

**Global ocean dynamical regions: Robust regimes in barotropic vorticity**

**19 March 2018**

Maike Sonnewald\*, Carl Wunsch

*Massachusetts Institute of Technology, 77 Massachusetts Ave, Cambridge, MA 02139, USA,*

*Harvard University, 26 Oxford Street, Cambridge, MA 02138, USA and The University of Texas*

*at Austin, 201 East 24th Street, Austin, TX 78712, USA*

& Patrick Heimbach

*The University of Texas at Austin, 201 East 24th Street, Austin, TX 78712, USA and*

*Massachusetts Institute of Technology, 77 Massachusetts Ave, Cambridge, MA 02139, USA*

\*Corresponding author address: Maike Sonnewald, Massachusetts Institute of Technology, 77 Massachusetts Ave, Cambridge, MA 02139, USA.

E-mail: maike\_s@mit.edu

## ABSTRACT

13 Global ocean dynamical regions are identified using a barotropic vorticity  
14 framework for a twenty year mean from the ECCO State Estimate. The  
15 unsupervised learning algorithm K-Means demonstrates five unambiguous  
16 and basin independent global regimes. The closure of the barotropic vorticity  
17 budget allows the classification and powerful dynamical interpretation.

18 The bottom pressure torque term is largely comparable to the Coriolis and  
19 stress terms, with dynamically important exceptions where the non-linear  
20 terms are significant. The “Momentum dominated” area covers  $56.7 \pm 1.8\%$ .

21 The surface and bottom stress terms are balanced by the bottom pressure  
22 torque and the non-linear terms. The “Transition zone” covers  $18.4 \pm 0.7\%$ ,  
23 and is found mainly in highly baroclinic subpolar regions. The surface  
24 and bottom stress is largely balanced by the Coriolis term and the bottom  
25 pressure torque. The “Subtropical gyres” cover  $10.9 \pm 0.5\%$ , characterized  
26 by a “quasi-Sverdrupian” regime where the Coriolis term is balanced by  
27 the wind and bottom stress term. The “Subpolar gyre” region characterised  
28 by baroclinic dynamics covers  $5.0 \pm 1.9\%$ , and is a muted version of the  
29 “Transition zone”. A similar gyre dynamic is seen in the “Southern Ocean  
30 gyre” covering  $1.4 \pm 0.7\%$ . The shift from the subpolar through the transition  
31 zone to the subtropical gyres is a shift from strong interior flow to one  
32 dominated by e.g. the western boundary currents. The remaining  $5.7 \pm 0.2\%$   
33 area is “Dominantly non-linear”, and found in key areas in the Southern  
34 ocean and along western boundaries.

<sup>36</sup> **1. Motivation**

<sup>37</sup> The global ocean is characterized using unsupervised learning classification on the depth-  
<sup>38</sup> integrated, barotropic vorticity (BV) equation. The adequacy of the BV framework is assessed  
<sup>39</sup> for decadal timescales. Characterizing the global ocean and reducing the overwhelming dimen-  
<sup>40</sup> sionality of observed dynamics has a long history in oceanography. Using theoretical conjectures,  
<sup>41</sup> Stommel (1948) and Munk (1950) assume topography is not important in the subtropics, while  
<sup>42</sup> topography has been key in understanding Southern Ocean dynamics (Munk and Palmén 1951).  
<sup>43</sup> Lorenz’s work from 1963 and 1987 highlighted the need for an “Occam’s razor” approach to  
<sup>44</sup> model development; resisting over-complicating descriptions of dynamics, but ensuring that  
<sup>45</sup> key components are present. Using theoretical conjectures and numerical models, it is not  
<sup>46</sup> obvious what level of complication is merited or if a satisfying global description is plausible.  
<sup>47</sup> General Circulation Models (GCM) solve the primitive equations in full detail, and using this the  
<sup>48</sup> success of simple characterizations of ocean features can be assessed. Using the BV equation,  
<sup>49</sup> the importance of non-linear terms illustrates how appropriate this highly simplified view is. A  
<sup>50</sup> simple interpretation of global ocean dynamics allows an intuitive interpretation of an otherwise  
<sup>51</sup> intractable system. The need for the complicated GCM approach can be assessed, and its  
<sup>52</sup> success and failure understood in terms of captured key dynamics, rather than overly generalized  
<sup>53</sup> quantities of interest such as the global overturning circulation. The existence of robust dynamical  
<sup>54</sup> regimes presented indicates what progress is key for predictive efforts, as it elucidates what  
<sup>55</sup> circulation features are key, and directly what forces are active to allow their emergence.

<sup>56</sup>

<sup>57</sup> The success of work such as Munk (1950) relies on a depth-integrated view, where lateral  
<sup>58</sup> friction provides closure in what is often referred to as Munk boundary layer structure for Western

59 Boundary Currents. A quasi-Sverdrupian framework is invoked, where the depth-integrated  
60 meridional flow is given by the local wind stress curl. The potential success of such frame-  
61 works suggests predictive skill based purely on the wind field, and suggests that a barotropic,  
62 depth-integrated view, is appropriate to characterize the subtropical ocean. GCM studies and  
63 observational work present a more intricate ocean, where baroclinic structure and subsequent  
64 non-linearities are key to ocean dynamics. If this is the case for large regions of the ocean, or  
65 if they govern key dynamics on relevant timescales, the simplified theoretical view would be  
66 inadequate.

67

68 The presence of dynamical regimes is suggested by the zonal nature of the wind stress forcing  
69 and the geometry of the ocean basins with underlying topography. Classifying and identifying  
70 regions in the world ocean is done using the BV equation, closing the global budget. Yeager  
71 (2015); Schoonover *et al.* (2016); Le Bras *et al.* (in review) have assessed the dynamics of the BV  
72 budget focusing on the North Atlantic. The link to the overturning and gyre circulation, as well as  
73 to the importance of mechanisms such as buoyancy forcing have been assessed. A global view  
74 allows the assessment of their conjectures and implications for dynamics in a global sense, which  
75 illuminates similarities and differences in governing dynamics.

76

77 Previous work assessing global ocean dynamics have been constrained to the surface. Xu and  
78 Fu (2012) used surface quasigeostrophic theory and sea surface height (SSH) data from altimeters  
79 to determine regimes governed by geostrophic turbulence. Major current systems are identified,  
80 but using altimeters the application to a comprehensive assessment of global dynamical regimes  
81 is limited. Hughes and Williams (2010); Sonnewald *et al.* (2018) use linear statistical models,  
82 finding global patterns in SSH. There global patterns are connected to circulation patterns and

83 planetary waves, and the statistical approach using autoregressive and moving average models is  
84 suggestive of global regimes. The work presented demonstrates this rigorously.

85

86 Using objective methods to identify patterns in data is commonly used in many fields ranging  
87 from pharmaceutical to engineering applications (). Applications to the earth system are being  
88 investigated both in the prognostic and diagnostic sense (Krasnopolksy *et al.* 2013; Schneider  
89 *et al.* 2017). To investigate global ocean dynamics, using objective methods adds credibility to  
90 global regions that confirm canonical ideas of e.g. gyre circulation regimes. Objective classifi-  
91 cation from machine learning has the potential to powerfully simplify the interpretation of ocean  
92 dynamics. The objective aspect allows an unbiased assessment of the success of the theoretical  
93 framework used, and the dimensionality reduction offered by using a classifier is an exciting  
94 and novel approach in physical oceanography. Ardyna *et al.* (2017) applied a similar method  
95 to identify regions with distinct biological activity, and Liang *et al.* (in review) used K-Means  
96 clustering to identify key regions for data collection to build maps of Nitrate in the Southern Ocean.

97

98 In section 2, the ECCOv4r2 State Estimate and the BV equation are introduced, and the details  
99 and considerations of applying the K-Means clustering are detailed. Section 3 presents the BV  
100 equation terms, highlighting the closure. The regions are presented, and the balance of the terms  
101 of the BV equation in these regions are assessed. Discussion and conclusions of the paper are in  
102 section 4. Further details and considerations of scaling and further interest in terms of robustness  
103 of the K-Means clustering algorithm are presented in appendix A1.

104 **2. Methods**

105 *a. ECCOv4 State Estimate*

106 The BV equation is used to assess the extent to which we can close this budget in the ECCOv4  
107 global bi-decadal state estimate, as described by Wunsch and Heimbach (2013); Forget *et al.*  
108 (2015) and others (see also ECCO Consortium (2017a,b)). The state estimate is global  $1^{\circ}$  with  
109 tropical mesh refinement. A least-squares with Lagrange multipliers approach is used to obtain  
110 the state estimate. The result is a *free-running* version of the MIT General Circulation Model  
111 (MITgcm, Adcroft *et al.* (2004)) with adjusted input variables solving the primitive equations. In  
112 contrast to most “reanalysis” products, the ECCO oceanic state satisfies basic conservation laws  
113 for enthalpy, salt, volume, and momentum remaining largely within error estimates of a diverse  
114 set of global data (Wunsch and Heimbach 2007, 2013; Stammer *et al.* 2016). Regions without  
115 data are filled in a dynamically consistent way using the dynamics, avoiding the use of untested  
116 statistical hypotheses e.g., Reynolds *et al.* (2013).

117 *b. Barotropic Vorticity*

118 To classify the ocean’s dynamical regions the BV equation is used, based on the solutions to  
119 the primitive equations from the ECCOv4 State Estimate. Illustrating the simplification, take the  
120 momentum equations of the ocean as a thin fluid on a rotating sphere:

$$\partial_t \mathbf{u} + f \mathbf{k} \times \mathbf{u} = -\frac{1}{\rho_0} \nabla p + \frac{1}{\rho_0} \partial_z \boldsymbol{\tau} + \mathbf{a} + \mathbf{b}, \partial_z p = -g\rho, \nabla \cdot \mathbf{v} = 0. \quad (1)$$

121 The pressure, acceleration due to gravity, density and vertical shear stress are denoted  $p$ ,  $g$ ,  $\rho$   
122 and  $\boldsymbol{\tau}$  respectively, with  $\rho_0$  the reference density; the three dimensional velocity field  $\mathbf{v} = (\mathbf{u},$   
123  $\mathbf{v}, \mathbf{w}) = (\mathbf{u}, \mathbf{w})$ ; the unit vector is denoted  $\mathbf{k}$ ; planetary vorticity as a function of latitude  $\phi$  in

<sub>124</sub>  $f\mathbf{k} = (0, 0, 2\Omega \sin\phi)$ ; the viscous forcing by vertical shear is denoted  $\partial_z \tau$ ; the non-linear terms  
<sub>125</sub> are  $\mathbf{a}$  and the horizontal viscous forcing  $\mathbf{b}$ . Assuming steady state, the vertical integral from the  
<sub>126</sub> surface  $z = \eta(x, y, t)$  to the depth below the surface  $z = H(x, y)$  gives

$$\beta V = \frac{1}{\rho_0} \nabla p_b \times \nabla H + \frac{1}{\rho_0} \nabla \times \tau + \nabla \times \mathbf{A} + \nabla \times \mathbf{B}, \quad (2)$$

<sub>127</sub> where  $\nabla \cdot \mathbf{U} = 0$ ,  $\mathbf{U} \cdot \nabla f = \beta V$ , the bottom pressure is denoted  $p_b$ ,  $\mathbf{A} = \int_H^\eta \mathbf{a} dz$  and  $\mathbf{B} = \int_H^\eta \mathbf{b} dz$ .  
<sub>128</sub> The LHS of equation (2) is the planetary vorticity advective term, while the RHS of equation (2) is  
<sub>129</sub> the bottom pressure torque, the wind and bottom stress curl, the non-linear torque and the viscous  
<sub>130</sub> torque, respectively. The non-linear torque is composed of three terms:

$$\nabla \times \mathbf{A} = \nabla \times \left[ \int_{-H}^0 \nabla \cdot (\mathbf{u} \mathbf{u}) dz \right] + [w \zeta]_{z=H}^{z=0} + [\nabla w \times \mathbf{u}]_{z=H}^{z=0}. \quad (3)$$

<sub>131</sub> The RHS on equation (3) represents the curl of the vertically integrated momentum flux diver-  
<sub>132</sub> gence, the non-linear contribution to vortex tube stretching and the conversion of vertical shear  
<sub>133</sub> to barotropic vorticity. The horizontal viscous forcing includes that induced by subgrid-scale  
<sub>134</sub> parameterizations. We work on the twenty year averaged fields of the BV equation after a  
<sub>135</sub> laplacian smoother is applied.

<sub>136</sub>

### <sub>137</sub> c. Unsupervised learning: K-Means clustering

<sub>138</sub> Assessing the presence of dominant global patterns in the BV equation the unsupervised learn-  
<sub>139</sub> ing algorithm K-Means clustering is used. The K-Means clustering method originates in signal  
<sub>140</sub> processing, and is used to determine if there are discernible patterns in the variability of the data.  
<sub>141</sub> It is a method of vector quantization, and associated analysis is used to confirm the robustness of  
<sub>142</sub> the clustering algorithm. Patterns of variability constitute areas in the function space, given by

143 the terms of the BV equation, that are more densely populated than others. If clusters are present  
 144 that are significantly different from each other, these can be robustly identified and are called  
 145 “clusters”. The clustering of configurations of terms and associated variability is identified by the  
 146 unsupervised learning K-Means algorithm. The data was scaled to have zero mean and unit vari-  
 147 ance. Referred to as an unsupervised learning algorithm (MacQueen 1967), a minimization of the  
 148 sum of squares (i.e variance) given by the partitioning of the hyperspace is iteratively approached:

$$J = \sum_{j=1}^K \sum_{i=1}^n \|x_i^j - c_j\|^2 \quad (4)$$

149 where the number of K ( $c_j$ ) is a parameter fixed a priori, randomly scattered among the  $n$  data  
 150 points ( $x$ ). The distance between a data point  $x_i^j$  and  $c_j$  is given by  $\|x_i^j - c_j\|^2$ . Initially, each  
 151 data point is associated with the closest K cluster. The position of  $c_j$  is recalculated, and the  
 152 association reassessed. This is iteratively repeated. The problem is computationally NP-hard,  
 153 and sensitive to the initialization and choice of K. As detailed in appendix A1, the appropriate  
 154 number of K is determined as  $K > 35$  using the Akaike and Bayesian Information Criteria (AIC  
 155 and BIC). The AIC and BIC indicate robust regimes asymptotic regimes, and we use  $K = 50$  for the  
 156 remaining analysis. Appendix A1 demonstrated the small sensitivity to the initial random seed of  
 157 the algorithm, demonstrating the high precision of the method.

158

### 159 3. Results

160 Figure 1 illustrates the closure in ECCOv4r2 for the 20 year average of the BV budget described  
 161 in equation (2). The residual is reconstructed and compared to the total temporal tendency  
 162 calculate at runtime, and is  $<< 1\%$  for 36% of the ocean, which is very small; Figure 1 illustrates  
 163 that the residual is largely smaller than  $\pm 10^{-12} ms^{-1}$ , while most terms are on the order of

<sub>164</sub>  $\pm 10^{-9} ms^{-1}$  in the BV equation. The relative size of the residual is demonstrated comparing  
<sub>165</sub> the magnitude of terms to the residual in the following analysis. The small size of the residual  
<sub>166</sub> confirms that the relevant terms are collected and assembled to give an accurate representation of  
<sub>167</sub> the dynamics. On this basis, we proceed with further analysis with confidence. Some numerical  
<sub>168</sub> issues on the shelf/shallow water have larger residuals, but these only amount to 2.9% of the  
<sub>169</sub> global ocean and are attributed to noise.

<sub>170</sub>

<sub>171</sub> Figure 1b illustrates the Coriolis term in ECCOv4r2 for the 20 year average from equation (2).  
<sub>172</sub> This is balanced well by the wind and bottom stress BV terms shown in Figure 1d (the bottom  
<sub>173</sub> stress term is small) and bottom pressure torque shown in Figure 1c. The remainder is largely  
<sub>174</sub> found in the non-linear BV contributions seen in Figure 1e, with the lateral viscous dissipation  
<sub>175</sub> largely being an order of magnitude smaller, apart from in localized regions in the Southern Ocean.

<sub>176</sub>

<sub>177</sub> The wind and bottom stress BV term in Figure 1d are largely zonally symmetric in the 20  
<sub>178</sub> year average, with large patterns of negative BV to the south in the Southern Ocean, large gyre  
<sub>179</sub> patterns visible in the Pacific and Atlantic basins. The bottom pressure torque term in Figure  
<sub>180</sub> 1c is associated with interactions with steep bathymetry, e.g. in the Southern Ocean we see a  
<sub>181</sub> large positive patch leading up to the Antarctic-Pacific ridge, and a negative patch following.  
<sub>182</sub> This is consistent with vortex stretching as the ridge is crossed, and we see a similar feature  
<sub>183</sub> associated with the Mid-Atlantic ridge in the Atlantic sector of the Southern Ocean. Along  
<sub>184</sub> Western Boundaries, the BPT is positive to the west and negative just adjacent to the east,  
<sub>185</sub> consistent with studies such as Myers *et al.* (1996). The BV of the non-linear terms is similarly  
<sub>186</sub> concentrated along the western edge of basins where western boundary currents are found, but  
<sub>187</sub> is less spatially coherent than the BPT term. The Southern Ocean stands out as a region of high

188 activity, particularly the Atlantic sector. The contribution of the lateral viscous dissipation to the  
189 BV is comparatively small, but also concentrated in the Southern Ocean.

190

191 Figure 2 illustrates the spatial extent of the dynamical regimes picked out by the K-Means  
192 algorithm. The numbering is arbitrary, and the variability is mainly found in 5 geographically  
193 coherent patches. There is a zonal pattern suggesting the wind stress contribution to the BV is  
194 key, but also much spatial structure in the Southern Ocean and along the western boundaries.  
195 Figures 3 and 4 isolate the geographical area and the associated area averaged magnitudes of the  
196 terms in the BV equation, ordered in decreasing percentage coverage of the global ocean. The bar  
197 charts represent the area averaged balance of the terms in the BV equation from the spatial areas  
198 determined as distinct by the K-Means clustering. Table 1 summarizes the findings, listing corre-  
199 sponding figure and dominant terms in the BV equation found in the K-Means determined clusters.

200

201 Figure 3a illustrates the region dominated by a balance of the surface and bottom stress terms  
202 covering 56% of the global ocean. This suggests a barotropic regime dominated by momentum.  
203 The momentum dominated region is found in zonal streaks in the Tropics, a thin ribbon in the  
204 Southern Ocean notably in the Pacific sector. In the Northern Hemisphere, the areas outline the  
205 edges of the subtropical and subpolar gyres. Large areas of the Arctic are also covered. Figure  
206 3b demonstrates the balance of terms that made the K-Means algorithm pick out these regions.  
207 The wind and bottom stress term is the major source of negative barotropic vorticity, while the  
208 BPT adds positive vorticity. The non-linear terms add negative vorticity. Figure 3c illustrates  
209 the spatial region covering the next dynamical region that covers 18% of the ocean area. In  
210 the Northern Hemisphere, these cover the southern region of the subpolar gyres. There is a  
211 zonal streak across the equator in both the Atlantic and Pacific, absent in the Indian Ocean. The

212 Southern Hemisphere show large expanses in both the Pacific and Atlantic, but the are covered  
213 in the Indian ocean is limited. The bar chart in Figure 3d highlight the wind as the major source  
214 of barotropic vorticity, with sinks in the Coriolis term and BPT. There is a small sink in the  
215 non-linear terms. Figure 3e illustrates the 11% of the ocean area selected by the next cluster. The  
216 subtropical gyres in the Northern Hemisphere Atlantic and Pacific stand out, together with thin  
217 streaks on the Equator. Isolated streaks are seen in the Southern Ocean, and a large area of the  
218 Southern Hemisphere Indian Ocean in the tropics. Figure 3f illustrates that these areas have a  
219 clear balance between the input of barotropic vorticity from the Coriolis term and a sink in the  
220 wind and bottom stress. This corresponds to a quasi-Sverdrupian regime in that we look at the  
221 depth integrated ocean. Figure 4a shows the area covered by the cluster covering 7% of the ocean.  
222 This is largely a complimentary poleward extension to the cluster covering 18% of the ocean seen  
223 in Figure 3c. In the Northern Hemisphere, the cluster largely represents the Northern edge of the  
224 subpolar gyre. In the Southern Hemisphere, the cluster is found on the Eastern edge of the Pacific  
225 and Atlantic basins, just to the south and flaring out westwards of the continental barrier. In the  
226 Indian Ocean, this barrier could be seen to be New Zealand or Australia, and the area filled by this  
227 dynamical regime fills the subtropical Indian Ocean down to the border with the Southern Ocean.  
228 This regime is absent from the Southern Ocean. The bar chart in Figure 4b illustrates that it is an  
229 amplified version of the barcharts seen in figure 3d, being an order of magnitude larger, but still  
230 having the wind as the major source of barotropic vorticity, with sinks in the Coriolis term and  
231 BPT. There is a small source in the non-linear terms. The Southern Ocean is better represented in  
232 the area covering 2% of the global world ocean seen in Figure 4c. This is seen mainly in a series  
233 of streaks in the Southern Ocean, mainly staying around 60°S where there is no continental block.  
234 Isolated areas are also seen in the Northern Hemisphere. The bar charts in Figure 4d illustrate that  
235 these are practically an inverse of the barcharts for the area representing the Northern extension of

236 the subtropical gyres covering 7% of the world ocean. The wind as the major sink of barotropic  
237 vorticity, with sources in the Coriolis term and BPT. There is a small sink in the non-linear terms.

238

239 The final map in Figure 4e is a summary of the areas terms that account for > 0.2%, these sum  
240 to 6.3% of the world ocean, and individual plots can be seen in appendix A1. These are mainly  
241 found in the Southern Ocean and along the Western Boundaries. Areas of rough bathymetry  
242 stand out, such as the Pacific-Antarctic Ridge and the Drake Passage area. The bar chart shown in  
243 Figure 4f illustrates that the non-linear contribution to the barotropic vorticity stands out, together  
244 with the Coriolis term. The different constituents are quite varied, but strong contributions from  
245 the non-linear terms are consistent. Unpicking these specifically is outside the current scope, but  
246 will be discussed in a following paper.

247

#### 248 **4. Discussion and Conclusions**

249 The BV equation is found to be a reasonable representation of large areas of the world ocean.  
250 Figure 1 demonstrated that the global ocean has large coherent regions, but that the dynamical  
251 balances implicit in the 20 year average are more nuanced than a balance between the wind  
252 stress, Coriolis and BPT terms. Areas where the non-linear terms are small suggest that the BV  
253 is a good approximation. Areas where the non-linear terms are important are found in western  
254 boundary regions, as well as the Southern Ocean where the ACC interacts with bathymetric  
255 obstacles. Figure 5 summarizes the identified regions. The momentum dominated area implies  
256 a coherent vertical structure. The subtropical gyre is unique due to lack of BPT. The transition  
257 zone has a stronger momentum driven portion of the BPT, and topographic interactions begin to  
258 become important. The subpolar gyre has a stronger baroclinic component to the BPT and feels

259 topography. The Southern Ocean gyre is like the subpolar, but with contributions of opposite sign.  
260 The remainder is dominated by non-linear contributions, and the barotropic interpretation is not  
261 appropriate.

262

263 Assessing the regimes that can be identified using altimeter data (Xu and Fu 2012; Hughes  
264 and Williams 2010) is limited to the surface, and identifying circulation patterns that leave a  
265 discernible surface imprint. Circulation patterns that have compensating effects are not discernible  
266 using surface observations. The classification based on the closure of the BV equation is based on  
267 the full primitive equations and adjusted to available data. Using the BV equations as our baseline  
268 constrains the dimensionality of our search space and facilitates the dynamical interpretation of  
269 the results. This allows a comprehensive assessment of the presence of dynamical regimes.

270

271 Yeager (2015) describes the North Atlantic circulation in terms of its BV equation breakdown.  
272 Following this, we interpret the regions in the North Atlantic as a shift where strong interior flow  
273 (baroclinic meridional, North Atlantic Current, NAC, and North Atlantic deep water, NADW,  
274 flow over the Mid Atlantic Ridge) transitions to more barotropic flow (deep western boundary  
275 current, DWBC). The geographical extent and transition between the 18.4% (Transition zone)  
276 and 110.9% (subtropical) areas make sense in light of this, particularly as they are mirrors of  
277 each other, but with the difference in wind sign and BPT interactions. The 5.0% (Subpolar gyre)  
278 region complements the subtropical region, being an intensified extension and both with the  
279 addition of small contributions of non-linear terms. Extrapolating this to the global picture, this  
280 suggests that strong interior flow is present in vast expanses of the Southern Hemisphere, as well  
281 as the North Pacific. The quasi-Sverdrupian regime we see in the Northern Hemisphere subtrop-  
282 ics is virtually absent in the South Pacific, suggesting the absence of a strictly barotropic flow

283 structure. This is found in the Southern and Indian Ocean, indicating that the wind field dominates.

284

285 The relative size of the areas classified should be considered in terms of the scale of dynamics  
286 present in the region and their impact on the global circulation. An area in the Southern Ocean  
287 can have a small spatial extent, but significant impact on the global overturning. The areas that  
288 are collated under the label “Dominantly non-linear” have a small spatial extent, but are found in  
289 regions that are known to be important for ocean dynamics. The Drake passage region as well as  
290 the Antarctic-Pacific ridge are areas where the circulation interacts with topography, and these  
291 regions are highlighted as regions where the non-linear effects are large. Similarly in the Northern  
292 Hemisphere, areas in the Labrador sea and shelf stand out.

293

294 The sign and spatial distribution of the wind stress term suggests the importance of Ekman  
295 pumping(negative)/suction(positive). The Equatorial and Southern Ocean regions have Ekman  
296 pumping, while the subpolar gyre areas have Ekman suction where modewaters are created. The  
297 BPT term mirrors the wind stress term, suggesting it acts as either a source or a sink in opposite  
298 compliment to the wind stress. A full description of the dynamics is outside the scope of the  
299 current study.

300

301 The area that is dominantly non-linear illustrates that key ocean regions may not conform to  
302 the simplification presented by the twenty year averaged BV equation. This region is a collection  
303 of smaller terms. Interactions with topographic barriers, as well as WBC regions stand out,  
304 suggesting that the vertically integrated momentum flux divergence, the non-linear contribution to  
305 vortex tube stretching and the conversion of vertical shear to barotropic vorticity are active and  
306 key to the dominant dynamics. A detailed description of balances present here will be the subject

307 of a future study.

308

309 The assumed variance structure by the K-Means clustering could mean a more computationally  
310 efficient solution is available by exploring other methods such as the more generalised gaussian  
311 mixture models. Doing this other covariance shapes can be explored, and we present initial results  
312 in figure 8. This was not found to impact results, but will likely be key for a robust scaling of the  
313 method to higher resolution data.

314

315 The present study found that the BV budget closes in the ECCOv4r2 State Estimate over  
316 twenty years. The balance observed deviated from a straightforward interpretation of the three  
317 dominant terms as the wind, Coriolis and BPT terms, with significant covariance and input from  
318 the non-linear terms in dynamically key regions. These regions were identified using the K-Means  
319 clustering method. Large areas of the ocean are well described by the BV equations. Five regions  
320 were identified covering 93% of the world ocean: momentum dominated (57%), transition zone  
321 (18%), subtropical gyre (11%), subpolar gyre (5%) and the Southern Ocean gyre (1.4%). The  
322 final area is dominated by the non-linear terms and will be the subject of a future study.

323

324 **5. Acknowledgments**

325 This work was funded by the US National Aeronautics and Space Administration Sea Level  
326 Change Team (contract NNX14AJ51G) and through the ECCO Consortium funding via the Jet  
327 Propulsion Laboratory. This work would not have been possible without the support of Anne  
328 Reinarz, Edward Doddridge, Roosa Tikkainen and Katherine Rosenfeld.

329 **APPENDIX**

330 **K-Means and influence of Information Criteria**

331 **A1. K-Means and influence of Information Criteria and covariance assumptions**

332 The robustness of the unsupervised learning algorithm K-Means is demonstrated and potential  
333 improvements discussed. The K-Means algorithm is initiated by scattering  $K$  guesses of where  
334 the parameters/clusters could be. This initial guess introduces a stochastic element. The success  
335 of the algorithm is sensitive to  $K$ , as this determines how the hyperspace given by the dimentions  
336 is partitioned. As with regression analysis, adding parameters can increase the accuracy, but over-  
337 fitting should be avioded. To determine the appropriate number ok  $K$ , the Akaike and Bayesian  
338 Information Criteria (AIC and BIC) are used; Minimizing the expectation of the prediction error:

$$\text{AIC} = 2K - 2\ln(\mathcal{L}),$$

$$\text{BIC} = K\ln(n) - 2\ln(\mathcal{L}),$$

339 where  $n$  is the number of datapoints and  $\mathcal{L}$  is the likelihood:

$$\mathcal{L} = \prod_{i=1}^N \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\sum_{i=1}^N \frac{(\zeta_i - \hat{\zeta}_i)^2}{2\sigma^2}\right).$$

340  $\zeta_i$  is the observed, and  $\hat{\zeta}_i$  is the prediction, so  $(\zeta_i - \hat{\zeta}_i)^2$  are the prediction residuals. In the  
341 estimate, the AIC value is minimized, which determines the smallest appropriate order to represent  
342 the time-series. As discussed by Priestley (1981) and Yang (2005), the AIC can overestimate the  
343 order. Figure 6 demonstrates that both the AIC and BIC stabilise at  $> 35K$ , and the asymptotic  
344 nature of the regime.

345

346 How robust the regions in terms of the stochastic initialisation is highlighted in Figure 7,  
347 where the K-Means clustering was run 100 times. The mean and  $2\sigma$  are used in Table 1. This  
348 demonstrates that the regimes identified are robust, with the extent of the Subpolar gyre being the  
349 main area where the algorithm shows appreciable variance.

350

351 To elucidate the impact of assumptions the algorithm makes for the classification, a more  
352 generalised form of clustering is also tested: Gaussian Mixture Models. These models are used  
353 to assess the impact of assumptions relating to the covariance structure; spherical, diagonal, tied  
354 or full covariance are assessed as seen in Figure 8. Using the BIC to assess the impact, this is  
355 found to likely not impact results, but could be key for scaling up the method for high resolution  
356 applications where computational efficiency becomes more important as the K-Means clustering  
357 problem is NP-hard.

358

## 359 References

360 Adcroft, A., Hill, C., Campin, J. M., Marshall, J., & Heimbach, P. (2004). Overview of the formulation and numerics of the MIT GCM (pp.

- 361 139-150). Presented at the ECMWF Conference Proceedings, Shinfield Park, Reading, UK.
- 362 Akaike, H. (1973), "Information theory and an extension of the maximum likelihood principle", in Petrov, B.N. ; Csiki, F., 2nd International  
363 Symposium on Information Theory, Tsahkadsor, Armenia, USSR, September 2-8, 1971, Budapest: Akademiai Kiad, p. 267-281.
- 364 Ardyna, M., H. Claustre, J. Sallee, F. DOvidio, B. Gentili, G. van Dijken, F. DOrtenzio, and K. R. Arrigo, 2017: Delineating environmental control  
365 of phytoplankton biomass and phenology in the southern ocean. *Geophys. Res. Lett.*, 44, 50165024, doi:10.1002/2016GL072428.
- 366 Church, J.A., et al., 2013: Sea Level Change. In: Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the  
367 Fifth Assessment Report of the Intergovernmental Panel on Climate Change [Stocker, T.F., D. Qin, G.-K. Plattner, M. Tignor, S.K. Allen, J.  
368 Boschung, A. Nauels, Y. Xia, V. Bex and P.M. Midgley (eds.)], Chapter 13, pp. 1137-1216, Cambridge University Press.
- 369 ECCO Consortium, 2017a, A Twenty-Year Dynamical Oceanic Climatology: 1994-2013. Part 1: Active Scalar Fields: Temperature, Salinity,  
370 Dynamic Topography, Mixed-Layer Depth, Bottom Pressure.
- 371 ECCO Consortium, 2017b, A Twenty-Year Dynamical Oceanic Climatology: 1994-2013. Part 2: Velocities, Property Transports, Meteorological  
372 Variables, Mixing Coefficients.
- 373 Forget, G., J.-M. Campin, P. Heimbach, C. N. Hill, R. M Ponte, and C. Wunsch, ECCO version 4: an integrated framework for non-linear inverse  
374 modeling and global ocean state estimation, *Geo. Sci. Model Dev.*, 8, 2015
- 375 Hirschi, J. J.-M., Blaker, A. T., Sinha, B., Coward, A., de Cuevas, B., Alderson, S., and Madec, G.: Chaotic variability of the meridional overturning  
376 circulation on subannual to interannual timescales, *Ocean Sci.*, 9, 805-823, doi:10.5194/os-9-805-2013, 2013.
- 377 Chris W. Hughes, Simon D. P. Williams, The color of sea level: Importance of spatial variations in spectral shape for assessing the significance of  
378 trends, *Journal of Geophysical Research*, 2010, 115, C10
- 379 Vladimir M. Krasnopol'sky, Michael S. Fox-Rabinovitz, and Alexei A. Belochitski, "Using Ensemble of Neural Networks to Learn Stochastic  
380 Convection Parameterizations for Climate and Numerical Weather Prediction Models from Data Simulated by a Cloud Resolving Model,"  
381 Advances in Artificial Neural Systems, vol. 2013, Article ID 485913, 13 pages, 2013. doi:10.1155/2013/485913
- 382 Le Bras, I., Toole, J.M., Sonnewald, M.: A bulk Potential Vorticity budget for the western North Atlantic based on observations. In review.
- 383 Liang, Y.-C. *et al.*, in review.
- 384 J. B. MacQueen (1967): "Some Methods for classification and Analysis of Multivariate Observations, Proceedings of 5-th Berkeley Symposium  
385 on Mathematical Statistics and Probability", Berkeley, University of California Press, 1:281-297.

- 386 Munk, W., 1950: On the wind-driven ocean circulation. *J. Meteor.*, 7, 7993, doi:10.1175/1520-0469(1950)007;0080:OTWDOC\_2.0.CO;2.
- 387 Munk, W.H., and Palmén, E.: Note on the dynamics of the Antarctic Circumpolar Current, *Tellus*, 3, 53-55, 1940.
- 388 Myers, Paul G., Augustus F. Fanning, Andrew J. Weaver, 1996: JEBAR, Bottom Pressure Torque, and Gulf Stream Separation. *J. Phys. Oceanogr.*,  
389 26, 671683. doi: [http://dx.doi.org/10.1175/1520-0485\(1996\)026<0671:JBPTAG>2.0.CO;2](http://dx.doi.org/10.1175/1520-0485(1996)026<0671:JBPTAG>2.0.CO;2)
- 390 Priestley, M. B. (1981) Spectral Analysis and Time Series, London: Academic Press.
- 391 Reynolds, R.W., D.B. Chelton, J. Roberts-Jones, M.J. Martin, D. Menemenlis, and C.J. Merchant, 2013: Objective Determination of Feature  
392 Resolution in Two Sea Surface Temperature Analyses. *J. Climate*, 26, 2514-2533, <https://doi.org/10.1175/JCLI-D-12-00787.1>
- 393 Schoonover, J., Dewar, W., Wienders, N., Gula, J., McWilliams, J.C., Molemaker, M.J., Bates, S.C., Danabasoglu, G. and Yeager, S.: North  
394 Atlantic Barotropic Vorticity Balances in Numerical Models, *Journal of Physical Oceanography* 2016 46:1, 289-303
- 395 Lan, S., Schneider, T., Stuartland, A., and Teixeira, J. (2017). Earth System Modeling 2.0: A Blueprint for Models That Learn From Observations  
396 and Targeted High-Resolution Simulations.
- 397 Sonnewald, M., C. Wunsch, and P. Heimbach, 2018: Linear Predictability: A Sea Surface Height Case Study. *J. Climate*, 31, 25992611,  
398 <https://doi.org/10.1175/JCLI-D-17-0142.1>
- 399 Stainforth, D., Aina, T., Christensen, C., Collins, M., Faull, N., Frame, D., Kettleborough, J., Knight, S., Martin, A., Murphy, J., Piani, C., Sexton,  
400 D., Smith, L., Spicer, R., Thorpe, A. and Allen, M.: 2005, Uncertainty in predictions of the climate response to rising levels of greenhouse  
401 gases, *Nature* 433(7024), 403-406. url:<http://oro.open.ac.uk/5025/>
- 402 D. Stammer, M. Balmaseda, P. Heimbach, A.Koehl, and A. Weaver, 2016: Ocean Data Assimilation in Support of Climate Applications: Status  
403 and Perspectives. *Ann. Rev. Mar. Sci.*, 8, 491-518.
- 404 Stommel, H. (1948), The westward intensification of wind-driven ocean currents, *Eos Trans. AGU*, 29(2), 202206, doi:10.1029/TR029i002p00202.
- 405 Yang, Y. (2005), "Can the strengths of AIC and BIC be shared?", *Biometrika*, 92: 937-950, doi:10.1093/biomet/92.4.937.
- 406 Yeager, S.: Topographic coupling of the Atlantic overturning and gyre circulations, *J. Phys. Oceanogr.*, 45, 1258-1284. doi: <http://dx.doi.org/10.1175/JPO-D-14-0100.1>, 2015.
- 408 Xu, Y. and L. Fu, 2012: The Effects of Altimeter Instrument Noise on the Estimation of the Wavenumber Spectrum of Sea Surface Height. *J. Phys.*  
409 *Oceanogr.*, 42, 22292233, <https://doi.org/10.1175/JPO-D-12-0106.1>
- 410 Wunsch, C., and P. Heimbach, 2007: Practical global oceanic state estimation. *Physica D*, 230, 197-208.

- 411 Wunsch, C. and P. Heimbach, 2013, Dynamically and kinematically consistent global ocean circulation and ice state estimates. In Ocean Circulation  
412 and Climate 2nd Edition, Siedler et al., Eds.
- 413 Wunsch, C. The interpretation of short climate records, with comments on the North Atlantic and Southern Oscillations. Bull. Am. Met. Soc. 80,  
414 245-255 (1999).
- 415 Wunsch, C. (2013), Covariances and linear predictability of the Atlantic Ocean, Deep Sea Res., Part II, 85, 228-243.
- 416 Wunsch, C. (2015), Modern Observational Physical Oceanography: Understanding the Global Ocean. Princeton University Press.

417 LIST OF TABLES

**Table 1.** Percentage of area covered by the area specific balance of the BV equation (2) and the corresponding map figure. Leading order terms are sorted by magnitude, colors indicating if barotropic vorticity is added (**red**) or removed (**blue**) by the leading order term, the corresponding bar chart figure shows the full breakdown. The quoted percentage coverage and StD is the mean of 100 runs of the algorithm.

Area	Leading terms
56.7±1.8%, Momentum dominated (Fig. 3a)	$\nabla \times \tau_{sb} + \nabla \times \mathbf{A} \approx \nabla p_b \times \nabla H$ (Fig. 3b)
18.4±0.7%, Transition zone (Fig. 3c)	$\nabla \times \tau_{sb} \approx \nabla p_b \times \nabla H + \nabla \cdot (\mathbf{fU})$ (Fig. 3d)
10.9±0.5%, Subtropical gyre (Fig. 3e)	$\nabla \times \tau_{sb} \approx \nabla \cdot (\mathbf{fU})$ (Fig. 3f)
5.0±1.9%, Subpolar gyre (Fig. 4a)	$\nabla \times \tau_{sb} \approx \nabla \cdot (\mathbf{fU}) + \nabla p_b \times \nabla H$ (Fig. 4b)
1.4±0.7%, Southern Ocean gyre (Fig. 4c)	$\nabla \times \tau_{sb} \approx \nabla \cdot (\mathbf{fU}) + \nabla p_b \times \nabla H$ (Fig. 4d)
5.7±0.2%, Dominantly non-linear (Fig. 4e)	$\nabla \cdot (\mathbf{fU}) \approx \nabla \times \mathbf{A} + \nabla \times \tau_{sb}$ (Fig. 4f)

TABLE 1: Percentage of area covered by the area specific balance of the BV equation (2) and the corresponding map figure. Leading order terms are sorted by magnitude, colors indicating if barotropic vorticity is added (**red**) or removed (**blue**) by the leading order term, the corresponding bar chart figure shows the full breakdown. The quoted percentage coverage and StD is the mean of 100 runs of the algorithm.

## 424 LIST OF FIGURES

425	<b>Fig. 1.</b> The breakdown of the barotropic vorticity budget ( $ms^{-1}$ ) over 1992-2013 in the 426 ECCOv4r2 State Estimate. . . . .	24
427	<b>Fig. 2.</b> The area selected by the clusters. The colors represent the clusters, and are in arbitrary order. 428 The momentum dominated ocean region (55.6%, Figure 3a) is in dark blue, the Transition 429 zone (17.6%, Figure 3c) is in light brown, the Subtropical Gyre (“Quasi-Sverdrup”, 11.2% 430 Figure 3e in light green, the Subpolar gyre (6.6%, Figure 4a) is in dark green, the Southern 431 Ocean gyre (2%, Figure 4c) in lighter blue and the remaining areas where the non-linear 432 terms are large amounting to 6.3% of the world ocean and are presented in Figure 4e.) . . . . .	25
433	<b>Fig. 3.</b> Maps of the selected locations (left) and corresponding area averaged his- 434 togram (right) of the terms in the BV equation. The colorbar is kept, but the 435 color/ordering of the map are arbitrary. Colors in the barchart indicate if BV is 436 added (red) or removed (blue). . . . .	26
437	<b>Fig. 4.</b> Maps of the selected locations (left) and corresponding area averaged his- 438 togram (right) of the terms in the BV equation. The colorbar is kept, but the 439 color/ordering of the map are arbitrary. Colors in the barchart indicate if BV is 440 added (red) or removed (blue). . . . .	27
441	<b>Fig. 5.</b> Schematic of identified regions. The momentum dominated area implies a coherent verti- 442 cal structure. The subtropical gyre is unique due to lack of BPT. The transition zone has 443 a stronger momentum driven portion of the BPT, and topographic interactions begin to be- 444 come important. The subpolar gyre has a stronger baroclinic component to the BPT and 445 feels topography. The Southern Ocean gyre is like the subpolar, but with contributions of 446 opposite sign. The remainder is dominated by non-linear contributions, and the barotropic 447 interpretation is not appropriate. . . . .	28
448	<b>Fig. 6.</b> The AIC and BIC asymptote and we choose a K of 50 for our analysis. Error bars represent 449 $2\sigma$ , capturing the stochastic start seen of the algorithm. . . . .	29
450	<b>Fig. 7.</b> The ocean area represented by the different areas from table 1, and 100 runs of the clas- 451 sification. Error bars represent $2\sigma$ , capturing the stochastic start seed of the classification 452 algorithm. The final point “Dominantly non-linear” represents the remainder in the 44 clus- 453 ters not shown. . . . .	30
454	<b>Fig. 8.</b> Trying different covariance models to check the convergence. . . . .	31

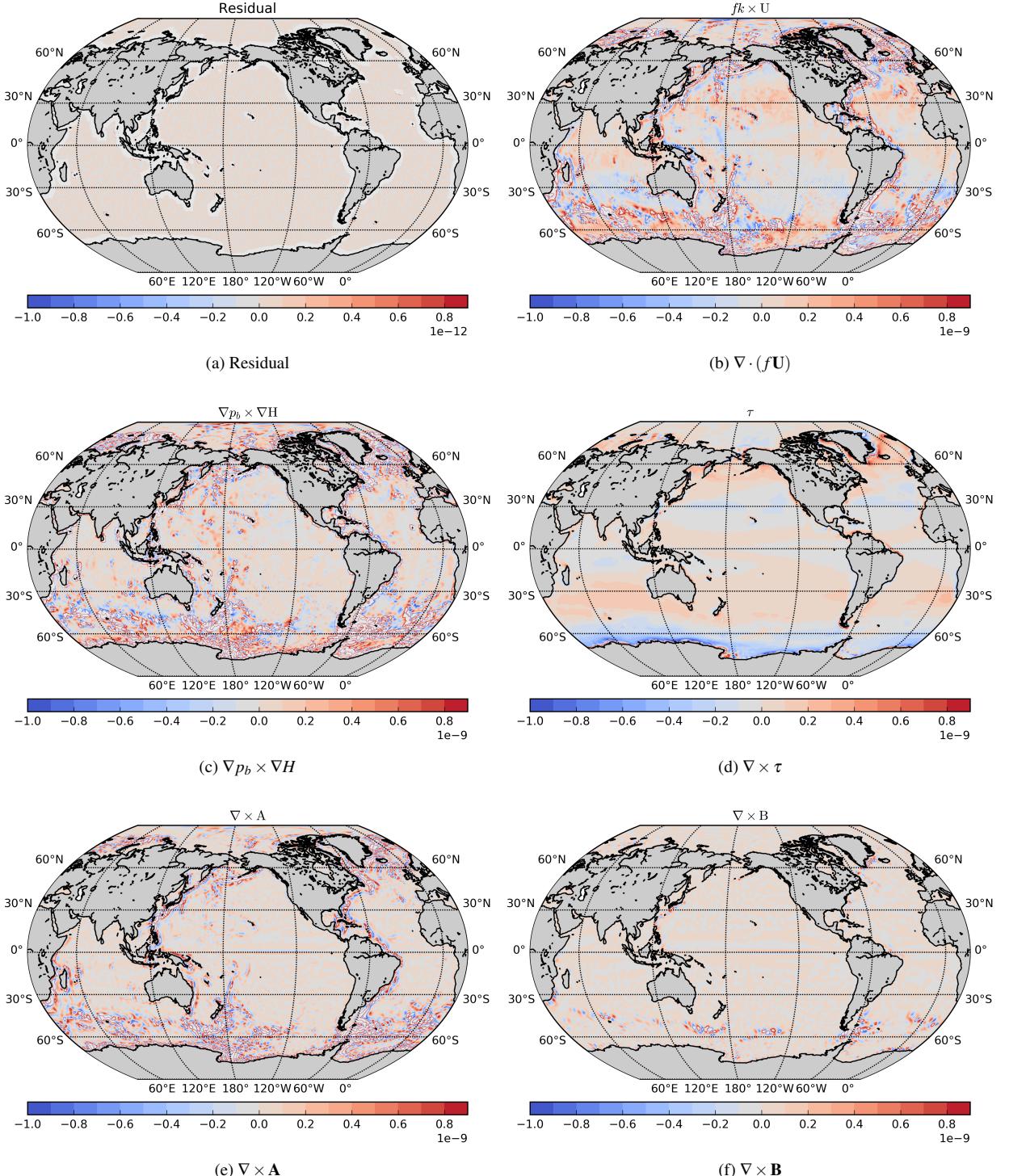


FIG. 1: The breakdown of the barotropic vorticity budget ( $m s^{-1}$ ) over 1992-2013 in the ECCOv4r2 State Estimate.

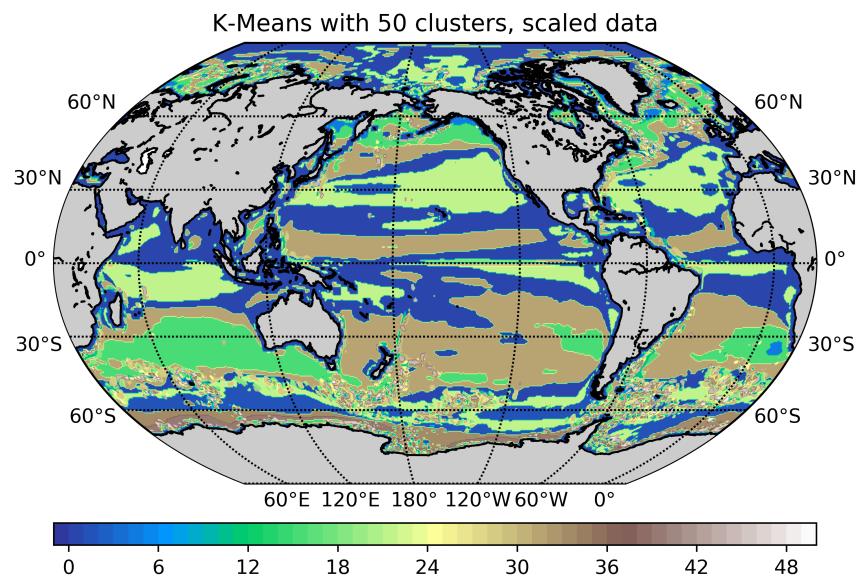
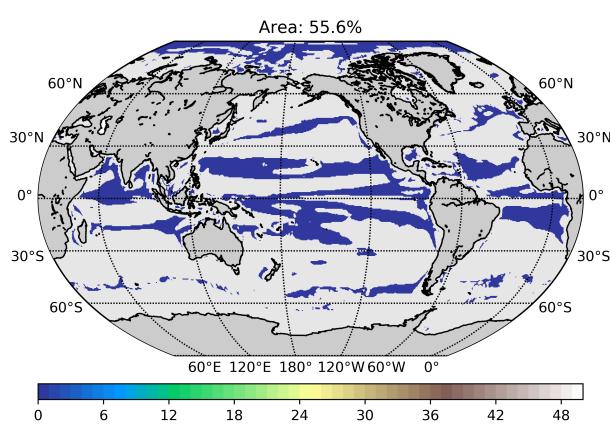
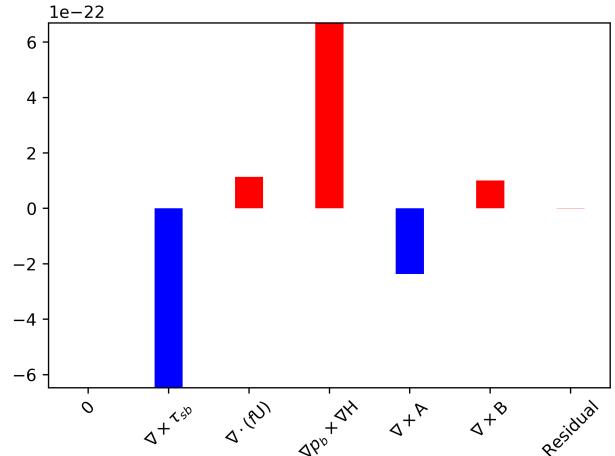


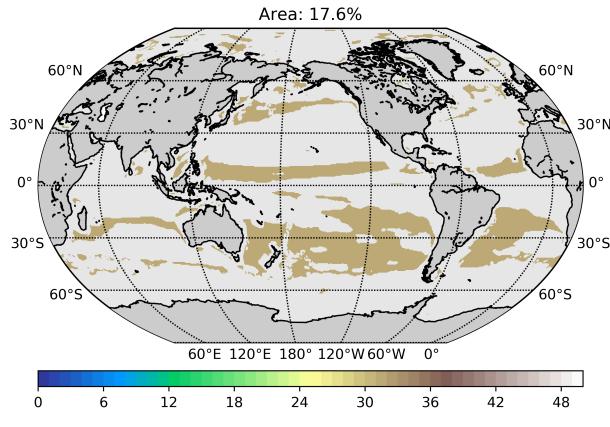
FIG. 2: The area selected by the clusters. The colors represent the clusters, and are in arbitrary order. The momentum dominated ocean region (55.6%, Figure 3a) is in dark blue, the Transition zone (17.6%, Figure 3c) is in light brown, the Subtropical Gyre (“Quasi-Sverdrup”, 11.2% Figure 3e in light green, the Subpolar gyre (6.6%, Figure 4a) is in dark green, the Southern Ocean gyre (2%, Figure 4c) in lighter blue and the remaining areas where the non-linear terms are large amounting to 6.3% of the world ocean and are presented in Figure 4e.)



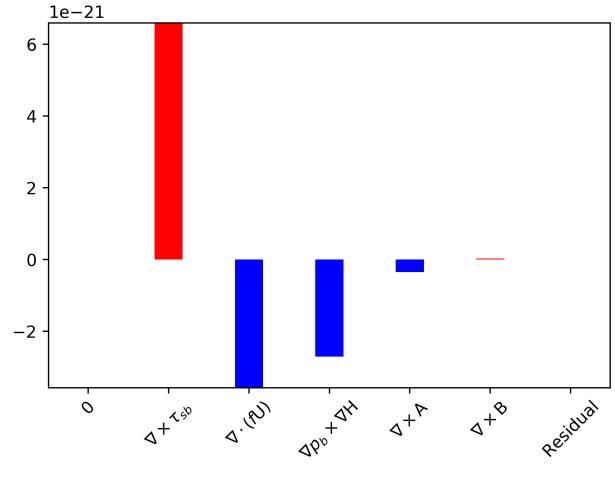
(a) Momentum dominated



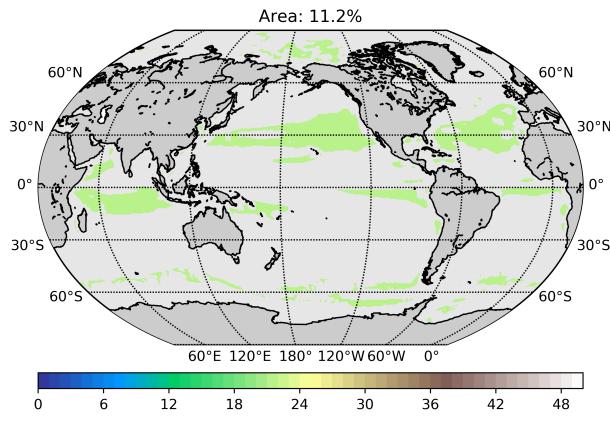
(b) Momentum dominated: Area averaged histogram



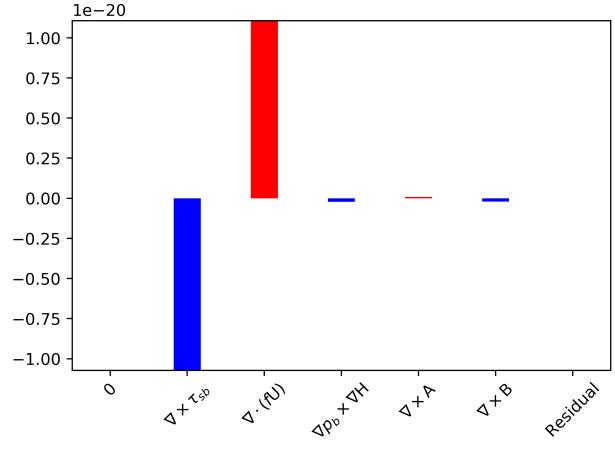
(c) Transition Zone



(d) Transition zone: Area averaged histogram

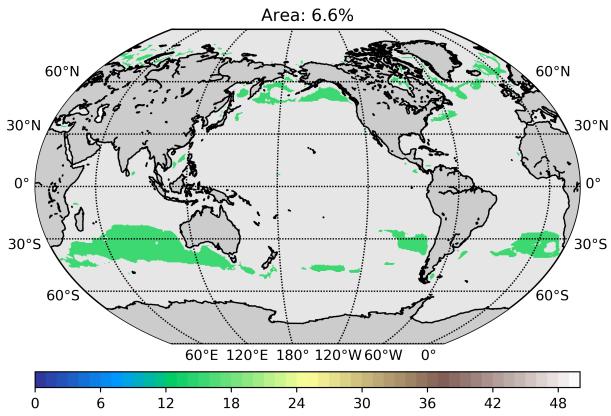


(e) Subtropical gyre

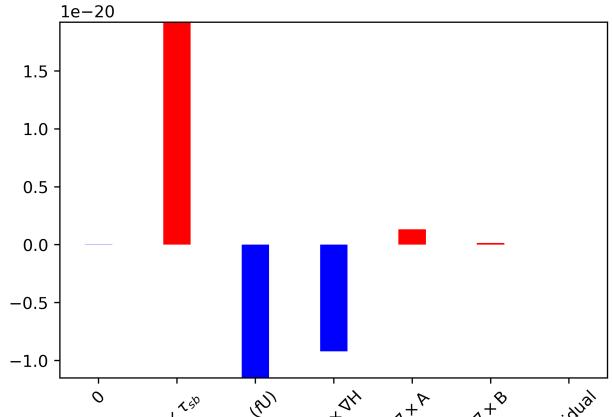


(f) Subtropical gyre: Area averaged histogram

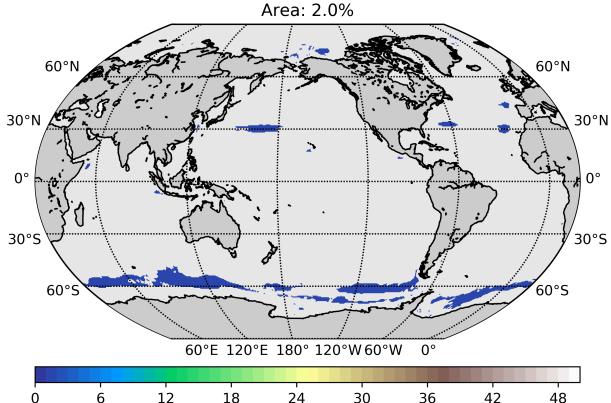
FIG. 3: Maps of the selected locations (left) and corresponding area averaged histogram (right) of the terms in the BV equation. The colorbar is kept, but the color/ordering of the map are arbitrary. Colors in the barchart indicate if BV is added (red) or removed (blue).



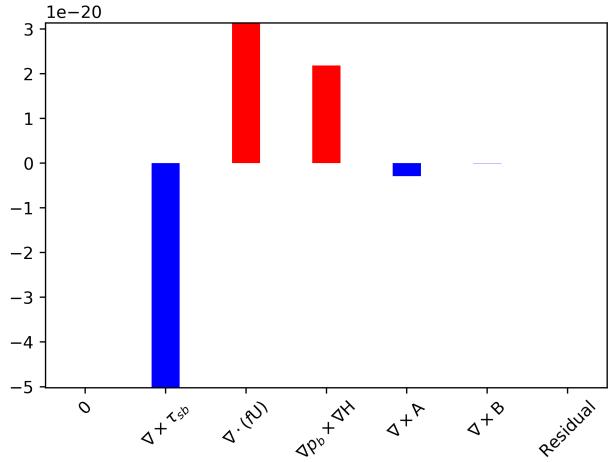
(a) Subpolar gyre



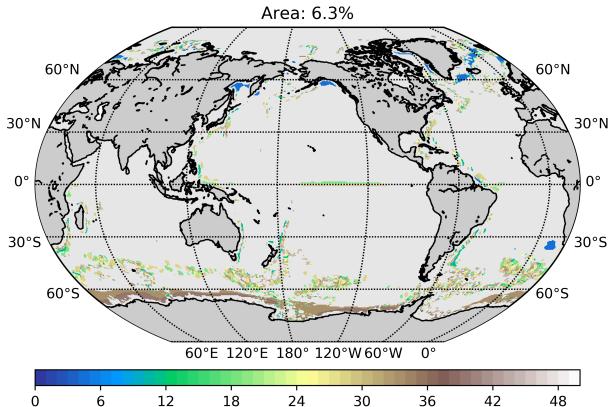
(b) Subpolar gyre: Area averaged histogram



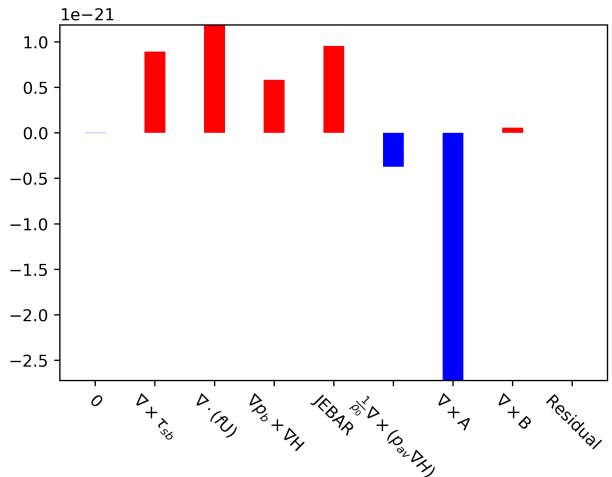
(c) Southern Ocean gyre



(d) Southern Ocean gyre: Area averaged histogram



(e) Dominantly non-linear



(f) Dominantly non-linear: Area averaged histogram

FIG. 4: Maps of the selected locations (left) and corresponding area averaged histogram (right) of the terms in the BV equation. The colorbar is kept, but the color/ordering of the map are arbitrary. Colors in the barchart indicate if BV is added (red) or removed (blue).

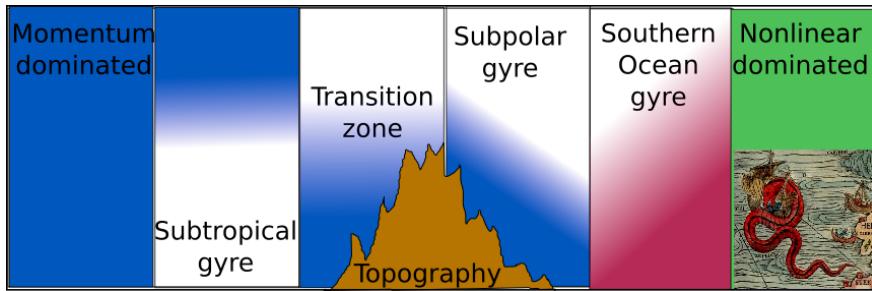


FIG. 5: Schematic of identified regions. The momentum dominated area implies a coherent vertical structure. The subtropical gyre is unique due to lack of BPT. The transition zone has a stronger momentum driven portion of the BPT, and topographic interactions begin to become important. The subpolar gyre has a stronger baroclinic component to the BPT and feels topography. The Southern Ocean gyre is like the subpolar, but with contributions of opposite sign. The remainder is dominated by non-linear contributions, and the barotropic interpretation is not appropriate.

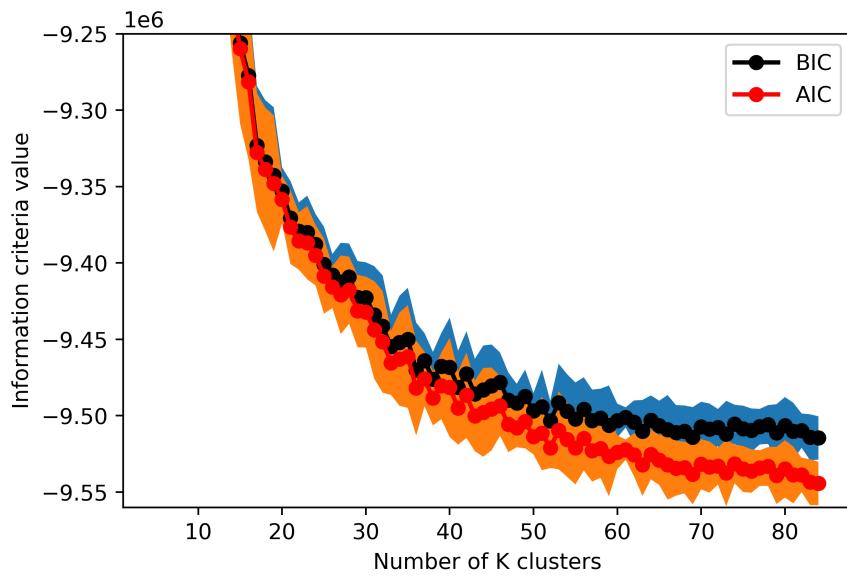


FIG. 6: The AIC and BIC asymptote and we choose a K of 50 for our analysis. Error bars represent  $2\sigma$ , capturing the stochastic start seen of the algorithm.

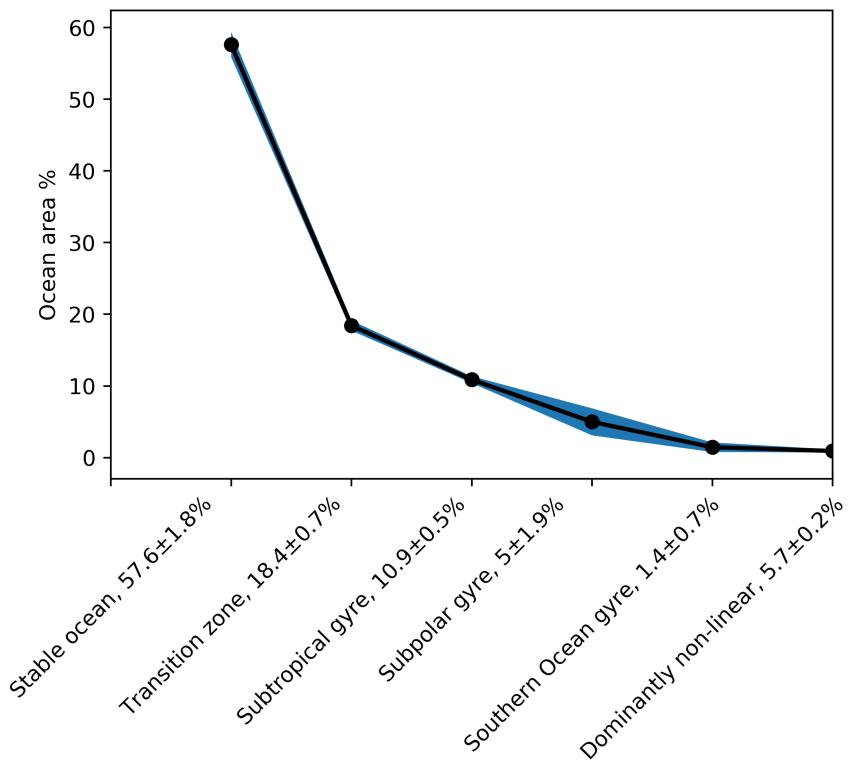


FIG. 7: The ocean area represented by the different areas from table 1, and 100 runs of the classification. Error bars represent  $2\sigma$ , capturing the stochastic start seed of the classification algorithm. The final point “Dominantly non-linear” represents the remainder in the 44 clusters not shown.

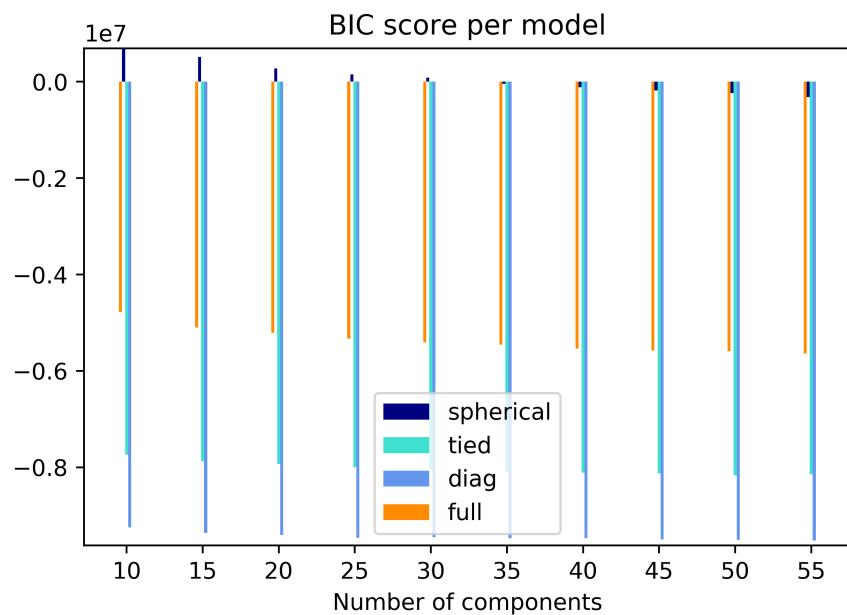


FIG. 8: Trying different covariance models to check the convergence.