



HAL
open science

Can We Reconstruct Mean and Eddy Fluxes from Argo Floats?

Christopher C. Chapman, Jean-Baptiste Sallée

► **To cite this version:**

Christopher C. Chapman, Jean-Baptiste Sallée. Can We Reconstruct Mean and Eddy Fluxes from Argo Floats?. Ocean Modelling, 2017, 120, pp.83-100. 10.1016/j.ocemod.2017.10.004 . hal-01618252

HAL Id: hal-01618252

<https://hal.sorbonne-universite.fr/hal-01618252>

Submitted on 17 Oct 2017

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.



Distributed under a Creative Commons Attribution - NonCommercial - NoDerivatives 4.0 International License

Accepted Manuscript

Can We Reconstruct Mean and Eddy Fluxes from Argo Floats?

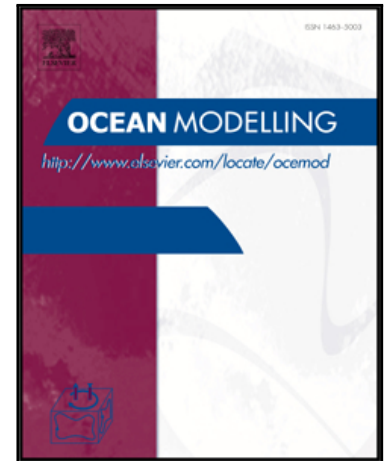
Christopher Chapman, Jean-Baptiste Sallée

PII: S1463-5003(17)30148-8
DOI: [10.1016/j.ocemod.2017.10.004](https://doi.org/10.1016/j.ocemod.2017.10.004)
Reference: OCEMOD 1251

To appear in: *Ocean Modelling*

Received date: 18 May 2017
Revised date: 13 October 2017
Accepted date: 15 October 2017

Please cite this article as: Christopher Chapman, Jean-Baptiste Sallée, Can We Reconstruct Mean and Eddy Fluxes from Argo Floats?, *Ocean Modelling* (2017), doi: [10.1016/j.ocemod.2017.10.004](https://doi.org/10.1016/j.ocemod.2017.10.004)



This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Highlights

- Virtual floats show that it is possible to reconstruct interior flow using Argo floats
- The reconstruction is sensitive to temporal sampling, number of floats and time span
- The cross-Polar Front heat flux is determined using Argo floats to be 2 p/m 0.5 TW
- Heat-flux is concentrated near large bathymetry

Can We Reconstruct Mean and Eddy Fluxes from Argo Floats?

Christopher Chapman^{a,*}, Jean-Baptiste Sallée^a

^a*LOCEAN-IPSL
Université de Pierre et Marie Curie, Paris CEDEX 75252, France.*

Abstract

The capacity of deep velocity estimates provided by the Argo float array to reconstruct both mean and eddying quantities, such as the heat flux, is addressed using an idealized eddy resolving numerical model, designed to be representative of the Southern Ocean. The model is seeded with 450 “virtual” Argo floats, which are then advected by the model fields for 10 years. The role of temporal sampling, array density and length of the float experiment are then systematically investigated by comparing the reconstructed velocity, eddy kinetic energy and heat-flux from the virtual Argo floats with the “true” values from the model output. We find that although errors in all three quantities decrease with increasing temporal sampling rate, number of floats and experiment duration, the error approaches an asymptotic limit. Thus, as these parameters exceed this limit, only marginal reductions in the error are observed. The parameters of the real Argo array, when scaled to match those of the virtual Argo array, generally fall near to, or within, the asymptotic region. Using the numerical model, a method for the calculation of cross-stream heat-fluxes is demonstrated. This methodology is then applied to 5 years of Argo derived velocities using the ANDRO dataset of Ollitrault & Rannou (2013) in order to estimate the eddy heat flux at 1000m depth across the Polar Front in the Southern Ocean. The heat-flux is concentrated in regions downstream of large

*Corresponding Author

Email address: chris.chapman.28@gmail.com (Christopher Chapman)

bathymetric features, consistent with the results of previous studies. 2 ± 0.5 TW of heat transport across the Polar Front at this depth is found, with more than 90% of that total concentrated in less than 20% of the total longitudes spanned by the front. Finally, the implications of this work for monitoring the ocean climate are discussed.

Keywords: Lagrangian floats, Southern Ocean, Idealized Modelling.

1. Introduction

Deep drifting floats, such the satellite tracked Argo floats and the Autonomous Lagrangian Circulation Explorer (ALACE) floats, or acoustically tracked Sound Fixing and Ranging (SOFAR) and RAFOS floats, provide direct measurements of the oceanic currents as they move with the flow. These floats provide the only direct measurements of the ocean's subsurface currents with broad spatial coverage and have been instrumental in shaping our comprehension of the structure of the ocean's interior (Roemmich et al., 2009; Riser et al., 2016). They have been shown to be capable of producing accurate and rich maps of the time mean interior currents (Davis, 1991a; Gille, 2003a; LaCasce, 2008; Ollitrault and Colin de Verdière, 2014) and measurements of features not readily inferred from remotely sensed surface measurements, such as deep jets and boundary currents (Richardson and Fratantoni, 1999; Fratantoni and Richardson, 1999; van Sebille et al., 2011, 2012). With the continued development of the Argo program, and the improved geographical and temporal coverage of the global ocean that comes with it, it is reasonable to ask: are we capable of observing robust, quantitative statistics of the oceanic meso-scale with current float deployments?

As a way of introducing the problem, Fig. 1a shows all Argo float positions in the Southern Ocean (south of 30° S) within 5 days of the 25th of December 2009. We have determined the mean distance between each of the points plotted in Fig. 1a and their closest neighbor is approximately 160km. Fig. 1b shows the trajectory of a single float, (World Meteorological Organization number

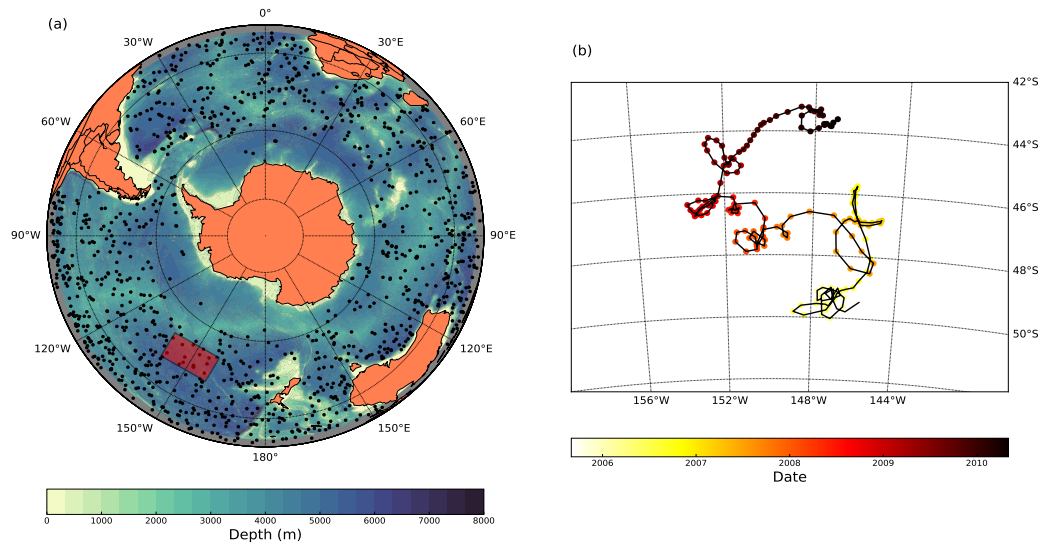


Figure 1: Spatial coverage of Argo floats in the Southern Ocean. (a) All reported Argo float positions at 1000db, within 5 days of the 25th of December, 2009 (points), overlaid over the topography from the ETOPO01 dataset (colored contours); (b) zoom on the highlighted region in panel (a), showing the trajectory of float # 5900777 from the 26th of April 2005 to the 26th of December 2009.

24 #5900777) over its lifetime. Numerous scales of motion are present in this
 25 trajectory, from very tight loops with a radius of order a few kilometers, to
 26 larger meanders with a effective radius of several hundred kilometers. The time
 27 series of float position is *non-stationary* (that is, the statistical properties of the
 28 motion change with time) and, although the Argo float has remained operational
 29 for approximately four years, the trajectory is limited to a relatively small part
 30 of the ocean, drifting only a few degrees throughout its operational life. As
 31 such, this float has repeatedly sampled the same geographic region.

32 Clearly, Argo float #5900777 ‘sees’ a number of features important to general
 33 circulation, including mesoscale eddies superimposed over a larger scale flow
 34 field. However, the geographic region sampled by this float is limited. Thus,

35 we pose the question: what characteristic must an array of these floats have in
36 order to resolve the oceanic meso-scale?

37 Argo floats, by their design, present several challenges for the accurate mea-
38 surements of deep currents. Argo floats must resurface to transmit their data,
39 leaving the currents inferred from their displacement subject to errors such as
40 delays in the surface location fix, shear in the water column and surface drift
41 (Ollitrault and Rannou, 2013). Although a substantial amount of work has been
42 undertaken to determine and control these errors in the measurement, less work
43 has been devoted to understanding the limitations of sampling and the sampling
44 density, particularly when compared with the large amount of work undertaken
45 to understand the limitations of the surface drifter array (Davis, 1982, 1987,
46 1991a,b; LaCasce, 2008). Surface drifters and Argo floats have several substan-
47 tial differences in their sampling characteristics. Due to the fact that surface
48 drifters do not need to complete a dive cycle, they report a position fix every
49 1-2 hours (Elipot et al., 2016), which is much more frequent than the standard
50 Argo position fixes of once every 5 to 15 days (with the vast majority of floats
51 reporting a position every 10 days). Additionally, the surface drifter dataset has
52 far denser sampling statistics than the Argo array. As such, work performed
53 using surface drifters may not translate directly to Argo floats.

54 In this study, we use a combined empirical/observational approach to study
55 the influence that sampling, both spatial and temporal, have on the ability to
56 reconstruct deep flows and the eddy fluxes associated with meso-scale motions,
57 treating the Argo float array as a array of “moving current meters” (Davis,
58 1991b). To do this, we will use a an idealized Observing System Simulation
59 Experiment (OSSE). OSSEs have become relatively common in climate science
60 since the 1980s (Hoffman and Atlas, 2016). The basic principle of an OSSE,
61 described in Hoffman and Atlas (2016), is to take the the output of a numerical
62 model as the “truth” and then sample this output with synthetic observations.
63 With the luxury of knowing the “truth” from the numerical model, the utility
64 of synthetic observing system can then be rigorously evaluated.

65 We will study the errors associated with the length of time between dive

66 and resurfacing of the floats, the spatial density of the Lagrangian array and
67 the duration of the float experiment will be in place by systematically modify-
68 ing the density and time span of the virtual float array, as well as the sampling
69 characteristics of the float derived velocities. Specifically, an idealized, eddy
70 resolving numerical model of the Southern Ocean is “observed” using “virtual”
71 Argo floats. By comparing the Lagrangian derived estimates of mean velocity,
72 eddy kinetic energy and heat flux to the “exact” results from the model solution,
73 we will demonstrate the utility and shortcomings of these Lagrangian measure-
74 ments. We will then use the understanding of the limitations of the Lagrangian
75 derived velocities gained from the model output in order to estimate the eddy
76 heat flux in the Southern Ocean from the existing array of Argo floats. We
77 limit our focus to the Southern Ocean for two primary reason: it is the prin-
78 ciple region of study for both authors of this paper; and the lack of available
79 “traditional” observations from ships means that a detailed investigation of the
80 Argo floats’ capacity to resolve meso-scale statistics is warranted. However, the
81 results obtained here are expected to apply quite generally.

82 The capacity of the Argo array to effectively represent important oceanic
83 variables, including current velocity, has already been subject to several OSSEs.
84 For example, Kamenkovich et al. (2011) used a “virtual” Argo float array, de-
85 signed to resemble the Argo array as it was at the time of publication, to sample
86 the output of a numerical model of the North Atlantic. The virtual Argo float
87 array performance was assessed in two model configurations: with and without
88 mesoscale variability present. Kamenkovich et al. (2009) and Kamenkovich et al.
89 (2011) found that the presence of mesoscale eddies had a profound effect on the
90 virtual Argo array’s data coverage, as eddies tend to efficiently disperse floats,
91 leading to broader spatial coverage. Puzzlingly, they find that poor data cover-
92 age is not consistently correlated with high reconstruction errors. In contrast,
93 errors are generally higher in regions dominated by strong advection, such as
94 the western boundary currents and the ACC. Focusing on the Southern Ocean,
95 Majkut et al. (2014) used a virtual array of Argo floats equipped with bio-
96 geochemical sensors sampling output from the GFDL-ESM2M climate model

97 to demonstrate the potential for these floats to provide useful data from the
 98 real ocean, and to suggest sampling strategies for future deployments. Roach
 99 et al. (2016) used virtual Argo and RAFOS arrays, advected in the data assim-
 100 ilating Southern Ocean State Estimate (SOSE) model to assess the fidelity of
 101 estimates of lateral diffusivity calculated with the real Argo array. However, to
 102 our knowledge, no study has investigated the influence of float sampling, array
 103 density or experiment time-span on the resulting reconstruction error, nor have
 104 the capacity of Argo floats to estimate quadratic quantities, such as heat flux,
 105 been rigorously assessed. In this paper it is our intention address these topics.

106 The remainder of this article is organized as follows: Section 2 discusses the
 107 spatial and temporal sampling characteristics of Argo floats and the effects of
 108 each on the estimation of underlying flow fields. The numerical model configura-
 109 tion and the method of advecting virtual Argo floats, as well as the observational
 110 datasets that will be used in the second part of this study will be described in
 111 Section 3. We will discuss the reconstruction of the numerical model fields from
 112 the virtual Argo floats in section 4, and the ability of Lagrangian observations
 113 to determine cross-frontal heat fluxes in the numerical model in section 5. Es-
 114 timates of the mean flow and the cross-stream eddy heat-flux in the Southern
 115 Ocean using the Argo float array will be presented in section 6, and the re-
 116 sults obtained will be discussed with reference to the numerical model results
 117 in section 7.

118 2. Sampling and Lagrangian Drifters

119 Here we briefly review the role of discrete sampling in reconstructing real-
 120 world signals, and its application to extracting information from Lagrangian
 121 drifters.

122 2.1. Temporal Sampling and Lagrangian Drifters

123 The position of an idealized Lagrangian float, $\mathbf{x} = (x(t), y(t), z(t))$, is related
 124 to the oceanic current velocity $\mathbf{u}(\mathbf{x}; t)$, by:

$$\dot{\mathbf{x}} = \mathbf{u}(\mathbf{x}; t). \quad (1)$$

125 However, due to their dive, drift and resurface cycle, Argo floats are sampled at
 126 discrete time intervals. To determine the current velocity from the Lagrangian
 127 measurements, for the m th float cycle, we follow Lebedev et al. (2007) and use
 128 the following difference equation:

$$\mathbf{u}^m(\mathbf{x}_{\text{deep}}, p_{\text{park}}; t_{\text{deep}}^m) = \frac{\mathbf{x}(t_{\text{asc}}^m) - \mathbf{x}(t_{\text{des}}^m)}{t_{\text{asc}}^m - t_{\text{des}}^m} \quad (2)$$

129 where \mathbf{x}_{deep} , t_{deep} and p_{park} are location, time and parking pressure of the
 130 deep velocity estimate, $\mathbf{x}(t_{\text{asc}})$ and t_{asc} are the location and time of the float
 131 ascension; and $\mathbf{x}(t_{\text{des}})$ and t_{des} are the location and time of the float descent for
 132 that cycle. For the majority of Argo floats, the parking pressure is approximately
 133 1000db. If the time between ascent and descent, Δt , remains constant over the
 134 lifetime of the float (which is true for the virtual floats by construction, and
 135 approximately true of Argo floats), and if the descent time of cycle $m+1$ is
 136 equal to the ascent time of cycle m , then Eqn. 3 can be written:

$$\mathbf{u}^{m+1/2}(\mathbf{x}^{m+1/2}, p_{\text{park}}; t^{m+1/2}) = \frac{\mathbf{x}^{m+1} - \mathbf{x}^m}{\Delta t}. \quad (3)$$

137 where $m+1/2$ represents some time between cycles m and $m+1$.

138 Eqns. 2 and 3 clearly represent a discrete approximation of the continuous
 139 circulation. This discretisation process induces both a truncation error, which
 140 is $O(\Delta t)$, but also an aliasing error, which occurs as a consequence of sampling
 141 with a frequency lower than twice that of the highest frequency present in the
 142 underlying flow (Smith, 1997, pgs. 39–44). Treatment of aliasing in Lagrangian
 143 measurements is not trivial, as the sampling rate provided by the floats depends
 144 on the velocity of the flow being sampled (Willis and Fu, 2008). An example
 145 of aliasing of the flow field is shown in Fig. 2, which compares the velocity
 146 obtained from a Lagrangian drifter using higher (red arrows) and lower (blue
 147 arrows) sampling rates. It is clear from Fig. 2 that a low sampling rate yields
 148 a velocity estimate that does not capture the structure of the underlying flow
 149 field, nor give an accurate velocity estimate. This picture is further complicated
 150 by the fact that if the velocity of the flow were to increase, then the Lagrangian
 151 float would be advected through the flow structure more rapidly, and the spatial

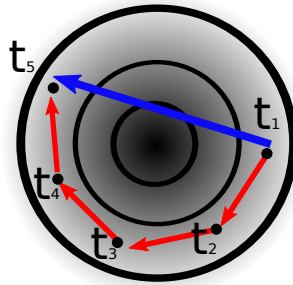


Figure 2: Schematic showing the aliasing of float observations. The shaded background and black, solid contours show the streamfunction of an idealised anticyclonic eddy. The black dots represent the float location at 5 separate times. The red arrows represent the velocity observations determined at the highest available sampling rate: $\Delta t = t_2 - t_1, t_3 - t_2, \dots$. The blue arrow represents velocity estimate made by sampling only the first and last float positions: $\delta t = t_5 - t_1$.

152 sampling would change in response.

153 In this paper we make no attempt to tease out the individual influences
 154 of aliasing and truncation on the overall error performance. Instead, we treat
 155 these two sources of error together as a ‘temporal sampling error’ and note that
 156 the truncation error is expected to grow linearly with Δt . The aliasing error is
 157 expected to be non-stationary: larger in regions dominated by intense features
 158 such as eddies and jets. We will explore how the background flow field modifies
 159 the error obtained using Lagrangian measurement in section 4.

160 2.2. Spatial Sampling provided by Argo floats in the Southern Ocean

161 Fig. 3 gives an indication of the geographical coverage provided by the
 162 dataset in the Southern Ocean. Fig. 3a shows the average number of observa-
 163 tions south of 40°S, binned by longitude and latitude with a bin size of $1^\circ \times 1^\circ$,
 164 between 2005 and 2011 (the time span of the ANDRO dataset). Throughout
 165 most of the Southern Ocean, the density of observations relatively homogeneous.
 166 However, there is a rapid reduction in data availability in high latitudes. For-
 167 tunately, regions with limited observational density typical occur south of the
 168 main ACC fronts (compare Fig. 3a with Fig. 14 or Fig. 4 of Dufour et al.

169 (2015)) and, as such, the ACC region which is the principle focus of this study,
170 can be considered to be approximately evenly sampled by the Argo floats.

171 To further understand the distribution of Argo based observations in the
172 Southern Ocean, we now plot the total number of observations south of 40°S ,
173 binned by longitude with a 1° bin size, in Fig. 3b (black line). The number of
174 observations is spatially variable, with a minimum in Drake Passage longitudes
175 (70°W to 60°W) of approximately 50 observations per degree of longitude over
176 the 5 year period, and a maximum of approximately 450 observations per degree
177 located upstream of Drake Passage at $\sim 90^{\circ}\text{W}$. An average of 220 observations
178 are taken in each longitude bin over the 2005-2011 period. Fig. 3b also shows
179 the average number of floats in each longitude during a 10 day window (red
180 line), which gives an approximation of how many simultaneous measurements
181 are taken in each “snapshot”. The curve in Fig. 3b broadly follows that of 3a,
182 with peaks and troughs in roughly the same longitudes. The average number
183 of floats available in a 10 day period over the entire Southern Ocean basin is
184 400 ± 30 .

185 The preceding analysis demonstrates that the data coverage provided by
186 Argo floats in the Southern Ocean is spatially variable, and that there are fre-
187 quently no Argo floats available to sample a particular region. With an aver-
188 age distance between ‘simultaneous’ measurements of ~ 160 km in the Southern
189 Ocean, which is an order of magnitude greater than the local Rossby deformation
190 radius, resolution of the instantaneous mesoscale field is impossible using Argo
191 floats. However, given that certain floats repeatedly sample the same region and
192 even the same feature (see Fig. 1b) it is difficult to infer a spatial resolution from
193 float distributions alone. Indeed in a similar OSSE Kamenkovich et al. (2011)
194 found only a weak correlation between the how well sampled a region was and
195 the underlying error of the reconstruction. How well the mesoscale statistics are
196 represented with the existing Argo array, and how their representation changes
197 with variations in array parameters, such as the number of floats and the length
198 of time of the float experiment, is the focus of the remainder of this paper.

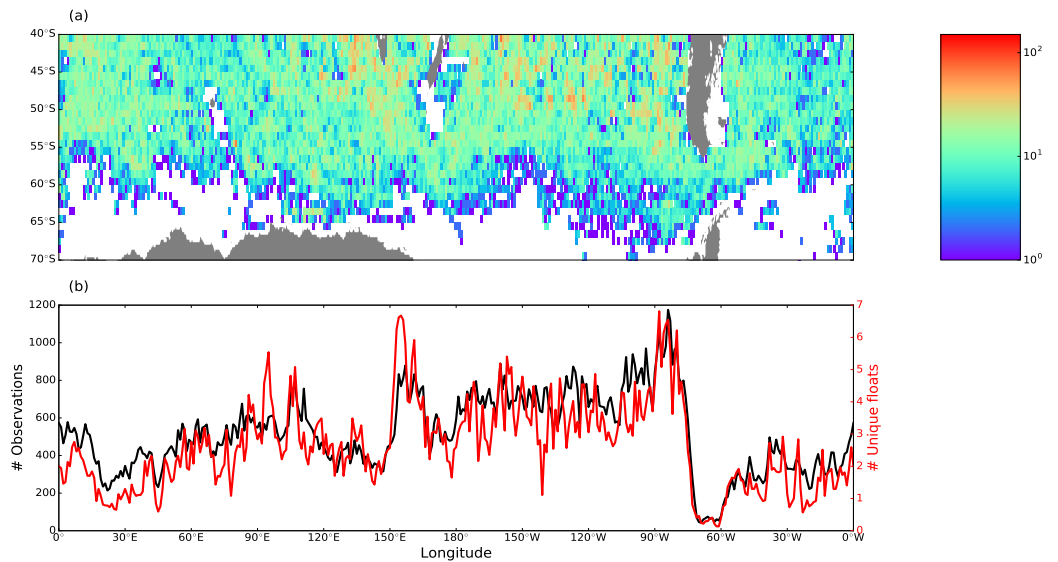


Figure 3: The Argo array spatial sampling characteristics in the Southern Ocean: (a) The number of velocity observations, between 2005 and 2010, south of 40° at 1000m depth, binned onto a 1° longitude/latitude grid; and (b) the total number of velocity observations south of 40° in each 1° longitude bin, across all latitudes (black) and the average number of individual Argo floats available in a 10 day “snapshot” period, in each 1° longitude bin (red).

199 3. Numerical Model, Argo Data, and Methods

200 In this section, we will introduce our idealized numerical model and our
 201 method of advecting numerical (‘virtual’) Argo floats. We will also describe
 202 the observational Argo float dataset (the ANDRO dataset) from which we will
 203 reconstruct the mean and eddy fluxes.

204 3.1. Numerical Model Configuration

205 The configuration of our numerical model is an idealized representation of
 206 the Southern Ocean, inspired by Abernathey et al. (2011). Here, we use the Nu-
 207 cleus for European Modelling of the Ocean (NEMO) model, version 3.6 (Madec,

208 2014), which solves the three-dimensional primitive equations on the β -plane,
 209 in standard vertical depth coordinates, using a C-grid for the spatial discretiza-
 210 tion and a linear equation of state with a constant salinity of 25 g.kg^{-1} . Our
 211 configuration is a zonally periodic Cartesian channel with a zonal length, L_x
 212 of 6000km and a meridional width, L_y of 2000km and a maximum depth H of
 213 4000m. Advection of both momentum and tracers is handled by the 3rd order
 214 upwind-biased scheme, which induces a resolution dependent implicit diffusion.
 215 Thus, no explicit horizontal diffusion or viscosity is applied. Vertical diffusion
 216 is handled using a Generic Length Scale (GLS) scheme. Surface forcing is sup-
 217 plied by a meridionally varying sinusoidal wind-stress $\tau(y) = \tau_0 \sin(\pi y/L_y)$ and
 218 by relaxing the surface to an imposed linear surface temperature distribution,
 219 with a relaxation coefficients of $30 \text{ W.m}^{-2} \text{ K}^{-1}$, as in Barnier et al. (1995). Ad-
 220 ditionally, following Abernathey et al. (2011) the temperature on the northern
 221 150km of the domain is relaxed to an exponential temperature profile, with a
 222 relaxation coefficient of 7 days^{-1} , which allows for the formation of a residual
 223 overturning.

224 We induce zonal assymetry in the model by the introduction of bottom
 225 bathymetry. As in Abernathey and Cessi (2014), we use a meridional ridge
 226 with a Gaussian cross-section described by:

$$h(x) = H_0 e^{-\frac{(x-L_x/2)^2}{\sigma_0^2}},$$

227 where h is the height of the bathymetry above the ocean floor, x is the zonal
 228 coordinate, $\sigma_0 = 150 \text{ km}$ is the topographic length scale and $H_0 = 2000 \text{ m}$ is the
 229 scale height of the topographic obstacle. The scale height and topographic
 230 length scales has been chosen to effectively block lower layer flow and induce a
 231 large stationary meander, thus effectively capturing some of the impacts of large
 232 bathymetric features, such as the Kerguelen Plateau, on the Southern Ocean
 233 circulation.

234 The model horizontal grid spacing is 5km and 50 vertical levels, distributed
 235 such that the vertical grid spacing is smaller towards the surface and deeper
 236 towards the ocean floor (minimum Δz of $\sim 5 \text{ m}$, maximum of $\sim 175 \text{ m}$). With an

Table 1: Parameter values used in the configuration of the numerical model.

Symbol	Value	Description
L_x	6000km	Zonal Domain Length
L_y	2000km	Meridional Domain Length
$\Delta x, \Delta y$	5km	grid-spacing
Δt	300s	barotropic time-step
H	4000m	Depth
H_0	2000m	Topography Scale Height
f_0	$-1.0 \times 10^{-4} \text{s}^{-1}$	Coriolis parameter
β	$1 \times 10^{-11} \text{s}^{-1} \text{m}^{-1}$	Meridional
τ_0	$1.5 \times 10^{-4} \text{N} \cdot \text{m}^{-2}$	Peak wind stress
r_D	$1.5 \times 10^{-3} \text{m} \cdot \text{s}^{-1}$	Linear bottom drag
κ_v	$0.5 \times 10^{-5} \text{m} \cdot \text{s}^{-2}$	Vertical diffusivity
T_S	7 days ⁻¹	Sponge layer relaxation time-scale
α	2.0×10^{-4}	Thermal expansion coefficient
g	$30 \text{W} \cdot \text{m}^{-2} \text{K}^{-1}$	Surface temperature relaxation coefficient

237 approximate Rossby deformation radius of 20km (verified by direct calculation
 238 after spin-up), this grid spacing is sufficient to explicitly resolve the meso-scale.
 239 The model is spun-up for 200 years, which is sufficient for the interior flow
 240 to attain statistical equilibrium, and then run for an additional 10 years. We
 241 output the snapshots of the model velocity (u , v , and w components) with daily
 242 temporal frequency. Additional parameter choices are noted in Table 1.

243 An example of the model output at 1000m depth is shown in Fig. 4. Fig. 4a
 244 shows the time-mean horizontal speed of the simulated currents at 1000m depth.
 245 Although highly idealized, our simulation captures a number of phenomena
 246 present in the ocean. As in the Southern Ocean, our simulation shows the flow

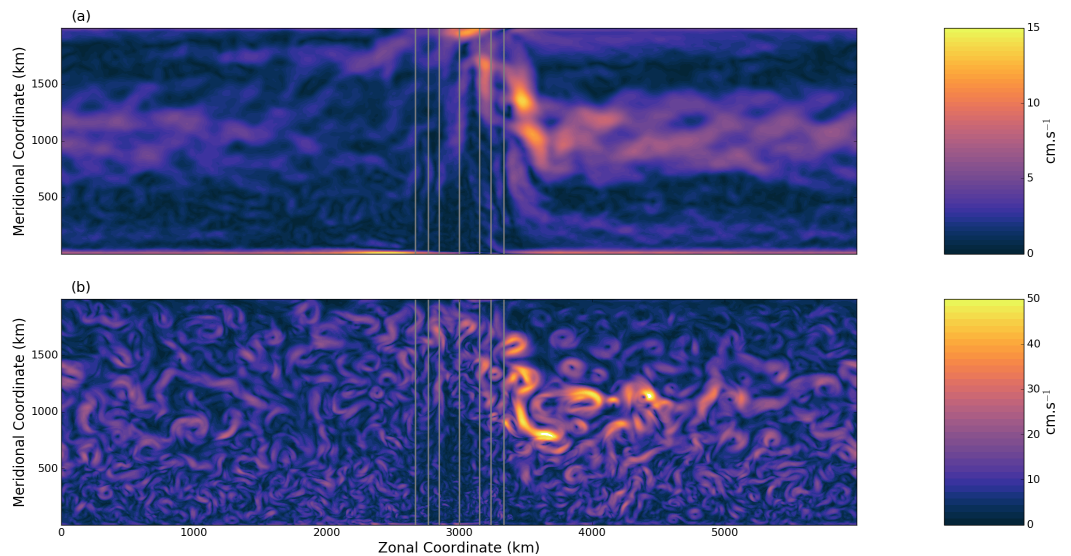


Figure 4: An example of the numerical model output at 1000m depth. (a) The mean current speed, taken over the 10 years of the model run; and (b) a snapshot of the current speed. The solid grey lines indicate the idealised topography depth (CI:500m)

247 organized into a series of zonal jets. The currents are steered by the bathymetry,
 248 being diverted to the north as they traverse the obstacle. Downstream of the
 249 bathymetry, a stationary meander is formed. Fig. 4b shows a snapshot of
 250 the current velocity at 1000m. Meso-scale features are evident throughout the
 251 domain, with an enhanced intensity downstream of the bathymetry, reminis-
 252 cent of an oceanic storm-track (Williams et al., 2007; Chapman et al., 2015).
 253 Characteristic mean velocities are found to be $10\text{--}15\text{cm.s}^{-1}$, with instantaneous
 254 velocities that can reach 60cm.s^{-1} , consistent with observations in the Southern
 255 Ocean (Ollitrault and Colin de Verdière, 2014).

256 3.2. Virtual Argo Float Advection

257 In this study we shall make extensive use of virtual Argo floats advected
 258 by the model fields. Hence, it is worthwhile to briefly discuss the numerical

259 implementation of the particle advection scheme and some of the assumptions
260 behind it.

261 We solve Eqn. 1 using a 4th order Runge-Kutta scheme with an adaptive
262 time-step, allowing us to specifically control the error of the solution while main-
263 taining computational efficiency. In practice, the truncation error of the solution
264 is required to be less than 10^{-3} (that is one part in 1000), although the true
265 computational error may be less than this value. Floats are advected “offline”,
266 using saved model output. The velocity at a particular virtual float time and
267 position is obtained by 3D linear interpolation.

268 The virtual Argo floats are advected on a constant depth surface: thus,
269 there is no vertical displacement of the particle. Additionally, we do not re-
270 quire our floats to undergo a surfacing/descending ‘dive’ cycle: the virtual float
271 positions are thus known exactly and there are no errors arising from vertical
272 shear in the water column, nor position fix delays. As such, the virtual Argo
273 floats can be considered to be “perfect” in the sense that the only source of
274 error is numerical. Roach et al. (2016) have tested how the Argo dive cycle
275 affects the estimations of diffusivity when compared to ‘perfect’ virtual floats
276 in a realistic numerical simulation of the Southern Ocean. They found that,
277 even with relatively pessimistic assumptions, the Argo dive cycle induced errors
278 that were small relative to natural variability within the ocean. Although this
279 calculation was performed in a different context, the results obtained by Roach
280 et al. (2016) allow us to assume that neglecting the Argo dive cycle will not
281 significantly affect the resulting reconstructions.

282 450 virtual Argo floats are advected in the model at a depth of 1000m.
283 The number of floats is selected by noting that there are, on average, $400 \pm$
284 30 floats in the Southern Ocean latitudes south of 40° at any particular time
285 (see Section 2.2). At 55°S , the earth’s circumference is $\sim 22 \times 10^3$ km, resulting
286 in ~ 0.02 floats per kilometer of zonal extent. With the model zonal basin
287 length of $L_x = 6000$ km, 110 ± 10 floats are required in the model to maintain an
288 equivalent number of floats per degree of longitude in the model. To test how the
289 reconstruction error changes with additional floats, we use 4 times the minimal

290 number floats required, hence 450. The location of each virtual float is saved
291 daily: thus there are 3650 x and y position records (10 years \times 365 days/year)
292 for each of the 450 floats in the experiment, giving a total of 1,642,500 virtual
293 float positions.

294 It is important to note that even with the relatively high rate at which the
295 model output is produced (1 day) the virtual Argo floats are liable to ‘overshoot’
296 (Keating et al., 2011) due to unresolved high frequency motions. Following
297 Keating et al. (2011), we have attempted reduce this error by maintaining a
298 maximum time step in the virtual Argo float integration of $\Delta t = 1$ hour and
299 linearly interpolating the model fields (both spatially and temporally) to the
300 virtual float location. At 5km grid spacing, this places our float experiment
301 within the ‘overshoot’ regime (see Fig. 14 of Keating et al. (2011)). However,
302 as noted by Keating et al. (2011) interpolation cannot eliminate the problem
303 of particle overshoot, and, as such, it is likely that our virtual particles show
304 spuriously high diffusivity due to this numerical effect.

305 *3.3. Deep Current Velocities, and Temperature and Salinity Profiles From Argo* 306 *Floats*

307 For the observational component of this study, we make use of the ANDRO
308 dataset (Ollitrault and Rannou, 2013), freely available for download (<http://wwz.ifremer.fr/lpo/>).
309 ANDRO provides estimates of the current velocity at the parking pressure of
310 the float and at locations that are estimated from the locations of the previous
311 two surface locations estimates, while controlling for, or estimating, sources of
312 error such as those due to vertical shear, surface fix delay, surface drift due to
313 inertial oscillations and uncertainty in the dive time. Unlike similar datasets
314 (for example the YoMaHa’07 dataset of Lebedev et al. (2007)) ANDRO also ex-
315 plicitly accounts for drift in the parking pressure that occur over the lifetime of
316 the float. We consider floats between the years 2005 and 2011. A total of 2440
317 floats are available south of 10°S, yielding a total of 217,065 independent esti-
318 mates of velocity at depths ranging from 500db to 2000db, although in practice,
319 we consider only velocity estimates near 1000db.

320 In section 6 of this paper, we will estimate the heat fluxes using Argo data.
 321 As such, knowledge of the temperature at the float parking depth is required. We
 322 obtain profiles of temperature, salinity and pressure from the surface to 2000db,
 323 for each of the floats in the ANDRO database from the various Argo Global Data
 324 Assembly Centers (Roemmich et al., 2009; Riser et al., 2016). The temperature
 325 and salinity are then used to determine the conservative temperature T using
 326 the TEOS-10 algorithm (McDougall and Barker, 2011). The value of T is then
 327 interpolated to the ANDRO velocity data locations using linear interpolation
 328 from adjacent float locations as in Elipot et al. (2016).

329 3.4. Reconstruction of Fields from Point Observations

330 The data provided by Lagrangian float observations are scattered and un-
 331 structured. As such, in order to estimate oceanographic fields on a regular
 332 grid, some mapping or ‘interpolation’ scheme must be employed. In most of
 333 the oceanographic literature, mapping is accomplished by optimal interpolation
 334 (Wunch, 2006, p. 163) or local least-squares fitting (Ridgway et al., 2002). Al-
 335 though powerful, these methods are computationally intensive. Since we will be
 336 performing numerous reconstructions, we chose to use the simpler procedure of
 337 geographic binning (Davis, 1991b; LaCasce, 2008). With this methodology, the
 338 domain is discretized into $N_x \times N_y$ points. All observations of some quantity,
 339 θ , that fall within some radius, R , of a particular grid point, are averaged to
 340 form a local ensemble mean:

$$\bar{\theta}(\mathbf{x}_i) = \sum_{d_{ij} < R} \theta_k(\mathbf{x}_j; t_j), \quad (4)$$

341 where d_{ij} is the distance from the grid-point \mathbf{x}_i to the float location \mathbf{x}_j . The
 342 geographical binning approach makes the implicit assumption that the mean
 343 and any residuals have distinctly different time scales, that there are sufficient
 344 observations to reliably estimate the mean, and that the binning radius, R is
 345 less than the decorrelation scale of the underlying data (LaCasce, 2008). Even
 346 assuming that these conditions have been met, geographic binning has numer-
 347 ous shortcomings. For example, the choice of radius R can influence the spatial

348 scale of the reconstructed flow. In addition, in situations where the number of
349 Lagrangian observations is variable in space, random background processes can
350 give rise to a spurious velocity down the gradient of the observational sampling
351 density (that is, the number of samples per unit area) (Davis, 1991b). In prin-
352 ciple, it is possible to correct for this effect, although it is technically difficult
353 (Davis, 1991b, 1998).

354 Despite these problems, we persist with this methodology due to its compu-
355 tational speed and since we are principally interested not in the absolute error
356 of the reconstruction, but instead the *relative errors* over the parameter space
357 to be explored. However, the reader should keep in mind the shortcomings of
358 the mapping procedure and recognize that absolute errors in fields produced in
359 this paper can be considered a “worst case” scenario and could be improved
360 through the application of more sophisticated methods.

361 4. Reconstruction of Mean and Eddy Fields in the Idealized Model

362 We now study the ability of velocities inferred from Lagrangian displacement
363 data to effectively reconstruct the large-scale flow field and the statistics of the
364 meso-scale using the virtual Argo floats advected in the numerical model. We
365 will test the sensitivity of the reconstruction to the number of virtual floats,
366 the length of time of the float experiment and their sampling characteristics.
367 On first glance, varying the number of floats and the length of time of the
368 experiment may seem redundant, as each parameter simply modifies the number
369 of observations. However, Argo floats are costly, and the absolute number of
370 Argo floats in the ocean is not expected to substantially increase in the next
371 few years, although there may be an increased focus of increasing the density
372 of observations selectively in certain regions (Riser et al., 2016). Additionally,
373 the number of floats must be sufficient to sample the majority of the domain. A
374 single float, for example, is unlikely to sample the entire model domain unless the
375 experiment is run for a prohibitively long time. As such, since number of Argo
376 floats is expected to remain somewhat fixed, it is certainly worth considering

377 how the fidelity of the reconstruction will change should the Argo array continue
 378 to operate with an unchanged number of floats.

379 4.1. Instantaneous Errors and the Effects of Temporal Sampling

380 In order to test the influence of the temporal sampling on the velocity errors,
 381 we determine the virtual float velocities using sampling periods of every 1, 2,
 382 5, 10, 20, 30, 40 and 50 days. In order to simulate the effects changing the
 383 temporal sampling rate, Eqn. 3 is modified to include the sampling interval,
 384 $K \in \mathbb{Z}$:

$$\mathbf{u}^{m+K/2}(\mathbf{x}^{m+K/2}, p_{\text{park}}; t^{m+K/2}) = \frac{\mathbf{x}^{m+K} - \mathbf{x}^m}{t^{m+K} - t^m} = \frac{\mathbf{x}^{m+K} - \mathbf{x}^m}{K\Delta t}. \quad (5)$$

385 By calculating the Lagrangian velocities in this manner, rather than simply
 386 sub-sampling the Lagrangian time series, the total number of velocity remains
 387 approximately constant, which avoids the problem of reducing the number of
 388 samples in the signal that would occur if it were sub-sampled naïvely. We note,
 389 that each virtual Argo float trajectory must be truncated by $K - 1$ points due to
 390 the finite length of the rolling window, although since the number of observation
 391 removed is small compared to the total number of observation the truncation
 392 has no discernible effect on the resulting statistics. In this section we use 110
 393 virtual floats, which is the number of virtual floats required to ensure that the
 394 number of floats per degree of longitude in the model is representative of the
 395 Argo array in the ACC.

396 The normalized histograms of the u and v velocity estimated from the vir-
 397 tual floats is shown in Fig. 5, where they are compared with the distributions
 398 calculated directly from the model output (thick black dashed line). Although
 399 the estimated distributions show a similar Gaussian character to the true dis-
 400 tribution, the virtual floats tend to produce distributions that underestimate
 401 the frequency of large magnitude velocities, as the tail of the estimated distri-
 402 butions fall below that of the true distribution for velocities with magnitudes
 403 larger than $\sim 7.5 \text{ cm.s}^{-1}$. As such, the virtual Argo floats tend to underestimate
 404 the magnitude of more extreme velocities produced by the model. Additionally,

405 the estimated distributions also tend to differ from the true distribution when
 406 velocities are weak. At high sampling rates, the velocity obtained from the vir-
 407 tual Argo floats tends to *underestimate* the frequency of weak velocities when
 408 the sampling rates are high (1–10 days), and *overestimate* their frequency when
 409 the sampling rates are low (20–50 days). There is also a significant asymmetry
 410 in the distribution of v that is most notable at strongly negative values. We are
 411 unable to definitively identify the cause of this asymmetry, although we spec-
 412 ulate that it may arise due to the fact that the storm track region, where the
 413 strongest eddy velocities are found, is collocated with strong southward mean
 414 flow induced by the stationary meander. We note also that the virtual argo floats
 415 are not able to capture this asymmetry, consistent with their underestimation
 416 of the true flow velocity in the tails of the distribution.

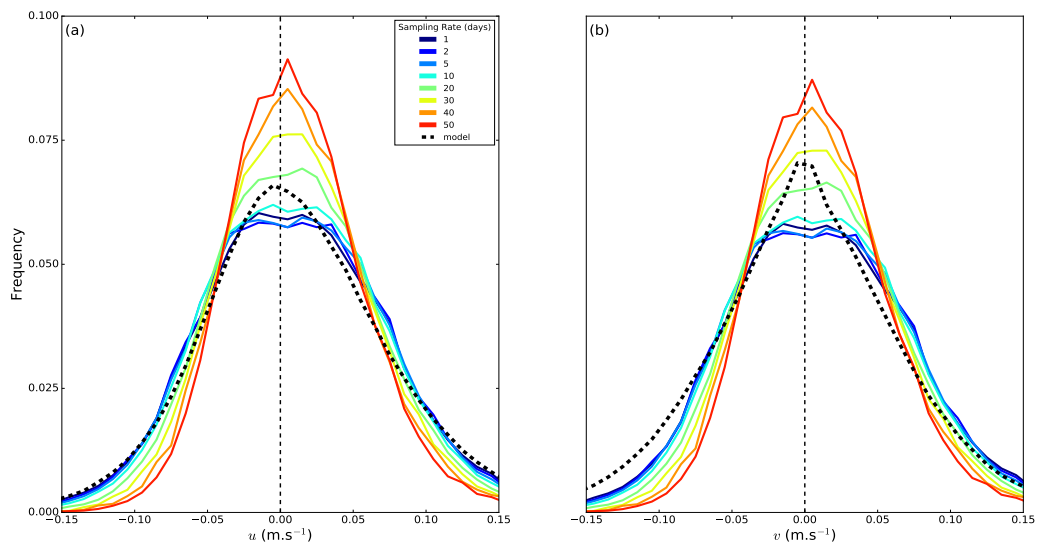


Figure 5: The normalized histograms for the zonal (a) and meridional (b) current velocities estimated by the virtual Argo floats, for each temporal sampling interval (see legend in panel (a)). The histogram computed directly from the model output at each of the float sampling points is indicated by the thick, dashed black curve.

417 To further explore the ability of the virtual Argo floats to estimate the
 418 modeled currents, at the location of each virtual float velocity measurement, we
 419 calculate the absolute velocity error:

$$\epsilon_{\mathbf{u}}^{\text{abs}}(\mathbf{x}_i; t) = \hat{\mathbf{u}}(\mathbf{x}_i; t) - \mathbf{u}(\mathbf{x}_i; t), \quad (6)$$

420 and the relative velocity error:

$$\epsilon_{\mathbf{u}}^{\text{abs}}(\mathbf{x}_i; t) = \frac{\hat{\mathbf{u}}(\mathbf{x}_i; t) - \mathbf{u}(\mathbf{x}_i; t)}{\mathbf{u}(\mathbf{x}_i; t)} \quad (7)$$

421 where $\hat{\mathbf{u}}(\mathbf{x}_i; t)$ is the velocity estimated from the virtual Argo floats and $\mathbf{u}(\mathbf{x}_i; t)$
 422 is the “true” velocity taken directly from the model at virtual float location
 423 $\mathbf{x}_i \forall i \in [1, N_{\text{obs}}]$, which is estimated at the virtual float locations by bilinear
 424 interpolation. These error estimates are averaged meridionally and binned by
 425 longitude with a bin size of 20km (4 grid cells). We have tested bin sizes
 426 from 10km to 50km, and found 20km to be a good compromise between the
 427 smoothness of the reconstructed fields and the ability of the our methodology
 428 to reconstruct important features. In each longitude bin, we compute the root-
 429 mean-squared-error:

$$\text{RMSE}_{\mathbf{u}}^{\text{abs,rel}} = \left[\frac{1}{M} \sum_{i \in \mathcal{B}} \left(\epsilon_{\mathbf{u}}^{\text{abs,rel}}(\mathbf{x}_i; t) \right)^2 \right]^{1/2} \quad (8)$$

430 where M is the number of observations in each longitude bin and \mathcal{B} represents
 431 the current longitude bin.

432 Fig. 6a shows the RMS of the absolute meridional velocity error ϵ_v^{abs} in each
 433 longitude bin (the zonal component shows very similar behavior). It is clear that
 434 the absolute error in the velocity estimated by the virtual floats, regardless of
 435 the sampling rate, increases downstream of the bathymetry (indicated by the
 436 dashed line in Fig. 6a). It is in the downstream “storm track” region that meso-
 437 scale eddies are the most intense (Chapman et al., 2015). The velocity error
 438 shows the greatest sensitivity to the sampling rate in the storm track. In the
 439 region upstream of the obstacle, the difference between the velocity estimates
 440 obtained using a sampling rate of 1 day and 50 days is approximately 5cm/s

441 in the less energetic upstream region, while the difference in errors increases
 442 to 1.25cm/s in the energetic storm track region. However, it is worth noting
 443 that the difference between errors obtained using a sampling rate of 1 day and
 444 those using a sampling rate of 10 days (the usual Argo sampling frequency) are
 445 indistinguishable upstream of the topography and the difference is limited to
 446 less than 2.5cm/s even in the storm track region. The virtual Argo floats are
 447 able to estimate the current speed with RMS errors of approximately 2.5cm/s
 448 upstream of the topography and approximately 5cm/s in the region downstream
 449 of the topography when the sampling rate is 10 days or less.

450 Investigation of the relative errors, plotted in Fig. 6b, reveals that the
 451 virtual Argo floats have errors between 20% with sampling periods less than
 452 20 days, rising to errors that are 80-100% for sampling rates of once every 50
 453 days. In contrast to the absolute errors, the relative error remains approximately
 454 constant throughout the domain for sampling rates more frequent than 20 days.
 455 For sampling rates more frequent than 20 days, the relative error in the velocity
 456 increases by approximately 20% in the storm track region. Since velocities
 457 are highest in the turbulent region downstream of the topography, the relative
 458 insensitivity of the relative error throughout the domain underscores the strong
 459 dependence of the instantaneous error on the velocity being observed.

460 Fig. 6 also shows the distributions of $\epsilon_{\mathbf{u}}$ for both the zonal (Fig. 6c) and
 461 meridional (Fig. 6d) components. The distributions for each velocity compo-
 462 nent are very similar, save for asymmetry that is present in the zonal error
 463 distribution. With changing sampling rate, both ϵ_u and ϵ_v distributions show a
 464 decreasing frequency of errors near zero and increasing standard deviation with
 465 increasing sampling period.

466 To understand how velocity errors manifest it is instructive to examine the
 467 scatter between the virtual Argo float velocity error and the true velocity, as in
 468 Fig. 7a for sampling rates of 1, 10 and 50 days for the v velocity component
 469 (the u component has a similar structure). Fig. 7a shows that, for all cases
 470 considered, there is a significant negative correlation between ϵ_v and the velocity
 471 being measured. As such, the virtual Argo floats tend to *underestimate* strongly

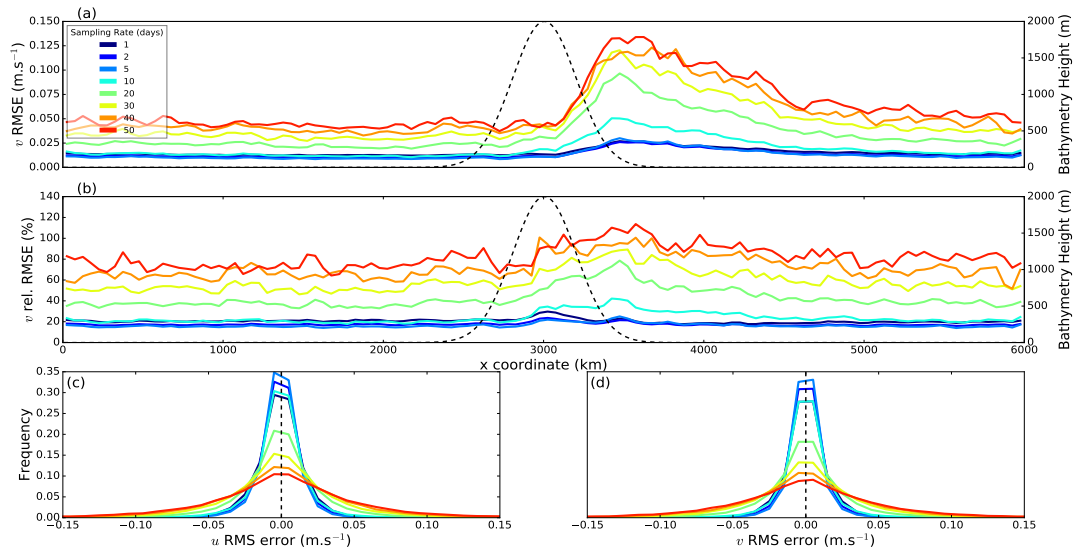


Figure 6: The effect of temporal sampling on the error in the velocity. (a) The *absolute* RMS errors in v , defined in equation 8 binned by longitude with a bin size 20km and averaged between $y = 500\text{km}$ and 1500km for each temporal sampling interval (colors) and; (b) as in panel (a) but for the *relative* RMS errors in v . The thin dashed line indicates the topography height. The (normalised) histogram of the errors in the u (c) and v (d), for each temporal sampling interval (see legend in panel (a)).

472 positive velocities and *overestimate* strongly negative velocities. A linear fit
 473 for each sampling rate is obtained using orthogonal regression (used in lieu of
 474 standard linear regression due to the increased density of points clustered near
 475 0), plotted in Fig. 7a (dashed black line) for the 10 day sampling rate. The
 476 slope of this linear fits is negative for all sampling rates more frequent than
 477 30 days, suggesting a consistent underestimation of high current speeds even
 478 at relatively high sampling rates. The scatter of points away from the best fit
 479 line increases as the sampling rate is decreased, particularly around 0m/s. In
 480 fact, with a sampling rate of 1 day, the points in Fig. 7 cluster about 0, giving
 481 the impression of data “funnelling” towards the axes center. For the 10 day

482 sampling period, the scatter of the error remains approximately constant about
 483 the best-fit line, while for the 50 day sampling, the error performance for slower
 484 current velocities (near $v=0$) deteriorates.

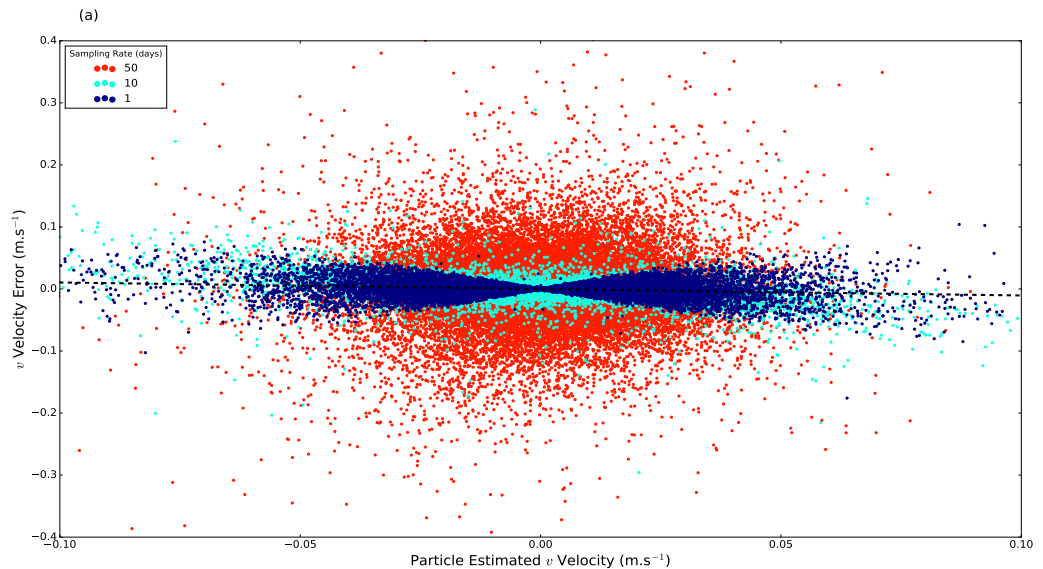


Figure 7: Influence of the current speed on the error. Velocity error (ordinate) vs. the estimated velocity (abscissa) for temporal sampling intervals 1 day (red), 10 days (turquoise) and 50 days (blue). The dashed line indicates the linear fit for the 10 day sampling period (slope $-0.3\text{m.s}^{-1}/\text{m.s}^{-1}$).

485 Are the virtual floats able to capture the dominant spatial and temporal
 486 scales of variability present in the model? Directly relating Lagrangian mea-
 487 surements to Eulerian is a complex task beyond the scope of this article, as a
 488 Lagrangian drifter moving through a flow field observes both spatial and tem-
 489 poral variations simultaneously (Middleton, 1985; Maas, 1989; Rupolo et al.,
 490 1996; Rupolo, 2007). However, we note that by assuming that the turbulence
 491 field evolves slowly on the advective time-scale (which is Taylor’s “frozen field”
 492 hypothesis), then Eulerian wavenumber (spatial) and frequency (temporal) spec-

493 trum should have the same slopes (Taylor, 1938; Arbic et al., 2012). As such,
 494 if the frozen field hypothesis holds, and we should note that there is now sub-
 495 stantial evidence that this hypothesis is only partially applicable to geostrophic
 496 turbulence (Arbic et al., 2012, 2014), if we are able to estimate the local Eulerian
 497 frequency spectrum from the Lagrangian observations, then we should also
 498 be able to develop qualitative understanding of the distribution of local spatial
 499 scales.

500 Middleton (1985) and Maas (1989) have shown that frequency spectra of
 501 Lagrangian observation will approximate the Eulerian frequency spectra when
 502 averaged over an ensemble of Lagrangian observations. Maas (1989) also showed
 503 that the Lagrangian spectra of an ensemble of floats well approximates the Eulerian
 504 spectra obtained by measurements fixed relative to a moving background
 505 flow, although the Lagrangian spectrum is ‘smeared’ when compared to the Eulerian
 506 spectrum. As such, we compute the Fourier transform of the complex
 507 velocities:

$$\tilde{v}(t) = u(t) + iv(t), \quad (9)$$

508 for all virtual floats with segments of at least one year within the two sub-
 509 domains shown by the black rectangles in Fig. 9. These sub-domains are chosen
 510 to be representative of the two dominant dynamical regimes in the model:
 511 energetic storm track region downstream of the topography, and the quieter
 512 region upstream of the topography. These individual virtual float spectra are
 513 then averaged together and compared with the complex velocity spectra computed
 514 directly from the model fields, area averaged over each individual region.
 515 The comparison between the PSDs is presented in Fig. 8a for the non-energetic
 516 eastern box, and in Fig. 8b for the energetic storm track region (a comparison
 517 between the spectra of the two regions is shown in the inset box). Note that
 518 as the PSDs are computed from complex time-series, the PSDs are asymmetric.
 519 Recall that highest frequency resolvable from discretely sampled observations is
 520 half the sampling frequency.

521 We find that the ensemble average of the virtual float spectra follow closely

522 the area-averaged Eulerian spectra taken directly from the model output. Over
 523 comparable frequency ranges there is little difference between the sampling
 524 rates. However, the virtual float spectra are generally too steep though the
 525 intermediate frequency ranges within in the non-energetic region (Fig. 8b) and
 526 over all frequencies higher than about 0.05 days^{-1} (~ 20 day periods) in the
 527 storm track region, indicating that over the temporal scales that contain some
 528 parts of the meso-scale field, some energy is not being captured by the virtual
 529 floats.

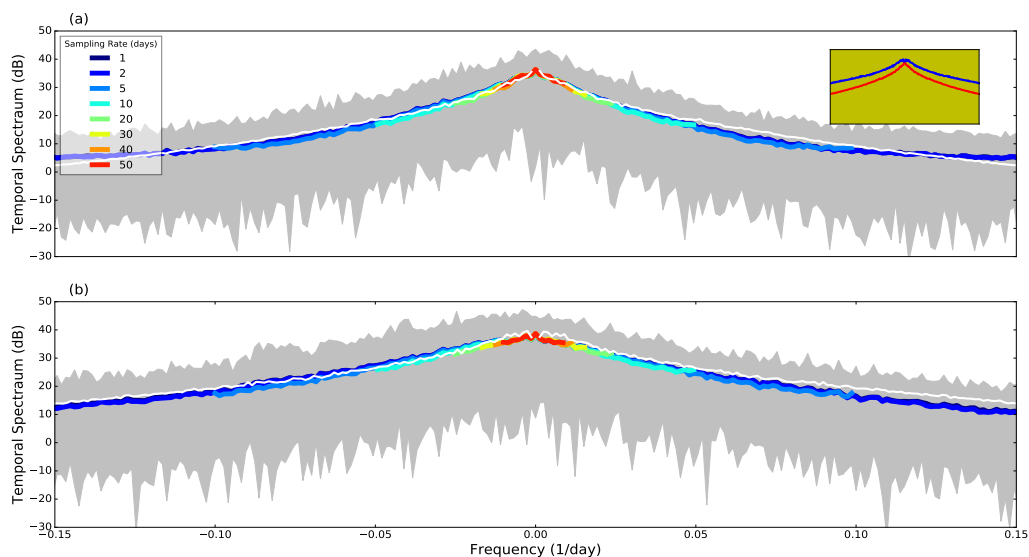


Figure 8: The temporal power spectral density $S(\omega)$ of the complex velocity $w = u + iv$, computed directly from the model output (white lines, inset) compared to ensemble average of all virtual float tracks longer than 1 year (colors) averaged over the (a) western (quiet) box; and (b) eastern (eddy) box. The individual colors correspond to velocity data calculated using different sampling frequencies. Grey shading shows the ensemble of float spectra. Note the log-linear axis scale. The inset box in panel (a) shows the average PSD for the western (red) and eastern (blue) regions computed directly from the model.

530 To summarize the results of this analysis virtual Argo float derived velocities
 531 will represent the modelled current velocities and their probability distributions

532 provided that the sampling rate remains more frequent than about 20 days.
 533 However, there is a notable tendency for the Lagrangian derived velocities to
 534 underestimate the magnitude of the current velocity, particularly as the speed
 535 increases, regardless of the sampling rate. The ability of Argo floats to accu-
 536 rately estimate the instantaneous flow velocity will have important implications
 537 for the Argo array to effectively determine meso-scale eddy statistics.

538 4.2. Reconstruction of Mean and Eddy Fields from Lagrangian Observations

539 We now discuss the problem of reconstructing mean and eddy fields from
 540 noisy Lagrangian drifter velocities. We approach the problem empirically, inves-
 541 tigating systematically the effects of changing the Lagrangian array parameters
 542 on the reconstructed fields.

543 4.2.1. Effect of Temporal Sampling

544 As shown in section 4.1, local estimates of the current velocity are sensitive
 545 to the temporal sampling rate. As such, we expect that the sampling rate would
 546 also affect the ability to reconstruct the large-scale flow fields.

547 As an example of the reconstruction of the model fields from 110 virtual Argo
 548 floats with a 10 days sampling rate (the standard Argo sampling rate) is shown
 549 in Fig. 9. For comparative purposes, the model fields are shown in panels (a)i–
 550 (c)i, and the equivalent fields reconstructed from the virtual Argo floats in panels
 551 (a)ii–(c)ii. We have chosen to investigate the time-mean meridional velocity v ,
 552 the eddy kinetic energy $EKE = 0.5 [\overline{u'u'} + \overline{v'v'}]$ and the meridional heat flux
 553 density $\rho c_p \overline{vT}$. The latter two quadratic quantities can give an indication of the
 554 ability of the virtual Argo floats to resolve eddy processes.

555 Fig. 9 shows good qualitative agreement between the model output and the
 556 reconstructed fields. The virtual Argo data is able to reproduce the standing
 557 wave produced by the interaction of the mean flow with topography, the magni-
 558 tude and the extent of the storm track produced downstream of the topography,
 559 and the response of the meridional heat flux to both. However, there is a ten-
 560 dency for the virtual Argo floats to underestimate these fields, consistent with

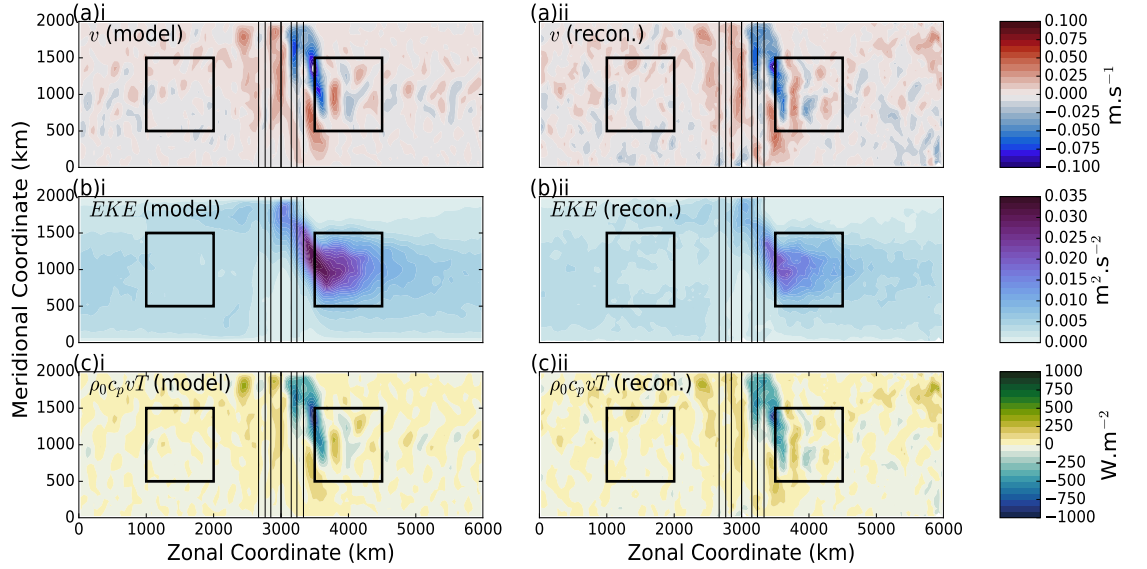


Figure 9: Comparison of the model fields at 1000m vs. fields reconstructed from virtual drifters with a 10 day sampling rate. The modelled (i) and the reconstructed (ii) fields of (a) time mean velocity \bar{v} ; (b) eddy kinetic energy; and (c) meridional heat flux $\rho_0 c_p \bar{v} T$. Thin black lines are bathymetric contours (CI: 500m) and the thick black lines show the boxes for the proceeding error calculations.

561 the analysis in section 4.1 which showed an increasing underestimation of the
 562 current speed as that speed increases.

563 To understand more quantitatively the sensitivity of the errors, we compute
 564 the RMSE, both absolute and relative of the difference between the true and the
 565 reconstructed fields within the two sub-domains shown in Fig. 9. The *absolute*
 566 RMSE for an arbitrary time-mean field $\theta(x, y)$ and its reconstruction $\hat{\theta}(x, y)$ is
 567 defined as:

$$\text{RMSE}^{\text{abs}} = \left\{ \frac{1}{N_x N_y} \sum_j^{N_y} \sum_i^{N_x} [\hat{\theta}(x_i, y_j) - \theta(x_i, y_j)]^2 \right\}^{1/2} \quad (10)$$

568 where $x_i, y_j \forall (i, j) \in \{(1, \dots, N_x, 1 \dots N_y)\}$ are the grid points inside the sub-
 569 domain, and N_x and N_y are the total number of grid points in the sub-domain.

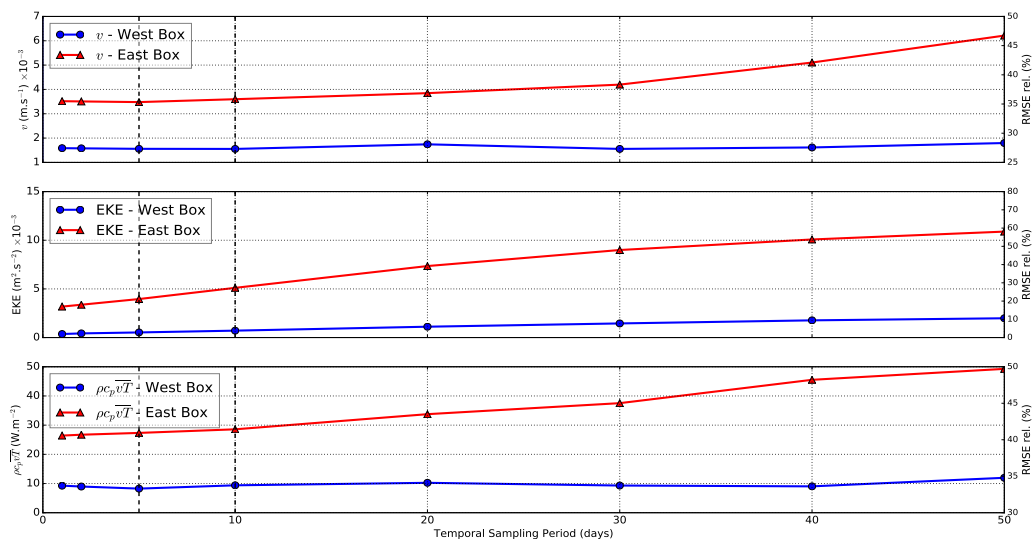


Figure 10: The area averaged RMSE computed over the eastern (blue) and western (red) as a function of sampling interval, Δt for the (a) meridional velocity v ; (b) the eddy kinetic energy EKE ; and (c) the meridional heat flux $\rho_0 c_p v T$. Dashed black lines indicate the parameter regime occupied by Argo floats.

570 The *relative* RMSE is the absolute RMSE normalized by the variance of the θ
 571 over the sub-domain.

572 The RMS errors for each of the quantities, integrated over the two subdo-
 573 mains, are shown in Fig. 10. For all quantities considered here, the error is
 574 relatively insensitive to changes in the temporal sampling rate in the western
 575 (quiet) region (blue line, circular markers). However, in the storm track region
 576 (red line, triangular markers) the error shows strong sensitivity to the sampling
 577 rate, with a non-linear response that accelerates when Δt increases over about
 578 20 days.

579 Despite the rapid increasing error with lower sampling rates, the RMSE
 580 remains relatively insensitive to sampling rate for the first few values used in

581 this study. For example, the RMSE in the eddy heat flux (Fig. 10c) increases
 582 from approximately 25W.m^2 at the most frequent sampling rate (1 day^{-1}), in
 583 the (turbulent) eastern sub-domain, corresponding to a relative error of $\sim 15\%$ to
 584 30W.m^2 (20% relative error) at the standard Argo sampling rate of $10.\text{days}^{-1}$.

585 When comparing errors between the eastern and western sub-domains, it is
 586 clear from Fig. 10 that the error is greater over the eastern storm-track box
 587 than the western box. The error over the storm track box is ~ 2 to ~ 3 times
 588 higher than the equivalent error over the eastern region. The larger errors in
 589 turbulent region downstream underscore the results of section 4.1 that showed
 590 an underestimate of the high magnitude motions.

591 4.2.2. Effect of the Number of Floats

592 We now repeat the analysis of section 4.2.1, this time investigating the in-
 593 fluence of the number of independent floats in the virtual array. The temporal
 594 sampling rate is held constant at the common Argo float sampling rate of 10
 595 days. As in section 4.2.1, we compute the RMS errors between the model and re-
 596 constructed time mean meridional velocity, EKE and heat-flux, integrated over
 597 the two regions shown Fig. 9. The number of virtual floats is controlled by ran-
 598 domly sampling a fraction of the float trajectories from the complete data set.
 599 The fractions of the total number of floats selected are $1/16, 1/8, 2/8, 3/8 \dots 1$.

600 The RMSE as a function of the total number of virtual floats for the each
 601 of the reconstructions are shown in Fig. 11. As should be expected, we find a
 602 decreasing RMSE for each of the fields considered with an increasing number
 603 of virtual floats. The RMSE, however, begins to approach a constant limit in
 604 both regions as the number of floats passes approximately 150. For example, as
 605 the number of floats increases from 28 (the smallest number used) to 112, the
 606 errors in the heat flux decrease from $\sim 120\text{W.m}^{-2}$ (a relative error of 50%) to
 607 20W.m^{-2} (20% relative error) in the western (quiet) box and $\sim 180\text{W.m}^{-2}$ (65%
 608 relative error) to 40W.m^{-2} (30% relative error) in the western (storm track)
 609 box, a decrease of approximately 80% in both cases. However, increasing the
 610 number of floats fourfold from 112 to 450 results in RMSE reductions of between

611 5 and 10% in each region. As such there exists a certain number of floats which
 612 could be considered ‘sufficient’, given the diminishing returns in RMSE with an
 613 increasing number of floats.

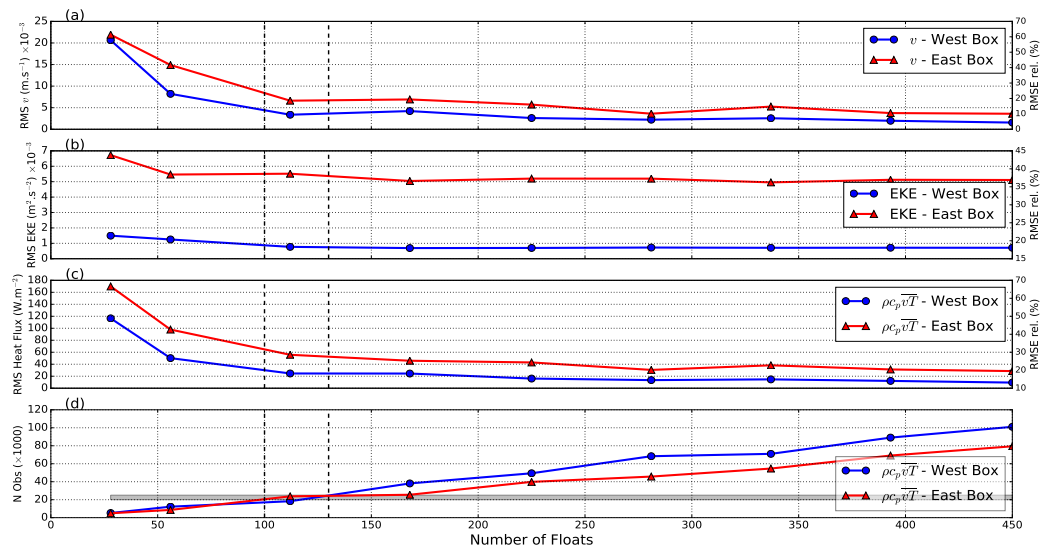


Figure 11: As in Fig. 10, but showing the change in the RMSE with variation in the number of virtual Lagrangian floats used in the reconstruction. Additional Panel (d) shows the total number of floats used in each reconstruction. Dashed black lines indicate the parameter regime occupied by Argo floats in the Southern Ocean, while the grey shaded region in panel (d) shows the equivalent number of observation density provided by Argo floats in the Southern Ocean

614 4.2.3. Effect of the Length of the Float Experiment

615 To conclude this section, we now investigate the effect of varying the length
 616 of float experiment from 1 to 10 years. The number of floats and temporal
 617 sampling rate are held constant at a value equivalent to the existing Argo array
 618 in the Southern Ocean (110 floats, which ensures that the number of floats
 619 per degree of longitude is similar to that of the current Argo array, with a
 620 sampling rate of 10 days). The RMSEs of the three chosen quantities are shown

621 in Fig. 12. As with the number of floats, we find that the RMSE approaches
 622 a limit for all three quantities after the experiment has been running for 4 to
 623 5 years, decreases from errors as high ranging from 90-100% in the eastern
 624 box in the case of the time-mean meridional current and eddy heat flux after a
 625 single year float experiment, to 20-30% after 5 years. There is some suggestion
 626 of improvement in the meridional heat flux error (Fig. 12c) in the energetic
 627 eastern box throughout the 10 years of the experiment, with a reduction in the
 628 relative error from $\sim 20\%$ to $\sim 10\%$. The decrease in the RMSE from 5 to 10
 629 year is certainly not as significant as during the first 4 years of the simulation,
 630 where the RMSE decreased by 60-70%.

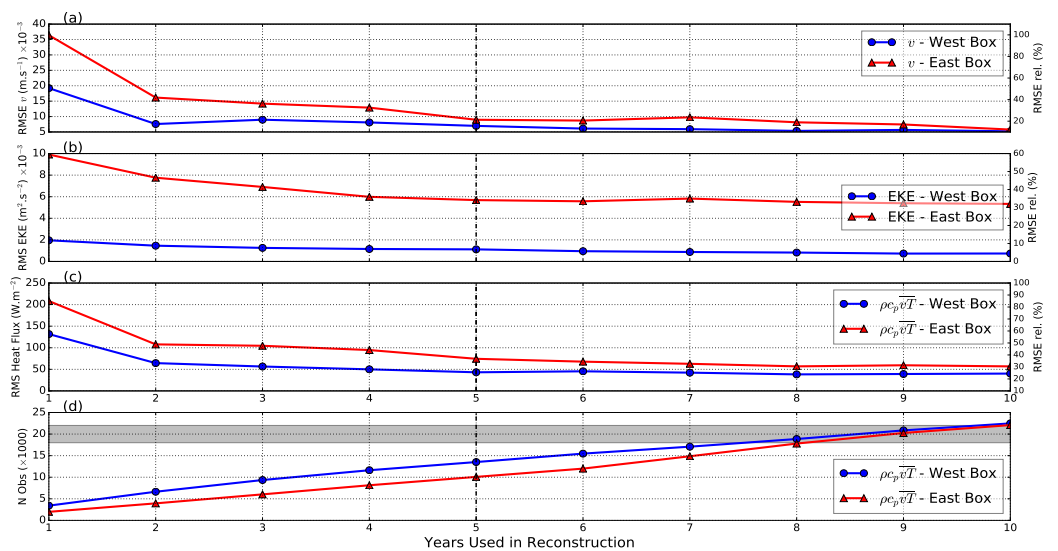


Figure 12: As in Fig. 10, but showing the change in the RMSE with variations in the number of years of Lagrangian data used in the reconstruction. Dashed black lines indicate the number of years in the ANDRO dataset Ocean, while the grey shaded region in panel (d) shows the equivalent number of observation density provided by Argo floats in the Southern Ocean

631 As with the previous discussion of the influence of the number of floats on

632 the capacity on large scale reconstructions of oceanographic quantities, there
 633 appears to be diminishing returns in RMSE after approximately 4 years, with
 634 the first 5 years providing approximately 90% of the reduction in heat-flux
 635 RMSE and the final five years providing an additional 10% of error reduction.

636 5. Cross-frontal eddy-fluxes from Lagrangian drifters

637 Is it possible to use Lagrangian observations to estimate the cross-stream
 638 fluxes? This question is complicated by the fact that it is necessary to estimate
 639 not only the fluxes themselves, but also the front or streamline, which must
 640 be computed in a manner consistent with the computed fluxes. To understand
 641 the importance of consistent estimation of the streamlines, consider the time-
 642 mean flux density of some tracer, θ , written as $\mathbf{F}^\theta = \mathbf{u}\theta$. We can now form the
 643 time-mean “flux-streamline” from the time-mean flux by solving the differential
 644 equation:

$$\frac{d\bar{\mathbf{X}}}{ds} = \bar{\mathbf{F}}^\theta, \quad (11)$$

645 with initial condition $\mathbf{X}(0) = \mathbf{X}_0 = (x_0, y_0)$. In Eqn. 11, $\mathbf{X} = (X(s), Y(s))$ are
 646 respectively the zonal and meridional coordinates of the streamline, parameter-
 647 ized by the arc-length, s . By construction, there can be no time-mean transport
 648 across this streamline, as $\mathbf{F}^\theta \cdot \boldsymbol{\eta} = 0$ everywhere along the curve, where $\boldsymbol{\eta}$ is the
 649 unit normal to \mathbf{X} . The streamline is not guaranteed to form a closed loop, even
 650 in a periodic domain, since integrating the y component of Eqn. 11 around the
 651 full circuit gives:

$$y_1 - y_0 = \oint \frac{dY}{ds} ds = \oint v\theta ds \quad (12)$$

652 where y_1 is the latitude of the streamline as it crosses its original longitude.
 653 $y_1 - y_0$ is not necessarily zero, as the flux $v\theta$ is not normal to the streamline and
 654 thus does not integrate to zero. Despite this fact, we note that in practice the
 655 difference between y_1 and y_0 is small. We take the latitude of the streamline
 656 as it crosses its original longitude x_0 to be y_1 , such that $X(s_1) = (x_0, y_1)$,
 657 where s_1 is the total arc-length of the streamline as it completes a circumpolar
 658 circuit, and $y_1 \neq y_0$. If we close this curve by artificially extending it from

659 (x_0, y_1) to (x_0, y_0) , then there can be a non-zero time mean flux across the
 660 curve, concentrated solely in the segment $(x_0, y_1) \rightarrow (x_0, y_0)$. Now, consider a
 661 new curve, \mathbf{X}' , with identical starting latitude and longitude as the streamline
 662 \mathbf{X} , that is $\mathbf{X}'(s_0) = (x_0, y_0)$, but constructed in such a way that it both forms
 663 a closed contour circling the domain, (i.e. it returns to (x_0, y_0)) and remains
 664 close to the original streamline \mathbf{X} . Since $\mathbf{X}' \cdot \boldsymbol{\eta} \neq 0$ as the new curve is no longer
 665 aligned with the streamline defined by Eqn. 11, there will be small, but non-
 666 zero cross-stream flux distributed along the contour. Since the curves \mathbf{X} and
 667 \mathbf{X}' enclose similar areas, as long as \mathbf{F}^θ is smooth, then by Green's theorem, the
 668 total flux across each contour should also be similar. However, the *distribution*
 669 of this flux along each curve is likely to be very different, with any non-zero flux
 670 across \mathbf{X} restricted to the $(x_0, y_1) \rightarrow (x_0, y_0)$ segment, while flux across \mathbf{X}' is
 671 likely to be distributed along the contour. The origin of these fluxes could be
 672 due to the contour X' passing through a new region of enhanced eddy activity
 673 or, more likely, due to a misalignment of the streamline path and the mean flux.
 674 While the former phenomena is interesting and worthy of further study, the
 675 later simply indicates that the definition of the front does not follow the mean
 676 path of the circumpolar current.

677 As such, to unambiguously identify the source of a flux, we now decompose
 678 the time-mean tracer flux density into time-mean and perturbation components:

$$\bar{\mathbf{F}}^\theta = \bar{\mathbf{u}}\bar{\theta} + \overline{\mathbf{u}'\theta'}. \quad (13)$$

679 We can now compute a new streamline, $\bar{\mathbf{X}}$, defined as:

$$\frac{d\bar{\mathbf{X}}}{ds} = \bar{\mathbf{u}}\bar{\theta}, \quad (14)$$

680 with the initial conditions $\bar{\mathbf{X}}(0) = (x_0, y_0)$. By construction, there can be
 681 no contribution to the total cross-stream flux from the mean component, as
 682 $\bar{\mathbf{u}}\bar{\theta} \cdot \boldsymbol{\eta} = 0$ at all points on the curve. However, since $\overline{\mathbf{u}'\theta'} \cdot \boldsymbol{\eta} \neq 0$, any significant
 683 fluxes must then arise solely from the eddy component. By using this special
 684 definition of a streamline, we are able to unambiguously identify the origins of
 685 the cross-frontal fluxes.

686 Local fluxes can be decomposed by the Helmholtz theorem into rotational
 687 and divergent components. The rotational fluxes necessarily non-divergent and
 688 they do not contribute to the to the local tracer balance. Thus, rotational
 689 fluxes have no direct dynamical effect on the flow field (Marshall and Shutts,
 690 1981). However, rotational fluxes can dominate any local flux (Marshall and
 691 Shutts, 1981; Griesel et al., 2009). In this paper, we do not attempt to perform
 692 a Helmholtz decomposition to remove the rotational fluxes, as it is not obvious
 693 how this should be done using our Lagrangian observations. Furthermore, in a
 694 singly periodic domain, such as our model domain, no unique decomposition of
 695 the flux exists (Fox-Kemper et al., 2003). On the other hand, rotational fluxes,
 696 although they dominate term in local tracer budgets, tend to transport as much
 697 tracer into a region as they do out of it (Jayne and Marotzke, 2002). Thus
 698 summing over a region tends to cancel out the non-divergent fluxes (Griesel
 699 et al., 2009). As such, rather than attempting a Helmholtz decomposition,
 700 instead, we follow Griesel et al. (2009) and Dufour et al. (2015) and compute
 701 the cumulative sum of the across-front tracer transport. Doing so has the effect
 702 of removing the majority of the dynamically inert rotational fluxes, as the act of
 703 summing positive and negative rotational fluxes with similar magnitudes results
 704 in a large degree of self-cancellation.

705 5.1. Cross-frontal heat-fluxes in the numerical model

706 With the argument made in the previous subsection in mind, we now attempt
 707 to determine the cross-frontal eddy heat fluxes at 1000m depth in the numerical
 708 model. The cross-frontal heat flux density calculated in this section is defined:

$$F_{\eta}^{\Theta}(x, y) = c_p \rho \Theta(x, y) v_{\eta}(x, y) \Delta s \Delta z. \quad (15)$$

709 In Eqn. 14 $c_p = 4.0 \text{ J} \cdot \text{kg}^{-1} \cdot \text{K}^{-1}$ is the specific heat of sea-water at constant
 710 pressure, and $\rho = 1024 \text{ kg} \cdot \text{m}^{-3}$ is the sea-water density.

711 For the purposes of comparison, two different frontal definitions are used.
 712 The first, which we call a ‘Lagrangian’ definition, is defined using the definition
 713 given in Eqn. 14, with initial conditions of $\mathbf{X}_0 = (0, L_y/2)$. The ‘mean’ flow

714 in Eqn. 14 is taken from the reconstructed mean velocity field, $\bar{\mathbf{u}}$ obtained
715 from the virtual Argo floats as described in section 4. Similarly, the mean
716 conservative temperature field, Θ , is determined using the virtual Argo float
717 temperature estimates (which are obtained by linearly interpolating the model
718 fields to the float velocity measurement locations) using the same geographic
719 binning methodology described in Sec. 3.4. The second definition, which we
720 call ‘Eulerian’ is defined by direct integration of the hydrostatic equation from
721 the surface to 1000m to obtain a geostrophic streamfunction, ψ_g , from which a
722 streamfunction contour is selected as the front. For easy comparison between
723 the Lagrangian and Eulerian fronts, we select the contour present at $\mathbf{X}_0 =$
724 $(0, L_y/2)$. The Lagrangian and Eulerian fronts, together with the geostrophic
725 streamfunction, are plotted in Fig. 13a. Although the definitions of each front
726 differ, Fig. 13 shows very similar trajectories. Differences in the location of the
727 fronts generally occur only at small scales. We also note that the Eulerian front,
728 by definition, returns to its initial location after a full circuit of the domain. In
729 contrast, the Lagrangian front does not exactly return to its starting location.
730 However, the difference between the front’s initial and final location is less than
731 20km.

732 The cross-frontal eddy heat flux is now estimated by the virtual Argo floats
733 in a manner almost identical to the method used to reconstruct the gridded
734 fields in Section 4: all float observations within 100km of a point on the front
735 are collected, resolved into along and across front components and the ensemble
736 is averaged. We compute the eddy heat flux from the virtual Argo floats across
737 both the Eulerian and Lagrangian contour, which allows us to evaluate the
738 influence of the choice of contour definition on the resulting reconstruction.
739 We also compute the eddy heat-flux across the Eulerian front directly from
740 the model output. This value is taken as the ‘true’ value for the purposes of
741 computing error statistics.

742 The structure of the cross-frontal heat flux is shown in Fig 13b for each of
743 our estimates. The directly computed heat flux (the ‘true’ value, red curve in
744 Fig. 13b) shows a very similar structure to that discussed by Abernathey and

745 Cessi (2014): there is an increased *southward* heat flux in the storm-track region
746 directly downstream of the topographic feature in the largest standing meander.
747 This localized southward heat-flux is somewhat moderated by a northward heat-
748 flux further upstream, consistent with the mechanism proposed by Abernathy
749 and Cessi (2014) (see their Fig. 3).

750 When estimating the cross-frontal heat-flux using the virtual Argo floats,
751 we report mixed results. While the estimates made using the virtual floats
752 capture the enhanced southward eddy heat-flux downstream of the topography,
753 only the flux estimated across the Eulerian contour captures the northward
754 heat flux further downstream. Investigating the source of this error reveals that
755 the Lagrangian contour does not produce a large enough secondary standing
756 meander and, as such, the northward heat flux is not represented.

757 The RMS error at each point on the contour is shown for the virtual Argo
758 float derived heat flux estimates in Figure 13c, where it is easily seen that
759 for both contours the error peaks in the storm-track region downstream of
760 the topography. Although the mean heat-flux error in this region is a fac-
761 tor of four larger than in the less energetic upstream region (defined over the
762 boxes described in Section 4) the error relative to the heat-flux magnitude,
763 $\epsilon_r = (F^{\text{estimated}} - F^{\text{exact}}) / F^{\text{exact}}$ remains roughly constant over the domain
764 (not shown). The fact that the heat-flux error scales with the magnitude of
765 the underlying heat flux is consistent with the discussion in Section 4, where
766 it was shown that both mean and eddy errors were significantly higher in the
767 storm-track region when compared with those in the quiet upstream region.

768 The cumulative fluxes, plotted in Fig. 13d, assuming a layer thickness of
769 100m, show that the basic qualitative spatial structure of the cross-frontal heat
770 flux captured by the Lagrangian observations, with the net southward heat
771 flux concentrated in the energetic region downstream of the topography, and
772 little significant heat flux outside of this region. Quantitatively, the southward
773 heat flux is underestimated by the Lagrangian observations. The heat-flux is
774 approximately 2 to 2.5 times smaller in the storm-track region when compared
775 to the heat flux estimated directly from the model output. It is also notable that

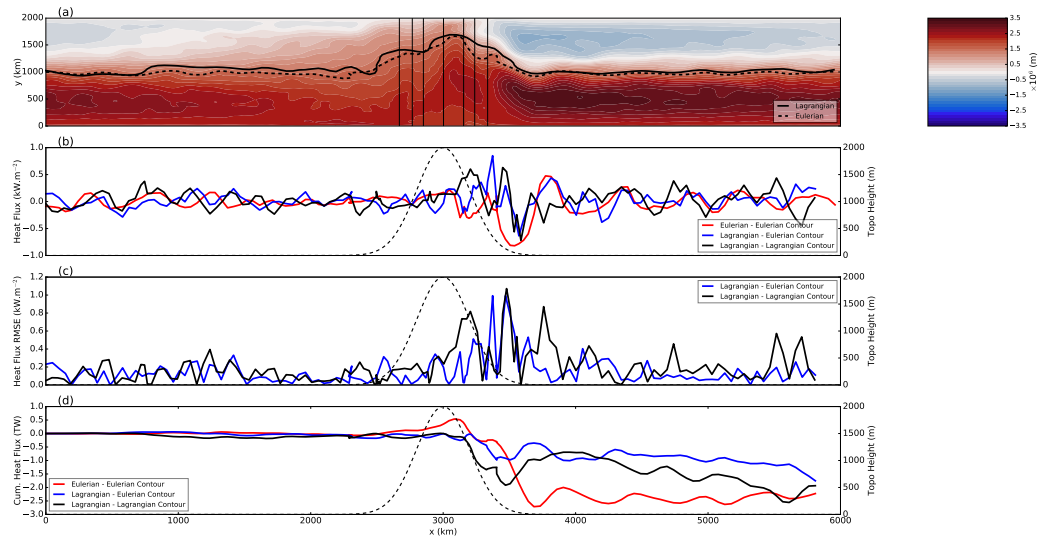


Figure 13: The cross stream heat flux at 1000m calculated directly from the numerical model and estimated from the virtual Argo floats. (a) The model time-mean geostrophic streamfunction ψ_g . The solid black line (labelled “Lagrangian”) indicates the streamline used for the heat-flux calculation determined from the virtual Argo floats, while the dashed line (labelled “Eulerian”) is the equivalent streamline computed directly from the numerical model fields; (b) cross-stream heat flux computed directly from the numerical model across the “Eulerian” contour (red); from the virtual Argo floats across the “Eulerian” contour (blue); and from the virtual Argo floats across the “Lagrangian” contour (blue); (c) the root mean squared error (RMSE) for the heat flux estimated by the virtual Argo floats across the Eulerian (blue) and Lagrangian (black) contour; and (d) the cumulative heat flux along the contours.

776 there is a slow drift in the heat flux estimated across the Lagrangian contour.
777 We have not been able to identify the source of this drift, but as the Lagrangian
778 contour and the Eulerian contour are not perfectly aligned, small fluxes across
779 this contour can easily accumulate into a significant net southward heat flux.

780 The results of this section indicate that the uncertainty in the cross-frontal
781 heat flux is as sensitive to the exact definition of the contour itself as it is to the
782 underlying errors from the use of finite number of Lagrangian observations in
783 its reconstruction. Small changes in a contour's location or orientation appear
784 to result in large localized differences in the flux across the contour. However,
785 despite these problems, the cross-frontal heat flux from the virtual Argo floats
786 captures the broad scale quantitative heat flux structure, correctly determining
787 the localisation of the heat-flux downstream of the bathymetry, as well as pro-
788 viding a quantitative estimate that correctly captures the heat-flux's order of
789 magnitude. These results provide some confidence that the existing Argo array
790 can be used to study heat-fluxes in the real ocean.

791 **6. Reconstruction of Deep Mean and Eddy Fluxes in the Southern** 792 **Ocean from Argo Floats**

793 We now employ the lessons learned from the numerical simulation to the
794 problem of estimating the mean and cross-frontal heat flux in the Southern
795 Ocean using the Argo array of floats between 2005 and 2011. Here, we make use
796 of the ANDRO dataset and the associated Argo hydrographic profiles, described
797 in Section 3. Additionally, we model the error in the velocity estimates as a
798 sum of instrumental error, $\epsilon_{\mathbf{u}_{\text{inst}}}$, which includes the error due to shear in the
799 water column and is included with the ANDRO dataset, and the sampling error,
800 $\epsilon_{\mathbf{u}_{\text{samp}}}$. The sampling error is simulated by direct Monte-Carlo methods. For
801 each velocity estimate, 1000 simulated velocity errors are drawn from a normal
802 distribution with mean and standard deviations determined from orthogonal
803 regression of the virtual Argo float errors described in Section 4.1 (see Fig. 7)
804 from the float experiment with parameters most appropriate to the Southern

805 Ocean Argo array (that is, 5 years experiment duration, 10 day sampling period
806 and 110 floats). Thus the error dependence on velocity is included in the error
807 model. The final error values are the instrumental and sampling errors summed
808 in quadrature.

809 *6.1. Time Mean Circulation and Heat Flux*

810 The time-mean speed and heat flux at 1000m depth are estimated using the
811 procedure described in Section 3 and displayed in Fig. 14. The mean speed
812 maps (Fig. 14a) are essentially identical to those produced by Ollitrault and
813 Colin de Verdière (2014) (see their Fig. 10) using the same dataset and show
814 numerous features, such as quasi-zonal jets associated with the Antarctic Cir-
815 cumpolar Current (ACC), topographic steering of those jets, strong boundary
816 currents and stationary meanders that are all known phenomena in the Southern
817 Ocean (Rintoul and Garabato, 2013). Current speeds of up to 25cm/s are found
818 in the boundary currents and in the ACC jet cores. The meridional heat flux
819 (Fig. 14b) shows enhanced values along the core of the ACC and downstream
820 of large bathymetric features where the heat flux is organized into a alternating
821 northward/southward bands due to the presence of standing meanders, remi-
822 niscent of the high resolution numerical simulations of Griesel et al. (2009) (see,
823 for example, their Fig. 3). The fact that these mean fields produce a large
824 number of the expected features of the Southern Ocean's circulation indicate
825 that there are sufficient observations within the ANDRO dataset, with suffi-
826 cient geographic coverage, that it is capable of producing at least qualitatively
827 accurate mean fields.

828 The error field, shown in Fig. 14c. Errors are limited to less than 30 cm.s⁻¹
829 throughout the Southern Ocean, and are found to be higher in regions associated
830 with strong jets or downstream of topographic features. However, the contrast
831 between regions is not large and and the estimated errors generally vary less
832 than 10 cm.s⁻¹ across the basin, consistent with the results of the idealized
833 numerical model.

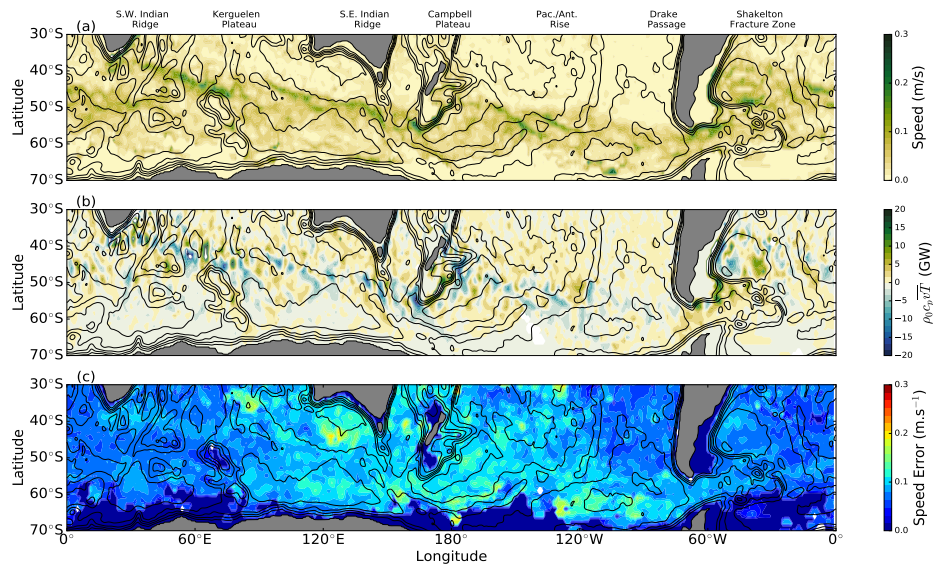


Figure 14: Time mean (a) speed; (b) meridional heat flux ($\rho_0 c_p \overline{vT}$) in the 950-1150db layer, reconstructed from the ANDRO float derived current velocities and Argo temperature profiles; and (c) estimated speed error including both instrumental and sampling errors. Thin black contours are the bathymetry (CI:1000m)

834 6.2. Cross-Frontal Eddy Heat Flux

835 We now compute the heat flux across a circumpolar contour that approxi-
 836 mates a mean streamline at this depth. To determine this streamline, we follow
 837 the procedure outlined in Section 5: we integrate Eqn. 14 numerically (as before
 838 with a 4th order Runge-Kutta scheme), using the time-mean velocity and con-
 839 servative temperature fields and assume a layer thickness of 150m. The latitude
 840 of the contour at 0° longitude is set to 48°S , corresponding to the approximate
 841 location of the polar front determined by Dufour et al. (2015) in a high resolution
 842 model. The location of this contour and bathymetry taken from the ETOPO01
 843 dataset (Amante and Eakins, 2009) is plotted in Fig. 15a. This contour follows
 844 a similar pathway to previous calculations of the polar front (e.g. Dufour et al.
 845 (2015)) and, as such, we take this contour to be the polar front (although it

846 should be noted that circumpolar ‘contour’ definitions of fronts have several
 847 limitations, e.g. Chapman (2014, 2017)). We note as well that although the
 848 observational sampling density along this streamline is approximately constant,
 849 there is a reduction of approximately 50% in the south west Pacific region, be-
 850 tween approximately 100°W and 80°W. As such, the sampling error estimates
 851 in this region are likely optimistic. We have repeated this calculation with more
 852 pessimistic parameter settings and obtained similar overall error estimates, in-
 853 dicating that the dominant source of uncertainty in the oceanic system is likely
 854 internal variability.

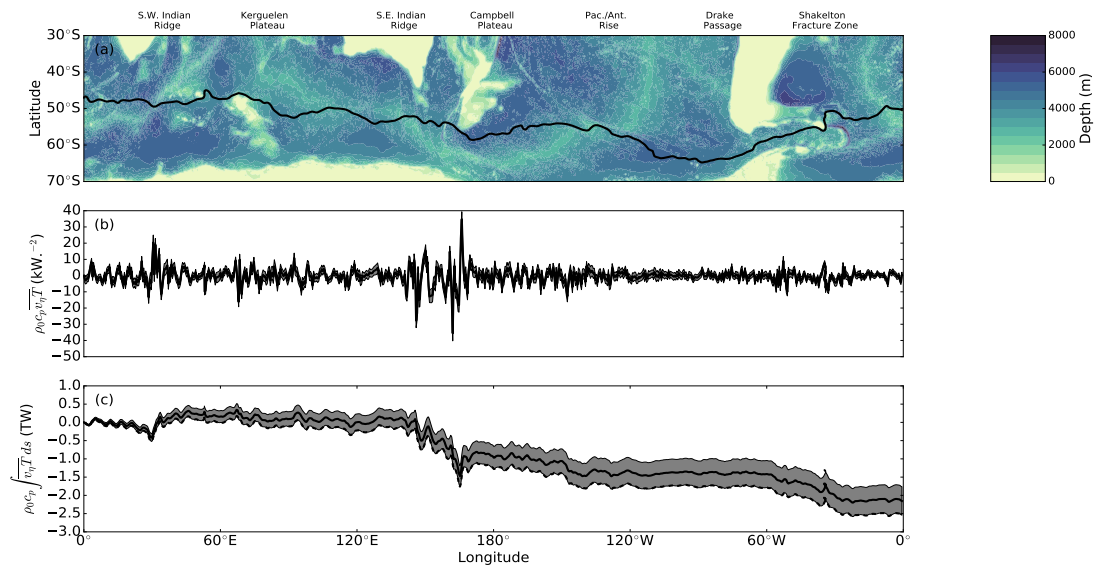


Figure 15: Heat-flux across the polar front estimated from the ANDRO dataset and the Argo temperature profiles. (a) the time-mean position of the polar front overlaying the bathymetry from the ETOPO01 dataset; (b) the cross frontal heat-flux ($\rho c_p \overline{vT} \Delta s \Delta z$); (c) the cumulative cross-frontal heat-flux. Shaded grey regions in panels (a) and (b) indicate the 3σ error bounds

855 The local eddy heat flux across the polar front is shown in Fig. 15b. As in
 856 Thompson and Sallée (2012) and Dufour et al. (2015), we find that the eddy
 857 heat flux is localized in ‘hot-spot’ regions where either the front crosses large

858 bathymetric features (labeled in Fig. 15a) such as the Campbell Plateau and
859 through Drake Passage, or in adjacent downstream regions. The magnitude of
860 the eddy heat flux averaged over regions where bathymetry is shallower than
861 1500 is approximately 2.5 greater than in deeper regions. Although the mag-
862 nitude of the cross-frontal heat flux increases in the regions with important
863 bathymetry, it is important to note that the eddy heat flux shows large positive
864 and negative fluctuations that cancel upon integration along the frontal contour.

865 Integrating along the polar front removes the rotational component of the
866 eddy flux and gives the cumulative transport (Fig. 15c), which further under-
867 scores the importance of hot-spots in the Southern Ocean heat-transport. Unlike
868 the local heat flux, the cumulative fluxes are organized into a series of gener-
869 ally southward step-changes (although a small northward heat flux is found
870 in the vicinity of the Southwest Indian Ridge at approximately 30°E). Large
871 southward heat transports are found near the Southeast Indian Ridge south
872 of Tasmania (longitude: $\sim 145^{\circ}\text{E}$, heat transport: $\sim 0.75\text{TW}$) the Campbell
873 Plateau ($\sim 170^{\circ}\text{E}$, $\sim 1.0\text{TW}$), the Pacific Antarctic Rise ($\sim 130^{\circ}\text{W}$, $\sim 0.25\text{TW}$)
874 and through Drake Passage and the nearby Shackleton Fracture Zone/Scotia
875 Arc ($\sim 50^{\circ}\text{W}$, $\sim 0.75\text{TW}$). In total, the ANDRO dataset reveals approximately
876 $2 \pm 0.5\text{TW}$ of heat transport across the polar front at 1000m depth. More than
877 90% of the total heat transport is occurs in less than 20% of the total longitudes
878 spanned by the contour.

879 As Dufour et al. (2015) found in their high resolution numerical model, the
880 ANDRO data reveal that the eddy heat flux is strongly concentrated in ‘hot-
881 spot’ regions near large bathymetric features. The concordance between our
882 results and those of Dufour et al. (2015) is remarkable, given the supposed
883 sparseness of the Argo float observations in the ocean. However, the results
884 of the modelling component of this study give us confidence that the results
885 presented in this section are valid, although subject to error. Improvement
886 of the mapping procedure, as well as the inclusion of additional deep drifter
887 datasets, such as RAFOS floats, could further increase confidence in the results
888 presented here.

889 7. Discussion and Conclusions

890 In this paper we have used ‘virtual’ Argo floats advected in an idealized
891 model of the Southern Ocean to critically assess the ability of the existing
892 Argo array to reconstruct both the time-mean and eddying quantities of im-
893 portance to the general circulation. Comparing time-mean and eddy quantities
894 reconstructed from the virtual Argo floats directly to the model fields reveals
895 that, at float observation densities similar to those available from the Argo ar-
896 ray, it is possible to robustly reconstruct several important quantities, including
897 quadratic perturbation quantities such as the EKE and eddy heat flux. We have
898 tested, systematically, the influence of temporal sampling frequency, the num-
899 ber of floats and the time span of the float experiment, and found, in all case,
900 that robust reconstructions of the these quantities is possible, even when rela-
901 tively ‘pessimistic’ values of these parameters are chosen. We have also shown
902 that is is also possible to reconstruct cross-frontal eddy heat fluxes using only
903 the Lagrangian floats, but only for specially defined frontal contours that may
904 not necessarily form closed circumpolar contours. As such, this study echoes
905 previous work (Davis, 1987, 1991b) who showed that comparatively few surface
906 drifters were required to resolve an idealized thin western boundary current.

907 The key result of this study is that, with a sufficient number of floats tracked
908 over a sufficiently long period of time, one can reconstruct with a high degree
909 of fidelity both time-mean fields and the local eddy statistics. The challenge
910 is, of course, to define how long a ‘sufficiently’ long time period is, and how
911 many floats are ‘sufficient’. There are no clear answers to these questions, as
912 any response would depend on the needs of the particular study. However, the
913 results of the numerical modeling portion of this study indicate that the cur-
914 rent observational coverage and sampling rates provided by Argo floats in the
915 Southern Ocean return reconstruction errors that are not substantially improved
916 by the addition of more floats or longer float experiments (although extending
917 the life of the Argo project is essential for long term climate monitoring), and
918 only marginally improved by increasing the sampling rate. The reasons for

919 the observed asymptotic error performance are not clear. However, a similar
920 OSSE performed by Kamenkovich et al. (2011) noted that reconstruction errors
921 were generally smaller in regions with higher absolute current speeds, where the
922 oceanic ‘signal’ is able to dominant the ‘noise’ introduced by the reconstruction
923 error. In the Southern Ocean, where currents are consistently strong, the signal
924 to noise ratio could well be large enough that the oceanic signal can be defined
925 with relatively few samples. In our model study, there are relatively small differ-
926 ences in the reconstruction error of the time mean meridional velocity between
927 the eastern (where time mean currents are weaker) and western (where time
928 mean currents are stronger) sub-domains, despite the enhanced variability in
929 the later region, which provides some limited evidence that the effect described
930 by Kamenkovich et al. (2011) may explain the relative insensitive of the re-
931 construction error to the number of samples - provided a minimum number of
932 samples has been obtained.

933 Although we have shown that increasing the float profiling rate (and hence
934 sampling rate) results a reduction in the error of the resulting estimates of
935 the both mean and eddying quantities, doing so would, in reality, reduce the
936 lifetimes of the floats which are generally inversely proportional to the number
937 of cycles (Roemmich et al., 2009). Although increasing the sampling rate would
938 not necessarily reduce the total number of profiles collected by a particular float,
939 it would reduce the length of the float experiment and, potentially, restrict the
940 geographical range sampled by the float.

941 With the results obtained from the numerical model in mind, we have then
942 used the existing array of Argo float to compute the eddy heat-flux across the
943 Polar Front in the Southern Ocean, building upon similar work using the smaller
944 ALACE float array (Gille, 2003a,b). The numerical model allows us to construct
945 a suitable model for the errors induced by the discrete temporal sampling for
946 inclusion alongside errors due to vertical shear in the water column and uncer-
947 tainty due to internal variability within the ocean which are obtained either from
948 the ANDRO dataset or estimated directly. We find that these errors, although
949 important, do not impede the calculation of the cross-frontal eddy heat-flux

950 at this depth. Our results are qualitatively very similar to those obtained by
951 Griesel et al. (2009) and Dufour et al. (2015) in a high resolution numerical
952 model, which, together with the results from our own modeling study, allow
953 us to place a fairly high degree of confidence in the capacity of Argo floats to
954 reconstruct the heat flux and other eddy quantities.

955 Our study does, however, contain some notable shortcomings. For exam-
956 ple, the idealized model configuration was used primarily for convenience and
957 although it produces a flow field reminiscent of the that in the Southern Ocean,
958 the actual ocean circulation is, in reality, far more complex and contains numer-
959 ous phenomena unrepresented in our model. Additionally, as shown by Rosso
960 et al. (2014) in a series of progressively higher resolution numerical models, 5km
961 grid spacing is not sufficient to completely resolve the oceanic mesoscale, and
962 certainly not the energetic sub-mesoscale. As such, the length scales of im-
963 portant features in the Southern Ocean are likely smaller than can be resolved
964 by our simulation, and it is still an open question if discretely sampled Argo
965 floats would be able to accurately represent the eddy fluxes under these condi-
966 tions. A similar analysis to the present work, conducted using the output of a
967 high-resolution realistic model configuration, could be illuminating.

968 With the shortcomings of this study noted, we finish on a note of optimism:
969 the evidence presented here suggests that the current Argo array is able to gener-
970 ate reliable eddy statistics and that the addition of additional floats to the
971 system are not strictly necessary for this purpose, as they are not likely to
972 dramatically improve the capacity of the array to represent meso-scale statis-
973 tics, although additional floats are likely to aid resolving important features in
974 undersampled regions. Thus, a promising avenue of future research is to ex-
975 ploit the Argo array to close local tracer budgets. In particular, the flux of
976 biogeochemical tracers across fronts, a quantity of great importance to the cli-
977 mate system, could be estimated using the developing array of 'bio-Argo' floats,
978 capable of measuring biogeochemical quantities such as carbon and nutrients.
979 Additionally, with the continuing improvement and maintenance of the Argo
980 array, long term monitoring of eddy quantities over broad regions may also

981 be possible.

982 **Acknowledgement**

983 C.C. was funded by a National Science Foundation Division of Ocean Sci-
984 ences Postdoctoral Fellowship 1521508. J.B.S. received support from Agence
985 Nationale de la Recherche (ANR), grant number ANR-12-PDOC-0001. Numer-
986 ical calculations were undertaken on the OCCIGEN supercomputer man-
987 aged by the French *Centre Informatique National de l'Enseignement Supérieur*.
988 Supercomputing resources were allocated by the *Grand Équipement National*
989 *de Calcul Intensif* under the SO-BRIGHT project. The authors acknowledge
990 two anonymous reviewers whose comments greatly improved the manuscript,
991 as well as a third anonymous reviewer on a previous version of this paper who
992 explained the importance of consistent flux streamline estimation.

993 **References**

- 994 Abernathey, R., Cessi, P., 2014. Topographic enhancement of eddy efficiency in
995 baroclinic equilibration. *Journal of Physical Oceanography* 44, 2107–2126.
- 996 Abernathey, R., Marshall, J., Ferreira, D., 2011. The dependence of southern
997 ocean meridional overturning on wind stress. *Journal of Physical Oceanogra-*
998 *phy* 41, 2261–2278. doi:10.1175/JPO-D-11-023.1.
- 999 Amante, C., Eakins, B., 2009. ETOPO1 1 Arc-Minute Global Relief Model: Pro-
1000 cedures, Data Sources and Analysis. NOAA Technical Memorandum NES-
1001 DIS NGDC-24.. National Geophysical Data Center, NOAA. doi:10.7289/
1002 V5C8276M.
- 1003 Arbic, B.K., Mller, M., Richman, J.G., Shriver, J.F., Morten, A.J., Scott,
1004 R.B., Srazin, G., Penduff, T., 2014. Geostrophic turbulence in the frequen-
1005 cywavenumber domain: Eddy-driven low-frequency variability. *Journal of*
1006 *Physical Oceanography* 44, 2050–2069. doi:10.1175/JPO-D-13-054.1.

- 1007 Arbic, B.K., Scott, R.B., Flierl, G.R., Morten, A.J., Richman, J.G., Shriver,
1008 J.F., 2012. Nonlinear cascades of surface oceanic geostrophic kinetic energy
1009 in the frequency domain. *Journal of Physical Oceanography* 42, 1577–1600.
1010 doi:10.1175/JPO-D-11-0151.1.
- 1011 Barnier, B., Siefridt, L., Marchesiello, P., 1995. Thermal forcing for a global
1012 ocean circulation model using a three-year climatology of ecmwf analyses.
1013 *Journal of Marine Systems* 6, 363–380.
- 1014 Chapman, C.C., 2014. Southern ocean jets and how to find them: Improving and
1015 comparing common jet detection methods. *Journal of Geophysical Research:*
1016 *Oceans* 119, 4318–4339. doi:10.1002/2014JC009810.
- 1017 Chapman, C.C., 2017. New perspectives on frontal variability in the south-
1018 ern ocean. *Journal of Physical Oceanography* 47, 1151–1168. doi:10.1175/
1019 JPO-D-16-0222.1.
- 1020 Chapman, C.C., Hogg, A.M., Kiss, A.E., Rintoul, S.R., 2015. The dynamics of
1021 southern ocean storm tracks. *Journal of Physical Oceanography* 45, 884–903.
- 1022 Davis, R., 1982. On relating eulerian and lagrangian velocity statistics: Single
1023 particles in homogeneous flows. *Journal of Fluid Mechanics* 114, 1–26. doi:10.
1024 1017/S0022112082000019.
- 1025 Davis, R.E., 1987. Modeling eddy transport of passive tracers. *Journal of Marine*
1026 *Research* 45, 635–666.
- 1027 Davis, R.E., 1991a. Lagrangian ocean studies. *Annual Review of Fluid Mechan-*
1028 *ics* 23, 43–64. doi:10.1146/annurev.fl.23.010191.000355.
- 1029 Davis, R.E., 1991b. Observing the general circulation with floats. *Deep Sea*
1030 *Research Part A. Oceanographic Research Papers* 38, S531 – S571. doi:10.
1031 1016/S0198-0149(12)80023-9.
- 1032 Davis, R.E., 1998. Preliminary results from directly measuring middepth cir-
1033 culation in the tropical and south pacific. *Journal of Geophysical Research:*
1034 *Oceans* 103, 24619–24639. doi:10.1029/98JC01913.

- 1035 Dufour, C.O., Griffies, S.M., de Souza, G.F., Frenger, I., Morrison, A.K., Palter,
1036 J.B., Sarmiento, J.L., Galbraith, E.D., Dunne, J.P., Anderson, W.G., et al.,
1037 2015. Role of mesoscale eddies in cross-frontal transport of heat and biogeo-
1038 chemical tracers in the southern ocean. *Journal of Physical Oceanography*
1039 45, 3057–3081.
- 1040 Elipot, S., Lumpkin, R., Perez, R.C., Lilly, J.M., Early, J.J., Sykulski, A.M.,
1041 2016. A global surface drifter data set at hourly resolution. *Journal of Geo-*
1042 *physical Research: Oceans* 121, 2937–2966. doi:10.1002/2016JC011716.
- 1043 Fox-Kemper, B., Ferrari, R., Pedlosky, J., 2003. On the indeterminacy of ro-
1044 tational and divergent eddy fluxes. *Journal of Physical Oceanography* 33,
1045 478–483. doi:10.1175/1520-0485(2003)033<0478:OTIORA>2.0.CO;2.
- 1046 Fratantoni, D.M., Richardson, P.L., 1999. Sofar float observations of an
1047 intermediate-depth eastern boundary current and mesoscale variability in the
1048 eastern tropical atlantic ocean. *Journal of Physical Oceanography* 29, 1265–
1049 1278. doi:10.1175/1520-0485(1999)029<1265:SF00AI>2.0.CO;2.
- 1050 Gille, S.T., 2003a. Float observations of the southern ocean. part i: Estimating
1051 mean fields, bottom velocities, and topographic steering. *Journal of Physical*
1052 *Oceanography* 33, 1167–1181.
- 1053 Gille, S.T., 2003b. Float observations of the southern ocean. part ii: Eddy
1054 fluxes. *Journal of Physical Oceanography* 33, 1182–1196.
- 1055 Griesel, A., Gille, S.T., Sprintall, J., McClean, J.L., Maltrud, M.E., 2009.
1056 Assessing eddy heat flux and its parameterization: A wavenumber per-
1057 spective from a 1/10 ocean simulation. *Ocean Modelling* 29, 248 – 260.
1058 doi:http://dx.doi.org/10.1016/j.ocemod.2009.05.004.
- 1059 Hoffman, R.N., Atlas, R., 2016. Future observing system simulation exper-
1060 iments. *Bulletin of the American Meteorological Society* 97, 1601–1616.
1061 doi:10.1175/BAMS-D-15-00200.1.

- 1062 Jayne, S.R., Marotzke, J., 2002. The oceanic eddy heat transport. *Journal*
1063 *of Physical Oceanography* 32, 3328–3345. doi:10.1175/1520-0485(2002)
1064 032<3328:T0EHT>2.0.CO;2.
- 1065 Kamenkovich, I., Cheng, W., Sarachik, E.S., Harrison, D.E., 2009. Simulation
1066 of the argo observing system in an ocean general circulation model. *Journal*
1067 *of Geophysical Research: Oceans* 114. doi:10.1029/2008JC005184. c09021.
- 1068 Kamenkovich, I., Cheng, W., Schmid, C., Harrison, D.E., 2011. Effects of
1069 eddies on an ocean observing system with profiling floats: Idealized simu-
1070 lations of the argo array. *Journal of Geophysical Research: Oceans* 116,
1071 n/a–n/a. URL: <http://dx.doi.org/10.1029/2010JC006910>, doi:10.1029/
1072 2010JC006910. c06003.
- 1073 Keating, S.R., Smith, K.S., Kramer, P.R., 2011. Diagnosing lateral mixing in
1074 the upper ocean with virtual tracers: Spatial and temporal resolution de-
1075 pendence. *Journal of Physical Oceanography* 41, 1512–1534. doi:10.1175/
1076 2011JP04580.1.
- 1077 LaCasce, J., 2008. Statistics from lagrangian observations. *Progress in Oceanog-*
1078 *raphy* 77, 1–29. doi:10.1016/j.pocean.2008.02.002.
- 1079 Lebedev, K.V., Yoshinari, H., Maximenko, N.A., Hacker, P.W., 2007. YoM-
1080 aHa'07: Velocity data assessed from trajectories of argo floats at parking
1081 level and at the sea surface. IPRC Technical Note 4.
- 1082 Maas, L.R.M., 1989. A comparison of eulerian and lagrangian current mea-
1083 surements. *Deutsche Hydrografische Zeitschrift* 42, 111–132. doi:10.1007/
1084 BF02226290.
- 1085 Madec, G., 2014. Nemo ocean engine. Note du Pôle de modélisation, Institut
1086 Pierre-Simon Laplace (IPSL) 27.
- 1087 Majkut, J.D., Carter, B.R., Frölicher, T.L., Dufour, C.O., Rodgers, K.B.,
1088 Sarmiento, J.L., 2014. An observing system simulation for southern ocean

- 1089 carbon dioxide uptake. *Philosophical Transactions of the Royal Society of*
1090 *London A: Mathematical, Physical and Engineering Sciences* 372. doi:10.
1091 1098/rsta.2013.0046.
- 1092 Marshall, J., Shutts, G., 1981. A note on rotational and divergent eddy
1093 fluxes. *Journal of Physical Oceanography* 11, 1677–1680. doi:10.1175/
1094 1520-0485(1981)011<1677:ANORAD>2.0.CO;2.
- 1095 McDougall, T.J., Barker, P.M., 2011. Getting started with TEOS-10 and the
1096 Gibbs Seawater (GSW) oceanographic toolbox. SCOR/IAPSO WG 127, 1–
1097 28.
- 1098 Middleton, J.F., 1985. Drifter spectra and diffusivities. *Journal of Marine*
1099 *Research* 43, 37–55. doi:doi:10.1357/002224085788437334.
- 1100 Ollitrault, M., Rannou, J.P., 2013. Andro: An argo-based deep displacement
1101 dataset. *Journal of Atmospheric and Oceanic Technology* 30, 759–788.
- 1102 Ollitrault, M., Colin de Verdière, A., 2014. The ocean general circulation near
1103 1000-m depth. *Journal of Physical Oceanography* 44, 384–409.
- 1104 Richardson, P.L., Fratantoni, D.M., 1999. Float trajectories in the deep western
1105 boundary current and deep equatorial jets of the tropical atlantic. *Deep Sea*
1106 *Research Part II: Topical Studies in Oceanography* 46, 305 – 333. doi:http:
1107 //dx.doi.org/10.1016/S0967-0645(98)00100-3.
- 1108 Ridgway, K., Dunn, J., Wilkin, J., 2002. Ocean interpolation by four-
1109 dimensional weighted least squares-application to the waters around australa-
1110 sia. *Journal of atmospheric and oceanic technology* 19, 1357–1375.
- 1111 Rintoul, S.R., Garabato, A.C.N., 2013. Chapter 18 - dynamics of the southern
1112 ocean circulation, in: Gerold Siedler, Stephen M. Griffies, J.G., Church, J.A.
1113 (Eds.), *Ocean Circulation and Climate A 21st Century Perspective*. Academic
1114 Press. volume 103 of *International Geophysics*, pp. 471 – 492. doi:10.1016/
1115 B978-0-12-391851-2.00018-0.

- 1116 Riser, S.C., Freeland, H.J., Roemmich, D., Wijffels, S., Troisi, A., Belbéoch, M.,
1117 Gilbert, D., Xu, J., Pouliquen, S., Thresher, A., et al., 2016. Fifteen years
1118 of ocean observations with the global argo array. *Nature Climate Change* 6,
1119 145–153.
- 1120 Roach, C.J., Balwada, D., Speer, K., 2016. Horizontal mixing in the southern
1121 ocean from argo float trajectories. *Journal of Geophysical Research: Oceans*
1122 121, 5570–5586. doi:10.1002/2015JC011440.
- 1123 Roemmich, D., Johnson, G.C., Riser, S., Davis, R., Gilson, J., Owens,
1124 W.B., Garzoli, S.L., Schmid, C., Ignaszewski, M., 2009. The argo pro-
1125 gram: Observing the global ocean with profiling floats. *Oceanography* 22.
1126 doi:10.5670/oceanog.2009.36.
- 1127 Rosso, I., Hogg, A.M., Strutton, P.G., Kiss, A.E., Matear, R., Klocker, A.,
1128 van Sebille, E., 2014. Vertical transport in the ocean due to sub-mesoscale
1129 structures: Impacts in the kerguelen region. *Ocean Modelling* 80, 10 – 23.
1130 doi:10.1016/j.ocemod.2014.05.001.
- 1131 Rupolo, V., 2007. A lagrangian-based approach for determining trajectories
1132 taxonomy and turbulence regimes. *Journal of Physical Oceanography* 37,
1133 1584–1609. doi:10.1175/JP03038.1.
- 1134 Rupolo, V., Artale, V., Hua, B.L., Provenzale, A., 1996. Lagrangian ve-
1135 locity spectra at 700 m in the western north atlantic. *Journal of Physi-
1136 cal Oceanography* 26, 1591–1607. doi:10.1175/1520-0485(1996)026<1591:
1137 LVSAMI>2.0.CO;2.
- 1138 van Sebille, E., Johns, W.E., Beal, L.M., 2012. Does the vorticity flux from
1139 agulhas rings control the zonal pathway of nadw across the south atlantic?
1140 *Journal of Geophysical Research: Oceans* 117. doi:10.1029/2011JC007684.
1141 c05037.
- 1142 van Sebille, E., Kamenkovich, I., Willis, J.K., 2011. Quasi-zonal jets in 3-d

- 1143 argo data of the northeast atlantic. *Geophysical Research Letters* 38. doi:10.
1144 1029/2010GL046267. 102606.
- 1145 Smith, S., 1997. *The Scientist and Engineer's Guide to Digital Signal Process-*
1146 *ing*. California Technical Pub. URL: [https://books.google.fr/books?id=](https://books.google.fr/books?id=rp2VQgAACAAJ)
1147 [rp2VQgAACAAJ](https://books.google.fr/books?id=rp2VQgAACAAJ).
- 1148 Taylor, G.I., 1938. The spectrum of turbulence. *Proceedings of the Royal*
1149 *Society of London A: Mathematical, Physical and Engineering Sciences* 164,
1150 476–490. doi:10.1098/rspa.1938.0032.
- 1151 Thompson, A.F., Sallée, J.B., 2012. Jets and topography: Jet transitions and
1152 the impact on transport in the antarctic circumpolar current. *Journal of*
1153 *physical Oceanography* 42, 956–972.
- 1154 Williams, R.G., Wilson, C., Hughes, C.W., 2007. Ocean and atmosphere storm
1155 tracks: The role of eddy vorticity forcing. *Journal of Physical Oceanography*
1156 37, 2267–2289.
- 1157 Willis, J.K., Fu, L.L., 2008. Combining altimeter and subsurface float data
1158 to estimate the time-averaged circulation in the upper ocean. *Journal of*
1159 *Geophysical Research: Oceans* 113. doi:10.1029/2007JC004690. c12017.
- 1160 Wunch, C., 2006. *Discrete Inverse and State Estimation Problems*. Cambridge
1161 University Press, New York. URL: www.cambridge.org/9780521854245.