Can We Reconstruct Mean and Eddy Fluxes from Argo Floats?

Christopher C. Chapman, Jean-Baptiste Sallée

Accepted Manuscript

Can We Reconstruct Mean and Eddy Fluxes from Argo Floats?

Christopher Chapman, Jean-Baptiste Sallée

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
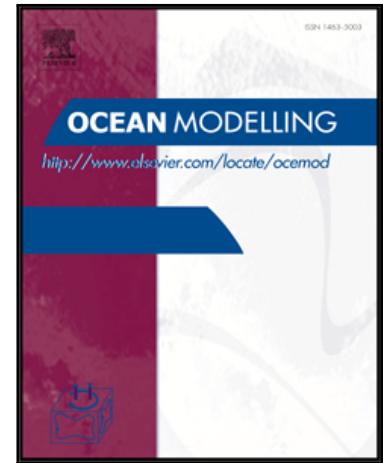
Highlights

- 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 re-construct 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.28gmail.com` (Christopher Chapman)

bathymetric features, consistent with the results of previous studies. 2±0.5TW 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. **1. Introduction**

2 Deep drifting floats, such the satellite tracked Argo floats and the Au-
3 tonomous Lagrangian Circulation Explorer (ALACE) floats, or acoustically
4 tracked Sound Fixing and Ranging (SOFAR) and RAFOS floats, provide direct
5 measurements of the oceanic currents as they move with the flow. These floats
6 provide the only direct measurements of the ocean's subsurface currents with
7 broad spatial coverage and have been instrumental in shaping our comprehen-
8 sion of the structure of the ocean's interior (Roemmich et al., 2009; Riser et al.,
9 2016). They have been shown to be capable of producing accurate and rich maps
10 of the time mean interior currents (Davis, 1991a; Gille, 2003a; LaCasce, 2008;
11 Ollitrault and Colin de Verdiére, 2014) and measurements of features not read-
12 ily inferred from remotely sensed surface measurements, such as deep jets and
13 boundary currents (Richardson and Fratantoni, 1999; Fratantoni and Richard-
14 son, 1999; van Sebille et al., 2011, 2012). With the continued development of
15 the Argo program, and the improved geographical and temporal coverage of
16 the global ocean that comes with it, it is reasonable to ask: are we capable of
17 observing robust, quantitative statistics of the oceanic meso-scale with current
18 float deployments?

19 As a way of introducing the problem, Fig. 1a shows all Argo float positions
20 in the Southern Ocean (south of 30°S) within 5 days of the 25th of December
21 2009. We have determined the mean distance between each of the points plotted
22 in Fig. 1a and their closest neighbor is approximately 160km. Fig. 1b shows
23 the trajectory of a single float, (World Meteorological Organization number

3
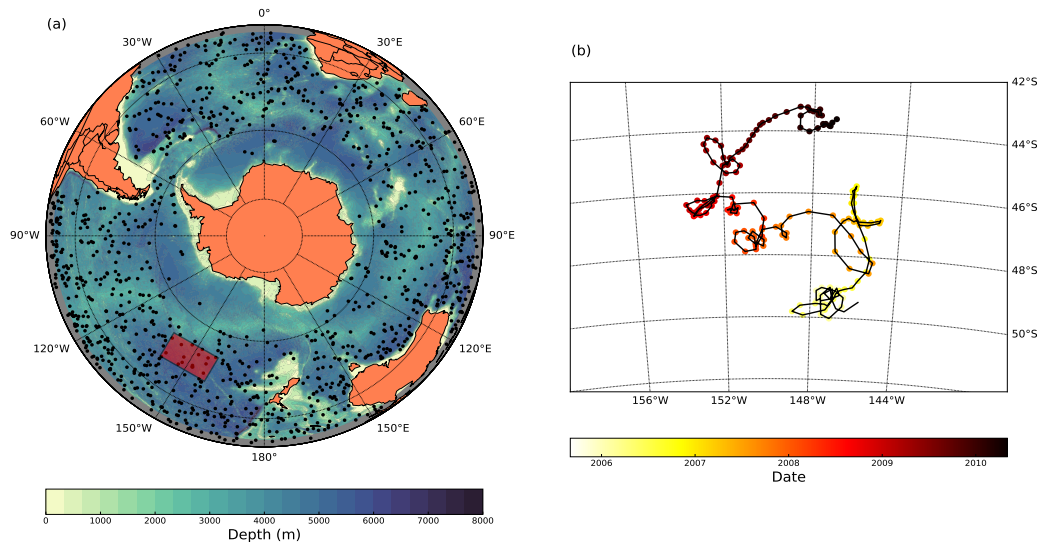
Figure 1: Spatial coverage of Argo floats in the Southern Ocean. (a) All reported Argo float positions at 1000db, within 5 days of the 25th of December, 2009 (points), overlayed over the topography from the ETOPO01 dataset (colored contours); (b) zoom on the highlighted region in panel (a), showing the 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,

4

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

5

and resurfacing of the floats, the spatial density of the Lagrangian array and the duration of the float experiment will be in place by systematically modifying the density and time span of the virtual float array, as well as the sampling characteristics of the float derived velocities. Specifically, an idealized, eddy resolving numerical model of the Southern Ocean is "observed" using "virtual" Argo floats. By comparing the Lagrangian derived estimates of mean velocity, eddy kinetic energy and heat flux to the "exact" results from the model solution, we will demonstrate the utility and shortcomings of these Lagrangian measurements. We will then use the understanding of the limitations of the Lagrangian derived velocities gained from the model output in order to estimate the eddy heat flux in the Southern Ocean from the existing array of Argo floats. We limit our focus to the Southern Ocean for two primary reason: it is the principle region of study for both authors of this paper; and the lack of available "traditional" observations from ships means that a detailed investigation of the Argo floats' capacity to resolve meso-scale statistics is warranted. However, the results obtained here are expected to apply quite generally.

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

6

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

The remainder of this article is organized as follows: Section 2 discusses the spatial and temporal sampling characteristics of Argo floats and the effects of each on the estimation of underlying flow fields. The numerical model configuration and the method of advecting virtual Argo floats, as wll as the observational datasets that will be used in the second part of this study will be described in Section 3. We will discuss the reconstruction of the numerical model fields from the virtual Argo floats in section 4, and the ability of Lagrangian observations to determine cross-frontal heat fluxes in the numerical model in section 5. Estimates of the mean flow and the cross-stream eddy heat-flux in the Southern Ocean using the Argo float array will be presented in section 6, and the results obtained will be discussed with reference to the numerical model results in section 7.

2. Sampling and Lagrangian Drifters

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

2.1. Temporal Sampling and Lagrangian Drifters

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

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

7

₁₂₅ However, due to their dive, drift and resurface cycle, Argo floats are sampled at

₁₂₆ discreet time intervals. To determine the current velocity from the Lagrangian

₁₂₇ measurements, for the mth float cycle, we follow Lebedev et al. (2007) and use

₁₂₈ the following difference equation:

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

₁₂₉ where $\mathbf{x}_{\text{deep}}$, 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:

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

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

₁₃₈ Eqns. 2 and 3 clearly represent a discreet 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
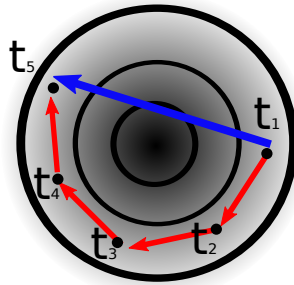
Figure 2: Schematic showing the aliasing of float observations. The shaded background and black, solid contours show the streamfunction of an idealised anticyclonic eddy. The black dots represent the float location at 5 seperate times. The red arrows represent the velocity observations determined at the highest availble sampling rate: $\Delta t = t_2 - t_1, t_3 - t_2, \ldots$. 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 are 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 $1° \times 1°$,
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.

9

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 ∼90°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 than 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.
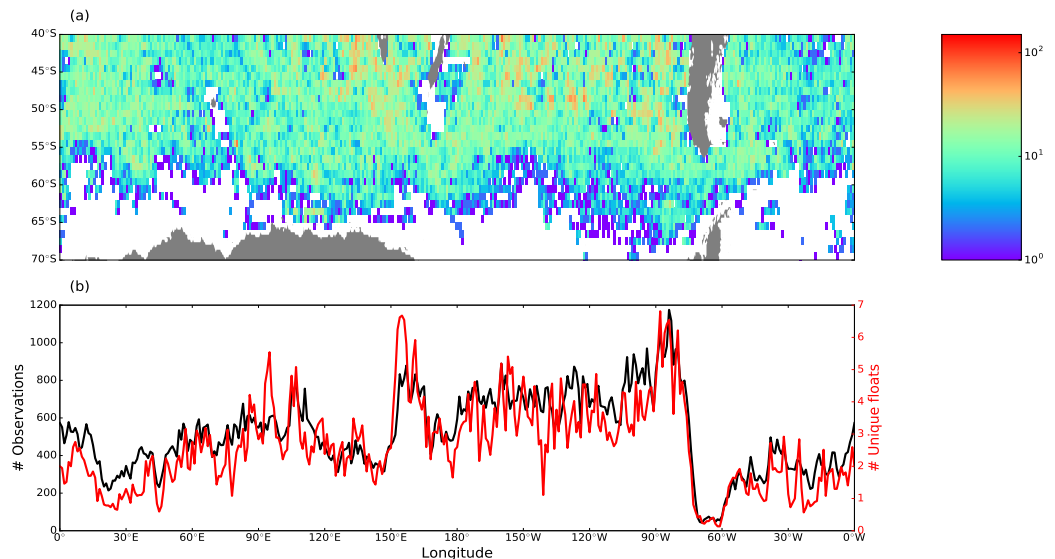
10

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 lgrid; 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 availble in a 10 day "snapshot" period, in each 1° longitude bin (red).

3. Numerical Model, Argo Data, and Methods

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

3.1. Numerical Model Configuration

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

11

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

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

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

where h is the height of the bathymetry above the ocean floor, x is the zonal coordinate, σ_0=150km is the topographic length scale and $H_0 = 2000m$ is the scale height of the topographic obstacle. The scale height and topographic length scales has been chosen to effectively block lower layer flow and induce a large stationary meander, thus effectively capturing some of the impacts of large bathymetric features, such as the Kerguelen Plateau, on the Southern Ocean circulation.

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

12

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
Δx, Δy	5km	grid-spacing
Δt	300s	barotropic time-step
H	4000m	Depth
H_0	2000m	Topography Scale Height
f_0	$-1.0\times10{-4}\mathrm{s}^{-1}$	Coriolis parameter
β	$1\times10^{-11}\mathrm{s}^{-1}\mathrm{m}^{-1}$	Meridional
τ_0	$1.5\times10^{-4}\mathrm{N.m}^{-2}$	Peak wind stress
r_D	$1.5\times10^{-3}\mathrm{m.s}^{-1}$	Linear bottom drag
κ_v	$0.5\times10^{-5}\mathrm{m.s}^{-2}$	Vertical diffusivity
T_S	7 days^{-1}	Sponge layer relaxation time-scale
α	2.0×10^{-4}	Thermal expansion coefficient
g	$30\mathrm{W.m}^{-2}\mathrm{K}^{-1}$	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
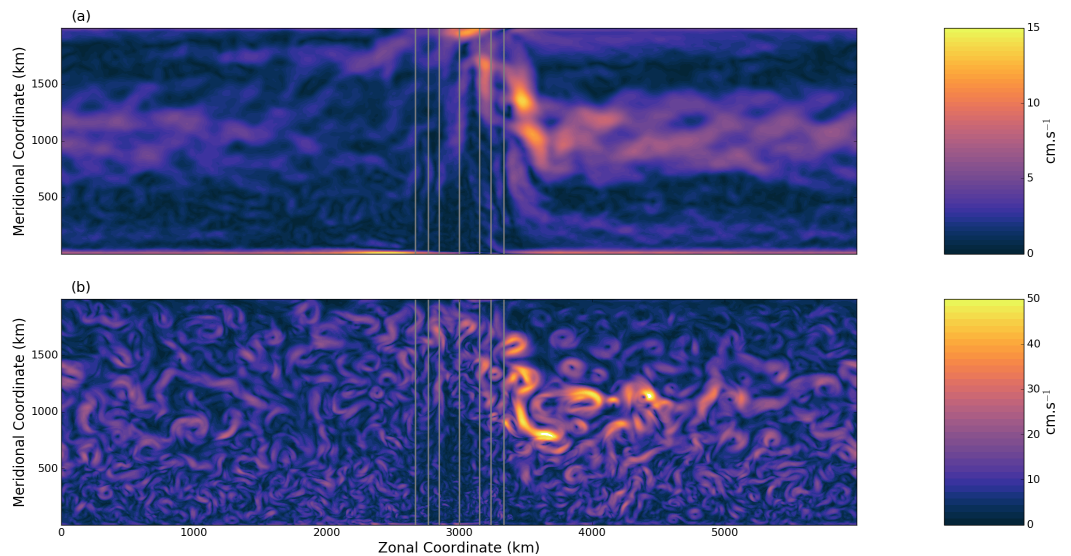
13

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)

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

3.2. Virtual Argo Float Advection

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

14

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" q 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$\pm$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

15

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.

16

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 discreetized 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:

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

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

17

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

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

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

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

18

377 how the fidelity of the reconstruction will change should the Argo array continue

378 to operate with an unchanged number of floats.

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 $K \in \mathbb{Z}$:

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

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,

19

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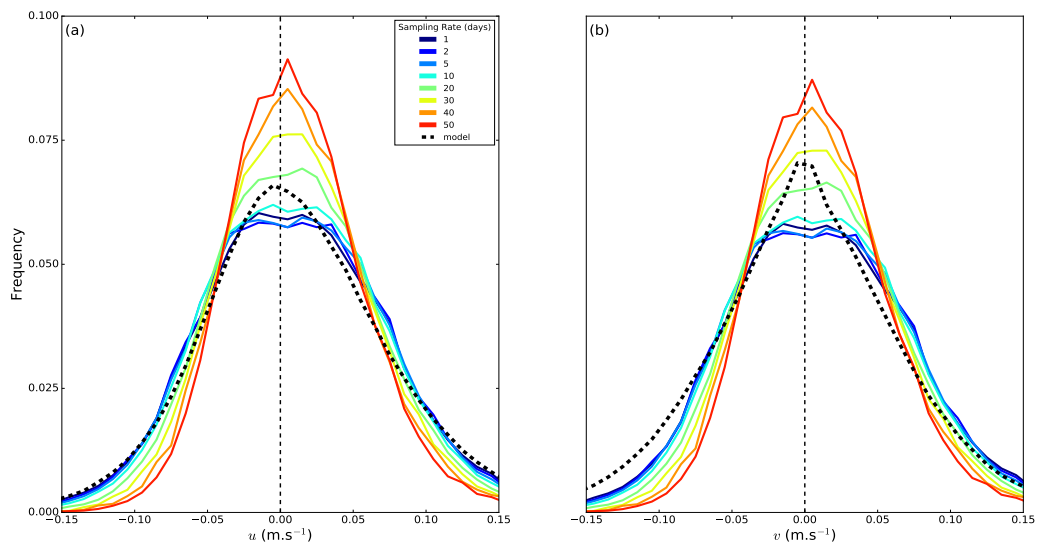

Figure 5: The normalized histograms for the zonal (a) and meridional (b) current velocities estimated by the virtual Argo floats, for each temporal sampling interval (see legend in panel (a)). The histogram computed directly from the model output at each of the float sampling points is indicated by the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

20

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:

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

₄₂₀ and the relative velocity error:

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

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 $\mathbf{x}_i \forall i \in \left[1, N_{\mathrm{obs}}\right]$, 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:

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

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 $\epsilon_v^{\mathrm{abs}}$ 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

21

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

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

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

To understand how velocity errors manifest it is instructive to examine the scatter between the virtual Argo float velocity error and the true velocity, as in Fig. 7a for sampling rates of 1, 10 and 50 days for the v velocity component (the u component has a similar structure). Fig. 7a shows that, for all cases considered, there is a significant negative correlation between ϵ_v and the velocity being measured. As such, the virtual Argo floats tend to *underestimate* strongly
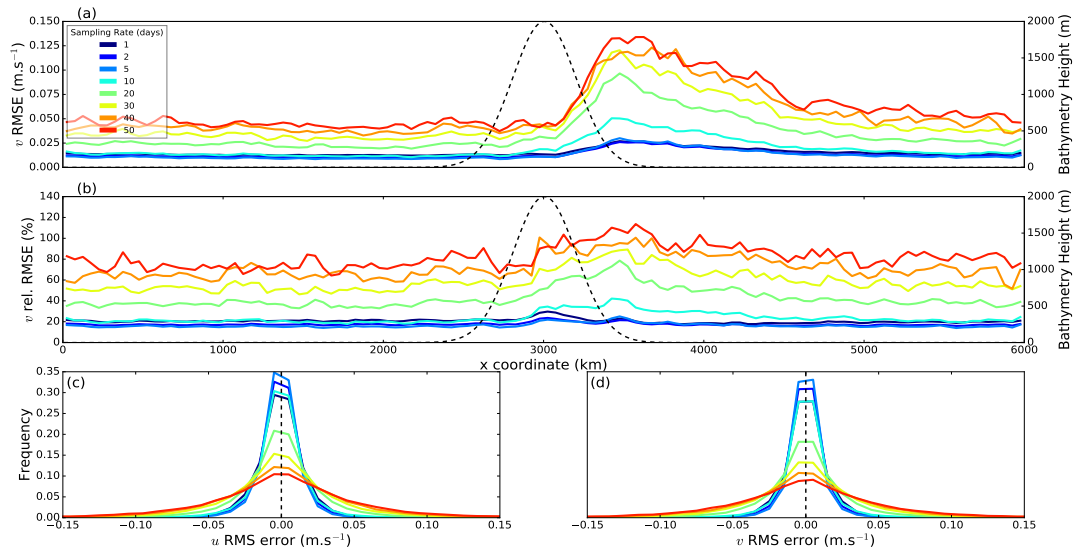
22

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 =500km and 1500km for each temporal sampling interval (colors) and; (b) as in panel (a) but for the the *relative* RMS errors in v. The thin dashed line indicates the topopgaphy 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

23

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

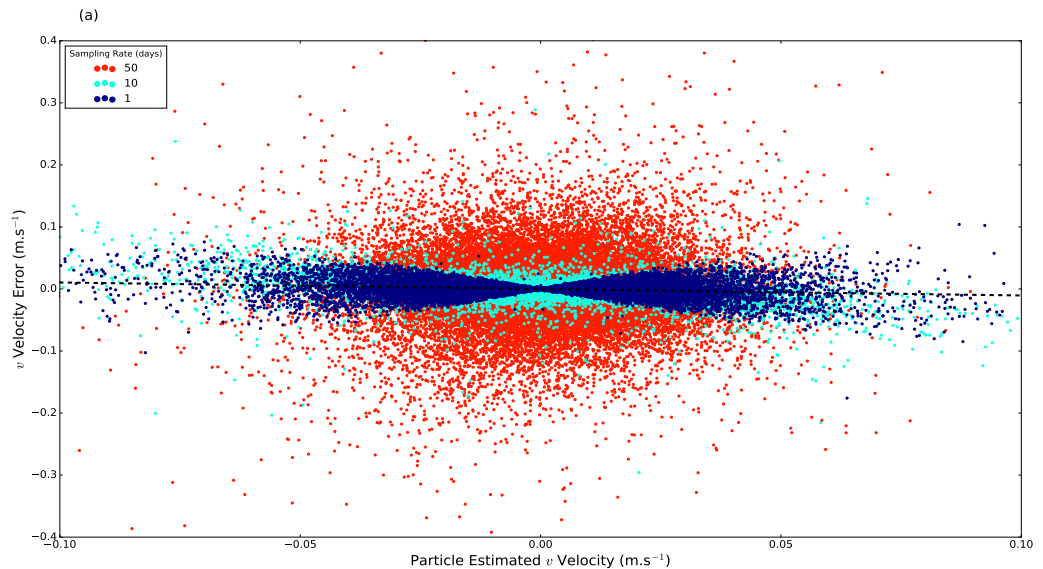⁴⁸⁴ current velocities (near $v{=}0$) deteriorates.

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 -0.3m.s^{-1}/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-

24

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

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

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

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

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

25

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} ($\sim$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
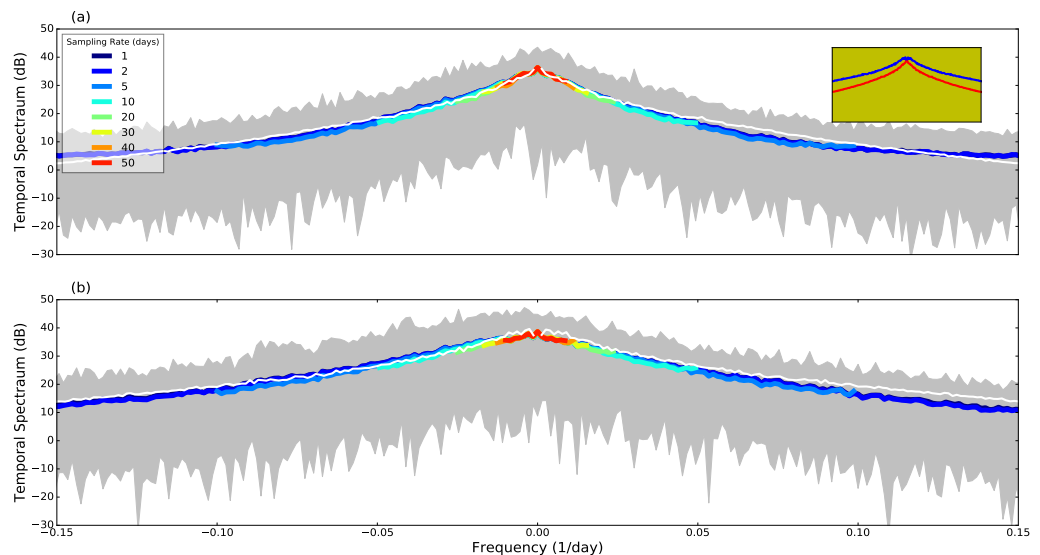⁵²⁹ floats.

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 (eddying) box. The indivdual 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 well represent the modelled current velocities and their probability distributions

26

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

4.2. Reconstruction of Mean and Eddy Fields from Lagrangian Observations

We now discuss the problem of reconstructing mean and eddy fields from noisy Lagrangian drifter velocities. We approach the problem empirically, investigating systematically the effects of changing the Lagrangian array parameters on the reconstructed fields.

4.2.1. Effect of Temporal Sampling

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

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

Fig. 9 shows good qualitative agreement between the model output and the reconstructed fields. The virtual Argo data is able to reproduce the standing wave produced by the interaction of the mean flow with topography, the magnitude and the extent of the storm track produced downstream of the topography, and the response of the meridional heat flux to both. However, there is a tendency for the virtual Argo floats to underestimate these fields, consistent with
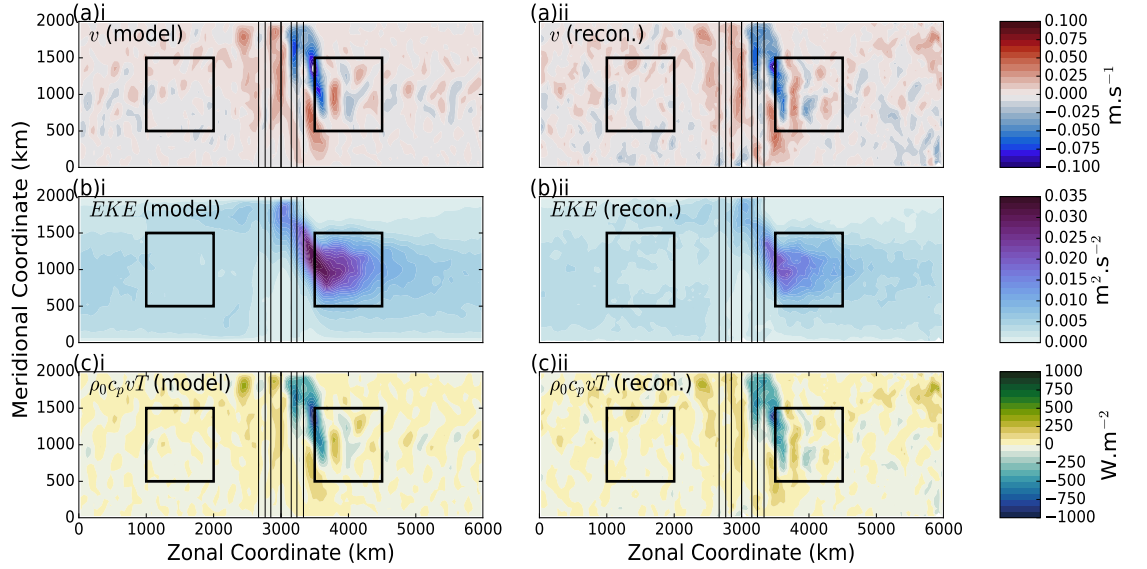
27

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 $\overline{v}$; (b) eddy kinetic energy; and (c) meridional heat flux $\rho_0 c_p \overline{vT}$ Thin black lines are bathymetric contours (CI: 500m) and the thick black lines show the boxes for the proceeding error calculations.

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

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

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

where $x_i, y_j \forall (i, j) \in \{(1, \ldots, N_x, 1 \ldots N_y)\}$ are the grid points inside the sub-domain, and N_x and N_y are the total number of grid points in the sub-domain.
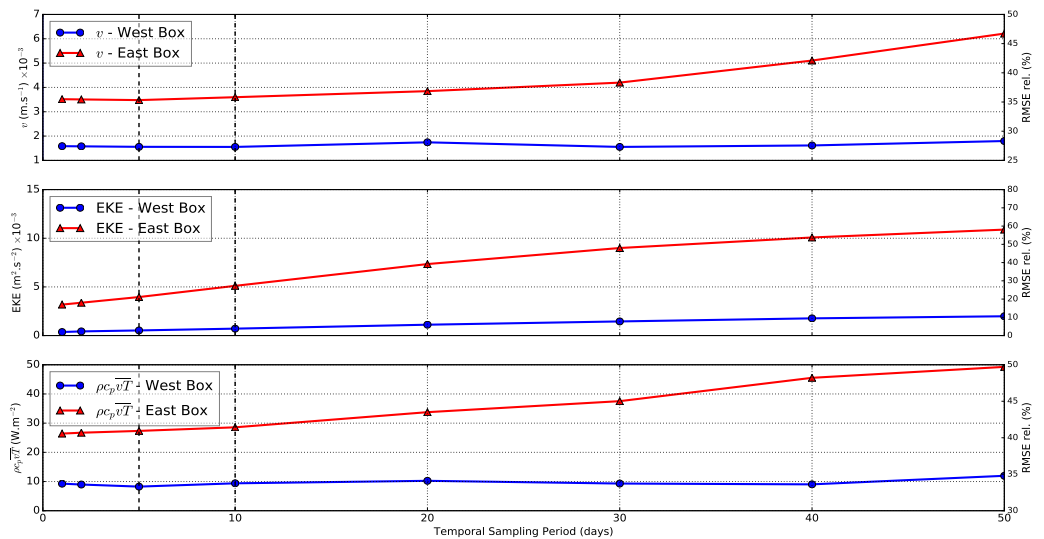
28

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 $\rho_0 c_p vT$. 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

29

this study. For example, the RMSE in the eddy heat flux (Fig. 10c) increases

from approximately 25W.m^2 at the most frequent sampling rate (1 day^{-1}), in

the (turbulent) eastern sub-domain, corresponding to a relative error of ~15% to

30W.m^2 (20% relative error) at the standard Argo sampling rate of 10.days^{-1}.

When comparing errors between the eastern and western sub-domains, it is

clear from Fig. 10 that the error is greater over the eastern storm-track box

than the western box. The error over the storm track box is ~2 to ~3 times

higher than the equivalent error over the eastern region. The larger errors in

turbulent region downstream underscore the results of section 4.1 that showed

an underestimate of the high magnitude motions.

4.2.2. Effect of the Number of Floats

We now repeat the analysis of section 4.2.1, this time investigating the in-

fluence of the number of independent floats in the virtual array. The temporal

sampling rate is held constant at the common Argo float sampling rate of 10

days. As in section 4.2.1, we compute the RMS errors between the model and re-

constructed time mean meridional velocity, EKE and heat-flux, integrated over

the two regions shown Fig. 9. The number of virtual floats is controlled by ran-

domly sampling a fraction of the float trajectories from the complete data set.

The fractions of the total number of floats selected are $1/16, 1/8, 2/8, 3/8 \ldots 1$.

The RMSE as a function of the total number of virtual floats for the each

of the reconstructions are shown in Fig. 11. As should be expected, we find a

decreasing RMSE for each of the fields considered with an increasing number

of virtual floats. The RMSE, however, begins to approach a constant limit in

both regions as the number of floats passes approximately 150. For example, as

the number of floats increases from 28 (the smallest number used) to 112, the

errors in the heat flux decrease from ~120W.m^{-2} (a relative error of 50%)to

20Wm^{-2} (20% relative error) in the western (quiet) box and ~180W.m^{-2} (65%

relative error) to 40Wm^{-2} (30% relative error) in the western (storm track)

box, a decrease of approximately 80% in both cases. However, increasing the

number of floats fourfold from 112 to 450 results in RMSE reductions of between

30

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

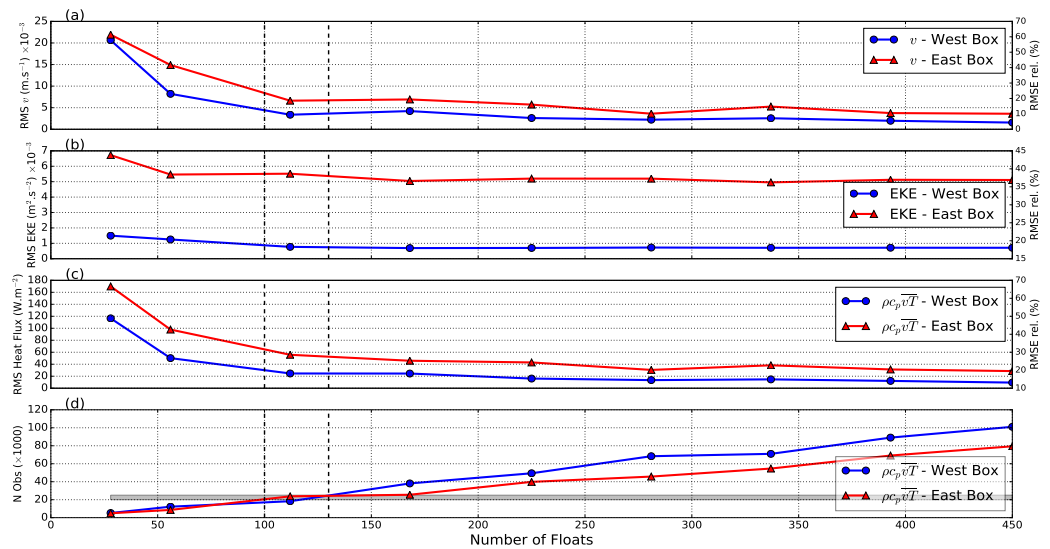₆₁₃ increasing number of floats.

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

31

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 meridonal 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,

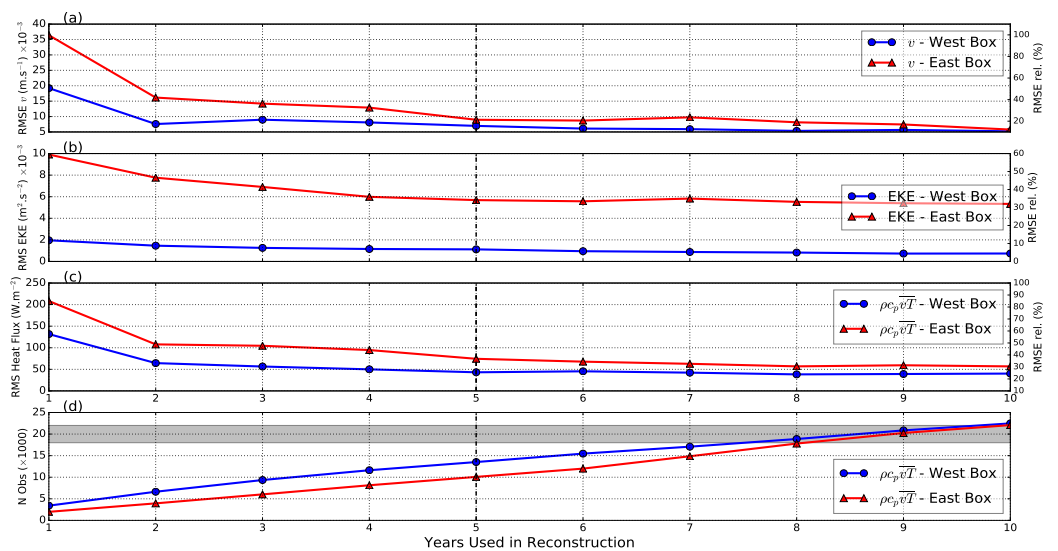₆₃₀ where the RMSE decreased by 60-70%.

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

32

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.

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:

$$\frac{d\overline{\mathbf{X}}}{ds} = \overline{\mathbf{F}}^{\theta}, \tag{11}$$

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 $\mathbf{F}^{\theta} \cdot \eta = 0$ everywhere along the curve, where η 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:

$$y_1 - y_0 = \oint \frac{dY}{ds} ds = \oint v\theta \, ds \tag{12}$$

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

33

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

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

$$\overline{\mathbf{F}}^\theta = \overline{\mathbf{u}}\overline{\theta} + \overline{\mathbf{u}'\theta'}. \tag{13}$$

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

$$\frac{d\overline{\mathbf{X}}}{ds} = \overline{\mathbf{u}}\overline{\theta}, \tag{14}$$

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

34

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:

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

709 In Eqn. 14 c_p =4.0J.kg^{-1}.K^{-1} is the specific heat of sea-water at constant

710 pressure, and ρ =1024kg.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

35

₇₁₄ in Eqn. 14 is taken from the reconstructed mean velocity field, $\overline{\mathbf{u}}$ 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

36

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 Abernathey

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

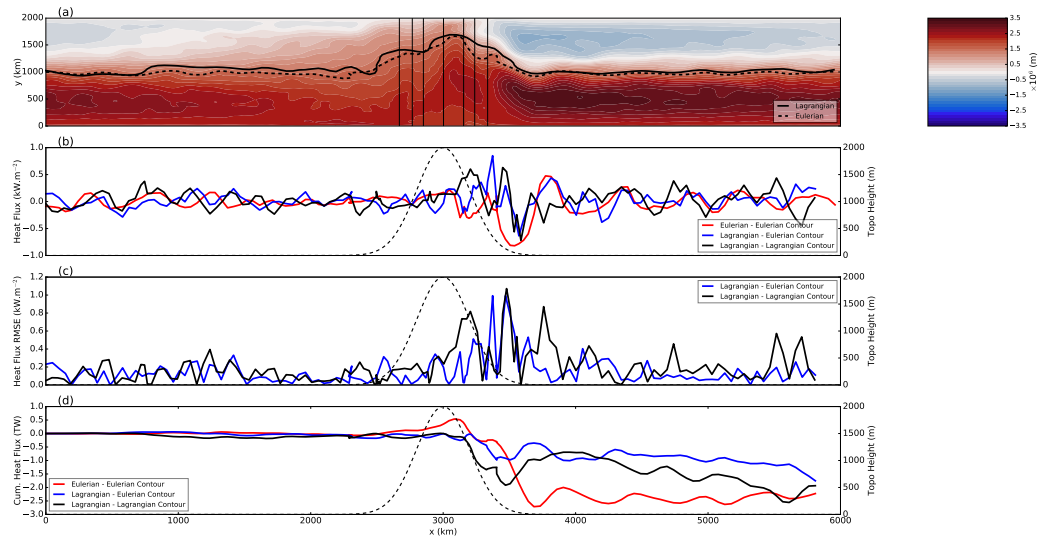⁷⁷⁵ to the heat flux estimated directly from the model output. It is also notable that

37

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 stream-function 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.

38

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

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

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

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

39

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 Verdiére (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
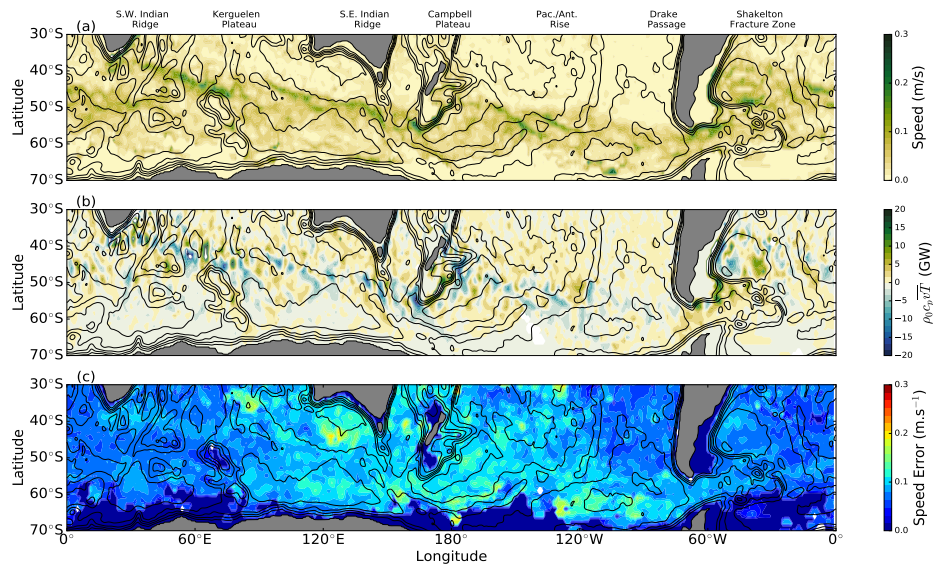833 numerical model.

40

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)

6.2. Cross-Frontal Eddy Heat Flux

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

41

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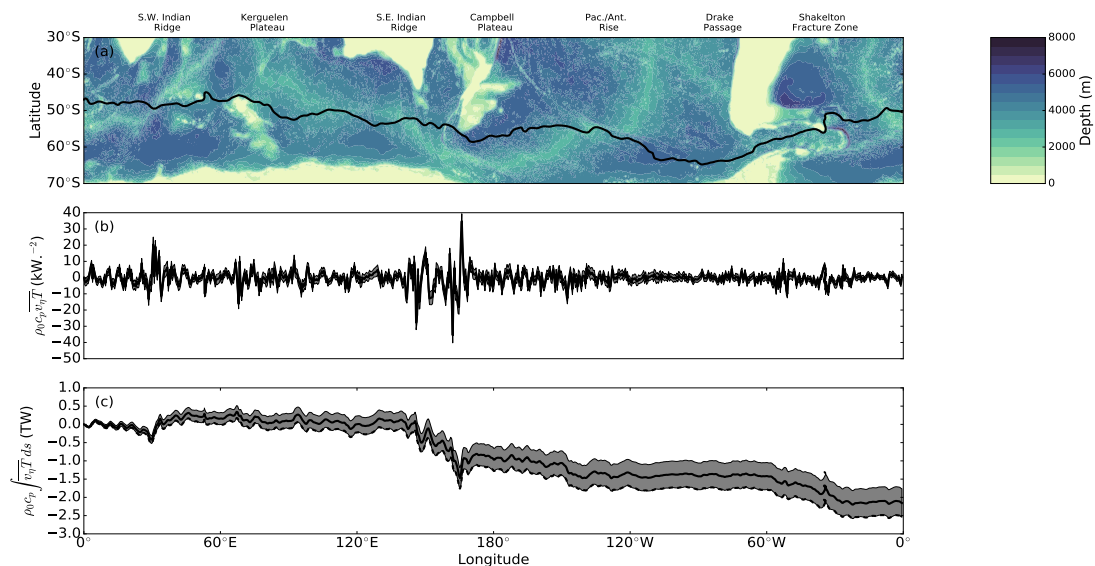
Figure 15: Heat-flux across the polar front estimated from the ANDRO dataset and the Argo temperature profiles. (a) the time-mean position of the polar front overlaying the bathymetry from the ETOPO01 dataset; (b) the cross frontal heat-flux ($\rho c_p \overline{vT} \Delta s \Delta z$); (c) the cumulative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

42

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

Integrating along the polar front removes the rotational component of the eddy flux and gives the cumulative transport (Fig. 15c), which further underscores the importance of hot-spots in the Southern Ocean heat-transport. Unlike the local heat flux, the cumulative fluxes are organized into a series of generally southward step-changes (although a small northward heat flux is found in the vicinity of the Southwest Indian Ridge at approximately $30°E$). Large southward heat transports are found near the Southeast Indian Ridge south of Tasmania (longitude: $\sim145°E$, heat transport: $\sim0.75TW$) the Campbell Plateau ($\sim170°E$, $\sim1.0TW$), the Pacific Antarctic Rise ($\sim130°W$, $\sim0.25TW$) and through Drake Passage and the nearby Shackleton Fracture Zone/Scotia Arc ($\sim50°W$, $\sim0.75TW$). In total, the ANDRO dataset reveals approximately $2\pm0.5TW$ of heat transport across the polar front at 1000m depth. More than 90% of the total heat transport is occurs in less than 20% of the total longitudes spanned by the contour.

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

43

7. Discussion and Conclusions

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

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

44

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

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

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

45

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

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

With the shortcomings of this study noted, we finish on a note of optimism: the evidence presented here suggests that the current Argo array is able to generate reliable eddy statistics and that the addition of additional floats to the system are not strictly necessary for this puropose, as they are not likely to dramatically improve the capacity of the array to represent meso-scale statistics, although additional floats are likely to aid resolving important features in undersampled regions. Thus, a promising avenue of future research is to exploit the Argo array to close local tracer budgets. In particular, the flux of biogeochemical tracers across fronts, a quantity of great importance to the climate system, could be estimated using the developing array of 'bio-Argo' floats, capable of measuring biogeochemical quantities such as carbon and nutrients. Additionally, with the continuing improvement and maintenance of the Argo array, long term monitoring of eddying quantities over broad regions may also

46

be possible.

References

Abernathey, R., Cessi, P., 2014. Topographic enhancement of eddy efficiency in baroclinic equilibration. Journal of Physical Oceanography 44, 2107–2126.

Abernathey, R., Marshall, J., Ferreira, D., 2011. The dependence of southern ocean meridional overturning on wind stress. Journal of Physical Oceanography 41, 2261–2278. doi:10.1175/JPO-D-11-023.1.

Amante, C., Eakins, B., 2009. ETOPO1 1 Arc-Minute Global Relief Model: Procedures, Data Sources and Analysis. NOAA Technical Memorandum NESDIS NGDC-24.. National Geophysical Data Center, NOAA. doi:10.7289/V5C8276M.

Arbic, B.K., Mller, M., Richman, J.G., Shriver, J.F., Morten, A.J., Scott, R.B., Srazin, G., Penduff, T., 2014. Geostrophic turbulence in the frequencywavenumber domain: Eddy-driven low-frequency variability. Journal of Physical Oceanography 44, 2050–2069. doi:10.1175/JPO-D-13-054.1.

47

Arbic, B.K., Scott, R.B., Flierl, G.R., Morten, A.J., Richman, J.G., Shriver, J.F., 2012. Nonlinear cascades of surface oceanic geostrophic kinetic energy in the frequency domain. Journal of Physical Oceanography 42, 1577–1600. doi:10.1175/JPO-D-11-0151.1.

Barnier, B., Siefridt, L., Marchesiello, P., 1995. Thermal forcing for a global ocean circulation model using a three-year climatology of ecmwf analyses. Journal of Marine Systems 6, 363–380.

Chapman, C.C., 2014. Southern ocean jets and how to find them: Improving and comparing common jet detection methods. Journal of Geophysical Research: Oceans 119, 4318–4339. doi:10.1002/2014JC009810.

Chapman, C.C., 2017. New perspectives on frontal variability in the southern ocean. Journal of Physical Oceanography 47, 1151–1168. doi:10.1175/JPO-D-16-0222.1.

Chapman, C.C., Hogg, A.M., Kiss, A.E., Rintoul, S.R., 2015. The dynamics of southern ocean storm tracks. Journal of Physical Oceanography 45, 884–903.

Davis, R., 1982. On relating eulerian and lagrangian velocity statistics: Single particles in homogeneous flows. Journal of Fluid Mechanics 114, 1–26. doi:10.1017/S0022112082000019.

Davis, R.E., 1987. Modeling eddy transport of passive tracers. Journal of Marine Research 45, 635–666.

Davis, R.E., 1991a. Lagrangian ocean studies. Annual Review of Fluid Mechanics 23, 43–64. doi:10.1146/annurev.fl.23.010191.000355.

Davis, R.E., 1991b. Observing the general circulation with floats. Deep Sea Research Part A. Oceanographic Research Papers 38, S531 – S571. doi:10.1016/S0198-0149(12)80023-9.

Davis, R.E., 1998. Preliminary results from directly measuring middepth circulation in the tropical and south pacific. Journal of Geophysical Research: Oceans 103, 24619–24639. doi:10.1029/98JC01913.

Dufour, C.O., Griffies, S.M., de Souza, G.F., Frenger, I., Morrison, A.K., Palter, J.B., Sarmiento, J.L., Galbraith, E.D., Dunne, J.P., Anderson, W.G., et al., 2015. Role of mesoscale eddies in cross-frontal transport of heat and biogeochemical tracers in the southern ocean. Journal of Physical Oceanography 45, 3057–3081.

Elipot, S., Lumpkin, R., Perez, R.C., Lilly, J.M., Early, J.J., Sykulski, A.M., 2016. A global surface drifter data set at hourly resolution. Journal of Geophysical Research: Oceans 121, 2937–2966. doi:10.1002/2016JC011716.

Fox-Kemper, B., Ferrari, R., Pedlosky, J., 2003. On the indeterminacy of rotational and divergent eddy fluxes. Journal of Physical Oceanography 33, 478–483. doi:10.1175/1520-0485(2003)033<0478:OTIORA>2.0.CO;2.

Fratantoni, D.M., Richardson, P.L., 1999. Sofar float observations of an intermediate-depth eastern boundary current and mesoscale variability in the eastern tropical atlantic ocean. Journal of Physical Oceanography 29, 1265–1278. doi:10.1175/1520-0485(1999)029<1265:SFOOAI>2.0.CO;2.

Gille, S.T., 2003a. Float observations of the southern ocean. part i: Estimating mean fields, bottom velocities, and topographic steering. Journal of Physical Oceanography 33, 1167–1181.

Gille, S.T., 2003b. Float observations of the southern ocean. part ii: Eddy fluxes. Journal of Physical Oceanography 33, 1182–1196.

Griesel, A., Gille, S.T., Sprintall, J., McClean, J.L., Maltrud, M.E., 2009. Assessing eddy heat flux and its parameterization: A wavenumber perspective from a 1/10 ocean simulation. Ocean Modelling 29, 248 – 260. doi:http://dx.doi.org/10.1016/j.ocemod.2009.05.004.

Hoffman, R.N., Atlas, R., 2016. Future observing system simulation experiments. Bulletin of the American Meteorological Society 97, 1601–1616. doi:10.1175/BAMS-D-15-00200.1.

49

Jayne, S.R., Marotzke, J., 2002. The oceanic eddy heat transport. Journal of Physical Oceanography 32, 3328–3345. doi:10.1175/1520-0485(2002)032<3328:TOEHT>2.0.CO;2.

Kamenkovich, I., Cheng, W., Sarachik, E.S., Harrison, D.E., 2009. Simulation of the argo observing system in an ocean general circulation model. Journal of Geophysical Research: Oceans 114. doi:10.1029/2008JC005184. c09021.

Kamenkovich, I., Cheng, W., Schmid, C., Harrison, D.E., 2011. Effects of eddies on an ocean observing system with profiling floats: Idealized simulations of the argo array. Journal of Geophysical Research: Oceans 116, n/a–n/a. URL: http://dx.doi.org/10.1029/2010JC006910, doi:10.1029/2010JC006910. c06003.

Keating, S.R., Smith, K.S., Kramer, P.R., 2011. Diagnosing lateral mixing in the upper ocean with virtual tracers: Spatial and temporal resolution dependence. Journal of Physical Oceanography 41, 1512–1534. doi:10.1175/2011JPO4580.1.

LaCasce, J., 2008. Statistics from lagrangian observations. Progress in Oceanography 77, 1–29. doi:10.1016/j.pocean.2008.02.002.

Lebedev, K.V., Yoshinari, H., Maximenko, N.A., Hacker, P.W., 2007. YoMaHa'07: Velocity data assessed from trajectories of argo floats at parking level and at the sea surface. IPRC Technical Note 4.

Maas, L.R.M., 1989. A comparison of eulerian and lagrangian current measurements. Deutsche Hydrografische Zeitschrift 42, 111–132. doi:10.1007/BF02226290.

Madec, G., 2014. Nemo ocean engine. Note du Pôle de modélisation, Institut Pierre-Simon Laplace (IPSL) 27.

Majkut, J.D., Carter, B.R., Frölicher, T.L., Dufour, C.O., Rodgers, K.B., 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 Verdiére, 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/JPO3038.1.

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

argo data of the northeast atlantic. Geophysical Research Letters 38. doi:`10.1029/2010GL046267`. l02606.

Smith, S., 1997. The Scientist and Engineer's Guide to Digital Signal Processing. California Technical Pub. URL: `https://books.google.fr/books?id=rp2VQgAACAAJ`.

Taylor, G.I., 1938. The spectrum of turbulence. Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences 164, 476–490. doi:`10.1098/rspa.1938.0032`.

Thompson, A.F., Sallée, J.B., 2012. Jets and topography: Jet transitions and the impact on transport in the antarctic circumpolar current. Journal of physical Oceanography 42, 956–972.

Williams, R.G., Wilson, C., Hughes, C.W., 2007. Ocean and atmosphere storm tracks: The role of eddy vorticity forcing. Journal of Physical Oceanography 37, 2267–2289.

Willis, J.K., Fu, L.L., 2008. Combining altimeter and subsurface float data to estimate the time-averaged circulation in the upper ocean. Journal of Geophysical Research: Oceans 113. doi:`10.1029/2007JC004690`. c12017.

Wunch, C., 2006. Discrete Inverse and State Estimation Problems. Cambridge University Press, New York. URL: `www.cambridge.org/9780521854245`.

53