Exploitation of error correlation in a large analysis validation: GlobCurrent case study

Richard E. Danielson[a,*], Johnny A. Johannessen[a], Graham D. Quartly[b], Marie-Hélène Rio[c], Bertrand Chapron[d], Fabrice Collard[e], Craig Donlon[f]

[a] *Nansen Environmental and Remote Sensing Center, Bergen, Norway*
[b] *Plymouth Marine Laboratory, Plymouth, United Kingdom*
[c] *Collecte Localisation Satellites, Ramonville Saint-Agne, France*
[d] *Ifremer, Plouzané, France*
[e] *OceanDataLab, Locmaria-Plouzané, France*
[f] *European Space Agency, Noordwijk, Netherlands*

Abstract

An assessment of variance in ocean current signal and noise shared by in situ observations (drifters) and a large gridded analysis (GlobCurrent) is sought as a function of day of the year for 1993-2015 and across a broad spectrum of current speed. Regardless of the division of collocations, it is difficult to claim that any synoptic assessment can be based on independent observations. Instead, a measurement model that departs from ordinary linear regression by accommodating error correlation is proposed. The interpretation of independence is explored by applying Fuller's (1987) concept of equation and measurement error to a division of error into shared (correlated) and unshared (uncorrelated) components, respectively. The resulting division of variance in the new model favours noise. Ocean current shared (equation) error is of comparable magnitude to unshared (measurement) er-

*Current affiliation: Danielson Associates Office Inc., Halifax, Canada
Email address: rickedanielson@gmail.com (Richard E. Danielson)

ror and the latter is, for GlobCurrent and drifters respectively, comparable to ordinary and reverse linear regression. Although signal variance appears to be small, its utility as a measure of agreement between two variates is highlighted.

Sparse collocations that sample a dense (high resolution) grid permit a first order autoregressive form of measurement model to be considered, including parameterizations of analysis-in situ error cross-correlation and analysis temporal error autocorrelation. The former (cross-correlation) is an equation error term that accommodates error shared by both GlobCurrent and drifters. The latter (autocorrelation) facilitates an identification and retrieval of all model parameters. Solutions are sought using a prescribed calibration between GlobCurrent and drifters (by variance matching). Because the true current variance of GlobCurrent and drifters is small, signal to noise ratio is near zero at best. This is particularly evident for moderate current speed and for the meridional current component.

Keywords: measurement model, ocean current, collocation, validation

1. Introduction

The idea that errors in two collocated estimates of ocean current could be independent of each other is, like geostrophy itself, both practical and instructive. The difficult implication is that only signal (or truth) is correlated while noise (or error) is not. Considering that all measurement models are approximate (Box, 1979), such a clean separation may be ideal in principle but is probably quite rare in practice. The purpose of this study is to assess the GlobCurrent analysis, but the need to accommodate cross-correlated errors between GlobCurrent and drifters is not matched by an existing framework for doing so. Thus, a new measurement model is called for.

Although there is no evidence that ocean current signal is dictated by drifters alone, drifters are employed to refine the mean dynamic topography (MDT; Rio and Hernandez 2004; Rio et al. 2014). Thus, measurement errors may be correlated because the MDT effectively determines GlobCurrent in a time-mean sense. Measurement error is not the only type of error, however. Perhaps the simplest measurement models (including all models of this study) assume that truth and error in a dataset are *additive* and the signal in two datasets can be *linearly* related. There is growing evidence that for datasets that do not conform exactly to such assumptions, an associated *equation error* term needs to be considered (Fuller, 1987; Carroll and Ruppert, 1996; Kipnis et al., 1999). It is precisely because equation error may be strongly correlated that datasets should not necessarily be considered independent, even if there is no apparent physical relationship between them.

This study represents an experiment in ocean surface current validation that draws on advances in measurement modelling, notably in hydrology and epidemiology, but contemporary surface current validation also informs this work. Johnson et al. (2007) attribute differences between the OSCAR five-day current analysis and in situ observations in part to dynamic processes that are difficult to resolve (e.g., tropical instability waves and high latitude eddies). Additionally, although larger signal and noise are resolved by OSCAR relative to an assimilative model, Johnson et al. highlight the existence of intrinsic challenges in capturing the meridional current near the equator and variability in both components near the poles.

Surface current validation by Blockley et al. (2012) and Sudre et al. (2013) similarly acknowledge in situ error. Blockley et al. highlight differences in the western equatorial Pacific between surface currents that they derive from in situ observations and the FOAM assimilative model. Global correlation between model and observations is again much better for the zonal current component (versus meridional), especially in the tropics and north Pacific (reduced correlation in the Atlantic is attributed to slightly greater coverage by eddies). Although the GECKO satellite-based analysis of Sudre et al. finds corresponding systematic variations (by latitude and current component), their combination of geostrophic and Ekman estimates is also significantly correlated with in situ estimates. It is the agreement between, and independence of, two such estimates that we wish to reconsider below.

It is convenient to speak of correlation either in terms of signal and noise, or equivalently, truth and error. It is also useful to distinguish between the (spatial or temporal) autocorrelation of a single variable and the cross-correlation of two variables. Geophysical modelling approaches (including this study) often assume that autocorrelation should be easy to find in high resolution (analysis) data, and for some (in situ) collocation subset, that an affine signal model with additive, orthogonal (or signal-uncorrelated) noise applies. More formally, if two collocated ocean current datasets (I and A) are divided parsimoniously into shared truth (t) and additive error (ϵ) such that

$$\begin{aligned} \text{in situ} \quad I &= t + \epsilon_I \\ \text{analysis} \quad A &= \alpha + \beta t + \epsilon_A, \end{aligned} \tag{1}$$

then the affine signal model is a linear calibration involving an unbiased intercept (α) and slope (β) that relates signal in the two datasets by $A_{signal} = \alpha + \beta I_{signal}$ (where $I_{signal} = t$). The measurement model (1) is known as a regression model with errors in the variables (I and A) but (with reference to a linear relationship between I_{signal} and A_{signal}) no error in the equation (Fuller, 2006). Note also that cross-correlation is only expected from truth, or perhaps error, that is shared between datasets and that (1) omits a partition of error into shared and unshared, or cross-correlated and uncorrelated, components.

If there is no obvious physical dependence between datasets, then there is no guarantee that shared error, or shared truth for that matter, exist. Because the geophysical interpretation of cross-correlated error continues to evolve, this concept of sharing is at least partly unfamiliar, even in the context of two datasets (1). An established explanation in the context of three datasets (Stoffelen, 1998; O'Carroll et al., 2008) focuses on the cross-correlated part of representativeness error: it is natural for correlation to exist between two higher resolution datasets on scales that a lower resolution dataset cannot resolve, but if there is a truth that is shared by all three datasets, then by definition, this truth is also low resolution and any high resolution correlation must be considered erroneous, albeit perfectly natural. Errors of representation in geophysics (e.g., mismatches that can be written as a component of ϵ_I or ϵ_A, as in Gruber et al. 2016b) refer to information that is beyond some true, or target, spatiotemporal resolution limit. However, if shared truth does exist, it follows that the most generic and inclusive definition of limitations in this truth is needed to define what remains in each individual dataset as error.

Stoffelen's introduction of the triple collocation model provides an important description, and one of the earliest quantifications, of representativeness error (see also Vogelzang et al. 2011). Nevertheless, the triple collocation model is just identified, so the parameters sought (see Appendix) are equal in number to the first and second moment equations that are available (cf. Gillard and Iles 2005). A familiar characteristic of this model (like simpler regression models) is its limited flexibility to identify more parameters. Hence, correlated representativeness error, and cross-correlated error in general, must either be known in advance or perhaps be justifiably small for a retrieval of the triple collocation parameters.

Caires and Sterl (2003) discovered a way to explore cross-correlated error (between altimeters) in comparative applications of the triple collocation model. They examined significant wave height and 10-m wind speed estimates from buoys and two altimeters, which were carefully averaged to be comparable in space and time with collocated ERA-40 estimates. Because representativeness errors were reduced by this averaging, it was postulated that any remaining ERA-40 cross-correlated errors could be neglected if ERA-40 did not assimilate an observational dataset. A bound on cross-correlated error was then estimated for the altimeters, whose uncorrelated error was found to be relatively low when retrieved together with ERA-40 rather than separately with ERA-40 and buoys. Consideration of this bound yielded an increase in altimeter error variance by a factor of two or more, but Caires and Sterl suggested that cross-correlated error may have been smaller.

Janssen et al. (2007) examined wave height data from two altimeters, buoys, and an ECMWF wave hindcast, and employed an iterative form of orthogonal regression (Gillard and Iles, 2005) with estimates of uncorrelated error from the triple collocation model. An important acknowledgement was given of the linear calibration in (1) being a potential source of cross-correlated error (i.e., where a nonlinear signal model might be appropriate instead). As in Caires and Sterl (2003), it was postulated that cross-correlated errors could be neglected if data (or systematic errors) were not assimilated, but uncorrelated altimetric error was again found to be relatively low when the triple collocation model was applied to both altimeters at once. Janssen et al. proposed additional model equations (using ECMWF first guess and analysis wave products) to quantify rather than just bound most errors, but found that altimetric error, including its cross-correlated component, was small.

Methods of collocating buoy, radiometer, and microwave SST estimates (e.g., O'Carroll et al. 2008) also point to cross-correlated error being small, but only insofar as representativeness error is tested, as above, by parameter comparisons. A novel assessment of cross-correlated error has also been given using a high resolution, rescaled in situ dataset as a proxy for truth. Yilmaz and Crow (2014) use this proxy to directly characterize terms of the triple collocation model based on soil moisture from an assimilative model and soil moisture retrievals from passive (AMSR-E) and active (ASCAT) satellites. The dependence of satellite retrievals is notable because significant cross-correlated errors are found. This study concludes that zero error cross-correlation is a tenuous assumption of the triple collocation model as its corresponding bias in parameter retrievals is systematic.

Contemporary calibration and validation studies have introduced a growing list of geophysical dataset differences, which taken together, define corresponding limitations on shared truth. However, perhaps the most generic characterization of these limitations is found in the measurement modelling literature: Fuller (1987) defines measurement error in the familiar sense of random data departures from a linear regression solution and distinguishes *equation error* as random departures from the linear signal model of (1), owing to nonlinearity in the signal model of interest. Carroll and Ruppert (1996) expose the importance of this refinement in a geophysical application and, as noted above, Janssen et al. (2007) highlight that such nonlinearity is a potential source of cross-correlated error.

The combination of measurement error and equation error is useful to better accommodate limitations in the scope of a shared truth. With reference to person-specific bias in epidemiology, Kipnis et al. (1999, 2002) introduce equation error as two additional terms (ϵ_{QI} and ϵ_{QA}) in (1) that lead to

$$\begin{aligned} \text{in situ} \quad I &= t + \epsilon_{QI} + \epsilon_I \\ \text{analysis} \quad A &= \alpha + \beta t + \epsilon_{QA} + \epsilon_A, \end{aligned} \tag{2}$$

where ϵ_I and ϵ_A are now random departures from a possibly nonlinear signal model. Carroll and Ruppert (1996) note that applications of (2) have been limited, possibly because if ϵ_{QI} and ϵ_{QA} are considered to be independent of other errors, they can be recombined with ϵ_I and ϵ_A to yield the simpler equation (1) with its original properties intact (Moberg and Brattström, 2011). Below, the same linear signal model as in (1) will be considered, with shared equation error defined by $\epsilon_{QI} = \epsilon_{QA}$ and total error involving both shared and unshared components. In other words, equation error is not independent so it is important to quantify this as a separate term in our application of (2).

In addition to the interpretation of cross-correlated errors, there remains the issue of identifying solutions to increasingly sophisticated statistical models. Increasing the number of collocated datasets (e.g., Janssen et al. 2007; Zwieback et al. 2012; Gruber et al. 2016a) is one approach. However, an important development in the geophysical literature is the recognition by Su et al. (2014) that three or more datasets may be unnecessary, as collocation models appear to belong to a broader family of instrumental variable regression models, and within this family, a precedent exists for using lagged variables as instruments. Following Su et al., this implies that by embracing autocorrelation, strategies should continue to emerge that depend on fewer datasets to identify a larger number of collocations and statistical model parameters. By comparison with the error-in-variables model (1), the novelty of the strategy proposed below is that it also permits the retrieval of variance in shared error and, in one ocean current experiment, also equation error.

The present study seeks to advance measurement modelling and parameter identification with the benefit of error correlation. The focus is on ocean surface current validation, but general supporting concepts and terms (such as *measurement model*) are provided in the Appendix. The next section describes the collocation of GlobCurrent and drifters and proposes a commonly prescribed linear relationship between them that addresses the difference in variance between these two datasets. Formulation of a measurement model that permits error correlation to be exploited is given in Section 3. We then describe the strong and weak constraints that allow a retrieval of all model parameters and assess the performance of GlobCurrent and drifter data in

Section 4. Throughout this paper, equal emphasis is placed on true variance and on the contributions to total error. Discussion of inferences based on the division of variance into shared truth and error are highlighted in Section 5 and Section 6 contains the conclusions.

2. Selection of a calibration

We begin with the idea that GlobCurrent and drifters provide estimates of fundamentally different ocean currents, but they also provide overlapping views of a true (or target) ocean current that can be represented at 15 m below the surface on a 6-h, $1/4°$ grid. By any definition of shared truth, both GlobCurrent and drifters have errors. GlobCurrent is an analysis that linearly combines the geostrophic and Ekman components. Drifters respond locally to a combination of geostrophic, Ekman, tidal, inertial, Stokes, and wind drift processes, including (erroneous) processes on scales smaller and faster than the GlobCurrent grid can resolve. In general, such differences can be considered a mismatch in their supports (see Appendix). Nearest-neighbour collocations of drifters (whose drogues move roughly with the 15-m current) and GlobCurrent (also at 15 m, with additional samples at daily intervals) are considered below.

Six-hourly drifter velocity has been estimated following Hansen and Poulain (1996). We restrict attention to drifters whose continuous drogue presence was confirmed by objective or subjective means (Rio, 2012; Lumpkin et al., 2013). The resulting geographic distribution for 1993-2015 (Fig. 1) yields more than eleven million drifter and GlobCurrent zonal and meridional velocity estimates (Danielson 2017; a comparable number of drifters lost their

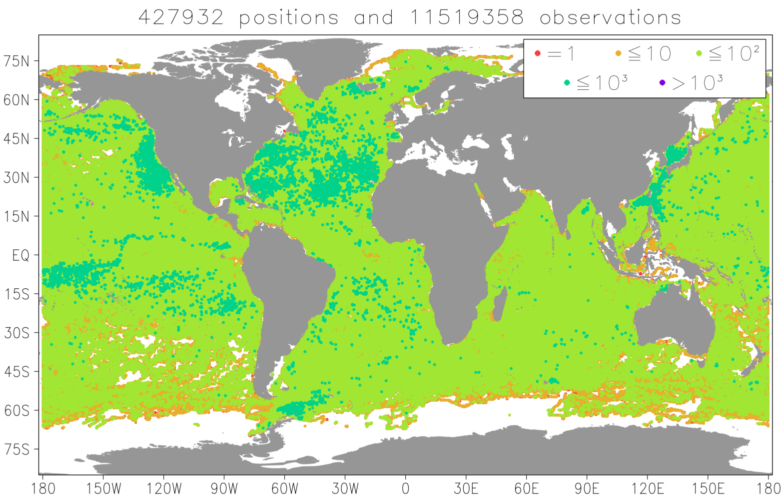

Figure 1: Number of surface drifter velocity observations between January 1993 and December 2015 (order of magnitude in colour) with drogues attached. Shown are values at the $1/4°$ resolution of the GlobCurrent grid (i.e., collocations are nearest neighbours).

drogues and, being more responsive to surface wind forcing, are ignored). It is convenient to divide collocations by even and odd year, with the latter subset permitting an independent check on calculations. Below, only the even-year subset is discussed but the same conclusions can be obtained from the results (available as supplementary material) of the odd-year subset.

Joint frequency of occurrence of current speed, including the full range of possible linear calibrations of GlobCurrent relative to drifters, is shown in Fig. 2. These two-dimensional histograms are rather well behaved following removal of about 10% of the most extreme current speeds (Hubert et al., 2012). Similar regression slopes are revealed in both the zonal and meridional distributions. Between the bounding ordinary and reverse linear regression reference slopes (dashed lines) is a slope defined by the ratio of total variance

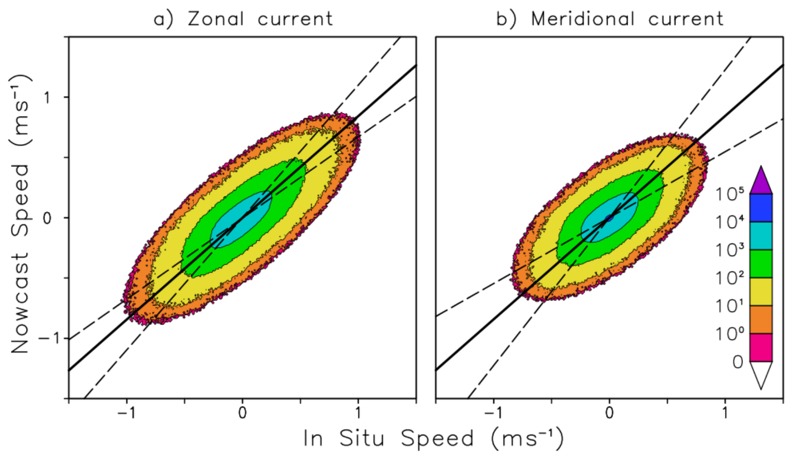

Figure 2: Two-dimensional histograms of a) zonal and b) meridional 15-m current component for 5310226 non-outlier collocations from the even years between 1993 and 2015 (approximately half the collocations of Fig. 1, after removing about 10% of these data as outliers following Hubert et al. 2012). The dashed lines are the ordinary (shallow slope) and reverse (steep slope) linear regression references for each current component. The slope of the solid line is defined by the GlobCurrent–drifter variance ratio (the same ratio for both current components; see next section). The logarithmic colourbar is number of values in 0.01-ms^{-1} bins.

between GlobCurrent and drifters (solid line; defined in the next section). Unfortunately, scatter away from these regression lines is a poor indication that there might be a component of error variance that is shared between GlobCurrent and drifters, or that total error variance might be greater than the variance in shared truth.

The corresponding one-dimensional (marginal) distributions (Fig. 3) highlight an unsurprising difference between current estimates: because drifters capture a greater range of physical processes at higher resolution, we find fewer low values and more high values than GlobCurrent (with an equal num-

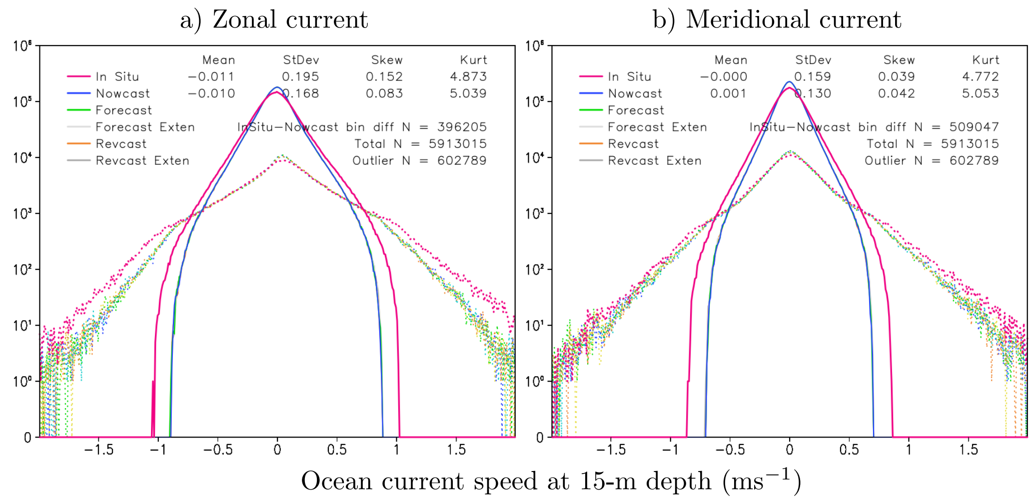

Figure 3: One-dimensional histograms of a) zonal and b) meridional 15-m current component, as in Fig. 2, but including outliers separately (dotted lines). Also shown are drifter (red) and GlobCurrent nowcast (blue), forecast (green and light grey), and revcast (orange and dark grey) histograms. Forecast and revcast data are taken one day (with extended data from two days) before and after each collocation, respectively. Statistical moments of the non-outlier in situ and nowcast distributions are given with a measure of difference between the two (i.e., one half of the in situ minus nowcast bin count difference). The logarithmic ordinate is number of values in 0.01-ms^{-1} bins.

ber at about ± 0.15 ms^{-1}). Also as expected, GlobCurrent samples at two days (extended forecast) and one day (forecast) before each drifter (in situ) observation, as well as one day (revcast) and two days (extended revcast) after, have the same distribution as the GlobCurrent collocations (nowcast). Outliers are shown separately by dotted lines in Fig. 3 and are identified by minimizing the covariance matrix determinant for the six estimates of zonal and meridional current (Hubert et al., 2012). Because covariance (and skewness and kurtosis) are sensitive to outliers (McColl et al., 2014; Su et al., 2014), collocation groups are trimmed by about 10% before other calculations are performed. Often this excludes extreme values in the zonal or meridional component and values close to zero in the opposite component.

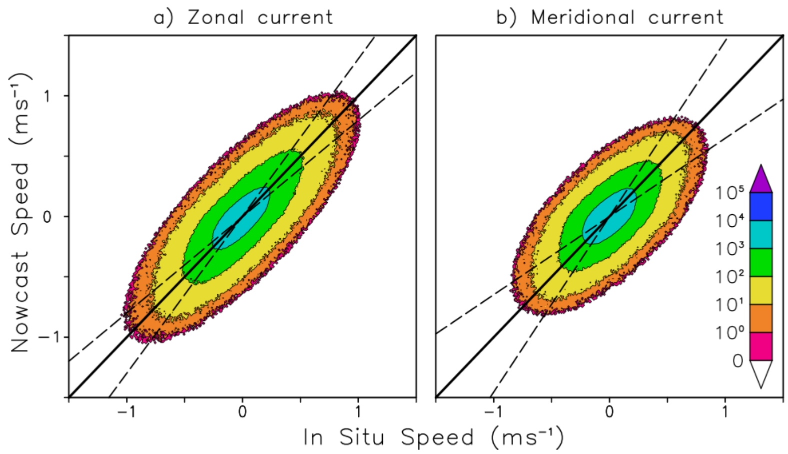

Figure 4: As in Fig. 2, but after dividing all GlobCurrent data by 0.84 (i.e., the ratio of nowcast to drifter standard deviation), where zonal and meridional components are expressed as complex numbers and the same variance match is applied to both components.

The distinction between cross-correlated and uncorrelated error is sufficiently novel that initial solutions of (2) benefit from the assumption of a fixed calibration that can be applied uniformly. (Subsequent work will seek a general, varying solution, but this simplification applies to all experiments below.) An assumption that would be consistent with the mismatch in GlobCurrent and drifter *support* (rather than a bias between them) is that both are already unbiased. However, we note in Section 4 that if calibration is bounded by ordinary and reverse linear regression (dashed lines in Fig. 2), then this assumption would not apply to all collocation subsets. An alternate assumption that can be applied uniformly, and whose bias is familiar in the context of (1), is known as variance matching (Fuller, 2006;

Yilmaz and Crow, 2013; Su et al., 2014). This calibration is marked by a lack of assumptions about relative error in GlobCurrent and drifters. It fixes regression slope midway between the bounding ordinary and reverse linear regression solutions (solid line in Fig. 2) and fixes GlobCurrent and drifter signal-to-noise ratio (SNR) to be equal. A definition and further implications are given in Section 3.

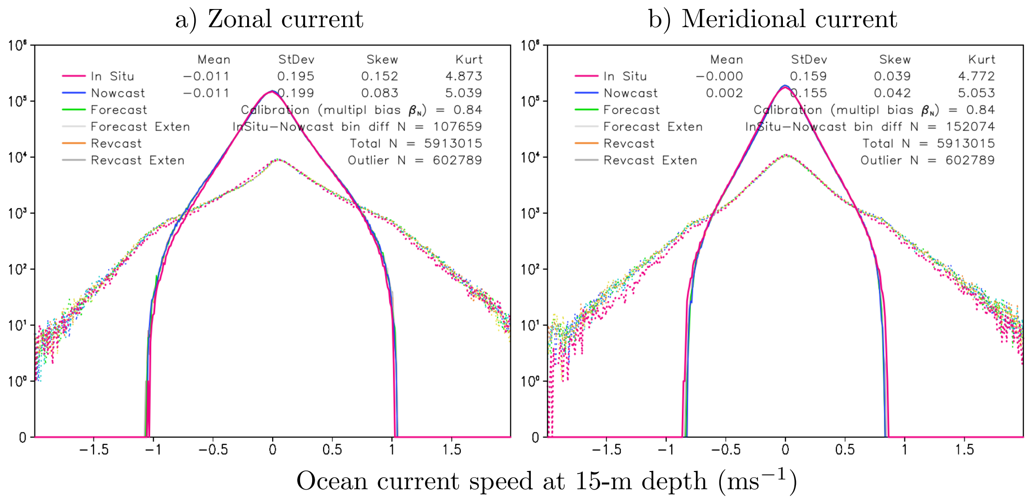

Figure 5: As in Fig. 3, but after dividing all GlobCurrent data by 0.84.

Figures 4 and 5 are the result of matching the variance of GlobCurrent to that of drifters. (Simultaneous matching of the zonal and meridional components is accomplished by expressing these two components as a complex number.) Dividing the GlobCurrent data by a standard deviation ratio of 0.84 reduces the number of weak values and increases the number of strong values, as expected. This calibration removes much of the cumulative difference in bin counts: from 7-8% in Fig. 3 to about 2% in Fig. 5. However, the distinction between calibrated GlobCurrent and drifters remains, as histogram shape is otherwise preserved (note that skewness and kurtosis are variance-normalized moments) and current direction is unchanged. Moreover, and notwithstanding important applications to assimilation and model validation (e.g., Stoffelen 1998; Tolman 1998), this distinction would remain at least under any affine calibration.

3. Measurement model development

A series of experimental models, based initially on the triple collocation approach (Stoffelen, 1998; McColl et al., 2014) with solutions sought by the method of moments (Gillard and Iles, 2005), have informed the measurement model that we will focus on. The first experimental model in this series (3) can be criticised for using extrapolated (forecast and revcast) GlobCurrent estimates assuming that extrapolated errors are independent. Gridded altimetric data are often based on a centered span of up to 12 days of Topex/Jason passes and a longer period for Envisat. Similarly for the Ekman (or Stokes) current estimates from a model-based analysis, if a model has the wind front in the wrong location or an incorrect initial storm intensity, it may retain a consistent bias for days. Thus, the assumption of independent errors ϵ in a slightly modified triple collocation model,

$$\begin{aligned} \text{in situ} \quad I &= t + \epsilon_I \\ \text{forecast} \quad F &= \alpha_F + \beta_F t + \epsilon_F \\ \text{revcast} \quad R &= \alpha_R + \beta_R t + \epsilon_R, \end{aligned} \tag{3}$$

can be considered experimental at best. Note that α, β, t, and ϵ are additive calibration, multiplicative calibration (or regression slope), truth, and error, respectively, and our use of drifters as a calibration reference implies that $\alpha_I = 0$ and $\beta_I = 1$. Here, F and R are obtained by extrapolation of GlobCurrent from outside a centered window of only a few days.

The form of (3) is recognizable in an intermediate (but still unsatisfactory) model (4) that includes both GlobCurrent and drifter collocations (I and N) and retains additive and multiplicative calibration parameters (α and β) for each GlobCurrent estimate. A notable simplification of (4) is that extrapolation is replaced by a persistence forecast/revcast, so F and R are just GlobCurrent samples taken one day before and after each collocation, respectively.

$$\begin{aligned}
\text{in situ} \quad I &= \qquad\qquad t + \epsilon_I \\
\text{nowcast} \quad N &= \alpha_N + \beta_N t + \epsilon_N \\
\text{forecast} \quad F &= \alpha_F + \beta_F t + \epsilon_N + \epsilon_F \\
\text{revcast} \quad R &= \alpha_R + \beta_R t + \epsilon_N + \epsilon_R.
\end{aligned} \tag{4}$$

The model (4) is overly constrained in its treatment of correlated error, however. There is no shared (equation) error between GlobCurrent and drifters and a complete sharing of N errors in F and R. In turn, it is perhaps unsurprising that there may be effectively no difference (in terms of physical insight) between parameter retrievals based on (4) and ordinary and reverse linear regression references based on I and N alone (Danielson et al., 2017).

Two further innovations are required to arrive at the measurement model of interest. One is that a first-order autoregressive (AR-1) parameterization is probably the simplest way to accommodate both GlobCurrent-drifter error cross-correlation as well as GlobCurrent error autocorrelation. Error propagation is parameterized in the same sense as it might occur in an ocean current analysis, with observational error having its biggest impact on an analysis at the time of observation, with a decreasing, but symmetric impact at times before and after. The AR-1 form accommodates autocorrelated errors (e.g., from altimetry) that also have a symmetric upstream and downstream impact (note that asymmetric error propagation may be appropriate in some applications).

The second innovation, following Su et al. (2014), is that additional, or extended, samples of GlobCurrent are beneficial, assuming these remain inside the autocorrelation envelope. The resulting model becomes

$$\begin{aligned}
\text{in situ} \quad I &= t + \epsilon_I \\
\text{nowcast} \quad N &= \alpha_N + \beta_N t + \lambda_N \epsilon_I + \epsilon_N \\
\text{forecast} \quad F &= \alpha_F + \beta_F t + \lambda_F(\lambda_N \epsilon_I + \epsilon_N) + \epsilon_F \\
\text{extended forecast} \quad E &= \alpha_E + \beta_E t + \lambda_E(\lambda_F(\lambda_N \epsilon_I + \epsilon_N) + \epsilon_F) + \epsilon_E \\
\text{revcast} \quad R &= \alpha_R + \beta_R t + \lambda_R(\lambda_N \epsilon_I + \epsilon_N) + \epsilon_R \\
\text{extended revcast} \quad S &= \alpha_S + \beta_S t + \lambda_S(\lambda_R(\lambda_N \epsilon_I + \epsilon_N) + \epsilon_R) + \epsilon_S,
\end{aligned} \tag{5}$$

where Fuller's (1987) equation error, corresponding in (2) to $\epsilon_{QI} = \epsilon_{QA}$ (Kipnis et al., 1999), is the shared (cross-correlated) error parameterization $\lambda_N \epsilon_I$. We return to the interpretation of shared and unshared error in ϵ_I below. The remaining errors are uncorrelated measurement errors, also denoted individual errors: ϵ_N, ϵ_F, ϵ_E, ϵ_R, and ϵ_S.

A so-called INFR model, whose name is taken from the data samples on the LHS of (4) but whose RHS is taken from (5), has parameters that are almost identifiable (in a statistical sense). That is, one can derive 10 covariance equations (given below) but there are 11 unknown parameters. The INFERS model (5) includes an extended forecast and revcast, which are GlobCurrent samples two days before and after each collocation. Under the assumption that GlobCurrent errors remain correlated at least over five days (e.g., as gauged by the product $\lambda_F\lambda_E\lambda_R\lambda_S$), INFERS is more attractive because there are more covariance equations (21) than unknown parameters (17). (Of course, with more samples further improvement in the ratio of these numbers is possible.) Standard assumptions of no correlation between truth and error (orthogonality) and among individual errors then allow all elements of the covariance matrix to be defined by

$$
\begin{aligned}
Var(I) &= \sigma_t^2 + \sigma_I^2 \\
Var(N) &= \beta_N^2\sigma_t^2 + \lambda_N^2\sigma_I^2 + \sigma_N^2 \\
Var(F) &= \beta_F^2\sigma_t^2 + \lambda_F^2\lambda_N^2\sigma_I^2 + \lambda_F^2\sigma_N^2 + \sigma_F^2 \\
Var(E) &= \beta_E^2\sigma_t^2 + \lambda_E^2\lambda_F^2\lambda_N^2\sigma_I^2 + \lambda_E^2\lambda_F^2\sigma_N^2 + \lambda_E^2\sigma_F^2 + \sigma_E^2 \\
Var(R) &= \beta_R^2\sigma_t^2 + \lambda_R^2\lambda_N^2\sigma_I^2 + \lambda_R^2\sigma_N^2 + \sigma_R^2 \\
Var(S) &= \beta_S^2\sigma_t^2 + \lambda_S^2\lambda_R^2\lambda_N^2\sigma_I^2 + \lambda_S^2\lambda_R^2\sigma_N^2 + \lambda_S^2\sigma_R^2 + \sigma_S^2 \\
Cov(I,N) &= \beta_N\sigma_t^2 + \lambda_N\sigma_I^2 \\
Cov(I,F) &= \beta_F\sigma_t^2 + \lambda_F\lambda_N\sigma_I^2 \\
Cov(I,E) &= \beta_E\sigma_t^2 + \lambda_E\lambda_F\lambda_N\sigma_I^2 \\
Cov(I,R) &= \beta_R\sigma_t^2 + \lambda_R\lambda_N\sigma_I^2 \\
Cov(I,S) &= \beta_S\sigma_t^2 + \lambda_S\lambda_R\lambda_N\sigma_I^2 \\
Cov(N,F) &= \beta_N\beta_F\sigma_t^2 + \lambda_F\lambda_N^2\sigma_I^2 + \lambda_F\sigma_N^2 \\
Cov(N,E) &= \beta_N\beta_E\sigma_t^2 + \lambda_E\lambda_F\lambda_N^2\sigma_I^2 + \lambda_E\lambda_F\sigma_N^2 \\
Cov(N,R) &= \beta_N\beta_R\sigma_t^2 + \lambda_R\lambda_N^2\sigma_I^2 + \lambda_R\sigma_N^2 \\
Cov(N,S) &= \beta_N\beta_S\sigma_t^2 + \lambda_S\lambda_R\lambda_N^2\sigma_I^2 + \lambda_S\lambda_R\sigma_N^2,
\end{aligned}
\tag{6}
$$

and

$$
\begin{aligned}
Cov(F,E) &= \beta_F\beta_E\sigma_t^2 + \lambda_E\lambda_F^2\lambda_N^2\sigma_I^2 + \lambda_E\lambda_F^2\sigma_N^2 + \lambda_E\sigma_F^2 \\
Cov(F,R) &= \beta_F\beta_R\sigma_t^2 + \lambda_F\lambda_R\lambda_N^2\sigma_I^2 + \lambda_F\lambda_R\sigma_N^2 \\
Cov(F,S) &= \beta_F\beta_S\sigma_t^2 + \lambda_F\lambda_S\lambda_R\lambda_N^2\sigma_I^2 + \lambda_F\lambda_S\lambda_R\sigma_N^2 \\
Cov(E,R) &= \beta_E\beta_R\sigma_t^2 + \lambda_E\lambda_F\lambda_R\lambda_N^2\sigma_I^2 + \lambda_E\lambda_F\lambda_R\sigma_N^2 \\
Cov(E,S) &= \beta_E\beta_S\sigma_t^2 + \lambda_E\lambda_F\lambda_S\lambda_R\lambda_N^2\sigma_I^2 + \lambda_E\lambda_F\lambda_S\lambda_R\sigma_N^2 \\
Cov(R,S) &= \beta_R\beta_S\sigma_t^2 + \lambda_S\lambda_R^2\lambda_N^2\sigma_I^2 + \lambda_S\lambda_R^2\sigma_N^2 + \lambda_S\sigma_R^2.
\end{aligned} \tag{7}
$$

The corresponding 17 unknowns are true variance (σ_t^2), multiplicative calibration for five datasets ($\beta_N, \beta_F, \beta_E, \beta_R, \beta_S$), and error variance for all six ($\sigma_I^2, \sigma_N^2, \sigma_F^2, \sigma_E^2, \sigma_R^2, \sigma_S^2$). There are also five parameters that gauge GlobCurrent-drifter error cross-correlation (λ_N is denoted shared error fraction below) and GlobCurrent error autocorrelation ($\lambda_F, \lambda_E, \lambda_R, \lambda_S$). An analytic solution of all parameters except σ_t^2 and β_N is possible using (6) as a strong constraint (i.e., using all variances and the covariances involving the GlobCurrent and drifter collocations I and N). The remaining equations (7) are denoted the autocovariance equations (i.e., covariances involving only GlobCurrent forecast and revcast samples $FERS$).

True variance (σ_t^2) and multiplicative calibration or regression slope (β_N) between GlobCurrent and drifters are key measurement model parameters. In the context of INFERS, these are both free parameters that can be sought numerically using the autocovariance equations as a weak constraint, that is, by approaching minima in the difference between the LHS and RHS of (7). Matching GlobCurrent variance to that of drifters (as in Section 2) provides all experiments with a fixed, but approximate, slope parameter β_N. In other words, our focus on a search for true variance is also limited by this assumption. It is important to note, moreover, that variance matching provides more freedom to retrieve large cross-correlated error because it is midway between the bounding ordinary and reverse linear regression solutions (i.e., where all variance in either GlobCurrent or drifters is assigned to truth and the possibility of cross-correlated error is excluded). It follows from this assumption that

$$\beta_N^2 = Var(N)/Var(I) \quad \Rightarrow \quad \sigma_N^2 = \sigma_I^2(\beta_N^2 - \lambda_N^2). \tag{8}$$

The remaining INFERS model parameters are retrieved once a solution for σ_t^2 is obtained. The weakly constrained minimization of (7) is sought between bounds for σ_t^2 that are given by $Var(I) = \sigma_t^2 + \sigma_I^2$ (i.e., between $\sigma_t^2 = 0$ and the ordinary linear regression solution of $\sigma_I^2 = 0$), with the additional strong constraint that all other variances $(\sigma_N^2, \sigma_F^2, \sigma_E^2, \sigma_R^2, \sigma_S^2)$ also remain non-negative. Just like the variance matched solution for β_N, each zonal and meridional current component is first expressed as a complex number so that 17 parameters are identified for both components at the same time (i.e., the covariances in (6) and (7) are real parts).

The remainder of this study is a diagnostic exploration of the parameters obtained from (5)-(8) given surface current variations that are jointly sampled by GlobCurrent and drifters. As required by INFERS, we also perform a simple check that GlobCurrent samples of truth and error (combined) remain inside their autocorrelation envelope: for any group of collocations, the minimum correlation between an NFERS pair (i.e., between E and S) is expected to be larger than about 0.7. All correlation estimates are obtained from the LHS of (6) and (7).

4. Performance assessment

We introduce a retrieval of measurement model parameters for all 5310226 non-outlier collocations from the even years between 1993 and 2015. This is followed by retrievals for subsets of this group as a function of day of the year and current speed. GlobCurrent and drifters appear to provide complementary information about ocean surface current. The SNR is near zero at best as variance in a shared true current tends to be smaller than the variance in total (shared and unshared) error. We also show that shared error fraction (λ_N) is quite high. A posteriori, this motivates our accommodation of cross-correlated error in (5). To the extent that cross-correlated error and equation error are the same (see Section 5), an important question is raised of whether a linear signal model and additive errors for GlobCurrent and drifters can be considered robust (and by what metric). Large individual (measurement) error is consistent with GlobCurrent and drifters as offering quite noisy estimates of shared true current variability (again subject to a linear calibration). In Section 5, we find that individual error is similar to the ordinary and reverse linear regression reference solutions. In other words, it is mainly a quantification of shared/correlated truth and error that require our attention.

Figure 6 depicts the absolute difference in LHS minus RHS of the forecast and revcast autocovariance equations (7). Differences are shown for all candidate values (i.e., true variance between zero and the variance of drifters), but model solutions are of interest only where variance is positive (unshaded region). The target minimum (open purple circle) is the average of three available local minima (i.e., no minima are associated with the ex-

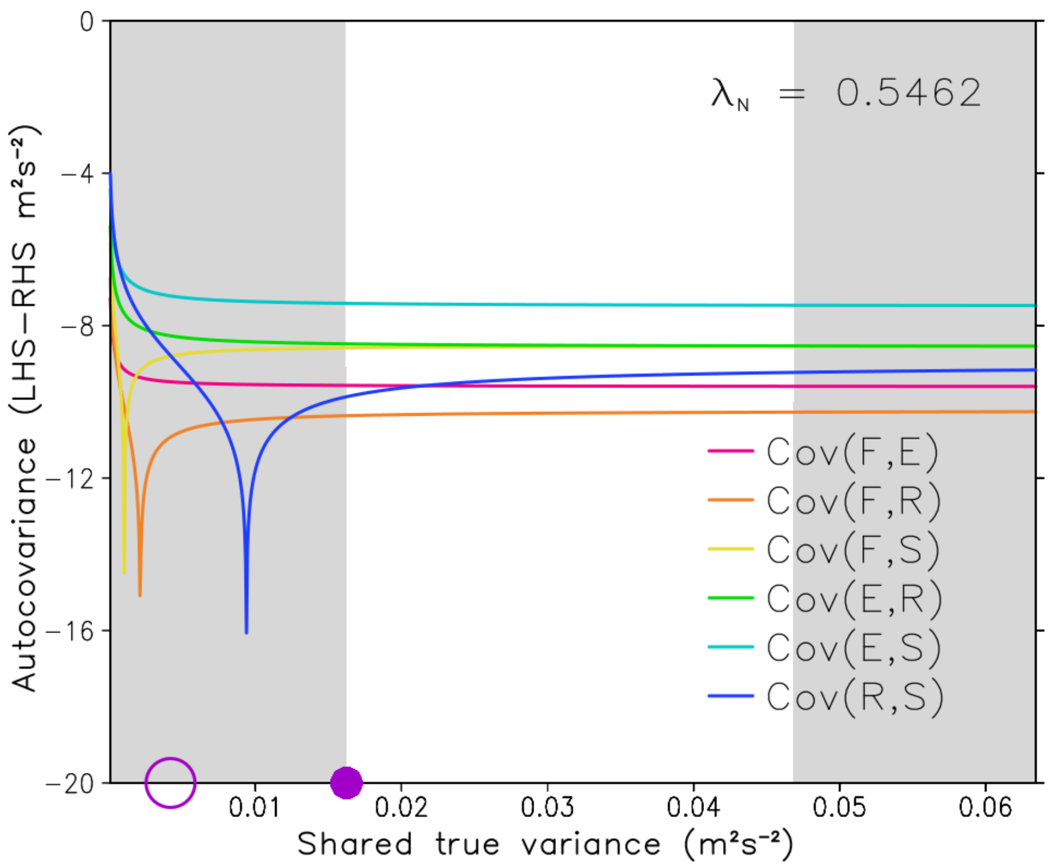

Figure 6: First demonstration of an INFERS parameter solution by weakly constrained minimization of the magnitude of differences between the LHS and RHS of the autocovariance equations (7) for the 5310226 non-outlier collocations from even years between 1993 and 2015 (roughly half of Fig. 3). The abscissa is true variance (σ_t^2) in m^2s^{-2} between zero and $Var(I)$. The ordinate is log of absolute difference (LHS minus RHS). Grey shading denotes regions of negative error variance retrieval. Included are the target minimum (open purple circle at the average of three local minima) and the chosen minimum on the unshaded region (closed purple circle). The GlobCurrent-drifter shared error fraction (λ_N) at the chosen minimum is also shown.

tended forecast E). Although this target minimum is not accessible (on the unshaded region), the chosen true variance solution is just to the right of this locus of three minima and about the same distance from them as they are from each other. This choice implies that at least one model variance estimate is zero. Here, shading on the left in Fig. 6 corresponds to negative shared true variance of the meridional current component (this is a derived quantity that varies with λ_N).

Whereas target solutions on the unshaded region can be seen as a reminder that models like (1), (2), and (5) are parsimonious (Box, 1979), the tendency of autocovariance minima to be found on the left side of Fig. 6 may be the most important aspect of accommodating error cross-correlation. This first demonstration indicates that true variance shared by GlobCurrent and drifters is as small as possible (given that retrieved variance should be positive). Visually, true and drifter error variance are the abscissa lengths to the left and right of the closed purple dot, respectively. True variance is thus smaller than drifter error variance when all collocations are considered.

Table 1 provides model parameters for the drifter (in situ) and GlobCurrent (nowcast) zonal (U) and meridional (V) current components. We find that truth and error are of similar magnitude and that GlobCurrent and drifters sample not only a shared truth but also shared error. However, this truth exists only in the zonal component (0.127 ms^{-1}). Negligible meridional amplitude (0.003 ms^{-1}) corresponds with a solution at the border of the shaded region in Fig. 6. The additive calibration of GlobCurrent (α_N) is also negligible and multiplicative calibration (β_N) is prescribed by variance matching (Fig. 5). Evidently, GlobCurrent samples are within their auto-

Table 1: Model parameters of the drifter (I) and GlobCurrent nowcast (N) zonal (U) and meridional (V) current components that are retrieved using 5310226 non-outlier collocations from the even years between 1993 and 2015 (cf. Fig. 3). Parameters include total standard deviation (σ), true standard deviation (σ_t), nowcast additive calibration (α_N), multiplicative calibration (β_N), shared error fraction (λ_N), individual ($[1-\lambda_N]^{1/2}\sigma_I$ and σ_N) and total (σ_I and $[\lambda_N^2\sigma_I^2+\sigma_N^2]^{1/2}$) error standard deviation as in (6), signal correlation (McColl et al., 2014), and signal to noise ratio (SNR; Gruber et al. 2016b). Standard deviation and additive calibration are given in ms^{-1} and SNR is in dB.

<table>
<tr><th></th><th>σ</th><th>σ_t</th><th>α_N</th><th>β_N</th><th>λ_N</th><th>σ_{indiv}</th><th>σ_{total}</th><th>Corr</th><th>SNR</th></tr>
<tr><td>U_I</td><td>0.195</td><td rowspan="4">U:
0.127
V:
0.003</td><td colspan="3" rowspan="2"></td><td>0.100</td><td>0.148</td><td>0.652</td><td>-1.3</td></tr>
<tr><td>V_I</td><td>0.159</td><td>0.107</td><td>0.159</td><td>0.021</td><td>-33.6</td></tr>
<tr><td>U_N</td><td>0.168</td><td>-0.001</td><td rowspan="2">0.843</td><td rowspan="2">0.546</td><td>0.100</td><td>0.129</td><td>0.640</td><td>-1.6</td></tr>
<tr><td>V_N</td><td>0.130</td><td>0.001</td><td>0.097</td><td>0.130</td><td>0.022</td><td>-33.3</td></tr>
</table>

covariance envelope as the minimum correlation for this sample is 0.91 and 0.83 for the zonal and meridional current components, respectively.

We obtain most of the individual error terms in (5) and (6) from the model retrievals of unshared (measurement error) variance (i.e., $\sigma_N^2, \sigma_F^2, \sigma_E^2, \sigma_R^2$, and σ_S^2). The exception is individual error for drifters ($[1-\lambda_N]\epsilon_I$), which follows from our definition of shared equation error (Kipnis et al., 1999). Diagnostic equations for shared and unshared drifter error variance can be written as $\lambda_N\sigma_I^2$ and $(1-\lambda_N)\sigma_I^2$, respectively (i.e., assuming an even split of the covariance between equation error $\lambda_N\epsilon_I$ and individual error $[1-\lambda_N]\epsilon_I$). Because over 50% of drifter error is shared by GlobCurrent (λ_N), the percentage of total variance in (6) that is shared equation error ranges from 23% (GlobCurrent zonal component) to 55% (drifter meridional component).

Individual and total error variance for the zonal and meridional components are both high (Table 1). Calibration by variance matching dictates that drifter and GlobCurrent correlation with truth (McColl et al., 2014) and SNR (Gruber et al., 2016b) are roughly the same by zonal or meridional component (Su et al., 2014). Meridional noise dominates signal (SNR is -33dB) and even zonal noise is larger than signal (SNR < 0). Note that SNR is calculated using total error (i.e., both correlated and uncorrelated; third column from the right in Table 1). A preliminary regional assessment (not shown; GlobCurrent project document 2017) is consistent with previous studies (Johnson et al., 2007; Blockley et al., 2012; Sudre et al., 2013) in highlighting that weak meridional SNR is a characteristic of the equatorial regions.

Figure 7 is a second demonstration that true variance shared by GlobCurrent and drifters is small. Parameters are retrieved as a function of day of the year, and to isolate one high latitude seasonal cycle, collocations north of 15°N latitude are selected. We focus on 2385232 collocations of this northern region from even years between 1993 and 2015 (i.e., 21% of those available, using about 6000 collocations per day and applying variance matching and outlier removal at daily intervals). Figure 7 depicts solutions of true variance for an arbitrary selection of 12 days, of which eight are consistent with Fig. 6 insofar as the locus of autocovariance minima (7) are at exceedingly small true variance. Only on day 240 (Fig. 7h) is true variance relatively large (as dictated by covariance involving F). An examination of all 364 days reveals a similar result: true variance is as small as possible on 250 of 339 days

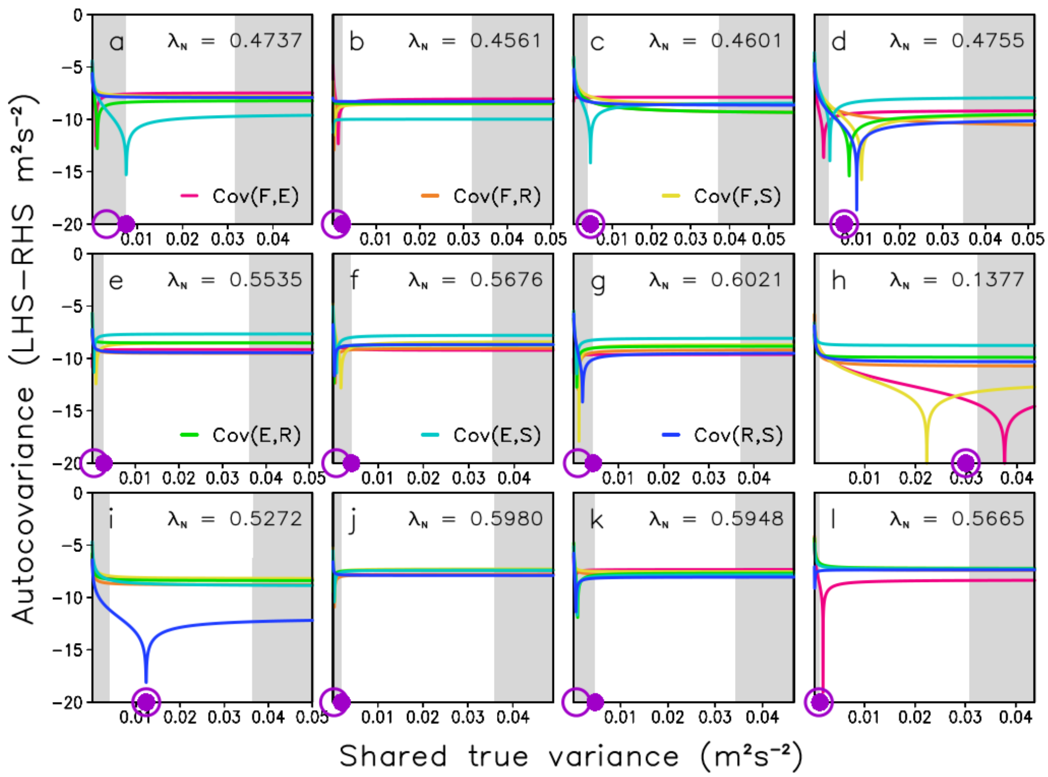

Figure 7: As in Fig. 6, but only for collocations north of 15°N on day a) 30, b) 60, c) 90, d) 120, e) 150, f) 180, g) 210, h) 240, i) 270, j) 300, k) 330, and l) 360 of the year for even years between 1993 and 2015.

(74%). No parameters are estimated on 25 of 364 days (7%) because no autocovariance minima are found.

Figure 8 depicts the Northern Hemisphere seasonal cycle by five-day running means for the full set of INFERS model parameters. There is an annual variation in the calibration and shared error parameters (c,d) that can be explained by (e,j) GlobCurrent and drifter variations being slightly more similar in amplitude toward the end of the year than at the beginning (e.g., solid lines tend to bracket the annual-average dashed lines in March and to

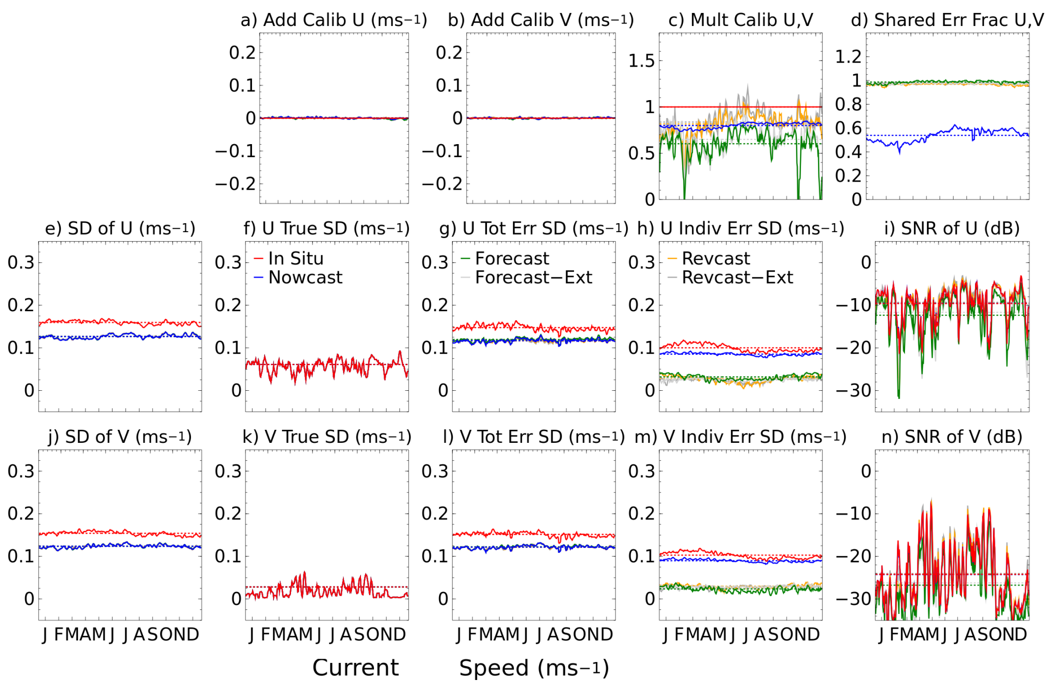

Figure 8: Retrieved model parameters as in Table 1, but for 339 days of the year using about 6000 collocations per day from north of 15°N and from even years between 1993 and 2015. Shown are the drifter (in situ/red) and GlobCurrent (nowcast/blue, forecast/green, revcast/orange, and extended forecast/light grey and revcast/dark grey) retrievals of a) zonal and b) meridional additive calibration (ms^{-1}) and c) multiplicative calibration and d) shared error fraction for both zonal and meridional components, and e,j) 15-m current, f,k) shared truth, g,l) total error, and h,m) individual error standard deviation (ms^{-1}), and i,n) signal to noise ratio (dB) for the zonal and meridional components, respectively. Solid lines are averages over five days and dashed lines are annual averages.

be bracketed by them in September). Of course, this similarity is largely superficial, based on a consistent retrieval throughout the year of small shared truth in the zonal component (f; Fig. 7), and as in Table 1, almost no signal in the meridional component (k).

Drifter noise in Fig. 8 appears to be greater during spring than fall whereas GlobCurrent signal (via seasonality in multiplicative calibration) is the opposite. As a result, signal to noise ratio is higher for both GlobCurrent and drifters in late summer compared to spring, even for the meridional current (despite its weak signal). A spatiotemporal refinement of this result (with specific attention to the role of mixed layer depth) seems to be required. This same seasonality in SNR is obtained for the forecast and revcast samples, although via a different allocation of variance (i.e., with total error being almost entirely defined by the GlobCurrent nowcast error). The range in multiplicative calibration (c) for the forecast and revcast data is an a posteriori justification for retaining separate calibrations in (5). All NFERS GlobCurrent samples again appear to be within their autocovariance envelope, as the minimum correlation among all days of the year is 0.88 and 0.84 for the zonal and meridional current components, respectively.

Figure 9 is the third demonstration that true variance shared by GlobCurrent and drifters is small. For a diagnosis of model parameters as a function of drifter current speed, we again apply variance matching and outlier removal (Hubert et al., 2012) as above, but to small groups of collocations. Tolman (1998) demonstrates that fine bin resolution (with sample sizes of O[100]) is useful to avoid bias in covariance estimates. Moreover, Zwieback et al. (2012) recommend at least 500 collocations based on idealized triple

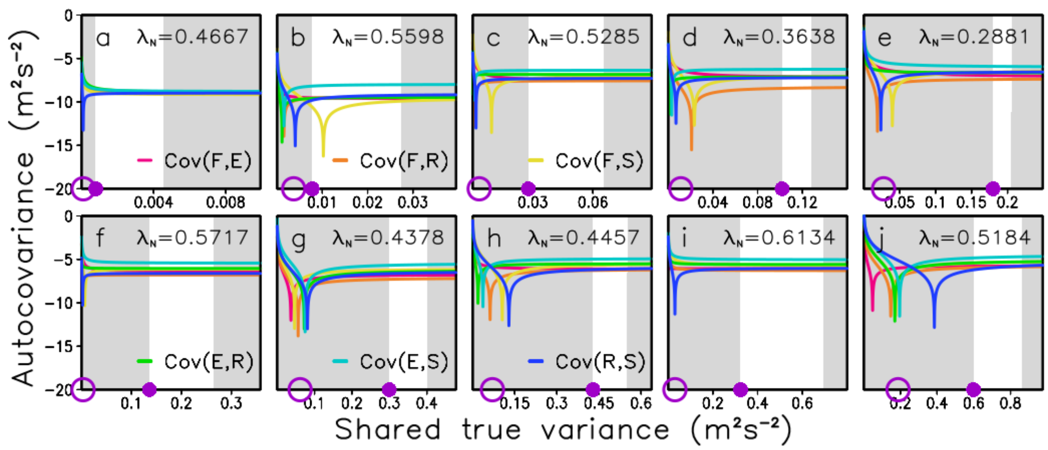

Figure 9: As in Fig. 6, but for subsets of 500 collocations whose drifter speed is nearest to a) 0.1 ms−1, b) 0.2 ms−1, c) 0.3 ms−1, d) 0.4 ms−1, e) 0.5 ms−1, f) 0.6 ms−1, g) 0.7 ms−1, h) 0.8 ms−1, i) 0.9 ms−1, and j) 1.0 ms^{-1}. Note that abscissa range varies with current speed.

collocation simulations. Solutions of true variance are thus obtained over a finely resolved (0.01-ms^{-1}) range in drifter speed using 500 collocations closest to each of 101 target speeds. (This sampling requires less than 1% of the available collocations.) Individual panels in Fig. 9 are again consistent with Fig. 6 in that all 10 loci of autocovariance minima (7) are at exceedingly small true variance. An examination of the 101 speed bins reveals that true variance is as small as possible for 90 of 92 bins (98%) and no parameters are estimated for 9 of 101 bins (9%) because no autocovariance minima are found.

Figure 10 illustrates the dependence of model parameters on current speed. There are weak trends in the calibration and shared error parameters (a-d) and strong trends in most variance parameters (e-n). As in Table 1,

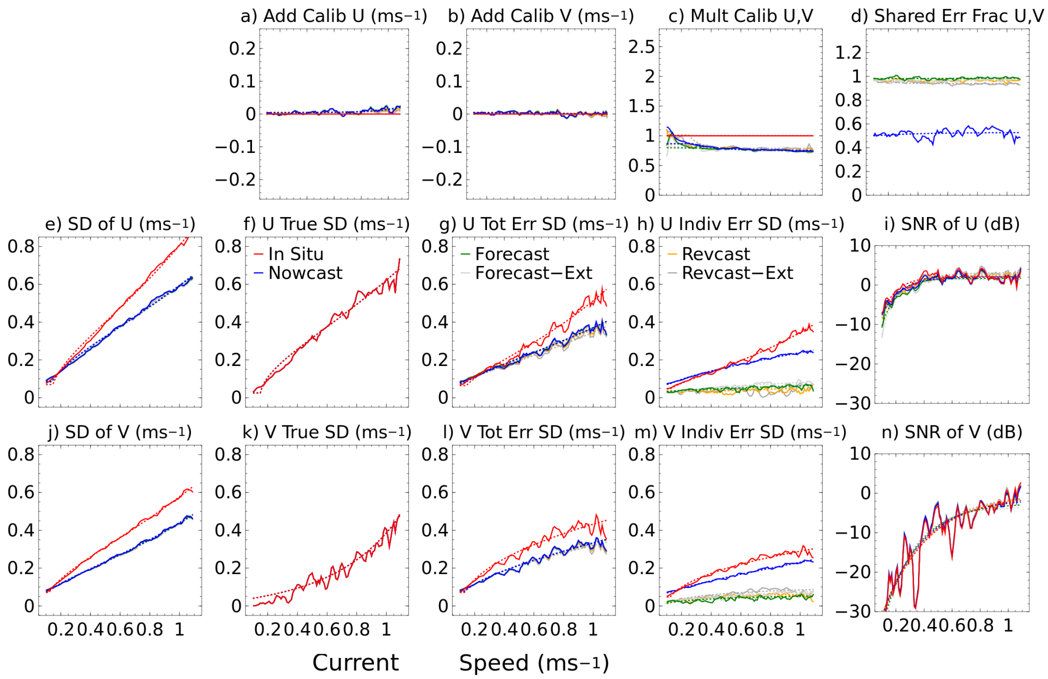

Figure 10: Model parameters as in Fig. 8, but for 92 subsets of 500 collocations whose drifter speed is nearest to target values between 0.1 ms^{-1} and 1.1 ms^{-1} at intervals of 0.01 ms^{-1} (excluding 9 solutions for which no autocovariance minima were found). Solid lines are averages of five adjacent intervals. Dashed lines are best fits of the form $y(x) = a+be^{cx}$ (Jacquelin, 2014), but for c) multiplicative calibration, this fit ignores target values less than 0.3 ms^{-1}.

GlobCurrent-drifter shared error fraction ($\lambda_N \approx 0.5$) is quite high, variance-matched multiplicative calibration (β_N) is about 0.85 beyond 0.3 ms^{-1}, and additive calibration of GlobCurrent (α_N) is negligible. Justification for our application of variance matching thoughout this study (rather than assuming no GlobCurrent bias) is that an upper bound on multiplicative bias, as given by reverse linear regression, falls below one at large current speed (not shown). In turn, the need to address strong current underestimation (perhaps locally in time and space, but at the resolution of the GlobCurrent analysis) may continue to exist (cf. Rio et al. 2014).

Errors in GlobCurrent samples separated by a day are basically the same in Fig. 10g,h,l,m. The product of the forecast and revcast shared error fraction parameters ($\lambda_F \lambda_E \lambda_R \lambda_S$) is thus close to unity, which implies that GlobCurrent error is being sampled within its autocovariance envelope. In effect, this justifies the use of the extended forecast and revcast samples in the INFERS model. Among all 92 subsets, the minimum correlation of combined truth and error (found at low speed between E and S) is 0.84 and 0.76 for the zonal and meridional current components, respectively.

Figure 10f,k reveals weak agreement between GlobCurrent and drifters on a shared truth at low current speed, but more agreement at higher current speed. This is dictated in part by current speed itself (Fig. 10e,j), but the meridional component of drifter error increases quickly with current speed (more so than the zonal component) and the opposite is the case for true variance. In contrast to negative SNR for the zonal component in Table 1, the GlobCurrent/drifter best fit SNR (Fig. 10i dashed lines; equivalent by variance matching) eventually exceed, but remain close to, 0 dB from about 0.3 ms^{-1}.

This section constitutes an introduction to the INFERS model featuring hundreds of parameter solutions. Our experiments are thus enabled by access to millions of drifter current estimates and a GlobCurrent analysis that is about three orders of magnitude larger. This is not to say that 500 collocations is small. In many contexts, including ours, a few hundred collocations may be ample. However, with the freedom afforded by large datasets to identify a range of solutions using appropriate instruments (cf. Kipnis et al. 2002), comes the opportunity to better characterize shared truth and error. The next section briefly explores shared truth as an updated measure of agreement between variates and clarifies shared error as an updated measure of dependence.

5. Discussion

It is sometimes the case in geophysics that only one truth (a so-called genuine truth) is of interest. Implicit in this concept is the idea that truth carries no information about particular datasets, which differ only in terms of their corresponding error, and this error is intrinsic (i.e., defined without reference to another dataset). Implicit in the definition of shared truth, on the other hand, is the idea that *if shared truth exists, then it contains information about an overlap in data supports* (see Appendix). Beyond the scope of this paper, but notable within geophysics, are formal inference theories that concern a conjunction of information and the problem of aggregated opinion (Tarantola, 2005). Here, it suffices to note that measurement models can provide a calibration by linear mapping, and a validation by shared/unshared error, but they can also provide a useful measure of agreement among datasets by shared truth.

One documented application of shared truth is an assessment by Bentamy et al. (2017) of various global ocean surface heat flux analyses. Using the INFERS model, Bentamy et al. experiment with shared truth as a metric of competitive validation (see Appendix). Following a recalibration of each gridded analysis to the same in situ reference, they observe that in situ and analysis total error becomes equal, whereas shared truth is invariant (their Table 2 thus provides a standardized ranking). This invariance of shared truth is a property of many measurement models and may not be well known, perhaps in part because shared truth itself is often undocumented. To be fair, all documented searches so far (including Bentamy et al.) assume a fixed calibration rather than seeking true variance and calibration together (cf. Section 3).

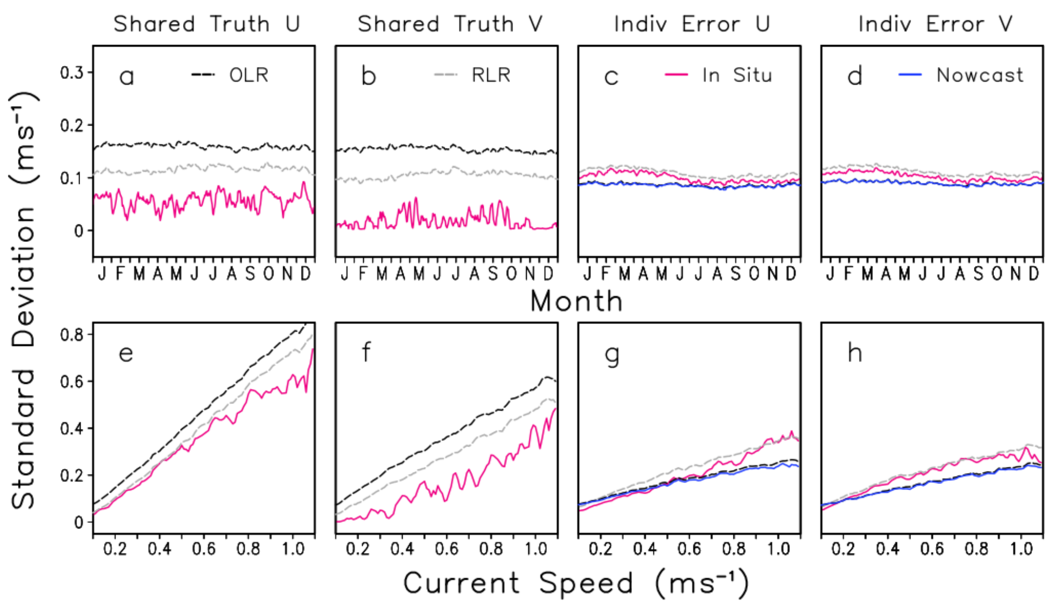

Figure 11: Shared truth (a,b,e,f) and individual error (c,d,g,h) as in Figs. 8 and 10 (f,k,h,m), but only for the drifter (in situ; red) and GlobCurrent (nowcast; blue) collocations. Included are the corresponding ordinary (OLR; dashed black) and reverse (RLR; dashed grey) linear regression reference solutions.

We propose that shared truth should have equal focus to error in typical validation efforts. Because INFERS introduces error correlation into the errors-in-variables regression model, a good comparison for INFERS is the full range of solutions consistent with (1), with familiar analytic solutions for ordinary (OLR) and reverse (RLR) linear regression being appropriate references. Solutions of the OLR and RLR models are identified by the method of moments with either drifter error (ϵ_I for OLR) or GlobCurrent error (ϵ_A for RLR) set to zero. INFERS estimates of truth and error from the previous section are placed alongside these two reference solutions in Fig. 11. It is notable that INFERS solutions of true standard deviation (Fig. 11a,b,e,f) are smallest. This is remarkable because the OLR and RLR references are understood to be the solutions that bound the range of true variance (and multiplicative calibration or regression slope) values that are consistent with the errors-in-variables model (1).

Further comparison between INFERS and the corresponding OLR and RLR reference solutions permit an interpretation of the unshared (measurement) errors that define much of the total error in this study. Figure 11c,d,g,h reveals that the magnitude of OLR error in GlobCurrent and RLR error in drifters appear to differ little from the unshared error shown in Fig. 8h,m and Fig. 10h,m. As noted in Section 4, some ambiguity is expected in a diagnostic estimate of drifter unshared error, but the overlapping agreement in GlobCurrent unshared error (i.e., black dashed and blue lines) is evident for all collocation divisions. Whereas OLR and RLR impose separate assumptions on (1) that provide hypothetical bounds on uncorrelated error, in this study a single model seems to provide both solutions.

Figure 11 reveals that total error in GlobCurrent and drifters can be interpreted as a combination of respective RLR and OLR upper limits in uncorrelated error. Subject to the caveat that a fixed calibration by variance matching allows more freedom for shared error in our INFERS solutions, the reason that shared true variance falls outside the OLR and RLR bounding reference solutions is because not only can the INFERS model accommodate bounds on unshared (measurement) error, as given by (1), but shared (equation) error is accommodated as well.

We conclude this initial characterization of model solutions by noting that shared error offers an updated measure of error dependence. It is important to recognize that any decision to exclude shared error from a measurement model, based on physical knowledge of the data alone, can always be challenged. In other words, *even if there is no apparent physical relationship between two datasets, independence of their errors should not be presumed without considering that the measurement model is only an approximation* (Box, 1979). Thus, it may be appropriate to accommodate sharing even if one cannot assume that shared error (or truth) exists. More specifically (Fuller, 1987), if the model assumes that truth and error are additive with a linearly related signal, as in (1), and this might not be strictly true of the data, then some form of both equation and measurement error (2) or correlated and uncorrelated error (5) should be included.

Equation error and correlated error are considered to be essentially the same in this study, as we now demonstrate, but they are not strictly the same error in general. For instance, Kipnis et al. (1999) allow for correlation in both equation error and measurement error. The Introduction acknowledges that GlobCurrent and drifters also may share a component of measurement error. This is because many of the same drifters that are employed to refine the CNES-CLS13 mean dynamic topography (MDT; Rio and Hernandez

2004; Rio et al. 2014) are employed above for validation. Although INFERS provides an estimate of error correlation that may include measurement error, one option for demonstrating its interpretation as equation error is a validation only after 2013. Instead, we opt to replace the CNES-CLS13 MDT in each GlobCurrent sample (NFERS) with a more approximate GOCE-only MDT (Rio et al., 2014). Drifter measurement error is thus removed from GlobCurrent and the remaining error correlation can be attributed entirely to equation error.

Table 2: As in Table 1, but for a measurement-error-independent comparison between GlobCurrent and drifters: GlobCurrent data exclude a velocity component associated with the CNES/CLS-2013 MDT and include instead a component associated with the GOCE-only geodetic MDT (Rio et al., 2014). Parameters of the drifter (I) and GlobCurrent nowcast (N) zonal (U) and meridional (V) current components are retrieved using 5280828 non-outlier collocations from the even years between 1993 and 2015.

<table>
<tr><th></th><th>σ</th><th>σ_t</th><th>α_N</th><th>β_N</th><th>λ_N</th><th>σ_{indiv}</th><th>σ_{total}</th><th>Corr</th><th>SNR</th></tr>
<tr><td>U_I</td><td>0.194</td><td rowspan="4">U:
0.119
V:
0.001</td><td colspan="3" rowspan="2"></td><td>0.106</td><td>0.153</td><td>0.612</td><td>-2.2</td></tr>
<tr><td>V_I</td><td>0.158</td><td>0.110</td><td>0.158</td><td>0.007</td><td>-43.7</td></tr>
<tr><td>U_N</td><td>0.161</td><td>-0.001</td><td rowspan="2">0.818</td><td rowspan="2">0.517</td><td>0.101</td><td>0.128</td><td>0.604</td><td>-2.4</td></tr>
<tr><td>V_N</td><td>0.127</td><td>0.002</td><td>0.097</td><td>0.127</td><td>0.007</td><td>-43.5</td></tr>
</table>

Table 2 provides a comparison between GlobCurrent (GOCE-only MDT) and drifters based on 5280828 non-outlier collocations from the even years between 1993 and 2015. With some of the strongest current components (i.e., most different in terms of MDT) again excluded as outliers, true standard deviation in the zonal component decreases slightly (0.127 ms^{-1} to 0.119 ms^{-1})

for an MDT that lacks drifter information. Otherwise, the results of Table 1 are reproduced, including small true variance in the meridional component and a large shared error fraction ($\lambda_N = 0.517$). Although this is a measurement-error-independent comparison, it is nevertheless clear that the two datasets are not independent. Shared error fraction in Table 1 is quite similar ($\lambda_N = 0.546$), as is the percentage of total variance in (6) that is shared error, again ranging from 24% for the GlobCurrent zonal component to 52% for the drifter meridional component. The implication is that there is little error correlation owing to drifter measurement error in the CNES-CLS13 MDT. There is instead large error correlation owing to equation error.

6. Conclusions

This study provides an approach to the challenge of introducing and, like any other model term, identifying cross-correlated error in linear regression models such as (1). Subject to the caveat that calibration is prescribed by variance matching (rather than being jointly retrieved with shared true variance), over 90% of all attempts to retrieve model parameters for GlobCurrent and drifters are successful. Perhaps the more surprising aspect is that, given two datasets, we require just a few additional samples of the GlobCurrent analysis around the time of each drifter observation. Compared to the frequency of these additional samples, necessary confirmation of slow changes in the evolution of GlobCurrent and its errors is also obtained.

Formulation of a new measurement model called INFERS (an acronym taken from data sample names) is inspired by instrumental variable regression (Su et al., 2014) and specifically the triple collocation approach (Stoffelen, 1998; Caires and Sterl, 2003; Janssen et al., 2007; O'Carroll et al., 2008; Vogelzang et al., 2011; Zwieback et al., 2012; McColl et al., 2014; Yilmaz and Crow, 2014; Gruber et al., 2016b). Error propagation through the data samples is modelled using a first-order autoregressive (AR-1) formula, except that propagation begins with the collocated sample equations (IN), which provide the cross-correlated error terms, and then includes a temporally symmetric application of AR-1 to error autocorrelation in the remaining equations (FERS). The most direct model comparison is to solutions of the linear errors-in-variables regression model (1) because this is the same model given by the collocated sample equations (IN) if cross-correlated errors are ignored. A search for true variance in a limited parameter space of the INFERS model (i.e., assuming the variance matching calibration) yields values smaller than for any solution of (1), as given by ordinary (OLR) and reverse (RLR) linear regression bounds. Over three quarters of these model solutions (Fig. 11) support the proposition that truth and signal, as defined in the INFERS model, are small (see also Table 2 of Bentamy et al. 2017).

If truth is considered a shared model variable just like error (ignoring its unshared component), then shared true variance can be considered a measure of agreement between GlobCurrent and drifters. Inferences about measurement model approximations as well as overlaps in data support are then possible. While it would be unfortunate to start with a true variance that is smaller than it actually is (i.e., the variance matching calibration may yield such a bias), to start with a truth that is larger than it actually is would likely be more worrisome. This study indicates that there is a potential to overstate the agreement between GlobCurrent and drifters based on an inflated true variance in the linear errors-in-variables model. Like the triple collocation model, OLR and RLR are just identified and necessarily lack a term for cross-correlated error. Because their solutions involve variance budgets with fixed total variance, as in the LHS of (6), if total error is increased by introducing a new error term (equation or correlated error), then true variance decreases by the same amount. Tables 1 and 2 reveal that roughly a quarter to a half of the total variance in GlobCurrent and drifters is shared error variance. Presumably, shared error is a first order term that could not be much larger and remain hidden. Subsequent studies are needed to confirm whether this masquerading of equation error as truth is common for other datasets and whether it should be attributed to limitations in the errors-in-variables model. However, this should *exclude prescribed calibration* and instead explore solutions in the full parameter space of the INFERS model.

Implications of measurement model assumptions (e.g., that truth and error are additive with a linearly related signal) are discussed in geophysics (e.g., Janssen et al. 2007; Zwieback et al. 2016), and moreso in the statistical literature, where notions are established regarding how to accommodate nonlinear signals in linear regression by including equation error (Fuller, 1987; Carroll and Ruppert, 1996). Furthermore, accommodation of equation error and measurement error *correlation* is given in sophisticated measurement models in epidemiology (Kipnis et al., 1999, 2002). In turn, it appears that the opportunity to simultaneously identify all parameters of such models can be taken up in part by studies like this one that incorporate an experimental sampling of large datasets.

A sufficient number of GlobCurrent samples is taken before and after each collocation (as persistence forecasts and revcasts, respectively) so that there are more covariance equations than model parameters. Retrieval of the 17 INFERS model parameters employs variance matching to first prescribe the calibration from GlobCurrent to drifters. Six autocovariance equations, involving the FERS samples, weakly constrain shared true variance and the remaining 15 covariance equations are a strong constraint on the remaining 15 unknown parameters. Insofar as true variance is weakly constrained, this study avoids a common assumption that real data be cast in the form of a simple measurement model.

Model solutions have been examined for collocation groups numbering about six million (from eleven years), 6000 (on each day of the year in the NH), and 500 (nearest drifter speeds at 0.01-ms^{-1} intervals). One must be cautious about groups of collocations both large (if in situ error is autocorrelated) and small (if parameter retrievals depend on individual collocations; cf. Zwieback et al. 2012). However, for all these subsets, SNR is near zero at best because the error in GlobCurrent and drifters is high, while variance of the true current is low. There are indications that the preferentially low SNR of the meridional component is a characteristic of equatorial regions (cf. Johnson et al. 2007; Blockley et al. 2012; Sudre et al. 2013). The interpretation of large individual error is also interesting in that the OLR and RLR reference bounds on uncorrelated error are reached by both GlobCurrent and drifters.

The last experiment of the Discussion is perhaps the most relevant for an interpretation of shared and unshared error in terms of equation and measurement error, respectively. A measurement-error-independent comparison between GlobCurrent (using a GOCE-only MDT) and drifters permits a diagnosis of just how large the correlation in equation error may be. There is little change in shared error fraction between the two MDT experiments, which suggests that correlated error in other comparisons of this study may be viewed as predominantly that of equation error rather than measurement error (in spite of a drifter error contribution to the CNES/CLS13 MDT). Good correspondence between equation error and correlated error provides further impetus for a review of common model assumptions.

The so-called genuine truth is not viewed in this study as the same true variable t that appears in most measurement models. The search for a genuine ocean surface current is ongoing, however, and iterative or comparative applications of a measurement model have a role to play (e.g., Bentamy et al. 2017). By analogy with efforts to validate SST, surface current depth should be useful to distinguish between a slower, quasi-balanced flow and interactions with the atmosphere. For example, both drifters and GlobCurrent may be good references for balanced flow experiments at the equator (cf. Chan and Shepherd 2014) and at higher latitudes (cf. Penven et al. 2014). High resolution analyses are expected to grow in number, and while validation is not a prescription for finding the genuine current, there is an opportunity to quantify improvements in two or more datasets (or versions of a single dataset) against one chosen reference dataset. This study documents variations in INFERS model parameters as a function of day of the year and current speed, but a high latitude flow experiment may benefit from distinctions between cyclonic and anticyclonic eddies, whereas an equatorial experiment may opt to treat the zonal and meridional components separately. With a view to mapping model parameters in the dimensions of large datasets, an important challenge involves selecting subsets of collocations according to an informed physical understanding.

This study is a contribution to efforts of the geophysical community to construct high resolution ocean surface current analyses using assimilative numerical models and a synergy of observations (this issue). Because observational coverage is sparse, especially over the ocean and in early years, a topical question remains whether to withhold reference observations from an analysis so as to later perform an independent validation. To respond to this question in the negative would imply that the same observations should be allowed to benefit both the construction of an analysis and its validation. In turn, shared signal and noise in observations and analyses need to be considered and measurement models that accommodate both equation error and measurement error are called for (cf. Caires and Sterl 2003; Gruber et al. 2016b). It appears that not only can a basis for understanding shared signal and noise be found in literature, but a year-on-year accumulation of geophysical observations and high resolution data is permitting more freedom, and slightly less parsimony, in experimental measurement modelling.

7. Acknowledgements

We are pleased to acknowledge an international effort over many years to collect, assemble, and analyze altimetric and drifter observations, as well as the support and discussions (commencing roughly in reverse time) with Bash Toulany, Will Perrie, Graham Dunn, Tim Williams, Igor Esau, Laurent Bertino, Ad Stoffelen, Abderrahim Bentamy, Svetla Hristova-Veleva, Bryan Stiles, Zorana Jelenak, Mike Brennan, Luc Fillion, Bridget Thomas, Hal Ritchie, and Mike Dowd. The opportunity to reflect on the comments of three reviewers has been invaluable in promoting a clearer presentation. Funding for this work (again in reverse time) was from the European Space Agency via the Nansen Center and Ifremer (Data User Element's GlobCurrent and Support to Science Element's Ocean Heat Flux projects, respectively). The first author was also supported by the U.S. National Aeronautics and Space Administration via the National Oceanic and Atmospheric Administration and University Corporation for Atmospheric Research (Hurricane Science Research, Ocean Vector Winds, and Visiting Scientist programs, respectively), and the Canadian Space Agency via Environment and Climate Change Canada (Government Related Initiatives program).

8. References

Bentamy, A., Piollé, J.-F., Grouazel, A., Danielson, R. E., Gulev, S. K., Paul, F., Azelmat, H., Mathieu, P.-P., von Schuckmann, K., Sathyendranath, S., Evers-King, H., Esau, I., Johannessen, J. A., Clayson, C. A., Pinker, R. T., Grodsky, S. A., Bourassa, M., Smith, S. R., Haines, K., Valdivieso, M., Merchant, C. J., Chapron, B., Anderson, A., Hollmann, R., Josey, S. A., 2017. Review and assessment of latent and sensible heat flux accuracy over global oceans. Remote Sens. Environ. 201, 196–218, doi:10.1016/j.rse.2017.08.016.

Blockley, E. W., Martin, M. J., Hyder, P., 2012. Validation of FOAM near-surface ocean current forecasts using lagrangian drifting buoys. Ocean Sci. 8, 551–565, doi:10.5194/os–8–551–2012.

Box, G. E. P., 1979. Robustness in the strategy of scientific model building, *Robustness in Statistics*, Academic Press, pp. 201–236 (accessed 4 August 2017 at http://www.dtic.mil/get-tr-doc/pdf?AD=ADA070213).

Caires, S., Sterl, A., 2003. Validation of ocean wind and wave data using triple collocation. J. Geophys. Res. 108, doi:10.1029/2002JC001491.

Carroll, R. J., Ruppert, D., 1996. The use and misuse of orthogonal regression in linear errors-in-variables models. The American Statistician 50, 1–6.

Chan, I. H., Shepherd, T. G., 2014. Diabatic balance model for the equatorial atmosphere. J. Atmos. Sci. 71, 985–1001, doi:10.1175/JAS–D–13–0224.1.

Danielson, R. E., 2017. Collocations of surface drifter and GlobCurrent (centered five-day) 15-m zonal and meridional velocity estimates. SEANOE online data archive at http://doi.org/10.17882/50225.

Danielson, R. E., Johannessen, J. A., Rio, M.-H., Collard, F., Donlon, C., Chapron, B., Quartly, G., 2017. An update on GlobCurrent calibration and validation activities, poster presentation at the third GlobCurrent User Consultation Meeting (UCM-3) in Frascati, Italy on 21-23 March 2017 (accessed May 2017 at https://www.researchgate.net/publication/315486597_Update_on_GlobCurrent_Calibration_and_Validation_Activities).

Fuller, W. A., 1987. Measurement Error Models. Wiley, New York.

Fuller, W. A., 2006. Errors in variables. *Encyclopedia of Statistical Sciences*, S. Kotz, C. B. Read, N. Balakrishnan, B. Vidakovic and N. L. Johnson, Eds., doi:10.1002/0471667196.ess1036.pub2.

Gillard, J., Iles, T., 2005. Variance covariance matrices for linear regression with errors in both variables. Cardiff School of Mathematics Technical Report (accessed August 2017 at http://mathsdemo.cf.ac.uk/maths/resources/Iles_Gillard_Tech_Report.pdf).

GlobCurrent project document, 2017. Fifth validation report (VAL-5).

Gruber, A., Su, C.-H., Crow, W. T., Zwieback, S., Dorigo, W. A., Wagner, W., 2016a. Estimating error cross-correlations in soil moisture data sets using extended collocation analysis. J. Geophys. Res. Atmos. 121, 1208–1219, doi:10.1002/2015JD024027.

Gruber, A., Su, C.-H., Zwieback, S., Crow, W. T., Dorigo, W. A., Wagner, W., 2016b. Recent advances in (soil moisture) triple collocation analysis. Int. J. Appl. Earth Obs. Geoinform. 45, 200–211, doi:10.1016/j.jag.2015.09.002.

Hansen, D., Poulain, P.-M., 1996. Quality control and interpolations of WOCE-TOGA drifter data. J. Atmos. Oceanic Technol. 13, 900–909, doi:10.1175/1520–0426.

Hubert, M., Rousseeuw, P. J., Verdonck, T., 2012. A deterministic algorithm for robust location and scatter. J. Comp. Grap. Stats 21, 618–637, doi: 10.1080/10618600.2012.672100.

Jacquelin, J., 2014. Régressions et équations integrals (accessed 8 july 2016 at https://www.researchgate.net/file.postfileloader.html?id=5549ee13f15bc7863c8b456d&assetkey=as%3a273771590553603%401442283641886).

Janssen, P. A. E. M., Abdalla, S., Hersbach, H., Bidlot, J.-R., 2007. Error estimation of buoy, satellite, and model wave height data. J. Atmos. Oceanic Technol. 24, 1665–1677, doi:10.1175/JTECH2069.1.

Johnson, E. S., Bonjean, F., Lagerloef, G. S. E., Gunn, J. T., Mitchum, G. T., 2007. Validation and error analysis of oscar sea surface currents. J. Atmos. Oceanic Technol. 24, 688–701.

Kipnis, V., Carroll, R. J., Freedman, L. S., Li, L., 1999. A new dietary measurement error model and its application to the estimation of relative risk: Application to four validation studies. Am. J. Epidemiol. 150, 642–651, doi:10.1093/oxfordjournals.aje.a010063.

Kipnis, V., Midthune, D., Freedman, L. S., Bingham, S., Day, N. E., Riboli, E., Ferrari, P., Carroll, R. J., 2002. Bias in dietary-report instruments and its implications for nutritional epidemiology. Public Health Nutr. 5(6A), 915–923, doi:10.1079/PHN2002383.

Lumpkin, R., Grodsky, S. A., Centurioni, L., Rio, M.-H., Carton, J. A., Lee, D., 2013. Removing spurious low-frequency variability in surface drifter velocities. J. Atmos. Oceanic Technol. 30, 353–360, http://dx.doi.org/10.1175/JTECH–D–12–00139.1.

McColl, K. A., Vogelzang, J., Konings, A. G., Entekhabi, D., Piles, M., Stoffelen, A., 2014. Extended triple collocation: Estimating errors and correlation coefficients with respect to an unknown target. Geophys. Res. Lett. 41, 6229–6236, doi:10.1002/2014GL061322.

Moberg, A., Brattström, G., 2011. Prediction intervals for climate reconstructions with autocorrelated noise – an analysis of ordinary least squares and measurement error methods. Palaeogeo., Palaeoclim., Palaeoecol. 308, 313–329, doi:10.1016/j.palaeo.2011.05.035.

O'Carroll, A. G., Eyre, J. R., Saunders, R. W., 2008. Three-way error analysis between AATSR, AMSR-E, and in situ sea surface temperature observations. J. Atmos. Oceanic Technol. 25, 1197–1207.

Penven, P., Halo, I., Pous, S., Marié, L., 2014. Cyclogeostrophic balance in the Mozambique Channel. J. Geophys. Res. Oceans 119, doi:10.1002/2013JC009528.

Rio, M.-H., 2012. Use of altimeter and wind data to detect the anomalous loss of SVP-type drifters drogue. J. Atmos. Oceanic Technol. 29, 1663–1674, doi:10.1175/JTECH–D–12–00008.1.

Rio, M.-H., Hernandez, F., 2004. A mean dynamic topography computed over the world ocean from altimetry, in situ measurements, and a geoid model. J. Geophys. Res. 109, doi:10.1029/2003JC002226.

Rio, M.-H., Mulet, S., Picot, N., 2014. Beyond GOCE for the ocean circulation estimate: Synergetic use of altimetry, gravimetry, and in situ data provides new insight into geostrophic and Ekman currents. Geophys. Res. Lett. 41, doi:10.1002/2014GL061773.

Stoffelen, A., 1998. Toward the true near–surface wind speed: Error modeling and calibration using triple collocation. J. Geophys. Res. 103, 7755–7766, doi:10.1029/97JC03180.

Su, C.-H., Ryu, D., Crow, W. T., Western, A. W., 2014. Beyond triple collocation: Applications to soil moisture monitoring. J. Geophys. Res. Atmos. 119, 6419–6439, doi:10.1002/2013JD021043.

Sudre, J., Maes, C., Garçon, V., 2013. On the global estimates of geostrophic and Ekman surface currents. Limnol. Ocean.: Fluids Environ. 3, doi:10.1215/21573689–2071927.

Tarantola, A., 2005. Inverse Problem Theory and Methods for Model Parameter Estimation. SIAM, Philadelphia, Pennsylvania.

Tolman, H. L., 1998. Effect of observation errors in linear regression and bin–average analyses. Quart. J. Roy. Meteor. Soc. 124, 897–917, doi:10.1002/qj.49712454712.

Vogelzang, J., Stoffelen, A., Verhoef, A., Figa-Saldaña, J., 2011. On the quality of high-resolution scatterometer winds. J. Geophys. Res. 116, doi:10.1029/2010JC006640.

Yilmaz, M. T., Crow, W. T., 2013. The optimality of potential rescaling approaches in land data assimilation. J. Hydrometeor. 14, 650–660, doi:10.1175/JHM–D–12–052.1.

Yilmaz, M. T., Crow, W. T., 2014. Evaluation of assumptions in soil moisture triple collocation analysis. J. Hydrometeor. 15, 1293–1302, doi:10.1175/JHM–D–13–0158.1.

Zwieback, S., Scipal, K., Dorigo, W., Wagner, W., 2012. Structural and statistical properties of the collocation technique for error characterization. Nonlin. Process. Geophys. 19, 69–80, doi:10.5194/npg–19–69–2012.

Zwieback, S., Su, C.-H., Gruber, A., Dorigo, W. A., Wagner, W., 2016. The impact of quadratic nonlinear relations between soil moisture products on uncertainty estimates from triple collocation analysis and two quadratic extensions. J. Hydrometeor. 17, 1725–1743, doi:10.1175/JHM–D–15–0213.1.

9. Appendix

Measurement models (defined below) are actively evolving in various fields, with geophysical applications that may be unfamiliar or are just beginning to have an impact. The solution of such models is called an inverse problem (Tarantola, 2005), by contrast with evolution equations for mass, motion, and constituents as a forward model. It should be noted that longstanding experience in the geo-physical/biological/chemical communities with forward modelling and with taking high resolution (so-called longitudinal) observations provide the basis for estimating error autocorrelation (e.g., using FERS). A brief clarification of other concepts relevant to this study is offered here as a complement to more formal definitions. Online sources (e.g., Wikipedia) also provide recent and useful collaborative summaries. Concepts relevant to this study include:

- Affine calibration: synonymous with a linear calibration by intercept (α_N) and slope (β_N) parameters. Adjustment of the nowcast data (N) by these parameters is a good test of the retrieval method, as the adjusted nowcast should be unbiased ($\alpha_N \approx 0$ and $\beta_N \approx 1$). Regardless of the method, however, it is important to note that no bias correction can fully address a mismatch in support.

- Autoregressive (AR) parameterization: an established expression of information propagation; used here to encompass not just error autocorrelation in time or space but also error cross-correlation between two ocean current variates. The first order (AR-1) form explored here is the simplest.

- Competitive validation: evaluation of two or more datasets (or versions of a single dataset) against one chosen reference dataset, where the metric of success is shared true variance. Even if linear calibration is postulated (as in this study, rather than estimated from a measurement model), removal of linear bias from one dataset has no impact on shared truth, but this is not so for error. This approach was first attempted by Bentamy et al. (2017) in a comparison of heat flux estimates.

- Footprint: target area (e.g., at the ocean surface) that contributes to radiation received by a satellite sensor during an imaging interval. Unless it is possible to combine views of the same target area to synthesize higher resolution, the footprint often defines a support scale lower bound.

- Instrumental variable: additional data is often required when the measurement model has too many unknown parameters to estimate. A conventional instrument, following Fuller (2006), is a variable that is taken to be correlated with truth but not with error. The forecast and revcast (FERS) lagged variables, by comparison, involve correlation of both truth and error, but this is accommodated by their model equations. As instruments, FERS play the required role of facilitating the identification of all model parameters.

- Measurement model: measurement *error* models accommodate error in all sources of information [i.e., both in the calibrated and uncalibrated data; this accommodation is known as (Fuller, 2006) an approach to errors in variables in econometrics and observation error or measurement error in other fields]. There is no intended distinction between a measurement model and measurement error model. The sole rationale for omitting the term "error" is that a more balanced focus on truth and error can be anticipated. In other words, a regression model is effectively a truth model as much as it is an error model. However, only if it is possible to claim that a model does not lack any broad category of error (i.e., equation error or correlated error), does it seem justifiable to explore inferences based on truth.

- Parsimony: synonymous with simplicity, especially in reference to measurement models that minimize the number of parameters to be identified. That is, non-technical definitions apply (e.g., to a careful collection or use of data with minimal extra assumptions).

- Shared variance: synonymous with correlation and involving a term that appears in more than one of the measurement model equations of interest (possibly multiplied by a parameter). The concept of sharing applies to both truth and error. It is central to the idea that there can be multiple truths, with each containing information about overlapping data supports, and that measurement model assumptions should be considered when determining statistical independence. It should be noted that standard metrics, including the coefficient of determination or percentage of explained variance, correlation with truth (McColl et al., 2014), and SNR (Gruber et al., 2016b) are all subject to interpretation in terms of shared variance.

- Strong constraint: as an example, many equations of the GlobCurrent and drifter covariance matrix (6) are satisfied exactly as part of any measurement model solution (cf. weak constraint).

- Support: a characterization of the type (e.g., range or quality) of information that a given platform or instrument is sensitive to. Often this is with reference to spatial and temporal scales that can be resolved, but any information sensitivity can be included, which implies that such information may exist as truth or perhaps as equation error, according to the measurement model.

- Synergy: an approach to combining information such that the whole is more valuable and informative than the sum of individual contributions. Measurement modelling is an unlikely tool to prescribe how synergy could be achieved, but may permit the quantitative exploration of both individual contributions and informed attempts to combine information.

- Triple collocation: following McColl et al. (2014), the model parameters sought are uncorrelated error variance of three independent datasets, and with one dataset as a reference, additive and multiplicative calibration of the other two. Following Stoffelen (1998), this measurement model implicitly includes cross-correlated error (e.g., representativeness error) because three different sources of information invariably have three different supports, so at least between two information sources with broader support (e.g., higher resolution), error cross-correlation would be expected.

- True variance estimation: curves of the LHS-RHS of the autocovariance equations (7) are each characterized by a single localized minimum and flatness elsewhere in the range of zero to $Var(I)$. The present study treats each available minimum as an equally good estimate of shared true variance and their average is taken. This is in contrast to a global minimum sought using the average of all such curves. However, minima are often not overlapping so the global minimum is effectively a selection among one of the six possible minima. This implies a reliance on the accuracy of each curve in representing its own (very small) minimum value, which might be ill advised.

- Weak constraint: as an example, the autocovariance equations provide different target estimates of shared true variance that cannot all be satisfied simultaneously; a solution close to the center of the ensemble is thus adopted (cf. strong constraint).