An ocean surface current analysis (GlobCurrent) calibration and validation

Conference Paper · April 2017

Rick Danielson
Nansen Environmental and Remote Sensing …

Johnny A. Johannessen
Nansen Environmental and Remote Sensing …

Fabrice Collard
OceanDataLab

Graham D. Quartly
Plymouth Marine Laboratory

Inference Error Model

Spatial and temporal coverage of the *intersection* of two (or more) datasets can be orders of magnitude *smaller* than the coverage of just one gridded dataset. In this sense, collocations only allow one to *infer* the bias and performance of a full dataset. However, such inferences are expected to be useful, as it is common for gridded atmospheric and oceanic data to benefit from observations in assimilation windows, across footprints, and within influence radii that are typically as large or larger than the grid interval on which a true or target variable is represented. Nonlocal, propagated, or shared signal and noise is the norm. Although some frameworks for assessing bias (e.g., ordinary least squares regression or triple collocation) assume independent errors, a corresponding framework for slowly varying and well resolved (correlated) error is worth exploring. There appears to be a family of error models (of which the following is a member) that provide the simplest possible framework for further exploration:

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

• The so-called INFERS error model consists of two datasets: an in situ estimate and a gridded estimate where the nowcast is collocated in space and time with the in situ estimate and the forecast and revcast are simply samples taken at adjacent locations on the grid (e.g., "persistence" over one or two days is our forecast/revcast method here). Each gridded estimate has its own additive and multiplicative bias (α and β), although only the nowcast bias (IN regression slope and intercept) is essential for a GlobCurrent recalibration.

• The in situ data is taken to be unassimilated above (i.e., GlobCurrent did not assimilate drifters, except in defining a time-invariant mean dynamic topography; Rio et al. 2014). However, a parameterization of shared or propagated error in the gridded data (NFERS) is quantified by a retrieval of the λ coefficients. Note that error in the INFERS model employs an AR-1 autoregressive form because this is arguably the simplest form. (The λ retrievals are discussed elsewhere.)

• It follows from our application of an AR-1 error model that the minimum number of equations (or samples of the gridded dataset) to match the total number of unknowns is four (NFER or NFRS). The symmetry of five samples (NFERS) appears to facilitate a retrieval of all parameters, which is done partly analytically and partly numerically.

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

$$
\begin{aligned}
Cov(I,N) &= \beta_N \sigma_t^2 \\
Cov(I,F) &= \beta_F \sigma_t^2 \\
Cov(I,E) &= \beta_E \sigma_t^2 \\
Cov(I,R) &= \beta_R \sigma_t^2 \\
Cov(I,S) &= \beta_S \sigma_t^2 \\
Cov(N,F) &= \beta_N \beta_F \sigma_t^2 + \lambda_F \sigma_N^2 \\
Cov(N,E) &= \beta_N \beta_E \sigma_t^2 + \lambda_E \lambda_F \sigma_N^2 \\
Cov(N,R) &= \beta_N \beta_R \sigma_t^2 + \lambda_R \sigma_N^2 \\
Cov(N,S) &= \beta_N \beta_S \sigma_t^2 + \lambda_S \lambda_R \sigma_N^2
\end{aligned}
$$

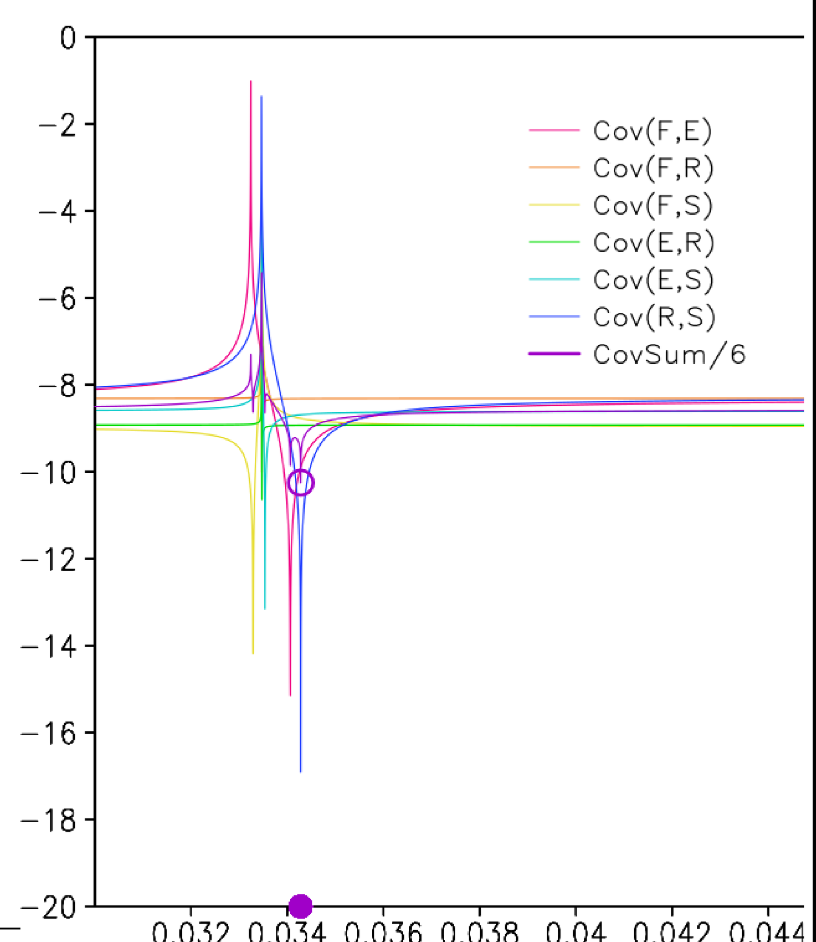
• All unknown parameters except true variance, σ_t^2, are determined analytically from the following subset of the full covariance matrix (cf. above):

• Remaining terms of the covariance matrix provide six equations, including the covariance between E and S (the extended forecast and revcast). In other words, the full covariance matrix is employed and the correlation between all samples (NFERS) is used to better estimate true variance σ_t^2.

• LHS minus RHS of each six covariance equations should be zero. A sum of the log of their absolute difference(divided by six) is minimized for our best estimate of true variance

• Simulations using known values of all error parameters yields pleasing retrievals (see the NERSC "Collocation" demonstration repository at https://github.com/nansencenter/Collocation)

• Global parameters are retrieved from **O[10^7]** collocations after separation into even and odd year groups. Zonal current retrievals are almost identical, whereas meridional current finds true variance minima at the boundary (ordinary or reverse least squares) search domain because the global minimum is associated with a negative error variance, which is not presently allowed (despite INFERS error model simplicity).



Even Years Odd Years	Drift RMSE	Drift SNR	Analys Bias Addit	Analys Bias Multi	Analysis RMSE	Analysis SNR
Zonal	0.098	11.5	0.000	0.892	0.075	14.6
Current	0.101	11.5	0.001	0.887	0.076	14.7
Meridional	0.115	-0.1	0.001	1.097	0.044	20.8
Current	0.012	52.9	0.002	0.547	0.103	-2.4

An ocean surface current analysis (GlobCurrent) calibration and validation

Rick Danielson, Ocean and Coastal Remote Sensing, Nansen Environmental and Remote Sensing Center,

Johnny Johannessen (NERSC), Marie-Hélène Rio (CLS), Fabrice Collard (ODL), Craig Donlon (ESA), Bertrand Chapron (Ifremer), and Graham Quartly (PML)

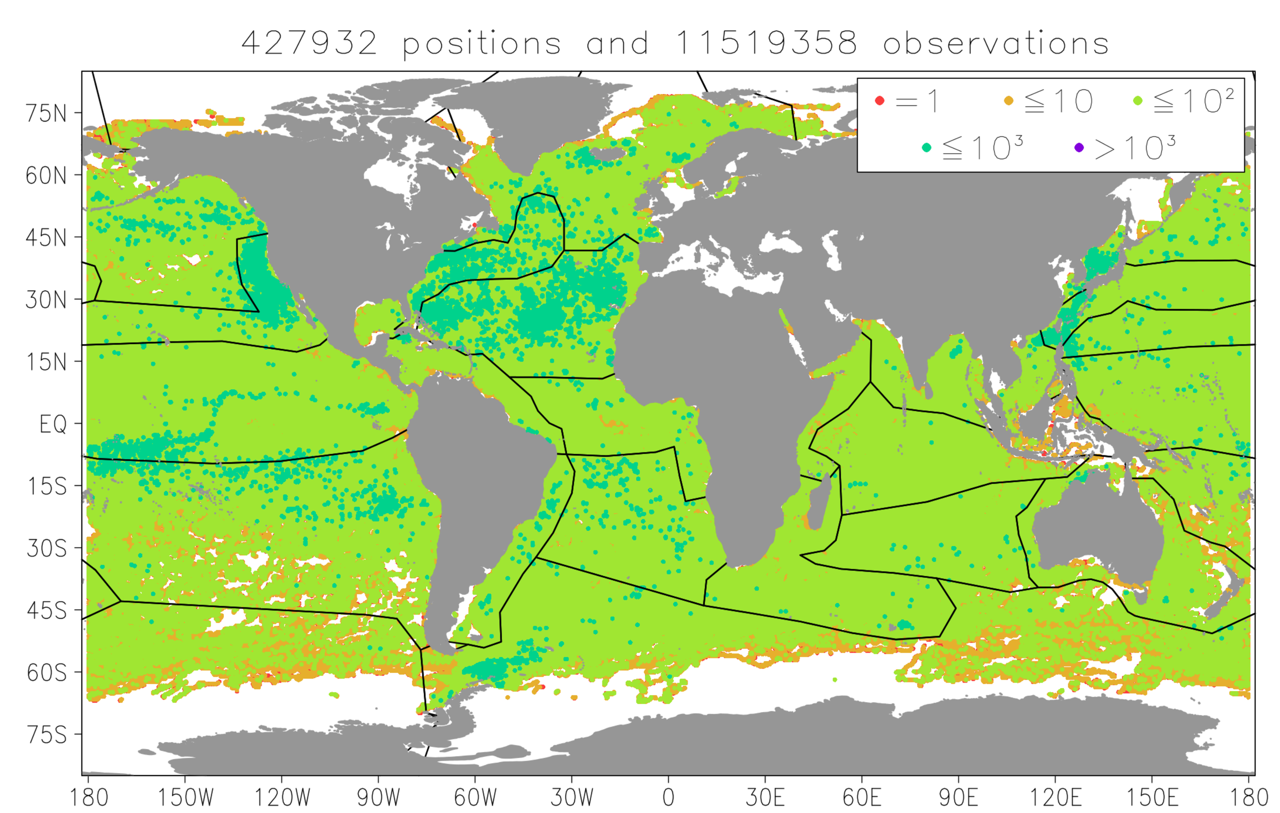
Introduction

Observations of extreme conditions, characterized by high heat flux, rapidly changing surface wind, or strong ocean current, are rare. Although analyses provide estimates of these conditions, because there are few observations to begin with, it is difficult to calibrate and validate an analysis or a retrieval assuming independent observations (cf. Stoffelen 1998). This requirement of independence may not be so dire, however, if we start with an error model that accommodates a nonlocal impact of observations. **We propose that a novel exploitation of error correlation may provide great freedom to improve analyses and retrievals.** In turn, we seek not only to assess performance, but also to suggest that a more complete analysis or retrieval is one that includes a first order calibration and validation against any other high quality reference. The example of ocean current (Rio et al. 2014) analysis calibration is given with reference to drifting buoys. The most obvious physical quantities (current and wind speed) are chosen to gauge measures of performance across an entire range, including both weak and extreme conditions.

Identification of Collocations

• The GlobCurrent analysis (Rio et al. 2014) is a combined estimate of the geostrophic and Ekman components (based on altimetric and scatterometric observations) that is valid at the surface and at 15-m depth. Here, drifters moving roughly with the 15-m current are used as a calibration reference for 15-m analyses that we sample at daily intervals. The complete Version-3 analysis (1993-2015) and all collocated drifters whose drogues remained attached are employed.

• More than eleven million velocity estimates are available from drifters that likely retained their drogues (Rio et al. 2012). Divisions of the global oceans are available for many purposes (e.g., economic and scientific divisions are given at http://www.marineregions.org) but a division by surface current regime is not widely available. Climatological ocean surface currents are generally well documented (e.g., on wikipedia and at http://oceancurrents.rsmas.miami.edu), but without proposing corresponding limits. Hence, we propose a division of the global ocean that is informed by such documented sources (in particular the RSMAS website) and whose limits follow the CNES-CLS2013 climatology (a two-decade mean geostrophic current derived from a combined satellite and in situ mean dynamic topography) described by Rio et al. (2014).



Acknowledgements

This work is funded by the FP7 E-GEM and ESA GlobCurrent and Ocean Heat Flux projects.

References

Gruber, A., C.H. Su, S. Zwieback, W. Crow, W. Dorigo, and W. Wagner, 2016: Recent advances in (soil moisture) triple collocation analysis. Int. J. Appl. Earth Obs. Geoinf., 45, 200-211.

McColl, K. A., J. Vogelzang, A. G. Konings, D. Entekhabi, M. Piles, and A. Stoffelen, 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.

Stoffelen, A., 1998: Toward the true near-surface wind speed: Error modeling and calibration using triple collocation, J. Geophys. Res., 103(C4), 7755–7766, doi:10.1029/97JC03180.

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

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

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

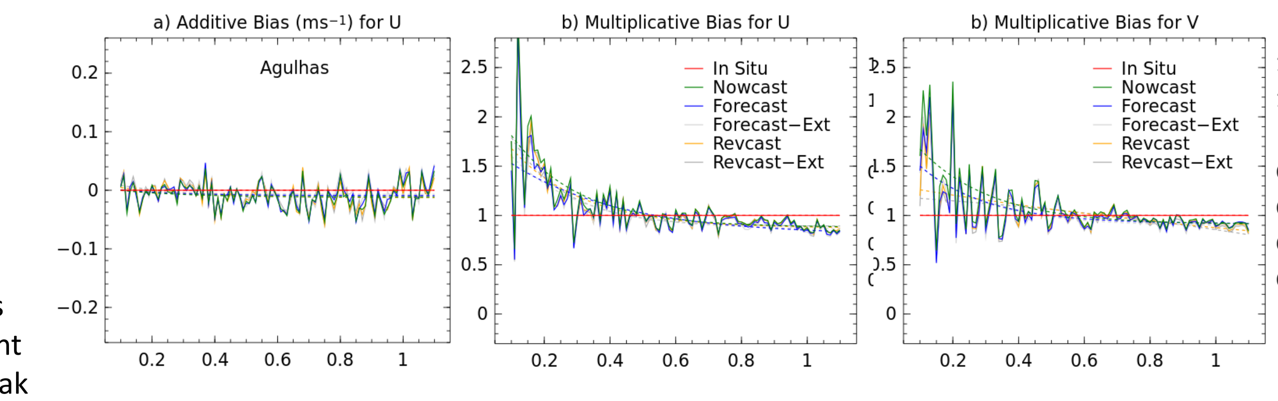
Vogelzang, J. and A. Stoffelen 2012: Triple collocation. EUMETSAT Report. [Accessed September 2015 at https://nwpsaf.eu/deliverables/scatterometer/TripleCollocation_NWPSAF_TR_KN_021_v1_0.pdf.

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

Regional Characterizations

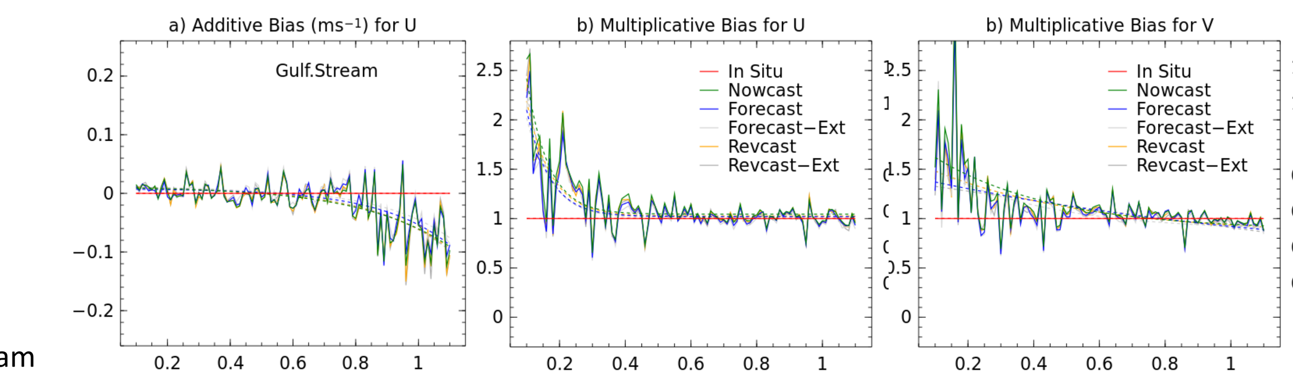
Consideration of all collocations (in the lower left table) reveals little additive bias, but a slight multiplicative bias indicating that GlobCurrent's 15-m current speed could be a bit faster (if drifters are the best calibration reference; note that they are expected to resolve current variability at scales that GlobCurrent cannot, which contributes to their higher RMSE also). The great number of collocations allows for a physically motivated subsetting, however. We opt to look at particular regions and at calibration as a function of drifter current speed. Thus, the nearest **200** collocations to current speed at intervals of 0.01m/s are mapped.

Agulhas



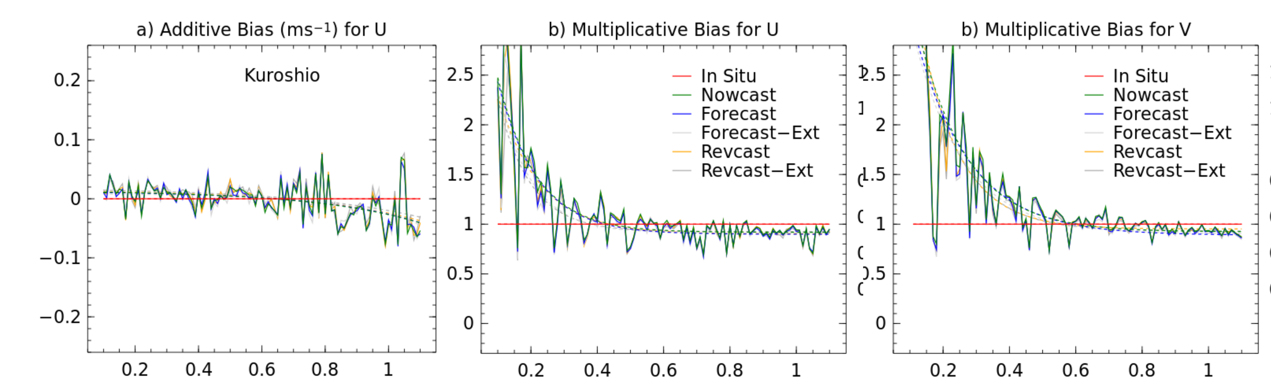
As with the global collocations, there is little additive bias in the Agulhas current analyses. Multiplicative bias is also weak at strong current speed, but weak currents may be too strong in this region. A recalibration of the combined geostrophic and Ekman current to be more consistent with the drifter response to the Agulhas could thus be performed.

Gulf Stream



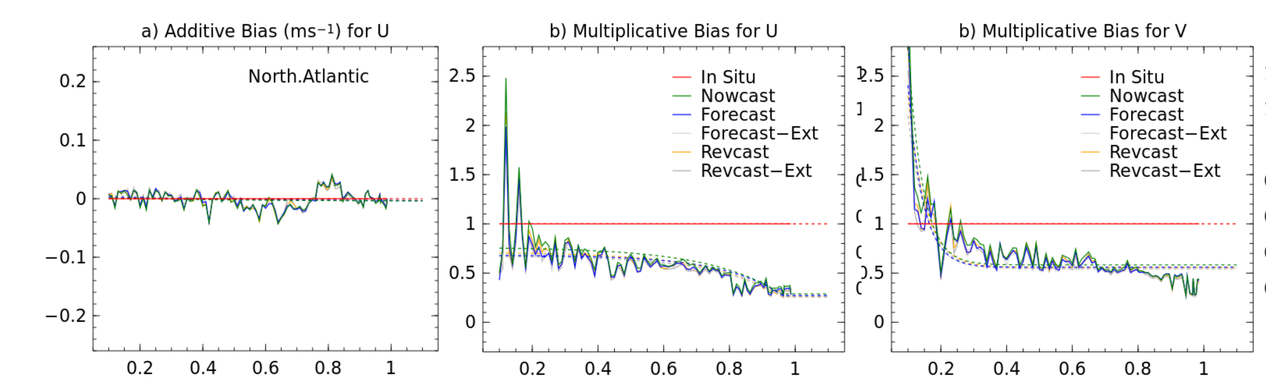
Additive bias is minor in the Gulf Stream region except at high current speed, where the bias is slight. Strong currents appear to be well resolved in the GlobCurrent version-3 analysis. As for the Agulhas, weak current may be too strong, but it can be noted that there is a wide range between ordinary and reverse regression at weak current speed, so the multiplicative bias is less clear than at higher currents.

Kuroshio



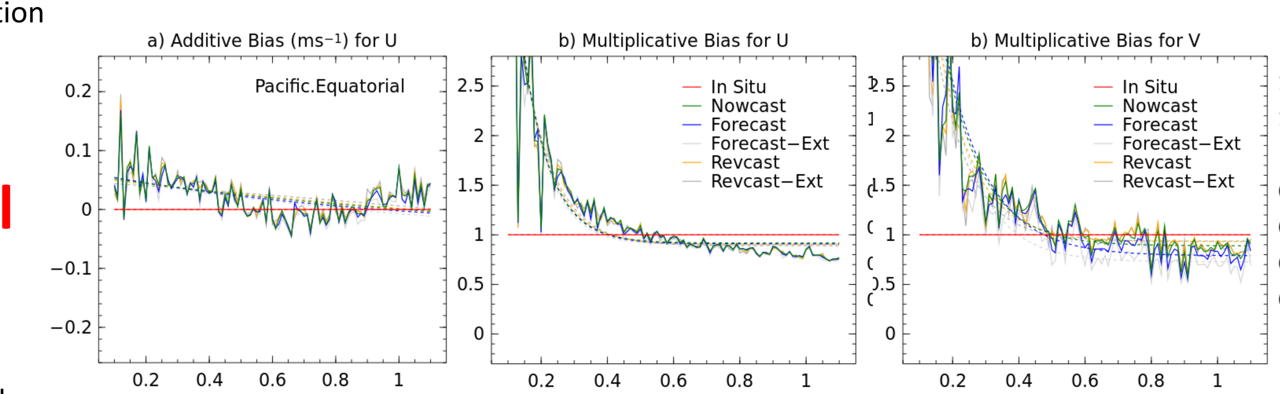
Multiplicative bias (an overprediction of weak current speed) seems more apparent in the Kuroshio than in the Gulf Stream, but again strong current shows little bias. (This is likely particular to the latest GlobCurrent version, with greater bias in earlier versions.)

North Atlantic



The North Atlantic region is associated with much weaker current speed overall than in the boundary currents above. Here, there appears to be an underprediction of current speed over a wide range and moreso at high speed. It is notable that ordinary and reverse regression slope are more consistent with each other and support this interpretation (not shown).

Pacific Equatorial



The Pacific Equatorial region has a weak additive bias trend and a strong overprediction of weak current. An assessment of bias for all analysis samples (NFERS) shown above is consistent across the regions. There is also consistency in both the zonal and meridional current component biases. Retrieval of bias using the INFERS method reveals trends as a function of current speed. Following Tolman (1998), the benefit of taking numerous small samples (and ignoring some) is apparent.

Conclusions

A model for regression with correlated error, with sufficient information to constrain a nonlinear relationship between two datasets by multiple samples from the GlobCurrent gridded analysis, has been proposed. The model does not suffer from a neglect of autocorrelated errors (see also Zwieback et al. 2012 and Su et al. 2014); not only does it benefit from them, it requires them! Evidence is given of regional differences in calibration and performance of the GlobCurrent analysis. The tentative conclusion is that either the GlobCurrent weak current is too strong or the strong currents are slightly weak, and sometimes bias exists in both senses. We propose to verify this by an improvement in trajectory calculation skill following recalibration.