https://doi.org/10.5194/amt-2020-95
Preprint. Discussion started: 23 April 2020

Atmospheric Measurement Techniques Discussions

Uncertainty Quantification for Atmospheric Motion Vectors with Machine Learning

Joaquim V. Teixeira[1], Hai Nguyen[1], Derek J. Posselt[1], Hui Su[1], Longtao Wu[1]

[1]Jet Propulsion Laboratory, California Institute of Technology

Abstract. Wind-tracking algorithms produce Atmospheric Motion Vectors (AMVs) by tracking clouds or water vapor across spatial-temporal fields. Thorough error characterization (also known as uncertainty quantification) of wind-tracking algorithms is critical in properly assimilating AMVs into weather forecast models and climate reanalysis datasets. Uncertainty quantification should yield estimates of two key quantities of interest: bias, the systematic difference between a measurement and the true value, and standard error, a measure of variability of the measurement. The current process of specification of the errors input into inverse modelling is often cursory and commonly consists of a mixture of model fidelity, expert knowledge, and need for expediency. The methods presented in this paper supplement existing approaches to error specification by providing an error-characterization module that is purely data-driven and requires few tuning parameters. This paper proposes an error-characterization method that combines the flexibility of machine learning (random forest) with the robust error estimates of unsupervised parametric clustering (using a Gaussian Mixture Model). Traditional techniques for uncertainty quantification through machine learning have focused on characterizing bias, but often struggle when estimating standard error. In contrast, model-based approaches such as k-means or Gaussian mixture modelling can provide reasonable estimates of both bias and standard error, but they are often limited in complexity due to reliance on linear or Gaussian assumptions. In this paper, a methodology is developed and applied to characterize error in tracked-wind using a high-resolution global model simulation, and it is shown to adequately capture the error features of the tracked wind.

1. Introduction

Reliable estimates of global winds are critical to science and application areas, including global chemical transport modeling and numerical weather prediction. One source of wind measurements consists of feature-tracking based Atmospheric Motion Vectors (AMVs), produced by tracking time sequences of satellite-based measurements of clouds or spatially distributed water vapor fields (Mueller et al., 2017; Posselt et al., 2019). The importance of global measurements of 3-dimensional winds was highlighted as an urgent need in the NASA Weather Research Community Workshop Report (Zeng et al., 2016) and was identified as a priority in the 2007 National Academy of Sciences Earth Science and Applications from Space (ESAS 2007) Decadal Survey and again in ESAS 2017. For instance, wind is used in the study of global CO_2 transport (Kawa et al., 2004), numerical weather prediction (NWP; Cassola and Burlando, 2012), as inputs into weather and climate reanalysis studies (Swail and Cox, 2000), and for estimating current and future wind-power outputs (Staffell and Pfenninger, 2016).

Thorough error characterization of wind-track algorithms is critical in properly assimilating AMVs into forecast models. Prior literature has explored the impact of ‘poor’ error-characterization in Bayesian-based approaches to remote sensing applications. Nguyen et al. (2019) proved analytically that when the input bias is incorrect in Bayesian methods (specifically, optimal estimation retrievals), then the posterior estimates would also be biased. Moreover, they proved that when the input standard error is 'correct' (that is, it is as close to the unknown truth as possible), then the resulting Bayesian estimate is 'efficient'; it has the smallest possible error. Additionally, multiple active and passive technologies are being developed to measure 3D winds, such as Doppler wind lidar (DWL) and radar and infrared/microwave sensors that derive AMVs using feature-tracking of consecutive images. Therefore, an accurate and robust uncertainty quantification methodology will allow for more accurate assessments of mission impacts, and the eventual propagation of data uncertainties for these instruments.

Velden and Bedka (2009) and Salonen et al. (2015) have shown that height assignment contributes a large component of uncertainty in AMVs tracked from cloud movement and from sequences of infrared satellite radiance images. However, height assignment is not the dominant portion of the error in AMVs obtained from water vapor profiling instruments (e.g., infrared and microwave sounders). As such, this study will focus on errors in the AMV estimates at a given height. Previous work has demonstrated several different approaches for characterizing AMV vector error. One common approach is to employ quality indicator thresholds, as described by Holmund et al (2001), which compare changes in AMV estimates between sequential timesteps and neighboring pixels, as well differences with model predictions, to produce a quality indicator to which a discrete uncertainty is assigned. The Expected Error approach, developed by Le Marshal et al. (2004), builds a statistical model using linear regression against AMV-radiosonde values to correct AMV observation error.

In this study, we detail a data-driven tool for building an AMV uncertainty model using observing system simulation experiment (OSSE) data. We build on the work by Posselt et al. (2019) in which a water vapor feature-tracking AMV algorithm was applied to a high-resolution numerical simulation, thus providing a global set of AMV estimates which can be compared to the reference winds produced by the simulation. In this case, a synthetic "true" state is available with which AMVs can be compared and errors are quantified, and it is shown that tracking errors in AMV estimates are state dependent. Our approach will use a conjunction of machine learning (random forest) and unsupervised parametric clustering (Gaussian mixture models) to build a model for the uncertainty structures found by Posselt et al. (2019). The realism and robustness of the resulting uncertainty estimates depend on the realism and representativeness of the reference dataset. This work builds upon the work of Bormann et al. (2014) and Hernandez-Carrascal and Bormann (2014), who showed that wind tracking could be divided into distinct geophysical regimes by clustering by cloud conditions. This study supplements that approach with the addition of machine learning, which, compared with traditional linear modeling approaches, should allow the model to capture more complex non-linear processes in the error function.

Traditional techniques for uncertainty quantification through machine learning have focused on characterizing bias but often struggle when estimating standard error. By pairing a random forest algorithm with unsupervised parametric clustering, we propose a data-driven, cluster-based approach for quantifying both bias and standard error from experimental data. According to the theory developed by Nguyen et al. (2019), these improved error characterizations

https://doi.org/10.5194/amt-2020-95
Preprint. Discussion started: 23 April 2020

Atmospheric Measurement Techniques Discussions

should then lead to improved error characteristics (e.g., lower bias, more accurate uncertainties) in subsequent analyses such as flux inversion or data assimilation.

The rest of the paper is organized as follows: In Section 2, we give an overview of the simulation which provides the training data for our machine learning approach and motivate and define the specific uncertainties this study aims to characterize. In Section 3, we describe the error characterization approach with the specifics of our error characterization model, including both the implementation of and motivations for employing the random forest and Gaussian mixture model. In Section 4, we provide a validation of our methods, attempting to assess the bias of our predictions. In Section 5, we discuss the implications of our error characterization approach, both on AMV estimation and data assimilation more broadly.

2. Experimental Set-up

2.1 Simulation and Feature-Tracking Algorithm

While our methodology in principle could be used to quantify uncertainties in any measurements used in data assimilation, in this paper we devote special emphasis to the use case of wind-tracking algorithms. In particular, we trained our model on the simulated data used by Posselt et al. (2019), in which they applied an AMV algorithm to outputs from the NASA Goddard Space Flight Center (GSFC) Global Modeling and Assimilation Office (GMAO) GEOS-5 Nature Run (G5NR; Putman et al. 2014). The Nature Run is a global dataset with ~7 km horizontal grid spacing that includes, among other quantities, three-dimensional fields of wind, water vapor concentration, clouds, and temperature. The AMV algorithm is applied on four pressure levels (300hPa, 500hPa, 700hPa, and 850hPa) at 6-hourly intervals, using three consecutive global water vapor fields spaced one hour apart, and for a 60-day period from 07/01/2006 to 08/30/2006. The water-vapor fields from GEOS5 were input to a local-area pattern matching algorithm that approximates wind speed and direction from movement of the matched patterns. The algorithm searches a pre-set number of nearby pixels to minimize the sum-of-absolute-differences between aggregated water vapor values across the pixels. Posselt et al. (2019) describes the sensitivity of the tracking algorithm and the dependency of the tracked winds on atmospheric states in detail.

It is important to note that the AMV algorithm tracks water vapor on fixed pressure levels. In practice, these would be provided by satellite measurements, whereas in this paper we use simulated water vapor from the GEOS-5 Nature Run. The height assignment of the AMVs is assumed to be perfectly known (or, at the very least, the pressure level uncertainty is captured by the satellite measurement uncertainty rather than the AMV estimate). As such, we focus solely on observational AMV error and not on height assignment error.

A snapshot of the dataset at 700hPa is given in Figure 1, where we display the true water vapor from Nature Run (top left panel), the true wind speed from Nature Run (top right panel), the tracked wind from the AMV-tracking algorithm (bottom right panel), and the difference between the true and tracked wind (bottom left panel). Note that the wind-tracking algorithm tends to have trouble in region where the true water vapor content is close to zero. It is clear that

https://doi.org/10.5194/amt-2020-95
Preprint. Discussion started: 23 April 2020

Atmospheric Measurement Techniques Discussions
Open Access
EGU

while the wind-tracking algorithm tends to perform well in most regions (we can classify these regions are areas where the algorithm is skilled), in some regions the algorithm is unable to reliably make a reasonable estimate of the wind speed (unskilled). We will examine these skilled and unskilled regimes (and their corresponding contributing factors) in the section 3.

2.2 Importance of Uncertainty Representation in Data Assimilation

Proper error characterization for any measurement, including AMVs, is important in data assimilation. Data assimilation often uses a regularized matrix inverse method based on Bayes' theorem, which, when all probability distributions in Bayes' relationship are assumed to be Gaussian, reduces to minimizing a least-squares (quadratic) cost function Eq (1):

$$\mathbf{J} = (\mathbf{x} - \mathbf{x_b})\mathbf{B}^{-1}(\mathbf{x} - \mathbf{x_b}) + \left((\hat{\mathbf{y}} - \mathbf{a}) - \mathbf{H}[\mathbf{x}]\right)^{\mathrm{T}}\mathbf{R}^{-1}\left((\hat{\mathbf{y}} - \mathbf{a}) - \mathbf{H}[\mathbf{x}]\right) \quad (1)$$

where x represents the analysis value, x_b represents the background field (first guess), B represents the background error covariance, y represents the observation, and H represents the forward operator that translates model space into observation space. This translation may consist of spatial and/or temporal interpolation if x and y are the same variable (e.g., if the observation of temperature comes from a radiosonde), or may be far more complicated (e.g., a radiative transfer model in the case of satellite observations). R represents the observation error covariance, and a represents the accuracy, or bias, in the observations. The right-most part of Eq. (1) can be interpreted as a sum of the contribution of information from the data (y – H[x] - a) and the contribution from the prior ($x - x_b$), which are weighted by their respective covariance matrices. In our analysis, the AMVs obtained from the wind-tracking algorithm is used as 'data' in subsequent analysis. That is, the tracked wind data $\hat{y}$ is a biased and noisy estimator of the true wind y, and might be assumed to follow the model Eq. (2):

$$\hat{y} = y + \epsilon \quad (2)$$

where ϵ is an error term, commonly assumed to be Gaussian with mean **a** and covariance matrix **R** (i.e., $\epsilon \sim N(\mathbf{a}, \mathbf{R})$), which are the same two terms that appear in Equation (1). As such, for data assimilation to function, it is essential to correctly specify the bias vector **a** and the standard error matrix $\mathbf{R}^{-1}$. Incorrect characterizations of either of these components could have adverse consequences on the resulting data assimilation analyses with respect to bias and/or the standard error (Nguyen et al., 2019).

3 Methodology

3.1 Generalized Error Characterization Model

An overview of our approach is outlined in Figure 2. Given a set of training predictors X, training responses $\hat{Y}$, and the true response Y, our approach begins with two independent steps. In one step, a Gaussian mixture model is trained on the set of X, Ŷ, and Y. This clustering algorithm identifies geophysical regimes where the nonlinear relationships between the three variables differ. In the other step, a random forest is used to model Y based on X and Ŷ. This step produces an estimate of the true response (we call this $\tilde{Y}$) using only the training predictors and response. We then employ the Gaussian mixture model to estimate the clusters which the set of X, Ŷ, and $\tilde{Y}$ pertain to. Subsequently, we compute the error characteristics of each cluster of X, Ŷ, and $\tilde{Y}$ in the training dataset. Therefore, given a new point consisting solely of X and Ŷ, we can assign it to a specific cluster and ascribe to it a set of error characteristics. This forms the basis for our error characterization model.

What follows in this paper is an implementation of the error characterization model obtained for a subsample of the GEOS-5 Nature Run at a fixed height of 700hPa. In particular, we trained the error characterization on a random sample from the first 1.5 months of the Nature Run, and show the results obtained when applying it to a test sample drawn from the subsequent 0.5 months of the Nature Run.

3.2 Error Regime

When examining the relationship between AMVs and simulated true winds in Figure 3**,** it is clear that there are two distinct 'error-regimes' present in the dataset. The majority of AMV estimates can be categorized as 'skilled', wherein their estimate lies clearly along a one-to-one line with the simulated true wind. However, there is also clearly an 'unskilled' regime, for which the AMV estimate is very close to zero when there are actually high or mid-level true wind values present. Our goal is to provide unique error characterizations for each error regime, because the error dynamics are different within each regime. Furthermore, when we analyze this error and its relationship to water vapor, we see that 'unskilled' regime correlates highly with areas of low water vapor in Figure 4. This matches the error patterns discussed in Posselt et al. (2019).

3.3 Gaussian Mixture Model

These distinct regimes present an opportunity to employ machine learning. Bormann et al. (2014) and Hernandez-Carrascal and Bormann (2014) demonstrated that cluster (also called regime) analysis is a successful approach for wind-tracking error characterization, and so we aim to train a clustering algorithm that is capable of determining whether any individual AMV estimate belongs in the 'skilled' or 'un-skilled' cluster. In particular, we use a clustering algorithm that can take advantage of the underlying geophysical dynamics, since we see the relationship between the error-regimes and water vapor content. To this end, we employ a Gaussian mixture model, a clustering algorithm based on estimating a training set as a mixture of multiple Gaussian distributions. A mathematical overview follows:

1. Define each location containing simulated true winds, water vapor, and AMV estimates as a random variable x_i
2. Define θ as the population that consists of all x_i in the training dataset
3. Model the distribution of the population P(θ) as:

https://doi.org/10.5194/amt-2020-95
Preprint. Discussion started: 23 April 2020

Atmospheric Measurement Techniques Discussions

$$P(\theta) = \sum_{j}^{K} \pi_j N(\mu_j, \Sigma_j) \quad (3)$$

Where $N(\mu_j, \Sigma_j)$ is the normal distribution with mean μ_j and covariance Σ_j,

K is the number of clusters, and π_j is the mixture proportion.

3. An Expectation–Maximization Algorithm determines π_j, μ_j, Σ_j for K clusters
4. Density estimation gives us $P(x_i \in k_j) = p_{ij}$
5. Maximum p_{ij} is the assigned cluster for point x_i

The mixture model clustering is based on the R package 'Mclust' developed by Fraley et al. (2012), which builds upon the theoretical work of Fraley and Raftery (2002) for model-based clustering and density estimation. The process uses an Expectation-Maximization algorithm to cluster the dataset, estimating a variable number of distinct multivariate Gaussian distributions from a sample dataset. Training the Gaussian mixture model on this dataset provides a clustering function which outputs a unique cluster for any data point with the same number of variables.

In one dimension, a Gaussian mixture model looks like the distributions depicted in Figure 5: instead of modelling a population as a single distribution (Gaussian or otherwise), the GMM algorithm fits multiple Gaussian distributions to a population. A key aspect is that this algorithm has the capability of assigning a new point to the most likely distribution. For example, in the 1-D figure, a normalized AMV estimate with a value of 10 would be more likely to originate from the broad cluster '2' than the narrow cluster '4'. In this case, we model the population as a Gaussian mixture model in five-dimensional space, which consists of two simulated true wind vector components (u and v), two AMV estimates of these wind components ($\hat{u}$ and $\hat{v}$), and the simulated water vapor values, all of which have been standardized. Each cluster has a 5-dimensional mean vector for the center and a 5x5 covariance matrix defining their multivariate Gaussian shape. The estimation of a covariance matrix allows for the characterization of the relationships between the different dimensions within each cluster, and as such the gaussian mixture model approach provides greater potential for understanding the geophysical basis of error regimes than other unsupervised clustering approaches.

In Figure 6, we applied the Gaussian mixture model to true u and v wind data using 9 clusters. Although Figure 4 indicates that the data tends to separate roughly into 'skilled' and 'unskilled' regimes, we opted to choose 9 clusters in the Gaussian mixture model after several sensitivity tests across all pressure levels found 9 to be the minimum number of clusters needed to ensure the separation into these separate regimes, as well as allowing for further stratification of sub-regimes within the skilled and unskilled regimes

By re-analyzing the AMV estimate in relation to the simulated true winds, separated into the cluster that each point has been assigned to (Figure 6), we find that the clustering approach successfully separates the AMV estimates according to their 'skillfulness'. Essentially, we repeat Figure 3 but divide the AMV estimates by cluster. We see that,

https://doi.org/10.5194/amt-2020-95
Preprint. Discussion started: 23 April 2020

Atmospheric Measurement Techniques Discussions

for example, clusters 4, 5, and 7 clearly represent cases in which the feature-tracking algorithm provides an accurate estimate of the true winds, with very low variance around the one-to-one line (i.e., low estimate errors). Clusters 1, 2, 3, and 9 are somewhat noisier than the low-variance clusters, with error characteristics similar to those of the entirety of the dataset. Clusters 6 and 8, on the other hand, are clearly unskilled in different ways. Cluster 6 is a noisy regime, which captures much of the more extreme differences between the AMV estimates and the simulated true winds. Cluster 8, on the other hand, represents the low AMV estimate, high true wind regime. This cluster is returning AMVs with values of zero where the true wind is clearly non-zero because of the very low water vapor present. We see that the clustering algorithm succeeds in capturing physically interpretable clusters without having any knowledge of the underlying physical dynamics. We further see the stratification of the regimes when analyzing the absolute AMV error in relation to the water vapor content (Figure 7). We see that clusters that have similar behaviors in the error pattern (such as 1, 2, and 3) represent different regimes of water vapor content.

3.5 Random Forest

The clustering algorithm requires the true wind vector component values (u and v) in order to classify the AMV error. When applying the algorithm in practice to tracked AMV wind from real observations, the true winds are unknown. Therefore, we develop a proxy for the true winds using only the AMV estimates and the simulated water vapor itself. This is an instance in which the application of machine learning is desirable, since machine learning excels at learning high-dimensional non-linear relationships from large training datasets. In this case, we specifically use random forest to create an algorithm which predicts the true wind values as a function of the tracked wind values and water vapor.

Random forest is a machine learning regression algorithm which, as detailed by Breiman (2001), employs an ensemble of decision trees to model a nonlinear relationship between a response and a set of predictors from a training dataset. Here, we chose random forest specifically because it possesses certain robustness properties that are more appropriate for our applications than other machine learning methods. For instance, random forest will not predict values that are outside the minimum and maximum range of the input dataset, whereas other methods such as neural networks can certainty exceed the training range, sometimes considerably so. Random forest, due to the sampling procedure employed during training, also tends to be robust to overtraining in addition to requiring fewer tuning parameters compared with methods such as neural networks.

We trained a random forest with 50 trees on a separate set of tracked winds and water vapor values to predict true winds using the ‘randomForest’ package in R. While the random forest estimate as a whole does not perform much better than the AMV values in estimating the true wind (2.89 RMSE for random forest vs 2.91 RMSE for AMVs), as shown in Figure 8, it does not display the same discrete regimentation as the AMV estimates in Figure 3. Relative to the AMV estimates, the error in each of the random forest estimates is closer to the mean of error of the entire dataset. As such, the random forest estimates can act as a proxy for true wind values in our clustering algorithm — they remove the regimentation which is a critical distinction between the AMV estimates and the true wind values.

https://doi.org/10.5194/amt-2020-95
Preprint. Discussion started: 23 April 2020

Atmospheric Measurement Techniques
Discussions

3.6 Finalized Error Characterization Model

The foundation of the error characterization approach is to combine the random forest and clustering algorithm. We apply the Gaussian mixture model, as trained on the true winds (in addition to the AMVs and water vapor), to each point of water vapor, AMV estimate, and associated random forest estimate. This produces a set of clusters which, when implemented, require no direct knowledge of the actual true state (Figure 9). We see that the algorithm manages to separate the AMV estimates into appropriate error clusters. Once again, clusters 6 and 8 manage to capture unskilled regimes, and clusters 4 and 5 remain extremely skillful. While there is some degradation in the performance relative to the classification algorithm on the training set, we see in Figure 9 and Figure 10 that the error characterization still discretizes the testing data set into meaningful error regimes.

By taking the mean and standard deviation of the difference between AMV estimates and true winds in each cluster, we develop error characteristics for each cluster (Figure 11); these quantities are precisely the bias and uncertainty that we require for the cost function J in Eq (1). We see that the unskilled clusters have very high standard errors and they correspond roughly to the areas of unskilled regimes in Figure 3. Since each cluster now has associated error characteristics (e.g., bias and standard deviation), it is then straightforward to assign the bias and uncertainty for any new tracked wind observation by computing which regime it is likely to belong to.

3.7 Experimental Set up

In this section we will describe our experimental setup for training the data and testing its performance on a withheld dataset. We divide the dataset into two parts: a training set consisting of the first 1.5 months of the GEOS-5 Nature Run, and a testing set consisting of the last 0.5 month of the Nature Run. Our training/testing procedure for the simulation data and tracked wind is as follows:

1. Divide the simulation data and tracked wind into two sets: training set of 1,000,000 points from the first 1.5 months of the Nature Run and a testing set of 1,000,000 points from the final 0.5 months of the Nature Run.
2. Using the 'density.Mclust' function, we train a Gaussian Mixture Model on a normalized random sample of observations from the training dataset of true winds (u and v direction), tracked winds (u and v direction), and water vapor with n=9 clusters.
3. We train two separate random forests on a different random sample of 750,000 observations from the training dataset. We use tracked wind (u and v direction) and water vapor to model, separately, true winds in both the u and v directions.
4. We apply the random forests to the dataset used for the Gaussian Mixture Model. This provides a random forest estimate for each point, which is used as a substitute for true wind values in the next step.
5. Using the 'predict.Mclust' function, we predict the Gaussian mixture component assignment for each point of water vapor, tracked winds, and random forest estimate.

https://doi.org/10.5194/amt-2020-95
Preprint. Discussion started: 23 April 2020

Atmospheric Measurement Techniques Discussions

6. We compute the mean and standard deviation of the difference between the tracked winds and the true winds, per direction, for each Gaussian mixture model cluster assignment. This provides a set of error characteristics that are specific to each cluster.
7. We can apply the random forest, and then the cluster estimation, to any set of water vapor and tracked AMV estimates. Thusly, any set of tracked AMV estimates and water vapor can be mapped to a specific cluster, and therefore its associated error characteristics.

4 Results and Validation

In this section, we compare our clustering method against a simple alternative, and we quantitatively demonstrate improvements that result from our error characterization. Recall that in Section 3, we divided the wind-tracking outputs into 9 regimes, which range from very skilled to unskilled. For each regime, we can quantify the uncertainty via a 95% confidence interval, which in the Gaussian case can easily be constructed as $[x_i - 2\,\sigma_i, x_i + 2\,\sigma_i]$, where x_i the predicted mean and σ_i is the predicted standard deviation of the i-th cluster. To test the bias of our confidence interval, we divide the dataset described in Section 2 into a training dataset (first 1.5 month) and a testing dataset (last 0.5 month). Having trained our model using the training dataset, we apply the methodology to the testing dataset, and we compare the performance of the predicted confidence intervals against the actual wind error (tracked winds - true winds). This is a type of probabilistic forecast assessment, and in this paper we assess the quality of the prediction using a scoring rule called continuous ranked probability score, which is defined as a function of a probabilistic forecast F (here represented by our confidence interval) and an observation x as follows:

$$\mathrm{CPRS}(\mathrm{F}, \mathrm{x}) = \int_{-\infty}^{\infty} (\mathrm{F}(\mathbf{x}) - \mathbb{1}(\mathbf{y}-\mathbf{x}))^2\, \mathbf{dy} \qquad (4)$$

Where $\mathbb{1}(\,)$ is the Heaviside step function and denotes a step function along the real line that is equal to 1 if the argument is positive or zero, and it is equal zero if the argument is negative. The continuous rank probability score here is strictly proper, which means that the function $\mathrm{CPRS}(\mathrm{F}, \mathrm{x})$ attains the maximum if the data x is drawn from the same probability distribution as the probabilistic forecast F. That is, if the data x is drawn from F, then $\mathrm{CRPS}(F, x) \leq \mathrm{CRPS}(G, x)$ for all $G \neq F$.

The alternative error characterization method that we test against is a simple marginal mean and marginal standard deviation of the entire track - true wind datasets. This is essentially equivalent to an error characterization scheme that utilizes one regime, and its confidence interval similarly could be constructed as $[x - 2\,\sigma, x + 2\,\sigma]$, where x and σ are the marginal mean and marginal standard deviation of the residuals (i.e., tracked wind minus true winds). Here, we use a negatively oriented version of the CRPS (i.e., Eq.(4) without the minus sign), which implies that lower is better, to evaluate the performance of our methodology against the naive error characterization method. We plot the histogram of the scores in Figure 12.

https://doi.org/10.5194/amt-2020-95
Preprint. Discussion started: 23 April 2020

The relative behavior of the CRPS is consistent between u and v winds. The CRPS tends to have to wider distribution when applied to the regime-based error characterization. Compared to the alternative error characterization scheme, our methodology produces a cluster of highly accurate predictions (low CRPS scores), in addition to some cluster of very uninformative predictions (high CRPS scores). These clusters likely correspond to the highly skilled cluster (e.g., Cluster 3) and the unskilled clusters (Cluster 5 and 8), respectively. Overall, the mean of the CRPS is lower for our methodology than it is for the alternative method, indicating that as a whole our method produces a more accurate probabilistic forecast.

Thus far we have shown that our method produces more accurate error-characterization than an alternative method based on marginal means and variance. Now, we assess whether our methodology provides valid probabilistic prediction; that is, we test whether the uncertainty estimates provided are consistent with the empirical distribution of the validation data. To assess this, we construct a metric in which we normalize the difference between the true wind and the tracked wind by the predicted variance. That is, we compute the normalized values for u and v using the following equations:

$$z_u = \frac{u - \hat{u}}{\sigma_u}$$

$$z_v = \frac{v - \hat{v}}{\sigma_v} \tag{5}$$

Where u is the true u wind from the Nature Run data, $\hat{u}$ is the tracked-wind, and σ_u is the error as assessed by our model. The values for the v-wind are defined similarly. The residuals in Eq (5) can be considered as a variant of the z-score, and it is straightforward to see that if our error estimates are valid (i.e., accurate), then the normalized residuals in Eq. (5) should have a standard deviation of 1. In Figure 12, we display the histogram of the normalized residuals z_u and z_v. It is clear that for both types of wind, our error characterization methodology produces highly accurate uncertainties (std = 1.003 and 1.009 for u and v, respectively).

5 Conclusion

Uncertainty quantification, which is the quantification of an imperfect or incomplete state of knowledge within a model, is an important component of data validation and scientific analysis. For wind-tracking algorithms, whose outputs (tracked u and v) are often used as observations in data assimilation analyses, it is necessary to accurately characterize the bias and standard error (e.g., see Section 2.2). Nguyen et al. (2019) illustrated that incorrect specification of these uncertainties (a and R in Eq. (1)) can adversely affect the assimilation results – mischaracterization of bias will assimilate an incorrect tracked wind, while an erroneous standard error could incorrectly weight the cost function.

In this paper, we develop an error-characterization scheme based on random forest and mixture model clustering. Here, the mixture of a parametric approach and a machine learning method allows us to combine the flexibility of machine learning with the interpretability of mixture modelling in an entirely data-driven framework. In theory, the fidelity of our method should scale with the number of training data observations, making the methodology well-suited for the massive datasets that are typical within remote sensing applications. Our error function has been applied to an AMV OSSE study using GEOS5 and its impact will be reported in a forthcoming paper.

We demonstrate that our methodology produces accurate error estimates (also called validity), and that it is able to identify and remove the biases within the wind-tracking algorithm's outputs. Particularly, the methodology is able to identify unskilled regimes that are physically meaningful — in our case, unskilled regimes related to regions of near-zero water vapor content. We note that our methodology is able to find this dependence between unskilled regimes and low water content without any prior knowledge or specification from the user, deducing the relationship from the underlying multivariate distribution of water vapor, true wind, and tracked wind. While we position the methodology as an error characterization tool, this property also makes it useful as an exploratory tool to aid in understanding the distribution of multivariate and potentially complex data.

Our algorithm consists of two parts: an emulator and a clustering algorithm. In this implementation, random forest and Gaussian mixture modelling are the approaches; in theory, these two steps could be accomplished using other algorithms belonging to the appropriate class. Future research includes replacing random forest with other machine learning methods such as neural networks or support vector machines, and investigating other methods of clustering, such as self-organizing networks. We note that the issue of bias removal in data assimilation and in remote sensing is certainly not limited to atmospheric motion vectors. The methods we have used to characterize uncertainties in AMVs are general, and can be applied to other inverse problems as well.

Author Contribution

Teixeira and Nguyen conceived of the idea. Teixeira performed the computation. Wu provided the experimental datasets. Posselt and Su provided subject matter expertise. All discussed the results. Teixeira wrote the manuscript. All authors contributed to the subsequent draft.

Competing Interest: The Authors declare no conflict of interest.

Funding Acknowledgment: The research was carried out at the Jet Propulsion Laboratory, California Institute of Technology, under a contract with the National Aeronautics and Space Administration (80NM0018D0004).

Atmospheric Measurement Techniques Discussions

References

Bormann, N., Hernandez-Carrascal, A., Borde, R., Lutz, H.J., Otkin, J.A. and Wanzong, S.: Atmospheric motion vectors from model simulations. Part I: Methods and characterization as single-level estimates of wind, Journal of Applied Meteorology and Climatology, 53(1), 47-64. https://doi.org/10.1175/JAMC-D-12-0336.1, 2014.

Breiman, L.: Random forests. Machine learning, 45(1), 5-32, 2001.

Cassola, F. and Burlando, M.: Wind speed and wind energy forecast through Kalman filtering of Numerical Weather Prediction model output, Applied Energy, 99, 154-166, 2012.

Fraley, C. and Raftery, A.E.: MCLUST: Software for model-based clustering, density estimation and discriminant analysis (No. TR-415). Washington University, Seattle Department of Statistics, 2002.

Fraley, C., Raftery, A.E., Murphy, T.B. and Scrucca, L: mclust version 4 for R: normal mixture modeling for model-based clustering, classification, and density estimation, Washington University, Seattle Department of Statistics, 2012

Hernandez-Carrascal, A. and Bormann, N.: Atmospheric motion vectors from model simulations. Part II: Interpretation as spatial and vertical averages of wind and role of clouds, Journal of Applied Meteorology and Climatology, 53(1), 65-82, 2014.

Holmlund, K., Velden, C. S., & Rohn, M.: Enhanced automated quality control applied to high-density satellite-derived winds, Monthly Weather Review, 129(3), 517-529, 2001.

Kawa, S.R., Erickson, D.J., Pawson, S. and Zhu, Z.: Global CO2 transport simulations using meteorological data from the NASA data assimilation system, Journal of Geophysical Research: Atmospheres, 109, https://doi.org/10.1029/2004JD004554, 2004.

Le Marshall, J., Rea, A., Leslie, L., Seecamp, R., & Dunn, M.: Error characterisation of atmospheric motion vectors, Australian Meteorological Magazine, 53(2), 2004.

Mueller, K.J., Wu, D.L., Horváth, Á., Jovanovic, V.M., Muller, J.P., Di Girolamo, L., Garay, M.J., Diner, D.J., Moroney, C.M. and Wanzong, S.: Assessment of MISR cloud motion vectors (CMVs) relative to GOES and MODIS atmospheric motion vectors (AMVs), Journal of Applied Meteorology and Climatology, 56(3), 555-572, https://doi.org/10.1175/JAMC-D-16-0112.1, 2017.

Nguyen, H., Cressie, N., and Hobbs, J.: Optimal Estimation retrievals: Implications and consequences when the prior's mean and covariance are misspecified, National Institute for Applied Statistics Research Australia Working Paper Series, 2019.

Posselt, D. J., L. Wu, K. Mueller, L. Huang, F. W. Irion, S. Brown, H. Su, D. , and C. S. Velden: Quantitative Assessment of State-Dependent Atmospheric Motion Vector Uncertainties. J. Appl. Meteor. Clim., In Press. https://doi.org/10.1175/JAMC-D-19-0166.1., 2019.

Putman, W., A.M. da Silva, L.E. Ott and A. Darmenov: Model Configuration for the 7-km GEOS-5 Nature Run, Ganymed Release (Non-hydrostatic 7 km Global Mesoscale Simulation). GMAO Office Note No.5 (Version 1.0), 18, 2014.

Salonen, K., J. Cotton, N. Bormann, and M. Forsythe: Characterizing AMV Height-Assignment Error by Comparing Best-Fit Pressure Statistics from the Met Office and ECMWF Data Assimilation Systems, J. Appl. Meteor. Climatol., 54, 225–242, https://doi.org/10.1175/JAMC-D-14-0025.1, 2015.

https://doi.org/10.5194/amt-2020-95
Preprint. Discussion started: 23 April 2020

Atmospheric Measurement Techniques
Discussions

Staffell, I. and Pfenninger, S.: Using bias-corrected reanalysis to simulate current and future wind power output, Energy, 114,1224-1239, 2016.

Swail, V.R. and Cox, A.T.: On the use of NCEP–NCAR reanalysis surface marine wind fields for a long-term North Atlantic wave hindcast, Journal of Atmospheric and oceanic technology, 17(4), 532-545, 2000.

Velden, C.S. and K.M. Bedka,: Identifying the Uncertainty in Determining Satellite-Derived Atmospheric Motion Vector Height Attribution. J. Appl. Meteor. Climatol., 48, 450–463, https://doi.org/10.1175/2008JAMC1957.1_2009.

Zeng, X., S. Ackerman, R.D. Ferraro, T.J. Lee, J.J. Murray, S. Pawson, C. Reynolds, and J. Teixeira: Challenges and opportunities in NASA weather research. Bull. Amer. Meteor. Soc., 97, 137–140, 2016.

https://doi.org/10.5194/amt-2020-95
Preprint. Discussion started: 23 April 2020

Atmospheric Measurement Techniques
Discussions

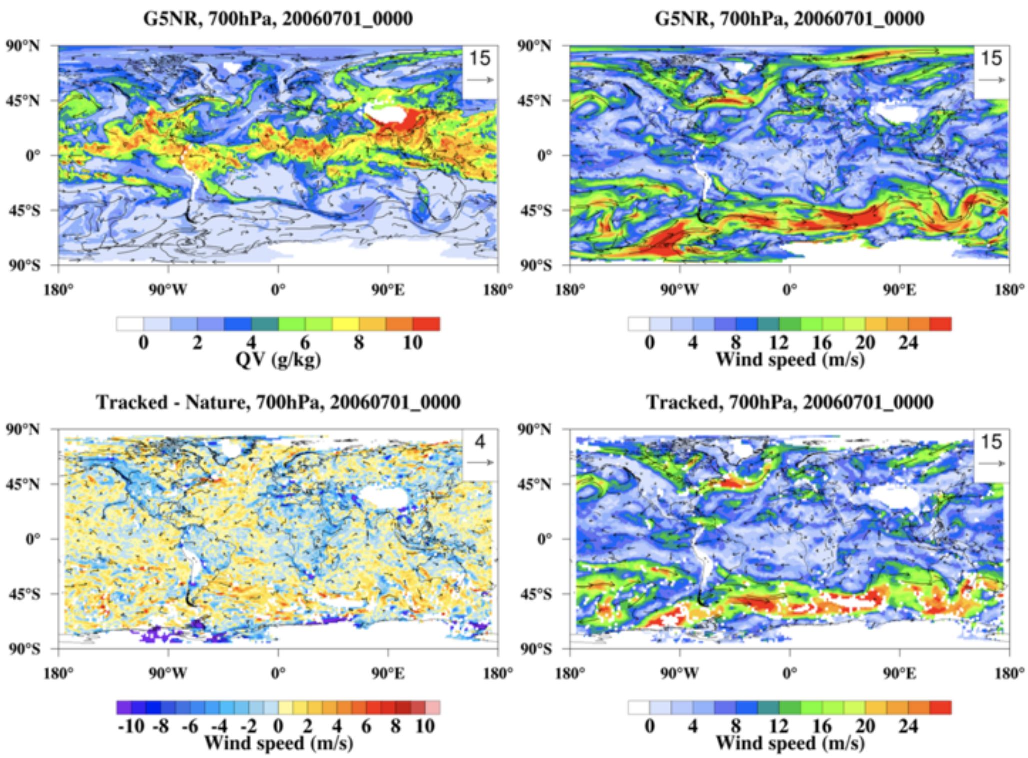
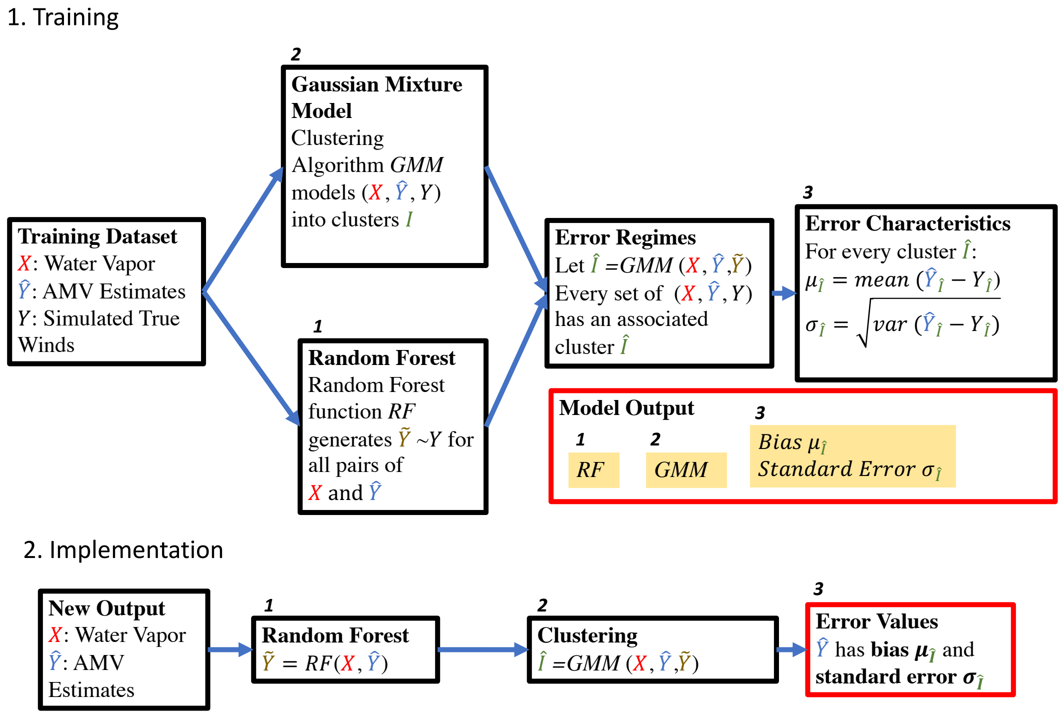


Figure 1: Map of Nature Run at one timestep at 700hPa (A): Water Vapor (B): True Wind Speed (C): Difference between True Wind Speed and AMV Estimate (D): AMV Estimate.

https://doi.org/10.5194/amt-2020-95
Preprint. Discussion started: 23 April 2020

Atmospheric Measurement Techniques Discussions

403



404

405 **Figure 2: Diagram of Training Approach and Diagram of Implementation steps.**

https://doi.org/10.5194/amt-2020-95
Preprint. Discussion started: 23 April 2020

Atmospheric Measurement Techniques
Discussions

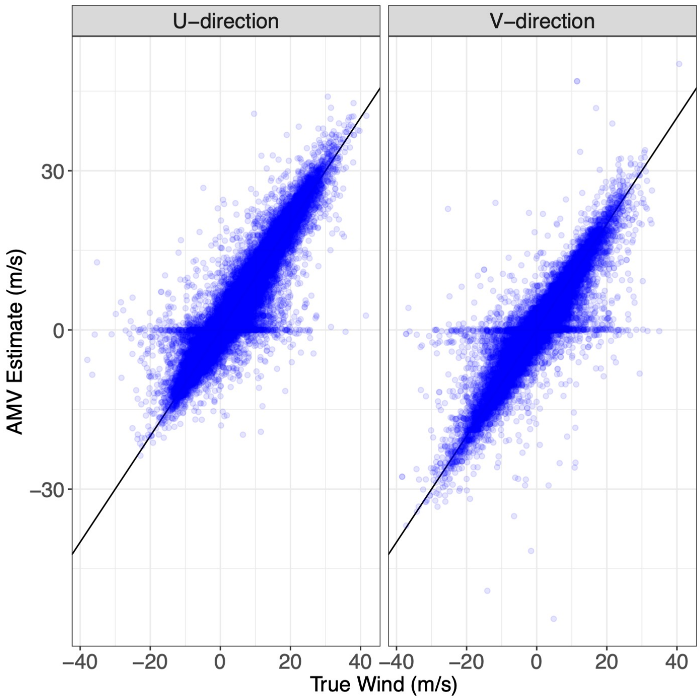


Figure 3: Scatter plot of the simulated true wind vs AMV estimates for u and v wind.

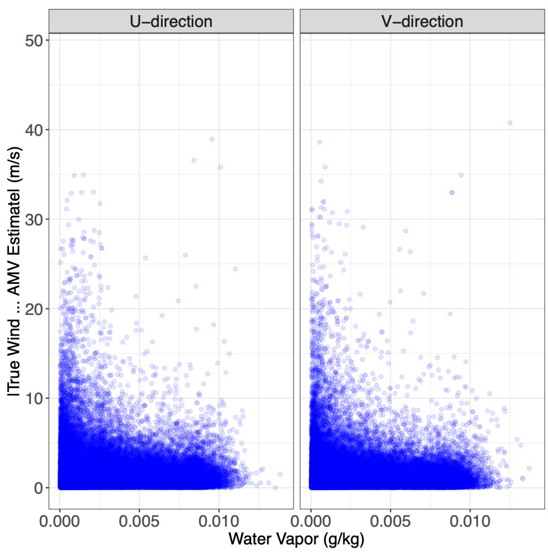


Figure 4: Simulated water vapor vs the absolute value of the difference between true and tracked winds.

https://doi.org/10.5194/amt-2020-95
Preprint. Discussion started: 23 April 2020

Atmospheric Measurement Techniques Discussions

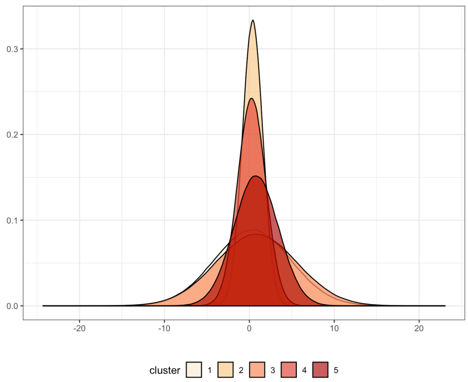


Figure 5: Example of Gaussian Mixture Model in one dimension. Density Figures for the U-Direction AMV Estimate dimension of fitted Gaussian mixture.

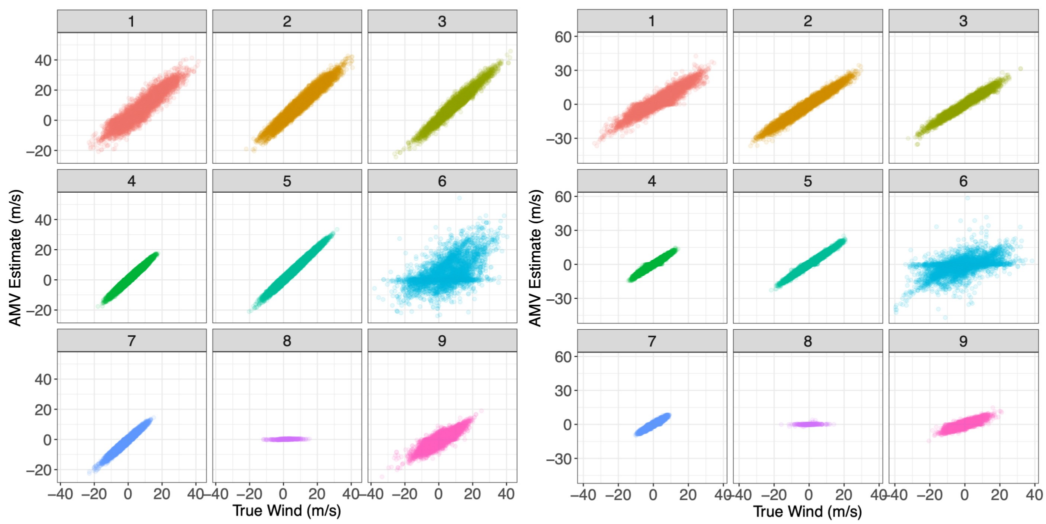


Figure 6: Scatterplot of simulated true wind vs AMV Estimates, each sub-panel corresponding to the specific Gaussian mixture component to which each point has been assigned. (A): U-Direction Wind (B): V-Direction Wind.

https://doi.org/10.5194/amt-2020-95
Preprint. Discussion started: 23 April 2020

Atmospheric Measurement Techniques
Discussions
Open Access
EGU

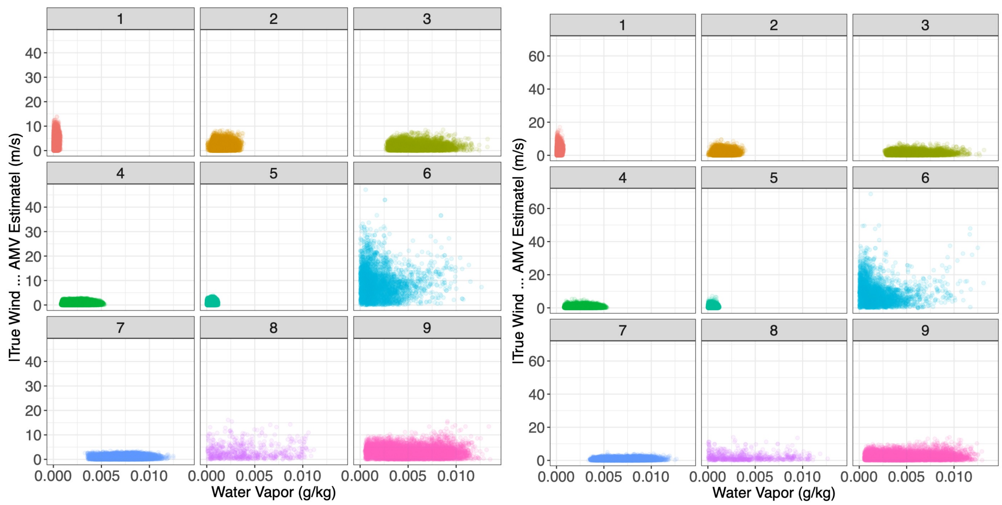


Figure 7: Scatterplot of Water Vapor vs Absolute Tracked Wind Error, each sub-panel corresponding to the specific Gaussian mixture component to which each point has been assigned. (A): U-Direction Wind (B): V-Direction Wind.

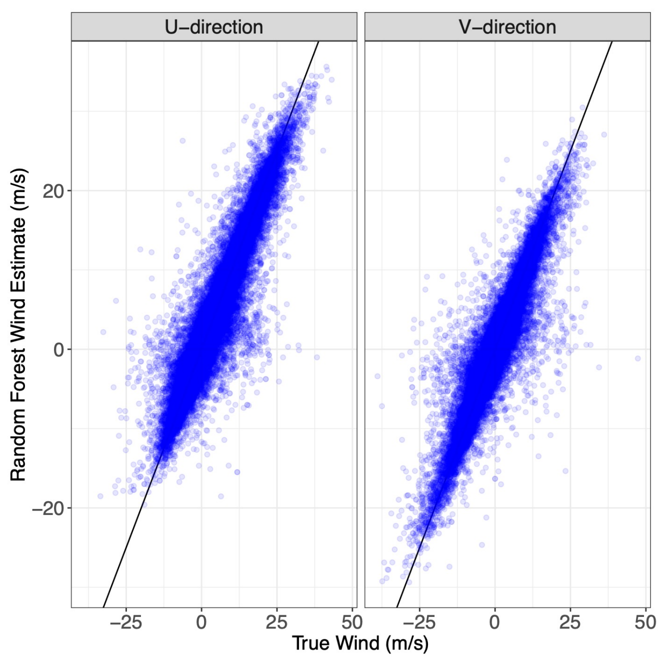


Figure 8: Scatterplot of true wind estimate vs random forest produced estimate. (A): U Direction (B): V Direction

https://doi.org/10.5194/amt-2020-95
Preprint. Discussion started: 23 April 2020

Atmospheric Measurement Techniques
Discussions

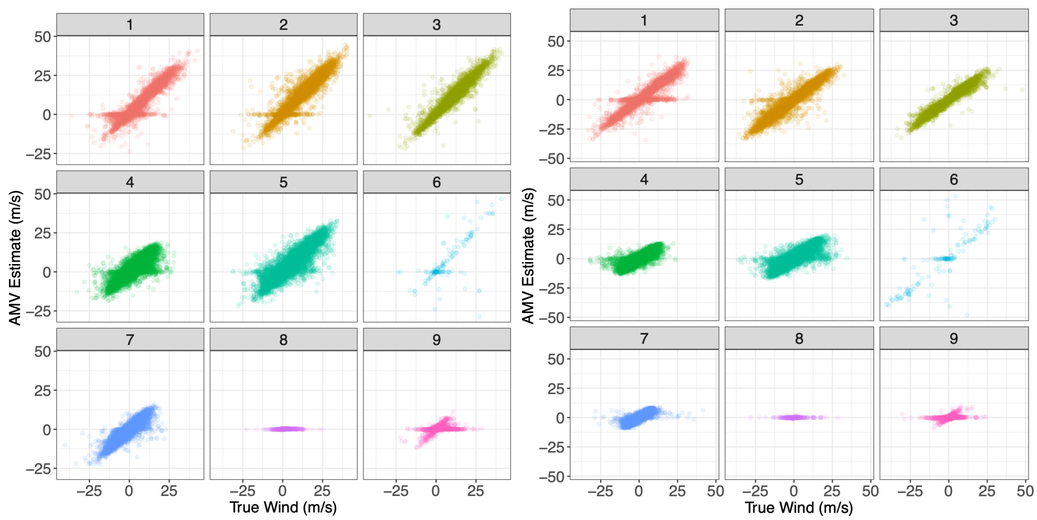


Figure 9: Scatterplot of simulated true wind vs AMV Estimates, each sub-panel corresponding to the specific Gaussian mixture component to which each point has been assigned when the true wind value has been substituted by the random estimate. (A): U-Direction Wind (B): V-Direction Wind

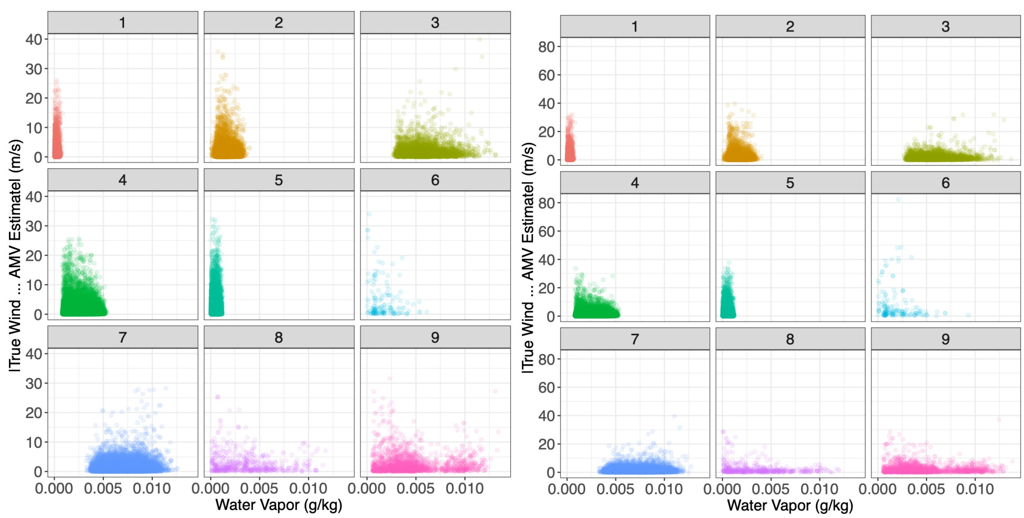


Figure 10: Water Vapor vs Absolute Tracked Wind Error, each sub-panel corresponding to the specific Gaussian mixture component each point has been assigned when the true wind value has been substituted by the random estimate. (A): U-Direction Wind (B): V-Direction Wind

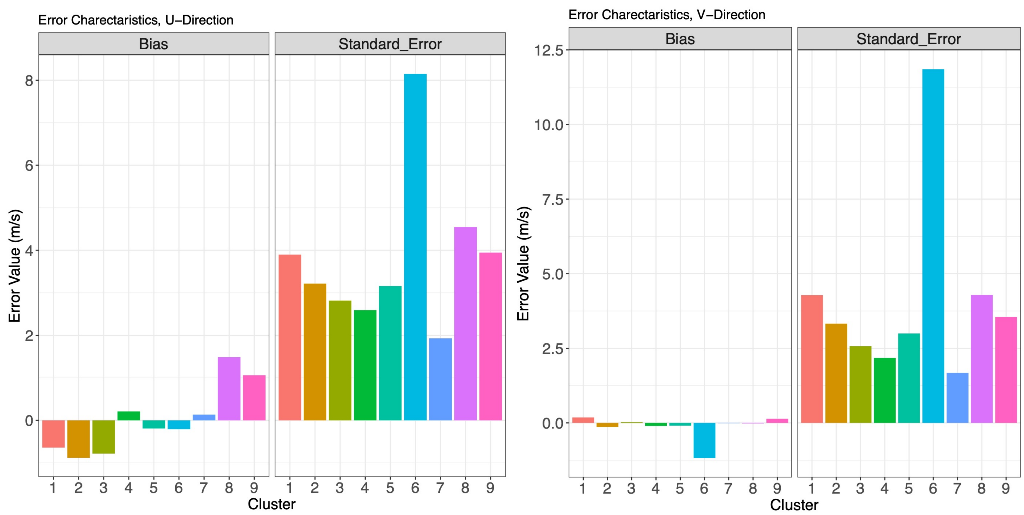


Figure 11: (A): Bias (Left Panel) and Standard Error (Right Panel) for each Gaussian mixture cluster in figure 6, U direction. (B): Same as (A) for V-direction

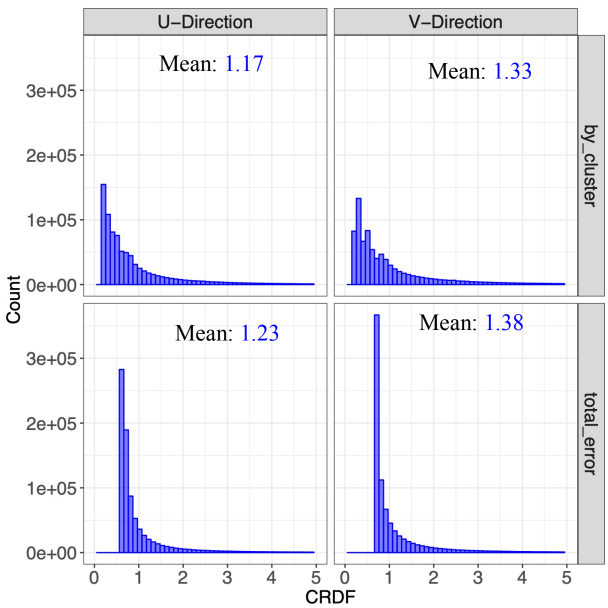


Figure 12: CRSP applied to different error approaches. (A): Cluster Errors for U Winds (B): Total Errors for U Winds (C): Cluster Errors for V Winds (D): Total Errors for V Winds.

https://doi.org/10.5194/amt-2020-95
Preprint. Discussion started: 23 April 2020

Atmospheric Measurement Techniques
Discussions
Open Access

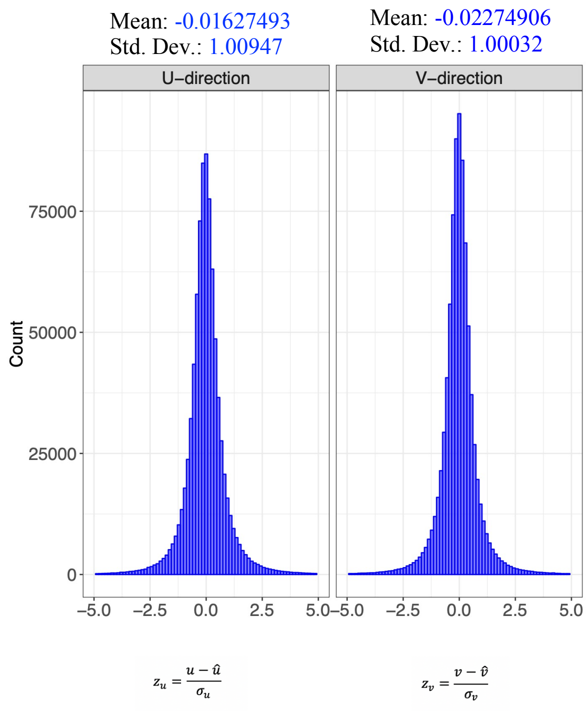


Figure 13: U and V winds normalized using Error Clusters