SeaWinds on QuikSCAT

Special Wind Vector Data Product: Direction Interval Retrieval with Thresholded Nudging (DIRTH)

Product Description

Version 1.1

Bryan W. Stiles

Jet Propulsion Laboratory

September 13, 1999

Contents

1 Algorithm Overview

The SeaWinds on QuikSCAT scatterometer (QSCAT) was developed by NASA JPL to measure the speed and direction of ocean surface winds. End-to-end simulations performed to estimate the performance of the QSCAT prior to its launch indicated that the directional accuracy of the wind vectors varies across the swath. Post-launch comparisons between scatterometer data and analytical wind fields (e.g., NCEP and ECMWF fields) support this conclusion, as does visual inspection of the scatterometer wind fields. The accuracy of the majority of the swath, and the size of the swath are such that QSCAT meets its science requirements despite shortcomings at certain cross track positions. Nonetheless, it is desirable to modify the baseline processing in order to improve the quality of the less accurate portions of the swath, in particular near the far swath and nadir [1]. Two disparate problems have been identified for far swath and nadir. At far swath, ambiguity removal skill is degraded due to the absence of inner beam measurements, limited azimuth diversity, and boundary effects. Near nadir, due to nonoptimal measurement geometry, (fore and aft looking measurement azimuths approximately 180^o apart) there is a marked decrease in directional accuracy even when ambiguity removal works correctly. Two algorithms were developed, direction interval retrieval (DIR) to address the nadir performance issue, and thresholded nudging (TN) to improve ambiguity removal at far swath. The two algorithms work independently, and need not be used together. However, both were used to obtain this special wind vector product.

DIR is a set theoretic estimation technique [1]. It is similar to the conventional (NSCAT) wind retrieval technique in that first a set of wind vectors are determined which are consistent with the data (*solution set*), then median filtering is used (spatial information incorporated) to select a solution vector from this set. DIR differs from the conventional method in that the solution set is not a finite set of vectors, but rather a set of disjoint 1-D curves in the 2-D space of wind speed and direction. The range of wind direction spanned by each of these curves is determined by a probabilistic analysis of the noise on the measurements and its effect on the directional discrimination information available. (See section 2.)

TN is a technique for optimizing the manner in which the ambiguity removal is initialized. In the baseline wind retrieval algorithm, the closest of the two most likely ambiguities to a co-located numerical weather product (NWP) wind vector is used to initialize the median filter. With TN, the number of ambiguities available for initialization is not limited

[1] In this context, nadir is taken to mean along the ground track of the satellite. Clearly the antenna is never actually pointed perpendicular to the ground.

to two, instead it is determined by thresholding the likelihood values associated with the ambiguities. In this manner, fewer ambiguities are considered in regions of high instrument skill, and thus the impact of the NWP field is lessened. On the other hand, in regions of lower instrument skill, more ambiguities are considered and the impact of the NWP field is heightened. (See section 3.)

The impact of the two techniques was studied in simulation and found to significantly improve wind direction accuracy. Improvements in RMS direction error were observed across the entire range of swath positions and wind speeds. Improvements as large as 10 degrees were obtained for low wind speeds and cross track positions near nadir. After launch similar studies were performed on real data, using ECMWF wind fields as truth, and similar results were obtained.

2 Direction Interval Retrieval

In order to discuss the DIR technique, some background information about the baseline wind retrieval algorithm is required. The baseline algorithm is composed of two parts: a pointwise maximum likelihood estimator to calculate a set of likely wind vectors and a median filter to select the best vector from the set. The maximum likelihood step has been shown to be insufficient to choose a unique wind vector [2]. For a small set of measurement azimuth angles, multiple wind vectors may yield the same set of σ_0 values. Even if there are enough measurements from enough different azimuth angles to preclude this possibility, the addition of noise can still lead to multiple solutions of significant likelihood. For this reason, the wind retrieval algortihm was designed to produce a discrete set of feasible solutions rather than a single solution. The solution set is the set of local maxima in the likelihood function. For NSCAT, this solution set resulted in acceptable directional accuracy. The likelihood function dropped off quickly in the neighborhood of the local maxima, so that the chance of the true wind vector being far away from every vector in the solution set was small. For QuikSCAT the rate at which the likelihood value drops off from the maxima varies with cross track distance. For wind vector cells near nadir, there are large ranges of direction over which the likelihood value is relatively similar, and it is inaccurate to represent the set of likely wind vectors by the maxima alone. The DIR method addresses this problem by calculating a solution set for each wind vector cell which includes a range of wind directions around each likelihood maxima. The extent of the ranges is determined independently for each wind vector cell according to the specific shape of the likelihood function for that cell.

The DIR technique is a set theoretic estimation technique [1] which incorporates information from the σ_0 measurements and a model of the noise on those measurements in order to construct the solution set. Allowing the technique to consider all possible sets of wind vectors would be time prohibitive, so a simplifying assumption must be made regarding the types of sets to be considered. For each wind direction ϕ there is a wind speed $u(\phi)$ which maximizes the likelihood function. We refer to the curve thus defined as the best speed ridge. In the baseline technique, solution sets are four or fewer points on the best speed ridge corresponding to local likelihood maxima. In DIR, solution sets are generalized to four or fewer segments of the best speed ridge, with each segment including a local maxima. This choice of solution set is justified by the observations that likelihood drops off sharply for speeds away from the best speed ridge, and that whenever the wind direction is determined accurately the wind speed is as well.

The endpoints of the segments are determined by estimating error bounds in a manner similar to techniques described in [3] and [4]. These techniques estimate probability distributions (and confidence intervals) for each measurement then combine information by intersecting solution sets derived from confidence intervals on each measurement. The DIR technique instead estimates a joint probability distribution for all the measurements and then directly computes the solution set, yielding a more accurate result. This technique is seldom employed due to computational efficiency concerns, but since most of the information needed for the calculation is already available from the maximum likelihood estimator and the search space is limited to one dimension (by the best speed ridge assumption) efficiency is not a problem.

We assume the noise on the measurements is mutually independent and Gaussian. The means and variance of the Gaussian noise used in the maximum likelihood estimator can be used to compute the conditional probability density of obtaining the σ_0 measurements given a wind vector represented by speed and direction (u, ϕ), $P(\{\sigma_{0i}\}|u, \phi)$. In fact the conditional probability is related to the likelihood estimate $f(u, \phi)$ by:

$$P(\{\sigma_{0i}\}|u, \phi) = k \exp(f(u, \phi)/2) \tag{1}$$

for some constant k. However, since the the purpose of wind retrieval is to find the most likely wind vector for a given set of σ_0 values rather than vice versa, a more relevant probability density function is $P(u, \phi|\{\sigma_{0i}\})$, the probability density of wind vectors given an observed set of σ_0 values. This function when integrated over any region in wind vector space yields the probability that a wind vector within that region has occurred given the observed data. The two probability density functions are related by Bayes' Theorem,

$$P(u,\phi|\{\sigma_{0i}\}) = \frac{P(\{\sigma_{0i}\}|u,\phi)P(u,\phi)}{P(\{\sigma_{0i}\})} \tag{2}$$

where $P(u,\phi)$ is the a priori probability density of wind vectors and $P(\{\sigma_{0i}\})$ is the a priori probability density of σ_0 observations. For a given set of measurements, $P(\{\sigma_{0i}\})$ is a constant.

In order to restrict the solution space to the best speed ridge as discussed earlier we let $P(u,\phi) = 1/2\pi$ for (u,ϕ) on the best speed ridge and 0 everywhere else. This choice also assumes that there are no wind directions which are preferred a priori.

By combining equations 1 and 2 and limiting consideration to wind vectors on the best speed ridge we get

$$P(\phi|\{\sigma_{0i}\}) = c\exp(f(u(\phi),\phi)/2)$$

for which the constant c is chosen to satisfy the probabilistic identity

$$\int_0^{2\pi} P(\phi|\{\sigma_{0i}\})d\phi = 1.$$

Now that the estimation of the probability density function (pdf) has been obtained, the solution set segments are determined by thresholding the probability. Given a threshold T, a set of directional intervals around each of the local maxima is selected such that the sum of the widths of the intervals is minimized and the integral of the pdf over the intervals is T.

The choice of the threshold T is an important consideration. A value that is too low i.e., 0.1, results in an a solution set which is too small to sufficiently represent the uncertainty in the measurements. In such a case the DIR technique will not go far enough in reducing the near nadir directional error. In fact, the baseline technique is identical to DIR with $T = 0$. A value which is too high i.e., 0.95, overestimates the uncertainty in the measurements allowing the ambiguity removal step to oversmooth the data. In simulation, $T = 0.8$, the value used in producing this product, was found to be a reasonable value. Performance was found to be insensitive to small changes in T. The chose of threshold T deserves further study because the simulation studies and model field comparisons are insufficient to determine its impact on mesoscale phenomena. Depending on how well mesoscale phenomena are preserved in the current product, T may be decreased to reduce smoothing or increased to improve noise removal.

Once the solution set has been calculated for each wind vector. Ambiguity removal is performed to select a unique solution vector from each solution set. A two step procedure is employed. First one of the disjoint segments which composes each solution set is selected by performing ambiguity removal in the usual manner [2]. Ambiguity removal is performed on the local likelihood maximas and the segment which encloses the selected maxima is chosen. Next, a unique vector within the chosen segment is selected by iteratively choosing the vector which is closest in direction to the median vector of the surrounding 7 x 7 window.[3] Each wind vector cell is initialized by the maxima within the selected segment. Wind vectors are not updated until after each median filtering pass is complete. Passes continue until no wind vectors change by more than a threshold amount (5 degrees) or a maximum number of passes (100) is exceeded. The author is unaware of the maximum number of passes ever being exceeded, and typically the vast majority of the wind vectors are determined by the fourth pass.

3 Thresholded Nudging

The baseline nudging algorithm, which is the same as that used for NSCAT, chooses an ambiguity to initialize the median filter. Currently, that algorithm only allows one of the two most likely ambiguities to be chosen. The rationale for that limit is based on NSCAT experience, where we assume that the scatterometer can choose the correct streamline, and want the nudging field to select the proper ambiguity from that line. The other reason for limiting to two the number of ambiguities from which the nudging field can choose is to limit the influence of the nudging field, and to use as much scatterometer information as possible. If all ambiguities are allowed to be selected by the nudging field, the retrieved wind field would be very close to the nudged wind field, defeating the point of making the measurement.

The QSCAT situation is somewhat different from the NSCAT situation. In the outer swath, the scatterometer can not always select the correct streamline. A significant percentage of the time (10-15 percent in simulation) the ambiguity closest to the truth is the third or fourth ranked ambiguity. Given that situation, one method that suggests itself is to use more ambiguities for nudging in the outer swath.

The likelihood function can be converted into an estimate of probability. (see previous section) Using equation 1 we calculate *relative likelihood* a quantity proportional to

[2] with the exception that the median filter is initialized using thresholded nudging. See next section for more detail.

[3] The window size was chosen to correspond to the size used by the baseline median filtering algorithm. Additional window sizes deserve further study both for DIR and the standard algorithm.

$P(\{\sigma_{0i}\}|u, \phi)$ normalized so that the relative likelihood of the first ranked ambiguity is one. The method by which we set the maximum rank for nudging is based on choosing the number of ambiguities above a certain threshold, M in relative likelihood. The threshold itself should be a function of the quality of the nudge field. The value used in this product, $M = 0.2$, was found to be an acceptable value in simulation and has been verified somewhat post-launch by comparisons of SeaWinds data with ECMWF wind fields. Thresholded nudging with $M = 0.2$ was found to outperform a number of other schemes for improving ambiguity removal.

4 Data Format and Guidelines for Use

The format of the data is the same as that of the official L2B Wind Vector Data Product, with the following exception: The last ambiguity in each wind vector cell is the solution vector obtained by using DIR and TN, and the selected index is set to point to this ambiguity. For wind vector cells which originally had less than 4 ambiguities, an extra ambiguity is added (the number of ambiguities value is incremented) to hold the DIRTH solution vector. For wind vector cells which already had all four ambiguities, the fourth ambiguity is overwritten by the solution vector. The overwriting was done, because there are only four slots available for ambiguities in the L2B format. The special product was designed to match that format in order to avoid requiring users to obtain or write new data reading routines. If the DIR or TN techniques are later incorporated into the official product, a new data set will be created to store the solution vector, while maintaining the original ambiguities and selected index.

5 Points of Contact

Questions concerning data distribution should be directed to PO.DAAC. Issues related to data quality or processing should be directed to Bryan Stiles. Specific contact information is listed below. Please note that e-mail is always the preferred means of communication.
PO.DAAC: Data Distribution Issues

JPL PO.DAAC User Services Office
Jet Propulsion Laboratory
Mail Stop Raytheon-299
4800 Oak Grove Drive
Pasadena, CA 91109, U.S.A.

Telephone:	(626) 744-5508
FAX:	(626) 744-5506
e-mail:	`qscat@podaac.jpl.nasa.gov`
Home Page:	`http://podaac.jpl.nasa.gov/quikscat`
FTP:	`ftp://podaac.jpl.nasa.gov`

Technical and Algorithmic Issues; Corrections and Updates to this Product Description

Dr. Bryan W. Stiles	e-mail:	`bstiles@acid.jpl.nasa.gov`
Jet Propulsion Laboratory, MS 300-319	Telephone:	(818) 354-5329
4800 Oak Grove Drive	FAX:	(818) 393-5184
Pasadena, CA 91109, U.S.A.		

References

[1] Combettes, P.L., "The Foundations of Set Theoretic Estimation," *Proceedings of the IEEE,* Vol. 81, No. 2, pp 182-208, 1993.

[2] Long, D. G. and Mendel, J. M., "Identifiability in Wind Estimation From Scatterometer Measurements," *IEEE Trans. on Geoscience and Remote Sensing,* Vol. 29, No. 2, pp 268-276, 1991.

[3] Combettes, P.L. and Trussell, H. J., " The Use of Noise Properties in Set Theoretic Estimation," *IEEE Trans. on Signal Processing* Vol. 39, No. 7, pp 1630-1641, 1991.

[4] Walter, E., and Piet-Lahanier, H., " Guaranteed Linear and Nonlinear Parameter Estimation from Bounded-Error Data: a Survey," *IEEE International Symposium on Circuits and Systems,* Vol. 1, pp 774-777, 1993.