SeaSonde Current Vector Uncertainties
The Start of QA/QC

With hundreds of SeaSondes in worldwide operation, most within large networks, increasing emphasis is being paid to data Quality Assurance & Quality Control (QA/QC). How good are the real-time vectors? Can one define metrics that will flag time-changing error or ***uncertainty variations*** to use in diagnoses or applications? Small-scale current-pattern variations in time and space are generally not considered useful (in fact, noise), in terms of current patterns deemed meaningful for most applications. Flagging these for user consideration has always been our goal at CODAR. This is one part of a comprehensive QA/QC process. To this end, we have provided "uncertainties" outputted to radial and total vector files for well over a decade. Yet many attempt, with good intentions, to contrive QA/QC methods, with a goal of identifying and removing "bad vectors" rather than using these CODAR-provided uncertainties. Without solid physical or statistical foundation, this could do more harm than good. With this in mind, we ask the question: why not start with the uncertainties provided by CODAR? Here we explain what they are what they tell us, and what are their limitations.

SeaSonde Output Uncertainties Defined

There are no models nor assumptions involved in calculating our two uncertainties. They are based on fluctuations in the data themselves: i.e., standard deviations. This is why having a lot of data over time (rather than five minutes out of the hour) provides more substantial and meaningful estimates of these uncertainties.

Spatial Uncertainties in Radials. Radial velocities are calculated from the data on a polar grid. For example, every ten minutes, there are generally many estimates of radial velocity for every range cell and bearing grid cell (typically 5°). These are averaged to produce a mean radial velocity. In the same calculation, the standard deviation about this mean is calculated and outputted, for that grid cell. Why do we call this a "spatial uncertainty"? Because if there were no random fluctuations in the data, these radial vector estimates would all be the same, i.e., their standard deviation would be zero. The fact that they are not tells us that they actually came from different bearings, where they probably did not belong. Hence the term "spatial uncertainty". These are written to a temporary file (in our example case, every 10 minutes), and then averaged over the (typical) hour "merging period", after which a "Radial (or Elliptical) File" is created and stored/transmitted to a central site.

Temporal Uncertainties in Radials. This is the very first uncertainty that we created and outputted. It is the standard deviation of the short-term radials (e.g., every 10 minutes) for each polar (or elliptical) grid cell. Over an hour, then, there could potentially be 7 that go into estimating this standard deviation. Because the statistic is calculated over an hour's time, it is called a "temporal uncertainty".

Uncertainties in Total Vectors. With two or more sets of radials (and/or ellipticals) combined to get the u/v components of a total vector, the radial uncertainties are "propagated" mathematically to get u/v uncertainties also (as well as a covariance between u/v). The temporal uncertainties (not spatial) are used for this calculation, and outputted as columns in the total vector files. The mathematics of this propagation calculation is found in papers by Lipa in our Web library.

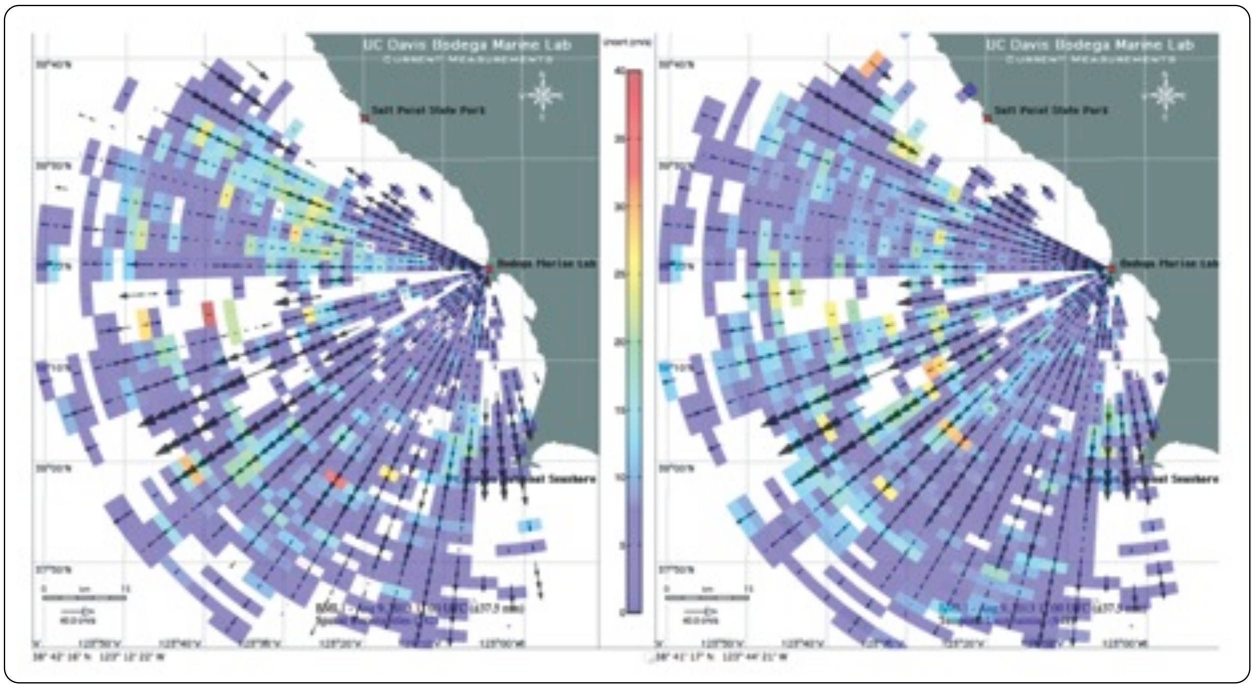
Are They Meaningful?

This is the first question a user should ask, because the fact that something is calculated and outputted does not guarantee it is meaningful or useful. The answer is yes, it is. A comprehensive study was done by an independent investigator, Kip Laws, in 2006 on the temporal uncertainty. Using many years of continuous data from Monterey Bay, he compared and presented at AGU the temporal uncertainties with baseline differences seen between two radars looking at the same spot from opposite sides of the bay. (With no fluctuations or errors, baseline differences should be zero, so this is a good metric for establishing a correlation with outputted uncertainties.) ***He found a coefficient of determination to be 92%, meaning a correlation coefficient of 96%. Yes indeed, it is meaningful.***

Sample CODAR Mapped Uncertainties: In addition to ASCII outputs, CODAR provides a simple, quick way to display and average uncertainties through its 'SeaDisplay' tool.

• **Example Radial Uncertainties Overlain on Vectors.** The figure below shows spatial (left) and temporal (right) uncertainties for a 13 MHz SeaSonde (from August 9, 2013) that has been operating at Bodega Marine Laboratories for 13 years. This is from an hourly merged Radial file. Many options in SeaDisplay for coloring and displaying uncertainties are available for viewer interpretation, both with and without vectors. Points to note here:

- The colorbar scale shows uncertainties for the most part fall below 5-10 cm/s.
- There are gaps in both the vectors and the uncertainties. Gaps are due to the fact that vectors were not found at that grid point. Calculation of uncertainties require (by definition) at least two points; if not a gap is shown.
- Patterns for the spatial (left) and temporal (right) uncertainties are generally similar.
- High uncertainties do not one-for-one always appear to be associated with wild vectors, and *vice versa*.



• **Example 2D Vector Uncertainties.** The figure on adjacent page shows temporal uncertainties for the same 13 MHz SeaSonde at the same time. To get these, radials from the Bodega Marine Lab site in the middle (shown earlier) were combined with simultaneous radials from the Southern site near Point Reyes (not shown here). Plotted at the left are the u (Eastward) uncertainties, with v (Northward) uncertainties at the right. Points to note:

- The colorbar scale has been changed, with a max now at 20 cm/s.
- Eastward (u) uncertainties (typically 2-3 cm/s) are lower than those Northward (v). The latter are greatest at the outermost coverage to the Northwest.
- Both vector and uncertainty maps for 2D vectors are completely filled in with no gaps, unlike radials. This is due to the broader averaging circle used for combination into totals.

What Is the Role of These Uncertainties in QA/QC?

Now we can make several points that put uncertainties from the data into perspective.

- **Can we weed out bad points using radial uncertainties?** Laws (cited earlier) has shown that -- taken in aggregate over long time periods -- uncertainties are a solid indicator of quality (or errors), with a correlation of 96%. Our plots above show that a high uncertainty does not always mean a bad vector, and *vice versa*. Thus it is dangerous to attempt to use a single radial uncertainty as an indicator that the vector at that grid point is good or bad. Why is this true? A couple examples:
  - Suppose a ship target got through. It would give a "wild" vector, because it would have been processed as though it were a current. But as a solid target, its variance (thence standard deviation) would yield a very low uncertainty.
  - Suppose two or three points went into creating a standard deviation, but one was a very large outlier. It would give a high standard deviation or uncertainty. But a merged radial vector created from the median of the three would be very reasonable.
  - Suppose second-order sea echo got mixed up with first-order (because of strong currents and/or high waves). This would give a very large "wild" vector. It would be great if it could be flagged and removed. But its uncertainty from its standard deviation could be quite low.

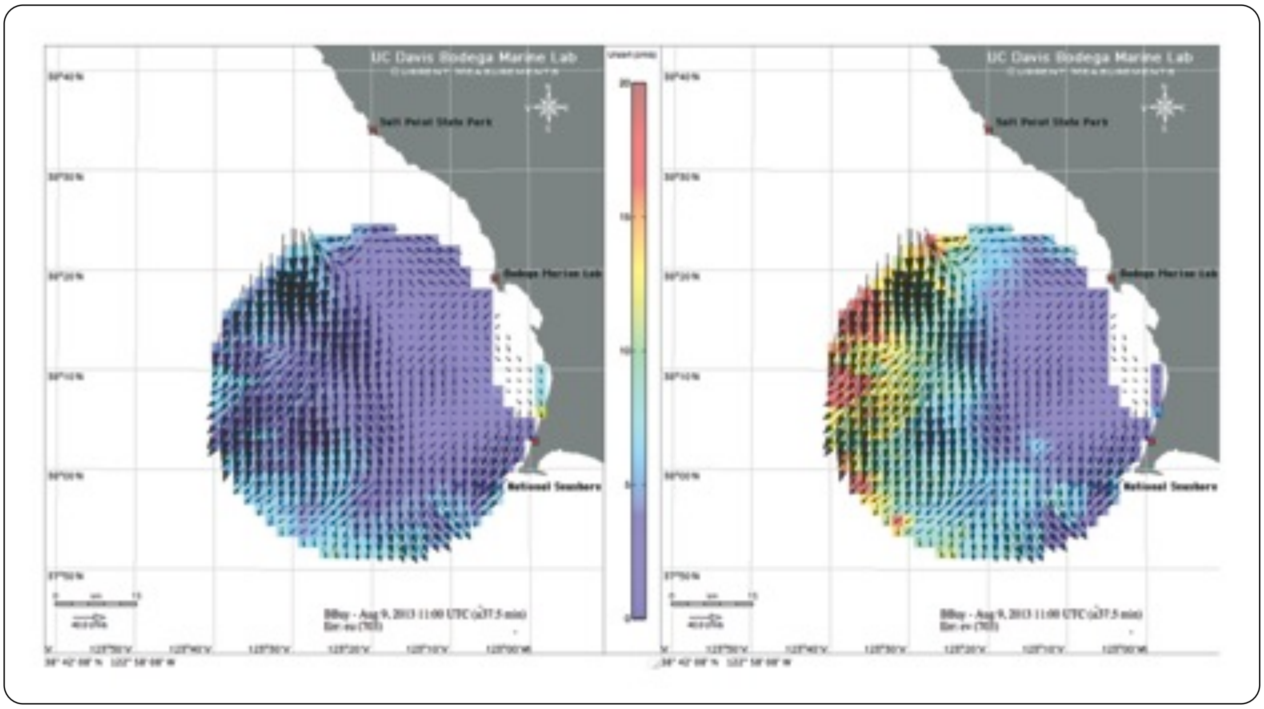
• **How about general quality vs. map position?** This is an area where the uncertainty maps can provide valuable insight. For example, the if the pattern of uncertainties found in the Bodega data above persists over time, it indicates that North/South(alongshore) accuracy is poorer than East/West (onshore), especially to the Northwest.

How Are They Useful in Circulation Models?

Numerical circulation models assimilate either radial or total vectors into them, in order to produce better forecasts; this improvement has been shown consistently by many investigators. These models need data covariances also. In the past, these have usually been ignored, or set to a constant because of lack of measured data. Our uncertainties (in both radials and totals) are actually the covariances needed by the models. Their value at any single space/time point is not always meaningful (as we have shown), but the models do not need this. They need an average over space and time, which is exactly the strength of the outputted SeaSonde uncertainties, as Laws and others have shown.

Role of SeaSonde Uncertainties in QA/QC: We have highlighted our uncertainties as a valuable resource that has been rarely used, in understanding QA/QC performance. They are easy to display and interpret with the SeaDisplay tool available to all clients. However, they are only one of several procedures and tools needed to ensure total QA/QC, because a bad uncertainty (or high possible error) does not tell the whole story. What else is needed?

• **Antenna Pattern Measurement.** Lack of proper antenna pattern measurement can bias output radial data. (Included with these patterns are amplitude and phase settings among the antenna element cables and receiver channels.) Such biases occur because the algorithms will misplace radial velocity vectors at the wrong bearings. The uncertainties described herein do not and cannot reflect the use of an improper antenna pattern and could be low, although vectors could be consistently wrong in certain regions.



• **Hardware and System Diagnostic Outputs.** If any hardware or system component fails or malfunctions, this will not necessarily show up as high uncertainties. However, CODAR has provided 10-minute updated diagnostic files on over 120 critical system parameters (e.g., temperatures, voltages, forward/reflected power at the transmit antenna, etc.). In addition, CODAR supplies a convenient tool -- 'DiagDisplay' --for plotting histories of any user-selectable combination of these diagnostics. These diagnostics should be routinely reviewed, as they offer QA/QC assistance that the uncertainties alone cannot provide.

• **Proper Attention to First/Second-Order Setup.** Site-specific settings are usually required in order to separate first-order echoes (used for currents) from second-order (used for waves). If this has been ignored or left at default settings, second-order echoes can mistakenly be processed, and will give totally erroneous radial vectors that are too large. As mentioned earlier, such "wild" vectors will not necessarily be associated with large uncertainties.