ASSIMILATION OF QUIKSCAT WIND OBSERVATIONS IN HIRLAM 3-D VAR

Frank Thomas Tveter, Norwegian Meteorological Institute

1 Introduction

This newsletter outlines the DNMI plan for the implementation of QuikScat scatterometer wind observations in HIRLAM 3-D Var. According to this plan, a recommendation for the method to be used in the assimilation of the ambiguous QuikScat observations in 3-D Var, should be ready by June this year. A new method that will be considered for this recommendation, is studied in some details in this newsletter.

2 The DNMI Project Plan

The following table shows the most important activities in the DNMI project for assimilation of QuikScat wind observations in HIRLAM 3-D Var.

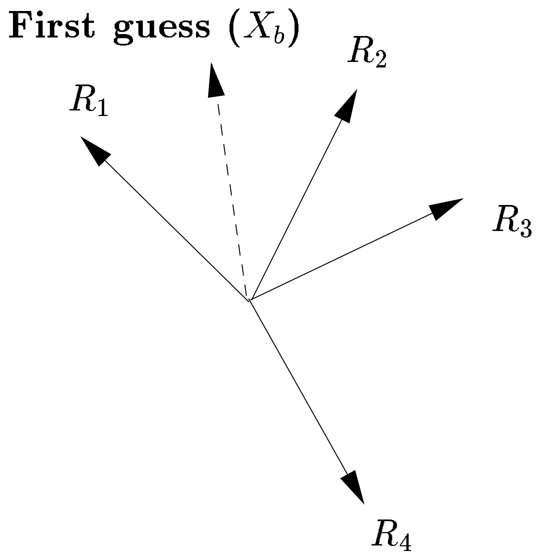
	2001		2002	
Activity	Spring	Autumn	Spring	Autumn
Research of methods	→			
Implementation in HIRLAM 3-D Var		→	→	
Impact studies				→

Note that the QuikScat observations will be implemented in HIRLAM 3-D Var by the middle of next year.

3 The current research at DNMI

The SeaWinds scatterometer on board the QuikScat satellite measures the radar backscatter from the sea surface. The backscatter properties of the sea surface depends on the surface wind. The backscatter measurements are transformed to a set of *possible* wind vector solutions, and it is these ambiguous solutions that we want to assimilate into HIRLAM 3-D Var.

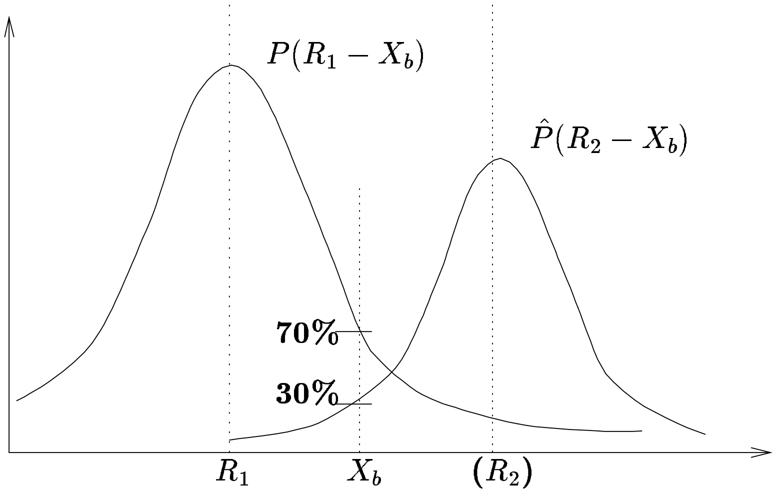
The following figure shows an example of a QuikScat observation with 4 wind vector ambiguities. A first guess wind vector is also indicated. The first guess is typically a prognosis from a weather prediction model. However, it could also be, say, the control variable in the 1st, 2nd, 3rd etc. iteration (note that the statistical properties of the deviation of the control variable from the truth change from one iteration to the next).



If we assimilated all ambiguities as independent observations, the analysis would move towards zero wind speed, since this is roughly the (least squares) average of one set of ambiguities. Assuming that the ambiguities are independent observations is therefore not a good strategy. We must make the system so that no more than one of the ambiguities in an observation influences on the analysis. The question is then: *Which ambiguity should we choose to use in the data assimilation?* A more general way of posing this question is: *Should we assimilate R_1?*

In this newsletter we simplify the problem to one dimension, and two ambiguities. Imagine that we have a large number of observations with exactly the same R_1 and R_2, and that we in each case know which ambiguity is the "correct" one. We make two distributions of the first guess, one for the cases where R_1 was the correct ambiguity, and one for the cases where

R_2 was the correct ambiguity. These may look something like this:



These density functions may be used together with the first guess to calculate the probability that a new observation has R_1 as the correct ambiguity. In the example above, for the X_b that is indicated, the most probable ambiguity is R_1 (70%). *But is "the most probable ambiguity" a sufficient criterium for assimilation purposes?* It is - if the average improvement in the analysis each time we are right is larger than the damage caused each time we are wrong (i.e. choose the wrong ambiguity). However, as we shall see later on, "being wrong" is more worse than "being right" is good.

Assimilating R_1 results in an "analysis", X_a, that is different from the first guess, X_b. On average, we want the analysis, X_a, to be closer to the "truth" than the first guess, X_b. Our next question is therefore: *what is the average effect of assimilating R_1?*

The analysis, X_a, is given by

$$X_a = X_b + K \cdot \gamma_1$$

where the innovation $\gamma_1 = R_1 - H \cdot X_b$, H is the (linear) forward model and K is the gain matrix.

The deviations from the unknown "truth", X_t, are then

$$\begin{aligned}\epsilon_b &= X_b - X_t \\ \epsilon_a &= X_a - X_t \\ \epsilon_o &= R_1 - H \cdot X_t = \gamma_1 + H \cdot \epsilon_b \\ \epsilon_a &= \epsilon_b + K \cdot \gamma_1 \ .\end{aligned}$$

Note that the analysis error, ϵ_a, and observation error, ϵ_o, can be written as a function of the unknown ϵ_b and known γ_1.

The difference between the analysis error and first guess error *on average* is given by

$$\delta r(\gamma_1) \quad = \quad \int_{-\infty}^{\infty} c' \cdot P_b(\epsilon_b) \cdot P_o(\gamma_1 + H \cdot \epsilon_b) \cdot \left(\epsilon_a^2 - \epsilon_b^2\right) d\epsilon_b$$

where c' is a normalization factor, P_b is the first guess error distribution, and P_o is the observation error distribution.

The term "$c' \cdot P_b(\epsilon_b) \cdot P_o(\gamma_1 + H \cdot \epsilon_b)$" gives the probability density of having a given ϵ_b and γ_1. The term "$(\epsilon_a^2 - \epsilon_b^2)$" represents the analysis improvement relative to the first guess for a given ϵ_b and γ_1. The integral appears since we average over the unknown first guess error, ϵ_b.

Note that a negative $\delta r(\gamma_1)$ indicates that the analysis is better than the first guess on average.

We assume that the first guess error distribution $p_b(\epsilon_b)$ is Gaussian.

In the cases where **$\mathbf{R_1}$ is the correct ambiguity**, assume that the observation error can be expressed as

$$P_o(\epsilon_o) \quad = \quad \frac{1}{\sqrt{2\pi R}} e^{-\frac{1}{2}\frac{\epsilon_o^2}{R}}$$

where R is the squared standard observation error.

If **$\mathbf{R_2}$ is the correct ambiguity**, then the observation error can be expressed as

$$\hat{P}_o(\epsilon_o) \quad = \quad \frac{1}{\sqrt{2\pi R}} e^{-\frac{1}{2}\frac{(\epsilon_o - \mu)^2}{R}}$$

where μ is the distance between the two ambiguities $(R_1 - R_2)$.

We can now estimate the average effect of assimilating $\mathbf{R_1}$.

When **$\mathbf{R_1}$ is the correct ambiguity**, we have

$$\delta r(\gamma_1) \quad \propto \quad -\gamma_1^2$$

The analysis is on average better than the first guess, since the right hand side (rhs) always is negative.

When **$\mathbf{R_2}$ is the correct ambiguity**, we get

$$\delta \hat{r}(\gamma_1) \quad \propto \quad -\gamma_1^2 + 2\gamma_1\mu = 3\gamma_1^2.$$

where we have substituted $\mu = 2\gamma_1$, since this is the case for our example.

The proportionality factor is the same for the two equations.

We observe that assimilating R_1 when in fact R_2 is the correct ambiguity, results in **3 times more damage** than the corresponding improvement in each case where R_1 is the correct ambiguity.

We must therefore be sure that the probability that R_1 is the correct ambiguity is **more than 75%** before we use this ambiguity in the assimilation. So in the example above, where the probability that R_1 is the correct ambiguity is only 70%, we should not use any of the ambiguities in the assimilation, since this on average would cause more damage to our analysis than improvement.

The next table shows the results from a simple experiment where the the theory above was used to select the ambiguity for a 1-D assimilation, X_a. The first guess, X_b, is based on a 6 to 12 hr forecast. Two references were used, the HIRLAM and ECMWF analysis. Observe that the QuikScat analysis is closer than the first guess to the two references for wind speed and wind direction.

	Wind dir (deg)		**Wind spd (m/s)**	
Difference type	**Bias**	**RMS**	**Bias**	**RMS**
X_b - X_a	0.127	5.919	-0.140	0.461
X_b - HIRLAM	-0.447	18.610	-0.123	1.450
X_a - HIRLAM	-0.610	**18.285**	0.018	**1.400**
X_b - ECMWF	-1.237	21.646	0.244	1.563
X_a - ECMWF	-1.376	**20.611**	0.384	**1.448**
HIRLAM - ECMWF	-0.568	19.327	0.367	1.502

DNMI research note no 53, "Variational Analysis and Contaminated Observations", discusses the theory above in more detail (contact f.t.tveter@dnmi.no for more information).

4 Conclusions

HIRLAM 3-D Var will be capable of assimilating QuikScat scatterometer observations by the middle of next year.

A recommendation for the method to be used in the assimilation of the ambiguous QuikScat observations in 3-D Var, will be ready by the end of June this year.