JP1.17 AN INTELLIGENT NN SYSTEM FOR QUALITY CONTROLLED RETRIEVING THE WIND SPEED FROM SSM/I MEASUREMENTS

Vladimir M. Krasnopolsky *
Science Applications International Corp. at the National Centers for Environmental Predictions

Abstract - The SSM/I application presented in this paper illustrates a NN based intelligent integral approach in satellite retrievals when the entire retrieval system, including the quality control block, is build as a combination of several specialized NNs. This approach offers significant advantages for operational applications. This intelligent retrieval system not only produces accurate retrievals, it also performs an analysis and quality control of these retrievals and environmental conditions, rejecting poor retrievals if they occur.

1. INTRODUCTION

Conventional methods for deriving geophysical parameters from satellite data (satellite retrievals) involve solving an inverse (or retrieval) problem and deriving a transfer function (TF), ***f***, which relates a geophysical parameter of interest, *G* (e.g. surface wind speed over the ocean, atmospheric moisture concentration, sea surface temperature, etc.), to a satellite measurement, *S* (e.g. brightness temperatures, radiances, reflection coefficients, etc),

$$G = \boldsymbol{f}\,(S) \qquad (1)$$

where both *G* and *S* may be vectors. The TF, ***f***, (it is also called a retrieval algorithm) can usually not be derived directly from the first principles because the relationship (1) does not correspond to a cause and effect principle. The relationship, however, can be written

$$S = \boldsymbol{F}\,(G) \qquad (2)$$

where ***F*** is a forward model (FM), which relates a vector *G* to a vector *S*. Forward models can usually be derived from physical considerations (e.g., radiative transfer theory) in accordance with the cause and effect principles because geophysical parameters affect the satellite measurements (not vice versa).

* *Corresponding author address:* Vladimir Krasnopolsky, SAIC at EMC/NCEP/NOAA, 5200 Auth Rd., Camp Spring, MD 20906, e-mail: Vladimir.Krasnopolsky@noaa.gov

NNs can be used to emulate FMs (2) and TFs (1) because FM and TF both are continuous mappings. There are many practical advantages (computational speed, accuracy, robustness) that can be achieved by using NNs for emulating FMs and TFs. A further advantage is the easiness and flexibility of incorporating into the NN any additional geophysical parameter known to influence the satellite measurements but not appearing in the original FM. This can improve the accuracy of the FM and can also help to regularize the inverse problem. To train NN, which emulates an explicit TF and/or FM, a training set, $\{G, S\}_{i=1,\ldots,N}$, ($S \in \boldsymbol{S}_T$) is required. Simulated or empirical data can be used to create the training set.

Well-constructed NNs have good interpolation properties; however, they may produce unpredictable outputs when forced to extrapolate. The NN training data (produced by a theoretical FM or constructed from empirical data collections) cover a certain manifold $\boldsymbol{S}_T$ (a subspace $\boldsymbol{S}_T \in \boldsymbol{S}$) in the full ***S*** space. Real data to be fed into the NN $\boldsymbol{f}_{NN}$, which emulates a TF (1), may not lie exactly in $\boldsymbol{S}_T$. There are many sources for such deviations of real data from the low dimensional manifold $\boldsymbol{S}_T$ of simulated data, e.g. simplifications made in the construction of the model, neglecting the natural variability of parameters occurring in the model and measurement errors in the satellite signal not taken into account during the generation of the training data. When empirical data are used, extreme events (highest and lowest values of geophysical parameters) are usually not sufficiently represented in the training set because they have a low frequency of occurrence in nature. That means that in the retrieval stage real data in some cases may force the NN $\boldsymbol{f}_{NN}$ to extrapolate. The error resulting from such forced extrapolation will increase with the distance of the input point from $\boldsymbol{S}_T$ and will also depend on the orientation of the input point relative to $\boldsymbol{S}_T$.

In order to recognize NN input not foreseen in the NN training phase and thus out of scope of the inversion algorithm, the validity check (Schiller and Krasnopolsky, 2001) can be used. Let the model $S = \boldsymbol{F}(G)$ have an inverse, $G = \boldsymbol{f}(S)$, then, by definition, $S = \boldsymbol{F}(\boldsymbol{f}(S))$.

Further, let f_{NN} be the NN emulating the inverse model in the domain $\mathbf{S_T}$. The result of G_0 = $f_{NN}(S_0)$ for $S_0 \notin \mathbf{S_T}$ may be arbitrary, and in general, $\mathbf{F}(f_{NN}(S_0))$ will not be equal to S_0. The validity of $S = \mathbf{F}(f_{NN}(S))$ is a necessary condition for $S \in \mathbf{S}$. Now, if in the application stage of the NN, f_{NN} , S is not falling into domain $\mathbf{S_T}$, the NN, f_{NN} , is forced to extrapolate. In such a situation the validity condition may not be fulfilled, and the resulting G in general is meaningless. For operational applications it is necessary to signal such events to the next higher evaluation level. In order to perform the validity test the FM must be applied after each inversion. This requires a fast but accurate FM. Such FM can be achieved by a NN emulating accurately the original FM, S = $\mathbf{F_{NN}}(G)$. So, the validity check algorithm consists of a combination of inverse and forward NNs that, in addition to the inversion, computes a quality measure for the inversion:

$$\delta = \| S - \mathbf{F_{NN}}(f_{NN}(S)) \| \qquad (3)$$

In conclusion, the solution to the problem of scope check is obtained by verifying the retrieved parameters using a NN emulating the FM and comparing the result with the measurement. This procedure (*i*) allows the detection of situations where the forward model is inappropriate, (*ii*) does an "in scope" check for the retrieved parameters even if the allowed region has a complicated geometry, (*iii*) can be adapted to all cases where a NN is used to emulate the inverse of an existing forward model.

2. SSM/I WIND SPEED RETRIEVALS

The OMBNN3 SSM/I retrieval algorithm (Krasnopolsky et al., 1999) is running as the operational algorithm in the data assimilation system at NCEP (NOAA) since 1998. Given five brightness temperatures, it retrieves four geophysical parameters: ocean surface wind speed, water vapor and liquid water concentrations, and sea surface temperature. At high levels of liquid water concentration the microwave radiation cannot penetrate clouds and surface wind speed retrievals become impossible.

Brightness temperatures for these occasions fall far outside the training domain $\mathbf{S_T}$. However, the retrieval algorithm in these cases, if not flagged properly, will produce wind speed retrievals, which are physically meaningless (i.e. not related to actual surface wind speed). Usually a statistically based retrieval flag is used to indicate such occurrences. Under complicated local conditions, however, this flag, because it is based on global statistics, produces significant amount of false alarms or does not produce alarms where needed.

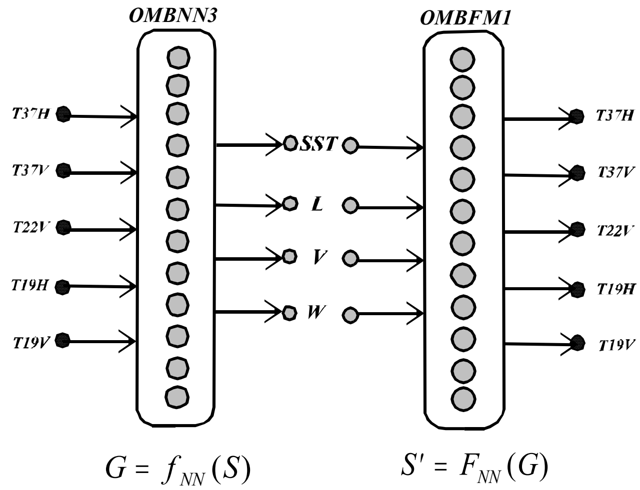

Fig. 1 SSM/I retrieval algorithm (OMBNN3) emulating the inverse model to retrieve vector G of four geophysical parameters: ocean surface wind speed (W), water vapor (V) and liquid water (L) concentrations, and sea surface temperature (SST) if given five brightness temperatures S = TXXY (XX – frequency in GHz, Y – polarization). This vector G is fed to the OMBFM1 emulating the forward model to get brightness temperatures S′ = TXXY′. The difference ΔS = |S - S′| is monitored and raises a warning flag if it is above a suitably chosen threshold.

An intelligent NN based system shown in Fig. 1 was developed to improve this situation. NN SSM/I forward model OMBFM1 (Krasnopolsky, 1997) is used in combination with the OMBNN3 retrieval algorithm. The validity check shown in Fig. 1, if added to standard retrieval flag, helps to indicate occurrences when S is outside the training domain.

For each satellite measurement S, geophysical parameters retrieved from brightness temperatures S are fed into NN SSM/I forward model, which produces another set of brightness temperatures S'. For S from training domain ($S \in \mathbf{S_T}$) the difference, $\Delta S = |S - S'|$, is sufficiently small, for S outside training

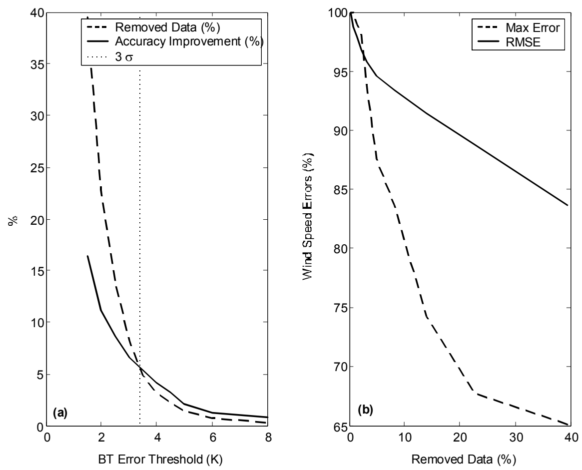

Fig. 2 (a) Percentage of removed data (dashed line) and wind speed accuracy improvement as functions of the threshold for BT discrepancy ΔS. The vertical line shows three standard deviations for ΔS. (b) Wind speed RMS and maximum errors (dashed) dependency on the percentage of the removed data.

domain the difference raises a warning flag if it is above a suitably chosen threshold. Statistical estimates for ΔS show that the standard deviation of this value is about σ = 1.1°K. Fig. 2a shows the percentage of removed data and improvements in the accuracy of the wind speed retrievals as functions of this threshold. The vertical dashed line shows the value of 3σ (about 3.3°K). Fig. 2b illustrates dependencies between the wind speed RMS error and maximum error and the percentage of the removed data. It shows that applying the generalization control reduces the RMS error significantly; the maximum error is reduced even stronger. For example, if the threshold value of about 2.5°K is selected, the validity check removes about 20% of data, which leads to about 10% improvement in the wind speed RMS error and to 30% decrease in the maximum wind speed error. Such a significant reduction in maximum errors means that the validity check approach is very efficient for removing outliers.

It is well known that a lot of outliers are located at the far ends of distributions of retrieved parameters; and it might happen that our procedure would simply remove events with extreme values of wind speeds, water vapor and liquid water concentration. It does not happen. Figs. 3, 4, and 5 show distribution of wind speeds, water vapor and liquid water concentrations for original data set and for data

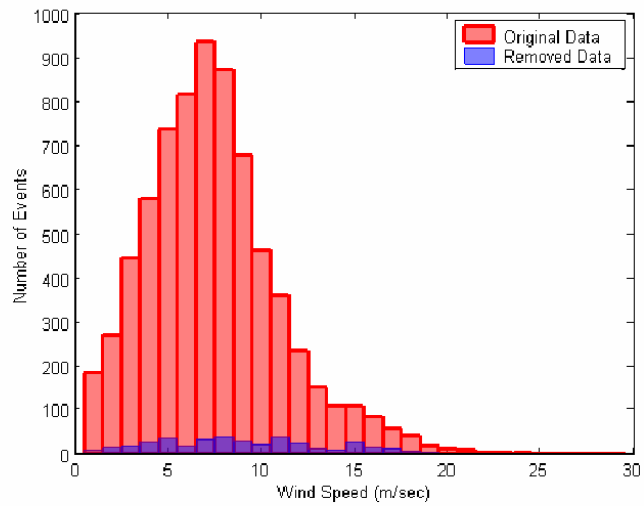

Fig. 3. Wind speed distribution of original data set (red) and of data removed by our quality control procedure (blue); the threshold is equal to 2σ.

removed by our quality control procedure. Distributions for removed data are similar to distributions of original data. Their maxima are located approximately in the same places where the maxima of original distributions are located. It means that this quality control procedure filters out a type of noise which does not depend on weather conditions.

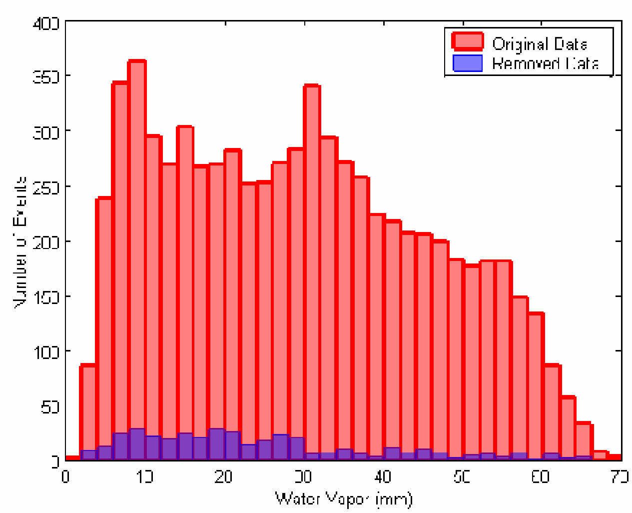

Fig. 4. Columnar water vapor distribution of original data set (red) and of data removed by our quality control procedure (blue); the threshold is equal to 2σ.

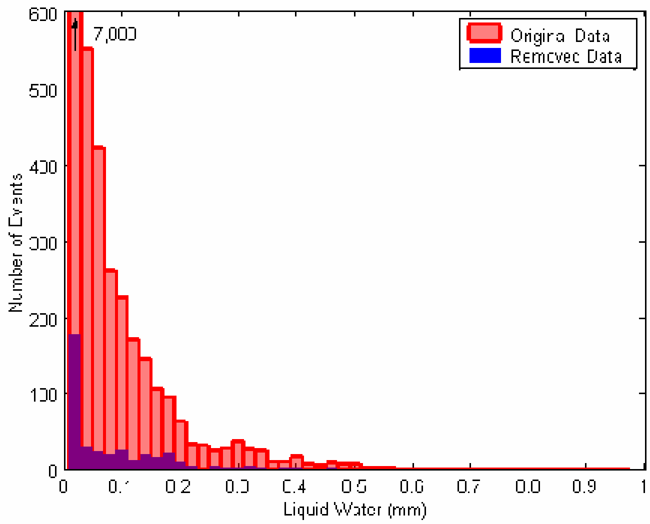

Fig. 5. Columnar liquid water distribution of original data set (red) and of data removed by our quality control procedure (blue); the threshold is equal to 2σ.

3. CONCLUSIONS

A NN based validity check technique (Schiller and Krasnopolsky, 2001) was applied to SSM/I wind speed retrievals. An intelligent NN system incorporating SSM/I NN forward model and SSM/I NN retrieval algorithm was developed and tested. This system successfully reduces both the wind speed RMS errors and maximum errors. Similar systems can be developed for other types of sensors and retrieval parameters.

References

Krasnopolsky, V.M., W.H. Gemmill, and L.C. Breaker, "A multi-parameter empirical ocean algorithm for SSM/I retrievals", *Canadian Journal of Remote Sensing*, 25, pp. 486-503, 1999

Krasnopolsky, V.M., "A neural network-based forward model for direct assimilation of SSM/I brightness temperatures", Technical Note, OMB contribution No. 140, NCEP/NOAA, 1997

Schiller H., and V.M. Krasnopolsky, "Domain check for input to NN emulating an inverse model", International Joint Conference on Neural Networks, Washington, DC, July 15-19, 2001, pp. 2150-2152