

2.4 NEURAL NETWORKS AS A GENERIC TOOL FOR SATELLITE RETRIEVAL ALGORITHM DEVELOPMENT AND FOR DIRECT ASSIMILATION OF SATELLITE DATA INTO NUMERICAL MODELS.

V.M. Krasnopolsky,
Environmental Modeling Center, NCEP, NOAA¹

1. INTRODUCTION: DERIVING GEOPHYSICAL PARAMETERS FROM SATELLITE MEASUREMENTS

Satellite remote sensing data are used by numerical weather prediction (NWP), field meteorology, fisheries, Coast Guard, the oil industry, the Navy and others. Users work with geophysical parameters such as pressure, temperature, wind speed and direction, water vapor, etc. Satellite sensors generate measurements in terms of radiances, sigma naughts, brightness temperatures, etc. Satellite retrieval algorithms which transform satellite measurements into geophysical parameters play the role of mediator between satellite measurements and users.

Conventional methods for using satellite data (standard 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), to a satellite measurement, s (e.g., SSM/I brightness temperatures),

$$g = f(s) \quad (1)$$

where both g and s may be vectors. The TF, f , may be derived explicitly or assumed implicitly. Standard retrievals have the same spatial resolution as the sensor measurements and produce instantaneous values of geophysical parameters over the areas where the measurements are available. Geophysical parameters derived using standard retrievals can be used for many applications, for example, in NWP data assimilation systems. In this case, a contribution to the analysis cost function χ_g from a particular retrieval, g^o , is:

$$\chi_g = \frac{1}{2}(g - g^o)^T (O + E)^{-1}(g - g^o) \quad (2)$$

where $g^o = f(s^o)$ is retrieved geophysical parameter (s^o - a sensor measurement), g - value of this geophysical parameter in analysis; O and E - expected error covariance of the observations and of the retrieval algorithm. Because standard retrievals are based on solution of inverse problem which is usually mathematically ill-posed, it has some rather subtle properties and error characteristics (Eyre, 1987), which cause additional errors and problems in retrievals (e.g.,

amplification of errors, ambiguities, etc.). As a result, high-quality sensor measurements are converted into lower-quality geophysical parameters.

This type of errors can be avoided or reduced, using variational retrievals (or inversion) through direct assimilation of satellite measurements (Lorenc, 1986; Parrish and Derber, 1992; Phalippou, 1996; Prigent et al., 1997).

In this case, due to direct assimilation of sensor measurements, the entire data assimilation system is used for inversion (as a retrieval algorithm). In this case, a contribution to the analysis cost function χ_s from a particular sensor measurement, s^o , is:

$$\chi_s = \frac{1}{2}(s - s^o)^T (O + E)^{-1}(s - s^o) \quad (3)$$

where

$$s = F(g) \quad (4)$$

F is a forward model (FM) which relates an analysis state vector g (vector of geophysical parameters in analysis) to a vector of simulated sensor measurements, s ; O and E - expected error covariance of the observations and of the forward model. The retrieval in this case is an entire field(global in the case of the global data assimilation system) for the geophysical parameter g which has the same resolution as the numerical model used in the data assimilation system. This resolution may be lower than the resolution of standard retrievals. The variational retrievals are also not instantaneous but averaged in time over the analysis cycle (several hours); however, the field is continuous and coherent. It is important to emphasize one very significant difference between using the TF for standard retrievals and the FM in variational retrievals. In standard retrievals the TF (1) is applied one time per sensor observation to produce a geophysical retrieval. In variational retrievals the FM and its partial derivatives (the number of derivatives is equal to $m \times n$, where m and n are the dimensions of the vectors g and s respectively) have to be estimated for each of k iterations performed during minimization of the cost function (3), that is $(m \times n + 1) \times k$ times (e.g., about 3000 times for SSM/I). Taking into account that an FM is often much more complicated than a TF, the requirements for simplicity of the FM used in the variational retrievals are very restrictive,

¹Corresponding author address: Dr. V. Krasnopolsky, EMC/NCEP, 5200 Auth Rd., Camp Springs, MD 20746;
e-mail: kvladimir@sun1.wwb.noaa.gov

and variational retrievals may require some special, simplified versions of FMs.

2. FORWARD AND INVERSE PROBLEMS IN REMOTE SENSING

2.1 Forward Models

The above consideration shows that both standard and variational retrievals require some kind of conversion procedure, either a TF (retrieval algorithm) or a FM, to relate geophysical parameters to satellite measurements. The FM and TF are solutions of a remote sensing forward or inverse problem respectively. A generic remote sensing forward problem is symbolically represented by eq. (4) where $s \in \mathcal{R}^n$ is a vector of satellite measurements (vector of BT's in the case of SSM/I) and $g \in \mathcal{R}^m$ is a vector of geophysical (atmospheric and surface) parameters which influence the measurement. For example, after some simplification and extensive empirical parametrization a radiative transfer FM for SSM/I BTs ($s = T_{v,\pi}$) may be reduced to a closed algebraic version (e.g., Wentz, 1997),

$$T_{v,\pi} = \varepsilon_{v,\pi} \tau_v T_s + T_{U'} + (1 - \varepsilon_{v,\pi}) \tau_v (\Omega_{v,\pi} T_D + \tau_v T_{BC}) \quad (5)$$

where all terms in eq. (5) are empirical functions of wind speed, W , columnar water vapor, V , columnar liquid water, L , and sea surface temperature, T_s . Such physically-based or radiative transfer-based forward models use many empirical data for parametrization. For example, Wentz (1997) used 35,650 buoy-SSM/I matchups and 35,108 radiosonde-SSM/I matchups to fit more than 100 empirical parameters contained in different terms of eq. (5). Finally, this SSM/I FM (5) may be formally written as a system of algebraic equations,

$$T_{v,\pi} = F_{v,\pi}(X), \text{ where } X = \{W, V, L, T_s\} \quad (6)$$

An alternative empirical approach can be applied to develop empirical forward models (or geophysical model function) based on empirical data. If a set of collocated in space and time satellite s and ground g observation - matchup data set $\{s, g\}$ - are collected or simulated, then an empirical FM can be developed based on this data set. Recently an empirical neural network model has been developed for five SSM/I BTs by Krasnopolsky (1997a) based on only about 3,500 matchups (see Section 4.1). This model can be formally described by eq. (6); however, the function F , in this case, is different. It is important to note that such an empirical model requires much less empirical data for development than the physically-based FM (5), it is more accurate (see Table 1 below) and much simpler (crucial for the direct assimilation) than the latter.

2.2 Retrieval Algorithms

A retrieval algorithm is a particular representation for a TF (1) and is also a solution of the inverse problem. Physically-based retrieval algorithm is an inversion of a physically-based forward model, therefore, it requires a physically-based FM as a necessary prerequisite, and, as a consequence, a large amount of empirical data for development (e.g., Wentz, 1997). An empirical algorithm does not require a FM to be developed; but a representative matchup data set is a prerequisite in this case. If we consider again the SSM/I as an example, most of SSM/I empirical wind speed algorithms (including the latest NN algorithm) have been developed using data sets of about 3,500 matchups (an order of magnitude less than for the physically-based algorithm by Wentz, 1997). For these empirical algorithms the resulting accuracies of retrievals are comparable or even better (for NN algorithm) than accuracies for the physically-based algorithm (see Table 2 below).

The inversion technique which is usually applied in physically-based retrieval algorithms to invert FM was described by Wentz (1997) for SSM/I retrieval algorithm. In this case, the TF, f , (1) is not determined explicitly, it is only determined implicitly for each BT vector $\{T_{v,m}\}$. Symbolically, the retrieval algorithm can be written as

$$X = f(T, T_s) \quad (7)$$

It is important to emphasize that the algorithm (7), by definition, is a multi-parameter algorithm, since it retrieves a vector X of several geophysical parameters (W , V , and L) simultaneously. In addition to BTs, T , this algorithm requires a SST value T_s as an input to produce retrievals.

Empirical algorithms are based on an approach which, from the beginning, assumes the existence of an explicit analytical representation for a TF, f . Some mathematical model, f_{mod} , is usually chosen (usually some regression) which contains a vector of empirical parameters $a = \{a_1, a_2, \dots\}$,

$$g_i = f_{mod}(T, a) \quad (8)$$

where these parameters are determined from empirical matchup data set $\{g_i, T\}$. The subscript i in g_i stresses the fact that most of empirical retrieval algorithms are single-parameter algorithms; they retrieve only wind speed (Goodberlet, 1989), or water vapor (Alishouse, 1990), or cloud liquid water (Weng and Grody, 1994), etc. Single-parameter algorithms have certain problems which are discussed below.

3. MATHEMATICAL TOOLS FOR DEVELOPING EMPIRICAL FORWARD MODELS AND RETRIEVAL ALGORITHMS

The above considerations show that both empirical forward models (4 and 5) and retrieval algorithms (1 and 8) can be considered as mappings which map a vector of sensor measurements, s (or T) $\in \mathcal{R}^n$, to a vector of geophysical parameters, g (or X) $\in \mathcal{R}^m$ (TFs, f) or vice versa (FMs, F). These mappings are built, using discrete sets of collocated vectors s and g (matchup data sets $\{s_i, g_i\}$). Single-parameter algorithms (8) may be considered as degenerate mappings where a vector is mapped onto a scalar (or a vector space onto a line).

The linear regression (LR) is the most attractive tool for developing empirical algorithms. It is simple; it has a well developed theoretical basis which enables a user to perform various statistical estimates. In the case of the LR, a linear model is built for FM or TF. For example for (8) we have:

$$g_i = LR(T, a) = \sum_j a_j T_j$$

here LR means linear regression and a is a vector of unknown parameters (regression coefficients). The problem with the LR is that it works with high accuracy in a broad range of the variability of arguments only if the problem and the function which it represents (FM or TF), is linear. If the problem is nonlinear, the LR can give only local approximation, or, if it is applied globally, this approximation has poor accuracy.

Because, in general, forward models and TFs are nonlinear functions of their arguments, nonlinear regressions (NRs) are better suited for modeling forward models and TFs. NR may be applied in many different ways. For example, f_{mod} in (8) can be chosen as a complicated NR function:

$$g_i = f_{NR}(T, a) \quad (9)$$

on the other hand, f_{mod} can be introduced as an expansion in a set of nonlinear functions $\{\varphi_j\}$:

$$g_i = \sum_j a_j \varphi_j(T) \quad (10)$$

In either case, if we use NR (9) or (10), we need to specify in advance a particular type of nonlinear function f_{NR} , or φ_j which we use. In other words, we need to introduce in advance a particular kind of nonlinearity, which we use to approximate the FM or TF under consideration. This may not always be possible, because we may not know in advance what kind of nonlinear behavior a particular FM or TF demonstrates, or this nonlinear behavior may be different in different regions of the FM's or TF's domain. If an improper NR is chosen (by chance), it

may represent a nonlinear FM or TF less accurate than a LR.

In the situation described above, where we do know that the TF or FM is nonlinear but do not know what kind of nonlinearity to expect, we need a flexible, self-adjustable approach that can accommodate various types of nonlinear behavior and represent a broad class of nonlinear mappings. Neural networks (NNs) are well suited for a very broad class of nonlinear approximations and mappings. It has been shown (e.g., Chen and Chen, 1995; Hornik, 1991; Funahashi, 1989; Gybenko, 1989) that a NN with one hidden layer can approximate any continuous mapping defined on compact sets in \mathcal{R}^n . Thus, any problem which can be mathematically reduced to a nonlinear mapping like (1), (4), (6), (8), etc. can be solved using a NN with one hidden layer.

Here several main properties of NNs are presented which make them a very suitable generic tool for nonlinear mapping (and, therefore, for algorithm development). Some of these properties have been illustrated above, others are described in the literature:

- NNs are able to model complicated nonlinear input/output relationships (any continuous nonlinear mapping).
- NNs are robust with respect to random noise and sensitive to systematic, regular signals (e.g., Kerlirzin and Réfrégier, 1995).
- NNs are fault-tolerant. An output value is created using all weights and biases so that an error in one of them usually causes only minor change in the output value (Cheng, 1996).
- NNs are well-suited for parallel processing and hardware implementations (Cheng, 1996) (all neurons in the same layer are completely independent and can be evaluated simultaneously).
- While training the NN is sometimes time consuming, its application is not. After the training is finished (it is usually performed only once), each application of the trained NN is practically instantaneous (several tens of floating point additions and multiplications -- microseconds on modern computers).

4. NEURAL NETWORKS FOR SSM/I DATA.

In previous sections we discussed theoretical possibilities and premises for using NNs for modeling TFs and FMs. In this section we illustrate these theoretical considerations using applications of the NN approach to the SSM/I forward and retrieval problems. Many different retrieval algorithms and several forward models have been developed for SSM/I sensor; and several different databases are available for the algorithm development and validation. For a detailed discussion of the database used here see Krasnopolsky et al., (1996), and Krasnopolsky, (1997a).

4.1 NN empirical forward model for SSM/I.

The empirical SSM/I FM represents the relationship between a vector of geophysical parameters \mathbf{X} and a vector of satellite BTs \mathbf{T} , where $\mathbf{T} = \{T19V, T19H, T22V, T37V, T37H\}$, $\mathbf{X} = \{W, V, L, T_s$ (or $SST)\}$. Four geophysical parameters were included in \mathbf{X} (wind speed, W , columnar water vapor, V , columnar liquid water, L , and SST) which are the main parameters determining satellite BTs, and which are used as inputs in the physically based FMs of Petty and Katsaros (1994) and Wentz (1997) (see Table 1). The NN, OMBFM1, which implements this FM has 4 inputs, $\{W, V, L, SST\}$, one hidden layer with 12 neurons, 5 standard BT outputs $\{T19V, T19H, T22V, T37V, T37H\}$, and 20 auxiliary outputs which produce derivatives of the outputs with respect to the inputs, or $\partial T_i / \partial x_j$. These derivatives constitute the Jacobian matrix $\mathbf{K}[\mathbf{X}] = \{\partial T_i / \partial x_j\}$ which emerges in the process of direct assimilation of the SSM/I BTs when the gradient of the SSM/I contribution to the cost function (3), χ_s , is calculated.

Fig. 1 shows the OMBFM1 architecture. Since these auxiliary outputs (Jacobian matrix \mathbf{K}) are not independent, we did not include them in the error function during the training, hence, only the standard outputs \mathbf{T} are involved in the training process. Including these additional outputs in the NN architecture simplifies the use of our NN FM for direct assimilation because, as we showed in Section 1, in the process of variational retrievals the FM and its derivatives have to be estimated about 3,000 times per satellite measurement.

The matchup database for the F11 SSM/I has been used for training (about 3,500 matchups) and testing (about 3,500 matchups) our forward model. Only matchups with $R \leq 15$ km and $t \leq 15$ min have been selected. The FM was trained on all matchups which correspond to clear + cloudy conditions in

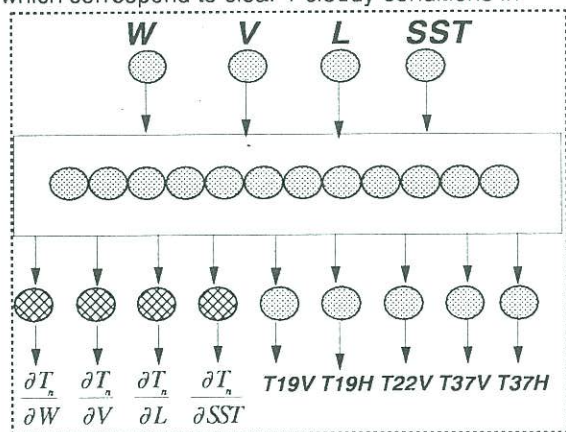


Fig. 1. NN forward model OMBFM1

Table 1. Comparison of physically based radiative transfer and empirical NN forward models for clear and clear+cloudy (in parentheses) conditions.

Author	Type	Inputs	BT RMS Error	
			V	H
Petty & Katsaros	Physically-based	$W, V, L, SST, \theta, PO^2, HWV^4, ZCLD^4, Ta^5, G^6$	1.9 (2.3)	3.3 (4.3)
Wentz (1997)	Physically-based	W, V, L, SST, θ	2.3 (2.8)	3.4 (5.1)
Krasnopolsky	NN, empirical	W, V, L, SST	1.5 (1.7)	3.0 (3.4)

¹ θ - incidence angle; ² PO - surface pressure

³ HWV - vapor scale height; ⁴ $ZCLD$ - cloud height

⁵ Ta - effective surface temperature; ⁶ G - lapse rate

accordance with the retrieval flags introduced by Stogryn et al. (1994).

More than 6,000 matchups for the F10 instrument have been used for validation and comparison of the NN FM with physically-based forward models by Petty and Katsaros (1994) and Wentz (1997)². For detailed training, testing and validation statistics see Krasnopolsky (1997a,b). Table 1 presents total statistics (RMS errors) for three FMs discussed here. RMS errors (in degrees Kelvin) are averaged over different frequencies for the vertical and horizontal polarizations separately for the F10 SSM/I data.

4.2 NN empirical SSM/I retrieval algorithms

About ten different SSM/I wind speed retrieval algorithms, both empirical and physically-based, have been developed using a large variety of approaches and methods. Here we perform a comparison of several of them in order to illustrate some properties of the different approaches mentioned in previous sections and some advantages of the NN approach.

The first global SSM/I wind speed retrieval algorithm (GSW) was developed by Goodberlet et al. (1989). This algorithm is a single-parameter algorithm (it retrieves only wind speed), and it is linear with respect to BTs (a multiple LR was used). This algorithm met specified accuracy criteria (2 m/s between 3 and 25 m/s) under rain-free and low moisture conditions. According to Goodberlet et al., however, this algorithm deteriorates rapidly in areas where rain and cloud cover occur. Wentz (1997) developed a physically-based approach to estimate surface wind speeds from the SSM/I. However, Wentz's approach requires some external inputs not available in the SSM/I data stream.

²The author coded both Wentz's FM and retrieval algorithm based on the detailed description published by Wentz (1997).

The first NN algorithm for SSM/I has been developed by Stogryn et al. (1994) for retrieving wind speed from the SSM/I BTs. This algorithm consists of two NNs, one of them performs retrievals under clear and another, under cloudy conditions. Krasnopolsky et al. (1994, 1995a) showed that a single NN (OMBNN1) with the same architecture can generate retrievals with the same accuracy as the two NNs developed by Stogryn et al. under both clear and cloudy conditions. This algorithm can be represented as:

$$W = f_{NN}(T) \quad (11)$$

where W is the wind speed, and $T = \{T19V, T22V, T37V, T37H\}$. Application of the OMBNN1 algorithm led to a significant improvement in wind speed retrieval accuracy for clear conditions. For higher moisture/cloudy conditions, the improvement was even far greater (25-30%) compared to the GSW algorithm. The increase in the areal coverage due to an improvements in accuracy was about 15% on average and higher in areas with higher meteorological activity.

First NN algorithms give very similar results because they have been developed using the same matchup database. This database, however, does not contain matchups with the wind speed higher than about 20 m/s and contains very few matchups with wind speeds higher than 15 m/s. These algorithms also are single-parameter algorithms, i.e. they retrieve only one parameter - wind speed, therefore they can not account for the variability of all related atmospheric (e.g., water vapor and liquid water) and surface (e.g., SST) parameters (especially important at higher wind speeds). This is why these NN algorithms demonstrate the same problem; they can not generate wind speeds higher than 18 - 19 m/s. The high wind speed performance has been improved in the OMBNN2 algorithm (Krasnopolsky et al., 1995b) by introducing new methods of NN training which enhance learning the high wind speed behavior and by using a bias correction. The OMBNN2 algorithm performs better than OMBNN1 for wind speeds higher than 15 m/s; however, it still can not generate wind speeds higher than 19 - 20 m/s without a bias correction because the same training set was used. It is also a single-parameter algorithm and is sensitive to the variability of related atmospheric and surface parameters at higher wind speeds.

The next generation NN algorithm - a multi-parameter NN algorithm (OMBNN3; Krasnopolsky et al., 1996) solved the high wind speed problem through three main advances. First, a new buoy/SSM/I matchup database, created by NRL and enriched by high latitude, high wind speed events (up to 26 m/s) from European OWS MIKE and LIMA, was used for the development of this algorithm. Second, the method of NN training which enhances learning the high wind speed behavior was used. Third, the variability of the

primary related atmospheric and surface parameters was taken into account: wind speed, columnar water vapor, columnar liquid water, and SST are retrieved simultaneously. In this case, the relation (11) is modified:

$$X = f_{NN}(T) \quad (12)$$

where $X = \{W, V, L, SST\}$ is now a vector, and W is the wind speed, V - columnar water vapor, L - columnar liquid water, and SST - sea surface temperature. The OMBNN3 algorithm uses five SSM/I channels: 19 GHz and 37 GHz (horizontal and vertical polarization) and 22 GHz (vertical polarization). It does not use any additional inputs. SST is an output here rather than additional input as in Wentz algorithms.

Fig. 2 illustrates the evolution of our NN algorithms from OMBNN1 to OMBNN3. Table 2 shows a condensed comparison of the GSW, Wentz, and OMBNN3 algorithms for more than 15,000 matchups from three different SSM/I instruments F08, F10, and F11.

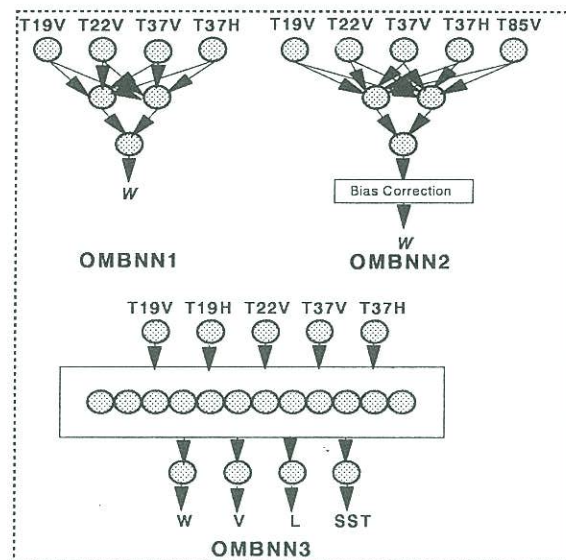


Fig.2 Evolution of the NN architecture from OMBNN1 to OMBNN3

Table 2. Biases and RMS errors (in m/s) for different SSM/I wind speed algorithms for clear and clear+cloudy (in parentheses) conditions.

Algorithm	Bias	RMS Errors
Linear Regression ¹	-0.2 (-0.5)	1.6(2.2)
Physically-Based ²	0.5(0.2)	2.0(2.3)
OMBNN3 ³	-0.1(-0.2)	1.6(1.7)

¹Goodberlet et al., 1989; ²Wentz, 1997;

³Krasnopolsky et al., 1996

