

## 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, NOAA<sup>1</sup>

### 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  $\chi_g$  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  $\chi_s$  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,

<sup>1</sup>Corresponding 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.