Assimilation the GPS Ionospheric Occulted Data

Zeng zhen   Hu xiong   Zhang xunjie
(Wuhan Institute of Physics and Mathematics, Chinese Academy of Sciences
P.O. Box 71010, 430071, Wuhan, Hubei, P.R. China
E-mail: chbo@wipm.whcnc.ac.cn, Fax: 0086-27-87869194)

Abstract: Artificial neural network (ANN) is put forward to assimilating of GPS ionospheric occulted data in order to take full advantage of the abundant GPS occulted data. In this paper, we chose a feedforward, full-connected network, which based on the back-propagation algorithm. Universal time, latitude, longitude, height, Kp index, and $F_{10.7}$ solar flux are chosen as the input vectors of the network as well as the electron density as the output vectors. The GPS observed data on May 24th, 1996 were taken as training samples to train an ANN, then using the well-trained ANN to predict the electron density on 25th. Comparison of the predicted results and observed data illustrated that ANN is a promising method in assimilating the GPS occulted data to establish the ionospheric weather prediction model. Furthermore, the accurate, rich occulted data are pivotal factor on ensuring the good performance of ANN.

Key word: assimilation occultation artificial neural network

Introduction

Nowadays, GPS/MET has showed great potential utility for weather observation and prediction in providing a unique combination of global coverage, high vertical resolution and accuracy, long-term stability, and all-weather capability [3]. However, the raw measurements of GPS limb soundings are the phase delays and amplitudes of the signals, which are different with the traditional meteorological variables. Additionally, the retrieved refractivity may contain significant errors because of the unresolved problems in inversing techniques. Effectively assimilating GPS occulted data to establish a weather prediction models is not a trivial matter.

Over the past several years, a considerable amount of effort has been devoted to the assimilation strategies. Simulations using variational data assimilation techniques by NCAR (National Center for Atmospheric Research) and ECMWF (European Center for Medium Range Forecast)[2][8] have shown that profiles of bending angles and refractivity from radio occultation experiments can strongly support climate and numerical weather predictions (NWP) analysis. Zou et al. [8] incorporated forward ray-tracing code into a three-dimensional variational (3DVAR) atmospheric data assimilation system.

In here, we wish to take advantage of ANN (Artificial Neural Network) to realize the assimilation of GPS ionospheric occulted data. ANN is a physical cellular system, which can acquire, store, and utilize experiential knowledge. Basically, ANN excels in solving problems of pattern recognition and identification, classification, pattern association, function approximation, forecasting and prediction, and so on. Recently, there also has been much interest in the use of ANN in ionospheric prediction models, such as the ionospheric parameter $f_0F_2$ model [1][6], and has achieved rather satisfied
This paper started with a description of GPS occulted data will be given including inversion the electric density with the signal frequency. Then, a strategy is described to deal with occulted data problems, especially the check of occulted data, the select of input neurons, the determination of the structure of ANN, the technique of fasting training, and so on. Finally, this method is illustrated on a series real data example, which is based on the GPS/MET occulted data on 24th, May 1996. The ANN was utilized to assimilate the full-day GPS occulted data. In additional, the occulted observed data on 25th, May were compared with the predicted data using the well-trained ANN. The comparison showed that they are quite in agreement. Some discussions will be presented in details.

**Process GPS Ionospheric Occulted Data**

The principle of this sky-based remote sensing technique is based on the influence of an atmospheric or ionospheric refractivity field on electromagnetic waves propagating through the field in limb sounding geometry. The radio wave path between an orbiting transmitter (GPS) and a low orbiting receiver traverses the Earth's atmosphere and is deflected primarily by the vertical gradient of atmospheric refractivity. In the occultation observation, with the radio wave path up or down, the effect can be described by the function \( \alpha(a) \). \( \alpha \) is bending angle, \( a \) is the impact parameter. This technique is described in Kursinski et al [3]. Because of GPS occulted observation have some merits that can't implement by ground observation, such as high vertical resolution, full global coverage, all weather. In this paper, used GPS data is level 2 ionospheric data available from UCAR. The data file contains 0.1Hz timing, orbit, raw carrier phase measurement L1 and pseudorange measurement P1 from the 1575.42 carrier for a specific ionospheric occultation event. The main idea of inversing the ionospheric parameter using single frequency information is briefly presented in here. From L1 and P1, the TEC can be approximately written as:

\[
S_P - S_L = -2 \frac{40.3}{f_1^2} \int \frac{N_v ds}{f_1^2} = 2 \frac{40.3}{f_1^2} TEC ,
\]

where \( S_P \) and \( S_L \) respectively are raw carrier phase measurement and pseudorange measurement, \( f_1 \) is carrier frequency, equal 1575.42MHz. Then, the bending angle \( \alpha \) can approximately be [7]:

\[
\alpha(a) = \frac{40.3}{f_1^2} \frac{dTEC}{da} ,
\]

according Abel inversion, the refractive index \( n \) is:

\[
n(x) = \exp\left[ \frac{1}{\pi} \int_{a}^{\infty} \frac{\alpha(a)}{\sqrt{(a^2 - x^2)}} da \right] ,
\]

where \( x \) denotes altitude. Then, the profile of the ionospheric electron density is given by:
Although the above deduce is based on some approximations and simplifications, this inverse technique is still feasibility in some extent.

**Assimilation Strategy**
In order to make optimal use of GPS occultation data in weather analysis and prediction, optimizing the strategy of the assimilation is needed. In here, we utilize the ANN to do some primary probe in this field. The main flow is listed in following.

**A. Verification GPS Occulted Data**
Checking GPS occulted data is the first step, also very important step in the assimilation process. We made use of the critical frequency on F\(_2\) layer data \(f_0F_2\) which obtained on the global distributing ionosonde stations as referent information. Since \(f_0F_2\) is directly related to the maximum electron density \(N_m\) in the ionosphere, through the relation \(N_m = 1.24 \times 10^{10} f_0F_2\), it can be used to verification the GPS inversion parameter \(N_m\).

Take occulted data on May 24th, 1996 as example. There're totally 23 occultation events. We downloaded \(f_0F_2\) data on the 28 ionosonde stations in the same day from Internet. Because the positions of the ground stations and the occultation are not same, ANN was employed to associate these ground data, then using well-trained ANN to verify \(f_0F_2\) on the positions of occulted observations. The verified results are shown in the following figure 1.

![Figure 1.](image)

The upper figure (Fig. 1.a) is the comparison of the geographical location between the 28 ground stations and the 23 GPS occulted tangent points corresponding the maximum electron density. It's clear that only three ground stations are located on the south hemisphere, so the sample sets contain rather limited information on the south hemisphere. The lower figure (Fig. 1.b) indicates the GPS observed data and ANN predicted results according the ground information. There have sharp differences in several points. Through careful analysis, we found that most of existing discrepancies' positions are where the GPS occultation events happened on the south hemisphere. Maybe due to the learning sets on the south hemisphere are so few that ANN can't make reliable prediction in such areas. Anyway, we decided to reject these data and omitted 5 occultation events, because the good learning set is crucial factor to ensure ANN to make good performance.

**B. Choice input and output neurons**
The ionosphere is a complicated system controlled by many parameters. As with all mathematical modeling, input parameters only that can characterize ionosphere more definitely and effectively should be chosen. In here, we chose universal time (UT), latitude, longitude, altitude, kp index and 10.7cm solar flux (F10.7), total six parameters as input neurons. UT is the primary index of diurnal variation; latitude, longitude, altitude represent spatial variation; Kp implies geomagnetic disturbances, which effects
Determination assimilation variable is also very important. Results from recent studies [2] have concluded the best strategy to assimilate GPS ionosphere occultation data is refractivity. Hence, we choose electron density as output neuron, which is similar to refractivity. In ANN, it is first assumed that the output can be calculated from the input using some kind of function. Our target is to utilize ANN to seek for a sound function mapping relationship between the input vector and the output vector.

C. Determination ANN Configuration

There are multitudes of different types of ANNs. The network used in here [5](Fig. 2) is a static feedforward network. It is full connected, not just every layer is only connected to the succeeding layer. The number of the hidden layer depends on the complexity in the relationship between the input data and the desired output. In order to simplify, only one hidden layer is applied in this study. Although no rules exist for determining the number of neurons of hidden layer, only by trial-and-error, the ratio of the number of training samples and the number of internal weights often is fixed on ten over one in the practice. The number of training samples is almost 3000, so we chose 30 hidden neurons.

During the training process the network is presented with values of the six inputs, as a six-component input vector, which produces a one-component output vector. This is then compared with measured archived values of the output vector corresponding to the inputs. The differences are then employed in the backpropagation (BP) algorithm to change the weights in such a way as to minimize these differences. This process can be denoted in mathematics by:

\[
E = \sum_{t=1}^{T} E(t) = \sum_{t=1}^{T} \sum_{i=1}^{n} \frac{1}{2} (\hat{Y}_i(t) - Y_i(t))^2
\]  

(5)

in here, \( \hat{Y}_i \) is generated output component using ANN, \( Y_i \) is the observed output component, and the square of their differences is defined as error function E. The constant T is the number of the training samples, and n is the number of output neurons. In basic backpropagation algorithm, the gradient descent is applied to adjust the connecting weights \( W_{ij} \) as to minimize the error E. Once the rms error E arrived the predestined threshold, the training is completed. And then, these weights represent the "memory" of the network, which represents "knowledge" acquired by the network about the relationship between the input and output quantities.

D. Training and Testing

Training of a BP ANN is achieved by presenting inputs to the network with the desired outputs. There are two problems we would emphasis in this part.

First, in order to circumvent over training the learning set, the selected set of examples is split in a training and a validating set. 17 events, totally 3148 samples were splited randomly into 2857 samples for training and 289 samples for validation. Weights
are optimized using the training set, and cross-validation with the validating set ensures an overall good performance. From figure 3, black and blue line separately denoted the rms error of training set and validating set varied with the iterative number using basic BP algorithm. 2200 epochs later, the rms error of validating set was almost invariable, though the error of training set kept decrease. Training procedure of ANN was stopped when the misfit of the validating set has reached a minimum. Hence, in this set, the threshold of error function tended to predestine in 0.06.

Second, how to accelerate the process of training ANN. In here, we tried two kinds of BP algorithm. One is basic BP algorithm, which based on the gradient descend of error function with respect to weights. The other is resilient BP (RPROP) method [4], which is one of ameliorative algorithm of basic BP algorithm. The direction of adjusting weights not depends on the value of the gradient of error function, but relies on the sign of the gradient. Comparison of the convergent speed under these two learning algorithms shows that the speed of training ANN using RPROP is much quicker than basic BP (Fig. 3). After 250 epochs, the error (green line) calculated by RPROP method has dropped into 0.06, while the error (black) using basic BP algorithm was still in 0.6. In contrast, the convergent speed of RPROP is almost the ten times of that using basic BP method. Undoubtedly, we prefer to RPROP learning algorithm.

Once the error arrived the predestined threshold, we thought that this ANN is well trained. Finally, we can utilize this well-trained ANN to do prediction.

**Case Studies**

We took the GPS occulted data on May 24, 1996 as learning sets. There're totally 17 occultation events distributed over the world, each events contains almost 190 samples. We split these 17 groups into two categories, 15 groups are for training, have 2857 samples; remained 2 groups, which have 289 samples, are for validation. According the number of samples and through trial-and-error respectively, we decided to choose 30 hidden layer neurons. Adopting resilient BP algorithm, after about 287 iterations, the training of ANN was completed.

Figure 4 illustrates the learning ability of ANN for training set. For figure 4.a. as an instance, the red line and blue line separately mark the output values of ANN and the observed GPS occulted results on an occultation event. These two electron density profiles agreed very well. The standard error of these two curses is 0.1165, and the correlative coefficient reaches 0.9907. Furthermore, even in the detailed variances, ANN can obtain good performance. Figure 5 shows the comparative results between the observed data on one occulted event which chosen as validated sets and the calculated results of ANN. The standard error and correlative coefficient are still satisfied. In order to test whether this ANN can perform the short-term ionospheric prediction, we chose the GPS occulted data observed on May 25 as predicted samples. There are two comparative results (Fig. 6). On the whole, these two curves have a good agreement. The correlative coefficients are all above 0.75, even better than the result of the validated data. From this point, we can reversibly infer that GPS occulted data is not
very ideal. If the GPS occulted data is more accuracy, we can affirm the predictive results will be more superior.

**Summary and Discussion**

We have investigated the application of artificial neural networks to the assimilation of GPS ionospheric occulted data and forecasting of the electron density in the ionosphere. Our research has led us to draw following conclusions:

1. From this primary attempt, we think ANN is a unique way to establish an ionosphere weather model using GPS occulted data. With the improvement of data quality and abundance of data resource, ANN will embody more powerful strength in associating data and short time forecasting.

2. The used prototype of ANN is a feedforward BP structure. The convergent speed of RPROP method is much quicker than basic BP algorithm.

3. Six input neurons and an output neuron compose the windows of the black box, the complicated nonlinear process is wrapped in the ANN. Input and output constitute the main parameter characterized the physical phenomena.

4. A representative data set is the critical factor to ensure the good performance of ANN. So, although assimilating GPS occulted data using ANN is feasible, but the attention should focus on the verification of GPS occulted data.

Acknowledgments. The authors thank UCAR for providing the GPS/MET occulted data.

References


FIGURES NOTE

Fig. 1. a) Comparison the geographic location of the ionosonde stations and the occultation tangent point corresponding the maximum electron densities, which separately marked by blue, crisscross symbols and red, starry symbols.

b) Verification results utilizing all available ground observed $f_0F_2$ data. The x-axis meant the sample number, and y-axis showed the normalized $f_0F_2$ value. The red line illustrated the assimilating results on the positions of each occulted observation using ANN according the ground observed data, while the blue line marked the occulted observed data.

Fig. 2. Network configuration characterized by feedforward, full connected, backpropagation. UT, Lat, Lon, h, Kp, F10.7 separately substituted universal time, latitude, longitude, height on the occulted tangent point, kp index, and 10.7cm solar flux on that moment.

Fig. 3. Comparison the convergent speed between the basic BP algorithm and the RPROP method as well as the basic strategy about how to circumvent overfitting. The green and red lines showed separately the convergent speed of the training and validating process using RPROP algorithm, and so did the blue and black lines using basic BP algorithm.

Fig. 4. Comparison the training results and observed data. The x-axis shows the natural logarithm of electron density, and y-axis is the height. There're totally four occulted events compared in here. The standard error and correlated coefficient between the training results and observed data are marked on each figure.

Fig. 5. Comparison the validating results using ANN and the observed data.

Fig. 6. Comparison the next day predicting results using ANN and the observed data on May 25th.
Fig. 1. a) Comparison the geographic location of the ionosonde stations and the occultation tangent point corresponding the maximum electron densities, which separately marked by blue, crisscross symbols and red, starry symbols.

b) Verification results utilizing all available ground observed \( f_0F_2 \) data. The x-axis meant the sample number, and y-axis showed the normalized \( f_0F_2 \) value. The red line illustrated the assimilating results on the positions of each occulted observation using ANN according the ground observed data, while the blue line marked the occulted observed data.

Fig. 2. Network configuration characterized by feedforward, full connected, backpropagation. UT, Lat, Lon, h, Kp, \( F_{10.7} \) separately substituted universal time, latitude, longitude, height on the occulted tangent point, kp index, and 10.7cm solar flux on that moment.
Fig. 3. Comparison the convergent speed between the basic BP algorithm and the RPROP method as well as the basic strategy about how to circumvent overfitting. The green and red lines showed separately the convergent speed of the training and validating process using RPROP algorithm, and so did the blue and black lines using basic BP algorithm.
Fig. 4. Comparison the training results and observed data. The x-axis shows the natural logarithm of electron density, and y-axis is the height. There're totally four occulted events compared in here. The standard error and correlated coefficient between the training results and observed data are marked on each figure.
Fig. 5. Comparison the validating results using ANN and the observed data.
Fig. 6. Comparison the next day predicting results using ANN and the observed data on May 25th.