

# Site-specific correlation of GPS height residuals with soil moisture variability using artificial neural networks

Clement Ogaja

*Geospatial and Earth Monitoring Division, Geoscience Australia, Australia*

(Received November 6, 2005; Revised January 19, 2006; Accepted January 27, 2006; Online published February 10, 2006)

Historical time series generated from GPS sites reveal significant seasonal variations in the vertical direction. It is well known that continental waters (soil moisture, snow, ground water) mass redistribution is one of the potential contributors to these observed seasonal variations although their actual loading effects on GPS results are least well understood. A number of hydrology model outputs exist with a fair degree of uncertainty. Studies of interrelations between anomalous vertical variations observed at geodetic sites and hydrology model outputs are useful, in particular, as the hydrology models continue to be refined. In this paper, artificial neural networks is proposed for correlating GPS height residuals with the soil moisture variability. Time series from eight sites of the global GPS network are used to correlate with the soil moisture information from the US National Oceanographic and Atmospheric Administration (NOAA) Climate Prediction Center's land data assimilation system (CPC LDAS). The results show the feasibility of the neural interpretation in terms of the correlation coefficients ( $\sim 0.6$ ) and root mean square errors (about 15% of residual range). Other geodetic time series can be used for the same purpose, such as from SLR, VLBI, and absolute gravity.

**Key words:** Geodesy, Global Positioning System (GPS), hydrology, time series.

## 1. Introduction

Historical time series generated from GPS sites reveal significant seasonal variations in the vertical direction, in particular, with an annual period. It is well known that continental waters (soil moisture, snow, ground water) mass redistribution is one of the potential contributors to these observed seasonal variations (Dong *et al.*, 2002; van Dam *et al.*, 2001). Studies of the interrelations between the hydrology model outputs and anomalous vertical movements observed at geodetic sites may be useful, in particular, as the global hydrology models continue to be refined. Improvements in the hydrologic (or water) load estimation would eventually lead to a recommendation on the best procedure for including water loading corrections into geodetic data, with an added potential benefit of being able to fill or predict periods of data gaps or large uncertainties.

Presently the IERS (International Earth Rotation Service) Special Bureau for Hydrology (SBH) distributes continental water storage data from some of the latest hydrology model developments at <http://www.csr.utexas.edu/research/ggfc/>. Examples include monthly outputs from CPC LDAS model and daily outputs from the NCEP (National Centers for Environmental Predictions)/NCAR (National Center for Atmospheric Research) models. This article adopts the monthly CPC LDAS model outputs to assess the neural network correlation between the water storage variations in top layers of soil (hence soil moisture) and long-term GPS results. The objective is to train and validate a neural net-

work methodology that transforms the input variables (soil moisture) into corresponding non-linear signals that closely approximates the seasonal trend of GPS height residuals. The hypothesis is that if there is a fair amount of correlation ( $\sim 50\%$  or greater) between the non-linear signals and the actual GPS data at the validation stage, the GPS height results are impacted by the soil moisture variability. The methodology, similar to finding the output from a linear regression function, is capable of directly correlating any number of input time series of forcing functions to the output variable through interconnected nodes with trainable weights and bias. In contrast to traditional harmonic analysis, which only predicts the periodic components, it can be trained to recognise and predict both non-linear and non-periodic signals.

It is intended that the neural methodology would be generally applicable to the tasks of correlating long-term geodetic height results (GPS, SLR, VLBI, absolute gravity) with soil moisture, in cases where the actual water loading corrections to the geodetic results have not been modelled or applied. Thus, it can be extended to water storage variations estimated by other models such as the NASA's Global Land Data Assimilation System (GLDAS, <http://ldas.gsfc.nasa.gov/>, Rodell *et al.*, 2004).

## 2. Soil Moisture and GPS Data

Data from eight different locations around the world, six at mid-latitudes (in Western Australia, North Eastern United States, Finland, Eastern Canada, Japan and China), and two equatorial sites (in Africa and South America) are used in the current study. Their geographic locations are depicted in Fig. 1(a).

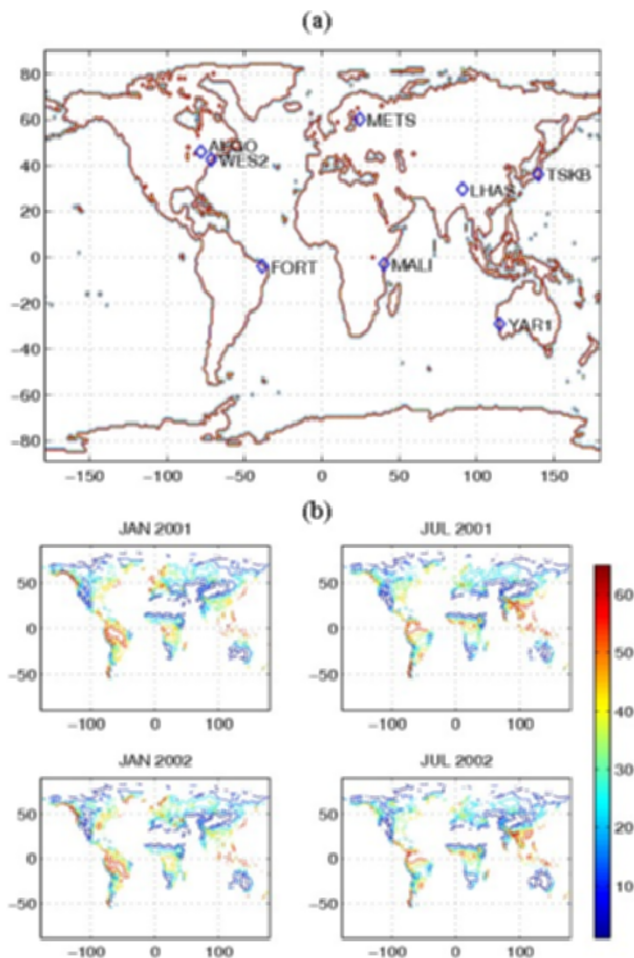


Fig. 1. (a) GPS sites used in the study. (b) Hydrology signals (in cm) at four epochs in the northern winter and summer of 2001 and 2002.

The soil moisture data were obtained from the CPC LDAS model in the IERS SBH Data Archive (<http://www.csr.utexas.edu/research/ggfc/>). These data were available on a 1-degree-by-1-degree grid of monthly temporal resolution (Jan. 1980–Dec. 2004) covering the continental areas and islands except Antarctica. The model is forced by observed precipitation, derived from CPC daily and hourly precipitation analyses, downward solar and long-wave radiation, surface pressure, humidity, temperature and horizontal wind speed from NCEP reanalysis (see also <http://www.cpc.ncep.noaa.gov/soilmst/>). The output consists of soil temperature and soil moisture, the soil moisture, in cm of equivalent water thickness, represents the net water mass stored over the continental areas and islands. Fig. 1(b) show winter and summer epochs of 2001 and 2002.

The time series of GPS height residuals were obtained from the Scripps Orbit and Permanent Array Center (SOPAC) in the form of daily solutions (available at <http://sopac.ucsd.edu>) in 9~10 year periods, mostly from January 1996 to December 2004. They represent the residuals after known geophysical sources associated with the response of the earth to surface mass loading have been removed. The known geophysical sources include pole tides, ocean tides, atmospheric mass loading, non-tidal ocean mass loading, and are well studied by Dong *et al.* (2002). However, the

loading-induced seasonal deformations from terrestrial water storage are the least well understood (van Dam *et al.*, 2001) and a reliable global hydrological model is not yet available (Takle *et al.*, 2005; Wilson and Chen, 2003). This article uses the term residuals to refer to the GPS height results with the known geophysical sources removed except those due to soil wetness (snow mass, soil moisture, ground water) and other unmodelled error sources.

The daily residuals were averaged into weighted monthly estimates that correspond to the epochs of the soil moisture data. Both the resulting GPS time series and the soil moisture data were mean-centered and linearly detrended before being used in the analysis. It is assumed that water storage loading may be the cause of some of the long-period variability, which is observed in the GPS time series, especially where the water storage loading effect is large. Hence for purposes of training the neural network, the first 60% (~6 years) of water storage data values were regarded as input and the corresponding GPS height residuals were treated as output, and vice versa. The subsequent 40% of the input and output pairs were then used to test (validate) the networks. A separate network was developed for each location.

### 3. Artificial Neural Networks

In this paper, artificial neural network (ANN) method is used to develop site-specific *GPS-soil moisture* models. Artificial neural networks are practical information processing systems that provide methods for “learning” functions from observations. An ANN roughly replicates the behaviour of the organic brain by emulating the operations and connectivity of biological neurons. This emulation, of course, is done in a mathematical form that is greatly simplified from the biological prototype. The advantage of ANNs in science and practical (engineering) applications lies in their ability to learn and capture information from data that describe the behaviour or pattern in a real system (Bose and Liang, 1998; Hammerstrom, 1993; Haykin, 1999; Wasserman, 1993). In contrast to standard statistical test procedures, they can correlate, both spatially and temporally, one or multiple input variables (e.g. soil moisture+other variables) with a single output signal (e.g. long-term GPS time series) through the interconnected neurons with trainable weights and bias signals.

A trained neural network can provide a much faster simulation for predicting long-term patterns than standard procedures since its calculation requires no computational iteration. Moreover, they often work well even when the data sets contain noise and measurement errors. Because of these important properties, ANNs can be applied to networks where large volumes of data are analysed. Examples of some relatively new applications in GPS and geodetic science include coordinate transformations, regional computation of GPS derived Total Electron Content (TEC), modelling of geodetic deformations such as prediction of subsidence due to underground mining, map-matching in GPS navigation and location management in mobile networks (see e.g. Ambrozic and Turk, 2003; Leandro, 2004; McKinnell, 2002; Winter and Taylor, 2003; Zaletnyik, 2005). In the context of this study, an ANN procedure is applied to represent temporal site-specific relationship be-

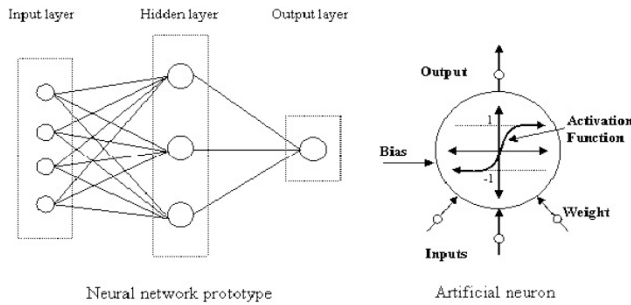


Fig. 2. Typical neural network prototype (left) and artificial neuron or node (right).

tween the soil moisture variations and GPS height residuals. There is no need to fully define the intermediate relationships between the soil moisture data and GPS residuals—the ANN identifies these during the “learning process”.

Figure 2 shows the structure prototype of an ANN model composed of a sequence of layers that are classified as input, hidden and output layers. Each layer of the network consists of a set of one or more neurons (or nodes). Each neuron has an activation function, which can be continuous, linear, or nonlinear [i.e., nonlinear function that saturates at finite value arguments like sigmoidal and tanh]. The neurons in the input layer receive information from the input variables, processes this information, and sends output to the next layer of neurons in the network. Each neuron is connected to neurons in the preceding layer, from which it receives inputs, and to the neurons in the subsequent layer, to which it passes its output.

The network “training” procedure can be described as follows:

The input layer data are multiplied by initial trial weights and a bias is added to the product. This weighted sum is then transferred through either linear or nonlinear transfer functions to yield an output. Thus, the output of a neuron in any layer is determined by applying a linear or nonlinear transformation (the activation function) to the sum of the weighted inputs it receives from the neurons in the previous layer. This output then becomes the input for the following hidden layers and the procedure is continued till the output layer is reached. The difference between the network output and the target is used to determine an error function, and the error is propagated back (“backpropagation”) to update the weights and biases using an optimization technique like the gradient descent which strives to minimise the error. The entire procedure is repeated for a number of epochs till the desired accuracy in outputs is achieved (“training”). Once the network is trained it can be used to validate against new data using the trained weights and biases. During validation, the model output variables are directly calculated, without iteration, from the input variables and the trained weights and biases, thus taking much less computation time.

Mathematically, the training process is similar to approximating a multivariable function (GPS height residuals),  $g(X)$ , by another function  $G(W, X)$ , where  $X = [x_1, x_2, \dots, x_n]$  is the input vector (hydrology signals), and  $W = [w_1, w_2, \dots, w_n]$  the coefficient or weight vec-

tor. The training task is to find the weight vector  $W$  that provides the best possible approximation of the function  $g(X)$  based on the training input  $[X]$ . By using the gradient descent method, weight changes move the weights in the direction where the error declines most quickly. Training is carried out by assigning random initial weights to each of the neurons (usually between 0.1 and 1.0) and then presenting sets of known input and *target (output)* values to the network. The network estimates the output value from the inputs, compares the model predicted output to the target value, and then adjusts the weights in order to reduce the mean squared difference between the network output and the target values. The complete input-output sets are run through the network for several iterations (or epochs) until either the mean square error is reduced to a given level or reaches a minimum, or until the network has been trained for a given number of iterations.

The structure of the ANN model is decided upon through the selection of the number of hidden layers and the specification of the number of nodes per layer. The latter is usually achieved by trial-and-error, as was the case in this study. A simple structure based on one hidden layer ( $H$ ) and one output layer ( $O$ ) was found to be optimal for forecasting both the residuals and water storage. Log-sigmoidal and tan-sigmoidal transfer functions were used for the hidden and output layers, respectively, and the input-output decks were scaled to a  $[-1, 1]$  range. A two-way analysis was done for each location. The first was done with an input deck ( $I$ ) of soil moisture and an output deck ( $O$ ) of GPS residuals, herein later referred to as the *water-to-GPS* model. The second analysis was in reverse order, hence referred to as *GPS-to-water* model. The performance functions were the root mean square errors ‘RMSE’ and the correlation coefficients ‘R’. All the networks were trained for 500 epochs over one data set ( $\sim 6$  years) and evaluated over another data set ( $\sim 3$  years) not included in the training. The trained weights and biases were retained for validation.

#### 4. Results and Discussion

All computations of the neural network model were done within the MATLAB 7.0 (Mathworks, 2004). There was fairly good correlation from training and validation. The final structure of the networks was typically of the type I30H1O1 (30 Input nodes 1 Hidden node 1 Output node). This architecture seem to be both time efficient and adequate to handle the rather small amount of data in the current study. Other parameters that were adjusted in order to achieve accurate results were the goal value of the error function of the network during training, calculated by the Mean Square Error (MSE), the learning rate of the network, and the number of epochs or network feeds. The need for adjustment of these parameters lies in the danger of over-training, an effect that is analogous to over-fitting a polynomial function. An interesting property of neural network models is that they often work well even when the training data sets contain noise and measurement errors (Hammerstrom, 1993). It is therefore prudent to select models that are more likely to generalize than over-parametrized models that may become tuned to noise within the training data.

Using the *water-to-GPS* model, ANN estimates of the

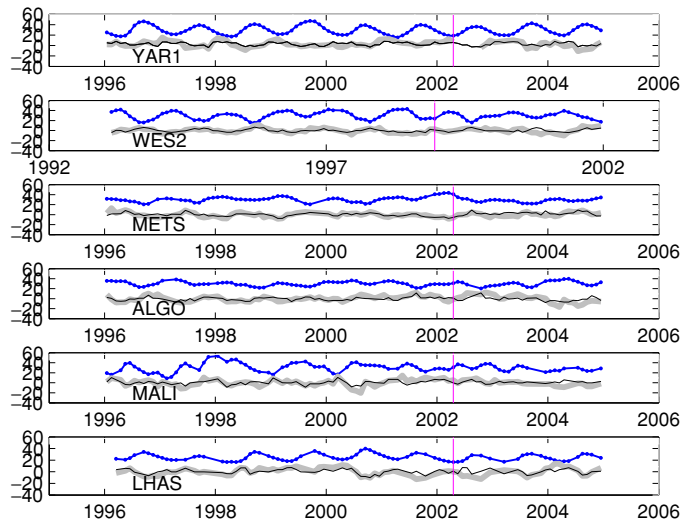


Fig. 3. Soil water variations (blue, cm), GPS height residuals (thick grey, mm) and ANN model (thin black, mm) at selected GPS sites. The time series are separated by vertical solid line for training and validation of the ANN model. The water residuals are offset by 30 cm for clarity.

GPS height residuals were derived with the soil moisture as input to the model. In general satisfactory results were produced with the correlation coefficients ‘R’ equal to  $\sim 0.6$  and RMSE of about 15% of residual range (residual range = 20.6 to 35.6 mm, mean  $\sim 27.4$  mm). In Fig. 3, we superimpose the ANN model (thin black lines) onto GPS height residuals (thick grey lines) for the selected sites. The ANN model tracks the observed residuals reasonably well at all of these sites, demonstrating that water loading may indeed be the cause of some of the seasonal variability which is observed in long term GPS time series (see also van Dam *et al.*, 2001). The water storage plots (bold blue lines) are included for visual comparison with the GPS height residuals. The mean RMSE for the *GPS-to-water* model was 4.27 cm, minimum and maximum residual ranges were respectively 19.3 cm and 44.8 cm (mean residual range was 28.2 cm).

The ANN model was able to successfully transform observed input vectors into reasonably accurate estimation of the outputs for all the locations. Its adequacy is demonstrated by the quality of the estimates as discussed above. The results demonstrate that there is an underlying significant relationship between the GPS height residuals and the CPC LDAS soil water storage variations at the sites that have been analyzed (in spite of the fact that the soil moisture data are numerically modelled (rather than observed) and thus have a fair degree of uncertainty). Other potential error sources contributing to the height residuals would include bedrock thermal expansion, errors in phase center variation models, errors in orbital modelling, the effects of employing different analysis strategies, and the unmodelled troposphere effects (see e.g. Dong *et al.*, 2002).

The neural methodology can be applied to other geodetic time series and hydrological models provided site-specific training and validation are carried out appropriately. It can be better tested if actual in-situ soil moisture data are observed over time with additional near-surface meteorological parameters.

**Acknowledgments.** The author is grateful to the IERS Global Geophysical Fluids Center (GGFC) and the Scripps Orbit and Permanent Array Center (SOPAC) for respectively sharing their data.

## References

- Ambrozic, T. and G. Turk, Prediction of subsidence due to underground mining by artificial neural networks, *Computers & Geosciences*, **29**(5), 627–637, 2003.
- Bose, N. and P. Liang, *Neural Network Fundamentals with Graphs, Algorithms and Applications*, Tata McGraw-Hill Publication, 478 pp, 1998.
- Dong, D., P. Fang, Y. Bock, M. K. Cheng, and S. Miyazaki, Anatomy of apparent seasonal variations from GPS-derived site position time series, *J. Geophys. Res.*, **107**(B4), doi:10.1029/2001JB000573, 2002.
- Hammerstrom, D., Working with neural networks, *IEEE Spectrum*, July, 46–53, 1993.
- Haykin, S., *Neural Network: A Comprehensive Foundation*, Prentice Hall—Upper Saddle River, New Jersey, 1999.
- Leandro, R. F., A new technique to TEC regional modelling using a neural network, Personal Communication, 9 pp, 2004.
- Mathworks, Inc., MATLAB 7.0 (Release 14), www.mathworks.com, 2004.
- McKinnell, L., A Neural Network based Ionospheric Model for the Bottomside electron density profile over Grahamstown, South Africa, Ph.D. Thesis, Rhodes University, 2002.
- Rodell, M., *et al.*, The global land data assimilation system, *Bull. Am. Meteorol. Soc.*, **85**(3), 381–394, 2004.
- Takle, E. S., M. Jha, and C. J. Anderson, Hydrological cycle in the upper Mississippi River basin: 20th century simulations by multiple GCMs, *Geophys. Res. Lett.*, **32**, L18407, doi:10.1029/2005GL023630, 2005.
- van Dam, T. M., J. M. Wahr, P. C. D. Milly, A. B. Shmakin, G. Blewitt, D. Lavallee, and K. M. Larson, Crustal displacements due to continental water loading, *Geophys. Res. Lett.*, **28**, 651–654, 2001.
- Wasserman, P., *Advanced Methods in Neural Computing*, Van Nostrand Reinhold, New York, 255 pp, 1993.
- Wilson, C. R. and J. Chen, The Hydrology Bureau of the Global Geophysical Fluids Center, in IERS Technical Note No. 30, edited by B. Richter *et al.*, Proceedings of IERS Workshop on Combination Research and Global Geophysical Fluids, ISBN 3-89888-877-0, 153–156, 2003.
- Winter, M. and G. Taylor, Modular neural networks for map-matched GPS positioning, Fourth International Conference on Web Information Systems Engineering Workshops (WISEW’03), 106–111, 2003.
- Zaletnyik, P., Coordinate transformation with neural networks and with polynomials in Hungary, Personal Communication, 9 pp, 2005.