INTRODUCTION

River flow forecasting deals with the estimation of future stages or flows at a single or multiple sites of a river system. Daily river flow forecasts are essential for water resources planning and management including potential water supply for domestic needs, irrigation scheduling, hydropower generation and regulating flows through reservoirs and barrages, whereas hourly water level forecasts are essential for issuing flood warning to mitigate natural disaster by undertaking appropriate evacuation and rehabilitation plans. The necessity of accurate and reliable river flow forecasts is increasingly being felt with increasing demands on water resources due to economic development and demographic expansion.

Classical time series models such as auto regressive integrated moving average (ARIMA) are widely used for hydrological time series forecasting as they are accepted as a standard representation of a stochastic time series (Maier and Dandy, 1997). However, these models are basically linear models which make use of classical statistics to analyse the historical data. They assume that data are stationary and have limited ability to represent non-stationary and nonlinear dynamics, if any, between the input and output variables. However, most hydrological processes exhibit high nonlinearity between the input and output variables, and hence, linear time series models may not always perform well. In the past, owing to the difficulties associated with nonlinear model structure identification and parameter estimation, the usual practice was to assume linearity or piecewise linearity in modelling nonlinear hydrological processes (Hsu et al., 1995).

To overcome the limitations of classical time series models, a wide variety of rainfall-runoff models have been developed and applied for river flow forecasting ranging from complex physically based to simple black box models. Black box models in the form of neural networks (NNs) have gained momentum in last few decades for river flow forecasting and have been accepted as a good alternative to physically based and conceptual models (ASCE, 2000a,b). The ability of NNs in extracting the complex nonlinear relationship between inputs and outputs without explicitly accounting for the physical processes has increased the number of applications in river flow forecasting. Most importantly NN models need limited inputs such as water level, discharge, rainfall or sometimes only a single input, whereas physically based models require several additional parameters which are difficult to measure because of temporal and spatial variability. This approach has also been criticised for making models overly complex which lead to problems of over parameterisation and equifinality (Beven, 2006) causing large prediction uncertainty.

NN models are computationally fast and efficient, which makes them a very suitable tool for river flow forecasting. Disadvantages related to NN models include interpretation of the NN structure ("black box") and their extrapolation capacity (Minns and Hall, 1996). Recently, researchers have been exploring different pre-processing approaches for inclusion of additional hydrological knowledge as input to NN models to improve the hydrological representation and generalisation (Corzo and Solomatine, 2007a,b). The quantification of the uncertainty associated with the results provided by NN models is essential for their confident and reliable use in practice as operational river flow management strongly depends on the accuracy and reliability of flow forecasts. Uncertainty assessment in real-time river flow forecasts is an important means for increasing the applicability of hydrological forecasts to mitigate the natural disaster. A probability distribution function of a predictand, or an ensemble of possible realisations of a predictand, enables users to make decisions that take the risk explicitly into account (Georgakakos and Krzysztofowicz, 2001).

The reliability of the model estimated discharge is affected by three sources of uncertainties (Bates and Townley, 1988): data uncertainty (quality and representativeness of data), model structure uncertainty (ability of the model to describe the catchment's response), and parameter uncertainty (adequate values of model parameters). It is difficult to assess the data uncertainty because the magnitude of data errors is often unknown and any attempt to model these deviations is ultimately based on a guess. Model structure uncertainty depends on the choice of the physical or statistical model. It cannot be reduced by increasing information (e.g. the sample size), but only by increasing the knowledge of the process, and by adopting more complex models. The parametric uncertainty is caused by variation in calibration datasets or sampling variability of a particular model structure. This uncertainty can be included in the predicted variables and presented in a stochastic or probabilistic framework. This provides for the range of possibilities and the likelihood that a given prediction will occur. The quantification of these uncertainties is important for practical decision making. When sufficiently large sets of examples (training patterns) are available, the sampling variability in weights can be approximated by bootstraps (Stone, 1974). Bootstrap technique (Efron and Tibshirani, 1993) has been successfully used in hydrological modelling. The bootstrap is a computational procedure that uses intensive resampling with replacement, in order to reduce uncertainty. In addition, it is the simplest approach since it does not require the complex computations of derivatives and Hessian-matrix inversion involved in linear methods or the Monte Carlo solutions of the integrals involved in the Bayesian approach (Dybowski and Roberts, 2000).

In spite of suitable flexibility of NNs in modelling hydrologic time series such as runoff (e.g. Hsu et al., 1995; Zhang and Dong, 2001) they fail to perform when signal fluctuations are non-stationary and hydrologic processes operate under a large range of scales varying from one day to several decades. In such a situation, NNs may not be able to cope with non-stationary data if pre-processing of the input and/or output data is not performed (Cannas et al., 2006). Recently, there has been increasing interest in the use of wavelet analysis in a wide range of fields related to water resources. The wavelet transformation (WT) provides information about the variation in a time series at different scales and locations. WT has positive effects on NN modelling performance, and the WT based NN models have performed very well in simulation and forecasting (Nourani et al., 2009; Adamowski and Sun, 2010).

Although several studies indicate that the data-driven models have proven to be potentially useful tools in hydrological modelling, two of the main issues that needs to be further explored before these models gain wider acceptability by researchers and practitioners are: (i) addressing the issues of nonlinearity and non-stationarity in NN modelling in a more effective way by considering them on a single platform, and (ii) identifying effective ways for assessing uncertainty in data-driven models, which in turn contributes to improving the reliability of such models (iii) developing a procedure for selection of appropriate model for a specific dataset profile.

The present study makes an attempt to address these issues to improve the credibility of the data-driven models among researchers and practitioners. Earlier studies on NN modelling have not given much emphasis on ensemble modelling and uncertainty assessment in hydrologic forecasting for different lead times. The potential of bootstrap resampling technique can be explored for uncertainty assessment and ensemble modelling whereas wavelet transformation can be used to handle the non-stationarity in the data by extracting the useful information from the dataset. The combined strength of wavelet transformation and bootstrap method for hydrologic forecasting has not been explored. Also, the uncertainty assessment for longer lead times has not been carried out. There is a need to develop a procedure for selection of an appropriate model based on the type of data profile. It is assessed from earlier review that the SOM technique can be used to develop a procedure to select a particular model for a particular dataset profile.

Keeping the above in view, the present research is undertaken with the following objectives:

1. To develop neural network (NN), bootstrap based NN (BNN), wavelet based NN (WNN) and wavelet-bootstrap based NN (WBNN) models for forecasting daily discharge and hourly water level at Naraj gauging site in Mahanadi river basin.

2. To compare the performance of the developed models for daily discharge and hourly water level forecasting.

3. To assess the uncertainty associated with daily discharge and hourly water level forecasts using BNN and WBNN models.