國立臺灣大學生物資源暨農學院生物環境系統工程學研究所

博士論文

Graduate Institute of Bioenvironmental Systems Engineering

College of Bioresources and Agriculture

National Taiwan University

Doctoral Dissertation

創建新穎動態類神經網路於水文環境系統

Development of Novel Dynamic Artificial Neural Networks for
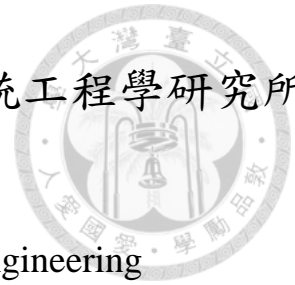
Hydro-Environmental Systems

陳品安

Pin-An Chen

指導教授: 張斐章 博士

Advisor: Fi-John Chang, Ph.D.

中華民國 103 年 7 月

July 2014

誌謝

摘要

水為人類賴以為生的重要資源，水與環境在複雜的交互作用下形成的水文環境系統孕育了各種生命，使得人類社會得以繁榮發展。然而近年來由於氣象、水文、地文及人為開發等各種錯綜複雜之因素影響，水文環境系統逐漸面臨失衡，加上全球氣候變遷現象日趨顯著，導致突發性極端水文事件發生之頻率提高，都市下游河川水質劣化。為能進一步分析掌握其變化，本論文發展新穎系統化動態類神經網路與相關之分析方法應用於水文環境中兩大重要議題：水文量與水質之推估。動態回饋式類神經網路(Recurrent neural networks, RNNs)具內部回饋連結，對於複雜且具有回饋特性的水文環境系統之模擬有較高之精確性，因此近年來受到相當地重視與大量地應用。本論文第一部分發展並推導多時刻強化型即時回饋線上學習演算法於回饋式類神經網路 (Reinforced real-time recurrent learning algorithm for RNNs, R-RTRL NN)。此演算法可充分利用最新的觀測值與網路過去之預測值，不斷地遞迴修正網路權重參數，改善多時刻預報之可靠度與精確度。而為了驗證此演算法之可靠性與效能，本研究針對著名的混沌時間序列與石門水庫颱洪時期之入流量進行二、四至六時刻之多階段預報。此外亦選用了三個常用之類神經網路模式(兩個動態與一個靜態類神經網路)進行效能比較。結果顯示發展之 R-RTRL 類神經網路較其他類神經網路於多時刻混沌時間序列與水庫入流量預報整體表現優異，且可有效減緩線上學習法於多時刻預報網路權重參數調整延遲之問題。本論文第二部分則發展一系統化動態類神經網路模擬架構 (Systematical dynamic-neural modeling, SDM)。SDM 主要包含 Gamma test 與非線性自回歸與外部輸入類神經網路(Nonlinear autoregressive with exogenous input, NARX)。Gamma test 可有效且快速地篩選影響目標變數最為顯著之輸入因子組合；NARX 類神經網路則具輸出層至輸入層之回饋連結，有優異的時間與空間推估能力。本論文運用發展之 SDM 於都市防洪抽水站多時刻前池水位預報，並探討回饋連結於不同模式
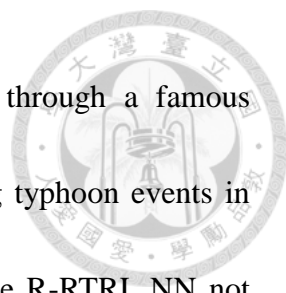
情境下之貢獻。而建構之水位預報模式之準確度與穩定性皆相當高，可有效率且準確地預測洪汛時期臺北市玉成抽水站之前池水位。此外，SDM 亦可應用於區域地下水砷濃度與河川總磷濃度之時空推估，精確度與可靠度皆較傳統倒傳遞類神經網路(BPNN)高，並有效解決進行區域水質推估常面臨之問題，如：影響因子組合之選取、資料稀少、過度描述與推估效果不佳等。而藉由 NARX 類神經網路之回饋連結與輸入因子資訊，可進一步將標的水質序列間隔較長之推估時間尺度轉換為與輸入因子相同且較短之觀測頻率，提供額外資訊進行水質狀況評估。整體而言，本論文創建之新穎動態類神經網路與分析方法：R-RTRL NN 與 SDM 皆具廣泛之應用性，十分適合分析處理或推估水文環境中水文量與水質變化等重要議題，提供政府有關單位於水庫、都市河川與地下水等經營管理之參考資訊。

關鍵詞: 回饋式類神經網路、強化型即時回饋線上學習演算法(R-RTRL)、Gamma test、非線性自回歸與外部輸入類神經網路(NARX-NN)、多時刻時間序列預測、都市防洪、砷、總磷。

Abstract

Water is a precious and scarce resource on the Earth and can be utilized by human beings. Due to the complex interaction between hydrological, meteorological, geographical factors and human activities with climate change effects, the hydro-environment that we live by is facing imbalanced conditions, such as intensive storms and typhoons with short durations and the degradation of the water quality in groundwater and urban rivers. Therefore, this dissertation is dedicated to the two main problems encountered in hydro-environmental systems: water quantity and water quality, and endeavors to develop novel dynamic artificial neural networks and modeling schemes to overcome problems for analyzing and estimating the dynamic variability of water quantity and water quality. Recurrent neural networks (RNNs) are computationally powerful nonlinear models that are capable of extracting dynamic behaviors from complex systems through internal recurrence and have attracted much attention for years. In the first part of this dissertation, a multi-step-ahead (MSA) reinforced real-time recurrent learning algorithm for RNNs (R-RTRL NN) is developed for adjusting connection weights by incorporating the latest observed values and model outputs into the online training process, and the sequential formulation of the R-RTRL NN is derived. To demonstrate its reliability and effectiveness, the proposed R-RTRL

NN is implemented to make 2-, 4- and 6-step-ahead forecasts through a famous benchmark chaotic time series and a reservoir inflow series during typhoon events in North Taiwan. Numerical and experimental results indicate that the R-RTRL NN not only achieves superior performance than the comparative networks but also significantly improves the precision of MSA forecasts with effective mitigation in time-lag problems for both chaotic time series and reservoir inflow case during typhoon events. In the second part of the dissertation, the systematical dynamic-neural modeling (SDM) scheme that consists of the Gamma test for input factor selection and the nonlinear autoregressive with exogenous input (NARX) network for spatio-temporal estimation is proposed. The SDM is then applied to urban flood control to explore the contribution of recurrent connections and provide reliable results for forecasting the floodwater storage pond (FSP) water level in the Yu-Cheng pumping station. And the SDM is further utilized to estimate the regional arsenic (As) and total phosphate (TP) concentrations in groundwater and river systems, respectively. Results demonstrate that the SDM satisfactorily overcomes the difficulty raised by traditional methods in estimating the temporal and spatial variability of water quality parameters, such as identification of key input factors, data scarcity issue, model over-fitting problem and poor estimation performance. In addition, the SDM bears the ability to reconstruct the time series of the estimated water quality parameter from the original monitoring scale

to a shorter monitoring scale through the recurrent connections of the NARX network.

In summary, the two developed novel techniques in learning algorithm and modeling scheme, the R-RTRL NN and SDM, have broad applicability and are suitable to deal with water quantity and water quality issues in hydro-environmental systems, which beneficially provides useful information to water authorities for the management of reservoir operation, river basin, urban flood control and groundwater contamination.

Contents

Figure contents

Table contents

XIII

1. Introduction

1.1 Motivation

Water resources are precious and can be utilized by human beings. The scarce sources of water on the Earth have gained more and more attention for decades. Water resources support the activities of agricultural, industrial, household, recreational, environmental and various sectors. Therefore, we all live in hydro-environmental systems and have complex interactions with these systems. In recent years, due to the comprehensive interactions between hydrological, meteorological, geographical factors and human activities with climate change effects, the hydro-environment that we live by is facing imbalanced conditions. The intensive storms and typhoons with short durations and the degradation of the water quality in groundwater and urban rivers are becoming serious but common disasters. To tackle such challenges, this dissertation focuses on two main problems that have occurred in hydro-environmental systems: water quantity and water quality issues, and endeavors to develop novel dynamic artificial neural networks and modeling schemes to solve these water-related problems.

Artificial neural networks (ANNs) are biologically motivated methods in which large numbers of neurons communicate with each other through weighted connections, and have the ability to approximate nonlinear functions for modeling time series. ANNs

are also considered as an alternative computational approach to modeling

physical-based problems. In the last decades, ANNs have been widely applied with

success to various water resources problems, such as rainfall-runoff modeling (Antar et

al., 2006; Chang et al., 2007; Chen and Yu, 2007; Chang et al., 2013; Chang et al., 2014;

Yang et al., 1999), flood control (Chang et al., 2008), evaporation estimation (Chang

and Sun, 2013; Chang et al., 2013), reservoir operation (Chang et al., 2010; Chang and

Wang, 2013; Wang et al., 2010), groundwater level prediction (Krishna, et al., 2008 ;

Nikolos, et al., 2008), and water quality estimation (Khalil et al., 2011; McNamara et al.,

2008; Sahoo et al., 2006). However, static neural networks may not be able to establish

reliable nonlinear models for predicting dynamical systems, especially for many

time-step-ahead forecasting or regional estimation. Recurrent neural networks (RNNs),

which belong to a class of dynamic ANNs, are powerful nonlinear tools capable of

extracting dynamic behaviors from complex systems through internal recurrence and

have attracted much attention for years (Assaad et al., 2005; Chang et al., 2012; Chiang

et al., 2010; Ma et al., 2008; Serpen and Xu, 2003). Nevertheless the batch training of

an RNN could be time consuming (Ahmad and Jie, 2002; Xie et al., 2006), and the

behaviors of recurrent connections in spatio-temporal estimation has not been fully

explored yet. Therefore, this dissertation develops a reinforced online-learning

algorithm for RNNs and explores the practical meaning as well as the importance of

recurrent connections of the NARX network through *topic 1: water quantity issues involving reservoir inflow forecasting and urban flood control;* and *topic 2: water quality issues involving the spatio-temporal estimation with respect to the arsenic (As) concentration in groundwater and the total phosphate (TP) concentration in a river basin.*

1.2 Research objectives

A. Topic 1: Water quantity issues for reservoir inflow forecasting and urban flood control

Accurate multi-step-ahead (MSA) forecasting is valuable and desired in many engineering problems, however it is a challenging task that is difficult to achieve. A common approach for improving the accuracy of MSA forecasting is to update network parameters through online learning techniques. Online learning is a supervised machine-learning framework, which adopts the latest information to adjust model parameters for a better mapping between instances and true values in an arbitrary system. Because most observational disciplines tend to infer the properties of an uncertain system from the analysis of time-dependent data, the analytical techniques for extracting the meaningful characteristics of time series data have certain inherent limitations, which have been widely discussed (Brockwell and Davis, 1987; Jaeger and Haas, 2004). Owing to the continual receipt of true values for adjusting model

parameters, online learning algorithms have several practical and theoretical advantages such as memory-efficient implementation, runtime-efficient implementation and strong guarantees on performance, even in a highly variable data structure of time series (Shalev-Shwartz et al., 2004). Nevertheless, the main defect of online learning can be attributed to the requirement for continual true values. Engineering problems often require models to predict many time steps into the future without the availability of measurements in the horizon of interest. The lack of true values makes it difficult to make MSA forecasts. In addition, many studies indicated that it is not an adequate strategy to recursively adopt single-step-ahead predictions for many time steps into the future because the errors of MSA predictors will be accumulated based on the single-step-ahead predictor (Parlos et al., 2000; Yong et al., 2010). Such time-lag problems may cause significant degradation in performance when dealing with MSA forecasting for real-world applications. For the MSA streamflow forecasting during typhoon events, models with time-lag problems (i.e., the latest observed values are unavailable) cannot keep flow trails, especially in peak flows, as the forecasting step increases. To mitigate time-lag phenomena that occur in online learning algorithms, it is argued that whether iterative adjustments of model parameters in consideration of additional information, such as the latest true values and/or antecedent model outputs, would be beneficial to MSA forecasting.

The real-time recurrent learning (RTRL) algorithm, proposed by Williams and Zipser (1989), is an effective and efficient online learning algorithm for training recurrent networks, in which real-time adjustments are made to the synaptic weights of recurrent networks. Several studies demonstrated that the RTRL algorithm for RNNs is very effective in modeling the dynamics of complex processes for providing accurate predictions (Chang et al., 2002; Chang et al., 2012; Hirasawa et al., 2000; Li et al., 2002).
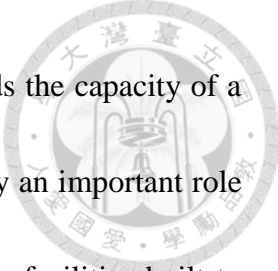
The first main goal of *topic 1* is to develop a reinforced RTRL algorithm for RNNs (R-RTRL NN) to mitigate time-lag effects for increasing the accuracy of MSA forecasting. The sequential formulation of the R-RTRL NN is derived, and its reliability and applicability are further demonstrated through two-step-ahead (2SA), four-step-ahead (4SA) and six-step-ahead (6SA) forecasting made for a famous benchmark chaotic time series and a reservoir inflow case in Taiwan. Comparative models consist of the original RTRL algorithm for RNNs (RTRL NN), the Elman neural network (Elman NN) (Elman, 1990; Liu and Wang, 2008; Liu et al., 2012) and the backpropagation neural network (BPNN, the most popular static ANN).

Urban flood control is a crucial and challenging task, particularly in developed cities. Urban floods are flashy in nature mainly due to severe thunderstorms and occur both on urbanized surfaces and in small urban creeks, which deliver mass water to cities.

On account of more impervious areas resulting from the rapid urbanization in metropolitan areas, less water infiltration has resulted in an increase in the flow rate and the amount of surface runoff over the last decades. Taiwan is located in the northwestern Pacific Ocean where subtropical air currents frequently introduce typhoons and convective rains. The urban flood hydrographs in Taiwan typically have large peak flows and fast-rising limbs in a matter of minutes, which could cause serious disasters. For example, Typhoon Nari brought massive rainfalls at an astonishing level of 500 mm/day on September 17th in 2001, which resulted in 27 deaths, inundations at some stations of the Taipei Metro System, and countless economic losses. The heavy rainfall event on June 12th in 2012 brought astonishing rainfalls with a cumulative amount of 54.1 mm/hr, which directly resulted in quick and wide surface flooding such that the transportation system collapsed in most of the southern Taipei City. It appears that floods cannot be prevented, but planning emergency measures through flood management might mitigate disastrous consequences.

In response to the flood threats to residents and property, the Taipei City Government has long-term endeavored to develop flood control-related infrastructures, such as increasing levee heights and enhancing sewerage systems, and therefore urban inundations have been significantly mitigated and controlled in recent years. As a result, the main threat to the city turns out to be the floodwater inside the levee system. A

surface inundation will inevitably take place if surface runoff exceeds the capacity of a

storm drainage system. To tackle this problem, pumping stations play an important role

in flood mitigation at metropolitan areas and are principal hydraulic facilities built to

manage internal stormwater flows at places under the condition that gravity drainage

cannot be achieved. The operation of a pumping station highly depends on the water

level information of its floodwater storage pond (FSP). Within the catchment of a

pumping station, surface runoff will drain to its FSP for storage and subsequent disposal

through gravity drainage. When the water level of the FSP reaches the start level of duty

pumps, the pumps will be activated according to operation rules for discharging the

stored floodwater into the nearby river of the pumping station. For floodwater control

management during heavy rainfall or typhoon events, it is imperative to construct an

efficient and accurate model to forecast many step-ahead FSP water levels by utilizing

the information of the current FSP water level and the rainfall measured at the

neighboring rainfall gauging stations of the pumping station. The proposed model is

expected to provide sufficient response time for warming up the pumps in advance for

enhancing secure pumping operations and urban flood control management.
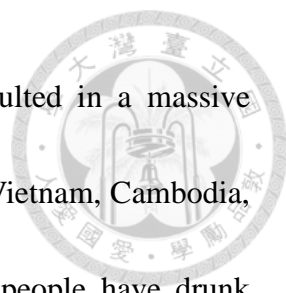
The greatest success in flood forecasting is commonly achieved on large rivers.

Nevertheless, flash urban floods associated with heavy thunderstorms in cities are often

very uncertain and are more difficult to predict due to complex dynamic phenomena
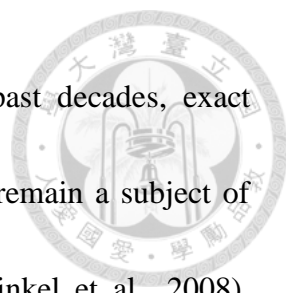
involved. Many studies demonstrated the predictability of streamflow through soft computation methods (Maity and Kumar, 2008) while only few papers investigated the prediction performance of inundation and/or sewerage systems in urban areas (Chiang et al., 2010). The second main goal of *topic 1* intends to investigate the reliability and accuracy of short-term (10- to 60-minute) forecasting models for the FSP of a sewer-pumping system in Taipei City. Multi-step-ahead FSP water level forecasting models for flood pumping control during heavy rainfall and/or typhoon events are tailored made through a static ANN (the BPNN) and two dynamic ANNs (the Elman NN; the NARX network). Consequently, the comparison results of these three ANN models are evaluated to identify the effectiveness of recurrent connections. The forecasting system is designed to anticipate the occurrence of flooding and to take measures necessary to reduce flood-induced losses. The study will give a boost to the efforts for urban flood disaster management and will strengthen the Taipei City Government with more proactive disaster preparedness.

B. Topic 2: Water quality issues for the spatio-temporal estimation with respect to the As concentration in groundwater and the TP concentration in a river basin

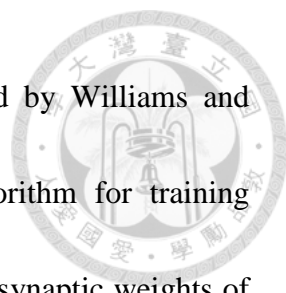The second topic of this dissertation focuses on water quality issues for which the stabilization and variation of concentrations are important tasks for preserving healthy human and hydro-environmental systems.

As contamination in groundwater has been reported and resulted in a massive

epidemic of As toxication in several countries such as Bangladesh, Vietnam, Cambodia,

China and Taiwan. It is estimated that approximately 57 million people have drunk

As-contaminated groundwater with concentrations exceeding the drinking water

standard recommended by the WHO (World Health Organization) (BGS-DPHE, 2001;

Chakraborti et al., 2010). As pollution affects not only crop productivity and water

quality but also the quality of water bodies, which threatens the health of animals and

human beings by way of food chains. Long-term exposure to As through drinking water

has been implicated in a variety of health concerns including cancers, cardiovascular

diseases, diabetes and neurological effects (National Research Council, 1999).

Blackfoot disease and cancers of the skin, bladder, lung and liver have been associated

with drinking As-contaminated groundwater (Chiou et al., 1997; Rahman, 1999).

As-contaminated groundwater is derived naturally from As-rich aquifer sediments, and

the geochemistry of As can be rather complex (Stollenwerk, 2003). Various

hydrogeological and biogeochemical factors affecting As concentration in groundwater

have been detected, such as sediment mineralogy, microbial oxidation or reduction of

As, groundwater recharge, groundwater flow paths (Ford et al., 2006; Wang et al., 2007

& 2011; Xie et al., 2012), and the presence of fractures in bedrock formations (Ayotte et

al., 2003; Liao et al., 2011). Even though the processes controlling the release of As into
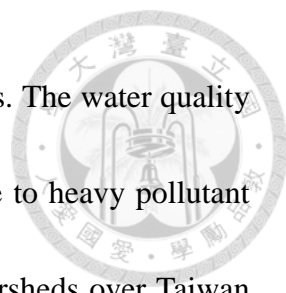
groundwater systems have been extensively discussed over the past decades, exact

chemical conditions and reactions leading to As mobilization still remain a subject of

intense debate (Goovaertset al., 2005; Polizzotto et al., 2006; Winkel et al., 2008).

Moreover, the high variability of As concentration can occur within a short distance

and/or in different depths of groundwater wells due to the diversity in geology and

geomorphology (Serre et al., 2003; Yu et al., 2003). Besides, the detection of As

contamination in groundwater by using graphite atomic absorption spectrophotometry

or inductively coupled plasma mass spectroscopy can be laborious and cost intensive.

Consequently, how to adequately estimate As concentrations in complex

hydro-geological systems is a crucial and challenging task.

The hyper-endemic blackfoot disease in the Yun-Lin County of Taiwan has been

verified to be associated with high As concentrations in groundwater (Chen et al., 1995;

Chiou et al., 1997). The residents have long-term exposed themselves to As through

various paths such as the ingestion of aquacultural and agricultural products, and thus

have dangerously posed carcinogenic risks to their health (Liu et al., 2008). Due to great

concern for the potential effects of As on human health, there is a growing need for

efficiently modeling the spatial distribution of As contamination in groundwater. One of

the popular modeling approaches in use is the multiple linear regression (MLR), this

approach, however, may fail to estimate the spatial distribution of As contamination due

to the great variability of As concentration and complex nonlinear processes involved in geology and geomorphology. Lately, using ANNs for the estimation of heavy metal concentration in groundwater has been attempted and gained a reasonably good degree of success (Chang et al., 2010; Cho, et al., 2011; Giri et al., 2011; Mondal et al., 2012; Purkait et al., 2008). The modeling results indicated that ANN techniques can produce higher estimation accuracy than conventional methods such as MLR. These studies were mostly dedicated to exploring the applicability of static ANNs, such as the BPNN, for building the relationship between As concentration in groundwater and hydro-geological parameters in As-affected areas. Nevertheless, the natural characteristics of hydrogeological processes are not only complex but also dynamic. The static neural networks might fail to establish reliable models for predicting the dynamical features, such that the delivered relationship might be simply the possible impacts of factors on temporal characteristics of local environments. Consequently, the comprehensive analysis of dynamic hydrogeological features and the estimation of As concentration variability over As-affected regions remains a great challenge that needs to be overcome.

The seasonal variation of steamflow in Taiwan is very high, where long-lasting low flows in drought seasons could dramatically increase the pollution levels in rivers. Pollution in the downstreams of rivers raises a major environmental issue because many

11
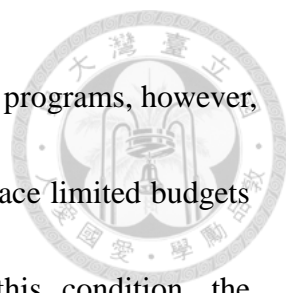
industrial facilities and large populated cities are located along rivers. The water quality

of the Dahan River in northern Taiwan has deteriorated rapidly due to heavy pollutant

loads from surrounding urban areas. Considering the scattered watersheds over Taiwan

and the high cost of field sampling, it is unlikely to obtain continuous water-quality time

series data with complete properties at all sampling locations. Alternatively, the Water

Quality Index (WQI) has been designed to assess the general conditions of water bodies

in rivers, lakes or reservoirs. The WQI is sensitive to light pollution, and therefore it is a

more suitable index adopted for water quality management. The WQI numerically

summarizes the information of multiple water quality parameters into a single value,

including dissolved oxygen (DO), coliform group, power of hydrogen (pH),

biochemical oxygen demand (BOD), ammonia nitrogen (NH_3-N), suspended solid (SS)

and total phosphate (TP). Except for TP (measured quarterly), the other water quality

parameters adopted in the WQI are measured monthly in Taiwan. Therefore, a monthly

WQI incorporated with TP would be more comprehensive and more beneficial to

short-term (monthly) water quality management.

TP, a combination of orthophosphate, polyphosphate and organic phosphate, is

regarded as an index used in representation of the phosphorus quantity in river water.

Phosphorus is an essential element for all life forms (Correll, 1998). When phosphorus

enters into a river, it is usually in the form of phosphate and can be transported from

upstream to downstream by flowing water. Excessive phosphorus is the most common

cause of eutrophication in freshwater lakes, reservoirs, streams, and headwaters of

estuarine systems. Orthophosphate chemicals are commonly used in agricultural

fertilizers, and thus enter surface water easily during rainfall periods. Many studies

reported that the phosphorus fertilizer form affects phosphorus loss to waterways

(Azevedo et al., 2013; Davis and Koop, 2006). Polyphosphate is a primary chemical

element added with considerable amount into detergents. Organic phosphates are

basically formed by biochemical procedures associated with excrement, kitchen waste,

water plants, etc. Phosphorus is one of the key elements essential for the growth of

plants and animals. Nevertheless, the anthropogenic nutrient enrichment of natural

water is of environmental importance as it can evoke declines in water quality, changes

in biotic population structures, and low dissolved oxygen concentrations in rivers

(Dodds et al., 2009; Austin et al., 1996). Excessive phosphorus has been shown to be a

main cause of eutrophication, for example, naturally-occurring nutrients in large

concentrations can often cause algae blooms (McDowell et al., 2010; Carpenter et al.,

1998).

Water quality models are useful tools for estimating the levels and risks of

chemical pollutants in a given water body (Duda, 1993). When building (or just

applying) a water quality model, it is necessary to have long and sufficient field data to

validate model applicability and reliability. Water quality monitoring programs, however, are expensive and time-consuming. Modeling practices commonly face limited budgets and time, and thus suffer a deficiency of field data. Under this condition, the implemented water quality models might fail to fit known hypotheses and/or assumptions or cause difficulties in making estimations within an acceptable range of errors or uncertainty. With the development of model theory and the fast-updating computer techniques, many artificial intelligent techniques have been developed with various analytical algorithms to overcome data scarcity issues and simultaneously increase model reliability.

The NARX network (Lin et al., 1996), a sub-class of RNNs, is suitable to build long-term temporal input-output patterns (Menezes Jr. and Barreto, 2008). The NARX network has been demonstrated to perform well in several nonlinear systems, such as waste water treatment plants (Su and McAvoy, 1991; Su et al., 1992) and time series forecasting (Shen and Chang, 2013). However, the dynamic feature and feasibility of the recurrent connections in the NARX network as a nonlinear tool for water quality time series modeling under limited data sets has not been fully explored yet. Therefore, *topic 2* will explore the practical meaning and importance of recurrent connections in the NARX network when dealing with spatio-temporal water quality estimation problems.

In *topic 2*, a systematical dynamic-neural modeling (SDM) scheme incorporated with a dynamic neural network and advanced statistical methods is developed for building spatio-temporal estimation models for (1) As concentration at decommissioned wells based on the easily-measured water quality parameters at nearby functioning wells to offer an applicable and useful reference to decision makers for dealing with groundwater management and preventing residents from drinking or using toxic groundwater; and (2) TP concentrations at seven sites along the Dahan River in a quarterly scale based on easily-measured water quality parameters. In addition, TP concentration data are reconstructed in a monthly scale through a process that adopts the dynamical neural architecture of the constructed NARX network, and thus the reconstructed monthly data can be used to produce the monthly WQI for short-term hydro-environmental management.

1.3 Dissertation layout

The framework of this dissertation is shown in Fig. 1.1. The novel learning algorithm (R-RTRL) and the modeling procedure (SDM) are developed to deal with two main topics in hydro-environmental systems: water quantity and water quality. First, The R-RTRL NN can repeatedly adjust model parameters through the reinforced process with the current information including the latest observed values and model outputs to enhance the reliability and the forecast accuracy of the proposed method. To

demonstrate its reliability and effectiveness, the proposed R-RTRL NN is implemented

to make 2-, 4- and 6-step-ahead forecasts in a famous benchmark chaotic time series

and a reservoir flood inflow series in North Taiwan. Second, the SDM which consists of

the Gamma test (GT) for input factor selection and the NARX network for

spatio-temporal estimation is proposed. The SDM is then applied to 1) urban flood

control problems to explore the contribution of recurrent connections and 2) the

spatio-temporal estimation of As and TP concentrations to provide useful information to

water authorities for dealing with groundwater and river basin management.
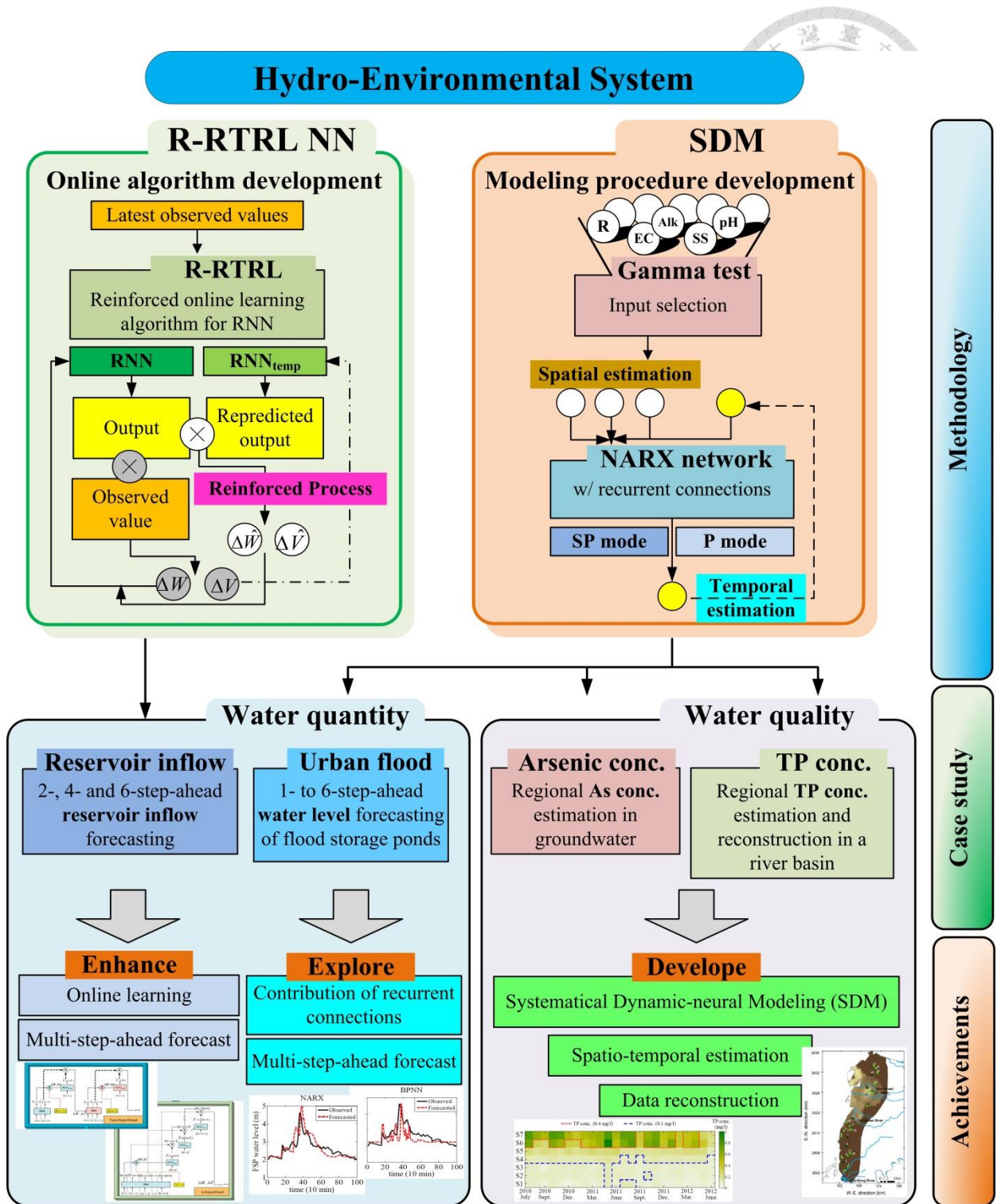
Fig. 1.1 Framework of the dissertation
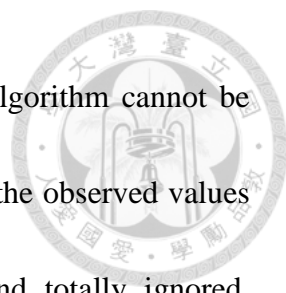
2. Methodology

2.1 MSA R-RTRL algorithm for RNNs

A. Rationale of MSA online learning algorithm

Two common strategies for MSA forecasting are the iterated prediction and the direct prediction. For n-step-ahead (nSA) prediction, the iterated method tackles the issue by iterating n times a one-step-ahead prediction whereas the direct method trains the model by conducting a direct forecast at time $t+n$. The debate on the superiority between these two methods still remains open; nevertheless both methods possess a common feature: the visibility of stochastic dependencies between future values becomes relatively vague as the time of prediction horizon increases, consequently the reliability and accuracy of predictions decreases. A possible way to remedy this shortcoming is to implement online learning techniques for repeatedly adjusting model parameters with the most current information including the latest true (observed) values and model outputs. An online learning algorithm proceeds in a sequence of trials through receiving an instance and making a prediction in each online-learning round to improve model performance.

The original RTRL algorithm, an online learning algorithm, was derived for one-step-ahead forecasting from the fact that real-time adjustments are made to the synaptic weights of a fully connected recurrent neural network (Williams and Zipser,

1989). For *n*SA forecasting, the weight adjustment of the RTRL algorithm cannot be conducted until obtaining the observed value at time *t+n*, in which the observed values and model outputs during time *t+1* and *t+n-1* are worthless and totally ignored. Therefore, the effectiveness of the original RTRL algorithm decreases considerably when time step *n* increases, which implies time lags occur in the weight adjustment process.

This dissertation proposes a novel reinforced RTRL algorithm based on RNN infrastructures (R-RTRL NN) for MSA forecasting through incorporating the latest antecedent forecasted and observed values into consecutive temporary networks for weight adjustments in the learning process. In other words, the R-RTRL algorithm repeatedly updates the synaptic weights by utilizing the most current obtainable information. The applicability and effectiveness of the R-RTRL NN is further investigated in Section 3.

The upper diagram of Fig. 2.1 shows the weight adjustment procedure of the R-RTRL algorithm for *2SA* forecasting (Chang et al., 2012). At time *t+2*, the weights are adjusted by the differences between observed and forecasted values. A reinforced process is introduced: the RNN_{temp} with adjusted weights can be used to produce a temp output $\hat{z}(t+3, \, 1)$ at time *t+1*, and the error between the temp output and forecasted output at time *t+1* can then be utilized to reinforce the weight adjustments, $\Delta \hat{W}_1(t)$ and

19
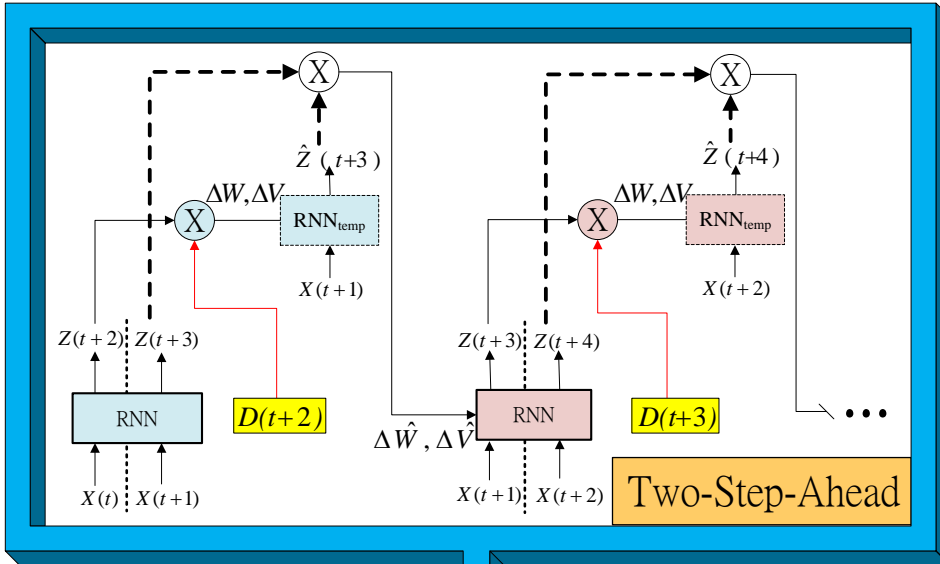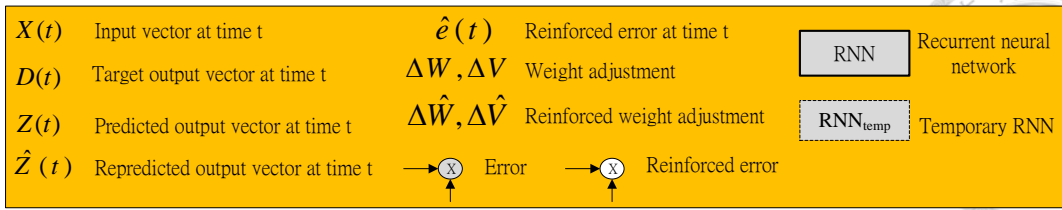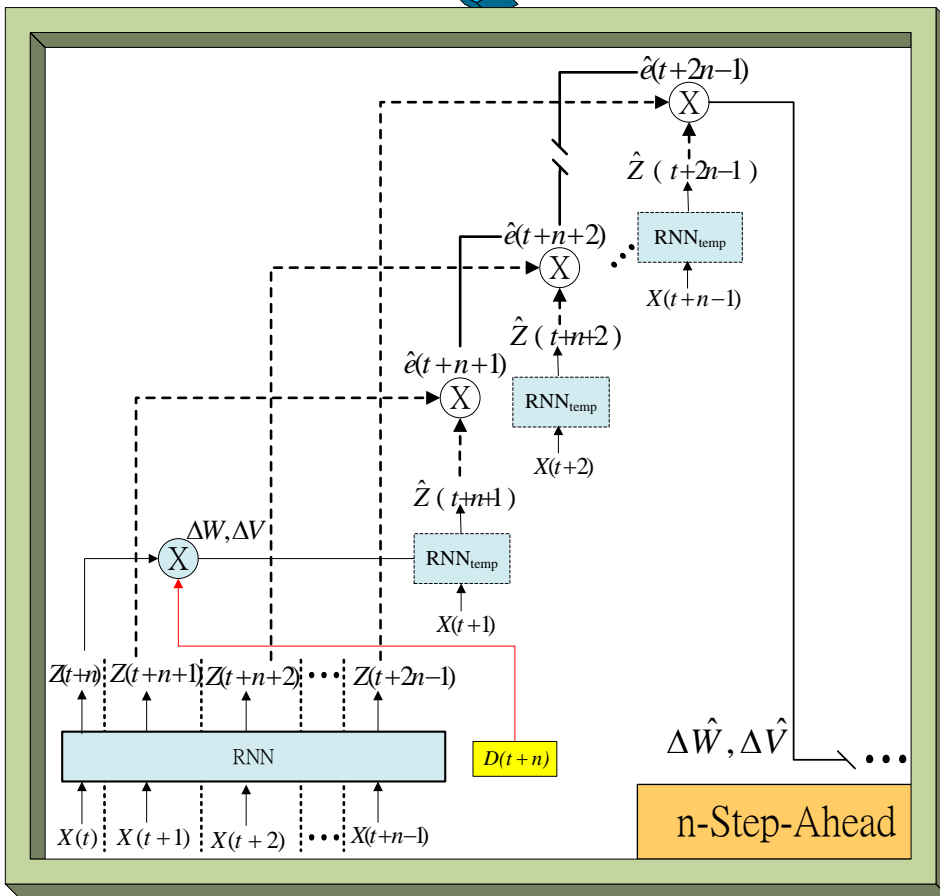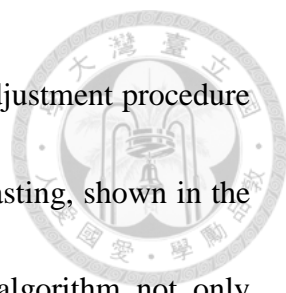
Fig. 2.1 Weight adjustment procedure for the n-step-ahead R-RTRL NN.

$\Delta \hat{V}_1(t+1)$. As this reinforced process repeats *n-1* times, the weight adjustment procedure can be extended to a general procedure for *n*SA ($n \geq 2$, $n \in \mathrm{N}$) forecasting, shown in the lower diagram of Fig. 2.1. In summary, the proposed R-RTRL algorithm not only adequately utilizes the up-to-date information of the observed values and their corresponding model outputs but also strengthens the usefulness of the latest observed values by the reinforced process to mitigate the time-lag phenomenon for MSA forecasting. The detailed sequential formulation of the R-RTRL algorithm is described as follows.

B. Deriving the MSA R-RTRL algorithm

Fig. 2.2 shows the MSA RNN architecture incorporated with the R-RTRL algorithm, in which there are M external inputs and one output. Let $X(t)$ denote the M × 1 input vector at discrete time *t*, $Y(t+1)$ denote the corresponding N × 1 vector at time *t+1* in the processing layer, and $Z(t + n)$ denote the corresponding output value for *n*SA ($n \geq 2$, $n \in \mathrm{N}$) forecasting.

The $X(t)$ and $Y(t)$ are concatenated to form the (M + N) × 1 vector $U(t)$, whose i^{th} element is denoted by $\mu_i(t)$. Let A denote the set of indices *i* for which $x_i(t)$ is an external input, and B denote the set of indices *i* for which $y_i(t)$ is the output of the processing layer. Thus, vector $\mu_i(t)$ can be represented as Eq. (1).

21

Output **Z***(t+n)*

Output Layer

Y*(t+1)*

Processing Layer

$W_{Nx(M+N)}$

Concatenated Input Layer

Y(t)

Input vector

X*(t)*

unit time delay
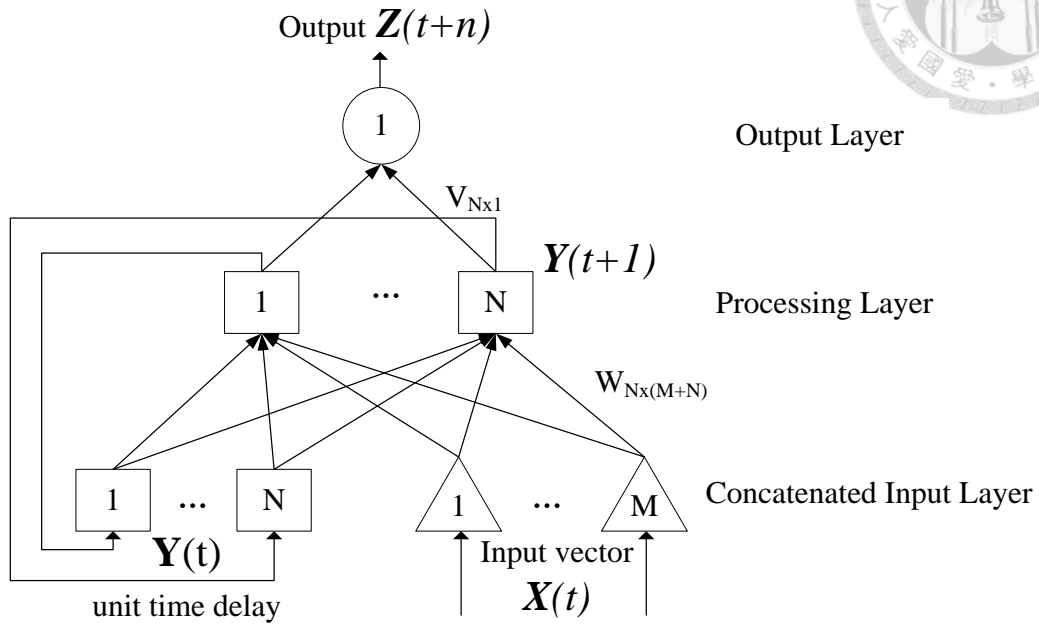
Fig. 2.2 Architecture of the multi-step-ahead RNN with the R-RTRL online learning algorithm.

$$\mu_i(t) = \begin{cases} x_i(t) & if \ i \in A \\ y_i(t) & if \ i \in B \end{cases} \tag{1}$$

W and V denote the weight matrices in the processing layer and the output layer, respectively. $W \leftrightarrow w_{ji}$ and $V \leftrightarrow v_j$ are of matrix forms. The output of neuron j in the processing layer that presents the transformation of information from the concatenated layer through nonlinear system f is given by:

$$y_j(t+1) = f\left(net_j(t+1)\right) = f\left(\sum_{i \in A \cup B} w_{ji}(t)\mu_i(t)\right) \tag{2}$$

The net output of the output layer at time $t + n$ through nonlinear system f is

22

computed by:

$$z(t + n) = f\big(net(t + n)\big) = f\left(\sum_j v_j(t + 1)y_j(t + 1)\right) \tag{3}$$

Let $d(t + n)$ denote the target value at time $t+n$. The time-varying error e and instantaneous error E is defined by:

$$E(t + n) = \frac{1}{2}e^2(t + n) = \frac{1}{2}[d(t + n) - z(t + n)]^2 \tag{4}$$

Then the weight adjustments can be computed by minimizing the instantaneous error at time $t + n$.

$$\Delta v_j(t + 1) = -\eta_1 \frac{\partial E(t + n)}{\partial v_j(t + 1)} \tag{5}$$

$$\Delta w_{ji}(t) = -\eta_2 \frac{\partial E(t + n)}{\partial w_{ji}(t)} \tag{6}$$

where η_1 and η_2 are learning-rate parameters. The detailed recurrent learning algorithm for one-step-ahead weight adjustment can be found in Williams and Zipser (1989), and that for two-step-ahead weight adjustment can be found in Chang et al. (2012). The entire antecedent information is considered crucial and could diminish time-lag effects. Consequently, the reinforced two-step weight adjustments (Chang et al. 2012) can be extended to n-step weight adjustments, and the information obtained from time $t+1$ to $t+n-1$ can contribute to weight adjustments. The adjusted weights are used to re-calculate the forecasted values at time from $t+1$ to $t+n-1$, and then the adjusted weights are further modified by minimizing the total error between original forecasted
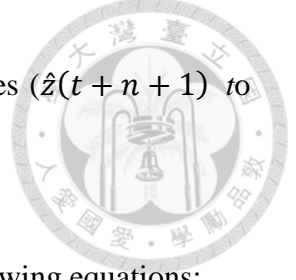
values ($z(t + n + 1)$ to $z(t + 2n - 1)$) and the re-forecasted values ($\hat{z}(t + n + 1)$ to $\hat{z}(t + 2n - 1)$).

The re-forecasted value $\hat{z}(t + n + p)$ is calculated by the following equations:

$$
\hat{y}_j(t + p + 1) = f\left(\widehat{net}_j(t + p + 1)\right)
$$

$$
= f\left(\sum_{i \in A \cup B} \left(w_{ji}(t) + \Delta w_{ji}(t)\right) \mu_i(t + p)\right)
\tag{7}
$$

$$
\hat{z}(t + n + p) = f\left(\widehat{net}(t + n + p)\right)
$$

$$
= f\left(\sum_j \left(v_j(t + 1) + \Delta v_j(t + 1)\right) \hat{y}_j(t + p + 1)\right)
\tag{8}
$$

where p denotes the time step (p = 1, 2, …, n-1). Therefore, the total reinforced error is defined by:

$$
\hat{E} = \frac{1}{2}\sum_{p=1}^{n-1} \hat{e}^2(t + n + p) = \frac{1}{2}[\hat{z}(t + n + p) - z(t + n + p)]^2
\tag{9}
$$

The reinforced weight adjustments can be expressed as:

$$
\Delta \hat{v}_j(t + 1) = -\eta_3 \frac{\partial \hat{E}}{\partial v_j(t + 1)}
\tag{10}
$$

$$
\Delta \hat{w}_{ji}(t) = -\eta_4 \frac{\partial \hat{E}}{\partial w_{ji}(t)}
\tag{11}
$$

where η_3 and η_4 are learning-rate parameters.

The weight adjustments of the R-RTRL algorithm for n-step-ahead RNNs are then shown as follows:

$$w_{ji}(t+1) = w_{ji}(t) + \Delta w_{ji}(t) + \Delta \widehat{w}_{ji}(t) \qquad (12)$$

$$v_j(t+2) = v_j(t+1) + \Delta v_j(t+1) + \Delta \widehat{v}_j(t+1) \qquad (13)$$

In sum, the reinforced process is implemented for nSA forecasting so that the adjusted weights are further modified through the comparison between the original forecasted values and the re-forecasted values.

2.2 Systematical Dynamic-neural Modeling (SDM)

The proposed SDM incorporates two core methods, the GT and the NARX network, with three optional statistical techniques to tackle rainfall-runoff modeling and spatio-temporal estimation problems, and its implementation procedure is shown in Fig. 2.3. The SDM first adopts the GT to effectively extract the non-trivial factors that significantly affect the fluctuations of the target variable (e.g., FSP water level, As or TP concentrations). The NARX network is then utilized to obtain forecasted or estimated values of the target variable with inputs consisting of the extracted non-trivial factors and the previous output values derived from recurrent connections. In addition, the time series of the target variable can be reconstructed in a minor scale (e.g., from a seasonal scale to a monthly scale) through the constructed NARX network for further evaluation of water quality. The scarcity of field data is commonly encountered in water quality modeling, and the following three optional methods can be incorporated into the
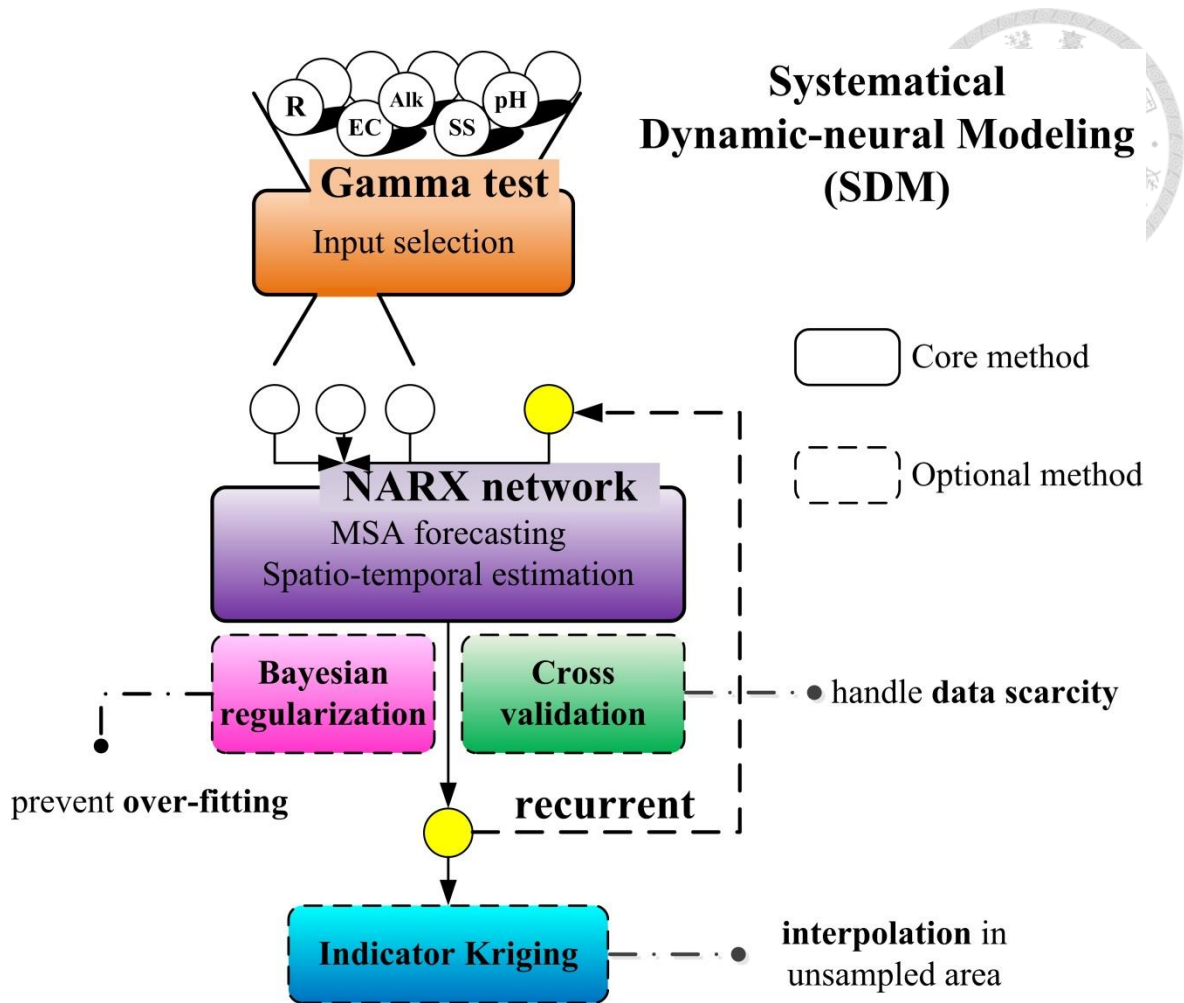
Fig. 2.3 Implementation procedure of the proposed SDM

proposed SDM to tackle data scarcity problems: the Bayesian regularization method; the cross validation technique; and the indicator kriging (IK). The Bayesian regularization method is configured to control the network complexity for preventing over-fitting. The cross validation technique is used to produce a low-bias estimator of the generalizability, and thus provides a sensible criterion for model selection in the calibration stage. Finally, the IK can be implemented to derive the probability map of the target variable in unsampled areas. The methods for use in the SDM are introduced

as follows:

A. Core techniques in the SDM

1) Gamma test (GT)

The GT, presented by Koncar (1997) and Agalbjorn et al. (1997), is a data analysis

technique for assessing the extent to which a given set of M data points can be modeled

by an unknown smooth non-linear function.

Suppose a set of input-output observation data is given in the form of:

$$[(x_i, y_i), 1 \leq i \leq M]$$ (14)

where vectors x_i are d dimensional vectors (with a record length of M) and the

corresponding outputs y_i are scalars. The underlying relationship of the system is

expressed as:

$$y = f(x_1....x_d) + r$$ (15)

where f is an unknown smooth function, and r denotes a random variable that represents

noise. The Gamma statistic (Γ) is an estimate of the variance of model outputs, which

cannot be accounted for through a smooth data model. The GT is assessed based on the

k^{th} $(1 \leq k \leq p)$ nearest neighbor $X_{N(i,k)}$ for each vector X_i, and then the GT can be

derived from the Delta function of input vectors:

$$\delta_M(k) = \frac{1}{M} \sum_{i=1}^{M} |X_{i,k} - X_i|^2 \quad (1 \leq k \leq p)$$ (16)

where $|\cdot\cdot|$ is the Euclidean distance, and the corresponding Gamma function of the

output values is given in Eq. (15). The number of p depends on the density of sampling

data (Koncar, 1997). In this dissertation, the number of p is determined as the value that

produces the minimum Γ value through trial and error (p ranges from 10 to 50).

$$\gamma_M(k) = \frac{1}{2M} \sum_{i=1}^{M} \left| y_{N(i,k)} - y_i \right|^2 \quad (1 \le k \le p) \tag{17}$$

where $y_{N(i,k)}$ is the corresponding y-value for the kth nearest neighbor of X_i, in Eq. (14).

For computing Γ, a least squares regression line is constructed for p points
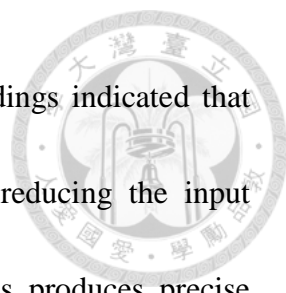
$(\delta_M(k), \gamma_M(k))$ as Eq. (16):

$$\gamma = A\delta + \Gamma \tag{18}$$

where A is the gradient.

Performing a single GT is a fast procedure, which can provide the noise estimate

(Γ value) for each subset of input variables. When the subset for which its associated

Γ value is the closest to zero, it can be considered as "the best combination" of input

variables. As the result, the GT is different from other input selection methods or

preprocessing methods, such as the correlation coefficient analysis or the principle

component analysis (PCA). The correlation coefficient analysis can only provide linear

estimation for the input-output datasets, and the principle components determined by

the PCA are calculated from all input variables and thus can neither extract the most

important factors nor reduce the dimension of input datasets.

Several studies discussed about the GT theory and its applications in time series

forecasting (Durrant, 2001; Tsui et al., 2002). Lately, research findings indicated that it is suitable and effective to combine ANNs with the GT for reducing the input dimension through identifying non-trivial input variables and thus produces precise outputs of ANNs (Moghaddamnia et al., 2009; Noori et al., 2010; Noori et al., 2011). Therefore, the NARX network combined with the GT is used to estimate water quality and forecast water level in this dissertation.

2) Nonlinear Autoregressive with eXogenous input (NARX) network

The NARX network is a recurrent network, which is suitable for time series prediction (Chang et al., 2013; Jiang and Song, 2011; Menezes Jr. and Barreto, 2008, Shen and Chang, 2013). Figure 2.4 shows the architecture of the NARX network used in this dissertation. The NARX network consists of three layers and produces recurrent connections from the outputs, which may delay several unit times to form new inputs. Therefore, this nonlinear system for estimation ($N = 0$) and *N*-step-ahead forecasting ($N \geq 1$) can be mathematically represented by the following equation:

$$z(t+N) = f\left[z(t+N-1),\ldots,z(t+N-q);U(t)\right] \tag{19}$$

where *U(t)* and *z(t+N)* denote the input vector and output value at the time step *t*, respectively. *f(·)* is the nonlinear function, and *q* is the output-memory order. There are
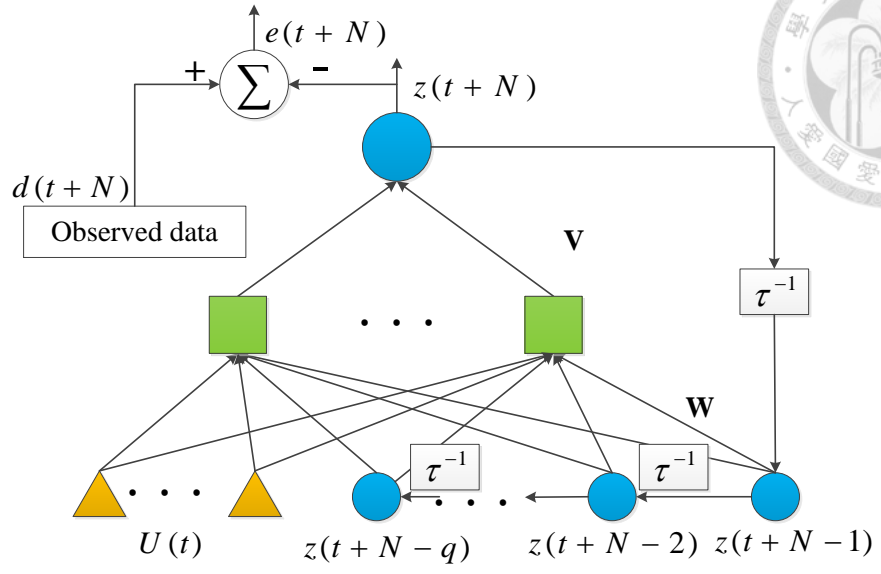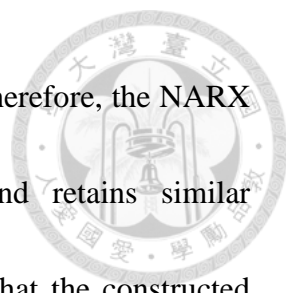
Fig. 2.4 Architectures of the NARX network

two input regressors: the regressor $z(t+N\text{-}i)$ (i ranges from 1 to q) plays the role of an

autoregressive model while $U(t)$ plays the role of an implicit exogenous variable in time

series.

There are two ways to train the NARX network. The first mode is the

Series-parallel (SP) mode, where the output's regressor in the input layer is formedonly

by the target (actual) values of the system, $d(t)$.

$$z(t+N) = f\big[d(t+N-1),\ldots,d(t+N-q);U(t)\big] \tag{20}$$

The other alternative is the Parallel (P) mode, where estimated outputs are fed back

into the output's regressor in the input layer, which can also be mathematically

represented as Eq. (19). In practice, when forecasting is conducted for more than

two-step-ahead ($N>1$), the q antecedent actual values ($d(t+N{-}1)$, $d(t+N{-}2)$,. . . ,

z(t+N−q)) are future data that cannot be obtained at current time. Therefore, the NARX network is trained in P mode with imperfect information and retains similar characteristics of input–output patterns in the testing stages such that the constructed NARX network can maintain similar capability of real time multi-step-ahead forecasting in both training and testing stages.

For water quality estimation, it is common that a regional model for estimating target variable (output) at unsampled times or at unmonitored sites often produces poor performance. The reasons are that the availability of the actual values of target variable may not always hold at certain monitoring stations and the consistency (autoregression) of the target variable is usually not as strong as that in rainfall-runoff modeling. Therefore, the relationship between actual and estimated values of the target variable can be established by the NARX network trained in the SP mode. Then the trained NARX network in the P mode can be applied at unrecorded periods or at ungauged sites for improving estimation accuracy through recurrent connections. It indicates that the recurrent connections of the NARX network are practically meaningful when modelling at unrecorded times or at ungauged sites.

B. Optional techniques in the SDM

1) Bayesian regularization

The regularization method proposed by MacKay (1992) can improve the generalizability of a neural network through minimizing an objective function that constrains the value of network weights. The idea is based on that the true underlying function is assumed to have a degree of smoothness controlled by the network parameters, and the network response will be smooth as the values of parameters are kept small. Thus the network is able to sufficiently represent the true function, rather than capture the noise. The objective function of the network in the regularization method is given by:

$$M(W) = \beta E_D + \alpha E_W \tag{21}$$

where E_D is the mean square error of network outputs:

$$E_D = \frac{1}{2}\sum_{t=1}^{n}(d(t) - z(t))^2 \tag{22}$$

where n is the number of samples, and E_W is a penalty term of network complexity in the regularization method, in which the smaller values of network weights imply lower connection complexity for network weights. E_W is given in Eq.(23).

$$E_W = \frac{1}{2}\sum_{i=1}^{m}w_i^2 \tag{23}$$

where w_i is the weight value of the network, and m is the number of weights. α and β
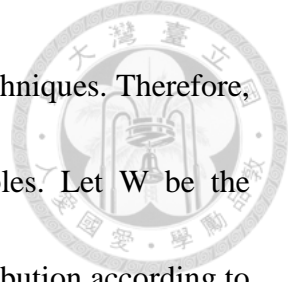
are regularization parameters that can be determined by Bayesian techniques. Therefore,

the weights of the network can be considered as random variables. Let W be the

network weight vector and D be the sample data. The posterior distribution according to

the Bayes' rule is shown as follows:

$$P(W \mid D, \alpha, \beta, \mathrm{H}) = \frac{P(D \mid W, \beta, \mathrm{H}) P(W \mid \alpha, \mathrm{H})}{P(D \mid \alpha, \beta, \mathrm{H})} \tag{24}$$

where $P(W \mid \alpha, \mathrm{H})$ is the prior density, $P(D \mid W, \beta, \mathrm{H})$ is the likelihood function,

$P(D \mid \alpha, \beta, \mathrm{H})$ is the normalization factor, and H presents the structure of the network.

Assuming the noise and the prior distribution for the weights are both Gaussian, the

probability densities can be written as:

$$P(W \mid \alpha, \mathrm{H}) = \frac{1}{Z_W(\alpha)} \exp(-\alpha E_W) \quad , Z_W(\alpha) = \left(\frac{2\pi}{\alpha}\right)^{\frac{m}{2}} \tag{25}$$

and

$$P(D \mid W, \beta, \mathrm{H}) = \frac{1}{Z_D(\beta)} \exp(-\beta E_D) \quad , Z_D(\beta) = \left(\frac{2\pi}{\beta}\right)^{\frac{n}{2}} \tag{26}$$

Substituting (25) and (26) into (24), it becomes:

$$P(W \mid D, \alpha, \beta, \mathrm{H}) = \frac{1}{Z_M(\alpha, \beta)} \exp(-\beta E_D - \alpha E_W) = \frac{1}{Z_M(\alpha, \beta)} \exp\left[-M(W)\right] \tag{27}$$

where

$$Z_M(\alpha, \beta) = \int_{-\infty}^{\infty} \left(-M(W)\right) dW \tag{28}$$

From the Bayesian framework, the optimal weights can be derived by maximizing

the posterior probability, which is equivalent to minimizing the regularized objective

function shown in Eq.(21).

The value of α and β can also be optimized by applying the Bayes' rule.

$$P(\alpha,\beta\,|\,D,\mathrm{H}) = \frac{P(D\,|\,\alpha,\beta,\mathrm{H})P(\alpha,\beta\,|\,\mathrm{H})}{P(D\,|\,\mathrm{H})} \qquad (29)$$

Therefore, the values of α and β can then be inferred at the minimum $\mathrm{W_{MP}}$ of $M(W)$, which are shown as:

$$\alpha_{MP} = \frac{\gamma_p}{2E_W(W_{MP})} \qquad (30)$$

$$\beta_{MP} = \frac{m-\gamma_p}{2E_D(W_{MP})} \qquad (31)$$

where γ_p presents the effective number of network parameters.

2) Cross Validation

Cross-validation, which requires partitioning data into training and testing sets, is commonly used to obtain a reliable estimate of the test error for model performance or for use as a model selection criterion. In the k-fold cross-validation for model selection, the first step is to assign a setting for model parameters (i.e., initial weights, epoch numbers, the number of neurons in the hidden layer, and the output-memory order of the NARX network), and then the original sample is partitioned into k subsamples. Among the k subsamples, a single subsample is retained as validation data for validating the model, and the remaining k-1 subsamples are used as training data. The cross-validation process is then repeated k times, with each of the k subsamples being

used exactly once as validation data. The k results from the folds in each round are averaged into a single estimation error, and the model in association with the minimum estimation error of multiple rounds is considered the most appropriate for use in the testing stage.

Cross validation can produce a low-bias estimator for the generalization ability of a model, and therefore provides a sensible criterion for model selection and performance comparison to tackle data scarcity (Kohavi, 1995; Stone, 1974), especially for samples that are hazardous, costly or difficult to collect, such as As or TP.

3) Indicator Kriging (IK) Method

In the application of As estimation, the IK is utilized to illustrate the variation of As concentration for the whole study area for reducing the influence of the extreme values of the estimated As concentration on the variogram and mitigating the uncertainty of the As estimation made by the NARX network model. The IK is a non-parametric geostatistical technique that involves the transformation of one variable to a binary response (0, 1) (Journel, 1983). The geostatistical spatial assumption generally holds on condition that the regionalized variable is second-order stationary. The geostatistical semivariogram of the data should be first determined, which quantifies the spatial variability of random variables between two areas. In practice, an

experimental semivariogram $\hat{\gamma}(h)$ is calculated by:

$$\hat{\gamma}(h) = \frac{1}{2N(h)} \left\{ \sum_{i=1}^{N(h)} z(u_i + h) - z(u_i)^2 \right\} \tag{32}$$

where $\hat{\gamma}(h)$ represents the semivariogram for a lag distance class h; *N(h)* denotes the set of pairs at a lag distance class h, and $z(u_i)$ and $z(u_i + h)$ are the values of the regionalized variables of interest at locations u_i and $u_i + h$, respectively.

The experimental semivariogram, $\hat{\gamma}(h)$, is fitted by a theoretical model, $\gamma(h)$, such as spherical, exponential or Gaussian functions, to determine three parameters: the nugget effect (c_o); the sill (*c*); and the range (*a*).

The geostatistical method is used to estimate the probability of exceeding a specific cut-off value (threshold), z_k, at a given location. Then the stochastic variable, *Z(u)*, is transformed into an indicator variable with a binary distribution shown as follows:

$$I(u; z_k) = \begin{cases} 1 \\ 0 \end{cases}, \text{ otherwise } \forall Z(u) \le z_k \tag{33}$$

The expected value of I(u; z_k), subject to n neighboring data, can be formed as:

$$E[I(u; z_k] = \text{Pr}ob\{Z(u) < z_k\} = F(u; z_k) \tag{34}$$

$$P(u; z_k) = 1 - F(u; z_k) \tag{35}$$

where $F(u; z_k)$ is the conditional cumulative distribution function of $Z(u) \le z_k$, and $P(u; z_k)$ is the probability that $Z(u) > z_k$. At an unsampled location (u_0), an estimation

must utilize both the IK and indicator estimator, $I^*(u_0; z_k)$, according to:

$$I^*(u_0; z_k) = \sum_{j=1}^{n} \lambda_j(z_k) I(u_j; z_k) \qquad (36)$$

where $I^*(u_0; z_k)$ represents the values of the indicator at the measurement locations, u_j

$j = 1, 2, ..., n$, and λ_j is the weighting factor of $I(u; z_k)$ when estimating $I^*(u_0; z_k)$. The IK

estimation must be unbiased with the minimum estimation error variance; that is,

$$E[I^*(u_0; z_k) - I(u; z_k)] \qquad (37)$$

$$\min\left\{\mathrm{var}\left[I^*(u_0; z_k) - I(u; z_k)\right]\right\} \qquad (38)$$

The weights, λ_j, are the solutions to the following system:

$$\begin{cases} \sum_{j=1}^{n} \lambda_j(z_k) \gamma_I(u_i - u_j; z_k) - \phi(z_k) = \gamma_I(u_i - u_0; z_k) \quad , \quad \forall i = 1 \ to \ N \\ \qquad\qquad \sum_{j=1}^{n} \lambda_j = 1 \end{cases} \qquad (39)$$

where $\phi(z_k)$ is the Lagrange multiplier $\gamma_I(u_i - u_j; z_k)$, which specifies the semivariogram

value between the indicator variables at the ith and j^{th} sampling points; and

$\gamma_I(u_i - u_{0.}; z_k)$ is the semivarigram value between the indicator variables at the ith

sampling point and u_0.

In the case study of modeling As concentration, the cut-off value is adequately set

as 10ugl^{-1}, the WHO drinking water standard for As concentration. Therefore, the IK

can derive the probability map that discloses the probability of As concentration

exceeding the WHO drinking water standard in the study area.

2.3 Comparative neural network models

In this dissertation, comparative models mainly consist of the most popular static ANN (i.e., BPNN), the original RTRL algorithm for RNNs (RTRL NN) and the Elman NN. The methodology of the BPNN and the Elman NN are briefly addressed as follows:

A. Backpropagation neural network (BPNN)

The BPNN is one of the most popular ANNs (Rumelhart, 1986). It belongs to a typical three-layered static feedforward neural network, which is comprised of multiple elements including nodes and weight connections (**W** and **V**) that link nodes. The network is divided into an input layer, a hidden layer and an output layer. Fig. 2.5 (a) shows the structure of the BPNN. In this dissertation, the BPNN is trained by the Levenberg–Marquardt back propagation algorithm based on the model output ($z(t+N)$) and observed data ($d(t+N)$).

B. Elman neural network (Elman NN)

The Elman NN is a three-layer RNN with internal time-delay feedback connections in the hidden layer. Each input neuron is connected to a hidden neuron, where each hidden neuron has its corresponding time-delay unit. The structure of the Elman NN is shown in Fig. 2.5 (b). A recurrent connection allows its time-delay unit to store the information of this hidden neuron as an additional input to all hidden neurons at the next time step.

(a)

$e(t + N)$

$+$ $\sum$ $-$ $z(t + N)$

$d(t + N)$

Observed data

V

W

$U(t)$

(b)

$e(t + N)$

$+$ $\sum$ $-$ $z(t + N)$

$d(t + N)$

Observed data

V

$Y(t + 1)$

τ^{-1}

W

$U(t)$

$Y(t)$
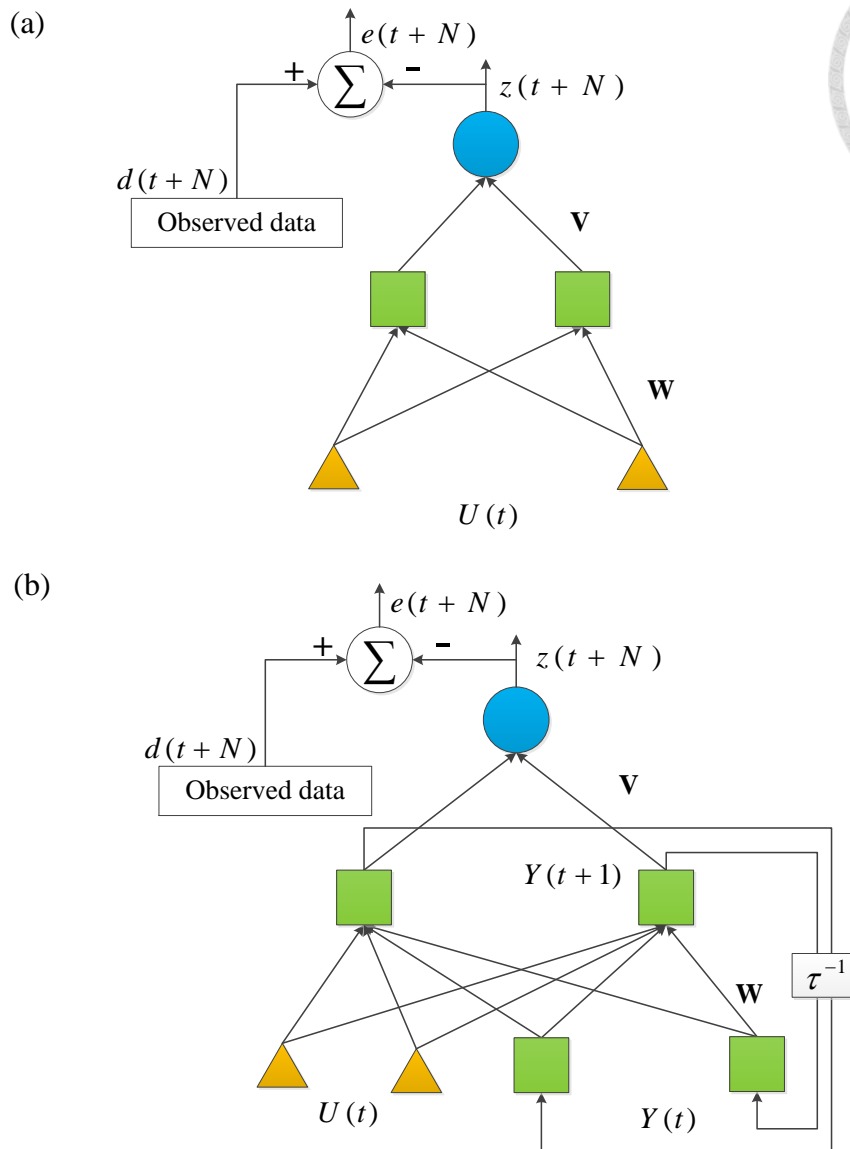
Fig. 2.5 Architectures of (a) the BPNN; and (b) the Elman NN.

Therefore, the Elman NN has an inherent dynamic memory given by the recurrent connections of the time-delay units, and its output depends not only on the current input information but also on the previous states of the network. In this dissertation, the Elman NN is trained by the Levenberg–Marquardt back propagation algorithm.

3. Case studies

3.1 Reinforced recurrent neural networks for multi-step-ahead flood forecasting

In this section, the reliability and applicability of the proposed R-RTRL NN is demonstrated through two-step-ahead (2SA), four-step-ahead (4SA) and six-step-ahead (6SA) forecasting made for a famous benchmark chaotic time series and a reservoir inflow case in Taiwan.

A. Numerical simulation

1) Phase space reconstruction

The first stage in the analysis of chaotic time series is to implement the phase space reconstruction theory, which can reconstruct a nonlinear model with a low-dimensional phase space to reflect the actual dynamic system. Based on the Takens' embedding theorem (Takens, 1981), a series of observations from a chaotic system can reconstruct an attractor, a subset of the phase space of the system with two parameters of time delay (T) and dimension (D). If appropriate T and D are selected, the attractor will retain topological properties and reveal the hidden information of the original dynamic system. The mutual information method (Fraser and Swinney, 1986) and the Cao's method (Cao, 1997) are used to determine the proper T and D, respectively.

Therefore, for an observed time series $x(t)$, the attractor $X(t) = [x(t), x(t - T), ..., x(t - (D - 1)T)]$ can be formed. Then the future value at time $t+n$ is

determined by the nonlinear function $\boldsymbol{F}$, which governs the system. The nonlinear

function $\boldsymbol{F}$ is defined as follows:

$$x(t + n) = F\{[x(t), x(t - T), \dots, x(t - (D - 1)T)]\} \tag{40}$$

The well trained ANN models can approximate the governing function $\boldsymbol{F}$. As a

result, the patterns of input-output data pairs of all ANN models in this study case are

shown as $\{X(t), x(t + n)\}$ and are applied to chaotic time series.

2) MSA forecasting of Mackey-Glass time series

To demonstrate and evaluate whether the proposed R-RTRL NN can construct a

reliable multi-step-ahead predictor by effectively utilizing the most current information,

the developed neural network model is first compared with a simulated chaotic system

(Mackey-Glass time series) for direct 2SA, 4SA and 6SA forecasting. Its performance is

then compared with that of the RTRL NN, the Elman NN and the BPNN. The Elman

NN and the BPNN configurations are used in many filtering and modeling applications

for time series, which have already been widely discussed (Chen et al., 2010; Zabiri et

al., 2009).

The Mackey-Glass differential delay equation (MacKey and Glass, 1977) is

defined below:

$$\frac{dx(t)}{dt} = \frac{0.2x(t - \tau)}{1 + x^{10}(t - \tau)} - 0.1x(t) \tag{41}$$

where the initial condition $x(0) = 1.2$ and $\tau = 17$. 1000 input-output data pairs are
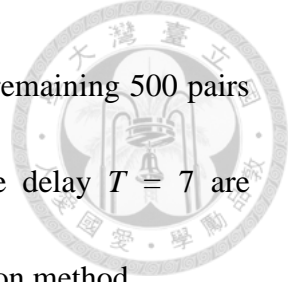
generated, where the first 500 pairs are used for training while the remaining 500 pairs are used for testing. The embedding dimension $D = 3$ and time delay $T = 7$ are determined, according to the Cao's method and the mutual information method.

For forecasting 2SA, 4SA and 6SA Mackey-Glass time series, the input and processing layers of the RTRL NN and the proposed R-RTRL NN consist of 3 and 8 neurons, respectively, while the two-layer Elman NN and BPNN trained by the Levenberg-Marquardt back propagation algorithm also consist of 3 and 8 neurons in the input and hidden layers, respectively. The numbers of the processing (hidden layer) neurons and layers determined above for these four models are identified as the best structures by trial and error. A training dataset is used to construct the aforementioned four neural network models for direct 2SA, 4SA and 6SA forecasting.

The root mean square error (RMSE), mean absolute error (MAE), and the goodness-of-fit with respect to the benchmark time series (G_{bench}) are used as performance criteria (Nash and Sutcliffe, 1970; Seibert, 2001) and are defined as:

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(Q_i - \hat{Q}_i)^2}{n}} \tag{42}$$

$$MAE = \frac{\sum_{i=1}^{n}\left|Q_i - \hat{Q}_i\right|}{n} \tag{43}$$

$$G_{bench} = 1 - \frac{\sum_{i=1}^{n}\left(Q_i - \hat{Q}_i\right)^2}{\sum_{i=1}^{n}\left(Q_i - Q_{i,bench}\right)^2} \tag{44}$$

where Q_i is the observed value in the ith step, $\hat{Q}_i$ is the forecasted value in the ith step, and n is the number of data points. $Q_{i,bench}$ is the antecedent observed value, i.e., $Q_{i,bench} = Q_{i-N}$ for NSA forecasting. To compare the proposed R-RTRL NN with comparative models, the criterion $G_{bench,II}$ is defined, where the benchmark $Q_{i,bench}$ is the forecasted value of the proposed R-RTRL NN in the ith step.

The results of model comparison are summarized in Table 1. It appears that: 1) all the models are suitably trained, and their training and testing results are quite consistent in all the cases; 2) the recurrent neural networks (i.e., R-RTRL NN, RTRL NN and Elman NN) can provide much better performance than the static time-delay BPNN, in terms of RMSE and MAE values; and 3) the proposed R-RTRL NN produces better performance than the RTRL NN, the Elman NN and the BPNN models for 2SA, 4SA and 6SA forecasting. Furthermore, the negative $G_{bench,II}$ values produced by the RTRL NN, the Elman NN and the BPNN reveal the superiority of the R-RTRL NN in MSA forecasting. In sum, the results demonstrate the proposed online learning algorithm (R-RTRL) that takes the closest antecedent information into consideration can effectively re-adjust synaptic weights in real time and the constructed model (R-RTRL NN) can provide reliable and accurate forecasts in real-time MSA forecasting.

Table 3.1 Model performance of two- to six-step-ahead forecasting for Mackey-Glass time series

		Training		Testing			
		RMSE	MAE	RMSE	MAE	G_{bench}	$G_{bench,II}$
R-RTRL NN	*2SA*	3.56E-03	2.79E-03	3.51E-03	2.77E-03	0.997	-
	4SA	4.23E-03	2.59E-03	4.06E-03	2.52E-03	0.999	-
	6SA	4.79E-03	3.33E-03	4.79E-03	3.35E-03	0.999	-
RTRL NN	*2SA*	6.69E-03	5.40E-03	6.51E-03	5.30E-03	0.990	-2.44
	4SA	7.43E-03	5.17E-03	7.40E-03	5.13E-03	0.997	-2.32
	6SA	9.09E-03	7.17E-03	8.94E-03	7.01E-03	0.998	-2.49
Elman NN	*2SA*	6.37E-03	4.44E-03	6.27E-03	4.31E-03	0.990	-1.38
	4SA	6.49E-03	5.08E-03	6.43E-03	5.00E-03	0.998	-0.80
	6SA	6.60E-03	5.17E-03	6.49E-03	5.10E-03	0.999	-2.42
BPNN	*2SA*	9.01E-03	6.16E-03	9.05E-03	6.07E-03	0.981	-5.66
	4SA	1.30E-02	1.03E-02	1.29E-02	1.02E-02	0.990	-9.08
	6SA	1.52E-02	1.32E-02	1.52E-02	1.31E-02	0.993	-9.03

B. Application of MSA reservoir inflow forecasting

1) Study area and datasets

Taiwan, located in the subtropical zone of the North Pacific Ocean, is an island with mountainous terrains and steep landforms, where typhoons usually couple with heavy rainfall and thus cause downstream flooding within a few hours. The Shihmen Reservoir, situated upstream of the Tahan Creek in northern Taiwan, has operated for multiple purposes including water supply, hydropower generation, flood mitigation, education and tourism. The reservoir has notably contributed to agricultural production, industrial development and the alleviation of drought disasters for decades. The basin area of the reservoir is about 763.4 km^2, and the effective reservoir capacity is 251.88 million cubic meters. The Shihmen Reservoir is for sure the most important water resources facility to
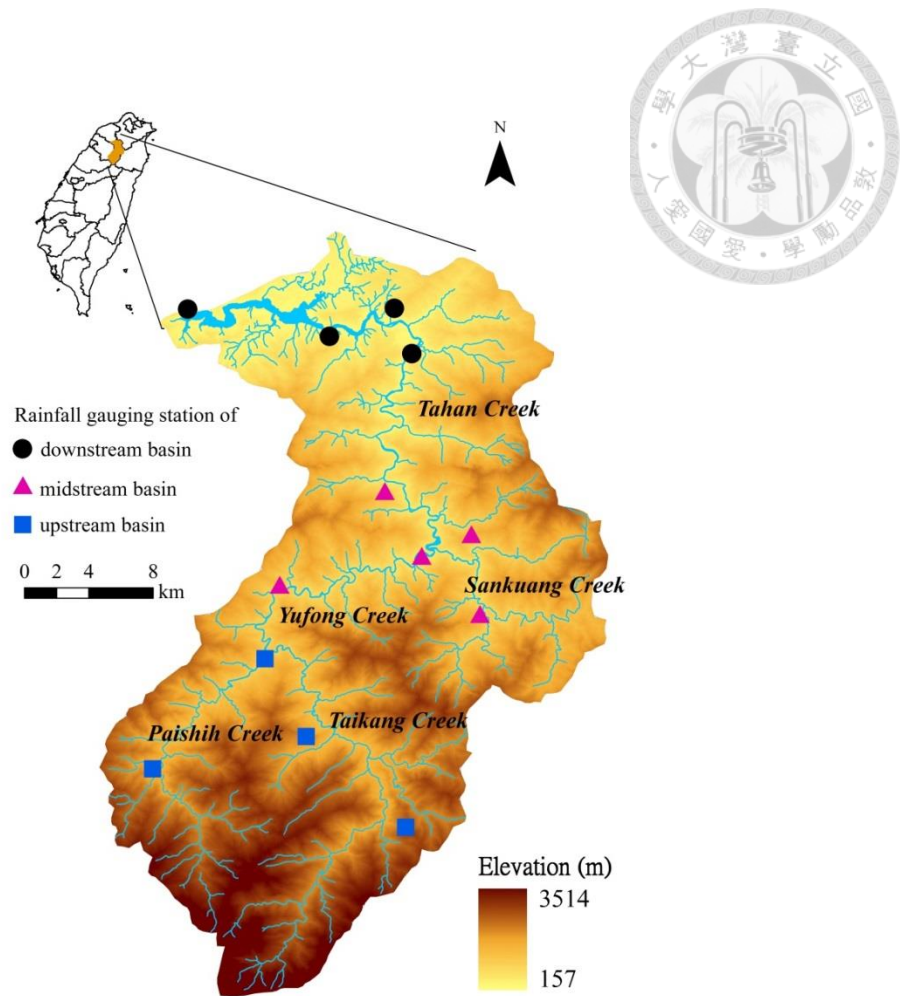
Fig. 3.1 Locations of the Shihmen Reservoir and rainfall gauging stations.

Table 3.2 Summary statistics of reservoir inflow and average hourly rainfall in training and testing datasets

	Training dataset				Testing dataset			
	Mean	**SD[1]**	**Max**	**Min**	**Mean**	**SD**	**Max**	**Min**
Inflow (cms)	725	1154	8594	1.26	720	910	5300	8.57
Average rainfall of downstream (mm h[-1])	4.0	7.9	55.1	0.0	4.0	7.1	45.7	0.0
Average rainfall of midstream (mm h[-1])	4.8	9.6	63.5	0.0	4.5	7.5	47.3	0.0
Average rainfall of upstream (mm h[-1])	4.4	8.8	62.0	0.0	4.5	7.1	46.6	0.0

[1]Standard deviation

Taipei metropolitan areas. An accurate inflow forecasting model for the reservoir is desired and crucial to flood control and water resources management (Huang and Hsieh, 2010). The longer the forecasting steps into the future, the more beneficial it is, in terms of time to adjust reservoir operation and reduce flood damages.

Figure 3.1 shows the locations of the basin and rainfall gauging stations in this case study. The hourly inflow and rainfall data of 22 typhoon events during 2001 and 2009 were collected. A total of 2136 datasets are used in this study, where 1296 datasets collected from 2001 to 2006 are used for training while the remaining 840 datasets collected from 2007 to 2009 are used to test the models independently. The Shihmen basin is divided into three areas (up-, mid- and downstream areas shown in Fig. 3.1), and the weighted average rainfall of each area is computed by the Thiessen polygon method (Thiessen, 1911). The summary statistics for reservoir inflow and average rainfall datasets are presented in Table 3.2, in which the extremely high inflow and the large variance of inflow implies that the MSA inflow forecasting is an important and challenging task in the Shihmen Reservoir.

2) *MSA forecasting of reservoir inflow time series*

The proposed R-RTRL NN is applied to the Shihmen Reservoir for forecasting reservoir inflow during typhoon events and is also compared with the other three models (RTRL NN, Elman NN and BPNN) for performance evaluation. Because the

transit time of flows moving from rainfall gauging stations to the Shihmen Reservoir is different, the Kendall's tau rank and Pearson's correlation coefficient are applied to the identification of the highest correlation among the different time lags between rainfall and reservoir inflow. The time lags of rainfall traveling from the up-, mid- and downstream basins to the reservoir are identified as seven, six and five hours, respectively. Therefore, the input layers of these different network models are established based on current inflow data (Q(t)) together with antecedent rainfall data (denoted by Ru(t-7), Rm(t-6) and Rd(t-5) accordingly) collected at up-, mid- and downstream basins with associated time lags. For direct 2SA inflow forecasting, the input and processing layers of the RTRL NN and R-RTRL NN consist of 4 and 6 neurons, respectively. The two-layer Elman NN and BPNN trained by the Levenberg-Marquardt back propagation algorithm also consists of 4 and 6 neurons in the input and hidden layers, respectively. The numbers of the processing (hidden layer) neurons and layers determined above for these models are identified as the best structures through a great number of trial-and-error processes. The numbers of neurons in the processing and hidden layers of these four models are increased to 8 for direct 4SA and 6SA inflow forecasting. The performance of these four models is evaluated by the criteria of RMSE, MAE, coefficient of efficiency (Nash Efficiency, or CE) (Nash and Sutcliffe, 1970), coefficient of correlation (CC), G_{bench} and $G_{bench,II}$. The CE and CC

are defined as:

$$CE = 1 - \frac{\sum_{i=1}^{n}\left(Q_i - \hat{Q}_i\right)^2}{\sum_{i=1}^{n}\left(Q_i - \overline{Q}\right)^2} \tag{45}$$

$$CC = \frac{\sum_{i=1}^{n}\left(\hat{Q}_i - \overline{\hat{Q}}\right)\left(Q_i - \overline{Q}\right)}{\sqrt{\sum_{i=1}^{n}\left(\hat{Q}_i - \overline{\hat{Q}}\right)^2 \sum_{i=1}^{n}\left(Q_i - \overline{Q}\right)^2}} \tag{46}$$

Summarized results are presented in Tables 3.3-3.5. Results indicate that the proposed R-RTRL NN model can produces acceptable RMSE values (184 to 490 cms) when compared with the mean and standard deviation (720.10±910.44 cms in the testing dataset of Table 3.2) of observed inflow data, and it has smaller RMSE and MAE values and higher CE, CC and G_{bench} values than the RTRL NN, the Elman NN and the BPNN models in both training and testing phases for 2SA, 4SA and 6SA inflow forecasting. It can be noticed that the proposed R-RTRL NN makes less difference in RMSE values between training and testing phases than comparative models, which demonstrates the impressive generalizability of the proposed R-RTRL NN. In addition, the proposed R-RTRL NN model produces positive G_{bench} values in the testing datasets for 2SA, 4SA and 6SA forecasting whereas the comparative models only produce either negative or near-zero G_{bench} values in the testing datasets. The results indicate the proposed online
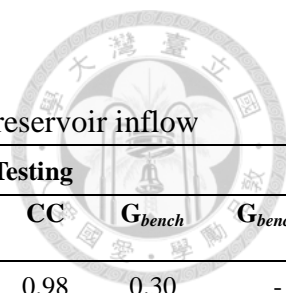
Table 3.3 Model performance of two-step-ahead forecasting for reservoir inflow

	Training					Testing					
	RMSE (cms)	MAE (cms)	CE	CC	G_{bench}	RMSE (cms)	MAE (cms)	CE	CC	G_{bench}	$G_{bench,II}$
R-RTRLNN	177	106	0.98	0.99	0.59	184	106	0.96	0.98	0.30	-
RTRL NN	214	112	0.97	0.98	0.40	228	124	0.94	0.97	-0.07	-0.53
Elman NN	211	107	0.97	0.98	0.42	220	118	0.94	0.97	-2.72E-04	-0.43
BPNN	220	113	0.96	0.98	0.37	232	125	0.93	0.97	-0.11	-0.59

Table 3.4 Model performance of four-step-ahead forecasting for reservoir inflow

	Training					Testing					
	RMSE (cms)	MAE (cms)	CE	CC	G_{bench}	RMSE (cms)	MAE (cms)	CE	CC	G_{bench}	$G_{bench,II}$
R-RTRL NN	308	161	0.93	0.96	0.63	314	172	0.88	0.94	0.26	-
RTRL NN	351	177	0.91	0.95	0.53	362	198	0.84	0.92	0.02	-0.33
Elman NN	345	172	0.91	0.95	0.54	360	198	0.84	0.92	0.03	-0.32
BPNN	357	181	0.91	0.95	0.51	365	200	0.83	0.92	1.82E-03	-0.35

Table 3.5 Model performance of six-step-ahead forecasting for reservoir inflow

	Training					Testing					
	RMSE (cms)	MAE (cms)	CE	CC	G_{bench}	RMSE (cms)	MAE (cms)	CE	CC	G_{bench}	$G_{bench,II}$
R-RTRL NN	440	238	0.86	0.93	0.62	490	302	0.71	0.85	0.07	-
RTRL NN	498	263	0.82	0.91	0.52	559	316	0.63	0.82	-0.22	-0.30
Elman NN	481	254	0.83	0.91	0.55	542	303	0.65	0.83	-0.15	-0.22
BPNN	525	276	0.80	0.89	0.46	600	324	0.57	0.82	-0.40	-0.50
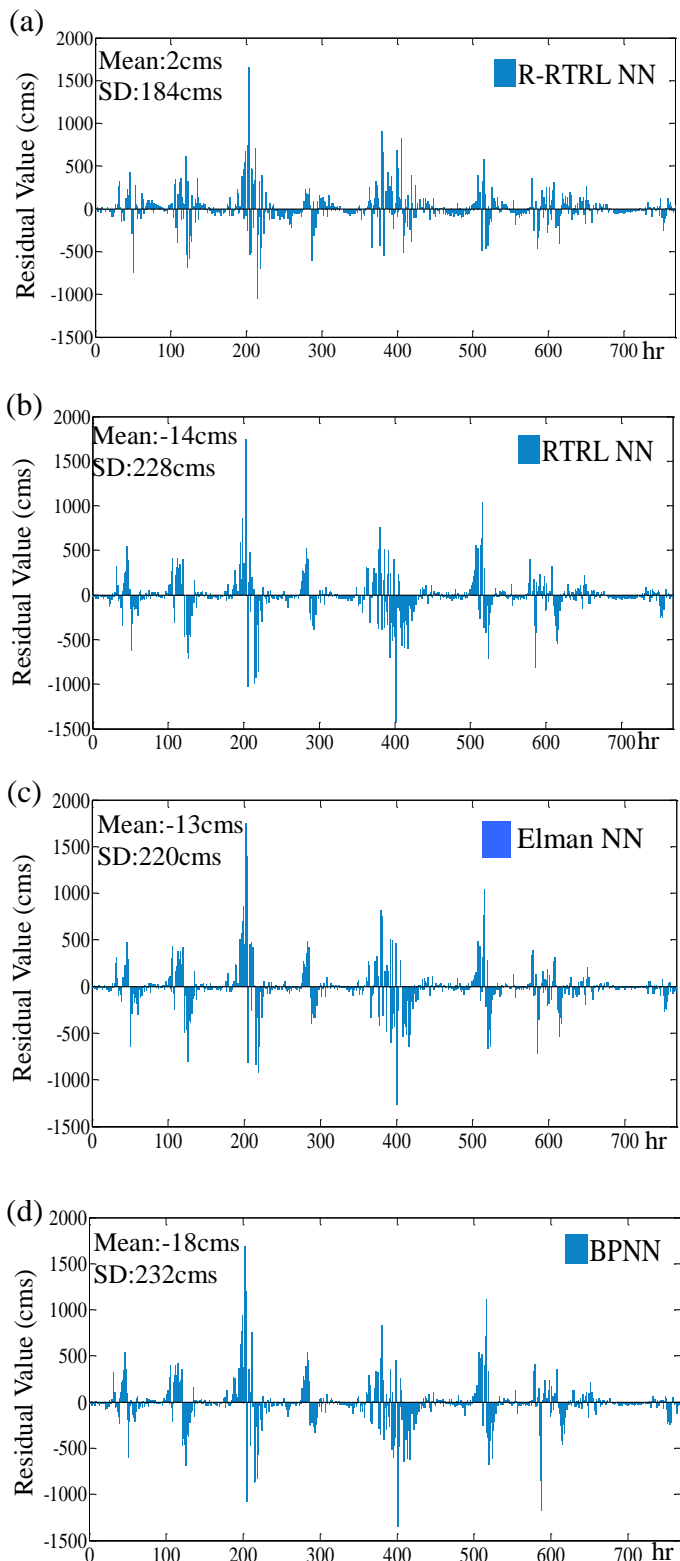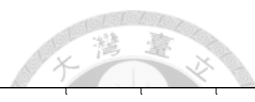
Fig. 3.2 *2*SA inflow forecast residuals (of testing data sets) based on (a) R-RTRL NN, (b) RTRL NN, (c) Elman NN and (d) BPNN, respectively.
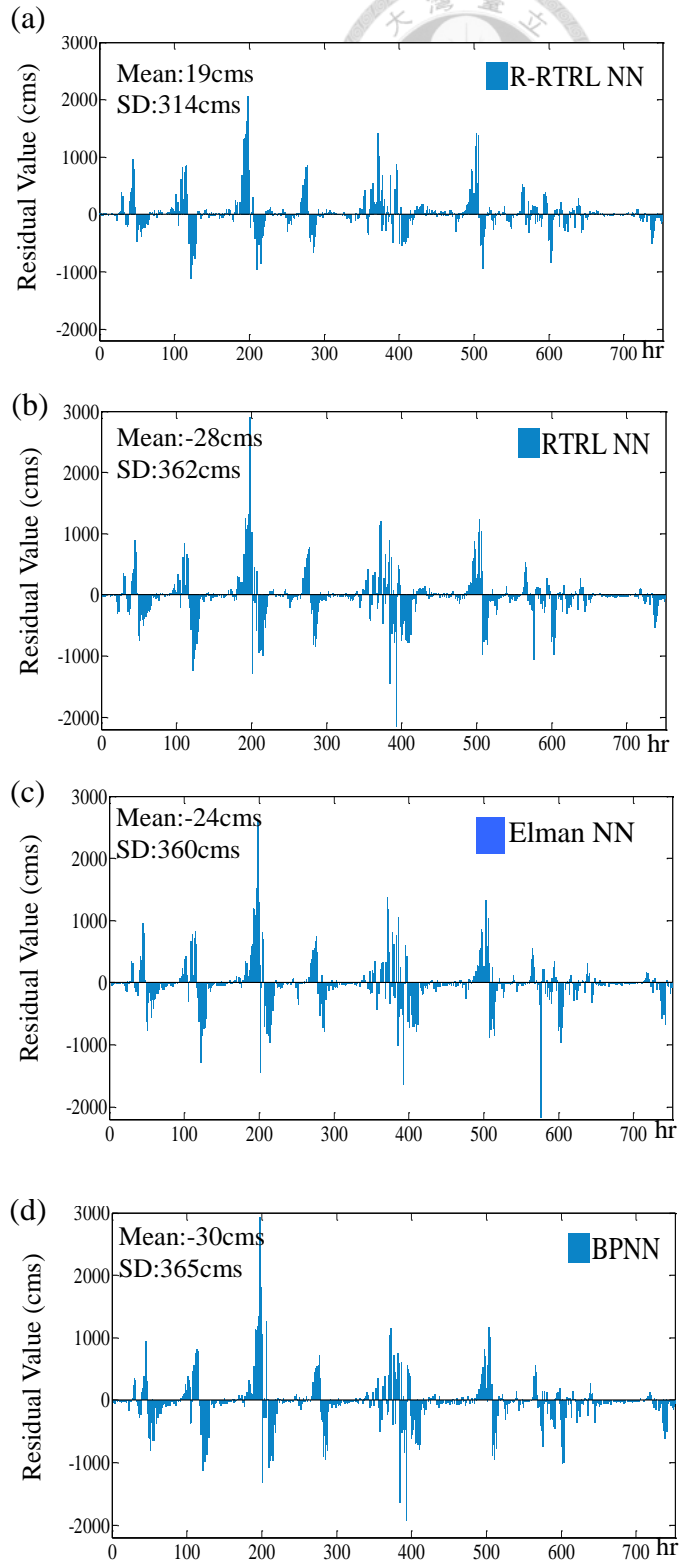
Fig. 3.3 *4*SA inflow forecast residuals (of testing data sets) based on (a) R-RTRL NN, (b) RTRL NN, (c) Elman NN and (d) BPNN, respectively.
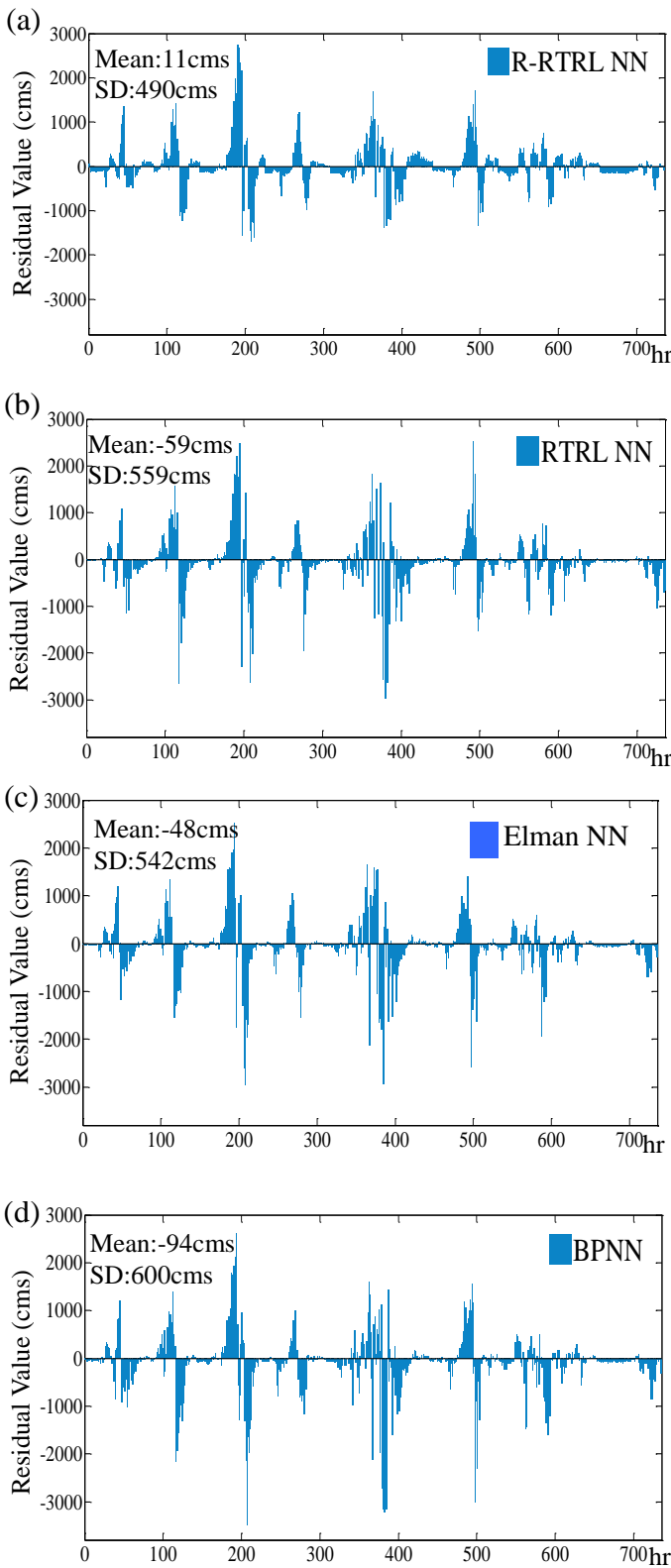
Fig. 3.4 *6*SA inflow forecast residuals (of testing data sets) based on (a) R-RTRL NN, (b) RTRL NN, (c) Elman NN and (d) BPNN, respectively.
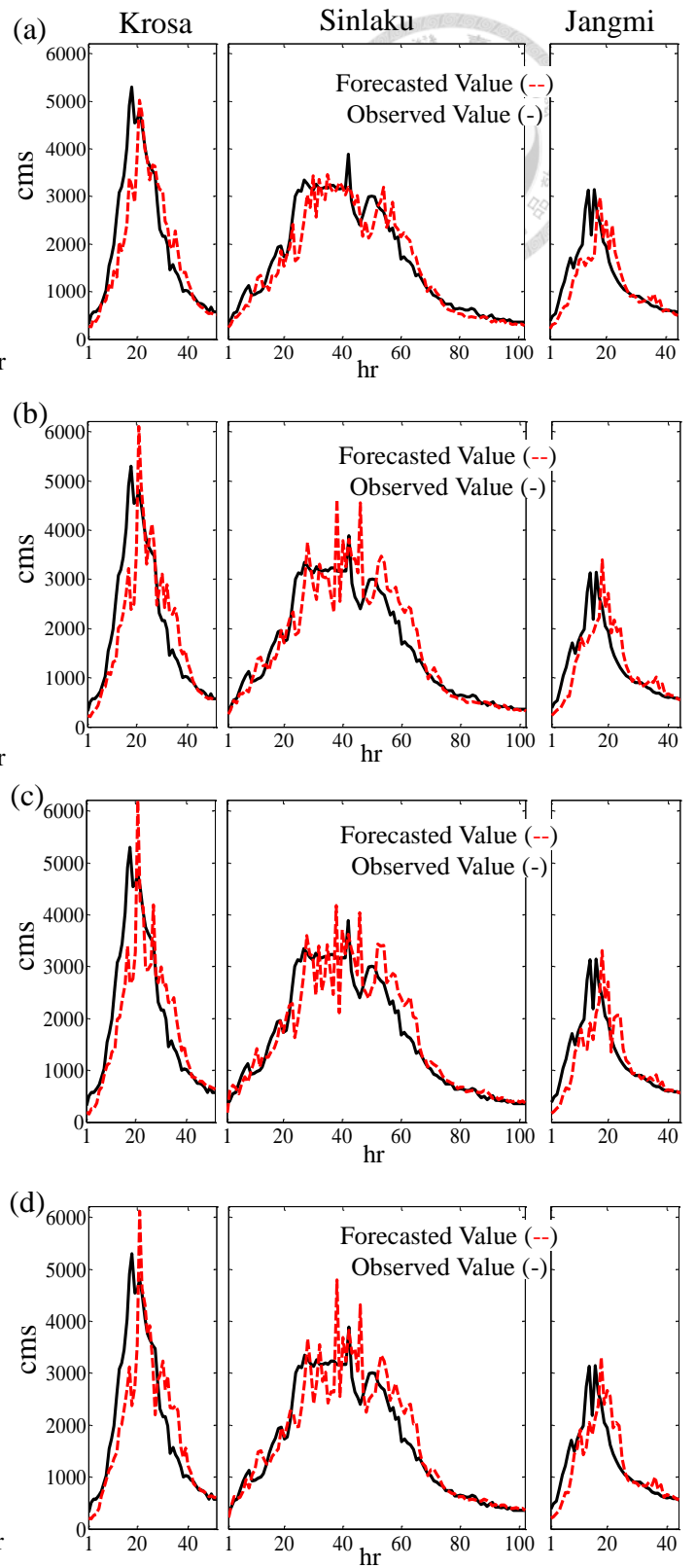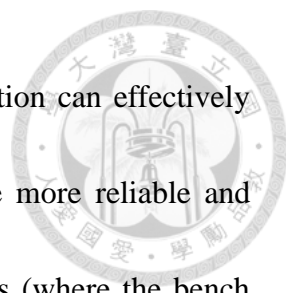
Fig. 3.5 *4*SA inflow forecasting (Typhoons Krosa, Sinlaku and Jangmi in testing data sets) based on (a) R-RTRL NN, (b) RTRL NN, (c) Elman NN and (d) BPNN, respectively.

learning algorithm (R-RTRL) that adopts the most current information can effectively

mitigate time-lag problems and the R-RTRL NN model can make more reliable and

accurate MSA forecasts. When closely assessing the $G_{bench,II}$ values (where the bench

series is the forecasted values of the proposed R-RTRL NN), all comparative models

produce negative values for 2SA, 4SA and 6SA inflow forecasting in the testing phases.

This result provides extra evidence that the proposed R-RTRL NN achieves superior

performance than comparative networks.

Figs. 3.2-3.4 show the corresponding residuals and their mean and standard

deviation produced by the R-RTRL NN, the RTRL NN, the Elman NN and the BPNN

models for 2SA, 4SA and 6SA forecasting in the testing datasets. The barplots of

residuals clearly indicate that the R-RTRL NN provides the smallest residuals and less

over-estimation results (the positive mean, smallest absolute mean and standard

deviation) as compared with the comparative models. To easily distinguish the

performance of these four models, three typhoon events (Typhoons Krosa, Sinlaku and

Jangmi) with high peak flow data (above 3000 cms) are extracted from the testing

dataset to illustrate the hydrographs of observed and 4SA forecasted inflow obtained

from the proposed R-RTRL algorithm and comparative models (Fig. 3.5). It

demonstrates that the proposed R-RTRL NN can mitigate time-lag problems and well

forecast 4SA inflow values whereas all comparative models not only have significant
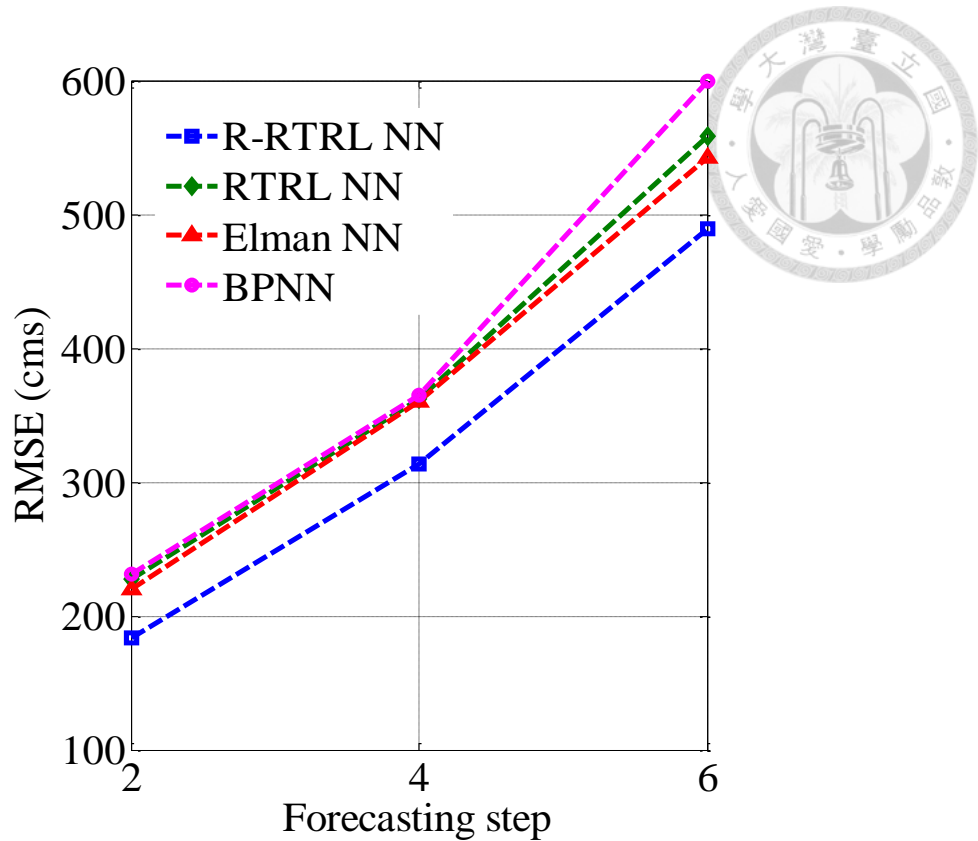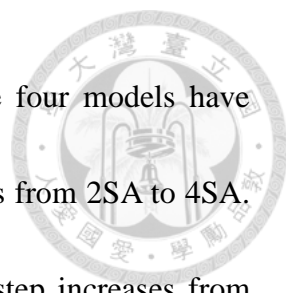
Fig. 3.6 Relationship between inflow forecast errors (RMSE) and forecasting steps of four neural network models.

time-lag phenomena but fail to well forecast 4SA inflow values (seriously over-estimate and oscillate at peak flows; and the G_{bench} values of comparative models are negative). In addition, the low G_{bench} value (0.07) in Table 3.5 indicates the maximum forecast time step for the proposed R-RTRL NN to reach is six (6 hrs) in this study case, which can provide sufficient response time to fully open the floodgates in the Shihmen Reservoir (usually take about two hours) and more response time for both reservoir flood control and flood warnings to downstream areas.

In summary, the relationship between forecast errors (RMSE) and forecasting steps

of these four models is presented in Fig. 3.6. It shows that these four models have

similar error-rising rates (slopes) when the forecasting step increases from 2SA to 4SA.

However, the BPNN has the steepest slope when the forecasting step increases from

4SA to 6SA, which indicates the static neural network fails to extract the dynamic

characteristics of time variation for MSA forecasting. Alternatively, the proposed

R-RTRL NN has similar error-rising rate to the Elman NN but has much smaller

forecast errors than the Elman NN for 2SA, 4SA and 6SA forecasts. It appears that the

proposed methodology can adequately utilize the closest antecedent information to

effectively re-adjust synaptic weights. As a result, the constructed R-RTRL NN

significantly diminishes time-lag effects and effectively provides much better and

adequate MSA forecasts.

3.2 Real-time multi-step-ahead water level forecasting by recurrent neural networks for urban flood control

In this case study, various ANNs are used to make water level forecasts for

representing the behavior of the rainfall-sewer flow processes in storm events. Flood

levels can be forecasted on the basis of (a) rainfall data; (b) previous water levels; and (c)

a combination of both data sets. Three ANNs coupled with statistical techniques are

adopted to construct real time multi-step-ahead FSP forecasting models. The
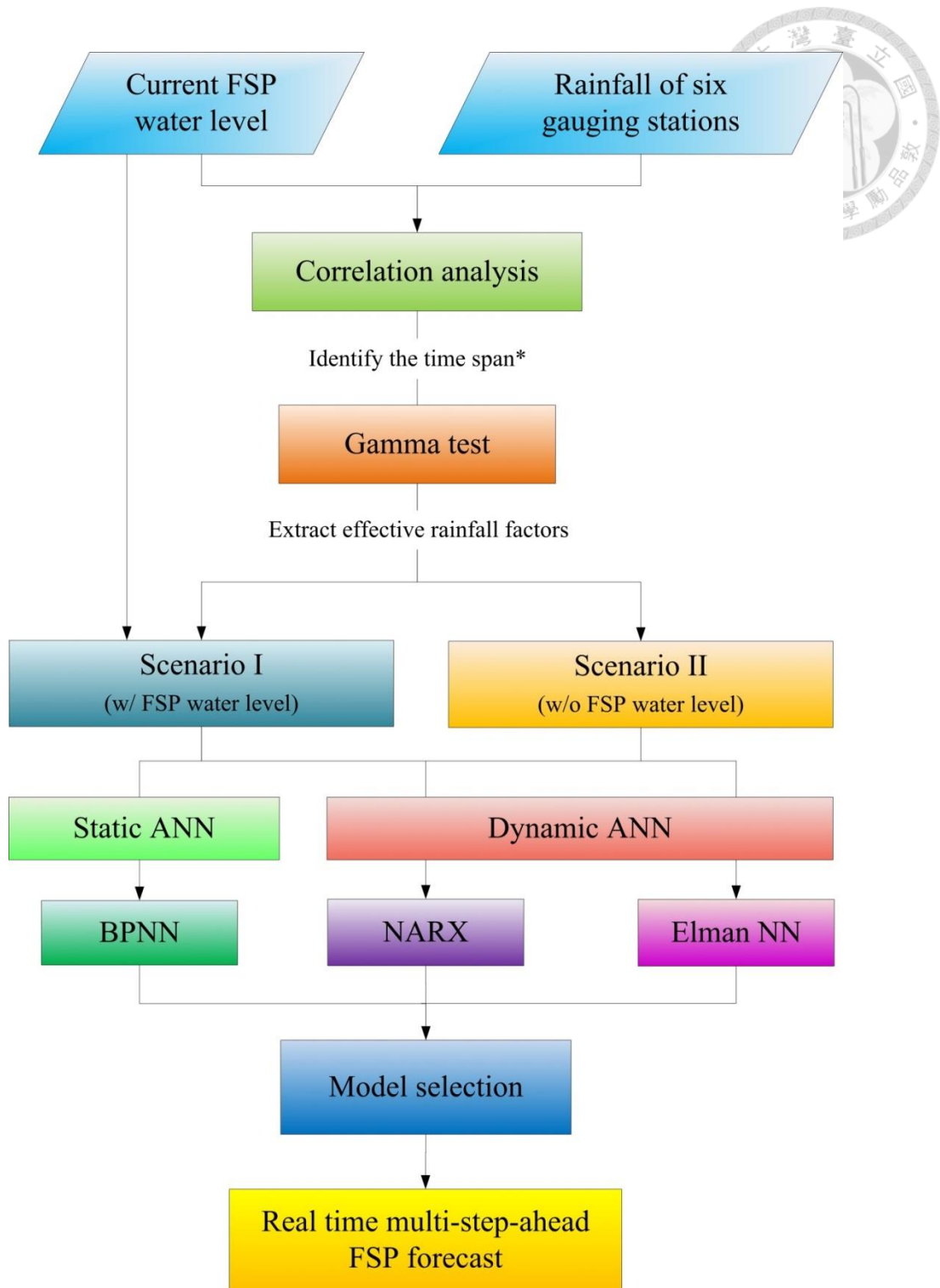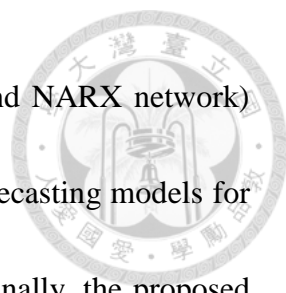
Fig. 3.7 Study flow of real-time MSA water level forecasting.

implementation procedure is shown in Fig. 3.7. The time span of rainfall affecting the

rise of FSP water level is first identified by the correlation analysis. Next the GT is

applied to extracting effective rainfall factors from all possible rainfall-related input

combinations. One static (BPNN) and two dynamic (Elman NN and NARX network) neural networks are proposed to construct MSA FSP water level forecasting models for two scenarios (w/ and w/o current FSP water level information). Finally, the proposed SDM with core techniques (GT and NARX network) and two other ANN models (Elman NN and BPNN) are evaluated by performance criteria. Because the main purpose of this case study is to identify the effectiveness between static and dynamic network models as well as different types of recurrent connections, the static BPNN and the dynamic Elman NN and NARX network that have different types of recurrent connections are selected for comparison purpose. These three network models are trained through batch learning algorithms. Therefore, the proposed R-RTRL NN that has the same type of recurrent connections as the Elman NN but is trained by a unique online learning algorithm is excluded in this case study.

A. Study area and dataset

Taiwan, an island located in the subtropical zone of the North Pacific Ocean, is covered with mountainous terrains and steep landforms. Taipei City, situated in the Taipei Basin of northern Taiwan, is surrounded by the Danshui River whose narrow estuary makes it difficult to discharge water effectively from the city. Consequently, the high levees along the Danshui River have been built to prevent outer flood into the city with a return period of two-hundred-year flood protection standard. Typhoons and/or
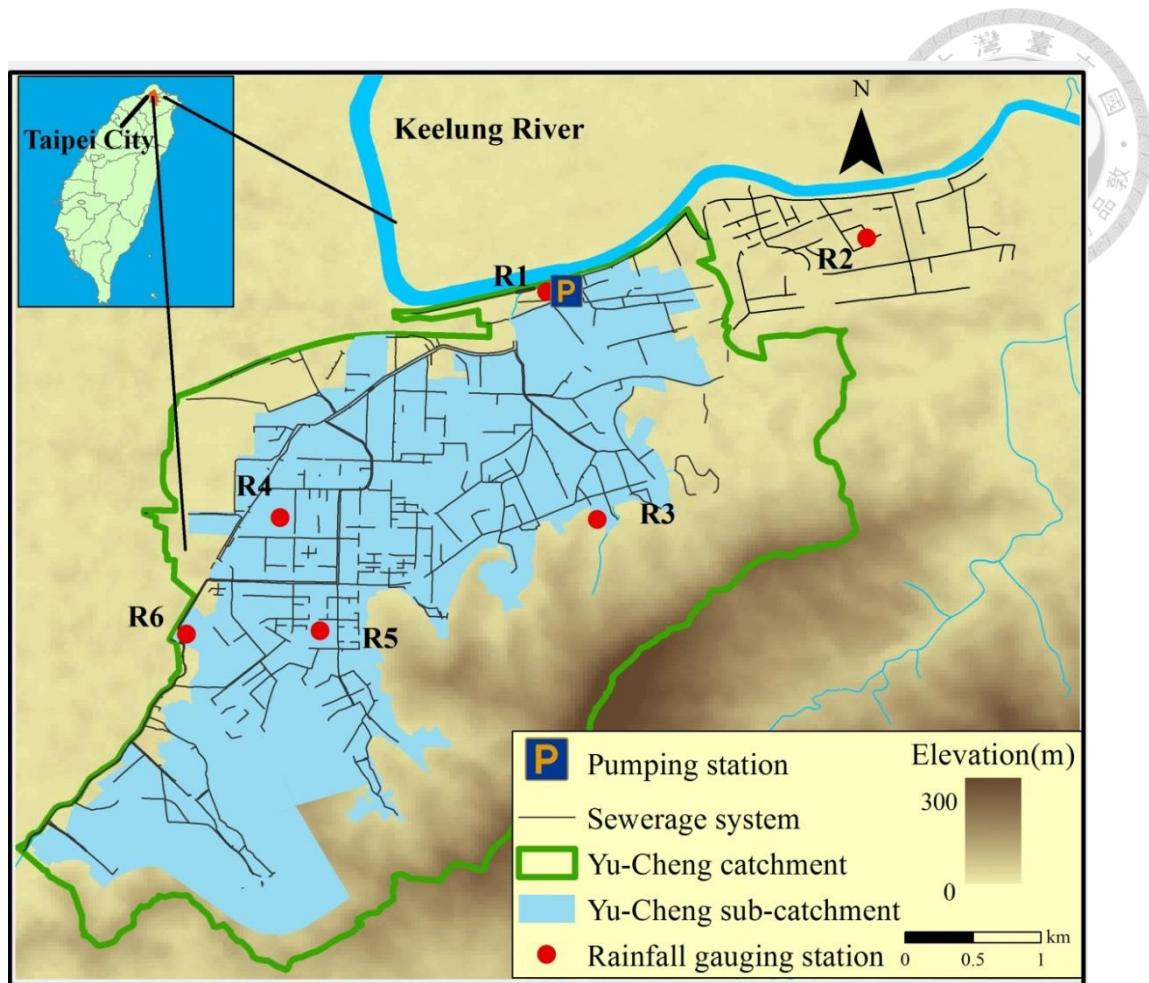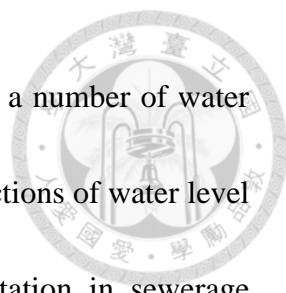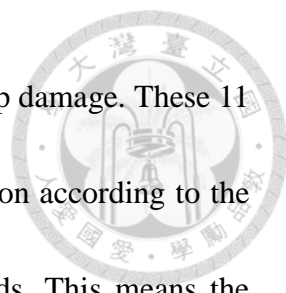
Fig. 3.8 Locations of the Yu-Cheng catchment and rainfall gauging stations.

heavy rainfall events are usually coupled with intensive rainfalls and thus easily cause urban flooding within a few hours, even within a few minutes, in Taipei City. Because of the high levees, the main threat to the city now turns out to be the floodwater inside the levee system. Therefore, pumping stations play an important role in managing internal stormwater flows for urban flood control. The Yu-Cheng catchment, located in southeastern Taipei, is selected as the study area (Fig. 3.8). There are six rainfall gauging stations (R1-R6, denoted as red dots in Fig. 3.8). Although station R2 is out of the catchment, it still belongs to a sewerage system that diverts floodwater to the

Yu-Cheng pumping station and the Keelung River. There also exist a number of water level gauging stations in this study catchment. However, the malfunctions of water level gauges caused by their collisions with unknown objects and siltation in sewerage systems raise the difficulty in the maintenance of water level gauges and their on-line monitoring. The water level data collected from the sewerage system are neither stable nor accurate, which means the FSP water level forecasting models for the Yu-Cheng Pumping Station would mostly rely on the rainfall information retrieved from its neighboring rainfall gauging stations.

The Yu-Cheng catchment occupies an area of about 1627 ha and owns the biggest drainage system in Taipei City. The Yu-Cheng Pumping Station was built in 1987 to drain or pump the internal stormwater flows into the Keelung River, a chief tributary of the Danshui River, and it was considered the most advanced and the largest pumping station in Asia in the 1980s. The pumping station is currently equipped with 11 pumps reaching a total pumping capacity of 234.1 cms, and the operation of the pumping station highly depends on the FSP water level information. If the FSP water level rises up to the warning level (1.8 m) during heavy rainfall or typhoon events, duty pumps are activated with a 3-minute warm up. Then stromwater starts to be pumped from the FSP into the Keelung River when the FSP water level reaches the start level (2.2 m). The start level is the lowest water level designed for the start of stormwater pumping as well

as for the prevention against the idle running of pumps to avoid pump damage. These 11

pump units operate independently and maintain a sequential operation according to the

laddered FSP water levels during typhoon or heavy rainfall periods. This means the

pumping operation begins with one pump unit, and only one pump unit, instead of all

remaining pump units, will join the operation at a time if the next higher rung of the

laddered water levels is reached. On the contrary, running pumps will be shut down

sequentially as the FSP water level decreases to the next lower rung. The operational

procedure of the Yu-Cheng Pumping Station is quite different from those of Hong Kong,

Tokyo or Singapore, where all the pump units are activated at the beginning if the FSP

water level exceeds the start level, then the pump units stop running as the FSP water

level drops to the stop level (DSD, 2000; PUB, 2013; Tamoto et al., 2008). It suggests

that the pumping operation in Taipei City is much more sensitive to the fluctuation of

FSP water level than those of big cities in Monsoon Asia.

Data of FSP water levels and rainfall at stations R1-R6 were collected with a

temporal resolution of 10 minutes from 13 typhoon and heavy rainfall events during

2004 and 2013. A total of 1985 datasets are used for constructing forecasting models in

this study, and the numbers of datasets allocated into training, validation and testing

stages are 826 (from 6 events), 651 (from 3 events) and 508 (from 4 events)

accordingly.

Table 3.6 Summary statistics for FSP water levels (m) and the peaks of average rainfall (mm/10 min)

Event Configuration	Model Stage	Number of data	Peak of average rainfall intensity (mm/10min)	Max FSP water level (m)	Mean FSP water level (m)	Standard deviation (m)
1			15.3	5.68	3.12	1.04
2			8.43	3.08	2.50	0.24
3	Training	826	8.03	2.73	2.25	0.28
4			3.37	2.40	2.07	0.14
5			2.26	2.41	2.13	0.09
6			5.17	2.57	1.79	0.37
7			11.0	3.69	2.23	0.48
8	Validation	651	10.1	2.68	2.05	0.39
9			5.91	2.85	2.07	0.31
10			5.47	2.84	2.17	0.18
11	Testing	508	6.99	2.69	2.25	0.19
12			4.92	2.50	2.08	0.26
13			12.4	4.59	2.57	0.55

Such allocation is made to maintain similar statistic characteristics in these three datasets in consideration of the summary statistics of the 13 events shown in Table 3.6. In addition, the weighted average rainfall (R_{avg}) over the Yu-Cheng catchment is computed by the Thiessen polygon method and is also considered as a potential input to the forecasting models. Furthermore, because the original FSP water level is indeed affected by the operation of pumping units, the original FSP water levels were recovered prior to model construction according to a recovery equation (provided by the Taipei City Government) that involves pumping capacity and pumping-affected area. The summary statistics for FSP water levels and rainfall datasets are presented in Table 3.6.
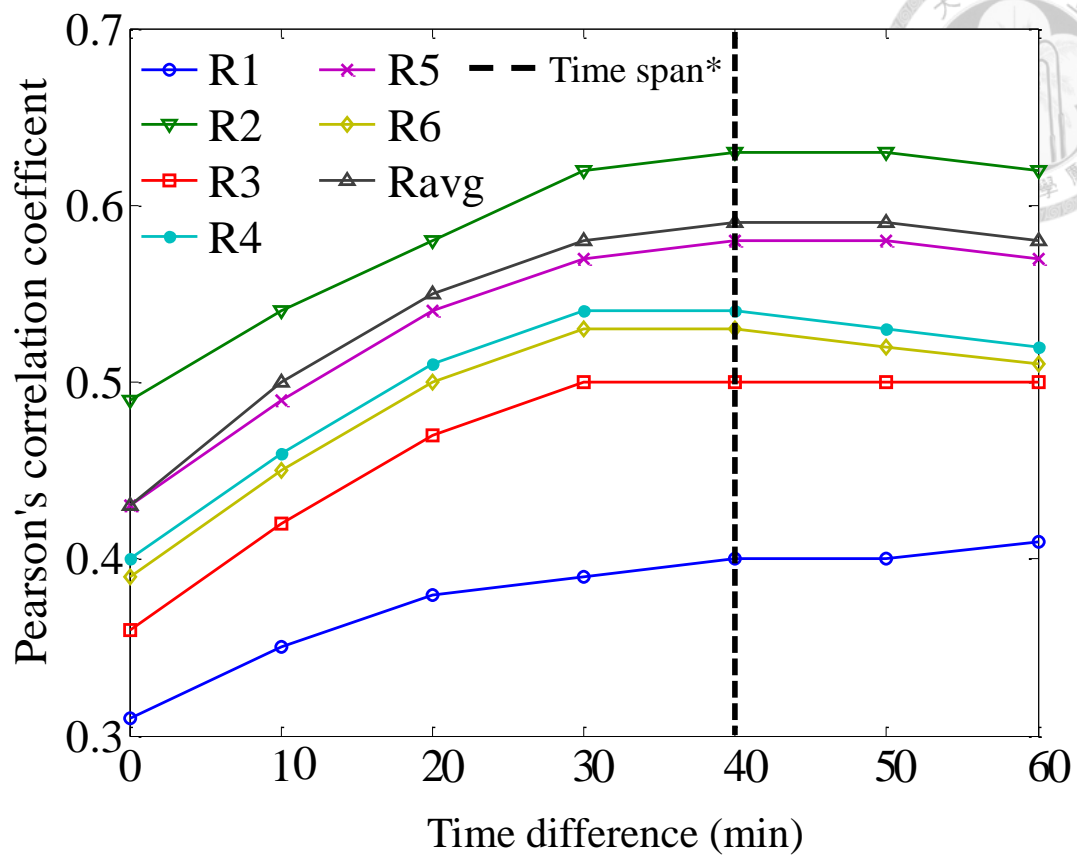
Fig. 3.9 Correlation analysis between FSP water levels and rainfall gauging stations in different time steps.

*time span of rainfall affecting the rise of FSP water level

B. Identification of the time span of rainfall affecting the rise of the FSP water level
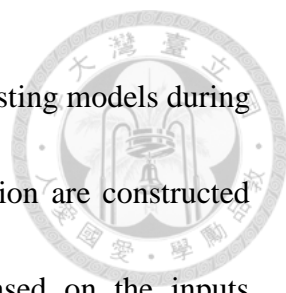
For constructing a rainfall-sewer flow model, the first step is to identify the temporal impacts of rainfall on the rise of FSP water level. In this study, the Pearson's correlation coefficient is applied to learning the linear relationship and the recognition of the highest correlations between FSP water level and rainfall at different time lags for each station (R1-R6) as well as the weighted average rainfall (R_{avg}) over the Yu-Cheng

catchment. The results shown in Fig. 3.9 indicate that it consistently takes about 40 minutes for rainfall at stations R1-R6 to cause an increase in the FSP water level, similarly for the R_{avg}. It is worthy to note that, in contrast to river channels, a sewerage system can be implicitly considered as a small-scale volume control system on account of the relatively small catchment with which the system was associated. The variation of the FSP water level is mainly affected by the rainfall aggregated within a short period of time in the catchment. As a result, the time span of rainfall affecting the rise of FSP water level at the Yu-Cheng Pumping Station is set as 40 minutes. It is noted that "time span" is used in this study while "concentration time" is usually used in river channel studies.

C. Extraction of effective rainfall factors and model construction

The Pearson's correlation coefficients between the FSP water level and rainfall at gauging stations R1-R6 as well as R_{avg} are not very high but quite similar (ranging from 0.41 to 0.63), which could be due to the lumped effect of rainfall falling to the catchment and the complex interactions between rainfall and sewer flow. In order to identify effective rainfall stations that significantly affect the fluctuations of FSP water level for modeling purpose, the GT is implemented in this study. That is to say, rainfall-related inputs to the estimation models of FSP water level is determined by the GT.

In this case study, one- to six-step-ahead FSP water level forecasting models during heavy rainfall and typhoon events for the Yu-Cheng Pumping Station are constructed through the BPNN, the Elman NN and the NARX network based on the inputs determined by the GT. The practical meaning and contribution of three forecasting models will be surveyed under two scenarios: (1) the information of current FSP water level is available (denoted as scenario I hereinafter); (2) the information of current FSP water level is not available (denoted as scenario II hereinafter). The applicability and reliability of these three constructed forecasting models at different forecasting steps are evaluated by the RMSE, CC and CE.

D. Results and discussion

This section presents the selection result of effective rainfall factors and the forecast performance of the static (BPNN) and dynamic (Elman NN and NARX network) neural networks in two scenarios (w/ and w/o current FSP water level information). The results and discussion are addressed in details, shown as follows:

1) Identification of effective rainfall stations

For extracting effective rainfall factors, data of the antecedent 40-minute rainfall collected at six gauging stations together with the average rainfall ($R1(t-4)$-$R6(t-4)$, $R_{avg}(t-4)$) are first scaled to [-1,1]. Then a total of 127 ($=2^7-1$) Γ values corresponding
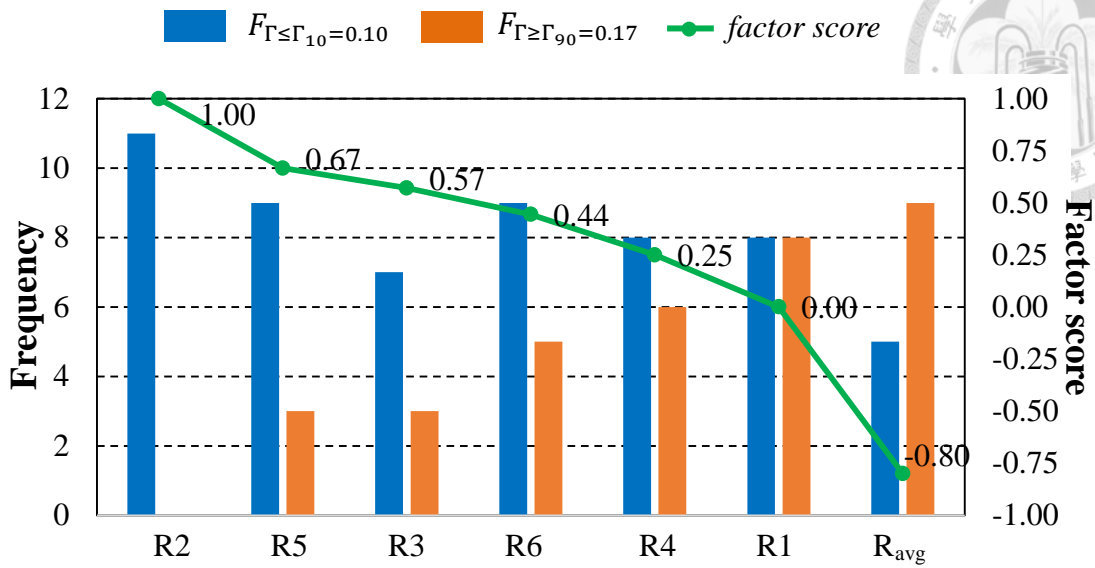
Fig. 3.10 Determination of effective rainfall stations by the GT results.

to all possible rainfall-related input combinations are calculated through the GT. The produced Γ values are next sorted in an ascending order, in which Γ values smaller than the 10th percentile ($\Gamma_{10} = 0.10$) are classified as the best group ($F_{\Gamma \leq \Gamma_{10}}$), whereas those bigger than the 90th percentile ($\Gamma_{90} = 0.17$) are classified as the worst group ($F_{\Gamma \geq \Gamma_{90}}$). Fig. 3.10 shows the result of the GT, where blue bars represent the occurrence frequency of factors in the best group ($F_{\Gamma \leq \Gamma_{10}}$) while orange bars represent the occurrence frequency of factors in the worst group ($F_{\Gamma \geq \Gamma_{90}}$), and factor scores calculated by Eq. (47) for each rainfall factors are drawn into a green dotted line.

$$factor\ score = 1 - \frac{F_{\Gamma \geq \Gamma_{90}}}{F_{\Gamma \leq \Gamma_{10}}} \tag{47}$$

where the factor score ranges from $-\infty$ to 1.
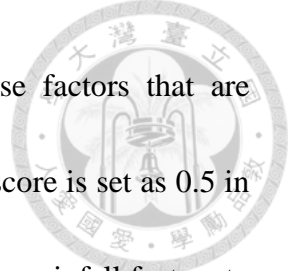
Therefore, effective rainfall factors can be identified as those factors that are associated with higher factor scores, and the threshold of the factor score is set as 0.5 in this study. Consequently, R2, R3 and R5 are identified as the effective rainfall factors to be used in the forecasting models.

2) *Performance of FSP water level forecasting in scenario I: current FSP water level is available*

In scenario I, data of the current FSP water level and rainfall of R2, R3 and R5 are utilized to construct 10- to 60-min-ahead (N = 1-6) FSP water level forecasting models through three ANNs. The input-output patterns of three ANN models can be represented as follows:

$$
\hat{WL}_{FSP}(t+N) = f\left[WL_{FSP}(t), R2(t+N-4), R3(t+N-4), R5(t+N-4)\right]
$$
$$
N \in 1-4(10\min)
$$
(48)

$$
\hat{WL}_{FSP}(t+N) = f\left[WL_{FSP}(t), R2(t), R3(t), R5(t)\right]
$$
$$
N \in 5-6(10\min)
$$
(49)

where $\hat{WL}_{FSP}(t+N)$ is the forecasted value at a lead time of N (10-min unit).

After implementing trial-and-error procedures for model configuration based on the training and validation data sets, the output-memory order q for NARX networks is 1 and all the three models are configured to have only one hidden layer with 2-4 nodes for different forecasting steps. Summarized results are presented in Tables 3.7.

Table 3.7 Model performance of one- to six-step-ahead forecasting for FSP water levels in scenario I

Time step	Number of nodes	Model stage	BPNN			Elman NN			NARX		
			RMSE (m)	CC	CE	RMSE (m)	CC	CE	RMSE (m)	CC	CE
t+1	2	Training	0.07	0.99	0.99	0.07	0.99	0.99	0.07	0.99	0.99
		Validation	0.07	0.98	0.96	0.07	0.98	0.97	0.07	0.98	0.96
		Testing	0.10	0.96	0.92	0.09	0.96	0.93	0.09	0.97	0.93
t+2	2	Training	0.12	0.98	0.96	0.11	0.98	0.97	0.11	0.98	0.96
		Validation	0.12	0.94	0.89	0.10	0.96	0.91	0.12	0.94	0.89
		Testing	0.17	0.90	0.77	0.15	0.91	0.82	0.16	0.91	0.79
t+3	2	Training	0.14	0.97	0.95	0.14	0.97	0.95	0.14	0.97	0.95
		Validation	0.14	0.91	0.83	0.14	0.92	0.84	0.15	0.90	0.81
		Testing	0.18	0.89	0.71	0.18	0.88	0.72	0.18	0.88	0.72
t+4	2	Training	0.15	0.97	0.94	0.15	0.97	0.94	0.18	0.96	0.91
		Validation	0.16	0.88	0.78	0.17	0.88	0.77	0.20	0.82	0.65
		Testing	0.21	0.87	0.62	0.19	0.88	0.67	0.18	0.86	0.70
t+5	3	Training	0.20	0.95	0.89	0.19	0.95	0.90	0.17	0.96	0.92
		Validation	0.19	0.85	0.70	0.17	0.87	0.75	0.19	0.85	0.70
		Testing	0.22	0.82	0.57	0.20	0.83	0.65	0.19	0.87	0.69
t+6	4	Training	0.24	0.92	0.84	0.22	0.94	0.87	0.16	0.96	0.93
		Validation	0.19	0.85	0.70	0.19	0.84	0.69	0.22	0.78	0.58
		Testing	0.23	0.77	0.52	0.22	0.79	0.58	0.19	0.88	0.67

Results indicate that the three comparative models perform rather consistently in the training and validation stages while both dynamic neural networks (the Elman NN and the NARX network) perform better than the static one (the BPNN) in the testing stages. Besides, the NARX network outperforms the Elman NN as the forecasting step exceeds
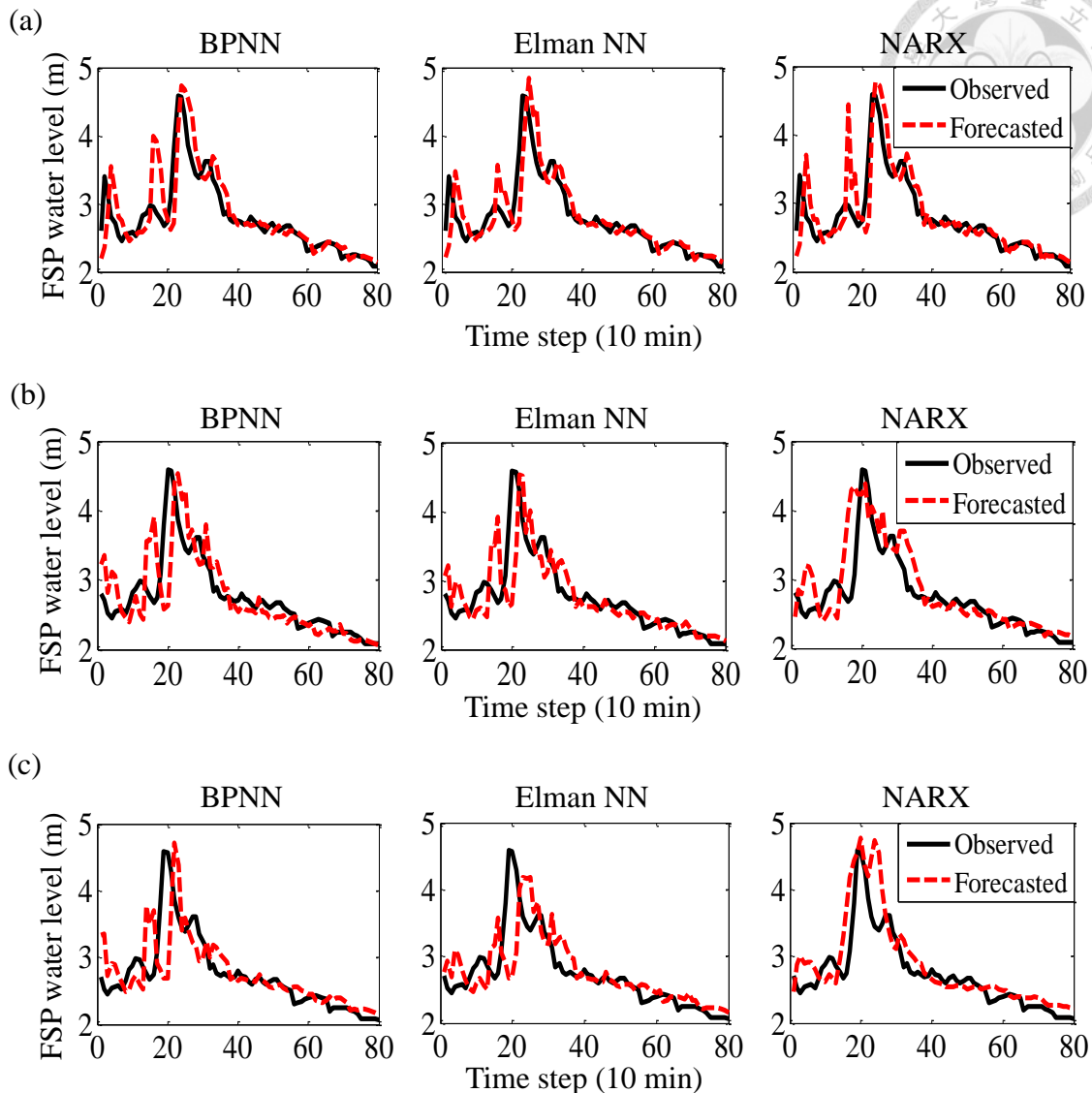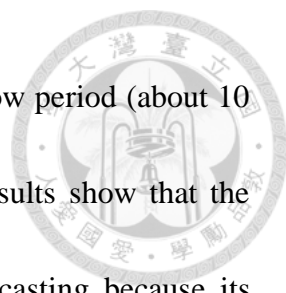
Fig. 3.11 (a) 20, (b) 50 and (c) 60-min-ahead forecasting of the 612 heavy rainfall event for scenario I with respect to the BPNN, the Elman NN and the NARX network.

four, and it even produces a high CE value (close to 0.7) as the forecasting step reaches

six (60-min-ahead forecasting). The 612 heavy rainfall event (12.4 mm/10 min; 54.1

mm/hr) with the highest peak FSP water level above 4.5 m is selected to illustrate the

hydrographs of observed versus 20-, 50- and 60-min-ahead forecasted FSP water levels

in the testing stages of three models (Fig. 3.11). During the peak flow period (about 10 hours) of this event, 6 up to 11 pumping units were operated. Results show that the Elman NN produces the best performance for 20-min-ahead forecasting because its over-estimation between 10 and 20 time steps is comparatively less serious. However, the NARX network can significantly mitigate time-lag problems at peak values and well forecast the 50- and 60-min-ahead FSP water level, whereas the other two comparative models not only have significant time-lag phenomena but fail to well forecast 50- and 60-min-ahead water levels, in which fluctuations occur near peak values.

Fig. 3.12 (a) shows the CE of 10- to 60-min-ahead forecasting in the testing stages of three models in scenario I. The three network models perform equally well for one- to three-step-ahead forecasting, whereas significant differences among their performances are found as the forecast time step exceeds four (40 min). The reason is that the time span of rainfall affecting the rise of the FSP water level is 40 minutes such that the rainfall-water level processes at 1-4 time steps could be suitably presented by rainfall input data and FSP water level output data (Eq. (7)). In addition to the fact that the persistence of FSP water level decreases as the time step increases, the time lag of rainfall becomes significant as the forecast time step exceeds four. As a result, both conditions would cause a reduction in forecast accuracy. It clearly indicates that the NARX networks produce much higher CE values than the other two models for

(a)

CE

(b)
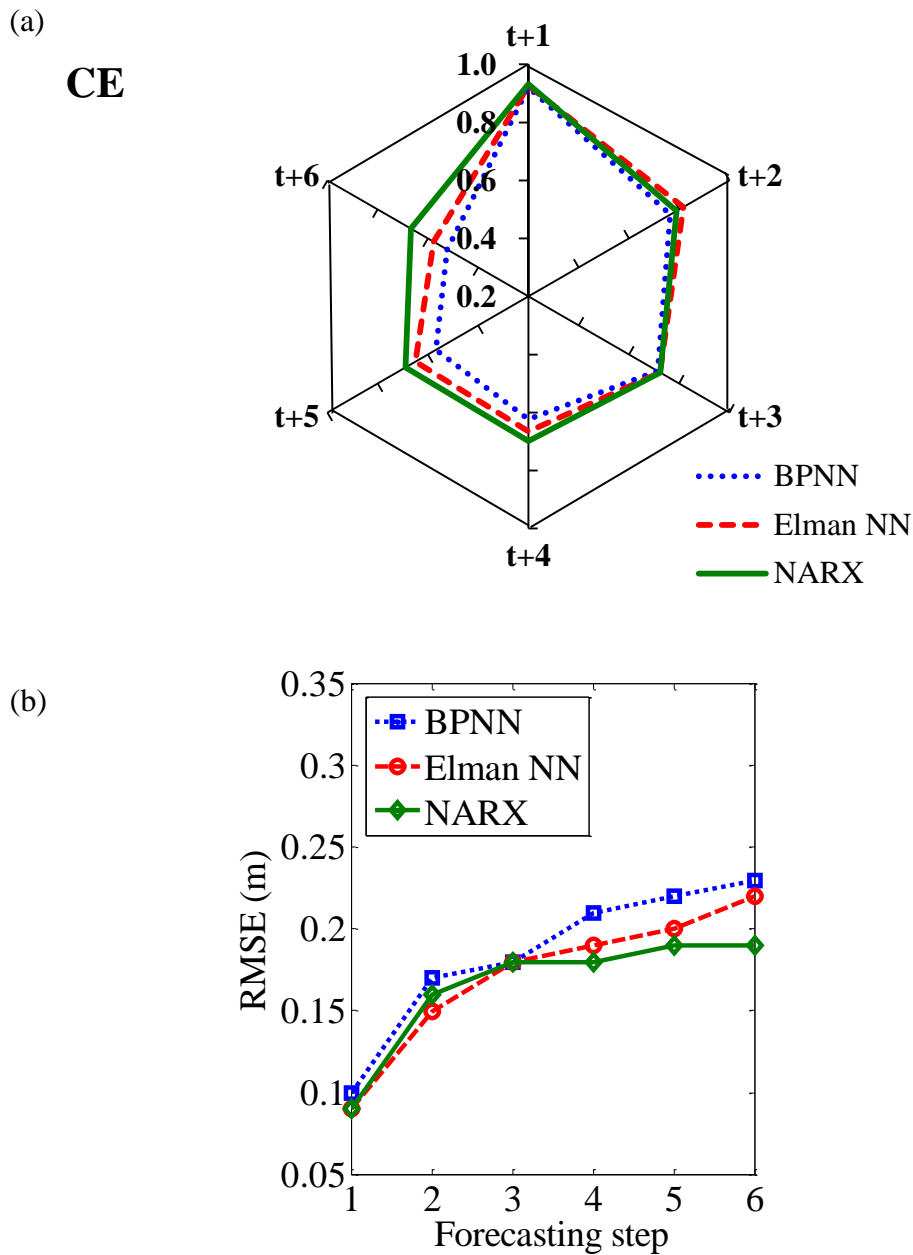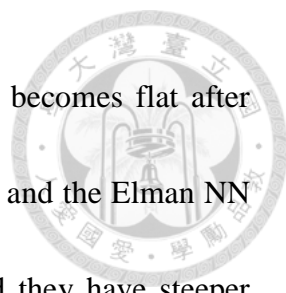
Fig. 3.12 (a) CE of 10- to 60-min-ahead forecasting and (b) relationship between FSP water level forecast errors (RMSE) and forecasting steps with respect to three forecasting models in the testing stages for scenario I.

four- to six-step-ahead forecasting, whereas the Elman NNs perform slightly better than

the BPNNs.

Finally, the relationship between forecast errors (RMSE) and forecasting steps of

these three models is presented in Fig. 12 (b). The RMSE trend of the NARX network

model increases gradually as the forecasting step increases, and it becomes flat after

three forecasting steps. Nevertheless the RMSE trends of the BPNN and the Elman NN

models significantly increase as the forecasting step increases, and they have steeper

slopes than that of the NARX network after three forecasting steps. The results provide

evidence that with the feedbacks of imperfect outputs representing the information the

closest to the forecasting horizon to the input layer, the NARX network can effectively

adopt extra information to promote the accuracy and reliability of multi-step-ahead FSP

water level forecasting.

3) Performance of FSP water level forecasting in Scenario II: current FSP water level

is unavailable

In the flood control centre of the Taipei City Government, the datasets of the

current FSP water levels at sixty-five pumping stations during typhoon events are

transmitted only through two channels of radio waves, and thus the backend system of

the flood control centre may not successfully receive the current FSP water level

information every ten minutes. Furthermore, the preliminary correlation analysis results

indicate that the time span between the FSP water level and rainfall over the study

catchment is about 40 minutes. An alternative to the forecasting model of scenario I is

considered essential for auxiliary purposes. In this scenario (II), FSP water level
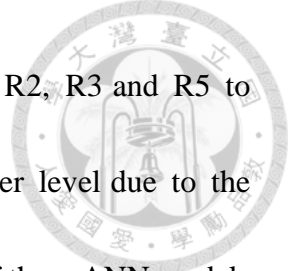
forecasting models are constructed based only on the rainfall of R2, R3 and R5 to

prevent any possible delay in the receipt of the current FSP water level due to the

unstable frequency of data transmission. The input-output patterns of three ANN models

can be represented as follows:

$$
\hat{WL}_{FSP}(t+N) = f\left[R2(t+N-4), R3(t+N-4), R5(t+N-4)\right]
$$
$$
N \in 1-4(10\,\text{min})
$$
(50)

$$
\hat{WL}_{FSP}(t+N) = f\left[R2(t), R3(t), R5(t)\right]
$$
$$
N \in 5-6(10\,\text{min})
$$
(51)

where $\hat{WL}_{FSP}(t+N)$ is the forecasted value at a lead time of N (10-min unit).

After implementing trial-and-error procedures for model configuration based on

the training and validation data sets, the output-memory order q for NARX networks

is 1 and all the three models are configured to have only one hidden layer with 2-4

nodes for different forecasting steps. Summarized results are presented in Tables 3.8.

Results indicate that the NARX networks significantly outperform the other two

network models in terms of lower RMSE values and higher CC and CE values in all

three stages (training, validation and testing) for one- to six-step-ahead forecasting. It is

noted that the performance of three ANN models in scenario II is not as good as that of

scenario I. The reason is that only rainfall information is utilized as model inputs in

scenario II, while another important factor, i.e., the persistent effect (auto-regression) of

Table 3.8 Model performance of one- to six-step-ahead forecasting for FSP water levels in scenario II

Time step	Number of nodes	Model stage	BPNN			Elman NN			NARX		
			RMSE (m)	CC	CE	RMSE (m)	CC	CE	RMSE (m)	CC	CE
t+1	3	Training	0.44	0.71	0.49	0.39	0.78	0.59	0.27	0.90	0.81
		Validation	0.35	0.46	0.01	0.33	0.42	0.13	0.31	0.56	0.25
		Testing	0.26	0.68	0.44	0.29	0.62	0.30	0.21	0.81	0.63
t+2	2	Training	0.48	0.74	0.39	0.39	0.78	0.59	0.27	0.90	0.81
		Validation	0.34	0.39	0.07	0.32	0.44	0.15	0.29	0.61	0.29
		Testing	0.26	0.66	0.43	0.27	0.66	0.37	0.20	0.84	0.67
t+3	2	Training	0.42	0.75	0.53	0.38	0.79	0.60	0.24	0.92	0.84
		Validation	0.33	0.44	0.06	0.32	0.43	0.13	0.31	0.62	0.21
		Testing	0.26	0.66	0.41	0.28	0.65	0.34	0.21	0.80	0.64
t+4	2	Training	0.47	0.74	0.40	0.39	0.78	0.60	0.25	0.91	0.83
		Validation	0.32	0.42	0.14	0.32	0.42	0.12	0.29	0.62	0.31
		Testing	0.27	0.64	0.38	0.27	0.66	0.35	0.19	0.83	0.67
t+5	3	Training	0.48	0.62	0.38	0.39	0.79	0.59	0.26	0.91	0.82
		Validation	0.37	0.38	-0.17	0.32	0.45	0.16	0.29	0.61	0.29
		Testing	0.28	0.58	0.30	0.30	0.60	0.21	0.21	0.82	0.59
t+6	4	Training	0.48	0.64	0.38	0.43	0.71	0.50	0.26	0.91	0.82
		Validation	0.36	0.41	-0.07	0.33	0.46	0.09	0.31	0.52	0.18
		Testing	0.28	0.56	0.27	0.29	0.51	0.25	0.24	0.71	0.47

the FSP water level, is not considered (or unavailable) in this circumstance. Under such condition, the NARX network equipped with recurrent connections from imperfect outputs can produce much more satisfactory results than the Elman NN and the BPNN.

Similar to that of scenario I, an analysis is conducted on the 612 heavy rainfall event for scenario II. The rainfall input datasets from three gauging stations and
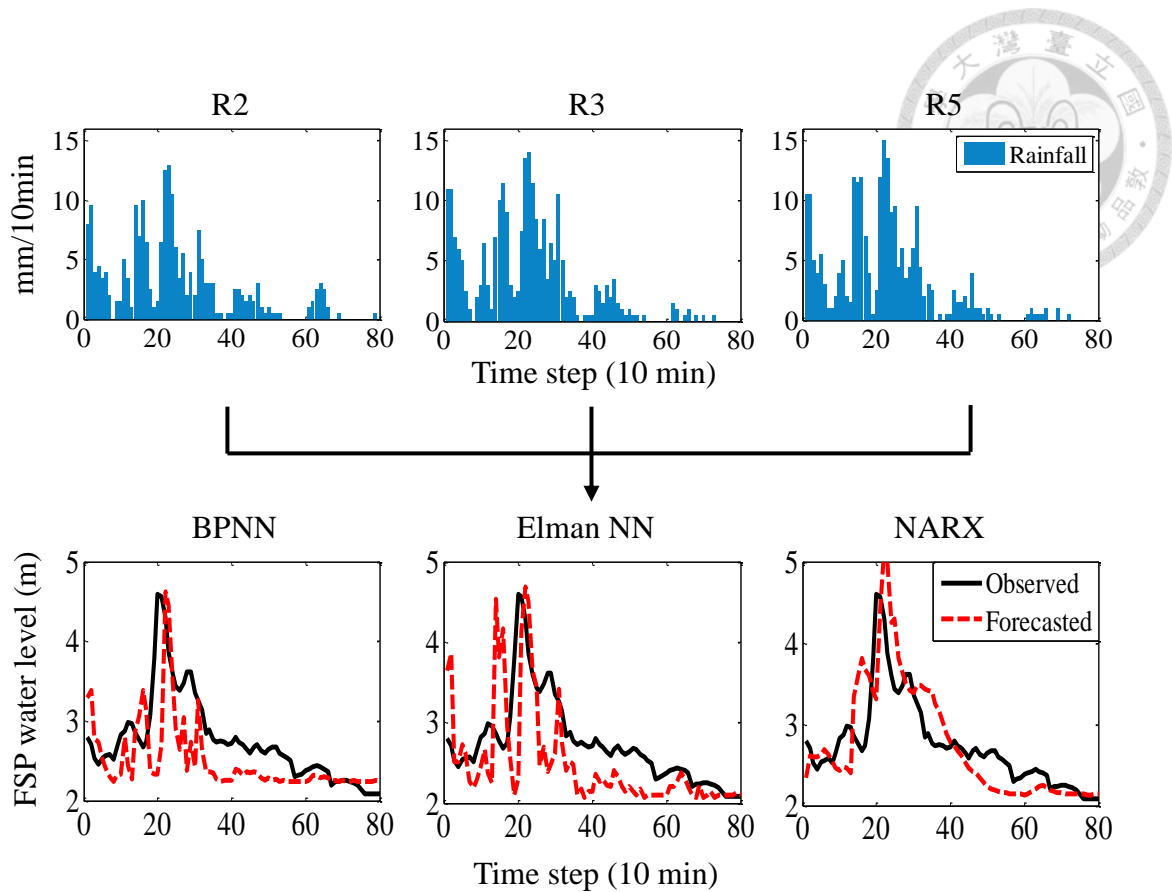
Fig. 3.13 Rainfall input datasets from three gauging stations and 50-min-ahead forecasting of the 612 heavy rainfall event for scenario II with respect to the BPNN, the Elman NN and the NARX network.

50-min-ahead forecasting are illustrated in Fig. 3.13. It demonstrates that the NARX network can well forecast the 50-min-ahead FSP water level and maintain the water level trail with less fluctuation than the BPNN and the Elman NN. The strong fluctuations occurring in the hydrographs associated with the BPNN and the Elman NN are mainly because these two models are driven only by the rainfall-related inputs that originally bear high variations, whereas the NARX network facilitates extra input information from the previous forecasted FSP water level to smooth the fluctuations of

the forecasted hydrograph.

Fig. 3.14(a) shows the CE of 10- to 60-min-ahead forecasting in the testing stages of three network models in scenario II. It clearly indicates that the NARX networks produce much higher CE values than the other two network models, while the Elman NNs perform even worse than the BPNNs, which implies the recurrent connections from the hidden layer of each Elman NNs magnify the highly variable rainfall information and thus do not increase the reliability of the Elman NNs.

Fig. 3.14(b) illustrates the relationship between forecast errors (RMSE) and forecasting steps of these three models. The results show that the NARX network produces much lower RMSE values than the other two models for one- to six-step-ahead forecasting, and the RMSE values of three models are relatively consistent (flat) for one- to four-step-ahead forecasting. For the NARX network, the RMSE value is the lowest at the 4th forecasting step and starts fast rising afterward. This is mainly because only current rainfall information is available as the forecasting step increases to five and six steps, which significantly causes the degradation of model performance at the 5th and 6th forecasting steps. This phenomenon is consistent with the 40-min time span of rainfall affecting the rise of the FSP water level, which is determined by the correlation analysis shown in Fig 3.9.

(a)

CE

(b)
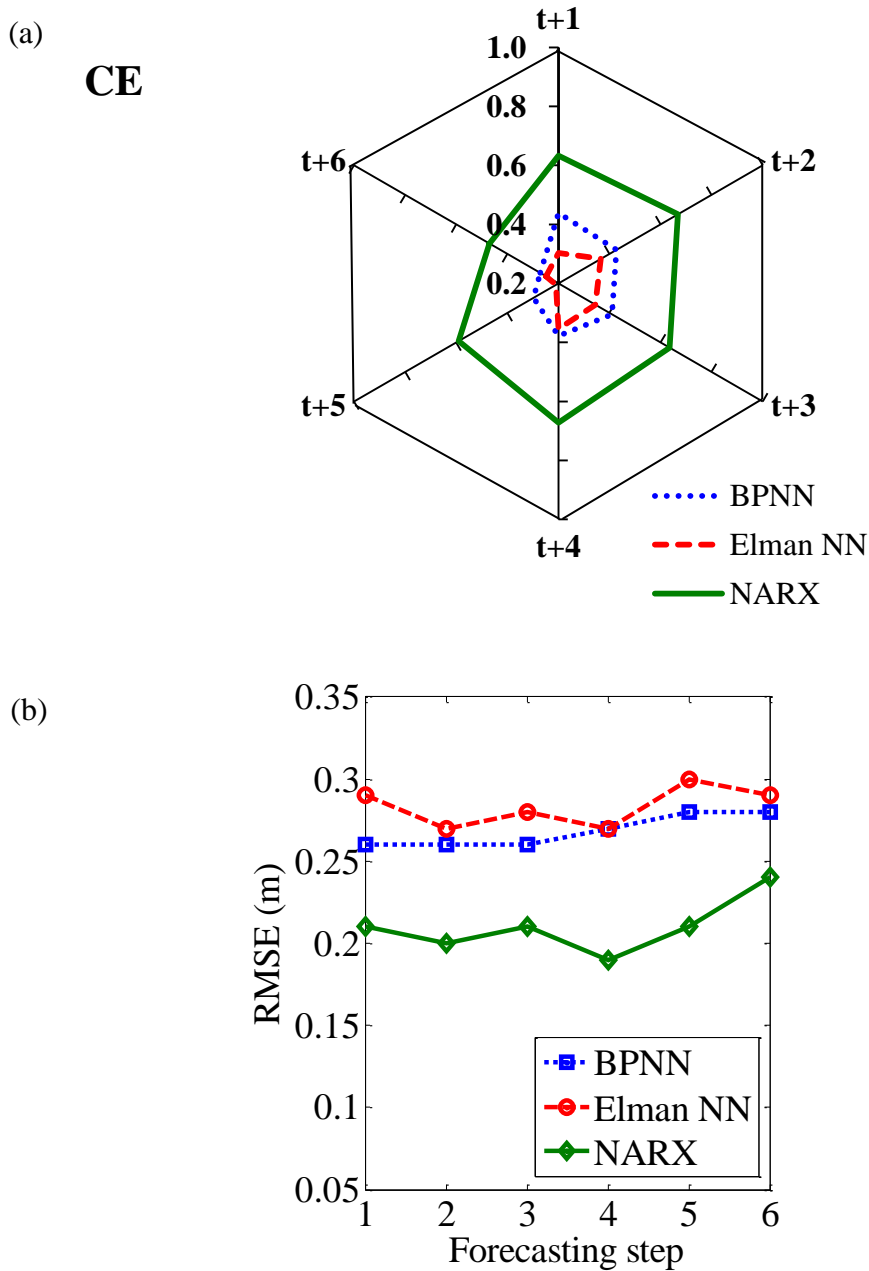
Fig. 3.14 (a) CE of 10- to 60-min-ahead forecasting and (b) relationship between FSP
water level forecast errors (RMSE) and forecasting steps with respect to three
forecasting models in the testing stages for scenario II.
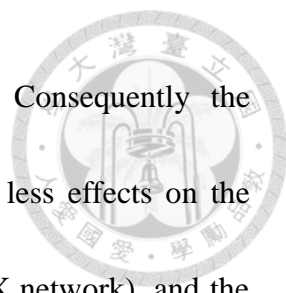
In brief, the NARX network implemented without the input of the current FSP
water level information can still provide reasonable and reliable multi-step-ahead FSP
water level forecasts, and therefore this model can be adopted for auxiliary purposes.

4) Summary of forecast performance

We explore the explanatory power in multi-step-ahead forecasting for FSP water levels through one static neural network (BPNN) and two dynamic neural networks (NARX network and Elmann NN). Results of scenario I demonstrate that the static network is inferior to the dynamic ones because the inputs of the static network depend solely on observed data, whereas those of the dynamic networks incorporate observed data with time delay units through recurrent connections either from the output layer (NARX network) or from the hidden layer (Elman NN), which makes significant contribution to the forecasted values.

We further explore the ability of these two recurrent neural networks (NARX network and Elman NN) to solve the problem of long–term dependencies in a time series. We find that although the NARX network may not completely circumvent this problem, it can much effectively discover the long–term dependencies through its recursive outputs and mitigate the fluctuation problem (instability) in its output. As comparing the results obtained from scenario II, it is easy to tell that the recurrent connection from the output layer of the NARX network is more effective in modeling the long-term dependencies than the recurrent connection from the hidden layer of the Elman NN. This result provides extra evidence that the backward connections from the hidden layer (Elman NN) could only maintain the previous values of the hidden units

with an emphasis on the highly variable rainfall information. Consequently the backward connections from the hidden layer (Elman NN) impose less effects on the output than the backward connections from the output layer (NARX network), and the generalizability of the Elman NN is weaker than that of the NARX network and the BPNN.

We notice that for the NARX and the Elman NN the validation results are worse than the testing ones in all the cases of scenario II and in the t+4, t+5 and t+6 cases of scenario I. This is mainly because the variability of rainfall intensity (which also results in the higher variability of FSP water level) in the validation set is higher than that in the testing set (Table 3.6), and rainfall effect would decrease as the forecasting step exceeds the time span of rainfall affecting the rise of the FSP water level over the catchment (in our case: 40 minutes).

From Tables 3.7 and 3.8, we notice that all the three forecasting models perform better in scenario I (w/ current FSP water level) than in scenario II (w/o current FSP water level). This reveals that the FSP water level is the dominant factor for the forecasting process. In terms of CE values produced by the NARX network in scenario II (Table 3.8), we notice that the training cases perform equally well (0.81-0.84) for one- to six-step-ahead forecasting whereas the testing cases perform relatively poor (0.47-0.67). This is mainly because the observed input-output data (i.e., rainfall and

water level) in the training stages are used to train and optimize the corresponding

weights of the networks, whereas similar strategies are not implemented in the

validation and testing stages.

3.3 Regional estimation of groundwater Arsenic concentration through Systematical Dynamic-neural Modeling

To reliably modelling water quality, it is important to understand the impacts of

factors on real competence, the interaction and evolvement of factors within an

operation system, and the measurements of factors. In this case study, we aim to present

a novel model of case competence with good accuracy and predictability, in which

certain assumptions are made for the nature of cases. The proposed SDM incorporates

the GT into a dynamic neural network accompanied with three optional statistical

techniques to tackle regional estimation problems for water quality management, and its

implementation procedure is shown in Fig. 3.15. The SDM first effectively extracts the

non-trivial factors that significantly affect the fluctuations of As concentrations through

the GT. The NARX network is then utilized to obtain As concentration at ungauged

sites with inputs consisting of the extracted non-trivial factors and the estimated As

concentrations obtained from recurrent connections. The Bayesian regularization

method is configured to control the network complexity for preventing over-fitting. The

cross validation technique is used to produce a low-bias estimator of model
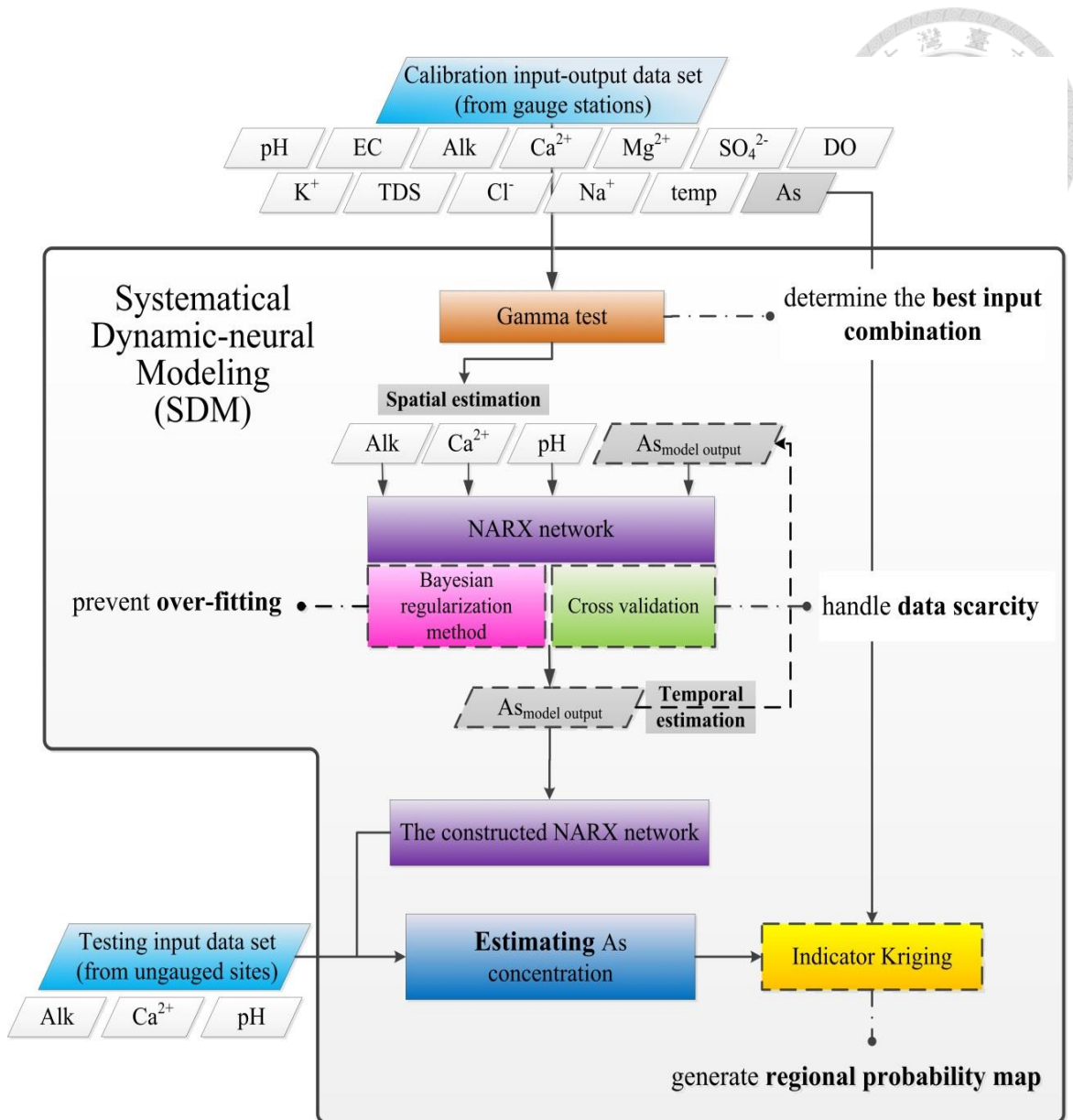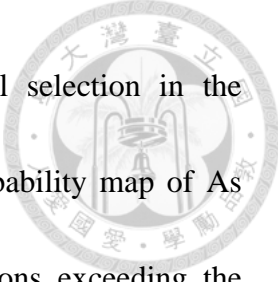
Fig. 3.15 Implementation procedure of the proposed SDM for regional estimation of As concentrations in groundwater

generalizability and thus provides a sensible criterion for model selection in the calibration stage. Finally, the IK is implemented to derive the probability map of As concentrations for detecting unsampled areas with As concentrations exceeding the WHO drinking water standard.

A. Study area

Yun-Lin County is located in the southwestern alluvial fan of the Chou-Shui River in Central Taiwan (Fig. 3.16). Based on hydrogeological settings, the southern Choushui River alluvial fan is classified mainly into the proximal-fan, the mid-fan and the distal-fan areas (Central Geological Survey, 1999), in which the coastal region of the Yun-Lin County is located in the distal-fan area. The hydrogeological formation of the distal-fan can be divided into six inter-layered sequences: three marine sequences; and three non-marine sequences. The non-marine sequences with coarse sediments (from medium sand to highly permeable gravel) are considered as aquifers, whereas the marine sequences with fine sediments are considered as aquitards. The annual average precipitation is 1417 mm, which mainly occurs during wet seasons (i.e., between May and September). Aquaculture is the primary revenue source for the inhabitants in the coastal region of Yun-Lin County. Due to high water demand but limited water supply,
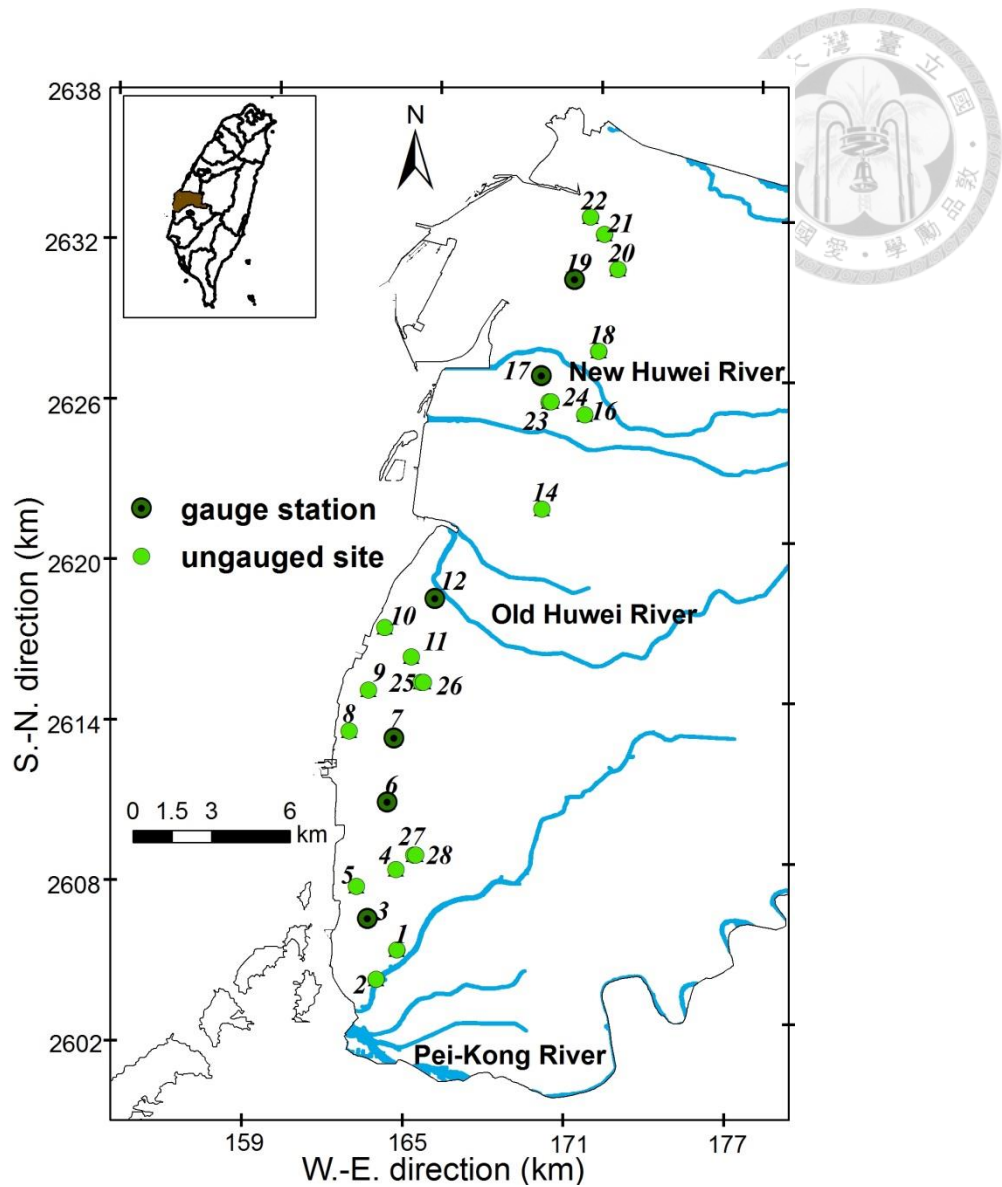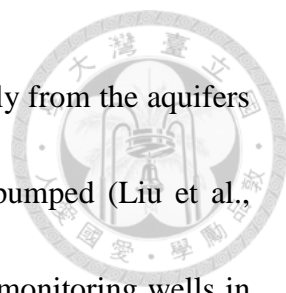
Fig. 3.16 Locations of twenty-six groundwater wells at the Yun-Lin coastal area, Taiwan.

groundwater has become a vital water resource in this area for decades. In 1992 the

Water Resources Agency installed 26 groundwater monitoring wells (well depths range

from 8m to 110m) distributed in this area for groundwater quality monitoring,

particularly for As pollution and other potential contamination in groundwater.

Approximately 757 million m^3 of groundwater was extracted annually from the aquifers in this area, of which 268 million m^3 was considered to be over-pumped (Liu et al., 2001). High As concentration (93.2±161 ugl^{-1}) was detected in the monitoring wells in this area (WHO drinking water standard: 10 ugl^{-1}). Liu et al. (2006) indicated that over-pumping groundwater induces dissolved oxygen and increases As mobility in water and thus the relatively high As content has accumulated and been deposited in the marine sequences with fine sediments.

B. Data collection and preliminary analysis

In this case study, sampling data of groundwater quality parameters were collected quarterly at 26 wells between 1992 and 1999, and the field sampling methods of As concentration was determined by hydride generation followed by atomic absorption spectroscopy, APHA Method 3500-arsenic Part B (APHA, 1992). The maintenance of groundwater monitoring wells is laborious and cost intensive, and therefore only six wells (#3, #6, #7, #12, #17 and #19) have continued monitoring groundwater quality after 1999. The proposed method intends to estimate the As fluctuations of 20 un-monitored wells based on other water quality parameters that are easier to measure. We assume that 20 un-monitored wells are ungauged sites and 6 monitored wells are gauge stations (Fig. 3.16).

Table 3.9 Statistics of groundwater quality parameters at six gauging stations (wells) during 1992 and 2005.
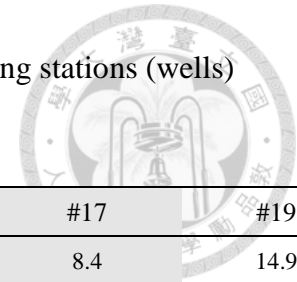
Item	unit	#3		#6		#7		#12		#17		#19	
Well depth	m	22.8		17.0		19.0		19.6		8.4		14.9	
		Mean	SD[1]	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
As	ug/L	75.9±67.6		177.0±109.5		450.4±314.3		43.7±30.7		39.5±47.6		38.1±30.7	
temp	℃	25.8±1.0		25.7±0.9		25.8±1.0		25.9±1.3		26.1±1.4		26.1±1.3	
pH		7.7±0.4		7.9±0.4		7.9±0.2		7.7±0.5		7.6±0.4		7.6±0.3	
EC	μmho/cm 25℃	23383±18221		16509±8317		2209±912.2		21795±13886		1408±970.8		17295±7677	
DO	uS/cm	1040±7347		419.0±2923		51.0±349.1		1.3±1.0		1.3±1.0		1.3±1.0	
Alk	ug/L	356.2±137.3		560.2±138.5		504.4±84.8		384.1±123.9		315.0±55.7		504.2±143.0	
TDS	ug/L	15822±10285		10963±4149		1432±626.2		13994±7302		885.9±637.7		11790±5426	
Cl⁻	ug/L	6851±4652		4479±1819		391.9±215.6		5945±3419		233.6v236.3		4937±2155	
SO_4^{2-}	ug/L	690.9±784.1		512.0±270.8		102.9±82.8		960.0±588.1		64.5±57.4		515.6±467.0	
Na^+	ug/L	3772±2604		2708±998.5		293.5±90.8		3297±1806		179.2v121.7		2756±1133	
K^+	ug/L	201.1±105.6		145.2±46.1		38.7±15.5		133.4±57.6		17.0v11.4		142.4±91.9	
Mg^{2+}	ug/L	598.8±954.6		254.7±205.5		73.2±30.3		427.5±296.4		32.2v22.6		323.8±156.3	
Ca^{2+}	ug/L	216.0±133.1		74.9±50.1		59.0±19.7		281.8±174.3		88.0±46.7		150.4±75.8	

[1] standard deviation

A total of 270 (=45x6 wells) data sets of twelve water quality parameters 【power of hydrogen (pH), alkalinity (Alk), cadmium ion (Ca^{2+}), chlorine ion (Cl^-), total dissolved solid (TDS), electrical conductivity (EC), sodium ion (Na^+), sulfate ion (SO_4^{2-}), potassium ion (K^+), dissolved oxygen (DO), magnesium ion (Mg^{2+}) and temperature (temp)】 were collected at six gauging stations (wells) between 1992 and 2005, which are used for model construction in this study. Table 3.9 shows the well

Table 3.10 Correlation matrix of As concentration and water quality parameters collected at six gauging stations (wells) during 1992 and 2005.

	As	pH	Alk	Ca^{2+}	Cl^-	TDS	EC	Na^+	SO_4^{2-}	K^+	DO	Mg^{2+}	temp
As	1.00	0.32	0.13	-0.34	-0.32	-0.31	-0.30	-0.30	-0.27	-0.26	0.19	-0.18	-0.07
pH		1.00	0.46	-0.56	-0.48	-0.43	-0.51	-0.44	-0.47	-0.37	0.26	-0.33	-0.03
Alk			1.00	-0.41	-0.27	-0.20	-0.28	-0.22	-0.33	-0.16	0.14	-0.17	0.04
Ca^{2+}				1.00	0.77	0.71	0.76	0.73	0.74	0.62	-0.22	0.52	0.02
Cl^-					1.00	0.93	0.97	0.97	0.88	0.87	-0.16	0.55	-0.05
TDS						1.00	0.93	0.93	0.81	0.84	-0.18	0.51	-0.07
EC							1.00	0.96	0.88	0.86	-0.19	0.53	-0.06
Na^+								1.00	0.84	0.88	-0.19	0.52	-0.09
SO_4^{2-}									1.00	0.72	-0.08	0.42	-0.06
K^+										1.00	-0.11	0.47	-0.07
DO											1.00	-0.12	-0.02
Mg^{2+}												1.00	0.02
Temp													1.00

depth, mean and standard deviation (SD) of groundwater quality parameters at these 6 gauging stations, in which high mean and variation of As concentration occur, especially at wells #6 and #7. The depth intervals of the three aquifers were <60m, 120-200m, and 280-350 m, respectively (Agricultural Engineering Research Center, 2008). This indicates that the 6 monitored wells (well depth: 8.4-22.8m) are in the same confined aquifer. Table 3.10 shows that all the correlation coefficients between As and twelve water quality parameters are smaller than 0.34 (in an absolute sense), which implies the difficulty in determining non-trivial factors that affect As concentration based solely on such traditional correlation analysis. Therefore, we adopt a more

sophisticated method to effectively extract non-trivial factors from water quality

parameters for building an As concentration estimation model.
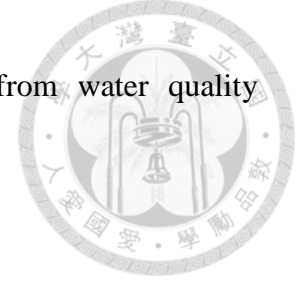
C. Results and discussion

1) Extracting effective water quality factors

The six wells (#3, #6, #7, #12, #17 and #19) that have sufficient water quality data

are assumed as gauging stations for As concentration and are utilized by the GT. Data

sets of twelve water quality factors are first scaled to [-1,1], and a total of 4095 ($=2^{12}$-1)

Γ values corresponding to all possible input combinations are derived through the GT.

The derived Γ values are next sorted in an ascending order, in which Γ values

smaller than the 10th percentile ($\Gamma_{10} = 0.0089$) are classified as the best group ($F_{\Gamma \leq \Gamma_{10}}$)

whereas Γ values bigger than the 90th percentile ($\Gamma_{90} = 0.136$) are classified as the

worst group ($F_{\Gamma \geq \Gamma_{90}}$). Figure 3.17 shows the results of the GT, where blue bars represent

the occurrence frequency of variables in the best group ($F_{\Gamma \leq \Gamma_{10}}$) and red bars represent

the occurrence frequency of variables in the worst group can then be identified as the

variables associated with higher blue bars and lower red bars simultaneously, and such

ratios are shown by the dotted line in Fig. 3.17. And therefore we can extract a subset of

input variables that ranks top three in the ratio of $F_{\Gamma \leq \Gamma_{10}}$ to $F_{\Gamma \geq \Gamma_{90}}$. The GT results

indicate that Alk, Ca^{2+} and pH value are the non-trivial factors for use in the estimation

models (the NARX network and the BPNN).
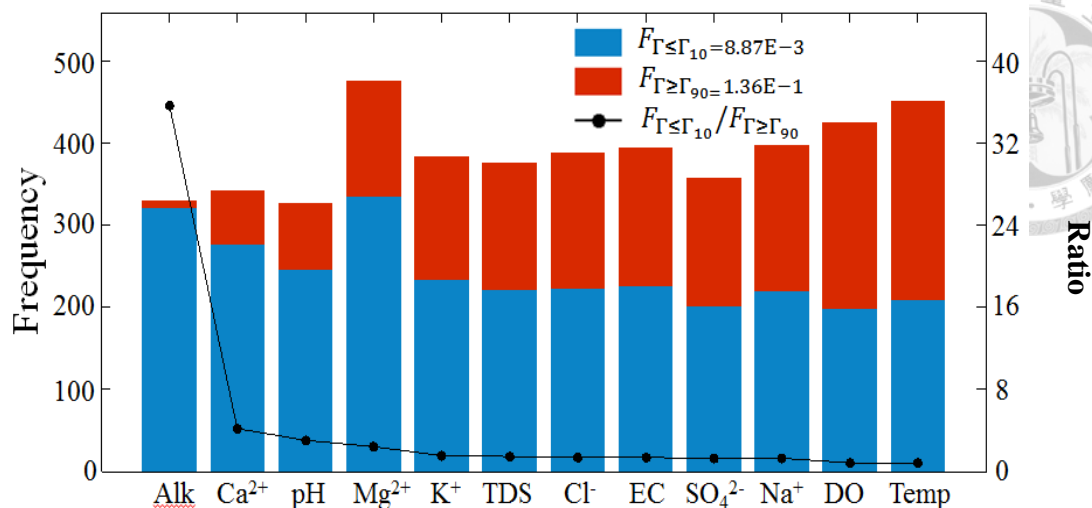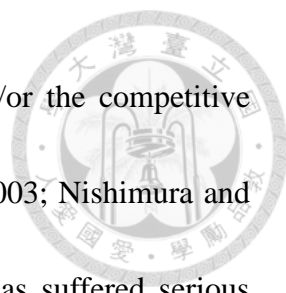
Fig. 3.17 Determination of non-trivial factors by the GT results.

These results are consistent with several studies, which indicated the increase in As leaching efficiency depends on high pH values and Alk concentration (Anawar et al., 2004; Kim et al., 2000; Kuo and Chang, 2010; Liu et al., 2003; Park et al., 2006; Pierce and Moore, 1982). The major C-containing species in the reducing condition in groundwater are HCO_3^- and H_2CO_3 , which cause high pH values and Alk concentration (Wang et al., 2007). In addition, salinization and As enrichment are two main hydro-geochemical characteristics in the Yun-Lin coastal area, and they were estimated by the factor analysis (Wang et al., 2007). Respectable cation, such as calcium ions, and anion contents carried by seawater intrusion initially increased the ion strength in groundwater and induced As desorption (Appelo et al., 2002; Keon et al., 2001). On the one hand, As anions could sorb or bind using carbonates in natural systems (Bauer et al., 2008; Rothwell et al., 2009). Therefore the relationship between As and calcium

ions might be caused by the dissolution of calcium arsenates and/or the competitive

desorption of calcium (Bothe and Brown, 1999; Mihaljevicet al., 2003; Nishimura and

Robins, 1998). In the Yun-Lin coastal area, the shallow aquifer has suffered serious

salinization that affects the concentration of the calcium ion due to the over-pumping of

groundwater. This exercise gives evidence that the GT can effectively identify

non-trivial and meaningful factors that affect the fluctuations of As concentration,

compared with the identification difficulty raised by the traditional correlation matrix

shown in Table 3.10.

2) Estimating As concentration at ungauged sites by the NARX network

In this case study, the NARX network is proposed to estimate the regional As

concentration in Yun-Lin County. Variables Alk, Ca^{2+} and pH determined by the GT are

used as exogenous inputs to the NARX network. The data sets collected at six gauging

stations between 1992 and 2005 are used for model calibration. Therefore, the NARX

network in the SP mode trained by the Bayesian regularization method is calibrated by a

30-fold cross validation. The log-sigmoid function and the linear function are the

transfer functions used in the hidden and output layers of the NARX network,

respectively. The most appropriate NARX network comprises two output-memory

orders and 20 neurons in the hidden layer by trial and error method, and the effective

number of network parameters (γ_p) is 23.74.

To demonstrate the effectiveness and usefulness of the NARX network established, the BPNN is implemented for comparison purpose. The constructed BPNN consists of the same input variables as those of the NARX network and six neurons in the hidden layer. The hyperbolic tangent sigmoid function and the linear function are the transfer functions used in the hidden and output layers of the BPNN, respectively. The BPNN trained with the Levenberg-Marquardt optimization algorithm is also calibrated by a 30-fold cross validation. The results show that the average RMSE of the NARX network in the training and validation phases are 95.11 and 106.13 ugl^{-1}, respectively, whereas the average RMSE of the BPNN in the training and validation phases considerably increases to 121.54 and 143.37 ugl^{-1}, respectively. The results demonstrate that the NARX network has much better performance than the BPNN. It is noted that both models produce large errors (average RMSE), this is mainly due to the high uncertainty attached to the sampled values, where the mean (138.26 ugl^{-1}) and standard deviation (205.25 ugl^{-1}) of As concentration contributes to the poor model accuracy.

It is worth noting that the effective number of network parameters (γ_p) for the NARX network has been optimized from 141 to 23.74 after the re-calibration of the network by using the Bayesian regularization method. This demonstrates that the Bayesian regularization method can significantly reduce the effective number of network parameters and thus avoid the over-fitting problem caused in a rather complex

network structure. As a result, the NARX network produces suitable results and has

similar performance in the training and validation phases (average RMSE: 95.11 ugl^{-1}

and 106.13 ugl^{-1} accordingly). In contrast, the BPNN requires fewer neurons in the

hidden layer to prevent the over-fitting problem but still performs poor in the validation

phase (average RMSE: 143.37 ugl^{-1}).

After model configuration, a total of 100 (=20x5 months) As concentration data

collected at twenty assumed ungauged sites in five monitoring months (January 1995,

October 1996, October 1997, September 1998 and January 1999) are utilized to test the

two constructed models. In addition to RMSE, the normalized mean squared error

(NMSE), R-square value (R^2) and F test are also used as performance criteria in the

testing phase. The NMSE is defined as:

$$NMSE = \frac{\sum_{i=1}^{n}(O_i - \hat{Z}_i)^2}{\sum_{i=1}^{n}(O_i - \overline{O})^2} \tag{52}$$

where O_i and $\hat{Z}_i$ are the observed and estimated As concentration from the ith assumed

ungauged sites in the same year, respectively, $\overline{O}$ represents the average of observed As

concentrations in a certain year, and n is the length of data.

The results of model comparison in the testing phase are summarized in Table 3.11,

which indicates the NARX network has much smaller RMSE as well as NMSE values

and higher R^2 values than the BPNN. Besides, when assessing the results of the

Table 3.11 Estimation performance of the NARX network and the BPNN for As concentration at 20 ungauged sites between 1995 and 1999 in the testing phase.

	Estimation time	RMSE (ugl[-1])	NMSE	R^2	F test p-value[2]	Data Mean (ugl[-1])	Data SD[1]. (ugl[-1])
NARX network	1995 Jan.	57.31	0.19	0.89	0.105	85.63	134.80
	1996 Oct.	91.53	0.29	0.89	<u>0.010</u>	90.71	173.61
	1997 Oct.	40.24	0.11	0.96	0.291	64.51	125.99
	1998 Sept.	48.34	0.26	0.91	0.731	47.83	97.02
	1999 Jan.	41.35	0.07	0.98	0.816	75.57	155.91
BPNN	1995 Jan.	109.02	0.69	0.41	<u>0.025</u>	85.63	134.80
	1996 Oct.	158.70	0.88	0.17	<u>0.002</u>	90.71	173.61
	1997 Oct.	142.42	1.35	0.18	0.070	64.51	125.99
	1998 Sept.	114.63	1.47	0.29	0.302	47.83	97.02
	1999 Jan.	140.21	0.85	0.25	<u>0.004</u>	75.57	155.91

[1]Standard deviation

[2]The null hypothesis is rejected at the 5% level (p-value < 0.05)

F test, the null hypothesis is rejected only at the 5% level of the estimation values in October 1996 for the NARX network, whereas the null hypothesis is rejected in January 1995, October 1996 and January 1999 for the BPNN. Figure 3.18 shows the scatter plots of observed and estimated As concentrations in five different months during 1995 and 1999 derived from the NARX network and the BPNN. The estimation values obtained from the NARX network are close to the ideal line and only have few underestimations at extremely high As concentrations, whereas the BPNN overestimates As concentrations at values lower than 200 ugl[-1] and seriously underestimates As
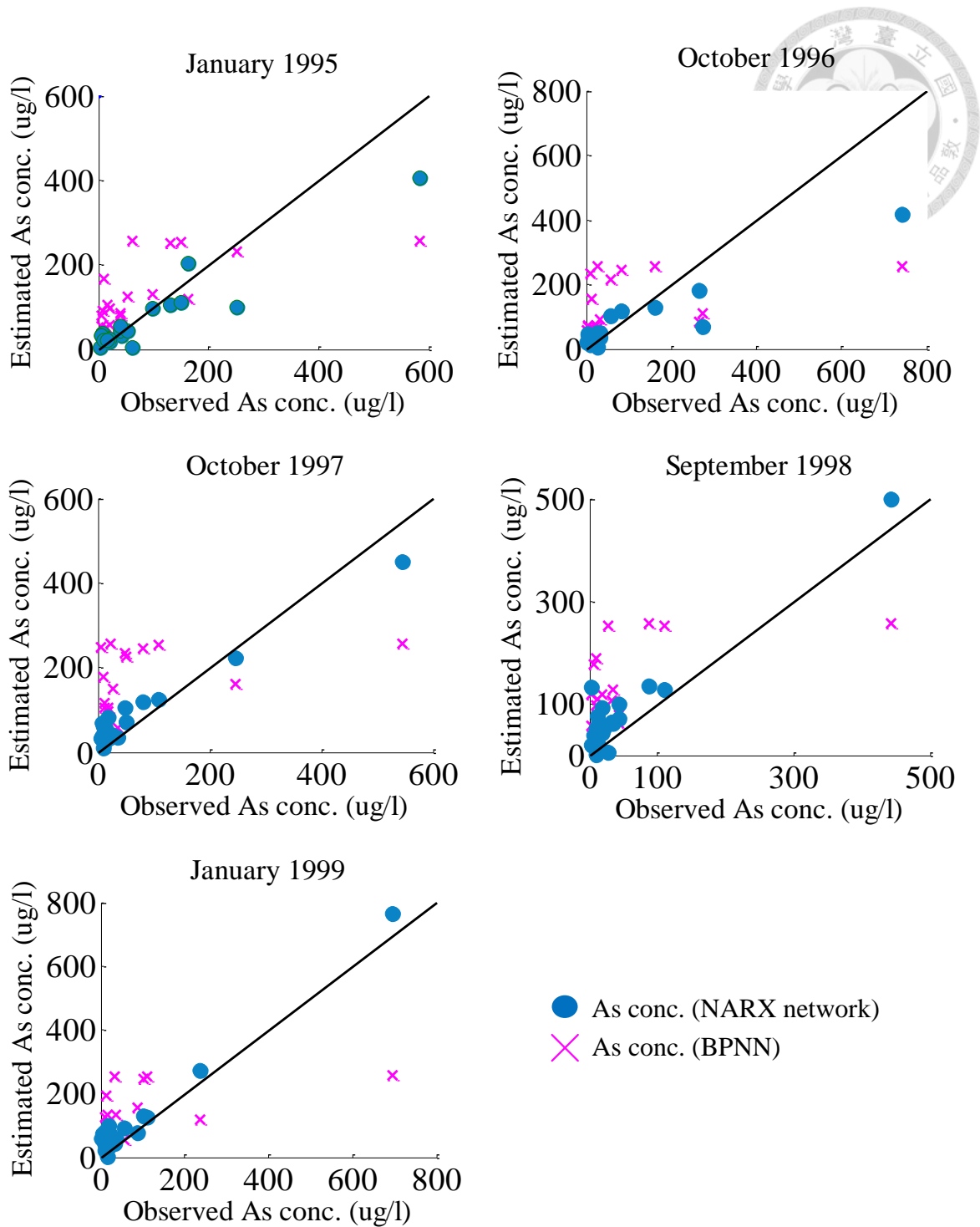
Fig. 3.18 Scatter plots of observed and estimated As concentration (conc.) derived from the NARX network and the BPNN at twenty ungauged sites (1995-1999).
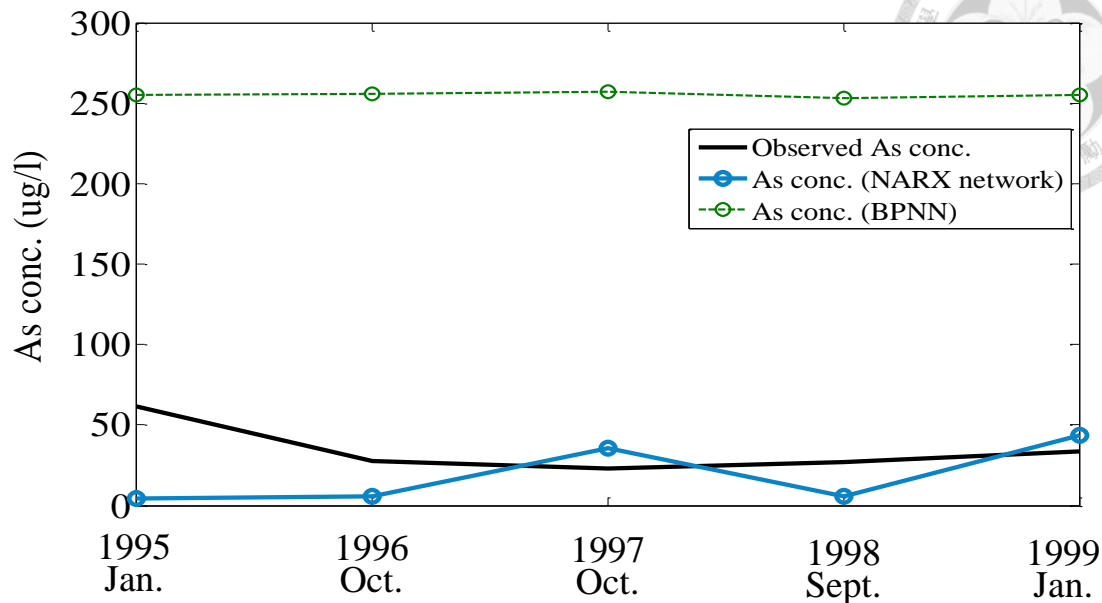
Fig. 3.19 Estimation results of As concentrations at ungauged well #14 during 1995 and 1999 in the testing phases of the NARX network and the BPNN.

concentrations at values higher than 200 ugl^{-1}. Furthermore, Fig. 3.19 shows the estimation results of As concentrations at the ungauged well #14, which is located at the central Yun-Lin coastal area and is far from other gauging and ungauged wells, during 1995 and 1999 in the testing phases of the NARX network and the BPNN. Results indicate that the NARX network only underestimates As concentration in 1995, whereas the BPNN highly overestimates As concentration in all five testing years.

In sum, the NARX network adequately utilizes the information of model outputs through recurrent connections to the network itself for producing reliable estimations of As concentrations at twenty ungauged sites. Owing to the implementation of the Bayesian regularization method into the NARX network, the constructed network

shows impressive generalizability and performs well in the testing phase, which can be

proved through the similar NMSE values in five testing years (Table 3.11).
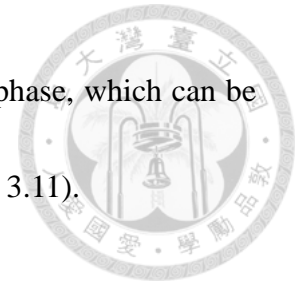
3) *Deriving the risk map of As concentration through the IK*

From the previous section, the NARX network can provide reliable point

estimation of As concentration at twenty ungauged sites. The IK is next employed to

estimate the regional spatial distribution and to compute the probability of the exposure

to high As pollution in unsampled areas. Because the WHO drinking water standard for

As concentration is $10ugl^{-1}$, the investigation of this study mainly focuses on the

threshold of $10ugl^{-1}$. Therefore, the estimated As concentration obtained from the

NARX network at twenty ungauged sites and the observed As concentration at six

gauging stations are utilized to construct the semivariogram models for the IK.

The NARX network coupled with the IK can illustrate the unknown probability of

the exposure to high As concentrations at neighboring areas of all 26 wells. If the

observed and estimated As concentrations exceed the threshold set in the adjacent

region, the IK will assign a high probability of concentration in the region of interest.

The probability maps of As concentration under the threshold of WHO drinking water

standard ($10ugl^{-1}$) in different time spans are shown in Fig. 3.20. In January 1995, high

exceeding probabilities (>10 ugl^{-1}) of As concentration occurred in northern and
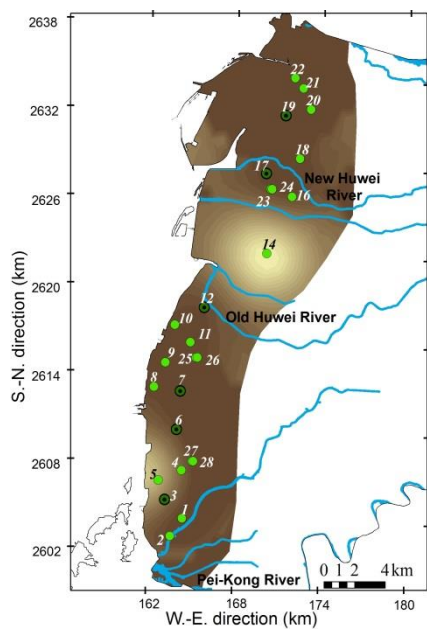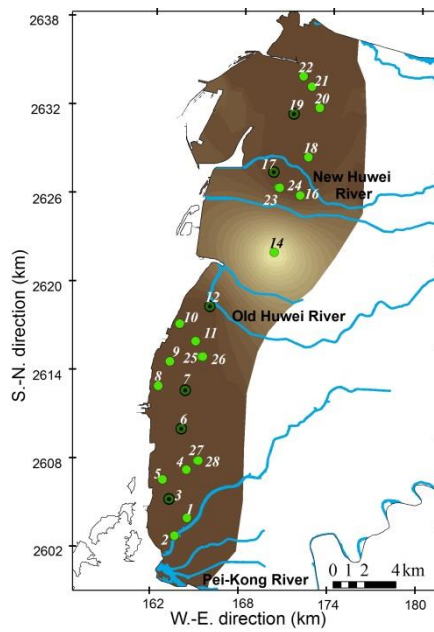
(a) January 1995 (b) October 1996 (c) October 1997
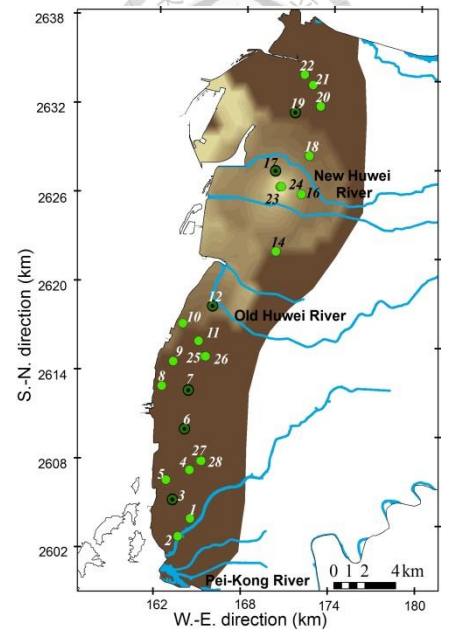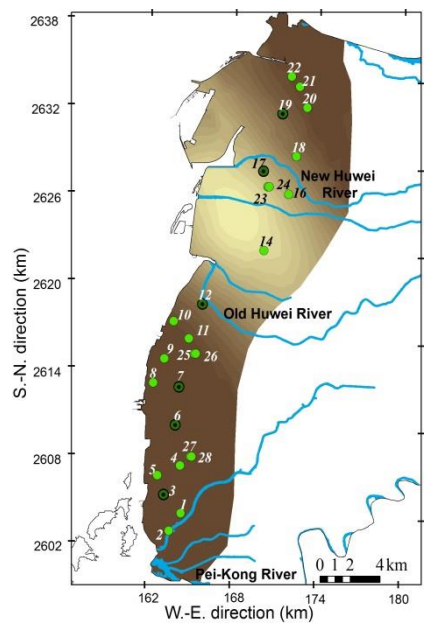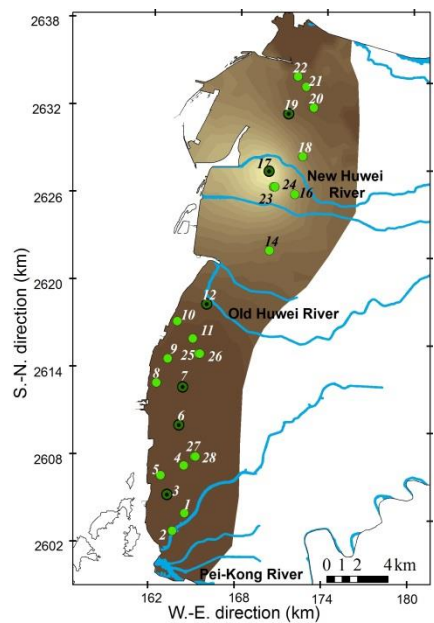
(d) September 1998 (e) January 1999
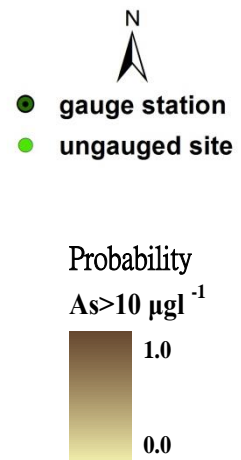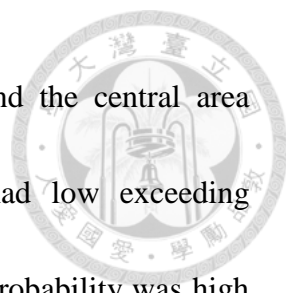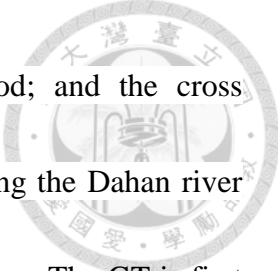
Fig. 3.20 Exceeding probability maps of As concentration under the threshold of WHO drinking water standard (10ugl^{-1}) between 1995 and 1999.

southern areas, whereas both the surrounding area of well #5 and the central area (located between Old Huwei River and New Huwei River) had low exceeding probabilities of As concentration. In October 1996, the exceeding probability was high in the southern area of the Old Huwei River. In contrast, the exceeding probability of As concentration was gradually and significantly mitigated in the central and northern areas from October 1997 to January 1999, and the Old Huwei River could be deemed as a clear boundary between high and low As concentrations in an exceeding probability sense. These risk maps reveal the high As-prone areas. As a result, the information of the risk maps derived from the IK of the proposed SDM can consequently help decision makers manage groundwater quality and thus prevent residents from drinking or using toxic groundwater.

3.4 Modeling spatio-temporal total phosphate (TP) concentration through Systematical Dynamic-neural Modeling

It is a great challenge to model the spatio-temporal variability of water quality along a highly dynamic river channel. Various difficulties may be commonly encountered in modeling, such as the scarcity of field data, the identification of key factors affecting the output, instability and unreliability. To tackle the possible difficulties, the proposed SDM that comprises a dynamic neural network with three

statistical techniques: the GT; the Bayesian regularization method; and the cross validation technique, is adopted to model the TP concentrations along the Dahan river channel based on the limited data sets of nine water quality parameters. The GT is first used to effectively extract key factors that significantly affect the fluctuations of TP concentrations. Then the NARX network is constructed to simultaneously estimate quarterly TP concentrations at seven monitoring stations based on the selected key factors, in which the network complexity is controlled automatically by the Bayesian regularization method to avoid over-fitting; and the cross validation technique is adopted as a sensible criterion for model selection in the calibration stage to overcome data scarcity. Through the constructed NARX network, TP concentrations can be reconstructed in a monthly scale for producing the monthly WQI (including TP). For the purpose of comparison, the BPNN is also implemented for comparison purpose. The implementation procedure is illustrated in Fig. 3.21.

A. Study area

The Dahan River is one of the most polluted rivers in northern Taiwan, and its water quality has decreased for the past decades. Figure 3.22 shows the location of the Dahan River basin, which is divided into two zones based on land-use morphology: the upstream zone (from Shihmen Dam to Yuanshan Weir: water quality monitoring stations S1-S3); and the downstream zone (from Yuanshan Weir to the confluence point of the
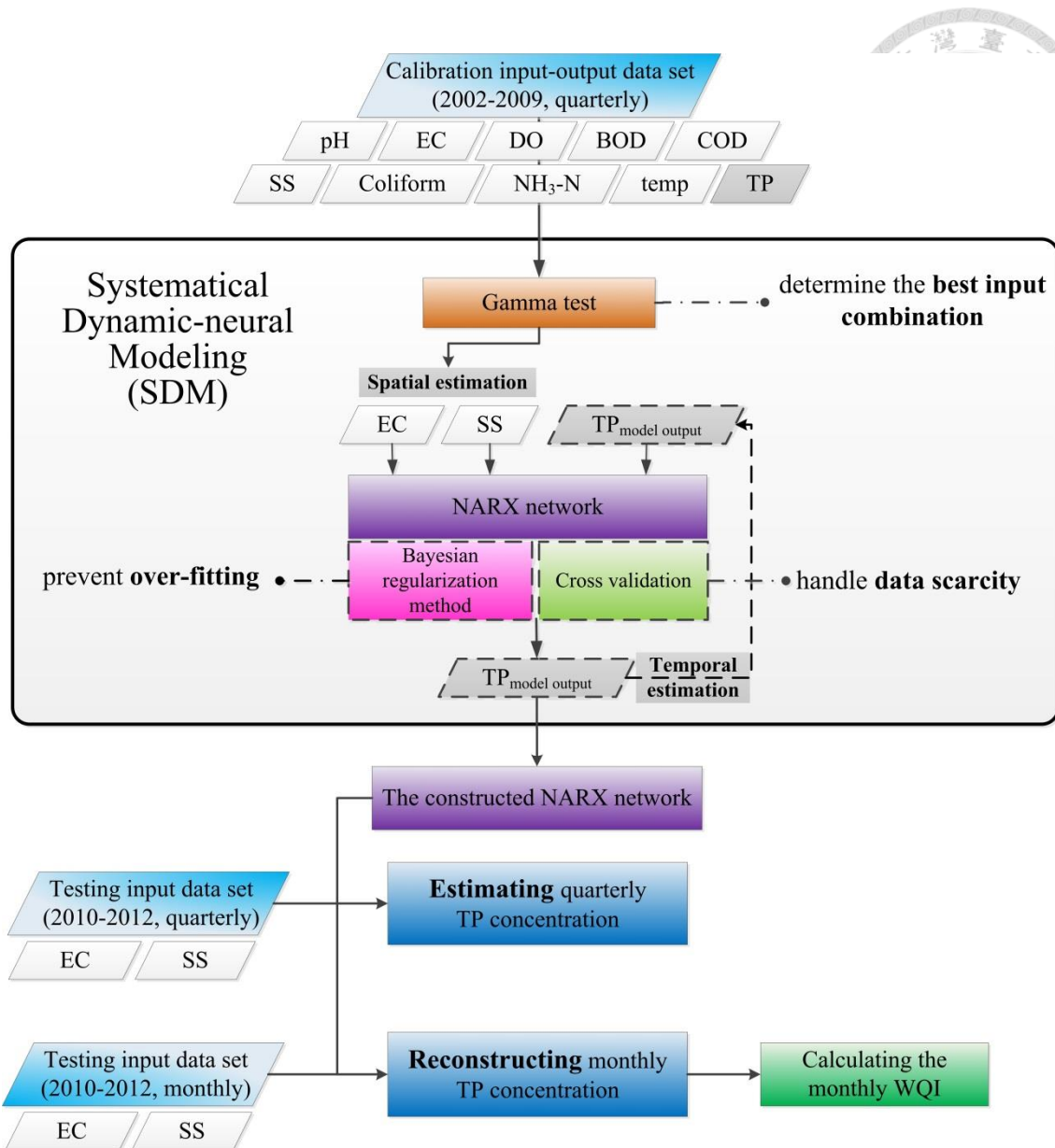
Fig. 3.21 Research flowchart of the proposed SDM for the estimation of TP concentration.

Dahan River and the Xintian River: S4-S7). According to the disclosure of the WQI, the

water quality of the upstream zone is better than that of the downstream one, which can

be attributed to the high degrees of urbanization and industrialization in the downstream

zone.
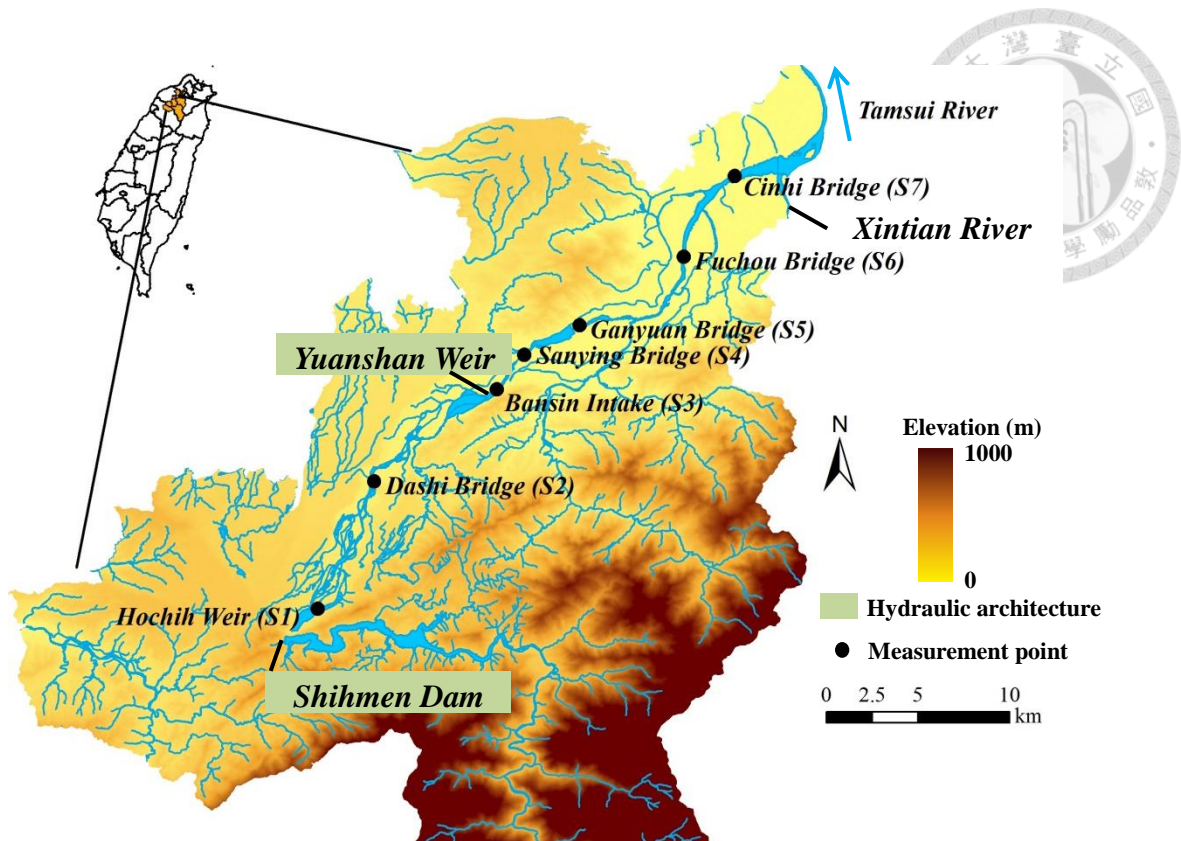
Fig. 3.22 Locations of the Dahan River basin and water quality monitoring stations S1-S7.

B. Data collection

A water quality survey was carried out at seven water quality monitoring stations (S1-S7) along the Dahan River channel from July 2002 to June 2012, which consisted of the quarterly-measured TP and nine monthly-measured water quality parameters: power of hydrogen (pH), electro conductivity (EC), dissolved oxygen (DO), biochemical oxygen demand (BOD), chemical oxygen demand (COD), suspended solid (SS), coliform group (Coliform), ammonia nitrogen (NH$_3$-N) and water temperature (temp). A total of 280 quarterly TP data sets collected at S1-S7 are separated into two

Table 3.12 Test methods and preliminary statistics of water quality parameters collected at water quality monitoring stations S1-S7 in the Dahan River basin during the model calibration period (2002-2009).
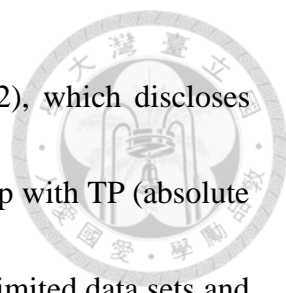
Variable	pH	EC µmho/cm 25℃	DO mgl^{-1}	BOD mgl^{-1}	COD mgl^{-1}	SS mgl^{-1}	Coliform CFU/100 mL	NH$_3$-N mgl^{-1}	temp ℃	TP mgl^{-1}
Test method	NIEA W424.52A	NIEA W203.51B	NIEA W455.52C	NIEA W510.55B	NIEA W515.54A	NIEA W210.58A	NIEA E202.55B	NIEA W448.51B	NIEA W217.51A	NIEA W427.53B
min	6.6	182	0.00	1.00	4.00	3.6	10	0.02	13.0	0.01
max	10.4	1200	14.10	22.80	535.00	13600	6.8E07	14.40	33.6	6.88
median	8.0	311	7.70	2.60	12.35	38.4	4750	0.36	23.5	0.13
mean	7.9	413	6.70	5.04	23.64	560	858244	1.94	22.9	0.40
SD[1].	0.6	218	3.18	5.20	44.28	1832	5499257	3.12	5.4	0.79
CC[2] with TP	0.19	0.56	-0.54	0.65	0.66	0.70	0.25	0.49	-0.02	-
N[3]	210	210	210	210	210	210	210	210	210	210

[1] Standard deviation
[2] Correlation coefficient
[3] Number of samples

stages for modeling: 210 data sets (30 x 7 stations) collected during 2002/07 and 2009/12 are for calibration purpose; and 70 data sets (10 x 7 stations) collected during 2010/03 and 2012/06 are for testing purpose. Table 3.12 shows the test methods and the statistical analysis of water quality parameters at the study area during the period of the model calibration (2002-2009), in which the maximum TP concentration exceeds 6 mgl^{-1}. The standard deviations of COD, SS, Coliform, NH$_3$-N and TP are high. Through checking the original data, it makes significant differences in the concentrations of these parameters between the upstream zone and the downstream zone, and such phenomena could be caused mainly by the human activities at the downstream zone. The relationship between TP and nine water quality parameters in the study area is
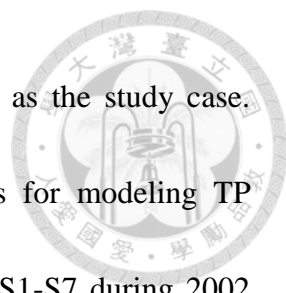
discovered by the correlation coefficient (CC) analysis (Table 3.12), which discloses that EC, BOD, COD, SS and DO have relatively stronger relationship with TP (absolute CC values > 0.5) than the remaining parameters. On account of the limited data sets and the weak linear relationships between the inputs (nine parameters) and the output (TP), a more sophisticated nonlinear analysis with parsimounous parameters for model input selection is expected.

C. Results and discussion

In this case study, TP concentration was measured in a quarterly scale, and the SDM is proposed to modeling spatio-temporal TP concentration. The NARX network is constructed to simultaneously produce quarterly TP estimates at individual water quality monitoring stations (S1-S7) based on water quality data measured at each station in the same months as TP was sampled during 2002 and 2012. As the estimation model is constructed, it can be utilized to reconstruct monthly TP concentrations by being fed with the monthly data of the selected water quality parameters at each monitoring stations (S1-S7). That is to say, the monthly reconstructions of seasonal TP data can be carried out, and thus the monthly WQI (including TP) can be obtained.

1) Determination of TP-affected water quality factors by the GT

For model input selection, it is difficult to identify key factors affecting TP concentrations by the traditional correlation matrix showyn in Table 1 because linear

correlations may not fully explain highly non-linear systems such as the study case. Alternatively, this study adopts the GT to detect the key factors for modeling TP concentration. Data collected at water quality monitoring stations S1-S7 during 2002 and 2009 are first utilized by the GT and then are used for the calibration of the spatio-temporal estimation model. The data sets of nine water quality parameters are scaled to [-1,1], and a total of 511 ($=2^9$-1) Γ values corresponding to all possible input combinations are calculated through the GT. A small subset of inputs significantly relevant to the output (the target variable-TP) can be extracted by examining the occurrence frequency of each input variable in the best results of the GT. Γ values smaller than the 10th percentile ($\Gamma_{10} = 0.028$) of all Γ values are defined as the best results of the GT, whereas Γ values bigger than the 90th percentile ($\Gamma_{90} = 0.217$) of all Γ values are defined as the worst results of the GT. Fig. 3.23 shows the GT results, where the blue bars represent the occurrence frequency of each variable in the best results ($F_{\Gamma \leq \Gamma_{10}}$) and the red bars represent the occurrence frequency of each variable in the worst results ($F_{\Gamma \geq \Gamma_{90}}$). As a consequence, key factors are those producing higher blue bars and lower red bars simultaneously, which can be easily identified by the ratios of $F_{\Gamma \leq \Gamma_{10}}$ (best) to $F_{\Gamma \geq \Gamma_{90}}$ (worst), as the dotted line shown in Fig. 3.23.

The results of the GT indicate that EC and SS are the desired key factors, while the rest of the parameters are not considered useful and/or important for modeling TP
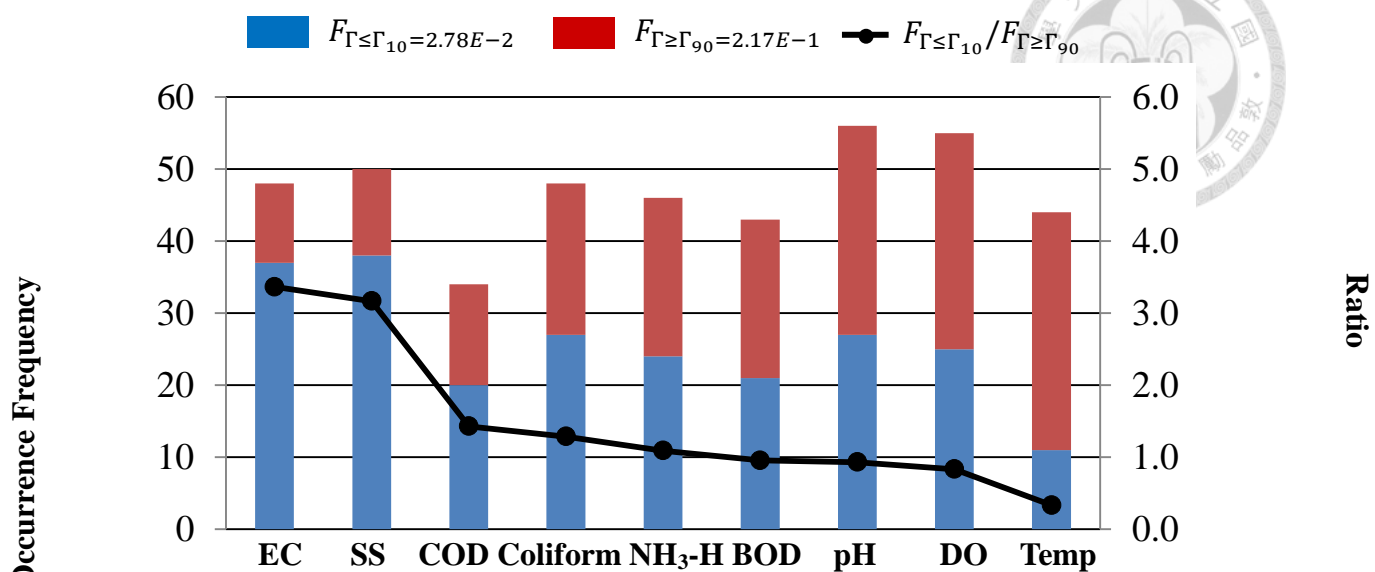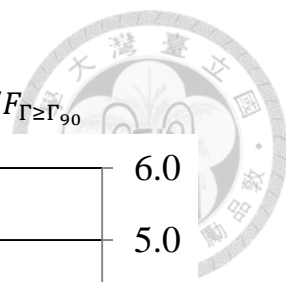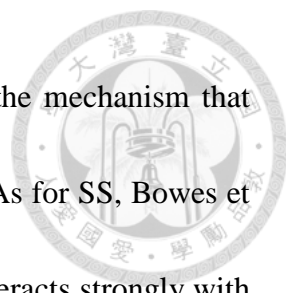
Fig. 3.23 Determination of key factors by the GT results.

concentration. We notice that Temp has the lowest absolute CC value and the lowest ratio of $F_{\Gamma \le \Gamma_{10}}$ to $F_{\Gamma \ge \Gamma_{90}}$ (as shown in Table 3.12), nevertheless DO and BOD have relatively high absolute CC values but lower ratios of $F_{\Gamma \le \Gamma_{10}}$ to $F_{\Gamma \ge \Gamma_{90}}$. The results suggest that if DO and BOD are used as input variables, the constructed model might not be as reliable (stable) as expected.

TP-related geochemical processes also support the selection of EC and SS as key factors. EC is a measure of the ability of water to pass an electrical current. The United States Environmental Protection Agency (USEPA) indicated that EC has certain relationship with TP concentrations because conductivity in water is affected by the presence of inorganic dissolved solids such as chloride, nitrate, sulfate, phosphate anions or sodium, magnesium, calcium, iron and aluminum cations (USEPA, 1987).
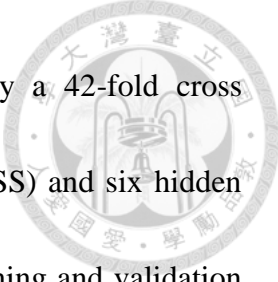
Therefore, EC and TP are closely related to each other owing to the mechanism that hydrolyzes phosphorus in water into the form of phosphate anions. As for SS, Bowes et al. (2003) indicated that dissolved inorganic phosphorus not only interacts strongly with sediments but involves various precipitation dissolution reactions, such as calcium carbonate phosphate and iron/aluminium oxide minerals, which reveals the reason why SS is also a key factor. In brief, although the GT cannot detect the specific chemical mechanism of TP, it can still effectively extract non-trivial and meaningful factors that affect the fluctuations of TP concentrations.

2) Spatio-temporal estimation of TP concentrations by the NARX network

The NARX network is employed in this study to estimate the spatio-temporal TP concentration in the Dahan River. EC and SS determined by the GT are used as exogenous inputs to the NARX network, and 210 data sets collected at water quality monitoring stations S1-S7 are used for model calibration. The NARX network in the SP mode trained by the Bayesian regularization method is calibrated by a 42-fold cross validation. The most appropriate NARX network comprises two output-memory orders and 10 hidden layer neurons.

To explore an effective and useful NARX network in this spatio-temporal estimation task, the BPNN is also implemented with its network structure (without recurrent connections) similar to that of the NARX network. The BPNN trained with the

Levenberg-Marquardt optimization algorithm is also calibrated by a 42-fold cross validation. The constructed BPNN consists of two inputs (EC and SS) and six hidden layer neurons. The results show that the average RMSEs in the training and validation phases of the NARX network are 0.188 and 0.375 mgl^{-1}, respectively, whereas those of the BPNN are significantly increased to 0.262 and 0.665 mgl^{-1}, respectively. It is worth noting that the effective number of network parameters has been optimized at 37.74 (notably dropped from 61 = ((2+2) x 10 + 10) parameters + (10+1) biases) as the Bayesian regulation method is employed. This demonstrates that the Bayesian regularization method can effectively restrain the network weights and equivalently reduce the number of network connections to avoid the over-fitting problem caused in a rather complex network structure. In brief, the NARX network produces acceptable results, compared with the mean and standard deviation (0.40±0.79mgl^{-1}) of the TP concentrations at seven monitoring stations, and similar performance is achieved in the training and validation phases. Compared with the NARX network, the BPNN requires fewer hidden layer neurons to prevent the over-fitting problem but performs poorly in both training and validation phases. The BPNN results suggest that this static neural network with two key input variables could not well simulate the spatio-temporal TP concentration and thus produce neither reliable nor stable results.
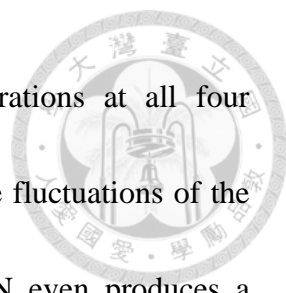
Table 3.13 Performance of the NARX network and the BPNN in the testing phases (2010-2012) for TP concentration estimation at water quality monitoring stations S1-S7.

Year	Model	RMSE (mgl^{-1})	MAE (mgl^{-1})	CE	CC	Data Mean (mgl^{-1})	Data SD[1]. (mgl^{-1})
2010	NARX network	0.075	0.052	0.86	0.95	0.222	0.210
	BPNN	0.230	0.144	-0.30	0.25		
2011	NARX network	0.151	0.102	0.73	0.92	0.237	0.297
	BPNN	0.276	0.181	0.11	0.40		
2012	NARX network	0.057	0.037	0.79	0.90	0.147	0.130
	BPNN	0.136	0.102	-0.18	0.25		

[1]Standard deviation

After finishing model calibration, 70 data sets collected at seven monitoring stations S1-S7 during 2010 and 2012 are utilized to test the two constructed spatio-temporal models, in which TP data collected in March and June of 2010 are used as the initial values for the output regressor of the NARX network. In addition to RMSE, MAE, CE, and CC are also used in the testing stages for comparison purpose.

The results of model comparison in the testing phases are summarized in Table 3.13 The results clearly indicate that the NARX remarkably outperforms the BPNN, in term of much smaller RMSE and MAE values but much higher CE and CC values. The estimations of TP concentrations at monitoring stations S4-S7 of the downstream zone (more polluted) in three testing years are shown in Fig. 3.24. Results indicate that the

NARX network can well capture the trend of the TP concentrations at all four monitoring stations each year, whereas the BPNN cannot reflect the fluctuations of the TP concentrations at most of the monitoring stations. The BPNN even produces a reverse trend as compared with the observed time series, in particular for peak values observed at the downstream stations (S5-S7). Results demonstrate that the NARX network can adequately utilize the information of model outputs through recurrently connecting to the network itself and capture the important dynamic and static features of the time series, which significantly increases estimation reliability. In sum, the NARX network coupled with the Bayesian regularization method shows an impressive generalizability and makes good contribution to model reliability.

3) *Reconstruction of the monthly TP time series and monthly WQI*

Following the previous section, the constructed NARX network can provide reliable point estimation and extract the trend of TP concentrations at seven monitoring stations S1-S7. Therefore, it can be utilized to reconstruct the time series of estimated TP concentrations from quarterly to monthly monitoring scales. Figures 3.25 and 3.26 show the reconstructed monthly time series and the distribution color map of TP concentrations at S1-S7 during 2010 and 2012, respectively. For the prevention of plant nuisances in streams or other flowing waters from being discharged directly to lakes or impoundments, USEPA has suggested 0.1 mgl^{-1} as the desired TP goal (Mackenthun,
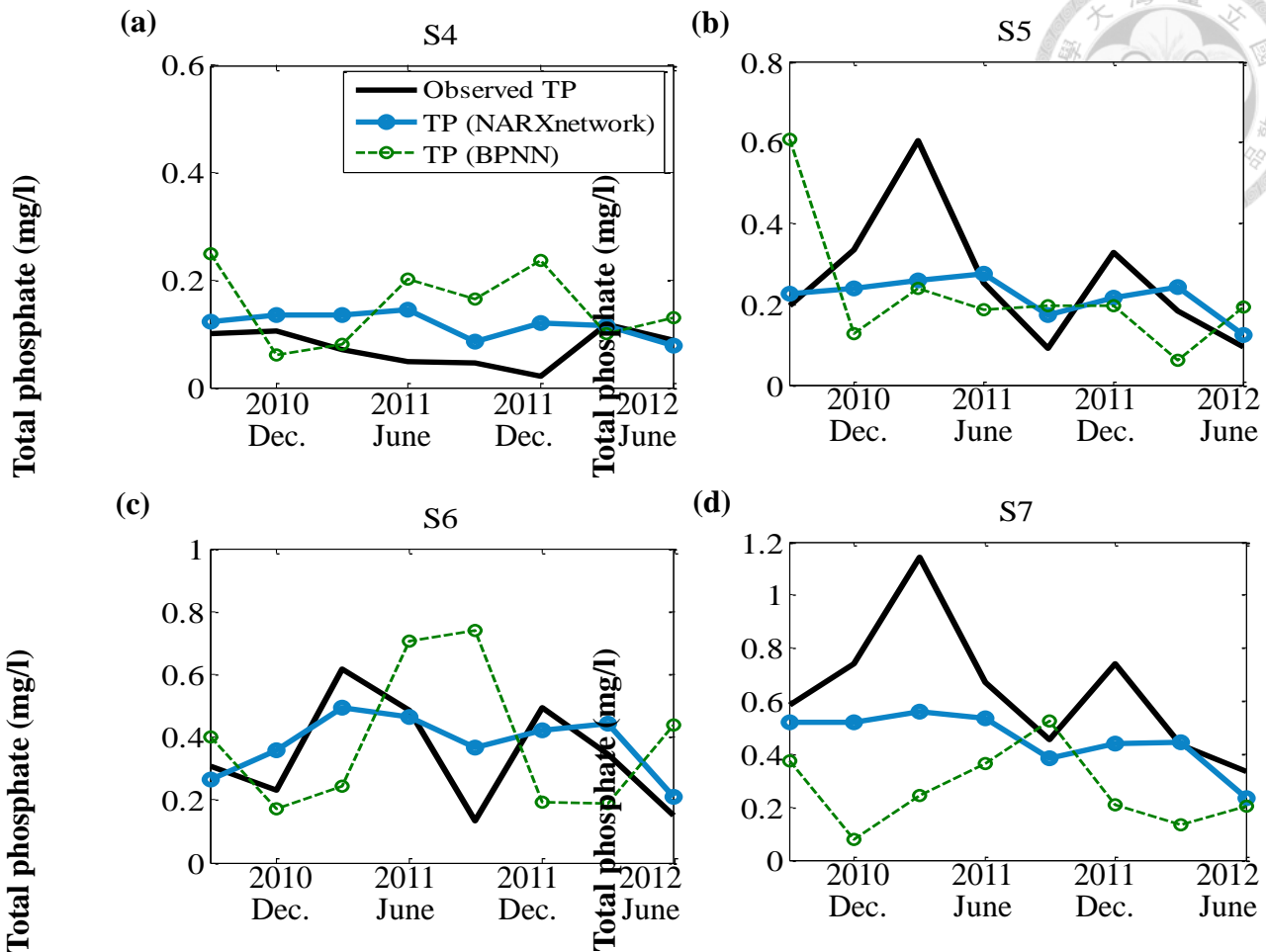
Fig. 3.24 Estimation results of TP concentrations at water quality monitoring stations (a) S4, (b) S5, (c) S6 and (d) S7 during 2010 and 2012 in the testing phases of the NARX network and the BPNN.

1973; USEPA, 1987) (the blue dashed line in Figure 3.25). Results indicate that TP concentration exceeds the quadruple of the desired TP goal at S4-S7 (downstream stations), especially in 2011. The reason for such distribution of TP concentration along the Dahan River might be ascribed to the channel dredging task executed by the New Taipei City Government during 2011, which affects the disturbance of benthal deposits
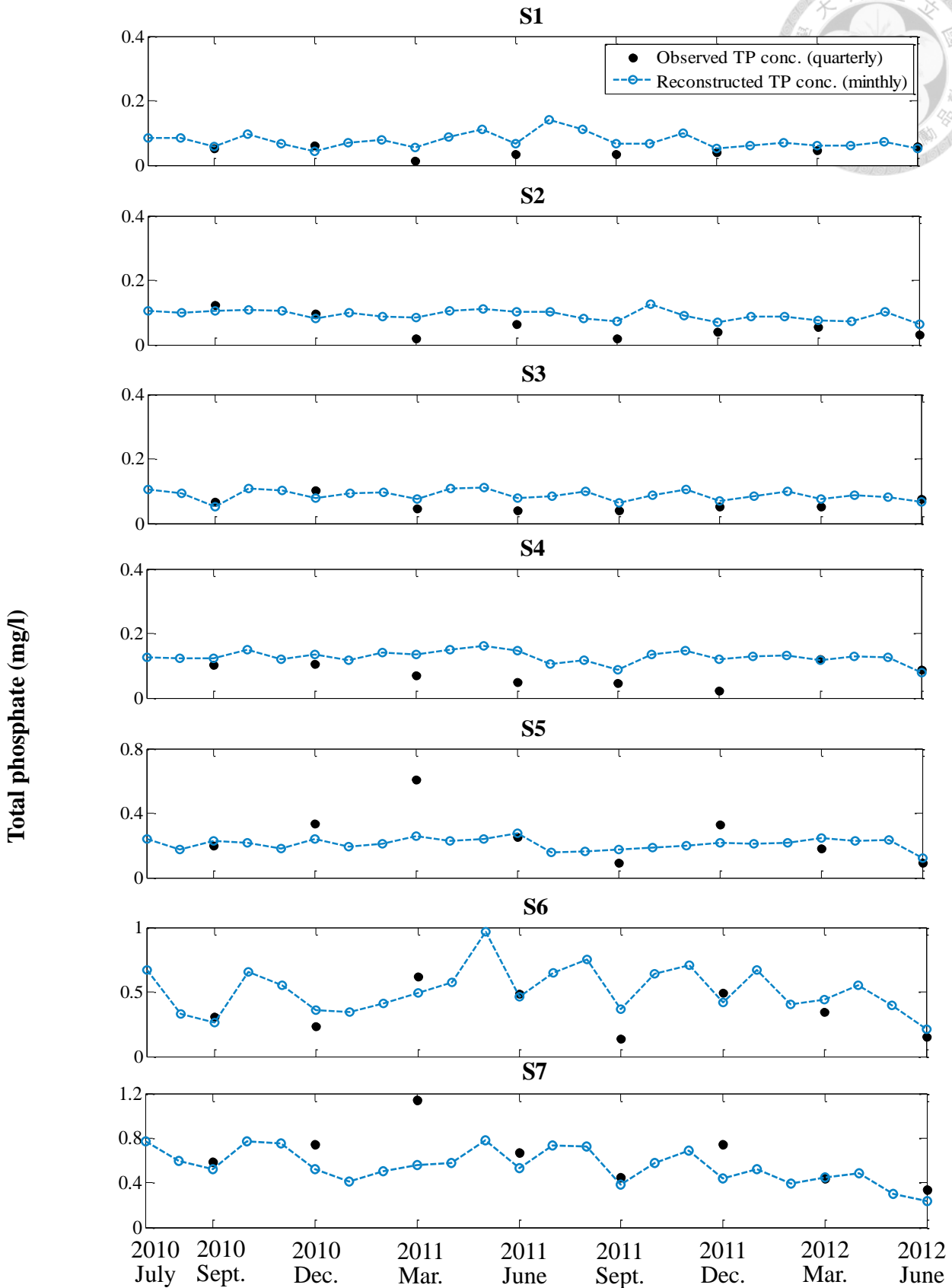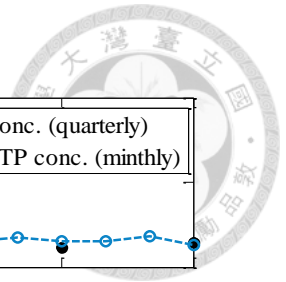
Fig. 3.25 Reconstructed monthly TP time series based on quarterly TP data estimated at water quality monitoring stations S1-S7 (2010-2012).
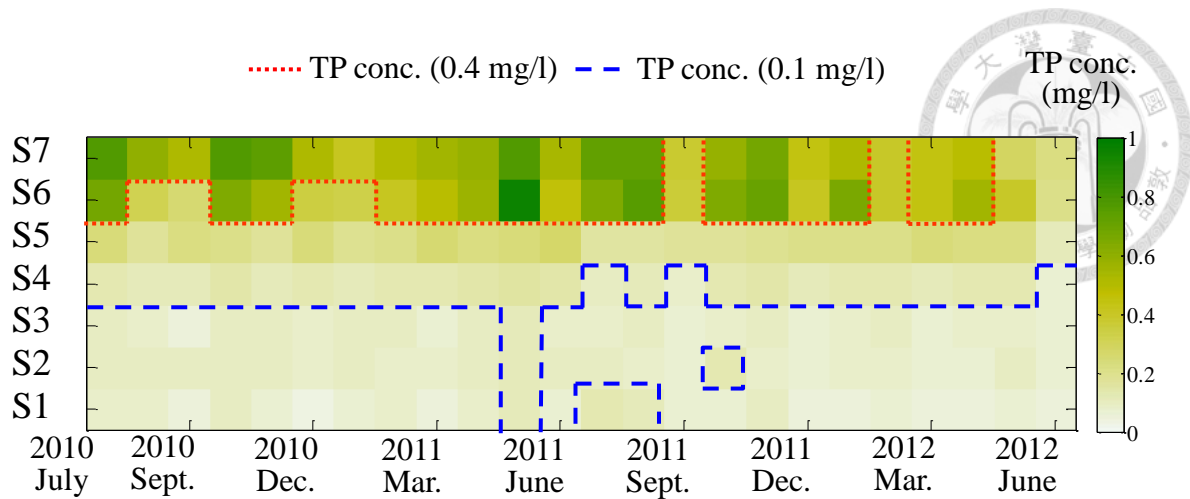
Fig. 3.26 Colormap of the reconstructed monthly TP concentrations at water quality monitoring stations S1-S7 (2010-2012).
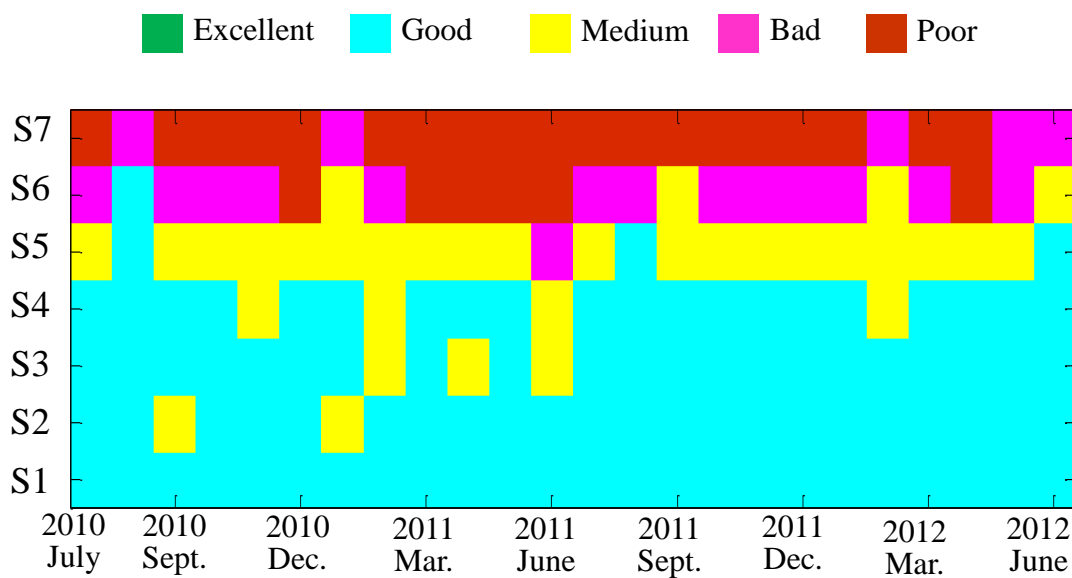
Fig. 3.27 Monthly WQI values (including TP) at water quality monitoring stations S1-S7 by incorporating the reconstructed monthly TP concentrations (2010-2012).

and thus causes the fluctuations of SS, EC and TP concentrations. According to the reconstructed TP time series (Fig. 3.25) and the threshold (0.4 mgl⁻¹) of TP concentration distribution (the red dotted line in Fig. 3.26), TP concentrations are the highest and fluctuate seriously at downstream monitoring stations S6 and S7. Finally,
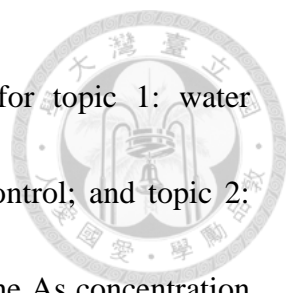
monthly WQI (including TP) can be calculated by incorporating the reconstructed monthly TP data, and the WQI then be converted into five water quality levels for convenience: 0-25 (poor); 25-50 (bad); 50-70 (medium); 70-90 (good); and 90-100 (excellent) (Fig. 3.27). Monthly WQI also indicates that water quality is either bad or poor at S6 and S7 in almost every month.

4. Conclusion and suggestion
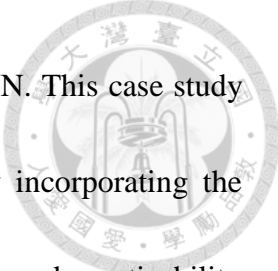
4.1 Conclusion

This dissertation focuses on the two main problems that are commonly encountered in hydro-environmental systems: water quantity and water quality management, and endeavors to develop novel dynamic artificial neural networks and modeling schemes to solve these problems. ANNs have the ability to approximate nonlinear functions as valuable tools for modeling time series and are considered an alternative computational approach to modeling physical-based problems. RNNs are a class of ANNs and are computationally powerful nonlinear models capable of extracting dynamic behaviors from complex systems through internal dynamic recurrence, which are suitable for the applications of hydro-environmental systems. Therefore, this dissertation develops the reinforced online learning algorithm for RNNs (R-RTRL NN)

and the systematical dynamic-neural modeling (SDM) scheme for topic 1: water quantity issues for reservoir inflow forecasting and urban flood control; and topic 2: water quality issues for spatio-temporal estimation with respect to the As concentration in groundwater and the TP concentration in a river basin. A total of four case studies are investigated in this dissertation.

The case study of the *"Reinforced recurrent neural networks for multi-step-ahead flood forecastings"* is devoted to dealing with the connatural limitation of online learning algorithms that is caused by a lack of future target values for MSA forecasting. A novel R-RTRL NN is proposed to not only adequately utilize the antecedent information of the observations as well as model outputs but also strengthen their usefulness to mitigate time-lag phenomena as well as increases model accuracy in MSA forecasting. The rigorous demonstration with respect to the superiority of MSA R-RTRL NN necessitates the use of a benchmark chaotic time series and the real-world application of the flood time series induced by typhoons at the Shihmen Reservoir in northern Taiwan. For comparison purpose, the original RTRL NN, the Elman NN and the BPNN are also performed. In the cases of benchmark time series, results indicate that the proposed R-RTRL NN has much better performance than comparative models for MSA forecasts. When modeling the flood time series, the proposed R-RTRL NN also shows great superiority on 2SA, 4SA and 6SA forecasts with significant reductions

in time-lag effects than the original RTRL NN, Elman NN and BPNN. This case study demonstrates that the developed R-RTRL algorithm for RNNs by incorporating the closest antecedent information into the online learning process has good practicability and produces high accuracy for MSA forecasts.

In the case study of the *"Real-time multi-step-ahead water level forecasting by recurrent neural networks for urban flood control"*, three ANN models (one static – the BPNN, two dynamic – the Elman NN and the NARX network) are developed to make forecasts on the evolution of water level at FSP as a function of current FSP water level and rainfall information based on the inputs extracted by an advanced factor selection method (GT) for allowing sufficient time advance to warm up the pumping system and enhancing secure pumping operations to prevent flooding of the city. The temporal resolution of water level and rainfall data is 10 minutes, and the forecasting horizon is 60 minutes (i.e., 6 time-step-ahead). The results demonstrate that the GT can efficiently identify the effective rainfall factors as important inputs to the ANNs for obtain promising forecasting results; and the NARX network has higher applicability than the BPNN and the Elman NN, in term of lower RMSE and higher CC and CE values for multi-step-ahead forecasts in both scenarios. The NARX network can well forecast the hydrograph of FSP water level and maintain the water level trail with less fluctuation.

In technical aspects, the outputs of the static network (the BPNN) depends solely on observed data, whereas the outputs of the dynamic networks incorporate observed data with time delay units through recurrent connections and thus significant contribution could be made to the output values. The dynamic networks have the merits to effectively discover the long–term dependencies through their recursive outputs and mitigate the fluctuation problem (instability) in their outputs.

In sum, the proposed SDM with two core methods of the GT and the NARX network can well construct multi-step-ahead hydrological water level forecasting models for urban flood control pumping. The results of this case study are beneficial to the identification of inundation risks induced by inner stormwater and can be incorporated into suitable operational strategies for enhancing the pumping efficiency at inundation-prone areas.
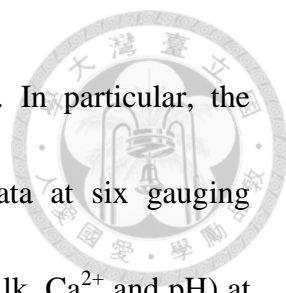
The third case study is the *"Regional estimation of groundwater Arsenic concentrations through Systematical Dynamic-neural Modeling"*. The blackfoot disease in the Yun-Lin County of Taiwan has been verified to be associated with high As concentrations in groundwater. Due to great concern for the potential effects of As on human health, there is a growing need for efficiently modeling the presence and the amount of As in groundwater. In this case study, the proposed SDM that incorporates the GT into the NARX network accompanied with three optional statistical techniques

is utilized to adequately estimate As concentration in the area of Yun-Lin County. The BPNN is adopted as a comparative model. The modeling processes and related results suggest that 1) the GT can effectively extract non-trivial factors that affect the target variable; 2) the Bayesian regularization method that constrains the weight values of the network does improve network generalizability; 3) the cross validation technique can produce a low-bias estimator of the generalization ability of networks; 4) the NARX network can provide reliable estimation of As concentration at both gauged and ungauged sites; and 5) the IK suitably derives the probability maps of As concentration under the threshold of WHO drinking water standard in the study area.

The results demonstrate that the NARX network produces much better performance than the BPNN. The configured NARX network can suitably and accurately estimate As concentrations at 20 ungauged sites in five testing years (all the R^2 values are high (0.82 to 0.95)), whereas the BPNN fails to provide suitable estimations (all the R^2 values are low (0.17 to 0.41)). It proves that the recurrent connections of output information (As concentrations) to the NARX network itself makes significant contribution to the accuracy of the spatio-temporal estimation model.

Finally, the IK can suitably derive the probability maps of As concentration under the threshold of the WHO drinking water standard in the study area, which is meaningful and useful for the authorities to manage water resources so that prevent

residents from using and drinking As-contaminated groundwater. In particular, the

construction of the proposed SDM requires As concentration data at six gauging

stations and data of three easily-measured water quality variables (Alk, Ca^{2+} and pH) at

six gauging stations and the other twenty ungauged sites. It merely requires data of three

three-easily measured water quality variables for the constructed SDM to effectively

and suitably estimate As concentrations at ungauged sites. This approach could

significantly reduce the manpower cost of monitoring wells and effectively provide

reliable estimation of As concentration at ungauged sites. In summary, the proposed

SDM scheme for the estimation of As concentration using on-site measurement data of

other water quality variables can be an alternative way to quantify As contamination

and to provide predictive information for better public health management.

In the fourth case study of the *"Modeling spatio-temporal total phosphate (TP)*

concentration through Systematical Dynamic-neural Modeling", the developed SDM is

utilized to estimate seasonal TP concentrations in a river channel as a function of

easily-measured water quality parameters with an extension to monthly TP estimation to

provide the monthly WQI (including TP) in response to the sudden changes in river

pollution levels caused by natural processes and/or human activities for timely

hydro-environmental management. Two different neural network models (one static/one

dynamic) are developed based on the inputs selected as significant factors from a

preliminary data analysis. Their performances are compared to derive guidelines for model selection. The prediction time horizon can be reduced from three months (seasonal) to one month. Results demonstrate that the NARX network (dynamic network) gives better estimation than the BPNN (static network). The proposed SDM can be beneficial to water authorities for hydro-environmental management with countermeasures in a shorter time scale (from quarterly to monthly). And it is believed that the proposed SDM can be reliably applied to spatio-temporal estimation for missing, hazardous or costly data of interest.

4.2 Suggestion

This dissertation have developed the reinforced online learning algorithm for RNNs (R-RTRL NN) and the systematical dynamic-neural modeling (SDM) scheme for water quantity and water quality issues in hydro-environmental systems.
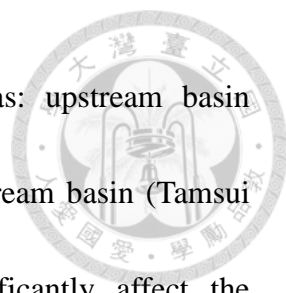
The sequential formulation of the R-RTRL NN is derived, and its reliability and applicability are demonstrated through two- to six-step-ahead forecasting made for a famous benchmark chaotic time series and a reservoir inflow case in Taiwan. In the future, it can be further applied to other rivers and/or urban basins for MSA stream flow and/or water level forecasting, and thus the practical conditions under various rainfall-runoff patterns suitable for adopting the online R-RTRL NN can be identified.

In the other way, although batch learning is more popular than online learning for training ANNs, online learning is still useful in certain situations. Online learning algorithms can be utilized when the input-output patterns of datasets change over time. For example, input data may not be available at every time step, so the estimated input values will usually be used under this circumstance. As a result, the input-output mapping pattern will change over time for different input data sources (observed input-output and estimated input-output). Moreover, the proposed reinforced online learning algorithm can be combined not only with the RNN but also with any arbitrary type of models for MSA forecasting, which can be further explored and/or applied.

Second, the SDM whose kernel consists of the GT for key factor selection and the NARX network for spatio-temporal estimation is proposed. The SDM is then applied to 1) urban flood control problems for exploring the contribution of recurrent connections of the NARX network; and 2) the spatio-temporal estimation of As and TP concentration for providing water authorities with useful information to deal with groundwater and river basin management. The proposed SDM has been applied with success to three different study areas: Yu-Cheng catchment; Yun-Lin coastal area; and Dahan River. It is believed that the SDM can also be easily and appropriately applied to larger-scale and more complex hydro-environmental systems, e.g., the Shihmen Reservoir-Tamsui estuary river system, for modeling target water quality parameters.
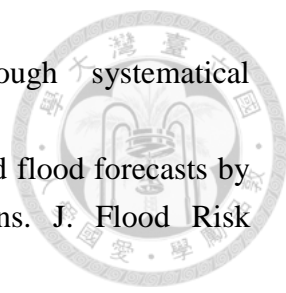
First, the river system can be divided into three sub-areas: upstream basin (Shihmen Reservoir); midstream basin (Dahan River); and downstream basin (Tamsui River). The GT can extract the non-trivial factors that significantly affect the fluctuations of target parameters in these three different sub-areas, respectively. Then a series-connection of three different NARX networks associated with these three sub-basins can be constructed. It means the estimated output of the NARX network associated with the upstream basin is one of the exogenous input to the NARX network associated with the midstream basin, and the estimated output of the NARX network associated with the midstream basin is also one of the exogenous input of the NARX network associated with the downstream basin. In this way, the SDM bears the ability in modeling the external (geographic dependency) and internal (temporal dependency) characteristics of the target variable in the whole river system.

5. Reference

1) Agalbjorn, S., Koncar, N., Jones, A. J., 1997. A note on the gamma test, Neural Comput. Appl. 5 131-133.

2) Agricultural Engineering Research Center, 2008. Survey, analysis and assessment of groundwater quality in Taiwan area. Water Resource Agency, Taiwan.

3) Ahmad, Z., Jie, Z., 2002. Improving long range prediction for nonlinear process modelling through combining multiple neural networks Control Applications. Proc. IEEE Conf. Control Appl., 962, 966-971.

4) Alvisi, S., Mascellani, G., Franchini, M., Bardossy, A., 2006. Water level forecasting through fuzzy logic and artificial neural network approaches. Hydro. Earth Syst. Sci. 10(1), 1-17.

5) Anawar, H.M., Akai, J., Sakugawa, H., 2004. Mobilization of arsenic from subsurface sediments by effect of bicarbonate ions in groundwater. Chemosphere. 54(6), 753-762.

6) Antar, M.A., Elassiouti, I., Allam, M.N., 2006. Rainfall-runoff modelling using artificial neural networks technique: a Blue Nile catchment case study. Hydrol. Processes. 20, 1201-1216.

7) APHA (American Public Health Association), 1992. Standard Methods for the Examination of Water and Waste Water, 18th ed., APHA, American Water Works Association, and Water Pollution Control Federation, Washington, DC.

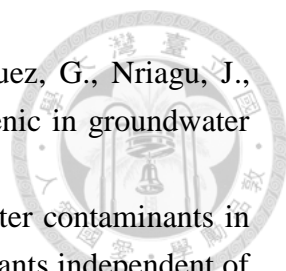8) Appelo, C.A.J., Van Der Weiden, M.J.J., Tournassat, C., Charlet, L., 2002. Surface Complexation of Ferrous Iron and Carbonate on Ferrihydrite and the Mobilization of Arsenic. Environ. Sci. Technol. 36(14), 3096-3103.

9) Ardalani-Farsa, M., Zolfaghari, S., 2010. Chaotic time series prediction with residual analysis method using hybrid Elman-NARX neural networks. Neurocomput. 73(13-15), 2540-2553.

10) Assaad, M., Boné, R., Cardot, H., 2005. Study of the behavior of a new boosting algorithm for recurrent neural networks. Lect. Notes Comput. Sci., pp. 169-174.

11) Austin, N. R., Prendergast, J. B., Collins, M. D., 1996. Phosphorus losses in irrigation runoff from fertilized pasture. J. Environ. Qual. 25(1), 63-68.

12) Ayotte, J.D., Montgomery, D.L., Flanagan, S.M., Robinson, K.W., 2003. Arsenic in Groundwater in Eastern New England: Occurrence, Controls, and Human Health Implications. Environ. Sci. Technol. 37(10), 2075-2083.

13) Azevedo, L. B., Henderson, A. D., van Zelm, R., Jolliet, O., Huijbregts, M.A., 2013. Assessing the importance of spatial variability versus model choices in life cycle impact assessment: the case of freshwater eutrophication in Europe. Environ. Sci. Technol. 47 (23), 13565-13570.

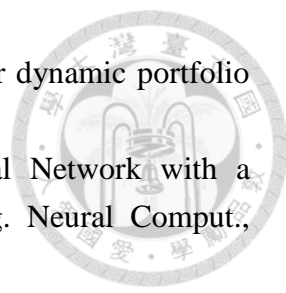14) Bargaouia, Z.K., Chebbib, A., 2009. Comparison of two kriging interpolation

methods applied to spatiotemporal rainfall. J. Hydrol. 365(1-2), 56-73.

15) Bauer, M., Fulda, B., Blodau, C., 2008. Groundwater derived arsenic in high carbonate wetland soils: Sources, sinks, and mobility. Sci. Total Environ. 401, 109-120.

16) Bauser, G., Franssen, H.-J. H., Kaiser, H.-P., Kuhlmann, U., Stauffer, F., Kinzelbach, W., 2010. Real-Time management of an urban groundwater well field threatened by pollution. Environ. Sci. Technol. 44(17), 6802-6807.

17) Bertin, M. J., Moeller, P., Guillette, Jr. L. J., Chapman, R. W., 2013. Using machine learning tools to model complex toxic interactions with limited sampling regimes. Environ. Sci. Technol. 47(6), 2728-2736.

18) Besaw, L.E., Rizzo, D.M., Bierman, P.R., Hackett, W.R., 2010. Advances in ungauged streamflow prediction using artificial neural networks. J. Hydrol. 386(1-4), 27-37.

19) BGS-DPHE, 2001. Arsenic Contamination of Groundwater in Bangladesh; BGS Technical Report WC/00/19. British Geological Survey, Keyworth, UK.

20) Bothe, J.V., Brown, P.W., 1999. The stabilities of calcium arsenates at 23±1 ∘C. J. Hazard.Mater. 69, 197-207.

21) Bowes M. J., House W. A., Hodgkinson R. A., 2003. Phosphorus dynamics along a river continuum. Sci. Total Environ. 313, 199-212.

22) Brockwell, P.J., Davis, R.A., 1991. Time Series: Theory and Methods. Springer-Verlag, New York.

23) Bruen, M., Yang, J.Q., 2006. Combined hydraulic and black-box models for flood forecasting in urban drainage systems, J. Hydrol. Eng. 11(6), 589-596.

24) Cao, L., 1997. Practical method for determining the minimum embedding dimension of a scalar time series. Phys. D. 110(1-2), 43-50.

25) Carpenter S. R., Caraco N. F., Correll D. L., Howarth R. W., Sharpley A. N., Smith V. H., 1998. Nonpoint pollution of surface waters with phosphorus and nitrogen. Ecol. Appl. 8(3), 559-568.

26) Central Geological Survey, 1999. Project of groundwater monitoring network in Taiwan during first stage-research report of Chou-Shui River alluvial fan, Taiwan. Water Resource Agency, Taiwan.

27) Chakraborti, D., Rahman, M.M., Das, B., Murrill, M., Dey, S., Mukherjee, S.C., Dhar, R.K., Biswas, B.K., Chowdhury, U.K., Roy, S., Sorif, S., Selim, M., Rahman, M., Quamruzzaman, Q., 2010. Status of groundwater arsenic contamination in Bangladesh:A 14-year study report. Water Res. 44, 5789-5802.

28) Chang, F. J., Chang, L.C., Huang, H.L., 2002, Real time recurrent learning neural network for streamflow forecasting. Hydrol. Processes. 16(13), 2577-2588.

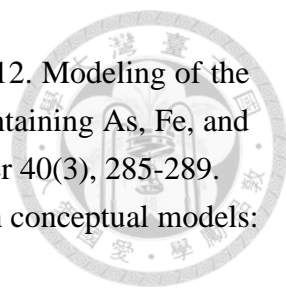29) Chang, F. J., Chen, P. A., Liu, C. W., Liao, V. H. C., Liao, C. M., 2013. Regional

estimation of groundwater arsenic concentrations through systematical dynamic-neural modeling. J. Hydrol. 499, 265-274.

30) Chang, F. J., Chiang, Y. M., Ho, Y. H., 2013. Multi-step-ahead flood forecasts by neuro-fuzzy networks with effective rainfall-runoff patterns. J. Flood Risk Manage..

31) Chang, F. J., Chiang, Y. M., Tsai, M. J, Shieh, M. C., Hsu, K. L., Sorooshian, S., 2014. Watershed rainfall forecasting using neuro-fuzzy networks with the assimilation of multi-sensor information. J. Hydrol. 508, 374-384.

32) Chang, F. J., Kao, L.S., Kuo, Y.M., Liu, C.W., 2010. Artificial neural networks for estimating regional arsenic concentrations in a blackfoot disease area in Taiwan. J. Hydrol. 388, 65-76.

33) Chang, F. J., Wang, K. W., 2013. A systematical water allocation scheme for drought mitigation. J. Hydrol. 507, 124-133.

34) Chang, F.J., Sun, W., 2013. Modeling regional evaporation through ANFIS incorporated solely with remote sensing data. Hydrol. Earth Syst. Sci. Discuss. 10, 6153-6192.

35) Chang, F.J., Sun, W., Chung, C. H., 2013. Dynamic factor analysis and artificial neural network for estimating pan evaporations at multiple stations in northern Taiwan. Hydrol. Sci. J. 58(4), 813-825.

36) Chang, L. C., Chen, P. A., Chang, F. J., 2012. A reinforced two-step-ahead weight adjustment technique for on-line training of recurrent neural networks. IEEE Trans. Neural Netw. Learn. Syst. 23(8), 1269-1278.

37) Chen, C.J., Hsueh, Y.M., Lai, M.S., Shyu, M.P., Chen, S.Y., Wu, M.M., Kuo, T.L., Tai, T.Y., 1995. Increased prevalence of hypertension and long-term arsenic exposure. Hypertension. 25,53-600.

38) Chen, C.S., Chen, B.P.T., Chou, F.N.F., Yang, C.C., 2010. Development and application of a decision group Back-Propagation Neural Network for flood forecasting. J. Hydrol. 385(1-4), 173-182.

39) Chen, P. A., Chang, L. C., Chang, F. J., 2013. Reinforced recurrent neural networks for multi-step-ahead flood forecasts. J. Hydrol. 497, 71-79.

40) Chiang, Y. M., Chang, L. C., Chang, F. J., 2004. Comparison of static-feedforward and dynamic-feedback neural networks for rainfall-runoff modeling. J. Hydrol. 290, 297-311.

41) Chiang, Y. M., Chang, L. C., Tsai, M. J., Wang, Y. F., Chang, F. J., 2010. Dynamic Neural Networks for Real-Time Water Level Predictions of Sewerage Systems-covering gauged and unguaged sites. Hydrol. Earth Syst. Sci., 14, 1309-1319.

42) Chiang, Y. M., Hsu, K. L., Chang, F. J., Yang, H., Soroosh, S., 2007. Merging

multiple precipitation sources for flash flood forecasting. J. Hydrol., 340, 183-196.

43) Chiou, H. Y., Hsueh, Y. M., Hsieh, L. L., Hsu, L. I., Hsu, Y. H., Hsieh, F. I., Wei, M. L., Chen, H. C., Yang, H. T., Leu, L. C., 1997. Arsenic methylation capacity, body retention, and null genotypes of glutathione S-transferase M1 and T1 among current arsenic-exposed residents in Taiwan. Mutat.Res. 386, 197-207.

44) Cho, K. H., Kao, S. S., Pachepsky Y. A., Kim, K. W., Kim. J. H., 2011. Prediction of contamination potential of groundwater arsenic in Cambodia, Laos, and Thailand using artificial neural network. Water Res. 45(17) , 5535-5544.

45) Correll, D. L., 1998. The role of phosphorus in the eutrophication of receiving waters: A review. J. Environ. Qual. 27(2), 261-266.

46) Coulibaly, P. Baldwin, C. K., 2005. Nonstationary hydrological time series forecasting using nonlinear dynamic methods. J. Hydrol., 307, 164-174

47) Coulibaly, P., 2010. Reservoir computing approach to Great Lakes water level forecasting. J. Hydrol. 381, 76-88.

48) Coulibaly, P., Evora, N.D., 2007. Comparison of neural network methods for infilling missing daily weather records. J. Hydrol. 341(1-2), 27-41.

49) Davis, J. R., Koop, K., 2006. Eutrophication in Australian rivers, reservoirs and estuariesa- A southern hemisphere perspective on the science and its implications. Hydrobiologia 559(1), 23-76.

50) Dodds, W. K., Bouska, W. W., Eitzmann, J. L., Pilger, T. J., Pitts, K. L., Riley, A. J., Schloesser, J. T., Thornbrugh, D. J., 2009. Eutrophication of U.S. freshwaters: analysis of potential economic damages. Environ. Sci. Technol. 43(1), 12-19.

51) DSD, Drainage Services Department, 2000. Stormwater drainage manual - planning, design and management, third ed. Hong Kong.

52) Duda A. M., 1993. Addressing nonpoint sources of water-pollution must become an international priority. Water Sci. Technol. 28(3–5), 1-11.

53) Durrant, P.J., 2001. winGamma: a Non-linear Data Analysis and Modeling Tool with Applications to Flood Prediction. PhD thesis. Department of Computer Science, Cardiff University, Wales, UK.

54) Elman, J.L., 1990. Finding structure in time. Cognit. Sci. 14, 179-211.

55) Ford, R.G., Fendorf, S., Wilkin, R.T., 2006. Introduction: Controls on arsenic transport in near-surface aquatic systems. Chem. 228, 1-5.

56) Fraser, A.M., Swinney, H.L., 1986. Independent coordinates for strange attractors from mutual information. Phys. Rev. A., 33(2), 1134.

57) Giri, A.K., Patel, R.K., Mahapatra, S.S., 2011. Artificial neural network (ANN) approach for modelling of arsenic (III) biosorption from aqueous solution by living cells of Bacillus cereus biomass. Chem. Eng. J. 178, 15-25.

58) Goovaerts, P., AvRuskin, G., Meliker, J., Slotnick, M., Jacquez, G., Nriagu, J., 2005. Geostatistical modeling of the spatial variability of arsenic in groundwater of southeast Michigan. Wat. Resour. 41.

59) Grčić, I., Puma, J. L., 2013. Photocatalytic degradation of water contaminants in multiple photoreactors and evaluation of reaction kinetic constants independent of photon absorption, irradiance, reactor geometry, and hydrodynamics. Environ. Sci. Technol. 47(23), 13702-1371.

60) Gronewold, A. D., Borsuk, M. E., 2010. Improving water quality assessments through a hierarchical bayesian analysis of variability. Environ. Sci. Technol. 44 (20), 7858-7864.

61) Hirasawa, K., Murata, J., Jinglu, H., Chunzhi, J., 2000. Universal learning network and its application to robust control. IEEE Trans. Syst., Man, Cybern. B, Cybern. 30, 419-430.

62) Hong, Y.S.T., 2012. Dynamic nonlinear state-space model with a neural network via improved sequential learning algorithm for an online real-time hydrological modeling. J. Hydrol. 468–469(0), 11-21.

63) Huang W. C., Hsieh C. L., 2010. Real-time reservoir flood operation during typhoon attacks. Water Resour. Res. 46, W07528.

64) Hung, N.Q., Babel, M.S., Weesakul, S., Tripathi, N.K., 2009. An artificial neural network model for rainfall forecasting in Bangkok, Thailand. Hydrol. Earth Syst. Sci. 13, 1413-1425.

65) Jaeger, H., Haas, H., 2004. Harnessing Nonlinearity: Predicting Chaotic Systems and Saving Energy in Wireless Communication. Sci., 304(5667), 78-80.

66) Jiang, C., Song, F., 2011. Sunspot forecasting by using Chaotic time series analysis and NARX network. J. Chem. Phys. 6, 1424-1429.

67) Jothiprakash, V., Magar R.B., 2012. Multi-time-step ahead daily and hourly intermittent reservoir inflow prediction by artificial intelligent techniques using lumped and distributed data. J. Hydrol., 450-451, 293-307

68) Journel, A.G., 1983. Nonparametric estimation of spatial distributions. J. Int. Assoc. Math. Geol. 15(3), 445-468.

69) Juang, K.W., Lee, D.Y., 1998. Simple indicator kriging for estimating the probability of incorrectly delineating hazardous areas in a contaminated site. Environ. Sci. Technol. 32, 2487-2493.

70) Keon N.E., Swartz C.H., Brabander D.J., Harvey C., Hemond H.F., 2001. Validation of an arsenic sequential extraction method for evaluating mobility in sediments. Environ. Sci. Technol. 35, 2778-2784.

71) Khalil, B., Ouarda, T.B.M.J., St-Hilaire, A., 2011. Estimation of water quality characteristics at ungauged sites using artificial neural networks and canonical

correlation analysis. J. Hydrol. 405(3-4), 277-287.

72) Kim, M.J., Nriagu, J., Haack, S., 2000. Carbonate ions and arsenic dissolution by groundwater. Environ. Sci. & Technol. 34(15), 3094-3100.

73) Kohavi, R., 1995. A study of cross-validation and bootstrap for accuracy estimation and model selection. Proc. 14th Int. Jt. Conf. Artif. Intell. 2, Quebec, Canada.

74) Koncar, N., 1997. Optimisation methodologies for direct inverse neurocontrol. PhD thesis, Department of computing, Imperial College of Science, Technology and Medicine, University of London.

75) Krishna, B., Rao, Y.R.S., Vijaya, T., 2008. Modelling groundwater levels in an urban coastal aquifer using artificial neural networks. Hydrol. Processes. 22, 1180-1188.

76) Kuo, Y.M., Chang, F.J., 2010. Dynamic Factor Analysis for Estimating Groundwater Arsenic Trends. J. Environ. Qual. 39, 176-184.

77) Li, C.U., He, S.B., Liao, X.F., Yu, J.B., 2002. Using recurrent neural network for adaptive predistortion linearization of RF amplifiers. Int. J. RF Microwave Comput.-Aided Eng., 12, 125-130.

78) Liao, V.H.C., Chu, Y.J., Su, Y.C., Hsiao, S.Y., Wei, C.C., Liu, C.W., Liao, C.M., Shen, W.C., Chang, F.J., 2011. Arsenite-oxidizing and arsenate-reducing bacteria associated with arsenic-rich groundwater in Taiwan. J. Contam. Hydrol. 123, 20-29.

79) Lin, T, Horne, B.G., Tino, P, Giles, C.L., 1996. Learning long-term dependencies in NARX recurrent neural networks. IEEE Trans. Neural Netw. 7 (6) ,1424-1438.

80) Liu, C.W., Huang, Y.K., Hsueh, Y.M., Lin, K.H., Jang, C.S., Huang, L.P., 2008. Spatiotemporal distribution of arsineic species of oysters (Crassostreagigas) in the coastal area of southwestern Taiwan. Environ. Monit. Assess. 138, 181-190.

81) Liu, C.W., Jang, C.S., Liao, C.M., 2004. Evaluation of arsenic contamination potential using indicator kriging in the Yun-Lin aquifer (Taiwan). Sci. Total Environ. 321(1-3), 173-188.

82) Liu, C.W., Lin, K.H., Kuo, Y.M., 2003. Application of factor analysis in the assessment of groundwater quality in a blackfoot disease area in Taiwan. Sci. Total Environ. 313(1-3), 77-89.

83) Liu, C.W., Lin, W.S., Shang, C., Liu, S.H., 2001. The effect of claydehydration on land subsidence in Yun-Lin coastal area, Taiwan. Environ. Geol. 40(4-5), 518-527.

84) Liu, C.W., Wang, S.W., Jang, C.S., Lin, K.H., 2006. Occurrence of Arsenic in Ground Water in the Choushui River Alluvial Fan, Taiwan. J. Environ. Qual. 35, 68-75.

85) Liu, Q.S., Guob, Z.S., Wang, J., 2012. A one-layer recurrent neural network for

constrained pseudoconvex optimization and its application for dynamic portfolio optimization. Neural Networks, 26, 99-109.

86) Liu, Q.S., Wang, J., 2008. A One-Layer Recurrent Neural Network with a Discontinuous Activation Function for Linear Programming. Neural Comput., 20(5), 1366-1383.

87) Liu, W., Moran, C. J., Vink, S., 2011. Quantitative risk-based approach for improving water quality management in mining. Environ. Sci. Technol. 45(17), 7459-7464.

88) Loke, E., Warnaars, E.A., Jacobsen, P., Nelen, F., Almeida, M.D., 1997. Artificial neural networks as a tool in urban storm drainage, Water Sci. Technol. 36(8-9), 101-109.

89) Ma, Q. L., Zheng, Q. L., Peng, H., Zhong, T. W., Qin, J. W., 2008. Multi-step-prediction of chaotic time series based on co-evolutionary recurrent neural network. Chin. Phys. B 17(2), 536

90) MacKay, D. J. C., 1992. Bayesian interpolation. Neural Comput. 4, 415-447.

91) Mackenthun, K. M., 1973. Toward a cleaner aquatic environment. United States Environmental Protection Agency: Washington, DC.

92) MacKey, M.C., Glass, L., 1977. Oscillation and chaos in physiological control systems, Sci., 197 (4300), 287-289.

93) Maity, R., Kumar, D.N., 2008. Basin-scale stream-flow forecasting using the information of large-scale atmospheric circulation phenomena, Hydrol. Process. 22(5), 643–650.

94) Matheron, G., 1963. Principles of geostatistics. Econ. Geol. 58(8), 1246-1266.

95) McDowell, R. W., Littlejohn, R. P., Blennerhassett, J. D., 2010. Phosphorus fertilizer form affects phosphorus loss to waterways: a paired catchment study. Soil Use Manage. 26(3), 365-373.

96) McNamara, J. P., Kane, D. L., Hobbie, J. E., Kling, G. W., 2008. Hydrologic and biogeochemical controls on the spatial and temporal patterns of nitrogen and phosphorus in the Kuparuk River arctic Alaska. Hydrol. Processes 22, 3294-3309.

97) Menezes Jr, J. M. P., Barreto, G. A., 2008. Long-term time series prediction with the NARX network: An empirical evaluation. Neurocomput. 71(16-18), 3335-3343.

98) Mihaljevic, M., Ponavic, M., Ettler, V., ˇSebek, O., 2003. A comparison of sequential extraction techniques for determining arsenic fractionation in synthetic mineral mixtures. Anal. Bioanal. Chem. 377, 723-729.

99) Moghaddamnia, A., Ghafari, M., Piri, J., Amin, S., Han, D., 2009. Evaporation estimation using artificial neural networks and adaptive neuro-fuzzy inference system techniques. Adv. Water Resour. 32, 88-97.

100) Mondal, P., Mohanty, B., Balomajumder, C., Saraswati, S., 2012. Modeling of the Removal of Arsenic Species from Simulated Groundwater Containing As, Fe, and Mn: A Neural Network Based Approach. Clean: Soil, Air, Water 40(3), 285-289.

101) Nash, J.E., Sutcliffe, J.V., 1970. River flow forecasting through conceptual models: 1. A discussion of principles, J. Hydrol., 10, 282-290.

102) Nasseri, M., Asghari, K., Abedini, M.J., 2008. Optimized scenario for rainfall forecasting using genetic algorithm coupled with artificial neural network. Expert Syst. Appl. 35(3), 1415-1421.

103) National Research Council, 1999. Arsenic in drinking water. Washington, DC, National Academy Press.

104) Nayak, P.C., Venkatesh, B., Krishna, B., Jain, S.K., 2013. Rainfall runoff modelling using conceptual, data driven and wavelet based computing approach. J. Hydrol. 493, 57-67.

105) Nikolos, I. K., Stergiadi, M., Papadopoulou, M. P., Karatzas, G. P., 2008. Artificial neural networks as an alternative approach to groundwater numerical modelling and environmental design. Hydrol. Processes 22, 3337-3348.

106) Nishimura, T., Robins, R.G., 1998, Are-evaluation of the solubility and stability regions of calciumarsenites and calcium arsenates in aqueous solution at 25 ∘C. Mineral Proc. and Extr.Met. 18, 283-308.

107) Noori, R., Karbassi, A.R., Moghaddamnia, A., Han, D., Zokaei-Ashtiani, M.H., Farokhnia, A., Gousheh, M.G., 2011. Assessment of input variables determination on the SVM model performance using PCA, Gamma test, and forward selection techniques for monthly stream flow prediction. J. Hydrol. 401, 177-189.

108) Noori, R., Sabahi, M.S., Karbassi, A.R., 2010. Evaluation of PCA and gamma test techniques on ANN operation for weekly solid waste predicting. J.ournal of Environ.mental Manage.ment 91, 767-771.

109) Nourani ,V., Komasi, M., Almai, M.T., 2011. A Hybrid Wavelet-Genetic Programming Approach to Optimize ANN Modelling of Rainfall-Runoff Process. J. Hydrol. Eng., 17, 724-741.

110) Nourani,V., Sayyah Frad, M., 2012. Sensitivity analysis of the artificial neural network outputs in simulation of the evaporation process at different climatologic regimes. Adv. Eng. Software, 47(1), 127-146.

111) Park, J.M., Lee, J.S., Lee, J.U., Chon, H.T. and Jung, M.C., 2006. Microbial effects on geochemical behavior of arsenic in As-contaminated sediments. J. Geochem. Explor. 88(1-3), 134-138.

112) Parlos, A.G., Rais, O.T., Atiya, A.F., 2000. Multi-step-ahead prediction using dynamic recurrent neural networks. Neural Networks, 13, 765-786.

113) Pierce, M.L., Moore, C.B., 1982. Adsorption of Arsenite and Arsenate on

Amorphous Iron Hydroxide. Water Res. 16(7), 1247-1253.

114) Polizzotto, M.L., Harvey, C.F., Li, G., Badruzzman, B., Ali, A., Newville, M., Sutton, S., Fendorf, S., 2006. Solid-phases and desorption processes of arsenic within Bangladesh sediments. Chemical Geology. 228(1-3), 97-111.

115) PUB., Singapore's National Water Agency, 2013. Code of Practice on Surface Water Drainage, sixth ed. Singapore.

116) Purkait, B., Kadam, S.S., Das, S.K., 2008. Application of artificial neural network model to study arsenic contamination in groundwater of Malda District, Eastern India. J. Environ. Inf. 12 (2), 140-149.

117) Quality Criteria for Water 1986 [The Gold Book], 1987. United States Environmental Protection Agency: Washington, DC.

118) Rahman, M., Tondel, M., Chowdhury, I.A., Axelson, O., 1999. Relations between exposure to arsenic, skin lesions, and glucosuria. Occup. Environ. Med. 56(4), 277-281.

119) Rothwell, J.J., Taylor, K.G., Ander, E.L., Evans, M.G., Daniels, S.M., Allott, T.E., 2009. Arsenic retention and release in ombrotrophic peatlands. Sci. Total Environ. 407, 1405-1417.

120) Rumelhart, D.E., Hinton, G.E., Williams, R.J., 1986. Learning representations by back-propagating errors. Nature 323(6088), 533-536.

121) Sahoo, G.B., Ray, C., De Carlo, E.H., 2006. Use of neural network to predict flash flood and attendant water qualities of a mountainous stream on Oahu, Hawaii. J. Hydrol. 327(3-4), 525-538.

122) Sahoo, G.B., Schladow, S.G., Reuter, J.E., 2009. Forecasting stream water temperature using regression analysis, artificial neural network, and chaotic non-linear dynamic models. J. Hydrol. 378(3-4), 25-342.

123) Seibert, J., 2001. On the need for benchmarks in hydrological modeling, Hydrol. Process., 15(6), 1063-1064.

124) Serpen, G., Xu, Y., 2003. Simultaneous recurrent neural network trained with non-recurrent backpropagation algorithm for static optimisation. Neural Comput. Appl. 12, 1-9.

125) Serre, M.L., Kolovos, A., Christakos, G., Modis, K., 2003. An Application of the Holistochastic Human Exposure Methodology to Naturally Occurring Arsenic in Bangladesh Drinking Water. Risk Anal. 23(3), 515-528.

126) Shalev-Shwartz, S., Singer, Y., Ng, A.Y., 2004. Online and batch learning of pseudo-metrics. Proc. Twenty-First Int. Conf. Mach. Learn., Alberta, Canada, 94.

127) Shen, H. Y., Chang, L. C., 2013. Online multistep-ahead inundation depth forecasts by recurrent NARX networks. Hydrol. Earth Syst. Sci. 17, 935-945.

128) Sorjamaa, A., Hao, J., Reyhani, N., Ji Y., Lendasse, A., 2007. Methodology for

long-term prediction of time series. Neurocomput. 70(16-18), 2861-2869.

129) Stamm, C., Jarvie, H. P., Scott, T., 2014. What's more important for managing phosphorus: loads, concentrations or both? Environ. Sci. Technol. 48 (1), 23-24.

130) Stollenwerk, K.G., 2003. Geochemical processes controlling transport ofarsenic in groundwater: a review of adsorption, in: Stollenwerk,K.G., Welch, A.H. (Eds.), Arsenic in Groundwater. Springer US,New York, pp. 67-100.

131) Stone, M., 1974. Cross-validatory choice and assessment of statistical predictions. J. R. Stat. Soc., Ser. B 36, 111-147.

132) Su, H. T., McAvoy, T. J., 1991. Identification of chemical processes using recurrent networks. Proc. Amer. Contr. Conf. 3, 2314-2319.

133) Su, H. T., McAvoy, T.,J., Werbos, P., 1992. Long-term predictions of chemical processes using recurrent neural networks: A parallel training approach. Ind. Eng. Chem. Res. 31, 1338-1352.

134) Takens, F., 1981. Detecting strange attractors in turbulence. Dynamical Systems and Turbulence. Lect. Notes Math., 898, 366-381.

135) Tamoto, N., Endo, J., Yoshimoto, K., Yoshida, T., Sakakibara, T., 2008. Forecast-based operation method in minimizing flood damage in urban area. Int. Conf. Urban Drain. 11th, Edinburgh, Scotland, UK.

136) Thiessen, A.H., 1911. Precipitation averages for large areas. Mon. Weather Rev. 39(7), 1082-1084.

137) Toth, E., 2009. Classification of hydro-meteorological conditions and multiple artificial neural networks for streamflow forecasting. Hydrol. Earth Syst. Sci. 13, 1555-1566.

138) Tsai, M. J., Abrahart, R. J., Mount, N. J., Chang, F. J., 2014. Including spatial distribution in a data-driven rainfall-runoff model to improve reservoir inflow forecasting in Taiwan. Hydrol. Processes 28, 1055–1070.

139) Tsui, A.P.M., Jones, A.J., Guedes de Oliveira, A.G., 2002. The construction of smooth models using irregular embeddings determined by a gamma test analysis. Neural Comput. Appl. 10, 318-329.

140) Wang, Q., Li, S., Jia, P., Qi, C., Ding, F., 2013. A review of surface water quality models. Sci. World J. 2013.

141) Wang, S.W., Kuo, Y.M., Kao, Y.H., Jang, C.S., Maji, S.K., Chang, F.J., Liu, C.W., 2011. Influence of hydrological and hydrogeochemical parameters on arsenic variation in shallow groundwater of southwestern Taiwan. J. Hydrol. 408(3-4), 286-295.

142) Wang, S.W., Liu, C.W., Jang, C.S., 2007. Factors responsible for high arsenic concentrations in two groundwater catchments in Taiwan. Appl. Geochem. 22(2), 460-476.

143) Webster, R., McBratney, A.B., 1987. Mapping soil fertility at Broom's Barn by simple kriging. J. Sci. Food Agric. 38(2), 97-115.

144) Williams, R.J., Zipser, D., 1989. A learning algorithm for continually running fully recurrent neural networks. Neural Comput. 1, 270-280.

145) Winkel, L., Berg, M., Amini, M., Hug, S. J., & Johnson, C. A., 2008. Predicting groundwater arsenic contamination in Southeast Asia from surface parameters. Nature Geoscience, 1(8), 536-542.

146) Xie, J. X., Cheng, C. T., Chau, K. W., Pei, Y. Z., 2006. A hybrid adaptive time-delay neural network model for multi-step-ahead prediction of sunspot activity. Int. J. Environ. Pollut., 28, 364-381.

147) Xie, X., Wang, Y., Ellis, A.,Su, C., Li, J., Lia M., Duan, M., 2013. Delineation of groundwater flow paths using hydrochemical and strontium isotope composition: A case study in high arsenic aquifer systems of the Datong basin, northern China. J. Hydrol. 476, 87-96.

148) Yang C.C., Chang L.C., Chen C.S., 1999. Comparison of integrated artificial neural network with time series modeling for flood forecast. J. Hydro sci. Hydraul. Eng. 17(2), 37-50.

149) Yong, S., Yibin, L., Qun, W., Caihong, L., 2010. Multi-steps prediction of chaotic time series based on echo state network. Proc. IEEE Conf. Bio-Inspired Comput.: Theor. Appl., Changsha, China, 669-672.

150) Yu, W. H., Harvey, C. M., Harvey, C. F., 2003. Arsenic in groundwater in Bangladesh: A geostatistical and epidemiological framework for evaluating health effects and potential remedies. Water Resour. Res. 39(6), 1146.

151) Zabiri, H., Maulud, A., Omar, N., 2009. NN-based algorithm for control valve stiction quantification. WSEAS Trans. Syst. 4 (2), 88-97.

Appendix

Acronym list

Acronym	Definition
ANN	Artificial neural network
RNN	Recurrent neural network
MSA	Mulit-step-ahead
R-RTRL	Reinforced real-time recurrent learning
SDM	Systematical dynamic-neural modeling
NARX	Nonlinear autoregressive with eXogenous input
SP (mode)	Series-parallel
P (mode)	Parallel
GT	Gamma test
IK	Indicator kriging
FSP	Floodwater storage pond
WQI	Water quality index

Publications

1) SCI Journal:

Chang, L.C., **Chen, P.A.**, Chang, F.J.*, 2012, "A Reinforced Two-Step-Ahead Weight Adjustment Technique for On-Line Training of Recurrent Neural Networks", IEEE Transactions on Neural Networks and Learning Systems, 23(8): 1269-1278. (NSC100-2313-B-002-011-MY3) (Appendix A)

Chen, P.A., Chang, L.C., Chang, F. J.*, 2013, "Reinforced Recurrent Neural Networks for Multi-Step-Ahead Flood Forecasts", Journal of Hydrology, 497: 71-79. (NSC100-2313-B-002-011-MY3) (Appendix A)

Chang, F. J.*, **Chen, P.A.**, Liu, C.W., Liao, V.H.C., Liao, C.M., 2013, "Regional Estimation of Groundwater Arsenic Concentrations through Systematical Dynamic-neural Modeling", Journal of Hydrology, 499: 265-274. (NSC100-2313-B-002-011-MY3) (Appendix A)

Chang, F. J.*, **Chen, P.A.**, Lu, Y.R., Huang, E., Chang, K.Y., 2014, "Real-time multi-step-ahead water level forecasting by recurrent neural networks for urban flood control", Journal of Hydrology, 517: 836-846. (H-102-03-102124) (Appendix A)

Chang, F. J.*, Chung, C.H., **Chen, P.A.**, Liu, C.W., Coynel, A., Vachaud, G., 2014, "Assessment of arsenic concentration in stream water using neuro fuzzy networks with factor analysis", Science of the Total Environment, 494-495(2014): 202-210.

Chang, F. J.*, **Chen, P.A.**, Tsai, Y.H., Chang, L.C., Huang, J.C., 2014, "Modeling spatio-temporal total phosphate concentration through systematical analysis scheme", Environmental Science & Technology. (Submitted)

Chang, F. J.*, Tsai, Y.H., **Chen, P.A.**, Coynel, A., Vachaud, G., 2014, "Modeling water quality in an urban river using hydrological factors by artificial intelligence techniques", Journal of Environmental Management. (Submitted)

2) Book chapter:

Chang, F.J.*, Lo, Y.J., **Chen, P.A.**, Chang, L.C., Shieh, M.C., 2014, "Multi-Step-Ahead Reservoir Inflow Forecasting by Artificial Neural Networks", Knowledge-based Information Systems in Practice, Springer. (in press)

Chang, F.J.*, Tsai, W.P., Wang, Y.C., **Chen, P.A.**, Chang, L.C., Coynel, A., Vachaud, G., 2014, "Optimal reservoir operation strategy for balancing ecosystem and human needs", Proceedings of FRIEND-Water 2014. (in press)

3) Conference proceedings:

Yang, S. N., Lu Y. R., Chen, P. A., Chang, L. C., Chang F. J.*, 2014, "Using ANFIS to construct real-time optimal operation for pumping station" , PAWEES 2014, Taiwan.

Chang, F. J.*, Lu, Y.R., Chen, P.A., Chen, Y.H., 2013, "Intelligent real-time water level forecast models for pumping stations", PAWEES 2013, Korea.

Chen, P.A., Chang, F.J.*, Tsai, Y.H., Chang, L.C., 2013, "Optimal Strategy for Water Quality Control through Suitable Discharge of Urban River BasinUnder Imbalanced Hydro-Environmental Conditions", AOGS 10th Annual Meeting (AOGS2013), Brisbane, Australia.

Chang, F.J.*, Chen, P.A., Tsai, Y.H., Tung, C.P., Lin, Y.P., Yam, R.S.W., Chang, L.C., Huang, J.C., Kao, S.J., Chang, T.W., Fang, W.T., 2012, "Estimating Regional Total Phosphate Concentration in a River Basin through the NARX network", PAWEES 2012, Thailand.

Chen, P.A., Chang, L.C., Chang, F.J.*, 2012, "A reinforced multi-step-ahead online learning algorithm for recurrent neural networks", HydroPredict 2012, Austria.

林政華, 陳品安, 江衍銘, 張斐章*, 2011, "以類神經網路建構區域地下水砷濃度推估", 水利工程研討會論文集.

Kao, L.S., Chen, P.A., Liu, C.W., Liao, H.C., Liao, C.M., Chang, F.J.*, 2011, "Applying Genetic Algorithm to optimizing Artificial Neural Networks for estimating arsenic pollution in groundwater systems", PAWEES 2011, Taipei, Taiwan.

Chen, P.A., Chang, L.C., Chang, F.J.*, 2010, "An enhanced two-step-ahead recurrent neural network for prediction of inflow in reservoir", AGU 2010 Fall Meeting, San Francisco, USA.

李婉君, 陳品安, 王國威, 張麗秋, 2008, "以類神經網路為基礎的X3D虛擬實境模擬水庫即時操作-以石門水庫為例", 農業工程研討會.

江衍銘, 陳品安, 張斐章, 2008, "遞迴式類神經網路於降雨-逕流之探討", 農業工程研討會.

Research projects involved

1) The Development of Novel Static and Dynamic ANNs for Hydro-Environmental Systems, funded by the National Science Council, Taiwan, under Grant number 100-2313-B-002-011-MY3. (Appendix B)

2) Sustainable and Integrated Water Resources Management on Watersheds Facing Industrial Pollutions and Urbanization: A Comparative Study between Lot River (France) and Danshuei River (Taiwan), funded by the National Science Council, Taiwan, under Grant number 101-2923-B-002-001-MY3, in collaboration with the 'Agence Nationale de la Recherche', project TWIN-RIVERS under Grant number ANR-11-IS56-0003. (Appendix B)

3) 臺北市水災危險潛勢地區保全計畫之智慧型支流河川水位預測系統研究，台北市政府水利工程處。 (Appendix B)

4) 抽水站排水系統水位預報及抽水機組智慧型操作策略評估工作，台北市政府水利工程處。 (Appendix B)

Awards and scholarship

1) 101 學年度「中興工程研究發展基金會」 博士班研究生獎助學金。

2) 102 學年度 國立臺灣大學生物資源暨農學院 博士班學生學術論文獎，*"Reinforced Recurrent Neural Networks for Multi-Step-Ahead Flood Forecasts"*。

3) 獲國科會 102 年度「補助國內研究生出席國際學術會議」補助，至澳大利亞布利斯班參加亞洲大洋洲地球科學學會年會 (AOGS2013)。

Appendix A

Reinforced Two-Step-Ahead Weight Adjustment Technique for Online Training of Recurrent Neural Networks

Li-Chiu Chang, Pin-An Chen, and Fi-John Chang

Abstract—A reliable forecast of future events possesses great value. The main purpose of this paper is to propose an innovative learning technique for reinforcing the accuracy of two-step-ahead (2SA) forecasts. The real-time recurrent learning (RTRL) algorithm for recurrent neural networks (RNNs) can effectively model the dynamics of complex processes and has been used successfully in one-step-ahead forecasts for various time series. A reinforced RTRL algorithm for 2SA forecasts using RNNs is proposed in this paper, and its performance is investigated by two famous benchmark time series and a streamflow during flood events in Taiwan. Results demonstrate that the proposed reinforced 2SA RTRL algorithm for RNNs can adequately forecast the benchmark (theoretical) time series, significantly improve the accuracy of flood forecasts, and effectively reduce time-lag effects.

Index Terms— Real-time recurrent learning (RTRL) algorithm, recurrent neural network (RNN), streamflow forecast, time series forecast.

I. INTRODUCTION

M OST observational disciplines tend to infer properties of an uncertain system from the analysis of its measured data. The analytical technologies for extracting the meaningful characteristics of time series data have been widely discussed for a long time [1]. Many mature techniques associated with time series analysis were used in many important applications such as environment and marketing [2]. Because observations closer together in time generally would be more closely related than observations further apart, it is more difficult to obtain a satisfactory multistep-ahead forecast. The recursive use of one-step-ahead forecasts for many time steps into the future is a commonly used strategy, which unfortunately has been shown to have shortcomings in real-world applications [3]. This is mainly because a small forecast error at the beginning could propagate into the future. To solve such a problem, it is argued whether an iterative adjustment of

the model's parameters based on additional information, such as antecedent observed values and/or model outputs, would be beneficial to multistep-ahead forecasts.

Over the last few decades, artificial neural networks (ANNs) have been recognized for modeling the underlying nonlinearities and complexities in artificial or physical systems. Many ANNs were developed to solve different problems, such as rainfall and streamflow forecasting [4]–[9], seismic [10], reservoir flood control [11], financial forecasts [12], sunspot activity [13], and many other disciplines for multistep-ahead forecasts [13]–[17]. Most of these applied with neural networks are classified into static neural networks and can simulate the short-term memory structures within processes, whereas the extraordinary time variation characteristics of time series might not be well retained.

Lately, recurrent neural networks (RNNs) have attracted much attention [18]–[24] for extracting dynamic time variation characteristics. Because of their dynamic nature, RNNs have been successfully applied to a wide variety of problems such as system identification [25]–[27], speech processing and plant control [28], and time series forecasting [29]–[33]. RNNs are capable of improving forecast accuracy [3], [34]–[36]. The training of an RNN, however, could be time consuming [13], [37], such as back-propagation through time (BPTT). BPTT, designed for training RNNs, can be derived by unfolding the temporal operation of the network into a multilayer feedforward network. The two familiar implementations of BPTT are the batch mode (epoch-wise BPTT) and real-time mode (truncated BPTT) [38]. A potential drawback of truncated BPTT is that the memory effects exceeding the truncation depth (duration) cannot be captured by RNNs.

The real-time recurrent learning (RTRL) algorithm, proposed by Williams and Zipser [39], is an effective and efficient algorithm for training recurrent networks, and its name is derived from the fact that real-time adjustments are made to the synaptic weights of an RNN. A number of previous studies demonstrated that the RTRL algorithm for RNNs is very effective in modeling the dynamics of complex processes and can provide accurate forecasts [40]–[42], while some studies further made efforts to reduce the time complexity of the RTRL algorithm [43]–[45].

Due to geophysical conditions, reservoirs in Taiwan are relatively small when considering the amount of water falling on watersheds during typhoon events. A controlled spillway is equipped with mechanical gates to control the water release

Reinforced recurrent neural networks for multi-step-ahead flood forecasts

Pin-An Chen [a], Li-Chiu Chang [b], Fi-John Chang [a,*]

[a] Department of Bioenvironmental Systems Engineering, National Taiwan University, Taiwan, ROC
[b] Department of Water Resources and Environmental Engineering, Tamkang University, Taiwan, ROC

ARTICLE INFO

SUMMARY

Considering true values cannot be available at every time step in an online learning algorithm for multi-step-ahead (MSA) forecasts, a MSA reinforced real-time recurrent learning algorithm for recurrent neural networks (R-RTRL NN) is proposed. The main merit of the proposed method is to repeatedly adjust model parameters with the current information including the latest observed values and model's outputs to enhance the reliability and the forecast accuracy of the proposed method. The sequential formulation of the R-RTRL NN is derived. To demonstrate its reliability and effectiveness, the proposed R-RTRL NN is implemented to make 2-, 4- and 6-step-ahead forecasts in a famous benchmark chaotic time series and a reservoir flood inflow series in North Taiwan. For comparison purpose, three comparative neural networks (two dynamic and one static neural networks) were performed. Numerical and experimental results indicate that the R-RTRL NN not only achieves superior performance to comparative networks but significantly improves the precision of MSA forecasts for both chaotic time series and reservoir inflow case during typhoon events with effective mitigation in the time-lag problem.

© 2013 Elsevier B.V. All rights reserved.

1. Introduction

Accurate multi-step-ahead (MSA) forecast is valuable and desired in many engineering problems, such as rainfall and flood forecasts, however it is a challenging task and difficult to achieve. A common approach to the MSA forecast is to update network parameters through online learning techniques. Online learning is a supervised machine-learning framework, which adopts the latest observed values to adjust model parameters for better mappings between instances and true values in a system. Because most observational disciplines tend to infer properties of an uncertain system from the analysis of time-dependent data, analytical technologies for extracting the meaningful characteristics of time series data have some inherent limitations, which has been a widely discussed issue for a long time (Brockwell and Davis, 1991; Jaeger and Haas, 2004; Jothiprakash and Magar, 2012; Nair et al., 2001). Online learning algorithms have several practical and theoretical advantages such as memory-efficient implementation, runtime-efficient implementation and strong guarantees on performance even in a highly variable data structure of time series (Shalev-Shwartz et al., 2004) owing to the continual receipt of true values for adjusting model parameters. Nevertheless, the main

defect of online learning is ascribed to the requirement for continual true values. Engineering problems frequently require models to predict many time-steps into the future without the availability of measurements in the horizon of interest. The lack of true values makes it difficult to achieve MSA forecasts. In addition, many studies indicated it is not an adequate strategy to recursively adopt single-step-ahead predictions for many time-steps into the future because the errors of MSA predictors will be accumulated based on the single-step-ahead predictor (Parlos et al., 2000; Yong et al., 2010). Such time-lag problems may cause significant performance degradation when dealing with MSA forecasts for real-world applications. For the MSA streamflow forecasts during typhoon events, models with time-lag problems (i.e. no updating latest observed values) cannot keep flow trails, especially in peak flows, as the forecasting step increases. To mitigate time-lag phenomena occurred in online learning algorithms, it is argued whether iterative adjustments of model parameters based on additional information, such as the latest true values and/or antecedent model outputs, would be beneficial to MSA forecasts.

Artificial neural networks (ANNs) have the ability to approximate nonlinear functions and therefore become valuable tools for various water resources problems (Cho et al., 2011; Nayak et al., 2005; Nikolos et al., 2008; Nourani et al., 2011; Nourani and Sayyah Frad, 2012). However, static neural networks might fail to establish reliable nonlinear models for predicting dynamical

* Corresponding author. Tel.: +886 2 23639461; fax: +886 2 23635854.
E-mail address: changfj@ntu.edu.tw (F.-J. Chang).
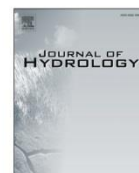
Regional estimation of groundwater arsenic concentrations through systematical dynamic-neural modeling

Fi-John Chang *, Pin-An Chen, Chen-Wuing Liu, Vivian Hsiu-Chuan Liao, Chung-Min Liao

Department of Bioenvironmental Systems Engineering, National Taiwan University, Taipei 10617, Taiwan, ROC

SUMMARY

Arsenic (As) is an odorless semi-metal that occurs naturally in rock and soil, and As contamination in groundwater resources has become a serious threat to human health. Thus, assessing the spatial and temporal variability of As concentration is highly desirable, particularly in heavily As-contaminated areas. However, various difficulties may be encountered in the regional estimation of As concentration such as cost-intensive field monitoring, scarcity of field data, identification of important factors affecting As, over-fitting or poor estimation accuracy. This study develops a novel systematical dynamic-neural modeling (SDM) for effectively estimating regional As-contaminated water quality by using easily-measured water quality variables. To tackle the difficulties commonly encountered in regional estimation, the SDM comprises of a neural network and four statistical techniques: the Nonlinear Autoregressive with eXogenous input (NARX) network, Gamma test, cross-validation, Bayesian regularization method and indicator kriging (IK). For practical application, this study investigated a heavily As-contaminated area in Taiwan. The backpropagation neural network (BPNN) is adopted for comparison purpose. The results demonstrate that the NARX network (Root mean square error (RMSE): $95.11\ \mu g\ l^{-1}$ for training; $106.13\ \mu g\ l^{-1}$ for validation) outperforms the BPNN (RMSE: $121.54\ \mu g\ l^{-1}$ for training; $143.37\ \mu g\ l^{-1}$ for validation). The constructed SDM can provide reliable estimation ($R^2 > 0.89$) of As concentration at ungauged sites based merely on three easily-measured water quality variables (Alk, Ca^{2+} and pH). In addition, risk maps under the threshold of the WHO drinking water standard ($10\ \mu g\ l^{-1}$) are derived by the IK to visually display the spatial and temporal variation of the As concentration in the whole study area at different time spans. The proposed SDM can be practically applied with satisfaction to the regional estimation in study areas of interest and the estimation of missing, hazardous or costly data to facilitate water resources management.

© 2013 Elsevier B.V. All rights reserved.

1. Introduction

Arsenic (As) contamination in groundwater has been reported and resulted in a massive epidemic of As toxication in several countries such as Bangladesh, Vietnam, Cambodia, China and Taiwan. It is estimated that approximately 57 million people drink As-contaminated groundwater with concentrations exceeding the drinking water standard recommended by the WHO (World Health Organization) (BGS-DPHE, 2001; Chakraborti et al., 2010). As pollution affects not only crop productivity and water quality but also the quality of water bodies, which threatens the health of animals and human beings by way of food chains. Long-term exposure to As in drinking water has been implicated in a variety of health concerns including cancers, cardiovascular diseases, diabetes and neurological effects (National Research Council, 1999). Blackfoot disease as well as cancers of the skin, bladder, lung and liver have been associated with drinking As-contaminated groundwater (Chiou et al., 1997; Rahman et al., 1999). As-contaminated groundwater is derived naturally from As-rich aquifer sediments, and the geochemistry of As can be rather complex (Stollenwerk, 2003). Various hydrogeological and biogeochemical factors affecting As concentration in groundwater have been detected, such as sediment mineralogy, microbial oxidation or reduction of As, groundwater recharge, groundwater flow paths (Ford et al., 2006; Wang et al., 2007, 2011; Xie et al., 2013), and the presence of fractures in bedrock formations (Ayotte et al., 2003; Liao et al., 2011). Even though the processes controlling the release of As into groundwater systems have been extensively discussed over the past decade, the exact chemical conditions and reactions leading to As mobilization still remain a subject of intense debate (Goovaerts et al., 2005; Polizzotto et al., 2006; Winkel et al., 2008). Moreover, the high variability of arsenic concentration can occur within a short distance and/or in different well depths due to the diversity in geology and geomorphology (Serre et al., 2003; Yu et al., 2003). Besides, the detection of As contamination in groundwater by using

* Corresponding author. Tel.: +886 2 23639461; fax: +886 2 23635854.
E-mail address: changfj@ntu.edu.tw (F.-J. Chang).

Real-time multi-step-ahead water level forecasting by recurrent neural networks for urban flood control

Fi-John Chang [a,*], Pin-An Chen [a], Ying-Ray Lu [a], Eric Huang [b], Kai-Yao Chang [b]

[a] Department of Bioenvironmental Systems Engineering, National Taiwan University, Taipei 10617, Taiwan, ROC
[b] Hydraulic Engineering Office, Public Works Department, Taipei City Government, Taipei 11008, Taiwan, ROC

SUMMARY

Urban flood control is a crucial task, which commonly faces fast rising peak flows resulting from urbanization. To mitigate future flood damages, it is imperative to construct an on-line accurate model to forecast inundation levels during flood periods. The Yu–Cheng Pumping Station located in Taipei City of Taiwan is selected as the study area. Firstly, historical hydrologic data are fully explored by statistical techniques to identify the time span of rainfall affecting the rise of the water level in the floodwater storage pond (FSP) at the pumping station. Secondly, effective factors (rainfall stations) that significantly affect the FSP water level are extracted by the Gamma test (GT). Thirdly, one static artificial neural network (ANN) (backpropagation neural network-BPNN) and two dynamic ANNs (Elman neural network-Elman NN; nonlinear autoregressive network with exogenous inputs-NARX network) are used to construct multi-step-ahead FSP water level forecast models through two scenarios, in which scenario I adopts rainfall and FSP water level data as model inputs while scenario II adopts only rainfall data as model inputs. The results demonstrate that the GT can efficiently identify the effective rainfall stations as important inputs to the three ANNs; the recurrent connections from the output layer (NARX network) impose more effects on the output than those of the hidden layer (Elman NN) do; and the NARX network performs the best in real-time forecasting. The NARX network produces coefficients of efficiency within 0.9–0.7 (scenario I) and 0.7–0.5 (scenario II) in the testing stages for 10–60-min-ahead forecasts accordingly. This study suggests that the proposed NARX models can be valuable and beneficial to the government authority for urban flood control.

© 2014 Elsevier B.V. All rights reserved.

1. Introduction

Urban flood control is a crucial and challenging task, particularly in developed cities. Urban floods are flashy in nature mainly due to severe thunderstorms and occur both on urbanized surfaces and in small urban creeks, which deliver mass water to cities. On account of more impervious areas resulting from the rapid urbanization in metropolitan areas, less water infiltration has resulted in an increase in the flow rate and the amount of surface runoff over the last decades. Taiwan is located in the northwestern Pacific Ocean where subtropical air currents frequently introduce typhoons and convective rains. The urban flood hydrographs in Taiwan typically have large peak flows and fast-rising limbs in a matter of minutes, which could cause serious disasters. For example, Typhoon Nari brought massive rainfalls at an astounding level of 500 mm/day on September 17th in 2001, which resulted in 27

deaths, inundations at some stations of the Taipei Metro System, and countless economic losses. The heavy rainfall event on June 12th in 2012 brought astonishing rainfalls with a cumulative amount of 54.1 mm/hr, which directly resulted in quick and wide surface flooding such that the transportation system collapsed in most of the southern Taipei City. It appears floods cannot be prevented, but planning emergency measures through flood management might mitigate disastrous consequences.

In response to the flood threat to residents and property, the Taipei City Government has long-term endeavored in developing flood control-related infrastructures, such as increasing levee heights and enhancing sewerage systems, and urban inundations have been significantly mitigated and controlled in recent years. As a result, the main threat to the city turns out to be the floodwater inside the levee system. A surface inundation will inevitably take place if surface runoff exceeds the capacity of a storm drainage system. To tackle this problem, pumping stations play an important role in flood mitigation at metropolitan areas and are principal hydraulic facilities built to manage internal stormwater

行政院國家科學委員會補助專題研究計畫 ■成果報告
□期中進度報告

創建新穎類神經網路於水文環境系統

計畫類別：■個別型計畫　　□整合型計畫

計畫編號：NSC 100-2313-B-002-011-MY3

執行期間：100 年 8 月 1 日至 103 年 7 月 31 日

執行機構及系所：國立臺灣大學生物環境系統工程學系暨研究所

計畫主持人：張斐章教授

共同主持人：

計畫參與人員：陳品安博士生、蔡文柄博士生、林政華碩士生

成果報告類型(依經費核定清單規定繳交)：□精簡報告　■完整報告

本計畫除繳交成果報告外，另須繳交以下出國心得報告：

□赴國外出差或研習心得報告

□赴大陸地區出差或研習心得報告

□出席國際學術會議心得報告

□國際合作研究計畫國外研究報告

處理方式：除列管計畫及下列情形者外，得立即公開查詢

　　　　　□涉及專利或其他智慧財產權，□一年□二年後可公開查詢

中　　華　　民　　國　　103 年 7 月 31 日

行政院國家科學委員會補助專題研究計畫 ■期中進度報告 □期末報告

永續水資源綜合管理應用於面臨工業污染及都市化之集水區 -- Lot River(法國)及淡水河(臺灣)之對照研究
Sustainable and integrated water resources management on watersheds facing industrial pollutions and urbanization: a comparative study between Lot River (France) and Danshuei River (Taiwan)

計畫類別：■個別型計畫　　□整合型計畫
計畫編號：NSC 101－2923－B－002－001－MY3
執行期間：101 年 1 月 1 日至 103 年 12 月 31 日

執行機構及系所：國立臺灣大學生物環境系統工程學系暨研究所

計畫主持人：張斐章教授
共同主持人：
計畫參與人員：童慶斌教授、林裕彬教授、任秀慧助理教授、黃誌川助理教授、
　　　　　　　高樹基博士、張麗秋副教授、張大偉博士、方文村博士、
　　　　　　　王國威博士、陳品安、蔡宇軒、王昱中、林恒玥

本計畫除繳交成果報告外，另含下列出國報告，共 ___ 份：
□移地研究心得報告
□出席國際學術會議心得報告
□國際合作研究計畫國外研究報告

處理方式：除列管計畫及下列情形者外，得立即公開查詢
　　　　　□涉及專利或其他智慧財產權，□一年■二年後可公開查詢
中　華　民　國　101 年 10 月 31 日

臺北市政府工務局
水利工程處
Hydraulic Engineering Office Public Works Department Taipei City

臺北市水災危險潛勢地區保全計畫
檢討工作

智慧型支流河川水位預測系統研究

採購名稱：臺北市水災危險潛勢地區保全計畫檢討工作

廠商名稱：　　環興科技股份有限公司

中華民國 101 年 7 月

抽水站排水系統水位預報及抽水機組智慧型操作策略評估工作

模式建置作業報告書 (定稿版)

國立臺灣大學生物環境系統工程學系
中華民國 103 年 3 月