Atmospheric Research 138 (2014) 166–178

Contents lists available at ScienceDirect

Atmospheric Research

journal homepage: www.elsevier.com/locate/atmos

Input selection and optimisation for monthly rainfall forecasting in Queensland, Australia, using artificial neural networks

John Abbot [1], Jennifer Marohasy *

School of Medical & Applied Sciences, Central Queensland University, Bruce Highway, Rockhampton, QLD 4702, Australia

ARTICLE INFO

Article history:
Received 20 August 2013
Received in revised form 21 October 2013
Accepted 1 November 2013

Keywords:
Rainfall
Climatology
ENSO
Statistical forecast
Seasonal forecast

ABSTRACT

There have been many theoretical studies of the nature of concurrent relationships between climate indices and rainfall for Queensland, but relatively few of these studies have rigorously tested the lagged relationships (the relationships important for forecasting), particularly within a forecast model. Through the use of artificial neural networks (ANNs) we evaluate the utility of climate indices in terms of their ability to forecast rainfall as a continuous variable. Results using ANNs highlight the value of the Inter-decadal Pacific Oscillation, an index never used in the official seasonal forecasts for Queensland that, until recently, were based on statistical models.

Forecasts using the ANN for sites in 3 geographically distinct regions within Queensland are shown to be superior, with lower Root Mean Square Errors (RMSE), Mean Absolute Error (MAE) and Correlation Coefficients (r) compared to forecasts from the Predictive Ocean Atmosphere Model for Australia (POAMA), which is the General Circulation Model currently used to produce the official seasonal rainfall forecasts.

1. Introduction

Until recently, the two government-based seasonal rainfall-forecasting programs for Queensland, Australia, used statistical models to provide the public with official forecasts (Fawcett and Stone, 2010). The first, produced by the Australian Bureau of Meteorology (BOM), commenced in 1989. The second, produced by the Queensland Government (QG) through the Department of Environment and Resource Management, commenced in 1994. Both programs issued seasonal (three-month) rainfall forecasts, in the format of the probability of exceeding the climatological seasonal median rainfall, and both were based on classification systems using climate indices representing broad scale atmospheric and oceanic circulation patterns.

In June 2013, the BOM moved to a new system based exclusively on The Predictive Ocean Atmosphere Model for Australia, POAMA, which is considered by the BOM to be a state-of-the-art seasonal to inter-annual seasonal forecast system based on a coupled ocean/atmosphere model and ocean/atmosphere/land observation assimilation systems (Australian Bureau of Meteorology, 2013). While POAMA generates quantitative seasonal and monthly rainfall predictions, the BOM chooses to continue to present its official forecasts simply as the probability of exceeding the long-term average value. A major limitation of such forecasts is that they provide no information about the magnitude of the expected deviation from the median rainfall value within the defined forecast period. For many practical purposes, such as management of water infrastructure or scheduling mine operations, the distribution of rainfall within the three-month period is more important than an averaged seasonal value (Sharma et al.,

* Corresponding author. Tel.: +61 7 4930 9622, +61 41 88 73 222 (Mobile); fax: +61 7 4930 9255.

E-mail addresses: j.abbot@cqu.edu.au (J. Abbot), jennifermarohasy@gmail.com (J. Marohasy).

[1] Tel.: +61 7 4930 9622, +61 41 88 73 222 (Mobile); fax: +61 7 4930 9255.

0169-8095/$ – see front matter
http://dx.doi.org/10.1016/j.atmosres.2013.11.002

2012). Furthermore, these programs failed to adequately forecast the exceptional wet summer of 2010 to 2011, with significant economic and social consequences (van den Honert and McAneney, 2011).

Researchers at the BOM acknowledge that there is presently a gap in rainfall prediction capability beyond 1 week and shorter than a season (Hudson et al., 2011). After about the first week the forecast system has typically lost most of the information from the atmospheric initial conditions, which are the basis for weather forecasts, while in the first month the ocean state probably has not changed much since the start of the forecast: hence it is difficult to beat persistence as a forecast (Vitart, 2004). Fawcett and Stone (2010) have reviewed the two major governmental statistical models used to forecast seasonal rainfall in Queensland, concluding that the skill level demonstrated for seasonal rainfall by both government agencies was "only moderate, although better than climatological and randomly guessed forecasts".

While official BOM seasonal forecasts are now based on POAMA, there is no evidence to suggest that this General Circulation Model (GCM) produces a more skilful forecast than the statistical system historically used. In fact results so far for POAMA have been disappointing (Hendon et al., 2012), and a review of 27 GCMs producing rainfall simulations for Queensland under the Intergovernmental Panel on Climate Change's Coupled Model Intercomparison Project Phase 5 produce widely divergent forecasts (Irving et al., 2012).

Artificial Neural Networks (ANNs) have been investigated for rainfall forecasting in many parts of the world including Greece (e.g. Nastos et al., 2013), China (e.g. Wu et al., 2001) and India (e.g. Venkatesan et al., 1997; Guhathakurta et al., 1999; Philip and Joseph, 2003; Chattopadhyay and Chattopadhyay, 2008; Shukla et al., 2011), but have been rarely applied in Australia (Mekanik et al., 2013), and are not used to generate the official forecasts (Abbot and Marohasy, 2012). The present study applies an ANN model with the objective of generating a more practical and skillful medium-term rainfall forecast for localities in Queensland than both the historical forecasts based on simple statistical models or the new method using the GCM, POAMA.

2. Theory

2.1. Artificial neural networks and rainfall forecasting

ANNs are powerful and versatile data-modelling tools that are able to capture and represent complex input and output relationships (ASCE Task Committee on Application of Artificial Neural Networks in Hydrology, 2000; Iseri et al., 2005). ANNs acquire knowledge through learning from multiple exemplars, storing that knowledge within inter-neuron connection strengths known as synaptic weights (Chakraverty and Gupta, 2008). A major advantage of neural networks lies in their ability to represent both linear and non-linear relationships and in their ability to learn these relationships directly from the data being modelled. Traditional linear models are simply inadequate for modelling data that contains non-linear characteristics.

The simplest types of neural network are based on multilayer perceptrons (MLPs), creating static models, where the input–output map depends only on the present input. However, if we want to process temporal data, each time sample has to be fed to a different input, requiring very large networks. With temporal problems, such as rainfall forecasting, the previous value of the input can potentially influence the current output. Jordan and Elman networks (Elman, 1990; Ding et al., 2013) can solve temporal problems by processing information over time using recurrent connections by incorporating context units. The context unit remembers the past of its inputs using a factor the unit forgets the past with an exponential decay. This means that events that just happened are stronger than the ones that have occurred further in the past. Other neural models, such as Time Lagged Recurrent Networks can also solve temporal problems by using memory units and are a type of dynamic neural network (ASCE Task Committee on Application of Artificial Neural Networks in Hydrology, 2000).

It is impossible to know in advance with any certainty which type of network architecture will give the best results for a particular problem. In the present investigation, through a process of trial and error it was found that Jordan and Elman networks generally gave superior results compared to other configurations tested. The context unit controls the 'forgetting factor' through a time constant set between 0 and 1. On the other extreme, a value of zero means that only the present time is factored in (i.e. there is no self-recurrent connection). The closer the value is to 1, the longer the memory depth and the slower the 'forgetting' factor. Based on experimentation, a Jordan network with one hidden layer and a time constant of 0.8 was selected for rainfall forecasts described in this study.

2.2. Input Variables, including Climate Indices

ANNs, like other statistical models, require a set of input predictor data. There are an infinite number of input variables potentially relevant to medium-term rainfall forecasting.

In practice however, we are limited to inputs that have been measured and, in particular, high-quality numerical values that extend back in time for a period long enough to enable pattern detection. In this study we considered relationships between lagged values for temperature, atmospheric pressure and rainfall as well as climate indices.

Climate indices describe recurrent patterns in sea surface temperatures (SST) and air pressures. The dominant concurrent phenomenon affecting Queensland rainfall is the El Niño–Southern Oscillation (ENSO) spanning the Pacific Ocean (Risbey et al., 2009). The Southern Oscillation Index (SOI) is a quantitative estimate of ENSO, defined as the normalised atmospheric pressure difference between Tahiti and Darwin. The simple statistical QG seasonal rainfall forecast model is based on the SOI, in particular a statistical technique of 'stratified climatology', employing five phases, or categories, derived from pairs of consecutive monthly values of the SOI (Stone and Auliciems, 1992; Stone et al., 1996).

Because relationships between rainfall and the SOI are weak throughout the year over the western third of Australia and during the late southern summer and autumn period over eastern Australia, the BOM historically chose to use climate indices based on sea surface temperature (SST) anomalies that span the Pacific and also Indian Oceans, designated as SST1 and SST2 (Drosdowsky and Chambers, 2001). The dominant mode of SST variability over the Indian–Pacific Ocean, SST1, strongly correlates with the SOI. Its skill as a predictor of seasonal rainfall with 1-month lag is comparable to that of the SOI with