Neural network modelling for environmental prediction

F. Murtagh\textsuperscript{a,*}, G. Zheng\textsuperscript{b}, J.G. Campbell\textsuperscript{b}, A. Aussem\textsuperscript{c}

\textsuperscript{a}School of Computer Science, The Queen's University of Belfast, Belfast BT7 1NN, Northern Ireland
\textsuperscript{b}Faculty of Informatics, University of Ulster, Magee College, Londonderry BT48 7JL, Northern Ireland
\textsuperscript{c}LIMOS, ISIMA, Université Blaise Pascal Clermont Ferrand II, Campus des Cézeaux - B.P. 125, 63173 Aubière Cedex, France

Received 30 September 1998; revised 14 February 1999; accepted 10 March 1999

Abstract

We describe the choice and assessment of neural network and statistical methods for data modelling, feature selection and forecasting. We deal in particular with how empirical environmental and Earth observation data can be used in conjunction with physical simulation models. © 2000 Elsevier Science B.V. All rights reserved.

1. Introduction

One theme in the environment and climate Neurosat (“Processing of Environmental Observing Satellite Data with Neural Networks”) project relates to prediction of oceanic upwelling off the Mauretanian coast, using sea surface temperature (SST) images, and real and model meteorological data for the year 1982. Upwelling is the periodic replenishment of coastal surface waters with cold deep water, which has various attendant effects.

The data available to us mainly consist of the following.

(i) Daily satellite SST data for 1982 using a geographic window covered by $82 \times 70$ 0.1’ (10 km) pixels off the northwest African coast. These are “sunshine data” (i.e., only available if the sky is not cloudy) at about 2 p.m. These data were from the advanced very high-resolution radiometer (AVHRR) on board satellites of the National Oceanographic and Atmospheric Administration (NOAA) series.

*Corresponding author. Tel.: + 44-1232-274620; fax: + 44-1232-683890.
E-mail address: f.murtagh@qub.ac.uk (F. Murtagh)

0925-2312/00/$ - see front matter © 2000 Elsevier Science B.V. All rights reserved.
PII: S 0 9 2 5 - 2 3 1 2 ( 9 9 ) 0 0 1 4 4 - 7
(ii) Daily wind and surface heat flux data, from the European Centre for Medium-Range Weather Forecasting, interpolated on a regular latitude-longitude grid with a resolution of 0.1°: zonal wind stress (N/m²), meridional wind stress (N/m²), total surface flux (latent + detectable + net longwave flux, W/m²), and short-wave flux (W/m²).

(iii) Data on the topography of the region.

(iv) Output from the Ispramix Ocean General Circulation Model (OGCM, see [1]), but currently with limited assimilation of empirical data.

The overall themes of our work are data and information fusion, empirical and model-based; handling data which is characterized by many uncertainties and numerous missing cases; and the development of “data-driven” pattern recognition and neural network methods. We are seeking an answer to the question as to whether such methods are an alternative to, or are complementary to, large physical simulation and modeling systems.

Data exploration and selection was discussed in [2,5]. Some aspects of this initial phase of the work were as follows. All empirical (observed) SST data were used, but we immediately faced the problem of missing values on a massive scale (over 70% of pixel values were missing, due directly or indirectly to cloud cover). Meteorological and environmental dynamics are mostly local, which led to the decision to cater for such local behaviour by nonlinear and locally piecewise methods. Hence we used a clusterwise regression method for imputation and nowcasting. Latency period of wind forcing of 10 or 11 d was assumed, which established the dimensionality of the problem. An approach to imputation of missing SST values was developed using clusterwise regression on wind/radiation, and spatial and temporal interpolation.

2. Learning from the Ispramix oceanographic model

Independently of the data imputation, another investigation was started which was based on ocean (physical) model output. The Ispramix model can provide complete output data, but at the cost of extensive computation time, and discrepancy from observed data. The latter difficulty can be mitigated by assimilation into the model of empirical data. The work reported below does not use assimilation, so that model output data approximates the reality. Use of assimilation is planned for the future.

If a neural network forecaster can be trained on model data, then we have proof of concept of the neural network capturing the dynamics of the environmental (ocean) phenomenon. This is our principal aim in this work. We also aim to “project” observed data into our trained neural network in the future, a task which will require validation and assessment.

In the work which will now be reported on, we took a small number of carefully selected datasets. We carried out a comprehensive evaluation schedule for (1) feature selection, and (2) one-step ahead forecasting. We found, and report on below, that model $\Delta T_i = T_i - T_{i-1}$ one-step ahead forecasts are to within $0.19^\circ$ or better, with
a range of neural net and related methods. (SST at time-point \(i\) is denoted \(T_i\)) The overall interval of \(\Delta T_s\) is approximately [0.5°, −0.6°].

Feature selection was carried out as follows:

- 15 different sets of experiments (input parameters, outputs) were defined.
- Two multilayer perceptron (MLP) architectures were used, denoted as: \(n - 3 - 1\) and \(n - 10 - 1\). The “embedding dimension”, \(n\), and the number of hidden layer units were arrived at heuristically, following exploratory experimentation.
- Four geographic locations were used. The pixel coordinates were (62, 19), (65, 23), (67, 16) and (71, 44).
- Four different times-of-day were available, and used.
- In all, 15 \(2 \times 4 \times 4 = 480\) training experiments were carried out. 50,000 iterations used in each case.

2.1. Feature selection

The 15 experiments carried out are summarized as follows. First we list variable names used below.

- \(\tau^{\text{WE}}, \tau^{\text{NS}}\): original W-E and N-S wind stresses.
- \(\tau^{\text{x}}, \tau^{\text{y}}\): orthogonal to shore, along-shore wind stresses.
- \(T\): model SST.
- \(q = q_{\text{sol}} + q_{\text{surf}}\) where these terms are short-wave flux and total surface flux, respectively.

\[
\begin{align*}
(i) & \quad \tau^{\text{WE}}_{t-5}, \ldots, \tau^{\text{WE}}_{t-1}, \Delta T_t \rightarrow \Delta T_{t+1} \\
(ii) & \quad \tau^{\text{NS}}_{t-5}, \ldots, \tau^{\text{NS}}_{t-1}, \Delta T_t \rightarrow \Delta T_{t+1} \\
(iii) & \quad \tau_{t-5}^{\text{WE}} 3/2, \ldots, \tau_{t-1}^{\text{WE}} 3/2, \Delta T_t \rightarrow \Delta T_{t+1} \\
(iv) & \quad \tau_{t-5}^{\text{NS}} 3/2, \ldots, \tau_{t-1}^{\text{NS}} 3/2, \Delta T_t \rightarrow \Delta T_{t+1} \\
(v) & \quad \tau^{\text{x}}_{t-5}, \ldots, \tau^{\text{x}}_{t-1}, \Delta T_t \rightarrow \Delta T_{t+1} \\
(vi) & \quad \tau^{\text{y}}_{t-5}, \ldots, \tau^{\text{y}}_{t-1}, \Delta T_t \rightarrow \Delta T_{t+1} \\
(vii) & \quad \tau_{t-5}^{\text{WE}} 3/2, \ldots, \tau_{t-1}^{\text{WE}} 3/2, \Delta T_t \rightarrow \Delta T_{t+1} \\
(viii) & \quad \tau_{t-5}^{\text{NS}} 3/2, \ldots, \tau_{t-1}^{\text{NS}} 3/2, \Delta T_t \rightarrow \Delta T_{t+1} \\
(ix) & \quad \tau^{\text{x}}_{t-2} 3/2 + \tau^{\text{x}}_{t-1} 3/2, \Delta T_t \rightarrow \Delta T_{t+1} \\
(x) & \quad \tau^{\text{y}}_{t-2} 3/2 + \tau^{\text{y}}_{t-1} 3/2, \Delta T_t \rightarrow \Delta T_{t+1} \\
(xi) & \quad \tau_{t-5}^{\text{WE}} 3/2 + \ldots + \tau_{t-1}^{\text{WE}} 3/2, \Delta T_t \rightarrow \Delta T_{t+1} \\
(xii) & \quad \tau_{t-5}^{\text{NS}} 3/2 + \ldots + \tau_{t-1}^{\text{NS}} 3/2, \Delta T_t \rightarrow \Delta T_{t+1} \\
(xiii) & \quad q_{t-5}, \ldots, q_{t-1}, \Delta T_t \rightarrow \Delta T_{t+1} \\
(xiv) & \quad \Delta q_{t-5}, \ldots, \Delta q_{t-1}, \Delta T_t \rightarrow \Delta T_{t+1} \\
(xv) & \quad T_{t-2}, T_{t-1}, T_t \rightarrow T_{t+1}
\end{align*}
\]

The rationale for some of these input datasets is as follows. The wind stresses were formative influences on changing SST. The along-shore wind stress was the more important, but the level of pixel resolution (i.e., relatively large pixels adjacent to the shoreline) made difficult the definition of the along-shore wind component. The power \(1/2\) was due to (i) the wind stresses being, dimensionally, second order to begin with, and
(ii) a power 3 expression yielding accumulated or integrated energy. The latency, e.g. 5 days here, was based on expert knowledge.

The feature selection experiments were carried out on all of the data, i.e. treating all data as the training set, since the best fit of neural network model and dataset was to be found. The 480 results were difficult to visually assess, which led to principal components analysis (PCA) being used simply to visualize this mass of data. Happily, a principal component opposing good fit to bad fit was found. We used this to select input dataset (iv) above as physically parsimonious, and proving one of the best results. Mapping (xv) above performed very significantly worse than all other mappings.

2.2. Forecasting method

Forecasting, based on the input dataset and mapping

\[ \tau_{t-5}, \ldots, \tau_t, \quad \Delta T_t \rightarrow \Delta T_{t+1} \]

made use of a training set of 299 values; and a test set 40 values (the latter were the first 20 d from months of March and September). Due to latency, a few values in early January were used indirectly only.

One-step ahead prediction at one location, with pixel coordinates (62,19), and one time of day, were studied in depth with many neural net prediction methods. The neural net and related methods used included the following.

- Weighted nearest neighbor regression.
- Generalized regression neural net (GRNN, [3,4]).
- Weighted nearest neighbor on principal components.
- GRNN on principal components.
- MLP – two architectures, 7–3–1 and 6–3–2.
- Kohonen self-organizing feature map (SOM) regression.

In some experiments, we mapped the \( \tau \) values onto two outputs, \( \Delta T_{t-1}, \Delta T_t \). The inherent dimensionality of the seven-dimensional data, i.e., \( \tau_s \) and \( \Delta T_{t-1} \), was assessed using PCA. The first and second principal components accounted for 52% and 16.7% of the variance. The weighted nearest-neighbour regression method can be described as follows: \( k \)-nearest neighbours using estimate \( \sum_{i \in N} w_i x_i / \sum_{i \in N} w_i \), with \( w_i = 1 / D(\cdot) \), \( D(\cdot) \) being the Euclidean distance, and \( N \) being defined by \( k \). The Kohonen SOM regression was as follows: construct SOM on training set; find winner nodes for test set; clusterwise (nearest mean) regression.

One-step ahead forecast results are summarized in Table 1 (note that equal RMSE results differed as expected when finer precision was taken into account) and one-step ahead MLP predictions for four locations are summarized in Table 2. We conclude that the root mean square error (RMSE) one-step-ahead prediction of \( \Delta T \) is 0.19° or better.
Table 1

<table>
<thead>
<tr>
<th>Method, input dataset</th>
<th>RMSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Distance weighted 1-NN, 7 dimensions</td>
<td>0.24</td>
</tr>
<tr>
<td>Distance weighted 3-NN, 7 dimensions</td>
<td>0.21</td>
</tr>
<tr>
<td>Distance weighted 5-NN, 7 dimensions</td>
<td>0.21</td>
</tr>
<tr>
<td>GRNN, Gaussian kernel, $\sigma = 0.1, 7$ dimensions</td>
<td>0.20</td>
</tr>
<tr>
<td>Distance weighted 1-NN, 1 principal component</td>
<td>0.27</td>
</tr>
<tr>
<td>Distance weighted 1-NN, 2 principal components</td>
<td>0.23</td>
</tr>
<tr>
<td>Distance weighted 1-NN, 3 principal components</td>
<td>0.25</td>
</tr>
<tr>
<td>GRNN, Gaussian kernel, $\sigma = 0.1, 1$ princ. comp.</td>
<td>0.20</td>
</tr>
<tr>
<td>GRNN, Gaussian kernel, $\sigma = 0.1, 2$ princ. comp.</td>
<td>0.20</td>
</tr>
<tr>
<td>GRNN, Gaussian kernel, $\sigma = 0.1, 3$ princ. comp.</td>
<td>0.20</td>
</tr>
<tr>
<td>SOM regression</td>
<td>0.21</td>
</tr>
<tr>
<td>MLP, 7-3-1, 5000 iterations</td>
<td>0.19</td>
</tr>
<tr>
<td>MLP, 6-3-2, 5000 iterations</td>
<td>0.19</td>
</tr>
</tbody>
</table>

Table 2

| Four pixel coordinate locations, MLP RMSE results, using 7–3–1 net, and 6–3–2 net |
|---------------------------------|------|------|
| (62,19)                         | 0.19 | 0.19 |
| (65,23)                         | 0.12 | 0.13 |
| (67,16)                         | 0.10 | 0.11 |
| (71,44)                         | 0.17 | 0.19 |

3. Conclusion

Further details on the work described here, including animations and background information, can be found at http://hawk.infm.ulst.ac.uk:1998/neurosat (or by emailing the authors, fmurtagh@qub.ac.uk).

We have explored both neural network modelling, and the coupling of such data processing with (i) empirically observed data, and (ii) model output of the physical dynamics. Our data-driven methodology therefore takes account of the physics of the domain studied.

We have noted two directions for future work. Firstly, we wish to redo our experiments with physical model data which uses assimilation of observed data. Secondly, we wish to use neural network forecast engines, trained on physical model data, on observed data in order to produce forecasts.

Acknowledgements

Particular acknowledgment is due to W. Eifler, M. Ouberdous and E. Demirov, Joint Research Centre, SAI – Marine Environment Unit Data Assimilation sector,
TP69, Ispra, Italy, for data, domain knowledge and many discussions. We also thank M. Crépon, S. Thiria and other consortium partners for feedback and comments received in discussions and presentations of this work.

References


