Geoderma 147 (2008) 1–7

Contents lists available at ScienceDirect

Geoderma

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

The wisdom of crowds — ensembles and modules in environmental modelling

L. Baker [a,*,1], D. Ellison [b]

[a] Dundee, UK
[b] University of Abertay Dundee, Bell Street, Dundee, DD1 1HG, Scotland, UK

ARTICLE INFO

Article history:
Received 21 October 2006
Received in revised form 15 May 2008
Accepted 14 July 2008
Available online 27 August 2008

Keywords:
Bias/variance trade-off
Ensemble
Module
Multi-model ensemble
Multi-model module

ABSTRACT

Predictive models that are composed of a number of combined models, although ubiquitous in climate prediction have not yet become popular in many other areas of environmental modelling, despite growing evidence that they are superior to single-model methods in many ways. These combined-model methodologies are termed ensembles and modules, and this paper reviews their concepts, advantages and how to create them. Additionally, they will be discussed in terms of the critically important bias/variance trade-off. Moreover, ensembles and modules will be discussed with reference to historical and current research papers within environmental modelling.

1. Introduction

> "If you ask a large enough group of diverse, independent people to make a prediction or estimate a probability, and then average those estimates, the errors that each of them makes in coming up with an answer will cancel themselves out. Each person's guess, you might say, has two components: information and error. Subtract the error, and you're left with the information."
>
> James Surowiecki, 2004
> The Wisdom of Crowds

As computing power expands, modellers are discovering ever more computer-intensive ways of yielding predictions, and environmental modellers are no different. No longer do they have to choose between different modelling methodologies, they can now combine models in ways that are far superior to single models — they are discovering that 'and' is better than 'or'. There are essentially two ways of combining models; ensembles and modules, where the former combines models that each predicts the same target, and the latter combines models predicting different targets.

Despite being used in climate forecasting for many years, model combinations have been mostly overlooked in many other areas of environmental modelling. Of those practitioners that have used ensembles and modules in their research, very few have provided explanations as to why these modelling methodologies produce superior results to their competing single model methods, and virtually none have provided analytical or anecdotal comparisons between single and combinatorial models. Moreover, the mathematical and theoretical bases for ensembles and modules are mainly absent in environmental publications. Consequently, there are few environmental modelling papers that can be cited with regards to ensemble and modular advantages, but we cite them here where found.

Although there are significant benefits to be gained by using these modelling methodologies, ensembles and modules have not yet become popular in environmental modelling, and it is not clear why these modalities have been little used. The aim of this paper is to introduce ensemble and modular methods, discuss their advantages, and show – both theoretically and with empirical examples – just why combinatorial models are superior to single modelled methods.

This paper introduces the concept of ensembles (Section 2), including multi-model ensembles, and explains why, mathematically, they are superior to single modelled methods. This entails how best to split a dataset in order to optimise the data selection process (create ensemble members), and then combine the resulting models (create ensembles) in order to exploit the critically important bias/variance trade-off that exists in all modelling methodologies. Empirical examples of ensemble improvements over single models are given where appropriate. Modular models (methods of creating and combining modular components) are also discussed (Section 3), including multi-model modules, and compared with the ensemble method. Methods of producing error estimates of individual predictions using the ensemble methodology is also discussed (Section 4).

Abbreviations: ANN, artificial neural network; ME, mean error; MSE, mean squared error; RMSE, root mean squared error.

* Corresponding author.
E-mail address: eelrekab@gmail.com (L. Baker).
[1] Private consultant.

0016-7061/$ – see front matter © 2008 Elsevier B.V. All rights reserved.
doi:10.1016/j.geoderma.2008.07.003

2. Ensembles

In general, different models perform differently, some providing more accurate predictions than others. It is therefore useful to develop a number of different models (perhaps by using different data subsets or utilising differing conditions within the modelling methodology of choice) to ensure that a good model is found (where the term 'good' is dependent upon the individual practitioner's requirements). However, selecting the 'best' model is not necessarily the ideal choice, because potentially valuable information may be wasted by discarding the results of less-successful models (Perrone and Cooper, 1993; Tumer and Ghosh, 1996). This leads to the concept of 'combining', where the outputs (individual predictions) of several models are pooled before a decision (collective prediction) is made (Tumer and Ghosh, 1996).

The word 'ensemble' is French, meaning 'together' or 'at the same time', and usually refers to a unit or group of complementary parts that contribute to a single effect. In predictive modelling an ensemble is a set of individual models, where the component models (also known as members) are redundant in that each provides a solution to the same task, even though this solution may be obtained by different means.

The main reasons for combining models in redundant ensembles are to improve the ability to provide accurate predictions and to guard against the failure of individual member models. Here, the term 'fail' refers to the fact that individual models will make predictions that will not usually be identical to the target function, and will usually under- or over-estimate the expected value(s).

Ensembles have a long history in the real world. The Condorcet Jury model proposed in 1786 that a democracy as a whole is more effective than any of its constituent members (Grofman and Owen, 1986), and in forecasting, it has been demonstrated that better results can be achieved by combining forecasts than by choosing the best one (Bates and Granger, 1969).

It has been shown mathematically (Perrone and Cooper, 1993) that the prediction error of an ensemble is related to the prediction error of its constituent members by:

$$MSE_{Ensemble} = \frac{1}{N}\overline{MSE} \quad (1)$$

where $\overline{MSE}$ is the average mean squared error, *MSE*, of the individual predictors and N is the number of members in the ensemble. Theoretically, this implies that by increasing the number of members in the population, the error of an ensemble's power of estimation can be made to be arbitrarily small when compared to the average error of the models when taken as individuals. In practice, however, as N becomes large only small improvements, if any, may be made to the predictive ability of the ensemble, mainly due to the correlations between data across the different members (further details may be found in Perrone and Cooper, 1993). Eq. (1)) is a very powerful result in the method of combining estimators in ensembles, and holds for any type of estimator, providing that an error estimate is given. To demonstrate this, we refer to the putative predictions and observed accuracies in Table 1 (the raison d'être and method of modelling are not important), which is used here to illustrate the utility of ensembles, and how improvements are made over individual models. The overall accuracies of each of the three individual models are equivalent, and yet by combining into two-membered ensembles, the accuracies – measured by the root mean squared error, *RMSE* – improve. Further, the error decreases to a greater extent for the three-membered ensemble (compared with two of the three two-membered ensembles — the differences between the three- and two-membered ensembles are discussed in greater detail later). The improvement seen is explained by the bias/variance trade-off.

Table 1
Predictions for three individual models

	Model predictions						
	Individual			Ensemble			
Expectation	A	B	C	A,B	A,C	B,C	A,B,C
0.401	0.428	0.369	0.375	0.399	0.402	0.372	0.391
0.412	0.420	0.394	0.400	0.407	0.410	0.397	0.405
0.422	0.434	0.412	0.405	0.423	0.420	0.409	0.417
0.345	0.350	0.315	0.335	0.333	0.343	0.325	0.333
0.313	0.339	0.301	0.285	0.320	0.312	0.293	0.308
0.453	0.489	0.432	0.420	0.461	0.455	0.426	0.447
0.477	0.467	0.458	0.482	0.463	0.475	0.470	0.469
0.532	0.527	0.528	0.538	0.528	0.533	0.533	0.531
0.341	0.343	0.339	0.335	0.341	0.339	0.337	0.339
0.379	0.388	0.385	0.365	0.387	0.377	0.375	0.379
RMSE	0.018	0.018	0.018	0.008	0.002	0.017	0.007
ME (bias)	0.011	−0.014	−0.014	−0.002	−0.001	−0.014	−0.006
Variance (x10^4)	1.93	1.33	1.55	0.55	0.02	0.94	0.14
RI (bias, %)	–	–	–	0.00	0.00	0.00	0.00
RI (variance, %)	–	–	–	66.04	98.79	34.79	91.37

Models *A* to *C* have equivalent accuracies (*RMSE*s), but different biases (*ME*s). Models *B* and *C* have equivalent bias, whilst model *A* has a similar bias to *B* and *C*, but in the opposite sense. The individual models are also combined into two- and three-membered ensembles. The accuracy, bias and variance are shown for all models. The final two rows are the relative improvement (*RI*) of the ensemble bias and variance over the average bias and variance of the constituent members.

The effect of Eq. (1) on ensemble predictions was noted by Jeong and Kim (2004) in their research on rainfall-runoff modelling, by Rallo et al. (2005) in chemical contamination research using ensemble artificial neural networks (ANNs), and investigated by Baker and Ellison (2008) in research on the water retention of European soils using an ensemble of ANNs. Baker and Ellison demonstrated that the improvement of two members (an ensemble with two ANNs) over one (the single ANN methodology) was 12.5%, whilst the improvement of 10 members over one was 17.6%. The conclusion of these authors was that, for the data investigated and the methods used, ensembles offer greatly improved results over methods using single models, even when the number of members is as few as two.

It should be noted that ensemble members need not be composed of the same type. Ensembles composed of constituent models of the same type may be considered as 'single-model ensembles' (but are usually referred to, simply, as ensembles), where those using members of different types are often referred to as 'multi-model ensembles'. Georgakakos et al. (2004) provided an investigation into the inter-comparisons of multi-model ensembles, single-model ensembles and single models in their research on streamflow simulations, and concluded that multi-model ensembles outperform both single models and single-model ensembles. Palmer et al. (2005) discussed the evolution of three different multi-model ensembles in their paper on climate prediction, whilst in the climate prediction paper of Krishnamurti et al. (1999), multi-model ensembles outperformed single-model ensembles. Indeed, these latter papers are just two multi-model ensemble papers of many in the field of climatology, where the use of multi-model ensembles is now commonplace (see also, for instance, Kharin and Zwiers, 2002; Hagedorn et al., 2005a,b and Thomson et al., 2006).

2.1. The bias/variance trade-off

The effect of combining models to reduce errors may be expressed in terms of the statistical terms bias and variance (for example, see Raviv and Intrator 1996, Sharkey 1999; Wang, 1998). If $\hat{\theta}$ is an estimator of the quantity θ, then the mean squared error (*MSE*) of an individual predictor (model) can be expressed as:

$$MSE = Variance\left(\hat{\theta}\right) + Bias^2\left(\hat{\theta}\right) \quad (2)$$

Fig. 1 is an illustration of Eq. (2), and shows that the model bias and model variance are decreasing and increasing functions, respectively,