Environmental Modelling & Software 80 (2016) 1–8

Contents lists available at ScienceDirect

Environmental Modelling & Software

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

Assessing spatial predictive models in the environmental sciences: Accuracy measures, data variation and variance explained

Jin Li*

Geoscience Australia, GPO Box 378, Canberra, ACT 2601 Australia

ARTICLE INFO

Article history:
Received 23 April 2015
Received in revised form
17 November 2015
Accepted 1 February 2016
Available online xxx

Keywords:
Predictive accuracy
Error measure
Data variance
Model assessment
Spatial interpolation methods
Spatial predictions

ABSTRACT

A comprehensive assessment of the performance of predictive models is necessary as they have been increasingly employed to generate spatial predictions for environmental management and conservation and their accuracy is crucial to evidence-informed decision making and policy. In this study, we clarified relevant issues associated with variance explained (*VEcv*) by predictive models, established the relationships between *VEcv* and commonly used accuracy measures and unified these measures under *VEcv* that is independent of unit/scale and data variation. We quantified the relationships between these measures and data variation and found about 65% compared models and over 45% recommended models for generating spatial predictions explained no more than 50% data variance. We classified the predictive models based on *VEcv*, which provides a tool to directly compare the accuracy of predictive models for data with different unit/scale and variation and establishes a cross-disciplinary context and benchmark for assessing predictive models in future studies.

1. Introduction

Spatial predictions are often required for and play a significant role in planning, risk assessment, and decision making in environmental management and conservation. With the advancement in computing technology and modelling techniques (Crawley, 2007; Hastie et al., 2009; Kuhn and Johnson, 2013) and development in remote sensing and geographic information systems, predictive models have been increasingly employed for making spatially continuous predictions in various disciplines in the environmental sciences (Dormann et al., 2007; Li et al., 2011a; Maier et al., 2014; Marmion et al., 2009). The accuracy of the predictions becomes important as it is crucial to evidence-informed decision making and policy. Hence it is necessary to have a comprehensive assessment of the performance of predictive models in the environmental sciences.

Various measures of predictive accuracy have been proposed for categorical data (e.g. presence and absence data) (Fielding, 2002; Liu et al., 2011) and for numerical data (e.g. continuous data, count data and proportion data) (Bennett et al., 2013; Han and Kamber, 2006; Krause et al., 2005; Li and Heap, 2008; Moriasi et al., 2007; Schloeder et al., 2001; Willmott et al., 2012). In this study, we focus on the accuracy measures for numerical data that is usually encountered in the environmental sciences. Many measures have been used for numerical data (Li and Heap, 2008), but their relationships are usually unknown. Definition of variance explained by predictive models or predictive accuracy measures sometimes varies with published studies. Different names and data have been used for variance explained or predictive accuracy measures, such as efficiency of a model or Nash–Sutcliffe efficiency (*NSE*) based on the difference of observed values and the model–derived computed values, simulated values or predicted values (Gupta et al., 2009; Krause et al., 2005; Nash and Sutcliffe, 1970; Willmott et al., 2015), model efficiency based on the difference of observed values and model–derived predicted values (Greenwood et al., 1985; Vicente-Serrano et al., 2003), and G-value based on the difference of observed values and predicted values derived from cross-validation (Schloeder et al., 2001). Some of these measures were often used as measure of goodness of fit instead of predictive accuracy in previous studies (Gupta et al., 2009; Krause et al., 2005). The key difference is that if the observations used to validate a predictive model were also used for developing the model, the resultant measures are for assessing the goodness of fit of the model; otherwise they are for assessing the predictive accuracy of the model. Consequently, these inconsistences in definitions and applications make the accuracy often incomparable among different studies.

Mean absolute error (*MAE*) and root mean squared error (*RMSE*)

* Corresponding author.
E-mail address: Jin.Li@ga.gov.au.

http://dx.doi.org/10.1016/j.envsoft.2016.02.004
1364-8152/

are the most commonly used measures to assess predictive accuracy in the environmental sciences (Li and Heap, 2008; Willmott, 1982), despite various advantages and disadvantages associated with them as being discussed in previous studies (Chai and Draxler, 2014; Willmott and Matsuura, 2005, 2006; Willmott et al., 2009). However their magnitude depends on the scale/unit of the variable predicted (Li and Heap, 2011), which make it impossible to compare the accuracy of predictive models for different disciplines and variables. Two unit/scale independent measures, relative *MAE* (*RMAE*) and relative *RMSE* (*RRMSE*), were then proposed (Li and Heap, 2011); and subsequently relationships of *RMAE* and *RRMSE* with data variation (i.e., coefficient of variation: *CV*) were preliminarily illustrated (Li and Heap, 2008; Li et al., 2011a, 2011b, 2012) and further depicted with individual methods and sampling designs (Li and Heap, 2008, 2011) based on information derived from cross validation in the environmental sciences, which provided the fundamental information for this study. The relationships were apparently linear, but not quantified. Moreover, the accuracy of predictive models in the environmental sciences has not been systematically assessed and the context and benchmark for assessing predictive models have not been developed, which prevent objective assessment of the performance of individual models.

The aim of this study is to systematically assess the accuracy of predictive models in the environmental sciences and provide a baseline and benchmark for scientists to evaluate the accuracy of their predictive models in the environmental sciences. The objectives of this study are to: 1) clarify relevant issues associated with variance explained by predictive models, 2) explore the relationships among various *MAE* and *RMSE* related predictive accuracy measures, and 3) quantify the relationships between the accuracy measures and *CV*.

2. Materials and methods

2.1. Source and description of dataset

The dataset for the predictive accuracy of 35 methods in relation to *CV* was from a published comprehensive review in the environmental sciences (Li and Heap, 2011). This dataset contained all 35 compared methods with 338 applications and is hereinafter referred to as 'all compared methods' (ACM). Of which, *RMAE* was derived for 210 applications, *RRMSE* for 296 applications, and *RMAE* and *RRMSE* jointly for 169 applications. There was neither *RMAE* nor *RRMSE* derived for one application in the dataset, so the sum of these numbers led to 337 instead of 338. We also added one extra column to ACM to indicate the most accurate method(s) for each variable predicted (Appendix A); and these methods were most likely recommended for generating spatial predictions. A sub-dataset of 83 predictive models was then resulted for these recommended methods and is hereinafter referred to as 'recommended methods' (RM). In RM, *RMAE* was available for 44 models, *RRMSE* for 70 models, and *RMAE* and *RRMSE* jointly available for 31 models.

2.2. Definition of variance explained

It is necessary to define and clarify relevant issues associated with variance explained by predictive models. In this study, it is defined as the proportion of variation in the validation data explained by the predicted values obtained from predictive models based on cross-validation; hence it is denoted as *VEcv*. Due to such predicted values, we don't consider the number of covariates used by the predictive models, so *VEcv* is different from the proportion of variation explained by covariates as for the regression models. A linear relationship with a slope of 1 is expected between the predicted and observed values when perfect predictions are generated, hence *VEcv* is calculated as:

$$VEcv = \left(1 - \frac{SSD}{SST}\right)100\ (\%) = \left(1 - \frac{\sum_1^n \left(y_i - \hat{y}_i\right)^2}{\sum_1^n (y_i - \bar{y})^2}\right)100\ (\%) \quad (1)$$

where n is the number of observations in a validation dataset, $\boldsymbol{y_i}$ is the observed value in the validation data, $\hat{\boldsymbol{y}}_{\boldsymbol{i}}$ is the predicted value, $\bar{\boldsymbol{y}}$ is mean of the observed values, *SSD* is sum of square departures (i.e. $\sum_1^{\boldsymbol{n}}(\boldsymbol{y_i} - \hat{\boldsymbol{y}}_{\boldsymbol{i}})^2$); and *SST* is total sum of squares (i.e. $\sum_1^{\boldsymbol{n}}(\boldsymbol{y_i} - \bar{\boldsymbol{y}})^2$). *SSD*/*SST* is also termed relative square error (*RSE*) (Han and Kamber, 2006).

2.3. Data analysis

Linear regression (lm), robust regression with M-estimation (rlm) and resistant regression (lqs) in the R package MASS (Venables and Ripley, 2002) were used to quantify the relationship of *RMAE* and *RRMSE* with *CV*. The latter two methods, rlm and lqs, were used as supplementary methods to lm to handle outliers in the data. Since *CV* data could not be derived for some applications, sample numbers for the regression models were usually smaller than above numbers for *RMAE* and RRMES. Linear regression was also used to quantify the relationship between *RMAE* and *RRMSE*. Given that we expected constant predictions when data variation is zero, the predictive error should be zero so we fitted these regressions with an intercept of zero. So did for the regression of *RRMSE* and *RMAE*. All these modelling and relevant plotting work were implemented in R 3.1.0 (R Development Core Team, 2012).

3. Results and discussion

3.1. Relationships between predictive accuracy measures and variance explained

3.1.1. VE_{CV} and other relevant accuracy measures

The definition of *VEcv* emphasises that it is an accuracy measure based on the predicted values that were derived from cross-validation. This *VEcv* is also called G-value (or goodness-of-prediction measure) (Schloeder et al., 2001). It is equivalent to model efficiency (Greenwood et al., 1985; Li and Heap, 2011; Vicente-Serrano et al., 2003) if the model–derived predicted values used in model efficiency were based on cross validation. It is also equivalent to efficiency of a model or Nash–Sutcliffe efficiency (Gupta et al., 2009; Krause et al., 2005; Nash and Sutcliffe, 1970; Willmott et al., 2015) if their model–derived computed values, simulated values or predicted values were based on cross validation, although they were often used for measuring the goodness of fit.

This *VEcv* is sometimes mistakenly referred to as R^2, the variance explained by regression models. However, they are fundamentally different concepts. This is because: 1) *VEcv* uses the predicted values as defined above, while R^2 uses fitted values (Draper and Smith, 1981) and measures correlation instead of accuracy (Kuhn and Johnson, 2013); 2) the predicted values are generated from cross-validation, while fitted values are not; and 3) R^2 is always positive, ranging from 0 to 1, whereas *VEcv* can be negative and with a maximum of 100%. For *VEcv*, a value of 100% suggests perfect predictions, a 0% indicates that the predictions are as accurate as using the mean of validation data as predictions, while negative values indicate that the predictions are less accurate than using the mean as predictions (Gotway et al., 1996; Schloeder et al., 2001).

3.1.2. Relationships between VE_{CV} and RMSE related measures

Among the accuracy measures, only *RMSE* and its related