ISPRS Journal of Photogrammetry and Remote Sensing 134 (2017) 110–121

Contents lists available at ScienceDirect

ISPRS Journal of Photogrammetry and Remote Sensing

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

Per-pixel bias-variance decomposition of continuous errors in data-driven geospatial modeling: A case study in environmental remote sensing

Jing Gao [a,b,*], James E. Burt [b]

[a] Climate and Global Dynamics Laboratory, National Center for Atmospheric Research, P.O. Box 3000, Boulder, CO 80307, USA
[b] Department of Geography, University of Wisconsin – Madison, 550 North Park Street, Madison, WI 53706, USA

ARTICLE INFO

Article history:
Received 26 April 2017
Received in revised form 31 October 2017
Accepted 1 November 2017

Keywords:
Model evaluation
Accuracy assessment
Bias-variance decomposition
Absolute error
Squared error

ABSTRACT

This study investigates the usefulness of a per-pixel bias-variance error decomposition (BVD) for understanding and improving spatially-explicit data-driven models of continuous variables in environmental remote sensing (ERS). BVD is a model evaluation method originated from machine learning and have not been examined for ERS applications. Demonstrated with a showcase regression tree model mapping land imperviousness (0–100%) using Landsat images, our results showed that BVD can reveal sources of estimation errors, map how these sources vary across space, reveal the effects of various model characteristics on estimation accuracy, and enable in-depth comparison of different error metrics. Specifically, BVD bias maps can help analysts identify and delineate model spatial non-stationarity; BVD variance maps can indicate potential effects of ensemble methods (e.g. bagging), and inform efficient training sample allocation – training samples should capture the full complexity of the modeled process, and more samples should be allocated to regions with more complex underlying processes rather than regions covering larger areas. Through examining the relationships between model characteristics and their effects on estimation accuracy revealed by BVD for both absolute and squared errors (i.e. error is the absolute or the squared value of the difference between observation and estimate), we found that the two error metrics embody different diagnostic emphases, can lead to different conclusions about the same model, and may suggest different solutions for performance improvement. We emphasize BVD's strength in revealing the connection between model characteristics and estimation accuracy, as understanding this relationship empowers analysts to effectively steer performance through model adjustments.

1. Introduction

Researchers and practitioners agree that careful assessment of analysis and modeling accuracy is necessary for using remote sensing products in policy and decision making (e.g. Campbell and Wynne, 2011; Congalton and Green, 2009). Effective model evaluations should go beyond describing the model, and provide insights for understanding and improving model performance. Environmental remote sensing (ERS) research in these directions has long focused on classification accuracy (e.g. Ma et al., 2017; Comber et al., 2012; Strahler et al., 2006; Congalton, 1991; Campbell, 1981), while remotely sensed data are also frequently used to estimate important continuous variables for environmental and ecological studies, such as land imperviousness (e.g. Xian et al., 2011), foliar nitrogen concentration (e.g. Eitel et al., 2014), land surface temperature (e.g. Weng and Fu, 2014) and soil moisture (e.g. Rodríguez-Fernández et al., 2016); sometimes although the response variable of interests is categorical, analysts may choose to use a model that estimates the likelihood of the occurrence of a category. In both cases, continuous instead of categorical errors are under investigation. Analyzing continuous errors is very different from analyzing categorical errors, due to their intrinsically different mathematical features. Current standard continuous model evaluations considerably rely on overall accuracy measures. Among the most commonly used are mean error, root mean squared error, and mean absolute error, with the average taken over pixels in a model evaluation dataset. Although such summary measures provide useful descriptions of overall model

* Corresponding author at: Climate and Global Dynamics Laboratory, National Center for Atmospheric Research, P.O. Box 3000, Boulder, CO 80307, USA.
E-mail addresses: jingg@ucar.edu (J. Gao), jeburt@wisc.edu (J.E. Burt).

https://doi.org/10.1016/j.isprsjprs.2017.11.001
0924-2716/

performance, they offer limited assistance for achieving important objectives of model evaluation – understanding and improving model performance. Especially, they do not inform spatially-explicit understandings about model performance, which has been proven beneficial by many authors (e.g. Khatami et al., 2017; Löw et al., 2015; McGwire and Fisher, 2001).

In ERS, data-driven geospatial models are widely used, including such supervised image processing algorithms as least-squares regressions, regression trees, for modeling continuous variables, and support vector machines, random forests, for modeling categorical variables. Many of these models originated from the fields of statistics and machine learning, where bias-variance error decomposition (BVD) is a proven-effective and commonly-used model evaluation method (e.g. Gao et al., submitted for publication; Domingos, 2000; Geman et al., 1992). BVD attributes expected modeling error to three sources: "bias", "variance", and "noise". These sources are independent from each other and require different methods to reduce. Hence, BVD can help analysts understand the effects of different model characteristics on estimation accuracy, and therefore develop effective plans altering model characteristics to achieve desired performance. Existing aspatial studies (e.g. Gao et al., submitted for publication; Schapire et al., 1998; Breiman, 1996) have demonstrated BVD's benefits in comparing and explaining the performances of data-driven models, addressing questions like why one model outperforms another, and why certain modeling practices work for some problems but not for others.

Additionally, BVD analyzes model performance at individual data points – if applied in ERS image processing, BVD can generate per-pixel model evaluations. This is different from how model "bias" and "variance" have conventionally been discussed in geospatial modeling, where spatial stationarity is often assumed (i.e. assuming the underlying process is uniform within a study area), and model bias and variance are summary descriptions over multiple model evaluation data points sampled across the study area. In contrast, BVD acknowledges that both underlying processes and model performances can vary over space, and the per-pixel analysis can reveal how modeling errors at different parts of the study area may be driven by different sources. Since different methods are needed to reduce different error sources, analysts can then effectively apply region-specific model-improvement strategies.

These traits of BVD indicate its potential for addressing the needs of ERS model evaluation, while limited work exists exploring the benefits of using BVD in spatially-explicit data-driven modeling. Gao et al. (2016) examined categorical-error BVD with data-driven geospatial classification models of a binary land cover change variable (new built-up development or not) using environmental variables, and found that BVD can help test the validity and mitigate the undesired impacts of spatial non-stationarity assumptions and can inform efficient training sample acquisition. However, continuous- and categorical-error BVDs differ substantially at both analytical and practical levels. Categorical-error-based insights do not necessarily apply to continuous errors, and continuous-error BVD has not been examined for geospatial modeling generally nor for ERS applications more specifically.

This study uses a regression tree model mapping land imperviousness (a continuous variable ranging 0–100%) using Landsat data as a showcase to examine the usefulness of continuous-error BVD for understanding and improving data-driven geospatial models. We investigate ways to interpret the spatial patterns of BVD error components, and explore avenues to use BVD insights for designing model improvement efforts, with a focus on the BVD insights that are not accessible through common model evaluation methods. The study area (Washtenaw County, MI, USA) is the same as Gao et al. (2016), allowing readers to synthesize continuous- and categorical-error BVDs through direct comparison.

2. Methods

2.1. Definitions of continuous errors

Predictive models estimate the response variable Y using the predictor vector $\boldsymbol{X}$. In this study, Y is land imperviousness ranging 0–100%, and $\boldsymbol{X}$ is a vector of reflectance values of various Landsat bands. A data-driven (a.k.a. supervised, empirical, or statistical) model, f, is constructed by a learning algorithm using a collection of observed $(\boldsymbol{X}, Y)$ pairs – a training set – with the goal to approximate the response Y with the model estimate $\hat{Y} = f(\boldsymbol{X})$. In this paper, we use upper-cased symbols for the names of variables, and lower-cased symbols for individual values. In the text, X, Y and x, y can sometimes be interpreted synonymously, but in mathematical equations, the distinction provides important information. For example, writing the statistical expectation E of a model f as $E(f(\boldsymbol{x}, Y))$ emphasizes that $\boldsymbol{x}$ is a constant vector and the expectation is taken with respect to the variable Y.

An error definition (a.k.a. a loss function) $L(Y, \hat{Y})$ is the cost of predicting $\hat{Y}$ when the response is Y. For continuous variables, squared error $L_{sq}(Y, \hat{Y}) = (Y - \hat{Y})^2$ and absolute error $L_{abs}(Y, \hat{Y}) = |Y - \hat{Y}|$ are widely used. They respectively are the squared or absolute value of the difference between the response and the estimate. The two error metrics gained popularity for different reasons: Squared error is mathematically tractable and leads to simplified forms in many theoretical and numerical analyses; absolute error directly measures the magnitude of the difference between response and estimate without changing the dimension of the response variable, hence is more natural for most estimation-focused applications. Some authors (e.g. Willmott et al., 2009) have advocated increasing the use of absolute error, considering that squared error's popularity arose from convenience rather than scientific justification. To be complete, we examine the BVD of both squared and absolute errors in this study, which also provides a way to compare the two error metrics.

Typical model training parameterizes a model to minimize a chosen loss function given training data, e.g., the ordinary least squares regression minimizes the sum of $L_{sq}(Y, \hat{Y})$ over the training set.

2.2. Definitions of BVD error components

BVD is most commonly used to evaluate data-driven models (e.g. linear regressions, decision trees), although it can be applied to any model with stochastic component(s). BVD was originally developed for analyzing squared errors in statistics (Geman et al., 1992), and was quickly recognized as an important tool for understanding inductive learning. In late 1990s, a number of studies attempted to extend the BVD definition to classification errors, but were all challenged by the mathematical differences between continuous and categorical variables. Uniting these diverged efforts, Domingos (2000) argued that instead of separately defining BVD for different loss functions, a better alternative is to develop a generalized BVD definition and examine its special cases for specific loss functions accordingly. His generalized definition has then been widely accepted. Later, James (2003) pointed out that in BVD analysis two sets of quantities are of interests: "characteristic quantities" and "effect quantities". For example, "variance" as a characteristic quantity measures the randomness of a variable, and "variance effect" measures the amount of estimation error caused by this randomness; similarly we have "bias" and "bias effect". Gao et al. (submitted for publication) recognized that understanding the connection between model "characteristics" and their "effects" on estimation accuracy can be essential for