CHARACTERIZING VARIABILITY AND MULTI-RESOLUTION PREDICTIONS OF VIRTUAL SENSORS

ASHOK N. SRIVASTAVA, PH.D. AND RAMA NEMANI, PH.D.*

Abstract. In previous papers, we introduced the idea of a Virtual Sensor, which is a mathematical model trained to learn the potentially nonlinear relationships between spectra for a given image scene for the purpose of predicting values of a subset of those spectra when only partial measurements have been taken. Such models can be created for a variety of disciplines including the Earth and Space Sciences as well as engineering domains. These nonlinear relationships are induced by the physical characteristics of the image scene. In building a Virtual Sensor a key question that arises is that of characterizing the stability of the model as the underlying scene changes. For example, the spectral relationships could change for a given physical location, due to seasonal weather conditions. This paper, based on a talk given at the American Geophysical Union (2005), discusses the stability of predictions through time and also demonstrates the use of a Virtual Sensor in making multi-resolution predictions. In this scenario, a model is trained to learn the nonlinear relationships between spectra at a low resolution in order to predict the spectra at a high resolution.

1. Introduction

In recent years, we have been developing the idea of a Virtual Sensor, which is a machine learning model that has learned the relationships between spectra for a given image scene. Once these relationships are known, the model can be used to predict values of the spectra using only a subset of the spectral measurements. This capability can prove to be very useful for scientific and engineering studies [5, 6, 7]. Related work from other groups include [2, 3] for both Earth Science and engineering applications.

We begin with a brief discussion of our approach to modeling remote sensing cubes from Earth observing satellites such as MODIS (Moderate Resolution Imaging Spectroradiometer). For purposes of the discussion presented here, we will model the observed data as a time series of matrices following the notation in [4]. The spatiotemporal random function $Z(\mathbf{u}, \lambda, t)$ is modeled as a finite number n of spatially correlated time series with the following representation: $Z(\mathbf{u}, \lambda, t) = [Z(\mathbf{u_1}, \lambda, t), Z(\mathbf{u}_2, \lambda, t), ..., Z(\mathbf{u}_n, \lambda, t)]^T$.

In this equation, $\mathbf{u}$ represents the two-dimensional spatial coordinate, λ represents the vector of measured wavelengths, and t represents time. The superscript ^{T} indicates the transpose operator. If multiple wavelengths are measured, then each Z_i is actually a matrix, and the function $Z(\mathbf{u}, \lambda, t)$ represents a data cube of size $(n \times \Lambda \times T)$, where these symbols represent the number of spatial locations, the total number of measured wavelengths, and the total number of time samples, respectively. An image scene $\mathbf{U}$ measured at Λ wavelengths corresponds to one data cube; if the same scene is sampled at T time steps, we get T such cubes.

*NASA Ames Research Center, Moffett Field, CA, 94035, ashok.n.srivastava@nasa.gov, rama.nemani@nasa.gov. This paper has been submitted to the Second NASA Data Mining Workshop: Issues and Applications in Earth Science, May 2006.

In this notation, the spatial coordinate $\mathbf{u}$ represents the coordinates (or index) of a measurement at a particular location in the field of view. Of course, this is not the only method available to model such time series. It is possible to model these time series as a set of temporally correlated spatial random functions, or as a deterministic or stochastic trend model. Kryiakidis and Journel, 1999 [4] have a more complete discussion of these models, along with their inter-relationships.

In building a Virtual Sensor, we estimate $P(Z(b_1,t)|Z(B_2,t))$ where b_1 is the 'target' wavelength band of the sensor of interest, and B_2 is the set of all other measured wavelengths, with $b_1 \cap B_2 = \emptyset$ and $b_1 \cup B_2$ being the entire set of wavelengths that are measured.

Although it is often possible to learn these relationships, the *stability of the relationship through time* and the ability of the estimator to *generate images at high resolution when only the low resolution measurements becomes important.* In this paper we examine these two areas. All the studies presented in this paper are based on 18 images taken during the year 2005 over Fresno CA. We chose this area because of the variability due to moisture and sunlight that the region experiences over a year. It is admittedly an easier region compared to deciduous forests where the degree of variation is expected to be much higher. Analysis of those areas will be discussed in future papers.

2. Assessing the Stability of the Spectral Relationships through Time

For a given spatial region $\mathbf{U}$ and time coordinate t_0, the spectral information can be correlated. [1] This correlation is induced by physical characteristics of the image scene. If those characteristics do not change significantly with time our model should be able to estimate $P(Z(b_1,t)|Z(B_2,t))$ with low error. If, however, the physical characteristics change significantly through time, the degree of correlation and more generally the interrelationships between the spectral measurements would change.

The choice of model used to perform the estimation of $P(Z(b_1,t)|Z(B_2,t))$ can make a difference in the overall prediction accuracy. For this study, we use two simple models, the first being a linear model and the second is an ensemble of bagged feedforward neural networks [1]. These models are widely available and are useful for benchmarking predictive models. They span the spectrum of the bias-variance tradeoff due to model complexity.

Figure 1 shows the observed variation in the prediction accuracy for a Virtual Sensor built with either one or three days worth of training data. The effects of seasonal variation are highest for the linear model trained with only one day of data. The data taken on 9/26 shows a high degree of variation. Inspection of the data indicates that this is likely due to a very low spectral reflectance in Channel 6 for that particular day, thus creating a poor signal to noise ratio. The normalized error shown on this figure was computed as $e = \frac{1}{N}\sum_i(|y_i - \hat{y}_i|/y_i)$ for N predictions.

3. Multi-Resolution Predictions

In the previous section we showed how predictive models can yield stable predictions through time given the appropriate model and training data. In this section we show how these models can be used to estimate the spectral intensity of a channel at a higher resolution than available from the sensor itself. In the MODIS instrument, Channel 6 is only available at 500 m, whereas the channels in the visible part of the spectrum are available at 250

[1]Although correlation generally implies linear correlation, we use the word to also include potentially nonlinear relationships.

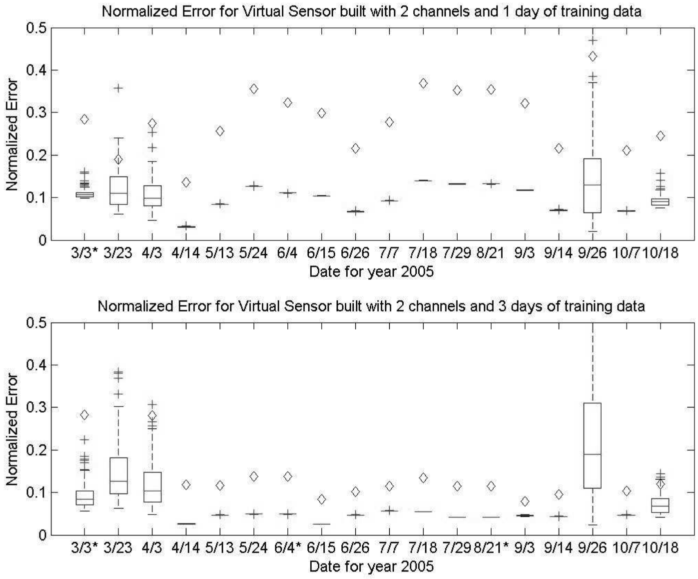

FIGURE 1. This figure shows the stability of model predictions as a function of 1 day of training data (top panel) and three days of training data (bottom panel). The box plot shows the variation due to model misspecification as a function of normalized error for a model built with an ensemble of 50 bagged neural networks. The diamonds indicate the predictions of a single linear model. Notice that the seasonal variation is more pronounced for the less-flexible linear model with one day of training data.

m. In this section, we show how three models: Gaussian process regression, bagged neural networks, and a linear model perform on this task.

Figure 2 shows the distributions of the input data at 500 m and 250 m resolutions. For this problem, we cannot compute an error measure between the actual Channel 6 values at 250 m and the predicted values, since the actual values do not exist at this resolution. Thus, we instead compare the distances between the distributions of Channel 6 at 500 m and the estimated distribution of Channel 6 at the 250 m resolution using the symmetric Kullback-Leibler distance. Since the low resolution measurements can be thought of as an aggregation of the high resolution measurements, the distributions can not be identical. For two distributions, p and q, this distance is defined as $d = \sum_i p_i \log(p_i/q_i) + q_i \log(q_i/p_i)$. The smaller this distance, the closer the distributions. For these models, the KL distance is $0.0167 \pm 0.0017, 0.0190 \pm 0.0003$, and 0.0260 ± 0.0000 for the GP, bagged neural network, and the linear model respectively. The GP thus has the best predictive accuracy according to this measure and a one-way ANOVA test performed over 50 trials confirms that these differences are statistically significant with $p < 0.001$.

4. Implications in the Earth Sciences

The monitoring component of Earth System Science depends on collecting observations in one of three domains: spatial, temporal and spectral. Significant progress has been made over the past three decades in all the domains. Three decades ago, for example, remote

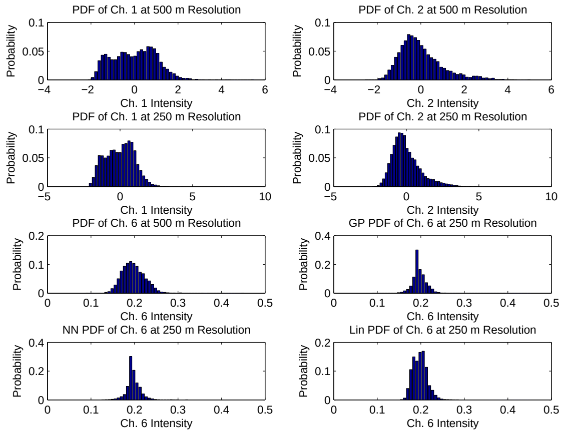

FIGURE 2. The first two rows in this figure show the distributions of the input data at 500 m resolution and at 250 m resolution. The third panel on the left shows the distribution of Channel 6 at 500 m resolution. The remaining three panels show the estimated distributions for Channel 6 at 250 m resolution using Gaussian process (GP) regression, bagged neural networks, and a linear model respectively. The Kullback-Leibler distance between these distributions and the true distribution shows that the GP regression performs the best.

sensing scientists were happy with data collected in four wavelengths, once every 16 days at 79m spatial resolution. Now orbiting satellites collect information nearly daily at 250-1000m in 36 wavelengths. Similarly hundreds of instrumented towers around the globe collect data every few minutes in support of verifying satellite data or for use in global climate change research. Operating these instruments continuously is a challenge. Similarly dealing with enormous volumes of data poses another set of problems that the community has not grappled with. We believe research on virtual sensors has the potential for contributing in a number of areas of earth sciences.

One such application is in the temporal domain dealing with filling gaps in CO2 flux data collected at FLUX tower sites around the globe. Gaps are common in these measurements, as CO2 instruments need to be calibrated frequently, while simultaneously collecting a set of environmental observations related to CO2 fluxes is much easier and robust. Therefore training virtual sensors on a full set of observations and using it to fill gaps is highly desirable.

Another area of earth sciences that can benefit from VS is in product generation from satellite data. Many of the algorithms used to generate geophysical products (Leaf area index, aerosols etc.) are complex and based on physical principles that need considerable resources for operational implementation. It is conceivable that one could directly learn the relationships between satellite data and products, thus simplifying the computational requirements. Such analysis may also help in determining the contribution or the need for a certain wavelength in producing a product or a set of products.

On-board processing and re-configurable sensors are two other areas of earth science research that VS have the potential to play a significant role. Since VS are data-driven models of a given phenomenon, one could implement such models for on-board processing so that important events such as detecting volcanoes, algal booms, and floods are detected early.

References

1. Leo Breiman, *Bagging predictors*, Machine Learning **24** (1996), no. 2, 123–140.
2. Maniscalco U. Ciarlini, P. and G. Regoliosi, *Validation of soft sensors for monitoring ambient parameters*, Proceedings of the International Conference on Advanced Mathematical and Computational Tools in Metrology (2005).
3. W. Dogget and S. Vazquez, *An architecture for real-time interpretation and visualization of structural sensor data in a laboratory environment*, Tech. report, NASA Langley Research Center, 2000.
4. P. C. Kyriakidis and A. G. Journel, *Geostatistical space-time models: A review*, Mathematical Geology **31** (1999), no. 6, 651–684.
5. A. N. Srivastava, N. C. Oza, and J. Stroeve, *Virtual sensors: Using data mining techniques to efficiently estimate remote sensing spectra*, IEEE Transactions on Geoscience and Remote Sensing **43** (2005), no. 3.
6. A. N. Srivastava and J. Stroeve, *Onboard detection of snow, ice, clouds, and other geophysical processes using kernel methods*, Proceedings of the ICML 2003 Workshop on Machine Learning Technologies for Autonomous Space Applications, 2003.
7. M. J. Way and A. N. Srivastava, *Novel methods for predicting photometric redshifts from broad band photometry using virtual sensors*, Submitted to Astrophysical Journal (2006).