

Modeling Spectral–Temporal Data From Point Source Events

Monica Reising, Max Morris, Stephen Vardeman & Shawn Higbee

To cite this article: Monica Reising, Max Morris, Stephen Vardeman & Shawn Higbee (2011) Modeling Spectral–Temporal Data From Point Source Events, *Technometrics*, 53:2, 183-195, DOI: [10.1198/TECH.2011.09014](https://doi.org/10.1198/TECH.2011.09014)

To link to this article: <http://dx.doi.org/10.1198/TECH.2011.09014>

 View supplementary material 

 Published online: 01 Jan 2012.

 Submit your article to this journal 

 Article views: 82

 View related articles 

Modeling Spectral–Temporal Data From Point Source Events

Monica REISING, Max MORRIS, and Stephen VARDEMAN

Department of Statistics
Iowa State University
Ames, IA 50011
(mmorris@iastate.edu)

Shawn HIGBEE

Air Force Research Laboratory
Sensors Directorate
Hanscom AFB
Bedford, MA 01731

In recent years, a great deal of effort has been invested in developing sensors to detect, locate, and identify “energetic” electromagnetic events. When observed through one type of imaging spectrometer, these events produce a data record that contains complete spectral and temporal information over the event’s evolution. This article describes the development of a statistical model for the data produced by a particular spectral–temporal sensor. While the application is unique in some ways, this approach to model building may be useful in other related contexts. Several plots, estimated parameters, and some additional details for an equation are provided in the Appendix which is available as supplementary material online.

KEY WORDS: Event discrimination; Infrared; Product correlation; Pseudo-imaging; Spectral imager.

1. INTRODUCTION

It is imperative in some military and national defense applications to quickly detect, locate, and identify short duration “energetic” electromagnetic events that have particular characteristic patterns of evolution over time. Spectral–temporal sensors (also known as pseudo-imagers) currently capture available information from energetic events and record observed intensities repeatedly in time for wavelengths ranging from the visible to the long-wave infrared (Pellegrini and Ewing 2004). Some of these important energetic events, particularly those associated with what are effectively point sources, can be difficult to detect, locate, and identify using traditional imaging cameras. Recently, considerable effort has been invested in pseudo-imaging sensors that use a direct vision prism to disperse incoming electromagnetic energy over a staring focal plane array depicted in Figure 1; see the work of Deming et al. (2006). Data examined in this article were collected using a sensor of this kind, operating in a restricted mode as described in Section 2.

While physical sensor technology is developing rapidly, there is a lag in the development of algorithms that can be used to identify and discriminate between types of energetic events in real time. There does not yet exist a coherent mathematical framework for discrimination of energetic electromagnetic events that (1) explicitly allows for many sources of variability (shot-to-shot variability, atmospheric variation, sensor noise, etc.), (2) can be implemented in real time, (3) cleanly handles “time registration of events” (recognizes that an incoming data stream may not contain data beginning at event initiation), and (4) can assess the likelihood that an incoming data stream matches the characteristic evolution of an “important” electromagnetic event. The phenomenology of an energetic event of interest will determine the optimal sensor and band pass (set of wavelengths to which it is sensitive) to use for detecting it. Although important for designing a sensor, such considerations are not of basic concern from a modeling point of view.

The research described here is in support of a longer-term effort that includes (1) the development of a “library” of signal characterizations corresponding to physical event types of

interest, (2) methods and algorithms for identifying, and discriminating among, data streams generated by event types in real time, based on comparison to the library of characterizations, and (3) an overall framework that supports quantifying the uncertainty and probabilities of error in such identification and discrimination applications. A systematic approach to event characterization, identification/discrimination, and uncertainty quantification can be undertaken only in the context of a *model* or a *family of models* that can effectively represent the structure (both signal and noise) in these data. This necessary process of modeling is itself complex, reflecting the complex character of spectral–temporal data, and our emphasis here is on the more fundamental issue of the *process* by which such models can be effectively developed.

In real applications, the signals of interest are very complex, reflecting the complex physical nature of their sources, and are subject to substantial environmental noise reflecting, among other things, variation in atmospheric conditions. To maintain a focus on *modeling methodology*, the development here is demonstrated using data collected in a small-scale experiment, performed under controlled conditions, with “events” defined by the striking of three kinds of matches. These restrictions simplify the overall scope of our description, but do not sacrifice the essential nature of temporal data collection for such problems. And, while the specific decisions that would be made in modeling a larger number of more complex events might be different in detail than those made here, the *approach* described can be used in more complicated settings of larger scale. Our intent is that the reader involved in spectral–temporal modeling applications might find this approach helpful whether his/her interest involves event discrimination or not.

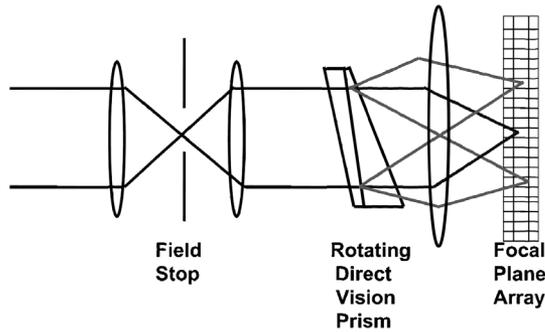


Figure 1. A schematic of the sensor used in data collection (rotation of the prism was disabled during our data collection).

2. DATA

Our test bed for developing statistical models consists of data on burns of three types of safety matches collected by a slightly modified spectral–temporal sensor. Safety matches (as opposed to other potential hot sources of radiation) were used so that we could easily control experimental conditions, and data for many event repetitions were collected.

The sensor we used typically employs a technique known as pseudo-imaging (or imaging) to provide spectral–temporal signatures and the locations of rapidly changing events within a given field of view. Pseudo-imaging is the process of taking data obtained from the focal plane array, creating a single spectral profile, and using the spectral profiles from multiple ‘snapshots’ to create a spectral–temporal signature from an observed event (Weeks et al. 2004). Figure 1 shows a schematic of the sensor we used from the work of Mooney et al. (1997) and Pellegrini and Ewing (2004).

Data were collected in May 2005 at Hanscom Air Force Base, MA. Three different types of safety matches were used to create realizations from three different classes of energetic events. Data were recorded for 20 matches of each type. The sensor used operated at a framing rate of 80 frames per second. After the sensor began recording, a match was ignited by hand and the sensor continued to collect data for nine seconds, producing 720 frames of measured intensities. A total of 720 frames were recorded for each test regardless of whether the match burned the entire time or not. For each of the 720 frames of data, a gray-scale image was produced. After all data were collected, the collection of pixels containing the brightest line in each gray-scale image was manually extracted. The intensity, represented as voltages resulting from photon counts, is recorded for each pixel along the line and compiled to create a matrix of data. Each row in this matrix corresponds to a single temporal frame, and each column to a single pixel on the focal plane array (an index for convenience corresponding to a wavelength in the spectrum). These are used to make what is commonly referred to as a “waterfall plot.” Figure 2 is such a plot.

In normal use, the sensor is equipped with a *rotating prism* and *rotating preprocessing algorithm* that can produce waterfall plots like that in Figure 2. However, for the purpose of initial development of modeling methods, we decided to not rotate the prism during data collection and began with raw, rather than

preprocessed, data from a single column of pixels on the focal plane array.

As mentioned previously, data were recorded for 20 matches of each of three types, say Class A, Class B, and Class C. In the process of collecting data for Class C, the prism was re-aligned after the first three observations. While preprocessing the data, we discovered that data for those three observations were extracted from a different column of pixels on the focal plane than the others. After careful consideration, these three observations were not used in the development of the model for Class C.

Class A and Class B models are expected to be similar, as a match from Class A was simply two matches from Class B physically bound together. The chemical composition of the Class A and Class B matches are identical. The matches from Class C are different in terms of chemical composition and the Class C model should be substantially different from those for Classes A and B.

Let the energy intensity observed by a sensor for one event be denoted by

$$S^*(t, l) \quad \text{for } t = 0, 1, \dots, 719 \text{ and } l = 1, 2, \dots, 149, \quad (1)$$

where l is an integer index used to represent a pixel on the focal plane array. $S_k^*(t, l)$ will be used to represent the measured energy at the t th time frame on the l th pixel for the k th observation for a given class of matches. Each value of l corresponds to a single pixel on the focal plane (since the spectral resolution of the data is a single pixel) and integer indexing was used for convenience. Note that a given value of l , say 80, corresponds to the same pixel on the focal plane for all time frames, and that any two consecutive values of l correspond to neighboring pixels on the focal plane. The raw, integer-valued data were preprocessed for analysis by “background subtraction,” pixel-for-pixel correction using the signal recorded just before the combustion event was initiated. Upon completion of the preprocessing, we had “useable” data like those portrayed in Figure 2. Let

$$S(t, l) \quad \text{for } t = 0, 1, \dots, 719 \text{ and } l = 1, 2, \dots, 149 \quad (2)$$

represent preprocessed data for a single event. Again, a value of l corresponds to the same pixel on the focal plane array for all event classes.

3. MODELING

To begin model development, suppose that a bright event of a particular type produces a sensed spectrum that evolves over continuous time (τ) and wavelength (λ) as

$$\theta(\tau, \lambda) \quad \text{for } \tau \in (0, T) \text{ and } \lambda \in (0, \infty). \quad (3)$$

We purposely use the notation $\theta(\tau, \lambda)$ here rather than $S(t, l)$, because for modeling purposes we wish to think of τ and λ as continuous variables and to do the modeling for all $\tau \in (0, T)$ and λ in some finite interval in $(0, \infty)$. The rationale here is that a point source has a signature that is continuous in time for some interval of wavelengths.

A model is needed for $\theta(\tau, \lambda)$ that explicitly allows for shot-to-shot, or realization-to-realization, variability. We will use a Gaussian random field model for a transformed version of

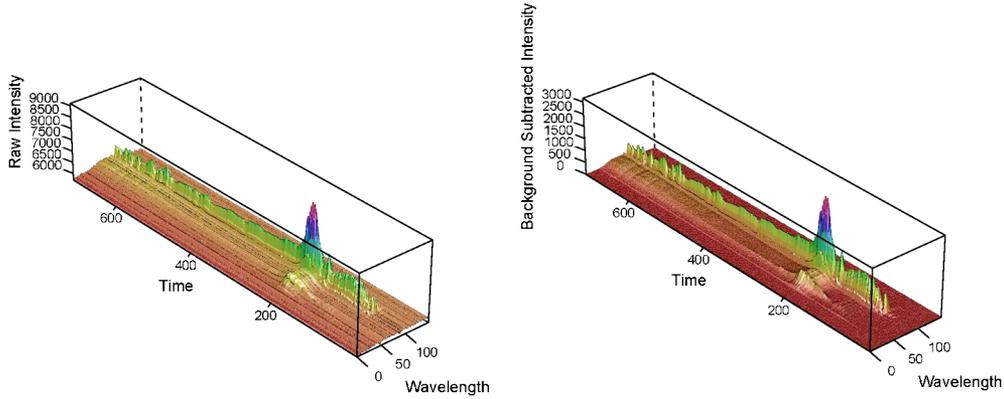


Figure 2. An example of raw and corresponding background subtracted data collected in the burn of a single match from Class A.

$\theta(\tau, \lambda)$ which we will call $\tilde{\theta}(\tau, \lambda)$. The exact transformation used in a particular case will be dictated by the data. In some applications it is possible that each class of events might require a different transformation.

Because physical intensities are nonnegative (read-out voltages as a function of photon counts) it is natural in this application to think of using a logarithmic transformation. A log transformation would convert any multiplicative effects, such as those produced by distance or atmospheric absorption, to additive effects and mitigate the “orders of magnitude” differences between responses seen across t and l pairs. However, due to the background subtraction and instrument noise, the preprocessed data (2) can include negative values. A transformation for θ was needed that would (1) have characteristics of the log transform where there is substantial signal, (2) be applicable to both positive and negative data values, and (3) be invertible to avoid information loss. Given these requirements, we selected the transform:

$$h(\theta) = \begin{cases} \ln(\theta) & \text{for } \theta \geq 2 \\ \theta(\ln(2)/2) & \text{for } -2 < \theta < 2 \\ -\ln(|\theta|) & \text{for } \theta \leq -2. \end{cases} \quad (4)$$

An integer value other than 2 might have been used as the threshold in (4), but a value of 1 would have violated requirement (3), and a value of 3 or more would have been less compatible with requirement (1).

Figure 3 shows the application of transform (4) to the values represented in Figure 2. These transformed values will further be referred to as intensities. It is important to call attention to the extreme wavelengths here. The pattern (and raw data) reveal that there is no signal specific to this event (or any other event in the class) in the extreme wavelengths displayed in Figure 3.

A Gaussian random field model for

$$\tilde{\theta}(\tau, \lambda) = h(\theta(\tau, \lambda)) \quad (5)$$

is characterized by a mean function $\mu(\tau, \lambda)$ and a covariance function $C((\tau_1, \lambda_1), (\tau_2, \lambda_2))$ defined for τ, τ_1 , and $\tau_2 \in (0, T)$ and λ, λ_1 , and λ_2 belonging to the interval of interest in $(0, \infty)$. For any finite set of time–wavelength pairs (τ_i, λ_i) , $i = 1, 2, \dots, m$, the joint distribution of values $\tilde{\theta}(\tau_i, \lambda_i)$ is modeled as multivariate normal with mean vector

$$E(\tilde{\theta}(\tau_i, \lambda_i)) = \mu(\tau_i, \lambda_i) \quad (6)$$

and covariance matrix with entries

$$\text{cov}(\tilde{\theta}(\tau_i, \lambda_i), \tilde{\theta}(\tau_j, \lambda_j)) = C((\tau_i, \lambda_i), (\tau_j, \lambda_j)). \quad (7)$$

3.1 Estimation Overview

There were many steps involved in obtaining the final models. We describe the details of our particular application in Section 4, but offer this preview here as a kind of “road map.” With some thought, these steps can be applied to model data that are collected from nearly any pseudo-imaging sensor operating in any part of the electromagnetic spectrum. Given below is a summary of the main steps we used to model our data once all preprocessing was complete.

The first step in estimating the model parameters was the time registration, described in Section 4.1. A temporal shift parameter for each observation within each class was estimated. Proper alignment of all events used in a training set was key to estimating the class mean and standard deviation functions. We note that time registration will also be important when models like this are used as the basis of discrimination and detection algorithms, as real events need not necessarily be observed from initiation.

After proper temporal alignment and all preprocessing was complete, the intensity values at each time–wavelength pair were averaged across all the observations within a class of

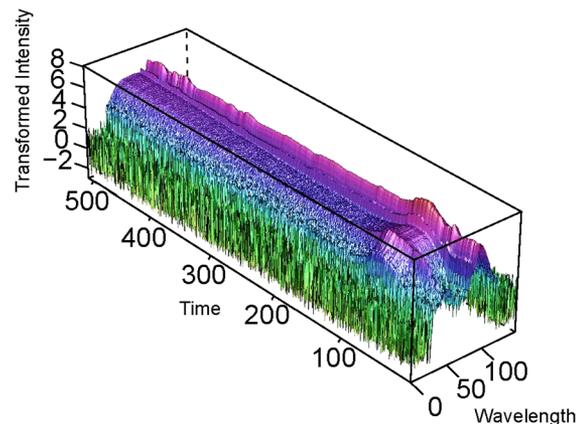


Figure 3. Transformed data corresponding to those represented in Figure 2.

data. The mean function represents the characteristic signature (for the preprocessed/transformed data) for the particular class of events and any observation from this class can be thought of as varying around the same characteristic signature. An observation-specific “intensity shift” parameter was estimated to allow for the modeling of objects from the same class of events observed at different distances. Recall that if two identical events were observed by the sensor through the same atmosphere at different distances, the event closer to the sensor would record higher intensity values. Next, the variance surface was estimated using the characteristic signature (mean surface) and the time- and intensity-shifted (preprocessed/transformed) observations. This captured the many sources of variability including, but not limited to, shot-to-shot, atmospheric interference, and sensor noise. Estimation of mean and variance surfaces and intensity shift parameters is discussed in Section 4.2.

The last step in the modeling was determining appropriate forms for the correlation structure and estimating parameters for those functions discussed in detail in Section 4.3. The process was complicated and only a sketch is provided here. The details that were used in modeling correlations for classes of safety matches may need to be altered for other classes of point source events observed through similar sensors. However, the main ideas used here should be applicable. The standardized residuals at each time index–wavelength index pair were used in the estimation of a correlation function in time and a separate correlation function in wavelength. Separate time and wavelength correlation functions were estimated for event initiation and steady-state phases and to the right and left sides of the undeviated wavelength. The data from the four corresponding time–wavelength regions were treated as independent. Next the standardized error process was decomposed into two components, one “white noise” uncorrelated component and the other involving correlation between intensities at different (t, l) pairs within the same observation. The variance function for the “white noise” was estimated for each (t, l) region within each class of events. After an adjustment for the “white noise” was made to the standardized error at each time index–wavelength index pair, empirical variograms were calculated to determine the form of the correlation structure and the parameter estimates for those correlation structures.

4. ESTIMATING MODEL PARAMETERS

Recall that the preprocessed data are

$$S(t, l) \quad \text{for } t = 0, 1, \dots, 719 \text{ and } l = 1, 2, \dots, 149 \quad (8)$$

as in (2), to be transformed as

$$\tilde{S}(t, l) = h(S(t, l)) \quad (9)$$

before modeling.

Following the rationale implicit in (5)–(9), we found that a plausible model for the k th observation within a given class is

$$\tilde{\theta}_k(\tau, \lambda) = d_k + \mu(\tau + \eta_k \Delta\tau, \lambda) + \sigma(\tau + \eta_k \Delta\tau, \lambda)\epsilon(\tau, \lambda). \quad (10)$$

In (10) the index k identifies a specific realization from the class, and d_k (intensity shift) and η_k (time shift) are parameters associated with the k th event and $\Delta\tau$ is the time difference

between two consecutive sampling points. Functions μ and σ are mean and standard deviation functions associated with the particular class of events and ϵ is a “spatial” stochastic process with mean zero, variance 1, and a correlation structure potentially specific to the class. We proceed to describe how we developed appropriate forms for $\mu(\cdot)$, $\sigma(\cdot)$, and the correlation structure for $\epsilon(\cdot, \cdot)$.

4.1 Training Event Time Registration

The raw data consist of 720 frames for each observation from each class of events, but the physical bright event did not start in the same frame of data for each observation. To estimate appropriate mean and standard deviation functions for characterizing a class, the bright events need to be aligned temporally. This alignment is accomplished by selecting values η_k in (10). The estimates of η_k for observations in a class were calculated from total intensity series. Let

$$I_k(t) = \sum_l S_k(t, l) \quad (11)$$

denote the total intensity at time t for event k from the class under consideration. The temporal alignment parameter, η_k , was estimated as

$$\hat{\eta}_k = \min_{k'} \left(\arg \max_t I_{k'}(t) \right) - \arg \max_t I_k(t). \quad (12)$$

That is, the realization with the earliest intensity “peak” was selected as a “base case,” and time-offsets for all other realizations are determined with reference to it. Figure 4 illustrates a subset of total intensity series for one class of events and the same subset of total intensity series aligned according to the values of the $\hat{\eta}_k$ computed using (12).

4.2 Estimating the Class Mean and Standard Deviation Surfaces and Event-Specific Intensity Shift Parameters

It is well known that the distance from which bright events are observed will affect the intensities registered by the sensor. If two identical events are observed by the camera through the same atmosphere at two different distances, the event closer to the sensor registers higher intensity values than the event that is observed from a greater distance. This phenomenon demonstrates the need for d_k , an intercept or an “intensity shift” parameter, in model (10). To estimate the intensity shift parameter, d_k , for an observation, the mean function for the class under

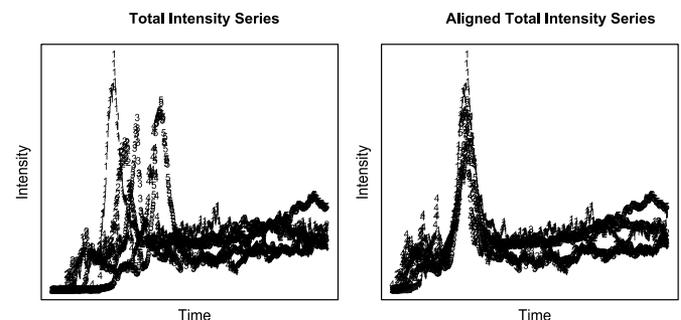


Figure 4. Total intensity and aligned total intensity series.

consideration, $\mu(\tau, \lambda)$, must first be estimated. For every (t, l) pair (after temporal alignment) an average was taken across the k observations within a class. The domain of the empirical mean function was trimmed so that only those intensities corresponding to times where all observations make a contribution to the average are included. (That is, we effectively reindex time for all realizations to match these with $\hat{\eta}_k = 0$ and only estimate mean response for such $t = 0, 1, \dots, 719 - \max_k |\hat{\eta}_k|$ and $l = 1, 2, \dots, 149$.) Let the estimated mean function be denoted by $\mu'(t, l)$ for such t and l .

With an estimated mean function in hand, we estimated the intensity shift, d_k , for each observation. We computed this as

$$\hat{d}_k = \frac{1}{149(719 - \max_k |\hat{\eta}_k| + 1)} \times \sum_{t=0}^{719 - \max_k |\hat{\eta}_k|} \sum_{l=1}^{149} (\tilde{S}_k(t + \hat{\eta}_k, l) - \mu'(t, l)). \quad (13)$$

After the mean function and intensity shift parameters are estimated, residuals may be computed as

$$\hat{e}_k(t, l) = \tilde{S}_k(t + \hat{\eta}_k, l) - (\hat{d}_k + \mu'(t, l)). \quad (14)$$

Note that for each (t, l) pair, the average of these residuals over all observations is zero. Then the class-specific variance at each observed time–wavelength pair (t, l) was estimated as

$$\hat{\sigma}^2(t, l) = \frac{1}{N - 1} \sum_k (\hat{e}_k(t, l))^2, \quad (15)$$

where N is the number of replicate trials from the class. Final estimates of the mean and variance functions are computed as smoothed (in both time and wavelength) versions of $\mu'(t, l)$ and $\hat{\sigma}^2(t, l)$. A loess smoothing technique was used with a smoothing parameter of 0.1. (With a smoothing parameter of 0.1, the smoother uses the nearest ten percent of observations and replaces the current observation with a tricubic weighted average of those nearest observations.) Figures A.1–A.3 in the online Appendix show the smoothed mean and standard deviation surfaces; the mean surface is displayed to the left of the standard deviation surface for each class. The first 35 wavelength values (the left side wavelengths) were removed prior to plotting. This was done so the surface could be “seen” for the middle wavelengths. These middle wavelengths are where the standard deviation surfaces vary the most from event class to event class.

Next, standardized residuals may be calculated using $\hat{\mu}(t, l)$ and $\hat{\sigma}(t, l)$. For each observed time–wavelength pair (t, l) the standardized residuals for the k th observation are

$$\hat{\xi}_k(t, l) = \frac{\tilde{S}_k(t + \hat{\eta}_k, l) - (\hat{d}_k + \hat{\mu}(t, l))}{\hat{\sigma}(t, l)}. \quad (16)$$

The standardized residuals, $\hat{\xi}_k(t, l)$, were constructed to have approximately mean 0 and standard deviation 1 for each (t, l) pair, across the events in the class. These are treated as partial realizations of the stochastic process $\epsilon(\tau, \lambda)$ in (10), and are used in estimating the covariance function for that process. Figure 5 is a plot of the standardized residuals $\hat{\xi}(t, l)$ for the raw data represented in Figure 2. It may appear that there is increased variability in the extreme wavelengths in Figure 5. As mentioned earlier, these extreme wavelengths do not contain

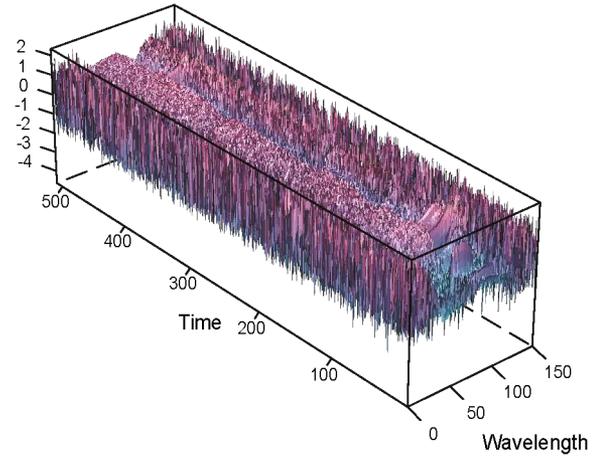


Figure 5. A single standardized residual surface (corresponding to the data represented in Figure 2).

signal specific to this event. The standardized errors in these extreme wavelengths appear to be white noise and that observations in adjacent pixels are uncorrelated. These regions are not further discussed in the modeling. In the estimation of the correlation structure, the discussion will only pertain to the (t, l) pairs in which signal was apparent. The total intensity in time for a particular wavelength was studied in making a determination of whether signal was apparent in a particular wavelength or not.

4.3 Estimating the Class Correlation Structure

Several sequential steps were needed in investigating and ultimately estimating the correlation in both time and wavelength.

4.3.1 Partitioning the Data Into Four Independent Regions. Data are recorded over a time interval that can naturally be divided into “initiation” and “steady-state” phases, and over a spectrum of frequencies that can be divided into “high” and “low” relative to a particular constant frequency (explained below). Based on this, we developed a model for each event type by segmenting the set of observed (t, l) pairs into four rectangular sets by what we will term class-specific “independence walls.” The corresponding four sets of $\epsilon(\tau, \lambda)$ ’s are assumed to be independent of each other.

(This assumption was confirmed by noting that calculated correlations between values from different regions tend to be relatively small.) Figure 6 gives a diagram of the independence walls and the four (τ, λ) regions they delineate. The left side corresponds to shorter wavelengths (small values of l in our indexing) and the right side corresponds to longer wavelengths. The vertical line represents an independence wall built into the model separating (t, l) pairs at the *undeviated wavelength*, the wavelength of energy that passes through the prisms undeflected. Our modeling assumption is that any $\epsilon(\tau, \lambda)$ with λ less than the undeviated wavelength is independent of any $\epsilon(\tau, \lambda)$ with λ greater than the undeviated wavelength. The horizontal line represents an independence wall built into the model in the time direction. This separates what we will call the event initiation phase (or early time) from what we will call the steady-state phase (or late time). In Figure 4, the steady-state phase is

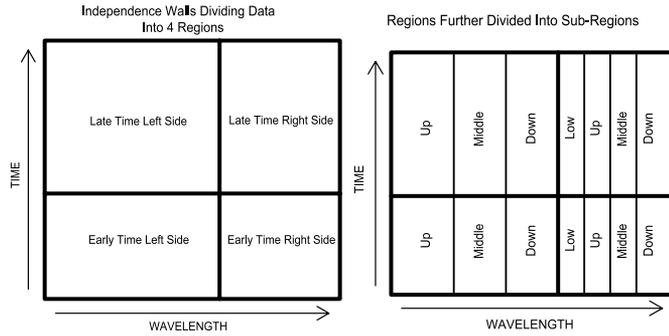


Figure 6. Independence walls in (t, l) space showing four regions and subregions.

represented by the part of the plot to the right of the sharp peak, where the total intensity series tends to vary relatively slowly. One can now think of the data as separated into four sets defined by the regions seen in Figure 6, right late time region, right early time region, left late time region, and left early time region.

4.3.2 Dividing Regions Into Subregions. Due to patterns remaining in the apparent “local roughness” of the standardized residual surfaces (see Figure 5), correlation structures were investigated separately in three or four subregions within each of the four regions (see Figure 6). “Local roughness” in realizations of a stochastic process is indicative of the nature of the correlation structure, smoothness indicating high correlation between responses with (τ, λ) in close proximity to each other. In the regions to the left of the undeviated wavelength, in both early and late time, (t, l) pairs were further divided into three subregions in the wavelength direction that we will call the “up” subregion, the “middle” subregion, and the “down” subregion according to the behavior of the estimated mean function $\hat{\mu}(t, l)$ in them. On the right side in the wavelength direction, in both early and late time, (t, l) pairs were further divided into four subregions, the “low” subregion, the “up” subregion, the “middle” subregion, and the “down” subregion according to the behavior of the estimated mean function in these regions. These subregions do not correspond to equal numbers of wavelength indices or corresponding subregions of equal size in different classes. All of the partitioning indicated in Figure 6 was done based on the empirical patterns seen in the training data for the three classes. The goal in dividing the (t, l) plane into subregions in the observed wavelength direction was to allow us to change the scales on the wavelength axis, subregion to subregion, so that a single stationary correlation function could be used across each region (after transforming wavelength).

4.3.3 Calculating Variograms and Decomposition of the Stochastic Process. In seeking a simple correlation structure for $\epsilon(\tau, \lambda)$, our hope was that a product form (in τ and λ) would be adequate. Accordingly, our attention turned to calculating variograms of the standardized residuals, $\hat{\xi}_k(t, l)$, for each class of objects in the time and wavelength directions separately. [See the book by Cressie (1993) beginning on page 58 for a discussion of variograms and their estimation.] A natural estimator (based on the method of moments) for the variogram was calculated in the wavelength direction, for each event and fixed time. Likewise, for each event and fixed wavelength, the empirical variogram was calculated in the time direction.

When seeking a simple functional form to fit to the empirical variograms in the wavelength direction for fixed time, we noticed that the “nugget” appeared to decrease in time in our test bed data. (The nugget is the size of the apparent discontinuity of a variogram at the origin, representing the scale of the uncorrelated “white noise” portion of variation in a stochastic process.) Because the variograms we calculated indicated both correlation between values and independent noise associated with each value, we decided to adopt a model that partitions the stochastic process into a white noise component for which the variance is a function of time and wavelength, independent of a second (correlated) component modeled using a product covariance structure. That is, we propose to employ a decomposition of the stochastic process as

$$\epsilon(\tau, \lambda) = \epsilon_1(\tau, \lambda) + \epsilon_2(\tau, \lambda), \tag{17}$$

where

$$\begin{aligned} E(\epsilon_1(\tau, \lambda)) &= 0 \quad \forall \tau, \lambda \quad \text{and} \\ \text{Var}(\epsilon_1(\tau, \lambda)) &= \delta^2(\tau, \lambda) \quad \forall \tau, \lambda, \\ \text{Cov}(\epsilon_1(\tau, \lambda), \epsilon_1(\tau', \lambda')) &= 0 \quad \text{for } \tau \neq \tau' \text{ or } \lambda \neq \lambda', \\ E(\epsilon_2(\tau, \lambda)) &= 0 \quad \forall \tau, \lambda \quad \text{and} \\ \text{Var}(\epsilon_2(\tau, \lambda)) &= 1 - \delta^2(\tau, \lambda) \quad \forall \tau, \lambda, \\ \text{Cov}(\epsilon_2(\tau, \lambda), \epsilon_2(\tau', \lambda')) &= R_\tau(|\tau - \tau'|)R_\lambda(|\lambda - \lambda'|)\sqrt{(1 - \delta^2(\tau, \lambda))(1 - \delta^2(\tau', \lambda'))} \\ &\quad \forall \tau, \lambda \text{ and } \tau', \lambda', \quad \text{and} \\ \text{Cov}(\epsilon_1(\tau, \lambda), \epsilon_2(\tau', \lambda')) &= 0 \quad \forall \tau, \lambda \text{ and } \tau', \lambda', \end{aligned}$$

where $R_\tau(|\tau - \tau'|)$ represents a nuggetless (i.e., continuous at the origin) correlation function in the time direction and $R_\lambda(|\lambda - \lambda'|)$ represents the nuggetless correlation function in the wavelength direction. We emphasize that R_λ is not assumed to be the same function in each subregion and that our model will eventually allow for nonzero correlation between standardized errors from subregions within the same region.

4.3.4 Estimating the Variance of the White Noise. In order to begin fitting the decomposition described in (17) to the standardized residuals in a given region, we first needed to estimate $\delta^2(\tau, \lambda)$. This cannot be done directly because there are no “true replicates” in the data, but we can obtain an estimate based on the degree of local variability near each (t, l) . To this end, let

$$Y_k(t, l) = \hat{\xi}_k(t, l) - \frac{1}{4}[\hat{\xi}_k(t, l - 1) + \hat{\xi}_k(t, l + 1) + \hat{\xi}_k(t - 1, l) + \hat{\xi}_k(t + 1, l)]. \tag{18}$$

The variable Y is, for each observed standardized residual, a difference between $\hat{\xi}$ and the average of its four nearest neighbors in time and wavelength. Intuition for this quantity was that it should roughly mimic ϵ_1 .

The squared residual, $Y_k^2(t, l)$, was calculated for each object in a class and each observed (t, l) combination other than at t - or l -edges, and these values were averaged across all observations within a class, for each (t, l) pair. The average was then smoothed using a loess smoothing technique, and each sequence of values for a fixed time was examined. A functional

form was fit for the small and large l values (for each fixed time point). A nonnegative Y_k^2 value for the first and last l values for each fixed time step was extrapolated. A similar procedure was used to extrapolate Y_k^2 for the first and last time step (for each fixed l). Let \check{Y}_k denote smoothed Y_k with the extrapolated edge values.

Notice that the expectation of $Y_k(t, l)$ in (18) is zero for all t, l . Hence,

$$E(Y_k^2(t, l)) = \text{Var}(Y_k(t, l)). \tag{19}$$

Therefore, it can be shown (see the online Appendix)

$$\text{Var}(Y_k(t, l)) \approx \frac{5}{4} \delta_k^2(\tau, \lambda). \tag{20}$$

Hence, the white noise variance function might be estimated for a particular region for a particular event class as

$$\hat{\delta}_k^2(t, l) = \min\left(\frac{4}{5} \check{Y}_k^2(t, l), 1\right) \quad \forall t, l. \tag{21}$$

It is worth noting that in our examples, $\hat{\delta}(t, l) < 1$ for all (t, l) . Figures A.4–A.6 in the online Appendix show estimated $\hat{\delta}_k^2$ surfaces for all four regions for each of the three classes of events. One will notice that these plots show that the values increase near the boundaries of each region. This is not surprising as these areas have the largest change in intensity within an observation.

4.3.5 Estimating the Correlation. With estimates (21) in hand, we turned our attention to estimating parametric forms for the nuggetless correlation structure [the correlation structure for ϵ_2 in (17)] in both the time and the wavelength directions. We based these estimates on the relationships

$$E(\epsilon_k(\tau, \lambda) - \epsilon_k(\tau, \lambda'))^2 = 2 - 2\sqrt{(1 - \delta_k^2(\tau, \lambda))(1 - \delta_k^2(\tau, \lambda'))} R_\lambda(|\lambda - \lambda'|) \tag{22}$$

and

$$E(\epsilon_k(\tau, \lambda) - \epsilon_k(\tau', \lambda))^2 = 2 - 2\sqrt{(1 - \delta_k^2(\tau, \lambda))(1 - \delta_k^2(\tau', \lambda))} R_\tau(|\tau - \tau'|). \tag{23}$$

To estimate a subregion correlation function in the wavelength direction specific to the data here, $R_\lambda(|l - l'|)$, first, for each fixed time index t and pair of wavelength indices l' and l'' for an observation within a class, we computed

$$X_{L,k}(t, l', l'') = \frac{2 - (\hat{\xi}_k(t, l') - \hat{\xi}_k(t, l''))^2}{2\sqrt{(1 - \hat{\delta}_k^2(t, l'))(1 - \hat{\delta}_k^2(t, l''))}}. \tag{24}$$

Similarly, to estimate the correlation function in the time direction specific to the data here, $R_\tau(|t - t'|)$, first, for each fixed wavelength index l and pair of time indices t' and t'' for an observation within a class, we computed

$$X_{T,k}(t', t'', l) = \frac{2 - (\hat{\xi}_k(t', l) - \hat{\xi}_k(t'', l))^2}{2\sqrt{(1 - \hat{\delta}_k^2(t', l))(1 - \hat{\delta}_k^2(t'', l))}}. \tag{25}$$

For each fixed time index t , values $X_{L,k}(t, l', l'')$ were then averaged according to

$$\hat{X}_{L,k}(t, \Delta l) = \frac{\sum_{\mathbf{N}(\Delta l)} X_{L,k}(t, l', l'')}{|\mathbf{N}(\Delta l)|}, \tag{26}$$

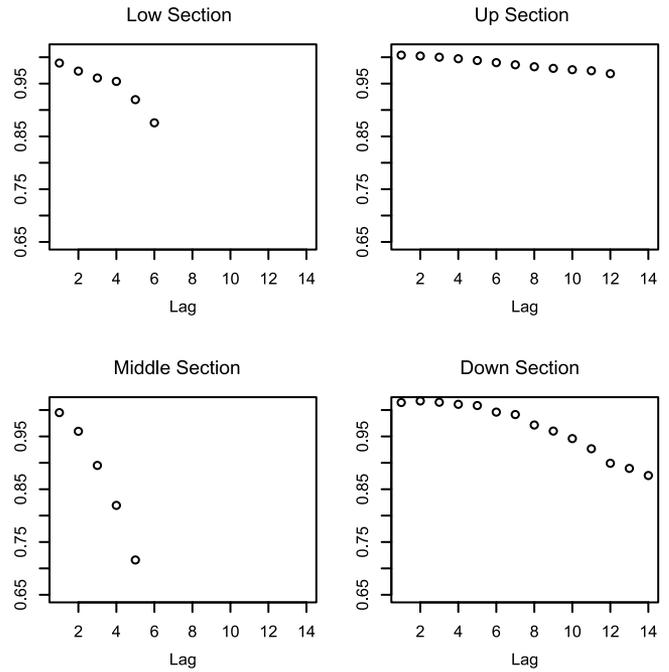


Figure 7. Empirical correlations $\hat{\mathbf{R}}_{L,k}$ at various wavelength lags averaged across time for all subregions within a single region for one event class (corresponding to the data represented in Figure 2).

where $\mathbf{N}(\Delta l) \equiv \{(l', l'') : |l' - l''| = \Delta l\}$ and $|\mathbf{N}(\Delta l)|$ is the number of distinct pairs in $\mathbf{N}(\Delta l)$, which varies from subregion to subregion within a class as seen in Figure 7. A similar averaging was done with the $X_{T,k}(t', t'', l)$ for each fixed wavelength index l resulting in $\hat{X}_{T,k}(\delta t, l)$.

The statistics used to estimate the correlation functions for a single observation were sequences of the above-calculated values, specifically

$$\hat{\mathbf{R}}_{L,k}(t) = (\hat{X}_{L,k}(t, 1), \hat{X}_{L,k}(t, 2), \dots, \hat{X}_{L,k}(t, \max \Delta l))$$

and

$$\hat{\mathbf{R}}_{T,k}(l) = (\hat{X}_{T,k}(1, l), \hat{X}_{T,k}(2, l), \dots, \hat{X}_{T,k}(\max \Delta t, l)).$$

To find a single parametric form for a correlation function in the wavelength direction, we averaged all of the values $\hat{\mathbf{R}}_{L,k}(t)$ across the appropriate t in a given observation. These empirical variograms in the wavelength direction were constructed separately for each subregion. Plots of these averages versus lag in wavelength index, Δl , were examined and a parametric form was chosen. Figure 7 shows plots of $\hat{\mathbf{R}}_{L,k}(t)$ averaged across time values for the four subregions within a region for a single observation from one class of events. In the wavelength direction, these plots were similar from observation to observation within a class and suggested that a reasonable single form for a correlation function might be Gaussian, that is, of the form $\exp(-\beta(\Delta l)^2)$.

The Gaussian correlation structure is associated with smooth realizations, which is consistent with the physical nature of our data.

We computed $\hat{\mathbf{R}}_{L,k}(t)$ for each t for each subregion within each region for each observation and then averaged across time to produce $\bar{\mathbf{R}}_{L,k}$ for each observation in the training data (for the given subregion for the relevant event class). Nonlinear least

squares was used to estimate the parameters of the Gaussian form expanded to include a nugget term,

$$\beta_1 + \exp(-\beta_2(\Delta l)^2). \quad (27)$$

The values of β_1 in the final forms used are set equal to zero because by our assumptions, the correlation function must approach a value of 1 as Δl approaches zero. However, in obtaining a good estimate of correlations at short lags, this fitted form was used to avoid biasing estimates of β_2 when there is some lack-of-fit at small wavelength differences. After the fitting, the empirical distributions of fitted β_1 and β_2 for a given subregion and event class across the observations for that class were examined. The distributions of observation-specific β_1 estimates were centered around zero, as one would expect. A final single estimate of β_2 used in the correlation function for modeling in a given subregion within a given region for a specific class of events was taken to be the median of observation-specific β_2 estimates; we denote this single value as $\hat{\beta}_2$.

The $\hat{\beta}_2$ values for the various subregions varied widely within a given region, indicating that even within a region a correlation structure appropriate to modeling the ϵ_2 of (17) cannot be stationary in l across an entire region. Nevertheless, by changing scale for wavelength indices (or introducing a “pseudo-wavelength” corresponding to each original wavelength index) it is possible to make use of a single (stationary) Gaussian correlation structure in wavelength (thereby assuring that what we have specified is indeed mathematically coherent/valid).

To illustrate, in the early-time-left-side region, there were estimated values for β_2 for each subregion, denoted as $\hat{\beta}_{2(\text{up})}$, $\hat{\beta}_{2(\text{middle})}$, and $\hat{\beta}_{2(\text{down})}$. Using these values, and setting $\hat{\beta}_{2(\text{up})}$ as a reference value, the size of the pseudo-distance (between adjacent wavelength indices) in a particular subregion of the region will be denoted as \tilde{l}^* :

$$\tilde{l}_{(\cdot)}^* = \sqrt{\frac{\hat{\beta}_{2(\text{up})}}{\hat{\beta}_{2(\cdot)}}}. \quad (28)$$

\tilde{l}^* represents the scaled “distance” between wavelength indices in a given subregion relative to a scaled “distance” of 1 in the reference subregion. To obtain an appropriate vector of pseudo-distances for a given subregion, the pseudo-wavelengths needed to be spaced appropriately. The (original and pseudo-) wavelength indices in the reference subregion, $(l_{1(\text{up})}, l_{2(\text{up})}, \dots, l_{\text{last}(\text{up})})$, are spaced so that consecutive values differ by 1, that is, $|l_i - l_j| = |i - j| \forall i, j$. The pseudo-wavelengths in the middle and down subregions are then defined as

$$\tilde{l}_{n(\text{middle})} = l_{\text{last}(\text{up})} + n * \tilde{l}_{(\text{middle})}^* \quad \text{for } n = 1, 2, \dots, N(\text{middle}) \quad (29)$$

and

$$\tilde{l}_{n(\text{down})} = l_{\text{last}(\text{middle})} + n * \tilde{l}_{(\text{down})}^* \quad \text{for } n = 1, 2, \dots, N(\text{down}), \quad (30)$$

respectively, where $N(\text{middle})$ is the number of wavelength indices in the middle subregion and $N(\text{down})$ is the number of wavelength indices in the down subregion.

Based on these pseudo-wavelengths a single stationary (in pseudo-wavelength) correlation function for the ϵ_2 process in the wavelength direction within each class and region was used. That is, for a given region, the correlation function for the ϵ_2 process in the wavelength direction is

$$\exp(-\hat{\beta}_{2(\text{up})}(\tilde{l}' - \tilde{l}'')^2) \quad (31)$$

for \tilde{l}' and \tilde{l}'' pseudo-wavelengths corresponding to wavelength indices l' and l'' . A similar process was used to estimate a single correlation function in the right side regions.

In a development parallel to the foregoing, we modeled correlations in the ϵ_2 process in the time direction. Ultimately, we decided to represent time correlation with a simple linear function. Least median squares was used (one observation at a time) to estimate the parameters of the linear form

$$\bar{\mathbf{R}}_T(\Delta t) \approx \alpha_1 + \alpha_2 \Delta t. \quad (32)$$

Note that in this parameterization, values of α_2 must be negative. Again, our assumption is that there is no nugget in this form, so the value of α_1 in the final form will be set equal to 1. The median of the fitted values of α_2 for a class, denoted by $\hat{\alpha}_2$, was used as the estimate for the entire region. Ultimately (so that the correlation at time lag 0 is 1 and no negative correlations are produced) we specify that for a given region, the correlation for the ϵ_2 process in the time direction for our data was (for $\hat{\alpha}_2$ specific to the class of events and region) $\max(0, 1 + \hat{\alpha}_2|t' - t''|)$.

We should note that choices for correlation forms other than our selections [(27) and (32)] might have been made. We have not fully evaluated a wide variety of parametric forms, but again note that our choices were made based on the examination of data plots.

4.3.6 Obtaining the Estimated Covariance Matrix. Having developed plausible wavelength and time correlation models for ϵ_2 from each region for each event class, it remained to combine the wavelength and time functions to create a single correlation function for a given region. One simple and mathematically coherent way to do this was to adopt a product correlation form. That is, if $\beta_{2(\text{reference})}$ and α_2 are the wavelength and time correlation parameters appropriate for a given region for a particular event class, then we adopted a model that specifies that the correlation for the ϵ_2 process for our data was, for time indices t' and t'' and wavelength indices l' and l'' ,

$$\text{corr}(\epsilon_2(t', l'), \epsilon_2(t'', l'')) = \exp(-\beta_{2(\text{reference})}(\tilde{l}' - \tilde{l}'')^2) \times \max(0, 1 + \alpha_2|t' - t''|), \quad (33)$$

where \tilde{l}' and \tilde{l}'' are pseudo-wavelength indices corresponding to l' and l'' , respectively. It should be noted that the adoption of this product, or “separable” correlation form, is a substantial modeling assumption. More general nonseparable forms exist (e.g., Schabenberger and Gotway 2005) and are necessary for model fidelity in some spatial applications. Besides functional simplicity, the product form can be exploited in our application to accelerate calculations needed for real-time comparison of new signals with a collection of class models.

We were now ready to assemble a single covariance structure for $\epsilon(\tau, \lambda)$ of displays (10) and (17) (the standardized error process) and finally a covariance structure for $\tilde{\theta}(\tau, \lambda)$. This is most

economically described in matrix terms. For a given region with time indices $t = t_{\min}, t_{\min} + 1, \dots, t_{\max}$ and wavelength indices $l = l_{\min}, l_{\min} + 1, \dots, l_{\max}$ there are $(t_{\max} - t_{\min} + 1)(l_{\max} - l_{\min} + 1)$ pairs of (t, l) indices of interest. Adopt a standard order in vector form as

$$\left((t_{\min}, l_{\min}), (t_{\min}, l_{\min} + 1), \dots, (t_{\min}, l_{\max}), \dots, (t_{\min} + 1, l_{\max}), \dots, (t_{\max}, l_{\min}), \dots, (t_{\max}, l_{\max}) \right)^T. \quad (34)$$

Then with rows and columns indexed by integers t' and t'' from t_{\min} to t_{\max} , define the $(t_{\max} - t_{\min} + 1) \times (t_{\max} - t_{\min} + 1)$ matrix \mathbf{T} , with elements $\max(0, 1 + \alpha_2 |t' - t''|)$. Similarly, with rows and columns indexed by integers l' and l'' from l_{\min} to l_{\max} and letting \tilde{l} stand for the pseudo-wavelength corresponding to wavelength index l , we define the $(l_{\max} - l_{\min} + 1) \times (l_{\max} - l_{\min} + 1)$ matrix \mathbf{F} with elements $\exp(-\beta_{2(\text{reference})}(\tilde{l}' - \tilde{l}'')^2)$. With this notation, the estimated correlation matrix for the ϵ_2 process over the (t, l) pairs listed in display (34) is compactly written as $\mathbf{T} \otimes \mathbf{F}$ (the Kronecker product of \mathbf{T} and \mathbf{F}).

Next, recall that the variances of the “white noise” quantities ϵ_1 are different at each (t, l) pair and are estimated as $\hat{\delta}^2(t, l)$. Then the estimated variance of the nuggetless component ϵ_2 must be $1 - \hat{\delta}^2(t, l)$. For (t, l) pairs in a region listed as in (34), let

$$\mathbf{\Delta} = \text{diag}(\hat{\delta}^2(t, l)). \quad (35)$$

Then the estimated covariance matrix for the nuggetless process ϵ_2 is

$$(\mathbf{I} - \mathbf{\Delta})^{1/2} \mathbf{T} \otimes \mathbf{F} (\mathbf{I} - \mathbf{\Delta})^{1/2} \quad (36)$$

and for

$$\epsilon(\tau, \lambda) = \epsilon_1(\tau, \lambda) + \epsilon_2(\tau, \lambda), \quad (37)$$

assuming the two components are independent, the estimated covariance matrix is

$$(\mathbf{I} - \mathbf{\Delta})^{1/2} \mathbf{T} \otimes \mathbf{F} (\mathbf{I} - \mathbf{\Delta})^{1/2} + \mathbf{\Delta}. \quad (38)$$

Then, with the estimated class variance function for the region of interest again $\hat{\sigma}^2(t, l)$, let

$$\mathbf{\Sigma} = \text{diag}(\hat{\sigma}^2(t, l)). \quad (39)$$

The estimated covariance matrix for the values of $\tilde{S}(t + \hat{\eta}_k, l)$ in the given region is

$$\mathbf{\Sigma}^{1/2} \{ (\mathbf{I} - \mathbf{\Delta})^{1/2} \mathbf{T} \otimes \mathbf{F} (\mathbf{I} - \mathbf{\Delta})^{1/2} + \mathbf{\Delta} \} \mathbf{\Sigma}^{1/2} = \mathbf{D}_1 [\mathbf{T} \otimes \mathbf{F} + \mathbf{D}_2] \mathbf{D}_1, \quad (40)$$

where diagonal matrices $\mathbf{D}_1 = \mathbf{\Sigma}^{1/2} (\mathbf{I} - \mathbf{\Delta})^{1/2}$ and $\mathbf{D}_2 = \text{diag}(\frac{\delta^2(t, l)}{1 - \delta^2(t, l)})$ are specific to the collection of (t, l) pairs being considered.

So finally, we have completely specified a model for the burns of each of three types of safety matches, Class A, Class B, and Class C. A partial description of the particular models developed for these three classes of safety matches is given in the online [Appendix](#). By using this modeling methodology, information from every indexed time and wavelength has been retained and modeled. No data reduction was necessary or used before completely modeling the three classes of safety matches. We have effectively modeled both the characteristic signatures

of the three event types and the observed variability from a variety of sources, including shot-to-shot, atmospheric interference, and sensor noise. Future work can use models like these in the development of detection and discrimination algorithms.

5. SIMULATING AND ASSESSING DATA FROM FITTED MODELS

We illustrate the effectiveness of our modeling by generating simulated sensor data from our fitted models and comparing it to the original (real) data. For a specified event class, data were simulated independently in the four regions indicated in Figure 6 and then put together to form a single dataset.

Let \mathbf{Q}_1 and \mathbf{Q}_2 be independent, multivariate normal random vectors with $E(\mathbf{Q}_1) = E(\mathbf{Q}_2) = \mathbf{0}$ and

$$\text{Var}(\mathbf{Q}_1) = \mathbf{T} \otimes \mathbf{F} \quad (41)$$

and

$$\text{Var}(\mathbf{Q}_2) = \mathbf{D}_2, \quad (42)$$

where \mathbf{T} , \mathbf{F} , and \mathbf{D}_2 are as defined in the previous section and are unique to a given region. (The class-specific estimated values for all parameters were used when simulating data.) $\mathbf{Q} = \mathbf{Q}_1 + \mathbf{Q}_2$ was simulated for each subregion independently. The final simulated data were generated as

$$\mathbf{D}_1 \mathbf{Q} \quad (43)$$

which is multivariate normal with mean $\mathbf{0}$ and covariance $\mathbf{D}_1 [\mathbf{T} \otimes \mathbf{F} + \mathbf{D}_2] \mathbf{D}_1$ as defined in (40). To obtain the data with correct mean, the values $\hat{\mu}(t, l)$ were added for each (t, l) location. We denote a simulated dataset by $\tilde{\theta}(t, l)$.

In simulating data, the software we used could not allocate enough memory to construct a covariance structure for the largest (t, l) region (which would have been an $18,855 \times 18,855$ matrix). Therefore, we simulated values only for every fourth time step to reduce the size of the matrix \mathbf{T} corresponding to each region. Figure 8 shows four realizations from Class A sampled at every fourth time index and Figure 9 shows corresponding simulated data. Visual comparison of the simulated data to the realizations suggests that our modeling has been effective. In the balance of this section we consider a more quantitative assessment of this effectiveness.

Recall that $\tilde{S}(t, l)$ represents preprocessed/transformed data, and that there are 20 realizations for Classes A and B and 17 realizations for Class C. It is convenient in the following to let $\tilde{S}(A; t, l)$, $\tilde{S}(B; t, l)$, and $\tilde{S}(C; t, l)$ denote these datasets for the three specific classes. Let the j th temporally- and intensity-shifted realizations from Class A be denoted by

$$\tilde{\underline{S}}_j(A; t, l) = \tilde{S}_j(A; t + \hat{\eta}_j(A), l) - \hat{d}_j(A),$$

where $\hat{\eta}_j(A)$ and $\hat{d}_j(A)$ denote the estimated time- and intensity-correction terms for the j th realization of Class A, as described in (12) and (13), and let $\tilde{\underline{S}}_j(B; t, l)$ and $\tilde{\underline{S}}_j(C; t, l)$ be similarly defined for Classes B and C. For our analysis, twenty datasets were simulated from each of the three models. We denote the k th Class A simulated dataset by $\tilde{\theta}_k(A; t, l)$, the intensity-shifted version of $\tilde{\theta}_k(A; t, l)$ by

$$\tilde{\underline{\theta}}_k(A; t, l) = \tilde{\theta}_k(A; t, l) - \hat{d}_k(A),$$

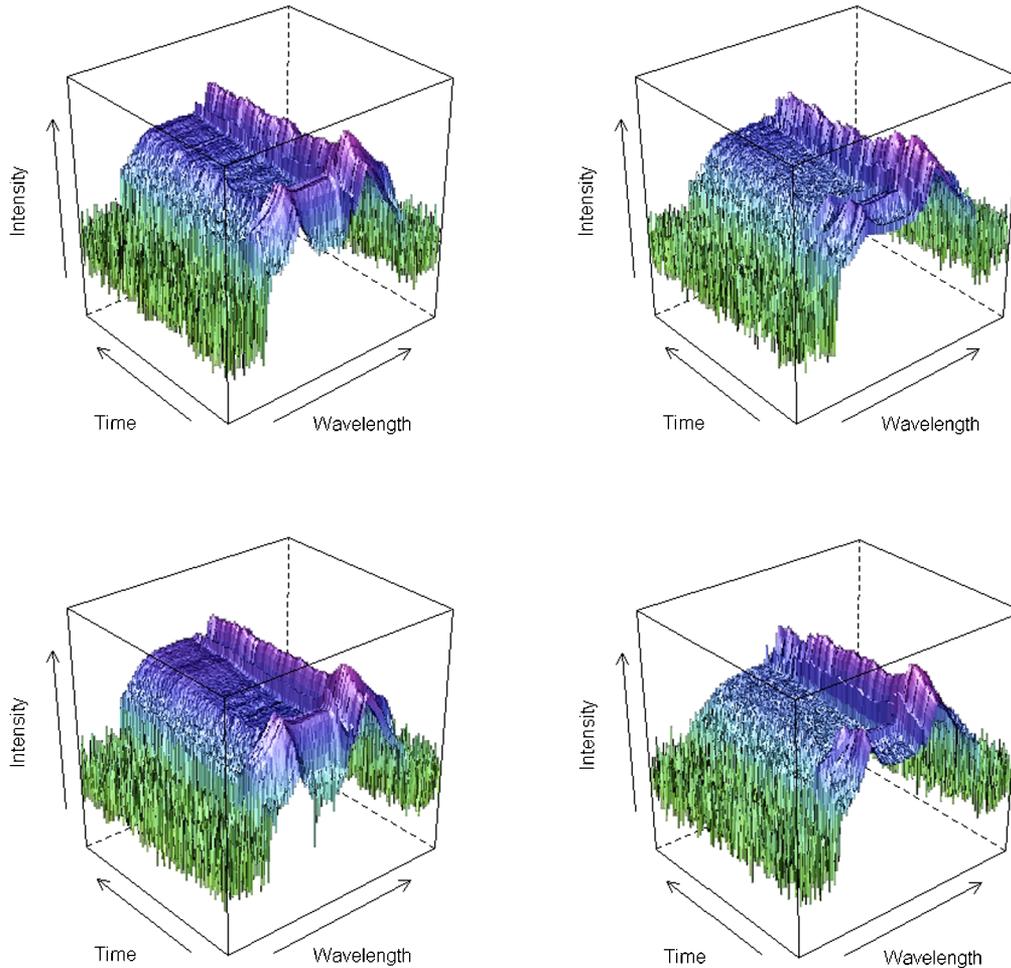


Figure 8. Four real observed datasets from Class A. The online version of this figure is in color.

and analogous intensity-shifted simulated datasets from Classes B and C by $\hat{\theta}_k(B; t, l)$ and $\hat{\theta}_k(C; t, l)$. Note that there is no need for temporal alignment with the simulated datasets; here \hat{d}_k is calculated as in (13), but with the known modeled peak intensity for the class used in place of $\hat{\eta}_k$.

Our interest is in looking at *differences* between datasets, both actual realizations and simulations. To that end, denote the absolute differences of values found, for example, in the j th real dataset from Class A and the k th simulated dataset from Class A as

$$\text{diff}(\tilde{S}_j(A), \tilde{\theta}_k(A); t, l) = |\tilde{S}_j(A; t, l) - \tilde{\theta}_k(A; t, l)|$$

and note that similar differences can be defined between any two datasets, real or simulated, and from the same class or not. For graphical displays, reduction is needed to focus attention on “typical” and “extreme” differences. For comparing Class A realizations to Class A simulations, we do this by computing the p th percentile of the entries in the $T \times L$ array $\text{diff}(\tilde{S}_j(A), \tilde{\theta}_k(A); t, l)$, denoting this scalar value by $\text{diff}^p(\tilde{S}_j(A), \tilde{\theta}_k(A))$. Finally, the empirical distribution of these values, over all comparisons of the 20 realizations and 20 simulations of Class A, is denoted by $F^p(\tilde{S}(A), \tilde{\theta}(A))$.

The top row of plots in Figures 10, 11, and 12 display empirical distribution functions of “within-class” differences:

$F^p(\tilde{S}(A), \tilde{S}(A))$, $F^p(\tilde{S}(A), \tilde{\theta}(A))$, $F^p(\tilde{\theta}(A), \tilde{\theta}(A))$, and analogous comparisons within Classes B and C; for $p = 0.05, 0.50$, and 0.95 , respectively. These can be thought of as portraying respectively how far apart a pair of waterfall plots/datasets tend to be in the regions where they are most alike, differ by a typical amount, and in the regions where they are most different. In all these plots, the empirical distribution of simulation-to-simulation differences is slightly shifted to the left of realization-to-realization differences and this is most apparent in the plot of the 5th percentiles as the range of values on the horizontal axis is small, amplifying the difference. These empirical simulation-to-simulation distributions are only slightly shifted to the left. This pattern was consistent across all the percentiles examined, and suggests that the simulated datasets are perhaps slightly smoother than the actual realizations. This is not surprising, as one would not expect a fitted model to capture every detail of the realizations. However, since the shift is not too extreme for any percentile, we feel the modeling has done an adequate job of describing the realizations.

As mentioned earlier, the Class A and Class B matches are very similar in chemical composition, because Class B matches are simply two Class A matches physically connected. Class C matches have a very different chemical composition than the other two classes of matches. The bottom row of

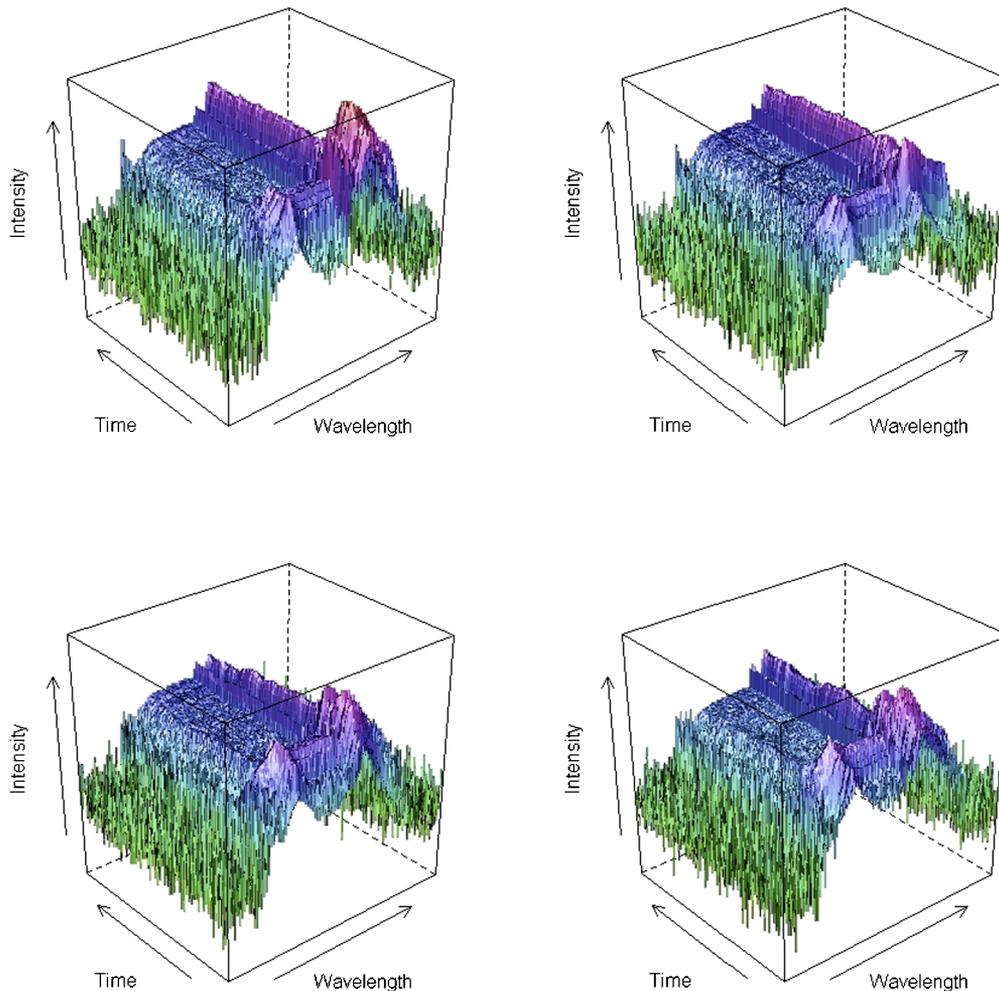


Figure 9. Four simulated datasets for Class A. The online version of this figure is in color.

plots in Figures 10, 11, and 12 display empirical distribution functions of “between-class” differences: $F^p(\tilde{S}(A), \tilde{S}(B))$, $F^p(\tilde{\theta}(A), \tilde{\theta}(B))$, $F^p(\tilde{S}(A), \tilde{\theta}(B))$, $F^p(\tilde{\theta}(A), \tilde{S}(B))$, and analogous comparisons between Classes A and C, and B and C; again for $p = 0.05, 0.50$, and 0.95 , respectively. The between-class comparisons for Class A and Class B take values very similar to those for the within-class comparisons for either Class A or Class B. This is what one would expect since these classes are very similar to one another in chemical composition. Further, the distribution functions for the between-class comparisons between Class C and either Class A or Class B place more weight on larger values. This is also to be expected since Class C is potentially very different in chemical composition from the other two classes. Again, the simulation-to-simulation comparisons take values slightly smaller than the realization-to-realization differences even if the comparison is made between classes, likely indicating slightly smoother simulations than realized datasets. The comparisons of one realization to one simulated dataset from two different classes are also slightly smaller in value than the realization-to-realization comparisons. Further, perhaps the most disparity between these distribution functions exists when comparisons are made with Class C.

The data models we have developed here are both subtle and complex. While they may not capture all real variability of the

processes, Figures 8–12 show that they can be used to generate data that reasonably represent characteristics common to each class, and within- and among-class variability.

6. CONCLUSIONS AND DISCUSSION

The work presented here focuses on modeling data collected using a spectral–temporal sensor that operates in the mid-wave infrared. Using relatively simple experimental data generated with three types of match burns, the modeling process described here includes correction for overall signal strength, temporal registration, separation of “white” measurement noise from physical (but random) “shot-to-shot” variation, and correlation of signal across wavelengths and time. While the details of these accommodations may be different for modeling a larger “library” of event types of greater interest, the *approach* outlined here provides an organized way of building the necessary model components. For example:

- The general form of the transform in (4) can be useful in other contexts where skewed and background-transformed data are used, but a threshold value other than 2 might be developed for data that are not integer-valued, or for which there is less skewness.

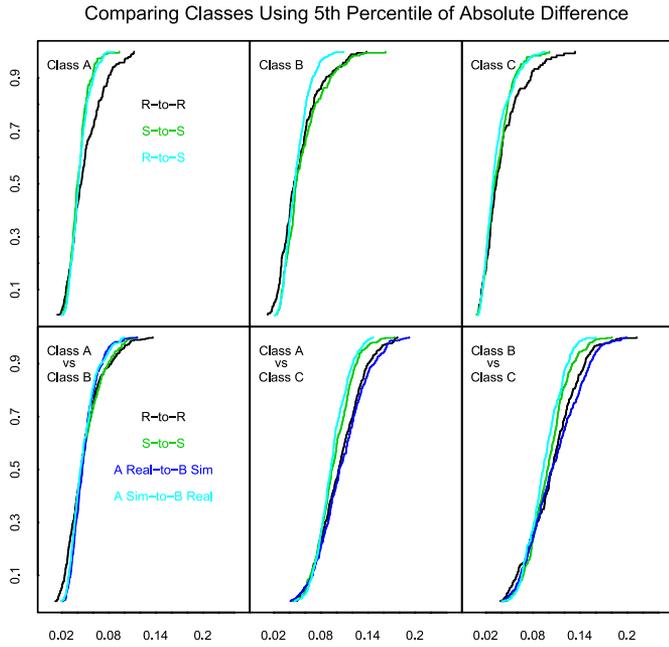


Figure 10. Within- and between-class comparisons using 5th percentile of absolute differences. The online version of this figure is in color.

- The idea of partitioning the domain with “independence walls,” and further into subregions (Figure 6) of homogeneous behavior can be useful in other contexts, but the specific physical details suggesting where the walls are placed and the size and number of subregions needed might vary.
- The model-free calculations described in Section 4.3 used to suggest appropriate temporal and spectral correlation forms should be generally useful, but in other cases might

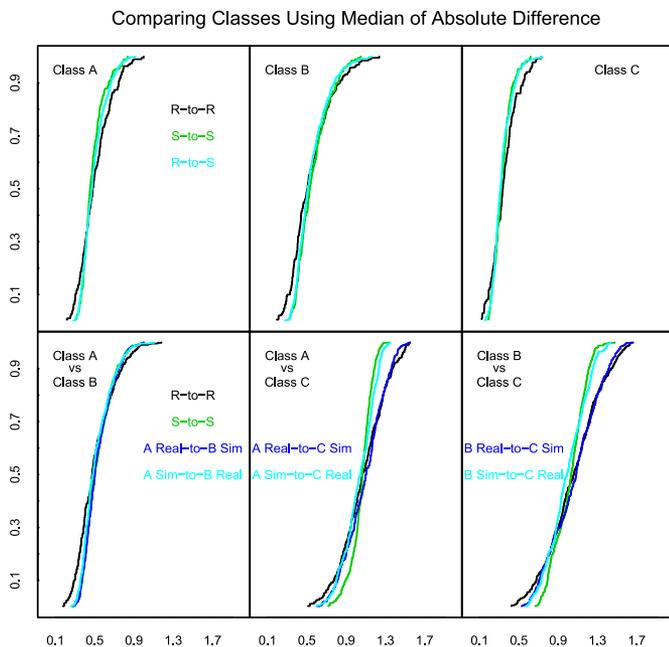


Figure 11. Within- and between-class comparisons using 50th percentile of absolute differences. The online version of this figure is in color.

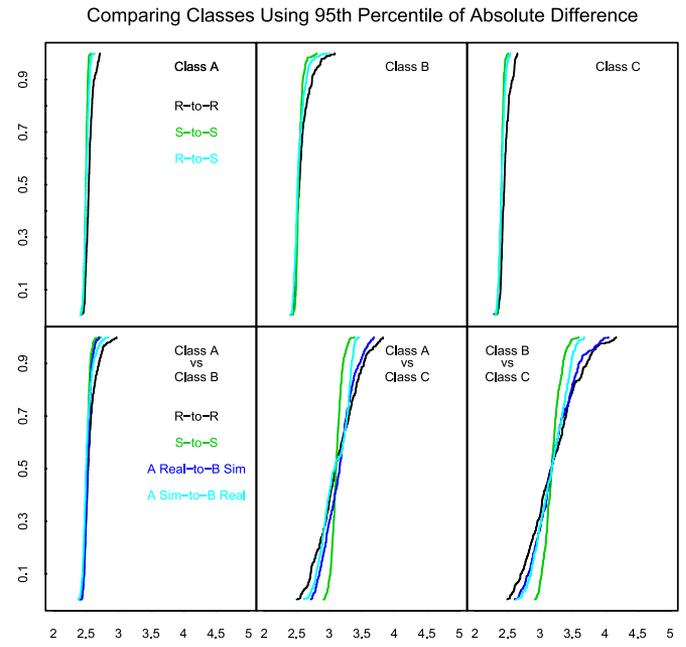


Figure 12. Within- and between-class comparisons using 95th percentile of absolute differences. The online version of this figure is in color.

lead to parametric forms other than those we selected [(27) and (32)].

“Next steps” in methodology development in our application area that motivated this work include data collection and modeling (again, through a process similar to that described here) for a larger collection of event types, and the development of useful identification and/or discrimination methods for those events based on the specific characteristics of the models derived.

The similarity between the observed data and data generated from the final models gives hope to development of algorithms that will allow for discrimination. Methods have been investigated for classification of events based on data reduction techniques (Gross et al. 2003; Orson, Bagby, and Perram 2003; Dills, Perram, and Gustafson 2004). To date no attempt has been made to simultaneously model the characteristic patterns of energetic events for both time and wavelength or to make use of all the information produced in both time and wavelength without some form of data reduction. The approach outlined here provides an organized way of building a model that will allow for unique discrimination algorithms that could be implemented in real time.

SUPPLEMENTARY MATERIAL

Appendix: The appendix is provided online as supplementary material which includes details for (20), plots of the mean and standard deviation surfaces, the δ^2 surface for each section within each class, and the parameter estimates for the correlations in time and wavelength. All plots provided here are also found online in color. (Appendix.pdf)

ACKNOWLEDGMENTS

We thank the Electromagnetic Technology Branch of the Sensors Directorate of Air Force Research Lab (AFRL/SNHI) for funding this project. The funding was administered by Solid State Scientific Corporation. The views expressed in this article are those of the authors and do not reflect the official policy or position of the United States Air Force, Department of Defense, or the United States Government.

[Received January 2009. Revised February 2011.]

REFERENCES

- Cressie, N. A. C. (1993), *Statistics for Spatial Data* (rev. ed.), New York: Wiley. [188]
- Deming, R., Higbee, S., Dwyer, D., Welser, M., Perlovsky, L., and Pellegrini, P. (2006), "Adaptive Estimation for Spectral-Temporal Characterization of Energetic Transient Events," in *World Congress on Computational Intelligence (WCCI)*, Vancouver, Canada. [183]
- Dills, A. M., Perram, G. P., and Gustafson, S. C. (2004), "Detonation Discrimination Techniques Using a Near-Infrared Focal Plane Array Camera," *Proceedings of SPIE—The International Society for Optical Engineering*, 5431, 77–86. [194]
- Gross, K. C., Dills, A. M., Perram, G. P., and Tuttle, R. (2003), "Phenomenology of Exploding Ordnance Using Spectrally and Temporally Resolved Infrared Emissions," *Proceedings of SPIE—The International Society for Optical Engineering*, 5075, 217–227. [194]
- Mooney, J. M., Vickers, V. E., An, M., and Brodzik, A. K. (1997), "High-Throughput Hyperspectral Infrared Camera," *Journal of the Optical Society of America A*, 14, 2951–2961. [184]
- Orson, J. A., Bagby, W. F., and Perram, G. P. (2003), "Infrared Signatures From Bomb Detonations," *Infrared Physics & Technology*, 44, 101–107. [194]
- Pellegrini, P. W., and Ewing, W. S. (2004), "A Novel Electro-Optical Sensor for the Find and Track Missions," available at <http://www.afrlhorizons.com/Briefs/Jun04/SN0309.html> (accessed 01 June 2006). [183,184]
- Schabenberger, O., and Gotway, C. A. (2005), *Statistical Methods for Spatial Data Analysis*, Boca Raton, FL: Chapman and Hall/CRC. [190]
- Weeks, M. M., Murguia, J. E., Mooney, J. M., Nelson, R. J., and Ewing, W. S. (2004), "Pseudo-Imaging the Spectral-Temporal Evolution of Energetic Events," in *Proceeding of the SPIE 26th International Congress on High-Speed Photography and Photonics*, Alexandria, VA. [184]