Multivariate methods to explore system sensitivities for hyperspectral subpixel target detection

Chase Cañas[a], John P. Kerekes[a], and Scott D. Brown[a]

[a]Rochester Institute of Technology, Center for Imaging Science, 1 Lomb Memorial Dr., Rochester, NY 14623

ABSTRACT

To explore system sensitivities in hyperspectral subpixel target detection, multivariate methods are applied to output detection metrics generated from a statistical target detection model. The Forecasting and Analysis of Spectroradiometric System Performance (FASSP) statistical model generates probabilities of detection (P_D) and false alarm (P_{FA}) using spectral libraries of target and background materials. This allows for the computation of the area under the receiver operating characteristic curve (AUC). To explore sensitivities within elements (e.g. scene, atmosphere, sensor) of the remote sensing system, ensembles of model-based scenarios are generated using combinations of the aerosol visibility, solar angle, and sensor viewing angle. Output detection metrics (P_D, AUC) from these scenarios were cached into a high-dimensional tensor, before utilizing multivariate methods (e.g. interpolation and regression) to explore sensitivities and correlations between system variables and detection. Inferences on limitations of detection within the system are drawn from multivariate contour regions which characterize joint parametric parameters required to exceed a desired threshold of detection. The outlined methods aim to provide an initial framework to investigate both specific and generalizable limitations of detection across various scenes (e.g. rural, urban, maritime, and desert), environmental conditions (e.g. solar angle, haze, clouds), sensor characteristics (e.g. noise, viewing angle) and processing configurations (e.g. feature selection, detector algorithm).

Keywords: spectral imagery, remote sensing, rare target detection, system limitations, novel targets, multivariable analysis, statistical modeling

1. INTRODUCTION

Hyperspectral subpixel target detection often applies statistical algorithms to identify subpixel objects of interest by exploiting unique spectral characteristics of materials. Model approaches offer user control over independent system variables, enabling the exploration of sensitivities in subpixel detection. In contrast, empirical methods provide real observations but face challenges due to limited accessibility of datasets containing diverse parameters related to the scene, atmosphere, and sensor. Additionally, processing large volumes of high-dimensional empirical data poses computational challenges. As a result, empirically validated models provide an optimal approach to explore systems sensitivities in hyperspectral data.

1.1 Multivariate Statistics

Multivariate statistics involves the analysis of multiple independent and/or dependent variables to explore relationships within complex, high-dimensional data. Independent variables can be categorical or continuous. For instance, a categorical variable might represent a generalized atmospheric model (e.g. rural, urban, maritime, desert). In contrast, a continuous variable is the aerosol visibility or meteorological range (e.g. 2 km, 4 km, 8 km, 16 km, 32 km). Multivariate methods (e.g. interpolation, regression) will be applied to system variables of the remote sensing chain to investigate relative sensitives on detection.

Further author information: (Send correspondence to C.C.)
C.C.: E-mail: cc1903@rit.edu

Algorithms, Technologies, and Applications for Multispectral and Hyperspectral Imaging XXX, edited by Miguel Velez-Reyes, David W. Messinger, Proc. of SPIE Vol. 13031, 1303104 · © 2024 SPIE · 0277-786X · doi: 10.1117/12.3013806

1.2 Forecasting and Analysis of Spectroradiometric System Performance

The Forecasting and Analysis of Spectroradiometric System Performance (FASSP) is a statistical tool used to predict the performance (P_D, P_{FA}) of a subpixel target detection scenario.[1] The model considers systems-level input parameters from the scene, atmosphere, sensor and post-processing stages of the imaging chain. Statistically, 1st and 2nd order statistics (mean and covariances) of target and background spectral reflectances which comprise the scene are transformed through the imaging chain. Spectral radiances are generated using MODTRAN, which are then transformed into digital signals using a physics-based noise sensor model. Subsequently, an atmospheric compensation algorithm is employed to derive transformed reflectance statistics. Finally, a detector algorithm (e.g. matched filter) is applied to produce the output detection values.

The FASSP model underwent empirical validation through the inclusion of novel subpixel targets in a hyperspectral data collection conducted in Rochester, NY.[2] The design of these lattice-based targets was further enhanced and incorporated into a recent data collection, showcasing their capability to capture abundant subpixel samples which consist of constant target fill percentages.[3] Subsequently, spectral libraries were generated from this data collection, supplying FASSP with input spectral reflectances for the current investigation. A recent reconfiguration of FASSP has been in ongoing development, written in a C++, with aim to function as a plugin tool for the Digital Imaging and Remote Sensing Image Generation Model (DIRSIG).[4] The reconfiguration has reduced computational run times and has enabled the ability to conduct the current research in multivariate trade studies.

2. METHODOLOGY

To explore variable impacts within the remote sensing system on subpixel target detection, multivariate analyses were conducted on detection outputs (P_D, P_{FA}) from the FASSP model, which were generated from large numbers of parameteric combinations of three continuous variables (aerosol visibility, solar zenith, and viewing zenith) and one categorical variable (target in background). The "target in background" variable refers to the background class which is mixed within the subpixel target. Reflectance statistics imported into FASSP were derived from atmospherically compensated hyperspectral data (UAV).[3] The baseline scene considers a 20 percent subpixel target (gray felt) surrounded by three background classes (grass, gravel and soil), with materials modeled as Lambertian reflectors. The model scene in FASSP was comprised of equal proportions of the 3 background classes. This was selected to ensure any inferences on background classes which "drive" the impedance of detection are not associated to the proportion or weight of samples in the scene. The spectral range for the investigation was constrained to the visible wavelengths (0.4 - 0.7 µm), to avoid erroneous band region of atmospheric absorption which may distort the conclusions between impacts of atmospheric variables (aerosol visibility and solar zenith).

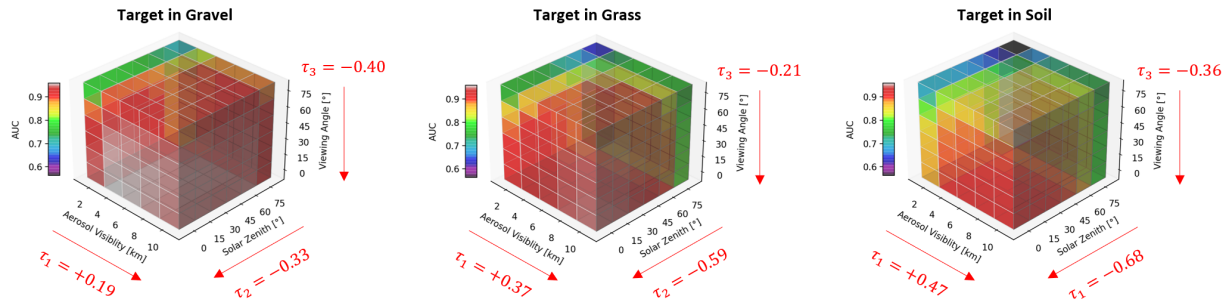


Figure 1: AUC detection outputs after iterative model runs of FASSP across parametric combinations of the aerosol visibility, solar zenith, and sensor viewing variables. The Kendall tau correlation coefficients, τ, represent the direction and magnitude between pairs of the independent and dependent (AUC) variables.

The parameters considered for each variable are as follows: aerosol visibility spans 2 to 10 km with a 2 km step, solar zenith spans 0 to 75 degrees with a 15 degree step, and sensor zenith spans 0 to 75 degrees with a 15 degree step. The output parameters from FASSP are the P_D and constant P_{FA} rates. To generalize all instances of this metric, the area under the ROC curve (AUC) was also considered. Given parametric combinations of

the 4 independent variables, unique detection outputs were generated and cached into a multidimensional tensor for subsequent multivariate analyses. Projections of the 4-dimensional tensor comprised of unique AUC output detection predictions for the defined scenario are shown in Fig. 1. It was determined the relationship between system variables (aerosol visibility, solar zenith, sensor zenith) and detection is non-linear. Therefore, a non-linear correlation coefficient (Kendall tau) was selected to assess the magnitude and direction of correlation between variables. Kendall tau, τ, is robust to outliers and the variant-b is effective in managing tie points, which are prevalent in our data.[5] We would like to acknowledge the P_D and specific P_{FA} rates were investigated, however, we concluded for this particular scenario, the direction of correlations, τ, between variables were often near-equivalent across all false alarm rates when compared with the AUC metric. For brevity, forthcoming analyses will consider only the AUC metric. It is advised more complex investigations should consider analysis of varying P_D and P_{FA} rates.

2.1 Multivariate Interpolation of Output Detection Tensors

Tensor product spline interpolation was implemented to resample detection outputs (AUC) and produce high-dimensional response surfaces. The method entails the product of uni-variate spline functions (piecewise polynomials) along each dimension of the grid. The generalized formalism is

$$F(x, y, z) = \sum_i \sum_j \sum_k \ldots c_{i,j,k} \cdot B_x(x_i) \cdot B_y(y_i) \cdot B_z(z_i) \cdot \ldots \tag{1}$$

where $B_x(x_i), B_y(y_i), B_z(z_i)$ are cubic spline basis functions and $c_{i,j,k}$ is a multidimensional array of coefficients for the piecewise cubic polynomials. The interpolated data improves visualizations and multivariate analyses, while reducing the computational time to run parametric combinations of FASSP. It is worth noting that additional independent variables induce an exponential increase in computational time.

When 2-dimensional "slices" are extracted from the 3-dimensional interpolated tensors, bi-variate contour and gradient plots enable assessment of variable sensitivities across the response surfaces, as shown in Fig. 2. More specifically, these visualizations provide detailed insight on detection sensitivities for joint combinations of system parameters, and assist in identifying regions of optimal performance. Similar multivariate techniques are often observed in response surface methodology.[6]

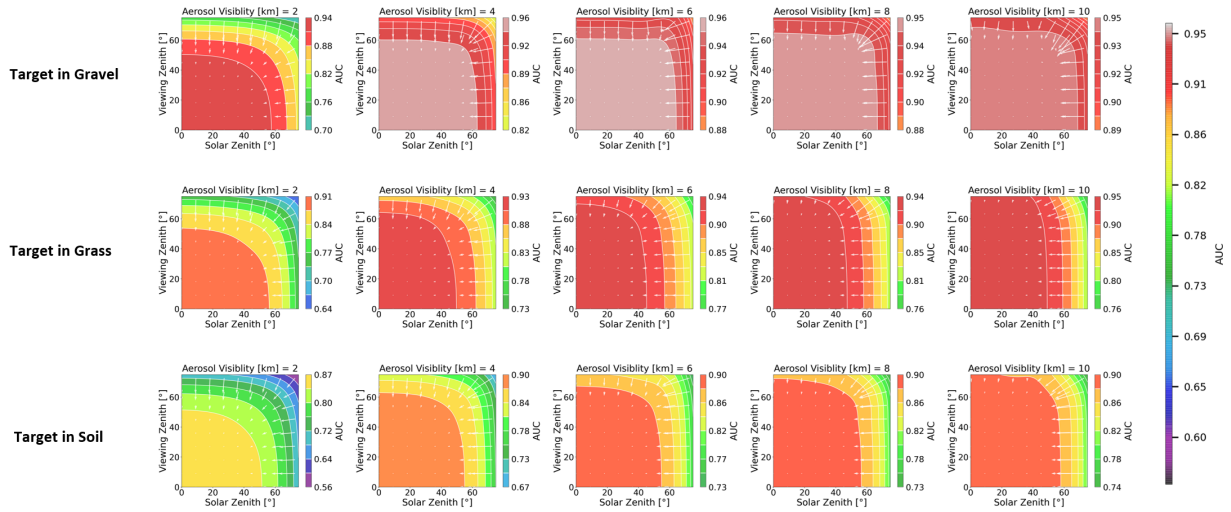


Figure 2: Contour and gradient plots of bi-variate projections of the 4-dimensional AUC interpolated tensor.

A contour region, denoted as $\mathcal{R}$, where detection exceeds a desired threshold, γ, can be characterized for sensitivity analyses. To demonstrate the methodology, we consider the prior example involving the contour image "target in grass" with aerosol visibility = 10 km, now highlighted on the left of Fig. 3. If we consider a test criteria where the AUC $\gtrapprox 0.94$, the corresponding bi-variate contour region is depicted on the right of Fig. 3.

After identifying the samples across the contour boundary, a regression model was fitted to derive an analytical inequality that characterizes the underlying contour region, $\mathcal{R}$. The inequality model is $x_2 < f(x_1)$ for $x_1 \in \mathcal{R}$, where in this case x_1 is the solar zenith and x_2 is the viewing zenith. This model may be used for predictive purposes to assess parametric requirements needed to achieve a desired level of detection.

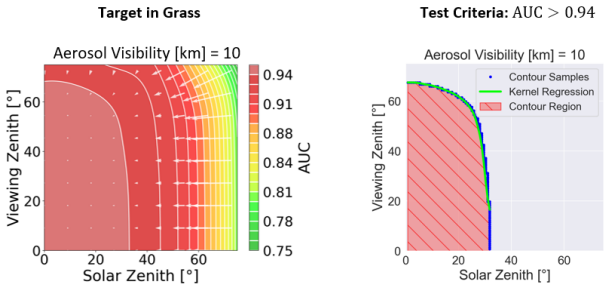
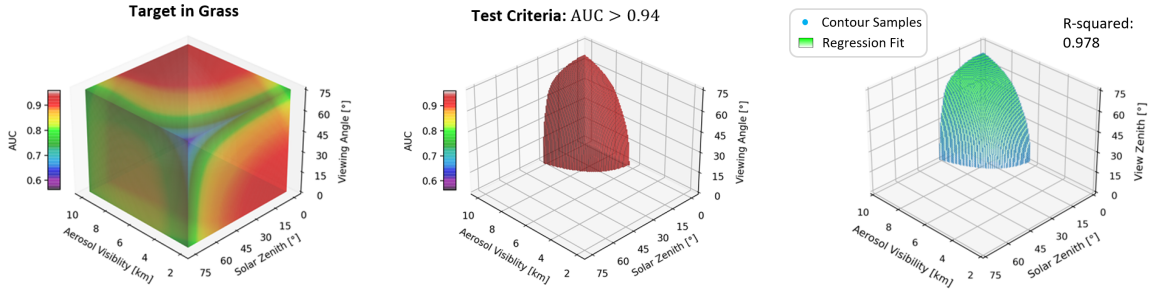


Figure 3: Contours of an AUC interpolated surface (left) and contour region, $\mathcal{R}$, where AUC $\gtrapprox 0.94$ (right).

3. RESULTS

The methodology for bi-variate contours can be applied to higher dimensional tensors. Effectively, contour surfaces become the multivariate extension to counter lines, which partition contour regions, $\mathcal{R}$, of exceeding detection. To demonstrate, the "target in grass" interpolated tensor is shown in Fig. 4a. Considering the same detection exceedence criteria as prior (AUC $\gtrapprox 0.94$), the bi-variate contour region, in Fig. 3, can be effectively extended into three dimensions around the aerosol visibility parameter of 10 km. The tri-variate contour region, $\mathcal{R}$, which satisfies the test criteria is shown in Fig. 4b. Boundary samples across the tri-variate contour surface are represented as blue points in Fig. 4c, along with a corresponding non-parametric regression fit.

Non-parametric regressions are effective at fitting complex multivariate non-linear surfaces. Examples include a kernel regression, Gaussian process regression, or random forest regression. In the demonstration, a Gaussian kernel regression was fitted to the tri-variate contour surface with an R-squared goodness-of-fit of 0.978, as shown in Fig. 4c. An algorithm was developed to identify interpolated samples across the contour, to serve as input data for the regression. First, the algorithm ensures contour samples are interconnected using a "connected component labeling" routine. If not interconnected, independent regressions are needed during fitting. Second, contour samples are indexed using a boundary extraction routine, while also discarding samples located on the outer perimeter of the grid. The regression is then applied to the indexed samples across the boundary contour surface within the grid.



(a) (b) (c)

Figure 4: (a) Interpolated AUC tensor for the "target in grass" variable, (b) contour region where AUC $\gtrapprox 0.94$, and (c) non-parametric kernel regression fitted to samples across the AUC $= 0.94$ contour surface.

The regression inequality model for system parameters x_1, x_2, x_3 characterizes the test criteria (AUC $> \gamma$) region and provides predictions across the parametric ranges they were trained. A generalized model is

$$x_n < f(x_1, x_2, \ldots, x_{n-1}) \qquad \text{for} \quad x_1, x_2, \cdots x_{n-1} \in \mathcal{R} \tag{2}$$

where x_i are independent variables of the system and $\mathcal{R}$ is the multivariate contour region. Projecting the boundary surface onto an independent variable plane enables assessment of bi-variate parametric combinations required to exceed a desired threshold of detection. Alternatively, tri-variate assessments are conducted using the visualizations shown in Fig. 5. For example, given a solar zenith of 15 degrees, aerosol visibility may range between approximately 9.0 and 10.0 km and viewing zenith between 0 to 70 degrees, to achieve an AUC $\gtrapprox 0.94$.

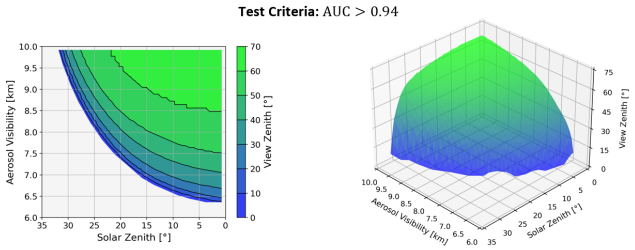


Figure 5: (a) 2D projection and (b) 3D projection of the kernel regression across the AUC = 0.94 contour surface, for the "target in grass" variable. Parametric combinations which reside below the surface $x_3 < f(x_1, x_2)$ satisfy the test criteria.

4. DISCUSSION

General inferences from the"target in grass" scenario co-agree with intuition that larger solar zenith angles and sensor zenith angles lead to lower detection rates of targets on the ground. This is attributed to an increase in path radiance (or atmospheric scattering) reaching the sensor for larger solar zenith angles. In addition, the increase in path length (or duration of atmospheric interaction) of radiance between the sensor and target for larger sensor zenith angles is a contributing factor.

Assessing the tri-variate contour surface in Fig. 5, the shape is bounded by minimum and maximum ranges of [6.4, 10] km for aerosol visibility, [0, 31.8] degrees for solar zenith, and [0, 66.7] degrees for sensor zenith. Given this observation, in addition to the correlation coefficients $\tau_2 = -0.59$ (solar) and $\tau_3 = -0.21$ (view) in Fig. 1, we can infer the solar angle is a more sensitive variable impacting detection for the "target in grass" parameter. This conclusion is not generalizable, however, as the difference between τ_2 and τ_3 changes direction in comparison to the "target in gravel" parameter. Therefore, future investigations are needed across large ensembles of alternate scenarios. The complexity of high-dimensional spectral signatures and covariances between of the target, background, and atmosphere emphasizes the value of systems-based modeling as a tool to identify trends in hyperspectral subpixel target detection, which may otherwise not possible be with empirical data.

5. CONCLUSION

We have developed a framework to explore system-level sensitivities using multivariate methods, after generating output detection predictions (AUC) from a hyperspectral subpixel target detection model. To demonstrate the multivariate approach, we considered parametric combinations of three continuous variables (aerosol visibility, solar zenith, and sensor zenith) and one categorical variable (target in background). Tensor product spline interpolation was used to resample and smooth the output detection AUC tensors. To investigate a parametric region, $\mathcal{R}$, where detection exceeds a desired threshold (e.g. $\gamma = 0.94$), an inequality model was defined for a multivariate contour surface characterized by non-parametric kernel regression.

Future analyses will encompass a broader range of scenarios, incorporating additional independent variables x_i $(i > 4)$ and materials classes within the scene. We will also consider the impacts of varying the proportions of background materials which comprise the scene. In addition, alternate machine learning models will be considered to infer generalizable information of system sensitivities across large ensembles of scenarios. The current research supports efforts to derive insights on fundamental sensitivities and limitations of detection.

ACKNOWLEDGMENTS

This research was supported by an academic grant from the National Geospatial-Intelligence Agency (Award No. HM04762110005, Project Title: Fundamental Research on Detection and Classification Limits in Spectral Imagery.) Any opinions, findings and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of NGA, DoD, or the US government. Approved for public release, NGA-U-2024-00613.

REFERENCES

[1] Kerekes, J. and Baum, J., "Spectral imaging system analytical model for subpixel object detection," *IEEE Transactions on Geoscience and Remote Sensing* **40**, 1088–1101 (May 2002).

[2] Cañas, C., Kerekes, J. P., Ientilucci, E. J., and Brown, S. D., "Empirical validation of a hyperspectral systems model for subpixel target detection using data from a new UAS field collection," in [*Imaging Spectrometry XXV: Applications, Sensors, and Processing*], **12235**, 98–103, SPIE (Sept. 2022).

[3] Cañas, C. and Kerekes, J. P., "Design and Demonstration of a Lattice-Based Target for Hyperspectral Subpixel Target Detection Experiments," *IEEE Transactions on Geoscience and Remote Sensing* **62**, 1–10 (2024).

[4] Goodenough, A. A. and Brown, S. D., "Dirsig5: Next-generation remote sensing data and image simulation framework," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* **10**(11), 4818–4833 (2017).

[5] Hollander, M. and Wolfe, D. A., "Nonparametric statistical methods," in [*Nonparametric statistical methods*], 503–503 (1973).

[6] Myers, R. H., Montgomery, D. C., and Anderson-Cook, C. M., [*Response Surface Methodology: Process and Product Optimization Using Designed Experiments*], John Wiley & Sons (Jan. 2016).

[7] Kendall, M. G., "A New Measure of Rank Correlation," *Biometrika* **30**(1/2), 81–93 (1938).