Intelligent instruments: discovering how to turn spectral data into information

W.L. Buntine*(RIACS) and Tarang Patel (RECOM)

Computational Sciences Division
Mail Stop 269-2
NASA Ames Research Center
Moffett Field, CA, 94035-1000

Abstract

The Optical Plume Anomaly Detection (OPAD) program at the NASA Marshall Space Flight Center is using plume spectroscopy for the diagnosis of the Space Shuttle Main Engines. A challenging part of this program is matching high resolution spectral data with a physicist's model of spectroscopy to produce estimates of metallic erosion through the plume. This paper describes the discovery process used in doing this. The physicist's model had to be debugged in order to discover the various instrument characteristics, discover critical elements of the data, and in general perform exploratory analysis to understand the instrument and the data it produces. This model gives us strong prior knowledge, however, this needs to be incorporated with care. We had to use a range of statistical techniques in our analysis, including one-dimensional super-resolution to determine the instrument response function. The paper concludes with a discussion of the role of discovery in building intelligent instruments that turn real-time data into useful information.

Introduction

Both Mission Control at Johnson Space Center for the Space Shuttle, and the Propulsion Laboratory operating the Technology Test Bed (TTB) stand at Marshall Space Flight Center (MSFC) currently have various instruments and sensors whose output is under the watchful eye of many operators. The TTB stand at MSFC has an experimental main engine for the Space Shuttle (SSME) used to test different configurations of the engine. The interpretation of these instruments may be as simple as "if the pressure goes below 25 p.s.i. then the feeder is failing so abort the test." Modern instruments such as vibration sensors, and the high resolution spectrometer considered in this paper are such that interpretation needs more than just care. Instead they require sophisticated post-processing of the data.

Intelligent instruments are an important facility in modern engineering and health management practice.

---

*Current address: Heuristicrats Research, Inc., 1678 Shattuck Avenue, Suite 310, Berkeley, CA 94709-1631, wray@Heuristicrat.COM.

This paper looks at the development of one intelligent instrument, the RAS spectrometer in use at the TTB stand at the Marshall Space Flight Center (Buntine et al. 1993) as part of the OPAD (Optical Plume Anomaly Detection) program. This development has been educational for the instrument designers, the physicists, and the data analysts (ourselves) as we discover the inherent interaction required between our disciplines in order to get the instrument and its software working. At the heart of the development, from the data analyst's perspective, is a discovery process: discovering what parts of the instrument were not performing to exact specification, discovering how the physicist's model and the data interpretation interact, and discovering how to make use of the physicist's model efficiently during analysis. Of course, what we are trying to determine is what kind of information can be got from the instrument, and how is that information to be extracted from the data. However, in order to do that, we have a discovery process to perform.

Most knowledge discovery to date has looked at large commercial data bases (transactions, expenses, finances, employee information), scientific data, sociological or medical data. This paper presents a new application of knowledge discovery concerned with the analysis of complex instruments. Here, a large part of the discovery process is concerned with understanding the way the instrument operates, and understanding how to interpret, use, and debug the physicist's model of the instrument.

This paper first introduces the RAS instrument, a spectrometer, and describes the context of plume spectroscopy in which it is used. Second, the specific tasks required of our analysis are discussed. Our immediate goal of the research is not to produce a real-time diagnostic system, but rather to understand how to process data from the instrument. An appropriate real-time diagnostic system can then be designed once we know what information the instrument is capable of producing. The remainder of the paper gives an account of our discovery experiences on this application. The paper concludes with a general discussion of intelligent instruments and their analysis.

Plume spectroscopy

Figure 1(Left) shows the engine firing during a test. The Mach disk is defined by the shock wave and seen as a bright thin layer at the very bottom. A high resolution spectrometer is positioned to have this disk in its field of view, although this is not known for sure. Unfortunately, the spectrometer also keeps the horizon in the field of view so atmospheric charges interfere with observations. However, correction for this is attempted by subtracting an average of a few scans prior to any test firing. A spectrum is generated every 0.5 seconds during a test which may last up to 250 seconds. The clock is possibly the only aspect of the instrument that is precisely calibrated. The RAS instrument is configured as 2×2048 pixels covering near UV to visible part of the spectrum (2850Å to 7880Å) on a grid approximately 1.1Å per pixel (1 Å = 1 Ångstrom = 0.1nm). The exact measurement at each pixel is a convolution
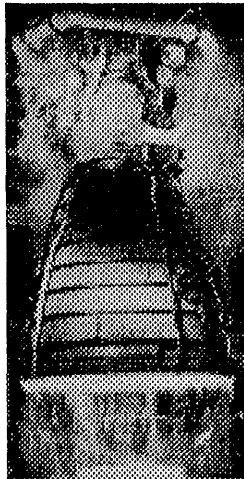

Figure 1: A view of the test engine showing the Mach disk.

of the instrument *response function* with the "true" light intensity in that region of the spectrum. The instrument response function is approximately a Gaussian with width at half its height of about 4 Å. For the specified optical elements, the wavelength *grid* is fixed from test to test, yet the actual grid is an unknown. So the pixels have their relative location fixed in the grid, but a single location parameter for the grid as a whole needs to be determined empirically for each test. A suitable location parameter is the exact center for the first bin, referred to as the *absolute registration*. Since measurements are made at approximately 1.1 Å intervals and the response function is wider, there is some overlap between measurements at neighboring bins. We return to these issues later.

Figure 2 shows how the physicist on our team, Tim Wallace, sees the problem of interpreting the data. Figure 2(a) shows a measured spectrum, referred to as a single *scan*. This is taken from TTB test number
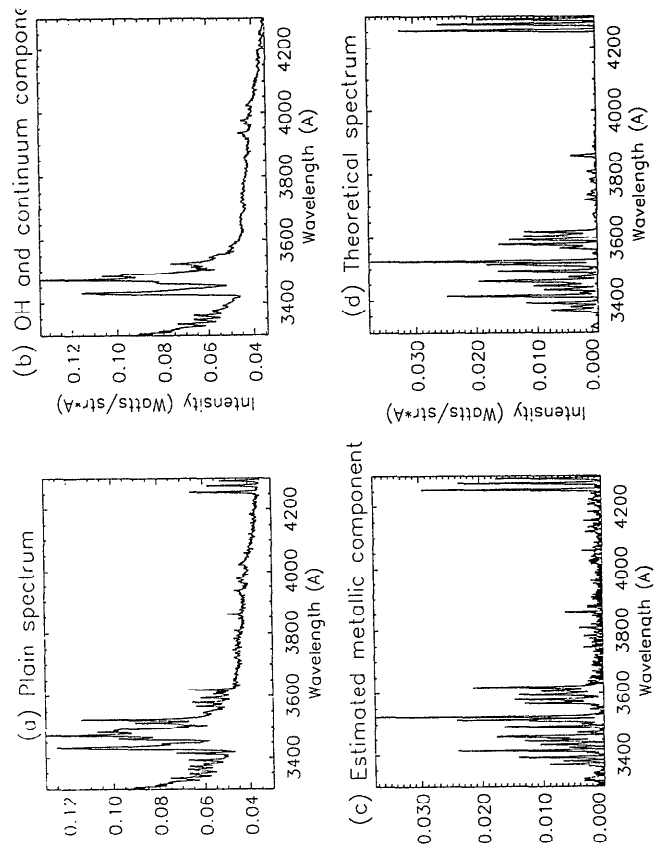

Figure 2: The four stages of an ideal analysis.

55, at 144.5 seconds, at which a significant metallic erosion event occurred. A single spectrum consists of an OH and its contribution to the continuum component, and it may include a metallic component due to some form of erosion in the motor. A first order approximation is that these two components, OH and metallic, add. The SSME is a hydrogen burning motor, and any level of metal in the spectrum indicates erosion of the motor, so most spectra are hoped to be metal free. Figure 2(b) shows the OH and continuum component of this spectrum, estimated by a process we describe later. The physics describing the OH and continuum spectrum is too complex for it to be modeled theoretically, instead it is up to us to determine it empirically. This OH component has, during equilibrium, approximately three degrees of freedom due to effects such as temperature, pressure, and fuel mixture, so it is not constant. Figure 2(c) shows a spectrum after the estimated OH and continuum component is subtracted. Notice, this estimated metallic component needs to be worked afresh for each scan.

Figure 2(d) shows a near match from the physicist's model. A new, advanced model of spectroscopy including line broadening, can be used to predict what metallic component should be expected, for given metallic inputs in parts per million. This model is implemented by our physicist Wallace as a program, SPECTRA6,

in a commercial matrix language called IDL from Research Systems, Incorporated (a relative of PV-Wave, with some similarities to Matlab). As well as metallic inputs, SPECTRA6 requires temperature which is only known approximately, so must be estimated as well. These are shown below:

```
Chromium:
    number density = 0.13e+09,
    broadening = 0.1
Nickel:
    number density = 0.40e+10,
    broadening = 0.1
Temperature = 2855 degrees Kelvin
```

The physicist's model takes a few minutes to run. By varying the metallic inputs and the temperature, we hope to create a spectrum that will be a good match with the measured metallic component from the RAS data. However, to find a good match can take upwards of an hour as the program iterates. It is desirable to run this matching process within the time frame of the data collection, once every half second.

Tasks for interpreting plume spectroscopy data

We can now summarize the tasks we must complete for plume spectroscopy to provide useful quantitative information about the metallic erosion rate in the plume. Bear in mind this is all in hindsight. Some of these pieces only came out as we discovered problems in trying to match the theoretical spectrum with the real data.

1. First, we wish to understand how to use the spectral model embodied in SPECTRA6 to turn spectral data into information about metallic erosion rates.

   (a) The fixed spectrometer characteristics need to be determined. These include the instrument response function and the relative registration (the grid). We only discovered these were a problem recently. Also, for each test, the absolute registration will be different.

   (b) The OH and continuum component of each spectrum needs to be estimated. This can only be done empirically.

   (c) The various inputs to the spectral program SPECTRA6 needs to be found that would provide a close match with the measured metallic component of the spectrum. Error bars on these estimated inputs also need to be computed.

2. Second, we need an approximate anomaly detector that, instead of returning estimated metallic erosion rates, would indicate that an anomalous level of a metal existed. Ideally, such levels need to be sensitive to the recognized engine stages such as startup and shutdown.

How would these information be used? When testing an engine, this kind of information, if generated in real time, can be used to provide warnings of serious metal erosion in the motor, for instance, indicating the test should be shut down. This information can also be analysed after a test to determine if a particular alloy was eroding in the motor, and therefore whether the engine needs to be stripped down. Such maintenance is expensive so reliable indication of erosion is required. Finally, this kind of health monitoring could be performed during flight. However, this would require a special purpose instrument be built. How sensitive should this instrument be? How should analysis be done? What kinds of erosion estimate could be obtained? Only by understanding the current RAS instrument and its processing, from data to metal erosion rates, can we hope to answer these questions. Our task then is to understand how to interpret the RAS data both so that more sophisticated health monitoring can be performed, and so that future plume spectrometers can be designed for different tasks such as in-flight health monitoring.

Our approach

It is tempting in this paper to follow the time-honored academic practice of reporting the "cleansed" version of our progress, as viewed *post hoc*. Unfortunately, progress comes in fits, starts, and occasional reversals, as one or another unexpected problem arose. We will try to present some of these in our discussion, because, from hindsight, they might have been avoided by a team with more experience. We extract the essential points in the conclusion.

There are two possible avenues for using SPECTRA6 in the analysis of the data.

Full theory: Develop a clear model of the OH and continuum component, and then perform the analysis described in Figure 2 directly. Since SPECTRA6 takes many minutes to run once, this kind of matching process could take at least one hour per spectrum.

Feature design: Develop a number of noise and error resilient features from the data, that can operate without precise quantification of the OH and continuum component, Use these features to predict metal erosion rates. The prediction is done by training, for instance with non-linear regression, from the generated SPECTRA6 data for which the features can be computed and the metal erosion rates are known. This approach could work in real time.

The first approach can be regarded as the theoretically clean approach—validate the theoretical model, and then invert the theoretical model to match the data. Another approach that is popular in the physics and image processing literature is deconvolution. While this would be possible, in our case it is not clear that it would produce useful information (apart from cleaning up the spectra). The spectra we have are complex, and

we wish to estimate the parameters such as temperature and metal number density, not the deconvolved spectra.

We began by attempting the full theory approach—better to validate the physicist's model first. Due to lack of space, this paper considers just this aspect of our study in detail. Our parallel progress on the feature design approach was an interesting study in expert interaction, so some comments are included in a later section.

Discovering the OH component of a spectrum

Each spectrum has 4000 points, and the OH component from physical chemistry, is expected to have a dimension of about 3 or 4. So we need to do a reduction of the spectral data down to some basic OH components. We used a variation of principle components (a standard unsupervised approach from pattern recognition (Rao 1965)) to perform this task. Unfortunately, this is not the right way to go. First, across different tests, the absolute registration is different, so that spectra from different tests do not have the same grid. Second, spectra vary dramatically in size and are always positive. In the first few seconds at start up, the magnitude is very small. So the Gaussian assumptions implicitly in the principle components methods are violated. Third, we do not not exactly which spectra are metal free, we had to eyeball the data to choose metal free spectra. We therefore need to apply a more sophisticated technique, which we are developing.

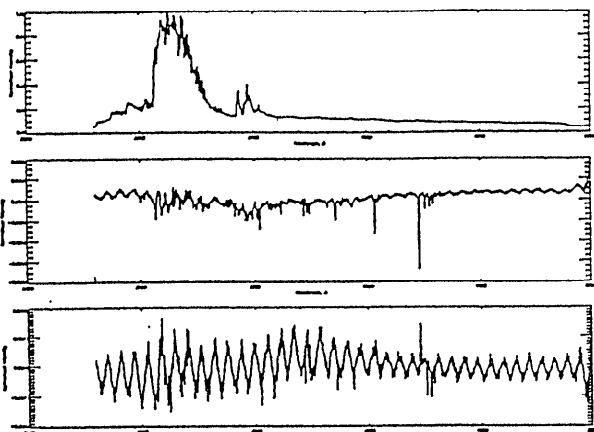One typical decomposition is demonstrated in Figure 3, taken from (Srivastava & Buntine 1995). This

Figure 3: Three extracted components for the OH and continuum.

decomposition has been used separately as a basis for predicting the engine rated power level from the spectrum in real-time. Notice in the second component there is some activity up around 4000 Å, to the right

of the plot. This is actually metal, indicating we chose spectra with some residue metal component. In the data the metal component was very small, but its contribution to the second component is large.

This data could not be used to "extract" OH and continuum components because the presence of the metal residue corrupted the match. We therefore attempted to discover some metal components as well. In this case, we generated a lot of metal spectrum using SPECTRA6 and generated components for the metal spectrum, a few for iron, a few for nickel, and so forth. We would then fit the OH and continuum component while concurrently extracting the metallic components, using an approach similar to (Hinton, Revow, & Dayan 1994) found to be successful in a simpler domain, digit recognition. While this approach showed promise, it turned out that the matches were out badly in some cases. After considerable work, we discovered that the theoretical instrument response function we were using for SPECTRA6 was a very poor fit to the "true" response function indicated by the data, and thus the fitting process would be badly skewed on some cases.

We learned two lessons from this:

- Carefully validate all aspects of your theoretical/computational model, even those aspects the physicist believes are correct. The truth ultimately lies with the data.

- Off the shelf data analysis methods (like principle components) never quite fit the task you have.

Rather than developing this more sophisticated OH and continuum model, we resorted (for now) to a simpler model based on interpolating from known metal free points. This simple model performs poorly in active regions of the OH and continuum spectrum, so is no good for metals like copper or silver which occur in the same region.

Matching the expert's SPECTRA6 model to the data

Our next task was to restore our confidence in the physicist's model by estimating the response function and the grid. Following our problems described above, we extracted some simple spectra and attempted to fit them. It was clear that the response function was lobsided and non Gaussian, and that the grid given to us by the instrument designers was wrong. Looking carefully at the data, and consultation with the physicist, made us believe that the instrument designer's specification for grid disagreed with the data by as much as 5 Å (0.5nm). We had previously discounted the anomalous peaks (which were out by 5 Å) as being water artifacts. This shows the importance of checking all anomalous results for an instrument carefully with all people concerned.

We extracted a set of 150 empirical peaks from the different spectra. This had to be done with care, because in many places peaks interact. In order to esti-

mate the response function one needs to look for clean, isolated metal lines, distinct metal lines as opposed to compound metal lines. We used super-resolution to estimate the response function. Again, there was no available data analysis software for this task. Each single peak, part of a spectrum, we extracted form the data, are measurements to indicate the shape of the response function. Each peak will be shifted in both the X and Y axis from (0,0), and will be scaled differently. If we know the scaling and location precisely, then the peak data corresponds to samples taken from the response function directly, so standard curve fitting can be used. Therefore, we implemented a program that repeatedly re-estimated the scaling and location for each peak given an estimate for the response function, and then re-estimated the response function (using curve fitting with 20-degree Legendre polynomials) using the estimated scaling and location for all the peaks. This algorithm is rather like the EM algorithm. The results for the RAS instrument is given in Figure 4. The data (indicated by dots) in Figure 4 gives
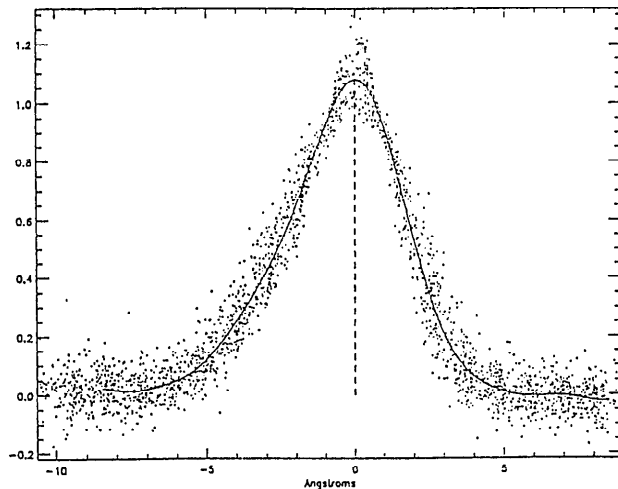

Figure 4: The empirical response function for RAS.

good support for our response fit. We also validated our process by running it for two other high resolution spectrometers in use at the TTB and showing the results to the physicist.

Discovering useful metal lines

How might we determine whether a metal is indicated by the spectrum? The physicists usually look for a few characteristic lines that are known to be indicative of the metal. Physicists tell us that when multiple metals appear in a spectra, the combined effect will be local. It is rare for two metals to emit at precisely the same place, however, if they emit nearby, for instance within one Å, then the peaks will be blended together by the instrument response function. Our first task is to discover those characteristic lines that indicate metals are present. The operator, Anita Cooper at

MSFC does this task routinely herself.

We talked with Cooper to discover the principles she uses. We then developed a visual discovery tool to perform this task automatically ourselves. We generated thousands of typical metal spectra using SPECTRA6. We then built a map indexed by wavelength to work out which metal lines appeared where. For several metals, different lines appear at different concentrations and temperatures, so the absence of a line does not necessarily indicate absence of the metal. We built a relative-minimum map from the data, which showed for each metal, what is the minimum relative size of a peak at each wavelength. i.e., every time a particular metal occurs, then at the given wavelength the peak must be at least this value, relative to the maximum peak for this metal. Likewise we built a relative maximum map. These maps are displayed as a complex plot with 3 axes, metal, wavelength, and relative size (in the range [0,1]) with the ability to zoom in on different regions. We then overlaid these to display "other" plots, which show, for instance for cobalt, the maximum size of the relative maximum map for all other metals except cobalt. A selected map for nickel and cobalt only is shown in Figure 5. First, the spectra for nickel are
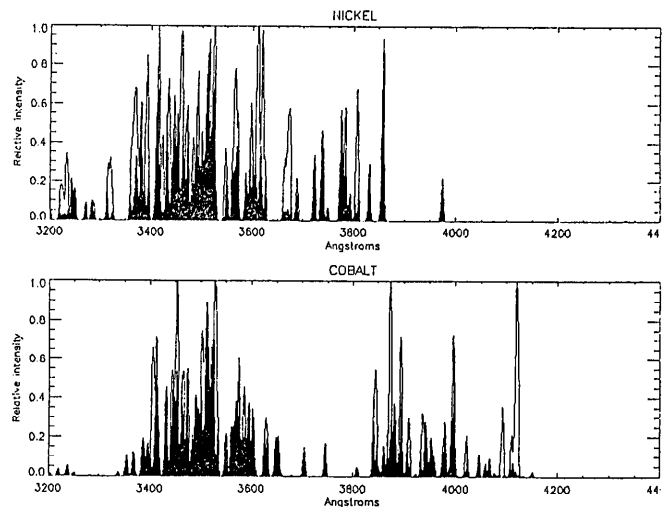

Figure 5: Relative metallic spectra for a selected range.

all normalized between 0 and 1. The spectra are then overlaid and the maximum value at each wavelength determined. This is the top curve plotted for nickel, a relative maximum curve. Second, the same curves are computed for all other metals except nickel. The maximum of this other curve is then plotted in black on top of where the nickel curve is. Notice for cobalt the two clear peaks at about 4120Å. The absense of black here shows that no other metals have peaks in this region that are over 0.05 of the metal's maximum peak. These two peaks are therefore useful indicators for cobalt. Likewise, nickel has a good clear peak at about 3670, but its peaks around 3500 interfere with other metals.

All these complex queries and plots, taken from the generated data, summarize which metallic lines are distinct and un-blended, and when they are blended, what the worst case interference would be. The interesting thing to note is that our approach for discovering indicative metal lines is more rigorous than that used by the physicist and the operator. We took their principles and embodied them in a program.

We then spent a day reviewing the results of our discovery features with the operator, Cooper. Our intent here is to make sure she had confidence in what we did. The data visualization tools where critical here. The modifications we made were at places where we had rushed through some of the earlier parts. Cooper suggested some lines of her own use, but if these were not indicated for use by our system, they were later dropped as they were found unreliable. This is not to criticize her set of lines, but rather she sometimes takes other information into account that was not considered by our automatic feature algorithm. Our algorithm follow a restricted but more rigorous approach than Cooper, and therefore the lines our algorithm uses are different.

Conclusion

Our current developments to date are as follows. We have installed the metal indicator, described above, and are currently testing it with Cooper during her post-test analysis. Other parts of the feature design approach are currently under development. We have validated the physicist's model reasonably well, and are currently linking it up to a robust optimization package that handles linear constraints, so the matching with the spectral data can be done automatically (we currently optimize "by hand"). We have also worked through some estimates of metallic erosion and got reasonable results. So we have hand-worked a basic demonstration that the software for the RAS intelligent instrument can be completed as desired using the full-theory approach. Also, with the spectral model generating validated metal spectrum, we can now restart the program described to better estimate the OH and continuum component.

More generally, we learned a number of important lessons from this work.

- Instruments can have errors in their manufacture that distort the analysis, and often times the physicists and instrument design teams will underestimate the magnitude of the problem. Their knowledge more often lies with the theoretical model, whereas, the true answer lies with the data.

- Uncovering these errors is a discovery process, requiring the interaction of targeted visualization and detailed, hand-worked analyses with the expert.

- Ideally, the instrument should be designed with this discovery process in mind—experiments should be devised in conjunction with the data analysts to explore the different error modes of the instrument and measure the empirical characteristics.

- On our task, the physicist was repeatedly surprised by the power of statistical and discovery methods. So involvement of the data analyst needs to be planned at the instrument design stage.

- Most instruments have physical models underlying their interpretation—strong prior knowledge—and it is important to both understand the limitations of these physical models, as well as incorporate them carefully into the analysis.

- In many cases, off the shelf discovery or data analysis packages did not address the problems we had. We had to design and code several new algorithms from scratch. This seems to be a common problem in the analysis of data from complex instruments (see, for instance, any of several papers in (Heidbreder 1994)).

It should be clear from this study that exact nature of the processing required for an instrument, and the potential sensitivity of an instrument are not always obvious beforehand. Increasingly complex instruments are coming out of the laboratories at an alarming rate: electrospray mass spectrometers, electron microscopes for medical imaging, high ($\gtrsim$200) band spectrometers mounted on satellites, and so forth. The Maximum Entropy conferences (Heidbreder 1994) are attended largely by physicists whose concern is the analysis of data from these complex instruments. It is not uncommon for the physicists to report an order of magnitude improvement in resolution when sophisticated post processing is done on data from radar, nuclear magnetic resonance (NMR) spectrometers, or one of the newer instruments above. Knowledge discovery can be used here to aid the development of software for these intelligent instruments.

References

Buntine, W.; Kraft, R.; Whitaker, K.; Cooper, A.; Powers, W.; and Wallace, T. 1993. OPAD data analysis. In *AIAA/SAE/ASME/ASEE 29th Joint Propulsion Conference*.

Heidbreder, G., ed. 1994. *Maximum Entropy and Bayesian Methods*. Kluwer Academic. Proceedings, at Santa Barbara, 1993.

Hinton, G.; Revow, M.; and Dayan, P. 1994. Recognizing handwritten digits using mixtures of linear models. In Tesauro, G.; Touretzky, D.; and Leen, T., eds., *Advances in Neural Information Processing Systems 7 (NIPS*93)*. Morgan Kaufmann.

Rao, C. 1965. *Linear Statistical Inference and its Applications*. Wiley.

Srivastava, A., and Buntine, W. 1995. Predicting engine parameters using the optical spectrum of the space shuttle main engine exhaust plume. In *AIAA Computing in Aerospace 10*.