Spring 5-15-2019

Quantitative Hyperspectral Imaging Pipeline to Recover Surface Images from CRISM Radiance Data

Linyun He
Washington University in St. Louis

WASHINGTON UNIVERSITY IN ST. LOUIS

Mckelvey School of Engineering

Department of Electrical & Systems Engineering

Dissertation Examination Committee:
Joseph A. O'Sullivan, Chair
R. Martin Arthur
Raymond E. Arvidson
Matthew D. Lew
Douglas A. Wiens

Quantitative Hyperspectral Imaging Pipeline to Recover Surface Images from CRISM Radiance Data

by

Linyun He

A dissertation presented to
The Graduate School
of Washington University in
partial fulfillment of the
requirements for the degree
of Doctor of Philosophy

May 2019
Saint Louis, Missouri

Contents

List of Figures

List of Tables

Acknowledgments

I was so lucky to have this great research opportunity working with multiple schools and scientists from different backgrounds for these five years.

First of all, I would like to express my deep gratitude to Professor O'Sullivan and Professor Arvidson, my research supervisors, for their patient guidance, enthusiastic encouragement and helpful critiques of this research work. They are not only talented researchers but also supportive leaders.

I would also like to thank Mr. Politte, for his technical collaboration in speeding up and releasing codes. My grateful thanks are also extended to Dr. Powell for her help in explaining planetary science and collecting basic library samples, to Mr. Condus, who worked with me on the application of neural networks to THEMIS data, which is described in Chapter 2 and to Dr. Vasavada for his support in the thermal model.

Many people, such as my current and previous lab members, have made valuable comments and suggestions during these five years which gave inspiration to improve this project. I want to thank all for their help directly and indirectly to complete this dissertation.

Personally, I wish to thank my parents and boyfriend for their emotional support and encouragement throughout my research, especially when I wanted to give up.

Financially, I'd like to thank these grants and groups: a NASA/GSFC Cooperative Agreement Notice NNX16AB15A for the Planetary Data System Geosciences Node and by Contract 110609 from the Johns Hopkins University Applied Physics Laboratory for Participation in the Mars Reconnaissance Orbiter CRISM science team.

Linyun He

Washington University in Saint Louis

May 2019

Dedicated to my parents and my fiancé.

ABSTRACT OF THE DISSERTATION

Quantitative Hyperspectral Imaging Pipeline to Recover Surface Images from CRISM Radiance Data

by

Linyun He

Doctor of Philosophy in Electrical Engineering

Washington University in St. Louis, 2019

Professor Joseph A. O'Sullivan, Chair

Hyperspectral data are important for remote applications such as mineralogy, geology, agriculture and surveillance sensing. A general pipeline converting measured hyperspectral radiance to the surface reflectance image can provide planetary scientists with clean, robust and repeatable products to work on.

In this dissertation, the surface single scattering albedos (SSAs) [2], the ratios of scattering efficiency to scattering plus absorption efficiences of a single particle, are selected to describe the reflectance. Moreover, the IOF, the ratio of measured spectral radiance (in the unit of watts per squared-meter and micrometer) to the solar spectral radiance (in the unit of watts per squared-meter and micrometer) at the observed time, is used to indicate the measurements. This pipeline includes two main parts: retrieving SSAs from IOF and reconstructing the SSA images from the SSA cube.

The Compact Reconnaissance Imaging Spectrometer for Mars (CRISM) on the Mars Reconnaissance Orbiter (MRO) helps scientists identify locations on Mars that may have exhibit

hydrated mineral phases. This dissertation mainly focuses on developing the pipeline for CRISM data. One should notice that pipelines for other hyperspectral spectrometers can also be developed based on almost the same idea.

Retrieving surface kinetic temperatures and SSA values from IOF data is challenging because the problem is under-determined and ill-posed, including modulating effects of atmospheric aerosols and gases, and surface scattering and emission properties. We introduce a general framework called STANN (Separating Temperature and Albedo using Neural Networks) to solve this kind of problem. STANN takes the hyperspectral IOF cube as inputs and outputs the retrieved temperature mapping and the corresponding SSA cube. Our STANN is derived using the Discrete Ordinates Radiative Transfer function to describe the forward model from SSA and temperature to IOF. In the STANN, we have a generator to generate more training samples based on limited library spectra and a neural network to approximate the inverse function based on enough generated training samples. This framework has been implemented for the Compact Imaging Spectrometer for Mars in a detailed manner.

SSA can be computed from IOF directly by modeling the thermal and solar reflectance together, based on retrieved temperatures. Because accurate retrieved temperature directly leads to the accurate SSA, we compare the accuracy of retrived temperatures from STANN.

The retrieved temperature has only 4 K error by one point validation (242 K) using the Curiosity Rover's thermal radiometer data. Our STANN temperature map is compared with a temperature map generated independently from a theoretical thermal model. The theoretical thermal model describes the relationship between Lambert albedo at the wavelength 1.0 µm, thermal inertia and the surface temperature. However, because the thermal inertia has pixel size larger than 100 m/pixel, the generated temperature also has the same pixel

size. Our STANN temperature is projected into the same pixel size (100 m/pixel) by the classic projection method. The two temperature maps have consistent global patterns.

Retrieved from an IOF cube, a noisy hyperspectral SSA cube needs to be denoised and reconstructed onto the Mars surface.

We propose a new algorithm, hypothesis-based estimation with regularization (HyBER), to reconstruct and denoise hyperspectral image data without extra statistical assumptions. The hypothesis test selects the best statistical model approximating measurements based on the data only. Gaussian and Poisson distributions are common respectively for continuous and integer random variables, due to the law of large numbers. Hyperspectral IOF data result from converting discrete photon counting data to continuous electrical signals after calibration. Thus, so far, Gaussian and Poisson are candidate distributions for our hypothesis tests. A regularized maximum log-likelihood estimation method is derived based on the selected model. A spatially dependent weighting on the regularization penalty is presented, substantially eliminating row artifacts that are due to non-uniform sampling. A new spectral weighting penalty is introduced to suppress varying detector-related noise. HyBER generates reconstructions with sharpened images and spectra in which the noise is suppressed, whereas fine-scale mineral absorptions are preserved. The performance is quantitatively analyzed for simulations with relative error 0.002%, which is better than the traditional non-statistical methods (baselines) and statistical methods with improper assumptions. When applied to the Mars Reconnaissance Orbiter's Compact Reconnaissance Imaging Spectrometer for Mars data, the spatial resolution and contrast are about 2 times better as compared to map projecting data without the use of HyBER.

So far, part of our results have enabled planetary scientists to identify minerals and understand the forming history of Mars craters. Some of these findings are verified by the

Opportunity Rover's measurements. In the future, results from this pipeline for CRISM are promising to play more and more critical roles in the planetary science.

Chapter 1

Introduction

1.1 Motivation and Background

Hyperspectral imaging sensors comprise a class of spectroscopy sensors, for which the sensed wavelengths are divided into hundreds of contiguous narrow frequency bands. Hyperspectral data are important for remote applications such as mineralogy, geology, agriculture and surveillance sensing [3].

Van et al. summarize how remote sensing data help us to understand surface materials [3]. Reflectance spectra of minerals are dominated in the visible-near infrared (VNIR) wavelength range by the presence or absence of transition metals (e.g., Fe, Cr, Co, Ni) resulting in absorption features due to electronic processes. The presence or absence of water and hydroxyl, carbonate and sulfate determine absorption features in the shortwave infrared (SWIR) region due to vibrational processes. In addition, absorption band depths are related to grain size, as the amount of light scattered and absorbed by a grain is dependent on grain size. A larger grain has a greater internal path where photons may be absorbed. On the contrary in smaller grains there are proportionally more surface reflections compared to internal photon path lengths, if multiple scattering dominates, the reflectance decreases

with increasing grain size. As the grain size becomes larger, more light is absorbed and the reflectance drops. In general, absorption band depth is correlated with the (relative) amount of material present. Mid infrared wavelengths from 3 µm to 5 µm (MIR) are areas that have great potential for geologic remote sensing studies (3 µm is related to the surface water), but have been under-investigated probably due to (1) the complexity of the physics (e.g., preserve of both solar and thermal emission) and (2) lack of field/laboratory spectrometers and (airborne/spaceborne) hyperspectral data [3].

Absorption of energy and thus the presence of absorption features in reflectance spectra due to electronic processes manifests in the VNIR part of the spectrum in broad features. Vibration gives rise to narrow and more pronounced features in the SWIR (as overtone features of mid infrared and longwavelength infrared absorption) roughly at 1.4 µm combined with 1.9 µm due to molecular water, 2.2 µm due to Al-OH, 2.3 µm due to Mg-OH [3].

The objective of hyperspectral remote sensing (also referred to as imaging spectrometry or imaging spectroscopy) is to measure quantitatively the components of the surface (radiance, reflectance or emissivity) spectra acquired as images in many narrow and contiguous spectral bands. For geologic applications this is done by using spectral absorption features to map planetary surface composition (in terms of mineralogy or lithology) or to quantify rock or soil chemistry or physics using derivative image products.

Based on the type of spatial information acquired, there are three basic types of hyperspectral sensor data: point scanning data, line scanning data and focal plane scanning data [4]. Fig. 1.1 shows how these three scanning ways work.

In this dissertation, we focus on hyperspectral images from the Compact Reconnaissance Imaging Spectrometer for Mars (CRISM) instrument [5] on the Mars Reconnaissance Orbiter (MRO) [5], which helps planetary scientists identify locations on Mars that may have hosted

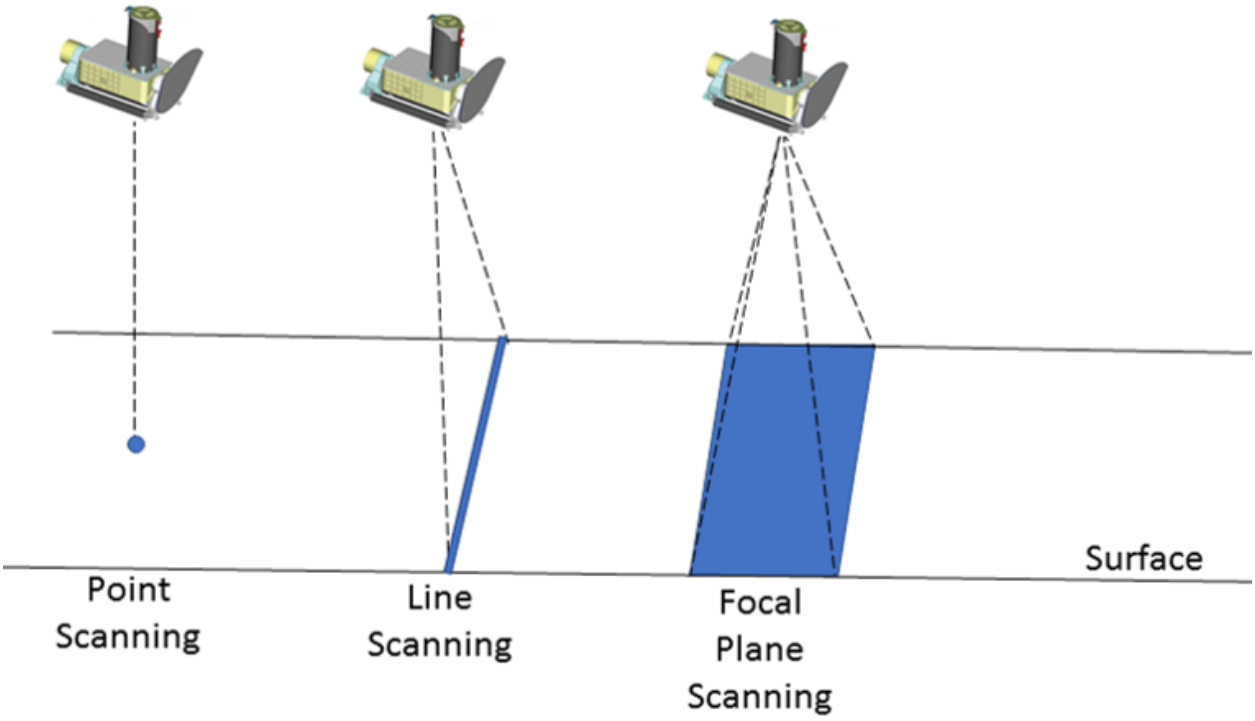


Figure 1.1: Three basic types of hyperspectral scanning.

water based on mapping hydrated mineral phases. The MRO has been in a 255 km by 320 km near-polar orbit around Mars since 2006 [5]. When the MRO spacecraft is at an altitude of 300 kilometers, CRISM, operating from 370 to 3920 nm (VNIR S-detector: 364-1055 nm; IR L-detector: 1001-3936 nm) with 6.55 nm spectral spacing, scans a sequence of narrow but long lines on the Martian surface; each location of the sensor measures a row of data from a region that is about 18 meters across and 10,800 meters long [5]. Beginning in 2010 a new gimbaled motion was implemented when requested and employs significant pixel overlap in the along-track direction [5]. Fig. 1.2 indicates how different sampling modes work and Fig. 1.3 gives two examples of CRISM data. Over the course of the MRO mission, CRISM has acquired over 276,000 individual observation segments (or mapping strips) with a variety of observing modes and data characteristics [6]. The aggregate VNIR and IR mapping coverage at the equator stand at $\sim 97\%$ and $\sim 78\%$, respectively, increasing poleward so that at higher latitudes there are typically multiple individual mapping strips that sample a given ground

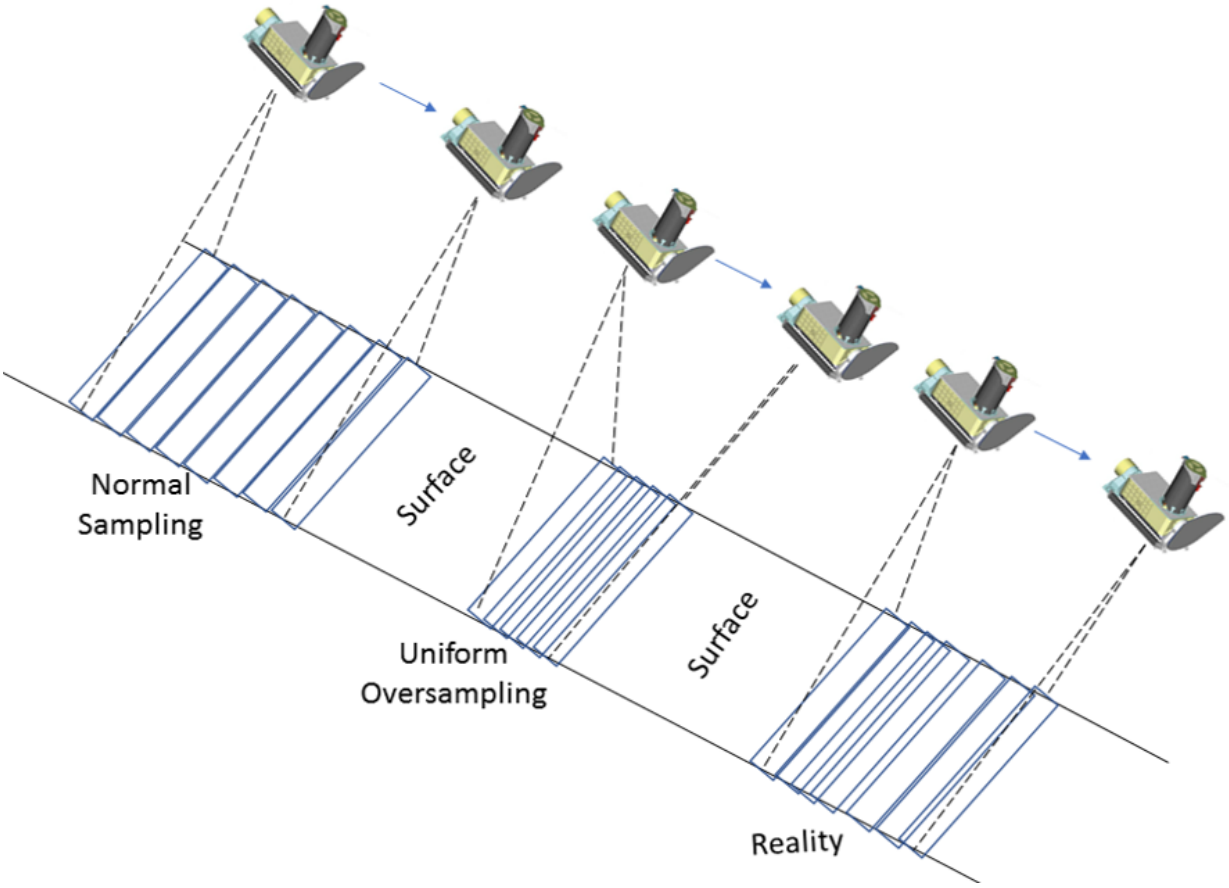
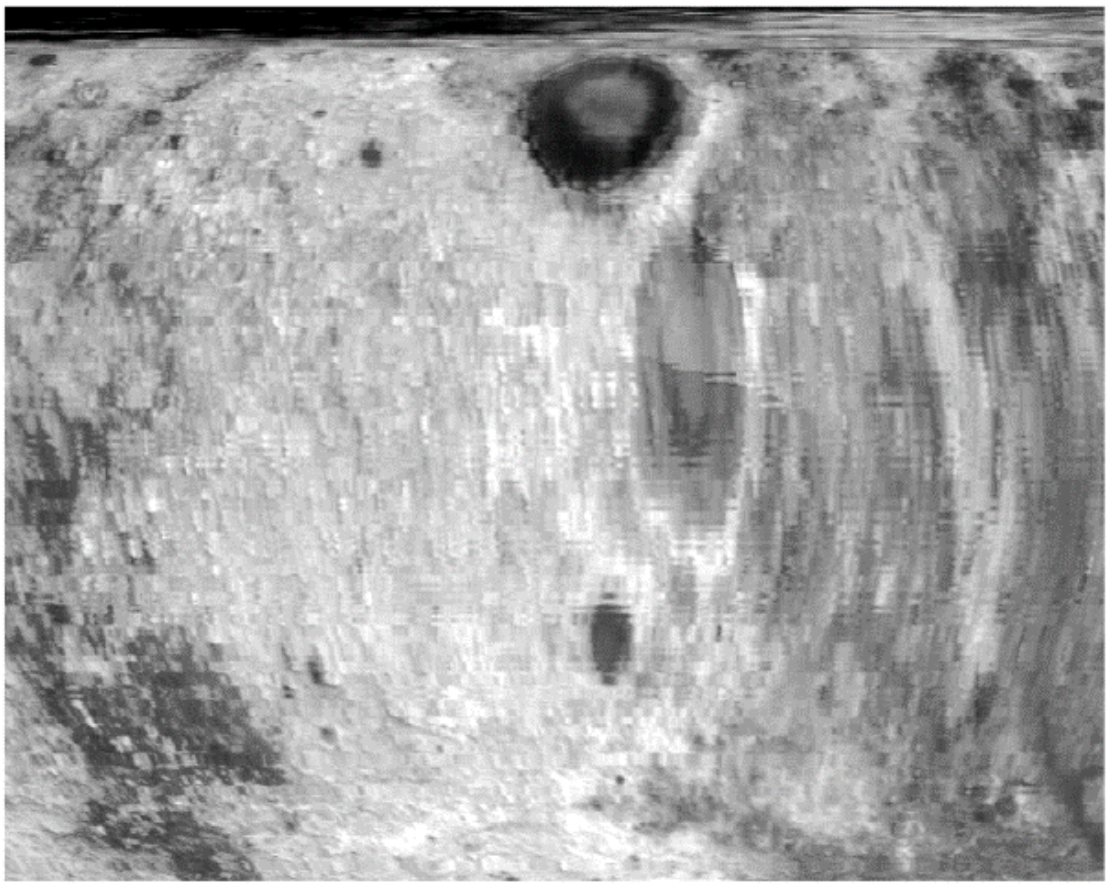
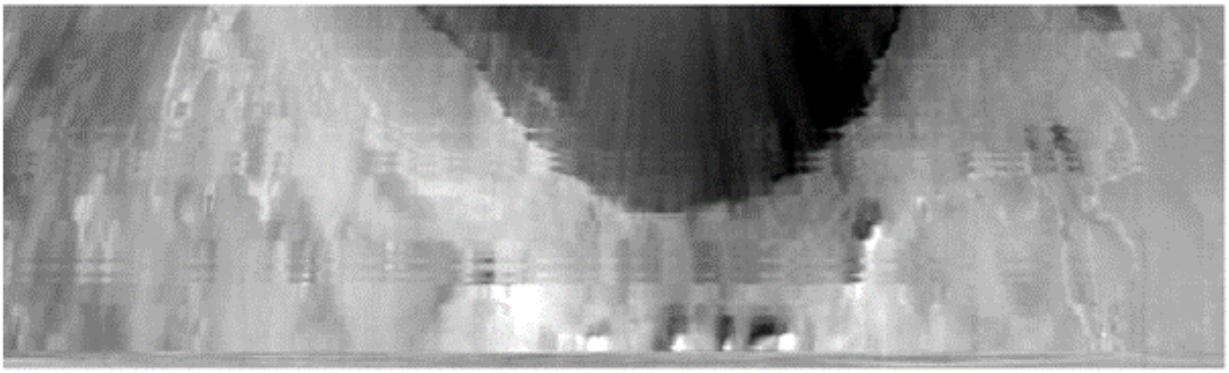


Figure 1.2: Examples of how CRISM always samples: normal sampling, uniform oversampling and real sampling.

location [6]. The quality of the CRISM VNIR mapping data has not changed substantially over the course of the mission, but the IR mapping data quality has varied significantly as function of the IR detector operating temperature. The CRISM IR detector is actively cooled and as the cryosystem efficiency has decreased with age, the IR detector temperature and associated noise level have increased accordingly [6]. The manifestation of the thermally driven IR noise is more challenging to address in the CRISM mapping data as compared to the hyperspectral targeted observation data.



(a) FRT000B6F1: no over-sampling



(b) ATO000D7D74: over-samplilng at target areas

Figure 1.3: Examples of collected data from different sampling modes. Data are shown at band 10 (wavelength 2.588 micrometers). Notice more obvious ripples in Fig. 1.3b than in Fig.1.3a, which are caused by over-sampling.

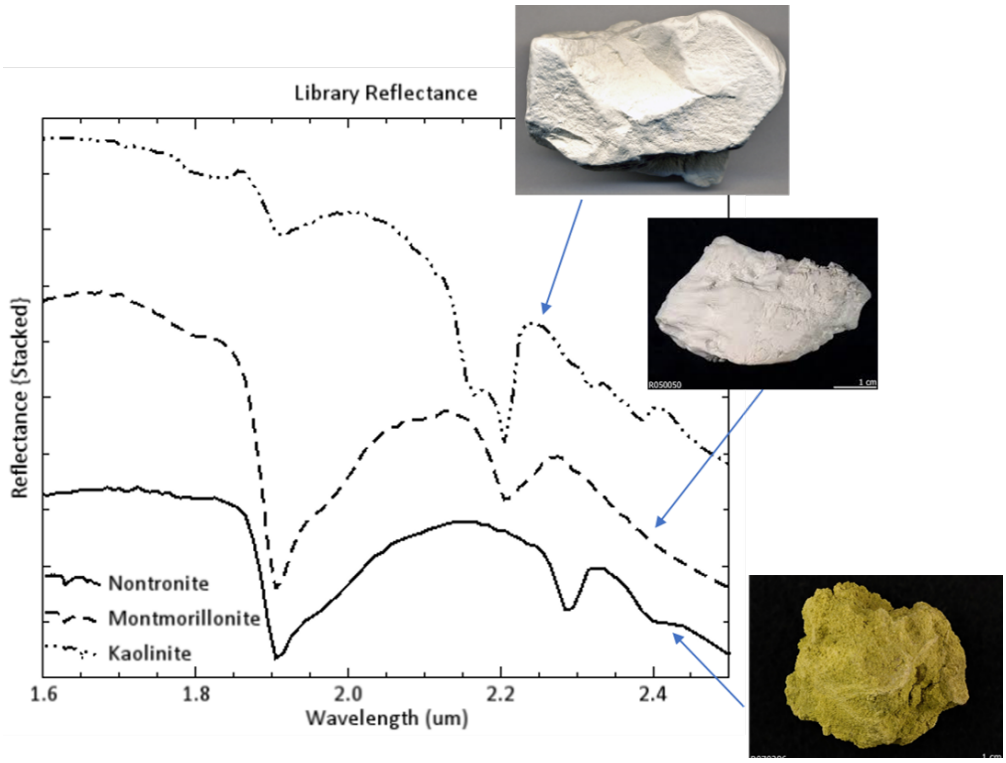


Figure 1.4: Different minerals have different spectra.

Every spatial pixel in one hyperspectral data cube has one spectrum, providing valuable information about the properties of this surface pixel. Fig. 1.4 shows one example that different minerals have different reflectance spectra, especially different absorption features.

Data directly collected by sensors are radiance instead of reflectance. In this dissertation, surface single scattering albedo (SSA) [2], the ratio of scattering efficiency to scattering plus absorption efficiencies of a single particle, is selected to describe the reflectance[1]. Compared with a measured hyperspectral radiance cube, a retrieved SSA cube is independent of the lighting and viewing conditions as well as atmospheric effects [2]. Thus, SSA is widely used in the planetary science analysis for the hyperspectral data. Fig. 1.5 shows a comparison between measured radiance and SSA.

[1]Other albedos such as Lambert albedo can also be used. Go to `https://en.wikipedia.org/wiki/Albedo` to know more.

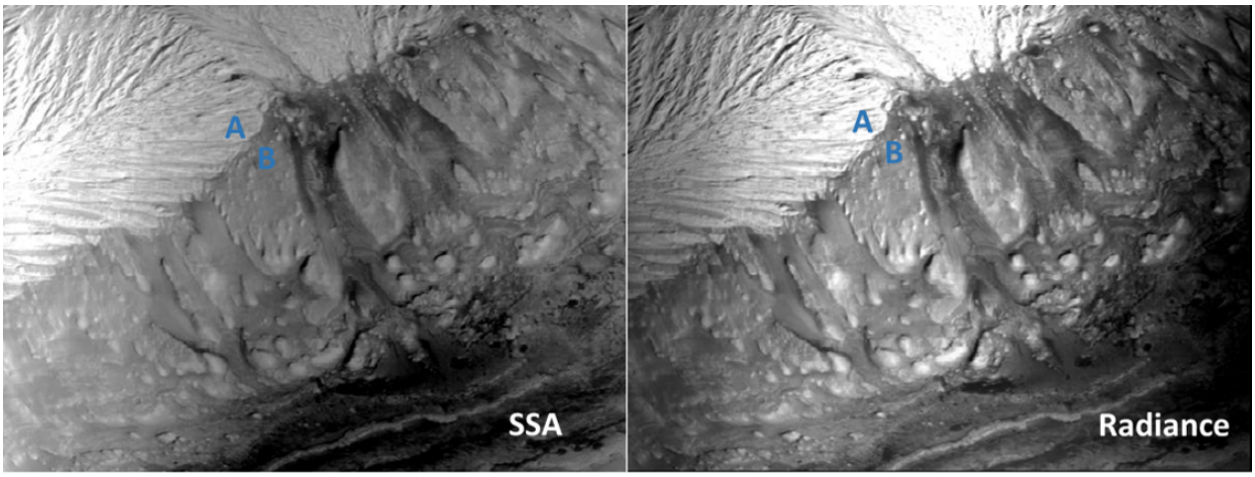


Figure 1.5: CRISM FRT000B6F1 is used as an example. The SSAs at 2 µm are shown in the left and the corresponding radiance are shown in the right. The location A and B should have similar minerals but the right radiance have obvious difference due to the lighting condition.

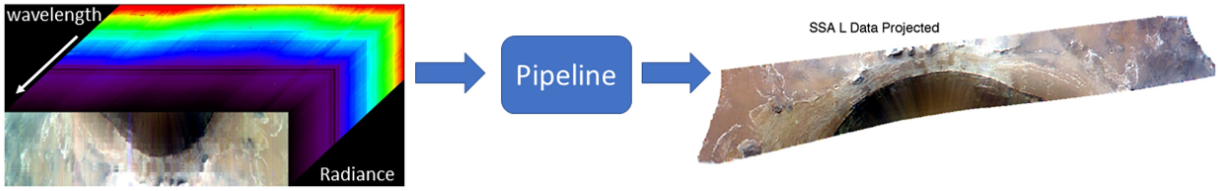


Figure 1.6: Big picture of the pipeline proposed by this dissertation. After inputting hyperspectral radiance cube, the pipeline outputs clean and well reconstructed hyperspectral SSA images.

This dissertation proposes and develops a hyperspectral data processing pipeline, which outputs clean and well reconstructed SSA images based on radiance data from hyperspectral sensors as shown in Fig. 1.6. A projected single scattering albedo image with better spatial resolution and denoised spectra enables planetary scientists' research [7–9].

1.1.1 SSA retrieval

A lot of research has sought to retrieve the spectral albedos from the radiance. For the spectral range lower than 2.6 µm, measured radiance is only related to the solar reflectance,

thus there exists a one-to-one relationship between SSA and radiance [10]. For the thermal spectral range ($> 2.6\,\mu m$), measured radiance starts to include the thermal emission, which makes the problem ill-posed because the surface temperature of the thermal radiance is also unknown. If the surface temperature is constant for one spatial pixel, hyperspectral data provide the potential opportunity to derive both the surface temperatures and the SSA spectra. However, algorithms for retrieving the surface temperatures and albedos are few, not only due to the difficulty in modeling the forward function from temperature and albedos to radiance (atmospheric effects should be included), but also the ill-posed property of estimating N+1 variables based on N measurements.

Discrete Ordinates Radiative Transfer (DISORT)-based processing [2] models radiative transfer streams associated with dust and ice aerosols, CO_2, H_2O and CO gases, and a surface boundary layer that is modeled by the Hapke function [11]. Beginning with dust and ice aerosols based on inferences from Opportunity or Curiosity rover sites, DISORT outputs a huge look-up table to describe the relationship between the SSA and the radiance for a given location (fixed emission, incident and phase angle), by modeling both the solar reflectance and the thermal emission based on the surface temperature.

Even if DISORT can provide this relationship with the atmosphere involved, the temperature and SSA inversion from radiance is still hard. N+1 unknown variables (SSAs at N bands plus the surface temperature) need to be estimated based on N measurements (radiance at the same N bands).

There exists an area of research Land Surface Temperature (LST) retrieval, which is key in a climate system. LST can also be used for weather or climate forecasting and studying. Besides the temperature and emissivity separation method [12], other physical methods to retrieve LST from satellite-based hyperspectral or multispectral thermal infrared data

include but are not limited to: two-step physical algorithm [13], split-window algorithm [14], double-viewing angle method [15] and Kalman physical filter [16].

Neural networks [17] have been well applied to remote sensing images, especially for the classification based on multi/hyperspectral data [18]. According to Zhu, et al. [19], the number of published papers related to neural networks in remote sensing in 2017 was more than 30 times the number in 2014. Supervised convolutional neural networks are developed to classify hyperspectral images in the spectral and spatial domains [20,21]. Li, et al. propose an anomaly detection method to identify weird pixels in hyperspectral data by addressing a convolutional neural network [22]. Neural networks are also applied to the interpretation of high-resolution hyperspectral images [19] such as pansharpening and superresolution, which will be introduced later.

Neural networks also play a role in the surface temperature retrieval. Aires, et al. proposed a neural network approach to retrieve the atmospheric water vapor in 2001 [23]. Later, Wu et al. applied a neural network to predict the sea surface temperature [24]. Mao, et al. used a neural network on Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER) imagery and the error can be controlled under 0.7 K [25]. Later Cheng et al. found that the performance of a neural network on FIIR datasets is also attractive [26].

Methods to retrieve the surface temperature and reflectance from radiance can be summarized into two categories: physical methods [27–30]; and neural networks methods [25, 26, 31–33]. The physical method retrieves a specific albedo based on prior knowledge and assumptions, such as assuming the albedo at longer wavelengths (thermal region) is the linear projection of ones at shorter wavelengths (non-thermal region) with the non-thermal albedos available [30]. However this assumption is not strictly true for all planets and sometimes it

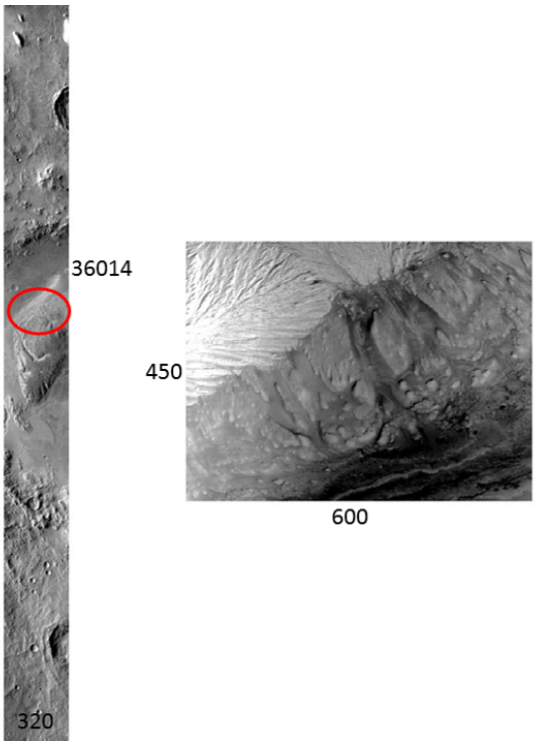
needs to involve much human intervention to obtain good performance for a specific spectrometer. A Neural networks method is applied to ASTER whose bands can be longer than 8 µm and it has the good performance [25, 26, 31]. However, there are limited studies on why neural networks can work so well and whether this idea can be extended to more instruments or planets.

1.1.2 Image Reconstruction

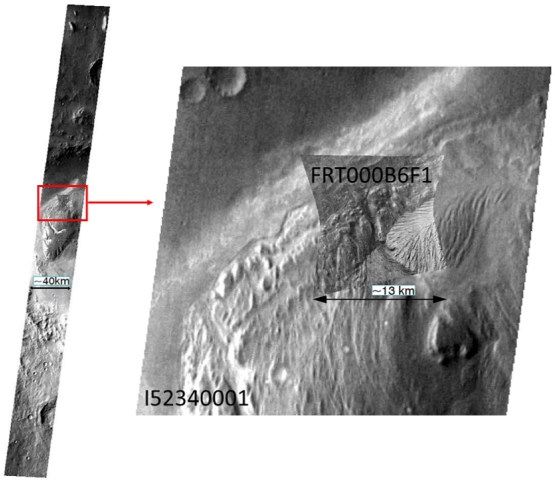
After retrieving an SSA cube, we also need to reconstruct the SSA into the projected map.

Most common planetary hyperspectral imaging instruments are line scanning as shown in Table 1.1. Image reconstruction is necessary for point and line scanning instruments because measurements are not mapped in areographic coordinates. Fig. 1.7 shows the necessity example.

In this thesis, the Mars equirectangular map projection [34] is used because data sets we employed as examples were acquired near the equator. The Mars equirectangular projection maps meridians from the center of the sphere and parallels with lines parallel to the equatorial plane. Moreover, spatially oversampled hyperspectral data can be reconstructed at a smaller pixel size, which can improve the spatial resolution by proper reconstruction methods. There are several signal processing approaches called Pansharpening [35] that can be used to enhance the spatial resolution, including component substitution [36], multiresolution methods [37], least squares estimation [38, 39] and statistical methods [40]. Maximum a posteriori (MAP) estimation has been proposed by Hardie et al. [41]. Zhang et al. introduced a method that works in the wavelet domain [42] and later published an expectation maximization algorithm to maximize the posterior distribution [43]. Convolutional neural networks have been applied to obtain superresolution of hyperspectral images [44].



(a) In the sensor space. THEMIS I52340001 (320 columns and 36014 rows) is shown in the left and FRT000B6F1 (600 columns and 450 rows) is shown in the right. The red circle in the I5234001 shows the area where FRT000B6F1 covers.



(b) After projected into the same coordinates. Both THEMIS I52340001 and FRT000B6F1 are shown in the left and the red box is zoomed in and shown in the right. One can find that FRT000B6F1 are well overlapped on the scene I52340001 because they are now projected into the same coordinate.

Figure 1.7: CRISM FRT000B6F1 SSA at 2 µm and THEMIS I52340001 are used as an example. One can find that in the sensor space, even we can know the relative locations of two scenes collected by different instruments, due to different pixel size, it is hard to compare them directly. After projecting into the same coordinates, it is much easier to analyze both of them jointly.

Table 1.1: Some common planetary hyperspectral imaging instruments

Planetary object	Hyperspectral Imaging Instrument	Corresponding Mission	Wavelengths (μm)	Scanning
Moon	Hyperspectral Imager (HySI) [45]	Chandrayaan-1, ISRO	0.42-0.96	Line
Moon	Moon Mineralogy Mapper [45]	Chandrayaan-1, ISRO	0.43-3.0	Line
Mars	Compact Reconnaissance Imaging Spectrometer for Mars (CRISM) [5]	Mars Reconnaissance Orbiter	0.4-4	Line
Mars	Observatoire pour la minéralogie, l'eau, les glaces et l'activité (OMEGA) [46]	Mars Express, European Space Agency [47]	0.36-5.2	Line
Earth	Airbone Visible/Infrared Imaging Spectrometer (AVIRIS) [48]	NASA's ER-2 Jet [2]	0.4-2.5	Line

1.1.3 Image Denoising

Even if the image can be reconstructed well, hyperspectral sensor data suffer from multiple sources of noise, which limits the utility of further processing and analysis, such as classification, unmixing, or object detection [7, 49, 50]. One example of how noisy data suppress small absorption features is shown in Fig.1.8. Therefore it is critical to reduce the noise in hyperspectral reconstructions when reconstructing them from measured data.

In recent decades, many hyperspectral image denoising algorithms have been proposed. For most statistical denoising methods, it is assumed that hyperspectral sensor data have additive Gaussian noise [51–55]; this assumption is often used for CRISM data [49,56]. This statistical assumption might work sometimes but previous classic hypothesis tests such as Kolmogorov-Smirnov [57] and χ^2 tests [58] will reject all distributions if true values are unknown; ground-truth is often missing for hyperspectral sensor data.

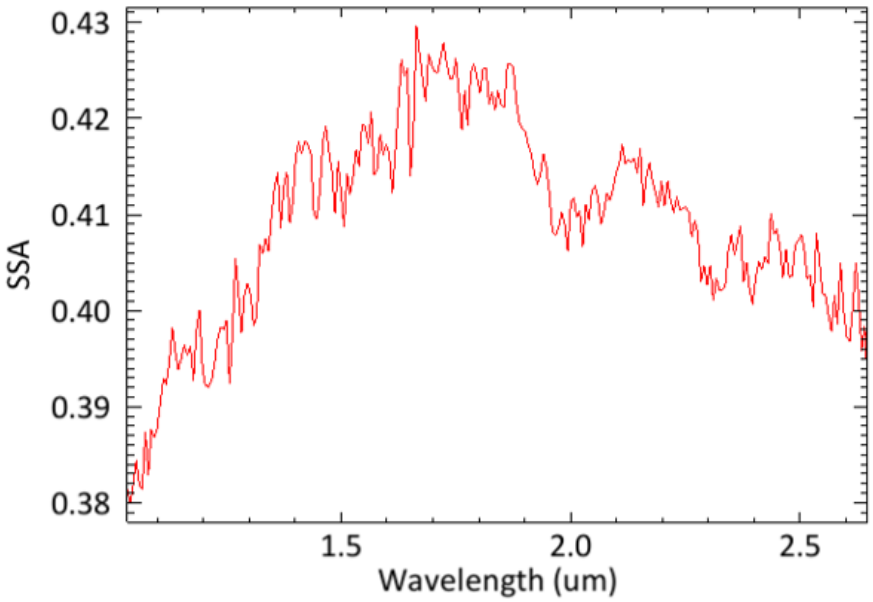


Figure 1.8: One SSA spectrum is shown. After analyzing, there should be two absorption features at 2.3 µm and 2.4 µm. Suppressed by noise, it is hard to tell now.

With statistical models assumed, classical algorithms can be employed to compute a penalized maximum likelihood image, such as the expectation-maximization (EM) algorithm. Regularization is necessary because the forward projection in the reconstruction model is ill-conditioned [59]. Yuan et al. proposed a spectral-spatial kernel regularization [60] and good regularized methods can remove phenomena noise such as striping noise [61].

As hyperspectral images may have hundreds of bands, and the noise intensity in each band varies, the suppression of noise should be adaptive based on data [53].

Moreover, due to the nonuniform spatial sampling of some hyperspectral sensors, some spatial areas are highly undersampled. If the reconstruction method does not account for it, spatial artifacts can be introduced.

In summary, our pipeline takes the form shown in Fig. 1.9. We combine the reconstruction and denoising as one processing step to avoid missing information. Although, CRISM images

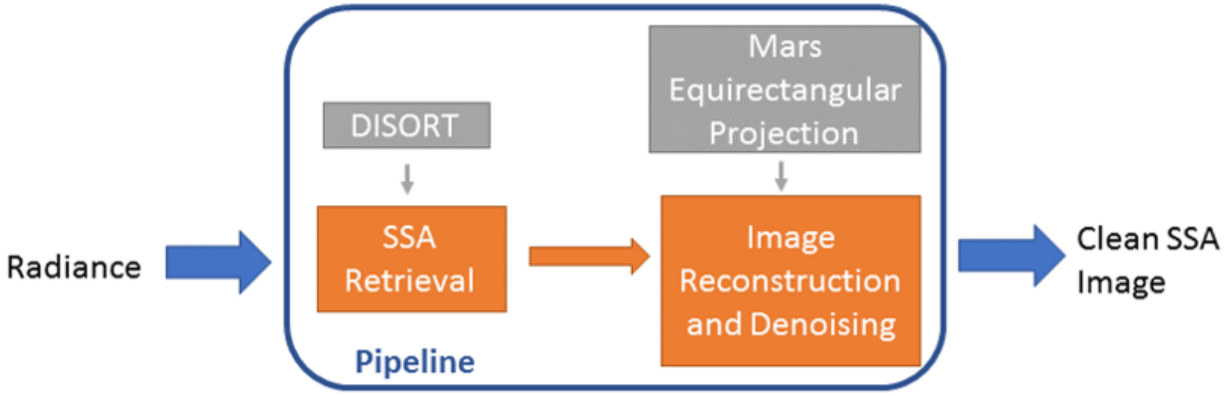


Figure 1.9: Pipeline of this dissertation.

are the main applications in this dissertation, the pipeline idea is general for a wide variety of hyperspectral data.

1.2 Problem Description

1.2.1 SSA retrieval

We call the function from SSA s_{mk} to the radiance r_{mk} the **forward model** f which has a general format for every spatial pixel m as shown

$$r_{mk} = f(s_{mk}, T_m, l_m, \lambda_k), \qquad k = 1, 2, ..., K, \tag{1.1}$$

where T_m, l_m are the surface temperature and the geometric information (such as incident angle, emission angle and phase angle) for this pixel. T_m is a scalar and l_m is a vector containing all geometric information. λ_k is the corresponding wavelength for the k^{th} band.

In this case, we have K bands in total. The forward model f is modeled by DISORT in [2] with simulated Mars atmospheric effects. f does not have a closed-form expression if it is modeled with atmospheric effects involved. Therefore, in practice, f is usually implemented as a huge look-up table with spline interpolation.

In general, we aim to solve the inverse function of this forward model as

$$[s_{mk}, T_m] = f^{-1}(r_{mk}, l_m, \lambda_k), \qquad k = 1, 2, ..., K. \tag{1.2}$$

Although we have K known equations ($k = 1, 2, ..., K$) from (1.1), we need to estimate $K + 1$ unknown variables s_{mk} and T_m for every spatial pixel. The ill-posed property leads to an inverse function f^{-1} which is one-to-many, in other words, there are several different combinations of s_{mk} ($k = 1, 2, ..., K$) and T_m that can generate the same r_{mk}. Therefore, the solutions of (1.2) are not unique theoretically.

If either s_{mk} or T_m is known, this inverse function can be mapping f_T^{-1} or f_S^{-1} with the unique solution

$$s_{mk} = f_S^{-1}(r_{mk}, l_m, \lambda_k, T_m), \tag{1.3}$$

or

$$\forall k, \quad T_m = f_T^{-1}(r_{mk}, l_m, \lambda_k, s_{mk}). \tag{1.4}$$

Also, if $\lambda_k < 2.5\mu m$ where the radiance does not depend on temperature, the unique SSA can be computed also by the one-to-one mapping f_S^{-1} as

$$s_{mk} = f_S^{-1}(r_{mk}, l_m, \lambda_k), \qquad \lambda_k < 2.5\mu m. \tag{1.5}$$

As we introduce in Section 1.1, previous methods to solve (1.2) can be summarized into two main ideas:

1. Physical method.
2. Neural network method.

Combining with the equations above, we can also summarize the mathematics of these two ideas[3].

Physical method The SSA at a specific band $k^{(1)}$ is computed from the function h with the available SSA ($s_{mk^{(2)}}$) at another band $k^{(2)}$. We can compute the temperature based on this SSA and its corresponding radiance from the one-to-one inverse functions f_T^{-1} and f_S^{-1}. When the surface temperature for this pixel is obtained, the SSA for other bands ($s_{mk}, k \neq k^{(1)}$) can be uniquely computed.

$$\begin{aligned}
&\text{Step 1:} \quad s_{mk^{(1)}} = h(s_{mk^{(2)}}) \qquad \text{for some } k^{(1)} \\
&\text{Step 2:} \quad T_m = f_T^{-1}(s_{mk^{(1)}}, r_{mk^{(1)}}, l_m, \lambda_{k^{(1)}}) \\
&\text{Step 3:} \quad s_{mk} = f_S^{-1}(T_m, r_{mk}, l_m, \lambda_k) \qquad k = 1, 2, ..., K
\end{aligned} \tag{1.6}$$

[3]Although other methods focus on albedos rather than SSA, we use SSA instead of albedos to compare with our method.

The key step of this method is to estimate $s_{mk^{(1)}}$ as accurately as possible. There has been a lot of research on this step [27–30]. For example, Clark et al. used a linear projection assumption to remove thermal effects from near-infrared multispectral data on the Moon [30]. They assume there exists a linear relationship between the emissivity at near-infrared wavelengths (less than 2.5 µm) and at short-infrared wavelengths (such as 3.7 µm). However, this assumption is not always strictly true and sometimes it needs to involve much human intervention (such as which wavelengths should be picked) to obtain good performance for a specific instrument or location. Erard et al. use a linear combination of spectra from brighter and darker areas on Mars [29].

Neural network method A new function g is learned by neural networks (NNs) to approximate the inverse function f^{-1}. Although g can output both s_{mk} and T_m, it is still common to retrieve a more accurate s_{mk} from the inverse function based on retrieved T_m as shown in (1.7).

$$\begin{aligned} &\text{Step 1:} \quad [T_m, s_{mk}] = g(r_{mk}, l_m, \lambda_k) \qquad \text{for all } k \\ &\text{Step 2:} \quad s_{mk} = f_T^{-1}(T_m, r_{mk}, l_m, \lambda_k) \end{aligned} \tag{1.7}$$

This idea is stimulated by the fact that neural networks can approximate almost all continuous functions [62, 63]. The goodness of approximation for a NN depends on the complexity of function f^{-1} and the number of training samples. The application of NN to ASTER performs very well as analyzed by previous authors [25, 26, 31]. In the following, we will explain the reason why it can work and whether it can be extended to other instruments and planets such as Mars.

1.2.2 Image Reconstruction and Denoising

Define $\mathbb{U} = \mathbb{R}^+$ as the positive real number space.

If we have I measurements in total (containing both spatial and spectral domains, thus $I = M * K$), SSA data $\mathbf{s} = [s_1, s_2, ..., s_I]'$ and their estimates $\mathbf{a} = [a_1, a_2, ..., a_I]'$ should belong to $\mathbb{U}^I$, that is $\mathbf{s}, \mathbf{a} \in \mathbb{U}^I$.

The reconstruction mapped image in local Cartesian axis $\mathbf{c} = [c_1, c_2, ..., c_J]'$ contains J pixels in total (also both spatial and spectral); $\mathbf{c} \in \mathbb{U}^J$.

The physical model provides the system matrix (forward model) $H \in \mathbb{R}^{I \times J}$, which describes how the mean data $\mathbf{a}$ is linearly generated from the mapped reconstruction $\mathbf{c}$, $\mathbf{a} = H\mathbf{c}$. Often, H is called a transfer function. We use a Gaussian-shape kernel as the transfer function for this hyperspectral image system.

If the probability density function of measurements SSA $\mathbf{s}$ is $p(\mathbf{s}|\mathbf{a})$ with $\mathbf{a}$ as the ground truths, then the reconstruction problem can be formulated as a statistical optimization problem. The maximum likelihood estimate achieves

$$\begin{aligned} &\max_{\mathbf{c}} \quad \ln p(\mathbf{s}|\mathbf{a}) \\ &\quad s.t. \ \mathbf{a} = H\mathbf{c}. \end{aligned} \tag{1.8}$$

The ill-posed nature of the forward model H and the noise in data $\mathbf{s}$ create instabilities in estimates for the mapped reconstruction image $\mathbf{c}$. Therefore a convex regularization function

$\Phi_{reg}(\mathbf{c})$ is introduced. Combining with (1.8), the reconstruction problem can be stated as an a penalized maximum likelihood problem.

$$\begin{aligned} \min_{\mathbf{c}} \quad & -\ln p(\mathbf{s}|\mathbf{a}) + \Phi_{reg}(\mathbf{c}) \\ & s.t.\ \mathbf{a} = H\mathbf{c}. \end{aligned} \tag{1.9}$$

The negative loglikelihood $-\ln p(\mathbf{s}|\mathbf{a})$ is a data fitting term because it is a measure of how close the measured data $\mathbf{s}$ are to the predicted mean data $\mathbf{a}$.

If $p(\mathbf{s}|\mathbf{a})$ is assumed to be Gaussian distributed with equal variances, (1.9) is equivalent to

$$\begin{aligned} \min_{\mathbf{c}} \quad & \sum_{i=1}^{I} (s_i - a_i)^2 + \Phi(\mathbf{c}) \\ & s.t.\ \mathbf{a} = H\mathbf{c}. \end{aligned} \tag{1.10}$$

Here, the data fit term reduces to square-error. If $p(\mathbf{s}|\mathbf{a})$ is assumed to be Poisson distributed, (1.9) is equivalent to

$$\begin{aligned} \min_{\mathbf{c}} \quad & \sum_{i=1}^{I} (s_i \ln a_i - a_i) + \Phi(\mathbf{c}) \\ & s.t.\ \mathbf{a} = H\mathbf{c}. \end{aligned} \tag{1.11}$$

In (1.10) and (1.11), $\Phi(\mathbf{c})$ is the equivalent regularization function from $\Phi_{reg}(\mathbf{c})$ in (1.9) after simplifying the data fitting term.

To reconstruct and denoise measured $\mathbf{c}$ well, an accurate probability density function is important for the data fitting term and a reasonable regularization needs to be adjusted for the penalty function.

1.3 Notation

$r, \mathbf{r}$	Radiance in scale (r) and in vector ($\mathbf{r}$)
$s, \mathbf{s}$	Single Scattering Albedo (SSA) in scale (s) and in vector ($\mathbf{s}$)
l	Spatially dependent parameters (such as emission and incident angles) excluding temperatures
λ	Wavelengths
$a, \mathbf{a}$	Ground truth SSA in the sensor space in scale (a) and in vector ($\mathbf{a}$)
$c, \mathbf{c}$	Reconstructed SSA in the projected coordinates in scale (c) and in vector ($\mathbf{c}$)
f	Forward model function from SSA to radiance
f_S^{-1}	Inverse function of f to calculate the unique s from r and T for $\lambda > 2.5\,\mu\text{m}$ or from s only for $\lambda < 2.5\,\mu\text{m}$
f_T^{-1}	Inverse function of f to calculate the unique T from r and s
Δr	Sensitivity of radiance with respect to temperatures
H	Linear transfer function from SSA in the projected coordinates to the sensor space
i, I	Index for a and s (I in total) including both spectral and spatial
j, J	Index for c (J in total) including both spectral and spatial
m, M	Spatial index for s, r, l and T (M in total)
k, K	Spectral index for s and r (K in total)
$'$	Transpose
Φ	Regularization function

p Probability density function

1.4 Contributions and Organization

The main contribution of this dissertation is to propose and develop a general pipeline from measured radiance to reconstructed surface albedos. The goal is to create new methods and tools that enable our scientific collaborators to perform planetary science research. Two main algorithms are developed to realize this pipeline:

- **STANN:** Retrieve single scattering albedos and temperature maps by neural networks.
- **HyBER:** The surface images are reconstructed.

STANN stands for separating temperature and albedo by neural networks, which is described in detail in Chapter 2. The main technical contributions for STANN are as follows.

- One-layer neural network model for longer and shorter wavelengths is proposed, trained and tested for temperature and albedos.
- Variational autoencoder is selected as the spectra generator used to generate enough training samples.
- The global pattern of the resulting temperature map is consistent with another temperature map independently generated from one thermal model.
- Local values of resulting temperature map are consistent with the local measurements by the Mars Rover Curiosity.

- Limitation of neural network approximation is quantitatively described by computing the sensitivity with respect to temperatures.

HyBER means hypothesis-based estimation with regularization, which is described in detail in Chapter 3. For HyBER, the main contributions are as follows.

- A new and general hypothesis testing method is proposed to automatically select a statistical model based on measured data without any extra assumptions.
- Based on the selected statistical model, a maximum loglikelihood algorithm is derived to reconstruct and denoise the hyperspectral image.
- A spatial and spectral weighting penalty is utilized to avoid artifacts introduced by non-uniform spatial sampling and varying noise levels.
- Quantitative performance is analyzed for HyBER results based on simulations with 0.002% error.
- Spatial image quality is analyzed for HyBER image reconstructions, which is 2 times better compared with the projection without HyBER.

This whole pipeline now has been released for planetary scientists and products of this pipeline have helped them to analyze the Mars mineralogy. The results of the mineralogy analysis are consistent with measurements by Mars Rovers. The products have been used to inform decision for Mars Rovers' landing locations and planned paths. It is very promising to see increasing impacts of this pipeline on Mars research projects.

Chapter 4 summarizes corresponding main works and impacts of this dissertation in the pipeline. Some future work and potential impacts for this pipeline are also described.

Instructions of hypothesis codes are described in Appendix A and some algorithm derivations are shown in detail in Appendix B.

Chapter 2

Separating Temperature and Albedo by Neural Networks

2.1 Mathematical Background

Any hyperspectral data cube has two domains: spatial domain and spectral domain. In this chapter, we use m to index the spatial pixels and k, j to index the spectral pixels (bands).

We call the function from SSA to the radiance the **forward model** f which has a general format for every spatial pixel m as shown

$$r_{mj} = f(s_{mj}, T_m, l_m, \lambda_j), \qquad j = 1, 2, ..., N. \tag{2.1}$$

In this case, we have N bands in total. The forward model f is modeled by Discrete Ordinates Radiative Transfer [2] with simulated Mars atmospheric effects. f does not have a closed-form expression if it is modeled with atmospheric effects involved. Therefore, in practice, f is usually implemented as a huge look-up table with spline interpolation.

In general, we aim to solve the inverse function of this forward model as

$$[s_{mj}, T_m] = f^{-1}(r_{mj}, l_m, \lambda_j), \qquad j = 1, 2, ..., N. \tag{2.2}$$

Although we have N known equations ($j = 1, 2, ..., N$) from (2.1), we need to estimate $N + 1$ unknown variables s_{mj} and T_m for every spatial pixel. The ill-posed property leads to an inverse function f^{-1} which is one-to-many, in other words, there are several different combinations of s_{mj} ($j = 1, 2, ..., N$) and T_m that can generate the same r_{mj}. Therefore, the solutions of (2.2) are not unique theoretically.

If either s_{mj} or T_m is known, this inverse function can turn out to be one-to-one mapping f_T^{-1} or f_S^{-1} with the unique solution $s_{mj} = f_S^{-1}(r_{mj}, l_m, \lambda_j, T_m)$ or $\forall j, \quad T_m = f_T^{-1}(r_{mj}, l_m, \lambda_j, s_{mj})$. Also, if $\lambda_j < 2.5\mu m$ where the radiance does not depend on temperature, the unique SSA can be computed also by the one-to-one mapping f_S^{-1} as

$$s_{mj} = f_S^{-1}(r_{mj}, l_m, \lambda_j), \qquad \lambda_j < 2.5\mu m. \tag{2.3}$$

To determine the sensitivity of the forward function f with respect to temperature, we define the sensitivity as the partial differential with respect to temperatures,

$$\Delta r = f(s^*, T + \Delta T, l^*, \lambda) - f(s^*, T, l^*, \lambda). \tag{2.4}$$

The higher the sensitivity is, the more sensitive the function f is to the temperature. For a fixed spatial pixel (l_m is fixed), this sensitivity can be seen as a function of the surface temperature and wavelengths, which is shown in Fig. 2.1 with $\Delta T = 1K, s^* = 0.6, l^* =$

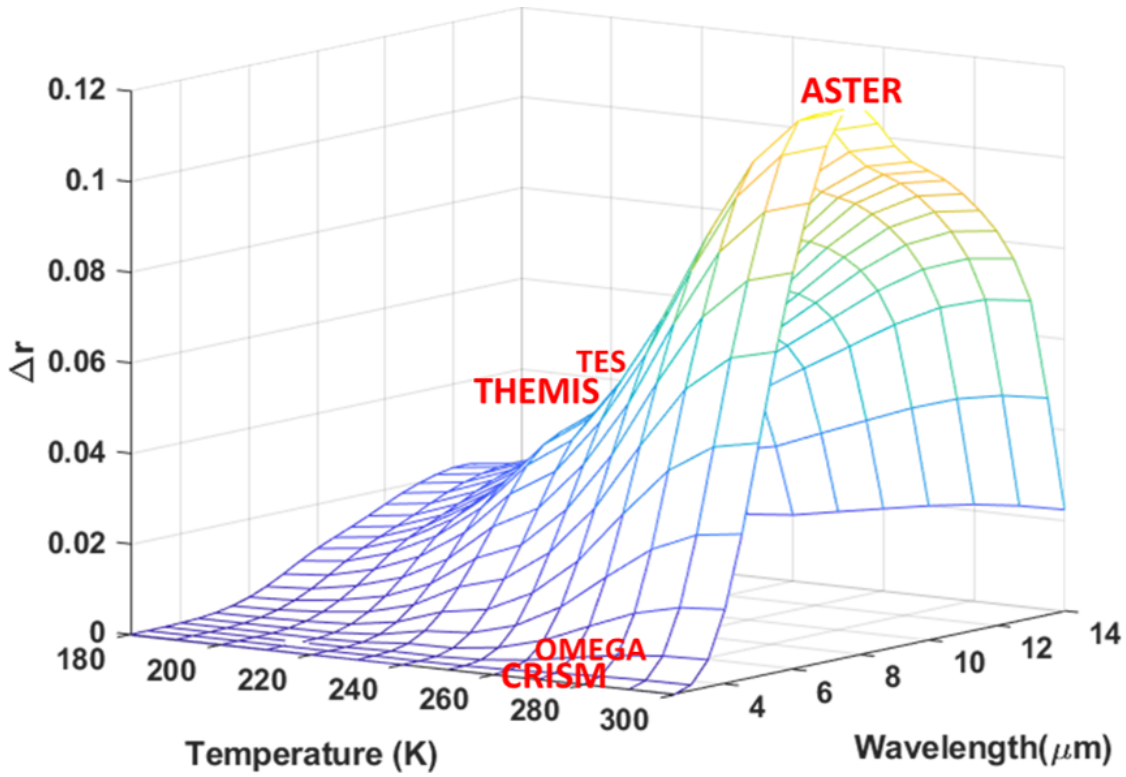


Figure 2.1: Sensitivity of radiance with respect to temperatures. Five instruments are labeled based on the covering spectral range and the estimated temperature range. THEMIS and TES cover the temperature from 240 to 280 K, which may be a little bit misleading in the figure.

$[1.62°, 82.91°, 108.79°, 0]$. In general, longer wavelengths (from 8 to 12 µm) have higher sensitivity than shorter wavelengths (from 3 to 5 µm). Previous NN methods [25,26,31] applied to ASTER in fact try to solve the inverse function located in the peak yellow mesh grid (sensitivity higher than 0.1) in Fig. 2.1 with estimated temperature range from 270 to 310 K and wavelengths from 8 to 12 micrometers. A trained NN for ASTER can approximate this function very well because a small error in the estimated temperature can lead the corresponding radiance to an unreasonable range (compared with given training samples), which can back-propagate to update the estimated temperature. However, the surface of Mars ($200 \sim 290K$) is not as hot as Earth ($> 270K$), therefore, for most Mars spectrometers, the inverse function we need to solve is much harder than ASTER. For example, THEMIS covering wavelengths from 7 to 15 µm has sensitivity on average less than 0.04. Even worse, CRISM covering thermal bands from 2.5 to 4 µm has the sensitivity less than 10^{-4}. Although solving this problem for Mars is harder than for Earth, Fig. 2.1 still shows that, it is still possible to use NN directly for longer thermal wavelengths and to apply NN with extra information for shorter thermal wavelengths. Therefore, we introduce our NN method, Separating Temperature and Albedo by Neural Networks (STANN) in Section 2.2.

2.2 Methodology

Fig. 2.2 shows how our method works in general. The radiance cube (or processed cube) of one scene and corresponding geospatial information are inputs. STANN can output the estimated temperature mapping of this scene and the corresponding SSA cube. In the following, we introduce how to generate the NN model and why a spectra generator is necessary for both longer and shorter thermal wavelengths.

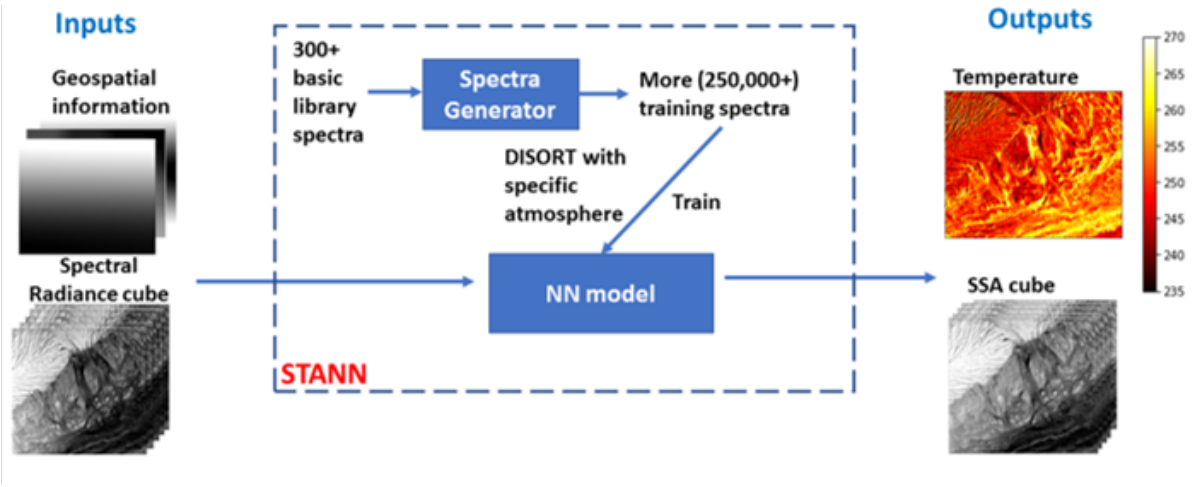


Figure 2.2: Framework of our Separating Temperature and Albedo by Neural Network (STANN) method. CRISM FRT0000B6F1 scene is used as an example.

2.2.1 Longer Thermal Wavelengths (>9 µm)

As explained in Section 1.1, the key idea is to approximate $f^{-1}(r_{mj}, l_m, \lambda_j)$ by some neural networks function $g(r_{mj}, l_m, \lambda_j)$ as closely as possible. Therefore, intuitively, the input nodes should be r_{mj}, l_m and λ_j. However, we can even make this model simpler because λ_j only depends on j and the function is applied independently for every m. Thus, the function we try to approximate can be written for all N bands ($j = 1, 2, ..., N$) in

$$[s_1, ..., s_N, T] = g(r_1, ..., r_N, e, i, \phi, p). \tag{2.5}$$

Mao et al. [25] in fact follows this framework but with a different forward model. We will introduce more details about how to apply this framework to a specific instrument such as THEMIS or TES in another paper [64] . Here, we only give a general model.

Input Nodes 11 nodes: Radiances at these 7 bands, emission angle, incident angle, phase angle, pressure.

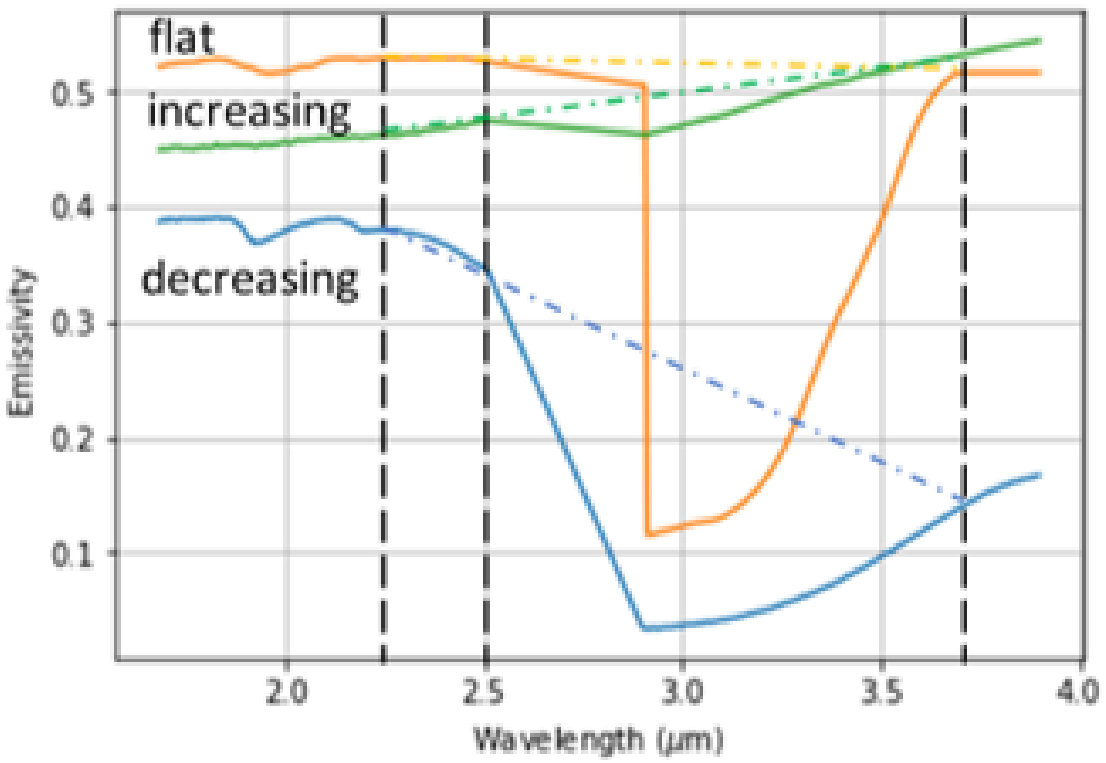


Figure 2.3: Three different spectra shapes based on the linear projection assumption.

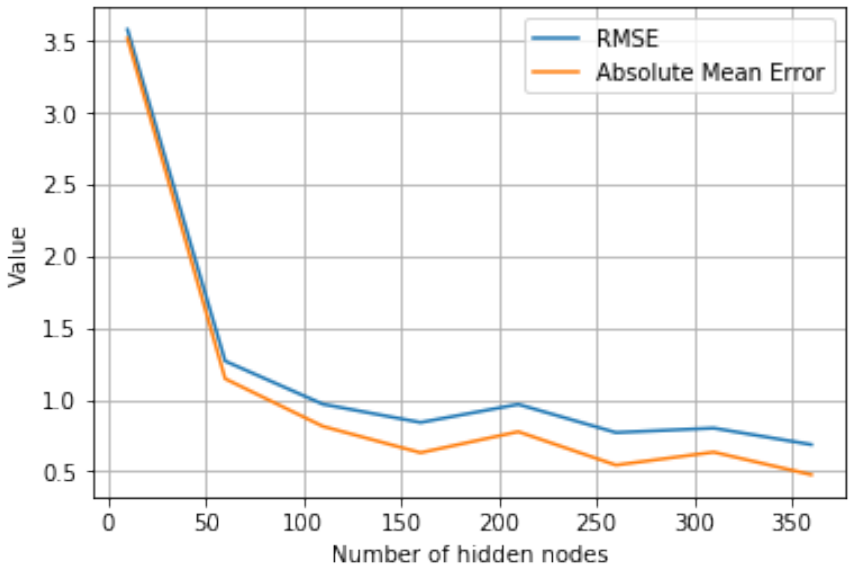


Figure 2.4: Simulation Performance of neural networks for THEMIS changes with the number of hidden nodes. RMSE means root mean square error of temperatures.

Output Nodes 8 nodes: SSAs at these 7 bands and temperature.

Hidden Nodes Fig. 2.4 shows that increasing hidden nodes reduces both RMSE and AME sharply for less hidden nodes, and finally converges to around 0.8 and 0.5 K for more hidden nodes. In another words, one-layer NN for the temperature range and wavelength range THEMIS covers, can perform as good as 0.5 K for the mean error. Combining with Fig. 2.1 where THEMIS has lower sensitivity than ASTER, it is not surprising that Mao et. al can have accuracy as low as 0.24 K for ASTER [25]. However, considering that more hidden nodes lead to a higher complexity of NN model which is not robust to noises [65], we select 150 as the number of hidden nodes for THEMIS.

Training Samples Most NN researches on ASTER ignore the importance of "enough" training samples because in practice, measured library spectra are limited. For example, Mao et. al used a complicated 2-layer neural network with 800 hidden nodes each layer, but

trained it on less than 8000 samples [25]. Fig. 2.5 shows how the performance changes over the number of training samples based on similar simulations above. Two curves of RMSE and AME are a little bit overlapped with each other due to a larger display range. We can find that for our one-layer neural network with 150 hidden nodes, at least 60,000 training samples are enough to guarantee the NN is well trained.

However, it is even harder to obtain measured spectra on Mars than on Earth. Therefore, we need to generate enough training samples (at least 60,000) based on limited measurements. In our case, the basic library spectra (363 in total) we use include. The basic 363 library spectra are from multiple sources including Mars analog materials by Ehlmann and Edwards [66], ASTER spectral library, USGS by Baldridge et al. [67] and some tephra and ash samples from Mauna Kea, basaltic glass, Martian meteorites, olivine, pyroxene, clays, and carbonates provided by Dr. Richard Morris.

We can generate more by randomly selecting Q library spectra and linearly adding them together with random generated proportions w_i as shown in (2.9). In practice, we generates 150,000 training samples, much more than 60,000, to increase the robustness of training process and Q is set to 10.

$$\begin{aligned} &\textit{Basic spectra: } A = a_1, a_2, ..., a_{363} \\ &\textit{Generated spectra: } B = b_1, b_2, ..., b_{60,000} \\ &\forall n \quad b_n = \sum_{i=1}^{Q} w_i a_i \ \textit{ with } \ \sum_{i=1}^{Q} w_i = 1 \end{aligned} \tag{2.6}$$

Framework of NN The one-layer neural network to separate the temperature and single scattering albedos for long thermal infrared wavelengths such as THEMIS has the framework

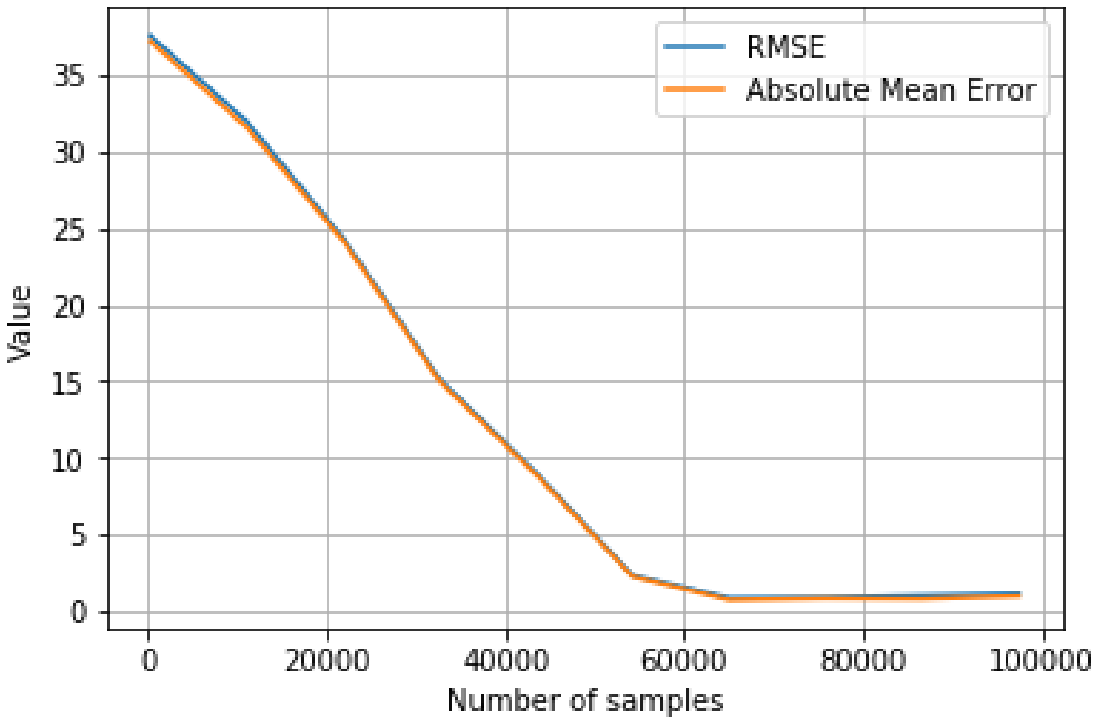


Figure 2.5: Simulation Performance of one-layer neural network for THEMIS changes with the number of training samples. RMSE means root mean square error of temperatures.

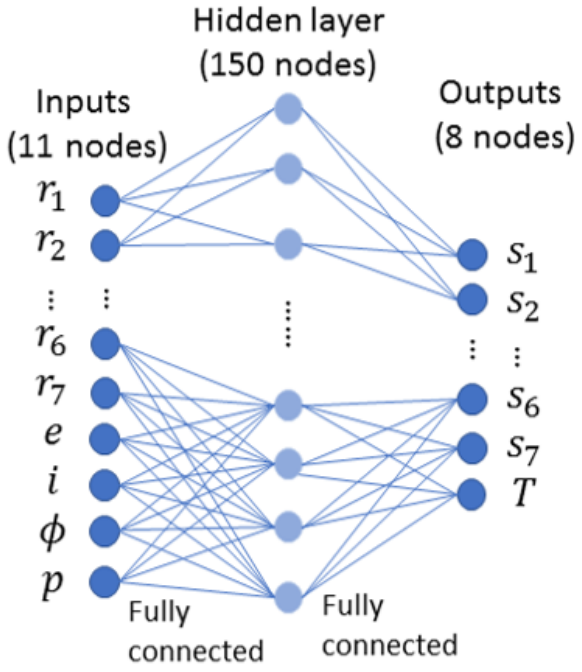


Figure 2.6: Framework of our one-layer neural network for THEMIS.

as shown in 2.6. The activation function is chosen as reLU defined as $f(x) = max(0, x)$ for all hidden nodes. Both inputs and outputs are scaled to the range [0,1] to guarantee the assumption of NN. Therefore unknown parameters for this framework are the weights on the edges.

Training Process of NN To train our one-layer NN for THEMIS, we estimate the 7.93 to 12.57μm SSA spectra and temperature for each spatial pixel m and compare these to input values, using a backpropagation method (shown in Fig. 2.10) to minimize the sums of squares of deviations between actual and predicted values. The gradient method we use is Adam [68] aimed to converge faster. After several iterations, the weights on the edges converge and this one-layer NN is ready to apply. Regularization is needed to avoid NN over-fitting of the training set. An L-2 norm regularization is used and the regularization weight is chosen by cross validation.

2.2.2 Shorter Thermal Wavelengths (from 2.6 to 9 µm)

Although THEMIS and TES have a lower temperature range than ASTER, due to longer thermal infrared wavelengths they cover, the relatively high sensitivity guarantees a theoretical good accuracy. Unlike for CRISM, according to the extremely low sensitivity (about 1/100 of THEMIS sensitivity) in Fig. 2.1, it is almost impossible to separate the surface temperature and SSA, just based on radiance only. Thus more prior knowledge is required.

Clark et al. [30] propose an assumption that there exists a linear projection relationship between shorter thermal wavelengths (from 2.6 to 5 µm) and non-thermal wavelengths (from 1.6 to 2.5 µm) as shown in Fig. 2.3. Although this assumption might not be strictly true, it is still worth considering this assumption as the prior knowledge. In our method, we use it as a soft constraint. In another words, we constrain the SSA spectra $[s_1, ..., s_N]$ retrieved by NN to have shapes similar to the training spectra.

Neural Network

We define the SSA spectra from 1.6 to 4 µm as $[s_1, s_2, ..., s_N]$ at N bands with wavelengths $\lambda_1 < \lambda_2 < ... < \lambda_N$ and the corresponding radiance as $[r_1, r_2, ..., r_N]$. To separate near-infrared and short infrared, we use 2.5 µm as the threshold to define two vectors: $SSA_{pre} = [s_1, ..., s_C]$ and $SSA_{post} = [s_{C+1}, ..., s_N]$ where C is an integer such that $\forall j < C$, $\lambda_j < 2.5$ µm and $\forall k \geq C$, $\lambda_k \geq 2.5$ µm. Different from Section 2.2.1, we use the pre-processed radiance, spectral IOF (radiance/solar radiance). Similarly, we can also have IOF_{pre} and IOF_{post}. As we discuss in Section 1.1, SSA_{pre} can be computed from (2.3), thus, we only need to estimate SSA_{post} and T. Therefore, with the same simplification used in Section 2.2.1 (2.2) can be

remodeled as

$$\begin{aligned} &[SSA_{post}, T] = f^{-1}(IOF_{post}, l) \\ &s.t. \quad [SSA_{pre}, SSA_{post}] \textit{ makes a real spectrum.} \end{aligned} \tag{2.7}$$

Similarly, the neural networks try to approximate the solution in (2.7) by a function g as

$$[SSA_{post}, T] = g(IOF_{post}, SSA_{pre}, e, i, \phi, p). \tag{2.8}$$

Thus, we can build a one-layer neural network to learn this function g.

For CRISM, we utilize 320 bands from 1.4 to 3.85 µm, thus $N = 320$ and $C = 169$.

Input Nodes 324 nodes: IOF_{post} at 151 bands, SSA_{pre} at 169 bands, emission angle, incident angle, phase angle, pressure.

Output Nodes 152 nodes: SSA_{post} 151 bands and temperature.

Hidden Nodes According to K. Funahashi [62] and K. Hornik [63], any continuous function on a bounded interval can be approximated by a single hidden layer neural network. It is reasonable to assume the inverse function f^{-1} we try to approximate is continuous and bounded, therefore, we think a one-layer NN is enough. The number of hidden nodes is selected by simulations as shown in Fig. 2.7. We generate two data sets individually and use one as the training set and another as the validation set. The root mean square error (RMSE) and the absolute mean error (AME) are calculated from the difference between

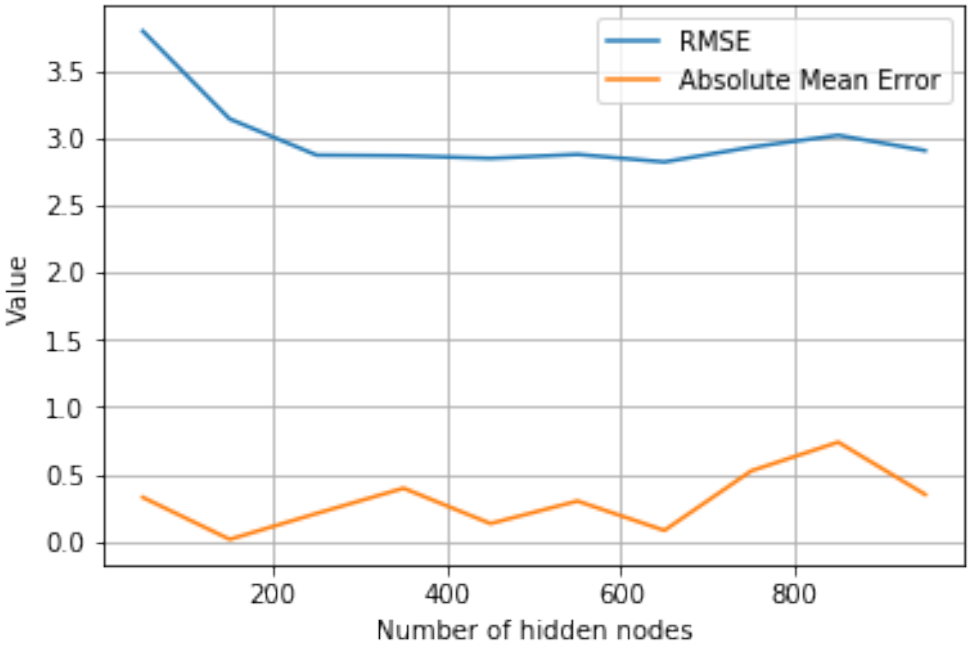


Figure 2.7: Simulation Performance of neural networks for CRISM changes with the number of hidden nodes. RMSE means root mean square error of temperatures.

estimates from NN based on the training set and truths of the validation set. Increasing the number of hidden nodes reduces the RMSE and finally converges to around 2.9 K. As for AME, it seems like 50 hidden nodes are enough to make error less than 1 K. In another words, the main problem of NN for CRISM is the uncertainty not the bias. In summary, a one-layer NN for the temperature range and wavelength range CRISM covers, can perform as good as 0.5 K for the mean error. Combining with Fig. 2.1 where CRISM has much lower sensitivity (10^{-4}) than ASTER (0.12), it is not surprising that Mao et. al can have accuracy as low as 0.24 K for ASTER [25]. However, considering that more hidden nodes lead to a higher complexity of NN model which is not robust to noise [65], we select 500 as the number of hidden nodes for CRISM.

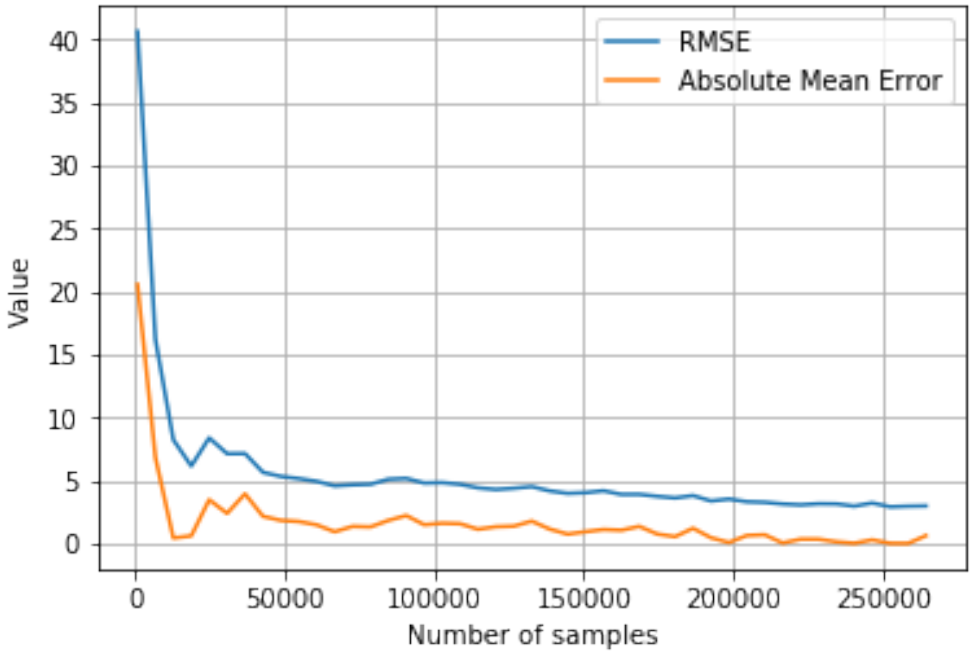


Figure 2.8: Simulation Performance of one-layer neural network for CRISM changes with the number of training samples. RMSE means root mean square error of temperatures.

Training Samples Most NN researches on ASTER ignore the importance of "enough" training samples because in practice, measured library spectra are limited. For example, Mao et al. used a complicated 2-layer neural network with 800 hidden nodes each layer, but trained it on less than 8000 samples [25]. Fig. 2.8 shows how the performance changes over the number of training samples. As we expect, more training samples are needed to guarantee the NN is well trained (at least 150,000).

Framework of NN The one-layer neural network to separate the temperature and single scattering albedos for long thermal infrared wavelengths such as CRISM has the framework as shown in Fig. 2.9. The activation function is chosen as reLU defined as $f(x) = max(0, x)$ for all hidden nodes. Both inputs and outputs are scaled to the range [0,1] to guarantee the assumption of NN.

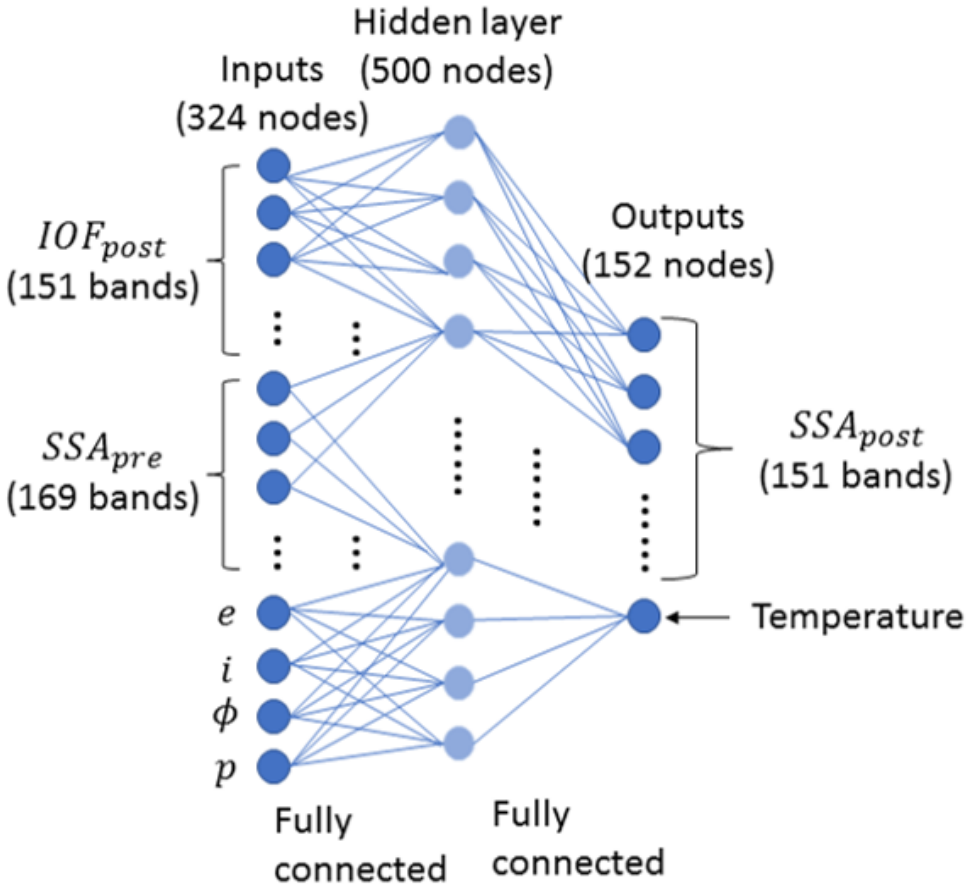
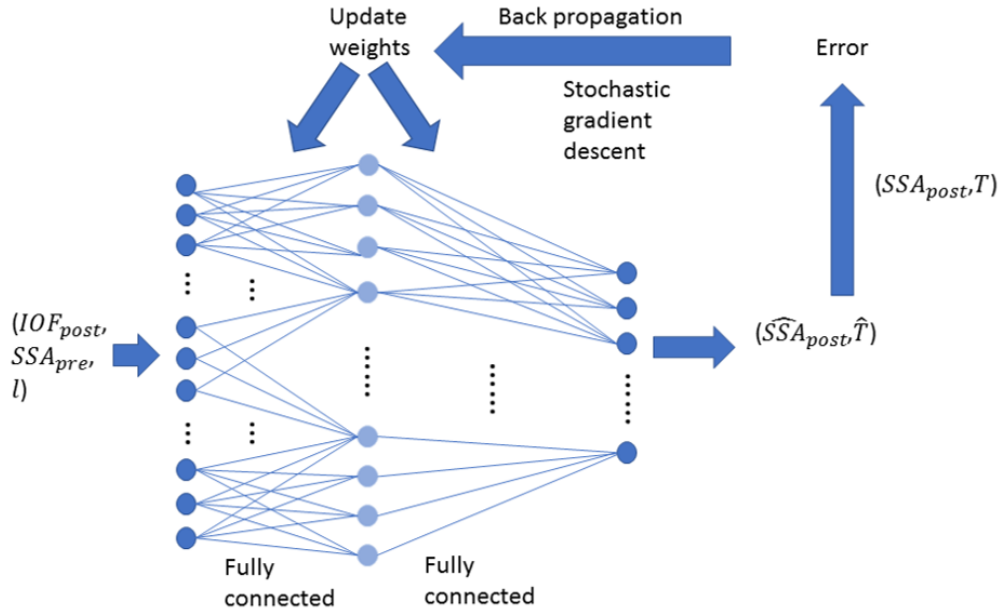


Figure 2.9: Framework of our one-layer neural network for CRISM.

Figure 2.10: Backpropagation of our one-layer neural network for CRISM.



Training Process of NN To train our one-layer NN for CRISM, we estimate the SSA spectra and temperature for each spatial pixel m and compare these to input values, using a backpropagation method (shown in Fig. 2.10) to minimize the sums of squares of deviations between actual and predicted values. The gradient method we use is Adam [68] whose goal is to converge faster. After several iterations, the weights on the edges converge and this one-layer NN is ready to apply. Regularization is needed to avoid NN over-fitting of the training set. An L-2 norm regularization is used and the regularization weight is chosen by cross validation.

2.3 Generator Selection

As we analyzed before, aimed to well train our neural network, we need enough training spectra. However, it is even harder to obtain measured spectra on Mars than on Earth.

Therefore, we need to generate enough training samples (at least 150,000) based on limited measurements.

2.3.1 Linear Generator

One intuitive way is to use the linear generator. We can generate more by randomly selecting Q library spectra and linearly adding them together with random generated proportions w_i as shown below,

$$\begin{aligned} &\textit{Basic spectra: } A = a_1, a_2, ..., a_{363} \\ &\textit{Generated spectra: } B = b_1, b_2, ..., b_{60,000} \\ &\forall n, \quad b_n = \sum_{i=1}^{Q} w_i a_i \ \textit{ with } \ \sum_{i=1}^{Q} w_i = 1. \end{aligned} \tag{2.9}$$

In practice, we generate 250,000 training samples, much more than 150,000, to increase the robustness of the training process and Q is set to 10. Fig. 2.11 compares the generated spectra with the original library spectra. The linear projection assumption seems to softly hold for all original library spectra (top left). The linear generator (top middle) generates varying spectra with similar shapes but due to the random generator of weights, it is possible to generate some spectra (like the orange and green curves) which do not follow the linear projection assumption.

2.3.2 AutoEncoder

Fig. 2.12 shows the framework of a single autoencoder [69]. It contains two parts: encoder and decoder. The encoder is a fully connected multiple-layer neural network with decreasing

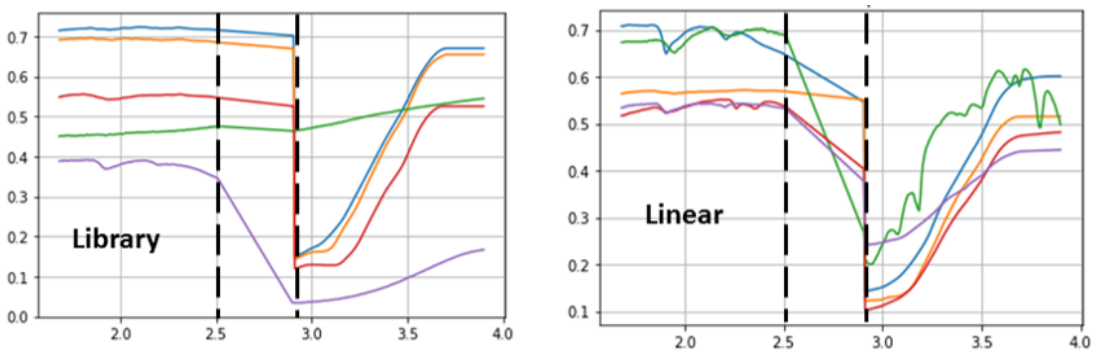


Figure 2.11: Comparison between the library spectra and generated spectra by the linear generator.

hidden nodes for each layer, usually half of tahe previous layer. The decoder is also a fully connected multiple-layer neural network with increasing hidden nodes, usually symmetric to the encoder. Between the encoder and the decoder, we have a latent layer with much less features. Typically, we input a library spectra and then the autoencoder is expected to output the same output. Our library spectra are sampled based on the CRISM bands, so the input and output nodes are 277. Thus we generate a five-layer encoder (including input and output layer) with 277, 128, 64, 32, 5 hidden nodes and a four-layer decoder (including output layer) with 5, 32, 64, 128, 277 nodes. Aimed to avoind the over-fitting of training process, we also attached a drop-out layer after every layer (except for the last layer) with 0.5 drop-out probability. A similar back-propagation method is used to update the weights in both encoder and decoder. After training, the decoder part is applied to generate more spectra by input the normal distributed latent features (with 0.5 mean and 0.1 variance).

The number of latent features is selected as 5 due to the cross-validation in Fig. 2.13. We notice that the sum of squared errors (SSE) decreases sharply around 4 and becomes stable at 5. Thus we choose 5 nodes in the latent layer.

Fig. 2.14 compares the generated spectra with the original library spectra. The autoencoder generates spectra all following the assumption, but only have the decreasing shape, which

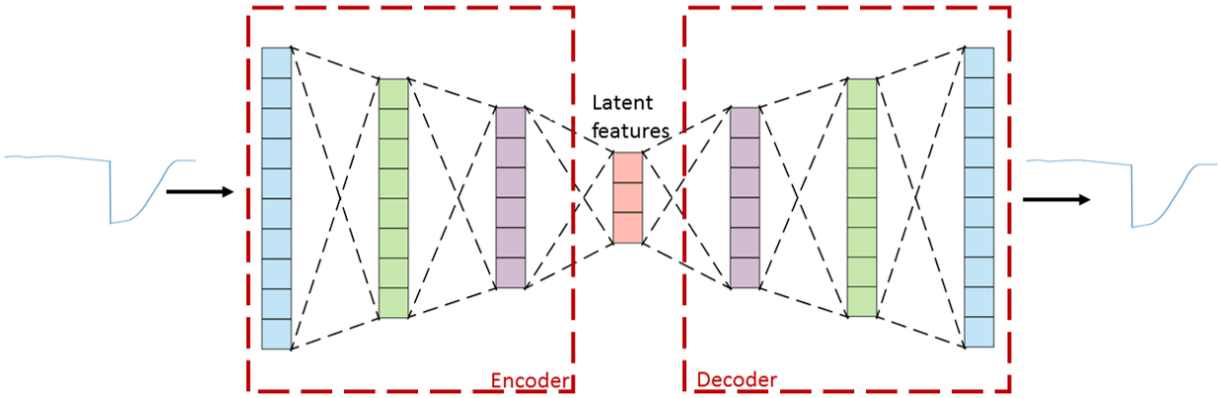


Figure 2.12: The framework of autoencoder for the one-dimensional spectrum.

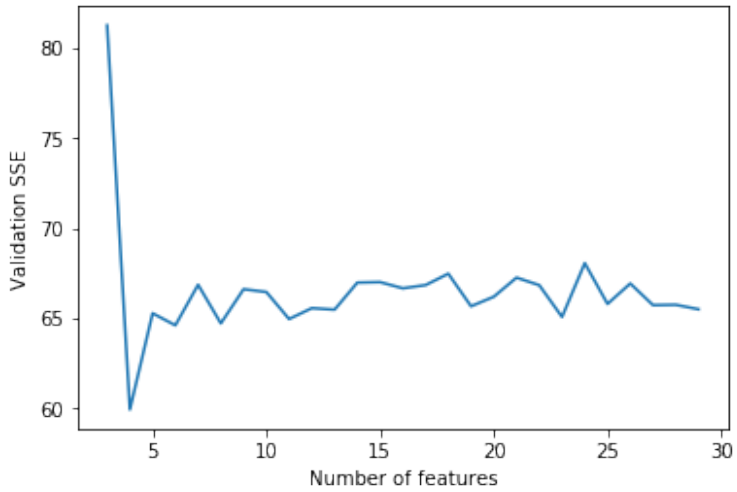


Figure 2.13: Performance changes with the increasing latent features.

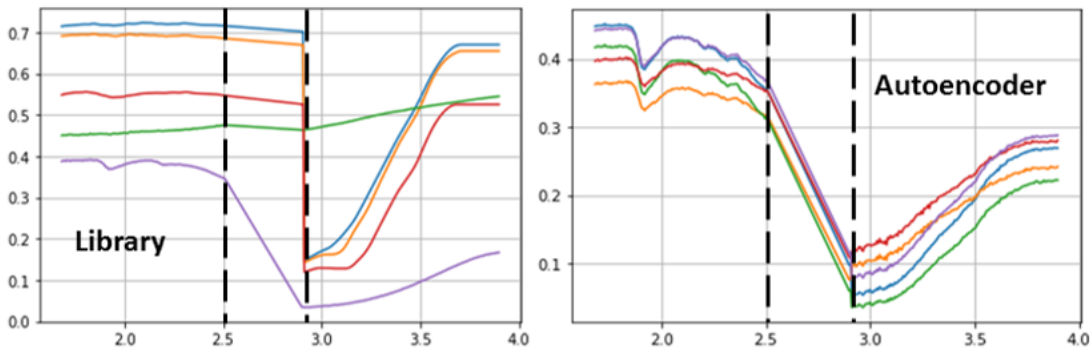


Figure 2.14: Comparison between the library spectra and generated spectra by the autoencoder generator.

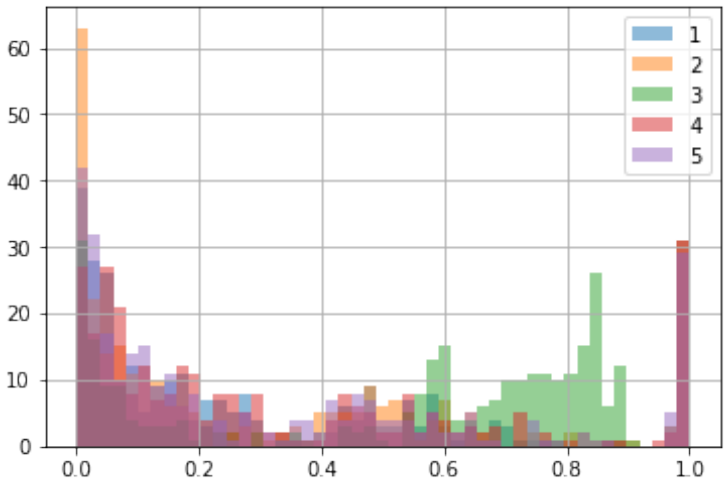


Figure 2.15: The histogram of five latent variables for the library set for one run.

is in fact rare in the original library. Combining with the latent features' histograms of the training set in 2.15, We can find that the distributions of five latent features are not normal. But when we generate the spectra, we assume the latent variables are normal distributed. This mismatch leads to too many decreasing-shape spectra, which in fact is rare in the library spectra.

2.3.3 Variational AutoEncoder

Different from the simple autoencoder, variational autoencoder (VAE) adds constraints to the lower dimensional space, which forces the latent features retrieved from the training set are (approximately) normal distributed (other distributions can also be assumed) in the lower dimensional space. The framework of VAE is shown in Fig. 2.16.

Both encoder and decoder are still symmetric fully connected neural networks as shown in Fig. 2.12. A training spectrum is also first input to the encoder. After the encoder, the mean and variance of latent features are retrieved and generated features based on these mean and

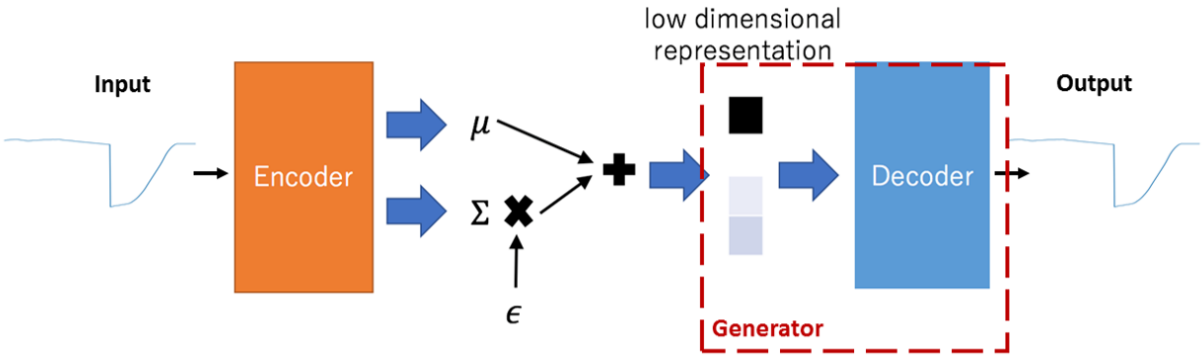


Figure 2.16: The framework of variational autoencoder for the one-dimensional spectrum.

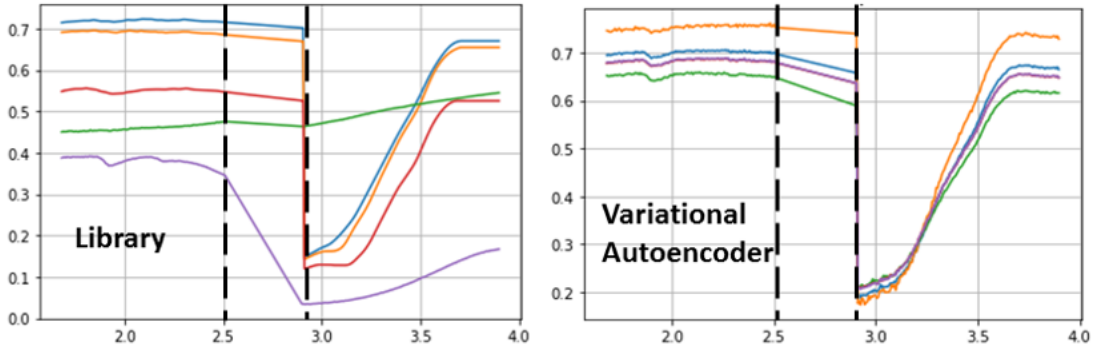


Figure 2.17: Comparison between the library spectra and generated spectra by the autoencoder generator.

variance are input to the decoder aimed to recover the input spectra. When training this VAE, we need to minimize the combination of the difference between output and input and the KL distance in the retrieved mean and variance [70].

After training, the VAE is ready to use. Thus we randomly generates normal distributed mean and variance and input them to the generator in Fig. 2.16. Fig. 2.17 compares the generated spectra with the original library spectra. As we expect, VAE generates spectra with the most common shape in the library, in our case, the flat shape. This is because VAE forces the latent features from training set be normal distributed and the most common shape will be very close to the center. And when we randomly generates latent features, we have high probability to generate shape close to the center (normal distribution), thus the

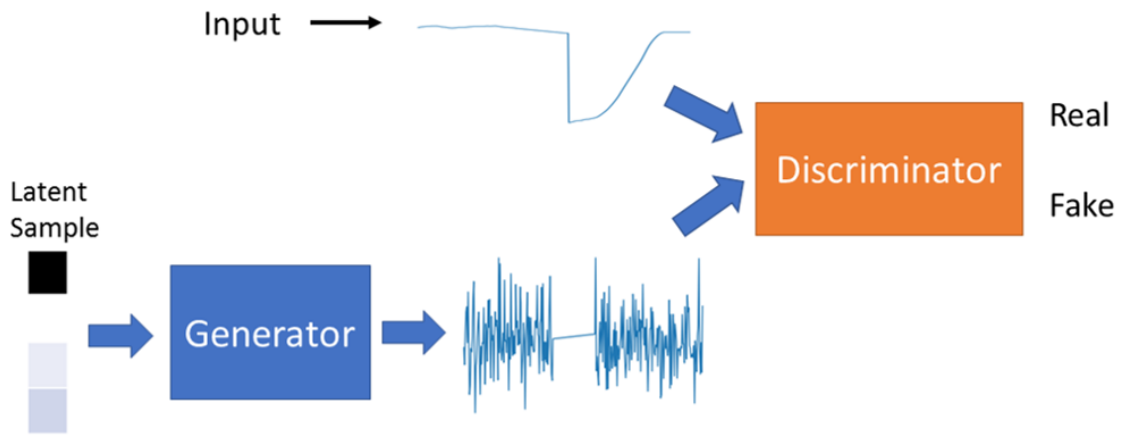


Figure 2.18: The framework of generative adversarial network for the one-dimensional spectrum.

generated spectra should have the most common shape in the training set. This property of VAE guarantees that generated spectra must have reasonable shape if the training set is reasonable selected, which makes up the shortage of autoencoder before. However VAE also has its own disadvantage that small absorption features (details in the spectrum) are similar for all generated spectra as shown in Fig. 2.17.

2.3.4 Generative Adversarial Network

Previous both autoencoder and variational autoencoder uses the difference between input and output as the quantity of cost function to run a generator. We uses L2 norm to quantify the difference but others can also be used. Different cost functions will lead to different generators with varying performances. Comparing all cost functions to get the best one will be very exhausting, thus generative adversarial network (GAN) proposes an idea that training both the generator and the cost function (which is discriminator in GAN) from data itself. Fig. 2.18 shows the framework of GAN for our case.

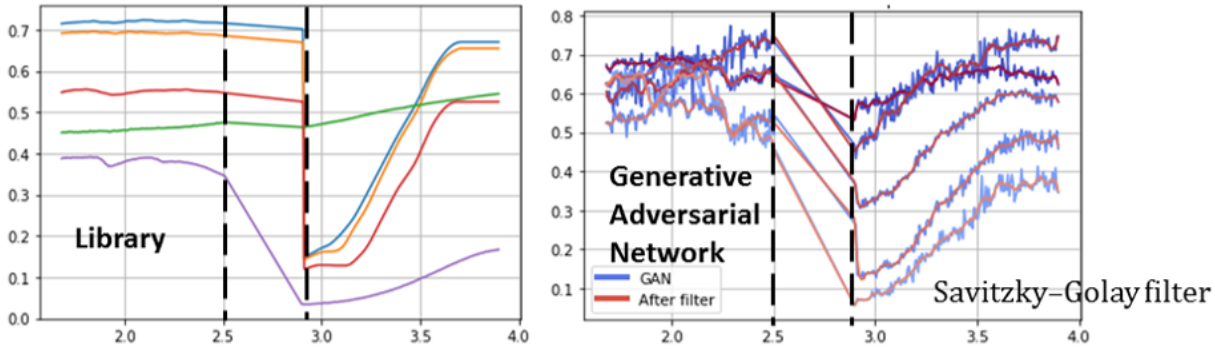


Figure 2.19: Comparison between the library spectra and generated spectra by the autoencoder generator.

GAN usually contains two main parts: generator and discriminator. We first randomly generates some "fake" spectra from generator and input them and real spectra to the discriminator. Then discriminator can be trained to distinguish real spectra and fake spectra well. When well trained, discriminator is used as the standard to train generator, aimed to generate spectra can "cheat" the discriminator in this iteration. Then when the generator is well trained for this discriminator, we train the discriminator again to increase the ability of separating generated spectra and real spectra. We can train discriminator and generator iteratively for enough iterations [71]. Then the GAN is ready to use.

Fig. 2.19 compares the generated spectra with the original library spectra. GAN generates spectra with correct and varying shapes and a lot more details (absorption features) even after being filtered by the Savitzky-Golay filter. Compared with VAE, generated spectra from GAN have more small absorption features and higher varying in shape.

2.3.5 Performance Comparison

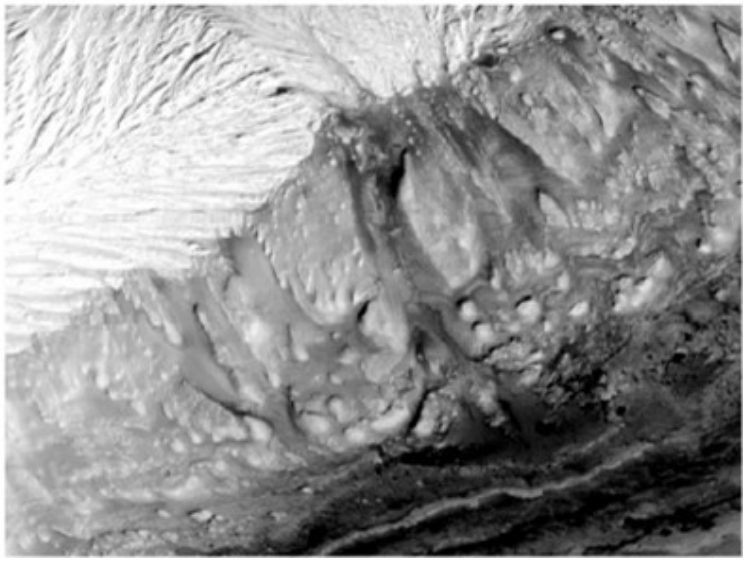
As we analyze before, we assume retrieved SSA spectra from NN have a similar shape as the library spectra, thus when generating more training spectra, the shape of generated

training spectra should be "real". Therefore, four generators: linear generator shown in (2.9), autoencoder [69], variational autoencoder (VAE) [70], and generative adversarial network (GAN) [72] are tried on the scene FRT0000B6F1 to test which one should be used in practice.

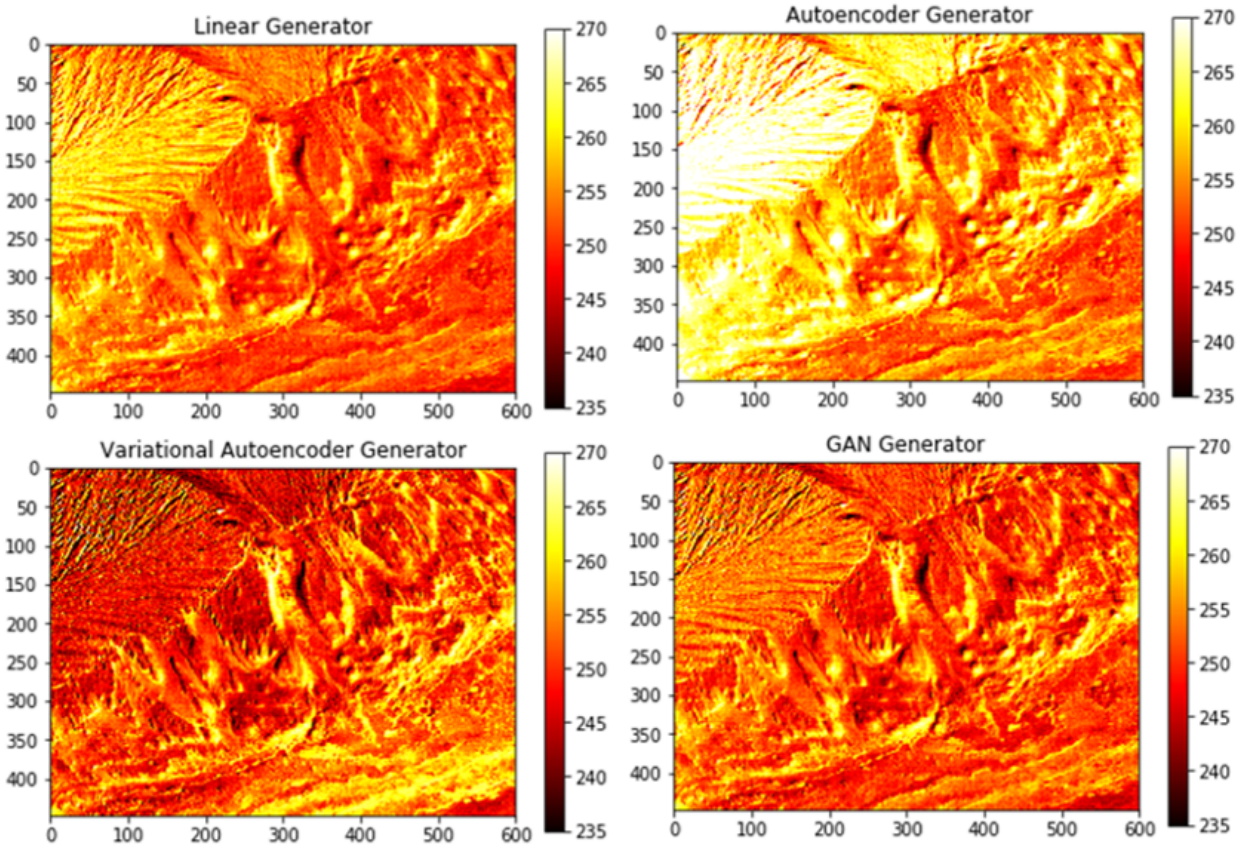
Fig. 2.20b shows the temperature mappings of the scene FRT0000B6F1 based on generated spectra in Fig. 2.20a. The image On the left is the single scattering albedo for this scene at near-infrared wavelengths (1.96 µm). It is clear to find that the top left area of this scene has higher reflectance and the bottom right area has lower reflectance. As we know, darker areas with lowers reflectances can absorb more heat, thus the surface temperatures should be higher. Comparing four temperature mappings on the right, we can find only VAE generator learns this information successfully. Both the linear generator and the autoencoder generator are misled by the reflectances (higher reflectances have warmer temperatures in their results). GAN successfully avoids this, but still fails in the learning of the most common spectral shape in the original library set. Combining the analysis of Fig. 2.20a, we can draw a conclusion that, VAE should be selected as the training set generator due to the good learning ability of the pattern and the most common shape of targeted spectra should be included in the basic library spectra.

2.4 Performance Analysis

CRISM FRT000B6F1 (acquired during 2008) used to validate the performance of STANN, covers the northwest part of Gale crater and the path of the Curiosity rover as shown in Fig. 2.21. In this section, we validate both robustness and consistency of STANN. The robustness is tested based on analysis of the standard derivation of multiple runs and the consistency is analyzed by comparing with the Curiosity point measure and another independent temperature retrieved from Ashwin's thermal model [73].



(a) FRT000B6F1 reflectance scene



(b) Retrieved temperature for the; scene FRT0000B6F1 based on generated spectra above

Figure 2.20: Performance comparison for four generators.

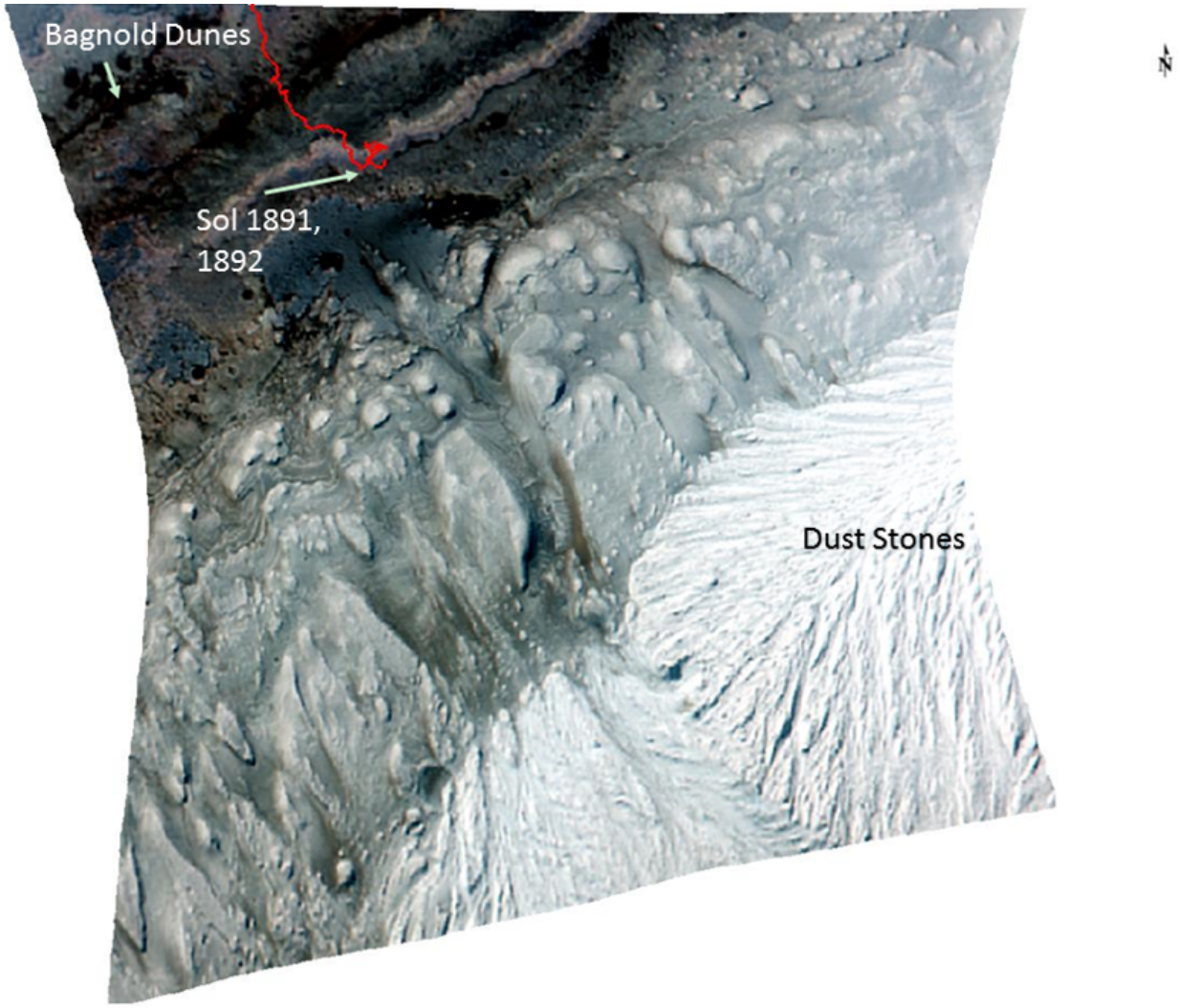


Figure 2.21: Relative Location of CRISM FRT000B6F1 and the Curiosity Rover Transverse Path. The red curve shows the Curiosity Rover path. The false colors RGB are set as 2.5, 1.5 and 1.0 µm. The labeled Sol 1891, 1892 are the location where the Curiosity Rover measured the diurnal surface temperature in Fig 2.28. The IOFs from Bagnold Dunes and Dust Stones are shown in Fig. 2.22.

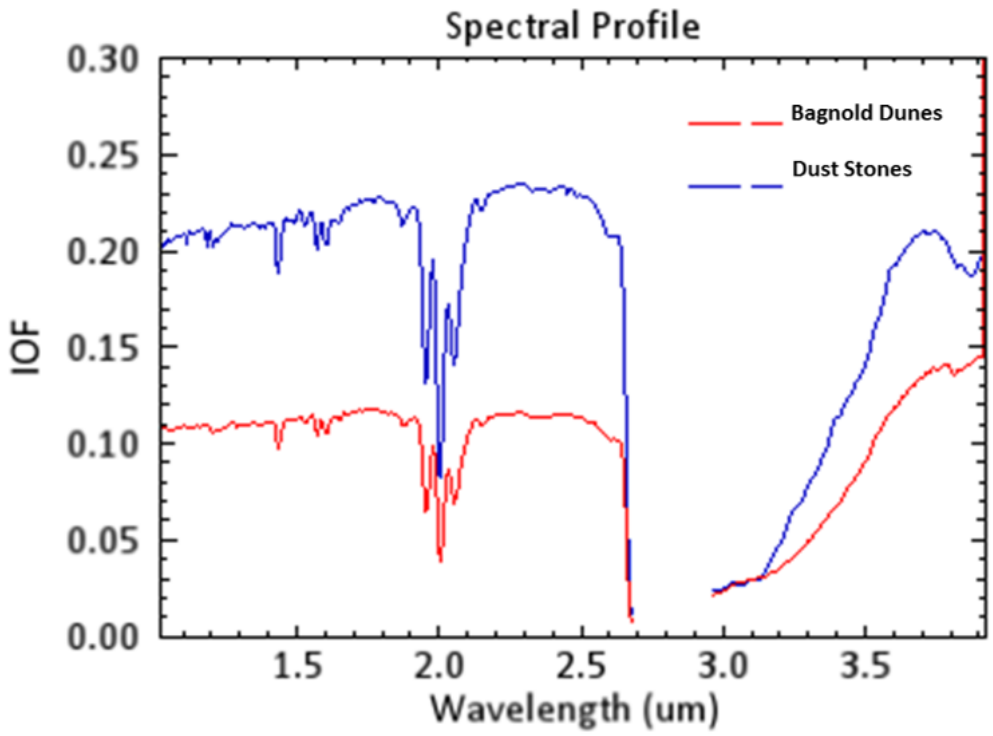


Figure 2.22: Two IOFs from Bagnold Dunes (red) and Dust Stones (blue) in Fig. 2.21 are plotted.

Fig. 2.22 shows IOFs from two areas in Fig. 2.21. One can find that the spectrum from Bagnold Dunes has higher value at 3.7 µm than 2.5 µm while the spectrum from dust stones has opposite tendency. In another words, compared with the their own solar IOFs before 2.5 µm, the IOF from Bagnold Dunes has much higher thermal radiance than one from the Dust Stones at the wavelengths from 3.0 µm to 4.0 µm. Therefore, the Bagnold Dunes should be warmer than the dust stones in the temperature mapping.

2.4.1 Robustness

Because both neural networks and VAE are not deterministic, STANN can output different results at every run for the same input. Therefore, we need to test the robustness of STANN for CRISM (FRT000B6F1).

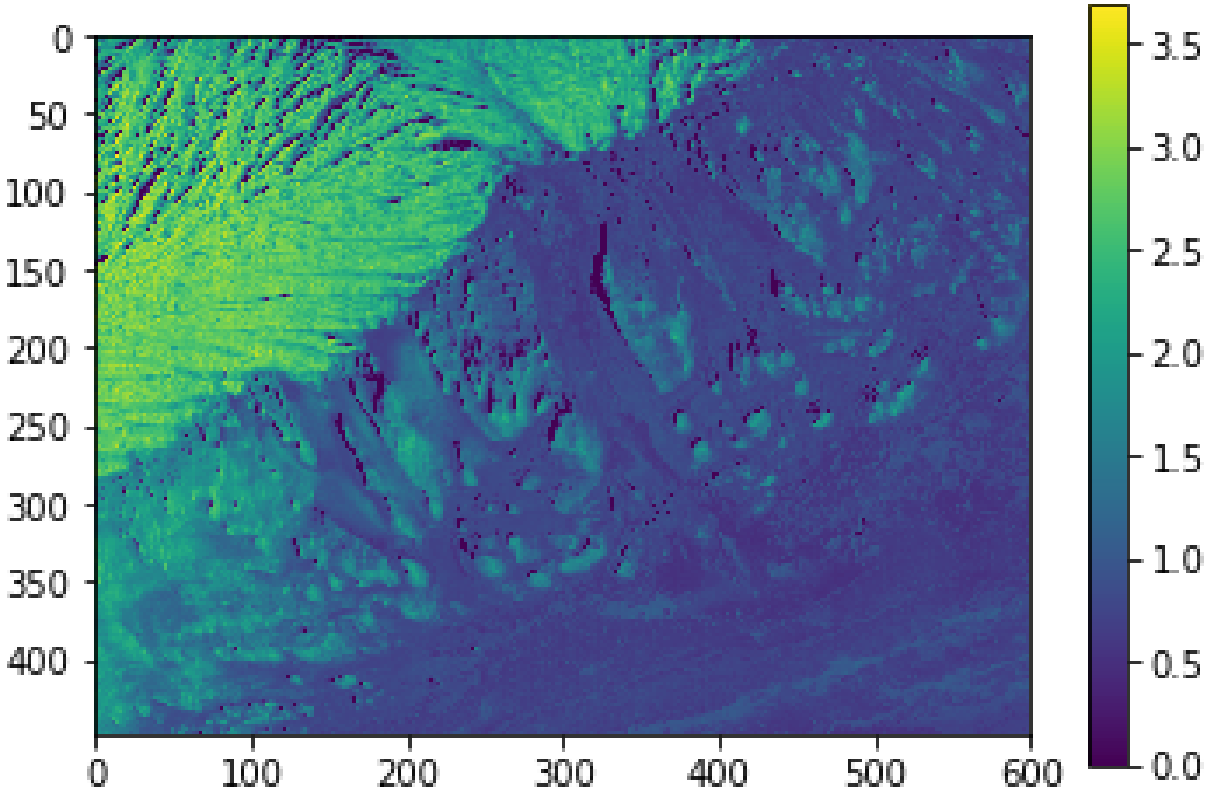


Figure 2.23: The mean standard derivation mapping for 20 runs of STANN.

Fig. 2.23 shows the mean standard derivation mappings for 20 runs of STANN for CRISM. The standard derivation of CRISM is small (sometimes up to 3.5 K) although the algorithm is not deterministic because the sensitivity figure in Fig. 2.1 shows CRISM has extremely small sensitivity. Moreover the standard derivations for most of areas are smaller than 1 K and the average is about 2 K, indicating that STANN for CRISM is also relatively very robust with the near-infrared spectra involved.

2.4.2 Consistency

Before consistency comparison, we first want to introduce how we retrieve another independent temperature mapping (called thermal model for short in the following) from the thermal model proposed by Vasavada et al. [73]. Their thermal model gives the relationship between the Lambert Albedo at the wavelength 1.0 µm, thermal inertia and the surface temperature. Fig. 2.24 shows the thermal inertia map from THEMIS and the corresponding Lambert albedo map at 1.05 µm. It is noticed that part of the scene does not have the corresponding thermal inertia. Thus the temperature mapping only has values in the overlapped area. The codes of this temperature generator can be found under my github.

Fig. 2.25 shows two temperature mappings generated independently from the thermal model and our STANN. In general, both temperature mappings have correct pattern and similar highest temperatures. However there also exist lots of differences in details, for example, thermal model T seems to be higher than STANN's T for some colder areas. Aimed to show it more clearly, we pick three horizontal lines (A, B, C in blue, red and green) and compare the horizontal values from these two temperatures in Fig. 2.26.

Both temperature maps show that the western area covered by line A is warmest and the overall area covered by line C seems to be coldest. For the warmest area, both CRISM and model data are in the range of 255 to 260 K. STANN temperature estimates for the regions covered by lines B and C are about 240 K while thermal model estimates are about 253 K. We believe this discrepancy is due to lack of discernable differences in thermal emission for the CRISM data, perhaps combined with the approximations used for thermal modeling, i.e., Lambert albedo assumption for surface properties, and retrieval of thermal inertias for these surfaces. We are pursuing sensitivity analyses to better understand the temperature differences.

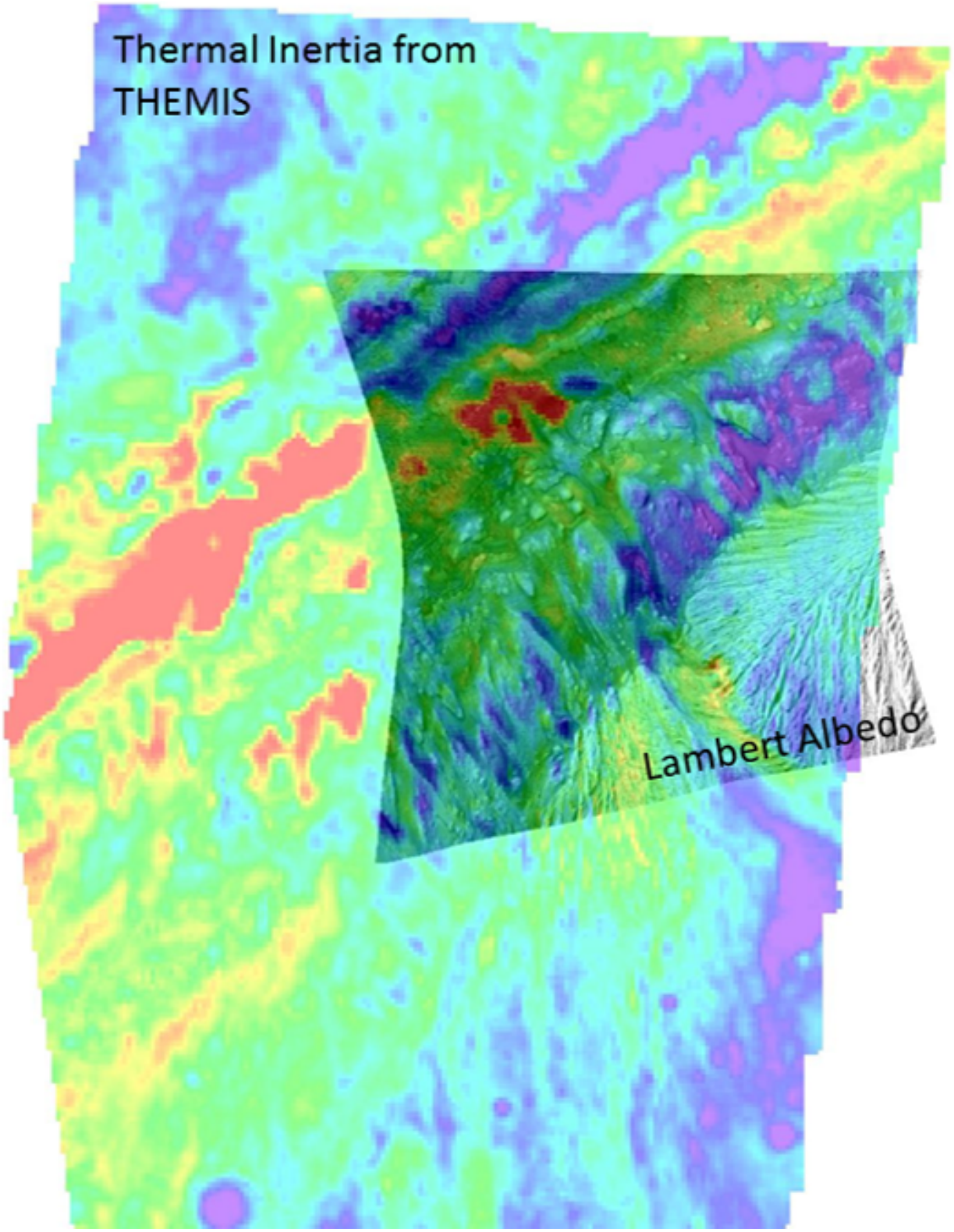


Figure 2.24: Thermal Inertia from THEMIS and the Lambert Albedo for FRT000B6F1. The thermal inertia has the range from 200 to 480 and the red and blue indicate the high and low thermal inertia areas. It is noticed that part of the scene does not have the corresponding thermal inertia. Thus the temperature mapping only has values in the overlapped area.

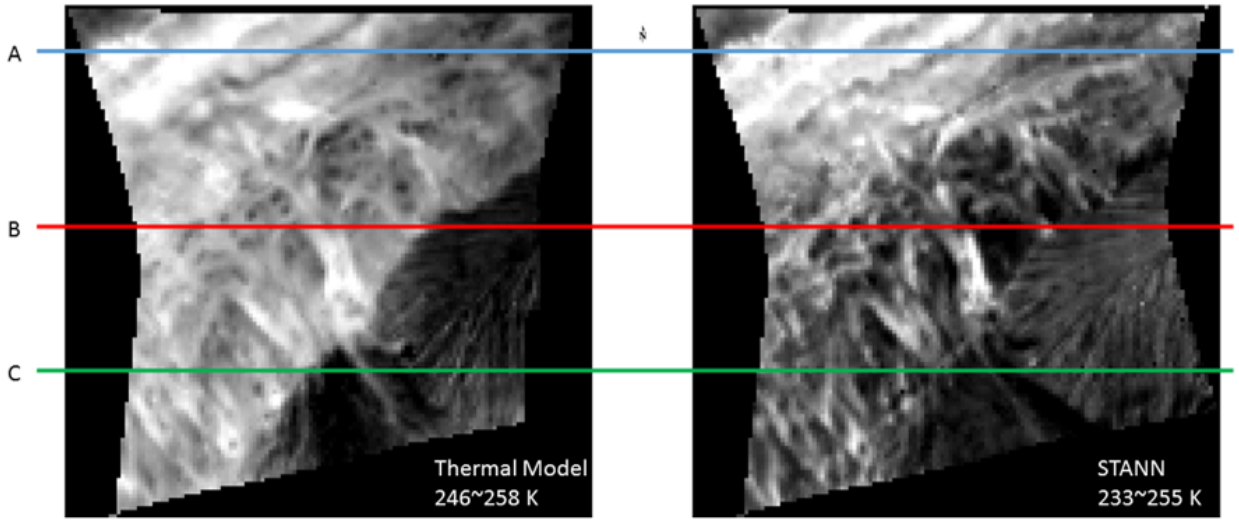


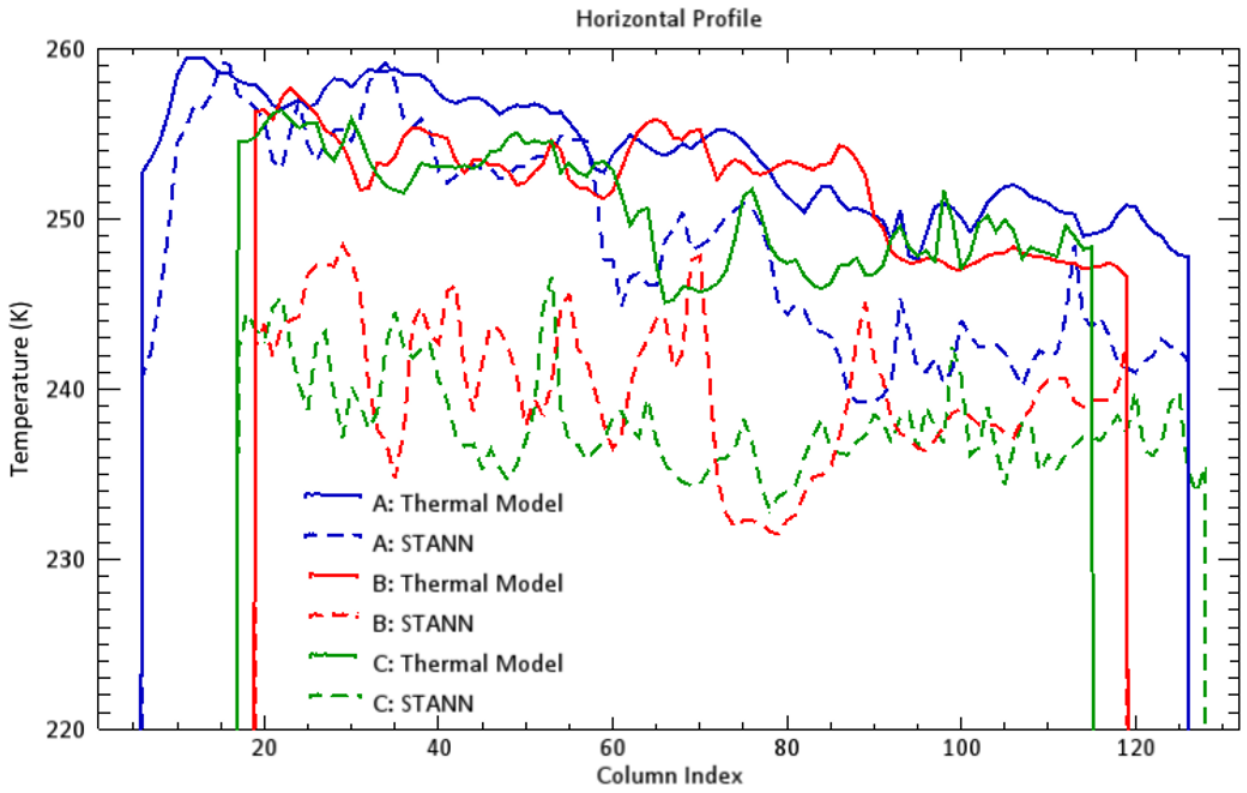
Figure 2.25: Two temperature mappings generated based on different methods independently are shown. The temperature in the left is from the thermal model proposed by Vasavada et al. and one in the right is from our STANN. Three horizontal lines are labeled to show locations where the horizontal comparison for these two mappings in Fig 2.26. One should notice that we do not model topography for scales less than 250 meters MOLA, thus the flat area (left area of the blue line) is more valid comparison than others.

One explanation for this difference is that because this area is very cold (colder than 255 K), 10 K difference in temperature can generate similar albedos due to the limitation of the Plank function property. Fig. 2.23) shows that for very cold area (about 240 K), even 15 K change in the estimated temperature can only lead a very small change in SSAs as shown in Fig. 2.27. This also consistent with our previous sensitivity analysis that for very cold area and CRISM spectral range, the sensitivity in Fig. 2.1 can be as small as 10^{-5}.

Fortunately, the colder area of FRT000B6F1 covers the location (Sol 1891, 1892) where the Curiosity rover measured the daily surface temperature curve.

Fig. 2.28 shows the diurnal surface temperature curves measured by the Curiosity rover at the point Sol 1891, 1892. Both curves are consistent with each other (similar pattern and temperature range). CRISM FRT0000B6F1 was observed at the solar time 15:29. Two cross

Figure 2.26: Three horizontal value comparisons are shown to see the consistency of the temperature from our STANN (in dashed line) and one from the thermal model (in solid line) proposed by Vasavada et al. The x-axis is the column index.



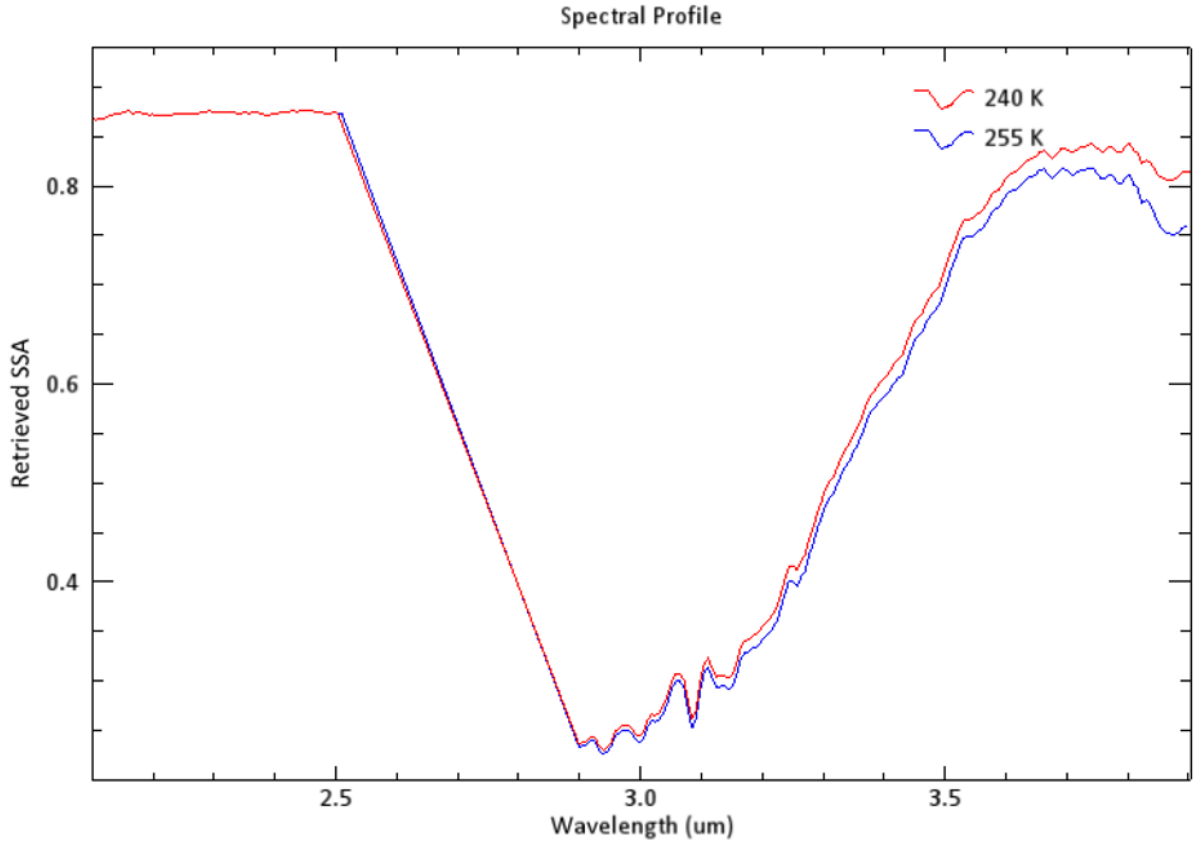


Figure 2.27: Two different SSAs retrieved from the same IOFs based on 240 K and 255 K.

points show estimates by STANN and the thermal model around Sol 1891, 1892. We can find that according to the Curiosity rover, the ground truth temperature should be around 242 K. STANN estimates it as 246 K, which has only 4 K difference while the thermal model estimate it as 255 K. It is partially verified that for the colder temperature, STANN can estimate better than the thermal model. Combining the analysis before (Section 2.4.1), 4 K difference is very hard to estimate at such a cold temperature but fortunately it would not affect on the retrieved SSAs too much(refer to Fig. 2.27).

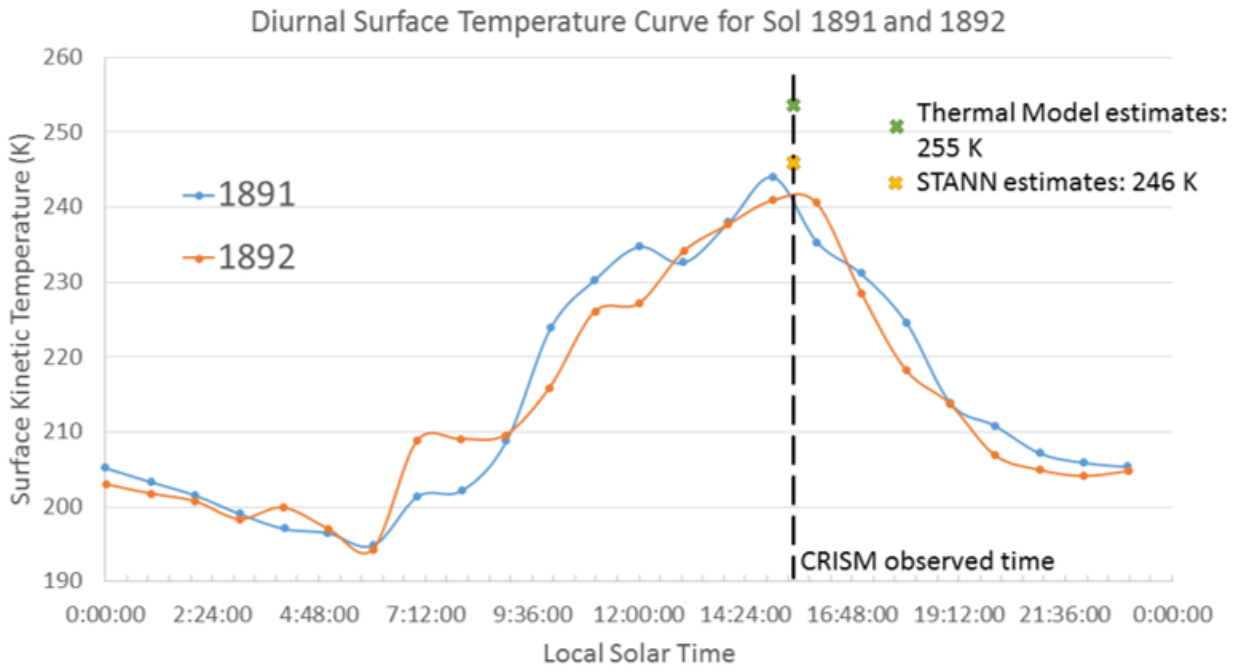


Figure 2.28: Diurnal surface temperature curve measured by the Curiosity Rover. The estimate from STANN is labeled in yellow cross and one from thermal model is labeled in green cross.

Aimed to verify this explanation, FRT0000C0EF is selected, a flat and warm scene where the Curiosity Rover has been [1].

In summary, we can draw a conclusion that in general our STANN can generate a robust temperature mapping, which is both consistent with the thermal model for warmer areas and with the Curiosity point measure for colder areas.

2.5 Discussion

We define the sensitivity of radiance with respect to temperatures, which describes how hard the neural network can approximate the inverse function to get the surface temperature. A general method, Separating Temperature and Albedo by Neural Networks (STANN), is proposed to retrieve the Single Scattering Albedo and the surface temperature from radiance. We also theoretically analyze the difficulty for different cases (different temperature ranges and different wavelengths) and explain the reason why previous similar methods can work well. Then the STANNs are derived and built for the hardest Mars spectrometer CRISM. Performances for both instruments are verified from the robustness and the consistency. It is noticed that STANN for CRISM can obtain a relatively robust and correct temperature mapping based on limited information. The local incidence angle can be computed from the elevation file, which corrects the spatial artifacts in generating albedos.

[1] More result analysis is coming for this scene

Chapter 3

Reconstruction and Denoising Method HyBER for Hyperspectral Image Data

3.1 Methodology

Fig. 3.1 shows the diagram of our hypothesis-based estimation with regularization (HyBER) for general cases. Hyperspectral data are input to HyBER, and based on our hypothesis testing method, the best statistical model is selected that describes the hyperspectral data (either Poisson or Gaussian distribution). A maximum loglikelihood algorithm is then applied for the selected distribution to produce a denoised and de-blurred hyperspectral mapped image.

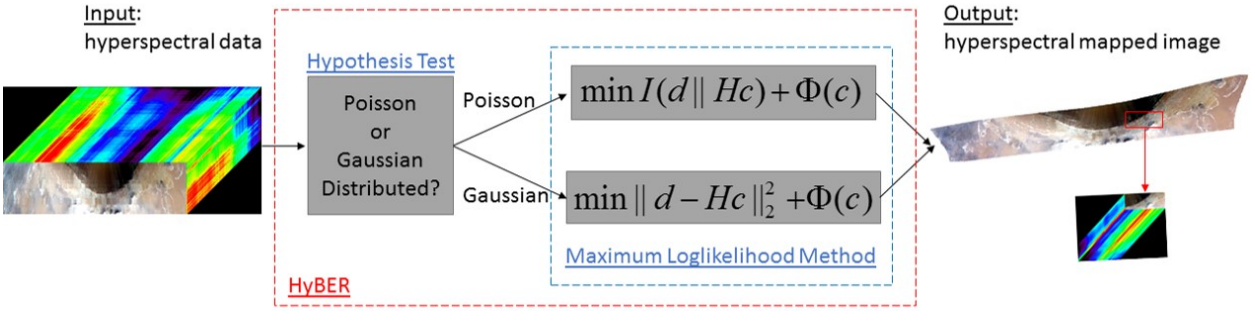


Figure 3.1: Diagram of the hypothesis-based estimation with regularization (HyBER). In HyBER, the hypothesis test is introduced in details in 3.1.1 and the maximum loglikelihood method is explained in 3.1.2.

3.1.1 Hypothesis Tests

Two statistical models for hyperspectral sensor data d_i are considered: Gaussian and scaled Poisson. Gaussian and Poisson distributions are common distributions respectively for continuous and integer random variables [65], due to the law of large number. Hyperspectral radiance data are discrete photon counting essentially as well as continuous electricity signals after collecting and calibration. Data might not be directly Poisson distributed due to some necessary calibration α in the instrument or preprocessing. Then two hypotheses are

$$\begin{aligned} \mathcal{H}_1 &: \ \alpha d_i \sim Poisson(\alpha a_i) \ i = 1, \cdots, M \\ \mathcal{H}_2 &: \ d_i \sim \mathcal{N}(\mu_i, \sigma^2) \ i = 1, \cdots, M, \end{aligned} \tag{3.1}$$

where $Poisson(\alpha a_i)$ is the Poisson distribution with mean αa_i and $\mathcal{N}(\mu_i, \sigma^2)$ is the Gaussian distribution with different means μ_i but the same variance σ^2. α is a positive constant for all $i = 1, \cdots, M$.

It can be shown in Harremoës and Tusnády [74] that for a Poisson distribution, the test G-statistic:

$$\begin{aligned} G(\alpha d_i) &= -2 \times \ln p_{\alpha a_i}(\alpha d_i) = 2 \times [\alpha a_i + \ln(\alpha d_i)! - \alpha d_i \times \ln \alpha a_i] \\ &\xrightarrow{\alpha d_i \to \infty} 2\alpha \times (d_i ln \frac{d_i}{a_i} - d_i + a_i) = 2\alpha I(d_i || a_i) \end{aligned} \tag{3.2}$$

is approximated by a χ^2-distribution with one degree of freedom.

For Gaussian distribution, the corresponding test G-statistic is,

$$G(d_i) = \frac{(d_i - \mu_i)^2}{\sigma^2} \tag{3.3}$$

which also has a χ^2-distribution with one degree of freedom.

We have stated that testing the original null hypotheses above is equivalent to testing the following hypothesis [10]:

$$\mathcal{H}_0 : \ pv_i = 1 - F_{\chi^2(1)}(G(d_i)) \sim U[0,1], \tag{3.4}$$

where $F_{\chi^2(1)}(\cdot)$ is the cumulative χ^2 probability with one degree of freedom. $U[0,1]$ is the uniform distribution on $[0,1]$.

We simplify a set of distribution tests with different parameters given one sample per parameter in (3.1) to the standard uniform distribution test given enough samples in (3.4). This is done by introducing the statistics $G(d_i)$ whose distribution is independent of parameters a_i or μ_i.

Given a float number between 0 and 1, if the P-value (pv_i) is less than this number, the test is rejected, and if the P-value is larger than this number, the test is accepted. Thus when

the number of tests becomes large enough (M is large enough), the float number indicates the probability that the test is rejected if the hypothesis is true,

$$Pr[\mathcal{H}_0 \text{ is rejected}|\mathcal{H}_0 \text{ is true}] \tag{3.5}$$

This probability is often reported as the type I error, or significance. Therefore when $M \to \infty$, given a significance, if the null hypothesis is true, the fraction of rejected d_i should be equal to this significance (Devore 2001). To demonstrate this method, Fig. 3.2 shows the results of a simulation of 100,000 random variables with mean values ranging over three orders of magnitude. When generating Poisson random variables, we first uniform-randomly generate 100,000 values as means and then generate 100,000 Poisson samples from corresponding mean (one sample for one mean). As for Gaussian random variables, means are generated the same while variance is chosen as the expectation of 100,000 means. A hypothesis that highly accurately describes the random variables should yield cumulative distributions that approximate a diagonal line through the plot. The lines for both a Poisson and Gaussian hypotheses yield lines along the y=x diagonal, whereas two tests that do not match the distribution yield curves that are not along the diagonal. One could notice that assuming Poisson when it is Gaussian (cyan line with circles) is worse than assuming Gaussian when it is Poisson (purple line with diamonds). One possible explanation is that when the mean goes large enough, Poisson distribution could be approximated by Gaussian, not vice versa.

For the Poisson hypothesis H_1, $G(d_i) \approx 2\alpha I(d_i||a_i) = 2\alpha(d_i \ln \frac{d_i}{a_i} - d_i + a_i)$ and for the Gaussian hypothesis H_2, $G(d_i) = \frac{(d_i-\mu_i)^2}{\sigma^2}$. We can find although the distribution of $G(d_i)$ is independent of parameters a_i or μ_i, the computation of $G(d_i)$ depends on them. However for real data, it is hard to find true parameters a_i or μ_i. Therefore in practice [10], what we

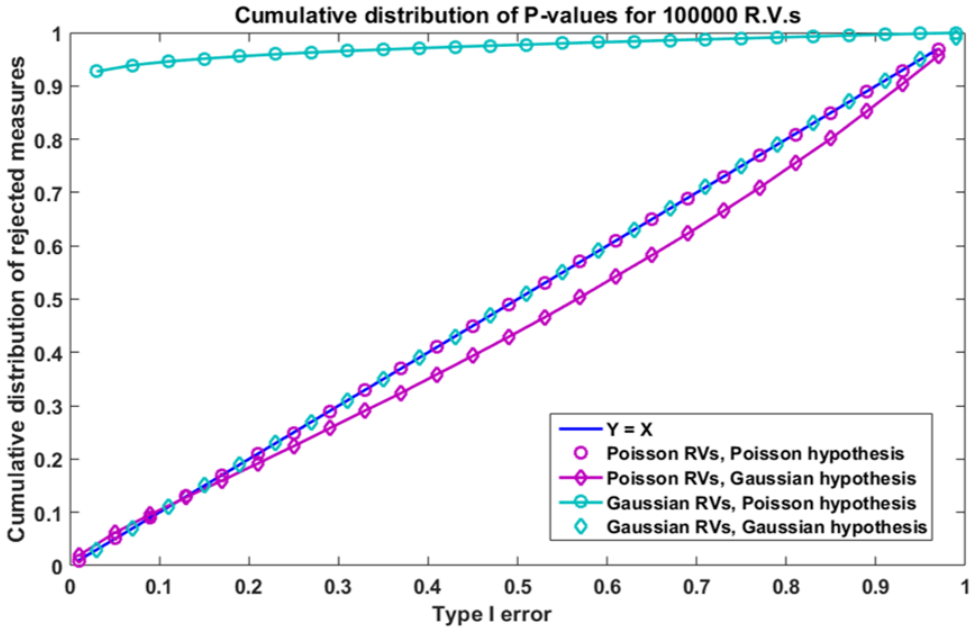


Figure 3.2: Cumulative distribution of P-values for various simulation cases (Simulation).

test is

$$\begin{aligned} \mathcal{H}_1 &: \ \alpha d_i \sim Poisson(\alpha \hat{a}_i) \ i = 1, \cdots, M \\ \mathcal{H}_2 &: \ d_i \sim \mathcal{N}(\hat{\mu}_i, \sigma^2) \ i = 1, \cdots, M, \end{aligned} \tag{3.6}$$

where $\hat{a}_i$ and $\hat{\mu}_i$ are the best estimates by minimizing $\sum_{i=1}^{M} I(d_i||a_i)$ (explained in detail in section 3.1.2) and $\sum_{i=1}^{M}(d_i - \mu_i)^2$ respectively.

To quantitatively analyze the goodness of fit, we introduce Kullback-Leibler (KL) divergence d_{KL} [75] to describe the discrepancy between the hypothesis uniform distribution $p(x)$ and the real distribution $q(x)$ of pv_i in (3.4).

$$\begin{aligned} d_{KL}(q||p) = \int_0^1 q(x) \ln \frac{q(x)}{p(x)} dx &= \int_0^1 q(x) \ln q(x) dx \\ &= -S(q), \end{aligned} \tag{3.7}$$

where $S(q)$ is the differential entropy of probability density $q(x)$.

Because we do not know the true $q(x)$, due to the law of large number, we use a Q-bin histogram $\frac{n_k}{M}$ to estimate $\frac{q(x_k)}{Q}$ where $x_k = \frac{2k-1}{2Q}, k = 1, 2, \cdots, Q$ and $n_k = \sum_{i=1}^{M} I_{[x_k - \frac{1}{2Q}, x_k + \frac{1}{2Q}]}(pv_i)$ with $I_{[x_k - \frac{1}{2Q}, x_k + \frac{1}{2Q}]}(pv_i) = 1$ if $pv_i \in [x_k - \frac{1}{2Q}, x_k + \frac{1}{2Q}]$, otherwise, $I_{[x_k - \frac{1}{2Q}, x_k + \frac{1}{2Q}]}(pv_i) = 0$. We also estimate d_{KL} (3.7) by

$$\hat{d}_{KL} = \sum_{k=1}^{Q} \frac{n_k}{M} \ln \frac{n_k}{M} + \ln(Q). \tag{3.8}$$

According to Moddemeijer [76], the bias error for $\hat{d}_{KL}$ is

$$\mathbf{E}[\hat{d}_{KL}] - d_{KL} \approx C_1 \frac{Q}{M} - C_2 \frac{1}{Q^2}, \tag{3.9}$$

where C_1, C_2 are constants in the order of 1. Since we have large M (around 10^7), $Q = 100$ makes this bias error close to 0. The variance of this estimate is $\sim 1/2M$ [76], which is $\sim 10^{-7}$ for our case. Therefore $\hat{d}_{KL}$ is used as the quantification that indicates the accuracy of the null hypothesis in (3.1).

Table 3.1: Hypothesis KL divergences for simulations

Data	Hypothesis	$\hat{d}_{KL}$
$\alpha d_i \sim Poisson(\alpha a_i)$	$\alpha d_i \sim Poisson(\alpha a_i)$	2.0×10^{-6}
$\alpha d_i \sim Poisson(\alpha a_i)$	$\alpha d_i \sim Poisson(\alpha \hat{a}_i)$	0.0089
$\alpha d_i \sim Poisson(\alpha a_i)$	$d_i \sim \mathcal{N}(\hat{\mu}_i, \hat{\sigma}^2)$	0.63
$d_i \sim \mathcal{N}(\mu_i, \sigma^2)$	$d_i \sim \mathcal{N}(\mu_i, \sigma^2)$	1.6×10^{-6}
$d_i \sim \mathcal{N}(\mu_i, \sigma^2)$	$d_i \sim \mathcal{N}(\hat{\mu}_i, \sigma^2)$	0.0011
$d_i \sim \mathcal{N}(\mu_i, \sigma_i^2)$	$\hat{\alpha} d_i \sim Poisson(\hat{\alpha} \hat{a}_i)$	3.4796

We generate two simulated data sets:

1. Data $d_i \sim Poisson(\alpha a_i)/\alpha$ for given a_i and α; a_i are randomly generated between 0 and 1, and $\alpha = 10^6$ to approximate real data

2. Data $d_i \sim \mathcal{N}(\mu_i, \sigma^2)$ for given μ_i and σ; μ_i are randomly generated between 0 and 1, and $\sigma^2 = 10^{-2}$.

Applying our hypothesis method to these two simulations, table 3.1 shows $\hat{d}_{KL}$ for different hypotheses and different datasets ($M > 10^7$). When the hypothesis distribution matches the real distribution and true values are known, the KL divergence is as small as 10^{-6}. It also verifies that the approximated distribution of $G(d_i) \approx 2\alpha I(d_i||a_i)$ [74] does not affect the final hypothesis results for large α. However, even based on estimated values, if the hypothesis distribution is true, the KL divergence is also very small (around 10^{-3}). Otherwise, if the hypothesis distribution is wrong, the KL divergence is relatively much larger. One can notice that the KL divergence for Gaussian distribution approximating Poisson data is clearly smaller than the KL for Poisson distribution approximating Gaussian data, which follows the Central Limit Law that implies the Gaussian distribution can approximate a Poisson distribution when the mean parameters are large enough. Our method succeed to select a better distribution from candidates based on data only.

3.1.2 Statistical Modelling

Define $\mathbb{U} = \mathbb{R}^+$ as the positive real number space. If we have M measurements, data $\mathbf{d} = [d_1, d_2, ..., d_M]^T$ and their estimates $\mathbf{a} = [a_1, a_2, ..., a_M]^T$ should belong to $\mathbb{U}^M$, that is $\mathbf{d}, \mathbf{a} \in \mathbb{U}^M$. The reconstruction mapped image in local Cartesian axis $\mathbf{c} = [c_1, c_2, ..., c_N]^T$ contain N pixels in total, then $\mathbf{c} \in \mathbb{U}^N$. The physical model provides the system matrix (forward model) $H \in \mathbb{R}^{M \times N}$, which describes how the mean data $\mathbf{a}$ is linearly generated from the mapped reconstruction $\mathbf{c}$, $\mathbf{a} = H\mathbf{c}$. Often, H is called a transfer function. We use a Gaussian kernel as the transfer function for this hyperspectral image system and the selection of the transfer function kernel is independent of the distribution of measured data.

We aim to maximize the loglikelihood based on the statistical model selected from section as in 3.1.1. The ill-posed nature of the forward model and the noise in data $\mathbf{d}$ create instabilities in estimates for the mapped reconstruction image $\mathbf{c}$. Therefore a convex regularization function $\Phi_{reg}(\mathbf{c})$, which is described in the section 3.1.3, is introduced. The reconstruction problem can be stated as an optimization problem

$$\begin{aligned} \min_{\mathbf{c}} \quad & -\ln p(\mathbf{d}|\mathbf{a}) + \Phi_{reg}(\mathbf{c}) \\ & s.t.\ \mathbf{a} = H\mathbf{c}. \end{aligned} \tag{3.10}$$

The negative loglikelihood $-\ln p(\mathbf{d}|\mathbf{a})$ is also called a data fitting term.

If data is selected as being Gaussian distributed, (3.10) is equivalent to

$$\begin{aligned} \min_{\mathbf{c}} \quad & \sum_{i=1}^{M} (d_i - a_i)^2 + \Phi(\mathbf{c}) \\ & s.t.\ \mathbf{a} = H\mathbf{c}. \end{aligned} \tag{3.11}$$

If data is modeled as being Poisson distributed, (3.10) is equivalent to

$$\begin{aligned} \min_{\mathbf{c}} \quad & \sum_{i=1}^{M}(-d_i \ln a_i + a_i) + \Phi(\mathbf{c}) \\ & s.t.\ \mathbf{a} = H\mathbf{c}. \end{aligned} \tag{3.12}$$

In (3.11) and (3.12), $\Phi(\mathbf{c})$ is the equivalent regularization function from $\Phi_{reg}(\mathbf{c})$ in (3.10) after simplifying the data fitting term.

We derive two iterative algorithms for two different distributions as shown in Algorithms 1 and 2 by decoupling the penalty as $\Phi(\mathbf{c}) \leq \Phi^D(\mathbf{c}, \mathbf{c}^{(t)}),\ \forall \mathbf{c}$ where $\Phi^D(\mathbf{c}, \mathbf{c}^{(t)})$ is decoupled using the surrogate method [10] by introducing known $\mathbf{c}^{(t)}$ at t iterations (see section 3.1.3). Minimizing the decoupled penalty minimizes the original penalty iteratively [59].

If HyBER verifies data are scaled Poisson distributed, Algorithm 1 is applied. If HyBER verifies data are Gaussian distributed, Algorithm 2 is utilized. It should be noticed that

Algorithm 2 is derived by taking advantage that H is a Gaussian kernel transfer function and details can be found in [77].

Algorithm 1: Regularized Maximum Log-likelihood Method for Poisson data

Data: measured data $\mathbf{d}$, system matrix H, maximum iteration *maxiter*, penalty parameters $\beta_1, \beta_2, \delta_1, \delta_2$

Result: reconstructed mapped image $\mathbf{c}$, estimated data $\mathbf{a}$

Initialize $c^{(1)} \in \mathbb{U}^N$ and $\mathbf{h} = H^T\mathbf{1}$ where $\mathbf{1} = [1, 1, ..., 1]^T \in \mathbb{U}^M$, $\mathbf{h} \in \mathbb{R}^N$;

for $i \leftarrow 1$ **to** *maxiter* **do**

Forward Projection: $\mathbf{a}^{(t)} \leftarrow H\mathbf{c}^{(t)}$ Data Error Ratio: $b_i^{(t)} \leftarrow d_i/a_i^{(t)}$ Backward Projection: $\mathbf{f}^{(t)} \leftarrow H^T\mathbf{b}^{(t)}$ Update using the Trust-Region Newton Method: $\mathbf{c}^{(t+1)} \leftarrow \arg\min_{\mathbf{c}} \sum_{j=1}^{N} -c_j^{(t)} f_j^{(t)} \ln c_j + h_j c_j + \Phi^D(\mathbf{c}, \mathbf{c}^{(t)}; \beta_1, \beta_2, \delta_1, \delta_2)$

end

Algorithm 2: Regularized Maximum Log-likelihood Method for Gaussian data

Data: measured data $\mathbf{d}$, system matrix H, maximum iteration *maxiter*, penalty parameters $\beta_1, \beta_2, \delta_1, \delta_2$

Result: reconstructed mapped image $\mathbf{c}$, estimated data $\mathbf{a}$

Initialize $c^{(1)} \in \mathbb{U}^N$ and $\mathbf{h} = H^T\mathbf{1}$ where $\mathbf{1} = [1, 1, ..., 1]^T \in \mathbb{U}^M$, $\mathbf{h} \in \mathbb{R}^N$;

for $i \leftarrow 1$ **to** *maxiter* **do**

Forward Projection: $\mathbf{a}^{(t)} \leftarrow H\mathbf{c}^{(t)}$ Data Error Difference: $b_i^{(t)} \leftarrow d_i - a_i^{(t)}$ Backward Projection: $\mathbf{f}^{(t)} \leftarrow H^T\mathbf{b}^{(t)}$ Update using the Trust-Region Newton Method: $\mathbf{c}^{(t+1)} \leftarrow \arg\min_{\mathbf{c}} \sum_{j=1}^{N} -2f_j^{(t)}(c_j - c_j^{(t)}) + h_j(c_j - c_j^{(t)})^2 + \Phi^D(\mathbf{c}, \mathbf{c}^{(t)}; \beta_1, \beta_2, \delta_1, \delta_2)$

end

3.1.3 Weighting Regularization

The regularization $\Phi(\mathbf{c}) = \Phi_1(\mathbf{c}) + \Phi_2(\mathbf{c})$ is a sum of spatial and spectral terms shown in (3.13). Our prior standard approach was to incorporate the penalty function with constant weights for all pixels and bands [59]. However, this approach produced artifacts in the resulting images and spectra. Fig. 3.8a shows the example how a constant weighting penalty introduces the line-dependent artifacts. Therefore model-based weights $w_{1,j}, w_{2,j}$ needed to be determined from the spatial sampling mode and wavelength-dependent spectral noise levels.

Line scanning hyperspectral sensors spatially sample data non-uniformly. $w_{1,j}$ is the spatially dependent weight to adjust the non-uniform sampling issue. Therefore we make it proportional to the sensitivity mapping $\mathbf{h}$ in algorithms which indicate the sampling degree of the instrument. The necessity of this weight is analyzed in the next section 3.2.3.

$$\begin{aligned}
\Phi_1(\mathbf{c}) &= \sum_{j=1}^{N} \Phi_{1,j}(\mathbf{c}) \\
&= \sum_{j=1}^{N} w_{1,j} \sum_{k \in NA(j)} \frac{1}{r_{kj}} \beta_1 \delta_1^2 \ln \cosh \left[\frac{c_j - c_k}{\delta_1} \right] \\
\Phi_2(\mathbf{c}) &= \sum_{j=1}^{N} \Phi_{2,j}(\mathbf{c}) \\
&= \sum_{j=1}^{N} w_{2,j} \sum_{k \in NE(j)} \frac{1}{r_{kj}} \beta_2 \delta_2^2 \ln \cosh \left[\frac{c_j - c_k}{\delta_2} \right]
\end{aligned} \tag{3.13}$$

where $NA(j)$ and $NE(j)$ are the spatial and spectral neighborhoods of j and r_{kj} is the distance between k and j.

Hyperspectral data with hundred of bands suffer varying noise intensities for different bands, especially for deteriorating detectors. To compensate for increased noise, we estimate wavelength dependent noise levels from the data set and suppress noisy data as part of the spectral regularization ($w_{2,j}$). First, define the squared magnitude of the short-wavelength Fourier transform as the spectrogram of a single spectrum. Let $s(\lambda)$ be the high frequency value of the mean spectrogram (of all spectra in the hyperspectral data). An example is shown in 3.9. In our approach, the spectral weights are set to increase as the uncertainty or noise in the measurement increases; the high frequency information $s(\lambda)$ quantifies this uncertainty. There must be a threshold on the weights corresponding to a maximum penalty. A Gaussian distribution function was found to work well with parameters chosen to give a rapid change in weights for low values of $s(\lambda)$. Given a range $[1, R]$, spectrally, $w_{2,j}$ is proportional to $F_{0,m(s(\lambda))}(s(\lambda))(R-1)+1$ where F_{μ_0,σ_0^2} is the Gaussian cumulative distribution function with mean μ_0 and variance σ_0^2 and $m(s(\lambda))$ is the median value of $s(\lambda)$. Then we make higher penalties associated with worse detectors. Also $w_{2,j}$ consider the non-uniform sampling issue for different areas too.

$$
\begin{aligned}
&\Phi_1(\mathbf{c}) \\
&\leq \sum_{j=1}^{N} \sum_{k \in NA(j)} \beta_1 \delta_1^2 \frac{w_{1,k} + w_{1,j}}{2d_{kj}} \ln \cosh \left[\frac{2c_j - c_k^{(t)} - c_j^{(t)}}{\delta_1} \right] \\
&= \Phi_1^D(\mathbf{c}, \mathbf{c^{(t)}}) \\
&\Phi_2(\mathbf{c}) \\
&\leq \sum_{j=1}^{N} \sum_{k \in NE(j)} \beta_2 \delta_2^2 \frac{w_{2,k} + w_{2,j}}{2d_{kj}} \ln \cosh \left[\frac{2c_j - c_k^{(t)} - c_j^{(t)}}{\delta_2} \right] \\
&= \Phi_2^D(\mathbf{c}, \mathbf{c^{(t)}})
\end{aligned}
\tag{3.14}
$$

Penalties in (3.13) contain unknown c_j, c_k coupled variables. To solve the optimization in (3.10) iteratively, we decouple these variables using a surrogate method as in (3.14) because the original penalties are convex. Replacing $\Phi^D(\mathbf{c}, \mathbf{c}^{(\mathbf{t})})$ in Algorithm 1 by $\Phi_1^D(\mathbf{c}, \mathbf{c}^{(\mathbf{t})}) + \Phi_2^D(\mathbf{c}, \mathbf{c}^{(\mathbf{t})})$, we now can solve the optimization iteratively.

3.2 Application and Performance Analysis

3.2.1 Data Introduction

The Compact Reconnaissance Imaging Spectrometer for Mars (CRISM) on the Mars Reconnaissance Orbiter (MRO) began operations in 2006 [5] as a gimballed push-broom (i.e., line sampling) hyperspectral imaging spectrometer operating from 362 to 3920 nm, with a 6.5 nm spectral band spacing, and a smallest ground pixel size of 18 m. Since 2010, images have occasionally been acquired using an along-track oversampled mode (ATO), with significant overlap in the along-track direction. ATOs allow the reconstruction of images with smaller pixel size (9-12 m/pixel). CRISM operates as S (0.362 to 1.030 μm) and L (1.036 to 2.650 μm) imaging spectrometers. We focus on bands from 0.362 to 2.650 μm here, because for wavelengths greater than $\sim$2.650 μm, the CRISM data contain both reflectance and undetermined thermal emission component. In this chapter, three ATO data sets (FRT0001EB5F, FRT0001DC22 and ATO00037D74) are employed[1]. The reader is referred to papers [78] [79] for scientific analyses using processed data.

CRISM data have been processed and placed in the Planetary Data System archives as two forms of sensor-space hyperspectral image cubes. For S data the raw data sets have

[1]The name "ATO" was used later than the ATO mode was applied. Before the name change, some scenes labeled "FRT" are actually acquired using the ATO mode.

been assembled to image cubes and calibrated to spectral radiance using a combination of pre-flight radiometric measurements and in-flight observations using calibrated light sources contained within the instrument. For L data the archives include image cubes calibrated to spectral radiance and also image cubes that have had known artifacts removed and converted to I/F, which is for each pixel and band the ratio of the spectral radiance from the scene divided by the spectral radiance of the sun at the heliocentric distance for the observation. We chose to use the latter form of the L data cubes as a start in our processing and will refer to this product as "CRISM baseline". For the map-projected version of the "CRISM baseline" we use a mapping algorithm that includes a 1/distance weighting of the mapped value of an output pixel, with no consideration of regularization or statistics.

Two processing steps for the hyperspectral ATO image cubes are necessary [10]:

1. *Retrieval of Single Scattering Albedo*: The primary derived CRISM hyperspectral image cube archived at NASA's Planetary Data System (PDS) has units of spectral radiance ($Wm^{-2}sr^{-1}\mu m^{-1}$). In our processing, spectral radiances are first converted to I/F by dividing by the solar spectral radiance at Mars at the time of observation. Discrete Ordinates Radiative Transfer (DISORT)-based processing incorporating modeling atmospheric gases and aerosols [2], and the Hapke function [11] for surface scattering, are then used to retrieve the surface single scattering albedo (SSA) [80]. SSA is the ratio of scattering efficiency to scattering plus absorption efficiencies for the Mars surface.

2. *Extrema Removal by Median Filter*: A median filter originally designed by Eliason and McEwen [81] is used next to remove extreme values, a consequence of aging detectors, and to replace their values by the median value of the neighborhood. This technique is similar to what was employed by Carter et al. for CRISM data [82]. The necessity of this step is verified by Kreisch, et al. [10]. These two processing steps remove the effects of atmospheric gases and aerosols, lighting and viewing conditions, and extreme noise spikes.

3.2.2 Hypothesis Results

Let's take a deeper look of $\mathcal{H}_1$. α is assumed to be constant, independent of spectral and spatial location. To test this assumption, we estimated this constant factor first. To test the shape of the distribution, we use these two homogeneous areas. For a homogeneous region, the statistics should be independent of spatial location. The hypothesized scale factor could depend on spectral value in this case $\alpha = \alpha(\lambda)$. A simple moment estimator yields a reasonable estimate for α. For a scaled Poisson random variable, the ratio of the variance to the mean equals the scale factor. We test on two homogeneous areas of CRISM ATO0002EC79 IR SSA data (shown in Fig. 3.3) and the estimate scalars depending on wavelength are plotted in the Fig. 3.4.

Fig. 3.4 shows that although the estimate scalars of SSA after median filter contains some noise, it is approximately independent of wavelength. By comparison, we note that the SSA after the median filter has a more constant estimated scale factor, with most spikes removed. It also verifies that the median filter in the preprocessing step is necessary if the scaled Poisson distribution is selected in the reconstruction step.

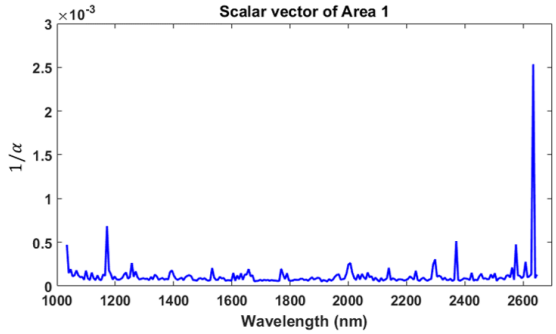
The average values from Fig. 3.4 are used to scale the data prior to testing the distribution. Fig. 3.5 shows a direct comparison of the Poisson and the cumulative distribution function for a Poisson compared to measured values (histogram and cumulative histogram), for each of the two areas in Fig. 3.3. While there is some difference between the shape of the Poisson distribution and the histogram near the peak, the cumulative distributions match closely.

After testing for the homogeneous areas, we apply the quantitative hypothesis method in 2.2 to some scenes: ATO00037D74, FRT0001EB5F, ATO0002DDF9 and ATO0002EC79.

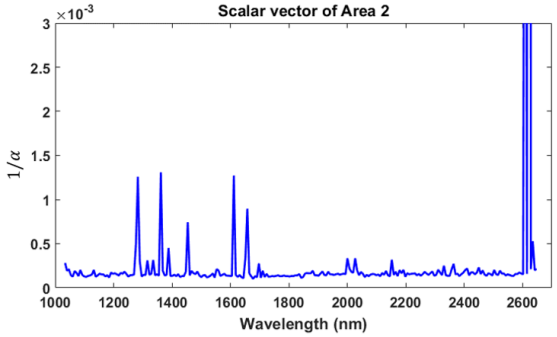
(a) Homogeneous Area 1.

(b) Homogeneous Area 2.

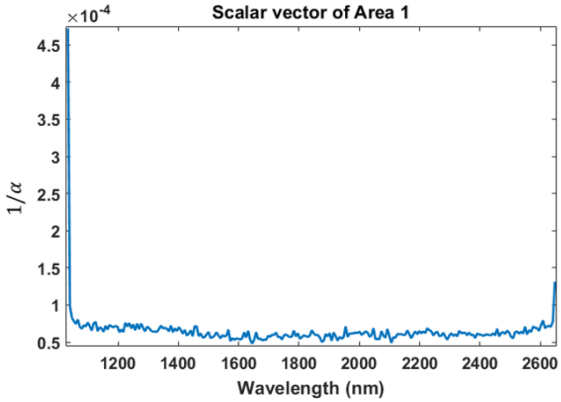
Figure 3.3: Two selected homogeneous areas of ATO0002EC79 IR SSA data (Area is marked in red).



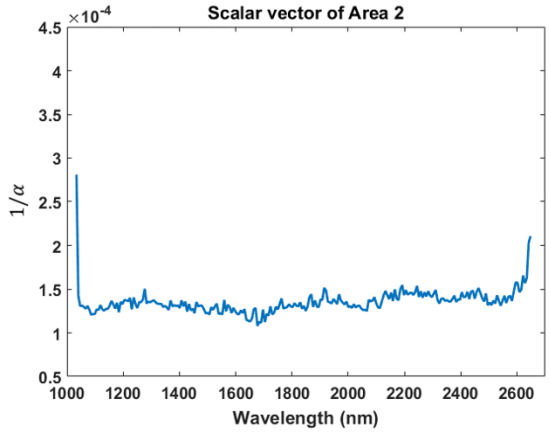
(a) Scalar vector of homogeneous area 1 before median filter.



(b) Scalar vector of homogeneous area 2 before median filter.

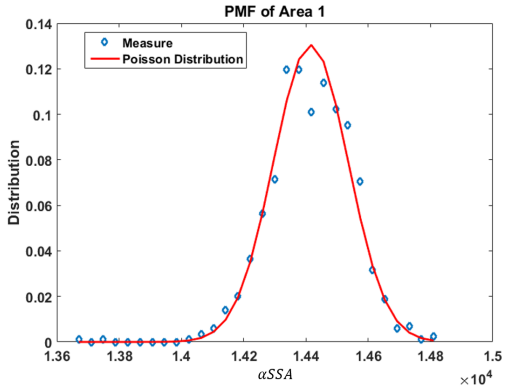


(c) Scalar vector of homogeneous area 1 after median filter.

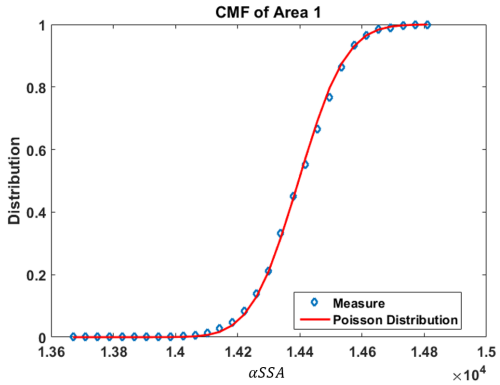


(d) Scalar vector of homogeneous area 2 after median filter.

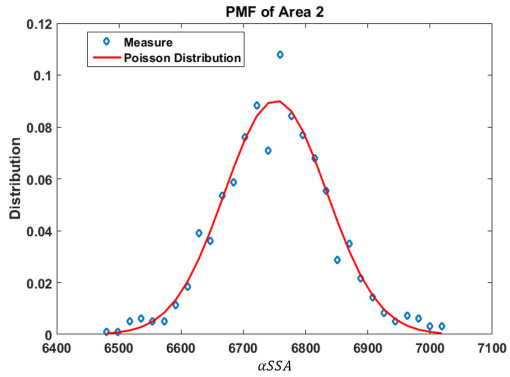
Figure 3.4: Scalar vector of homogeneous areas before and after the median filter.



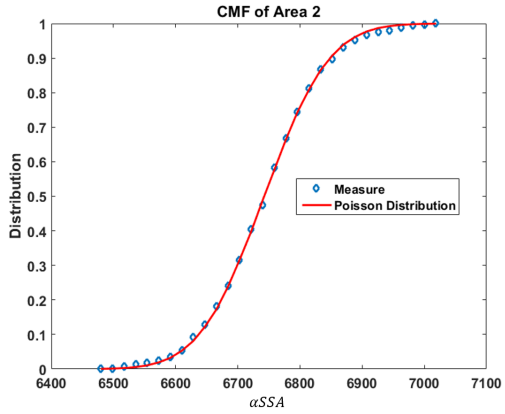
(a) Probability mass function of the homogeneous area 1.



(b) Cumulative mass function of the homogeneous area 1.



(c) Probability mass function of the homogeneous area 2.



(d) Cumulative mass function of the homogeneous area 2.

Figure 3.5: Theoretical scaled Poisson distribution approximates the real distribution of measured values.

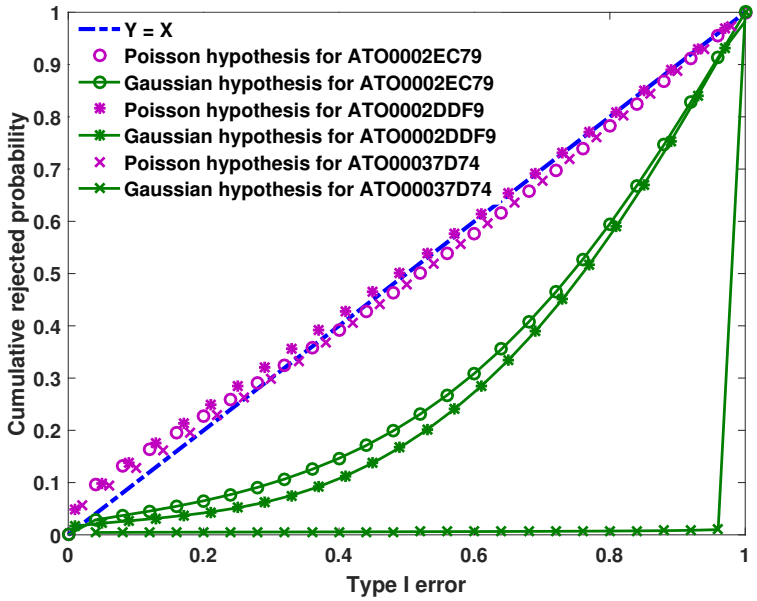


Figure 3.6: Scalar vector of homogeneous area before and after median filter.

P-value curves are drawn in Fig. 3.6, and curves closer to the diagonal line indicate better approximations. Fig. 3.6 indicates that for both scenes the scaled Poisson distribution is closer than the additive white Gaussian distribution to the data distribution.

Compute the d_{KL} from the hypothesis method in 3.1.1 to CRISM data (ATO00037D74 and FRT0001EB5F, collected four years apart), KL divergences in table 3.2 for two scenes show that the scaled Poisson distribution is closer to the real distribution than the additive white Gaussian hypothesis. We also note that the KL divergence for the scaled Poisson distribution based on real data (10^{-2}) is much higher than the scaled Poisson distributed simulation (10^{-3}) in table 3.1. This might be due to higher variation in α_i.

Table 3.2: Hypothesis KL divergences for CRISM data

Data	Hypothesis	$\hat{d}_{KL}$
ATO00037D74	$\hat{\alpha}d_i \sim Poisson(\hat{\alpha}\hat{a}_i)$	0.034
ATO00037D74	$d_i \sim \mathcal{N}(\hat{\mu}_i, \hat{\sigma}^2)$	0.54
FRT0001EB5F	$\hat{\alpha}d_i \sim Poisson(\hat{\alpha}\hat{a}_i)$	0.081
FRT0001EB5F	$d_i \sim \mathcal{N}(\hat{\mu}_i, \hat{\sigma}^2)$	4.53

3.2.3 Weighting Penalty Performance

Spatially dependent weights

Regularized reconstruction methods typically use penalty functions with constant weights $w_{1,j} = 1$ for all spatial neighborhood pixels in (3.13). This approach introduced row artifacts in CRISM map projected reconstructions because in undersampled areas the penalty term was weighted higher than the data fitting term (3.10), resulting in over-smoothing. For example, when reconstructing the hyperspectral sensor data FRT0001DC22, the sensitivity mapping of it is shown in Fig. 3.7. Without prior knowledge of this sampling informtion, the constant spatial penalty leads to the reconstructed map projected reconstruction shown in Fig. 3.8a. The sensitivity $\mathbf{h}$ (independent with the statistical model) in Algorithm 1 accounts for varying sampling of different spatial regions and its values range over 20 orders of magnitude (shown in Fig. 3.7), which shows that some areas are oversampled while some are highly undersampled. Making the weights in the spatial penalty $\Phi_1(\mathbf{c})$ proportional to sensitivity values equalizes the balance between the data fitting term and penalty terms in (3.10), and substantially eliminates the row artifacts associated with high degree of undersampling as shown in Fig. 3.8b.

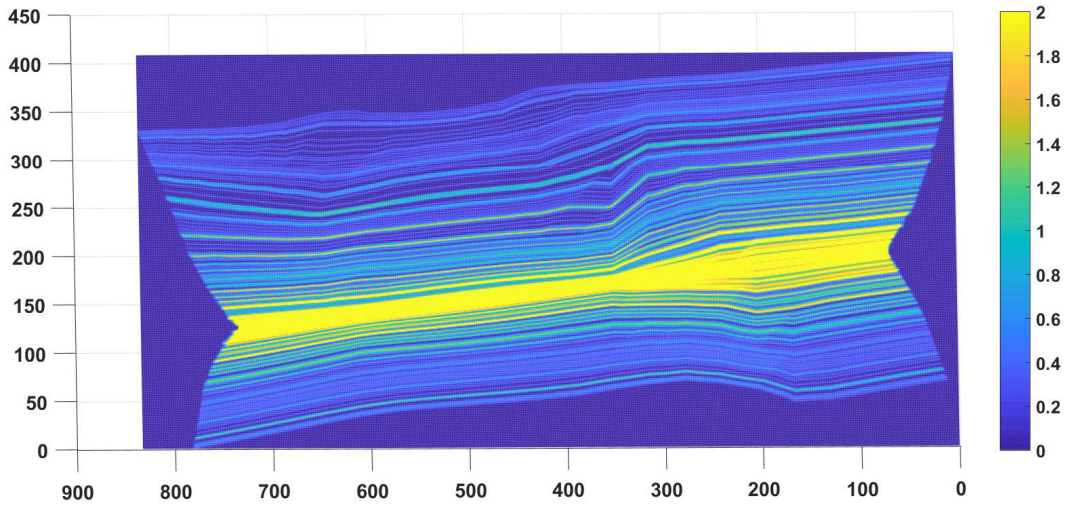
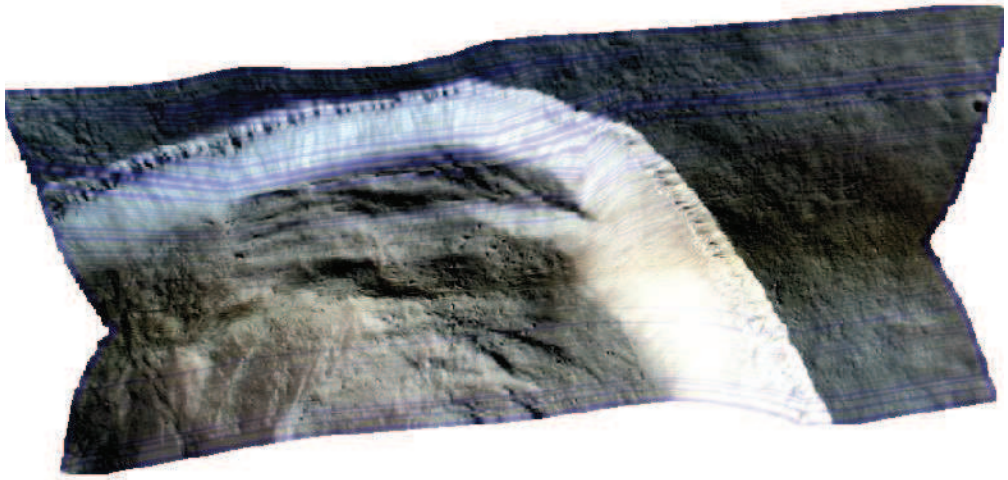


Figure 3.7: An example of the spatial sensitivity mapping for the hyperspectral reconstruction CRISM FRT0001DC22 on Mars. Values in this sensitivity mapping vary from 10^{-20} to 20. Data are mapped to the Mars equirectangular projection.

Spectrally dependent weights

Fig. 3.9 shows spectrograms of two examples: hyperspectral data FRT0001EB5F collected by CRISM (see 3.2.5 for details) during June 2011 and ATO00037D74 acquired in August 2015. Spectrogram values of FRT0001EB5F are smaller than those of ATO00037D74 at high frequencies, which matches the fact that CRISM in 2011 had higher signal to noise ratios (SNRs) than in 2015. The increased noise for the most recent scenes is associated with aging detectors and thermoelectric coolers. The high frequency information in Fig. 3.9b shows relatively lower values between 1.7 to 1.9 and 2.1 to 2.5 μm, which is consistent with the predicted instrument performance [5].

(a) Constant Weighting Penalty.



(b) Spatially Dependent Weighting Penalty.

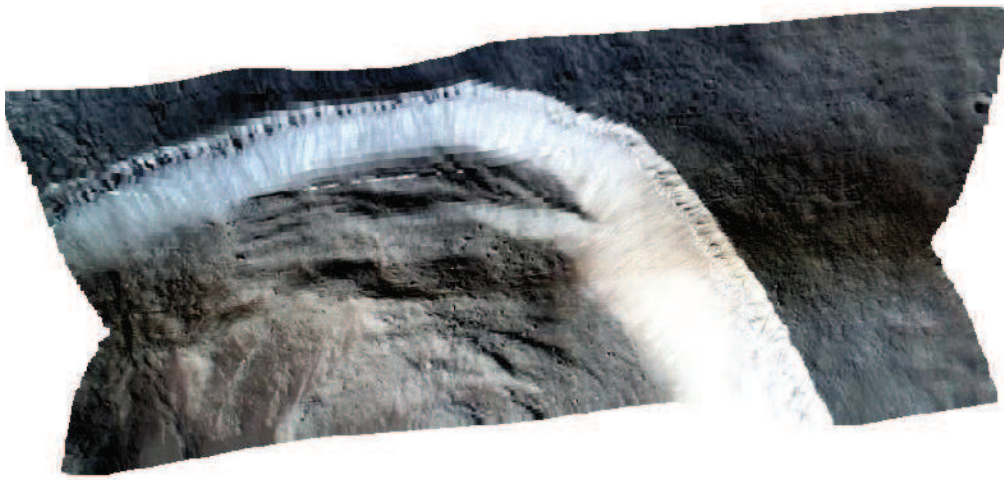
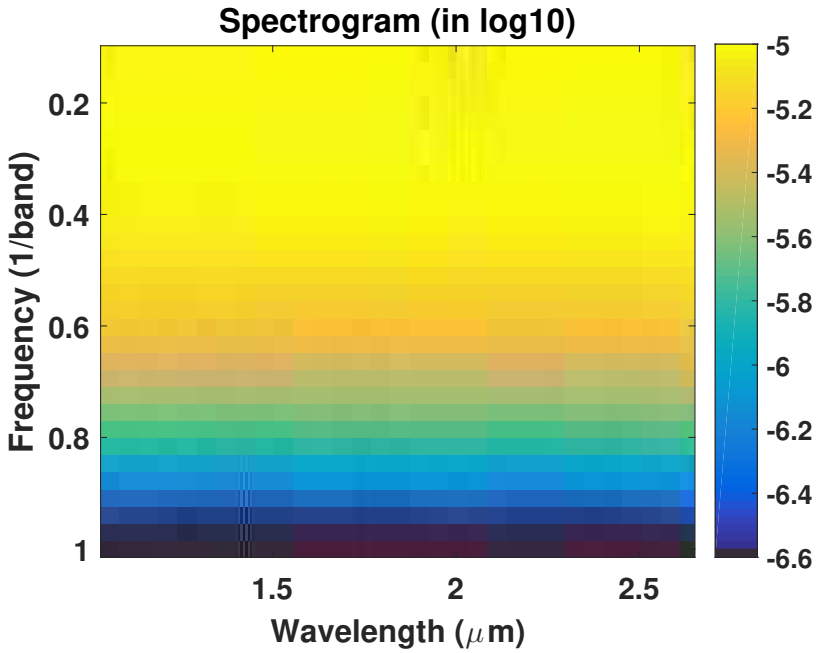


Figure 3.8: Results with different penalties for CRISM FRT0001DC22 on Mars in Fig 3.7. Reconstructions are shown in RGB with wavelengths 0.71 μm, 0.599 μm and 0.534 μm. Frames are $\sim$10 km across. Note the row artifacts in Fig. 3.8a is corresponding to the sensitivity mapping in Fig 3.7.

(a) Spectrogram of the CRISM FRT0001EB5F in 2011.



(b) Spectrogram of the CRISM ATO00037D74 in 2015.

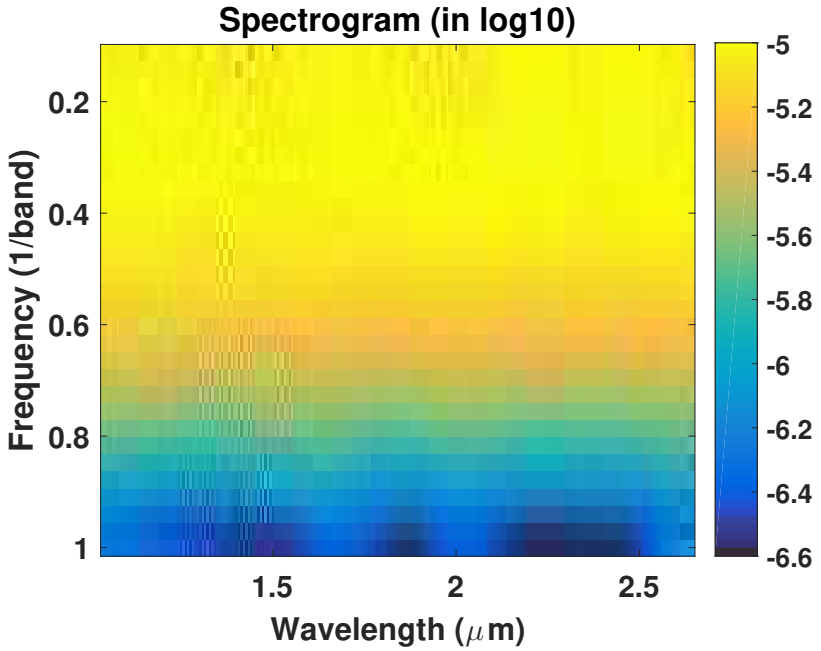


Figure 3.9: Mean spectrograms of L data for older and recent scenes are shown with color-coded relative logarithmic noise power. The vertical axis is the normalized Fourier frequency. For the scene ATO00037D74, relatively lower noise levels at higher frequency are found between 1.7 to 1.9 and 2.1 to 2.5 μm, which is consistent with the recent performance of the instrument.

3.2.4 Simulations

CRISM often observes areas on Mars where no landed spacecraft has been to provide "ground truth". To quantitatively evaluate the performance of the algorithm, we therefore use simulations. Our simulations use a hyperspectral reconstruction as known truth. The hyperspectral mapped image chosen is the output of application of our algorithm to real data from the scene ATO00037D74. This choice is arbitrary and any random scene with reasonable continuity and smoothness properties could have been chosen. Using a scene from Mars allows visualization of the results.

Random hyperspectral data were generated by applying the system matrix to this hyperspectral mapped image, scaling the result by α, then generating independent Poisson random variables with these means. These Poisson variables were then divided by the scalar α. In the true scenes, the scalar α varies over a range from 10^3 to 10^4. Do account for this, a second set of simulations was run with varying α_i. The varying scalar is chosen randomly for every spectral band, uniformly in the range from 10^3 to 10^4. The results below are for 32 independent realizations of the data for each of the two cases (constant and varying α) [2].

Table 3.3 compares relative errors for our method HyBER, the Gaussian method (incorrect assumption) and the projected baseline (non-statistical method). Using HyBER, the Poisson model can be selected over the Gaussian model (from table 3.1). Then Algorithm 1 can be applied with penalty parameters chosen as $\beta_1 = 0.01, \delta_1 = 4, \beta_2 = 0.1, \delta_2 = 0.9$. Using an incorrect statistical model (Gaussian method in table 3.3), Algorithm 2 can be applied with the same penalty parameters. The projected baseline method (also in table 3.3) is the projection without noise suppression and using one over distance weighting to calculate projected pixel values from surrounding pixels.

[2]Simulation codes can be found under my github (No.1): `https://github.com/linyunhe/hypothesis_test_PG_for_CRISM`

Table 3.3: Relative errors for scaled Poisson distributed simulations

Methods	Constant scalar	Varying scalar
HyBER (mean)	1.704×10^{-5}	1.982×10^{-5}
HyBER (standard derivation)	0.0056	0.0085
Gaussian method (mean)	-0.0025	-0.269
Gaussian method (standard derivation)	0.0107	0.733
Projected baseline (mean)	-0.138	-0.138
Projected baseline (standard derivation)	0.0054	0.0081

For simulations with both constant and varying scalars, HyBER performs the best in both means and standard derivations of errors. HyBER is also robust to the choice of scalars (constant or varying), while the Gaussian method works much better with a constant scalar (10^4) than a varying scalar ($10^3 - 10^4$). One possible explanation is that a constant and larger scalar generates simulated data that are approximately Gaussian distributed according to the central limit theorem. Moreover, for varying scalar simulations, the Gaussian method generates more biased results than the projected baseline does, but for constant scalar simulations, the Gaussian method leads to much more unbiased reconstructions than the projected baseline does. Therefore, with a wrong statistical assumption (Gaussian distribution), the statistical method can even perform much worse than the projected baseline.

Then we want to analyze errors in details to figure out why the wrong statistical assumptions will lead to bad estimates. Fig. 3.10 shows the boxplot of all mean relative errors for constant-scalar simulations. We can find that compared with HyBER, Gaussian method has more outliers (lower than -100), which are certainly wrong. If zooming in the quantiles range, we can find that both HyBER and Gaussian has median which is very close to zeros(-1.91×10^{-5}

for Gaussian and -1.90×10^{-5} for HyBER). However, Gaussian errors still have a little bit larger range between quantiles than HyBER, but not as large as the difference in Table 3.3. Thus, the large mean and variance of Gaussian method errors seem to be caused by the outliers. Fig. 3.11a shows the spatial locations of these outliers. We can find that

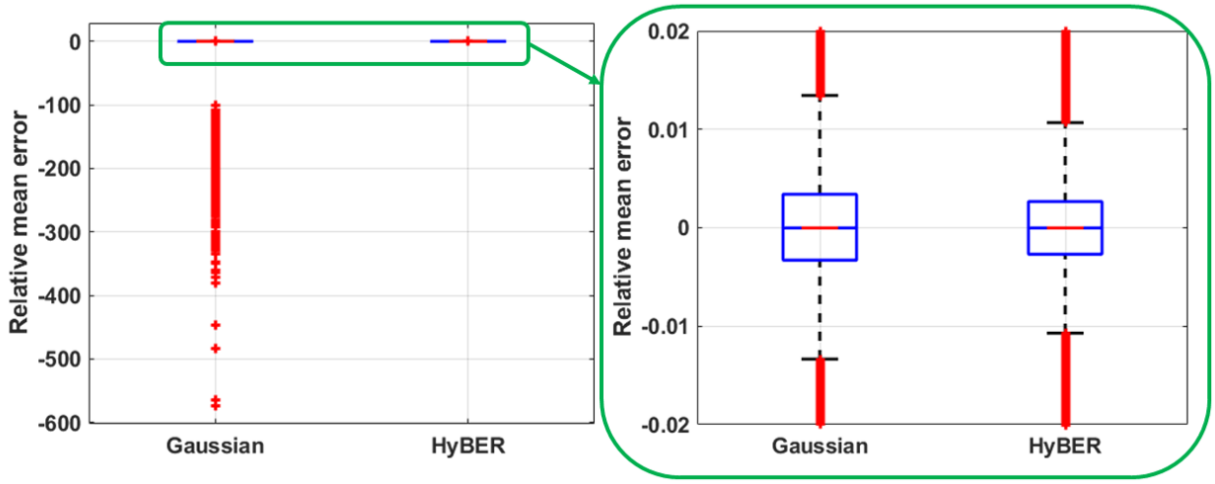
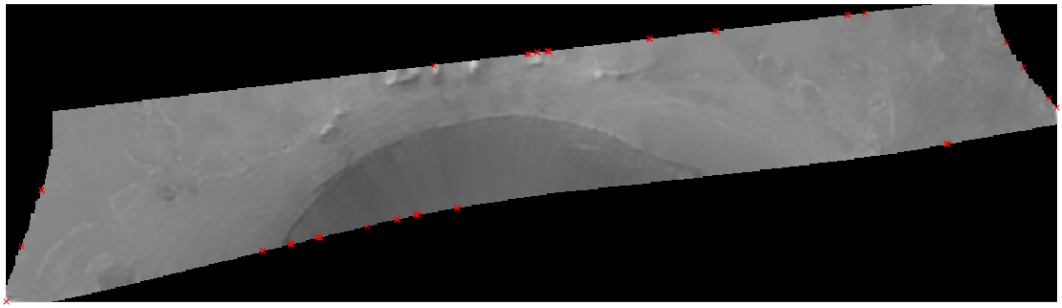


Figure 3.10: Boxplot of all mean relative errors for Gaussian methods and HyBER for constant scalar simulations.

it seems like the outliers all appears in the boundary of the image where we have limited sampled information. Therefore it seems like if we cut the boundary of Gaussian method, Gaussian algorithms will be more robust and accurate. Moreover Fig. 3.11b shows the ratio of errors (after removing outliers) from Gaussian and HyBER in $\log 10$. Thus if the pixel value is larger than 0, Gaussian method has higher errors than HyBER. Overall, it seems like Gaussian method in general a little bit worse than HyBER (which verifies the conclusion from Fig. 3.10). One interesting finding is that in the boundary of crater (half ring shape in the middle of the figure), Gaussian seems to have around 2 times errors than HyBER. Because the penalty parameters have been adjusted based on overall cost function, this boundary bias seems to be related with the wrong statistical assumptions.

(a) Gaussian method estimations (in average). The red crosses show pixels who have extremely weird estimates in the spectral domain (for example, negative values).



(b) The ratio of Gaussian relative errors and HyBER relative errors in $\log 10$.

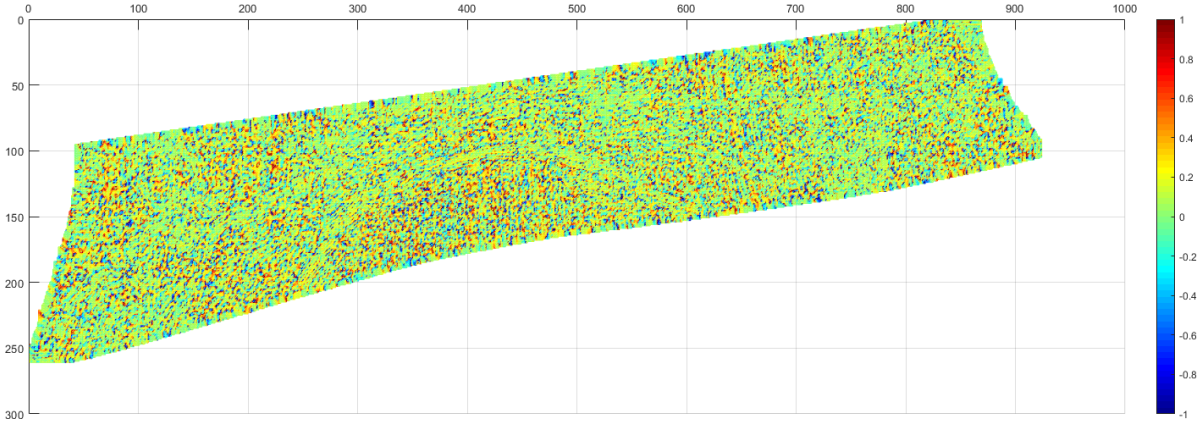


Figure 3.11: Spatial visualization of Gaussian method errors for constant scalar simulations.

In general, statistical method in general still works better than non-statistical method. However the wrong assumption of data can lead to extremely large boundary bias. If we remove these outliers, estimates from the wrong assumption still have higher boudary bias than HyBER. In other words, our model selection is essential to the reconstruction problem.

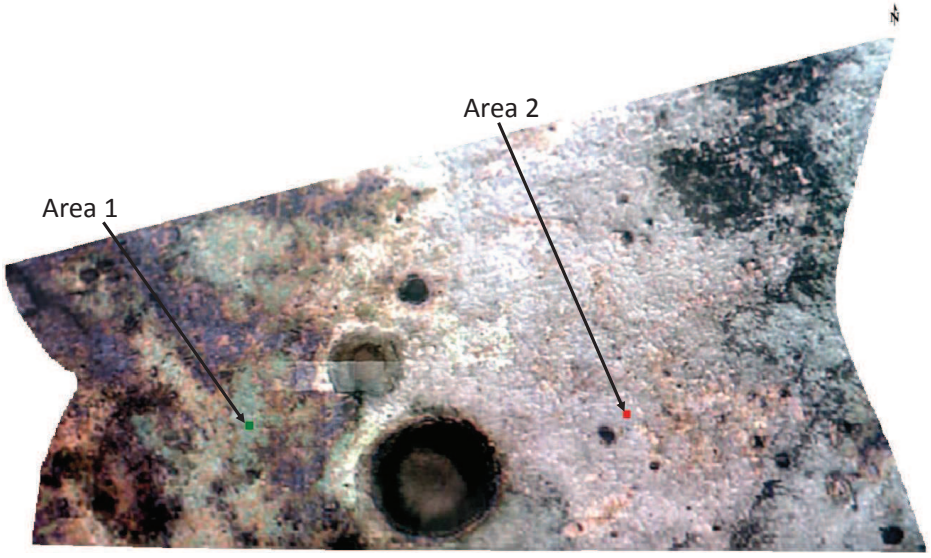
3.2.5 Application to CRISM data

The projected image from the sensor data $\mathbf{d}$ in (3.10) with the same desired pixel size based on the projected baseline (non-statistical) is compared with reconstructions from HyBER. Spectral comparisons verify the fidelity of original and denoised results. Spatial analysis quantifies the fidelity of different reconstructions.

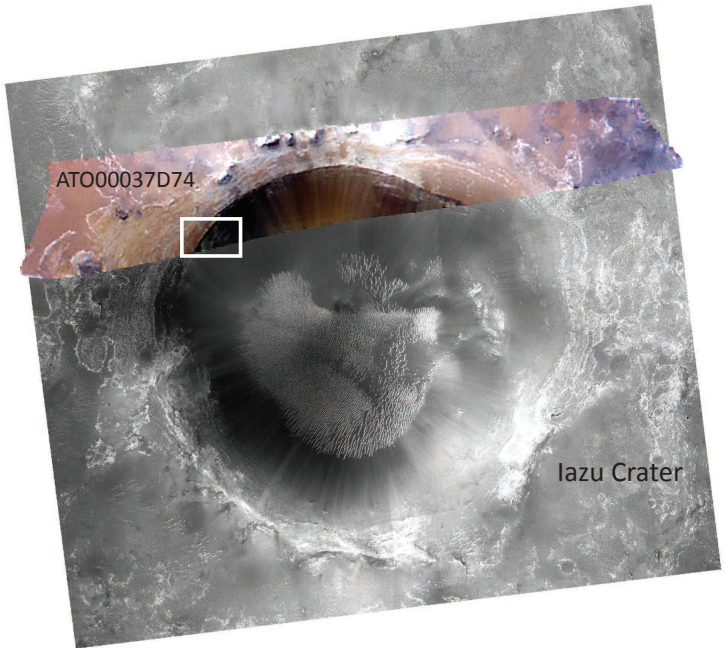
Spectral Analysis

To validate spectral features after application of our algorithm, FRT0001EB5F acquired over Mawrth Vallis is used. The area exposes nontronite and montmorillonite clay minerals [83–86]. ATO00037D74 was acquired over Iazu Crater in Meridiani Planum, south of the Mars Exploration Rover (Opportunity) landing site. This place shows evidence of nontronite and saponite [79] which has also been reported ~ 20 km away at Endeavour Crater [7].

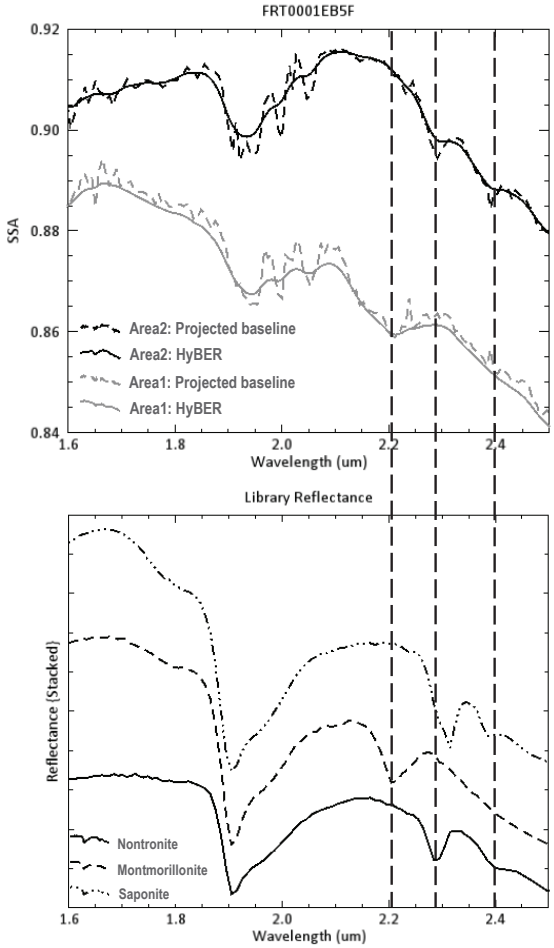
Fig. 3.12a shows the first scene with locations of two spectrally interesting areas delineated. In Fig. 3.13a, the top graph shows the mean single scattering albedos of two areas from the baseline and HyBER reconstructions, and the bottom shows the library spectra of nontronite, montmorillonite and saponite. The HyBER spectrum is denoised without removing



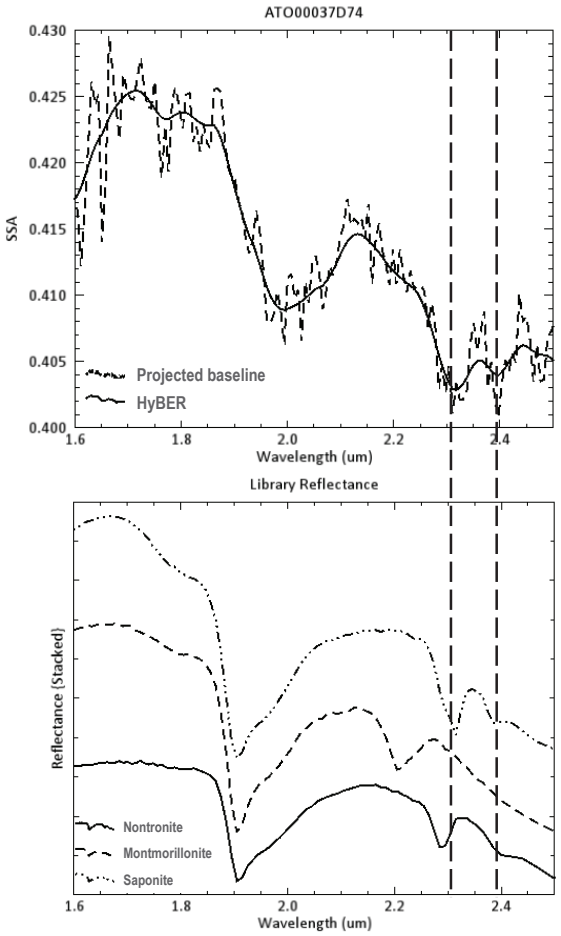
(a) The scene FRT0001EB5F over Mawrth Vallis and the spectra shown in (b) are from the two labeled areas. Reconstructions are shown in RGB with 2.529 μm, 1.506 μm and 1.060 μm. The frame is ~13.5 km across and data are regularized and projected at 12 m/pixel.



(b) Location of scene ATO00037D74 over Iazu Crater on Mars and the following spectral products are from the rectangular area. Reconstructions are shown in RGB with 2.529 μm, 1.506 μm and 1.060 μm. The frame of ATO00037D74 is ~10.5 km across and data are regularized and projected at 12 m/pixel.



(a) Spectral comparison between the projected baseline and HyBER reconstructions for FRT0001EB5F.



(b) Spectral comparison between the projected baseline and HyBER reconstructions for ATO00037D74.

Figure 3.13: Spectral comparison and analysis. The spectra shown is the mean spectra of 3 by 3 pixels and the vertical lines are centered on laboratory-based absorption features. The bottom graph is for laboratory-based spectra from USGS spectral library [1]. Nontronite, montmorillonite and saponite are candidate minerals for this area. Features at $1.9 \sim 2.0$ μm are OH/H_2O-related absorptions and smaller features at $2.2 \sim 2.3$ μm and 2.4 μm indicate the present of metal-OH.

diagnostic absorption features (2.2, 2.3 and 2.4 μm). Area 1 is a good match for the phyllosilicate mineral montmorillonite (matched absorption feature at around 2.2 μm) whereas area 2 contains nontronite (matched absorption features at around 2.3 and 2.4 μm).

Fig. 3.12b indicates the location of the scene ATO00037D74 over Iazu Crater. The mean single scattering albedos from the rectangular area are shown in Fig. 3.13b. The top of Fig. 3.13b indicates the spectral comparison between the baseline and HyBER reconstructions. Compared with Fig. 3.13a, baseline processed data from ATO00037D74 are noisy and it is harder to distinguish small features at 2.3 and 2.4 μm. Due to the spectral weighting regularization, fine signatures are evident with suppressed noise, even for the more recent scene suffering from enhanced noise relative to earlier scenes. Fig. 3.13b shows the evidence of saponite.

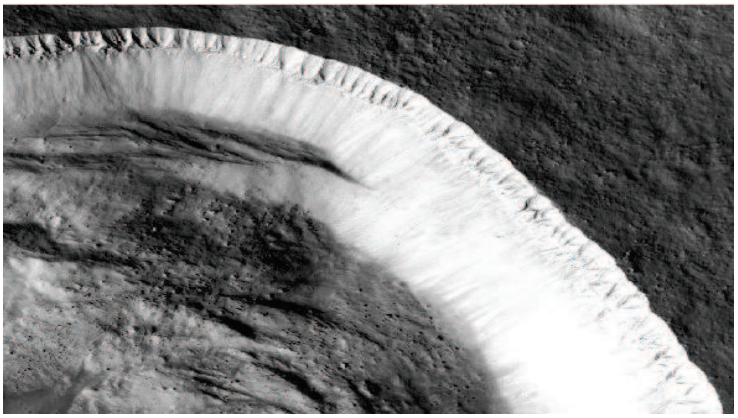
In summary, HyBER provides spectra in which noise is suppressed, whereas fine-scale mineral absorptions are preserved, both for relatively low-noise data acquired during 2011, and higher noise data collected during 2015.

Spatial Analysis

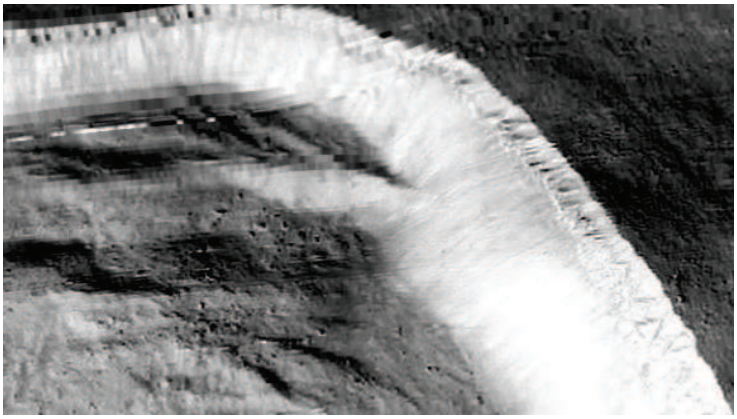
We use FRT0001DC22 acquired over Pangboche, a young impact crater on the slopes of the Olympus Mons volcano, using large boulders on its floor to characterize HyBER spatial fidelity. We validate the spatial characteristics using MRO Context Imager data (CTX) acquired at 6 m/pixel at wavelength $\sim$0.6 μm. Visual comparison in Fig. 3.14 shows that compared with the projected baseline, HyBER depicts more fine spatial details, i. e., discrete boulders [3].

[3]Codes can be obtained under my GIThub: `https://github.com/linyunhe/image_quality_CRISM_CTX`

(a) Context camera projected at 6 m/pixel at ~ 0.6 μm.



(b) HyBER reconstruction projected at 9 m/pixel at band 0.592 μm.



(c) Projected baseline at 9 m/pixel at band 0.592 μm.

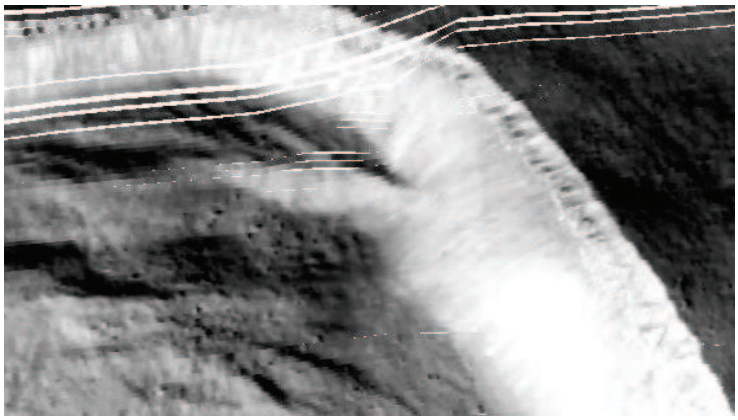


Figure 3.14: Part of Pangboche crater from three different reconstructions. FRT0001DC22 is reconstructed. Frames are $\sim$1.5 km across.

To quantify the quality of the reconstructed CRISM image data, both local and global comparisons are employed.

Spatial Local Analysis

Define the Rock Spread Function $RSF(x, y)$ to be the two-dimensional spline interpolation for pixels covering a spatial rock, where (x, y) is the relative spatial distance to the maximum or minimum of $RSF(x, y)$. $RSF_{CTX}(x, y)$, $RSF_{HyBER}(x, y)$ and $RSF_{Baseline}(x, y)$ are shown in 3.15. To quantify the reconstruction quality, we introduce two indexes:

$$
\begin{aligned}
&(1) Contrast: \\
&\quad contrast = \frac{\max_{x,y} RSF(x,y) - \min_{x,y} RSF(x,y)}{\frac{1}{2}\left(\max_{x,y} RSF(x,y) + \min_{x,y} RSF(x,y)\right)} \\
&(2) Sharpness: \\
&\quad sharpness = \frac{\max_{x,y} RSF(x,y) - \min_{x,y} RSF(x,y)}{width(0.1 \sim 0.9)}
\end{aligned}
\tag{3.15}
$$

where $width(0.1 \sim 0.9) = \min ||(x, y)_{0.9} - (x, y)_{0.1}||_2$ with $(x, y)_\eta = \{(x, y)|RSF(x, y) = \eta \max_{x,y} RSF(x, y) + (1 - \eta \min_{x,y} RSF(x, y))\}$, that is, the minimal distance between two brightness contours. The distance between maximum and minimum, $width(0.1 \sim 0.9)$ is chosen because the distance between two points is more sensitive than one between two contours. These two indexes can describe the local quality of an RSF since the contrast shows the amplitude range and the sharpness measures the slope.

In Fig. 3.15, the RSF of CTX has a 1.27 contrast and 0.014 sharpness, the RSF of HyBER has 0.63 contrast and 0.0059 sharpness, and the RSF of the projected baseline has 0.21 contrast and 0.0021 sharpness. The local quality of other rocks is compared in Table 3.4.

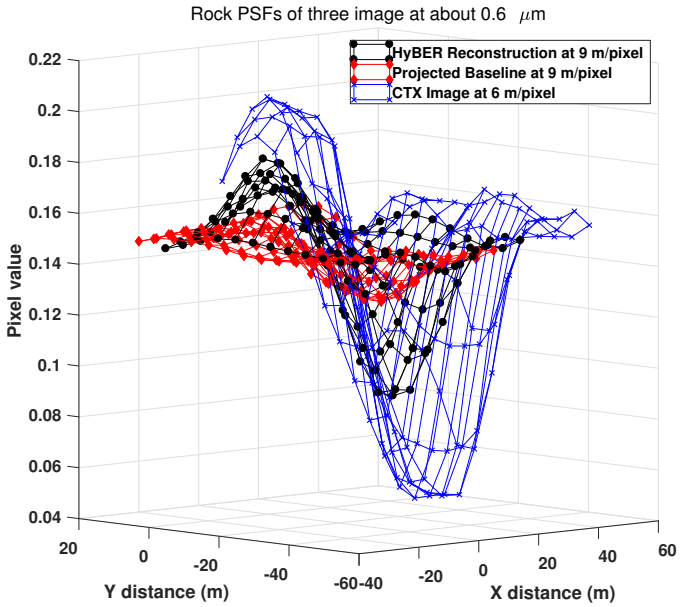


Figure 3.15: Rock Spread Functions (RSFs) of the same rock from CTX, HyBER and the projected baseline.

Table 3.4: Local goodness of four different rocks

Reconstruction Image	Rock 1		Rock 2		Rock 3	
	sharpness	*contrast*	*sharpness*	*contrast*	*sharpness*	*contrast*
CTX (6m)	0.01770	1.5595	0.01704	1.4937	0.01322	1.4396
HyBER (9m)	0.00787	0.7901	0.00602	0.5985	0.00348	0.4828
Baseline (9m)	0.00337	0.3634	0.00319	0.2702	0.00196	0.2210
HyBER (12m)	0.00643	0.6835	0.00516	0.5054	0.00372	0.5180
Baseline (12m)	0.00205	0.2567	0.00285	0.2630	0.00110	0.1388
HyBER (18m)	0.00284	0.4484	0.00271	0.3746	0.00286	0.4953
Baseline (18m)	0.00086	0.1438	0.00137	0.1726	0.00059	0.1097

Three conclusions can be drawn from Table 3.4:

1. Smaller pixel sizes lead to higher contrast and sharpness for both HyBER and projected baselines;

2. HyBER has at least twice the contrast and sharpness of the baseline method and the larger the pixel size is, the greater the improvement of HyBER is over the projected baseline method;

3. There necessarily is a gap between CTX at 6 m/pixel and HyBER at 9 m/pixel. This is not surprising given that the ATO mode only achieves up to 9 m/pixel for limited areas (non-uniform sampling).

Spatial Global Analysis

The global analysis is complicated because the hyperspectral image reconstruction algorithm is nonlinear, iterative, and spatially varying. As described above, the regularization is designed to remove some spatially varying artifacts with a goal of uniform spatial resolution properties across the scene. The spatial frequency analysis of a scene is used to quantify the global performance. The basis of comparison is the spatial resolution determined from the relative frequency context of the scene.

Let $F(\mu,\nu) = \sum_{k_y=0}^{M-1}\sum_{k_x=0}^{N-1} f\left(k_x,k_y\right) e^{-2\pi j(\mu k_x+\nu k_y)}$ be the Discrete Fourier Transform of the 2D image $f\left(k_x,k_y\right)$. Let $\rho = \sqrt{\mu^2+\nu^2}$ be the radial frequency and $J(\rho)$ be the average,

$$J(\rho) = \left\langle \frac{|F(\mu,\nu)|}{F(0,0)} \right\rangle, \text{ for } |\mu^2+\nu^2-\rho^2| < \triangle\rho \tag{3.16}$$

where $\langle\cdot\rangle$ denotes the average and $\triangle\rho$ is small enough.

For a given threshold τ, the spatial resolution r (in meters) is defined as the largest value that satisfies,

$$J(\rho) \geq \tau, \text{ for all } \rho < \frac{1}{r}. \tag{3.17}$$

Fig. 3.16 shows how $J(\rho)$ changes with the spatial radial frequency ρ for three reconstructions. It is clear that CTX is much better than the baseline and HyBER, which matches results shown in Fig. 3.14. The HyBER $J(\rho)$ is higher than one of the projected baseline

Table 3.5: Global resolution for different areas (unit: meter)

Reconstruction Image	Rocks	Crater Edge	Plain	Combination
CTX ($6m$)	11.0	11.5	10.4	10.6
HyBER ($9m$)	14.2	15.2	14.2	14.3
Baseline ($9m$)	20.3	23.5	20.3	20.9
HyBER ($12m$)	16.8	18.1	16.0	17.5
Baseline ($12m$)	25.2	28.5	24.9	27.3
HyBER ($18m$)	22.4	32.3	18.7	27.2
Baseline ($18m$)	34.4	42.7	33.4	39.3

when $0.018 < \rho < 0.095$. In other words, for small spatial features (smaller than 55 but larger than 11 meters), HyBER depicts them more clearly than the projected baseline.

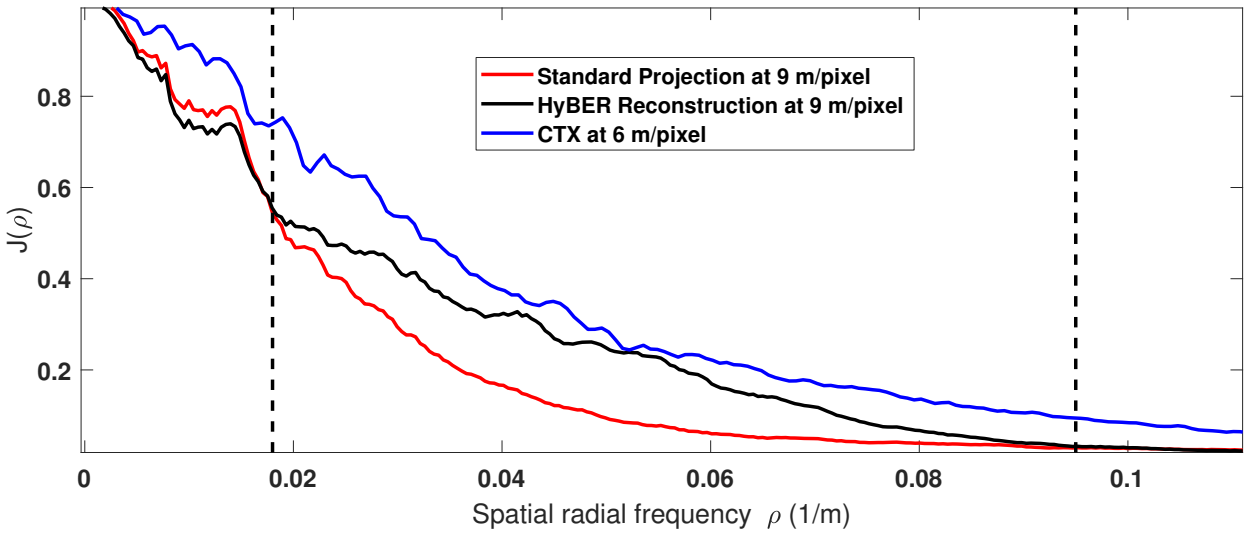


Figure 3.16: $J(\rho)$ depending on spatial radial frequency ρ. HyBER works much better than the baseline method when spatial frequency is inside the range of 0.018 to 0.095, that is, 10.5 to 55 meters spatially.

Table 3.5 shows the spatial resolution of three reconstructions for different areas, given the distinguishing threshold 0.1 in (3.17). Two conclusions are summarized:

1) CTX has better resolution than both HyBER and the projected baseline method at different projected pixel sizes (9, 12 and 18 m/pixel);

2) For both HyBER and projected baseline methods, decreasing the pixel size leads to better

the image resolution. Also, HyBER has better resolution than the projected baseline for all pixel sizes.

3.3 Summary

We propose a general method hypothesis-based estimation with regularization (HyBER) to reconstruct and denoise hyperspectral reconstruction data. First, a new hypothesis testing method is used to select a better data statistical distribution (Poisson and Gaussian distributions as candidates). Based on the selected model, a maximum log-likelihood method is used to reconstruct and denoise hyperspectral data. Algorithms for both Gaussian and Poisson models are stated. Moreover, a spatially and spectrally dependent weighting penalty is proposed to avoid artifacts introduced by non-uniform sampling issues and unstable noise environments.

Simulated Poisson data are generated to test the performance of HyBER, a statistical method with a wrong assumption and a non-statistical (the projected baseline) method. It is found that HyBER can reconstruct images with relative error (to ground truths) less than 0.002%, which is much smaller than others. Furthermore, reconstructions based on real CRISM data from HyBER for both FRT0001EB5F and ATO00037D74, which were collected four years apart, provide spectra in which noise is well suppressed but fine-scale mineral features are preserved. Moreover the quantitative comparison of contrast, sharpness and spatial resolution for the scene FRT0001DC22 show that the HyBER reconstruction is on average twice as good as the projected baseline.

Chapter 4

Conclusions, Impacts and Future Work

4.1 Conclusions and Future Work

We developed and evaluated a general pipeline for hyperspectral data and implemented it for Compact Reconnaissance Imaging Spectrometer for Mars from radiance measurements to denoised and projected single scattering albedos (SSAs). This pipeline includes two main parts:

1. Generating single scattering albedos and the surface temperature based on radiance (or IOF);

2. Reconstructing and denoising the single scattering albedos.

In Chapter 2, we introduce a general method, Separating Temperature and Albedo by Neural Networks (STANN), to retrieve the Single Scattering Albedo and the surface temperature from radiance. We also theoretically analyze the difficulty for different cases (different temperature ranges and different wavelengths) and explain the reason why previous similar

methods can work well. Then the STANN is derived and built for the hardest Mars spectrometer, CRISM. Performances for both instruments are verified in term of the robustness and the consistency. It is noticed that STANN for CRISM can obtain a relatively robust and correct temperature mapping based on limited information. The local incidence angle can be computed from the elevation file, which corrects spatial artifacts in generating albedos.

In Chapter 3, we propose a general method hypothesis-based estimation with regularization (HyBER) to reconstruct and denoise hyperspectral reconstruction data. First, a new hypothesis testing method is used to select a better data statistical distribution (Poisson and Gaussian distributions as candidates). Based on the selected model, a maximum log-likelihood method (HyBER) reconstructs and denoises hyperspectral data. Algorithms for both Gaussian and Poisson models are presented. Moreover, a spatially and spectrally dependent weighting penalty is proposed to avoid artifacts introduced by non-uniform sampling issues and unstable noise environments.

Simulated Poisson data are generated to test the performance of HyBER, a statistical method with a wrong assumption and a non-statistical (the projected baseline) method. It is found that HyBER can reconstruct images with relative error (to ground truths) less than 0.002%, which is much smaller than others. Furthermore, reconstructions based on real CRISM data from HyBER for both FRT0001EB5F and ATO00037D74, which were collected four years apart, provide spectra in which noise is well suppressed but fine-scale mineral features are preserved. The quantitative comparison of contrast, sharpness and spatial resolution for the scene FRT0001DC22 shows that the HyBER reconstruction is on average twice as good as the projected baseline.

To make this study more general and applicable, there are also several tasks to be considered in future work. First of all, so far this pipeline mainly works for CRISM and it is possible

to develop similar pipelines for other instruments such as OMEGA, THEMIS and TES. Other parts can also be added to make this pipeline more useful. For example, more spatial artifacts caused by the viewing conditions can be reduced if the local emission angle for the DISORT model is also computed. The projected and denoised SSA images can also be used to compute and classify the surface mineral compositions and their corresponding grain sizes, which in fact has been under research [8] for the Bagnold Dunes and compared with data from the Curiosity Rover.

Further analysis of temperatures retrieved from STANN and other thermal model (with smaller pixel size) can be considered to verify or modify the performance.

Fig. 4.1 summarizes the main work of this dissertation and some future work for the hyperspectral data processing pipeline. The results of the dissertation contribute mainly to the darker blue boxes and partially contribute to the light blue box. These results cannot work well without the significant help of other's work labeled in grey boxes. Potential future work to make this pipeline better is indicated in the white boxes. A similar pipeline including STANN and HyBER for OMEGA and THEMIS will be released soon.

We have proposed a hypothesis test to verify the best approximation of data distribution assuming that measurements are independent, thus it is promising to develop a method to test whether data are independently collected for both Poisson distribution and Gaussian distribution.

4.2 Impacts

Results from HyBER have helped planetary scientists to predict minerals on the Mars surface, which were proved to be real when the Opportunity rover arrived there. Marathon

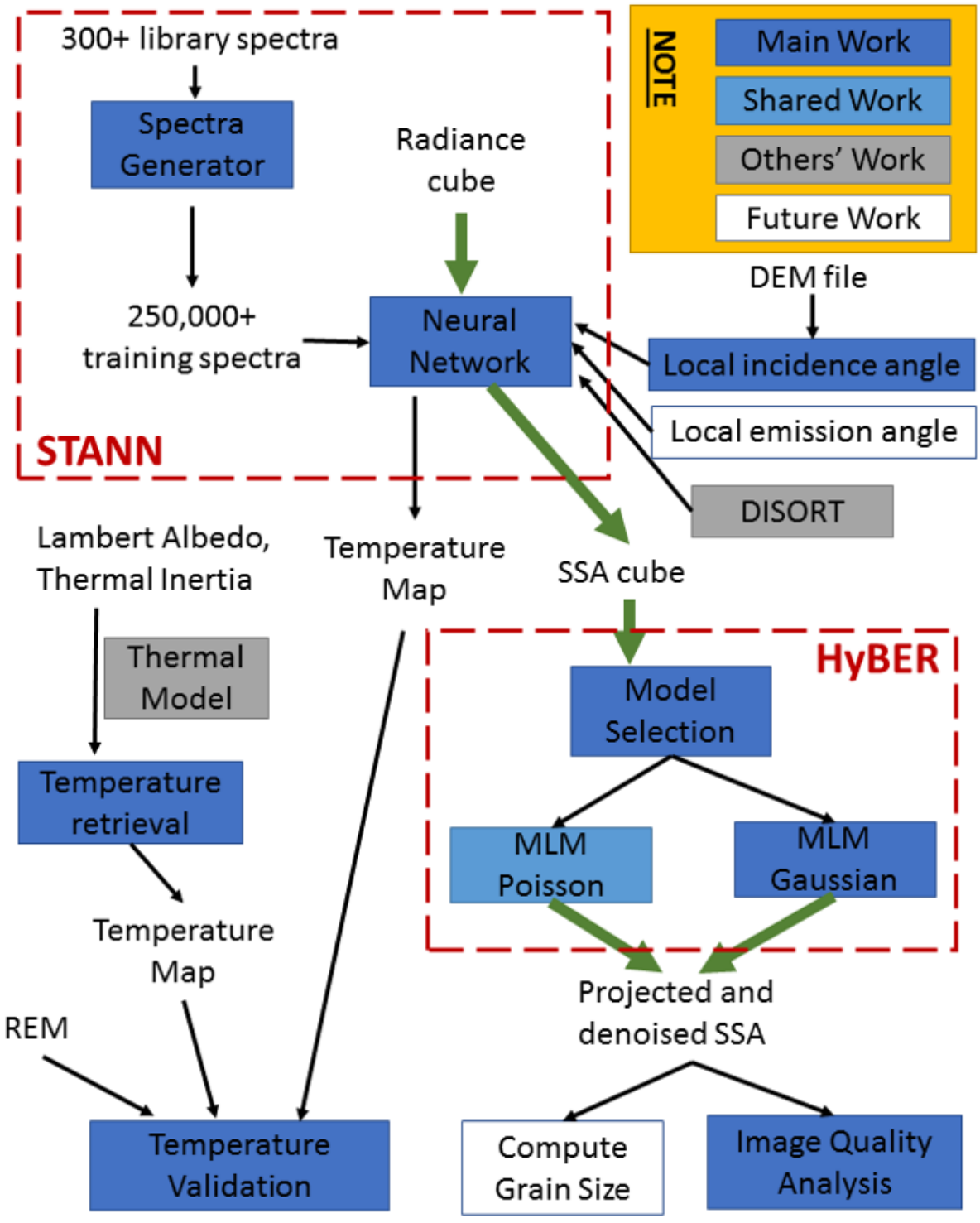


Figure 4.1: Summary of the dissertation work to the hyperspectral data (CRISM) processing pipeline. DISORT: Discrete Ordinates Radiative Transfer. DEM: Digital Elevation Map. REM: Rover Environmental Monitoring.

Valley, about 100 meters wide, crosscuts the Cape Tribulation rim segment of the 22 km diameter, Noachian-age Endeavour crater on Mars [7]. Fox, et al. found that resulting data from HyBER predicts the large areal extent of light-toned breccias within Marathon Valley, are the carriers of the $Fe^{3+} - Mg^{2+}$ smectite signatures. Given this prediction, the Opportunity rover was directed to explore Marathon Valley, characterized the outcrops with the smectite spectral signature and arrived in August 2015 [7]. Compositions of rocks scanned by the rover within the breccias are consistent with isochemical alteration forming $Fe^{3+} - Mg^{2+}$ smectites, with the exception of minor exposure of red pebbles within some of the fractures separating the light-toned breccia outcrops [7]. More compositional evidences by the rover are analyzed in detail by Mittlefehldt et al. [87].

Furthermore, HyBER outputs also enabled scientists to interpret the formation history of craters on Mars. Iazu is a 6.8 km diameter crater located about 25 km south of Noachian-age Endeavour crater on Mars [79]. Powell et al. drew a conclusion that Noachian basaltic crust changes through evaporative environments and finally, the current anhydrous environment dominated by aeolian processes, by analyzing HyBER results over Iazu's rim and walls [79]. Similarly, Rampe et al. compare our outputs of Gale crater [88] with observations from the Curiosity Rover.

More applications of HyBER outputs can be found in [8, 89, 90].

Before STANN, most scientists only focus on some bands for CRISM (from 0.6 µm to 2.5 µm), which are just 57% of the whole CRISM hyperspectral data due to the thermal effects. STANN provides a promising potential to increase the usage of CRISM data. Powell et al. have started to analyze the broad OH- and H_2O- related absorption features near 3.0 µm over the Curiosity [91] traverse in Gale crater [9, 92]. Hopefully, this analysis can be verified soon when the Curiosity rover arrives.

Moreover the temperature maps from STANN provide more thermal information about the Mars surface, which can work as supplementary data for other thermal instruments such as THEMIS.

Appendix A

Hypothesis

A.1 Poisson Distribution Test for Homogeneous Area

In this section, we introduce our codes[1] for the homogeneous area test in Chapter 3. As introduced in 3.1.1, we test whether the distribution of a homogeneous area is close to the scaled Poisson distribution and whether the scalars retrieved from different bands are similar.

Here we use the scene FRT0001EB5F as the example. Fig. A.1 shows the M instruction steps for MATLAB codes. We first run the main function "homogenous_scalar_poisson_test()" on the MATLAB command window and then three browsers show up, asking for SSA input file, WA input file and output folder respectively. Then the forced color image (input) is displayed and the user needs to select the homogeneous area by drawing any shape on the image (see the blue circle in the middle). After the selection of the homogeneous area, the scalars for different bands are plotted. If the user is satisfied with this selection, input "0" in the box and click "OK". Otherwise the user can input anything except for "0" to re-select the area. Then all necessary steps by users are done and the code computes distribution.

[1]Codes and instructions can also be found on my GIThub (No.3): `https://github.com/linyunhe/hypothesis_test_PG_for_CRISM`

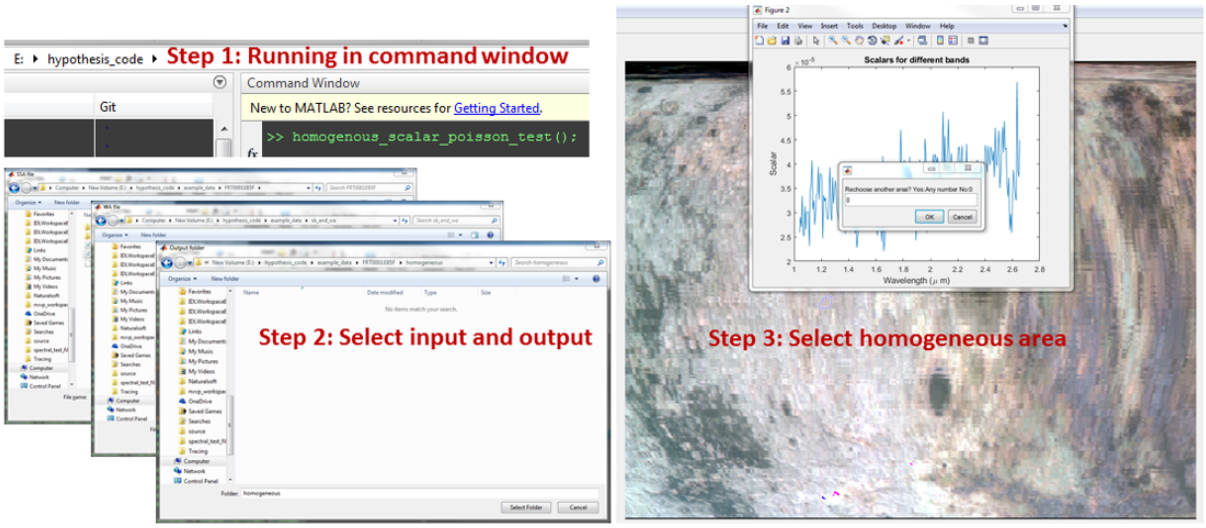


Figure A.1: Instruction for the input and running of the homogeneous area test codes.

After a few minutes, outputs are all saved into the output folder as shown in Fig. A.2. There are three outputs: one video named "Poisson_test_plot_cdf.avi" and two MATLAB figures named "Scalar_bands.fig" and "Spat_area_homo.fig". "Poisson_test_plot_cdf.avi" records all plots of distributions of this homogeneous area for all bands (for CRISM L-data, 246 bands). The figure in the green box shows the example of one plot at band 33. "Spat_area_homo.fig" saves the homogeneous area you selected before in red overlaid with the original image and "Scalar_bands.fig" plots retrieved scalars for different bands of this area. For our example scene, we can find that this scene contains too many spatial features and it is very hard to find a clean and large homogeneous area. Even for a small area, the scalars we retrieved show it is not as stable as Fig. 3.4. Also the Poisson approximation of this area is also not as good as Fig. 3.5.

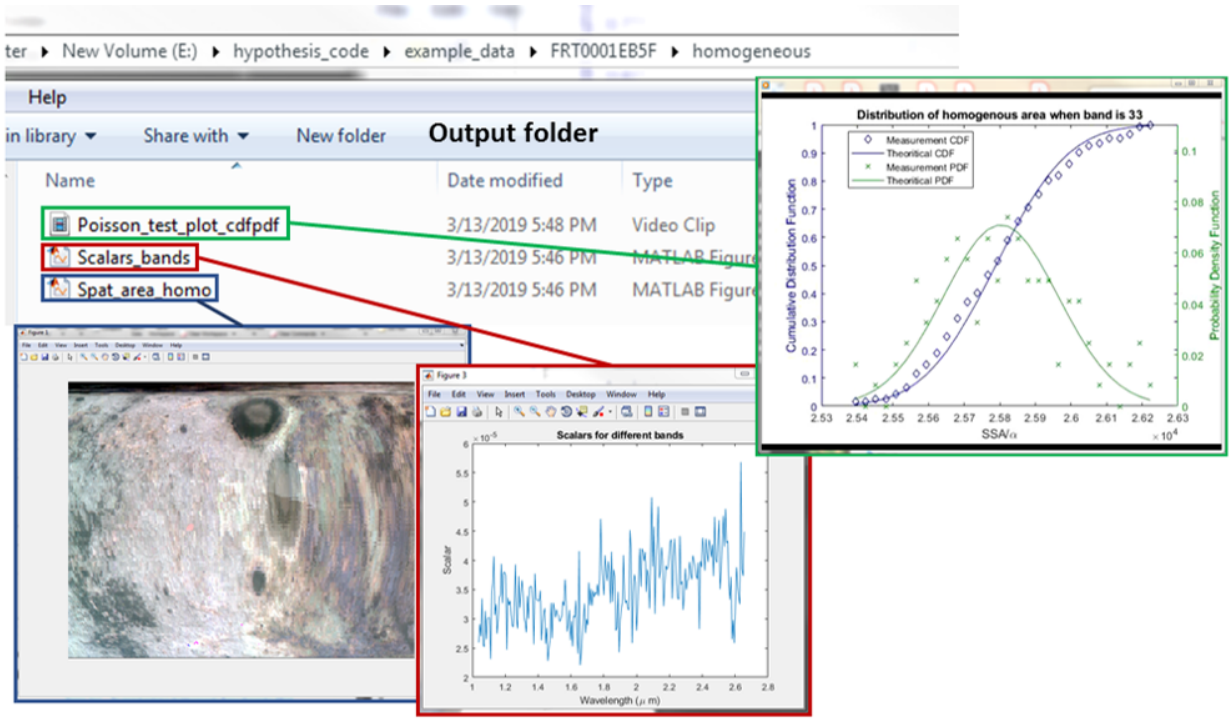


Figure A.2: Explanation for the outputs of the homogeneous area test codes.

A.2 Model Selection between Scaled Poisson and Gaussian

In this section, we introduce our model selection codes[2] in details.

As introduced in 3.1.1, our model selection is based on the best estimates retrieved from the maximum loglikelihood method for candidate distributions such as scaled Poisson and White Gaussian for our case. Then we compute the d_{KL} or plot the p-value figure to see which assumption works better.

Here we also use the scene FRT0001EB5F as the example. Fig. A.3 shows the instruction for this scene from the input folder to the MATLAB code running. The input folder must

[2]Codes and instructions can also be found on my GIThub (No.2): `https://github.com/linyunhe/hypothesis_test_PG_for_CRISM`

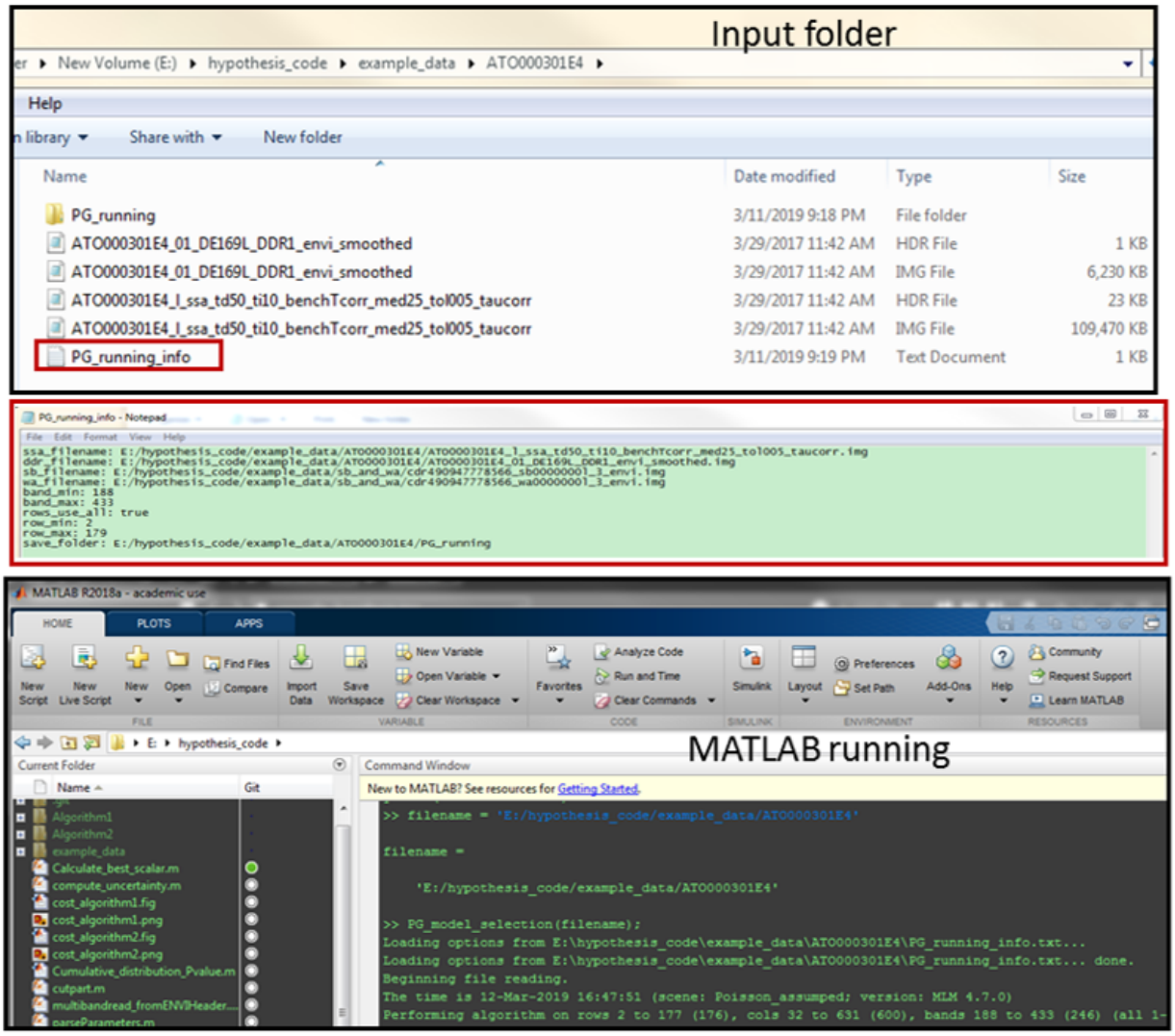


Figure A.3: Instruction for the input and running of model selection codes.

contain the SSA, DDR, an empty folder named "PG_running" to save the outputs and an input text file named "PG_running_info .txt." This file records the directory of SSA, DDR, SB and WA files [3], and the bands (consistent with the SSA data) you want to run. In the example, we only run for the L-band data. We also provide the selection of sub-rows. If you want to run the sub-scene, just set the "rows_use_all" as false and insert the subset of rows you want to run in "row_min" and "row_max."

[3]SB and WA files are always the same for the same instrument

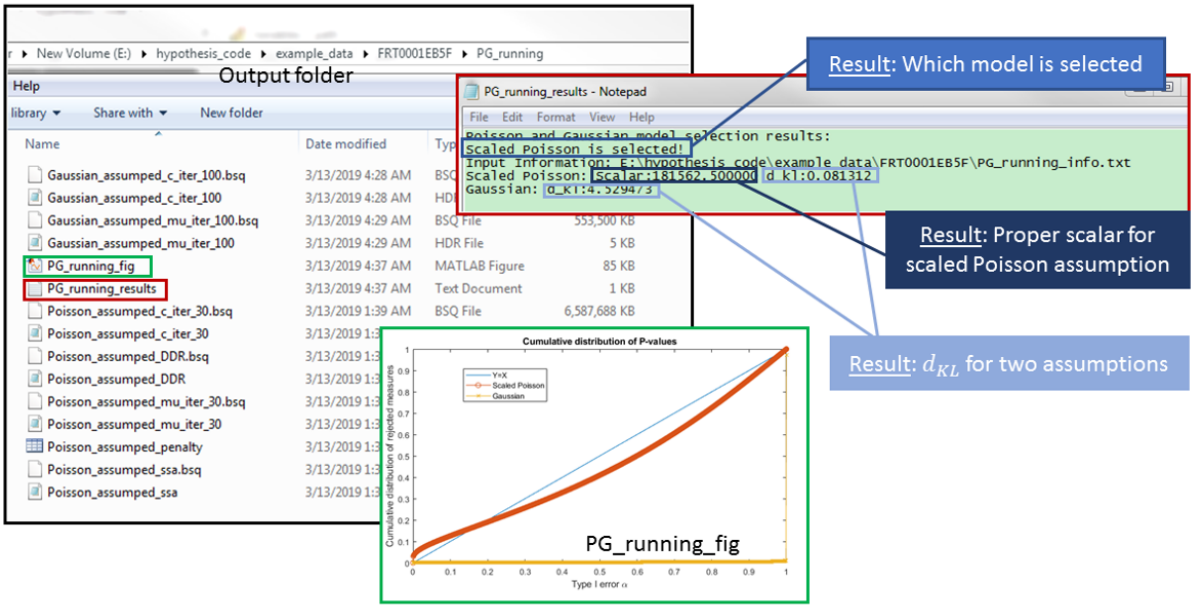
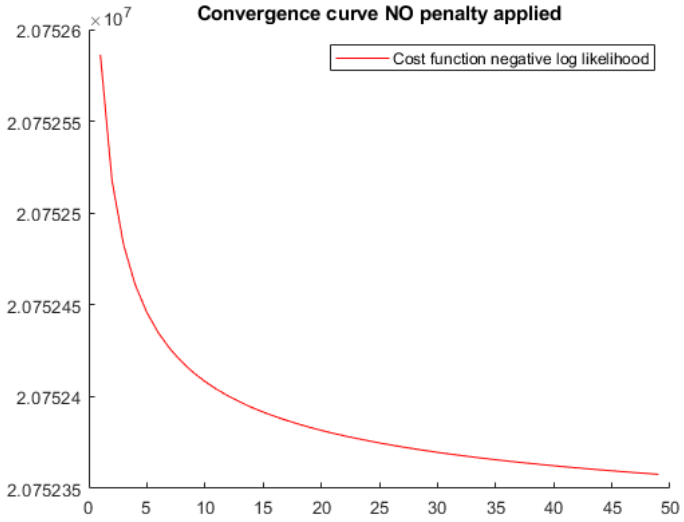


Figure A.4: Explanation for the outputs of model selection codes.

After running[4], outputs are all saved into the output folder as shown in Fig.∼A.4. The main results are saved in the text file named "PG_running_results.txt": the second line indicates which statistical model is selected and the corresponding d_{KL} for two statistical models are saved. The p-value figure is saved as "PG_running_fig.fig": closer the curve is to the diagonal line, the better the statistical assumption is to describe the input data. For this example scene FRT0001EB5F, the scaled Poisson distribution is selected and the scaled Poisson distribution has small $d_{KL} = 0.08$ while the Gaussian distribution has $d_{KL} = 4.53$. Also we notice that Poisson p-value curve is very close to the diagonal line but the Gaussian p-value curve is really far away. One explanation is that because this scene has too many features and the inputted L-band SSAs vary from 0.7 to 0.9 (larger variance compared with other scenes), the scaled Poisson distribution can not be approximated by the Gaussian distribution especially for the large scalar (1.8×10^5).

[4]About 1 hour and 20 minutes for the 110 MB scene on the MATLAB 2018a, Windows 7, i7-5820K CPU, 32GB RAM

(a) Cost function curve for Algorithm 1 with no penalty.



(b) Cost function curve for Algorithm 2 with no penalty.

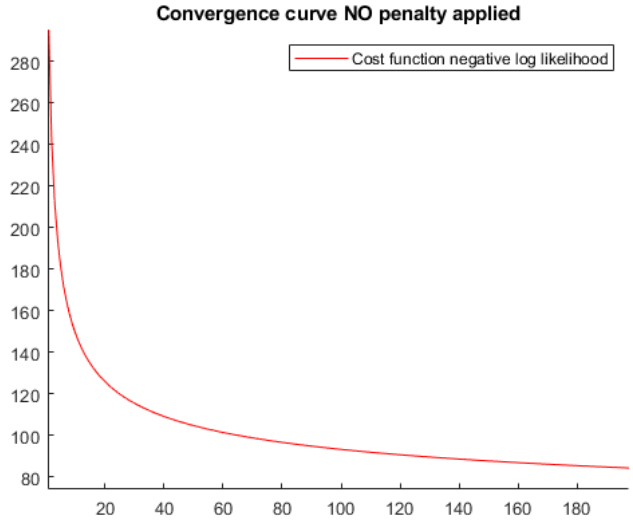


Figure A.5: Cost function decreasing curve for two algorithms.

The numbers of iterations for Algorithm 1 and Algorithm 2 are defaulted as 30 and 100 based on the cost function curve shown in Fig.∼A.5. Around 30 iterations and 100 iterations, our algorithms converge. We also find that Algorithm 1 converges faster than Algorithm 2.

Appendix B

Convergence Proof of Algorithms

The derivation of Algorithm 1 and 2 is included in the following appendix for the non-penalty data fitting term, because the decoupling of the penalty is included in section 3.1.3.

B.1 Algorithm 1

Lemma B.1 (Expectation Maximization Lemma)**.** *Let* $p(\mathbf{d}|\mathbf{c})$ *be the probability of measurements* $\mathbf{d}$ *given the image* $\mathbf{c}$*. Then for any intermediate random variable* x*,*

$$\ln p(\mathbf{d}|\mathbf{c}) \geq \sum_x p(x|\mathbf{d}, \mathbf{c}^{(t)}) \ln \frac{p(x, \mathbf{d}|\mathbf{c})}{p(x|\mathbf{d}, \mathbf{c}^{(t)})}, \tag{B.1}$$

where based on guessed image from last iteration $\mathbf{c}^{(t)}$*, and the equality holds iff* $\mathbf{c} = \mathbf{c}^{(t)}$*.*

Proof of the Expectation Maximization Lemma. Based on the concavity of $\ln p$ and Jensen's inequality, we have

$$\ln p(\mathbf{d}|\mathbf{c}) = \ln \sum_x p(\mathbf{d}, x|\mathbf{c}) \geq \sum_x q(x) \ln \frac{p(\mathbf{d}, x|\mathbf{c})}{q(x)}, \tag{B.2}$$

for all $q(x)$ where $\sum_x q(x) = 1$.

For the known $\mathbf{c}^{(t)}$, we can obtain $q^{(t)}(x)$ by maximizing $\sum_x q(x) \ln \frac{p(\mathbf{d},x|\mathbf{c}^{(t)})}{q(x)}$, thus,

$$q^{(t)}(x) = \frac{p(\mathbf{d}, x|\mathbf{c}^{(t)})}{\sum_x p(\mathbf{d}, x|\mathbf{c}^{(t)})} = p(x|\mathbf{d}, \mathbf{c}^{(t)}).$$

By replacing $q(x)$ in (B.2) by $q^{(t)}(x)$, Lemma B.1 is proved. □

From Lemma B.1, we can know that after t iterations, the loglikelihood is bounded by the decoupled function $\sum_x q(x) \ln \frac{p(\mathbf{d},x|\mathbf{c})}{q(x)}$, thus we can achieve the global minimum by iteratively solving (after removing the constant term):

$$\max_{\mathbf{c}} \sum_x p(x|\mathbf{d}, \mathbf{c}^{(t)}) \ln p(x, \mathbf{d}|\mathbf{c}) = \max_{\mathbf{c}} E_{x|\mathbf{d},\mathbf{c}^{(t)}} \left[\ln p(x, \mathbf{d}|\mathbf{c})\right].$$

If we select the intermediate random variable be $x_{ij} = H_{ij}c_j$, based on the Poisson additive property, we have $x_{ij} \sim Poisson(H_{ij}c_j)$, $d_i = \sum_j x_{ij}$. Replacing $\ln p(x_{ij}, \mathbf{d}|\mathbf{c}) = -H_{ij}c_j + x_{ij} \ln H_{ij}c_j - \ln X_{ij}!$ above, we can simplify it as:

$$\max_{\mathbf{c}} E_{x|\mathbf{d},\mathbf{c}^{(t)}} \left[\ln p(x, \mathbf{d}|\mathbf{c})\right] = \max_{\mathbf{c}} \sum_{ij} \left[-H_{ij} + E_{x|\mathbf{d},\mathbf{c}^{(t)}}[x_{ij}] \ln H_{ij}c_j\right].$$

Because for a given sum d_i (Poisson distributed), the conditional probability of x_{ij} is multinomial, we have $E_{x|\mathbf{d},\mathbf{c}^{(t)}}[x_{ij}] = \frac{H_{ij}c_j^{(t)}d_i}{\sum_k H_{ik}c_k^{(t)}}$. Finally,

$$\begin{aligned} c_j^{(t+1)} &= \arg\max_{c_j} \left[\sum_i \frac{H_{ij}d_i}{\sum_k H_{ik}c_k^{(t)}}\right] c_j^{(t)} \ln c_j - \left(\sum_i H_{ij}\right) c_j \\ &= \arg\max_{c_j} f_j c_j^{(t)} \ln c_j - h_j c_j. \end{aligned} \tag{B.3}$$

B.2 Algorithm 2

As we showed before, maximizing the loglikelihood is equivalent to minimizing the cost function (without regularization):

$$\min_{\mathbf{c}} \sum_i (\sum_j H_{ij} c_j - d_i)^2 \tag{B.4}$$

where H_{ij} is the entry of the matrix H.

This cost function is hard to minimize is due to the coupled c_j in the term $(\sum_j H_{ij} c_j - d_i)^2$. There exists a lemma [93]:

Lemma B.2 (Convex Decomposition Lemma)**.** *Let $f(\mathbf{x})$ be a convex function with $\mathbf{x} \in \mathcal{R}^P$ be a reference point. Let $r_j \geq 0$ such that $\sum r_j <= 1$. Then*

$$f(\mathbf{x}) \leq \sum_j r_j f(\mathbf{x}^* + \frac{x_j - x_j^*}{r_j} \mathbf{e_j}), \tag{B.5}$$

where $\mathbf{e_j}$ denotes a column vector with all zeros except 1 at the j^{th} entry. Equality is obtained iff $(x_j - x_j^*)/r_j$ is independent of j.

Based on Lemma B.2, we can decouple $(\sum_j H_{ij} c_j - d_i)^2$ in (B.4) as:

$$\begin{aligned}
(\sum_j H_{ij} c_j - d_i)^2 &\leq \sum_j r_j (\sum_k H_{ik} c_k^{(t)} - d_i + H_{ij} \frac{c_j - c_j^{(t)}}{r_j})^2 \\
&= \sum_j r_j (a_i^{(t)} - d_i + H_{ij} \frac{c_j - c_j^{(t)}}{r_j})^2 \\
&\sim \sum_j 2 H_{ij} (a_i^{(t)} - d_i)(c_j - c_j^{(t)}) + \sum_j \frac{H_{ij}^2}{r_j} (c_j - c_j^{(t)})^2.
\end{aligned} \tag{B.6}$$

Thus the cost function in (B.4) bounded by decoupled function:

$$\min_{\mathbf{c}} \sum_j \left[\sum_i H_{ij}(a_i - d_i) \right] (c_j - c_j^{(t)}) + \sum_j \frac{\sum_i H_{ij}^2}{r_j} (c_j - c_j^{(t)})^2, \tag{B.7}$$

after t iterations. Thus this optimization function can be computed iteratively.

We select $r_j = \frac{\sum_i H_{ij}^2}{\sum_i H_{ij}}$, thus (B.7) can be written as:

$$\min_{\mathbf{c}} \sum_j [\sum_i H_{ij}(a_i - d_i)](c_j - c_j^{(t)}) + h_j(c_j - c_j^{(t)})^2]. \tag{B.8}$$

In the following, we will prove the selection of r_j can guarantee the decreasing. By minimizing (B.8), the update rule based on last iteration t can be retrieved:

$$\mathbf{c}^{(t+1)} = \mathbf{c}^{(t)} + A^{-1}H^T(\mathbf{d} - \mathbf{a}^{(t)}), \tag{B.9}$$

where A is a diagonal matrix with $\mathbf{d}$ as the diagonal entries.

Then the error function $||\mathbf{d} - H\mathbf{c}^{(t+1)}||^2$ can be computed as:

$$\begin{aligned} ||\mathbf{d} - H\mathbf{c}^{(t+1)}||^2 &= ||\mathbf{d} - H\mathbf{c}^{(t)} - HA^{-1}H^T(\mathbf{d} - H\mathbf{c}^{(t)})||^2 \\ &= ||(I - HA^{-1}H^T)(\mathbf{d} - H\mathbf{c}^{(t)})||^2. \end{aligned} \tag{B.10}$$

Theorem B.1 (Perron-Frobenius Theorem)**.** *Let $M = (m_{ij})$ be an $n \times n$ positive definite matrix. The largest eigenvalue of M should be no more than* $\max_i \sum_j m_{ij}$.

Let $M = I - HA^{-1}H^T$, according to the Perron-Frobenius theorem, the largest eigenvalue should be no more than the maximum entry of $(I - HA^{-1}H^T)\mathbf{1} = \mathbf{1} - H\mathbf{1}$. Because H is the Gaussian-shape kernel, all entries in $H\mathbf{1}$ should be positive and less than 1. Thus the

largest eigenvalue of $(I - HA^{-1}H^T)$ should be no more than 1. In another words, if we eigen-decompose $(I - HA^{-1}H^T) = Q\Gamma Q^T$, $\Gamma \leq I$. Thus,

$$\begin{aligned}
||\mathbf{d} - H\mathbf{c}^{(t+1)}||^2 &= ||\mathbf{d} - H\mathbf{c}^{(t)} - HA^{-1}H^T(\mathbf{d} - H\mathbf{c}^{(t)})||^2 \\
&= ||(I - HA^{-1}H^T)(\mathbf{d} - H\mathbf{c}^{(t)})||^2 \\
&= (\mathbf{d} - H\mathbf{c}^{(t)})^T Q\Gamma Q^T Q\Gamma Q^T(\mathbf{d} - H\mathbf{c}^{(t)}) \\
&= (\mathbf{d} - H\mathbf{c}^{(t)})^T Q\Gamma^2 Q^T(\mathbf{d} - H\mathbf{c}^{(t)}) \\
&\leq (\mathbf{d} - H\mathbf{c}^{(t)})^T QIQ^T(\mathbf{d} - H\mathbf{c}^{(t)}) = (\mathbf{d} - H\mathbf{c}^{(t)})^T(\mathbf{d} - H\mathbf{c}^{(t)}) = ||(\mathbf{d} - H\mathbf{c}^{(t)})||^2.
\end{aligned} \tag{B.11}$$

In another words, Algorithm 2 can decrease the original cost function and converges to the global minimum.

Appendix C

Local Incidence Angle

C.1 Motivation and Methodology

In Chapter 2, for fine spatial features less than 250 meters, we cannot model them well based on current incident angles (around 250 meters per pixel). Therefore a local incident angle in smaller pixel size can help us to remove the small spatial artifacts caused by the rough incident angles. In this appendix, we introduce the idea of computing the local incident angles from the digital elevations.

The digital elevation map (DEM) used to populate a high spatial resolution local incidence angle DDR plane was derived from stereo HiRISE image data sets. It is the product produced by the United States Geological Survey Astrogeology Science Center and posted as `https://astrogeology.usgs.gov/search/map/Mars/MarsScienceLaboratory/Mosaics/MSL_Gale_DEM_Mosaic_10m`. Fig.∼C.1 indicates areas that the elevation file DEM covers and the CRISM FRT000B6F1 covers. For every pixel in CRISM FRT000B6F1, you can find the corresponding elevation value and compute a more accurate local incident angle.

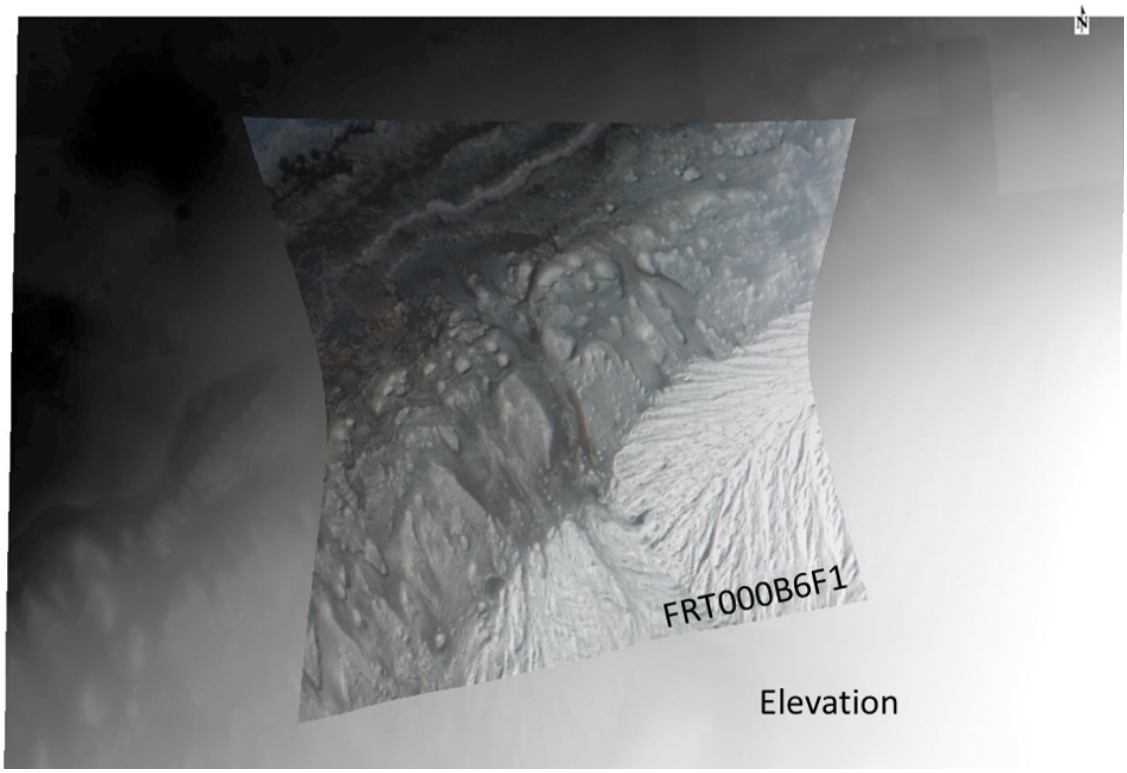


Figure C.1: After registration, the elevation file DEM can be overlapped with CRISM FRT000B6F1 in the projection. DEM elevation is in 1 m/pixel and CRISM FRT000B6F1 is in 12 m/pixel.

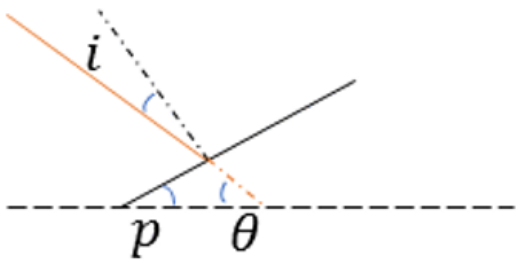
Let i be the local incident angle of the solar radiance, p be the elevation angle of the surface above the local horizon (along the solar incident direction). The sun is located at an azimuth of S degree clockwise from north, and has an elevation angle above the local horizon of θ.

Fig.∼C.2 shows four different cases of how the incident angles i can be computed from p and θ. When the surface has relative small elevation angle p, the incident angle $i = 90° - (p + \theta)$ as shown in Fig.∼C.2A. When the surface has large p as shown in B, the reflected solar radiance still goes back the sky, thus the incident angle $i = (p + \theta) - 90°$. However when the surface has elevation angle larger than 90° and also very precipitous as shown in C, it is possible that the reflected radiance cannot go back to the sky. For this case, although we can still compute the incident angle, in fact we get nothing back from the surface. Because the surface elevation angle is defined along the solar direction (at an azimuth of S degree clockwise from north), it is possible that the p is very large but the surface is likely horizontal as shown in D. For this case, the incident angle can be obtained as $i = 270° - (p + \theta)$. In general, the incident angle can be computed by

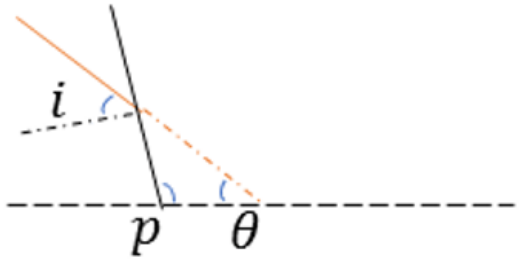
$$i = |90° - (\theta + p)\%180°| \tag{C.1}$$

for all cases where % is the modulo operator.

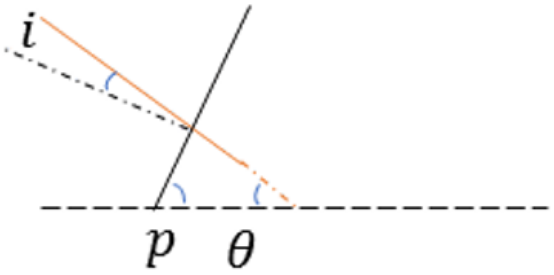
Then we need to compute the surface elevation angle p along the solar direction (at an azimuth of S degree clockwise from north). Fig.∼C.3 indicates the top view of the horizontal surface (the dash line in Fig.∼C.2). For every pixel in CRISM, we can know the local spatial location axes (x, y). Then along the solar direction, we can find the previous point (x_{pre}, y_{pre}) and the post point x_{post}, y_{post} with the northern distance $l/2$ where l is the target pixel size in meter. Let the elevation for the point (x, y) be $dem(x, y)$, then the elevation surface degree



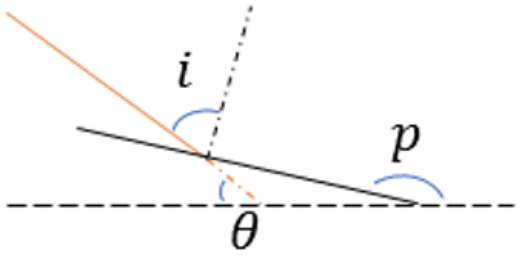
A. $i = 90° - (p + \theta)$



C. $i = (p + \theta) - 90°$



B. $i = (p + \theta) - 90°$



D. $i = 270° - (p + \theta)$

Figure C.2: The vertical slice display along the solar incident direction at an azimuth of S degree clockwise from north. Four different cases of the incident angles are shown. The red solid line indicates the incidence of the solar light and the black solid line is the local flat surface. The dash-dot line indicates the vertical direction of the local surface and the dash line is the horizontal surface.

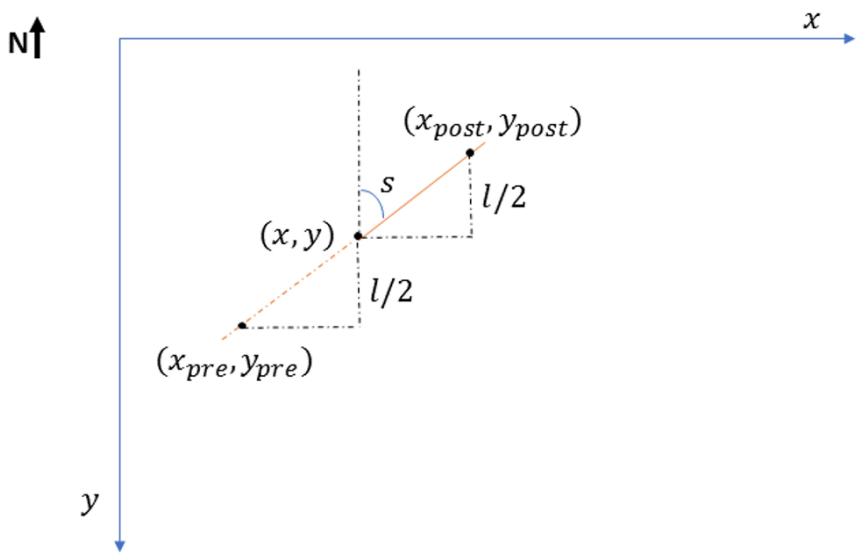


Figure C.3: The top view of the horizontal surface. (x, y) show the spatial location of one point. l is the target pixel size length in meter. S is an azimuth degree of the solar clockwise from north.

along the solar direction is

$$p = \arctan\left[\frac{dem(x_{pre}, y_{pre}) - dem(x_{post}, y_{post})}{\sqrt{1 + \tan S^2} \times l/2}\right]. \tag{C.2}$$

C.2 Results Comparison

Now we applied the method before to the scene FRT000B6F1[1] and Fig.∼C.4 shows the comparison between the original incident angles and our new local incident angles at the same pixel size 250 m/pixel.

It is noticed that our results have similar value range and the general pattern with the original ones. Moreover our new local incident angles have more details (corresponding to

[1]Open-source Python codes can be found under my GIThub: `https://github.com/linyunhe/temperature_on_thermal_model_Ashwin`.

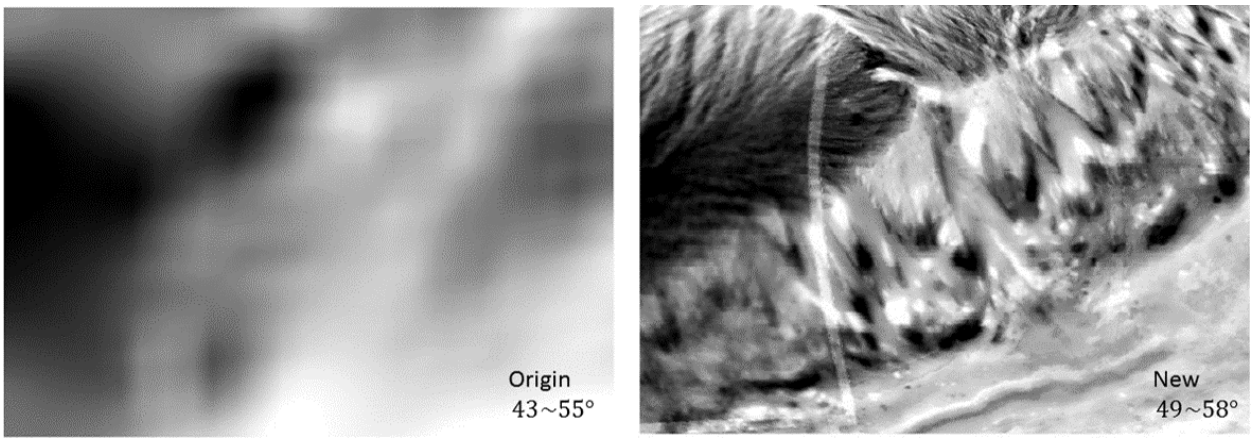


Figure C.4: Results comparison of the original incident angles and our new local incident angles. The original incident angles from the pixel size around 250 m/pixel is in the left and our generated local incident angles at the same pixel size is shown in the right.

the local shape of the surface) compared with the original angles. For example, the original incident angles almost have no information in the middle area where there exist lots of small features, while our new local incident angles contain these details.

In summary, although we still have some mismatch and artifacts for generating local incident angles, it is still promising to generate local angles by computing the elevations. All these disadvantages will be modified after this dissertation and updated codes and results can be found under my GIThub `https://github.com/linyunhe/temperature_on_thermal_model_Ashwin`.

Bibliography

[1] R. N. Clark, G. A. Swayze, R. Wise, K. E. Livo, T. Hoefen, R. F. Kokaly, and S. J. Sutley, "Usgs digital spectral library splib06a," *US geological survey, digital data series*, vol. 231, p. 2007, 2007.

[2] K. Stamnes, S.-C. Tsay, W. Wiscombe, and K. Jayaweera, "Numerically stable algorithm for discrete-ordinate-method radiative transfer in multiple scattering and emitting layered media," *Applied optics*, vol. 27, no. 12, pp. 2502–2509, 1988.

[3] F. D. Van der Meer, H. M. Van der Werff, F. J. Van Ruitenbeek, C. A. Hecker, W. H. Bakker, M. F. Noomen, M. Van Der Meijde, E. J. M. Carranza, J. B. De Smeth, and T. Woldai, "Multi-and hyperspectral geologic remote sensing: A review," *International Journal of Applied Earth Observation and Geoinformation*, vol. 14, no. 1, pp. 112–128, 2012.

[4] C.-I. Chang, *Hyperspectral imaging: techniques for spectral detection and classification*. Springer Science & Business Media, 2003, vol. 1.

[5] S. L. Murchie, F. P. Seelos, C. D. Hash, D. C. Humm, E. Malaret, J. A. McGovern, T. H. Choo, K. D. Seelos, D. L. Buczkowski, M. F. Morgan *et al.*, "Compact reconnaissance imaging spectrometer for mars investigation and data set from the mars reconnaissance orbiter's primary science phase," *Journal of Geophysical Research: Planets*, vol. 114, no. E2, 2009.

[6] F. Seelos, G. Romeo, C. Hash, S. Murchie, and E. Garhart, "Next generation crism multispectral map of mars: Noise reduction and radiometric reconciliation," in *Lunar and Planetary Science Conference*, vol. 48, 2017.

[7] V. Fox, R. Arvidson, E. Guinness, S. McLennan, J. Catalano, S. Murchie, and K. Powell, "Smectite deposits in marathon valley, endeavour crater, mars, identified using crism hyperspectral reflectance data," *Geophysical Research Letters*, vol. 43, no. 10, pp. 4885–4892, 2016.

[8] M. G. Lapotre, B. Ehlmann, S. E. Minson, R. Arvidson, F. Ayoub, A. Fraeman, R. Ewing, and N. Bridges, "Compositional variations in sands of the bagnold dunes, gale crater, mars, from visible-shortwave infrared spectroscopy and comparison with ground truth from the curiosity rover," *Journal of Geophysical Research: Planets*, vol. 122, no. 12, pp. 2489–2509, 2017.

[9] K. Powell, R. Arvidson, L. He, D. Politte, J. O'Sullivan, S. Murchie, and R. Morris, "Retrieval of crism single scattering albedos from 1-3.8 μm over the curiosity rover traverse," 2018.

[10] C. Kreisch, J. O'Sullivan, R. Arvidson, D. Politte, L. He, N. Stein, J. Finkel, E. Guinness, M. Wolff, and M. Lapôtre, "Regularization of mars reconnaissance orbiter crism along-track oversampled hyperspectral imaging observations of mars," *Icarus*, vol. 282, pp. 136–151, 2017.

[11] B. Hapke, *Theory of reflectance and emittance spectroscopy*. Cambridge University Press, 2012.

[12] A. Gillespie, S. Rokugawa, T. Matsunaga, J. S. Cothern, S. Hook, and A. B. Kahle, "A temperature and emissivity separation algorithm for advanced spaceborne thermal emission and reflection radiometer (aster) images," *IEEE transactions on geoscience and remote sensing*, vol. 36, no. 4, pp. 1113–1126, 1998.

[13] X. L. Ma, Z. Wan, C. C. Moeller, W. P. Menzel, L. E. Gumley, and Y. Zhang, "Retrieval of geophysical parameters from moderate resolution imaging spectroradiometer thermal infrared data: evaluation of a two-step physical algorithm," *Applied Optics*, vol. 39, no. 20, pp. 3537–3550, 2000.

[14] O. Rozenstein, Z. Qin, Y. Derimian, and A. Karnieli, "Derivation of land surface temperature for landsat-8 tirs using a split window algorithm," *Sensors*, vol. 14, no. 4, pp. 5768–5780, 2014.

[15] A. Chedin, N. Scott, and A. Berroir, "A single-channel, double-viewing angle method for sea surface temperature determination from coincident meteosat and tiros-n radiometric measurements," *Journal of Applied Meteorology*, vol. 21, no. 4, pp. 613–618, 1982.

[16] G. Masiello, C. Serio, I. De Feis, M. Amoroso, S. Venafra, I. Trigo, and P. Watts, "Kalman filter physical retrieval of surface emissivity and temperature from geostationary infrared radiances," *Atmospheric Measurement Techniques*, vol. 6, no. 12, pp. 3613–3634, 2013.

[17] M. T. Hagan, H. B. Demuth, M. H. Beale, and O. De Jesús, *Neural network design*. Pws Pub. Boston, 1996, vol. 20.

[18] P. Du, J. Xia, W. Zhang, K. Tan, Y. Liu, and S. Liu, "Multiple classifier system for remote sensing image classification: A review," *Sensors*, vol. 12, no. 4, pp. 4764–4792, 2012.

[19] X. X. Zhu, D. Tuia, L. Mou, G.-S. Xia, L. Zhang, F. Xu, and F. Fraundorfer, "Deep learning in remote sensing: A comprehensive review and list of resources," *IEEE Geoscience and Remote Sensing Magazine*, vol. 5, no. 4, pp. 8–36, 2017.

[20] W. Hu, Y. Huang, L. Wei, F. Zhang, and H. Li, "Deep convolutional neural networks for hyperspectral image classification," *Journal of Sensors*, vol. 2015, 2015.

[21] N. Kussul, M. Lavreniuk, S. Skakun, and A. Shelestov, "Deep learning classification of land cover and crop types using remote sensing data," *IEEE Geoscience and Remote Sensing Letters*, vol. 14, no. 5, pp. 778–782, 2017.

[22] W. Li, G. Wu, and Q. Du, "Transferred deep learning for anomaly detection in hyperspectral imagery," *IEEE Geoscience and Remote Sensing Letters*, vol. 14, no. 5, pp. 597–601, 2017.

[23] F. Aires, C. Prigent, W. Rossow, and M. Rothstein, "A new neural network approach including first guess for retrieval of atmospheric water vapor, cloud liquid water path, surface temperature, and emissivities over land from satellite microwave observations," *Journal of Geophysical Research: Atmospheres*, vol. 106, no. D14, pp. 14 887–14 907, 2001.

[24] A. Wu, W. W. Hsieh, and B. Tang, "Neural network forecasts of the tropical pacific sea surface temperatures," *Neural Networks*, vol. 19, no. 2, pp. 145–154, 2006.

[25] K. Mao, J. Shi, H. Tang, Z.-L. Li, X. Wang, and K.-S. Chen, "A neural network technique for separating land surface emissivity and temperature from aster imagery," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 46, no. 1, pp. 200–208, 2008.

[26] J. Cheng, Q. Xiao, X. Li, Q. Liu, Y. Du, and A. Nie, "Multi-layer perceptron neural network based algorithm for simultaneous retrieving temperature and emissivity from hyperspectral ftir dataset," in *Geoscience and Remote Sensing Symposium, 2007. IGARSS 2007. IEEE International*. Ieee, 2007, pp. 4383–4385.

[27] C. Gao, Z.-L. Li, S. Qiu, B.-H. Tang, H. Wu, X. Jiang *et al.*, "An improved algorithm for retrieving land surface emissivity and temperature from msg-2/seviri data." *IEEE Trans. Geoscience and Remote Sensing*, vol. 52, no. 6, pp. 3175–3191, 2014.

[28] M. Pivovarník, S. J. S. Khalsa, J. C. Jiménez-Muñoz, and F. Zemek, "Improved temperature and emissivity separation algorithm for multispectral and hyperspectral sensors," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 55, no. 4, pp. 1944–1953, 2017.

[29] S. Erard and W. Calvin, "New composite spectra of mars, 0.4–5.7 μm," *Icarus*, vol. 130, no. 2, pp. 449–460, 1997.

[30] R. N. Clark, C. M. Pieters, R. O. Green, J. Boardman, and N. E. Petro, "Thermal removal from near-infrared imaging spectroscopy data of the Moon," *Journal of Geophysical Research: Planets*, vol. 116, no. E6, 2011.

[31] Z. Shao-bin, W. Xu-nan, and B. Yun-fei, "The study of the land surface temperature retrieval with the proposed neural network model," in *2008 International Conference on Neural Networks and Signal Processing*, 2008.

[32] Y. Bai, M. Wong, W.-Z. Shi, L.-X. Wu, and K. Qin, "Advancing of land surface temperature retrieval using extreme learning machine and spatio-temporal adaptive data fusion algorithm," *Remote sensing*, vol. 7, no. 4, pp. 4424–4441, 2015.

[33] S.-B. Duan, Z.-L. Li, and P. Leng, "A framework for the retrieval of all-weather land surface temperature at a high spatial resolution from polar-orbiting thermal infrared and passive microwave data," *Remote sensing of environment*, vol. 195, pp. 107–117, 2017.

[34] J. P. Snyder, *Map projections–A working manual*. US Government Printing Office, 1987, vol. 1395.

[35] G. Vivone, L. Alparone, J. Chanussot, M. Dalla Mura, A. Garzelli, G. A. Licciardi, R. Restaino, and L. Wald, "A critical comparison among pansharpening algorithms," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 53, no. 5, pp. 2565–2586, 2015.

[36] V. Shettigara, "A generalized component substitution technique for spatial enhancement of multispectral images using a higher resolution data set," *Photogrammetric Engineering and remote sensing*, vol. 58, no. 5, pp. 561–567, 1992.

[37] A. E. Iverson and J. R. Lersch, "Adaptive image sharpening using multiresolution representations," in *SPIE's International Symposium on Optical Engineering and Photonics in Aerospace Sensing*. International Society for Optics and Photonics, 1994, pp. 72–83.

[38] J. C. Price, "Combining panchromatic and multispectral imagery from dual resolution satellite instruments," *Remote sensing of environment*, vol. 21, no. 2, pp. 119–128, 1987.

[39] C. K. Munechika, J. S. Warnick, C. Salvaggio, and J. R. Schott, "Resolution enhancement of multispectral image data to improve classification accuracy," *Photogrammetric Engineering and Remote Sensing*, vol. 59, no. 1, pp. 67–72, 1993.

[40] R. Nishii, S. Kusanobu, and S. Tanaka, "Enhancement of low spatial resolution image based on high resolution-bands," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 34, no. 5, pp. 1151–1158, 1996.

[41] R. C. Hardie, M. T. Eismann, and G. L. Wilson, "Map estimation for hyperspectral image resolution enhancement using an auxiliary sensor," *IEEE Transactions on Image Processing*, vol. 13, no. 9, pp. 1174–1184, 2004.

[42] Y. Zhang, S. De Backer, and P. Scheunders, "Noise-resistant wavelet-based Bayesian fusion of multispectral and hyperspectral images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 47, no. 11, pp. 3834–3843, 2009.

[43] Y. Zhang, A. Duijster, and P. Scheunders, "A Bayesian restoration approach for hyperspectral images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 50, no. 9, pp. 3453–3462, 2012.

[44] Y. Yuan, X. Zheng, and X. Lu, "Hyperspectral image superresolution by transfer learning," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 10, no. 5, pp. 1963–1974, 2017.

[45] J. Goswami and M. Annadurai, "Chandrayaan-1: India's first planetary science mission to the moon," *Current Science*, pp. 486–491, 2009.

[46] J.-P. Bibring, A. Soufflot, M. Berthé, Y. Langevin, B. Gondet, P. Drossart, M. Bouyé, M. Combes, P. Puget, A. Semery *et al.*, "Omega: Observatoire pour la minéralogie, l'eau, les glaces et l'activité," in *Mars Express: the scientific payload*, vol. 1240, 2004, pp. 37–49.

[47] J.-P. Bibring, Y. Langevin, J. F. Mustard, F. Poulet, R. Arvidson, A. Gendrin, B. Gondet, N. Mangold, P. Pinet, F. Forget *et al.*, "Global mineralogical and aqueous mars history derived from omega/mars express data," *science*, vol. 312, no. 5772, pp. 400–404, 2006.

[48] R. O. Green, M. L. Eastwood, C. M. Sarture, T. G. Chrien, M. Aronsson, B. J. Chippendale, J. A. Faust, B. E. Pavri, C. J. Chovit, M. Solis *et al.*, "Imaging spectroscopy and the airborne visible/infrared imaging spectrometer (aviris)," *Remote sensing of environment*, vol. 65, no. 3, pp. 227–248, 1998.

[49] M. Parente, H. D. Makarewicz, and J. L. Bishop, "Decomposition of mineral absorption bands using nonlinear least squares curve fitting: Application to martian meteorites and crism data," *Planetary and Space Science*, vol. 59, no. 5, pp. 423–442, 2011.

[50] K. Schmidt, A. Skidmore, E. Kloosterman, H. Van Oosten, L. Kumar, and J. Janssen, "Mapping coastal vegetation using an expert system and hyperspectral imagery," *Photogrammetric Engineering & Remote Sensing*, vol. 70, no. 6, pp. 703–715, 2004.

[51] N. Acito, M. Diani, and G. Corsini, "Signal-dependent noise modeling and model parameter estimation in hyperspectral images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 49, no. 8, pp. 2957–2971, 2011.

[52] G. Chen and S.-E. Qian, "Denoising of hyperspectral imagery using principal component analysis and wavelet shrinkage," *IEEE Transactions on Geoscience and remote sensing*, vol. 49, no. 3, pp. 973–980, 2011.

[53] Q. Yuan, L. Zhang, and H. Shen, "Hyperspectral image denoising employing a spectral–spatial adaptive total variation model," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 50, no. 10, pp. 3660–3677, 2012.

[54] Y.-Q. Zhao and J. Yang, "Hyperspectral image denoising via sparse representation and low-rank constraint," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 53, no. 1, pp. 296–308, 2015.

[55] X. Lu, Y. Yuan, and P. Yan, "Sparse coding for image denoising using spike and slab prior," *Neurocomputing*, vol. 106, pp. 12–20, 2013.

[56] X. Ceamanos and S. Douté, "Spectral smile correction of crism/mro hyperspectral images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 48, no. 11, pp. 3951–3959, 2010.

[57] F. J. Massey Jr, "The kolmogorov-smirnov test for goodness of fit," *Journal of the American statistical Association*, vol. 46, no. 253, pp. 68–78, 1951.

[58] K. Pearson, "X. on the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling," *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, vol. 50, no. 302, pp. 157–175, 1900.

[59] A. R. De Pierro, "A modified expectation maximization algorithm for penalized likelihood estimation in emission tomography," *IEEE transactions on medical imaging*, vol. 14, no. 1, pp. 132–137, 1995.

[60] Y. Yuan, X. Zheng, and X. Lu, "Spectral–spatial kernel regularized for hyperspectral image denoising," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 53, no. 7, pp. 3815–3832, 2015.

[61] X. Lu, Y. Wang, and Y. Yuan, "Graph-regularized low-rank representation for destriping of hyperspectral images," *IEEE transactions on geoscience and remote sensing*, vol. 51, no. 7, pp. 4009–4018, 2013.

[62] K.-I. Funahashi, "On the approximate realization of continuous mappings by neural networks," *Neural networks*, vol. 2, no. 3, pp. 183–192, 1989.

[63] K. Hornik, M. Stinchcombe, and H. White, "Multilayer feedforward networks are universal approximators," *Neural networks*, vol. 2, no. 5, pp. 359–366, 1989.

[64] T. Condus, R. Arvidson, L. He, J. O'sullivan, M. Wolff, and R. Morris, "A neural network approach to retrieve single scattering albedos from tes data," in *50th Lunar and Planetary Science Conference*, 2019.

[65] C. M. Bishop, *Pattern recognition and machine learning*. springer, 2006.

[66] B. L. Ehlmann and C. S. Edwards, "Mineralogy of the martian surface," *Annual Review of Earth and Planetary Sciences*, vol. 42, pp. 291–315, 2014.

[67] A. Baldridge, S. Hook, C. Grove, and G. Rivera, "The aster spectral library version 2.0," *Remote Sensing of Environment*, vol. 113, no. 4, pp. 711–715, 2009.

[68] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.

[69] G. E. Hinton and R. R. Salakhutdinov, "Reducing the dimensionality of data with neural networks," *science*, vol. 313, no. 5786, pp. 504–507, 2006.

[70] D. P. Kingma and M. Welling, "Auto-encoding variational bayes," *arXiv preprint arXiv:1312.6114*, 2013.

[71] A. Savitzky and M. J. Golay, "Smoothing and differentiation of data by simplified least squares procedures." *Analytical chemistry*, vol. 36, no. 8, pp. 1627–1639, 1964.

[72] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial nets," in *Advances in neural information processing systems*, 2014, pp. 2672–2680.

[73] A. R. Vasavada, S. Piqueux, K. W. Lewis, M. T. Lemmon, and M. D. Smith, "Thermophysical properties along curiosity's traverse in gale crater, mars, derived from the rems ground temperature sensor," *Icarus*, vol. 284, pp. 372–386, 2017.

[74] P. Harremoës and G. Tusnády, "Information divergence is more χ 2-distributed than the χ 2-statistics," in *Information Theory Proceedings (ISIT), 2012 IEEE International Symposium on*. IEEE, 2012, pp. 533–537.

[75] S. Kullback and R. A. Leibler, "On information and sufficiency," *The annals of mathematical statistics*, vol. 22, no. 1, pp. 79–86, 1951.

[76] L. Paninski, "Estimation of entropy and mutual information," *Neural computation*, vol. 15, no. 6, pp. 1191–1253, 2003.

[77] S. Degirmenci, *A General Framework of Large-Scale Convex Optimization Using Jensen Surrogates and Acceleration Techniques*. Washington University in St. Louis, 2016.

[78] L. He, J. O'Sullivan, D. Politte, and R. Arvidson, "Resolution analysis of regularized maximum log-likelihood reconstruction method for crism hyperspectral data," in *Propagation Through and Characterization of Atmospheric and Oceanic Phenomena*. Optical Society of America, 2017, pp. JTu5A–16.

[79] K. Powell, R. Arvidson, M. Zanetti, E. Guinness, and S. Murchie, "The structural, stratigraphic, and paleoenvironmental record exposed on the rim and walls of iazu crater, mars," *Journal of Geophysical Research: Planets*, 2017.

[80] R. Arvidson, S. Squyres, J. Bell, J. Catalano, B. Clark, L. Crumpler, P. De Souza, A. Fairén, W. Farrand, V. Fox *et al.*, "Ancient aqueous environments at endeavour crater, mars," *Science*, vol. 343, no. 6169, p. 1248097, 2014.

[81] E. M. Eliason and A. S. McEwen, "Adaptive box filters for removal of random noise from digital images," *Photogrammetric Engineering and Remote Sensing*, vol. 56, no. 4, pp. 453–458, 1990.

[82] J. Carter, F. Poulet, J.-P. Bibring, N. Mangold, and S. Murchie, "Hydrous minerals on mars as seen by the crism and omega imaging spectrometers: Updated global view," *Journal of Geophysical Research: Planets*, vol. 118, no. 4, pp. 831–858, 2013.

[83] N. K. McKeown, J. L. Bishop, E. Z. Noe Dobrea, B. L. Ehlmann, M. Parente, J. F. Mustard, S. L. Murchie, G. A. Swayze, J.-P. Bibring, and E. A. Silver, "Characterization of phyllosilicates observed in the central mawrth vallis region, mars, their potential formational processes, and implications for past climate," *Journal of Geophysical Research: Planets*, vol. 114, no. E2, 2009.

[84] J. R. Michalski, J.-P. Bibring, F. Poulet, D. Loizeau, N. Mangold, E. N. Dobrea, J. L. Bishop, J. J. Wray, N. K. McKeown, M. Parente *et al.*, "The mawrth vallis region of mars: A potential landing site for the mars science laboratory (msl) mission," *Astrobiology*, vol. 10, no. 7, pp. 687–703, 2010.

[85] J. Wray, B. Ehlmann, S. Squyres, J. Mustard, and R. Kirk, "Compositional stratigraphy of clay-bearing layered deposits at mawrth vallis, mars," *Geophysical Research Letters*, vol. 35, no. 12, 2008.

[86] J. L. Bishop, E. Z. N. Dobrea, N. K. McKeown, M. Parente, B. L. Ehlmann, J. R. Michalski, R. E. Milliken, F. Poulet, G. A. Swayze, J. F. Mustard *et al.*, "Phyllosilicate diversity and past aqueous activity revealed at mawrth vallis, mars," *Science*, vol. 321, no. 5890, pp. 830–833, 2008.

[87] D. W. Mittlefehldt, R. Gellert, D. W. Ming, A. S. Yen, B. C. Clark, R. V. Morris, C. Schröder, L. S. Crumpler, J. A. Grant, B. L. Jolliff *et al.*, "Diverse lithologies and alteration events on the rim of noachian-aged endeavour crater, meridiani planum, mars: In situ compositional evidence," *Journal of Geophysical Research: Planets*, vol. 123, no. 5, pp. 1255–1306, 2018.

[88] E. Rampe, M. Lapotre, T. Bristow, R. Arvidson, R. Morris, C. Achilles, C. Weitz, D. Blake, D. Ming, S. Morrison *et al.*, "Sand mineralogy within the bagnold dunes, gale crater, as observed in situ and from orbit," *Geophysical Research Letters*, vol. 45, no. 18, pp. 9488–9497, 2018.

[89] A. Fraeman, B. Ehlmann, R. Arvidson, C. Edwards, J. Grotzinger, R. Milliken, D. Quinn, and M. Rice, "The stratigraphy and evolution of lower mount sharp from spectral, morphological, and thermophysical orbital data sets," *Journal of Geophysical Research: Planets*, vol. 121, no. 9, pp. 1713–1736, 2016.

[90] Y. Liu, T. D. Glotch, N. A. Scudder, M. L. Kraner, T. Condus, R. E. Arvidson, E. A. Guinness, M. J. Wolff, and M. D. Smith, "End-member identification and spectral

mixture analysis of crism hyperspectral data: A case study on southwest melas chasma, mars," *Journal of Geophysical Research: Planets*, vol. 121, no. 10, pp. 2004–2036, 2016.

[91] D. M. Hassler, C. Zeitlin, R. F. Wimmer-Schweingruber, B. Ehresmann, S. Rafkin, J. L. Eigenbrode, D. E. Brinza, G. Weigle, S. Böttcher, E. Böhm *et al.*, "Mars' surface radiation environment measured with the mars science laboratory's curiosity rover," *science*, vol. 343, no. 6169, p. 1244797, 2014.

[92] K. E. Powell, "Spectral and stratigraphic mapping of layered sulfate deposits on mars using advanced crism data processing techniques," Ph.D. dissertation, Washington University in St. Louis, 2018.

[93] Y. Kaganovsky, S. Han, S. Degirmenci, D. G. Politte, D. J. Brady, J. A. O'Sullivan, and L. Carin, "Alternating minimization algorithm with automatic relevance determination for transmission tomography under poisson noise," *SIAM Journal on Imaging Sciences*, vol. 8, no. 3, pp. 2087–2132, 2015.

Vita

Linyun He

Degrees

Ph.D. Electrical Engineering, Washington University in St. Louis, Missouri, U.S., May 2019

M.S. Engineering Data Analytics and Statistics, Washington University in St. Louis, Missouri, U.S., December 2018

B.S. Electrical Engineering, University of Electronic Science and Technology of China, Chengdu, China, July 2014

Publications

Journal Publications:

L. He, J. A. O'Sullivan, D. V. Politte, K. E. Powell, and R. E. Arvidson, "Quantitative reconstruction and denoising method HyBER for hyperspectral image data and its application to CRISM," to appear in *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 2019.

L. He, R. E. Arvidson, J. A. O'Sullivan, D. V. Politte, T. Condus, and A. R. Vasavada, "Retrieving temperature and single scattering albedos from hyperspectral data using neural networks," preparing for *Icarus*, 2019.

C. D. Kreisch, J. A. O'Sullivan, R. E. Arvidson, D. V. Politte, **L. He**, N. T. Stein, J. Finkel, E. A. Guinness, M. J. Wolff, and M. G. A. Lapotre, "Regularization of Mars Reconnaissance Orbiter CRISM along-track oversampled hyperspectral imaging observations of Mars," *Icarus*, vol. 282, pp. 44-56, 2017.

Conference Presentations:

L. He, R. E. Arvidson, D. V. Politte, T. Condus, J. A. O'Sullivan, and A. R. Vasavada, "Retrieving temperature and single scattering albedos from hyperspectral data using neural networks," presented at *50th Lunar and Planetary Science Conference*, Houston, TX, 2019.

T. Condus, R. E. Arvidson, **L. He**, J. A. O'Sullivan, M. J. Wolff and R. V. Morris, "A neural network approach to retrieve single scattering albedos and temperatures from THEMIS and TES infrared data over Gale Crater," presented at *50th Lunar and Planetary Science Conference*, Houston, TX, 2019.

D. V. Politte, R. E. Arvidson, J. A. O'Sullivan, and **L. He**, "Pipeline for retrieval of surface temperatures and single scattering albedos," presented at *50th Lunar and Planetary Science Conference*, Houston, TX, 2019.

L. He, R. E. Arvidson, and J. A. O'Sullivan, "Retrieving single scattering albedos and temperatures from CRISM hyperspectral data using neural networks," presented at *Planetary Science Informatics and Data Analytics Conference*, St. Louis, MO, 2018.

K. E. Powell, R. E. Arvidson, **L. He**, D. V. Politte, J. A. O'Sullivan, S. L. Murchie, and R. V. Morris, "Retrieval of CRISM single scattering albedos from 1-3.8 micrometers over the Curiosity Rover Traverse," presented at *49th Lunar and Planetary Science Conference*, Houston, TX, 2018.

D. V. Politte, R. E. Arvidson, J. A. O'Sullivan, **L. He**, K. E. Powell, and E. A. Guinness "End-to-end processing of CRISM Along-Track Oversampled Observations with atmosphere and temperature corrections," presented at *49th Lunar and Planetary Science Conference*, Houston, TX, 2018.

L. He, J. A. O'Sullivan, D. V. Politte, and R. E. Arvidson , "Resolution analysis of regularized maximum log-likelihood reconstruction method for CRISM hyperspectral data," presented at *Propagation Through and Characterization of Atmospheric and Oceanic Phenomena, Optical Society of America*, San Francisco, CA, 2017.

L. He, R. E. Arvidson, J. A. O'Sullivan, and D. V. Politte, "Regularization in maximum log-likelihood method for CRISM hyperspectral image cube," presented at *48th Lunar and Planetary Science Conference*, Houston, TX, 2017.

L. He, R. E. Arvidson, J. A. O'Sullivan, and D. V. Politte, "Performance analysis of regularized MLM hyperspectral imaging for Mars Reconnaissance Orbiter CRISM Along-Track Oversampled data," presented at *The 43rd Annual North American Meeting of the Federation of Analytical Chemistry and Spectroscopy Societies*, Minneapolis, MN, 2016.

Awards

FACSS First Price Student Poster Award at *The 43rd Annual North American Meeting of the Federation of Analytical Chemistry and Spectroscopy Societies*, Minneapolis, MN, 2016

May 2019