Inel 6007
Introduction to Remote Sensing
Chapter 4
Data Models

Prof. Vidya Manian
Dept. of Electrical and Computer Engineering
Introduction

• What Factors Affect Remote Sensing Imagery?
  – atmospheric conditions
  – characteristics of the earth’s surface
  – sensor characteristics

• These factors can be modeled for their effects on remote sensing data

• Many image processing algorithms, in turn, rely on these data models
Data Models

• The simplest data model is to model it as random
  – Random variables
  – Random Fields (Random processes in 2D)

• Why?
  – Natural way to deal with variability
  – Too many factors cause changes in intensities (scalar or vectorial) in an image
Data Models

• Univariate Data Statistics
• Multivariate Data Statistics
• Scattergrams and Scatterplots
• Noise Models
• Scene and Sensor Effects
Univariate Data Statistics

• Each pixel in a single band is modeled as sample of the same random variable
• Band-by-band statistics of a multispectral image
• Do not measure interrelationships among bands
• Can, in rare cases, be applied to the multivariate case
  – Bands are statistically independent

• Definition:
  – Digital Number (DN) – numerical value of a pixel, either a scalar (univariate, single band) or vector (multivariate, multiband)
DN Histogram

• Measures brightness distribution
  – Define equally sized DN “bins”,
  – Count the number of pixels in each bin and divide by the total number of pixels in the image

\[
\text{hist}_{DN} = \frac{\text{count}(DN)}{N}
\]

• Properties
  – Analogous to the continuous Probability Density Function (PDF) of statistics

\[
\text{hist}_{DN} \approx \text{PDF}(DN)
\]
  – Histogram contains no information about the spatial distribution of pixels
  – Usually skewed, with a “tail” towards higher DNs

• Applications
  – contrast enhancement
  – DN thresholding
Histograms

- A plot of the number of pixels at each signal value within each band (wavelength).
- Usually it is convenient to make 256 bins to create the histogram.
Histograms for Añasco Landsat TM Simulator Image Band 1
The Histogram Explained

The actual data is only spread over part of the total possible range.

Fraction of Total Pixels on this axis
DN Moments

- **DN mean**
  \[
  \mu = \frac{1}{N} \sum_{p=1}^{N} DN_p = \sum_{DN=DN_{min}}^{DN_{max}} DN \times hist_{DN}
  \]
  - \( \mu \) is a measure of average **brightness**
  - first moment (centroid) of the histogram

- **DN variance**
  \[
  \sigma^2 = \frac{1}{N-1} \sum_{p=1}^{N} (DN_p - \mu)^2 = \frac{N}{N-1} \sum_{DN=DN_{min}}^{DN_{max}} (DN - \mu)^2 \times hist_{DN}
  \]
  - Standard deviation \( \sigma \) is a measure of average **contrast**
  - second central moment of the histogram

- Efficient to calculate if histogram is already available
More Histogram Features

• Higher order moments
  – skewness measures asymmetry of the histogram
  – kurtosis measures sharpness of the histogram peak relative to a normal distribution

• Other statistics
  – mode has the maximum count
  – median divides histogram area in half

Example image histogram compared to equivalent Gaussian distribution

\[ N(f; \mu, \sigma) = \frac{1}{\sigma \sqrt{2\pi}} e^{- \frac{(f-\mu)^2}{2\sigma^2}} \]
More on moments:
Skewness and Kurtosis.

Negative Skew
Elongated tail at the left
More data in the left tail than would be expected in a normal distribution

Positive Skew
Elongated tail at the right
More data in the right tail than would be expected in a normal distribution

\[
skewness = \frac{1}{N} \sum_{p=1}^{N} \left( \frac{DN_p - \mu}{\sigma} \right)^3 = \sum_{DN=DN_{\text{min}}}^{DN_{\text{max}}} \left( \frac{DN - \mu}{\sigma} \right)^3 \times \text{hist}_{DN}
\]

Measures asymmetry
Zero for a Symmetric Histogram
Positive for histograms with long tails toward positive numbers
More on moments: Skewness and Kurtosis.

Sharpness of the peak relative to the gaussian distribution
Is zero for normal distribution
Positive for sharper than gaussian peak
Negative for less sharper
Histograms for Añasco Landsat TM Simulator Image Band 1

\[ \mu = 112.8370 \]
\[ \sigma = 18.6028 \]
\[ \text{kurtosis} = 4.6097 \]
\[ \text{skewness} = 0.8146 \]
Types of Distributions

- Normal Distribution
- Multimodal Distribution
- Negatively Skewed Distribution
- Positively Skewed Distribution
- Uniform Distribution
Gaussian pdf

• A R.V. X that is normally distributed has density function:

\[ f_X(x) = \frac{1}{\sqrt{2\pi\sigma_X^2}} \exp\left\{ -\frac{(x - \eta_X)^2}{2\sigma_X^2} \right\} \]
Uniform pdf

- A R.V. $X$ that is uniformly distributed between $x_1$ and $x_2$ has density function:

$$f_X(x) = \begin{cases} 
\frac{1}{b-a} & \text{for } a \leq x \leq b \\
0 & \text{otherwise}
\end{cases}$$
Multimodal Distribution
Negatively Skewed Distribution

![Graph showing a negatively skewed distribution with mean, median, and mode indicated.]

INEL6007(Spring 2010)  ECE, UPRM  19
Positively Skewed Distribution
Cumulative Histogram

- Cumulative fraction of pixels with value less than or equal to DN
- Properties
  - Analogous to the Cumulative Distribution Function (CDF) of statistics
    - Integral of the pdf
    - Monotonic function of DN
- Applications
  - contrast enhancement (histogram equalization)
  - relative radiometric normalization of multitemporal images
  - noise removal (destriping)
Histograms for Añasco Landsat TM Simulator Image Band 1

Median 112.8
LANDSAT TM Simulator, Añasco/Mayaguez, PR

Band 1

Band 2

Band 3

Band 4
LANDSAT TM Simulator, Añasco/Mayaguez, PR

Band 5

Band 6

Band 7

INEL6007(Spring 2010) ECE, UPRM
Histograms for Añasco Landsat TM Simulator Image Band 5

\[ \mu = 111.1417 \quad \text{kurtosis} = 2.3416 \]

\[ \sigma = 47.1913 \quad \text{skewness} = -0.7048 \]
Data Models

• Univariate Data Statistics
• Multivariate Data Statistics
  – Brief Review of Probabilities
  – Back to Multispectral
• Scattergrams and Scatterplots
• Noise Models
• Scene and Sensor Effects
Two Random Variables

- Defined as real value functions, $X(\zeta)$ and $Y(\zeta)$, over the same sample space $S$.

Figure 2.15  Random variables $X$ and $Y$ defined on the sample space $S$. 
Total Characterization: Joint CDF

- The joint cumulative distribution function (JCDF) for a random variables $X$ and $Y$ is

$$F_{XY}(x, y) = P\{X \leq x, Y \leq y\}$$
Properties: Joint CDF

• Note that \( F_{XY}(x,y) \) is **bounded** from the above and below

\[
0 \leq F_{XY}(x,y) \leq 1
\]

• Note that \( F_{XY}(x,y) \) is **non-decreasing** in \( x \) and \( y \), i.e.

\[
F_{XY}(x_2, y) \geq F_{XY}(x_1, y) \quad \text{for all } x_2 > x_1, \text{ and all } y \\
F_{XY}(x, y_2) \geq F_{XY}(x, y_1) \quad \text{for all } y_2 > y_1, \text{ and all } x
\]
Properties: Joint CDF

• $F_{XY}(x,y)$ is **continuous** from the right

$$\lim_{\varepsilon \to 0^+} F_{XY}(x + \varepsilon, y + \delta) = F_{XY}(x, y)$$

$$\lim_{\delta \to 0^+} F_{XY}(x, y + \delta) = F_{XY}(x, y)$$

• $F_{XY}(x,y)$ can be used to **calculate** probabilities

$$P\{x_1 < X \leq x_2, y_1 < Y \leq y_2\} = F_X(x_2, y_2) - F_X(x_1, y_1)$$
Total Characterization: Joint pdf

• $F_{XY}(x,y)$ is related to the joint probability density function $f_{XY}(x,y)$ by

$$f_{XY}(x, y) = \frac{\partial^2 F_{XY}(x, y)}{\partial x \partial y}$$

$$F_{XY}(x, y) = \int_{-\infty}^{x} \int_{-\infty}^{y} f_{XY}(u, v) \, du \, dv$$
Properties of the Joint pdf

(1) Positivity,

\[ f_{XY}(x, y) \geq 0 \quad \text{for all } x \text{ and } y \]  
\[ (2.83) \]

(2) Integral over all \( x \) and \( y \),

\[ \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{XY}(x, y) \, dx \, dy = 1 \]
\[ (2.84) \]

(3) \( f_{XY}(x, y) \) can be used to calculate probability of rectangular events as

\[ P[x_1 < X \leq x_2, y_1 < Y \leq y_2] = \int_{x_1^+}^{x_2^+} \int_{y_1^+}^{y_2^+} f_{XY}(x, y) \, dx \, dy \]
\[ (2.85) \]

where \( x_1^+, x_2^+, y_1^+ \) and \( y_2^+ \) are limits from the positive sides, or any event \( A \) as

\[ P((X, Y) \in A) = \int_{A} f_{XY}(x, y) \, dx \, dy \]
\[ (2.86) \]
Partial Characterization

• Marginal densities
• Conditional densities
• Joint moments
  – Conditional means
  – Correlation, covariance
  – Higher order moments
Marginal Densities

\[ f_X(x) = \int_{-\infty}^{\infty} f_{XY}(x, y) \, dy \]

\[ f_Y(y) = \int_{-\infty}^{\infty} f_{XY}(x, y) \, dx \]
Conditional pdf

\[ f_X(x/y) = \frac{f_{XY}(x, y)}{f_Y(y)} = \frac{f_{XY}(x, y)}{\int_{-\infty}^{\infty} f_{XY}(x, y) dx} \]

\[ f_Y(y/x) = \frac{f_{XY}(x, y)}{f_X(x)} = \frac{f_{XY}(x, y)}{\int_{-\infty}^{\infty} f_{XY}(x, y) dy} \]

\[ f_{XY}(x, y) = f_Y(y/x)f_X(x) = f_X(x/y)f_Y(y) \]
Statistical Independence

\[ f_{X}(x/y) = \frac{f_{XY}(x, y)}{f_{Y}(y)} = f_{X}(x) \]

\[ f_{Y}(y/x) = \frac{f_{XY}(x, y)}{f_{X}(x)} = f_{Y}(y) \]

\[ f_{XY}(x, y) = f_{Y}(y)f_{X}(x) \]
Show that if $X$ and $Y$ are independent then

$$E(X/y) = E(X)$$

$$E(Y/x) = E(Y)$$
Correlation $R_{XY}$

- The correlation between two random variables $X$ and $Y$ is defined as
  $$R_{XY} = E[XY]$$

- Two random variables are defined as uncorrelated if
  $$R_{XY} = E[X] E[Y]$$

- Two random variables are called orthogonal if
  $$R_{XY} = E[XY] = 0$$
Covariance $\sigma_{XY}$

- The covariance between two random variables $X$ and $Y$ is defined as
  $$\sigma_{XY} = E[(X - \eta_X)(Y - \eta_Y)] = R_{XY} - \eta_X \eta_Y$$

- If $X$ and $Y$ are uncorrelated then
  $$\sigma_{XY} = R_{XY} - \eta_X \eta_Y = 0$$
Correlation Coefficient $\rho_{XY}$

- The correlation coefficient $\rho_{XY}$ between the two random variables $X$ and $Y$ is defined as

$$
\rho_{XY} = \frac{\sigma_{XY}}{\sigma_X \sigma_Y}
$$

- It provides a measure of linear relationship between two RVs $X$ and $Y$
Correlation Coefficient $\rho_{XY}$ (cont.)

\[-1 \leq \rho_{XY} \leq +1\]

- The closer $\rho_{XY}$ is to -1 or +1 the more the random variables X and Y are said to be linearly related.
Correlation Between 2 Variables

\[ \rho_{XY} \approx 1 \]

\[ 0 \leq \rho_{XY} < 1 \]

\[ \rho_{XY} \approx 0, \quad \sigma_X^2 = \sigma_Y^2 \]

\[ \rho_{XY} \approx -1 \]

\[ -1 \leq \rho_{XY} \leq 0 \]

\[ \rho_{XY} \approx 0, \quad \sigma_X^2 > \sigma_Y^2 \]
Correlation Coefficient $\rho_{XY}$

- Useful relations

$$\rho_{XY} = \frac{\sigma_{XY}}{\sigma_X \sigma_Y}$$

$$\sigma_{XY} = \rho_{XY} \sigma_X \sigma_Y$$

$$r_{XY} = \sqrt{1 - \rho_{XY}^2}$$
Correlation, $\rho_{XY}$ is a Measure of Linear Relation
CDF and PDF More than 2 RV

CDF

\[ F_{X_1, X_2 \ldots X_n}(x_1, x_2, \ldots, x_n) = P\{X_1 \leq x_1, X_2 \leq x_2, \ldots, X_n \leq x_n \} \]

pdf

\[ f_{X_1, X_2 \ldots X_n}(x_1, x_2, \ldots, x_n) = \frac{\partial^n F_{WZ}(x_1, x_2, \ldots, x_n)}{\partial x_1 \partial x_2, \ldots \partial x_n} \]

\[ F_{X_1, X_2 \ldots X_n}(x_1, x_2, \ldots, x_n) = \int_{-\infty}^{x_1} \int_{-\infty}^{x_2} \ldots \int_{-\infty}^{x_n} f_{X_1, X_2 \ldots X_n}(x_1, x_2, \ldots, x_n) \, dx_1 dx_2 \ldots dx_n \]
More than 2 RV: Vector Notation

CDF  \[ F_X(x) = P\{X \leq x\} \]

pdf  \[ f_X(x) = \frac{\partial^n F_X(x_1, x_2, \ldots, x_n)}{\partial x_1 \partial x_2, \ldots \partial x_n} \]

\[ F_X(x) = \int_{-\infty}^{x} f_X(x) \, dx \]

where  \[ x = \begin{bmatrix} x_1 & x_2 & \cdots & x_n \end{bmatrix}^T \]
Multidimensional Gaussian pdf

\[
f_x(x) = \frac{1}{(2\pi)^{n/2}|K_x|^{1/2}} \exp \left[ -\frac{1}{2}(x - \eta_x)^T K_x^{-1} (x - \eta_x) \right]
\]

\[
f_x(x) \sim N(\eta_x, C_x)
\]

\[
\eta_x = E[X] = [E[X_1] \ E[X_2] \ \ldots \ E[X_n]]^T
\]

\[
C_x = E[(X - \eta_x)(X - \eta_x)^T]
\]
Multidimensional Gaussian Density
Partial Characterization

Mean

$$\eta_x = E[X] = [E[X_1] \ E[X_2] \ \cdots \ E[X_n]]^T$$

Autocorrelation

$$R_x = E[XX^T], \ [R_x]_{ij} = E[X_iX_j]$$

Covariance Matrix

$$C_x = E[(X - \eta_x)(X - \eta_x)^T] = R_x - \eta_x\eta_x^T$$

$$c_{ij} = [C_x]_{ij} = \sigma_{X_iX_j}, \ c_{ii} = [C_x]_{ii} = \sigma_{X_i}^2$$
Correlation Matrix, $R$
(sorry for the abuse of notation)

$$\rho_{X_iX_j} = [R]_{ij} = \frac{\sigma_{X_iX_j}}{\sqrt{\sigma_{X_i}^2 \sigma_{X_j}^2}}$$
Multivariate Data Statistics

- K-band multispectral image
- Each pixel is modeled as a sample of a random vector
- Measures statistical relationships among bands
Vector Representation

• Each multispectral pixel is a K-dimensional column vector
  – Components of the vector are the DN values in each of the K bands
  – At pixel (i,j):

\[
DN_{ij} = \begin{bmatrix} DN_{ij_1} & DN_{ij_2} & \ldots & DN_{ij_K} \end{bmatrix}^T = \begin{bmatrix} DN_{ij_1} \\ DN_{ij_2} \\ \vdots \\ DN_{ij_K} \end{bmatrix}
\]

• Quantization
  – Q = number of bits/pixel/band
  – Data space is quantized into \((2^Q)^K\) cells
• With 3 bands and 8 bits/pixel/band —> \(2^{563} = 16,777,216\) possible data vectors
Moments

- **DN mean vector**
  \[
  \mu = \begin{bmatrix} \mu_1 & \cdots & \mu_K \end{bmatrix}^T = \langle DN \rangle \quad \mu_k = \langle DN_k \rangle = \frac{1}{N} \sum_{p=1}^{N} DN_{pk}
  \]
  - centroid of the data in K-D space

- **DN covariance matrix**
  \[
  C = \begin{bmatrix} c_{11} & \cdots & c_{1K} \\ \vdots & \ddots & \vdots \\ c_{K1} & \cdots & c_{KK} \end{bmatrix} = \langle (DN - \mu)(DN - \mu)^T \rangle
  \]
  - where \( c_{mn} \) is the covariance between bands \( m \) and \( n \)
  - diagonal element \( c_{kk} \) is the DN variance in band \( k \)
  - \( C \) is a measure of the “spread” of the distribution about the mean vector, similar to the 1-D variance
For the Añasco image

```
>> whos
Name      Size            Bytes  Class          Value
C         7x7             392    double array
R         7x7             392    double array
number_of_bands  1x1     8       double array
number_of_columns  1x1     8       double array
number_of_rows    1x1     8       double array
pixels         188308x7   10545248 double array

Grand total is 1318257 elements using 10546056 bytes
```
Covariance matrix for the Añasco TM image

```
>> C=cov(pixels)

C =

1.0e+003 *

0.3461  0.2299  0.3122  0.2272  -0.0859  -0.1854  0.0037
0.2299  0.2167  0.3344  0.3210  0.2484  0.2327  0.0931
0.3122  0.3344  0.5760  0.5875  0.4950  0.4858  0.2150
0.2272  0.3210  0.5875  0.6855  0.8683  0.9566  0.3192
-0.0859  0.2484  0.4950  0.8683  2.2270  2.7072  0.5727
-0.1854  0.2327  0.4858  0.9566  2.7072  3.3229  0.6732
 0.0037  0.0931  0.2150  0.3192  0.5727  0.6732  0.2193
```
Correlation

• DN correlation matrix

\[ R = \begin{bmatrix} 1 & \cdots & \rho_{1K} \\ \vdots & \ddots & \vdots \\ \rho_{K1} & \cdots & 1 \end{bmatrix}, \quad -1 \leq \rho_{mn} \leq 1 \text{ or } |\rho_{mn}| \leq 1 \]

• Correlation coefficient between bands \( m \) and \( n \):

\[ \rho_{mn} = \frac{c_{mn}}{(c_{mm} \cdot c_{nn})^{1/2}} \]

– Covariances are normalized by standard deviation of each band

• Interpretation of correlation

– Measures the amount of linear relationship between pairs of bands
Correlation Between 2 Variables

- $\rho_{mn} \approx 1$
- $0 < \rho_{mn} < 1$
- $\rho_{mn} = 0$, $c_{mm} = c_{nn}$
- $\rho_{mn} \approx -1$
- $-1 \leq \rho_{mn} \leq 0$
- $\rho_{mn} = 0$, $c_{mm} < c_{nn}$

high correlation  moderate correlation  no correlation
Correlation matrix for the Añasco TM image

```matlab
>> R = corrcov(pixels);
>> R = corrcov(pixels)

R =

1.0000    0.8394    0.6994    0.4664   -0.0978   -0.1729    0.0135
0.8394    1.0000    0.9465    0.8327    0.3575    0.2743    0.4270
0.6994    0.9465    1.0000    0.9350    0.4371    0.3511    0.6048
0.4664    0.8327    0.9350    1.0000    0.7027    0.6338    0.8233
-0.0978    0.3575    0.4371    0.7027    1.0000    0.9952    0.8195
-0.1729    0.2743    0.3511    0.6338    0.9952    1.0000    0.7887
 0.0135    0.4270    0.6048    0.8233    0.8195    0.7887    1.0000
```
C and R Matrix Properties

• Properties of C and R
  – Symmetric, i.e., \( c_{mn} = c_{nm} \) and \( \rho_{mn} = \rho_{nm} \) and positive definite (their eigenvalues are real and positive)
  – If \( C \) and \( R \) are diagonal, the pixel values in bands m and n are uncorrelated

\[
C = \begin{bmatrix}
c_{11} & \cdots & c_{1K} \\
\vdots & \ddots & \vdots \\
c_{K1} & \cdots & c_{KK}
\end{bmatrix} \quad R = \begin{bmatrix}
1 & \cdots & \rho_{1K} \\
\vdots & \ddots & \vdots \\
\rho_{K1} & \cdots & 1
\end{bmatrix}
\]

– Multiband images can be transformed by the Principal Components Transform (PCT) so that \( C \) and \( R \) are diagonal, i.e. the transformed data are uncorrelated

• Applications
  – Designing decision boundaries for pattern recognition
  – Removing redundancy among spectral bands (data compression)
  – Designing color contrast enhancement methods
Multivariate Histogram

- Count the number of pixels in each DN “bin” and divide by the total number of pixels in the image

\[ \text{hist}_{DN} = \frac{\text{count}(DN)}{N} \]

- **Scalar** function of a vector
  - Each value is the pixel count for a particular DN vector, divided by the total number of pixels
  - Not easily visualized for \( K > 2 \)

- **K-D Histogram Models** (for reference)
  - Gaussian (Normal) distribution usually assumed for mathematical convenience

\[ N(f; \mu, \sigma) = \frac{1}{\sigma \sqrt{2\pi}} e^{-\frac{(f-\mu)^2}{2\sigma^2}} \]
\[ N(DN; \mu, C) = \frac{1}{|C|^{1/2}} (2\pi)^{K/2} e^{-\frac{(DN-\mu)^T C^{-1}(DN-\mu)}{2}} \]

1-D or Scalar Gaussian pdf      N-D or Multidimensional Gaussian pdf
Data Models

• Univariate Data Statistics
• Multivariate Data Statistics
• Scattergrams and Scatterplots
• Noise Models
• Scene and Sensor Effects
Scatterplot

• Scatterplot is a **binary** plot which shows a dot if a particular multispectral vector has a histogram count of at least one.
  – One way to visualize 2D & 3D data
  – Loses pixel counts: number of pixels with a particular multispectral vector is not shown.
  – Project 3-D scatterplot onto 2-D,

Example for $K = 3$, viewed from different directions
2-D Scatterplots

band 3 vs. band 2
band 4 vs. band 2
band 4 vs. band 3
Scatterplot Combinations

All possible 2-D scatterplots for a 7-band TM image
Figure 1

Scatter plot between Bands 1 and 5

\[ \rho_{15} = -0.098 \]
Scatter plot between Bands 1 and 3

\[ \rho_{13} = 0.694 \]
Figure 1: Scatter Plots for Añasco data using plotmatrix.
TM Bands 1-4 of the San Pablo (left) and Briones Reservoirs (right) north of Berkeley.
Scattergrams

• Graylevels used to represent frequency 0 to 1.
• Retains information about pixel counts in 2-D “clusters” from the Briones and San Pablo Reservoirs in Fig. 2–14 of the textbook.

“z-axis” is the fraction of pixels with a given $(DN_m, DN_n)$ vector

INEL6007(Spring 2010)  ECE, UPRM
From univariate to multivariate

- Using scalar histogram analysis for vector case
- Only when the channels are independent, we get that

\[ PDF(DN) = PDF(DN_1) \cdot PDF(DN_2) \cdot \ldots \cdot PDF(DN_K) \]
Data Models

• Univariate Data Statistics
• Multivariate Data Statistics
• Scattergrams and Scatterplots
• Noise Models
• Scene and Sensor Effects
Signal vs Noise

- **Signal**: varying quantity that carry information
- **Noise**:
  - In signal processing or computing it can be considered data without meaning; that is, data that is not being used to transmit a signal, but is simply produced as an unwanted by-product of other activities.
  - In Information Theory, however, noise is still considered to be information.
  - Noise can block, distort, or change the meaning of a message in both human and electronic communication.
It is problem dependent!

• RS over coastal regions
  – For someone studying optical properties of the water the contribution of the bottom is noise while the water column signal is information
  – For someone doing benthic habitat classification the water column contribution is noise while the bottom signal is information
Lets focus!

• Random variations created within the sensor system itself must surely be regarded as noise if the data is intended to provide information about the ground scene.
Sources of Noise

- **Atmospheric effects**
  - Atmosphere uniform and constant over the region of interest

- **Detector/Preamplifier noise**
  - Shot noise-creation and recombination of charge carriers
    - Zero mean Gaussian with variance directly related to the signal level.
  - Thermal Noise – motion of electrons above absolute zero
    - Zero mean Gaussian with variance, $v^2 = 4kT\Delta f$

- **Quantization noise**
  - Uniformly distributed over one discrete quantization interval.
Simple Models

• Additive
  \[ DN_p = \text{int}[a_p + n_p] \]

• Signal dependent noise
  \[ DN_p = \text{int}[a_p + n_p(a_p)] \]
System Noise Model

\[ E(U^2) = 4kT \Delta \lambda \]

Figure 9.4. System Noise Model.
Figure 9-5. Equivalent Noise Radiance vs. Radiance for the 0.52-0.60 μm band
Figure 9-6. Equivalent Noise Radiance vs. Radiance for the 0.75-0.91 μm band.
CALCULATED THEMATIC MAPPER NOISE
1.55-1.75 μm Band

Figure 9-7. Equivalent Noise Radiance vs. Radiance for the 1.55-1.75 μm band.
Figure 9-8. Probability of error vs. meteorological range.
Figure 29-9. Probability of error vs. quantization precision.
Experiment: Effects of Noise

• Agricultural area
• White Gaussian noise added
  – $s_{\text{meas}} = s + n$, $n \sim \mathcal{N}(128 I, \sigma^2 I)$
  – $\sigma = \text{brightness levels (8 bits, 256 levels)}$
• Bands used
  – 0.40-0.44 μm, 0.52-0.55 μm, 0.66-0.72 μm, 0.80-1.00 μm
• 10 classes
Figure 9-1. Examples in image form of noise added to data.
Effects in Classification Accuracy.
Figure 9-3. Classification accuracy vs. noise level.
Noise Dependent on Signal Level: Photographic Film
Push-broom Sensor Accumulation
(HYDICE, NVIS, Hyperion)

Image for single spectrum (i.e. fixed $\lambda$)

$y$, the second spatial dimension, is obtained in time as sensor is moved across the scene. Sensor can also acquire data using a whiskbroom scan procedure to increase FOV.
Stripping Noise: Pushbroom

Image with Stripping Noise
INEL6007(Spring 2010)

De-stripped image
ECE, UPRM
Whiskbroom Sensor Accumulation (AVIRIS, AVHRR/MODIS, LANDSAT)

Image for single spectrum (i.e. fixed $\lambda$)

$x$, the first spatial dimension is obtained by scanning orthogonal to the direction of sensor motion.

$y$, the second spatial dimension, is obtained in time as sensor is moved across the scene.
Badline scan noise in TM

*Landsat-1 MSS bad scanline noise*
TM banding noise

Landsat-4 TM banding noise
ASTER SWIR Bands Ghost Images

Bottom row is a contrast stretched version of the first row

INES6007(Spring 2010)  ECE, UPRM
Sensor Comparison

- Whiskbroom: Landsat Enhanced Thematic Mapper ETM+
- Pushbroom: Earth Observer – 1 Advanced Land Imager ALI

Signal-to-Noise Ratio (SNR) – Alaska low-light image, both 30m GIFOV

Landsat ETM+ (November 2000)  EO-1 ALI (December 2000)
AVIRIS 2004 Anomaly

Enrique

El Palo

Turrumote
Figure 9-10. Noise Sources.
Figure 9.11. Signal sources.
Statistical Measures of Image Quality

• Contrast
  – $C_{\text{ratio}} = \frac{DN_{\text{max}}}{DN_{\text{min}}}$
  – $C_{\text{range}} = DN_{\text{max}} - DN_{\text{min}}$
  – $C_{\text{std}} = \sigma_{DN}$

• Modulation
  – $M = \frac{DN_{\text{max}} - DN_{\text{min}}}{DN_{\text{max}} + DN_{\text{min}}}$
Signal to Noise Ratio

- $\text{SNR}_{\text{amplitude}} = \frac{C_{\text{signal}}}{C_{\text{noise}}}$
- $\text{SNR}_{\text{power}} = (\text{SNR}_{\text{amplitude}})^2$
- $\text{SNR}_{\text{var}} = \frac{\sigma_{\text{signal}}^2}{\sigma_{\text{noise}}^2}$
- $\text{SNR}_{\text{dB}} = 10 \log_{10}(\text{SNR})$

- Not a single definition. Most of the time estimated from the data itself.
Hyperion SNR

Radiometric performance model based on 60° Solar zenith angle, 30% albedo, standard scene

<table>
<thead>
<tr>
<th>Wavelength (nm)</th>
<th>Hyperion Measured SNR</th>
</tr>
</thead>
<tbody>
<tr>
<td>550 nm</td>
<td>161</td>
</tr>
<tr>
<td>650 nm</td>
<td>144</td>
</tr>
<tr>
<td>700 nm</td>
<td>147</td>
</tr>
<tr>
<td>1025 nm</td>
<td>90</td>
</tr>
<tr>
<td>1225 nm</td>
<td>110</td>
</tr>
<tr>
<td>1575 nm</td>
<td>89</td>
</tr>
<tr>
<td>2125 nm</td>
<td>40</td>
</tr>
</tbody>
</table>
Effect of Random Noise on Image Quality

$SNR_{std} = 1$
$SNR_{var} = 1$
$SNR_{dB} = 0$

$SNR_{std} = 2$
$SNR_{var} = 4$
$SNR_{dB} = 6$

$SNR_{std} = 5$
$SNR_{var} = 25$
$SNR_{dB} = 14$

$SNR_{std} = 10$
$SNR_{var} = 100$
$SNR_{dB} = 20$
Comparison of Random and Stripping Noise

noiseless

global random noise
detector striping noise
National Imagery Interpretability Scale

• NIIRS: Graduated 10 point scale designed to quantify and communicate the potential interpretability of panchromatic imagery

• Relate system parameters to the task to be performed
<table>
<thead>
<tr>
<th>rating level</th>
<th>example criterion</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>cannot be interpreted due to clouds or poor quality</td>
</tr>
<tr>
<td>1</td>
<td>distinguish airport taxiways and runways</td>
</tr>
<tr>
<td>2</td>
<td>detect large buildings</td>
</tr>
<tr>
<td>3</td>
<td>identify large ship type</td>
</tr>
<tr>
<td>4</td>
<td>identify individual tracks in railroad yard</td>
</tr>
<tr>
<td>5</td>
<td>identify individual railcars by type</td>
</tr>
<tr>
<td>6</td>
<td>identify automobiles as sedans or station wagons</td>
</tr>
<tr>
<td>7</td>
<td>identify individual railroad ties</td>
</tr>
<tr>
<td>8</td>
<td>identify vehicle windshield wipers</td>
</tr>
<tr>
<td>9</td>
<td>detect individual railroad tie spikes</td>
</tr>
</tbody>
</table>
Data Models

- Univariate Data Statistics
- Multivariate Data Statistics
- Scattergrams and Scatterplots
- Noise Models
- Spatial Statistics
- Scene and Sensor Effects
Spatial Statistics

• Spatial statistics → Geostatistics
• Random Fields: generalization of random variable/vector to include spatial relations
• Will focus on partial characterization
  – Second order statistics
    • Covariance and semivariograms
On histograms and scattergrams

\[ \text{hist}_h = \frac{\text{count}(DN_1, DN_2)}{N_h} \]

Look at scattergrams for this image

Joint PDF of two pixels separated by a distance \( h \)
Semivariogram Function

- The variogram \(2\gamma(x,y)\) is a function describing the degree of spatial dependence of a spatial random field or stochastic process.
- The variogram between two spatial locations \(r_1=(x_1,y_1)\) and \(r_2=(x_2,y_2)\)

\[2\gamma(r_1,r_2)=E[|Z(r_1)-Z(r_2)|^2]\]

- Its estimator

\[\hat{\gamma}(h) = \frac{1}{2N(h)} \sum_{h_{ij}=h} (z_i - z_j)^2\]
Covariance Function for a given lag $h_{ij} = h$

$$R(h) = \frac{1}{N(h)} \sum_{(i,j)|h_{ij}=h} DN_i DN_j$$

$$C(h) = \frac{1}{N(h)} \sum_{(i,j)|h_{ij}=h} DN_i DN_j - \mu_- h \mu_+ h$$
Describing the spatial variation: the semi-viariogram

Variogram cloud: a plot of a measure of height differences against the distance $d_{ij}$ between the control points for all possible pairs of points.

$$P_{ij}(d) = (z_i - z_j)^2$$
Describing the spatial variation: the semi-variogram

Example of variogram cloud

There is a trend such that height differences increase as the separation distance increases.

Indicating the farther apart two control points are, the greater is the likely difference in their value.
Describing the spatial variation: the semi-variogram

Spatial dependence can be described more concisely by the experimental semivariogram function as follows

\[ \hat{\gamma}(h) = \frac{1}{2N(h)} \sum_{h_{ij}=h} (z_i - z_j)^2 \]

\( N(h) \) is the number of pair of points at separation \( h \) and \( \hat{\gamma} \) is the estimated semi-variogram
Describing the spatial variation: the semi-variogram

\[ \hat{\gamma}(h) = \frac{1}{2N(h)} \sum_{h_{ij}=h} (z_i - z_j)^2 \]

This is the theoretical equation for variogram estimation and it is not straightforward in applications.

e.g. for a given distance h, it is more likely that there will be no pair of observations at precisely that separation.
Describing the spatial variation: the semi-variogram

In reality, variogram is estimated for different bands (or lags) rather than continuously at all distances.

\[ \hat{\gamma}(h) = \frac{1}{2N(h)} \sum_{h_j = h-\Delta/2}^{h+\Delta/2} (z_i - z_j)^2 \]

\( \Delta \) is the lag width
\( N(h) \) is the number of point pairs within (h- \( \Delta/2 \), h+ \( \Delta/2 \))
Figure 2.1: Example of data on a grid for the calculation of an experimental semi-variogram – iron ore.
Figure 2.3: Identifying all the pairs at 100ft apart in the east-west direction.

\[ \gamma^*(100) = \left[ (40 - 42)^2 + (42 - 40)^2 + (40 - 39)^2 + (39 - 37)^2 ight. \\
+ (37 - 36)^2 + (43 - 42)^2 + (42 - 39)^2 + (39 - 39)^2 \\
+ (39 - 41)^2 + (41 - 40)^2 + (40 - 38)^2 + (37 - 37)^2 \\
+ (37 - 37)^2 + (37 - 35)^2 + (35 - 38)^2 + (38 - 37)^2 \\
+ (33 - 37)^2 + (37 - 33)^2 + (33 - 34)^2 + (35 - 38)^2 \\
+ (35 - 37)^2 + (37 - 36)^2 + (36 - 36)^2 + (36 - 35)^2 \\
+ (36 - 35)^2 + (35 - 36)^2 + (36 - 35)^2 + (35 - 34)^2 \\
+ (34 - 33)^2 + (33 - 32)^2 + (32 - 29)^2 + (29 - 28)^2 \\
+ (38 - 37)^2 + (37 - 35)^2 + (29 - 30)^2 \\
\left. + (30 - 32)^2 \right] \times 36 \]

\[ \gamma^*(100) = 1.46(\%)^2 \]
\[ \gamma^*(200) = [(44 - 40)^2 + (40 - 40)^2 + (42 - 39)^2 + (40 - 37)^2 \\
+ (39 - 36)^2 + (42 - 43)^2 + (43 - 39)^2 + (42 - 39)^2 \\
+ (39 - 41)^2 + (39 - 40)^2 + (41 - 38)^2 + (37 - 37)^2 \\
+ (37 - 35)^2 + (37 - 38)^2 + (35 - 37)^2 + (38 - 37)^2 \\
+ (37 - 33)^2 + (37 - 34)^2 + (38 - 35)^2 + (35 - 36)^2 \\
+ (37 - 36)^2 + (36 - 35)^2 + (36 - 36)^2 + (35 - 35)^2 \\
+ (36 - 34)^2 + (35 - 33)^2 + (34 - 32)^2 + (33 - 29)^2 \\
+ (32 - 28)^2 + (38 - 35)^2 + (35 - 30)^2 + (30 - 29)^2 \\
+ (29 - 32)^2] \] 

\[ \text{ECE: } (\text{PRM3}) \]

\[ \gamma^*(200) = 3.30(\%)^2 \]
<table>
<thead>
<tr>
<th>Direction</th>
<th>Distance between samples (ft)</th>
<th>Experimental semi-variogram</th>
<th>Number of pairs</th>
</tr>
</thead>
<tbody>
<tr>
<td>East-west</td>
<td>100</td>
<td>1.46</td>
<td>36</td>
</tr>
<tr>
<td></td>
<td>200</td>
<td>3.30</td>
<td>33</td>
</tr>
<tr>
<td></td>
<td>300</td>
<td>4.31</td>
<td>27</td>
</tr>
<tr>
<td></td>
<td>400</td>
<td>6.70</td>
<td>23</td>
</tr>
<tr>
<td>North-south</td>
<td>100</td>
<td>5.35</td>
<td>36</td>
</tr>
<tr>
<td></td>
<td>200</td>
<td>9.87</td>
<td>27</td>
</tr>
<tr>
<td></td>
<td>300</td>
<td>18.88</td>
<td>21</td>
</tr>
</tbody>
</table>

Table 2.1: Calculation of experimental semi-variogram values in two major directions for iron ore example on square grid
Figure 2.5: Experimental semi-variograms in the two major directions for the iron ore example.
Table 2.2: Calculation of semi-variogram in diagonal direction for iron ore
Describing the spatial variation: the semi-variogram

\[ \Delta = 0.5 \quad \text{What is the value of } \gamma(0.5)? \]
\[ h = 0.5 \quad \text{What is the value of } \gamma(1.5)? \]

\[
\begin{bmatrix}
    a & b & c & d & e \\
    a & 0 & 1 & 1 & 3 & 2.5 \\
    b & 1 & 0 & 2 & 2 & 3 \\
    c & 1 & 2 & 0 & 3 & 1 \\
    d & 3 & 2 & 3 & 0 & 3 \\
    e & 2.5 & 3 & 1 & 3 & 0
\end{bmatrix}
\]

Distance matrix
Describing the spatial variation: the semi-v variogram
Parameters used to Describe the Semivariogram

- **Nugget** ($c_0$): variance at zero distance
- **Range** ($a$): the distance at which the semivariogram levels off and beyond which the semivariance is constant
- **Sill** ($c_0 + c_1$): the constant semivariance value beyond the range
Summarize the spatial variation by a regular mathematical function

Having approximated the semivariograms by mean values at a series of lags, the next step is to summarize the experimental variogram using a mathematical function.
TABLE 4-2. Some 1-D continuous models for the discrete spatial covariance and semivariogram (Isaaks and Srivastava, 1989; Pratt, 1991; Carr, 1995). Note each of the semivariogram models is normalized to a sill value of one.

<table>
<thead>
<tr>
<th>spatial statistic</th>
<th>model</th>
<th>equation</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>exponential</td>
<td>[ C(h) = C(0)e^{-\alpha h} ]</td>
</tr>
<tr>
<td></td>
<td>(Markov model)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Gaussian</td>
<td>[ \gamma(h) = 1 - e^{-3(h/\alpha)^2}, \text{ where } \alpha = \gamma(0.95\text{sill}) ]</td>
</tr>
<tr>
<td></td>
<td>exponential</td>
<td>[ \gamma(h) = 1 - e^{-3(h/\alpha)}, \text{ where } \alpha = \gamma(0.95\text{sill}) ]</td>
</tr>
</tbody>
</table>
|                   | spherical        | \[ \gamma(h) = \begin{cases} 
1.5h/\alpha - 0.5(h/\alpha)^3, & h \leq \alpha \\
1, & h > \alpha 
\end{cases}, \text{ where } \alpha = \gamma(\text{sill}) \] |
Three Normalized Semivariogram Models
Summarize the spatial variation by a regular mathematical function

**Nugget model**: A constant variance model

\[ \gamma = c_0 \]
Summarize the spatial variation by a regular mathematical function

Linear model: Variances change linearly with the change of distance

\[ \gamma = d \quad \text{When } d < a \]
Summarize the spatial variation by a regular mathematical function

Spherical model starts from a nonzero variance \((c_0)\) and rise as an elliptical arc to a maximum value \((c_0+c_1)\) at distance \(a\).

If \(d \leq a\) then

\[
\gamma(d) = c_0 + c_1 \left[ \frac{3d}{2a} - 0.5 \left( \frac{d}{a} \right)^3 \right]
\]

If \(d > a\) then

\[
\gamma(d) = c_0 + c_1
\]
Covariance Function

• The covariance function \(C(r_1, r_2)\) gives the covariance of the values of the random field at the two locations \(r_1\), and \(r_2\).

\[
C(r_1, r_2) = \text{Cov}(Z(r_1), Z(r_2))
\]

• When \(r\) is a scalar as with time, this corresponds to the autocorrelation function of a stochastic process
Covariance Function

- 1D Estimator: Single Direction

\[ C(h) = \sum_{i,j \mid h_{ij} = h} DN_i DN_j \]

\[ C(h) = \frac{1}{N(h)} - \mu_h \mu_h \]

can use MATLAB Autocovariance function estimator

- Relation to semivariogram

\[ \gamma(h) = C(0) - C(h) \]
FIGURE 4-22. Scanned aerial image used to illustrate spatial statistics. Three groups of three transects each are extracted from regions of distinctly different vegetation density and plant size: no vegetation (top), low density (middle), and high density (bottom). Each transect is 100 pixels long.
FIGURE 4-23. Covariance functions and semivariograms for the transects of Fig. 4-22. The average function for the three transects in each category is plotted.
Normalized Covariance Function
\[
\frac{C(h)}{C(0)} = Ke^{-\alpha h}
\]

Figure 6.35: Exponential model fitted to covariance functions from Fig. 4-24. Note that the horizontal axis is scaled differently in each graph.
TABLE 4-4. Correlation lengths obtained from the exponential model fits in Fig. 4-25.

<table>
<thead>
<tr>
<th>category</th>
<th>$\alpha$ (pixels$^{-1}$)</th>
<th>correlation length (pixels)</th>
</tr>
</thead>
<tbody>
<tr>
<td>no vegetation</td>
<td>0.114</td>
<td>8.78</td>
</tr>
<tr>
<td>low density</td>
<td>0.422</td>
<td>2.37</td>
</tr>
<tr>
<td>high density</td>
<td>0.560</td>
<td>1.79</td>
</tr>
</tbody>
</table>
Quiz

- What are univariate statistics
- Calculate histogram, mean, variance, skewness and kurtosis for given data
- Compute covariance, correlation and scatterplot between the bands for the given data.
- What are the different types of noise and how they affect image quality
- What are statistical measures of image quality
- Explain contrast, modulation, SNR
- Compute semivariogram for the matrix shown at $h=1,2,3$ in the E-W direction and plot it.
- Interpret the plot of covariance function
- Do exercise 4-1, 4-4, 4-7
Going 2D: Isotropic

\[ C(r) = C(0)e^{-\alpha r} \]

\[ r = (h_x^2 + h_y^2)^{1/2} \]
Going 2D: Anisotropic

\[ C(x, y) = C(0)e^{-\alpha|x|}e^{-\beta|y|} = C(0)e^{-(\alpha|x| + \beta|y|)} \]

\[ C(h_x, h_y) = C(0)e^{-\sqrt{\alpha h_x^2 + \beta h_y^2}} \]
FIGURE 7-26. Three possible forms for 2-D exponential covariance models, in perspective and as contour maps. For the anisotropic model, $\alpha = \beta/4$. 
Other Possibilities

- Co-Ocurrence Matrix and Texture Features
- Fractal Geometry
Data Models

• Univariate Data Statistics
• Multivariate Data Statistics
• Scattergrams and Scatterplots
• Noise Models
• Spatial Statistics
• Scene and Sensor Effects
Topography Effects Study

• Model: radiance of scene is proportional to cosine of solar-surface angle (Chapter 2)

\[ L(x, y) = a \rho(x, y) \cos \theta(x, y) + b \]
Shaded Relief and Class Maps

• Create shaded-relief image from DEM using cosine irradiance model

• Create class masks for spectral classes soil and vegetation
  – class masks are mutually exclusive, i.e. a pixel is either soil or vegetation

\[ \cos \theta(x, y) \]

\[ \text{solar irradiance spatial variation} \]

\[ \text{spatial distribution of classes} \]

soil

vegetation
Synthetic Scene With Relief

- Create randomized spectra for each class (Gaussian)
- Set each pixel to a sample from the appropriate class distribution
  - \( \rho(x,y) \) for each class
- Multiply spectral maps by shaded-relief image
- Combine classes in each band

\[ \rho(x, y) \cos \theta(x, y) \] for each class

\[
\begin{align*}
\text{Red Band} & \quad \rho(x, y) \\
\text{NIR Band} & \quad \cos \theta(x, y)
\end{align*}
\]
Final Simulated Images

soil pixels

red band +

vegetation pixels

NIR band +

final simulated images
Scattergram Analysis

- Without topography (flat terrain): two circular distributions, zero correlation
- With topography: two elliptical distributions, high correlation

Spectral scattergrams without and with topography

\[ \rho(x, y) \]

\[ \rho(x, y) \cos \theta(x, y) \]

Without topography

With topography

Conclusion: topography creates spectral correlation
Influence of Sensor in Spatial Statistics: Noise and GIFOV

noisy sensor

larger GIFOV
FIGURE 4-39. Covariance and semivariogram functions for the original image and with added spatially-uncorrelated noise. Note that the vertical scale is different for each category (compare to Fig. 4-23).
FIGURE 4-40. Covariance functions for the original image with a simulated GIFOV of 5-by-5 pixels. Note that the image scale is different for each category.
Normalized covariance and semi-variance results for the original image with a simulated GIFOV of 5-by-5 pixels. Note that the vertical scale is different for each category (compare to Fig. 4.23).
Sensor Spatial Response Study

- Model: Sensor spatial response causes image blurring and mixing of spectral signatures (Chapter 3)

**Sensor Spatial Effects**

- Create synthetic scene containing soil, vegetation, and water
- Create randomized spectra for each class
- Convolve with 5 x 5 pixel Gaussian spatial response
- Compare before/after spectral scattergrams
Synthetic Scene

- Create class maps for spectral classes soil, vegetation, water
  - class maps are mutually exclusive, i.e. a pixel is only one class
  - no topography
- Create randomized spectra (random texture) for each class (Gaussian)
- Set each pixel to a sample from the appropriate class distribution
  - $\rho(x,y)$ for each class
- Combine classes in each band

\[ \rho(x, y) \text{ for each band} \]

\[ \text{red band} \quad \text{NIR band} \]

\[ \text{class maps, no texture} \]

\[ \text{uncorrelated texture} \]

\[ \text{spatially-correlated texture} \]
More on the Synthetic Scene
Simulate Blurring

- Convolved with 5 x 5 pixel Gaussian PSF
  - represents image from simulated sensor
Blurring: Scattergram Analysis

red band

NIR band

uncorrelated texture

mixed vegetation and soil

mixed soil, vegetation, and asphalt

mixed soil and asphalt

red

spatially correlated texture

INEL6007(Spring 2010)  ECE, UPRM
Conclusion: spatial integration creates “mixing” of spectral signatures
Blurring + Topography

red band

NIR band

NIR

vegetation

mixed vegetation and soil

soil

red
Summary

• Common models used in image data analysis were described and illustrated with examples. The connection between radiation, sensor models and data models was discussed and explored by simulation.

• Spectral statistics are influenced by topography
  – Induce correlation between spectral bands

• Spectral statistics are influenced by the sensor’s spectral passband locations and widths, and noise characteristics.

• Spatial and spectral characteristics are influenced by the sensor’s PSF
  – Increase spatial correlation length
  – Reduce within class variance
  – Created mixed pixels