Information Content of Ocean Colour Data

Dr. A. Neumann, H. Krawczyk
German Aerospace Centre DLR
Remote Sensing Technology Institute
Marine Remote Sensing

• What is Information Content?
• Instrument and Target Aspects
• Analysis of Information Content
• Retrievability of Parameters
• Multivariate Analysis - Principal Components
  and Intrinsic Dimensionality

What is Information Content?

• ... the attempt to measure (i.e. quantify, enumerate) the “valuable information” in a set or stream of data

• different approaches worked out in information theory, statistics and applied sciences, depending on the problem to be solved (optimal coding strategies, data compression, parameter retrieval, ...)

• in the context of ocean colour remote sensing: requirements to spectral characteristics (resolution, number of channels, SNR) from the interpretation/applicational perspective
**Parameters to Quantify or Handle Information Content:**

- entropy (statistics, data coding)
- correlation (spectral, spatial)
- dominant factors (Factor Analysis)
- significant Ortho-Components (Principal Component Analysis, Karhunen-Loewe Transformation)

- *in our context leading to*
  
  ⇒ intrinsic dimensionality analysis
  ⇒ “retrievability” of parameters
  ⇒ error analysis of retrieval algorithms

**Information Content of Remote Sensing Data**

**Instrument Aspects**

- Spectral Resolution
- Number of Channels
- Spectral Range
- Spatial Resolution
- Radiometric Accuracy
- Instrument Noise

**Target Aspects**

- Target Spatial Variation
- Target Spectral Variation
- "Noise" of Target
- Target Dynamics
Analysis of Information Content

The Remote Sensing Problem

- The inverse task is analytically not solvable
- Which and how many parameters can be retrieved
- Which and how many channels contribute how much to each parameter
- How to construct an "optimal" algorithm
How to Perform an Information Analysis?

- Using experimental data, i.e. satellite and in-situ measurements:
  - quality/accuracy of the data often not known
  - influencing physical parameters not under control
  - problem: large number (>10^4) of data sets for various combinations of parameters necessary

- Using bio-optical and radiative transfer models:
  - clear understanding of mathematical and physical interdependencies
  - variability and number of parameters under control
  - problem: adequateness of the model for practical application

Retrievability of Parameters - Basics

- We have measurements/simulations of spectral signatures, i.e. values of radiance/reflectance at certain \( \lambda \) (channels) with certain \( \Delta \lambda \) (spectral halfwidth) and a given SNR and corresponding (bio-) physical parameters to be retrieved from the data \{L_1, ..., L_N, C, S, Y, \tau_A, ...\}

- generally no link between ONE parameter and ONE channel
- usually significant spectral correlation, at least in spectral subranges
- influence of the parameters spectrally overlapping
- statistical behaviour and interdependencies of parameters not predictable a priori

\( \Rightarrow \) first (and simple) test: numerical analysis
\( \Rightarrow \) second investigation: multivariate analysis
Spectral Overlap of Different Constituents

<table>
<thead>
<tr>
<th>Spectral range</th>
<th>Chlorophyll</th>
<th>Gelbstoff</th>
<th>Sediment</th>
</tr>
</thead>
<tbody>
<tr>
<td>blue</td>
<td>absorption</td>
<td>strong absorption</td>
<td>scattering</td>
</tr>
<tr>
<td>green</td>
<td>scattering</td>
<td>weak absorption</td>
<td>scattering</td>
</tr>
<tr>
<td>red</td>
<td>scattering,</td>
<td>absorption</td>
<td>scattering</td>
</tr>
<tr>
<td></td>
<td>absorption,</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>fluorescence</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Atmosphere/aerosol: scattering at all wavelengths

Retrievability of Parameters - Numerical Analysis

- Generate a large number of M spectra \{L_1, ..., L_N\} = f(C, S, Y, \tau_A) using radiative transfer model, N - number of available spectral channels
- degrade the "ideal" values by a given NEΔL or relative "measurement" error
- anlyse, how many spectra can be distinguished, i.e. are at least in one channel different to each other
- example for MOS channels: data set of 140,000 spectra for independently varying parameters of C = 0..100 µg/l, a_Y(440nm) = 0..1, b_S(550nm)=0..10, \tau_A = 0..0.25

<table>
<thead>
<tr>
<th>Relative Error [%]</th>
<th>0</th>
<th>1</th>
<th>3</th>
<th>5</th>
<th>7</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>Discriminated Spectra [%]</td>
<td>97</td>
<td>69</td>
<td>25</td>
<td>12</td>
<td>8</td>
<td>4</td>
</tr>
</tbody>
</table>
Numerical Analysis - Conclusion

• for theoretical case of zero error the result is not too bad, disillusioning for realistic error values
• underlines high requirements to measurement accuracy
• more detailed analysis of error propagation and influence of single parameters possible, however, this seems not the adequate tool to handle correlations and interdependencies

⇒ ill-defined, multivariate problem which needs more sophisticated tools for analysis

Retrievability of Parameters - Multivariate Analysis

• Multivariate analysis (e.g. factor analysis, principal component analysis) provide an effective mathematical tool to analyse large data sets in the presence of correlation and interdependencies

• the principle is an “optimal” transformation of the original data set to an orthogonal system containing identical information, accounting for (spectral) correlation

• leads to diagonalisation of the covariance matrix

• is a purely mathematically defined tool, not accounting for physical background represented by the data
Principal Component Analysis (PCA)

- Let $\text{Cov}\{L_j/\Delta L_j\}$, $j=1..N$, be the error normalised spectral covariance matrix of the data set, the principal components are computed as:

$$PC_k = \sum_{m=1}^{N} \frac{U_{km}(L_m-\overline{L}_m)}{\Delta L_m}$$

- Where:
  - $PC_k$ - kth principal component, k from 1 to N
  - $N$ - number of spectral bands used in the original data set
  - $U_{km}$ - kth eigenvector of the covariance matrix
  - $L_m$ - radiance in the mth spectral band
  - $\overline{L}_m$ - mean radiance value in the mth spectral band
  - $\Delta L_m$ - noise-equivalent radiance, measurement error.

- Due to the transformation principle (diagonalisation of the covariance matrix) the PC are uncorrelated (orthogonal) and contain successively maximal information in the sense of signal variance:

$$\lambda_i \geq \lambda_j$$ for $i>j$

PCA - SNR and Intrinsic Dimensionality

- Since the computation was done using error-normalised radiances, the eigenvalue gives directly the SNR of the corresponding PC:

$$\lambda_k = \text{SNR}^2(PC_k)$$

with $\lambda_k$ being the kth eigenvalue of $\text{Cov}\{L_j/\Delta L_j\}$

- Based on this we can define the intrinsic dimensionality of the analysed data set as:

$$D = \max(k) \text{ with } \sqrt{\lambda_k} > 1, D < N.$$

- The intrinsic dimensionality gives a measure, how many PCs contain significant information (in the sense of variance), i.e. are necessary to reconstruct the full information content of the original data set. It is, however, an unsharp definition.
What does Intrinsic Dimensionality Tell Us?

- It tells us how many statistically independent parameters can be retrieved from the data set in principle (but not which ones and how!)
- It tells us whether the inverse task is solvable in general (D ≠ number of desired parameters)
- It does not say, that we need only D spectral channels, because:
  - each parameter retrieval from the spectral measurement can be seen as a statistical estimate ⇒ general rule: the larger the number of measurements, the higher the achievable accuracy!
- Examples:
  - for simulated MOS-data (varying C, S, Y, τ, α) we get D up to 6
  - real MOS-data over water D ~ 5

Appendix: Technical Parameters MOS-IRS

<table>
<thead>
<tr>
<th>Parameter</th>
<th>MOS-A</th>
<th>MOS-B</th>
<th>MOS-C</th>
</tr>
</thead>
<tbody>
<tr>
<td>Spectral Range [nm]</td>
<td>755 - 768</td>
<td>408 - 1010</td>
<td>SWIR</td>
</tr>
<tr>
<td>No. of Channels</td>
<td>4</td>
<td>13</td>
<td>1</td>
</tr>
<tr>
<td>Wavelengths [nm]</td>
<td>756.7; 760.6; 763.5; 766.4 O2A-band</td>
<td>408; 443; 485; 520; 570; 615; 650; 685; 750; 870; 1010; 815; 945 (H2O-vapor)</td>
<td>1600</td>
</tr>
<tr>
<td>spectral halfwidth [nm]</td>
<td>1.4</td>
<td>10</td>
<td>100</td>
</tr>
<tr>
<td>Field of View</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>along track x [deg]</td>
<td>0.344</td>
<td>0.094</td>
<td>0.14</td>
</tr>
<tr>
<td>across track [deg]</td>
<td>13.6</td>
<td>14.0</td>
<td>13.4</td>
</tr>
<tr>
<td>Swath Width [km]</td>
<td>195</td>
<td>200</td>
<td>192</td>
</tr>
<tr>
<td>No. of Pixels</td>
<td>140</td>
<td>384</td>
<td>299</td>
</tr>
<tr>
<td>Pixel Size x*y [km²]</td>
<td>1.57x1.4</td>
<td>0.52x0.52</td>
<td>0.52x0.64</td>
</tr>
<tr>
<td>Measuring Range</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lmin [µWcm⁻²nm⁻¹sr⁻¹]</td>
<td>0.1</td>
<td>0.2</td>
<td>0.5</td>
</tr>
<tr>
<td>Lmax [µWcm⁻²nm⁻¹sr⁻¹]</td>
<td>40</td>
<td>65</td>
<td>18</td>
</tr>
</tbody>
</table>