Signal Processing and Time-Series Analysis

1. Signal Processing
   A. Analytical Signals are recorded as:
      Spectra, chromatograms, voltammograms or titration curves
      (monitored in frequency, wavelength, time)

   B. Signal processing is used to distinguish between signal and noise.

   ![Diagram of signal processing]

2. Signal Processing
   C. Methods of Evaluating Analytical Signals
      1) Transformation
      2) Smoothing
      3) Correlation
      4) Convolution
      5) Deconvolution
      6) Derivation
      7) Integration

   Important as data is usually processed digitally
Signal Processing and Time-Series Analysis

D. Digital smoothing and Filtering

1) Moving Average Filtering – smoothes data by replacing each data point with the average of the neighboring data points:

\[ y_s(i) = \frac{1}{2N+1}[y(i+N) + y(i+N-1) + \ldots + y(i-N)] \]

Where \( y_s(i) \) is the smoothed value for the ith data point, \( N \) is the # of neighboring data points on either side of \( y_s(i) \), and \( 2N+1 \) is the span (filter width).

Signal Processing and Time-Series Analysis

D. Digital Smoothing and Filtering

1. Moving Average Filtering – Rules for selecting the most appropriate filter:

- When applied repetitively, the largest smoothing effect (>95%) is observed in the first application (single smoothing usually sufficient).

- Filter width should correspond to the full width at half maximum of a band or a peak.
  - Too small a width results in unsatisfactory smoothing.
  - Too large of a width leads to distortion of the original data structure.

- Distortion of data structure is more severe in respect of the area than of the height of the peaks.
  - Filter width selected must be smaller if the height rather than the area is evaluated.
Signal Processing and Time-Series Analysis

D. Digital Smoothing and Filtering
1. Moving Average Filtering

Note: The influence of the filter-width on the distortion of the peaks can be quantified by means of the relative filter width, $b_{\text{relative}}$:

$$b_{\text{relative}} = \frac{b_{\text{filter}}}{b_{0.5}}$$

Where $b_{\text{filter}}$ is the filter width, and $b_{0.5}$ is the full width at half maximum.

Signal Processing and Time-Series Analysis

E. Savitzky-Golay Filter (Polynomial smoothing)

→ smoothing that seeks to preserve shapes of peaks

-After deciding on the filter width, the filtered value for the $k$th data point is calculated from:

$$y_k = \frac{1}{N\text{ORM}} \sum c_j y_{k+j}$$

where NORM is a normalization factor obtained from the sum of the coefficients $c_j$. 
**Signal Processing and Time-Series Analysis**

F. Kalman Filter
- Estimate the state of a system from measuring which contain random errors
- Based on two models:
  1) Dynamic System model (Process)
     \[ x(k) = x(k-1) + w(k-1) \]
  2) Measurement Model
     \[ y(k) = H^T(h) x(h) + v(h) \]
- where \( x \) = state vector, \( y \) = the measurement, \( F \) = system transition matrix and \( H \) = the measurement vector (matrix).
- \( w \) = signal noise vector, \( v \) = measurement noise vector
- \( h \) = denotes the actual measurement or time

**Signal Processing and Time-Series Analysis**

F. Kalman Filter
1) only matrix operations allowed

\[
\begin{bmatrix}
X_n \\
Y_k
\end{bmatrix} = \begin{bmatrix}
1 & 0 & X_{n-1} \\
0 & 1 & Y_{n-1}
\end{bmatrix} + \begin{bmatrix}
\sim V_{k-1} \\
\sim Y_{k-1}
\end{bmatrix}
\]

\( X_n \) = state transition
\( Y_k \) = state noise
F. Kalman Filter
b) Measurement Model

\[
\begin{bmatrix}
U_k \\
V_k
\end{bmatrix} = \begin{bmatrix}
H_x & 0 \\
0 & H_y
\end{bmatrix} \begin{bmatrix}
X_n \\
Y_n
\end{bmatrix} + \begin{bmatrix}
\sim U_k \\
\sim Y_k
\end{bmatrix}
\]

Measurement measurement state noise

matrix

G. Signal Derivatives

→ useful for eliminating background noise, determining peak position and improving the visual resolution of peaks.
**Signal Processing and Time-Series Analysis**

G. Signal Derivatives

Ex: Noise characteristics for derivatives of signals (y-signal around an observation point h).

<table>
<thead>
<tr>
<th></th>
<th>$y_{h-2}$</th>
<th>$y_{h-1}$</th>
<th>$y_h$</th>
<th>$y_{h+1}$</th>
<th>$y_{h+2}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.2</td>
<td>0.5</td>
<td>0.7</td>
<td>0.4</td>
<td>0.1</td>
<td></td>
</tr>
</tbody>
</table>

- can calculate filtered values and their standard deviations by means of the tabulated filter coefficients.
Signal Processing and Time-Series Analysis

H. Transformations
→ useful for filtering of data, convolution and deconvolution of analytical signals, integration, background correction and reducing data points.

1) Fourier Transforms – integral transform that re-expresses a function
\[ x(k) = x(k-1) + w(k-1) \]
2) Measurement Model
\[ y(k) = H^T(k) x(\text{h}) + v(\text{h}) \]

- where x = state vector, y = the measurement, F = system transition matrix and H = the measurement vector (matrix).

- w = signal noise vector, v = measurement noise vector

- h = denotes the actual measurement or time

---

Figure: Fourier Transform of a One Dimensional Signal

Source: Nikos Drakos, Computer Based Learning Unit, University of Leeds & Kristian Sandberg, University of Colorado.
3. Hadamard Transformation

→ Based on the Walsh Function in contrast to the sine and cosine functions of FT.

\[ y^* = H_y \]

where \( H \) is the \((n \times n)\) Hadamard matrix, \( y \) is the vector of the original \( n \) signal values and \( y^* \) is the vector of the transformed \( n \) signal values.

The \( h^{th} \) Hadamard Transform matrix

\[
H_h = \begin{bmatrix}
H_{h-1} & H_{h-1} \\
H_{h-1} & -H_{h-1}
\end{bmatrix}
\]

Ex: Four data points are to be treated with HT. with \( n = 2^h = 4 \) we have \( h=2 \). If we set \( H_0 = 1 \) then we would obtain for the matrices \( H_1 \) and \( H_2 \):

\[
H_1 = \begin{pmatrix}
1 & 1 \\
1 & -1
\end{pmatrix}
\]

\[
H_2 = \begin{pmatrix}
H_1 & H_1 \\
H_1 & -H_1
\end{pmatrix} = \begin{bmatrix}
1 & 1 & 1 & 1 \\
1 & -1 & 1 & -1 \\
1 & 1 & -1 & -1 \\
1 & -1 & -1 & 1
\end{bmatrix}
\]
3. Hadamard Transformation

Transformation Equation (according to \( y^* = Hy \)):

\[
\begin{pmatrix}
    y^*_1 \\
    y^*_2 \\
    y^*_3 \\
    y^*_4
\end{pmatrix} =
\begin{pmatrix}
    1 & 1 & 1 & 1 \\
    1 & -1 & 1 & -1 \\
    1 & 1 & -1 & -1 \\
    1 & -1 & -1 & 1
\end{pmatrix}
\begin{pmatrix}
    y_1 \\
    y_2 \\
    y_3 \\
    y_4
\end{pmatrix}
\]

multiplication of the equations = transformed signal

\[
y^*_1 = y_1 + y_2 + y_3 + y_4
\]

\[
y^*_2 = y_1 + y_2 + y_3 + y_4
\]

…… and so on.

*Insert own #’s to transform your signals.

Advantages over FT

a. Simple arithmetic operations (addition & subtraction)
b. Faster algorithm
c. Real (no imaginary transformations)

Applications

a. Signal filtering – suppresses high frequency noise or drift
b. Convolution and Deconvolution – restoration of signal distorted by instrument function or overlapping signals
c. Integration – of area (how does this differ from peak height?)
d. Data reduction and background correction
4. Time-Series Analysis

-characterization of a set of measurements as a function of time
e.g. Phosphorus concentrations in rivers:

**Time-Series Analysis**

1. Field-Based/Submersible Approaches
2. Laboratory-Based Approaches
3. Chemometric Approach

Experimental Design and Multivariate Data Analysis
**Time-Series Analysis**


<table>
<thead>
<tr>
<th>Month</th>
<th>EpCO₂</th>
<th>Alkalinity (m Eq L⁻¹)</th>
<th>Calcium (m Eq L⁻¹)</th>
<th>pH</th>
<th>Conductivity (µ S)</th>
<th>Discharge (m³ s⁻¹)</th>
<th>FRP (µ M)</th>
<th>TP (µ M)</th>
</tr>
</thead>
<tbody>
<tr>
<td>January</td>
<td>10.56</td>
<td>3.94</td>
<td>4.77</td>
<td>8.32</td>
<td>514</td>
<td>8.40</td>
<td>4.38</td>
<td>7.38</td>
</tr>
<tr>
<td>February</td>
<td>12.61</td>
<td>4.01</td>
<td>4.80</td>
<td>7.75</td>
<td>507</td>
<td>8.03</td>
<td>4.04</td>
<td>6.88</td>
</tr>
<tr>
<td>March</td>
<td>10.72</td>
<td>4.11</td>
<td>4.90</td>
<td>7.86</td>
<td>521</td>
<td>4.87</td>
<td>3.44</td>
<td>5.97</td>
</tr>
<tr>
<td>April</td>
<td>9.87</td>
<td>4.05</td>
<td>4.86</td>
<td>7.89</td>
<td>508</td>
<td>4.24</td>
<td>2.87</td>
<td>5.54</td>
</tr>
<tr>
<td>May</td>
<td>9.43</td>
<td>4.14</td>
<td>4.84</td>
<td>7.91</td>
<td>515</td>
<td>6.33</td>
<td>3.38</td>
<td>6.22</td>
</tr>
<tr>
<td>June</td>
<td>8.33</td>
<td>4.13</td>
<td>4.81</td>
<td>7.86</td>
<td>515</td>
<td>5.61</td>
<td>4.67</td>
<td>8.23</td>
</tr>
<tr>
<td>July</td>
<td>11.10</td>
<td>4.07</td>
<td>4.82</td>
<td>7.83</td>
<td>514</td>
<td>5.71</td>
<td>5.34</td>
<td>7.88</td>
</tr>
<tr>
<td>August</td>
<td>11.32</td>
<td>4.12</td>
<td>4.82</td>
<td>7.84</td>
<td>510</td>
<td>6.84</td>
<td>5.47</td>
<td>8.87</td>
</tr>
<tr>
<td>September</td>
<td>13.67</td>
<td>3.96</td>
<td>4.65</td>
<td>7.74</td>
<td>514</td>
<td>7.37</td>
<td>5.70</td>
<td>9.40</td>
</tr>
<tr>
<td>October</td>
<td>13.82</td>
<td>3.91</td>
<td>4.62</td>
<td>7.73</td>
<td>511</td>
<td>6.90</td>
<td>5.70</td>
<td>8.44</td>
</tr>
<tr>
<td>November</td>
<td>12.53</td>
<td>3.72</td>
<td>4.56</td>
<td>7.75</td>
<td>505</td>
<td>6.92</td>
<td>4.89</td>
<td>7.88</td>
</tr>
<tr>
<td>December</td>
<td>10.22</td>
<td>3.84</td>
<td>4.62</td>
<td>7.83</td>
<td>489</td>
<td>7.09</td>
<td>4.36</td>
<td>8.29</td>
</tr>
</tbody>
</table>

Any Relationships?
A. Correlation Methods

1) Autocorrelation or Autocovariance – correlations within a time series.

2) Cross Correlation – correlations between two different time series.

-Correlations can be found if the data are plotted against successive values:
**Time-Series Analysis**

Ex: Figure 3-14, monthly sulfur concentrations \( y(t) \) is plotted against time, \( t \).

Correlations obtained by plotting the measurement at \( t \), \( y(t) \), against the value at time \( t+1 \), i.e. \( y(t+1) \), at time \( y(t+2) \) or in general at time \( y(t + \tau) \), \( \tau \) represents the Log time.

![Graphs showing correlations](image)

**Time-Series Analysis**

- Empirical autocorrelation is applied to measure amount of correlation

\[
r(\tau) = \frac{\sum_{t=1}^{n-\tau} (y_t - \bar{y})(y_{t+\tau} - \bar{y})}{\sum_{t=1}^{n} (y_t - \bar{y})^2}
\]

Where \( \bar{y} \) = arithmetic mean

Note: Denominator expression is a measure of variance, \( s^2 \), because:

\[
s^2 = \frac{\sum_{t=1}^{n} (y_t - \bar{y})^2}{n - 1 - \tau} \quad \text{or} \quad \sum_{t=1}^{n} (y_t - \bar{y})^2 = (n - 1 - \tau)s^2
\]
### Time-Series Analysis

- Individual values for the time series in Fig. 3-14

<table>
<thead>
<tr>
<th>$t$</th>
<th>Month/Year</th>
<th>$y(t)$</th>
<th>$t$</th>
<th>Month/Year</th>
<th>$y(t)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>8/93</td>
<td>0.400</td>
<td>15</td>
<td>10/93</td>
<td>0.684</td>
</tr>
<tr>
<td>2</td>
<td>9/93</td>
<td>0.540</td>
<td>16</td>
<td>11/93</td>
<td>0.920</td>
</tr>
<tr>
<td>3</td>
<td>10/93</td>
<td>0.640</td>
<td>17</td>
<td>12/93</td>
<td>0.140</td>
</tr>
<tr>
<td>4</td>
<td>11/93</td>
<td>1.280</td>
<td>18</td>
<td>1/94</td>
<td>0.096</td>
</tr>
<tr>
<td>5</td>
<td>12/92</td>
<td>0.250</td>
<td>19</td>
<td>2/94</td>
<td>0.100</td>
</tr>
<tr>
<td>6</td>
<td>1/93</td>
<td>0.160</td>
<td>20</td>
<td>3/94</td>
<td>0.300</td>
</tr>
<tr>
<td>7</td>
<td>2/93</td>
<td>0.200</td>
<td>21</td>
<td>4/94</td>
<td>0.452</td>
</tr>
<tr>
<td>8</td>
<td>3/93</td>
<td>0.248</td>
<td>22</td>
<td>5/94</td>
<td>0.540</td>
</tr>
<tr>
<td>9</td>
<td>4/93</td>
<td>0.404</td>
<td>23</td>
<td>6/94</td>
<td>1.364</td>
</tr>
<tr>
<td>10</td>
<td>5/93</td>
<td>0.744</td>
<td>24</td>
<td>7/94</td>
<td>0.570</td>
</tr>
<tr>
<td>11</td>
<td>6/93</td>
<td>0.700</td>
<td>25</td>
<td>8/94</td>
<td>0.720</td>
</tr>
<tr>
<td>12</td>
<td>7/93</td>
<td>0.730</td>
<td>26</td>
<td>9/94</td>
<td>0.360</td>
</tr>
<tr>
<td>13</td>
<td>8/93</td>
<td>0.560</td>
<td>27</td>
<td>10/94</td>
<td>0.660</td>
</tr>
<tr>
<td>14</td>
<td>9/93</td>
<td>0.510</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Time-Series Analysis

Ex.: Time series of sulfur concentrations (Empirical autocorrelation)

Lag time $\tau = 12$ from the 27 individual data.

$\bar{y}$ mean = 0.530

$$r(12) = \frac{\sum_{i=1}^{12} (y_i - \bar{y})(y_{i+12} - \bar{y})}{\sum_{i=1}^{27} (y_i - \bar{y})^2}$$

$$= \frac{[(0.40 - 0.53)(0.56 - 0.53) + (0.54 - 0.53)(0.51 - 0.53)]}{(0.40 - 0.53)^2 + (0.54 - 0.53)^2 + \ldots + (0.66 - 0.53)^2} = 0.3$$

Note: The lower the calculated value, the more random the residuals are.
**Time-Series Analysis**

Cross-Correlation
- Correlation between two different time series, \( y(t) \) and \( x(t) \)
- Empirical cross-correlation:

\[
 r_{xy}(\tau) = \frac{\sum_{t=1}^{n-|\tau|} x_t y_{t-\tau}}{\sqrt{\sum_{t=1}^{n} x_t^2} \sqrt{\sum_{t=1}^{n} y_t^2}}
\]

**Time-Series Analysis**

Random series with drift
- Deviations from the stationary behavior of the time series, presence drift in the signal.

![Absorbance (A.U.) vs Time (s)](image)
Time-Series Analysis

Drift in time series

Fig 11. (a) Time series with a periodicity (F = 0.125). (b) Autocorrelation function of 11.

Fig 11. (c) Time series with a periodicity (F = 0.125). (d) Autocorrelation function of 11.

Fig 11. (e) Time series with a periodicity (F = 0.125). (f) Autocorrelation function of 11.