# ASME

# DTOG 2024

**Digital Horizons: Energizing Transformation in Oil, Gas, and Beyond**

Hyatt Regency Houston West
Houston, TX

# Digital Transformation of Oil & Gas and Beyond

## Short Course: Design of Experiments and Machine Learning

ASME DTOG 2024
Digital Horizons: Energizing Transformation in Oil, Gas, and Beyond

# Outline of Short Course

**Data Cleansing**

Raw data is messy due to sensor noise and dropped data. This section presents data cleansing due to sensor noise.

**Design of Experiments**

Large databases make process understanding challenging. This section offers a structured approach to identify relationships among measurements for optimal analysis.

**Machine Learning**

This section is and engineering efficiency approach to obtain predictive models and algorithms that learns the data enabling make real-time decisions.

# Digital Transformation of Oil & Gas and Beyond

## Data Cleansing

# About the Instructors

**Matthew Franchek, Ph.D.**

**Model Based Solutions, LLC**
Owner
**University of Houston**
Mechanical Engineering
Professor (2002-Present)

**Malek Rekik, Ph.D. Candidate**

**University of Houston**
Mechanical Engineering

**ChampionX**
Data Scientists

**Amal Chebbi, Ph.D.**

**University of Houston**
Mechanical Engineering
Data Scientist

# Learning Outcomes

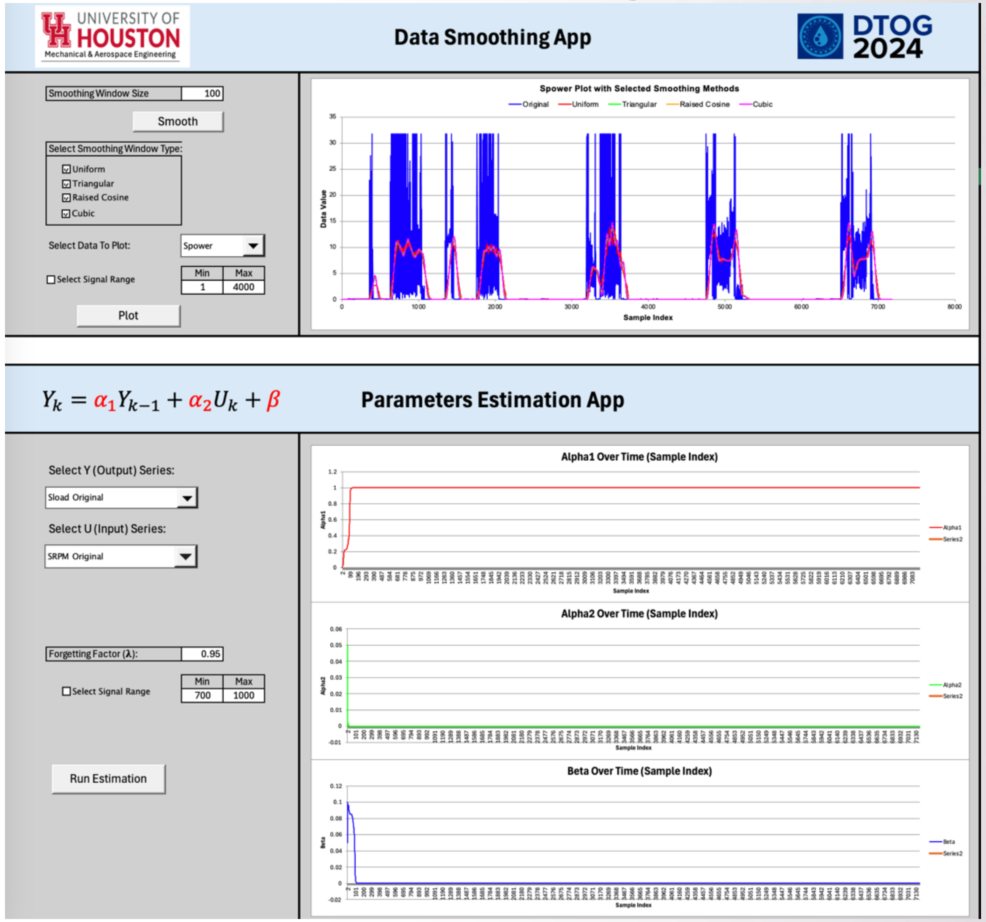
At the completion of this data cleansing learning module, you will **(1)** understand the cleansing algorithm, **(2)** learn the GUI driven cleansing tool provided to you, and **(3)** practice regression analysis on cleansed data.

# Approach to Learning Outcomes

## Data Cleansing

| | |
|---|---|
| Supporting Mathematics | Smoothing Algorithms |
| Linear Algebra | Programming Syntax & Logic |
| Using Linear Algebra in Coding | Analyzing Case Studies |

# A Data Cleansing Tool for You

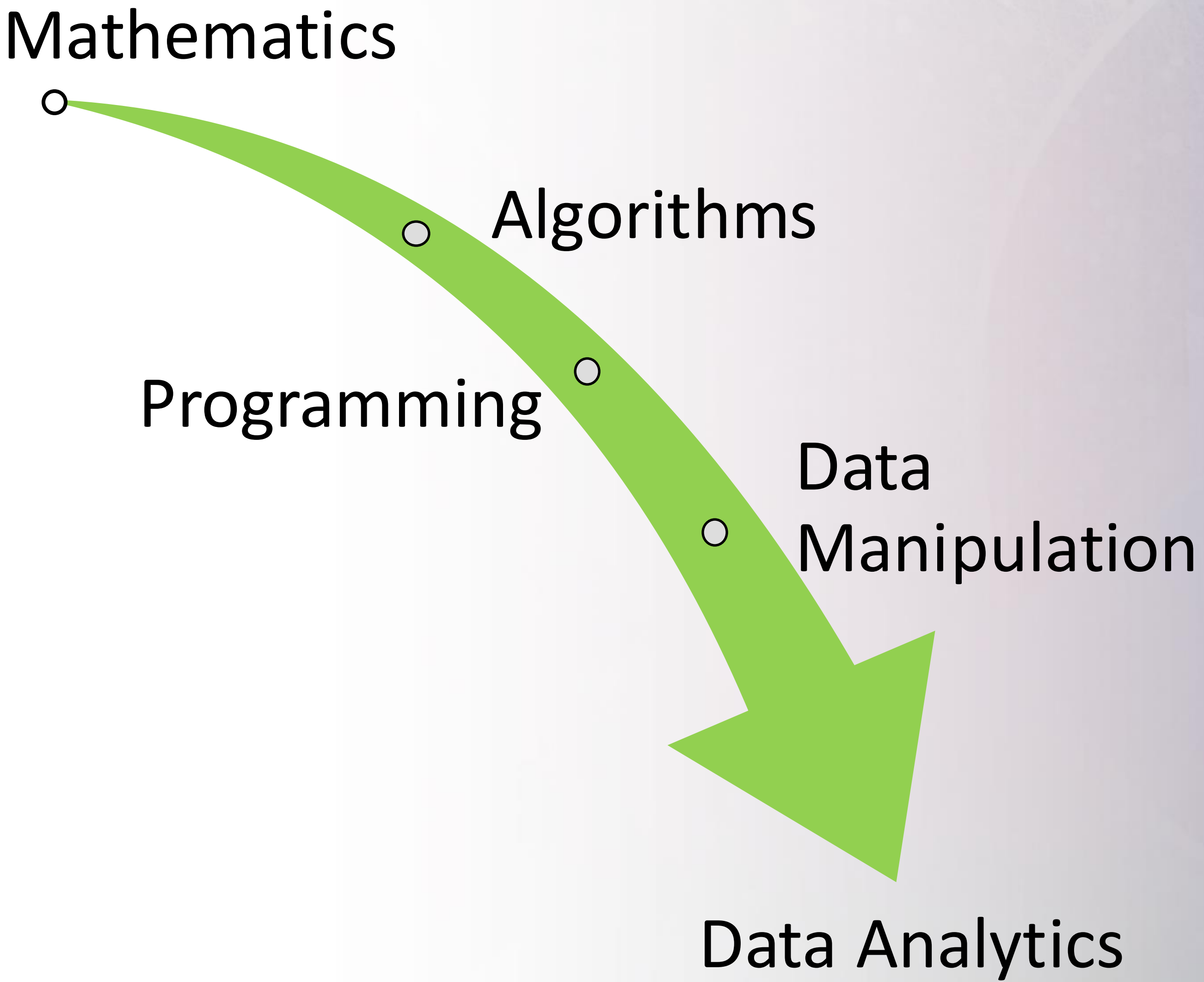



## University of Houston Data Cleansing and Regression Analysis Tool

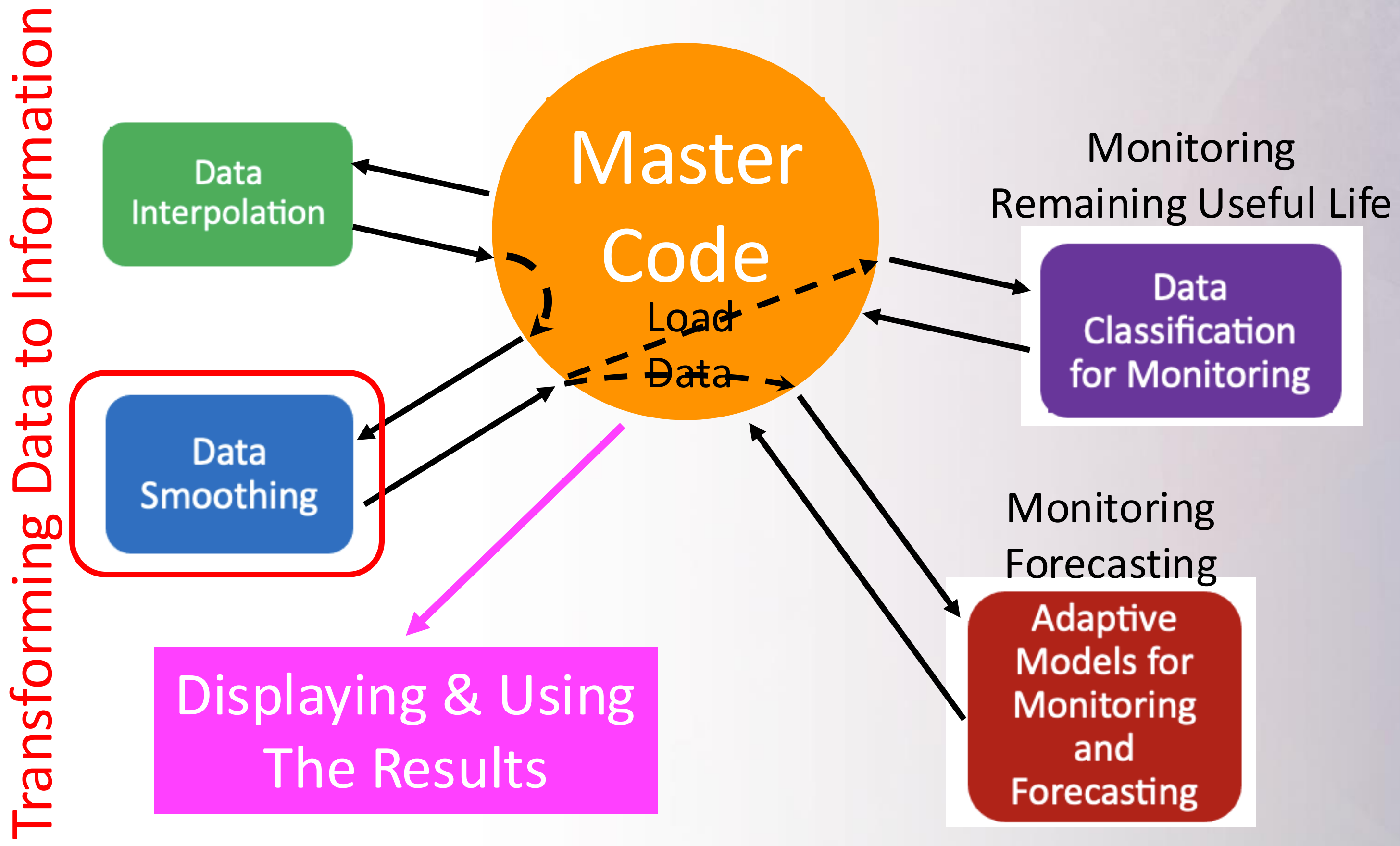
*Developers: Malek Rekik and Amal Chebbi*

ASME DTOG 2024
Digital Horizons: Energizing Transformation in Oil, Gas, and Beyond

# You are Learning a Process

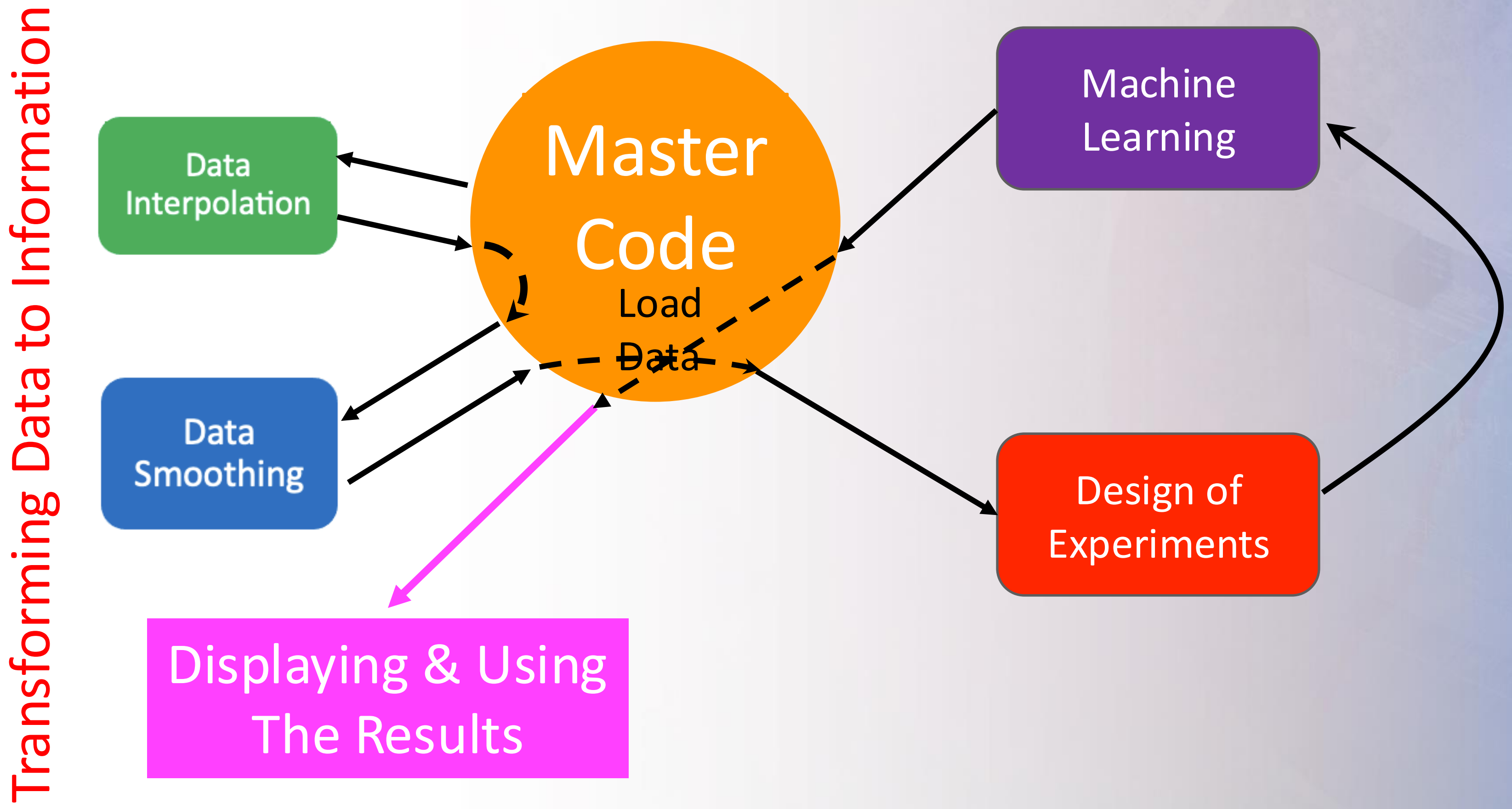


# The Data Analytics Ecosystem



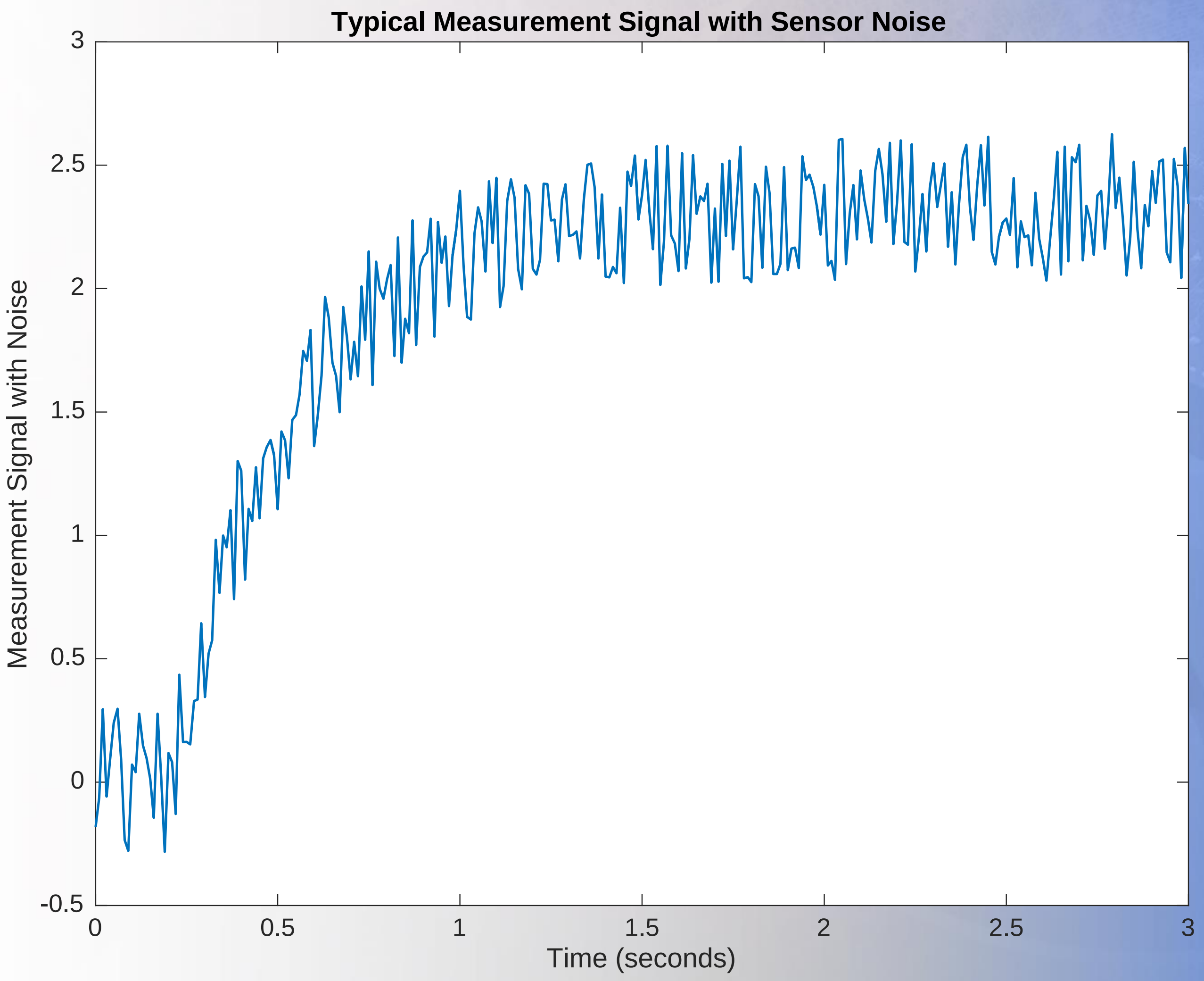
# The AI Driven Data Analytics Ecosystem





# Data Cleansing Algorithms and Processes

# Sensor Measurements with Noise



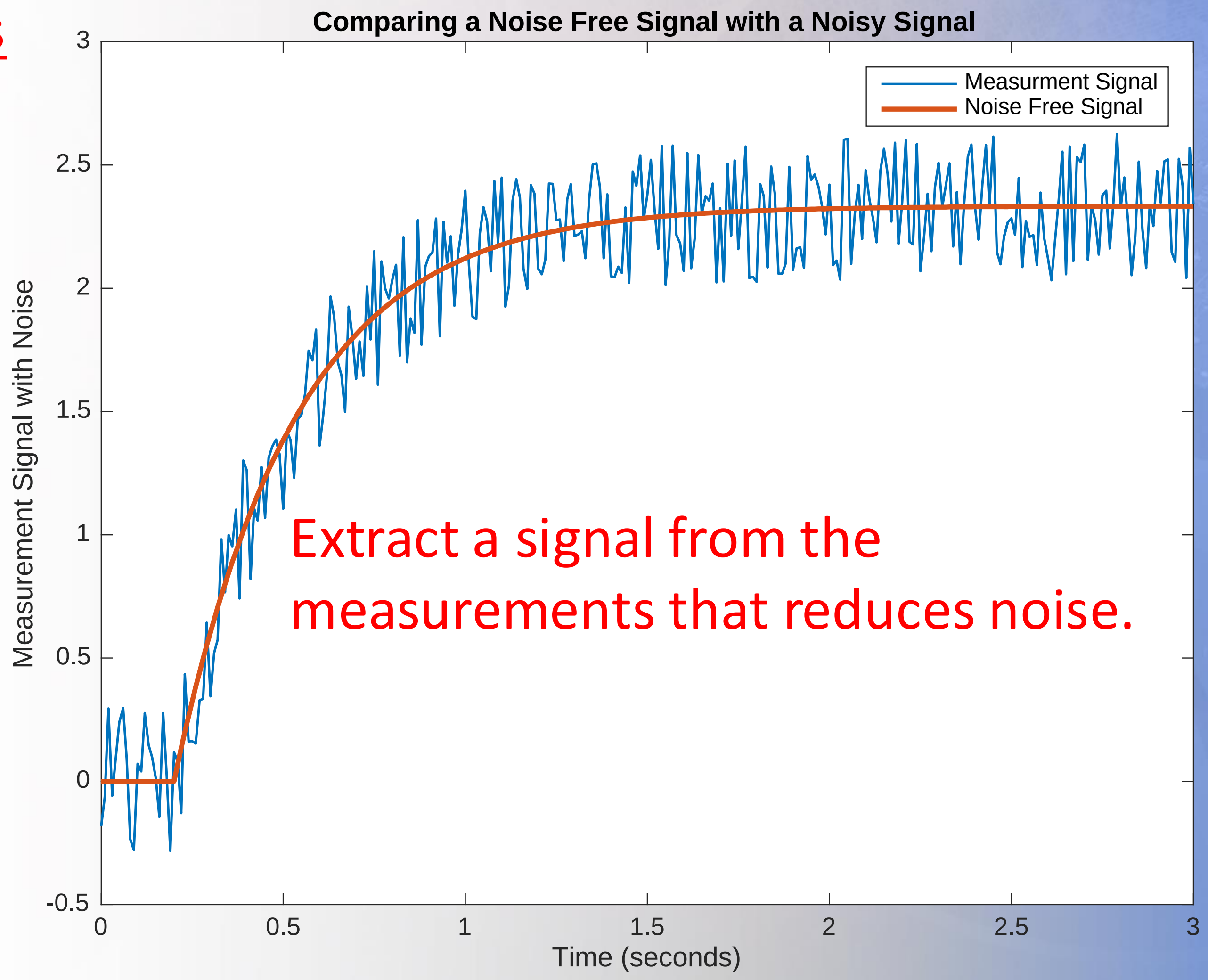
ASME DTOG 2024
Digital Horizons: Energizing Transformation in Oil, Gas, and Beyond

# Goal of Data Smoothing

## Initial Challenges in Data Analytics

1. Sensor Noise   Smoothing

2. Dropped Data   Interpolation




Extract a signal from the measurements that reduces noise.

# Smoothing Method of Solution

## Weighted Central Data Smoothing

Most versatile smoothing algorithm.

ASME DTOG 2024
Digital Horizons: Energizing Transformation in Oil, Gas, and Beyond

# Weighted Central Smoothing Algorithm

Smoothed Value for $y(i)$

*npts* = Odd number of points used to calculate $\bar{y}(i)$

Weighting of $y_{i+k}$ when calculating $\bar{y}_i$

Data at sample $(i + k)$

$$\bar{y}_i = \frac{\sum_{k=-(npts-1)/2}^{(npts-1)/2} W_k * y_{i+k}}{\sum_{k=-(npts-1)/2}^{(npts-1)/2} W_k}$$

# Weighted Central Data Smoothing

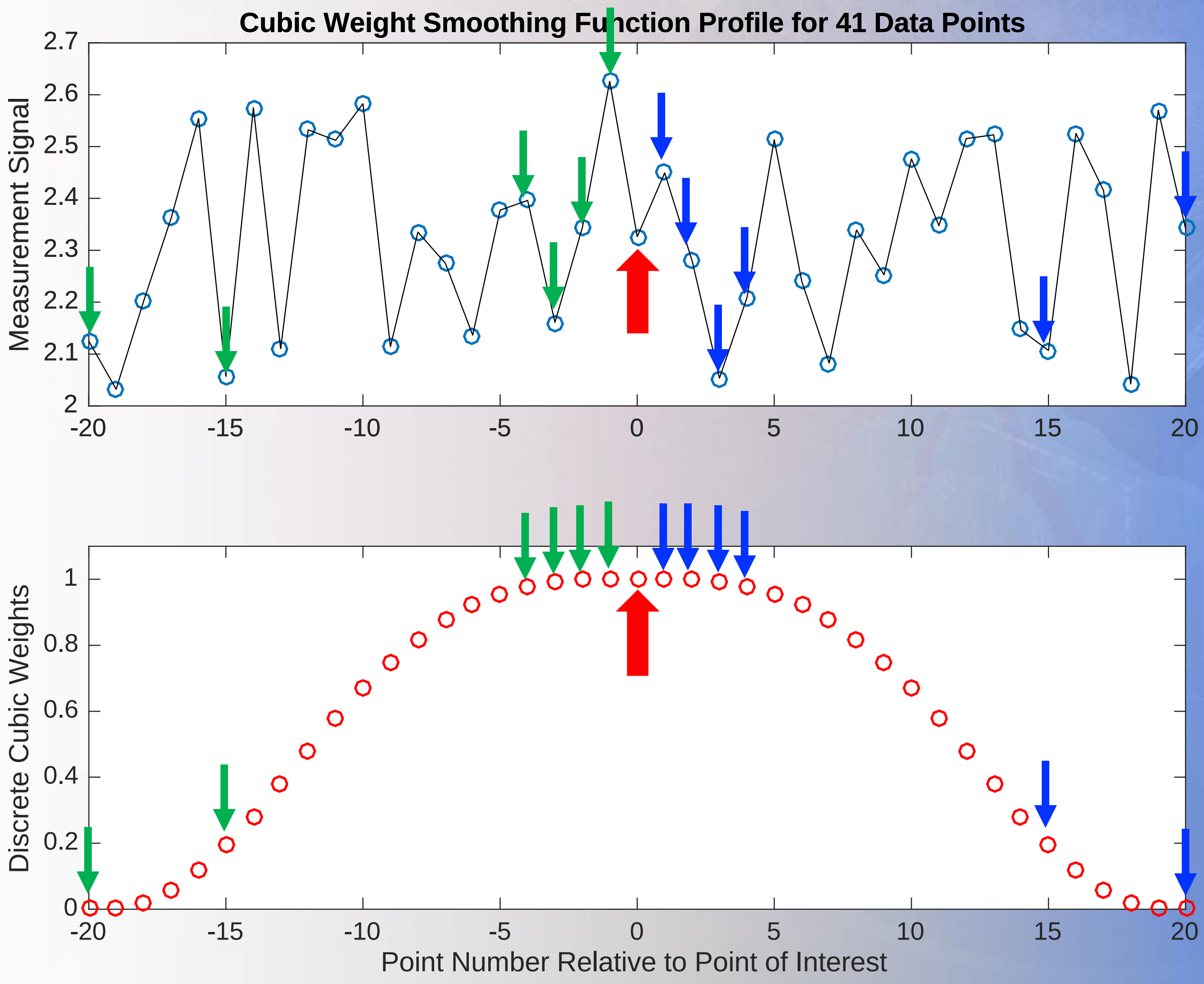
Expanded Version of $\bar{y}_i = \dfrac{\sum_{k=-(npts-1)/2}^{(npts-1)/2} W_k * y_{i+k}}{\sum_{k=-(npts-1)/2}^{(npts-1)/2} W_k}$ for *npts* = 5

$$\bar{y}_i = \frac{\underbrace{W_{-2}\, y_{i-2} + W_{-1}\, y_{i-1}}_{\text{Back samples used to calculate } \bar{y}_i} + W_0\, y_i + \underbrace{W_1\, y_{i+1} + W_2\, y_{i+2}}_{\text{Forward samples used to calculate } \bar{y}_i}}{W_{-2} + W_{-1} + W_0 + W_1 + W_2}$$

# Weighted Central Data Smoothing*

* Example Window Size = 41 pts

$$\bar{y}_0 = \frac{\left[\begin{array}{l} W_0 * y_0 + W_{-1} * y_{-1} + W_{-2} * y_{-2} \\ + W_{-3} * y_{-3} + W_{-4} * y_{-4} \cdots + W_{-15} * y_{-15} \\ \cdots + W_{-20} * y_{-20} + W_1 * y_1 + W_2 * y_2 \\ + W_3 * y_3 + W_4 * y_4 \cdots + W_{15} * y_{15} \\ \cdots + W_{20} * y_{20} \end{array}\right]}{W_{-20} + \cdots + W_0 + \cdots + W_{20}}$$

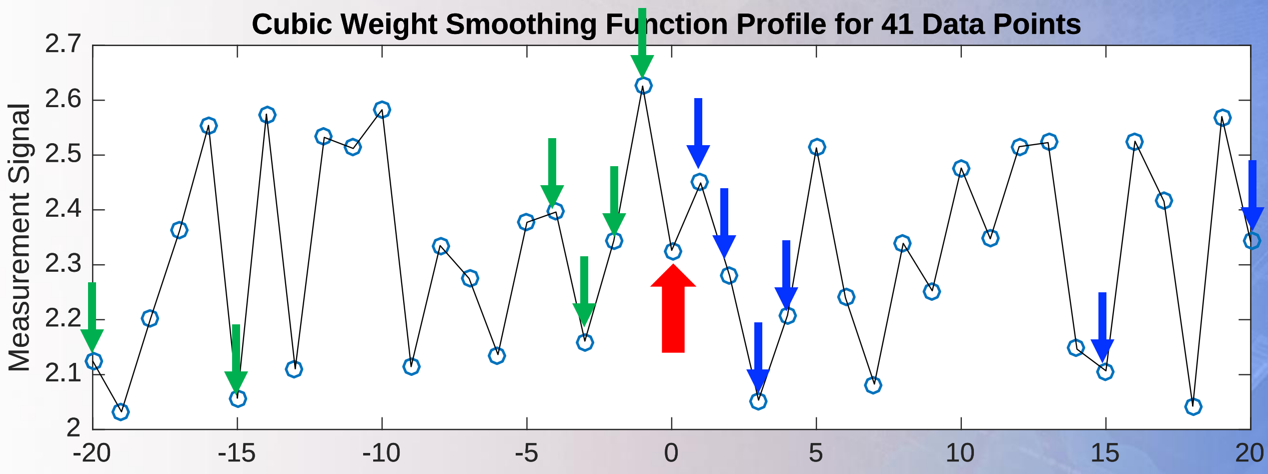


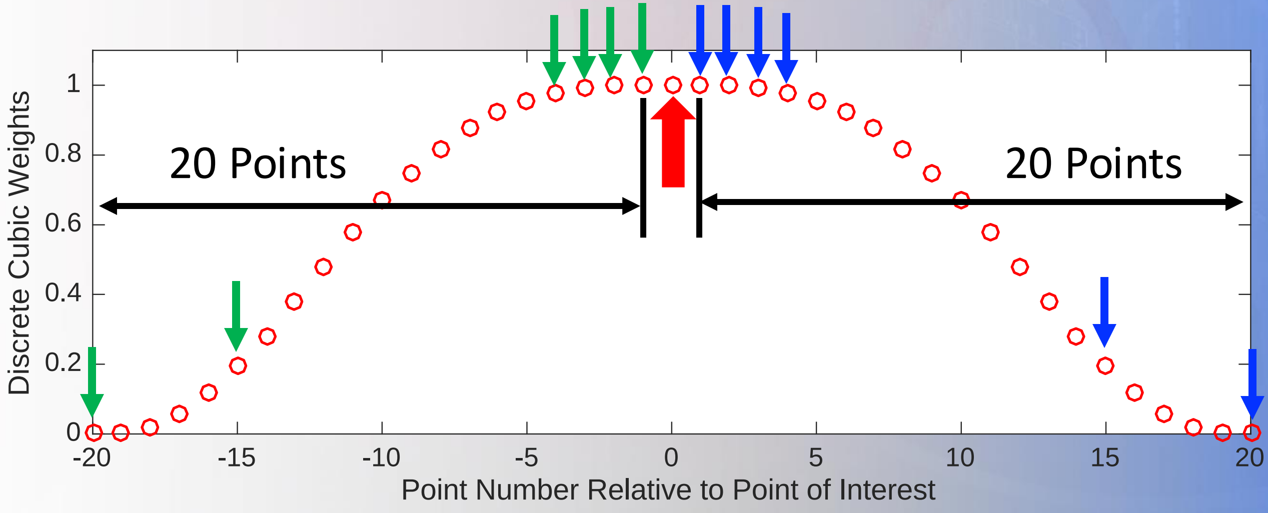
# Weighted Central Data Smoothing*

* Example Window Size = 41 pts

| | | |
|---|---|---|
| | Center Point | = 1 |
| | Back Points | = 20 |
| + | Forward Points | = 20 |
| | Number of Points | = 41 |

$$\bar{y}_0 = \frac{\left[\begin{array}{l} W_0 * y_0 + W_{-1} * y_{-1} + W_{-2} * y_{-2} \\ + W_{-3} * y_{-3} + W_{-4} * y_{-4} \cdots + W_{-15} * y_{-15} \\ \cdots + W_{-20} * y_{-20} + W_1 * y_1 + W_2 * y_2 \\ + W_3 * y_3 + W_4 * y_4 \cdots + W_{15} * y_{15} \\ \cdots + W_{20} * y_{20} \end{array}\right]}{W_{-20} + \cdots + W_0 + \cdots + W_{20}}$$

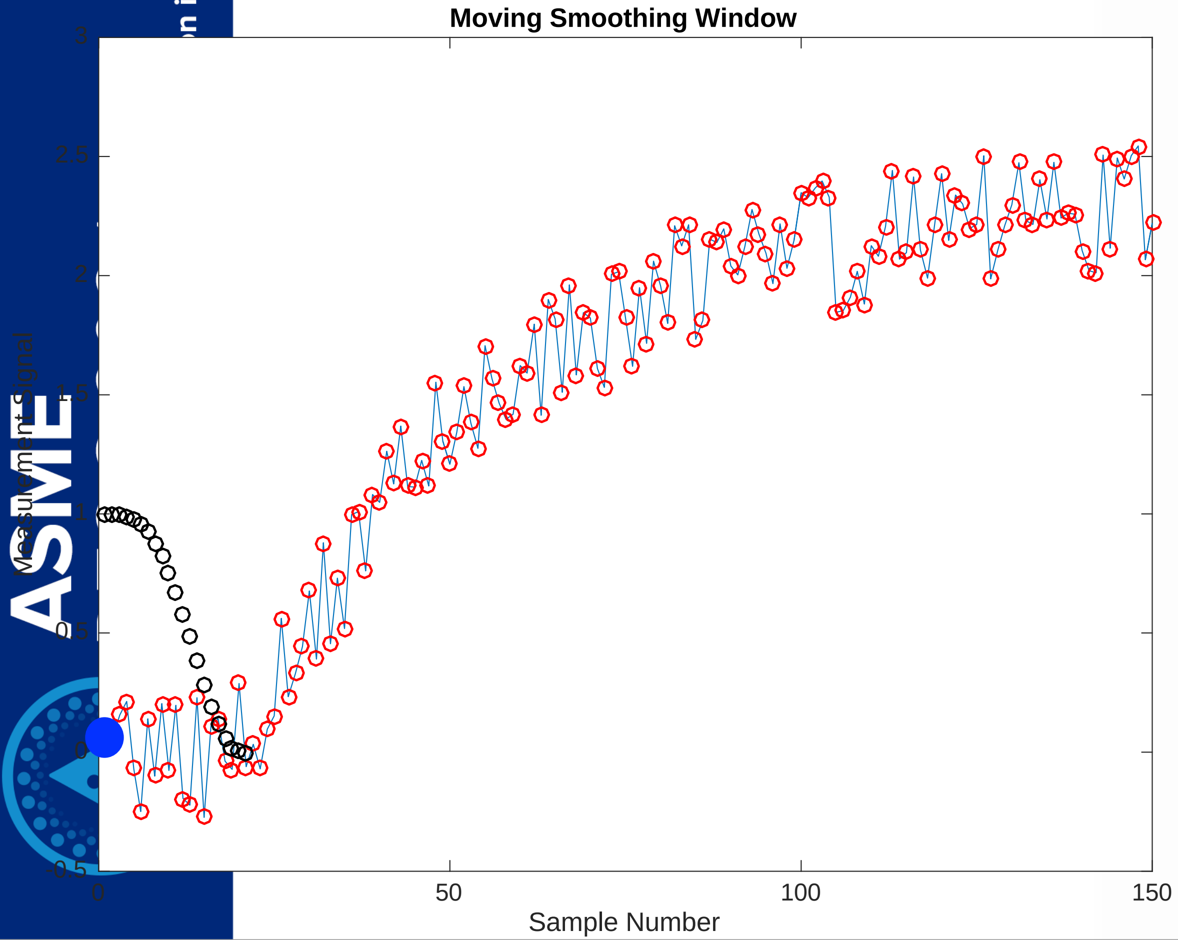


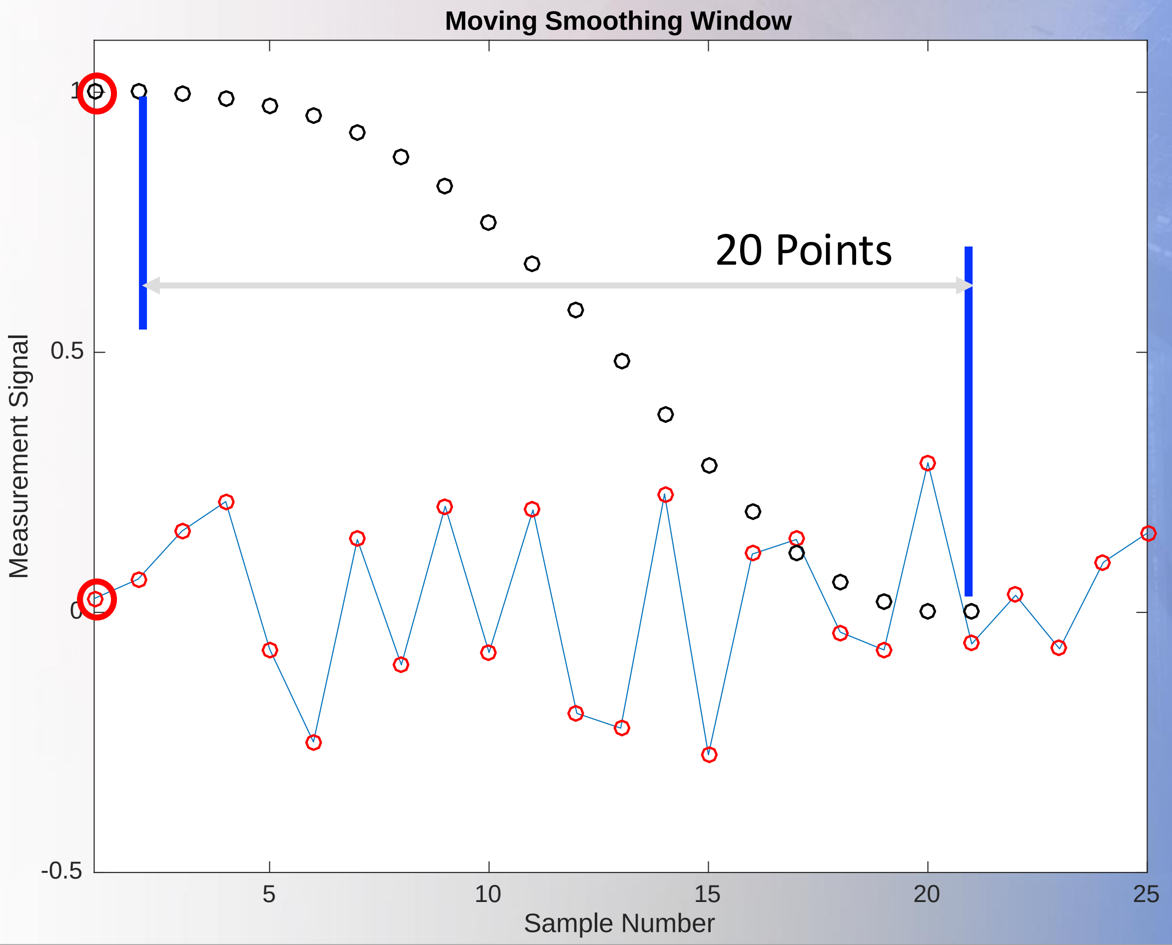



# Smoothing Window at Sample 1

npts = 41

$$\bar{y}_1 = \frac{\boxed{W_1\, y_1} + \boxed{W_2\, y_2 + \cdots + W_{20}\, y_{20} + W_{21}\, y_{21}}}{W_1 + W_2 + \cdots + W_{20} + W_{21}}$$

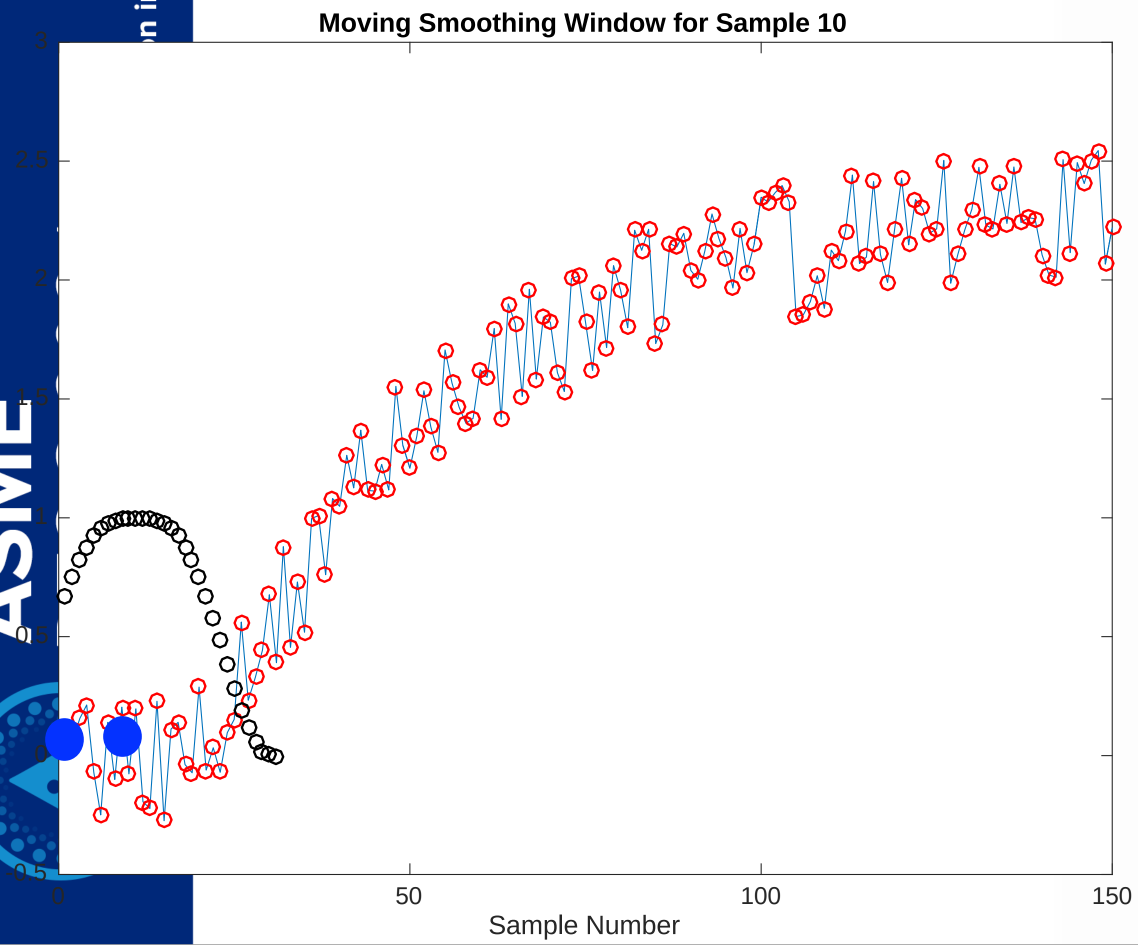


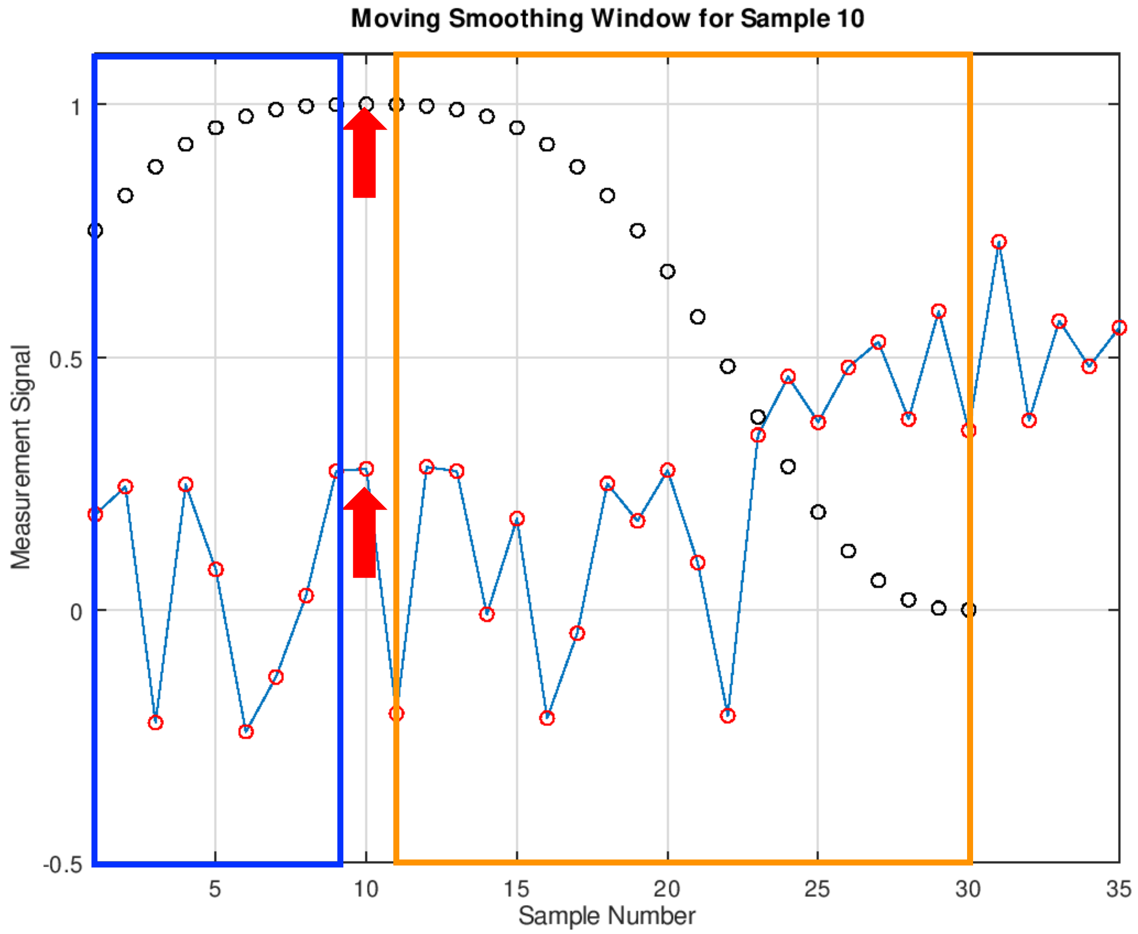



# As the Window Moves to Sample 10

npts = 41

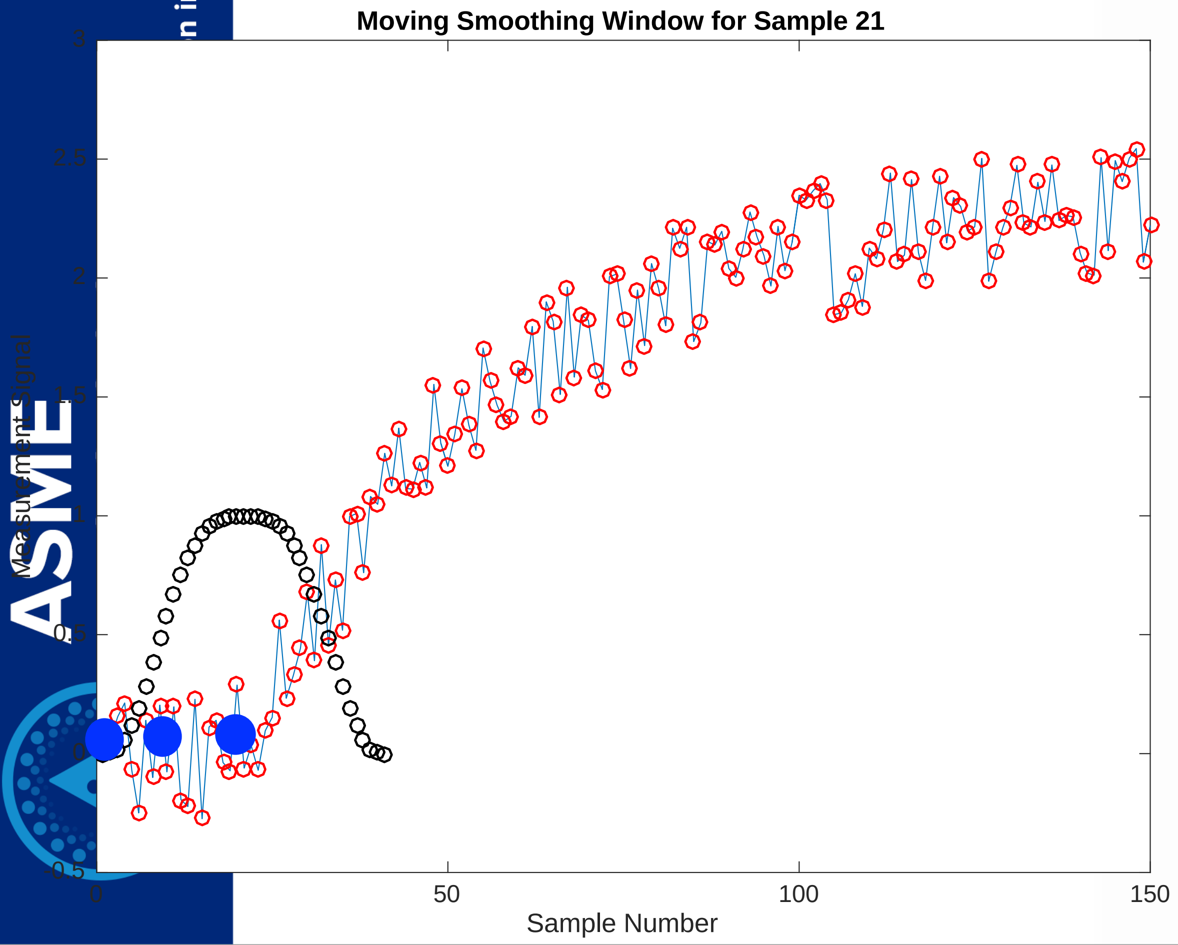
$$\bar{y}_{10} = \frac{\boxed{W_1\, y_1 + W_2\, y_2 + \cdots W_9\, y_9} + \boxed{W_{10} y_{10}} + \boxed{W_{11} y_{11} \cdots + W_{29}\, y_{29} + W_{30}\, y_{30}}}{W_1 + W_2 + \cdots + W_{29} + W_{30}}$$

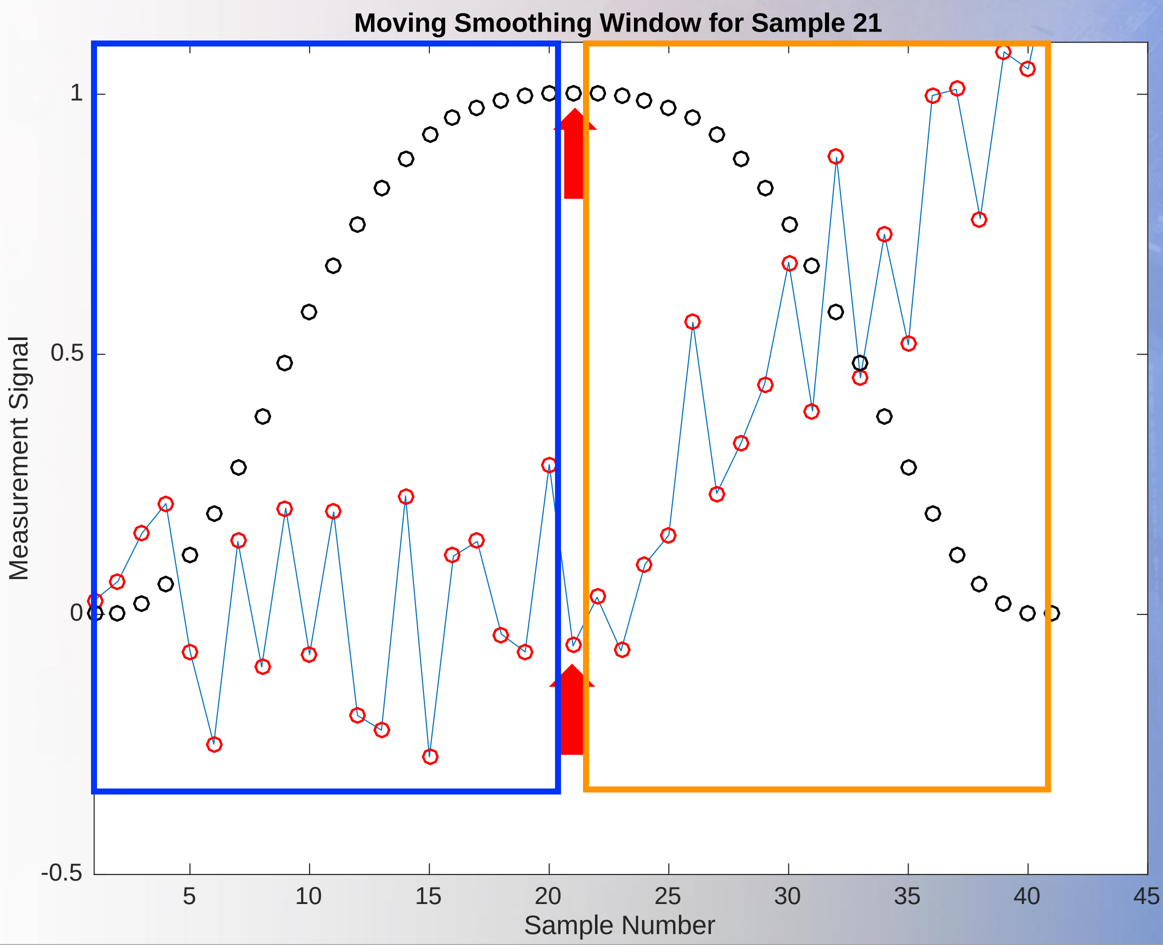


# As the Window Moves to Sample 21

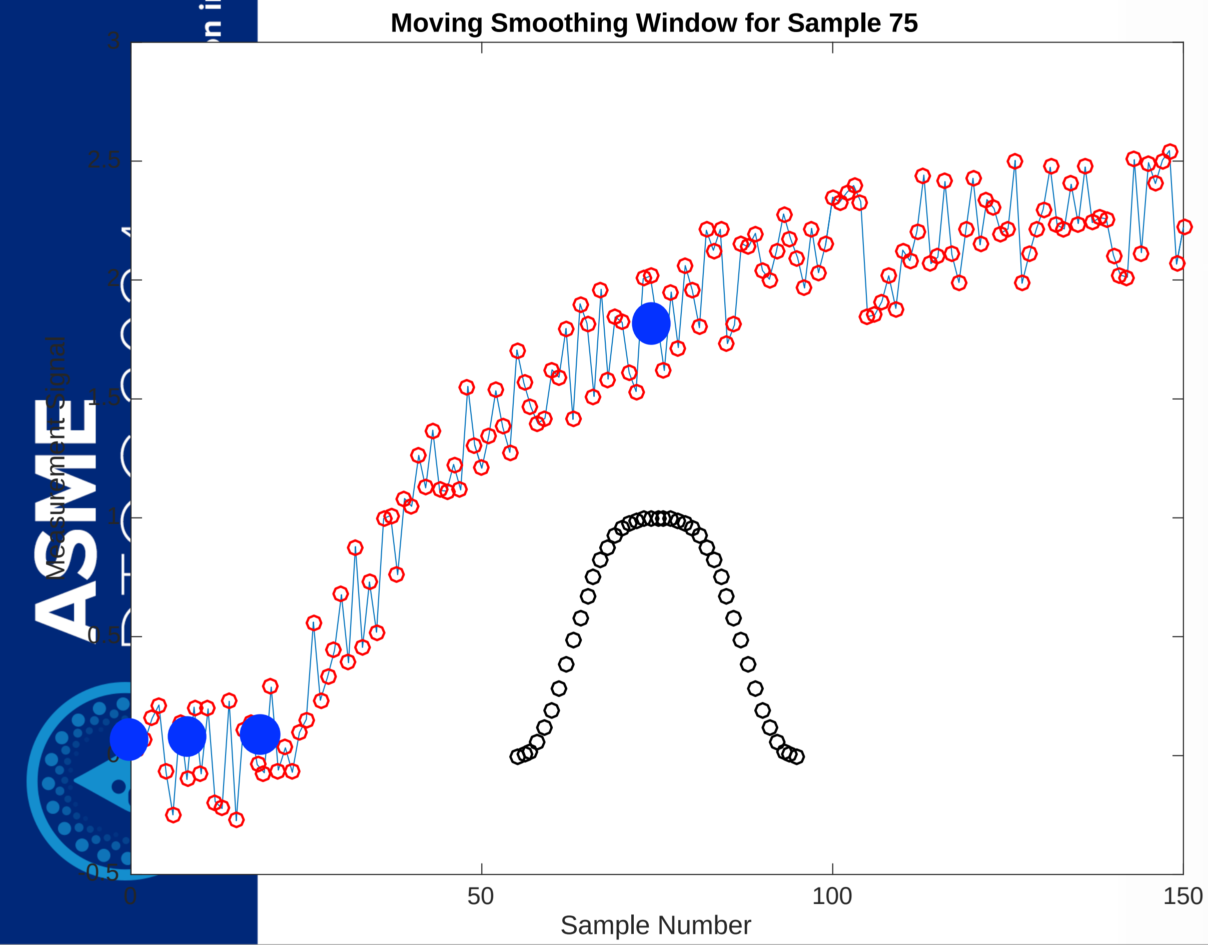
$$\bar{y}_{21} = \frac{\boxed{W_1\, y_1 + W_2\, y_2 + \cdots} \boxed{W_{21} y_{21}} \boxed{+ \cdots + W_{40}\, y_{40} + W_{41}\, y_{41}}}{W_1 + W_2 + \cdots + W_{40} + W_{41}}$$

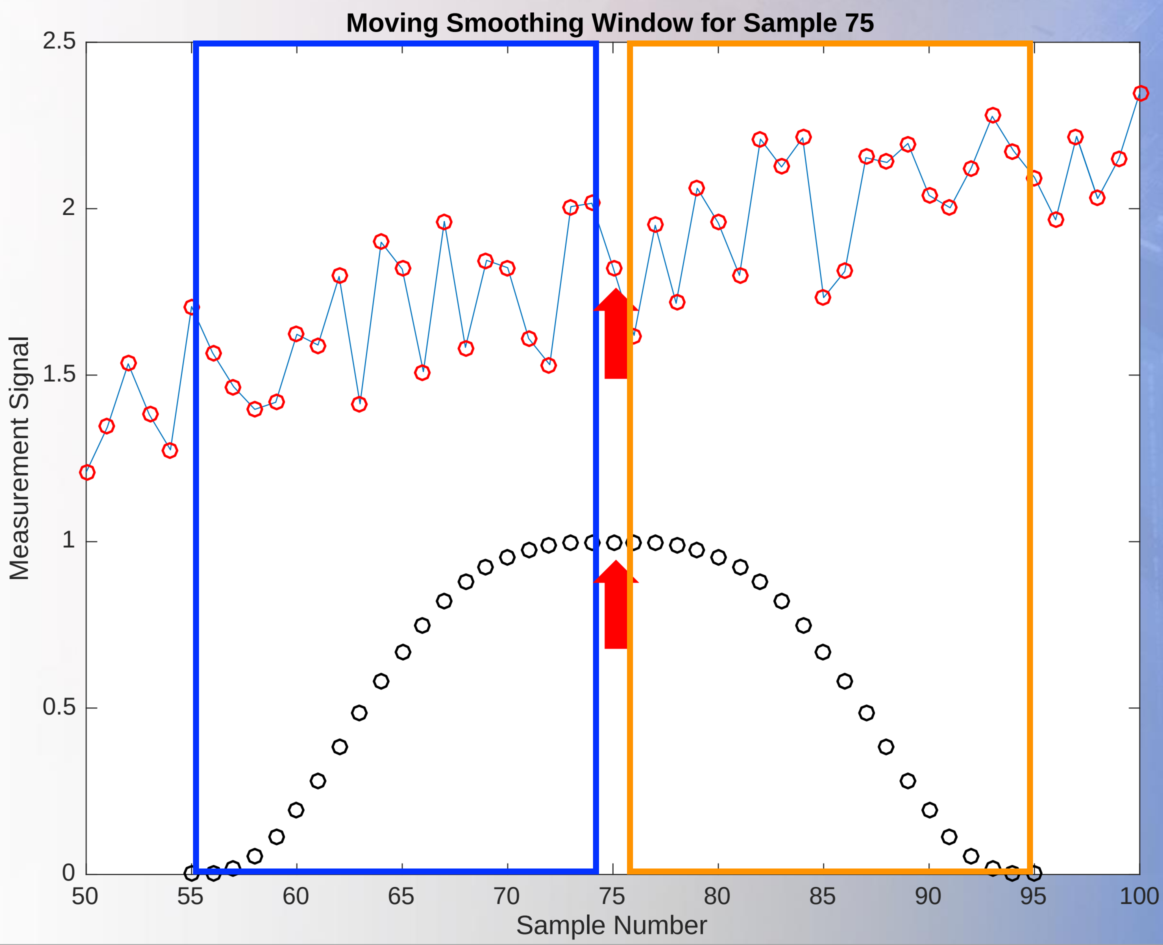


# As the Window Moves to Sample 75

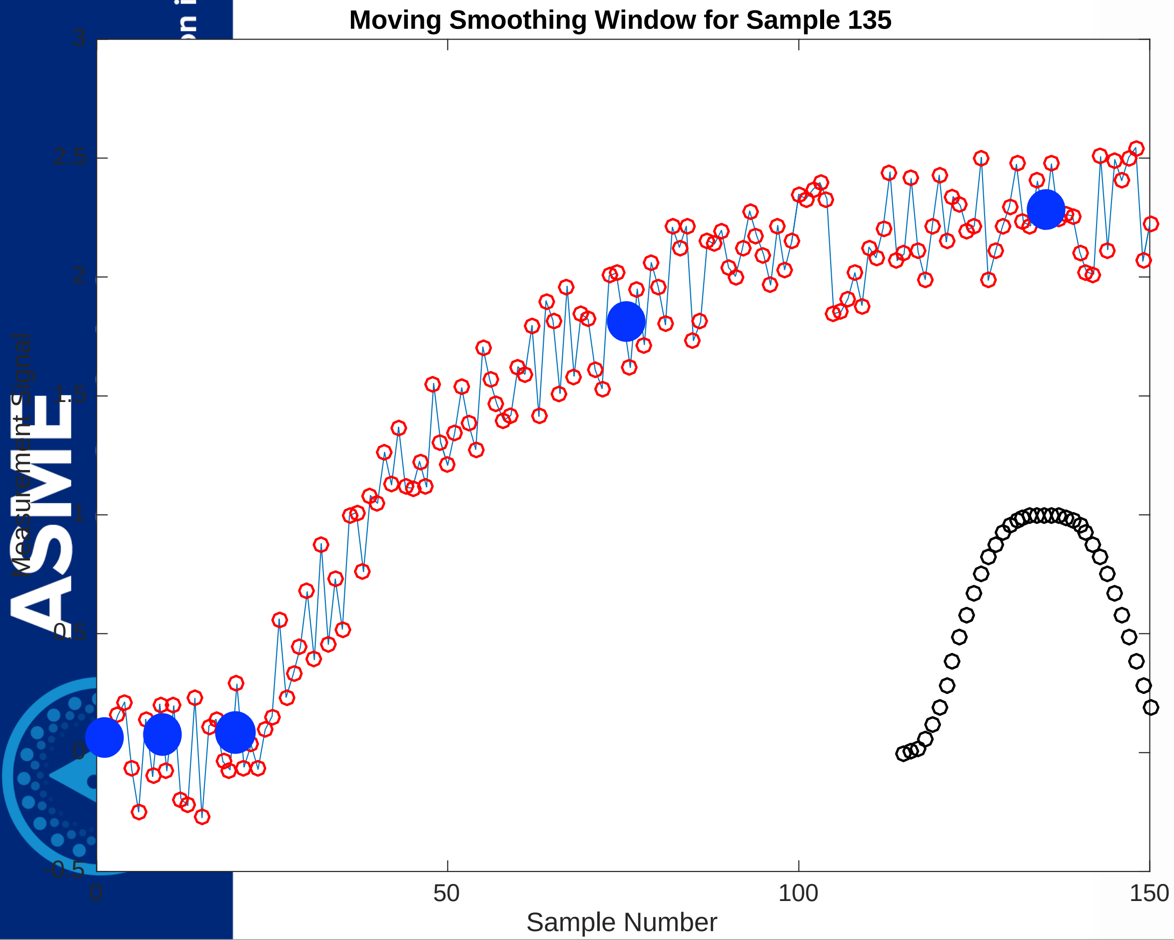
$$\bar{y}_{75} = \frac{\boxed{W_{55}\, y_{55} + W_{56}\, y_{56} + \cdots} + \boxed{W_{75} y_{75}} \boxed{+ \cdots + W_{94}\, y_{94} + W_{95}\, y_{95}}}{W_{55} + W_{56} + \cdots + W_{94} + W_{95}}$$

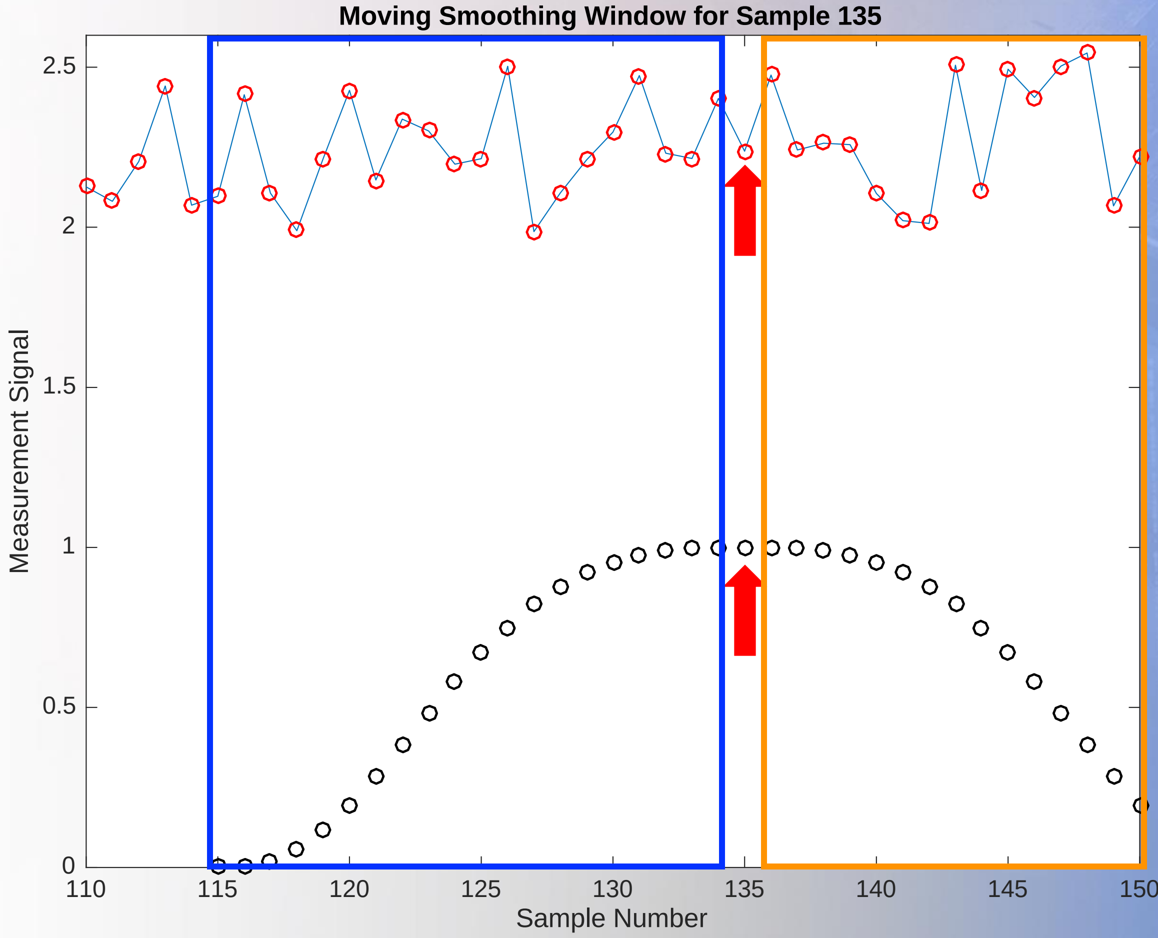


# As the Window Moves to Sample 135

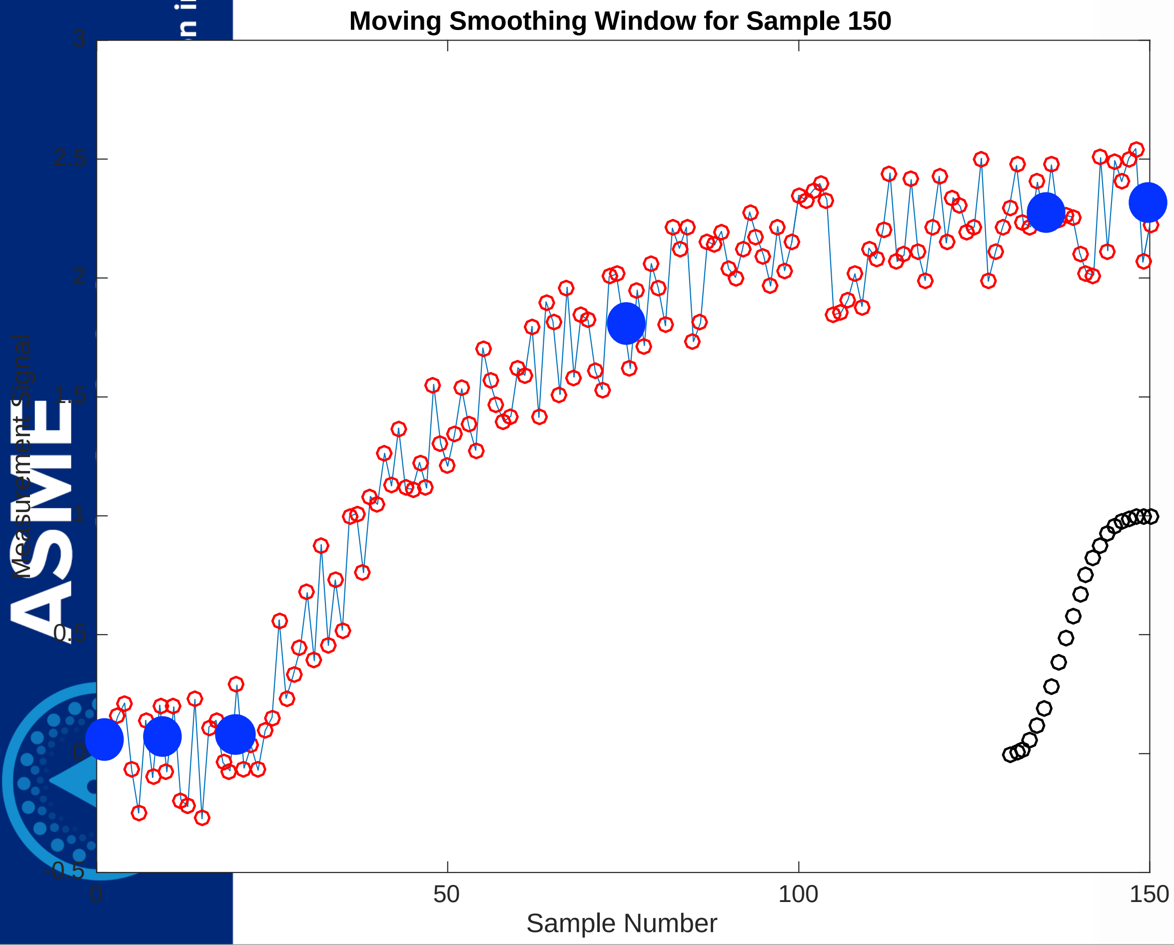
$$\bar{y}_{135} = \frac{\boxed{W_{115}\, y_{115} + W_{116}\, y_{116} + \cdots} + \boxed{W_{135} y_{135}} \boxed{+ \cdots + W_{149}\, y_{149} + W_{150}\, y_{150}}}{W_{115} + W_{116} + \cdots + W_{149} + W_{150}}$$

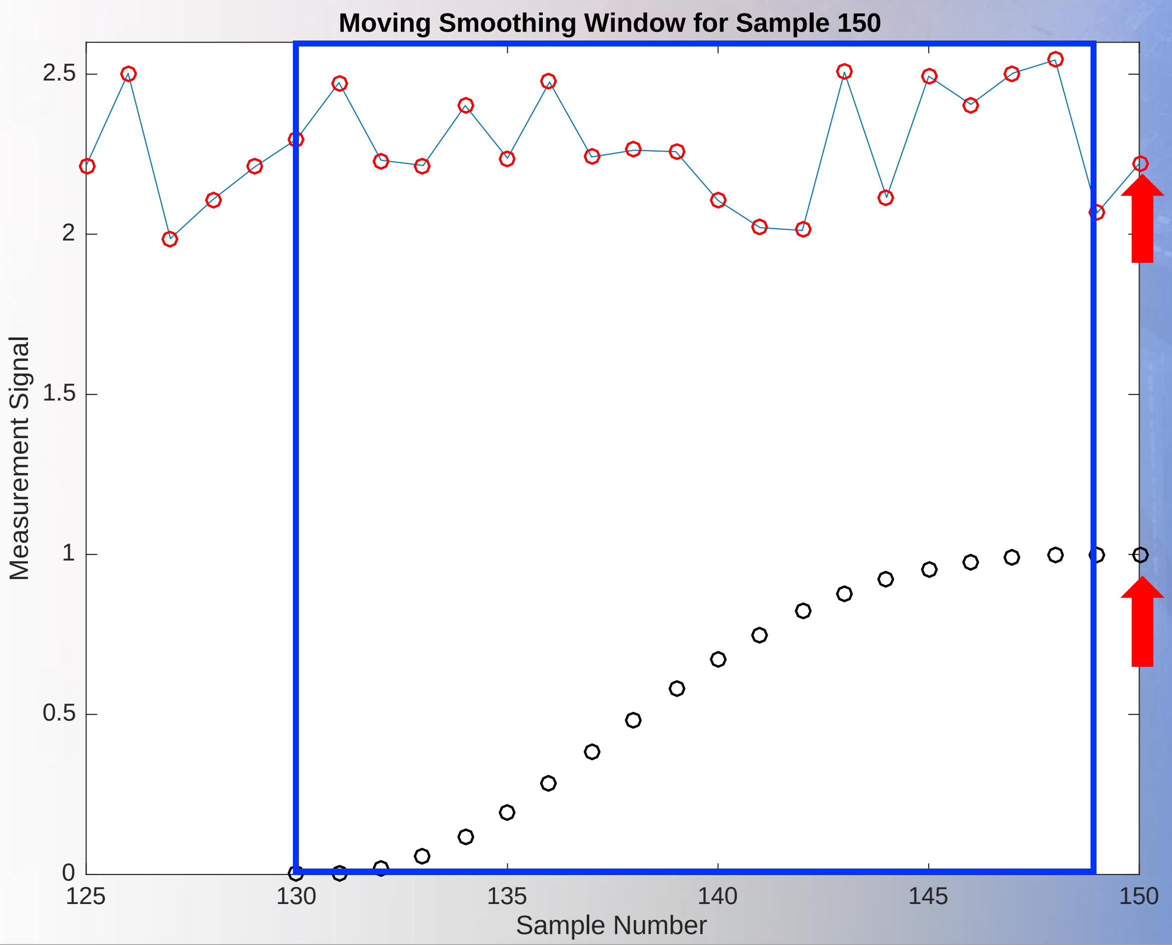


# As the Window Moves to Sample 150

$$\bar{y}_{150} = \frac{\boxed{W_{130}\, y_{130} + W_{131}\, y_{131} + \cdots + W_{149}\, y_{149}} + \boxed{W_{150}\, y_{150}}}{W_{130} + W_{131} + \cdots + W_{149} + W_{150}}$$



# Common Weights for Central Data Smoothing

- Cubic
- Triangular
- Raised Cosine
- Uniform Weights

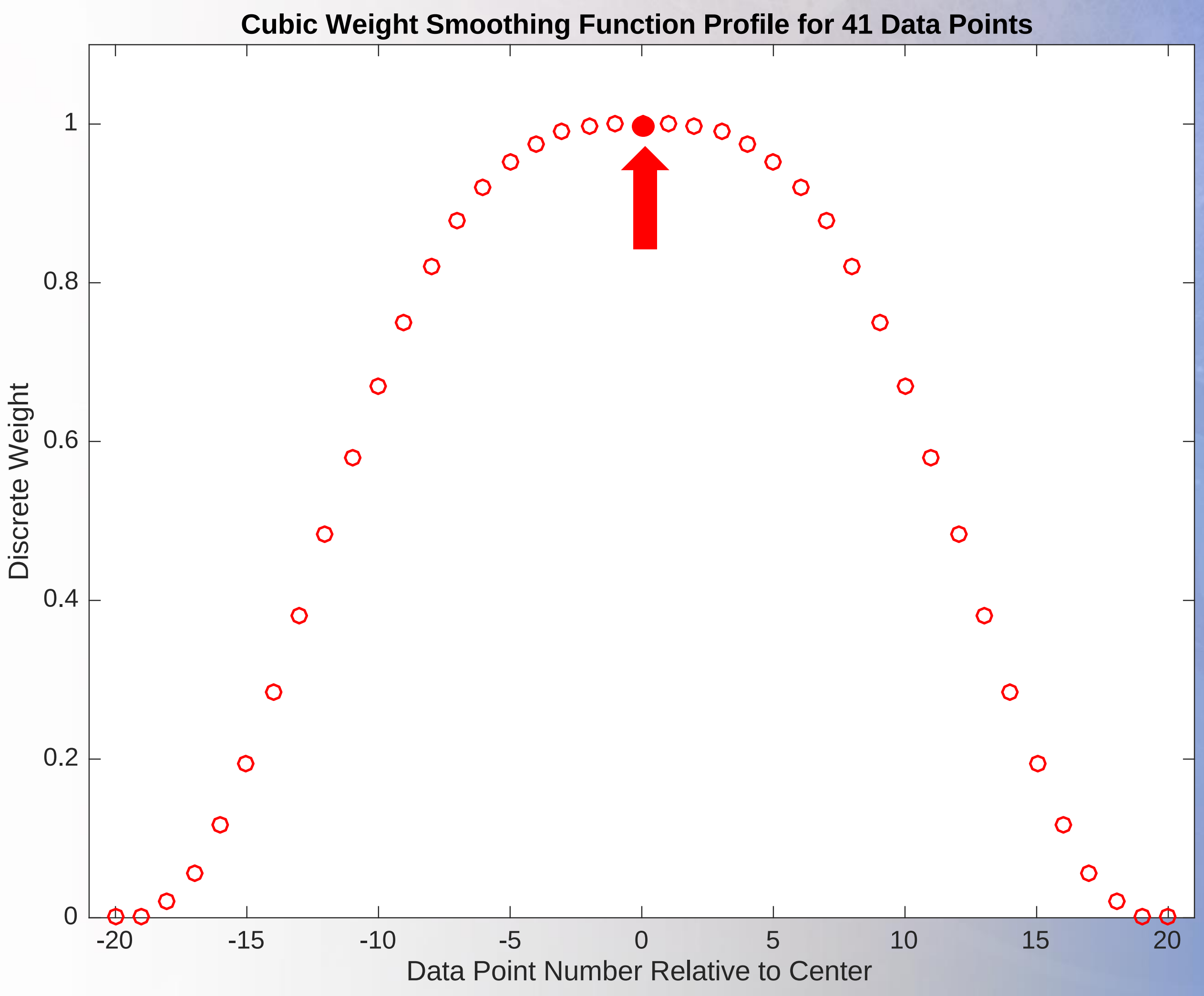
# Cubic Weights for Central Data Smoothing

$$W_i = \left[1 - \left(\frac{abs(i)}{Side}\right)^3\right]^3$$

$npts$ = Odd Number
= Number of Points

$$Side = \frac{npts - 1}{2}$$

$$i \in [-Side, \underbrace{Side}_{\text{Number of Data Points Per Side}}]$$



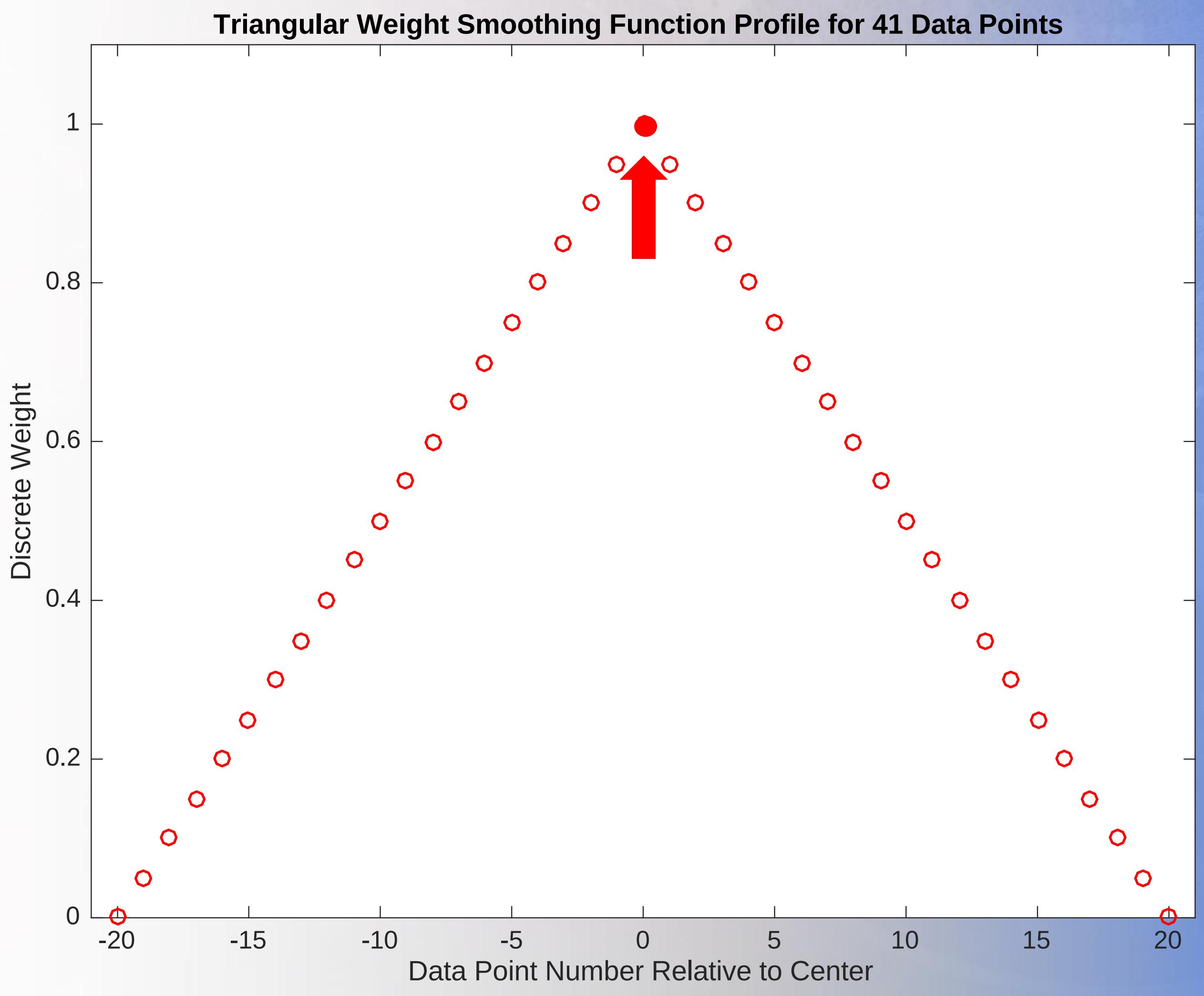
# Triangular Weights for Data Smoothing

$$\boxed{W_i = 1 - \left(\frac{abs(i)}{Side}\right)}$$

$npts$ = Odd Number
= Number of Points

$$Side = \frac{npts - 1}{2}$$

$$i \in [-Side, \underbrace{Side}_{\text{Number of Data Points Per Side}}]$$



# Raised Cosine Weights for Data Smoothing

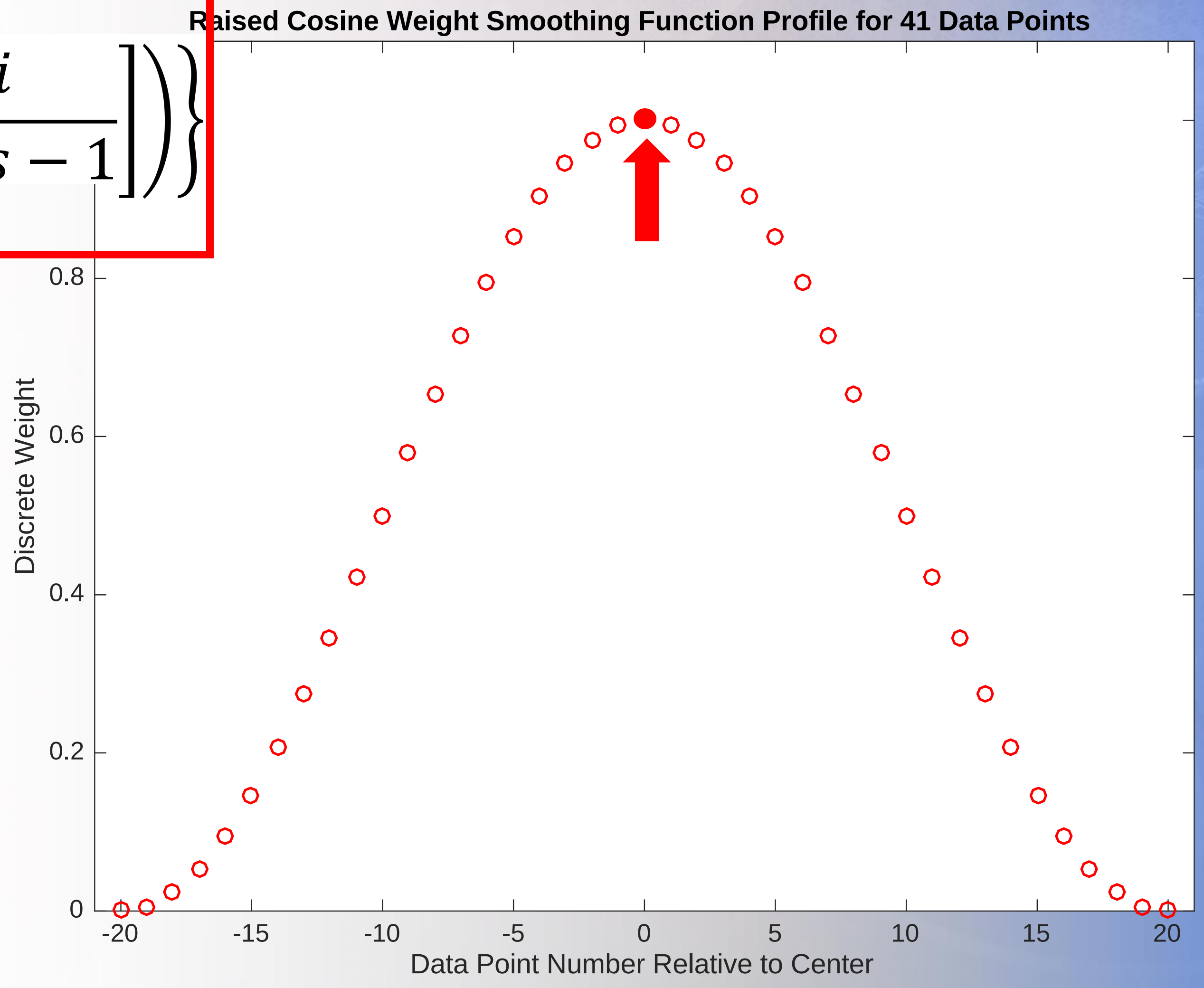
$$W_i = \frac{1}{2}\left\{1 - \cos\left(2\,\pi\left[\frac{1}{2} + \frac{i}{npts-1}\right]\right)\right\}$$

$npts$ = Odd Number
= Number of Points

$$Side = \frac{npts-1}{2}$$

$$i \in [-\underbrace{Side, Side}]$$

Number of Data Points Per Side



# Uniform Weights for Central Data Smoothing

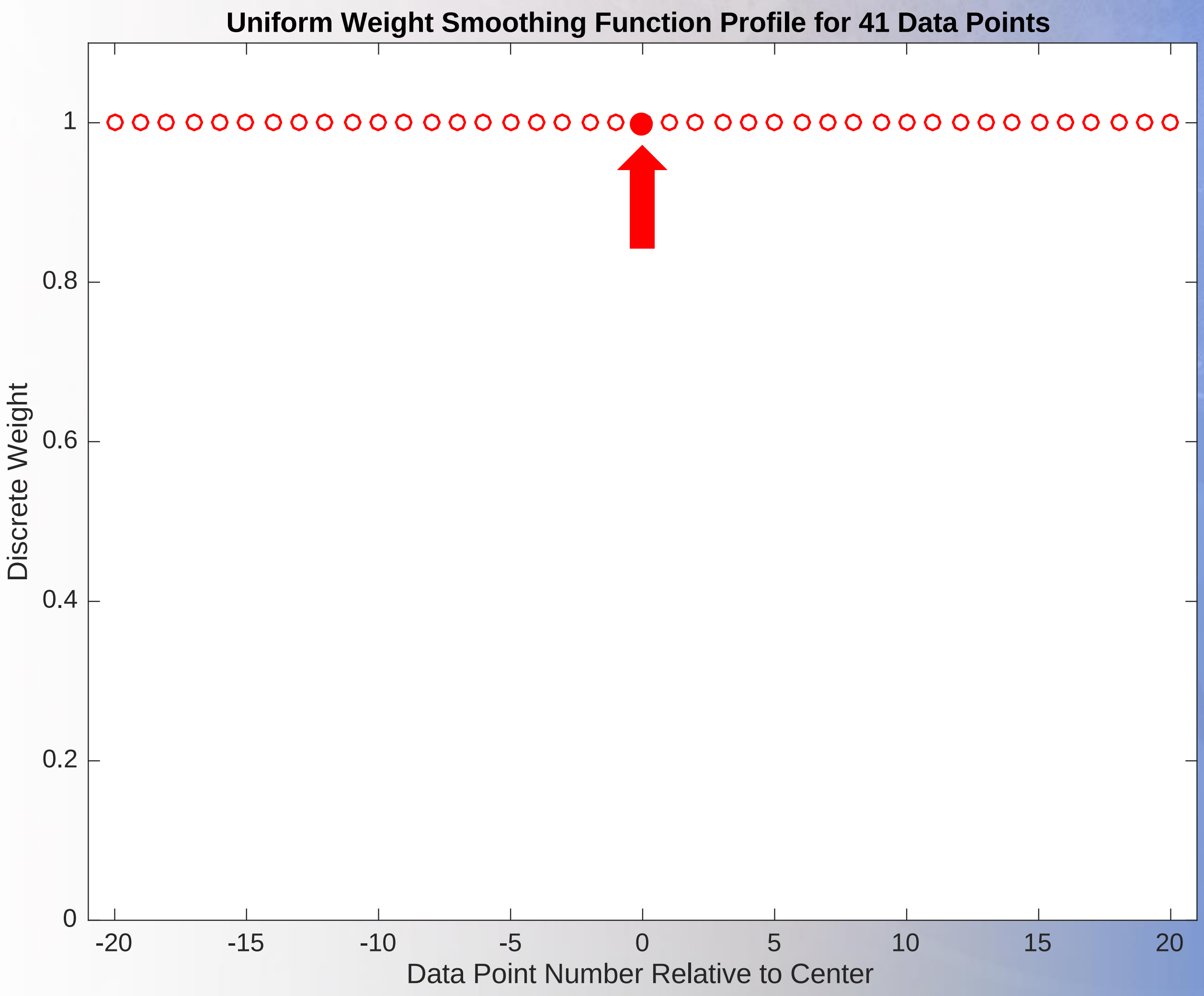
$$W_i = 1$$

(This is the mean value calculation)

$npts$ = Odd Number
= Number of Points

$$Side = \frac{npts - 1}{2}$$

$$i \in [-Side, \underbrace{Side}_{\text{Number of Data Points Per Side}}]$$



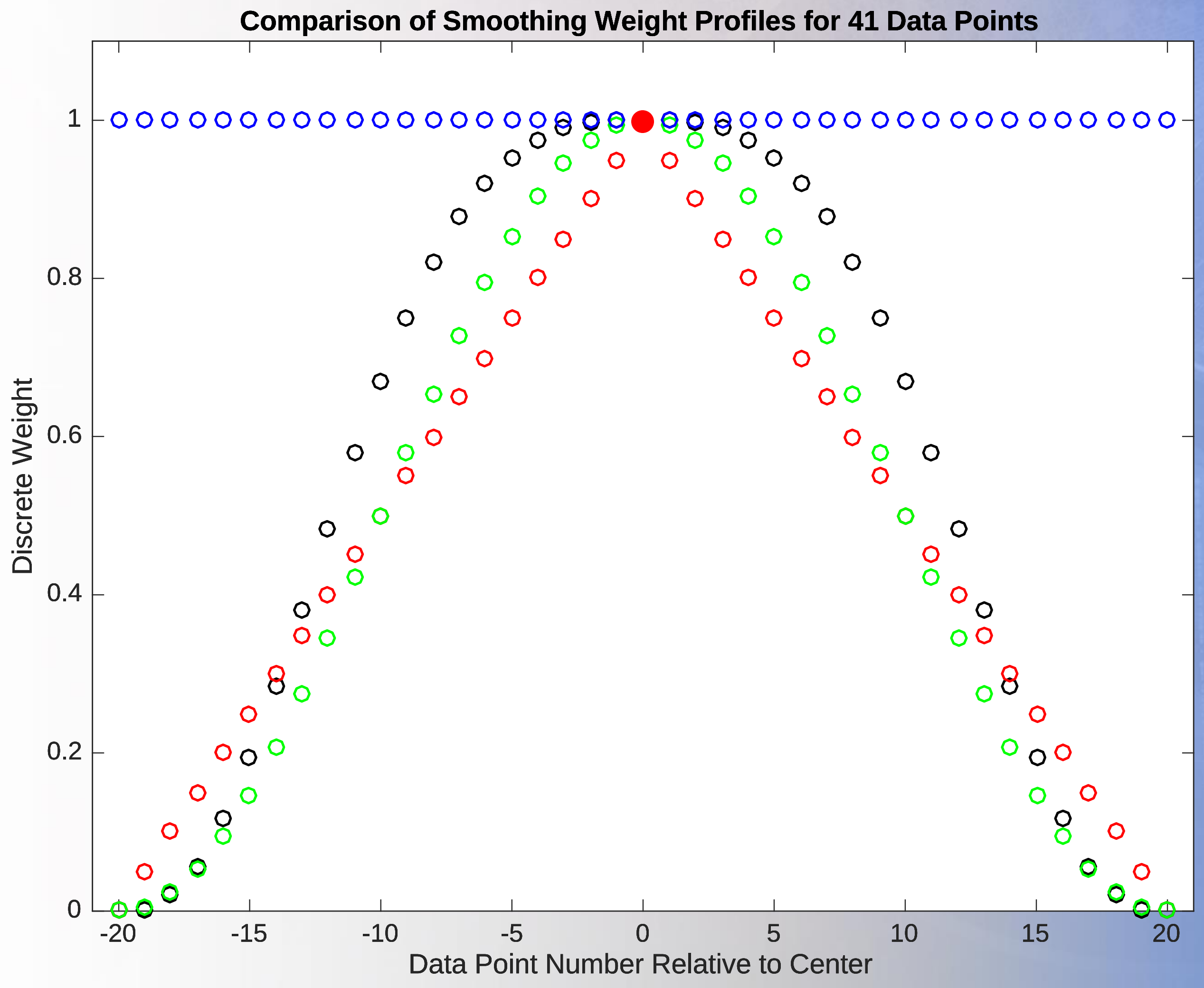
# Smoothing Function Weight Comparisons

Cubic Weights

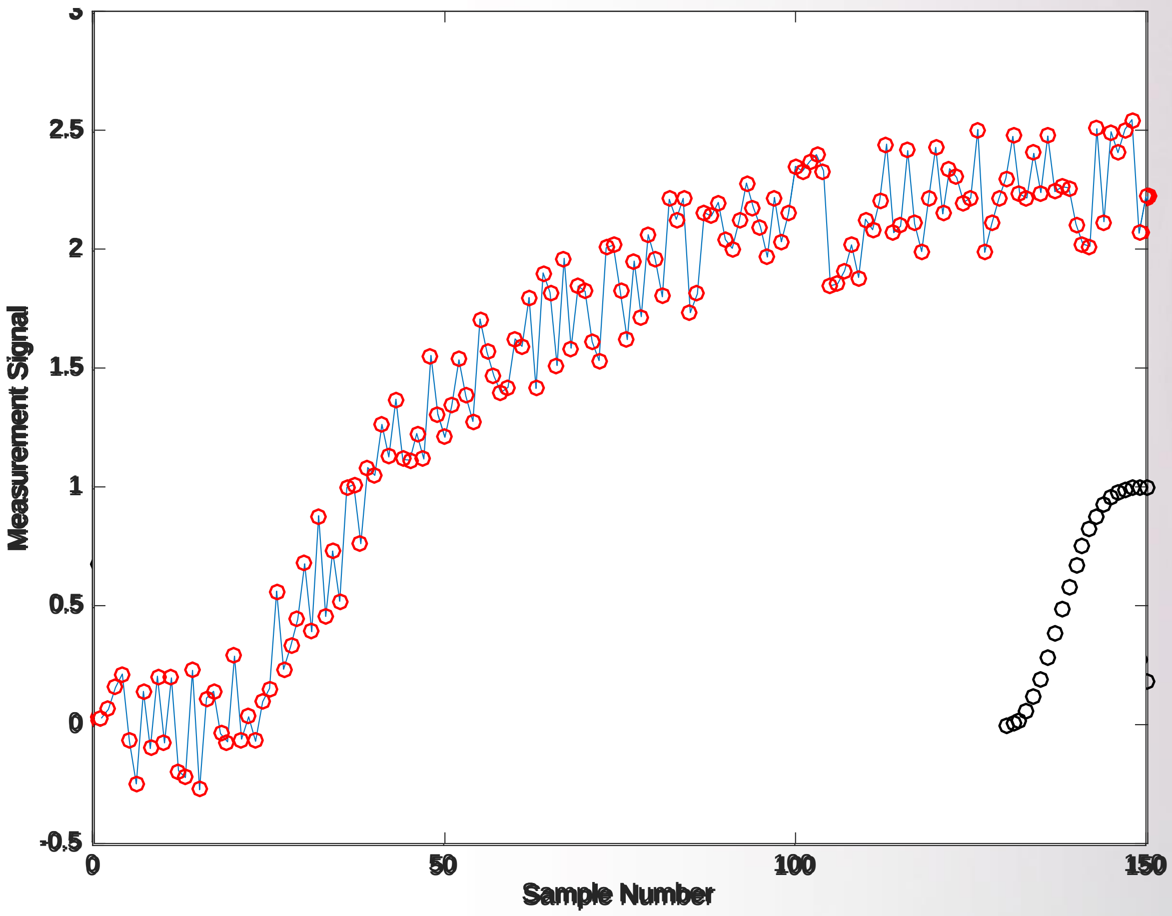
Triangular Weights

Raised Cosine Weights

Uniform Weights



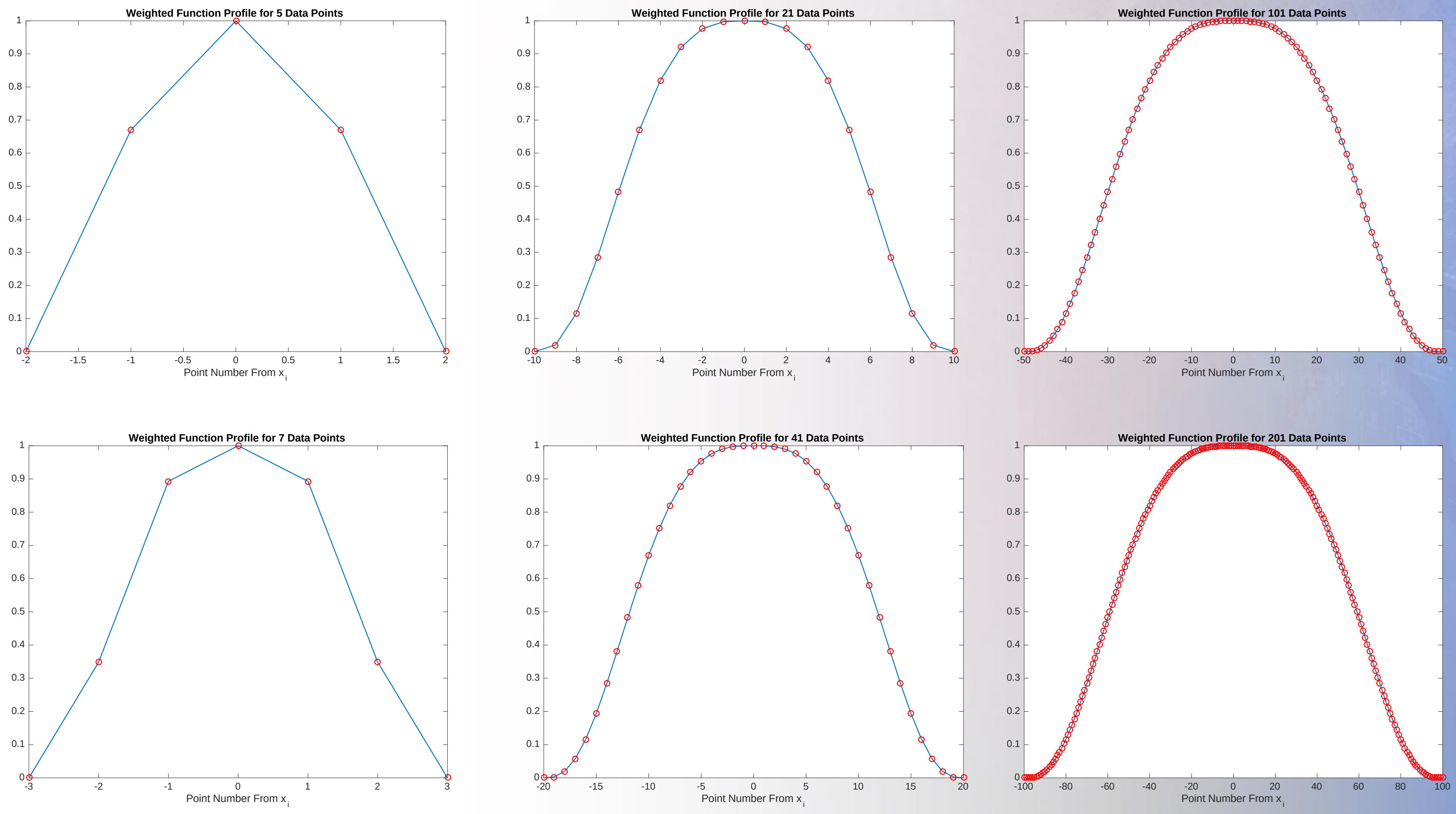
# Moving the Weighting Window



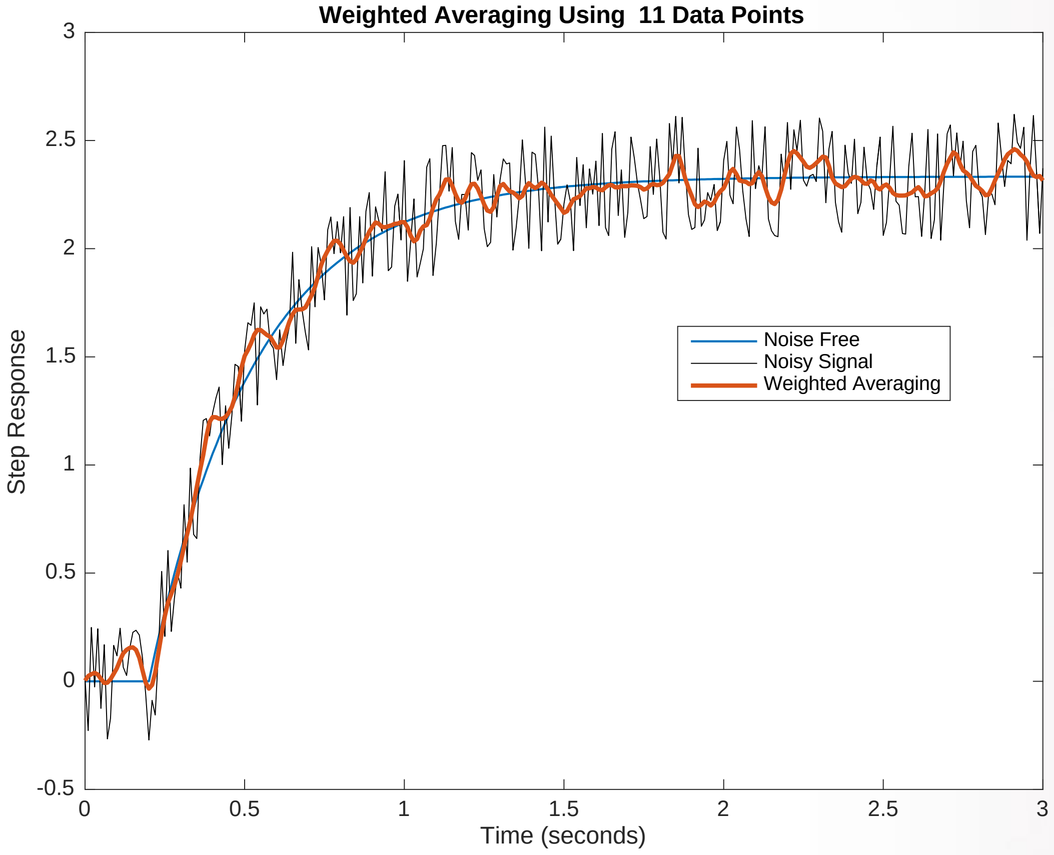

$$\begin{bmatrix} 0 \\ 0 \\ y_1 \\ y_2 \\ y_3 \\ y_4 \\ y_5 \\ y_6 \\ y_7 \\ y_8 \\ y_9 \\ y_{10} \\ 0 \\ 0 \end{bmatrix}$$

# Looking at Different Number of Window Points

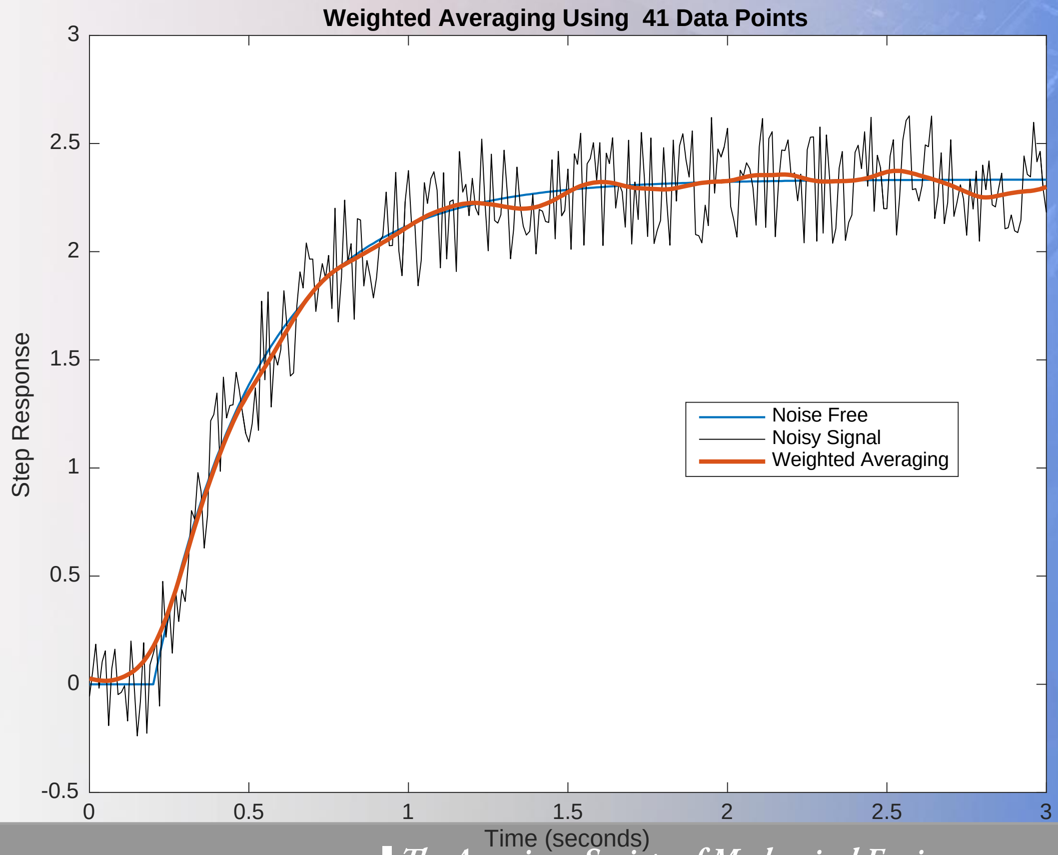
Cubic Weights



# Weighted Central Smoothing




Selecting *npts* is a balance between attenuating the noise and not suppressing the signal features or dynamics.



# Comparison of Smoothing Functions

Cubic

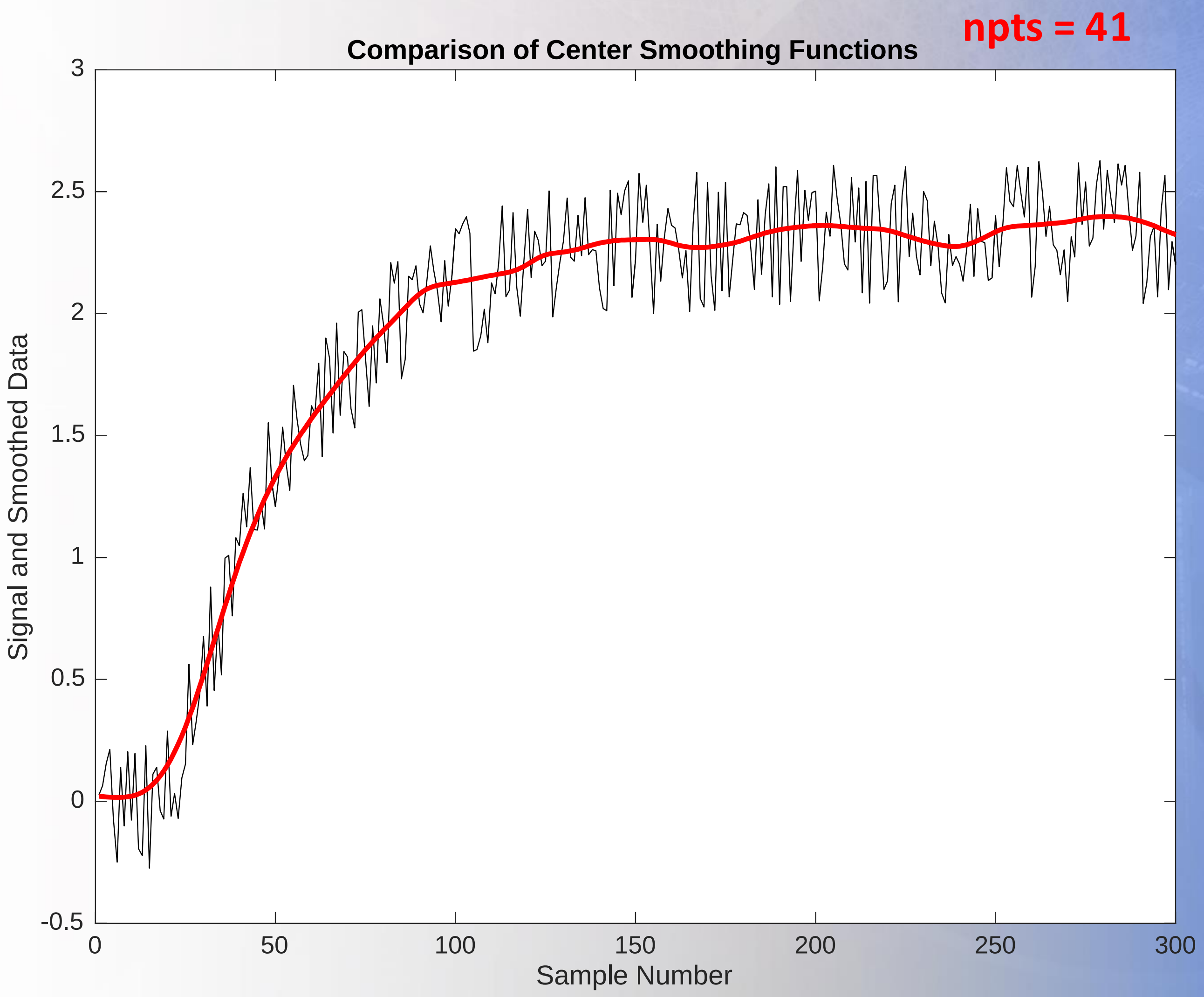
Triangular

Raised Cosine

Uniform

# Comparison of Smoothing Functions

Cubic

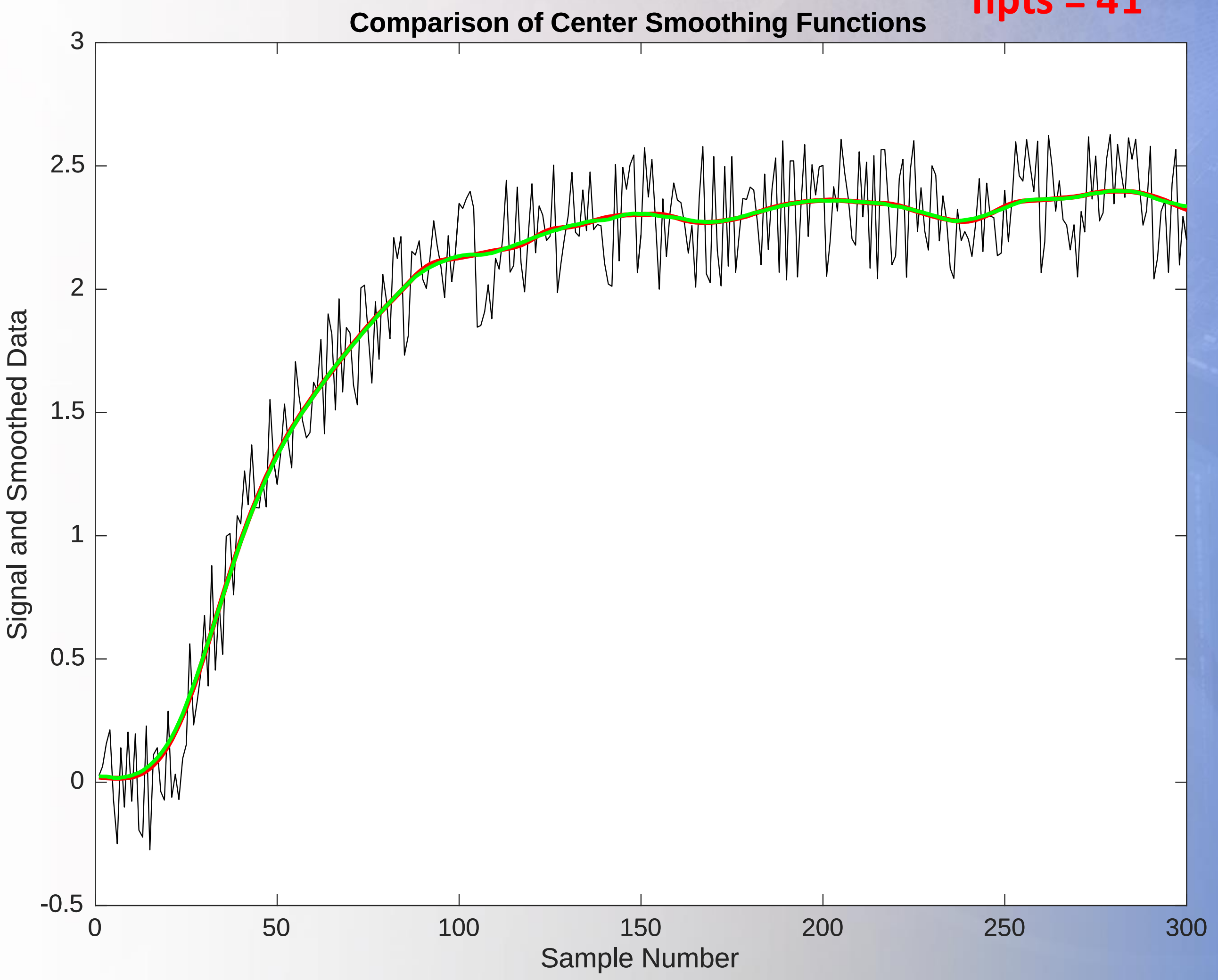


# Comparison of Smoothing Functions

Cubic

Triangular

npts = 41



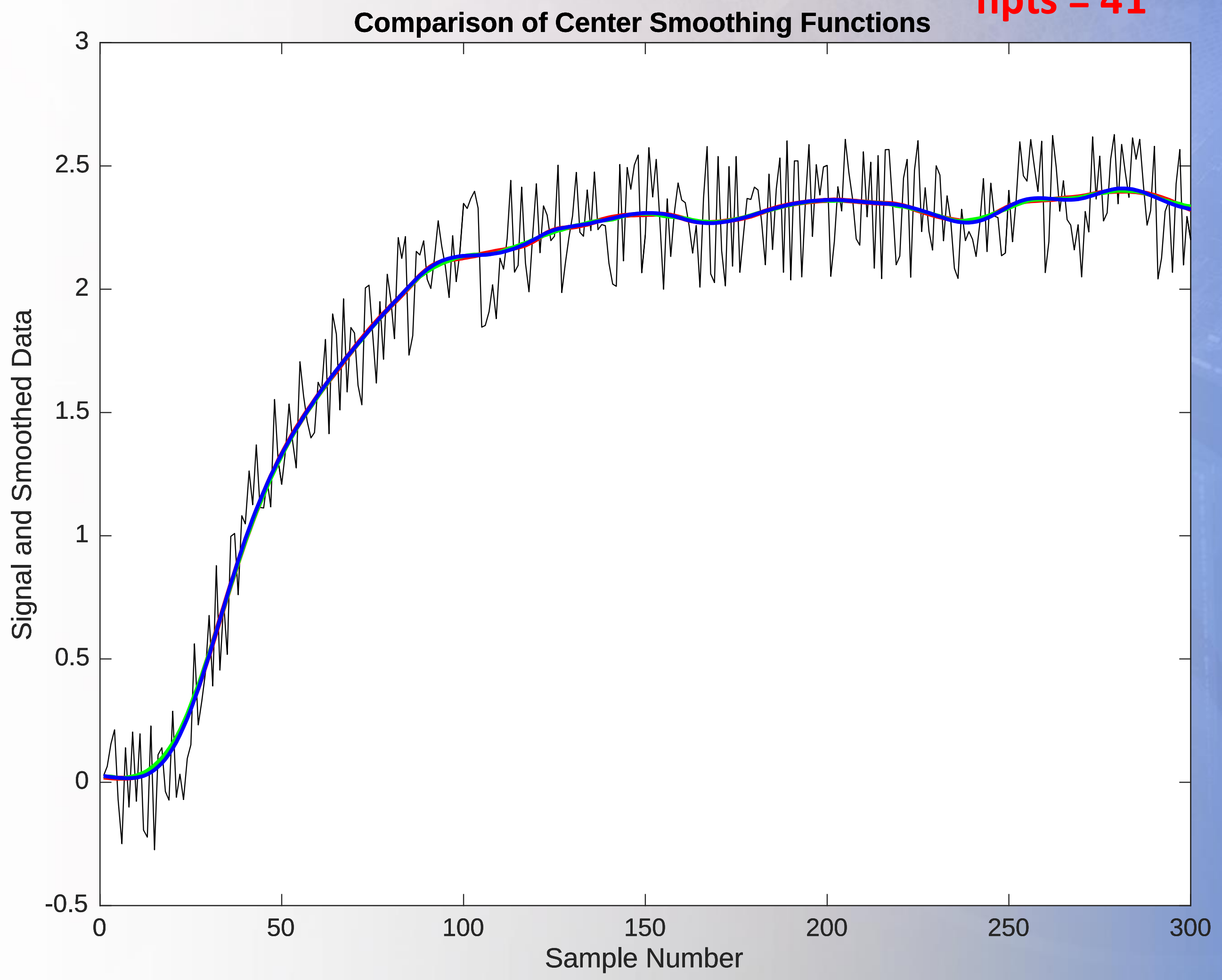
# Comparison of Smoothing Functions

Cubic

Triangular

Raised Cosine

npts = 41



# Comparison of Smoothing Functions

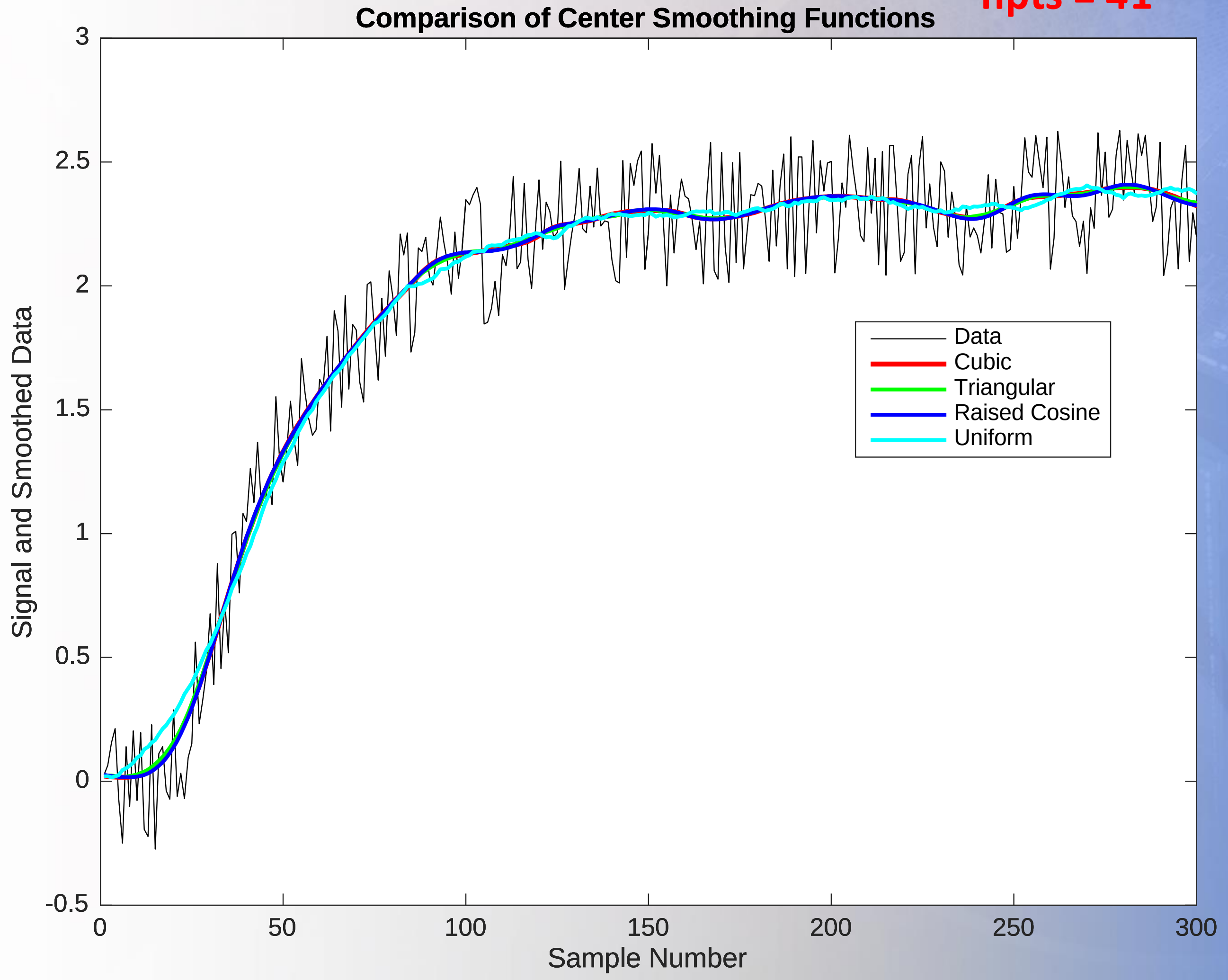
Cubic

Triangular

Raised Cosine

Uniform

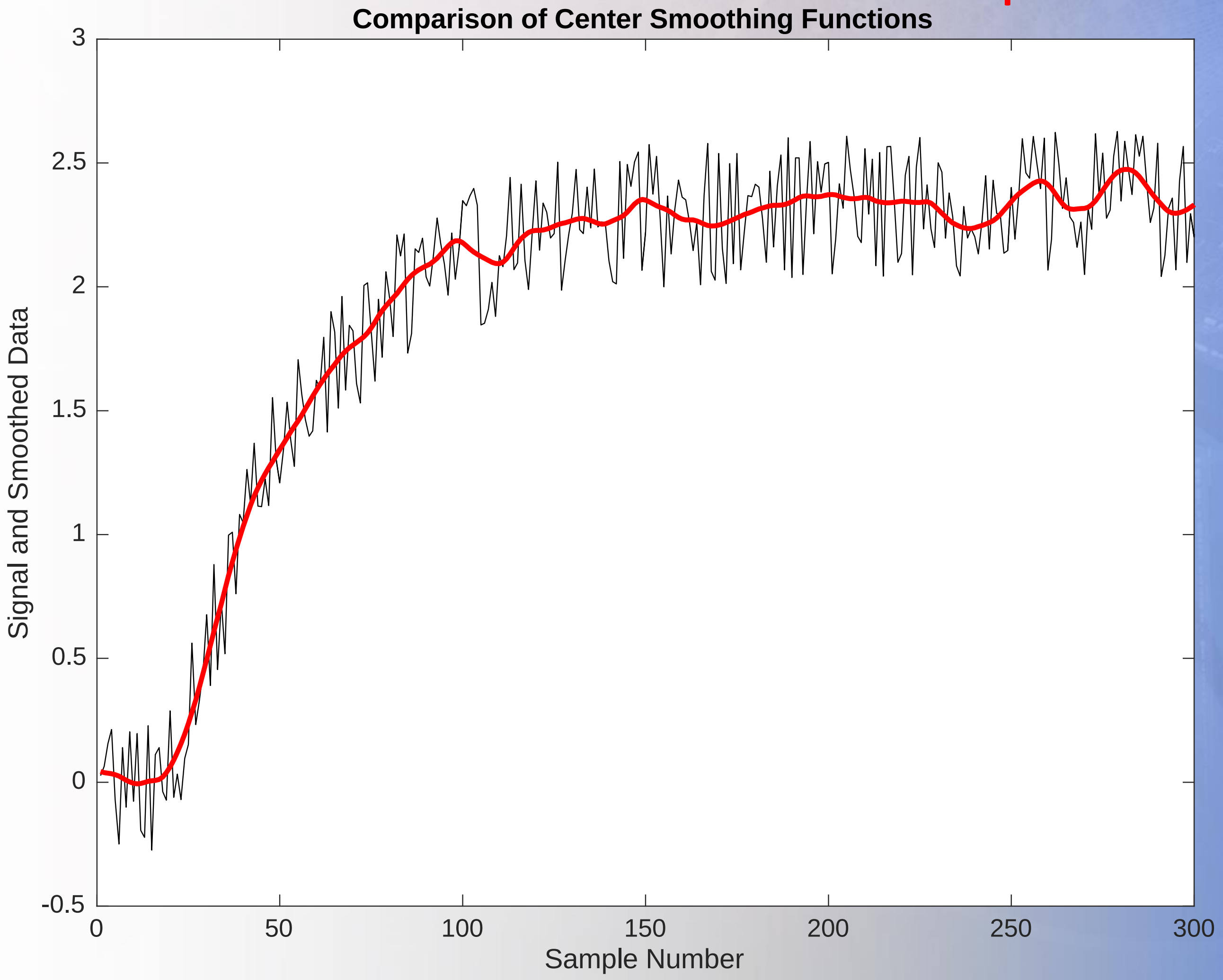
npts = 41



# Comparison of Smoothing Functions

Cubic

npts = 21

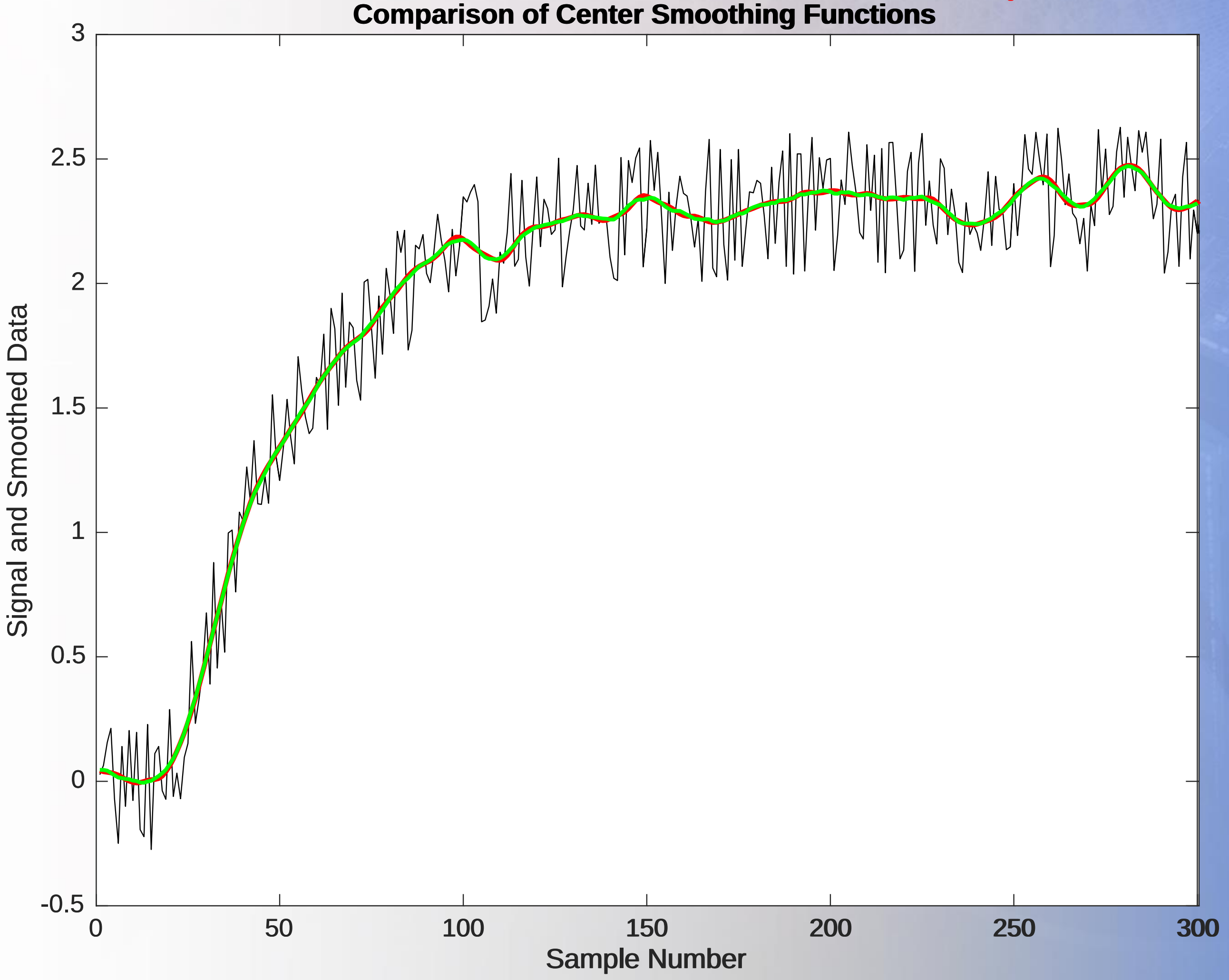


# Comparison of Smoothing Functions

Cubic

Triangular

npts = 21



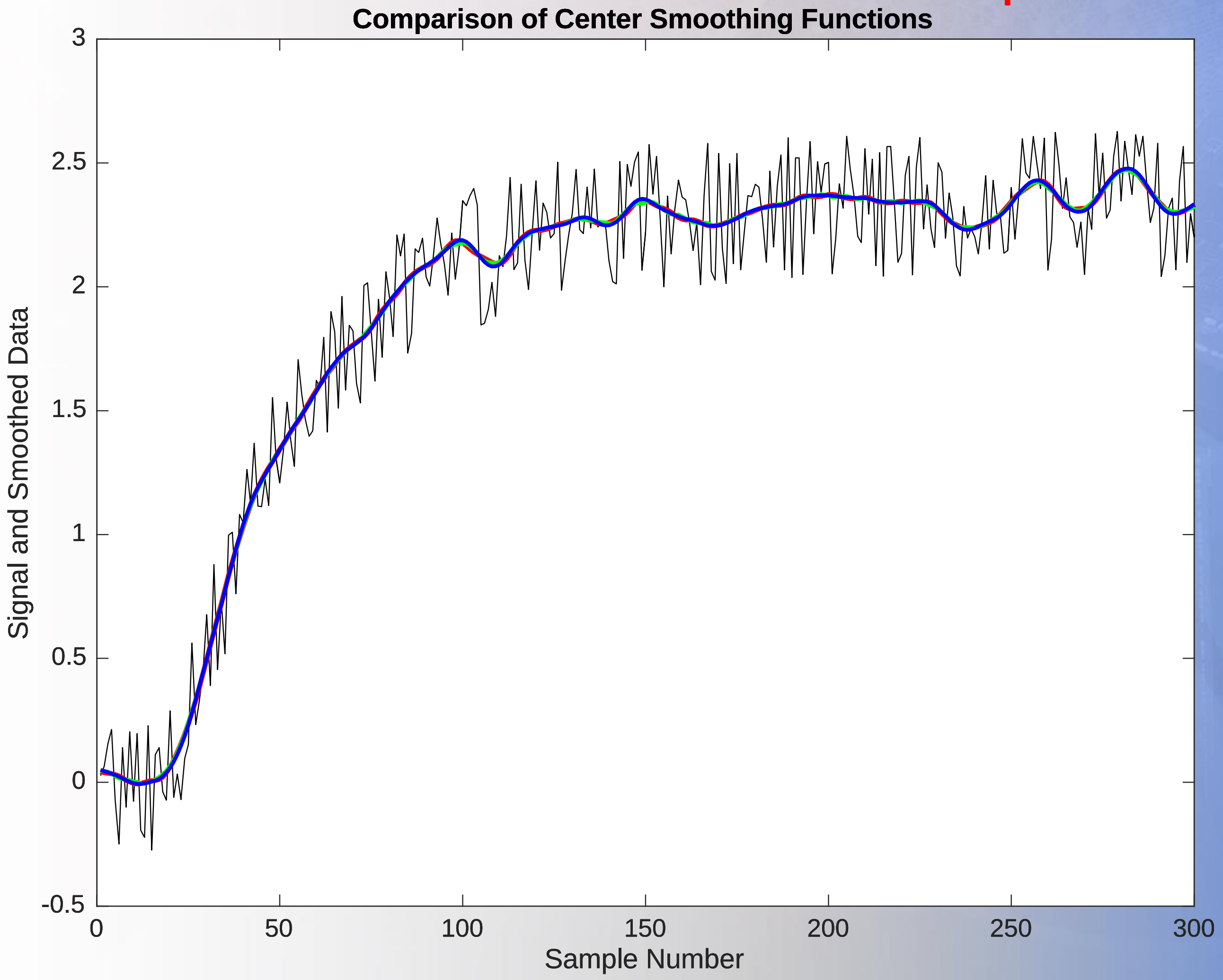
# Comparison of Smoothing Functions

Cubic

Triangular

Raised Cosine

npts = 21



# Comparison of Smoothing Functions

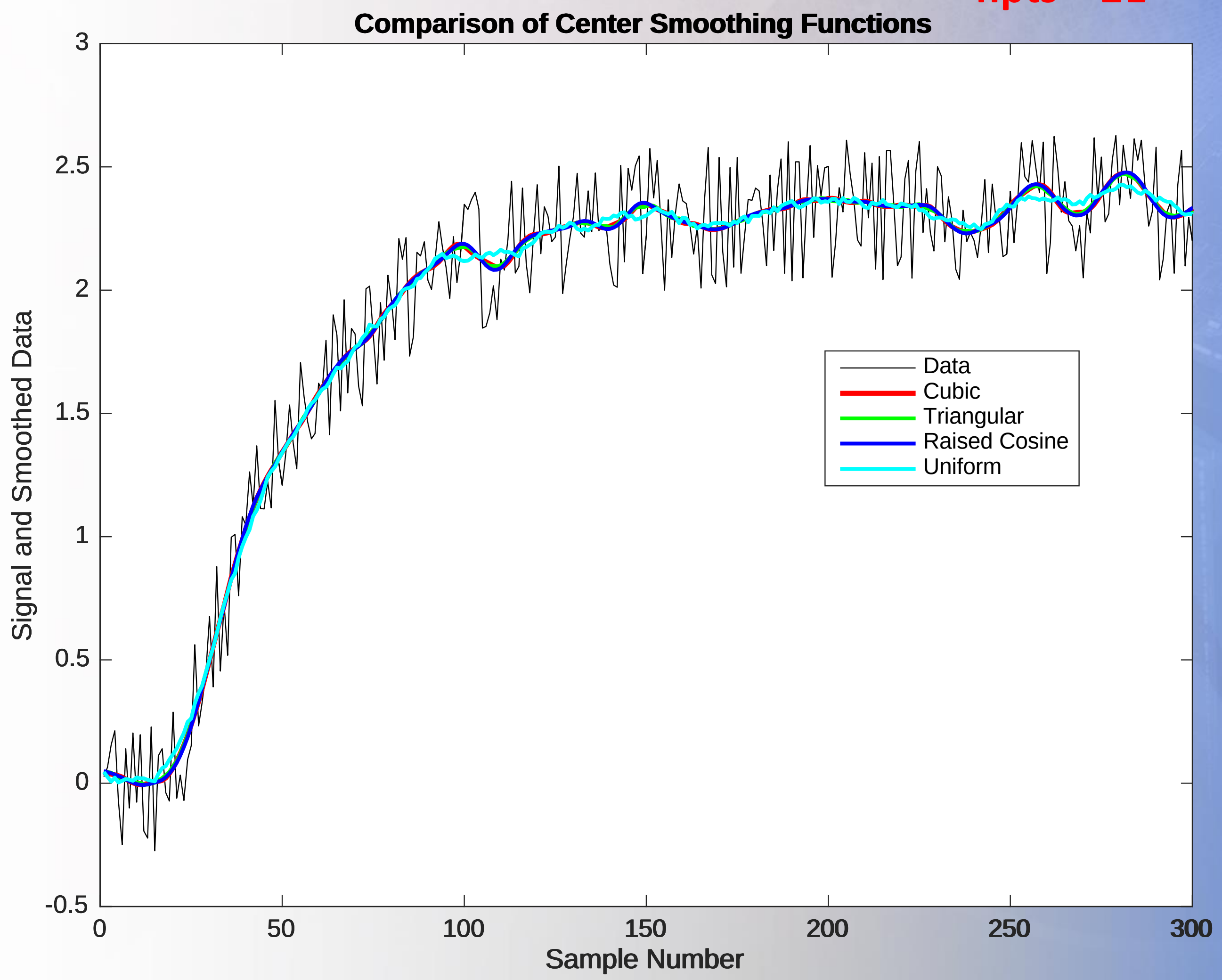
Cubic

Triangular

Raised Cosine

Uniform

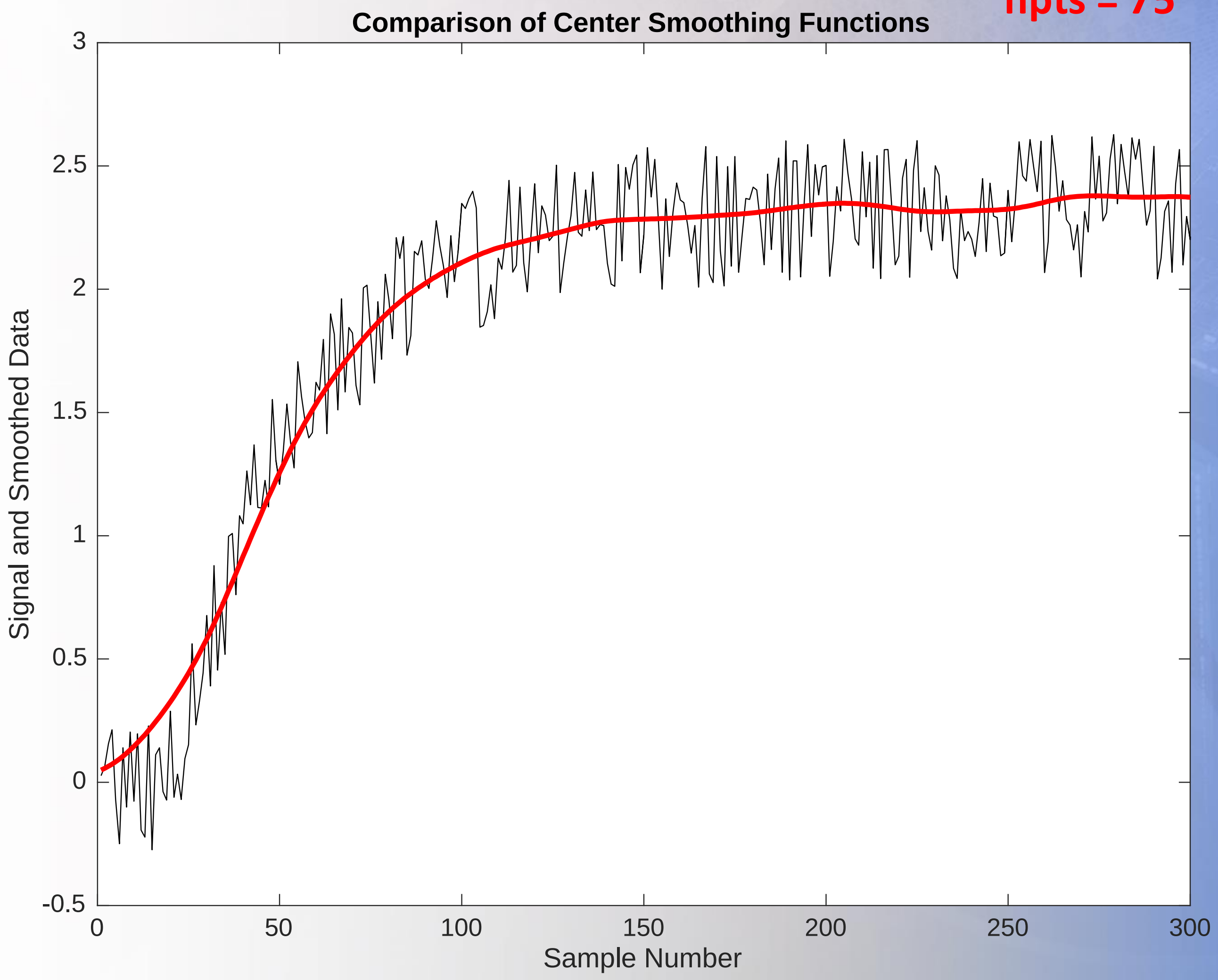
npts = 21



# Comparison of Smoothing Functions

Cubic

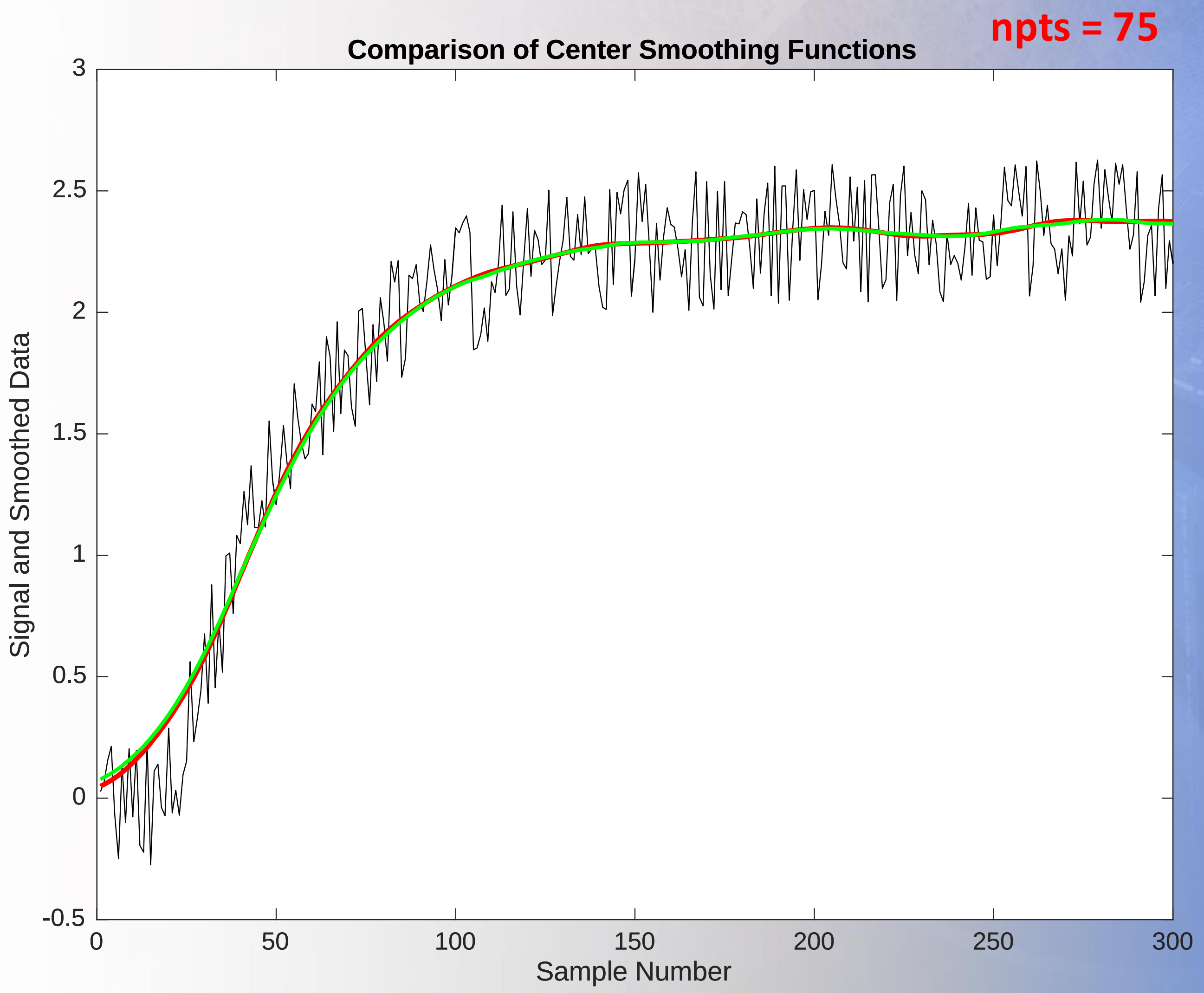
npts = 75



# Comparison of Smoothing Functions

Cubic

Triangular



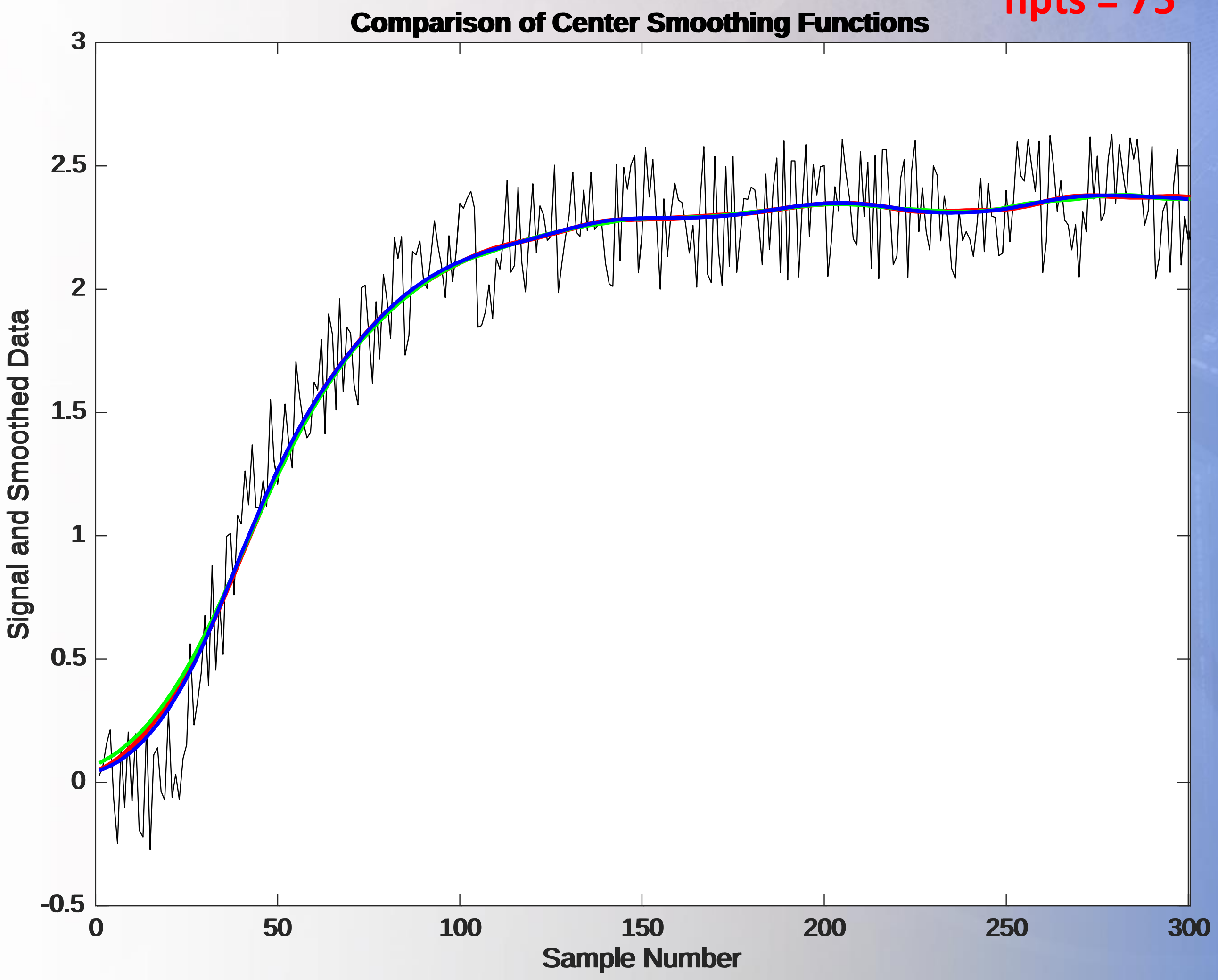
# Comparison of Smoothing Functions

Cubic

Triangular

Raised Cosine

npts = 75



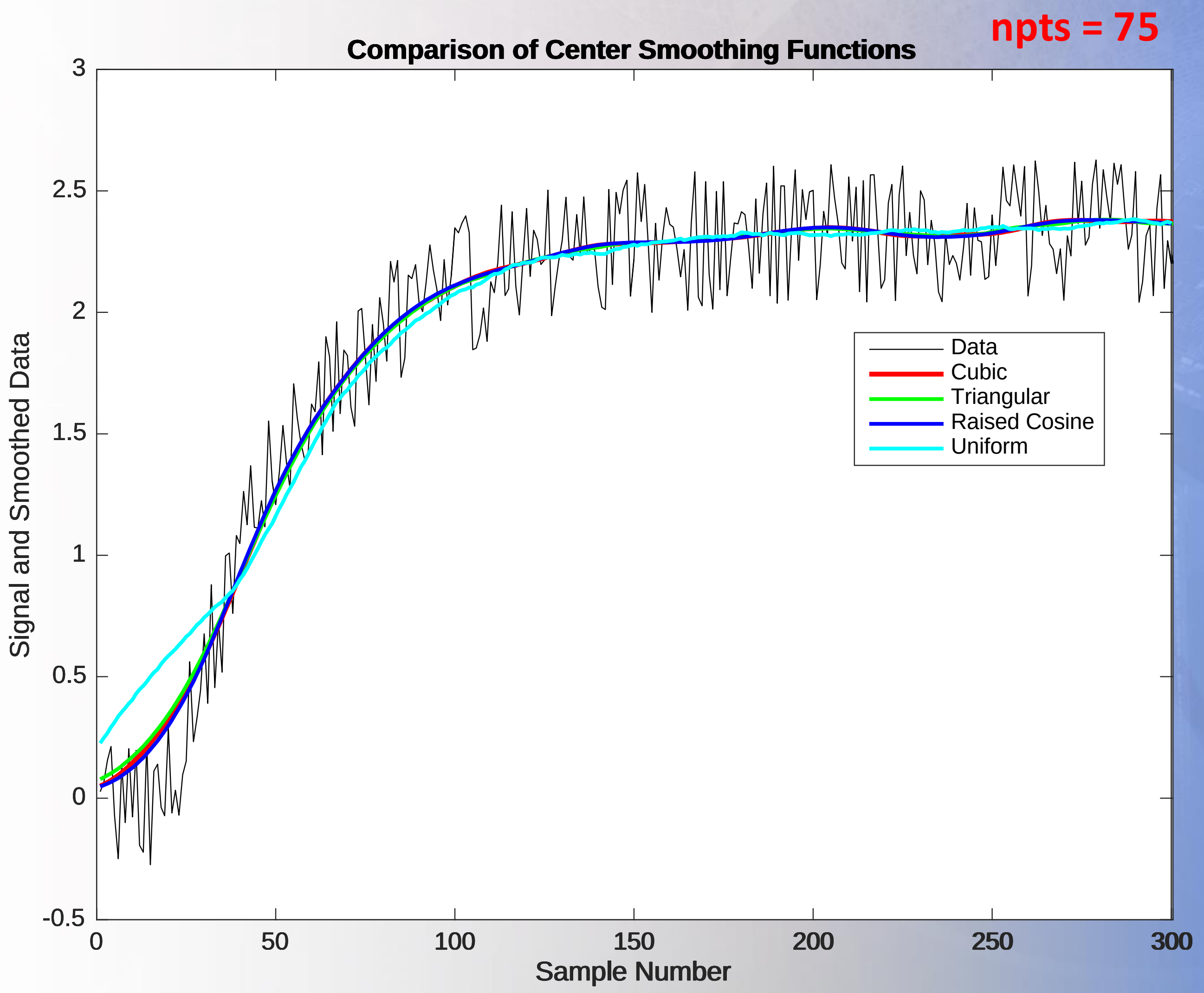
# Comparison of Smoothing Functions

Cubic

Triangular

Raised Cosine

Uniform

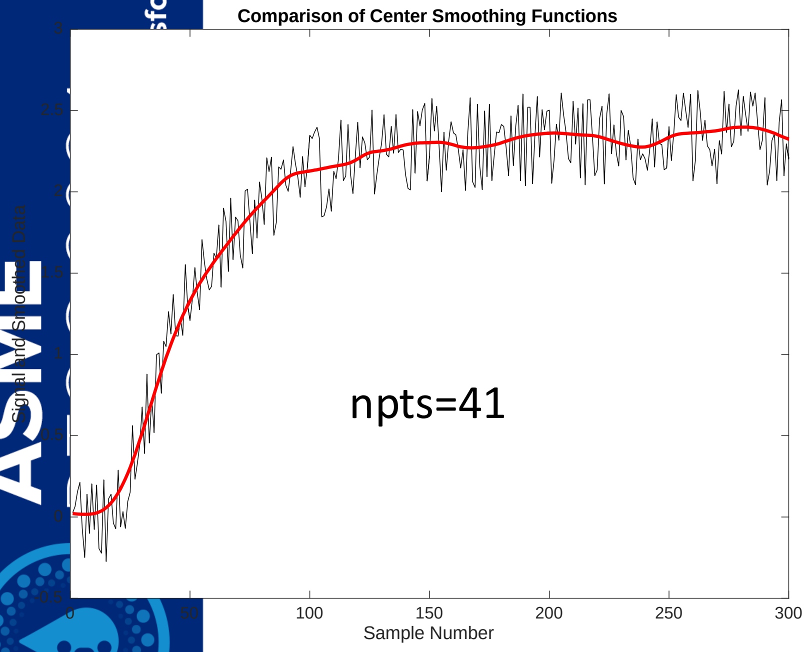


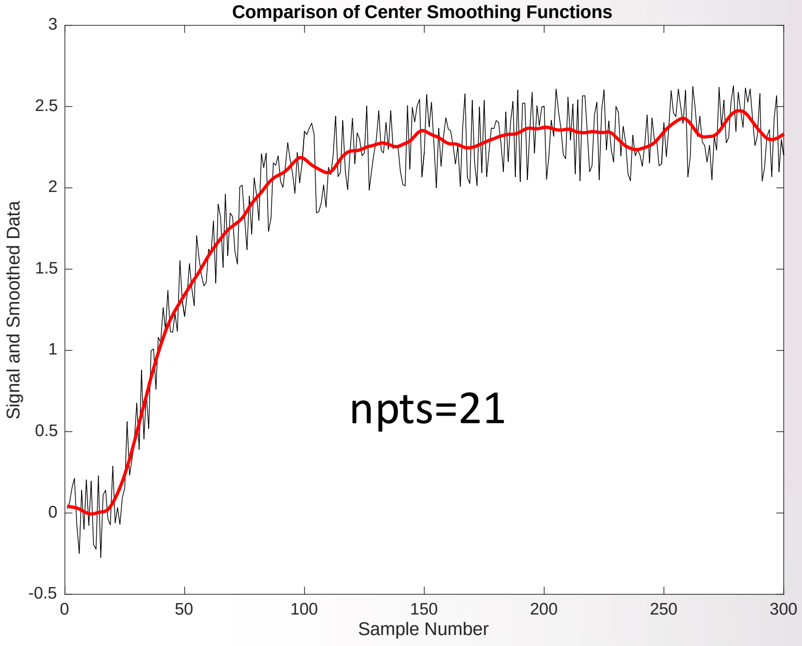
# Best Practices

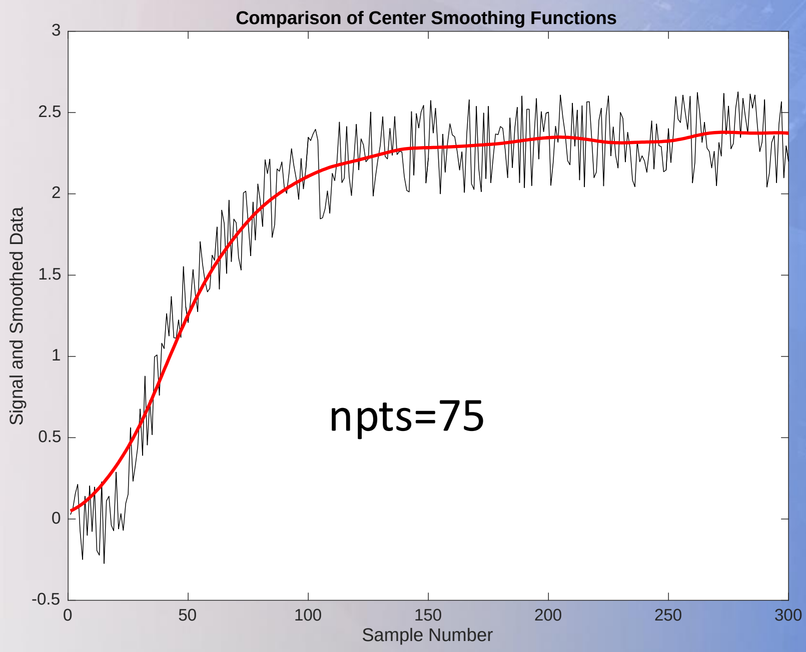
## Selecting the Window Size (*npts*) involves tradeoffs.

Suppressing the noise vs suppressing the signal features or dynamics.

Selecting *npts* is qualitative and based on what you want to do with the signal.



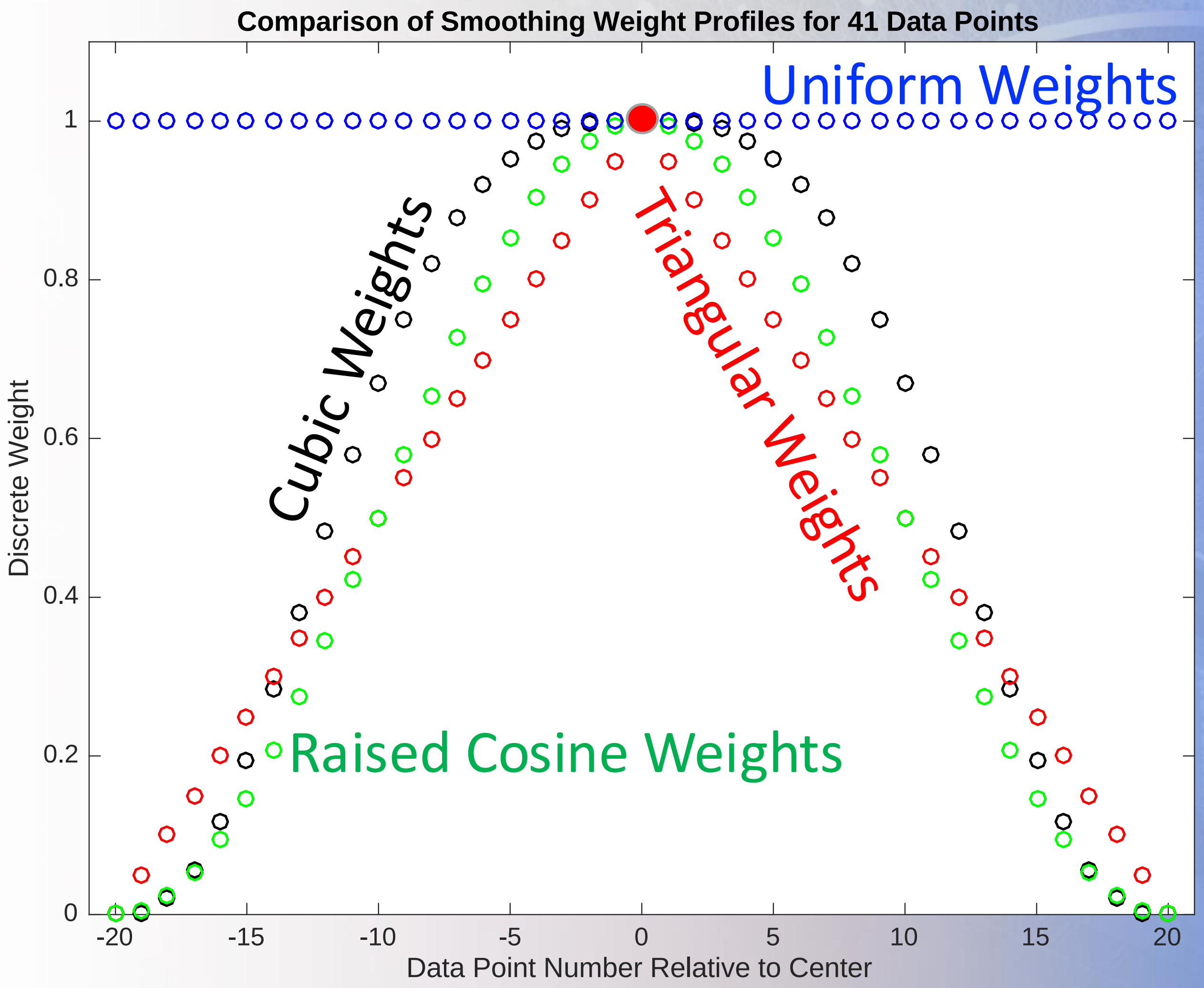



# Best Practices

Selecting Window Type is based on what degree you want to weigh data points near the the data point of interest to those points further away.

Cubic weights are the most used.

Raised Cosine weights emphasize those data points closer to the point of interest in comparison to Cubic weights.




Triangular weights de-emphasizes data points near the data point of interest while emphasizing data points further away.

Uniform weights recovers the mean value.

# The UH Data Cleansing and Regression GUI

# App Tutorial

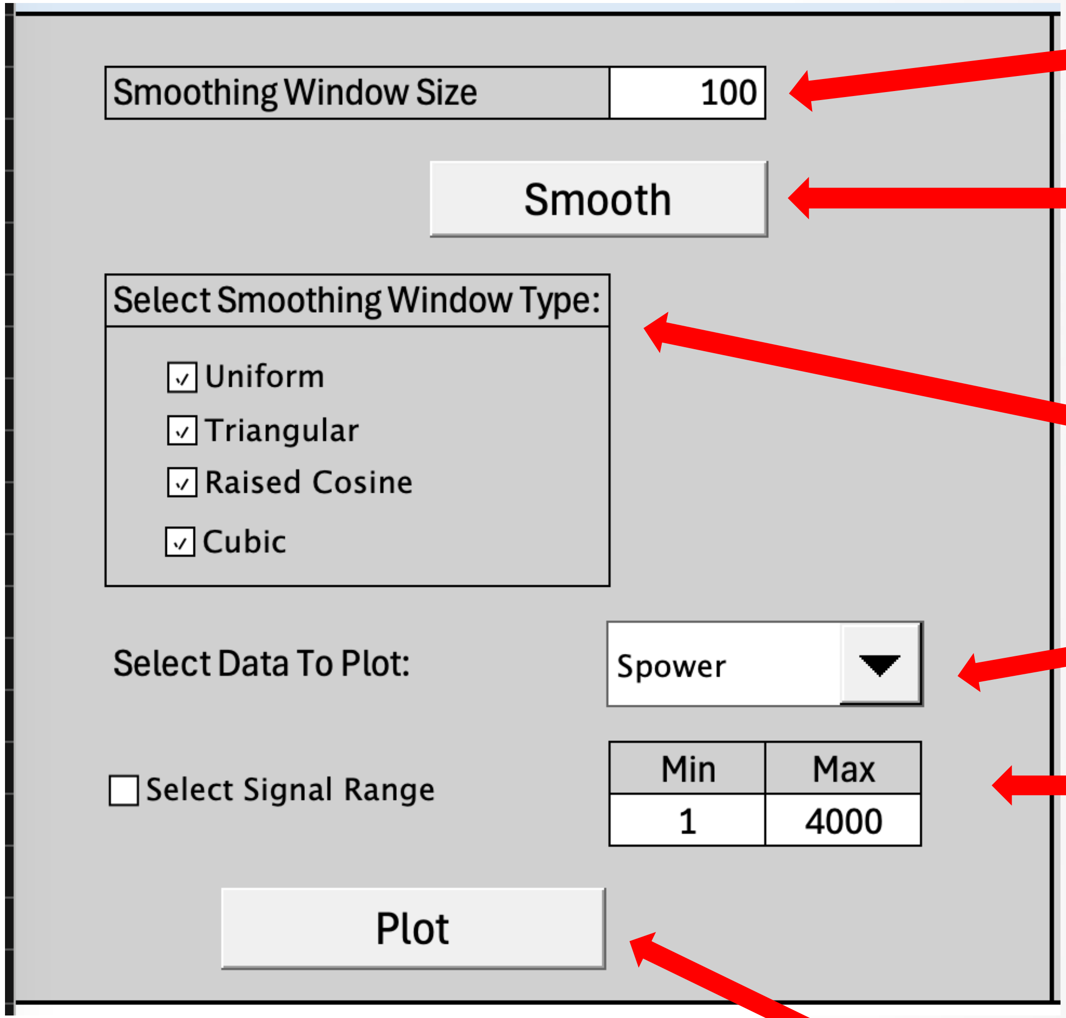

App sheet

Paste your data in the Raw Data Sheet

App results for smoothing

Parameters Estimation App results

# App Tutorial: Data Smoothing App




1- Choose Window Size

2- Press Smooth to smooth all the data in the Raw Data sheet

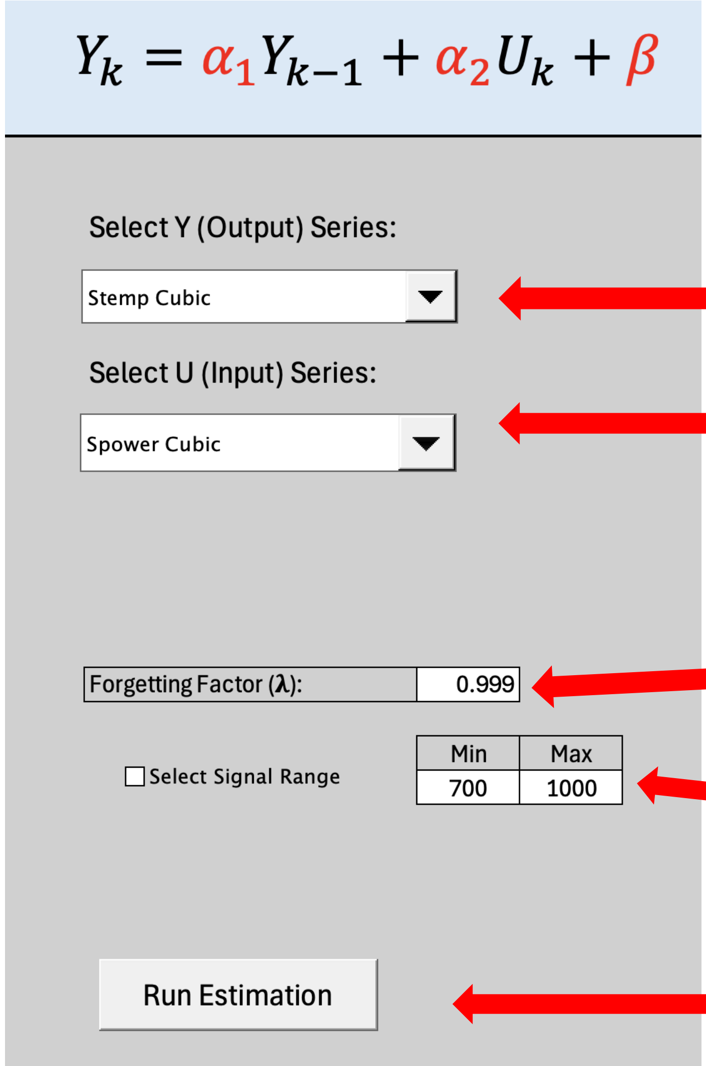
3- Choose the smoothing window type for plotting

4- Choose the data for plotting

5- Optional: You can select a range of samples to zoom in by checking “Select Signal Range” and assign a min and max range.

6- Press plot to plot the data and the smoothed data on the same plot

# App Tutorial: Parameters Estimation App




Regression dynamic model

1-Choose an output variable (target variable $Y$)

2-Choose an input variable (target variable $U$)

3- Assign a forgetting factor (preferably closer to 1)

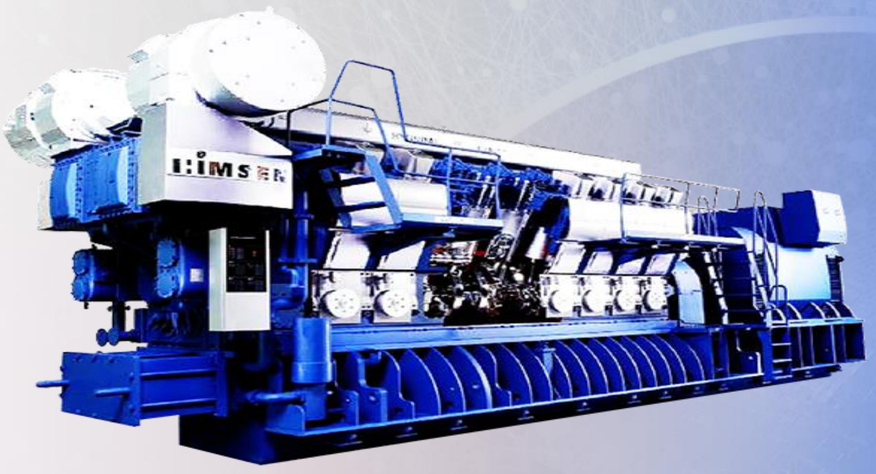
4- Optional: Select a range of data to zoom in for plotting

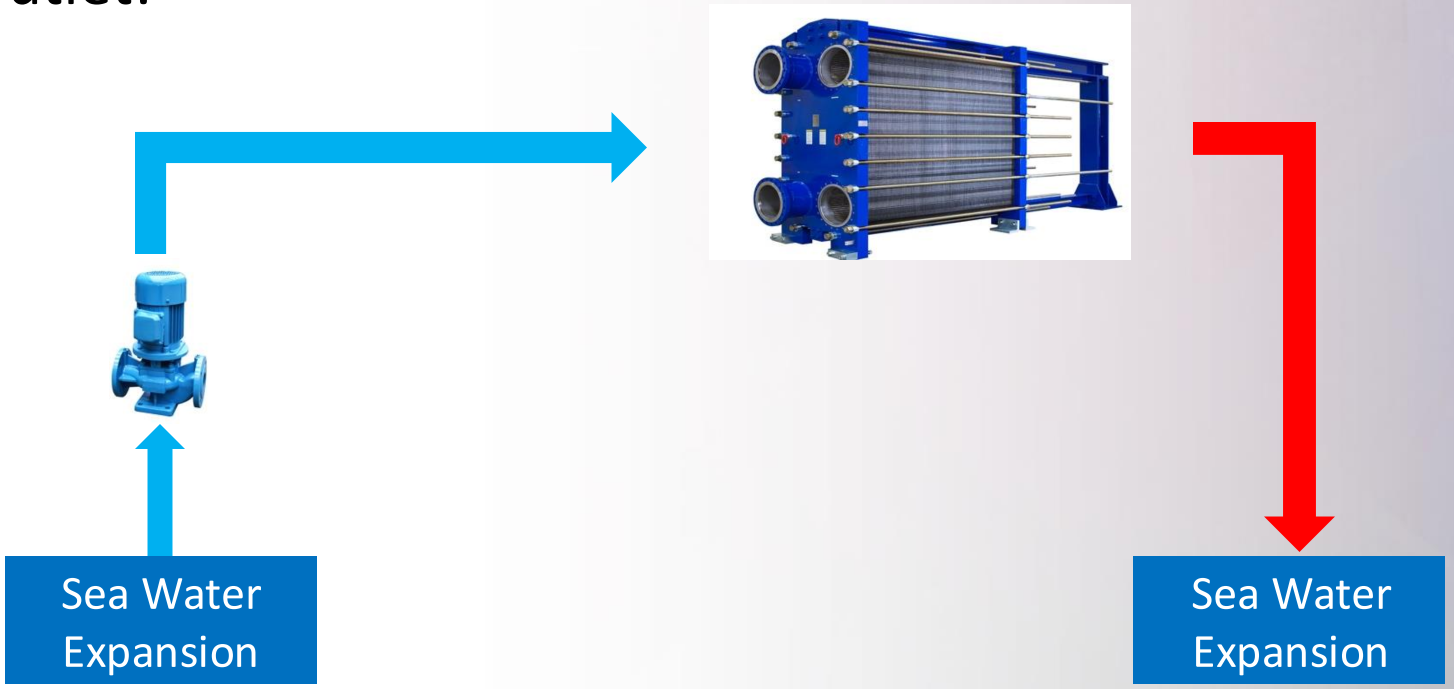
5- Press "Run Estimation" to get results in the "Parameters Estimation sheets" and plot the results.

# Data Cleansing Practice Examples and Regression Analysis

# Practice Example

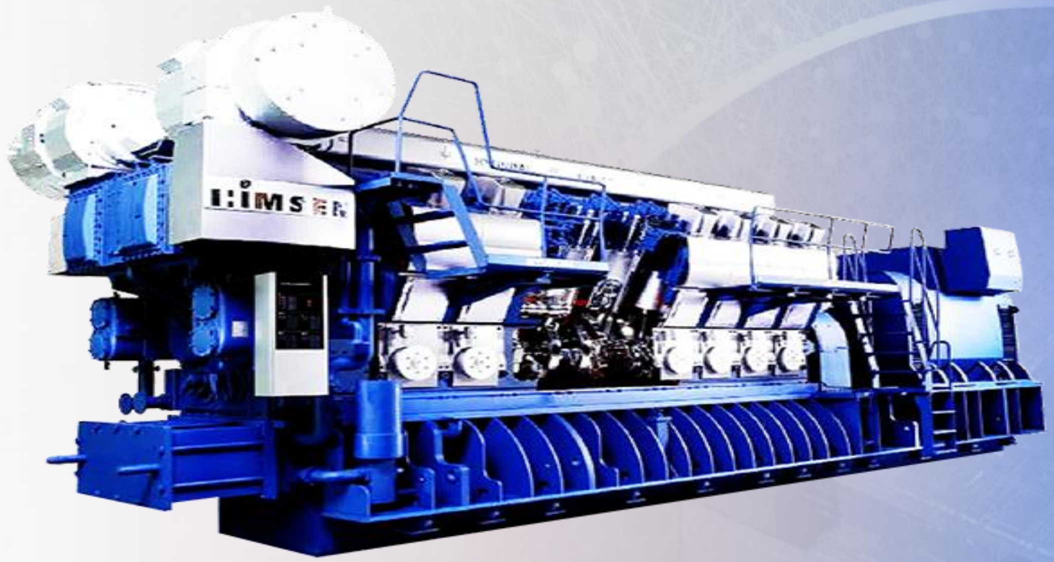
- Cooling water temperature for engine inlet.
- Cooling water temperature for engine outlet.

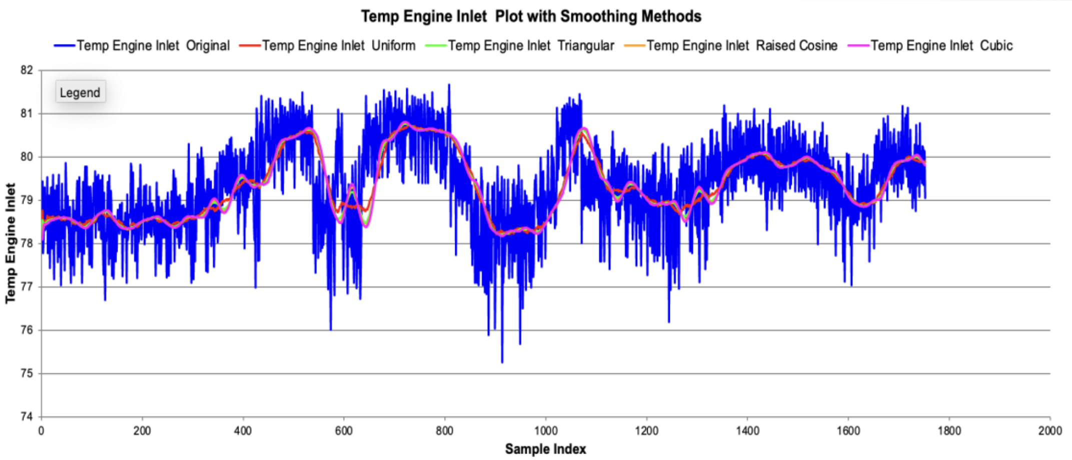


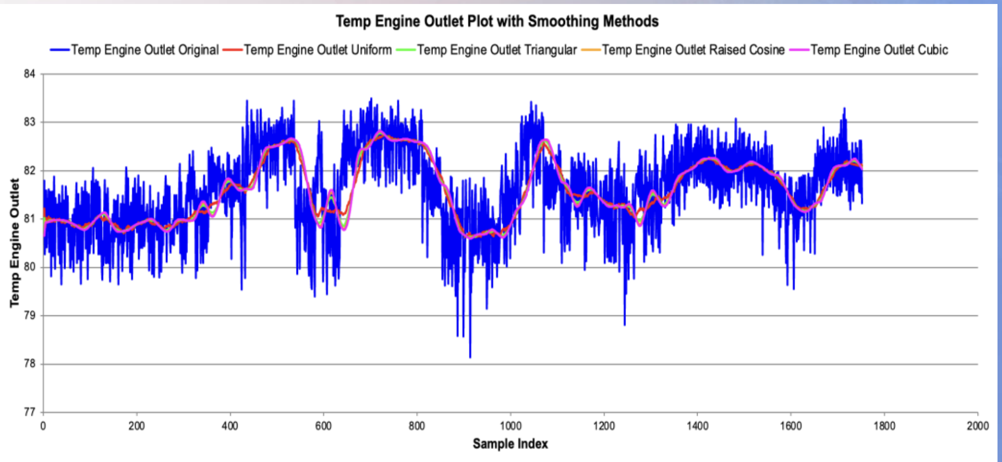


# Practice Example

- Cooling water temperature for engine inlet.
- Cooling water temperature for engine outlet.

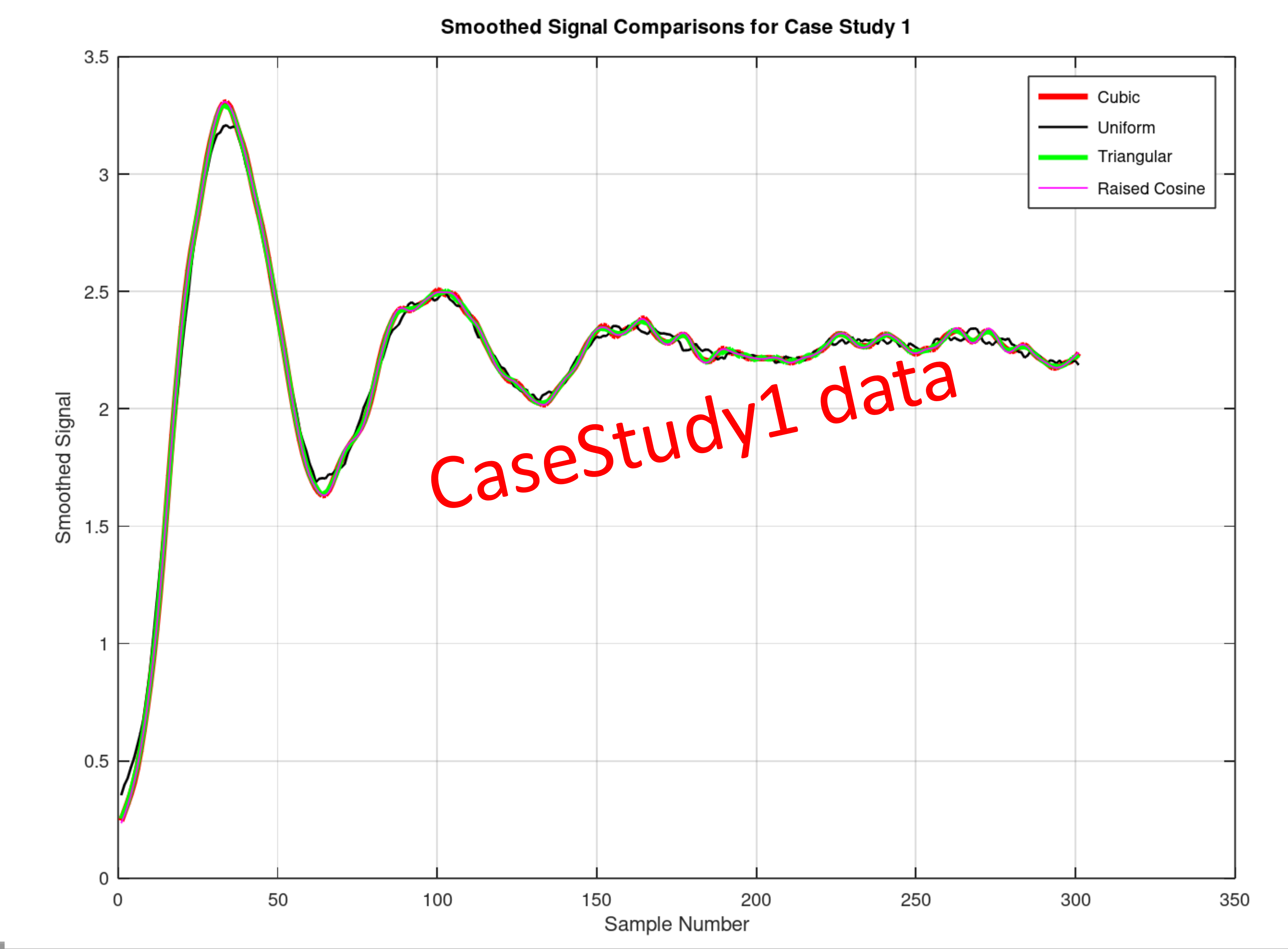





$$T_{out} = a_1 T_{out} + a_2 T_{in}$$

# Plotting All 4 Smoothing Solutions



# Thank you for your attention

# Questions?