# Model Agnostic

## Extreme Sub-pixel Visual Measurement and Model Characterization

By Michael R. Munroe

# Thank you

- Micah Munroe, my wife
- Frank and Sabina Peyton, who said “go”
- Stew and Becky Bruner, who made a bridge
- Jeff Lavell, who taught me how to look around
- Michele Milano, who opened the door
- Garth Bowen, Mike MacGregor and Ibrahim Bekar who said keep going

And the Providence that brought me here today.

# Agenda

Introduction
- Three  Scientific Agnosticisms
- Visual Metrology
- The Starter Problem

Extreme Sub-pixels
- Framework
- Exploring “knobs”
- Results

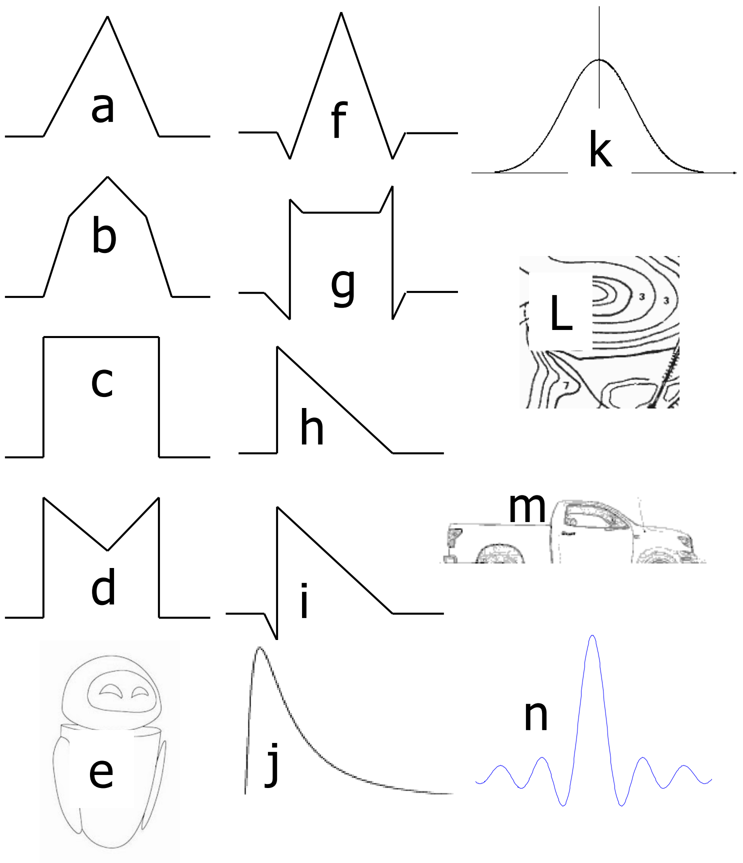
Global Smoothing
- Statistical back-story
- Drunkards walk revisited
- Agnostic Optimal Smoothing

Model Variation
- The incomplete mean
- Smoothing revisited
- Empirical Kalman Formulation

Test Cases
- Line Tracking
- Ball Tracking

# THREE SCIENTIFIC AGNOSTICISMS

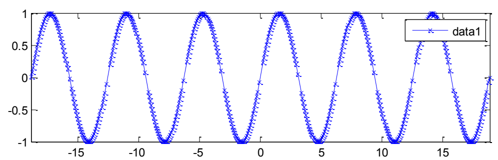
# Instant: Is the edge a line?



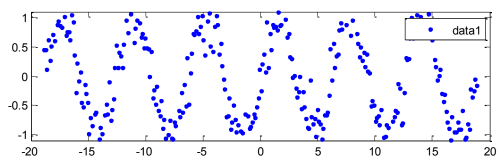

- Think of these as intensity profiles.
  - They have different Fourier Series, and different energy spectra.
  - They are a trivial subset of what can be found in images
- How do you work with them
  - If they represent the shape of a mountain you are climbing
  - If you have the wrong map
- We assume the real world is shaped something like this every day.
  - Processing the stock market
  - Making assumptions about the shape of a machined parts or processes.
  - Construction of Autopilot and cruise control systems

Our systems might be better if they could handle more, whatever that means.

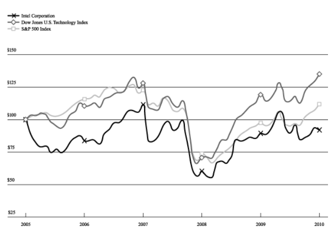
# Over time: measurement isn’t information

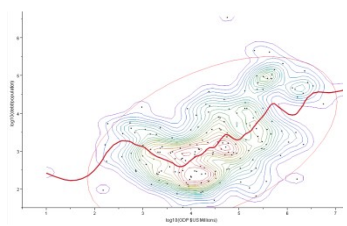



$y=\sin(t)$

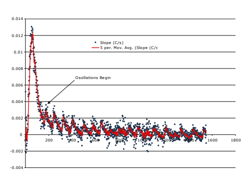



$y=\sin(t) + \varepsilon$





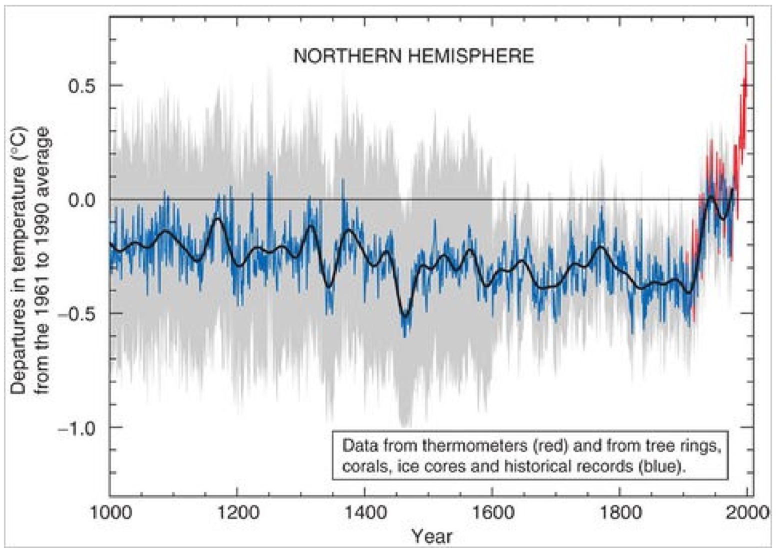
Y=???



- When we measure:
  - We measure over time
  - The measurements are comprised of signal and noises
  - They are fixed in count – the real world is continuous the measurements are not.
- Questions
  - How many terms in the Fourier series until we are over-fitting?
  - What if we used a Polynomial?
- Points
  - We have a finite zoo of fit-functions here too.
  - Many of our functions come with baggage.

Our systems might be better they could get at the information, whatever that means

# Variation: The mean is nonphysical




Q: Am I a light-year (LY) away from you?
A: Yes, one LY ± one LY.

- Mean is best estimator of "central tendency"
- A measurement of mean is incomplete – the real world has both variety and noise.
- Physically valid measurements characterize uncertainty/variation.
- There are multiple classes of sources of variation: phenomena, sensor, and model

Our systems might be better they could characterize their own uncertainties.

# By Agnosticisms I mean:

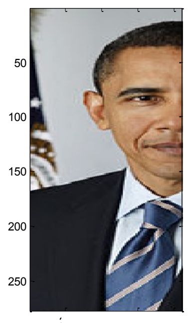
1. It can work with many profiles because it makes the data supply the profile instead of approaching it with a mean, or an a-priori model. *Assumes signal energy spectrum doesn't change (much) over time.*

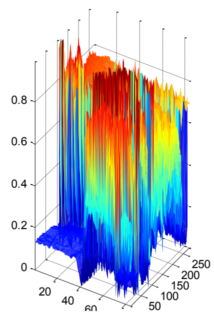
2. It can automatically remove noise from a system using non-parametric/hyper-parametric basis functions with a useful fit metric. Model Agnostic is the best description for this combination. *Assumes regularity conditions for Information Criterion are met.*

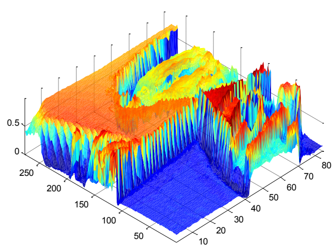
3. It can automatically account for its own variation (and enable lots of fancy processing) by interpolation, and model-agnostic basis functions, and an extension to a very useful fit metric. *Imposes Discrete Kalman Filter form on process.*

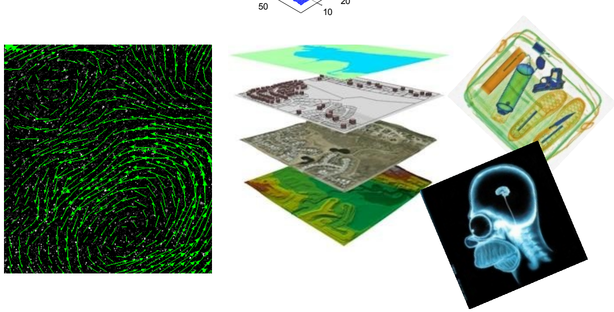
# VISUAL METROLOGY

# Measuring things using images



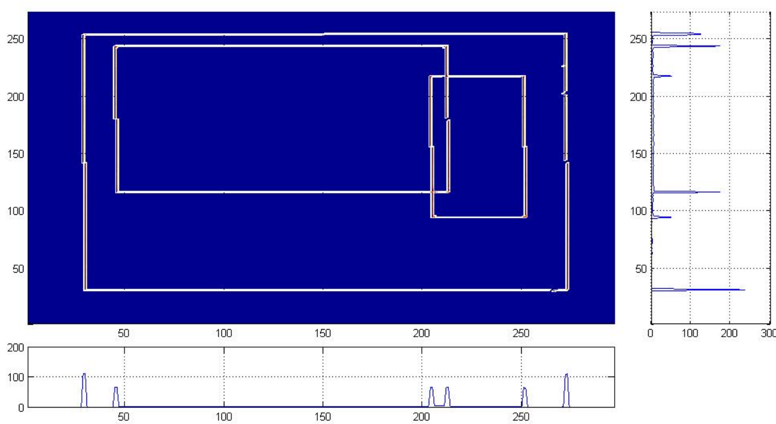



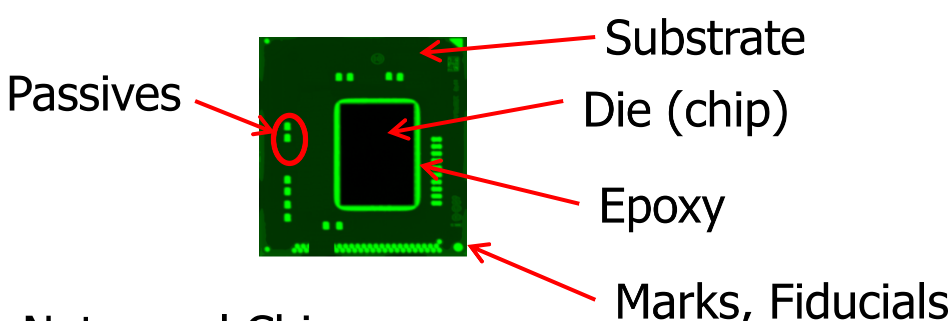



- Most human-useful visual information is complex, and noisy.
- Procedures that depend on pristine phenomena fail.  How do you measure the edge of the shoulder?
- Paradigm
  - We use intensity values/transforms to trigger a "measure".
  - We convert pixel-coordinates to a translated phenomena coordinates.
- Current applications for visual metrology
  - PIV, flow-field metrology
  - Agriculture, Ecology, Geography, …
  - Manufacturing Quality Control, defect capture, measurement
  - Health (X-ray, CAT, MRI, …), Security

# What I measure with Images

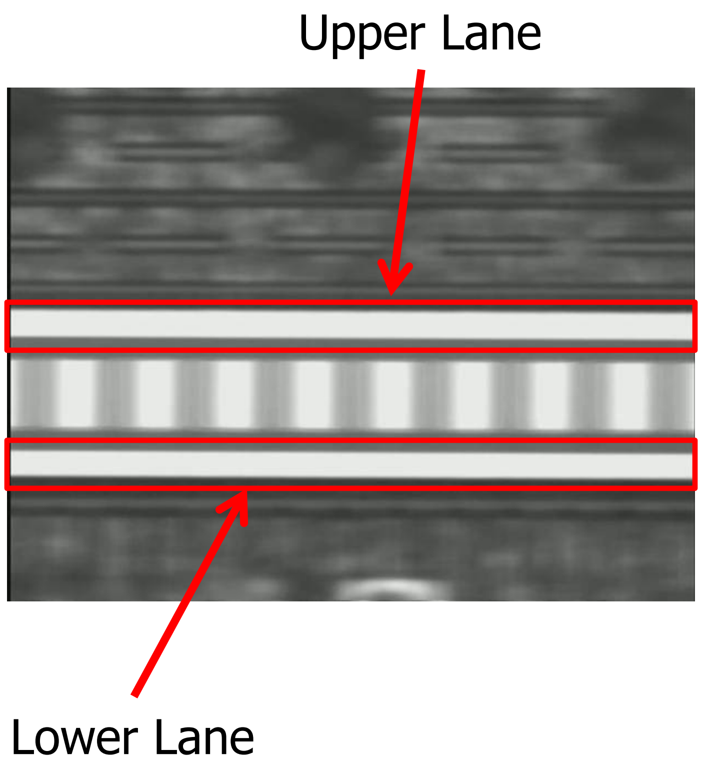





Not a real Chip

- Part phenomena
  - Substrate and Chip: Center, Edge, size
  - Ball Height, Diameter, Offset
  - Ball field Coplanarity
  - Passives, Lids, Pins, Pads, Fiducials, Other marks
  - Surface damage, foreign material
- Some approaches
  - Threshold by intensity
  - Rigid rotation/translation
  - Row and column sums
  - Intensity based triggers

# A Measurement Problem




- Given the picture
  - A movie containing hours of variations of this picture.
  - This picture is a calibration wafer on an xy table that is part of a laser-scribe.
- Measure the position of the "lanes"
  - Only approximate pixel size is known (~700 microns per pixel)
  - We don't care about each one, we care about their center location.
  - We don't know the y-position where it starts.
  - The camera goes on and off the die
- to the best accuracy and precision possible
  - Try to beat a pixel in resolution.

# General Approach

- Procedure:

  *Informed by Particle Image Velocimetry (Dr. Adrian, MAE 504)*

  - Smooth as needed
  - Find a “particle at each column of pixels in each image”
  - Convolute it (frequency domain) it with what it should be at the next column
  - Repeat until done with all images
- But
  - Skip images if information is redundant,
  - Don’t miss anything important
  - Make it run “fast” on the computer – we can’t wait weeks to process
- Desired form of results
  - Give statistics on the wafer, and on each “chip” as defined by “between intersections.
  - Give plots of centered data – we don’t care if it starts at pixel x, we care about how it changes position over the traverse of the table.

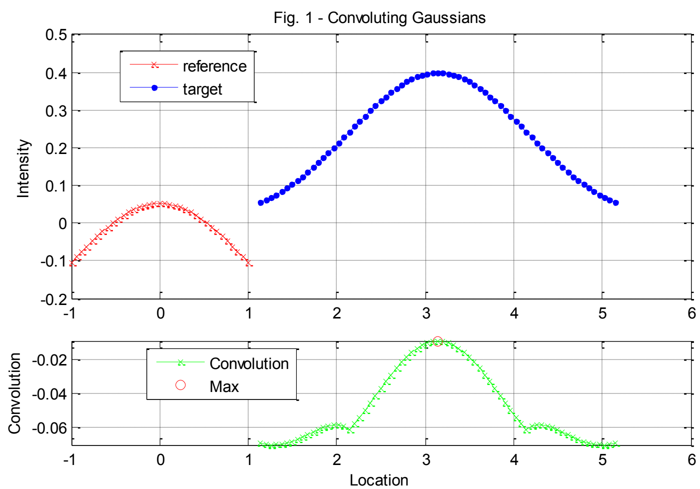
# General results

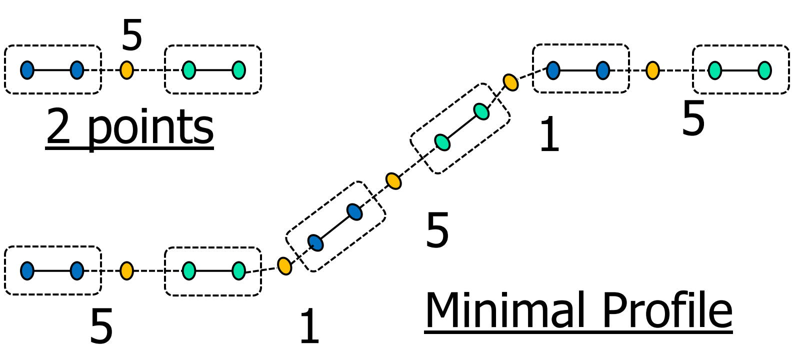
- Useful, but involved “voodoo”, “art” or other forms of non-science.
- Detailed (extensively) later in this presentation
- Drove me to explore, and that is a good thing.

I have been lucky enough to gain a reputation as a solver of exotic problems. This is useful because it brings me more exotic problems to solve than I could ever have managed alone.

# EXTREME SUB-PIXEL VISUAL METROLOGY

# Framework




*Informed by neurophysiology.*

## Tools

- Use Gaussians, it is a good basis function
- Use constants that I know [0, 1, pi, sqrt(2)] so I can test it. Make sure it can't be accidentally "perfect".
- Use common sample densities, traverse 17 – *human neural processing, recognition, physical intuition.*
- Test conclusions in "good" synthetic example (damped nonlinear spring)

# The Four Knobs

- Initial Sampling
  - How many patches of spatial averaged intensity traverse the domain?
  - Spatial discretization in image.
- Offset of intensities

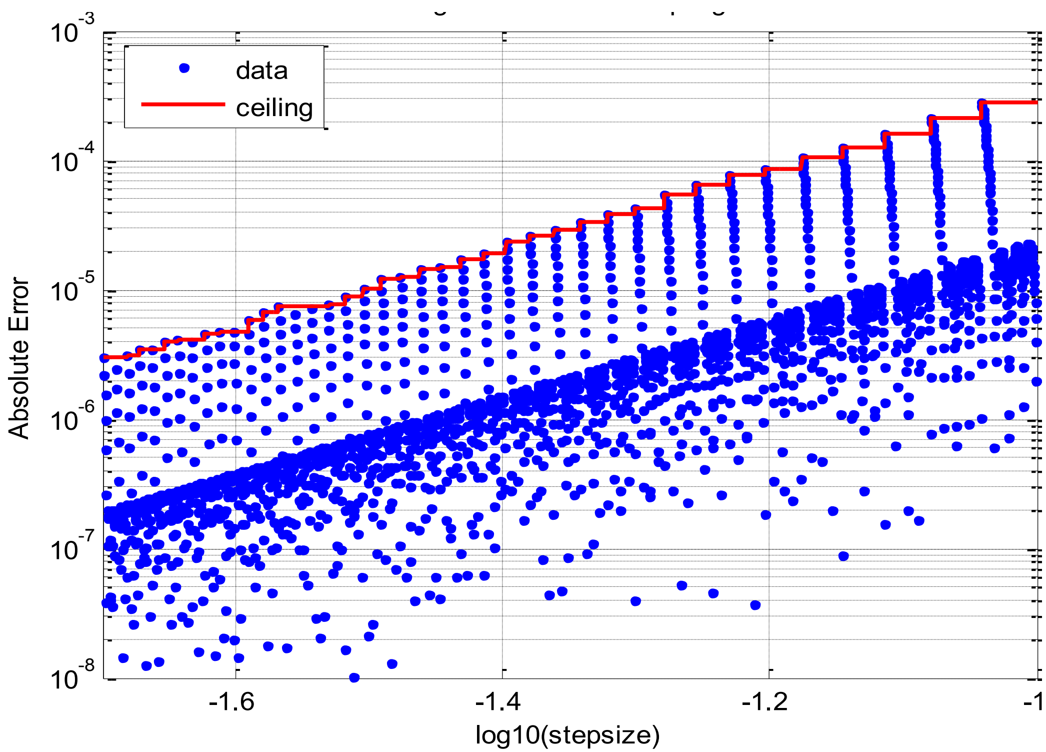
  Rigid translation in intensity profile domain.

  Somewhat non-physical for pdfs, but good calibration for image intensities.
- Resampling

  This improved the spatial discretization of convolution surprisingly.
- Smoothing

  to handle the intensity discretization.

# Knob: Initial Sampling Density

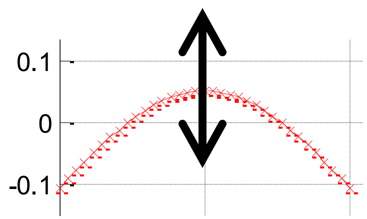



$$y = 0.2738 \cdot \Delta x^3$$

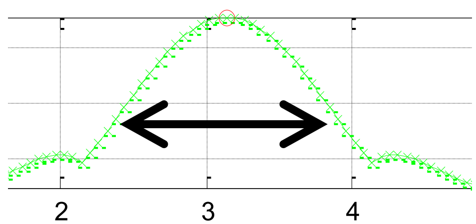
Observations

- Analysis is contrived in terms of minimizing maximum error.
- Max error decreases as cube of initial sampling.
- Mean error is about 16x smaller (1.2 decades)
- For 17 samples per reference the expected maximum error is 0.044%.
- Error in terms of pixel size is found by dividing the expression by $\Delta$x. Max Error per pixel is 0.378%

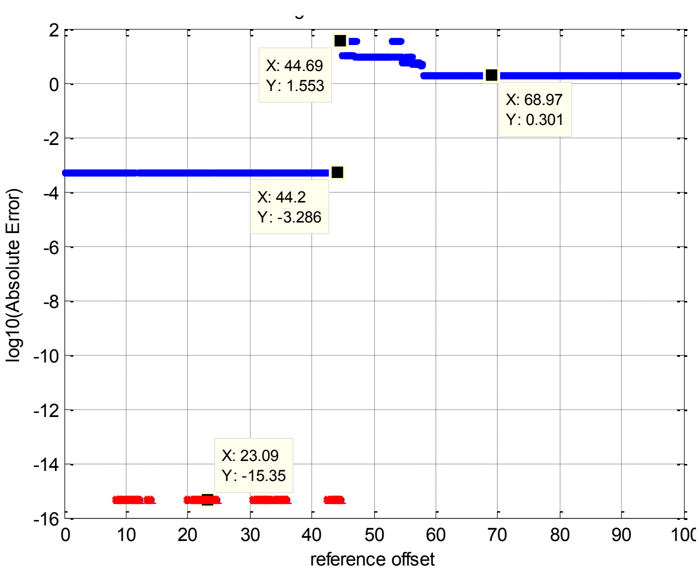
# Knob: Reference intensity offset

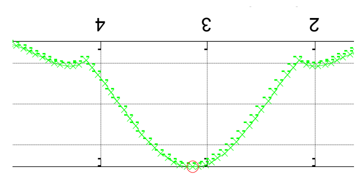



this




makes this



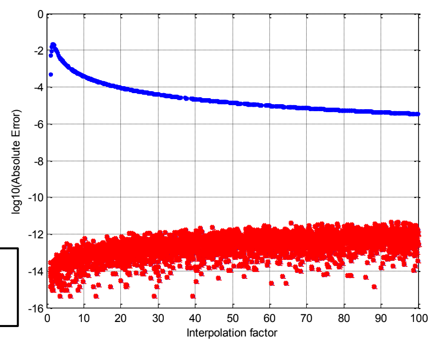



or this

Observations

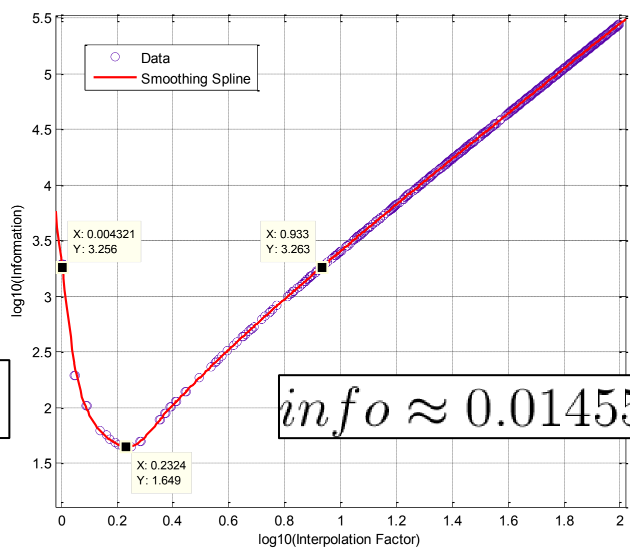
- Negative offset makes the convolution have more curvature (smaller central variation) than either input.
- The curvature affects how "quadratic" the top is, and makes the analytic root a better estimate of the true root.
- Offset was set at 23rd percentile (or less) of reference intensity.

# Knob: Interpolation Sampling Density




1

| Resampling Ratio | Scaled Values | |
|---|---|---|
| | Error | Information |
| 1.00 | 1.000 | 1.000 |
| 1.71 | 43.34 | 0.023 |
| 8.78 | 1.00 | 0.999 |
| 10 | 0.759 | 1.317 |
| 20 | 0.179 | 5.575 |
| 50 | 0.0278 | 36.014 |
| 100 | 0.00687 | 145.613 |



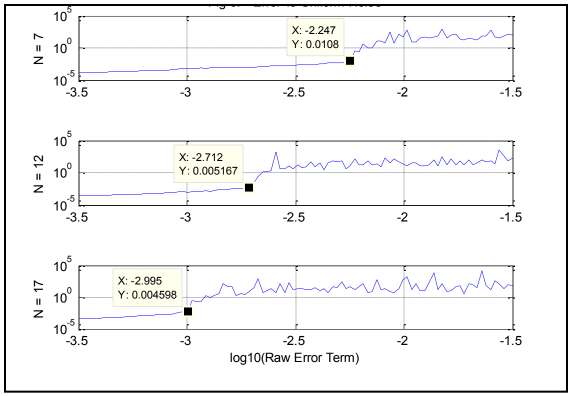

2

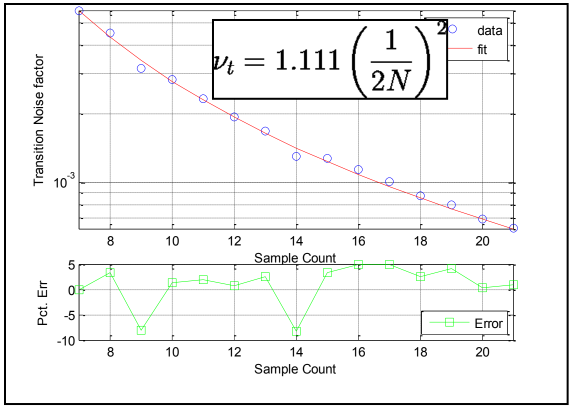
$$info \approx 0.01455 \cdot ratio^2$$

## Observations

- The analytic result has many "roundoff" results (shown in red).
- Information is defined here as inverse of Error.
- Dimensionless error is useful because it eliminates a need to convert to error per pixel.
- Critical value where resampling starts improving values is around 8.78.
- A resampling ratio of 10 was a good accident.

# Knob: Smoothing

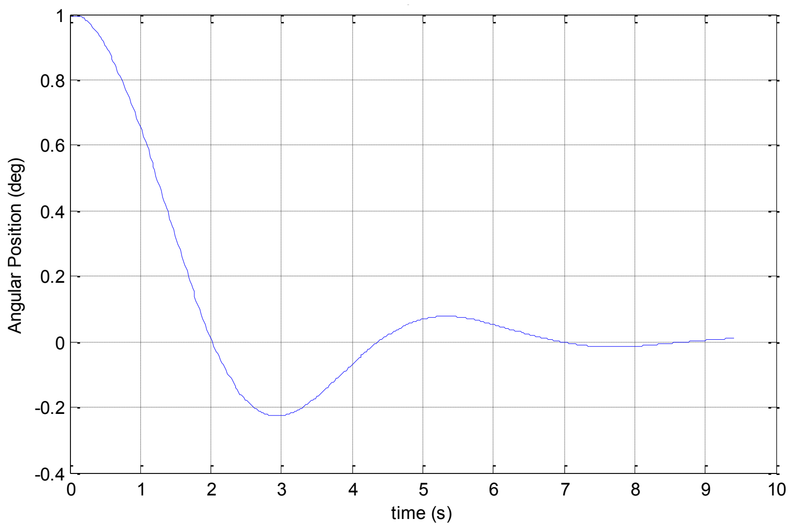


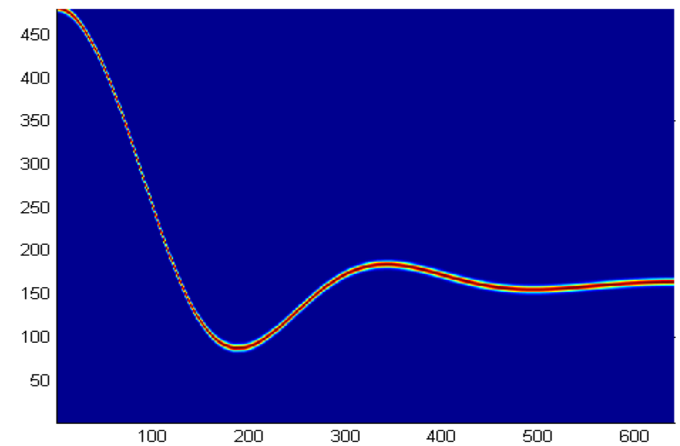




Observations:

- Added noise in this case was uniform, not centered. This is therefore an error-ceiling approach.
- The "nu" is the scaling factor multiplied to the noise term where the transition occurs.
- Discretizing the value from 64 bit (IEEE 784) to 8-bit (in images) is the same as adding noise.
- A small smoothing applied to the discretization (x≤6% loess) was found to "undo" the effect of adding the discretization noise.
- This should be further investigated for its implications in multi-precision computing.

# Synthetic Case: Setup




$$\ddot{\theta} + h\,\dot{\theta} + k\,sin\left(2\theta\right) = 0$$

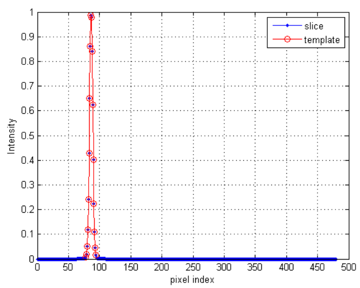
$$h = 1 \quad k = 1 \quad \theta\left(0\right) = 1$$

$$0 \leq t \leq 10\ (sec)$$

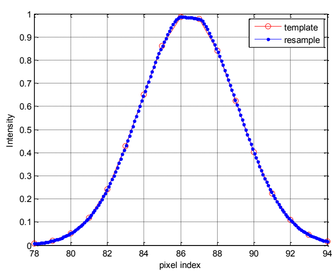
RK4/5 ='ode45'

- Each vertical slice is a Gaussian with mean equal to $\theta$ and standard deviation of 0.00825.
- The image has 480 rows and 640 columns.
- **Initial column is set to zero mean for calibration purposes.**

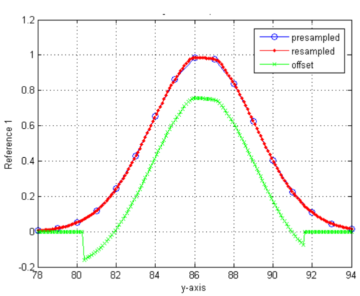
# Synthetic Case: Processing




17 samples



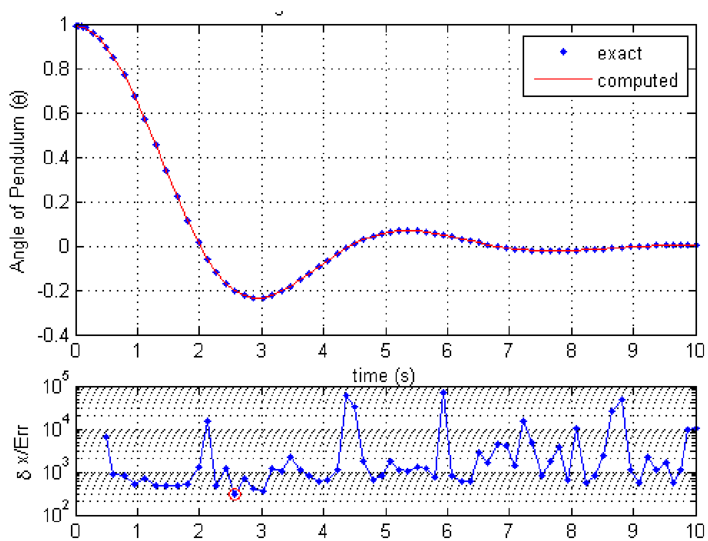

Ratio = 10:1




Offset = 23%

- Parameters are consistent with values determined above
- A Loess smooth of 2.3% was used between the raw sampling and the super-sampling.
- It was found useful to express error in terms of information per pixel.

# Synthetic Case: Results

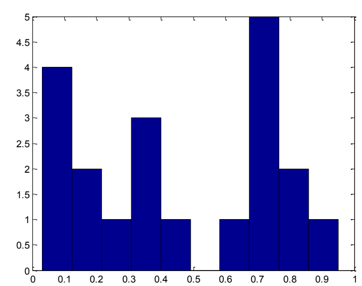



| | Statistics (Absolute valued) | |
|---|---|---|
| | Error | Δx/err |
| mean | 3.1259E-06 | 5694.7 |
| median | 2.7618E-06 | 1131.5 |
| std/sqrt(n) | 2.8534E-07 | 1633.9 |
| iqr/1.35 | 2.5036E-06 | 1324.6 |
| range | 9.9724E-06 | 68074 |

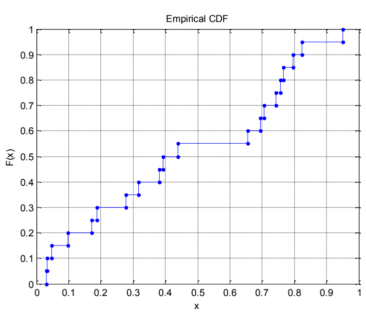
- The 640 lines resampled to the 66 time-steps from the numeric solver using Hermite interpolation
- Zero-crossings give error artifacts
  - Indicated by mean-median mismatch and iqr vs. population stdev estimate mismatches.
  - Visible in error subplot – outliers are all toward right side of plot.
- Max error (red circle) at 0.32%. Median error was much better.
- In 66 samples this compares well to the expected ceiling of 0.37%

# GLOBAL SMOOTHING

# Smoothing: First problem




trimodal?
0, 0.3, 0.75



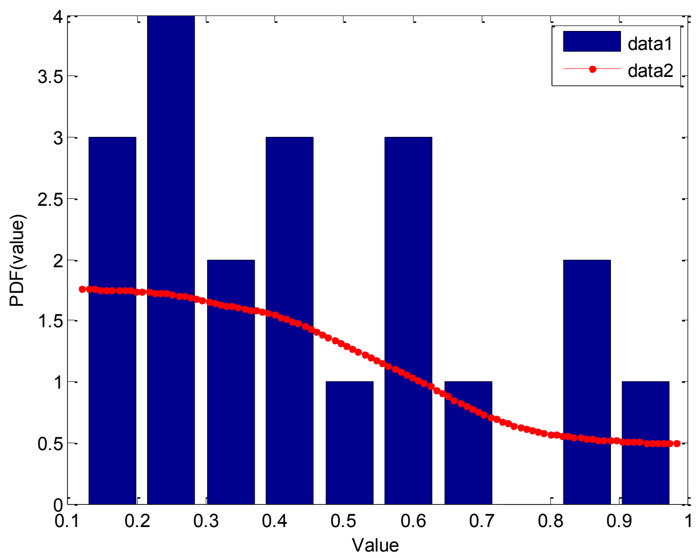

Not as multimodal
Flat between modes

- Small sample size and histograms don't work well together.
- Empirical CDF works better
  - Centered errors cancel
  - Overall trend is easier to perceive.

Uniform Random
[0, 1]
20 samples

Both
Actual

# AIC+Spline on CDF: The approach




- Convert data to cumulative domain
- Use Akaike Information Criteria (AIC) to find best smoothing value.
- Interior minimum is “best”.
- Take analytic derivative to convert the fit to non-cumulative domain.
- Model is in Cumulative domain
- Cubics go well with many CDF’s
  - Handles the tails
  - Taylor series error at fourth-order term

Bottom line: it is a model-optimal sanity check on my histogram.

# Akaike Information Criterion (AIC)

- One of many "Information Criterion"
- Derived from Kullback-Leibler divergence.
- Has very useful form (takes inputs that are convenient outputs)
- Minimum AIC indicates "best" candidate model.

$$AIC(\nu) = n \cdot \left( \frac{RSS(\nu)}{n} \right) + 2 \cdot k(\nu)$$

RSS = Sum of Squared Error

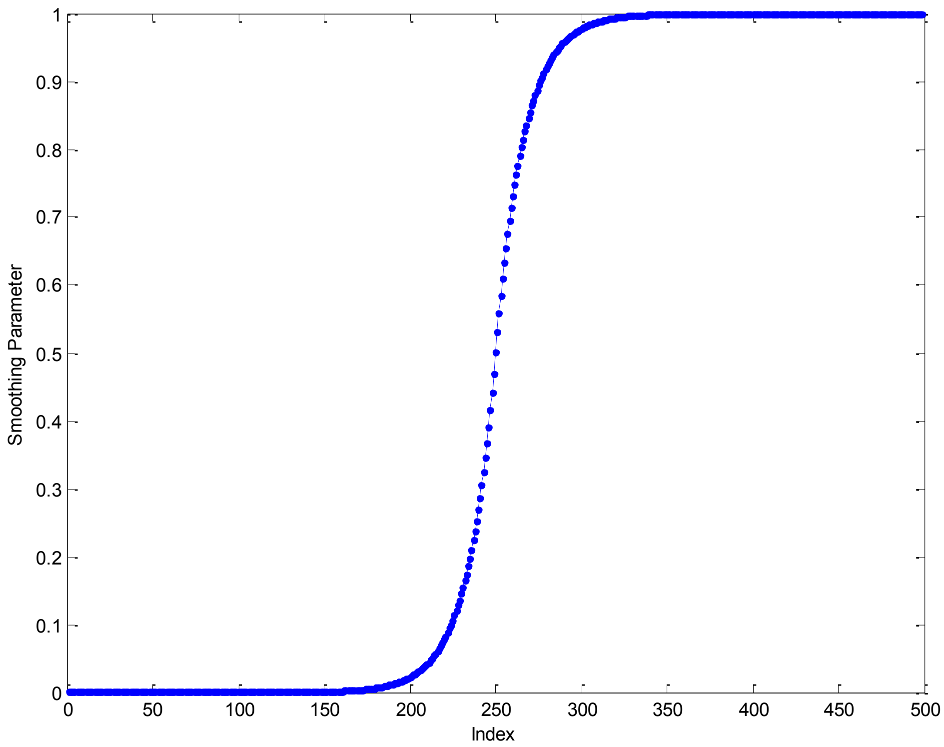
K = number of parameters in model

n = number of samples

$\nu$ = smoothing parameter used in spline

H. Akaike, "An information criterion (AIC)", Math. Sci., 14(153):5-9 (1976).

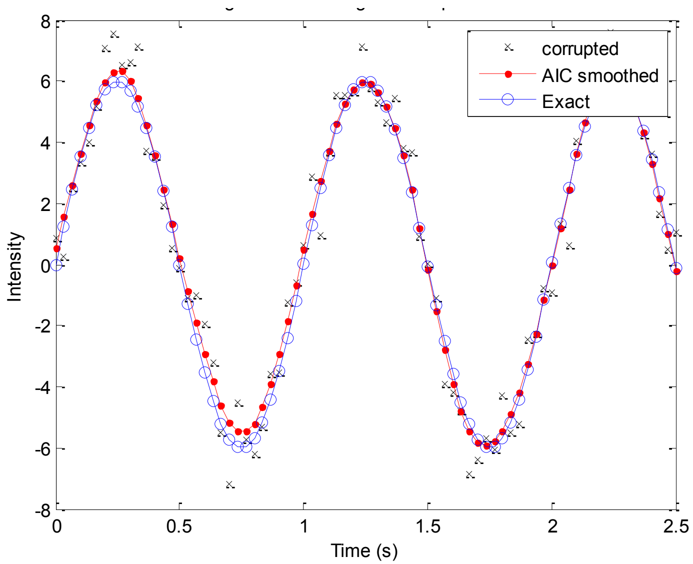
# About the Parameter




Observations

- Not uniformly sampled because experience in CDF's taught me all the action happens at the end.
- Density is "high enough" to "sufficiently" characterize domain.

# Smoothing: Analytic Model




$$y = 6sin(t \cdot 2\pi) + \epsilon$$

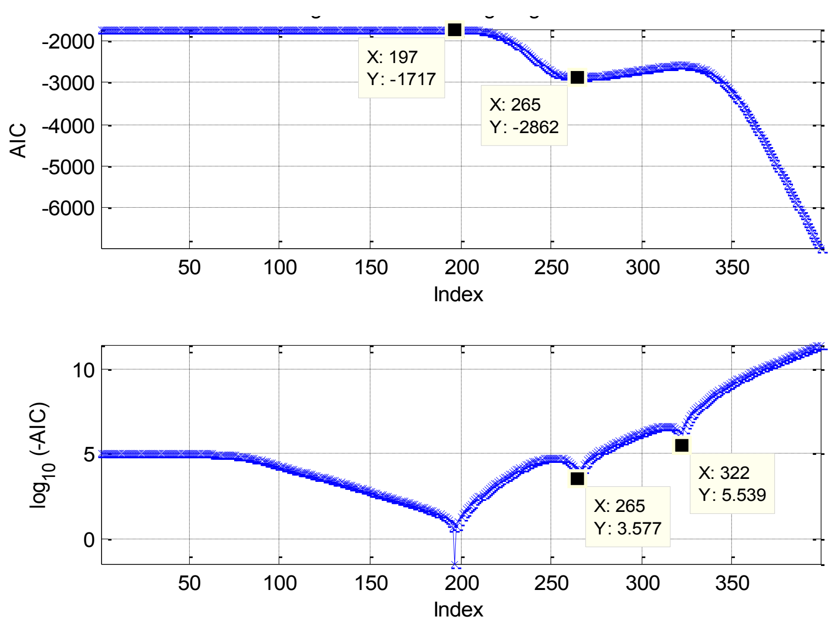
$$\epsilon \sim N(0, 1)$$

$$0 \leq t \leq 30$$

- Sampling
  - There are 30 samples taken per second
  - There repetitions of the cycle that are sampled.
  - This is required to characterize "capability".
- Statistics folks prefer 30 as a minimum sample size.

http://www.itl.nist.gov/div898/handbook/mpc/mpc.htm

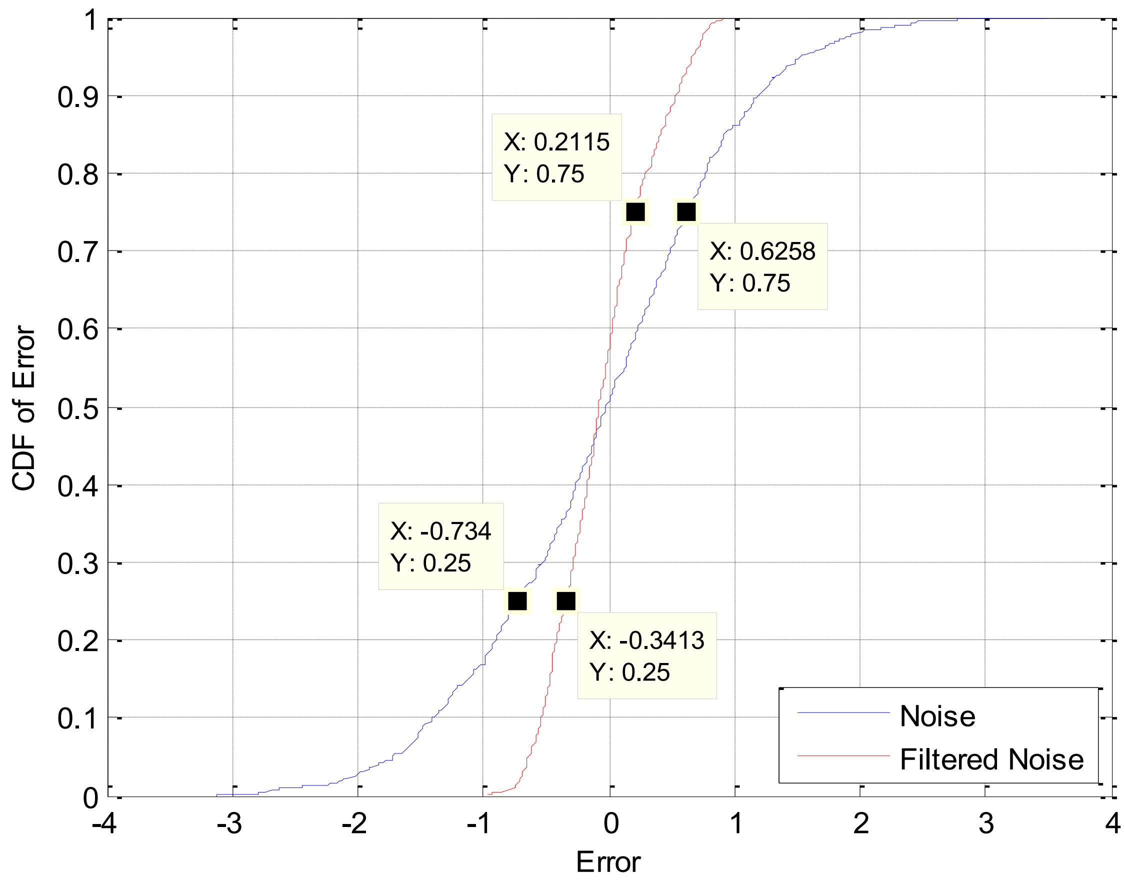
# Smoothing: Finding the AIC




- Model is in non-cumulative domain so AIC is calculated there.
- There is a clear interior minimum at 265
- If treated in cumulative domain the "drunkards-walk" is confounded with the "model".
- Interior is critical – the right end perfectly interpolates and is a false-positive.
- AIC derived parameter is 0.039% off of a least-squares reduction to the exact one – its effectively identical

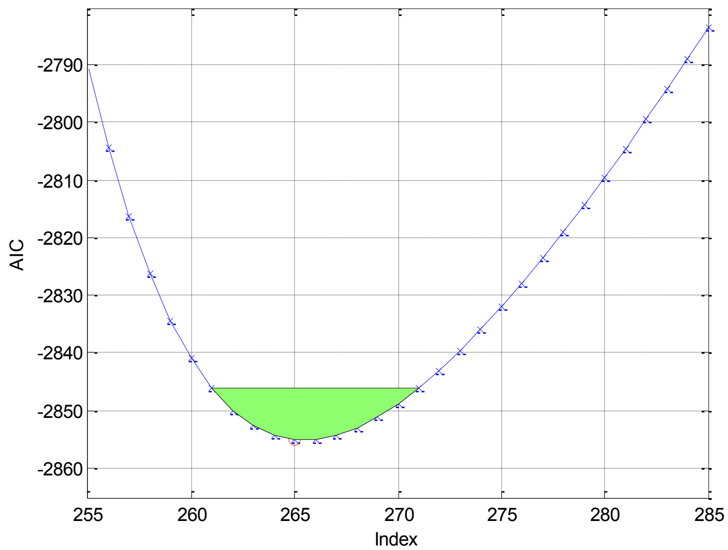
# Smoothing: Analytic Model

- Mean of error is increased (from 0)
- Variation is decreased by 59.3%
- Decrease is non-uniform, big errors are fixed more than small ones.



# MODEL VARIATION

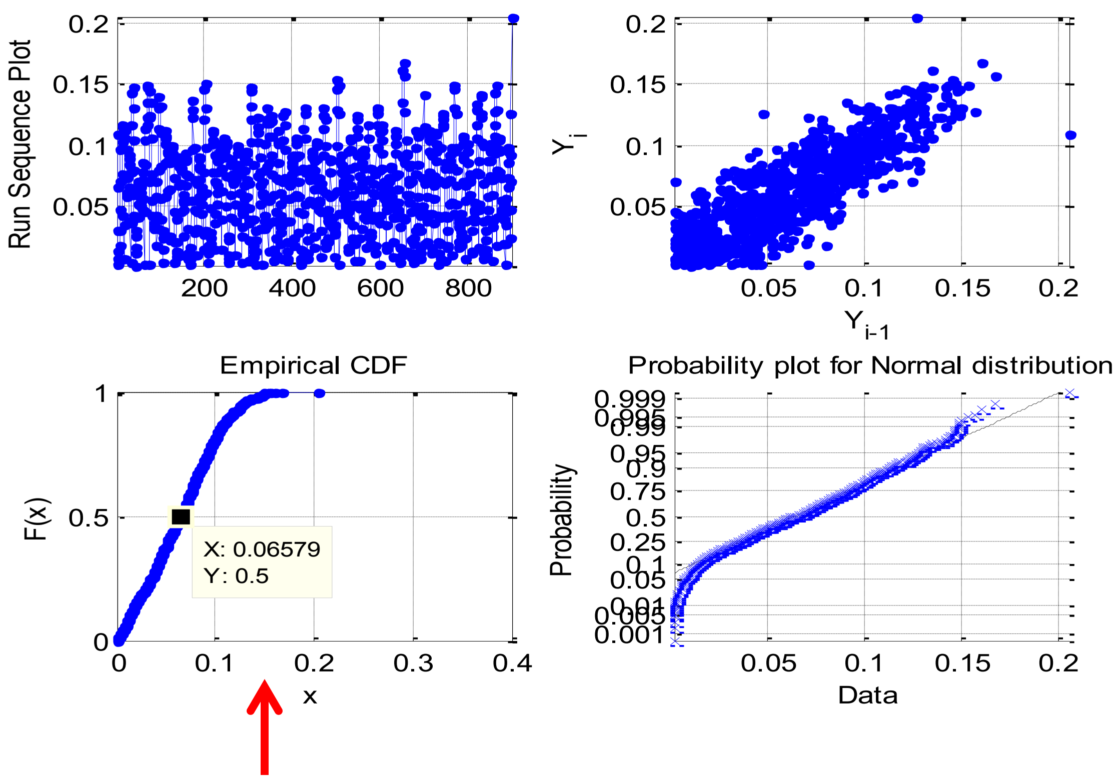
# Variation in the model




$$L\left(model|data\right) \propto \exp\left(-0.5\frac{AIC\left(\nu\right)}{AIC\left(\nu^{*}\right)}\right)$$

- AIC is an approximation of the log-likelihood.
- There is a "puddle" of candidate models.
- Using the ratio of AIC to optimal AIC (Akaike weights) the following set the size of the puddle:
  - 0 to 2 : substantial support
  - 4 to 7 : considerably less support
  - Over 10 : essentially no support
- This family of candidate models describes the model uncertainty

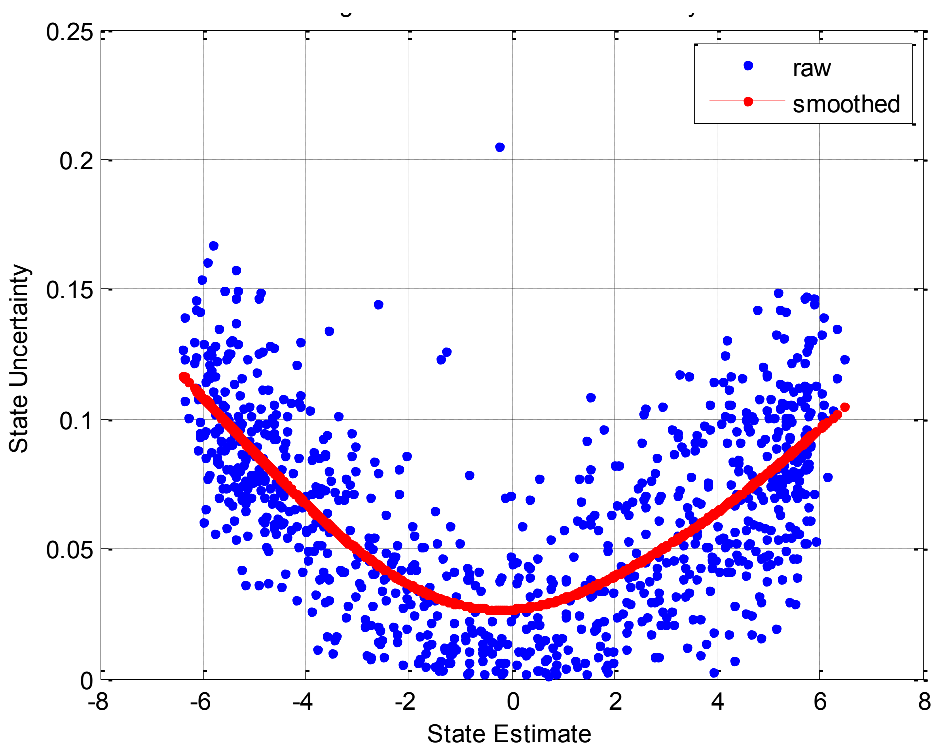
# Evaluated over domain (EDA)




Replace Histogram with eCDF

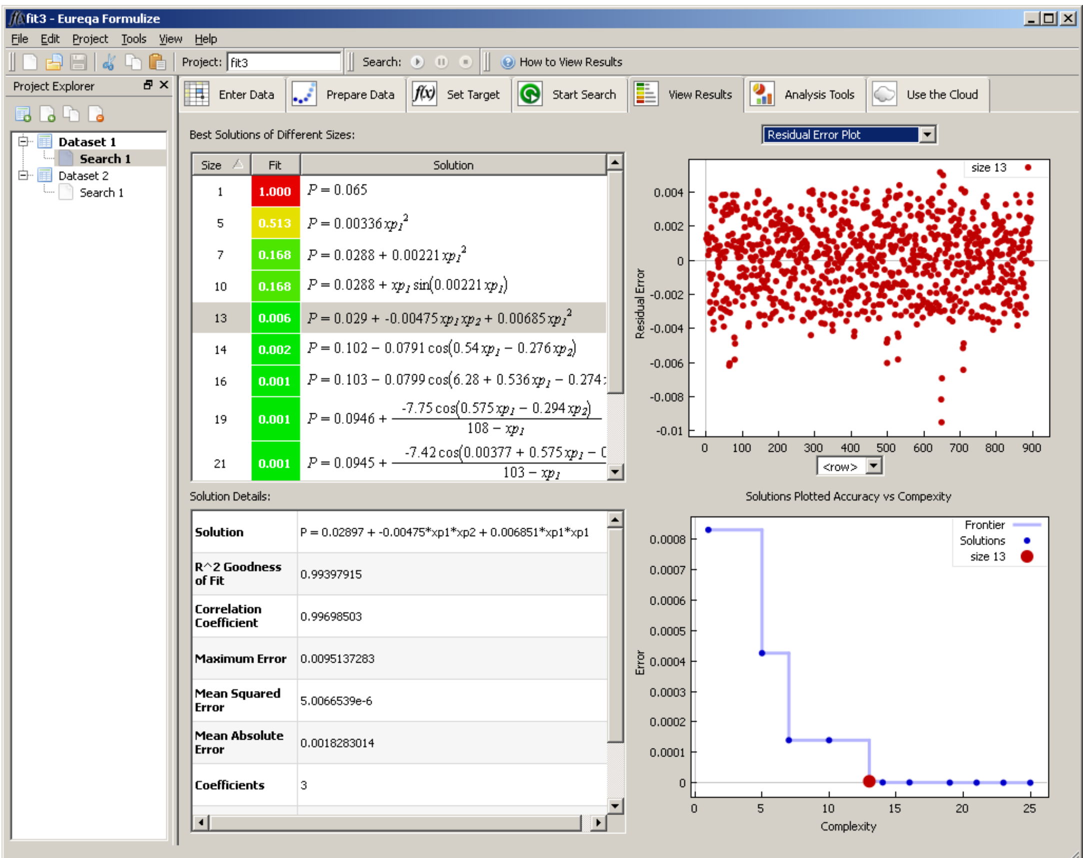
- Used 200 candidate models uniformly spaced across puddle.
- Computed standard deviation of ensemble at every point
- Observations
  - High correlation in the lag plot
  - Exponential-like distribution in eCDF

# As a function of State



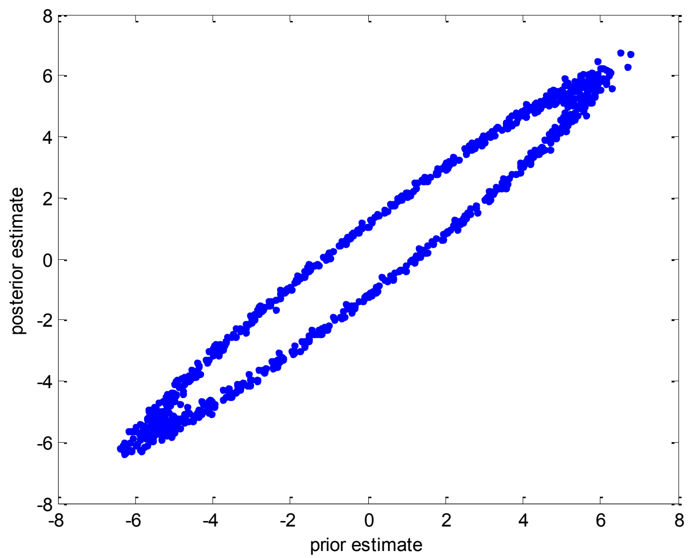

- Informed by General form of Kalman filter (to follow)
- Reasonable trend suggests use of smoother
- Smoothed form looks quadratic
- Smoothed for is analytic already, but using a fit tool can make it more user-friendly.

# Simpler Variance function




- Inputs were last 4 state variables: x(t-1)… x(t-4)
- Output variable was smoothed state standard deviation

# State Estimate



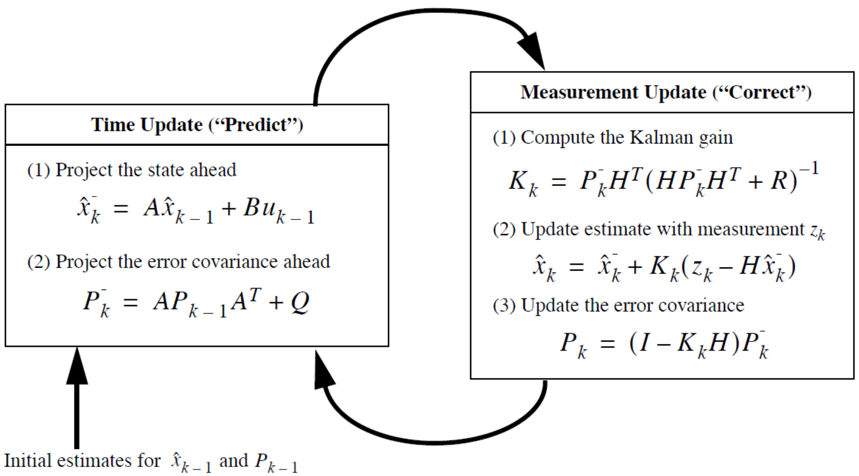

$$\hat{x}_k^- = 1.95621 \cdot \hat{x}_{k-1}^+ - \hat{x}_{k-2}^+, \quad R^2 = 99.993\%$$

- Vertical line test is failed
  - Non-Markov function
  - Needs two priors to specify state
- Using Eureqa/Formulize a simpler (non-spline) analytic form was found.
- $R^2$ is the fit statistic, not the variance.

# Kalman Filtering

Taken From Welch (2006)

http://www.cs.unc.edu/~welch/kalman/




$$E[(x_k - \hat{x}_k)(x_k - \hat{x}_k)^T] = P_k$$

$$p(w) \sim N(0, Q) \qquad x_k = Ax_{k-1} + Bu_{k-1} + w_{k-1}$$

$$p(v) \sim N(0, R) \qquad z_k = Hx_k + v_k$$

- Process:
  - Predict using model
  - Update estimate using measurement
- Assumptions:
  - Markov – last state estimate holds all information needed.
  - Uncorrelated noise
  - Update model is good

# Our System

Filter Parameters

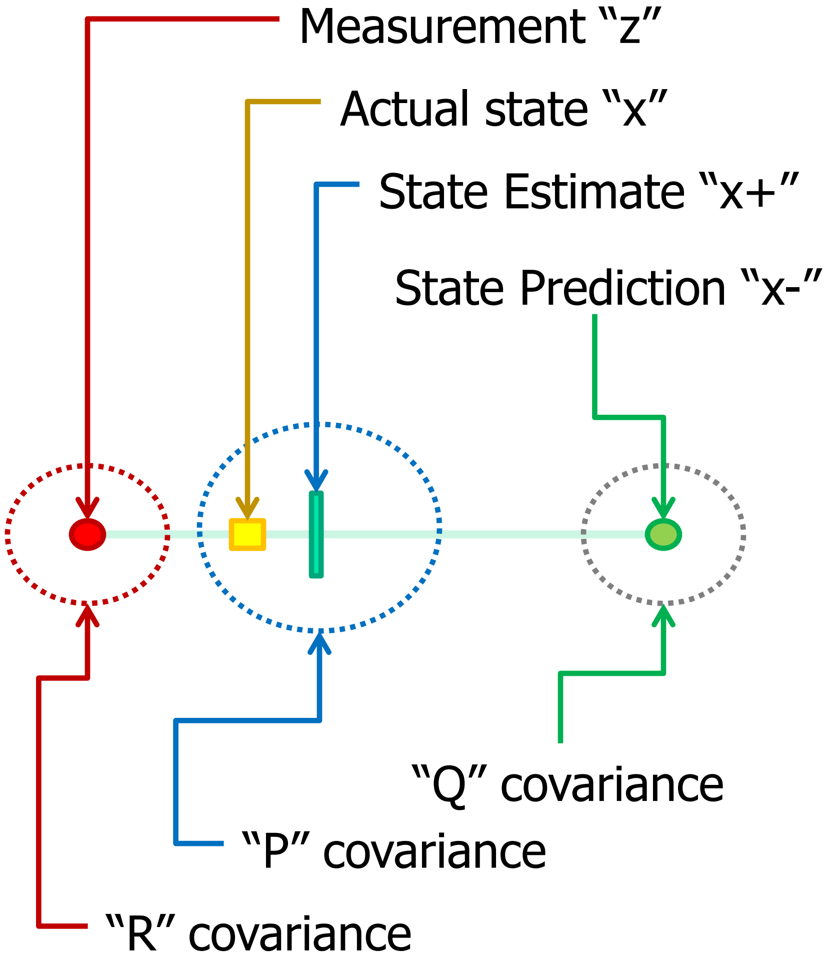
1. H = 1
2. B = 0

Reasoning

1. Direct state measurement
2. This Parameter estimation does not require control input.

Assumptions:

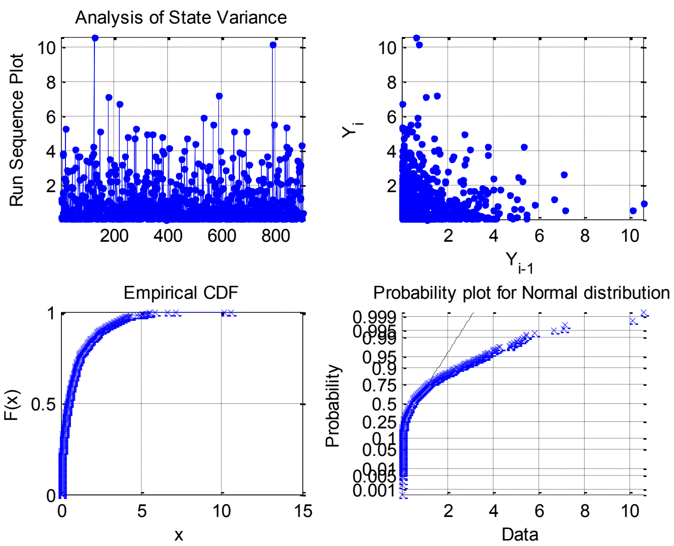
1. All new information is provided in most recent step.
2. State update function sufficiently describes underlying physics
3. Noise is
   a) Centered
   b) Uncorrelated
4. Underlying state is interior to span between model prediction and measurement



# Variation due to Measurement

$$z_k = Hx_k + v_k \longrightarrow z_k - Hx_k = \nu_k$$

$$R = E\left(\nu_k \nu_k^T\right) \longrightarrow R = E\left(z_k - Hx_k\right)$$

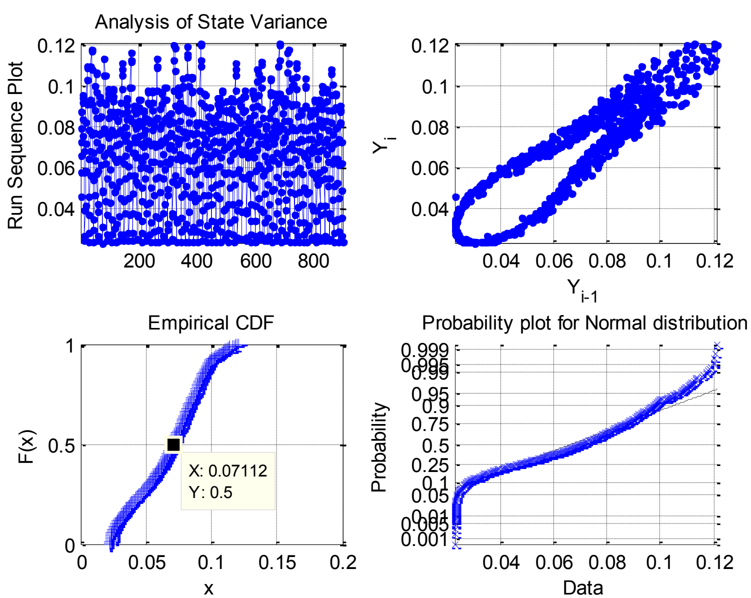



- The measurement process introduces noise and discretization error.
- The measurement variance is shown on the left.
- EDA within the lag plot shows in log-log scale the distribution is gaussian.
- The mean, or expectation, of this is 0.88402.

# The Kalman Gain

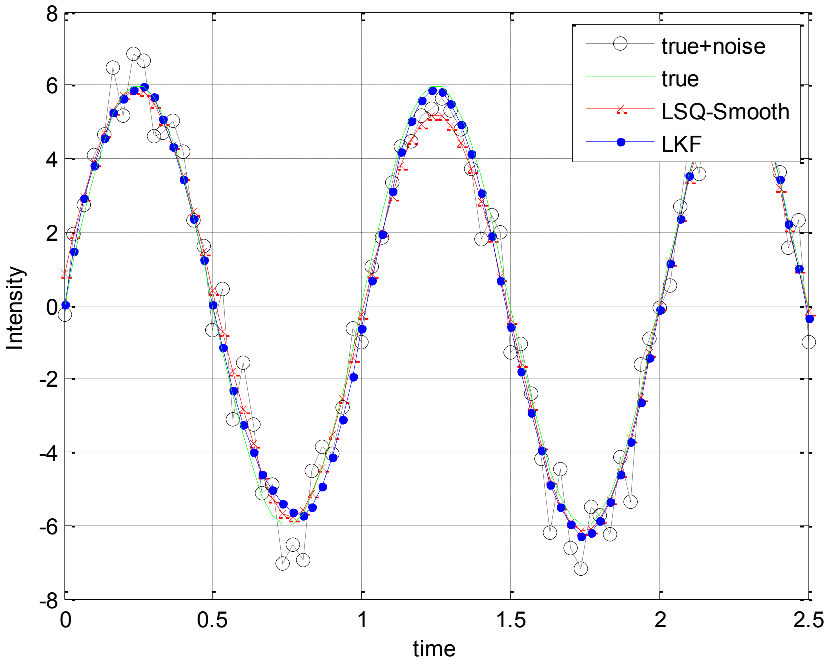
$$K = P_k H^T \left( H P_k H^T + R \right)^{-1}$$

$$K_k = P_k^- \left( P_k^- + 0.88402 \right)^{-1}$$




- The function is simplified using our expressions for "R" and "P"
- The minus means "prior", a before-estimate. Plus would be "posterior" or "after".
- Lag function suggests deterministic model in x (unsurprisingly)

# Empirical Kalman Filter Results




- Over an 10 runs the mean reduction in norm of errors using
  - the AIC fit was 60.7% ± 1.2%
  - The Kalman Filter 34.6% ± 2.4%
- However the Kalman filter generalized the behavior better, and stages for other KF-derived tools (like smoothers and data assimilation)

The really important part: It worked reasonably well "out the gate" without tuning.

# RESULTS

# Case 1: Lane Tracking

Microsoft Office
owerPoint Presentatio

Brief Embedded PowerPoint

# Case 2: Ball Tracking

Microsoft Office
owerPoint Presentatio

Brief Embedded PowerPoint

# Summary

- A method has been demonstrated that is useful for extreme sub-pixel measurement and is model agnostic in terms of the nature of the feature profile, in 1d or 2d. Its real-world results are compatible with analytic cases.
- A method has been demonstrated that is useful for removing the noise from the signal and is model-agnostic to the underlying model. This method was extended into a framework for approaching the model uncertainty, and was applied to in a demonstration of a Kalman Filter.

# QUESTIONS?