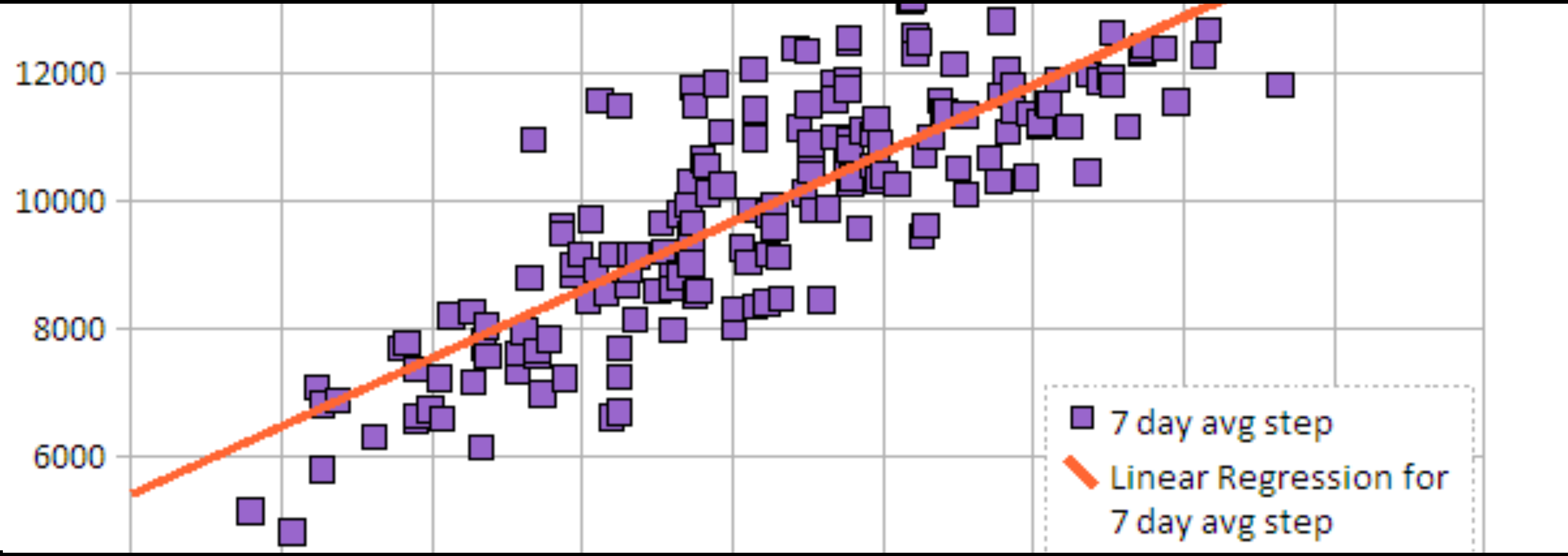
E80

Experimental Engineering

Lecture 2

Data Fitting



Jan. 24, 2013

Christopher M. Clark

Outline

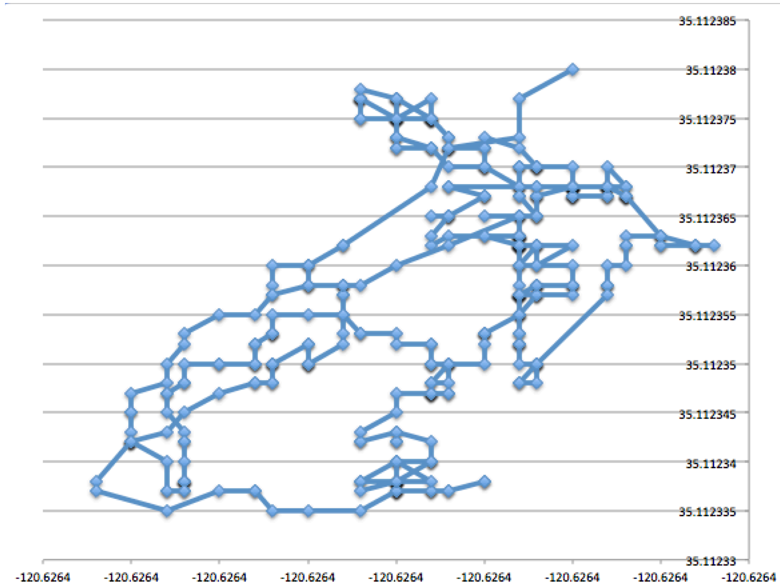
- Confidence in Measurements
- Linear Regression
- Error Propagation
- Quantization Error

Confidence in Measurements

- Consider the GPS measurements of my AUV…

Confidence in Measurements

- Zoom in, and you can see the AUV when it isn't moving



What is the AUV's location?

Confidence in Measurements

- When we take measurements, we want to know how "good" they are.
  - We establish measures of confidence in our measurements
  - E.g. confidence limits, standard deviation, variance, etc.

Confidence in Measurements

- Lets just consider the case where we want to estimate the **actual** longitude

$$\mu_x$$

- Consider N **measurements** of longitude

$$x_1, x_2, x_3, \ldots x_N$$

Confidence in Measurements

- For these N measurements, the **sample mean** is

$$\bar{x} = \frac{1}{N}\sum_{i=1}^{N} x_i ,$$

- We define the **residual error**, for each measurement, to be

$$e_i = x_i - \bar{x}$$

Confidence in Measurements

- We characterize our residual errors using the **sample variance**:

$$S^2 \equiv \frac{1}{N-1}\sum_{i=1}^{N} e_i^2 = \frac{1}{N-1}\sum_{i=1}^{N}\left(x_i - \bar{x}\right)^2$$

- The **sample standard deviation** can be defined as:

$$S = \sqrt{S^2}$$

Confidence in Measurements

- The **sample variance** S^2 characterizes the spread of the measurements.
- We estimate how far the sample mean is from the actual value using the **estimated standard error**:

$$ESE = \frac{S}{\sqrt{N}}$$

Confidence in Measurements

- Lets get back to our AUV's GPS measurements of longitude. Here are N measurements:

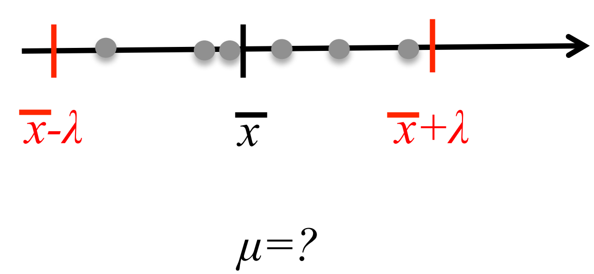
- The corresponding sample mean, sample variance, and estimated standard error are:

$$\bar{x} = -120.626368 \quad S = 7.71967E\text{-}06 \quad ESE = 3.8265E\text{-}07$$

Confidence in Measurements

- We can use the **estimated standard error**, to determine confidence intervals.
- That is, we can say

μ is within $+/-\ \lambda$ of $\bar{x}$
with $P\%$ confidence



Confidence in Measurements

- For example:

actual longitude is within *+/- 0.00000075°* of *-120.626368°*
with *95%* confidence

Confidence in Measurements

- Where does $\lambda = 0.00000075$ come from?

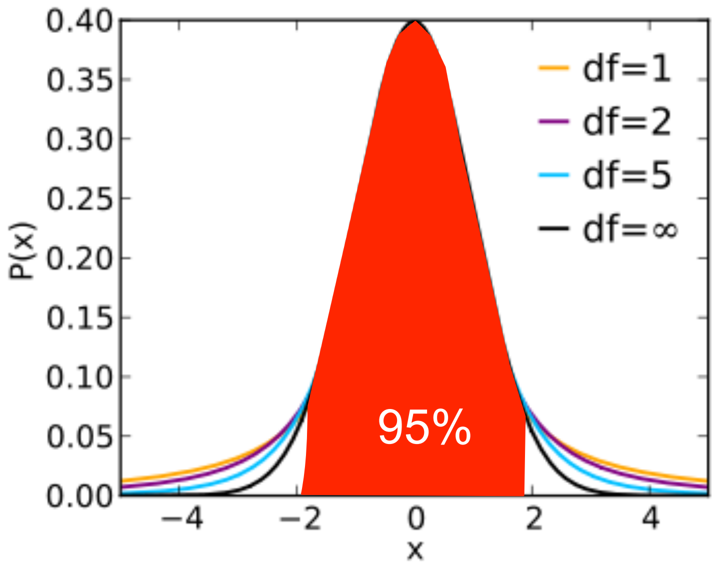
$$\lambda = ESE*t$$

- "Students t value"
  - The value of t relates the confidence interval to the area under a standard distribution

Confidence in Measurements

- What is t ?
  - Represents the limits of a standard distribution that encapsulates $1\text{-}P$ of area under the distribution
  - The df are the degrees of freedom, which is the number of samples N minus number of parameters estimated

SIGNIFICANCE LEVEL FOR TWO-TAILED TEST						
df	.20	.10	.05	.02	.01	.001
1	3.078	6.314	12.706	31.821	63.657	636.619
2	1.886	2.920	4.303	6.965	9.925	31.598
3	1.638	2.353	3.182	4.541	5.841	12.941
4	1.533	2.132	2.776	3.747	4.604	8.610
5	1.476	2.015	2.571	3.365	4.032	6.859
10	1.372	1.812	2.228	2.764	3.169	4.587
20	1.325	1.725	2.086	2.528	2.845	3.850
30	1.310	1.697	2.042	2.457	2.750	3.646
40	1.303	1.684	2.021	2.423	2.704	3.551
60	1.296	1.671	2.000	2.390	2.660	3.460
120	1.289	1.658	1.980	2.358	2.617	3.373
∞	1.282	1.645	$\boxed{1.960}$	2.326	2.576	3.291



Confidence in Measurements

- Summary:
  1. Calculate your mean $\bar{x}$
  2. Calculate your estimated standard error ESE
  3. For a given df and significance level $= 1 - P$, find t from table
  4. Calculate $\lambda = ESE * t$

Outline

- Confidence in Measurements
- Linear Regression
- Error Propagation
- Quantization Error

Linear Regression

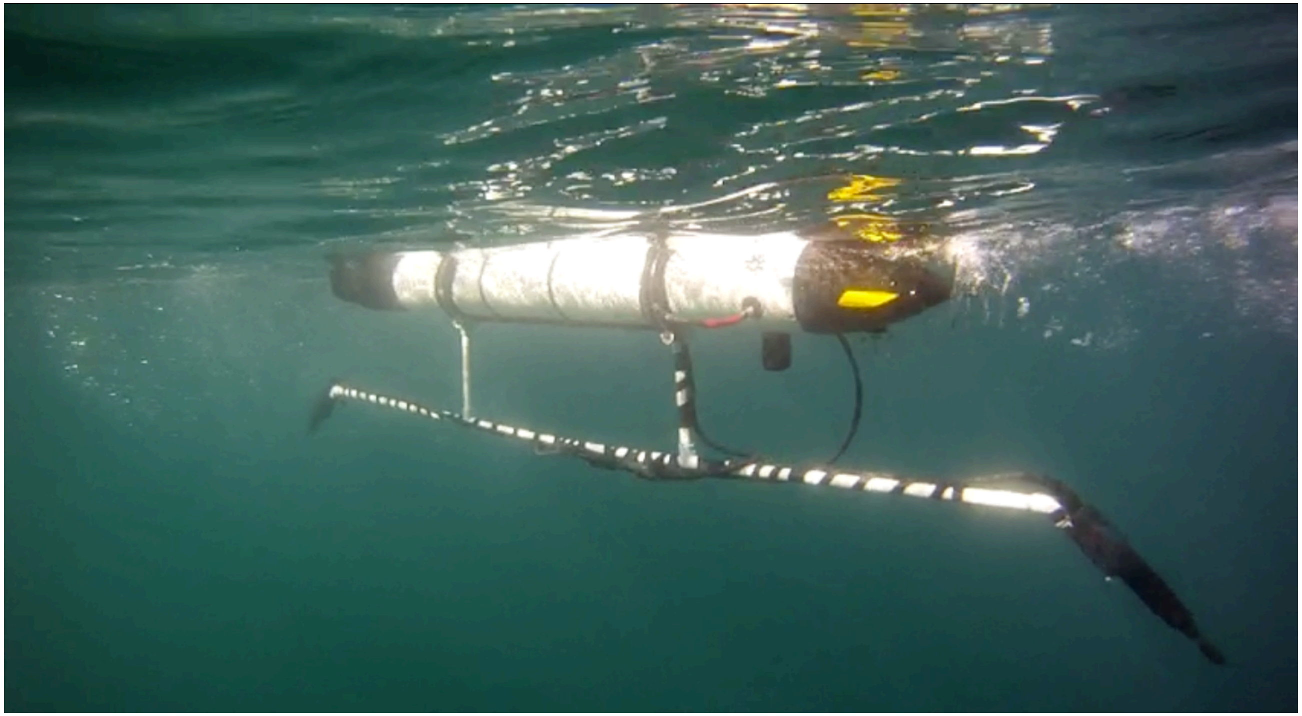
- Sometimes we measure one variable x, but are interested in another variable $y = f(x)$

- Often, $f()$ is assumed to be linear

$$y = \beta_0 + \beta_1 x$$

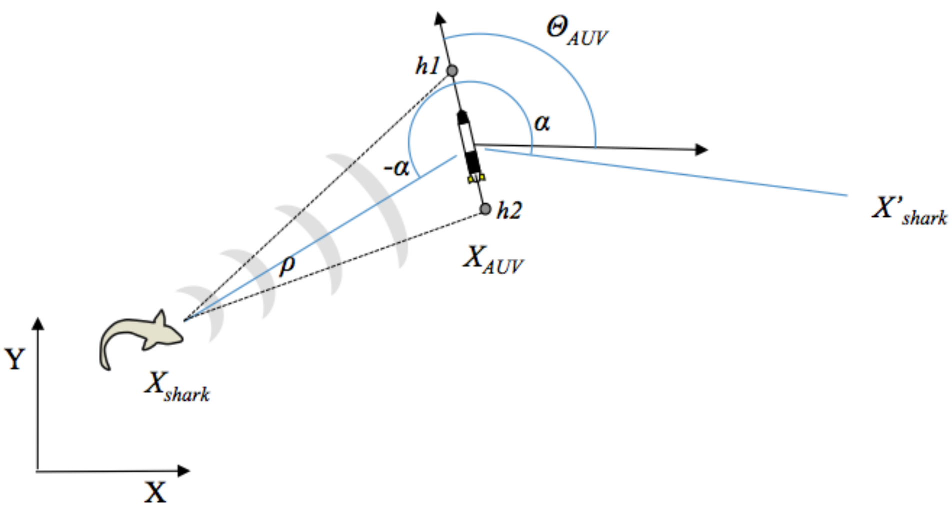
Linear Regression

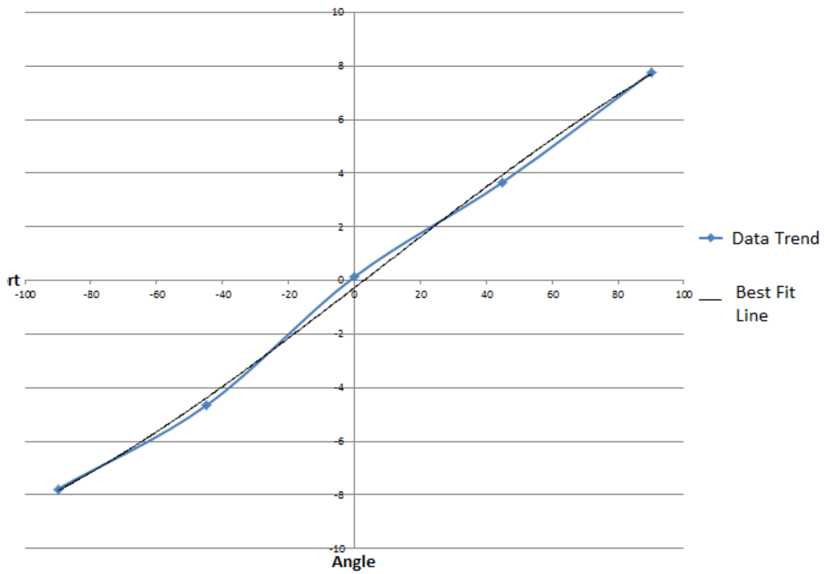
- Shark tracking Example…



Linear Regression

- Shark tracking Example…





Linear Regression

- We usually must estimate the coefficients β_0 and β_1. from a data set:

$$(x_1, y_1), (x_2, y_2), \dots (x_N, y_N)$$

- Our model becomes

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x$$

Linear Regression

- To estimate β_0 and β_1 we minimize the sum of squared errors:

$$SSE = \sum_{i=1}^{N} e_i^2 = \sum_{i=1}^{N} \left[y_i - \left(\hat{\beta}_0 + \hat{\beta}_1 x_i \right) \right]^2$$

- This minimization results in

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} \qquad \hat{\beta}_1 = \frac{\sum_{i=1}^{N} (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{N} (x_i - \bar{x})^2}$$

Linear Regression

- How much confidence do we have in $\hat{\beta}_0$ and $\hat{\beta}_1$?

$$S_e = \sqrt{\frac{SSE}{N-2}} = \sqrt{\frac{\sum_{i=1}^{N} e_i^2}{N-2}}\ .$$

Linear Regression

- How much confidence do we have in $\hat{\beta}_0$ and $\hat{\beta}_1$?

$$S_{\beta_0} = S_e \sqrt{\frac{1}{N} + \frac{\bar{x}^2}{\sum_{i=1}^{N}(x_i - \bar{x})^2}}.$$

$$S_{\beta_1} = S_e \sqrt{\frac{1}{\sum_{i=1}^{N}(x_i - \bar{x})^2}}.$$

$$\lambda_{\beta 0} = t\, S_{\beta 0}$$

$$\lambda_{\beta 1} = t\, S_{\beta 1}$$

Linear Regression

- How much confidence do we have in y ?

$$S_y = S_e \sqrt{\frac{1}{N} + \frac{(x-\bar{x})^2}{\sum_{i=1}^{N}(x_i-\bar{x})^2}}$$

$$\lambda_y = t\, S_y$$

Linear Regression

- Quick Summary:
  - Given a set of (x, y) pairs, we can calculate

1. The coefficient estimates $\hat{\beta}_{0,}\hat{\beta}_1$ of the linear regression
2. The confidence limits $\lambda_{\beta 0,}\lambda_{\beta 1}$ on the coefficients
3. The confidence limits λ_y on the y values

Outline

- Confidence in Measurements
- Linear Regression
- Error Propagation
- Quantization Error

Error Propagation

- Given a function $F(x, y, z, \ldots)$, and known error in variables $x, y, z, \ldots$, what is the error in F ?

Error Propagation

- Lets assume that errors are small and that we can make a first order Taylor series approximation:

$$F - F_{true} = \frac{\partial F}{\partial x}\left(x - x_{true}\right) + \frac{\partial F}{\partial y}\left(y - y_{true}\right) + \frac{\partial F}{\partial z}\left(z - z_{true}\right) + \cdots .$$

Error Propagation

- For errors $\varepsilon = x - x_{true}$, that are systematic, known, and small, we can rewrite as:

$$\varepsilon_F = \frac{\partial F}{\partial x}\varepsilon_x + \frac{\partial F}{\partial y}\varepsilon_y + \frac{\partial F}{\partial z}\varepsilon_z + \cdots .$$

Error Propagation

- If errors of $x, y, z, \ldots$ are independent random variables, then standard errors are assumed related by:

$$\varepsilon_F = \sqrt{\left(\frac{\partial F}{\partial x}\right)^2 \varepsilon_x^2 + \left(\frac{\partial F}{\partial y}\right)^2 \varepsilon_y^2 + \left(\frac{\partial F}{\partial z}\right)^2 \varepsilon_z^2 + \cdots}$$

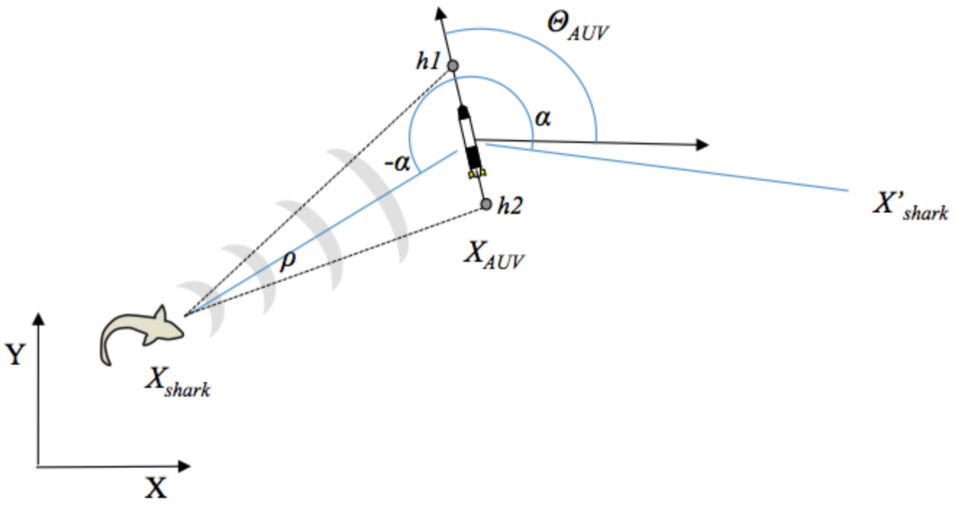
Error Propagation

- Example:
  - We model the range to a shark tag ρ as a function of the strength of the received acoustic signal s.

$$\rho = K_s s^a$$

where

$a < 1$ is constant



Error Propagation

- Example cont':
  - If we know the sample variance S_s^2 in signal strength measurements, and the variance S_K^2 in K_s, we can calculate the corresponding variance in range S_ρ^2

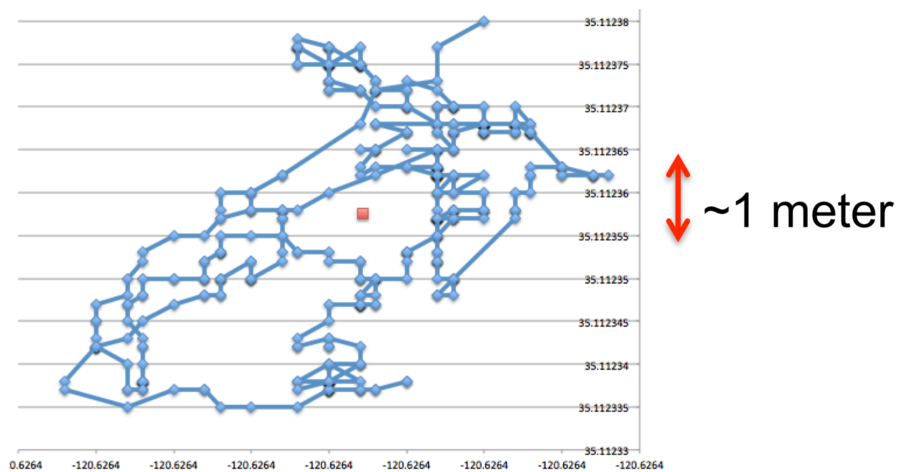
$$S_\rho^2 = (d\rho / ds)^2 S_s^2 + (d\rho / dK_s)^2 S_K^2$$
$$= (aK_s s^{a-1})^2 S_s^2 + (s^a)^2 S_K^2$$

Outline

- Confidence in Measurements
- Linear Regression
- Error Propagation
- Quantization Error

Quantization Error

- Lets revisit the static AUV plot of positions…



We have ~0.1 meter resolution

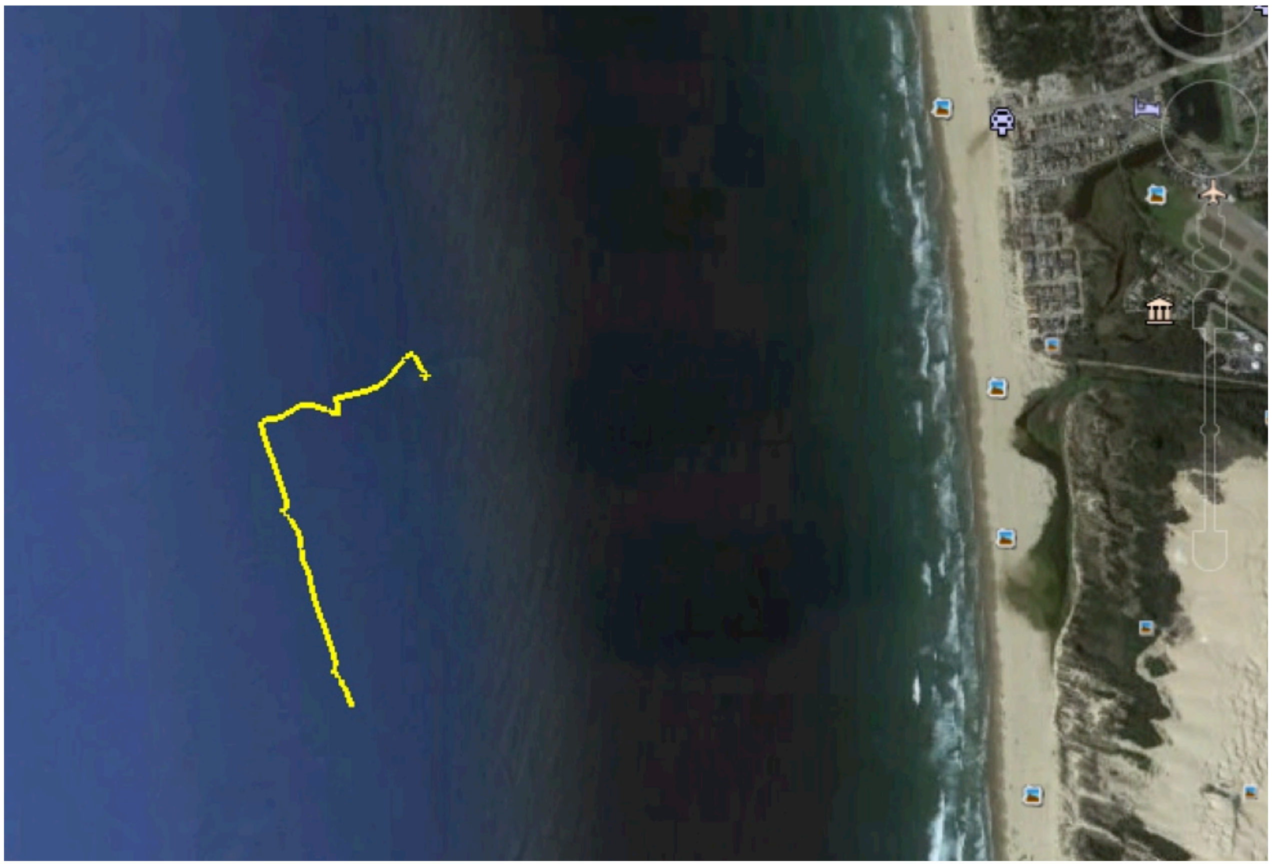
Quantization Error

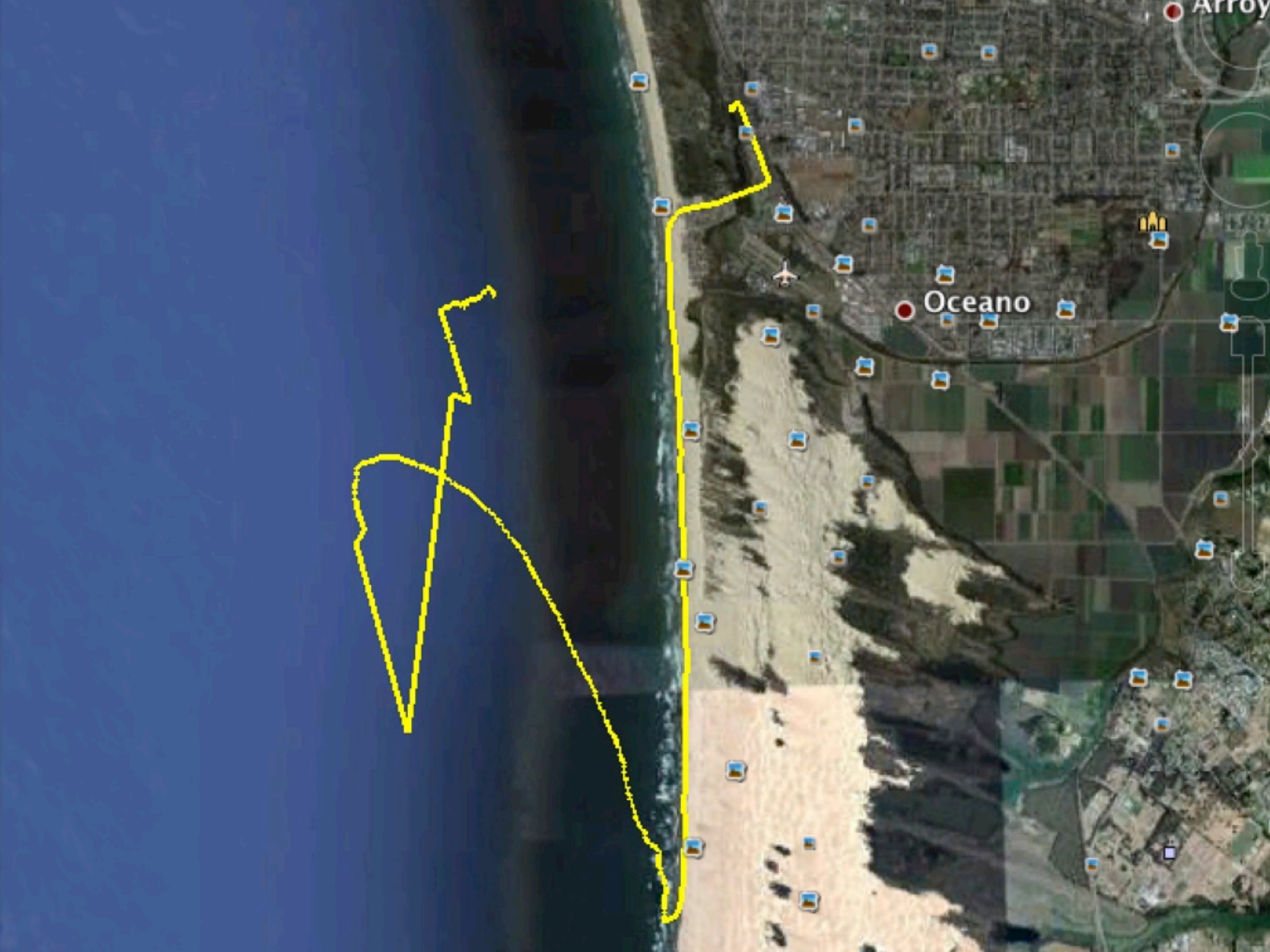
- We often witness finite precision in our sensors.
  - If the sample standard deviation S of our measurements is much larger than the quantization error (e.g. 10 times greater), we can ignore the quantization error.
  - For the GPS longitude measurements which have $S > 1$ meter, we can ignore the quantization error.

Summary

- We can calculate confidence intervals for parameters being measured
- We can construct linear models relating two parameters, along with their confidence intervals
- We can approximate how the error of one parameter affects a function of that parameter
- We can check that the quantization error is insignificant

To the Rescue




Oceano

Quantization Error

- DMM Example:
  - For a 12 bit DAQ, set to +/- 5V, the quantization range q is:

$$q = 10V * 1/2^{12} = 0.027V$$

  - The standard deviation of measurements within q is

$$S_q = q / \sqrt{12}$$

If $S > 10S_q$, ignore S_q

If $S_q > S > 10S_q$, let $S^2_{used} = S^2 + S^2_q$