AOE 3054. Statistical Analysis of Experimental Data

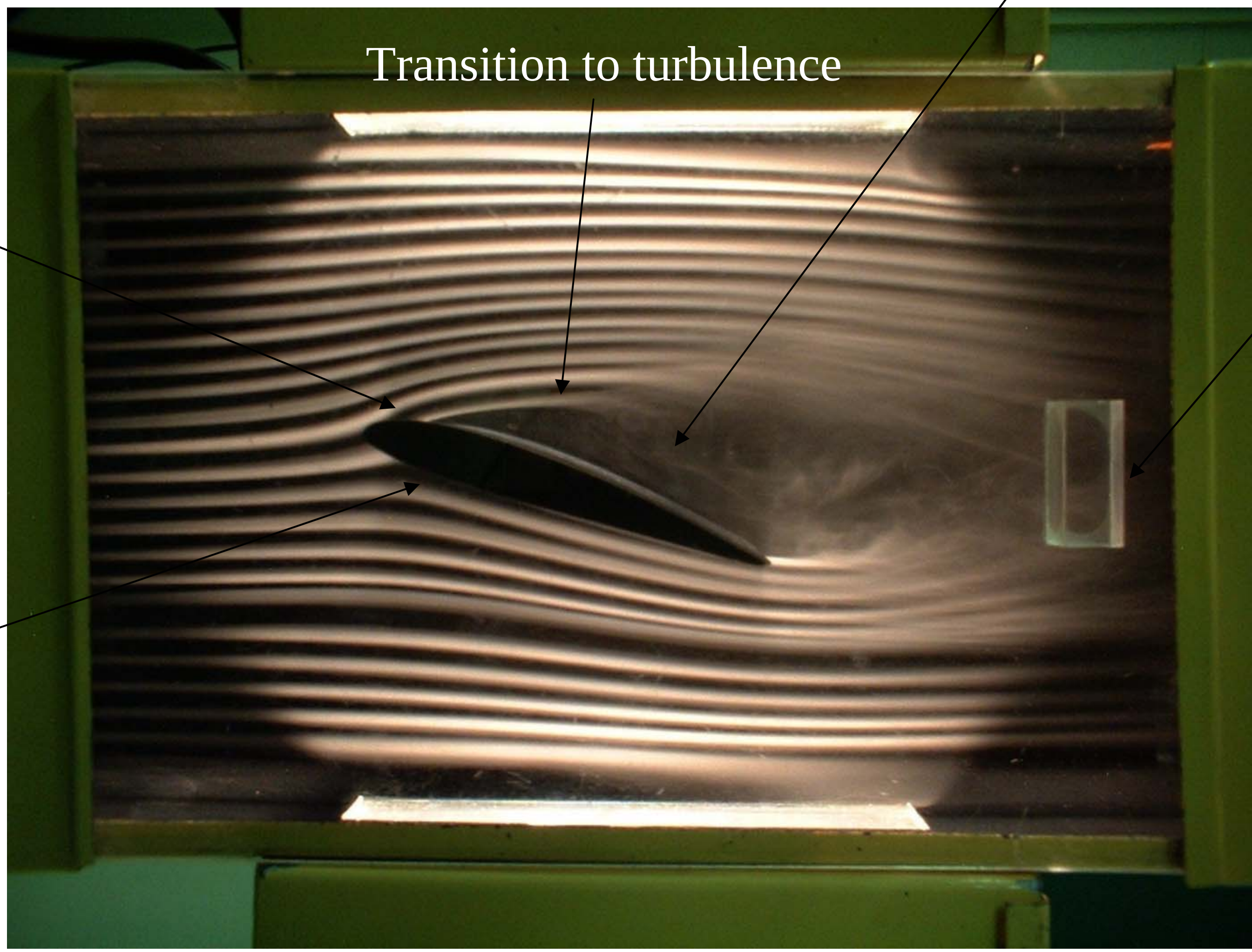
- Relevant to all experiments, IDLab and uncertainty analysis
- See lab manual, "Basic Concepts in Experiments"

1. Mean, Variance, Histograms
2. Probability density functions and the normal distribution
3. Correlation and Regression

Analysis?

Stall zone

Separation



Wake closes

Slowing of flow around stagnation point

Example

USS Jefferson City (SSN 759) navigates past Orote Point as the submarine enters Apra Harbor, Guam.

Wind Tunnel Model

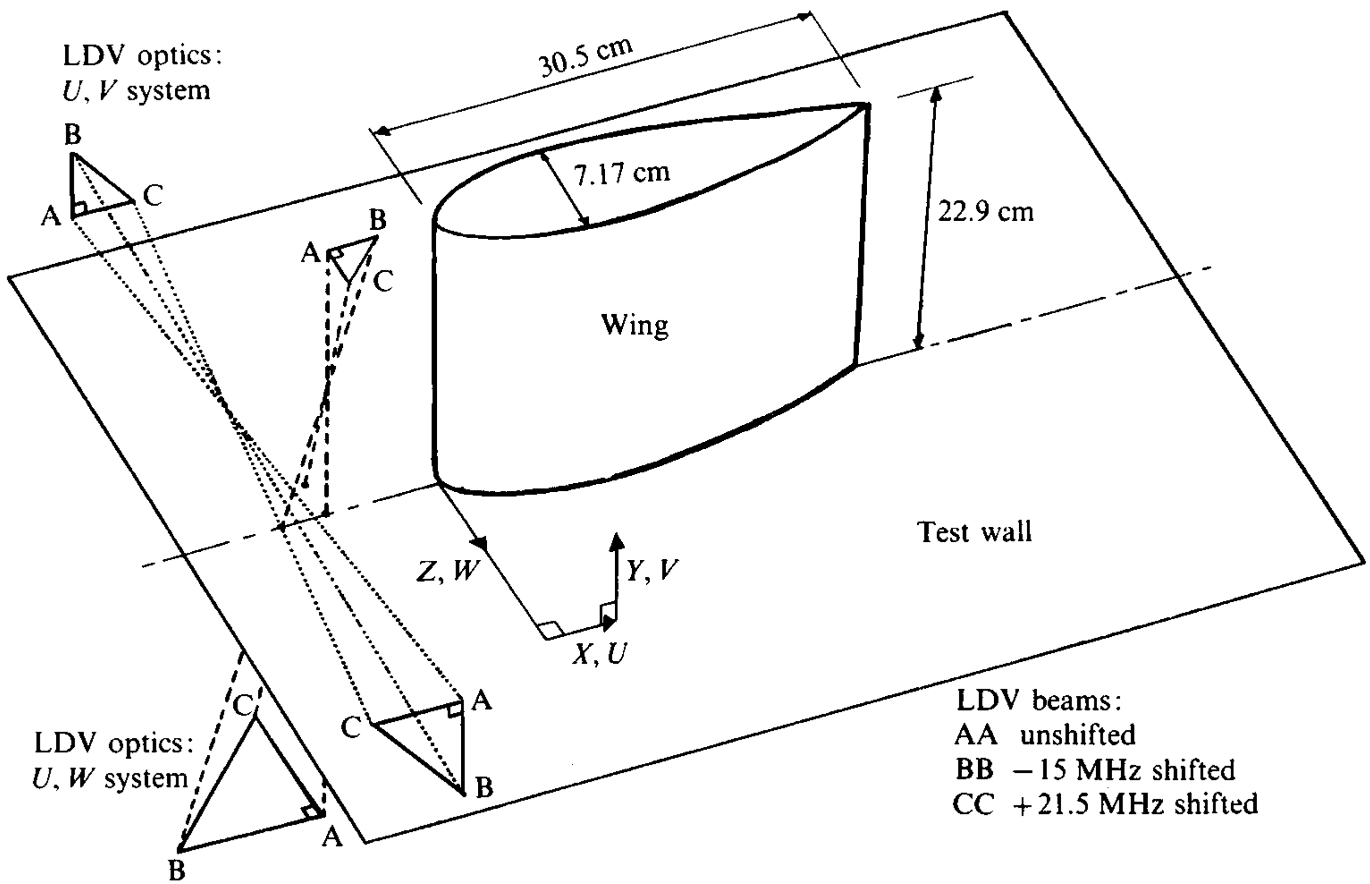
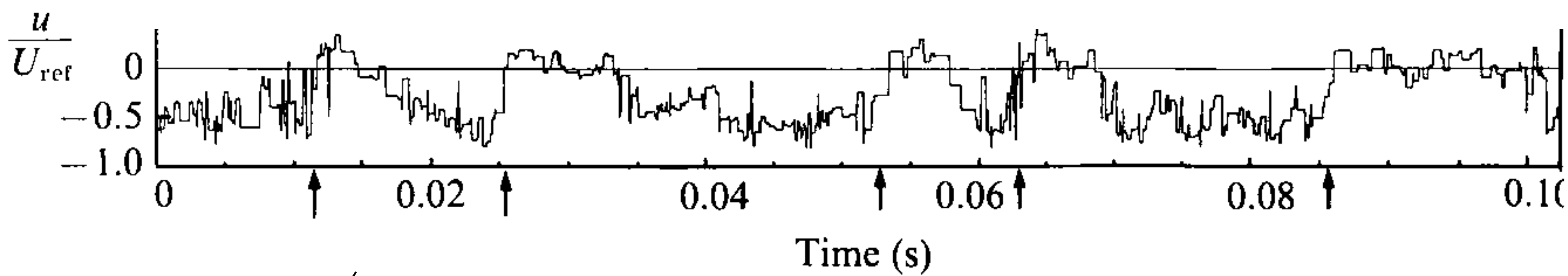
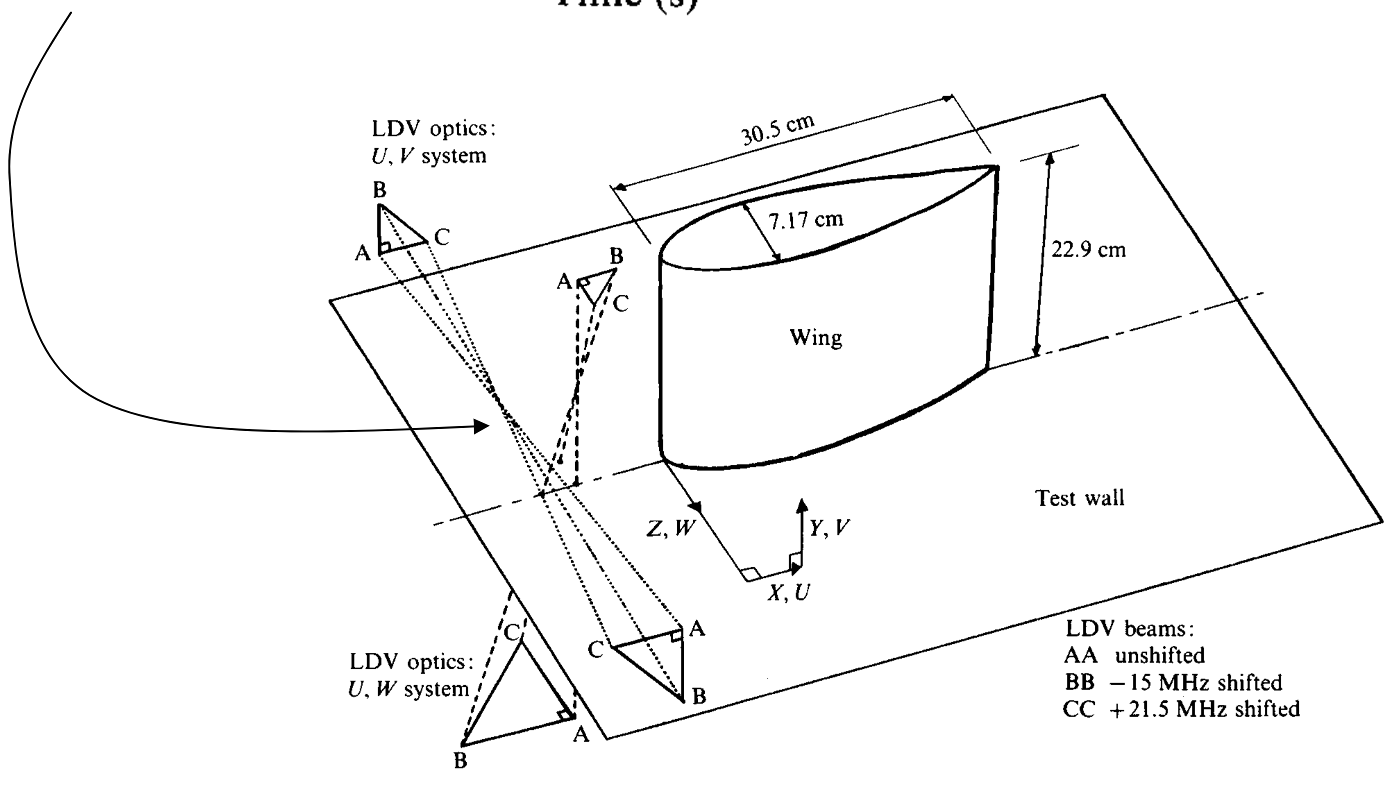


FIGURE 1. Perspective view of the wing-body junction showing the three-component LDV system.

Typical Velocity Signal





Mean, Variance, Histograms
Definitions

Consider a series of measurements x_1, x_2, x_3 … x_N,

Mean

Variance

Standard deviation

For a continuous function $x(t)$, where t could be anything ...

Mean

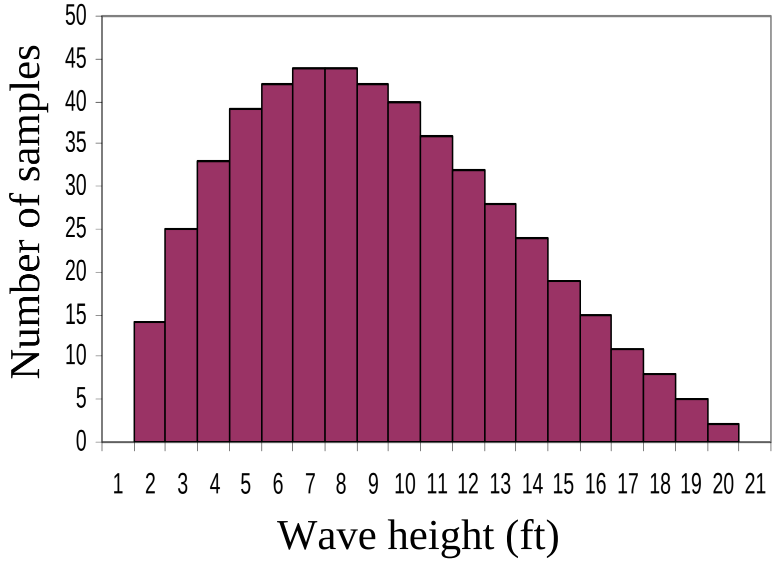
Variance

Standard deviation

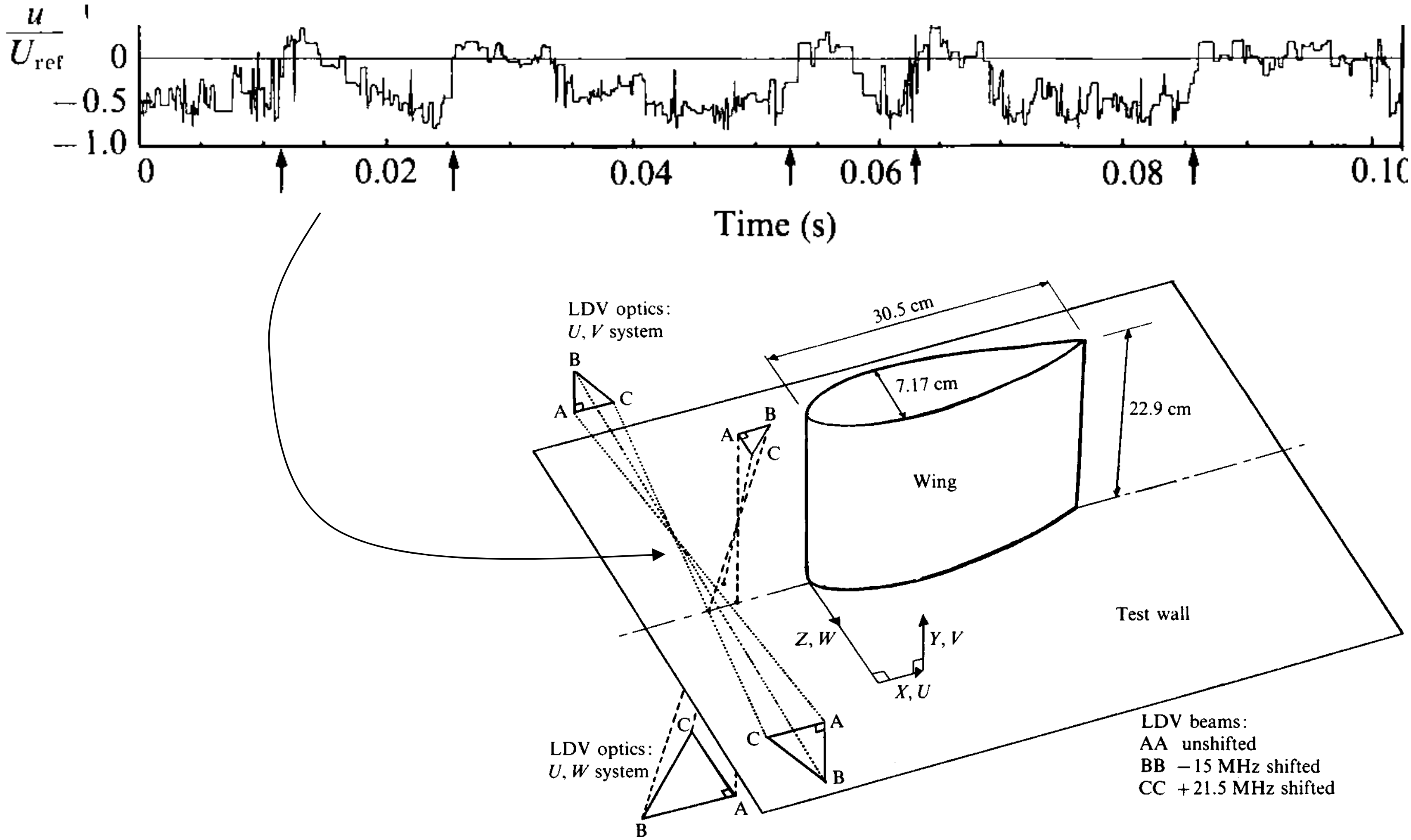
Mean, Variance, Histograms *Definitions*

Histogram: A chart showing the distribution of values in a series of samples

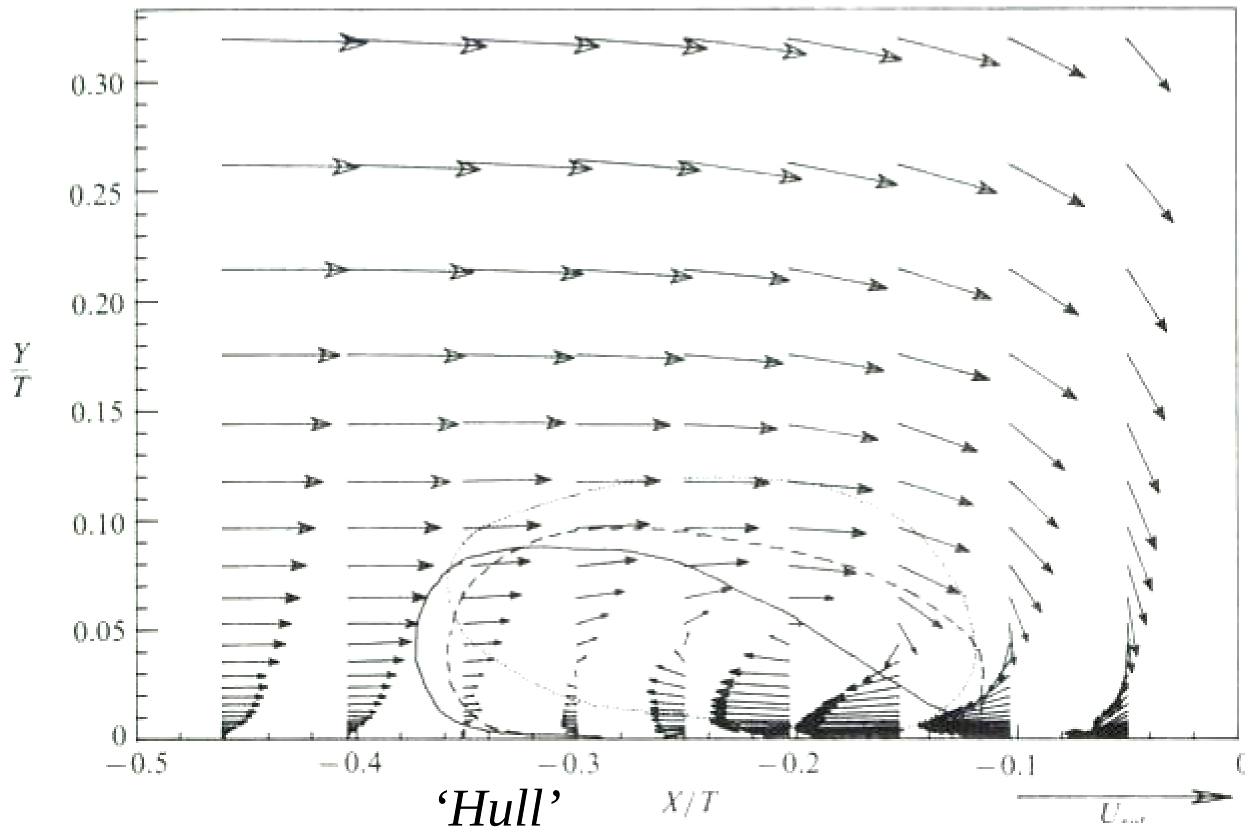
- *Divide up range of the measurements into a series of equal intervals (called bins)*
- *Add up the number of measurements falling in each bin*
- *The mean usually falls near the center of the histogram*
- *The standard deviation is typically 1/4 to 1/6th of the spread*
- *Histograms may be used to estimate probability*



Typical Velocity Signal



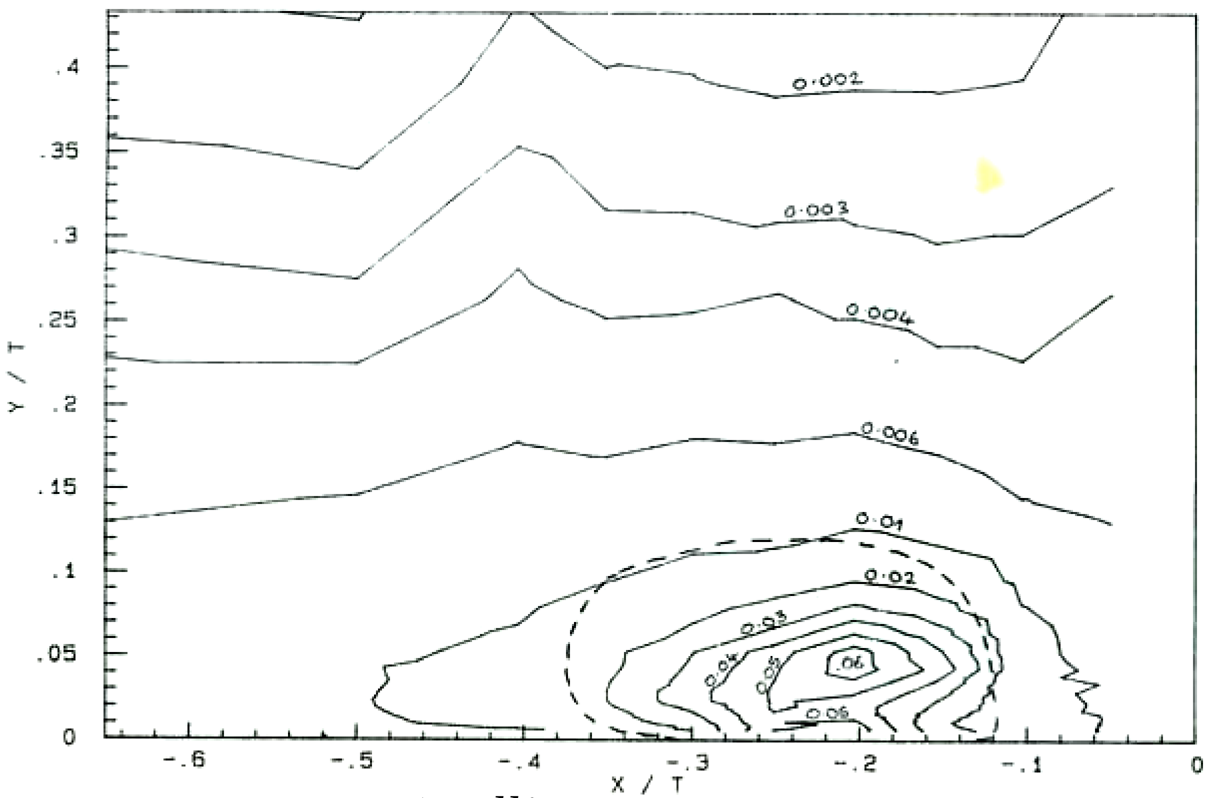
Mean Velocity on Center-plane



'Hull'

'Conning Tower Leading Edge'

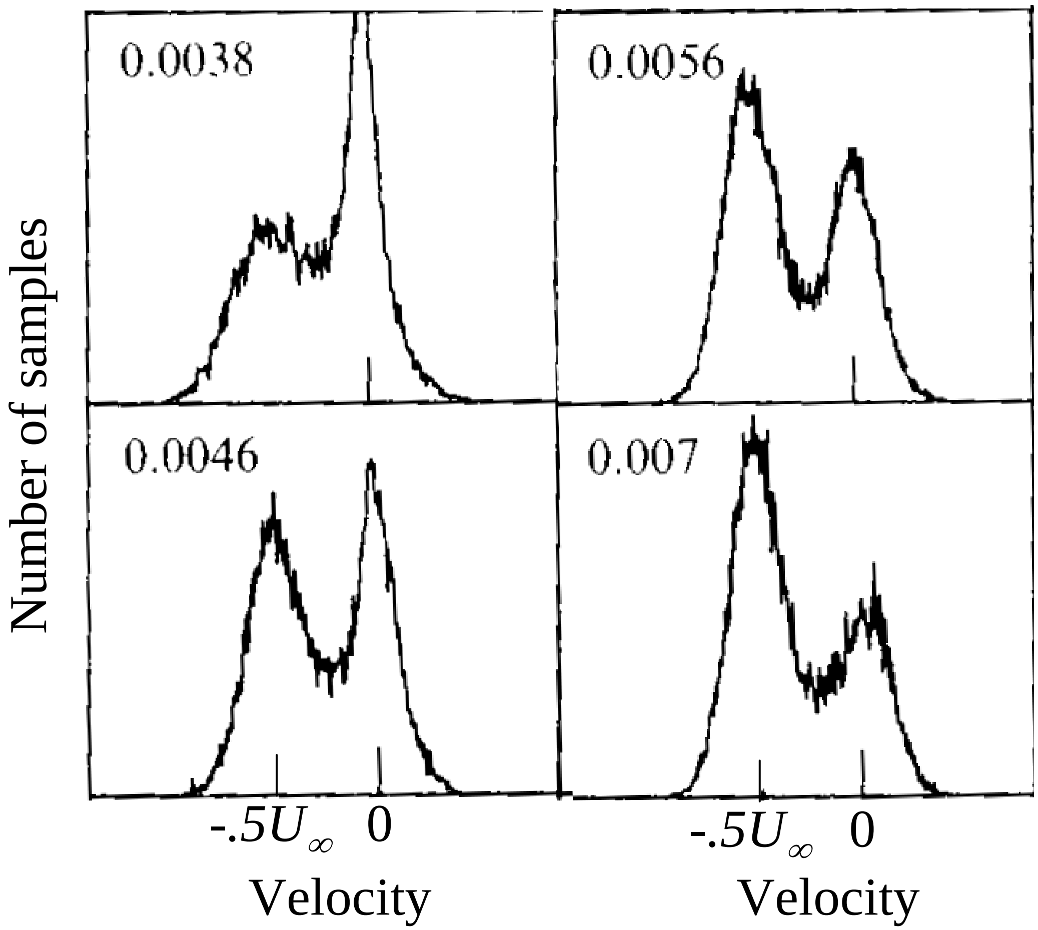
Velocity Variance



'Hull'

'Conning Tower Leading Edge'

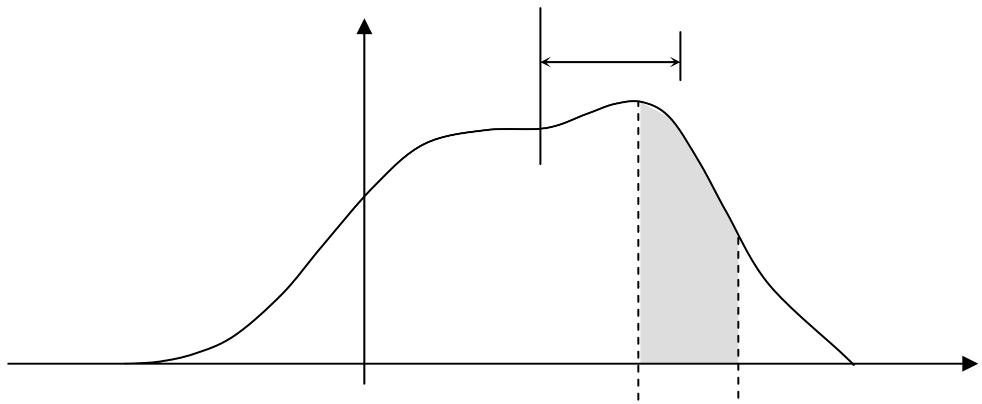
Velocity Histograms From in the Vortex



USS Seawolf (SSN 21), completing its initial sea trial

Probability Density Functions
Definition

(Add labels)

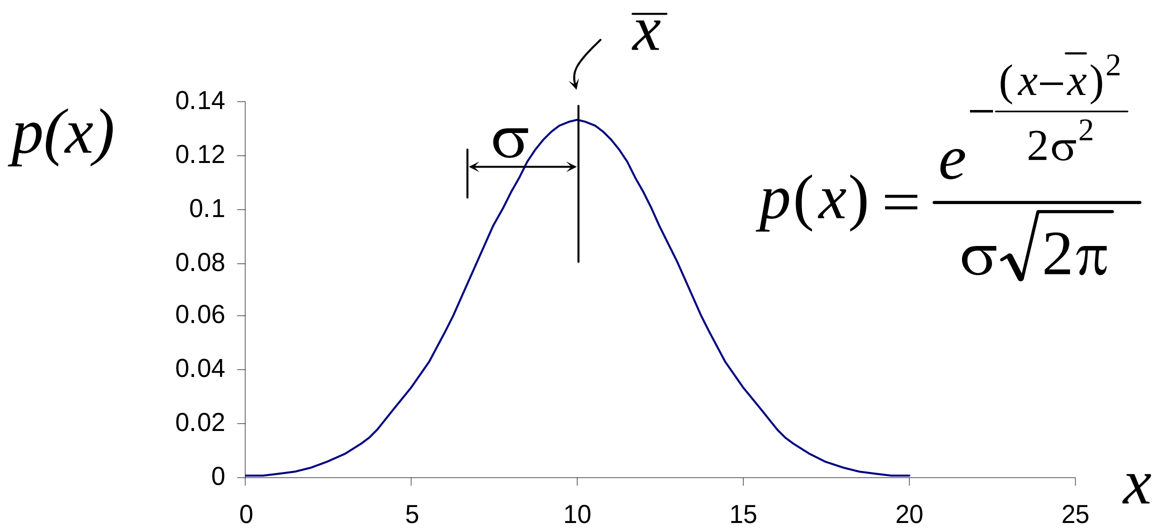


Definition: The area under $p(x)$ between two values x_0 and x_1 is the probability P that a given sample of x will fall between these values.

Mathematically:

Normal Distribution

Definition



$$p(x) = \frac{e^{-\frac{(x-\bar{x})^2}{2\sigma^2}}}{\sigma\sqrt{2\pi}}$$

The probability P that a given sample of x will fall between x_0 and x_1 is:

Which can be re-written as

Where

=

Matlab

Excel

Normal Distribution
Example of Use

A sensor is used to detect the flow rate of fuel to a jet engine. The flow rate fluctuates and the following are 21 successive readings readings (in arbitrary units),

(a) Determine the mean and standard deviation.

(b) Estimate the probability that a flow rate will have a value between .5 and .6.

(c) What percentage of the time is the flow rate likely to lie above a value of .8?

(d) What percentage of a large number of readings is likely to lie further than two standard deviations from the mean?

Reading	Flow rate
1	0.512
2	0.477
3	0.794
4	0.672
5	0.713
6	0.588
7	0.621
8	0.734
9	0.771
10	0.486
11	0.559
12	0.614
13	0.687
14	0.722
15	0.627
16	0.701
17	0.573
18	0.721
19	0.802
20	0.553
21	0.605

(a) Determine the mean and standard deviation

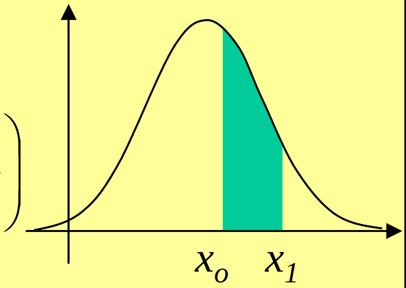
$$\overline{x} = \tfrac{1}{21}(.512 + .477 + .794 + \ldots) = 0.644$$

$$\sigma_x = \sqrt{\tfrac{1}{20}((.512 - .644)^2 + (.477 - .644)^2 + \ldots)} = 0.098$$

Normal Distribution

Example of Use

$$P(x_0 < x < x_1) = I\left(\frac{x_1 - \bar{x}}{\sigma}\right) - I\left(\frac{x_0 - \bar{x}}{\sigma}\right)$$



(b) Estimate the probability that the flow rate will have a value between .5 and .7.

Reading	Flow rate
1	0.512
2	0.477
3	0.794
4	0.672
5	0.713
6	0.588
7	0.621
8	0.734
9	0.771
10	0.486
11	0.559
12	0.614
13	0.687
14	0.722
15	0.627
16	0.701
17	0.573
18	0.721
19	0.802
20	0.553
21	0.605

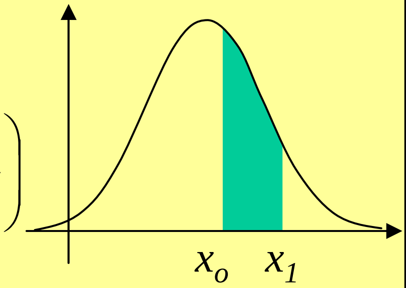
$\bar{x} = 0.644$

$\sigma = 0.098$

Normal Distribution

Example of Use

$$P(x_0 < x < x_1) = I\left(\frac{x_1 - \bar{x}}{\sigma}\right) - I\left(\frac{x_0 - \bar{x}}{\sigma}\right)$$



(c) What percentage of the time is the flow rate likely to lie above a value of .8?

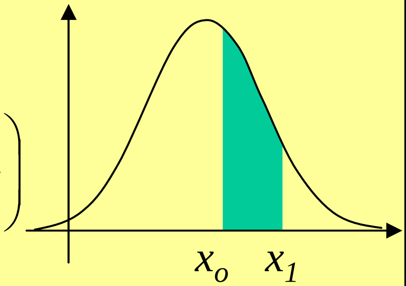
Reading	Flow rate
1	0.512
2	0.477
3	0.794
4	0.672
5	0.713
6	0.588
7	0.621
8	0.734
9	0.771
10	0.486
11	0.559
12	0.614
13	0.687
14	0.722
15	0.627
16	0.701
17	0.573
18	0.721
19	0.802
20	0.553
21	0.605

$\bar{x} = 0.644$

$\sigma = 0.098$

Normal Distribution
Example of Use

$$P(x_0 < x < x_1) = I\left(\frac{x_1 - \overline{x}}{\sigma}\right) - I\left(\frac{x_0 - \overline{x}}{\sigma}\right)$$



(d) What percentage of a large number of readings will lie further than two standard deviations from the mean?

Reading	Flow rate
1	0.512
2	0.477
3	0.794
4	0.672
5	0.713
6	0.588
7	0.621
8	0.734
9	0.771
10	0.486
11	0.559
12	0.614
13	0.687
14	0.722
15	0.627
16	0.701
17	0.573
18	0.721
19	0.802
20	0.553
21	0.605

$\overline{x} = 0.644$

$\sigma = 0.098$

So, on average, there is close to a 95% chance that the value will lie within 2 standard deviations of the mean. This result is important when we come to estimating experimental error.

Correlation (co-variance)
Definition

Consider two series of measurements of related quantities

$x_1, x_2, x_3 \ldots x_N,$ $y_1, y_2, y_3 \ldots y_N,$

Correlation

For continuous functions $x(t)$ and $y(t)$, where t could be anything ...

Correlation

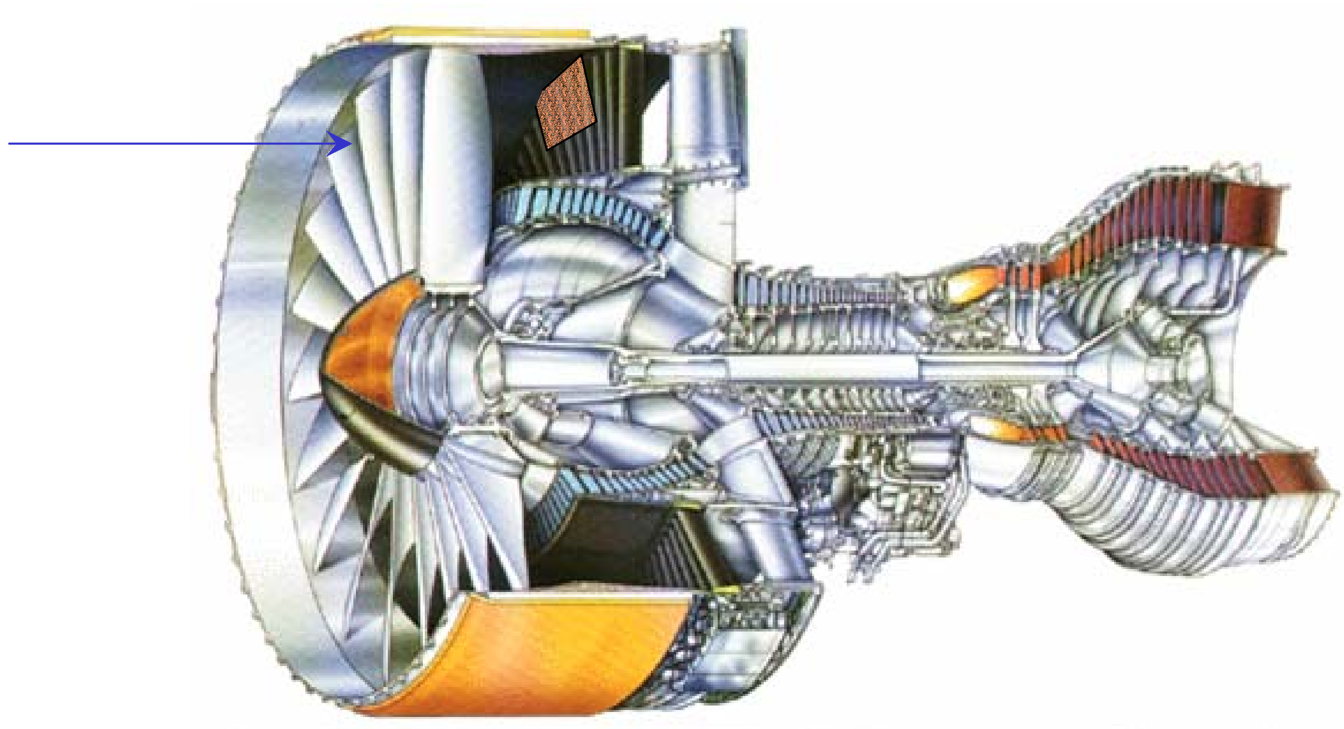
Correlation coefficient

Varies from -1 to 1

We also define the Correlation function

where τ is a shift in time, distance of whatever t represents, and the

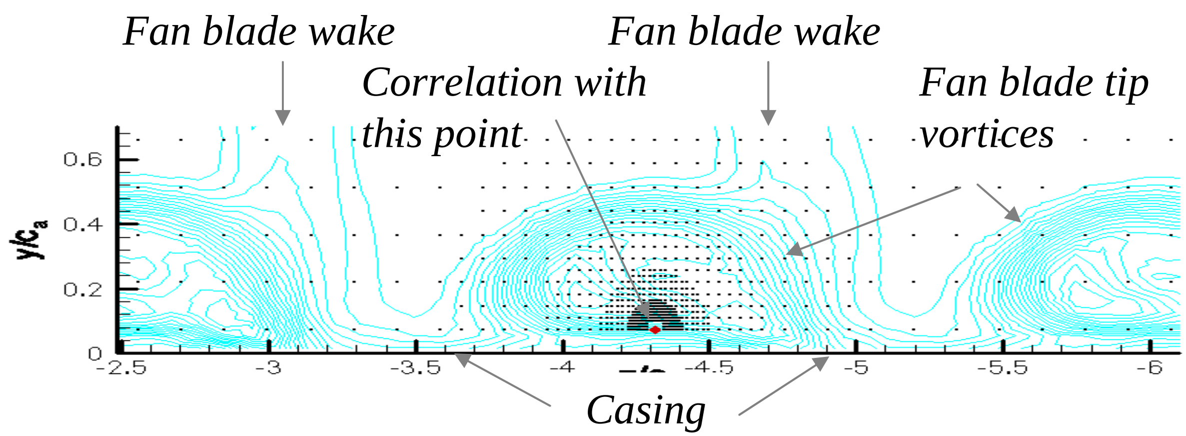
Correlation (co-variance)
Application Example

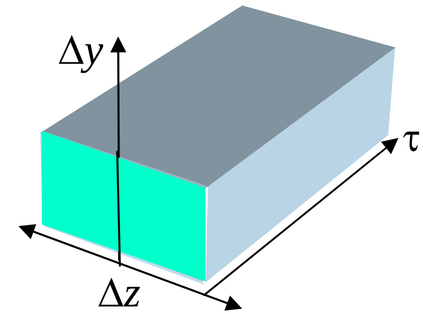


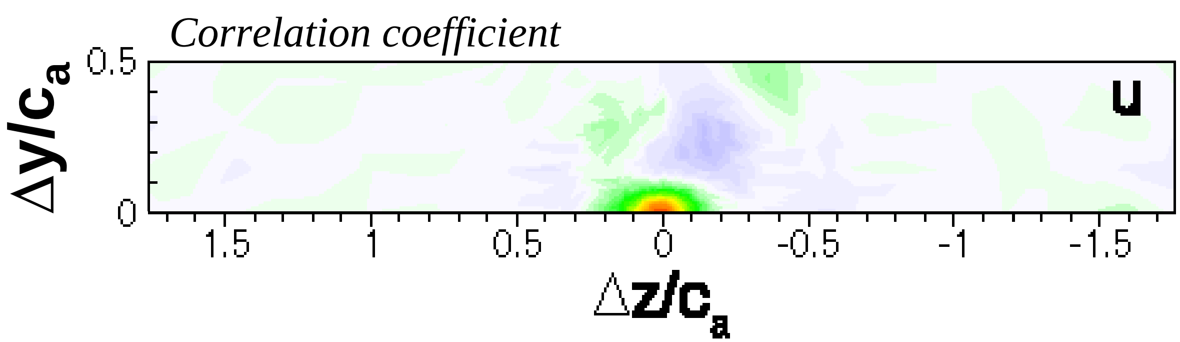
- GE 90 Aircraft Engine
- Eddies from fan tip entering stator row impinge in stators generating noise

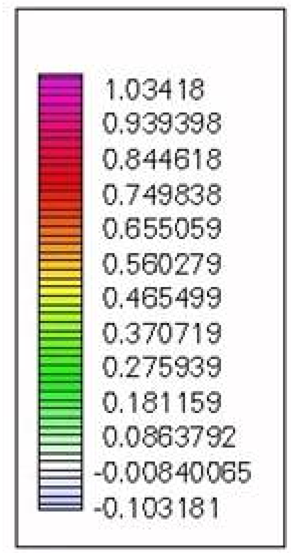
Space Time Correlations

Zero time-delay correlation function, U



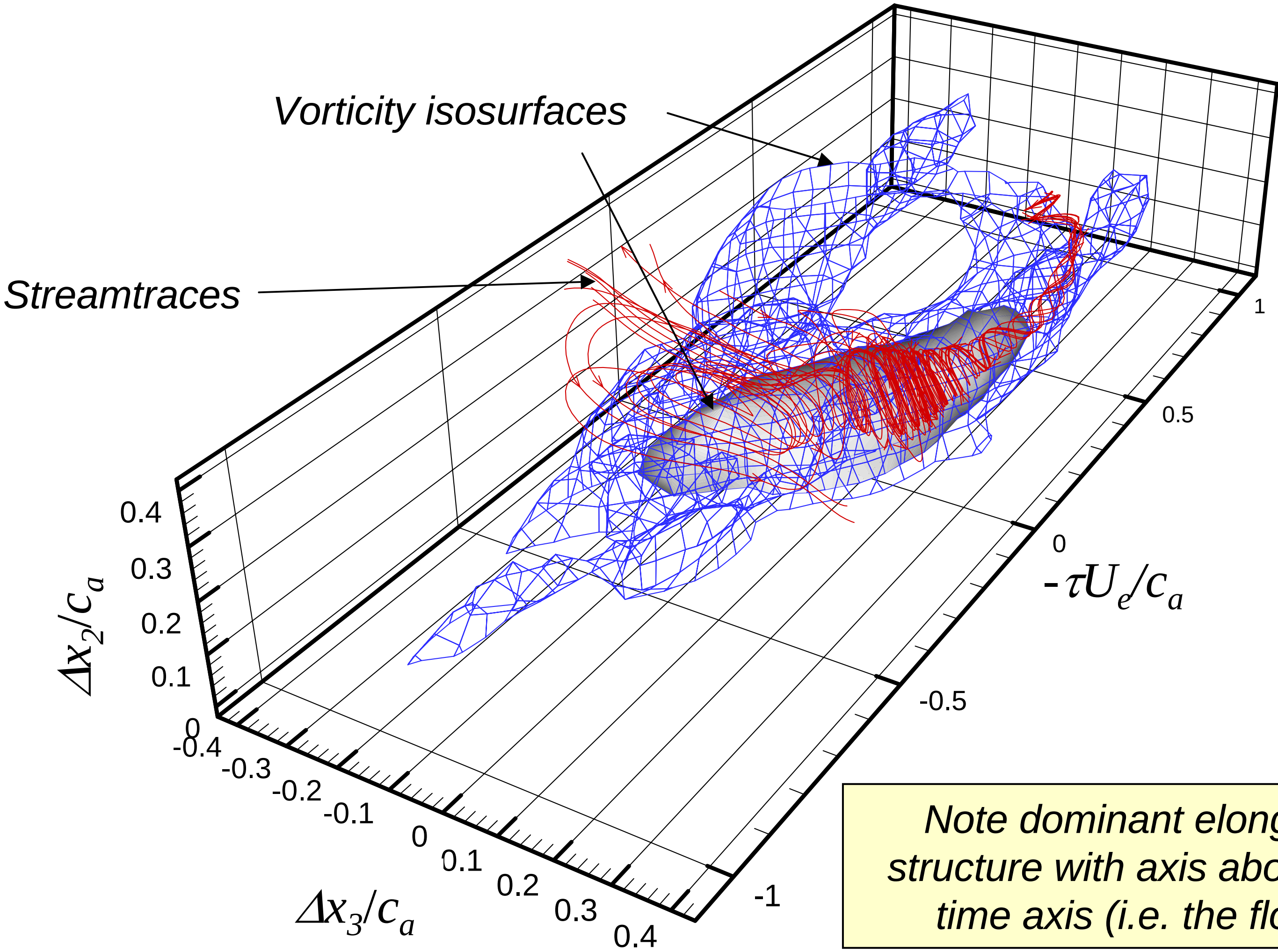






Reconstruction of 3D field from correlations

Linear stochastic estimate of flowfield around location 2



Note dominant elongated vortical structure with axis about 20º from the time axis (i.e. the flow direction)

Linear Regression

Consider a series of measured points (x_i,y_i) describing a relationship between two quantities x and y. We wish to find a straight line of the form $y = A+Bx$ that lies as closely as possible to the data. That is, we wish to choose A and B to minimize the error

This is done by setting

Which gives us two simultaneous equations for A and B, which may be solved to yield

Linear Regression

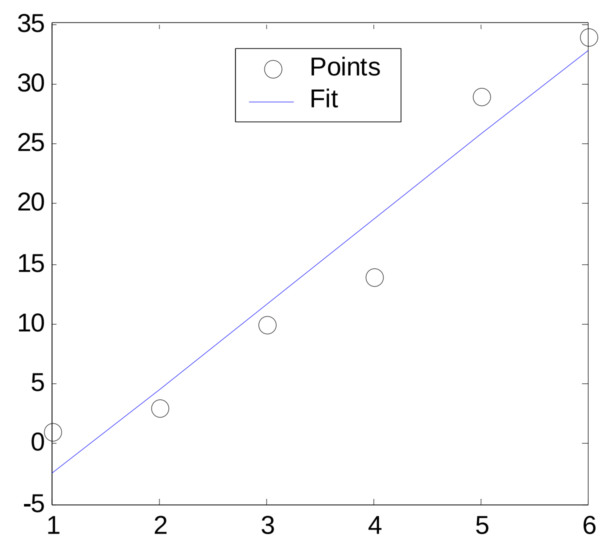
- Using Excel:
  - Under *Tools>Data Analysis…* (may have to add this feature by going to *Tools>Add-ins…* and checking *Analsys Toolpak*
  - Alternatively …

- Using Matlab (preferred)

Matlab Example

$$\begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ \vdots \end{pmatrix} = \begin{pmatrix} 1 & x_1 \\ 1 & x_2 \\ 1 & x_3 \\ \vdots & \vdots \end{pmatrix} \begin{pmatrix} A \\ B \end{pmatrix}$$

```
x=[1; 2;  3;  4;  5;  6];
y=[1; 3; 10; 14; 29; 34];
c=[ones(size(x)) x];
a=c\y;
plot(x,y,'ko',x,a(1)+a(2)*x,'b-');
```



Matlab Example – *Generalized Regression*

$$\begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ \vdots \end{pmatrix} = \begin{pmatrix} 1 & x_1 & x_1^2 \\ 1 & x_2 & x_2^2 \\ 1 & x_3 & x_3^2 \\ \vdots & \vdots & \vdots \end{pmatrix} \begin{pmatrix} A \\ B \\ C \end{pmatrix}$$

```
x=[1; 2;  3;  4;  5;  6];
y=[1; 3; 10; 14; 29; 34];
c=[ones(size(x)) x x.^2];
a=c\y;
plot(x,y,'ko',x,a(1)+a(2)*x+a(3)*x.^2,'b-');
```