Environmental Fluid Dynamics: Lecture 16

Dr. Jeremy A. Gibbs

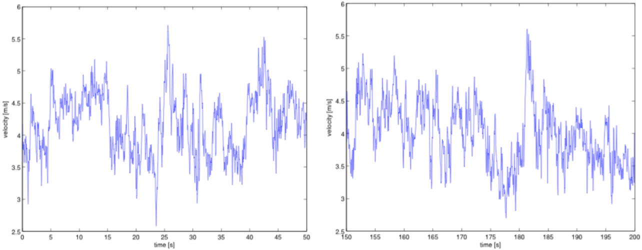
Department of Mechanical Engineering
University of Utah

Spring 2017

Overview

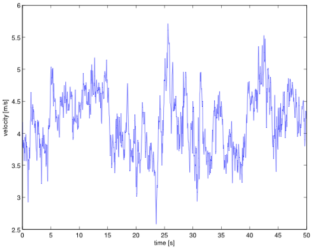
Statistical Tools for Turbulence

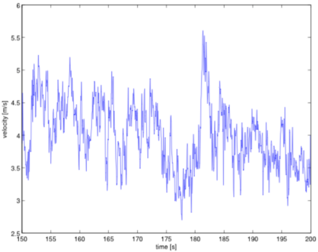
- Consider the velocity field U.



- Since U is a random variable, its value is unpredictable for a turbulent flow.
- Thus, any theory used to predict a particular value for U will likely fail.
- As we saw, however, certain statistical measures (histogram) appear to be reproducible.

- Consider the velocity field U.





- In this time trace, notice that velocity is bounded, so there is some defined measure of turbulence intensity.
- The ability to pick a mean value means that the flow field is not entirely random.
- The time trace shows the existence of multiple time scales, which suggests that there are spatial features of different sizes and durations (i.e. a spectrum)

Turbulence Spectrum

- Example spectrum of wind speed near the ground

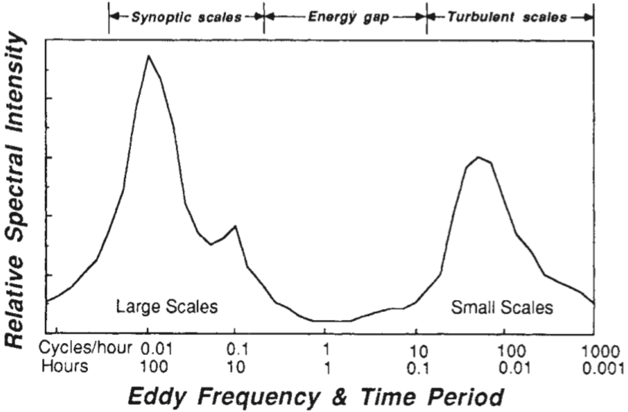


Fig. 2.2 Schematic spectrum of wind speed near the ground estimated from a study of Van der Hoven (1957).

From Stull (1988)

Turbulence Spectrum

- In the example spectrum, the vertical axis gives the contribution to the total turbulence energy by a particular eddy size, while the horizontal axis gives the size of eddies contained in the spectrum in terms of their duration.
- Peaks show scales with the largest contribution.
- The left peak, associated with a period of 100 hours, is associated with wind speed variations caused by frontal passages and weather systems.
- The next peak is at 24 hours, which highlights the diurnal variations in wind speed.
- The right peak, located between 10 seconds and 10 minutes, is evidence of small-scale eddies associated with turbulence.

Spectral Gap

- The relative lack of variation between the synoptic and microscale peaks is called the spectral gap.
- Motions to the left are considered the *mean flow*, and motions to the right are considered the *turbulent flow*.
- The spectral gap allows us to separate the flow into turbulent and non-turbulent parts.

Turbulence Spectrum

Mean and fluctuating parts

- To isolate large scales from small scales, we can average over a period of 30 minutes to 1 hour
- This allows us to average out the positive and negative fluctuations about the mean
- The mean ($\overline{U}$; more later) is subtracted from the instantaneous value (U) to obtain the turbulent part

$$u' = U - \overline{U}$$

Turbulence Spectrum

Mean and fluctuating parts

- You can think of the turbulent fluctuation u' as a gust of wind associated with variations lasting longer than an hour, while the mean wind $\overline{U}$ is the part of the wind with variations that last longer than an hour (or so).

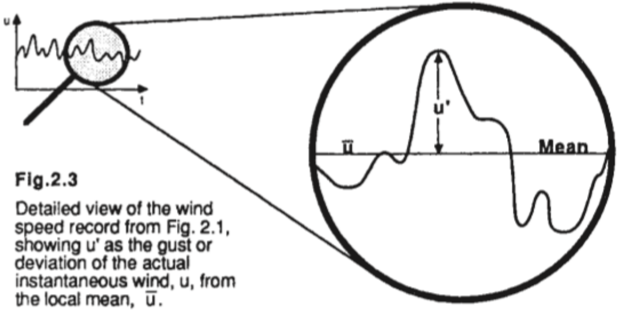


Fig.2.3

Detailed view of the wind speed record from Fig. 2.1, showing u' as the gust or deviation of the actual instantaneous wind, u, from the local mean, $\overline{u}$.

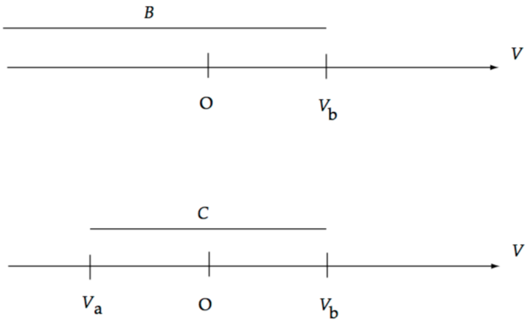
From Stull (1988)

- This underscores the need to describe turbulence statistically.

Sample Space

- In reality, a velocity field $U(\vec{x}, t)$ is more complicated than a single random variable.
- We need a wide range of statistical tools to characterize random variables.
- In order to consider more general events, we need to think in terms of *sample space*.
- Consider an independent velocity variable V, which is the sample-space variable for U.

Sample Space



$$B \equiv \{U < V_b\}$$
$$C \equiv \{V_a \leq U < V_b\}$$

B and C are events (or values) that correspond to different regions of the sample space (*i.e.*, velocity field).

Probability

- Using the previous example, the probability of event B is given as:

$$p = P(B) = P\{U < V_b\}$$

- This is the likelihood of B occurring $(U < V_b)$.
- p is a real number, $0 \le p \le 1$.
- $p = 0$ is an impossible event.
- $p = 1$ is a certain event.

Cumulative Distribution Function

- The probability of any event is determined by the cumulative distribution function (CDF)

$$F(V) \equiv P\{U < V\}$$

- For event B:

$$P(B) = \{U < V_b\} = F(V_b)$$

- For event C:

$$\begin{aligned} P(C) &= \{V_a \leq U < V_b\} = P\{U < V_b\} - P\{U < V_a\} \\ &= F(V_b) - F(V_a) \end{aligned}$$

Cumulative Distribution Function

Three basic properties of CDF:

- $F(-\infty) = 0$, since $\{U < -\infty\}$ is impossible.
- $F(\infty) = 1$, since $\{U > \infty\}$ is impossible.
- $F(V_b) \geq F(V_a)$, for $V_b > V_a$, since $p > 0$. Thus, F is a non-decreasing function.

Probability Density Function

The probability density function (PDF) is the derivative of the CDF

$$f(V) \equiv \frac{dF(v)}{dV}$$

Based on the properties of the CDF, it follows that:

- $f(V) \geq 0$
- $\int_{-\infty}^{\infty} f(V) dV = 1$
- $f(-\infty) = f(\infty) = 0$

Probability Density Function

The probability that a random variable is contained within a specific interval is the integral of the PDF over that interval

$$\begin{aligned} P(C) = P\{V_a \leq U < V_b\} &= F(V_b) - F(V_a) \\ &= \int_{V_a}^{V_b} f(V)dV \end{aligned}$$

Or, for a very small interval dV_s:

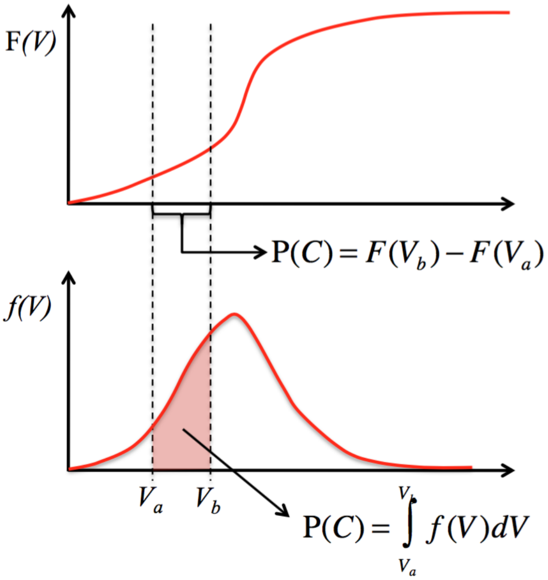
$$\begin{aligned} P(C_s) = P\{V_a \leq U < V_a + dV_s\} &= F(V_a + dV_s) - F(V_a) \\ &= f(V_a)dV_s \end{aligned}$$

Probability Density Function

More details about the PDF $f(V)$:

- $f(V)$ is the probability per unit distance in the sample space – hence, the term *density*.
- $f(V)$ has dimensions of U^{-1}, while the CDF is dimensionless.
- The PDF fully characterizes the statistics of a signal (random variable).
- If two or more signals have the same PDF, then they are considered to be statistically identical.

CDF (top) vs. PDF (bottom)



The Mean

We can also define a signal by its individual statistics, which collectively describe the PDF.

The mean (or expected) value of a random variable U is given by:

$$\overline{U} \equiv \int_{-\infty}^{\infty} V f(V) dV$$

or in discrete form:

$$\overline{U} \equiv \frac{1}{N} \sum_{i=1}^{N} V_i$$

The mean represents the probability-weighted sum of all possible values of U.

The Mean: Temporal Averaging

- For a continuous function at point in space $s = (x, y, z)$ over a time period P, where the turbulence is assumed to be *stationary* (the statistics are **NOT** changing over the averaging period), the time average is given by:

$$\overline{u}^t(s) = \frac{1}{P}\int_{t=t_0}^{t=t_0+P} u(t, s)dt$$

- For discrete data uniformly spaced in time, where $P = N\Delta t$:

$$\overline{u}^t(s) = \frac{1}{N}\sum_{i=1}^{N} u(t, s)$$

The Mean: Spatial Averaging

- Applies at in instant in time and is given as an integral of the spatial domain S:

$$\overline{u}^s(t) = \frac{1}{S}\int_S u(t,s)dS$$

- Over some volume $S = \Delta x \Delta y \Delta z$:

$$\overline{u}^s(t) = \frac{1}{\Delta x \Delta y \Delta z}\int\int_S\int u(t,x,y,z)dxdydz$$

- For line averaging uniformly spaced data in space, where $Y = N\Delta y$:

$$\overline{u}^s(t) = \frac{1}{Y}\int_{y=0}^{y=Y} u(t,s)dy$$

  or discretely

$$\overline{u}^s(t) = \frac{1}{Y}\sum_{i=1}^{N} u(t,s)$$

The Mean: Ensemble Averaging

- Averaging over a number (N) of experiments at a point in space. This method of averaging tends to minimize random experimental errors by repeating an experiment.

$$\overline{u}^e(t,s) = \frac{1}{N}\sum_{i=1}^{N} u_i(t,s)$$

The Mean: Ergodicity

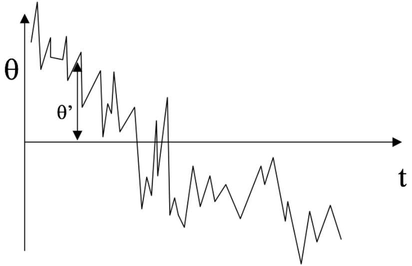
- If the turbulence is both *homogeneous* and *stationary*, the time, space, and ensemble averages should be the same, namely

$$\overline{u}^t = \overline{u}^s = \overline{u}^e = \overline{u}$$

- homogeneous turbulence $\Rightarrow$ statistics are invariant of coordinate translation
- stationary turbulence, $\Rightarrow$ the statistics are invariant to the choice of time window
- In the atmosphere there are many occasions when turbulence is neither homogeneous (e.g., around trees or buildings) or stationary (e.g., evening decay of the CBL)
- We must be very careful in applying the various types of averaging.

The Mean: Ergodicity

- Consider this time series of potential temperature during the evening transition. These data clearly show the diurnal decrease of temperature in evening.
- If we choose an averaging time of 1 hr, we will be including this diurnal variation in our fluctuations.
- This problem may be avoided by using smaller averaging times or using linear detrending techniques



Reynolds Averaging

- We saw that the spectral gap allows the separation of a flow field into mean and perturbation parts.
- The formal procedure of applying this separation is called **Reynolds decomposition** or **Reynolds averaging**.
- Accordingly, it is important to discuss the properties of the mean in this context.
- We will use this procedure and rules when deriving the turbulence equations (next lecture).

Reynolds Averaging

Rules of averaging:

- $\overline{c} = c$, where c is a constant
- $\overline{(c\,A)} = c\,\overline{A}$
- $\overline{(\overline{A})} = \overline{A}$
- $\overline{(\overline{A}\,B)} = \overline{A}\,\overline{B}$
- $\overline{(A + B)} = \overline{A} + \overline{B}$
- $\overline{\left(\frac{dA}{dt}\right)} = \frac{d\overline{A}}{dt}$

Reynolds Averaging

- Let's apply these rules to variables that have been split into their mean and perturbation parts.
- Let $A = \overline{A} + a'$ and $B = \overline{B} + b'$
- Start with A

$$\overline{(A)} = \overline{(\overline{A} + a')} = \overline{(\overline{A})} + \overline{a'} = \overline{A} + \overline{a'}$$

The only way this is true is if $\overline{a'} = 0$, which makes sense if we consider the definition of the mean (i.e., the sum of positive perturbations from the mean equals the sum of the negative perturbations).

Reynolds Averaging

- Consider another example:

$$\overline{(\overline{B}\,a')} = \overline{B}\;\overline{a'} = \overline{B} \cdot 0 = 0$$

- Similarly, $\overline{(\overline{A}\,b')} = 0$
- Lastly:

$$\begin{aligned}
\overline{A\,B} &= \overline{(A+a')(B+b')} = \overline{\overline{A}\;\overline{B} + \overline{A}\,b' + a'\,\overline{B} + a'\,b'} \\
&= \overline{\overline{A}\;\overline{B}} + \underbrace{\overline{\overline{A}\,b'}}_{=0} + \underbrace{\overline{\overline{B}\,a'}}_{=0} + \overline{a'\,b'} \\
&= \overline{A}\;\overline{B} + \overline{a'\,b'}
\end{aligned}$$

Where $\overline{a'\,b'}$ is covariance (more later), or variance if $a' = b'$.
Note: although $\overline{a'} = 0$ and $\overline{b'} = 0$, $\overline{a'\,b'}$ is **not** necessarily 0.

Variance

Recall that a *fluctuation* (or perturbation) from the mean:

$$u' \equiv U - \overline{U}$$

The *variance* is just the mean-square fluctuation:

$$\begin{aligned}\sigma_u^2 = \mathsf{var}(U) &= \overline{u'^2} \\ &= \int_{-\infty}^{\infty} (V - \overline{U})^2 f(V) dV\end{aligned}$$

Or, in discrete form:

$$\sigma_u^2 = \frac{1}{N - 1^*} \sum_{i=1}^{N} (V_i - \overline{U})^2$$

Variance essentially measures how far a set of (random) numbers are spread out from their mean.

*note the $(N-1)$. This is the Bessel correction – used to correct for bias.

Standard Deviation

The *standard deviation*, or root-mean square (rms) deviation, is just the square-root of the variance:

$$\sigma_u \equiv \mathsf{sdev}(U) = \sqrt{\sigma_u^2} = \sqrt{\overline{u'^2}}$$

The standard deviation basically measures the amount of variation of a set of numbers.

Other moments

The n^{th} central moment is defined as:

$$\mu_n \equiv \overline{u'^n} = \int_{-\infty}^{\infty} (V - \overline{U})^n f(V) dV$$

Standardized Moments

It is often advantageous to express variables as standardized random variables. These standardized variables have zero mean and unit variance.

The standardized version of U (centered and scaled) is given by:

$$\hat{U} \equiv \frac{U - \overline{U}}{\sigma_u}$$

Accordingly, the n^{th} standardized moments are expressed as:

$$\hat{\mu}_n \equiv \frac{\overline{u'^n}}{\sigma_u^n} = \frac{\mu_n}{\sigma_u^n} = \int_{-\infty}^{\infty} \hat{V}^n \hat{f}(\hat{V}) d\hat{V}$$

Other moments

Different moments each describe an aspect of the shape of the PDF:

- $\mu_1 =$ mean (expected value)
- $\mu_2 =$ variance (spread from the mean)
- $\hat{\mu}_3 =$ skewness (asymmetry of PDF)
- $\hat{\mu}_4 =$ kurtosis (sharpness of the PDF peak)

Example PDFs

Read Pope (Chapter 3.3) for descriptions of different PDFs.

Examples include:

- uniform
- exponential
- Gaussian
- log-normal
- gamma
- Delta-function
- Cauchy

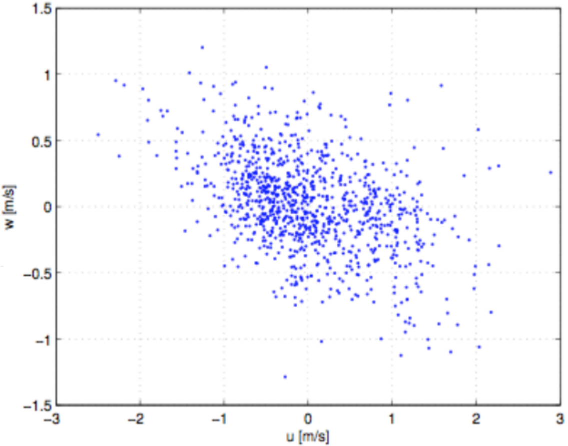
Joint Random Variables

So far, our statistical description has been limited to single random variables. However, turbulence is governed by the Navier-Stokes equations, which are a set of 3 coupled PDEs.

We expect this will result in some correlation between different velocity components.

Joint Random Variables

Example: turbulence data from the ABL: scatter plot of horizontal (u) and vertical (w) velocity fluctuations.



The plot appears to have a pattern (*i.e.*, negative slope).

Joint Random Variables - Sample Space

We will now extend the previous results from a single velocity component to two or more.

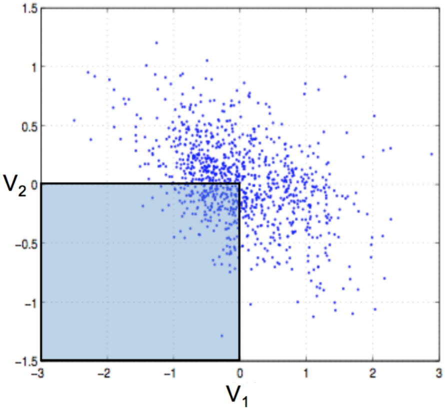
The sample-space variables corresponding to the random variables $U = \{U_1, U_2, U_3\}$ are given by $V = \{V_1, V_2, V_3\}$.

Joint Random Variables - CDF

The joint CDF (jCDF) of the random variables (U_1, U_2) is given by:

$$F_{12}(V_1, V_2) \equiv P\{U_1 < V_1, U_2 < V_2\}$$

This is the probability of the point (V_1, V_2) lying inside the shaded region



Joint Random Variables - CDF

The jCDF has the following properties:

- $F_{12}(V_1 + \delta V_1, V_2 + \delta V_2) \geq F_{12}(V_1, V_2)$ (non-decreasing) for all $\delta V_1 \geq 0$, $\delta V_2 \geq 0$

- $F_{12}(-\infty, V_2) = P\{U_1 < -\infty, U_2 < V_2\} = 0$ since $\{U_1 < -\infty\}$ is impossible.

- $F_{12}(\infty, V_2) = P\{U_1 < \infty, U_2 < V_2\} = P\{U_2 < V_2\}$
  $F_{12}(\infty, V_2) = F_2(V_2)$, since $\{U_1 < \infty\}$ is certain.

In the last example, $F_2(V_2)$ is called the *marginal CDF*.

Joint Random Variables - PDF

The joint PDF (jPDF) is defined as:

$$f_{12}(V_1, V_2) \equiv \frac{\partial^2}{\partial V_1 \partial V_2} F_{12}(V_1, V_2)$$

If we integrate over V_1 and V_2, we get the probability:

$$P\{V_{1a} \leq U_1 < V_{1b}, V_{2a} \leq U_2 < V_{2b}\} = \int_{V_1 a}^{V_1 b} \int_{V_2 a}^{V_2 b} f_{12}(V_1, V_2) dV_2 dV_1$$

Joint Random Variables - PDF

Based on the jCDF, the jPDF has the following properties:

- $f_{12}(V_1, V_2) \geq 0$
- $\int_{-\infty}^{\infty} f_{12}(V_1, V_2) dV_1 = f_2(V_2)$
- $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{12}(V_1, V_2) dV_1 dV_2 = 1$

In the middle example, $f_2(V_2)$ is called the *marginal PDF*. Practically speaking, we find the PDF of a time (or space) series by:

- Create a histogram of the series(group values into bins)
- Normalize the bin weights by the total $\#$ of points

Joint Random Variables - Means

Similar to the single variable form, if we have $Q(U_1, U_2)$:

$$\overline{Q(U_1, U_2)} \equiv \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} Q(V_1, V_2) f_{12} dV_2 dV_1$$

We can use this equation to define a few important statistics.

Joint Random Variables - Covariance

We can define *covariance* as:

$$\begin{aligned}\mathrm{cov}(U_1, U_2) &\equiv \overline{u_1' u_2'} \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(V_1 - \overline{U_1})(V_2 - \overline{U_2}) f_{12}(V_1, V_2) dV_2 dV_1\end{aligned}$$

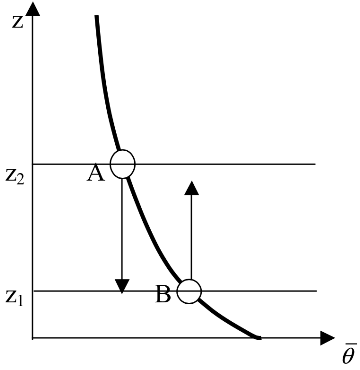
Or for discrete data

$$\begin{aligned}\mathrm{cov}(U_1, U_2) &\equiv \overline{u_1' u_2'} \\ &= \frac{1}{N-1}\sum_{j-1}^{N}(V_{1j} - \overline{U_1})(V_{2j} - \overline{U_2})\end{aligned}$$

Covariance is basically a measure of how much two random variables change together.

Joint Random Variables - Covariance

Example: cov(w,θ) $= \overline{w'\theta'}$. This is the eddy flux concept.



Under convective conditions

- Parcel A: $w' < 0$ and $\theta' < 0 \rightarrow \overline{w'\theta'} > 0$ (positive heat flux)
- Parcel B: $w' > 0$ and $\theta' > 0 \rightarrow \overline{w'\theta'} > 0$ (positive heat flux)

Joint Random Variables - Correlation Coefficient

The *correlation coefficient* is given by:

$$\rho_{12} \equiv \frac{\overline{u_1' u_2'}}{\sqrt{\overline{{u_1'}^2}\,\overline{{u_2'}^2}}}$$

Correlation coefficient has the following properties:

- $-1 \leq \rho_{12} \leq 1$
- Positive values indicate correlation.
- Negative values indicate anti-correlation.
- $\rho_{12} = 1$ is perfect correlation.
- $\rho_{12} = -1$ is perfect anti-correlation.

Two-point Statistical Measures

Autocovariance measures how a variable changes with different lags, s.

$$R(s) \equiv \overline{u(t)u(t+s)}$$

or the *autocorrelation function*

$$\rho(s) \equiv \frac{\overline{u(t)u(t+s)}}{u(t)^2}$$

Or for the discrete form

$$\rho(s_j) \equiv \frac{\sum_{k=0}^{N-j-1}(u_k u_{k+j})}{\sum_{k=0}^{N-1}(u_k^2)}$$

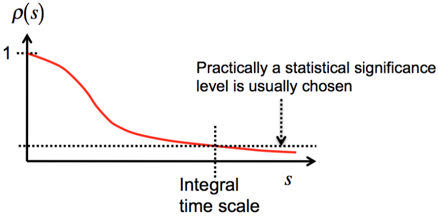
Two-point Statistical Measures

Notes on autocovariance and autocorrelation

- These are very similar to the covariance and correlation coefficient
- The difference is that we are now looking at the linear correlation of a signal with itself but at two different times (or spatial points), i.e. we lag the series.
- We could also look at the cross correlations in the same manner (between two different variables with a lag).
- $\rho(0) = 1$ and $|\rho(s)| \leq 1$

Two-point Statistical Measures

- In turbulent flows, we expect the correlation to diminish with increasing time (or distance) between points
- We can use this to define an integral time (or space) scale. It is defined as the time lag where the integral $\int \rho(s)ds$ converges.
- It can also be used to define the largest scales of motion (statistically).



Two-point Statistical Measures

The *structure function* is another important two-point statistic.

$$D_n(r) \equiv \overline{[U_1(x+r,t) - U_1(x,t)]^n}$$

- This gives us the average difference between two points separated by a distance r raised to a power n.
- In some sense it is a measure of the moments of the velocity increment PDF.
- Note the difference between this and the autocorrelation which is statistical linear correlation (*i.e.*, multiplication) of the two points.

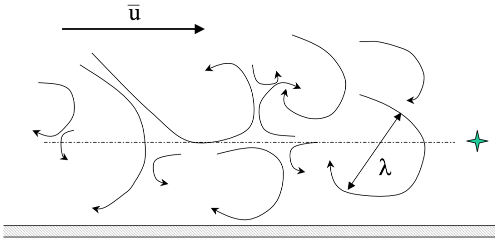
Taylor's Frozen Turbulence Hypothesis

- It is very difficult to produce highly spatially resolved measurements of temperatures and velocities over a large spatial region at one instant in time
- Thus, we sually measure over large time periods at very few points in space (i.e., a sonic anemometer mounted on a tower or a hot-wire probe in a wind tunnel).
- G.I. Taylor (1938) proposed an idea that for some special cases, turbulence might be considered "frozen" as it advects pass our measuring device.

Taylor's Frozen Turbulence Hypothesis

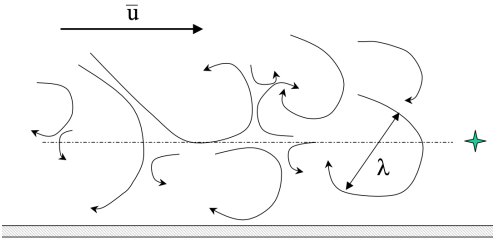
- As a result turbulence measurements that are made as a function of time can be translated into a corresponding spatial measurement.
- This hypothesis is useful for cases where turbulent eddies evolve with a timescale longer than the time scale it takes the eddy to be advected past the sensor.

Taylor's Frozen Turbulence Hypothesis



- Consider schematic of turbulence in a boundary layer.
- One way to measure the velocities along the line shown at an instant in time, would be to place sensors all along the line.

Taylor's Frozen Turbulence Hypothesis



- Another way would be to move the probe very quickly through the flow at some known velocity assuming the flow doesn't change much while you traverse it.
- One other way would be to leave the probe in one place and allow the fluid to advect past the probe.
- The last two ways utilize Taylor's Frozen Turbulence hypothesis

Taylor's Frozen Turbulence Hypothesis

- Following Stull (1988), the substantial derivative is zero for Taylor's Hypothesis
- Thus,

$$\frac{\partial \zeta}{\partial t} = -\overline{u}\frac{\partial \zeta}{\partial x} - \overline{v}\frac{\partial \zeta}{\partial y} - \overline{w}\frac{\partial \zeta}{\partial z}$$