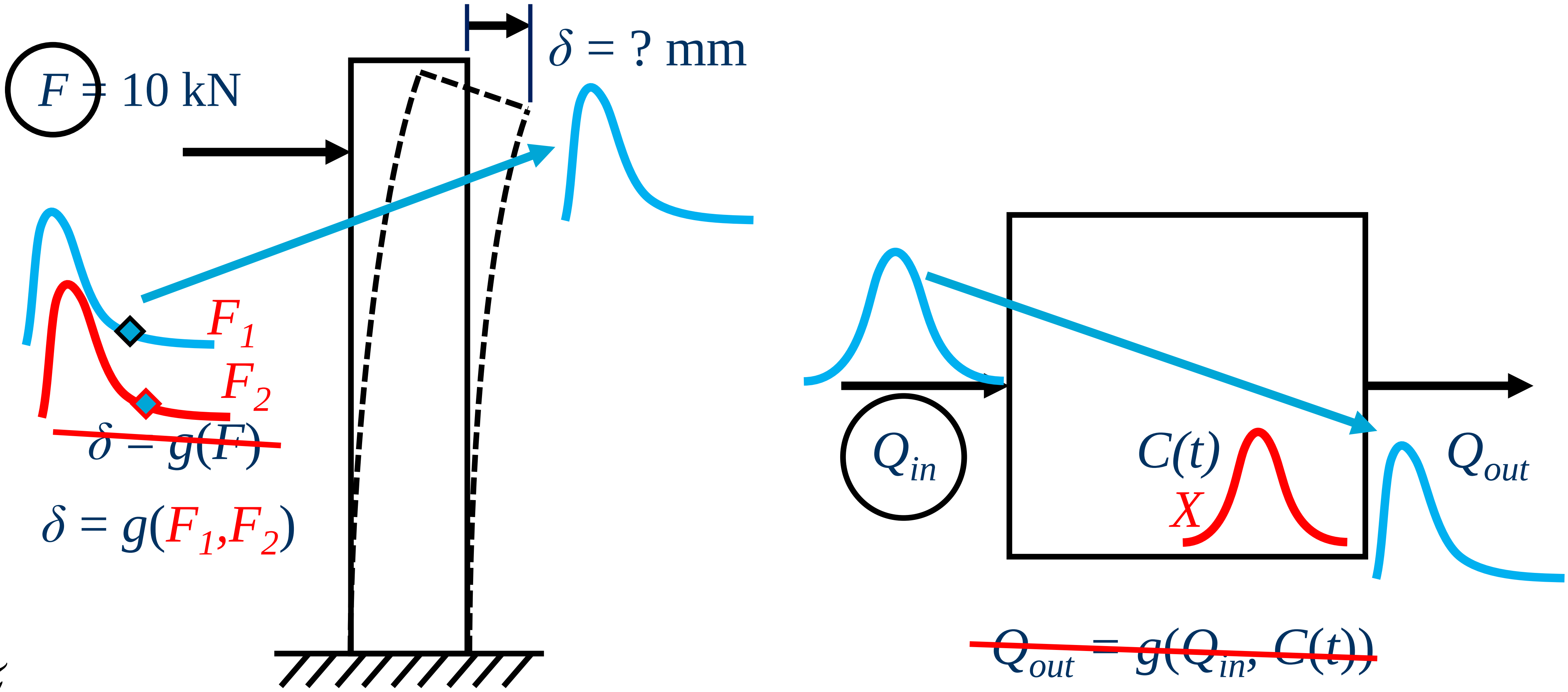
Modelling, Uncertainty and Data for Engineers (MUDE)

Week 1.8 : Multivariate Distributions

There was a lot of work on the board in this lecture. Some slides have been added as a supplement/quick reference (with a box like this), but the recording should be watched to get the complete story; a similar example to the island case here is worked out in the textbook (Chapter 6).
https://collegeramavideoportal.tudelft.nl/catalogue/cegm1000

TU Delft

Incorporating (More!) Uncertainty in our Models



$\rho(X_1, X_2) = ???$

$Q_{out} = g(Q_{in}, C(t, X))$

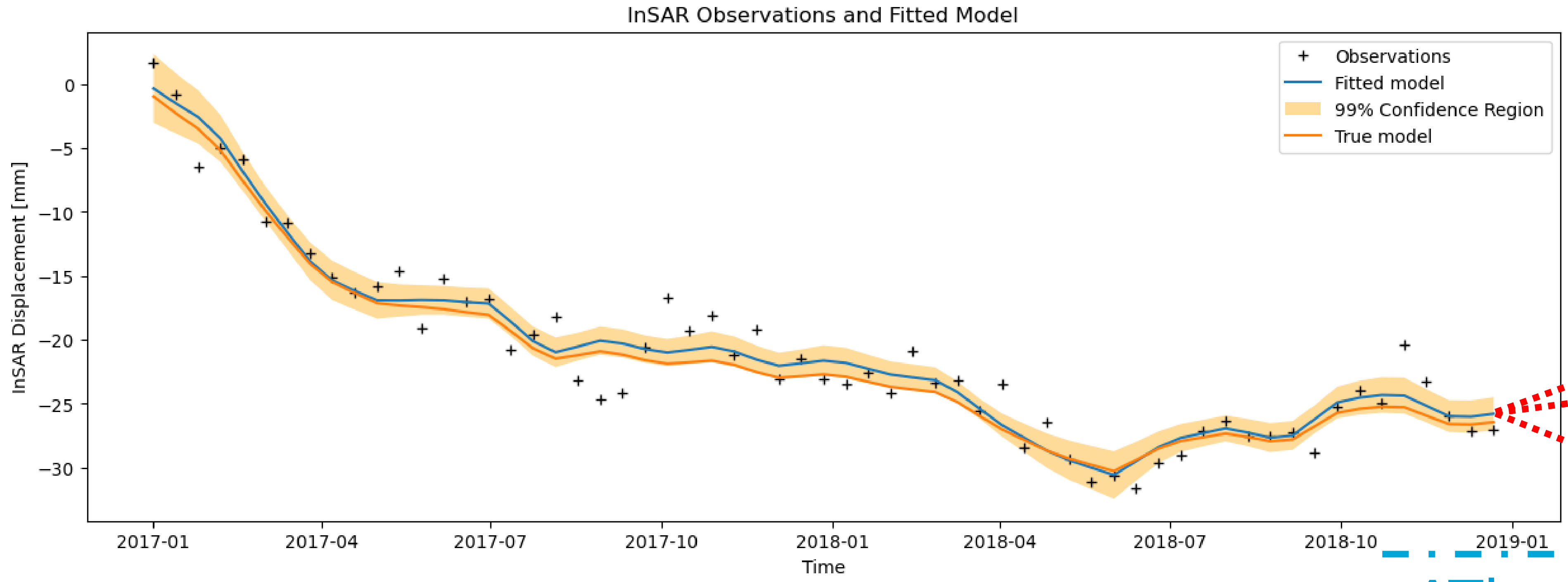
Project 2

$$d = d_0 + vt + k\ \mathrm{GW}$$

- d_0: initial displacement at t_0
- v: displacement velocity
- k: soil response to change in GW

$$d = d_0 + R(1 - \exp -\frac{t}{a}) + k\ \mathrm{GW}$$

- d_0: initial displacement at t_0
- R: soil response to road
- k: soil response to change in GW



To use model, we need:

- **Multivariate model, incorporating probability**
- **Non-Gaussian Dist.'s**
- **Non-linear functions**

?

?

^Threshold?

?

P(*d* exceeds threshold) = ?

TU Delft

Today: tools for creating a multivariate model

- Covariance, dependence
- Functions of random variables
- Monte Carlo simulation

- We've already done these things before!
- A bit of context + math to get more comfortable with probability

Why do we want to use probability?

A simple example

See Collegerama Recording

Oceanic island example:

- Coral atoll, susceptible to sea level rise and settlement due to dissolution of coral (calcium carbonate)
- Current elevation = 10 m + mean sea level (MSL)
- Settlement, D ~ N(μ = 3m, σ = 1.0 m)
- Sea level rise, S ~ N(μ = 5m, σ = 1.6 m)
- Observations: from a few atolls around the world (440 in total)
- Distributions: prediction of what will happen over the next 100 year period

- How can we evaluate the probability that the island is underwater?

  See Chapter 6 in textbook for similar example

Explanation slide of board work in lecture.

Settlement Due to Ocean Acidification

Corals Are Dissolving Away

- ..researchers monitored 57 locations at five coral reefs around the world… [and] found a strong correlation between the dissolving process and calcium carbonate levels in the water…the dissolving process … [is] sensitive to ocean acidification…
- By Chelsea Harvey & E&E News, February 23, 2018
- https://www-scientificamerican-com.tudelft.idm.oclc.org/article/corals-are-dissolving-away1/

- Bradley D. Eyre et al. ,Coral reefs will transition to net dissolving before end of century.Science **359**,908-911(2018).DOI:10.1126/science.aao1118
- https://www-science-org.tudelft.idm.oclc.org/doi/10.1126/science.aao1118

Image: The atoll of Teti'aroa in French Polynesia. Supertoff. https://en.wikipedia.org/wiki/Atoll#/media/File:Tetiaroa_from_sky.JPG

Probability Calculations – a summary of the lecture

- First note that we are after $P(D \cap S)$, where $P(D) = 0.2$ and $P(S) = 0.4$.
- $P(D \cap S) = P(S)P(D) = 0.08$ **iff** they are independent. This is not always a good assumption.
- Using empirical (data on next slide), we see 0.10 compared to 0.08.
- Challenge: → how do we get $P(D \cap S)$?
  - → a multivariate distribution!

- Regardings these distributions, we covered this in lecture:
  - How to integrate density in 2D to get a probability
  - What the contours are and what they tell us about probability
  - Correlation coefficient and how that influences contours/probability

Explanation slide of board work in lecture.

Data*: Observations of Coral Atolls

D → [2.1, 2.6, 4.3, 3.8, 2.5, 4.7, 1.4, 1.9, 3.6, 3.1]

S → [5.1, 3.2, 7.2, 4.8, 6.5, 4.1, 2.4, 6.2, 6.9, 3.6]

Elevation of island is 10 m above sea level.

Find probability that it is underwater after 4 m settlement and 6 m sea level rise.

*data on this page is carefully selected to have nice properties for this lecture. In other words: it is total BS.

Data*: Observations of Coral Atolls

D → [2.1, 2.6, **4.3**, 3.8, 2.5, 4.7, 1.4, 1.9, 3.6, 3.1]

S → [5.1, 3.2, **7.2**, 4.8, 6.5, 4.1, 2.4, 6.2, 6.9, 3.6]

$$P[D \cap S] = \frac{1}{10} \quad \rightarrow \quad \neq P(A)P(S) \;\; and \;\; \neq (1 - F_D(d = 4))(1 - F_S(s = 6))$$

Why?!

*data on this page is carefully selected to have nice properties for this lecture. In other words: it is total BS.

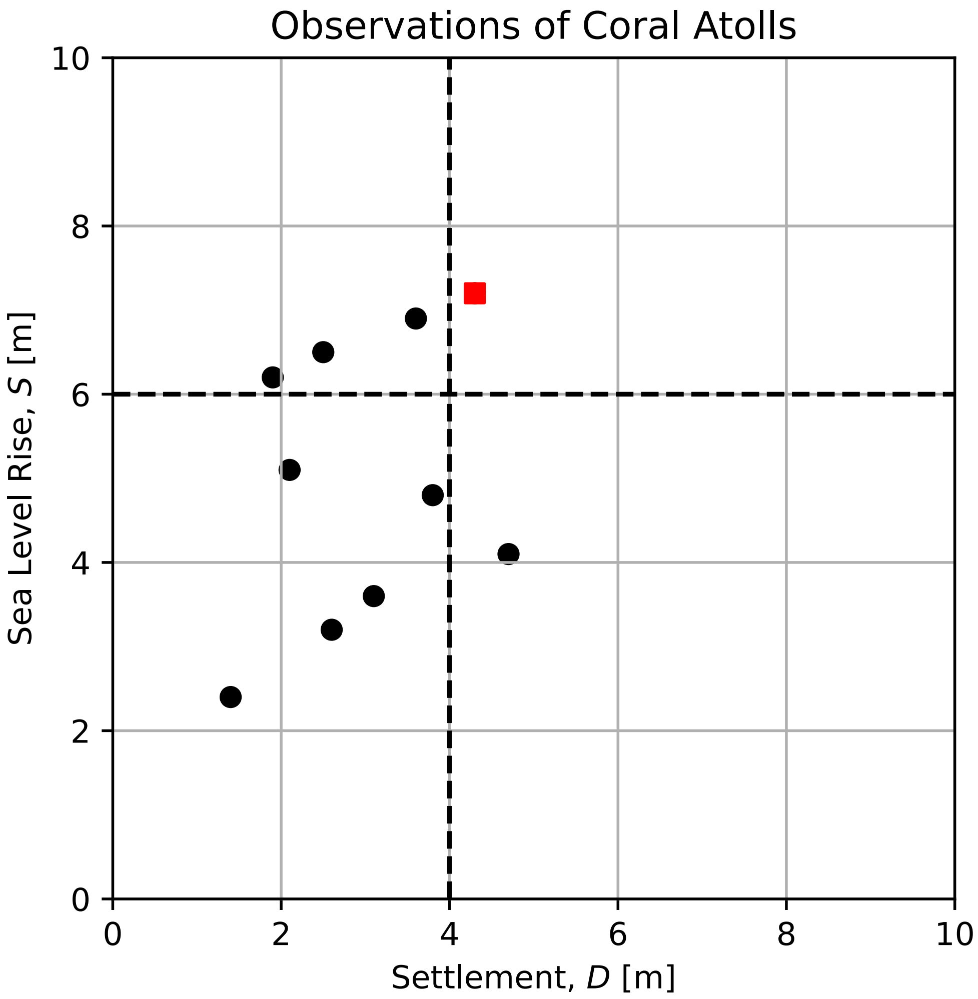
Data*: Observations of Coral Atolls

D → [2.1, 2.6, **4.3**, 3.8, 2.5, 4

S → [5.1, 3.2, **7.2**, 4.8, 6.5, 4

$$P[D \cap S] = \frac{1}{10} \quad \rightarrow \quad \neq P(A)P(S) \quad and \quad \neq$$

Why?!

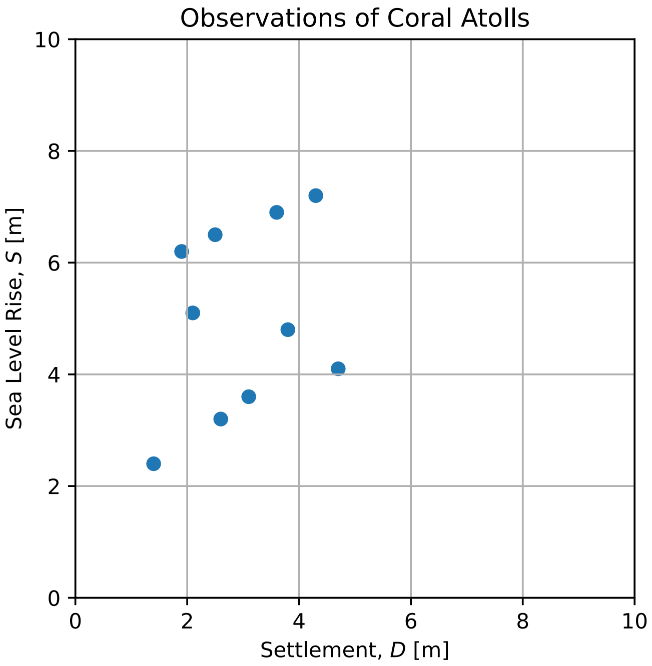


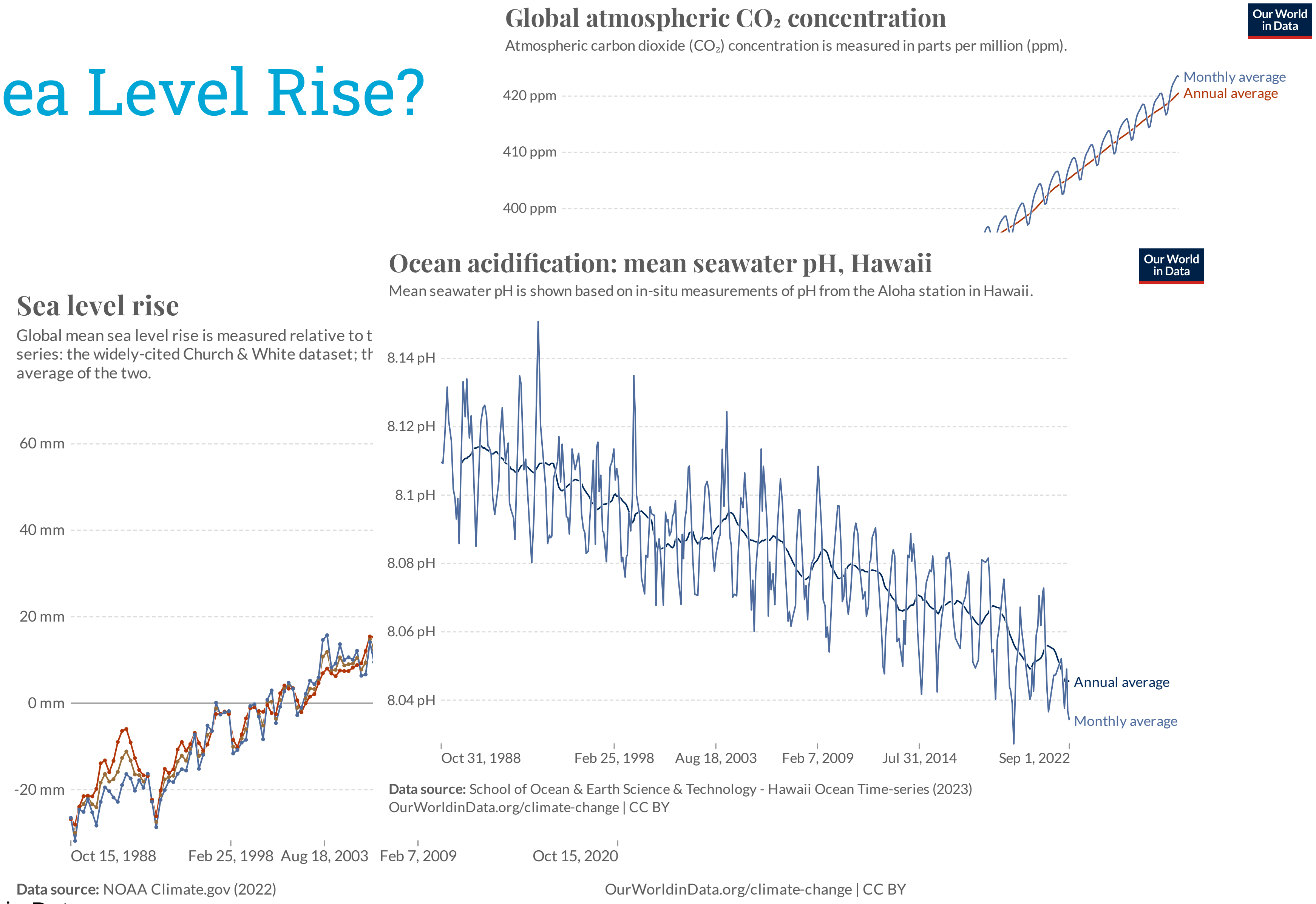
*data on this page is carefully selected to have nice properties for this lecture. In other words: it is total BS.

Covariance, Dependence

- Definitions (refresher)
- Influence of (Pearson) correlation coefficient, ρ
- Bivariate perspective

Settlement v. Sea Level Rise?





Not covered in lecture.
Illustration that dependence plays a role.

Source: Our World in Data
Ocean acidification: mean seawater pH, Hawaii: https://ourworldindata.org/grapher/seawater-ph
Global atmospheric CO_2 concentration: https://ourworldindata.org/grapher/global-co2-concentration?time=1988-09-15..latest
Sea level rise: https://ourworldindata.org/grapher/sea-level?time=1988-10-15..latest

D and S as Random Variables (not events)

$$D \sim \mathrm{N}(\mu = 3.0, \sigma = 1.0)$$

$$S \sim \mathrm{N}(\mu = 5.0, \sigma = 1.6)$$

Elevation $\mathrm{Z} = 10$ m above sea level

Function of random variables

- Define a function $Z(D,S)$ to describe underwater case

$$Z(D,S) = 10 - D - S$$

Find probability $P[Z < 0]$

Function of random variables

$$Z(D, S) = 10 - D - S$$

$$D \sim \mathrm{N}(\mu = 3.0, \sigma = 1.0) \quad S \sim \mathrm{N}(\mu = 5.0, \sigma = 1.6)$$

$$\rho(X_1, X_2) = 0.3$$

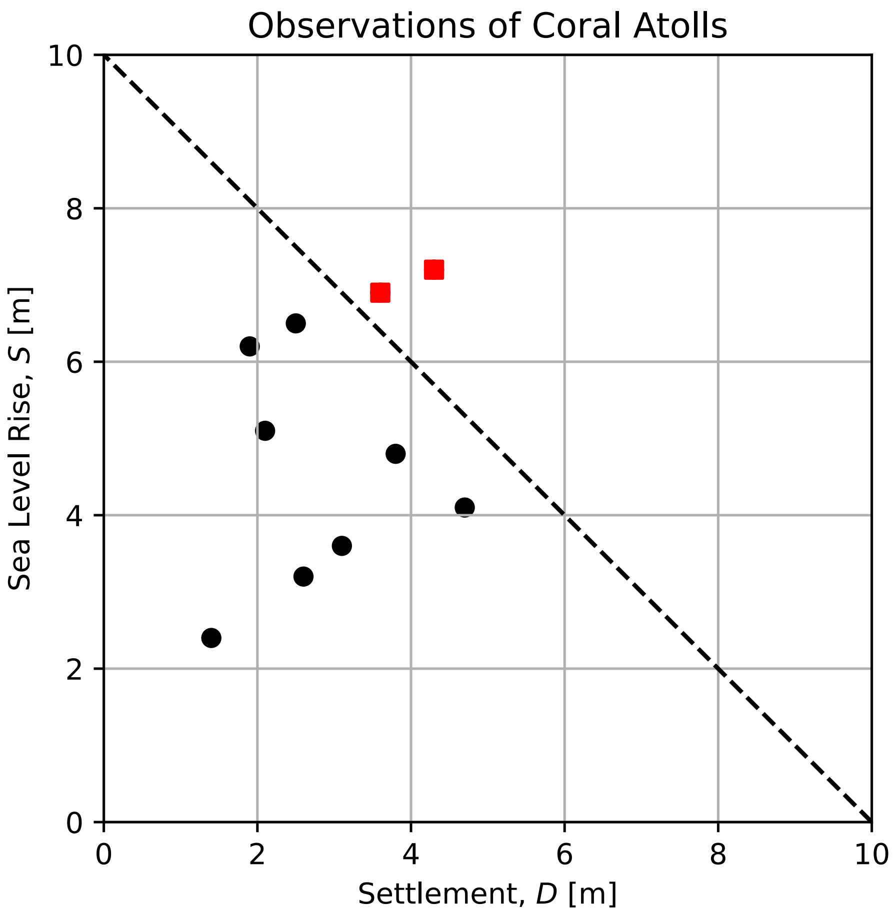
Not covered in lecture.
Can evaluate analytically!

Find probability $P[Z < 0]$

Function of random variables

$$Z(D,S) = 10 - D - S$$

Not covered in lecture.
Can evaluate analytically!



Multivariate Distributions

- Independent Marginals
- Multivariate Gaussian
- Copulas

→ Density contours and samples in 2D are useful!

Copula Page in Book

Definition of bivariate copula

The definition of copula for the bivariate case is given by

$$F_{X_1,X_2}(x_1, x_2) = C\left[F_{X_1}(x_1), F_{X_2}(x_2)\right]$$

where $F_{X_1,X_2}(x_1, x_2)$ for $(x_1, x_2) \in \mathbb{R}^2$ is a joint distribution with marginals $F_{X_1}(x_1)$ and $F_{X_2}(x_2)$ in [0, 1] and C is a copula in the unit square $I^2 = ([0,1] \times [0,1])$, being this equation satisfied for all $(x_1, x_2) \in \mathbb{R}^2$.

Therefore, the joint density is given as the product of the density of the copula, c, and the densities of the marginals as

$$f_{X_1,X_2}(x_1, x_2) = f_{X_1}(x_1) f_{X_2}(x_2) c(F_{X_1}(x_1), F_{X_2}(x_2))$$

Non-linearity

- Functions
- Marginal Distributions
- Dependence

TUDelft

Programming Aspects

- Multivariate Distributions
- Monte Carlo Simulation

TUDelft

Copula in Python (pyvinecopulib)

- `X = stats.multivariate_normal(mean=[3, 5],`
- `cov=[[1.00, rho],`
- `[rho, 2.60]])`

$$X = [X_1 \quad X_2 \quad \dots \quad X_m]^T$$

$$f_X(x) = \frac{1}{\sqrt{\det(2\pi\Sigma_X)}} \exp(-\frac{1}{2}(x - \mu_X)^T \Sigma_X^{-1} (x - \mu_X))$$

TU Delft

Copula in Python (pyvinecopulib)

- X1: scipy.stats.rv_continuous object. Marginal distribution
- X2: scipy.stats.rv_continuous object. Marginal distribution
- rho: float. Pearson correlation coefficient.
- copula: cop.Bicop(family=pyvinecopulib.BicopFamily.gaussian, parameters = [rho])

- x1 = x[0]
- x2 = x[1]
- u =X1.cdf(x1)
- v =X2.cdf(x2)
- Return copula_uv.cdf(np.array([[u, v]]))[0]

$$F_{X_1,X_2}(x_1,x_2) = C\left[F_{X_1}(x_1), F_{X_2}(x_2)\right]$$

$$f_{X_1,X_2}(x_1,x_2) = f_{X_1}(x_1) f_{X_2}(x_2) c(F_{X_1}(x_1), F_{X_2}(x_2))$$

TUDelft

Monte Carlo Simulation

Associated terms: propagation of uncertainty, realizations, samples

Steps:

1. Define distributions for random variables (probability density over a domain)

   Hint: `xxx.rvs(size=N)`

   A method of parent class **`rv_continuous`**

2. Generate random samples

3. Do something with the samples (deterministic calculation)

4. Evaluate the results: e.g., "empirical" PDF, CDF of samples, etc.

When RV's are *not* independent, make sure you sample from the joint distribution!
→ **`scipy.stats.multivariate_normal.rvs()`**

Wednesday and Friday: build on last week, add…

- Bivariate joint distributions:
  → linear (Gaussian) dependence structure
  → non-Gaussian marginals
- Region of interest: thresholds, functions, etc.
- Contour plots
- Evaluating the effect of dependence (qualitative and quantitative)

Thanks for joining the…

Modelling, Uncertainty, and Data for Engineers

Last Q1 Lecture!

TU Delft