Exam Q2

CEGM1000 Modelling, Uncertainty and Data for Engineers EXAM 24/25 · 10 exercises · 90.0 points

1 Programming

9.0 points · 3 questions

A colleague sends you a piece of code in the file class.py, with instructions to import it using the following line of code:

```
from class import Class
```

Text

a Which line below would be the proper way to instantiate a class contained in class.py ?

3.0 points · Multiple choice · 5 alternatives

○ C = class	0.0
○ class	0.0
◉ C = Class()	3.0
◉ Class()	1.0
○ Instantiate(Class)	0.0

Feedback

Feedback when the question is answered correctly

Feedback when the question is answered partially correctly

Feedback when the question is answered incorrectly

Consider the code snippet below, then answer the following **2** questions. Note that the file data.csv contains the date and daily measurements of an unknown property over a total period of 4.5 years.

```
D = pd.read_csv('data.csv', delimiter = ',', parse_dates = True)
indices = D.groupby(pd.DatetimeIndex(D['Date']).year)['olddata'].idxmax()
newdata = D.loc[indices]
print(newdata.shape)
```

Text

b The output from running this cell will **most likely** be:

3.0 points · Multiple choice · 6 alternatives

- ○ (5,)
- ○ (5, 1)
- ◉ **(5, 2)**
- ○ (54,)
- ○ (54, 1)
- ○ (54, 2)

Feedback

Feedback when the question is answered correctly

Feedback when the question is answered partially correctly

Feedback when the question is answered incorrectly

c Select all that are true:

3.0 points · Multiple choice · 5 alternatives

- [x] **D is an object**
- [ ] D is an array
- [x] **D is a data frame**
- [ ] D is a dictionary
- [ ] D is a dinosaur

Feedback

Feedback when the question is answered correctly

Feedback when the question is answered partially correctly

Feedback when the question is answered incorrectly

2 Finite Volume Methods

14.0 points · 4 questions

a Select the correct option to fill in the "[blank]" in the following sentence:

In [blank] equations, the information travels at a finite speed.

Hint: remember that for our purposes pure advection is hyperbolic; pure diffusion can be elliptic or parabolic; and, the classification of diffusion-advection is problem-specific.

3.0 points · Multiple choice · 5 alternatives

- (●) **Hyperbolic**
- ( ) Elliptic or parabolic
- ( ) Hyperbolic, elliptic **or** parabolic
- ( ) Hyperbolic, elliptic **and** parabolic
- ( ) None of the above

Feedback

Feedback when the question is answered correctly

Feedback when the question is answered partially correctly

Feedback when the question is answered incorrectly

b The equation:

$$\frac{\partial \phi}{\partial t} + c\frac{\partial \phi}{\partial x} = 0$$

is solved numerically using central differences in space and the resulting ODE is solved with forward Euler in time. Which type of may occur?

3.0 points · Multiple choice · 4 alternatives

- ○ Numerical diffusion
- ◉ **Numerical instability**
- ○ Spatial accuracy order $O\left(\Delta x\right)$
- ○ All of the above

Feedback

Feedback when the question is answered correctly

Feedback when the question is answered partially correctly

Feedback when the question is answered incorrectly

c Recall that Gauss's theorem can be written as:

$$\int_V \nabla \cdot \overrightarrow{r} \mathrm{dV} = \int_S \overrightarrow{n} \cdot \overrightarrow{r} dS$$

Integrate the diffusion equation:

$$\frac{\partial \phi}{\partial t} = D \frac{\partial^2 \phi}{\partial x^2}$$

over an arbitrary volume and apply Gauss's theorem to the diffusion term. There is no need to solve the surface integral.

5.0 points · Open · 7/10 Page

+1 point

+5 points

Integrate the diffusion equation:

$\frac{\partial \phi}{\partial t} = D \frac{\partial^2 \phi}{\partial x^2}$

over an arbitrary volume and apply Gauss's theorem to the diffusion term. There is no need to solve the surface integral.

$\int_V \frac{\partial \phi}{\partial t} \, dV = \int_V D \frac{\partial^2 \phi}{\partial x^2} \, dV$

Using Gauss's theorem:

$\int_S D \frac{\partial \phi}{\partial x} \cdot \mathbf{n} \, dS$

+2 points

Setting up the volume integral: $\int_V \frac{\partial \phi}{\partial t} \, dV = \int_V D \frac{\partial^2 \phi}{\partial x^2} \, dV$

+3 points

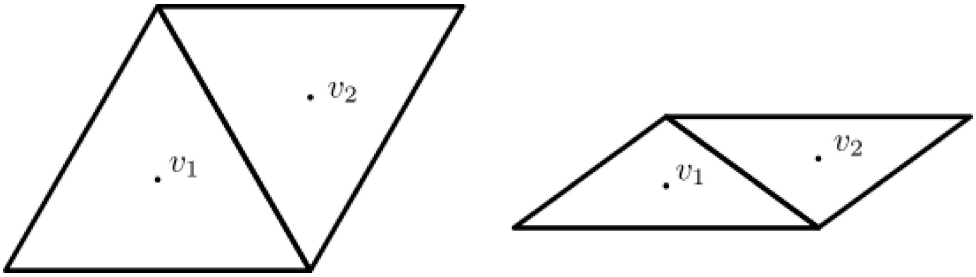
Using Gauss's theorem correctly:

$\int_S D \frac{\partial \phi}{\partial x} \cdot \mathbf{n} \, dS$

-1 points

Introducing non-relevant or "strange" operations

The flux due to diffusion in orthogonal meshes, as shown below. can be approximated adequately with a central difference scheme. However, in non-orthogonal meshes, new error sources may appear.



Text

d
State the two main reasons why errors would appear.

3.0 points · Open · 1/2 Page

+3 points

both reasons are good:

- The approximated gradient may not be defined at the midpoint of the face connecting the volumes.
- The gradient/flux may not be normal to the face.

-1 points

a little error or reason not quite right

+1.5 points

one reason is right: The approximated gradient may not be defined at the midpoint of the face connecting the volumes.

+1.5 points

one reason is right:
The gradient/flux may not be normal to the face.

3 Finite Element Methods

7.0 points · 2 questions

Consider the following code for the computation of one contribution to the discrete system of equations with the finite element method in 1D.

Text

Screenshot 2025-01-20 103513.png

Image 3/20 Page

```python
def get_element_vector(value, dx):
    locations = [(dx - dx/(3**0.5))/2, (dx + dx/(3**0.5))/2]
    weights = [dx/2, dx/2]

    elvec = np.zeros((3,1))

    for x, w in zip(locations, weights):
        N = evaluate_N(x,dx)
        elvec += np.transpose(N)*value*w

    return elvec
```

This function is part of a finite element code that solves the PDE $a\frac{\partial^2 u}{\partial x^2} + bu = c$ for $u(x)$ with given parameters a, b and c.

Text

a Which of the three parameters (a, b or c) is present in this function under the variable name "**value**"?

3.0 points · Multiple choice · 4 alternatives

- ○ a
- ○ b
- ◉ c
- ○ The same function can be used for a, b and c.

Feedback

Feedback when the question is answered correctly

Feedback when the question is answered partially correctly

Feedback when the question is answered incorrectly

b If at one end of the domain, no boundary condition is implemented in the finite element code,
which boundary condition will be imposed by default? You do not need to write more than one equation and 1-2 sentences.

4.0 points · Open · 3/10 Page

+4 points
homogeneous Neumann BC, or ∂u/∂x = 0 + explanation

+2 points
No reasoning but correct expression, or vice versa

+1 point
only name

4 Signal Processing

5.0 points · 1 question

We start from a signal which is the sum of two cosines, both with unit amplitude and zero phase, one with a frequency of 4 Hz, and the other with 80 Hz. The signal is sampled at $f_s = 100$ Hz, for a duration of $T = 2$ seconds.

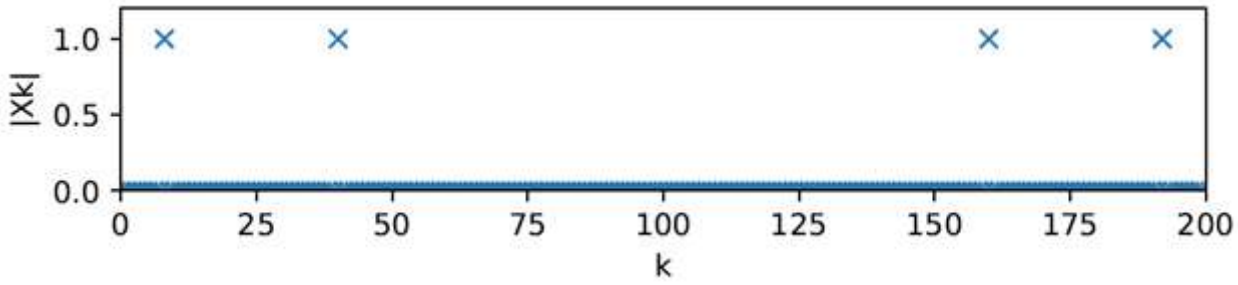
The $N = 200$ discrete time samples are input to the Discrete Fourier Transform (DFT). We directly plot the magnitude (modulus) of the output, hence $|Xk|$, multiplied by $D_t = \frac{1}{f_s}$, with $k = 0, 1, \ldots, N-1$.
(i.e. Xk=np.fft.fft(xt), and plt.plot(abs(Xk)/fs) in code).

The resulting graph is shown below:

Text

Picture1.png

Image 1/10 Page



Provide an adequate and correct labelling of the horizontal axis, such as the minimum and maximum bounds, **and** annotate the four peaks.

Please note that we have provided two images for you to use in case you need to start over. **Please cross over the image you do NOT want to submit.** If it is unclear which image you want to submit, we will only grade the bottom one.

5.0 points · Image · 2/5 Page

+1 point
correct labelling horizontal axis [0, 100), or [-50, 50)

+4 points
All four labels correct: 4 Hz, 20 Hz, 80 Hz, 96 Hz **OR** -4 Hz, -20 Hz, -80 Hz, -96 Hz

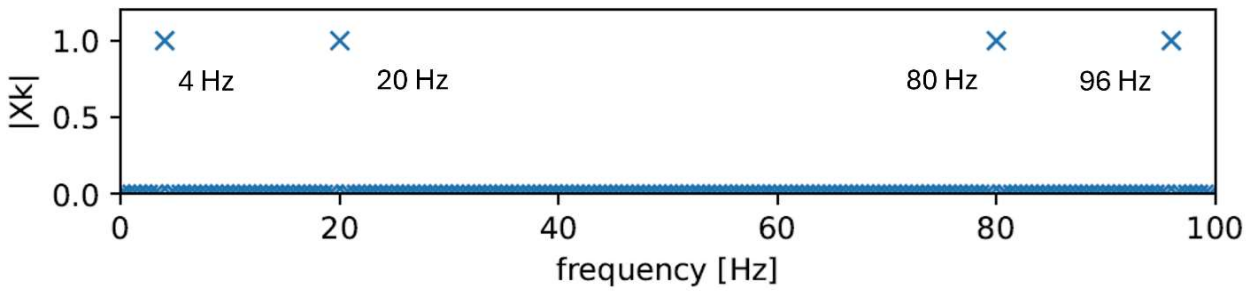
+3 points
Any of the following:
- 4 Hz, 80 Hz, 120 Hz, 196 Hz
- 4 Hz, 20 Hz, 30 Hz, 46 Hz

+2 points
Any of the following:
- -20 Hz, -4 Hz, 4 Hz, 20 Hz
- -6 Hz, -4 Hz, 4 Hz, 6 Hz (provided that hor. axis label is in line, e.g. [-7, 7] approx)
- only 4 Hz and 80 Hz labelled (at correct positions)

+1 point
Any of the following:
- -80 Hz, -4 Hz, 4 Hz, 80 Hz
- only 4 Hz (correctly) labelled
- only 80 Hz (correctly) labelled

+2.5 points
only 4 Hz and 20 Hz labelled (at correct positions)

+4 points



5 Time Series Analysis

12.0 points · 3 questions

The ice thickness at a specific location is measured on a weekly basis during 10 weeks $(t = 0, 1, \ldots, 9)$, and is assumed to increase linearly in time.

Least-squares is applied to fit a linear trend, which provides $\hat{\mathbf{x}} = [\hat{x}_0,\ \hat{v}]^T = [1.6,\ 0.12]^T$.($\hat{x}_0$ is the estimated intercept, $\hat{v}$ the estimated slope).

The estimated linear trend is then subtracted from the observations $\mathbf{y}$ to obtain the stationary time series:

$$S := \hat{\epsilon} = y - A\hat{x} = \begin{bmatrix} 0.23 \\ 1.27 \\ 1.16 \\ -0.52 \\ -1.90 \\ -0.50 \\ -1.60 \\ -0.62 \\ 0.04 \\ 1.23 \end{bmatrix}$$

The stochastic process underlying this stationary time series can be modeled as an AR(1) process with $\phi_1 = 0.45$.

Text

Below you can find important formulas that you may need to use.

AR(p) process: $S_t = \sum_{i=1}^{p} \phi_i S_{t-i} + e_i$

with zero mean random noise, $\mathbb{E}(e_t) = 0$

Autocovariance function for AR(1) process: $c_r = \sigma^2 \phi^T$

Normalized autocovariance function: $\rho_r = \dfrac{c_r}{c_0}$

Prediction formula: $\hat{Y}_p = \mathrm{A}_p \hat{X} + \hat{\epsilon}_p$

Note that this information is for the next three questions on the following pages.

Text

a Make a sketch (no calculation needed) of the ACF for an AR(1) process with $\phi_1 = 0.45$.

4.0 points · Open · 1/2 Page

+1 point
1st lag has ac of +/- .45; 0 lag 1 if present

+1 point
ACF is strictly decreasing in x

+1 point
Values are significant up to +/- lag 3. Or decaying nature is consistent

+1 point
When not significant any more acf shows small random fluctuations around 0. Not showing insignificant points is also marked as correct

-1 points
small mistake

-1 points
no clear y or x axis scale/labeling

+1 point
Large mistake, such as perfect correlation with the first lag (instead of lag 0) but still some merit to the plot

b Assume $e_{10} = 0$, what is the predicted value of $\hat{\epsilon}_{10}$?

4.0 points · Open · 1/2 Page

+4 points
$\hat{\epsilon}_{10} = \phi_1 \hat{\epsilon}_9 = 0.45 \times 1.23 = 0.55$

c What is the predicted value of the ice thickness in week 10?

4.0 points · Open · 1/2 Page

+2 points
$\mathrm{A}_p \hat{x} = 1.6 + 0.12 \times 10$

+2 points
Add the predicted noise from question b
$\hat{Y}_{10} = \mathrm{A}_p \hat{x} + \hat{\epsilon}_{10} = 2.8 + 0.55 = 3.35$

-1 points
small error

6 Optimization

12.0 points · 2 questions

A logistics company manages the transportation of goods between two warehouses (A and B) and three distribution centers (C1, C2, and C3). The company must transport at least 500 kg of goods to C1, 400 kg to C2, and 300 kg to C3.

The transport vehicles have the following constraints:

- The total weight transported from A to all centers cannot exceed 700 kg.
- The total weight transported from B to all centers cannot exceed 800 kg.
- For operational reasons, the weight transported from A to C2 must not exceed 300 kg.

The cost of transporting 1 kg of goods between locations is given below:

From/To	C1	C2	C3
A	2	3	1
B	4	2	5

Text

a Formulate the problem of deciding the amount that should be transported between warehouses and centers. In your formulation please include the decision variables, the objective function and the constraints.

8.0 points · Open · 1 1/2 Page

+2 points

Solution:

Decision Variables:

Let:

- x_{AC1}: Weight transported from A to C1 (kg)
- x_{AC2}: Weight transported from A to C2 (kg)
- x_{AC3}: Weight transported from A to C3 (kg)
- x_{BC1}: Weight transported from B to C1 (kg)
- x_{BC2}: Weight transported from B to C2 (kg)
- x_{BC3}: Weight transported from B to C3 (kg)

+1 point

Objective Function:

Minimize the total transportation cost:

$$Z = 2x_{AC1} + 3x_{AC2} + 1x_{AC3} + 4x_{BC1} + 2x_{BC2} + 5x_{BC3}$$

+5 points

Constraints:

1. Meet demand at each distribution center:

$$x_{AC1} + x_{BC1} \geq 500$$

$$x_{AC2} + x_{BC2} \geq 400$$

$$x_{AC3} + x_{BC3} \geq 300$$

2. Capacity limits at warehouses:

$$x_{AC1} + x_{AC2} + x_{AC3} \leq 700$$

$$x_{BC1} + x_{BC2} + x_{BC3} \leq 800$$

3. Operational constraint for A to C2:

$$x_{AC2} \leq 300$$

4. Non-negativity:

$$x_{AC1}, x_{AC2}, x_{AC3}, x_{BC1}, x_{BC2}, x_{BC3} \geq 0$$

-2 points

Forgetting non-negativity constraint

b One of the distribution centers (e.g., C1) has an urgent demand for goods. Therefore you want to impose that at least 60% of the total weight transported to C1 must come from Warehouse A. Write the constraint that you need to add to the problem.

4.0 points · Open · 1/2 Page

+4 points

If the total weight transported to C1 is:

$$x_{AC1} + x_{BC1}$$

Then the constraint becomes:

$$x_{AC1} \geq 0.6(x_{AC1} + x_{BC1})$$

This can be rewritten as:

$$0.4x_{AC1} \geq 0.6x_{BC1}$$

Or equivalently:

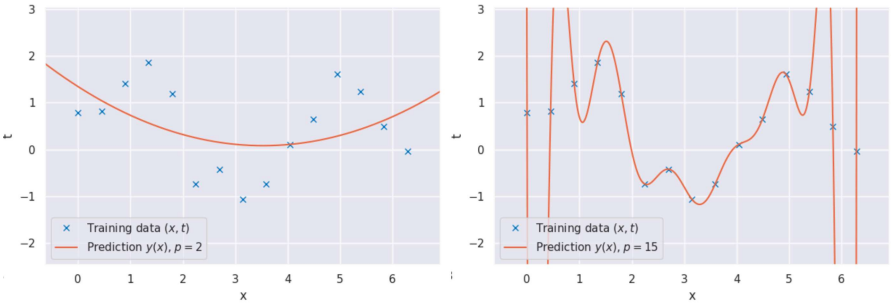
$$2x_{AC1} \geq 3x_{BC1}$$

7 Machine Learning

12.0 points · 3 questions

a

Consider the two following models of the form $y(x) = \mathbf{w}^{\mathrm{T}}\boldsymbol{\phi}(x)$, where $\boldsymbol{\phi}$ is a set of polynomial basis functions up to degree p:



(a) Model A: Polynomial with degree $p = 2$

(b) Model B: Polynomial with degree $p = 15$

which are both fitted to a dataset with $N = 15$ noisy observations of target t. The prediction line shows the regression function $y(x)$. Regarding the suitability of these two models in making new predictions as part of an engineering decision-making process, which **ONE** of the following statements is true?

3.0 points · Multiple choice · 4 alternatives

- Decomposing the expected loss $\mathbb{E}[L]$ into bias, variance and noise parts, Model A would have higher variance (and therefore lower bias) than Model B
- By adding an L_2 regularization term $\frac{\lambda}{2}\mathbf{w}^{\mathrm{T}}\mathbf{w}$ to the error function of Model A, we could reach a similar level of complexity as Model B as we increase the value of λ
- **Starting with Model B and adding an L_2 regularization term, we could reach a suitable model by increasing λ until enough bias has been added to the model**
- Model B is a classic example of an overfit machine learning model. Assuming we are using a validation set to perform model selection, the overfit nature of Model B could be detected by noticing that both training and validation losses go to zero after training

Feedback

Feedback when the question is answered correctly

Feedback when the question is answered partially correctly

Feedback when the question is answered incorrectly

b Still regarding the two polynomial regression models above, which **ONE** of the following statements is true?

3.0 points · Multiple choice · 4 alternatives

- Even though the regression functions have a complex shape, both A and B are still linear regression models in the original feature space $\mathbf{x}$
- Since the models are nonlinear, there is no closed-form (analytical) solution for $\mathbf{w}$, with SGD being therefore required for training
- Model B is equivalent to a neural network with a single hidden layer with 15 units (neurons) with sigmoid activation
- **Both models assume Gaussian observation noise, which means they would be unsuitable to problems with two or more distinct ranges of t values for the same x**

Feedback

Feedback when the question is answered correctly

Feedback when the question is answered partially correctly

Feedback when the question is answered incorrectly

The information here is for us in answering 7c, which is written in the code that follows.
Now assume we would like to fit a neural network to the same dataset with $N = 15$ observations in the figure above. We opt for a Stochastic Gradient Descent optimizer and pre- and postprocess the data accordingly. Consider the following list of operations:

A) Fit input and target normalizers to the full dataset and then normalize training and validation sets separately.
B) Compute the loss over the full training and validation sets and append them to separate lists.
C) Shuffle the training dataset and split it into minibatches withN_b samples each.
D) Plot the evolution of training and validation losses against epochs.
E) Fit a single set of normalizers (input/target) to the training dataset and use it to normalize both training and validation sets.
F) Use the network to get predictions $y(x)$for all validation samples, and then plot them against the corresponding targets t.
G) Shuffle the dataset and split it into a training set and a validation set.
H) For each minibatch, compute $L = \frac{1}{N_b}\sum_{n=1}^{N_b}(t_n - y_n)^2$ and its gradient and update network weights.
I) Use the network to get predictions $y(x)$ for all validation samples, denormalize them, and then plot them against the corresponding targets t.

Text

c Fill in the three blanks in the code block with letters referring to the operations in the list above. Not all operations need to be used, and each blank contains more than one operation. Within each blank, make sure **the operations appear in the correct order** in the code.

Please note that we have provided two images for you to use in case you need to start over. **Please cross over the image you do NOT want to submit.** If it is unclear which image you want to submit, we will only grade the bottom one.

6.0 points · Free formatted question

+1.5 points
G -> E in the correct order and spot

+1.5 points
Either C -> H or H -> B in the correct spot

+3 points
C -> H -> B in the correct order and spot

+1.5 points
D -> I or I -> D in the correct spot

+6 points

```
def train_model(data_X, data_t, n_epochs):
  model = init_network()

  # G -> E

  for epoch in range(n_epochs):
    # C -> H -> B

  # D -> I (or I -> D)

  return model

```

8 Extreme Value Analysis

13.0 points · 3 questions

You are investigating extreme wind speeds and their impacts on buildings along the Netherlands together with a team. A colleagues studying a region close to the German border asked for your help.

You are told that the design wind speed (W) for the buildings in that region is 40 km/h which corresponds to a return period of 50 years and that the design life of the buildings is 25 years.

Text

a Choose ALL THE CORRECT statements. There might be more than one.

3.0 points · Multiple choice · 6 alternatives

- [x] **The expected time between two events equal or higher than 40km/h is 50 years.** 1.0
- [ ] An event equal or higher than 40km/h will not be observed in the next 50 years. -0.5
- [ ] An event equal or higher than 40km/h will not be observed in the next 25 years. -0.5
- [x] **Every year $P[W > 40] = \frac{1}{50}$** 1.0
- [ ] Along a design life of 25 years, $P[W > 40] = \frac{1}{50}$ -0.5
- [x] **The probability of $P[W > 40]$ along the design life of 25 years is higher than the yearly probability of $P[W > 40]$** 1.0

Feedback

Feedback when the question is answered correctly

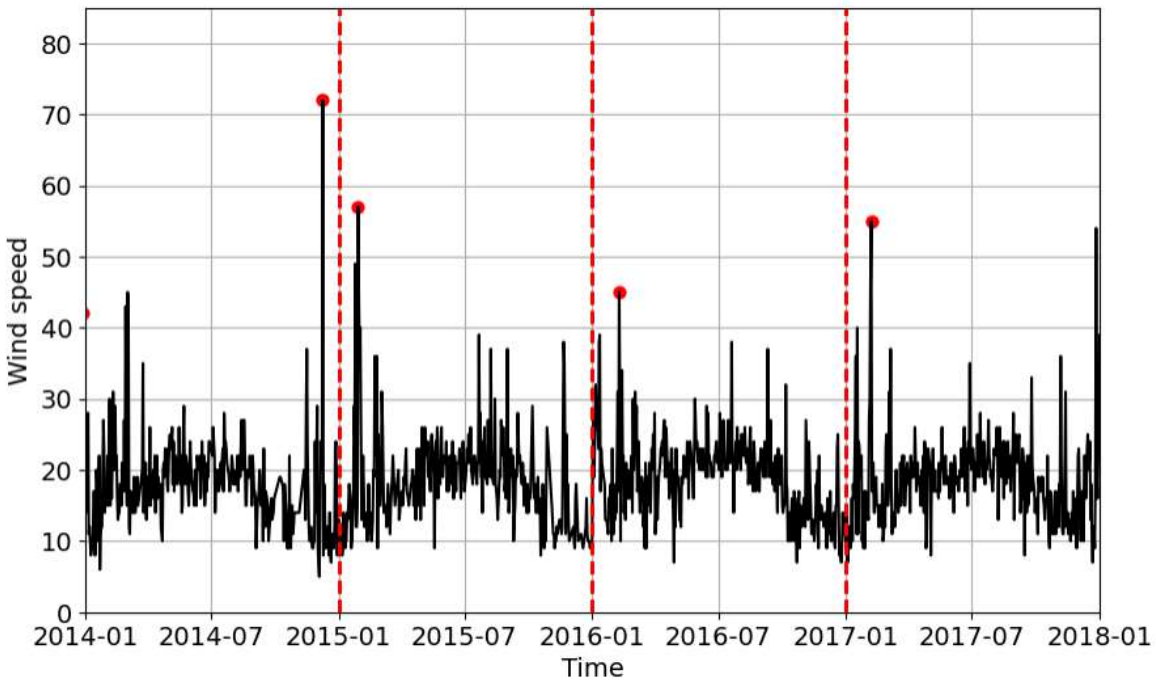
Feedback when the question is answered partially correctly

Feedback when the question is answered incorrectly

b Given the following figure of the time series between 2014 to 2017, how would you find the yearly maxima?
First, describe ONLY the sampling process briefly in the space below; **second,** use the plot provided to sketch it. Please note that we have provided two images in case you need to start over. **Please cross over the image you do NOT want to submit.** If it is unclear which image you want to submit, we will only grade the bottom one.

4.0 points · Open · 2/5 Page

Model answer



The timeseries is divided in blocks of 1 year and the maximum observation in each year is selected.

+4 points
Both answer and sketch are correct.

+2 points
Written answer is fully correct:
- Yearly, so use block maxima
- Have data for 4 years, so 4 maxima

+2 points
Sketch fully correct

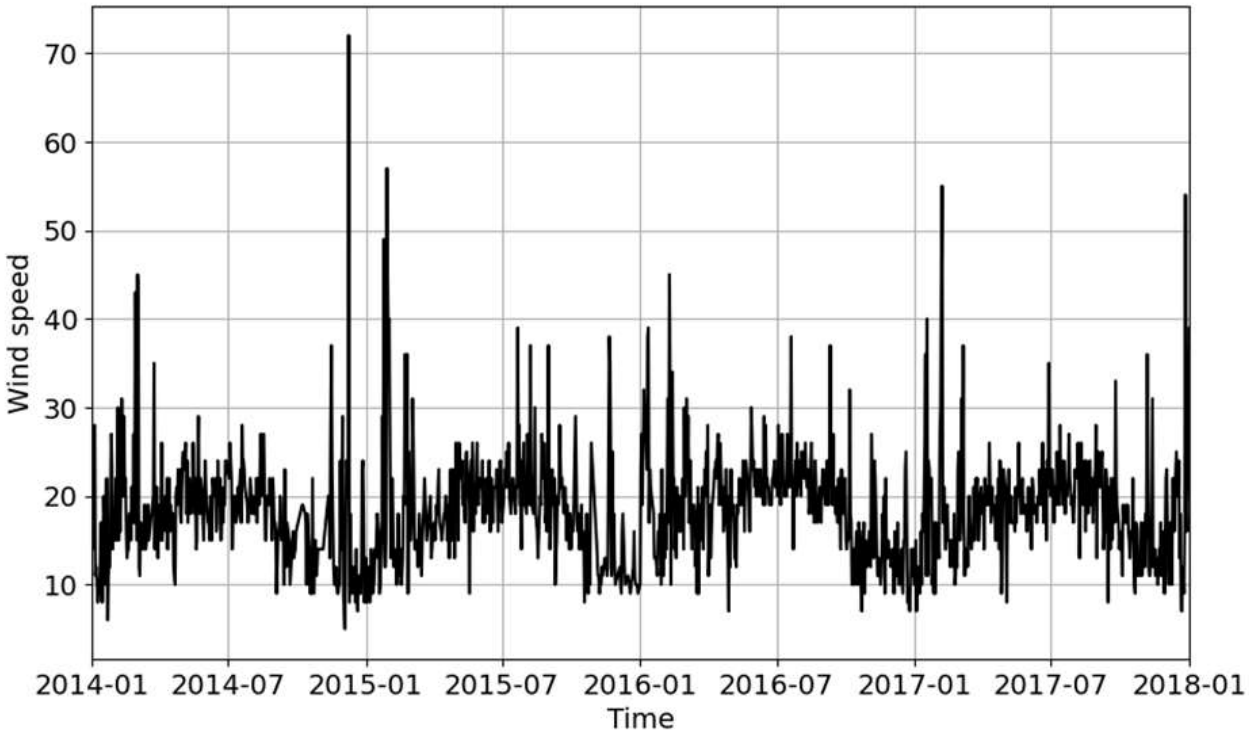
+1 point
Sketch incomplete: blocks are defined but sampled maxima is not indicated

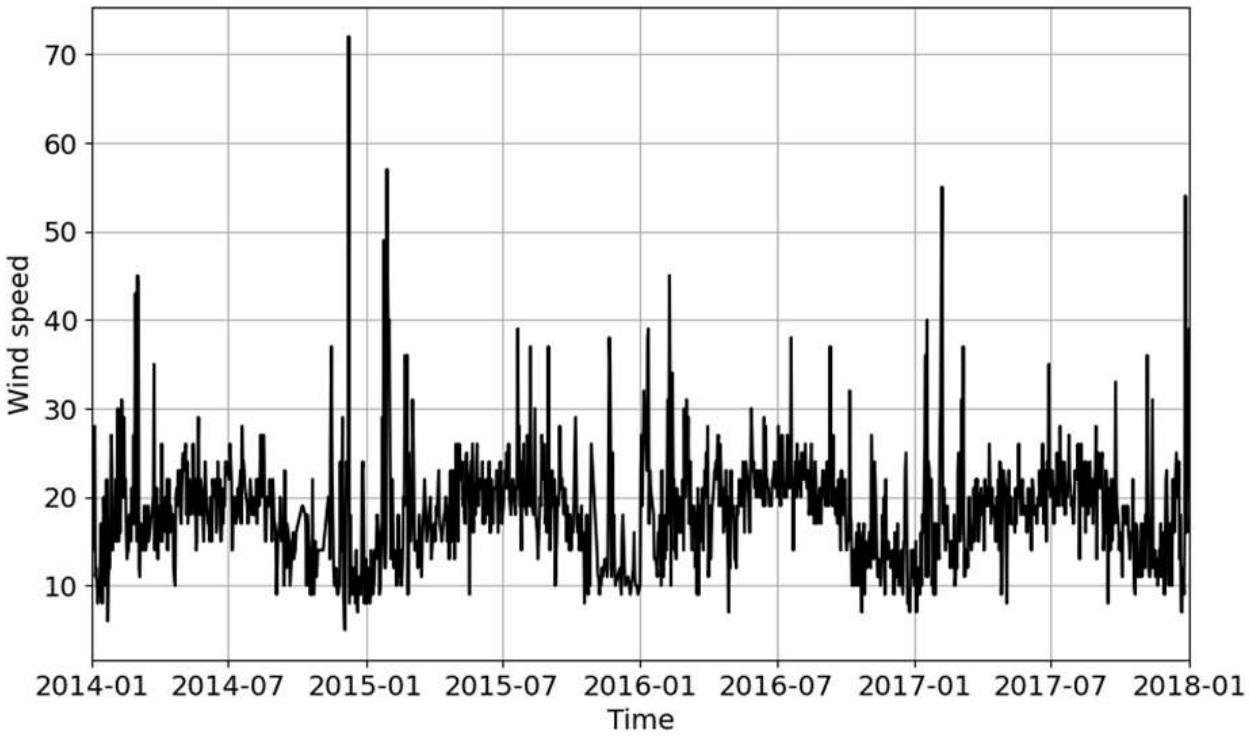
+1 point

Written explanation incomplete

-0.5 points

Small typo: sampling the wrong maxima although right explanation and drafting of the blocks





Text

c Another colleague asked you for your help as probability expert. They have already performed yearly maxima on the observations of wind speed (km/h) and have fitted a Generalized Extreme Value distribution (GEV) so:

$$W \sim \mathrm{GEV}(\mu = 25, \sigma = 10, \xi = -1)$$

The GEV probability density function is given by:

$$g(x) = \frac{1}{\sigma}\left[1 + \xi\frac{x-\mu}{\sigma}\right]^{-\frac{1}{\xi}(\xi+1)} \exp\left(-\left[1 + \xi\frac{x-\mu}{\sigma}\right]^{-\frac{1}{\xi}}\right)$$

and the GEV cumulative distribution function is given by:

$$G(x) = \exp\left(-\left[1 + \xi\frac{x-\mu}{\sigma}\right]^{-\frac{1}{\xi}}\right).$$

Compute the wind speed associated with a return period of 50 years.

6.0 points · Open · 3/5 Page

Model answer

$RT = 50years \rightarrow P[W > x] = 1/50 = 0.02 \rightarrow P[W \leq x] = 1 - 0.02 = 0.98$
$G(x) = exp(-[1 + \xi(x-\mu)/\sigma]^{(}-1/\xi))$
$-ln(G(x)) = [1 + \xi(x-\mu)/\sigma]^{(}-1/\xi)$
$-ln\ [\![(G(x))^{(}-\xi)]\!]\ -1 = \xi(x-\mu)/\sigma$
$x = \sigma/\xi[-ln\ [\![(G(x))^{(}-\xi)]\!]\ -1] + \mu$
$x = 10/(-1)[-ln\ [\![(0.98)^{1}]\!]\ -1] + 25 \approx 34.8km/h$

+6 points
Fully correct

+2 points
Going from return period to non-exceedance probability

RT=50years→P[W>x]=1/50=0.02→P[W≤x]=1−0.02=0.98

+2 points
Identifying the right equation: the CDF of the GEV

+2 points
Using the inverse of the GEV and compute the design wind speed

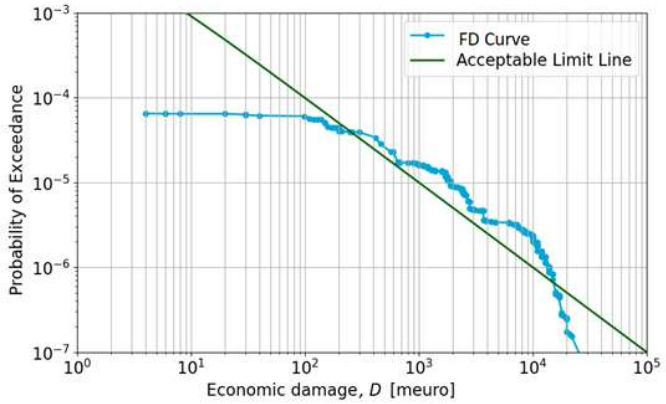
-1 points
Error when applying the GEV (using an exceedance probability or the return period instead of a non-exceedance probability.

-0.5 points
small calculation error

9 Risk & Reliability

6.0 points · 2 questions

When answering the following 2 questions , consider the figure here, which illustrates a risk analysis for a flood protection area.



Text

a Which of the following would be appropriate ways to *meet* the safety requirement? (more than one can be selected)

3.0 points · Multiple choice · 4 alternatives

- [x] Remove people from the area.
- [x] Make the flood protection stronger
- [ ] Convince policy makers to allow for a higher chance of fatalities
- [ ] None of the above

Feedback

Feedback when the question is answered correctly

Feedback when the question is answered partially correctly

Feedback when the question is answered incorrectly

b Identify the false statement in the list below (there is only one). The statement should complete the sentence: *The FD curve for the flood protection system...*

3.0 points · Multiple choice · 5 alternatives

- ...quantifies probability of system performance
- **...quantifies probability of component failure in the system**
- ...illustrates consequences of system failure
- ...can be used to compute expected value of consequences
- ...illustrates exceedance probability

Feedback

Feedback when the question is answered correctly

Feedback when the question is answered partially correctly

Feedback when the question is answered incorrectly

3332868 (6)_Page_24.png

Image 1 Page

Use the grid below if you run out of space for any exercise.
In that case, please indicate so at the original answer field.