Surname, First name

1	1	1	1	1	1	1
2	2	2	2	2	2	2
3	3	3	3	3	3	3
4	4	4	4	4	4	4
5	5	5	5	5	5	5
6	6	6	6	6	6	6
7	7	7	7	7	7	7
8	8	8	8	8	8	8
9	9	9	9	9	9	9
0	0	0	0	0	0	0

Modelling, Uncertainty and Data for Engineers EXAM (CEGM1000)
Resit Q1
31 Jan, 2024, 13:30-16:30

Do not open the exam or turn to the back page until given permission by the instructor!
(you may write your name and student ID)

Before you start the exam, a few remarks:

- Write down your first and last name in the field on the top left corner of this page
- Fill in your student number on the top right corner of this pages. Fill in the number in the boxes on top, and mark the corresponding number. Fill the corresponding circle as indicated above.
- You may use pen or pencil, a scientific (non-graphing) calculator and the attached formula sheet. Any other tools and sources of information are not allowed.
- Exam should remain stapled.
- Scratch paper is available to use during the exam, but will not be collected or graded.
- On the following pages, some questions have a specific box for you to answer: anything written outside the boxes will not be graded. Note that we have provided a lot of space for answers.
- In case you want to erase and rewrite your answer, ask an invigilator for a white sticker to cover your incorrect answer.
- The answer space size is not an indicator of how long we expect your answers to be! Shorter is generally better.
- In case you want to correct your answer for a multiple choice question, follow the instructions above. If you mess up, put an arrow to the answer you think is correct. If you mess up again, add a comment.
- A summary of points and questions is provided on the last page, as well as examples of how to correct your multiple choice answers.

Good luck!

Exercise 1: Programming

2p **1a** You are discretizing a single-valued differential equation $f(x,t)$ in time and are pre-allocating a list or array to store the results at each time step. Which of the following commands is appropriate?

(a) np.linspace(0, time_max, n_point)
(b) np.discretize(0, time_max, n_point)
(c) range(0, time_max)
(d) range(0, time_max, n_point)

2p **1b** Which of the following methods is the best method to provide information on the rows and columns of a Python object?

(a) sort()
(b) shape()
(c) size()
(d) dimensions()
(e) rows_and_columns()

4p **1c** Sketch the file structure of a typical weekly MUDE repository that includes *at least* three key file types or directories. Include a short explanation and description of each item listed. Your answers should demonstrate how files in the MUDE repository were generally organized and used.

Note: if you are from the 2022-23 (MUDE year 1) you do not need to answer this question (write your year in the space provided).

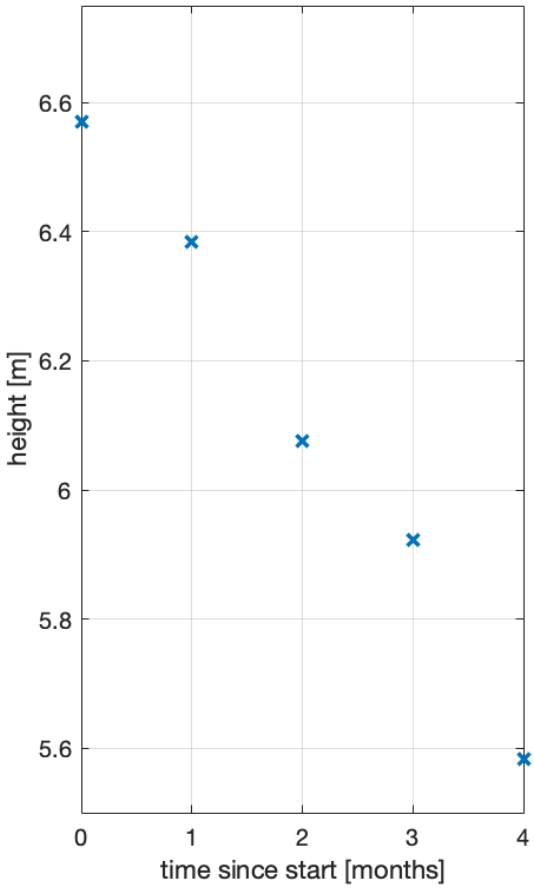
Exercise 2: Estimation and uncertainty propagation

The height of a point is observed 5 times at a monthly interval, all observables are independent; see figure. It is assumed that the point is subsiding at a constant rate v.

5p **2a** In the figure below, sketch the estimated linear trend line **and** confidence intervals assuming:

- the initial height x_0 at $t_1 = 0$ is **unknown** and estimated with best linear unbiased estimation; and
- the subsidence rate is **known** to be $v = -0.2$ m/month

Note that no calculations are required for this question. You may assume an arbitrary confidence level, and importance should be placed on the shape and relation to trend line.



Assume that again the initial height is unknown, but now the subsidence rate is estimated with an independent set of observations, the precision of the estimated rate is $\sigma_{\hat{V}} = 0.1$ m/month. The 'corrected' observables are thus given by:

$$Y_i = Y_{i,\mathrm{o}} - \hat{V} \cdot t_i.$$

The $Y_{i,\mathrm{o}}$ are independent and have a precision of 0.1 m.

5p **2b** Specify the stochastic model with $Y = [Y_1, \ldots, Y_5]^T$ as the observables.

Hint: first write the new observable vector as:

$$Y = \mathrm{T} \begin{bmatrix} Y_{1,\mathrm{o}} \\ Y_{2,\mathrm{o}} \\ \vdots \\ Y_{5,\mathrm{o}} \\ \hat{V} \end{bmatrix}$$

where you have to define T.

Exercise 3: Observation Theory

The concentration of particulate matter (PM10) is observed 6 times in a row: 3 times with device A and 3 times with device B. The observations are all assumed to have the same standard deviation of 1 μg/m^3 and to be independent. Furthermore, it is assumed that the actual (true) concentration of PM10 did not change during the time that the 6 observations were made.

The observations from device A are: 26, 27 and 23 μg/m^3, respectively.
The observations from device B are: 30, 29 and 27 μg/m^3, respectively.

It needs to be tested whether or not there is a constant offset between the observations of the 2 instruments. The null-hypothesis $\mathcal{H}_0$ is that there is **no** offset and the PM10 is constant during the time of the observations. The alternative hypothesis $\mathcal{H}_a$ reads that PM10 is constant during the time of the observations but that there is a constant offset ∇ between the observations of both devices.

6p **3a** Specify the functional and stochastic model under the two hypotheses.

6p **3b** Apply the appropriate test to check the validity of $\mathcal{H}_0$ with false alarm probability $\alpha = 0.005$. Explain whether or not $\mathcal{H}_0$ is accepted?

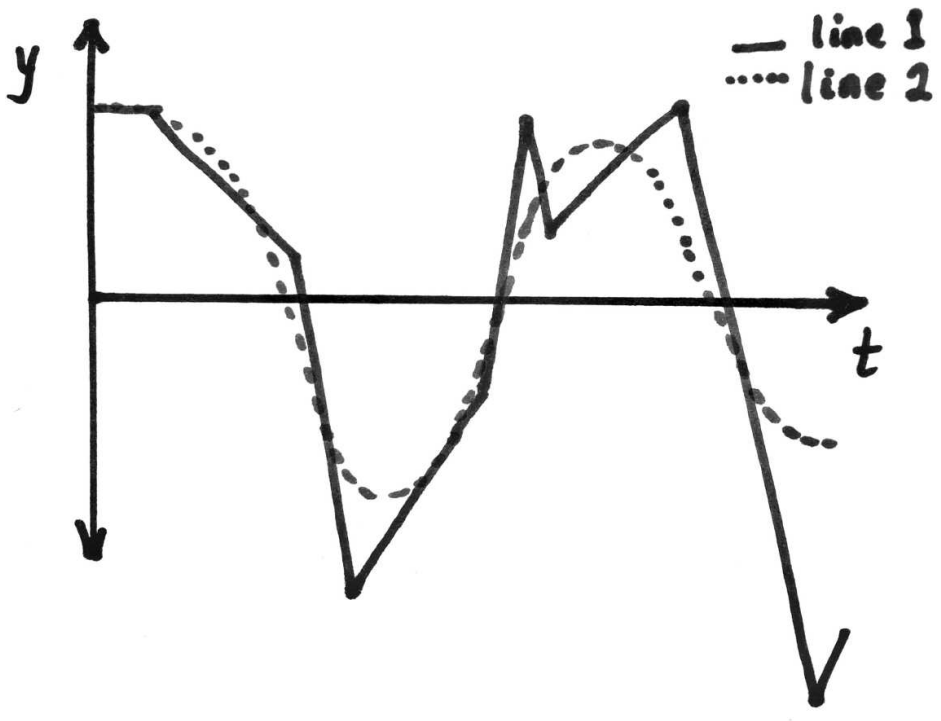
6p **3c** Assume now that we used a too optimistic standard deviation, both devices in fact have a precision of 3 μg/m^3. Show / explain how this affects the decision in the previous question.

Exercise 4: Numerical modelling

12p **4a** For an arbitrary function $f(x)$ that is continuously differentiable, derive the Forward Euler approximation, beginning with a Taylor series expansion. Express your answer in a series of steps with a short description (1 sentence max) of each step to go along with the derivation. Include a sketch that illustrates how the Forward Euler expression approximates the arbitrary function $f(x)$; label the figure appropriately.

The next questions are completely unrelated to the previous question.

A colleague has mentioned that they are working on processing data from a sensor that appears to behave like a damped harmonic oscillator. They decided to model the sensor output using an explicit method. They mention that there is some instability, but don't know how to fix it. Consider the sketch provided by your colleague and then answer the following questions. The figure shows a numerical analysis ("Line 1") compared with the sensor output $y(t)$ ("Line 2").



Your colleagues sketch.

6p **4b** Describe what you would suggest to your colleague to help with their model, and justify your answer briefly (2 sentences max.).

3p **4c** Which of the following should you evaluate to gain additional understanding about stability of the numerical approach.

(a) Initial conditions of the system
(b) Eigenvalue of the system
(c) Time-step size
(d) Order of the polynomial being modelled

4p **4d** Recalling that the general expression for a damped oscillator is of the form

$$m\frac{d^2y(t)}{dt^2} + c\frac{dy(t)}{dt} + ky(t) = 0$$

which method(s) might help to numerically model the signal? *(you may select more than one, or none)*

- ☐ Implicit (backward) Euler method
- ☐ Riemann method
- ☐ Runge-Kutta method
- ☐ Trapezoidal rule
- ☐ Gauss integration method
- ☐ Simpson's rule

Exercise 5: Probability & Reliability

A climate scientist is performing a study about rainfall and the ice content in the clouds. X is the daily rainfall in a city in mm. After several years of measurements, it has been discovered that the cumulative distribution function of X is given by an exponential distribution with parameter $\lambda = 0.05$.

2p **5a** What is $P[X \leq 20$ mm$]$? Round to two decimal figures.

2p **5b** The researcher wants to design for the value of X which is exceeded with a probability of 0.01. What is the design value? Round to two decimal figures.

The researcher also wants to investigate the content of ice in the clouds (kg/m^2), denoted here by Y. After a field campaign, the following statistics of Y are calculated.

Mean	55.5
Standard deviation	11.2
Minimum value	3.1
P25%	50.7
P50%	57.8
P75%	63.2
Maximum value	76.1

1p **5c** Which of the following distributions would be the best fit to the variable Y based on the previous statistics:

(a) Gaussian
(b) Left-tailed Gumbel
(c) Uniform

3p **5d** Justify your choice with at least one reason that justifies your decision.

Now the researcher wants to dive into the relationship between X and Y.

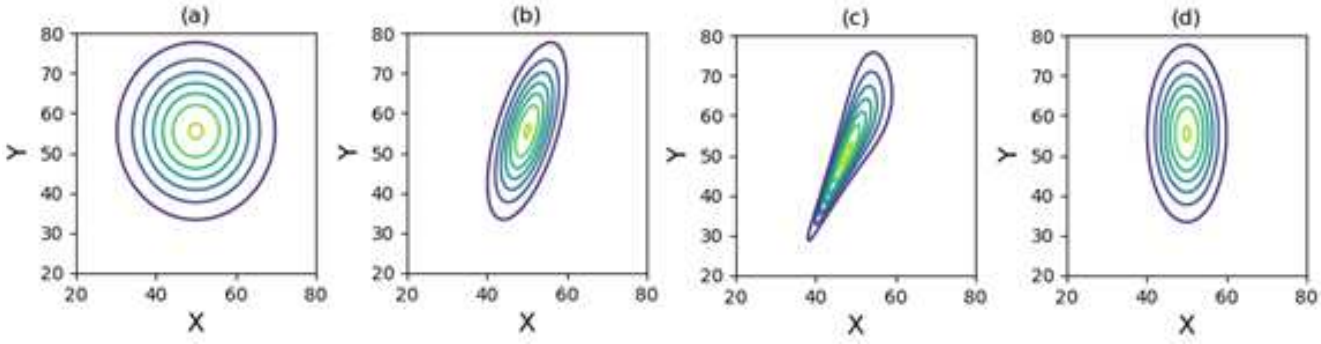
4p **5e** A high positive correlation between both random variables is observed. Provide an explanation of what that means in 1 or 2 sentences. Give a numeric range for positive correlation.

3p **5f** The researcher models the joint probability of X and Y using a bivariate Gaussian distribution. Which probability is equivalent to the multivariate Gaussian cumulative distribution function, $F_{X,Y}(X = x, Y = y)$?

(a) $P[X \geq x \text{ AND } Y \geq y]$
(b) $P[X \geq x \text{ OR } Y \geq y]$
(c) $P[X \leq x \text{ AND } Y \leq y]$
(d) $P[X \leq x \text{ OR } Y \leq y]$

The following plots show the possible contour plots of the bivariate PDF of X and Y.

Use them to answer the following two questions.



2p **5g** In which plots are X and Y independent? *(you may select more than one, or none)*

- ☐ Plot a
- ☐ Plot b
- ☐ Plot c
- ☐ Plot d

2p **5h** In which plots do X and Y follow a multivariate Gaussian distribution? *(you may select more than one, or none)*

- ☐ Plot a
- ☐ Plot b
- ☐ Plot c
- ☐ Plot d

This page left intentionally blank.

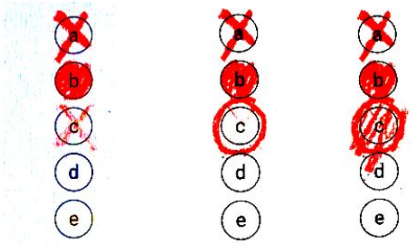
Exam Overview

No.	Topic	Number of sub-parts	Point
1	Programming	3	8
2	Uncertainty Propagation	2	10
3	Observation Theory	3	18
4	Numerical Methods	4	25
5	Probability and Reliability	8	19
Total			80

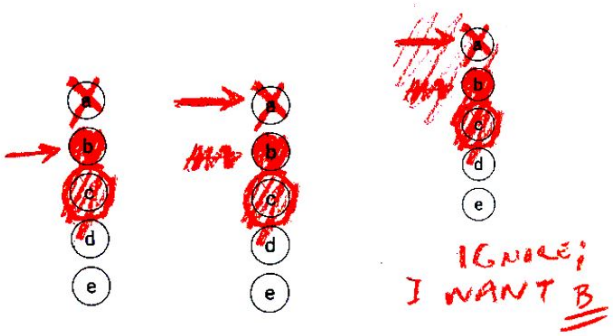
Use the table above to help plan your time during the exam.

In case you want to correct your answer for a multiple choice question put an ARROW in front of your final answer. If you also make a mistake with your arrow, write a clear message on the page. Here are a few examples:

Examples of UNCLEAR multiple choice response:



Examples of CLEAR multiple choice response:



Answer: B Answer: A Answer: B