Exam Q2 R

CEGM1000 Modelling, Uncertainty and Data for Engineers EXAM 24/25 · 9 opgaven · 80.0 punten

Finite Volume Method #47275467

12 ptn · Laatst bijgewerkt 9 apr, 20:20

Which of the following statements correctly distinguishes the Finite **Difference** Method (FDM) from the Finite **Volume** Method (FVM)?

3 ptn · Meerkeuzevraag · 5 alternatieven

- FDM is more suitable for solving conservation laws, whereas FVM is better for problems involving high-order derivatives.
- **FDM is based on approximating differential equations at discrete points, while FVM enforces conservation laws over control volumes.**
- In FDM, fluxes are explicitly conserved over a control volume, while in FVM, derivatives are computed using Taylor series expansions.
- FVM requires a structured grid, while FDM can be applied to both structured and unstructured grids.
- In FVM, integration over control volumes leads to the direct calculation of pointwise derivatives, similar to FDM.

Feedback

Feedback als de vraag juist is beantwoord

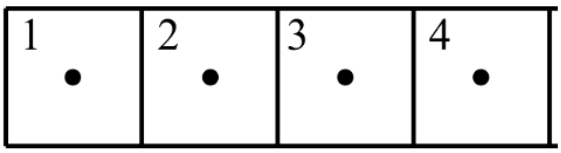
Feedback als de vraag gedeeltelijk juist is beantwoord

Feedback als de vraag onjuist is beantwoord

For a purely advective problem in 1D:

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

After schematizing the problem with the following mesh:



And applying the Finite Volume Method, we obtain:

$$\frac{\partial \phi}{\partial t} = \frac{c(\phi_e - \phi_w)}{\Delta x}$$

Tekst

What is the physical interpretation of the variable c, in one sentence?

2 ptn · Open · 3/10 Pagina

The velocity of propagation of information (phi) or how fast phi moves spatially	2 ptn
mentioning both components: - mention velocity/speed - mention what is being transported/propagated/moved	
mention only one of the two ideas but not both (either "velocity" or "transport of information" but not both)	1 pt
rate of change of phi. the student connects that c has something to do with changes of phi	0,5 ptn

The values of ϕ at the centre of the cell/volume are known at time $t = 0$ and vary across the mesh.

- Apply the **Forward Euler** method to discretize the time derivative.
- Approximate the surface values ϕ_e andϕ_w using **central differences**.

Write the discrete expression for ϕ_2^1, which represents the value at cell 2 after one time step.

(Here, the subscript refers to the cell number and the superscript to the time step.)

4 ptn · Open · 1/2 Pagina

full marks for both correct discretization in time and space: $\phi_2^1 = \phi_2^0 - \frac{c\Delta t}{2\Delta x}\left(\phi_3^0 - \phi_1^0\right)$	4 ptn
either forward Euler's correct $\frac{\phi_2^1 - \phi_2^0}{\Delta t}$ or central difference $-\frac{c}{2\Delta x}\left(\phi_3^0 - \phi_1^0\right)$, but not both	2 ptn
wrong sign	-1 ptn
mistake with index	-1 ptn

Consider a case where the mesh is very large (order of 1000000000 cells) and the propagation of the quantity ϕ is in the order of 10-100 cells. What will occur with this solution after many time steps? Limit your answer to 2-3 sentences.

3 ptn · Open · 2/5 Pagina

The solution becomes unstable due to Forward Euler and wiggles appear	3 ptn
Mentions instability, unstable behaviour, or solution blowing up	1 pt
Correctly links instability to method (FE)	1 pt
Describes observable behaviour (wiggles/oscillations)	1 pt

Finite Element Method #47275667

10 ptn · Laatst bijgewerkt 18 apr, 15:30

Consider the homogeneous steady-state advection/diffusion equation in 1D with given velocity $v\,(x)$ and diffusion coefficient k:

$$v\,\frac{\partial u}{\partial x} - \frac{\partial}{\partial x}\left(k\,\frac{\partial u}{\partial x}\right) = 0$$

To arrive at the weak form, integration by parts is not applied on the advection term. What is then the form of the stiffness matrix contributions...

Tekst

With v?

3 ptn · Meerkeuzevraag · 4 alternatieven

- ○ $\int_{\Omega} \mathbf{N}^T v \mathbf{N} d\Omega$
- ◉ $\int_{\Omega} \mathbf{N}^T v \mathbf{B} d\Omega$
- ○ $\int_{\Omega} \mathbf{B}^T v \mathbf{N} d\Omega$
- ○ $\int_{\Omega} \mathbf{B}^T v \mathbf{B} d\Omega$

Feedback

Feedback als de vraag juist is beantwoord

Feedback als de vraag gedeeltelijk juist is beantwoord

Feedback als de vraag onjuist is beantwoord

With k?

3 ptn · Meerkeuzevraag · 4 alternatieven

- ○ $\int_{\Omega} \mathbf{N}^T k \mathbf{N} d\Omega$
- ○ $\int_{\Omega} \mathbf{N}^T k \mathbf{B} d\Omega$
- ○ $\int_{\Omega} \mathbf{B}^T k \mathbf{N} d\Omega$
- ◉ $\int_{\Omega} \mathbf{B}^T k \mathbf{B} d\Omega$

Feedback

Feedback als de vraag juist is beantwoord

Feedback als de vraag gedeeltelijk juist is beantwoord

Feedback als de vraag onjuist is beantwoord

For a single 2-node element with the first node at $x = 0$ and the second node at $x = 10$, give a mathematical expression for $\mathbf{N}$.

2 ptn · Open · 3/10 Pagina

correct solution: $\mathbf{N}(x) = \left[1 - \frac{x}{10}, \quad \frac{x}{10}\right]$ or $\mathbf{N}(x) = \left[\frac{10 - x}{10}, \quad \frac{x - 0}{10}\right]$	2 ptn
The sign is incorrect in the shape function expressions.	-0,5 ptn
Correct vector form (correct shape of N vector) but wrong shape functions	0,5 ptn
numerical mistake	-0,5 ptn

For the same element, give a mathematical expression for $\mathbf{B}$.

2 ptn · Open · 3/10 Pagina

correct solution: $\mathbf{B}(x) = \left[-\frac{1}{10}, \quad \frac{1}{10}\right]$	2 ptn
the sign is incorrect	-0,5 ptn
correct vector notation and shape, but wrong derivatives	0,5 ptn
little mistake	-0,5 ptn

Signal Processing #41774687

5 ptn · Laatst bijgewerkt 9 apr, 20:41

We start from a signal which is the sum of two cosines, both with unit amplitude and zero phase, one with a frequency of 10 Hz, and the other with 60 Hz. The signal is sampled at $f_s = 100$ Hz, for a duration of $T = 2$ seconds.

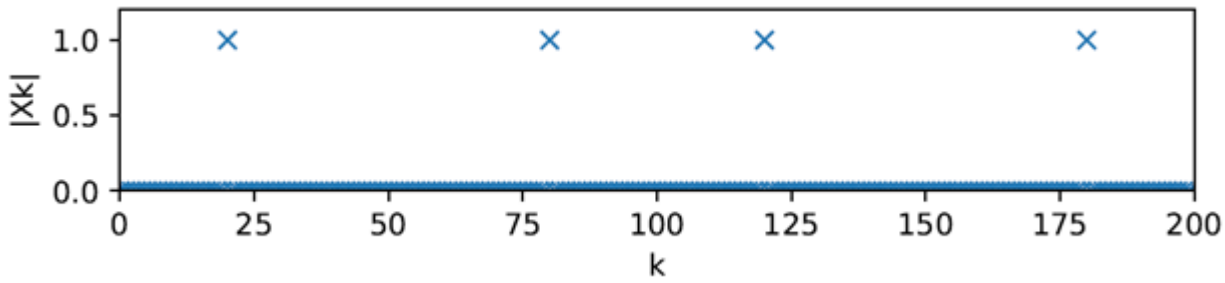
The discrete time samples are input to the Discrete Fourier Transform (DFT) and we directly plot the magnitude (modulus) of the output, hence $|Xk|$, multiplied by $\Delta t = 1/f_s$, with $k = 0, \ldots, N-1$. *(i.e. Xk=np.fft.fft(xt), and plt.plot(abs(Xk)/fs) in code),* using crosses as markers.

The resulting graph is shown below:

Tekst

Picture3.png

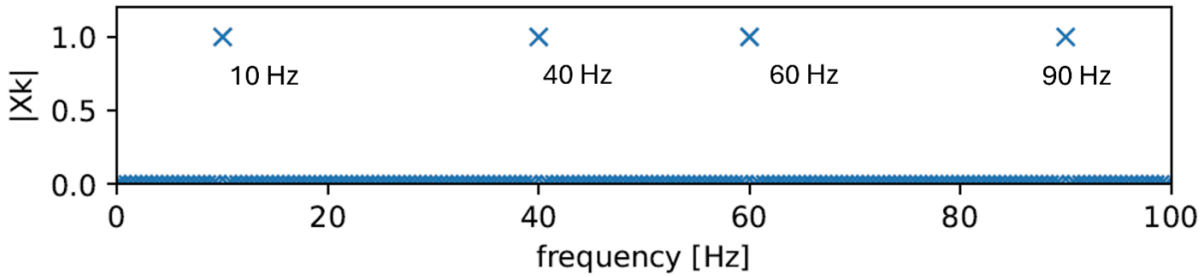
Afbeelding 1/10 Pagina



Provide an adequate and correct labelling of the horizontal axis, in terms of frequency expressed in [Hz], 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 ptn · Afbeelding · 2/5 Pagina



4 ptn

All labels correct, any of the following:
- 10 Hz, 40 Hz, 60 Hz, 90 Hz
- -40 Hz, -10 Hz, 10 Hz, 40 Hz
- -10 Hz, -40 Hz, -60 Hz, -90 Hz

4 ptn

Any of the following:
- 10 Hz, 60 Hz, 140 Hz, 190 Hz
- 10 Hz, 40 Hz, 60 Hz

3 ptn

only 10 Hz and 40 Hz labelled (at correct positions) 2,5 ptn

Any of the following:
- 10 Hz and 60 Hz correctly labelled
- 40 Hz and 60 Hz correctly labelled
- -60 Hz, -10 Hz, 10 Hz, 60 Hz

2 ptn

Any of the following:
- only 10 Hz (correctly) labelled
- only 60 Hz (correctly) labelled

1 pt

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

10 Hz, 40 Hz, -40 Hz, -10 Hz (provided axis is also correctly labelled [0,50)[-50,0)) 5 ptn

Time Series Analysis #47314205

12 ptn · Laatst bijgewerkt 17 apr, 10:49

The magnitude of the vertical stress in a rock mass has been observed at times $t = 0, 1, \ldots, 4$ months and is assumed to follow a linear trend; after detrending the data, the following time series is obtained:

$$S := \hat{\epsilon} = \mathrm{y} - \mathrm{A}\hat{\mathrm{x}} = \begin{bmatrix} 1.8 \\ 10.4 \\ -1.1 \\ 5.4 \\ -8.7 \end{bmatrix} MPa$$

The estimated linear trend parameters are $\hat{\mathrm{x}} = [\hat{x_0}, \hat{v}]^T = [3.0, 0.8]^T$ ($\hat{x_0}$ [MPa] is the estimated intercept, $\hat{v}$ the estimated slope [MPa/month]).

The stochastic process underlying time series S is modelled as a **zero-mean AR(1) process**.

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

Tekst

Time Series Analysis formulas:

AR(p) process:

$$S_t = \sum_{i=1}^{p} \phi_i S_{\{t-i\}} + e_t$$

Autocovariance function for AR(1) process:

$$c_{\{\tau\}} = \sigma^2 \phi_1^\tau$$

Normalized autocovariance function:

$$\rho_\tau = c_\tau / c_0$$

Least-squares estimate of autocovariance:

$$\hat{c}_\tau = \frac{1}{m-\tau} \sum_{t=1}^{m-\tau} (S_t - \mu)(S_{t+\tau} - \mu) \text{ with}$$

$$\mu = \mathbb{E}(S)$$

Prediction formula:

$$\hat{\mathrm{y}}_p = \mathrm{A}_p \hat{\mathrm{x}} + \hat{\epsilon}_p$$

Tekst

Estimate the normalized autocovariance function ρ_1 for the AR(1) process based on the stationary time series S.

4 ptn · Open · 4/5 Pagina

$C_0 = \frac{1.8^2 + 10.4^2 + 1.1^2 + 5.4^2 + 8.7^2}{5 - 0} = 43.49$ 1 pt

$C_1 = \frac{1.8 * 10.4 + 10.4 * -1.1 - 1.1 * 5.4 + 5.4 * -8.7}{5 - 1} = -11.41$ 2 ptn

$\rho_1 = \frac{C_1}{C_0} = -\frac{11.41}{43.49} = -0.26$ 1 pt

Make a plot (no calculation needed) of the ACF for the AR(1) process (if you did not find a solution for question a, you may use $\hat{\rho}_1 = -0.1$).

4 ptn · Open · 1/2 Pagina

for AR(1) we have that ρ1=ϕ1, the ACF at lag 0 is equal to 1, for lag 1 it is estimated to be −0.26, and for all other lags it is close to 0. 4 ptn

What is the predicted value of the vertical stress at $t = 5$ months?

4 ptn · Open · 9/20 Pagina

e5=ϕ1*e4=−0.26×−8.7=2.26 MPa 2 ptn

y5=Apx+ϵ5=3.0+0.8×5+2.26=9.26 MPa 2 ptn

Optimization #47312643

10 ptn · Laatst bijgewerkt 17 apr, 10:27

Solve the following mathematical programming problem using the graphical solution method. Do not forget to state clearly the solution and the value of the objective function in your answer, as well as how you obtain such a solution.

1. Minimize:

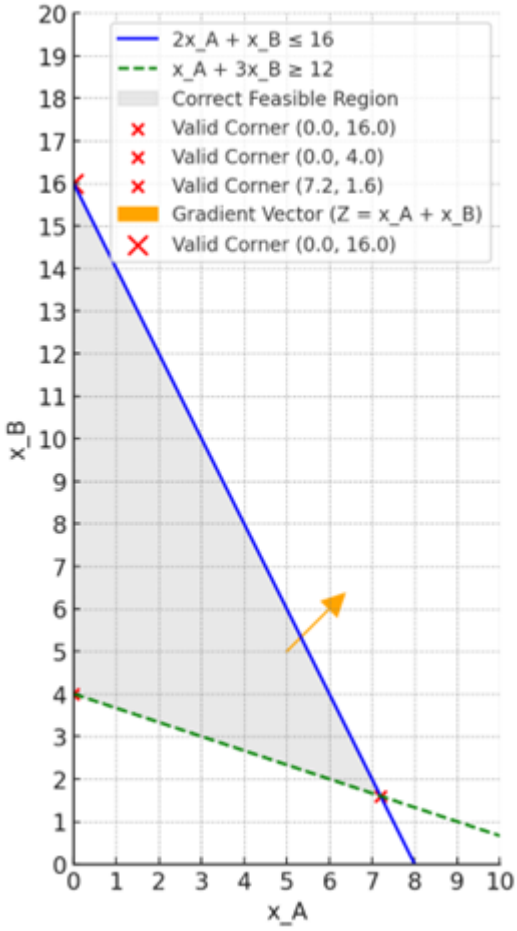
$$Z = x_A + x_B$$

2. Subject to:

$$2x_A + x_B \leq 16 \quad \text{(Constraint 1)}$$
$$x_A + 3x_B \geq 12 \quad \text{(Constraint 2)}$$
$$x_A, x_B \geq 0 \quad \text{(Non-negativity)}$$

7 ptn · Open · 13/20 Pagina

correct graph



3 ptn

The gradient vector is: [1,1]'

The point where the green constraint meets the xB axis is the point that minimizes the Objective function because it's the last point that one can find if going into the oppositive direction of the gradient (minimization). 2 ptn

solution (0,4) 1 pt

objective = 4 1 pt

Big mistake, such as interpreting constraint incorrect, but apart from that correct 3 ptn

If eventually you were told that the variables must be integer, what would the solution to the same problem be if the objective function became: $\max(Z) = x_A$?

Please mark this in the figure you made in **part a**. In addition to marking this on your figure, use the textbox below to explain your reasoning briefly.

3 ptn · Open · 2/5 Pagina

(7,2)	3 ptn
Correct answer using wrong region from A	2 ptn
Only considering xA	1 pt

Machine Learning #47155133

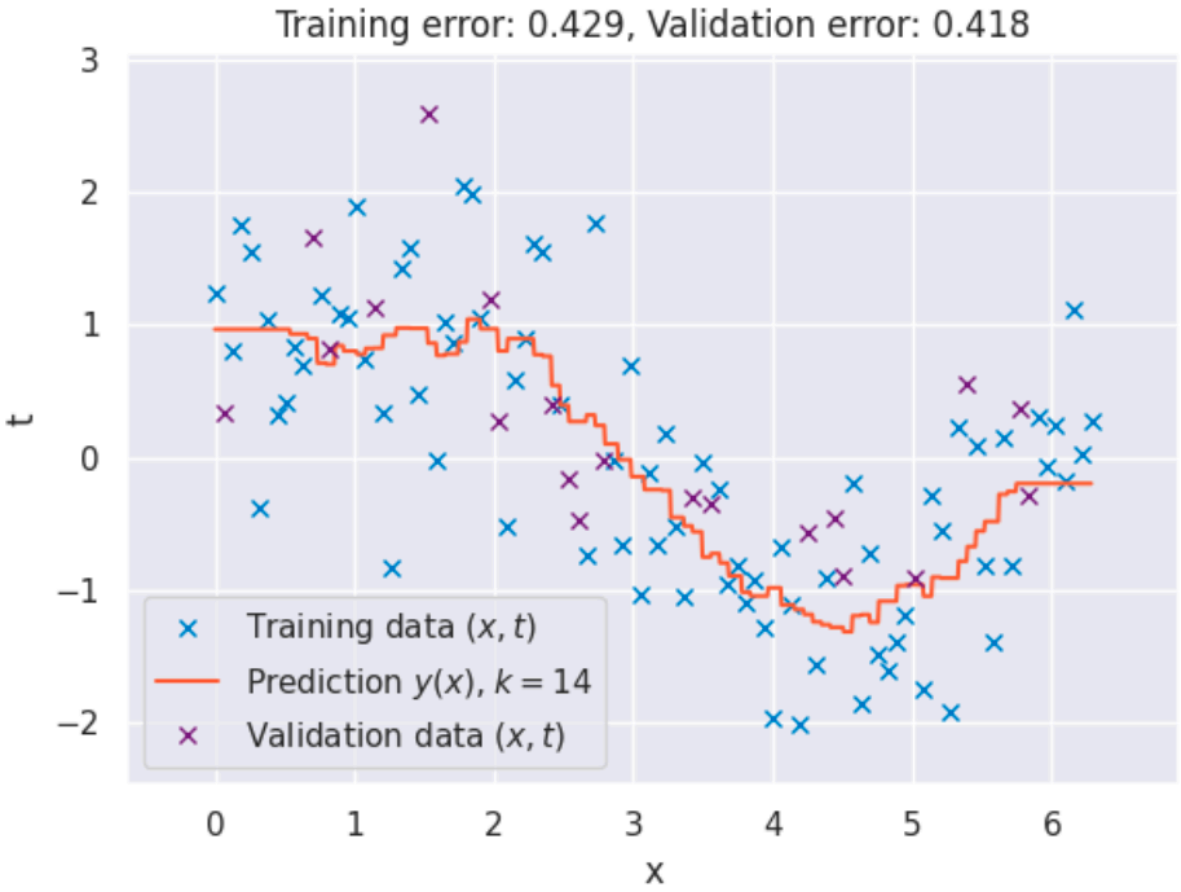
9 ptn · Laatst bijgewerkt 9 apr, 20:23

Consider the following kNN model trained on 80 noisy observations, with 20 observations left out for validation:

Tekst

Screenshot 2025-04-01 at 8.11.55 PM.png

Afbeelding 3/10 Pagina



Regarding decision theory and model selection for kNN models, which **ONE** of the following statements is true?

3 ptn · Meerkeuzevraag · 4 alternatieven

- Decision theory dictates that our regression function should be $y(x) = \int tp(t|x)dt$. In kNN we have direct access to $p(t|x)$ but cannot integrate it exactly, so we opt to define $y\,(x)$ as the average of all data points located at position x
- **For datasets containing only a single observation at each location $\boldsymbol{x}$, kNN shows extreme overfitting for $\boldsymbol{k = 1}$, with the training loss being exactly zero. By computing the loss on a validation set, we can have a more realistic idea about the performance of the model and adjust k accordingly**
- In order to avoid overfitting when using kNN models, we use only the training set to obtain k and average over only validation samples when computing $y\,(x)$
- In kNN, k can be seen as a hyperparameter, since it controls model complexity: as we increase k we decrease model complexity, as the neighbourhood we use to compute $y\,(x)$ shrinks in size and, therefore, contains less and less observation noise.

Feedback

Feedback als de vraag juist is beantwoord

Feedback als de vraag gedeeltelijk juist is beantwoord

Feedback als de vraag onjuist is beantwoord

Regarding the specific kNN model above, its losses $L_{\text{train}} = 0.429$ and $L_{vali} = 0.418$ and considering that $k = 14$ **leads to the best possible model for this dataset**, which **ONE** of the following statements is true

3 ptn · Meerkeuzevraag · 4 alternatieven

- ○ Starting from $k = 14$, further increasing k will lead to models with higher variance (and therefore lower bias), leading to a reduction of L_{train}
- ○ Regardless of the value of k, neither L_{train} nor L_{vali} can be reduced below the irreducible loss $L_{\text{noise}} = \int \int \left(\mathbb{E}\left[t \,|x|\right] - t\right)^2 p\left(x, t\right) \mathrm{d}x dt$
- ◉ **Moving away from $k = 14$ in any direction should makeL_{vali} increase. For $k < 14$ we expect L_{train} to decrease, while for $k > 14$ we expect L_{train} to increase.**
- ○ From the situation above, we could further regularize the kNN model by adding an L_2 regularization term of the form $\frac{\lambda}{2} \mathbf{w}^T \mathbf{w}$ to L_{train}, where $\mathbf{w}$ are the weights of the kNN model.

Feedback

Feedback als de vraag juist is beantwoord

Feedback als de vraag gedeeltelijk juist is beantwoord

Feedback als de vraag onjuist is beantwoord

Imagine you are working on your thesis and come across a regression problem you would like to solve with a **feedforward neural network.** Regarding dataset preparation, training and model selection, which **ONE** of the following modeling steps should be **AVOIDED**?

3 ptn · Meerkeuzevraag · 6 alternatieven

- ○ Split the dataset into 70% for training, 20% for validation and 10% for testing.
- ○ Normalize only the training dataset first and then use the same resulting normalization coefficient for the other two datasets
- ○ Monitor the validation loss during training and stop when its value does not decrease for a number of epochs (*early stopping*)
- ○ Shuffle the full dataset before splitting, in order to remove unintended biases during data acquisition and cleaning
- ◉ **Pick a set of network architectures (layer sizes, number of neurons, activation functions), train them all separately and then pick the one with the lowest error on the test dataset**
- ○ Include an L_2 regularization term to the loss function in order to minimize the risk of overfitting and calibrate λ using the validation dataset

Feedback

Feedback als de vraag juist is beantwoord

Feedback als de vraag gedeeltelijk juist is beantwoord

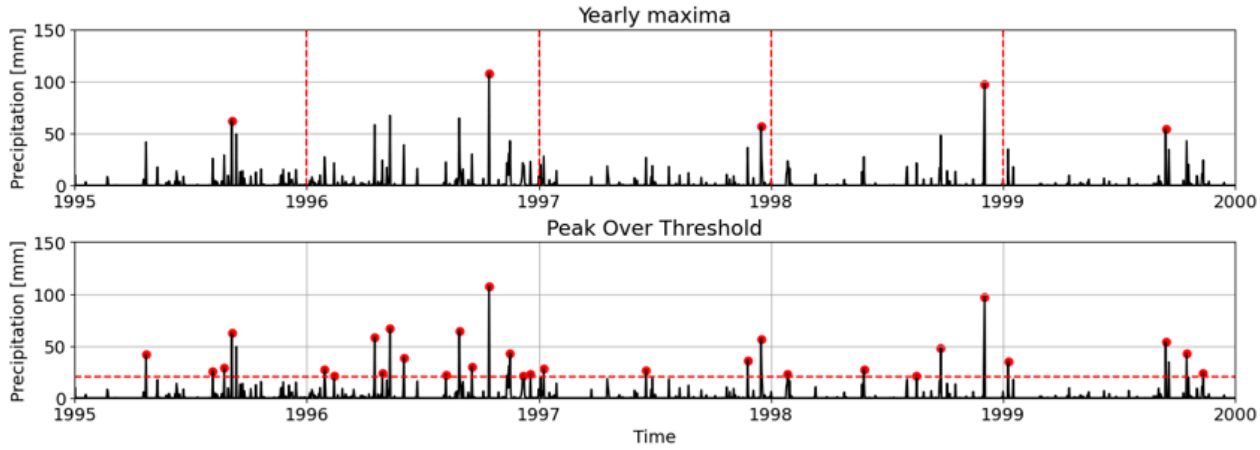
Feedback als de vraag onjuist is beantwoord

Extreme Value Analysis #47260769

12 ptn · Laatst bijgewerkt 10 apr, 14:01

Your team is investigating the efficiency and performance of drainage systems in cities along the South of Europe. Drainage systems collect rainfall water during extreme rainfall events to help control the flooding in the city.

One of your colleagues is analyzing the extreme precipitations used to design the drainage system in Barcelona (Spain). They obtained a timeseries of 5 years of precipitation and have performed both Yearly Maxima and Peak Over Threshold to select the extremes in the timeseries. They want your advice on which method to use for further analysis. Consider **ONLY** the two approaches shown in the figure below.



Tekst

Which option would you recommend to your colleague?

1 pt · Meerkeuzevraag · 2 alternatieven

- ○ Yearly Maxima
- ◉ Peak Over Threshold

Feedback

Yearly maxima is not an appropriate method in this case as the timeseries is very short and, thus, not enough extremes are selected (only 5). Peak Over Threshold draws much more extremes.

Feedback

Feedback als de vraag juist is beantwoord

Feedback als de vraag gedeeltelijk juist is beantwoord

Feedback als de vraag onjuist is beantwoord

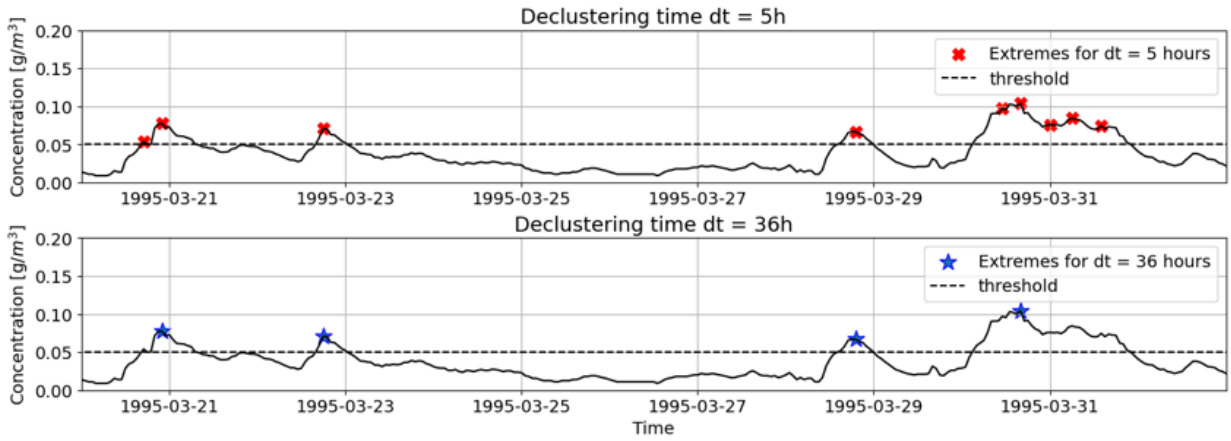
Justify your answer briefly.

2 ptn · Open · 3/10 Pagina

Yearly maxima is not an appropriate method in this case as the timeseries is **very short** and, thus, **not enough extremes are selected** (only 5). Peak Over Threshold draws much more extremes. 2 ptn

Justification is partially correct. -1 ptn

Another colleague is studying the water quality of the rainfall water that is received by the drainage system during extreme events. During extreme rainfall events, the water collected by the drainage system "cleans" the streets. They are working with a timeseries of the concentration of nitrogen in the water. They have applied Peak Over Threshold using a threshold value of $0.05g/m^3$. However, they are not sure about the declustering time (dt) they should use. In the figure below, you can see a small part of the timeseries and the extremes extracted using two declustering times, $dt = 5h$ and $dt = 36$.



Tekst

Which dt would you recommend to your colleague to use?

1 pt · Meerkeuzevraag · 2 alternatieven

- dt = 5h
- **dt = 36h**

Feedback

Using dt=5h you sample several extremes from the same extreme event (a series of observations above the threshold). With dt = 32h, you only sample one value per event, so you sample extremes that are independent, which is one of the assumptions of extreme value analysis.

Feedback

Feedback als de vraag juist is beantwoord

Feedback als de vraag gedeeltelijk juist is beantwoord

Feedback als de vraag onjuist is beantwoord

Justify your answer briefly.

2 ptn · Open · 3/10 Pagina

Using dt=5h you sample several extremes from the same extreme event (a series of observations above the threshold). With dt = 32h, you only sample one value per event, so you sample extremes that are **independent,** which is one of the assumptions of extreme value analysis.	2 ptn
Partially correct justification	-1 ptn
	0 ptn

A third colleague asked you for your help as probability expert. They are studying the extreme discharges in a drainage system. They have sampled 60 extremes using Peak Over Threshold with a threshold 50 l/s on 6 years of observations of discharges (l/s) and have fitted a Generalized Pareto distribution (GPD) so $Q \sim GPD\left(\sigma_{\text{th}} = 20, \xi = 0.25\right)$.

Compute the discharge associated with a return period of 10 years. The GPD probability density function is given by

$$h(x) = \begin{cases} \left(1 + \frac{\xi x}{\sigma_{th}}\right)^{-\frac{\xi+1}{\xi}} & \text{for } \xi \neq 0 \\ \exp\left(-\frac{x}{\sigma_{th}}\right) & \text{for } \xi = 0 \end{cases}$$

And the GPD cumulative distribution function is given by

$$H(x) = \begin{cases} 1 - \left(1 + \frac{\xi x}{\sigma_{th}}\right)^{-\frac{1}{\xi}} & \text{for } \xi \neq 0 \\ 1 - \exp\left(-\frac{x}{\sigma_{th}}\right) & \text{for } \xi = 0 \end{cases}$$

6 ptn · Open · 3/5 Pagina

Computes the probability of failure $RT = 10\,years \rightarrow P\left[Q > x\right] = \frac{1}{10} = 0.1$	1 pt
Compute the non-exceedance probability: $P\left[Q < x\right]_{event-wise} = 1 - 0.01 = 0.99$	1 pt
Compute the event-wise probability of failure: $P\left[Q > x\right]_{event-wise} = 0.1\,\frac{6}{60} = 0.01$	2 ptn
Evaluate the GPD distribution for a shape parameter different to 0 to compute the excess: $H\left(x\right) = 1 - (\,1 + \xi\,\frac{x}{\sigma_{\text{th}}})^{-\frac{1}{\xi}}$ $(\,1 + \xi\,\frac{x}{\sigma_{\text{th}}})^{-\frac{1}{\xi}} = 1 - H\left(x\right) 1 + \xi\,\frac{x}{\sigma_{\text{th}}} = (1 - H\left(x\right))^{-\xi} x = \left(\left(1 - H\left(x\right)\right)^{-\xi} - 1\right)\,\frac{\sigma_{\text{th}}}{\xi}$ $x = \left(\left(1 - 0.99\right)^{-0.25} - 1\right)\right)\,\frac{20}{0.25} \approx 173\;\frac{l}{s}$	2 ptn
Add the threshold to obtain the value of the random variable: $Q = 173 + 50 = 223\;\frac{l}{s}$	
Minor mistake in any of the 4 steps outlined above	-1 ptn
Forgets to add the th value to final answer	-0,5 ptn

Risk and Reliability #47398000

10 ptn · Laatst bijgewerkt 19 apr, 17:00

Your are asked to perform a risk analysis of a chemical spill into a groundwater aquifer which can occur in two ways:

A) 'A small spill'
B) 'A big spill'

Spill A has a 10% probability of occuring each year, which would cause 10 fatalities.
Spill B has a 1% chance of ocurring, which would cause 100 fatalities.

Assume that the regulating authority is risk averse ($\alpha = 2$) and that the tolerable risk limit is governed using a constant of $C = 1$.
Recall that the risk limit line can formulated as $1 - F_N(n) = C^\alpha$

Tekst

Compute the expected fatalities caused by the system

2 ptn · Open · 3/10 Pagina

$$E(N) = \sum P_i \cdot N_i = P_1 \cdot N_1 + P_2 \cdot N_2 = 0.1 \cdot 10 + 0.01 \cdot 100 = 2$$

2 ptn

If the chemical industry improved the safety of its systems, what would be the **maximum allowable number of fatalities** that would be considered acceptable due to a small spill?

2 ptn · Open · 3/10 Pagina

Using the limit curve$1 - F_N(N) = \frac{C}{n^\alpha}$
$0.1 = \frac{1}{n^2}$
$n = \sqrt{10} = 3.16 \rightarrow n = \lfloor \sqrt{10} \rfloor = 3$

2 ptn

What is the **maximum value for the probability** of a big spill that would be considered acceptable by the regulating authority?

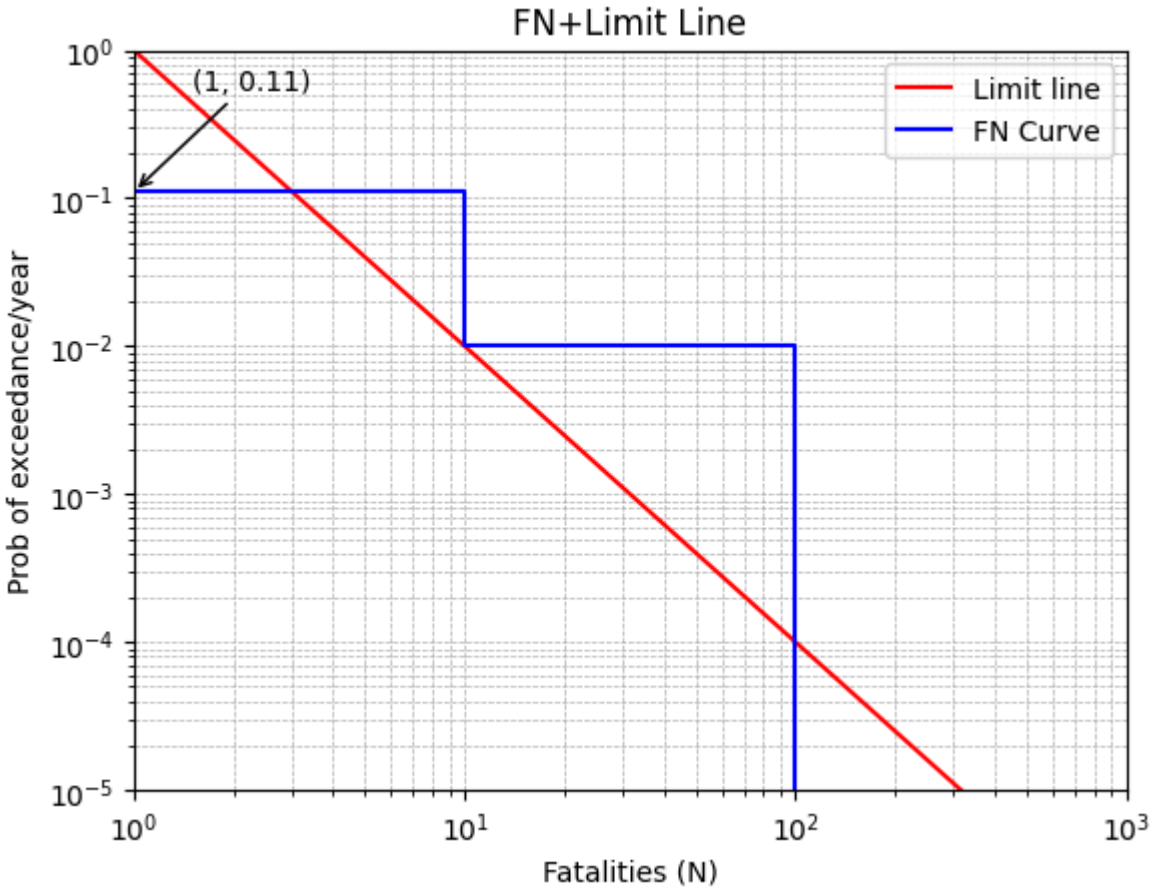
2 ptn · Open · 3/10 Pagina

$$1 - F_N(100) = \frac{C}{(100)^2} \rightarrow 10^{-4}$$

2 ptn

Create an FN curve for this situation, including the limit line. Be sure to clearly label the axes and the values of key points on your diagram.

4 ptn · Open · 3/5 Pagina



4 ptn

No/wrong axes	-0,5 ptn
No/wrong limit line	-1 ptn
Minor mistake	-0,5 ptn
Wrong values	-1 ptn