Werkplan Masterproef

Titel	Detection of plasma instabilities in tokamaks with machine learning
Naam Student	<name student 1>, <name student 2>
Email	<email student 1>, <email student 2>
Bedrijf/ Onderzoeksgroep	Research Unit Nuclear Fusion (infusion)
Promotoren	<name promotor>
Begeleiders	<name begeleider>

Opsplitsing per semester:

<name student 1>

	Semester 1	**Semester 2**
# studiepunten vakken	21	18
# studiepunten masterproef	6	12

<name student 2>

	Semester 1	**Semester 2**
# studiepunten vakken	24	21
# studiepunten masterproef	6	12

Bestaande situatie en probleemstelling

Nuclear fusion is an alternative to nuclear fission to generate massive amounts of energy to power our society. Unlike fission, fusion does not produce a lot of dangerous waste products and there is no way for the fuel, Hydrogen, to be converted into weapons. There is also no possibility of a runaway reactor meltdown since the reactor will operate in pulses.

To get a fusion reaction, Hydrogen gas must be heated to extreme temperatures under high pressure, becoming a plasma. This state of matter allows the hot atoms to be contained by powerful magnets, protecting the donut shaped vessel called the “tokamak.” In the 80s a new mode of operation in tokamaks, H-mode, was found, with doubled performance and better confinement. However, as a side effect, it created periodic releases of energy (like how a pressure cooker's valve releases steam every now and then). The risk is the wear and tear to the reactor walls, as materials are repeatedly exposed to short bursts of heat.

This phenomenon is called Edge-Localised Modes (ELMs), in relation to the source of these instabilities: turbulence occurring near the edge of the plasma, right by the vessel walls. And like turbulence, ELMs are difficult to model with fundamental physics principles. Therefore, studies rely on empirical data from large international experiments, such as JET (the Joint European Torus).

Doelstelling van de masterproef

The Infusion group at Ghent University is specialized in data science and contributes to fusion research and Eurofusion projects. While studies of these ELMs typically compare a few experiments at a time, the group is able to aggregate all the available data, by characterising each experiment with statistical features extracted from the ELM signals.

A key step is therefore detecting when an ELM has occurred. This relies on detecting light emission which follows the release of energy from the plasma. Simply put, whenever there is a peak in this light emissivity diagnostic, there should be an ELM. Previously, the group developed a simple but robust algorithm to set an optimal threshold and filter out noise from the emissivity signal. Their method was successfully applied to data from a variety of fusion machines, and several papers were published using the results.

One recent area of interest is studying plasma configurations that create smaller, more manageable ELMs. The emissivity patterns which they create however, is not as clear cut, and so the detection algorithm needs to be improved. Which is the goal of this thesis: to evaluate the performance of the existing algorithm, and to propose improvements using state of the art tools, such as machine learning and artificial neural networks.

Besides improving existing ELM detection methods we also propose some novel approaches. The first is a convolution-based detection method based on the blob detection used in the SIFT computer vision algorithm. Because SIFT is invariant to changes in illumination we hope this will translate well to the different amplitudes encountered in unstable ELM signals.

The second method is a rolling Z-score threshold. This method is promising because of its simplicity. Not a lot of prior knowledge about statistics is needed to understand the rolling Z-score. Which can improve the chance of adoption in the near future. The rolling Z-score threshold is a statistical measurement and indicates how far a datapoint is from the rest of the dataset.

Important note

The data that will be used during the thesis to train and evaluate ELM detection methods originated from recent experiments on the JET tokamak in the UK. Because of this the data is very sensitive and expensive. **The researchers at JET require us to wait a minimum of two years before the thesis can be published**.

Planning en mijlpalen

Task categories:
A : Data exploration and tools
B : Anomaly detection models
C : Validation methods
D : Writing and presentation

Task 1	1 week	*Deadline:* 3 October 2022	**Data exploration and tools: Develop early version of first technique to evaluate core concepts**
Responsible: <name student 1>			
Content Developing and implementing an initial attempt at peak detection to check viability of core concepts of our first method. (Peak detection based on convoluting with a Laplacian of gaussian kernel similar to the blob detection in SIFT)			
Important results, deliverables or insights after this phase: Presenting technique to supervisor in first meeting. Further evaluation will follow after the technique is finished during the thesis.			

Task 2	2 weeks	*Deadline:* 21 October 2022	**Data exploration and tools: explain project structure, discuss goals and 'werkplan'**
Responsible: <name student 1>, <name student 2>			
Content Setting up communication channels, discussing existing GIT-repo structure, explaining how fusion data can be accessed and making the 'werkplan'.			
Important results, deliverables or insights after this phase: Making sure all team members can communicate using Slack and have a technical understanding of the in-house data libraries. The 'werkplan' should be done.			

Task 3	3 weken	*Deadline:* 11 November 2022	**Data exploration and tools: EDA of database**
Responsible: <name student 1>, <name student 2>			
Content Data exploration of the ELM database. Exploring existing tooling to extract ELM time series from databases. Looking at what features are available for different research labs.			
Important results, deliverables or insights after this phase: Writing a report for the supervisor			

Task 4	2 weeks	*Deadline:* 25 November 2022	**Validation methods: Prepare data**
Responsible: <name student 1>, <name student 2>			
Content Prepare the data and mark peaks so our tools can compare the results.			
Important results, deliverables or insights after this phase: The data should be fully prepared to be used for training and comparing the different methods.			

Task 5	2 weeks	*Deadline:* 9 December 2022	**Validation methods: Developing evaluation methods**
Responsible: <name student 1>, <name student 2>			
Content Developing methods to compare our peak detection methods against existing fusion peak detection methods. (False positive, accuracy, retention, ROC-curves, ...)			
Important results, deliverables or insights after this phase: Presenting code to supervisor and writing a report on how our methods can be compared to existing fusion techniques. The evaluation methods can still be extended if needed for certain methods.			

Task 6	2 weeks	*Deadline:* 19-23 December 2022	**Writing and presentation: Presentation**
Responsible: <name student 1>, <name student 2>			
Content Presentation: Making and practising the presentation			
Important results, deliverables or insights after this phase: Presentation to the promotor and supervisor			

Task 7	3 weeks	*Deadline:* 3 March 2023	**Validation methods: Implement existing methods**
Responsible: <name student 2>			
Content Implement existing methods that can be used to validate our results. Some interesting techniques are: - OMFIT - CIEMAT SVM - gSPRT Note that these tools are possibly not available for us to compare. We can try to get permission from the authors. Otherwise, we have to find other methods to implement.			
Important results, deliverables or insights after this phase: At least one of the existing methods should be implemented. Writing a report on these techniques.			

Task 8	3 weeks	*Deadline:* 3 March 2023	**Anomaly detection models: First method**
Responsible: <name student 1>			
Content Implementing first technique: Laplacian of gaussian kernel convolution based on SIFT			
Important results, deliverables or insights after this phase: Writing a report on the first technique. Give a detailed comparison with the current methods.			

Task 9	3 weeks	*Deadline:* 24 March 2023	**Anomaly detection models: Second method**
Responsible: <name student 2>			
Content Implementing second technique: Basic rolling Z-score threshold			
Important results, deliverables or insights after this phase: Writing a report on the second technique. Give a detailed comparison with the current methods.			

Task 10	3 weeks	*Deadline:* 24 March 2023	**Anomaly detection models: Third method**
Responsible: <name student 1>			
Content Implementing third technique: R-Peak-Detection-1D-CNN			
Important results, deliverables or insights after this phase: Writing a report on the third technique. Give a detailed comparison with the current methods.			

Task 11	1 week	*Deadline:* 31 march 2023	**Writing and presentation: Write at least 25 pages**
Responsible: <name student 1>, <name student 2>			
Content Compile al the current data and reports in the scription.			
Important results, deliverables or insights after this phase: At this stage, the scription should be at least 25 pages long.			

Task 12	1 day	*Deadline:* 1 April 2023	**Submit administrative data on Plata**
Responsible: <name student 1>, <name student 2>			
Content Submit the remaining required data on Plato.			
Important results, deliverables or insights after this phase: Submit the remaining required data on Plato.			

Task 13	3 weeks	*Deadline:* 21 April 2023	**Anomaly detection models: Fourth method**
Responsible: <name student 2>			
Content Implementing fourth technique: Peak detection and baseline correction using a convolutional neural network			
Important results, deliverables or insights after this phase: Writing a report on the fourth technique. Give a detailed comparison with the current methods.			

Task 14	3 weeks	*Deadline:* 21 April 2023	**Anomaly detection models: Fifth method**
Responsible: <name student 1>			
Content Implementing fifth technique: AMPD: An Efficient Algorithm for Automatic Peak Detection in Noisy Periodic and Quasi-Periodic Signals			
Important results, deliverables or insights after this phase: Writing a report on the fifth technique. Give a detailed comparison with the current methods.			

Task 15 (Optional)	3 weeks	*Deadline:* 21 April 2023	**Validation methods: review of evaluation methods**
Responsible: <name student 1>, <name student 2>			
Content If there is any time left before this deadline: Take a deep review of our evaluation methods. Look what can be refined and update the reports based on the changes.			
Important results, deliverables or insights after this phase: Improve the evaluation methods based on the newly gained insights and update the reports.			

Task 16	5 weeks	*Deadline:* 25 may 2023	**Writing and presentation: 95% version of the scription**
Responsible: <name student 1>, <name student 2>			
Content Write at least 95% of the scription and present it to the promotor and mentor.			
Important results, deliverables or insights after this phase: The scription should be almost fully completed. After this point, only minor changes should be done.			

Task 17	1 week	*Deadline:* 8 June 2023	**Writing and presentation: Submit scription**
Responsible: <name student 1>, <name student 2>			
Content Submit the final scription.			
Important results, deliverables or insights after this phase: The final version of the scription is submitted on Plato and in paper to every required person.			

Task 18	2 weeks	*Deadline:* 28-30 June 2023	**Writing and presentation: Public defence**
Responsible: <name student 1>, <name student 2>			
Content Prepare and rehearse the presentation to defend the thesis.			
Important results, deliverables or insights after this phase: Present and defend the thesis.			

Contactmomenten

There is a weekly meeting with the supervisor on Tuesday at 13h30 at the infusion offices. Both students will also send a weekly progress report in the form of an email to the supervisor and the promotor on Friday. To track the work on the tasks a Jira server is used. A Slack server facilitates all other communication between the students and the supervisor.

Gantt chart:

	Task 1	Task 2	Task 3	Task 4	Task 5	Task 6	Task 7	Task 8	Task 9	Task 10
Week 1	X									
Week 2										
Week 3		X								
Week 4		X								
Week 5			X							
Week 6			X							
Week 7			X							
Week 8				X						
Week 9				X						
Week 10					X					
Week 11					X					
Week 12						X				
Week 13						X				
Kerstvakantie										
Examenperiode januari										
Week 1							X	X		
Week 2							X	X		
Week 3							X	X		
Week 4									X	X
Week 5									X	X
Week 6									X	X

	Task 11	Task 12	Task 13	Task 14	Task 15	Task 16	Task 17	Task 18
Week 7	X	X						
Paasvakantie Week 1			X	X	X			
Paasvakantie Week 2			X	X	X			
Week 8			X	X	X			
Week 9						X		
Week 10						X		
Week 11						X		
Week 12						X		
Week 13						X		
Week 14								
Week 15							X	
Week 16								
Week 17								X
Week 18								X