

# Eindrapport

March 30, 2025

## Preface

Presented before you is my bachelor thesis with the title: “The Impact of Climate Change on the Hydrology of the Wien River and the Implications for Flooding of the Adjacent U4 Subway Line”. During my Bachelor’s in Civil Engineering, I learned about a wide range of topics, but hydrology has been my main area of interest. With climate change effects being of growing significance, researching its impact on a river seemed like a meaningful and interesting choice to write my thesis about.

Writing this thesis has learned me a lot about using a hydrological model. It also expanded my programming skills and challenged me to critically assess data and results.

I would like to thank dr. ir. Rolf Hut and ir. Vincent Hooglander for their excellent guidance throughout my research. Your support with helping me narrow down my topic, structure my research and critically evaluate my results has been of great importance. I would like to thank ir. Mark Melotto as well for helping me with issues concerning the eWaterCycle platform. Additionally, I would like to thank Lars Kramer, for providing me with peer feedback.

My hope for this research is that it will contribute to a broader understanding of the impact of climate change on rivers. Furthermore, I aim to provide the city of Vienna with recommendations on whether or not their flood protection infrastructure for the U4 Subway Line will need improvements.

I hope you find this thesis an interesting read.

Thirza van Esch  
Delft, 30-3-2025

## Abstract

Climate change will cause precipitation to increase in the Northern Hemisphere, impacting river discharge and flood risk. This research assesses the impact of climate change on the hydrology of the Wien River and the implications for flooding of the adjacent U4 subway line. This is done with use of the HBV model forced with CMIP6 data.

First, the threshold for flooding of the U4 subway line is determined, after which observation data is used to determine the current return period of threshold exceedance. The HBV model is then calibrated and forced with CMIP6 data to model future discharges.

The discharge threshold before flooding is determined by looking at other studies to be  $530 \text{ m}^3/\text{s}$ . Looking at observation data, the current return period of threshold exceedance is 641 years. Under the SSP119 scenario, a sustainability scenario, the future return period of threshold exceedance is 697 years. The four other investigated scenarios result in future return periods ranging from 191 to 272 years.

This means that for four of the five SSP scenarios, extreme discharge events will become more frequent. Considering the current return period already being lower than the design return period of 1000 years, the return period for flooding of the U4 subway line will get even lower under four out of five climate change projections.

Due to the limited availability of observation data, a period of around 40 years, and only using one ensemble of the 25 ensemble members of the MPI model, uncertainties exist in return period estimation and model calibration. Future research should investigate all the 25 ensemble members to reduce uncertainties. A follow-up study could investigate the current flood protection infrastructure and find potential reinforcements.

# Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
1.1	Motivation . . . . .	1
1.2	Problem Analysis . . . . .	1
1.3	Objective . . . . .	1
1.4	Approach . . . . .	2
1.5	Reading Guide . . . . .	2
<b>2</b>	<b>Current maximum discharge before flooding of the U4</b>	<b>4</b>
2.1	Background . . . . .	4
2.2	Determination of flooding threshold . . . . .	4
2.2.1	Literature Study . . . . .	4
2.2.2	Conclusion . . . . .	5
<b>3</b>	<b>Current frequency of threshold exceedance</b>	<b>7</b>
3.1	Observations . . . . .	7
3.2	Extrapolation . . . . .	7
3.3	Conclusion . . . . .	9
<b>4</b>	<b>Calibration of the HBV model</b>	<b>10</b>
4.1	Model Selection . . . . .	10
4.2	Calibration . . . . .	11
4.2.1	Root Mean Squared Error . . . . .	11
4.2.2	Nelder-Mead Optimization . . . . .	12
4.2.3	Calibration Results . . . . .	12
4.3	Conclusion . . . . .	15
<b>5</b>	<b>Future frequency of threshold exceedance under different climate change scenarios</b>	<b>16</b>
5.1	Historical Comparison . . . . .	16
5.2	Future Scenarios . . . . .	17
5.3	Conclusion . . . . .	19
<b>6</b>	<b>Discussion</b>	<b>20</b>
6.1	Key Findings . . . . .	20
6.1.1	Current Threshold for Flooding . . . . .	20
6.1.2	Current Frequency of Threshold Exceedance . . . . .	20
6.1.3	Future Frequency of Threshold Exceedance under Climate Change Scenarios	20
6.2	Interpretation of Results . . . . .	21
6.3	Implications of Findings . . . . .	21
6.4	Limitations . . . . .	21
6.5	Recommendations . . . . .	22
<b>7</b>	<b>Conclusion</b>	<b>23</b>
<b>8</b>	<b>References</b>	<b>24</b>
<b>9</b>	<b>Appendix I</b>	<b>26</b>



# 1 Introduction

## 1.1 Motivation

Last September, the Wien River, normally a small trickle, swelled into a raging torrent, having a serious impact on public transport. Damage to the public transport was kept to a minimum thanks to the flood alarm plan, but still almost one-million litres of water had to be pumped out of the subway lines (Krone.At, 2024b).

The Wien River is one of the largest rivers of Vienna and has a catchment area of 230 km<sup>2</sup>. Along a reach of over 8 km, the subway line U4 follows the river in an open section on the right bank. Flooding of the Wien River is critical due to its rapid onset and potential for severe damage. The large area of impervious surfaces in the catchment area make Wien River floodings typically flash floodings. The threat regarding flooding in Vienna is caused by large channel slopes and flow velocities, rapid increase of discharge and the absence of natural retention areas. Furthermore, the low hills and mountains in the area intensify storm events compared to plain areas through intensified convective air movement (Compton et al., 2009, pp. 13–14).

Dore (2005) states in his article about climate change and changes in global precipitation patterns that due to climate change, precipitation is expected to increase in the Northern Hemisphere. The wet areas will get wetter, and the rainfall will get more intense. Changes in precipitation due to climate change may affect the water level and discharge of the Wien River, posing potential risks to the operation and safety of the adjacent U4 subway line.

## 1.2 Problem Analysis

Flood risk management is critical for urban resilience, especially where natural waters are closely linked to infrastructure. Compton et al. (2009) highlight how an approach based on catastrophe modelling can provide a useful framework for comparing different mitigation strategies as well as integrate the risk perspectives of different technical disciplines. Their study uses hydrological models to obtain probability of failure for different storm return periods and states of the flood control reservoirs.

This report also examines flood risk but shifts focus to how climate change will alter discharge and water levels in the Wien River, affecting flood risk for the U4 subway line. A hydrological model forced with different climate change scenarios will be implemented to simulate the future discharge scenarios of the Wien River. The report will focus solely on changes in flood frequency due to climate change and excludes the current flood protection infrastructures and adaption measures these might require. This research will contribute to Vienna's flood risk management and is relevant for the transportation authorities, policymakers and urban developers in Vienna who are seeking to make the city future-proof for increasing threats due to climate change.

## 1.3 Objective

In this report the following research question will be answered: “What is the impact of predicted climate change on the hydrology of the Wien River and what are the implications for flooding of the adjacent U4 subway line?” This is done using the following sub-questions:

- What is the current maximum discharge (m<sup>3</sup>/s) in the Wien above which the U4 subway line will flood?
- How often does the discharge currently exceed the maximum?
- How often will the discharge exceed this maximum in the future under different climate change scenarios?

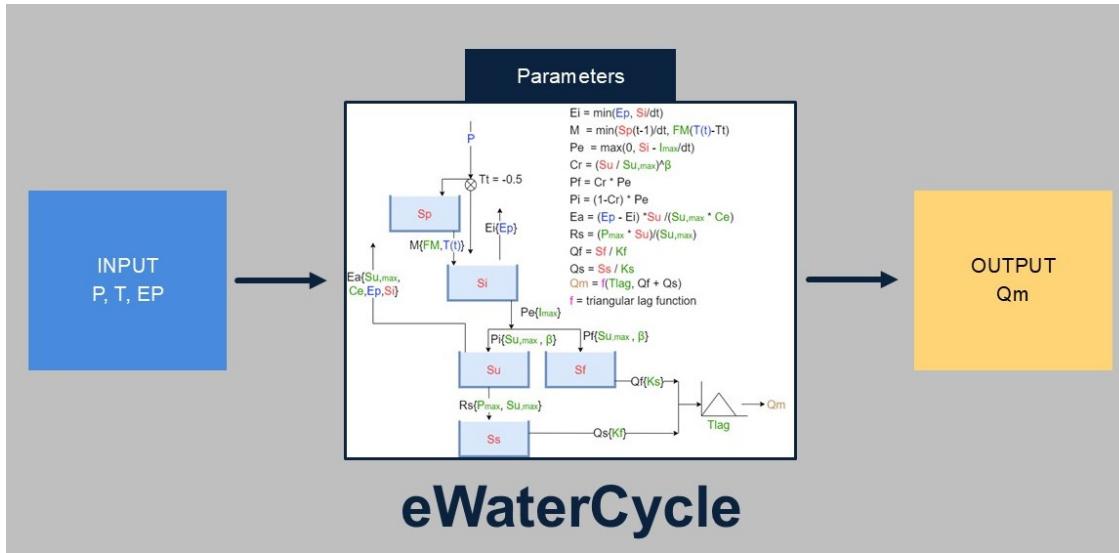
## 1.4 Approach

To determine the maximum discharge threshold for flooding of the U4, literature study needs to be done. Historical events, flood risk maps, research papers and monitoring stations need to be analysed.

The current frequency at which the discharge exceeds this maximum can be determined using eWaterCycle. Past streamflow can be simulated with a hydrological model. The simulated discharge data then can be compared to the flood threshold determined before. The model should be calibrated using real-world data. The frequency of flooding of the U4 with current discharge data can be conducted with return period calculations.

The future discharge can be simulated with the use of future climate projections. Climate models provided by CMIP6 will be used. Different IPCC scenarios give different projected climate variables which can be used in hydrological models using eWaterCycle to simulate future discharge. These modelled future discharge scenarios should be compared to the flood threshold to estimate the future frequency of exceedance for different climate change scenarios.

For this research a model through the eWaterCycle platform will be used. The eWaterCycle platform provides the hydrological community with models that can all be accessed in a similar manner, through the Jupyter notebook environment in eWaterCycle. A visual representation of the eWaterCycle platform can be seen below in Figure 1. The platform provides access to input, forcings, and to models, in this case the HBV model. The model can easily be changed to another model available through the eWaterCycle platform (Hut, 2022).



*Figure 1: Visual representation of the eWaterCycle platform. The eWaterCycle provides access to data and models. Parameters need to be calibrated. In this research the HBV model is used, but this model can easily be replaced for a different model. (Hrachowitz, z.d.; Hut, 2022)*

## 1.5 Reading Guide

In Chapter 2 the current maximum discharge ( $\text{m}^3/\text{s}$ ) is determined above which the U4 subway line will flood. Chapter 3 determines how often the discharge currently exceeds this maximum using the observation data accessible through eWaterCycle. In Chapter 4 the hydrological model

is chosen and calibrated. In Chapter 5 different climate change scenarios are projected through the hydrological model to simulate future discharge. These future discharge scenarios are compared to the flooding threshold to estimate the future flooding frequency of the U4 subway line. The discussion and conclusion can be found in Chapter 6 and Chapter 7.

This research can also be found on <https://www.ewatercycle.org/projects/main/intro.html>. The code used to obtain the results can be found through <https://github.com/eWaterCycle/projects>.

## 2 Current maximum discharge before flooding of the U4

### 2.1 Background

The Wien River finds its origin in the Wienerwald, west of Vienna, and enters the city after approximately 20 km. The river discharges into the Donaukanal. The historical hydrology of the Wien River cannot be reconstructed with certainty. However, before the construction of intercepting sewers along the river in the 1830s and its regulation and channelization at the end of the 19th century, the estimated mean annual discharge was approximately  $2 \text{ m}^3/\text{s}$  (Pollack et al., 2016). Since the flood retention basins that were made in the early 1800's, the 10-year return flood was estimated at  $140 \text{ m}^3/\text{s}$ , and the 100-year return flood at  $200 \text{ m}^3/\text{s}$ . Due to the high potential losses in the city of Vienna, the Wien River is designed to withstand a 1000-year discharge return period (Faber & Nachtnebel, 2002).

The catchment area of the Wien River can be seen in Figure 2. The observation data used in this research is obtained from the measuring station at the Kennedybrücke. Upstream from the Kennedybrücke the catchment area is rural, and downstream it is urban.

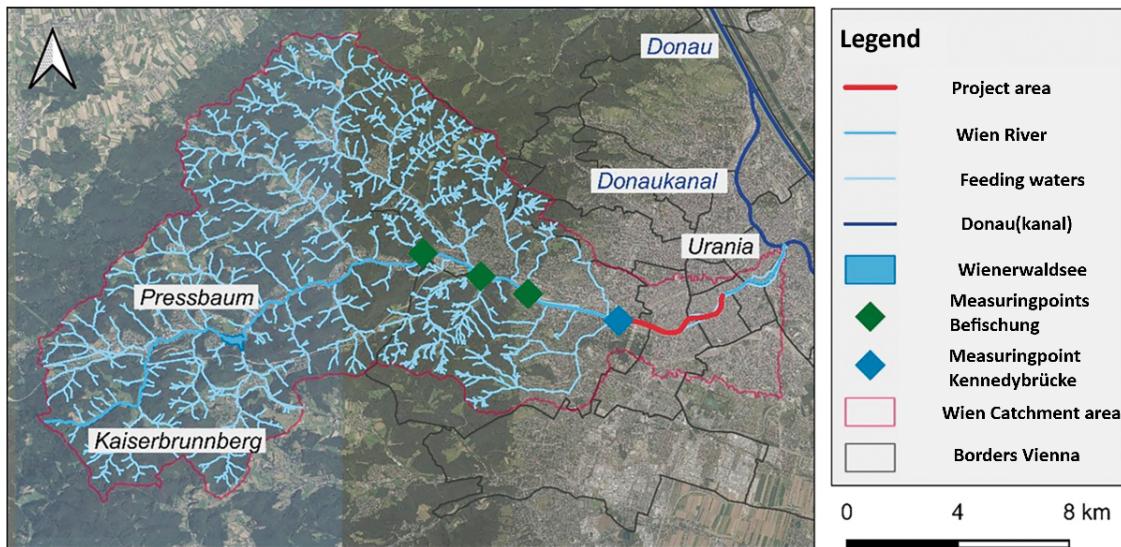


Figure 2: The catchment area of the Wien River, with the measuring station at the Kennedybrücke from which the observation data is obtained. The part upstream from the Kennedybrücke is the rural catchment, and downstream the urban catchment. (Sindelar et al., 2024)

### 2.2 Determination of flooding threshold

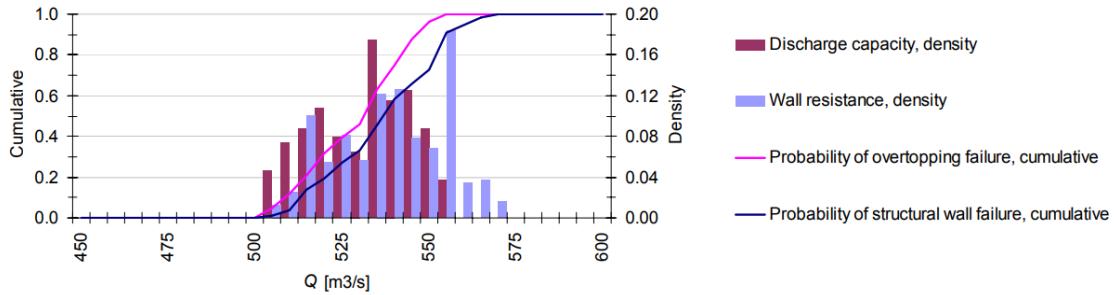
#### 2.2.1 Literature Study

Compton et al. (2009, p. 54) state: “A failure that results in the release of water to the U4 occurs when the discharge into the Wien River exceeds the given threshold, resulting in either overtopping of the floodwall or collapse of the floodwall due to either foundation scouring or hydrostatic pressure.” While failure due to overtopping is a function of the flow rate in the channel, uncertainties in the water flow rate are expected to be minimal since the Wien River is a channelized river with well characterized geometry. More uncertainty is expected in the erosive failure and wall collapse which are a function of the shear at the channel bed and the shear strength of the invert.

According to the report, the failure leading to overflowing of the U4 is expected to occur at a discharge of  $530 \text{ m}^3/\text{s}$ . Given the uncertainties in the floodwall's resistance parameters, this critical discharge is modelled as a normal distribution with mean value of  $530 \text{ m}^3/\text{s}$  and a standard deviation of  $10 \text{ m}^3/\text{s}$ . This means that at a discharge of around  $510 \text{ m}^3/\text{s}$  failure could occur with a probability of 5 percent, and at a discharge of around  $550 \text{ m}^3/\text{s}$  with a probability of 95 percent. (Compton et al., 2009)

The study of Faber (2006) also analyzed flood risk in an Austrian context, and specifically for the Wien River. He estimated peak flow frequencies using the rainfall-runoff model IHW for the rural catchment and ITWH for the urban catchment, and he used Monte Carlo simulations to account for uncertainties. A total of 7000 simulations were performed within the critical range of 400 to  $600 \text{ m}^3/\text{s}$  where failures were most likely to occur. Figure 3 shows that the flood walls can handle discharges up to  $500 \text{ m}^3/\text{s}$ , while overtopping of the floodwall is almost certain at a discharge of  $560 \text{ m}^3/\text{s}$ . The mean value of the discharge capacity before overtopping amounts to  $534 \text{ m}^3/\text{s}$ , with a standard deviation of  $14 \text{ m}^3/\text{s}$ . The mean value of discharge for structural floodwall failure is  $541 \text{ m}^3/\text{s}$  with a standard deviation of  $16 \text{ m}^3/\text{s}$  (Faber, 2006).

Faber also analyzed the probabilities of the different failure events, overtopping and structural damage of the flood wall, individually and in combination. The probability of structural flood wall failure without overtopping was not observed in any simulation. The overall system reliability is more than 99 percent, which indicates that failure of the flood wall is extremely rare with the used past peak flows. He further analyzed the return periods of the failure events. The installation of the controlled retention basins in 1998 increased the return period of failure from approximately 550 years to 1100 years. This return period exceeds the 1000-year discharge return period the Wien River is designed for (Faber & Nachtnebel, 2002). This deviation is due to limitations of the return period-based design, which does not fully account for uncertainties in flood frequency and magnitude.



*Figure 3: “Distribution of the resistance of the hydraulic system in terms of the bankfull discharge and the flow related to structural floodwall failure” (Faber, 2006)*

## 2.2.2 Conclusion

It can be concluded that the critical event for flooding of the U4 subway line is overtopping rather than structure failure of the flood wall. According to Faber (2006), the mean failure discharge at which overtopping happens is  $534 \text{ m}^3/\text{s}$ , with a 5 percent probability at  $511 \text{ m}^3/\text{s}$ , and a 95% probability at  $557 \text{ m}^3/\text{s}$ . These values align with the threshold values determined by Compton et al. (2009), who estimated a mean failure discharge of  $530 \text{ m}^3/\text{s}$ , with 5 and 95 percent probabilities of failure at discharges of approximately  $510$  and  $550 \text{ m}^3/\text{s}$ . The threshold values are normally distributed, so looking at one critical threshold value would be a simplification. The probability

of exceeding a threshold is heavily dependent on where that threshold is situated in the normal distribution. For this reason the return periods of all discharges in the normal distribution are calculated. This will give an indication of the return period of a certain discharge, and the probability that this specific discharge will lead to flooding of the U4 subway line. To make comparison easier, the graphs in the research will be illustrated with a threshold value of  $530 \text{ m}^3/\text{s}$ .

### 3 Current frequency of threshold exceedance

In this chapter the current frequency of exceeding the threshold determined in Chapter 2 will be analysed. This is done by looking at the available observation data of the catchment area of the Wien River. The eWaterCycle platform provides access to this data from camels in Austria. As determined in Chapter 2, the Wien River is designed for a 1000-year discharge return period. The observation data is unlikely to cover a period of 1000 years, so the data will need to be extrapolated to be able to estimate the return period of the previously established threshold.

#### 3.1 Observations

First of all, the available observation data will be investigated. This data is available from camels in Austria to which eWaterCycle provides access. This dataset contains data on precipitation, surface air temperature, potential evapotranspiration and discharge. The discharge observations can be seen in the graph in Figure 4 below.

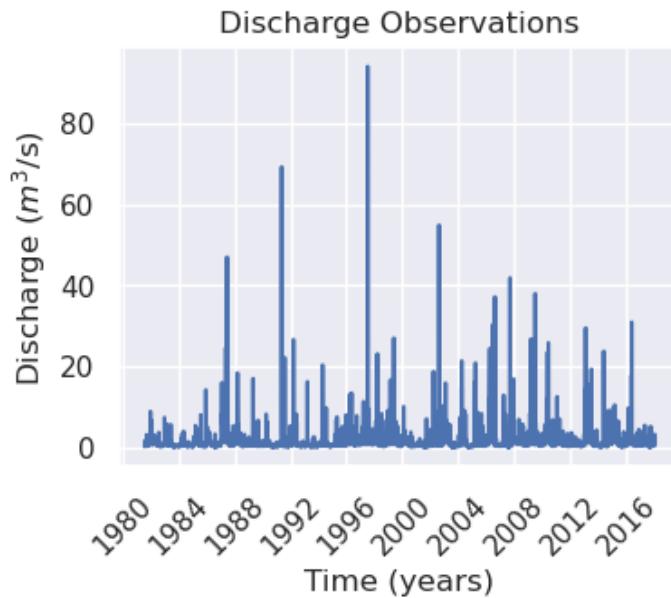


Figure 4: Discharge observations. The discharge ( $m^3/s$ ) is plotted against the time (years). The observation data is obtained from camels in Austria which eWaterCycle provided access to.

The observation data is available from 01-08-1981 to 01-01-2018, which covers a period of 36 years and 5 months. The highest peak present in this dataset has a discharge of  $94.05\text{ m}^3/\text{s}$ . Since there is no peak present in this dataset with a discharge value of the threshold value for flooding, the observation data will need to be extrapolated.

#### 3.2 Extrapolation

The observation data is plotted in Figure 5 in the form of a scatter plot, with on the y-axis the discharge and on the x-axis the corresponding return period. The blue line is fitted through the observation data, and is extrapolated to find out the return period for the threshold value.

The return periods corresponding to the discharges are calculated according to the Weibull equation with the following formula (Baaqeel et al., 2016):

$$T = \frac{n + 1}{\text{Rank} \cdot 365.25}$$

T = return period in years

n = total datapoints

rank = position of datapoint in sorted data list

A line is fitted through the observation data with the power law function:

$$f(x) = a \cdot x^b$$

parameters a and b are fitted with the curve\_fit function from scipy.optimize.

x represents the return periods

The fit 1 to fit 5 lines represent fitted lines where one or more of the observed data points are missing.

- fit 1 misses the highest peak
- fit 2 misses the second highest peak
- fit 3 misses the two highest peaks
- fit 4 misses the five highest peaks
- fit 5 misses the discharges lower than 0.5 m<sup>3</sup>/s

The lines from fit 1 to fit 5 are plotted as well to give an indication of the influence of certain peak discharges on the extrapolation. This will be useful later in the research, when the significance of the difference in future return periods with different climate change scenarios compared to the current return periods needs to be assessed.

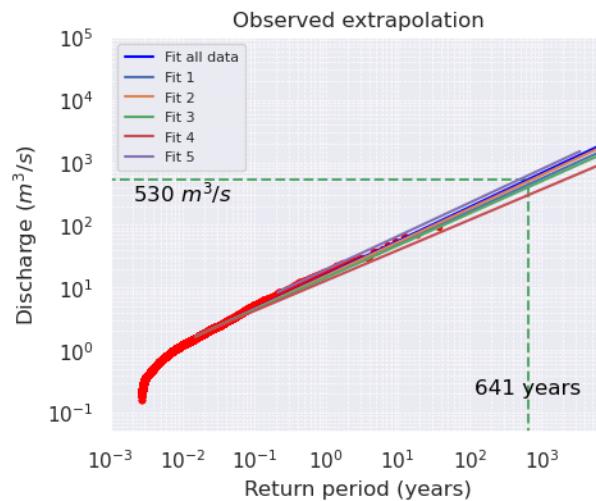


Figure 5: Observation Extrapolation. The observed data and corresponding return periods are plotted as red dots.

They are fitted and extrapolated using the power law function.

The different fits represent datasets without certain datapoints, to show the influence of certain discharges on returnperiod calculations.

In Figure 5 it can be seen that the presence of certain peak discharges in the data have a big influence on the extrapolation. Using all data, the key discharges from the normal distribution determined in Chapter 2 have the following return periods:

- $510 \text{ m}^3/\text{s}$ : 597 years
- $511 \text{ m}^3/\text{s}$ : 599 years
- $530 \text{ m}^3/\text{s}$ : 641 years
- $534 \text{ m}^3/\text{s}$ : 650 years
- $550 \text{ m}^3/\text{s}$ : 686 years
- $557 \text{ m}^3/\text{s}$ : 702 years

Removing the highest discharge value results in a return period of 923 years for a discharge of  $530 \text{ m}^3/\text{s}$ , removing the five highest peaks results in a return period of 2100 years for a discharge of  $530 \text{ m}^3/\text{s}$ , and removing the discharges lower than  $0.5 \text{ m}^3/\text{s}$  results in a return period of 487 years. These influences of certain discharge values on the return period calculation are important to consider when comparing the model output to the observations.

In Figure 6 both the normal distribution of the exceedance threshold values and the extrapolated return periods are plotted. A discharge value more to the right of the normal distribution, has a higher return period, but also has a higher probability of actually causing flooding of the U4 subway line.

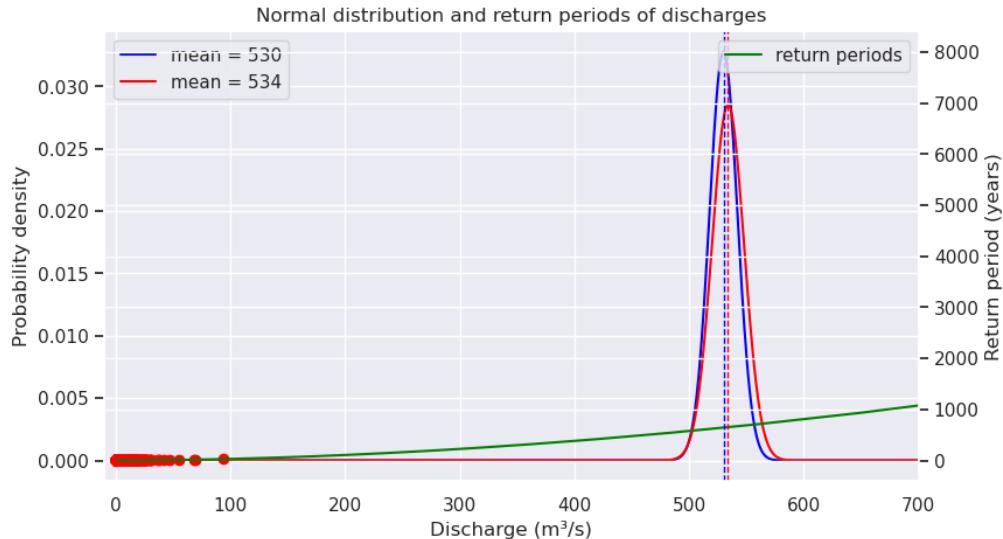


Figure 6: Normal distribution and return periods of discharges.  
The observation data and corresponding return periods are shown as red dots, and the fitted extrapolation is the green line.  
This figure shows how higher discharges have higher return periods, but have a higher probability of actually causing flooding of the U4

### 3.3 Conclusion

It can be concluded that the current return period for the mean threshold value of  $530 \text{ m}^3/\text{s}$  is 641 years, and for  $534 \text{ m}^3/\text{s}$  is 650 years. This means that the current return period of flooding is shorter than the return period the Wien River was designed for, 1000 years.

## 4 Calibration of the HBV model

In this chapter the parameters used in the HBV model are calibrated using the observation data. A set of parameters is used by the HBV model to model the discharge. These parameters are calibrated to make the model work for the catchment area of the Wien River.

### 4.1 Model Selection

To be able to model the expected future discharge, the Hydrologiska Byråns Vattenbalansavdelning (HBV) model is chosen. The developer of the model, Sten Bergström (1992), says the HBV model can best be classified as a semi-distributed conceptual model. The model represents a catchment area by using interconnected storage reservoirs to simulate the movement of water. By adjusting nine parameters, the HBV model can be calibrated to different catchment areas. The model requires 3 inputs, the forcings, which are precipitation, temperature and potential evapotranspiration.

The nine parameters that have to be calibrated are **I<sub>max</sub>**, **C<sub>e</sub>**, **S<sub>umax</sub>**, **B<sub>eta</sub>**, **P<sub>max</sub>**, **T<sub>lag</sub>**, **K<sub>f</sub>**, **K<sub>s</sub>**, and **F<sub>M</sub>**. The parameters represent the following:

Parameter	Description
<b>I<sub>max</sub></b>	Maximum intensity
<b>C<sub>e</sub></b>	Coefficient of evaporation
<b>S<sub>umax</sub></b>	Field capacity
<b>B<sub>eta</sub></b>	Shape coefficient
<b>P<sub>max</sub></b>	Maximum percolation rate
<b>T<sub>lag</sub></b>	Time lag
<b>K<sub>f</sub></b>	Fast run-off parameter
<b>K<sub>s</sub></b>	Slow run-off parameter
<b>F<sub>M</sub></b>	Degree-day factor

*Table 1: The 9 parameters of the HBV model that need to be calibrated for the investigated catchment area. This table shows the meaning of the individual parameters. (Wawrzyniak et al., 2017; Seibert, 2000)*

A visual representation of the HBV model can be seen in Figure 7.

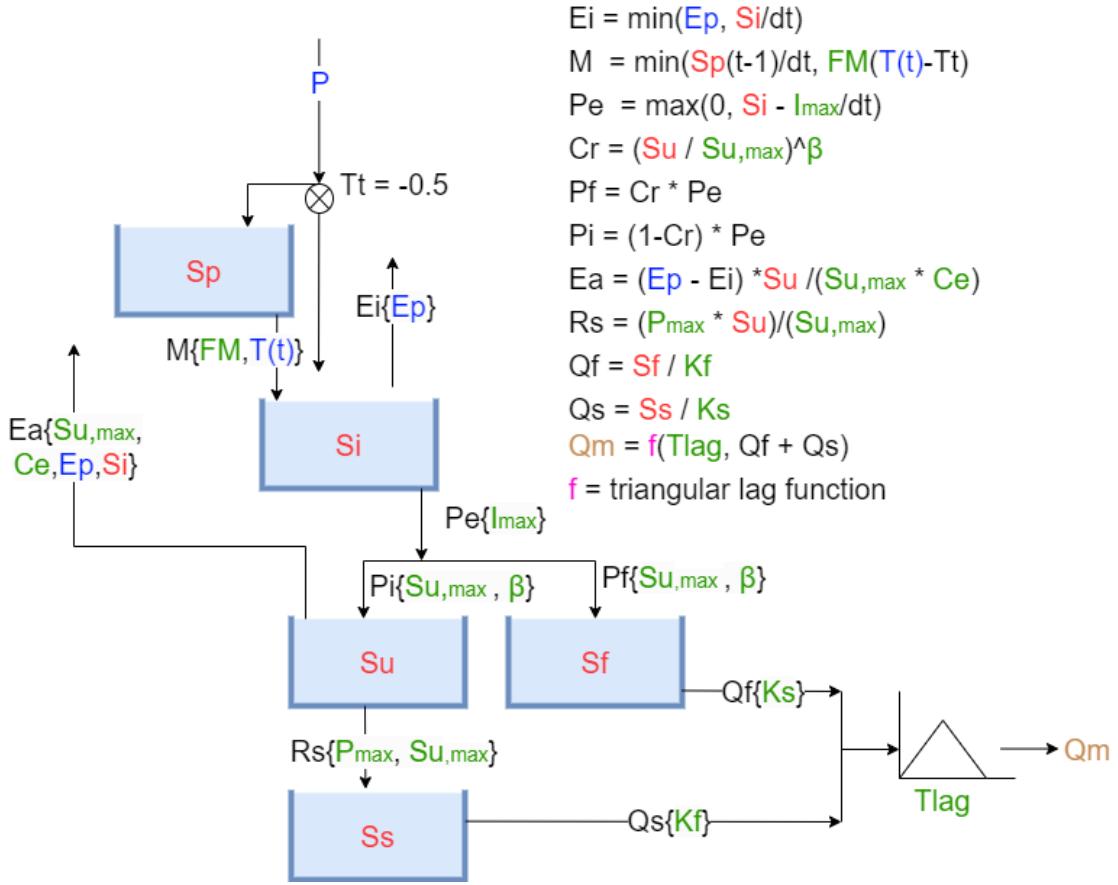


Figure 7: Layout of the HBV Model (Hrachowitz, z.d.)

## 4.2 Calibration

The model's performance with a certain parameter combination is assessed by evaluating how well it represents the distribution of discharges, since for my research the exact timing of certain discharges and peaks is not relevant. To measure the difference between the observed and modelled discharges, the Root Mean Squared Error (RMSE) is used. The Nelder-Mead Optimization is then applied to find the best set of parameters that results in the lowest RMSE. The model is trained on the first 75 percent of the observation data, and is then tested on the last 25 percent of the data. The test shows how well the model performs on data which it has never seen before.

### 4.2.1 Root Mean Squared Error

The RMSE is used to determine the difference between the observed and modelled discharge. First, the difference between the observed and modelled discharge is calculated for each day, called the error. The RMSE squares this error to give more weight to large differences, such as when observed peaks are not modelled well. This method causes peaks to contribute more heavily to the RMSE. After this, the mean of all squared values is taken, before the square root is applied to get the RMSE.

#### 4.2.2 Nelder-Mead Optimization

The optimal parameter combination that results in the lowest RMSE is calculated using Nelder-Mead Optimization. The Nelder-Mead algorithm is a way to find the minimum of a function without needing to calculate its derivative. It does this by adjusting a group of points through steps like reflecting, expanding, squeezing, or shrinking until it gets closer to the best solution (Chen, 2023).

#### 4.2.3 Calibration Results

The parameter combination that resulted in the best performing modelled discharge is as follows:

Parameter	Value
<b>I<sub>max</sub></b>	0.000
<b>C<sub>e</sub></b>	1.324
<b>S<sub>umax</sub></b>	100.048
<b>Beta</b>	3.894
<b>P<sub>max</sub></b>	0.666
<b>T<sub>lag</sub></b>	0.043
<b>K<sub>f</sub></b>	1.005
<b>K<sub>s</sub></b>	1.940
<b>F<sub>M</sub></b>	0.459

Table 2: This table shows the values of the 9 parameters of the HBV model after calibration for the catchment area of the Wien River

The distributions of the observed and modelled discharges for the calibration period can be seen below in Figure 8.

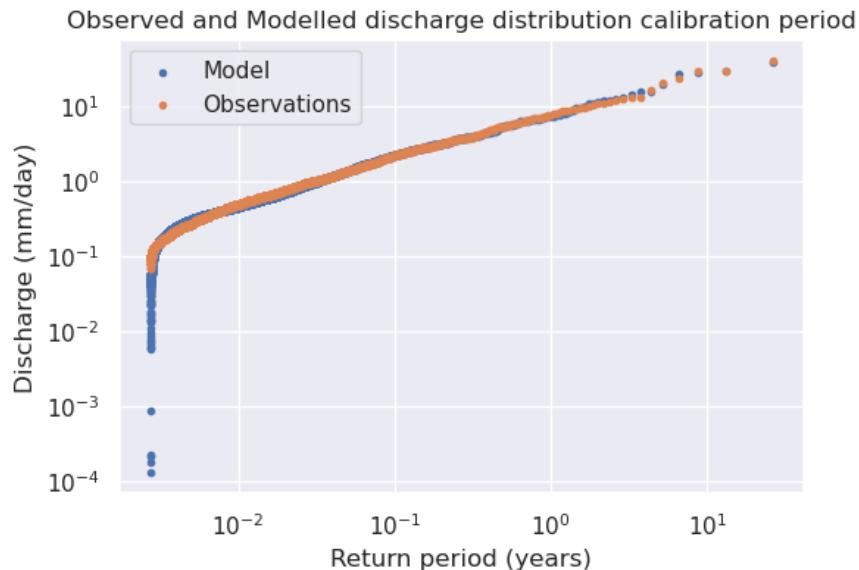
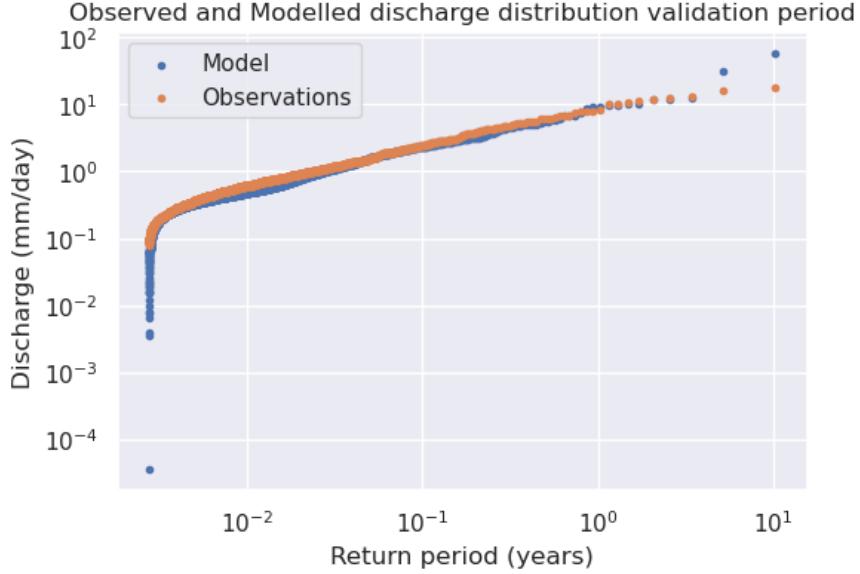


Figure 8: Distribution of the observed and modelled discharges for the calibration period.  
The distribution of the peaks is more important for my research than  
the exact timing of the peaks, which is why the distribution is investigated

To test the ability of the model to predict discharges for periods of which it has never seen data before, the modelled discharge for the validation period is compared to the observed discharge. The distribution of discharges for modelled and observed data for the validation period can be seen in Figure 9.



**Figure 9: Distribution of the observed and modelled discharges for the validation period.**  
The calibration needs to be tested on data it has never seen before,  
to make sure the parameters are not overfitted on the data they were trained on.

The RMSE for both the calibration and validation period can be calculated and compared. The RMSE of the calibration period is 0.0632, and is 0.7020 for the validation period. The RMSE of the validation period is around 11 times larger than the one for the calibration period. The cause of the RMSE being larger for the validation period is visible in Figure 9 above. The model predicts more larger peaks than observed. This is caused by the available observation data. The last period of the observation data contains no peaks, while the rest of the data does contain peaks. This causes the model to overestimate the actual amount of peaks present in the validation period. In both the calibration and validation period the model predicts lower discharges than are observed. A better insight in the differences between the modelled and observed discharges for the calibration and validation period can be found in Appendix I. Later in this chapter the return period of the threshold value is calculated by extrapolation of the modelled data. If the model works well enough with this parameter combination for the catchment area of the Wien River will be assessed in this part.

To be able to compare the model output to the observation data, the model output is converted from mm/day to  $\text{m}^3/\text{s}$  as well. This is plotted below in Figure 10a for both the calibration and validation period. To check whether or not the modelled hydrograph is realistic, the plot in Figure 10b zooms in on a period of two years. It can be seen that the model follows the discharge patterns

quite well, which means the modelled discharge can be assessed as realistic.

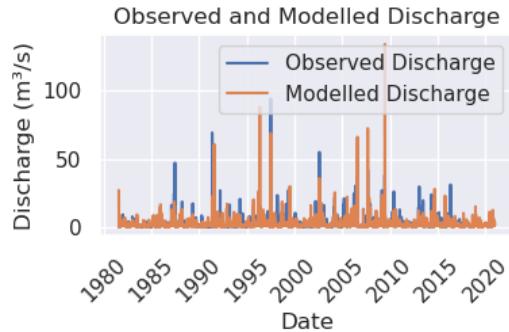


Figure 10a: The modelled and observed discharge for the entire period.

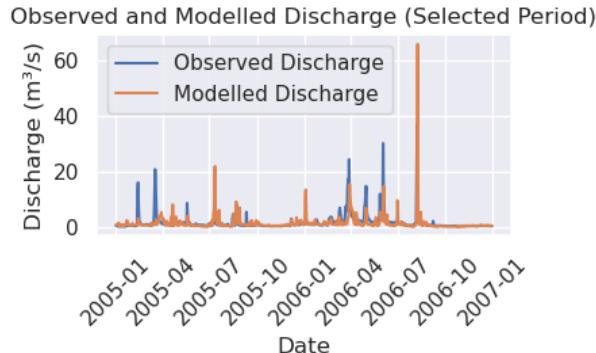


Figure 10b: The modelled and observed discharge zoomed in on a 2-year period.  
The modelled discharge is realistic

Below, in Figure 11, the distribution of the modelled discharge is plotted. The blue line is fitted through the modelled data, and is extrapolated to be able to determine the return period for the threshold value. It can be seen that a discharge of  $530 \text{ m}^3/\text{s}$  has a return period of 398 years. In Chapter 3, the observed data resulted in a return period of 641 years for a discharge of  $530 \text{ m}^3/\text{s}$ . These two extrapolated return periods are not exactly the same. In Chapter 3, the influence of peaks in the data was considered as well. Removing the highest peaks resulted in a return period of 923 years. Removing peaks from the data increases the return period, and thus it can be assumed that adding peaks will decrease the return period. Removing the lowest discharges decreases the return period. The model overestimates the amount of peaks in the data, resulting in a smaller return period. Overestimating peaks is safer than underestimating peaks, because it prevents underdesigning of flood protection infrastructure and improves the preparedness for flooding. Since it can not be expected from the model to perfectly predict the discharges, a return period of 398 years for a discharge of  $530 \text{ m}^3/\text{s}$  is considered close enough to the return period of 641 years for the observed data.

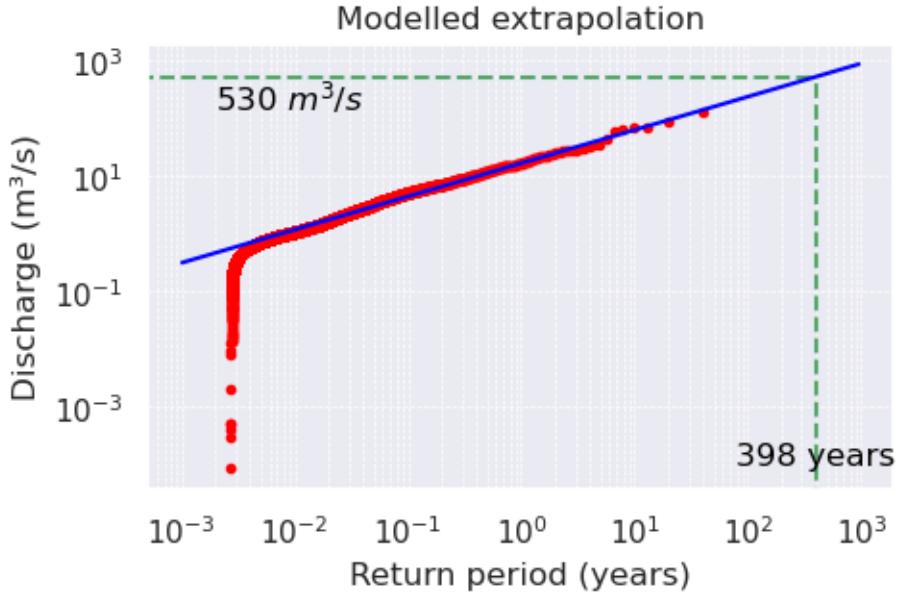


Figure 11: Distribution, fit and extrapolation of the modelled discharge.  
The return period for the threshold value is 398 years.

### 4.3 Conclusion

It can be concluded that the best parameter combination for the catchment area of the Wien River is:

Parameter	Value
I <sub>max</sub>	0.000
C <sub>e</sub>	1.324
S <sub>umax</sub>	100.048
Beta	3.894
P <sub>max</sub>	0.666
T <sub>lag</sub>	0.043
K <sub>f</sub>	1.005
K <sub>s</sub>	1.940
FM	0.459

Table 3: This table shows the values of the 9 parameters of the HBV model after calibration for the catchment area fo the Wien River

In the next chapter, the HBV model will be run using forcings generated by CMIP6 in combination with these parameters.

## 5 Future frequency of threshold exceedance under different climate change scenarios

In this chapter the future frequency of threshold exceedance is determined. This is done using forcings based on different climate change scenarios, the Shared Socioeconomic Pathways (SSPs). SSPs are a set of narratives that outline potential future scenarios for human society, focussing on the use of fossil fuels and the social and economic factors that influence the consumption of fossil fuel (ClimateData.ca, 2025). The five main SSPs (Meinshausen et al., 2020) are being considered:

Climate Scenario (SSP)	Description
<b>SSP1-1.9</b>	Sustainability scenario, aligns with a 1.5°C target under the Paris Agreement.
<b>SSP1-2.6</b>	Sustainability scenario with a radiative forcing level of 2.6 W/m <sup>2</sup> by 2100.
<b>SSP2-4.5</b>	“Middle of the road” scenario, assumes current development patterns continue, leading to a 4.5 W/m <sup>2</sup> forcing level by 2100.
<b>SSP3-7.0</b>	Medium-high reference scenario in the ‘regional rivalry’ socio-economic family.
<b>SSP5-8.5</b>	High-emission scenario in the ‘fossil-fueled development’ socio-economic family.

*Table 4: The five main SSPs with descriptions (Meinshausen et al., 2020; ClimateData.ca, 2025)*

The future discharges are modelled with the HBV model using forcings from CMIP6 (Coupled Model Intercomparison Project) (Azote, 2025) as inputs. The forcings used in this research are generated by the German MPI model. Only one of the 25 ensembles is used, because in this research there is not enough time to look at all of the ensemblers. The used ensemble is called ‘r1i1p1f1’. The nine parameters for the HBV model are calibrated in Chapter 4.

### 5.1 Historical Comparison

To be able to determine how well the model predicts discharge data using CMIP6, it is important to compare the model output to observation data. This shows us the uncertainty of the model. The distribution of this modelled discharge and the extrapolation can be seen in Figure 12.

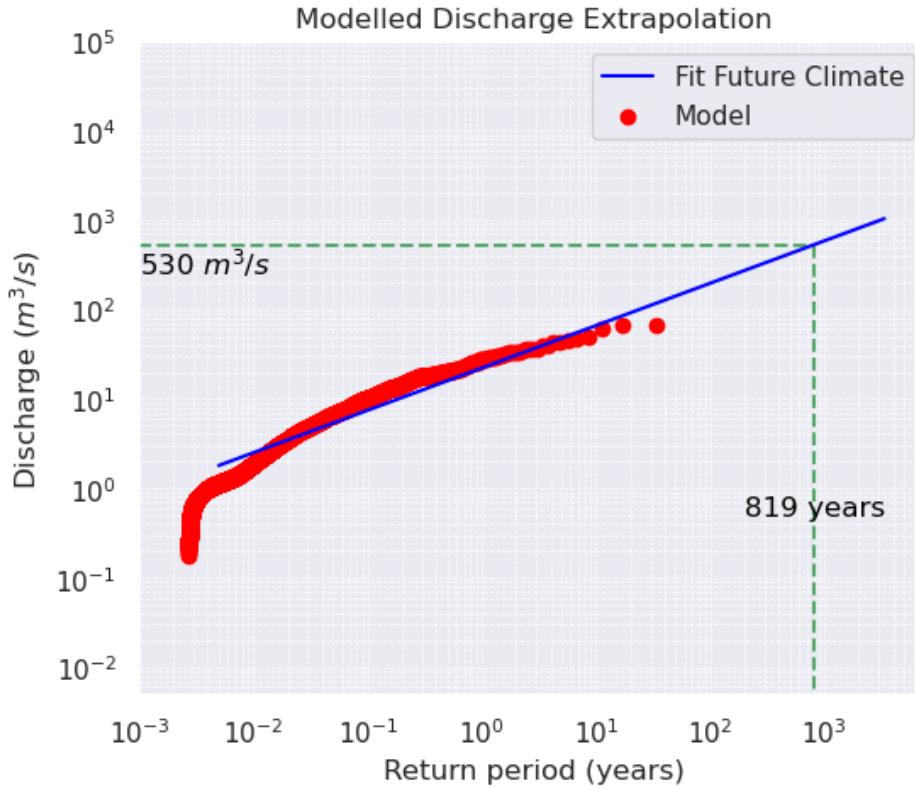


Figure 12: Historical modelled discharge distribution and extrapolation using CMIP.  
This return period can be compared to the one obtained from the observed data to assess the model prediction uncertainty.

In Figure 12 it can be seen that the mean threshold value of  $530 \text{ m}^3/\text{s}$  has a return period of 819 years. The extrapolated return period of the observed discharge determined in Chapter 3 amounted to 641 years. This means that the modelled and observed data result in a 27.7% difference. Therefore, with determining the return periods with different climate change scenarios, a model uncertainty of 27.7% is taken into account.

## 5.2 Future Scenarios

Next, the future discharges for different climate change scenarios can be modelled using forcings from CMIP. The future discharge is projected for a period of 40 years, from 2060 to 2100. The observation data is available for a period of about 40 years, so to be able to make a good comparison between the historical observation data and the future projected data, it is imported to use periods of equal lengths. This ensures that the comparison reflects differences in discharge rather than differences in trends caused by unequal time spans. It eliminates the influence of longer-term variations or trends that may occur over extended periods. A period at the end of the century is used rather than the upcoming 40 years, since climate change projections are expected to get more extreme. By modeling future discharge under the most extreme expected climate scenarios, recommendations can be provided for the development of infrastructure designed for long-term resilience and protection.

In Figure 13 the future modelled discharges are plotted for the different climate change scenarios. In Figure 14 the distribution and extrapolation to the threshold values can be seen.

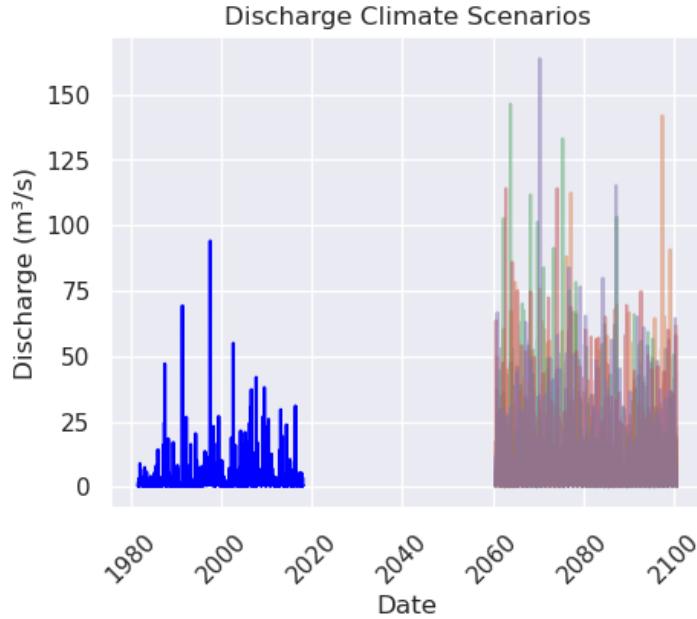


Figure 13: Historical observed discharge and modelled future discharge for different climate scenarios.  
This figure shows that the future discharges seem to get higher than the observed discharges

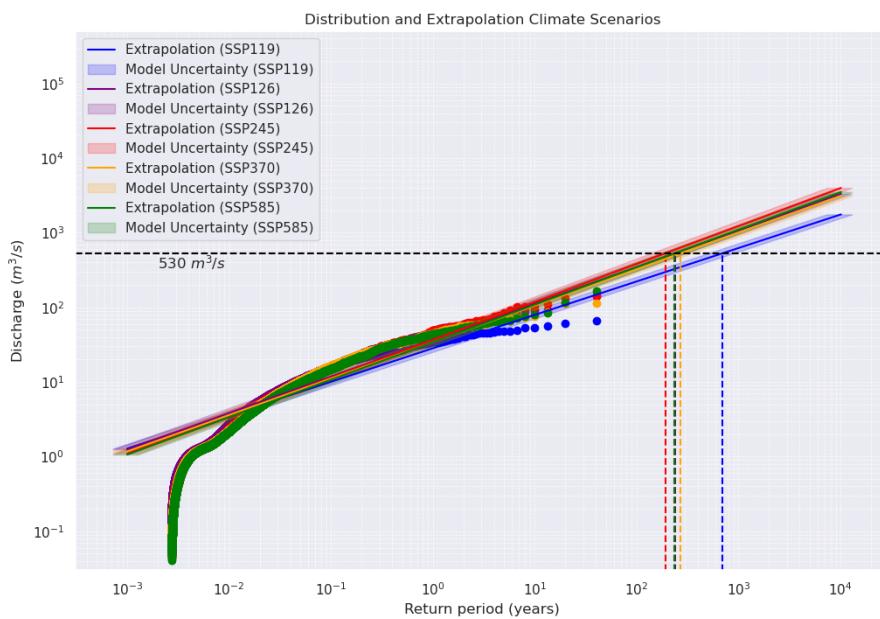


Figure 14: Modelled discharge distribution and extrapolation for different climate change scenarios  
The projected discharges with corresponding return periods for different climate change scenarios are plotted, fitted and extrapolated. The model uncertainty of 27.7% is shown as well. The return periods for the threshold value for the different climate change scenarios can be extracted and are shown with the vertical dashed lines.

### 5.3 Conclusion

From Figure 14, it can be concluded that the future return periods for a discharge of  $530 \text{ m}^3/\text{s}$  considering the different climate change scenarios are:

Climate Scenario (SSP)	Return Period (Years)	Model Uncertainty Range (Years)
SSP119	697	504 – 890
SSP126	238	172 – 292
SSP245	191	138 – 234
SSP370	272	197 – 334
SSP585	235	170 – 288

*Table 5: The SSPs with their return periods for the threshold value for flooding of the U4. The uncertainty range of 27.7% is calculated as well.*

## 6 Discussion

### 6.1 Key Findings

This research aimed to investigate the impact of predicted climate change on the hydrology of the Wien River and the implications for flooding of the adjacent U4 subway line. The study focussed on determining the flooding threshold, estimating the current frequency of threshold exceedance, and projecting future frequency of exceedance under different climate change scenarios using the HBV model.

The findings indicate that climate change will increase the frequency of extreme discharges, reducing the return periods for flooding.

#### 6.1.1 Current Threshold for Flooding

In Chapter 2, the critical discharge at which the U4 subway line is expected to flood is determined to be at a mean value of 530 or 534 m<sup>3</sup>/s. These thresholds are based on studies by Faber (2006) and Compton et al. (2009), who determined that overtopping is the primary mechanism leading to flooding, rather than structural failure of the floodwall. Given that the threshold values are normally distributed, overtopping could happen with a probability of 5% at discharge values of 510 and 511 m<sup>3</sup>/s, or with a 95% probability of actually causing overtopping at discharge values of 557 and 550 m<sup>3</sup>/s.

#### 6.1.2 Current Frequency of Threshold Exceedance

The observation data used in Chapter 3 shows that the Wien River has not exceeded the threshold value in the period of 1981 to 2018. The return period of the threshold value is estimated using extrapolation. The return periods of the key discharges from the normal distributions are as followed:

Discharge (m <sup>3</sup> /s)	Return Period (Years)	Probability of Flooding
510	597	5%
511	599	5%
530	641	50%
534	650	50%
550	686	95%
557	702	95%

Table 6: Key discharges with return periods and probability of actually causing flooding of the U4

These return periods imply that under current climate conditions, the return period of the discharge severe enough to cause flooding of the U4 subway line is already more frequent than the 1000 year return period the Wien River was designed for.

#### 6.1.3 Future Frequency of Threshold Exceedance under Climate Change Scenarios

Using the HBV model with CMIP6 climate projections, future return periods for the 530 m<sup>3</sup>/s flooding threshold were estimated under different SSP scenarios. These results show that climate change scenarios have significant impact on the return periods of the discharge threshold value.

The return periods for different climate change scenarios for a threshold value of  $530 \text{ m}^3/\text{s}$  are determined:

Climate Scenario (SSP)	Return Period (Years)	Model Uncertainty Range (Years)
SSP119 (sustainability)	697	504 – 890
SSP126 (sustainability)	238	172 – 292
SSP245 (middle of the road)	191	138 – 234
SSP370 (regional rivalry)	272	197 – 334
SSP585 (fossil-fueled development)	235	170 – 288

*Table 7: The SSPs with their return periods for the threshold value for flooding of the U4. The uncertainty range of 27.7% is calculated as well.*

## 6.2 Interpretation of Results

If the return periods under the different climate change scenarios are compared to the current return period of the threshold value, it can be seen that only SSP119 results in a higher return period. The rest of the climate change scenarios all result in a lower return period and thus a higher frequency of flooding of the U4 subway line. A shocking outcome is the return period of scenario SSP126, this is a sustainability scenario, but results in a lower return period than the current return period. This means that increased floodrisk is likely even with climate mitigation efforts.

## 6.3 Implications of Findings

The findings of this research have significant implications for the urban flood risk management in Vienna. The increasing frequency of extreme discharges increases the risk of flooding of the U4 subway line. The Wien River is currently designed to withstand a discharge with a return period of 1000 years, but in reality the Wien Rivers flooding threshold currently has a return period of 641 years. Under four of the five investigated climate change projections this return period of flooding will only decrease. This means that the current flood protection infrastructure of the Wien River is insufficient to realize the designed return period for flooding of 1000 years, and needs reinforcement.

## 6.4 Limitations

While this research provides a valuable assessment of future flood risk of the U4 subway line, some limitations must be acknowledged. First of all, the observed data covers only a period of about 40 years and does not contain a discharge as high as the threshold value, resulting in the need for extrapolation. This means that estimating the return periods for extreme discharges relies on statistical extrapolation, introducing the risk for potential errors.

The data only being available for a period of around 40 years also influences the parameter calibration. The calibration period contains peaks, but the validation period does not. This causes the model to overestimate the amount of peaks in the validation period. It is decided to choose the calibration and validation periods like this, and not to alter them. Altering them, so choosing a calibration period with the period without peaks, causes the model to underestimate the amount of

peaks in the validation period. Overestimating peaks is more ‘safe’ than underestimating peaks. It is important to consider that the model is calibrated on a limited period of data, and it is uncertain how representative this data is.

Furthermore, the HBV model, like all hydrological models, is a simplification of reality. The model has inherent uncertainties caused by the parameter calibration, data quality and simplification in the representation of the catchment area and the inherent physical processes.

The use of climate projections also brings its uncertainties. Climate models are biased, they are focussed on perfecting the energy balance. This means they are not focussed on specifically the waterbalance. Projections for precipitation may thus not be as good as temperature projections. Furthermore, in this research only one of the 25 ensemlemembers of the MPI model is used because of a lack of time. Other ensemlemembers could give other projections that result in different return periods. Using all the 25 ensemlemembers would reduce uncertainties.

Despite these limitations, the research provides a good understanding of the future flood risks under different climate change scenarios, and highlights the need for action in urban flood protection infrastructure.

## 6.5 Recommendations

To improve the accuracy and applicability of flood risk assessment of the Wien River, future studies should consider using all the 25 ensemlemembers of the MPI climate model. In this way uncertainty is reduced, and the results of the return period calculations will be more reliable.

Furthermore, the flood protection infrastructure can be investigated. Due to the time limitation of this research this could not be done, but it would be of great relevance to investigate the current flood protection infrastructure of the Wien River and look for where and what reinforcements can be applied.

## 7 Conclusion

This research investigated the impact of climate change on the hydrology of the Wien River and its implications for flooding of the adjacent U4 subway line. This is investigated by using the HBV model with CMIP6 climate change projections. The current discharge above which the Wien River will flood is  $530\text{--}534 \text{ m}^3/\text{s}$ , with overtopping as the primary failure mechanism.

The current return period of this threshold exceedance is determined to be 641 years, which is already more frequent than the design return period of 1000 years originally intended for the Wien Rivers flood management system. Under climate change scenario SSP119, the return period of threshold exceedance is expected to increase to 697 years, suggesting a slight decrease in flood risk. However, under all other projected climate change scenarios the return periods are expected to decrease, ranging from 191 to 272 years. This suggests a significant increase in the frequency of extreme discharge events.

Overall, it can be concluded that the impact of climate change on the hydrology of the Wien River is that extreme discharges will become more frequent in four out of the five SSP scenarios. Its implications for flooding of the adjacent U4 subway line are that its return period for flooding will be far more frequent than the intended design period of once every 1000 years. Given these projections, reinforcing flood protection infrastructure is of great importance to maintain safety and functionality of the U4 subway line.

## 8 References

- Azote. (2025, 26 februari). CMIP - Coupled Model Intercomparison Project. Coupled Model Intercomparison Project. <https://wcrp-cmip.org/>
- Baaqeel, A. H., Quliti, S. A. E., Daghreri, Y. A., Hajlaa, S. A. B., & Yami, H. H. A. (2016). Estimating the Frequency, Magnitude and Recurrence of Extreme Earthquakes in Gulf of Aqaba, Northern Red Sea. *Open Journal Of Earthquake Research*, 05(02), 135–152. <https://doi.org/10.4236/ojer.2016.52011>
- Bergström, S. (1992). THE HBV MODEL - its structure and applications. In SMHI REPORIS HYDROLOGY (p. 32). SMHI.
- Chen, E. (2023, 9 april). The Simplex Approach of the Nelder-Mead Algorithm - Ethan Chen - Medium. Medium. <https://medium.com/%40ethanwchen/the-simplex-approach-of-the-nelder-mead-algorithm-c582556d8d84>
- ClimateData.ca. (2025, 28 februari). Understanding Shared Socio-economic Pathways (SSPs) — ClimateData.ca. <https://climatedata.ca/resource/understanding-shared-socio-economic-pathways-ssps/#:~:text=Shared%20Socioeconomic%20Pathways%20are%20a,which%20drive%20fossil%20fuel%20use.>
- Compton, K. L., Faber, R., Ermolieva, T. Y., Linnerooth-Bayer, J., & Nachtnebel, H.-P. (2009). Uncertainty and Disaster Risk Management Modeling the Flash Flood Risk to Vienna and Its Subway System (RR-09-002). International Institute for Applied Systems Analysis.
- Dore, M. H. I. (2005). Climate change and changes in global precipitation patterns: What do we know? *Environment International*, 31, 1167–1181. <https://www.elsevier.com/locate/envint>
- Faber, R. (2006). FLOOD RISK ANALYSIS: RESIDUAL RISKS AND UNCERTAINTIES IN AN AUSTRIAN CONTEXT (Door University of Natural Resources and Applied Life Sciences, Vienna).
- Faber, R., & Nachtnebel, H. P. (2002). Flood Risk Assessment in Urban Areas: Development and application of a stochastic hydraulic analysis method considering multiple failure types. In J. Linnerooth-Bayer (Ed.), Proceedings Second Annual IIASA-DPRI Meeting, Integrated Disaster Risk Management, Held On 29–31 July 2002, Laxenburg, Austria.
- Hrachowitz, M. (z.d.). TU Delft Couse ENVVM1502 - “River Basin Hydrology”.
- Hut, R., Drost, N., van de Giesen, N., van Werkhoven, B., Abdollahi, B., Aerts, J., Albers, T., Alidoost, F., Andela, B., Camphuijsen, J., Dzigan, Y., van Haren, R., Hutton, E., Kalverla, P., van Meersbergen, M., van den Oord, G., Pelupessy, I., Smeets, S., Verhoeven, S., de Vos, M., and Weel, B.: The eWaterCycle platform for open and FAIR hydrological collaboration, *Geosci. Model Dev.*, 15, 5371–5390, <https://doi.org/10.5194/gmd-15-5371-2022>, 2022.
- Krone.At. (2024a, september 15). “Landfall” in the city - Dramatic pictures: Vienna in the grip of the floods. Kronen. <https://www.krone.at/3526665>
- Krone.At. (2024b, september 19). Vienna public transport balance sheet - Flood had “unprecedented impact on public transport”. Kronen. <https://www.krone.at/3531234>
- Meinshausen, M., Nicholss, Z. R. J., Lewis, J., Gidden, M. J., Vogel, E., Freund, M., Beyerle, U., Gessner, C., Naueis, A., Bauer, N., Canadell, J. G., Daniel, J. S., John, A., Krummel, P. B., Luderer, G., Meinshausen, N., Montzka, S. A., Rayner, P. J., Reimann, S., . . . Wang, R. H. J.

(2020). The SSP greenhouse gas concentrations and their extensions to 2500. In Geosci. Model Dev. (Vols. 13–13, pp. 3571–3605). <https://doi.org/10.5194/gmd-13-3571-2020>

Pollack, G., Gierlinger, S., Haidvogl, G., & Winiwarter, V. (2016). Using and abusing a torrential urban river: the Wien River before and during industrialization. *Water History*, 8(3), 329–355. <https://doi.org/10.1007/s12685-016-0177-7>

Seibert, J. (2000). Multi-criteria calibration of a conceptual runoff model using a genetic algorithm. *Hydrology And Earth System Sciences*, 4(2), 215–224. <https://doi.org/10.5194/hess-4-215-2000>

Sindelar, C., Lasinger, N., Buchinger, M., Leutgöb, E., & Habersack, H. (2024). Fish passability through innovative groynes in combined scaled and etho-hydraulic 1:1 model tests using the example of the Wienfluss river. *Österreichische Wasser- Und Abfallwirtschaft*, 76(3–4), 151–157. <https://doi.org/10.1007/s00506-024-01028-5>

Wawrzyniak, T., Osuch, M., Nawrot, A., & Napiorkowski, J. J. (2017). Run-off modelling in an Arctic unglaciated catchment (Fuglebekken, Spitsbergen). *Annals Of Glaciology*, 58(75pt1), 36–46. <https://doi.org/10.1017/aog.2017.8>

## 9 Appendix I

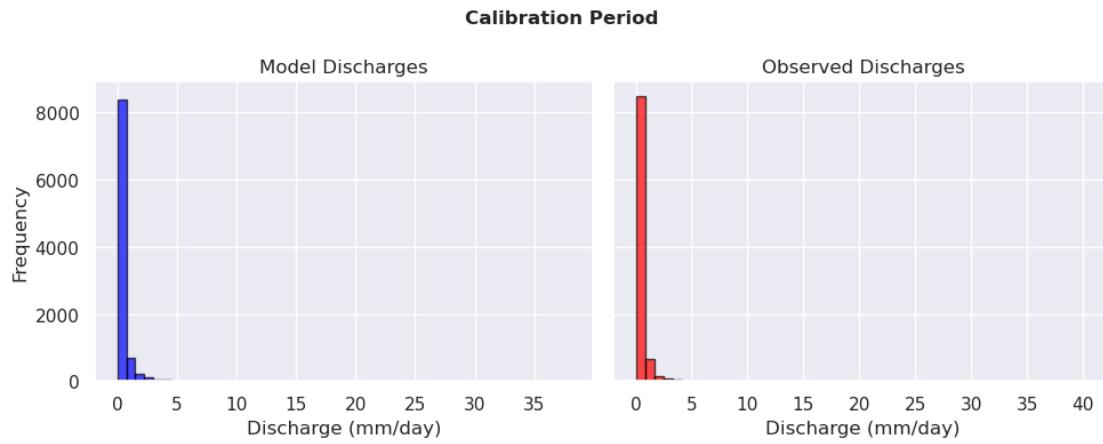


Figure I1: Histogram of the observed and modelled discharges for the calibration period.  
It can be seen that the modelled discharges are similarly distributed as the observed discharges.

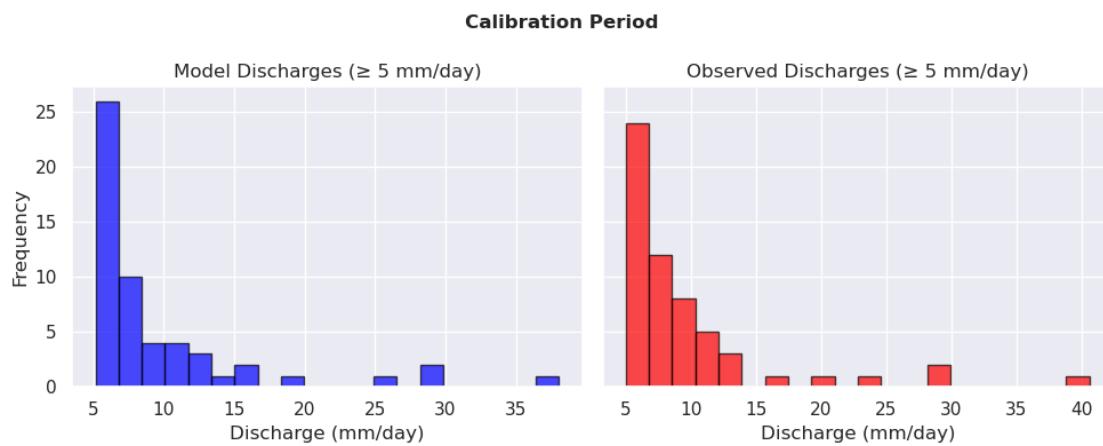


Figure I2: Histogram of the observed and modelled discharges for the calibration period.  
Only values  $\geq 5$  mm/day are included to focus on relevant discharges.  
The distribution of the modelled and observed discharge are very similar

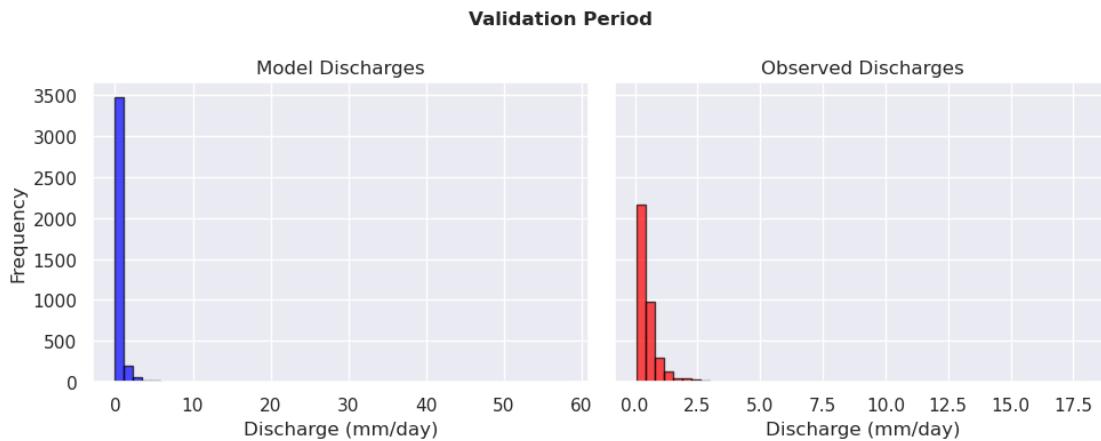


Figure I3: Histogram of the observed and modelled discharges for the validation period.  
It can be seen that the model overestimates the amount of very small discharge values.

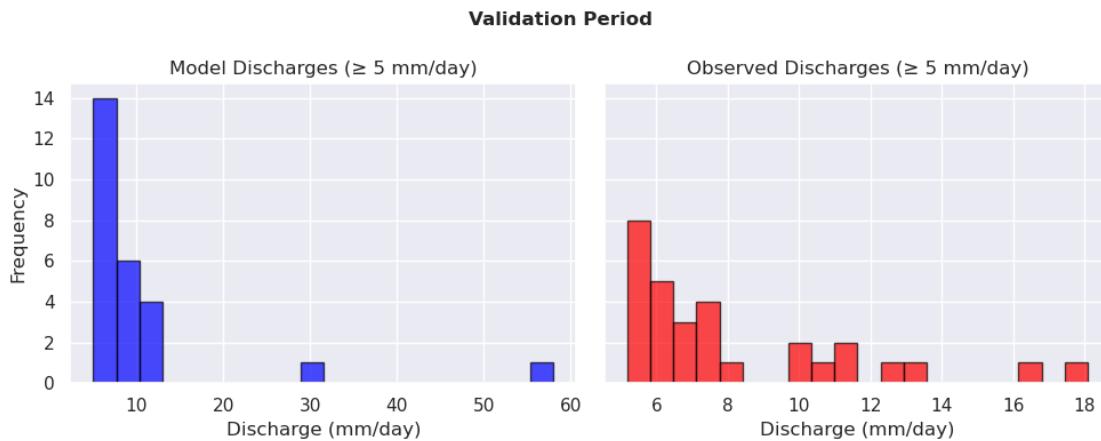


Figure I4: Histogram of the observed and modelled discharges for the validation period.  
Only values  $\geq 5$  mm/day are included to focus on relevant discharges.  
It can be seen that the distribution of modelled discharge is very different than the one of the observed discharge

## 10 Appendix II: AI Prompts

how do i import rich

To import and use Rich in Python, follow these steps:

---

### 1. Install Rich (if not already installed)

If you haven't installed Rich yet, run the following command in your terminal or command prompt:

bash

Kopieren Bewerken

```
pip install rich
```

---

### 2. Importing Rich

Once installed, you can import different parts of Rich depending on what you need.

#### Basic Import for Colored Printing

python

Kopieren Bewerken

```
from rich import print

print("[bold cyan]Hello, Rich![/bold cyan]")
```

hoe krijg ik waarden boven een bepaalde waarde uit een list in python

Je kunt eenvoudig waarden boven een bepaalde drempel uit een lijst filteren met **list comprehensions** of de **filter()**-functie in Python. Hier zijn een paar manieren:

### 1. List Comprehension (Meest gebruikt)

python

 Kopiëren  Bewerken

```
waarden = [10, 25, 30, 5, 60, 15]
drempel = 20

gefilterde_waarden = [x for x in waarden if x > drempel]
print(gefilterde_waarden) # Output: [25, 30, 60]
```

<class 'xarray.core.dataarray.DataArray'>  
dit is de type data, het is in dagen, ik wil graag de dag met de  
maximale discharge per jaar eruit krijgen

Je hebt een `xarray.DataArray` met dagelijkse waarden en je wilt voor elk jaar de dag met de maximale discharge bepalen. Dit kun je doen met `xarray` door de data te groeperen per jaar en vervolgens de maximale waarde te vinden. Hier is hoe je dat doet:

◆ **Stap-voor-stap aanpak:**

1. **Converteren van de tijdsdimensie** (als die nog geen datetime-objecten zijn).
2. **Groeperen per jaar.**
3. **Zoeken naar de dag met de maximale waarde** binnen elk jaar.

❖ **Codevoorbeeld:**

```
python                                     ⌂ Kopiëren    ⌂ Bewerken

import xarray as xr

# Voorbeeld DataArray met tijd als coördinaat
time = xr.cftime_range(start="2000-01-01", periods=365 * 3, freq="D") # 3 jaar
import numpy as np
data = xr.DataArray(np.random.rand(len(time)), coords=[time], dims=["time"], name="discharge")

# Zorg ervoor dat de tijdsdimensie een echte datetime is
data["time"] = data["time"].astype("datetime64[ns]")

# Groeperen per jaar en de dag met de maximale discharge vinden
max_per_year = data.groupby("time.year").idxmax("time")

print(max_per_year)
```

when i code this: `max_discharge = ds_forcing["Q"].groupby("time.year")`  
`print(max_discharge)`  
i get the result: <DataArrayGroupBy, grouped over 1 grouper(s), 40 groups in total:  
'year': 40/40 groups present with labels 1981, 1982, 1983, ..., 2019, 2020>  
But i want to see the maximum discharge values in these years

You're on the right track! To get the maximum discharge values for each year, you need to apply the `.max()` function after grouping by `"time.year"`. Try this:

python

Kopiëren Bewerken

```
max_discharge = ds_forcing["Q"].groupby("time.year").max()
print(max_discharge)
```

AttributeError: 'Timestamp' object has no attribute 'ndim'

It looks like `xarray` is returning a scalar `Timestamp` object rather than an `xarray.DataArray`, which causes issues inside `.map()`. Instead, we should ensure that we're working with an `xarray.DataArray` by using `.where()` and `dropna()`.

### Corrected Solution:

python

Kopiëren Bewerken

```
import xarray as xr

# Find the maximum discharge per year
max_discharge = ds_forcing["Q"].groupby("time.year").max()
print(max_discharge) # Max discharge values per year

# Find the corresponding dates of max discharge
def get_max_date(group):
    return group.time.where(group == group.max(), drop=True)

max_dates = ds_forcing["Q"].groupby("time.year").map(get_max_date)
print(max_dates) # Dates of maximum discharge per year
```

KeyError: 'Indexing with a boolean dask array is not allowed. This will result in a dask array of unknown shape. Such arrays are unsupported by Xarray. Please compute the indexer first using .compute()'

The error happens because `xarray` is using **Dask** internally, and **boolean indexing with Dask arrays isn't supported directly**. To fix this, we need to compute the indexer **before** applying `.where()`.

## ✓ Final Working Solution

```
python                                     ⌂ Kopieren ⌂ Bewerken

import xarray as xr

# Find the maximum discharge per year
max_discharge = ds_forcing["Q"].groupby("time.year").max()
print(max_discharge) # Maximum discharge values per year

# Find the corresponding dates of max discharge
def get_max_date(group):
    # Compute the max value and its corresponding index
    max_val = group.max()
    indexer = (group == max_val).compute() # Compute before applying
    return group.time.where(indexer, drop=True)

max_dates = ds_forcing["Q"].groupby("time.year").map(get_max_date)
print(max_dates) # Dates of maximum discharge per year
```

hoe plot ik een normale distributie in python met een mean van 530 en 5% prob van 510 en 95 van 550

Je wilt een normale distributie plotten met een gemiddelde (mean) van 530, waarbij 510 de 5e percentiel en 550 de 95e percentiel is.

#### Stap 1: Bereken de standaarddeviatie ( $\sigma$ )

In een normale verdeling geldt:

$$X = \mu + Z \cdot \sigma$$

waarbij:

- $\mu = 530$  (het gemiddelde),
- $X = 510$  bij 5% en  $X = 550$  bij 95%,
- $Z$ -waarden uit de standaard normale verdeling zijn -1.645 (5e percentiel) en 1.645 (95e percentiel).

We lossen op voor  $\sigma$ :

$$\sigma = \frac{510 - 530}{-1.645} = \frac{550 - 530}{1.645}$$

Nu kunnen we de distributie plotten in Python! ↗

#### Stap 2: Plot de normale verdeling

Hier is de code om dit te doen:

```
python
import numpy as np
import matplotlib.pyplot as plt
import scipy.stats as stats

# Gegeven parameters
mean = 530
p5 = 510 # 5e percentiel
p95 = 550 # 95e percentiel

# Bereken standaarddeviatie
z_5 = stats.norm.ppf(0.05) # -1.645
z_95 = stats.norm.ppf(0.95) # 1.645
std_dev = (p95 - mean) / z_95 # σ = (550 - 530) / 1.645

# X-waarden voor de plot
x = np.linspace(490, 570, 1000)
y = stats.norm.pdf(x, mean, std_dev)

# Plot de normale verdeling
plt.figure(figsize=(8, 5))
plt.plot(x, y, label="Normale verdeling", color='blue')

# Markeer de 5e en 95e percentielen
plt.axvline(p5, color='red', linestyle='dashed', label="5e percentiel (510)")
plt.axvline(p95, color='green', linestyle='dashed', label="95e percentiel (550)")

# Labels en titel
plt.xlabel("Waarde")
plt.ylabel("Kansdichtheid")
plt.title(f"Normale verdeling (μ={mean}, σ={std_dev:.2f})")
plt.legend()
plt.grid()

# Toon de plot
plt.show()
```

hoe kan ik in python bij deze normale verdeling nog iets plotten  
met op de x-as dezelfde eenheid maar op de y-as een andere

Om in Python een grafiek te maken waarbij je op de x-as dezelfde eenheid hebt als de normale verdeling, maar op de y-as een andere eenheid (bijvoorbeeld een andere schaal of een ander type data), kun je gebruik maken van een tweede y-as. Dit kun je doen met de `twinx()` functie van matplotlib.

Hier is een voorbeeld van hoe je een normale verdeling kunt plotten met een tweede y-as:

#### Code voorbeeld:

Stel je voor dat je naast de normale verdeling ook een andere functie (bijvoorbeeld een lineaire functie) wilt plotten met een andere y-as eenheid:

```
python                                ⌂ Kopieren ⌂ Bewerken

import numpy as np
import matplotlib.pyplot as plt
import scipy.stats as stats

# Gegeven parameters
mean = 530
p5 = 510 # 5e percentiel
p95 = 550 # 95e percentiel

# Bereken standaarddeviatie
z_5 = stats.norm.ppf(0.05) # -1.645
z_95 = stats.norm.ppf(0.95) # 1.645
std_dev = (p95 - mean) / z_95 # σ = (550 - 530) / 1.645

# X-waarden voor de plot
x = np.linspace(490, 570, 1000)

# Normale verdeling
y_normal = stats.norm.pdf(x, mean, std_dev)

# Stel een tweede y-as in
fig, ax1 = plt.subplots(figsize=(8, 5))

# Plot de normale verdeling op de eerste y-as
ax1.plot(x, y_normal, label="Normale verdeling", color='blue')
ax1.axvline(p5, color='red', linestyle='dashed', label="5e percentiel (510)")
ax1.axvline(p95, color='green', linestyle='dashed', label="95e percentiel (550)")
ax1.set_xlabel("Waarde")
ax1.set_ylabel("Kansdichtheid", color='blue')
ax1.tick_params(axis='y', labelcolor='blue')

# Maak een tweede y-as
ax2 = ax1.twinx()

# Stel een andere functie voor de tweede y-as in (bijvoorbeeld lineaire functie)
y_linear = 0.1 * x # Voorbeeld lineaire functie

# Plot de Lineaire functie op de tweede y-as
ax2.plot(x, y_linear, label="Lineaire functie", color='orange')
ax2.set_ylabel("Andere eenheid", color='orange')
ax2.tick_params(axis='y', labelcolor='orange')

# Titel en legenda
plt.title("Normale verdeling en lineaire functie")
ax1.legend(loc="upper left")
ax2.legend(loc="upper right")

# Toon de plot
plt.grid(True)
plt.show()
```

ik wil graag iets in python logaritmisch plotten

Je kunt een logaritmische plot maken in **Python** met **Matplotlib**. Afhankelijk van je behoeften, kun je een logaritmische schaal gebruiken op de **x-as**, **y-as**, of beide.

### Voorbeeld: Logaritmische plot

python

 Kopiëren  Bewerken

```
import numpy as np
import matplotlib.pyplot as plt

# Data genereren
x = np.linspace(1, 100, 100) # x-waarden van 1 tot 100
y = x ** 2 # Kwadratische relatie

# Plot maken
plt.figure(figsize=(8, 6))
plt.plot(x, y, label="y = x^2", color="blue")

# Logaritmische schaal instellen (y-as)
plt.yscale("log")

# Labels en titel
plt.xlabel("X-waarden")
plt.ylabel("Y-waarden (log)")
plt.title("Logaritmische plot van y = x^2")
plt.legend()
plt.grid(True, which="both", linestyle="--", linewidth=0.5)

# Plot tonen
plt.show()
```



ik wil graag uit een dataset met heel veel data per dag de herhalingstijden van deze waarden plotten



Je wilt een herhalingstijddanalyse uitvoeren en de herhalingstijden plotten. Dit wordt vaak gedaan met een exceedance probability plot of Gumbel-plot, waarbij de data worden geordend en de terugkeerperiodes worden berekend.

#### Stappen om herhalingstijden te plotten

1. Laad de dataset met tijdreeksen van afvoer, waterstanden, of een andere variabele.
2. Selecteer de piekwaarden per dag, maand of jaar (afhankelijk van de resolutie).
3. Sorteer de waarden van hoog naar laag.
4. Bepaal de herhalingstijden ( $T$ ) met de formule:

$$T = \frac{n+1}{m}$$

Waarbij:

- $n$  = totaal aantal waarnemingen
- $m$  = rangnummer van de gesorteerde waarde

5. Maak een logaritmische plot van herhalingstijd ( $T$ ) versus piekwaarde.

#### Voorbeeld in Python

```
python Kopieren Bewerken

import numpy as np
import pandas as pd
import matplotlib.pyplot as plt

# 1. Simuleer een dataset (vervang dit met je eigen data)
np.random.seed(42)
data = np.random.weibull(a=2, size=1000) * 100 # Simuleer piekwaarden

# 2. Sorteer de data van hoog naar laag
sorted_data = np.sort(data)[::-1] # Sorteer aflopend

# 3. Bereken de herhalingstijden
n = len(sorted_data)
rank = np.arange(1, n + 1) # Rangnummers
return_periods = (n + 1) / rank # Herhalingstijd

# 4. Plotten in Log-Log schaal
plt.figure(figsize=(8, 6))
plt.scatter(return_periods, sorted_data, color="blue", label="Waarden")
plt.xscale("log") # Logaritmische schaal voor herhalingstijd
plt.xlabel("Herhalingstijd (jaar)")
plt.ylabel("Piekwaarde")
plt.title("Herhalingstijddanalyse van piekwaarden")
plt.grid(True, which="both", linestyle="--", linewidth=0.5)
plt.legend()
plt.show()
```

2010-08-31T00:00:00Z hoe kan ik hiervan alleen de datum krijgen  
en niet de tijd



Je kunt de tijd verwijderen en alleen de datum behouden met **Pandas** of **Python's datetime module**.

### Oplossing in Pandas (voor een DataFrame)

Als je een **DataFrame** hebt met een kolom vol datums in dit formaat:

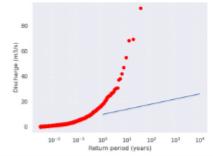
```
python                                ⬤ Kopiëren      ⚙ Bewerken

import pandas as pd

# Voorbeeld dataset
df = pd.DataFrame({'datetime': ['2010-08-31T00:00:00Z', '2020-05-15T12:30:45Z']})

# Converteer naar datetime en behoud alleen de datum
df['date'] = pd.to_datetime(df['datetime']).dt.date

print(df)
```



this is the result, i want to make a better fit

It looks like your current fit is not capturing the trend of the data properly. Since your data appears to be following a nonlinear trend (possibly exponential or power-law), here are a few approaches to improve the fit:

### 1. Try a Power-Law Fit (Log-Log Regression)

Since your x-axis is in log scale, a power-law relationship of the form:

$$y = a \cdot x^b$$

can be fitted by taking the logarithm of both sides:

$$\log(y) = \log(a) + b \log(x)$$

**Python Code:**

```
python
import numpy as np
import matplotlib.pyplot as plt
from scipy.optimize import curve_fit

# Sample data (replace with your actual data)
x = np.array([...]) # Return period (years)
y = np.array([...]) # Discharge (mA/s)

# Define power-law function
def power_law(x, a, b):
    return a * np.power(x, b)

# Fit the data
params, _ = curve_fit(power_law, x, y)
a, b = params

# Generate fitted curve
x_fit = np.logspace(np.log10(min(x)), np.log10(max(x)), 100)
y_fit = power_law(x_fit, a, b)

# Plot data and fit
plt.figure(figsize=(8, 5))
plt.scatter(x, y, color='red', label="Data")
plt.plot(x_fit, y_fit, color='blue', label=f"Fit: y = {a:.2e} x^{b:.2f}")
plt.grid(True, which="both", linestyle="--", linewidth=0.5)
plt.show()
```

TypeError: Invalid comparison between dtype=datetime64[ns] and str



The error "TypeError: Invalid comparison between dtype=datetime64[ns] and str" happens because `validation_start_time` and `validation_end_time` are likely strings, while `df.index` and `Q_pandas.index` are datetime objects (`datetime64[ns]`).

### 🔧 Solution: Convert to Datetime Format

Ensure that `validation_start_time` and `validation_end_time` are in the correct `datetime` format before filtering:

python

Kopieren Bewerken

```
import pandas as pd

# Convert validation_start_time and validation_end_time to datetime
validation_start_time = pd.to_datetime(validation_start_time)
validation_end_time = pd.to_datetime(validation_end_time)

# Check the index type (should be datetime)
print(df.index.dtype)
print(Q_pandas.index.dtype)
```

ValueError: The index must be timezone aware when indexing with a date string with a UTC offset

Deze fout betekent dat je een **tijdzone-bewuste** index moet hebben in je `model_output`, omdat je een datum met een **UTC-offset** gebruikt (bijv. `2025-03-20T00:00:00+00:00`).

### Oplossing 1: Zorg ervoor dat de index een tijdzone heeft

Je kunt de index van je `pd.Series` tijdzone-bewust maken met `.tz_localize()` of `.tz_convert()`:

python

Kopiëren

Bewerken

```
model_output.index = model_output.index.tz_localize("UTC")
```

💡 Gebruik `tz_convert()` als de index al een tijdzone heeft maar niet UTC is:

python

Kopiëren

Bewerken

```
model_output.index = model_output.index.tz_convert("UTC")
```

Daarna zou je deze code zonder problemen kunnen uitvoeren:

python

Kopiëren

Bewerken

```
model_output_filtered = model_output.loc[experiment_start_date:experiment_end_date]
```

ik maak mijn `x_fit` en `y_fit` hier, hoe kan ik dan een onzekerheid over de x-as plotten

Als je onzekerheid wilt toevoegen over de x-as (de `return_periods_years`, in jouw geval) terwijl je je `x_fit` en `y_fit` plot, kun je de onzekerheid over de x-waarden (de return perioden) berekenen en vervolgens een "band" rondom de x-as plotten, vergelijkbaar met hoe je de onzekerheid over de y-as hebt toegevoegd.

Hier is een aangepaste versie van je code met onzekerheid voor de x-as toegevoegd:

#### Stappen:

1. Bereken de onzekerheid voor de x-as (de return perioden):  
Dit kun je bijvoorbeeld doen door een procentuele onzekerheid (bijv. 22%) toe te voegen aan de `x_fit` waarden.
2. Plot de onzekerheid voor de x-as:  
Dit doe je door een gebied (zoals `plt.fill_between`) te maken tussen de geschatte `x_fit` waarden met een onzekerheidsband.

#### Code:

```
python Kopiëren Bewerken  
  
import numpy as np  
import matplotlib.pyplot as plt  
from scipy.optimize import curve_fit  
  
# Example data and power law function  
sorted_model_data = np.sort(model_data)[::-1] # Sorted data from high to low  
  
def power_law(x, a, b):  
    return a * x ** b  
  
def calculate_return_periods(sorted_model_data):  
    n = len(sorted_model_data)  
    rank = np.arange(1, n + 1)  
    return_periods_days = (n + 1) / rank  
    return_periods_years = return_periods_days / 365.25  
  
    # Only keep data for return periods >= 0  
    return_periods_years_subset = return_periods_years[return_periods_years >= 0]  
    sorted_data_subset = sorted_model_data[return_periods_years >= 0]  
  
    # Fit the power law function  
    params, _ = curve_fit(power_law, return_periods_years_subset, sorted_data_subset, maxfev=1000)  
    a_opt, b_opt = params  
  
    # Create a fit for the return period  
    x_fit = np.logspace(np.log10(min(return_periods_years_subset)), np.log10(max(return_periods_years_subset)), 100)  
    y_fit = power_law(x_fit, a_opt, b_opt)  
  
    return x_fit, y_fit, a_opt, b_opt, return_periods_years_subset, sorted_data_subset = calculate_return_periods()  
  
# Add uncertainty over the x-axis (return periods)  
uncertainty_percentage = 0.22 # 22% uncertainty  
x_uncertainty = uncertainty_percentage * x_fit # Calculate the uncertainty for the x-axis  
  
# Upper and lower bounds for the uncertainty on the x-axis  
x_upper_bound = x_fit + x_uncertainty  
x_lower_bound = x_fit - x_uncertainty  
  
# Plotting the results  
plt.plot(x_fit, y_fit, label='Fit Future Climate', color='blue')  
  
# Fill the uncertainty region for the x-axis  
plt.fill_between(x_fit, x_lower_bound, x_upper_bound, color='blue', alpha=0.3, label="22% Uncertainty")  
  
# Customize labels and legend  
plt.xscale('log')  
plt.yscale('log')  
plt.xlabel('Return Period (years)')  
plt.ylabel('Discharge')  
plt.legend()  
  
plt.show()
```

- \*\*SSP119 (sustainability):\*\* 697 years, with a model uncertainty ranging from 504 to 890 years
- \*\*SSP126 (sustainability):\*\* 238 years, with a model uncertainty ranging from 172 to 292 years
- \*\*SSP245 (middle of the road):\*\* 191 years, with a model uncertainty ranging from 138 to 234 years
- \*\*SSP370 (regional rivalry):\*\* 272 years, with a model uncertainty ranging from 197 to 334 years
- \*\*SSP585 (fossil-fueled development):\*\* 235 years, with a model uncertainty ranging from 170 to 288 years

Kan ik hier een tabel van maken in jupyter notebook in een markdowncell

8

Ja, je kunt een tabel maken in een **Markdown-cell** in Jupyter Notebook met de volgende Markdown-syntax:

markdown

Kopiëren Bewerken

<b>Climate Scenario (SSP)</b>	<b>Return Period (Years)</b>	<b>Model Uncertainty Range (Years)</b>
**SSP119 (sustainability)	697	504 – 890
**SSP126 (sustainability)	238	172 – 292
**SSP245 (middle of the road)	191	138 – 234
**SSP370 (regional rivalry)	272	197 – 334
**SSP585 (fossil-fueled development)	235	170 – 288

```

ik wil dit ook graag in 1 subplot:
plt.figure(figsize=(4, 3))
ds_forcing_filtered.plot(label="Observed discharge")
model_output_filtered.plot()
plt.legend()
plt.ylabel("Discharge (mm/d)")
plt.title("Observed and Modelled discharge validation period")
plt.text(0.5, -0.4, "Figure 8: Modelled and observed discharge validation period",
        ha="center", fontsize=9, transform=plt.gca().transAxes);
en

# Sort data from high to low
sorted_model_data = np.sort(model_output_filtered)[::-1] # Sorteer aflopend
ds_forcing_filtered = np.sort(ds_forcing_filtered)[::-1]

#calculate return periods
m = len(sorted_model_data)
rank = np.arange(1, m + 1)
return_periods_days_model = (m + 1) / rank
return_periods_years_model = return_periods_days_model / 365.25

m = len(ds_forcing_filtered)
rank = np.arange(1, m + 1)
return_periods_days_obs = (m + 1) / rank
return_periods_years_obs = return_periods_days_obs / 365.25
plt.figure(figsize=(4, 3))
plt.xscale('log')
plt.scatter(return_periods_years_model, sorted_model_data,
           label='Model', marker='.')
plt.scatter(return_periods_years_obs, ds_forcing_filtered,
           label='Observations', marker='.')
plt.legend()
plt.xlabel('Return period (years)')
plt.ylabel('Discharge (mm/day)')
plt.title("Observed and Modelled discharge validation period")
plt.text(0.5, -0.3, "Figure 8: Modelled and observed discharge validation period",
        ha="center", fontsize=9, transform=plt.gca().transAxes);

# Eerste subplot: Observed vs. Modelled discharge
fig, axes = plt.subplots(1, 2, figsize=(12, 5)) # Twee subplots naast elkaar
axes[0].plot(ds_forcing_filtered, label="Observed discharge")
axes[0].legend()
axes[0].set_ylabel("Discharge (mm/d)")
axes[0].set_title("Observed and Modelled Discharge Validation Period")
axes[0].text(0.5, -0.4, "Figure 8: Modelled and observed discharge validation period",
            ha="center", fontsize=9, transform=axes[0].transAxes)
axes[1].plot(model_output_filtered, label="Modelled discharge")
axes[1].legend()
axes[1].set_xlabel("Return period (years)")

# Tweede subplot: Return period plot
sorted_model_data = np.sort(model_output_filtered)[::-1] # Sorteer aflopend
sorted_ds_forcing_filtered = np.sort(ds_forcing_filtered)[::-1]
m_model = len(sorted_model_data)
rank_model = np.arange(1, m_model + 1)
return_periods_years_model = (m_model + 1) / rank_model / 365.25
m_obs = len(sorted_ds_forcing_filtered)
rank_obs = np.arange(1, m_obs + 1)
return_periods_years_obs = (m_obs + 1) / rank_obs / 365.25

axes[1].set_xscale('log')
axes[1].scatter(return_periods_years_model, sorted_model_data, label='Model', marker='.')
axes[1].scatter(return_periods_years_obs, sorted_ds_forcing_filtered, label='Observations', marker='.')
axes[1].legend()
axes[1].set_xlabel('Return period (years)')
axes[1].set_ylabel('Discharge (mm/day)')
axes[1].set_title('Observed and Modelled Discharge Distribution Validation Period')
axes[1].text(0.5, -0.3, "Figure 9: Modelled and observed discharge distribution validation period",
            ha="center", fontsize=9, transform=axes[1].transAxes)

plt.tight_layout()
plt.show()

```

sla aan het einde van je notebook per scenario de output data die je voor deze grafieken nodig hebt op een handige manier op en maak één notebook waarin je die data weer inleest en vervolgens alleen figuur 15 en 16 maakt. Dat scheelt veel model-runs. hoe doe ik dit

Het klinkt alsof je de output van verschillende berekeningen en grafieken in een **opslagbestand** wilt opslaan, zodat je later, zonder opnieuw de berekeningen te doen, de data kunt inladen en de grafieken opnieuw kunt genereren. Een handige manier om dit te doen is door gebruik te maken van **bestandsopslag** voor de outputdata, bijvoorbeeld door de gegevens op te slaan in een `.csv`, `.pickle` of `.json` bestand.

#### Stappenplan:

1. **Sla de outputdata op in een bestand.** Je kunt de data opslaan met bijvoorbeeld `pickle`, wat handig is voor Python-objecten (zoals arrays, dictionaries, enz.), of `csv` voor tabulaire gegevens.
2. **Laad de data later in een ander notebook.** In het nieuwe notebook kun je de opgeslagen data inladen en grafieken genereren op basis van die data zonder de berekeningen opnieuw te hoeven uitvoeren.

Laten we deze stappen verder toelichten:

---

#### Stap 1: Sla de data op in een bestand

##### 1.1 Gebruik `pickle` om je data op te slaan:



```
python                                ⌂ Kopiëren ⌂ Bewerken

import pickle

# Laten we zeggen dat je de parameters en resultaten wilt opslaan (bijv. 'x_fit', 'y_fit', enz.)
output_data = {
    'x_fit': x_fit,
    'y_fit': y_fit,
    'a_opt': a_opt,
    'b_opt': b_opt,
    'return_periods_years_scatter': return_periods_years_scatter,
    'sorted_data_scatter': sorted_data_scatter
}

# Sla de data op in een pickle-bestand
with open("output_data.pkl", "wb") as f:
    pickle.dump(output_data, f)

print("Data succesvol opgeslagen in output_data.pkl")
```

## Stap 2: Laad de opgeslagen data in een ander notebook en maak de grafieken

### 2.1 Laad de data in een nieuw notebook met `pickle`:

```
python  
  
import pickle  
  
# Laad de opgeslagen data  
with open("output_data.pkl", "rb") as f:  
    loaded_data = pickle.load(f)  
  
# Haal de benodigde gegevens uit de geladen data  
x_fit = loaded_data['x_fit']  
y_fit = loaded_data['y_fit']  
a_opt = loaded_data['a_opt']  
b_opt = loaded_data['b_opt']  
return_periods_years_scatter = loaded_data['return_periods_years_scatter']  
sorted_data_scatter = loaded_data['sorted_data_scatter']  
  
print("Data succesvol geladen!")
```

ik exporteer een jupiter notebook naar pdf



Ah, je exporteert een Jupyter Notebook naar PDF! In dat geval kunnen afbeeldingen die in de notebook zijn ingevoegd via Markdown of code-cellen (zoals met Matplotlib) soms problemen veroorzaken bij het exporteren naar PDF, vooral wanneer je de optie `--no input` gebruikt. Dit kan gebeuren omdat de afbeeldingen niet goed worden ingesloten in de geëxporteerde PDF.

Hier zijn enkele stappen die je kunt volgen om ervoor te zorgen dat afbeeldingen correct worden meegenomen bij het exporteren van een Jupyter Notebook naar PDF:

#### 1. Gebruik `nbconvert` met de juiste opties:

Wanneer je een Jupyter Notebook naar PDF exporteert, gebruik je waarschijnlijk de `nbconvert` tool. Als je de `--no input` optie gebruikt, betekent dit meestal dat je alleen de uitvoer van de cellen (zonder de code) wilt exporteren. Als de afbeeldingen niet goed worden opgenomen, kun je proberen een andere benadering te gebruiken.

**Oplossing:** Gebruik `nbconvert` zonder `--no input` of controleer of je afbeeldingen correct zijn ingevoegd in je notebook voordat je de export uitvoert.

Gebruik bijvoorbeeld deze opdracht om een notebook naar PDF te exporteren:

```
bash  
  
jupyter nbconvert --to pdf --output naam_van_je_pdf notebook_naam.ipynb
```

Dit zorgt ervoor dat de afbeeldingen in je notebook correct worden meegenomen in de PDF.