The Effect of Hydraulic and Meteorological Correlations on Dike Stability during Extreme Events

Author: Bagas Mahadika Abri Putra
Date: 20 August 2025

TU Delft

The Effect of Hydraulic and Meteorological Correlations on Dike Stability during Extreme Events

By

Bagas Mahadika A. P.

In partial fulfilment of the requirements for the degree of:

Master of Science

in Civil Engineering

at the Delft University of Technology,

to be defended publicly on Friday, 29 August 2025.

Supervisor:	dr.ir. C. (Cong) Mai Van	TU Delft
Thesis committee:	dr.ir. Olivier Hoes,	TU Delft
	Prof.dr.ir. P.H.A.J.M. van Gelder,	TU Delft
	ir. A. (Anton) van der Meer,	TU Delft
	dr.ir. M. (Mark) van der Krogt,	Deltares

TU Delft

Deltares

Summary

This study investigates how hydraulic and meteorological variables act together to affect dike stability, with emphasis on inner-slope failure during extreme hydrometeorological events. The variables influencing dike stability are first identified through a literature review, after which their correlations are examined and analysed using copula theory. The effect of these correlations is then assessed in more detail by applying a groundwater model developed with Pastas and performing slope stability calculations using the Bishop method.

The analysis focuses on the period from December 2023 to January 2024, when the Netherlands experienced high cumulative rainfall, elevated river discharges, and restricted outflow due to sea storm surges, leading to prolonged high water levels in the IJsselmeer–Markermeer system. The case study is a lakeside dike along Markermeer between Hoorn and Enkhuizen (about 17.8 km), where data availability enables detailed hydrological and geotechnical modelling. Two cross-sections (*raai_2* and *raai_3*) are instrumented with multiple observation wells from phreatic to deep sand layers. Inputs combine hourly lake levels from Krabbersgat Zuid and the nearby Drieban pumping station, hourly precipitation and daily evapotranspiration from Berkhout station, and local groundwater measurements from 10 Nov 2023 to 25 Feb 2025. These data are used to calibrate and validate Pastas groundwater models and to evaluate slope stability for representative hydraulic loading conditions.

The literature indicates that phreatic levels around Markermeer and IJsselmeer are governed by external hydraulic loads, climate, internal soil properties, dike geometry, and local lake dynamics. Using statistical analysis, copula modelling, and time-series groundwater simulations, this study examines how precipitation and lake level jointly influence the phreatic surface within the dike. Results show a moderate positive correlation between cumulative local precipitation and lake water levels. Copula models, particularly the BB8 family, capture asymmetric dependence between rainfall and water level, highlighting an increased likelihood of joint extremes.

For stability evaluation, observed groundwater data were first used as input to D-Stability to compute the factor of safety (FoS) over selected periods. This “dependent” case reflects the real, correlated relationship between precipitation and water level and shows a moderate negative correlation with FoS, meaning increases in either driver reduce stability. An “independent” case is then constructed by generating a new water-level series from the fitted bivariate copula using conditional sampling with rank-exact back-mapping, so that water level is statistically independent of precipitation while preserving the marginal (univariate) distributions. The Pastas model is re-fitted with this synthetic water-level series to produce new groundwater heads, and FoS is recomputed. Under the observed (dependent) case, peak external water levels coincided with prolonged high precipitation, producing higher phreatic levels and a lower minimum FoS (1.745). When the same marginals were used but the drivers were made independent, peak water levels were lower and the minimum FoS improved (1.768). By evaluating the correlation, the dependent case shows stronger negative correlation for both precipitation (-0.67) and water level (-0.49) versus FoS. In the case of independent variable, the correlation between water level and FoS strengthened to −0.85, while the correlation between precipitation and FoS weakened to −0.18.

Based on the previous results, it can be concluded that during the wet season, the correlation between precipitation and water level leads to a more conservative outcome, expressed as a lower factor of safety (FoS) for dike stability. This finding is consistent with real-world conditions, where periods of higher rainfall typically occur together with higher local lake water levels caused by runoff from around the lake, direct rainfall itself, and polder drainage pumping into the lake. This conclusion is based on the assumption that the water level dataset used in this study represents local water level observations, where in reality the actual local water level at this specific dike section may differ slightly depending on wind magnitude and direction. The main recommendation for dike assessment based on this study is that the correlation between precipitation and water level should be explicitly considered in stability analyses, since neglecting this dependence may underestimate phreatic levels within the dike and result in a less conservative estimate of the factor of safety.

Acknowledgements

This research was a collaboration between the Hydraulic Structures and Flood Risk Sections, Faculty of Civil Engineering and Geosciences, TU Delft, and the Geotechnical Engineering Risk and Reliability team at Deltares. I gratefully acknowledge the support of the Ministry of Public Works (Kementerian Pekerjaan Umum), which granted me permission to pursue my studies, and of LPDP as the scholarship provider—without which this project would not have been possible.

I would like to express my sincerest gratitude to my supervisors. To **Cong Mai Van**, thank you for your unwavering support throughout this project, beginning with my research internship. Your kindness and trust motivated me to give my best effort despite setbacks. To **Anton van der Meer**, my daily supervisor, thank you for supporting me at every step and for looking out for me—especially during a personal difficulty. You consistently encouraged me, monitored my progress, and provided highly constructive feedback. To **Mark van der Krogt**, thank you for entrusting me with this project. Even before starting my Master's at TU Delft, I wanted to learn how dike safety is approached in the Netherlands and to bring that knowledge back to Indonesia. The opportunity to carry out this project has been invaluable, broadening my understanding and inspiring my future work in Indonesia.

To **Pieter van Gelder** and **Olivier Hoes,** thank you for accepting my invitation as my thesis committee member and for the expertise that improved this thesis beyond my expectations. I also thank **Teun Hanzen** from HHNK for providing data for the Markermeer dike section, specifically dike geometry, soil parameters, and groundwater measurements.

Special thanks to my parents, who have always supported me and never stopped believing in me. Finally, heartfelt thanks to my friends in the Netherlands for being with me through every high and low.

Table of Contents

Chapter 1. Introduction

1.1. Research Context

In December 2023 and January 2024, rivers and lakes in Netherlands experienced relatively high-water levels (Expertise Netwerk Waterveiligheid, 2024), leading to increased alertness from water authorities and significant media attention. According to Expertise Netwerk Waterveiligheid (ENW) Report, there are three main events that caused significant load on Dutch dikes during this period: sea storm surge, high water level on the main and regional river, and prolonged rainfall. Based on observational data from Lobith, the peak water level in the Rhine River was relatively low at around 1432 cm NAP (Figure 1.1), which corresponds to a return period of approximately 5 years. The peak discharge at the end of December 2023 was about 7500 m^3/s (Figure 1.2), which corresponds to a return period of 4 years. In comparison, the Overijssel Vecht River experienced a peak discharge of approximately 400 m^3/s, which corresponds to a return period of 50 years. Additionally, according to GRADE simulations at the Lobith station, the average discharge over 51 days for the Rhine River was estimated at 5102 m^3/s, corresponding to a return period of 40 years (Expertise Netwerk Waterveiligheid, 2024). Although the flood event did not cause substantial problems, it emphasized the critical role of flood defences, when faced with multiple extreme events at the same time.

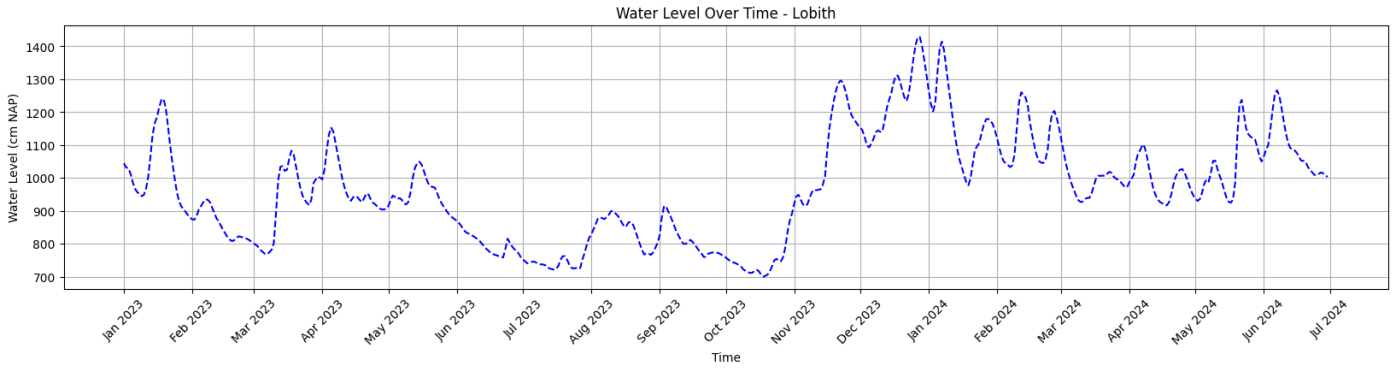


Figure 1.1 Water Level Measurement at Lobith (from Rijkswaterstaat Waterinfo)

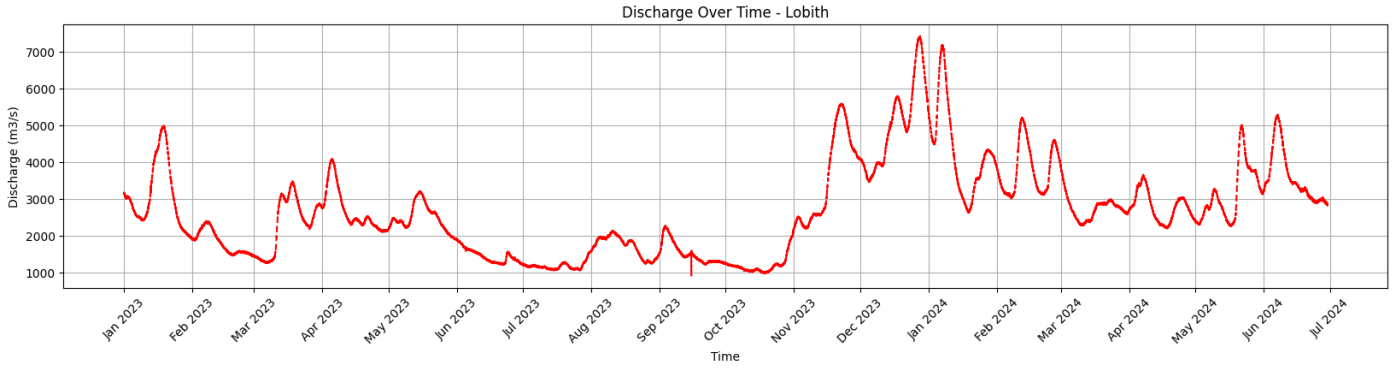


Figure 1.2 Discharge Measurement at Lobith (from Rijkswaterstaat Waterinfo)

Dike safety standards in the Netherlands are assessed based on the probability of failure of the entire flood defence system, defined as the loss of the dike's ability to retain water and protect the surrounding area. Various failure mechanisms must be considered during the assessment, one of which is dike inner slope instability. This failure mechanism occurs when water infiltrates the dike body due to groundwater recharge, local rainfall, or high river/lake water levels on the outer side of the dike.

During the December 2023 – January 2024 period, exceptionally high amount of rainfall in Rhine basin led to increased river discharge. According to daily rainfall record in Heino (Figure 1.3), while no extreme peak observed during November 2023 - January 2024, the prolonged rain with moderate to high intensity plus sharp peak at the end of December 2023 caused high river discharge for longer than usual duration. At the same time, multiple storms raised sea levels outside the Afsluitdijk (shown in Figure 1.4, where increase of sea level observed during December 2023), limiting discharge capacity. This resulted in sustained high-water levels in both the IJssel River

and IJssel Lake (as shown in Figure 1.5, where increased water level reaching beyond 60 cm+NAP is observed starting from middle of December 2023 and decreasing at first half of January 2024). The simultaneous occurrence of high water levels and high precipitation increased pore water pressure, represented by the phreatic line and head lines within the dike, potentially leading to slope instability.

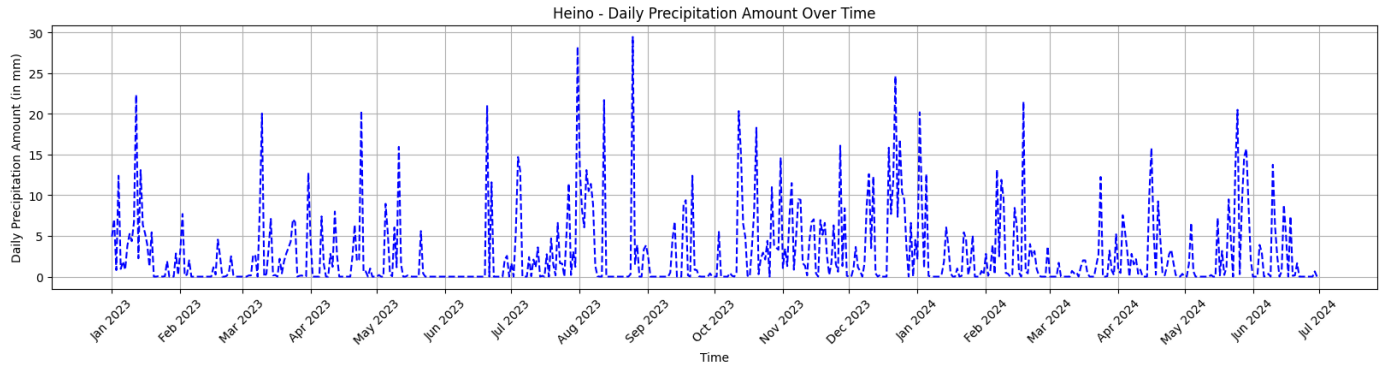


Figure 1.3 Daily Rain Intensity at Heino Rain Station (from Rijkswaterstaat Waterinfo)

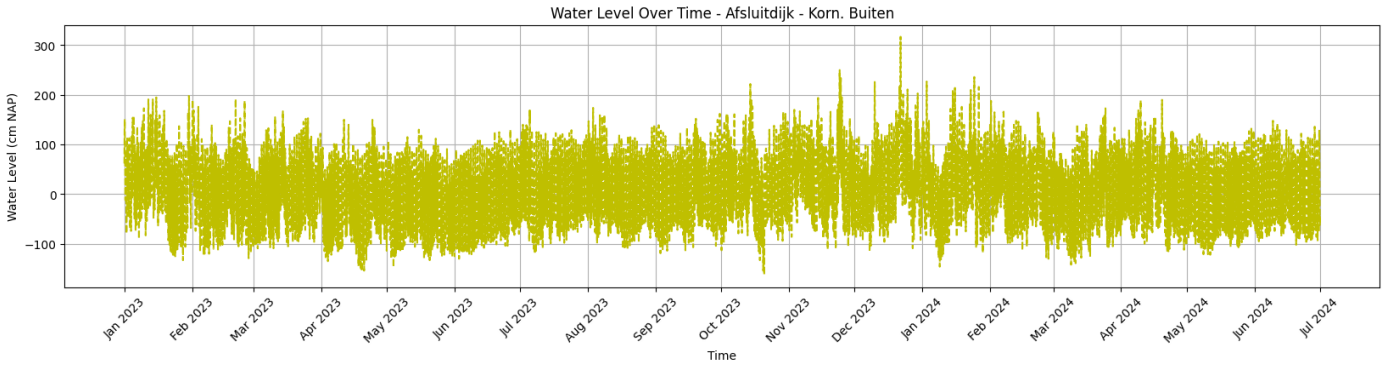


Figure 1.4 Water Level Measurement at Kornwerderzand buiten (Afsluitdijk – Wadden Sea side) (from Rijkswaterstaat Waterinfo)

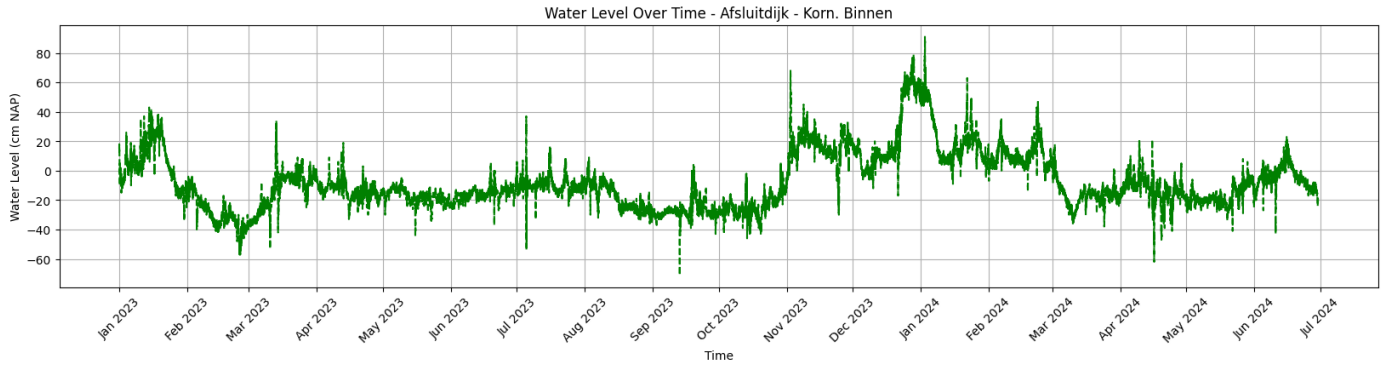


Figure 1.5 Water Level Measurement at Kornwerderzand binnen (Afsluitdijk – IJsselmeer side) (from Rijkswaterstaat Waterinfo)

Although no dike breaches were observed during this period, concerns have arisen about the combination of loading scenarios—specifically, the compound event between river/lake water levels and local precipitation intensity—that contributed to increased pore water pressure and potential dike instability. Evaluating these load combinations in the context of the December 2023 – January 2024 flood event can provide valuable insights into the event's extremity based on failure probability estimations.

1.2. Research Scope

This research will focus on a dike section built around Markermeer, which influenced by lake water level, local precipitation and evapotranspiration, IJssel river discharge, storm surge at the Afsluitdijk, and polder water level. The dike section that will be used is chosen according to data availability. Also, between multiple dike failure mechanisms, inner slope instability case of failure probability is assessed in this research.

1.3. Research Objective and Research Questions

From the research context, the following main research objective is formulated: “To assess how the Effect of Hydraulical and Meteorological Correlations towards Dike Stability during Extreme Events.”

And the following sub research questions are proposed to answer the research objective:

1. What are the variables influencing the phreatic line, which also influence dike stability in dike section, in Markermeer and IJsselmeer region?
2. Is there any correlation between hydraulic and meteorological variables that influence dike stability?
3. Based on the proposed stochastic variables, what is the effect of correlation between water level and precipitation for the phreatic level in extreme conditions for dike stability in Markermeer during the wet season (Dec 2023 – Jan 2024)?

1.4. General Approach

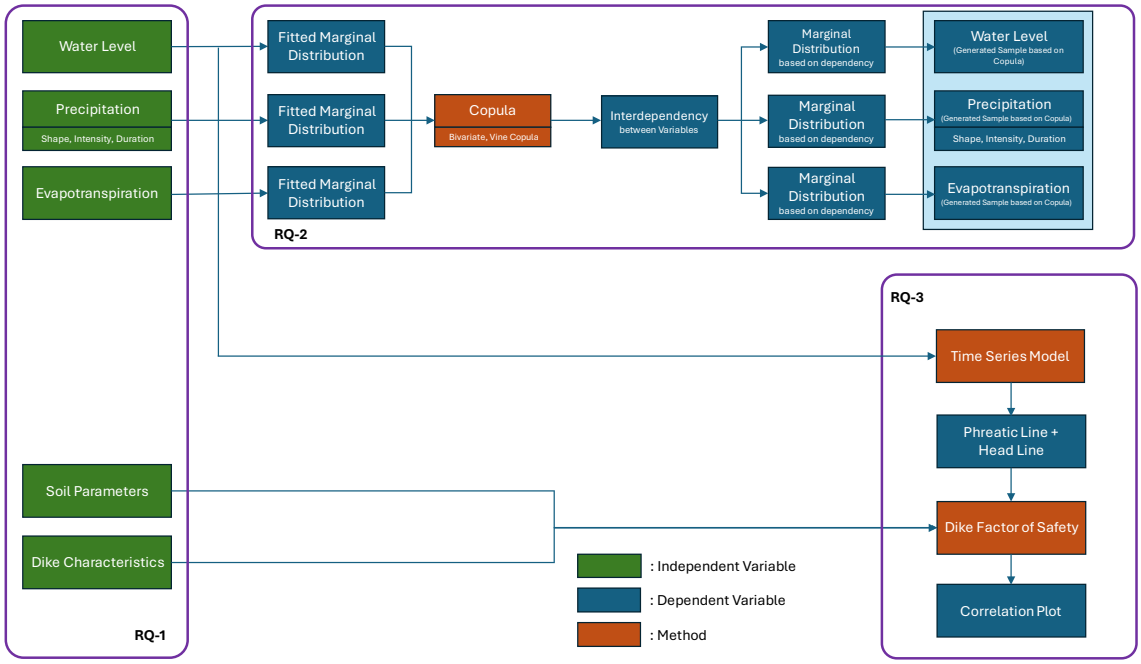


Figure 1.6 Theoretical Framework

To answer the previous Research Question, theoretical framework is composed (Figure 1.6). First, few variables that influence phreatic level is proposed, focused on lake water level, local precipitation and evapotranspiration, with soil parameters and dike characteristics to support calculation of Factor of Safety. Then, to answer Research Question 2 the interdependency will be modelled using Copula. Research Question 3 will be answered using time series model to create and simulate Phreatic Line, which then used as an input to calculate dike factor of safety. The result then compared to determine return period plot, which will be used to answer the main research objective to determine how extreme the high-water event was.

1.5. Outline

After the Introduction in Chapter 1, the theoretical background and analysis of existing relevant research are presented in Chapter 2 – Literature Study. The proposed theoretical framework is then detailed and explained further in Chapter 3 – Methodology and Approach. The location used for this study is described in Chapter 4 – Case Study and Datasets. The results of the analysis are shown in Chapter 5 – Result, and further discussion, including limitations and assumptions, is provided in Chapter 6 – Discussion. Lastly, the conclusion and recommendations based on this study are presented in Chapter 7 – Conclusion and Recommendation.

Chapter 2. Literature Study

2.1. Hydrology and Meteorological Variables

The stability of dike slopes depends on the condition of the phreatic line, which is defined as the groundwater table within the dike. This phreatic line is crucial because its location, along with pore pressures, directly impacts the inner slope macro stability of a dike. An increase in groundwater pressure, commonly observed during flooding or high-water events, reduces the effective stress in the soil and can lead to lower shear forces and increased driving forces, ultimately threatening the dike's structural integrity and potentially causing slope instability. (Hassan & Ismail, 2018)

The phreatic level inside a dike, and its interactions with surrounding water bodies like the IJssel Lake and River system, are governed by a combination of external hydraulic loads, internal dike material properties, and climatic conditions.

2.1.1. Water Level Fluctuations and Hydraulic Loads

The most significant influence on the phreatic line's position comes from the water levels on both the inner and outer sides of the dike. These water levels dictate the entry and exit points of the phreatic line within the dike body. During high water waves, there is a delay in the response of pore pressures inside the dike, a phenomenon known as the "lingering effect," which is particularly pronounced in low-permeability clay dikes.

Studies have shown that during high water waves, the estimated phreatic line schematizations can be exceeded at the inner toe of homogeneous clay dikes. This suggests that current schematic estimations may underestimate the actual phreatic line in these areas, especially when sub-soil thickness is thinner (e.g., 1 meter vs. 4 meters), with exceedance occurring as early as 12 to 20 days into a normative water level simulation. For sandy dikes, however, these schematic estimations are generally not exceeded. (Woerkom, 2023)

The rate of water level fluctuation can affect the changing rate of the water table within piezometer tubes. While some models reasonably estimate the phreatic line under uniform lowering conditions, they may underestimate it when the lowering rate is very large. The peak hydraulic heads in canal dikes, for instance, are influenced by various weather events and the duration of precipitation. (Yan et al., 2010)

2.1.2. Precipitation and Evapotranspiration

Hydrological processes like precipitation and evapotranspiration are key determinants of water inflow and outflow in a dike, thus controlling the phreatic line's height.

Precipitation

Not all precipitation infiltrates directly into the dike body; a considerable portion of rainfall can result in surface runoff, particularly during intense or prolonged events. However, the water that does infiltrate contributes to raising the groundwater table inside the dike, leading to an elevated phreatic line. This effect tends to be most noticeable at the inner and outer toes of the dike, where water accumulates and infiltration pathways are concentrated.

The extent of infiltration is influenced significantly by the composition of the dike’s surface layers. Dikes with less permeable surfaces, such as asphalt covers, dense clay layers, or compacted soils, tend to resist water penetration, limiting infiltration and reducing the potential for a rise in the phreatic level. Conversely, dikes with more permeable surfaces such as sand layers or grass covers allow rainwater to infiltrate more readily. Grass-covered surfaces, in particular, can facilitate infiltration through porous soil structures and preferential flow paths created by root systems or small cracks in the surface.

The properties of the unsaturated zone above the phreatic line also play a critical role in determining how quickly infiltrated rainfall reaches the groundwater table. The pore size distribution and capillary characteristics of this zone control the rate of downward water movement. Coarse-grained materials enable faster percolation, while fine-grained soils slow infiltration due to higher water retention in small pores.

During periods of heavy or prolonged rainfall, the infiltration capacity of the dike material may be exceeded, causing a significant rise in the phreatic level within the dike. According to *Technische Rapport Waterspanningen bij Dijklichamen* (TRWD), under extreme rainfall conditions, the phreatic line can rise by approximately 0.5 to 1.0 meters, particularly between the inner toe and the inner crest of the dike. Simulations further show that in such conditions, the phreatic surface near the inner crest can be up to 1 meter higher compared to dry periods, reflecting the combined effect of increased infiltration and lateral groundwater flow within the dike body. (Denkers, 2021)

Evapotranspiration

The interaction between soil, vegetation, and atmospheric conditions has been identified as a critical factor affecting the hydraulic and mechanical behavior of dike structures. This soil–vegetation–atmosphere (SVA) system influences not only the moisture content within the dike body but also the overall water balance, flow patterns, and pore water pressure distribution. In particular, processes such as evapotranspiration and drought-induced loss of moisture significantly alter the hydraulic response of dikes during varying climatic conditions. (Jamalinia et al., 2021)

Extended periods of drought lead to progressive drying of the dike's surface layers through evaporation and transpiration. As moisture is lost from fine-grained soils, shrinkage occurs, leading to the formation of surface and subsurface cracks. These shrinkage cracks represent a significant structural weakening of the dike material, creating discontinuities that compromise cohesion and shear strength. Moreover, these cracks function as preferential flow paths during subsequent rainfall or high-water events, effectively bypassing the natural filtration capacity of the unsaturated zone. The presence of such cracks increases the hydraulic conductivity of the dike body, allowing infiltrating water to move more rapidly and reach deeper soil layers where it contributes to elevated pore water pressures. (Jamalinia et al., 2021)

The development of high pore pressures within the dike due to rapid infiltration through cracks can substantially reduce the effective stress in the soil structure. This reduction in effective stress lowers the soil's shear strength and, consequently, the overall factor of safety of the slope, increasing the risk of slope failure or internal erosion processes. According to Wang et al. (2022), the hydraulic response of cracked soil is markedly different from that of intact structures, with infiltration and seepage patterns being governed by the presence, orientation, and extent of the crack network. As crack network grew, the water flow more rapidly in the path formed by the crack and infiltrate the subsoil quicker compared to less permeable soil material.

In addition to its impact on hydraulic behavior, drought-induced cracking also affects the vegetation cover on the dike surface. Reduced soil moisture and physical disruption of the root zone from cracking typically result in a decline in vegetation density. This is reflected in a lower Leaf Area Index (LAI), which indicates diminished canopy coverage and a reduced capacity for evapotranspiration. As vegetation thins, the storage capacity of the root zone decreases, leading to further soil drying and potentially accelerating crack development. (Sammis et al., 1986)

2.1.3. Internal Dike Material Properties (Soil Layers)

The behavior of groundwater within dikes is strongly influenced by the permeability and hydraulic conductivity of the soil layers that compose the structure. Permeability governs the ease with which water moves through the soil, and it plays a critical role in determining the shape and position of the phreatic line, particularly during fluctuations in river or groundwater levels. Clay, often characterized as a low-permeability material, typically exhibits anisotropic hydraulic conductivity, with horizontal conductivity generally exceeding vertical conductivity due to its laminated structure and depositional environment (Clennell et al., 1999). This anisotropy becomes especially important in understanding the hydraulic response of a dike. A low hydraulic conductivity limits both infiltration and drainage rates, causing the phreatic line to remain relatively high and stable throughout the year. This slow response time is a critical consideration during flood events, where the rapid rise of river water may not be immediately matched by internal pore water adjustments. In such cases, horizontal hydraulic conductivity becomes the dominant factor influencing the lateral flow of water, whereas vertical conductivity is more relevant in scenarios involving direct infiltration from precipitation. (Adji et al., 2023)

Another important parameter affecting groundwater dynamics within dikes is the volumetric specific storage capacity. This property determines the amount of water a soil can store or release in response to changes in hydraulic head. Soils with lower storage capacity tend to exhibit a more stable and elevated phreatic line, as they have a reduced ability to absorb transient fluctuations in water levels (Pešková & Štibinger, 2015). In combination with hydraulic conductivity, specific storage capacity governs the rate at which the phreatic surface migrates through the dike body in response to external hydraulic stimuli, such as high river stages or heavy rainfall. A reduced storage capacity, therefore, can lead to prolonged saturation within the dike, contributing to elevated pore pressures and reduced effective stress, particularly during extended high-water periods.

The internal structure of a dike further complicates the groundwater behavior due to inherent soil heterogeneity. Unlike engineered embankments constructed with uniform materials, many dikes have evolved over time through a series of repairs and reinforcements using locally available soils. This historical layering and variability lead to significant heterogeneity in both hydraulic and mechanical properties across the dike body. Such variability must be accounted for in stability assessments, as ignoring it can result in misleading estimates of dike safety. For instance, the combination of a highly permeable dike fill material overlying a relatively impermeable foundation or subsurface layer can restrict the vertical dissipation of pore pressures. This configuration can trap water within the dike body, particularly after high river levels recede, and lead to sustained high pore water pressures. These conditions pose a serious risk to slope stability, especially on the landward side of the dike, where delayed dissipation of pore pressure can result in slope failure during rapid drawdown events.

2.1.4. Dike Geometry and Sub-soil Conditions

The geometry of the dike, including its slope, height, and width, in conjunction with the characteristics of the underlying sub-soil, plays a significant role in controlling the behavior of the phreatic line and, by extension, the overall stability of the dike structure. These physical attributes influence both the hydraulic response of the dike to external water level changes and the internal distribution of pore pressures within the dike body.

The steepness of the dike slope is particularly critical in determining slope stability. Steeper dikes tend to facilitate more efficient surface and internal drainage, resulting in generally lower phreatic levels. This reduced saturation in the upper part of the slope contributes to improved stability under both static and dynamic loading conditions. In contrast, flatter slopes may retain more moisture, potentially leading to higher pore pressures and reduced shear strength, thereby increasing the risk of slope instability. The height of the dike also plays a supporting role, as it governs the hydraulic gradient and influences the total volume of water that can be stored or transmitted within the structure. (Denkers, 2021)

Dike width is another influential geometric factor. Wider dikes typically exhibit higher phreatic lines due to their greater internal volume and storage capacity. This larger volume results in slower drainage rates and a more gradual response to external hydraulic forcing, such as river water level fluctuations or rainfall infiltration. As a result, wide dikes may experience prolonged periods of saturation, especially under sustained high-water conditions, which can elevate internal pore pressures and impact slope stability.

In addition to geometry, the sub-soil conditions beneath the dike exert a substantial influence on groundwater dynamics. The behavior of the phreatic line can vary depending on both the composition and thickness of the sub-soil layers. For clay-dominated sub-soils, greater thickness tends to result in a higher and more variable phreatic line. The low permeability of clay restricts vertical drainage, causing water to accumulate and persist within the dike structure. Over time, this can lead to significant seasonal variations in saturation and pore pressure. In contrast, for sand dikes, which generally exhibit higher permeability, the influence of underlying clay sub-soil thickness on the phreatic line appears to be negligible. The more permeable nature of sandy soils allows for rapid infiltration and drainage, resulting in a more dynamic and responsive phreatic behavior. (Nishiie et al., 2019)

2.1.5. Markermeer and Ijsselmeer Condition

The phreatic levels of dikes surrounding the Markermeer and IJsselmeer are strongly affected by the water levels of these adjacent lakes. In the Markermeer, precipitation serves as the primary source of water input,

supplemented by inflows through sluices and limited river discharge. In contrast, the IJsselmeer primarily receives its water from river discharges, particularly from the IJssel, Meppelerdiep, Eem, and Laak rivers, which together contribute approximately 85% of its total inflow. Regarding water outflow, the Markermeer predominantly loses water through evaporation and sluice operations. Meanwhile, the IJsselmeer primarily discharges water via gravity flow towards the North Sea, representing its main mechanism for water level regulation. (Goossens, 2022)

Other than water source itself, wind also plays a significant role in shaping the hydrodynamics and water level fluctuations within large, shallow lakes such as the Markermeer and IJsselmeer. Due to their extensive surface areas and shallow depths, these lakes are particularly sensitive to wind-driven forces. Wind set-up, also known as wind skewness or storm effect, occurs when persistent winds push water toward one side of a lake, raising water levels on the windward side and lowering them on the leeward side, creating a sloping water surface. The strength of this effect depends on wind speed, water depth, and fetch length—the distance over which wind blows across the water. In the IJsselmeer, strong south-westerly winds can raise water levels by over 1.5 meters on the windward side while lowering them by about 1 meter on the opposite side. In the Markermeer, easterly winds can cause level differences of up to 15 centimeters. This wind-driven effect weakens with increasing water depth. During storms, water levels can rise and fall rapidly, sometimes changing by over a meter in a short period, while waves of up to 1.5 meters can develop quickly due to the lakes' large fetch. (Rijkswaterstaat, 2025; Goossens, 2022)

2.2. Interdependence and Correlation Measures

Interdependence describes relationship between two or more variables, and how a change in variable(s) affect other variable(s). One approach that can be used to determine interdependence is by calculating correlation. Correlation is a statistical measure that quantifies the degree of dependency between two variables. A positive correlation indicates that an increase in one variable is associated with an increase in the other, whereas a negative correlation signifies that an increase in one variable corresponds to a decrease in the other. A value close to zero suggests little to no linear relationship between the variables. (Czado, 2019).

To assess interdependency between each related variable, correlation between variables will be calculated. There are three measures that will be applied: Pearson's ρ; Spearman's ρ_s; Kendall's τ.

Table 2.1 Correlation Measures Formula

Correlation measures	Formula	Note
Pearson's	$\hat{\rho}(X_1, X_2) = \frac{\sum_{i=1}^{n}(x_{i1} - \overline{x_1})(x_{i2} - \overline{x_2})}{\sqrt{\sum_{i=1}^{n}(x_{i1} - \overline{x_1})^2}\sqrt{\sum_{i=1}^{n}(x_{i2} - \overline{x_2})^2}}$	Range of interval [-1,1] For random sample x_{i1}, x_{i2} with size n and $i = 1, \dots, n$ $\overline{x_1} = \frac{1}{n}\sum_{i=1}^{n} x_{i1}$ $\overline{x_2} = \frac{1}{n}\sum_{i=1}^{n} x_{i2}$
Spearman's	$\hat{\rho}_s(X_1, X_2) = \frac{\sum_{i=1}^{n}(r_{i1} - \overline{r_1})(r_{i2} - \overline{r_2})}{\sqrt{\sum_{i=1}^{n}(r_{i1} - \overline{r_1})^2}\sqrt{\sum_{i=1}^{n}(r_{i2} - \overline{r_2})^2}}$	Range of interval [-1,1] For random sample x_{i1}, x_{i2} with size n; ranks r_{i1}, r_{i2} and $i = 1, \dots, n$ $\overline{r_1} = \frac{1}{n}\sum_{i=1}^{n} r_{i1}$ $\overline{r_2} = \frac{1}{n}\sum_{i=1}^{n} r_{i2}$
Kendall's	$\hat{\tau}(X_1, X_2) = \frac{N_c - N_d}{\sqrt{N_c + N_d + N_1} \times \sqrt{N_c + N_d + N_2}}$	Range of interval [-1,1] For random sample x_{i1}, x_{i2} with size n and $i = 1, \dots, n$ N_c = concordant pairs N_d = discordant pairs $N_{1,2}$ = number of extra pairs for $x_{1,2}$

Pearson's rho describes whether the two variables linearly correlated as long as the dataset follows normal distribution. If nonlinear transformation is applied to the variables, the Pearson's rho does not stay the same.

Value of -1 describe perfectly decreasing linear relationship and +1 for perfectly increasing linear relationship. Value of 0 means no linear relationship.

Meanwhile, Spearman's rho and Kendall's tau are known as rank-based correlation measure. Both methods use order of the value to determine the relation between two variables instead of the value itself. Due to this characteristic, the value does not depend on distribution of each variable (marginal distribution) and only depend on the parameters of its copula distribution.

2.3. Copula Theory and Surrogate Time Series

2.3.1. Copula Theory

While correlation can quantify the strength and direction of the relationship between two variables, correlation by itself is insufficient for capturing complex dependencies, especially in non-linear relationships and extreme events in the tails of distributions. Therefore, joint distribution function is introduced in the form of copula. Following Sklar's Theorem, copula can be used to characterize the dependency between non-dimensional variables, or to build multivariate distribution (Czado, 2019).

According to Sklar's Theorem, for a set of random variables $X_1, X_2, \dots, X_n$ with each variable has their own marginal distribution $F_{X_1}(x), F_{X_2}(x), \dots., F_{X_n}(x)$, then the multivariate joint cumulative distribution function (CDF) can be represented as copula C:

$$F(x_1, x_2, \dots, x_n) = C\left(F_{X_1}(x_1), F_{X_2}(x_2), \dots., F_{X_n}(x_n)\right) \quad (4-1)$$

and the joint probability distribution function (PDF) written as:

$$f(x_1, x_2, \dots, x_n) = c\left(F_{X_1}(x_1), F_{X_2}(x_2), \dots., F_{X_n}(x_n)\right) f_1(x_1) f_2(x_2) \dots f_n(x_n) \quad (4-2)$$

where c is copula density and f is the PDF of X.

To determine absolute limits of dependence between random variables, Fréchet-Hoeffding bound theorem is applied:

For any bivariate copula $C(u, v)$ with $u, v \in [0,1]$, the following inequation should be satisfied:

$$max(0, u + v - 1) \leq C(u, v) \leq min(u, v) \quad (4-3)$$

The lower bound represent perfect negative dependence (when one variable increase, other variable decrease) and the upper bound represent perfect positive dependence (both variables increase or decrease together).

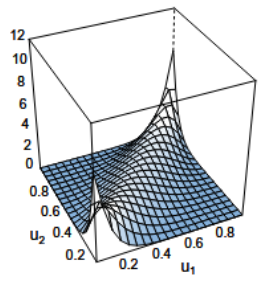
Here is the bivariate copula family used in this study, following the availability from *pyvinecopulib* library:

Table 2.2 Bivariate Family (adapted from Czado, 2019 and Joe, 1997)

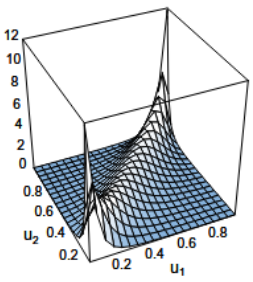
Bivariate Family	Copula CDF	Parameters
Gaussian	$C(u_1, u_2; \rho) = \Phi_2(\Phi^{-1}(u_1), \Phi^{-1}(u_2); \rho)$	$\Phi_2(x, y; \rho)$ = CDF of standard bivariate normal distribution with correlation ρ $\Phi^{-1}(\cdot)$ = inverse CDF of standard normal distribution **No** tail dependence
Student's t	$C(u_1, u_2; \nu, \rho) = \int_0^{u_1}\int_0^{u_2} \frac{t(T_\nu^{-1}(v_1), T_\nu^{-1}(v_2); \nu, \rho)}{t_\nu(T_\nu^{-1}(v_1)) t_\nu(T_\nu^{-1}(v_2))} dv_1 dv_2 = \int_{-\infty}^{b_1}\int_{-\infty}^{b_2} t(x_1, x_2; \nu, \rho) dx_1 dx_2$	$t(x, y; \nu, \rho)$ = CDF of standard bivariate Student's t-distribution with correlation ρ and degrees of freedom ν $T_\nu^{-1}(\cdot)$ = inverse CDF of the standard univariate Student's t-distribution **Upper and Lower** tail dependence with coefficient $2t_\nu\left(-\sqrt{\nu+1}\sqrt{\frac{1-\rho}{1+\rho}}\right)$
Clayton	$C(u_1, u_2; \delta) = \left(u_1^{-\delta} + u_2^{-\delta} - 1\right)^{-1/\delta}$	$-1 < \delta < \infty$

		$\delta \to 0$ = independence copula where $C(u_1, u_2) = u_1 \cdot u_2$ $\delta \to \infty$ = positive dependence $\delta \to -1$ = negative dependence **Lower** tail dependence with coefficient $2^{-1/\delta}$
Gumbel	$C(u_1, u_2; \delta) = \exp\left(-\left((-\ln u_1)^{\delta} + (-\ln u_2)^{\delta}\right)^{1/\delta}\right)$	$\boldsymbol{\delta \geq 1}$ $\delta \to \infty$ = positive dependence $\delta = 1$ = independence **Upper** tail dependence with coefficient $2 - 2^{1/\delta}$
Frank	$C(u_1, u_2; \delta) = -\frac{1}{\delta}\ln\left(1 + \frac{(e^{-\delta u_1} - 1)(e^{-\delta u_2} - 1)}{e^{-\delta} - 1}\right)$	$-\infty \leq \delta \leq \infty$ $\delta \to 0$ = independence copula where $C(u_1, u_2) = u_1 \cdot u_2$ $\delta \to \infty$ = positive dependence $\delta \to -\infty$ = negative dependence **No** tail dependence
Joe	$C(u_1, u_2; \delta) = 1 - \left((1-u_1)^{\delta} + (1-u_2)^{\delta} - (1-u_1)^{\delta}(1-u_2)^{\delta}\right)^{1/\delta}$	$\boldsymbol{\delta \geq 1}$ $\delta \to \infty$ = positive dependence $\delta = 1$ = independence **Upper** tail dependence with coefficient $2 - 2^{1/\delta}$
BB1	$C(u_1, u_2; \theta, \delta) = \left[1 + \left(\left({u_1}^{-\theta} - 1\right)^{\delta} + \left({u_2}^{-\theta} - 1\right)^{\delta}\right)^{1/\delta}\right]^{-1/\theta}$	$\theta > 0;\ \delta \geq 1$ $\theta \to 0; \delta \to 1$ = independence $\theta \to \infty; \delta \to \infty$ = positive dependence **Upper** tail dependence with coefficient $2 - 2^{1/\delta}$ **Lower** tail dependence with coefficient $2^{-1/(\theta\delta)}$
BB6	$C(u_1, u_2; \theta, \delta) = 1 - \left(1 - \exp\left\{-\left[\left(-\log\left(1 - {u_1}^{\theta}\right)\right)^{\delta} + \left(-\log\left(1 - {u_2}^{\theta}\right)\right)^{\delta}\right]^{\frac{1}{\delta}}\right\}\right)^{\frac{1}{\theta}}$	$1 \leq \theta; \delta < \infty$ $\theta \to 1; \delta \to 1$ = independence $\theta \to \infty; \delta \to \infty$ = positive dependence **Upper** tail dependence with coefficient $2 - 2^{1/(\theta\delta)}$
BB7	$C(u_1, u_2; \theta, \delta) = 1 - \left[1 - \left(\left(1 - (1-u_1)^{\theta}\right)^{-\delta} + \left(1 - (1-u_2)^{\theta}\right)^{-\delta} - 1\right)^{-1/\delta}\right]^{1/\theta}$	$\delta > 0;\ \theta \geq 1$ $\delta \to 0; \theta \to 1$ = independence $\delta \to \infty; \theta \to \infty$ = positive dependence **Upper** tail dependence with coefficient $2^{-1/\delta}$ independent of θ **Lower** tail dependence with coefficient $2 - 2^{1/\theta}$ independent of δ
BB8	$C(u_1, u_2; \theta, \delta) = \delta^{-1}\left[1 - \left\{1 - (1-\delta)^{\theta}\right\}^{-1}\left[1 - (1-\delta u_1)^{\theta}\right]\left[1 - (1-\delta v_1)^{\theta}\right]\right]^{\frac{1}{\theta}}$	$0 < \delta \leq 1;\ \theta \geq 1$ $\delta \to 0; \theta \to 1$ = independence $\delta \to 1; \theta \to \infty$ = positive dependence **No** tail dependence except when $\delta = 1$

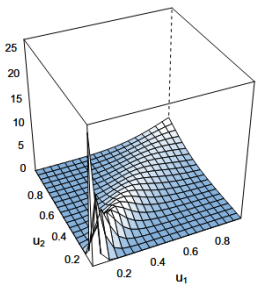
Some copula family like Gaussian and Frank has no tail dependence, which means that extreme events in one variable do not change the probability of extreme events in other variable. Illustration of tail dependence for some copula family is shown in the following figure:



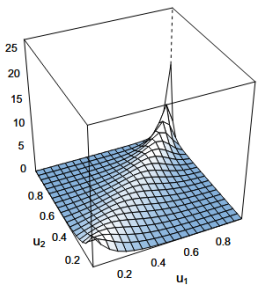
Gaussian



Student's t



Clayton



Gumbel

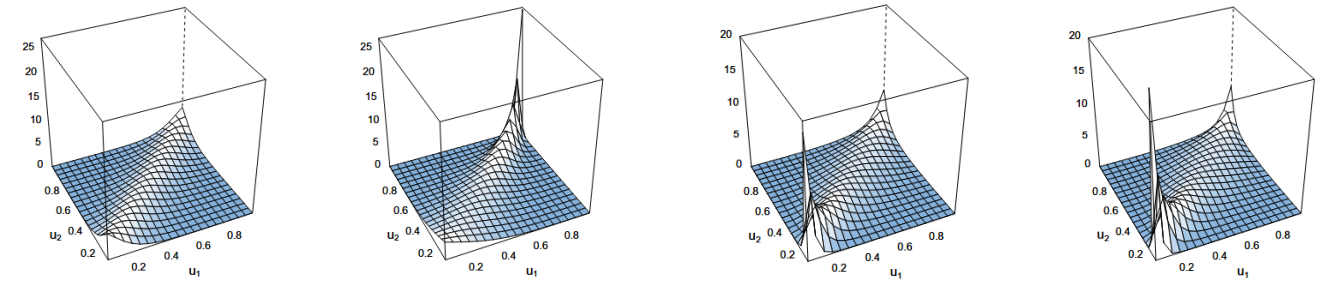


Figure 2.1 Bivariate Copula density of some known Copula Family (taken from Czado, 2019)

2.3.2. Goodness-of-fit Measurement

To evaluate and determine goodness-of-fit, two statistical measures are used: Akaike Information Criterion (AIC) and Cramér-von Misses criteria.

AIC check the quality of each copula family related to other copula family, with AIC value calculated either from maximum likelihood value (formula 4-4) or from mean square error (formula 4-5 and 4-6):

$$AIC_{c,MLE} = 2m - 2ln(L) \tag{4-4}$$

$$\mathrm{MSE_C} = \frac{1}{n}\sum_{i=1}^{n}\left(C_n(u_i, v_i) - C_{\Theta_n}(u_i, v_i)\right)^2 \tag{4-5}$$

$$\mathrm{AIC_{C,\,MSE}} = 2m + n\log(\mathrm{MSE_C}) \tag{4-6}$$

where m is dimension of parameters, L is maximum likelihood estimate from the copula model with parameter θ in the form of $L(\theta; u, v) = \prod_{i=1}^{n} c(u_i, v_i; \theta)$, n is sample size, $C_n(u_i, v_i)$ is the empirical CDF and $C_{\Theta_n}(u_i, v_i)$ is theoretical CDF (Chen, 2019).

Cramér-von Mises criteria is a non-parametric test to estimate goodness-of-fit from empirical CDF compared to theoretical CDF:

$$D_n^2 = \int_{-\infty}^{\infty}\left[C_n(u_i, v_i) - C_{\Theta_n}(u_i, v_i)\right]^2 dC_{\Theta_n}(u_i, v_i) \tag{4-7}$$

with empirical value of the test statistic based on the sample is:

$$d_n = nD_n^2 = \frac{1}{12n} + \sum_{i=1}^{n}\left[\frac{2i-1}{2n} - C_n(u_i, v_i)\right]^2 \tag{4-8}$$

Hypothesis H_0 where $C_n(u_i, v_i) = C_{\Theta_n}(u_i, v_i)$ will be accepted if $d_n \geq d_\alpha$ where d_α is 1 minus α-quantile order of D_n^2 (OpenTURNS, n.d.).

2.3.3. Conditional CDF in Bivariate Copula (h-function)

From the copula theory explained in Section 2.3.1, conditional distributions can be described directly as partial derivative of bivariate copula $C(u, v)$:

$$P(U \leq u | V = v) = h(u, v) = \frac{\partial C(u, v)}{\partial v} \tag{4-10}$$

In the case of conditional exceedance probability, then the formula can be written as:

$$P(U > u | V = v) = 1 - h(u, v) = 1 - \frac{\partial C(u, v)}{\partial v} \tag{4-11}$$

Proof of the bivariate conditional distribution (h-function) can be found in Czado (2019).

2.3.4. Surrogate Time Series via Rank-exact Back Mapping

Surrogate time series is synthetic time series that preserve the original marginal distribution, while modifying or removing certain properties or structures of its dataset. According to Schreiber and Schmitz (1999), this approach is often termed constrained randomization where specific features are held constant and others deliberately randomized. Surrogate series are commonly used in climatology, hydrology, and nonlinear time series analysis to test hypotheses about system behaviour, like the presence of nonlinearity or dependence.

One method to generate surrogate time series is by using rank-exact back mapping. This method involves permuting observed values according to a new ordering derived from an independent ranking variable, thereby preserving the full marginal distribution (quantiles, extremes, histogram) while altering temporal alignment or dependence structures. For example, one study applied a rank-based surrogate generation technique to test multifractality in water-level records by remapping data via a surrogate Gaussian process that preserved empirical order statistics (Wu et al., 2018)

In this study, rank-exact back mapping is applied together with conditional exceedance probability to remove the dependence between two variables, while keeping the marginal distribution. For each pair of u and v where both variables are correlated, the conditional exceedance probability calculated from h-function is used to mapping the value and finding its unconditional equivalents under independence (uncorrelated) variable:

$$P_{correlated}(U > u|V = v) = P_{uncorrelated}(U > \boldsymbol{u}) \qquad (4-12)$$

To extend this mapping across the entire time series, both the correlated variable u and their conditional exceedance probabilities $P_{correlated}(U > u|V = v)$ are sorted, and the smallest u is assigned to the smallest conditional exceedance probability, the second smallest to the second conditional probability, and so forth. This ensures that the recreated time series exactly preserves the marginal distribution while removing the statistical dependence on v. The resulting surrogate time series thus represents the uncorrelated case against which the effect of dependence can be evaluated.

2.4. Time Series Analysis by Transfer Function Model

Time series analysis is described as method to describe relationship between variables by analyzing the past, sequentially recorded observations. These observations can be either independent or dependent on previous values, with the core idea being to understand the relationship between successive data points. Further, by applying stochastic process (evolution of a system over time influenced by random variables), probability distribution for the unknown future or past values can be described.

2.4.1. Transfer Function Model

According to STOWA (2021), there are two distinct time-series approach used in groundwater modeling: statistically oriented Box-Jenkins Transfer Noise, which based on Auto Regressive Moving Average (ARMA) model, and physically oriented Predefined Impulse Response Function In Continuous Time (PIRFICT), which combines statistic approach with physical hydrology process. Pastas is one example that apply PIRFICT model for groundwater hydrology analysis.

Both models make use of transfer function model, which describes possible output ($Y_{1,2,...}$) of dynamic system in response to input variable ($X_{1,2,...}$) using impulse response function. This formulation may also include residual noise component (N_t), as illustrated in Figure 2.2 Transfer Function Noise Model (adapted from Manzione, 2017). The Box-Jenkins Transfer Noise model represents transfer function in an ARMA model form which requires time discretization. As a result, the interval sample for both input and output should be identical and therefore influence the frequency of whole dataset.

In contrast, PIRFICT model differs by using integral function to capture the dynamic system in continuous time. This integral equation is applied to both impulse response function and noise function. The accuracy and performance of this model depends on which type of impulse response function is chosen, if the solution is known. The type of functions can be empirical solutions, such as Gamma or Exponential distribution function, or physically deterministic function. Due to the continuous nature of the integral approach, it is not necessary to synchronize the frequency of the input and output observations. (von Asmuth, 2002)

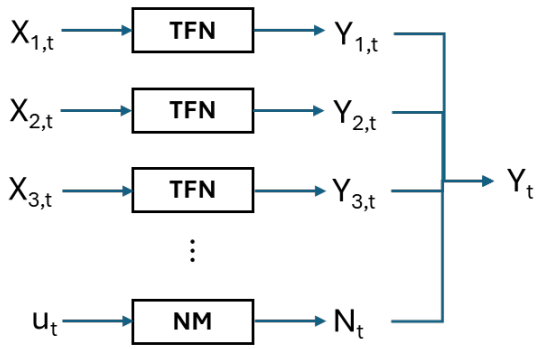


Figure 2.2 Transfer Function Noise Model (adapted from Manzione, 2017)

The basic model of PIRFICT model is written as:

$$h(t) = h^*(t) + r(t) + d \tag{5-1}$$

$$h^*(t) = \int_{-\infty}^{t} S_m(\tau)\theta_m(t-\tau)d\tau \tag{5-2}$$

$$r(t) = \int_{-\infty}^{t} \phi(t-\tau)dW(\tau) \tag{5-3}$$

where:

$h(t)$ = observed value at timestep t (in this case, groundwater head)

$h^*(t)$ = contribution of stress to the observed value, relative to d (i.e. stress due to precipitation and/or evapotranspiration)

$r(t)$ = residual series

d = base elevation of the model

$S_m(t)$ = time series of related stress

$\theta_m(t)$ = transfer impulse response function of the related stress

$\phi(t)$ = noise transfer function

$W(t)$ = continuous white noise (Wiener) process with $E\{dW(t)\} = 0; E[\{dW(t)\}^2] = dt; E\{dW(t_1)\}dW(t_2)\} = 0, t_1 \neq t_2$

The noise model used for PIRFICT model should capture irregular spacing of residual series, which is equivalent to exponential decay function based on AR(1) model:

$$\phi(t) = \sqrt{2\alpha\sigma_n^2}e^{-\alpha t} \tag{5-4}$$

where α = decay rate parameter and σ_n^2 = residual variance.

In Pastas, there are multiple response functions that can be used during groundwater analysis, each with different usage depending on case study, shown in Table 2.3.

Table 2.3 Response Functions in Pastas

No	Name	Formula	Parameters (p)
1	Exponential	$\theta(t,p) = \frac{A}{a} \cdot \exp(-\frac{t}{a})$	A (scaling), a (shape)
2	Gamma	$\theta(t,p) = A\frac{t^{n-1} \cdot \exp(-\frac{t}{a})}{a^n \cdot \Gamma(n)}$	A (scaling), a (shape), n (shape)

3	Hantush	$\theta(t,p) = \frac{A}{2t \cdot K_0(2\sqrt{b})} \cdot \exp(-\frac{t}{a} - \frac{ab}{t})$	A (scaling), a (shape), b (shape), K_0 = modified Bessel function of second kind, order zero
4	Double Exponential	$\theta(t,p) = A \cdot (\frac{1-\alpha}{a_1} \cdot \exp\left(\frac{-t}{a_1}\right) + \frac{\alpha}{a_2} \cdot \exp(-\frac{t}{a_2}))$	A (scaling), a_1 (shape), a_2 (shape), α (distribution)
5	Polder	$\theta(t,p) = A \cdot \sqrt{\frac{ab}{\pi}} t^{-1.5} \cdot \exp(-\frac{t}{a} - \frac{ab}{t})$	A (scaling), a (shape), b (shape)
6	Four Parameter	-	Combination of Exponential, Gamma, Hantush and Polder

Gamma function is the most common one to use, because it can model the delay depending on value of n, with $n > 1$ shows the delay in response. Gamma response function often used for precipitation and evaporation stress, while Exponential can be used as simplified Gamma function with less parameters, which can be helpful during model optimization and processing time, or with stresses that immediately affect the groundwater head (no delay). Hantush response function mainly used in case of pumping well scenario, and Polder response function is used for the case of surface water level variation boundary. Double exponential function is intended for case of multi-aquifer system.

2.4.2. Water Balance Model

The minimum input variable needed to build the PIRFICT groundwater model in Pastas, following STOWA guideline, are groundwater level measurement, precipitation and evaporation. Both precipitation and evaporation influence toward groundwater storage were proved by Milly (1994) using simple water balance model, depicted in Figure 2.3, which shows that local annual water balance is controlled by the distribution of supply (precipitation) and demand (evapotranspiration) and balanced by soil water storage (groundwater). In the water balance model, precipitation serves as the main source of water input to the land surface, with a portion infiltrating the soil profile and contributing to groundwater recharge, depending on soil properties, land cover, and antecedent moisture conditions. In contrast, evaporation, together with transpiration from vegetation (collectively referred to as evapotranspiration), constitutes the principal mechanism for atmospheric water loss.

To improve model accuracy, few additional variables can be considered. Artificial groundwater extraction from pumping well drilled into the aquifer will influence the phreatic line, especially if the infiltration rate is lower than the extraction. Conversely, artificial groundwater recharge via surface flooding or from reclaimed wastewater also helps to balance the water demand while keeping the soil water storage constant. In the area with hydraulic connection such as rivers and lakes, surface discharge and/or water level can have larger impact toward groundwater head via seepage flux mechanism. Han et al. (2019) also shows that there is both discharge and recharge relationship between lake and groundwater, which vary by season. The cumulative effect of high-frequency pumping and recharge activities over extended periods can lead to pronounced long-term shifts in groundwater trends.

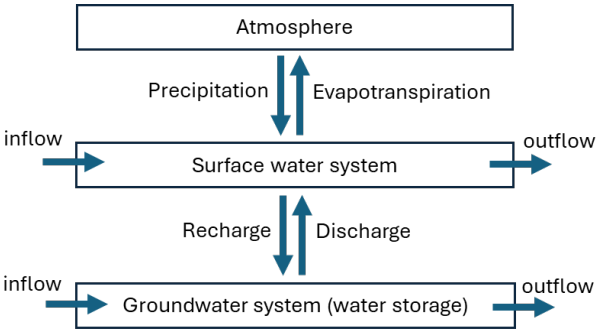


Figure 2.3 Simple Water Balance Model (Han et al., 2019)

2.4.3. Recharge Model

During setting up model, there are few ways to incorporate precipitation and evapotranspiration as an input, either separate or combined. Recharge model is a stress model that combines both variables to model

groundwater fluctuations. There are four types of Recharge model options included in Pastas library: Linear; Berendrecht; FlexModel; Peterson (see Table 2.4).

Table 2.4 Recharge model formula and their notation

Recharge model	Recharge formula (R)	Notation
Linear	$R = P - f \cdot E$	P = precipitation rate f = evaporation factor, default = 1 E = evaporation
Berendrecht	$\frac{dS_e}{dt} = \frac{1}{D_e}(P_e - E_a - R)$ $R(S_e) = K_s S_e^{\lambda}\left(1 - \left(1 - S_e^{\frac{1}{m}}\right)^m\right)^2$	S_e = effective degree of water saturation $(0 < S_e \leq 1)$ D_e = effective thickness of root zone, calculated as root zone depth multiplied by difference between saturated and residual soil water content P_e = net precipitation (gross precipitation multiply by interception factor) E_a = actual evapotranspiration K_s = saturated hydraulic conductivity λ, m = empirical shape factors with $\lambda m \geq -2$
FlexModel	$\frac{dS}{dt} = P_e - E_a - R$ $R = K_s\left(\frac{S}{S_u}\right)^{\lambda}$	S_u = maximum storage capacity
Peterson	$\frac{dS}{dt} = P_e - E_a - R$ $P_e = P\left(1 - \frac{S}{10^{S_u}}\right)^{\alpha}$ $E_a = E_p\left(\frac{S}{10^{S_u}}\right)^{\gamma}$ $R = 10^{K_s}\left(\frac{S}{10^{S_u}}\right)^{10^{\beta}}$	E_p = potential evapotranspiration rate α, γ, β = dimensionless parameter

Berendrecht, FlexModel and Peterson use non-linear model for calculation based on soil-water balance equation. These non-linear models have their own strength, with Berendrecht model is one of the earliest non-linear models based on degree of saturation at the root zone and can predict extreme groundwater event better than Linear model. Peterson model improves on the previous model with main goal to simulate episodic recharge and droughts. FlexModel use less input (precipitation, potential evaporation, and groundwater levels) but still shows remarkable improvement compared to linear model. In this study, both Linear and FlexModel will be used and compared.

2.5. Slope Stability Analysis

How does groundwater level influence slope stability in dike? To answer this question, the basic concept of slope stability should be introduced. Slope instability occurs when shear stress of soil is larger than its shear strength. Shear strength of soil is the maximum amount of resistance toward sliding along failure plane within soil mass, which is controlled by effective stresses. The Mohr-Coulomb (drained) criteria states:

$$s = c' + \sigma_n' \tan \phi \qquad (5-5)$$

where s is shear strength, c' is effective cohesion, σ_n' is effective stress and ϕ is angle of internal friction. This equation assumed soil dilatancy equal to angle of internal friction. Effective stress can be calculated from total

stress of the soil, which is sum of all forces acting on soil (i.e. self-weight, water, external load, etc.) F divided by its area A, minus pore water pressure u:

$$\sigma = \frac{F}{A} = \sigma' + u \qquad (5-6)$$

2.5.1. Pore Water Pressure

Pore water pressure u consists of hydrostatic water pressure (pressure due to gravity) u_0 and additional water pressure due to external load Δu. Hydrostatic water pressure at depth d below water surface:

$$u_0 = \gamma_w d \qquad (5-7)$$

with γ_w is unit water weight (9.81 kN/m^3).

While the total stress is same, the balance between effective stress and pore water pressure will be different depending on soil conditions during loading: undrained and drained. In the case of undrained soil, there is no change in soil volume such that resistance force from the additional loading is fully carried by pore water pressure. However, if the soil is drained, the water inside the soil is allowed to flow over time. Therefore, the additional loading will be resisted by the soil granule itself, and the pore water pressure left is only from hydrostatic water pressure. (Duncan et al., 2014)

For soil with cross section consists of multiple layers of clay and sand, there are another factor to be considered where water flows vertically upwards from sand layer to clay layer, called seepage. This phenomenon occurs when the water pressure in confined sand layer is higher than the clay layer above, causing pressure gradients to occur. In the case of upward flow, the calculation of pore water pressure changes to:

$$u = \gamma_w d(1 - i) \qquad (5-8)$$

with i as hydraulic gradient from Darcy's law, $i = -\frac{k}{q}$ where q is specific discharge and k is hydraulic conductivity (Verruijt, 2018).

To determine pore water pressure in soil cross section, there are two lines that must be determined: phreatic line, based on hydraulic head measured at unconfined aquifer, and piezometric line, measured at confined aquifer.

2.5.2. Bishop Method

To determine safety factor against slope failure, Bishop limit equilibrium method is chosen. This method determines failure surface as a circular with specific radius, beginning from some distance before the slope starts and finished at some distance after the slope ends. Determination of which failure surface causes the most critical conditions can be done using trial-and-error or optimization algorithm. After that, the soil is divided into multiple slices, each with its own properties.

Bishop method works with assumption that for each slide the equilibrium vertical force is zero and shear stresses between each slice is neglected such that only resultant force is horizontal (Figure 2.4).

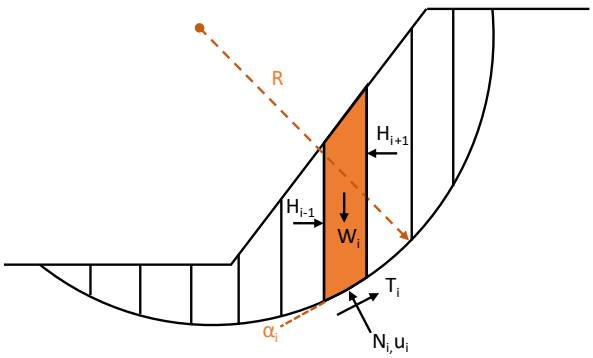


Figure 2.4 Bishop slices and force direction

The vertical equilibrium per slices calculated as:

$$\Sigma F_v = 0 \rightarrow -W_i + N_i \cos \alpha_i + u_i \cos \alpha_i + T_i \sin \alpha_i = 0 \qquad (5-9)$$

and moment equilibrium calculated as:

$$\Sigma M_i = 0 \rightarrow W_i R \sin \alpha_i - T_i R = 0 \qquad (5-10)$$

The shear force along the slip plane is equal to maximum shear force divided by Factor of Safety FoS, and shear strength follows Mohr-Coulomb equation on (5-5) such that the shear force equation become:

$$T_i = \frac{s_i l_i}{FoS} = \frac{c' l_i + (N_i - u_i l_i) \tan \phi}{FoS} \qquad (5-11)$$

Substitute Equation (5-11) into (5-10) and solve for N_i and then sum for all slices such that FoS can be calculated as:

$$F = \sum_{i=1}^{n} \left(c' l_i + \tan \phi \left(\frac{W_i - \frac{c' l_i \sin \alpha_i}{F}}{\cos \alpha_i \left(1 + \frac{\tan \phi \tan \alpha_i}{F}\right)} \right) \right) \times \frac{1}{\sum_{i=1}^{n} W_i \sin \alpha_i} \qquad (5-12)$$

Chapter 3. Methodology and Approach

This chapter explain the detailed steps taken during this study. Flowchart of whole processes is shown in Figure 3.1 and Figure 3.2.

3.1. Hydrology and Meteorological Variables

To determine the hydrology and meteorological variables, literature review approach is used based on Chapter 2.1. Relevant scientific articles, technical reports, and government publications were collected using databases such as Scopus, Google Scholar, and official Dutch water management agency sources (e.g., Rijkswaterstaat reports). Publications were selected based on their relevance to phreatic line behavior and dike stability. Data extracted from these sources were systematically categorized to identify variables affecting the phreatic level, their interactions, and their impact on slope stability.

To support the literature review finding, a simple calculation will be executed by calculating discharge volume over time from IJsselmeer to North Sea on the Afsluitdijk, which then will be compared with wind direction and magnitude from relevant weather station. To calculate the hydraulic discharge, simplified version of formula to calculate discharge over submerged weir will be used:

$$Q = B\,(h_{outside} - d)\sqrt{2g(h_{inside} - h_{outside})}$$

where B is the width of sluice, h_{inside} is the water level from IJsselmeer side, $h_{outside}$ is the water level at the North Sea side of Afsluitdijk, d is the elevation of the base of sluice and g is gravitational acceleration.

3.2. Correlation and Copula Fitting

This analysis focuses on exploring the dependence structure between water level and precipitation data using correlation measures and copula fitting techniques using *pyvinecopulib*. The water level dataset, recorded every 10 minutes, is resampled to hourly frequency by computing the hourly mean, while the precipitation dataset, recorded hourly, is processed by rounding up rainfall values between 0 and 0.05 mm to 0.05 mm.

Correlation between the hourly water level and the smoothed hourly precipitation is evaluated using three statistical measures: Pearson, Spearman, and Kendall, as outlined in Table 2.1. Also, correlation between hourly water level and each of the following scenario: hourly precipitation, 7-day sum, 14-day sum, 21-day sum, 28-day sum is calculated, and then compared to choose the highest correlation. A scatter plot is generated to visualize the relationship between both variables. To isolate the dependence structure from the marginal distributions, the Probability Integral Transform (PIT) is applied to both datasets, transforming them into standard uniform distributions. This transformation involves ranking each dataset and dividing the ranks by the total number of observations plus one, resulting in a u-v plot (union space) where copula fitting is conducted.

Several copula families are tested to model the dependence structure, including Gaussian, Student's t, Clayton, Gumbel, Frank, Joe, BB1, BB6, BB7, and BB8, as listed in Table 2.2. For each fitted copula, the Akaike Information Criterion (AIC) is calculated to evaluate the trade-off between model fit and complexity, with the lowest AIC indicating the best balance. Additionally, the Cramer-von Mises (CvM) statistic is computed to assess the goodness-of-fit by comparing the empirical and theoretical copula cumulative distribution functions (CDFs). The empirical CDF is calculated by counting the proportion of observation pairs less than or equal to each data point in the union space, while the theoretical CDF is derived from the fitted copula. The copula with the smallest CvM value is considered the best fit.

Finally, random samples are generated from the best-fitting copula model to visually validate the fit. These samples are plotted alongside the observed data in both the uniform (union) space and the original variable space. The original space is reconstructed through the inverse PIT process, allowing a direct comparison between the generated and observed datasets.

3.3. Groundwater Modelling

Due to limited availability of groundwater observation data, it is necessary to define a model to simulate longer dataset, which will be used as input during dike stability calculation. To create the model, an open-source

Python package called Pastas will be used. Pastas model is a type of time-series analysis model where the output series is influenced by multiple input series and added noise such as autocorrelated residuals from autoregressive (AR) and moving average (MA) process. In the context of groundwater modeling, by combining multiple factors such as precipitation, evapotranspiration and observed water level, a groundwater head model can be generated and fitted with existing groundwater observation.

Pastas is chosen for this research because it can simulate synthetic groundwater head time series beyond available dataset based on the previously defined model. The model then can be used to perform hindcasting (backward projection) or forecasting (future projection) and its uncertainty, which also can be improved by using longer time series of precipitation and evapotranspiration.

To build the Pastas model, the following datasets are prepared: historical groundwater level observations, hourly precipitation data (where rainfall values between 0 and 0.05 mm are rounded up to 0.05 mm), hourly evapotranspiration data, and hourly water level records. The groundwater observations are divided into two subsets: a calibration set and a validation set. To ensure seasonality is represented despite the limited observation period, the calibration set covers at least one year of data, while the remaining data are reserved for validation stage.

3.3.1. Determining model configuration

The first step in the modeling process is to determine the most suitable model configuration. The primary goal is to identify which configuration results in the best performance based on statistical metrics. A total of eight configurations are evaluated:

- Configuration A_Lin: using Linear recharge model with response function Exponential
- Configuration A_Flex: using FlexModel recharge model with response function Exponential
- Configuration B_Lin: using Linear recharge model with response function Gamma
- Configuration B_Flex: using FlexModel recharge model with response function Gamma
- Configuration C: using separate stress model for Precipitation and Evapotranspiration, with each stress model using response function Gamma
- Configuration D: same as Configuration C, with addition of Water Level measurement from Markermeer as stress model with response function Polder
- Configuration E_Lin: same as Configuration B_Lin, with addition of Water Level measurement from Markermeer as stress model with response function Polder
- Configuration E_Flex: same as Configuration B_Flex, with addition of Water Level measurement from Markermeer as stress model with response function Polder

In general, the Polder or One response function is typically used for modeling the influence of water levels. The One response function represents an immediate response with no delay and is more appropriate for observation wells located very close to the water surface. The Polder response function, on the other hand, introduces a delayed reaction and is better suited for wells situated further from the Markermeer.

All model configurations use the same input datasets, which include hourly precipitation and evapotranspiration data and groundwater level measurements from observation well. For configurations that include water level as a stress model, the data is taken from the representative station. This dataset, originally recorded at 10-minute intervals, is resampled to 15-minute intervals for consistency.

In calibration stage, each configuration is run once with simulated frequency is set to hourly. The initial result then evaluated using two performance metrics: the coefficient of determination (R^2) and the corrected Akaike Information Criterion (ΔAICc). R^2 quantifies how well the model predictions align with observed values by calculating the proportion of explained variance. Although a higher R^2 indicates a better fit, it may also suggest overfitting, particularly when the model contains a large number of parameters. $\Delta AICc$, on the other hand, introduces a correction term to the standard AIC, imposing a stronger penalty for complexity when the sample size is small. A $\Delta AICc$ value of zero indicates the best configuration among those compared.

During validation, each model configuration generates samples for the same timeframe as the validation dataset. The results are then evaluated again, this time using R^2 and the Root Mean Squared Error (RMSE) metrics. RMSE measures the average magnitude of prediction errors, with an emphasis on larger errors due to squaring before averaging. Consequently, a smaller RMSE value indicates fewer large deviations, reflecting a better overall fit.

The configuration then evaluated based on both calibration and validation result. Configuration with low $\Delta AICc$ value, low RMSE value and acceptable R^2 for both calibration and validation stage is selected as the default for the next modeling steps. However, if the default configuration does not provide a satisfactory fit during next modeling, it may be modified and re-evaluated to achieve improved results.

3.3.2. Groundwater scenario

Several groundwater modeling scenarios are developed, each based on different dike cross-sections and associated water level observation stations. While the spatial configuration varies, all scenarios consistently use hourly precipitation and evapotranspiration data from the same reference station, and groundwater observation datasets are also recorded at hourly intervals. The corresponding water level datasets have a higher resolution, recorded every 15 minutes. The purpose of these scenarios is to examine and compare groundwater behavior patterns across various locations.

Each scenario initially uses the default model configuration determined from the earlier model selection step. Similar to previous step, each model is generated once for calibration and then validated. However, if the resulting model fit is unsatisfactory where R^2 is below 0.5 after validation stage, then the configuration is modified. If R^2 remains below 0.5 after adjustment, the configuration with the highest achievable R^2 is chosen.

The final results after adjustment are visualized in a mosaic plot format. The upper-left panel displays a comparison of observed and modeled groundwater levels, while the next panel below it shows the residuals. Additional panels on the lower-left present the contribution of each input variable to the model prediction, indicating whether the influence is positive or negative. On the upper-right side, the first table summarizes the model's goodness-of-fit metrics, followed by a second table listing the calibrated model parameters. The remaining panels on the right illustrate the step response of each variable, showing the dynamic impact of each input on the groundwater head over time.

Using the calibrated Pastas model for each scenario, a hindcasted daily random sample is generated to fill in missing groundwater data from. A one-year warm-up period is applied to stabilize the model before simulation begins. In the case when hindcasted sample shows irregular behavior, parameters used in the model might be manually modified until reasonable behavior is observed.

3.4. Slope Stability Modelling

The slope stability modeling process begins with the preparation of essential inputs, including soil parameters and dike section geometry. A single dike cross-section is selected for analysis due to the availability of detailed soil stratigraphy at that location.

Two main tools are used together in this process: D-Stability and the *geolib* Python module. D-Stability is employed to define the geometry of the dike cross-section, soil layer structure, and relevant geotechnical properties. The *geolib* module is used to automate the assignment of phreatic and head lines and to run stability calculations in a looped structure for multiple groundwater conditions.

The geometry of the selected dike cross-section is configured in D-Stability, along with the corresponding soil layer properties. The Bishop limit equilibrium method is chosen as the stability solver, combined with a brute-force search algorithm to identify the critical slip surface with the lowest safety factor for each input condition.

Groundwater data are derived from a measured time series. To reduce computational load during simulation, shorter period were chosen, specifically during high water season (end of 2023 – start of 2024).

The groundwater datasets of the 4 stand pipes in the considered cross-section are used as input in the *geolib* Python workflow. While the geometry and soil layer data are saved in a standardized format (.stix) from D-

Stability, the groundwater time series and observation well positions are defined directly in the Python code. The slope stability solver is then executed in a loop, where the phreatic and head lines are updated for each time step. The results are visualized in both Q-Q plots and time series plot together with best rolling window sum of precipitation (from Chapter 3.2) and hourly water level to observe overall trends and identify extreme event when the safety factor is lowest.

To indicate the effect of correlated versus uncorrelated variable towards dike safety factor, synthetic water level time series is generated from marginal statistics. This independent water level time series should have the same exceedance probability of the observed peak water level as the dependent one. By calculating the probability of observed peak water level during dependent case given local precipitation during same timeframe $P_{correlated}(W > w_{max}|R)$ using previously established copula model in Chapter 3.2, the marginal statistics for independent water level can be used to determine specific water level **w** with same exceedance probability value, $P_{uncorrelated}(W > \boldsymbol{w})$. This conditional exceedance probability is calculated for each pair in whole observed time series. To form the synthetic water level time series, rank-exact back-mapping method is applied. By sorting both observed water level value and the conditional exceedance probability, then assign the smallest water level to largest conditional exceedance probability, the second smallest to second largest, and so on. This step ensure that the dependence of precipitation is removed.

The synthetic water level time series then used as an input to Pastas, together with existing precipitation and evapotranspiration dataset to generate new hydraulic head time series while keeping the parameters the same. Then, factor of safety is calculated again using D-Stability python module with similar workflow as previous run. The result then analysed and compared with correlated run.

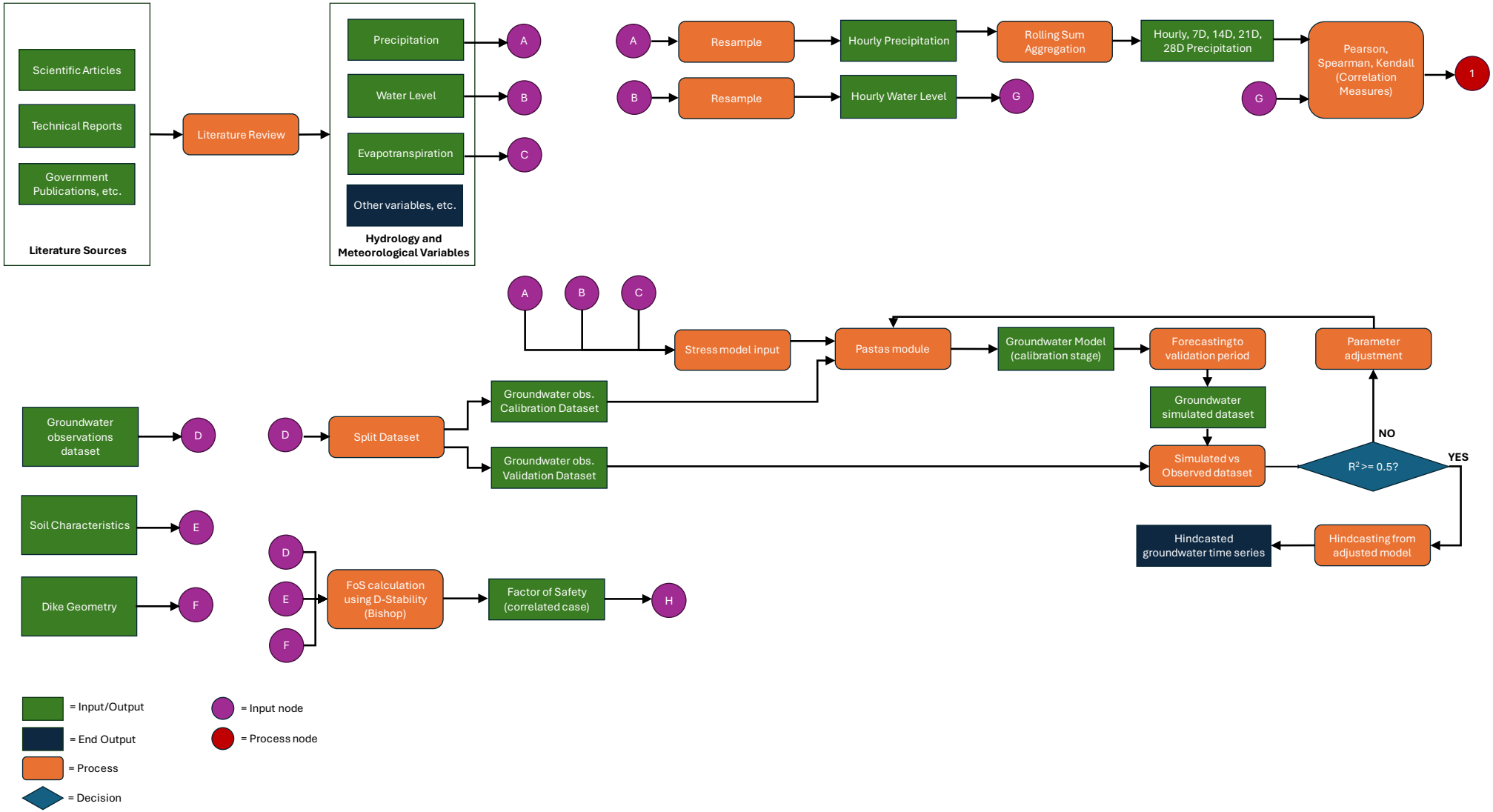


Figure 3.1 Flowchart of whole process (part 1 of 2)

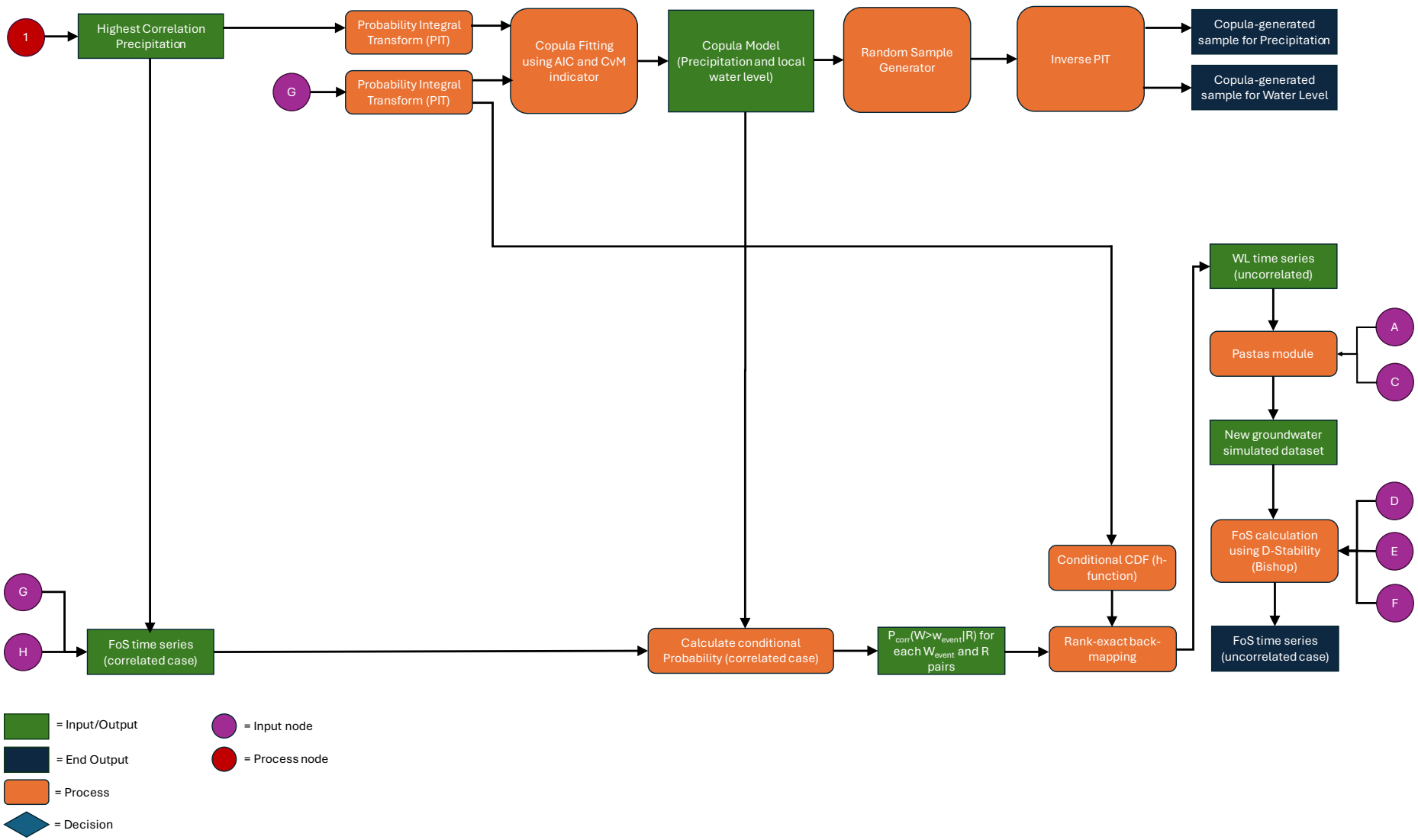


Figure 3.2 Flowchart of whole process (part 2 of 2)

Chapter 4. Case Study and Datasets

4.1. IJsselmeer System Area

IJsselmeer system area is divided into four major subsystems: IJsselmeer, Markermeer, IJssel-Vecht delta and peripheral lake around Veluwe area. Outside of the system, there are three large polders where the run-off directly impacts the Ijsselmeer system as shown in Figure 4.1: Wieringermeer (A), Flevoland (B) and Nordoostpolder (C).

The main function of Ijsselmeer system is divided into three: flood defence, mainly to control balance between water from Ijssel river and regional runoff versus high water from storm surge; freshwater supply, mainly for agriculture, drinking water and salination flushing; aquatic ecosystem as habitat for existing flora and fauna. Due to smaller storage capacity of Markermeer (around 700 km^2) compared to Ijsselmeer (around 1200 km^2), Ijsselmeer receive more water discharge especially during Winter period. The permitted fluctuation for Ijsselmeer is around -0.40 to -0.05 m NAP, where Markermeer around -0.40 to -0.20 m NAP. During summer, Ijsselmeer area maintain water level around -0.10 to -0.30 m NAP to keep up with freshwater demands.

In the schematic of Ijsselmeer Area (Figure 4.2), water from Ijssel river and Ijssel-Vecht delta is contained in Ijsselmeer lake area. The discharge is controlled from Afsluitdijk with two main gates: Stevinsluis and Lorentzsluizen. This structure closed off Ijsselmeer area from directly exposed to North Sea, and therefore the impact of high water level due to storm surge is alleviated. Also, Houtribdijk, together with two gates: Krabbersgat and Houtribsluizen, split the Ijsselmeer from Markermeer, which also affect the dominant hydraulic load toward dike in Ijsselmeer area. Besides IJsselmeer, the water level fluctuation in Markermeer area is also controlled from North Sea Canal using two lock complexes: Ijmuiden locks between North Sea and North Sea Canal, and Schellingwoude locks between North Sea Canal and Markermeer. Due to main function of North Sea Canal as shipping route to Port of Amsterdam, the water level at North Sea Canal is maintained around -0.45 m NAP. Both lock complexes also function to control salt intrusion from North Sea to Markermeer. Based on this system, Markermeer dike is mostly affected by changes in water level, while Ijsselmeer dike and IJssel-Vecht delta are more prone to onshore wind effect from North Sea. (Deltares, 2022)

Figure 4.1 Hydrological Map of Netherlands (from Water Management in the Netherlands, Rijkswaterstaat, 2019). The main run-off toward IJssel Lake comes from Wieringermeer (A), Flevoland (B) and Nordoostpolder (C), while discharge control in Afsluitdijk is handled

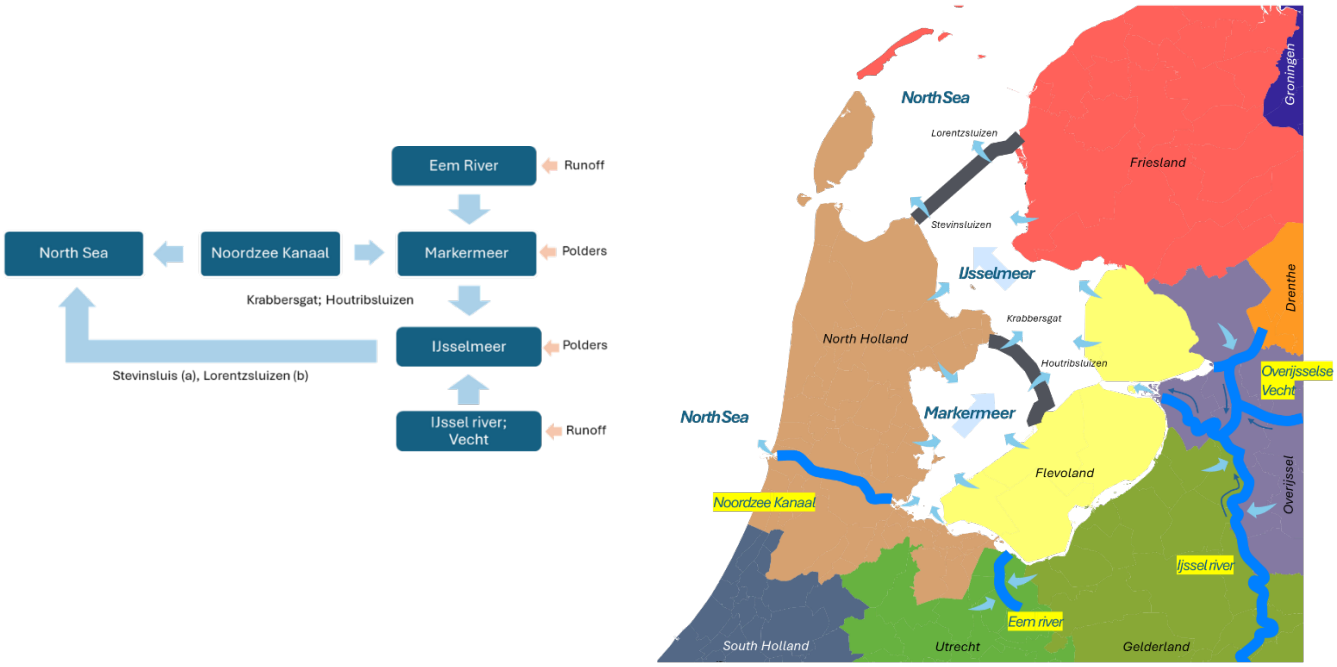


Figure 4.2 Schematic of Ijsselmeer Area (left), and Map Illustration (right). Arrows show the main flow direction of water in case of high water.

4.2. Dike Section

The dike section chosen for this study is located between Hoorn and Enkhuizen, with total length for this sector is around 17.8 km (Figure 4.3). Markermeer dike provides protection from flood for North Holland population that living behind the dike. Markermeer lake itself now is a freshwater lake after the construction of Afsluitdijk that closed off the water from North Sea. After dike assessment in 2006, Markermeer dike needs to be reinforced according to new standard from Water Act 2017. Now, Markermeer dike is under reinforcement project executed by local water board Hoogheemraadschap Hollands Noorderkwartier (HHNK). According to HHNK in their booklet *Naar veilige Markermeerdijken* (2014), the dike protects around 1.2 million people with economic value reaching up to 25 billion Euros. Compared to dike in Ijsselmeer and North Sea coastal area which highly influenced by high waves, Markermeer dike is prone to failure from precipitation and high water level from lake. (HHNK, 2014)

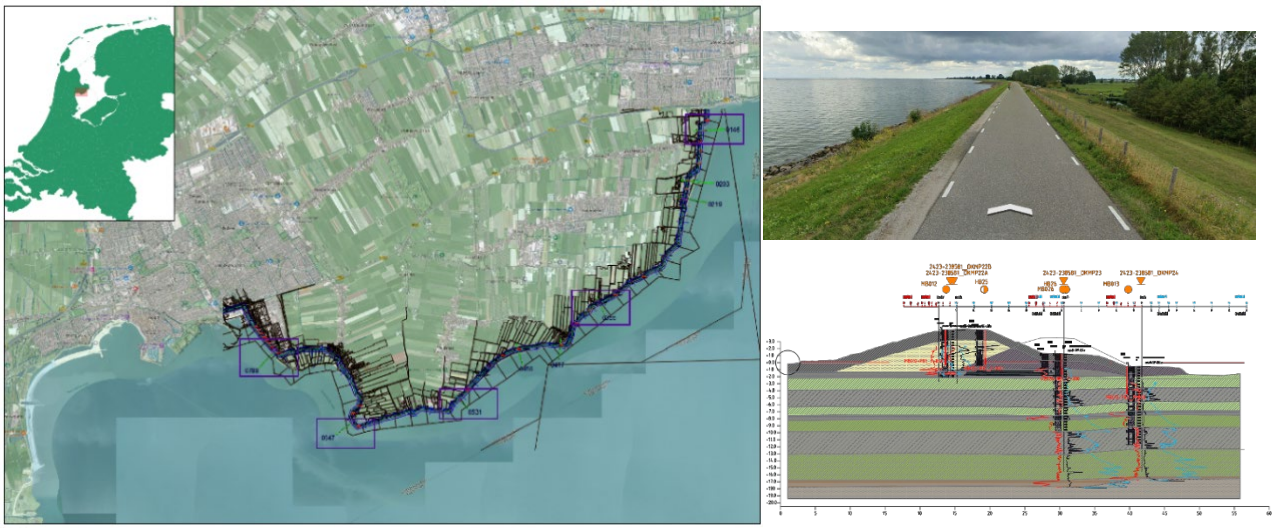


Figure 4.3 Location of Dike Section (left), Photo from Google Street View – left side is Markermeer (upper right), Typical Dike Cross Section (lower right)

In this Hoorn-Enkhuizen dike section, which is part of Markermeer flood control system, the water level is tightly maintained with the sluices and canal operation. The change on water level is relatively low but during storms the capacity to discharge to sea and IJsselmeer is constrained and wind set-up/seiche can temporarily raise lake levels, causing outer slope loading and overtopping which directly affects dike safety. Meanwhile in the land

area, during prolonged rainfall the polder fills and pumping increases, yet if lake levels are high or structures are restricted, backwater effects keep polder levels elevated. Higher landward water levels reduce the lake to polder head difference (slightly easing piping risk), but they increase pore pressures in the dike's inner slope, lowering effective stress and shear strength such that macro-stability/micro-instability become more critical. After the storm, if the lake is drawn down quickly, the dike can remain water-logged on the inside, which is an unfavorable rapid drawdown condition. Rapid post-storm drawdown of the lake can also leave pore pressures high in the dike body, unfavourable for inner-slope stability.

Typical cross section of Markermeer dike is shown in Figure 4.3 at lower right. For calculation of dike safety factor, only *raai_2* cross section is chosen due to availability of soil data layer. Soil characteristic of the dike consist of sand as main dike body covered in clay while the inner berm mainly built from clay. The first soil layer is peat with thickness around 1 m, followed with organic clay, sand mixed with clay, and then organic clay again before peat layer shown again. This layer is repeated once again until reaching deep sand layer. After the inner berm, there is a *polderpeil* or ditch with water level maintained at -1.85 m NAP (HHNK, 2023).

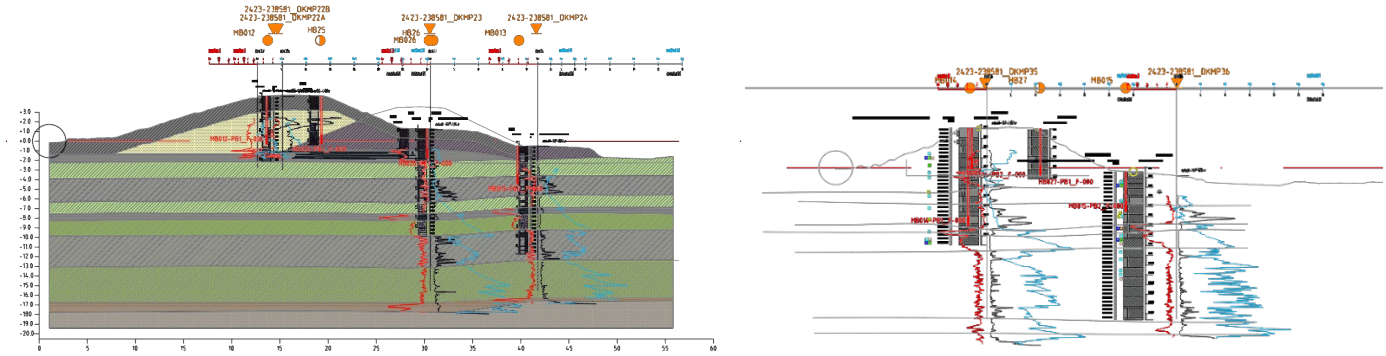


Figure 4.4 Cross section of the considered dikes, raai_2 (left) and raai_3 (right)

There are two dike cross-section that will be analyzed further: cross section *raai_2* which have 6 groundwater observation stations and cross section *raai_3* which have 4 groundwater observations (Figure 4.4). All these observation stations differ based on the location point from closest to furthest from lake, and from the depth of observation well in correlation with soil layer under the dike (Table 4.1).

Table 4.1 Observation Well per Cross Section

Cross Section	Observation Well	Soil Layer
raai_2	MB012-PB1	Phreatic
	HB025-PB1	Phreatic
	MB026-PB2	Phreatic
	MB026-PB1	Sandy intermediate layer
	MB013-PB2	Sandy intermediate layer
	MB013-PB1	Deep sand layer
raai_3	MB014-PB1	Sandy intermediate layer
	MB014-PB2	Phreatic
	HB027-PB1	Phreatic
	MB015-PB2	Deep sand layer

4.3. Dataset

The overview of variables used in this study and data availability is shown in Table 4.2.

Table 4.2 Variable used and data availability

Variables	Data	Availability
Water level	Hourly water level (mm)	Rijkswaterstaat

Precipitation	Hourly precipitation (mm)	KNMI
Evapotranspiration	Daily evapotranspiration (mm)	KNMI
Wind Magnitude and Direction	Daily wind magnitude and its direction (m/s)	KNMI
Groundwater Measurement	Local groundwater measurement in Markermeer dike section (see Chapter 4.2)	Local water board (HHNK)
Dike Geometry	Cross section of the dike	Local water board (HHNK)
Soil Characteristic	Soil parameters (unit weight, drained Mohr-Coulomb parameters)	Local water board (HHNK), and assumptions were used in the case of unknown parameters.

4.3.1. Hydrological and Meteorological Variable Dataset

Dataset that will be used to support the literature review analysis: Water level from Kornwerderzand Buiten and Binnen station (2023-01-01 until 2024-07-31) per 10 minutes frequency, wind direction and magnitude from The Kooy station, same period, with daily frequency.

4.3.2. Correlation and Copula Fitting Dataset

There are two variables to be analyse: water level and hourly precipitation. Water Level dataset is taken from Krabbersgat Zuid measurement station, period 1999-03-22 – 2025-05-23 from *Rijkswaterstaat Waterinfo*, while daily precipitation dataset is taken from Berkhout weather station with same period from KNMI.

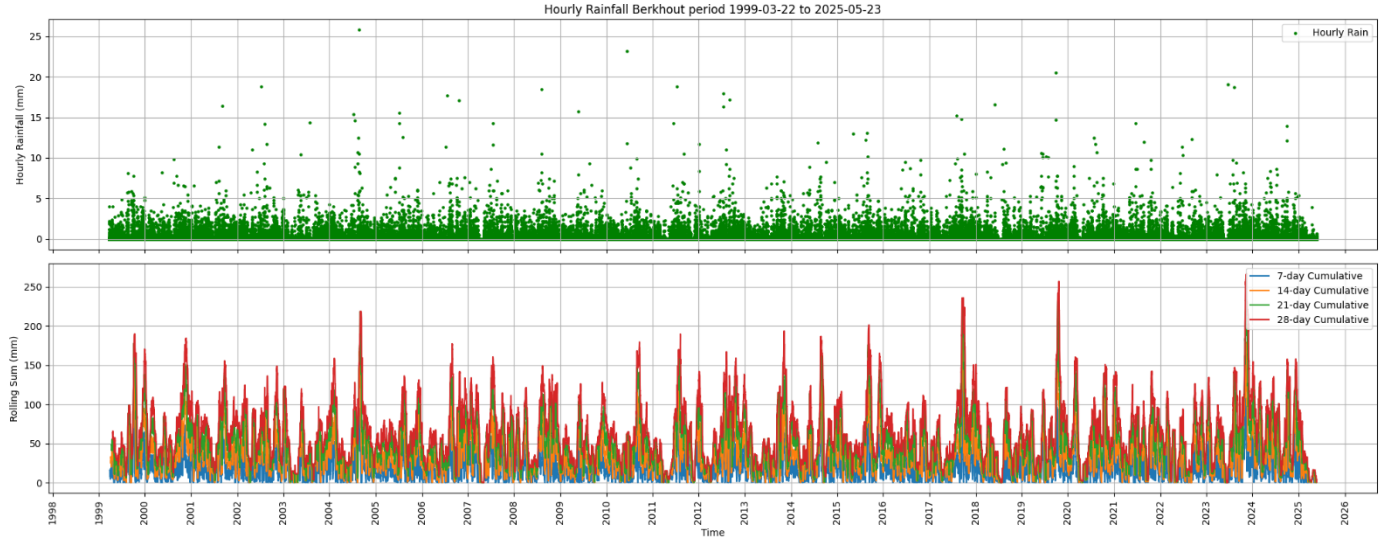


Figure 4.5 Hourly Rainfall Berkhout, period 1999 – 2025. Bottom plot shows 7-day, 14-day, 21-day and 28-day Rolling Window Cumulative (total rainfall).

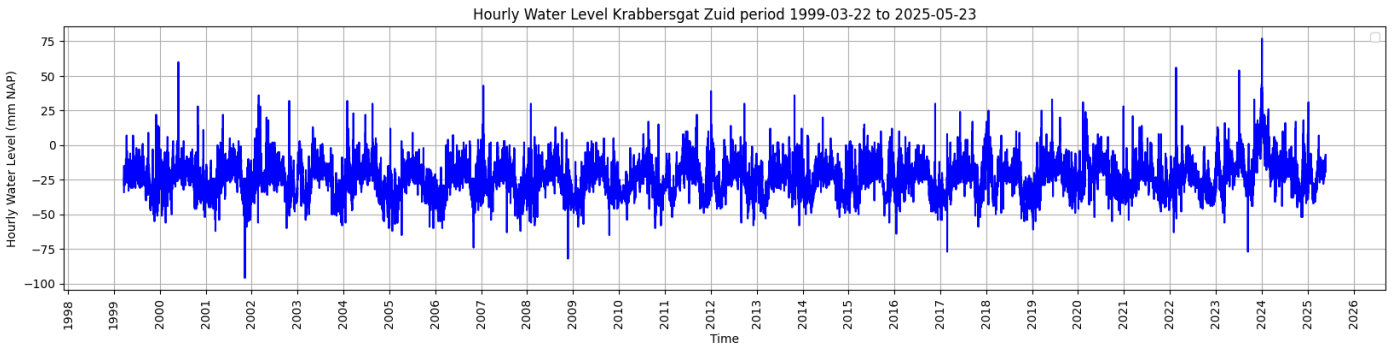


Figure 4.6 Hourly Water Level in Krabbersgat Zuid, period 1999 – 2025.

4.3.3. Groundwater Model and Slope Stability Dataset

Dataset that will be used for groundwater model and slope stability analysis are:

- Historical dataset of groundwater measurement from two chosen dike cross-section, which is taken from dike section around Markermeer area, location between Hoorn and Enkhuizen. The measurement is taken between 10/11/2023 until 25/02/2025.
- Hourly precipitation dataset taken from Berkhout Station, start from 23/03/99 until 06/05/25. Data source from KNMI.
- Daily evapotranspiration dataset taken from Berkhout Station, start from 23/03/99 until 06/05/25. Data source from KNMI.
- Hourly period water level recording from two sources:
    - Water level at pumping station Drieban (start from 16/10/23 until 26/02/25)
    - Water level recording at Krabbersgat Zuid taken from Rijkswaterstaat (from 23/03/99 until 06/05/25 following the availability data of precipitation and evapotranspiration).
- Soil parameter and dike cross section of *raai_2*.

While Drieban pumping station is closer to dike cross section location, the length of observation is more limited compared to Krabbersgat Zuid. To check the difference of water level between Drieban pumping station and Krabbersgat Zuid station, visual inspection during same period is plotted at Figure 4.9. While the general shape and height is similar, there are more 'spikes' at Krabbersgat Zuid, which can be caused from wind effect. It is also interesting to see that during the highest measured water level at the beginning of January 2024, Drieban Pumping Station record 0.571 m while Krabbersgat Zuid record 0.77 m, which shows that the wind direction during that day is around northeast.

Table 4.3 Soil Layer Properties of raai_2 Cross Section

Soil Layer	Unit Weight (kN/m3)	Cohesion (kN/m2)	Friction Angle (deg)
Clay – Top Layer	16.32	5	20
Sand	18	0	35
Clay – Shallow	14.92	5	20
Peat	10.89	5	15
Organic Clay	13.76	5	20
Sand with Clay	18	0	35
Clay – Deep	14	5	20

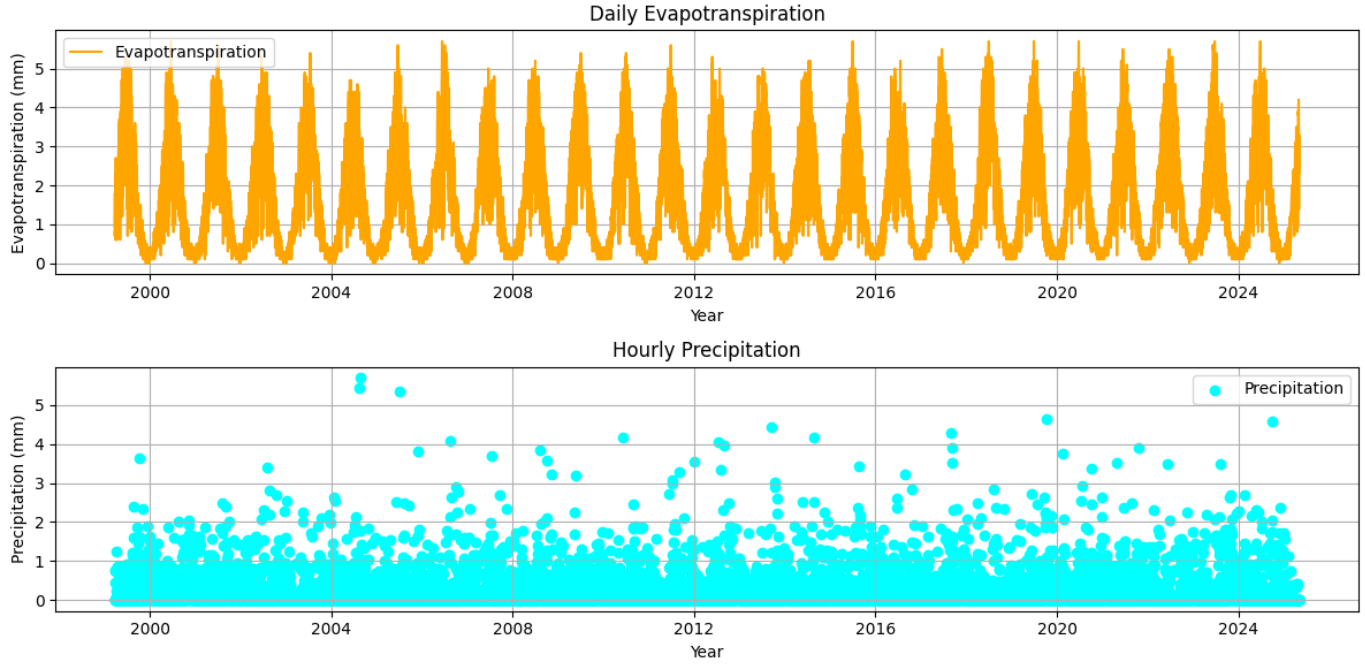


Figure 4.7 Dataset from Berkhout Station, daily evaporation in mm (top) and hourly precipitation in mm (bottom)

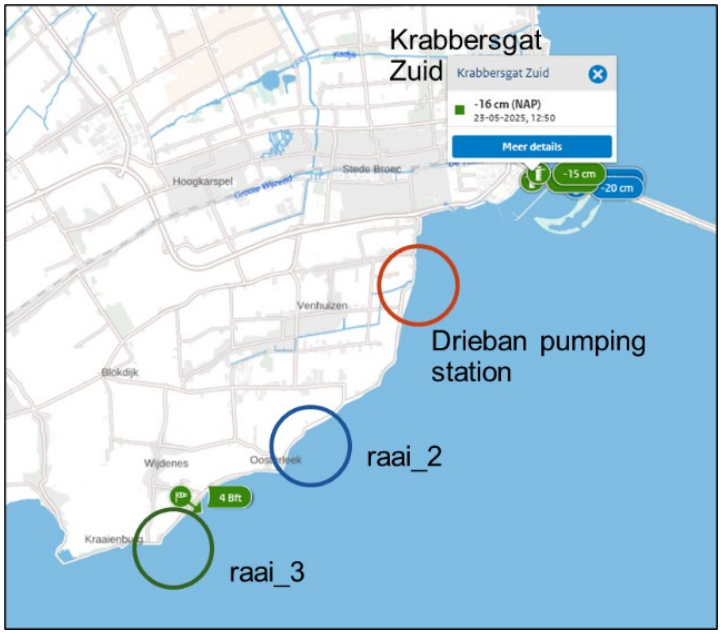


Figure 4.8 Location of raai_3, raai_2, Drieban pumping station and Krabbersgat Zuid observation

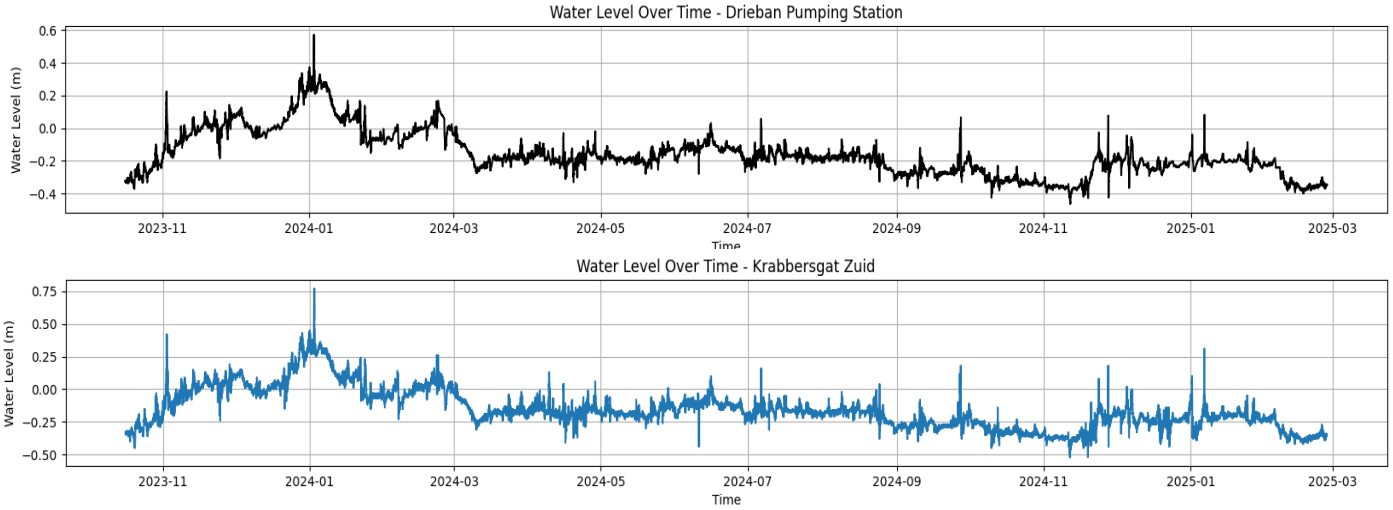


Figure 4.9 Water Level Observation, Drieban (top) and Krabbersgat Zuid (bottom), during same time period (16 Oct 2023 – 26 Feb 2025)

Chapter 5. Results

5.1. Hydrological and Meteorological Variables

Based on the reviewed literature in Chapter 2.1, the phreatic level in dikes surrounding Markermeer and IJsselmeer is influenced primarily by water level fluctuations, precipitation patterns, evapotranspiration rates, internal material permeability, dike geometry, and regional hydrodynamic conditions. Each of these variables interacts to determine groundwater movement within the dike body, affecting pore pressures and thus inner slope stability.

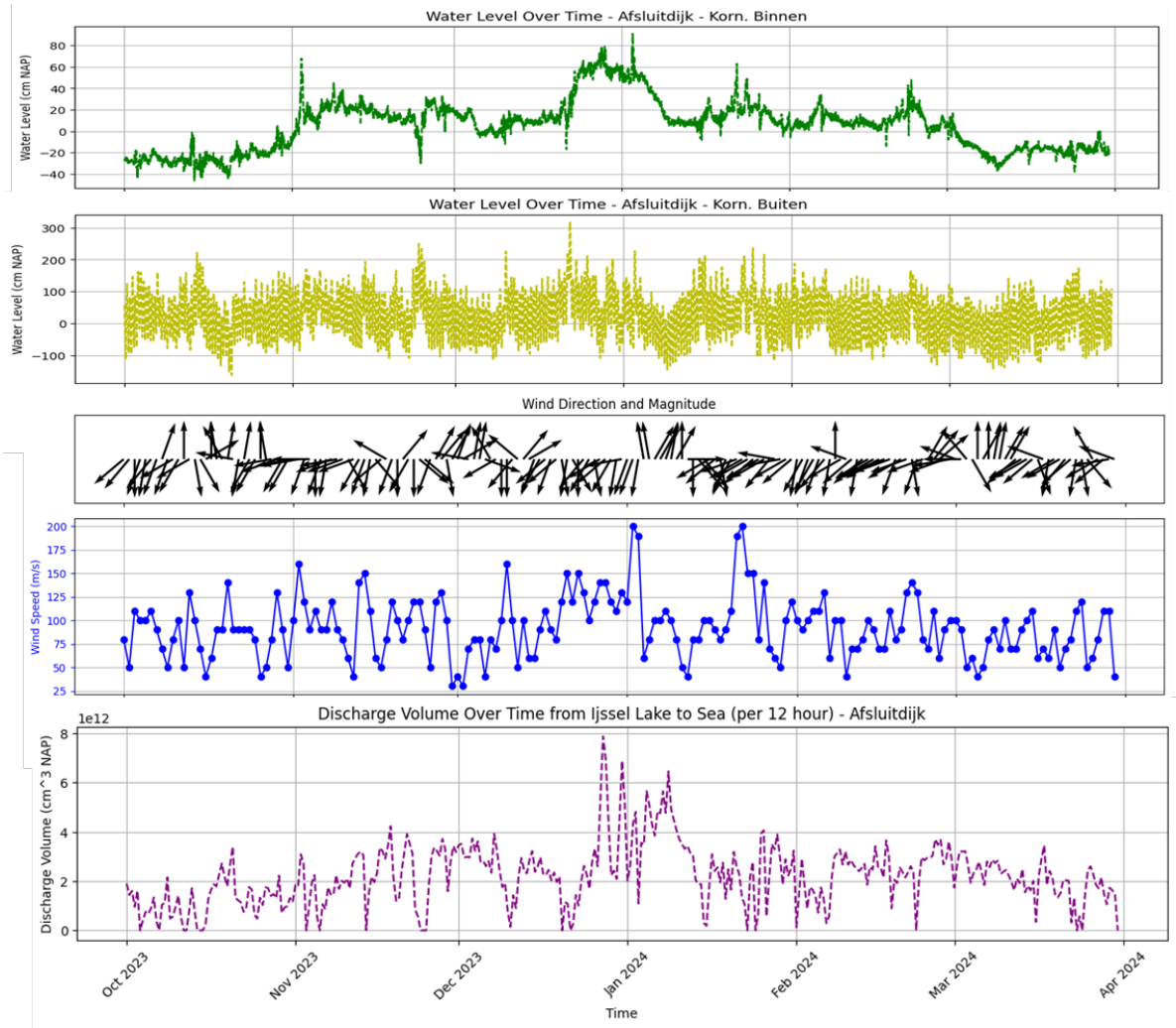


Figure 5.1 Discharge Volume over Time compared with Water Level inside Afsluitdijk (Korn. Binnen), outside of Afsluitdijk (Korn. Buiten) and Wind Direction and Magnitude from The Kooy weather Station. All dataset is in period October 2023 – April 2024.

Figure 5.1 shows the discharge volume over time between October 2023 until April 2024 in Afsluitdijk. Interesting pattern is observed during end of December 2023 when the water level inside the Afsluitdijk increase rapidly and stay high until decreasing during first-second week of January 2024. At that moment, the water level outside of Afsluitdijk is relatively stable and slightly decreasing during end of first week January 2024 while the wind speed is lower. Sharper movement at the discharge volume together with high wind speed and wind direction towards IJsselmeer means that the amount of water flowing outside of IJsselmeer is limited and controlled by sluice. After the wind speed is lower (first-second week of January 2024), the discharge is gradually decrease together with Korn. Binnen water level measurement which indicates water inside IJsselmeer can flow outside of sluice.

5.2. Correlation and Copula Fitting

Correlation is calculated between hourly rainfall and hourly water level, and for subsequent cumulative day.

The result of correlation calculation is shown below:

Table 5.1 Correlation Comparison between Pearson, Spearman and Kendall based on Statistics value and p-value. All comparison were calculated compared to hourly water level.

Precipitation Type	Coefficients			p-value		
	Pearson	Spearman	Kendall	Pearson	Spearman	Kendall
Hourly	0.0663	0.033	0.0261	1.25E-221	2.00E-56	1.79E-55
7-day sum	0.1753	0.1316	0.089	0	0	0
14-day sum	**0.1973**	**0.147**	**0.0994**	**0**	**0**	**0**
21-day sum	0.1892	0.1316	0.0891	0	0	0
28-day sum	0.1694	0.105	0.0712	0	0	0

Based on the result, all coefficients value is positive and statistically significant (p-value less than 0.05 such that null hypothesis is rejected) which shows evidence of linear relationship. However, the coefficient value is relatively low (less than 0.2) which means precipitation alone doesn't fully explain the changes in Markermeer water level. Between five precipitation type, 14-days sum shows strongest correlation which suggests that Markermeer water level responds to cumulative precipitation.

To check whether there is lag effect, for each precipitation type, a lag was imposed from 0 until 14 days. The result is shown in Appendix A. Overall result shows that the correlation decrease steadily with each increasing lag, and by lag larger than 7 days most correlations are close to zero or even slightly negative. This suggests the lake responds to recent cumulative precipitation, but not significantly delayed rainfall.

After calculating the correlation, copula analysis is executed for 14-day sum condition. Both rainfall and water level datasets are converted to uniformly distributed data by dividing rank of the data with total sample plus one (Figure 5.2).

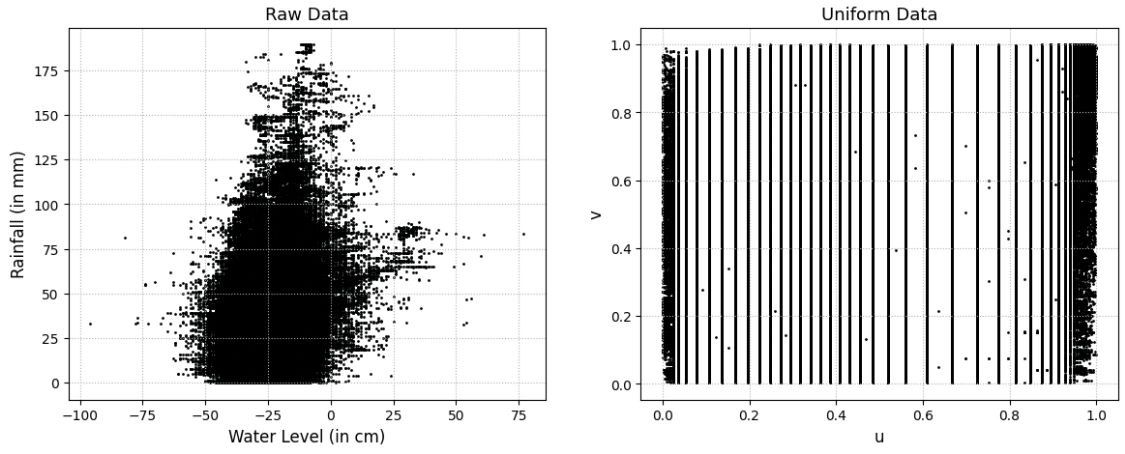


Figure 5.2 Data before transformed (left) and after transformed to uniformly distributed data (right)

In the previous figure, the pattern before transformed shows dense cluttering near the center, with water level spread around -50 cm until 0 cm NAP and rainfall spread between 0 and 10 mm. The right figure shows uniformly transformed dataset in u and v space evenly spread in [0,1] range. Strong vertical banding is observed due to many repeated values in water level measurement. The vertical bandings happen when rank transformation assigns the same rank to duplicate values, and many identical water level values collapse into the same vertical strip when converted to uniform values.

From the uniformly distributed data, multiple bivariate copula family then fitted to existing dataset using *pyvinecopulib* module. The result then checked and compared to determine best copula shown in Table 5.2.

Table 5.2 Model selection test using AIC between chosen Copula family

Rank	Copula Family	AIC
1	**BB8**	**-9249.3872**
2	BB6	-8085.9881
3	Joe	-8067.712
4	BB7	-8013.2814
5	Gumbel	-7783.4721
6	BB1	-7781.4546
7	Gaussian	-6981.3706
8	Student (t)	-6560.7665
9	Frank	-4943.7569
10	Clayton	-1141.4677

Based on Table 5.2, copula BB8 with parameter value of θ = 1.3 and δ = 0.97 is the best model according to AIC test where lower AIC shows better tradeoff between fit and complexity. Compared to other copula families, the BB8 copula is capable of capturing both upper and lower tail dependence, as well as asymmetric dependence. In the context of rainfall and Markermeer water levels, this asymmetric dependence helps explain events where heavy rainfall can lead to a rise in water level, but the reverse is not necessarily true. The presence of upper and lower tail dependence allows the BB8 copula to model the joint occurrence of extreme rainfall and extreme water levels. However, due to the asymmetric nature of the dependence, these co-occurring extreme events may not consistently appear.

On the other hand, the Gaussian and Student's t copulas, which assume symmetric and elliptical dependence, show a poor fit. This suggests that Markermeer water levels do not respond uniformly across the entire range of precipitation values. Among the tested copulas, the Clayton copula performed the worst. This is likely because Clayton focuses on modeling lower tail dependence, such as extreme droughts. This does not align with the characteristics of the Markermeer, where water levels are more sensitive to extreme or cumulative rainfall events (upper tail) rather than to moderate or symmetric changes.

To represent how well the copula fitting, BB8 copula model with previous parameter value is used to generate samples, which then compared with original observations.

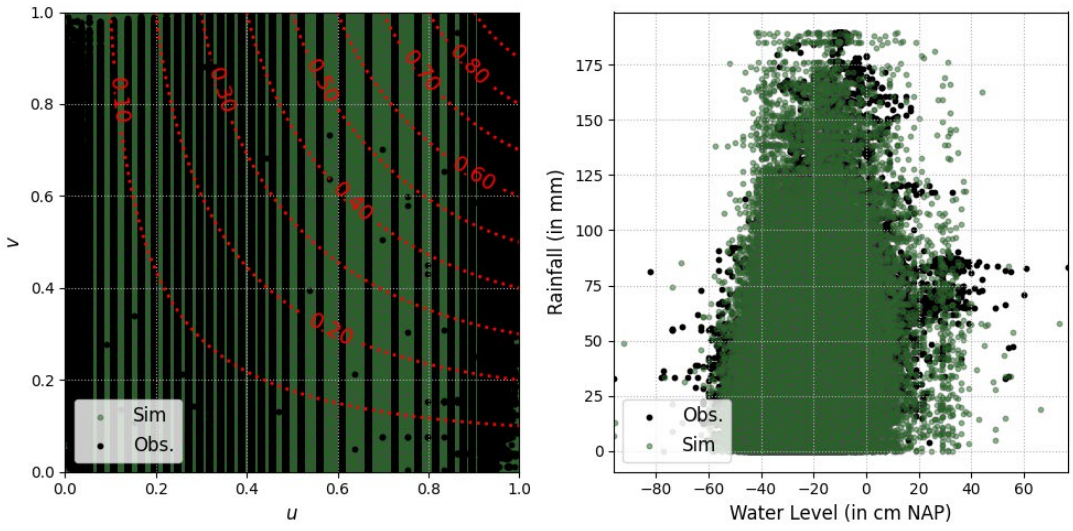


Figure 5.3 Simulated vs Observed Dataset, in union space (left) and normal space (right)

Based on Figure 5.3, the simulated dataset shows no vertical banding compared to observation which is expected from continuous function. The contour curves reflect the dependency structure captured by the

copula which in this case is asymmetric and nonlinear. On the right plot, the spread of simulated dataset follows closely to observation, which shows that the copula model preserves the marginal distributions, since it simulates in (u, v) space and transforms back using real data distributions.

Marginals of water level and precipitation are compared with Q-Q and EDCF plot:

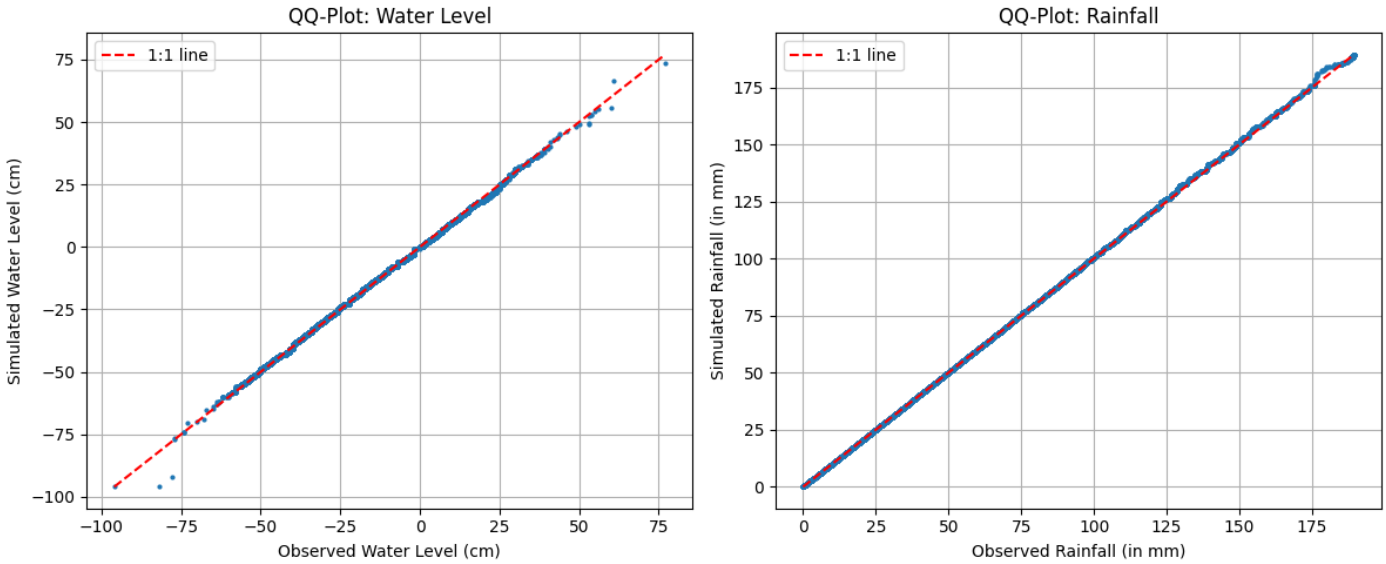


Figure 5.4 Q-Q Plot between simulated dataset and observed dataset

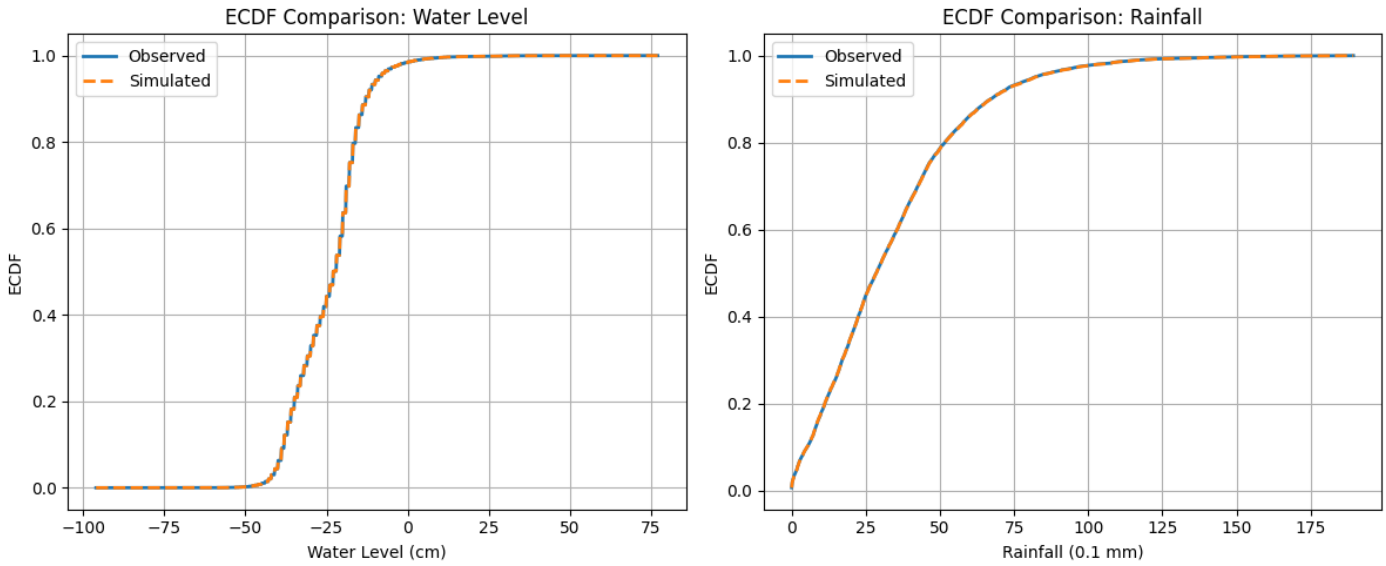


Figure 5.5 EDCF Comparison between simulated dataset and observed dataset

From Figure 5.4, the empirical quantile pairs closely follow the red diagonal line, which means simulated water level distribution matches the observed distribution very well, suggesting an accurate marginal fit and effective inverse-transformation from the copula. In Figure 5.4, the empirical distributions of both observed and simulated water levels and rainfall are nearly identical, further demonstrating that the BB8 copula provides a good representation of the marginal distributions for both variables. Overall, BB8 copula simulation with empirical marginals has successfully reproduced the marginal distributions of both water level and rainfall.

To visualize the empirical BB8 copula, a 3-dimensional figures are plotted between observed dataset and simulated dataset, and goodness-of-fit is calculated and compared by Cramer-von-Mises test:

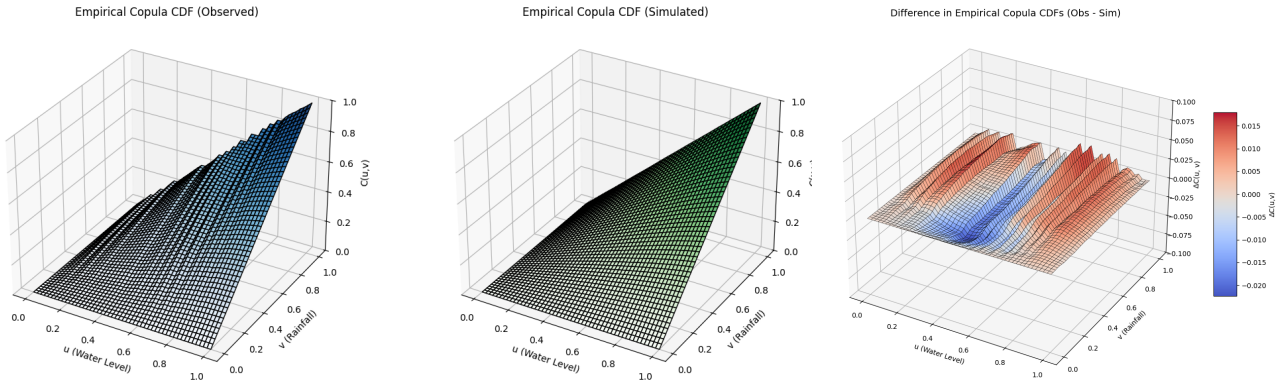


Figure 5.6 3-dimensional representation of BB8 copula between observation (left plot), simulated (middle plot) and the difference (right plot)

Table 5.3 Goodness-of-Fit test using Cramer-von Mises (CvM) Criterion

Rank	Copula	CvM Criterion
1	**BB8**	**7.4e+06**
2	Gumbel	8.8e+06
3	Clayton	2.2e+07
4	Gaussian	1.0e+07

In the Figure 5.6, the BB8 copula model can represent the empirical dependence well enough with similar shape and peak. The difference shown in right plot are mostly centered around zero, with some slight overestimation (blue shadow) in the mid-range region and underestimation in the water level marginal due to asymmetry and rounded observation numbers. No major bias or structural mismatch is observed.

In Table 5.3, BB8 has the lowest CvM score, confirming it's the best at capturing the joint structure of rainfall and water level, including possible tail asymmetry and nonlinear patterns. Gumbel distribution also fits reasonably well, especially in the upper tail, but still less flexible than BB8. Meanwhile, Gaussian does not model tail dependence or asymmetry, and Clayton focuses on lower tail dependence only, explaining the poor fit for both model.

5.3. Groundwater Model

5.3.1. Determining Model Configuration

The result of each configuration model predictions compared with real observation from HB27_PB1 well during calibration stage is shown in Figure 5.7.

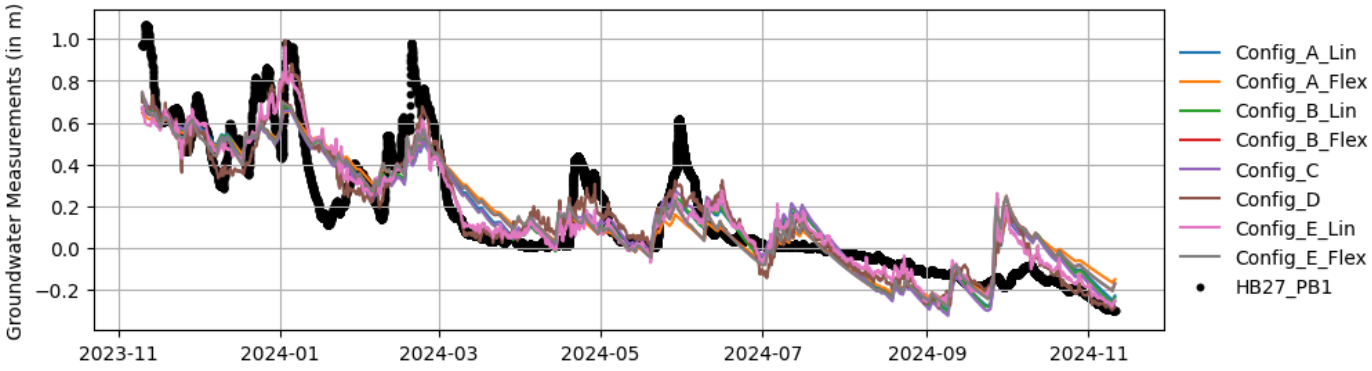


Figure 5.7 Time series plot of observation from HB27_PB1(black markers) vs model prediction for each Configuration (line graphs) in Calibration stage.

Figure 5.7 presents the observed groundwater measurements for well HB27_PB1 (black markers) alongside the simulated groundwater levels from eight different model configurations (*A_Lin* to *E_Flex*, represented by colored lines). The dataset spans from 2023-11-10 until 2024-11-10 following calibration dataset period, with groundwater levels reported in meters.

The observed groundwater levels show quick rise followed by highest level measurements around the middle of November 2023 reaching 1.1 m. This spike pattern also observed repeatedly during November 2023 until March 2024, with notable peaks around 1.0 m occurred in January 2024 and end of February 2024. After this period, the groundwater tends to decline gradually, with smaller peaks observed in end of April 2024 and beginning of June 2024.

The prediction models generally follow the trend of observed groundwater levels. All models can reproduce the four observed peak of January 2024, end of February 2024, end of April 2024 and June 2024, with varying level of deviations. All models consistently predict higher value during the rapid decrease in the middle of January 2024, and lower peak value during last three observed peaks. Only configuration *D* and *E_Lin* capture the peak of January 2024 and rapid decline around March-April 2024. Configuration *D* also predict lower groundwater level in December 2023 just below 0.3 m. In the low season starting from end of June 2024, all prediction models show more erratic behavior with higher peak during middle of July 2024, lower value in September 2024 and rapid increase at October 2024.

No significant difference is observed between Lin configuration (using Linear recharge model) and Flex configuration (using FlexModel recharge model).

The goodness-of-fit metric from calibration stage is shown in Table 5.4.

Table 5.4 Goodness-of-Fit Comparisons between each Configuration in Calibration stage. Bold number shows the best value for each metric.

Configuration Setting	R^2	$\Delta AICc$
A_Lin	0.77	2740.42
A_Flex	0.74	3664.08
B_Lin	0.77	2643.59
B_Flex	0.75	3294.25
C	0.77	2476.12
D	**0.83**	**0.0**
E_Lin	0.8	1188.27
E_Flex	0.75	3303.28

Table 5.4 presents the goodness-of-fit comparisons for each configuration during the calibration stage. Configuration D indicates the best fit among other configurations with highest R^2 value of 0.83 and lowest $\Delta AICc$ value of 0. Meanwhile, Configuration A_Flex performs the worst with lowest R^2 of 0.74 and highest $\Delta AICc$ value of 3664.08. The value of R^2 for other configuration settings ranging from 0.75 to 0.80 and $\Delta AICc$ values between 1188.27 and 3294.25.

Validation test for each configuration is set to run using groundwater observation at period 2024-11-11 until 2025-02-25. The result of each configuration model predictions compared with real observation from HB27_PB1 well during observation stage is shown in Figure 5.8.

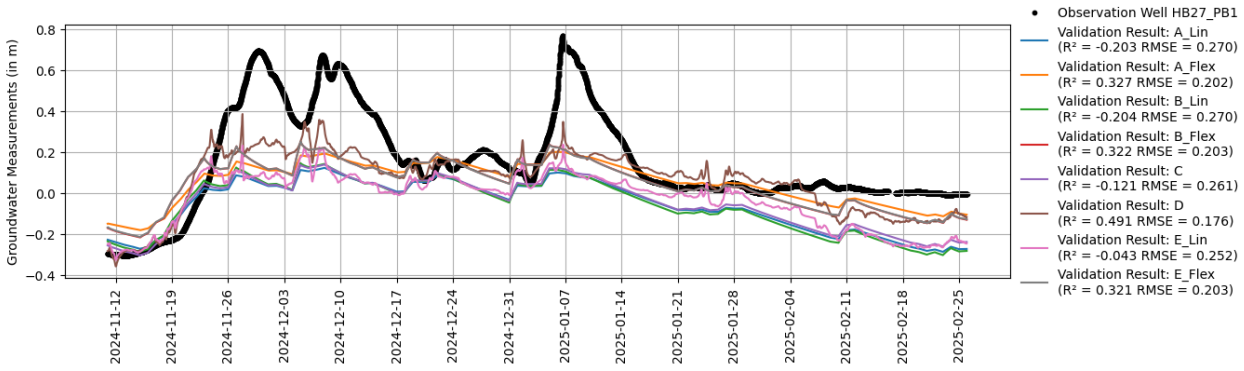


Figure 5.8 Observed Groundwater Measurement from HB27_PB1 versus Model Prediction from all configurations during Validation Stage

Figure 5.8 displays the observed groundwater measurements together with prediction from eight model configurations, with goodness-of-fit metrics R^2 and $RMSE$ included. The original measurement shows a clear increasing trend from late November 2024, peaking around early January 2025, and then declining through February 2025. The highest observed levels reach approximately 0.75 m, while the lowest fall observed at 2024-11-12 below -0.2 m.

While all models can replicate the gradual increase in groundwater level at the middle of November 2024, they consistently underestimate the groundwater level during observed peaks at December 2024 and early January 2025. For example, the observed peak near 0.75 m is consistently underpredicted by all models, with simulated peaks generally not exceeding 0.4 m. Following this peak, models such as *A_Lin* (blue line), *B_Lin* (green line) and *C* (purple line) show consistently lower simulated levels compared to observations from end of November 2025 through February 2025.

With regards to model performance, R^2 metric show relatively low value, with some configurations in the negative range, starting from -0.204 (configuration *B_Lin*) with the highest value is 0.491 (configuration *D*). For $RMSE$ values, all configurations show considerable deviations ranging from 0.176 m from configuration *D* until 0.270 m measured at configuration *A_Lin* and *B_Lin*. Comparison between all configurations shows Lin configuration (using Linear recharge model) models consistently measured at negative R^2 value and higher value of $RMSE$ from 0.252 until 0.270. Meanwhile, configuration D in validation stage also indicates the best fit with highest r-squared value among other configurations at 0.491 and lowest $RMSE$ at 0.176.

5.3.2. Creating the Groundwater Model

Four scenarios were proposed during the development of the groundwater model to assess the influence of different nearby water level sources on the selected cross-sections. In Scenario 1, a groundwater model was created for cross-section *raai_3*, incorporating an additional water level stress model based on data from Drieban. Scenario 2 also focused on *raai_3* but used the water level stress model from Krabbersgat Zuid instead. For Scenario 3, the groundwater model was developed for cross-section *raai_2*, with the water level stress model taken from Drieban. Lastly, Scenario 4 involved modeling *raai_2* using the water level stress model from Krabbersgat Zuid.

Scenario 1 vs Scenario 2

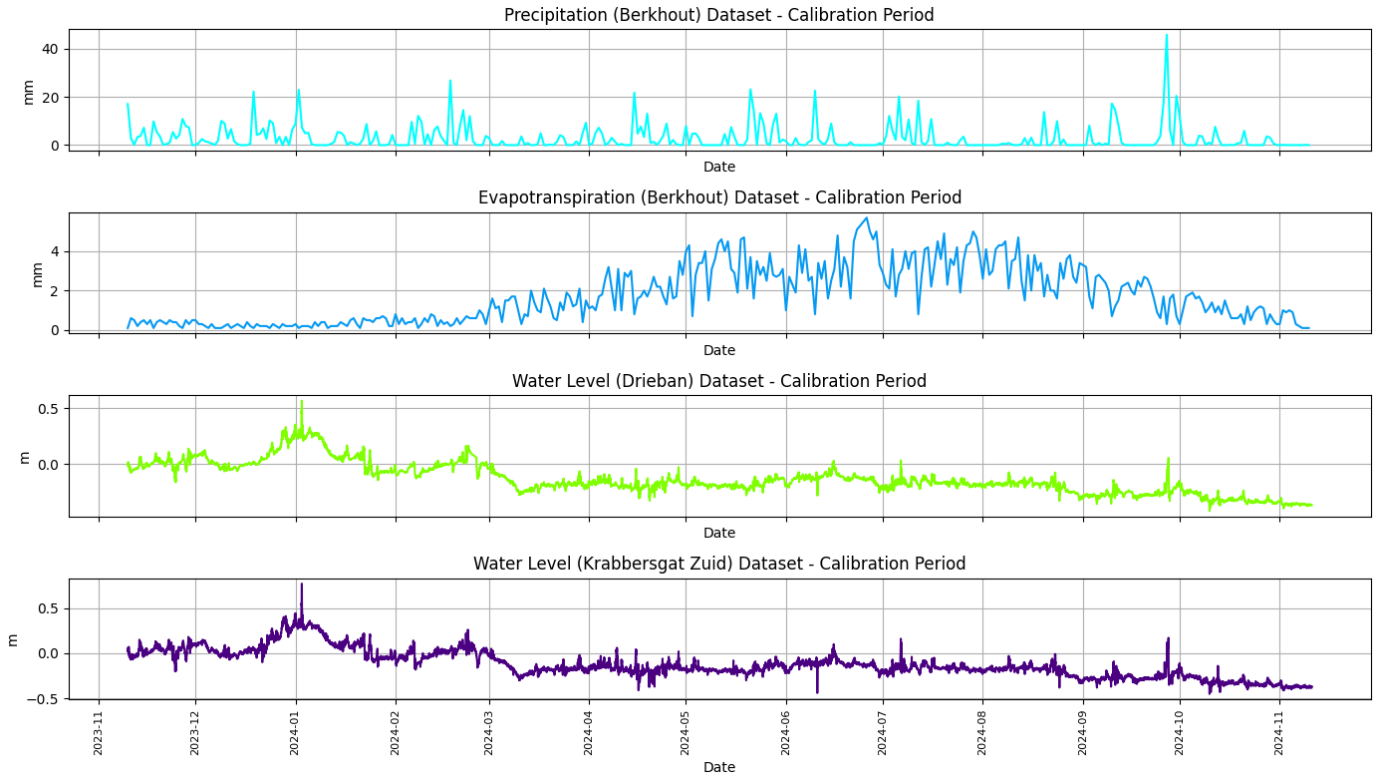


Figure 5.9 Precipitation, Evapotranspiration and Water Level plot during Calibration period (2023-11-10 until 2024-11-10)

The groundwater observation wells in *raai_3* for both scenario 1 and 2 is set up using configuration D: separate stress model with water level measurement as additional stress model. Similar to previous run, the groundwater level dataset is split between calibration stage and validation stage. In this first stage, the result is presented as mosaic graphs which combine both Scenario 1 and Scenario 2 result per each observation well. Input dataset for precipitation, evapotranspiration and water level used in calibration stage is also presented in Figure 5.9.

After calibrating the Pastas models for Scenario 1 and Scenario 2, comparative mosaic plots were generated. Each mosaic displays the observed and simulated groundwater heads with residuals in the upper-left panel; the upper-right panel reports goodness-of-fit statistics and calibrated parameter values for both scenarios. The left column shows the time-varying stress contributions of precipitation, evapotranspiration, and chosen local water level (Drieban pumping station for Scenario 3, Krabbersgat Zuid for Scenario 4), while the right column presents the corresponding step-response functions. Typical mosaic plot is shown in the Figure 5.10, and the detailed result is shown in Appendix C.

Across four observation wells (MB014-PB1, MB014-PB2, HB027-PB1, MB015-PB2), both scenarios reproduce the seasonal groundwater dynamics and many sub-seasonal fluctuations. Deviations are concentrated around sharp peaks or rapid drawdowns; residuals for MB014-PB1 are mostly within ±0.1 m, with a single excursion near −0.12 m. Visual fit is high at MB014-PB1 (R^2 = 0.96 for S1; 0.95 for S2), moderate at MB014-PB2 (0.58; 0.59), and good at HB027-PB1 (0.83; 0.85). For MB015-PB2, the narrative indicates accurate predictions with some underestimation between March–August 2024; the R^2 value was not reported in the excerpt.

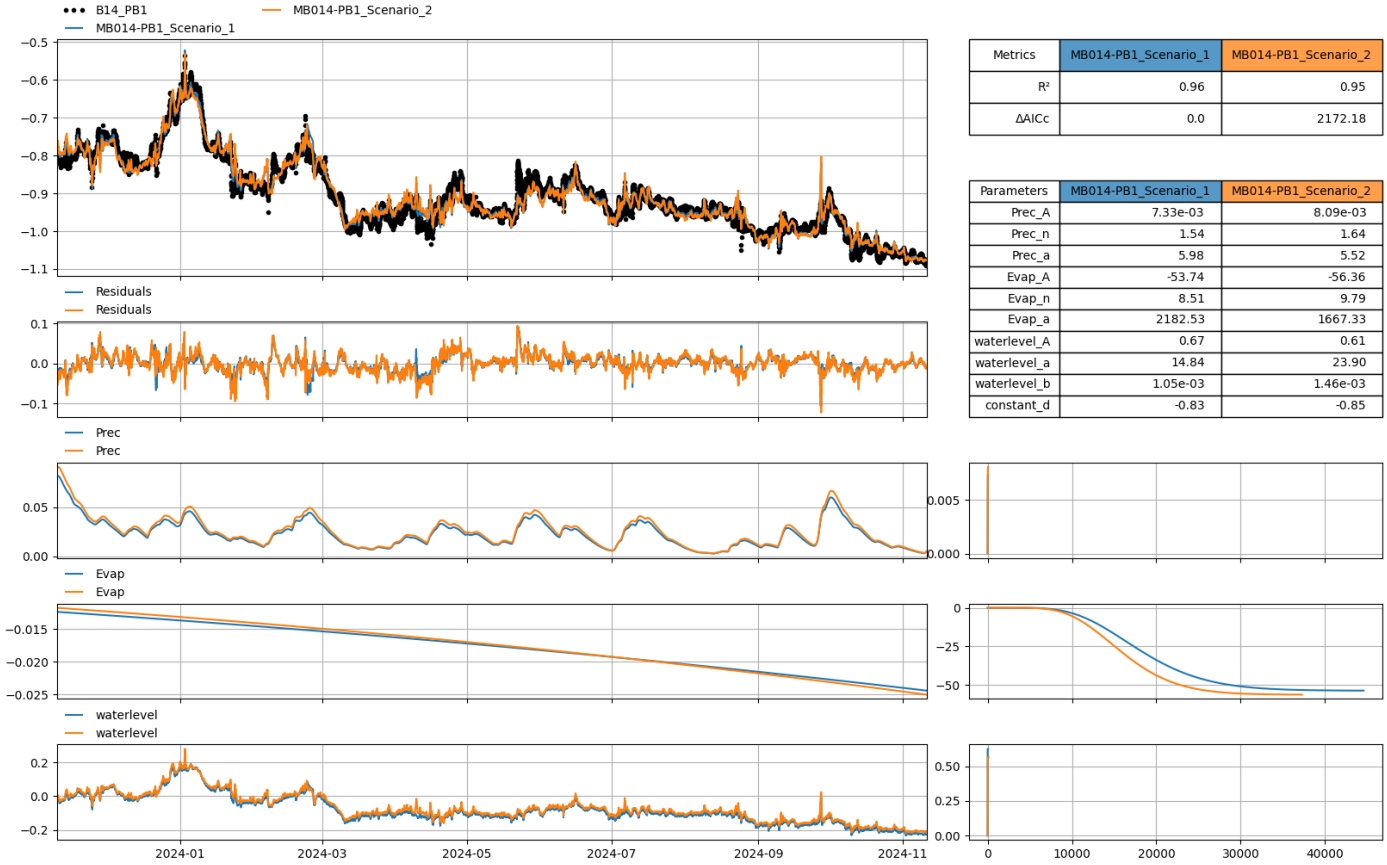
Process attribution is consistent with hydrogeological expectations. Precipitation and surface-water level changes contribute positively to heads, whereas evapotranspiration exerts a negative influence. Step-response plots for precipitation and water level are near-vertical (very short characteristic times; shallow system behavior), while evapotranspiration responses are slow and negative, approaching a stable level (e.g., ~−50 in the MB014-PB1 step response). Scenario-specific differences are subtle but systematic: at MB014-PB1 the Scenario 1 system is slightly "slower" (blue curve) than Scenario 2 (orange), whereas at MB014-PB2 the precipitation contribution is much larger and smoother under Scenario 2; Scenario 1's precipitation amplitude is near zero (*Prec_A* ≈ 4.54 × 10^{-3}) with larger shape and smaller scale parameters. At HB027-PB1, evapotranspiration is far less negative and much faster in Scenario 1 (~150 days to plateau) than in S2 (~6000 days), implying more persistent evapotranspiration effects under Scenario 2. For MB015-PB2, precipitation and water-level step responses are short (near-instantaneous), and evapotranspiration shows a sinusoidal, net-negative contribution.

R^2 is relatively similar between Scenario 1 and Scenario 2, with small advantages: Scenario 1 at MB014-PB1; Scenario 2 at MB014-PB2 and HB027-PB1. Differences arise primarily from how each scenario partitions variance between precipitation and evapotranspiration and from their effective response times. In particular, Scenario 2 often attributes a smoother, stronger role to precipitation (MB014-PB2) and a more persistent evapotranspiration influence (HB027-PB1), whereas Scenario 1 tends toward faster evapotranspiration dynamics and, at some sites, slightly closer tracking of abrupt rises (MB014-PB1).

Overall, all models in this calibration stage show relatively good approximation toward training dataset, with R^2 value all higher than 0.5. However, stress contribution pattern for some models is atypical, such as precipitation pattern for MB014-PB2, and evapotranspiration pattern for MB014-PB1, MB014-PB2 and HB027-PB1. Statistical comparisons (R^2 and $\Delta AICC$) for Scenario 1 and 2 is shown in Table 5.5.

Table 5.5 Statistical comparisons by groundwater well for Scenario 1 vs Scenario 2. Bold letters indicate better result between scenarios.

Groundwater Well	Scenario 1		Scenario 2	
	R^2	$\Delta AICc$	R^2	$\Delta AICc$
MB014-PB1	**0.96**	**0.00**	0.95	2172.18
MB014-PB2	0.58	250.49	**0.59**	**0.00**
HB027-PB1	0.83	1145.81	**0.85**	**0.00**
MB015-PB2	**0.83**	**0.00**	0.83	192.24



Metrics	MB014-PB1_Scenario_1	MB014-PB1_Scenario_2
R^2	0.96	0.95
ΔAICc	0.0	2172.18

Parameters	MB014-PB1_Scenario_1	MB014-PB1_Scenario_2
Prec_A	7.33e-03	8.09e-03
Prec_n	1.54	1.64
Prec_a	5.98	5.52
Evap_A	-53.74	-56.36
Evap_n	8.51	9.79
Evap_a	2182.53	1667.33
waterlevel_A	0.67	0.61
waterlevel_a	14.84	23.90
waterlevel_b	1.05e-03	1.46e-03
constant_d	-0.83	-0.85

Figure 5.10 Typical Mosaic Plot Comparison between Scenarios

After calibration stage, the validation simulation is executed with dataset from 2024-11-11 until 2025-02-25, which is illustrated in Figure 5.11. The result for validation is shown in Figure 5.12.

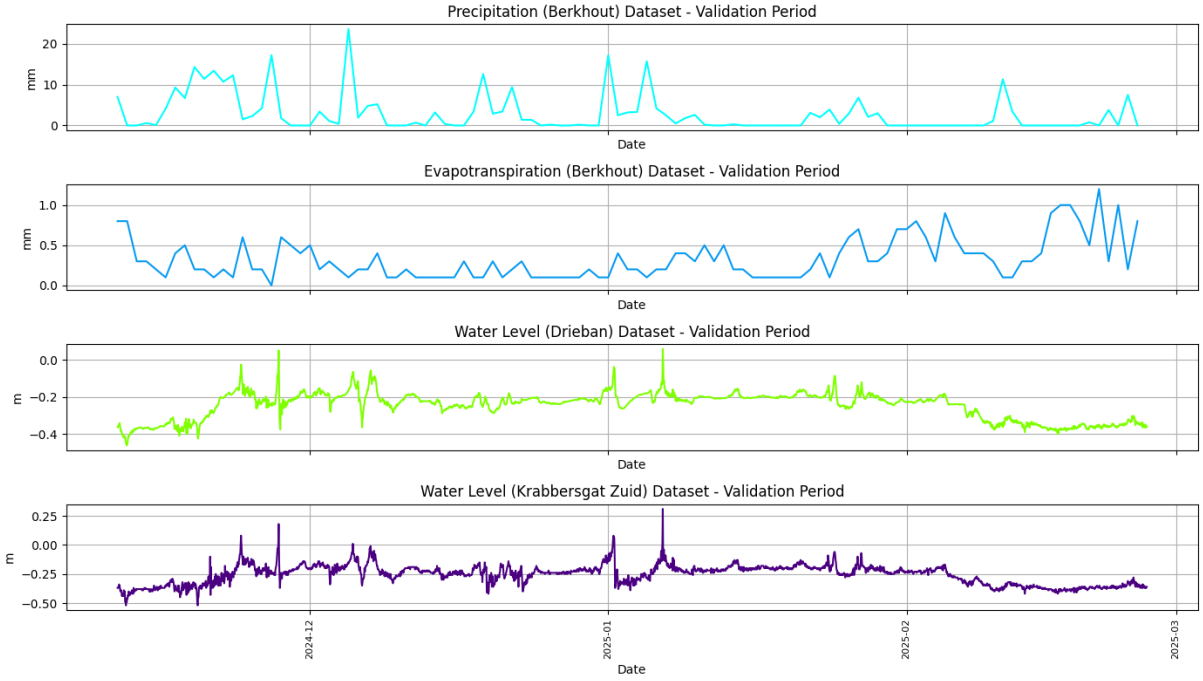


Figure 5.11 Precipitation, Evapotranspiration and Water Level plot during Validation period (2023-11-10 until 2024-11-10)

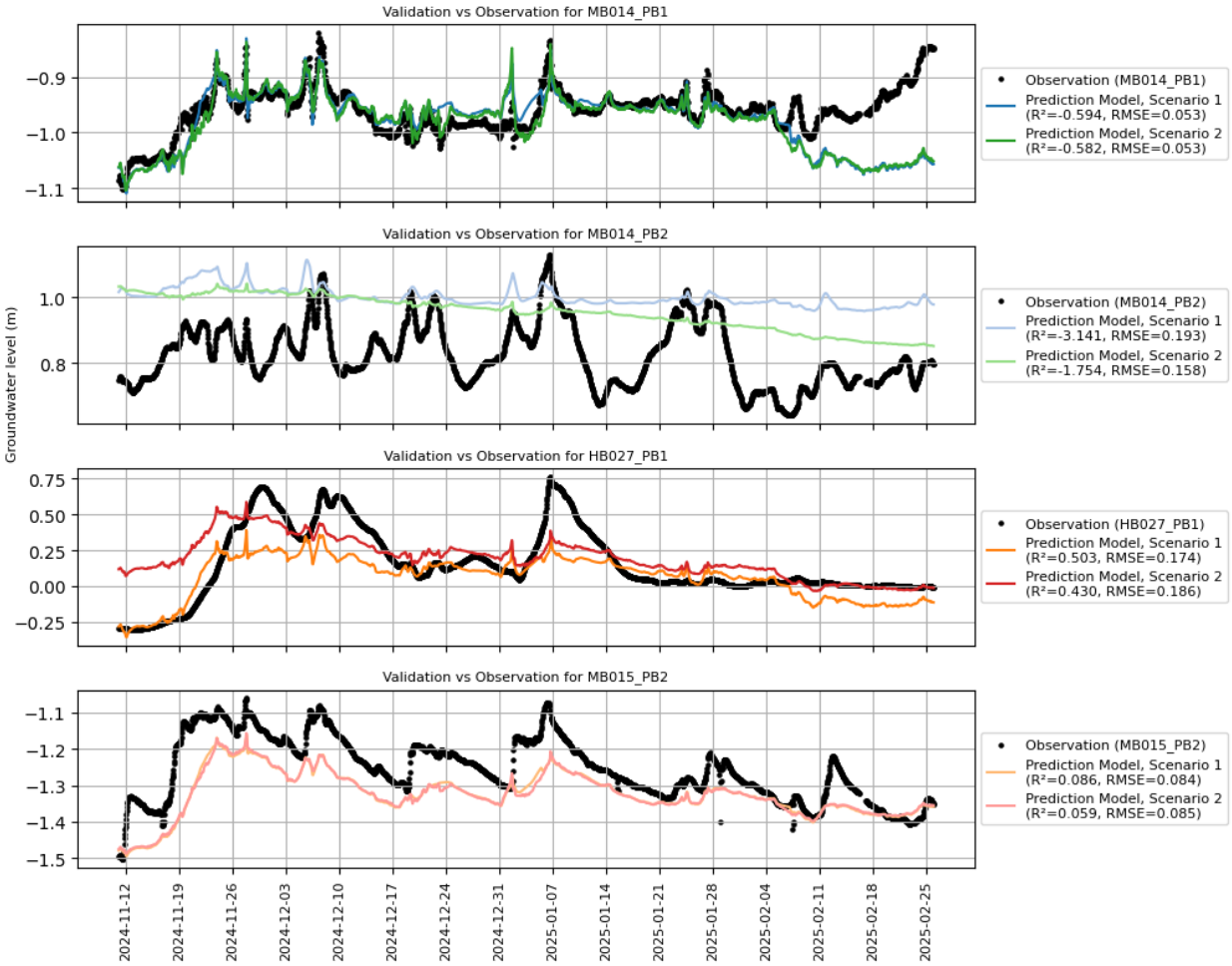


Figure 5.12 Observation vs prediction for validation stage, Scenario 1 vs 2.

In the first plot, the prediction models follow the observation data very closely but starting from 2025-02-04 major deviation is detected where downward trend for both models occur which is opposite of upward trend from original observation. This also reflects on the R^2 value at negative value of -0.786 for Scenario 1 and -0.571 for Scenario 2. In the second plot, the validation dataset shows rapid and significant fluctuations which both model struggle to capture correctly. Scenario 1 veers over higher average value around 1.0 m and almost consistently overpredicted, such that the R^2 value is significantly low at -3.141. Scenario 2 fare a little bit better with higher but still negative R^2 and downward trend observed, but both model failed to predict correct peak and troughs. In third plot (HB027-PB1), both models show better performance with highest R^2 value compared to other groundwater location. Nonetheless, the models remain insensitive to extreme events, and Scenario 2 model shows significant overpredict at the beginning. For fourth plot, while both models follow the general trend of the observed dataset reasonably well, they consistently underpredict groundwater levels throughout the entire period.

To improve model performance, parameter adjustments were added to the model, with main focus to maintain physical properties especially evapotranspiration stress contribution. Parameter adjustment is done by setting initial, minimum value and maximum value for specific parameters during model solver. The adjusted model parameters are shown in Appendix B.

The hindcasting simulation result before parameter calibration is shown in Figure 5.13, and after parameter calibration in Figure 5.14.

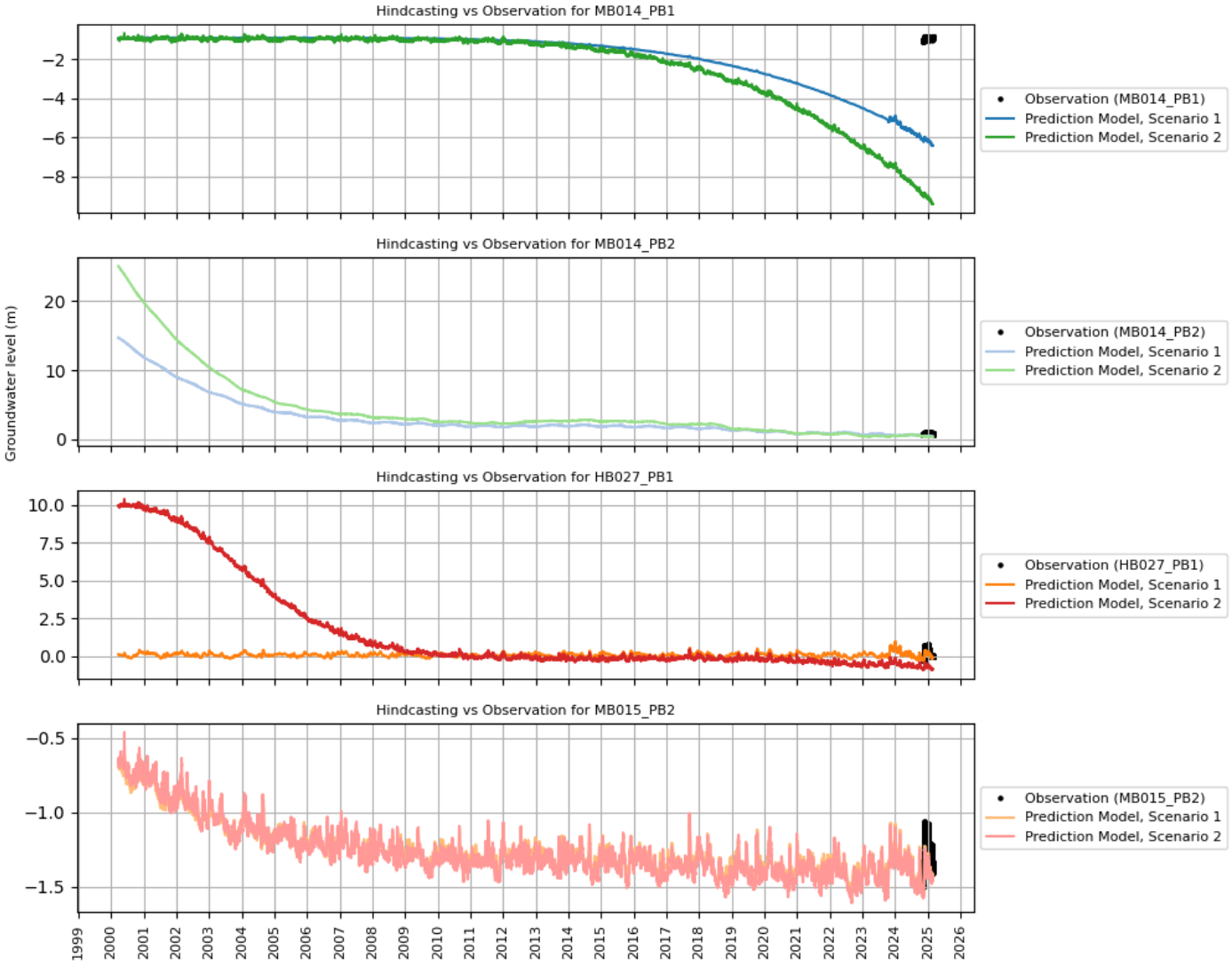


Figure 5.13 Hindcasting simulation result for Scenario 1 and 2 before manual calibration

In the first set of hindcast plots (before calibration), both Scenario 1 and Scenario 2 reproduce the broad groundwater trends but often miss the observed values: Scenario 1 is closer at MB014_PB1, HB027_PB1, and MB015_PB2, while Scenario 2 better matches MB014_PB2. After manual adjustment of model parameters (second set of plots), the simulated curves from both scenarios fall much closer to the black-dot observations at all four sites. In MB014_PB1 and MB014_PB2, there is a slightly downward trend of prediction from 1999 to 2025, and both prediction model show almost exact same prediction value. As the only difference between Scenario 1 and 2 is the water level dataset, it shows that after manual calibration, the stress contribution from water level to the groundwater level prediction is almost negligible, and the prediction mainly influenced by precipitation and evapotranspiration stress. Calibration reduces the bias and variance of the predictions, so that the models now capture both the mean levels and short-term fluctuations almost equally well across every monitoring location.

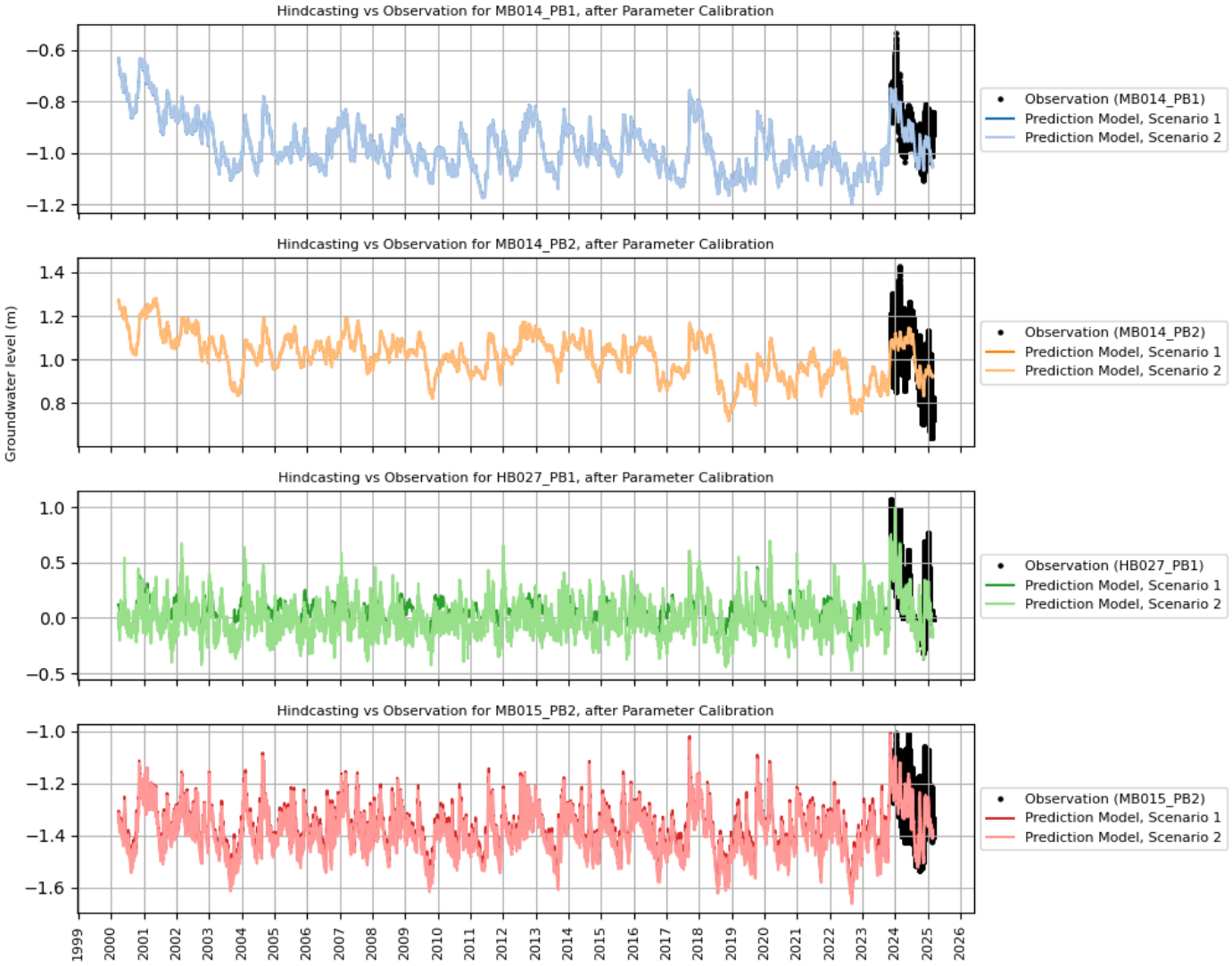


Figure 5.14 Hindcasting simulation result for Scenario 1 and 2 after manual calibration

Scenario 3 vs Scenario 4

Both scenario 3 and scenario 4 use groundwater data from *raai_2* cross section, which consist of 6 groundwater well: MB012-PB1, HB025-PB1, MB026-PB2, MB026-PB1, MB013-PB2, MB013-PB1. The setup is mostly similar to scenario 1 and 2, where scenario 3 use water level dataset from Drieban and scenario 4 from Krabbersgat Zuid. The calibration dataset is shown previously in Figure 5.9, and validation dataset refers to Figure 5.11.

Similar to previous scenario comparison, the typical mosaic plot is same as Figure 5.10, and the detailed result is shown in Appendix C.

Across six observation wells, both Scenario 3 and Scenario 4 reproduce the seasonal groundwater dynamics and many sub-seasonal fluctuations. Residuals are typically small, around ±0.1–0.2 m, with the largest deviations coinciding with sharp, isolated peaks or drops in head. Erratic, short-lived observations are not consistently captured, suggesting local rapid processes (e.g., runoff/infiltration or site-specific boundary effects) that are difficult to resolve by the model.

In terms of process, stress-contribution and step-response plots show a fast, positive head response to precipitation and local water level changes and a slower, negative response to evaporation. Scenario 4 generally approaches equilibrium marginally sooner than Scenario 3, indicating slightly leaner/stronger weighting on water level as additional boundary stresses.

Where Scenario 4 is favoured (MB026-PB2, MB013-PB2), the calibration typically down-weights precipitation (lower precipitation amplitude/scale) and places relatively greater weight on evaporation and/or boundary

stresses, while achieving a lower corrected $\Delta AICc$ for comparable or higher R^2. At sites where Scenario 3 is preferred, precipitation shows a stronger influence and the $\Delta AICc$ value favours its slightly higher dynamics.

Model efficiency (R^2) is generally similar between scenarios within each well, so model selection is primarily governed by $\Delta AICc$. Scenario 3 provides the better fit vs complexity balance at four wells (MB012-PB1, HB025-PB1, MB026-PB1, MB013-PB1), whereas Scenario 4 is superior at two wells (MB026-PB2, MB013-PB2). This mixed outcome indicates spatial variability in dominant stresses across the network, justifying per-well model choice.

Overall results show that most model prediction in *raai_2* cross section can reproduce general trend of real-world observation but struggling to capture extreme and erratic movement which is to be expected. All models also show correct physical interpretation of evapotranspiration negative sinusoidal pattern. Water level contribution, compared to Scenario 1 and 2, is less prone to rapid movement which reflected in the initial step response and smoother shape. Statistical comparisons (R^2 and $\Delta AICC$) for Scenario 3 and 4 is shown in Table 5.6.

Table 5.6 Statistical comparisons by groundwater well for Scenario 3 vs Scenario 4. Bold letters indicate better result between scenarios.

Groundwater Well	Scenario 3		Scenario 4	
	R^2	$\Delta AICc$	R^2	$\Delta AICc$
MB012-PB1	**0.76**	**0.00**	0.76	4.93
HB025-PB1	**0.50**	**0.00**	0.50	57.64
MB026-PB2	0.89	1846.24	**0.91**	**0.00**
MB026-PB1	**0.76**	**0.00**	0.76	63.1
MB013-PB2	0.71	1432.76	**0.75**	**0.00**
MB013-PB1	**0.50**	**0.00**	0.49	46.83

After calibration, validation stage is performed for all groundwater wells in *raai_2*.

Figure 5.15 presents the validation results comparing groundwater observations against two predictive model scenarios across six groundwater monitoring wells. For MB12-PB1, both scenarios significantly overpredict groundwater levels, particularly during peak events, resulting in strongly negative R^2 values (-3.318 for Scenario 1 and -5.190 for Scenario 2). In HB25-PB1, sharp peaks observed above 2 meters are poorly represented by both scenarios, which predict relatively stable levels around 1 meter, yielding low R^2 values (0.177 for Scenario 1 and 0.182 for Scenario 2). The validation for MB26-PB2 indicates challenges in capturing observed peaks, with Scenario 1 notably underpredicting (R^2 = -0.208), while Scenario 2 shows modest improvement but still lacks accuracy (R^2 = 0.438). Conversely, MB26-PB1 demonstrates good model performance, with both scenarios closely following observed fluctuations and achieving higher R^2 values (0.617 for Scenario 1 and 0.605 for Scenario 2). For MB13-PB2, both models reasonably track the overall observed groundwater trends, although Scenario 2 (R^2 = 0.473) better captures fluctuations than Scenario 1 (R^2 = 0.406). Lastly, at MB13-PB1, both scenarios show limited predictive accuracy, as neither model accurately replicates short-term fluctuations and the observed downward trend, resulting in low or negative R^2 values (Scenario 1 = -0.048; Scenario 2 = 0.170). Overall, the models vary in predictive skill depending on location, particularly struggling with short-term peaks and rapid groundwater fluctuations.

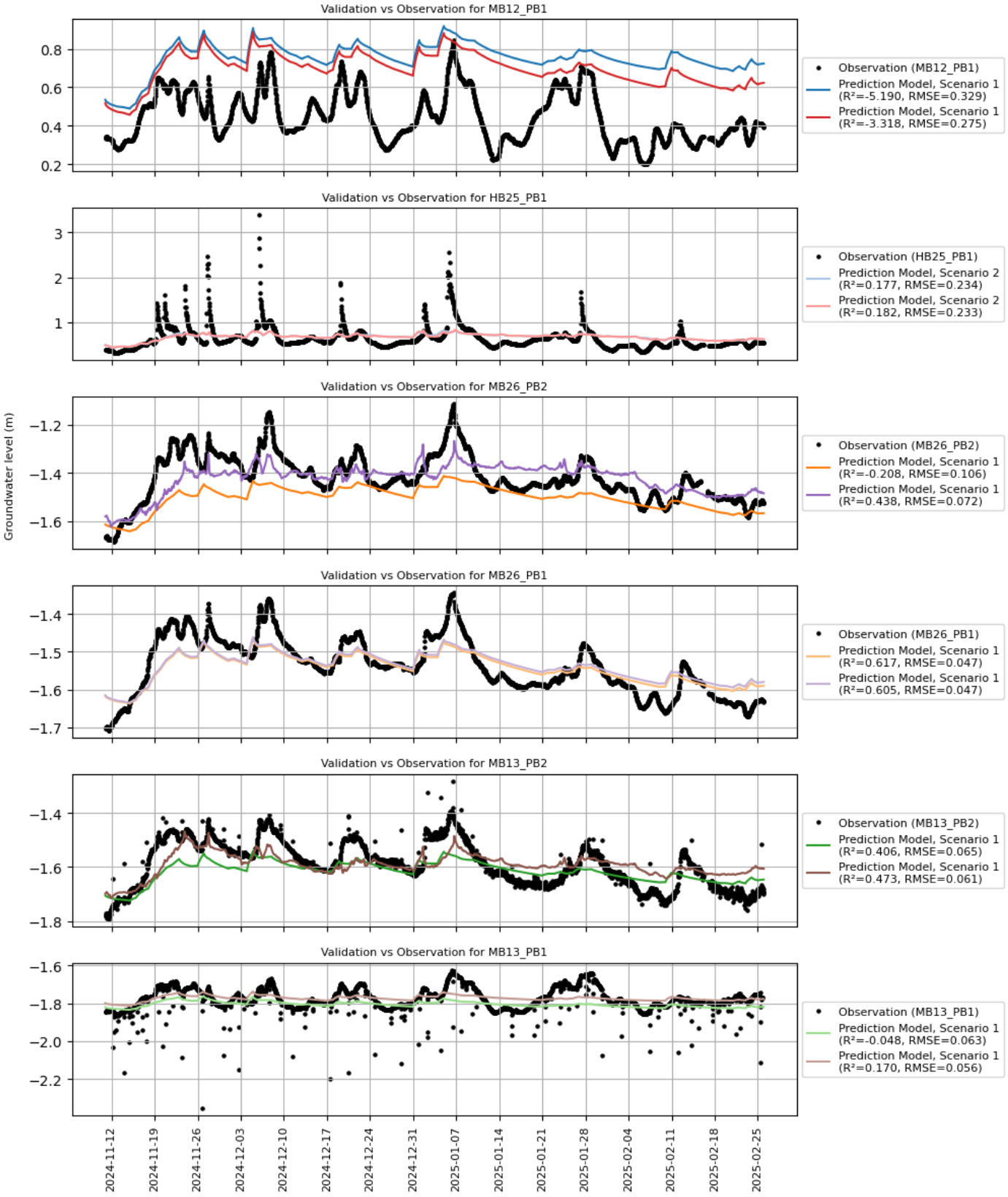


Figure 5.15 Observation vs prediction for validation stage, Scenario 3 vs 4

To improve model performance, parameter adjustments were added to the model. Unlike Scenario 1 and 2, the evapotranspiration properties already show correct physical properties during calibration stage such that this parameter adjustment focus is to maintain consistent pattern for evapotranspiration scale and shape.

Parameter adjustment is done by setting initial, minimum value and maximum value for specific parameters during model solver. The adjusted model parameters are shown in Appendix B.

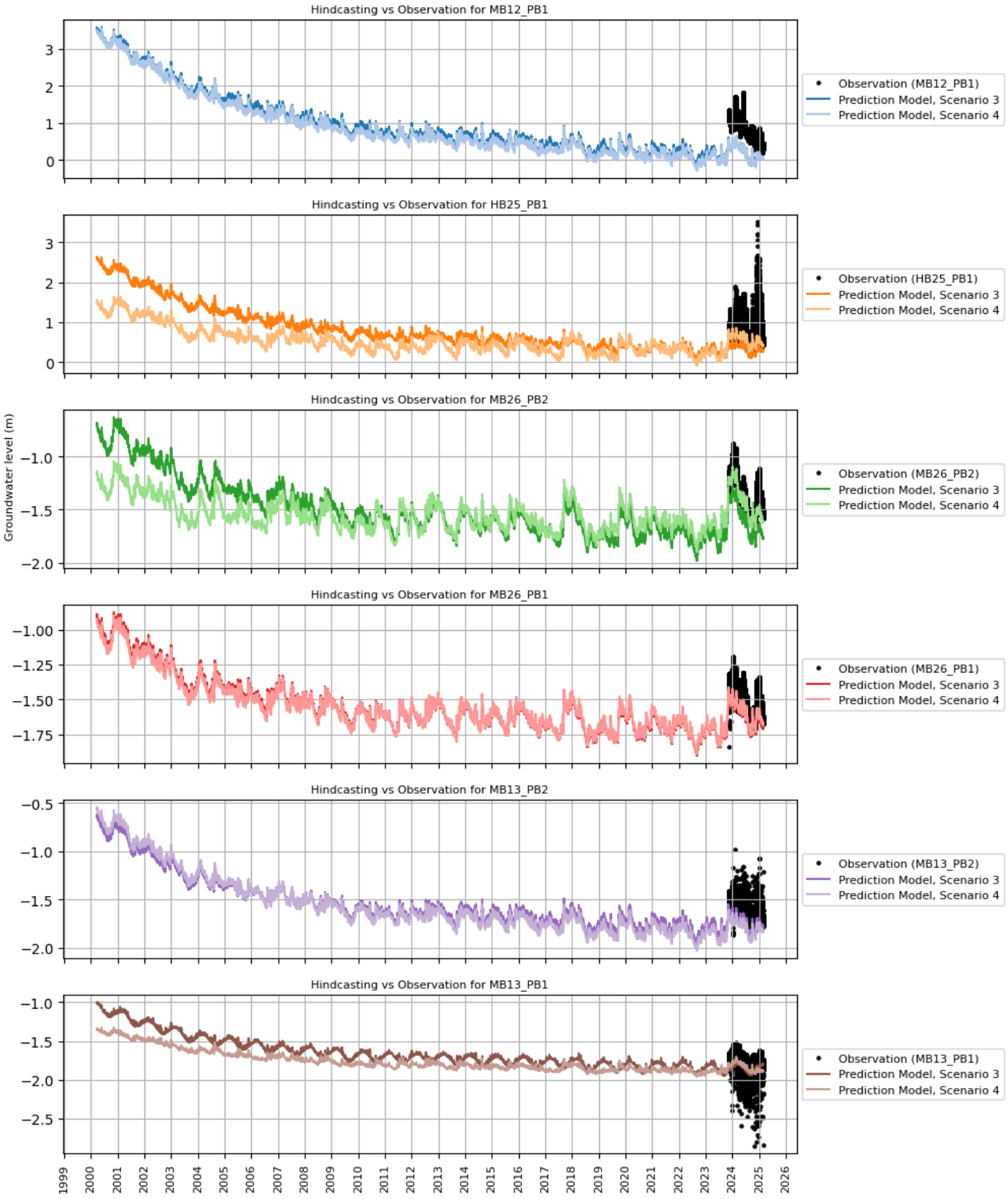


Figure 5.16 Hindcasting simulation result for Scenario 3 and 4 before manual calibration

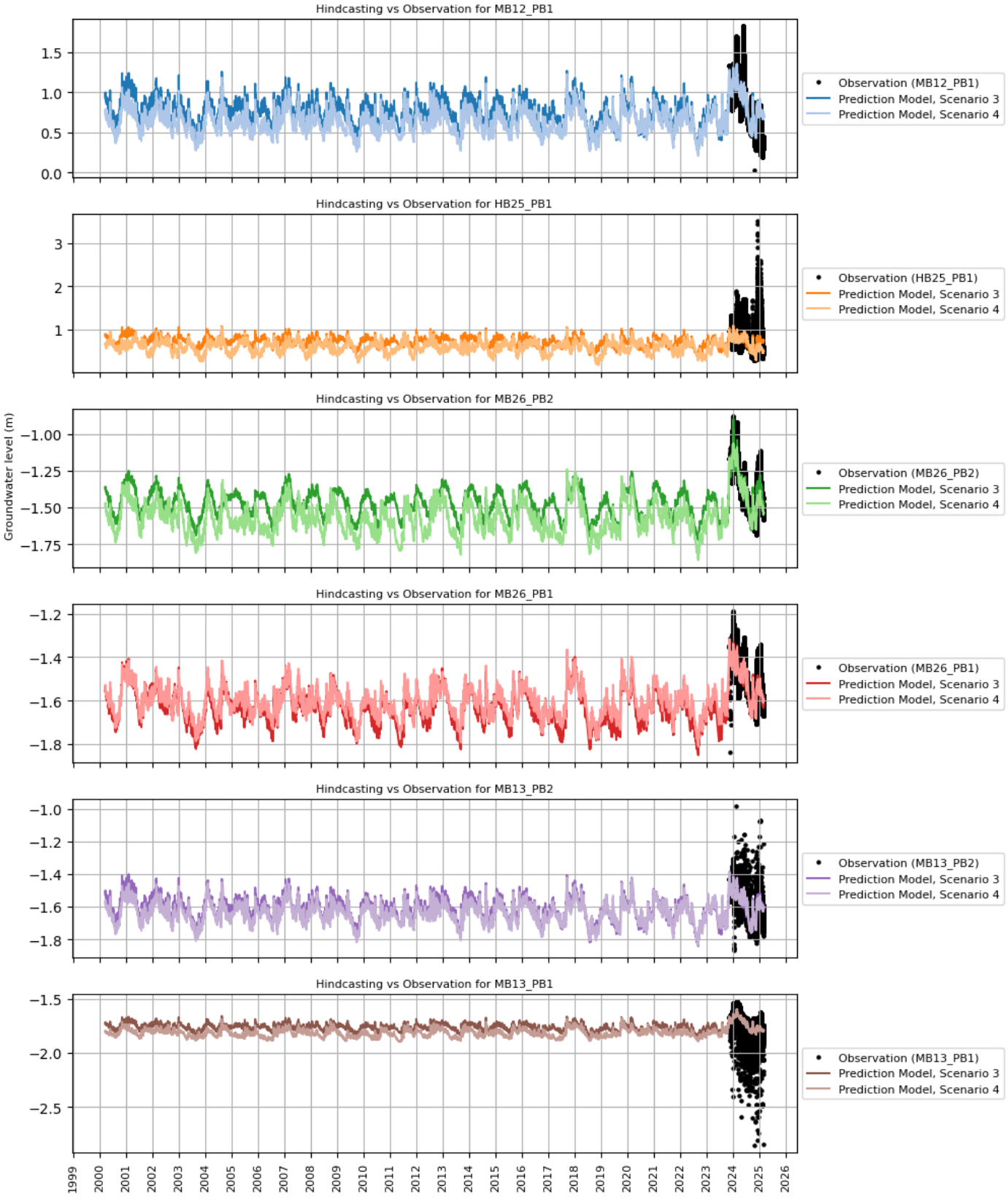


Figure 5.17 Hindcasting simulation result for Scenario 3 and 4 after manual calibration

Figure 5.16 and Figure 5.17 compare hindcasting results of groundwater levels at six monitoring wells before and after parameter calibration. Before calibration, model predictions across most wells exhibit substantial deviations from observed groundwater heads, particularly evident at wells MB12-PB1, HB25-PB1, and MB26-PB2, where model scenarios either significantly underpredict or overpredict observed data trends. Additionally, at wells MB26-PB1, MB13-PB2, and MB13-PB1, initial predictions align closer to observations, although considerable biases and mismatches remain visible. After parameter calibration, predictions show marked improvements at all monitoring locations. For example, at MB12-PB1, the calibrated models closely capture

observed seasonal fluctuations, substantially reducing bias. Similar improvements occur at HB25-PB1 and MB26-PB2, where calibrated predictions more accurately track the observed variability and peaks. Moreover, at MB26-PB1, MB13-PB2, and MB13-PB1, calibration significantly enhances model performance by reducing deviations and aligning predictions closely with observations throughout the entire historical period. Overall, parameter calibration results in visually improved accuracy, even though both models still cannot capture extreme event observed especially in HB25-PB1 and MB13-PB1. This suggest that there is another unknown stress contribution that not yet included in the Pastas model.

Comparing both Scenario 1 vs 2 and Scenario 3 vs 4, it should be noted that extreme downward trend shown in some prediction is not shown after parameter fixing, as the parameters specifically for evapotranspiration is set up such that the stress contribution mirrors the physical, real world phenomena which described as sinusoidal, negative stress contribution, and step response limited to one year following the yearly seasonality of daily evapotranspiration time series.

5.4. Dike Stability

Given the availability of the soil characteristics, for *raai_2* cross section the dike stability is assessed. The dike stability id determined with the D-Stability software, using the Bishop brute force method. The cross sectional profile with the soil layers is shown in Figure 5.18.

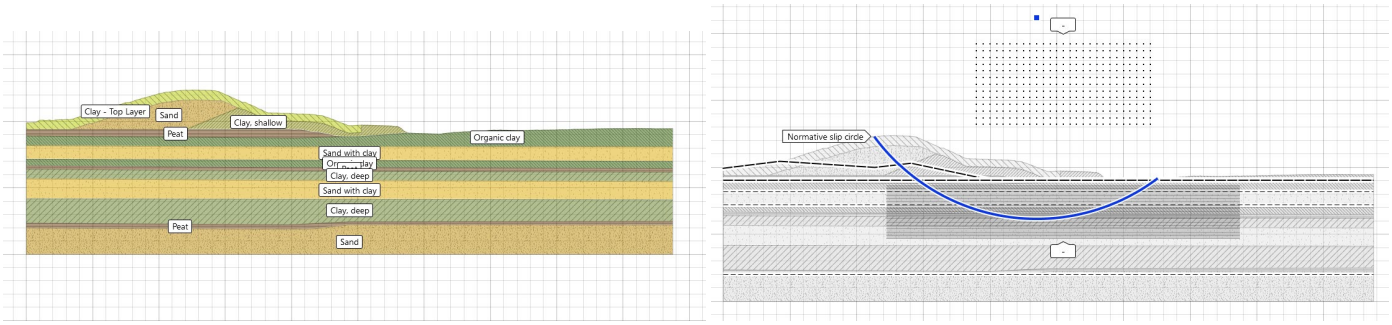


Figure 5.18 Cross Section of raai_2, Soil Layer and resulting slip plane on one of the calculations with Bishop Brute Force.

First Run: Correlated variables, Dec 2023-Jan 2024 event

The groundwater dataset used in this model follows from Scenario 4 result (*raai_2* cross section and Krabbersgat Zuid water level observation). First run is set up using hourly simulated groundwater result from 2023-12-01 00:00 until 2024-01-16 00:00, which amounts to 1127 rows of data. The relationship between related variables (precipitation, Markermeer water level) and dike stability (factor of safety) is illustrated and calculated using correlation measures (Figure 5.19 and Table 5.7).

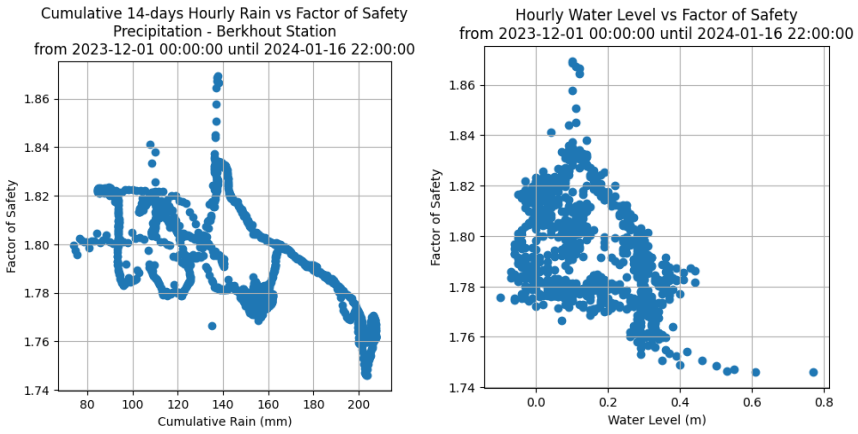


Figure 5.19 Relationship between multiple variables that influence groundwater level and factor of safety (FoS) of the dike – Correlated Case. Left plot: cumulative 14-days hourly rain vs FoS, right plot: hourly water level vs FoS.

Table 5.7 Correlation coefficient between variables and factor of safety – Correlated Case

Correlation (correlated case)	Cum. Precipitation	Water Level
Pearson	-0.67 (p≈0.0)	-0.49 (p=0.0)
Spearman	-0.68 (p≈0.0)	-0.45 (p=0.0)
Kendall	-0.48 (p≈0.0)	-0.30 (p≈0.0)

The time series of the dike stability, Markermeer water level and precipitation is shown in Figure 5.28. Overall, the comparison between the Markermeer water level and the factor of safety shows moderately negative correlation where lower water levels correlate with safer dikes. Similar moderate negative correlation is observed between 14-days cumulative precipitation and the Factor of Safety. This is supported by the correlation coefficients in Table 5.7 which indicates negative correlation for water level at -0.30 until -0.49, and negative correlation for the cumulative sum water level at -0.48 until -0.67.

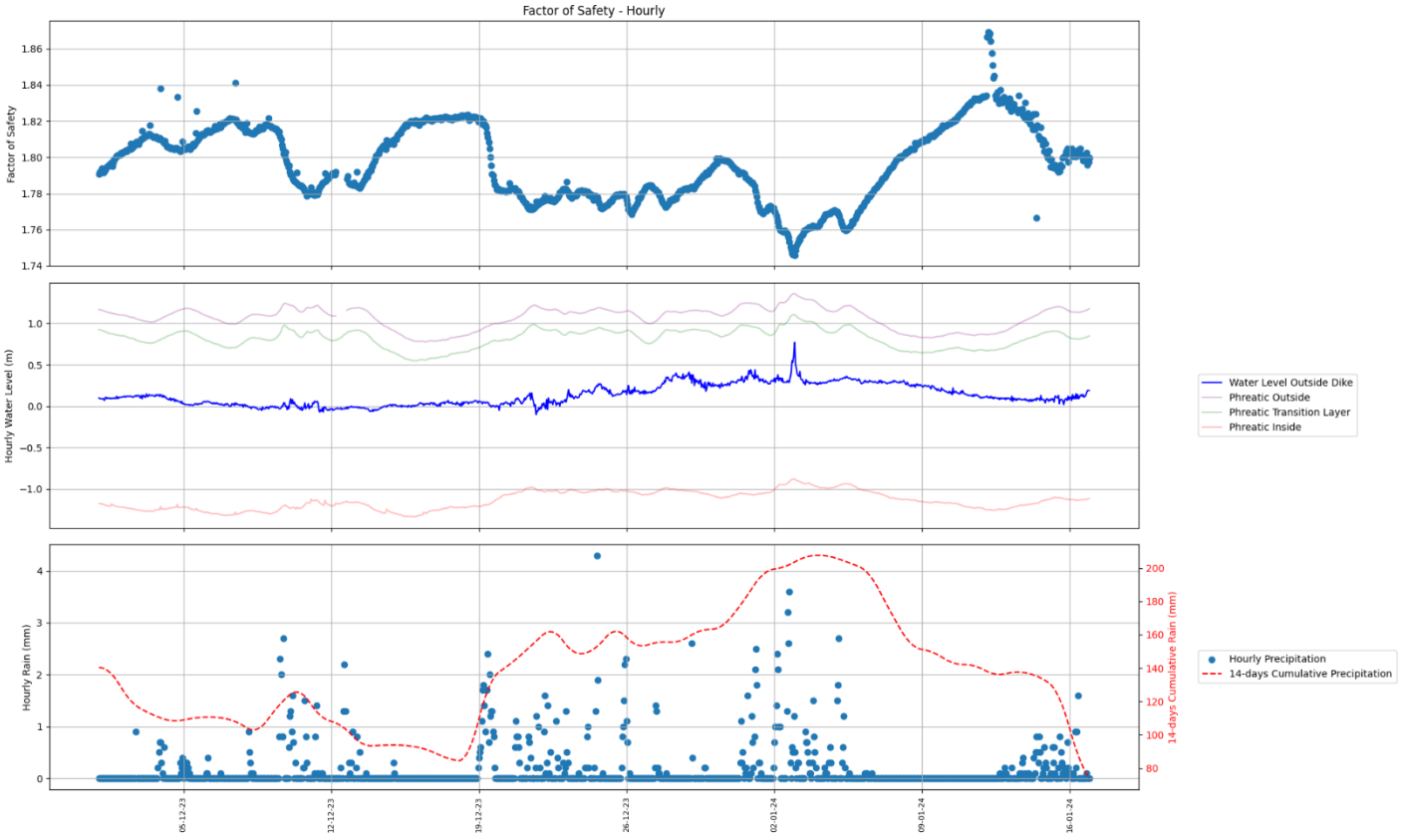


Figure 5.20 Time Series Plot of First Run (based on Hourly Dataset from 2023-12-01 00:00 until 2024-01-16 00:00). Red line shows 14-days cumulative rain based on rolling window sum.

The time series plot illustrates the relationship between slope stability (Factor of Safety), Markermeer water levels, and precipitation dynamically. The top panel shows the hourly Factor of Safety (FoS), which exhibits noticeable declines, particularly around late December to early January, indicating periods of reduced stability. The middle panel displays water levels, the local Markermeer water level and internal piezometric responses at various layers. The peak in external water level and corresponding rise in internal pore pressures align with the drop in FoS, suggesting that hydraulic loading from the water body contributes significantly to instability. In the bottom panel, observed hourly precipitation (blue dots) and its 14-days cumulative version (red dashed line) are shown. The cumulative rainfall appears to align more closely with the observed drops in FoS, indicating that prolonged wetting rather than isolated rain bursts is the dominant trigger, because the 14-day accumulation better captures soil saturation and the delayed rise of phreatic pressures that weaken the slope. The gap in FoS around middle of December is caused by the gap of observed pore pressures, specifically at the phreatic line outside of the dike. The FoS value around 12 January 2024 shows significant outlier with strong jump of FoS from 1.835 to 1.865 and then decreasing fast to 1.835 few hours later while there was no significant event on the local

water level, phreatic level or precipitation observed around that timeframe. This can be explained due to the significant change of Bishop slip plane at the specific time event compared to typical slip plane shown in Figure 5.18 (right side), as the method to determine the safety factor is repeatedly create the slip plane and calculate the FoS (brute force) with the predetermined search grid as the boundary condition.

Second run: Uncorrelated (independent) variables, synthetic water level time series

Both the Markermeer water level and the local 14-day precipitation have a negative correlation with dike stability (as shown Figure 5.28 and Table 5.5). As derived in Chapter 5.2 (Table 5.1), there is a positive albeit low correlation between the Markermeer water level and the local precipitation. This positive correlation makes low stability factors more likely, compared to a zero-correlation case. Therefore, an uncorrelated case is analysed by reconstructing Markermeer water level time series so that its water level has the same exceedance probability under the assumption of independent drivers as the observed water level (for example, NAP +0.7 m; see Figure 5.29) has under dependent drivers (see Section 2.3.4).

The conditional exceedance probability during observed peak water level is calculated as:

$$P_{correlated}(W > w_{event}|R = r_{event}) = 1 - P_{correlated}(W \leq w_{event}|R = r_{event}) = 1 - \frac{\partial C(u,v)}{\partial v}$$

where $\frac{\partial C(u,v)}{\partial v}$ is conditional distribution function of bivariate copula, called h-function, u is the empirical CDF value (quantile) of the critical water level, and v is the empirical CDF value of the critical cumulative precipitation. Both u and v are in [0,1] uniform space, independent of R. The copula is previously defined as BB-8 copula with parameter θ = 1.3 and δ = 0.97 (see Chapter 5.2).

For w_{event} = 0.7 m, R = 203.92 mm such that:

$$P_{correlated}(W > w_{event}|R = r_{event}) = P_{correlated}(W > 0.7|R = 203.92) = 0.000022$$

which is equivalent to $P_{uncorrelated}(W > \boldsymbol{w})$ with $\boldsymbol{w}$ = 0.54 m. The conditional exceedance probability is calculated for each pair in whole time series. To form the synthetic water level time series, rank-exact back-mapping method is applied. By sorting both observed water level value and the conditional exceedance probability, then assign the smallest water level to largest conditional exceedance probability, the second smallest to second largest, and so on. This step ensure that the dependence of precipitation is removed. Recreated time series is shown in Figure 5.21.

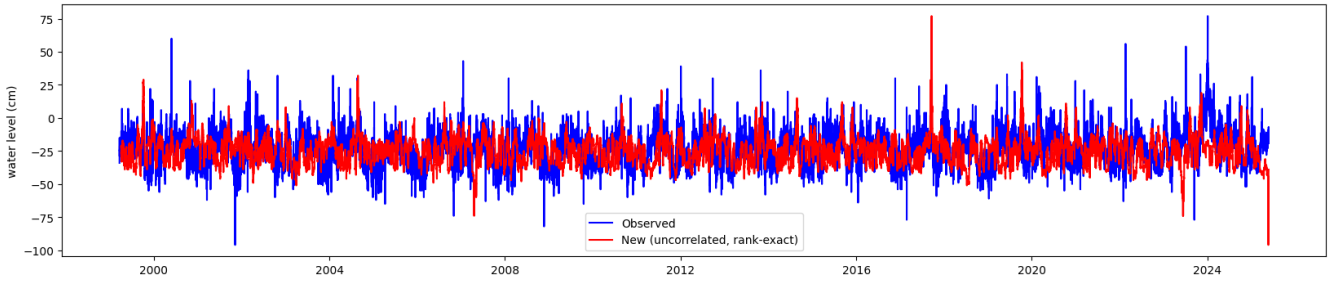


Figure 5.21 Observed Markermeer water level (blue line) and generated water level time series based on its marginal distribution (red line)

Together with observed precipitation time series and evapotranspiration, the generated water level time series is used to obtain the time series of the hydraulic head in the dike using Pastas, following the parameters previously defined (see Chapter 5.3). From the Pastas hindcasted hydraulic head, the time series of FoS is calculated using D-Stability, and the result is shown in Figure 5.22 – 5.23 and Table 5.8.

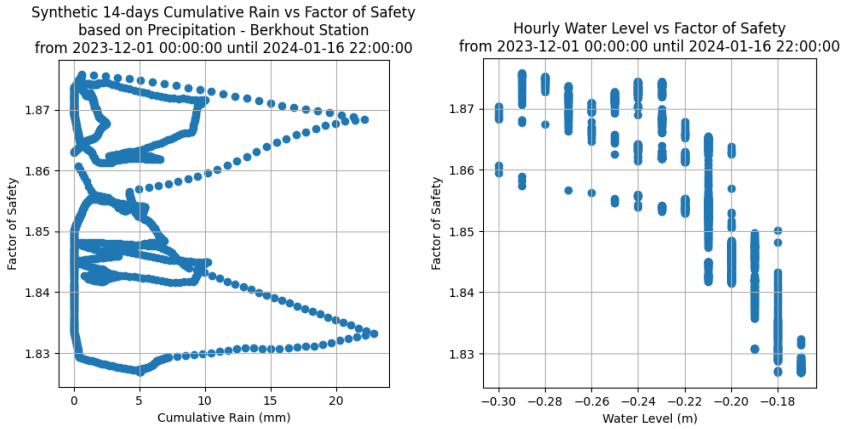


Figure 5.22 Relationship between multiple variables that influence groundwater level and factor of safety (FoS) of the dike – Uncorrelated Case. Left plot: cumulative 14-days hourly rain vs FoS, right plot: hourly water level vs FoS.

Table 5.8 Correlation coefficient between variables and factor of safety – Uncorrelated Case

Correlation (uncorrelated case)	Cum. Precipitation	Water Level
Pearson	-0.18 (p≈0.0)	-0.85 (p=0.0)
Spearman	-0.17 (p≈0.0)	-0.91 (p=0.0)
Kendall	-0.12 (p≈0.0)	-0.76 (p≈0.0)

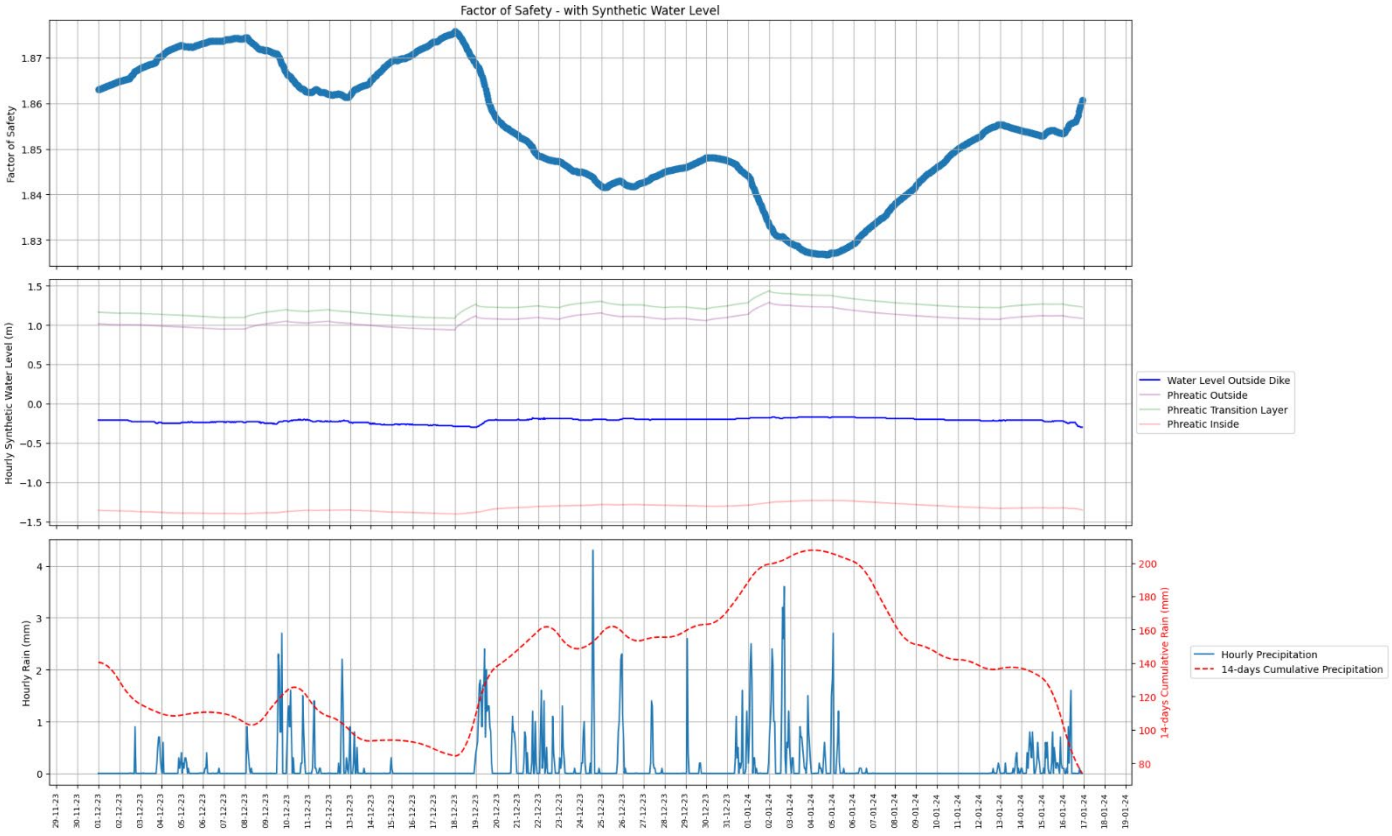


Figure 5.23 Time Series Plot of Second Run (based on Hourly Dataset from 2023-12-01 00:00 until 2024-01-16 00:00). Red line shows 14-days cumulative rain based on rolling window sum.

Based on Figure 5.23, the peak water level is much lower compared to first run (correlated), around -0.2 m. However, in this independent case, the overall factor of safety is much higher, with lowest recorded FoS is 1.827 for water level **w** > -0.2 m compared to FoS of 1.745 in the dependent case. By looking at the range, the recorded FoS range is from 1.878 until 1.827, compared to correlated case range from 1.84 until 1.745, which is 86.3%

larger. It is also interesting to observe that due to relatively flat local water level, the shape of safety factor time series is close to the inverse of the 14-days cumulative precipitation, which support the negative correlation between FoS and precipitation. Therefore, this result confirms previous statement that positive correlation makes low stability factors more likely, compared to a zero-correlation case.

Correlated vs Uncorrelated results

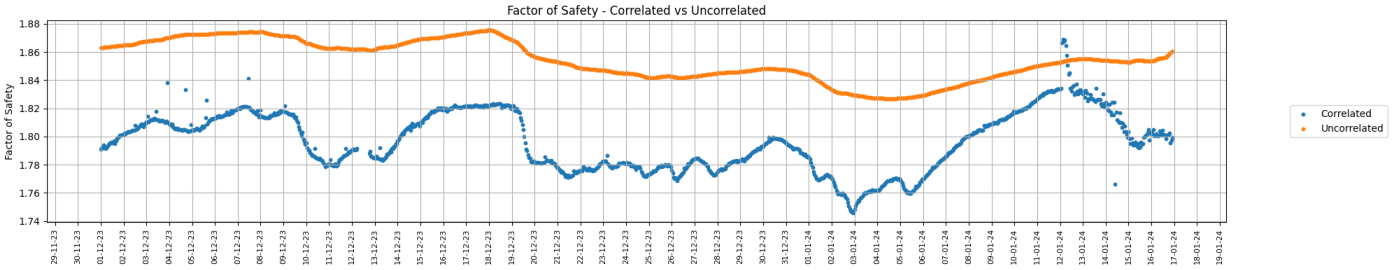


Figure 5.24 Comparison of Factor of Safety over time, correlated case (First run, blue color) vs uncorrelated case (Second run, orange color)

Following the results from first run and second run, a comparison on FoS has been made in Figure 5.24. During this period, the uncorrelated case yields a consistently higher and smoother factor of safety compared to the correlated case. In the correlated case, clear fluctuations and deeper dips are observed, particularly during periods of high water levels and rainfall, whereas the uncorrelated case remains relatively stable with only gradual variations. The minimum values of FoS are visibly lower in the correlated series, while the uncorrelated series maintains higher FoS throughout the entire time window.

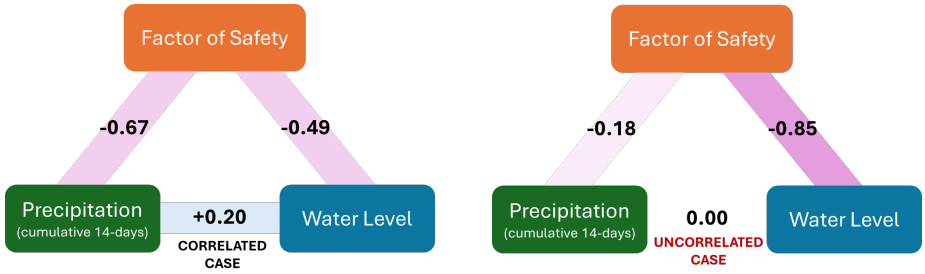


Figure 5.25 Comparison of Pearson coefficient measures, correlated case (left side) vs uncorrelated case (right side)

The comparison of coefficient measures is illustrated in Figure 5.25. There are few points to consider from this figure:

- In the dependent (correlated) case, both precipitation and lake water level show strong negative correlations with the factor of safety (FoS). When the dependence between precipitation and water level is removed, water level becomes the dominant driver. This indicates that water level largely controls the stability response, while precipitation mainly amplifies the FoS reduction when it occurs together with high water levels.
- There are two implications from this:
    - Accounting for dependence between precipitation and water level yields a lower FoS, which is more conservative and aligns with many real situations for dikes along rivers and lakes. For river dikes, weak or no local dependence can occur, for example when rain or snowmelt happens upstream so that a high river stage develops downstream without local precipitation. For large lakes, correlated conditions are more commonly found in storm events, where the rain often cover both the lake and the adjacent dike and directly affect the local water level.
    - At the design stage, using the dependent combination tends to be more conservative during wet-season conditions (as shown for December 2023 to January 2024), since it generally produces a lower FoS. If a combination-factor approach is used, apply a site-specific multiplier for precipitation conditional on high lake level rather than a fixed, independent factor.

Chapter 6. Discussions

In this chapter, several assumption and limitation made during this study will be explained.

While many factors and variables influence the phreatic level within the dike, only a few were selected for further analysis: precipitation, evapotranspiration, water level, dike cross-section, and soil layer properties. The wind factor was not considered due to the lack of local water level measurements near the selected dike cross-section. Although wind setup significantly affects water level fluctuations in large bodies of water such as the Markermeer and IJsselmeer, predicting local water levels in time-series form is challenging. However, for extreme condition calculations, it is possible to estimate the effect of wind setup using the Young and Verhagen (1996) equation, which can be used to calculate the significant wave height to be added to the normal water level.

Copula analysis is used to demonstrate not only the correlation between variables but also their dependency structure. The main purpose of using a copula model is to capture and simulate extreme values that cannot be adequately represented by simple linear models. By combining samples generated from the copula model with seasonal time-series models, such as Seasonal ARIMA, it is possible to create synthetic time series for input into groundwater modeling. However, in this study, the generated copula samples were not used as input for Pastas because the available dataset was sufficient to address the research questions within the scope of this study.

During groundwater modeling for all cross-sections, the groundwater measurements at both raai_2 and raai_3 show a large difference between the start and end of the calibration period, although sharp peaks and troughs are still observed throughout the intermediate period. Although the calibration period was set to one full year, this large difference suggests that the yearly seasonality (high and low seasons) is shifted and does not follow a consistent annual pattern. Extending the calibration period could improve the model's accuracy; however, due to data availability limitations, only one year of data could be used for calibration, with an additional three months reserved for validation.

During validation stage, there are lots of model with poor performance where R^2 value is negative (worse compared to mean value prediction of target variable). While the model parameter itself is chosen based on comparison in section 5.3.1, the poor performance is due to training dataset used during calibration stage is not enough to capture the possible pattern during validation stage. This is also relevant with previous paragraph, where the major trend of observed groundwater level during calibration stage is decreasing but when it comes to validation stage the dataset is increased sharply then flat (ignoring the short-term seasonality).

The extreme divergence of the model tail especially before model calibration arises because the model was never constrained by observations in that early period, and then during hindcasting simulation the small error is building up over time, therefore longer period of hindcasting will cause larger accumulated error. Therefore, longer historic data for calibration are needed to bring the model tail closer to reality.

To assess whether the limited dataset contributes to the model's poor performance, walk-forward validation is used as a comparison to static validation. Walk-forward validation, also known as rolling origin cross-validation, is a time series-specific technique that evaluates forecasting models by continuously retraining them as new observations become available (Siami-Namini & Namin, 2018). This method is particularly suitable for time series datasets, as it simulates a real-world forecasting scenario where the model is regularly updated with the most recent data.

Compared to a static split, walk-forward validation offers an iterative approach that better reflects the dynamic nature of time series. In this process, the model is first trained using n months of data, then validated using the data from month $n + 1$. The chosen error metrics (R^2 and RMSE, in this case) are recorded after each run. In the next step, the model is retrained using $n + 1$ months of data and validated on month $n + 2$. This process repeats iteratively until the final step, where the model is trained using the entire dataset minus the last month, and the final month is used as the validation period. This approach allows for a more robust evaluation of the model's performance over time, providing insights that a static validation split might overlook.

Walk-forward validation model is applied to Scenario 2 of groundwater model, specifically on MB014-PB1 groundwater well location. For the first part, the model is trained using 3 months of data, from 2023-11-10 until 2024-02-10. The result is shown in the following graph:

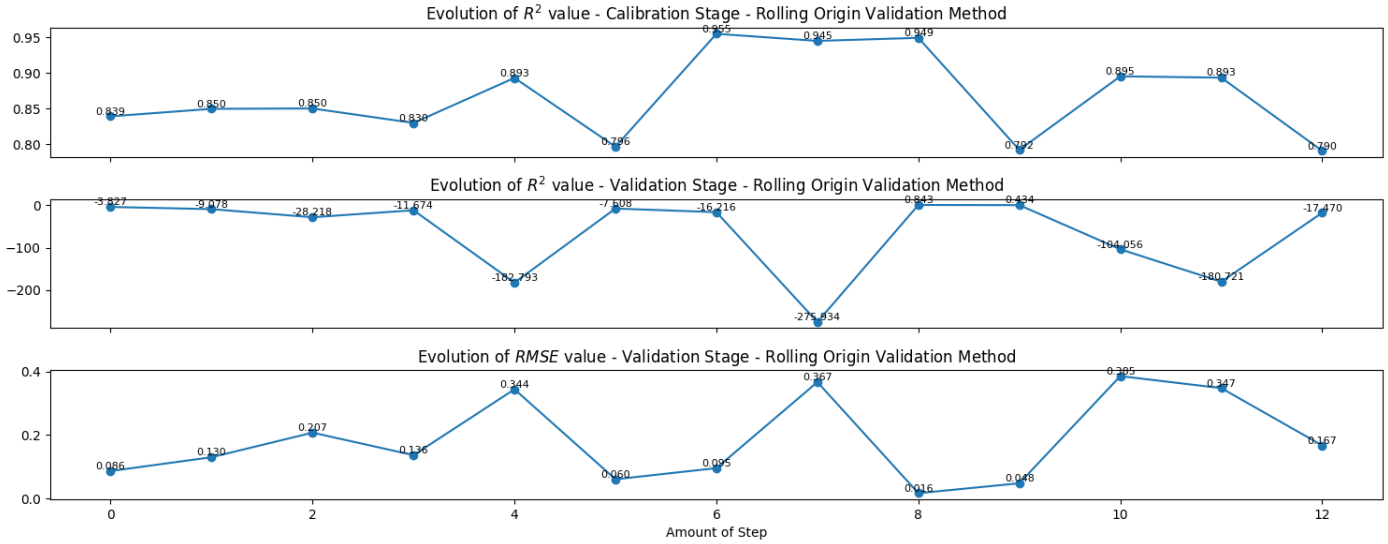


Figure 6.1 Walk-forward Validation Model for MB014-PB1 Scenario 2 with 12 steps.

The results of the rolling-origin validation show significant difference between the model's performance during calibration and validation stage. Specifically, the model reaches consistently high R^2 value during the calibration stage, indicating a strong fit to the calibration dataset. However, during validation, the R^2 values become negative, and in some cases very significant, suggesting that the model's predictive accuracy is inferior to that of a simple mean-based predictor. This pattern is further supported by the $RMSE$ values observed during validation, which shows inconsistent and large prediction errors.

The possible reason why the groundwater model failed to predict during validation stage is most likely due to data extrapolation error, where the model may operate outside its trained range due to extreme events, trends or seasonality that have not yet captured during calibration stage. This argument further supported based on the result during step 8 and 9. In those steps, the model performs well, with a high R^2 during calibration and a positive R^2 during validation. This corresponds to the period from July 2024 to September 2024, when the observed data show a linear decreasing trend. To improve the model's accuracy, the response functions could be enhanced by using a more complex model, such as a Recharge model. Alternatively, adding additional datasets that include more diverse events could help improve the training process and overall model performance.

During model improvement for hindcasting, parameter fixing in the evapotranspiration stress model was chosen as the primary modification. This approach was selected to ensure that the evapotranspiration stress follows a typical sinusoidal pattern and maintains a negative contribution to the groundwater head, reflecting the physical reality that increased evapotranspiration leads to a decrease in groundwater levels. However, because the Pastas solver uses least squares optimization, where the main goal is to minimize the sum of squared residuals between observed and simulated groundwater heads by adjusting the response function parameters, the initial calibration results tend to focus on maximizing the R^2 value. This can sometimes come at the expense of maintaining the correct physical representation of evapotranspiration. By specifically fixing the evapotranspiration parameters, the R^2 value may decrease slightly, but the model preserves a more realistic physical behavior.

The overall diagnostic plots (see Appendix A) for all scenarios generally show no strong systematic bias. The Q-Q plots align well with the theoretical distribution, and the histograms indicate an approximately normal distribution of residuals. However, the ACF plots reveal strong seasonality and significant autocorrelation, suggesting that the models tend to miss short-term processes in groundwater fluctuations.

In some cases, the probability plots show a poor fit, reflected by low R^2 values—such as for well MB12-PB1 in Scenarios 3 and 4. The observed groundwater levels at this location exhibit highly erratic behavior over time, which is difficult for the model to replicate. This could be due to the specific location of MB12-PB1, which is situated relatively far from Markermeer on the landward side of the dike. Additionally, the base of the well is located in a deep sand layer with high hydraulic conductivity, making it more susceptible to external influences such as groundwater pumping or localized recharge occurring within the same aquifer. These factors may explain the model's limited ability to capture the observed fluctuations at this particular well.

The Bishop brute force method was chosen to calculate the safety factor instead of other methods available in D-Stability, such as Particle Swarm (Uplift-Van method) or Genetic Algorithm (Spencer method). This choice was made because the slope is relatively simple, and the dike geometry and soil layer properties are constant. As a result, the grid area remains fixed and manageable, allowing the Bishop method to provide fast computational times and a less complex iterative solution compared to the other available methods.

A long-term dike stability analysis was not conducted in this study due to the limited quality of the hindcasting results. The groundwater model generated using Pastas was based on a relatively short period of real groundwater observations, which lacked the temporal variation needed to accurately represent long-term trends. As a result, the model was unable to capture the full range of seasonal and extreme conditions that are essential for a reliable long-term stability assessment. To perform such an analysis, a longer and more comprehensive dataset would be required to ensure that the groundwater fluctuations and their effects on dike stability are realistically represented over time.

Overall, despite these limitations and assumptions, the model depicts the phenomenon that influence the dike stability in real life. Groundwater levels rise after rainy periods or when the lake level is high, and they fall in dry periods when evapotranspiration is strong, which is consistent with what is observed in practice. However, there are still some variables that not yet captured in the model that influence the prediction in specific scenario. Local wind effect in the dike section is not captured due to no measurement available, but this wind effect can push groundwater up near the dike temporarily, explaining the short, erratic spikes in the observed dataset. Local pumping and draining also influence the head measurement in short to medium timeframe. Also, the model only trained on a year of validation observed dataset with limited trends and seasonality. Longer groundwater measurements (more than 5 years) can improve the model parameters and even changing the stress model configuration, which in turn improve the prediction.

In the context of dike stability, the soil parameters are selected deterministically to produce a single factor of safety (FoS) each time step due to computational time limitation. Due to this approach, there is risk of FoS overconfidence due to uncertainty in related variable such as dike geometry, unit weight, etc. This can be improved by either adding sensitivity analysis or cross-check with other method outside of Bishop (i.e. Spencer or Uplift-Van model) as comparison.

The framework used in this study can also be applied to the other dike section, such as IJsselmeer dike sections or even dike section that directly facing the North Sea. However, in these cases the influence of other variables might be stronger, such as wind setup for dike facing North Sea, or pump operation for dike facing polder, such that the correlation between other variables can be explored further and might even have larger influence dike stability compared to precipitation and local water level.

Chapter 7. Conclusions and Recommendations

7.1. Conclusions

Conclusion will be formulated based on the research objective to assess how the Effect of Hydraulical and Meteorological Correlations towards Dike Stability during Extreme Events, by answering the following research questions:

1. What are the variables that influence phreatic level, which also influence dike stability in dike section around Markermeer and IJsselmeer region?

   The phreatic level within dikes around the Markermeer and IJsselmeer region is influenced by a combination of external hydraulic loads, climatic conditions, internal soil properties, dike geometry, and local lake dynamics, all of which impact dike stability. A key external influence is the fluctuation of lake water levels, affected by storm conditions, sluice operations, and wind-driven water movement. As previously illustrated in Figure 5.1, in the case of Afsluitdijk the wind storm from northeast will increase the water level in the outer dike but simultaneously decrease the water level in the inner dike (IJsselmeer side). In this case, the sluice will closed and no water can get out from IJsselmeer, which when combined with heavy rainfall and runoff will increase the water level in the IJsselmeer and in turn will increase the load on the dike. When the wind goes from southwest, the inverse happens, and water from IJsselmeer can be discharged to the North Sea. In the context of local dike, strong winds across the IJsselmeer and Markermeer can push water toward the shorelines, causing temporary rises in local water levels that increase hydraulic pressure on nearby dikes. Rapid changes in water levels cause delayed pore pressure responses within low-permeability clay dikes, increasing the risk of internal instability. Climatic variables such as precipitation and evapotranspiration also play a crucial role where rainfall can infiltrate and raise the phreatic line, especially at the inner toe and crest, while drought conditions lead to shrinkage cracks that act as rapid infiltration pathways during the following wet periods. Internally, the permeability and specific storage capacity of soil layers govern how quickly water moves and is stored, with low-permeability, clay-ey soil type, or heterogeneous soils tending to maintain higher and more persistent phreatic levels. Dike geometry, including slope, width, and height, further influences drainage and storage behavior, with wider or flatter dikes retaining more water and elevating pore pressures. Sub-soil conditions, particularly clay thickness beneath the dike, also affect drainage and saturation levels. Collectively, these variables determine the phreatic behavior and, in turn, the macro-stability of dikes in this low-lying region.

2. Is there any correlation between hydraulic and meteorology variables that influence dike stability?

   Yes, there is correlation between hydraulic and meteorology variables that influence dike stability, as demonstrated in the result of correlation measurement between precipitation and water level in Chapter 5.2. The analysis shows that cumulative precipitation, particularly over a 14-day period, correlates positively with rising water levels, indicating that sustained rainfall contributes to increased hydraulic load. The copula fitting results strengthen this understanding. The best-fit copula (BB8) successfully captures asymmetric dependence and tail co-occurrence, which indicates that extreme rainfall events and high water levels tend to occur together, though not always symmetrically. This suggests a nonlinear and complex relationship between meteorological (rainfall) and hydraulic (water level) variables that jointly influence dike safety.

3. Based on the proposed stochastic variables, what is the effect of correlation between water level and precipitation for the phreatic level in extreme conditions for dike stability in Markermeer during the wet season (Dec 2023 – Jan 2024)?

   The analysis in Chapter 5.4 shows that during the wet season of December 2023 to January 2024, both Markermeer water level and 14-day cumulative precipitation exhibit negative correlations with the dike's factor of safety while the correlation between water level and precipitation is slightly positive, making

simultaneous extreme conditions more probable. Under the observed (dependent) case, peak external water levels coincided with prolonged high precipitation, producing higher phreatic levels within the dike and a lower minimum factor of safety (1.745). In contrast, when the same marginal distributions were used but the drivers were made independent, peak water levels were lower, and the minimum factor of safety improved (1.768).

By looking at the Pearson correlation measures, the correlation between 14-days cumulative precipitation and hourly water level is 0.20, while the correlation between cum. precipitation vs Factor of Safety (FoS) is -0.67 and hourly water level vs FoS is -0.49. After removing the correlation between cum. precipitation and water level, the correlation between water level and FoS is stronger (-0.85) compared to precipitation case (-0.18). This result confirms that while main external load on Markermeer dike phreatic levels is the lake water level, in the case of wet season during December 2023 – January 2024, precipitation and water level tend to rise together (correlation = 0.2), such that the cum. precipitation has more influence on the dike stability.

7.2. Recommendations

In light of the findings presented in this study, the following recommendations are proposed to improve the assessment and modelling of dike stability under varying hydrometeorological conditions:

1. Extend groundwater observation records
   The availability of long-term groundwater data is essential for enhancing model calibration and validation. Extended datasets would allow better representation of seasonal variability, extreme events, and long-term trends, thereby increasing the reliability of groundwater and stability modelling.
2. Incorporate detailed soil and hydraulic properties
   The current approach used generalized soil characteristics. More detailed and spatially variable data on soil permeability, specific storage, and stratigraphy are recommended to better capture subsurface flow responses, particularly in areas with heterogeneous or highly permeable soils.
3. Advance time-series modelling techniques
   Seasonal ARIMA and walk-forward validation proved valuable for identifying model limitations. Future studies should continue to apply adaptive time-series techniques, including recharge-based models and rolling-origin cross-validation, to address extrapolation errors and improve model generalization under changing conditions.
4. Integrate copula-based stochastic simulation
   Copula modelling revealed significant nonlinear dependencies between rainfall and water level. Integrating copula-generated time series into groundwater models would allow for more realistic scenario generation, particularly for stress testing under joint extreme conditions.
5. Maintain physically meaningful parameter constraints
   While achieving a high goodness-of-fit is important, it should not override the need for physical realism. Fixing evapotranspiration parameters to ensure negative contributions to groundwater head reflects actual processes and improves the interpretability and reliability of model outcomes.
6. Investigate site-specific anomalies
   For locations where model performance was poor, such as well MB12-PB1, further investigation is recommended. Site-specific influences such as groundwater abstraction, recharge from local features, or deep aquifer interactions may need to be considered to improve simulation accuracy.

References

Adji, F. F., Yulianti, N., Damanik, Z., Candra, B., Sajarwan, A., Winerungan, S., Darung, U., Teguh, R., Sitohang, G. J., & Sitorus, M. R. (2023). Soil hydraulic conductivity and soil water retention of inland peat on various land covers (Case study: Natural peat and burnt peat). *OnLine Journal of Biological Sciences, 23*(3), 251–263. https://doi.org/10.3844/ojbsci.2023.251.263

Berendrecht, W., Heemink, A., van Geer, F., & Gehrels, J. (2006). A non-linear state space approach to model groundwater fluctuations. *Advances in Water Resources, 29*(7), 959–973. https://doi.org/10.1016/j.advwatres.2005.08.009

Chen, L., & Guo, S. (2019). Copula theory. In *Copulas and its application in hydrology and water resources* (pp. 11–36). Springer. https://doi.org/10.1007/978-981-13-0574-0_2 Clennell, M. B., Dewhurst, D. N., Brown, K. M., & Westbrook, G. K. (1999). Permeability anisotropy of consolidated clays. *Geological Society London Special Publications, 158*(1), 79–96. https://doi.org/10.1144/gsl.sp.1999.158.01.07

Collenteur, R. A., Bakker, M., Klammler, G., & Birk, S. (2021). Estimation of groundwater recharge from groundwater levels using nonlinear transfer function noise models and comparison to lysimeter data. *Hydrology and Earth System Sciences, 25*, 2931–2949. https://doi.org/10.5194/hess-25-2931-2021

Czado, C. (2019). *Analyzing dependent data with vine copulas: A practical guide with R* (Vol. 222). Springer. https://doi.org/10.1007/978-3-030-13785-4

Denkers, A. J. (2021). *Analysis of the phreatic line on various types of primary flood defences* [Master's thesis]. TU Delft.

Duncan, J. M., Wright, S. G., & Brandon, T. L. (2014). *Soil strength and slope stability* (2nd ed.). Wiley.

Expertise Netwerk Waterveiligheid. (2024). Water tegen de dijk: Het functioneren van waterkeringen tijdens het winterhoogwater 2023/2024 [Report]. Expertise Netwerk Waterveiligheid.

Goossens, L. M. (2022). *Effects of flexible water level management on flood risk in areas outside primary levees around the IJsselmeer and Markermeer* [Master's thesis]. TU Delft.

Han, Z., Shi, X., Jia, K., Sun, B., Zhao, S., & Fu, C. (2019). Determining the discharge and recharge relationships between lake and groundwater in Lake Hulun using hydrogen and oxygen isotopes and chloride ions. *Water, 11*(2), 264. https://doi.org/10.3390/w11020264

Hassan, M. A., & Ismail, M. A. M. (2018). Influence of dike slope on the development of infiltration water and erosion processes during overtopping tests. *International Journal of Engineering & Technology, 7*, 520–525. https://doi.org/10.14419/ijet.v7i2.9308

HHNK - Hoogheemraadschap Hollands Noorderkwartier. (2014). *Naar veilige Markermeerdijken*.

HHNK - Hoogheemraadschap Hollands Noorderkwartier. (2023). *Waterschapsblad van Hoogheemraadschap Hollands Noorderkwartier*. Jamalina, E., Vardon, P. J., Steele-Dunne, S. C., & Tavakol, E. (2021). The stability of dikes subject to soil-vegetation-atmosphere interaction.

Ilyés, C., Tóth, M., & Szűcs, P. (2023). Calculating the connection between precipitation and shallow groundwater. Multidiszciplináris Tudományok, 13(3), 292–299. https://doi.org/10.35925/j.multi.2023.3.29

Joe, H. (1997). *Multivariate models and multivariate dependence concepts* (1st ed.). Chapman and Hall/CRC. https://doi.org/10.1201/9780367803896

Manzione, R. L. (2018). Physical-based time series model applied on water table depths dynamics characteristics simulation. *RBRH, 23*. https://doi.org/10.1590/2318-0331.0318170071

Milly, P. C. D. (1994). Climate, soil water storage, and the average annual water balance. *Water Resources Research, 30*(7), 2143–2156. https://doi.org/10.1029/94WR00586

Nishiie, S., Nishimura, S., & Yamazoe, N. (2019). Long- and short-term pore water pressure variations in sandy river dike interpreted with 1- and 2-phase seepage flow analysis. *Japanese Geotechnical Society Special Publication, 7*(2), 648–651. https://doi.org/10.3208/jgssp.v07.099

OpenTURNS. (n.d.). Cramer-von Mises test - OpenTURNS 1.25 documentation. https://openturns.github.io/openturns/latest/theory/data_analysis/cramer_vonmises_test.html

Pešková, J., & Štibinger, J. (2015). Computation method of the drainage retention capacity of soil layers with a subsurface pipe drainage system. *Soil and Water Research, 10*(1), 24–31. https://doi.org/10.17221/119/2013-swr

Peterson, T. J., & Western, A. W. (2014). Nonlinear time-series modeling of unconfined groundwater head. *Water Resources Research, 50*(12), 8330–8355. https://doi.org/10.1002/2013WR014800

Rijkswaterstaat. “The IJsselmeer and Markermeer: A Natural Water Tank.” Rijkswaterstaat.nl, Rijkswaterstaat, 18 July 2025, www.rijkswaterstaat.nl/en/news/archive/2025/07/the-ijsselmeer-and-markermeer-a-natural-water-tank. Accessed 21 July 2025.

Sammis, T. W., Williams, S., Smeal, D., & Kallsen, C. E. (1986). Effect of soil moisture stress on leaf area index, evapotranspiration and modeled soil evaporation and transpiration. *Transactions of the ASAE, 29*(4), 956–962. https://doi.org/10.13031/2013.30255

Schreiber, T., & Schmitz, A. (2000). Surrogate time series. Physica D: Nonlinear Phenomena, 142(3–4), 346–382. https://doi.org/10.1016/s0167-2789(00)00043-9

Siami-Namini, S., & Namin, A. S. (2018). Forecasting Economics and Financial Time Series: ARIMA vs. LSTM. arXiv (Cornell University). https://doi.org/10.48550/arxiv.1803.06386

van Ginkel, K., Klijn, F., Mens, M., & ter Maat, J. (2022). *Verkennende systeemanalyse IJsselmeergebied*. Deltares.

Verruijt, A. (2018). *An introduction to soil mechanics*. In *Theory and Applications of Transport in Porous Media*. Springer. https://doi.org/10.1007/978-3-319-61185-3

von Asmuth, J. R., Bierkens, M. F. P., & Maas, K. (2002). Transfer function-noise modeling in continuous time using predefined impulse response functions. *Water Resources Research, 38*(12), 1287. https://doi.org/10.1029/2001WR001136

Wang, L., Wang, Y., Vuik, C., & Hajibeygi, H. (2022). Accurate modeling and simulation of seepage in 3D heterogeneous fractured porous media with complex structures. *Computers and Geotechnics, 150*, 104923. https://doi.org/10.1016/j.compgeo.2022.104923

Woerkom, T. A. A. van. (2023). *Improving dike slope failure assessments related to groundwater hydrology*. https://doi.org/10.33540/1819

Wu, L., Chen, L., Ding, Y., & Zhao, T. (2018). Testing for the source of multifractality in water level records. Physica A: Statistical Mechanics and Its Applications, 508, 824–839. https://doi.org/10.1016/j.physa.2018.05.148

Yan, Z., Wang, J., & Chai, H. (2010). Influence of water level fluctuation on phreatic line in silty soil model slope. *Engineering Geology, 113*, 90–98. https://doi.org/10.1016/j.enggeo.2010.02.004

Appendix

A. Correlation Coefficients for Each Precipitation Type with Lag up to 14 days

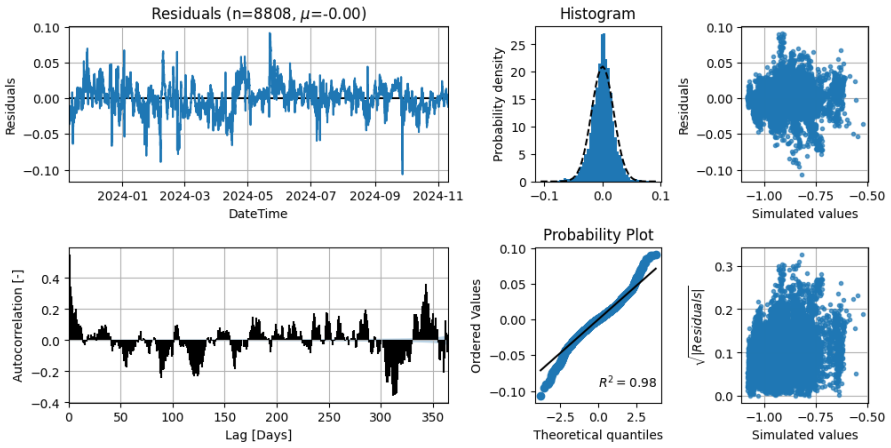
Precip Type	Lag (days)	Coefficients			p-value		
		Pearson	Spearman	Kendall	Pearson	Spearman	Kendall
Hourly	**0**	**0.0663**	**0.033**	**0.0261**	**1.25E-221**	**2.00E-56**	**1.79E-55**
	1	0.0306	0.001	0.0008	1.39E-48	0.616074928	0.638696946
	2	0.0407	0.0152	0.0121	8.64E-85	3.19E-13	4.01E-13
	3	0.038	0.0114	0.009	6.96E-74	5.45E-08	5.73E-08
	4	0.0338	0.0048	0.0038	6.60E-59	0.021530584	0.020720059
	5	0.032	0.0007	0.0006	3.69E-53	0.741129917	0.732169171
	6	0.0325	0.0014	0.0011	1.11E-54	0.514292596	0.507712069
	7	0.0305	-0.004	-0.0032	2.59E-48	0.054670269	0.053577907
	8	0.0263	-0.0042	-0.0033	2.01E-36	0.043518798	0.044055919
	9	0.0253	-0.0061	-0.0049	6.68E-34	0.003555716	0.003467042
	10	0.0264	-0.0082	-0.0065	9.78E-37	8.23E-05	8.46E-05
	11	0.0235	-0.0135	-0.0108	2.89E-29	9.26E-11	8.77E-11
	12	0.0252	-0.0131	-0.0105	1.22E-33	3.21E-10	2.63E-10
	13	0.0231	-0.0137	-0.0109	1.69E-28	5.86E-11	5.52E-11
	14	0.0217	-0.0214	-0.017	3.24E-25	1.62E-24	1.18E-24
7-day sum	**0**	**0.1753**	**0.1316**	**0.089**	**0**	**0**	**0**
	1	0.1693	0.1226	0.0831	0	0	0
	2	0.1627	0.1119	0.0761	0	0	0
	3	0.1522	0.0983	0.067	0	0	0
	4	0.1453	0.0897	0.0611	0	0	0
	5	0.1388	0.0807	0.0551	0	0	0
	6	0.1335	0.0753	0.0513	0	1.84E-285	3.35E-287
	7	0.1272	0.0693	0.0471	0	5.72E-242	1.06E-242
	8	0.1216	0.0629	0.0426	0	1.65E-199	3.35E-199
	9	0.1132	0.0535	0.0361	0	6.58E-145	2.07E-143
	10	0.105	0.0446	0.03	0	4.33E-101	1.08E-99
	11	0.094	0.0361	0.0242	0	5.13E-67	1.11E-65
	12	0.0844	0.0266	0.0179	0	4.44E-37	1.01E-36
	13	0.0735	0.0147	0.01	1.74E-271	2.03E-12	1.77E-12
	14	0.0651	0.0067	0.0046	3.11E-213	1.30.E-03	1.10.E-03
14-day sum	**0**	**0.1973**	**0.147**	**0.0994**	**0**	**0**	**0**
	1	0.1898	0.1366	0.0927	0	0	0
	2	0.18	0.1234	0.0838	0	0	0
	3	0.1678	0.1085	0.0736	0	0	0
	4	0.1562	0.095	0.0643	0	0	0
	5	0.1457	0.0831	0.0562	0	0	0
	6	0.1351	0.0723	0.0489	0	6.31E-263	1.07E-261
	7	0.1256	0.0637	0.043	0	2.05E-204	2.42E-203
	8	0.1161	0.0549	0.0369	0	4.74E-152	4.69E-150
	9	0.1072	0.0465	0.0311	0	1.24E-109	2.30E-107
	10	0.0982	0.0377	0.0252	0	1.01E-72	7.14E-71
	11	0.0887	0.0291	0.0195	0	3.90E-44	5.44E-43
	12	0.0794	0.0201	0.0133	7.12E-317	8.44E-22	5.05E-21

	13	0.0685	0.0093	0.006	3.56E-236	8.78E-06	2.42E-05
	14	0.0576	-0.0016	-0.0015	4.18E-167	0.431697938	0.299962387
21-day sum	**0**	**0.1892**	**0.1316**	**0.0891**	**0**	**0**	**0**
	1	0.1787	0.1186	0.0805	0	0	0
	2	0.1683	0.1054	0.0716	0	0	0
	3	0.1558	0.0908	0.0617	0	0	0
	4	0.1448	0.0782	0.0531	0	3.35E-307	9.87E-308
	5	0.1341	0.0668	0.0453	0	1.34E-224	1.15E-224
	6	0.1228	0.0561	0.0379	0	9.11E-159	1.05E-157
	7	0.1109	0.0446	0.0299	0	6.85E-101	3.25E-99
	8	0.0995	0.0327	0.0216	0	5.36E-55	8.52E-53
	9	0.0888	0.021	0.0136	0	9.32E-24	5.59E-22
	10	0.0782	0.0099	0.006	9.51E-307	2.28E-06	2.13E-05
	11	0.0678	-0.0007	-0.0012	2.49E-231	0.753843884	0.408427402
	12	0.058	-0.0108	-0.008	2.18E-169	2.60E-07	1.48E-08
	13	0.0466	-0.0221	-0.0157	4.89E-110	4.95E-26	1.45E-28
	14	0.0353	-0.0329	-0.023	6.41E-64	9.47E-56	2.02E-59
28-day sum	**0**	**0.1694**	**0.105**	**0.0712**	**0**	**0**	**0**
	1	0.1572	0.0902	0.0613	0	0	0
	2	0.1453	0.0752	0.0511	0	5.47E-284	3.15E-285
	3	0.1319	0.0599	0.0406	0	1.32E-180	4.74E-181
	4	0.1202	0.0467	0.0315	0	1.07E-110	8.49E-110
	5	0.109	0.0347	0.0233	0	8.19E-62	1.23E-60
	6	0.0972	0.0231	0.0153	0	1.84E-28	4.57E-27
	7	0.0849	0.0116	0.0074	0	2.86E-08	1.74E-07
	8	0.0736	0.0009	0.0001	1.09E-271	0.65175688	0.960639372
	9	0.0625	-0.0094	-0.0071	1.20E-196	6.38E-06	5.02E-07
	10	0.0514	-0.0197	-0.0142	1.48E-133	3.87E-21	1.49E-23
	11	0.0403	-0.0299	-0.021	7.27E-83	1.98E-46	5.88E-50
	12	0.0301	-0.0393	-0.0274	5.61E-47	1.32E-78	2.31E-83
	13	0.0191	-0.0498	-0.0345	6.61E-20	2.17E-125	3.57E-131
	14	0.0083	-0.0601	-0.0414	7.05E-05	8.87E-182	3.45E-188

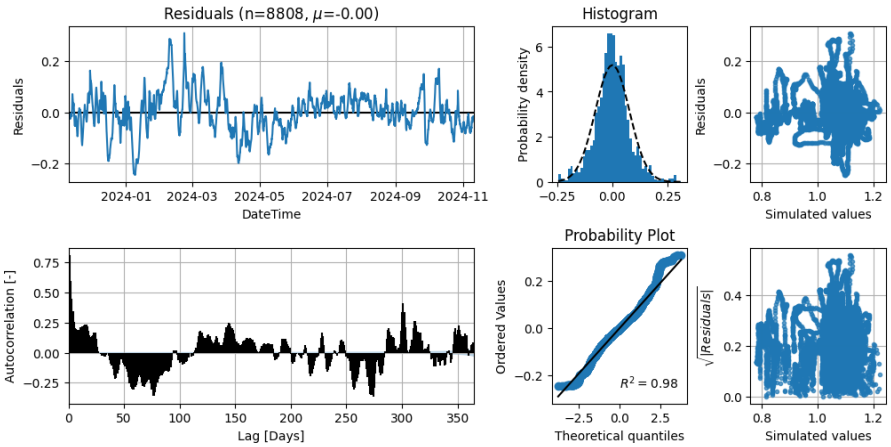
B. Diagnostic Result

Scenario 1 – Calibration Stage

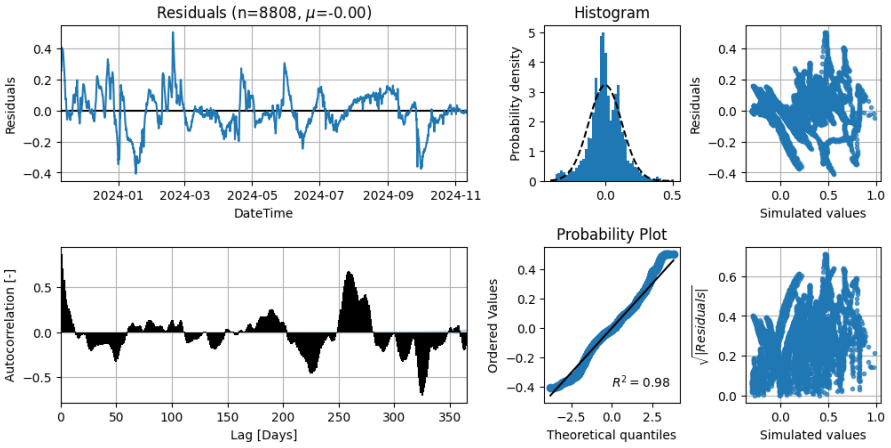
MB014-PB1



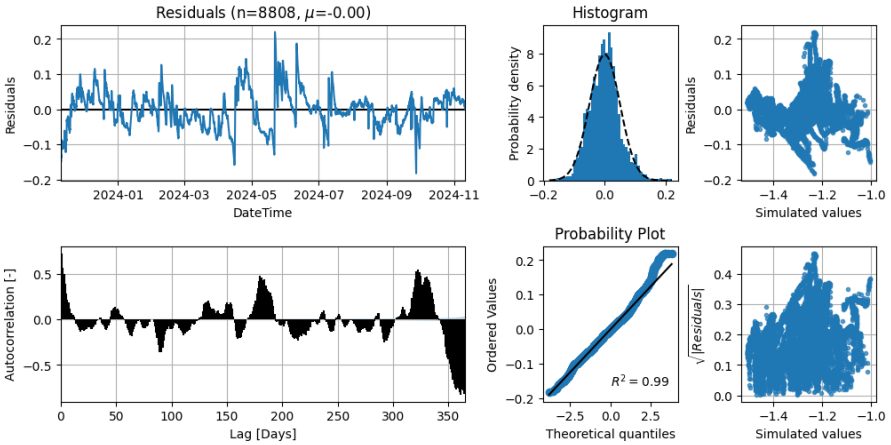
MB014-PB2



HB027-PB1

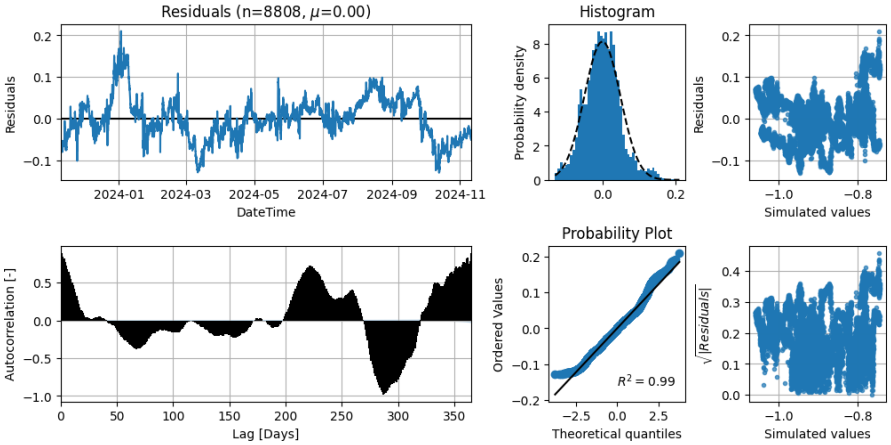


MB015-PB2

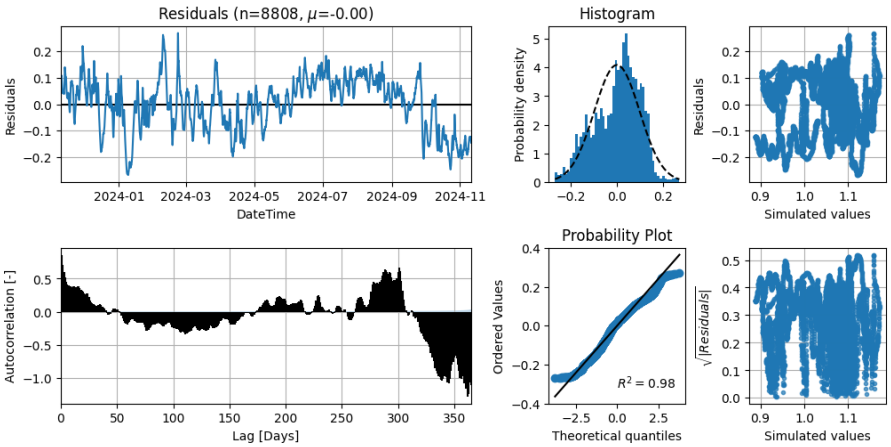


Scenario 1 – After Parameter Fixing

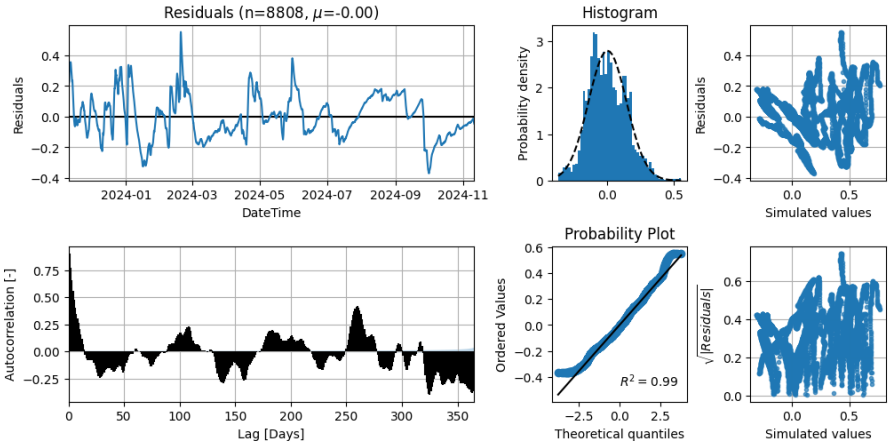
MB014-PB1



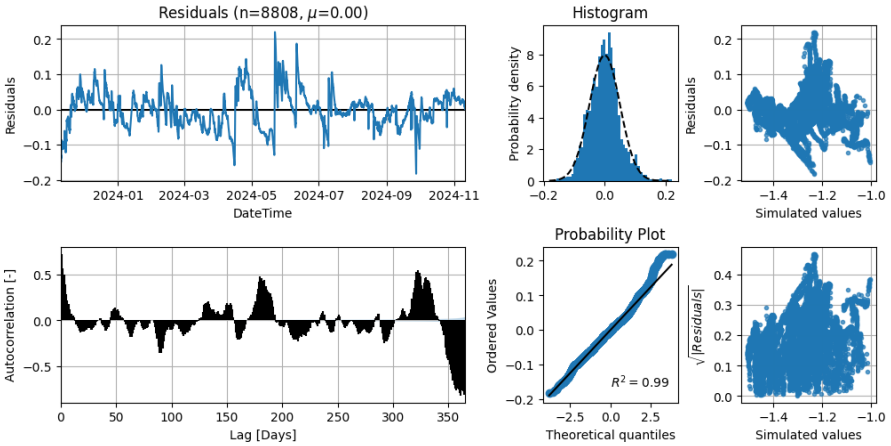
MB014-PB2



HB027-PB1

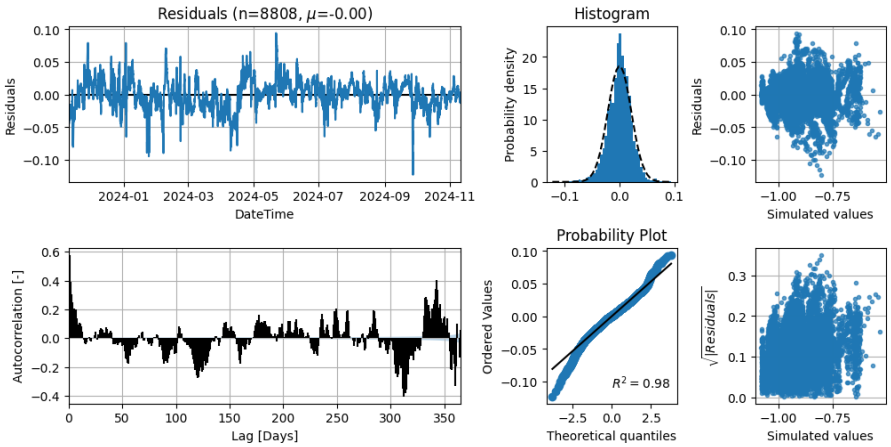


MB015-PB2

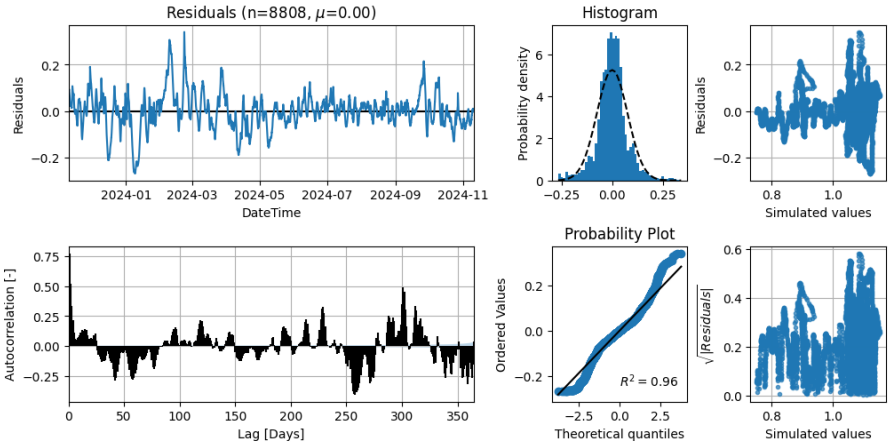


Scenario 2 – Calibration Stage

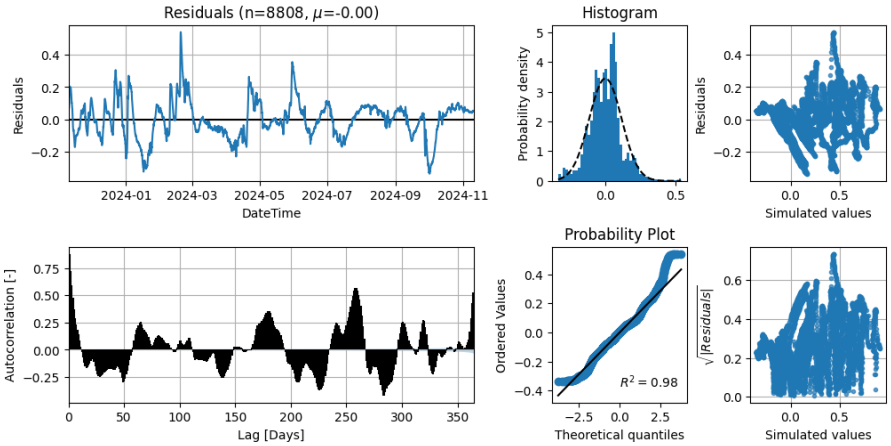
MB014-PB1



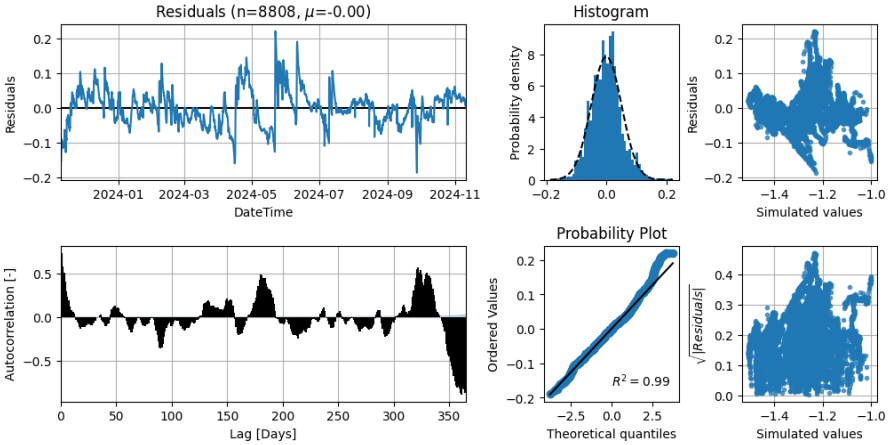
MB014-PB2



HB027-PB1

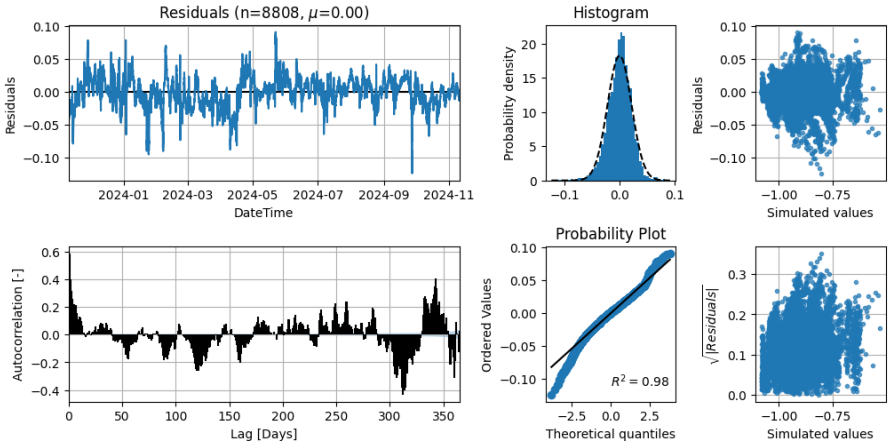


MB015-PB2

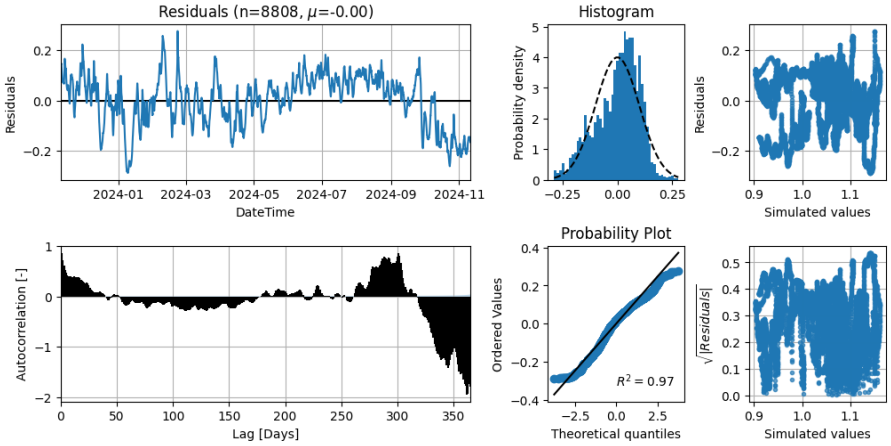


Scenario 2 – After Parameter Fixing

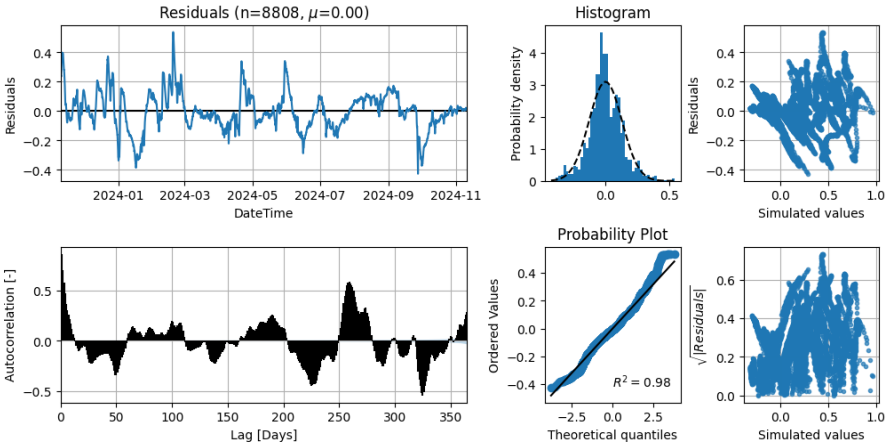
MB014-PB1



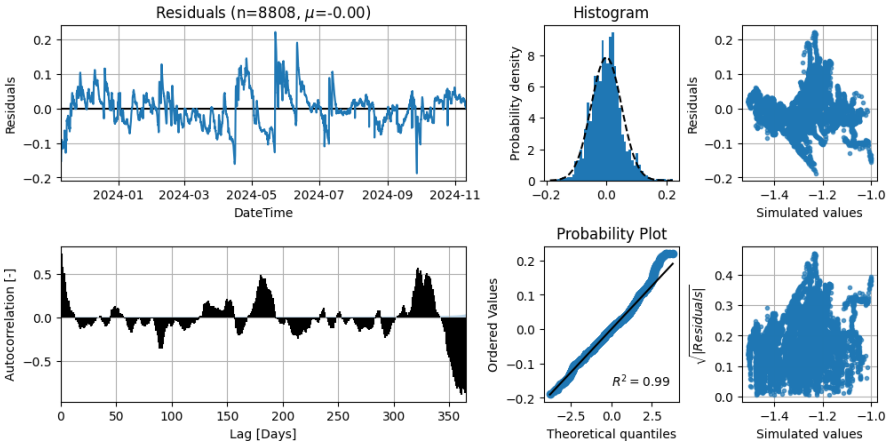
MB014-PB2



HB027-PB1

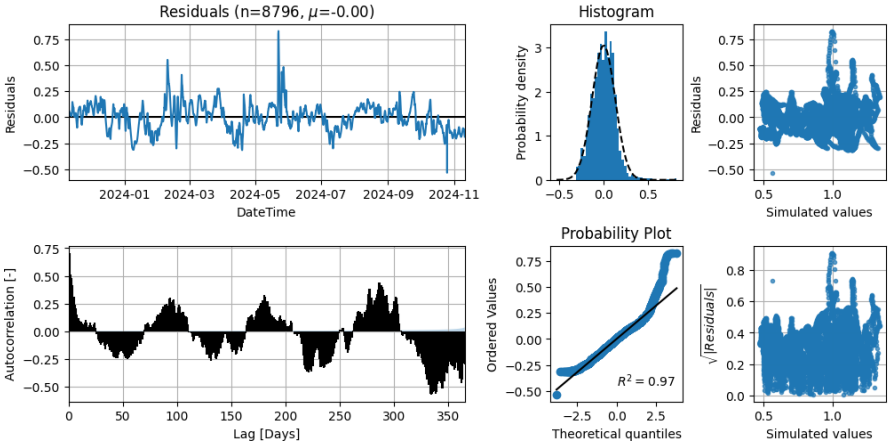


MB015-PB2

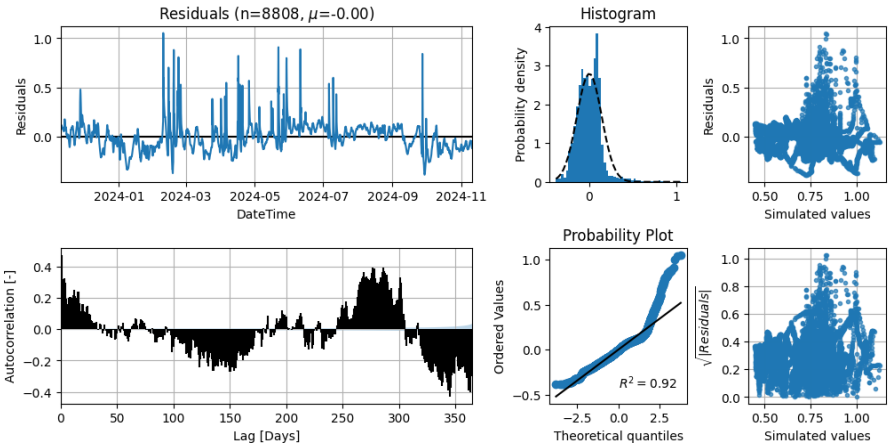


Scenario 3 – Calibration Stage

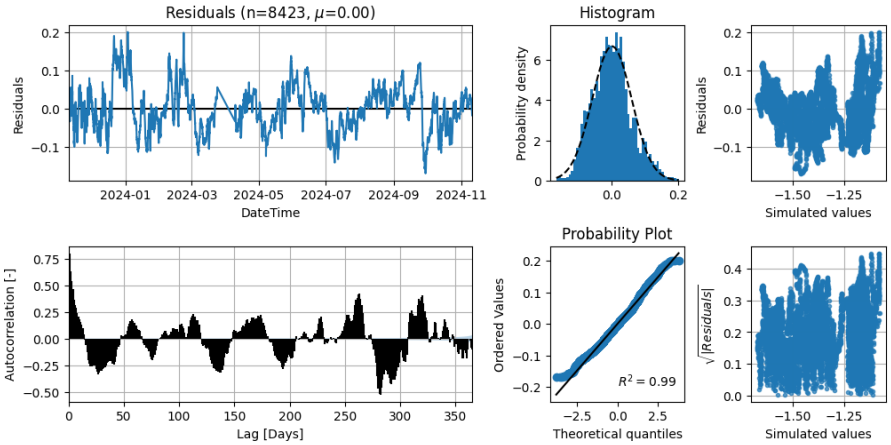
MB12-PB1



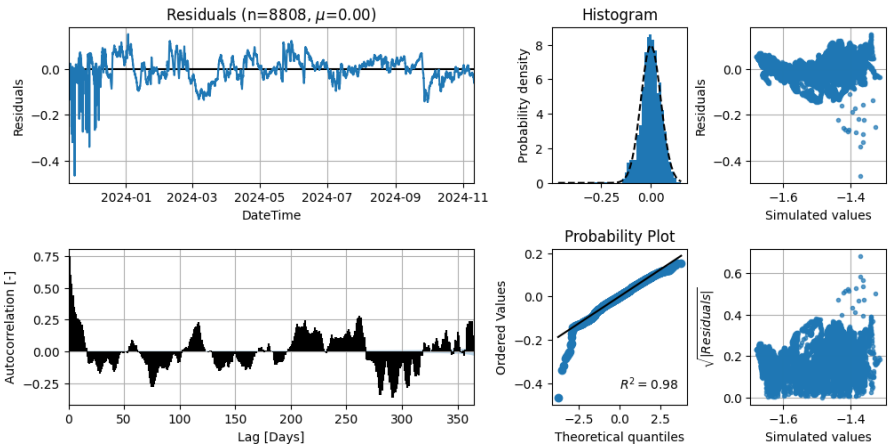
HB25-PB1



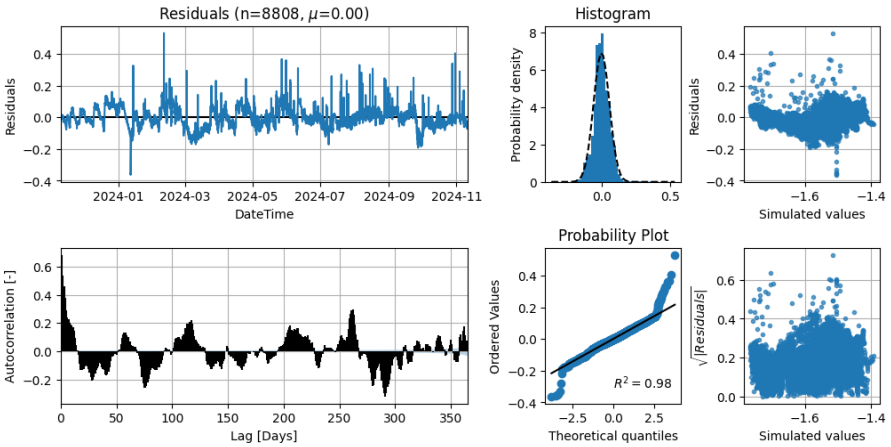
MB26-PB2



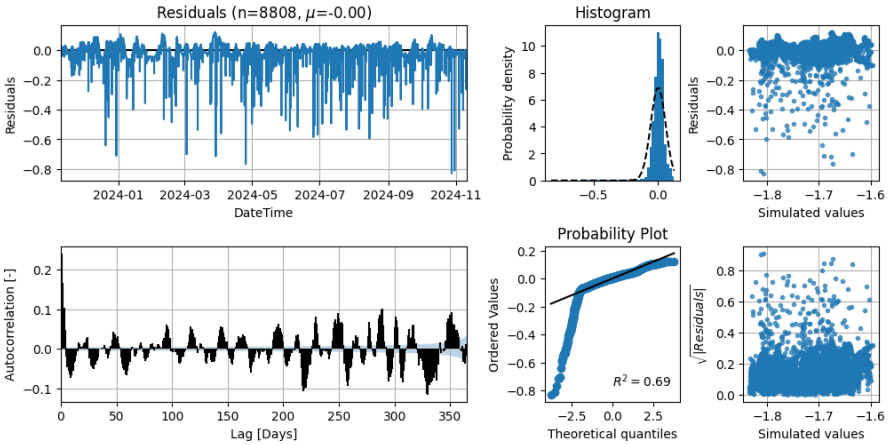
MB26-PB1



MB13-PB2

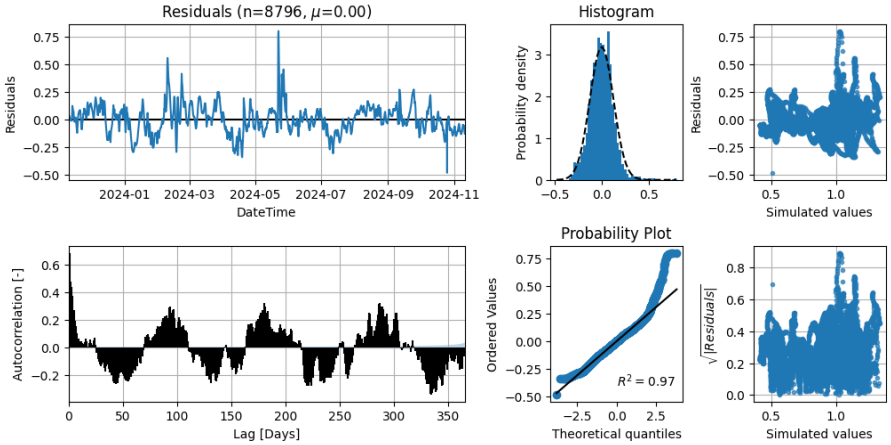


MB13-PB1

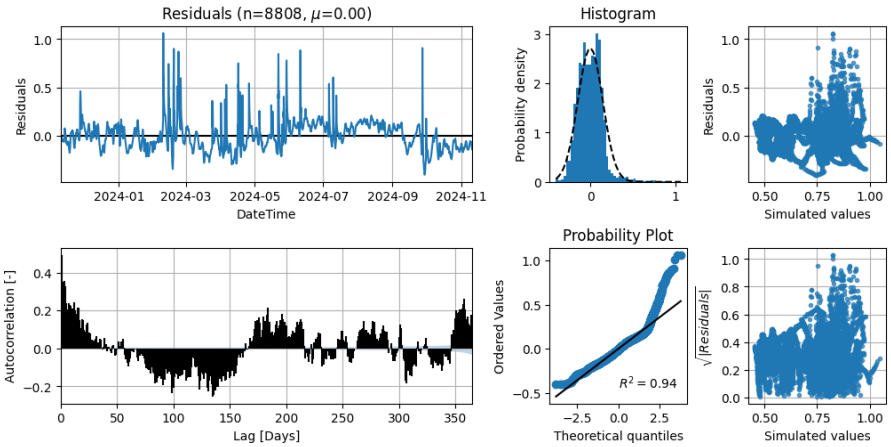


Scenario 3 – After Parameter Fixing

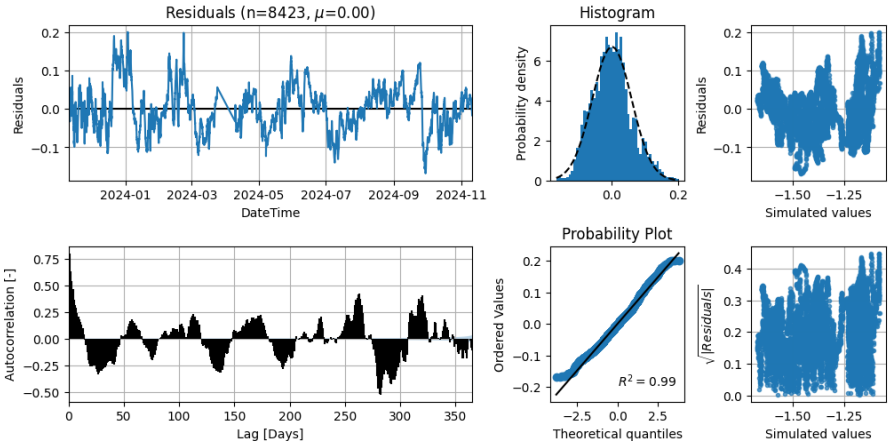
MB12-PB1



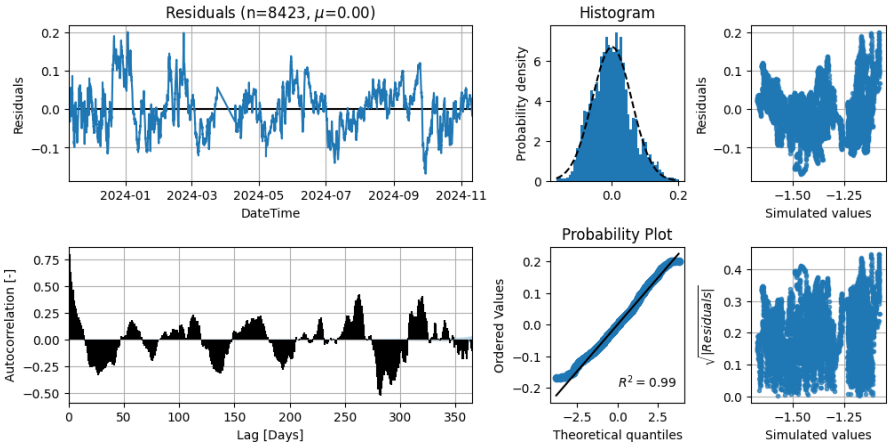
HB25-PB1



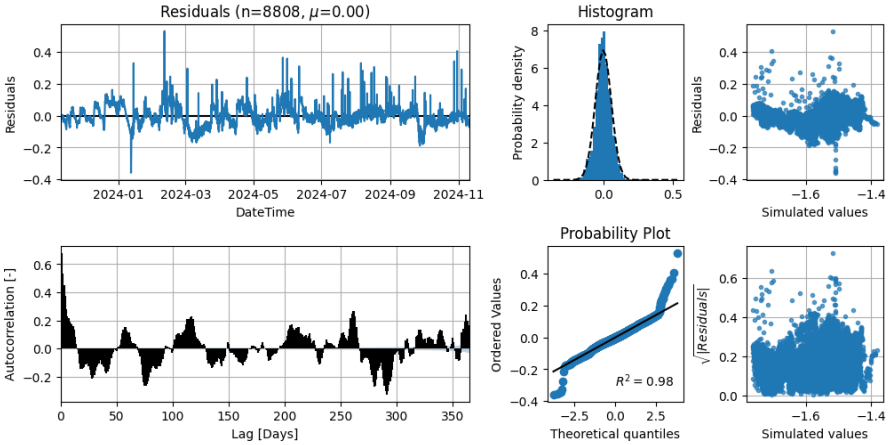
MB26-PB2



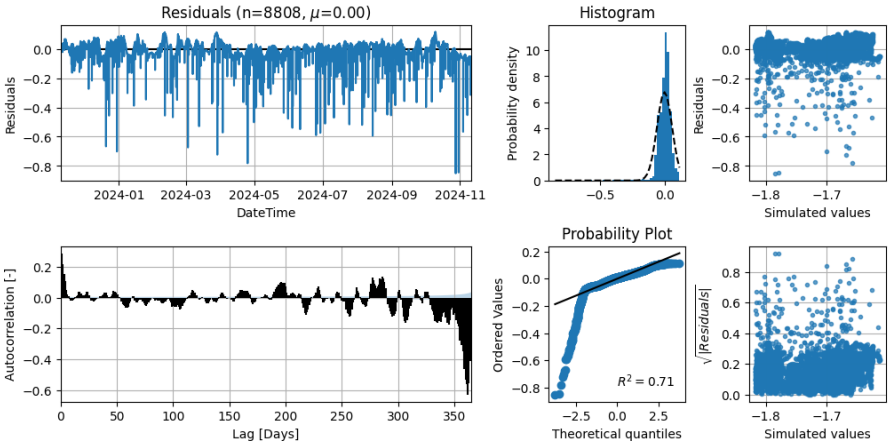
MB26-PB1



MB13-PB2

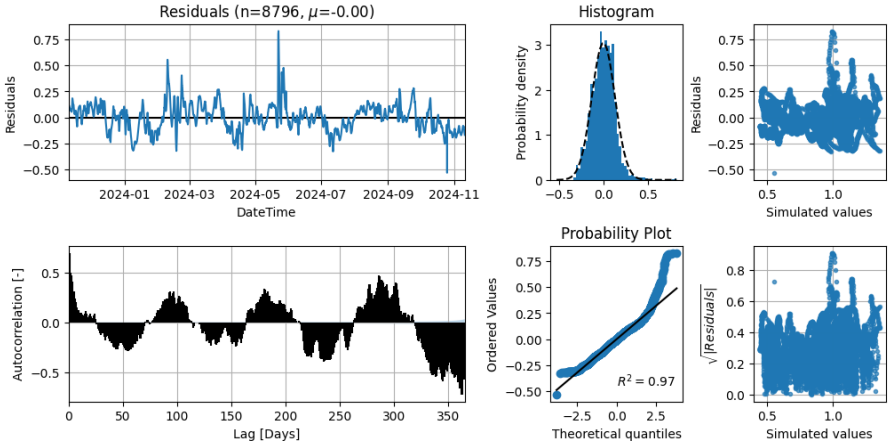


MB13-PB1

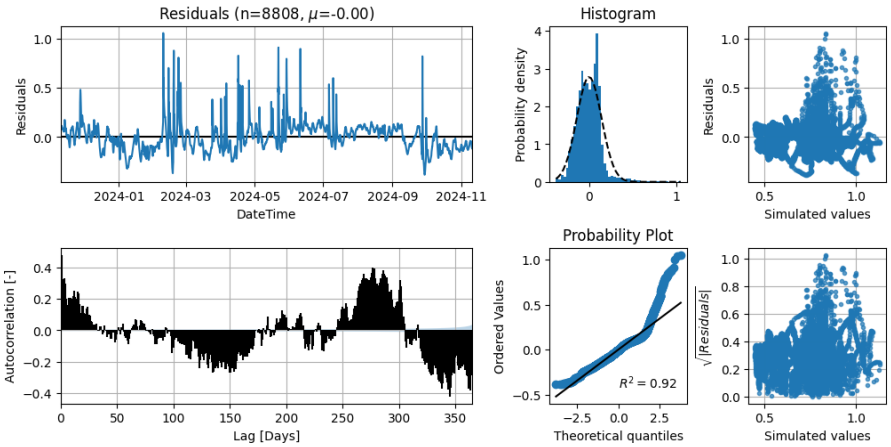


Scenario 4 – Calibration Stage

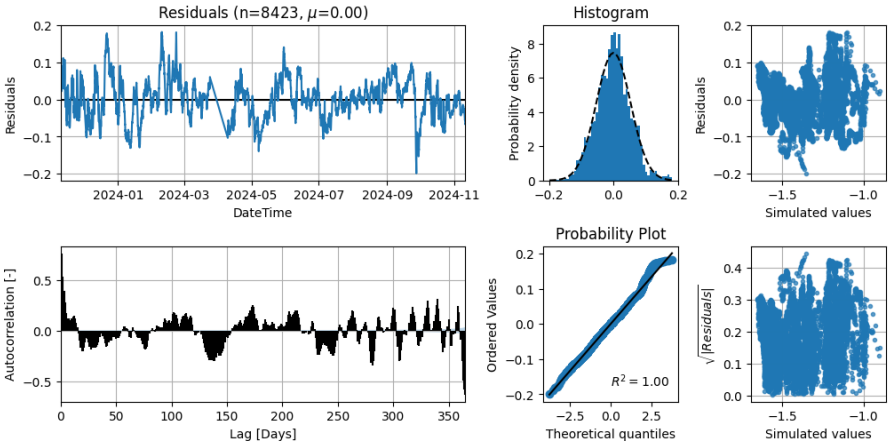
MB12-PB1



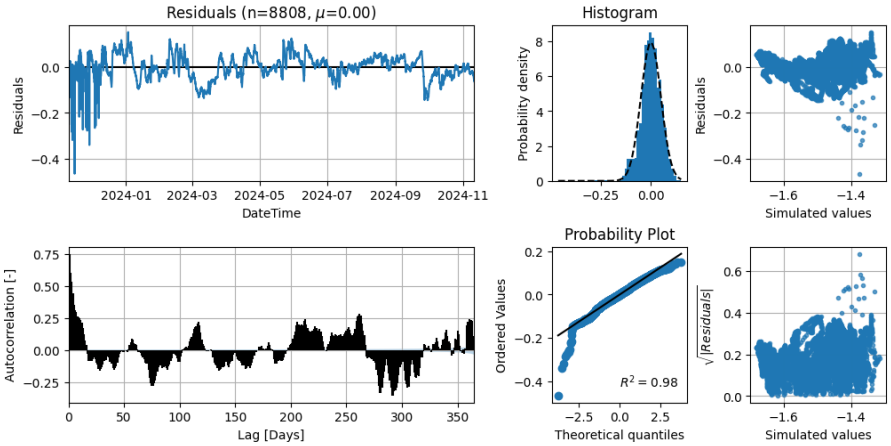
HB25-PB1



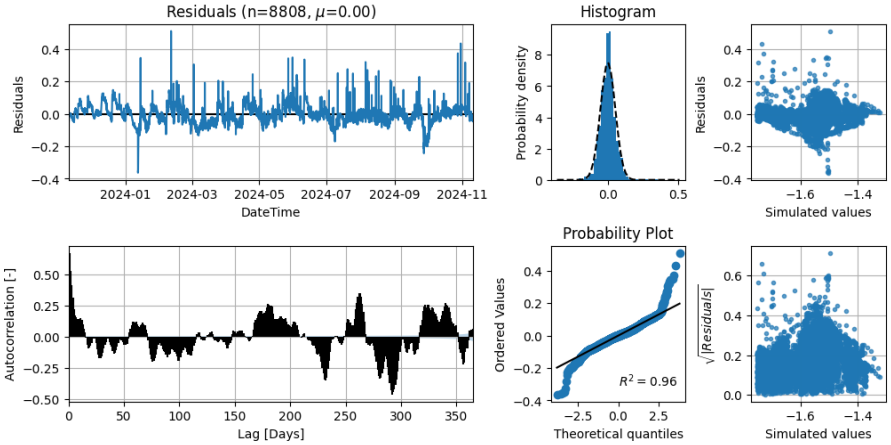
MB26-PB2



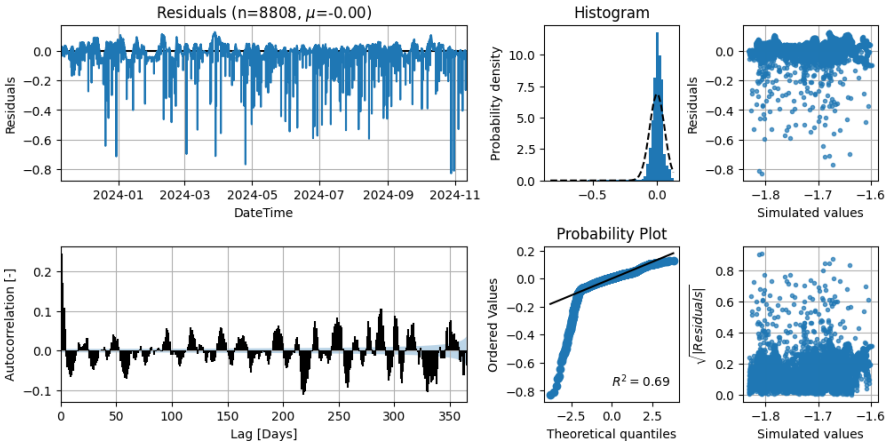
MB26-PB1



MB13-PB2

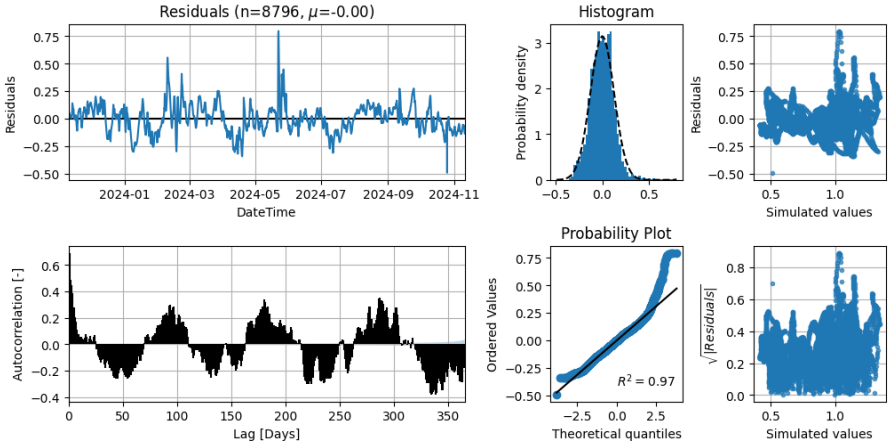


MB13-PB1

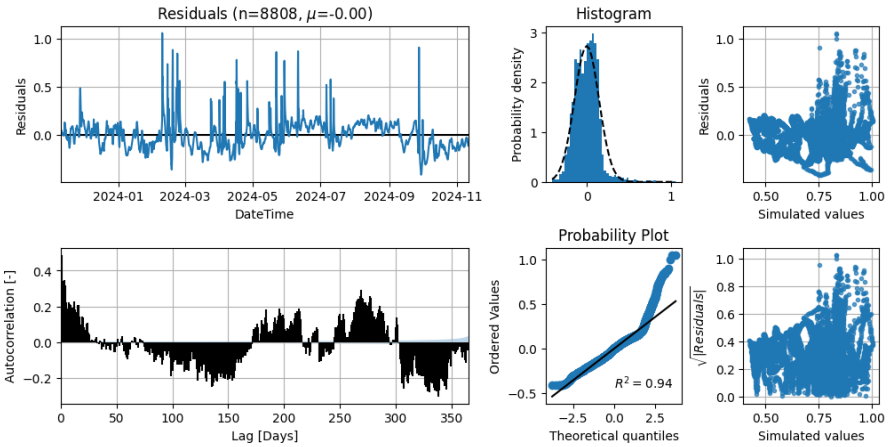


Scenario 4 – After Parameter Fixing

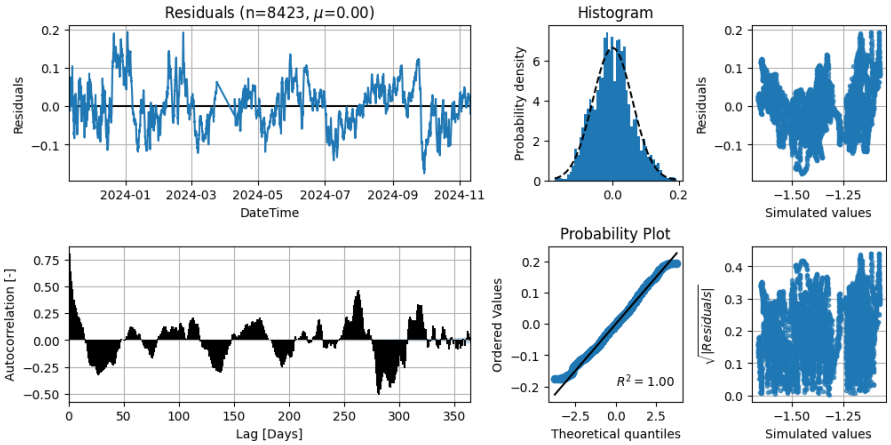
MB12-PB1



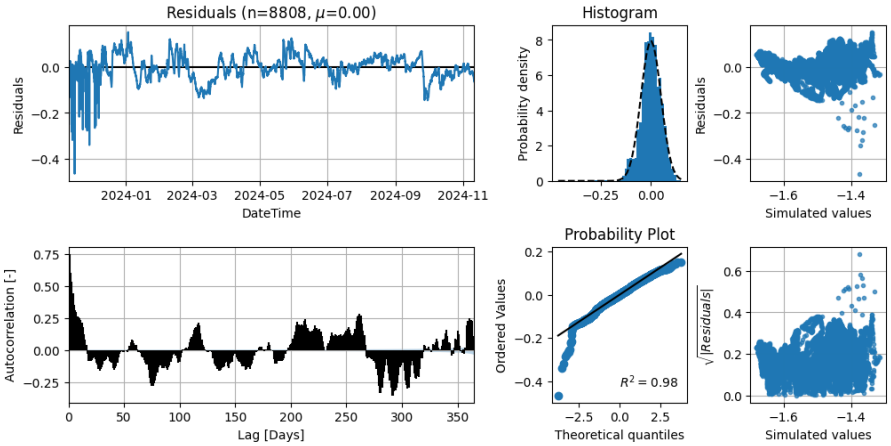
HB25-PB1



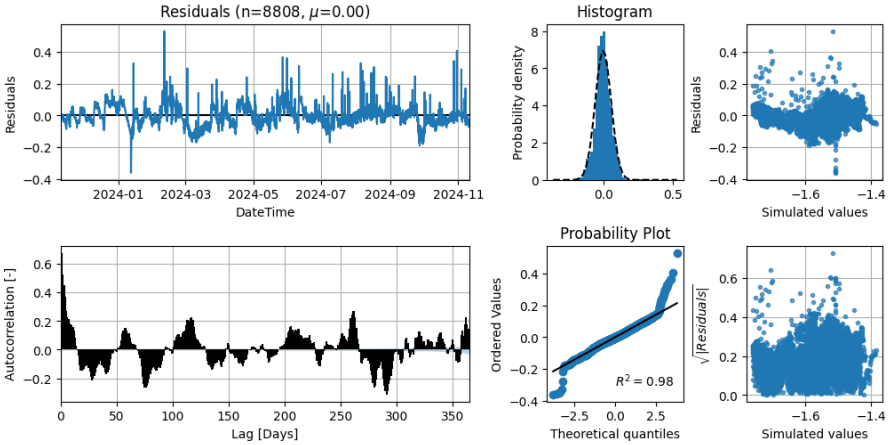
MB26-PB2



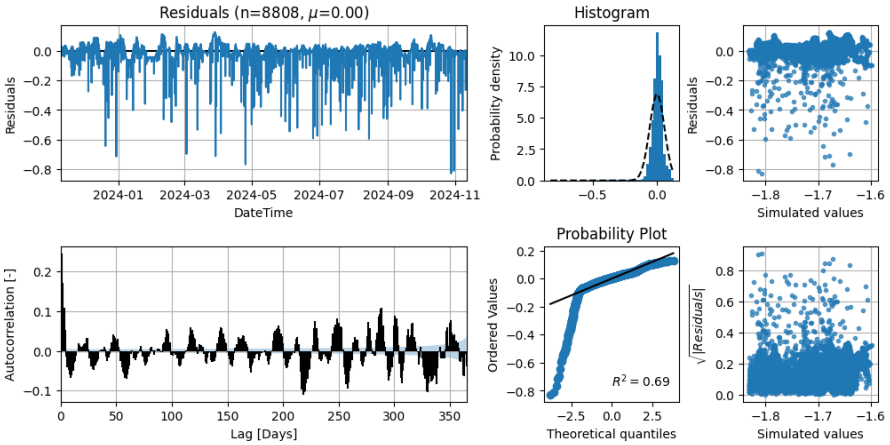
MB26-PB1



MB13-PB2

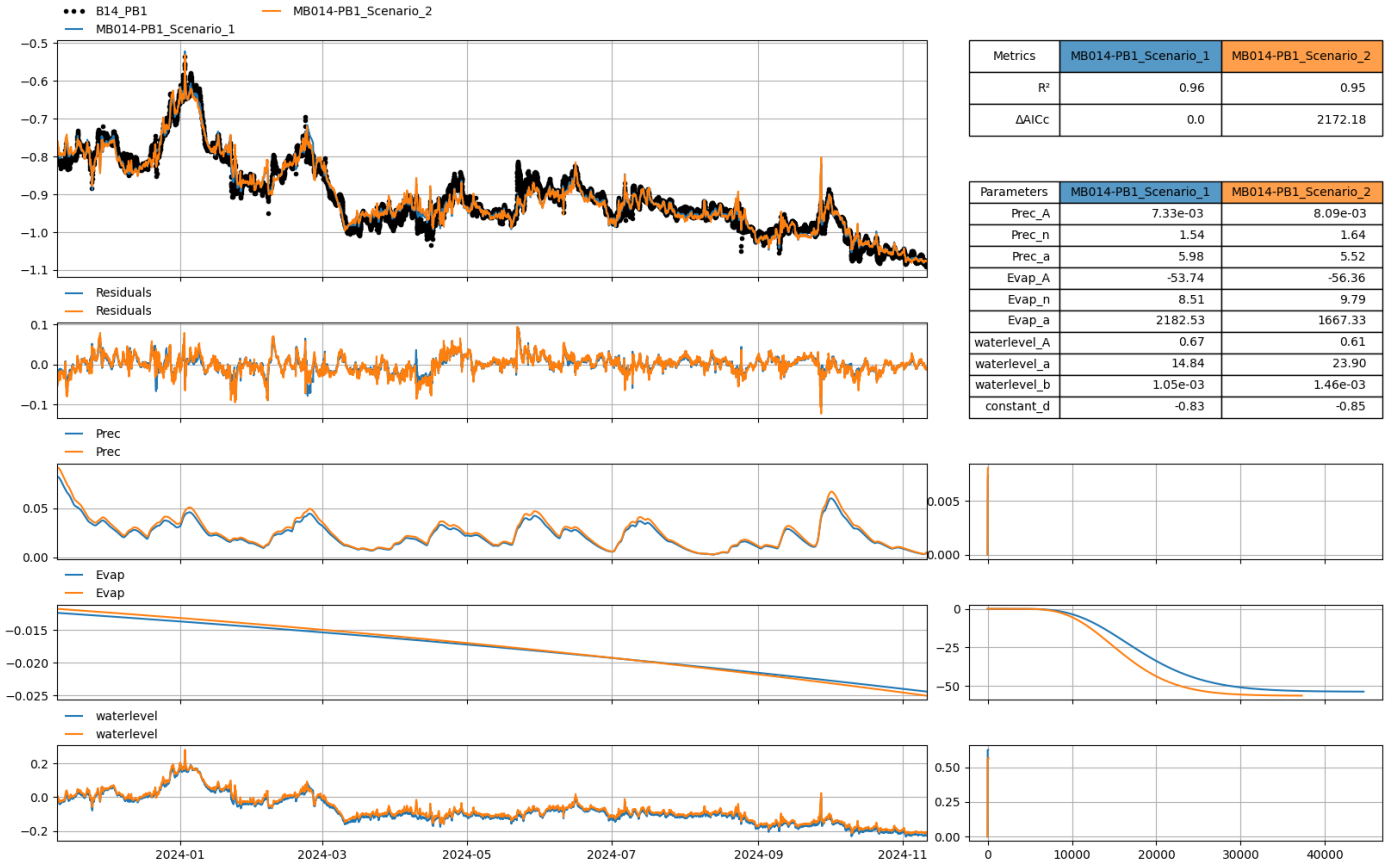


MB13-PB1


Residuals (n=8808, μ=0.00)
Residuals
DateTime
2024-01
2024-03
2024-05
2024-07
2024-09
2024-11
Histogram
Probability density
Residuals
Simulated values
Autocorrelation [-]
Lag [Days]
Probability Plot
Ordered Values
Theoretical quantiles
$R^2 = 0.69$
$\sqrt{|Residuals|}$
Simulated values

C. Groundwater Model Comparisons

Scenario 1 vs 2



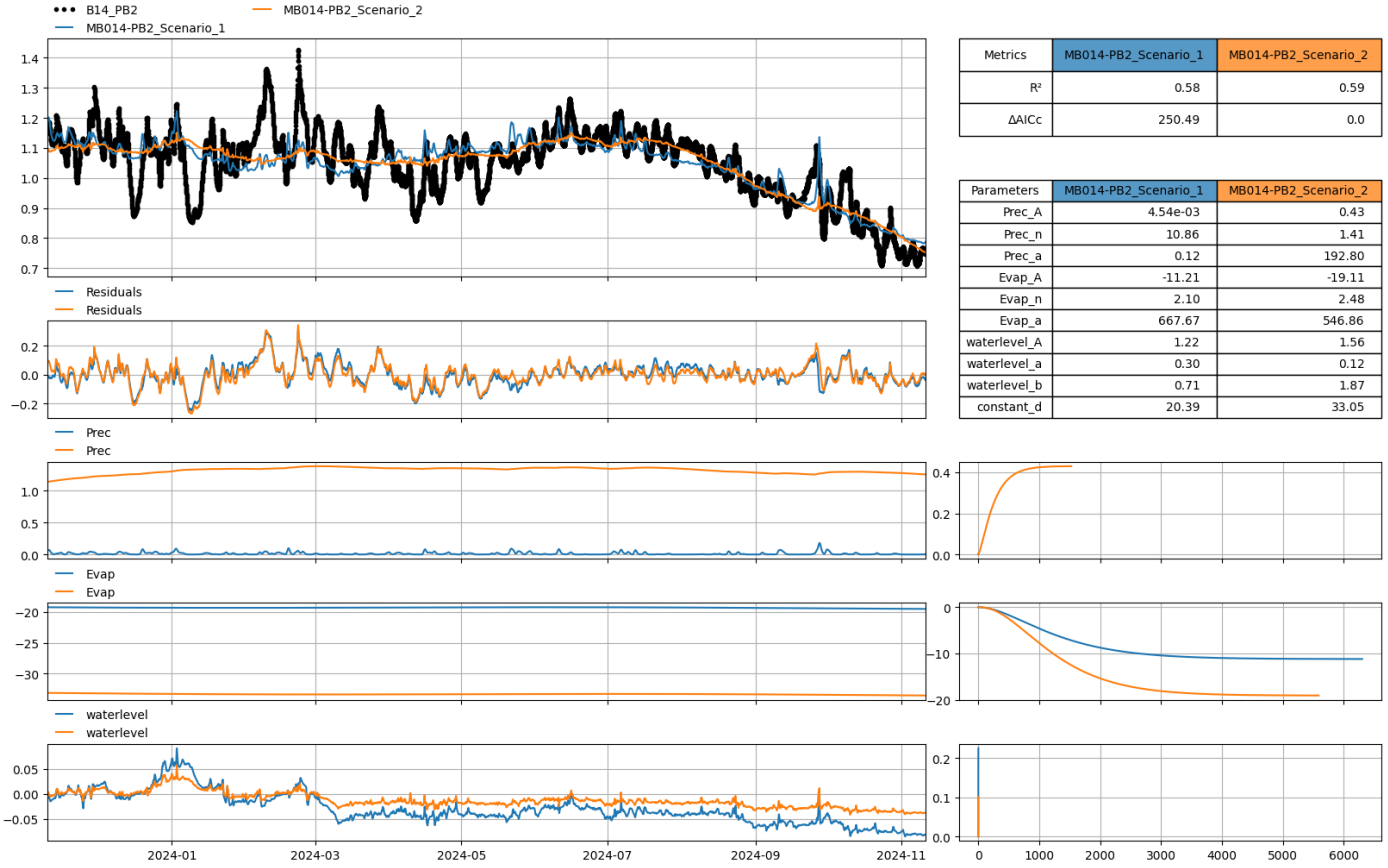
Metrics	MB014-PB1_Scenario_1	MB014-PB1_Scenario_2
R^2	0.96	0.95
ΔAICc	0.0	2172.18

Parameters	MB014-PB1_Scenario_1	MB014-PB1_Scenario_2
Prec_A	7.33e-03	8.09e-03
Prec_n	1.54	1.64
Prec_a	5.98	5.52
Evap_A	-53.74	-56.36
Evap_n	8.51	9.79
Evap_a	2182.53	1667.33
waterlevel_A	0.67	0.61
waterlevel_a	14.84	23.90
waterlevel_b	1.05e-03	1.46e-03
constant_d	-0.83	-0.85

Figure C.1 Mosaic plot result for MB014-PB1 observation well (calibration stage) between Scenario 1 and 2. Coloured lines correspond to model-predicted measurement and black markers for observations.

Figure C.1 shows the result of groundwater modeling analysis using Pastas for monitoring well MB014_PB1 between Scenario 1 and 2. The top-left graph illustrates the comparison between observed groundwater heads together with model predictions from Scenario 1 and 2. Visually, both Scenario 1 and 2 follows the observed trend well with only few significant departures: slight under-predicted during end of 2024-05, slight over-predicted at the end of 2024-09 and sharp peak over-predicted around 2024-10. This deviation can also be seen in the residual plot directly under the comparison plot, with most variation hovering between 0.1 and -0.1 m except for that one sharp peak with -0.12 m. The right-top table presents statistical measure of fit for both models, with R^2 value shows 0.96 for Scenario 1 and 0.95 for Scenario 2, indicating high correlation for both scenarios.

The last three plots illustrate the effect of the corresponding stresses (precipitation, evaporation, and water level) on the left side, and their corresponding step response plots in the right side. For precipitation stress, the contribution toward model prediction is overall positive, with no significant difference between Scenario 1 and Scenario 2. Evapotranspiration stress shows linear negative trend and keeps decreasing for both scenario 1 and 2. Water level contribution plot shows erratic behavior, with distinct positive peak at 2024-01 followed by gradual decline that stabilizes at negative range value beginning in 2024-03. From step response plot, both precipitation and water level stresses show almost vertical line and no plateau which indicates very quick initial response which is the characteristic of shallow aquifers, while evapotranspiration step response for both scenario shows no initial response, and over time both responses drop until reaching stability around -50, which indicates negative step response. Blue curve (Scenario 1) tends to be slower system compared to orange curve (Scenario 2) which drops faster and earlier.



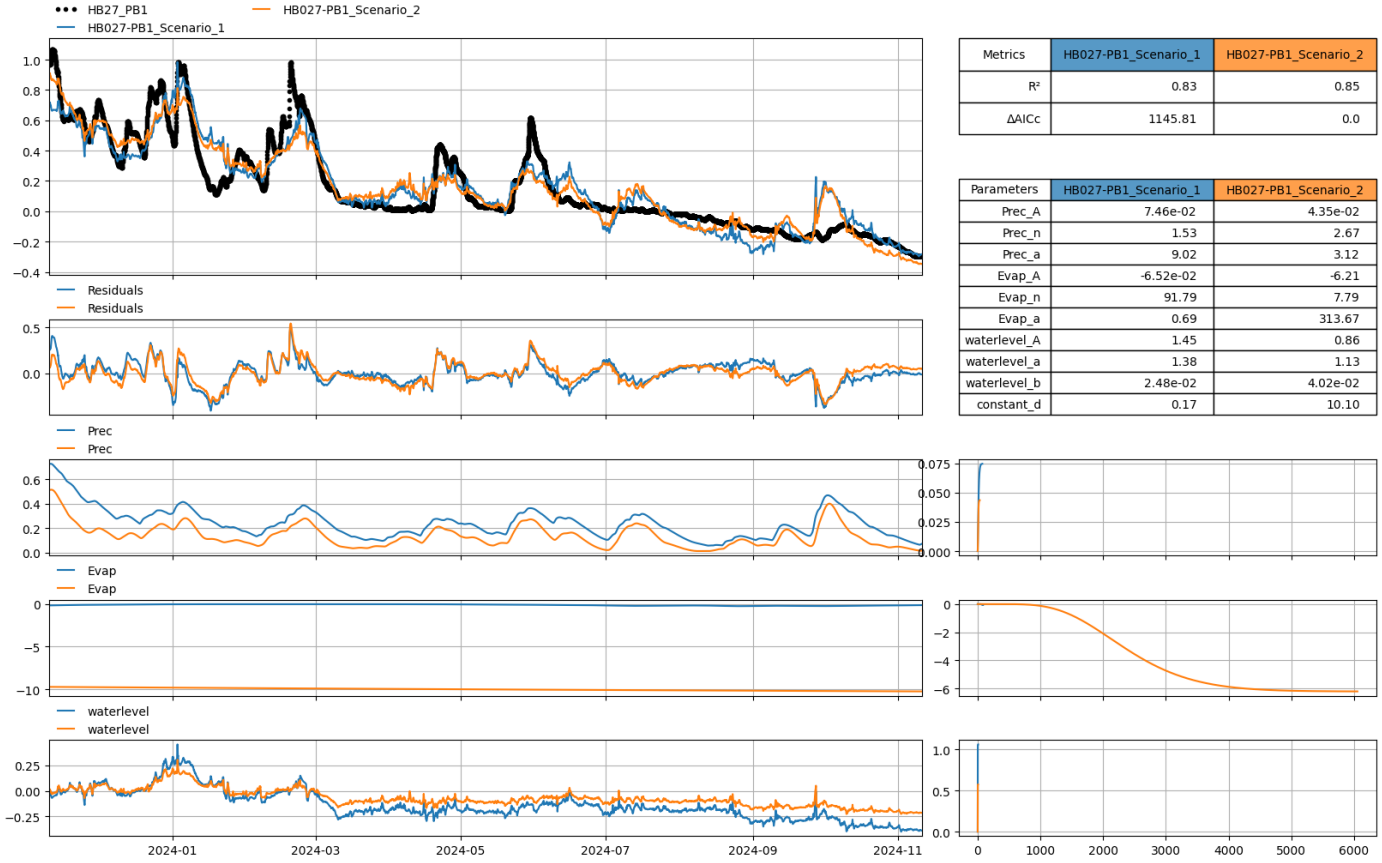
Metrics	MB014-PB2_Scenario_1	MB014-PB2_Scenario_2
R^2	0.58	0.59
ΔAICc	250.49	0.0

Parameters	MB014-PB2_Scenario_1	MB014-PB2_Scenario_2
Prec_A	4.54e-03	0.43
Prec_n	10.86	1.41
Prec_a	0.12	192.80
Evap_A	-11.21	-19.11
Evap_n	2.10	2.48
Evap_a	667.67	546.86
waterlevel_A	1.22	1.56
waterlevel_a	0.30	0.12
waterlevel_b	0.71	1.87
constant_d	20.39	33.05

Figure C.2 Mosaic plot result for MB014-PB2 observation well (calibration stage) between Scenario 1 and 2. Coloured lines correspond to model-predicted measurement and black markers for observations.

Figure C.2 illustrate the model result for MB014-PB2 groundwater well. For observation versus model prediction plot, both Scenario generally follow the overall trend from the observation with R^2 value of 0.58 for Scenario 1 and slightly higher for Scenario 2 with 0.59. However, both scenarios tend to underpredict the rapid declines and sharp peaks. Scenario 1 predicts sharp increase in groundwater level around 2024-10, which is closer to actual observations compared to Scenario 2. The residual plot overall confirms the previous findings, showing that the largest residuals occur at the highest peaks and deepest troughs of the observation. Precipitation contribution shows different pattern, with Scenario 1 shown a very small amplitude teetering at zero, as shown in Prec_A model parameters of 4.54e-03, and more unstable behavior compared to Scenario 2 which shows smoother line and higher contribution. This is further confirmed by Scenario 1's larger shape (Prec_n) parameter and smaller scale (Prec_a) parameter.

Evaporation contribution plot depicts an almost horizontal line at large negative value, with smaller amplitude for Scenario 2 (-19.11) compared to Scenario 1 (-11.22). From step response plot, precipitation for Scenario 2 shows gradual increase which then plateau until reaching steady state at t = 1500 days. Scenario 1 precipitation step response has very small magnitude compared to Scenario 2 such that it cannot be shown in the plot. For evaporation step response, Scenario 1 exhibit gradual decline over an extended period, reaching plateau at approximately t = 6500 days, while Scenario 2 reaches its plateau early and achieve steady state around t =5500 days.

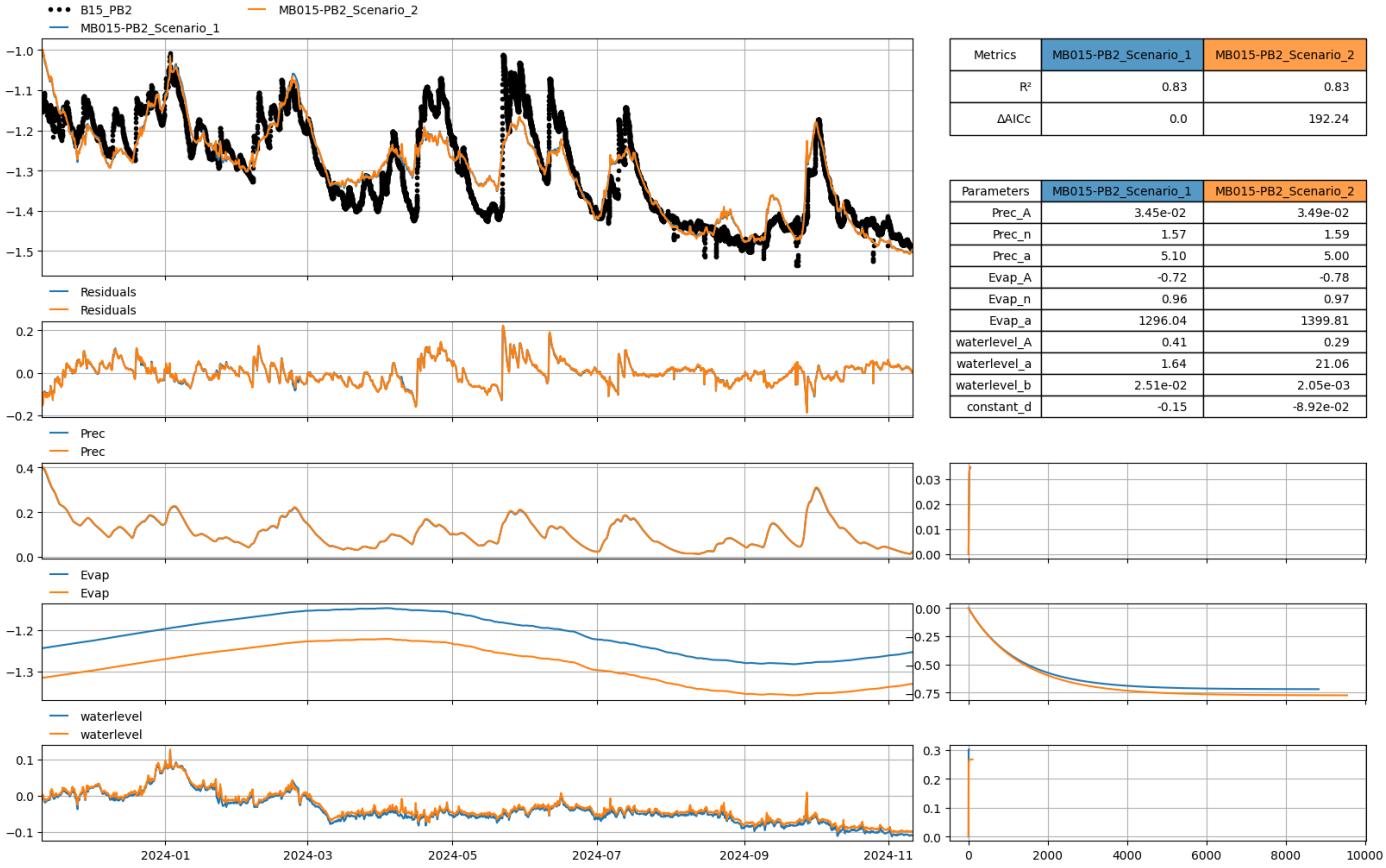


Metrics	HB027-PB1_Scenario_1	HB027-PB1_Scenario_2
R^2	0.83	0.85
ΔAICc	1145.81	0.0

Parameters	HB027-PB1_Scenario_1	HB027-PB1_Scenario_2
Prec_A	7.46e-02	4.35e-02
Prec_n	1.53	2.67
Prec_a	9.02	3.12
Evap_A	-6.52e-02	-6.21
Evap_n	91.79	7.79
Evap_a	0.69	313.67
waterlevel_A	1.45	0.86
waterlevel_a	1.38	1.13
waterlevel_b	2.48e-02	4.02e-02
constant_d	0.17	10.10

Figure C.3 Mosaic plot result for HB027-PB1 observation well (calibration stage) between Scenario 1 and 2. Coloured lines correspond to model-predicted measurement and black markers for observations.

Figure C.3 depicts the HB027-PB1 groundwater level observation and model prediction. For the first plot, with R^2 value of 0.83 for Scenario 1 and 0.85 for Scenario 2, the general shape and trend of actual observation can be captured by both models, with some deviations. Underpredictions observed at some peaks, such as during end of 2023-12, 2024-03 and middle of 2024-06. Different pattern emerges on 2024-10, where both models overestimated the increased level of groundwater by around 0.25 m. The residual show slightly more often variation predominantly less than zero which indicates more overpredictions. The precipitation stress contribution follows a similar pattern to that of MB014-PB1 with higher contribution from Scenario 1 compared to Scenario 2. While both scenarios show a negative contribution from evapotranspiration stress, Scenario 1's effect is less pronounced, with an amplitude of -6.52e-02 compared to -6.21 for Scenario 2. Additionally, the response time for evapotranspiration in Scenario 1 is much shorter, around 150 days, compared to approximately 6000 days for Scenario 2.



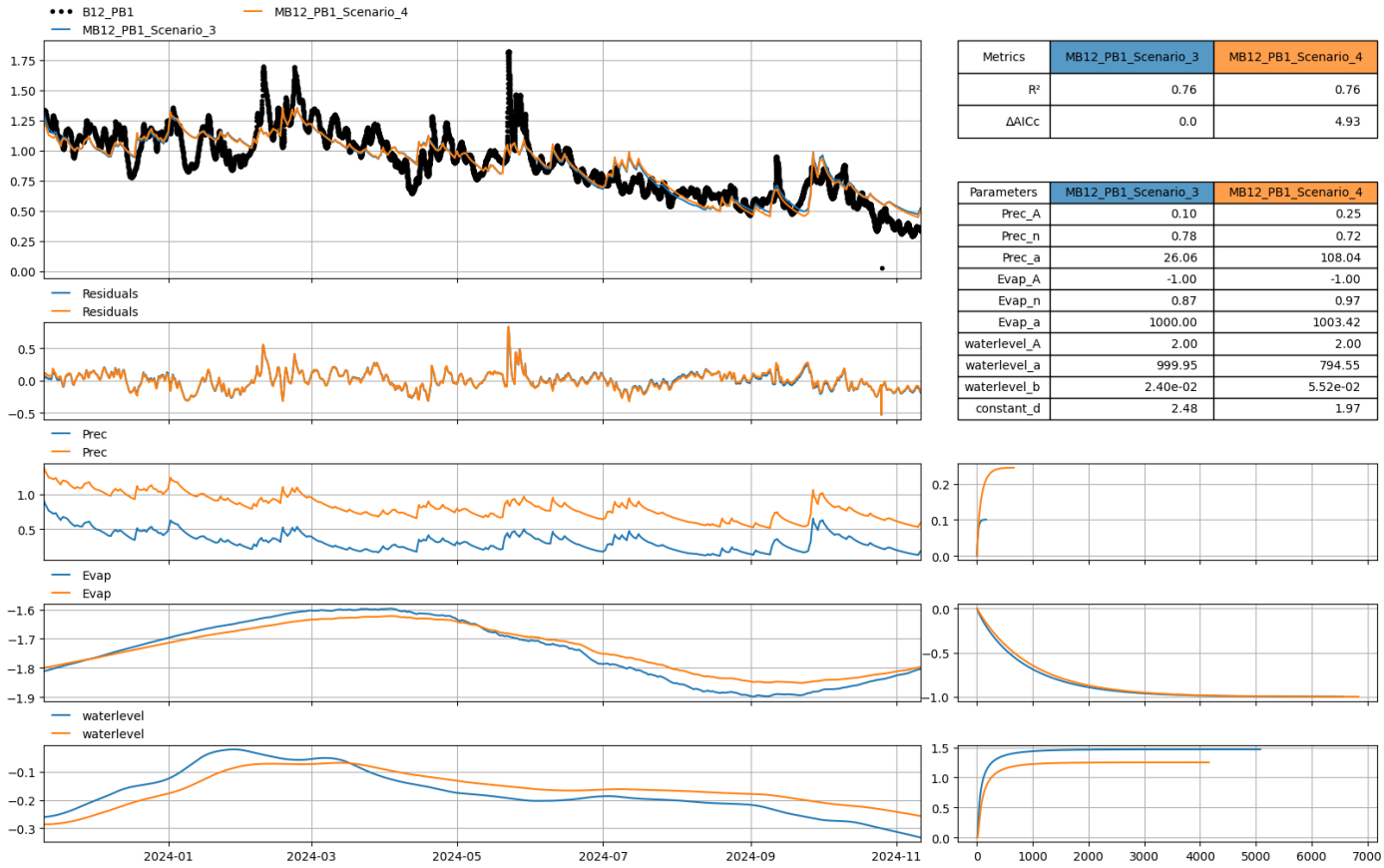
Metrics	MB015-PB2_Scenario_1	MB015-PB2_Scenario_2
R^2	0.83	0.83
ΔAICc	0.0	192.24

Parameters	MB015-PB2_Scenario_1	MB015-PB2_Scenario_2
Prec_A	3.45e-02	3.49e-02
Prec_n	1.57	1.59
Prec_a	5.10	5.00
Evap_A	-0.72	-0.78
Evap_n	0.96	0.97
Evap_a	1296.04	1399.81
waterlevel_A	0.41	0.29
waterlevel_a	1.64	21.06
waterlevel_b	2.51e-02	2.05e-03
constant_d	-0.15	-8.92e-02

Figure C.4 Mosaic plot result for MB015-PB2 observation well (calibration stage) between Scenario 1 and 2. Coloured lines correspond to model-predicted measurement and black markers for observations.

For MB015-PB2 (Figure C.4), the model prediction shows relatively accurate prediction, however between 2024-03 and 2024-08 there are some underpredict tendency. Precipitation pattern between Scenario 1 and 2 follows very similar shape, which reflected on the parameters. Evaporation pattern shows sinusoidal pattern and negative contribution. Water level contribution is quite detail and ranging between +0.1 m and -0.1 m. Both precipitation and water level step response show similar pattern of short, almost instantaneous response time while evapotranspiration step response plot itself show decrease with similar drop between blue line (Scenario 1) and orange line (Scenario 2), which is consistent with physical process when groundwater head drops gradually in response to increased evapotranspiration.

Scenario 3 vs 4



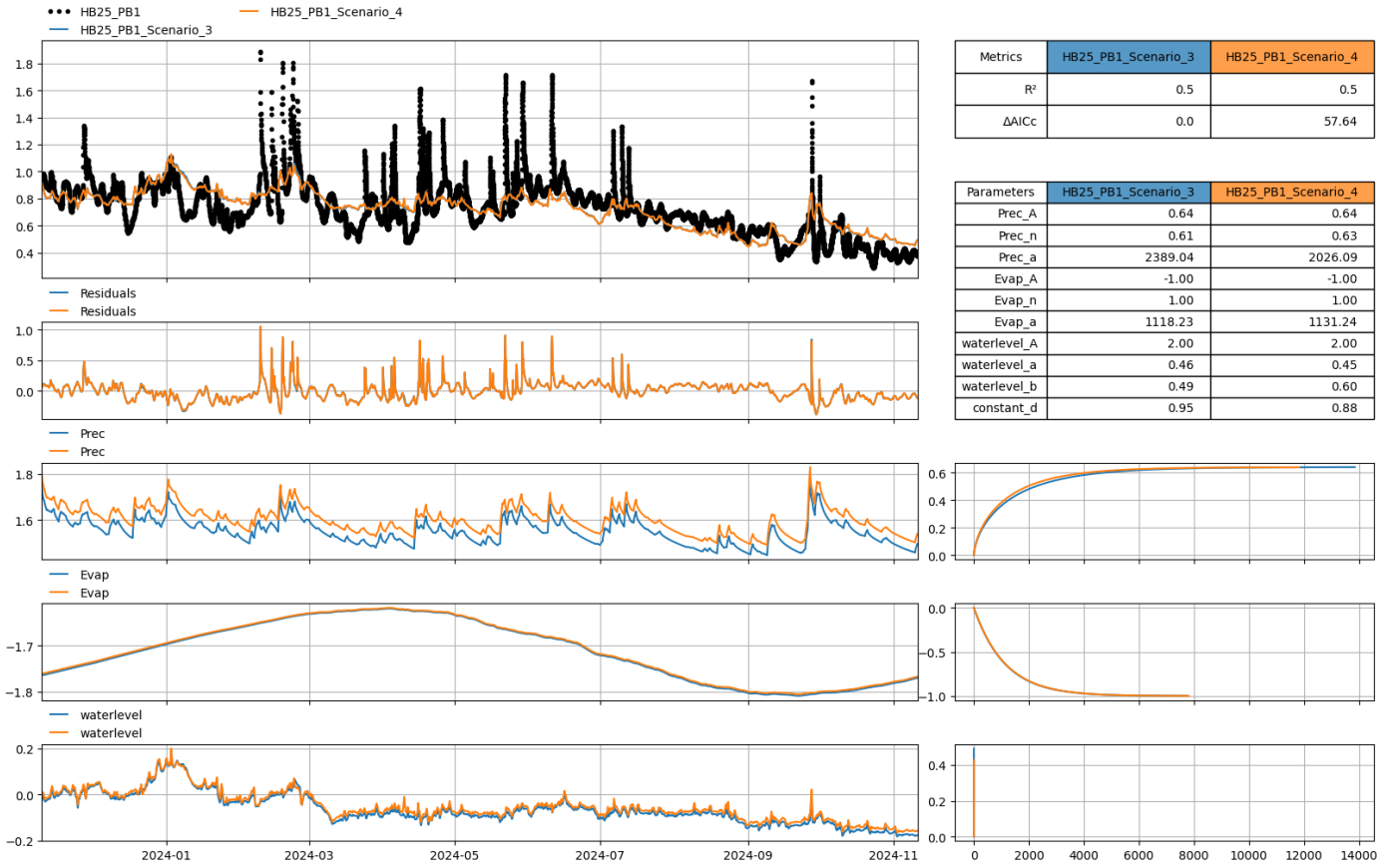
Metrics	MB12_PB1_Scenario_3	MB12_PB1_Scenario_4
R^2	0.76	0.76
ΔAICc	0.0	4.93

Parameters	MB12_PB1_Scenario_3	MB12_PB1_Scenario_4
Prec_A	0.10	0.25
Prec_n	0.78	0.72
Prec_a	26.06	108.04
Evap_A	-1.00	-1.00
Evap_n	0.87	0.97
Evap_a	1000.00	1003.42
waterlevel_A	2.00	2.00
waterlevel_a	999.95	794.55
waterlevel_b	2.40e-02	5.52e-02
constant_d	2.48	1.97

Figure C.5 Mosaic plot result for HB025-PB1 observation well (calibration stage) between Scenario 3 and 4. Coloured lines correspond to model-predicted measurement and black markers for observations.

Figure C.5 compares observed groundwater heads at well MB12_PB1 (black dots) with model predictions under Scenario 3 (blue) and Scenario 4 (orange). In the top-left panel, both scenarios reproduce the overall seasonal decline and short-term fluctuations from January through November 2024, with only minor under-prediction and over-prediction toward the end of the period. The residual plot below shows that most errors remain within ±0.1 m, apart from one spike around mid-2024. The table at upper right reports identical coefficients of determination (R^2 = 0.76) for both scenarios; however, Scenario 3 achieves a lower corrected AIC ($\Delta AICc$ = 0.0 versus 4.93), indicating a better trade-off between goodness of fit and model complexity. The parameter values reveal that Scenario 4 model is fitted using larger precipitation amplitude (Prec_A = 0.25 versus 0.10) and scale parameter (Prec_a = 108.04 versus 26.06) which then compensated by slightly smaller constant offset (constant_d = 1.97 versus 2.48) which resulting in similar value of R^2.

On the left side, the stress-contribution plots illustrate that precipitation inputs raise groundwater levels while both evaporation and water level draw them down. Their corresponding step-response curves on the right confirm a very rapid initial response for both rainfall and water level from Markermeer inputs, typical of a shallow aquifer, while evaporation produces a slower, steadily negative response. Scenario 4 reaches its equilibrium response slightly sooner than Scenario 3, reflecting its stronger recharge coefficient.



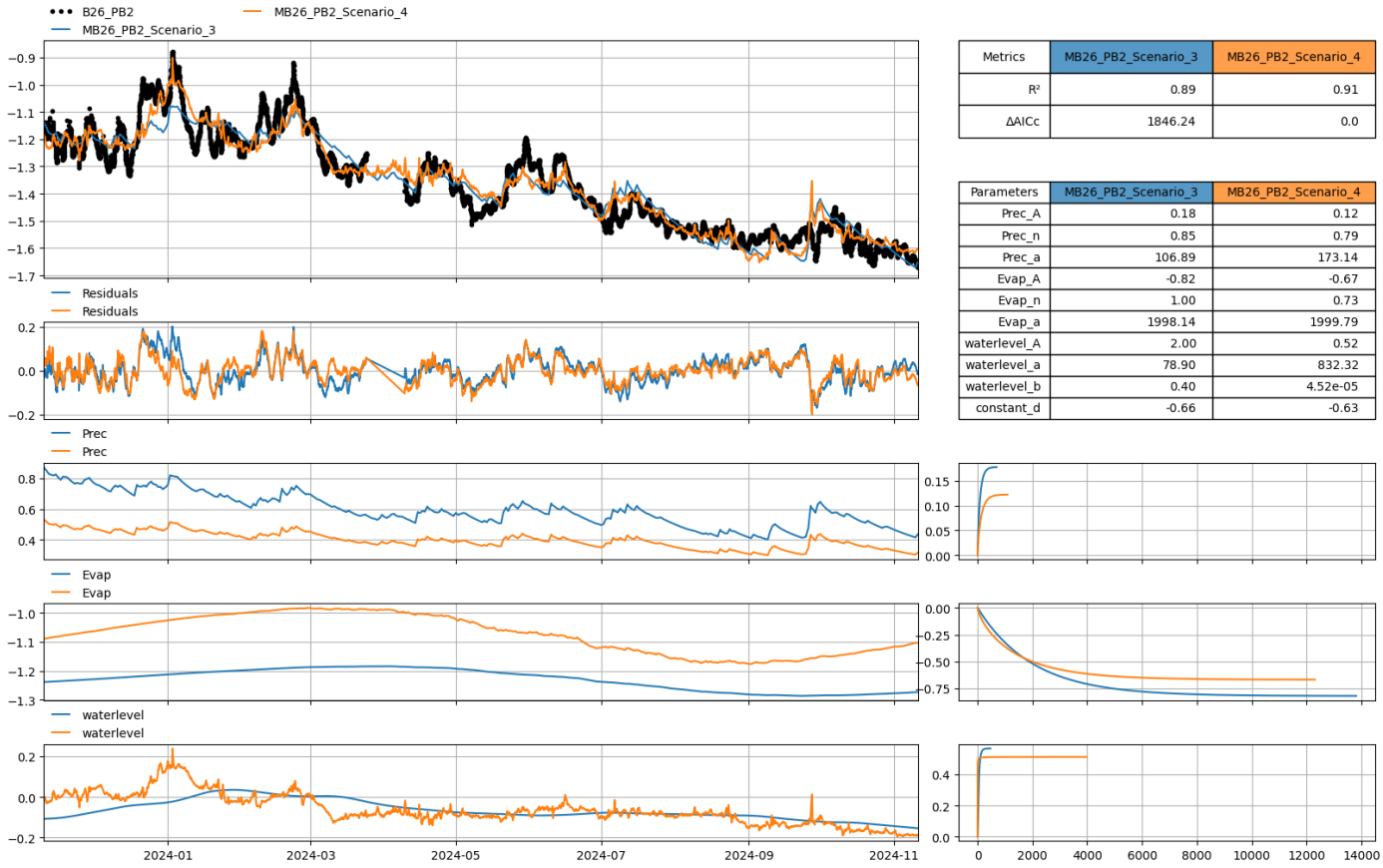
Metrics	HB25_PB1_Scenario_3	HB25_PB1_Scenario_4
R^2	0.5	0.5
ΔAICc	0.0	57.64

Parameters	HB25_PB1_Scenario_3	HB25_PB1_Scenario_4
Prec_A	0.64	0.64
Prec_n	0.61	0.63
Prec_a	2389.04	2026.09
Evap_A	-1.00	-1.00
Evap_n	1.00	1.00
Evap_a	1118.23	1131.24
waterlevel_A	2.00	2.00
waterlevel_a	0.46	0.45
waterlevel_b	0.49	0.60
constant_d	0.95	0.88

Figure C.6 Mosaic plot result for HB025-PB1 observation well (calibration stage) between Scenario 3 and 4. Coloured lines correspond to model-predicted measurement and black markers for observations.

Figure C.6 presents the Pastas hindcast for well HB25_PB1 under Scenario 3 (blue) and Scenario 4 (orange). In the top-left panel, both model runs reproduce the broad seasonal decline and many of the short-term fluctuations seen in the black-dot observations, but neither captures the sharp, isolated spikes in water level. The residual plot shows that the largest errors occur at those peak events, while most residuals remain within ±0.2 m. The metrics table at upper right gives an R^2 of 0.50 for both scenarios, indicating only moderate agreement with the observations. Scenario 3 has the lower corrected AIC ($\Delta AICc$ = 0.0 versus 57.64), suggesting it provides a better compromise between fit and complexity. The parameter table shows that both scenarios share the same precipitation amplitude (Prec_A = 0.64) but Scenario 4 has a slightly smaller precipitation scale (Prec_a = 2026.09 versus 2389.04) and a lower constant offset (constant_d = 0.88 versus 0.95).

On the left, the precipitation contribution plot shows irregular positive pulses tied to rainfall events, with Scenario 3 peaks generally a bit lower than Scenario 4. The evaporation contribution is essentially a steady negative drawdown for both scenarios, with only minor differences in amplitude. The boundary-condition (water-level) contribution follows a seasonal rise and fall, and again the two scenarios overlap closely. The step-response curves on the right confirm these dynamics: rainfall and boundary inputs produce a rapid initial rise that plateaus quickly, while evaporation yields a slower, steadily negative response. Scenario 4 reaches its equilibrium response slightly sooner than Scenario 3, reflecting its marginally stronger weighting of the recharge and boundary stresses.



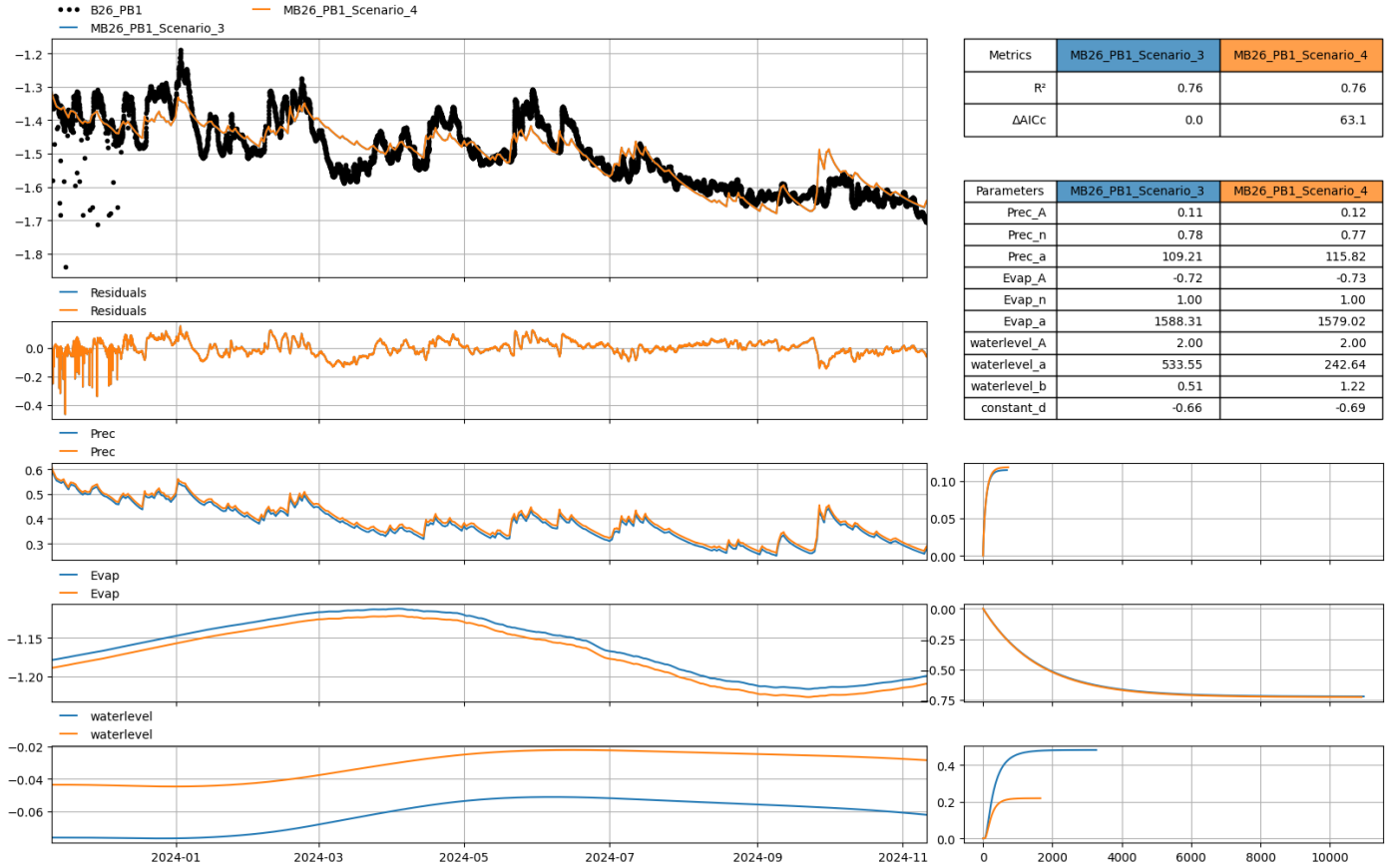
Metrics	MB26_PB2_Scenario_3	MB26_PB2_Scenario_4
R^2	0.89	0.91
ΔAICc	1846.24	0.0

Parameters	MB26_PB2_Scenario_3	MB26_PB2_Scenario_4
Prec_A	0.18	0.12
Prec_n	0.85	0.79
Prec_a	106.89	173.14
Evap_A	-0.82	-0.67
Evap_n	1.00	0.73
Evap_a	1998.14	1999.79
waterlevel_A	2.00	0.52
waterlevel_a	78.90	832.32
waterlevel_b	0.40	4.52e-05
constant_d	-0.66	-0.63

Figure C.7 Mosaic plot result for MB026-PB2 observation well (calibration stage) between Scenario 3 and 4. Coloured lines correspond to model-predicted measurement and black markers for observations.

Figure C.7 shows the Pastas hindcast for well MB26_PB2 under Scenario 3 (blue) and Scenario 4 (orange). In the top-left panel, both scenarios capture the overall seasonal decline and most short-term fluctuations from January to November 2024, but neither reproduces all of the sharp peaks seen in the black-dot observations. The residual plot directly below indicates that most errors lie within ±0.2 m. The table at upper right reports an R^2 of 0.89 for Scenario 3 and 0.91 for Scenario 4, while the corrected AIC strongly favors Scenario 4, suggesting it attains a better balance of fit and complexity. The parameter values reveal that Scenario 4 uses a slightly lower precipitation amplitude (Prec_A = 0.12 versus 0.18) and shape (Prec_n = 0.79 versus 0.85), a weaker negative evapotranspiration coefficient (Evap_A = –0.67 versus –0.82), and a larger boundary-stage amplitude (waterlevel_A = 2.00 versus 0.52) with an effectively zero shape factor (waterlevel_b ≈ 4.5×10^{-5}) compared to Scenario 3 which loses most of its characteristic time scale and shape. Scenario 4 also has a slightly smaller constant offset (constant_d = –0.63 versus –0.66).

On the left, the stress-contribution plots show that rainfall events generate positive pulses that decline in magnitude over the year, evaporation produces a nearly steady negative drawdown in sinusoidal pattern, and imposed water level stress drive a smooth long-term fall in head. The step-response curves on the right confirm that both rainfall and water level provoke a rapid rise that plateaus within a few hundred days, while evaporation yields a slower, steadily negative response. In each case Scenario 4 reaches its equilibrium response a little sooner than Scenario 3.



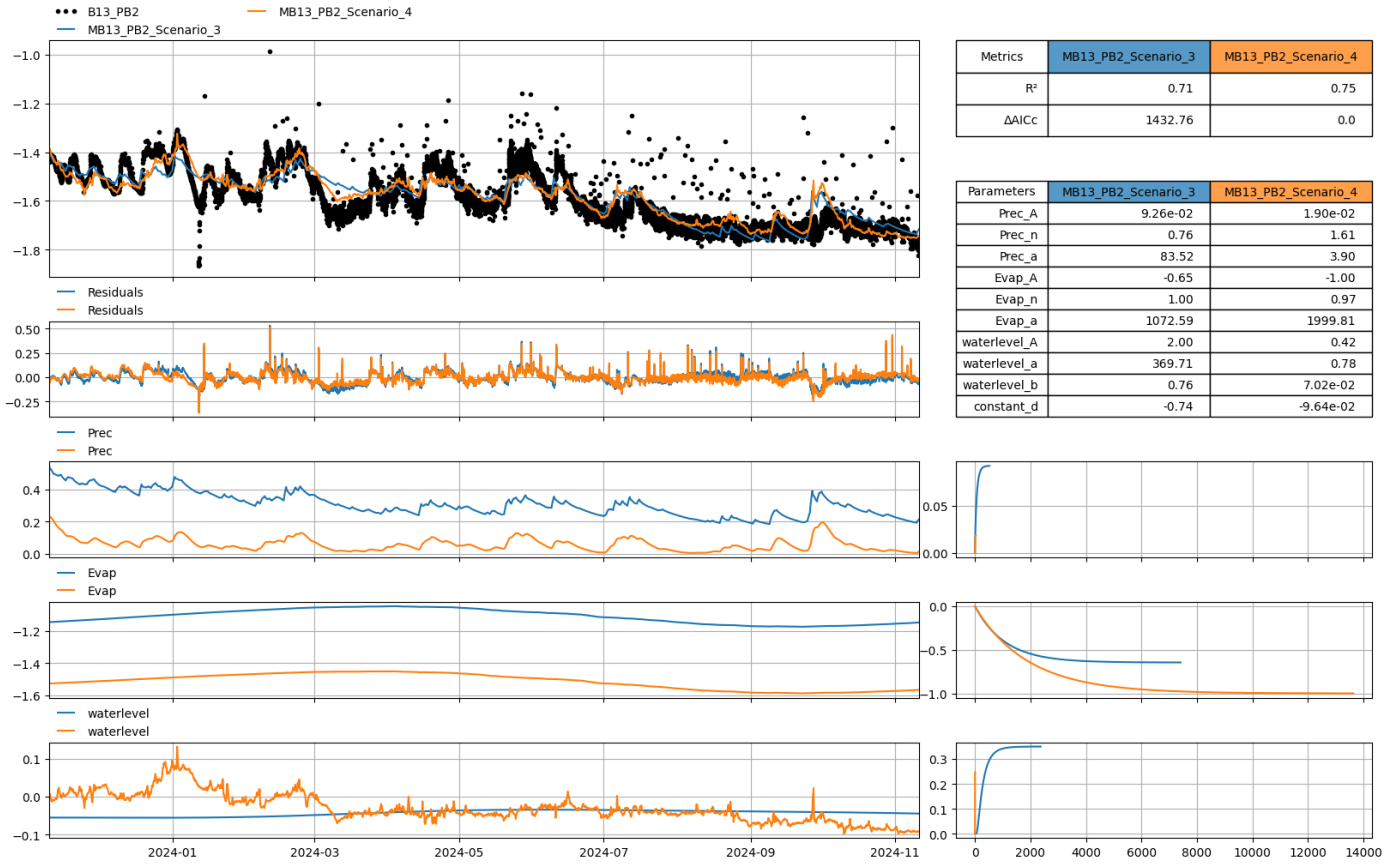
Metrics	MB26_PB1_Scenario_3	MB26_PB1_Scenario_4
R^2	0.76	0.76
ΔAICc	0.0	63.1

Parameters	MB26_PB1_Scenario_3	MB26_PB1_Scenario_4
Prec_A	0.11	0.12
Prec_n	0.78	0.77
Prec_a	109.21	115.82
Evap_A	-0.72	-0.73
Evap_n	1.00	1.00
Evap_a	1588.31	1579.02
waterlevel_A	2.00	2.00
waterlevel_a	533.55	242.64
waterlevel_b	0.51	1.22
constant_d	-0.66	-0.69

Figure C.8 Mosaic plot result for MB026-PB1 observation well (calibration stage) between Scenario 3 and 4. Coloured lines correspond to model-predicted measurement and black markers for observations.

Figure C.8 shows the hindcast results for well MB26_PB1 under Scenario 3 (blue) and Scenario 4 (orange). In the top-left panel, both scenarios trace the broad seasonal decline and most short-term fluctuations in the black-dot observations from January to November 2024, although neither reproduces every sharp peak. The residuals directly below indicate that most errors lie within ±0.2 m, with no systematic bias. Interestingly, there are some rapid groundwater movements at the beginning of calibration period, which might be explained by effect of surface runoff or infiltration in the groundwater well. The metrics table at upper right reports identical R^2 values of 0.76 for both scenarios, but Scenario 3 achieves a lower corrected AIC ($\Delta AICc$ = 0.0 versus 63.1), indicating it offers the better balance between fit and complexity. The parameter table shows only minor differences: Scenario 4 has a marginally larger precipitation amplitude (Prec_a = 115.82 versus 109.21) and a slightly stronger precipitation coefficient (Prec_A = 0.12 versus 0.11), while the water level shape factor of a and b is higher in Scenario 4, causing more stretched out pattern and higher scale factor.

On the left, the stress-contribution plots reveal that rainfall induces modest positive pulses, evaporation produces a steady negative drawdown, and water level impact shows smooth long-term fall. Their step-response curves on the right show a rapid initial rise from rainfall that plateaus within a few hundred days, a slowly developing negative response from evaporation that stabilizes over thousands of days, and a fast decline from water level. Scenario 3's step responses are slightly stronger and slower to stabilize than Scenario 4's, consistent with its higher boundary amplitude and lower precipitation scale.



Metrics	MB13_PB2_Scenario_3	MB13_PB2_Scenario_4
R^2	0.71	0.75
ΔAICc	1432.76	0.0

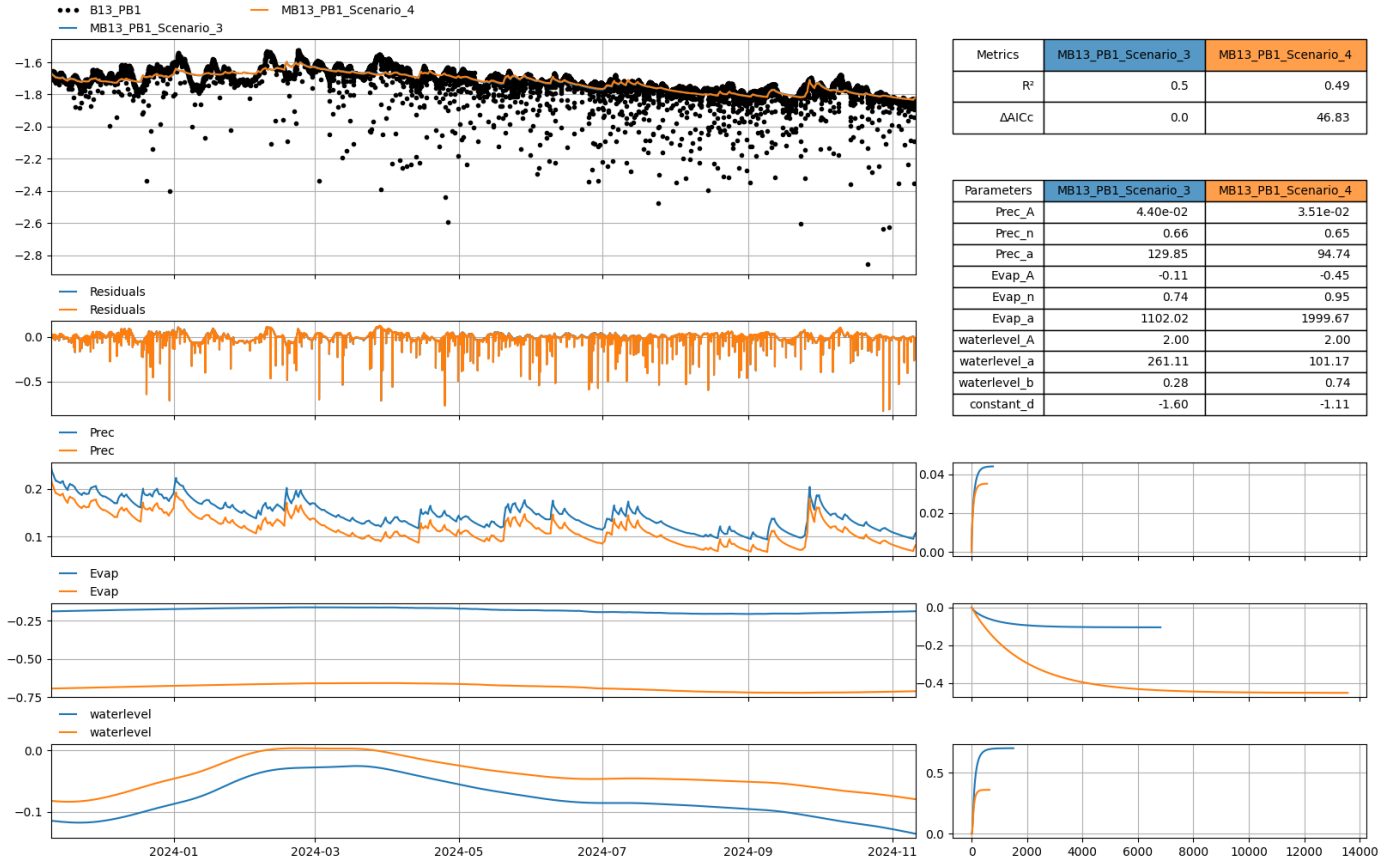
Parameters	MB13_PB2_Scenario_3	MB13_PB2_Scenario_4
Prec_A	9.26e-02	1.90e-02
Prec_n	0.76	1.61
Prec_a	83.52	3.90
Evap_A	-0.65	-1.00
Evap_n	1.00	0.97
Evap_a	1072.59	1999.81
waterlevel_A	2.00	0.42
waterlevel_a	369.71	0.78
waterlevel_b	0.76	7.02e-02
constant_d	-0.74	-9.64e-02

Figure C.9 Mosaic plot result for MB013-PB2 observation well (calibration stage) between Scenario 3 and 4. Coloured lines correspond to model-predicted measurement and black markers for observations.

Figure C.9 compares observed groundwater heads at well B13_PB2 (black dots) with Pastas model predictions under Scenario 3 (blue) and Scenario 4 (orange). In the top-left panel, both scenarios follow the general seasonal decline and many of the short-term fluctuations from January through November 2024, but neither captures all of the scattered peak events seen in the observations. The residual plot below shows most errors lie within ±0.2 m, with larger deviations coinciding with those sharp peaks. Based on the observed heads, there are some rapid groundwater movements, which might be explained by the location of groundwater well in sandy intermediate layer such that it respons faster to external recharge or pumping factor.

The metrics table at upper right reports an R^2 of 0.71 for Scenario 3 and 0.75 for Scenario 4, while the corrected AIC ($\Delta AICc$) is much lower for Scenario 4 (0.0 versus 1432.76), indicating that Scenario 4 achieves a better balance of fit and model parsimony. The parameter table shows that Scenario 3 has a higher precipitation amplitude (Prec_A ≈ 0.093) and scale (Prec_a ≈ 83.5) compared to Scenario 4 (Prec_A ≈ 0.019, Prec_a ≈ 3.9), so it responds more strongly to rainfall inputs. By contrast, Scenario 4 relies less on precipitation but more on evaporation (Evap_A = –1.00 versus –0.65) and has a much smaller boundary-stage amplitude (waterlevel_a ≈ 0.42 versus 369.7).

On the left side, the stress-contribution plots show that rainfall spikes raise heads more under Scenario 3, whereas evaporation draws down heads more sharply under Scenario 4. The boundary input produces a smooth seasonal signal only in Scenario 3. The step-response curves on the right confirm this behavior: Scenario 3 shows a rapid, high-magnitude rise from rainfall before plateauing, and a slow negative response from evaporation that stabilizes over several thousand days. Scenario 4's rainfall response is almost negligible, its evaporation response reaches its steady negative drawdown more quickly, and its stage response is minimal. These differences explain why Scenario 4 fits the scattered observations better despite its simpler stress weighting.



Metrics	MB13_PB1_Scenario_3	MB13_PB1_Scenario_4
R^2	0.5	0.49
ΔAICc	0.0	46.83

Parameters	MB13_PB1_Scenario_3	MB13_PB1_Scenario_4
Prec_A	4.40e-02	3.51e-02
Prec_n	0.66	0.65
Prec_a	129.85	94.74
Evap_A	-0.11	-0.45
Evap_n	0.74	0.95
Evap_a	1102.02	1999.67
waterlevel_A	2.00	2.00
waterlevel_a	261.11	101.17
waterlevel_b	0.28	0.74
constant_d	-1.60	-1.11

Figure C.10 Mosaic plot result for MB013-PB1 observation well (calibration stage) between Scenario 1 and 2. Coloured lines correspond to model-predicted measurement and black markers for observations.

Figure C.10 shows groundwater heads at B13_PB1 (black dots) alongside Pastas model outputs for Scenario 3 (blue) and Scenario 4 (orange). In the main panel both scenarios follow the gradual long-term decline from January through November 2024, but they fail to reproduce many of the large downward spikes in the observations. The residual plot beneath confirms that most errors lie below 0.2 m, with the largest under-predictions occurring at those sharp drops. Similar to the previous well, the groundwater head observations also exhibit erratic and rapid drops, indicating sudden fluctuations that are difficult for the model to capture accurately.

The table at upper right reports R^2 values of 0.50 for Scenario 3 and 0.49 for Scenario 4, while Scenario 3 has the lower corrected AIC (ΔAICc = 0.0 versus 46.83), indicating a slightly better balance of fit and simplicity. Parameter values show that Scenario 3 responds more strongly to precipitation (Prec_A = 0.044 versus 0.035; Prec_a = 129.9 versus 94.7) and more weakly to evaporation (Evap_A = –0.11 versus –0.45). Scenario 4, by contrast, has a higher evaporation shape factor (Evap_n = 0.95 versus 0.74) and a larger boundary-stage shape (waterlevel_b = 0.74 versus 0.28).

The stress-contribution plots on the left reveal that rainfall produces a modest positive pulse, more pronounced in Scenario 3, while evaporation delivers a steady negative drawdown that is stronger under Scenario 4. The boundary condition input shows a smooth seasonal signal for both models. Finally, the step-response curves on the right confirm that rainfall and water level trigger a rapid initial change followed by quick stabilization, while evaporation causes a slow negative response that only levels off after several thousand days. Scenario 4's responses reach equilibrium slightly faster, reflecting its higher evaporation weighting and lower precipitation scale.

D. Comparison of Parameters after Parameter Fixing

Scenario 1

Observation Well	Model Parameter	Initial Value	Min. Value	Max. Value	Adjusted Value
MB014_PB1	Prec_A	-	-	-	3.48e-02
	Prec_n	-	-	-	1.03
	Prec_a	-	-	-	135.44
	Evap_A	-0.1	-1	-0.1	-0.52
	Evap_n	0.1	0.1	2	2.00
	Evap_a	5	5	365	365.00
	waterlevel_A	-	-	-	0.61
	waterlevel_a	-	-	-	9.68
	waterlevel_b	-	-	-	8.77e-04
	constant_d	-	-	-	-1.38
MB014_PB2	Prec_A	-	-	-	0.22
	Prec_n	-	-	-	1.30
	Prec_a	-	-	-	86.59
	Evap_A	-0.1	-1	-0.1	-1.00
	Evap_n	0.1	0.1	2	2.00
	Evap_a	5	5	365	238.49
	waterlevel_A	-	-	-	2.00
	waterlevel_a	-	-	-	782.87
	waterlevel_b	-	-	-	6.65e-02
	constant_d	-	-	-	-7.27e-02
HB027_PB1	Prec_A	-	-	-	0.35
	Prec_n	-	-	-	1.89
	Prec_a	-	-	-	43.27
	Evap_A	-0.1	-1	-0.1	-0.10
	Evap_n	0.1	0.1	2	1.41
	Evap_a	5	5	365	62.81
	waterlevel_A	-	-	-	2.00
	waterlevel_a	-	-	-	187.95
	waterlevel_b	-	-	-	0.47
	constant_d	-	-	-	-1.55
MB015_PB2	Prec_A	-	-	-	0.28
	Prec_n	-	-	-	1.22
	Prec_a	-	-	-	101.20
	Evap_A	-0.1	-1	-0.1	-0.17
	Evap_n	0.1	0.1	2	1.60
	Evap_a	5	5	365	125.87
	waterlevel_A	-	-	-	2.00
	waterlevel_a	-	-	-	200.74
	waterlevel_b	-	-	-	0.54
	constant_d	-	-	-	-2.62

Scenario 2

Observation Well	Model Parameter	Initial Value	Min. Value	Max. Value	Adjusted Value
MB014_PB1	Prec_A	-	-	-	4.48e-02
	Prec_n	-	-	-	1.66
	Prec_a	-	-	-	106.05
	Evap_A	-0.1	-1	-0.1	-0.52
	Evap_n	0.1	0.1	2	2.00
	Evap_a	5	5	365	365.00
	waterlevel_A	-	-	-	0.53
	waterlevel_a	-	-	-	19.79
	waterlevel_b	-	-	-	9.51e-04

	constant_d	-	-	-	-1.38
MB014_PB2	Prec_A	-	-	-	0.27
	Prec_n	-	-	-	1.80
	Prec_a	-	-	-	75.54
	Evap_A	-0.1	-1	-0.1	-1.00
	Evap_n	0.1	0.1	2	2.00
	Evap_a	5	5	365	238.07
	waterlevel_A	-	-	-	0.10
	waterlevel_a	-	-	-	10.00
	waterlevel_b	-	-	-	1.11e-06
	constant_d	-	-	-	-0.34
HB027_PB1	Prec_A	-	-	-	0.36
	Prec_n	-	-	-	1.75
	Prec_a	-	-	-	52.41
	Evap_A	-0.1	-1	-0.1	-0.10
	Evap_n	0.1	0.1	2	1.40
	Evap_a	5	5	365	63.80
	waterlevel_A	-	-	-	0.22
	waterlevel_a	-	-	-	10.06
	waterlevel_b	-	-	-	1.13e-22
	constant_d	-	-	-	-1.60
MB015_PB2	Prec_A	-	-	-	0.32
	Prec_n	-	-	-	1.33
	Prec_a	-	-	-	105.04
	Evap_A	-0.1	-1	-0.1	-0.23
	Evap_n	0.1	0.1	2	1.49
	Evap_a	5	5	365	169.63
	waterlevel_A	-	-	-	0.10
	waterlevel_a	-	-	-	10.00
	waterlevel_b	-	-	-	1.00e-02
	constant_d	-	-	-	-2.80

Scenario 3

Observation Well	Model Parameter	Initial Value	Min. Value	Max. Value	Adjusted Value
MB012_PB1	Prec_A	-	-	-	9.51e-02
	Prec_n	-	-	-	0.78
	Prec_a	-	-	-	25.16
	Evap_A	-0.1	-10	-0.1	-0.53
	Evap_n	0.1	0.1	1	1.00
	Evap_a	5	5	365	365.00
	waterlevel_A	-	-	-	2.00
	waterlevel_a	-	-	-	1000.00
	waterlevel_b	-	-	-	2.96e-02
	constant_d	-	-	-	5.02
HB025_PB1	Prec_A	-	-	-	7.51e-02
	Prec_n	-	-	-	0.67
	Prec_a	-	-	-	41.01
	Evap_A	-0.1	-10	-0.1	-0.47
	Evap_n	0.1	0.1	1	1.00
	Evap_a	5	5	365	365.00
	waterlevel_A	-	-	-	2.00
	waterlevel_a	-	-	-	880.69
	waterlevel_b	-	-	-	0.21
	constant_d	-	-	-	3.65
MB026_PB2	Prec_A	-	-	-	0.18
	Prec_n	-	-	-	0.85
	Prec_a	-	-	-	106.71

	Evap_A	-0.1	-10	-0.1	-0.13
	Evap_n	0.1	0.1	2	0.95
	Evap_a	5	5	365	137.29
	waterlevel_A	-	-	-	2.00
	waterlevel_a	-	-	-	78.94
	waterlevel_b	-	-	-	0.40
	constant_d	-	-	-	-0.66
MB026_PB1	Prec_A	-	-	-	0.12
	Prec_n	-	-	-	0.77
	Prec_a	-	-	-	116.17
	Evap_A	-0.1	-10	-0.1	-0.13
	Evap_n	0.1	0.1	2	1.00
	Evap_a	5	5	365	338.70
	waterlevel_A	-	-	-	2.00
	waterlevel_a	-	-	-	411.54
	waterlevel_b	-	-	-	0.66
	constant_d	-	-	-	-0.73
MB013_PB2	Prec_A	-	-	-	6.98e-02
	Prec_n	-	-	-	0.76
	Prec_a	-	-	-	60.09
	Evap_A	-0.1	-10	-0.1	-0.26
	Evap_n	0.1	0.1	2	1.00
	Evap_a	5	5	365	364.92
	waterlevel_A	-	-	-	2.00
	waterlevel_a	-	-	-	329.97
	waterlevel_b	-	-	-	0.96
	constant_d	-	-	-	-0.13
MB013_PB1	Prec_A	-	-	-	5.66e-03
	Prec_n	-	-	-	2.71
	Prec_a	-	-	-	0.73
	Evap_A	-0.1	-10	-0.1	-0.10
	Evap_n	0.1	0.1	2	0.99
	Evap_a	5	5	365	364.92
	waterlevel_A	-	-	-	2.00
	waterlevel_a	-	-	-	954.70
	waterlevel_b	-	-	-	3.41
	constant_d	-	-	-	-6.55

Scenario 4

Observation Well	Model Parameter	Initial Value	Min. Value	Max. Value	Adjusted Value
MB012_PB1	Prec_A	-	-	-	0.11
	Prec_n	-	-	-	0.78
	Prec_a	-	-	-	28.42
	Evap_A	-0.1	-10	-0.1	-0.11
	Evap_n	-	-	-	8.38
	Evap_a	5	5	365	11.03
	waterlevel_A	-	-	-	2.00
	waterlevel_a	-	-	-	293.76
	waterlevel_b	-	-	-	0.11
	constant_d	-	-	-	5.02
HB025_PB1	Prec_A	-	-	-	0.37
	Prec_n	-	-	-	0.71
	Prec_a	-	-	-	297.68
	Evap_A	-0.1	-10	-0.1	-0.16
	Evap_n	-	-	-	4.82
	Evap_a	5	5	365	32.83
	waterlevel_A	-	-	-	2.00

	waterlevel_a	-	-	-	626.61
	waterlevel_b	-	-	-	0.29
	constant_d	-	-	-	1.49
MB026_PB2	Prec_A	-	-	-	0.19
	Prec_n	-	-	-	0.88
	Prec_a	-	-	-	97.99
	Evap_A	-0.1	-10	-0.1	-0.04
	Evap_n	-	-	-	10.00
	Evap_a	5	5	365	7.82
	waterlevel_A	-	-	-	2.00
	waterlevel_a	-	-	-	70.34
	waterlevel_b	-	-	-	0.47
	constant_d	-	-	-	-1.18
MB026_PB1	Prec_A	-	-	-	0.12
	Prec_n	-	-	-	0.77
	Prec_a	-	-	-	116.45
	Evap_A	-0.1	-10	-0.1	-0.05
	Evap_n	-	-	-	3.64
	Evap_a	5	5	365	34.77
	waterlevel_A	-	-	-	2.00
	waterlevel_a	-	-	-	210.92
	waterlevel_b	-	-	-	1.41
	constant_d	-	-	-	-0.76
MB013_PB2	Prec_A	-	-	-	7.29e-02
	Prec_n	-	-	-	0.75
	Prec_a	-	-	-	66.96
	Evap_A	-0.1	-10	-0.1	-0.05
	Evap_n	-	-	-	8.02
	Evap_a	5	5	365	11.92
	waterlevel_A	-	-	-	2.00
	waterlevel_a	-	-	-	266.93
	waterlevel_b	-	-	-	1.22
	constant_d	-	-	-	-0.05
MB013_PB1	Prec_A	-	-	-	3.53e-02
	Prec_n	-	-	-	0.65
	Prec_a	-	-	-	98.48
	Evap_A	-0.1	-10	-0.1	-0.02
	Evap_n	-	-	-	1.93
	Evap_a	5	5	365	43.04
	waterlevel_A	-	-	-	2.00
	waterlevel_a	-	-	-	98.00
	waterlevel_b	-	-	-	0.76
	constant_d	-	-	-	-1.08