UNIVERSITY OF COPENHAGEN
INSTITUTE OF GEOSCIENCE AND NATURAL RESSOURCE MANAGEMENT

Hydrological analysis of present and future hydrology in Ahlergaarde Subcatchment

Picture of Skjern Å, source: (Larsen, no date)

Elfrida Marie Jacobsen, cqv134
Lucy Dunshea, fzg264
Mette Iversen, vmg876

Submitted: 14/01/2025

Index

1. Introduction

This hydrological analysis is based on a 133-year dataset for the Ahlergaarde Subcatchment (Fg. 1), covering 1050 km^2 (Jensen and Refsgaard, 2018). Situated in western Denmark, this subcatchment provides a context for understanding historical, present, and projected hydrodynamics. Key hydrological variables are investigated to identify trends, variability, and potential future impacts on water resource management. Daily data on temperature (T) and precipitation (P), available from 1875, and discharge (Q) data, recorded since 1922, form the basis of this study. The dataset includes information on potential evapotranspiration (ET), enabling the assessment of patterns and trends in historical, present, and future hydrological regimes and climate conditions.

Various statistical methods and hydrological modeling approaches are employed, such as homogeneity testing, trend analysis, flow duration curves, and an extreme value analysis. The NAM rainfall-runoff model is also applied to simulate historical processes and project future changes under different climate scenarios derived through the delta change (DC) and distribution-based scaling (DBS) methods. Together, these analyses aim to assess temporal changes in the hydrological behavior in the Ahlergaarde Subcatchment, contributing to adaptive water management and climate resilience planning.

1.1 Study site

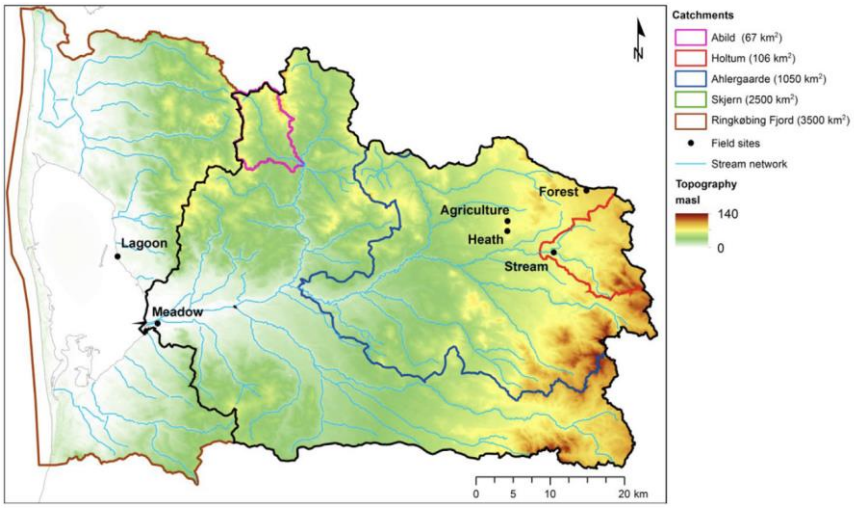


Figure 1: Map of the study area (Jensen and Refsgaard, 2018)

The Ahlergaarde Subcatchment (Fg. 1) is connected to the Skjern Catchment in western Denmark. Skjern River has the largest mean annual Q in Denmark and connects to Ringkøbing Fjord, which is subsequently connected to the North Sea (Karlsson *et al.*, 2014; Jensen and Refsgaard, 2018). The

area is relatively flat, and the sediment type and geology dominating the area is partly due to the maximum extension of the Weichel Ice Age and later Saale Ice Age (Karlsson *et al.*, 2014; Jensen and Refsgaard, 2018). The eastern part of the catchment is highly affected by moraine till, resulting in clayey sand. The topsoil is, hence, highly permeable, resulting in a domination of groundwater inflow regarding stream flow (Jensen and Refsgaard, 2018). In contrast, the south and central part of the catchment is partly dominated by washout sand and gravel from the advance of the Weichel. The area is mainly utilized for agriculture (60,5%), with additional land use including grassland, forest, heath (37,7%), and urban area (1,8%) (Karlsson *et al.*, 2014).

2. Theory and methods

2.1 Statistical methods

The following section covers the methods used to investigate the hydrological system at the Ahlergaarde Catchment, including homogeneity tests and correction, trend tests, autocorrelation, flow duration curves, and extreme value analysis. The following time periods are used throughout this report: historical period (years 1875-1904), reference period (1971-2000), and the far future (2071-2100).

2.2 Homogeneity test

Inhomogeneity in a hydrological data series collected in a catchment can occur due to a myriad of reasons, including changes in instrumentation, exposure, station environment, location, measurement techniques, and many others (Searcy and Hardison, 1960). Inhomogeneity is especially profound in long climate data periods, where there is a greater chance that there has been either a change in the physical conditions of the basin or in the data acquisition methods. Using inhomogeneous data to perform hydrological modeling can significantly impact the model's accuracy and result in inaccurate projections of the future conditions of a catchment. Therefore, it is essential to assess and correct the data using a homogeneity test (Karlsson *et al.*, 2014).

Homogeneity tests compare the accumulation of one measurement against the accumulation of a reference measurement of the same variable (Searcy and Hardison, 1960). In the case of hydrology, this theory is relevant to recorded variables such as P and Q. In this case study, a homogeneity test has been performed using the double mass curve method, comparing homogeneous reference datasets of P and Q to the original recorded data for each variable. A double mass curve was produced for P

between 1876 and 2007 and Q between 1922 and 2007. Both variables show inhomogeneity due to breaks occurring in the dataset. A break in the slope for P was identified in 1959 and for Q in 1964.

The slope is adjusted using a correction factor, which is calculated using the ratio of the trendline slope before and after the break. For P and Q, the correction factor was applied to data before the break. Adjusting the data with a low correction factor is preferred over using a higher one, as over-adjustment can distort the underlying trends in the data (Searcy and Hardison, 1960). The homogeneous datasets produced by applying this method were carried forward into the subsequent steps of this assessment.

2.3 Trend test

A trend test can determine if the changes observed in a given variable over the study period are statistically significant or could just be considered noise. Trend tests are used in statistics to identify if there is an association or correlation between two or more variables (Meals, Dressing and Harcum, 2011). The tests can be parametric or non-parametric. Parametric tests rely on specific assumptions about the underlying distribution of the variable being examined, i.e., that T follows a Gaussian (normal) distribution. Non-parametric tests do not rely on these specific assumptions and are sometimes referred to as distribution-free methods (Totaro, Gioia and Iacobellis, 2019). Non-parametric tests are often used in climate assessments, where it is difficult to control the variables being tested, and it cannot be assumed that they follow a particular distribution. Non-parametric tests also require a smaller sample size to reject the null hypothesis, which can help model complex environmental systems. Trends in datasets can either be monotonic, which is a gradual change over time in a consistent direction, or a step change, which is an abrupt shift at a certain point in time (Meals, Dressing and Harcum, 2011).

The Mann-Kendall test was used to assess the changes to P, Q, T, and ET in this case study. The Mann-Kendall test uses time series data, and it assumes that the data is independent and homogenous

$$S = \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} sign(x_j - x_i) \qquad (1)$$

(Karlsson *et al.*, 2014). The test determines whether there is a monotonic trend in the data by comparing every data point to every data point preceding it in the time series, giving a total of n(n-1)/2 pairs of data points. Mann-Kendall tests can be one-sided or two-sided and identify if the variable is consistently increasing or decreasing over time, although this does not have to be a linear process (Birkens, n.d.). The statistic for the Mann-Kendall test is calculated using the equation below:

The variance of the statistic S is given by:

For this assessment, a two-sided Mann-Kendall test was applied to both annual and monthly climate datasets. This gives an understanding of how the climate variables are changing over an extended

$$var = \frac{1}{18}\left[n(n-1)(2n+5) - \sum_t f_t(f_t-1)(2f_t+5)\right] \quad (2)$$

period and within the year. Mann-Kendall tests with a yearly timestep tested annual sums of P, annual average T, Q, and ET, daily maximum values of P, T, and Q, and finally monthly maximum values of ET. The monthly Mann-Kendall tests were applied to the annual sum of P by month and monthly averages of T and Q. The significance of the test statistic calculated by the Mann-Kendall test is determined by the equation below, where *se* is the standard error:

$$z = \begin{cases} (S-1)/se, & S > 0 \\ 0, & S = 0 \\ (S+1)/se, & S < 0 \end{cases} \quad (3)$$

A positive z indicates a positive monotonic trend in the data, and a negative z suggests a negative monotonic trend. A z score further from zero indicates a more statistically significant trend. The statistical significance level for the trend test is α=0.1 - 0.001, respectively indicating a strong and low trend. If the significance level is greater than 0.1, there is no statistical significance and the null hypothesis of no trend cannot be rejected (Salmi, 2002).

2.4 Autocorrelation

Autocorrelation is used to describe the relation of a time series with itself. The goal of the autocorrelation for time series of stochastic processes is to understand at which time scale the measurements of P, Q, and T are correlated. Periodic variability and short-term trends can result in autocorrelation in these time series, however, differences in the processes which drive these weather patterns mean that each variable is likely to have a unique relationship to itself (Sepp and Jaagus, 2002). The equation used to calculate autocorrelation for P, Q, and T is below:

$$\rho_{zz,k} = \frac{COV[Z_t, Z_{t+k}]}{COV[Z_t, Z_t]} = \frac{E[\{Z_t - \mu_z\}\{Z_{t+k} - \mu_z\}]}{\sigma_z^2} \quad k = -\infty,\ldots,-1,0,1,\ldots,\infty \quad (4)$$

The autocorrelation can be between -1 and 1, with values close to 1/-1 meaning a perfect correlation, and a value of 0 meaning an absence of correlation (Birkens, n.d). The autocorrelation decreases over

time as the measurement of the previous day affects the new measurement less. The time lag therefore represents the correlation between these measurements over time (Birken, n.d).

2.5 Flow duration

The flow duration curve (FDC) is used to understand the flow rate of a specific catchment. It is a graphical representation of the percentage of time that a certain flow level is exceeded in a catchment, as taken from a specific measurement point. The use of FDCs is broad, ranging from estimating the occurrence of floods and droughts to informing the location of hydropower stations (Westerberg *et al.*, 2014). Estimations of the high and low-flow values in a specific catchment can be derived from FDCs. The Q values with certain exceedance percentages (i.e. 90% for extreme low flows, 10% for extreme high flows) are often used as threshold values when calibrating hydrological models (Birkens, n.d.). For the Ahlergaarde Catchment, the FDC is used to inform the value used in the NAM model to represent peak flows. The FDC for the Ahlergaarde Catchment was calculated using all the available continuous Q data from 1942-2007.

Some caution must be taken when selecting the range of data used in an FDC. Since an FDC does not display data as a time series, the shape of the curve is heavily determined by the time range it represents (Searcy and Hardison, 1960). Due to both seasonality in climate patterns and more long-term variance due to climate change, the curve is also influenced by Δt. Choosing the range of data used depends on the goal of the modeling, for example, estimation of the average flow of the Skjern River during winter is likely to be more accurate using an FDC with data from the winter months.

Variance is used to compare two different FDCs, which could come from a different part of the catchment, using data over a different period, or taken from another catchment altogether. High variance is characteristic of smaller catchments, and low variance is common for larger catchments with rivers with a high base flow (Searcy, 1959). FDCs also cannot be used to estimate the frequency of peak or low flows, they are only an indication of the magnitude of possible flow levels at a certain point of the catchment. Using historic data to construct FDCs also means that they cannot be relied on to predict future catchment conditions, given the changing patterns of climatic variables caused by increasing atmospheric average Ts (Jensen and Refsgaard, 2018).

2.6 Extreme value analysis

Extreme value analysis in hydrology is a method of exploring the extreme events that can occur within a catchment. The process evaluates the probability of an event of a specific magnitude occurring,

which is critical for determining maximum flood levels to be accounted for when surveying catchments (Jensen, n.d). Floods profoundly affect both anthropogenic and ecological functions of a catchment; therefore, understanding these events is essential to mitigate their adverse impacts (Suchara, 2019). Extreme value analyses can be conducted on data in an annual maximum series (AMS) or partial duration series (PDS). An AMS uses the highest value recorded within a year, and a PDS consists of values exceeding a certain threshold, which is determined subjectively. For this analysis, an AMS was taken from the reference period and calculated future projections.

When using an AMS in an extreme value analysis of Q, P, and T, it is assumed that the maximum values are independent, there is no temporal trend (i.e., stationarity is maintained), and that the maximum values in the series follow the same probability distribution (Madsen, 2000). A cumulative frequency distribution is calculated using the exceedance probability, representing the likelihood of values being equaled or exceeded. This distribution determines if a value from the AMS is higher or lower than a given threshold. The inverse of the exceedance probability gives the return period of a specific event. This would mean that when using annual data series, an event with a 1 percent exceedance probability corresponds to a return period of 100 years (Madsen, 2000; Jensen, n.d).

The distribution of minimum and maximum values can be explained by a Gumbel distribution, which is slightly skewed to better account for extreme outliers (Jensen, n.d). The equation for the Gumbel distribution is shown below:

$$F_X(x) = e^{-e^{-\alpha(x-\beta)}} \quad (5)$$

The mean and variance of the Gumbel distribution determine the value of an extreme event given its return period, while equations for the frequency factor (Kt) and standard deviation (St) quantify the uncertainty of this extreme value. As the return period of a projected event increases, the uncertainty of the projection also tends to increase (Jensen, n.d). Communicating the uncertainty associated with extreme value predictions to ensure effective decision-making around risk.

AMS must be evaluated with care when evaluating data susceptible to autocorrelation. For example, maximum values recorded on the 31st of December and the 1st of January can be linked to the same flood event. However, these observations will appear independent from each other in a time series, even though they may be correlated events. Understanding the autocorrelation of each variable can reduce the chances of misinterpreting these extreme values.

3. Hydrological modeling

A hydrological model aims to simulate and predict changes in the water flux, flow, or water storages in the terrestrial components of the hydrological cycle. These models are utilized to predict future trends, extreme flow events, and their return times. They are useful for assessing and predicting the hydrological effects of climate change, enabling qualified estimations for mitigation strategies for flooding or changes in sediment beds affecting bridge constructions (Karlsson *et al.*, 2016).

Modeling of the runoff processes within a catchment can be used to quantify water balance components. By doing so, predictions of extreme flow events, flow intensities, and how the P is being divided into different hydrological processes, can be interpreted. The most important components for modeling are Precipitation (P), Evapotranspiration (ET), surface runoff, discharge (Q), infiltration, unsaturated flow (upper soil layer with partly air-filled pores), and groundwater flow (lower soil layer with almost no air content). The relative contributions depend on the characteristics of the specific catchment. Therefore, changes in topography, soil porosity, and moisture conditions are important considerations in case of breaks or abrupt changes in the historical data ('NAM model documentation', 1999).

Because all the parameters are mutually connected, changes in one parameter are expected to alter other components. This relation between parameters is investigated through manual- and auto-calibration to understand the complexity of modeling precise estimations for each parameter contribution (Madsen, 2000; Kumar, Lohani and Nema, 2019).

3.1 NAM Model

To simplify this complex interconnected hydrological cycle, the conceptual model, NAM, is applied, to model the rainfall-runoff processes quantitatively ('NAM model documentation', 1999). As mentioned, NAM focuses on the near- and underground water flows and fluxes, leaving out the circulation to and from the atmosphere. It accounts for the water circulation with four different sinks: snow storage, surface storage, root zone storage, and groundwater storage. The flows and fluxes drain or recharge each sink, as demonstrated in fg. 2 (Kumar, Lohani and Nema, 2019).

Models are classified based on their complexity and accuracy. The NAM model is categorized as a lumped model where each input parameter is represented by a single value, accounting for the entire catchment area, thus being one of the simplest models. By representing a catchment by one value, thus establishing uncertainty in the model. Lumped models implement flows calculated from empirical relations, meaning the water flow is estimated based on mathematical and theoretical relationships derived from previous experiments, rather than exact measurements. Because they are empirical, the model parameters are changeable to find the best water balance and statistical fit, compared to historical data. This study uses data from the reference period to calibrate and validate the NAM model.

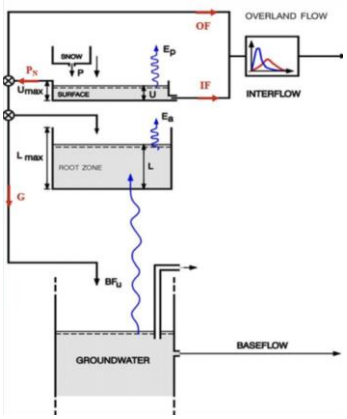


Figure 2: ***Graphical representation of the NAM model, showing the connection and modeled movement of water between each storage sink (Kumar et al., 2019)***

Lumped conceptual models focus on simulating the stream Q as a function of time. The simplicity of this type of model means few parameters and initial data are needed to produce a model output. This means that even with very few parameters, the model is relatively robust. In this specific model, model parameters, initial conditions, meteorological data (P, potential ET, and T) and stream flow data are input requirements ('NAM model documentation', 1999).

3.2 Calibration and Validation Principles

In hydrological modeling, model calibration and validation can ensure more robust and reliable predictions of the future stream Q of a catchment. For the NAM model, the objective of calibration is to optimize the model parameters LMAX, UMAX, CQOF, CKIF, TIF, TOF, TG, CK1, and CKBF so that the modeled hydrological processes represent the actual behavior as precisely as possible. By doing the calibration, we ensure that the model behavior aligns with the hydrological processes it seeks to represent and reduces the uncertainties following parameter estimation ('NAM model documentation', 1999; Karlsson *et al.*, 2014).

It is not possible to collect exhaustive data on the features and processes in a catchment, so instead the parameter values are determined by mathematically minimizing the discrepancy between the simulated and the observed historical Q. To achieve this, we aim to use the fewest possible parameters to accurately simulate the predicted Q. Using few parameters also reduces the issue of overfitting the

model. This can be executed with different techniques: Trial-and-Error manual calibration, Intelligent manual calibration, and autocalibration ('NAM model documentation', 1999).

Firstly, an iterative process of trial-and-error calibration gives an insight into how the model reacts to changes in parameters, informing its general behavior. Secondly, to narrow down which parameters display the greatest changes. A simple sensitivity analysis is conducted to evaluate the influence of each parameter. By adding and subtracting a fixed percentage to all parameter values and calculating the Δ root mean square (RMS) between them, the sensitivity becomes apparent for each parameter. A high ΔRMS indicates high sensitivity and conversely for a low ΔRMS.

Based on this analysis, the five most sensitive parameters making the greatest impact on the simulated Q are identified and can be used as a foundation for further trial-and-error calibration. For this study, adjustments of the parameters Umax, CQOF, TIF, TOF, and TG had the greatest impact on Q. These parameters are then tweaked using auto-calibration techniques to further refine the performance of the model (Karlsson et al., 2014).

3.3 Auto-calibration

Auto-calibration is a mathematical and semi-complex parameter optimization, based on complex algorithms enabling automated exploration of a larger range of parameter combinations. It is less time-consuming than manual calibration but requires greater computational strength. It is a form of inverse modeling where known outcomes (observed data) are used to infer the optimal parameter values, which were originally unknown (Karlsson *et al.*, 2014). The algorithm automates this iterative adjustment process. Autocalibration can be optimized on single or multiple objectives depending on which objective is of importance for the application of the model. Efficient multi-objective functions are distributed along a Pareto front, which can help to focus the decision within an efficient set by identifying parameter combinations that optimize competing goals, such as minimizing errors in peak flow predictions while maintaining accuracy in low-flow conditions. For the NAM model, optimization is generally done for (1) overall water balance, (2) overall shape of the hydrograph, (3) peak flows, and (4) low flows (Madsen 2000). Auto-calibration improves model performance but typically cannot produce a model with zero error (Birkens, n.d).

3.4 Performance Criteria

Performance criteria represent the statistical evaluation of how well the simulation captures the observed streamflow patterns. In this study, three performance metrics are used: Mean error (ME),

Root Mean Squared Error (RMS), and the Nash-Sutcliffe coefficient (R^2). Firstly, calibration is done for individual optimization, and secondly as a multiple objective function. The calibration aim is to achieve ME and RMS values as close to zero as possible to minimize errors and provide an assessment of the model performance. The ME calculation focus is detecting the average error in the model prediction, indicating whether the simulation over- or underestimates the streamflow in general (positive or negative value respectively). RMS on the other hand, evaluates the magnitude of errors, with extra sensitivity towards larger error deviations, highlighting the fit of peak flows. Finally, the R^2 evaluates the overall efficiency of the model, with values close to 1 when the observed and simulated data have a strong alignment, and a value of zero indicates that the model does not explain any of the variability in the simulated data around its mean. Together the three metrics give statistical insights to the accuracy and reliability of the streamflow simulation. These are used in the iterative process to evaluate whether the parameter composition performs better than previous attempts (Madsen, 2000; Karlsson *et al.*, 2014).

3.5 Calibration issues

Even though calibration is performed to accurately represent the physical processes of the catchment, several problems and uncertainties arise during the calibration process. The observation data brings uncertainties to the model, as P, ET, and Q measurements can be biased for several reasons, including incomplete datasets, non-stationary observation points, and limited observation points. However, for this lumped conceptual model, the number of stations is not highly important, as the input values are uniform for the entire catchment ('NAM model documentation', 1999). Despite this, the geographical location of the stations where data is recorded is still necessary to consider to contextualize the model.

The project aims to optimize the model parameters to best fit the observed data. However, every model simplifies reality, meaning the model will never fit perfectly with what is observed. Real-world complexity is impossible to describe with a model, especially in a simple lumped model. This aligns with the concept of "equifinality" which suggests that many different parameter combinations can yield acceptable solutions, indicating that a unique optimal set of parameters is highly unlikely (Madsen, 2000). Hence, it is difficult to optimize for all three metrics at once, as trade-offs between different objectives are inherent in the calibration process.

3.6 Validation

Because the model structure is expected to have errors, the process of validation is important to incorporate when modeling hydrology (Seaby *et al.*, 2013). Validation aims to determine whether the model can accurately simulate the physical behavior of the environment it represents, using data that was not involved in the calibration. This is done to ensure that the model not only fits the calibration data but applies to the catchment in the future. If the model cannot be generalized for other similar uses or datasets, the issue of uniqueness occurs. Another reason to validate a model is to check for non-identifiability, as it is important to be able to identify if the chosen value combination is unique. If it is non-identifiable, the same result is obtainable from several value combinations (Madsen, 2000). We therefore aim to find a trade-off between accuracy and generalization potential by adjusting and identifying the most important theoretical parameters we know affect catchment hydrology. To further minimize the risks of uniqueness and non-identifiability the model could be validated through a split sample test which splits a dataset into two sub-samples. This allows the model to be calibrated on one data set while saving a set of observations that can be compared to the NAM model to evaluate its performance after calibration.

As the temporal change of climate data is relevant, the reference period sample was split at a specific date, as opposed to samples being randomly allocated to calibration and validation sets (Van Roosmalen, Christensen and Sonnenborg, 2007). For the split sample test method to be relevant, the data in the validation series must represent similar conditions in the dataset. Hence, observations later than 2000, which are more likely to be influenced by climate change were omitted from the validation set. The calibration set must be long enough to ensure that natural variability in the observations, especially for P, is not mistaken for a trend (Motavita *et al.*, 2019). Thus, the model was calibrated on data ranging from 1970-1990 and validated using data from 1991-2000.

4. Climate modeling

Climate change is expected to impact the hydrological system, resulting in changes to T, P, and ET (Van Roosmalen, Christensen and Sonnenborg, 2007). The change in the parameters and the effect on Q is of great interest since due to its impact on environmental, agricultural, and economic systems (Thodsen, 2007). These changes are expected to occur globally and on a regional scale, thus modeling future change is of great importance (Van Roosmalen, Christensen and Sonnenborg, 2007). In Denmark, the hydrological conditions are expected to cause more flooding and faster groundwater

recharge in the winter months, decreased water tables, and depletion of low flows in the summer months due to changes in timing and delivery of P (Seaby *et al.*, 2013).

Global circulation models (GCM) have been used to generate projections of climate change based on different emission scenarios with described greenhouse gas and aerosol concentrations (Van Roosmalen, Christensen and Sonnenborg, 2007). The GCM is a numerical model, representing ocean-atmosphere circulation. Its coarse resolution means that downscaling is necessary to capture local climate change dynamics. One method to downscale a GCM is by combining it with a regional climate model (RCM). Lateral boundary conditions from the GCM over a limited area are used in the RCM model. The RCM is better at capturing the climate of smaller regions, such as country level, thus implementing the RCM results in higher resolution outputs (Van Roosmalen, Christensen and Sonnenborg, 2007; Seaby *et al.*, 2013; Pasten-Zapata, Sonnenborg and Refsgaard, 2019). P is a climatic parameter which especially requires further downscaling before using it in hydrological models. This is due to P having high spatial and temporal variability, as P events can occur over very limited areas and a short time scale (Madsen, 2000; Seaby *et al.*, 2013).

The climate model coupling of the GCM-RCM used in the report is the ARPEGE-RM5.1 is known to simulate dry conditions and is one of many combinations from the EU ENSEMBLES project. The project is an initiative to obtain better climate projections in Europe (Seaby *et al.*, 2013). The climate projection is based on the IPCC AR4 SRES A1B (Karlsson *et al.*, 2016). The A1 scenario describes a future world with rapid economic growth, peak population in mid-century, and a fast introduction to new and more efficient technologies. The A1 scenario is further distinguished into different groups according to the direction of the energy system. A1B describes a system in balance, meaning the energy system does not rely heavily on one particular energy source (IPCC, n.d). Since RCM outputs are still subject to systematic errors and biases, correction methods such as the indirect delta change (DC) and the direct distribution-based scaling (DBS) methods, should be applied before using the output in hydrological simulations (Seaby *et al.*, 2013).

4.1 Correction Methods

Correction methods can be indirect, where corrections are based only on observed data, or direct, where projections from an RCM are used. The DC bias correction method is indirect, where mean meteorological variables for two periods, representing a control and a simulated future climate period, are used. The difference in mean meteorological variables between the reference and future period

gives the delta change factor (Van Roosmalen, Christensen and Sonnenborg, 2007; Seaby *et al.*, 2013). The reference period is represented from 1971-2000 and the future climate is represented in 2071-2100, thus representing 30 years. The delta change factor is applied to P, Q, T, and potential ET, to obtain climate data for the far future. The DC method makes the critical assumption that system dynamics observed in the reference period will remain the same in the future period (Seaby et al., 2013). Since the method uses the mean for the control and future period, changes in extremes and amount of rain days in the future climate are not captured in this downscaling method (Van Roosmalen, Christensen and Sonnenborg, 2007; Seaby *et al.*, 2013). Another method used is DBS. By this method, the function of the RCM outputs is adjusted to match the function of observed data. A transfer function is found from the simulated and observed cumulative distribution functions. A main difference between the two downscaling methods is the ability to reproduce extreme events in the future. Here DBS is more suitable than DC, as it preserves the variance and mean of the RCM outputs, hence keeping the statistical distribution of observed data (Seaby *et al.*, 2013). Both methods have been applied to the reference data to evaluate the two downscaling methods.

4.2 Climate Models and Uncertainty

Karlsson et al (2016), argues that the effect of climate change can be viewed as a cascade of uncertainties. The uncertainty arises from the top when estimating emission scenarios and propagates to the bottom of the climate impact chain. Uncertainty sources in the use of a climate scenario include the lack of knowledge of human development in the future and climate sensitivity in feedback mechanisms due to increasing emissions (Van Roosmalen, Christensen and Sonnenborg, 2007). Another uncertainty is of course the chosen climate model ensemble. Karlsson et al (2016) argue that another climate model ensemble would represent another emission scenario leading to a different result. The GCM-RCM pairing also has its uncertainties and limitations. The GCM is effective in capturing large-scale climate features, but due to the coarse resolution and inability to resolve sub-grid processes, it is ineffective on a regional scale. Since the RCM uses the lateral boundary conditions from the GCM, the uncertainties effectively propagate to the dynamic downscaling (Seaby *et al.*, 2013). As previously discussed the downscaling method DC uses the reference climate as a baseline for the future projection of the climate, which also has its uncertainties since the future climate has been argued to become more extreme (Van Roosmalen, Christensen and Sonnenborg, 2007; Seaby *et al.*, 2013).

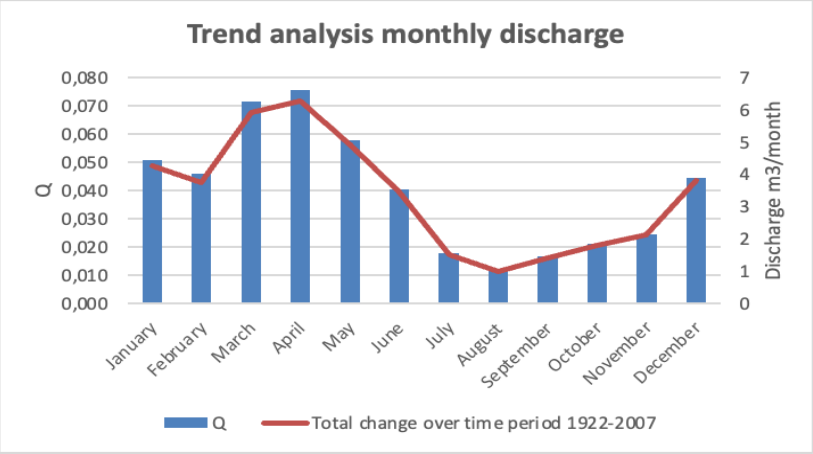
5. Results and discussion

5.1 Homogeneity test

The homogeneity test was applied for P and Q data by using the double mass curve method. A break for P was identified in 1959, and for Q in 1964. The inhomogeneity was corrected by using the correction factor for P, a correction factor of 1.33, and for Q, the correction factor was 0.67. Consequently, homogenous data for the whole period was obtained and carried forward into the assessment.

5.2 Trend test

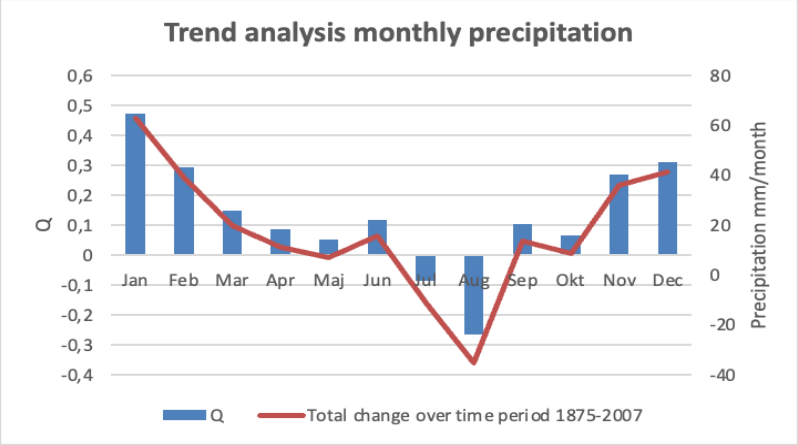
The trend test was applied for P, T, Q, and ET to obtain the trend over time. The trend test was done on monthly climate data. Q has been analyzed from 1922-2007, whereas P and T have been analyzed from 1875-2007. The trend test for annual maximum values was done for P, T, Q, and ET.



Month	Significance
January	+
February	+
March	***
April	***
May	***
June	***
July	+
August	
September	
October	
November	
December	*

Figure 3: ***Left; Trend analysis for monthly Q. Q (blue) and total change over the time period for Q per. Month (red). Right; Significance level for individual months from the trend analysis***

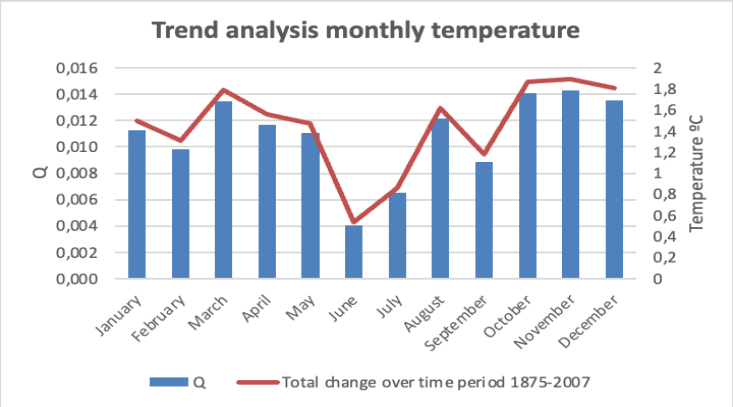
The yearly trend analysis for the whole period 1922-2007 showed a total increase in yearly average Q of 4 m3/yr, with a strong statistical significance (α=0.001). The trend analysis for Q (Fg.3) shows an increasing trend for all months. However, this is most pronounced from December to June, reflected by the total change in discharge and the corresponding significance level. The highest statistical significance is found from March to June, indicating a strong increase in Q in the respective months. For the other months, there is also a statistical significance, although lower. For August to November, there is no implied statistical significance in the monthly trend.



Month	Significance
January	***
February	***
March	*
April	
May	
June	
July	
August	*
September	
October	
November	**
December	***

Figure 4: ***Left; Trend analysis for monthly P, Q (blue), and total change over the period for P per. month (red). Right; Significance level for individual months from the trend analysis***

The trend analysis for the annual P data during the historical period 1875-2007 showed a total increase in yearly average of 232 mm, with a strong statistical significance (α=0.001). When looking at the trend test for individual months there is an increasing trend for all months, with the highest increase from Nov. to Feb. (Fg.4). This is also reflected by the total change per month and supported by the high statistical significance (Fg. 4, right). There is a decreasing trend for August, although with a weaker statistical significance (a=0.1). The findings of the trend analysis correspond with the findings of Karlsson (2014). The annual pattern shows strong seasonal changes, as the winter gets wetter and the summer gets drier, also argued in Karlsson (2014). The seasonal pattern for P is reflected by the monthly trend in Q. The trend analysis for Q reflects an increase for all months, but lower in the summer. Although there is a decrease in the summer P, Q still has an increase. The increase in Q in summer can be attributed to the buttressing effect, which is where the P mainly occurs during the winter when the groundwater fills up and gives a delayed effect of Q (Karlsson et al., 2014).



Month	Significance
January	*
February	
March	**
April	***
May	***
June	
July	*
August	***
September	***
October	***
November	***
December	**

Figure 5.: ***Left; Trend analysis for monthly T, Q (blue), and total change over the time period for T per. month (red). Right; Significance level for individual months from the trend analysis***

The trend analysis for monthly T shows an increasing trend for all months with a corresponding high significance for most months (Fg. 5). The total increase for the T over the whole study period is 1.5 ºC, with a strong statistical significance (α=0.001). The trend analysis shows general warming over the whole study period, with the smallest increase occurring for June and July, reflected by low total change for these months. The trend analysis for monthly and yearly T increases was also found by Karlsson (2014). The general increase of T is reflected by a yearly increase of ET of 2 mm over the whole study period (1875-2007), which also had a corresponding statistical significance of 0.01. Monthly trend analysis for ET was not included, but this could have indicated a change in the seasonal trend. The maximum daily values for P, T, and Q and the maximum monthly ET for each year were analyzed using the Mann-Kendall test. The purpose is to see if there is only a trend for annual values, and if this is also reflected by an increase in the maximum values. The trend test for maximum values, all showed an increasing trend, but with various statistical significance.

5.3 Autocorrelation

Autocorrelation values for changes in homogeneous P, T, and Q datasets were calculated at incrementing lag periods between 1 and 10 days to analyze their temporal persistence and variability (Fg. 6). At a time step of 1 day between observations, T and Q were strongly autocorrelated, with respective values of 0.98 and 0.96. This indicates that over short time periods, there is a significant coupling between past and present measurements for T and Q. However, as the lag period is extended to 10 days, autocorrelation for Q diminishes to a value of 0.64, more than T, which maintains a higher value of 0.86. The stronger persistence of T patterns can be attributed to large-scale atmospheric and seasonal processes that govern T variation, such as solar radiation and synoptic weather systems. These processes exhibit consistent temporal structures that extend over days to weeks (Sepp and Jaagus, 2002). Some variation in these processes, particularly on shorter time scales, is caused by highly localized weather patterns, explaining the reduced autocorrelation value. The stronger decay seen in the Q dataset demonstrates the heterogeneous response of the Ahlergaarde Catchment to rainfall events, as the observations can be further influenced by several geological, hydrological, and climate factors. Catchment-specific factors such as soil type and permeability, vegetation coverage, catchment size, and topography can increase the rate at which autocorrelation diminishes (Hornberger *et al.,* 1998). This is relevant in this study as the Ahlergaarde Catchment is sandy.

Hydrological time series are strongly autocorrelated due to memory effects such as soil moisture content, baseflow levels, and interactions between groundwater and surface water (De Lavenne *et al.*,

2022). However, these factors become less influential as lag periods increase, because of the stochasticity of rainfall events and shifts in initial conditions. Unlike T and Q, P had much lower autocorrelation overall, with a value of 0.29 for a one-day lag time, quickly decaying to 0.03 for a 10-day period. This suggests that P can be seen as a stochastic variable, particularly over longer durations, which is not significantly coupled to past observations. Rainfall is episodic and determined by large-scale climate patterns and localized weather systems, resulting in significant temporal variability.

Due to the increasing energy in the atmosphere from climate change, rainfall patterns in Denmark are projected to become more variable and intense in the future (Karlsson et al, 2014). This includes more frequent and intense storm events interspersed with longer dry periods, a trend observed in the Mann-Kendall test results (Fg.4). Such changes are expected to reduce P autocorrelation values even further, reflecting heightened temporal variability. This variability is also likely to impact T and Q values, which are linked to processes impacted by climate change as well.

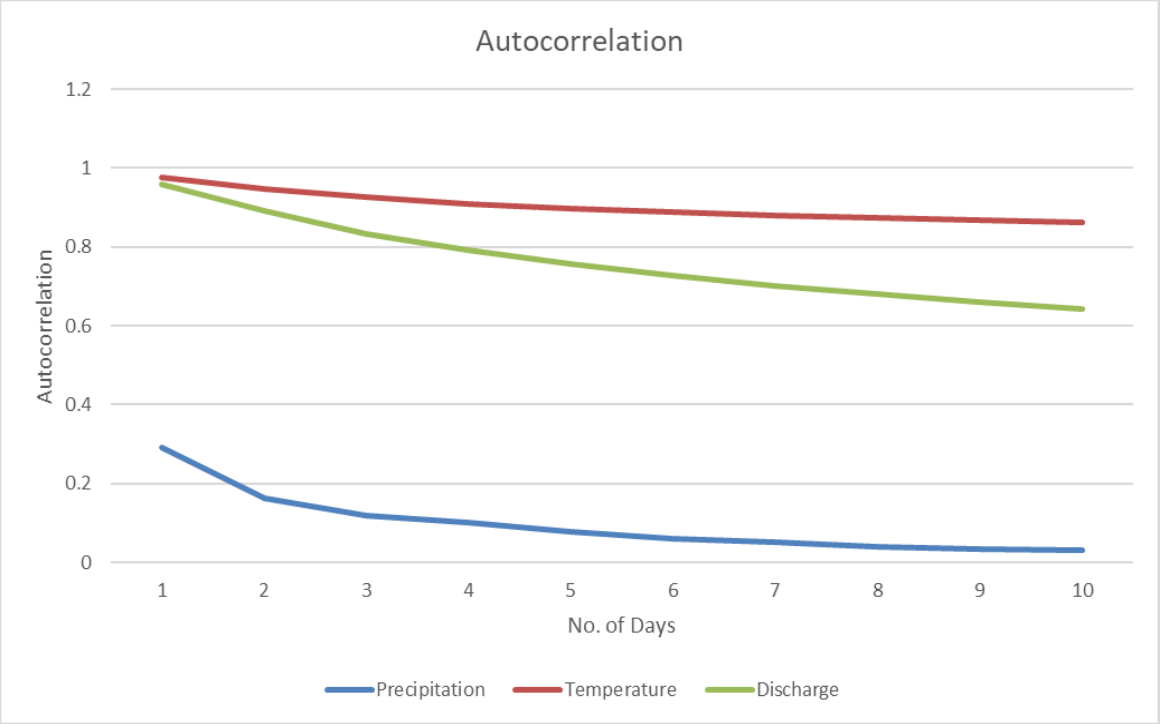


Figure 6: ***Diminishing autocorrelation values for P, T, and Q over a time period of 10 days***

5.4 Flow Duration Curve

The FDC for the Ahlergaarde Catchment (Fg. 7) shows that between the period of 1942 to 2007, the Q levels measured ranged between 119 m^3/s and 5.7 m^3/s. There is a large variance of Q levels that were exceeded less than 10% of the period, with events ranging from approximately 24 m^3/s to 119 m^3/s. This variance indicates the uncertainty associated with projecting extreme flows in catchments.

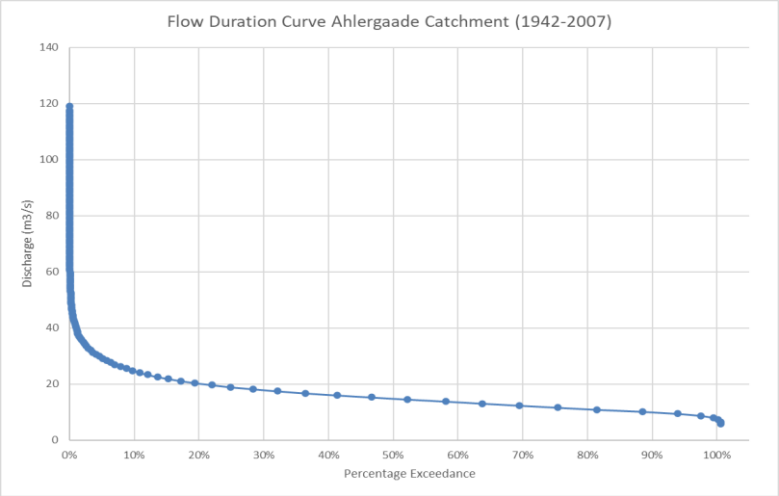


Figure 7: Flow duration curve of daily stream Q measured at the Ahlergaarde Catchment between 1942 and 2007

This FDC has been created using daily measurements taken over a period of 65 years. Although this can give a relatively good representation of the range of potential flow levels in the Ahlergaade Catchment, care should be taken when using this data to inform any decisions made about the future management of the area. The curve does not clearly show any variance in how exceedance probabilities have changed over time in response to climate change, which is notable given the significant increases in discharge as revealed by the Mann-Kendall test.

5.5 Extreme value analysis

To estimate the likelihood of extreme minimum and maximum events occurring over a certain period, extreme value analyses were carried out for P, T, and Q. As the analyses are carried out similarly for all three parameters, the results for Q are here assessed in detail and then compared to the results for P and T. The analysis of annual maximum Q (Qmax) reveals that the widening of confidence intervals for longer return periods highlights the increasing uncertainty in predicting rare and extreme events. Since data points on 200-year events are scarce, the uncertainty of the magnitude is greater than for the more documented events of a 50-year return period.

Return Period (Years)	Q_{min} ± Interval (m^3/s)	Q_{max} ± Interval (m^3/s)	T_{min} ± Interval (C)	T_{max} ± Interval (C)	P_{max} ± Interval (mm)
50	4.52 ± 0.85	95.05 ± 11.41	−15.29 ± 0.84	26.42 ± 1.06	57.39 ± 5.31
100	3.77 ± 0.95	105.32 ± 74.25	−17.24 ± 0.98	27.38 ± 19.30	35.46 ± 25.00
200	3.01 ± 1.12	115.55 ± 93.03	−19.19 ± 1.12	28.33 ± 22.81	35.56 ± 28.63

Table 1: ***Flow rates (Qmin, Qmax), Ts (Tmin, Tmax), and P (Pmax) with confidence intervals for 50, 100, and 200-year return periods. Made on the reference period 1971-2000.***

For the Ahlergaarde Catchment (table 1), the 50-year return period has a predicted Qmax magnitude of 95,05 m3/s with a confidence interval of ± 11.41. In contrast, the 100-year return period has a predicted magnitude of 105.32m3/s ± 74.25, and the 200-year return period has a Qmax of 115.55m3/s ± 93.03. The results for Qmax show that longer return periods exhibit greater magnitudes, but they also come with much greater uncertainties, as indicated by the confidence intervals (table 1). A confidence interval constituting 80.5% of the Qmax extreme events means that the results of 200-year magnitudes are highly unreliable, but showcase the issue of predicting rare events.

A small confidence interval generally is preferred, as the long-term estimates are used for practical applications, such as flood risk assessments and the design of infrastructure to withstand extremes. In the case of the Qmax 200-year return period, the results are not reliable for practical applications. These patterns of uncertainty are also applicable to the results of Tmax and Pmax (table 1) and could be recalculated for a longer reference period if available.

6. Hydrological modeling

6.1 Sensitivity Analysis

The NAM model was initially calibrated manually using a trial-and-error process to identify the most sensitive parameters influencing the Nash-Sutcliffe coefficient, water balance, and RMSE values. The trial-and-error process revealed that UMAX was inversely proportional to water balance, meaning a higher UMAX parameter value achieved a water balance closer to zero. A higher UMAX value also caused the correlation coefficient and R^2 values to increase slightly. This process also identified that changes to CQOF had the most influence on the representation of peak flow values. These relationships helped to identify the CQOF parameter as the most sensitive, followed by TOF, TG, UMAX, and TIF. The parameter values identified in the manually calibrated model and its performance against the validation period are shown in Fg. 8. The manual calibration produced a correlation coefficient of 0.84, a water balance of 0.9, and a Nash-Sutcliffe coefficient of 0.65.

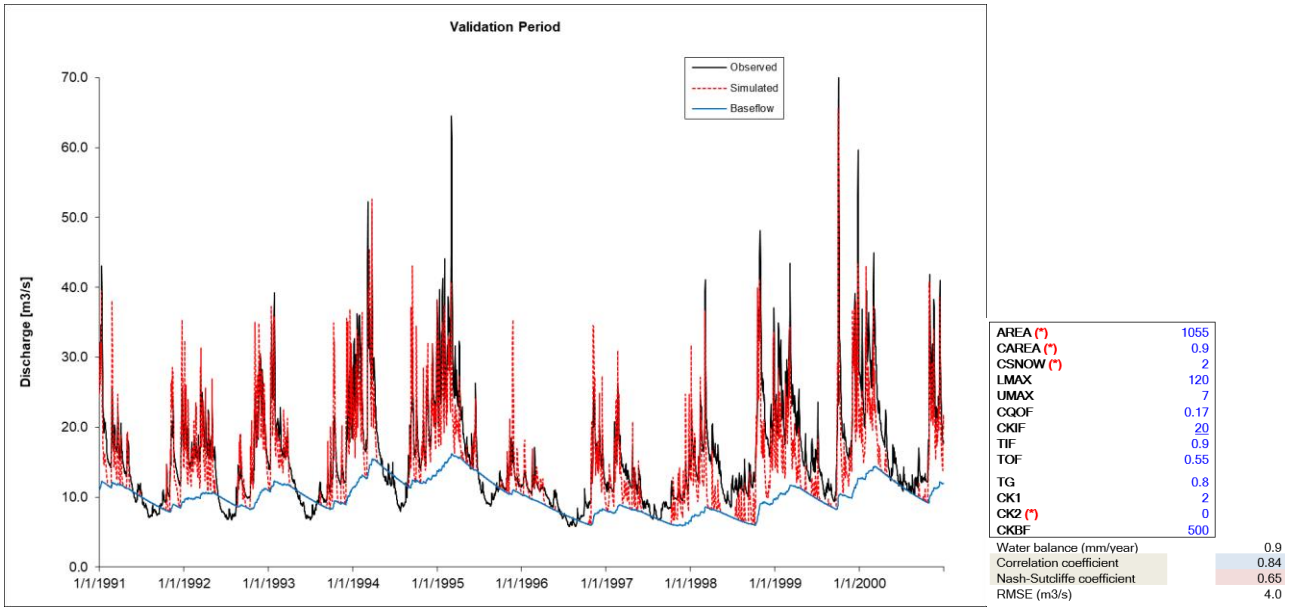


Figure 8*: Left: The simulated NAM model Q from the chosen model compared with observed Q from the validation period (1991-2000) Right: The manually selected parameter values which resulted in equally strong performance criteria values*

The resultant model from the manual calibration process was compared to the unused validation data from the period between 1991 and 2000. This showed that despite resulting in relatively good values for the performance criteria, the model struggled to accurately represent peak flow values (Fg. 8). However, since the timing of the Q was accurately represented and the performance criteria were met in a balanced way, the manually selected parameter values were used for autocalibration.

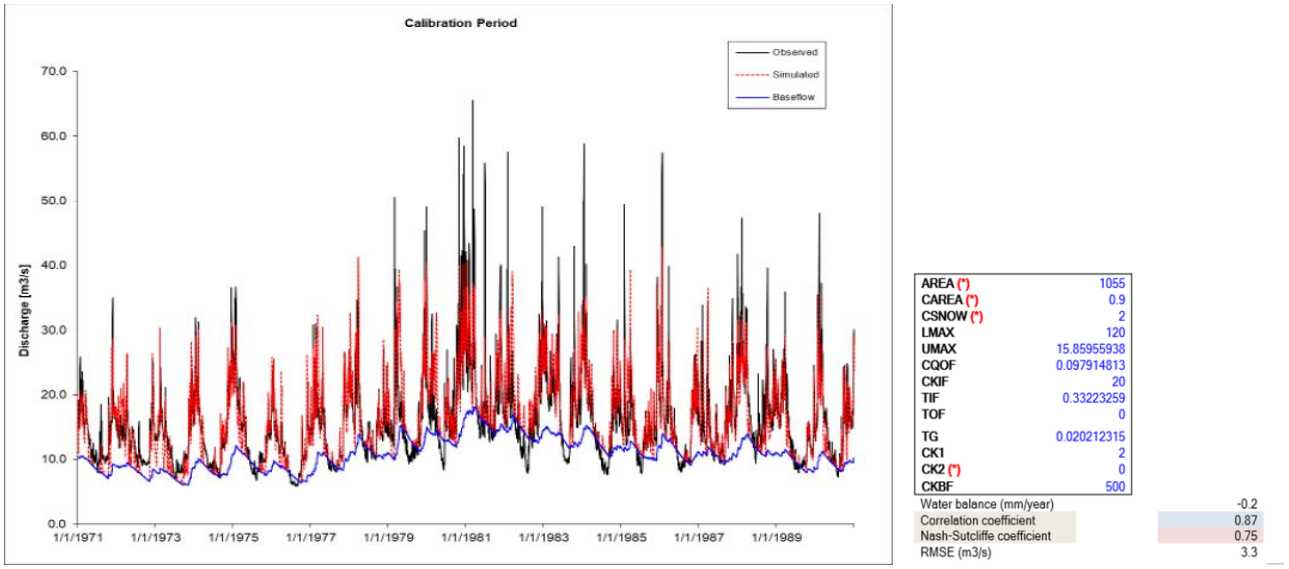


Figure 9*: Left: Multiple objective functions with regard to water balance and the Nash-Sutcliffe coefficient (auto-calibration). Right: Hydrological model based on the parameter combination from the multiple objective function. Auto calibration during the calibration period.*

Compared to the manual-focused calibration, the auto-calibration model showed a slight overall improvement in the parameter values. The primary change is related to the water balance and the Nash-Sutcliffe coefficient, which the model was optimized for in the multiple objective function (Fg. 9). Although this value combination gives the overall best statistical fit, the calibration period shows an underestimation of peak flows and low flows for the years 1980-2000 (Fg. 9). This indicates that the peak estimations in the model do not accurately represent the observed flow dynamics. Hence the peak flows in the validation data are also underestimated greatly (Fg. 10). Here, however, the low flows are well represented. The seasonal variance in Q is modeled with relative precision in both the calibration and validation. Consequently, the auto-calibrated model demonstrates greater utility for hydrological analyses of seasonal variability compared to its application in analyzing peak flow events.

With the autocalibration approach, six parameters are identified as optimal for achieving the best model performance, whereas only five were adjusted in the manual calibration process. This suggests that autocalibration explores a broader parameter composition, potentially leading to a more accurate or optimized model outcome. It systematically tests and evaluates additional parameters. This refinement allows the model to account for nuances in the data that might be missed during a manual calibration, which often relies on subjective judgment and limited testing of parameter adjustments. However, it is still possible to get the most accurate model based on manual calibration (Madsen, 2000).

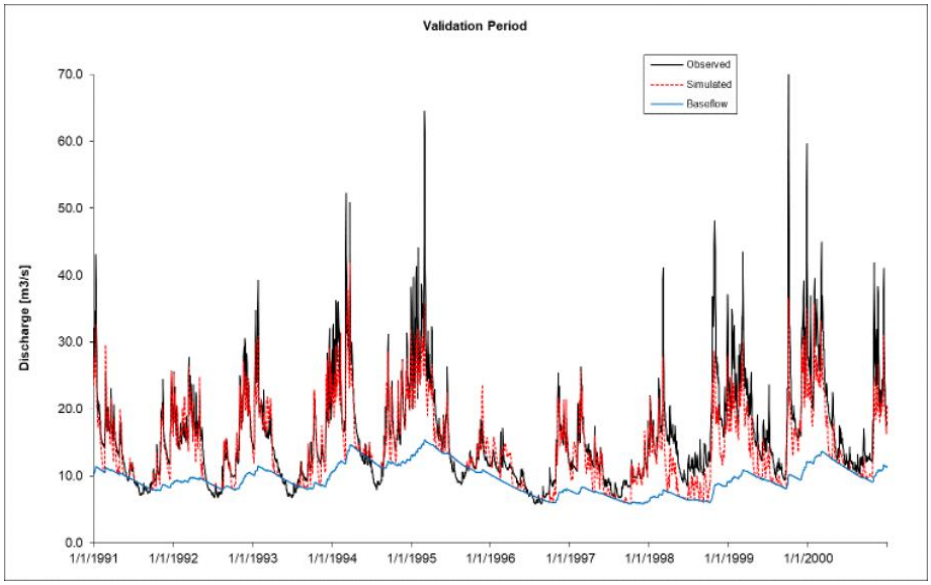


Figure 10: ***Validation model based on the parameter combination output from the multi-objective function***

As closure of the water balance has been a significant challenge in modeling water circulation at a catchment scale in Denmark (Jensen and Refsgaard, 2018), producing a model which optimized for a water balance close to zero was prioritized in the manual calibration process. When water balance was chosen as the objective function a value close to zero (-0.2) was achieved (Fg. 9). For the calibration period, water balance was maintained within a reasonable range. However, for the validation period, an ongoing decrease in water balance, particularly after 1999 was produced (Fg. 11). This issue was also observed in the single component objective function optimizing for water balance and the Nash-Sutcliffe coefficient; however, the multi-objective calibration method produced the most precise simulations when accounting for all statistical measures, and hence was carried forward into the following stages of the study.

The water balance closure problem largely stems from uncertainties in data collection and processing during the model calibration stage (Jensen and Refsgaard, 2018). The water balance discrepancies, particularly during the validation period, may occur due to unmodeled spatial changes in soil characteristics. As NAM is a lumped conceptual model, it does not integrate the spatial variability known to distinguish the north from the south of the catchment. Also, without incorporating urbanized areas, agricultural practices, and runoff patterns, the model's ability to maintain water balance in future simulations deteriorates over time (Karlsson *et al.*, 2016). The uncertainties may also arise from measurement methods or correction errors. This poses significant implications for the future application of the calibrated NAM model, as it is highly unlikely that the model will be able to maintain an ideal water balance when making projections, especially when using climate conditions projected in the late-century period.

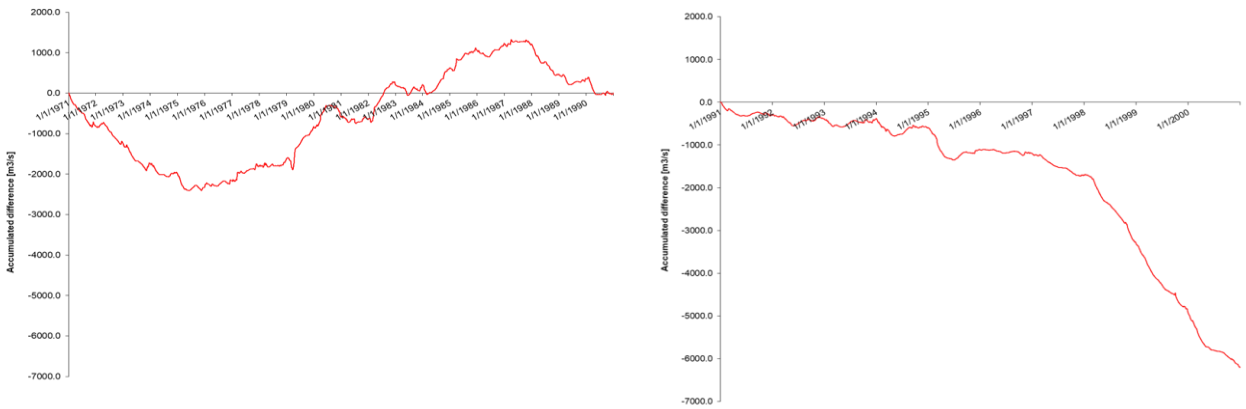


Figure 11: ***Left: Water balance for the calibration period for the auto-calibrated model optimized for water balance. Right: Water balance for the validation period***

7. Climate and hydrological projections

7.1 Correction Method Results

To see if there was a significant change in annual P, T, Q, and ET between the historical period and the future period, the t-test was performed with the null hypothesis that there *is no difference between the historical data and the future data* (table 2). The significance level is 0.05 for a two-tailed test. The degrees of freedom for the DBS and DC method for all climatic parameters are 27 and 28 and the values of significance are 2.0518 and 2.0484 respectively.

Table 2: T-test results for DBS and DC correction methods

Time series	P	T	Q	ET
2071-2100 DBS	2.4	-9.81	1.79	-3.4
2071-2100 DC	1.9	-9.30	5	-19

The t-test implies that there is a statistical difference between the historical and future period, for all climatic parameters except for future P with the DC method and Q with the DBS method. Changes in P between the reference and future period were not significant for both DC and DBS methods. For the T, the average monthly temperature is expected to be higher in the future compared to the reference period for both downscaling methods (Fg. 12). The increase of T in general and a peak in August is also found in other studies (Van Roosmalen, Christensen and Sonnenborg, 2007; Jensen and Refsgaard, 2018). For the average ET, the reference and future follow the same overall pattern, but there is a peak in the average for both the DC and DBS methods in August, which could be

connected to the increased T. The T also reflects the higher ET in August, since we have the peak Ts suggested in future climate for both methods (Fg.13).

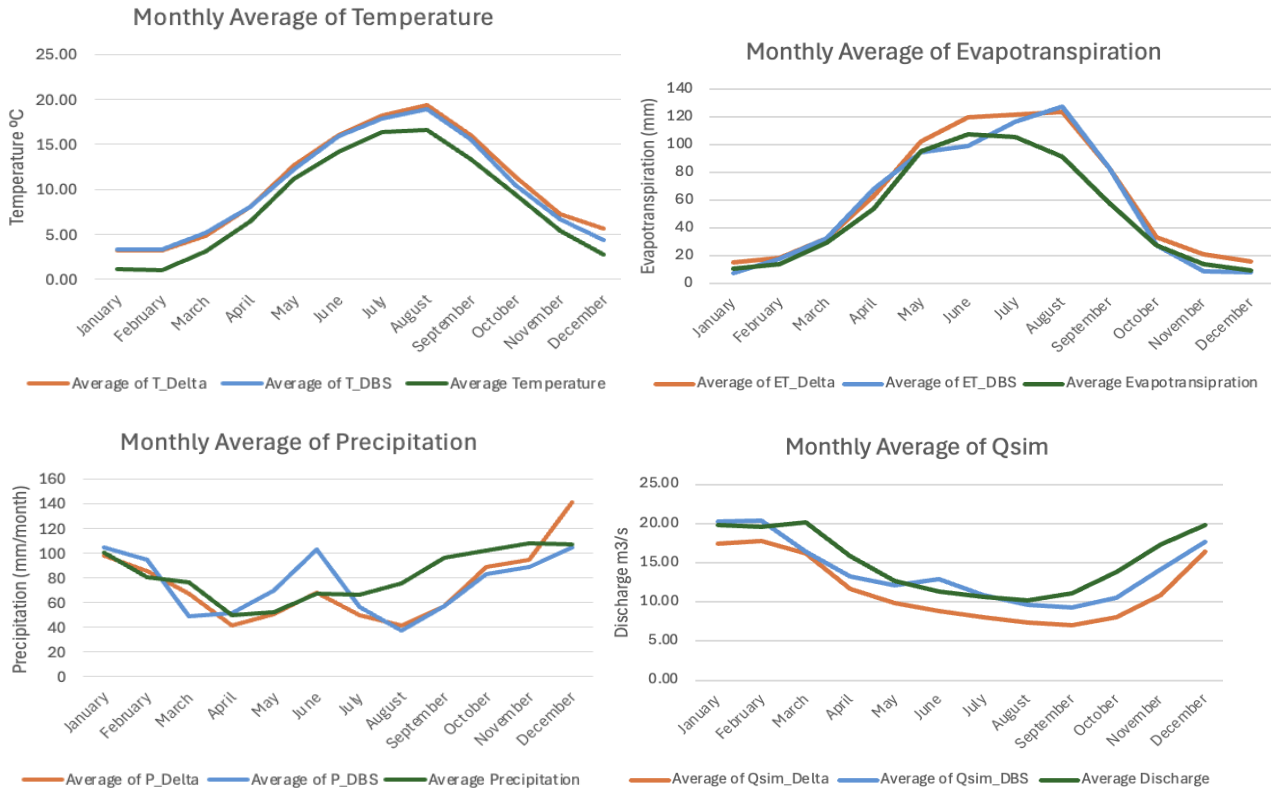


Figure 12*: Top left: Average T for the future period with the DC (orange), DBS (blue), and the reference period (green). Top right: Average ET for the future period with the DC (orange), DBS (blue), and the reference period (green). Bottom left: Average P for the future period with the DC (orange), DBS (blue), and the reference period (green). Bottom right: Average simulated Q (m3/s) for the future period with the DC (orange), DBS (blue), and the reference period (green).*

The average for the daily P does not follow the same pattern when comparing the reference climate data with the future climate. For the reference period, the P is around 60 mm, but for the winter it is approximately 100 mm with a maximum amount of P in November and December of 110 mm. For the future climate, the DC method represents the same pattern of monthly P from January to June of the reference period, but a lower P for the rest of the year although suggesting a peak in December exceeding the P for the reference period (Fg. 12). In contrast, the DBS method suggests a much higher P in summer with a peak in June, and a preceding lower P compared to the reference period with a minimum in August. The peak in P in June is not represented in a higher potential ET for the DBS method. This could be due to the type of rainfall. As discussed earlier the total amount of P does not differ a lot between the reference period and the future period, but the low ET could indicate a high cloud cover in this month, and thereby a persistent amount of little rain throughout June in the future

period (Teuling *et al.*, 2019; Calvin *et al.*, 2023). The minimum P for August is also reflected in the potential ET. The peak for ET in August could be connected to low P, as this could result in low cloud cover.

For the average of Q, the historical and the future time series follows the overall same pattern over the year. The DC and DBS methods both generally estimate a lower Q than the historical time series, except for June where we see a peak in P with the DBS method (Fg. 12). Other findings in general suggest a general increase in P in the future, and also a more pronounced seasonal pattern with drier summers and wetter winters (Thodsen, 2007). However, for this study, these are not the findings in the modeled future climate. Other reports such as Karlsson et al (2016), have made a comparable analysis of the same catchment area as used in this report. In the findings of Karlsson, the ARPEGE-RM5.1 reflects a decrease in mean annual P by 11% (Karlsson *et al.*, 2016). This is compared to our results with a decrease in P from the reference period to our future period of 10% for DC and an 8% decrease for DBS.

7.2 Annual Results

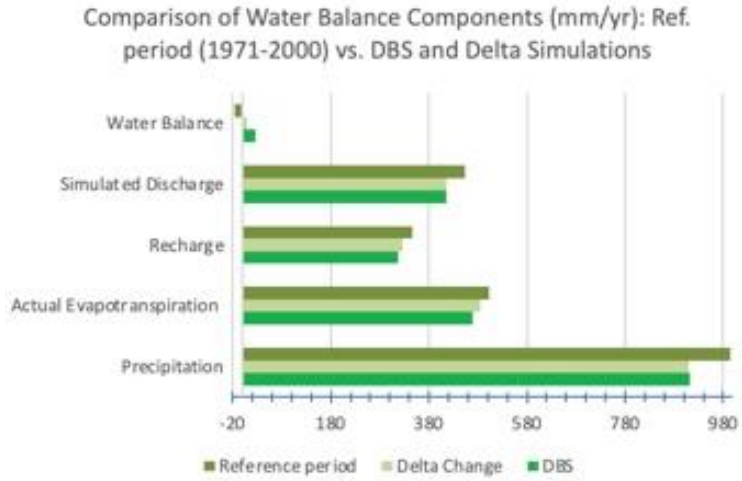


Figure 13*: comparison bar chart: Water balance components modeled by NAM for present hydrological processes compared to two future simulations (Delta Change and DBS)*

Both DBS and DC predict changes in the water balance components compared to the reference period, with ET and recharge showing notable differences (Fg. 13). Based on the water balance, the DC model demonstrates a more balanced simulation with a lower residual compared to the DBS model. Generally, both simulation models predict an annual average decrease in P, compared to the reference data, leading to an underestimation of all hydrological components, based on the general scientific understanding of future patterns. (Fg. 12) (Andersen *et al.*, 2006; Thodsen, 2007). We would expect

the components to increase in both future models, compared to the reference period. However, none of the models effectively capture the anticipated precipitation increase in the components. This is because the underestimation of future precipitation levels has a cascading effect on all other hydrological processes, leading to their subsequent underestimations in the simulations (Thodsen, 2007).

7.3 Extreme Value Analysis

Table 3: Extreme value analysis results for DBS (2071-2099)

Return Period (Years)	Q min ± Interval (m^3/s)	) max ± Interval (m^3/s	T min ± Interval (C)	T max ± Interval (C)	P max ± Interval (mm)
50	4.38 ± 0.93	61.46 ± 6.10	-10.44 ± 1.09	34.36 ± 1.72	85.27 ± 9.86
100	3.57 ± 1.08	66.77 ± 7.11	-11.39 ± 1.27	35.86 ± 2.00	93.86 ± 11.49
200	2.76 ± 1.24	72.06 ± 8.11	-12.34 ± 1.45	37.35 ± 2.29	102.41 ± 13.12

Table 4: ***Extreme value analysis results for DC (2071-2100)***

Return Period (Years)	Q min ± Interval (m^3/s)	) max ± Interval (m^3/s	T min ± Interval (C)	T max ± Interval (C)	P max ± Interval (mm)
50	2.21 ± 1.03	47.78 ± 4.10	-13.00 ± 2.20	29.36 ± 1.24	59.18 ± 6.21
100	1.30 ± 1.20	51.41 ± 4.78	-14.95 ± 2.57	30.46 ± 1.44	64.68 ± 7.24
200	0.39 ± 1.37	55.03 ± 5.46	-16.90 ± 2.93	31.55 ± 1.65	70.17 ± 8.27

The extreme analyses carried out for both the DBS and DC corrected models for the period between 2071–2100 showed a notable increase in the extreme values projected when compared to the reference periods. Although the projections of average values for P, Q, and T only increase slightly (Fg. 3, 5, 6), the changes in distribution indicate that rainfall events may become more dispersed but significantly more intense.

The DC model showed more extreme minimums in measured Q compared with the DBS method and the historical reference period (Table 3, 4). This could be attributed to the increased sensitivity of the DC model to changes in seasonal patterns. This indicates that, if changes in the climate follow the trajectory of the DC model, drought or low-flow events may increase in frequency in the Ahlergaarde Catchment in the late-century period. In contrast, the DBS model suggests that minimum Q will become less extreme in the same period. For instance, the minimum annual Q with a return period of 200 years is projected at 2.76 m^3/s in the DBS model, compared to a more consequential value of 2.21 m^3/s under the DC model. This suggests the DC model may better emphasize changes in seasonal flows, while the DBS model projections are less variable.

For both late-century projections, minimum annual Ts are expected to be higher than those of the reference period (reference period 50-year return value of -19.27°C compared to -12.34°C for the DBS method and -16.90°C for the DC method). Warmer minimum Ts are likely to impact the current

hydrological behavior of the Ahlergaarde Catchment, with decreases in snowpack and frost days altering groundwater recharge rates and seasonal flows (Healy et al., 2007). The DC model projects that a maximum annual T expected to occur once in 200 years (28.44℃) in the reference period will have a return period of less than 50 years in the late-century period. The DBS model amplifies this further, projecting the 200-year return period T at 37.35℃, nearly 10℃ higher than the reference period. This indicates heat waves of a greater magnitude occurring in the late-century period.

Under both correction methods, P events are expected to become more extreme. The DBS model projects a more extreme annual maximum P than the DC model, however with greater uncertainty. A 1-in-200-year rainfall event projected using the DBS model is expected to be 102.41 mm with a deviation of ±13.12 mm, whereas an event with the same return period for the DC method is only 70.17 mm, with a lower deviation of ±8.27 mm. In the near future, uncertainty in the precipitation results is mainly attributed to natural variability in rainfall, however, for the late century period, variance in the results is caused predominantly by the climate model (Hawkins and Sutton, 2009).

The DC model produces extreme values with a lower margin of error (Karlsson *et al.*, 2014), which may indicate greater robustness for projecting averages and capturing gradual trends. However, its lower sensitivity to variability can under-represent the frequency and magnitude of tail-end events. This could mean it is less suitable for scenarios requiring detailed analysis of rare, high-impact occurrences.

The DBS method, which is designed to preserve statistical distributions, is effective at modeling tail-end events (Seaby *et al.*, 2013). Its ability to model the intensity of rare extremes provides insights for planning infrastructure and response strategies. However, the greater uncertainty associated with the DBS model's projections, particularly for variables like precipitation, suggests comparing the outputs with other climate models in an ensemble can refine the results and reduce uncertainty.

7.4 Ensembles and Uncertainty

Uncertainty is a core issue affecting the reliability of climate and hydrological projections. It arises at multiple stages of modeling, including the choice of climate models, the underlying climate scenarios, and the downscaling methods used (Pasten-Zapata, Sonnenborg and Refsgaard, 2019). Statistical approaches such as constructing ensembles from a variety of GCM-RCM combinations can be used to quantify this uncertainty, improving the robustness of the analyses. By aggregating the results from each GCM-RCM combination by month, T, Q, and T projections from the DBS, DC,

and reference period datasets can be compared with each other using a model ensemble chart (Fg. 14, 15, 16).

Most model combinations are close to reaching an agreement of the approximate values of T for the far future. The mean in T is in general following the same pattern as for the reference period, although the ensemble average for the climate projections is in general suggesting a higher year-round temperature. When taking the maximum values into account, the ensemble average for the climate projections is in alignment suggesting an increase in maximum summer temperatures higher than for the reference period (Fg. 14). The most variability between the model combinations is seen in the projections for maximum T in the far future period using the DBS method (Fg. 14).

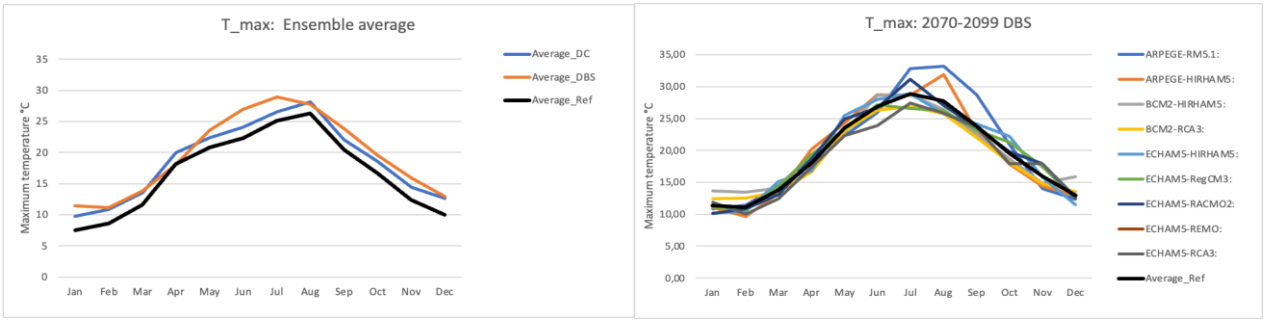


Figure 14*: Left: Ensemble averages for maximum temperature for each month from the reference period, DBS model and DC model. Right: monthly maximum T projections from nine GCM-RCM combinations using the DBS method on the far future period.*

The mean for the ensemble for discharge suggests a generally higher yearly Q for the DBS method, while the DC suggests the pattern of Q is relatively unchanged from the reference period (Fg. 15). This is also reflected by the higher SNR value for mean discharge (DBS), indicating a lower uncertainty (table 5). A higher mean Q for the DBS model is likely due to its ability to model extreme events with better accuracy than the DC method (Seaby *et al.*, 2013).

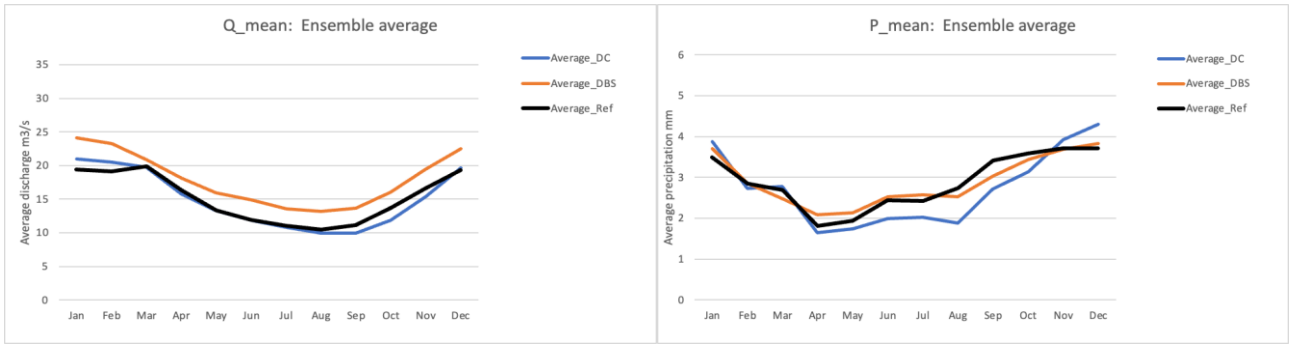


Figure 15*: Left: Monthly mean average values of Q taken from nine GCM-RCM combinations. Right: Monthly mean average values of P taken from nine GCM-RCM combinations.*

When evaluating the future period with the ensemble of climate projections, the variation from the reference period is most evident when looking at precipitation. The different climate projections both suggest an average mean precipitation which is lower and higher than the reference period. The ensemble average of the maximum precipitation for the DBS method generally suggests a higher maximum precipitation for most of the year, whereas the DC method in contrast follows a similar pattern as the reference period (Fg. 16). This difference is likely due to the DBS methods' susceptibility to the drizzle effect, where RCMs tend to have a wet day bias which results in excessive low intensity precipitation on a high amount of days (Seaby *et al.*, 2013). Since DBS is based on the distribution and frequency of wet days, it is likely that the DBS projections have been impacted by the RCM error more than the DC projections.

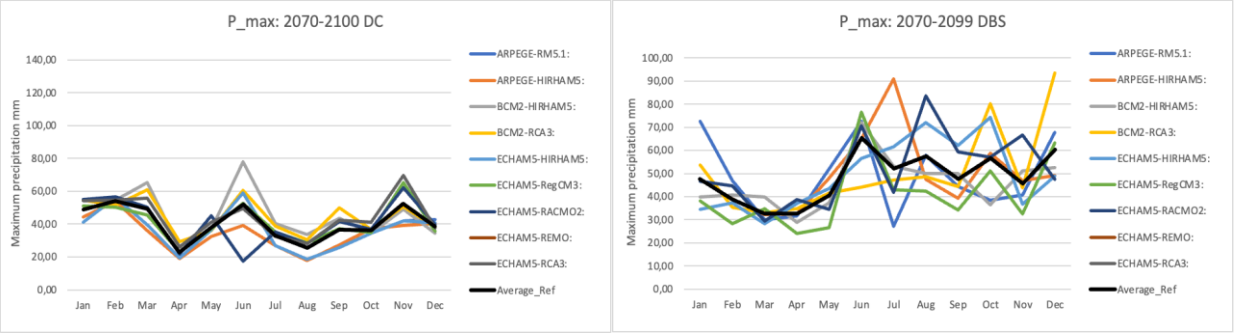


Figure 16*: Left: Maximum precipitation DC method. Right: Maximum precipitation DBS method*

Table 2*: SNR mean values for DC and DBS methods*

	Mean discharge	Max discharge	Min discharge	Mean precipitation	Max precipitation	Mean temperature	Min temperature	Max temperature
Total SNR DC	3,70	4,57	7,10	2,90	3,36	1,88	-7,15	3,02
Total SNR DBS	4,46	3,31	7,09	4,19	3,12	1,88	-6,80	2,99

To assess the impact of uncertainty related to downscaling, we compare the average of the signal-to-noise ratio (SNR) quantified for nine different climate models. SNR, determined by dividing the ensemble mean by the standard deviation, reflects uncertainty, where a lower SNR indicates high uncertainty (Pasten-Zapata, Sonnenborg and Refsgaard, 2019). This was done on the monthly mean, minimum, and maximum values for the future period and assessed using both DC and DBS (table 5). Comparing the two methods, DC and DBS produce similar SNR values, implying that choosing one method that produces less uncertainty and provides more reliable projections is challenging. The SNR values are calculated on all nine RCM-GCM combinations, which may disguise some differences in the performance of the individual combinations. The similarity between the two downscaling methods can indicate that the variability is relatively small compared to what is inherent across the climate models themselves, making it difficult to definitively favor one method over the other.

Conclusion

This report on the hydrological processes of the Ahlergaarde Subcatchment underlines the importance of consecutive long-term datasets when modeling hydrological processes with the objective of understanding the catchment dynamics and predicting future changes. Historical trends indicate significant increases in T, P, and Q along with increased seasonal variability, including wetter winters and drier summers. Through DC and DBS downscaling methods, these seasonal patterns are found to intensify under most future climate scenarios, implying a need to update local politics and plans for water management, especially for flood risk mitigation and agricultural practices.

The NAM model has proved to be effective in simulating key water balance components and streamflow although having limitations in model calibration and peak flow representation as a cause of the model simplicity. Projections for the far future (2071-2100) indicate an increased frequency and magnitude of extreme hydrological events, however, these results are followed by large uncertainties, which brings implications for decisions around managing and constructing infrastructure in the catchment area.

The primary sources of uncertainty influencing the results of this report have been identified as model parameterization, data quality, and climate projections. To mitigate these uncertainties, validation procedures and bias correction methods have been implemented. Although modeling is followed by uncertainties, this study is successful in assessing future hydrological dynamics and the implications hereof.

Reference list

Andersen, H.E. *et al.* (2006) 'Climate-change impacts on hydrology and nutrients in a Danish lowland river basin', *Science of The Total Environment*, 365(1–3), pp. 223–237. Available at: https://doi.org/10.1016/j.scitotenv.2006.02.036.

Birkens (n.d), MFP, van Geer, FC, Stochastic Hydrology, Department of Physical Geography, Utecht University

Calvin, K. *et al.* (2023) *IPCC, 2023: Climate Change 2023: Synthesis Report. Contribution of Working Groups I, II and III to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change [Core Writing Team, H. Lee and J. Romero (eds.)]. IPCC, Geneva, Switzerland.* First. Intergovernmental Panel on Climate Change (IPCC). Available at: https://doi.org/10.59327/IPCC/AR6-9789291691647.

De Lavenne, A. *et al.* (2022) 'Quantifying multi-year hydrological memory with Catchment Forgetting Curves', *Hydrology and Earth System Sciences*, 26(10), pp. 2715–2732. Available at: https://doi.org/10.5194/hess-26-2715-2022.

Hawkins, E. and Sutton, R. (2009) 'The Potential to Narrow Uncertainty in Regional Climate Predictions', *Bulletin of the American Meteorological Society*, 90(8), pp. 1095–1108. Available at: https://doi.org/10.1175/2009BAMS2607.1.

Healy, RW, Winter, TC, LaBaugh, JW, Franke, OL, 2007, Water budgets: Foundations for effective water-resources and environmental management, U.S. Geological Survey Circular 1308,

Hornberger, G. *et al.* (1998) 'Catchment Hydrology: The Hillslope-Stream Continuum, Ch.9'.

IPCC (n.d) 'Emissions Scenarios'. Available at: https://archive.ipcc.ch/ipccreports/sres/emission/index.php?idp=03 (Accessed: 9 January 2024).

Jensen, K.H. (no date) *Short note on Gumbel extreme vakue analysis*.

Jensen, K.H. and Refsgaard, J.C. (2018) 'HOBE: The Danish Hydrological Observatory', *Vadose Zone Journal*, 17(1), pp. 1–24. Available at: https://doi.org/10.2136/vzj2018.03.0059.

Karlsson, I.B. *et al.* (2014) 'Historical trends in precipitation and stream discharge at the Skjern River catchment, Denmark', *Hydrology and Earth System Sciences*, 18(2), pp. 595–610. Available at: https://doi.org/10.5194/hess-18-595-2014.

Karlsson, I.B. *et al.* (2016) 'Combined effects of climate models, hydrological model structures and land use scenarios on hydrological impacts of climate change', *Journal of Hydrology*, 535, pp. 301–317. Available at: https://doi.org/10.1016/j.jhydrol.2016.01.069.

Kumar, P., Lohani, A.K. and Nema, A.K. (2019) 'Rainfall Runoff Modeling Using MIKE 11 Nam Model', *Current World Environment*, 14(1), pp. 27–36. Available at: https://doi.org/10.12944/CWE.14.1.05.

Larsen, R. (no date) *Naturnationalpark Skjern Floddelta*, *Naturnationalpark Skjern Floddelta*. Available at: https://danarige.dk/naturnationalpark-skjern-floddelta/.

Madsen, H. (2000) 'Automatic calibration of a conceptual rainfall–runoff model using multiple objectives', *Journal of Hydrology*, 235(3–4), pp. 276–288. Available at: https://doi.org/10.1016/S0022-1694(00)00279-1.

Meals, D., Dressing, S. and Harcum, J. (2011) 'Statistical analysis for monotonic trends, Tech Notes 6'.

Motavita, D.F. *et al.* (2019) 'The comprehensive differential split-sample test: A stress-test for hydrological model robustness under climate variability', *Journal of Hydrology*, 573, pp. 501–515. Available at: https://doi.org/10.1016/j.jhydrol.2019.03.054.

'NAM model documentation' (1999).

Pasten-Zapata, E., Sonnenborg, T.O. and Refsgaard, J.C. (2019) 'Climate change: Sources of uncertainty in precipitation and temperature projections for Denmark', *Geological Survey of Denmark and Greenland Bulletin*, 43. Available at: https://doi.org/10.34194/GEUSB-201943-01-02.

Seaby, L.P. *et al.* (2013) 'Assessment of robustness and significance of climate change signals for an ensemble of distribution-based scaled climate projections', *Journal of Hydrology*, 486, pp. 479–493. Available at: https://doi.org/10.1016/j.jhydrol.2013.02.015.

Searcy, J.K. (1959) *Flow-duration curves*. Available at: https://doi.org/10.3133/wsp1542A.

Searcy, J.K. and Hardison, C.H. (1960) ‘Double-Mass Curves Manual of Hydrology: Part 1. General Surface-Water Techniques’, 1960. Available at: https://pubs.usgs.gov/wsp/1541b/report.pdf.

Sepp, M. and Jaagus, J. (2002) ‘Frequency of circulation patterns and air temperature variations in Europe’, *BOREAL ENVIRONMENT RESEARCH* [Preprint]. Available at: https://www.borenv.net/BER/archive/pdfs/ber7/ber7-273.pdf.

Suchara, I. (2019) ‘The Impact of Floods on the Structure and Functional Processes of Floodplain Ecosystems’, *Journal of Soil and Plant Biology*, pp. 28–44. Available at: https://doi.org/10.33513/JSPB/1801-03.

Teuling, A.J. *et al.* (2019) ‘Climate change, reforestation/afforestation, and urbanization impacts on evapotranspiration and streamflow in Europe’, *Hydrology and Earth System Sciences*, 23(9), pp. 3631–3652. Available at: https://doi.org/10.5194/hess-23-3631-2019.

Thodsen, H. (2007) ‘The influence of climate change on stream flow in Danish rivers’, *Journal of Hydrology*, 333(2–4), pp. 226–238. Available at: https://doi.org/10.1016/j.jhydrol.2006.08.012.

Totaro, V., Gioia, A. and Iacobellis, V. (2019) ‘Power of parametric and non-parametric tests for trend detection in annual maximum series’. Available at: https://doi.org/10.5194/hess-2019-363.

Van Roosmalen, L., Christensen, B.S.B. and Sonnenborg, T.O. (2007) ‘Regional Differences in Climate Change Impacts on Groundwater and Stream Discharge in Denmark’, *Vadose Zone Journal*, 6(3), pp. 554–571. Available at: https://doi.org/10.2136/vzj2006.0093.

Westerberg, I.K. *et al.* (2014) ‘Regional water balance modeling using flow-duration curves with observational uncertainties’, *Hydrology and Earth System Sciences*, 18(8), pp. 2993–3013. Available at: https://doi.org/10.5194/hess-18-2993-2014.