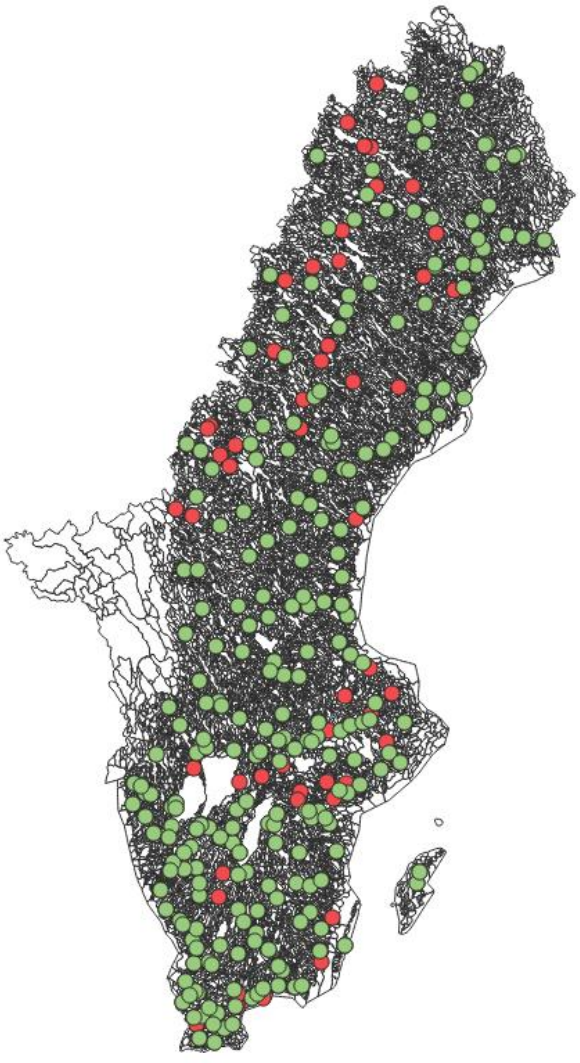
Master Thesis
TVVR 26/5019

Adjustment of peak flows in water courses without gauging from S-HYPE data

A comparison between observed and modelled flow in Sweden

Jacob Johnsson



Division of Water Resources Engineering
Department of Building and Environmental Technology
Lund University

Adjustment of peak flows in water courses without gauging from S-HYPE data

A comparison between observed and modelled flow in Sweden

By:
Jacob Johnsson

Master Thesis

Division of Water Resources Engineering
Department of Building & Environmental Technology
Lund University
Box 118
221 00 Lund, Sweden

Water Resources Engineering
TVVR-26/5019
ISSN 1101-9824

Lund 2026
www.tvrl.lth.se

Master Thesis
Division of Water Resources Engineering
Department of Building & Environmental Technology
Lund University

Swedish title:	Korrigering av toppflöden i omätta vattendrag med S-HYPE-data: Jämförelse mellan observerade och modellerade flöden i Sverige
English title:	Adjustment of peak flows in water courses without gauging from S-HYPE data: A comparison between observed and modelled flow in Sweden
Author(s):	Jacob Johnsson
Supervisor:	Magnus Persson Daniel Erdal
Examiner:	Linus Zhang
Language	English
Year:	2026
Keywords:	S-HYPE; Peak flow estimation; Flood frequency analysis; Random Forest; Multiple Linear Regression; Bias correction

Abstract

Extreme meteorological events are projected to increase in both frequency and intensity as climate change progresses, placing greater demands on accurate hydrological modelling for flood risk assessment and infrastructure design. In Sweden, the process-based S-HYPE model is widely used to simulate streamflow across approximately 26340 catchments, but like many regional rainfall-runoff models it systematically underestimates peak flows during extreme events, which poses a structural risk when designing culverts, retention basins, and urban drainage systems in ungauged catchments.

This study develops and evaluates a regression-based post-processing framework to quantify and correct this bias for small-to-medium Swedish watercourses with 100-year return period flows below 1000 m^3/s. Daily observed and S-HYPE-simulated streamflow data from 2010 to 2025 were compiled from an nationwide network of active gauging stations. Annual maximum daily flows were extracted and subjected to flood frequency analysis. The Generalized Extreme Value distribution fitted via L-moments (GEV-Lmom) was identified as the best framework, outperforming the Gumbel, Log-Normal, and Log-Pearson type III distributions across Anderson–Darling and Probability Plot Correlation Coefficient goodness-of-fit tests.

Baseline comparison confirmed that unadjusted S-HYPE outputs underestimate peak flows by a mean bias of -7.3%, despite maintaining strong spatial–temporal correlation ($R^2 = 0.964$). To correct this systematic underestimation while preserving the model's ranking performance, three correction configurations, Full, Reduced, and Flow–only, were evaluated using Multiple Linear Regression (MLR) and Random Forest (RF) algorithms across return periods from 2 to 100 years. Feature selection via permutation testing isolated five statistically significant physiographic predictors for the Reduced configuration: modelled flow, catchment area, upstream area, regulation fraction, and agricultural land–use fraction.

Both MLR and RF successfully reduced the baseline bias, reducing absolute peak flow errors to below $+1.4\%$ across all configurations. The RF model consistently achieved higher predictive accuracy ($R^2 = 0.990$) with tighter error distributions across all return periods, reflecting its capacity to capture non-linear catchment interactions. The log-linear MLR formulation, while slightly less accurate, produced an explicit and transferable scaling equation well suited for operational application at ungauged sites. The framework demonstrates that integrating statistical post-processing with physical catchment descriptors provides a scalable tool for improving extreme peak flow estimation across ungauged Swedish basins and while some uncertainties remain a future study could include a Peaks Over Threshold and a hold-out validation to improve on the results.

Preface

This master's thesis was carried out during the spring semester of 2026 at Lund University as part of the requirements for the degree of Master of Science in Water Resources Engineering. The work was conducted in collaboration with Tyréns and focuses on improving estimates of extreme flows in ungauged watercourses through statistical correction of S-HYPE model outputs.

I would like to express my gratitude to my supervisors, Magnus Persson and Daniel Erdal, for their valuable guidance, constructive feedback, and continuous support throughout this project. I am especially grateful to Daniel, who introduced me to the subject and made the thesis possible.

Finally, I would like to thank my family and friends for their encouragement and support during my studies and throughout the completion of this thesis.

Lund, June 2026

Jacob Johnsson

Abbreviation list

AD = Anderson–Darling
API = Application Programming Interface
CI = Confidence interval
FFA = Flood Frequency Analysis
GEV = Generalized Extreme Value
GUI = Graphical User Interface
HBV = Hydrologiska Byråns Vattenbalansavdelning
HYPE = Hydrological Predictions for the Environment
IQR = Interquartile Range
KGE = Kling–Gupta Efficiency
L–mom = L–moments
MLE = Maximum Likelihood Estimation
MLR = Multiple Linear Regression
NaN = Not a Number
NSE = Nash–Sutcliffe Efficiency
Pbias = Percent Bias
PDF = Probability Density Function
PPCC = Probability Plot Correlation Coefficient
RF = Random Forest
RP = Return Period
S–HYPE = Swedish implementation of the HYPE model
SMHI = Swedish Meteorological and Hydrological Institute

Table of contents

1. Introduction

1.1 Background and motivation

Extreme weather events are projected to increase in both frequency and intensity as climate change progresses[1]. In addition to mitigation measures, societies must adapt infrastructure, civil engineering design standards, and water management systems to cope with changing hydrological conditions and more severe precipitation events[2]. Hydrological modelling therefore plays a central role in understanding future flow conditions, anticipating extreme events, and delineating flood risk zones.

In Sweden, hydrological modelling has developed substantially over the past decades. One of the earliest operational models, the HBV model, was successfully implemented in 1972[3,4]. Despite initial criticism regarding its conceptual simplicity, HBV became widely adopted both nationally and internationally. Advances in computational capacity and increasing safety requirements for water resource management and hydropower infrastructure subsequently motivated the development of more physically based and spatially detailed models.

The continuous scaling up of infrastructure design constraints necessitated regional-scale tools capable of characterizing ungauged basins and providing continuous discharge simulations across vast, interconnected river networks. The Hydrological Predictions for the Environment (HYPE) model, developed by SMHI in the early 2000s, represents such an advancement[5,6]. The Swedish implementation, S-HYPE, applies the model across approximately 26 340 catchments covering Sweden, and parts of Norway and Finland. The model simulates water and nutrient transport across connected catchments, as seen in Figure 1.1, and has demonstrated good agreement with observed stream flow data under typical conditions.

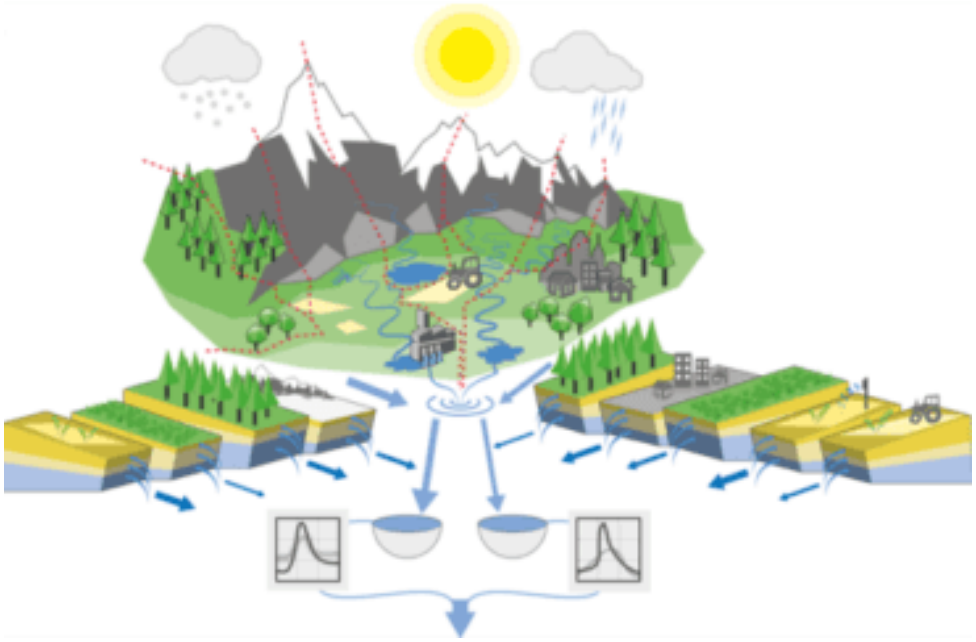


Figure 1.1: A schematic description of the HYPE model in a single catchment. Source: SMHI: About the HYPE model.

However, studies have shown that peak flows during extreme precipitation events are consistently underestimated by regional conceptual models[7,8]. This limitation is particularly critical for hydraulic infrastructure safety and flood risk assessment, motivating further investigation into model performance under extreme hydrological conditions. Because hydrological models are generally calibrated and validated on historical data the performance has started to deteriorate due to changing climate conditions, especially for peak flow[9,10].

Consequently, civil engineers cannot rely solely on raw uncorrected regional model outputs for local Flood Frequency Analysis (FFA). FFA traditionally relies on fitting extreme value probability distributions, such as the Generalized Extreme Value (GEV) or Generalized Pareto distribution, to observed annual maximum series using robust parameter estimation methods like L-moments[11]. At unmonitored or ungauged catchments, however, observed records do not exist, rendering localized empirical fitting impossible.

To resolve this dilemma, statistical post-processing framework regionalization can be employed. By pairing physical catchments that have long observed gauge records with their corresponding S-HYPE simulated outputs, a mathematical correction function can be established. By incorporating spatial covariates, such as upstream catchment area, land–use fractions, and regulation degrees, parametric statistical regressions or non-parametric machine learning models can learn to predict the systematic bias. This allows the structural underestimation of regional models to be corrected at any arbitrary, ungauged node across the nation, providing a scalable and hydraulically accurate tool for engineering design.

1.2 Research objectives and questions

The overarching goal of this study is to bridge the systematic gap between regional conceptual hydrological modelling outputs and actual observed localized hydraulic realities

under extreme flow conditions. Specifically, the objectives of this study are to:

1. Identify the most statistically robust extreme value distribution framework for modelling annual maximum streamflows across Swedish gauging stations.
2. Quantify the systematic underestimation bias of the S-HYPE framework across a broad range of return periods ($T = 2$ to 100 years).
3. Develop and cross-validate regionalized correction models that leverage catchment physiographic characteristics and anthropogenic factors to adjust modelled peak flows.

To satisfy these core objectives, the thesis addresses the following specific research questions:

- **RQ1:** Which extreme value probability distribution configuration offers the most mathematically robust framework for characterising annual maximum daily peak flows across diverse Swedish catchments?
- **RQ2:** To what extent can parametric log-linear Multiple Linear Regression (MLR) and non-parametric Random Forest (RF) frameworks resolve the structural underestimation of S-HYPE peak flow return levels?
- **RQ3:** How do geomorphological parameters (e.g., catchment area, forest fraction, lake fraction) and anthropogenic impacts (e.g., regulation degree) influence the magnitude of the model bias, and which serve as the primary predictors in the reduced regionalization formula?

1.3 Thesis outline

The remainder of this thesis is structured as follows: Chapter 2 summarizes the theoretical foundations of FFA, extreme value theory, and the statistical algorithms evaluated. Chapter 3 outlines the data acquisition process, station screening criteria, and spatial join mechanics executed within QGIS. Chapter 4 presents the methodological layout governing the empirical fitting, feature elimination routines, and k-fold cross-validation loops. Chapter 5 displays the results of the distribution rankings, baseline bias profiles, and corrected regression models. Chapter 6 provides a discussion regarding the physical meaning of the selected predictors, structural limitations, and practical engineering trade-offs. Finally, Chapter 7 provides concluding remarks and operational recommendations for future hydrologic adjustments.

2. Literature review

2.1 Hydrological modelling in Sweden

Hydrological simulation in Sweden transitioned from lumped, conceptual conceptualizations toward highly semi-distributed, process-based setups over the last fifty years [4,5]. The foundational benchmark established by the conceptual HBV model relied on generalized sub-basin routing variables to compute runoff [4]. While computationally efficient, lumped conceptual configurations struggle to isolate local spatial variations in soil composition, localized rainfall distribution, and land cover deviations [6].

To overcome these spatial constraints and provide high-resolution nutrient and water tracking across the nation's hydrographic networks, SMHI engineered the HYPE framework. S-HYPE discretizes large river basins into highly structured sub-basins, explicitly accounting for spatial heterogeneity.

2.2 S-HYPE model overview

As structurally outlined in Figure 1.1, S-HYPE operates on a daily time step, establishing land use and soil layers to calculate e.g. vertical water profiles, macro-pore infiltration, snowmelt dynamics, and evapotranspiration matrices [12]. Despite its widespread regulatory use, process-based rainfall-runoff configurations frequently smooth hydrograph responses, leading to a systematic underestimation of peak flows during extreme precipitation events [7]. This structural smoothing behaviour likely stems from the spatial-temporal averaging of rainfall inputs over expansive sub-basins and the volume-conservation constraints inherent to larger soil reservoir storages.

2.3 Previous studies on peak flow estimation

To quantify these extremes, hydrological practice frequently relies on fitting probability density functions (PDFs) to annual maximum flows. Because the tail behaviour of these distributions varies significantly based on the chosen mathematical framework, selecting a PDF that accurately fits the empirical data is essential for robust risk estimation.

In a study by Coronado-Hernández et al., four different distributions: Generalized Extreme Value (GEV), Gumbel, Normal, and Pearson III were fitted and adjusted with maximum likelihood and method of moments to find the best fitting PDF for streams in Colombia [13]. This methodology was adopted and adapted for this thesis, examining the structural performance of the GEV, Gumbel, Log-Normal, and Log-Pearson type III distributions.

2.3.1 Generalized Extreme Value distribution

The GEV distribution unifies the three extreme value distributions into one general framework, which provides an increased adaptability to the data [14,15]. The combined distributions are the Gumbel, Fréchet, and Weibull, extreme value distribution I, II, and III

respectively. The shape parameter κ determines which type is represented: $\kappa = 0$ yields the Gumbel, $\kappa > 0$ the Fréchet, and $\kappa < 0$ the Weibull distribution.

$$y = f(x|\kappa, \mu, \sigma) = \frac{1}{\sigma} \exp\left(- \left(1 + \kappa \frac{x - \mu}{\sigma} \right)^{-\frac{1}{\kappa}} \right) \left(1 + \kappa \frac{x - \mu}{\sigma} \right)^{-1-\frac{1}{\kappa}} \tag{2.1}$$

for

$$1 + \kappa \frac{x - \mu}{\sigma} > 0 \tag{2.2}$$

where x is the observed value, κ is the shape parameter, μ is the location parameter, and σ is the scale parameter.

2.3.2 Gumbel distribution

The Gumbel distribution is, as mentioned, a special case of the GEV distribution where $\kappa = 0$, and it approximates the GEV well when $|\kappa| < 0.3$ [14,15]. Due to its simplicity and simple parameter estimation, it has been widely used in flood frequency analysis [16].

$$y = f(x|0, \mu, \sigma) = \frac{1}{\sigma} \exp\left(- \frac{x - \mu}{\sigma} - \exp\left(- \frac{x - \mu}{\sigma} \right) \right) \tag{2.3}$$

where x is the observed value, μ is the location parameter, and σ is the scale parameter.

2.3.3 Log-Normal distribution

The two-parameter log-normal distribution assumes that the logarithm of the variables follows a normal distribution [14,15]. It is commonly applied in hydrology because streamflow data are strictly positive and often positively skewed.

$$y = f(x|\mu, \sigma) = \frac{1}{x\sigma\sqrt{2\pi}} \exp\left(\frac{-(\ln x - \mu)^2}{2\sigma^2} \right) \tag{2.4}$$

where x is the observed value, μ is the mean of $\ln x$, and σ is the standard deviation of $\ln x$.

2.3.4 Log-Pearson type III distribution

The Log-Pearson type III distribution is fitted to the logarithms of the observed data using a three-parameter gamma distribution, where κ, μ, and σ represent the shape, scale, and location, respectively [14,15].

$$y = f(x|\kappa, \mu, \sigma) = \frac{1}{x\mu\,\Gamma(\kappa)} \left(\frac{\ln x - \sigma}{\mu} \right)^{\kappa-1} \exp\left(- \frac{\ln x - \sigma}{\mu} \right), \quad x > 0 \tag{2.5}$$

$$\Gamma(\kappa) = \int_0^\infty e^{-t} t^{\kappa-1} dt \tag{2.6}$$

where x is the observed value, κ is the shape parameter, μ is the scale parameter, σ is the location parameter, and $\Gamma(\kappa)$ is the gamma function evaluated at κ.

3. Data and study area

3.1 Data sources and quality control

All data used in this study were downloaded from SMHI via public application programming interfaces (APIs) and graphical user interfaces (GUIs). Raw streamflow gauge readings and regional hydrological simulation outputs were cross-referenced to compile the core dataset.

To ensure statistical consistency and filter out lower-quality records, a three-tiered quality control sequence was enforced. Selected stations had to satisfy three criteria:

1. The observation station had to be active at the time of data download.
2. Continuous daily observations had to be fully available for the period 2010–2025 to match the temporal window of the model.
3. A corresponding, operational S-HYPE dataset must encompass the targeted station coordinate.

3.2 Catchment selection criteria

Station coordinates were matched to S-HYPE subcatchments by intersecting a point layer of station locations with a polygon layer of subcatchments in the software QGIS. Because spatial coordinates of physical river gauges do not always perfectly align with digital elevation model stream networks due to grid resolution or database limitations, a robust hydrological validation framework was established. This framework verified the baseline correspondence between daily observed and simulated time series under typical flow conditions before any FFA was conducted.

To quantify this baseline agreement, three distinct hydrological metrics were evaluated simultaneously across all candidate stations: percent bias (P_{bias}), Pearson's correlation coefficient (r), and the Kling–Gupta Efficiency (KGE). Percent bias measures the average percentage tendency of the simulated data to be larger or smaller than their observed counterparts. Mathematically defined as $P_{\text{bias}} = [\sum(S_i - O_i)/\sum O_i] \times 100$, where O_i and S_i represent the observed and simulated daily flows, it acts as a primary metric for identifying systemic volume errors.

The Pearson correlation coefficient measures the linear strength and temporal synchronicity between the observed and simulated hydrographs. It is mathematically defined as:

$$r = \frac{\sum_{i=1}^{n}(O_i - \bar{O})(S_i - \bar{S})}{\sqrt{\sum_{i=1}^{n}(O_i - \bar{O})^2 \sum_{i=1}^{n}(S_i - \bar{S})^2}} \tag{3.1}$$

where n represents the total number of paired daily discharge observations, O_i is the observed streamflow on day i, and S_i is the corresponding S-HYPE simulated streamflow. The terms $\bar{O}$ and $\bar{S}$ denote the calculated temporal means of the observed and simulated time series over the entire evaluation period, respectively.

To provide a comprehensive, balanced assessment of overall hydrograph behaviour, KGE was utilized to overcome peak-flow inflation biases, associated with Nash–Sutcliffe

Efficiency (NSE)[17]. The KGE integrates three independent aspects of model performance by computing the Euclidean distance from an ideal simulation point in a three-dimensional criteria space:

$$KGE = 1 - \sqrt{(r-1)^2 + (\alpha-1)^2 + (\beta-1)^2} \tag{3.2}$$

where r is the correlation coefficient, $\alpha = \sigma_s/\sigma_o$ represents the variability ratio (the ratio of simulated to observed standard deviations), and $\beta = \mu_s/\mu_o$ represents the bias ratio (the ratio of simulated to observed means).

Strict acceptance thresholds were enforced based on established literature guidelines for daily hydrological modelling. A station was deemed acceptable only if its baseline daily series yielded a percent bias within $\pm 20\%$, a correlation coefficient $r \geq 0.5$, and an integrated $KGE \geq 0.5$[18]. A P_{bias} within $\pm 20\%$ ensures that total catchment water balance errors are sufficiently constrained. Meanwhile, thresholds of 0.5 for both r and KGE ensure that the model behaves with meaningful predictive power over a benchmark based purely on historical observed mean flow[19]. Stations failing one or more of these criteria were automatically flagged for further spatial inspection.

For these flagged stations, modelled data from all subcatchments within a 1 km geographical radius were extracted and re-evaluated using the identical validation triad. This step corrected for instances where a physical gauge was erroneously snapped to an adjacent or minor tributary in the spatial database instead of the main stem. The subcatchment yielding the highest baseline agreement was selected as the true hydrological match. Stations with no satisfactory match within this radius were deemed unrepresentative of the local flow realities and excluded from the study.

Furthermore, because the ultimate design objective of this study is to derive reliable predictive adjustments for unmonitored small-to-mid-sized headwater basins, a final scale-based filter was implemented. Any station whose calculated 100-year return period peak flow (Q_{100}) exceeded an upper threshold of 1000 m^3/s was filtered out. While arbitrary, this constraint effectively prevented massive, highly regulated regional river systems—whose extreme peak flows are heavily modulated by major reservoir operations and complex upstream routing from overly distorting and skewing the regression models. Note that these stations were included during the distribution calibration and selection, and were excluded for the regressions.

3.3 Description of study area

The resulting network of valid stations spans across Sweden's diverse hydroclimatic zones, stretching from the cold, snow-dominated alpine catchments in the north to the flatter, temperate basins in the south. The structural distribution covers a balanced mix of undisturbed natural headwaters, high-density forested subcatchments, and agricultural zones impacted by intensive tile drainage.

In Figure 3.1, all subcatchments are displayed with the locations of the downloaded stations. In total 298 stations were downloaded, of which 246 stations were used for the analysis.

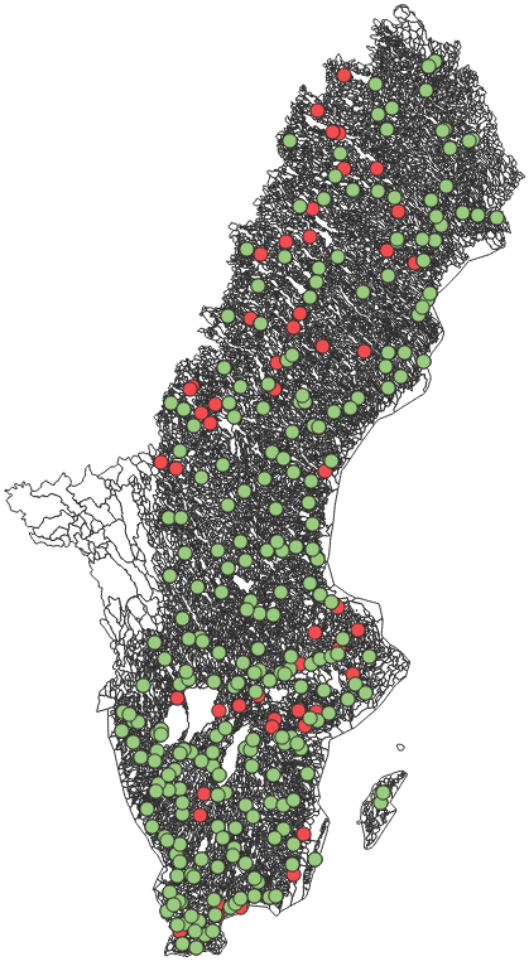


Figure 3.1: Overview of all S-HYPE catchments with dots representing selected stations. Green stations passed all tests and were used for the distribution selection, whereas red dots fulfilled the three criteria but did not have a matching subcatchment.

After quality control, annual maximum daily flows were extracted from both observed and modelled datasets according to the structural categories detailed in Table 3.1.

Table 3.1: Dataset categories extracted from SMHI data and used in the analysis.

Dataset	**Description**
Observed Full	Continuous observed daily streamflow spanning all complete historical calendar years available per station. Utilized for extreme value distribution calibration and candidate selection to maximize fitting sample sizes.
Observed Limited	Observed streamflow records truncated to 2010–2025. Used for analysis against corresponding simulated flows.
Modelled Normal	S-HYPE simulated daily streamflow for the 2010–2025 period under standard operational configurations. Represents the primary baseline simulated dataset evaluated throughout the core analysis. Originally designated as *Total Flow* within the SMHI database but renamed here for clarity.
Modelled Statistical	S-HYPE simulated daily flow (2010–2025) structurally adjusted by SMHI using localized observed data. Because this interpolation requires local physical gauges, it is omitted from the regionalization analysis and restricted to comparative evaluations in Appendix C.
Modelled Natural	Estimated natural daily streamflow (2010–2025) simulated by S-HYPE, where human interventions (e.g. dam operations, reservoir storage, etc.) are mathematically removed to recreate unregulated conditions. Omitted from the main regionalization models but included in Appendix C.

4. Methodology

4.1 Extraction of annual maximum flows

The empirical analysis began by analysing the continuous, daily time series arrays from 2010–2025 to isolate the single highest peak flow event recorded within each operational calendar year. For the majority of high-latitude Swedish stations, this annual maximum was driven primarily by spring snowmelt events, while smaller agricultural catchments in the south frequently recorded autumn and winter peaks, likely caused by low-velocity frontal systems.

4.2 Flood frequency analysis and parameter estimation

The parameters defining each candidate probability distribution must be estimated to achieve the best fit to the full observed dataset. Before executing these parameter estimation algorithms, a robust framework was established to quantify the underlying sampling uncertainty of the fitted extreme value distribution curves. Two primary methods were applied for the central fitting, while a statistical resampling technique was used to estimate uncertainty.

4.2.1 Uncertainty estimation via parametric bootstrapping

To evaluate the statistical stability of the distribution fits and derive robust empirical confidence intervals for the return level curves, a parametric bootstrapping routine was executed[20,21]. Unlike non-parametric resampling which redraws elements directly from historical records, a parametric bootstrap utilizes the initial fitted distribution parameters (θ) to generate entirely new synthetic realizations of the annual maximum series. Given an original sample of length N at a specific gauging station, a synthetic sample of identical size N was simulated in each iteration using the continuous random variable sampler corresponding to the candidate distribution family ($X_{\text{synth}} \sim f(x \mid \theta)$). For log-transformed distributions, such as the Log-Pearson Type III, synthetic series were drawn natively in base-10 logarithmic space and subsequently back-transformed into original flow units.

This synthetic sampling loop was executed for $B = 500$ replicates. For each simulated realization, the candidate distribution was refitted using either Maximum Likelihood Estimation or the Method of L-Moments to yield a vector of bootstrap parameters (θ_b^*). To prevent unphysical tail explosions or numerical anomalies typical of small sample sizes, a rigorous parameter plausibility filter was applied to the refitted vectors. A bootstrap replicate was structurally rejected from the matrix if parameter optimization fails to converge, if any calculated return level was non-finite, or if shape-defining parameters violate hydrological bounds. Specifically, a replicate was omitted if the Generalized Extreme Value shape parameter satisfies $|\xi| > 0.5$, marking the threshold where variance becomes infinite or the upper bound truncates observed realities, if the Log-Pearson Type III absolute skewness exceeds 4.0, or if the Log-normal scale-free standard deviation drops below zero or exceeds 3.0[11].

If a station retained fewer than 10 valid, unrejected replicates across the simulation loop, the confidence intervals was deemed numerically unstable and assigned as non-numeric values (NaN). For robust stations, the compiled return level iterations form an empirical bootstrap matrix. The 5-th and 95-th percentiles extracted along the axis of each targeted return period, T, defined the lower and upper bounds of a 90% confidence interval, providing an uncertainty range.

4.2.2 Maximum Likelihood Estimation

Maximum likelihood estimation (MLE) determines the parameter vector $\hat{\theta}$ that maximises the likelihood function, i.e., the probability of observing the given data under the assumed distribution[22].

$$L(\theta \mid \mathbf{x}) = \prod_{i=1}^{n} f(x_i \mid \theta) \tag{4.1}$$

where n is the number of observations, x_i is the i-th observed value, and $f(x_i \mid \theta)$ is the probability density evaluated at x_i for parameter vector θ. In practice, it is equivalent and numerically preferable to maximise the log-likelihood:

$$\ell(\theta) = \sum_{i=1}^{n} \ln f(x_i \mid \theta) \tag{4.2}$$

The MLE $\hat{\theta}$ is then obtained as $\hat{\theta} = \arg\max_{\theta} \ell(\theta)$.

4.2.3 L-Moments method

L-moments are linear combinations of order statistics and provide robust parameter estimates, particularly for heavy-tailed distributions and small sample sizes[23]. The r-th L-moment is defined as:

$$\lambda_r = \int_0^1 Q(u)\, P^*_{r-1}(u)\, du \tag{4.3}$$

where $u = F(x) \in [0, 1]$ is the non-exceedance probability, $Q(u) = F^{-1}(u)$ is the quantile function, and $P^*_{r-1}(u) = (r-1)$-th shifted Legendre polynomial. Equivalently, via probability-weighted moments $\beta_r = E[X \cdot F(X)^r]$:

$$\lambda_1 = \beta_0 \tag{4.4}$$
$$\lambda_2 = 2\beta_1 - \beta_0 \tag{4.5}$$
$$\lambda_3 = 6\beta_2 - 6\beta_1 + \beta_0 \tag{4.6}$$
$$\lambda_4 = 20\beta_3 - 30\beta_2 + 12\beta_1 - \beta_0 \tag{4.7}$$

The scale-free L-moment ratios are defined as $\tau_r = \lambda_r/\lambda_2$ for $r \geq 3$, yielding L-skewness (τ_3) and L-kurtosis (τ_4), where $-1 < \tau_4 < 1$.

4.2.4 Goodness-of-Fit testing

The distribution selection was based on the full observed dataset, evaluated through two parallel goodness-of-fit frameworks:

- **Anderson–Darling (AD) Test:** Assigns higher mathematical weights to the tails of distributions, rendering it ideal for extreme value evaluations[15,24].
- **Probability Plot Correlation Coefficient (PPCC) Test:** Measures the linear correlation between ordered empirical points and theoretical quantile locations[25,26].

4.3 Regression-based correction frameworks

Following candidate distribution identification, regression structures were constructed to quantify correction mappings between simulated S-HYPE flow points and geographical features. Return periods (RPs) $T = 2, 5, 10, 20, 50$, and 100 years were compiled from the limited flow data sets for the analysis. The one-year RP was excluded because fitting distributions to annual maximum series is mathematically unreliable at such short RPs and corresponds to the lower bound of the annual-maximum distribution[27,28].

Three distinct structural configurations were compiled across both statistical backends:

1. **Full Model:** Includes all eleven extracted physiographical and raw flow parameters.
2. **Reduced Model:** Eliminates non-performing predictors to optimize structural simplicity.
3. **Flow-Only Model:** Relies strictly on the raw modelled flow output.

4.3.1 Multiple Linear Regression

The MLR model utilized a log-linear transformation to preserve positive flow constraints and stabilize variance across disparate scale horizons[11]:

$$\log(Q_{\text{obs}}) = \beta_0 + \beta_1 \log(Q_{\text{mod}}) + \beta_2 \log(A_c) + \beta_3 \log(A_u) + \cdots + \epsilon \tag{4.8}$$

Percentage-based parameters (e.g., land use fractions and regulation factors) entered the formulation linearly, resulting in the following exponential multiplicative adjustment structure[29]:

$$Q_{\text{obs}} = \exp(\beta_0) \cdot Q_{\text{mod}}^{\beta_1} \cdot A_c^{\beta_2} \cdot A_u^{\beta_3} \cdot \exp(\beta_4 \cdot \text{R}) \cdots \tag{4.9}$$

where Q_{obs} is observed flow, Q_{mod} is modelled flow, A_c is catchment area, A_u is upstream area, R is regulation percentage, and β_i represent the fitted coefficients.

4.3.2 Random Forest

As a non-parametric alternative to the linear models, a Random Forest (RF) ensemble machine learning algorithm was evaluated[30]. The RF framework constructs a multitude of individual decision trees trained on bootstrapped data matrices, ultimately averaging their independent outputs to counteract local overfitting and reduce predictive variance. Because this non-linear mapping operates without predefined functional or log-linear constraints, the ensemble structure is inherently capable of capturing complex, high-order interactions among catchment variables automatically.

This non-parametric nature fundamentally alters the data preprocessing requirements compared to the parametric regression baseline. While the MLR architecture strictly dictates logarithmic transformations to stabilize variance and prevent scale distortions

across continuous variables, the RF algorithm requires no prior feature scaling or attribute normalization[31]. Because the foundational decision trees within the ensemble isolate split thresholds based on coordinate-wise, orthogonal partitioning rules, the optimization criteria rely solely on the rank-order statistics of the data rather than their absolute numerical magnitudes[32]. Consequently, the RF model allows raw, native-unit parameters to be used without introducing scaling biases.

4.4 Feature selection process

To refine the full parameter list detailed in Table 4.1, variables with weak statistical significance were iteratively dropped using a permutation process. Covariates exhibiting an empirical p-value greater than the target significance threshold ($\alpha = 0.05$) were sequentially removed[33]. This sequence was repeated until the variable pool stabilized on the central estimate.

Table 4.1: Parameters used as input variables in the regression analysis. All catchment information was obtained from S-HYPE data provided by SMHI.

Parameter	**Description**
Modelled flow	Included as the primary predictor to be corrected by the regression model.
Catchment area	Influences the total volume of water contributing to streamflow.
Upstream area	Larger upstream areas contribute greater cumulative volumes through the catchment.
Regulation	Regulated watercourses exhibit altered flow signatures compared to natural systems.
Land use	Fractions for agriculture, clear-cut areas, forest, mountain, urban areas, water bodies, and wetlands.

4.5 Model validation and cross-validation approach

To guarantee generalized structural performance on ungauged locations, a repeated k-fold cross-validation was performed[32]. k-fold cross validation is particularly beneficial to use with small-to-mid-sized data sets as it provides sufficient training and testing of the model, where there otherwise would be an insufficient number of data points[31]. The complete data block was randomly and partitioned into $k = 4$ folds of equal size. Three of these folds served as training arrays while the remaining segment was held back for testing, as outlined in Figure 4.1. This was repeated 20 times with a new random split each iteration. Performance was quantified using the coefficient of determination (R^2) and is discussed.

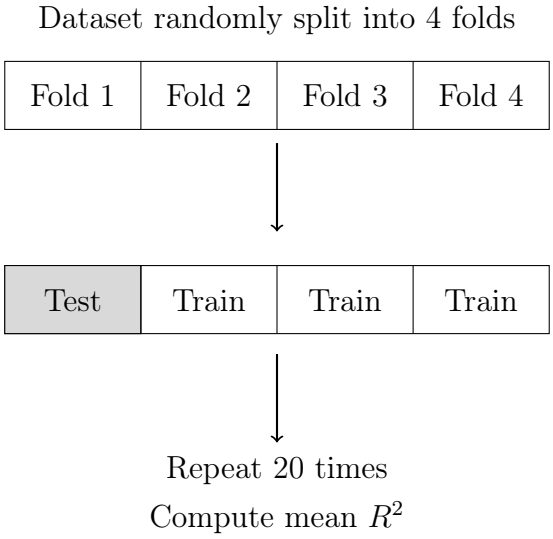


Figure 4.1: Schematic of the repeated 4-fold cross-validation procedure used for model evaluation.

4.6 Clarification regarding confidence intervals

It was decided early on during the project that a CI of 90 % would be used for the RPs to give a good estimation of the range the flow can have for a certain RP in a catchment, without having excessive ranges. This CI extracted from the bootstrapping method, as described in Section 4.2.1, and used as the lower and upper inputs to the regression analysis. However, the functions in the numpy and scipy Python packages used for the analysis use a 95 % CI as a standard, which was not changed. Thus, there are two ranges for CI used in the thesis. All analysis of regression performance used a CI of 95 % and the RPs used as input for the regressions had a 90 % CI.

5. Results

5.1 Performance of S-HYPE model

A probability distribution was selected using the Anderson-Darling (AD) and Probability Plot Correlation Coefficient (PPCC) tests described previously. The results are shown in Figure 5.1, where the Generalized Extreme Value distribution fitted with L-moments (GEV-Lmom) provided the best overall performance for both tests, yielding 45.9% and 30.5% of the best fits for the AD and PPCC tests, respectively. Since the distribution was fitted to each station individually, the parameters $\kappa, \mu,$ and σ will differ for each fit and the values of these parameters were therefore not extracted. GEV-Lmom was utilized in all subsequent analyses.

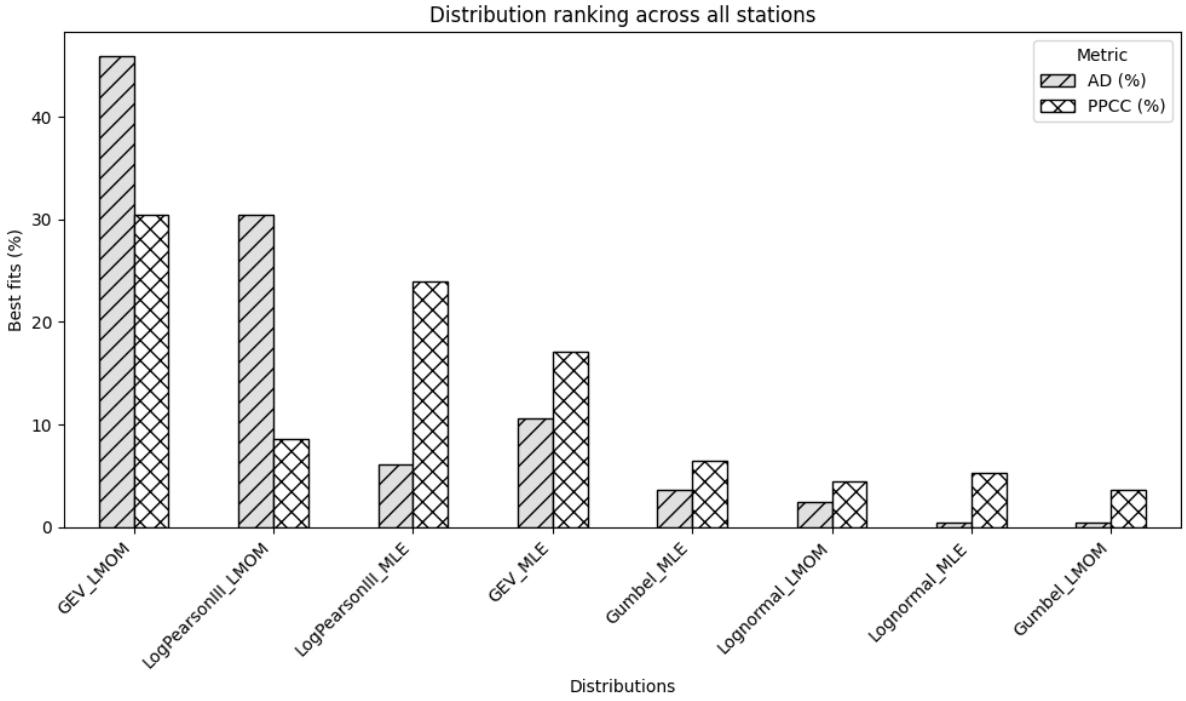


Figure 5.1: Total share of best fits each distribution achieved with the full observed data series when performing AD and PPCC tests across all stations.

A clear trend indicates that S-HYPE systematically underestimates peak flows. As illustrated in Figure 5.2, 75% of the uncorrected data points fall above the 1:1 threshold line for normal flows. In Appendix B Figure 9.1, a multi-panel overview show the 1:1 plot split by return period, with CIs for both modelled and observed peak flows. It becomes evident that the accuracy of the distribution decreases rapidly after the 20–year RP. The model also underestimates less, based on the central point estimates, which could be explained by the GEV distribution may fit the observed and modelled flows differently and in turn causing return period flows to grow at different rates.

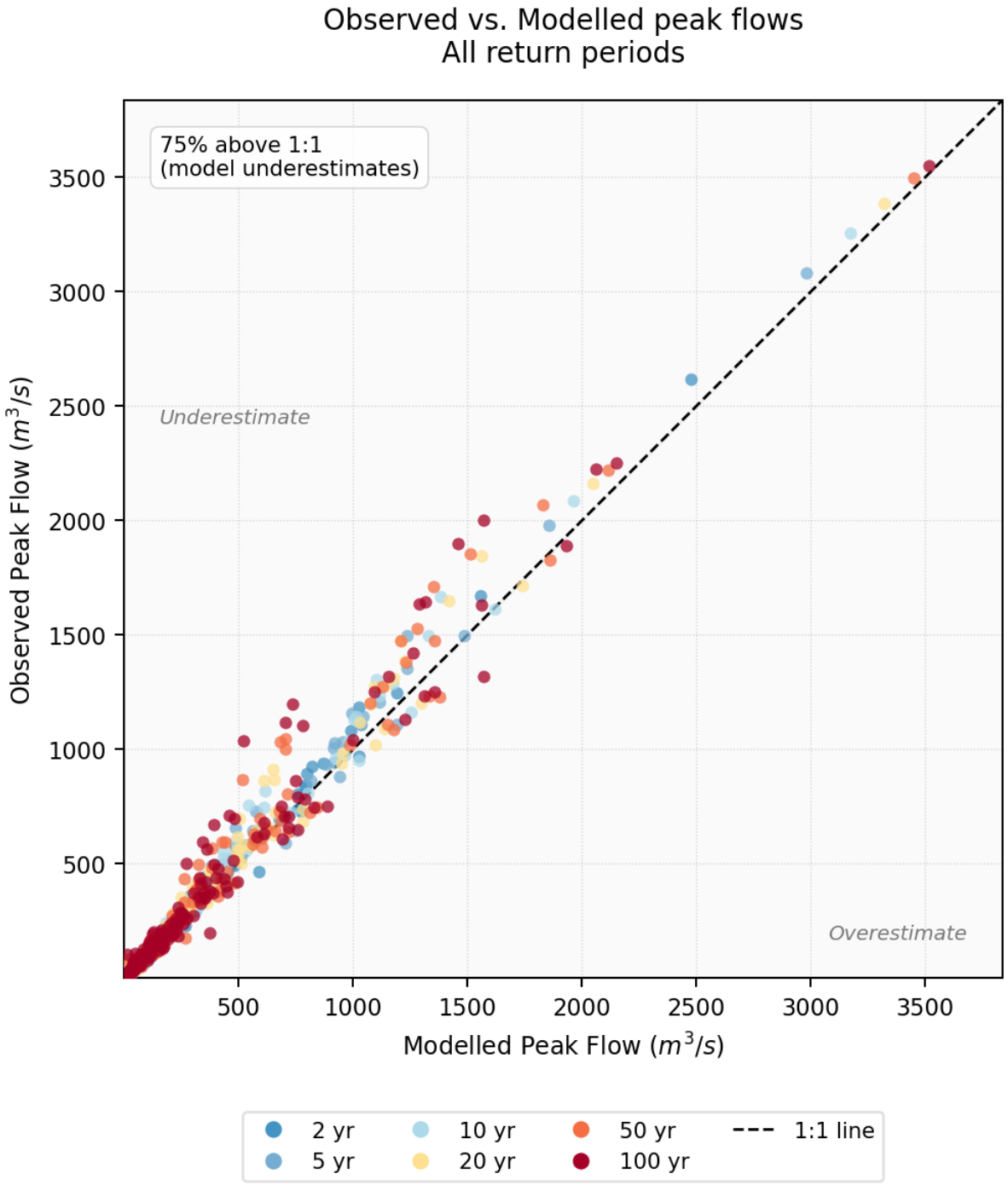


Figure 5.2: Observed versus modelled peak flows for the normal flow scenario, combining all target return periods ($T = 2$ to 100 years) into a single 1:1 validation plot. The ideal match is indicated by the dashed diagonal 1:1 line; data points falling above this line reveal that observed flows exceed modelled values, illustrating that the uncorrected S-HYPE model systematically underestimates peak flows across approximately 75% of all station records (based on central point estimates).

5.2 Regression model outcomes

To evaluate how catchment characteristics influence this relationship, regressions were performed for the three distinct model configurations. The resulting parameter weights for the Full configuration are illustrated in Figure 5.3. In this figure, there are several parameters which has a very large CI, most of which does not show any correlation with the predicted flow according to the permutation test ($p \geq 0.05$) on the central estimate.

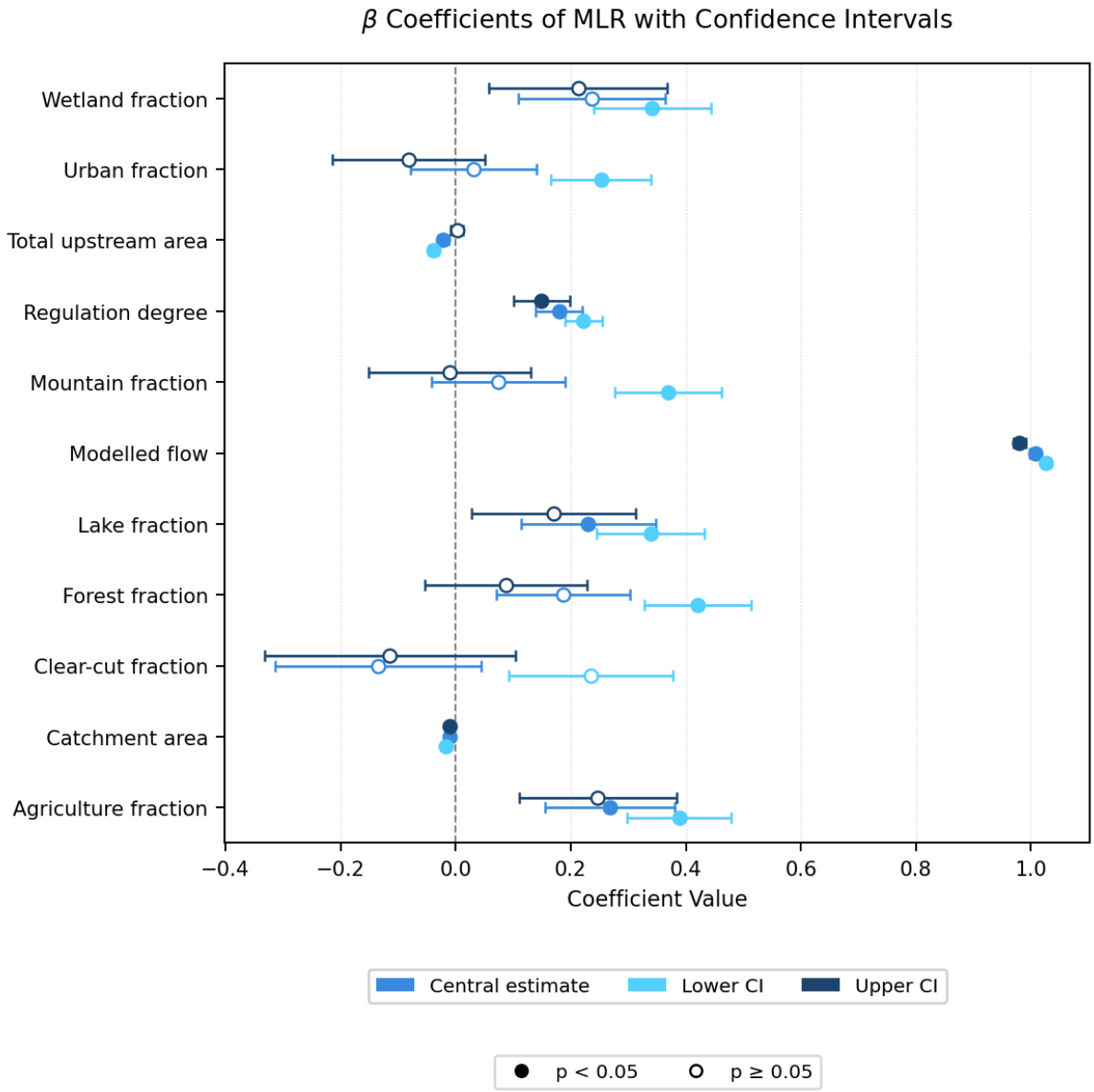


Figure 5.3: Estimated β coefficients for the eleven Multiple Linear Regression (MLR) catchment parameters following the initial regression analysis under the normal flow scenario. Central markers indicate coefficient point estimates; filled circles denote statistically significant predictors ($p < 0.05$) and open circles denote non-significant ones determined by a permutation test. Error bars represent the 95% confidence intervals. Large confidence intervals that widely cross the zero line visually highlight parameters with weak, highly uncertain correlations to the predicted flow, serving as candidates for feature removal.

Seven of the original eleven parameters exhibited weak correlations with the predicted flow on the central estimate and were subsequently excluded from the reduced model during the feature selection. The remaining five β coefficients are displayed in Figure 5.4 and are: modelled flow, catchment area, upstream area, regulation fraction, and agricultural land use. Although minor shifts in coefficient magnitudes occurred, the remaining parameters retained stable relationships with the central estimate. No additional parameters were removed after this step and this was regarded as the reduced model.

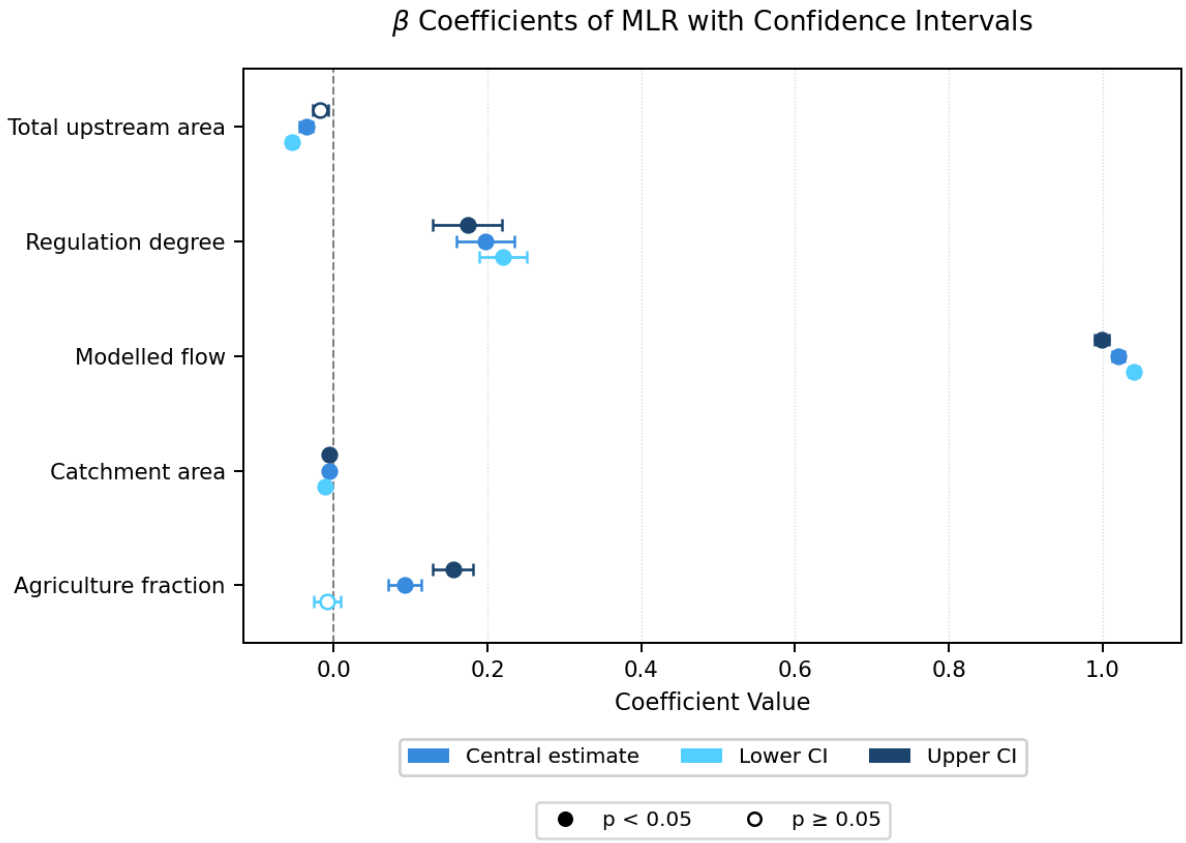


Figure 5.4: Estimated β coefficients for the five remaining parameters (modelled flow, catchment area, upstream area, regulation fraction, and agricultural land use fraction) in the reduced MLR model following variable selection. Central markers denote point estimates (filled circles for $p < 0.05$, open circles for non-significant), and error bars represent 95% confidence intervals. The narrow, stable confidence bounds across return periods demonstrate that these primary predictors maintain a uniform, scale-invariant correction mechanism with low structural uncertainty.

The numerical stability of the β-weights across the varying target return periods indicates that the underlying log-linear correction mechanism acts uniformly across scales, maintaining tight, non-diverging bounds for the primary predictors. In Table 5.1 the values of the reduced models β–coefficients are displayed and can be used with Equation 4.9 and Monte Carlo simulation to calculate a predicted observed flow in a catchment. For the remaining analysis the central estimates β-values was used as a benchmark evaluation.

Table 5.1: Regression coefficients values (β) and associated standard errors (σ) for the parameters of the reduced MLR model. Columns represent separate regression calibrations fitted using the lower uncertainty bounds, central point estimates, and upper uncertainty bounds of the peak flow return levels, respectively.

	Lower		Central		Upper	
Parameter	β	σ	β	σ	β	σ
Constant	0.28773	0.02440	0.22296	0.02820	0.18280	0.03222
Total upstream area	-0.05464	0.00707	-0.03596	0.00842	-0.01748	0.00974
Regulation degree	0.22040	0.03072	0.19706	0.03749	0.17393	0.04516
Modelled flow	1.04060	0.00665	1.02016	0.00800	0.99873	0.00926
Catchment area	-0.01081	0.00220	-0.00649	0.00273	-0.00664	0.00331
Agriculture fraction	-0.00913	0.01733	0.09196	0.02164	0.15498	0.02632

5.3 Bias correction effectiveness

The application of both the MLR and RF frameworks successfully eliminated the S-HYPE's trend toward underestimation. In contrast to the unadjusted baseline, the statistically corrected flow performs considerably better, aligning closer to the 1:1 line as seen in Figure 5.5 for the reduced model.

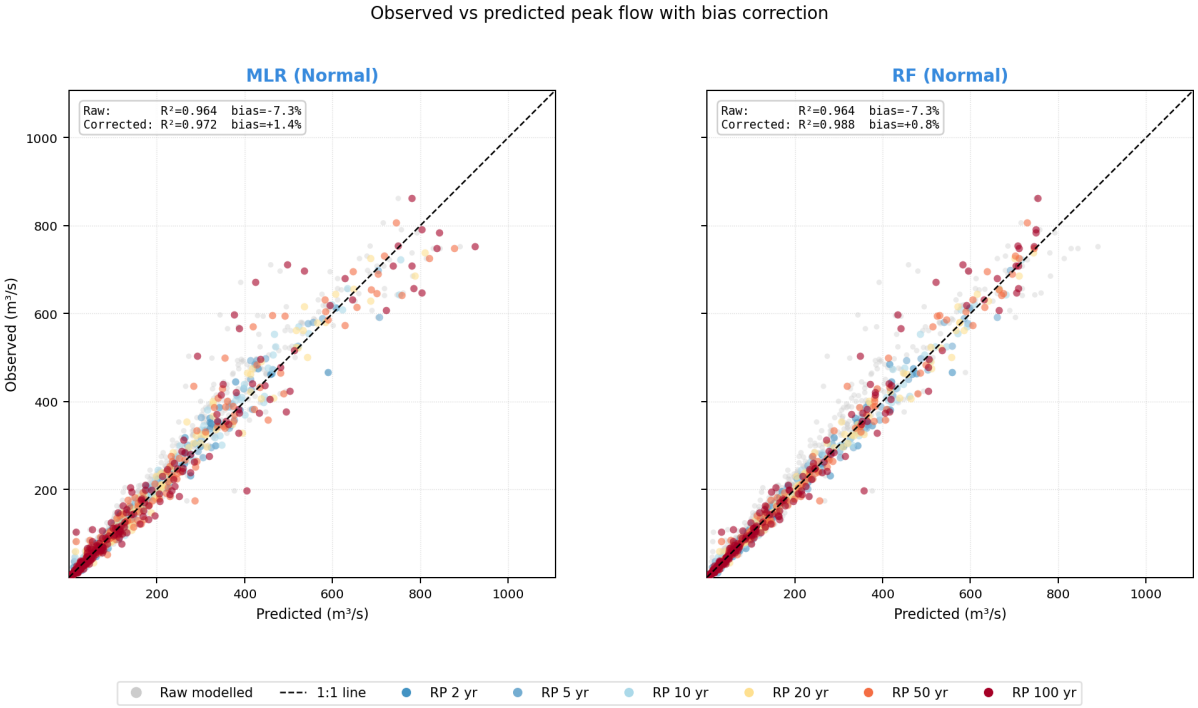


Figure 5.5: Observed versus predicted peak flows (Q) comparing the uncorrected S-HYPE baseline against the corrected MLR and Random Forest (RF) reduced frameworks across target return periods (T = 2 to 100 years). Light grey background points show the uncorrected baseline, which systematically trends toward underestimation. The overlaid coloured markers show that both correction frameworks successfully eliminate this systemic underestimation, shifting the data tightly onto the dashed 1:1 reference line to yield high R^2 values and minimal relative bias (as detailed in the embedded performance metric boxes).

The mean relative bias decreased significantly from a baseline underestimation to within ±2% across the validation stations, confirming that incorporating catchment area, upstream boundaries, regulation percentages, and targeted agricultural land-use factors effectively standardizes the simulated output. In Table 5.2 the R^2 and bias is shown for all three model configurations, as well as the modelled flows original fit. Across all configurations both metrics significantly improved, especially for RF, which performed very well with the full model. It is expected since more parameters makes the decision tree make more founded choices.

MLR was surprisingly close to RF in ability to predict the flow and performed the best with the reduced model. However, when taking CI into account, RF is clearly better as seen in Figure 5.6. For all sampled return periods RF has a smaller interquartile range (IQR) and shorter whiskers, making the predicted flow more likely to be correct. The median relative bias is close to 0% for both models, which indicates that the models both work very well if the uncertainties can be minimized.

Table 5.2: R^2 coefficients and percentage biases categorized by regression configuration. Values gathered from Figures 5.5, 9.3 and 9.4.

Configuration	**Method**	Normal R^2	Normal **Bias [%]**
Reduced model	MLR	0.972	+1.4
	RF	0.988	+0.8
Flow-only model	MLR	0.969	+1.4
	RF	0.980	+1.1
Full model	MLR	0.971	+1.4
	RF	0.990	+0.6
Modelled flow	–	0.964	-7.3

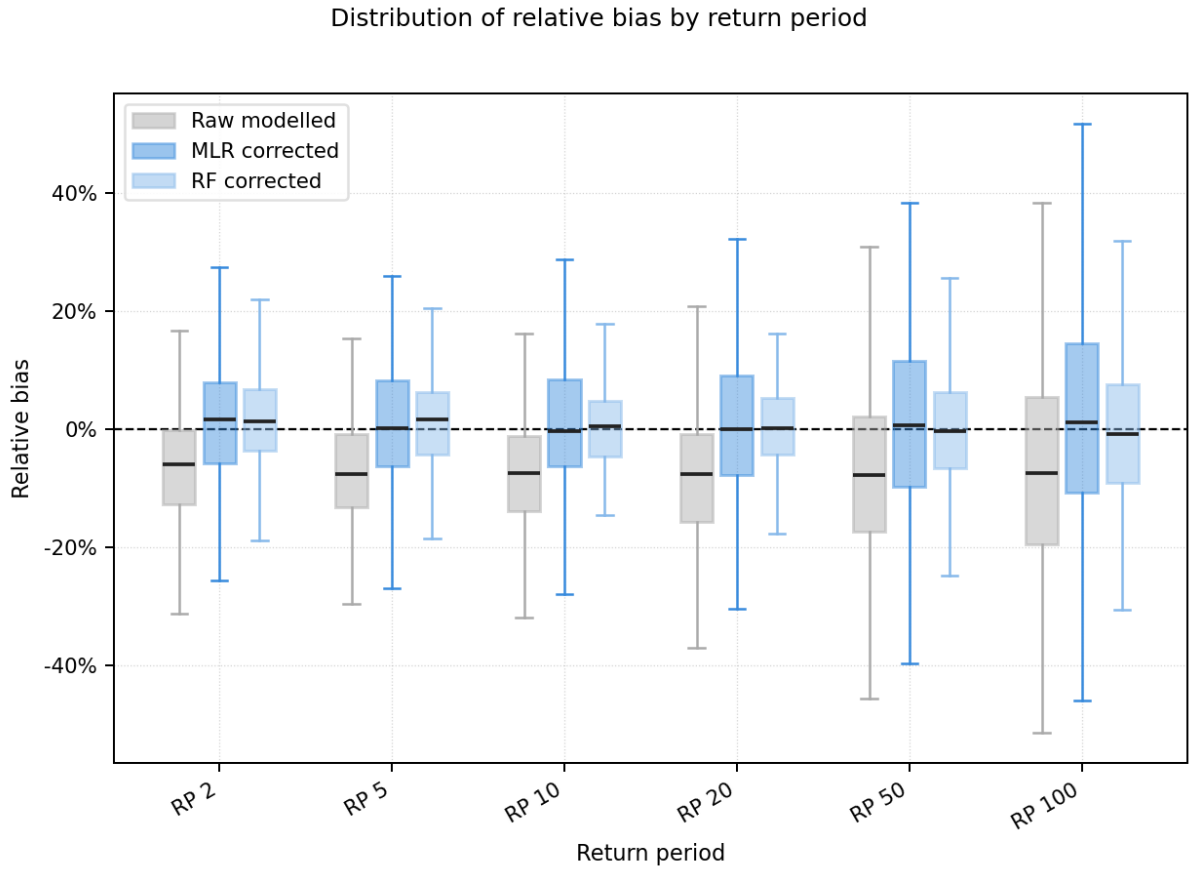


Figure 5.6: Distribution of relative peak flow bias across target return periods ($T = 2$ to 100 years) for the uncorrected baseline, MLR, and RF reduced models. The internal horizontal lines show the median relative bias, while the shaded boxes capture the interquartile range (IQR, 25th to 75th percentiles) and whiskers represent the 2.5th and 97.5th percentiles (extreme outliers omitted). While both correction models bring the median bias close to 0%, the noticeably smaller IQRs and shorter whiskers for the RF model across all return periods demonstrate its superior capability in minimizing overall prediction uncertainty compared to the MLR framework.

6. Discussion

6.1 Interpretation of results and hydrological processes

6.1.1 Distribution selection

The superior fitting performance of the GEV distribution paired with L-moments (45.9% and 30.5% best-fit shares for the AD and PPCC tests, respectively) is consistent with theoretical expectations. Because the GEV distribution is specifically designed for block series, such as annual maximum series it was expected to perform quite well [14,34]. The other distributions were likely limited by:

- **Gumbel:** Simplifies GEV by forcing a constant exponential tail due to $\kappa = 0$ and cannot adjust to unpredictable shapes, which short-to-moderate data sets likely has [34].
- **Log-Normal:** Designed to model continuous variables skewed to the right and bounded by zero at the lower end [34]. More suitable for daily or monthly hydrological variables.
- **Log-Pearson III:** Flexible three parameter distribution, like GEV, and had the second best fit. Skewness parameter is sensitive to sampling noise and requires a longer data record to stabilize [34]. This is why the skewness was bound during the analysis, to prevent the upper tail to explode or collapse.

The superiority of L-moments over MLE in this context aligns with previous findings that L-moment estimators are more robust for small-to-moderate sample sizes, like the 16-year data set used in this thesis [23].

6.1.2 Performance of S-HYPE and regression model outcomes

The empirical results confirm that the unadjusted S-HYPE model exhibits a systematic underestimation of peak discharges, with 75% of uncorrected data points plotting above the 1:1 threshold line (Figure 5.2) and a bias of -7.3% for the normal modelled flow (Table 5.2). Both by Bergstrand et al. and Brunner et al., also reported that process-based hydrological models tend to underestimate extreme discharge events due to limitations in representing rapid surface runoff and saturation-excess processes during intense precipitation [7,8]. The underestimation is physically plausible because the conceptual structure of HYPE, like many rainfall-runoff models, smooths the hydrograph response and may not fully capture localised, short-duration events that produce the highest annual peaks [35].

Despite this bias, the high R^2 values of 0.964 for unadjusted modelled flow demonstrate that the model captures the spatial and temporal variability of peak flows reasonably well. S-HYPE does provide a meaningful ranking of flow magnitudes across catchments, even if absolute values are systematically low. This property is what makes a regression-based correction approach viable. Because the model reproduced the correct ranking but wrong magnitudes, a scaling correction can be meaningful.

Thus the underestimation can be mitigated by incorporating specific physiographical variables. The retained predictors in the reduced MLR model (Table 5.1) provide insights into these correction models:

- **Total Upstream Area** ($\beta = -0.03596$) **and Catchment Area** ($\beta = -0.00649$)**:** Intriguingly, both spatial scaling parameters yielded negative coefficients in the log-linear formulation. In physical hydrology, this presumably represents the spatial attenuation and flood-peak routing delays that occur over large drainage networks. This accounts for the natural attenuation of peak waves as they move through expanding river networks, preventing the correction framework from over-inflating flood magnitudes in larger catchments.
- **Regulation Degree** ($\beta = 0.19706$)**:** The positive relationship between regulation intensity and the required correction indicates that S-HYPE over-attenuates flows in managed catchments. Process-based models often utilize rigid, highly conservative reservoir operation rules that assume maximum water retention during high-flow conditions. In reality, anthropogenic reservoir storage and dam retention strategies can vary due to power generation goals or limited pre-release capacity. The model over-dampens these systems, requiring a positive scalar correction to match observed realities.
- **Agriculture Fraction** ($\beta = 0.09196$)**:** The positive coefficient directly indexes the impacts of artificial field drainage (tile drainage) and reduced surface roughness relative to natural forest canopies. Agricultural lands accelerate rainfall-to-runoff transformation rates by removing natural surface storage opportunities, generating higher and flashier hydrograph peaks that the generalized routing components of S-HYPE fail to fully resolve.

6.2 Methodological comparison: MLR vs. RF

A critical outcome of this study is the comparative performance between the parametric MLR and the non-parametric RF frameworks. As shown in Table 5.2, RF model consistently outperformed the MLR across all configurations, reaching an R^2 of 0.990 in the full model setup. This variance in capability highlights the fundamental architectural differences between the two algorithms.

The MLR framework is bound by a strict, pre-defined log-linear formulation that assumes a continuous, monotonic relationship between the physical attributes and the peak flow corrections. RF, as an ensemble of randomized decision trees, is entirely unconstrained by functional forms. It identifies multi-dimensional, step-wise partitions across the predictor space, allowing it to capture non-linear variations and high-order interactions automatically without user-defined cross-product terms[30].

This divergence in architecture explains the performance trends observed between the full and reduced model configurations. For RF, the full model yielded the highest predictive capacity ($R^2 = 0.990$, Bias = +0.6%). Because tree architectures partition data orthogonally, the inclusion of weaker or slightly redundant variables does not degrade model performance; the algorithm simply isolates and ignores non-informative parameters during tree-node splitting. This aligns with standard machine learning literature establishing that ensemble decision trees are structurally resilient against multicollinearity and feature inflation[32].

Conversely, the MLR model achieved its optimal and most stable configuration when pruned down to five highly significant features ($R^2 = 0.972$, Bias = +1.4%). Including non-significant variables with large confidence intervals (as seen in Figure 5.3) introduces multicollinearity into a parametric framework, which inflates the variance of the estimated β-weights and decreases global predictive stability out-of-sample [31].

However, looking at the central point estimates alone can be misleading, as the MLR was surprisingly close to RF in its raw R^2 capabilities. The true superiority of the RF model is revealed when examining the distribution of relative validation bias across target return periods (Figure 5.6). For every evaluated return period from 2 to 100 years, the RF model exhibits a significantly tighter IQR and substantially shorter whiskers than the MLR counterpart. This smaller error distribution demonstrates that the ensemble mechanics of RF yields a model that is significantly less sensitive to sampling noise, making its predictions consistently more reliable across independent watersheds.

6.3 Limitations, scale constraints, and uncertainties

A primary constraint identified in the results is the systematic decrease in prediction accuracy observed beyond the 20-year return period ($T > 20$). As shown by the multi-panel analyses, the model underestimation diminishes while the variance around the central estimates expands rapidly. This trend is a consequence of the finite temporal window available for this study. A Peaks Over Threshold analysis could provide additional data points and will therefore likely decrease the uncertainties regarding the distribution fit and RP flows.

Extrapolating a 16-year observation window (2010–2025) to derive a 100-year flood event introduces statistical uncertainty. At short record lengths, the empirical sample variance is highly sensitive to the presence or absence of a single historic event. When fitting a GEV distribution via L-moments, this limited sample size introduces wide confidence intervals in the tail-defining shape parameter (ξ), directly projecting structural uncertainty onto long-term return level estimations.

This constraint justifies the exclusion of flows exceeding ~ 1000 m^3/s from the regression modelling framework. This threshold strategy was driven by two core considerations:

1. **Hydrological Focus:** Massive river systems in Sweden are universally regulated and heavily gauged by hydropower consortiums or SMHI monitoring networks. The true challenge for water resource management lies in small-to-medium ungauged catchments, which are highly susceptible to sudden flash events but lack dedicated data streams.

2. **Regression Leverage Points:** Because extreme high-flow data points are statistically sparse, including a small number of massive regional flows would exert an over-proportional influence on the parametric least-squares fitting of the MLR model. These few points would act as statistical leverage handles, warping the entire regression plane to minimize extreme errors at a single high-magnitude station at the expense of prediction accuracy across smaller watercourses [31].

Another limitation of the thesis is that the regression was trained and validated on the same set of stations using cross-validation [32]. While repeated k-fold cross-validation reduces the risk of overfitting, the true test of the method would be application to genuinely ungauged catchments. Future work could include a hold-out validation on stations

not used during model development to provide a more realistic estimate of out-of-sample performance.

The model would likely perform better locally if grouped based on either physical parameters or geographically. In Figure 3.1 it is also evident that there are fewer stations in northern Sweden, which is also where many stations failed to meet the quality control thresholds. This will undoubtedly cause the predictive correction model to have a certain bias towards the conditions in southern/central Sweden.

Finally, the study is temporally bounded by the 2010–2025 period. The observed increase in extreme precipitation events projected under climate change means that correction relationships derived from this period may not remain stable under future hydrological conditions[1]. Periodic recalibration of the correction formula as new observations become available is recommended and may also increase correction performance, because of the larger data series available.

6.4 Implications for flood risk assessment and infrastructure design

For civil engineers, municipal planners, and risk managers working in unmonitored catchments, relying on unadjusted S-HYPE outputs introduces a structural vulnerability. An underestimation bias of 7.3% across return periods means that critical infrastructure designed using uncorrected data will be systematically under-sized.

Implementing the log-linear scaling equations (Equation 4.9) or deploying the trained RF pipeline offers a structured method to correct these outputs before initiating hydraulic calculations. Incorporating these dynamic bias adjustments into local flood risk assessments promotes a design which is able to withstand actual physical peak event magnitudes in urban drainage networks, retention basins, culverts, etc., thus increasing climate resiliency across ungauged Swedish catchments.

7. Conclusions and recommendations

7.1 Summary of key findings

This thesis substantially reduced the systematic S-HYPE peak flow bias under cross-validation. The core conclusions are summarized below:

1. The Generalized Extreme Value distribution fitted via L-moments provides the most reliable mathematical framework for evaluating annual maximum flow distributions across the targeted stations.

2. The unadjusted S-HYPE baseline exhibits a systematic peak underestimation trend (-7.3% bias), which was successfully corrected using multi-variable regression frameworks, reducing relative bias to within $\pm 1.4\%$.

3. The feature selection process identified modelled flow, catchment area, upstream area, regulation status, and agricultural land-use percentages as the primary physical predictors required to regionalize peak corrections.

4. While Multiple Linear Regression provides transparent, easily scalable parametric correction equations, the Random Forest model offers higher non-linear interpolation accuracy within heavily monitored regions.

7.2 Recommendations for future work

To expand upon these findings, future research should test regionalized parameter groupings based on explicit hydroclimatic zones rather than a single nationwide regression. This could capture unique localized behaviours, such as alpine snowmelt dynamics or agricultural tile drainage effects. Future studies could also use a Peaks Over Threshold analysis, scaled to the individual catchments, to get a more robust data set, decreasing the uncertainties.

8. References

[1] IPCC. Chapter 11: Weather and Climate Extreme Events in a Changing Climate, 2021. URL https://www.ipcc.ch/report/ar6/wg1/chapter/chapter-11/.

[2] Marie Buhl and S. Markolf. A review of emerging strategies for incorporating climate change considerations into infrastructure planning, design, and decision making. *Sustainable and Resilient Infrastructure*, 8(sup1):157–169, January 2023. ISSN 2378-9689. doi: 10.1080/23789689.2022.2134646.

[3] Jan Seibert and Sten Bergström. A retrospective on hydrological catchment modelling based on half a century with the HBV model. *Hydrology and Earth System Sciences*, 26(5):1371–1388, March 2022. ISSN 1027-5606. doi: 10.5194/hess-26-1371-2022.

[4] Sten Bergström and Gunlög Wennerberg. Om 100 år Den svenska hydrologiska tjänsten 1908 – 2008. Technical report, SMHI, 2008.

[5] Göran Lindström, Charlotta Pers, Jörgen Rosberg, Johan Strömqvist, and Berit Arheimer. Development and testing of the HYPE (Hydrological Predictions for the Environment) water quality model for different spatial scales. *Hydrology Research*, 41(3-4):295–319, June 2010. ISSN 0029-1277, 2224-7955. doi: 10.2166/nh.2010.007.

[6] SMHI. HYPE - our hydrological model, December 2025. URL https://www.smhi.se/en/research/research-units/hydrology/hype---our-hydrological-model.

[7] Marie Bergstrand, Sara-Sofia Asp, and Göran Lindström. Nationwide hydrological statistics for Sweden with high resolution using the hydrological model S-HYPE. *Hydrology Research*, 45(3):349–356, June 2014. ISSN 0029-1277, 2224-7955. doi: 10.2166/nh.2013.010.

[8] Manuela I. Brunner, Lieke A. Melsen, Andrew W. Wood, Oldrich Rakovec, Naoki Mizukami, Wouter J. M. Knoben, and Martyn P. Clark. Flood spatial coherence, triggers, and performance in hydrological simulations: large-sample evaluation of four streamflow-calibrated models. *Hydrology and Earth System Sciences*, 25(1):105–119, January 2021. ISSN 1607-7938. doi: 10.5194/hess-25-105-2021.

[9] Manuela I. Brunner, Lieke A. Melsen, Andrew W. Wood, Oldrich Rakovec, Naoki Mizukami, Wouter J. M. Knoben, and Martyn P. Clark. Flood hazard and change impact assessments may profit from rethinking model calibration strategies, May 2020.

[10] Mariana Castaneda-Gonzalez, Annie Poulin, Rabindranarth Romero-Lopez, and Richard Turcotte. Hydrological models weighting for hydrological projections: The impacts on future peak flows. *Journal of Hydrology*, 625:130098, October 2023. ISSN 0022-1694. doi: 10.1016/j.jhydrol.2023.130098.

[11] J. R. M. Hosking and James R. Wallis. *Regional Frequency Analysis: An Approach Based on L-Moments*. Cambridge University Press, April 1997. ISBN 978-0-521-43045-6. Google-Books-ID: gurAnfB4nvUC.

[12] SMHI. HYPE variables [HYPE Model Documentation], February 2026. URL `http://hype.smhi.net/wiki/doku.php?id=start:hype_file_reference:info.txt:variables`.

[13] Óscar E. Coronado-Hernández, Julio Ramos-Urzola, Luis F. Julio-Amigo, Jairo R. Coronado-Hernández, Gustavo Gatica, and Nohora Mercado-Caruso. A preliminary analysis for selecting the best hydrological probability density functions of annual peak flows associated to various return periods in some rivers of Colombia. *Procedia Computer Science*, 198:560–565, 2022. ISSN 18770509. doi: 10.1016/j.procs.2021.12.286.

[14] Jery R Stedinger, Richard M Vogel, and Efi Foufoula-Georgiou. Frequency analysis of extreme events. In *Handbook of Hydrology*, pages 679–754. McGraw-Hill, 1993.

[15] Asif Iqbal Shah and Nibedita Das Pan. Evaluation of probability distribution methods for flood frequency analysis in the Jhelum Basin of North-Western Himalayas, India. *Cleaner Water*, 2:100044, December 2024. ISSN 2950-2632. doi: 10.1016/j.clwat.2024.100044.

[16] Cristian Gabriel Anghel. Revisiting the Use of the Gumbel Distribution: A Comprehensive Statistical Analysis Regarding Modeling Extremes and Rare Events. *Mathematics*, 12(16):2466, January 2024. ISSN 2227-7390. doi: 10.3390/math12162466.

[17] Hoshin Gupta, Harald Kling, Koray Yilmaz, and Guillermo F. Martinez. Decomposition of the Mean Squared Error and NSE Performance Criteria: Implications for Improving Hydrological Modelling | Request PDF. *Journal of Hydrology*, 377(1):80–91, October 2009. doi: 10.1016/j.jhydrol.2009.08.003.

[18] D. N. Moriasi, J. G. Arnold, M. W. Van Liew, R. L. Bingner, R. D. Harmel, and T. L. Veith. Model Evaluation Guidelines for Systematic Quantification of Accuracy in Watershed Simulations. *Transactions of the ASABE*, 50(3):885–900, 2007. ISSN 2151-0040. doi: 10.13031/2013.23153.

[19] Wouter J. M. Knoben, Jim E. Freer, and Ross A. Woods. Technical note: Inherent benchmark or not? Comparing Nash–Sutcliffe and Kling–Gupta efficiency scores. *Hydrology and Earth System Sciences*, 23(10):4323–4331, October 2019. ISSN 1027-5606. doi: 10.5194/hess-23-4323-2019.

[20] Tanmaya Kumar Sahoo and Rachita Panda. Estimating Uncertainty in Flood Frequency Analysis Due to Limited Sample Size Using Bootstrap Method. In B. B. Das, Hiroshan Hettiarachchi, Prasanta Kumar Sahu, and Satyajeet Nanda, editors, *Recent Developments in Sustainable Infrastructure (ICRDSI-2020)—GEO-TRA-ENV-WRM: Conference Proceedings from ICRDSI-2020 Vol. 2*, pages 653–661. Springer, Singapore, 2022. ISBN 978-981-16-7509-6. doi: 10.1007/978-981-16-7509-6_51.

[21] Weiqiang Zheng, Shuguang Liu, Zhengzheng Zhou, and Yiping Guo. Flood frequency analysis: confidence interval estimations with generalized extreme value distributions using special pivotal quantities. *Stochastic Environmental Research*

and Risk Assessment, 39(11):4933–4948, November 2025. ISSN 1436-3259. doi: 10.1007/s00477-025-03055-4.

[22] Larry A. Wasserman. *All of Statistics: A Concise Course in Statistical Inference*. Springer, December 2003. ISBN 978-0-387-40272-7.

[23] J. R. M. Hosking. L-Moments: Analysis and Estimation of Distributions Using Linear Combinations of Order Statistics. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 52(1):105–124, September 1990. ISSN 1369-7412, 1467-9868. doi: 10.1111/j.2517-6161.1990.tb01775.x.

[24] M. A. Stephens. EDF Statistics for Goodness of Fit and Some Comparisons. *Journal of the American Statistical Association*, 69(347):730–737, September 1974. ISSN 0162-1459. doi: 10.1080/01621459.1974.10480196.

[25] James J. Filliben. The Probability Plot Correlation Coefficient Test for Normality. *Technometrics*, 17(1):111–117, February 1975. ISSN 0040-1706. doi: 10.1080/00401706.1975.10489279.

[26] Fuladipanah Mehdi and Jorabloo Mehdi. Determination of Plotting Position Formula for the Normal, Log-Normal, Pearson(III), Log-Pearson(III) and Gumble Distributional Hypotheses Using The Probability Plot Correlation Coefficient Test. *World Applied Sciences Journal*, 15(8):1181–1185, 2011. ISSN 1818-4952.

[27] Richard W Katz, Marc B Parlange, and Philippe Naveau. Statistics of extremes in hydrology. *Advances in Water Resources*, 25(8-12):1287–1304, August 2002. ISSN 03091708. doi: 10.1016/S0309-1708(02)00056-8.

[28] M. Lang, T. B. M. J. Ouarda, and B. Bobée. Towards operational guidelines for over-threshold modeling. *Journal of Hydrology*, 225(3):103–117, December 1999. ISSN 0022-1694. doi: 10.1016/S0022-1694(99)00167-5.

[29] Richard M. Vogel, Wilbert O. Thomas, and Thomas A. McMahon. Flood-Flow Frequency Model Selection in Southwestern United States. *Journal of Water Resources Planning and Management*, 119(3):353–366, May 1993. ISSN 0733-9496, 1943-5452. doi: 10.1061/(ASCE)0733-9496(1993)119:3(353).

[30] Leo Breiman. Random Forests. *Machine Learning*, 45(1):5–32, October 2001. ISSN 1573-0565. doi: 10.1023/A:1010933404324.

[31] Max Kuhn and Kjell Johnson. *Applied Predictive Modeling*. Springer, 2013. doi: 10.1007/978-1-4614-6849-3.

[32] Trevor Hastie, Robert Tibshirani, and Jerome Friedman. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer Science & Business Media, second edition, 2009. doi: 10.1007/978-0-387-84858-7.

[33] Peter Bruce and Andrew Bruce. *Practical Statistics for Data Scientists*. O'Reilly Media, 1 edition, May 2017. ISBN 1-4919-5296-2.

[34] Ven Te Chow, David R. Maidment, and Larry W. Mays. *Applied Hydrology*. McGraw-Hill, New York, 1988.

[35] András Bárdossy and Faizan Anwar. Why do our rainfall–runoff models keep underestimating the peak flows? *Hydrology and Earth System Sciences*, 27(10):1987–2000, May 2023. ISSN 1027-5606. doi: 10.5194/hess-27-1987-2023.

9. Appendices

9.1 Appendix A: Additional station information

Table 9.1: All downloaded stations with their Station ID and matched Catchment SUBID. Stations shown in red were downloaded but not included in the final analysis due to mismatching data.

Station Name	Station ID	Catchment SUBID
Abisko	2357	17633
Alfta Krv	1890	13645
Alttajärvi	2356	16290
Ankarvattnet	1537	18333
Arnestorp	2311	6507
Arödån	1881	23643
Assmebro	1166	24978
Attorp	2371	24998
Baggebol Nedre 3	2250	2046
Berg	2218	1923
Bergeforsens Krv	2185	10439
Björkborns Krv	2419	25156
Björna Krv	2376	10268
Blankaström	1806	5574
Boasjöbäck	2318	4319
Bodens Krv	2131	20629
Borga Nedre	1581	18219
Borgviks Krv	2033	24970
Bosgården	2211	24836
Brattforsens Krv	2420	25191
Brusafors	1622	4893
Bräkne-Hoby	2189	5218
Bröddebacken	2127	4555
Buskbäcken	2227	7871
Byske	2284	21527
Bössbo	2396	13477
Carlfors	2439	5357
Christineholms Regl.Damm	2020	2705
Collins Mölla 2	2444	300
Dalkarlshyttan	2206	2628
Dalkarlså	2409	16751
Dansjön	2335	5046
Dömsta	2219	2617

Continued on next page

Table 9.1 – *continued*

Station Name	Station ID	Catchment SUBID
Dönje Krv	2384	13726
Edsforsens Krv	1703	25289
Edsvalla Krv	2036	25250
Eggelstad	2125	5140
Eggvena	2213	24815
Ellinge	2126	4972
Emsfors	20002	5582
Ersbo	654	10903
Fellingsbro	2205	2424
Finsta	2429	2291
Flåsjön Nedre	1519	16660
Flötemarken	1375	24701
Fors 2	2611	2543
Forsmöllans Krv	2372	25055
Fryele	2362	24959
Fyrås	2117	10220
Fättjaure	1425	9941
Galgbacken	2443	24974
Gauträsk	1630	20755
Getebro	855	5499
Gimdalsby	97	10194
Gransel	1387	21567
Gransjön	2605	9831
Granåker	2237	21620
Grea	2309	9092
Gredeby	2188	5251
Gränvad	2217	7344
Gråda Krv	1949	13716
Grötsjön	1171	13428
Gullspångs Krv	1610	5559
Gävunda Krv	2367	12265
Göstad	2328	4756
Hallamölla	2146	2768
Hallbosjön	1273	2675
Halvfari Krv	2383	12897
Halvorsbyn	1243	4793
Hammarby 2	2153	2604
Hammarforsen	558	10434
Harrsele Krv	1733	16682
Hassela	2273	13456
Hebylund Kvarnsjön	2614	2033
Hestraviken	2606	25029
Heåkra	2128	167
Hillared	364	25013

Continued on next page

Table 9.1 – *continued*

Station Name	Station ID	Catchment SUBID
Hissmofors Krv	1357	14645
Hjälta Krv	1557	14416
Holmen	2445	5603
Holmfors	2596	21161
Hotagströmmen	1909	10271
Hulubäcken	2305	3096
Hyttingsheden	2410	12428
Hälla	2331	5258
Hällered	2422	23476
Härnevi	2248	2432
Hålabäck	736	2765
Högsmölla	2171	5477
Höljes Krv	2224	25276
Hörsne	2365	5819
Hörvik	10450	247
Idafors	1706	21494
Idre 3	2258	13612
Igelfors	2617	5087
Junosuando	4	17697
Järnsbäcken	10458	4869
Järpströmmens Krv	1495	15225
Kaalasjärvi	1456	524
Kallio 2	2395	17707
Karats	1403	17542
Karesuvanto	10006	17677
Karlslund 2	2139	2643
Kedevad 2	2465	4488
Kila Övre	2595	2292
Kilforsens Krv	1944	15382
Killingi	2159	24561
Kilvamsklippen	1685	10226
Klinggården	2770	2507
Klippan 2	1635	24724
Klockarträsk	2597	21230
Konstdalsströmmen	2240	13508
Kringlan	2229	2434
Krokbräcke	2119	4815
Krokfjorden	2482	5422
Krokfors Kvarn	2215	24454
Kukkasjärvi	1160	17459
Kukkolankoski Övre	16722	17733
Kvarntorp	2310	6918
Kvistforsens Krv	1870	18424
Källstorp 2	1962	5235

Continued on next page

Table 9.1 – *continued*

Station Name	Station ID	Catchment SUBID
Kåbtåjaure	1790	23961
Kåge 2	2235	21385
Kåtaselet	1480	21499
Köpinge	1697	212
Laholms Krv	2477	25281
Laisvall	2414	20936
Lajksjön 2	2323	10022
Landösjön Nedre	1701	10178
Lannavaara	5	25024
Lasele Krv	1736	15304
Lennartsfors Krv	1764	6666
Letsi Krv	2039	22483
Lillglän	1083	25345
Lillänget	2283	21440
Lima Krv	1762	13656
Lisjöbäcken	1882	23922
Lissbro	2336	4811
Ljusne Strömmar Krv	2257	13735
Ljusnedal Övre	1169	13362
Loversbro	2619	5280
Lännässjön	20035	10238
Långhag	1643	13745
Långsjön	1690	19954
Magnor	10016	25669
Malgomajs Krv	2369	17762
Mariefors	2400	5409
Marstad	2406	4432
Medstugan Nedre	2152	15212
Melby	2289	5096
Mellankvarn	2214	25021
Mertajärvi	1780	17421
Mesjön Övre	2255	18058
Mockfjärd Krv	2203	13693
Moholm	1221	25084
Motala Krv	1950	4671
Munkedal 2	257	25125
Mälaren	20040	2739
Männikkö	11	17191
Möckeln	1069	5458
Mörrum	186	5558
Nedra Bägby	2200	4389
Nedre Skärvagen	2467	13613
Niavve	591	23199
Niemisel	20	17646

Continued on next page

Table 9.1 – *continued*

Station Name	Station ID	Catchment SUBID
Nissaström	2471	1365
Nordmark 2	2408	7710
Norekvarn	2202	851
Norrefors 2	2621	4729
Norrsjön	2053	13431
Nybro	740	13602
Nymölla	2256	206
Nytorp	2330	5230
Näs 2 Tämnarån	2612	5714
Odensvibron 2	2221	7081
Oppli	2399	12808
Orgvätar	2304	3215
Ottsjön Nedre	1336	10098
Pajala Pumphus	2012	17708
Pepparforsen	2341	24849
Porjus Kraftverk	1702	23351
Ramnäs Krv	2343	2687
Ransta	2247	7457
Rengen	1341	10134
Riebnes Krv	2302	22273
Ringsjödammen Övre	2176	188
Risnäs	2398	9459
Ritsems Krv	2298	25280
Rolfsta	1159	13565
Ryggesbo	2417	11154
Ryttarbacken	2170	2762
Räktfors	17	17711
Råda Krv	1446	8101
Rörvik	200	24566
Röån	1378	15569
Saras Fors	2405	13476
Saxbro	2402	2098
Seitevare Krv	1969	23541
Sibro 2	2587	2538
Sikfors Krv	1788	21591
Simlången	1575	1162
Skabram	1675	16330
Skallböle Krv	1650	10406
Skangselforsen	2598	21465
Skattungen	1567	13571
Skeens Krv	1952	25163
Skirknäs	2275	20175
Skogsliden	2404	13241
Skye Kvarn	2337	5187

Continued on next page

Table 9.1 – *continued*

Station Name	Station ID	Catchment SUBID
Skällnora Övre	2610	2016
Skärblacka	2454	5599
Skärsboda	1809	2757
Slöta	2347	4435
Snapparp	2340	24910
Snogeröd2	10501	21667
Solberg 2	2608	21417
Sollefteå Krv	2133	14358
Solnen	2291	2797
Solveden	887	3421
Sorsele 2	2238	21579
Spjutmo Krv	2436	13687
Sporrsjön Vängelälv 2	20025	15634
Spånga Kvarn	2212	24904
Spånga Nedre	2143	347
Stabby	1742	710
Staloluokta	2353	24320
Stensjön 2	2342	24972
Stensåkra	2361	5379
Stenudden	37	21498
Storavan	20061	20397
Storjuktan Totalt	20047	21463
Stormyra	1835	6179
Stornorrfors Krv	1734	21632
Storuman 2	20010	21558
Storvågen	2767	9967
Strömfors	1609	21359
Strömsborg	2292	5122
Strömsvattudal 3	20014	15867
Stugusjön	2604	9790
Stupforsen	2607	17501
Sundstorp	1236	24993
Sved	2359	3768
Svedala	1879	47
Svegs Krv	2434	13690
Svärdsjö	2407	13569
Sädva Krv	2382	22149
Säffle Övre	2616	25259
Sällsjön Totalt	20033	14469
Söderköping	2432	5184
Tolhem	2293	5520
Tolvfors Krv	1940	13620
Torpshammar Krv	1471	13069
Torrböle 2	2506	21509

Continued on next page

Table 9.1 – *continued*

Station Name	Station ID	Catchment SUBID
Torröns Krv	2208	15746
Torsebro Övre	2525	5567
Trolleberg 2	2768	5169
Tvärsjön 2	2278	22105
Tängvattnet	1673	21062
Tärendö 2	2358	17691
Tärnsjö	2299	5634
Tånemölla	2129	46
Törnestorp	1639	4196
Uppsala Bärbyleden	2609	2630
Vagelö	2618	5478
Vagge	2308	25064
Valex	2118	24164
Valtorp	2345	24794
Vargöns Krv	1954	4409
Vassbotten	751	24952
Vattholma 2	2244	38
Velen 2	1662	4531
Vietas (Satisdelen)	20019	24354
Vietas (Suorvadelen)	20018	24701
Vojmsjöluspen	20031	18415
Vombsjön Övre	2018	125
Vrångebäcken	1623	24771
Vuoddasbäcken	1963	16517
Värmeshults Krv	1907	1774
Vässinkoski Krv	2435	11164
Västersel	1583	10161
Yngeredsforsen	2472	1679
Ytterholmen	1123	17545
Ytterträsk Nedre	2236	21347
Yxered 2	2290	2967
Äcklingen	1309	15694
Älvkarleby Krv	2423	13771
Ängabäcks Krv	1953	25272
Ärrarp	2325	24755
Åbromölla	2196	285
Åkers Krutbruk	2249	6298
Åkesta Kvarn 3	2773	2560
Åsbro 3	2201	25181
Åtorp-Stavån	2613	2101
Östavallselet	20036	10328
Östra Norn	1328	10218
Övre Abiskojokk	957	25812
Övre Hyndevad	138	2713

Continued on next page

Table 9.1 – *continued*

Station Name	**Station ID**	**Catchment SUBID**
Övre Lansjärv	1740	17584

Table 9.2: Dataset categories used in the analysis.

Dataset	**Median**	**Lowest**	**Highest**	**Average**
Catchment area	13	0.01	485	35
Upstream area	660	3	28900	1880
Regulation	0.0	0.0	0.841	0.104
Agriculture	0.078	0.0	0.912	0.185
Forest	0.537	0.0	0.868	0.491
Water bodies	0.029	0.0	0.483	0.076
Wetland	0.045	0.0	0.327	0.068

9.2 Appendix B: Additional result figures

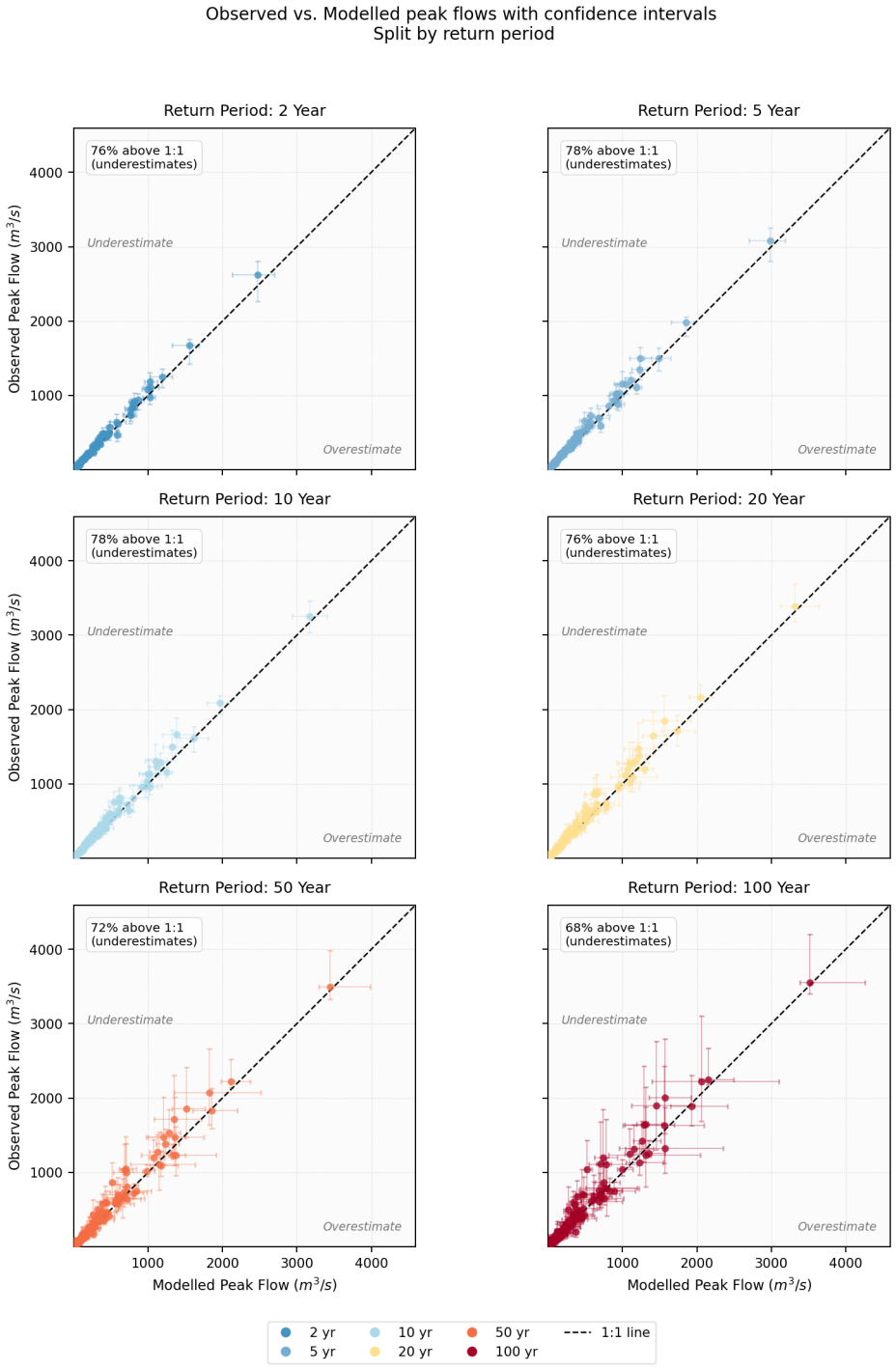


Figure 9.1: Multi-panel 1:1 validation plot displaying the 90% confidence intervals (CIs) split by individual target return period ($T = 2$ to 100 years) for all stations. Horizontal error bars represent the 90 % CI of the modelled peak flow, and vertical error bars represent the CI 90 % of the observed peak flow. Points falling above the dashed 1:1 line indicate model underestimation. The progressive widening of the error bars and increased scatter at higher return periods (especially post-$T = 20$) visually demonstrate that prediction accuracy degrades as the scale of the extreme event increases.

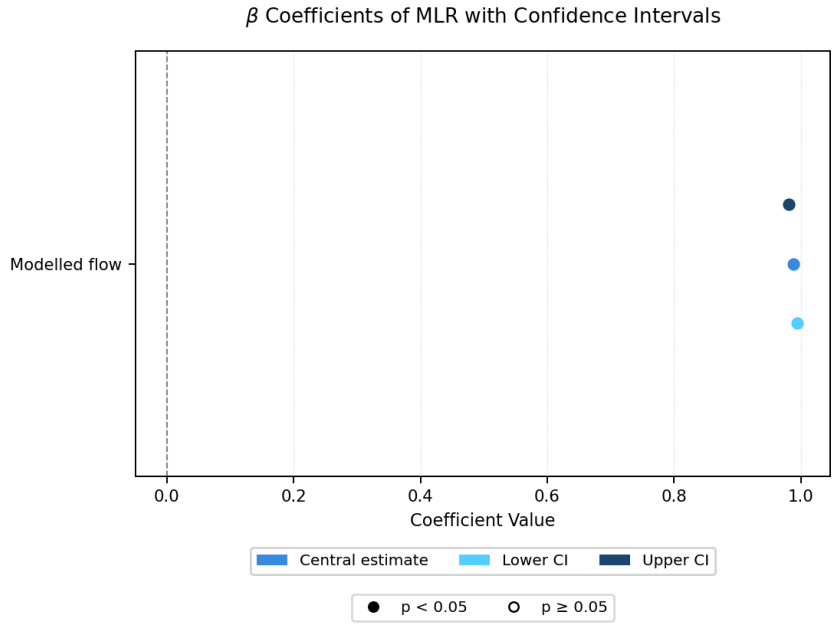


Figure 9.2: Estimated β coefficient of the modelled flow parameter within the flow-only MLR model configuration across varying return periods. Central markers denote the coefficient point estimates (filled circles indicate statistically significant predictors at $p < 0.05$; open circles indicate non-significant ones). The error bars represent the 95% confidence intervals. The stability of this coefficient across all target return periods demonstrates that the scaling relationship between simulated and observed baseline flow remains uniform regardless of event extremity.

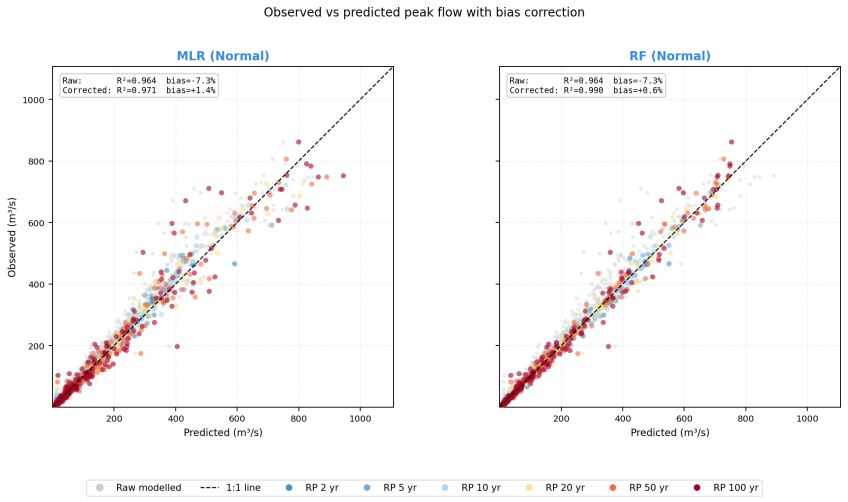


Figure 9.3: Observed versus predicted peak flows (Q) plotted against a dashed 1:1 reference line for the MLR and RF adjustment frameworks utilizing the **full model** configuration (all eleven catchment parameters). Background points in light grey represent the uncorrected S-HYPE baseline, showing a stark trend of underestimation. The tightly clustered, colored markers show that both correction frameworks successfully eliminate this bias across all return periods ($T = 2$ to 100 years), shifting the data onto the 1:1 line to achieve high R^2 values and near-zero relative bias.

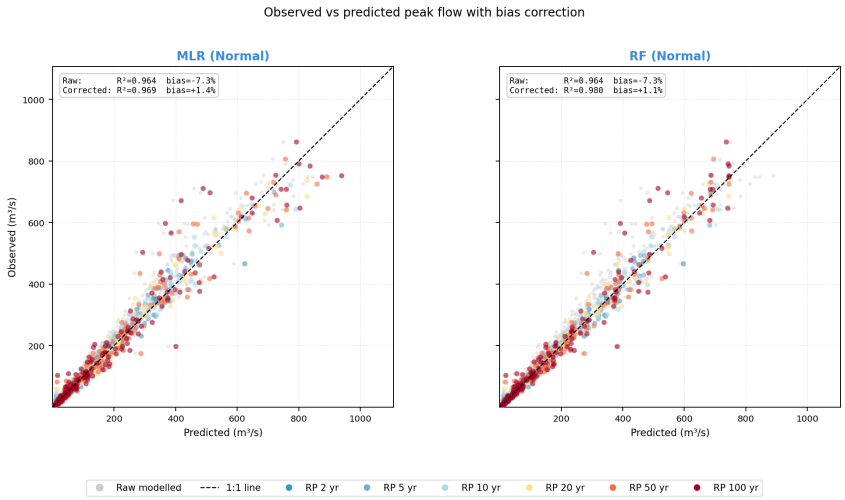


Figure 9.4: Observed versus predicted peak flows (Q) plotted against a dashed 1:1 reference line for the MLR and RF adjustment frameworks utilizing the **flow-only model** configuration (excluding catchment physical characteristics). Light grey background points show the uncorrected S-HYPE baseline underestimation. While the coloured correction markers align well with the 1:1 line and successfully eliminate systemic bias, the slightly wider scatter compared to the full model highlights the marginal performance penalty when omitting catchment-specific physical features.

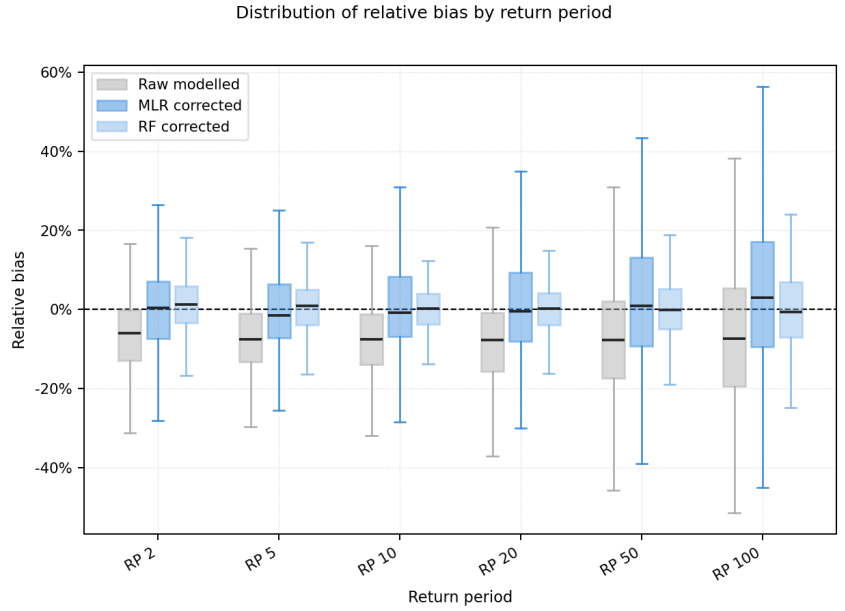


Figure 9.5: Distribution of relative peak flow bias across target return periods ($T = 2$ to 100 years) for the uncorrected baseline, alongside the MLR and RF **full models**. Internal horizontal lines denote the median relative bias, shaded boxes encompass the interquartile range (IQR, 25th to 75th percentiles), and whiskers extend to the 2.5th and 97.5th percentiles (extreme outliers omitted). The noticeably compressed IQRs and shorter whiskers of the RF full model over the MLR highlight the tree-based framework's superior capacity to leverage multiple catchment parameters to minimize prediction variance.

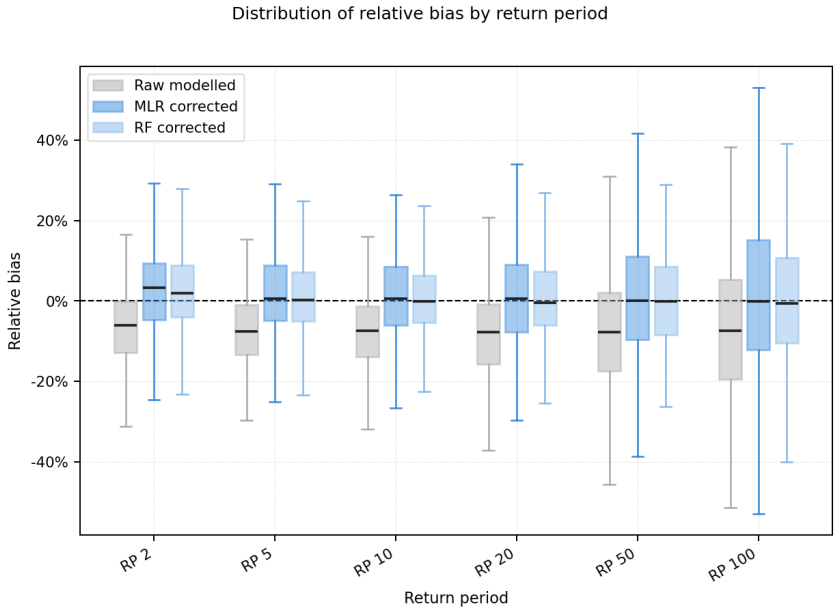


Figure 9.6: Distribution of relative peak flow bias across target return periods ($T = 2$ to 100 years) for the uncorrected baseline, alongside the MLR and RF **flow-only models**. Internal horizontal lines indicate median bias, shaded boxes represent the IQR, and whiskers track the 2.5th/97.5th percentiles. Although both corrected models center the median bias near 0%, the wider IQRs and extended whiskers compared to their full-model counterparts visually demonstrate the increased predictive uncertainty that arises when catchment physical properties are missing from the workflow.

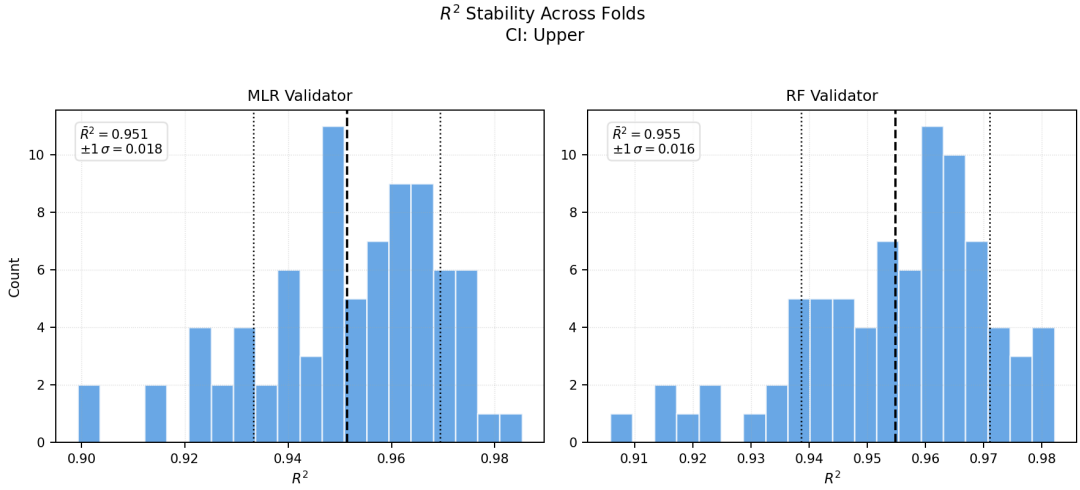


Figure 9.7: Cross-validation stability profiles showing the R^2 performance metrics across all 20 4-folds for the **upper confidence interval limit** using both MLR and RF frameworks. The black dashed line marks the median value, while the dotted lines indicate ± 1 standard deviation from this median. The tight grouping of individual fold markers and narrow standard deviation bands confirm high model stability and low vulnerability to sample overfitting at the upper uncertainty boundary.

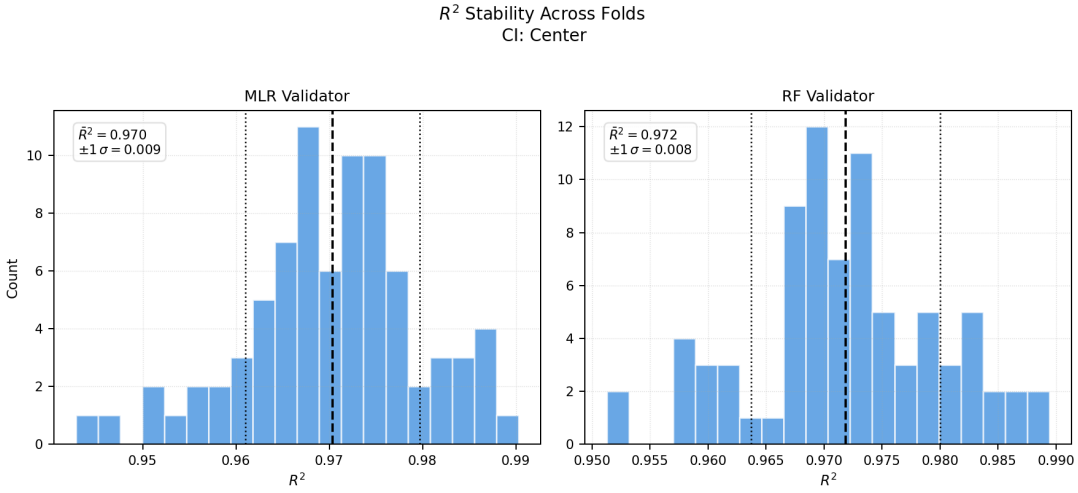


Figure 9.8: Cross-validation stability profiles showing the R^2 performance metrics across all 20 4-folds for the **central point estimate** using both MLR and RF frameworks. The black dashed line marks the median value, and the dotted lines show ± 1 standard deviation. The minimal variance across the folds demonstrates excellent consistency and generalizability of the primary prediction mechanism when applied to independent data subsets.

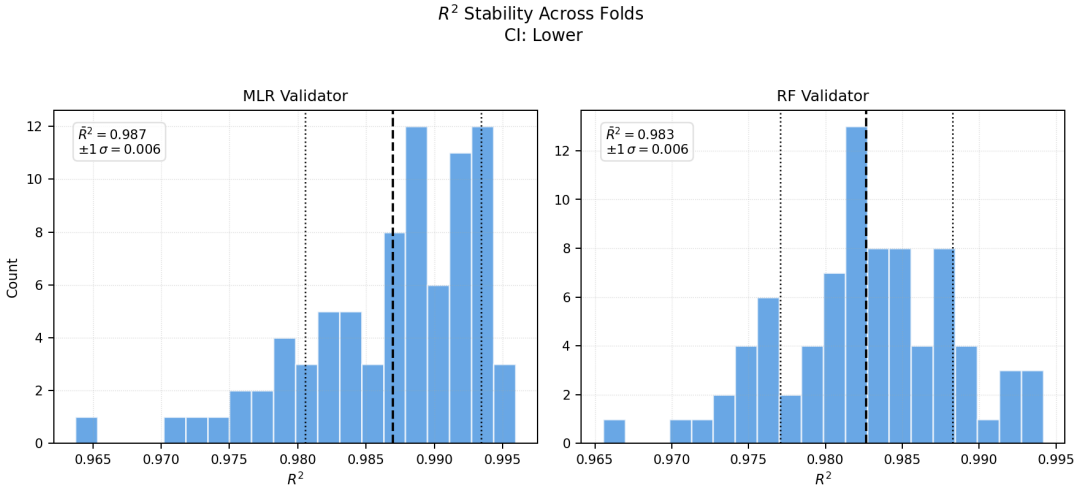


Figure 9.9: Cross-validation stability profiles showing the R^2 performance metrics across all 4 evaluation folds for the **lower confidence interval limit** using both MLR and RF frameworks. The black dashed line indicates the median performance value, and the dotted lines mark ± 1 standard deviation. The low dispersion among the validation folds establishes that the lower uncertainty bounds maintain stable mathematical relationships across distinct catchment sub-samples.

9.3 Appendix C: Figures for comparison to statistically corrected and natural flow

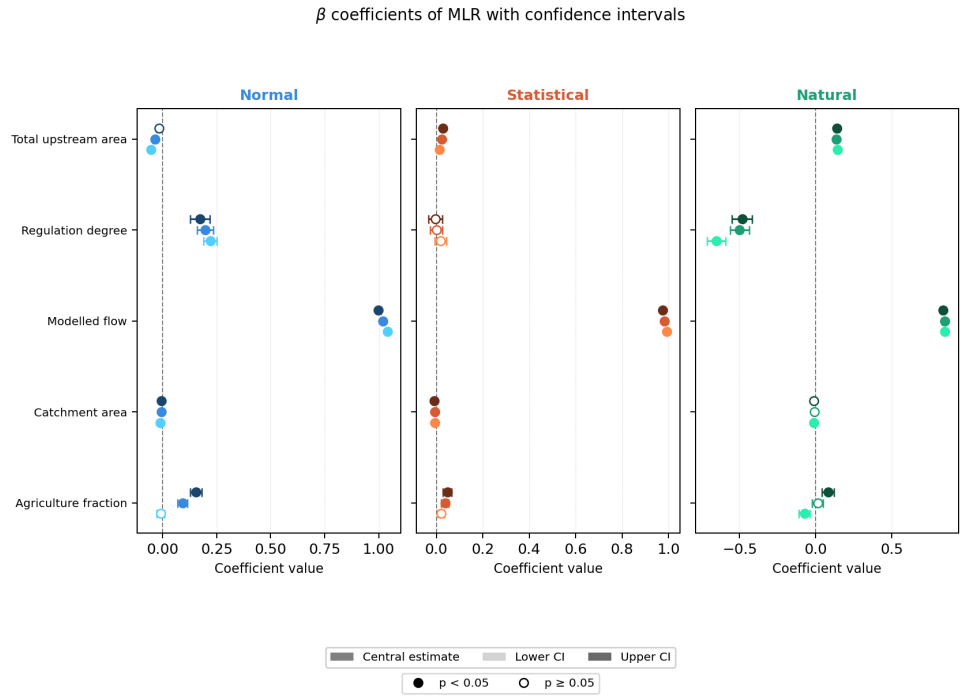


Figure 9.10: Estimated β coefficient weights for the remaining parameters of the reduced MLR configuration, evaluated across all S-HYPE flow series (including both regulated and natural catchment variations). Central markers denote the point estimates, and error bars show the 95% confidence intervals. The only parameter with a large variation is the regulation degree, otherwise the error bounds don't vary much and verify that the selected catchment features maintain robust physical relationships with peak flows under diverse hydrological regimes.

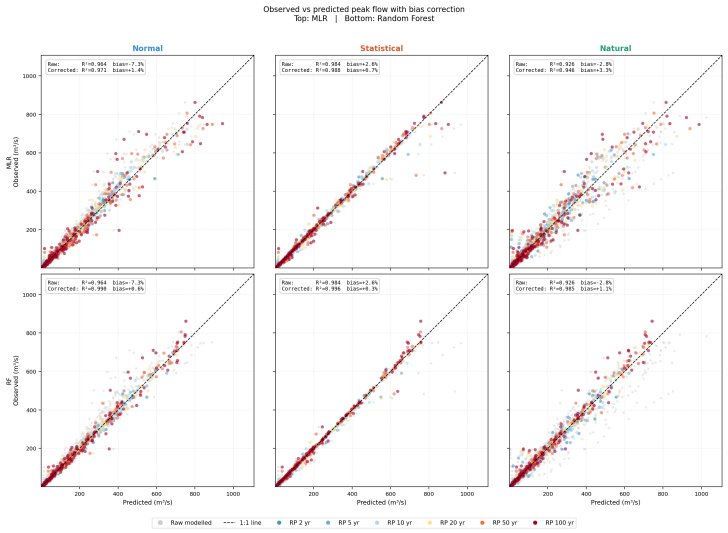


Figure 9.11: Observed versus predicted peak flows (Q) plotted against a 1:1 reference line for the MLR and RF adjustment frameworks of the **full model**, validated across all S-HYPE flow configurations (natural and statistically adjusted). Background light grey points highlight the uncorrected baseline underestimation trend. The statistically adjusted flow has the best baseline R^2 value and benefit the least from the correction, while the opposite is true for the natural flow. RF still performs the best for all flow types.

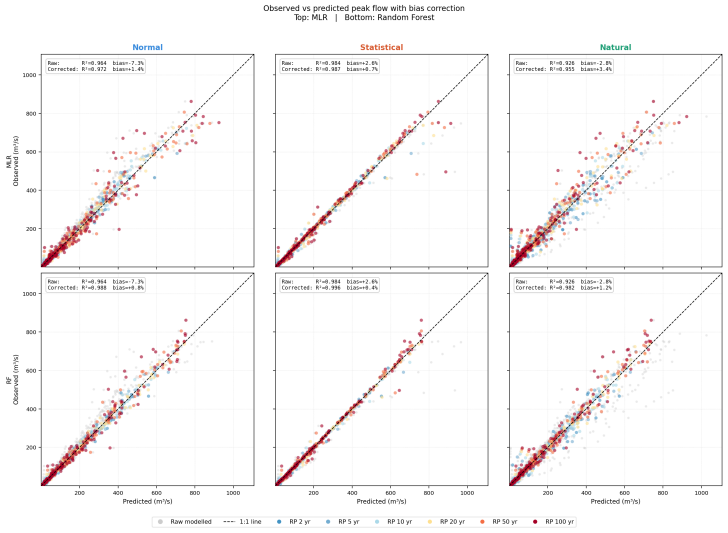


Figure 9.12: Observed versus predicted peak flows (Q) plotted against a 1:1 reference line for the MLR and RF adjustment frameworks of the **reduced model**, validated across all S-HYPE flow configurations. Background light grey points show the uncorrected baseline. The coloured markers illustrate that stripping the model down to its five primary parameters preserves excellent predictive accuracy (R^2) and leaves minimal residual bias, confirming that feature reduction does not compromise performance under natural or modified regimes. The statistically adjusted flow has the best baseline R^2 value and benefit the least from the correction, while the opposite is true for the natural flow. RF still performs the best for all flow types.

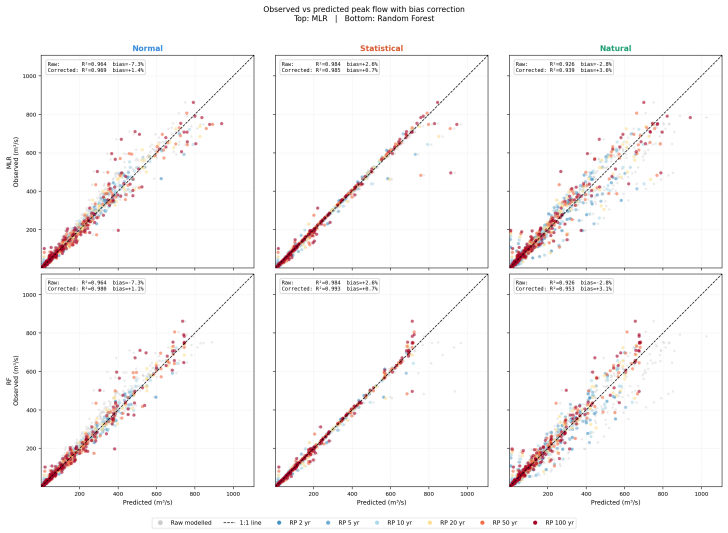


Figure 9.13: Observed versus predicted peak flows (Q) plotted against a 1:1 reference line for the MLR and RF adjustment frameworks of the **flow–only model**, validated across all S-HYPE flow configurations. Light grey background points map the uncorrected baseline. While the coloured correction markers successfully align better with the 1:1 line to fix the systemic underestimation, the visibly higher data dispersion compared to Figures 9.11 and 9.12 underscores the value of including catchment physical data to resolve complex, mixed-flow regimes. Like the previous figures the statistically adjusted flow has the best baseline R^2 value and benefit the least from the correction, while the opposite is true for the natural flow. RF still performs the best for all flow types.

9.4 Appendix D: Disclosure of Artificial Intelligence Usage

This appendix outlines the specific applications and limitations of Artificial Intelligence (AI) tools utilized during the thesis. AI technologies were employed strictly as auxiliary aids to enhance research efficiency and code optimization, while all core conceptual development, and data interpretation remain the sole work of the author and the supervisors. The primary applications of AI in this study include:

- **Code Optimization and Troubleshooting:** Google Gemini and Anthropic Claude were utilized to streamline, debug, and optimize the Python scripts developed for data collection and statistical analysis.
- **Literature Grounding:** Google NotebookLM was employed as a preliminary brainstorming and literature-mapping tool. This platform assisted in identifying relevant thematic connections and verifying that the empirical methods and statistical tests aligned with established academic literature.

Note on Academic Integrity: All final references, methodologies, and code blocks were independently verified and critically evaluated by the author to ensure factual accuracy and compliance with academic standards.