



Royal Netherlands  
Meteorological Institute  
*Ministry of Infrastructure  
and Water Management*

# Comparison of tail models and data for extreme value analysis of high tide water levels along the Dutch coast

C.F de Valk, H.W. van den Brink

De Bilt, 2023 | Scientific report; WR-23-01

## Samenvatting

Voor de beoordeling van de veiligheid van waterkeringen in Nederland zijn waarden van windsnelheid en zeewaterstand voor terugkeertijden tot (plaatselijk) enkele miljoenen jaren nodig. Dit is een uitdaging, gegeven dat reeksen van betrouwbare windmetingen niet verder teruggaan dan ongeveer 70 jaar, en de metingen van waterstand niet verder dan 70-140 jaar.

Momenteel wordt door KNMI en Deltares in opdracht van Rijkswaterstaat gewerkt aan een oplossing van dit probleem. Deze is gebaseerd op het gebruik van omvangrijke datasets van simulaties door numerieke weermodellen en daaraan gekoppelde waterbewegingsmodellen; zie van den Brink (2018, 2020); de Valk and van den Brink (2020a). Deze aanpak staat of valt met de kwaliteit van de modellen. Bovendien, zelfs met gebruik van grote datasets zoals het archief van ECMWF ensemble seizoenvoorspellingen blijft er een fors verschil in terugkeerperiode dat moet worden overbrugd. Daarom wordt gekeken naar de mogelijkheid om de extrapolatie van staarten van kansverdelingen over een breed bereik van terugkeertijden te verbeteren door gebruik van de Gegeneraliseerde Weibull (GW) staart, de log-Gegeneraliseerde Weibull (log-GW) staart, of de Weibull staart.

Voor deze modellen alsmede voor twee klassieke staart-modellen, de Gegeneraliseerde Pareto (GP) staart en de exponentiële staart, evalueren we de nauwkeurigheden van schattingen van de staart van de kansverdeling van de hoogwaterstand en de scheve opzet op 21 kuststations op basis van simulaties met het WAQUA DCSM-5 waterbewegingsmodel aangedreven door schuifspanning en luchtdruk uit de ECMWF SEAS5 ensemble seizoenvoorspellingen, waarvan ongeveer 5800 jaar aan data is gebruikt.

Daarvoor gebruiken we twee methoden: de methode uit van den Brink and Können (2008), en Monte-Carlo simulatie van de schatting van terugkeerwaarden voor zeer grote terugkeertijden op basis van plausibele modellen van de staart, bepaald uit de gesimuleerde gegevens. De laatste methode verbetert de methode gebruikt in de Valk and van den Brink (2020a), maar de resultaten van beide methodes wijken niet veel af.

Voor waterstand en opzet blijken zowel de GW staart als de GP staart goede resultaten te geven wanneer de vormparameters nauwkeurig worden geschat uit de volledige set van SEAS5/DCSM-5 gegevens. De GW staart presteert overall het beste, met name als alle parameters (inclusief de vormparameter) uit een kleine dataset van 78 jaar worden geschat. Eerder bleek dit type staart voor wind en (uit schuifspanning afgeleide) pseudo-wind de meest geschikte optie (de Valk and van den Brink, 2020a). Een bijkomend voordeel van het gebruik van de GW staart voor zowel windsnelheid/pseudo-windsnelheid als opzet/waterstand is dat een machtswet voor de wind-opzet relatie (fysisch enigszins te verdedigen als benadering) behouden blijft in GW staarten.

Daarnaast zijn vergelijkingen gemaakt tussen de staartverdelingen van waterstand en opzet geschat uit de SEAS5/DCSM-5 modelsimulaties en de staartverdelingen geschat uit de meetgegevens van 6 stations. De resultaten wijzen in de richting van enige onderschatting van de vormparameter van de GW staart bij de wat minder hoge waterstanden/opzetten (vaker overschreden dan in 1 op de 100 hoogwaters).

---

Alleen voor de waterstand en opzet van meetstation Delfzijl is duidelijk sprake van onderschatting; tegelijk lijken voor hogere waterstanden en opzetten de vormparameters uit de SEAS5/DCSM-5 modelsimulaties voor Delfzijl meer realistisch dan de schattingen uit de meetgegevens.

Om deze afwijkingen bij Delfzijl beter te begrijpen en te bepalen hoe hiermee om te gaan in de afleiding van terugkeerwaarden en in de foutenanalyse wordt aanbevolen om eerst de geplande analyse van de effecten van modelresolutie en andere aspecten van de atmosferische en waterbewegingsmodellen op de gesimuleerde waterstand en opzet uit te voeren.

Deze studie is uitgevoerd binnen een Maatwerk I&W opdracht, onder begeleiding van Marcel Bottema en Robert Slomp (WVL). Prof. dr. Pieter van Gelder heeft een review van dit rapport uitgevoerd waarvan dankbaar gebruikt is gemaakt voor de eindversie van dit rapport en voor vervolgstudies.

## Summary

To assess the reliability of flood protection in the Netherlands, return values of wind speed and coastal water level for return periods up to several million years are needed. This is a major challenge, given that records of reliable wind measurements do not go back further than about 70 years, and water level measurements do not go back further than 70-140 years.

Several ideas are currently explored to tackle this problem. One idea is to increase data volume by utilizing large datasets of simulations by numerical weather prediction models and hydraulic models forced by these simulations; see van den Brink (2018, 2020); de Valk and van den Brink (2020a). This approach relies heavily on the quality of these models. Furthermore, even large datasets such as the archived ECMWF seasonal ensemble forecasts leave a considerable gap in return period to be overcome. Therefore, a parallel effort is made to improve the extrapolation of the tails of distribution functions over a wide range of return periods: the Generalized Weibull (GW) tail, the log-Generalized Weibull (log-GW) tail, or the 1-parameter Weibull tail.

For these models and for two classical tail models, the Generalized Pareto (GP) tail and the exponential tail, we compare estimates of the tails of the distributions of high-tide water level and skew surge at 21 tide gauge stations derived from simulations by the WAQUA DCSM-5 shallow-water flow model driven by wind stress and surface pressure from the ECMWF SEAS5 seasonal ensemble forecast archive. For each tide gauge station, approximately 5800 years of data is used.

Two methods are used to assess the estimates: the method from van den Brink and Können (2008), and Monte-Carlo simulation of the estimation of return values for very high return periods based on plausible models of the tails, derived from the SEAS5/DCSM-5 data. The latter method improves the method used in de Valk and van den Brink (2020a), but the results of both methods are not very different.

For water level and surge, both the GW tail and the GP tail give accurate estimates if their shape parameters are estimated accurately from the complete set of SEAS5/DCSM-5 data. Overall, the GW tails performs best, in particular if all parameters (including the shape parameter) are estimated from a small 78-

year subset. Earlier, this type of tail was found to be the most suitable one for wind and for pseudo-wind derived from stress (de Valk and van den Brink, 2020a). An additional advantage of using the GW tail both for (pseudo)wind speed and for surge/water level is that a power law for the wind-surge relation (a reasonable simplification) is preserved.

In addition, tail estimates for water level and surge from SEAS5/DCSM-5 model simulations are compared to estimates from measurements at 6 tide gauge stations. The comparison is focused on the shape parameter (as errors in shape estimates are the main source of error in estimates of return values), regarded as a function of water level or surge.

The results indicate that the estimates of the shape parameter of the GW tail from the model simulations are somewhat lower than the estimates from measurement data for relatively low water levels or surges (exceeded in more than 1 in 100 high tides).

Only for low water levels and surge at Delfzijl, the shape is clearly underestimated by the simulated data. However, in the higher range of water levels or surges, the shape estimates from the SEAS5/DCSM-5 simulations for Delfzijl look more realistic than the estimates from the measurements.

To understand these deviations at Delfzijl better, we need the results of the ongoing analysis of the effects of model resolution and other aspects of the atmospheric and hydraulic models on the simulated water levels and surges.

This study was carried out within Maatwerk I&W under supervision by Marcel Bottema and Robert Slomp (WVL). Prof. dr. Pieter van Gelder carried out a review of this report, which was used in the final version and in the work that followed.

---

## Contents

<b>I</b>	<b>Report</b>	9
1	Introduction . . . . .	9
2	The SEAS5/DCSM-5 dataset . . . . .	12
3	Comparison of estimates from the full sample . . . . .	15
4	Statistics of transformed maxima over subsamples . . . . .	19
5	Checking the extrapolation from subsamples by Monte-Carlo simulation from plausible distribution functions . . . . .	27
5.1	Introduction . . . . .	27
5.2	Method . . . . .	27
5.3	Results . . . . .	31
6	Verification of the tail shapes of high-tide water level and skew surge data from SEAS5/DCSM-5 . . . . .	45
7	Implications for simulation model bias . . . . .	51
8	Conclusions . . . . .	53
	Bibliography . . . . .	55
<b>II</b>	<b>Appendix</b>	57
A	Return level estimates of skew surge . . . . .	58
B	Estimation of the GW shape parameter as a function . . . . .	60
C	Error statistics of return values of wind speed estimated by Monte Carlo simulation . . . . .	65
D	Comparison of estimates of the GP shape parameter from SEAS5/DCSM-5 data and measurements . . . . .	68
E	Accounting for serial dependence . . . . .	71

## List of Figures

1	Locations of coastal tide gauge stations in the North Sea. Station OS11 is close to Roompot. . . . .	13
2	Difference between the quantiles of high-tide water level and skew surge from SEAS5/DCSM-5 at the 21 tide gauge stations as functions of the sample fraction exceeding the quantile; the indicated sample fractions correspond to 0.007, 0.7 and 70/year. . . . .	14
3	Return level estimates of high-tide water level for a return period $R = 10^7$ years as function of location index. Estimates are based on sample fractions indicated above the panels (these correspond to frequencies of exceedance of 0.7, 1.4, 3.4, 7, 14 and 35/year). . . . .	17
4	Return level estimates of high-tide water level for a return period $R = 10^4$ years as function of location index. Estimates are based on sample fractions indicated above the panels (these correspond to frequencies of exceedance of 0.7, 1.4, 3.4, 7, 14 and 35/year). . . . .	18
5	Gumbel plots of transformed maxima over the subsamples of the data of water level (left) and surge (right), for different tails (see legend) and stations (rows). Dashed: standard Gumbel line; dotted: 95% confidence bounds. Tail estimates based on values exceeded 25/year (sample fraction of 0.036). . . . .	20
5	Continued from last page. . . . .	21
6	Gumbel plots of transformed maxima over the subsamples of the data of water level (left) and surge (right), for different tails (see legend) and stations (rows). Dashed: standard Gumbel line; dotted: 95% confidence bounds. Tail estimates based on values exceeded 5/year (sample fraction of 0.007). . . . .	22
6	Continued from last page. . . . .	23
7	Gumbel plots of transformed maxima over the subsamples of the data of water level (left) and surge (right), for different tails (see legend) and stations (rows). Dashed: standard Gumbel line; dotted: 95% confidence bounds. Tail estimates based on values exceeded 1.0/year (sample fraction of 0.0014). . . . .	24
7	Continued from last page. . . . .	25
8	GW-shape parameter estimates $\hat{\rho}_p$ from SEAS5/DCSM-5 data of skew surge at Cuxhaven as function of sample fraction $p$ (full), with their two-sided 95% confidence intervals. The dashed line is $\hat{\gamma}_p(\log 1/p) + 1$ , with $\hat{\gamma}_p$ the GP- shape parameter estimates from the same data. The indicated sample fractions correspond to 0.007, 0.07, 0.7, 7.0 and 70/year. . . . .	28

9	GW-shape parameter estimates $\hat{\rho}_p$ (full) from SEAS5/DCSM-5 data of skew surge at Cuxhaven as function of sample fraction $p$ , with their two-sided 95% confidence intervals. Dots indicate the estimated model (3), fitted to the values of $\hat{\rho}_p$ for $p \leq 0.01$ . The indicated sample fractions correspond to 0.007, 0.07, 0.7, 7.0 and 70/year.	30
10	Statistics of return level estimates of high-tide water level at Vlissingen and Hoek van Holland for a return period $R = 10^7$ years. The sample fractions correspond to 0.7, 7.0 and 70/year.	32
11	As Figure 10 for IJmuiden and Den Helder.	33
12	As Figure 10 for Harlingen and Huibertgat.	34
13	As Figure 10 for Delfzijl and Cuxhaven.	35
14	Statistics of return level estimates of high-tide water level for a return period $R = 10^7$ years and a fixed sample fraction of 0.005 (3.5/year).	36
15	As Figure 14 for a sample fraction of 0.01 (7/year).	37
16	As Figure 14 for a sample fraction of 0.05 (35/year).	38
17	Statistics of return level estimates of high-tide water level for a return period $R = 10^4$ years and a fixed sample fraction of 0.005 (3.5/year).	40
18	As Figure 17 for a sample fraction of 0.01 (7/year).	41
19	As Figure 17 for a sample fraction of 0.05 (35/year).	42
20	Bias and RMS error (left and centre) of return level estimates of high-tide water level and skew surge at Vlissingen for a return period $R = 10^7$ years computed as in de Valk and van den Brink (2020b) with an estimated reference return level ("bias-corrected return value") (top 2 rows), and using the Monte-Carlo based method of this report (bottom 2 rows). The indicated sample fractions correspond to 0.7, 7.0 and 70/year.	43
21	Bias and RMS error (left and centre) of return level estimates of simulated high-tide water level and skew surge at Hoek van Holland for a return period $R = 10^7$ years, computed as in de Valk and van den Brink (2020b) with an estimated reference return level ("bias-corrected return value") (top 2 rows), and using the Monte-Carlo based method of this report (bottom 2 rows). The indicated sample fractions correspond to 0.7, 7.0 and 70/year.	44
22	Empirical probabilities of exceedance of high-tide water level (top) and skew surge (bottom) from measurements (thin lines) and SEAS5/DCSM-5 simulations (thick lines) for six tide gauge stations: IJmuiden (orange), Hoek van Holland (blue), Vlissingen (black), den Helder (magenta), Harlingen (green) and Delfzijl (cyan). The indicated fractions of time correspond to 0.0007, 0.07, 7.0, and 700/year.	46
23	GW shape estimate vs. sample fraction with 95% confidence interval from measurements (thin line, light shading) and from SEAS5/DCSM-5 simulations (thick line, dark shading) of high-tide water level (left) and skew surge (right) at three tide gauge stations. The indicated sample fractions correspond to 0.07, 0.7, 7.0 and 70/year.	49

24	As Figure 23 for three other tide gauge stations. .....	50
25	Return level estimates of skew surge for a return period $R = 10^7$ years as function of location index. Estimates are based on sample fractions indicated above the panels (these correspond to frequencies of exceedance of 0.7, 1.4, 3.4, 7, 14 and 35/year). . . . .	58
26	Return level estimates of skew surge for a return period $R = 10^4$ years as function of location index. Estimates are based on sample fractions indicated above the panels (these correspond to frequencies of exceedance of 0.7, 1.4, 3.4, 7, 14 and 35/year). . . . .	59
27	GW-shape parameter estimates $\hat{\rho}_p$ (full) from SEAS5/DCSM-5 data of high-tide water level as function of sample fraction $p$ , with their two-sided 95% confidence intervals, for all stations. Dots indicate the estimated model (3), fitted to the values of $\hat{\rho}_p$ for $p \leq 0.01$ . The indicated sample fractions correspond to 0.007, 0.07, 0.7, 7.0 and 70/year. .....	61
28	Continued from Figure 27. . . . .	62
29	Same as Figure 27, but for skew surge. .....	63
30	Continued from Figure 29. . . . .	64
31	Error statistics of $10^7$ -year return level estimates of wind speed from System 4 and SEAS5 data (top and centre) and pseudo-wind from SEAS5 (bottom) derived directly from the data using the method from de Valk and van den Brink (2020a). The indicated sample frac- tions correspond to 0.7, 7.0 and 70/year. . . . .	66
32	As Figure 31, but derived from Monte-Carlo simulation based on plausible tails (see Section 5). . . . .	67
33	GP shape estimate vs. sample fraction with 95% confidence interval from measurements (thin line, light shading) and from SEAS5/DCSM- 5 simulations (thick line, dark shading) of high-tide water level (left) and skew surge (right) at three tide gauge stations. The indicated sample fractions correspond to 0.07, 0.7, 7.0 and 70/year. .....	69
34	As Figure 24 for three other tide gauge stations. .....	70

---

## List of Tables

1	Approximations of the probability of exceedance $1 - F(z)$ of level $z$ and of the tail quantile $Q(p')$ exceeded with probability $p'$ by five different types of tails (GP, EXP, GW, log-GW, Weibull): $p$ is the probability of exceedance of the fixed threshold $Q(p)$ . $a$ , $f$ and $g$ are positive functions and $\gamma$ , $\rho$ and $\theta$ are real numbers. . . . .	10
2	Estimates of return levels of high-tide water level (m) for a return period of $10^4$ years from different sources (see also Chbab (2017), Table 3.15); the estimates from SEAS5/DCSM5 are from a GW tail fitted to a sample fraction of 0.01 (7/year).  . . . . .	16
3	Increase in return level of high-tide water level (m) in response to an increase $\Delta\rho$ in the GW shape parameter for several return periods (header row); increases are maximized over all tide gauge sites. . . . .	51
4	Increase in return level of skew surge (m) in response to an increase $\Delta\rho$ in the GW shape parameter for several return periods (header row); increases are maximized over all tide gauge sites. . . . .	51

---

# Part I. Report

## 1 Introduction

This report describes a comparison of statistical models and methods for estimating return values of high-tide water level and skew surge along the coast of the Netherlands, undertaken as part of the development of new estimates of the statistics of hydraulic loads on the flood protection infrastructure of the Netherlands.

These statistics include return values of high-tide water level and skew surge corresponding to return periods of up to  $10^7$  years, which is the range relevant for flood protection in the Netherlands (see e.g. the maximum allowed annual probabilities of failure of the flood protection in Waterwet Bijlage II).

The estimates currently in use (Chbab, 2017) are based on measurement records covering less than 150 years, which limits their precision. Furthermore, the methods employed involve assumptions and simplifications needed to obtain estimates at sites where no measurements are available and to extrapolate to very high return periods. As a result, the estimates may be biased and physically inconsistent. The resulting uncertainty in return values is particularly high for the highest return periods in the range considered.

To improve the accuracy of the statistics of wind, wind stress, water level and surge, we are currently exploring the use of very large datasets of simulated weather and water levels; in particular,

- the SEAS5 seasonal ensemble weather reforecast archive of ECMWF, containing for each of its grid points effectively 8000 years of weather conditions representative for the present climate. and
- water level simulations created by running the DCSM-5 shallow-water flow model with surface stress and pressure forcing from the SEAS5 data; see van den Brink (2020).

Together, these two datasets will be referred to as the SEAS5/DCSM-5 dataset.

Furthermore, we explore the use of new statistical models for estimating return values, making use of the SEAS5/DCSM-5 dataset to compare statistical models for estimating return values of high-tide water level and skew surge, as we did earlier in de Valk and van den Brink (2020a,b) for wind and pseudo-wind speed.

The statistical models considered are parametric approximations of the tail of a distribution function for values above some threshold value, based on extreme value theory de Haan & Ferreira (2006). The purpose of these models is to provide a stable extrapolation of the distribution function beyond this threshold, to determine return values for very high return periods. An overview of these tail models and their background is given in de Valk and van den Brink (2020a) (see also Gardes and Girard (2016); de Valk (2016a,b)). They are

- the Generalized Pareto (GP) tail,
- the Generalized Weibull (GW) tail,

- the log-Generalized Weibull (GW) tail,
- the Weibull tail (a special case of the former two), and
- the exponential (EXP) tail (a special case of all the former).

Table 1 gives an overview. In this table,  $F$  is the distribution function of high-tide water level or skew surge ( $F(z)$  being the mean fraction of high waters in which the level  $z$  is not exceeded),  $Q$  is the inverse of  $1 - F$  (so  $1 - F(Q(p')) = p'$  for all  $p' \in [0, 1]$ ), and  $p$  is the probability of exceedance of a threshold value  $Q(p)$  (so  $p$  is small, but large enough that  $Q(p)$  can be estimated accurately by an empirical quantile).

Tail	$1 - F(z) =$	$Q(p') = (1 - F)^{-1}(p') =$
GP	$p \left( 1 + \gamma \left( \frac{z - Q(p)}{a(p)} \right) \right)^{-1/\gamma}$	$Q(p) + a(p) \frac{1}{\gamma} \left( \left( \frac{p'}{p} \right)^\gamma - 1 \right)$
EXP	$p e^{(Q(p) - z)/a(p)}$	$Q(p) + a(p) \log \left( \frac{p'}{p} \right)$
GW	$p^{(1 + \rho \left( \frac{z - Q(p)}{f(p)} \right))^{1/\rho}}$	$Q(p) + f(p) \frac{1}{\rho} \left( \left( \frac{\log p'}{\log p} \right)^\rho - 1 \right)$
log-GW	$p^{(1 + \theta \left( \frac{\log z - \log Q(p)}{g(p)} \right))^{1/\theta}}$	$Q(p) \exp \left( g(p) \frac{1}{\theta} \left( \left( \frac{\log p'}{\log p} \right)^\theta - 1 \right) \right)$
Weibull	$p \left( \frac{z}{Q(p)} \right)^{1/\rho}$	$Q(p) \left( \frac{\log p'}{\log p} \right)^\rho$

Tab. 1: Approximations of the probability of exceedance  $1 - F(z)$  of level  $z$  and of the tail quantile  $Q(p')$  exceeded with probability  $p'$  by five different types of tails (GP, EXP, GW, log-GW, Weibull):  $p$  is the probability of exceedance of the fixed threshold  $Q(p)$ .  $a$ ,  $f$  and  $g$  are positive functions and  $\gamma$ ,  $\rho$  and  $\theta$  are real numbers.

Note that the Weibull and the exponential tails are 1-parameter tails (not counting the threshold  $Q(p)$  in the formulas in Table 1), and the GP, GW and log-GW tails are 2-parameter tails.

As in de Valk and van den Brink (2020a), we apply two different approaches to compare the performances of the tail models: one approach (Chapter 4) is based on the statistics of maxima of high-tide water level or skew surge over 78-year subsamples of the SEAS5/DCSM-5 data, and the other approach (Chapter 5) focuses directly on the error statistics of return values of high-tide water level and skew surge estimated from these subsamples and/or from the full SEAS5/DCSM-5 dataset. However, in this report, a different method is used to assess these error statistics. A comparison is made with the results of the method of de Valk and van den Brink (2020a) to demonstrate that both methods are compatible, but that the new method works better.

Besides providing guidance on the type of tail to use for extrapolation, the analysis offers other valuable insight, for example about the potential increase in accuracy if the shape parameter of the GP, GW or log-GW tail is estimated from the full SEAS5/DCSM-5 dataset instead of from a subsample. This gives a conservative assessment of the potential increase in precision due to the use of SEAS5/DCSM-5 data for these tails: it is foreseen that the location and scale parameters will be estimated after calibration of the SEAS5/DCSM-5 data to high-resolution simulations

of a limited subset of storms from the SEAS5/DCSM-5 archive, so the statistical extrapolation after this calibration may be somewhat less precise than before (which is the price to be paid for the expected reduction in simulation model bias).

In the comparison of the performances of tail models, we ignore simulation model bias, defined as systematic error in the distributions of SEAS5/DCSM-5 model output variables.

Simulation model bias is analyzed separately, focusing on the shape parameters of the GW and GP tails (as simulation model bias in the scale and location parameters is expected to be corrected by calibration on high-resolution simulations). Shape estimates from these data are compared to estimates from measurement data for six tide gauge stations, and an assessment is made of the order of magnitude of the bias in return values resulting from simulation model bias in the shape parameter. Possible implications for the use of calibrated SEAS5/DCSM-5 data for estimation of return values for high return periods are indicated.

The discussions of the results are included in the chapters.

Deze studie is uitgevoerd binnen een Maatwerk I&W opdracht onder begeleiding van Marcel Bottema en Robert Slomp (WVL). Prof. dr. Pieter van Gelder heeft een review van dit rapport uitgevoerd waarvan dankbaar gebruikt is gemaakt voor de eindversie van dit rapport en voor vervolgstudies.

## 2 The SEAS5/DCSM-5 dataset

The bulk of the archive of SEAS5 weather data consists of seasonal ensemble reforecasts computed starting at the first day of every month in 1981-2016, each forecast running over at least 7 months. In addition, operational forecasts are available from 2017 to present.

The output time step is 6 hours. Further information about the SEAS5 data can be found in [Implementation of Seasonal Forecast SEAS5](#) (ECMWF, 2018a) and [SEAS5 User Guide](#) (ECMWF, 2018b).

Water levels were computed by the DCSM-5 flow model for shallow tidal seas, forced by pressure and wind from SEAS5; see van den Brink (2020). High-tide water level maxima were extracted from these data. In addition, the skew surge was determined by the method implemented in the DCSM-5 model.

No corrections for limitations on resolution etc. (van den Brink, 2020) have yet been applied to the SEAS5/DCSM-5 data.

For the SEAS5 reforecast data, the ensemble size is 25 and for Feb, May, Aug and Nov, the size is 50. The ensemble size is also 50 for the operational forecasts. Only 25 ensemble members of each forecast were used in order to retain the same data volume for every year/month (out of caution, but this may not have been necessary). Furthermore, data of the first month of each forecast run were discarded to ensure independence among the members of an ensemble. This resulted in about 5800 years of data for each station. Subsequently, 150 subsamples of these data were generated by combining ensemble members with the same ensemble member label and with starting dates at regular 6-month intervals. Pairs of subsamples were combined, resulting in 75 subsamples, each one covering about 78 years.

In this study, we analyse high-tide water level and skew surge data computed for the 21 coastal tide-gauge stations Duinkerke, Oostende, Zeebrugge, Cadzand, Vlissingen, Westkapelle, OS11, Roompot buiten, Lichteiland Goeree, Hoek van Holland, Scheveningen, Meetpost Noordwijk, IJmuiden, Den Helder, Texel Noord, Harlingen, West Terschelling, Huibertgat, Delfzijl, Cuxhaven, and Esbjerg. These stations are aligned along the southern and eastern boundaries of the North Sea from N. France to Denmark; see Figure 1.

For each station, Figure 2 shows the difference between the empirical quantiles of water level and skew surge as a function of the probability of exceedance. As observed in van den Brink (2020)[Figure 9], these are fairly close to constant, but in particular for sites to the south of Den Helder, some curvature can be seen. A possible explanation for this curvature is a systematic dependence between the skew surge (high-tide water level minus undisturbed tidal maximum) and the height of the undisturbed tidal maximum, due to surge-tide interaction (for example: high astronomical maxima are reduced more by the surge than low astronomical maxima). The effect should be higher in the south, where the tidal range is larger and storm surges tend to be lower (see Appendix A). This is indeed observed in Figure 2.

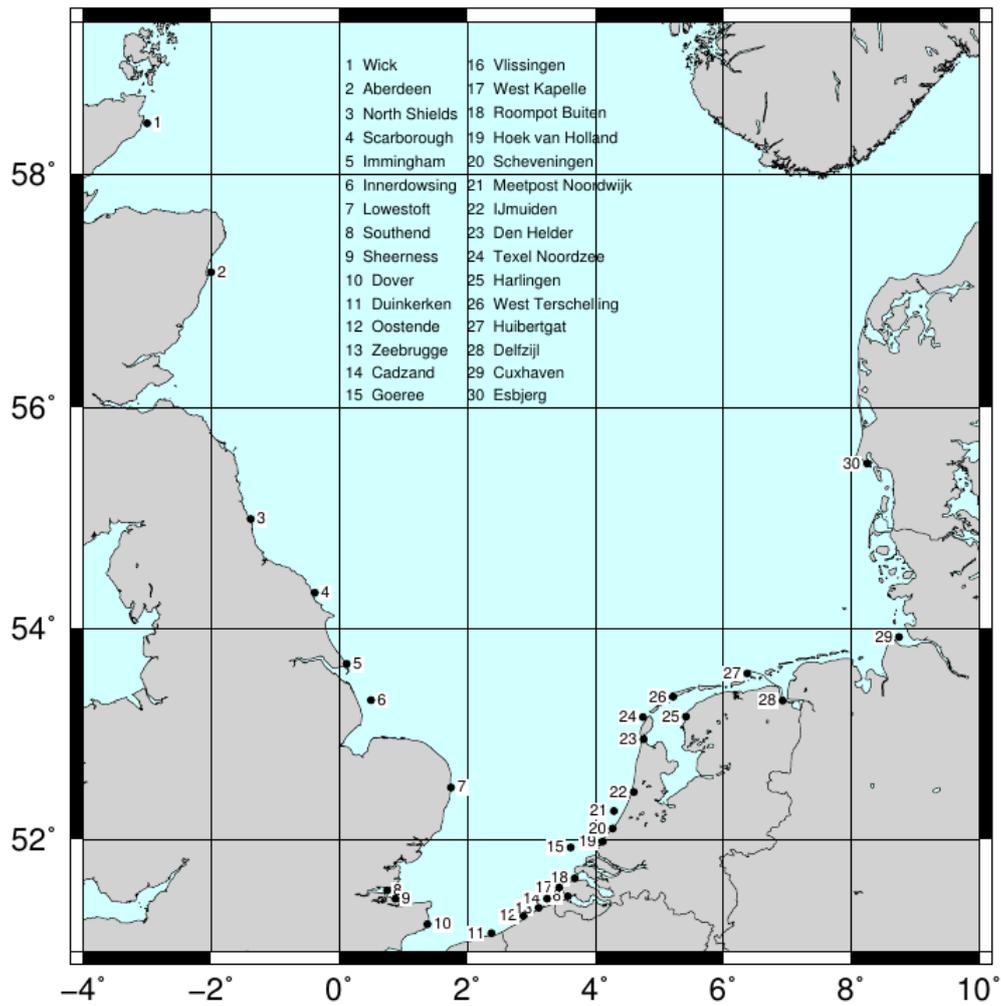


Fig. 1: Locations of coastal tide gauge stations in the North Sea. Station OS11 is close to Roompot.

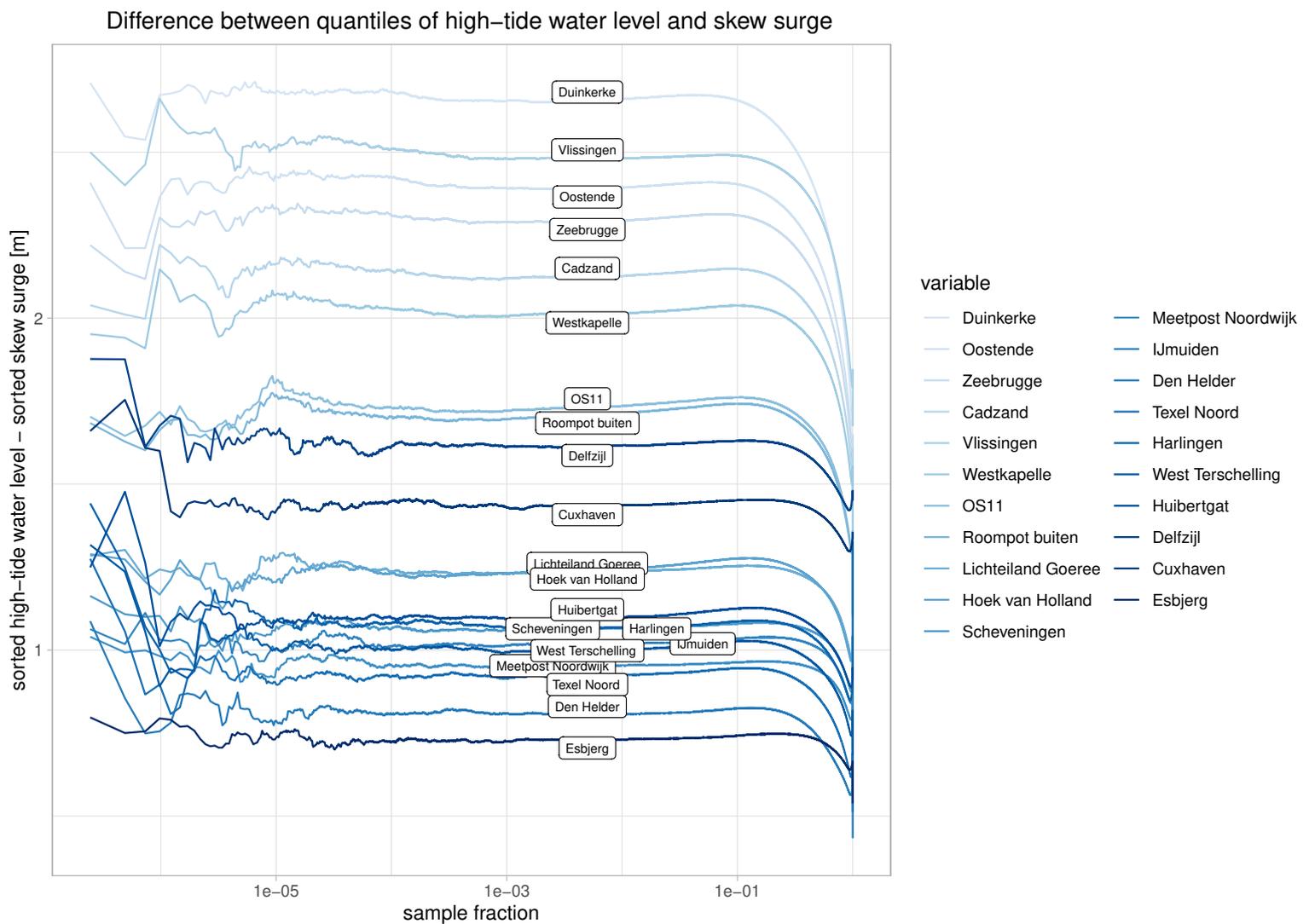


Fig. 2: Difference between the quantiles of high-tide water level and skew surge from SEAS5/DCSM-5 at the 21 tide gauge stations as functions of the sample fraction exceeding the quantile; the indicated sample fractions correspond to 0.007, 0.7 and 70/year.

### 3 Comparison of estimates from the full sample

Estimates of return levels of water level were made for return periods of  $10^7$  years and of  $10^4$  years

- a) directly from high-tide water level data, and
- b) indirectly, from data of skew surge, with the return levels of skew surge converted to water level by adding a site-dependent constant determined from Figure 2 at a sample fraction of 0.005 (3.5/year), in the range where the curves are relatively stable and precise.

Figure 3 shows these estimates for a return period of  $10^7$  years and for 6 different threshold probabilities (sample fractions) as functions of the location index from South (Duinkerke) to North (Esbjerg). The estimates for a return period of  $10^4$  years are shown in Figure 4.

The patterns in the estimates and their dependence on location, data source (water level or skew surge) and tail type are similar for both return periods, but the estimates for a return period of  $10^4$  years show less variation when varying tail type and data source. Therefore, we will focus on the return period of  $10^7$  years.

Return level estimates based on corrected skew surge data (dashed) and water level data (full) are very different for the Weibull tail. As for wind (de Valk and van den Brink, 2020a), this tail is clearly too rigid and cannot adapt sufficiently to the data. For the exponential tail, the estimates based on water level data and surge data agree much more closely. Among the 2-parameter tails, only the log-GW tail shows large differences between the estimates based on water level data and surge data; they are of the order of 0.5-1.0 m for threshold sample fractions of 0.001-0.02 (frequencies of exceedance of 0.7-14/year). This is not surprising: the difference between the tails of water level and skew surge is approximately a shift, but the log-GW tails are not shift-invariant<sup>1</sup>, unlike the GW and GP tails. For the GW and GP tails (which are shift-invariant 2-parameter tails), the estimates based on water level data and surge data agree quite closely for sample fractions of 0.005-0.02. At lower sample fractions and in the South, they deviate more. The differences between GW and GP-based estimates is somewhat larger in the Waddenzee than further to the West and South. This is likely related to the larger curvatures of the tails for stations in the Waddenzee; see Chapter 6. For all tails except the Weibull, the estimates vary considerably with sample fraction in the range 0.01-0.05, but are rather stable for sample fractions of 0.01 and lower. From the Flemish coast up to Hoek van Holland/IJmuiden, the estimates based on GW and GP tails are similar to the estimates based on the exponential tail. This is not the case further North, where the exponential gives higher return values.

Generally, return value estimates of high-tide water level based on skew surge data are lower than estimates derived directly from water level data (but for the

---

<sup>1</sup> A shift-invariant family of tails has the property that if the random variable (e.g. water level) is expressed relative to a different offset (e.g. mean sea level instead of NAP), its tail still belongs to the same family of tails.

Source	Vlissingen	Hoek v H	Den Helder	Harlingen	Delfzijl
Dillingh et al (1993)	5.55	4.75	4.15	4.40	6.15
Eilander (2012)	5.32	4.66	3.76	4.44	6.16
SEAS5/DCSM5	5.48	4.26	3.79	4.56	5.62

Tab. 2: Estimates of return levels of high-tide water level (m) for a return period of  $10^4$  years from different sources (see also Chbab (2017), Table 3.15); the estimates from SEAS5/DCSM5 are from a GW tail fitted to a sample fraction of 0.01 (7/year).

Weibull tail, it is the opposite). The reason is dependence between the skew surge and the height of the astronomical maximum due to surge-tide interaction, as explained in Section 2. Indeed, as expected, the largest differences are found for stations in the South, where the tidal range is high. This could be a reason to prefer direct estimates from from water level data. For the log-GW tail and in particular, the Weibull tail, the differences between estimates derived from water level data and from surge data are unrealistically large, in particular at the higher sample fractions (these differences should be considerably smaller than the tidal amplitude). This indicates serious bias.

In van den Brink (2020), the empirical distributions of annual maxima of skew surge and water level from the same SEAS5/DCSM-5 dataset are compared to distributions of annual maxima of water level and skew surge from measurements. For most stations, the estimates from SEAS5/DCSM-5 data are considerably lower than those from measurements.

In Table 2, the estimates of return levels for a return period of  $10^4$  year in Figure 4 for five stations are compared to two different reference estimates from Chbab (2017), Table 3.15. For the stations Vlissingen, Den Helder and Harlingen, estimates from SEAS5/DCSM5 data are compatible with the estimates from the other studies. For Hoek van Holland and Delfzijl, estimates from SEAS5/DCSM5 data are up to 10% lower. Possibly, the limited resolution of DCSM5 has a relatively large impact at these two sites.

All together, we may conclude that the estimates from SEAS5/DCSM5 are not far off the earlier estimates from measurement data. It is too early for a closer examination of the differences, since we have not yet addressed corrections of the SEAS5/DCSM5 data for resolution effects etc. (van den Brink, 2020).

In Appendix A), return value estimates are shown for skew surge, prepared in the same way as for high-tide water level.

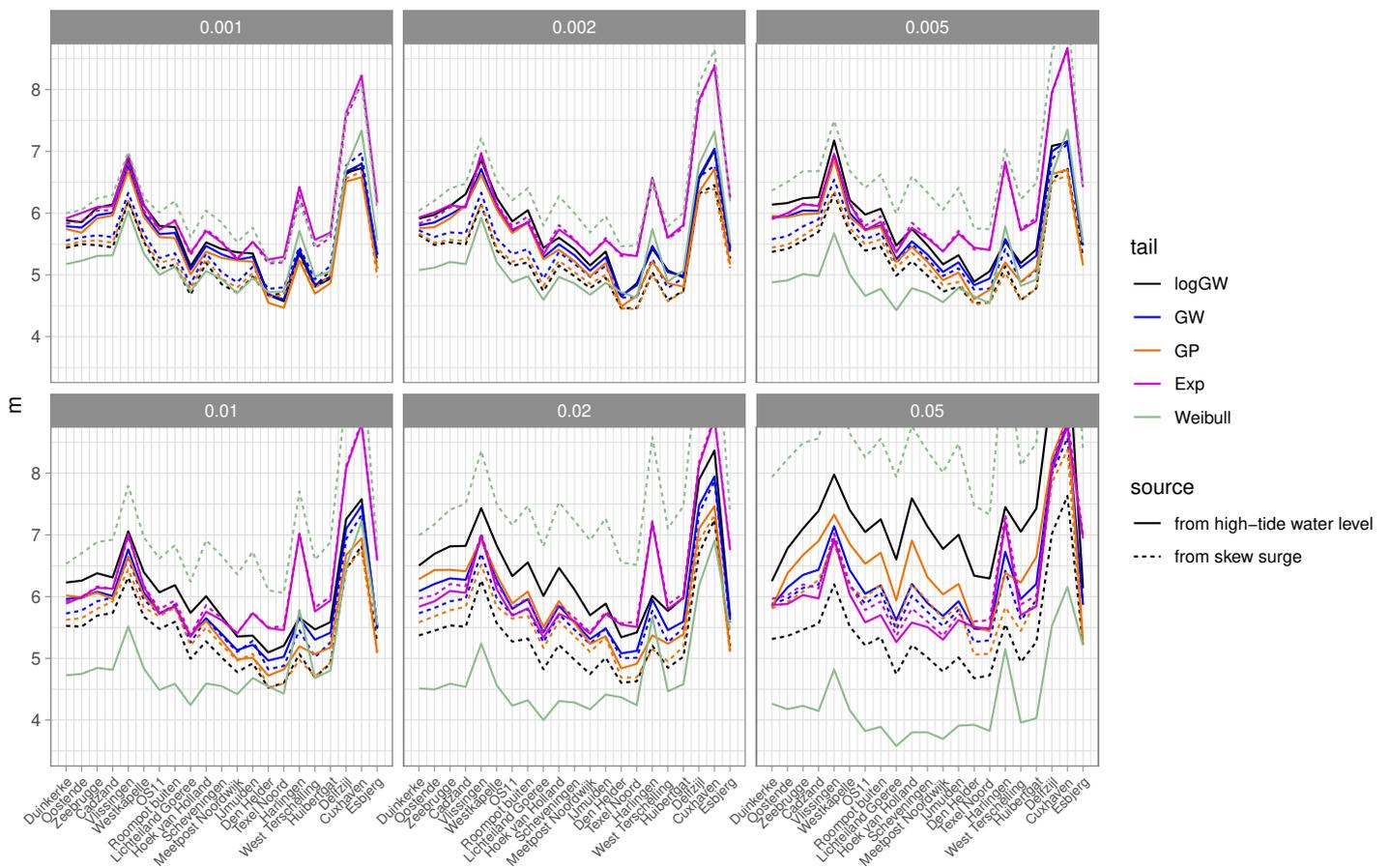


Fig. 3: Return level estimates of high-tide water level for a return period  $R = 10^7$  years as function of location index. Estimates are based on sample fractions indicated above the panels (these correspond to frequencies of exceedance of 0.7, 1.4, 3.4, 7, 14 and 35/year).

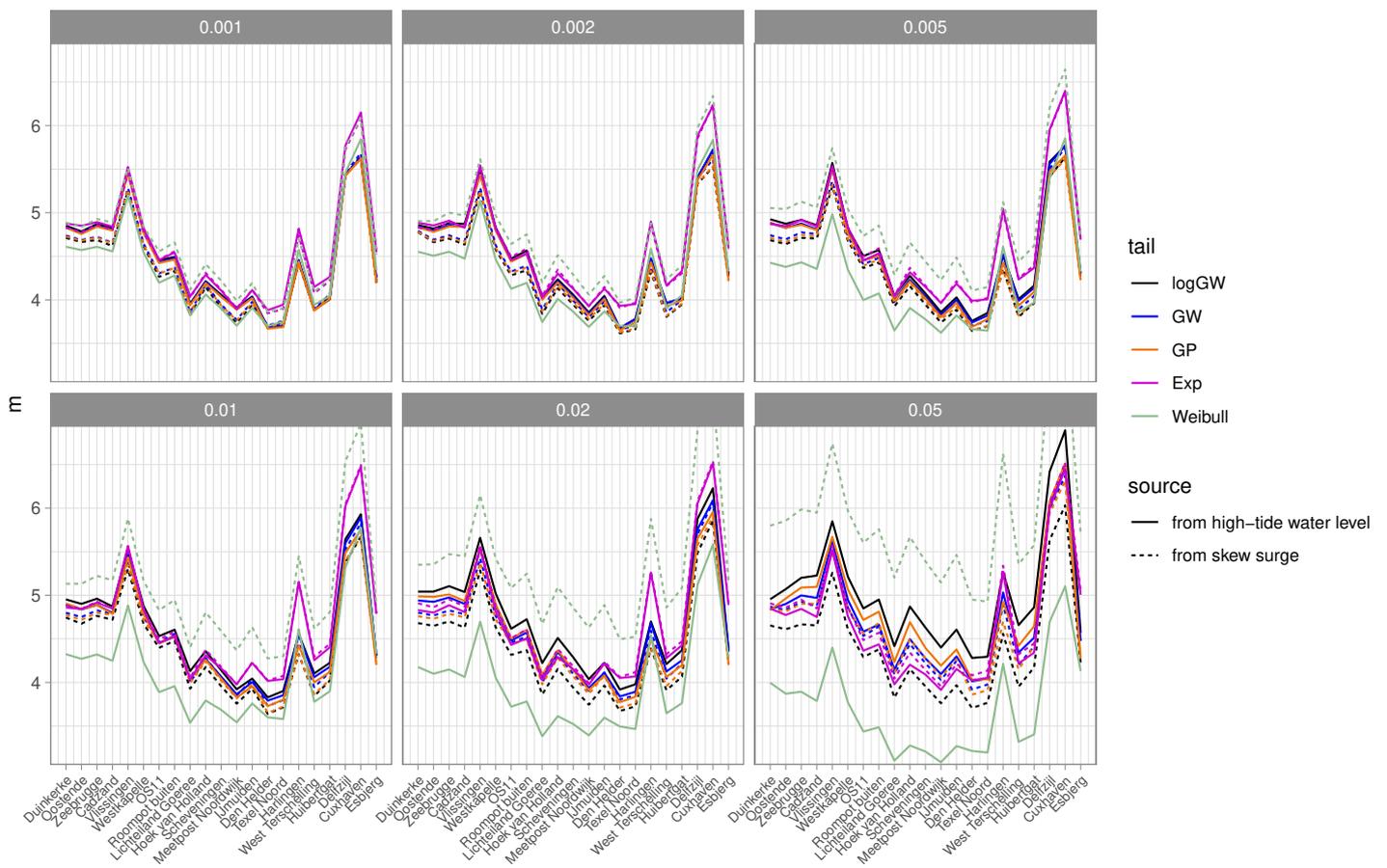


Fig. 4: Return level estimates of high-tide water level for a return period  $R = 10^4$  years as function of location index. Estimates are based on sample fractions indicated above the panels (these correspond to frequencies of exceedance of 0.7, 1.4, 3.4, 7, 14 and 35/year).

## 4 Statistics of transformed maxima over subsamples

A first check of the extrapolation skills of different tail models is from van den Brink and Können (2008, 2011). With a chosen tail model, we fit the empirical tail of the seasonal wind speed forecasts, using the upper  $100p\%$  of values. This tail fit allows us to transform the upper  $100p\%$  of values in each of the  $m = 75$  subsamples monotonically to render their tail approximately standard exponential. If the fitted tail matches the true tail closely, then the distribution function of the transformed wind speed should be close to the standard exponential distribution. Let  $n$  be the size of a subsample, and  $Y$  the highest transformed wind speed in the subsample minus  $\log(np\alpha)$ , with  $\alpha$  the extremal index (e.g. de Valk and van den Brink (2020a), Ch 2). For high tide water level and skew surge, a value of 1 was found for  $\alpha$ . Then the distribution function of  $Y$  should be close to a standard Gumbel distribution: for  $np$  large,  $\mathbb{P}(Y \leq x) \approx (1 - \exp(-x - \log(np\alpha)))^{np\alpha} \approx \exp(-\exp(-x))$ .

Since we have  $m$  subsamples, we can check whether this is true. Let  $Y_1, \dots, Y_m$  be the highest transformed wind speeds from each of the  $m$  subsamples; they can be sorted to obtain the order statistics  $Y_{1:m} \leq \dots \leq Y_{m:m}$ . Then the plot of  $Y_{i:m}$  against  $-\log(-\log(i/(m+1)))$  for  $i = 1, \dots, m$  (known as the Gumbel plot) should be close to a straight line through the origin with slope 1. Furthermore, assuming that  $Y$  has an exact Gumbel distribution and that  $Y_1, \dots, Y_m$  are independent, pointwise confidence bounds can be derived from e.g. Czörgő & Révész (1978). The accuracy of the approximation provides a measure of the accuracy of the fitted tail used to transform the data to unit exponential.

The method described above differs from the one applied in de Valk and van den Brink (2020a) in that there, the tails of individual subsamples were fitted, whereas here, the fit was computed from the full sample. The latter method is valid (bias in the estimates from subsamples and from the full sample is virtually identical), and is much less affected by noise.

Figure 5 shows the results for a sample fraction  $p = 0.036$  (the values exceeded less than 25/year). It should be kept in mind that the results for different stations are not independent; they are generated by surge events affecting the whole southern North Sea. For most stations and tails, a substantial part of the Gumbel plot is outside the pointwise 95% confidence interval. This indicates that none of these tails fits the data very well over the entire range from the quantile with probability of exceedance  $p = 0.036$  to the maximum over the subsample. However, the sample fraction  $p = 0.036$  is rather large, so this is not necessarily a negative result. The plots for the Weibull tail deviate far from the diagonal in almost all cases. For the exponential tail, the deviation is substantial in particular for surge at stations in the North.

With a lower sample fraction  $p = 0.007$  (the values exceeded less than 5/year), the diagonal is approximated much more closely; see Figure 6. Values a little outside the 95% confidence interval are observed mainly for Weibull and water level in the South, for Weibull and surge in the North, and for the exponential in the North (both water level and surge). For the other tails, the curves are mostly within the confidence intervals.

The approximations are even better with a lower sample fraction  $p = 0.0014$ (the

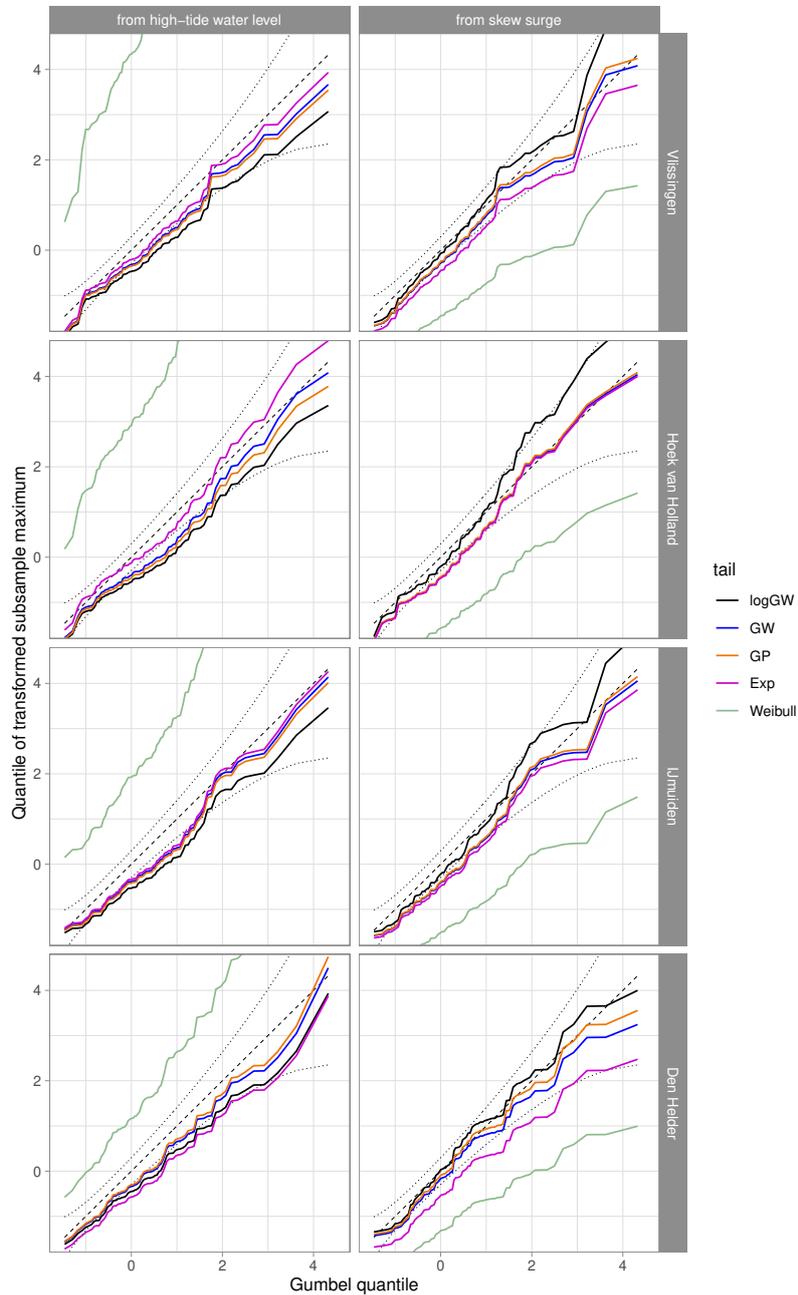


Fig. 5: Gumbel plots of transformed maxima over the subsamples of the data of water level (left) and surge (right), for different tails (see legend) and stations (rows). Dashed: standard Gumbel line; dotted: 95% confidence bounds. Tail estimates based on values exceeded 25/year (sample fraction of 0.036).

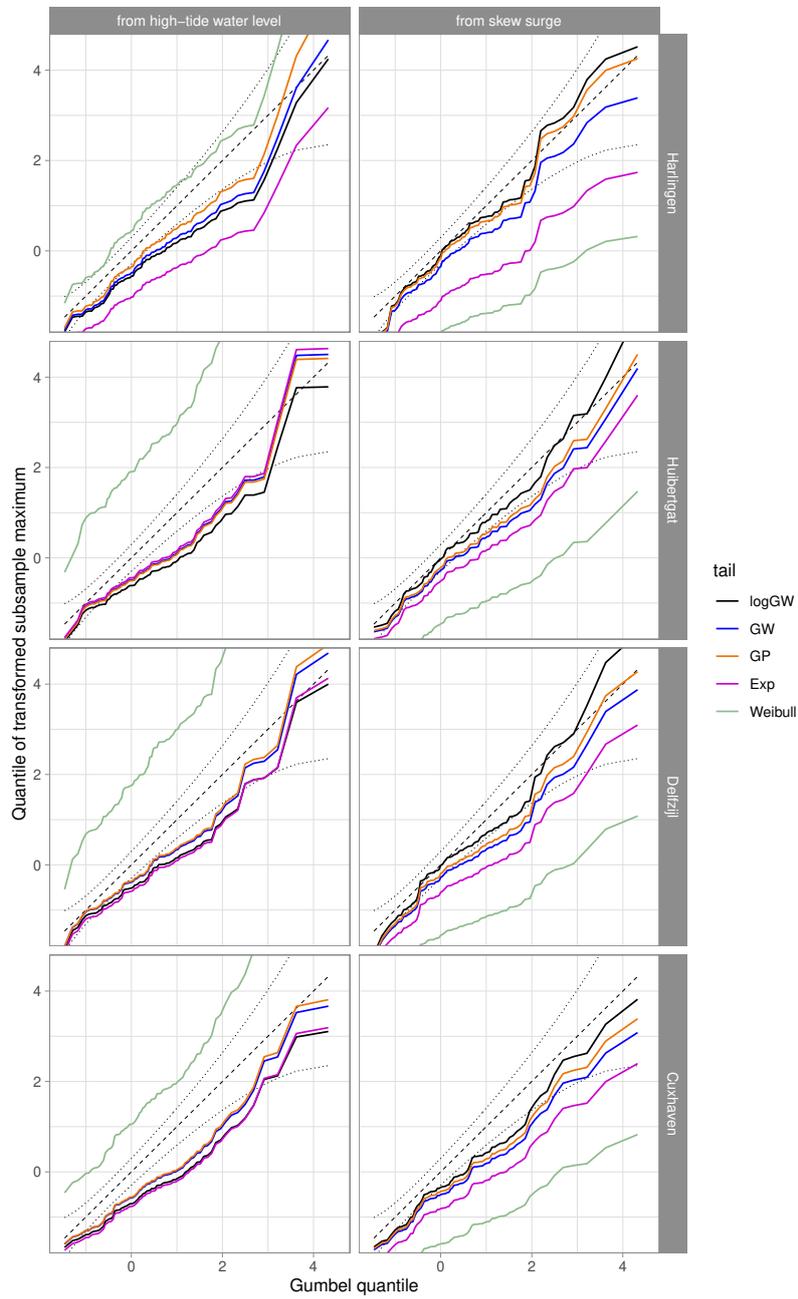


Fig. 5: Continued from last page.

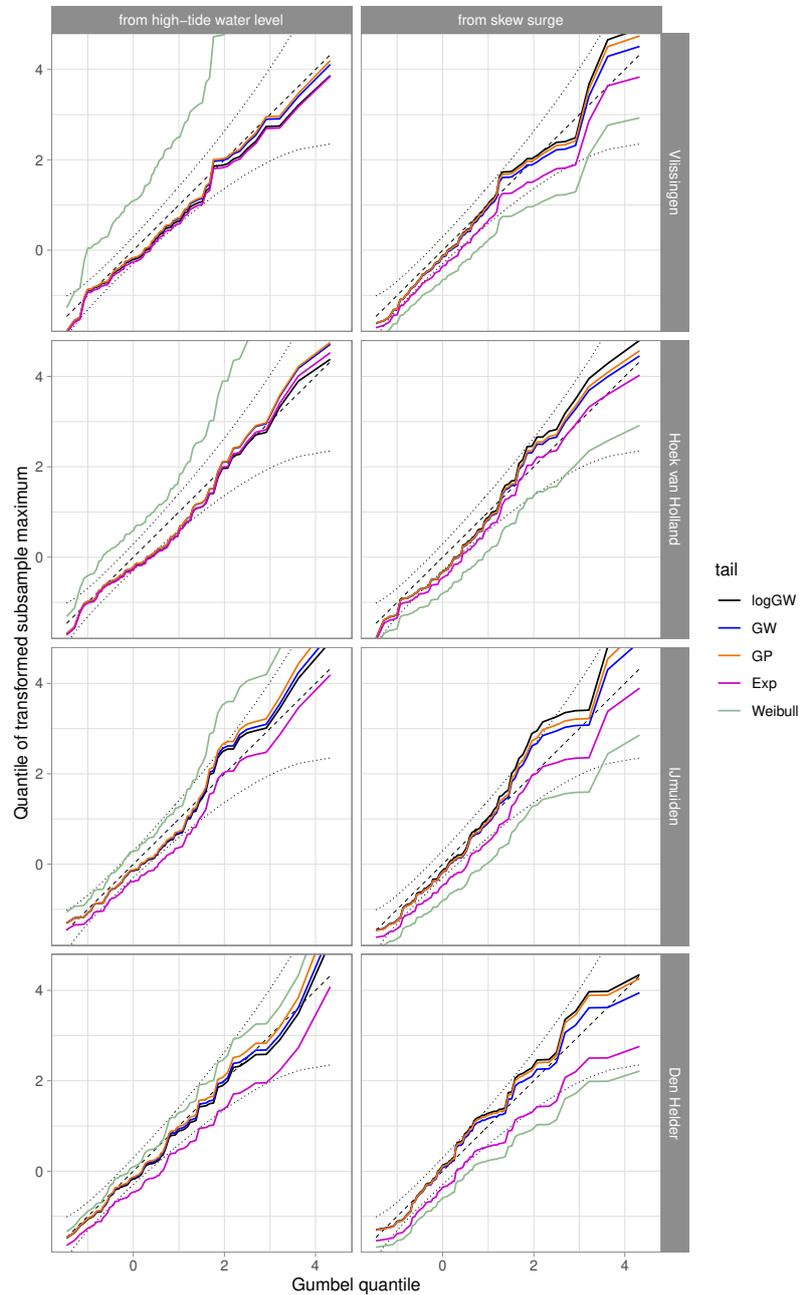


Fig. 6: Gumbel plots of transformed maxima over the subsamples of the data of water level (left) and surge (right), for different tails (see legend) and stations (rows). Dashed: standard Gumbel line; dotted: 95% confidence bounds. Tail estimates based on values exceeded 5/year (sample fraction of 0.007).

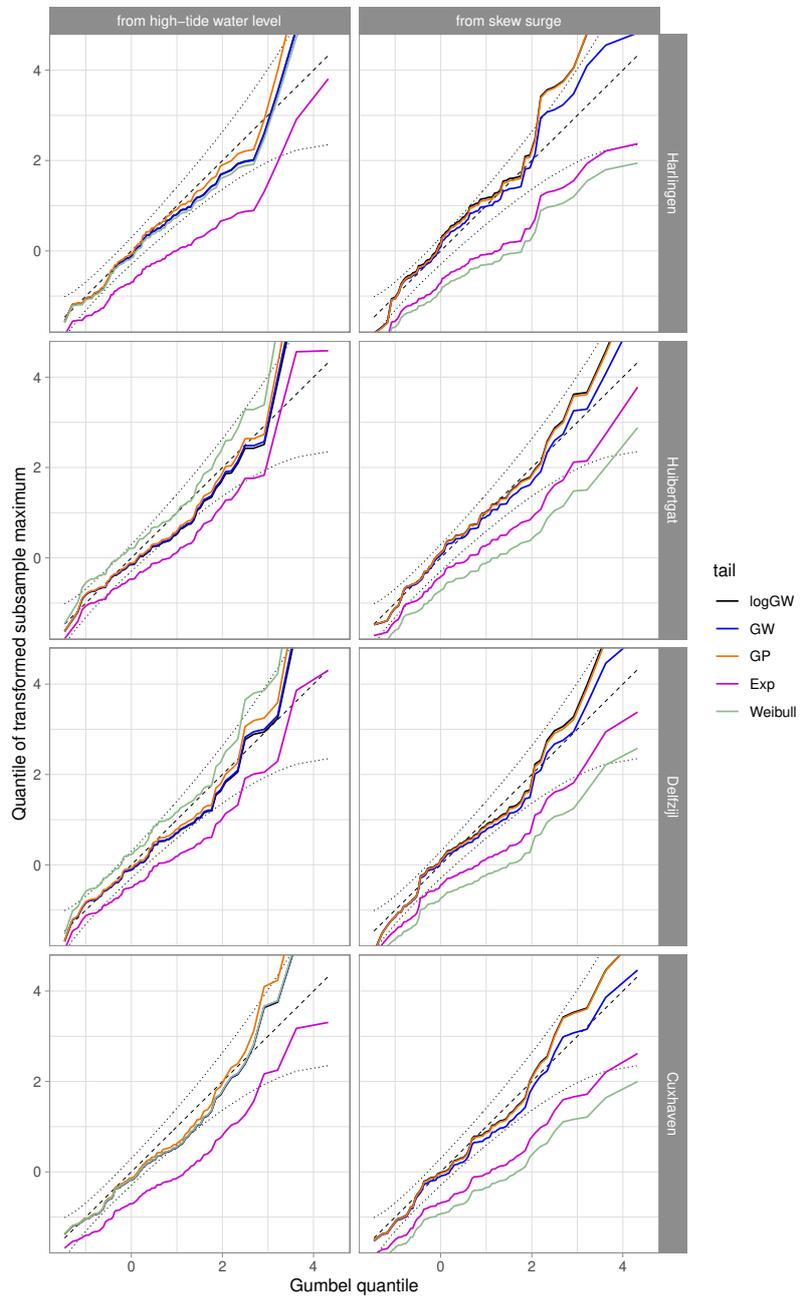


Fig. 6: Continued from last page.

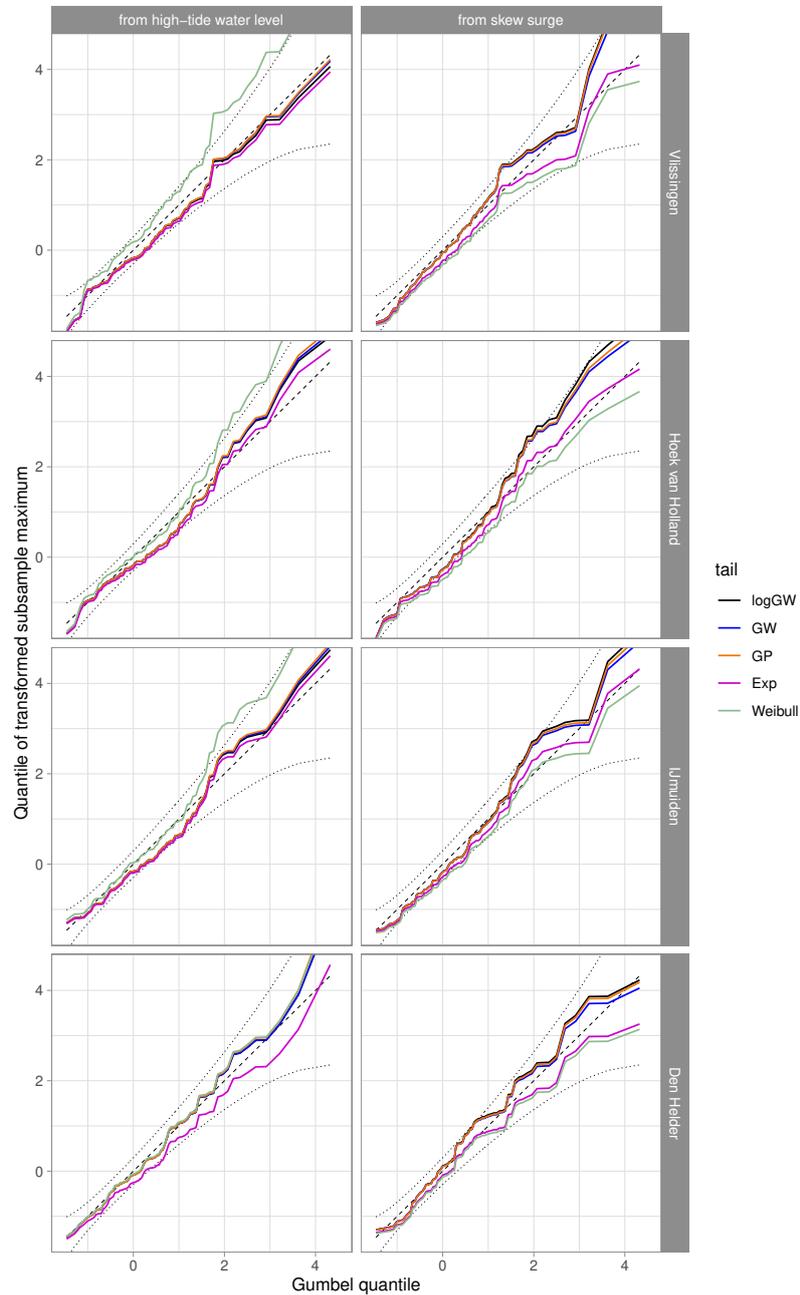


Fig. 7: Gumbel plots of transformed maxima over the subsamples of the data of water level (left) and surge (right), for different tails (see legend) and stations (rows). Dashed: standard Gumbel line; dotted: 95% confidence bounds. Tail estimates based on values exceeded 1.0/year (sample fraction of 0.0014).

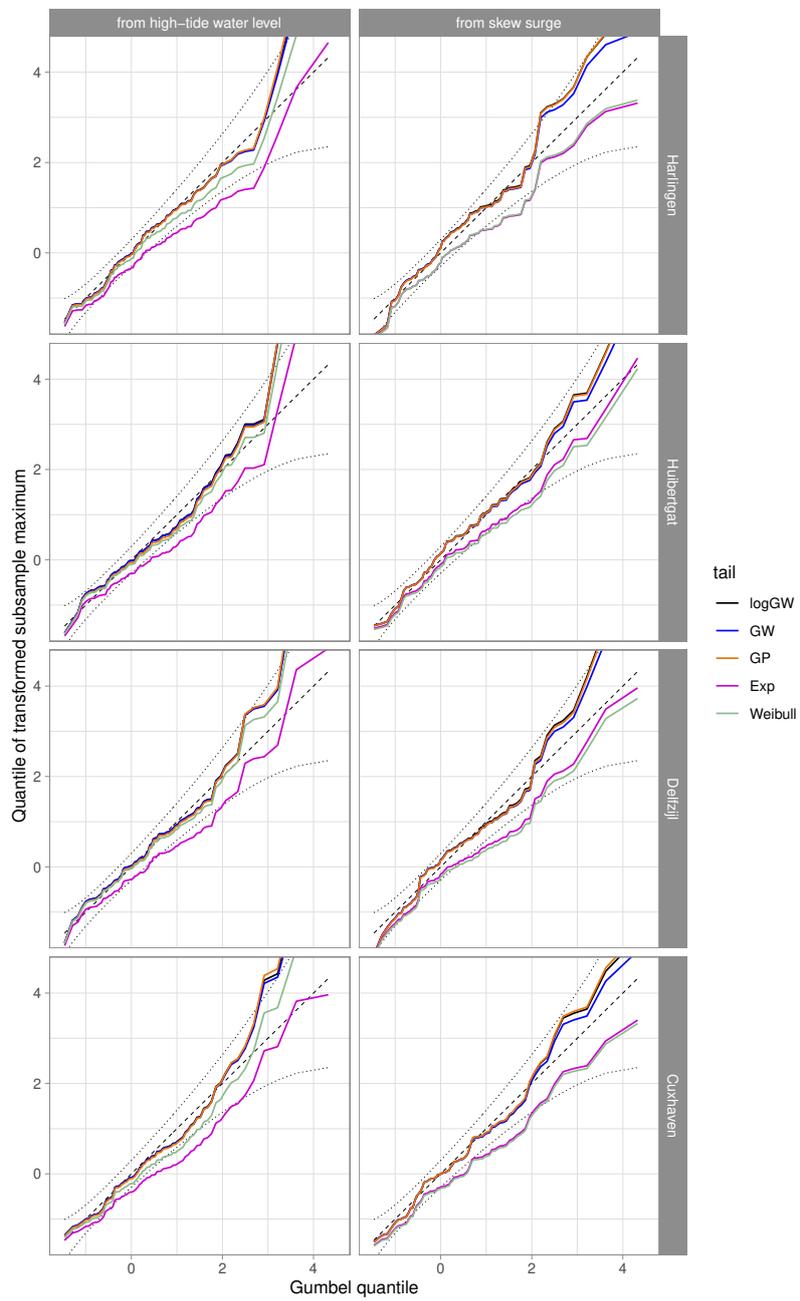


Fig. 7: Continued from last page.

values exceeded less than 1/year); see Figure 7. Part of this is explained by the fact that with a lower sample fraction  $p$ , the subsample-maxima are closer to the threshold exceeded by this sample fraction, so we are testing extrapolation to a larger fraction of  $p$ ;  $0.013p$  in the latter case.

We may conclude that 2-parameter tails (GP, GW and log-GW) are performing satisfactorily for  $p \leq 0.01$ ; the 1-parameter tails (Weibull and exponential) show clear biases.

## 5 Checking the extrapolation from subsamples by Monte-Carlo simulation from plausible distribution functions

### 5.1 Introduction

In de Valk and van den Brink (2020a), the bias in return value estimates of (pseudo) wind speed from subsamples is estimated from the difference between their mean value and a reference estimate derived from the full sample at a much higher threshold. This bias estimate is used for the estimation of the root mean square (RMS) error.

Here, a modified method is applied which is less sensitive to noise, and provides estimates of bias relative to an absolute reference value. For a given distribution function  $F$ , the error statistics of quantile estimates based on different types of tails can be determined straightforwardly by Monte Carlo simulation. The advantage of this approach is that the reference quantile is known exactly, as was stressed by prof. Pieter van Gelder in his review of de Valk and van den Brink (2020a). However, the results of Monte Carlo simulation depend sensitively on the choice of the distribution function(s), and therefore often tend to be subjective and ambiguous.

To avoid this, we construct plausible tails representing the tails of high tide water level and skew surge at all tide gauge stations considered in this study. These tails are derived from the estimates of the GW, log-GW or GP shape parameter for all threshold probabilities  $p$  obtained from SEAS5/DCSM5 data. This is feasible thanks to the large size of the SEAS5/DCSM5 dataset available for each tide gauge station<sup>2</sup>.

### 5.2 Method

Suppose for simplicity that the distribution function  $F$  is smoothly increasing, and let  $q$  be the quantile function defined as the function satisfying

$$1 - F(q(y)) = e^{-y}$$

for every  $y \in [0, \infty)$  (so  $q(y) = Q(e^{-y})$  with  $Q$  defined in Section 1).

Define the function  $\tilde{\rho}$  by

$$\tilde{\rho}(y) = y(\log(yq'(y)))', \quad (1)$$

with  $'$  indicating differentiation with respect to  $y$ ;  $\tilde{\rho}$  is a dimensionless measure of curvature. If  $\tilde{\rho}(y)$  tends to a constant  $\rho$  when  $y \rightarrow \infty$ , then the distribution function  $F$  satisfies a GW tail limit with GW shape parameter equal to  $\rho$ . However, we can model  $\tilde{\rho}$  regardless of whether a GW tail limit applies or not, as the quantile function  $q$  can be determined by solving (1). Moreover, we can estimate  $\tilde{\rho}$  regardless of whether a GW tail limit applies or not: most GW shape estimators  $\hat{\rho}_p$  at a threshold exceeded by the fraction  $p$  of the sample are really estimators for  $\tilde{\rho}(\log(1/p))$ , so applying them for different  $p$ , we can estimate the function  $\tilde{\rho}$ .

<sup>2</sup> The next section is rather technical; it may be skipped or browsed in a first reading.

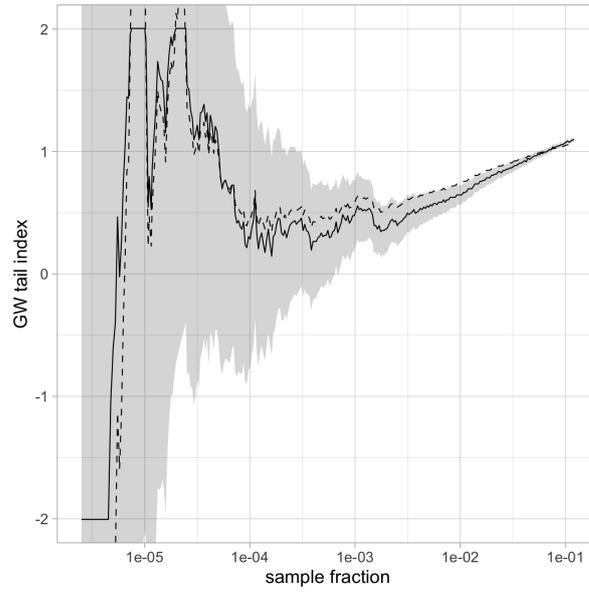


Fig. 8: GW-shape parameter estimates  $\hat{\rho}_p$  from SEAS5/DCSM-5 data of skew surge at Cuxhaven as function of sample fraction  $p$  (full), with their two-sided 95% confidence intervals. The dashed line is  $\hat{\gamma}_p(\log 1/p) + 1$ , with  $\hat{\gamma}_p$  the GP- shape parameter estimates from the same data. The indicated sample fractions correspond to 0.007, 0.07, 0.7, 7.0 and 70/year.

Alternatively, we can consider another type of dimensionless curvature  $\tilde{\gamma}$ , defined by

$$\tilde{\gamma}(y) = (\log(q'(y)))' = (\tilde{\rho}(y) - 1)/y. \quad (2)$$

If  $\tilde{\gamma}(y)$  tends to a constant  $\gamma$ , then  $F$  satisfies as GP tail limit with shape parameter equal  $\gamma$ . But again, this is not the issue here.

In principle, it does not matter whether the function  $\tilde{\rho}$  or the function  $\tilde{\gamma}$  is used to describe the quantile function  $q$ ; either will do.

From (2), we see that  $\tilde{\rho}(y) - 1$  is the magnification of  $\tilde{\gamma}(y)$  by a factor  $y$ . This is illustrated by the estimates  $\hat{\rho}_p$  of  $\tilde{\rho}(\log 1/p)$  (full line) and  $\hat{\gamma}_p/(\log 1/p) + 1$  of  $(\log 1/p)\tilde{\gamma}(\log 1/p)$  (dashed line) in Figure 8, derived from the SEAS5/DCSM-5 data of skew surge at Cuxhaven<sup>3</sup>. Therefore,  $\tilde{\rho}$  is more suitable for modelling of the finer details of the tail than  $\tilde{\gamma}$ , as the former magnifies these details.

The following procedure is used to determine a plausible quantile function  $q$ :

<sup>3</sup> The small differences between the two curves are due to different types of estimators being used, so the estimates do not exactly satisfy (2).

1. fix a threshold probability  $p_0$  such that  $q(y)$  can be determined accurately for all  $y \in [0, \log 1/p_0]$  directly from the empirical quantiles with probabilities of exceedance  $[p_0, 1]$ ; let  $y_0 := \log(1/p_0)$ ;
2. estimate a plausible profile for  $\tilde{\rho}(y)$  for all  $y \geq y_0 := \log(1/p_0)$  from the estimates  $\hat{\rho}_p$  for  $p \leq p_0$ ;
3. estimate  $q'(y_0)$  (in fact,  $q'(y_0) = f(y_0)/y_0$ , with  $f(y)$  the local scale parameter of the GW tail at  $y$ , which can be estimated directly);
4. numerically integrate the differential equation (1) starting from the initial values  $(q(y_0), q'(y_0))$  to obtain  $q(y)$  for every desired value of  $y \geq y_0$ .

The large size of the SEAS5/DCSM5 dataset for each tide gauge station makes it possible to make accurate shape estimates  $\hat{\rho}_p$  over a much wider range of sample fractions  $p$  than would be possible using measurements or reanalysis data.

One candidate model for the profile of  $\tilde{\rho}(y)$  at  $y \geq y_0$  is

$$\tilde{\rho}(y) = \tilde{\rho}(y_0) + (c/\delta)((y/y_0)^\delta - 1) \quad (3)$$

with  $\delta < 0$  and  $c$  a real number. In this model,  $\tilde{\rho}(y)$  relaxes toward the constant  $\rho(y_0) - c/\delta$  as  $y \rightarrow \infty$ , which can be close to  $\tilde{\rho}(y_0)$  or far away from  $\tilde{\rho}(y_0)$ . This model implies that  $F$  satisfies the GW tail limit (eq. (9) in de Valk and van den Brink (2020a)) with shape parameter  $\rho = \tilde{\rho}(y_0) - c/\delta$  and with second-order regularity given by eq. (24) in Appendix A.1 of de Valk and van den Brink (2020a). The relaxation to a constant gives the model of the tail a certain stability, which helps to reduce the sensitivity of estimates of the model to noise. However, if the true  $\tilde{\rho}(y)$  tends to  $\pm\infty$  when  $y$  tends to infinity, then estimates  $\hat{\rho}_p$  will also diverge as  $p$  tends to 0 (i.e., estimates of  $\delta$  will be close to 0 and/or estimates of  $|c|$  will be large). If this is observed, then we may consider other classes of profiles for  $\tilde{\rho}(y)$  which diverge as  $y$  tends to infinity, such as, for example,  $\tilde{\rho}(y) = \tilde{\rho}(y_0) + c \log(y/y_0)$ . But if the model (3) matches the estimates  $\hat{\rho}_p$  well, then there is no need for this.

The estimator of  $\hat{\rho}_p$  is from de Valk & Cai (2018) (the latter addresses estimation of the log-GW shape parameter and tail; it is straightforwardly adapted the GW shape parameter and tail). The estimation of  $\delta$  and  $c$  from the estimates  $\hat{\rho}_p$  or different sample fractions  $p$  is a delicate issue, because these estimates are strongly dependent, and the variance of  $\hat{\rho}_p$  increases strongly with decreasing  $p$ . A new approach was developed in this study, based on the large-sample approximation of the error statistics of  $\hat{\rho}_p$ ; see Appendix B.

This estimator appears to be very effective. As an example, Figure 9 shows the shape estimates  $\hat{\rho}_p$  with their confidence intervals as in Figure 8. The dots in Figure 9 indicate the estimated curve given by (3) with  $\delta$  and  $c$  estimated from the estimates  $\hat{\rho}_p$  for  $p \leq p_0 = 0.01$ . In view of the confidence bands for  $\hat{\rho}_p$ , the estimated curve seems quite plausible. Plausible estimates are also obtained for water level and surge at the other sites of coastal tide gauges, all with the same threshold probability  $p_0 = 0.01$ . All estimates are shown in the figures in Appendix B.

It should be noted that such refined estimation of the function  $\hat{\rho}_p$  would not be possible without a very large dataset.

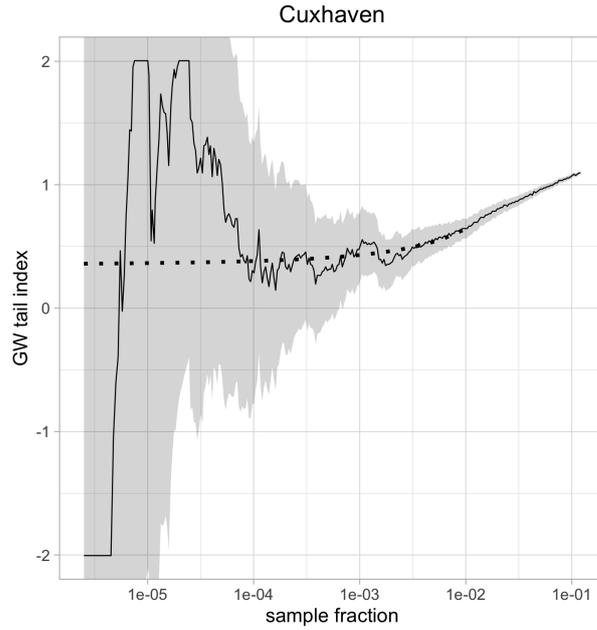


Fig. 9: GW-shape parameter estimates  $\hat{\rho}_p$  (full) from SEAS5/DCSM-5 data of skew surge at Cuxhaven as function of sample fraction  $p$ , with their two-sided 95% confidence intervals. Dots indicate the estimated model (3), fitted to the values of  $\hat{\rho}_p$  for  $p \leq 0.01$ . The indicated sample fractions correspond to 0.007, 0.07, 0.7, 7.0 and 70/year.

After computing the quantile function  $q$  from the estimated function  $\tilde{\rho}$  and other estimates by the four-step procedure outlined earlier, it is straightforward to simulate synthetic datasets of independent random high-tide sea level or skew surge values, each covering a period of 5800 years.

Subsequently, return levels are estimated from the full dataset, from 78-year subsamples, and from a combination of these (using the full dataset to estimate the GP, GW or log-GW shape parameter, and a subsample to estimate the scale and location parameters).

From these, we can compute the error statistics (bias, variance, and from these, the root mean square (RMS) error) directly using the exact return level obtained from the known quantile function  $q$ . We simulate 15 realisations of the full dataset and draw 50 subsamples from each. This is sufficient to obtain precise estimates of bias and RMS error from the full samples as well as from the (much smaller) subsamples.

One issue remains: the synthetic data are independent, but the original high-tide water level and skew surge values from SEAS5/DCSM-5 data are only independent far in the tail of the sample. For return value estimates based on large sample fractions, serial dependence increases the variance. To account for this, the estimated variance of a return level estimate from the synthetic data is multiplied by a threshold-dependent correction factor estimated from the SEAS5/DCSM-5 data; see Appendix E.

### 5.3 Results

Figures 10-13 show the bias and RMS error of the return level estimates from subsamples for different tails (colours) as functions of sample fraction, both for estimates from high-tide water level data (top) and for estimates from data of skew surge (bottom), for a representative subset of eight tide gauge stations.

Dashed lines are the bias and RMS for the GP, GW and log-GW tails with shape parameter fixed to the estimate from the full sample. Comparing these to the bias and RMS of estimates from a subsample gives a conservative assessment of the potential increase in precision due to the use of SEAS5/DCSM-5 data for these tails: it is foreseen that the location and scale parameters will be estimated after calibration of the SEAS5/DCSM-5 data to high-resolution simulations of a limited subset of storms from the SEAS5/DCSM-5 archive, so these will be less precise than estimates from the original set of SEAS5/DCSM-5 data (how much less, is difficult to say at this moment).

For a different view of the results, the same statistics are plotted in Figures 14-16 as functions of the station from South (Duinkerke) to North (Esbjerg), for three values of the sample fraction (0.005, 0.01, 0.05, corresponding to 0.7, 7.0 and 70/year). To read the names of stations along the horizontal axis, it may be necessary to zoom in.

The following can be seen in these plots:

1. The 1-parameter tails (Weibull, exponential) do not consistently perform well, due to bias. For water level, the Weibull tail performs well for water level data at sites to the North of IJmuiden, but poorly for sites further to the South. The opposite is seen with the exponential tail. For skew surge data, both tails show considerable bias, in particular the Weibull tail. This indicates that 1-parameter tails are not reliable.
2. At low thresholds exceeded by sample fractions above 0.01 (frequencies above 7/year), bias and RMS error are generally high, in particular for water level (see Figure 16).
3. The estimates of GP and GW tails with shape parameter estimated from the full dataset (emulating the use of the full SEAS5/DCSM-5 dataset calibrated using a much smaller subset, in this case of 87 years; see above) achieve in most cases the lowest RMS error of all tails. For sample fractions of 0.01 (frequencies above 7/year) or lower, this is consistently so in all cases; RMS errors well below 0.5 m are found for the best performing GW tail at sample fractions of 0.005 and 0.01 (3.5/year and 7/year; see Figures 14-15).

Although these quantitative outcomes cannot be directly translated to the situation of using calibrated SEAS5/DCSM-5, we can conclude that for these two tails, random errors in the shape and location parameters due to calibration of the data on a relatively small subset are likely to be small; most of the error is associated to the shape parameters.

The RMS error for the log-GW tail with shape parameter estimated from the full dataset is overall slightly higher and the results are less consistent.

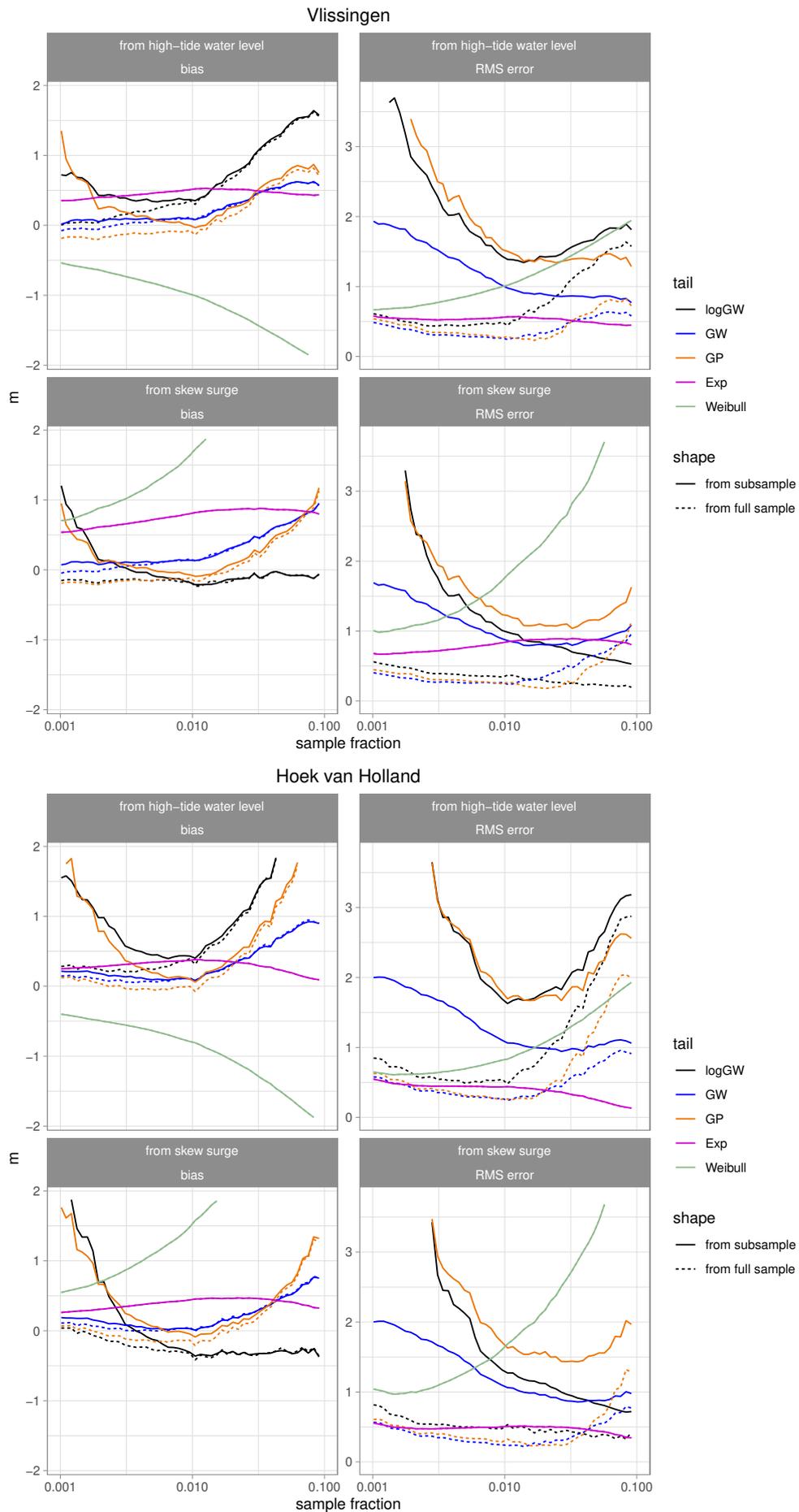


Fig. 10: Statistics of return level estimates of high-tide water level at Vlissingen and Hoek van Holland for a return period  $R = 10^7$  years. The sample fractions correspond to 0.7, 7.0 and 70/year.

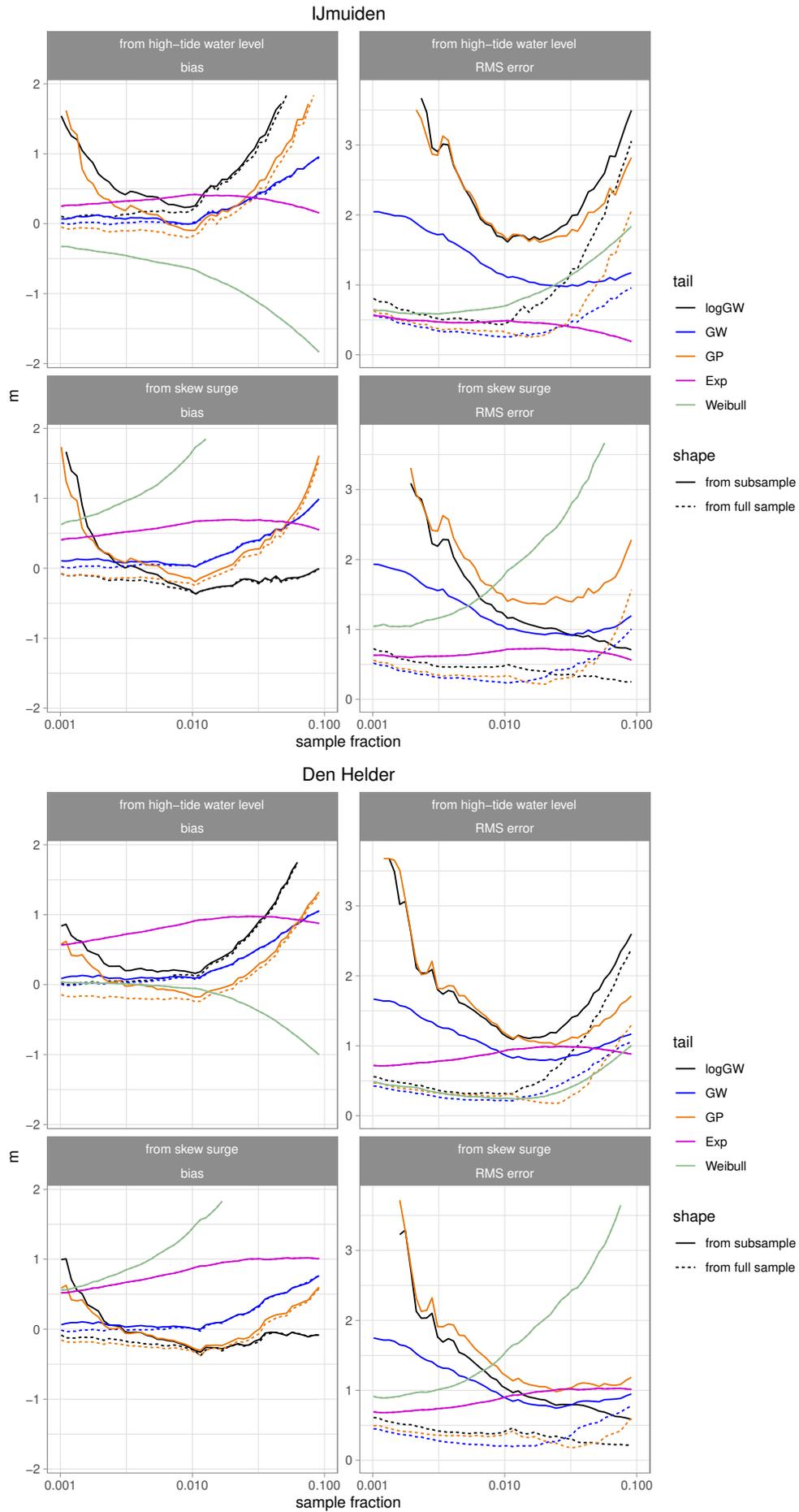


Fig. 11: As Figure 10 for IJmuiden and Den Helder.

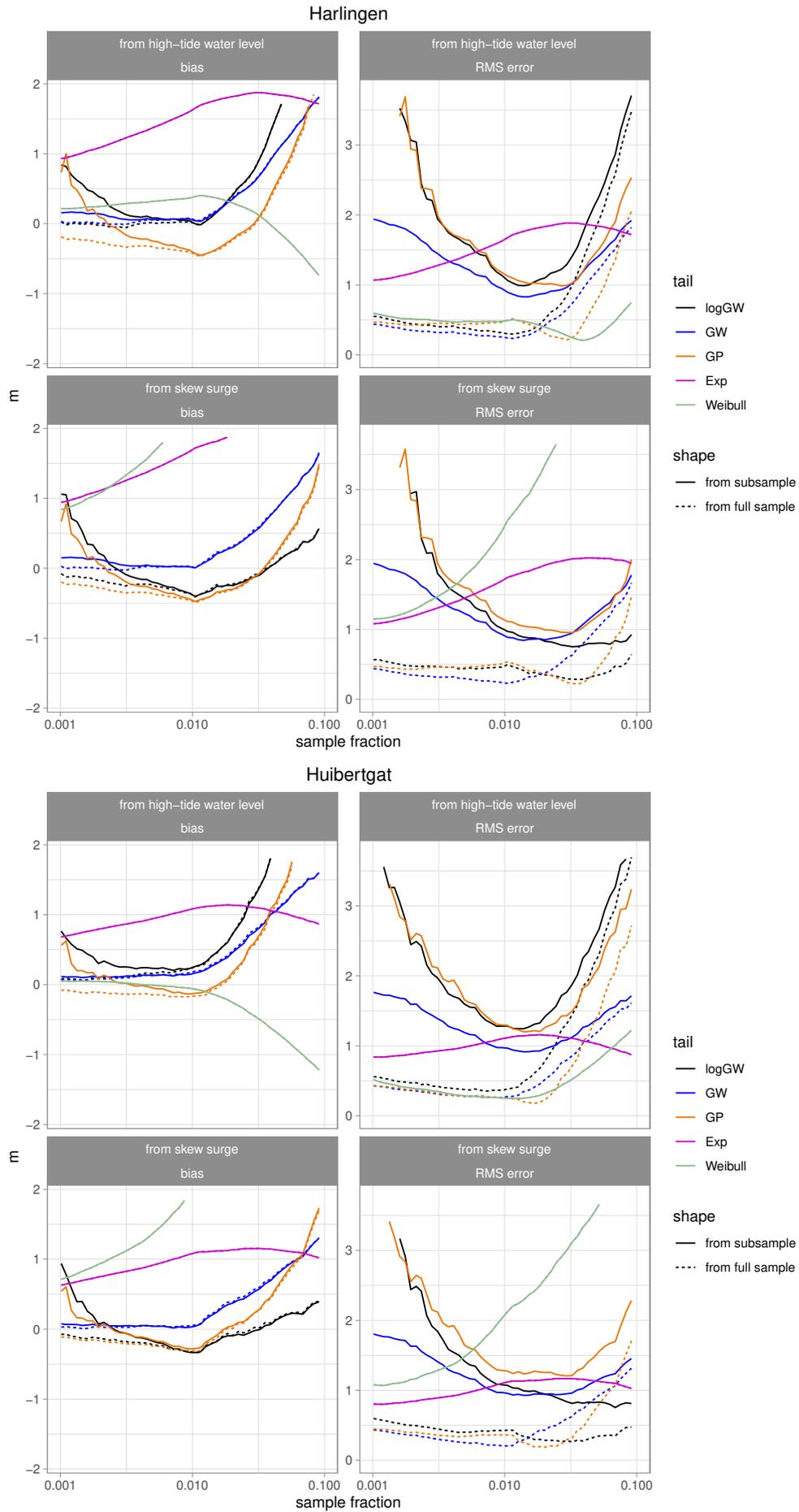


Fig. 12: As Figure 10 for Harlingen and Huibertgat.

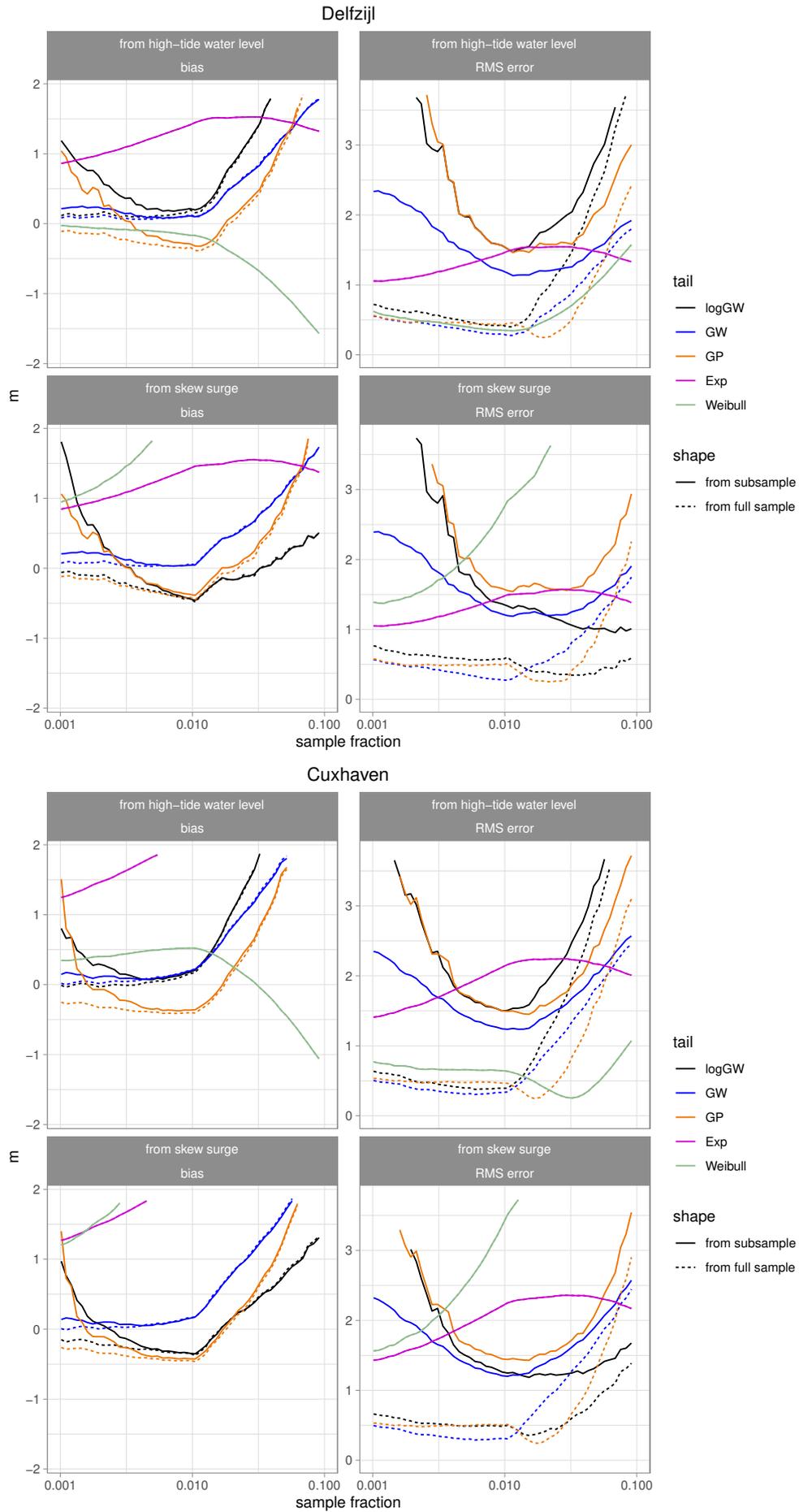


Fig. 13: As Figure 10 for Delfzijl and Cuxhaven.

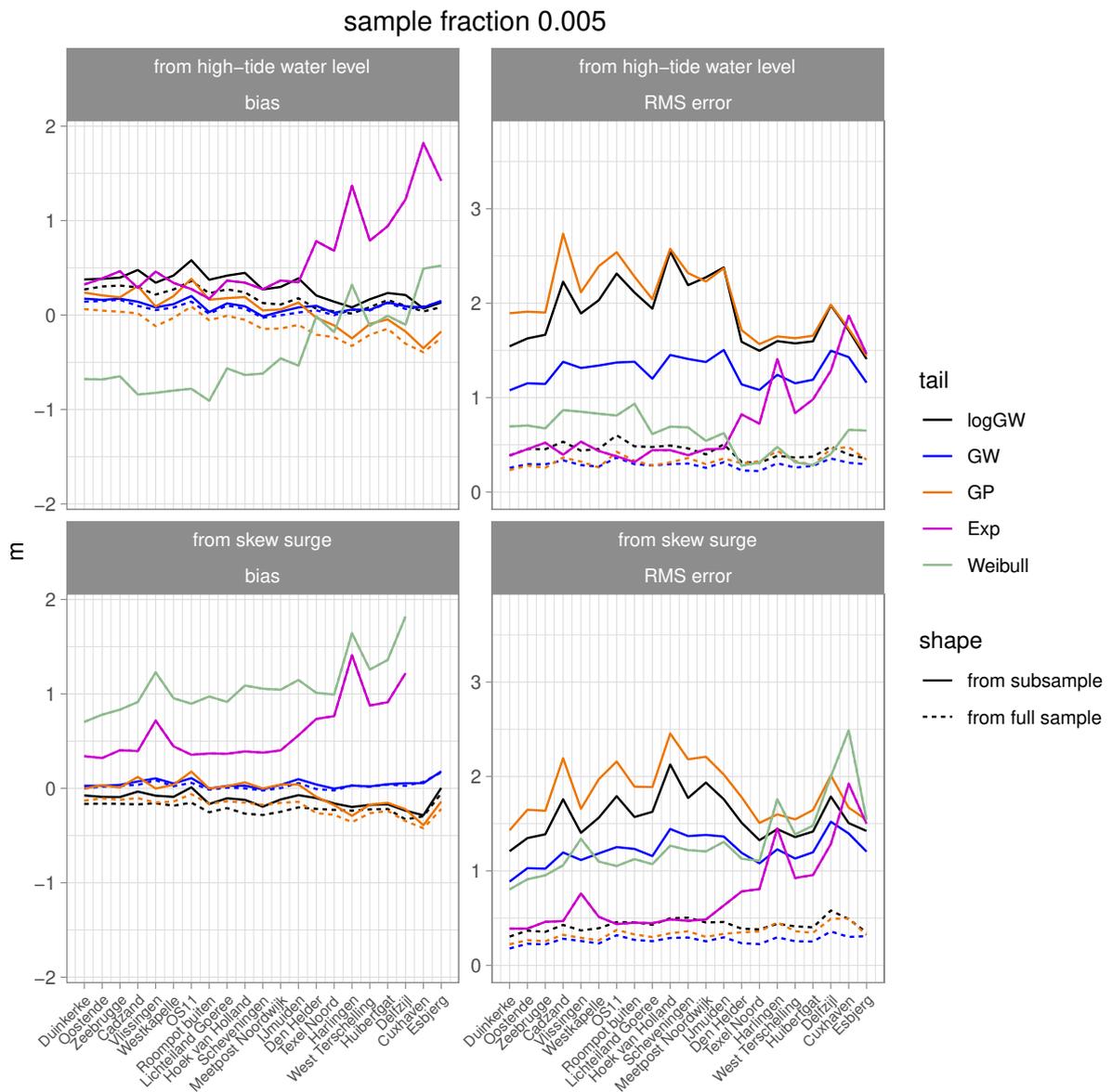


Fig. 14: Statistics of return level estimates of high-tide water level for a return period  $R = 10^7$  years and a fixed sample fraction of 0.005 (3.5/year).

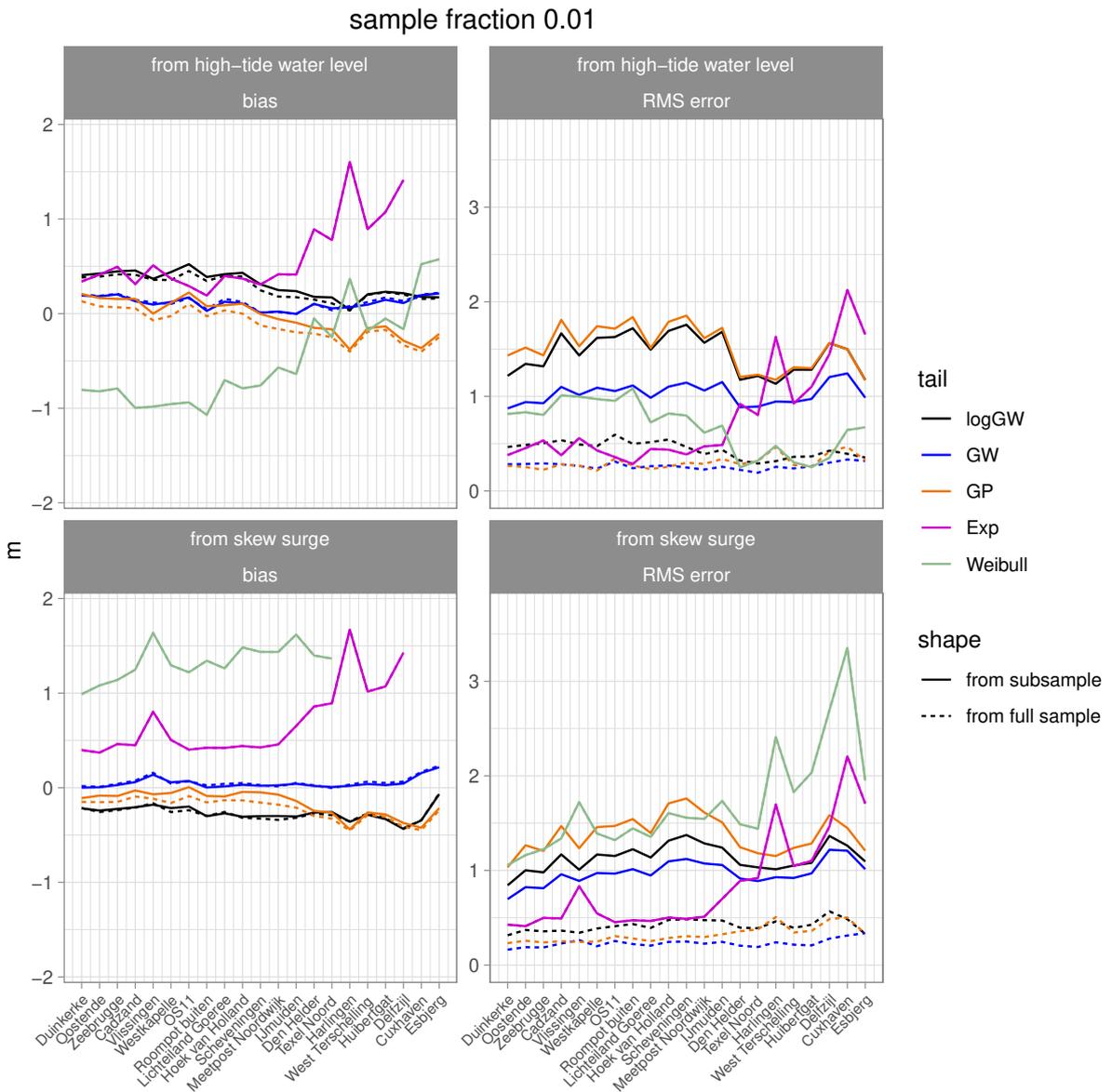


Fig. 15: As Figure 14 for a sample fraction of 0.01 (7/year).

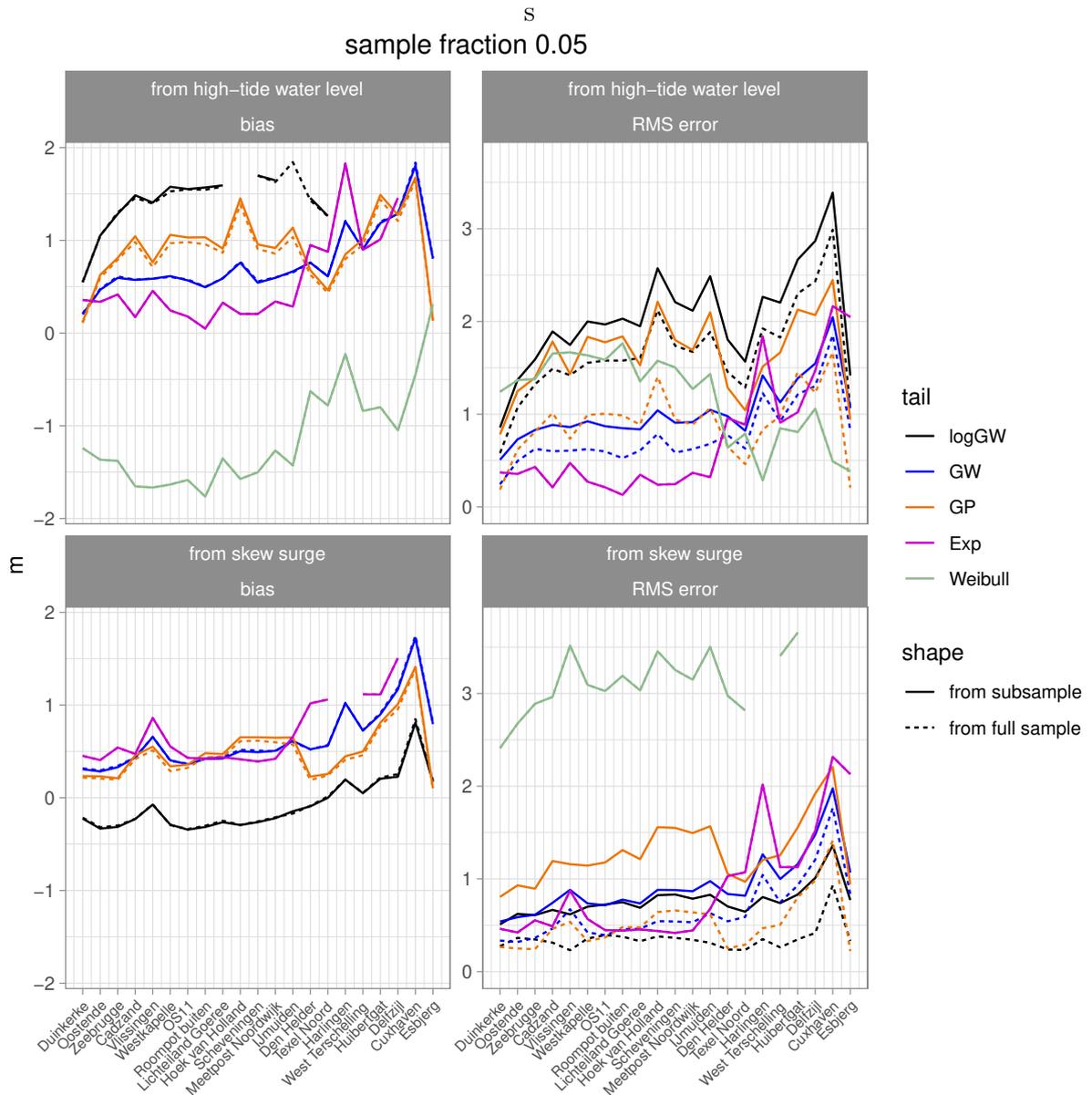


Fig. 16: As Figure 14 for a sample fraction of 0.05 (35/year).

4. If the GP, GW or log-GW shape parameter is estimated from the subsamples (emulating estimates from measurements or reanalysis data), then the RMS error tends to be much higher: of the order of 1 metre or more.

This shows the potential value of using SEAS5/DCSM-5 data for estimation of the shape parameter, and using much smaller datasets for calibration of the location and scale parameter. In the application, model-related bias in the shape parameter from the SEAS5/DCSM-5 data (not taken into account in the present analysis) will also contribute to the RMS error. This is addressed in Section 6.

5. The GW tail appears to perform slightly better than the GP tail, because the GP tail has some negative bias for tide gauge stations North of IJmuiden. If the shape parameter is estimated from the subsample, then the difference in performance is larger, because return value estimates based on the GW tail have lower variance.
6. Using skew surge data instead of high-tide water level data results in approximately the same accuracy, according to the Monte Carlo analysis.

The same analysis was performed with return level estimates for a return period of  $10^4$  years, for which the reference estimates are more tightly constrained by the data; see Figures 17-19. The patterns do not differ from those for a return period of  $10^7$  years.

To see how the estimates of bias and RMS error from Monte-Carlo simulation compare to estimates derived directly from the data using the method from de Valk and van den Brink (2020a), the estimates for Vlissingen and Hoek van Holland from both methods are plotted in Figures 20 and 21.

For sample fractions larger than about 0.01 (frequencies above 7/year), the estimates of bias and RMS error from the two methods are remarkably similar. The main difference is that the estimates based on de Valk and van den Brink (2020a) are much more noisy, in particular for sample fractions below 0.01. Note in particular the large differences between reference estimates ("bias-corrected return value") at different sample fractions and for different tails. In fact, where there are large differences between both types of estimates (e.g. the bias and RMS error of the  $10^7$ -year water level and surge at Hoek van Holland estimated using the GP tail, at sample fractions above 0.01), there are also large differences between the reference values ("bias-corrected return values") for different tails, so the results of the method from de Valk and van den Brink (2020a) are ambiguous. In these cases, the error statistics based on Monte-Carlo simulation appear to be more reliable.

These results show that for high-tide water level and skew surge, the advantages of the modified approach appear to outweigh its potential drawbacks.

A similar comparison is shown in Appendix C for wind speed from the SEAS4 and SEAS5 seasonal ensemble reforecasts and for pseudo-wind speed from SEAS5 at the location 55N, 3E in the central North Sea. The results are similar to those for high-tide water level and skew surge above, and confirm the conclusions.

The finding that two completely different methods for assessing error statistics give similar results gives confidence in the conclusions derived from them.

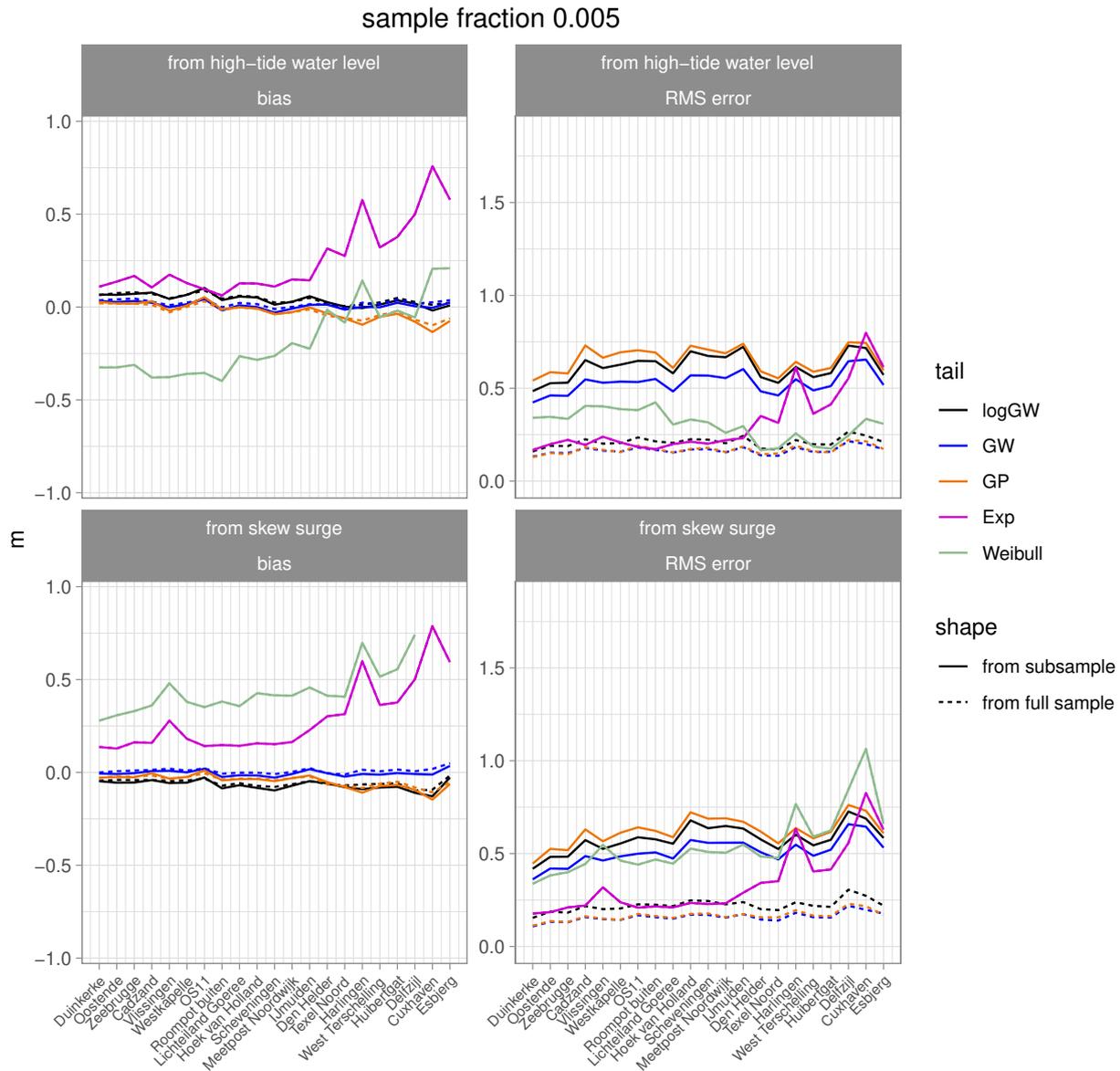


Fig. 17: Statistics of return level estimates of high-tide water level for a return period  $R = 10^4$  years and a fixed sample fraction of 0.005 (3.5/year).



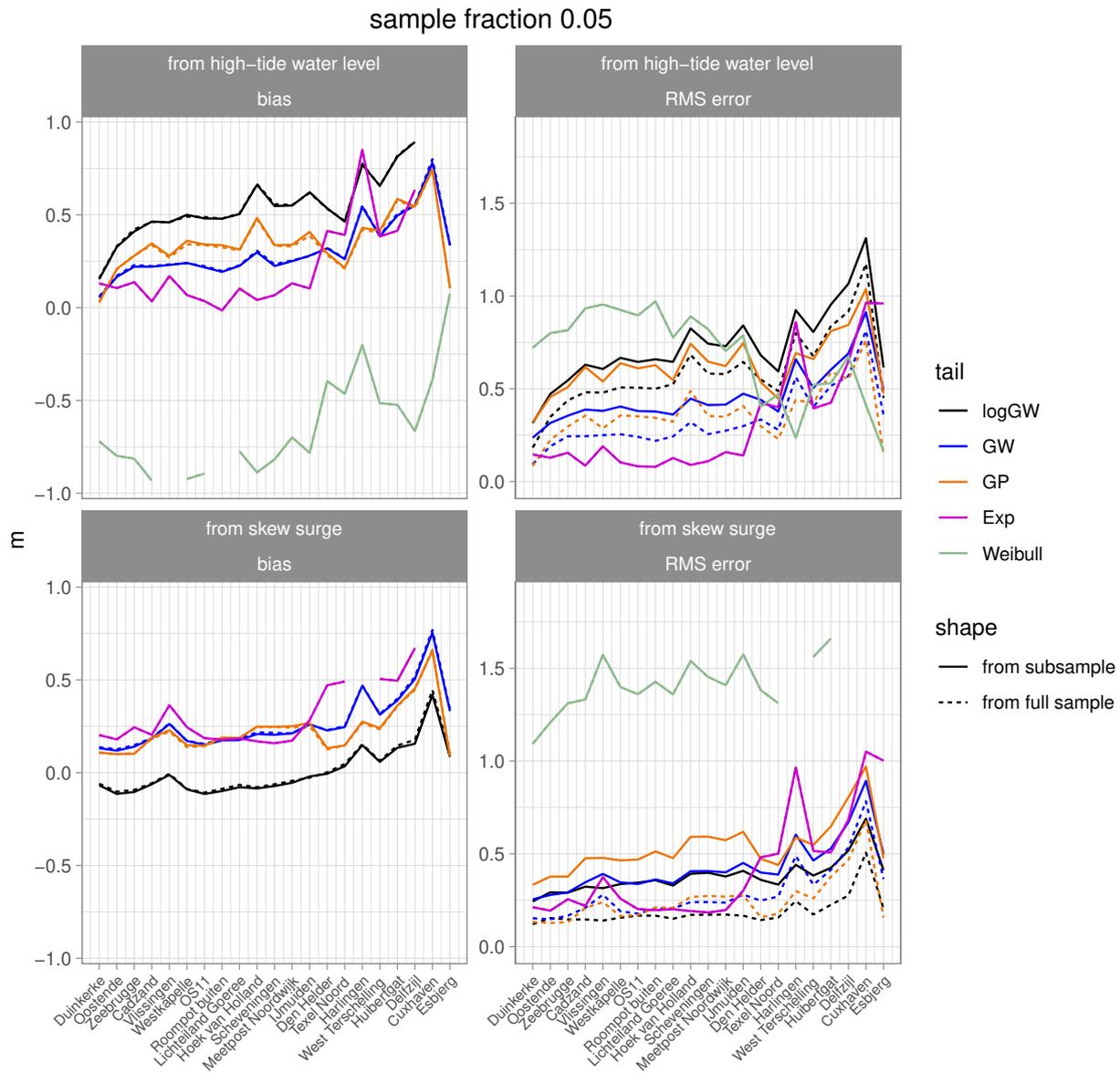


Fig. 19: As Figure 17 for a sample fraction of 0.05 (35/year).

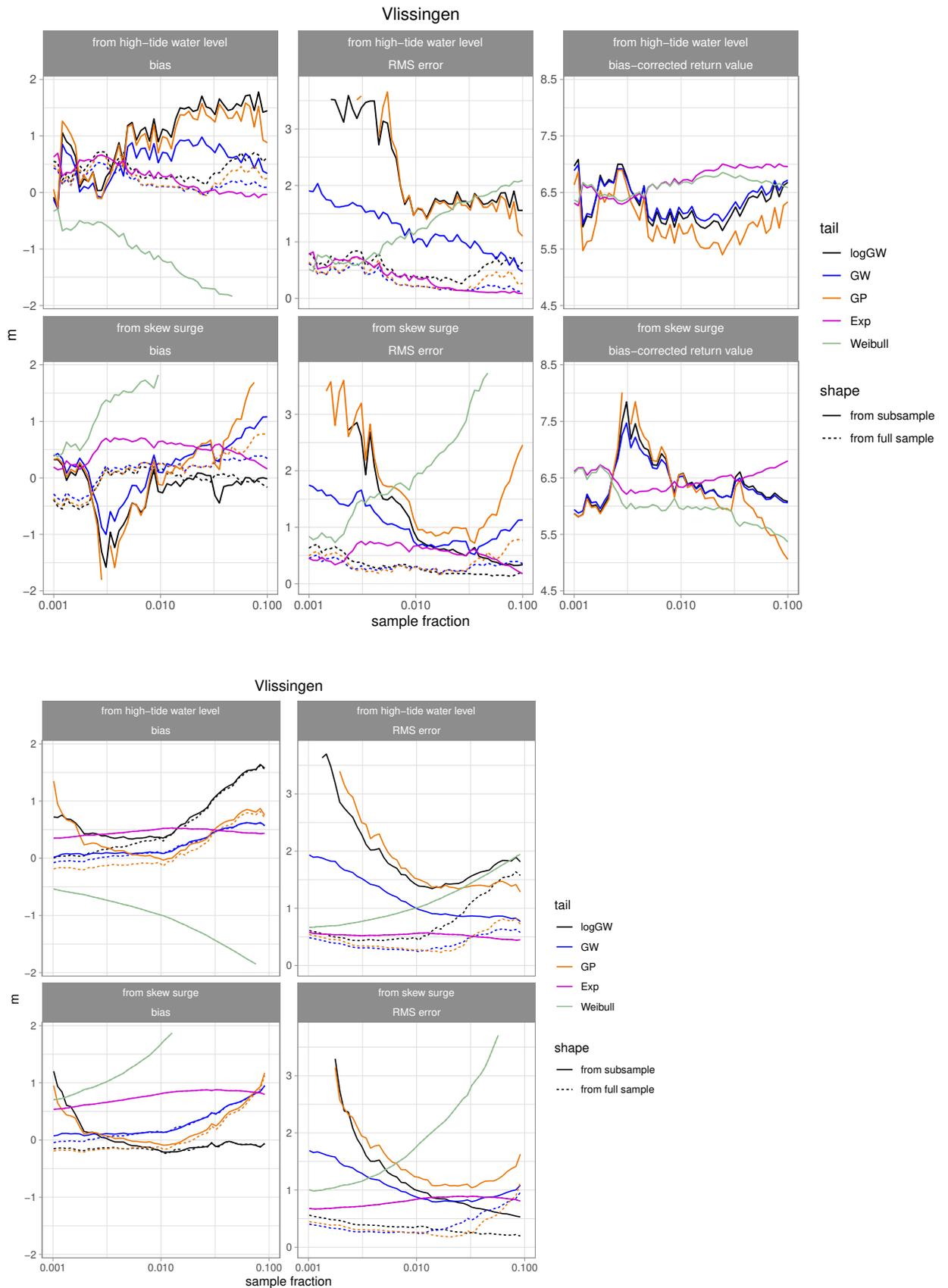


Fig. 20: Bias and RMS error (left and centre) of return level estimates of high-tide water level and skew surge at Vlissingen for a return period  $R = 10^7$  years computed as in de Valk and van den Brink (2020b) with an estimated reference return level ("bias-corrected return value") (top 2 rows), and using the Monte-Carlo based method of this report (bottom 2 rows). The indicated sample fractions correspond to 0.7, 7.0 and 70/year.

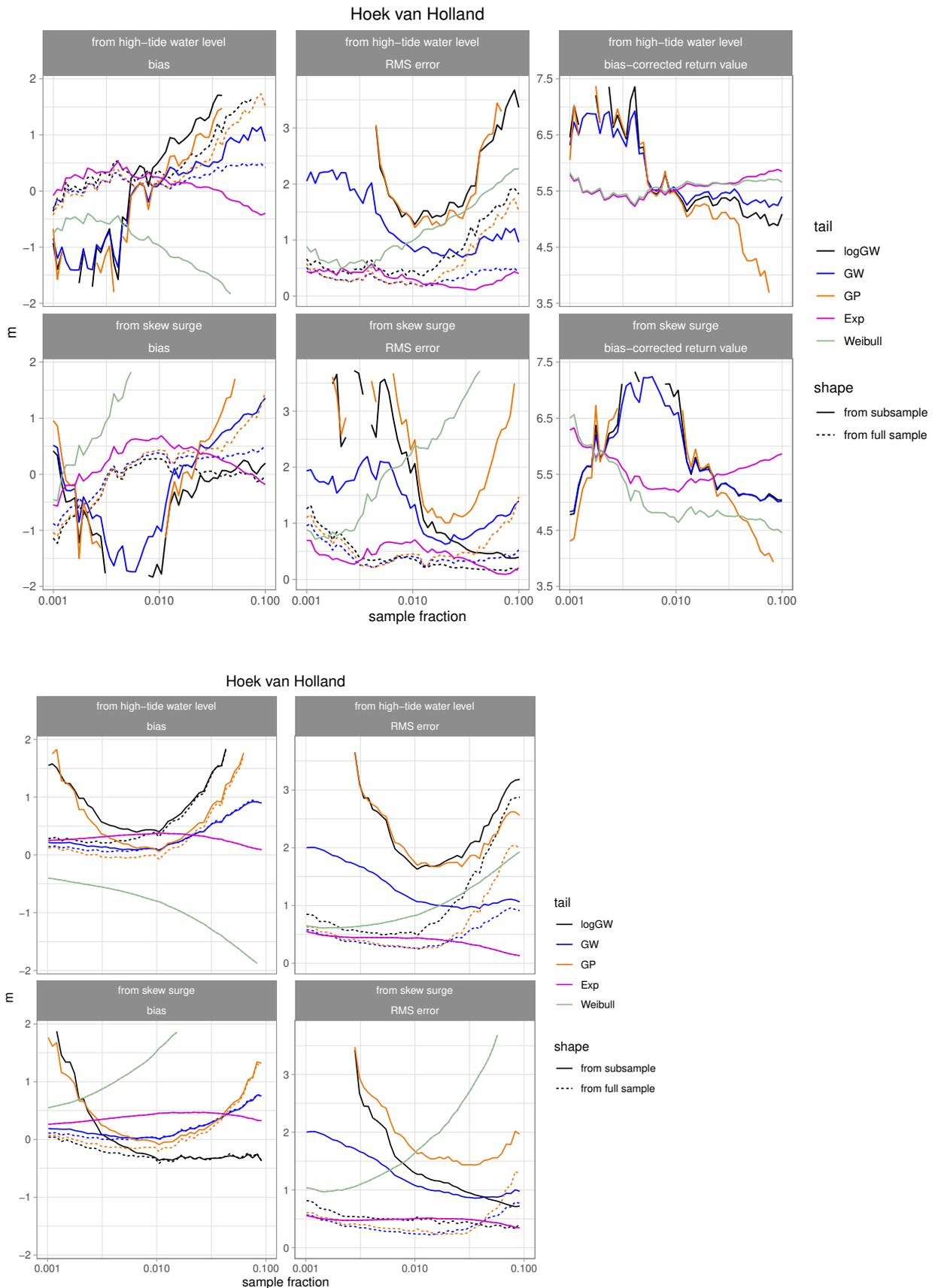


Fig. 21: Bias and RMS error (left and centre) of return level estimates of simulated high-tide water level and skew surge at Hoek van Holland for a return period  $R = 10^7$  years, computed as in de Valk and van den Brink (2020b) with an estimated reference return level ("bias-corrected return value") (top 2 rows), and using the Monte-Carlo based method of this report (bottom 2 rows). The indicated sample fractions correspond to 0.7, 7.0 and 70/year.

## 6 Verification of the tail shapes of high-tide water level and skew surge data from SEAS5/DCSM-5

The high-tide water level and skew surge data from SEAS5/DCSM-5 can be used for estimating return values for high return periods if there is enough confidence that the bias in the tail of their distribution function is acceptable. Errors in offset and scale are not considered a problem, as these can be corrected by calibration. The current plan is to correct bias in offset and scale using the output of models of higher resolution than of the SEAS5/DCSM-5 suite, because this makes it possible to derive the corrections also for sites where no measurements are available.

Therefore, a comparison between the empirical tails from SEAS5/DCSM-5 data and measurements should be focused primarily on the shape parameter of the tail to be used for estimation of return values. A comparison of shape parameter estimates of the GW and GP tails from SEAS5/DCSM-5 data and from measurements was carried out for the tide gauge stations at Vlissingen, Hoek van Holland, IJmuiden, den Helder, Harlingen and Delfzijl.

The high-tide water level measurements relative to NAP (as well as various types of low-tide water levels) were provided by the Helpdesk Water of Rijkswaterstaat. For each station, a smooth trendline for high-tide water level was determined by fitting a local linear regression (loess) curve to the annual means as in de Valk (2020). This trendline was normalised by subtracting its value for the last complete year (2019), and was subsequently subtracted from the high-tide water level values.

Skew surge data from measurements at the same six stations were used without adjustments<sup>4</sup>. Only the values above 0.30 m are available.

The tails of the empirical distribution functions of high-tide water level and skew surge are displayed in Figure 22. For the water level exceeded 10% of the time, we see positive differences between SEAS5/DCSM-5 and measurements at all stations except IJmuiden and Hoek van Holland. This indicates that the tidal maxima at Vlissingen, Den Helder, Harlingen and Delfzijl are overestimated. This is confirmed by the plots for skew surge, which do not show an offset error. The overestimation of tidal maxima may in part be caused by the limited resolution of DCSM-5, since all four stations where it occurs are adjacent to estuaries.

This finding calls for the use of a shallow-water flow model in which the tides are represented more accurately, such as DCSM6 or DCSM7. Further work is currently undertaken to downscale 250 storms from the SEAS5 archive and compute the resulting water levels and surges in order to map meteorological and hydraulic simulation model biases.

The slopes of the lines in Figure 22 for SEAS5/DCSM-5 and measurements agree only for the station Vlissingen; for other stations, the slopes from SEAS5/DCSM-5 data are steeper than the slopes from measurements. The difference in slope appears to increase along the coast from SW to NE. The same was found earlier in Appendix B of Chbab (2017). The cause is not yet known.

Estimates of the GW shape parameter from measurements and SEAS5/DCSM-5 data of high-tide water level and skew surge are shown in Figures 23 and 24. As

---

<sup>4</sup> For Delfzijl, the values for dates before Aug 3, 1881 were discarded, because they are apparently erroneous.

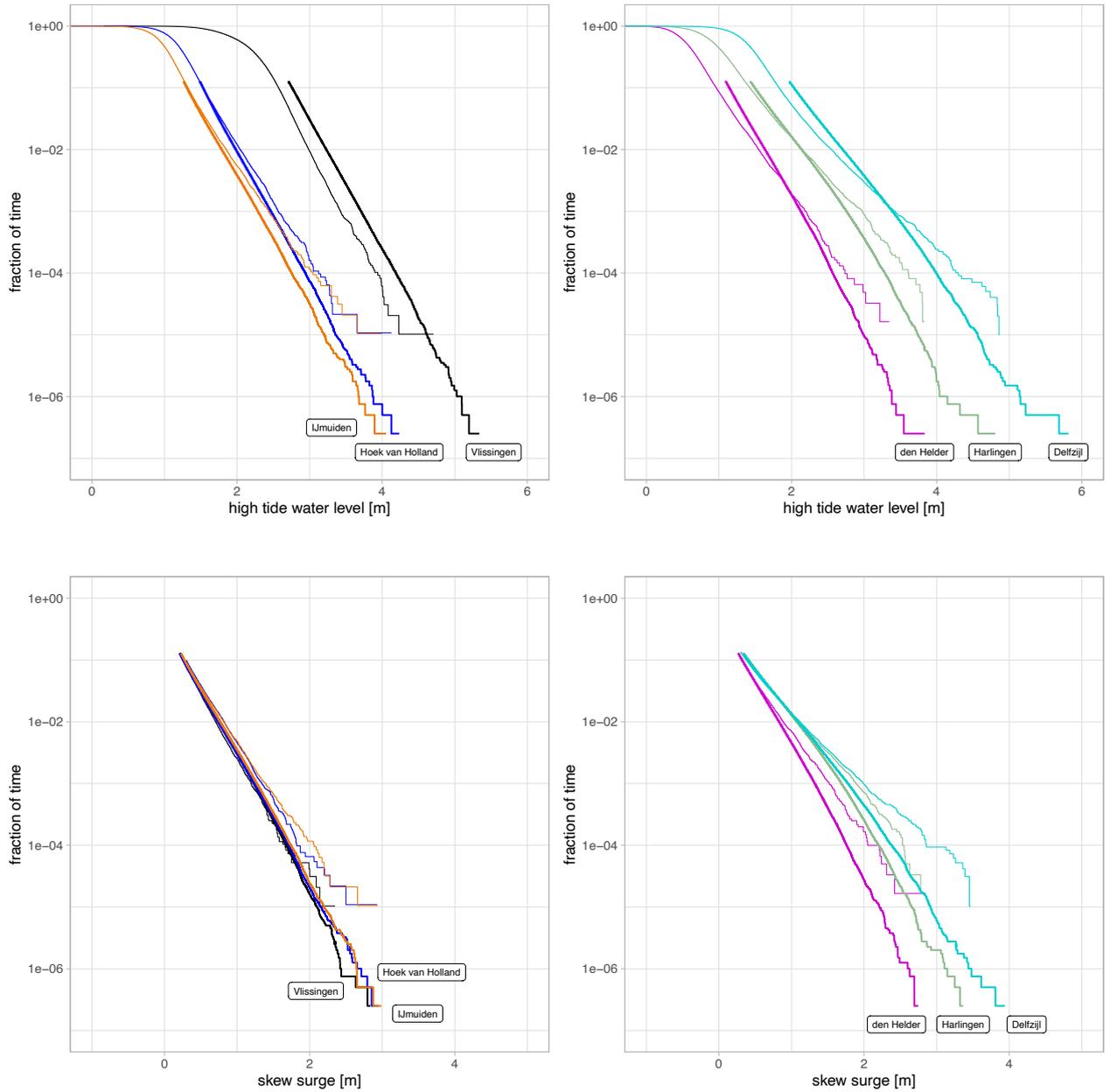


Fig. 22: Empirical probabilities of exceedance of high-tide water level (top) and skew surge (bottom) from measurements (thin lines) and SEAS5/DCSM-5 simulations (thick lines) for six tide gauge stations: IJmuiden (orange), Hoek van Holland (blue), Vlissingen (black), den Helder (magenta), Harlingen (green) and Delfzijl (cyan). The indicated fractions of time correspond to 0.0007, 0.07, 7.0, and 700/year.

in Section 5, the shape parameter is regarded as a function of the probability of exceedance  $p$  of the threshold ('sample fraction' in the plots), perturbed by noise which increases with decreasing  $p$ .

For relatively high  $p$ , the shape parameter from SEAS5/DCSM-5 data is lower than the shape parameter from measurements at all stations except Hoek van Holland. However, for all other stations except Delfzijl, the simulated shape values are above or near the lower boundary of the 95% confidence intervals of the shape estimates from measurement data. For Delfzijl, the differences are larger.

For this station and for Harlingen, the estimates from measurements drop steeply as  $p$  decreases below about 0.01. However, the estimates from SEAS5/DCSM-5 data decrease only slightly. Because the confidence intervals of the estimates from measurements are very wide, they contain almost all estimates from SEAS5/DCSM-5, so the latter are not invalidated in this range. Figure 22 suggests that the drop may have been enhanced by the small differences between the top three surges and water levels at Harlingen and Delfzijl.

It seems unlikely that the drop in the shape estimates from measurements is a real feature of the tail, because one would expect that for exceptionally high storm surges, the large-scale surge on the North Sea would dominate, so one would not expect that the tails at Harlingen and Delfzijl are radically different from those along the West coast.

Therefore, for the lowest sample fractions, the curves from SEAS5/DCSM-5 for Harlingen and Delfzijl in Figure 24 seem more realistic than those from the measurements. This already shows the value of the SEAS5/DCSM-5 dataset, even with all its limitations: it imposes physical consistency between the tail estimates for different sites. In fact, it is virtually impossible to choose a value of the shape parameter to be used for extrapolation from the estimates from measurements (as they vary strongly with sample fraction), but for the estimates from the much larger SEAS5/DCSM-5 dataset, this is much easier.

The apparent underestimation by SEAS5/DCSM-5 of the shape parameter (upward curvature) at sample fractions above about 0.01 (frequencies above 7/year) is directly connected with the substantial overestimation of the magnitudes of the slopes at lower sample fractions in the panels on the right in Figure 22. For relatively low water levels (exceeded by high sample fractions above say 0.1, corresponding to 70/year), the slopes of the curves in this figure are relatively steep<sup>5</sup>. Because the curvature is the derivative of the slope, the steeper slope of SEAS5/DCSM-5 at high water/surge levels therefore implies a less positive curvature at intermediate levels, i.e., a lower GW shape parameter. Both findings (the underestimation of the shape parameter and the overestimation of the magnitude of the slope) may therefore have the same explanation(s). This implies that the apparent underestimation by SEAS5/DCSM-5 of the shape parameter at sample fractions above about 0.01 does not need to invalidate these estimates.

In de Valk and van den Brink (2020b), no mismatch is found between estimates of the omnidirectional and directional GW shape parameter of the SEAS5 wind and

---

<sup>5</sup> For high-tide water level, this may be expected, because the astronomical tide has a substantial effect on water levels in this range and it has a very light tail. The tail of skew surge is similar to the tail of high-tide water level except for a level shift, and it shows this feature as well.

pseudo-wind speeds and estimates from reanalysis data. Therefore, we should first look at limitations of the DCSM-5 hydraulic model (such as resolution effects) for an explanation of observed mismatches in shape parameter. An ongoing detailed analysis of the effects of improved resolution and physics of the atmospheric and hydrodynamic models on the water levels reached during storms covering a wide range of severities (including the storms causing the highest surges in the SEAS5/DCSM-5 dataset) may help to understand these mismatches better.

Comparing the shape estimates for IJmuiden and Hoek van Holland, we see that for measurements, the estimates for IJmuiden are mostly higher than those for Hoek van Holland, but for SEAS5/DCSM-5 data, it is just the opposite. The cause is not understood.

Appendix D presents a comparison of estimates of the GP shape parameter from SEAS5/DCSM-5 data and measurements. It does not provide additional insight; mismatches are qualitatively similar.

In summary, the results indicate that for most stations, the shape parameter may be somewhat underestimated by SEAS5/DCSM-5 simulations, but with the exception of Delfzijl, we cannot conclude that the shape parameter estimates from the SEAS5/DCSM-5 simulations are inconsistent with the estimates from measurements. For the interior of the Waddenzee (and in particular for Delfzijl), there is a need for better physical and numerical understanding of the considerable mismatch between tail estimates from measurements and from the SEAS5/DCSM-5 data.

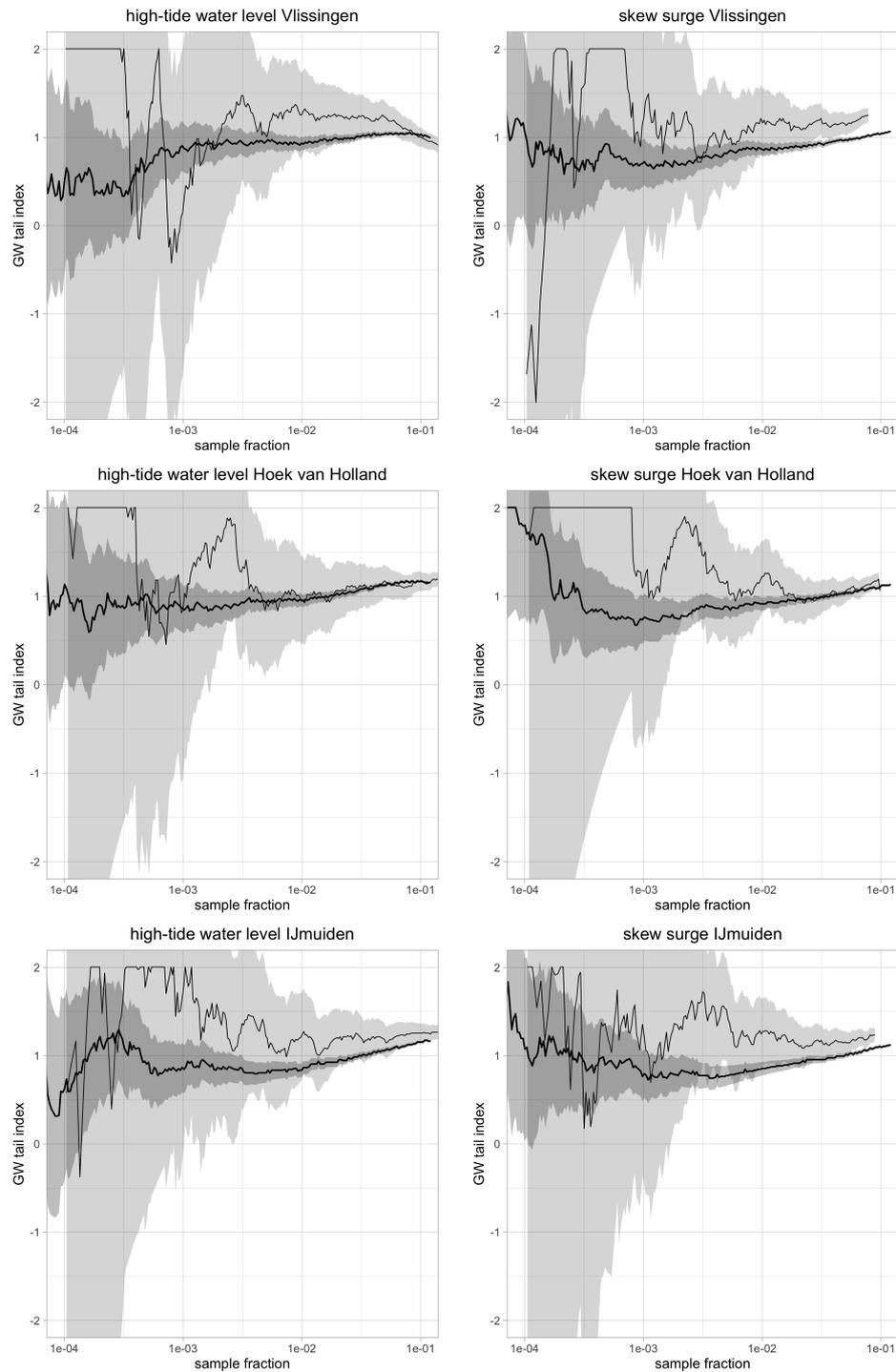


Fig. 23: GW shape estimate vs. sample fraction with 95% confidence interval from measurements (thin line, light shading) and from SEAS5/DCSM-5 simulations (thick line, dark shading) of high-tide water level (left) and skew surge (right) at three tide gauge stations. The indicated sample fractions correspond to 0.07, 0.7, 7.0 and 70/year.

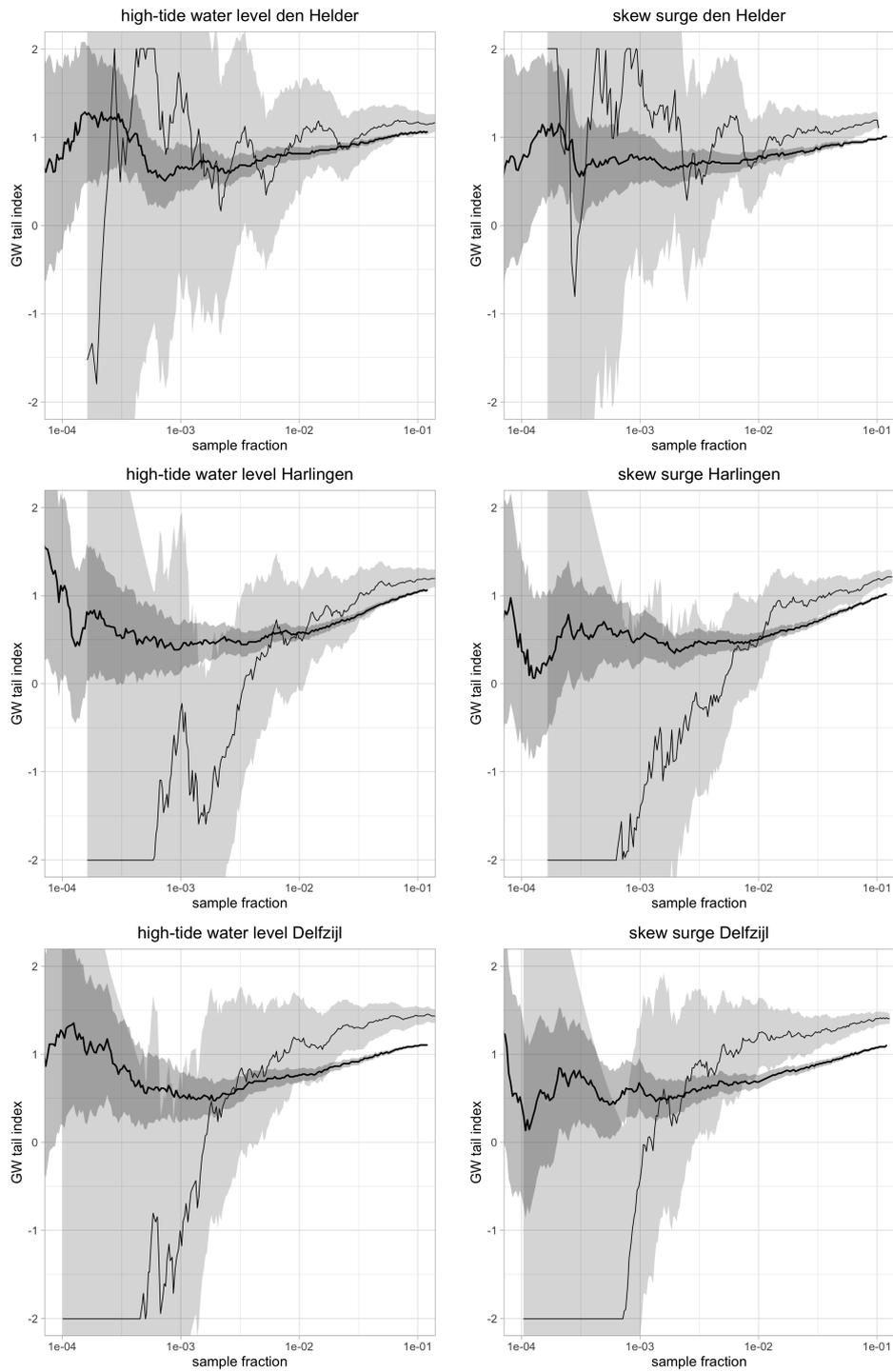


Fig. 24: As Figure 23 for three other tide gauge stations.

## 7 Implications for simulation model bias

To put the preceding comparison of GW shape parameters in context, Tables 3 and 4 show the effect of a given increase in the shape parameter estimated from the SEAS5/DCSM-5 data on the return values of high-tide water level and skew surge, showing the highest increase over all tide gauge sites. Increases of the shape parameter of 0.1 and 0.2 are considered, which are typical for the deviation of the shape parameter of the measurements from the shape parameter of the SEAS5/DCSM-5 data in Figures 23 and 24 (of course, the estimates from measurement data have wide confidence intervals, so these values are indicative at best).

$\Delta\rho$	$10^7$	$10^6$	$10^5$	$10^4$	$10^3$	$10^2$	$10^1$	years
0.1	0.38	0.30	0.24	0.19	0.14	0.09	0.05	m
0.2	0.80	0.65	0.50	0.39	0.28	0.18	0.09	m

Tab. 3: Increase in return level of high-tide water level (m) in response to an increase  $\Delta\rho$  in the GW shape parameter for several return periods (header row); increases are maximized over all tide gauge sites.

$\Delta\rho$	$10^7$	$10^6$	$10^5$	$10^4$	$10^3$	$10^2$	$10^1$	years
0.1	0.34	0.29	0.24	0.19	0.14	0.09	0.05	m
0.2	0.72	0.61	0.50	0.39	0.28	0.18	0.09	m

Tab. 4: Increase in return level of skew surge (m) in response to an increase  $\Delta\rho$  in the GW shape parameter for several return periods (header row); increases are maximized over all tide gauge sites.

To compare the simulation model bias to the estimation error analyzed in Chapter 5, we may compare the numbers in Table 4 to the RMS errors in Figures 14-15: both are "total" errors (the former consists of bias only).

Comparing the values in Table 4 for a return period of  $10^7$  year to the RMS errors in Figures 14-15 for GW and GP tails with shape estimated from the full dataset (dashed lines), which are of the order of 0.3 m, we see that a model uncertainty in the shape parameter of 0.1 gives a similar uncertainty in the return value as the estimated RMS error. With a model uncertainty in the shape parameter of 0.2, this becomes the dominant source of error in return value estimates. Comparing the tables to Figures 17-18, the same is found for a return period of  $10^4$  year.

On the positive side, the effects of model uncertainties in the GW shape parameter of up to 0.2 are still smaller than the RMS errors of return value estimates from 78-year long datasets using the GW tail (full lines), which are of the order of 1 m for a return period of  $10^7$  year and 0.5 m for a return period of  $10^4$  year (see e.g. Figures 14, 15, 17 and 18), and even higher for estimates based on the GP tail. This supports the use of SEAS5/DCSM-5 data (at least for the shape parameter of the tail).

In the uncertainly analysis, model uncertainty in the shape cannot be ignored and may well be the dominant source of error in the end. Figures 23 and 24 suggest that this is the case in particular along the Waddenzee coast, for sample fractions above about 0.01 for Harlingen or 0.005 for Delfzijl (7/year, resp. 3.5/year). However, for

the final estimates based on calibrated SEAS5/DCSM-5 data, we only need the shape parameter at a relatively low sample fraction smaller or equal to 0.01 (see Chapter 5). It is possible that in this range, the shape parameter from SEAS5/DCSM-5 is reliable, but we cannot check this because the corresponding shape estimates from measurements have very wide confidence bands.

After performing additional runs with higher-resolution models (HARMONIE, and more recent versions of DCSM), we will have a better understanding of the shortcomings of the SEAS5/DCSM-5 model suite, and possibly an improved dataset matching the tails from the measurement better. At that stage, we may be able to interpret the observed mismatches much better than we are able to right now.

## 8 Conclusions

**Tail estimation error** An analysis of the bias and RMS error of return value estimates of simulated skew surge and high-tide water level at 21 tide gauge stations (which ignores simulation model error) shows that:

- a) If the shape parameter is estimated from a very large dataset (in this case, of effectively 5800 years), then both the GW and GP tail produce accurate return value estimates for skew surge and high-tide water level: root mean square errors are of the order of 0.3 m for a return period of  $10^7$  year.
- b) This supports the use of (calibrated) SEAS5/DCSM-5 or similar large datasets.
- c) Overall, the GW tail performs best for skew surge and high-tide water level at all sites: it produces slightly higher return values than the GP tail (in particular, no negative bias at sites to the north of IJmuiden), performs significantly better than the GP tail if noise is large (if the shape parameter is estimated from the smaller 78-year datasets), and clearly performs best for wind speed/pseudo-wind speed (Appendix C and de Valk and van den Brink (2020a)). In addition, the refined tail fits for the purpose of Monte-Carlo simulation in Section 5 and Appendix B support a GW tail.
- d) Monte-Carlo simulation based on plausible tails estimated by a refined analysis of very large and realistic datasets of wind, water level and surge data, like the SEAS5/DCSM-5 data in this study, appears to be a good method for comparing statistical models and methods: results are compatible with the method in de Valk and van den Brink (2020a) but are much less affected by noise.
- e) Use of the same (GW) tail for (pseudo-) wind speed and surge/water level has the added benefit that a power law for the wind speed/surge relation (a reasonable simplification on physical grounds) is preserved in GW tails, but not in GP or log-GW tails, for example. A practical demonstration was provided in de Valk and van den Brink (2020a)).
- f) A sample fraction of about 0.01 (threshold exceeded 7/year) for tail estimation from the SEAS5/DCSM-5 water level and surge data appears to be a good choice: higher values result in significant bias, and much lower values in loss of precision. The same was found for wind speed/pseudo-wind speed from SEAS5.
- g) For estimation, skew surge offers no benefit over high-tide water level. Return value estimates of high-tide water level derived from data of skew surge tend to be slightly lower than direct estimates from high-tide water level data, probably as a result of surge-tide interaction. Therefore, there is a slight benefit in using water level data directly, because it tends to result in somewhat higher return values for the water level.

### Simulation model error: comparison to shape parameter estimates from measurements

- h) For most stations, the GW shape parameter may be somewhat underestimated by SEAS5/DCSM-5 simulations, but with the exception of Delfzijl, we cannot conclude that the shape parameter estimates from the SEAS5/DCSM-5 simulations are inconsistent with the estimates from measurements.

However, even for Delfzijl, the shape estimates from SEAS5/DCSM-5 data using small fractions of the data (high thresholds) are not invalidated by the measurements; in fact, they seem more realistic than the (very imprecise) estimates from the measurements.

- i) Earlier (de Valk and van den Brink, 2020b), it was found that for wind speed and pseudo-wind speed, estimates of the GW shape parameter from SEAS5 data are consistent with estimates from reanalysis data. Therefore, the apparent bias in the GW shape parameter of SEAS5/DCSM-5 water level and surge simulations is likely related to the hydraulic model.
- j) Therefore, it is recommended to investigate in detail the impact of hydraulic model refinement (in particular, improvement of the spatial resolution) on the simulated high-tide water level and surge extremes.
- k) For Delfzijl and Harlingen, it is virtually impossible to derive a value of the GW or GP shape parameter from the measurement data, due to ambiguity and high uncertainty. Such problems are not encountered with the SEAS5/DCSM-5 data, due to their much larger sample size.
- l) Realistic values of simulation model bias in the GW shape parameter from SEAS5/DCSM-5 simulations lead to errors in return values which are considerably smaller than estimation errors for 77-year data subsamples and therefore, smaller than estimation errors from measurement records. This indicates that the shape parameter estimates from SEAS5/DCSM-5 simulations are suitable for estimating return values of high-tide water level. The results of simulations of selected storms from SEAS5 using more refined/higher-resolution models may provide additional insight.

The present results also indicate that in the uncertainty analysis, simulation model bias in the shape parameter cannot be neglected, as it may be as large or larger than the estimation error from the full set of SEAS5/DCSM-5 data. This means that we will need to make at least an educated guess of the simulation model bias in the shape parameter, based on all information then available.

- m) There is a need for better physical and numerical understanding of the observed mismatch between tail estimates from measurements and from the SEAS5/DCSM-5 data within the Waddenzee, and in particular for Delfzijl. When the results of downscaling of 250 storms from the SEAS5 archive and high-resolution simulation of water levels and surges become available, the comparison should be revisited, as these results will help us to learn more about the causes of the mismatch and how it should be corrected.

## References

- van den Brink, H. W. and Können, G. P. (2008), The statistical distribution of meteorological outliers. *GRL* **35**, L23702, doi: 10.1029/2008GL035967.
- van den Brink, H. W. en Können, G. P. (2011), Estimating 10000-year return values from short time series. *Int. J. Climatol.* **31**(1), 115–126.
- van den Brink, H.W. (2018), Extreme wind en druk in de ECMWF seizoenverwachtingen. Report TR-364, KNMI, de Bilt (in Dutch).
- van den Brink, H.W. (2020), Het gebruik van de ECMWF seizoenverwachtingen voor het berekenen van de klimatologie van extreme waterstanden langs de Nederlandse kust. Report TR-385, KNMI, de Bilt (in Dutch).
- Chhab, H. (2017), Basisstochasten WTI-2017 - Statistiek en statistische onzekerheid. *Report 1209433-012-HYE-0007*, Deltares, Delft (in Dutch).
- Czörgő, M. and Révész, P. (1978), Strong Approximations of the Quantile Process. *Ann. Stat.* **6**(4), 882–894.
- Dillingh, D., De Haan, L., Helmers, R., Können, G.P., Van Malde, J. (1993), De basispeilen langs de Nederlandse kust, Statistisch onderzoek, *Rapport DGW-93.023*, Rijkswaterstaat Dienst Getijdewateren (in Dutch).
- Drees, H. (2000), Weighted approximations of tail processes for  $\beta$ -mixing random variables. *Ann. Appl. Prob.* **10**(4), 1274–1301.
- Drees, H. (2003), Extreme quantile estimation for dependent data, with applications to finance. *Bernoulli* **9**(4), 617–657.
- Gardes, L. and Girard, S. (2006), Comparison of Weibull tail-coefficient estimators. *REVSTAT* **4**(2), 163–188.
- de Haan, L., Ferreira, A. (2006), *Extreme value theory - An introduction*. Springer.
- ECMWF (2018a), Implementation of Seasonal Forecast SEAS5. <https://confluence.ecmwf.int/display/FCST/Implementation+of+Seasonal+Forecast+SEAS5>, ECMWF, Reading.
- ECMWF (2018b), SEAS5 User Guide. [https://www.ecmwf.int/sites/default/files/media\\_library/2017-10/System5\\_guide.pdf](https://www.ecmwf.int/sites/default/files/media_library/2017-10/System5_guide.pdf), ECMWF, Reading.
- Eilander, D. (2014), Herberekening basispeilen in het kader van WTI-2017. *Memo 1209431-003-ZWS-0002*, Deltares, Delft.
- de Valk, C. (2016), Approximation of high quantiles from intermediate quantiles. *Extremes* **19**, 661–686.
- de Valk, C. (2016). Approximation and estimation of very small probabilities of multivariate extreme events. *Extremes* **19**, 687–717.

- 
- de Valk, C. and Cai, J.J. (2018), A high quantile estimator based on the log-Generalised Weibull tail limit. *Econometrics and Statistics* **6**, 107–128.
- de Valk, C. (2020), Standard method for determining a climatological trend. *Report TR-389*, KNMI, De Bilt.
- de Valk, C.F. and van den Brink, H.W. (2020), Estimation of wind speeds with very high return periods from large datasets generated by weather prediction models : statistical aspects. *Report WR-2020-01*, KNMI, De Bilt.
- de Valk, C.F. and van den Brink, H.W. (2020), Evaluation of tail models and datasets for analysis of the extreme North Sea wind climate. *Submitted manuscript, available upon request.*

## Part II. Appendix

## A Return level estimates of skew surge

The following figures are the analogues of Figures 3 and 4 for skew surge instead of high-tide water level.

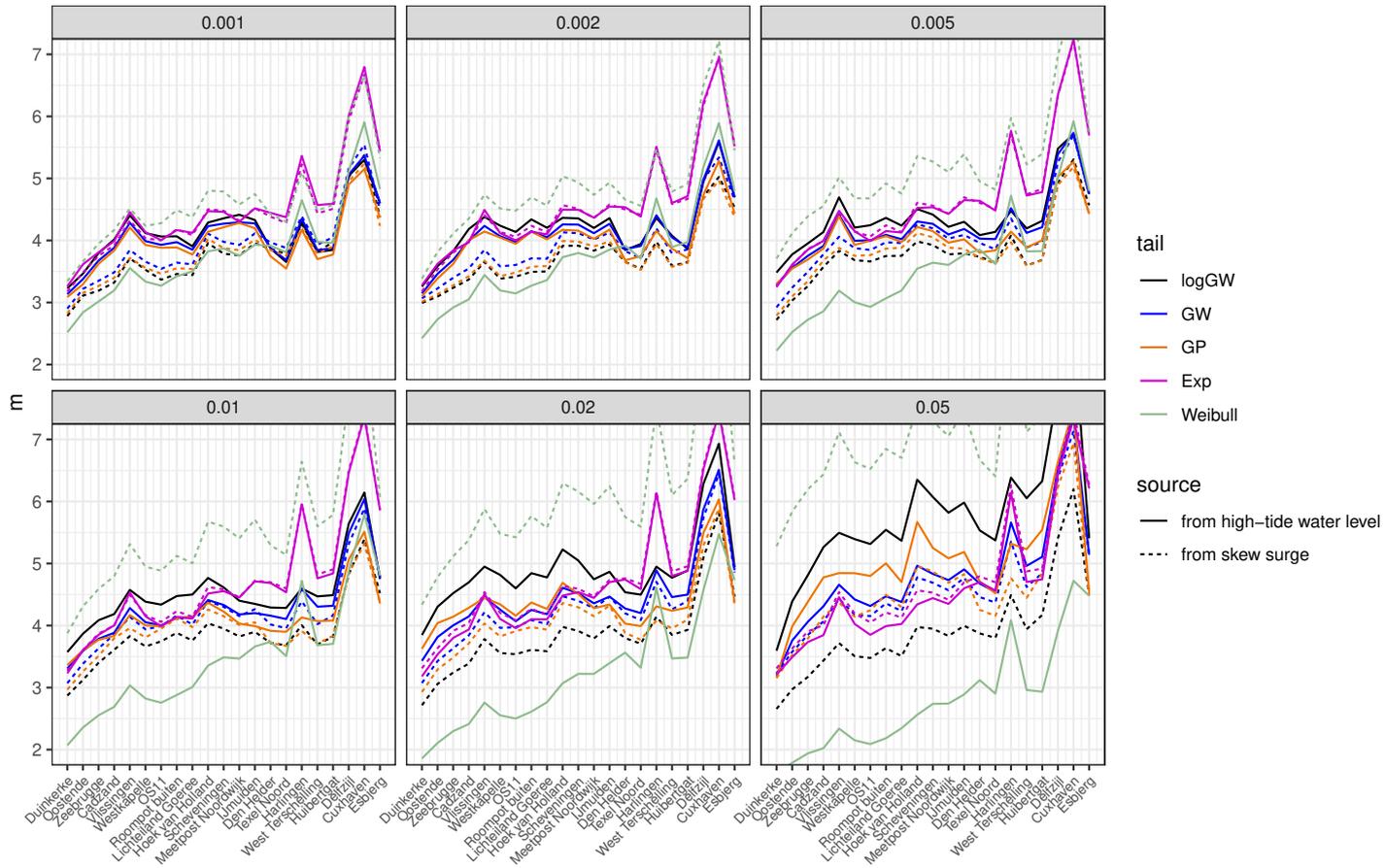


Fig. 25: Return level estimates of skew surge for a return period  $R = 10^7$  years as function of location index. Estimates are based on sample fractions indicated above the panels (these correspond to frequencies of exceedance of 0.7, 1.4, 3.4, 7, 14 and 35/year).



## B Estimation of the GW shape parameter as a function

For large sample size  $n$  and a suitable range of  $p$  (depending on sample size  $n$ ), the following approximation follows from eq. (27) in de Valk & Cai (2018):

$$n^{1/2} \frac{\hat{\rho}_p - \tilde{\rho}(\log(1/p))}{\log(1/p)} = p^{-1} B(p) \quad (4)$$

in which  $B$  is a Brownian motion<sup>6</sup>: Brownian motion is the continuous analogue of a random walk: increments of  $B$  over disjoint intervals of  $[0, 1]$  are independent zero-mean normal random variables with variances equal to the lengths of these intervals.

According to (4), the variance  $\mathbb{E}|\hat{\rho}_p - \tilde{\rho}(\log(1/p))|^2$  of  $\hat{\rho}_p$  is  $n^{-1}p^{-1}(\log(1/p))^2$ , so estimates  $\hat{\rho}_p$  fluctuate strongly at low  $p$ . This is indeed observed in practice, for example in Figure 9.

We only have values of  $\hat{\rho}_p$  at discrete values  $p_1 < \dots < p_n$  of  $p$ . By (4), the differences  $\Delta\hat{\rho}_{p_i} := \hat{\rho}_{p_{i+1}} - \hat{\rho}_{p_i}$  for different  $i$  are independent normal random variables with mean  $\tilde{\rho}(\log(1/p_{i+1})) - \tilde{\rho}(\log(1/p_i))$  and variance

$$\text{Var}(\Delta\hat{\rho}_{p_i}) = n^{-1}(p_i^{-1}(\log(1/p_i))^2 - p_{i+1}^{-1}(\log(1/p_{i+1}))^2).$$

From this information, it is straightforward to derive the maximum likelihood estimator for the parameters  $\delta$  and  $c$  of the function  $\tilde{\rho}$  given by (3).

---

<sup>6</sup> In the application, a small correction is applied to (4) to account for the serial dependences of successive high-tide water level or skew surge values.

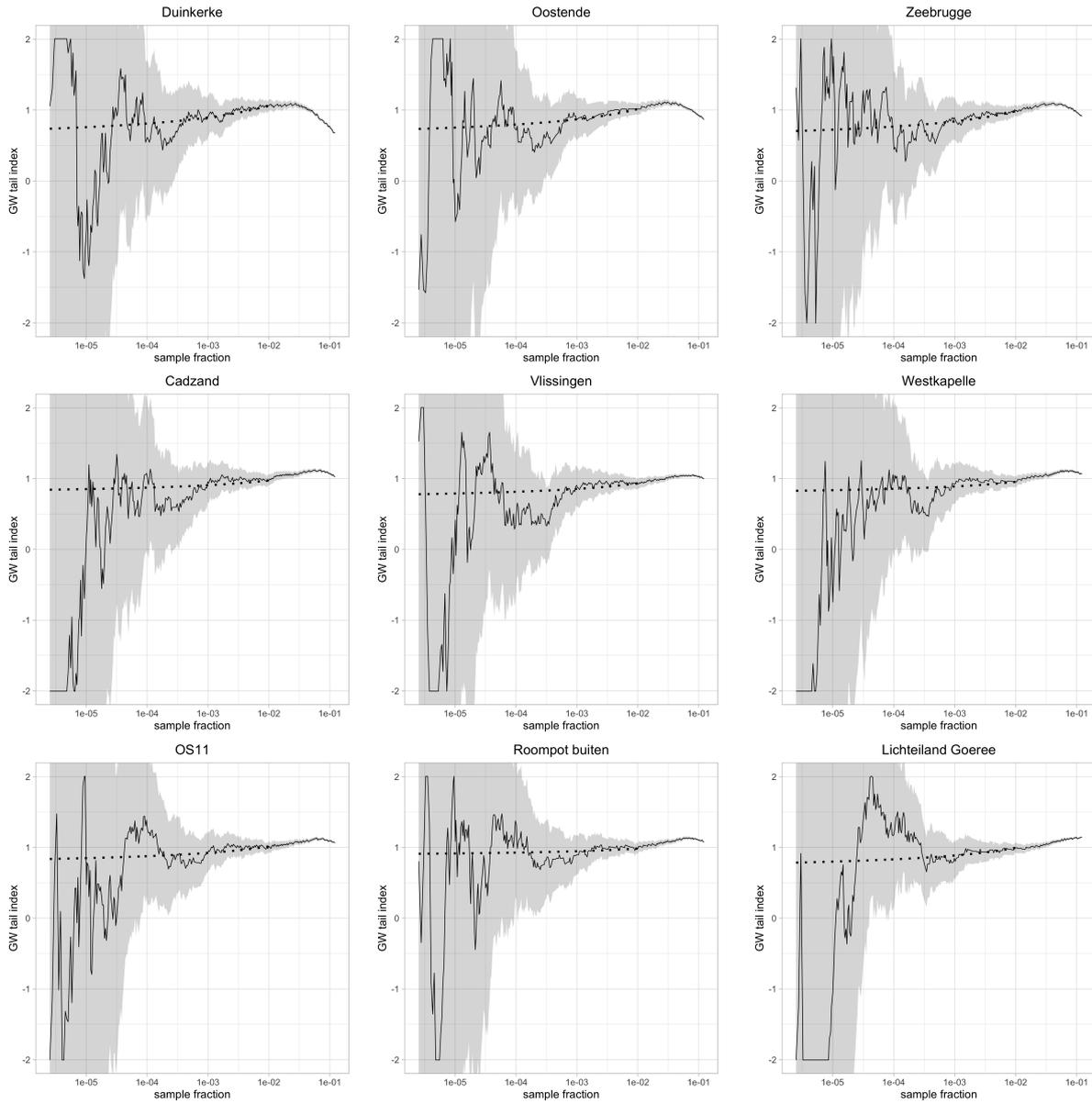


Fig. 27: GW-shape parameter estimates  $\hat{\rho}_p$  (full) from SEAS5/DCSM-5 data of high-tide water level as function of sample fraction  $p$ , with their two-sided 95% confidence intervals, for all stations. Dots indicate the estimated model (3), fitted to the values of  $\hat{\rho}_p$  for  $p \leq 0.01$ . The indicated sample fractions correspond to 0.007, 0.07, 0.7, 7.0 and 70/year.

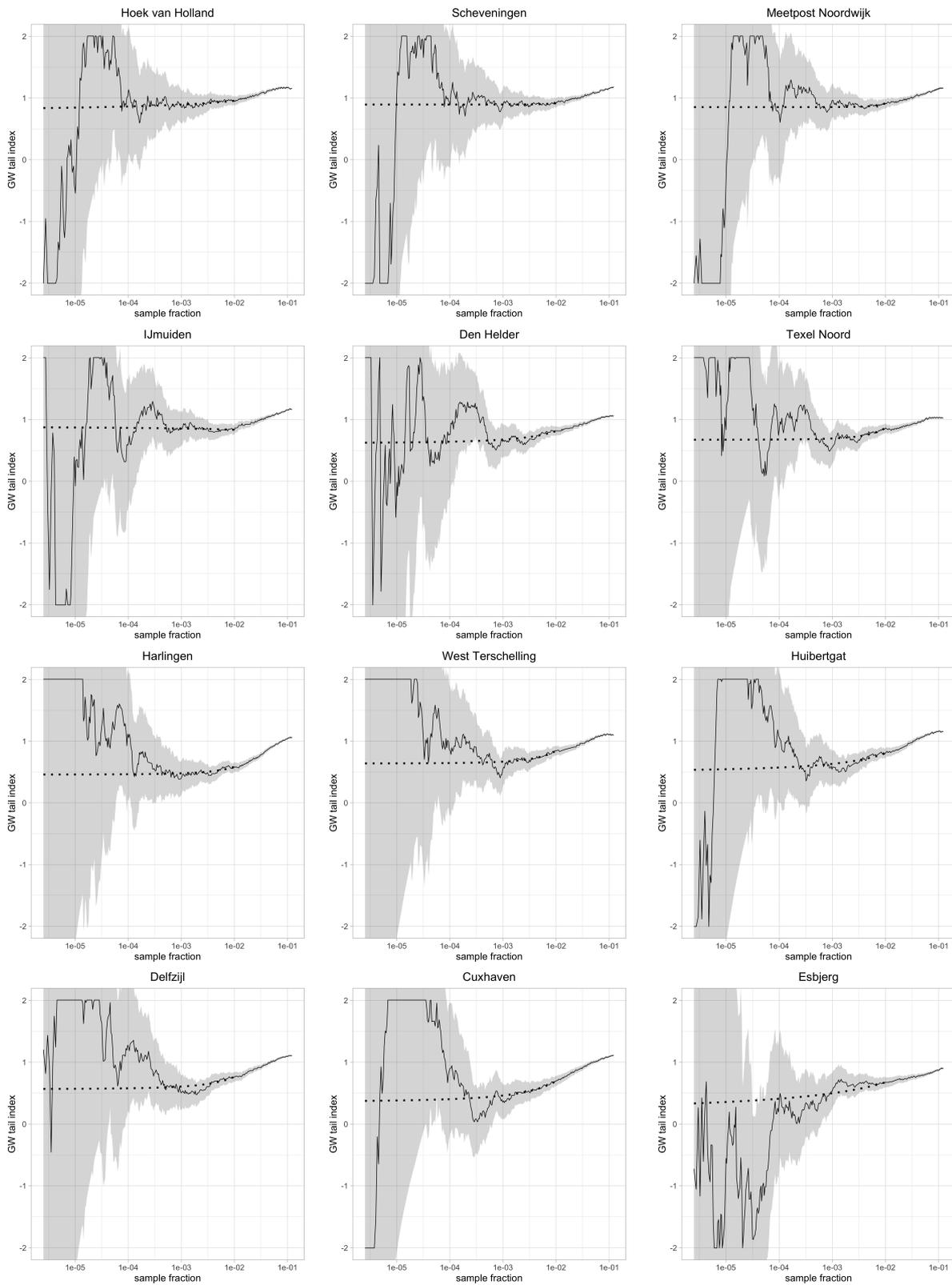


Fig. 28: Continued from Figure 27.

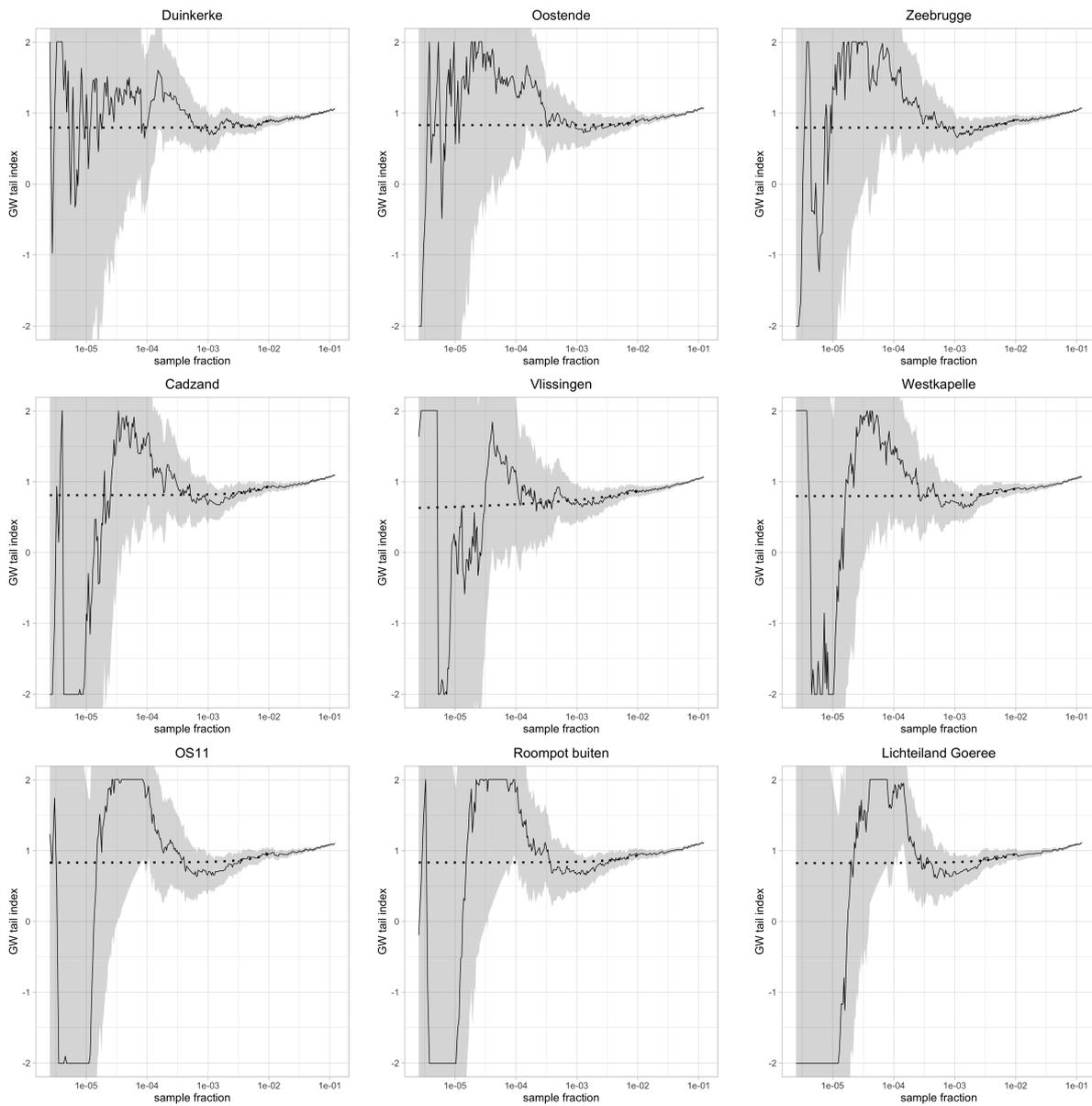


Fig. 29: Same as Figure 27, but for skew surge.

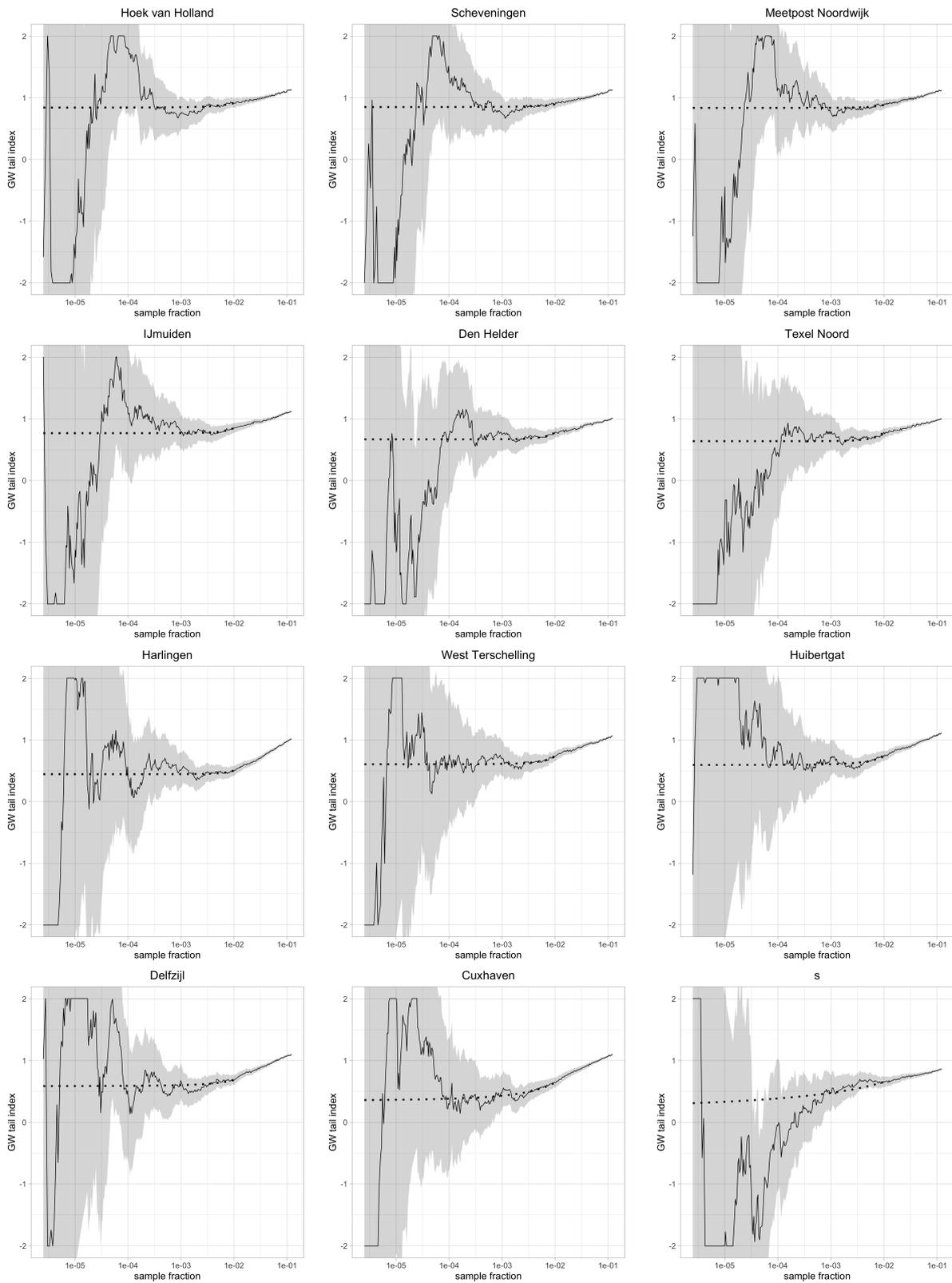


Fig. 30: Continued from Figure 29.

## C Error statistics of return values of wind speed estimated by Monte Carlo simulation

For wind speed and pseudo-wind speed in the central North Sea (analysed before in de Valk and van den Brink (2020b)), Figures 31 and 32 show the estimates for the error statistics of the  $10^7$  year return value of wind speed and pseudo-wind speed derived directly from data as in de Valk and van den Brink (2020b) and derived by the Monte Carlo method from Section 5, respectively.

The most notable difference between these figures is that estimates based on Monte Carlo simulation are much less noisy than those based on direct estimation from data. Note in particular the large variation in the reference values ("bias-corrected return values") on the right of Figure 31.

Another feature is that the minima of the RMS error from the MC simulation are larger than those directly estimated from the data. The difference between the curves of RMS error estimated directly from the data and from MC simulation can largely be explained by the variation in the reference values for the former on the right of Figure 31, which are caused by noise in the return value estimates. This indicates that the MC-based estimates are more reliable. However, qualitatively, both methods give similar results, and the conclusions from de Valk and van den Brink (2020b) are confirmed by the MC simulation.

Figure 32 also shows estimates of bias and RMS error for return values based on the GP tail with shape parameter estimated from the full sample (this was not yet shown in de Valk and van den Brink (2020b)). We see that the bias and RMS error are considerably larger than using the same method with the GW tail. Therefore, the GP tail is clearly not suitable for estimating return values of (pseudo) wind speed, even if the shape parameter can be estimated accurately. This conclusion contrasts with what was found for high-tide water level and skew surge.

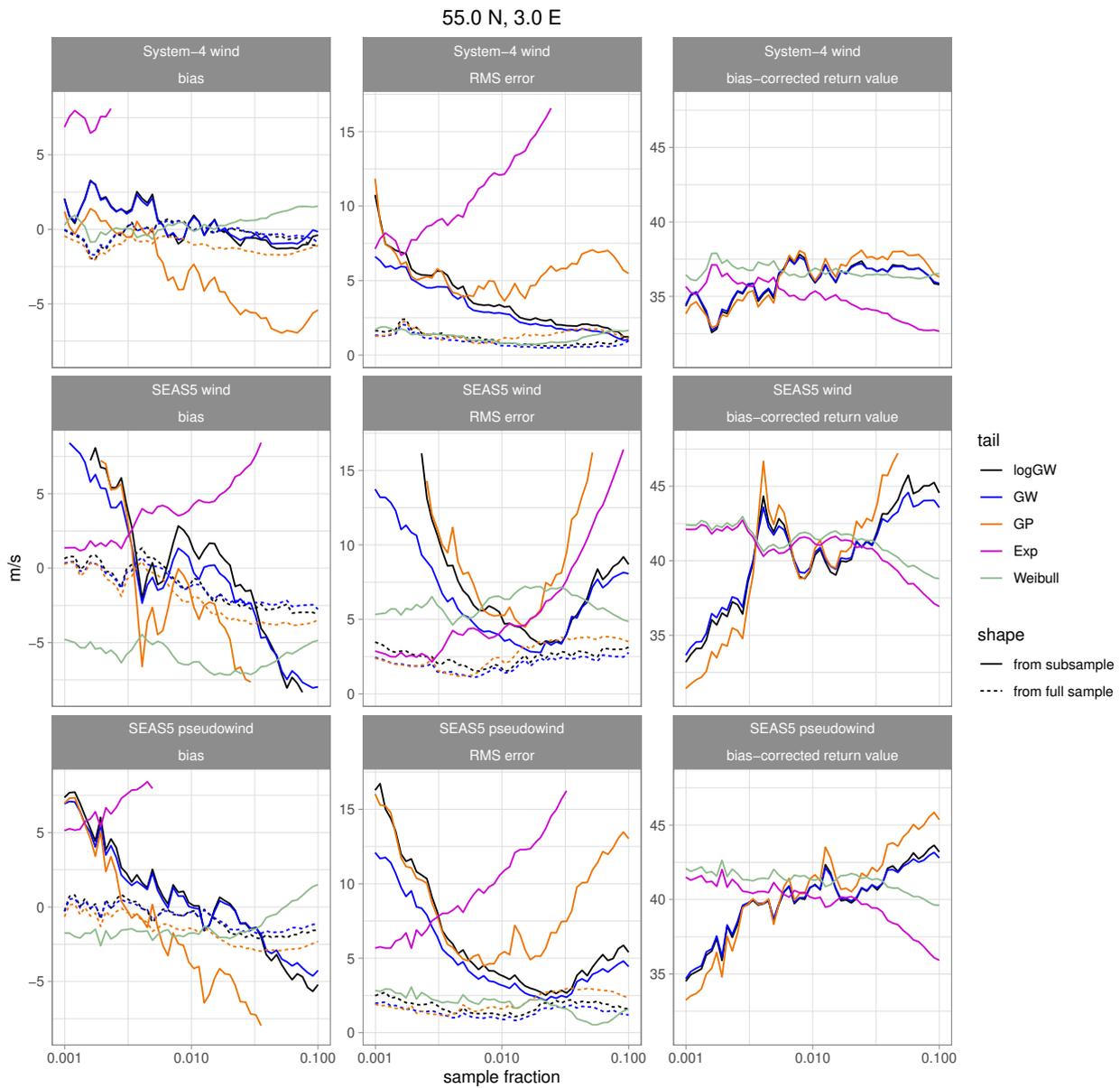


Fig. 31: Error statistics of  $10^7$ -year return level estimates of wind speed from System 4 and SEAS5 data (top and centre) and pseudo-wind from SEAS5 (bottom) derived directly from the data using the method from de Valk and van den Brink (2020a). The indicated sample fractions correspond to 0.7, 7.0 and 70/year.

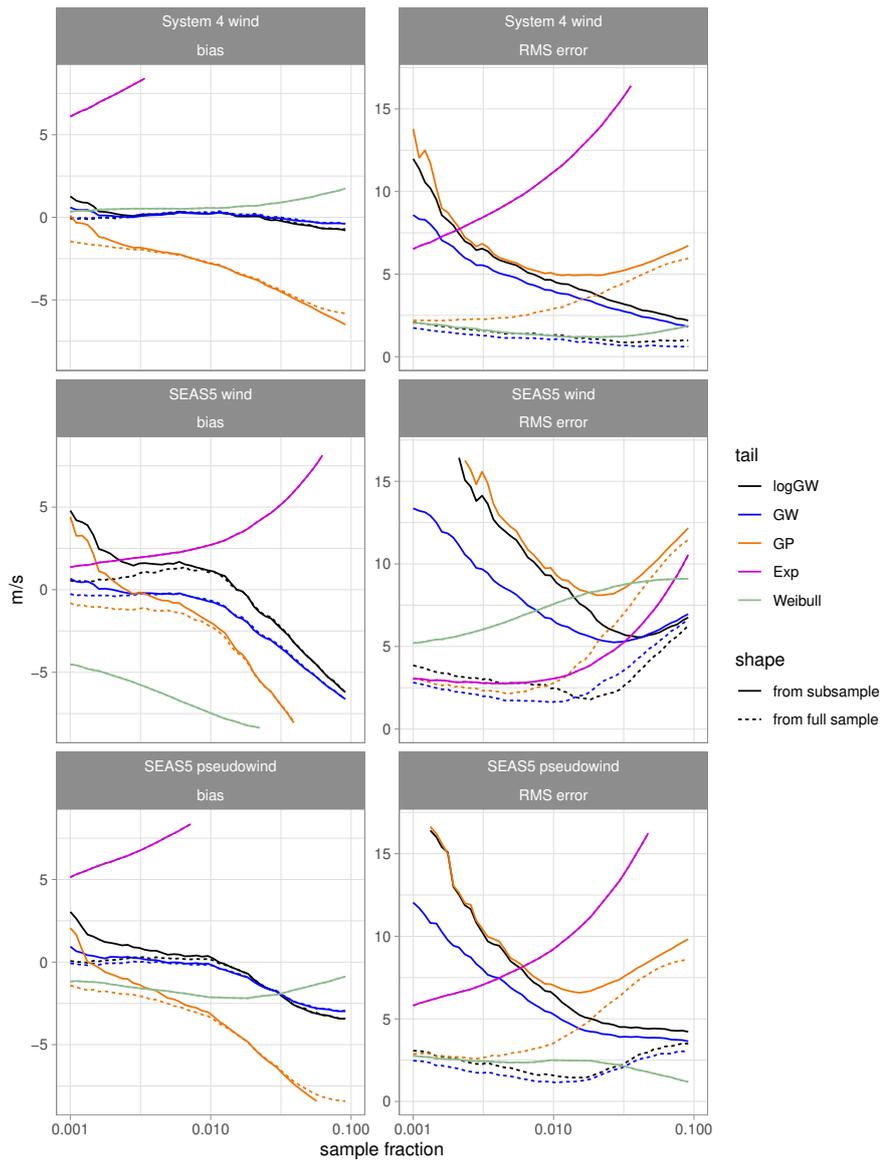


Fig. 32: As Figure 31, but derived from Monte-Carlo simulation based on plausible tails (see Section 5).

## D Comparison of estimates of the GP shape parameter from SEAS5/DCSM-5 data and measurements

In Figures 23 and 24, estimates of the GW shape parameter from SEAS5/DCSM-5 data and measurements are compared (see Chapter 6). Figures 33 and 34 show the same type of plots for the shape parameter of the GP tail.

Based on the relationship (2), we should expect qualitatively similar outcomes. Indeed, for sample fractions above 0.01 (frequencies above 7/year), GP shape estimates from SEAS5/DCSM-5 data are also generally lower than estimates from measurements.

A difference with the GW shape estimates is that for Harlingen and Delfzijl, for low sample fractions, the 95% confidence intervals of the the GP shape estimates from measurements drop below the estimates from SEAS5/DCSM-5 data. The narrow confidence bands for the GP shape estimates from measurements are caused by the high sensitivity of the GP tail to the shape parameter if it deviates substantially from zero, and indicates that a GP model is too crude for these tails.

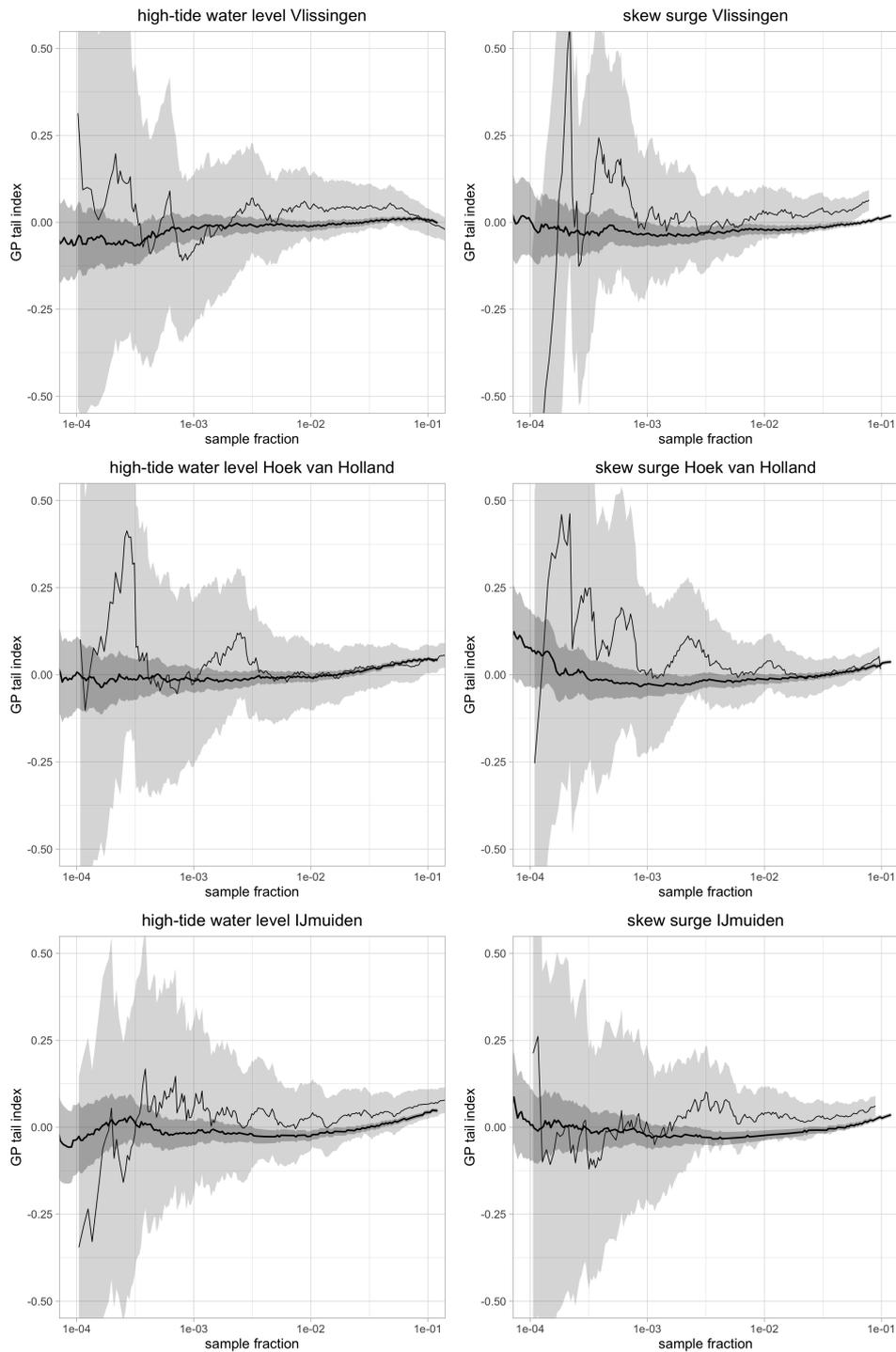


Fig. 33: GP shape estimate vs. sample fraction with 95% confidence interval from measurements (thin line, light shading) and from SEAS5/DCSM-5 simulations (thick line, dark shading) of high-tide water level (left) and skew surge (right) at three tide gauge stations. The indicated sample fractions correspond to 0.07, 0.7, 7.0 and 70/year.

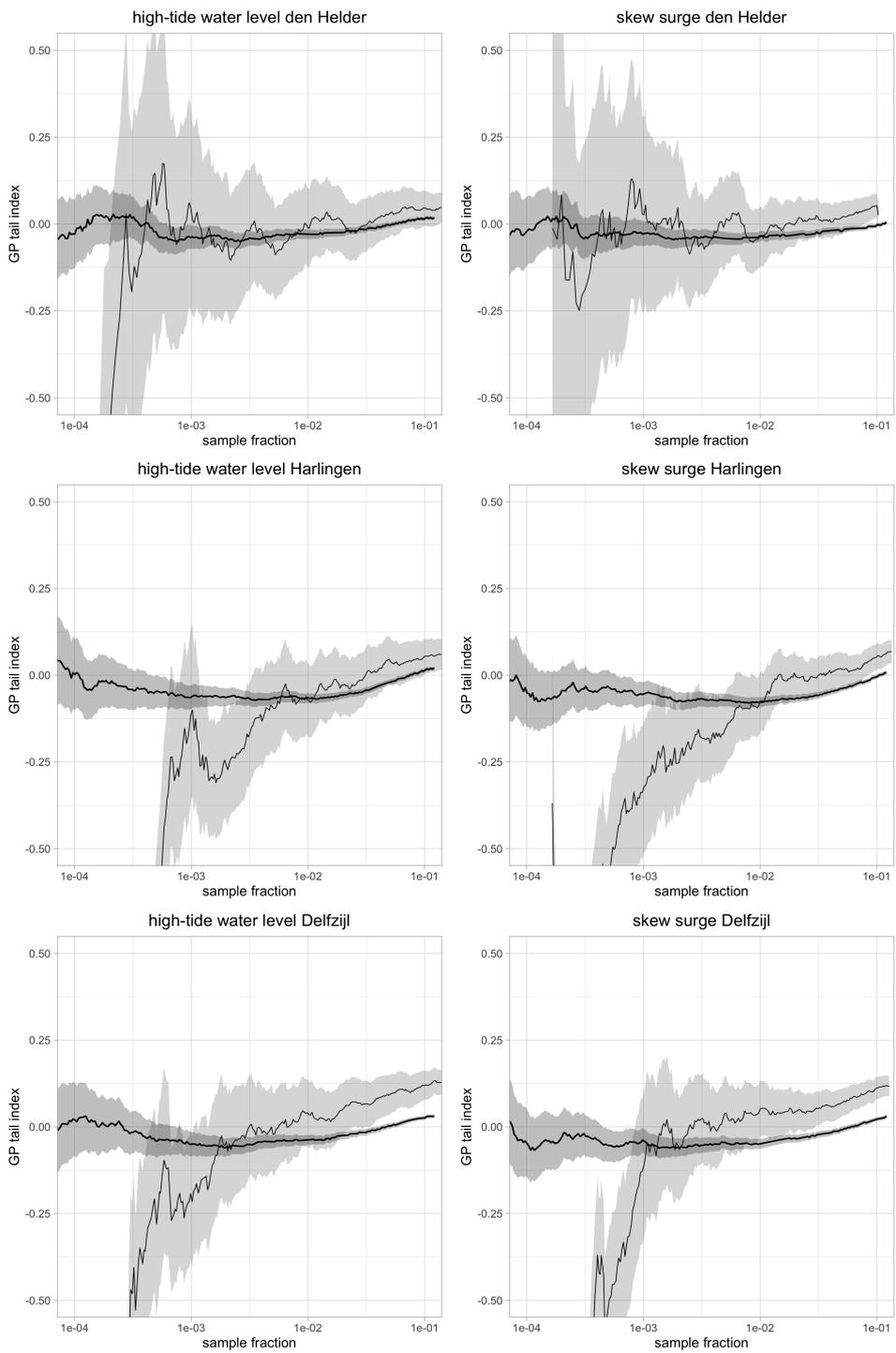


Fig. 34: As Figure 24 for three other tide gauge stations.

## E Accounting for serial dependence

Serial dependence in the SEAS5/DCSM-5 may increase the variance of return value estimates (its effect is the same as of a reduction of the sample size).

For very high water levels or surges, it is weak (for example, estimates of the extremal index are close to 1). However, when using a large sample fraction for tail estimation, the effect of serial dependence on the variance of parameter and return value estimates is not negligible. To account for this, the estimated variance of a return level estimate from the synthetic data is multiplied by a threshold-dependent correction factor estimated from the SEAS5/DCSM-5 data. This factor is identical to the factor derived in Section 4 of Drees (2000) for the Hill and maximum likelihood estimators of the shape parameter of the GP tail. This approach is based on a heuristic argument that it also applies to the estimators of the parameters of the GW, log-GW and Weibull tails used in the present study. Most of the theory in Drees (2000, 2003) is also applicable in our context, with a slightly modified assumption on the serial dependence. Furthermore, in the large-sample limit, the parameter estimators GW, log-GW and Weibull tails are localized (they converge to local values of the quantile function and of differential operators applied to it), just like the Hill and maximum likelihood estimators considered in Drees (2000, 2003). A rigorous analysis of this problem would of course be more satisfying than the heuristic approach sketched above.

The proposed correction factor is implemented in the module `r11.R` of the R-package `EVTools` (<https://github.com/ceesfdevalk/EVTools>).

**Royal Netherlands Meteorological Institute**

PO Box 201 | NL-3730 AE De Bilt  
Netherlands | [www.knmi.nl](http://www.knmi.nl)