

Bias correction of NORA3 wave hindcast data using wave buoy observations along the Norwegian coast

S Ameer^{1,2}, P S Reed³, L Jiao³, Z Jiang² and L Cappiatti¹

¹Department of Civil and Environmental Engineering, University of Florence, Italy

²Department of Engineering Sciences, University of Agder, Grimstad, Norway

³Department of Information and Communication Technology, University of Agder, Grimstad, Norway

E-mail: shoaib.ameer@unifi.it, per.simen.reed@outlook.com, lei.jiao@uia.no, zhiyu.jiang@uia.no, lorenzo.cappiatti@unifi.it

Abstract. Accurate estimates of extreme wave heights are essential for the design and operation of coastal and offshore infrastructure. Although long-term wave hindcasts provide valuable spatial and temporal coverage, systematic biases can affect their reliability, particularly in nearshore environments. This study presents a bias-correction framework for significant wave height (H_s) from the Norwegian reanalysis (NORA3) using wave-buoy observations from Fauskane, Fedjeosen, and Vestfjorden along the Norwegian coast. Multiple statistical and machine-learning correction methods were combined using a Mixture-of-Experts (MoE) approach, in which an XGBoost-based gating model assigned input-dependent weights to each correction method. The combined MoE model was subsequently applied to three selected Norwegian cities (i.e., Kristiansund, Bergen, and Stavanger) to investigate the influence of bias correction on extreme-wave return level estimates. Results showed that NORA3 biases varied across locations, highlighting the need for site-specific correction. Local calibration consistently improved hindcast accuracy, with the largest improvement observed at Fedjeosen, where the MoE approach reduced the root mean square error by 26.6%. In cross-location transfer, the MoE model trained at Fedjeosen reduced the upper-tail root mean square error at Fauskane by 29.2%. The corrected wave series also affected extreme-value estimates, changing 50-year return levels by up to 0.57 m. These findings demonstrate the potential of data-driven bias correction to improve hindcast-based assessments of extreme wave conditions.

1 Introduction

Reliable prediction of coastal wave conditions is fundamental to the design and operation of marine infrastructure. The significant wave height (H_s) is a key statistical parameter for the characterization of wave heights and can be used to assess structural loads, operational safety, and extreme-wave hazards for offshore structures including breakwaters and aquaculture facilities [1]. As coastal developments expand into increasingly energetic environments, the demand for accurate long-term wave information has become increasingly important for engineering design and risk assessment. Along the Norwegian coast, the combination of complex bathymetry, irregular coastlines, and highly variable exposure produces spatially heterogeneous wave conditions that are difficult to represent accurately through numerical simulation models[2]. Consequently,

improving the accuracy of coastal wave predictions is critical for supporting reliable engineering design and operational decision-making.

Long-term wave hindcast datasets have become a valuable source of information for coastal and offshore engineering by providing continuous spatial and temporal wave records over extended periods. Among these, the Norwegian reanalysis (NORA3) offers high-resolution wave hindcasts for the Norwegian, North and Barents Seas at a spatial resolution of 3 km, making it well suited for the assessment of wave climate, the design of coastal infrastructure, and extreme-value analysis (EVA) [3]. Despite these advantages, hindcast products may exhibit systematic discrepancies relative to local observations, particularly in nearshore environments where local-scale processes that influence wave evolution cannot be fully resolved. These discrepancies can bias wave statistics and, consequently, affect estimates of design wave conditions and extreme return levels [4]. Therefore, improving agreement between hindcast data and buoy observations is essential to enhance the reliability of coastal wave predictions and support robust engineering applications.

To reduce these systematic discrepancies, researchers have developed a wide range of statistical methods (SM) and machine-learning (ML) approaches to correct bias in wave hindcasts. Here, bias refers to a systematic difference between hindcast estimates and corresponding buoy observations. Conventional SM modify model outputs to align their statistical characteristics with observational data and have been widely adopted for their simplicity and computational efficiency [5]. More recently, ML approaches have demonstrated considerable potential to capture complex nonlinear relationships between hindcast data and observations, offering greater flexibility in modeling systematic errors [6]. Despite these advances, the performance of individual bias correction (BC) methods remains dependent on factors such as local wave characteristics, data availability, and model complexity. Consequently, no single approach consistently performs best across different coastal environments and wave conditions, highlighting the need for more robust and flexible bias correction frameworks.

Despite progress with both SM and ML approaches, several challenges remain. Existing studies have focused primarily on improving prediction accuracy using individual bias-correction methods, while comparatively little attention has been paid to frameworks that integrate multiple correction models to exploit their complementary strengths. Furthermore, the transferability of bias correction models between coastal locations has received limited investigation, despite its importance for extending corrections to regions where observational data are unavailable. In addition, the influence of bias correction on extreme-wave return-level estimation has not been comprehensively assessed, even though these estimates underpin the design and safety assessment of coastal and offshore infrastructure. Addressing these challenges requires a robust bias-correction framework that not only improves hindcast accuracy but also enhances the reliability of engineering-relevant extreme-wave characterization.

To address these challenges, this study develops a Mixture-of-Experts (MoE)-based framework to correct biases in significant wave height from NORA3 using in-situ buoy observations. The proposed framework combines multiple BC models through a data-driven weighting strategy to improve forecast accuracy under varying wave conditions. Its performance is evaluated for both local bias correction and cross-location transferability using observations from Norwegian coastal wave buoys. Furthermore, the corrected hindcast datasets are used to investigate the influence of bias correction on extreme-wave return level estimation through EVA. The main contributions of this study are: (i) the development of a Mixture-of-Experts framework for NORA3 wave hindcast bias correction, (ii) a systematic evaluation of local and cross-location model performance, and (iii) an assessment of the impact of bias correction on engineering-relevant extreme-wave return levels.

2 Methodology

2.1 Study area and datasets

This paper focuses on the Norwegian coast, a region characterized by complex coastal morphology and highly variable wave conditions (Figure 1). The primary numerical dataset used in this study was NORA3, a high-resolution regional wave hindcast with a spatial resolution of approximately 3 km, covering the Norwegian, North, and Barents Seas [3]. NORA3 provides continuous long-term wave information and was used as the baseline dataset for bias correction and subsequent extreme-wave analysis [7]. To develop and evaluate the bias correction models, significant wave height (H_s) data from NORA3 were paired with quality-controlled in-situ observations obtained from the Norwegian Meteorological Institute's MOTUS wave-buoy network [8] over overlapping time periods. As shown in Figure 1, Fauskane and Fedjeosen were designated as the core buoys and used for model development, calibration, and cross-location transfer assessment. Vestfjorden was excluded from model training and calibration and served as an external validation buoy to independently evaluate the spatial transferability and generalization capability of the proposed framework. The trained models were subsequently applied to three additional study-area locations, namely Kristiansund, Bergen, and Stavanger, to investigate the influence of bias correction on engineering-relevant extreme-wave return-level estimates through EVA.

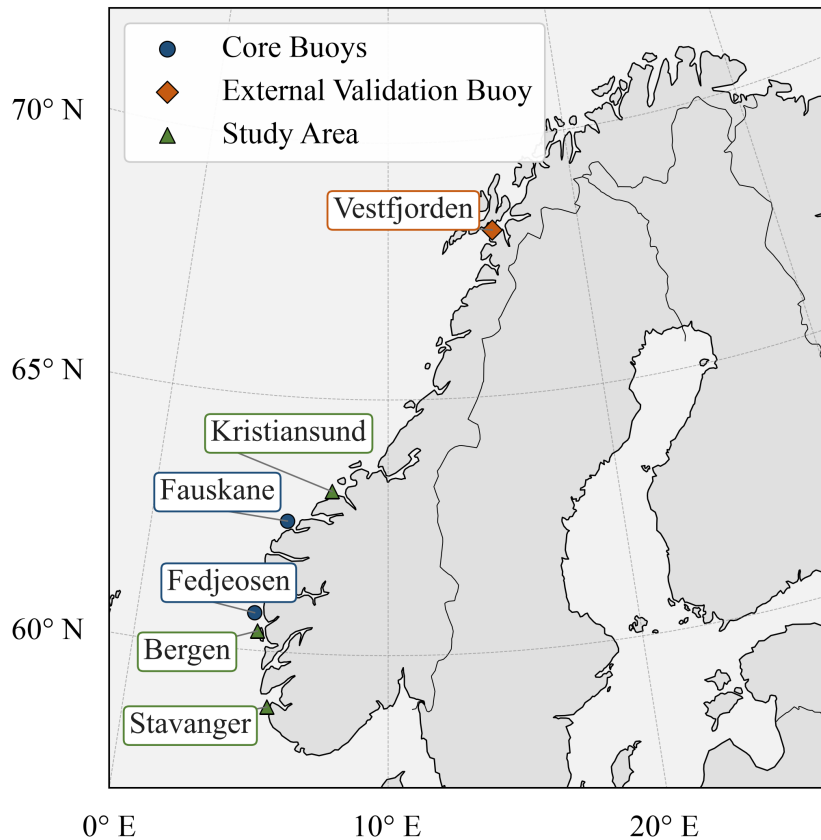


Figure 1. Study area showing the locations of the wave buoy stations used in this study.

2.2 Bias correction framework

Figure 2 illustrates the overall framework of the proposed methodology. The workflow begins by preparing paired NORA3 hindcast and wave buoy observations through preprocessing and temporal alignment. These paired datasets are then used to develop and evaluate multiple bias-correction models for H_s . The corrected predictions from the individual models are then integrated into a MoE framework, where a gating network adaptively combines the outputs to produce a single bias-corrected estimate. Finally, the corrected hindcast series are evaluated against buoy observations using statistical performance metrics and are applied to EVA to quantify the influence of bias correction on return-level estimation and associated uncertainties.

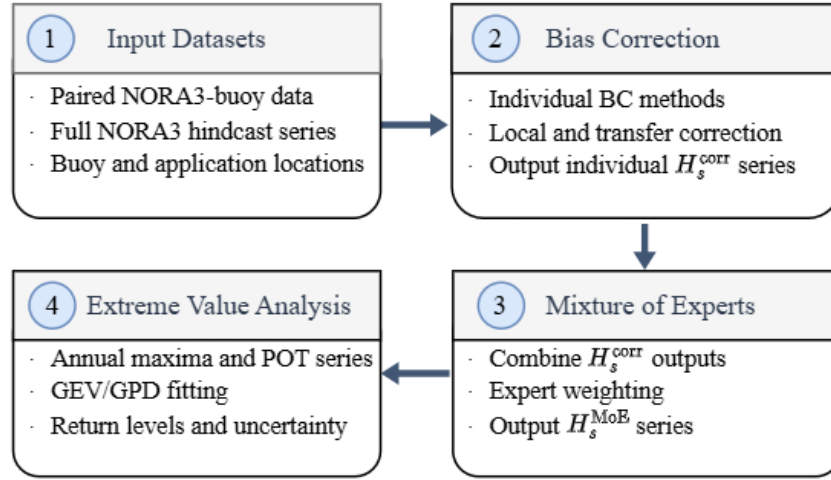


Figure 2. Workflow of the methodology employed in this study.

2.3 Statistical bias correction

Statistical bias correction methods aim to reduce systematic differences between hindcast simulations and buoy observations by establishing statistical relationships between the two datasets. Unlike machine-learning approaches, these methods rely mainly on predefined mathematical relationships or distribution-based transformations to adjust the hindcast data. This study implemented three SM: Linear Regression (LR), Parametric Quantile Mapping (PQM), and the Directional Adjusted Gumbel Quantile Mapping (DAGQM) [9]. These methods range from a simple regression-based correction to distribution-based approaches designed to correct biases across the wave-height distribution.

LR provides a baseline correction by fitting a linear relationship between NORA3 hindcast and observed H_s [9]. Although computationally efficient and easy to interpret, it has limited ability to represent complex or nonlinear bias patterns.

Distribution-based approaches were adopted to correct biases that vary across different quantiles of the wave-height distribution [5]. PQM adjusts the cumulative distribution of the hindcast data to match that of the observations by fitting theoretical probability distributions to both datasets [9]. Building upon this principle, DAGQM incorporates wave-direction information into the correction process, allowing the statistical transformation to account for directional variations in wave characteristics [9]. This directional adjustment is intended to improve bias correction under varying wave climates, particularly for extreme wave conditions where directional effects become more pronounced.

2.4 *Machine learning-based bias correction*

ML approaches formulate bias correction as a supervised learning problem by learning the relationship between hindcast variables and buoy observations directly from data [6, 10]. Compared with conventional SM, these approaches can capture complex nonlinear relationships and interactions among predictors without relying on predefined functional relationships. In this study, three ML models, Gaussian process regression (GPR) [11], Extreme gradient boosting (XGBoost) [12], and Transformer [13], were implemented to investigate their ability to improve the correction of H_s under varying wave conditions.

GPR is a probabilistic machine-learning approach based on a Bayesian non-parametric framework for modelling nonlinear relationships while providing predictive uncertainty [11, 14]. Its flexible kernel formulation enables it to capture smooth and complex patterns that may not be adequately represented by simple linear models. In this study, GPR was applied to model the nonlinear relationship between hindcast and observed wave heights and to correct systematic biases in the hindcast data.

XGBoost is an ensemble learning algorithm based on gradient-boosted decision trees that constructs a sequence of weak learners to minimize prediction error [12]. Its ability to model nonlinear relationships, feature interactions, and heterogeneous data makes it well suited for bias correction of wave hindcasts. In this study, the model was trained on paired NORA3 hindcast and buoy observations, along with relevant meteorological and oceanographic predictors, to estimate the corrected H_s . Regularization and optimized hyperparameters were used to improve model generalization and reduce overfitting.

The Transformer model was adopted to exploit its ability to learn long-range dependencies in sequential data through a self-attention mechanism [13]. Unlike recurrent neural networks, the Transformer processes the entire input sequence simultaneously, enabling efficient modelling of temporal relationships in wave conditions. To improve the representation of extreme events, the Transformer was implemented using a tail-aware residual learning strategy, in which the model predicts the residual correction relative to a baseline estimate while giving increased emphasis to the upper tail of the wave-height distribution. This approach enhances the model's ability to represent rare, high-impact wave conditions that are particularly important for engineering applications and EVA.

2.5 *Mixture-of-experts framework*

Individual bias correction (BC) methods have different strengths under varying wave conditions. Statistical methods generally provide robust corrections for systematic biases, whereas machine-learning models can capture more complex nonlinear relationships. The proposed Mixture-of-Experts (MoE) framework therefore combines six BC methods, namely LR, GPR, PQM, DAGQM, XGBoost, and Transformer, to exploit their complementary capabilities [15, 16].

Each pre-trained BC method acts as an expert and independently produces a corrected estimate of H_s . These predictions, together with NORA3 sea-state variables and features describing the agreement and spread among the experts, are provided to an XGBoost-based gating model. For each sample, the gate assigns non-negative, normalized weights according to the prevailing wave conditions and the relative performance of the experts [17, 18]. The final MoE-corrected H_s is calculated as the weighted sum of the six expert predictions. Unlike a fixed ensemble average, the assigned weights vary between samples, allowing the framework to adapt to both general and high-wave conditions. A tail-aware adjustment is applied above the Q_{95} and Q_{99} thresholds to increase the contribution of experts that perform more effectively under high-wave conditions. A more detailed description of the framework is provided by Reed [19]. The resulting corrected H_s series was subsequently used for performance evaluation and EVA.

2.6 Experimental design

The proposed framework was evaluated using paired NORA3 hindcast and MOTUS wave buoy observations from three locations along the Norwegian coast. Fauskane and Fedjeosen were selected as the primary training sites for developing the BC models, while Vestfjorden was reserved for independent validation to assess the framework's transferability. To investigate model generalization, both local and cross-location correction experiments were performed. In the local experiments, models were trained and evaluated using data from the same buoy location. In the cross-location experiments, models trained at one location were applied to another location without retraining, thereby assessing their capability to transfer bias correction between coastal sites with different wave characteristics.

Model performance was evaluated by comparing the corrected H_s against buoy observations using Root Mean Square Error (RMSE), Mean Absolute Error (MAE), Mean Bias Error (MBE), and the coefficient of determination (R^2). Because the primary objective of the study is to improve the representation of extreme wave conditions, additional upper-tail diagnostics, including quantile-based error statistics and exceedance rate bias (ERB), were used to evaluate model performance during high-wave events.

To assess the practical implications of bias correction for coastal engineering applications, the corrected hindcast datasets were subsequently used in EVA. Annual maxima of significant wave height were extracted and fitted using the Generalized Extreme Value (GEV) distribution, whose cumulative distribution function is given by

$$G(z) = \exp \left\{ - \left[1 + \xi \left(\frac{z - \mu}{\sigma} \right) \right]^{-1/\xi} \right\}, \quad (1)$$

where μ , $\sigma > 0$, and ξ represent the location, scale, and shape parameters, respectively, subject to $1 + \xi(z - \mu)/\sigma > 0$ [20]. The shape parameter determines the upper-tail behaviour of the distribution. For a return period T , the corresponding return level H_s was estimated as

$$z_T = \mu + \frac{\sigma}{\xi} \left[\left\{ -\ln \left(1 - \frac{1}{T} \right) \right\}^{-\xi} - 1 \right], \quad \xi \neq 0, \quad (2)$$

with the Gumbel limit applied as ξ approaches zero. Return levels derived from the BC datasets were compared with those obtained from the original NORA3 hindcast. This comparison enabled the influence of bias correction on long-term extreme-wave estimates and their associated uncertainty to be quantified.

3 Results and discussion

3.1 Assessment of the raw H_s from NORA3 bias

NORA3 hindcast was evaluated against MOTUS buoy observations to characterize systematic errors in H_s before applying bias correction. As shown in Figure 3, quantile-based analysis showed that prediction errors increased toward higher wave-height percentiles at all sites, indicating reduced hindcast accuracy during extreme wave conditions. This reduced accuracy under extreme wave conditions profoundly affects the determination of design waves for survivability studies based on numerical hindcasts. Enhancing model precision is therefore critical to improving the reliability of wave inputs and subsequent structural resistance assessment. Fedjeosen exhibited the largest overall error, followed by Fauskane, while Vestfjorden showed comparatively smaller errors despite increasing RMSE in the upper tail.

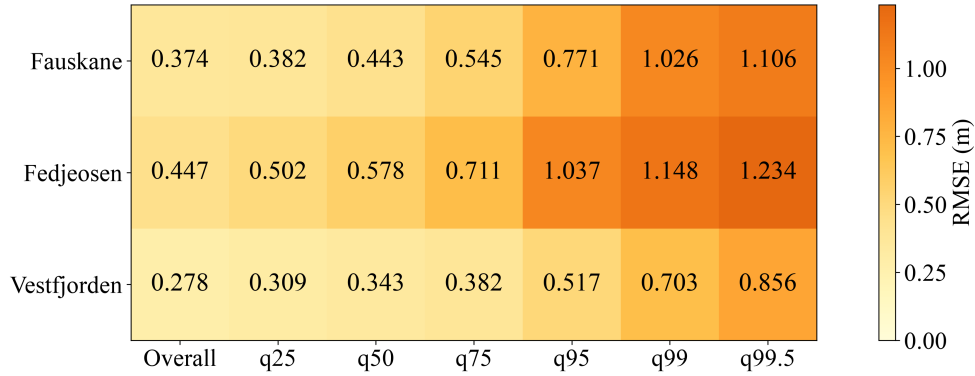


Figure 3. RMSE of raw NORA3 H_s across different quantiles at the buoy locations during the overlapping period.

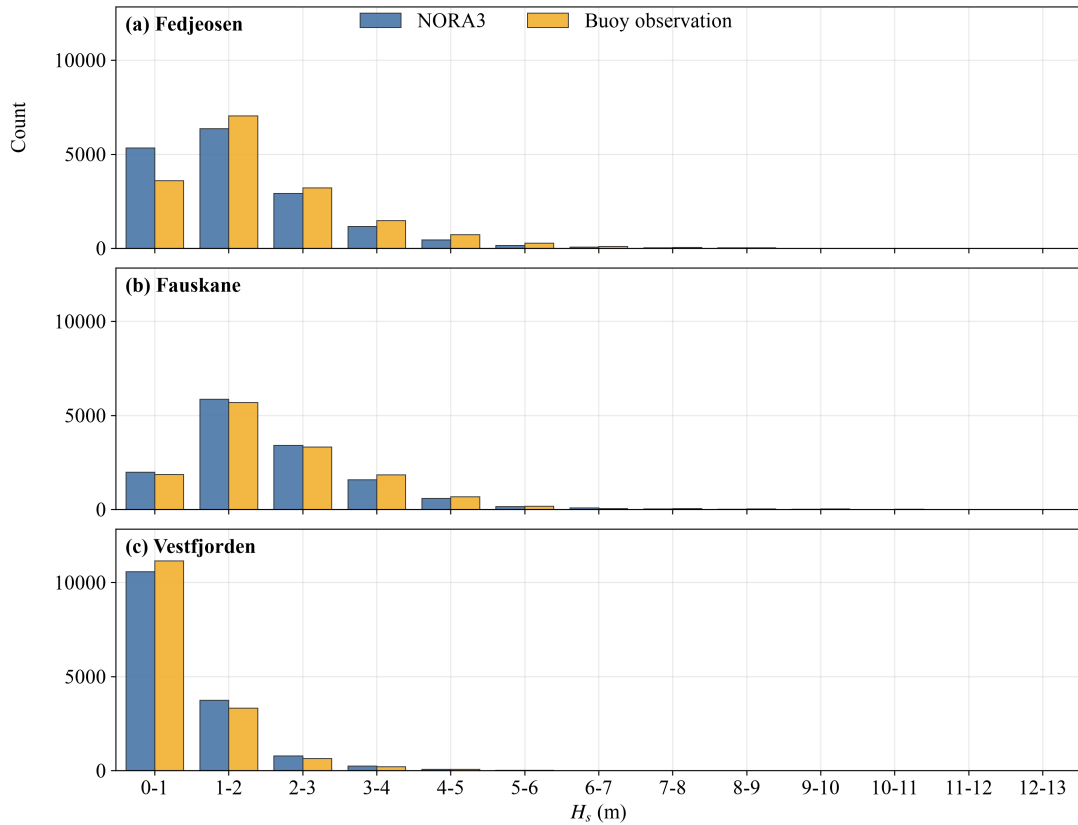


Figure 4. Comparison of H_s distributions from buoy observations and NORA3 during the overlapping periods at (a) Fauskane (2018–2019 and 2020–2021), (b) Fedjeosen (2023–2025), and (c) Vestfjorden (2022–2024).

The bias characteristics were strongly site-dependent. As illustrated in Figure 4, NORA3 consistently underestimated wave heights at Fauskane and Fedjeosen, with the underestimation becoming more pronounced for larger waves, whereas Vestfjorden exhibited a weak positive bias. The increasing underestimation at Fauskane and Fedjeosen may reflect limitations in representing storm winds and the resulting wave growth. The contrasting behavior at Vestfjorden is likely associated with its more sheltered fjord environment and lower wave-energy conditions

compared with the two exposed coastal sites. In this environment, incomplete representation of sheltering by islands and headlands may allow excessive wave energy to reach the model location, potentially contributing to the positive bias. The approximately 3 km model grid may also limit the representation of local bathymetry and its influence on wave propagation. However, further analysis is needed to confirm the contribution of these processes to the observed biases. These findings indicate that the NORA3 hindcast does not exhibit a uniform bias across the study area. The spatial variability and increasing upper-tail errors therefore highlight the need for site-specific bias correction to better represent extreme wave conditions before EVA.

3.2 Performance of bias correction methods

The performance of the SM and ML bias correction methods was evaluated under both local and cross-location validation using RMSE across the full distribution and the upper quantiles (q95 and q99). Overall, all methods improved the raw NORA3 hindcast to varying degrees, with the greatest improvements observed at Fedjeosen, where the initial bias was strongest (Figure 5).

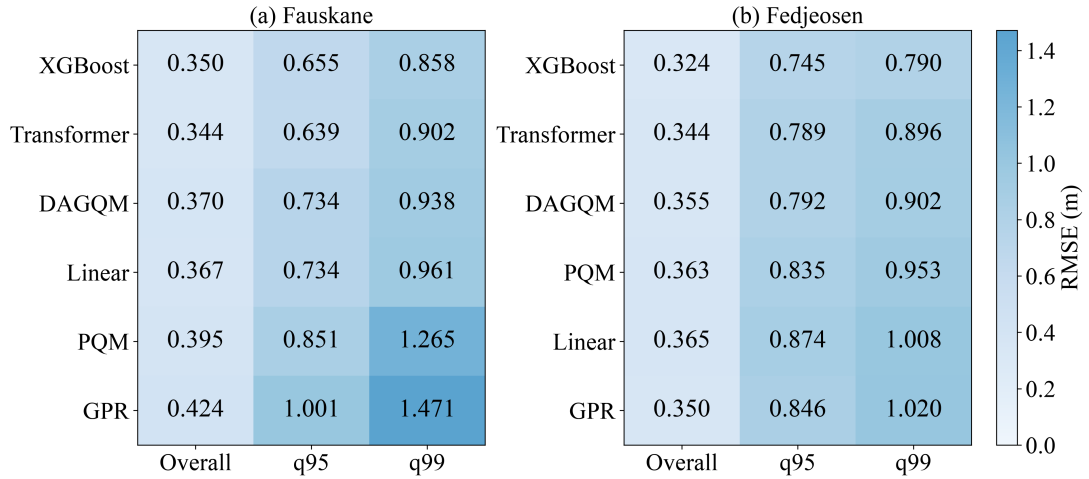


Figure 5. RMSE-based evaluation of local bias-correction methods at Fauskane and Fedjeosen, compared with raw NORA3 across the overall H_s distribution and selected upper quantiles.

Under local validation, XGBoost consistently performed best among the ML methods, while DAGQM was the strongest SM approach. As summarized in Table 1, XGBoost reduced the overall RMSE by 6.4% and the q99 RMSE by 16.4% at Fauskane. Larger improvements were obtained at Fedjeosen, where the overall and q99 RMSE decreased by 27.5% and 31.2%, respectively (Table 1). These results indicate that locally trained models effectively capture site-specific bias, particularly for extreme wave conditions.

Table 1. RMSE-based assessment of local BC performance for the overall distribution and upper tail.

Location	Method	RMSE	Δ RMSE	RMSE q_{95}	Δq_{95}	RMSE q_{99}	Δq_{99}
Fauskane	Best SM	0.370	-1.2%	0.734	-4.8%	0.938	-8.6%
Fauskane	Best ML	0.350	-6.4%	0.655	-15.0%	0.858	-16.4%
Fedjeosen	Best SM	0.355	-20.6%	0.792	-23.6%	0.902	-21.4%
Fedjeosen	Best ML	0.324	-27.5%	0.745	-28.1%	0.790	-31.2%

Cross-location validation produced more variable results (Figure 6). Models transferred between Fauskane and Fedjeosen generally improved upper-tail accuracy but were less consistent across the full distribution. The best transfer performance was obtained from Fauskane to Fedjeosen, where XGBoost reduced the overall RMSE by 9.8% and the q99 RMSE by 11.2% (Table 2). In contrast, correction models applied to Vestfjorden failed to improve the hindcast, reflecting the site’s distinct wave climate and bias characteristics. These findings suggest that transferable bias correction is feasible between locations with similar wave conditions but becomes less reliable when environmental characteristics differ substantially.

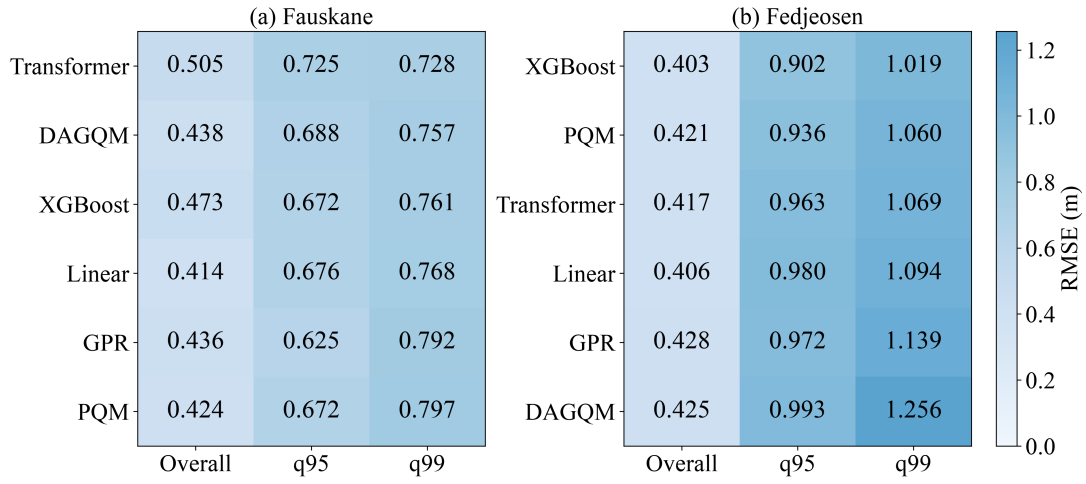


Figure 6. RMSE-based evaluation of cross-location bias-correction methods transferred between Fauskane and Fedjeosen, compared with raw H_s from NORA3 for overall performance and at selected upper quantiles.

Table 2. Cross-location BC performance relative to raw H_s from NORA3. The source site provides the data used to calibrate the BC model, whereas the target site indicates where the transferred model is assessed. The best SM and ML methods are selected based on the lowest transferred q_{99} RMSE.

Target	Source	Method	RMSE	Δ RMSE	RMSE _{q_{95}}	Δq_{95}	RMSE _{q_{99}}	Δq_{99}
Fauskane	Fedjeosen	Best SM	0.438	+17.1%	0.688	-10.8%	0.757	-26.2%
Fauskane	Fedjeosen	Best ML	0.505	+35.0%	0.725	-6.0%	0.728	-29.1%
Fedjeosen	Fauskane	Best SM	0.421	-5.8%	0.936	-9.7%	1.060	-7.7%
Fedjeosen	Fauskane	Best ML	0.403	-9.8%	0.902	-13.0%	1.019	-11.2%

3.3 Mixture-of-experts performance

The MoE framework was evaluated using the same RMSE and ERB metrics as the individual bias-correction methods. As shown in Figure 7, the local MoE consistently improved the raw NORA3 hindcast at both core buoy locations. At Fauskane, the local MoE reduced the overall RMSE by 8.0%, while the corresponding reduction at Fedjeosen reached 26.6%, confirming that adaptive combinations of multiple correction models can effectively reduce site-specific bias.

The transferred MoE produced more variable results. Transfer between Fauskane and Fedjeosen generally improved upper-tail accuracy but was less consistent across the full distribution. In the combined-transfer configuration, the MoE integrates information from both core buoy sites, Fauskane and Fedjeosen, to predict H_s at Vestfjorden. This configuration evaluates whether combining information from multiple source sites improves transfer performance at a target site without local calibration data. However, the combined-transfer configuration did not improve accuracy at Vestfjorden. These results indicate that the effectiveness of the MoE depends on the similarity between the source and target locations, with transfer becoming less reliable for sites exhibiting different wave climates and bias characteristics.

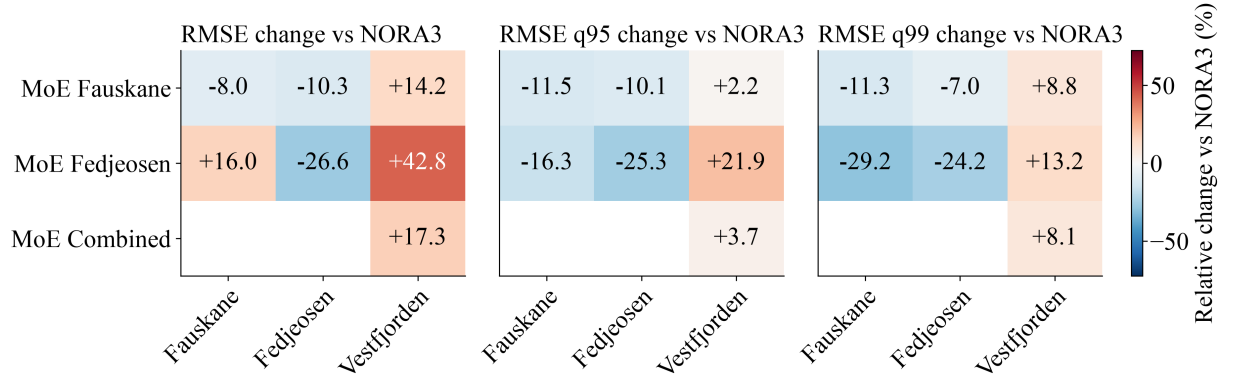


Figure 7. Relative change in RMSE for local, single-source transfer, and combined-transfer MoE configurations with respect to raw NORA3 H_s .

The transferred MoE produced more variable results. Transfer between Fauskane and Fedjeosen generally improved upper-tail accuracy but was less consistent across the full distribution. In the combined-transfer configuration, the MoE integrates information from both core buoy sites, Fauskane and Fedjeosen, to predict H_s at Vestfjorden. This configuration evaluates whether combining information from multiple source sites improves transfer performance at a target site without local calibration data. However, the combined-transfer configuration did not improve accuracy at Vestfjorden. These results indicate that the effectiveness of the MoE depends on the similarity between the source and target locations, with transfer becoming less reliable for sites exhibiting different wave climates and bias characteristics.

ERB analysis (Figure 8) further showed that the local MoE substantially improved the representation of high-wave exceedance frequencies, making it well suited for subsequent extreme value analysis. Examination of the expert weights (Figure 9) revealed that the gating network assigned relatively balanced weights under moderate conditions, while increasing the contribution of XGBoost and Transformer during high-wave events. Rather than consistently outperforming every individual model, the MoE provided a robust and adaptive correction across a range of sea states, supporting its selection for the extreme value analysis presented in the following section.

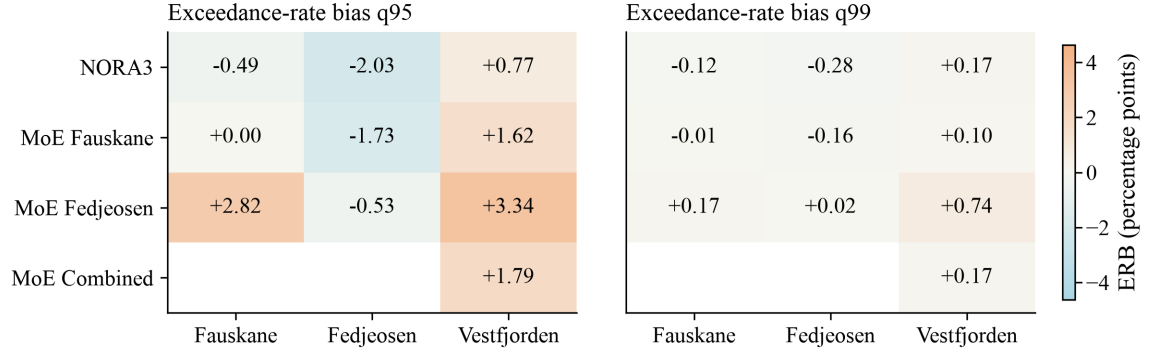


Figure 8. ERB-based assessment of the MoE configurations at the q_{95} and q_{99} exceedance levels of observed H_s .

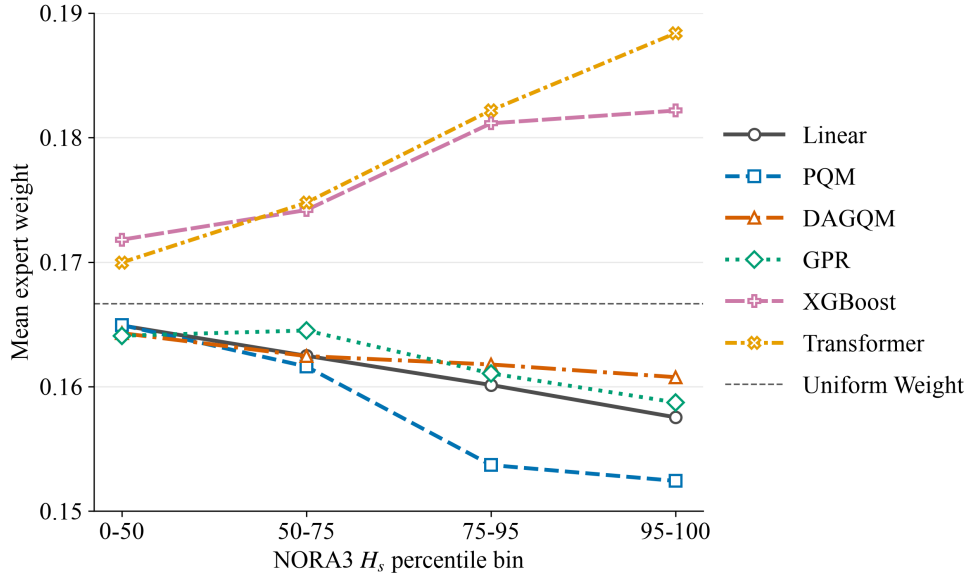


Figure 9. Expert weights for the local MoE at Fauskane across the NORA3 H_s percentile bins. The dashed line represents equal weighting among the six experts.

3.4 Impact on extreme value analysis

The GEV model fitted to annual maxima was used for the main return-level comparison. This approach was considered appropriate for consistently comparing the NORA3 and MoE-corrected H_s datasets. Further details on the selection and evaluation of the EVA approach are provided in the Reed [19]. For Fauskane and Fedjeosen, the locally trained MoE models were used for the EVA, while the combined-transfer MoE model was used for Kristiansund as a representative study-area location. Results for the other study-area locations are provided in the Reed [19]. As shown in Figure 10, the corrected datasets generally produced higher GEV return-level curves than the raw NORA3 hindcast, indicating that bias correction increased the estimated extreme significant wave heights. The largest increase was observed at Fauskane, while Fedjeosen exhibited a more moderate change despite its larger validation-period bias. This suggests that improvements in bias correction do not necessarily translate directly into proportional changes

in return-level estimates.

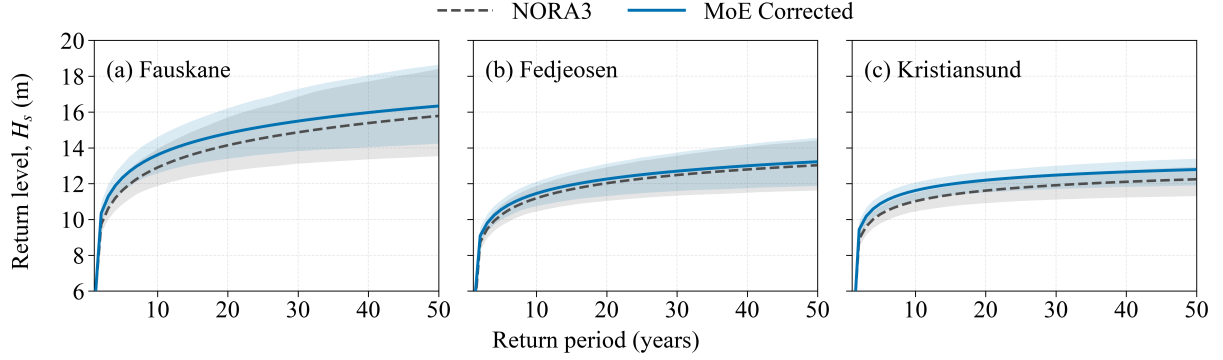


Figure 10. GEV-based return levels for raw NORA3 and MoE-corrected H_s at the selected locations, shaded areas denoting the 95% confidence intervals.

The changes in 50-year return levels are summarized in Table 3. Compared with the raw hindcast, the corrected datasets increased the estimated 50-year return level by 0.55 m at Fauskane, 0.19 m at Fedjeosen, and by 0.49–0.57 m at the study area locations. Confidence intervals (CI) were also slightly reduced after correction, although these reductions were modest and should be interpreted with caution.

Table 3. Estimated 50-year return levels of significant wave height ($H_{s,50}$), associated CI widths, and deviations from raw NORA3. Values are reported in metres.

Location	Source	$H_{s,50}$	CI lower	CI upper	CI width	$\Delta H_{s,50}$	ΔCI
Fauskane	NORA3	15.78	13.54	18.41	4.87	–	–
	MoE Fauskane	16.34	14.24	18.65	4.41	0.55	-0.46
Fedjeosen	NORA3	13.03	11.62	14.42	2.80	–	–
	MoE Fedjeosen	13.22	11.87	14.56	2.69	0.19	-0.11
Kristiansund	NORA3	12.25	11.31	12.94	1.63	–	–
	MoE Combined	12.79	11.90	13.40	1.50	0.54	-0.13
Bergen	NORA3	13.19	11.92	14.34	2.42	–	–
	MoE Combined	13.68	12.48	14.68	2.20	0.49	-0.22
Stavanger	NORA3	12.09	10.88	13.13	2.26	–	–
	MoE Combined	12.66	11.43	13.68	2.25	0.57	-0.01

The influence of bias correction on the annual maximum significant wave heights is illustrated in Figure 11, which compares the raw NORA3 hindcast with the corresponding MoE-corrected datasets across all study locations. In general, the corrected datasets produced higher annual maxima than the raw hindcast, consistent with the increase observed in the estimated return levels.

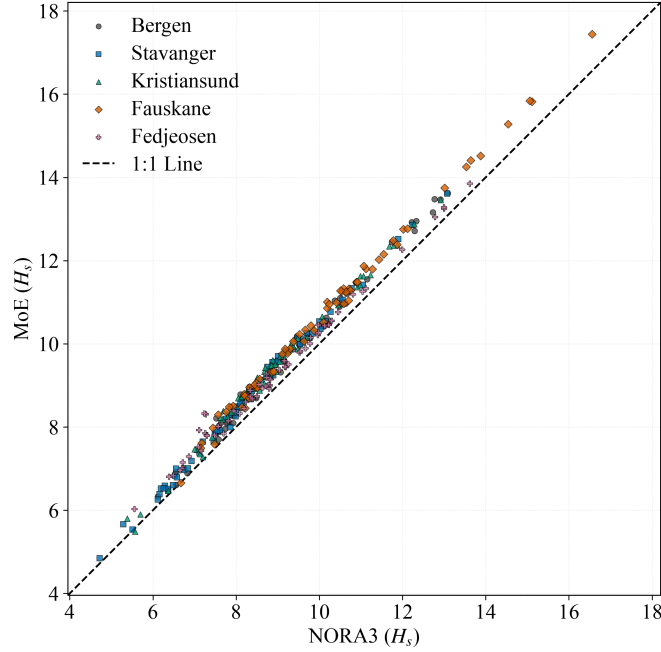


Figure 11. Scatter plots of annual maxima H_s values derived from the raw NORA3 hindcast and MoE-corrected series at all locations.

The corresponding GEV parameters are presented in Table 4. Compared with the raw NORA3 hindcast, the MoE-corrected datasets generally showed higher location parameter (μ) values and, at most locations, an increase in the scale parameter (σ). Although the shape parameter (ξ) generally became more negative after correction, the combined changes in the fitted parameters resulted in consistently higher return level estimates. Overall, these results demonstrate that bias correction influences both the magnitude and the statistical representation of extreme waves, highlighting its importance for reliable coastal EVA.

Table 4. GEV parameter comparison between raw NORA3 and MoE-corrected H_s series.

Location	Source	ξ	σ	μ
Fauskane	NORA3	0.021	1.638	9.127
	MoE Fauskane	-0.026	1.788	9.698
Fedjeosen	NORA3	-0.089	1.466	8.198
	MoE Fedjeosen	-0.089	1.407	8.583
Kristiansund	NORA3	-0.233	1.512	8.374
	MoE Combined	-0.274	1.637	8.868
Bergen	NORA3	-0.128	1.496	8.599
	MoE Combined	-0.175	1.633	9.061
Stavanger	NORA3	-0.152	1.532	7.575
	MoE Combined	-0.174	1.665	7.943

4 Conclusions

This study evaluated statistical methods (SM), machine learning (ML), and mixture-of-experts (MoE) approaches to correct systematic biases in significant wave height of Norwegian reanalysis

hindcasts (NORA3). All methods improved agreement with buoy observations, with the greatest improvements achieved under local calibration. Extreme Gradient Boosting showed the best overall performance among the individual methods, and Directional Adjusted Gumbel Quantile Mapping was the most effective SM. Cross-location validation indicated that transferability depends on similarities in wave climate and bias characteristics between locations. The proposed MoE framework further improved prediction accuracy by adaptively combining SM and ML models, and applying the corrected hindcasts to extreme value analysis produced higher and more reliable return-level estimates than raw NORA3. These findings highlight the importance of bias correction before extreme value analysis and provide a practical framework for improving wave data used in coastal and offshore engineering applications. Future work will examine the framework at additional locations and extend its application to other wave parameters and metocean datasets.

Acknowledgments

The authors acknowledge support by the EU Sustainable Blue Economy Partnership co-funded project FAMOS, www.famos-project.eu (Research Council of Norway project number: 351720).

Code availability

The source code supporting the main analyses presented in this study is publicly available at https://github.com/persimenreed/Norwegian_Coast_Extreme_Waves.

References

- [1] Law Chune S, Aouf L, Levier B and Dalphinet A 2024 *Global High Resolution Production Centre: GLOBAL_MULTIYEAR_WAV_001-032* Copernicus Marine Environment Monitoring Service Europe quality Information Document (QUID) for the global wave reanalysis product. Approved 25 October 2024; published 31 May 2024. URL <https://documentation.marine.copernicus.eu/QUID/CMEMS-GLO-QUID-001-032.pdf>
- [2] Wang W, Pákozdi C, Kamath A and Bihs H 2023 *J. Ocean Eng. Mar. Energy* **9** 567–586
- [3] Haakenstad H, Breivik Ø, Furevik B R, Reistad M, Bohlinger P and Aarnes O J 2021 *J. Appl. Meteorol. Climatol.* **60** 1443–59
- [4] Fanti V, Ferreira Ó, Kümmeler V and Loureiro C 2023 *Commun. Earth Environ.* **4** 151
- [5] Parker K and Hill D F 2017 *Ocean Model.* **110** 52–65
- [6] Hieronymus M and Hieronymus F 2023 *Tellus A: Dyn. Meteorol. Oceanogr.* **75** 129–144
- [7] Breivik Ø, Carrasco A, Haakenstad H, Aarnes O J, Behrens A, Bidlot J R, Björkqvist J V, Bohlinger P, Furevik B R, Staneva J and Reistad M 2022 *J. Geophys. Res.: Oceans* **127** e2021JC018196
- [8] Furevik B R, Aarnes O J and Borg A L Analysis of co-located buoy measurements and wave model results MET Report 01/2019 Norwegian Meteorological Institute Oslo, Norway
- [9] Lemos G, Menendez M, Semedo A, Camus P, Hemer M, Dobrynin M and Miranda P M A 2020 *Glob. Planet. Change* **186** 103109
- [10] Maraun D 2016 *Curr. Clim. Change Rep.* **2** 211–220

- [11] Seeger M 2004 *Int. J. Neural Syst.* **14** 69–106
- [12] Chen T and Guestrin C 2016 XGBoost: A scalable tree boosting system *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* pp 785–794
- [13] Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez A N, Kaiser L and Polosukhin I 2017 Attention is all you need *Advances in Neural Information Processing Systems* vol 30 pp 5998–6008
- [14] Wang J 2023 *Comput. Sci. Eng.* **25** 4–11
- [15] Jacobs R A, Jordan M I, Nowlan S J and Hinton G E 1991 *Neural Comput.* **3** 79–87
- [16] Dietterich T G 2000 Ensemble methods in machine learning *Multiple Classifier Systems (Lecture Notes in Computer Science* vol 1857) (Springer) pp 1–15
- [17] Ré Henriques M, Silva D, Yanchin I, Latas M and Guedes Soares C 2025 *Ocean Eng.* **327** 120925
- [18] Callens A, Morichon D, Abadie S, Delpey M and Liquet B 2020 *Appl. Ocean Res.* **104** 102339
- [19] Reed P S 2026 *Bias Correction Framework for Extreme Wave-Height Estimation on the Norwegian Coast Using NORA3 Hindcast and Buoy Observations* Master's thesis University of Agder Grimstad, Norway
- [20] Coles S 2001 Classical extreme value theory and models *An Introduction to Statistical Modeling of Extreme Values* (London: Springer) pp 45–73