POLITECNICO MILANO 1863

SCUOLA DI INGEGNERIA INDUSTRIALE E DELL'INFORMAZIONE

Executive Summary of the Thesis

COMET-QR: Conformal Multi-fidelity Transformer with Quantile Regression for Antarctic Sea-Ice Concentration Forecasting

Laurea Magistrale in Mathematical Engineering - Ingegneria Matematica

Author: Giorgio Aurina

Advisor: Prof. Andrea Manzoni

Co-advisor: Dr. Paolo Conti

Academic year: 2024-2025

1. Introduction

Antarctic Sea-Ice constitutes a fundamental component of the global climate system, playing a critical role in regulating ocean circulation, surface albedo, and atmosphere–ocean coupling. Despite decades of analysis via satellite observations and climate models, the underlying dynamics governing its variability remain poorly understood. Traditional physical and statistical models have historically struggled to reconcile these interactions, particularly when attempting to integrate heterogeneous datasets characterised by differing spatial and temporal resolutions. To bridge the gap between data-driven inference and physically based modelling, this study proposes a novel, two-phase framework. We combine a physics-informed offline algorithm—Bagging Optimised Dynamic Mode Decomposition—with a deep learning framework based on a Transformer architecture [2]. By leveraging the attention mechanism, this model captures complex spatiotemporal dependencies and cross-modality relationships. This approach allows for the fusion of low- and high-fidelity representations, exploiting complementary information from satellite observations, reanalysis products, and climate model outputs to improve sea-ice forecasting. While preliminary analyses suggest that Transformer-based approaches offer superior flexibility in representing large-scale variability and regional anomalies, challenges regarding causal interpretation and prediction resolution persist. This work specifically highlights the application of COMET-QR (COnformal Multi-fidElity Transformer with Quantile Regression) to address these complex climate problems. A key objective of our research is not only to enhance prediction accuracy but also to produce calibrated confidence intervals. This uncertainty quantification is essential for understanding Antarctic Sea-Ice behaviour, particularly when accounting for extreme events and long-term forecasting reliability.

2. Methodology

Inspired by recent advances as presented in [3], this work introduces a hybrid, data-driven framework for Sea-Ice prediction that leverages an Agnostic Multi-modal Learning paradigm to overcome these challenges. The approach reconceptualises forecasting by decomposing system dynamics into two distinct regimes: a coherent linear structure and a nonlinear stochastic residual.

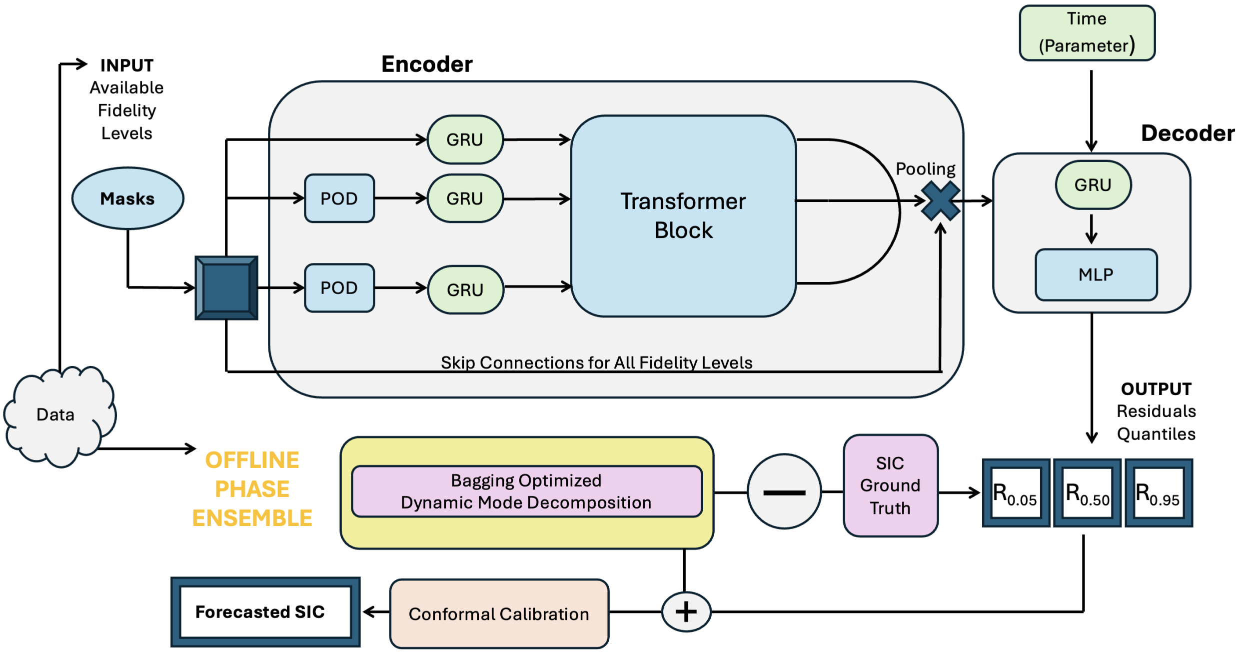


Figure 1: End-to-end inference pipeline of the COMET-QR framework

2.1. Offline Surrogate Model

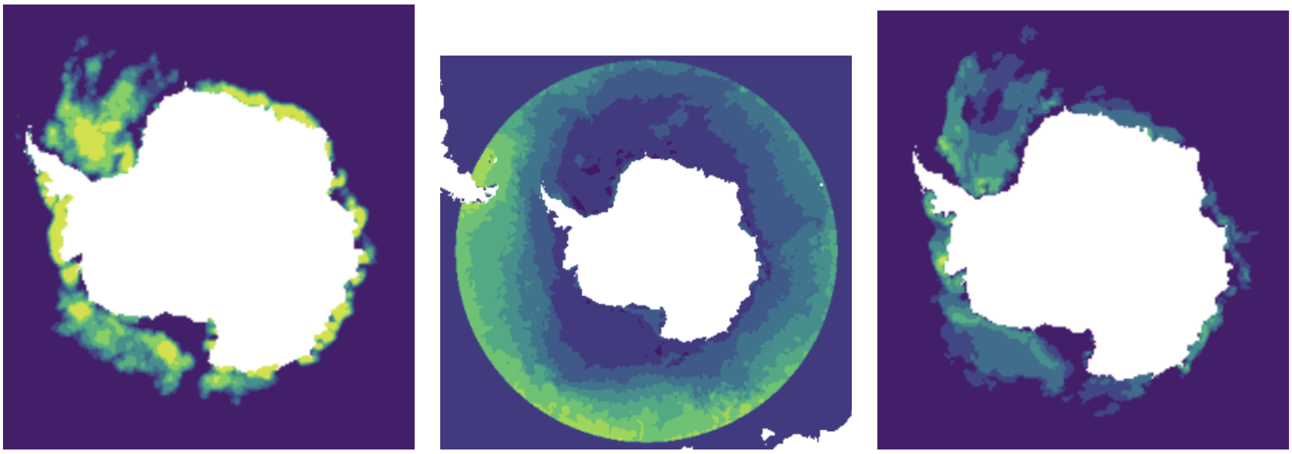
The linear baseline is captured using Bagging Optimised Dynamic Mode Decomposition (BOPDMD) [1], a physics-grounded surrogate model. A key advantage of this spectral approach is the decoupling of computational costs; while extracting modes from historical data is intensive, this task is confined to an offline training phase, allowing for $\mathcal{O}(1)$ online inference. This provides instantaneous, physics-consistent baselines without the need for expensive time-stepping integration. Unlike standard DMD, which yields a single spectral decomposition that is often sensitive to measurement noise and transients, BOPDMD utilises bootstrap aggregating (bagging) to generate a statistical ensemble of models. The system state is reconstructed as a linear superposition of spatiotemporal modes, which are defined by their spatial structures, frequencies, growth or decay rates, and amplitudes. This local approximation acts as the physics-based prior for the subsequent year. The pipeline reshapes spatiotemporal data into snapshots and employs Singular Value Decomposition to extract the dominant linear modes. Specifically, we exploit a bootstrap ensemble where 100 distinct DMD models are trained on resampled data, taking their mean prediction as the baseline forecast. The model is restricted to a rank of 5, isolating the five most dominant coherent structures; stability is enforced by requiring modes to be stable or decaying (non-growing). This ensures physical consistency and guarantees that forecasts remain strictly real-valued. To further capture the latent non-Markovian dynamics, we augment the input space using Time-Delay Embeddings by constructing a matrix representation of the system state that incorporates historical context. Having established the offline BOPDMD baseline, the effective dataset for the residual learning phase spans 31 years (after initialisation). We compute the daily spatial residuals $r_{i,t}$ as the difference between the Sea-Ice Concentration (SIC) ground truth and the DMD forecast, bounded to $[0, 1]$:

$$r_{i,t} = n_{i,t} - y_{i,t}^{\text{DMD}} \quad \forall i, \forall t. \tag{1}$$

The final hybrid model combines the zero-cost inference of the spectral prior with a probabilistic correction from the deep learning module:

$$\hat{n}_t = \underbrace{y_t^{\text{DMD}}}_{\text{DMD Prior}} + \underbrace{f_\theta(x_{1,t}, x_{2,t}, x_{3,t})}_{\text{MF-Transformer Correction}} \tag{2}$$

where f_θ denotes the outcome of a Multi-Fidelity Transformer trained to predict conditional quantiles of the residual, ensuring a calibrated proba-



(a) Sea-Ice Concentration (b) Sea Surface Temperature (c) Sea-Ice Thickness

Figure 2: SIC (Target), SST and SIT (available modalities)

bilistic forecast. To provide the necessary physical context for correction, we define three distinct Low-Fidelity (LF) input modalities, which are extracted from raw observations and processed into standardised temporal sequences. To address high dimensionality, we employ Proper Orthogonal Decomposition (POD) on Sea-Ice Thickness (SIT) and Sea Surface Temperature (SST) (x_1, x_2), thus projecting the dynamics onto a low-dimensional space. Unlike the global POD features, the Sparse Residuals modality (x_3) acts as a real-time error map, obtained by sampling the BOPDMD residual field at random locations. To bridge the gap between data types, each modality x_i is passed through a dedicated Gated Recurrent Unit (GRU) encoder. The GRU processes a temporal lookback sequence of length 365 (with a 30-day frequency) to generate a latent token that encapsulates the historical evolution of that specific source before it enters the Transformer's attention mechanism.

2.2. Online Residual Learning

To address the limitations of DMD in capturing fine spatial features and nonlinearities, the framework employs deep learning architectures, such as GRUs and Transformers, to model the stochastic residuals, effectively correcting biases and enhancing overall predictive accuracy through multi-fidelity data assimilation. We utilise a dedicated GRU for each of the three fidelity levels to create a hidden state acting as a memory embedding before the Transformer block. The core of the framework is based on the standard Transformer block [5], which preserves dimensionality and facilitates gradient flow via residual connections and Layer Normalisation. The attention mechanism contextualises information across the fidelity sequences, and the outputs from the heads H are concatenated and projected back to the original dimension. To handle varying data availability, we employ a masking strategy within the attention mechanism, ensuring that the attention weights for missing sources become exactly zero, allowing the model to dynamically aggregate only valid information without structural changes. For the decoding stage, we adopt a recurrent architecture to model the temporal evolution of forecast residuals. Since the offline BOPDMD baseline already provides the primary spatial structure, the network's role is to track error propagation. Finally, acknowledging the stochastic nature of the target, a probabilistic field of ice presence, we move beyond deterministic point forecasts. We formulate the learning objective as Conformalised Quantile Regression (CQR) [4]. This ensures that our forecasts are not merely accurate in the mean squared sense, but are statistically calibrated, providing rigorous confidence intervals that explicitly account for the uncertainty inherent in both the satellite retrieval process and the chaotic dynamics of the system, as shown in Figure 3.

2.3. Conformal Prediction

We present a theoretically sound conformal prediction framework designed to handle the high-

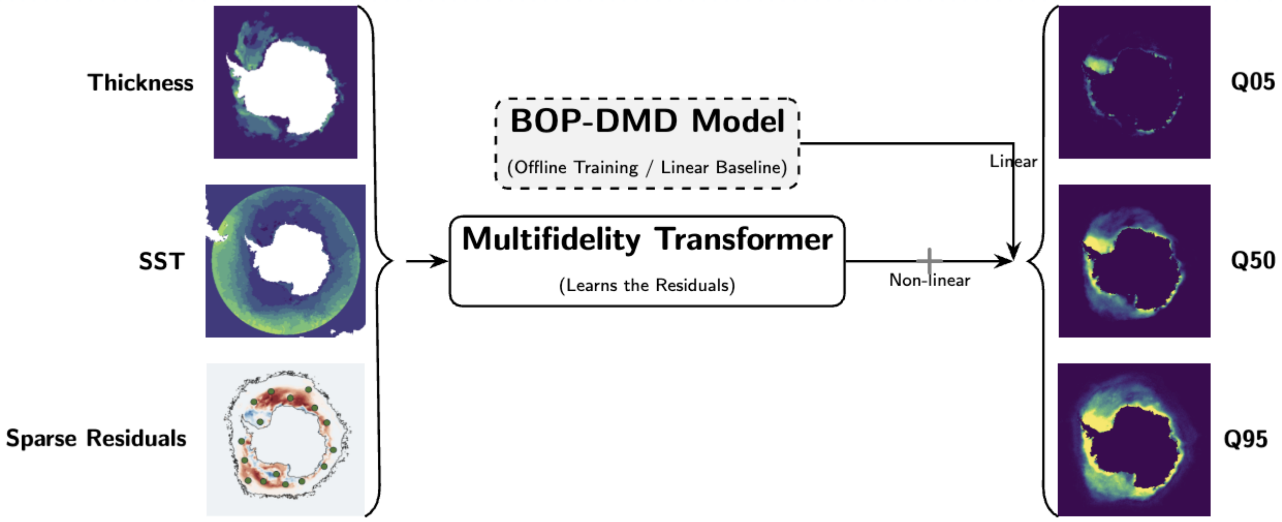


Figure 3: The final output is a probabilistic envelope defined by the 5th, 50th, and 95th SIC quantiles.

dimensional nature of Antarctic Sea-Ice forecasting (p pixels over T daily time steps). A critical insight of this approach is that conformity scores must be computed on the final Sea-Ice Concentration (SIC) predictions rather than on intermediate residuals. The forecasting system builds upon a physics-informed baseline (DMD) by learning to predict the residual error. For each pixel i and time t, the framework defines:

$$\begin{cases} y_{i,t}^{\mathrm{DMD}} \in [0,1], & \text{(DMD baseline)} \\ n_{i,t} \in [0,1], & \text{(True SIC)} \\ r_{i,t} = n_{i,t} - y_{i,t}^{\mathrm{DMD}}, & \text{(True residual)} \\ \hat{r}_{i,t} \in [-1,1]. & \text{(Predicted residual).} \end{cases}$$

To preserve temporal dependencies and prevent information leakage, the 31-year dataset is partitioned into three disjoint temporal blocks. The validation set is reserved strictly for *conformal calibration* to compute conformity scores, ensuring theoretical guarantees for prediction intervals, and is never seen by the neural network during training. The model predicts the 5th, 50th, and 95th percentiles, and the weights are optimised using the Pinball Loss (or Quantile Loss), which asymmetrically penalises errors to enforce the learning of specific quantiles:

$$\mathcal{L}(\theta) = \frac{1}{|\mathcal{T}_{\mathrm{train}}|} \sum_{t,i,q} \rho_q(\tilde{r}_{i,t} - \tilde{r}_{i,t}^{(q)}), \tag{3}$$

where ρ_q is the tilted absolute value function and $\tilde{r}_{i,t}$ is the standardised residual.

While quantile regression provides an estimate of uncertainty, it lacks formal statistical guarantees. To address this, we apply Conformal Prediction using a hold-out validation set to construct calibrated prediction intervals that satisfy $\mathbb{P}(n_{i,t} \in C_{i,t}) \geq 1 - \alpha$. A critical theoretical distinction in our framework is the computation of conformity scores, which quantify the miscoverage of the uncalibrated interval $[\hat{n}_{i,t}^{\mathrm{low}}, \hat{n}_{i,t}^{\mathrm{high}}]$:

$$E_{i,t} = \max\left(\hat{n}_{i,t}^{\mathrm{low}} - n_{i,t},\, n_{i,t} - \hat{n}_{i,t}^{\mathrm{high}}\right) \tag{4}$$

Finally, to account for the spatial and temporal heterogeneity of Antarctic ice dynamics, we employ an adaptive calibration strategy, stratifying the calibration process spatially (per pixel) and temporally (per season). To account for the distinct climatological patterns of the Antarctic, we define four seasons based on Day of Year (DOY). This stratification allows the calibration process to adapt to seasonal predictability; transitional periods like Autumn (freeze-up) and Spring (melt onset), which typically exhibit higher uncertainty, will naturally yield larger Q-scores, thereby widening the prediction intervals to maintain coverage. We compute Q-scores by collecting all conformity scores $E_{i,t}$ from validation samples for each unique combination of pixel i and season s. Once calibrated, prediction intervals for new test points are constructed by adjusting the model's raw predicted quantiles (5th and 95th percentiles).

Scenario	MAE Mod.	Imp. DMD	Imp. Clim	Cov.	Width
3L_all	0.0746	21.95%	7.63%	0.895	0.248
2L_noSensors	0.0797	16.55%	1.24%	0.870	0.242
2L_noThickness	0.0751	21.42%	7.00%	0.894	0.248
2L_noSST	0.0750	21.52%	7.12%	0.900	0.251
1L_Sensors	0.0794	16.84%	1.58%	0.886	0.253
1L_SST	0.0822	13.92%	-1.87%	0.859	0.241
1L_Thickness	0.0818	14.36%	-1.36%	0.874	0.246

Table 1: COMET-QR performance comparison across different scenarios

3. Results

We summarise in this section our results, mainly shown in Table 1. In particular:

1. The model predicts residuals, which are denormalised and added to the DMD baseline.
2. These "uncalibrated" SIC predictions are clipped to the physical range $[0, 1]$.
3. The Q-score is applied to expand these bounds, ensuring that the true value falls within the interval with a probability of at least $1 - \alpha$.

The final test evaluation confirms the superiority of the multi-fidelity fusion approach over traditional baselines. The complete 3-level configuration—incorporating SST (91 modes), SIT (95 modes), and 80 Sparse Sensors—achieved the best overall performance with a Mean Absolute Error (MAE) of 0.0746. This translates to a 21.95% improvement over the surrogate BOPDMD model (MAE 0.0955) and a 7.63% improvement over the static Climatology baseline (MAE 0.0807). All the numerical results are reported in Table 1, while in Figure 8 we show the predictions of our model over 4 distinct days of the test years, along with calibrated confidence intervals. The empirical results drawn from a rigorous 31-year temporal split demonstrate the superiority of our approach. Key findings of this research include:

- The Power of Multi-Fidelity Fusion: Relying on single-source data is insufficient for complex climate modelling. By tokenising Sea Surface Temperature (SST), Sea-Ice Thickness (SIT), and Sparse Sensors into a shared latent space, the Transformer's attention mechanism successfully learned the cross-correlations between oceanic forcing and ice mechanics. The full multi-fidelity model reduced the Mean Absolute Error by nearly 22% compared to the purely physical DMD surrogate.
- The Necessity of Localised Error Proxies: Our ablation studies revealed that global thermodynamic features alone cannot correct localised prediction drift. The inclusion of sparse point-wise residuals proved mandatory to guide the network in correcting the high-frequency errors of the baseline model.
- Robust Uncertainty Quantification: By combining Quantile Regression with Conformal Prediction, the framework transitioned from deterministic point estimates to statistically calibrated probabilistic bounds. Applying the conformal correction directly in the clipped physical space ($\text{SIC} \in [0, 1]$) guaranteed a valid 90% coverage rate. The resulting uncertainty maps proved physically interpretable, correctly identifying the Marginal Ice Zone and transitional seasons as the primary sources of predictive instability.

4. Conclusions

While COMET-QR represents a significant step forward in operational Sea-Ice forecasting, it also opens several avenues for future research. From an architectural standpoint, scaling the Transformer blocks and increasing the number of attention heads could further enhance the model's capability to resolve finer spatial heterogeneities. From a physical perspective, includ-

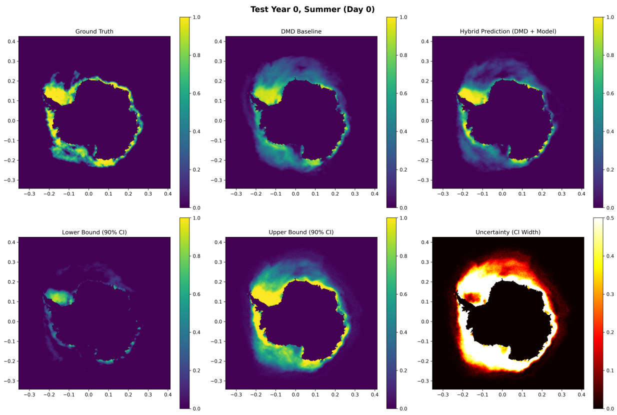


Figure 4: Year 1: Summer (Day 1)

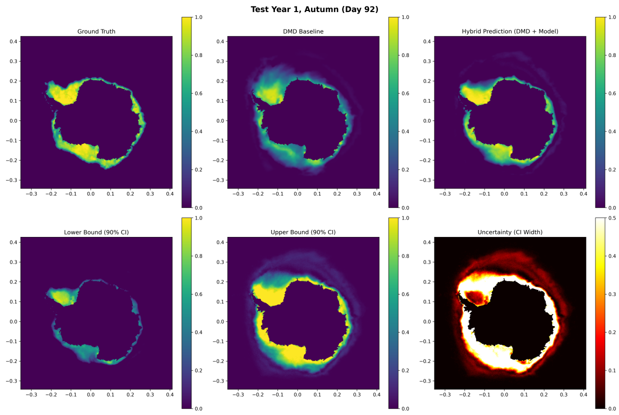


Figure 5: Year 2: Autumn (Day 93)

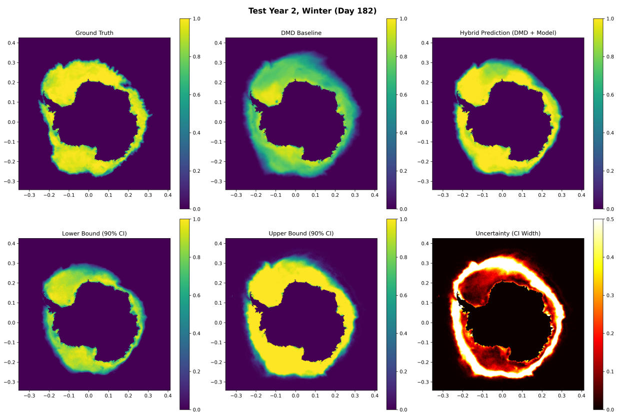


Figure 6: Year 3: Winter (Day 183)

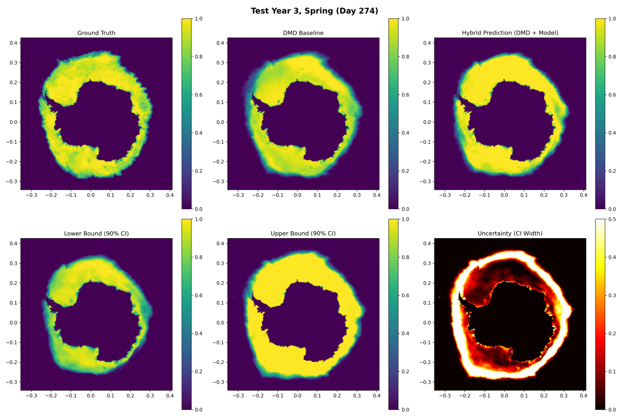


Figure 7: Year 4: Spring (Day 275)

Figure 8: Seasonal predictions showing the evolution of Sea-Ice Concentration over different years.

ing other atmospheric modalities, such as surface wind stress vectors, could improve results even more.
Ultimately, this study demonstrates that the future of climate forecasting does not lie in choosing between theoretical physics and machine learning, but rather in their systematic integration. By embedding physical constraints within the loss landscape and treating diverse observational datasets as interconnected tokens of a unified environment, COMET-QR provides a scalable, risk-aware framework capable of navigating the volatility of the Antarctic cryosphere.

References

[1] J. Nathan Kutz, Steven L. Brunton, Bingni W. Brunton, and Joshua L. Proctor. *Dynamic Mode Decomposition: Data-Driven Modeling of Complex Systems.* SIAM, Philadelphia, PA, 2016.

[2] Marco Da Mommio. MELTER: a transformer architecture for multi-fidelity problems. Master thesis, Politecnico di Milano, 2024.

[3] Liam Parker, Francois Lanusse, Jeff Shen, Ollie Liu, Tom Hehir, Leopoldo Sarra, et al. AION-1: Omnimodal foundation model for astronomical sciences. *arXiv preprint*, 2024.

[4] Yaniv Romano, Evan Patterson, and Emmanuel J. Candès. Conformalized quantile regression. In *Advances in Neural Information Processing Systems*, volume 32, pages 3538–3548. Curran Associates, Inc., 2019.

[5] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. In *Advances in neural information processing systems*, pages 5998–6008, 2017.