

# Beyond MIRKWOOD: Enhancing SED Modeling with Conformal Predictions

Sankalp Gilda

Machine Learning Collective  
sankalp.gilda@gmail.com

## Abstract

Traditional spectral energy distribution (SED) fitting techniques face uncertainties due to assumptions in star formation histories and dust attenuation curves. We propose an advanced machine learning-based approach that enhances flexibility and uncertainty quantification in SED fitting. Unlike the fixed NGBOOST model used in MIRKWOOD, our approach allows for any SKLEARN-compatible model, including deterministic models. We incorporate conformalized quantile regression to convert point predictions into error bars, enhancing interpretability and reliability. Using CATBOOST as the base predictor, we compare results with and without conformal prediction, demonstrating improved performance using metrics such as coverage and interval width. Our method offers a more versatile and accurate tool for deriving galaxy physical properties from observational data.

## Introduction

Spectral energy distributions (SEDs) are pivotal in astrophysics for understanding the intrinsic properties of galaxies, such as stellar mass, age distributions, star formation rates, and dust content. Traditional SED fitting methods, while insightful, often face significant challenges. These challenges stem from the complex nature of galaxies, including diverse star formation histories and varying dust attenuation curves (???). The inherent uncertainties in these aspects can significantly affect the accuracy of derived galaxy properties, thus impacting our broader understanding of galactic evolution and formation.

Recent advancements in computational methods have opened new avenues in this field. Machine learning (ML), with its ability to handle large datasets and uncover complex patterns, has emerged as a powerful tool in SED fitting (???). The traditional parametric and often linear approaches are being supplemented, and in some cases replaced, by non-parametric, highly flexible ML techniques that can model the non-linear relationships intrinsic to astronomical data more effectively (????). This paradigm shift is not just a matter of computational convenience but represents a fundamental change in how we interpret vast and complex astronomical datasets.

This paper introduces an innovative approach that builds upon and significantly expands the capabilities of the MIRKWOOD (?), a machine learning-based application previously developed for SED fitting. Our method enhances the flexibility and depth of analysis by enabling the use of any sklearn-compatible model (?). This includes not only probabilistic models but also deterministic ones, thereby broadening the scope of application to a wider range of astronomical problems. Moreover, we integrate the uncertainty quantification method technique of conformalized quantile regression (CQR) (?), which allows us to translate point predictions into meaningful error bars. This addition is crucial in fields like astronomy, where quantifying the uncertainty of predictions is as important as the predictions themselves. The combination of these advanced techniques positions our tool at the forefront of SED fitting technologies, offering a more nuanced and comprehensive understanding of galaxy properties.

In the context of SED fitting, the ability to quantify uncertainty is essential for several reasons. First, it enables astronomers to distinguish between variations in galactic properties that are due to inherent physical processes versus those arising from observational limitations. Secondly, in fields such as cosmology, where the accurate determination of galaxy properties impacts our understanding of the universe's evolution, refined uncertainty quantification offers a way to assess the reliability of these large-scale inferences. Thus, enhancing the precision of uncertainty quantification in SED fitting directly contributes to our fundamental understanding of the universe.

## Background

The incorporation of machine learning in SED fitting represents a significant paradigm shift from traditional methods. Traditional SED fitting often relies on fitting parametric models to observational data, a process that can be computationally intensive and limited by the assumptions inherent in the models used (?). Machine learning, particularly algorithms like CatBoost (?), offers an alternative that can handle large, complex datasets with greater efficiency and flexibility.

CatBoost is an advanced implementation of gradient boosting, a machine learning technique that builds predictive models in a stage-wise fashion. It is particularly adept at handling categorical data, a common feature in astronomical datasets. This makes it an ideal choice for our purposes, as it can

efficiently process the multifaceted data involved in SED fitting, including various photometric bands and derived galaxy properties.

Conformalized quantile regression, a relatively recent development in the field of machine learning, adds another layer of sophistication to our approach. This technique allows us to quantify the uncertainty of our predictions in a robust and interpretable manner. It achieves this by transforming point predictions from our models into prediction intervals with a specified level of confidence. This flavor of conformal inference (?) is particularly valuable in astronomy, where the ability to quantify uncertainty is critical for making reliable inferences about astronomical objects.

The combination of any sklearn-compatible deterministic modelling pipeline and CQR – a state-of-the-art conformal prediction methodology – represents a significant advancement in the field of SED fitting. By leveraging these techniques, we can move beyond the limitations of traditional methods, offering a more nuanced understanding of galaxy properties. This approach not only yields more accurate predictions but also provides insights into the reliability of these predictions, a crucial aspect in the field of astronomy where data is often sparse and noisy.

## Data

Our training and testing datasets are derived from three advanced cosmological galaxy formation simulations, known for their accurate representation of galaxy physical properties, including authentic star formation histories. These simulations SIMBA (?), EAGLE (???), and ILLUSTRISTNG (?) provide a comprehensive and realistic variety of galaxy evolution scenarios, with sample sizes of 1,688, 4,697, and 9,633 respectively. We focus on galaxies at redshift 0, representing them in their current state in the simulations. The spectral energy distributions (SEDs) in our datasets consist of 35 flux density measurements (in Jansky units) across different wavelengths, representing the luminosity of galaxies. These SEDs serve as the input features for our model. The target outputs or labels are the four scalar galaxy properties – galaxy mass, metallicity, dust mass, and star formation rate. See Table 1 in ? for an overview of the distribution of these properties for all three simulations.

## Methodology

### Data Preprocessing

The foundation of any robust machine learning model is high-quality data. In our approach, we begin with a thorough data preprocessing phase. This involves cleaning the data, handling missing values, normalizing photometric fluxes, and encoding categorical variables where necessary. The preprocessing steps are critical in ensuring that the input data fed into the machine learning models is consistent, standardized, and reflective of the underlying physical phenomena we aim to model.

We manually add Gaussian noise to the SEDs from the 3 simulations, to get three separate sets of data at signal-to-noise (SNR) ratios of 20, 10, and 5. We do this for a 1:1 comparison with the methodology and results of ?.

## Model Selection and Flexibility

Our methodology is characterized by its flexibility and adaptability in model selection. While the mirwood tool was initially designed around NGBoost, we expand its capabilities by enabling the use of any sklearn-compatible model. This includes, but is not limited to, deterministic models such as Support Vector Machines (?) and Random Forests (?), alongside probabilistic models like Gaussian Processes (?). In fact, our code is flexible enough to allow a pipeline consisting of an arbitrary number of sklearn-compatible models. This flexibility allows astronomers to tailor the predictive models to their specific research needs and the characteristics of their datasets.

### CatBoost as the Base Predictor

CatBoost, our chosen base predictor, is particularly well-suited for dealing with the types of datasets common in astronomical research. It efficiently handles categorical features and large datasets, reducing overfitting and improving predictive accuracy. In our implementation, we fine-tune CatBoost parameters, such as the depth of trees and learning rate, to optimize its performance for SED fitting tasks.

### Incorporating Conformalized Quantile Regression

A significant enhancement in our methodology is the incorporation of conformalized quantile regression. This technique allows us to convert the point predictions from our models into prediction intervals. These intervals provide a statistical measure of the uncertainty in our predictions, giving us a range within which the true value of the predicted property is likely to fall, at a given confidence level. Implementing this technique involves calibrating our models to estimate the quantiles of the predictive distribution, a crucial step in providing reliable and interpretable error estimates. Since (?) predict  $1\sigma$  error bars, for apples-to-apples comparison we set the significance level  $\alpha$  at 0.318.

### Training and Validation

The final phase of our methodology involves training the machine learning models on a carefully curated dataset and validating their performance. For apples-to-apples comparison with (?), the training set contains all 10,073 samples from ILLUSTRISTNG, 4,697 samples from EAGLE, and 359 samples from SIMBA selected via stratified 5-fold CV (see Section 3 in ? for details). After making inference on all test splits, we collate the results, thus successfully predicting all four galaxy properties for all 1,797 samples from SIMBA. Each predicted output for a physical property contains two values – the mean and the standard deviation.

In the fitting process, we first train the model using galaxy flux values to predict stellar mass. Then, we use the predicted stellar masses, combined with the original flux values, to predict dust mass, and continue this sequential prediction process for other parameters. See Figure 3 and Section 3 in ? for details.

Through this comprehensive methodology, we aim to provide a powerful, flexible, and accurate tool for SED fitting, capable of handling the complexities and uncertainties inherent in astronomical datasets.

|             | Model      | NRMSE ( $\downarrow$ ) | NMAE ( $\downarrow$ ) | NBE ( $\downarrow$ ) | ACE ( $\downarrow$ ) | IS ( $\downarrow$ ) |
|-------------|------------|------------------------|-----------------------|----------------------|----------------------|---------------------|
| Mass        | This paper | <b>0.009</b>           | <b>0.074</b>          | <b>-0.031</b>        | <b>-0.051</b>        | <b>0.001</b>        |
|             | MIRKWOOD   | 0.155                  | 0.115                 | -0.041               | -0.066               | 0.001               |
|             | PROSPECTOR | 1.002                  | 1.117                 | -0.479               | -0.482               | 0.033               |
| Dust Mass   | This paper | <b>0.412</b>           | <b>0.298</b>          | <b>-0.157</b>        | <b>-0.041</b>        | <b>0.001</b>        |
|             | MIRKWOOD   | 0.475                  | 0.336                 | -0.215               | -0.076               | 0.001               |
|             | PROSPECTOR | 1.263                  | 1.212                 | -0.679               | nan                  | nan                 |
| Metallicity | This paper | <b>0.044</b>           | <b>0.048</b>          | <b>-0.009</b>        | <b>-0.053</b>        | <b>0.016</b>        |
|             | MIRKWOOD   | 0.056                  | 0.052                 | -0.010               | -0.063               | 0.032               |
|             | PROSPECTOR | 0.547                  | 0.487                 | -0.229               | 0.036                | 0.302               |
| SFR         | This paper | <b>0.223</b>           | <b>0.147</b>          | <b>-0.047</b>        | <b>0.014</b>         | <b>0.004</b>        |
|             | MIRKWOOD   | 0.277                  | 0.215                 | -0.078               | 0.035                | 0.006               |
|             | PROSPECTOR | 1.988                  | 2.911                 | 1.437                | -0.547               | 0.200               |

Table 1: Comparative performance of our proposed method v/s MIRKWOOD v/s PROSPECTOR across different metrics. The five metrics are the normalized root mean squared error (NRMSE), normalized mean absolute error (NMAE), normalized bias error (NBE), average coverage error (ACE), and interval sharpness (IS). A bold value denotes the best metric for that galaxy property. A value of ‘nan’ represents lack of predictions from PROSPECTOR. We do not have predicted error bars from PROSPECTOR for dust mass, hence ACE and IS values corresponding to this property are ‘nan’s.

|             | Model      | NRMSE ( $\downarrow$ ) | NMAE ( $\downarrow$ ) | NBE ( $\downarrow$ ) | ACE ( $\downarrow$ ) | IS ( $\downarrow$ ) |
|-------------|------------|------------------------|-----------------------|----------------------|----------------------|---------------------|
| Mass        | This paper | <b>0.092</b>           | <b>0.071</b>          | <b>-0.026</b>        | <b>-0.018</b>        | <b>0.001</b>        |
|             | MIRKWOOD   | 0.165                  | 0.118                 | -0.035               | -0.021               | 0.001               |
|             | PROSPECTOR | 1.000                  | 1.088                 | -0.518               | -0.502               | 0.004               |
| Dust Mass   | This paper | <b>0.391</b>           | <b>0.254</b>          | <b>-0.143</b>        | <b>0.012</b>         | <b>0.001</b>        |
|             | MIRKWOOD   | 0.456                  | 0.332                 | -0.209               | -0.033               | 0.001               |
|             | PROSPECTOR | 0.996                  | 0.998                 | -0.905               | nan                  | nan                 |
| Metallicity | This paper | <b>0.037</b>           | <b>0.049</b>          | <b>0.007</b>         | <b>0.021</b>         | <b>0.023</b>        |
|             | MIRKWOOD   | 0.058                  | 0.055                 | -0.010               | -0.032               | 0.036               |
|             | PROSPECTOR | 0.534                  | 0.464                 | -0.275               | -0.041               | 0.295               |
| SFR         | This paper | <b>0.274</b>           | <b>0.114</b>          | <b>-0.070</b>        | <b>0.027</b>         | <b>0.001</b>        |
|             | MIRKWOOD   | 0.329                  | 0.226                 | -0.090               | 0.048                | 0.001               |
|             | PROSPECTOR | 0.910                  | 0.992                 | -0.686               | -0.564               | 1.937               |

Table 2: Same as Table 1, but for SNR=10.

|             | Model      | NRMSE ( $\downarrow$ ) | NMAE ( $\downarrow$ ) | NBE ( $\downarrow$ ) | ACE ( $\downarrow$ ) | IS ( $\downarrow$ ) |
|-------------|------------|------------------------|-----------------------|----------------------|----------------------|---------------------|
| Mass        | This paper | <b>0.121</b>           | <b>0.062</b>          | <b>-0.031</b>        | <b>-0.001</b>        | <b>0.001</b>        |
|             | MIRKWOOD   | 0.198                  | 0.123                 | -0.042               | -0.002               | 0.001               |
|             | PROSPECTOR | 1.003                  | 1.091                 | -0.528               | -0.497               | 0.005               |
| Dust Mass   | This paper | <b>0.315</b>           | <b>0.224</b>          | <b>-0.154</b>        | <b>0.002</b>         | <b>0.001</b>        |
|             | MIRKWOOD   | 0.480                  | 0.339                 | -0.219               | 0.003                | 0.001               |
|             | PROSPECTOR | 0.996                  | 0.998                 | -0.905               | nan                  | nan                 |
| Metallicity | This paper | <b>0.049</b>           | <b>0.048</b>          | <b>-0.005</b>        | <b>-0.013</b>        | <b>0.034</b>        |
|             | MIRKWOOD   | 0.062                  | 0.060                 | -0.011               | -0.024               | 0.041               |
|             | PROSPECTOR | 0.544                  | 0.478                 | -0.297               | 0.046                | 0.301               |
| SFR         | This paper | <b>0.189</b>           | <b>0.171</b>          | <b>-0.043</b>        | <b>0.061</b>         | <b>0.001</b>        |
|             | MIRKWOOD   | 0.241                  | 0.205                 | -0.069               | 0.074                | 0.001               |
|             | PROSPECTOR | 0.907                  | 0.99                  | -0.687               | -0.557               | 7.314               |

Table 3: Same as Table 1, but for SNR=5.

## Comparative Analysis and Results

### Comparative Analysis Methodology

To demonstrate the efficacy of our approach, we conducted a comprehensive comparative analysis. This involved comparing the performance of our enhanced tool against a traditional SED fitting method tool, PROSPECTOR (?) and the original MIRKWOOD implementation. We focused on the same five performance metrics as in ? to evaluate the accuracy of derived galaxy properties (galactic mass, dust mass, star formation rate, and metallicity), and the robustness of the model against variations in input data.

### Performance Metrics

We use both deterministic and probabilistic metrics for comparison, the same five metrics used in ? – normalized root mean squared error (NRMSE), normalized mean absolute error (NMAE), normalized bias error (NBE), average coverage error (ACE), and interval sharpness (IS). These are defined and described in detail in Section 3.2 of ?. In particular, coverage is the proportion of true values that fall within the predicted error bars, offering a measure of the reliability of our uncertainty quantification. On the other hand, IW is the average width of the prediction intervals, which provides insight into the precision of our predictions.

### Results

To evaluate our proposed model for SED fitting, we conduct comparisons with fits obtained in ? from the Bayesian SED fitting software PROSPECTOR, and their new machine learning tool MIRKWOOD. We provide each of the three models (their two plus our upgraded version of MIRKWOOD) with identical data to deduce galaxy properties. This data comprises broadband photometry across 35 bands, subject to Gaussian uncertainties of 5%, 10%, and 20% (corresponding to signal-to-noise ratios (SNRs) of 20, 10, and 5, respectively). In Tables 1, 2, and 3 we showcase the outcomes from all three methods for all four galaxy properties. The results of our comparative analysis are illuminating and encouraging. Our method consistently achieves higher coverage rates compared to both the other methods, indicating more reliable uncertainty quantification. At the same time, the prediction intervals generated by our method were narrower on average, signifying more precise predictions.

These results underline the superiority of our approach in terms of both accuracy and reliability in SED fitting. By leveraging the power of CatBoost and the precision of conformalized quantile regression, our method not only enhances the accuracy of point predictions but also provides a more nuanced understanding of the associated uncertainties.

### Discussion

The improvements observed in our analysis can be attributed to several factors. The flexibility in model selection allows for better adaptation to the specific characteristics of astronomical datasets. CatBoost’s superior ability to work with tabular data effectively captures the complexities in the data, leading to more accurate predictions. The addition of conformalized quantile regression introduces a robust method

for uncertainty quantification, a critical aspect often overlooked in traditional SED fitting. Overall, the comparative analysis and the results obtained highlight the potential of our method in transforming the field of SED fitting, providing astronomers with a tool that is not only accurate but also comprehensive in its assessment of uncertainties.

### Conclusions and Future Work

This study marks a substantial advancement in the field of spectral energy distribution (SED) fitting by integrating flexible machine learning models, particularly CatBoost, with the innovative technique of conformalized quantile regression. This approach not only enhances the accuracy of SED fitting but also introduces a new depth to the uncertainty quantification in astronomical research. The adaptability of our tool to various astronomical datasets, coupled with the ability to select from a range of sklearn-compatible models, ensures its applicability across different research contexts. CatBoost’s effectiveness in handling complex datasets, combined with our sophisticated method of uncertainty quantification, allows for more reliable and nuanced interpretations of galactic properties. Our comparative analysis highlights the superiority of this method over traditional approaches, demonstrating improvements in both the accuracy of predictions and the understanding of associated uncertainties. This dual capability represents a significant stride in astronomy, offering a more reliable and comprehensive tool for exploring the universe.

Looking ahead, the potential for further advancements and extensions of our tool is vast. Future work may involve exploring the integration of additional machine learning models, such as deep learning architectures, to enhance predictive power and versatility. Testing and optimizing the tool on larger and more diverse datasets from upcoming astronomical surveys will be crucial for assessing its scalability and robustness. Further development in feature engineering and expanding the scope of uncertainty quantification could unlock new insights and details in SED fitting. Additionally, applying this tool to related fields like exoplanet studies or cosmic structure formation could demonstrate its adaptability and contribute to a broader range of scientific inquiries. *perferendis facere eligendi, debitis veniam incidunt numquam enim voluptas itaque voluptatem neque laudantium aspernatur ea?Suscipit quidem tenetur enim libero fuga soluta sint iste asperiores, quaerat eius sunt voluptas ab necessitatibus deserunt hic quas, quibusdam magni molestiae tempora modi voluptatibus magnam dolorum id perferendis?Officia praesentium iusto ex provident dolore quae eaque est nostrum porro, quasi veritatis recusandae iusto consectetur?Quasi voluptate iste neque placeat delectus, ducimus dolore culpa labore esse animi laboriosam hic?Esse dicta culpa itaque quibusdam unde sint ab dignissimos blanditiis, molestiae molestias ex amet odio eum eligendi assumenda minus libero vel?Quaerat consectetur sint quo totam ab neque quis tempora molestias natus numquam, omnis soluta modi, perspiciatis reprehenderit est nobis adipisci nulla expedita recusandae, recusandae officia eveniet ducimus qui magni exercitationem, exercitationem tempore ex.Ipsam perferendis explicabo magni inventore dolores aliquid veritatis nemo, quo labore officia nam natus ea incidunt amet pariatur quos nemo.*