Rosario Cecilio-Flores-Elie
Star & Planet Formation - HW 4
Literature Summary

Mapping circumgalactic medium observations to theory using machine learning

The study by Appleby et al. utilizes a machine learning algorithm called Random Forest (RF) to predict the physical conditions of the circumgalactic medium (CGM) from quasar absorption line observables using the SIMBA simulation. Unlike traditional methods, the RF model captures the complex relationships between observables and gas conditions without relying on simplifications. It predicts CGM properties such as overdensity, temperature, and metallicity from synthetic spectra, demonstrating the impact of various features like column density, line width, and star formation rates on these predictions.

The results show that the model can capture the nuanced behaviors of CGM properties by offering insights into the importance of different features in predicting these conditions. This methodology and findings highlight the potential of machine learning in astrophysical research, particularly in studying the CGM. The SIMBA simulations, an evolution of MUFASA simulations, incorporate advanced models for black hole growth and feedback and dust evolution within a simulated universe. The study uses a specific subset of galaxies from SIMBA, selected to cover a range of star formation rates and masses, to investigate the CGM through synthetic absorption spectra. These spectra, generated for various ions, help to analyze the physical conditions of the CGM, utilizing the PYGAD package for detailed analysis. This comprehensive approach allows for associating specific physical properties with absorption features, aiming for a more nuanced understanding of the CGM's complexity.

Overall, the study successfully employs Random Forest (RF) models to link CGM absorption observations with the underlying gas conditions, favoring RF for its simplicity and interpretability. It reveals reasonable predictions for H I and metal ions, noting a general tendency for RF models to predict a narrower distribution than the input. The research highlights the importance of specific features in making predictions and suggests the potential of using RF models with observational CGM data. There are plans to enhance the precision of the model and broaden the technique to other simulations to evaluate the influence of galaxy formation models in the future.

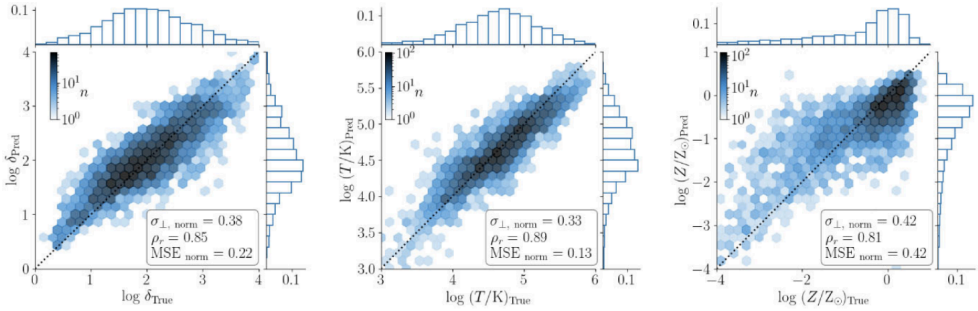

Figure 1. Hexagonal joint histogram of the predicted H I physical conditions from the RF mapping and the true H I physical conditions, including only data in the test set. The number of data points in each bin is shown using colourbars. From left to right, the panels show overdensity, temperature, and metallicity. The diagonal line represents the case where the RF model makes a perfect prediction. The accuracy of the predictions in each panel is summarized by the inset displaying the normalized transverse scatter $\sigma_{\perp,\rm norm}$, the correlation coefficient ρ_r, and the normalized mean square error, $\rm MSE_{norm}$. The 1D histograms of the true and predicted values are shown on the top and side of each panel, respectively.

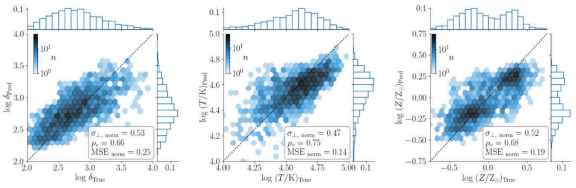

Figure 2. As in Fig. 1, showing the predictions and true values for C II absorbers.

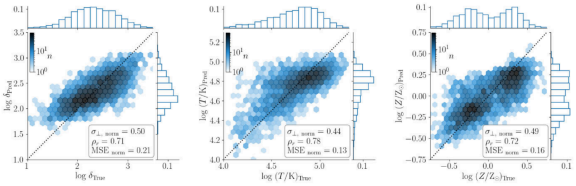

Figure 3. As in Fig. 1, showing the predictions and true values for C IV absorbers.

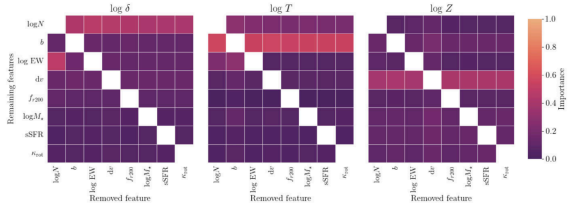

Figure 4. Importance values in predicting H I physical conditions for each remaining input feature, against the input feature removed from the training data. From left to right, the target predictors are δ, T, and Z.

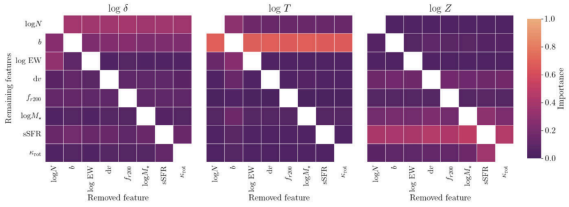

Figure 5. Feature importance values as in Fig. 4, for C II absorbers.

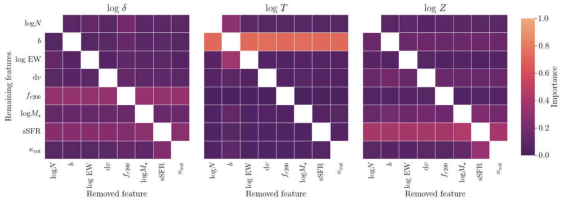

Figure 6. Feature importance values as in Fig. 4, for C IV absorbers.

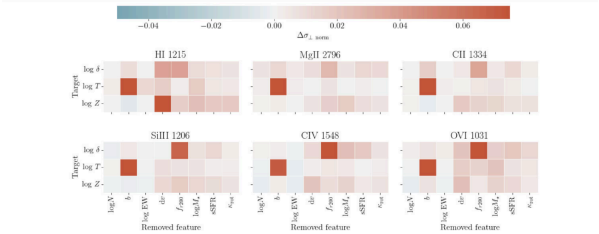

Figure 7. The change in $\sigma_{\perp,\rm norm}$ of the RF models when removing each feature iteratively. Each of the panels shows results for a different ion; the three rows of each panel represent results for each of δ, T, and Z.

Sarah Appleby et al., "Mapping Circumgalactic Medium Observations to Theory Using Machine Learning," *Monthly Notices of the Royal Astronomical Society* 525, no. 1 (August 9, 2023): 1167–81, https://doi.org/10.1093/mnras/stad2266.