

# Identifying Galaxy Properties That Govern Gas-Phase Metallicity in IllustrisTNG with Machine Learning

Authors: S. Savić, M. Pleansri, M. Mićić, A. Mitrašinović  
Affiliation: Astronomical Observatory Belgrade

## INTRODUCTION

Gas-phase metallicity is regulated by different processes in galaxies, including supernova feedback, AGN activity, and the inflow of pristine gas.

Different galaxy parameters can trace these processes.

Here, we examine which parameters are most important for predicting gas-phase metallicity using different machine learning models.



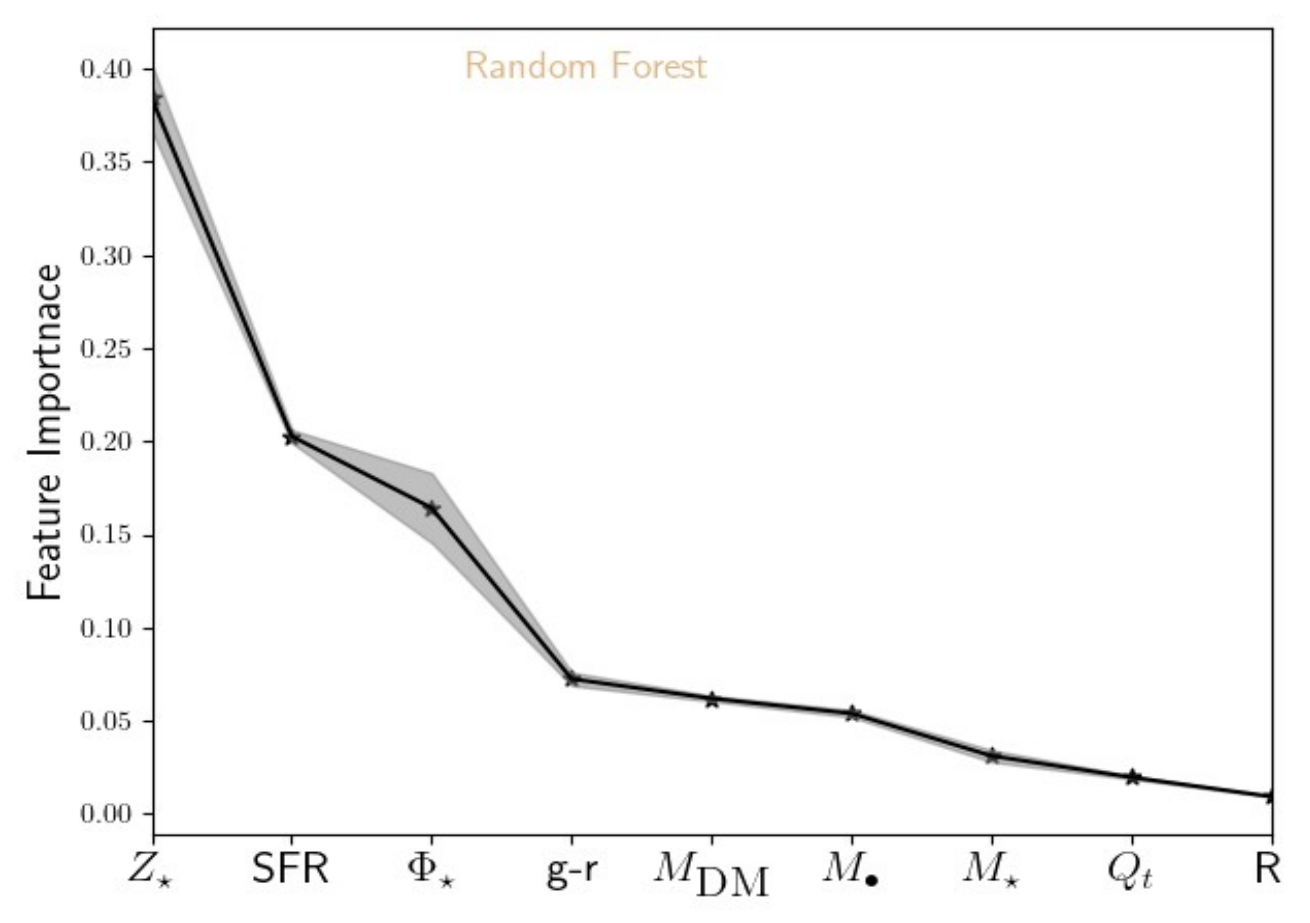
## METHOD

We use results from the large-scale cosmological magnetohydrodynamical simulation of galaxy formation IllustrisTNG300.

Our sample consists of 32510 galaxies at  $z=0$ . For this preliminary study, we select galaxy properties that have been investigated in observational studies as potential drivers of gas-phase metallicity: stellar metallicity ( $Z_*$ ), stellar mass ( $M_*$ ), supermassive black hole mass ( $M_\bullet$ ), dark matter mass ( $M_{DM}$ ), star formation rate (SFR), color ( $g-r$ ), stellar potential ( $\Phi_*$ ) and proxy for overdensity of galaxies ( $Q_t$ ).

We apply machine learning methods: Random Forest (RF), Principal Component Analysis (PCA) and Gradient Boosting (GB) to investigate the relative importance of these properties in determining gas-phase metallicity and compare the results obtained with different methods.

## RESULTS

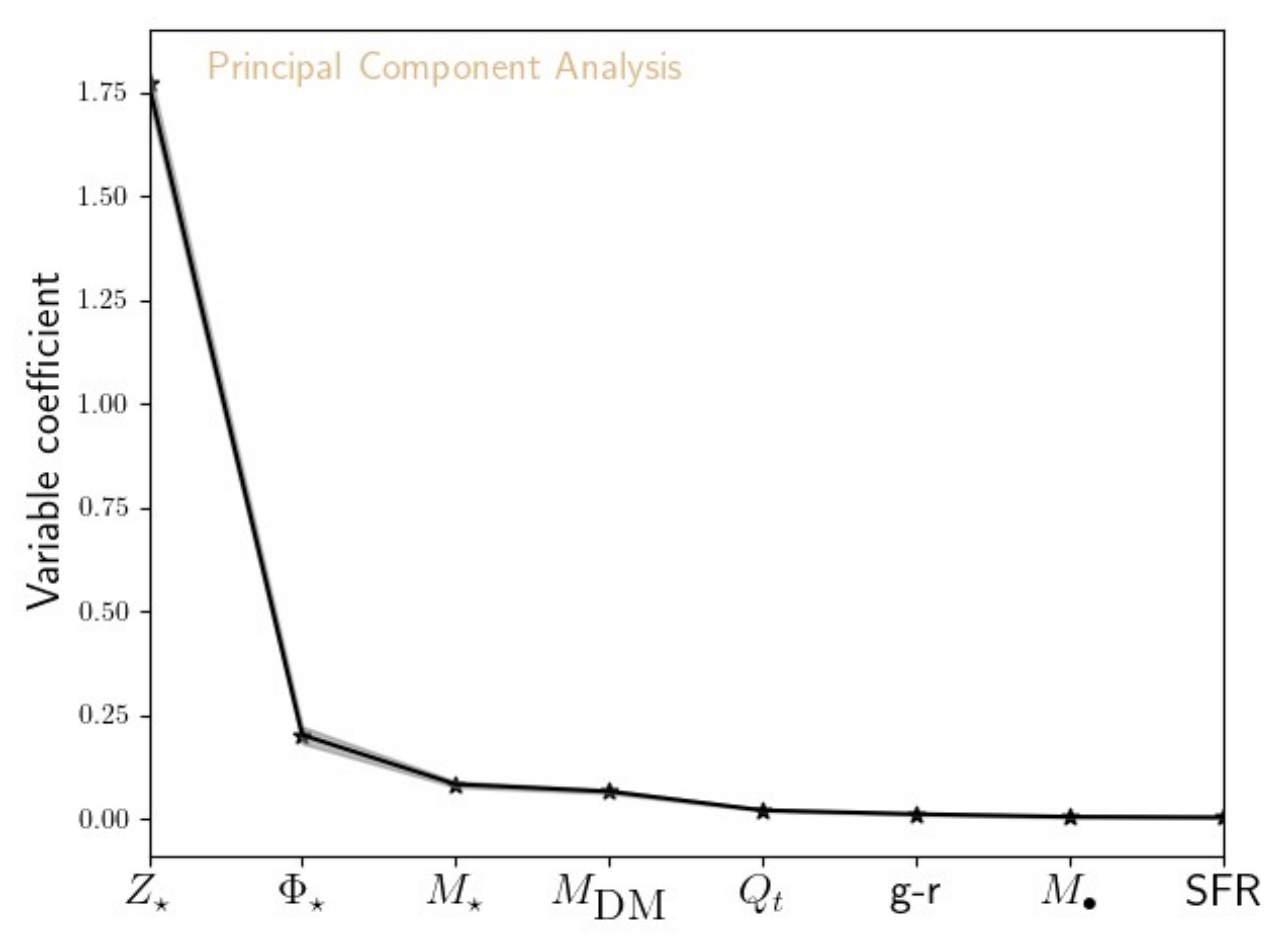


### Random forest regression

X-axis represents variables and y-axis represents their importances in predicting gas-phase metallicity.

Model is created with 200 estimators. Maximum depth of the tree is 14. Minimum samples in leaf is 5. Maximum number of features used in a tree is 7. Feature importance errors are calculated as the standard deviation of coefficients across 50 model iterations.

$R^2 = 80\%$ .

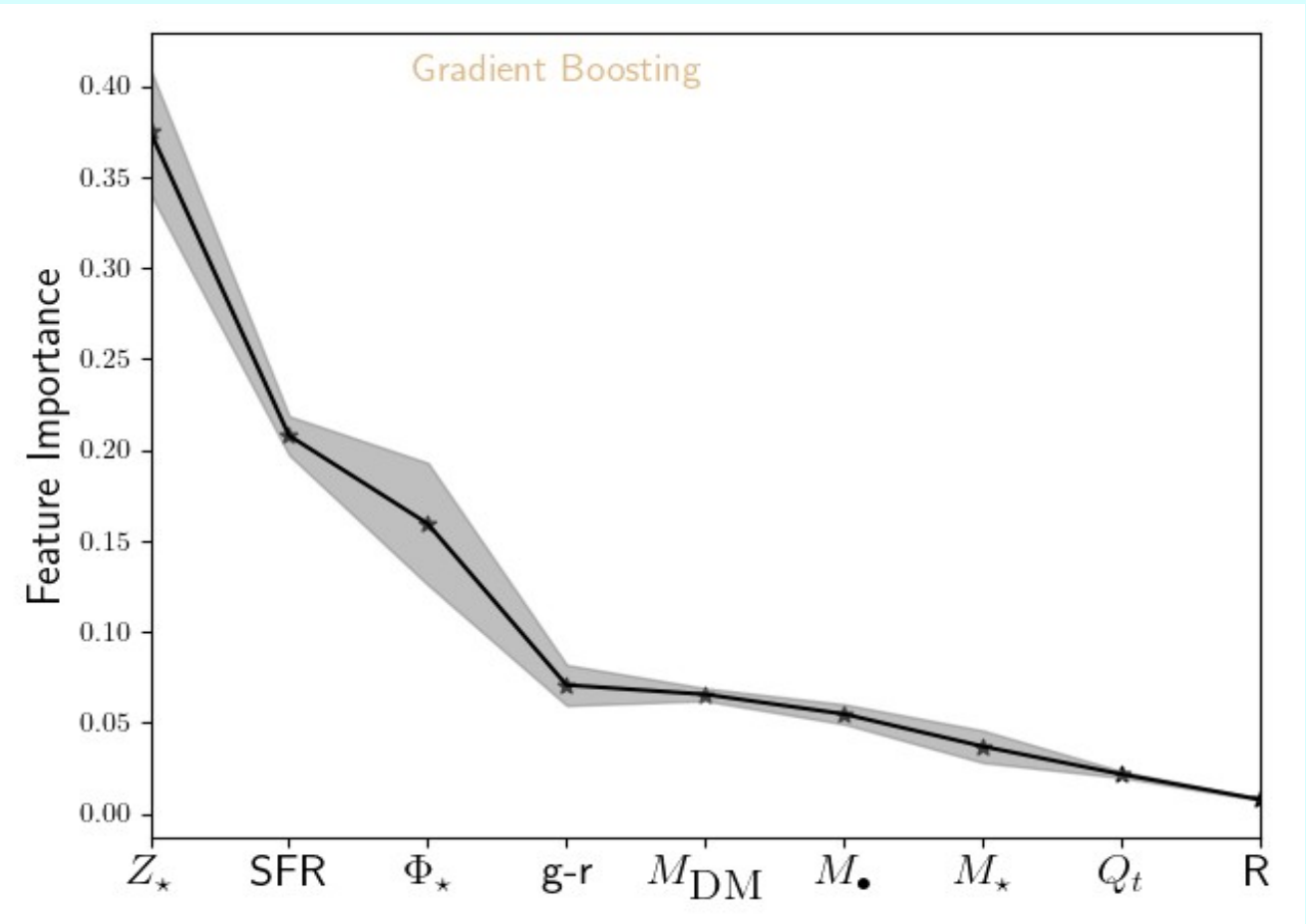


### Principal component regression

X-axis represents variables and y-axis represents their coefficients in predicting gas-phase metallicity.

Number of principal components used is 7.

$R^2 = 79\%$



### Gradient boosting regression

X-axis represents variables and y-axis represents their importances in predicting gas-phase metallicity.

Model is created with 200 estimators and learning rate 0.1. Maximum depth of the individual estimators is 10. Minimum samples in leaf is 5. Maximum number of features used in a tree is 7.

$R^2 = 80\%$

## DISCUSSION

All three models identify stellar metallicity as the most important feature for predicting gas-phase metallicity

Stellar potential is highly ranked by all three models, while stellar mass and dark matter mass show moderate importance, suggesting a connection between the chemical properties of galaxies and their mass and gravitational potential

The largest discrepancy between the models is found for SFR, suggesting that SFR contains information that is useful for predicting gas-phase metallicity in a tree-based model (RF, GB)

All three models achieve similar predictive performance ( $R^2 \sim 80\%$ ), indicating that the selected galaxy properties collectively capture a substantial fraction of the variation in gas-phase metallicity.

## FUTURE WORK

We estimated the variable importance of selected galaxy parameters for predicting gas-phase metallicity in a galaxy sample from IllustrisTNG300.

Our preliminary results suggest that discrepancies in the reported importance of different parameters may partly result from the use of different machine learning models.

In future work, we will incorporate all available galaxy parameters from IllustrisTNG.