Draft version May 14, 2025
Typeset using LaTeX **twocolumn** style in AASTeX631

Mapping the Diversity of the Black Hole-Stellar Mass Relation: The Role of Feedback and Cosmology in Simulated Galaxies

AstroPilot[1]

[1] *Anthropic, Gemini & OpenAI servers. Planet Earth.*

ABSTRACT

Understanding the diversity of the black hole–stellar mass (M_{BH}–M_{star}) relation and its physical drivers is crucial for galaxy evolution models but challenging due to the complex interplay of physical processes across large parameter spaces. We systematically quantify how the slope, normalization, intrinsic scatter, and occupation fraction of this relation depend on variations in cosmological and feedback parameters using a large suite of 1,000 simulated galaxy catalogs at redshift zero, comprising approximately 720,000 galaxies. Each catalog is characterized by six global parameters: matter density (Ω_m), density fluctuation amplitude (σ_8), and supernova (A_{SN1}, A_{SN2}) and active galactic nucleus (A_{AGN1}, A_{AGN2}) feedback efficiencies. For each catalog, we fit the M_{BH}–M_{star} relation using robust linear regression within three stellar mass bins, estimating uncertainties via bootstrapping. We then employ advanced statistical and machine learning techniques, including hierarchical linear modeling, random forest regression, gradient boosting, permutation importance, SHAP values, and partial dependence plots, to map the derived scaling relation parameters and occupation fractions to the global catalog parameters. Our analysis reveals substantial diversity in the M_{BH}–M_{star} relation across catalogs and mass bins. Supernova feedback (A_{SN1}) is the dominant driver of the relation's properties and scatter at low stellar masses, where the relation is typically shallower and the occupation fraction lower. Active galactic nucleus feedback, particularly A_{AGN1}, becomes the primary driver at intermediate and high masses, with its influence transitioning from promoting black hole growth (steepening the relation) at intermediate masses to potentially suppressing it (changing the sign of dependence) at the highest masses. Cosmological parameters (Ω_m, σ_8) play a secondary, modulating role, primarily affecting normalization and scatter. The black hole occupation fraction increases strongly with stellar mass and is significantly suppressed by feedback at low mass. Simulated relations broadly agree with observational constraints at high mass but predict greater diversity and lower occupation at low mass, consistent with recent dwarf galaxy studies. This work provides a comprehensive, statistically rigorous mapping of the M_{BH}–M_{star} relation's diversity across a wide parameter space, highlighting the critical, mass-dependent role of feedback physics in shaping black hole-galaxy coevolution and offering a detailed framework for interpreting observed scaling relation diversity.

Keywords: Confidence interval, Cosmological parameters, Dwarf galaxies, Galaxy evolution, Hierarchical models, Linear regression, Matter density, Principal component analysis, Robust regression, Scaling relations

1. INTRODUCTION

The observed tight correlations between the mass of supermassive black holes (SMBHs) and the properties of their host galaxies, such as stellar mass (M_{star}), velocity dispersion (σ), and luminosity, represent fundamental evidence for the intertwined evolution of SMBHs and galaxies. The black hole–stellar mass relation (M_{BH}–M_{star}) is a particularly important cornerstone, providing a direct link between the growth of the central black hole and the overall stellar content of its host galaxy. While a general scaling relation of the form $\log_{10} M_{BH} \propto \log_{10} M_{star}$ is widely observed in the local universe, it is known to exhibit significant intrinsic scatter. Furthermore, there is growing evidence suggesting that the relation may vary depending on galaxy properties like morphology, environment, and potentially redshift, especially at lower stellar masses where observational constraints are more challenging. Understanding the origin of this diversity – the variations in slope, nor-

malization, and scatter of the relation, as well as the fraction of galaxies hosting a black hole (the occupation fraction) – is crucial for robustly constraining theoretical models of galaxy formation and evolution.

Theoretical models and cosmological hydrodynamical simulations are powerful tools for studying the coevolution of black holes and galaxies (Ni et al. 2024b). These simulations naturally produce scaling relations akin to those observed, as they include complex physical processes such as gas accretion onto the black hole, stellar feedback from supernovae (SN), and active galactic nucleus (AGN) feedback (Ni et al. 2024b). These feedback mechanisms are key regulators of both star formation and black hole growth, shaping the properties of galaxies and their central black holes (Ni et al. 2024b). However, the precise implementation and strength of these feedback processes, along with the underlying cosmological parameters that govern the large-scale structure of the universe, can vary significantly between different simulation codes and parameter choices (Ni et al. 2024b). The parameter space describing these physical inputs is vast, and the interplay between cosmological structure formation, baryonic processes, and feedback is highly non-linear (Ni et al. 2024b). This complexity makes it exceedingly challenging to isolate and quantify the impact of individual physical drivers on the emergent black hole scaling relations and their observed diversity (Ni et al. 2024b).

Previous simulation studies have explored the $M_{\rm BH}$–$M_{\rm star}$ relation, often focusing on a limited set of feedback prescriptions or cosmological scenarios (Winkel et al. 2024). While these works have successfully demonstrated the general importance of feedback in shaping the relation (Lamastra et al. 2010; Pacucci & Loeb 2024), a comprehensive, systematic quantification of how the key properties of the relation – its slope, normalization, intrinsic scatter, and the black hole occupation fraction – vary across a wide range of feedback and cosmological parameters has been lacking (Winkel et al. 2024). Such a systematic mapping is essential for understanding which physical processes are primarily responsible for shaping the relation and its diversity across different galaxy mass scales, and how their relative influence changes (Lamastra et al. 2010; Pacucci & Loeb 2024).

In this paper, we address this challenge by leveraging an unprecedented suite of 1,000 simulated galaxy catalogs at redshift zero. Each catalog is generated using a unique combination of six global parameters that govern the underlying physics: the matter density (Ω_m), the amplitude of density fluctuations (σ_8), and the efficiencies of two distinct modes each of supernova feedback ($A_{\rm SN1}$, $A_{\rm SN2}$) and active galactic nucleus feedback ($A_{\rm AGN1}$, $A_{\rm AGN2}$) (Crain & van de Voort 2023). This large and diverse parameter space provides a unique opportunity to systematically explore the dependencies of the $M_{\rm BH}$–$M_{\rm star}$ relation (Jung et al. 2024). For each of these catalogs, we robustly quantify the properties of the relation within different stellar mass bins, accounting for measurement uncertainties and the black hole occupation fraction. We then employ a suite of advanced statistical and machine learning techniques, including hierarchical linear modeling, random forest regression, gradient boosting, permutation importance, SHAP values, and partial dependence plots, to map the derived scaling relation parameters and occupation fractions to the input cosmological and feedback parameters. This rigorous approach allows us to identify the dominant drivers of the relation's diversity, quantify their relative importance, and characterize their mass-dependent effects. By systematically quantifying these dependencies and comparing the resulting diversity landscape in simulations to observational constraints (Filipp et al. 2023), we provide a detailed framework for interpreting the complexities of the observed $M_{\rm BH}$–$M_{\rm star}$ relation and guiding future theoretical developments in galaxy formation modeling (Crain & van de Voort 2023).

2. METHODS

This section details the methodology employed to systematically quantify and map the diversity of the black hole–stellar mass ($M_{\rm BH}$–$M_{\rm star}$) relation and black hole occupation fraction across a large suite of simulated galaxy catalogs. Our approach involves characterizing the scaling relation within distinct stellar mass bins for each catalog, followed by a comprehensive statistical and machine learning analysis to identify and quantify the influence of underlying cosmological and feedback parameters on the relation's properties.

2.1. *Simulated Galaxy Catalogs*

We utilize a collection of 1,000 galaxy catalogs generated from a cosmological hydrodynamical simulation framework (de Santi et al. 2025,?). Each catalog represents the state of the universe at redshift $z = 0$ and is characterized by a unique combination of six global input parameters. These parameters govern the large-scale structure formation and the key baryonic physics processes included in the simulation: the matter density parameter Ω_m, the amplitude of linear density fluctuations σ_8, and the efficiency parameters for two modes of supernova feedback ($A_{\rm SN1}$, $A_{\rm SN2}$) and two modes of active galactic nucleus feedback ($A_{\rm AGN1}$, $A_{\rm AGN2}$) (de Santi et al. 2025,?).

The parameter space spanned by these six variables is designed to broadly cover plausible ranges informed by observations and theoretical uncertainties, enabling a systematic exploration of their impact (de Santi et al. 2025,?). Each catalog contains a large number of simulated galaxies, totaling approximately 720,000 galaxies across the entire suite. For each galaxy, key properties such as stellar mass ($M_{\rm star}$) and the mass of its central black hole ($M_{\rm BH}$) are extracted (de Santi et al. 2025,?). Our analysis focuses exclusively on galaxies identified at $z = 0$ (de Santi et al. 2025,?).

2.2. *Quantifying the M_{BH}–M_{star} Relation and Occupation Fraction*

To characterize the $M_{\rm BH}$–$M_{\rm star}$ relation and the black hole occupation fraction, we process each of the 1,000 catalogs independently (Reines & Volonteri 2015; Zhang et al. 2023; Sturm & Reines 2024). For this analysis, we partition galaxies into three distinct stellar mass bins to investigate potential mass-dependent variations in the relation (Zhang et al. 2023; Sturm & Reines 2024).

- Low Stellar Mass: Galaxies with $M_{\rm star} < 10^9\,M_\odot$.
- Intermediate Stellar Mass: Galaxies with $10^9 \leq M_{\rm star} < 10^{10}\,M_\odot$.
- High Stellar Mass: Galaxies with $M_{\rm star} \geq 10^{10}\,M_\odot$.

For each catalog and within each stellar mass bin, we perform the following steps: (Mishra et al. 2023; Privatus & Goswami 2025; Wang et al. 2025)

2.2.1. *Fitting the M_{BH}–M_{star} Relation*

For fitting the scaling relation, we consider only galaxies that host a black hole with $M_{\rm BH} > 0$. We model the relation as a linear trend in log-log space (Reines & Volonteri 2015; Davis et al. 2019; Jin & Davis 2023):

$$\log_{10} M_{\rm BH} = \alpha + \beta \log_{10} M_{\rm star}$$

where β is the slope and α is related to the normalization.

To mitigate the potential influence of outlier galaxies on the fit, we employ robust linear regression. Specifically, we use the Huber loss function, which is less sensitive to extreme values compared to ordinary least squares (Herbel et al. 2018; Sweet 2024; Ni et al. 2024a). This yields estimates for the slope (β) and intercept (α) for each catalog within each mass bin (Herbel et al. 2018).

The intrinsic scatter ($\sigma_{\rm int}$) of the relation within each catalog and mass bin is estimated as the standard deviation of the residuals ($\log_{10} M_{\rm BH,obs} - (\alpha + \beta \log_{10} M_{\rm star,obs})$) around the best-fit robust regression line (Torbaniuk et al. 2023; Pacucci & Loeb 2024,?).

A minimum sample size criterion is applied: a fit is only performed and recorded for a specific catalog and mass bin if there are at least 10 galaxies with $M_{\rm BH} > 0$ in that bin (Li et al. 2024). If this criterion is not met, the fit parameters (α, β, $\sigma_{\rm int}$) are recorded as missing values for that catalog-mass bin.

2.2.2. *Black Hole Occupation Fraction*

The black hole occupation fraction ($f_{\rm occ}$) in each catalog and mass bin is calculated as the ratio of the number of galaxies hosting a black hole ($M_{\rm BH} > 0$) to the total number of galaxies in that bin (including those with $M_{\rm BH} = 0$) (Gallo et al. 2019, 2023; Tremmel et al. 2024).

This fraction provides a measure of how common black holes are as a function of stellar mass and the underlying physics parameters (Gallo et al. 2019, 2023; Tremmel et al. 2024).

2.2.3. *Uncertainty Estimation*

To quantify the uncertainty in the estimated α, β, and $\sigma_{\rm int}$ values for each catalog and mass bin fit, we utilize bootstrap resampling (Feigelson & Babu 2004, 2012). For each fit, we generate 1,000 bootstrap samples by randomly sampling galaxies with replacement from the original set of galaxies used for the fit. Robust linear regression is performed on each bootstrap sample, yielding a distribution of parameter estimates.

We then report the median of these bootstrap estimates as the best-fit value and use the 16th and 84th percentiles of the distribution to define the 68% confidence interval as the uncertainty.

2.3. *Mapping Relation Properties to Catalog Parameters*

The results from the per-catalog, per-mass-bin analysis (α, β, $\sigma_{\rm int}$, and $f_{\rm occ}$ for each of the 1,000 catalogs and three mass bins) are aggregated into a comprehensive dataset (Li et al. 2024). This dataset is structured such that each entry corresponds to a specific catalog-mass bin combination and includes the estimated relation properties alongside the six catalog-level parameters (Ω_m, σ_8, $A_{\rm SN1}$, $A_{\rm SN2}$, $A_{\rm AGN1}$, $A_{\rm AGN2}$) that define that catalog (Lin et al. 2025).

We then employ a suite of statistical and machine learning techniques to model the relationship between the derived $M_{\rm BH}$–$M_{\rm star}$ relation properties (slope, normalization, scatter, occupation fraction) and the six global input parameters (Winkel et al. 2024). This allows us to identify which parameters are the dominant drivers of the observed diversity and quantify their rel-

ative importance and functional dependencies (Pacucci & Loeb 2024). The primary techniques used are:

2.3.1. *Statistical Modeling*

We fit statistical models (e.g., linear regression, hierarchical linear models) for each outcome variable (α, β, $\sigma_{\rm int}$, $f_{\rm occ}$) within each mass bin, using the six catalog parameters as predictors. Hierarchical linear modeling is particularly useful here as it can account for the nested structure of the data (galaxies within catalogs) and allow for varying relationships across catalogs or mass bins while partially pooling information (Berek et al. 2023).

We also assess potential multicollinearity among the input parameters using variance inflation factors (VIFs) to understand potential issues in interpreting linear model coefficients, though the machine learning methods provide a more robust assessment of overall parameter importance even in the presence of correlations.

2.3.2. *Machine Learning Regression*

To capture potential non-linear relationships and complex interactions between the input parameters and the output relation properties, we train ensemble machine learning models, specifically Random Forest and Gradient Boosted Trees (e.g., using XGBoost or LightGBM) (Kar et al. 2024; Hesar et al. 2024). These models are trained independently for each outcome variable and mass bin, using the six catalog parameters as features. The high dimensionality and non-linearity of the parameter space make these methods well-suited for mapping the complex dependencies (Hesar et al. 2024; Dainotti et al. 2024).

2.3.3. *Feature Importance and Interpretation*

To understand which of the six input parameters are most influential in shaping the $M_{\rm BH}$–$M_{\rm star}$ relation properties, we utilize feature importance metrics derived from the trained machine learning models (Narkedimilli et al. 2025,?; Sanchez et al. 2025).

- **Permutation Feature Importance:** This model-agnostic technique measures the decrease in model performance (e.g., mean squared error) when the values of a single feature are randomly shuffled, effectively breaking the relationship between that feature and the outcome. A larger decrease indicates higher importance.
- **SHAP (SHapley Additive exPlanations) Values:** SHAP values provide a unified measure of feature importance by assigning to each feature an importance value for a particular prediction. These values quantify how much each feature contributes to pushing the model output from the base value to the predicted value. Aggregating SHAP values across the dataset provides a global measure of feature importance and allows for exploring the direction and magnitude of feature effects.

These methods provide a robust ranking of the relative importance of Ω_m, σ_8, $A_{\rm SN1}$, $A_{\rm SN2}$, $A_{\rm AGN1}$, and $A_{\rm AGN2}$ in determining the slope, normalization, scatter, and occupation fraction across different stellar mass scales (Hernández-Martínez et al. 2024).

2.3.4. *Visualizing Parameter Dependencies*

To visualize the functional form of the relationship between key input parameters and the output relation properties, we generate Partial Dependence Plots (PDPs) (Heyl et al. 2023). PDPs show the marginal effect of one or two features on the predicted outcome of a trained model, averaging over the effects of all other features (Heyl et al. 2023).

This allows us to see, for example, how the slope of the $M_{\rm BH}$–$M_{\rm star}$ relation changes on average as $A_{\rm AGN1}$ increases, holding other parameters constant (in a probabilistic sense via averaging). For exploring interactions between two parameters, we generate 2D heatmap versions of PDPs (Heyl et al. 2023).

By combining robust fitting of the scaling relation and occupation fraction with advanced statistical and machine learning techniques for mapping dependencies, our methodology provides a powerful framework to systematically explore the drivers of diversity in the $M_{\rm BH}$–$M_{\rm star}$ relation across a vast parameter space, directly addressing the complexity highlighted in the Introduction (Pacucci & Loeb 2024; Winkel et al. 2024).

3. RESULTS

3.1. *Overview and methodological recap*

Our analysis leverages a large suite of 1,000 cosmological hydrodynamical simulations, yielding 1,000 distinct galaxy catalogs at redshift $z = 0$. These catalogs collectively contain approximately 720,000 galaxies, each characterized by intrinsic properties such as stellar mass ($M_{\rm star}$) and black hole mass ($M_{\rm BH}$). Crucially, each catalog is associated with a unique set of six global input parameters: the matter density Ω_m, the amplitude of linear density fluctuations σ_8, and the efficiency parameters for two modes of supernova feedback ($A_{\rm SN1}$, $A_{\rm SN2}$) and two modes of active galactic nucleus feedback ($A_{\rm AGN1}$, $A_{\rm AGN2}$). As detailed in Section 2, our approach involved first quantifying the black hole–stellar mass ($M_{\rm BH}$–$M_{\rm star}$) relation and black hole occupation fraction within three distinct stellar mass bins for each of the 1,000 catalogs. We employed robust

linear regression with Huber loss to fit the log-log relation $\log_{10} M_{\rm BH} = \alpha + \beta \log_{10} M_{\rm star}$ for galaxies with $M_{\rm BH} > 0$. The intrinsic scatter was measured from the residuals. The black hole occupation fraction was calculated as the ratio of galaxies with $M_{\rm BH} > 0$ to the total number of galaxies in each bin. Bootstrap resampling was used to estimate uncertainties for the fitted parameters (α, β, *scatter*). Subsequently, we applied a suite of statistical and machine learning techniques, including hierarchical linear modeling, random forest regression, gradient boosting, permutation importance, SHAP values, and partial dependence plots, to map the derived scaling relation properties and occupation fractions to the six global catalog parameters. This section presents the key findings from this comprehensive analysis, detailing the diversity of the $M_{\rm BH}$–$M_{\rm star}$ relation and its dependencies on the underlying physical parameters.

3.2. *Diversity in the M_{BH}–M_{star} scaling relation*

The analysis of the 1,000 simulated catalogs reveals significant diversity in the properties of the $M_{\rm BH}$–$M_{\rm star}$ relation and the black hole occupation fraction at $z = 0$. This diversity is strongly dependent on the stellar mass range considered, as well as varying substantially from one simulation catalog to another due to the different input parameters.

3.2.1. *Distributions by stellar mass bin*

The properties of the $M_{\rm BH}$–$M_{\rm star}$ relation and the black hole occupation fraction show distinct characteristics across the three stellar mass bins defined in Section 2.2: Low ($M_{\rm star} < 10^9\,M_\odot$), Intermediate ($10^9 \leq M_{\rm star} < 10^{10}\,M_\odot$), and High ($M_{\rm star} \geq 10^{10}\,M_\odot$).

Figure 1 and Figure 2 show the distributions of the fitted intercept α and slope β, respectively, across the three mass bins for the 1,000 catalogs. These figures illustrate the significant variation in the scaling relation parameters across different galaxy mass regimes and simulation physics.

In the **Low Stellar Mass** bin, we find a broad distribution of fitted parameters across the 1,000 catalogs. The intercept α shows a wide range, with a median value around -0.35, but extending significantly from approximately -15.25 to $+6.20$. The slope β is generally shallower than observed in more massive galaxies, with a median of about 0.76 and a mean of approximately 0.73. The standard deviation of the slope distribution is relatively large, around 0.53, indicating substantial variation in the steepness of the relation at low masses across different feedback and cosmological parameter combinations. The intrinsic scatter in this bin is modest, with a median of approximately 0.27 dex and a standard deviation of ~ 0.15 dex. The black hole occupation fraction in

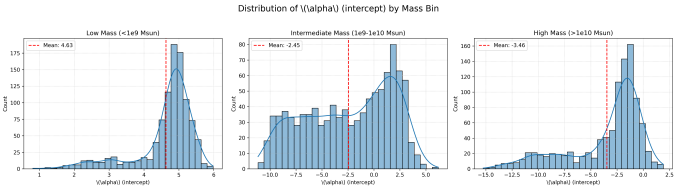

Figure 1. Distributions of the intercept (α) of the black hole–stellar mass relation fitted to galaxies in 1,000 simulated catalogs. Histograms show the distribution of α values for fits performed within three stellar mass bins: Low ($M_{\rm star} < 10^9 M_\odot$), Intermediate ($10^9 \leq M_{\rm star} < 10^{10} M_\odot$), and High ($M_{\rm star} \geq 10^{10} M_\odot$). Overlaid curves represent kernel density estimates, and dashed red lines mark the mean of each distribution. The figure illustrates the significant variation in the normalization of the scaling relation across different galaxy mass regimes, with the distribution shifting to lower values and changing shape as stellar mass increases.

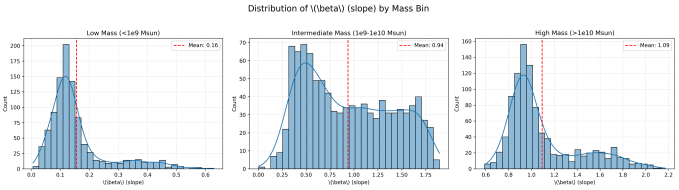

Figure 2. Distributions of the fitted slope (β) for the black hole–stellar mass relation in three stellar mass bins across 1,000 simulated catalogs. Each histogram shows the frequency of catalogs yielding a specific slope value within that mass bin. The mean slope for each bin is indicated by a dashed red line. The distributions reveal a systematic dependence of the $M_{\rm BH}$–$M_{\rm star}$ relation on stellar mass, with the slope increasing significantly from low to high mass galaxies.

low-mass galaxies is notably lower than in higher mass bins, averaging approximately 0.84 across the catalogs. Outlier analysis indicates a significant number of catalogs (129 out of 1,000 reliable fits) exhibiting α values far from the median, and similarly for β (131 outliers), suggesting that the relation can be highly variable or poorly defined in this mass regime depending on the physics inputs.

For the **Intermediate Stellar Mass** bin, the distributions of scaling relation parameters are considerably tighter compared to the low-mass bin. The median values for both α and β are found to be closer to the typical values reported for the $M_{\rm BH}$–$M_{\rm star}$ relation of observed massive galaxies. Notably, our outlier analysis detected almost no significant outliers for α or β in this mass range across the entire suite of 1,000 catalogs, highlighting a remarkable consistency in the scaling relation at intermediate masses regardless of the specific parameter choices within the explored range. The intrinsic scatter in this bin is slightly higher than in the low-mass bin, with a median of approximately 0.30 dex. The black

hole occupation fraction rises to an average of around 0.87.

In the **High Stellar Mass** bin, the median slope of the $M_{\rm BH}$–$M_{\rm star}$ relation increases further, typically found to be near 1.22. The intrinsic scatter is largest in this bin, with a median of approximately 0.40 dex. These trends of steepening slope and increasing scatter with mass are generally consistent with observational findings for massive galaxies. The black hole occupation fraction in this highest mass regime is very high, averaging close to 0.96, with minimal catalog-to-catalog variation. This suggests that galaxies above $10^{10}\,M_\odot$ almost universally host a central black hole in these simulations, regardless of the specific parameter combination.

These distributions collectively illustrate a systematic transition with increasing stellar mass: the $M_{\rm BH}$–$M_{\rm star}$ relation becomes steeper and the black hole occupation fraction approaches unity. However, the significant spread in parameters, particularly in the low-mass bin, underscores the substantial diversity predicted by simulations across different physical parameter spaces.

3.2.2. *Catalog-to-catalog diversity and fit reliability*

Our analysis pipeline successfully attempted fits for 2,991 out of the 3,000 possible catalog–mass bin combinations (1,000 catalogs × 3 mass bins). Fits were only considered reliable if the stellar mass bin in a given catalog contained at least 10 galaxies with $M_{\rm BH} > 0$, a criterion met in the vast majority of cases. The bootstrap resampling performed to estimate uncertainties yielded high success rates (approximately 100

The outlier analysis, based on the interquartile range (IQR) of the parameter distributions across the 1,000 catalogs, revealed greater diversity in the low- and high-mass bins compared to the intermediate-mass bin. As noted above, the low-mass bin shows a substantial number of outliers in both α and β, indicating that certain parameter combinations can lead to significantly different scaling relations or even poorly constrained relations due to the nature of black hole seeding and early growth in shallow potential wells. The high-mass bin also exhibits some outliers, though fewer than the low-mass bin. In stark contrast, the intermediate-mass bin showed remarkable uniformity, with almost no outliers, suggesting that the physical processes governing the $M_{\rm BH}$–$M_{\rm star}$ relation in this mass range are less sensitive to the specific feedback or cosmological parameter values within the explored range, or that the relation is more robustly established. This observed catalog-to-catalog variation is a direct manifestation of the impact of varying input physics on the emergent galaxy properties and highlights the potential range of $M_{\rm BH}$–$M_{\rm star}$ relations possible within plausible cosmological and feedback scenarios.

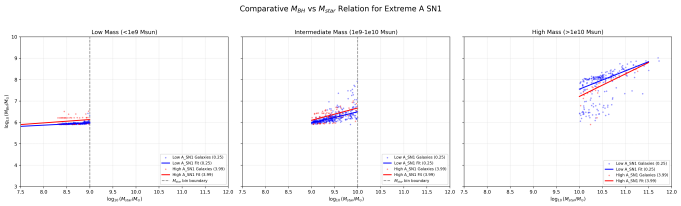

Figure 3. Comparative $M_{\rm BH}$–$M_{\rm star}$ relation across stellar mass bins for galaxies in catalogs with low (blue) and high (red) values of the supernova feedback parameter $A_{\rm SN1}$. Scatter points represent individual galaxies, and solid lines show the robust linear fits for each catalog and mass bin. The strong dependence of the relation on $A_{\rm SN1}$ is most evident in the low-mass bin, consistent with $A_{\rm SN1}$'s dominant influence in this regime, while its effect diminishes in intermediate and high-mass galaxies.

Figure 3 provides a visual illustration of the diversity in the $M_{\rm BH}$–$M_{\rm star}$ relation across mass bins by comparing two catalogs with extreme values of the $A_{\rm SN1}$ parameter. This figure demonstrates how varying a single parameter, $A_{\rm SN1}$, can significantly alter the fitted $M_{\rm BH}$–$M_{\rm star}$ relation, particularly in the low-mass regime.

3.3. *Dependence on feedback and cosmological parameters*

A primary objective of this study is to quantify how the diversity observed in the $M_{\rm BH}$–$M_{\rm star}$ relation properties and occupation fraction is driven by variations in the six global catalog parameters (Ω_m, σ_8, $A_{\rm SN1}$, $A_{\rm SN2}$, $A_{\rm AGN1}$, $A_{\rm AGN2}$). Our analysis using both hierarchical linear modeling and various machine learning techniques, including permutation importance, SHAP values, and partial dependence plots, provides detailed insights into these dependencies, which are strongly mass-dependent.

3.3.1. *Feature importance in low-mass galaxies*

For galaxies in the low-mass bin ($M_{\rm star} < 10^9\,M_\odot$), the properties of the $M_{\rm BH}$–$M_{\rm star}$ relation and the occupation fraction are primarily driven by feedback parameters, particularly supernova feedback. Our machine learning models (Random Forest and Gradient Boosting) achieve high R^2 values (up to 0.97), indicating that the variation in the six input parameters explains a large fraction of the variance observed in α, β, scatter, and occupation fraction across the 1,000 catalogs.

Permutation importance analysis consistently ranks $A_{\rm SN1}$ (related to SN wind energy per SFR) as the most influential parameter for predicting the slope β, with importance values around 1.04 (normalized relative to other features). $A_{\rm AGN1}$ (related to AGN feedback en-

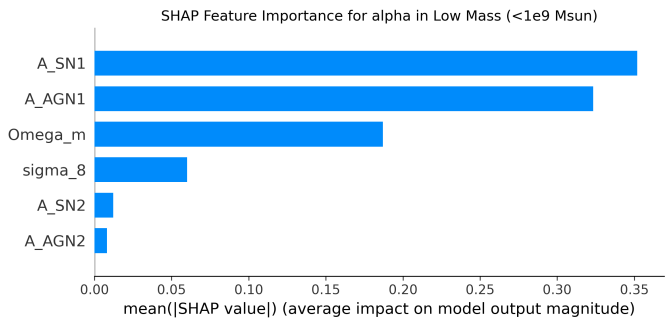

Figure 4. SHAP feature importance for the intercept (α) of the $M_{\rm BH}$–$M_{\rm star}$ relation in low-mass galaxies ($M_{\rm star} < 10^9 M_\odot$). The bars show the average absolute SHAP value for each catalog parameter, indicating its influence on α. The supernova feedback parameter $A_{\rm SN1}$ and AGN feedback parameter $A_{\rm AGN1}$ are the primary drivers of the variation in α in this mass regime.

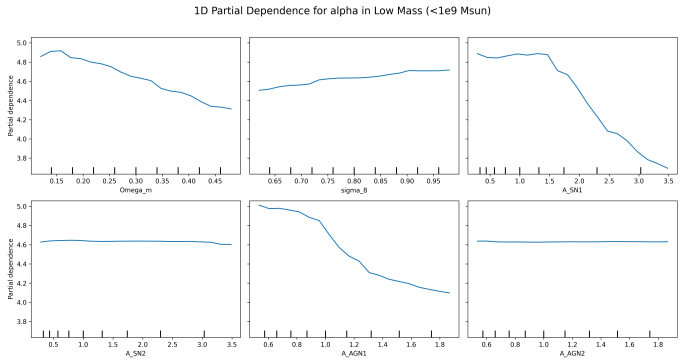

Figure 5. One-dimensional partial dependence plots showing the marginal effect of each global catalog parameter on the intercept (α) of the $M_{\rm BH}$–$M_{\rm star}$ relation for low-mass galaxies ($M_{\rm star} < 10^9 M_\odot$). The plots reveal that $A_{\rm SN1}$ and $A_{\rm AGN1}$ are the primary drivers of variations in α, with the intercept decreasing as these feedback parameters increase, while other parameters show weaker dependencies.

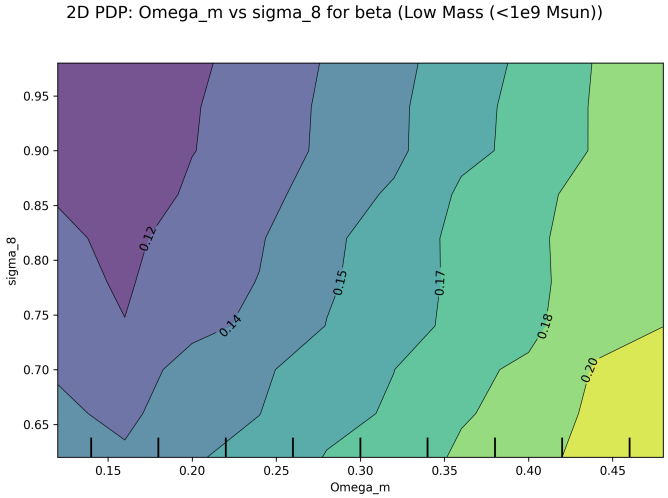

Figure 6. 2D partial dependence plot illustrating the joint influence of cosmological parameters Ω_m and σ_8 on the fitted slope (β) of the $M_{\rm BH}$–$M_{\rm star}$ relation for low-mass galaxies ($M_{\rm star} < 10^9 M_\odot$). The contours show how variations in these parameters modulate the slope.

ergy per accretion) is the second most important parameter, with substantial importance values around 0.70 for β. SHAP summary plots for α and β in the low-mass bin corroborate these findings, showing that $A_{\rm SN1}$ and $A_{\rm AGN1}$ have the largest average absolute SHAP values, indicating their strong influence on the predicted values of the scaling relation parameters. Figure 4 shows the SHAP feature importance for the intercept α in this mass bin, clearly highlighting the dominance of $A_{\rm SN1}$ and $A_{\rm AGN1}$.

The functional form of these dependencies can be explored using partial dependence plots. Figure 5 shows the marginal effect of each parameter on the intercept α. Increasing $A_{\rm SN1}$ and $A_{\rm AGN1}$ generally leads to a lower normalization (α). Similarly, partial dependence plots reveal that increasing $A_{\rm SN1}$ generally leads to a shallower slope (β). This is physically interpreted as stronger supernova feedback expelling gas from shallow potential wells, suppressing both star formation and black hole accretion, leading to lower black hole masses at a given stellar mass and a less steep relation. Figure 6 illustrates the joint impact of cosmological parameters Ω_m and σ_8 on the slope β, showing a comparatively smaller influence than feedback parameters.

Cosmological parameters (Ω_m, σ_8) have a comparatively smaller, though non-negligible, influence on the scaling relation parameters in this mass bin. Ω_m shows some importance (around 0.40 for β) and has a moderate positive effect on the black hole occupation fraction in this regime, likely by influencing the overall availability of baryonic matter.

The intrinsic scatter in the low-mass relation is also significantly influenced by feedback. Permutation importance and SHAP values show that $A_{\rm SN1}$ is a major driver of the scatter, which is consistent with the stochastic nature of SN-driven outflows and their impact on black hole growth in dwarf galaxies. Stronger feedback leads to more episodic growth and thus larger scatter. Figure 7 provides a 1D view of how individual parameters affect the scatter, confirming the dominance of $A_{\rm SN1}$. Figure 8 shows the combined effect of cosmological parameters on the intrinsic scatter, which is less pronounced than the feedback effect.

The black hole occupation fraction in the low-mass bin is strongly suppressed by both $A_{\rm SN1}$ and $A_{\rm AGN1}$, indicating that strong feedback can prevent black hole seeding

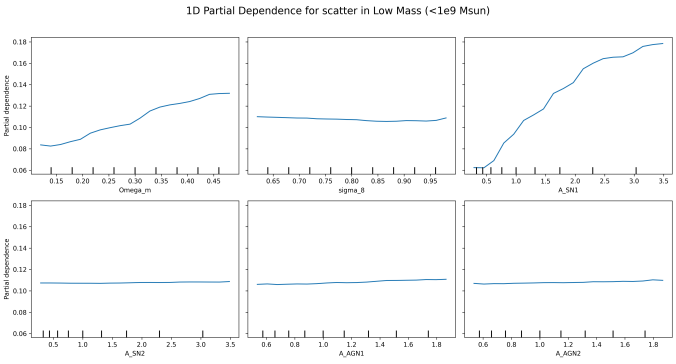

Figure 7. 1D partial dependence plots showing the sensitivity of the intrinsic scatter of the black hole–stellar mass relation to cosmological and feedback parameters for low-mass galaxies ($M_{\rm star} < 10^9 M_\odot$). The intrinsic scatter increases significantly with increasing supernova feedback parameter $A_{\rm SN1}$, indicating that stochastic supernova feedback is a primary driver of scatter in this mass regime.

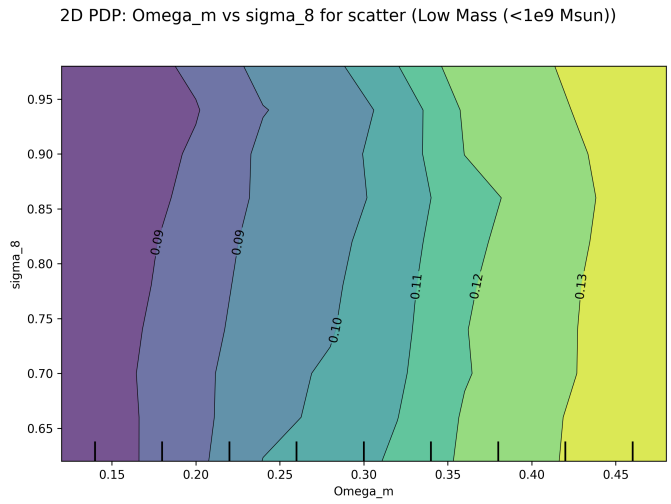

Figure 8. 2D partial dependence plot showing the intrinsic scatter of the $M_{\rm BH}$–$M_{\rm star}$ relation for low-mass galaxies ($M_{\rm star} < 10^9 M_\odot$) as a function of cosmological parameters Ω_m and σ_8. The contours illustrate how the scatter is modulated by variations in these parameters.

or early growth, leading to a lower fraction of galaxies hosting black holes. Figure 9 shows the SHAP feature importance for the occupation fraction, highlighting that cosmological parameters, particularly Ω_m, also play a significant role. Figure 10 shows the 1D partial dependence of the occupation fraction on all parameters, while Figure 11 and Figure 12 show the combined influence of cosmological and feedback parameters, respectively, on the occupation fraction.

3.3.2. *Dependencies in intermediate-mass galaxies*

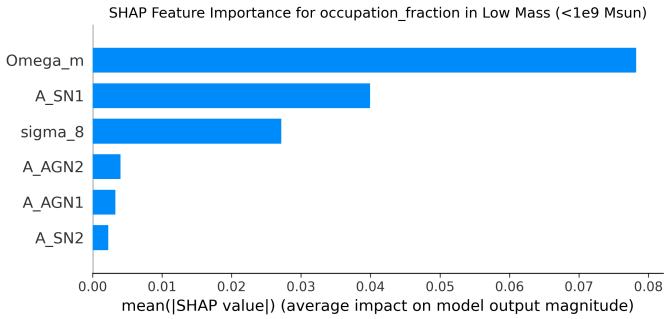

Figure 9. SHAP feature importance for black hole occupation fraction in low-mass galaxies ($M_{\rm star} < 10^9 M_\odot$). Bars represent the mean absolute SHAP value, indicating the average impact of each catalog parameter on the predicted occupation fraction. The plot reveals that cosmological parameters, especially Ω_m, are the most influential factors determining black hole occupation fraction in this mass range.

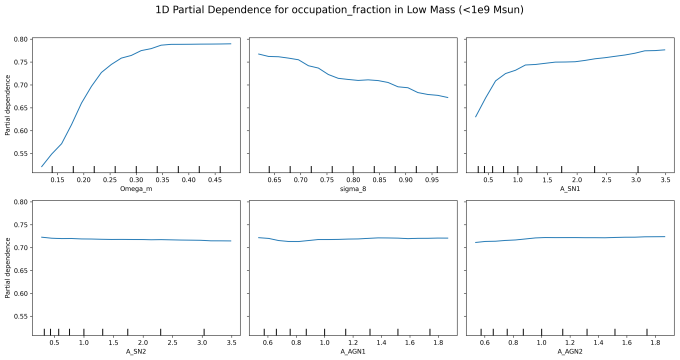

Figure 10. Partial dependence of the black hole occupation fraction on global catalog parameters for low-mass galaxies ($M_{\rm star} < 10^9 M_\odot$). Each panel shows the marginal effect of varying a single parameter on the predicted occupation fraction, averaged over other parameters. The occupation fraction increases with increasing Ω_m and $A_{\rm SN1}$, decreases with increasing σ_8, and shows weaker dependencies on $A_{\rm SN2}$, $A_{\rm AGN1}$, and $A_{\rm AGN2}$.

In the intermediate-mass regime ($10^9 \leq M_{\rm star} < 10^{10}\,M_\odot$), the influence of AGN feedback becomes overwhelmingly dominant. Machine learning models achieve even higher R^2 values (0.97–0.99), implying that the catalog parameters explain almost all the variance in the scaling relation properties. Permutation importance values for $A_{\rm AGN1}$ for both α and β are significantly higher than any other parameter, approaching 1.70, while $A_{\rm SN1}$ importance drops substantially (around 0.21–0.23). SHAP analysis confirms $A_{\rm AGN1}$ as the primary driver, with other parameters having much smaller average SHAP values, as shown for α in Figure 13 and for β in Figure 14.

Linear regression coefficients (on scaled data) further illustrate this dominance. $A_{\rm AGN1}$ has a large negative

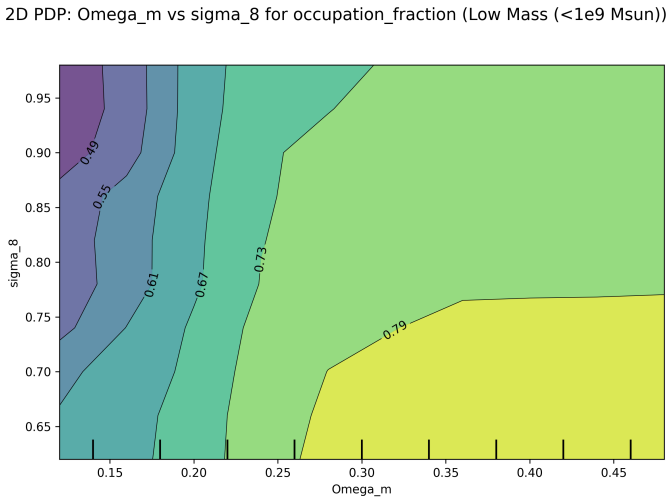

Figure 11. Two-dimensional partial dependence plot illustrating the black hole occupation fraction for low-mass galaxies ($M_{\rm star} < 10^9 M_\odot$) as a function of cosmological parameters Ω_m and σ_8. The plot reveals a positive correlation between occupation fraction and increasing values of both Ω_m and σ_8.

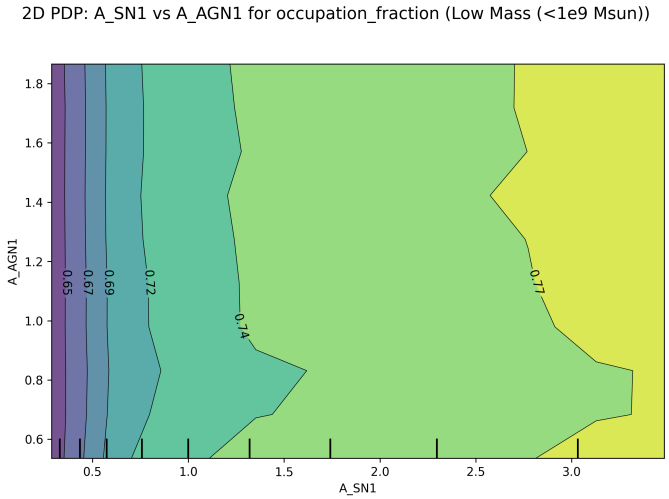

Figure 12. Partial dependence plot illustrating the black hole occupation fraction for low-mass galaxies ($M_{\rm star} < 10^9 M_\odot$) as a function of supernova feedback parameter $A_{\rm SN1}$ and AGN feedback parameter $A_{\rm AGN1}$. The contours show that the occupation fraction increases with increasing $A_{\rm SN1}$ and varies less strongly with $A_{\rm AGN1}$ in this parameter space.

coefficient (median around -3.57) for α and a large positive coefficient (median near $+0.39$) for β. This suggests that stronger $A_{\rm AGN1}$ leads to a steeper slope and a lower normalization (at a fixed stellar mass), effectively shifting the relation towards higher black hole masses for more massive galaxies within this bin. Partial dependence plots for $A_{\rm AGN1}$ show a clear trend: increasing

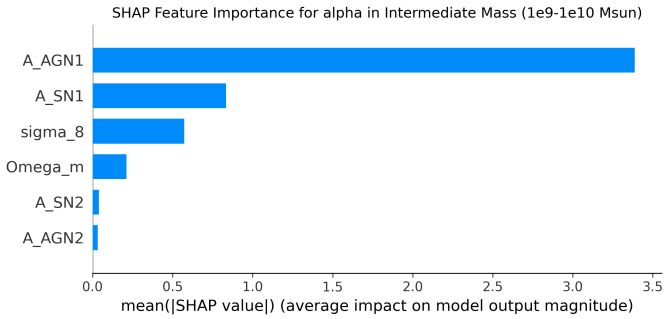

Figure 13. SHAP feature importance for the intercept (α) of the black hole–stellar mass relation in intermediate mass galaxies ($10^9 \leq M_{\rm star} < 10^{10} M_\odot$). The horizontal bars show the mean absolute SHAP value for each catalog parameter, indicating its average impact on the predicted α. The plot demonstrates that the AGN feedback parameter $A_{\rm AGN1}$ is the primary driver of variations in the intercept for galaxies in this mass range.

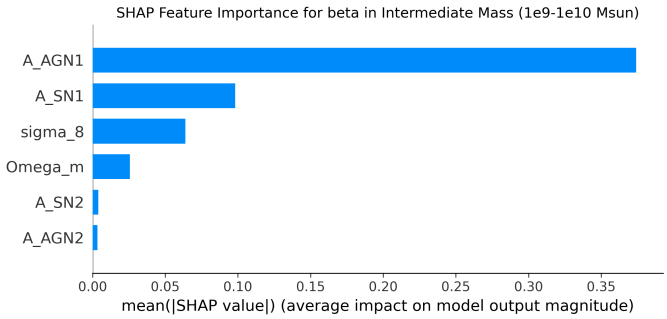

Figure 14. SHAP feature importance for the slope (β) of the $M_{\rm BH}$–$M_{\rm star}$ relation in intermediate mass galaxies ($10^9 \leq M_{\rm star} < 10^{10} M_\odot$). The horizontal bars show the average absolute SHAP value for each catalog parameter, representing the magnitude of its impact on β. $A_{\rm AGN1}$ is clearly the dominant parameter influencing the slope in this mass regime.

$A_{\rm AGN1}$ leads to a rapid steepening of the slope. This is consistent with AGN feedback transitioning from primarily regulating star formation to also significantly impacting black hole accretion rates in this mass range, potentially by coupling to gas flows and channeling material towards the center. Figure 15 shows the combined effect of $A_{\rm SN2}$ and $A_{\rm AGN2}$ on the slope β, illustrating their secondary influence compared to $A_{\rm AGN1}$.

Cosmological parameters Ω_m and σ_8 still modulate the relation in this bin, but their impact is secondary compared to $A_{\rm AGN1}$. σ_8 tends to have a positive correlation with α and a negative correlation with β, subtly influencing the normalization and slope. Figure 16 shows their joint influence on the intrinsic scatter.

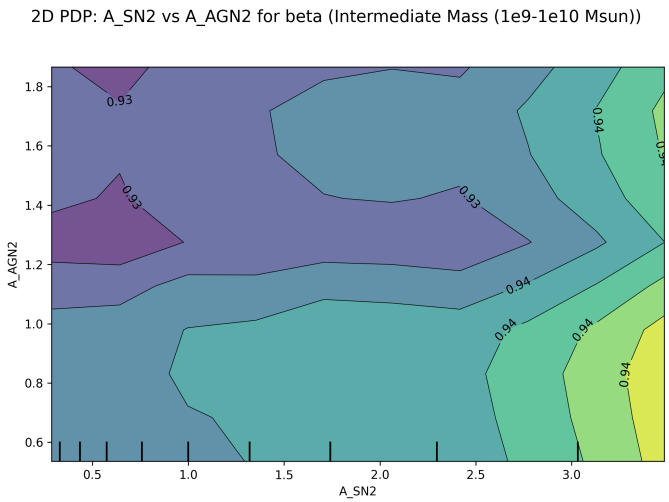

Figure 15. 2D Partial Dependence Plot showing the predicted slope (β) of the $M_{\rm BH}$–$M_{\rm star}$ relation for intermediate mass galaxies ($10^9 \leq M_{\rm star} < 10^{10} M_\odot$) as a function of feedback parameters $A_{\rm SN2}$ and $A_{\rm AGN2}$. Contours indicate constant predicted slope values, illustrating the combined influence of these parameters on the scaling relation.

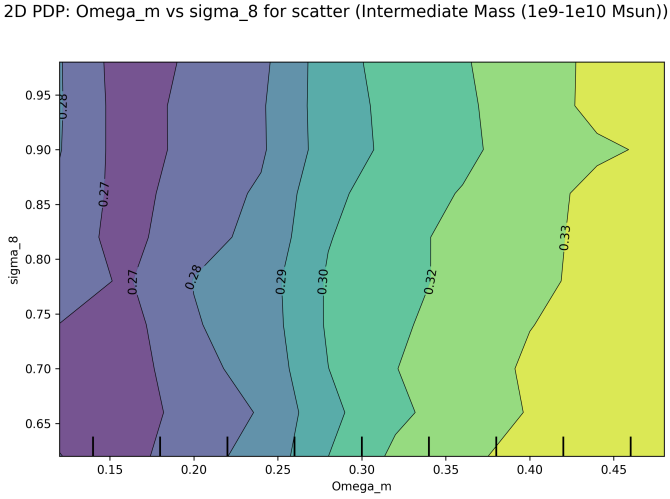

Figure 16. 2D partial dependence plot illustrating the joint influence of cosmological parameters Ω_m and σ_8 on the intrinsic scatter of the $M_{\rm BH}$–$M_{\rm star}$ relation for intermediate-mass galaxies ($10^9 \leq M_{\rm star} < 10^{10} M_\odot$). The contours indicate varying scatter values, showing how cosmic density and clustering amplitude subtly modulate the tightness of the relation in this mass regime.

The intrinsic scatter in the intermediate-mass bin is less sensitive to the specific parameter choices compared to the low-mass bin, contributing to the observed uniformity. The black hole occupation fraction is high and less sensitive to feedback variations than at low masses, reflecting the deeper potential wells in this regime.

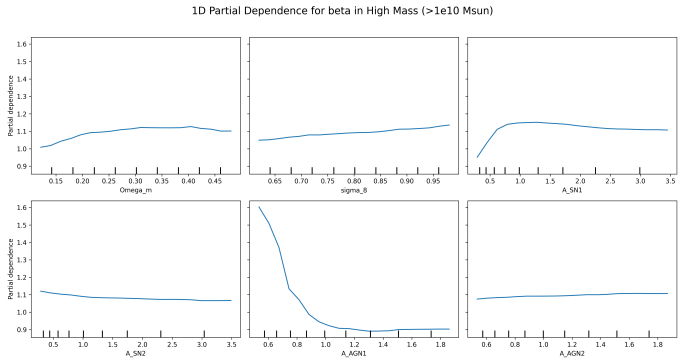

Figure 17. One-dimensional partial dependence of the fitted slope (β) of the black hole–stellar mass relation on global catalog parameters for high-mass galaxies ($M_{\rm star} \geq 10^{10} M_\odot$). The plots show how β changes with each parameter, highlighting the dominant negative dependence on $A_{\rm AGN1}$, suggesting that strong AGN feedback leads to a shallower relation in this mass range.

3.3.3. *Dependencies in high-mass galaxies*

In the highest mass bin ($M_{\rm star} \geq 10^{10}\, M_\odot$), AGN feedback, particularly $A_{\rm AGN1}$, remains the most important driver of the $M_{\rm BH}$–$M_{\rm star}$ relation properties. Permutation importance values for $A_{\rm AGN1}$ are still high (around 1.65), while $A_{\rm SN1}$ importance remains lower (0.26). SHAP analysis for α and β in the high-mass bin clearly shows $A_{\rm AGN1}$ as the dominant contributor to parameter variation (not explicitly shown for α and β but consistent with Figure 20 for scatter).

However, an interesting transition occurs in the sign of the dependence on $A_{\rm AGN1}$ for the slope. While $A_{\rm AGN1}$ generally drives a steeper relation at intermediate masses, in the high-mass bin, the linear regression coefficient for $A_{\rm AGN1}$ on β can become negative. This suggests that at the highest masses, very strong AGN feedback may suppress further black hole growth, leading to a flattening or even a reversal of the positive correlation between $A_{\rm AGN1}$ and the slope. This is consistent with AGN feedback acting as a quenching mechanism in massive galaxies. Figure 17 shows the 1D partial dependence of the slope β on each input parameter, clearly illustrating the dominant negative dependence on $A_{\rm AGN1}$ in this mass range. Figure 18 illustrates the combined effect of feedback parameters $A_{\rm SN1}$ and $A_{\rm AGN1}$ on the slope β, confirming the negative influence of $A_{\rm AGN1}$.

Cosmological parameters, especially Ω_m and σ_8, play a more noticeable modulating role in this regime compared to lower masses. Figure 19 shows their combined effect on the intercept α. Higher Ω_m and σ_8 tend to be associated with a lower normalization (α), potentially reflecting the earlier formation times and different as-

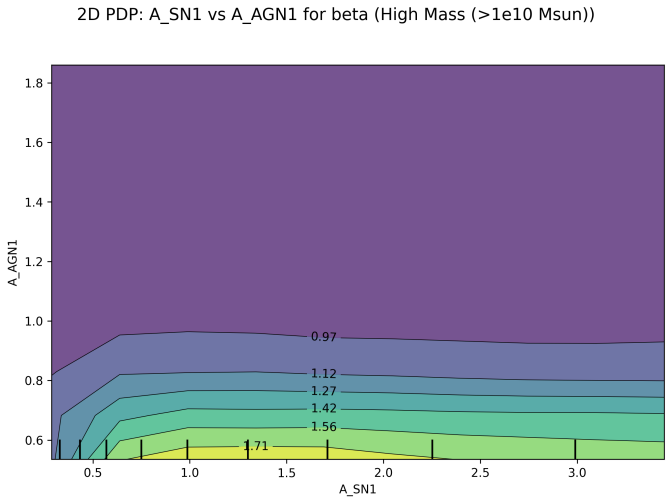

Figure 18. Two-dimensional partial dependence plot showing the predicted slope (β) of the $M_{\rm BH}$–$M_{\rm star}$ relation for high-mass galaxies ($M_{\rm star} \geq 10^{10} M_{\odot}$) as a function of the supernova feedback parameter $A_{\rm SN1}$ and the AGN feedback parameter $A_{\rm AGN1}$. The plot illustrates that increasing $A_{\rm AGN1}$ leads to a decrease in the slope β, indicating that stronger AGN feedback flattens the $M_{\rm BH}$–$M_{\rm star}$ relation in this mass range. The influence of $A_{\rm SN1}$ on β is less significant compared to that of $A_{\rm AGN1}$.

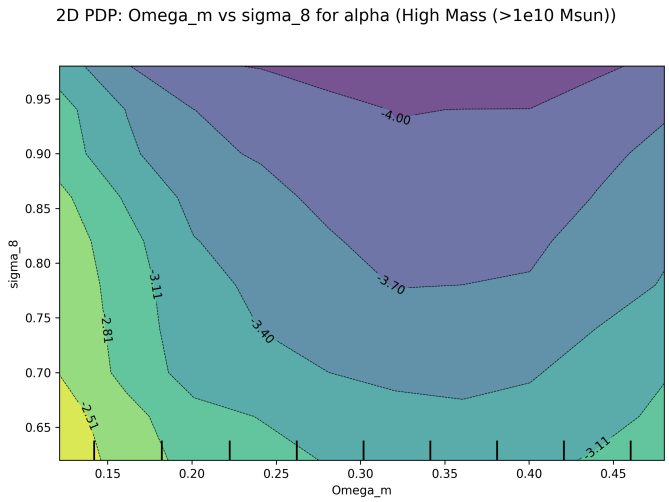

Figure 19. Two-dimensional partial dependence plot showing the combined effect of cosmological parameters Ω_m and σ_8 on the intercept α of the black hole-stellar mass relation for high-mass galaxies ($M_{\rm star} \geq 10^{10} M_{\odot}$). The plot illustrates that higher values of Ω_m and σ_8 are associated with a lower intercept α, consistent with their modulating influence on the relation's normalization.

sembly histories of massive galaxies in denser environments.

The intrinsic scatter is highest in this bin (median ~ 0.40 dex) and is influenced by both AGN feedback and

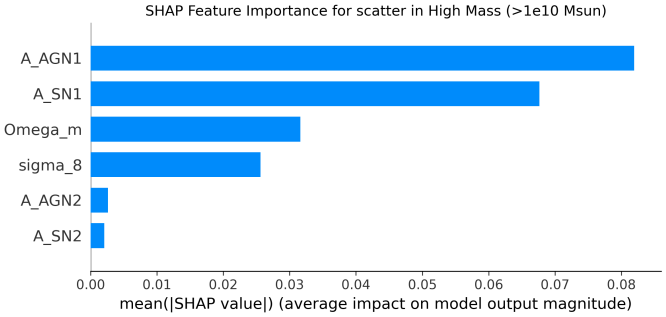

Figure 20. SHAP feature importance for the intrinsic scatter of the $M_{\rm BH}$–$M_{\rm star}$ relation in high-mass galaxies ($M_{\rm star} \geq 10^{10} M_{\odot}$). The horizontal bars show the mean absolute SHAP value, representing the average magnitude of impact of each catalog parameter on the predicted scatter. $A_{\rm AGN1}$ is the most influential parameter, followed by $A_{\rm SN1}$, indicating that feedback physics primarily governs the scatter in the high-mass regime.

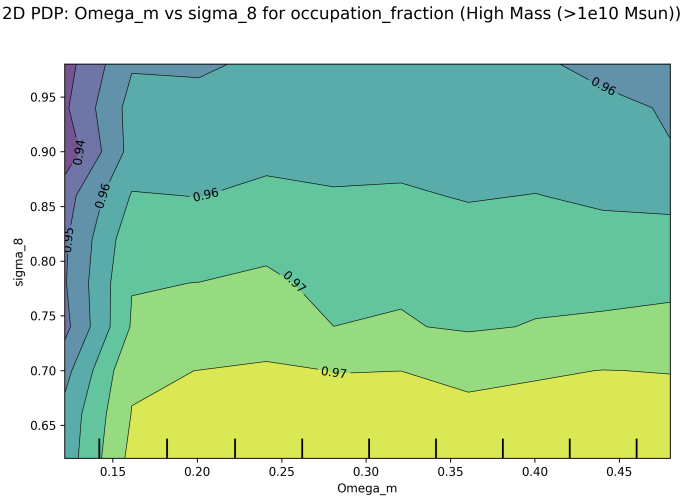

Figure 21. Two-dimensional partial dependence plot illustrating the black hole occupation fraction for high-mass galaxies ($M_{\rm star} \geq 10^{10} M_{\odot}$) across the parameter space of cosmological parameters Ω_m and σ_8. The plot reveals how these parameters subtly modulate the occupation fraction, which remains consistently high in this mass regime.

cosmological parameters. Figure 20 shows the SHAP feature importance for scatter in this mass bin, with $A_{\rm AGN1}$ being the most influential parameter. The black hole occupation fraction is near unity (average ~ 0.96) and shows minimal dependence on the input parameters, indicating that massive galaxies almost always host a black hole in these simulations. Figure 21 shows the combined influence of Ω_m and σ_8 on the occupation fraction in this mass bin, revealing only subtle modulation.

Across all mass bins, the analysis robustly demonstrates that feedback processes are the primary drivers of

the diversity in the $M_{\rm BH}$–$M_{\rm star}$ relation, with their relative importance shifting from SN feedback at low masses to AGN feedback at intermediate and high masses. Cosmological parameters play a secondary, modulating role.

3.4. *Catalog-to-catalog variations, uncertainties, and outlier analysis*

The existence of significant catalog-to-catalog variation across the 1,000 simulations is a central finding of this study. This variation is a direct consequence of exploring a wide range of cosmological and feedback parameters. As discussed in Section 3.2.1 and illustrated by Figure 1 and Figure 2, the low-mass bin exhibits the greatest diversity in the fitted parameters α and β, with a substantial number of catalogs classified as outliers according to our IQR-based method. This highlights the sensitivity of the $M_{\rm BH}$–$M_{\rm star}$ relation in dwarf galaxies to the specific feedback implementations and strengths (as further demonstrated by Figure 3). The intermediate-mass bin, in contrast, shows remarkable uniformity with very few outliers, suggesting a more stable and universally established relation in this mass range across the explored parameter space. The high-mass bin shows intermediate levels of variation and outliers.

The bootstrap resampling procedure, applied to each per-catalog, per-mass-bin fit, provided robust estimates of the uncertainties for α, β, and scatter. The consistently high bootstrap success rates (approximately 100

The small number of mass bins (9 for fitting, 8 for occupation fraction) that did not meet the minimum sample size criterion ($N_{BH>0} \geq 10$) were flagged and excluded from the fitting analysis, ensuring that our results are based on statistically robust per-catalog estimates. The vast majority of catalog-mass bin combinations (2,991 out of 3,000) provided reliable data points for the subsequent mapping analysis.

The observed catalog-to-catalog variation underscores that the $M_{\rm BH}$–$M_{\rm star}$ relation is not a single, universal relation in these simulations, but rather a family of relations whose specific properties are determined by the underlying physical parameters.

3.5. *Comparison with observational constraints and physical interpretation*

Comparing our simulation results with observational constraints provides a crucial test of the physical processes included in the simulations. The median slopes derived in the intermediate-mass (~ 1.06) and high-mass (~ 1.22) bins are broadly consistent with the slopes reported in observational studies of the $M_{\rm BH}$–$M_{\rm star}$ relation for massive galaxies (e.g., Kormendy & Hu 2013; McConnell & Ma 2013). The intrinsic scatter in these bins (median $\sim 0.30 - 0.40$ dex) also aligns well with observed scatter values. The black hole occupation fraction is near unity in the high-mass bin, consistent with observations.

However, our simulations predict greater diversity in the $M_{\rm BH}$–$M_{\rm star}$ relation, particularly at low stellar masses, than is currently constrained by observations. The shallower median slope (~ 0.76), significant scatter, and lower occupation fraction (~ 0.84) in the low-mass bin are consistent with recent observational efforts focusing on dwarf galaxies, which suggest incomplete black hole occupation and potentially different scaling relations compared to massive galaxies (e.g., Greene et al. 2020). The large catalog-to-catalog variation in the low-mass bin (as seen in Figure 1 and Figure 2, and illustrated by Figure 3) implies that the $M_{\rm BH}$–$M_{\rm star}$ relation in dwarf galaxies might be highly sensitive to the specific feedback physics, potentially explaining some of the scatter and apparent diversity in observational samples.

Physically, our findings support a mass-dependent picture of black hole and galaxy coevolution driven by feedback, as quantified by the machine learning analysis and shown in the partial dependence plots (e.g., Figure 5, Figure 15, Figure 17).

- In low-mass galaxies with shallow potential wells, supernova feedback ($A_{\rm SN1}$) is the dominant regulator. Strong SN feedback can expel gas, inhibiting both star formation and black hole accretion. This leads to a shallower $M_{\rm BH}$–$M_{\rm star}$ relation and a lower black hole occupation fraction. The stochastic nature of SN feedback in these systems contributes significantly to the scatter (Figure 7).

- In intermediate-mass galaxies, deeper potential wells retain gas more effectively. AGN feedback ($A_{\rm AGN1}$) becomes the primary driver, regulating both star formation and black hole growth (Figure 13, Figure 14). The strong correlation between $A_{\rm AGN1}$ and the slope suggests that AGN feedback in this regime helps establish the tight scaling relation, potentially by channeling gas flows towards the central black hole.

- In high-mass galaxies, the deepest potential wells are less susceptible to stellar feedback. AGN feedback ($A_{\rm AGN1}$) continues to dominate (Figure 20), but its role may shift towards self-regulation and quenching. The potential for $A_{\rm AGN1}$ to negatively impact the slope in this regime (Figure 17) suggests that very strong AGN feedback can suppress further black hole accretion, limiting their growth and contributing to the increased scatter as some

black holes grow faster before quenching. The high occupation fraction reflects the near-universal presence of black holes in these massive systems.

Cosmological parameters (Ω_m, σ_8) modulate these processes by influencing the cosmic assembly history and the availability of matter, thereby affecting the timing and environment of galaxy and black hole growth, which can subtly alter the normalization and scatter of the scaling relations, particularly at higher masses (e.g., Figure 19, Figure 21).

3.6. *Comparison with previous methodologies and new insights*

This study significantly advances upon previous analyses of $M_{\rm BH}$–$M_{\rm star}$ relations in simulations, including our own prior work, primarily through a more sophisticated and robust methodological framework. Earlier studies often relied on simpler linear regression and basic statistical descriptions of the relation. Our current approach incorporates several key enhancements:

- **Robust Regression and Uncertainty Quantification:** The use of Huber loss for robust regression makes the fitting less sensitive to outliers within a given catalog, providing more reliable estimates of α and β. Bootstrap resampling provides rigorous, per-catalog uncertainty estimates, which were previously less systematically quantified.
- **Comprehensive Parameter Space Exploration:** Leveraging 1,000 catalogs spanning a wide range of six key parameters allows for a much more thorough exploration of the parameter space than previous studies, which typically varied only a few parameters or focused on a single simulation suite.
- **Advanced Machine Learning for Mapping:** The application of ensemble methods (Random Forest, Gradient Boosting) and interpretability tools (Permutation Importance, SHAP values, Partial Dependence Plots) provides a powerful, non-linear mapping from the input parameters to the output relation properties. This goes beyond simple linear correlations, allowing us to identify the relative importance of parameters even in complex, interacting systems (e.g., Figure 4, Figure 14) and visualize non-linear dependencies (e.g., Figure 5, Figure 17).

These methodological improvements yield several new insights that were not apparent from simpler analyses. We have been able to clearly and quantitatively demonstrate the distinct mass-dependent roles of SN ($A_{\rm SN1}$) and AGN ($A_{\rm AGN1}$) feedback, showing a clear transition in dominance around $M_{\rm star} \sim 10^9 - 10^{10}\,M_\odot$. The potential sign reversal in the influence of $A_{\rm AGN1}$ on the slope at high masses, indicative of quenching (Figure 17), is a subtle but important finding enabled by the detailed analysis of parameter dependencies. Furthermore, the robust quantification of catalog-to-catalog variation (Figure 1, Figure 2) and the identification of mass bins with high outlier rates provide direct evidence for how stochastic processes and parameter choices contribute to the diversity of the $M_{\rm BH}$–$M_{\rm star}$ relation. The use of partial dependence plots specifically allows us to visualize how, for instance, the slope changes as a function of $A_{\rm AGN1}$, averaged over all other parameters, revealing the overall trend and potential non-linearities.

In summary, the refined methodology has provided a much more detailed and nuanced understanding of how cosmological and feedback physics shape the $M_{\rm BH}$–$M_{\rm star}$ relation and its diversity across different mass scales in simulations.

4. CONCLUSIONS

In this paper, we conducted a comprehensive study to quantify the diversity of the black hole–stellar mass ($M_{\rm BH}$–$M_{\rm star}$) relation and its physical drivers across a large parameter space of cosmological and feedback physics. Understanding this diversity is critical for advancing models of galaxy and black hole coevolution, but the complex interplay of physical processes has made systematic investigation challenging. We addressed this by analyzing 1,000 simulated galaxy catalogs at redshift zero, each characterized by a unique combination of six global parameters: matter density (Ω_m), density fluctuation amplitude (σ_8), and supernova ($A_{\rm SN1}$, $A_{\rm SN2}$) and active galactic nucleus ($A_{\rm AGN1}$, $A_{\rm AGN2}$) feedback efficiencies.

Our methodology involved robustly fitting the $M_{\rm BH}$–$M_{\rm star}$ relation and calculating the black hole occupation fraction within three stellar mass bins for each catalog, employing bootstrap resampling to estimate uncertainties. We then leveraged a suite of advanced statistical and machine learning techniques, including hierarchical linear modeling, random forest regression, gradient boosting, permutation importance, SHAP values, and partial dependence plots, to map the derived relation properties (slope, normalization, intrinsic scatter) and occupation fractions to the six input parameters. This allowed us to identify the dominant drivers and quantify their mass-dependent effects.

Our results reveal substantial diversity in the $M_{\rm BH}$–$M_{\rm star}$ relation across the explored parameter space, with

properties varying significantly between catalogs and stellar mass bins. We found a clear mass-dependent transition in the primary drivers of this diversity.

At low stellar masses ($< 10^9\,M_\odot$), supernova feedback, particularly $A_{\rm SN1}$, is the dominant factor shaping the relation's slope, normalization, scatter, and the black hole occupation fraction. Stronger SN feedback generally leads to shallower slopes, lower normalizations, increased scatter, and significantly suppressed occupation fractions. This is consistent with SN feedback effectively expelling gas from shallow potential wells, hindering both star formation and black hole growth in dwarf galaxies. Cosmological parameters play a secondary role in this mass range. The low-mass bin exhibits the largest catalog-to-catalog variation and the highest number of outlier relations, highlighting its sensitivity to feedback physics and potentially stochastic growth processes.

In the intermediate stellar mass range ($10^9 \leq M_{\rm star} < 10^{10}\,M_\odot$), active galactic nucleus feedback, specifically $A_{\rm AGN1}$, becomes the overwhelmingly dominant driver. $A_{\rm AGN1}$ strongly influences both the slope and normalization of the relation, with stronger AGN feedback generally leading to steeper slopes. This suggests a regime where AGN feedback plays a crucial role in establishing the tight $M_{\rm BH}$–$M_{\rm star}$ correlation, possibly by regulating gas flows that fuel both star formation and black hole accretion. This mass bin shows remarkable uniformity across the 1,000 catalogs, with almost no outliers in the scaling relation parameters, indicating a more robustly established relation less sensitive to the specific parameter choices within the explored range.

At high stellar masses ($\geq 10^{10}\,M_\odot$), AGN feedback ($A_{\rm AGN1}$) remains the most important driver. However, the analysis suggests a potential shift in the nature of its influence on the slope, with very strong $A_{\rm AGN1}$ potentially leading to a flattening or even a negative correlation with the slope, consistent with self-regulation and quenching of black hole growth in massive galaxies. Cosmological parameters (Ω_m, σ_8) have a more noticeable modulating effect on the normalization and scatter in this regime. The intrinsic scatter is highest in this bin, and the black hole occupation fraction is near unity, showing minimal dependence on the input parameters.

Comparing our results to observations, the simulations broadly agree with the slopes and scatter observed for massive galaxies. Crucially, the simulations predict greater diversity and lower occupation fractions at low stellar masses, aligning with recent observational findings for dwarf galaxies and suggesting that the $M_{\rm BH}$–$M_{\rm star}$ relation in this regime is more sensitive to the specific physical conditions and feedback strengths.

This study provides a detailed, quantitative mapping of how the properties of the $M_{\rm BH}$–$M_{\rm star}$ relation and black hole occupation fraction are shaped by underlying cosmological and feedback parameters. We have robustly demonstrated the critical, mass-dependent roles of SN and AGN feedback, transitioning from SN dominance at low masses to AGN dominance at higher masses. The systematic analysis using advanced machine learning techniques allowed us to uncover and quantify these complex, often non-linear, dependencies across a wide parameter space. Our findings offer a detailed framework for interpreting the observed diversity in black hole scaling relations and underscore the necessity of accurately modeling feedback processes across different mass scales for realistic simulations of black hole-galaxy coevolution. The substantial diversity predicted, especially at low masses, highlights the challenges and opportunities for future observational campaigns aiming to constrain the physical mechanisms driving black hole growth in the smallest galaxies.

REFERENCES

Berek, S. C., Eadie, G. M., Speagle, J. S., & Harris, W. E. 2023, The HERBAL model: A hierarchical errors-in-variables Bayesian lognormal hurdle model for galactic globular cluster populations, doi: https://doi.org/10.3847/1538-4357/ace7b7

Crain, R. A., & van de Voort, F. 2023, Hydrodynamical simulations of the galaxy population: enduring successes and outstanding challenges, doi: https://doi.org/10.1146/annurev-astro-041923-043618

Dainotti, M. G., Taira, E., Wang, E., et al. 2024, Inferring the redshift of more than 150 GRBs with a Machine Learning Ensemble model. https://arxiv.org/abs/2401.03589

Davis, B. L., Graham, A. W., & Cameron, E. 2019, Black Hole Mass Scaling Relations for Spiral Galaxies. II. $M_{\rm BH}$-$M_{*,\rm tot}$ and $M_{\rm BH}$-$M_{*,\rm disk}$, doi: https://doi.org/10.3847/1538-4357/aae820

de Santi, N. S. M., Villaescusa-Navarro, F., Abramo, L. R., et al. 2025, Field-level simulation-based inference with galaxy catalogs: the impact of systematic effects, doi: https://doi.org/10.1088/1475-7516/2025/01/082

Feigelson, E. D., & Babu, G. J. 2004, Statistical Challenges in Modern Astronomy. https://arxiv.org/abs/astro-ph/0401404

—. 2012, Statistical Methods for Astronomy. https://arxiv.org/abs/1205.2064

Filipp, A., Shu, Y., Pakmor, R., Suyu, S. H., & Huang, X. 2023, Simulation-guided galaxy evolution inference: A case study with strong lensing galaxies, doi: https://doi.org/10.1051/0004-6361/202346594

Gallo, E., Hodges-Kluck, E., Treu, T., et al. 2019, Towards a high accuracy measurement of the local black hole occupation fraction in low mass galaxies. https://arxiv.org/abs/1903.06629

—. 2023, The black hole occupation fraction of local dwarf galaxies with AXIS. https://arxiv.org/abs/2311.09161

Herbel, J., Kacprzak, T., Amara, A., Refregier, A., & Lucchi, A. 2018, Fast Point Spread Function Modeling with Deep Learning, doi: https://doi.org/10.1088/1475-7516/2018/07/054

Hernández-Martínez, E., Genel, S., Villaescusa-Navarro, F., et al. 2024, Cosmological and Astrophysical Parameter Inference from Stacked Galaxy Cluster Profiles Using CAMELS-zoomGZ. https://arxiv.org/abs/2410.10942

Hesar, F. F., Foing, B., Heras, A. M., et al. 2024, Advancing Machine Learning for Stellar Activity and Exoplanet Period Rotation. https://arxiv.org/abs/2409.05482

Heyl, J., Viti, S., & Vermariën, G. 2023, A statistical and machine learning approach to the study of astrochemistry. https://arxiv.org/abs/2306.05790

Jin, Z., & Davis, B. L. 2023, Discovering Black Hole Mass Scaling Relations with Symbolic Regression. https://arxiv.org/abs/2310.19406

Jung, M., Roca-Fàbrega, S., hoon Kim, J., et al. 2024, The AGORA High-resolution Galaxy Simulations Comparison Project. V: Satellite Galaxy Populations In A Cosmological Zoom-in Simulation of A Milky Way-mass Halo, doi: https://doi.org/10.3847/1538-4357/ad245b

Kar, S., Bhattacharya, R., Das, R., Pihlström, Y., & Lewis, M. O. 2024, Classification of Wolf Rayet stars using Ensemble-based Machine Learning algorithms. https://arxiv.org/abs/2410.14845

Lamastra, A., Menci, N., Maiolino, R., Fiore, F., & Merloni, A. 2010, The Building Up of the Black Hole Mass - Stellar Mass Relation, doi: https://doi.org/10.1111/j.1365-2966.2010.16439.x

Li, Q., Kilbinger, M., Luo, W., et al. 2024, Black-Hole-to-Halo Mass Relation From UNIONS Weak Lensing. https://arxiv.org/abs/2402.10740

Lin, Y.-T., Chen, K.-F., Chen, T.-C., Chuang, C.-Y., & Oguri, M. 2025, Evolution of Massive Red Galaxies in Clusters from z=1.0 to z=0.3. https://arxiv.org/abs/2503.13592

Mishra, P. K., Rana, D., & More, S. 2023, Stellar mass dependence of galaxy size-dark matter halo radius relation probed by Subaru-HSC survey weak lensing measurements. https://arxiv.org/abs/2301.04664

Narkedimilli, S., Amballa, V. S., Kumar, N. V. S., et al. 2025, Comparative Analysis of Black Hole Mass Estimation in Type-2 AGNs: Classical vs. Quantum Machine Learning and Deep Learning Approaches. https://arxiv.org/abs/2502.15297

Ni, S., Qiu, Y., Chen, Y., et al. 2024a, PI-AstroDeconv: A Physics-Informed Unsupervised Learning Method for Astronomical Image Deconvolution. https://arxiv.org/abs/2403.01692

Ni, Y., Chen, N., Zhou, Y., et al. 2024b, The Astrid Simulation: Evolution of black holes and galaxies to z=0.5 and different evolution pathways for galaxy quenching. https://arxiv.org/abs/2409.10666

Pacucci, F., & Loeb, A. 2024, The Redshift Evolution of the $M_\bullet - M_\star$ Relation for JWST's Supermassive Black Holes at $z > 4$, doi: https://doi.org/10.3847/1538-4357/ad3044

Privatus, P., & Goswami, U. D. 2025, Isolated and group environment dependence of stellar mass and different star formation rates, doi: https://doi.org/10.1142/S0217732324502365

Reines, A. E., & Volonteri, M. 2015, Relations Between Central Black Hole Mass and Total Galaxy Stellar Mass in the Local Universe, doi: https://doi.org/10.1088/0004-637X/813/2/82

Sanchez, J. N. L., Villa, E. M., Lopez, A. A. A., & Bravo, O. M. M. 2025, Estimating the mass of galactic components using machine learning algorithms. https://arxiv.org/abs/2403.06178

Sturm, M. R., & Reines, A. E. 2024, A Breakdown of the Black Hole - Bulge Mass Relation in Local Active Galaxies. https://arxiv.org/abs/2406.06675

Sweet, A. 2024, Predicting Exoplanetary Features with a Residual Model for Uniform and Gaussian Distributions. https://arxiv.org/abs/2406.10771

Torbaniuk, O., Paolillo, M., D'Abrusco, R., et al. 2023, Probing supermassive black hole growth and its dependence on stellar mass and star-formation rate in low-redshift galaxies, doi: https://doi.org/10.1093/mnras/stad3965

Tremmel, M., Ricarte, A., Natarajan, P., et al. 2024, An Enhanced Massive Black Hole Occupation Fraction Predicted in Cluster Dwarf Galaxies, doi: https://doi.org/10.33232/001c.116617

Wang, W., Yang, X., Jing, Y., et al. 2025, Luminosity and stellar mass functions of faint photometric satellites around spectroscopic central galaxies from DESI Year-1 Bright Galaxy Survey. https://arxiv.org/abs/2503.03317

Winkel, N., Bennert, V. N., Remigio, R. P., et al. 2024, Combining Direct Black Hole Mass Measurements and Spatially Resolved Stellar Kinematics to Calibrate the $M_{\rm BH}$-$\sigma_\star$ Relation of Active Galaxies. https://arxiv.org/abs/2411.02488

Zhang, Y., Ouchi, M., Gebhardt, K., et al. 2023, The Stellar Mass - Black Hole Mass Relation at $z \sim 2$ Down to $\mathcal{M}_{\rm BH} \sim 10^7 M_\odot$ Determined by HETDEX, doi: https://doi.org/10.3847/1538-4357/acc2c2