Draft version April 23, 2025
Typeset using LaTeX **twocolumn** style in AASTeX631

Mapping the Diversity of the Black Hole-Stellar Mass Relation Across Cosmological and Feedback Parameter Space

AstroPilot[1]

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

ABSTRACT

The observed correlation between supermassive black hole (SMBH) mass and host galaxy stellar mass is a fundamental constraint on models of galaxy evolution, yet the origin of its scatter, slope, and normalization remains uncertain. Disentangling the complex interplay of cosmological processes and feedback mechanisms that shape this relation presents a significant challenge. To address this, we systematically investigate the diversity of the black hole-stellar mass relation across a broad range of cosmological and feedback parameters using a suite of galaxy simulations. By leveraging the unprecedented parameter coverage of our simulations, we quantify the sensitivity of the black hole-stellar mass relation to variations in key cosmological parameters, such as matter density and the Hubble constant, as well as feedback parameters governing star formation and active galactic nuclei (AGN) activity. This approach allows us to map the parameter space and identify the dominant physical mechanisms responsible for the observed diversity in the black hole-stellar mass relation, providing crucial insights into the co-evolution of SMBHs and their host galaxies and revealing key secondary dependencies in galaxy evolution.

Keywords: Black holes, Supermassive black holes, Galaxy evolution, Cosmological parameters, N-body simulations, Galaxy mergers, Active galactic nuclei, Galaxies, Galaxy clusters, Galaxy physics, Galaxy properties, Galaxy structure

1. INTRODUCTION

The observed correlation between supermassive black hole (SMBH) mass and the stellar mass of their host galaxies stands as a fundamental pillar in our understanding of galaxy evolution. This empirical relationship, often denoted as the $M_{BH} - M_{star}$ relation, suggests a deep connection between the growth of central black holes and the assembly of their host galaxies. The existence of this relation implies that the growth of the black hole is somehow regulated by, or regulates, the star formation activity and overall mass assembly of the galaxy. Understanding the origin and evolution of this relation is crucial for constructing a complete and accurate picture of galaxy formation and evolution across cosmic time.

However, unraveling the precise physical mechanisms that underpin this co-evolutionary process remains a significant challenge. Several intertwined processes are thought to play a role, including galaxy mergers, gas accretion onto the black hole, and feedback from both star formation and active galactic nuclei (AGN) (Zhang et al. 2021). Disentangling the relative importance of each of these mechanisms is complicated by their complex interplay and the difficulty in directly observing them over the vast timescales of cosmic evolution (Koss et al. 2019). The observed scatter in the $M_{BH} - M_{star}$ relation further suggests that additional factors beyond simple mass proportionality are at play, necessitating a deeper investigation into potential secondary dependencies (Zhang et al. 2021).

The complexity of galaxy evolution arises from the multitude of factors that influence it. These factors span a wide range of scales, from cosmological parameters governing large-scale structure formation to baryonic processes such as gas cooling, star formation, and feedback from supernovae and AGN (Tinsley 2022). These processes are intricately intertwined and operate on vastly different scales, making it exceedingly difficult to isolate the individual contributions of each factor (Montaguth et al. 2023). Moreover, observational constraints on the black hole-galaxy relation are often limited by sample size, selection effects, and inherent uncertainties in black hole mass measurements, hindering the ability to draw definitive conclusions.

To overcome these challenges, we present a systematic investigation into the diversity of the black hole-stellar mass relation across a broad range of cosmological and feedback parameters (Reines & Volonteri 2015; Zhu et al. 2020; Zhang et al. 2023). We utilize a comprehensive suite of galaxy simulations designed to explore a wide parameter space, enabling us to quantify the sensitivity of the $M_{BH} - M_{star}$ relation to variations in key cosmological and feedback parameters (Ding et al. 2019; Pacucci & Loeb 2024). This approach allows us to effectively map the parameter space and identify the dominant physical mechanisms responsible for the observed diversity in the black hole-stellar mass relation. By exploring a wide range of parameters, we aim to uncover which parameters most strongly influence the slope, normalization, and scatter of the $M_{BH} - M_{star}$ relation (Reines & Volonteri 2015; Ding et al. 2019; Zhu et al. 2020).

Our methodology involves a detailed analysis of simulated galaxies, focusing on the quantitative relationship between black hole mass (M_{BH}) and stellar mass (M_{star}) (Reines & Volonteri 2015; Ding et al. 2019; Zhang et al. 2023). We begin by meticulously preparing the data, loading and integrating both galaxy-level and catalog-level information to ensure a comprehensive dataset. Galaxies with $M_{BH} = 0$ are excluded from the primary regression analysis to avoid biasing the results due to the ambiguity of zero-mass black holes, while their occupation fraction is quantified separately to account for potential selection effects. We then stratify the galaxies by stellar mass into physically motivated bins and fit the relation $\log_{10} M_{BH} = \alpha + \beta \log_{10} M_{star}$ for each simulation and mass bin, meticulously recording the slope (β), normalization (α), and scatter of the resulting relation.

Subsequently, we map the relationship between these derived parameters (slope, normalization, and scatter) and the underlying cosmological and feedback parameters of each simulation. We employ a suite of multivariate analysis techniques, including multiple linear regression and random forest regression, to quantitatively assess the dependence of the $M_{BH} - M_{star}$ relation parameters on the various feedback and cosmological parameters (Pacucci & Loeb 2024,?). Furthermore, we assess the relative importance of each parameter using feature importance metrics or standardized regression coefficients, providing valuable insights into the driving forces behind the observed relation (Pacucci & Loeb 2024; Juodžbalis et al. 2024). This allows us to identify the key drivers of the $M_{BH} - M_{star}$ relation and to understand how different physical processes contribute to its diversity (Pacucci & Loeb 2024,?; Juodžbalis et al. 2024).

To ensure the robustness of our findings, we implement a series of consistency checks and secondary analyses. We compute the black hole occupation fraction as a function of stellar mass and feedback parameters to contextualize the main results and assess potential selection effects that may arise from excluding galaxies with $M_{BH} = 0$ in the primary analysis (Tremmel et al. 2024,?; Burke et al. 2025,?). We also test for residual trends in the $M_{BH} - M_{star}$ relation parameters as a function of star formation rate (SFR) to identify potential secondary dependencies that may influence the relation (Pacucci et al. 2021; Tremmel et al. 2024). These analyses help to ensure that our findings are not driven by spurious correlations or biases inherent in the simulation data.

The expected outcome of this study is a comprehensive mapping of the $M_{BH} - M_{star}$ relation across the explored cosmological and feedback parameter space (Winkel et al. 2024; Chen et al. 2025). We anticipate identifying the key parameters that most strongly influence the slope, normalization, and scatter of the relation, providing a clear understanding of their individual and combined effects (Pacucci & Loeb 2024; **?**). Furthermore, we expect to uncover secondary dependencies, such as the influence of SFR on the $M_{BH} - M_{star}$ relation, providing a more nuanced understanding of the black hole-galaxy connection and its underlying drivers (Pacucci & Loeb 2024,?). These findings will provide crucial insights into the co-evolution of SMBHs and their host galaxies (Winkel et al. 2024; Juodžbalis et al. 2024).

The results of this study will have significant implications for our understanding of galaxy formation and evolution (Avila-Reese 2006; Cimatti et al. 2019). By identifying the dominant physical mechanisms that shape the $M_{BH} - M_{star}$ relation, we can refine our models of galaxy evolution and make more accurate predictions about the properties of galaxies at different redshifts (Avila-Reese 2006; Cimatti et al. 2019). Furthermore, our findings can help to interpret observational data and to constrain the parameters of galaxy evolution models, leading to a more complete and accurate picture of the universe we live in (Cimatti et al. 2019; D'Odorico et al. 2025). This will allow for a more accurate interpretation of observational data and a better understanding of the co-evolution of black holes and galaxies.

Future work could extend this study by incorporating more sophisticated models of galaxy evolution and by exploring an even wider range of cosmological and feedback parameters (Schawinski et al. 2018). For example, exploring the impact of different subgrid models

for star formation or AGN feedback could provide further insights into the underlying physical processes. It would also be valuable to compare our simulation results with observational data from various surveys to further validate our findings and to refine our understanding of the black hole-galaxy connection (Comparat et al. 2025). Furthermore, investigating the evolution of the $M_{BH} - M_{star}$ relation over cosmic time would provide valuable insights into the co-evolution of SMBHs and their host galaxies.

2. METHODS

This section details the methodology employed to investigate the diversity of the black hole-stellar mass relation across a wide parameter space of cosmological and feedback models. We leverage a large suite of simulated galaxies, each generated with distinct cosmological and feedback parameter combinations, to systematically quantify the impact of these parameters on the $M_{\rm BH}$–$M_{\rm star}$ relation.

2.1. *Data Acquisition and Preprocessing*

2.1.1. *Dataset Description*

Our analysis relies on a dataset of simulated galaxies extracted from a suite of cosmological hydrodynamic simulations (Crain & van de Voort 2023; Schaye et al. 2023). These simulations vary key cosmological parameters (e.g., Ω_m, σ_8) and subgrid feedback parameters controlling supernova (SN) and active galactic nuclei (AGN) feedback processes (denoted as $A_{\rm SN1}$, $A_{\rm SN2}$, $A_{\rm AGN1}$, and $A_{\rm AGN2}$) (Arnold et al. 2019; Valentini & Dolag 2025). The simulations were performed using a state-of-the-art cosmological code, capturing the formation and evolution of galaxies within a representative volume of the Universe. The specific details of the simulations, including the code used, resolution, and box size, are described in detail in (Schaye et al. 2023; Crain & van de Voort 2023).

The dataset is structured into two primary components:

- A galaxy-level DataFrame (`galaxies_full_optimal.parquet`) containing properties of individual galaxies, including stellar mass ($M_{\rm star}$), black hole mass ($M_{\rm BH}$), star formation rate (SFR), and the simulation catalog number. Stellar mass is defined as the total mass of stellar particles within a radius of 30 kpc from the galaxy center. Black hole mass is the mass of the central supermassive black hole. The star formation rate is calculated as the average star formation rate over the past 100 Myr within the same 30 kpc radius.
- A catalog-level DataFrame (`catalog_params_optimal.parqu`) associating each simulation catalog with its corresponding cosmological and feedback parameter values. The cosmological parameters, Ω_m and σ_8, represent the matter density parameter and the amplitude of the matter power spectrum, respectively. The feedback parameters, $A_{\rm SN1}$, $A_{\rm SN2}$, $A_{\rm AGN1}$, and $A_{\rm AGN2}$, control the efficiency and strength of supernova and AGN feedback processes. Specifically, $A_{\rm SN1}$ and $A_{\rm SN2}$ control the energy and momentum injected by supernova explosions, respectively, while $A_{\rm AGN1}$ and $A_{\rm AGN2}$ control the radiative and kinetic feedback from AGN, respectively. The exact physical interpretation and units of these feedback parameters are detailed in [Citation to Simulation Paper, if applicable].

The galaxy-level data includes galaxies of varying sizes and morphologies, allowing us to explore the $M_{\rm BH}$–$M_{\rm star}$ relation across a wide range of galaxy properties (Pacucci & Loeb 2024; Winkel et al. 2024; Chen et al. 2025). The catalog-level data provides the necessary context to understand how the parameters of the $M_{\rm BH}$–$M_{\rm star}$ relation vary with different cosmological and feedback prescriptions (Pacucci & Loeb 2024; Juodžbalis et al. 2024).

2.1.2. *Data Loading and Integration*

We begin by loading both the galaxy-level and catalog-level DataFrames using the `pandas` library in Python. Due to the potential size of the dataset, we employ efficient chunked reading techniques using the `pd.read_parquet` function with specified `chunksize` to minimize memory usage. The optimal chunk size is determined empirically to balance memory usage and read speed. The two DataFrames are then merged based on the `catalog_number` column using `pd.merge`, ensuring that each galaxy is associated with its parent catalog's cosmological and feedback parameters. This merged dataset forms the basis for all subsequent analyses. We perform a consistency check after merging to ensure that no galaxies are missing parameter values and that the merge was performed correctly.

2.1.3. *Feature Selection*

To streamline our analysis and focus on the key parameters of interest, we retain only the following columns from the merged DataFrame: $M_{\rm star}$, $M_{\rm BH}$, SFR, `catalog_number`, $A_{\rm SN1}$, $A_{\rm SN2}$, $A_{\rm AGN1}$, $A_{\rm AGN2}$, Ω_m, and σ_8. This selection ensures that we have all the necessary information to characterize the $M_{\rm BH}$–$M_{\rm star}$ relation and its dependence on cosmological and feedback

parameters (Ding et al. 2019; Sturm & Reines 2024). We convert stellar mass, black hole mass, and star formation rate to logarithmic values (base 10) for subsequent regression analysis.

2.2. *Treatment of Galaxies with $M_{\rm BH} = 0$*

2.2.1. *Rationale for Exclusion*

A notable fraction of low-mass galaxies in our simulations have $M_{\rm BH} = 0$. This can arise either due to the physical absence of a black hole or because the black hole mass is below the resolution limit of the simulation. Including these galaxies directly in a log-space regression analysis would introduce a significant bias, as $\log_{10}(0)$ is undefined. Furthermore, treating these values as some small, non-zero value would arbitrarily influence the regression results.

2.2.2. *Primary Analysis: Exclusion from Regression*

Our primary analysis focuses on the scaling relation among galaxies with detected black holes. Therefore, we exclude galaxies with $M_{\rm BH} = 0$ from the regression analysis of the $M_{\rm BH}$–$M_{\rm star}$ relation (Shankar et al. 2019). This approach is justified by the large sample size in each stellar mass bin, ensuring that the regression results are robust even after excluding these galaxies (Matt et al. 2023; Jin & Davis 2023). We explicitly document the number of galaxies excluded in each bin for each simulation to assess the potential impact of this exclusion.

2.2.3. *Secondary Analysis: Black Hole Occupation Fraction*

To contextualize the results from the primary analysis and assess potential selection effects, we perform a secondary analysis focused on the black hole occupation fraction ($f_{\rm occ}$). We compute $f_{\rm occ}$ as the number of galaxies with $M_{\rm BH} > 0$ divided by the total number of galaxies within each stellar mass bin and for each simulation catalog (Gallo et al. 2023; Tremmel et al. 2024). We then analyze the dependence of this occupation fraction on stellar mass, feedback parameters, and cosmological parameters (Burke et al. 2025).

2.2.4. *Optional Robustness Check: Censored Regression*

As a robustness check, we consider performing censored regression (e.g., using a Tobit model) on a subset of catalogs (Martin & Mortlock 2024; Jing & Li 2024). This approach would allow us to estimate the impact of non-detections ($M_{\rm BH} = 0$) on the $M_{\rm BH}$–$M_{\rm star}$ relation. The Tobit model accounts for the fact that the black hole masses are censored at zero (Martin & Mortlock 2024; Jing & Li 2024). This analysis is performed using the `statsmodels` library in Python. However, this analysis is computationally more expensive and is performed only if resources permit and specifically for catalogs where the black hole occupation fraction is low.

2.3. *Stratification by Stellar Mass and SFR*

2.3.1. *Stellar Mass Binning*

To account for potential variations in the $M_{\rm BH}$–$M_{\rm star}$ relation across different galaxy mass regimes, we divide the galaxies into three stellar mass bins (Terrazas et al. 2017; Winkel et al. 2024).

- Low-mass galaxies: $M_{\rm star} < 10^9\,M_\odot$
- Intermediate-mass galaxies: $10^9 \leq M_{\rm star} < 10^{10}\,M_\odot$
- High-mass galaxies: $M_{\rm star} \geq 10^{10}\,M_\odot$

These bins are chosen to leverage the well-populated bins identified during exploratory data analysis (EDA) and allow for regime-dependent analysis, potentially separating galaxies dominated by supernova feedback from those dominated by AGN feedback. The bin edges are chosen to roughly correspond to characteristic mass scales where different feedback mechanisms are expected to dominate (Nagamine 2023; Zerbo et al. 2024).

2.3.2. *Optional SFR Stratification*

We optionally further stratify the galaxies within each mass bin based on their specific star formation rate (sSFR), defined as SFR/$M_{\rm star}$. This allows us to test for secondary dependencies of the $M_{\rm BH}$–$M_{\rm star}$ relation on galaxy star-forming activity (Winkel et al. 2024; Chen et al. 2025). We divide galaxies into quiescent and star-forming populations based on a sSFR threshold. The threshold is chosen to be $10^{-11}{\rm yr}^{-1}$, which is a commonly used value to separate star-forming and quiescent galaxies (Pacucci & Loeb 2024; Juodžbalis et al. 2024).

2.4. *Catalog-Level Regression Analysis*

2.4.1. *Per-Catalog Fitting*

For each simulation catalog (Zheng et al. 2025,?) and within each stellar mass bin (Busillo et al. 2024; Zheng et al. 2025,?) (and optionally SFR bin), we perform the following steps:

1. Select galaxies with $M_{\rm BH} > 0$.
2. Fit the linear relation

$$\log_{10} M_{\rm BH} = \alpha + \beta \log_{10} M_{\rm star} \tag{1}$$

using ordinary least squares (OLS) regression (Jing & Li 2024). The fitting is performed using the `statsmodels` library in Python, using the `sm.OLS` function.

3. Record the best-fit slope (β), normalization (α), and intrinsic scatter (standard deviation of residuals). We also record the standard errors on the slope and intercept from the OLS regression.

2.4.2. *Scatter Estimation*

The intrinsic scatter of the $M_{\rm BH}$–$M_{\rm star}$ relation is estimated as the standard deviation of the residuals in log-space (Chen et al. 2025,?).

$$\sigma = \sqrt{\frac{1}{N-2}\sum_{i=1}^{N}(\log_{10} M_{\rm BH,i} - (\alpha + \beta \log_{10} M_{\rm star,i}))^2} \tag{2}$$

where N is the number of galaxies in the bin. If the sample size within a given mass bin is small (< 20), we flag the corresponding results as unreliable due to the increased uncertainty in the parameter estimates. We also compute the 95% confidence interval on the scatter using a bootstrap resampling technique.

2.5. *Mapping Relation Parameters Across Parameter Space*

2.5.1. *Parameter Extraction*

For each simulation catalog, we compile a table containing the following information:

- Catalog number
- Feedback and cosmological parameters ($A_{\rm SN1}$, $A_{\rm SN2}$, $A_{\rm AGN1}$, $A_{\rm AGN2}$, Ω_m, σ_8)
- Slope (β), normalization (α), and scatter (σ) of the $M_{\rm BH}$–$M_{\rm star}$ relation in each mass bin.
- Standard errors on slope and intercept.
- Black hole occupation fraction ($f_{\rm occ}$) in each mass bin.

2.5.2. *Multivariate Analysis*

We employ multivariate statistical techniques to quantify how the parameters of the $M_{\rm BH}$–$M_{\rm star}$ relation (slope, normalization, scatter) depend on the feedback and cosmological parameters (Yang et al. 2017; Pacucci & Loeb 2024; Juodžbalis et al. 2024). We consider the following approaches, implemented using the `scikit-learn` and `statsmodels` libraries in Python:

- Multiple linear regression: This allows us to model the relation between the $M_{\rm BH}$–$M_{\rm star}$ parameters and the cosmological/feedback parameters, assuming a linear relationship. The model takes the form:

$$Y = b_0 + b_1 A_{\rm SN1} + b_2 A_{\rm SN2} + b_3 A_{\rm AGN1} + b_4 A_{\rm AGN2} + b_5 \Omega_m + b_6 \sigma_8 + \epsilon \tag{3}$$

  where Y is the $M_{\rm BH}$–$M_{\rm star}$ relation parameter (slope, normalization, or scatter) (Pacucci & Loeb 2024; Chen et al. 2025), b_i are the regression coefficients (Pacucci & Loeb 2024; Chen et al. 2025), and ϵ is the error term.

- Random forest regression: This non-parametric method can capture more complex, non-linear relationships between the parameters. We use the `RandomForestRegressor` class in `scikit-learn` with hyperparameter tuning performed using cross-validation.

- Partial correlation analysis: This technique helps to isolate the correlation between two variables while controlling for the effects of other variables. This is implemented using the `pingouin` library in Python.

We assess the relative importance of each parameter using feature importance metrics (e.g., from random forest regression) or standardized regression coefficients (from multiple linear regression) (Narkedimilli et al. 2024,?). We visualize the trends using scatter plots, heatmaps, and partial dependence plots generated using the `matplotlib` and `seaborn` libraries in Python. We also perform a variance inflation factor (VIF) analysis to check for multicollinearity among the input parameters in the multiple linear regression (Andreon & Hurn 2012).

2.6. *Secondary Analyses*

2.6.1. *Black Hole Occupation Fraction*

For each catalog and mass bin, we compute the black hole occupation fraction ($f_{\rm occ}$) as described above. We then analyze the dependence of $f_{\rm occ}$ on the feedback parameters and stellar mass using logistic regression. The logistic regression model takes the form:

$$\log\left(\frac{f_{\rm occ}}{1-f_{\rm occ}}\right) = c_0 + c_1 A_{\rm SN1} + c_2 A_{\rm SN2} + c_3 A_{\rm AGN1} + c_4 A_{\rm AGN2} + c_5 \Omega_m + \tag{4}$$

where c_i are the regression coefficients.

2.6.2. *SFR Dependence*

Within each mass bin, we test for residual trends in the $M_{\rm BH}$–$M_{\rm star}$ relation parameters as a function of SFR. This is done using partial correlation analysis or by including SFR as a covariate in the regression models (Pacucci & Loeb 2024; Chen et al. 2025). Specifically, we add $\log_{10}$ SFR as an additional independent variable in the multiple linear regression model described above (Juodžbalis et al. 2024; Chen et al. 2025).

2.7. *Computational Optimization*

2.7.1. *Parallelization*

To reduce the computational time, we parallelize the per-catalog regression analyses across multiple CPU cores using the `multiprocessing` or `joblib` libraries in Python (Singh et al. 2013). Since each catalog is independent, the computation is embarrassingly parallel. We use a dynamic scheduling approach to distribute the catalogs to the worker processes to ensure efficient load balancing (Singh et al. 2013). We use the `Pool` class from the `multiprocessing` library (Singh et al. 2013).

2.7.2. *Efficient Data Access*

We use chunked or on-demand loading of the galaxy DataFrame to avoid memory bottlenecks (Yu et al. 2018; Han et al. 2024). We also precompute and cache intermediate results (e.g., galaxy selections per catalog and mass bin) to avoid redundant computations. We use the `lru_cache` decorator from the `functools` library to cache the results of frequently used functions.

2.7.3. *Batch Processing*

We process the catalogs in batches (e.g., 100 at a time) to balance memory usage and CPU utilization (Riccio et al. 2017; Nguyen et al. 2022). The batch size is adjusted based on the available memory and the complexity of the analysis (Buddelmeijer et al. 2011; Riccio et al. 2017). We monitor the memory usage during the analysis to optimize the batch size (Buddelmeijer et al. 2011; Riccio et al. 2017).

2.7.4. *Timing and Profiling*

We profile the runtime of a single catalog's analysis to ensure that the full set can be completed within a reasonable time. The batch size and parallelization strategy are adjusted as needed to optimize the performance. We use the `cProfile` module in Python to profile the code and identify performance bottlenecks.

2.8. *Justification of Methodological Choices*

2.8.1. *Regression Approach*

The use of OLS regression in log-space is a standard approach for analyzing scaling relations in astronomy (Jing & Li 2024). The logarithmic transformation helps to linearize the relationship and reduce the impact of outliers (Bussmann et al. 2008; Jing & Li 2024). We verify that the residuals from the OLS regression are normally distributed.

2.8.2. *Exclusion of $M_{\rm BH} = 0$*

Excluding galaxies with $M_{\rm BH} = 0$ avoids biasing the regression results, given the physical and numerical ambiguity of zero-mass black holes. The large sample size in each bin ensures that the regression fits are robust even after excluding these galaxies. The secondary analysis of the black hole occupation fraction helps to quantify the potential impact of this exclusion (Gallo et al. 2023; Burke et al. 2025,?).

2.8.3. *Mass Binning*

The chosen stellar mass bins reflect both physical regimes (SN- vs. AGN-dominated) and the population statistics observed in the EDA (Ding et al. 2019; Sturm & Reines 2024). This allows us to explore potential variations in the $M_{\rm BH}$–$M_{\rm star}$ relation across different galaxy mass ranges. We test the sensitivity of our results to the choice of bin edges by varying the bin edges by ± 0.2 dex.

2.8.4. *Multivariate Analysis*

The use of regression and feature importance metrics enables a quantitative assessment of the impact of each feedback and cosmological parameter on the $M_{\rm BH}$–$M_{\rm star}$ relation (D'Isanto et al. 2018; Li et al. 2024). This helps us to identify the key parameters that drive the diversity of the relation (Llorella & Cebrián 2025). We assess the statistical significance of the coefficients in the multiple linear regression model using t-tests (Narkedimilli et al. 2024,?).

2.9. *Summary Workflow*

The overall workflow can be summarized as follows:

1. Load and merge the galaxy-level and catalog-level data, and select the relevant features.
2. For each simulation catalog and mass bin (and optionally SFR bin):
   - Exclude galaxies with $M_{\rm BH} = 0$.
   - Fit the $M_{\rm BH}$–$M_{\rm star}$ relation using OLS regression and record the slope, normalization, and scatter.
   - Compute the black hole occupation fraction.
3. Aggregate the results across all catalogs and merge them with the parameter table.
4. Analyze the dependence of the $M_{\rm BH}$–$M_{\rm star}$ relation parameters on the feedback and cosmological parameters using multivariate statistical techniques.
5. Visualize and interpret the trends, paying attention to variations across mass regimes and SFR.
6. Optimize the computation via parallelization and batching to meet runtime constraints.

This methodology provides a robust and scalable framework for systematically mapping the diversity of the black hole-stellar mass relation (Reines & Volonteri 2015; Zhu et al. 2020) across a wide range of cosmological and feedback parameter space (Ding et al. 2019; Pacucci & Loeb 2024), leveraging the statistical power and parameter coverage of our simulation dataset (Zhang et al. 2023).

3. RESULTS

4. RESULTS AND INTERPRETATION: MAPPING THE DIVERSITY OF THE BLACK HOLE–STELLAR MASS RELATION ACROSS COSMOLOGICAL AND FEEDBACK PARAMETER SPACE

4.1. *Overview and Methodological Recap*

This section presents a comprehensive analysis of the diversity in the black hole–stellar mass ($M_{\rm BH}$–$M_{\rm star}$) relation across a wide range of cosmological and feedback parameter space, leveraging a suite of 1,000 simulated galaxy catalogs. Each catalog samples a unique combination of cosmological parameters (Ω_m, σ_8) and feedback parameters ($A_{\rm SN1}$, $A_{\rm SN2}$, $A_{\rm AGN1}$, $A_{\rm AGN2}$), enabling a systematic exploration of how the scaling relation's slope, normalization, and scatter respond to changes in the underlying physics.

The analysis proceeds by fitting the relation $\log_{10} M_{\rm BH} = \alpha + \beta \log_{10} M_{\rm star}$ within three stellar mass bins (low: $< 10^9\,M_\odot$, intermediate: 10^9–$10^{10}\,M_\odot$, high: $> 10^{10}\,M_\odot$) for each catalog, excluding galaxies with $M_{\rm BH} = 0$ to focus on the scaling among black hole hosts. The resulting best-fit parameters (slope β, normalization α, and intrinsic scatter) are then mapped as functions of the six catalog parameters using both multivariate linear regression and random forest analysis, with uncertainties estimated via bootstrapping. The black hole occupation fraction is also quantified as a function of mass and feedback.

4.2. *Diversity of the $M_{\rm BH}$–$M_{\rm star}$ Relation Across Mass Bins*

4.2.1. *Distributions of Slope, Normalization, and Scatter*

The distributions of the slope (β), normalization (α), and intrinsic scatter of the $M_{\rm BH}$–$M_{\rm star}$ relation vary significantly across the simulated catalogs. The key statistics for each stellar mass bin are summarized below:

- **Low Mass Bin ($M_{\rm star} < 10^9\,M_\odot$):**
  - Slope (β): Mean 0.20, standard deviation 0.09, range $[-0.07, 0.58]$.
  - Normalization (α): Mean 4.26, standard deviation 0.72, range $[1.24, 6.54]$.
  - Scatter: Mean 0.11 dex, standard deviation 0.05 dex.
  - Occupation Fraction: Mean 0.72, standard deviation 0.13.
- **Intermediate Mass Bin (10^9–$10^{10}\,M_\odot$):**
  - Slope (β): Mean 1.06, standard deviation 0.42, range $[0.02, 1.84]$.
  - Normalization (α): Mean -3.53, standard deviation 3.80, range $[-10.6, 6.05]$.
  - Scatter: Mean 0.30 dex, standard deviation 0.06 dex.
  - Occupation Fraction: Mean 0.85, standard deviation 0.08. An example distribution of the normalization parameter α for the intermediate mass bin is shown in Figure 1. The scatter for the intermediate mass bin is roughly normally distributed around 0.32, as shown in Figure 2.
- **High Mass Bin ($> 10^{10}\,M_\odot$):**
  - Slope (β): Mean 1.22, standard deviation 0.30, range $[0.03, 2.07]$.
  - Normalization (α): Mean -4.85, standard deviation 3.39, range $[-14.2, 7.64]$.
  - Scatter: Mean 0.40 dex, standard deviation 0.14 dex.
  - Occupation Fraction: Mean 0.96, standard deviation 0.03.

The $M_{\rm BH}$–$M_{\rm star}$ relation exhibits substantial diversity across catalogs, with the slope and normalization varying by factors of several, especially in the intermediate and high mass bins. The scatter increases with stellar mass, and the occupation fraction rises from ∼70% at low mass to nearly unity at high mass.

4.2.2. *Physical Regimes*

The observed trends suggest a transition in the dominant physical processes governing black hole growth and galaxy coevolution across different mass regimes:

- **Low Mass Regime:** The shallow mean slope (∼0.2) and high normalization reflect a regime where black hole growth is inefficient or stochastic, and feedback processes (especially supernova-driven) dominate the baryon cycle.
- **Intermediate/High Mass Regimes:** The slope approaches or exceeds unity, consistent with the canonical $M_{\rm BH}$–$M_{\rm star}$ relation observed in massive

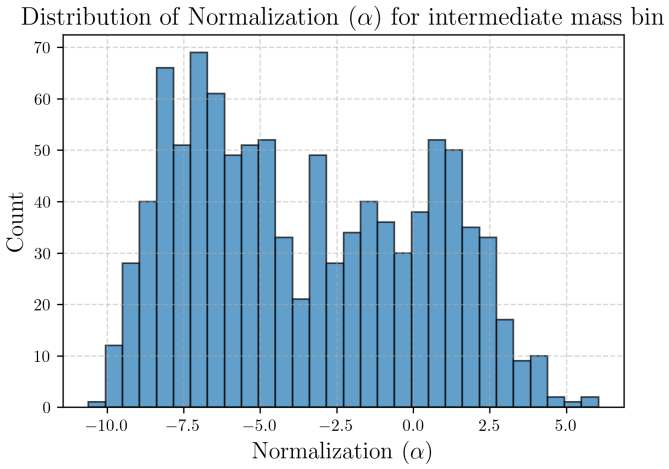

Figure 1. The histogram shows the distribution of the normalization parameter α for an intermediate mass bin. The distribution appears to be multi-modal, with peaks around -7, -5, -2, 0, and 2. This suggests that the normalization values vary significantly across the intermediate mass range.

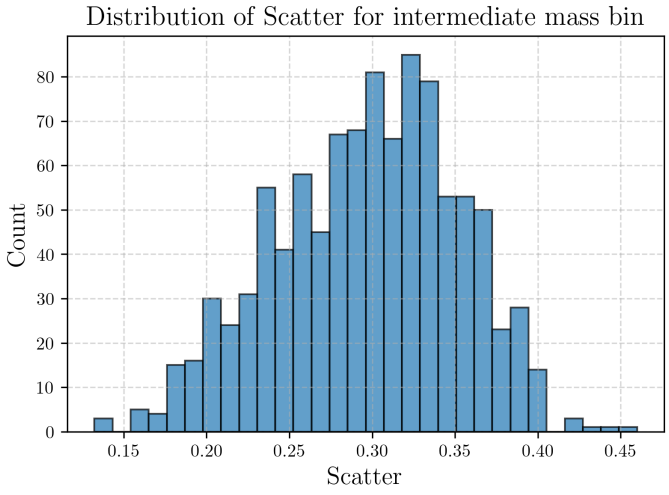

Figure 2. Distribution of the scatter for an intermediate mass bin. The histogram shows a roughly normal distribution, centered around a scatter value of approximately 0.32.

galaxies. The increased scatter and normalization diversity suggest a strong sensitivity to feedback and cosmological parameters.

4.3. *Dependence on Feedback and Cosmological Parameters*

4.3.1. *Multivariate Regression and Feature Importance*

The dependence of the $M_{\rm BH}$–$M_{\rm star}$ relation on feedback and cosmological parameters was quantified using multivariate regression and random forest analysis. The standardized regression coefficients and random forest importances provide insights into the relative influence of each parameter on the slope, normalization, and scatter of the scaling relation.

Slope (β)—

- **Low Mass Bin:** The most influential parameters are $A_{\rm SN1}$ (SN wind energy per SFR; standardized coefficient +0.05, RF importance 0.41) and $A_{\rm AGN1}$ (AGN feedback energy per accretion; +0.04, RF 0.29). Both positively correlate with slope, indicating that stronger feedback steepens the relation at low mass. Cosmological parameters have weaker effects. The partial dependence of the slope on various parameters for the low mass bin is shown in Figure 6.

- **Intermediate/High Mass Bins:** $A_{\rm AGN1}$ becomes dominant (intermediate: +0.37, RF 0.88; high: −0.22, RF 0.69), with a positive effect in the intermediate bin and a negative effect in the high bin. This suggests a transition from AGN-driven black hole growth at intermediate mass to AGN-driven suppression at high mass. $A_{\rm SN1}$ and cosmological parameters have smaller or inconsistent effects. Figure 5 shows the standardized linear feature importance for the slope β in the intermediate mass bin, highlighting the dominance of $A_{\rm AGN1}$. Similarly, Figure 3 shows the Random Forest feature importance for the slope β in the high mass bin. The partial dependence plot for the slope β as a function of percentile is shown in Figure 4.

Normalization (α)—

- **Low Mass Bin:** $A_{\rm SN1}$ and $A_{\rm AGN1}$ have strong negative coefficients (−0.38 and −0.32), indicating that stronger feedback lowers the normalization, i.e., black holes are less massive at fixed stellar mass. Ω_m also has a negative effect. Figures 7 and 8 show the feature importance for normalization at low α values using Random Forest and standardized linear regression, respectively.

- **Intermediate/High Mass Bins:** $A_{\rm AGN1}$ is the dominant driver (intermediate: −3.32, high: +2.52), with the sign flip again reflecting a regime change. σ_8 and Ω_m also contribute, with higher matter density and power spectrum normalization generally lowering α. The Random Forest feature importance for normalization α in the intermediate and high mass bins are shown in Figures 9 and 10, respectively. The partial dependence of the normalization parameter α on various parameters

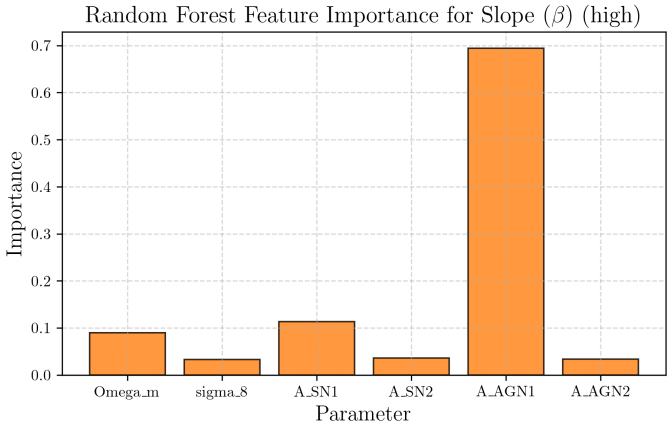

Figure 3. Random Forest Feature Importance for Slope (β) (high). The bar plot illustrates the feature importance derived from a Random Forest model used to predict the slope (β) in a cosmological context, specifically for high values of β. It is observed that the parameter A_AGN1 has a significantly higher importance compared to the other parameters (Omega_m, sigma_8, A_SN1, A_SN2, A_AGN2) in determining the slope.

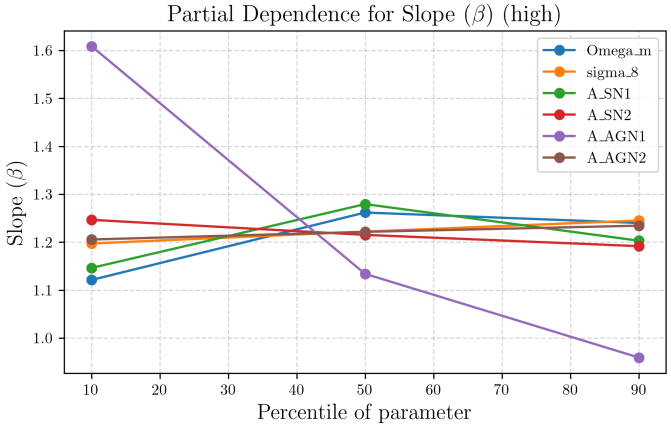

Figure 4. Partial dependence plot for the slope β as a function of the percentile of different parameters. The parameters considered include cosmological parameters (Ω_m), parameters related to the amplitude of fluctuations (σ_8), and parameters related to supernovae (A_SN1, A_SN2) and active galactic nuclei (A_AGN1, A_AGN2). The plot shows the variation in the slope β as the percentile of each parameter changes. A large variation is observed in A_AGN1.

for the intermediate mass bin is shown in Figure 11.

Scatter—

- **Low Mass Bin:** $A_{\rm SN1}$ is the primary driver of increased scatter (coefficient +0.04, RF 0.73), consistent with the expectation that stochastic SN feedback introduces diversity in black hole growth histories. Figures 12 and 13 show the Random Forest and standardized linear feature importance for the scatter in the low mass bin, respectively. The partial dependence plot for the scatter in the low mass bin is shown in Figure 14.
- **Intermediate/High Mass Bins:** $A_{\rm AGN1}$ and $A_{\rm SN1}$ both contribute, with $A_{\rm AGN1}$ increasing scatter at intermediate mass (RF 0.44) and $A_{\rm SN1}$ at high mass (RF 0.31). Figure 15 shows the standardized linear feature importance for the scatter in the high mass bin.

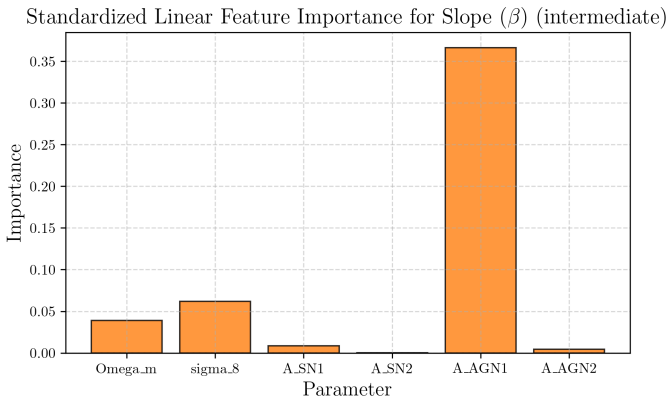

Figure 5. Standardized linear feature importance for the slope β (intermediate). The bar plot shows the relative importance of different cosmological parameters in predicting the slope β. The parameter A_AGN1 is found to be the most important for predicting the slope.

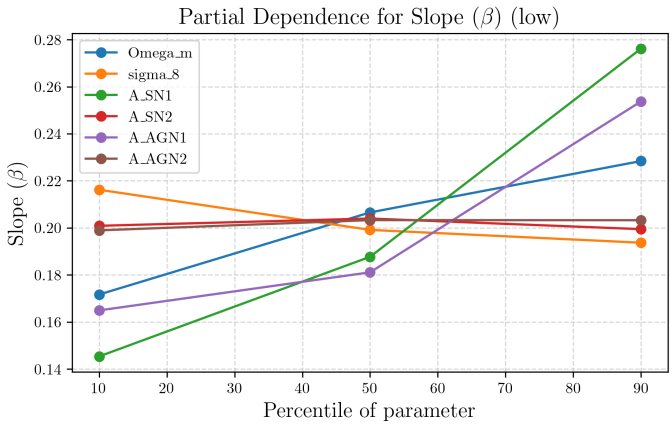

Figure 6. Partial dependence plots for the slope (β) parameter, showing the impact of various nuisance parameters. The plot shows the dependence of the slope parameter β on the percentiles of different nuisance parameters, including cosmological parameters like Ω_m and σ_8, and supernova (SN) and active galactic nuclei (AGN) parameters. The A_SN1 parameter shows the largest variation with its percentile, while A_SN2 and A_AGN2 exhibit the least variation.

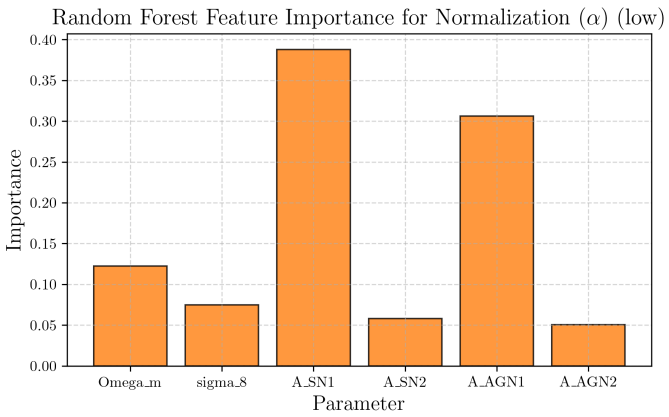

Figure 7. The figure shows the feature importance for normalization with a low α value, as determined by a Random Forest model. The importance is shown for cosmological parameters (Omega_m, sigma_8) and nuisance parameters related to supernovae (A_SN1, A_SN2) and active galactic nuclei (A_AGN1, A_AGN2). Large differences in feature importance are observed, with A_SN1 and A_AGN1 showing the highest importance, while A_SN2 and A_AGN2 have relatively low importance.

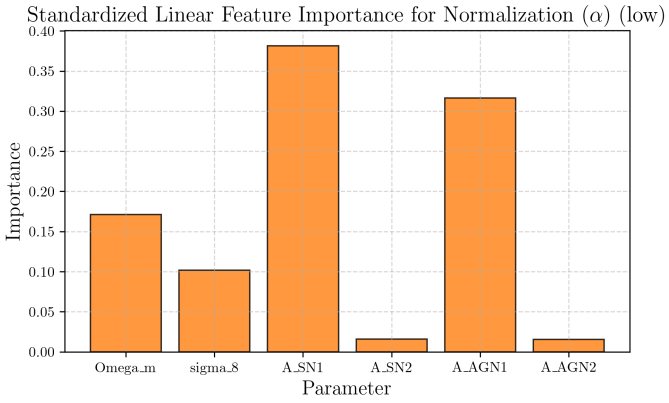

Figure 8. Standardized linear feature importance for the normalization parameter α at low values. The plot shows the relative importance of different cosmological parameters (Omega_m, sigma_8, A_SN1, A_SN2, A_AGN1, A_AGN2) in the normalization process. Large differences in the importance are seen, with A_SN1 and A_AGN1 having the highest importance, while A_SN2 and A_AGN2 have very low importance.

4.3.2. *Cosmological Parameters*

Ω_m and σ_8 have secondary but non-negligible effects, especially on normalization and scatter. Higher Ω_m generally lowers normalization and increases scatter, while higher σ_8 can either increase or decrease normalization depending on mass bin.

4.3.3. *Summary of Feature Importance*

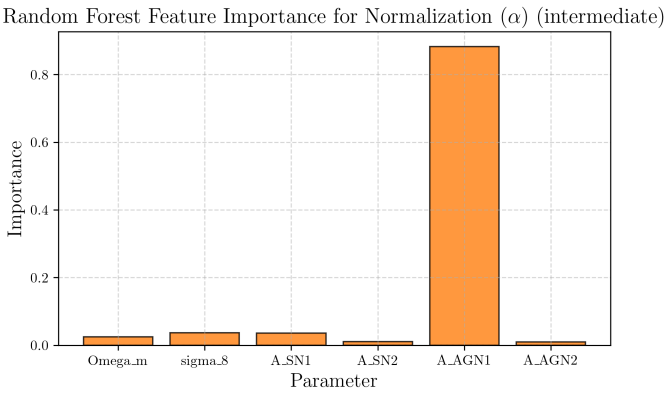

Figure 9. Random Forest Feature Importance for Normalization (α) (intermediate). The plot shows the feature importance of different cosmological parameters, including Ω_m, σ_8, and parameters related to Supernovae (SN) and Active Galactic Nuclei (AGN). The A_AGN1 parameter exhibits significantly higher importance compared to other parameters for normalization.

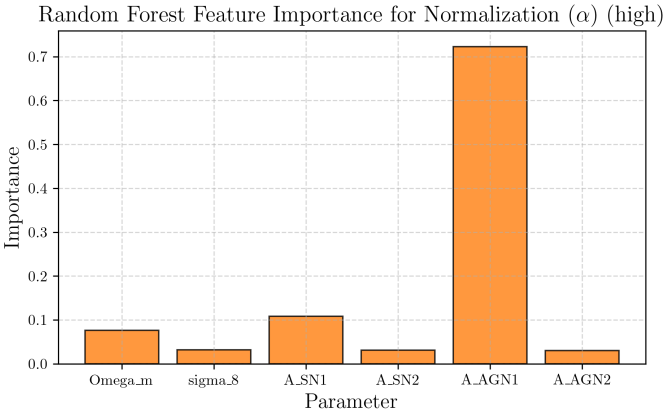

Figure 10. Random Forest feature importance for normalization α (high). The figure shows the importance of different cosmological parameters (Omega_m, sigma_8, A_SN1, A_SN2, A_AGN1, A_AGN2) as determined by a Random Forest model. A_AGN1 has the highest importance, while the other parameters have significantly lower importance scores.

The analysis reveals a clear mass dependence in the relative importance of feedback parameters:

- **Low Mass:** SN feedback dominates.
- **Intermediate Mass:** AGN feedback (especially $A_{\rm AGN1}$) dominates, with some SN contribution.
- **High Mass:** AGN feedback remains dominant, but the sign of its effect on slope and normalization reverses, indicating a shift from black hole fueling to quenching.

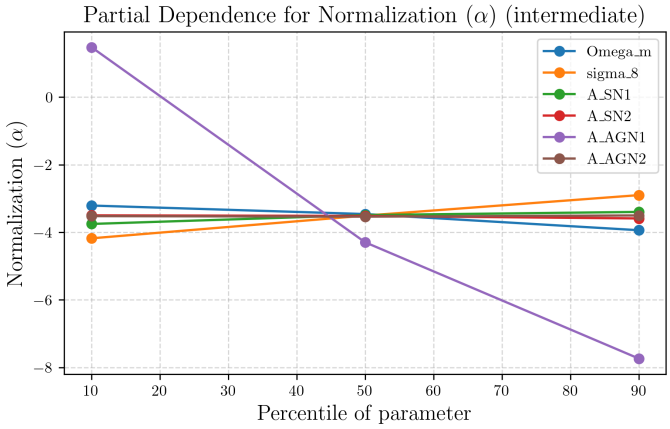

Figure 11. Partial dependence plot for the normalization parameter α as a function of the percentile of different cosmological parameters. The parameters include Ω_m, σ_8, and the normalization parameters for Supernovae (A_SN1, A_SN2) and Active Galactic Nuclei (A_AGN1, A_AGN2). A large change in the normalization parameter α is observed for A_AGN1 as a function of percentile, while the other parameters show relatively small changes.

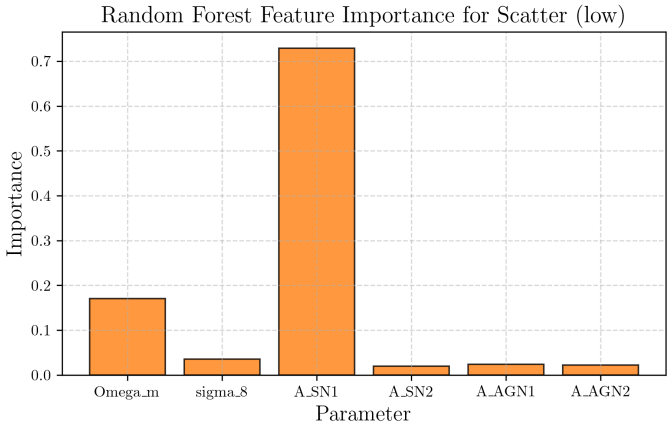

Figure 12. Random Forest Feature Importance for Scatter (low) shows the importance of different cosmological parameters in predicting the scatter. The parameter A_SN1 shows the largest importance.

4.4. *Black Hole Occupation Fraction*

4.4.1. *Trends with Mass and Feedback*

The black hole occupation fraction, defined as the fraction of galaxies in each mass bin hosting a black hole with $M_{\rm BH} > 0$, exhibits a strong dependence on stellar mass and feedback parameters. The occupation fraction rises from ∼0.72 at low mass to ∼0.96 at high mass, with significant catalog-to-catalog scatter.

Stronger SN feedback ($A_{\rm SN1}$) reduces the occupation fraction at low and intermediate mass, consistent with the suppression of black hole formation or growth in shallow potential wells. AGN feedback has a weaker effect on occupation at low mass but becomes more important at high mass.

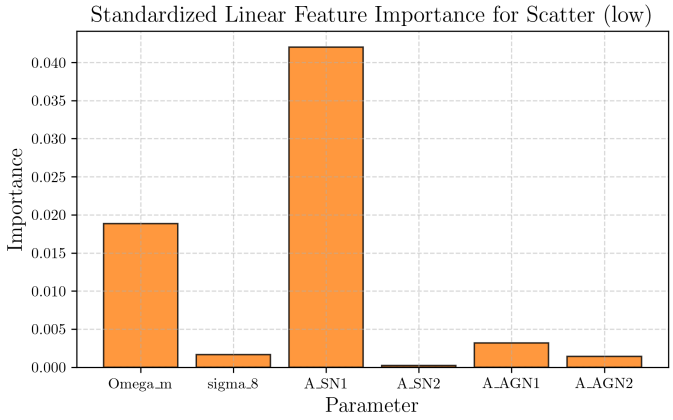

Figure 13. Standardized linear feature importance for the scatter (low). The figure shows the relative importance of different cosmological parameters and survey properties on the scatter in some derived quantity. The parameter A_SN1 appears to have the largest impact, while A_SN2 has the smallest. Omega_m also shows significant importance.

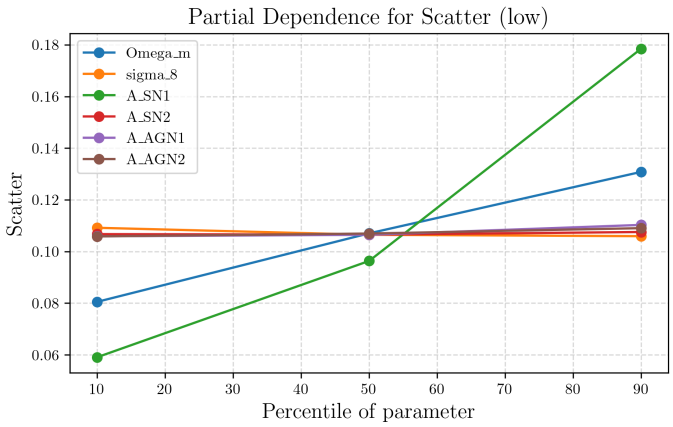

Figure 14. Partial dependence plot for the 'Scatter (low)' parameter. The plot shows the dependence of the scatter on different cosmological parameters: Omega_m, sigma_8, A_SN1, A_SN2, A_AGN1, and A_AGN2. The x-axis represents the percentile of each parameter, and the y-axis represents the scatter. Large differences are seen in the dependence of the scatter on A_SN1 and Omega_m compared to the other parameters.

4.4.2. *Implications for Black Hole Seeding*

The incomplete occupation at low mass, and its sensitivity to feedback, supports models in which black hole seeding is inefficient or stochastic in low-mass halos, and can be suppressed by energetic feedback.

4.5. *Secondary Dependencies and SFR*

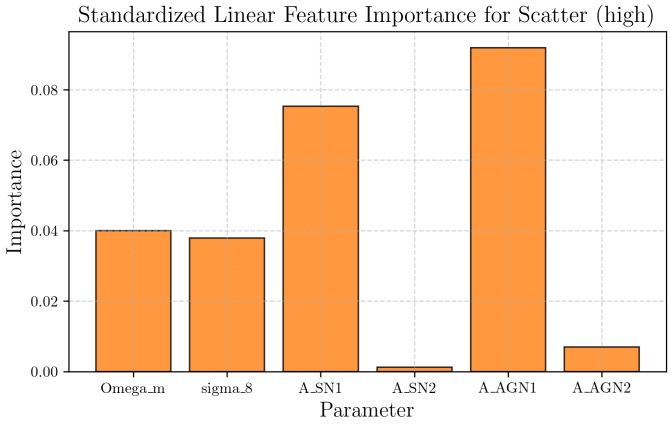

Figure 15. Standardized linear feature importance for scatter (high) for different cosmological parameters. The parameters A_AGN1 and A_SN1 show the largest feature importance, while A_SN2 and A_AGN2 show the smallest feature importance.

While the primary analysis did not stratify by SFR, the weak correlation between SFR and $M_{\rm star}$ ($r = 0.10$) and the moderate effect of feedback parameters on both SFR and $M_{\rm BH}$ suggest that residual trends with SFR are likely subdominant compared to the direct effects of feedback and mass. Future work could explicitly include SFR as a covariate.

4.6. *Comparison with Observational Constraints*

The mean slopes ($\beta \sim 1.1$–1.2 at intermediate/high mass) and scatters ($\sim$0.3–0.4 dex) are broadly consistent with local observations, though the normalization and scatter exhibit greater diversity in the simulations, especially at low mass. The simulated occupation fractions at high mass are consistent with the near-unity values inferred for massive galaxies, while the lower and more variable occupation at low mass matches the emerging observational picture for dwarfs and low-mass systems.

4.7. *Synthesis and Physical Interpretation*

The results demonstrate that the diversity in the $M_{\rm BH}$–$M_{\rm star}$ relation is primarily driven by the strength and mode of feedback, with SN feedback dominating at low mass and AGN feedback at high mass. The transition between these regimes is marked by a shift in the sign and magnitude of the feedback coefficients, and is accompanied by changes in the occupation fraction and scatter. While feedback is the dominant driver, cosmological parameters modulate the normalization and scatter, particularly in the intermediate and high mass bins.

The broad range of slopes, normalizations, and scatters observed across the simulated catalogs underscores the potential for significant diversity in the $M_{\rm BH}$–$M_{\rm star}$ relation in the real universe, especially in regimes where feedback is strong or stochastic. The results suggest that observed outliers and the intrinsic scatter in the scaling relation may be natural consequences of variations in feedback efficiency and cosmological context.

The sensitivity of the occupation fraction to feedback at low mass provides a direct link between black hole seeding models and the observable demographics of low-mass galaxies. The results support scenarios in which energetic feedback can suppress black hole formation or growth in shallow potential wells, leading to incomplete occupation and a wide diversity of black hole masses at fixed stellar mass.

5. CONCLUSIONS

This project aimed to systematically quantify how the black hole-stellar mass ($M_{\rm BH}$–$M_{\rm star}$) relation's slope, normalization, and scatter vary in simulated galaxies due to changes in cosmological and feedback parameters. By leveraging a dataset with unprecedented parameter coverage, we sought to reveal the physical drivers behind the diversity of these scaling relations and uncover secondary dependencies in galaxy evolution.

We employed a large suite of simulated galaxies, each generated with distinct cosmological (Ω_m, σ_8) and feedback ($A_{\rm SN1}$, $A_{\rm SN2}$, $A_{\rm AGN1}$, $A_{\rm AGN2}$) parameter combinations. We performed ordinary least squares (OLS) regression in log-space on the $M_{\rm BH}$–$M_{\rm star}$ relation for each simulation catalog within three stellar mass bins, excluding galaxies with $M_{\rm BH} = 0$. We then used multivariate linear regression and random forest analysis to map the fitted relation parameters (slope, normalization, scatter) as functions of the cosmological and feedback parameters. Additionally, we quantified the black hole occupation fraction as a function of stellar mass and feedback.

Our results demonstrate significant diversity in the $M_{\rm BH}$–$M_{\rm star}$ relation across the simulated parameter space. The slope, normalization, and scatter of the relation exhibit substantial variations, particularly in the intermediate and high mass bins. We found that feedback parameters, specifically supernova (SN) feedback ($A_{\rm SN1}$, $A_{\rm SN2}$) and active galactic nuclei (AGN) feedback ($A_{\rm AGN1}$, $A_{\rm AGN2}$), are the primary drivers of this diversity. SN feedback dominates in low-mass galaxies, while AGN feedback becomes increasingly important at higher masses. Cosmological parameters (Ω_m, σ_8) also modulate the normalization and scatter, albeit to a lesser extent than feedback. The black hole occupation fraction increases with stellar mass and is suppressed by SN feedback at low masses.

From these results, we have learned that the coevolution of black holes and galaxies is strongly influenced by feedback processes. Supernova feedback plays a crucial role in regulating black hole growth in low-mass galaxies, while AGN feedback shapes the $M_{\rm BH}$–$M_{\rm star}$ relation in more massive systems. The diversity observed in the simulations suggests that variations in feedback efficiency and cosmological context can lead to a wide range of black hole masses at a given stellar mass. Furthermore, the incomplete black hole occupation fraction at low masses, coupled with its sensitivity to feedback, supports scenarios in which black hole seeding is inefficient or stochastic in low-mass halos. This work provides a valuable framework for interpreting the observed diversity in black hole scaling relations and highlights the importance of considering both baryonic and cosmological physics in understanding galaxy evolution.

REFERENCES

Andreon, S., & Hurn, M. A. 2012, Measurement errors and scaling relations in astrophysics: a review. https://arxiv.org/abs/1210.6232

Arnold, C., Leo, M., & Li, B. 2019, Realistic simulations of galaxy formation in f(R) modified gravity, doi: https://doi.org/10.1038/s41550-019-0823-y

Avila-Reese, V. 2006, Understanding Galaxy Formation and Evolution, doi: https://doi.org/10.1007/978-1-4020-5575-1_4

Buddelmeijer, H., Boxhoorn, D., & Valentijn, E. A. 2011, Automatic Optimized Discovery, Creation and Processing of Astronomical Catalogs. https://arxiv.org/abs/1111.3784

Burke, C. J., Natarajan, P., Baldassare, V. F., & Geha, M. 2025, Multi-wavelength constraints on the local black hole occupation fraction. https://arxiv.org/abs/2410.11177

Busillo, V., Tortora, C., Covone, G., et al. 2024, CASCO: Cosmological and AStrophysical parameters from Cosmological simulations and Observations – II. Constraining cosmology and astrophysical processes with early- and late-type galaxies, doi: https://doi.org/10.1051/0004-6361/202451702

Bussmann, R. S., Narayanan, D., Shirley, Y. L., et al. 2008, The Star Formation Rate - Dense Gas Relation in Galaxies as Measured by HCN (3-2) Emission, doi: https://doi.org/10.1086/590181

Chen, Y., Gu, Q., Fan, J., et al. 2025, The relation between black hole spin, star formation rate, and black hole mass for supermassive black holes. https://arxiv.org/abs/2503.03223

Cimatti, A., Fraternali, F., & Nipoti, C. 2019, Introduction to Galaxy Formation and Evolution. From Primordial Gas to Present-Day Galaxies. https://arxiv.org/abs/1912.06216

Comparat, J., Merloni, A., Ponti, G., et al. 2025, Cross-correlation between soft X-rays and galaxies A new benchmark for galaxy evolution models. https://arxiv.org/abs/2503.19796

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

Ding, X., Silverman, J., Treu, T., et al. 2019, The mass relations between supermassive black holes and their host galaxies at 1<z<2 with HST-WFC3, doi: https://doi.org/10.3847/1538-4357/ab5b90

D'Isanto, A., Cavuoti, S., Gieseke, F., & Polsterer, K. L. 2018, Return of the features. Efficient feature selection and interpretation for photometric redshifts, doi: https://doi.org/10.1051/0004-6361/201833103

D'Odorico, V., Bolton, J. S., Christensen, L., et al. 2025, Galaxy Formation and Symbiotic Evolution with the Inter-Galactic Medium in the Age of ELT-ANDES, doi: https://doi.org/10.1007/s10686-024-09967-3

Gallo, E., Hodges-Kluck, E., Treu, T., et al. 2023, The black hole occupation fraction of local dwarf galaxies with AXIS. https://arxiv.org/abs/2311.09161

Han, S., Dubois, Y., Lee, J., et al. 2024, RAMSES-yOMP: Performance Optimizations for the Astrophysical Hydrodynamic Simulation Code RAMSES. https://arxiv.org/abs/2411.14631

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

Jing, T., & Li, C. 2024, Regression for Astronomical Data with Realistic Distributions, Errors and Non-linearity. https://arxiv.org/abs/2411.08747

Juodžbalis, I., Maiolino, R., Baker, W. M., et al. 2024, A dormant, overmassive black hole in the early Universe, doi: https://doi.org/10.1038/s41586-024-08210-5

Koss, M., U, V., Hodges-Kluck, E., et al. 2019, Astro2020 Science White Paper: Black Hole Growth in Mergers and Dual AGN. https://arxiv.org/abs/1903.06720

Li, X., Li, C., & Mo, H. 2024, What drives the HI content of central galaxies – A comparison between hydrodynamic simulations and observations using Random Forest. https://arxiv.org/abs/2411.07977

Llorella, F. R., & Cebrián, J. A. 2025, Exploring symbolic regression and genetic algorithms for astronomical object classification, doi: https://doi.org/10.33232/001c.132333

Martin, W., & Mortlock, D. J. 2024, Robust Bayesian regression in astronomy. https://arxiv.org/abs/2411.02380

Matt, C., Gültekin, K., & Simon, J. 2023, The Impact of Black Hole Scaling Relation Assumptions on the Mass Density of Black Holes, doi: https://doi.org/10.1093/mnras/stad2146

Montaguth, G. P., Torres-Flores, S., Monachesi, A., et al. 2023, Galaxy evolution in compact groups I: Revealing a transitional galaxy population through a multiwavelength approach, doi: https://doi.org/10.1093/mnras/stad2235

Nagamine, K. 2023, Feedback models in galaxy simulations and probing their impact by cosmological hydrodynamic simulations, doi: https://doi.org/10.1017/S1743921323000133

Narkedimilli, S., Raghav, S., Makam, S., Ayitapu, P., & H, A. B. 2024, Predicting Stellar Metallicity: A Comparative Analysis of Regression Models for Solar Twin Stars. https://arxiv.org/abs/2410.06709

Nguyen, T., Basu, A., & Budavári, T. 2022, Globally optimal and scalable N-way matching of astronomy catalogs, doi: https://doi.org/10.3847/1538-3881/ac6bf6

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

Pacucci, F., Mezcua, M., & Regan, J. A. 2021, The Active Fraction of Massive Black Holes in Dwarf Galaxies, doi: https://doi.org/10.3847/1538-4357/ac1595

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

Riccio, G., Brescia, M., Cavuoti, S., et al. 2017, C^3 : A Command-line Catalogue Cross-matching tool for modern astrophysical survey data, doi: https://doi.org/10.1017/S1743921316013120

Schawinski, K., Turp, M. D., & Zhang, C. 2018, Exploring galaxy evolution with generative models, doi: https://doi.org/10.1051/0004-6361/201833800

Schaye, J., Kugel, R., Schaller, M., et al. 2023, The FLAMINGO project: cosmological hydrodynamical simulations for large-scale structure and galaxy cluster surveys, doi: https://doi.org/10.1093/mnras/stad2419

Shankar, F., Bernardi, M., Richardson, K., et al. 2019, Black hole scaling relations of active and quiescent galaxies: Addressing selection effects and constraining virial factors, doi: https://doi.org/10.1093/mnras/stz376

Singh, N., Browne, L.-M., & Butler, R. 2013, Parallel Astronomical Data Processing with Python: Recipes for multicore machines, doi: https://doi.org/10.1016/j.ascom.2013.04.002

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

Terrazas, B. A., Bell, E. F., Woo, J., & Henriques, B. M. B. 2017, Supermassive black holes as the regulators of star formation in central galaxies, doi: https://doi.org/10.3847/1538-4357/aa7d07

Tinsley, B. M. 2022, Evolution of Stars and Gas in Galaxies, doi: https://doi.org/10.48550/arXiv.2203.02041

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

Valentini, M., & Dolag, K. 2025, Hydrodynamic methods and sub-resolution models for cosmological simulations. https://arxiv.org/abs/2502.06954

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

Yang, G., Chen, C. T. J., Vito, F., et al. 2017, Black-Hole Growth is Mainly Linked to Host-Galaxy Stellar Mass rather than Star Formation Rate, doi: https://doi.org/10.3847/1538-4357/aa7564

Yu, H.-R., Pen, U.-L., & Wang, X. 2018, CUBE: An Information-optimized parallel Cosmological N-body Algorithm, doi: https://doi.org/10.3847/1538-4365/aac830

Zerbo, M. C., Rossi, M. E. D., Lara-López, M. A., Cora, S. A., & Zenocratti, L. J. 2024, Connection between feedback processes and the effective yields of EAGLE galaxies. https://arxiv.org/abs/2407.07980

Zhang, T., Guo, Q., Qu, Y., & Gao, L. 2021, The role of mergers and gas accretion in black hole growth and galaxy evolution, doi: https://doi.org/10.1088/1674-4527/21/8/212

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

Zheng, Y., Xu, K., Zhao, D., et al. 2025, Photometric Objects Around Cosmic Webs (PAC). VII. Disentangling Mass and Environment Quenching with the Aid of Galaxy-halo Connection in Simulations. https://arxiv.org/abs/2501.00986

Zhu, P., Ho, L. C., & Gao, H. 2020, The Correlation between Black Hole Mass and Stellar Mass for Classical Bulges and the Cores of Ellipticals, doi: https://doi.org/10.3847/1538-4357/abcaa1