Draft version September 8, 2025
Typeset using LaTeX **preprint** style in AASTeX7.0.1

# Research Statement

Benjamin L. Davis [1]

[1] *Center for Astrophysics and Space Science, New York University Abu Dhabi, PO Box 129188, Abu Dhabi, UAE*

## 1. MY MOTIVATIONS

### 1.1. *Accurately Measuring Galaxies*

From the very beginning of my research as a graduate student, I have always been motivated by studying black hole mass scaling relations. Initially, I was inspired by the work of Seigar et al. (2008) to further study the relationship between supermassive black holes (SMBHs) and the geometry of their host spiral galaxies. This pursuit first led to an improved software tool to measure the winding angle of spiral arms (Davis et al. 2012, 2016; Seigar et al. 2018), and was followed by an updated relation (Berrier et al. 2013) and application to the black hole mass function (Davis et al. 2014; Mutlu-Pakdil et al. 2016; Fusco et al. 2022). My work on logarithmic spiral-arm pitch angle also yielded papers dealing with dark matter (Seigar et al. 2014), density wave theory (Davis et al. 2015; Pour-Imani et al. 2016; Miller et al. 2019; Abdeen et al. 2020, 2022), and an alternate measurement method (Shields et al. 2015, 2022). I finished my Ph.D. with a thesis that encapsulated my work on pitch angle and black holes (Davis 2015).

### 1.2. *Subsequent Scaling Relations and Related Discoveries*

I began my first postdoc with pitch-angle related work I was familiar with (Koliopanos et al. 2017; Davis et al. 2017) before moving on to new areas of research. My focus then turned to measuring accurate light profiles of galaxies, decomposing their components, and producing new black hole mass scaling relations for spiral galaxies (Davis et al. 2018, 2019a,b). This work led to several studies that used our new scaling relations to identify intermediate-mass black hole candidates (Graham et al. 2019, 2021a,b; Davis & Graham 2021). Our work then focused on decomposing the light profiles of early-type galaxies and producing their scaling relations in combination with the previous work from spiral galaxies (Sahu et al. 2019a,b, 2020, 2022a,b). Such work led me to join the LISA Consortium and co-authoring their white paper (Amaro-Seoane et al. 2023). We also investigated the possible disc cloaking of compact red nugget galaxies (Hon et al. 2022).

During my second postdoc, I began working in a galaxy formation group, focused on numerical simulations of galaxies (Waterval et al. 2024). We also used direct $N$-body simulations to tackle hot topics like "little red dots" (Khan et al. 2025). Additionally, I learned a lot about artificial intelligence (AI) and machine learning (ML). My adoption of these new technologies led to numerous, interdisciplinary studies (Pasquato et al. 2023; Jin & Davis 2023; Jin et al. 2024, 2025; Davis & Jin 2023; Davis et al. 2024, 2025a,b), which I will describe in greater detail in the subsequent sections. From spiral-arm pitch angles to causal discovery, my research has consistently sought fundamental links between black holes and galaxies.

Email: ben.davis@nyu.edu

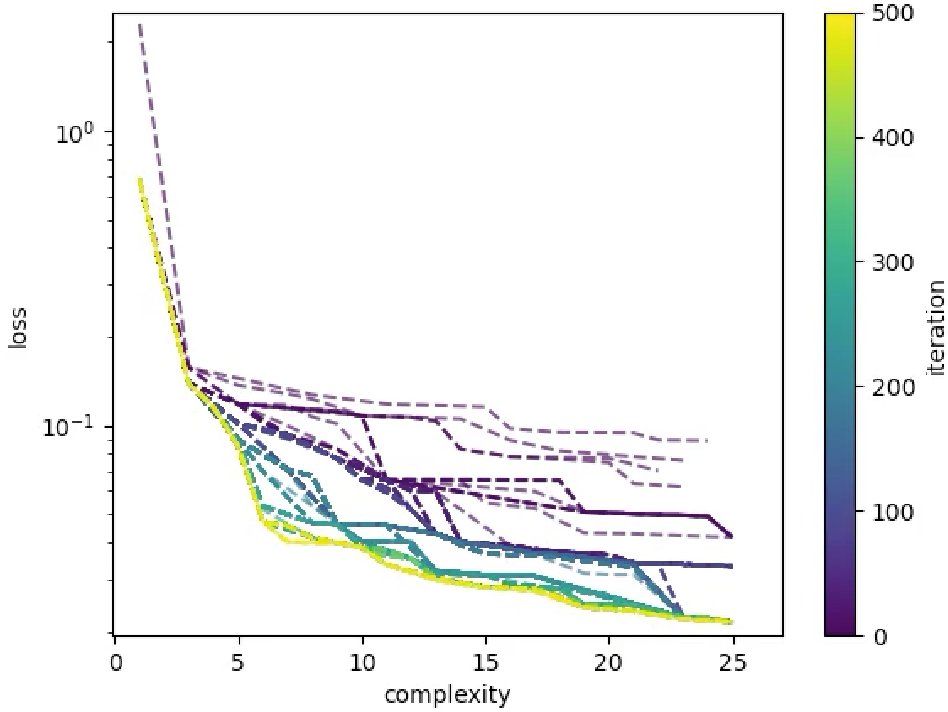


**Figure 1.** Evolution toward the Pareto front. Here, we plot loss (mean squared error) in the predicted black hole mass vs. the complexity (number of variables plus operators) of the formulae to predict black hole mass. We restrict the maximum complexity to 25. Each run of `PySR` (Cranmer 2023) evolves 2000 mutations and produces a front of ∼10 formulae that sample the population. In total, we run 500 iterations of `PySR` recursively to migrate (from --- to --- ) toward the true Pareto front. Finally, we select formulae from the final iteration ( --- ) to further refine into our optimal black hole mass scaling relations via our Bayesian framework.

## 2. WHERE MY MOTIVATIONS HAVE LED ME

### 2.1. *Finding Better Correlations with Symbolic Regression*

As I have described, my career was inspired by, and largely spent in the pursuit of scaling relations. Given the abundance of scaling relations, the challenge lies in finding improved ones by leveraging multi-dimensional data. This is where ML can be beneficial, especially the subfield of Symbolic Regression (SR). SR is an ML technique that aims to find a mathematical expression that best fits the data.

Based on Occam's razor, which states that the simplest model that fits the data is the best model, SR finds a Pareto front that best optimizes the trade-off between model complexity and accuracy. There are many scaling relations in astrophysics, such as 2-D ones including the Tully-Fisher relation, the Faber-Jackson relation, and the $M$–$\sigma$ relation, or 3-D ones such as the fundamental plane of elliptical galaxies. SR can be used to find novel (higher dimensional) scaling relations and provide new insights into the astrophysical processes behind these relations. Figure 1 is from a work in preparation that uses SR to discover Pareto optimal black hole mass scaling relations.

### 2.2. *Understanding Correlations via Causal Discovery*

As an observational science, most studies in astrophysics rely on correlations (hence, the scaling relations I have spent my career studying). Indeed, the natural way of determining causation is

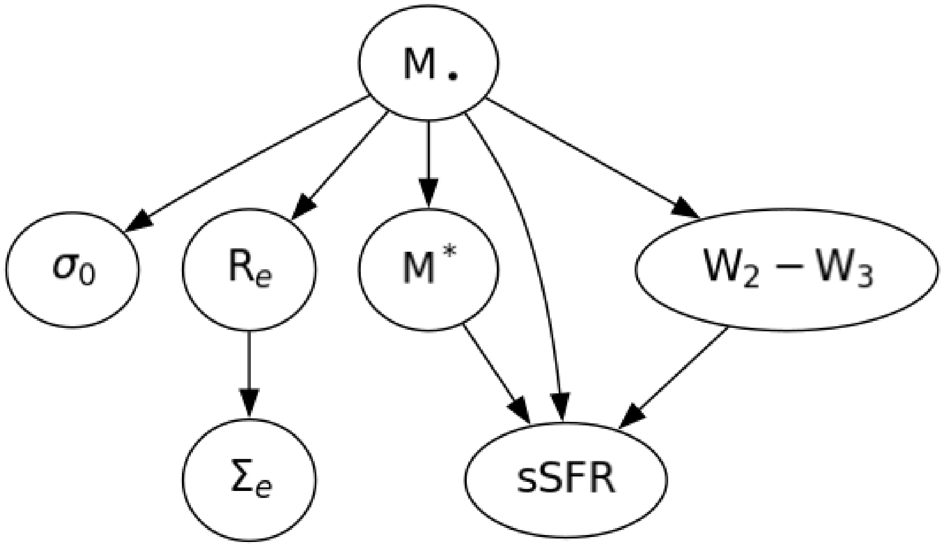


**Figure 2.** Example causal relationship identified in real spiral galaxy data shown in a directed acyclic graph (DAG). With black hole mass ($M_\bullet$) positioned at the top of the DAG, this suggests that black hole mass is a causal parent of its host galaxy properties: central stellar velocity dispersion ($\sigma_0$), spheroid effective radius ($R_\mathrm{e}$), galaxy stellar mass ($M^*$), $\mathrm{W}_2$−$\mathrm{W}_3$ color, average projected density within $R_\mathrm{e}$ ($\Sigma_\mathrm{e}$), and specific star-formation rate (sSFR).

through interventions, e.g., controlled experiments in labs. However, such interventions are impossible in astrophysics as humans have no control over the Universe, thus most studies are limited to correlations. This conundrum can be resolved by the field of causal discovery, a field that aims to infer causal relationships from observational data. This is accomplished mainly through examining the conditional independence relationships between variables, which can be represented by a causal graph. I am working on bringing causality to astrophysics, with the aim to understand the causal relationships behind astrophysical data. Figure 2 illustrates from our recent work that tries to solve the long-standing debate about the causal interpretation of the correlations between the mass of central SMBHs and the properties of their host galaxies (Jin et al. 2025).

## 3. WHERE MY MOTIVATIONS CONTINUE TO LEAD ME

### 3.1. *Cross-pollination of Astrophysics and Computer Science*

To date, there has been very little adoption of causal discovery in astronomy. Because causal discovery lies in the realm of AI and ML, with advances coming predominantly from computer science, I have made an effort to bring the two worlds together: astronomy and computer science. Over the past several years, I have presented papers at some of the world's most prominent computer science conferences (Pasquato et al. 2023; Jin & Davis 2023; Jin et al. 2024; Davis et al. 2025a). My goal is to cross-pollinate the two sciences so that astronomers become aware of the useful tools of computer scientists, and computer scientists learn about the suitability of applying their methods to astrophysical data that is ripe for new discoveries.

### 3.2. *Fertile Ground for Discoveries*

With great tools at my disposal, I am looking for problems to solve across multiple disciplines. Ongoing work is using causal discovery and $N$-body simulations to identify causal reversals across

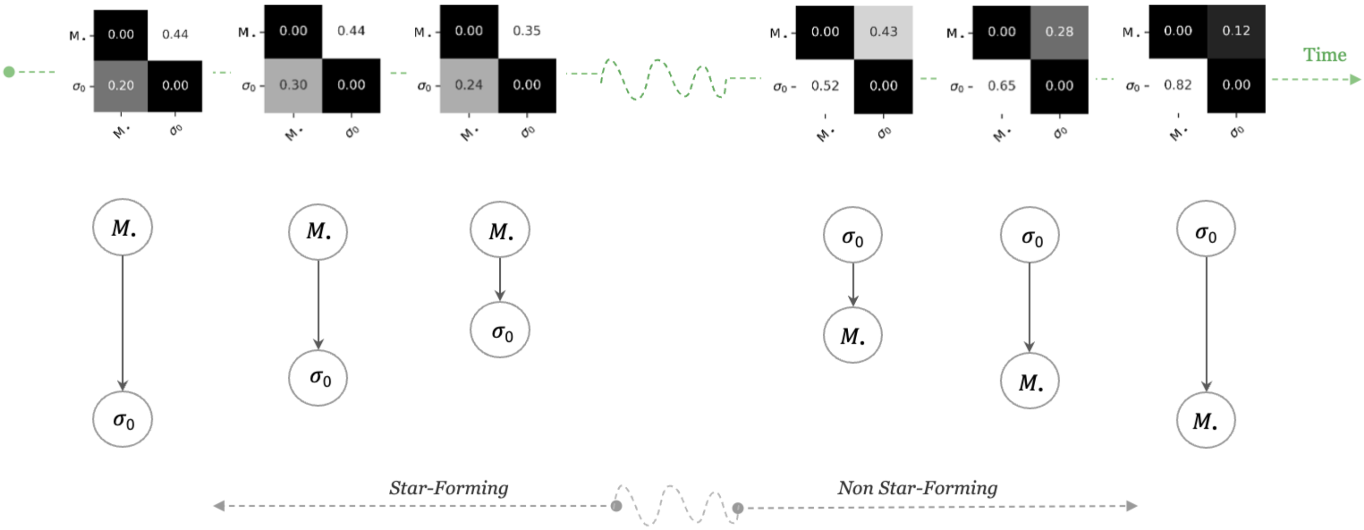


**Figure 3.** Time evolution of the causal $M_\bullet$–$\sigma_0$ relation. Here, we reduce the full edge marginal matrices to only the $2 \times 2$ matrices concerning SMBH mass and central stellar velocity dispersion. The $x$-axes depict causal children and the $y$-axes depict causal parents. We select six matrices, arranged from youngest (on the left) to oldest (on the right). The three matrices on the left are from successive epochs progressing towards the peak of star formation. The three matrices on the right are from successive epochs receding progressively away from the peak of star formation. The matrices demonstrate a monotonic decrease of the Bayesian probabilistic advantage of $M_\bullet \to \sigma_0$ vs. $\sigma_0 \to M_\bullet$ with time. Explicitly, the off-diagonal elements of the first matrix (left) begin with $P(M_\bullet \to \sigma_0) = 44\%$ vs. $P(\sigma_0 \to M_\bullet) = 20\%$, and the off-diagonal elements of the last matrix (right) end with $P(M_\bullet \to \sigma_0) = 12\%$ vs. $P(\sigma_0 \to M_\bullet) = 82\%$. The DAGs depict the net strength (represented by arrow length) and direction of the dominant causal direction in the $M_\bullet$–$\sigma_0$ relation as a function of time. In our study, we generally consider the *star-forming* period to be $\geq$1 Gyr *before* the peak of star formation and the *non-star-forming* period to be $\geq$1 Gyr *after* the peak of star formation.

the temporal evolution of galaxies (Figure 3). I am even using these tools to find discoveries in planetary science. Our recent publication (Davis et al. 2025b) discovered the causal origin of colors in trans-Neptunian objects (Figure 4), and ongoing work is looking at causal connections between exoplanets and their host stars.

## 4. MY DESIRED FUTURE IN A TENURE-TRACK FACULTY POSITION

My research journey has been driven by a desire to understand the fundamental connections in our Universe, from the structure of galaxies to the causal relationships that govern their evolution. This path has led me from observational astronomy to the cutting edge of computational science, where I have focused on developing and applying novel ML techniques to astrophysical data, yielding new discoveries and insights. The next phase of my career will be focused on establishing a strong, independent research program that continues this cross-pollination between astrophysics and computer science. The synthesis of scaling relations plus ML/causality is a brand new field of research that I have largely created and I am well positioned to guide its growth toward wide-spread adoption.

My future research plans are centered on fertile ground for new discoveries. With the new Vera C. Rubin Observatory online, the amount of new data will flood astronomy and my methods will be well-positioned to take advantage of this era of rapid growth in astronomical data. I will leverage the powerful tools of symbolic regression and causal discovery to not only uncover new, higher-

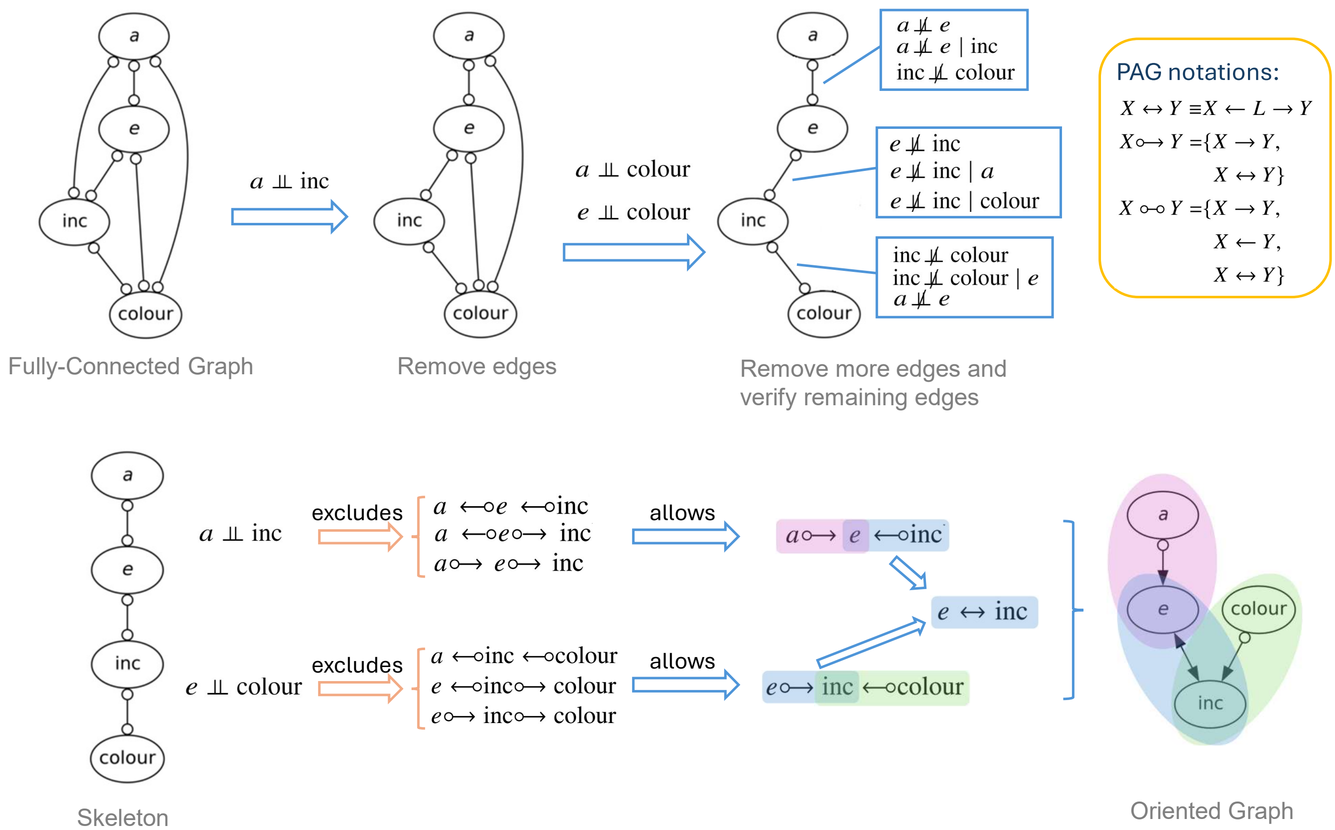


**Figure 4.** The visualization of the Fast Causal Inference (FCI) algorithm. In FCI stage i (upper panel), the algorithm starts with a fully-connected graph where all nodes are interconnected with ∘—∘ to allow all possible cases. Then, the algorithm goes through every edge and removes an edge when two nodes are independent, and when two nodes become conditionally independent given some subset of other variables. In FCI stage ii (lower panel), the remaining edges are oriented by constraining the ending symbols of each edge according to conditional independencies.

dimensional scaling relations but to understand their physical origins. This research program is perfectly suited for mentoring students. My projects naturally lend themselves to student engagement, from developing AI tools to applying them to galaxy and exoplanet data. This approach has a proven track record of producing high-impact publications and is well-positioned to attract the extramural funding necessary to build and sustain a vibrant research group.

In addition to my research, I am enthusiastic about teaching, mentoring, and advising students/postdocs. My broad background, spanning observational techniques, numerical simulations, and ML, prepares me to teach a wide range of courses across the physics and astronomy curriculum. I am eager to share my passion for discovery with the next generation of scientists and to become a valued faculty member. I am excited by the prospect of becoming a professor and establishing a legacy of excellence in research and education.

## REFERENCES


Abdeen, S., Kennefick, D., Kennefick, J., et al. 2020, MNRAS, 496, 1610

Abdeen, S., Davis, B. L., Eufrasio, R., et al. 2022, MNRAS, 512, 366

Amaro-Seoane, P., Andrews, J., Arca Sedda, M., et al. 2023, Living Reviews in Relativity, 26, 2

Berrier, J. C., Davis, B. L., Kennefick, D., et al. 2013, ApJ, 769, 132

Cranmer, M. 2023, arXiv e-prints, arXiv:2305.01582

Davis, B. 2015, PhD thesis, University of Arkansas

Davis, B. L., Ali-Dib, M., Zheng, Y., et al. 2025a, arXiv e-prints, arXiv:2507.03760

—. 2025b, MNRAS

Davis, B. L., Berrier, J. C., Shields, D. W., et al. 2012, ApJS, 199, 33

—. 2016, 2DFFT: Measuring Galactic Spiral Arm Pitch Angle

Davis, B. L., & Graham, A. W. 2021, PASA, 38, e030

Davis, B. L., Graham, A. W., & Cameron, E. 2018, ApJ, 869, 113

—. 2019a, ApJ, 873, 85

Davis, B. L., Graham, A. W., & Combes, F. 2019b, ApJ, 877, 64

Davis, B. L., Graham, A. W., & Seigar, M. S. 2017, MNRAS, 471, 2187

Davis, B. L., Graham, A. W., Soria, R., et al. 2024, ApJ, 971, 123

Davis, B. L., & Jin, Z. 2023, ApJL, 956, L22

Davis, B. L., Berrier, J. C., Johns, L., et al. 2014, ApJ, 789, 124

Davis, B. L., Kennefick, D., Kennefick, J., et al. 2015, ApJL, 802, L13

Fusco, M. S., Davis, B. L., Kennefick, J., Kennefick, D., & Seigar, M. S. 2022, Universe, 8, 649

Graham, A. W., Soria, R., Ciambur, B. C., Davis, B. L., & Swartz, D. A. 2021a, ApJ, 923, 146

Graham, A. W., Soria, R., & Davis, B. L. 2019, MNRAS, 484, 814

Graham, A. W., Soria, R., Davis, B. L., et al. 2021b, ApJ, 923, 246

Hon, D. S. H., Graham, A. W., Davis, B. L., & Marconi, A. 2022, MNRAS, 514, 3410

Jin, Z., & Davis, B. L. 2023, arXiv e-prints, arXiv:2310.19406

Jin, Z., Pasquato, M., Davis, B. L., Macciò, A. V., & Hezaveh, Y. 2024, arXiv e-prints, arXiv:2410.14775

Jin, Z., Pasquato, M., Davis, B. L., et al. 2025, ApJ, 979, 212

Khan, F. M., Davis, B. L., Macciò, A. V., & Holley-Bockelmann, K. 2025, ApJL, 986, L1

Koliopanos, F., Ciambur, B. C., Graham, A. W., et al. 2017, A&A, 601, A20

Miller, R., Kennefick, D., Kennefick, J., et al. 2019, ApJ, 874, 177

Mutlu-Pakdil, B., Seigar, M. S., & Davis, B. L. 2016, ApJ, 830, 117

Pasquato, M., Jin, Z., Lemos, P., Davis, B. L., & Macciò, A. V. 2023, arXiv e-prints, arXiv:2311.15160

Pour-Imani, H., Kennefick, D., Kennefick, J., et al. 2016, ApJL, 827, L2

Sahu, N., Graham, A. W., & Davis, B. L. 2019a, ApJ, 876, 155

—. 2019b, ApJ, 887, 10

—. 2020, ApJ, 903, 97

—. 2022a, Acta Astrophysica Taurica, 3, 39

—. 2022b, ApJ, 927, 67

Seigar, M. S., Davis, B. L., Berrier, J., & Kennefick, D. 2014, ApJ, 795, 90

Seigar, M. S., Kennefick, D., Kennefick, J., & Lacy, C. H. S. 2008, ApJL, 678, L93

Seigar, M. S., Mutlu-Pakdil, B., Hewitt, I. B., & Treuthardt, P. 2018, P2DFFT: Parallelized technique for measuring galactic spiral arm pitch angles

Shields, D. W., Boe, B., Pfountz, C., et al. 2015, Spirality: Spiral arm pitch angle measurement

—. 2022, Galaxies, 10, 100

Waterval, S., Macciò, A. V., Buck, T., et al. 2024, MNRAS, 533, 1463