Contribution ID: **237**   Type: **Oral Presentation**

Machine-Learnt Star-formation laws: Symbolic Regression with FIRE-2 galaxies

Thursday 11 July 2024 11:50 (20 minutes)

Whilst star formation (SF) in the interstellar medium (ISM) and the physics that govern it are some of the most fundamental mechanisms needed to paint a nuanced understanding of galaxy evolution, attempts to construct closed-form analytic expressions that connect SF and physical variables that have been observed to influence it, such as the density and turbulence properties of surrounding gas, still exhibit substantial intrinsic scatter. In this work we leverage recent advancements in machine learning (ML) and use neural network symbolic regression (SR) techniques to produce the first data-driven, ML-discovered analytic relations for SF using the publicly available FIRE-2 simulation suites, which have no explicit numerical sub-grid recipe for SF. We employ a genetic algorithm-based SR pipeline that assembles analytic functions to model a given dataset called PySR, training it to predict symbolic representations of a model for the star formation rate surface density (ΣSFR) at both 10 mega-years (Myr) and 100 Myr based on extracted variables from FIRE-2 galaxies. These variables include those dominated by small-scale characteristics such as gas surface densities, gas velocity dispersions and surface density of stars, as well as large-scale environmental properties like the dynamical time and the potential of gas. The stochastic nature of short-scale SF is reflected in the functional forms of the equations we find via PySR for ΣSFR at 10 Myr, which are heavily dependent on properties that show more fluctuation on a local scale such as the surface density and velocity dispersion of gas. These results indicate the critical need to consider turbulence and validate "bottom-up" approaches in recipes of SF over short timescales. The equations that PySR finds to describe ΣSFR at 100 Myr exhibit more influence from properties that show more consistency over the entire galaxy such as the dynamical time, pointing to the efficacy of "top-down" analytic models for describing SF when considering longer timescales. Furthermore, the equations found for the longer SFR timescale better capture the intrinsic physical scatter of the data within the Kennicutt-Schmidt plane, indicating that training on longer SFR timescales leads to less overfitting. Future applications of this process include investigation of equations found for observational data and input of such found equations into semi-analytic models of galaxies.

Presenter: SALIM, Diane

Session Classification: Cosmology & Simulations