

A statement of broader impacts is included in Section A. Section B provides additional details on the relevant astrophysical background for readers interested in research around Giant Molecular Clouds (GMCs) and star formation. Further experimental details, including extended qualitative comparisons across different methods, are presented in Section C. We have also included our **core code** as part of the supplementary material, with the full codebase and datasets to be released upon acceptance.

## A Statement of broader impacts

We apply the Diffusion Schrödinger Bridge framework to observational inverse prediction problems in astronomy. We do not anticipate any immediate societal risks or ethical concerns associated with this research. Our work focuses on advancing scientific understanding in a domain with minimal direct human impact, and all data used in this study are derived from publicly available astronomical observations or licensed simulations.

Beyond its immediate scientific goals, this research contributes positively to the broader advancement of interdisciplinary science. By demonstrating the application of a principled generative modeling framework in a physically grounded domain, our work helps bridge the gap between data-driven machine learning and theory-driven physical modeling. This approach not only promotes interpretability and robustness in ML research but also supports the long-term vision of using AI to accelerate scientific discovery. We hope this study encourages further collaboration between the machine learning and astrophysics communities, especially in leveraging modern ML techniques to understand complex natural systems.

## B Astrophysical background

We provide extended details about GMCs and star formation research for readers interested in the astrophysical background related to this work. In this work, we focus on studying the physical conditions and processes within giant molecular clouds (GMCs), by inferring the precise distributions of various critical physical quantities from limited observational data.

Giant Molecular Clouds (GMCs) are key structures within the interstellar medium (ISM) of galaxies, composed primarily of cold gas and dust. The ISM, which fills the space between stars, also contains cosmic rays and magnetic fields and plays a central role in galactic evolution by regulating star formation rates (Spitzer Jr, 2008). GMCs are of particular interest because they harbor most of the dense gas in the ISM, which is essential for the formation of stars (McKee and Ostriker, 2007). The physical conditions within GMCs are highly variable and complex, exhibiting spatial and temporal variations in density, temperature, velocity, and magnetic field strength and orientation. These factors give rise to a wide array of astrophysical phenomena, including the formation of protostars (Heyer and Dame, 2015), star clusters (Krumholz et al., 2019), and complex organic molecules (Bisbas et al., 2023). Among all physical quantities, we study two fundamental quantities in this work, which are the gas density ( $\rho$ ) and the magnetic field ( $B$ ).

### B.1 Gas density

The gas density (i.e., mass per unit volume), often measured as the number density of hydrogen nuclei (H), is considered one of the most fundamental quantities in GMC studies. Under the assumption of an abundance of one He nucleus for every 10 H nuclei in interstellar gas, we have a mass per H nucleus of  $\mu_H = 1.4m_H = 2.34 \times 10^{-24}g$ , leading to  $n_H = 1cm^{-3}$  equivalent to  $\rho = 2.34 \times 10^{-24}g cm^{-3}$ .

The gas density directly governs gravitational collapse, sets the local timescale for star formation, and strongly influences the thermal and chemical evolution of the gas. For example, denser regions within GMCs are more likely to become gravitationally unstable, leading to the formation of stars or star clusters. Moreover, the density controls the rates of radiative cooling and chemical reactions, such as molecule formation and dust grain processes, which are crucial for understanding the lifecycle of interstellar gas and the conditions for planet formation. Accurately inferring the spatial distribution of density within GMCs is therefore essential for modeling the early stages of stellar evolution.

In astronomical studies, it is usually very challenging to precisely quantify the number density of GMCs from observations. The traditional astrostatistical approach relies on the observations of column density and certain assumptions on the geometry of the clouds. For example, a cylindrical

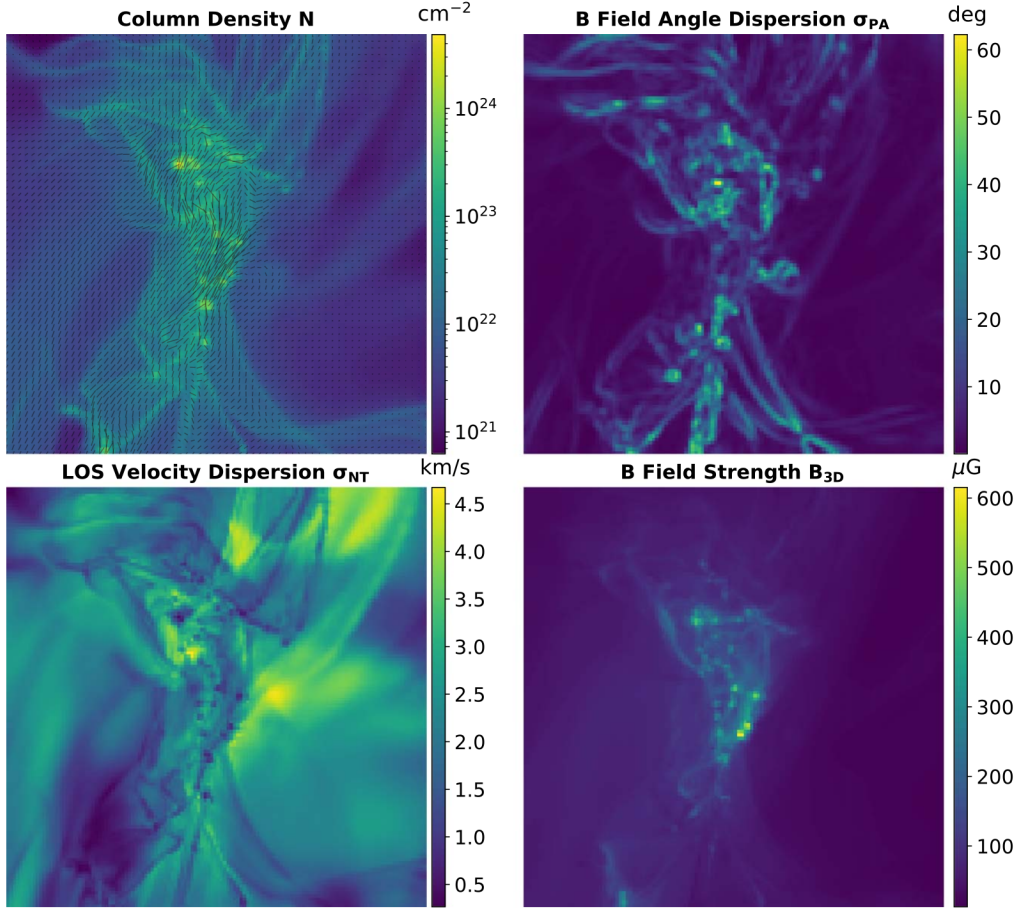


Figure 5: **Illustration of different synthetic observables from (Xu et al., 2025) for density and magnetic field estimation.** Top left: map of column density with magnetic field directions; Top right: magnetic field angle dispersion; Bottom left: LOS velocity dispersion; Bottom right: projected magnetic field strength.

834 geometry for filamentary structures or spherical geometry for dense cores (Palmeirim et al., 2013).  
835 Bisbas et al. (2021, 2023) later propose an empirical power law to convert the observed column  
836 density to the mean number density of GMCs based on MHD simulations (Wu et al., 2015), which has  
837 also been adapted in our work as two baseline methods (i.e., two-component power-law conversion  
838 and three-component power-law conversion). Despite being effective in similar simulation data, the  
839 prediction performance of these methods often suffers from a significant drop when new observations  
840 present a large distributional shift compared to the original fitting data (i.e., OOD testing cases).  
841 In our work, we aim to introduce a machine learning based approach with better robustness and  
842 generalization ability.

## 843 B.2 Magnetic field

844 Magnetic fields, often measured in microgauss ( $\mu G$ ), represent another critical physical quantity  
845 within the interstellar medium (ISM), playing a fundamental role in shaping the structure and regulat-  
846 ing the dynamics of GMCs (Crutcher, 1999, 2012). They can provide support against gravitational  
847 collapse, guide anisotropic gas flows, and influence the morphology of star-forming regions through  
848 magnetic tension and pressure. In combination with turbulence, magnetic fields are believed to  
849 significantly affect the efficiency and fragmentation of star formation across spatial scales.

850 Similar to the previous density estimation, estimating the magnetic field strength within GMCs is a  
851 challenging task. The Davis–Chandrasekhar–Fermi (DCF) method is a widely adopted technique

in astrophysics to estimate the plane-of-sky magnetic field strength  $B_{\text{POS}}$  in molecular clouds based on observed dust polarization (Davis, 1951; Chandrasekhar and Fermi, 1953; Beck, 2016), which relates the magnetic field strength to the polarization angle and other observable properties. Its core assumption is that an equipartition exists between the magnetic field energy and the turbulent kinetic energy of the gas. Specifically, the DCF method states that the relation between the gas density  $\rho$ , the nonthermal velocity dispersion  $\sigma_V$ , and the polarization angle dispersion  $\sigma_{PA}$  gives the POS magnetic field strength  $B_{\text{POS}}$  as:

$$B_{\text{POS}} = f \sqrt{4\pi\rho} \frac{\sigma_V}{\sigma_{PA}}, \quad (7)$$

where  $f$  is a correction factor accounting for projection and geometric effects. More recently, a variant of DCF method (Skalidis and Tassis, 2021) takes the compressible modes into account and modifies the previous Eq. 7 to:

$$B_{\text{POS}} = \sqrt{2\pi\rho} \frac{\sigma_V}{\sqrt{\sigma_{PA}}}. \quad (8)$$

While widely used, this method suffers from several limitations: (1) the equipartition assumption may not hold in strongly magnetized or shock-dominated regions; (2)  $\sigma_\phi$  becomes unreliable in regions with low signal-to-noise ratio or complex field morphology; and (3) accurate density and turbulence estimates are required for robust magnetic field inference, which are themselves uncertain. Moreover, DCF provides only an average value over large regions and lacks the ability to resolve fine-scale spatial structures.

In this work, we aim to bypass the limitations of traditional analytic approximations by leveraging machine learning to directly infer the underlying magnetic field distribution from observables, jointly with other key physical quantities such as gas density. Our proposed framework thus offers a more flexible and data-driven solution, especially under conditions with limited observability or strong distribution shifts.

In Fig. 5, we visualize an example of molecular cloud data used in machine learning methods from Xu et al. (2025), including the map of column density with magnetic field directions, the magnetic field angle dispersion, the LOS velocity dispersion, and the projected magnetic field strength.

## C Additional experimental results and analysis

We provide extended details about the experiments in this section.

### C.1 More details about the magnetohydrodynamics data simulations

We follow the same simulation setups used in prior works that apply conditional DDPMs to astrophysical observational prediction (Xu et al., 2023b, 2025; Wu et al., 2020; Hsu et al., 2023). While the design and analysis of physical simulations represent an important research direction on their own that is distinct from the machine learning focus of this work, we provide only the essential background here and refer interested readers to the original astrophysical literature for further details.

**Density simulation.** For the density simulations, we adopt the magnetohydrodynamic (MHD) simulations of colliding and non-colliding GMCs following prior works (Wu et al., 2020; Hsu et al., 2023), performed using the adaptive mesh refinement (AMR) code Enzo (Bryan et al., 2014). The simulations include self-gravity, magnetic fields, and thermal processes under a photodissociation radiation field ( $G_0 = 4$  Habings) and cosmic-ray ionization rate  $\zeta = 10^{-16} \text{ s}^{-1}$ . Two clouds with radii of 20 pc are initialized in a  $128^3 \text{ pc}^3$  domain with a  $256^3$  grid resolution.

Each cloud starts with  $n_H = 83 \text{ cm}^{-3}$  and  $T = 15 \text{ K}$ , while the ambient gas is set at  $n_H = 8.3 \text{ cm}^{-3}$  and  $T = 150 \text{ K}$  to maintain pressure balance. A multiphase temperature structure develops over time, with high-density regions ( $n_H \gtrsim 10^3 \text{ cm}^{-3}$ ) cooling to 10–20 K, and low-density regions ( $n_H \lesssim 10 \text{ cm}^{-3}$ ) reaching up to 1000 K. Magnetic fields are initialized at 10–50  $\mu\text{G}$ , oriented  $60^\circ$  to the collision axis. The refinement strategy ensures that the local Jeans length is resolved by at least eight cells.

The simulation runs span 5 Myr, with snapshots taken at 3 and 4 Myr. Column density maps and corresponding line-of-sight (LOS) mass-weighted number density maps are extracted at multiple

physical scales (32, 16, 8, and 4 pc), and projected to  $128 \times 128$  pixel images. The resulting dataset contains 7179 samples, with an approximate 80/20 split between training and testing. For more details, please refer to (Wu et al., 2020; Hsu et al., 2023).

**Magnetic field simulation.** For magnetic field prediction tasks, we adopt ideal magnetohydrodynamics (MHD) simulations following the setup in prior work (Wu et al., 2020; Hsu et al., 2023), using the MUSCL-Dedner method and HLLD Riemann solver within the AMR code ENZO (Bryan et al., 2014). These simulations incorporate self-gravity, magnetic fields, and thermal processes under a photodissociation radiation field with  $G_0 = 4$  and a cosmic-ray ionization rate of  $\zeta = 10^{-16} \text{ s}^{-1}$ . Each simulation initializes two clouds (radius 20 pc) in a  $128 \text{ pc}^3$  domain at  $256^3$  resolution, with initial hydrogen number density  $n_{\text{H}} = 83 \text{ cm}^{-3}$ , temperature  $T = 15 \text{ K}$ , and solenoidal turbulence satisfying  $v_k^2 \propto k^{-4}$  for  $2 \leq k \leq 20$ . The magnetic field is initialized at a  $60^\circ$  angle to the collision axis with strengths of 10, 30, and  $50 \mu\text{G}$  across different cases, and four additional refinement levels are used to resolve the local Jeans length. For each magnetic field configuration, we model both colliding and non-colliding GMC scenarios, where colliding clouds are offset by  $0.5 R_{\text{GMC}}$  and have a relative velocity of  $10 \text{ km/s}$ . These simulations run up to 4.1 Myr and capture early evolutionary phases before star formation begins.

To generate realistic observables, we compute maps of column density, line-of-sight (LOS) mass-weighted polarization angle, LOS nonthermal velocity dispersion, and projected 3D magnetic field strength across varying AMR levels. The final dataset consists of 25,479 images at  $128 \times 128$  resolution, sampled across physical scales of 32, 16, 8, and 4 pc. To enhance diversity, we include multiple projection angles and aggregate LOS effects without manually disentangling contributions such as turbulence or cloud-scale motions, consistent with observational limitations.

## C.2 Additional training details

As we have mentioned in the main paper, our proposed *Astro-DSB* benefits from a faster convergence speed and thus less computational resources compared to conditional DDPMs and vanilla DSB models. In Fig. 6, we compare the training curves among different variants of DSB as additional ablation analysis.

Compared to previous works that leverage conditional DDPMs as the prediction framework (Xu et al., 2023a, 2025), our *Astro-DSB* converges around 25k iterations as shown in Fig. 6(d), which corresponds to approximately 10 GPU hours. In contrast, conditional DDPMs take around 40 GPU hours to train.

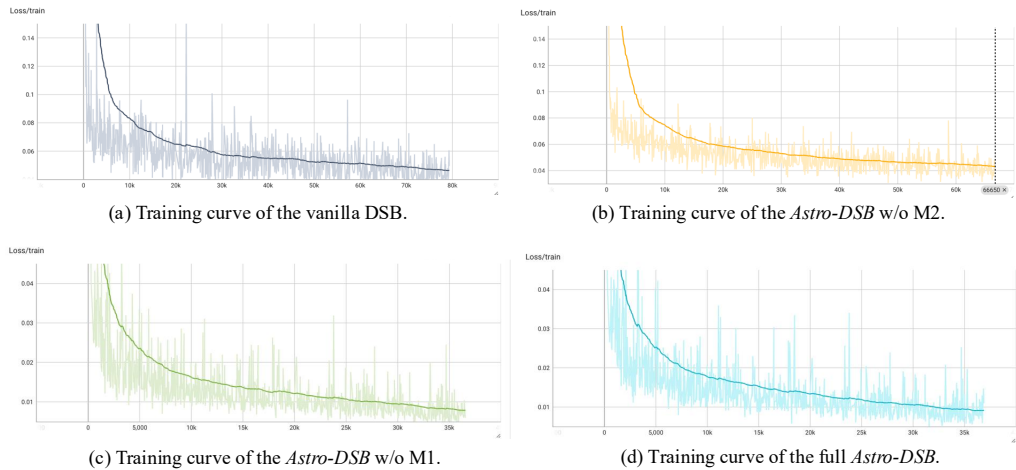


Figure 6: **Training curve comparison of different variants of DSB for ablation analysis.** We show that the proposed M2 component on the observable enhancement contributes to a faster training convergence speed. Best viewed in color and with zoom-in.



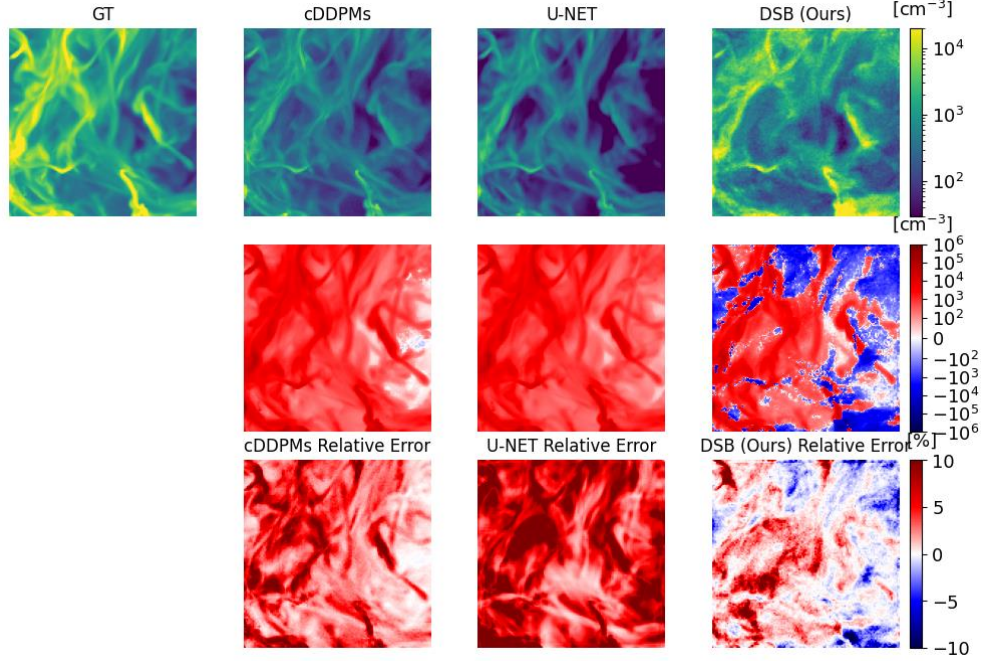


Figure 7: **Additional visualization of relative error patterns from different machine learning methods for the density prediction task on OOD testing cases.** In the first row, we qualitatively show the GT and prediction results from conditional DDPMs, U-Net, and our proposed *Astro-DSB* method. The second and third rows present the relative errors computed by “GT - pred” in absolute values and the values after log transformation, respectively. Compared to other ML methods that tend to underpredict in OOD testing cases, our *Astro-DSB* achieves less biased and more robust prediction performance.

### 929 C.3 More details about experiments on simulation data

930 As explained in the main paper, evaluations in astrophysical prediction particularly emphasize the  
 931 relative error patterns exhibited between the prediction results and the ground-truth distributions. In  
 932 Fig. 7 we present additional visualization of relative error patterns from different machine learning  
 933 methods for the density prediction task on OOD testing cases. The first row shows the ground truth  
 934 and prediction results from conditional DDPMs, U-Net, and our proposed *Astro-DSB*. The second  
 935 and third rows visualize the corresponding relative errors, computed as “GT - Pred” in absolute values  
 936 and after applying log transformation, respectively. Compared to DDPMs and U-Net, which tend to  
 937 significantly underpredict dense structures in the OOD regime, *Astro-DSB* produces more balanced  
 938 and spatially consistent residuals. These visual results further support the claim that our method  
 939 achieves improved robustness and lower bias when generalizing beyond training distributions.

940 To better understand the impact of each design component, we compare the distribution of relative  
 941 errors for different methods in Fig. 8. Traditional power-law baselines (p2, p3) and vanilla DDPMs  
 942 (cdm) exhibit large biases and heavy tails, especially under distribution shifts. The vanilla DSB  
 943 (psbv1) slightly reduces error spread, while our proposed components—M1 and M2—each contribute  
 944 to sharpening and centering the PDF. The full model (psbv4) demonstrates the narrowest and most  
 945 symmetric error distribution, highlighting its robustness and well-calibrated behavior in OOD settings.

### 946 C.4 More details about experiments on real observations

947 We further compare our proposed *Astro-DSB* with previous machine learning methods on real  
 948 observational data from the Taurus B213 region, focusing on the inferred number density distribution.  
 949 As shown in Fig. 4, the prediction of *Astro-DSB* closely matches the spatial structures and filament  
 950 morphology present in the original Herschel column density map. In contrast, the prediction from

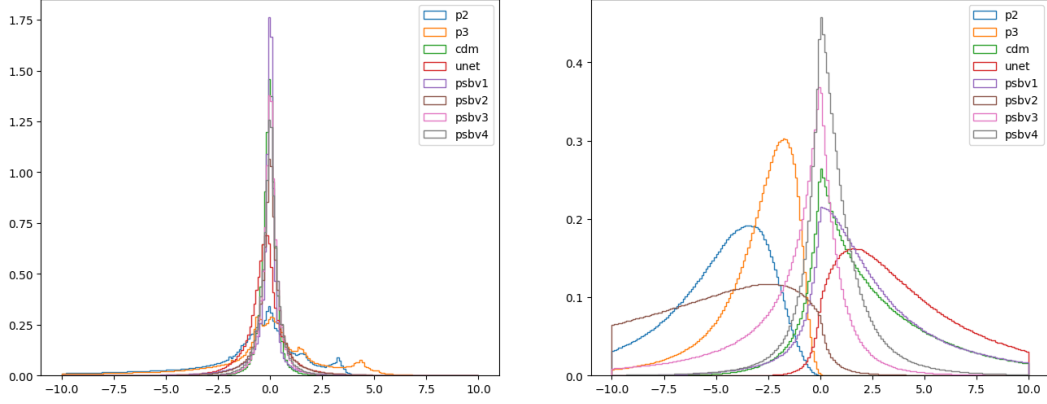


Figure 8: **PDF comparisons across different methods for the density prediction task on ID and OOD data.** Methods  $p2$  and  $p3$  correspond to two- and three-component power-law conversions, respectively.  $cdm$  denotes conditional DDPMs.  $psbv1$  is the vanilla diffusion Schrödinger bridge without our proposed modules.  $psbv2$  and  $psbv3$  remove M1 (noise alignment) and M2 (observable enhancement), respectively, while  $psbv4$  represents the full *Astro-DSB* model.

951 CASI-2D (Xu et al., 2020a) fails to capture fine-scale density variation and underestimates extended  
 952 dense regions. The conditional DDPMs produce more coherent predictions but still smooth out  
 953 localized dense structures. Overall, *Astro-DSB* exhibits better spatial continuity, structure recovery,  
 954 and dynamic range, demonstrating its superior robustness and generalization capability on real-world  
 955 astrophysical observations.

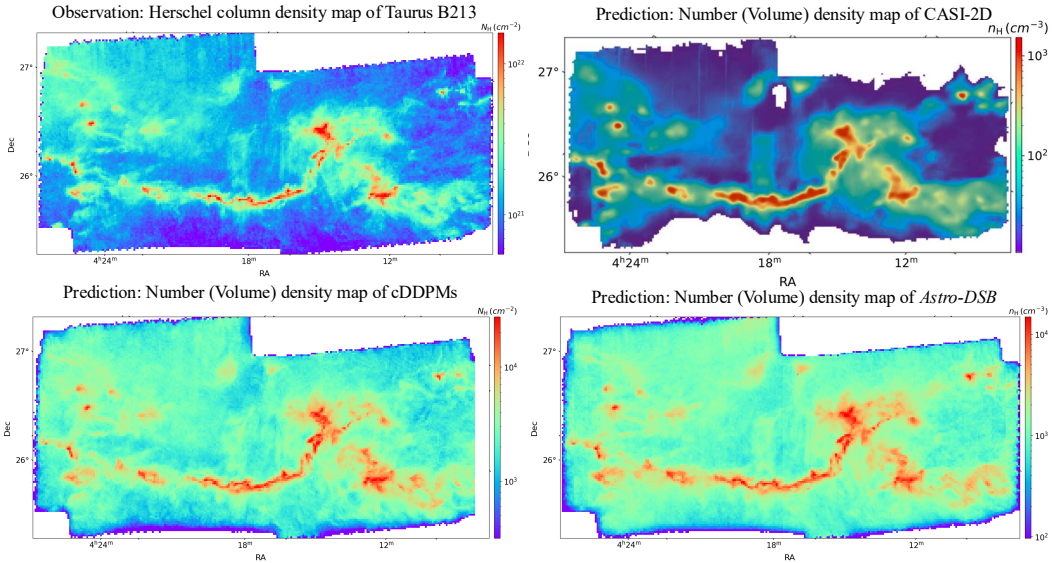


Figure 9: **Additional qualitative comparisons among different machine learning methods on Taurus B213 number density prediction.** CASI-2D (Xu et al., 2020a) is a previous ML method with U-Net architecture trained with discriminative reconstruction loss.