Exoplanet formation inference using conditional invertible neural networks

Anonymous Author(s)
Affiliation
Address
email

Abstract

The interpretation of the origin of observed exoplanets is usually done only qualitatively due to uncertainties of key parameters in planet formation models. To allow a quantitative methodology which traces back in time to the planet birth locations, we train recently developed conditional invertible neural networks (cINN) on synthetic data from a global planet formation model which tracks growth from dust grains to evolved final giant planets. In addition to deterministic single planet formation runs, we also include gravitationally interacting planets in multiplanetary systems, which include some measure of chaos. For the latter case, we treat them as individual planets or choose the two or three planets most likely to be discovered by telescopes. We find that training on multiplanetary data, each planet treated as individual point, is promising. The single-planet data only covers a small range of planets and does not extrapolate well to planet properties not included in the training data. Extension to planetary systems will require more training data due to the higher dimensionality of the problem.

1 Motivation

Planet formation theory has advanced thanks to the discovery of thousands of exoplanets. In an optimistic view, successful planet formation models exist and can reproduce the general outline of the exoplanet demographics. However, some planet formation parameters are notoriously difficult to constrain, such as a parameter characterizing turbulence in the disk, while others are known to vary significantly within a certain range, such as the initial disk gas and solid mass. It is therefore interesting to try to constrain these values by retrieving planet formation parameters from the exoplanet data (Mollière et al. 2022). Therefore, we attempt here a first step towards this goal, by starting with planetary mass and distance to the star as observables. The overarching goal is to apply the model to the exoplanet data to retrieve the likelihood of formation parameters given a model as well as to improve the quantitative interpretation of individual systems.

2 Methods

2.1 Physical model

To study planet formation from a global perspective, we use an established global model (described in detail by Emsenhuber et al. 2021) with extensions from Voelkel et al. (2020, 2022). We briefly summarize the key ingredients here. The model starts at an early stage once a relatively massive disk has formed with a mass fraction $M_{\mathrm{disk}}/M_{\star}$. The disk is mostly gaseous, with a percent-level dust content, that is, the solid disk mass $M_{\mathrm{solid,disk}}$, set by the dust-to-gas ratio. Its initial radius follows the empirical relation found by Tobin et al. (2020) for an early disk stage: $R_{\mathrm{disk}} =$

$70\,\mathrm{au} \times (M_{\mathrm{solid,disk}}/(100 M_{\oplus}))^{1/4}$. Another varied initial condition, is the disk inner edge defined by magnetospheric accretion onto the star.

A central free parameter controlling several aspects of planet formation is the non-dimensional viscous α parameter (Shakura & Sunyaev 1973), assumed to arise from turbulence but still poorly constrained. By assuming a value for α, the model evolves the one-dimensional viscous diffusion equation numerically (Pringle 1981). The parameter also influences disk temperatures and planetary migration.

The model evolves the solid material by balancing coagulation and fragmentation as well as radial drift driven by aerodynamic breaking of the orbiting particles (Birnstiel et al. 2012). Gas turbulence sets the relative speed of particles and thus whether they fragment. For both this process and vertical dust settling, we use a reduced value of $\alpha_{\mathrm{dust}} = 0.1\alpha$, not physically motivated but chosen to allow sufficient pebble accretion, which would otherwise be suppressed. It is however conceivable that turbulence on small scales and in vertical direction differs from the global, radial one (Lesur et al. 2023).

Radial drift implies a mass flux of solids, mainly centimeter-sized *pebbles*, from which a fraction of ϵ_{plts} (set to 0.01) is converted into larger bodies called *planetesimals* over a characteristic length scale of five disk scale heights (Lenz et al. 2019). Theory and Solar System evidence suggest planetesimals are at least 50 km in diameter (Polak & Klahr 2023), which we assumed here. The planetesimal size is important for planetesimal collision rates, which we treat in two stages. In the early runaway regime, we assume that a largest planetesimal – given a head-start with a larger initial diameter of 1320 km – accretes smaller planetesimals and pebbles (following Ormel 2017). Once it reaches 0.01 Earth masses, the body is promoted to a protoplanet in the disk model, from which point onward it influences planetesimal dynamics (Fortier et al. 2013), removes mass from the solid and gas disk via accretion, and interacts with other protoplanets gravitationally which is a source of chaos (using the symplectic N-body code by Chambers 1999). This two-stage process reduces computation for bodies with negligible mass.

The protoplanets can also accrete gas, where the rate is obtained from solving the one-dimensional interior structure of the gaseous envelope and limited by accretion rates obtained from detailed three-dimensional calculations (Bodenheimer et al. 2013). Ultimately, gas accretion ceases when the disk dissipates after several million years through so-called *photoevaporation*, thermal mass loss driven by heation from stellar X-ray (Picogna et al. 2019; Ercolano et al. 2021) and external ultraviolet irradiation (Haworth et al. 2018), with parameters listed in Table 1.

2.2 Data

Two datasets are generated with this set-up, one where at most one protoplanet forms and injection of further protoplanets is suppressed (single-planet case), and one which allows up to 100 protoplanets per disk (nominal). The initial conditions are varied in logarithmic space in the four dimensions listed in Table 1. For both cases, 1000 simulations were started. In some disks, the conditions did not allow a single protoplanet to grow to the threshold mass which results in 707 planets for the single-planet data and 15777 planets in 690 disks for the multi-planet case, where some simulations did not successfully complete due to known numerical issues. From the multi-planet case, we also extracted the two and three planets with the highest radial velocity signal imposed on their stars ordered by this radial velocity semi-amplitude. This gives us two additional datasets with 679 (two-planet) and 658 (three-planet) entries. We note that for training an invertible neural network with four input dimensions, these datasets are relatively small which makes the task challenging and further data generation is conceivable in the future if motivated by the early results presented below. For slight data augmentation, we re-draw noise from a Gaussian model with deviation of 0.01 in normalized units for both parameters (disk mass fraction, α, dust/gas ratio, and inner edge, all logarithmic) and observations (the planets' mass and semi-major axis, also logarithmic) at each training epoch.

2.3 Neural network design and training

In this work, we make use of conditional invertible neural networks (cINN) (Ardizzone et al. 2019, 2021) by extending the public Framework for Easily Invertible Architectures (FrEIA, , https://github.com/vislearn/FrEIA Ardizzone et al. 2018/2022). The structure of cINN deviates from invertible neural networks (INN) presented in Ardizzone et al. (2018) in that the training or

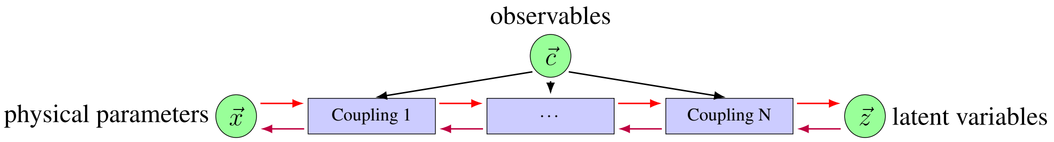

Figure 1: cINN information flow. Red: forward direction; Purple: inverse; Black: conditions (required for both). Observables $\vec{c}$, planetary mass and semi-major axis, are from the training, test, or real data; model parameters $\vec{x}$ are here M_{disk}, α, inner edge orbital period, and dust-to-gas ratio.

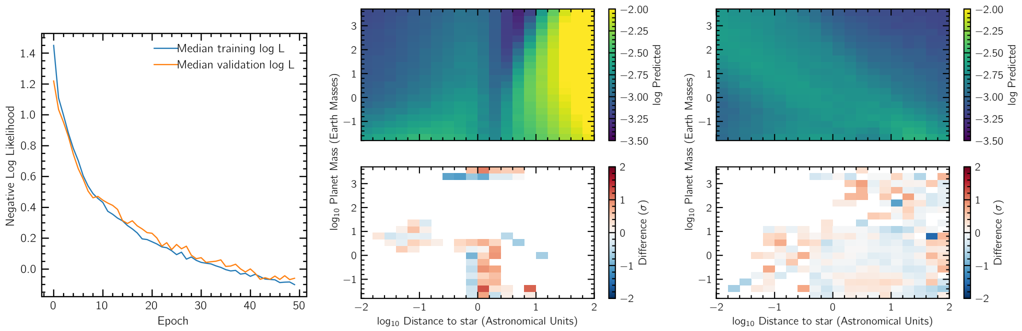

Figure 2: Loss function during training and comparison of predicted and training data on the viscous α. Left: Validation and training Log Likelihood as a function of training epoch. Center: single-planet histogram on inferred α; Right: as Center but for the nominal data. Difference of the posterior mean to training data is normalized by standard deviations as $\sigma = \sqrt{\sigma^2_{\text{train}} + \sigma^2_{\text{posteriors}}}$.

real data is not part of the mapping of parameters to observables of the INN, but instead, it is given to all conditional blocks in the network structure. Figure 1 schematically depicts the approach. A commodity of both approaches is that the model is trained to map parameters ($\vec{x}$) to a latent variable space $\vec{z}$ without physical meaning but with a distribution in $\vec{z}$ space following a multidimensional Gaussian with identity covariance matrix. Upon successful training, this property can be used to draw $\vec{z}$ values from the Gaussian, transverse the network in inverse direction under conditions of true observed systems and infer physical parameters.

The detailed structure of the network trained here follows Ksoll et al. (2020) based on GLOW-style (Kingma & Dhariwal 2018) affine coupling layers with conditioning. Compared to Ksoll et al. (2020), we reduced the network width and optimized training parameters for our data. We use 16 coupling blocks with random permutations between them, each containing 3 hidden layers with 8 units per layer and rectified linear unit activation, motivated by our low-dimensional $\vec{c}$ and $\vec{x}$. Training is performed by minimizing with the Adam optimizer ($\beta_1 = 0.8$ and $\beta_2 = 0.8$, learning rate starting at 0.001 and decreasing using the StepLR function with $\gamma = 0.99$ and step sizes of one epoch) the negative log likelihood from the forward passes as well as the mean squared errors between recovered parameters $\hat{\vec{x}}$ and true parameters $\vec{x}$. For the multiplanet sample treated as individual planets, a validation set of 10 batches of 64 points is visualized during training and learning parameters were tuned manually using this measure. A test set of 1577 points (10%) is held out from training and training parameter tuning completely until testing of the results.

3 Results

Visual inspection reveals that the loss function reaches its optimum after about 50 epochs, at which point we stopped training (left panel in Fig. 2). Accuracy is measured in several ways. First, maximum a posteriori estimates (Ksoll et al. 2020) are compared with the ground truth data from the test set with satisfying results. For example, for the nominal data, we find average deviations of 0.2 in standardized units, centered on zero. For all datasets, both the training data and the cINN results are insensitive (within the achieved accuracy) to the inner-edge orbital period; we therefore omit this parameter from the following discussion.

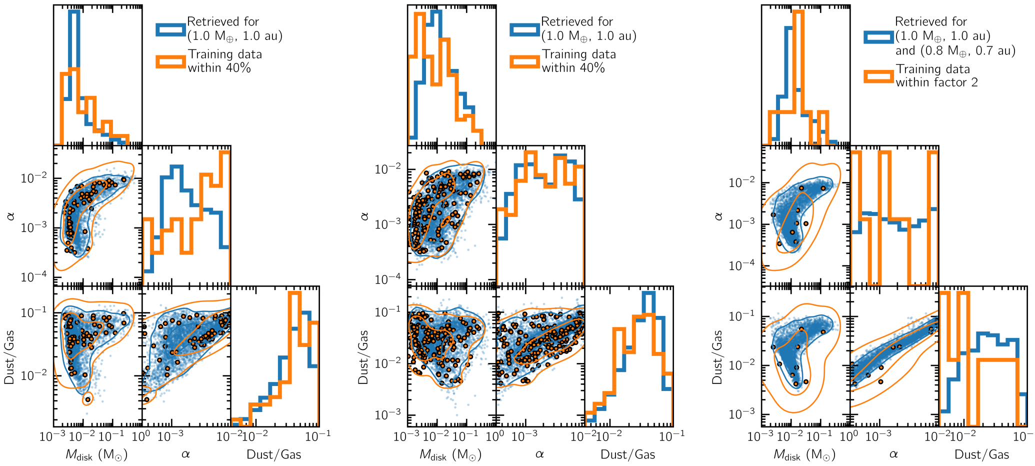

Figure 3: Retrieved and training disk parameters for Earth-analogues. Left: single-planet per disk; Center: nominal, Right: two-planet systems. Training data is shown if both the planet mass and semi-major axis lies within 40% of the Earth (factor 2 of an Earth + Venus system).

For the single-planet data, the observation parameter space, from which $\vec{c}$ are drawn, is poorly sampled, with typical planet formation outcomes clustering in restricted regions. The cINN therefore extrapolates to unsampled areas, as seen in the central panel of Figure 2: performance is reasonable where data exists, but at long orbital distances the model predicts large turbulence parameters. Such a dependency is not physically expected and the results from multiplanet (nominal) data, which better sample this space, does not show such behavior. More concerning than the extrapolation itself is that the posterior width narrows in this region. The model therefore predicts with confidence an unexpected and unwarranted extrapolation which is concerning in applications. Although this effect is weaker in other parameters, we conclude that single-planet simulations are unsuitable for building a robust model applicable to real exoplanet data which includes exoplanets at these locations.

The nominal data shows a more expected dependency on α. The cINN predicts a diagonal of enhanced values in distance against mass space (Fig. 2, right), which we attribute to the interplay of the effects of α on both migration (influencing more massive planets) and dust properties (determining the low-mass growth).

As a first application and further test, we let the models predict the disk parameters for selected planets and contrast them against the full test and training data in corner plots (e.g. for an Earth-like planet in Fig. 3). The retrievals appear qualitatively reasonable and reveal correlations between important parameters, $M_{\rm disk}$, α, and Dust/Gas ratio, accross all tested observation $\vec{c}$ combinations (0.1, 1, 10, and 4000 Earth mass planets at 0.1, 1, or 10 au). Comparisons with training data are limited to $\vec{c}$ regions with sufficient samples. For the two-planet systems (four dimensional $\vec{c}$), only a few hundred test cases are available, leading to sparse coverage and necessitating a broader range of included training data for the plot. The issue becomes more severe for three-planet systems (six-dimensional). Nevertheless, training seems to converge to an optimum, and the maximum a posteriori estimates remain similarly close to the test samples as in the two cases with individual planets (single-planet and nominal).

In summary, we conclude that a cINN can be trained on the outcomes of a global dust-to-planet formation model. Multiplanet simulations sample more evenly the parameter space and yield improved results and applicability. In these simulations, the introduced chaos is not detrimental and might even lead to a more robust training, indicating that the imprint of disk parameters survives planet-planet interactions. To accommodate planetary system data, focusing on the most observable planets is practical, and data generation efficiency could be improved by raising the mass threshold at which protoplanets are individually resolved.

References

Ardizzone, L., Bungert, T., Draxler, F., et al. 2018/2022, Framework for Easily Invertible Architectures (FrEIA)

Ardizzone, L., Kruse, J., Lüth, C., et al. 2021, Conditional Invertible Neural Networks for Diverse Image-to-Image Translation

Ardizzone, L., Kruse, J., Wirkert, S., et al. 2018, Analyzing Inverse Problems with Invertible Neural Networks

Ardizzone, L., Lüth, C., Kruse, J., Rother, C., & Köthe, U. 2019, Guided Image Generation with Conditional Invertible Neural Networks

Birnstiel, T., Klahr, H., & Ercolano, B. 2012, A&A, 539, A148

Bodenheimer, P., D'Angelo, G., Lissauer, J. J., Fortney, J. J., & Saumon, D. 2013, ApJ, 770

Chambers, J. E. 1999, MNRAS, 304, 793

Emsenhuber, A., Mordasini, C., Burn, R., et al. 2021, A&A, 656, A69

Ercolano, B., Picogna, G., Monsch, K., Drake, J. J., & Preibisch, T. 2021, MNRAS, 508, 1675

Fortier, A., Alibert, Y., Carron, F., Benz, W., & Dittkrist, K.-M. 2013, A&A, 549, A44

Haworth, T. J., Clarke, C. J., Rahman, W., Winter, A. J., & Facchini, S. 2018, MNRAS, 481, 452

Hunter, J. D. 2007, Computing in Science & Engineering, 9, 90

Kingma, D. P. & Dhariwal, P. 2018, Glow: Generative Flow with Invertible 1x1 Convolutions

Ksoll, V. F., Ardizzone, L., Klessen, R., et al. 2020, MNRAS, 499, 5447

Lenz, C. T., Klahr, H., & Birnstiel, T. 2019, ApJ, 874, 36

Lesur, G., Ercolano, B., Flock, M., et al. 2023, in Protostars and Planets VII, Astronomical Society of the Pacific Conference Series, ed. Inutsuka, S. and Aikawa, Y. and Muto, T. and Tomida, K. and Tamura, M., Vol. 534, Kyoto, 465

Mollière, P., Molyarova, T., Bitsch, B., et al. 2022, ApJ, 934, 74

Mordasini, C. 2014, A&A, 572, A118

Ormel, C. W. 2017, in Astrophysics and Space Science Library, Vol. 445, Formation, Evolution, and Dynamics of Young Solar Systems, ed. {Pessah}, Martin and {Gressel}, Oliver,, 197

Paszke, A., Gross, S., Massa, F., et al. 2019, PyTorch: An Imperative Style, High-Performance Deep Learning Library

Picogna, G., Ercolano, B., Owen, J. E., & Weber, M. L. 2019, MNRAS, 487, 691

Polak, B. & Klahr, H. 2023, ApJ, 943, 125

Pringle, J. E. 1981, Annual Review of Astronomy and Astrophysics, 19

Shakura, N. I. & Sunyaev, R. A. 1973, A&A, 24, 337

Tobin, J. J., Sheehan, P. D., Megeath, S. T., et al. 2020, ApJ, 890, 130

Voelkel, O., Klahr, H., Mordasini, C., & Emsenhuber, A. 2022, A&A, 666, A90

Voelkel, O., Klahr, H., Mordasini, C., Emsenhuber, A., & Lenz, C. 2020, A&A, 642, A75

Waskom, M. 2021, Journal of Open Source Software, 6, 3021

A Technical Appendices and Supplementary Material

We provide a table of physically relevant parameters for the generation of the dataset in Table 1. These parameters are required for reproducibility of our results since they partially deviate from prior published works.

Table 1: Key physical model parameters

Varied parameters (uniform sampling)		
$\log_{10} M_{\rm disk}/M_\star$	Gas disk mass fraction	[-3,-0.5]
$\log_{10} \alpha$	Viscous parameter	[-3.5,-2]
$\log_{10} M_{\rm dust}/M_{\rm gas}$	Dust/Gas ratio	[-2.4,-1]
$\log_{10} P_{\rm in}$	Inner edge orbital period	[0, 1.3]*
Constants		
$N_{\rm p,max}$	Maximum number of planets	[1,100]
$M_\star$	Stellar mass	1 Solar mass
$R_{\rm plts}$	Planetesimal radius	25 km
$R_{\rm l,plt}$	Largest planetesimal radius	660 km
$v_{\rm frag}$	Pebble fragmentation velocity	1 m/s
$L_{\rm X}$	X-ray luminosity for disk photoevaporation	5×10^{28} erg/s
F_{UV}	External ultraviolet field strength	10 G_0†
$\epsilon_{\rm plts}$	Planetesimal formation efficiency	0.01
$T_{\rm min}$	Minimum disk temperature	10 K
$f_{\rm opa}$	Opacity reduction factor in planet envelope	0.003‡
$\rho_{\rm dust}$	Dust bulk density	1.675 g/cm^3
a_0	Dust monomer (minimum) size	10^{-5} cm

*Corresponds to a range from 1 to 20 days

†Measured in local interstellar FUV (912-2400 Å) radiation field strength $G_0 = 1.6 \times 10^{-3}$ erg/(s cm^2)

‡Following the dust growth argument by Mordasini (2014)