# ExoMDN: Rapid characterization of exoplanet interior structures with mixture density networks ${ }^{\star, \star \star}$ 

Philipp Baumeister ${ }^{1,2}$ (1) and Nicola Tosi ${ }^{1}{ }^{(0)}$<br>${ }^{1}$ Institute of Planetary Research, German Aerospace Center (DLR), Rutherfordstraße 2, 12489 Berlin, Germany e-mail: philipp.baumeister@dlr.de<br>${ }^{2}$ Department of Astronomy and Astrophysics, Technische Universität Berlin, Hardenbergstraße 36, 10623 Berlin, Germany

Received 22 February 2023 / Accepted 14 June 2023


#### Abstract

Aims. Characterizing the interior structure of exoplanets is essential for understanding their diversity, formation, and evolution. As the interior of exoplanets is inaccessible to observations, an inverse problem must be solved, where numerical structure models need to conform to observable parameters such as mass and radius. This is a highly degenerate problem whose solution often relies on computationally expensive and time-consuming inference methods such as Markov chain Monte Carlo. Methods. We present ExoMDN, a machine-learning model for the interior characterization of exoplanets based on mixture density networks (MDN). The model is trained on a large dataset of more than 5.6 million synthetic planets below 25 Earth masses consisting of an iron core, a silicate mantle, a water and high-pressure ice layer, and a $\mathrm{H} / \mathrm{He}$ atmosphere. We employ log-ratio transformations to convert the interior structure data into a form that the MDN can easily handle. Results. Given mass, radius, and equilibrium temperature, we show that ExoMDN can deliver a full posterior distribution of mass fractions and thicknesses of each planetary layer in under a second on a standard Intel i5 CPU. Observational uncertainties can be easily accounted for through repeated predictions from within the uncertainties. We used ExoMDN to characterize the interiors of 22 confirmed exoplanets with mass and radius uncertainties below 10 and $5 \%$, respectively, including the well studied GJ 1214 b , GJ 486 b , and the TRAPPIST- 1 planets. We discuss the inclusion of the fluid Love number $k_{2}$ as an additional (potential) observable, showing how it can significantly reduce the degeneracy of interior structures. Utilizing the fast predictions of ExoMDN, we show that measuring $k_{2}$ with an accuracy of $10 \%$ can constrain the thickness of core and mantle of an Earth analog to $\approx 13 \%$ of the true values.


Key words. planets and satellites: interiors - planets and satellites: composition - methods: numerical - methods: statistical

## 1. Introduction

In the past decade, the number of discovered exoplanets has been growing rapidly, with more than 5000 planets confirmed to date. Characterizing the interior structures of these planets, that is, the size and mass of their main compositional reservoirs, is critical to understanding the processes that govern their formation, evolution, and potential to support life (Spiegel et al. 2014; Van Hoolst et al. 2019). Numerical models are commonly used to compute interior structures that fit to observed mass and radius of the planet (e.g., Sotin et al. 2007; Valencia et al. 2007; Fortney et al. 2007; Wagner et al. 2011; Zeng \& Sasselov 2013; Unterborn \& Panero 2019; Baumeister et al. 2020; Huang et al. 2022). However, unlike planets in the Solar System for which a wealth of observational data is available ranging from geodetic observations to in situ seismic measurements, for exoplanets, mass and radius are often the only parameters that can be determined. As a result, the interior structure is highly degenerate, with many qualitatively different interior compositions that can match the observations equally well (Rogers \& Seager 2010; Dorn et al. 2015, 2017b; Brugger et al. 2017). Probabilistic inference methods, such as Markov chain Monte Carlo (MCMC) sampling, are

[^0]regularly utilized to obtain a comprehensive picture of possible planetary interiors, while also taking into account observational uncertainties (Rogers \& Seager 2010; Dorn et al. 2015, 2017b; Dorn \& Heng 2018; Acuña et al. 2021). Given prior estimations of interior parameters, probabilistic inference methods allow for the determination of posterior probabilities that best fit the observations. However, in general, MCMC methods are computationally intensive and time-consuming, requiring calculations of hundreds of thousands of interior structure models. The interior inference of a single exoplanet can therefore take from hours to days. Furthermore, a dedicated framework combining both a forward interior structure model and an MCMC scheme is necessary, which can limit the large-scale applicability of these techniques due to the need for specialized expertise in planetary interior modeling. To fully exploit the ever-increasing number of exoplanet detections, a fast alternative to MCMC inference is needed.

Here, we present ExoMDN, a standalone machine-learningbased model that is capable of providing a full inference of the interior structure of low-mass exoplanets in under a second without the need for a dedicated interior model. We have made both the trained models and the training routines available in a GitHub repository ${ }^{1}$. The purpose of ExoMDN is to provide a rapid first general characterization of an exoplanet interior, which can then be investigated further with more detailed, specialized models.

[^1]
## 2. Machine Learning for interior characterization

In recent years, machine-learning based methods have become increasingly relevant in planetary science because of their ability to facilitate and speed up otherwise very time-consuming calculations. Deep neural networks in particular have been applied to the detection of transits (Chaushev et al. 2019; Malik et al. 2021; Valizadegan et al. 2022), for atmospheric retrievals (Márquez-Neila et al. 2018; Zingales \& Waldmann 2018; Himes et al. 2022), in geodynamic simulations (Atkins et al. 2016; Agarwal et al. 2021a,b), in planet formation models (Alibert \& Venturini 2019; Cambioni et al. 2019; Emsenhuber et al. 2020; Auddy et al. 2022), as well as for characterizing exoplanet interiors (Baumeister et al. 2020; Zhao \& Ni 2021; Haldemann et al. 2023).

In an earlier work (Baumeister et al. 2020), we presented a proof-of-concept method to characterize exoplanet interiors using mixture density networks (MDNs; Bishop 1994), which can predict the full probability distribution of parameters by approximating these with a linear combination of Gaussian kernels. We trained an MDN to infer the range of plausible thicknesses of compositional layers in a planet based on mass and radius inputs. However, this was not a full characterization of the interior, as our network could only predict the marginals of the posterior distribution. While this gives an accurate estimation of the range of admissible parameter values, it does not allow us to pinpoint specific interior structures that fit observed mass and radius nor determining correlations between the various layers. For this purpose, the prediction of the full, multidimensional posterior distribution is required.

ExoMDN builds upon our previous work and is capable of providing a full inference of the entire posterior distribution of interior structures for a planet in a fraction of a second (e.g., on a standard Intel i5 CPU). In addition, we include the equilibrium temperature of the planet as an input parameter to the network in addition to mass and radius, by improving on the atmosphere and water layers in the underlying forward model used to generate the training data. In particular, we used the full water phase diagram compiled by Haldemann et al. (2020) in place of the previous simple isothermal, high-pressure ice layer and we modeled an isothermal atmosphere instead of the previous zero-temperature approach. The use of the equilibrium temperature thus implicitly includes the orbital distance as an observable parameter. We further improved the robustness of the underlying forward model at high pressures by adopting updated high-pressure equations of state (EoS) for the silicate mantle and iron core. A comparison of the old and new forward models can be found in Fig. B.5.

We used our interior model to first generate a dataset of $\approx 5.6$ million synthetic planets spanning the desired parameter space of interior structures, planet masses, and equilibrium temperatures. We then trained a mixture density network to predict the parameters of a mixture of multivariate normal distributions, with the aim of approximating the posterior distribution for a given set of input parameters, namely, mass, radius, and equilibrium temperature. In order for the MDN to handle multidimensional predictions, we applied log-ratio transformations on the training data to convert the interior structures into new coordinates that the MDN can easily handle. We present two trained models: one trained on planetary mass, radius, and equilibrium temperature and the second including the fluid Love number $k_{2}$ as an additional input. Fluid Love numbers describe the shape of a rotating planet in hydrostatic equilibrium. The seconddegree Love number $k_{2}$ is particularly interesting for exoplanet interior characterization, as it depends solely on the interior
density distribution (Kellermann et al. 2018; Padovan et al. 2018; Baumeister et al. 2020). In a body with $k_{2}=0$, the entire mass is concentrated in the center, while $k_{2}=1.5$ corresponds to a fully homogeneous body. For a number of exoplanets, $k_{2}$ is potentially measurable through either second-order effects on the shape of the transit light curve (Hellard et al. 2019; Akinsanmi et al. 2019), or through the apsidal precession of the orbit (Csizmadia et al. 2019).

## 3. Methods

### 3.1. Interior model

We compute planetary interior structures with our code TATOOINE (Baumeister et al. 2020; MacKenzie et al. 2023). Each planet consists of compositionally distinct layers. The model takes as input the planet mass, $M_{\mathrm{p}}$, the mass fractions of each layer, $w_{i}$, and the equilibrium temperature, $T_{\text {eq }}$ (defined at the top of the atmosphere). From the top of the planet toward the center, the model calculates radial profiles of mass, $m$, pressure, $P$, and density, $\rho$, by solving the equations for mass conservation (1a), hydrostatic equilibrium (1b), as well as the equation of state (EoS, 1c) relating pressure, density, temperature, $T$, and composition, $c$ :
$\begin{aligned} \frac{\mathrm{d} m(r)}{\mathrm{d} r} & =4 \pi r^{2} \rho(r), \\ \frac{\mathrm{d} P(r)}{\mathrm{d} r} & =-\frac{G m(r) \rho(r)}{r^{2}},\end{aligned}$
$P(r)=f(\rho(r), T(r), c(r))$,
where $G$ is the gravitational constant. The planet radius, $R_{\mathrm{p}}$, is iteratively adjusted until the mass at the planet center approaches zero. This yields a final planet radius and the radius fractions of each layer $d_{i}$. We fix the pressure at the top of the atmosphere to 10 mbar . We focus here on planets below $25 M_{\oplus}$. We consider four distinct layers: an iron core, a silicate mantle, a water layer, and an $\mathrm{H} / \mathrm{He}$ atmosphere.

### 3.1.1. Iron core

We assumed that the core consists of pure, solid, hcp-iron. We used the temperature-dependent, high-pressure EoS by Bouchet et al. (2013) for pressures below 234.4 GPa. At higher pressures, we switch to the high-pressure EoS from Hakim et al. (2018), valid up to 10 TPa .

The presence of lighter elements in the core such as sulfur or hydrogen can significantly reduce the density of the core, which in turn can have large effects on the core size and consequently on the planet radius (Hakim et al. 2018). The amount of lighter elements in an exoplanet's core is hard to constrain, as it not only depends on the initial abundances in the protoplanetary disk, but also on the processes of core formation and magma ocean cooling (Hirose et al. 2021). For a proper treatment of the interior inference, the amount of light elements should be taken as a free parameter, which will increase the degeneracy of interior structures even more. For simplicity and to better illustrate our method, here we neglect the presence of lighter elements in the core, following an approach commonly used in the exoplanet community (e.g., Seager et al. 2007; Wagner et al. 2011). Nevertheless, we note that the uncertainty in the light elements budget of the core can be easily incorporated into our method by sampling from a range of core compositions upon creating the training dataset (Sect. 3.4).

### 3.1.2. Silicate mantle

The silicate layer consists of an upper mantle composed of olivine $(\mathrm{Mg}, \mathrm{Fe})_{2} \mathrm{SiO}_{4}$ and pyroxene $\left(\mathrm{Mg}, \mathrm{Fe}_{2}\right)_{2} \mathrm{Si}_{2} \mathrm{O}_{6}$, a lower mantle composed of magnesiowüstite ( $\mathrm{Mg}, \mathrm{Fe)O}$ and bridgmanite $(\mathrm{Mg}, \mathrm{Fe}) \mathrm{SiO}_{3}$, and a high-pressure phase of magnesiowüstite and post-perovskite. The transition from upper to lower mantle is assumed to occur at a fixed pressure of 23 GPa , and the transition to post-perovskite at a pressure of:
$P(T)=89.184 \mathrm{GPa}+13.3 \mathrm{MPa} \mathrm{K}^{-1} T$,
following Tateno et al. (2009), where $T$ is the adiabatic temperature in the mantle. We model the upper and lower mantle with a modified Tait EoS from Holland \& Powell (2011) and the postperovskite phase with the generalized Rydberg EoS described in Wagner et al. (2011).

Similar to the composition of the core, the mantle composition could be varied, for example, using stellar abundances as proxies of planet composition (e.g., Dorn et al. 2017a; Hinkel \& Unterborn 2018). However, since the relation between star and planet composition is not straightforward (Plotnykov \& Valencia 2020), for simplicity, we assume the silicate mantle to have an Earth-like composition with a molar $\mathrm{Mg} / \mathrm{Si}$ ratio of 1.131 and a magnesium number (Mg\#) of 0.9 (Sotin et al. 2007). The $\mathrm{Mg} / \mathrm{Si}$ ratio determines the mixing ratio of the respective mantle minerals, with the Mg \# determining the ratio of the respective Mg and Fe end members.

### 3.1.3. Water layer

For the water layer, we used the tabulated AQUA EoS from Haldemann et al. (2020), spanning a wide temperature and pressure range and including gas, liquid, and solid water phases. Liquid and solid layers are assumed to be fully convective, with an adiabatic temperature profile calculated from the adiabatic gradient given by the AQUA table for any given temperature and pressure. Water vapor is assumed to be part of an isothermal atmosphere at the equilibrium temperature.

### 3.1.4. $\mathrm{H} / \mathrm{He}$ atmosphere

The low densities of many exoplanets hint at extended primordial envelopes composed of hydrogen and helium (e.g., Jontof-Hutter 2019). We therefore include an outer gaseous $\mathrm{H} / \mathrm{He}$ envelope of solar-like composition ( $71 \%$ hydrogen, $29 \%$ helium by mass) based on the EoS from Saumon et al. (1995). We treat the atmosphere as isothermal with a temperature equal to the equilibrium temperature, an approach also employed, for example, by Dorn et al. (2017b) and Zeng et al. (2019). While an isothermal atmosphere certainly does not capture the full complexities of exoplanet atmospheres, more detailed atmosphere models would require the inclusion of additional parameters such as infrared and optical opacities, as well as the intrinsic temperature of the planet (e.g., Guillot 2010). Since the goal of this work is to explore the machine-learning method and its applications, we have chosen a simplified atmosphere model to limit the overall model complexity of the model and of the training data. However, more complex atmosphere models, in particular those specifically designed to treat gas giant planets (e.g., Fortney et al. 2007; Nettelmann et al. 2011; Leconte \& Chabrier 2012), can be easily incorporated into ExoMDN by producing suitable training data.

### 3.2. Compositional data

Ideally, an inference model for planets should provide a set of desired parameters that fully describe the interior, such as the thickness or mass of each interior layer. These quantities represent a type of compositional data, where the sum of the $D$ components is always constant (e.g., the planet radius or mass). In our case, we are interested in the relative mass and thickness of each layer in the planet, so that
$\sum_{i}^{D} x_{i}=1$,
where $x_{i}$ is the relative thickness or mass of the $i$ th planet layer. This restricted space is known as the simplex $\mathcal{S}^{D}$, and is commonly represented in the form of a ternary diagram (e.g., Rogers \& Seager 2010).

The nature of compositional data can make the statistical treatment cumbersome. The constraints imposed by Eq. (3) can give rise to spurious correlations. The shapes of probability distributions can be distorted and skewed, and trying to fit distributions to sample data may lead to points lying outside the simplex (Aitchison 1982; Pawlowsky-Glahn \& Egozcue 2006). In particular, this means that Gaussian distributions, which are commonly used to represent continuous data, cannot be utilized directly to describe distributions of compositional data, as parts of the distribution would fall outside the closed space. This point is especially relevant for this work: the simple parameterization of Gaussian distributions makes them convenient candidates for components in mixture distributions in order to approximate arbitrary posterior distribution with neural networks, as the entire mixture is described by only a few parameters. It is therefore highly desirable to extend the usefulness of Gaussian mixtures to the analysis of compositional data, while retaining the convenience of their simple parameterization.

One solution is to introduce a set of coordinate changes called log-ratio transformations (Aitchison 1982), which transforms the data coordinates from the simplex into (unconstrained) real space by way of logarithmic ratios between coordinates. We focus here on the additive log-ratio transformation alr: $\mathcal{S}^{D} \rightarrow$ $\mathbb{R}^{D-1}$, which takes the logarithm of pairwise ratios between $D-1$ coordinates and an arbitrarily chosen $D$ th coordinate $\left(x_{D}\right)$, thereby reducing the dimension of the new space by one:
$\operatorname{alr}\left(x_{i}\right)=y_{i}=\ln \frac{x_{i}}{x_{D}}$.
The back-transformation onto the simplex is given by (Aitchison 1982)
$\operatorname{alr}^{-1}\left(y_{i}\right)=x_{i}= \begin{cases}\frac{\exp \left(y_{i}\right)}{1+\sum_{j}^{D} \exp \left(y_{j}\right)} & (i=1, \ldots, D-1), \\ \frac{1}{1+\sum_{j}^{D} \exp \left(y_{j}\right)} & (i=D) .\end{cases}$

### 3.3. Mixture density networks

Neural networks are a widely used tool in machine learning due to their ability to learn complex, nonlinear mappings between input variables $\mathbf{x}$ and output variables $\mathbf{t}$. Neural networks can model this mapping by learning from a set of training data which provide concrete examples of the output values corresponding to each set of input values. Conventionally, neural networks are


Fig. 1. Schematic overview of the MDN architecture and inference procedure.
trained by minimizing the mean squared error between known values from the training data and predicted outputs from the neural network. However, this approach tends to be wholly inadequate for inverse problems, where one set of input values may correspond to multiple output values, or more generally, to some posterior probability density $p(\mathbf{t} \mid \mathbf{x})$, (i.e., the probability density of $\mathbf{t}$ given some input $\mathbf{x}$ ). To preserve the practicality of neural networks and extend their functionality to include arbitrary probability functions, Bishop (1994) introduced a class of neural networks called mixture density networks, which combine a conventional neural network with a mixture density model. The posterior $p(\mathbf{t} \mid \mathbf{x})$ can be approximated by a linear combination of $m$ kernel functions $\phi_{i}(\mathbf{t} \mid \mathbf{x})$,
$p(\mathbf{t} \mid \mathbf{x})=\sum_{i=1}^{m} \alpha_{i}(\mathbf{x}) \phi_{i}(\mathbf{t} \mid \mathbf{x})$,
where $\alpha_{i}$ are mixture weights. Various functions can be chosen for $\phi_{i}(\mathbf{t} \mid \mathbf{x})$. We focus here on a mixture model with Gaussian kernels of the form:

$$
\begin{align*}
\phi_{i}(\mathbf{t} \mid \mathbf{x})= & \frac{1}{(2 \pi)^{c / 2} \operatorname{det}\left(\boldsymbol{\Sigma}_{i}\right)^{1 / 2}} \\
& \times \exp \left\{-\frac{1}{2}\left(\mathbf{t}-\boldsymbol{\mu}_{i}\right)^{\top} \boldsymbol{\Sigma}_{i}^{-1}\left(\mathbf{t}-\boldsymbol{\mu}_{i}\right)\right\}, \tag{7}
\end{align*}
$$

where $c$ is the dimension of $\mathbf{t}$ (i.e., the number of output variables), and $\mu_{i}$ is the center of the $i$ th Gaussian kernel with a diagonal covariance matrix, $\Sigma_{i}$ :
$\boldsymbol{\Sigma}_{i}=\operatorname{diag}\left(\boldsymbol{\sigma}_{i}\right)=\left[\begin{array}{lll}\sigma_{i, 1} & & \\ & \ddots & \\ & & \sigma_{i, c}\end{array}\right]$.

The conditional probability distribution $p(\mathbf{t} \mid \mathbf{x})$ is completely described by weights, $\alpha_{i}$, means, $\boldsymbol{\mu}_{i}$, and variance, $\boldsymbol{\sigma}_{i}$. Training the MDN to predict these outputs therefore allows for the distribution to be reconstructed. With $m$ mixtures and $c$ parameters, the total number of network outputs is $(2 c+1) m$. A mixture density network is built as a conventional feedforward neural network, where the last layer approximates the distribution parameters (Fig. 1). The model can then be trained with a maximum likelihood approach by minimizing the average negative

Table 1. Prior distributions of model parameters for training data generation.

| Parameter | Range | Shape |
| :--- | :--- | :--- |
| $M_{\mathrm{p}}$ | $0.1-25 M_{\oplus}$ | Uniform |
| $T_{\text {eq }}$ | $100-1000 \mathrm{~K}$ | Uniform |
| $w_{\text {Core }}$ | $0-1$ | Uniform (simplex) |
| $w_{\text {Mantle }}$ | $0-1$ | Uniform (simplex) |
| $w_{\text {Water }}$ | $0-1$ | Uniform (simplex) |
| $w_{\text {Gas }}$ | $10^{-8}-1$ | Uniform in $\ln w_{\text {Gas }}$ (simplex) |

$\log$-likelihood $\mathcal{L}$ across the training data set:

$$
\begin{align*}
\mathcal{L} & =-\frac{1}{N} \sum_{k=1}^{N} \ln \mathcal{L}_{k} \\
& =-\frac{1}{N} \sum_{k=1}^{N} \ln \left(\sum_{i=1}^{m} \alpha_{i}\left(\mathbf{x}_{k}\right) \phi_{i}\left(\mathbf{t}_{k} \mid \mathbf{x}_{k}\right)\right), \tag{9}
\end{align*}
$$

where $N$ is the size of the training data set.

### 3.4. Training data and network architecture

We created a data set of 5.6 million synthetic planets randomly sampled from the prior distributions, summarized in Table 1. The planet mass was chosen uniformly between 0.1 and $25 M_{\oplus}$. Each planet was set at a specific equilibrium temperature ranging from 100 to 1000 K . The mass fraction of each planetary layer was sampled from the simplex so that they add to one. The gas mass fraction $w_{\text {Gas }}$ was sampled logarithmically with a lower limit of $10^{-8}$, while the other mass fractions were sampled uniformly. Given these inputs, the TATOOINE model calculates planet radius and thickness of each layer. For each planet, we also calculated the fluid Love number, $k_{2}$, using the matrix-propagator approach from Padovan et al. (2018).

To prepare the training data, we log-ratio transformed both mass fractions and radius fractions according to Eq. (4) using the core mass and radius as a base coordinate ( $x_{D}$ in Eq. (4)). This is the key difference to our previous work, which enables the prediction of multivariate distributions of mass and radius fractions. The log-ratio transformation enforces the condition that the mass and radius fractions add up to one and allows the network to
operate in unbounded real space, $\mathbb{R}$, instead of the simplex, $\mathcal{S}$, which permits the use of Gaussian kernels as described above.

As a preprocessing step before training and prediction, we also log-transformed the planet mass, which we found to slightly improve the training performance. We took $70 \%$ of the data set for training, using the remaining $30 \%$ to evaluate the performance of the MDN during training. In addition, we retained a small set of data for final model validation (see Sect. 4).

We trained the MDN to predict the parameters of the posterior distributions of the log-ratio-transformed mass fractions ( $\left.\ln w_{\text {Mante }} / w_{\text {Core }}, \ln w_{\text {Water }} / w_{\text {Core }}, \ln w_{\text {Gas }} w_{\text {Core }}\right)$ and radius fractions $\left(\ln d_{\text {Mante }} / d_{\text {Core }}, \ln d_{\text {Water }} / d_{\text {Core }}, \ln d_{\text {Gas }} / d_{\text {Core }}\right)$. We trained two models with different sets of inputs: Model 1 with $M_{\mathrm{p}}, R_{\mathrm{p}}, T_{\text {eq }}$ and Model 2 with $M_{\mathrm{p}}, R_{\mathrm{p}}, T_{\text {eq }}, k_{2}$.

The MDN is built from a feedforward neural network using the Keras framework (Chollet et al. 2015) and TensorFlow (Abadi et al. 2016), with the MDN output layer adapted from Martin \& Duhaime (2019). The best MDN architecture was found through hyperparameter optimization using the KerasTuner framework (O'Malley et al. 2019). We optimized for the number of hidden layers, the number of units per layer, the learning rate, as well as the batch size. We kept the number of mixture components fixed at $m=50$, because we noticed that the tuner would always optimize for the highest available number of mixtures, but with very small mixture weights for most components. We found 50 components to be a good middle ground where training accuracy was good, but without too many components contributing little to the posterior distribution.

The architecture that yielded the best training performance for Model 1 consists of three hidden layers with 384 nodes per layer with a batch size of 750 . For Model 2, the best architecture consists of three hidden layers with 896 nodes per layer with a batch size of 1000 . Models with a base learning rate of 0.001 performed best in both cases.

Each hidden layer is activated with a rectified linear unit (ReLU), which is a commonly used activation function in deep learning models (Nair \& Hinton 2010; Goodfellow et al. 2017). To ensure that the variances are always positive, we activated $\sigma_{i}$ in the output layer with a nonnegative exponential linear unit (NNELU) after Brando (2017):
$\operatorname{NNELU}(x)= \begin{cases}x+1 & \text { for } x \geq 0, \\ \exp (x) & \text { for } x<0 .\end{cases}$
The nodes for mixture weights, $\alpha_{i}$, and means, $\boldsymbol{\mu}_{i}$, are activated with a linear function to allow for unrestricted output values. To avoid overfitting, we applied an early stopping of the learning algorithm once the validation loss did not improve for eight consecutive training epochs. To improve training performance, we reduced the learning rate by a factor of ten every time the validation loss stopped improving for more than four epochs during training, down to a lower bound for the learning rate of $10^{-8}$. This helps fine-tune the model weights once a near-optimal set of parameters has been learned. The MDN was trained on a GPU workstation with eight NVIDIA RTX A5000 graphics cards. The (wall clock) training time for a model was around three hours.

### 3.5. Backtransformation to mass and radius fractions

From the predicted parameters of the MDN, the approximate posterior distribution of the log-ratio transformed mass and radius fractions corresponding to the given inputs of the MDN can be reconstructed according to Eq. (6). The log-ratio space
is not particularly useful for interpreting the inferred interior structure distributions. However, the back-transformation of the Gaussian mixture onto the simplex (Eq. (5)) is mathematically unwieldy, as the normal distributions are highly deformed when in the compositional space. Instead, we randomly sample a sufficiently large number of points from the log-ratio posterior probability distribution and transform these back into compositional space. This is conceptually similar to MCMC sampling and gives a good approximation of the posterior distribution.

### 3.6. Incorporating measurement uncertainties

The current network architecture is built on the assumption that the input parameters are known exactly without uncertainties. However, except for Solar System planets, observations of exoplanets will always come with considerable measurement uncertainties. With ExoMDN, measurement uncertainties can be taken into account in a straightforward way by repeatedly sampling $n$ times from within the error bars of the input parameters, predict the interior distribution for each sample, and combining the results into a single posterior distribution. This can be either done via summing up each Gaussian mixture in log-ratio space and then normalizing the resulting distribution, or by first subsampling from each prediction $n^{\prime}$ times and then merging the samples (for a final dataset size of $n \times n^{\prime}$ samples). Subsampling first and then merging is considerably less memory and processing intensive, as the full posterior distribution can be built up sequentially from each planet sample. Summing up all predicted posterior distributions first requires loading the entire posterior distribution, consisting of $n \times m$ multivariate Gaussian kernels, into memory. Sampling from this mixture distribution can be computationally very expensive for large sample sizes, $n$, which are needed to treat the measurement uncertainties well. We find that both approaches display no functional difference in the predicted full posterior distributions (Fig. B.4). We therefore chose the approach of sampling first from each prediction and then merging, as it is also easy to implement in the current prediction pipeline. However, $n^{\prime}$ should be chosen significantly smaller than $n$ to avoid oversampling of specific mass-radius-temperature pairs.

## 4. Validation

To establish the accuracy of the trained MDN, we validated it in two ways: by forward modeling and by independent inference. In the first case, we used the predicted mass fractions as inputs to the forward model and recomputed the interior structures of planets to investigate how well the planet radius can be retrieved from the predictions. This allows us to put constraints on systematic errors in the MDN outputs. In the second case, we ensured that the MDN predictions are accurate and consistent with other inference methods by comparing the predicted posterior distributions with those obtained by an independent inference approach.

### 4.1. Radius accuracy

We used the MDN to predict the interior structures distributions of 500 randomly selected planets out of the test data set. We took 200 samples of interior structures for each prediction and model these planets with the TATOOINE forward model by taking the mass fractions of the layers as inputs (i.e., 10000 sample points in total). We then compared the relative error $\frac{\Delta R}{R_{\mathrm{p}}}=\frac{\left(R_{\text {val }}-R_{\mathrm{p}}\right)}{R_{\mathrm{p}}}$


Fig. 2. Radius accuracy of Model 1 after recalculating the planet interior based on the MDN prediction. Panel a shows the distribution of the relative radius error of 10000 sample points. The blue line marks the median, with the blue area showing the range where $80 \%$ of values lie. Panels b-e show the standard deviation in relative radius errors $\sigma$ for a variety of planet parameters: bulk density (b), planet mass (c), equilibrium temperature (d), and average atmosphere thickness $\bar{d}_{\text {Gas }}$ of recomputed planet samples (e). Each point represents one of 500 planets from the test data set (see the text for more details).
between the true planet radius $R_{\mathrm{p}}$ and the recomputed planet radius $R_{\text {val }}$ obtained from the MDN predictions (Fig. 2).

We find that the recomputed planet radii of Model 1 fit closely to the expected ones, with the MDN introducing a slight overestimation of the radius of about $0.4 \%$ (Fig. 2a). The MDN introduces a small amount of noise into the recomputed radii, with $80 \%$ of planets having a radius error of less than $1.5 \%$. The MDN does not perform equally well across the parameter space. Low-density planets tend to have a wider spread in radius errors (Fig. 2b), largely independent of planet mass (Fig. 2c). Higher equilibrium temperatures increase the error slightly (Fig. 2d). We attribute the larger errors mainly to the atmosphere. The recomputed radius errors tend to be the largest in planets with extensive gas envelopes (Fig. 2e). Small errors in the prediction of the gas mass fraction are amplified into larger radius errors due to the low density of the atmosphere. In addition, the transformation from log-ratios to mass and radius fractions amplifies any small uncertainty present in the atmosphere-core log-ratio predictions.

For Model 2, we find that the planetary radii can be reproduced very accurately with a relative radius error of less than $0.55 \%$ (Fig. 3a). As with Model 1, the radius is slightly overestimated by $0.4 \%$. We additionally check how well the fluid Love number $k_{2}$ is reproduced by computing the relative error $\frac{\Delta k_{2}}{k_{2}}=\frac{\left(k_{2, \text { val }}-k_{2}\right)}{k_{2}}$, where $k_{2 \text {,val }}$ is the Love number of the validation planet to be reproduced, and $k_{2}$ is the fluid Love number calculated from the predicted interior mass fractions. We find that $k_{2}$ is reproduced well, with more than $80 \%$ of the points falling within $2.3 \%$ of the true $k_{2}$ value (Fig. 3b). While the median of the data set sits at zero error, the data set is slightly skewed toward low $k_{2}$ values. This is most likely caused by the atmosphere. The fluid Love number, $k_{2}$, is highly sensitive to the density structure of the planet, especially in the upper layers. Slight overestimations of the atmosphere mass fractions result in larger underestimations of $k_{2}$.

### 4.2. Independent inference

We randomly selected 20 planets from a test dataset that the MDN did not see during training and ran an independent


Fig. 3. Radius accuracy (a) and $k_{2}$ accuracy (b) of model 2 after recalculating the planet interior based on the MDN prediction. Panel a shows the distribution in the relative radius error, panel b of the relative $k_{2}$ error of 10000 sample points. Blue lines mark the median, with the blue areas showing the range where $80 \%$ of values lie.
inference of their interior structures using a straightforward Monte-Carlo sampling method, assuming a radius uncertainty of $1 \%$. We assessed how well the predicted posterior distributions, $P$, fit to the posterior distributions from the validation set, $Q$, by calculating the Hellinger distance $H(P, Q)$ for each marginal distribution following the approach by Haldemann et al. (2023). The Hellinger distance is an integrated metric bounded between 0 and 1 that measures the similarity of two probability distributions. Two identical probability distributions have a Hellinger distance of 0 , while a Hellinger distance of 1 is reached when there is no overlap between the two distributions. We binned the data into $n=20$ bins with sample frequencies $p_{i}$ and $q_{i}$ in each bin. The (squared) Hellinger distance is then given by
$H^{2}(P, Q)=\frac{1}{2} \sum_{i}^{n}\left(\sqrt{p_{i}}-\sqrt{q_{i}}\right)^{2}$.
The average Hellinger distance $\bar{H}$ over the 20 validation planets is shown in Table 2 for both the log-ratio outputs from the MDN and the transformed compositional mass and radius fractions. We find that the predicted log-ratio distributions compare

Table 2. Average Hellinger distance $\bar{H}$ for 20 randomly selected validation planets of all MDN (log-ratio) output distributions and of their corresponding transformed parameters.

| Parameter | $\bar{H}$ |
| :--- | :---: |
| $\ln d_{\text {Mantle }} / d_{\text {Core }}$ | $1.02 \times 10^{-3}$ |
| $\ln d_{\text {Water }} / d_{\text {Core }}$ | $8.84 \times 10^{-4}$ |
| $\ln d_{\text {Gas }} / d_{\text {Core }}$ | $1.30 \times 10^{-3}$ |
| $\ln w_{\text {Mantle }} / w_{\text {Core }}$ | $8.03 \times 10^{-4}$ |
| $\ln w_{\text {Water }} / w_{\text {Core }}$ | $6.87 \times 10^{-4}$ |
| $\ln w_{\text {Gas }} / w_{\text {Core }}$ | $6.07 \times 10^{-4}$ |
| $d_{\text {Core }}$ | $4.35 \times 10^{-3}$ |
| $d_{\text {Mantle }}$ | $2.91 \times 10^{-3}$ |
| $d_{\text {Water }}$ | $3.31 \times 10^{-3}$ |
| $d_{\text {Gas }}$ | $5.03 \times 10^{-3}$ |
| $w_{\text {Core }}$ | $2.72 \times 10^{-3}$ |
| $w_{\text {Mantle }}$ | $2.76 \times 10^{-3}$ |
| $w_{\text {Water }}$ | $3.84 \times 10^{-3}$ |
| $w_{\text {Gas }}$ | $2.16 \times 10^{-2}$ |



Fig. 4. Example of the Hellinger distances for a well-predicted planet ( $4.722 M_{\oplus}, 1.82 R_{\oplus}$ ) from the 20 validation planets. The blue line marks the independent validation, the orange line shows the ExoMDN prediction.
very well to the validation set, with Hellinger distances around $1 \times 10^{-3}$. This corresponds to two normal distributions differing in their means by about $3 \times 10^{-3}$ units (assuming a standard deviation of 1 ), or in the standard deviation by $0.2 \%$ (assuming the same mean). Figure 4 shows an example for a well predicted validation planet with small Hellinger distances.

The transformed mass and radius fractions also fit well, albeit with slightly higher Hellinger distances around $3 \times 10^{-3}$. The gas mass fraction $w_{\text {gas }}$ is the least well constrained parameter here with $\bar{H}=2.16 \times 10^{-2}$. This mirrors the effect already discussed in Sect. 4.1.

## 5. Results

### 5.1. Earth and Neptune

We demonstrate the ability of ExoMDN to perform an interior characterization of Earth and Neptune by treating them as if they were exoplanets, where only the mass and radius are measured, and the equilibrium temperature is set according to their orbital distance. Earth represents the archetypical rocky planet whose internal structure is best known of all the planets in the Solar System. Neptune lies on the upper end of the mass range


Fig. 5. Predicted log-ratios of the thickness of interior layers for an Earth-like planet with $1 M_{\oplus}$ and $1 R_{\oplus}$. The ellipses in the top right plots mark the location and covariance of each of the 50 Gaussian kernels, with the colors showing the mixture weight of each kernel.
we investigate and is a representative example of volatile-rich planets.

The MDN prediction takes the form of a six-dimensional distribution of the log-ratios of masses and thicknesses of the planetary layers, which can be transformed back to layer mass and thickness, as described in Sect. 3.5. For clarity, we will focus in this section only on the thickness of the layers. Figures showing the mass fractions can be found in the appendix.

Figure 5 shows the posterior distribution of log-ratios for Earth, as approximated by the MDN, given Earth's mass, radius, and equilibrium temperature of 255 K . The ellipses in the upper right plots show the location and covariance of each Gaussian kernel, with the colors marking the respective mixture weights, $\alpha_{i}$. The kernels are well spaced with little overlap and most mixture weights are similar. This indicates that the MDN is able to efficiently leverage all its 50 kernels to construct the posterior distribution.

We sampled 200000 points from the log-ratio distribution to construct the posterior distribution of the actual layer thickness, which are shown in Fig. 6. Given only three observables (mass, radius, and equilibrium temperature), the prediction indicates a mostly rocky planet composed of an iron core which makes up at least $50 \%$ of the planet, and with only little water and gas. The colored symbols mark end-member interior structures with only two layers. In this case, only three of these exist, namely: (1) the actual structure of the Earth with an iron core making up $55 \%$ of the interior, and a silicate mantle on top; (2) an iron-water planet composed of a relative core size of $73 \%$ and an ice layer; and (3) an iron-gas planet composed of a massive iron core making up $80 \%$ of the planet, and a $\mathrm{H} / \mathrm{He}$ envelope taking up the rest. In cases 2 and 3 , the iron core needs to be very large to compensate for the low density of the water and gas layers. Although these two cases are probably not likely to occur in nature, they demonstrate that the interior cannot be fully constrained without additional constraints. In this example, the thickness of the silicate mantle in particular can barely be constrained.

It should be noted that due to our assumption of uniform priors, the most commonly predicted interior structures encompass


Fig. 6. Predicted thickness of interior layers for an Earth-like planet with $1 M_{\oplus}$ and $1 R_{\oplus}$. The colored points mark possible end-member compositions, which are illustrated on the right. The red circle corresponds to Earth's true interior structure. The diagonal plots show the marginal distributions of each layer, with the blue dashed lines marking the median value and the dotted lines the 5th and 95th percentiles.
a combination of all four layers. For this reason, the distributions presented should not be understood as definitive probabilities, but rather as the number of potential solutions for each given layer thickness fraction. Consequently, the actual interior of Earth lies outside the bulk of the predicted distribution. In fact, only a single solution exists that matches Earth's mass and radius with only an iron core and a silicate mantle, and no water or extended atmosphere.

For Neptune, the MDN predicts a substantial atmosphere between 20 and $70 \%$ of the planet's radius (Fig. 7) and only a small iron core ( $\leq 40 \%$ of the radius). We note here that rather than the actual temperature of 51 K , we used an equilibrium temperature of 100 K , which is the lowest temperature for which the AQUA EoS used for the water layer is valid. The predicted interior structures lie well within previously published results, which generally agree on Neptune having a small iron-silicate core of about $20 \%$ of the planet's radius and an atmosphere of about $30-40 \%$, with a water-rich envelope in between (e.g., Hubbard et al. 1991; Podolak et al. 1995; Nettelmann et al. 2013; Neuenschwander \& Helled 2022).

### 5.2. Application to exoplanets

One of the main advantages of using neural networks for interior structure inference over other methods such as MCMC sampling is the speed at which the posterior distribution can be obtained. The inference process for MCMC sampling can take several hours per individual planet (e.g., Haldemann et al. 2023). The MDN can perform the same prediction in fractions of a second. In addition, the MDN model is optimized for bulk processing of inputs owing to the Keras framework, allowing for multiple

Table 3. Average inference time for different numbers of planets.

| Planets | $t_{\mathrm{MDN}}(\mathrm{s})$ | $t_{\text {sampling }}(\mathrm{s})$ | Total (s) | Time/ <br> planet (ms) |
| :--- | :--- | :--- | :--- | :--- |
| 1 | 0.265 | 0.470 | 0.734 | 734.079 |
| 10 | 0.257 | 0.524 | 0.781 | 78.126 |
| 100 | 0.257 | 0.978 | 1.236 | 12.356 |
| 1000 | 0.314 | 4.790 | 5.104 | 5.104 |

Notes. $t_{\text {MDN }}$ shows the time needed to retrieve the mixture parameters from the MDN; $t_{\text {sampling }}$ shows the time needed to sample 1000 points from each input planet and transform from log-ratio space to compositional data. For each row, we performed ten inferences and averaged the computation times.
input samples to be predicted simultaneously. Between 1 and 1000 input data points, we find little difference in the computation time needed by the MDN for a prediction ( $t_{\mathrm{MDN}}$, Table 3). The sampling from the predicted distribution and transformation to mass and radius fractions (see. Sect. 3.5) takes up most of the time ( $t_{\text {sampling }}$ ). Even so, predicting and sampling a thousand planets is possible in under six seconds on a conventional laptop processor ${ }^{2}$. In fact, the main limitation to predicting a large number of planets simultaneously is the amount of available computer memory.

These fast prediction times mean that interior structures can be inferred for every exoplanet for which mass, radius, and equilibrium temperature are known. To demonstrate this, we selected planets from the NASA Exoplanet Archive ${ }^{3}$ that lie in the parameter space of our training data (Table 1) and for which upper and lower mass and radius uncertainties are given. We used ExoMDN to infer the interior structure of each planet, incorporating the mass, radius, and equilibrium temperature uncertainties according to Sect. 3.6. For each planet, we sampled 5000 mass, radius, and equilibrium temperature points from within a normal distribution given by the uncertainties, and predicted the posterior distribution for each point. From each of these posterior distributions, we then generated an additional ten random samples. In total, this yields 50000 samples of interior structures per planet, forming the full posterior distribution and spanning the range of measurement uncertainties.

Table A. 1 shows the 22 planets from this dataset where mass uncertainties are below $10 \%$ and radius uncertainties are below $5 \%$. This includes the well studied planets GJ 1214 b , GJ 486 b , and the TRAPPIST-1 planets, among others. For each planet, we provide the predicted median thickness of each interior layer, alongside the ranges in which $90 \%$ of the solutions are found. The total time to produce this data was $\approx 30 \mathrm{~s}$. A more extensive data set of 75 planets with radius and mass uncertainties of 10 and $20 \%$, respectively, including both mass fractions and thickness, is available online at the CDS.

Upcoming exoplanet missions such as PLATO (Rauer et al. 2014) will significantly increase the number of exoplanets with well-determined masses and radii. PLATO in particular will allow for the radii of Earth-sized planets to be determined within an accuracy of up to $3 \%$, while follow-up ground based observations are expected to constrain the mass of these planets with an accuracy of $10 \%$ or better. We can leverage the fast prediction times of ExoMDN to investigate the degree to which the

[^2]

Fig. 7. Predicted thickness of interior layers for a Neptune-like planet with $17.1 M_{\oplus}$ and $3.865 R_{\oplus}$. The diagonal plots show the marginal distributions of each layer, with the blue dashed lines marking the median value and the dotted lines the 5th and 95 percentiles. ${ }^{(*)}$ Instead of Neptune's equilibrium temperature of 51 K , a value of 100 K was used to be in line with the parameter range of the training data.
interior of a planet could be constrained based on the accuracy of the mass and radius determination. Taking Earth and Neptune as examples, we imposed a $10 \%$ mass uncertainty and varied the radius uncertainty between 1 and $20 \%$ (Fig. 8). As above, in each case we sample 10000 times from within mass and radius uncertainties, and take ten random samples from each predicted posterior distribution for a total of 100000 samples.

We find that with a radius accuracy of $3 \%$, the core radius fraction of Earth can be constrained to $d_{\text {Core }}=0.69_{-0.14}^{+0.10}$ (error bars are the 5 th and 95 th percentiles, respectively), which is close to the value we found assuming a perfect knowledge of mass and radius (Fig. 6). With a radius accuracy of $10 \%$, the core size is significantly less well constrained ( $d_{\text {Core }}=$ $0.68_{-0.27}^{+0.20}$. Similarly, with a radius accuracy of $3 \%$, the predicted atmosphere thickness of a Neptune analog (Fig. 8b) is $d_{\text {Gas }}=$ $0.45_{-0.12}^{+0.12}$, which is again close to the value obtained assuming no error in mass and radius (Fig. 7). With a radius accuracy of $10 \%$, the uncertainty of $d_{\text {Gas }}$ grows to $d_{\text {Gas }}=0.45_{-0.21}^{+0.16}$. Increasing the radius accuracy has little effect on the possibility to constrain the layers below the atmosphere. Due to the low density of the atmosphere, different planet radii can be easily accommodated by small changes in atmosphere mass without significantly affecting the other layers. In a sense, the presence of a large atmosphere obscures the inference of the thickness of the deeper layers.

The radius accuracy controls to a large extent the uncertainties in the predicted thickness of the various layers. For completeness, we show in Fig. B. 3 predictions of Earth- and Neptune-like interiors obtained when both radius and mass accuracies are varied simultaneously (from 1 to $20 \%$ and from 3 to $40 \%$, respectively). Indeed, the inferred structures are very similar to those we obtained by fixing the mass accuracy to $10 \%$.


Fig. 8. Effect of radius uncertainty, $\delta R$, on the ability to constrain the interior for Earth (a) and Neptune (b) analogs. Each panel shows the marginal distributions for each interior layer with increasing amounts of radius uncertainty. An uncertainty of $10 \%$ in mass and $2 \%$ in $T_{\text {eq }}$ has been assumed for both planets in all cases. ${ }^{(*)}$ Instead of Neptune's equilibrium temperature of 51 K , a value of 100 K was used to be in line with the parameter range of the training data.

### 5.3. Constraining the interior with $k_{2}$

Mass, radius, and equilibrium temperature alone are not sufficient to fully constrain the interior of a planet, as demonstrated above. The fluid Love number $k_{2}$ is a potential direct link from observation to interior structure, as it only depends on the density distribution in the planet. This stands in contrast to for example the elemental abundances of the host star, which may be representative of the bulk abundances of the planet and its atmosphere (Dorn et al. 2015, 2017a; Brugger et al. 2017; Spaargaren et al. 2020), but which necessitate additional assumptions about the planet formation and evolution history.

Figure 9 shows the MDN prediction of the interior of Earth, given knowledge of Earth's value of $k_{2}=0.933$ (Lambeck 1980). With this added information, the MDN is capable of fully constraining Earth's actual interior (particular in comparison to Fig. 6). In fact, the constraints from $k_{2}$ are strong enough that the composition of the iron core becomes important. The planets in the training data are modeled with a pure iron core, while Earth has about $10-15 \mathrm{wt} . \%$ of lighter elements in its core (Poirier 1994). Thus, the MDN predicts a smaller core radius of $51 \%$, while the true core size is about $54.5 \%$ of the total radius. In practice, of course, the measurements $k_{2}$ for exoplanets will be associated with considerable uncertainties. Constraining a planet's interior to the degree shown in Fig. 9 is therefore unlikely in the near future. Nevertheless, we can utilize the fast predictions of the MDN to estimate the accuracy that would be needed to properly constrain the interior. We performed a number of predictions for Earth and Neptune analogs with increasing


Fig. 9. Predicted thickness of interior layers for an Earth-like planet with $1 M_{\oplus}$ and $1 R_{\oplus}$, where also $k_{2}$ is known ( $k_{2}=0.933$ ) in addition to mass and radius. The red circle marks Earth's true interior structure. The diagonal plots show the marginal distributions of each layer, with the blue dashed lines marking the median value and the dotted lines the 5th and 95th percentiles. Compared to Fig. 6, the axis limits have been adjusted to better show the model results.
$k_{2}$ uncertainties, in addition to mass, radius, and equilibrium temperature uncertainties of 5,3 , and $2 \%$, respectively). These values are representative of a very well studied and characterized exoplanet, which would likely be needed for an accurate measurement of the Love number. For Neptune, we take a value of $k_{2}=0.392$, which we calculated after Hubbard (1984) from the gravitational moment $J_{2}=3.40843 \times 10^{-3}$ (Jacobson 2009). The predicted results are detailed in Fig. 10. As the uncertainty in $k_{2}$ grows, the interior of both planets becomes less and less constrained. We find that with a $k_{2}$ uncertainty of $10 \%$, Earth's core and mantle thickness could be constrained to about $\pm 13 \%$ of their actual value (within the 5 and $95 \%$ percentiles). With a $k_{2}$ uncertainty of $20 \%$, mantle and core can be constrained to within $\pm 17 \%$. Even with large $k_{2}$ uncertainties, Earth could be clearly identified as a rocky planet with very little water and a thin atmosphere. The uncertainties of mass and radius put a limit on how well the interior can be determined. With the given mass and radius uncertainties, we find that in the Earth-like case, $k_{2}$ uncertainties lower than $10 \%$ do not constrain the interior further. For Neptune, a $10 \%$ uncertainty in $k_{2}$ could help constrain the atmospheric thickness to $22_{-13}^{+10} \%$ of the planet's radius.

## 6. Discussion and conclusions

MDNs can provide a reliable way to rapidly characterize the interior structure of exoplanets within fractions of a second. Compositional data such as mass fractions of individual interior layers can be easily accommodated in the network by using log-ratio transformations. An additional benefit of a machine learning approach over other inference methods is that the forward model computations are decoupled from the actual inference process.


Fig. 10. Effect of $k_{2}$ uncertainty on the ability to constrain the interior for Earth (a) and Neptune (b) analogs. Each panel shows the marginal distributions for each interior layer with increasing amounts of $k_{2}$ uncertainty. An uncertainty of 5 in mass, 3 in radius, and $2 \%$ in $T_{\text {eq }}$ has been assumed for both planets in all cases. ${ }^{(*)}$ Instead of Neptune's equilibrium temperature of 51 K , a value of 100 K was used to be in line with the parameter range of the training data.

The training data are calculated separately before training and the training process encodes the information from the training data into the network weights. The trained network itself is standalone and interior inferences can be performed without requiring the training data, a dedicated interior model, a separate inference scheme, or prior expertise about exoplanet interior modeling. This stands in contrast to MCMC sampling, where running the data-generating forward model during the inference is an integral part of exploring the posterior distributions.

While in this work the training data were generated from a single forward model, this is actually not necessary for the training of the network. Since the forward model data generation is separated from the inference itself, different parts of the data set can be modeled by different dedicated forward models, for example to include both Jupiter-like and low-mass planets which may require different modeling approaches. Importantly, this means that the training data can be computed, collected, and combined from multiple sources without much overhead and without the need to integrate different numerical codes into a single model, as would be needed for MCMC sampling. Furthermore, this means that this method is easily extendable to different models and applicable to other inverse problems. However, the necessary prior generation of training data locks the model assumptions of the forward model into the training data. Changing the forward model therefore requires computing a new set of training data and training of a new neural network. This may be a drawback if the forward model assumptions often change (e.g., with different atmosphere compositions).

Our method is best suited for problems where the number of constraining parameters is relatively small. The required number of training samples increases (potentially exponentially)
with each additional parameter, a phenomenon which has been termed the "curse of dimensionality" (Bellman 1966). This can make the generation and handling of training data cumbersome and time-consuming for larger numbers of constraining parameters, as it is generally necessary to sample from the entire investigated parameter space to achieve good MDN performance. A potential way to alleviate this issue could be to generate the training data "on the fly" and train the network with an incremental learning approach (van de Ven et al. 2022), where the network learns continuously with new incoming data, thus reducing the need to save large amounts of data.

Conditional invertible neural networks (cINN) may be another alternative, as demonstrated by Haldemann et al. (2023). These potentially work better with higher-dimensional data while requiring comparatively less training data, with the tradeoff that the network setup is more complex and predictions are generally slower.

As with other machine learning methods, the nature of the training process introduces a small amount of intrinsic noise into the model. However, we have shown that the errors introduced by this are generally small (see Sect. 4.1), particularly for exoplanets where uncertainties in the observable quantities are relatively large.

The file size of the fully trained model is only $4 \approx 6.8 \mathrm{MB}$, which facilitates sharing and distribution online. The posterior distributions predicted by ExoMDN provide a first characterization of newly observed planets, which can then be further explored with dedicated models. ExoMDNs posterior distributions could also be employed as advanced priors for MCMC inferences based on more sophisticated forward models to help speed up their convergence. We believe that ExoMDN is a valuable tool for the exoplanet science community to gain access to a rapid first characterization of the possible interiors of low-mass planets.

Acknowledgements. We thank Heike Rauer for important suggestions and discussions on the role of uncertainties in mass and radius, and an anonymous referee for their comments, which helped improve a previous version of the manuscript We acknowledge the support of the DFG priority program SPP 1992 "Exploring the Diversity of Extrasolar Planets" (TO 704/3-1) and of the research unit FOR 2440 "Matter under planetary interior conditions" (PA 3689/1-1). Training of ExoMDN was performed on a GPU Workstation also sponsored by the DFG research unit FOR 2440 (grant number RE 882/19-2), which is gratefully acknowledged.

## References

Abadi, M., Agarwal, A., Barham, P., et al. 2016, ArXiv e-prints [arXiv:1603.04467]
Acuña, L., Deleuil, M., Mousis, O., et al. 2021, A\&A, 647, A53
Agarwal, S., Tosi, N., Kessel, P., Breuer, D., \& Montavon, G. 2021a, Phys. Rev Fluids, 6, 113801
Agarwal, S., Tosi, N., Kessel, P., et al. 2021b, Earth and Space Science, 8, 4 Aitchison, J. 1982, J. R. Stat. Soc. Ser. B Methodol., 44, 139
Akinsanmi, B., Barros, S. C. C., Santos, N. C., et al. 2019, A\&A, 621, A117
Alibert, Y., \& Venturini, J. 2019, A\&A, 626, A21
Atkins, S., Valentine, A. P., Tackley, P. J., \& Trampert, J. 2016, Phys. Earth Planet. Int., 257, 171
Auddy, S., Dey, R., Lin, M.-K., Carrera, D., \& Simon, J. B. 2022, ApJ, 936, 93
Baumeister, P., Padovan, S., Tosi, N., et al. 2020, ApJ, 889, 42
Bellman, R. 1966, Science, 153, 34
Bishop, C. M. 1994, Mixture Density Networks, Tech. Rep., Aston University
Bouchet, J., Mazevet, S., Morard, G., Guyot, F., \& Musella, R. 2013, Phys. Rev B, 87, 094102
Brando, A. 2017, Master's Thesis, Universitat Politecnica de Catalunya, Spain
Brugger, B., Mousis, O., Deleuil, M., \& Deschamps, F. 2017, ApJ, 850, 93

Cambioni, S., Asphaug, E., Emsenhuber, A., et al. 2019, ApJ, 875, 40
Chaushev, A., Raynard, L., Goad, M. R., et al. 2019, MNRAS, 488, 5232
Chollet, F. et al. 2015, Keras, https://keras.io/
Csizmadia, Sz., Hellard, H., \& Smith, A. M. S. 2019, A\&A, 623, A45
Dorn, C., \& Heng, K. 2018, ApJ, 853, 64
Dorn, C., Khan, A., Heng, K., et al. 2015, A\&A, 577, A83
Dorn, C., Hinkel, N. R., \& Venturini, J. 2017a, A\&A, 597, A38
Dorn, C., Venturini, J., Khan, A., et al. 2017b, A\&A, 597, A37
Emsenhuber, A., Cambioni, S., Asphaug, E., et al. 2020, ApJ, 891, 6
Fortney, J. J., Marley, M. S., \& Barnes, J. W. 2007, ApJ, 659, 166
Goodfellow, I., Bengio, Y., \& Courville, A. 2017, Deep Learning (Cambridge MIT Press)
Guillot, T. 2010, A\&A, 520, A27
Hakim, K., Rivolidini, A., Van Hoolst, T., et al. 2018, Icarus, 313, 61
Haldemann, J., Alibert, Y., Mordasini, C., \& Benz, W. 2020, A\&A, 643, A105
Haldemann, J., Ksoll, V., Walter, D., et al. 2023, A\&A, 672, A180
Hellard, H., Csizmadia, S., Padovan, S., et al. 2019, ApJ, 878, 119
Himes, M. D., Harrington, J., Cobb, A. D., et al. 2022, Planet. Sci. J., 3, 91
Hinkel, N. R., \& Unterborn, C. T. 2018, ApJ, 853, 83
Hirose, K., Wood, B., \& Vočadlo, L. 2021, Nat. Rev. Earth Environ., 2, 645
Holland, T. J. B., \& Powell, R. 2011, J. Metamorph. Geol., 29, 333
Huang, C., Rice, D. R., \& Steffen, J. H. 2022, MNRAS, 513, 5256
Hubbard, W. B. 1984, Planetary Interiors (New York, N.Y.: Van Nostrand Reinhold)
Hubbard, W. B., Nellis, W. J., Mitchell, A. C., et al. 1991, Science, 253, 648
Jacobson, R. A. 2009, AJ, 137, 4322
Jontof-Hutter, D. 2019, Ann. Rev. Earth Planet. Sci., 47, 141
Kellermann, C., Becker, A., \& Redmer, R. 2018, A\&A, 615, A39
Lambeck, K. 1980, The Earth's Variable Rotation: Geophysical Causes and Consequences, Cambridge Monographs on Mechanics (Cambridge: Cambridge University Press)
Leconte, J., \& Chabrier, G. 2012, A\&A, 540, A20
MacKenzie, J., Grenfell, J. L., Baumeister, P., et al. 2023, A\&A, 671, A65
Malik, A., Moster, B. P., \& Obermeier, C. 2021, MNRAS, 513, 5505
Márquez-Neila, P., Fisher, C., Sznitman, R., \& Heng, K. 2018, Nat. Astron., 2, 719
Martin, C., \& Duhaime, D. 2019, https://zenodo.org/record/2578015
Nair, V., \& Hinton, G. E. 2010, in Proceedings of the 27th International Conference on International Conference on Machine Learning, ICML' 10 (Omnipress), 807
Nettelmann, N., Fortney, J. J., Kramm, U., \& Redmer, R. 2011, ApJ, 733, 2
Nettelmann, N., Helled, R., Fortney, J. J., \& Redmer, R. 2013, Planet. Space Sci., 77, 143
Neuenschwander, B. A., \& Helled, R. 2022, MNRAS, 512, 3124
O'Malley, T., Bursztein, E., Long, J., et al. 2019, https://github.com/ keras-team/keras-tuner
Padovan, S., Spohn, T., Baumeister, P., et al. 2018, A\&A, 620, A178
Pawlowsky-Glahn, V., \& Egozcue, J. J. 2006, Geol. Soc. London Spec. Pub., 264, 1
Plotnykov, M., \& Valencia, D. 2020, MNRAS, 499, 932
Podolak, M., Weizman, A., \& Marley, M. 1995, Planet. Space Sci., 43, 1517
Poirier, J.-P. 1994, Phys. Earth Planet. Interiors, 85, 319
Rauer, H., Catala, C., Aerts, C., et al. 2014, Exp. Astron., 38, 249
Rogers, L. A., \& Seager, S. 2010, ApJ, 712, 974
Saumon, D., Chabrier, G., \& van Horn, H. M. 1995, ApJS, 99, 713
Seager, S., Kuchner, M., Hier-Majumder, C., \& Militzer, B. 2007, ApJ, 669, 1279
Sotin, C., Grasset, O., \& Mocquet, A. 2007, Icarus, 191, 337
Spaargaren, R. J., Ballmer, M. D., Bower, D. J., Dorn, C., \& Tackley, P. J. 2020, A\&A, 643, A44
Spiegel, D. S., Fortney, J. J., \& Sotin, C. 2014, Proc. Natl. Acad. Sci., 111, 12622
Tateno, S., Hirose, K., Sata, N., \& Ohishi, Y. 2009, Earth Planet. Sci. Lett., 277, 130
Unterborn, C. T., \& Panero, W. R. 2019, J. Geophys. Res. Planets, 124, 1704
Valencia, D., Sasselov, D. D., \& O'Connell, R. J. 2007, ApJ, 665, 1413
Valizadegan, H., Martinho, M. J. S., Wilkens, L. S., et al. 2022, ApJ, 926, 120
van de Ven, G. M., Tuytelaars, T., \& Tolias, A. S. 2022, Nat. Mach. Intell., 4, 1185
Van Hoolst, T., Noack, L., \& Rivoldini, A. 2019, Adv. Phys. X, 4, 1630316
Wagner, F., Sohl, F., Hussmann, H., Grott, M., \& Rauer, H. 2011, Icarus, 214, 366
Zeng, L., \& Sasselov, D. 2013, PASP, 125, 227
Zeng, L., Jacobsen, S. B., Sasselov, D. D., et al. 2019, Proceedings of the Natl. Acad. Sci., 116, 9723
Zhao, Y., \& Ni, D. 2021, A\&A, 650, A177
Zingales, T., \& Waldmann, I. P. 2018, AJ, 156, 268
Appendix A: Exoplanet predictions
Table A.1. Predicted thicknesses $d_{i}$ of interior layers for exoplanets with well known masses and radii.
Notes: Here, we give the median of each predicted distribution and the $5 \%$ and $95 \%$ percentiles. For planets with nonsymmetric mass or radius uncertainties, the larger value was used. Separate references for $T_{\text {eq }}$ are given in parentheses. For planets where no uncertainty of $T_{\text {eq }}$ was given, an uncertainty of $2 \%$ is assumed. A more extensive data set of 75 planets with radius and mass uncertainties of $10 \%$ and $20 \%$, respectively, including both mass fractions and thickness, is available in electronic form at the CDS.

## Appendix B: Additional figures



Fig. B.1. Predicted mass fraction of interior layers for an Earth-like planet with $1 M_{\oplus}$ and $1 R_{\oplus}$. The red circle corresponds to Earth's true interior structure. Due to the presence of lighter elements in Earth's core, the actual core mass of Earth $33 \%$ is slightly lower than what is predicted by ExoMDN for a solution with no water and atmosphere (39\%). The diagonal plots show the marginal distributions of each layer, with the blue dashed lines marking the median value and the dotted lines the 5th and 95th percentiles.


Fig. B.2. Predicted mass fraction of interior layers for a Neptune-like planet with $17.1 M_{\oplus}$ and $3.865 R_{\oplus}$. The diagonal plots show the marginal distributions of each layer, with the blue dashed lines marking the median value and the dotted lines the 5th and 95 th percentiles. ${ }^{(*)}$ Instead of Neptune's equilibrium temperature of 51 K , a value of 100 K was used to be in line with the parameter range of the training data.


Fig. B.3. Effect of radius $(\delta R)$ and mass $(\delta M)$ uncertainty on the ability to constrain the interior for Earth (a) and Neptune (b) analogs. Each panel shows the marginal distributions for each interior layer with increasing amounts of radius and mass uncertainty. A $T_{\text {eq }}$ uncertainty of $2 \%$ has been assumed for both planets. ${ }^{(*)}$ Instead of Neptune's equilibrium temperature of 51 K , a value of 100 K was used to be in line with the parameter range of the training data.


Fig. B.4. Comparison of the two possible approaches to incorporate measurement uncertainties (Sect. 3.6) for an Earth-like planet with 5\% radius and $10 \%$ mass uncertainty. 5000 posterior distributions were predicted from random mass, radius, and $T_{\text {eq }}$ inputs within the uncertainties. The red line shows the predicted thickness of each interior layer obtained by first summing up all 5000 posterior distributions and then taking 5000 random samples from the combined mixture, while the dark blue, light blue, and green lines show those obtained by first taking 1 , 10 , and 100 samples, respectively, from each of the 5000 posteriors and then combining the samples.


Fig. B.5. Illustration of the differences in interior models between the previous work (Baumeister et al. 2020, black line) and this work (blue and red lines, for two different equilibrium temperatures $T_{\text {eq }}=100 \mathrm{~K}$ and $T_{\text {eq }}=1000 \mathrm{~K}$, respectively). The figure shows density profiles of a representative $5 M_{\oplus}$ planet with $w_{\text {Core }}=0.2, w_{\text {Mantle }}=0.49, w_{\text {Water }}=0.3$, and $w_{\text {Gas }}=0.01$ (Panel a). Panel b shows a zoomed-in view of only the water and atmosphere layers. The dashed lines mark the respective planets' radii.


[^0]:    * Full Table A. 1 is only available at the CDS via anonymous ftp to cdsarc.cds.unistra.fr (130.79.128.5) or via https:// cdsarc.cds.unistra.fr/viz-bin/cat/J/A+A/676/A106
    ** ExoMDN is freely accessible through the GitHub repository https://github.com/philippbaumeister/ExoMDN

[^1]:    1 https://github.com/philippbaumeister/ExoMDN

[^2]:    ${ }^{2}$ All predictions were performed on an Intel ${ }^{\circledR}$ Core ${ }^{\mathrm{TM}} \mathrm{i} 5-8250 \mathrm{U}$ CPU.
    ${ }^{3}$ https://exoplanetarchive.ipac.caltech.edu/

