Neural network-based emulation of astrophysical models

Application to the Meudon PDR code

L. Einig[1,2], P. Palud[3,4], J. Chanussot[2], J. Pety[1], F. Le Petit[4], E. Bron[4] and the ORION-B consortium[5]

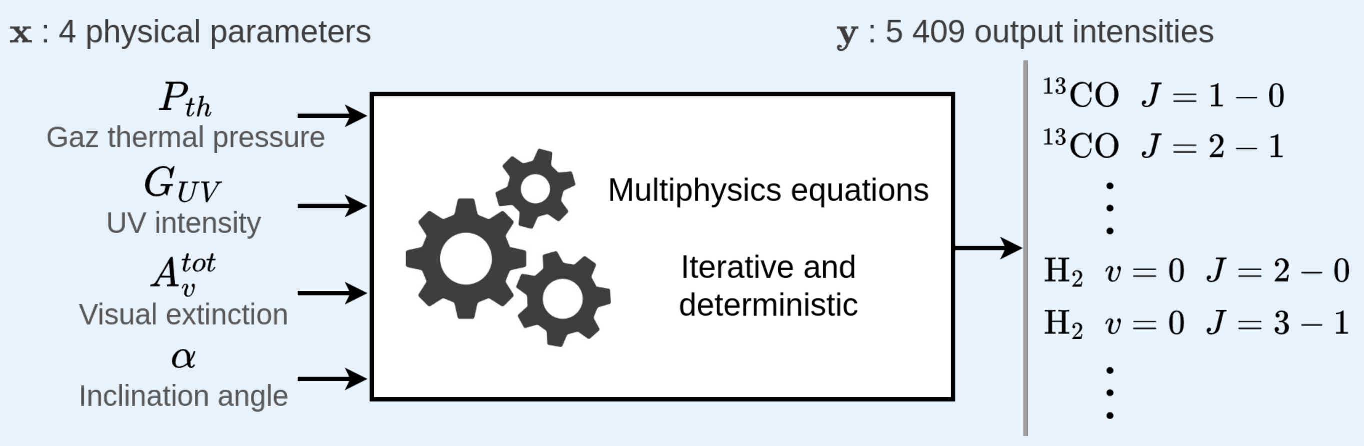
[1] *Institut de Radioastronomie Millimétrique*
[2] *GIPSA-Lab, Université Grenoble Alpes*
[3] *CRIStAL, Université de Lille*
[4] *LERMA, Observatoire de Paris*
[5] https://www.iram.fr/~pety/ORION-B/

1/ Interstellar medium models

❒ Numerical simulations are used to model **complex astrophysics phenomena**. Most realistic models can take into account a wide range of **multiphysics** aspects.
✦ Computing **time and memory resources** requirements can be very demanding, limiting their usability.
✦ Often replaced by **reduced models**, approximating the original behavior with lower complexity.
➙ Usually **interpolation methods**.

❒ We present a **reduction of the Meudon PDR code** [**1**] based on an artificial neural network (ANN).
✦ Numerical model of photodissociation regions (PDRs) (*e.g.*, the Horsehead Nebula, illuminated by Alnitak).

✦ From a **few parameters**, the model calculates the intensity of **numerous emission lines** for various species.



✦ Process a combination of inputs in about **six hours**.

2/ Challenges

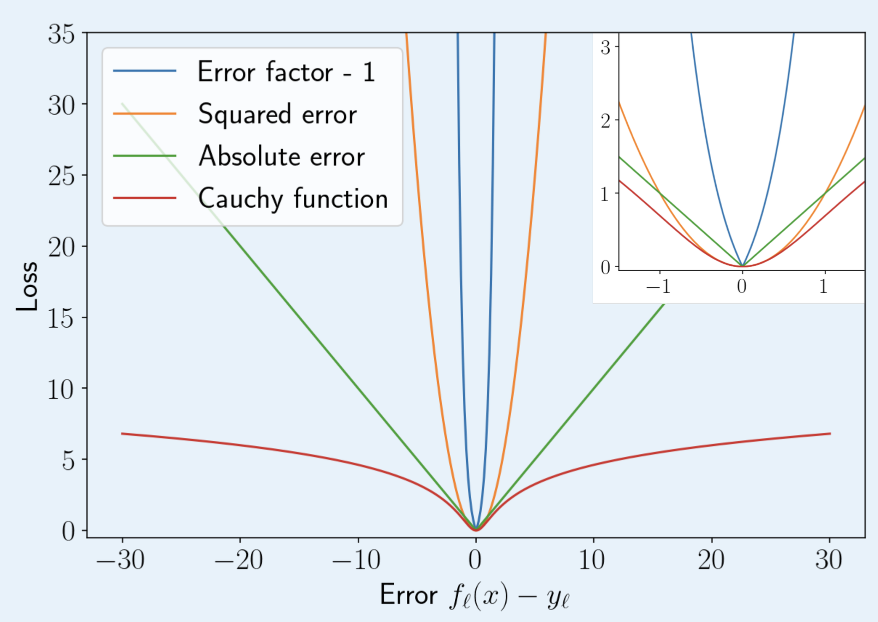
❒ Most works on ANNs consider many independent inputs and few outputs while the opposite is true here.

❒ The PDR code can produce **anomalies** ($\neq$ outliers).
➙ We want to train a model that does not learn them.

❒ For **Bayesian inference** [**2**], the network must be
✦ **Differentiable** (at least to order 2) from end to end.
✦ As **regularized** as possible to avoid significant errors on successive derivatives (*e.g.*, due to oscillations).

❒ Comparison of ANNs with interpolations in terms of:
✦ **Speed:** Evaluation time for a batch of 1 000 entries.
✦ **Memory:** Number of parameters to describe the model.
✦ **Accuracy:** Average and 99% percentile error factor.

$$\mathrm{EF}\left(\hat{y}, y\right) = 10^{\left|\log_{10}\hat{y} - \log_{10} y\right|} - 1 = \max\left(\frac{\hat{y}}{y}, \frac{y}{\hat{y}}\right) - 1$$



3/ Training set cleaning and architecture design

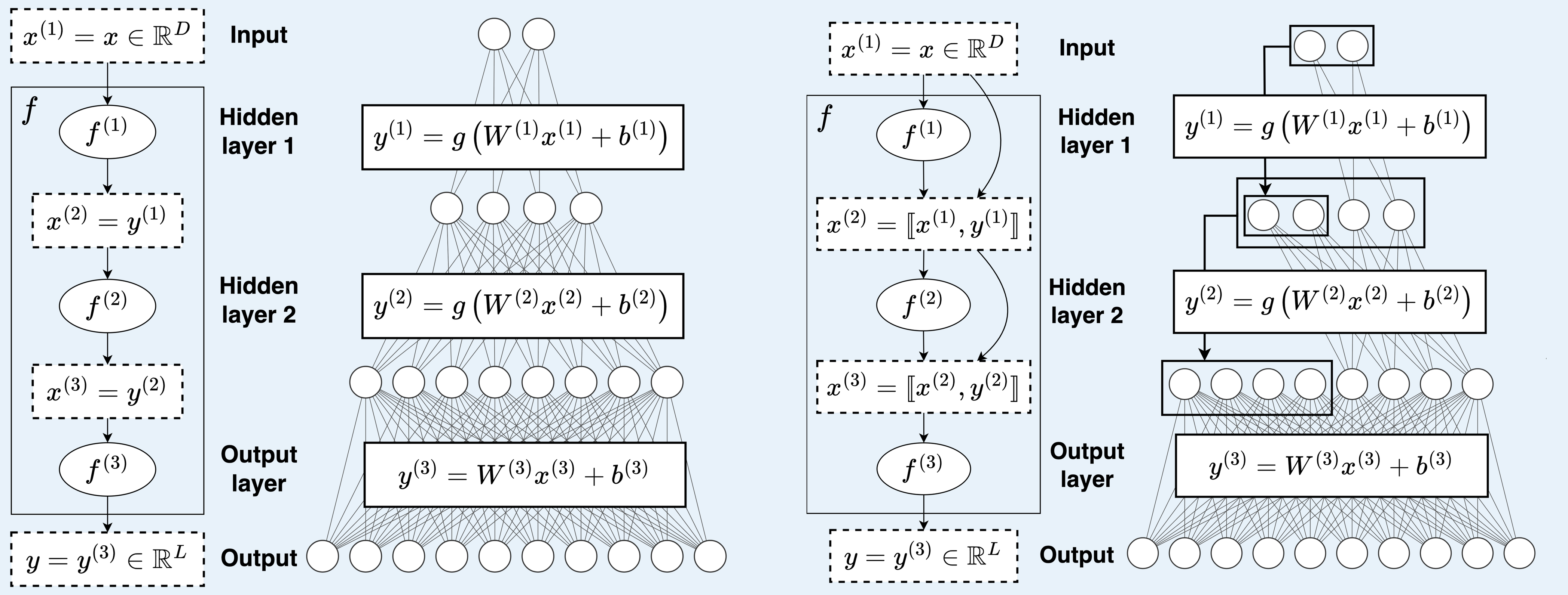
➙ Our approach: improve on network **training procedure** (point 1) and on its **architecture design** (points 2, 3, 4).

1. **Anomalies identification:** we perform a three-stage procedure.
✦ We train a network with a **Cauchy loss** (robust to outliers).
✦ Training points with highest errors: reviewed on the basis of **physics knowledge**.
✦ Another training with a **masked squared error**: abnormal points are ignored.

$$\mathcal{L}_m(\mathbf{f}, (\mathbf{x}, y, m)) = m \cdot (f(\mathbf{x}) - y)^2 \quad \text{with} \quad m \in \{0, 1\}$$

2. **Polynomial expansion of inputs:** helps the network create non-linearities from few inputs. Calculated at runtime to ensure the model differentiability.

3. **Dense architecture:** exploits inputs and intermediate results to predict outputs.
➙ We use an architecture inspired by the convolutive network DenseNet [**3**].



Example of fully connected ANN vs. densely connected ANN

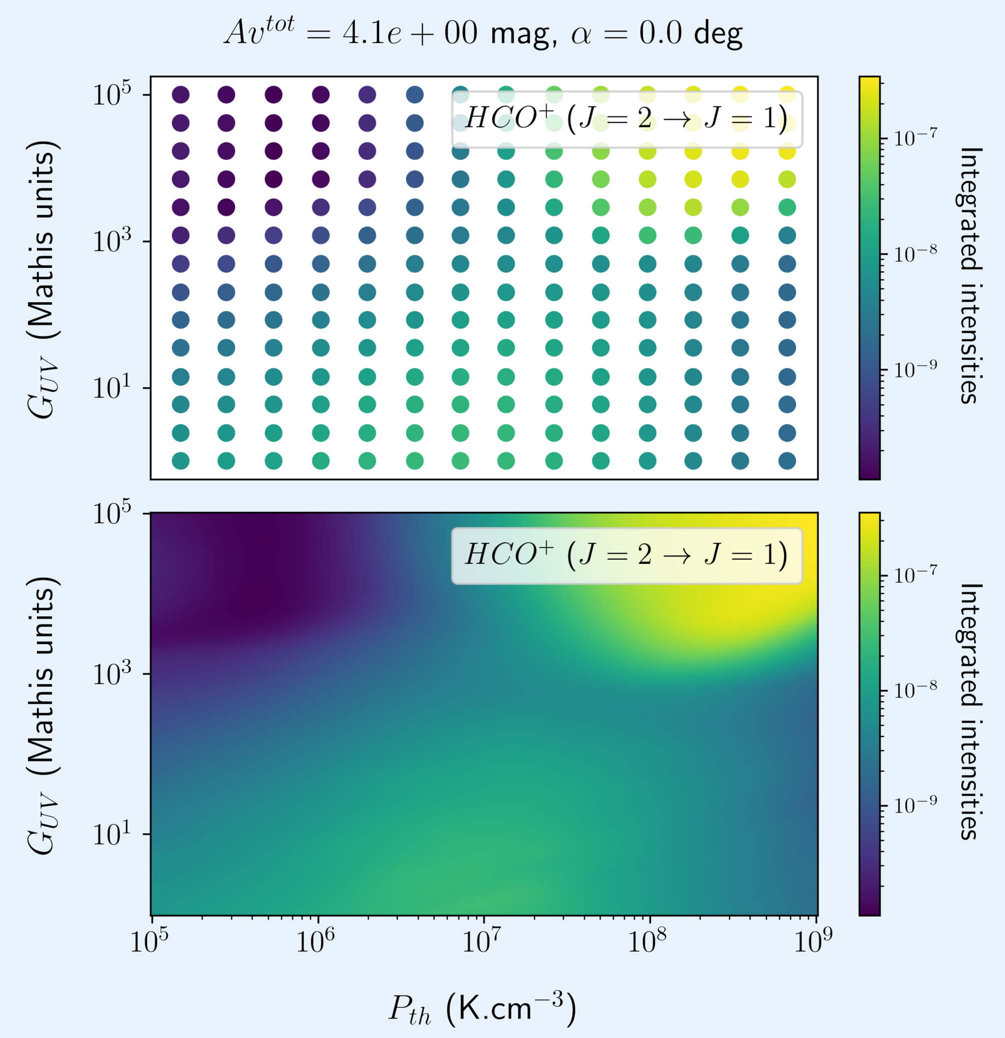
4. **PCA to size hidden layers:** we exploit the linear redundancy between outputs to adequately size the penultimate layer and significantly reduce the number of parameters.

4/ Regression results

Performance of interpolation methods and of proposed neural networks

	Method	Error factor Mean	Error factor 99% pc.	Memory (MB)	Speed (ms)
Interp.	Near. neigh.	1 310%	1 130%	1 650	62
	Linear	15.7%	230%	1 650	1.5e3
	Cubic spline	11.2%	220%	1 650	-
	RBF linear	10.2%	96.8%	1 650	1.1e4
ANN	S	7.3%	64.8%	**118**	12
	S+P	6.2%	49.7%	**118**	13
	S+P+A	5.5%	42.3%	**118**	13
	S+P+A+D	**4.5%**	**33.1%**	125	**11**

S: *sizing of the last hidden layer using PCA*
P: *polynomial transform of inputs*
A: *anomalies removal on the training set*
D: *dense architecture*



5/ Conclusion

✓ ANNs outperform interpolation methods on every metrics.
✓ Detection of anomalies and robust learning.
✓ Efficient computation of successive derivatives.
✓ Paves the way to efficient inferences on large multi-line maps.

6/ References

[**1**] Le Petit *et al.*, *A model for atomic and molecular interstellar gas: The Meudon PDR code*, 2006
[**2**] Palud *et al.*, *Estimating physical parameters in the ISM from hyperspectral observations*, in prep.
[**3**] Huang *et al.*, *Densely connected convolutional networks*, 2017