# Variational Inference for Data-Driven Galaxy Population Priors


Maxime RONCERAY[1,2], François Lanusse[2], Samuel Farrens[2]

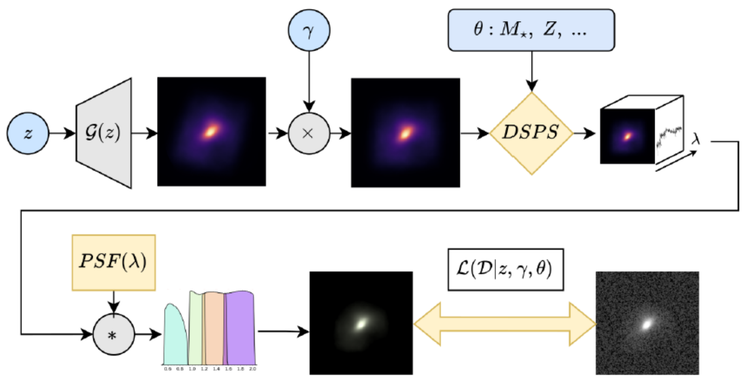
1 : CEA - 2 : Université Paris-Saclay - 3 :Université Paris Cité - 4 : CEA - 5 : CNRS - 6 : AIM


## 1 - MOTIVATIONS

**Why do we need a good SED generator ?** Pixel level inference for shear estimation requires a chromatic PSF
**Goal:** provide galaxy SEDs for the chromatic image model.

### SHINE

- **Pixel level inference** pipeline for unbiased shear estimation

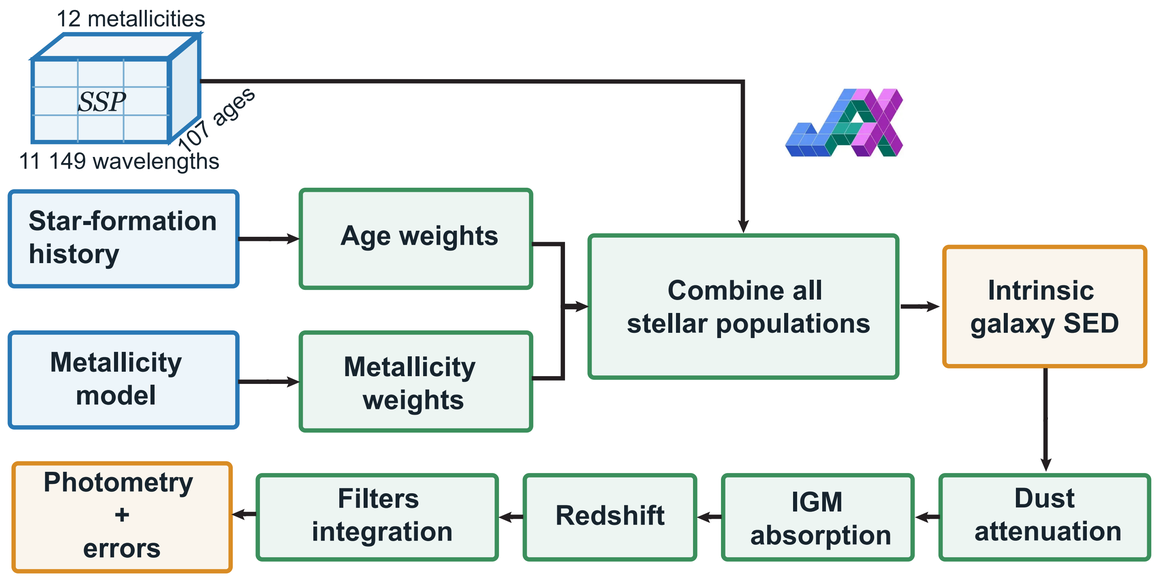



- **JAX-powered :** Automatic differentiation

### CONTEXT

**1.** Broadband colours do not properly constrain galaxy properties.

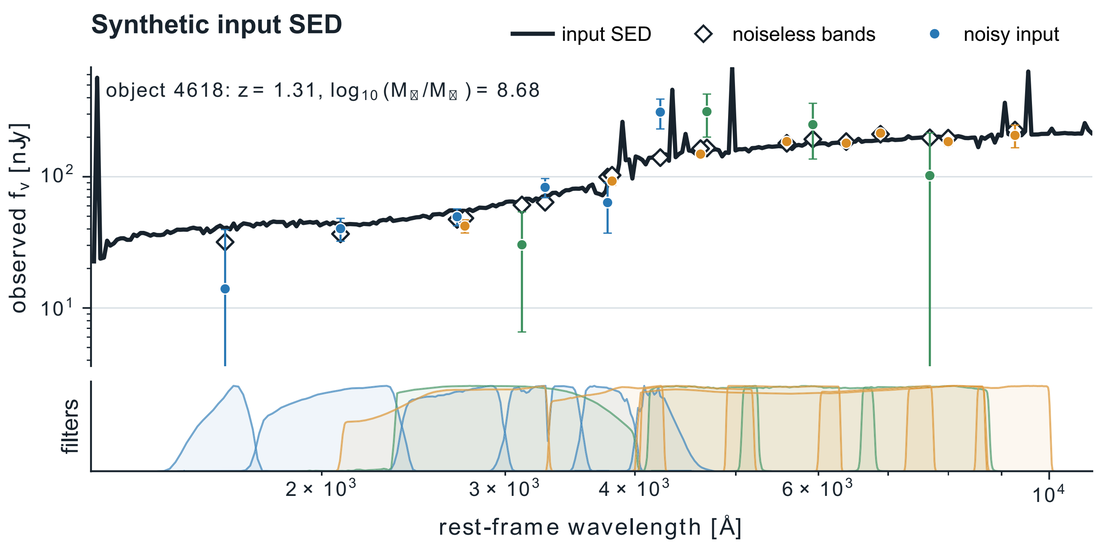
**2.** DSPS connects **degenerate combination** of physical parameters to spectra and photometry.



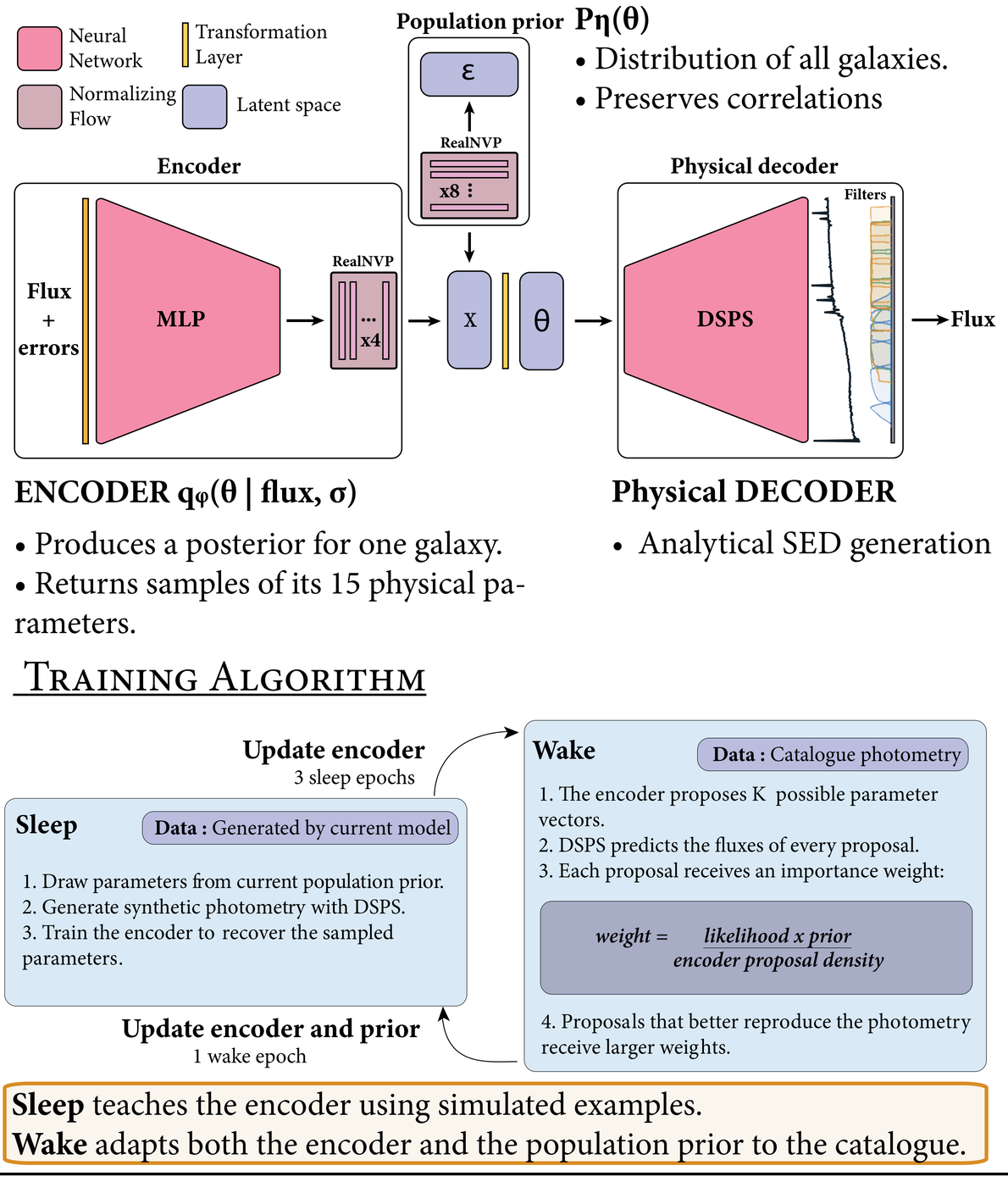

**3.** Two distributions learnt:
- a **prior** for the full galaxy population;
- a **posterior** for each observed galaxy.

## 2 - DATA

- **Synthetic dataset** generated with Diffsky [2]
- Observed through 18 filters: 6 **Rubin** + 4 **Euclid** + 8 **Roman.**
- **15 Inferred parameters:**
  - redshift
  - stellar mass
  - metallicity
  - 2 dust parameters
  - 10 star-formation-history nodes




## 3 - ARCHITECTURE




- Distribution of all galaxies.
- Preserves correlations

- Produces a posterior for one galaxy.
- Returns samples of its 15 physical parameters.

- Analytical SED generation

### Training Algorithm

**Update encoder**
3 sleep epochs

**Sleep** — **Data :** Generated by current model

1. Draw parameters from current population prior.
2. Generate synthetic photometry with DSPS.
3. Train the encoder to recover the sampled parameters.

**Update encoder and prior**
1 wake epoch

**Wake** — **Data :** Catalogue photometry

1. The encoder proposes K possible parameter vectors.
2. DSPS predicts the fluxes of every proposal.
3. Each proposal receives an importance weight:

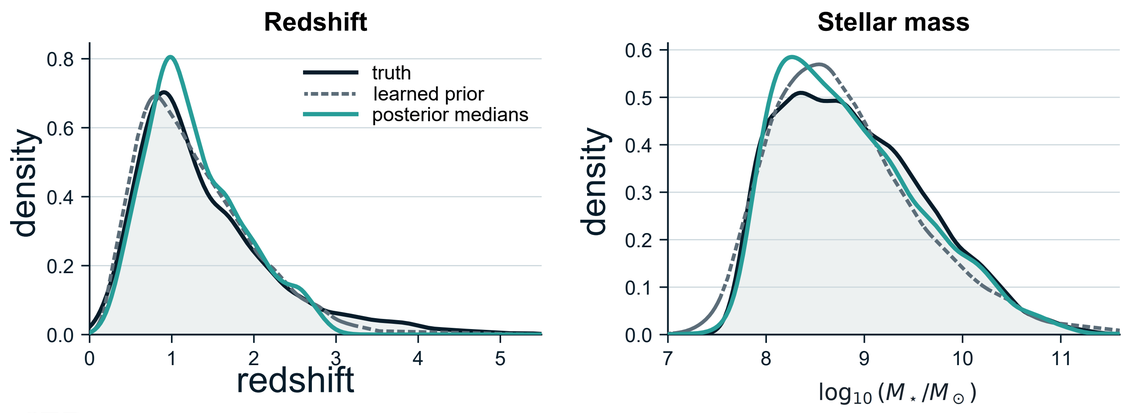
$$weight = \frac{likelihood \times prior}{encoder\ proposal\ density}$$

4. Proposals that better reproduce the photometry receive larger weights.

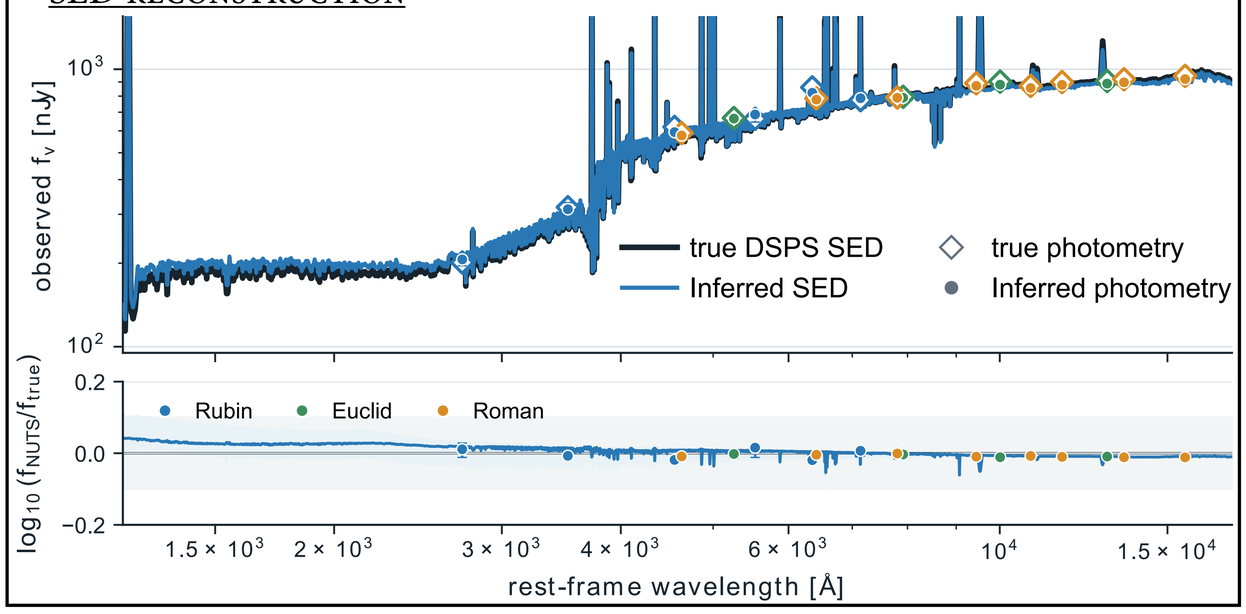
**Sleep** teaches the encoder using simulated examples.
**Wake** adapts both the encoder and the population prior to the catalogue.

## 4 - PRELIMINARY RESULTS

### Aggregate posterior




### SED reconstruction




## 5 - TAKEAWAYS

- Differentiable forward model connects galaxy physics to SEDs and photometry with end-to-end gradients.
- We jointly learns a shared population prior and fast per-galaxy posteriors from catalogue photometry.

## References


[1] Hearin et al. (2023), DSPS: Differentiable Stellar Population Synthesis
[2] Hearin et al. (2021), A Differentiable Model of the Assembly of Individual and Populations of Dark Matter Halos; Alarcon et al. (2023), Diffstar: A Fully Parametric Physical Model for Galaxy Assembly History.
[3] Bornschein & Bengio (2015), Reweighted Wake-Sleep; Le et al. (2020), Revisiting Reweighted Wake-Sleep for Models with Stochastic Control Flow.
[4] Alsing et al. (2024), pop-cosmos: A Comprehensive Picture of the Galaxy Population from COSMOS Data; Thorp et al. (2024), pop-cosmos: Scaleable Inference of Galaxy Properties and Redshifts with a Data-driven Population Model.