# Spectrospatial models for LVM

## SDSS-V Meeting 2026

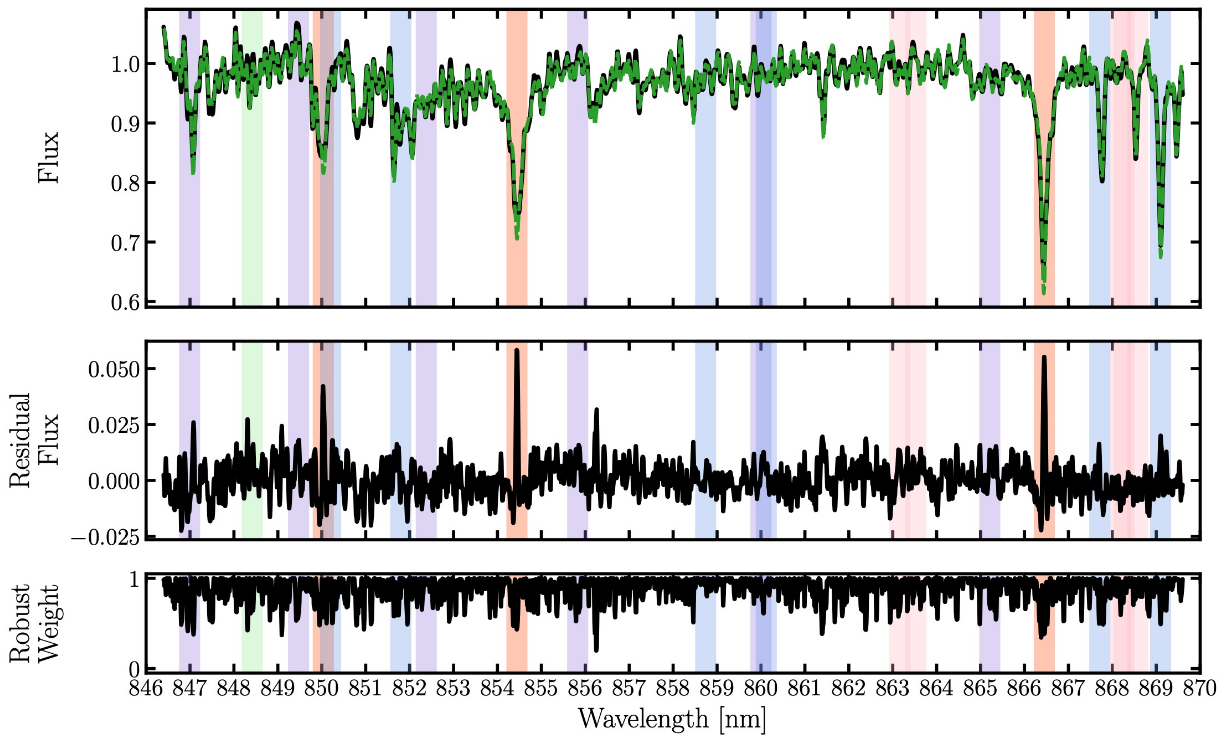
spectracles

SDSS-V

Tom Hilder (Monash)
& Andy Casey (CCA, Monash)
+ Collaborators

MONASH University

# Aside: PCA with outliers and bad data




Draft version July 10, 2026

Typeset using LaTeX default style in AASTeX631

**Robust Heteroskedastic Matrix Factorization:**
**A Generalization of PCA that Flags Outliers and Handles Missing Data**

Thomas Hilder [1], David W. Hogg [2,3,4] Andrew R. Casey [1,4] and Hans-Walter Rix [3]

[1] School of Physics and Astronomy, Monash University VIC 3800, Australia

[2] Center for Cosmology and Particle Physics, Department of Physics, New York University, 726 Broadway, New York, NY 10012, USA

[3] Max-Planck-Institut für Astronomie, Königstuhl 17, D-69117 Heidelberg, Germany

[4] Center for Computational Astrophysics, Flatiron Institute, 162 Fifth Ave., New York, NY 10010, USA

ABSTRACT

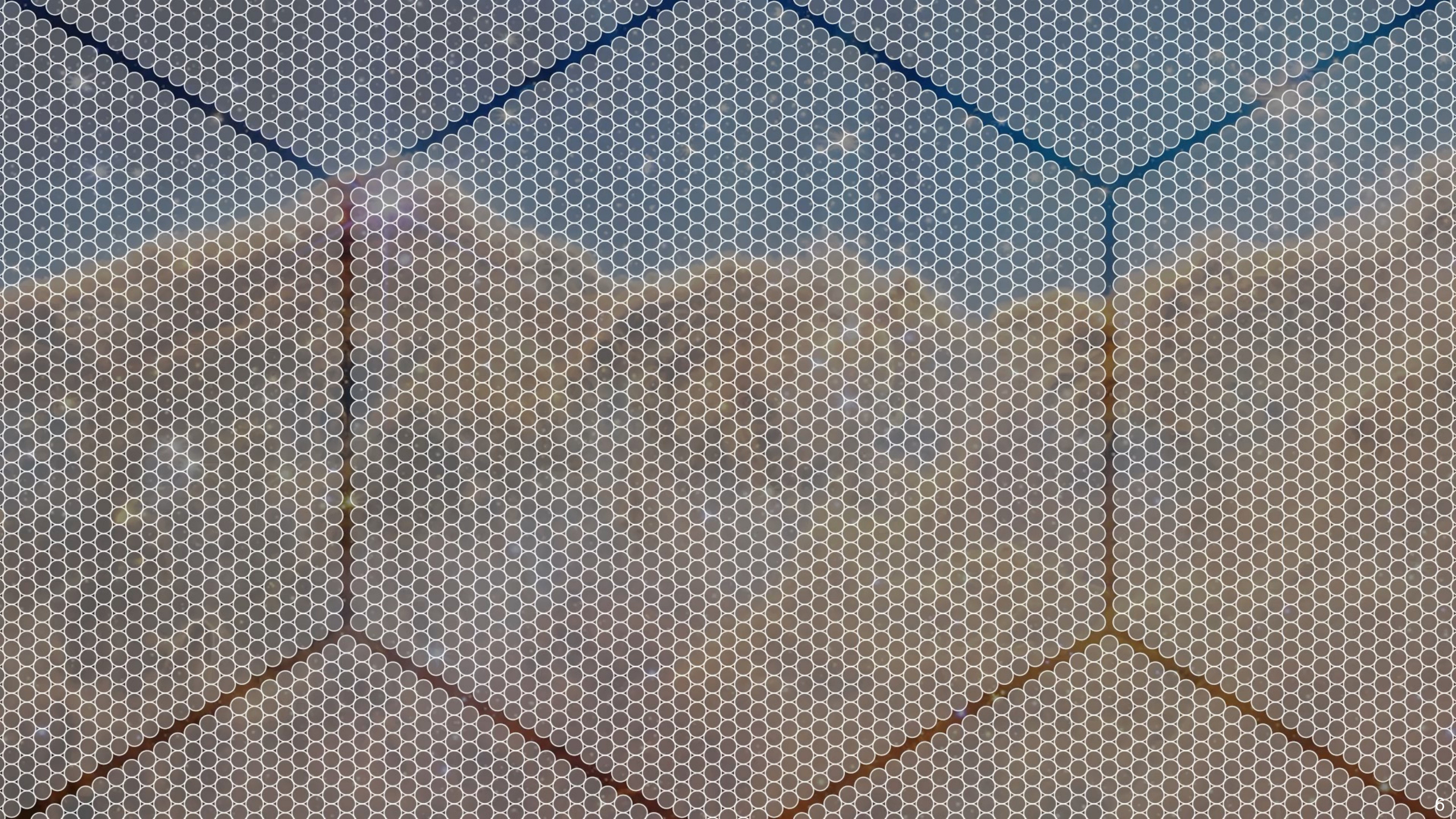
We present Robust Heteroskedastic Matrix Factorization (RHMF), a generalization of Principal Component Analysis (PCA) that is robust to outliers, handles per-feature uncertainties and missing data, and automatically flags per-feature and per-object anomalies. RHMF is useful both in recovering a low-dimensional embedding unspoiled by bad data or anomalies, and in identifying those anomalies. It utilises an iterative reweighting algorithm that implicitly maximizes a Student-t likelihood. This admits an equivalent probabilistic interpretation as fitting a hierarchical model with per-data-point latent variances. We deliver a fast `JAX` implementation, `Robusta-HMF`, and practical guidance for users. We demonstrate the ability of the model to identify and mitigate outliers of different classes. Identification accuracy is contingent on the choice of hyperparameters, but we show that these can be set reliably by cross-validation. We also apply RHMF to RVS spectra from Gaia DR3 to find main-sequence stars that are strange relative to their neighbors in color-magnitude space. We highlight specific examples, including a known binary hosting a Be star, and M-dwarfs with subtle emission in the Ca II triplet lines, indicative of accretion or magnetic activity, which would not be obvious to identify by eye.

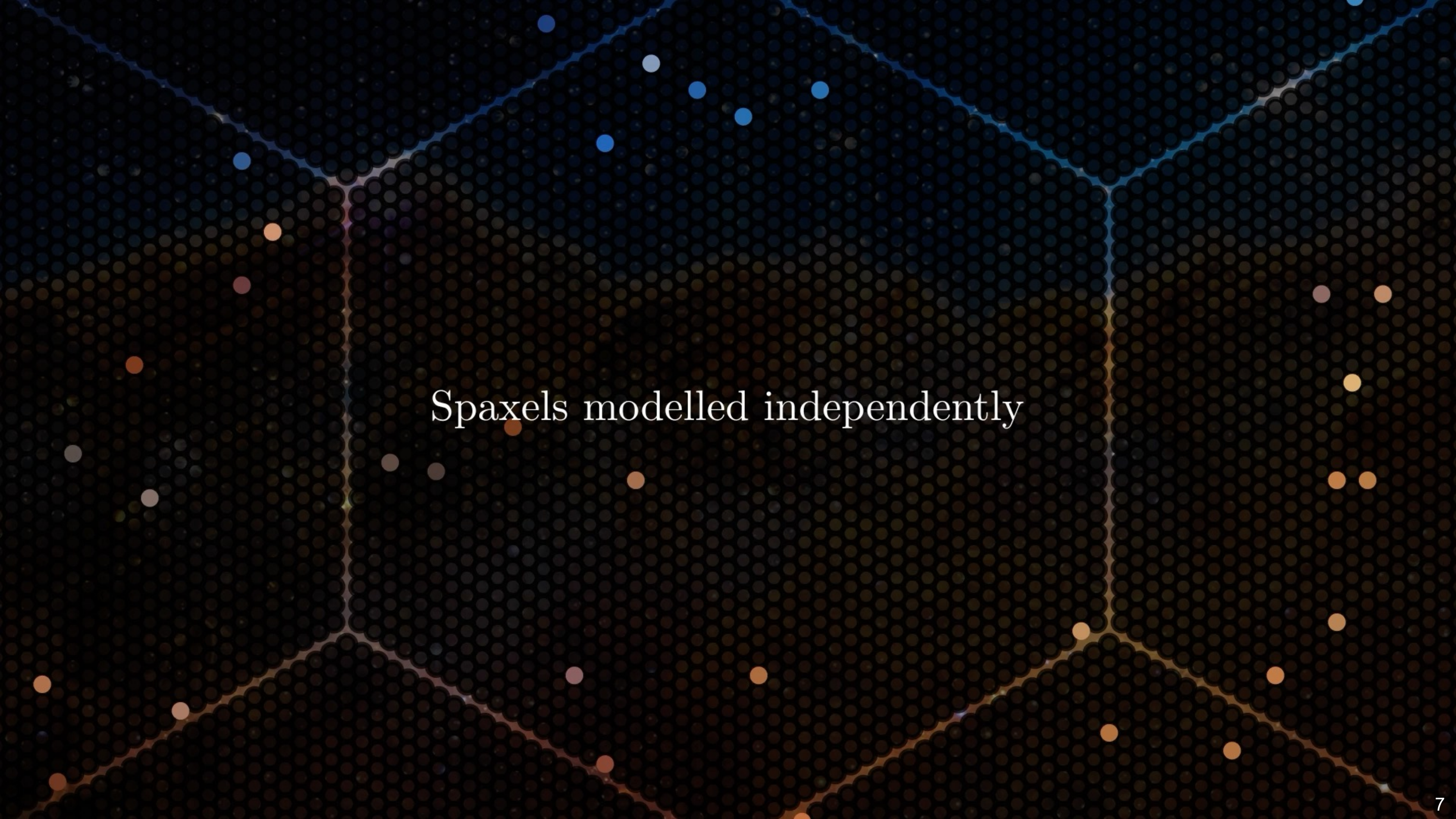
## 1. INTRODUCTION

Most stellar spectra are well described by a low-dimensional manifold. There are a handful of physical parameters—effective temperature, surface gravity, metallicity, $\alpha$-element enhancement, and perhaps projected rotation and a continuum shape—that account for most of the variation across thousands of pixels of a given spectrum, for most stars of most types. This is both an expectation of stellar atmosphere theory (Hubeny & Mihalas 2014) and an empirical finding from data-driven analyses of large spectroscopic surveys (e.g. Ness et al. 2015; Casey et al. 2016; Price-Jones & Bovy 2018).

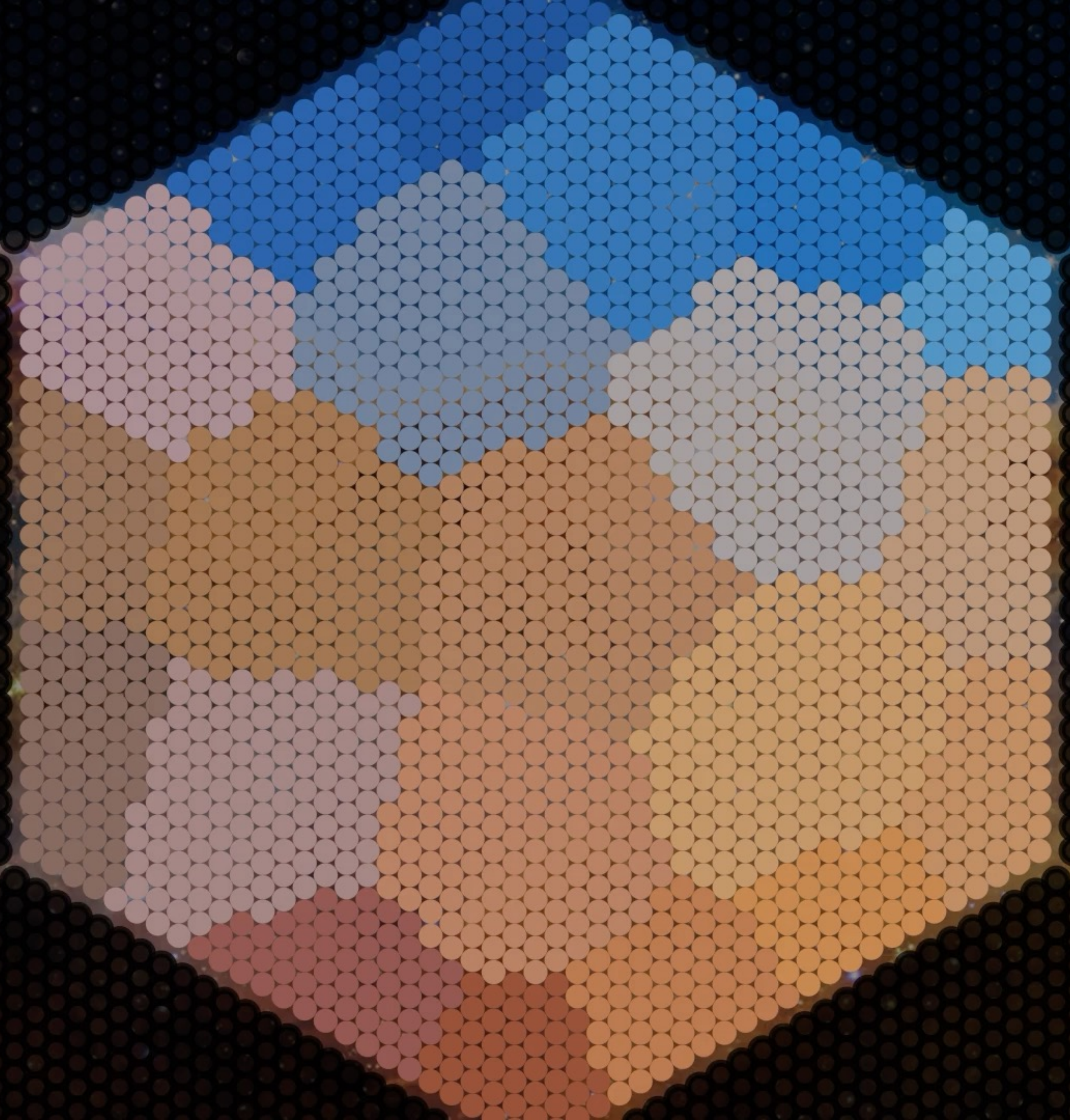
# Spectrospatial models

for mapping weak emission lines in LVM & beyond



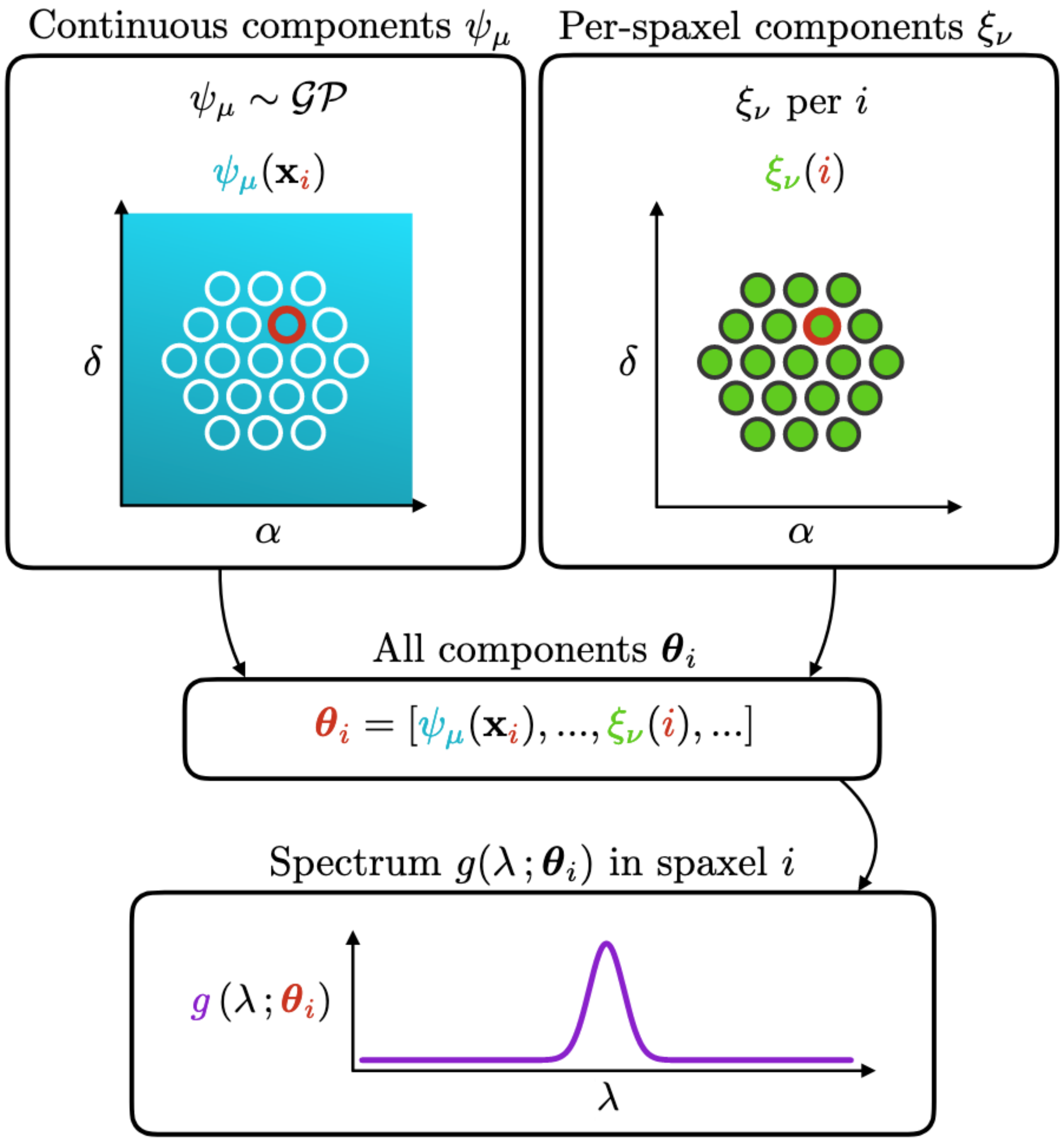

Spaxels modelled independently

# Binning?



# Model underlying **fields**

- Line strength
- Radial velocity
- Velocity dispersion

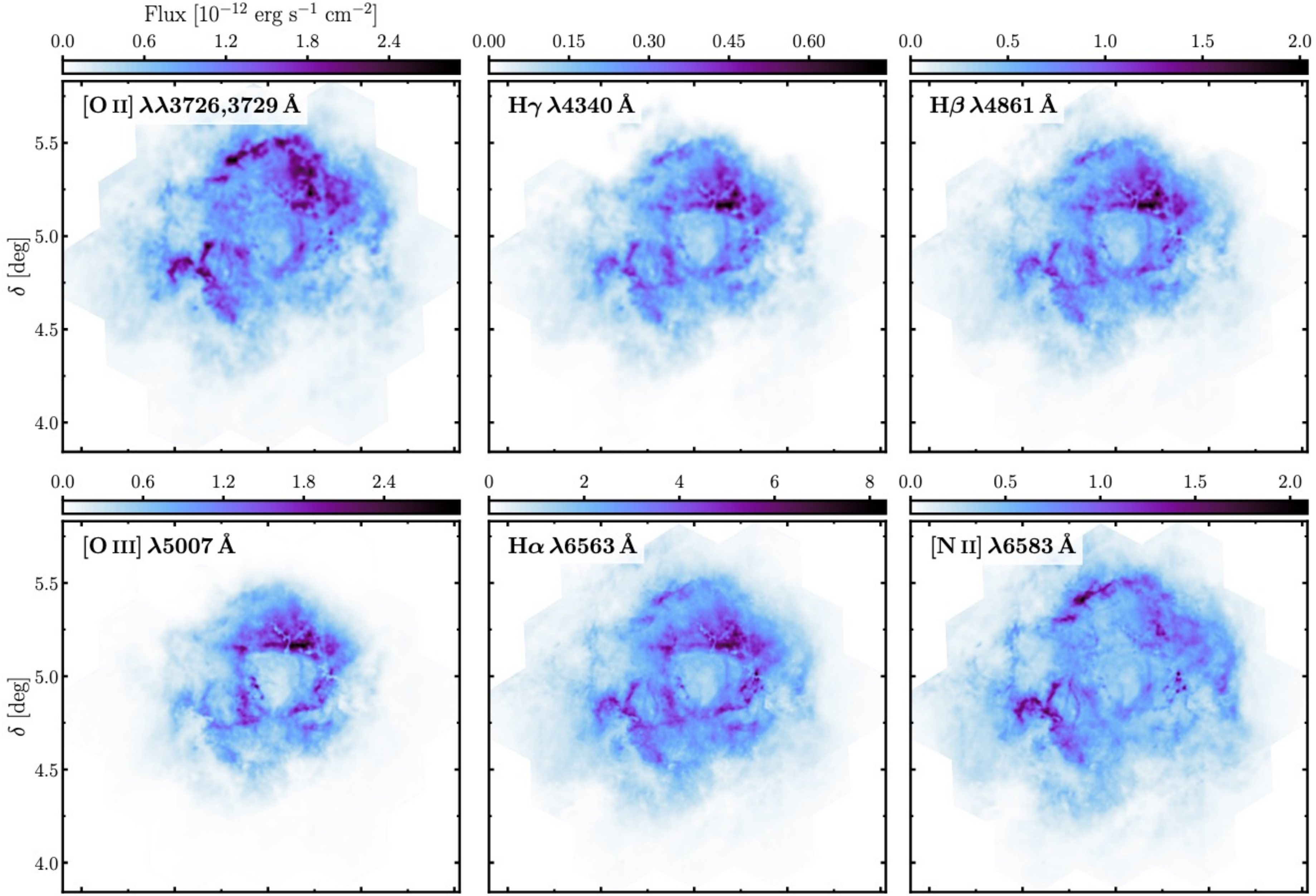



+ discrete bits

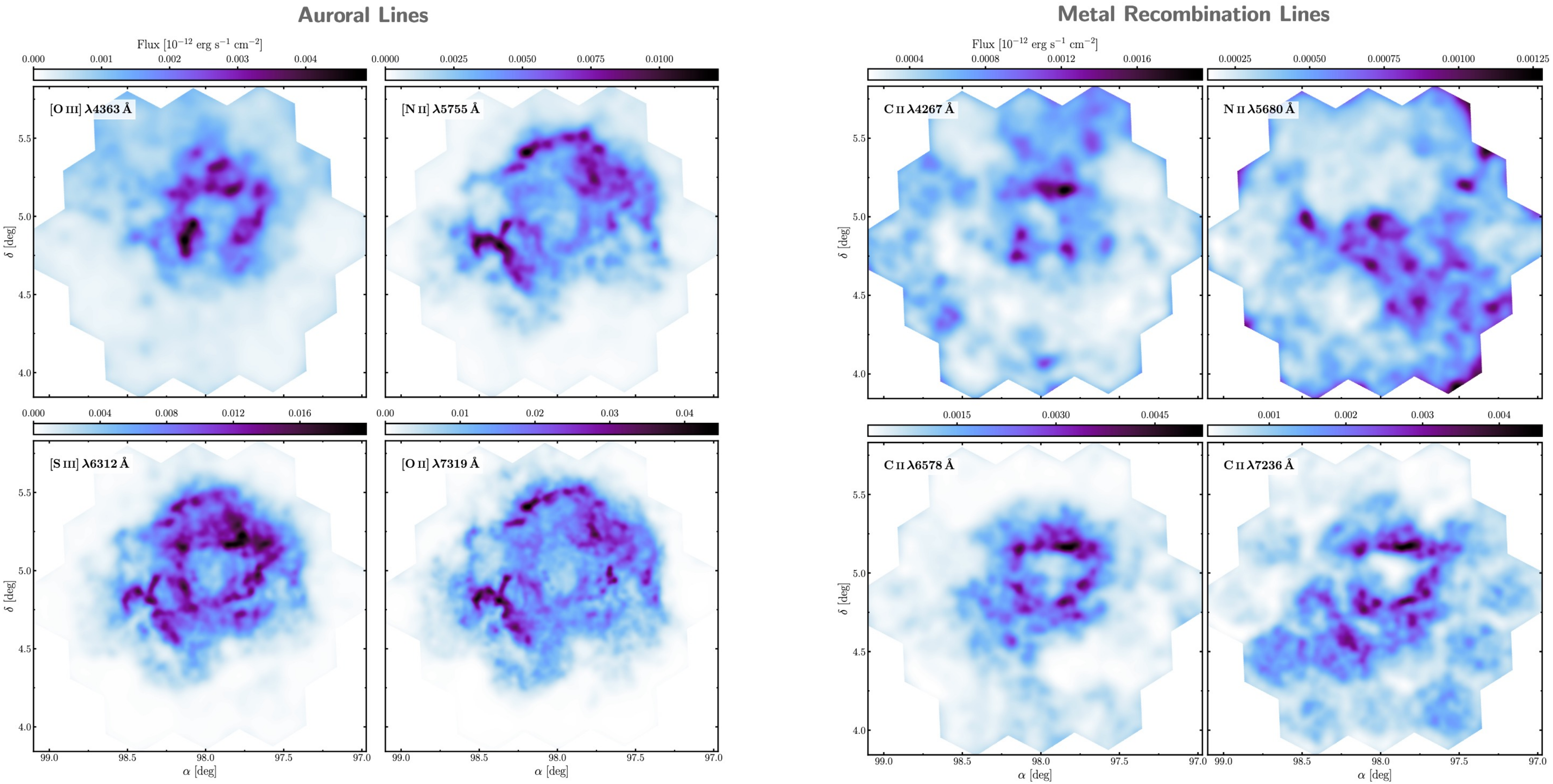
(e.g. Stellar Continuum)

spectracles

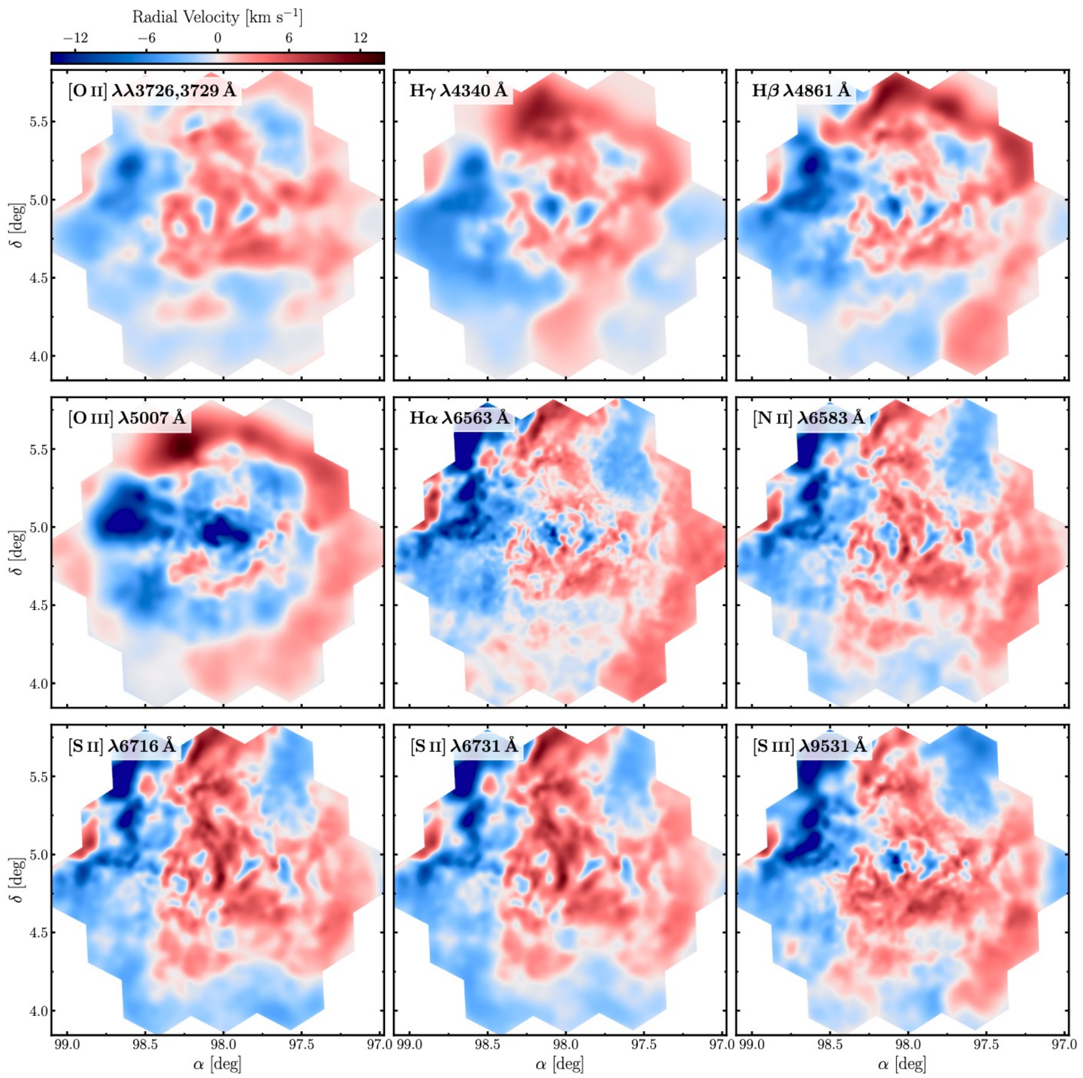
# First application – Rosette Nebula



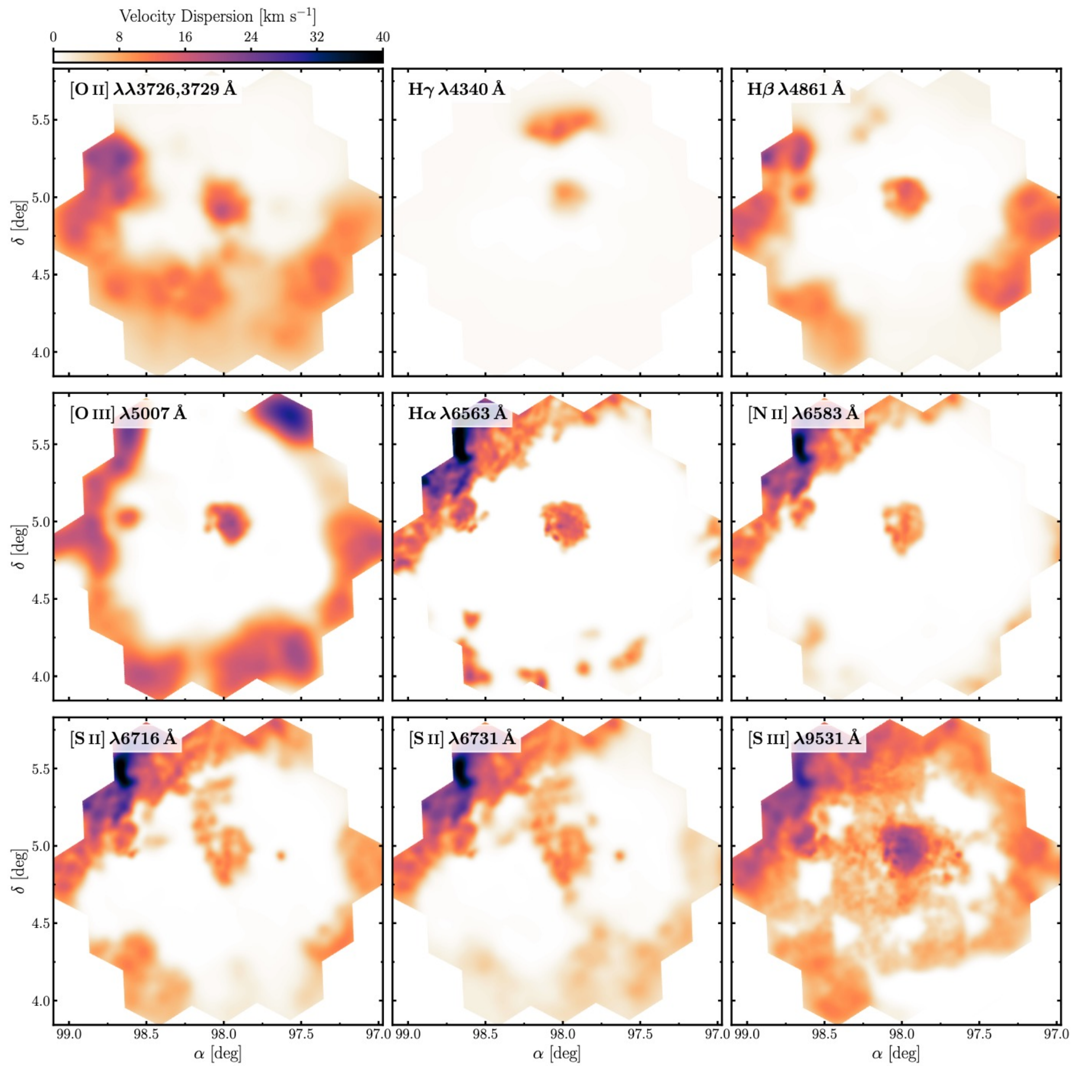
# First application – Rosette Nebula



# Radial velocities




# Dispersions (LSF corrected)

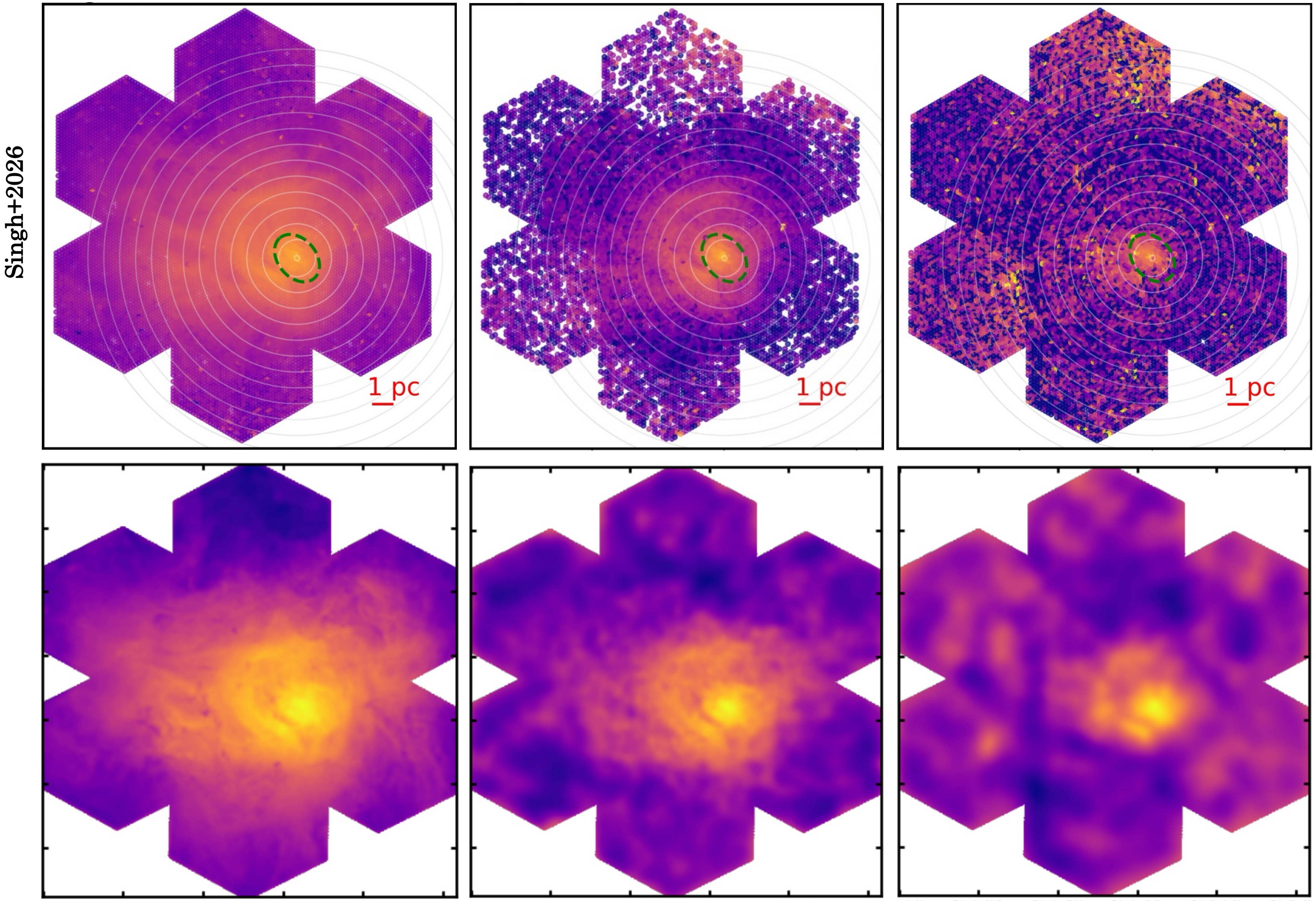


# Lagoon / M8

- w/ Amrita Singh

(Vera Rubin Obs.)

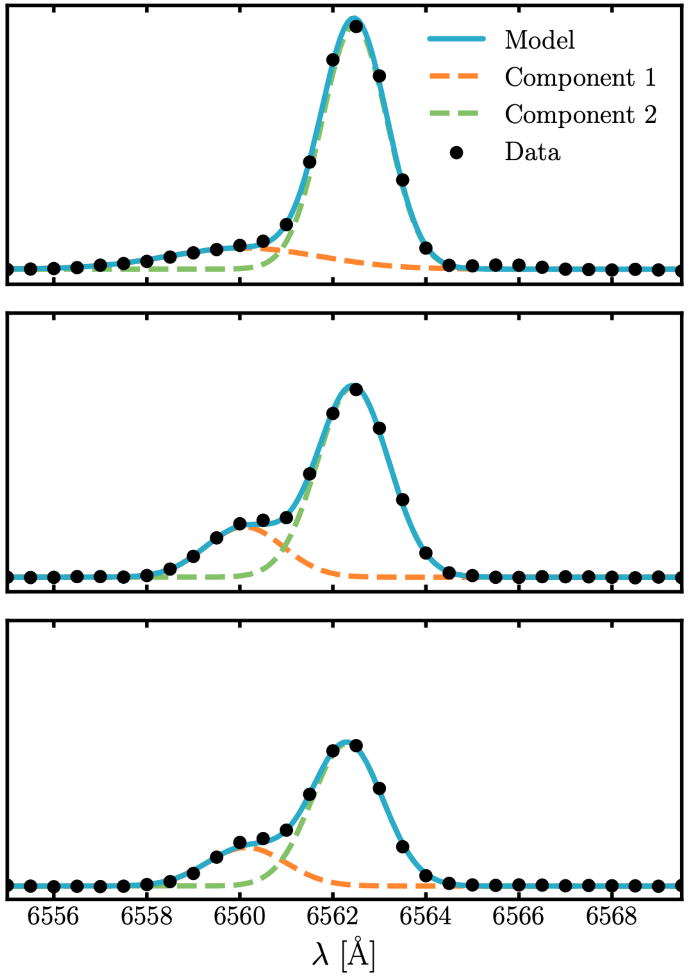
[O III] 5007 | [O III] 4363 | O II 4676



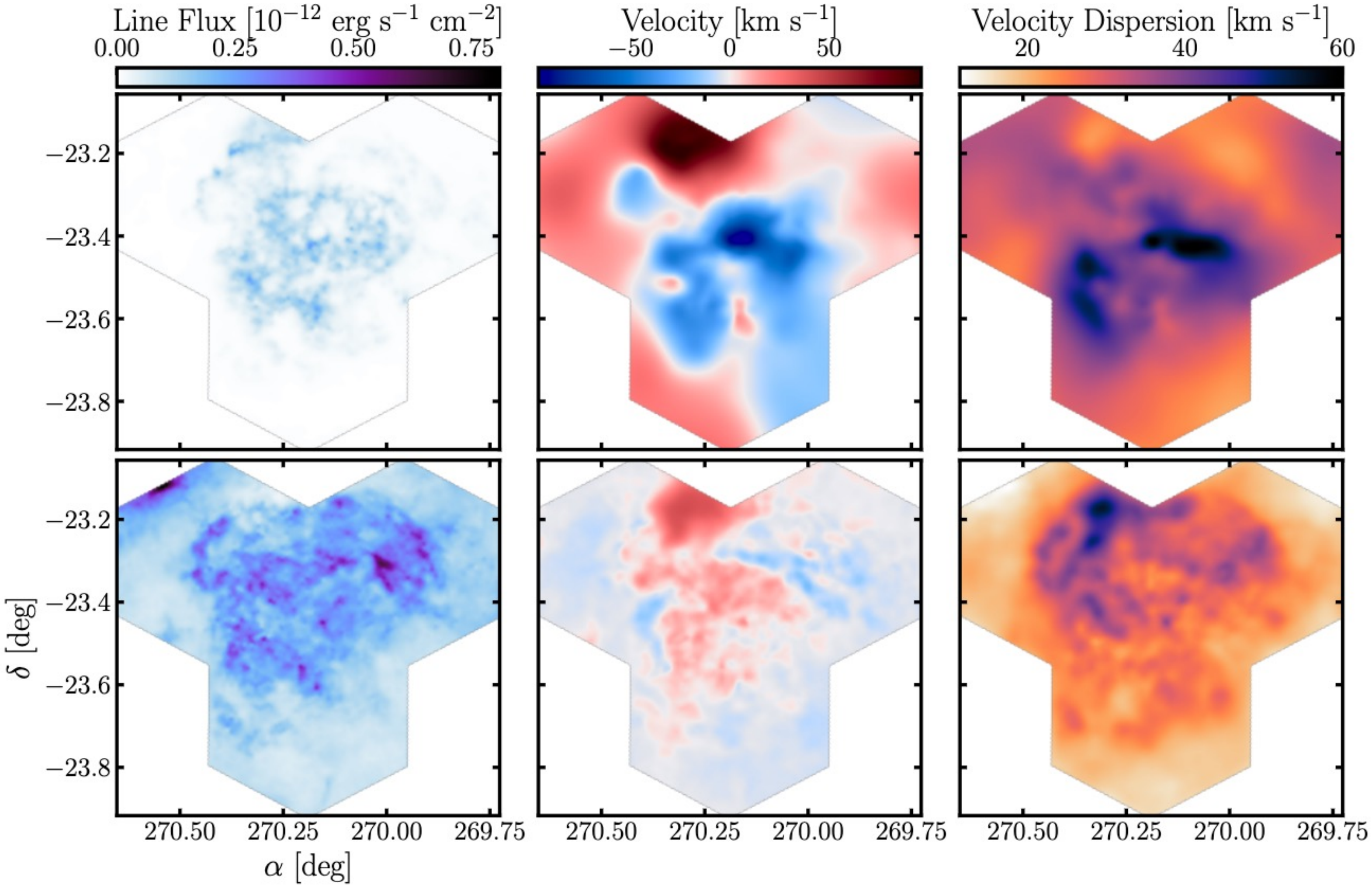

Strong Line | Auroral Line | Metal RL

# W28 Supernova Remnant

- w/ **Knox Long**
- Run this yourself on BinderHub

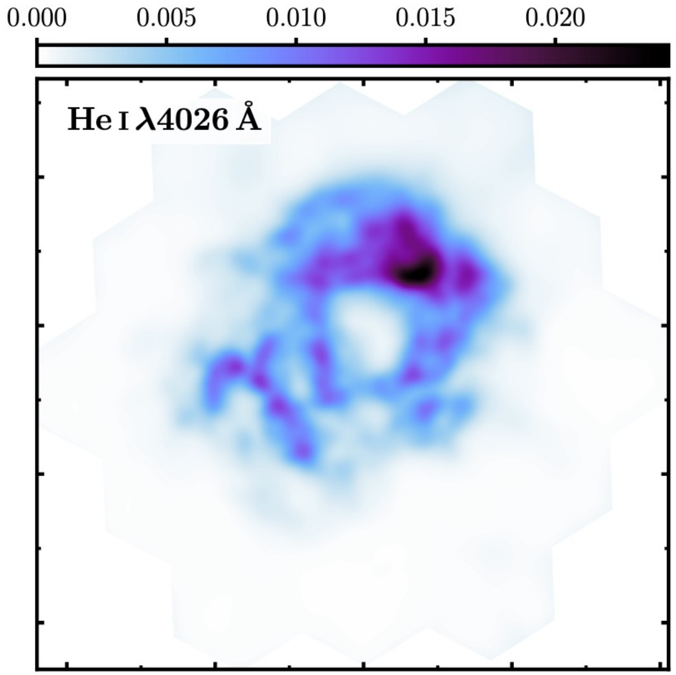


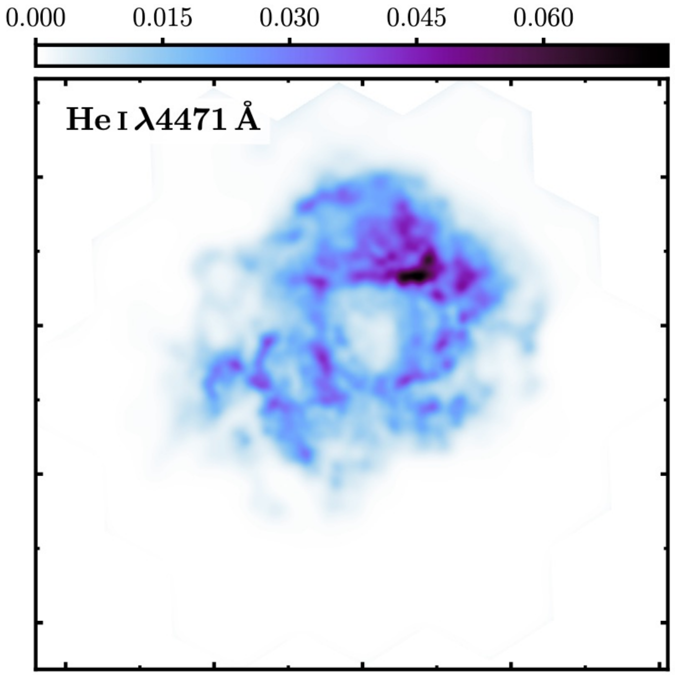

## [N II] 6583



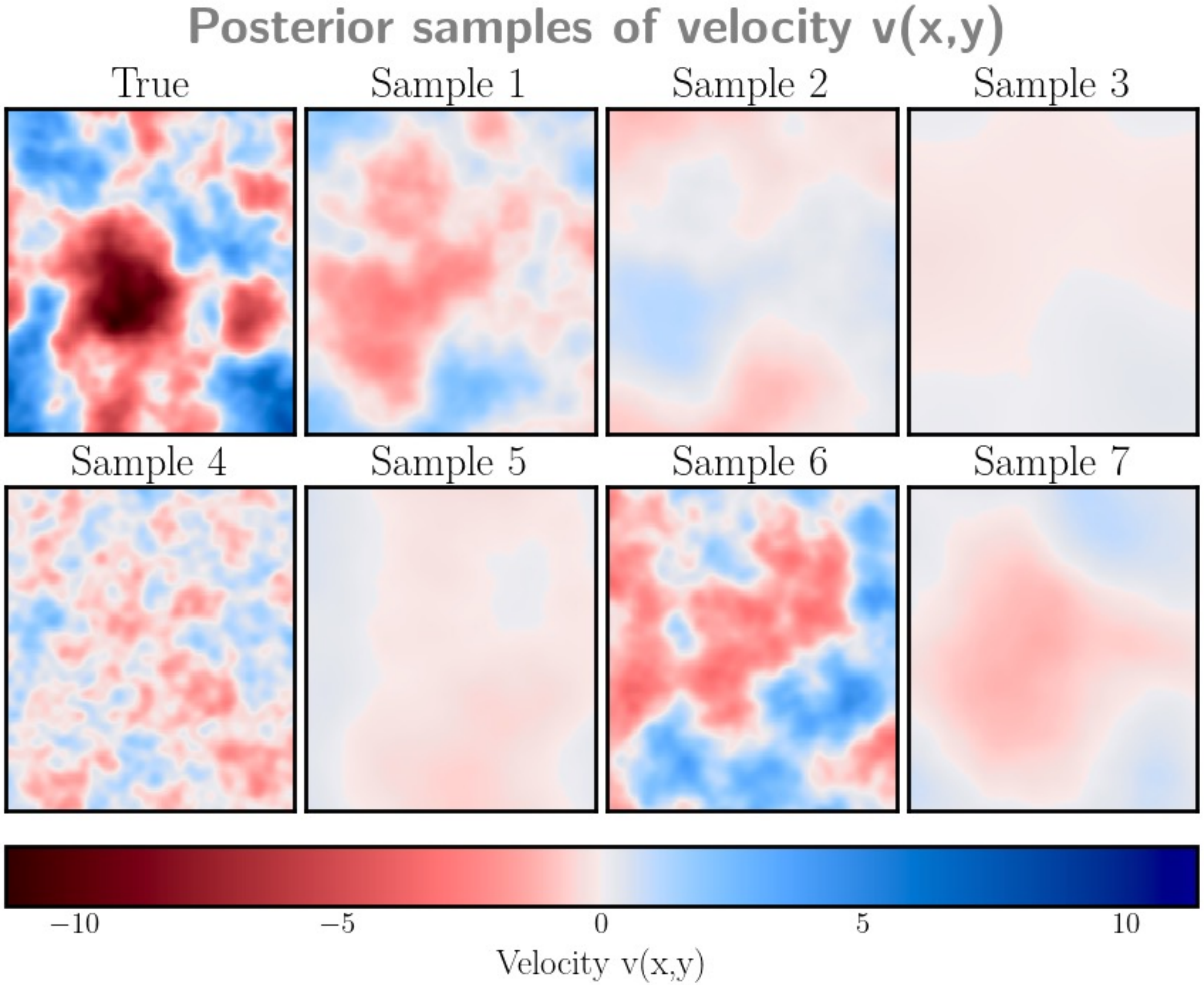
# Caveats?

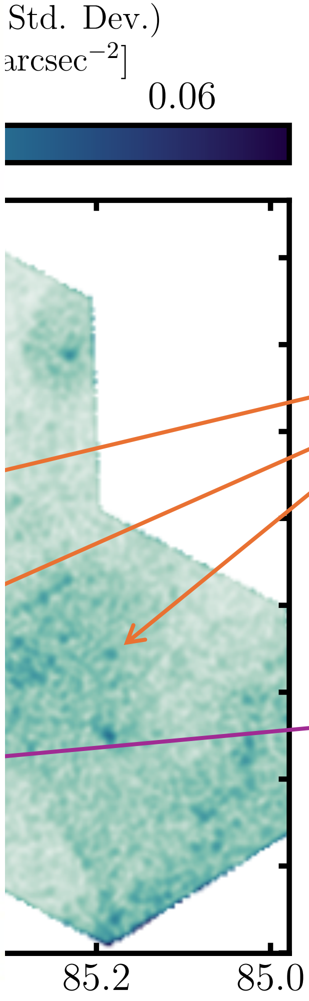
- Variable resolution at different S/N
  - Bias ratio maps?
- No uncertainties



# Unc



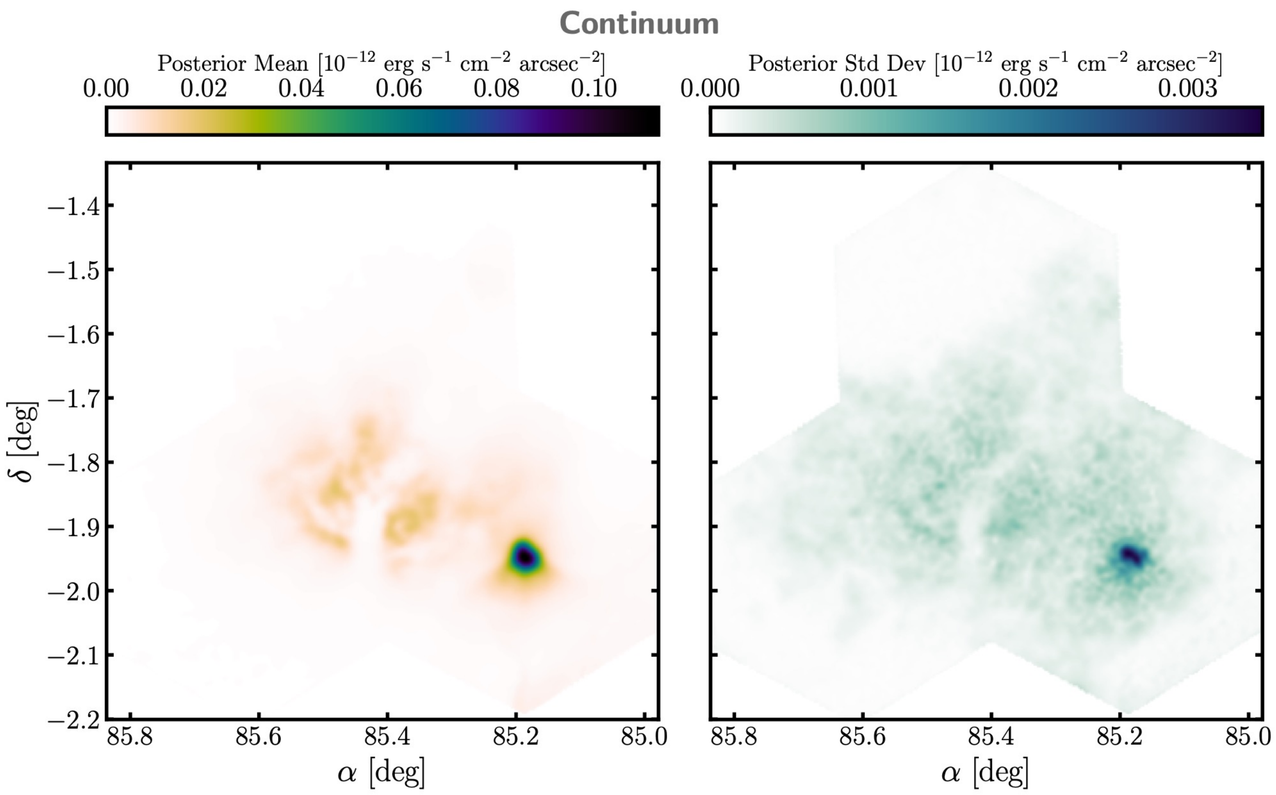




Flame ne
w/ **Lucian**
**& Amy St**

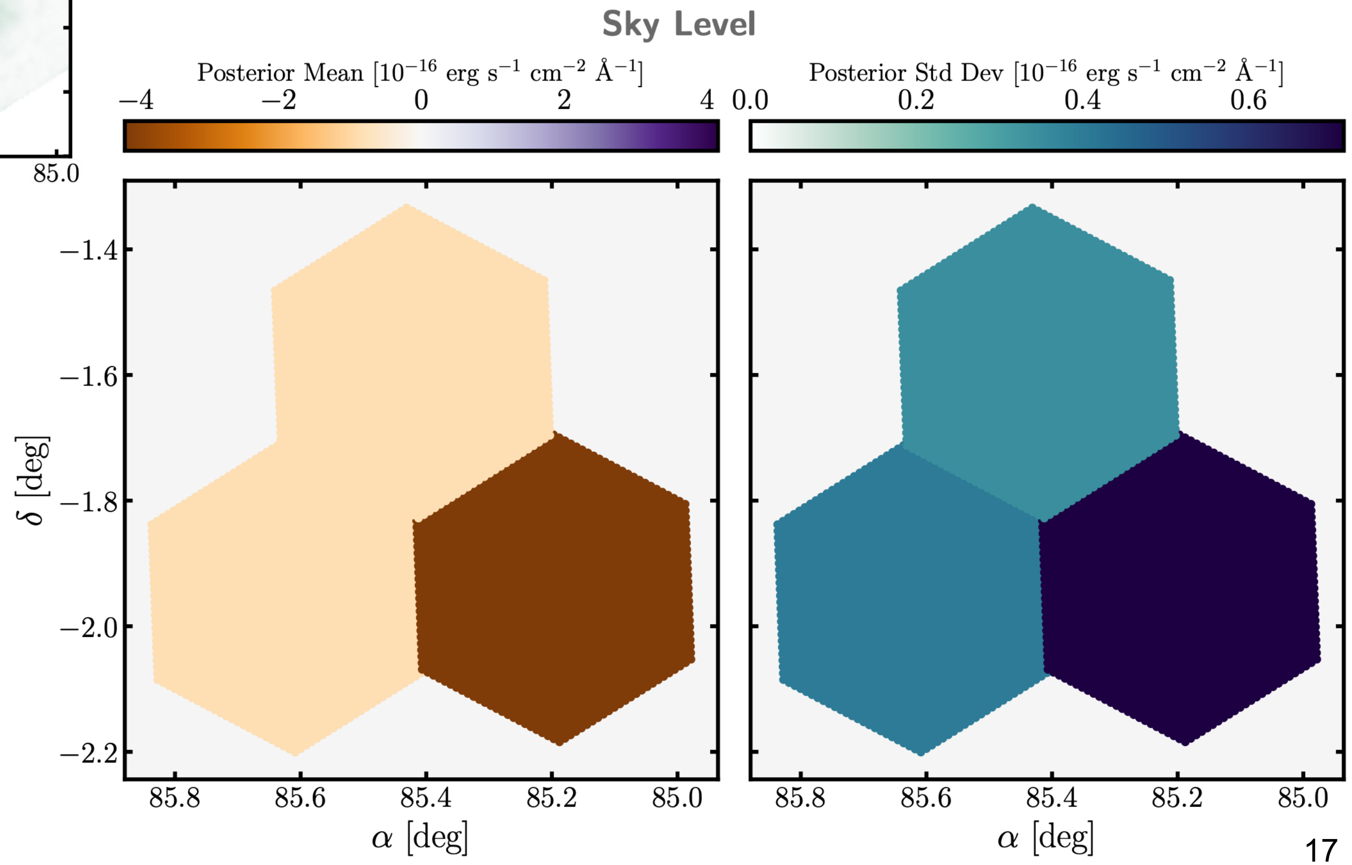
Higher uncertainty at excluded fibres

Line of partially excluded fibres here

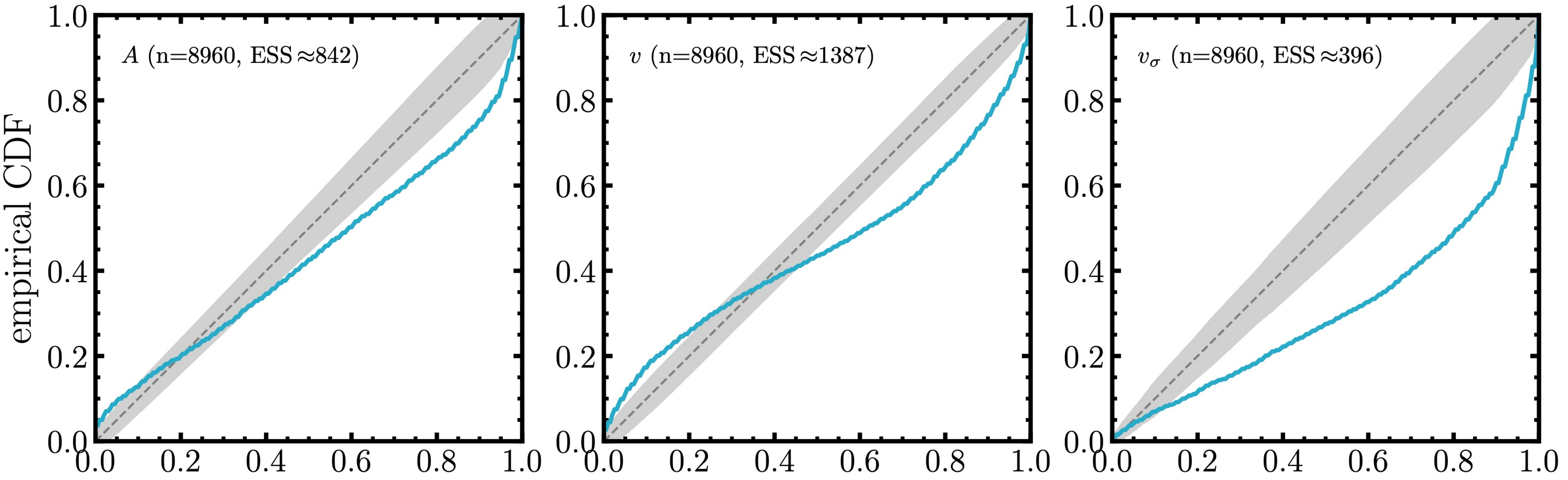
We are sampling over **all** model parameters, including the GP kernel parameters




# **Marginalise** over small mistakes in data reduction

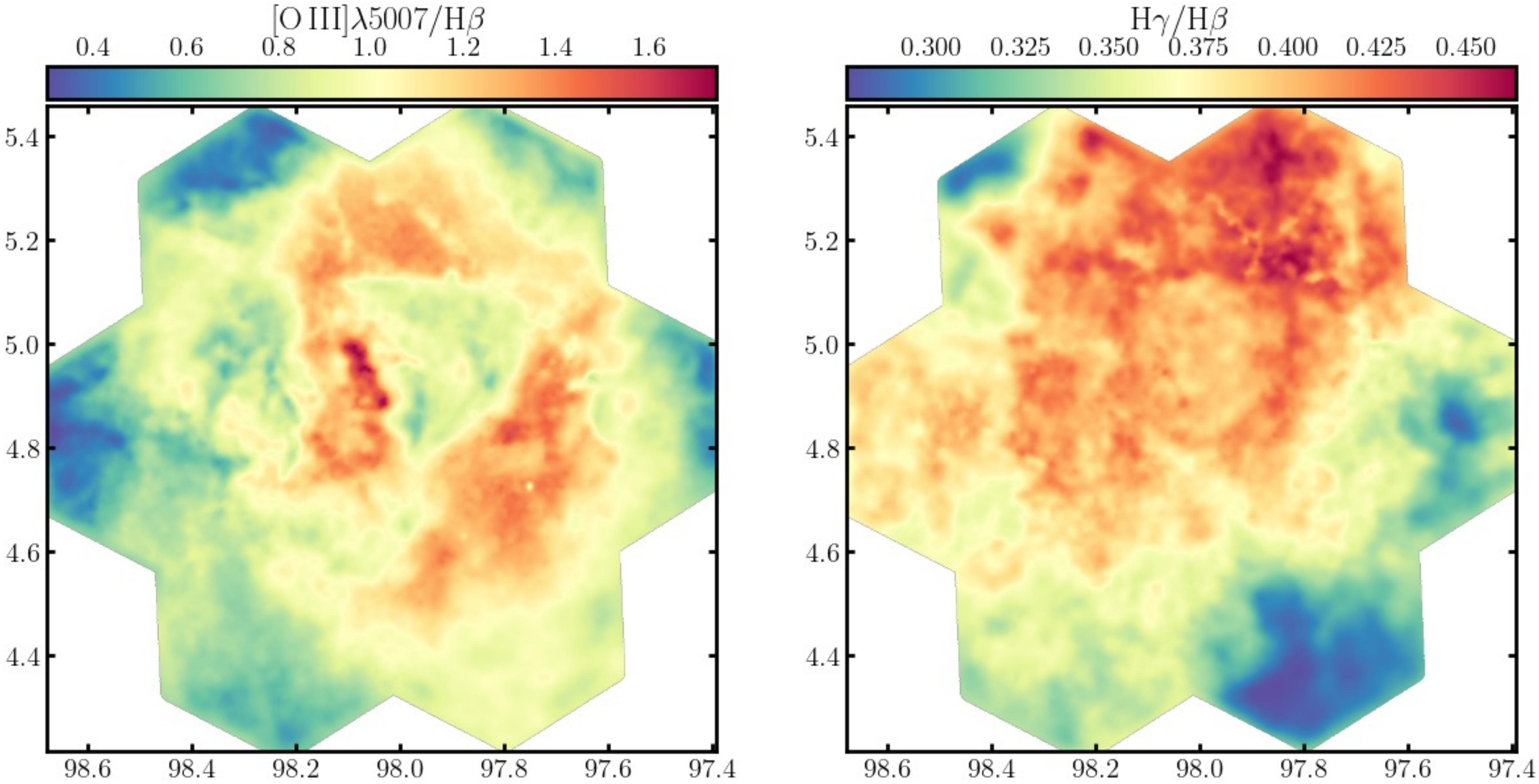


# Are the uncertainties correct?

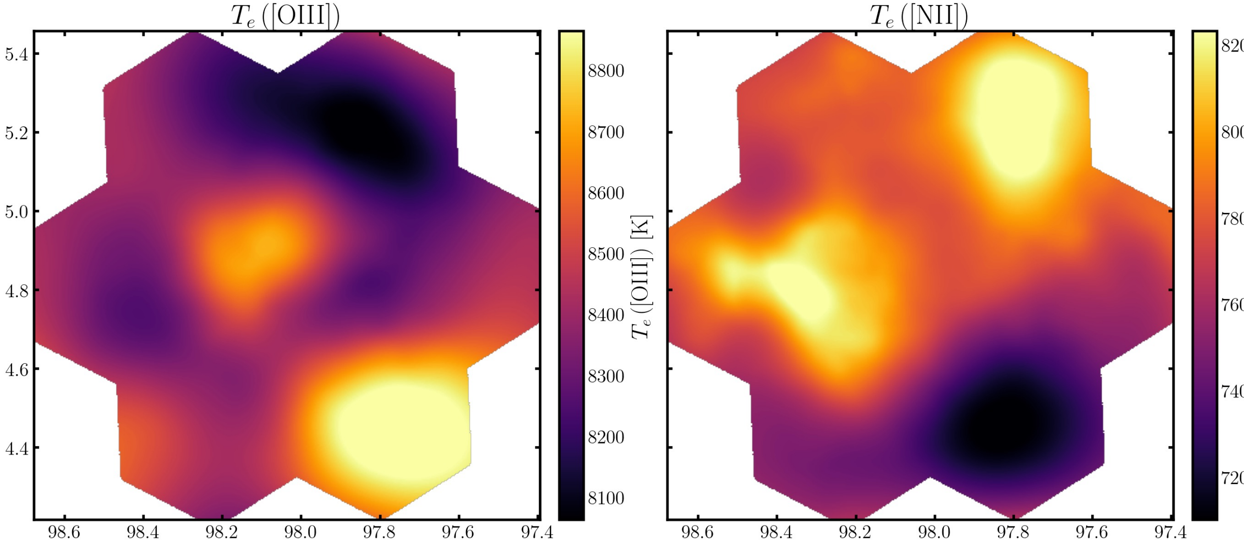


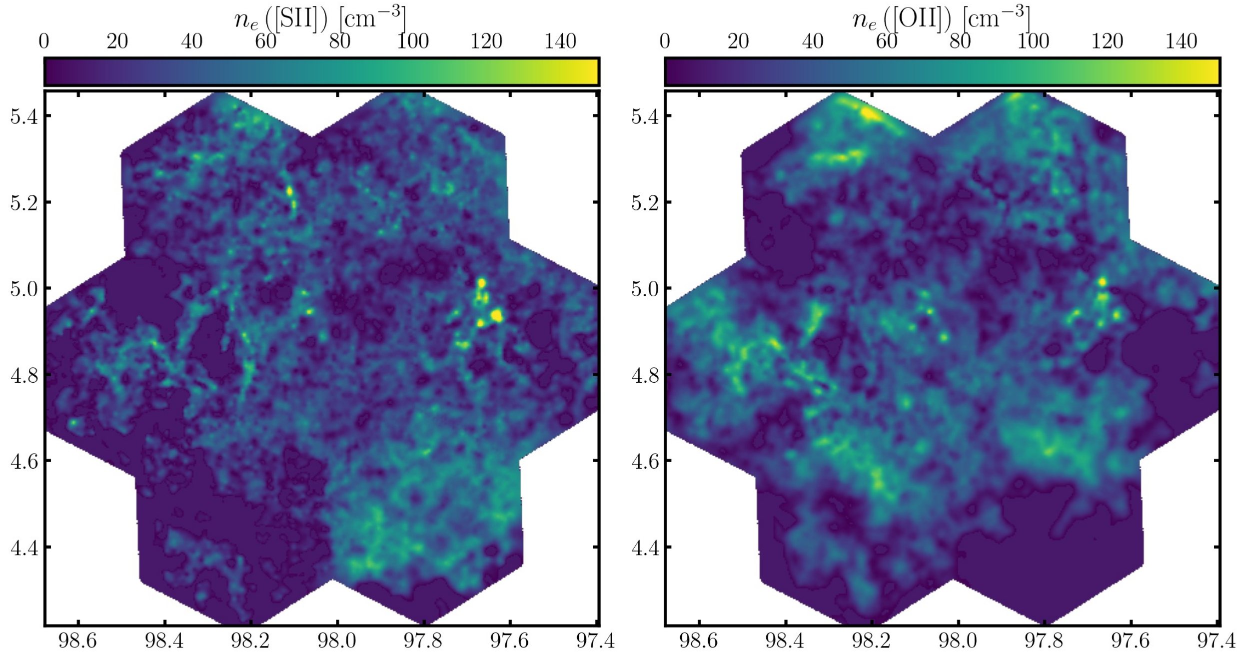

Mis-calibrated when there are **multiple fields** (sad)

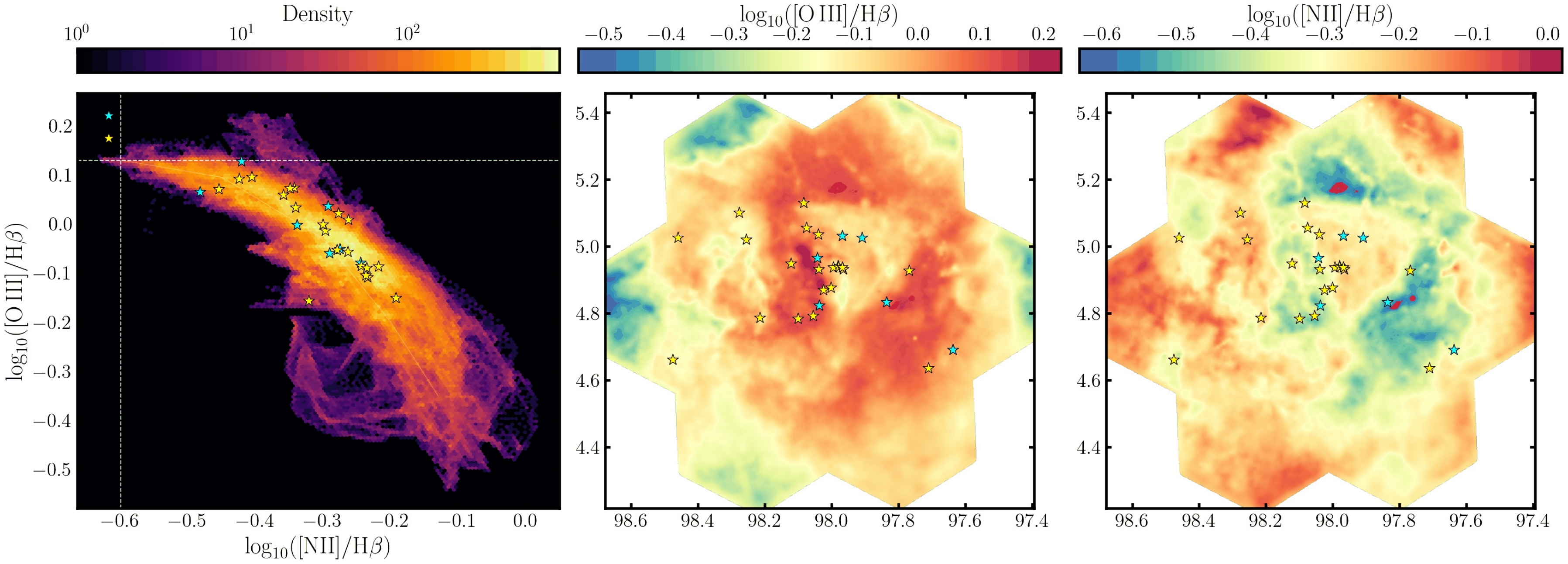
# **Line ratio field** as part of model



# Nebular conditions?

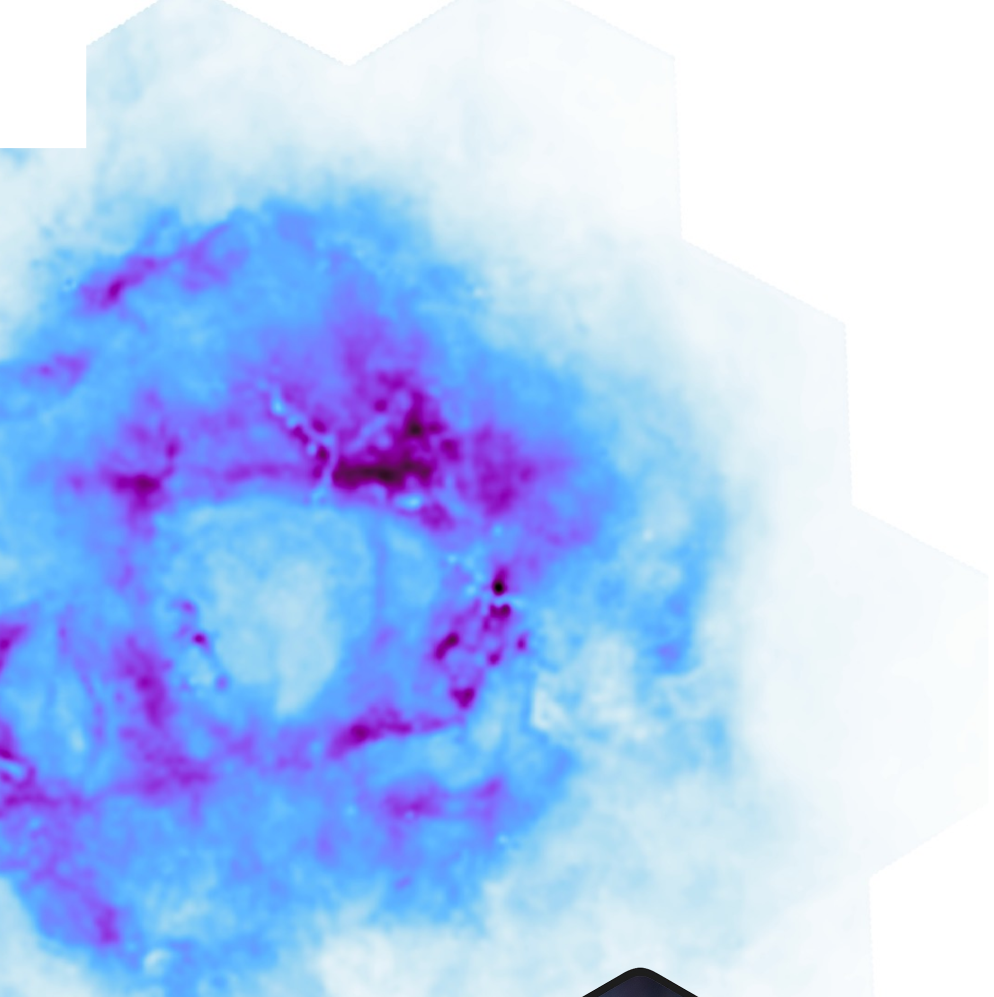


# Future

- ❑ Submit PhD thesis (…in 3 weeks)
- ❑ Calibrated uncertainties
- ❑ Applications to LVM H II regions
- ❑ Synergies via simultaneously modelling multiple surveys (SPHEREx, AMASE)
- ❑ Model **<u>physical quantities</u>** ($T_e$, $n_e$) as **<u>fields</u>**
- ❑ Continuing to work on LVM (LVM2?) at CCA



spectracles