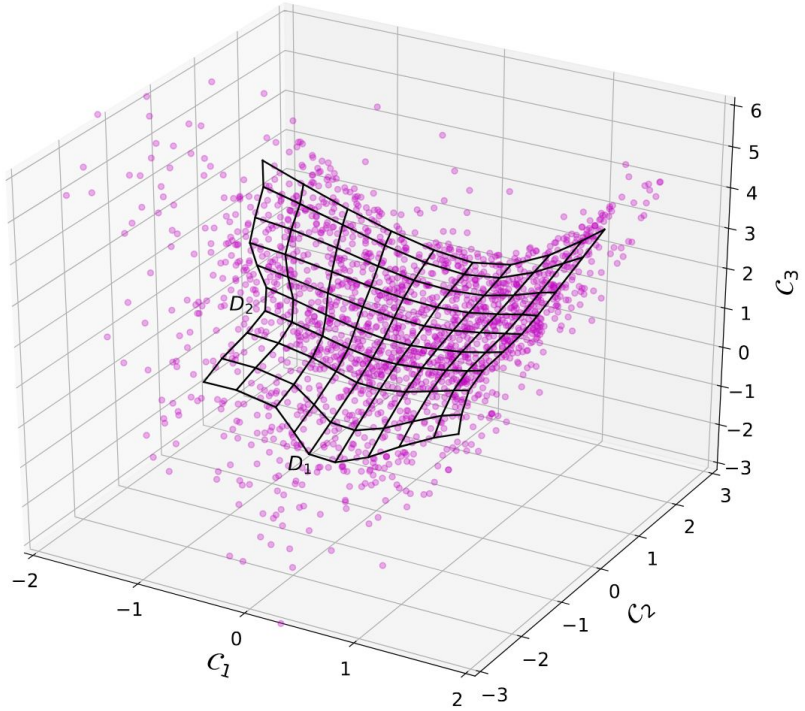
# Manifold learning to explore the galaxy parameter space

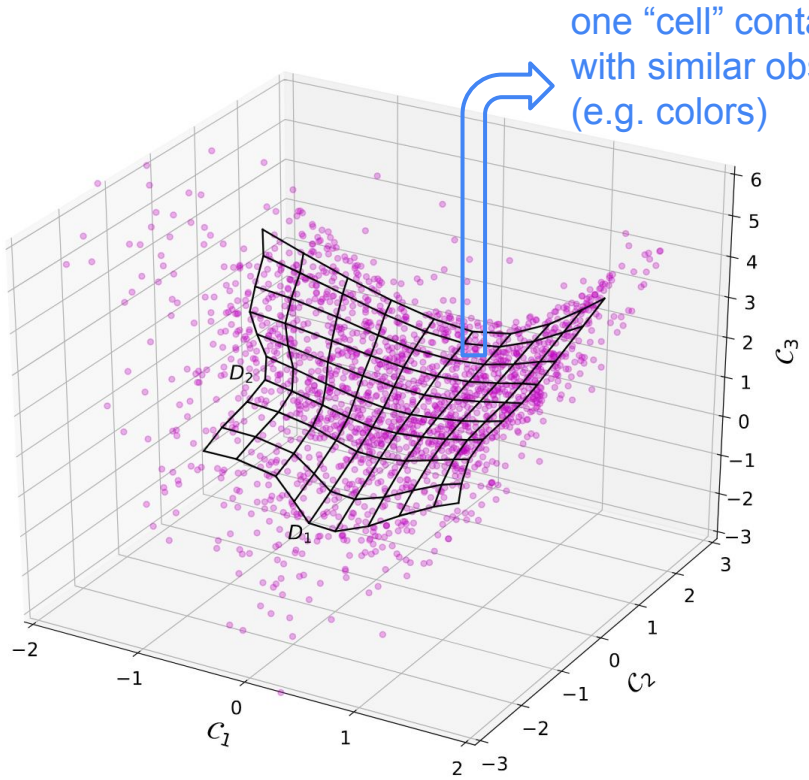
Tom Charnock, **Iary Davidzon**, Olivier Ilbert, Keerthana Jegatheesan, Clotilde Laigle, Marko Shuntov and the COSMOS team

# **Self-organizing maps**
(SOM, Kohonen 1981)



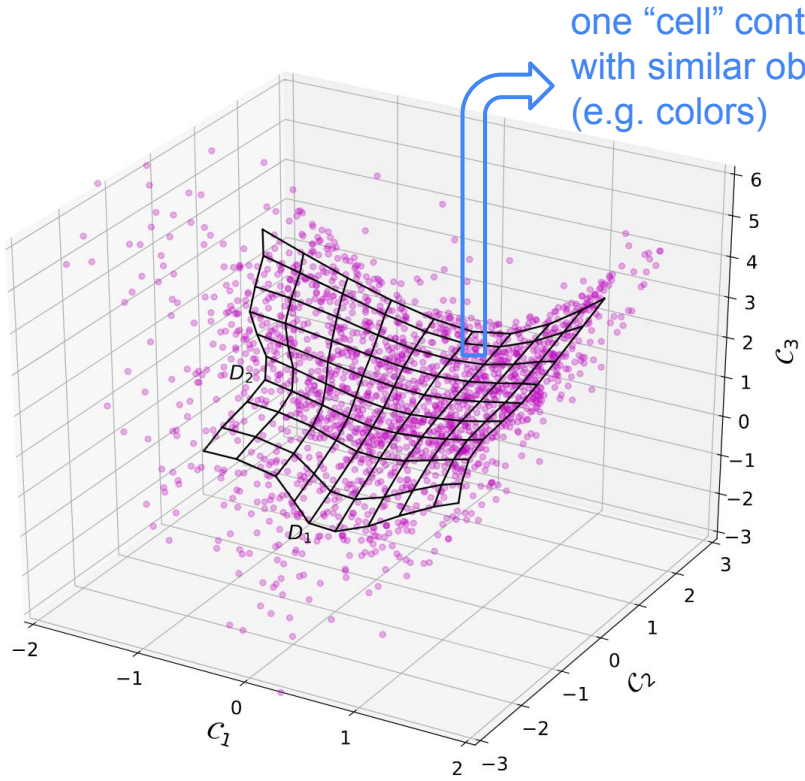

A dataset of photometric galaxies with observer's frame colors as features

# **Self-organizing maps**
(SOM, Kohonen 1981)

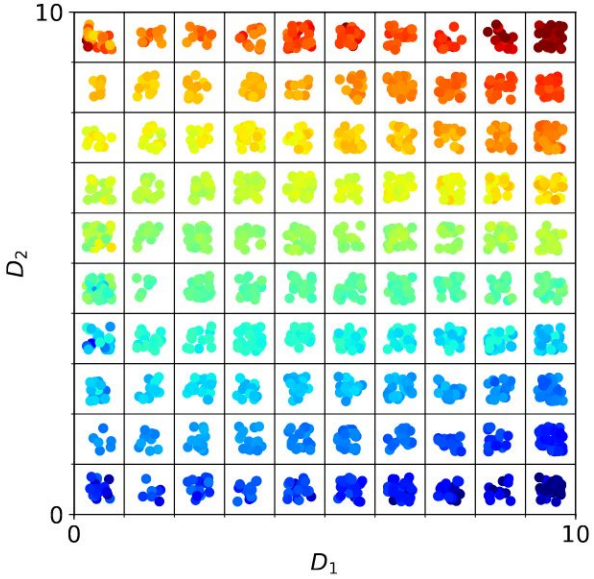



A dataset of photometric galaxies with observer’s frame colors as features

# Self-organizing maps
(SOM, Kohonen 1981)




dimensionality reduction

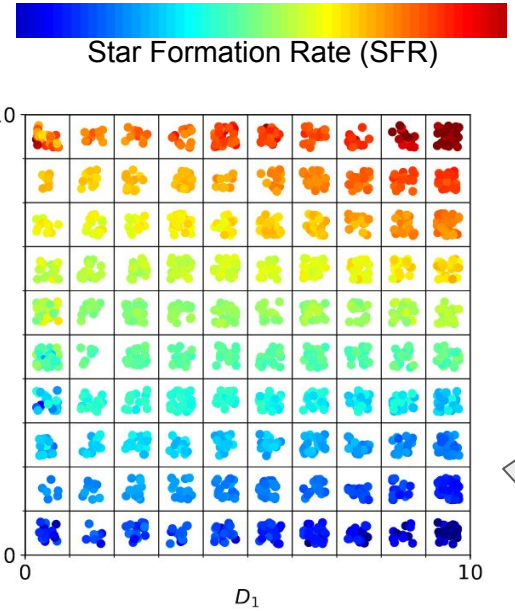



This is an **empirical SED classification *without* model templates**

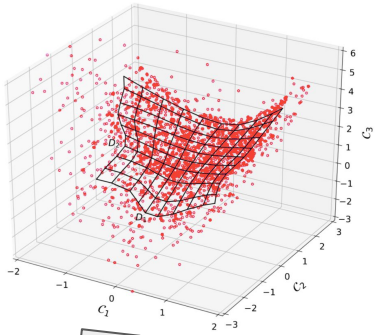
I.D., Laigle, et al. 2019

# One application: measuring physical parameters

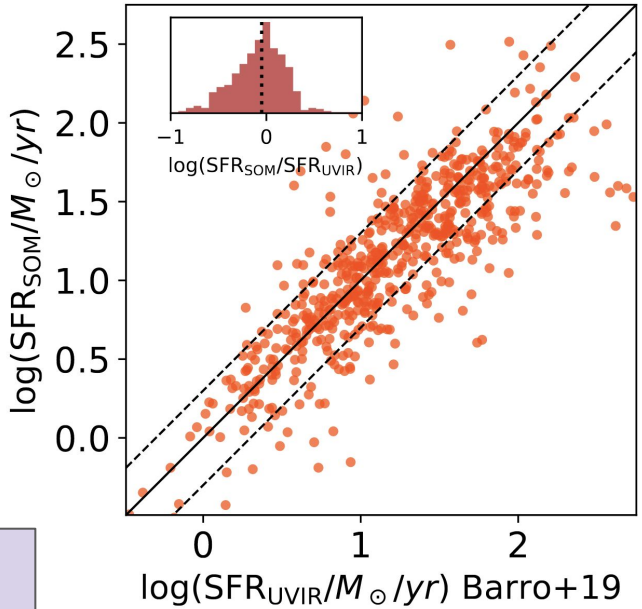
1. Labelling the grid with a calibration sample.




2. Other galaxies can be mapped...

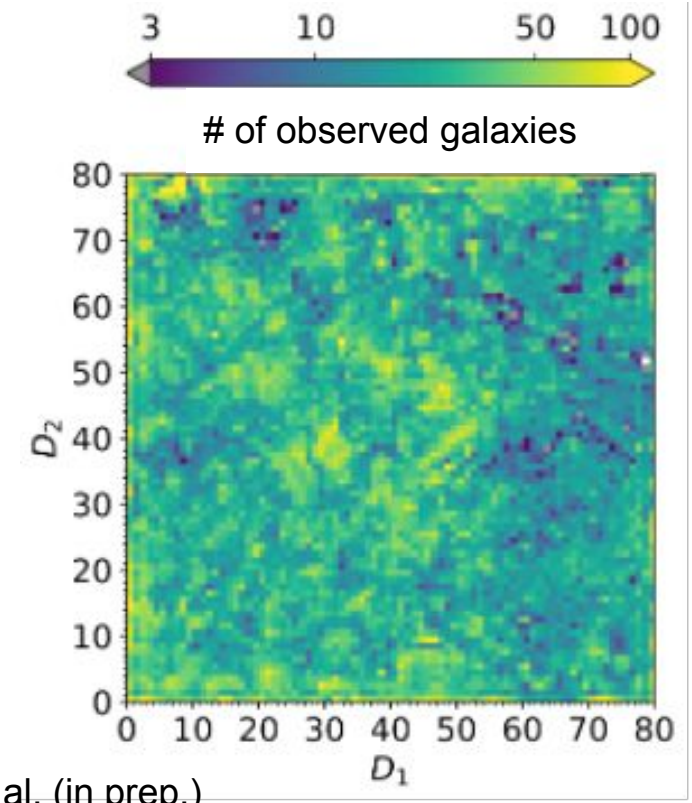


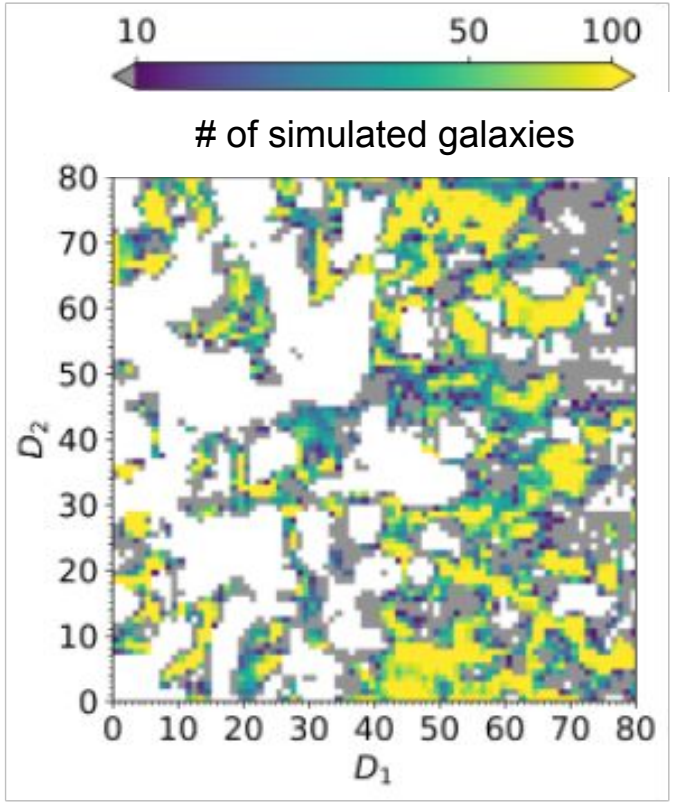
3. ...and they get *z*, M*, SFR estimates




See Dan Masters' talk, Hemmati et al. (2019) and soon I.D. et al. (in prep.)

# Another application: comparing two (n-dim.) distributions




Laigle et al. (in prep.)

# Earth mover distance to compare obs. and sim. in astrophysics

Something discussed with **Tom Charnock and Marko Shuntov**, but we never started working seriously on this...

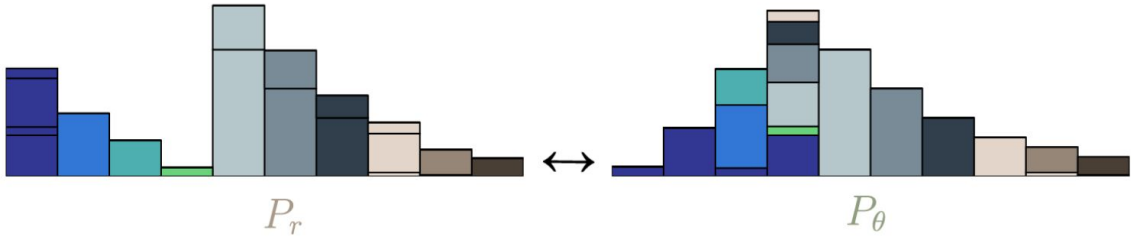



Fig.3: Optimal transportation between $P_r$ and $P_\theta$

Credit: Vincent herrmann

## On Wasserstein Two Sample Testing and Related Families of Nonparametric Tests

Aaditya Ramdas, Nicolas Garcia, Marco Cuturi

Nonparametric two sample or homogeneity testing is a decision theoretic problem that involves identifying differences between two random variables without making parametric assumptions about their underlying distributions. The literature is old and rich, with a wide variety of statistics having being intelligently designed and analyzed, both for the unidimensional and the multivariate setting. Our contribution is to tie together many of these tests, drawing connections between seemingly very different statistics. In this work, our central object is the Wasserstein distance, as we form a chain of connections from univariate methods like the Kolmogorov-Smirnov test, PP/QQ plots and ROC/ODC curves, to multivariate tests involving energy statistics and kernel based maximum mean discrepancy. Some connections proceed through the construction of a \textit{smoothed} Wasserstein distance, and others through the pursuit of a "distribution-free" Wasserstein test. Some observations in this chain are implicit in the literature, while others seem to have not been noticed thus far. Given nonparametric two sample testing's classical and continued importance, we aim to provide useful connections for theorists and practitioners familiar with one subset of methods but not others.

Comments: 18 pages
Subjects: **Statistics Theory (math.ST)**; Machine Learning (stat.ML)
Cite as: arXiv:1509.02237 **[math.ST]**