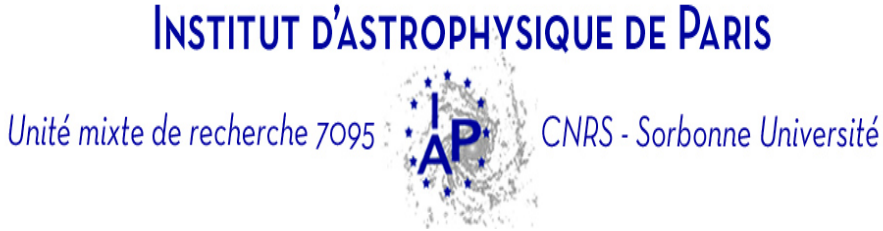
Generative Topographic Mapping for Redshift estimates

Is there merit in being principled ?

Grégoire Aufort

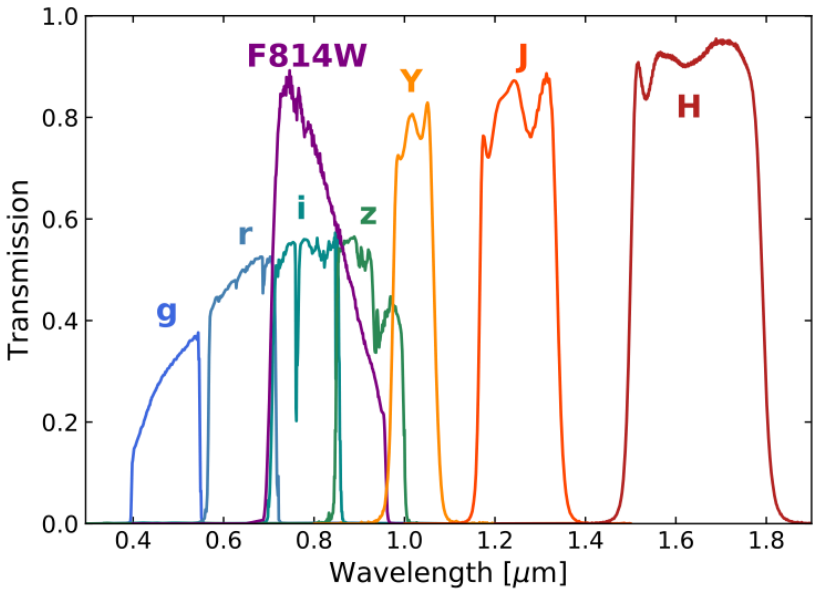
Joint work with Olivier Ilbert and Afshine Daurelle

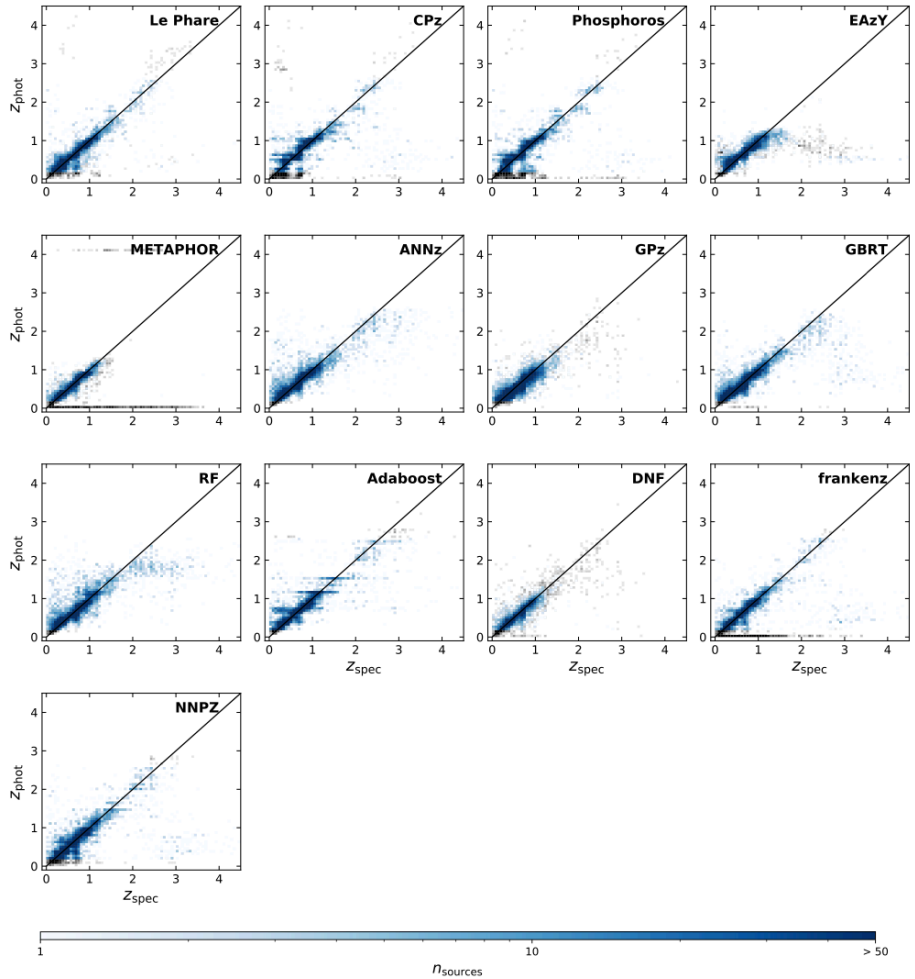
December 1st 2023



The Euclid photometric-redshift challenge

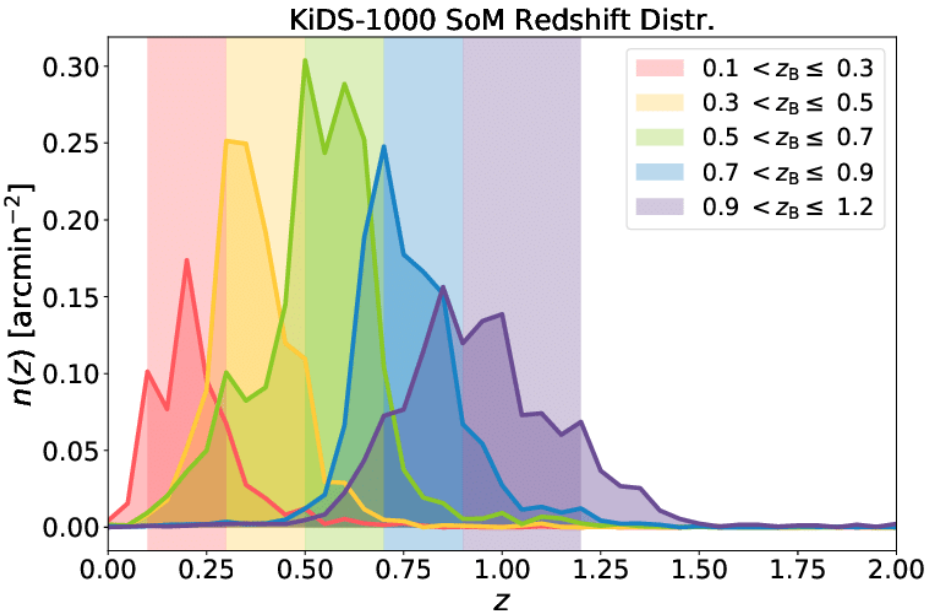
[Euclid collab : Desprez +22]



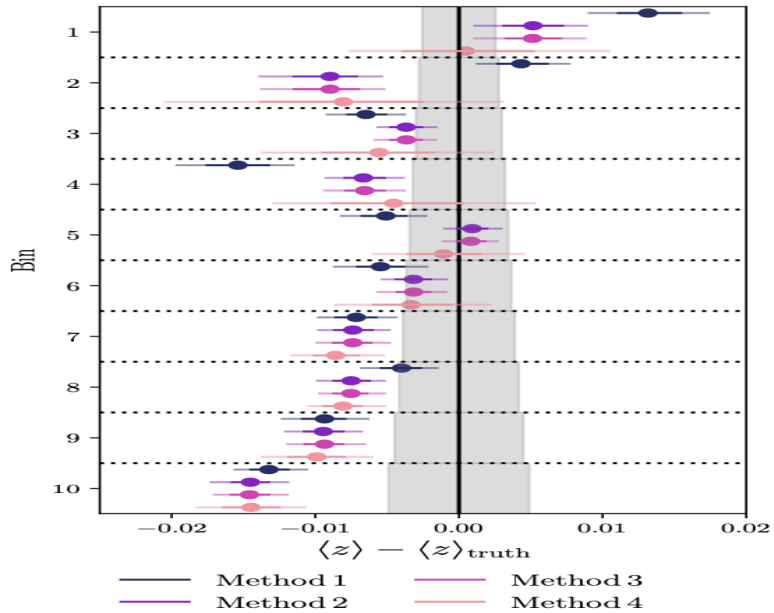


Tomographic bins average redshift challenge

[Euclid KIDS : Loureiro +21]



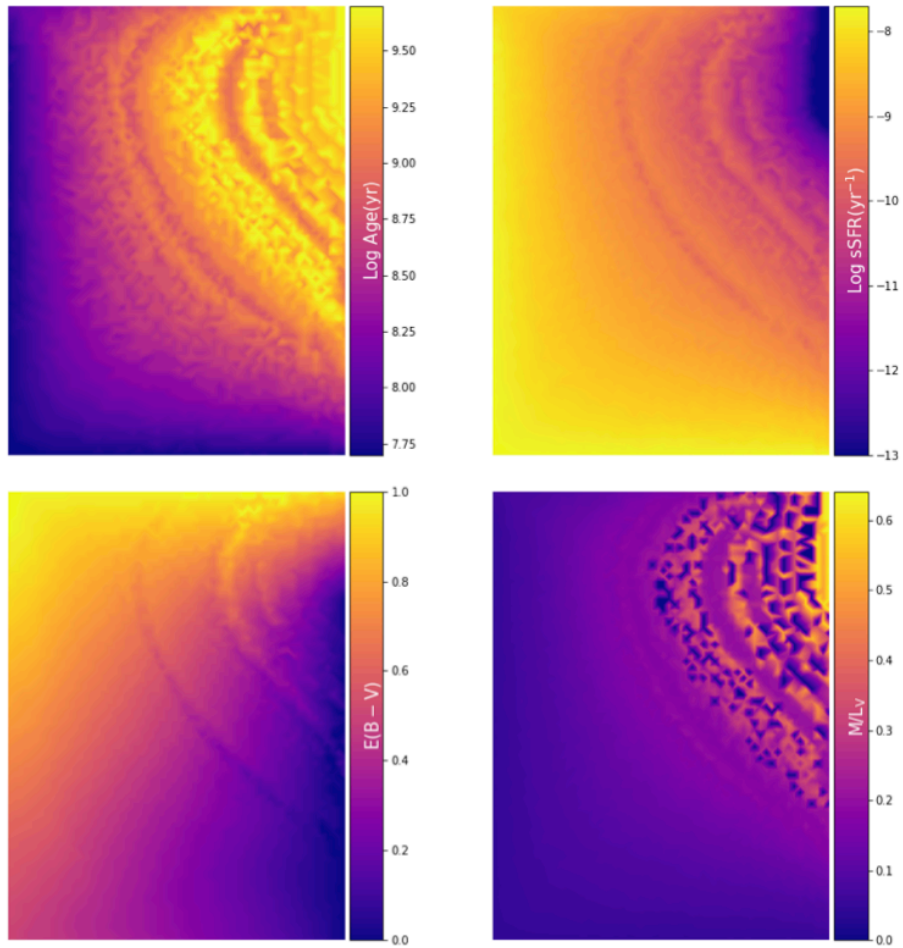
[Euclid collab : Naidoo +22]



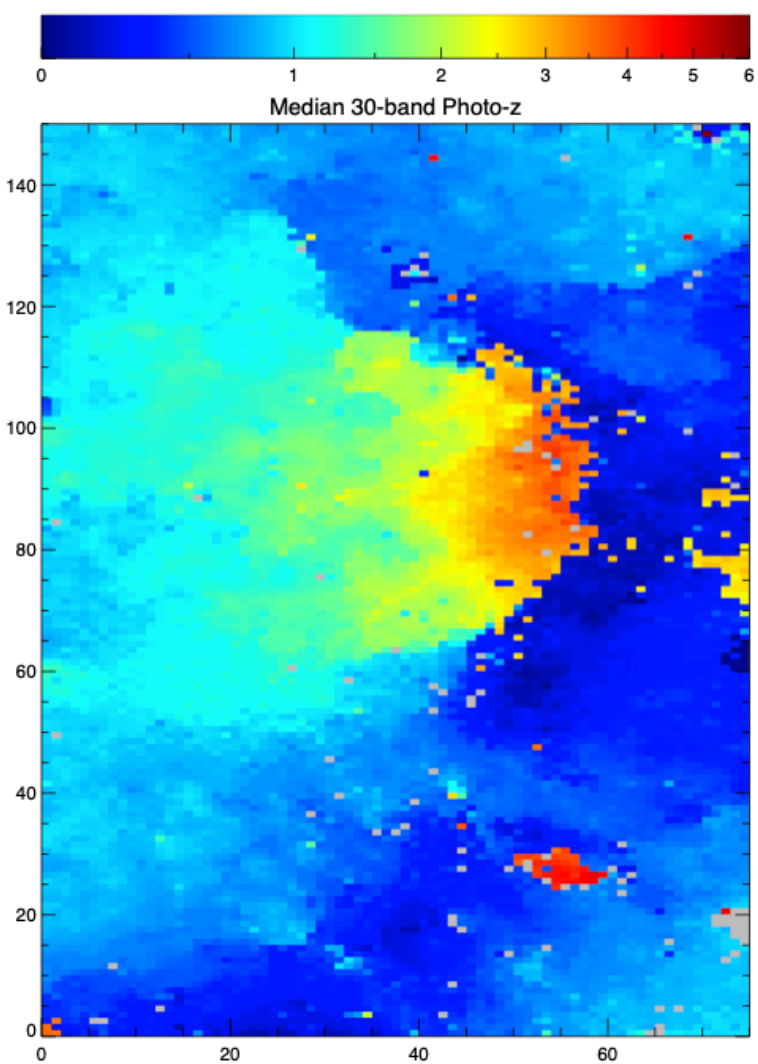
Self-Organizing Maps

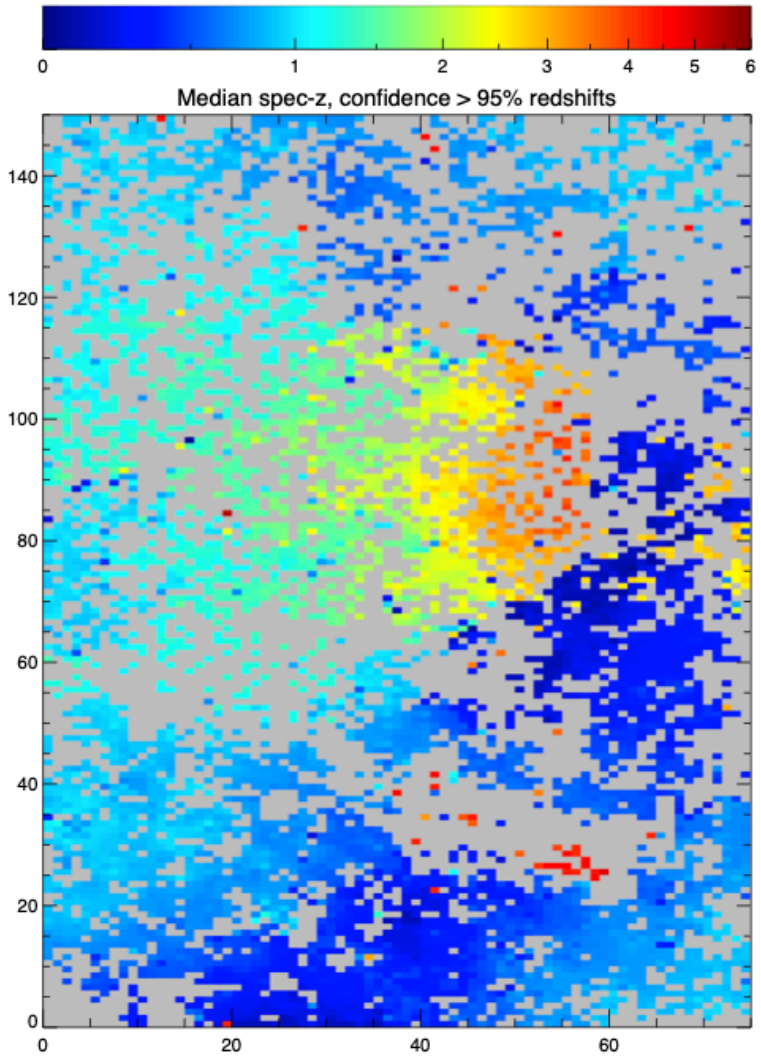
[Hemmati +19]

- Kohonen +82
- Dimensionality reduction algorithm
- Based on biologically inspired heuristic
- extremely popular for visualization



Self-Organizing Maps: redshift

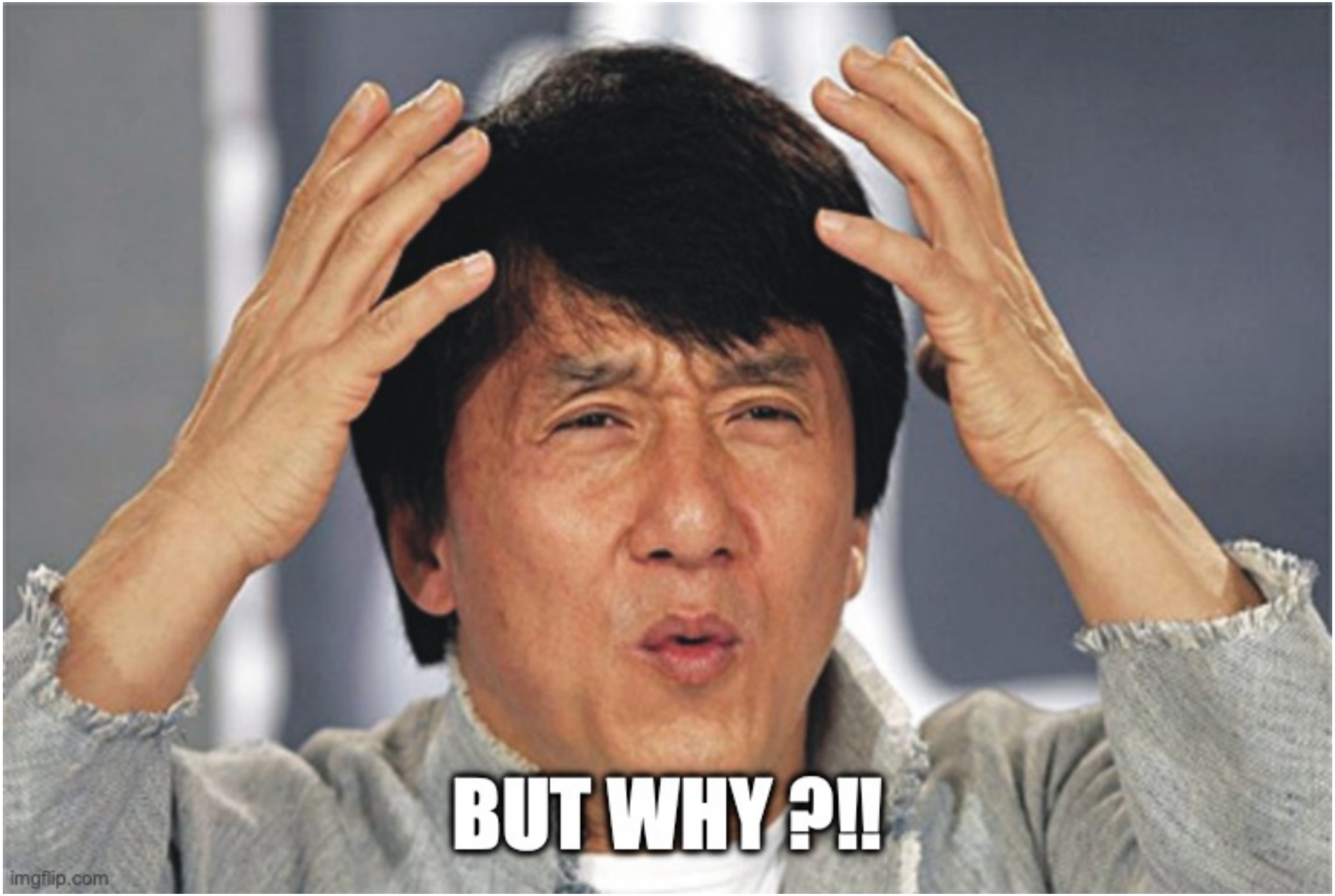




[Masters +17]

Baseline methodology: a SOM recipe

- Train a Self Organizing Map on the colors of a catalogue (with unknown spec-z)
- In each cell of the SOM, get a spec-z
- Label each cell of the SOM with the associated spec-z
- Label every galaxy in the original catalogue with the label of the cell they fall in
- For each tomographic bin :
  - For each galaxy in the bin :
    - If the spec-z label is in the interval give it weight 1
    - Otherwise discard it (give it weight 0)
  - Compute the weighted average


BUT WHY ?!!
imgflip.com


"It just works"

Questions

- Why work in this latent space ?
- Point estimates, no uncertainties ?
- Do we have to discard observations ?
- Can we mitigate spec-z errors ?
- what if we need n(z) ?

Questions

- Point estimates, no uncertainties ?
- What happens to unlabelled cells ?
- Can we mitigate spec-z errors ?
- what if we need n(z) ?

Can't we approximate the posterior distribution for each galaxy ?

A probabilistic SOM alternative

[Bishop +98]

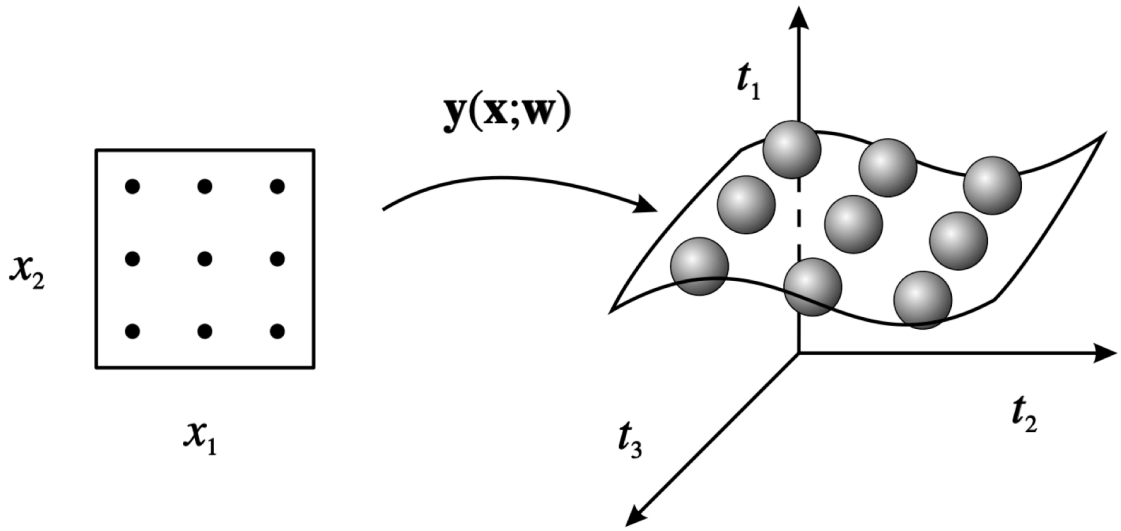


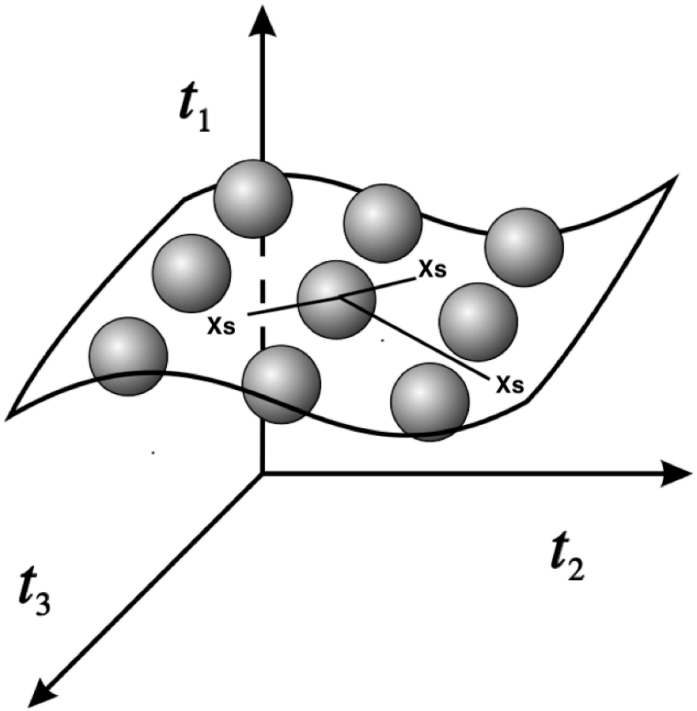
Figure 2: In order to formulate a latent variable model which is similar in spirit to the SOM, we consider a prior distribution $p(\mathbf{x})$ consisting of a superposition of delta functions, located at the nodes of a regular grid in latent space. Each node $\mathbf{x}_i$ is mapped to a corresponding point $\mathbf{y}(\mathbf{x}_i; \mathbf{W})$ in data space, and forms the centre of a corresponding Gaussian distribution.

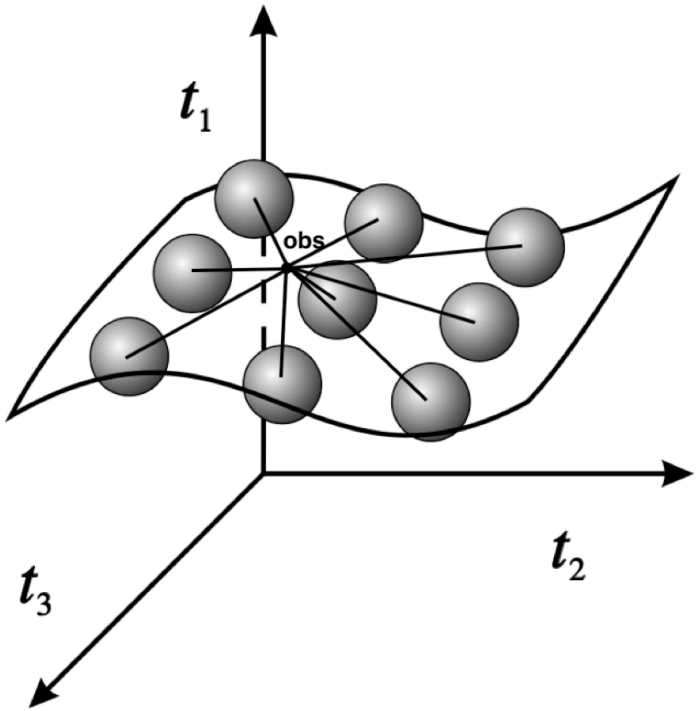
$$p(x) = \frac{1}{N} \sum_{i=1}^{K} \delta(x - x_i)$$

$$p(t|x, W, \beta) = \left(\frac{\beta}{2\pi}\right) \exp\left\{-\frac{\beta}{2}||y(x; W) - t||^2\right\}$$

$$L(W, \beta) = \sum_{n=1}^{N} ln\left\{\frac{1}{N} \sum_{i=1}^{K} p(t_n|x_i, W, \beta)\right\}$$

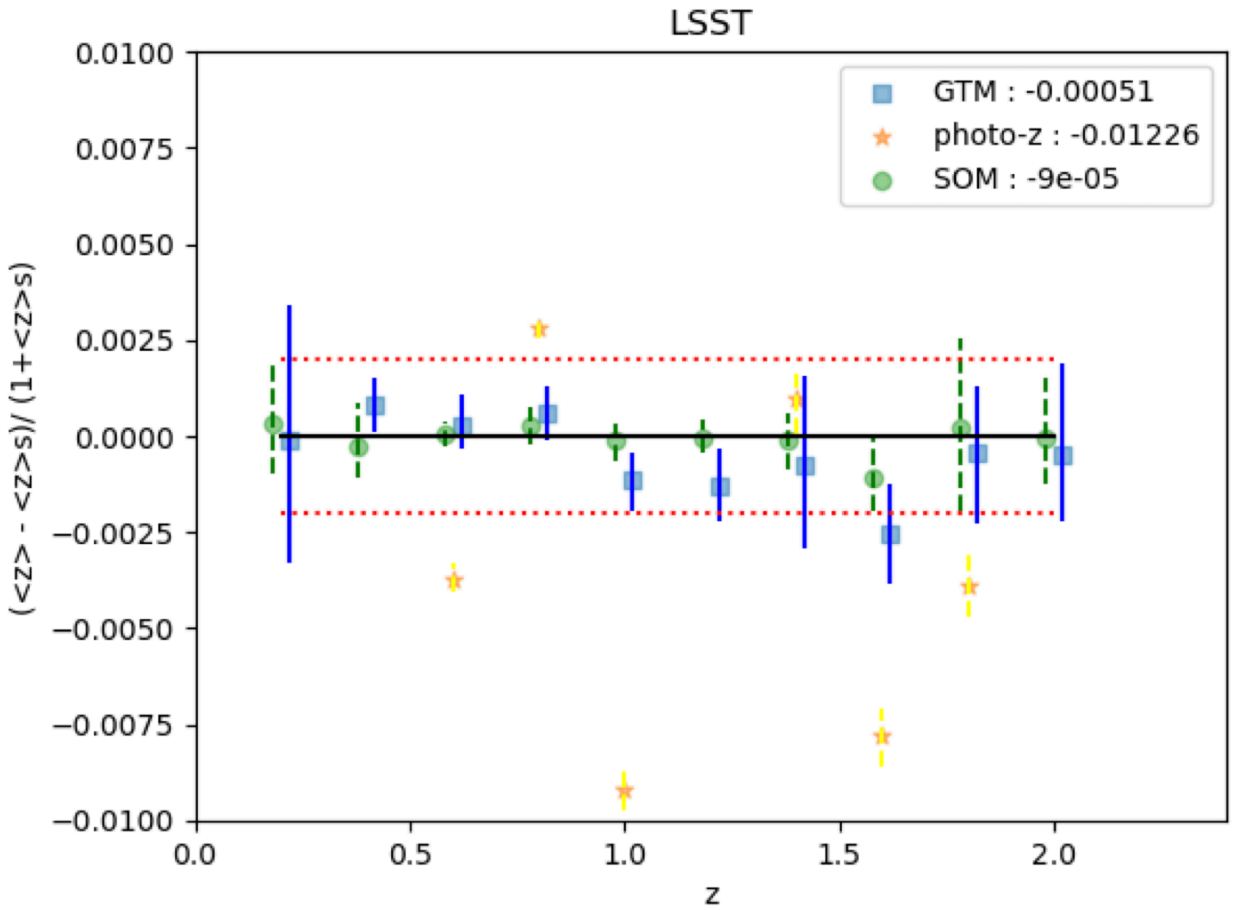
Estimating Redshift from the GTM



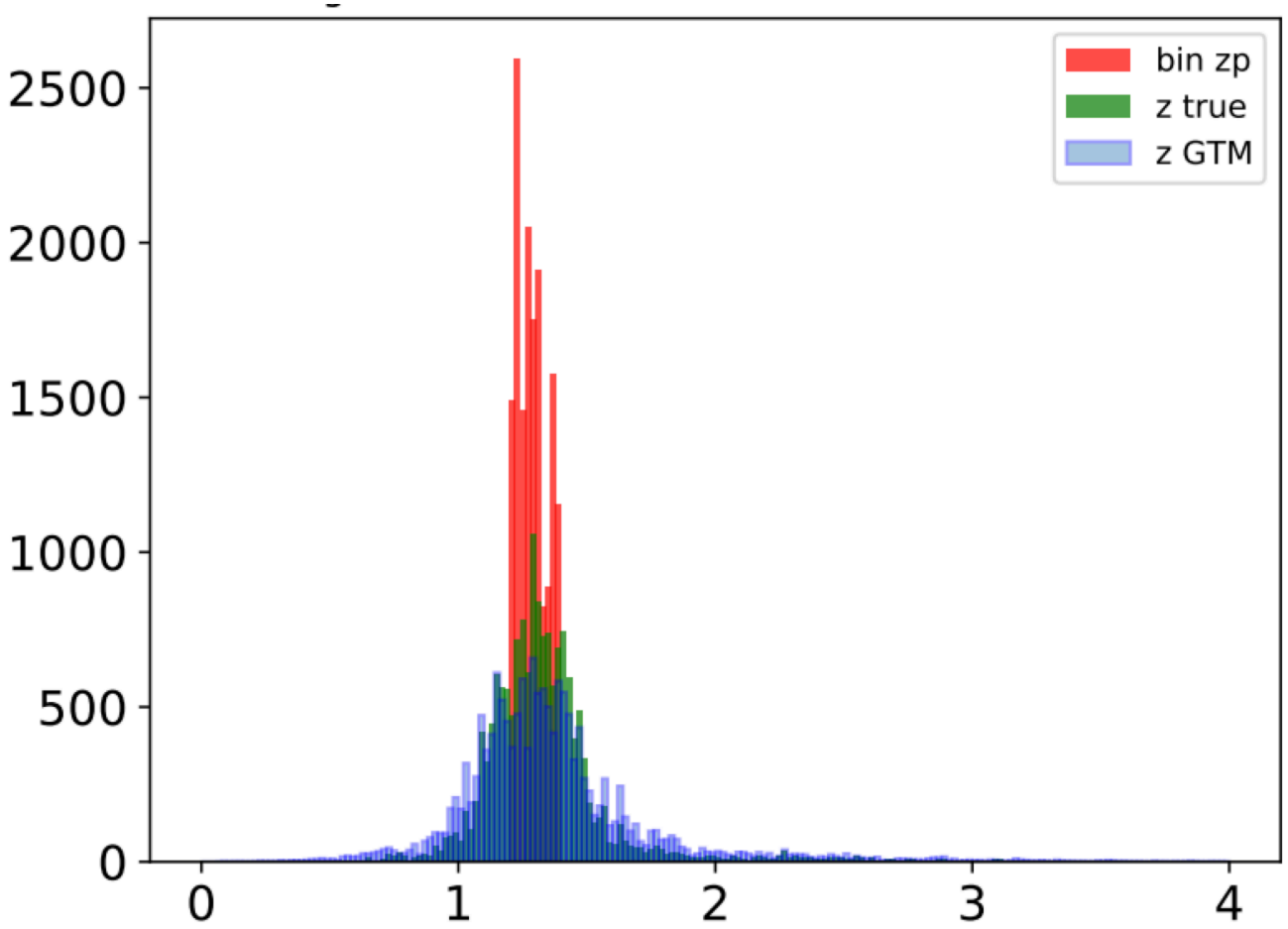


$$p(z|obs) = \sum_{k=1}^{K} p(z|t_k)p(t_k|obs)$$

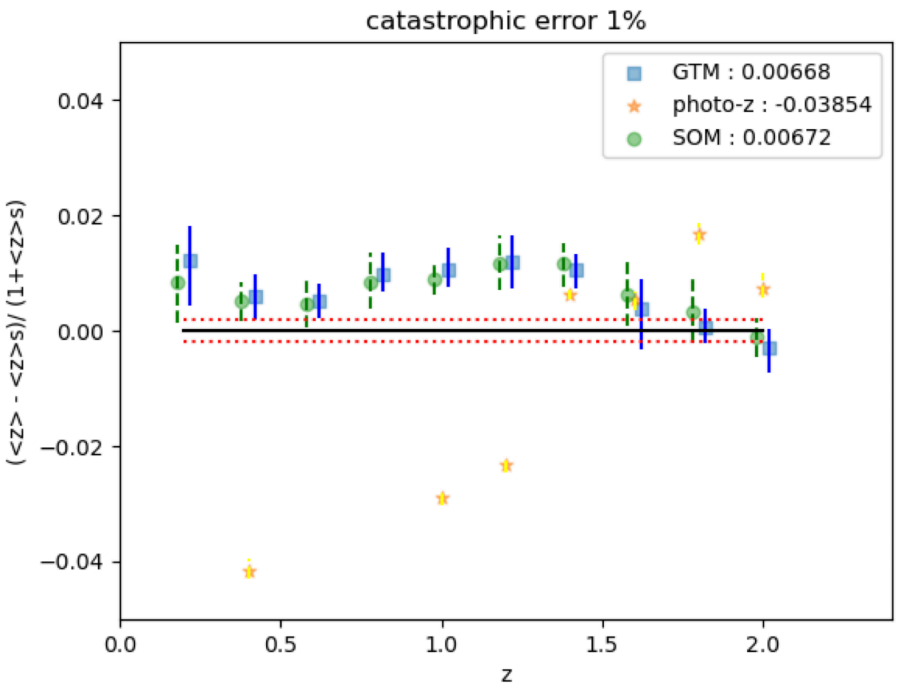
Preliminary Results : $\langle z \rangle$

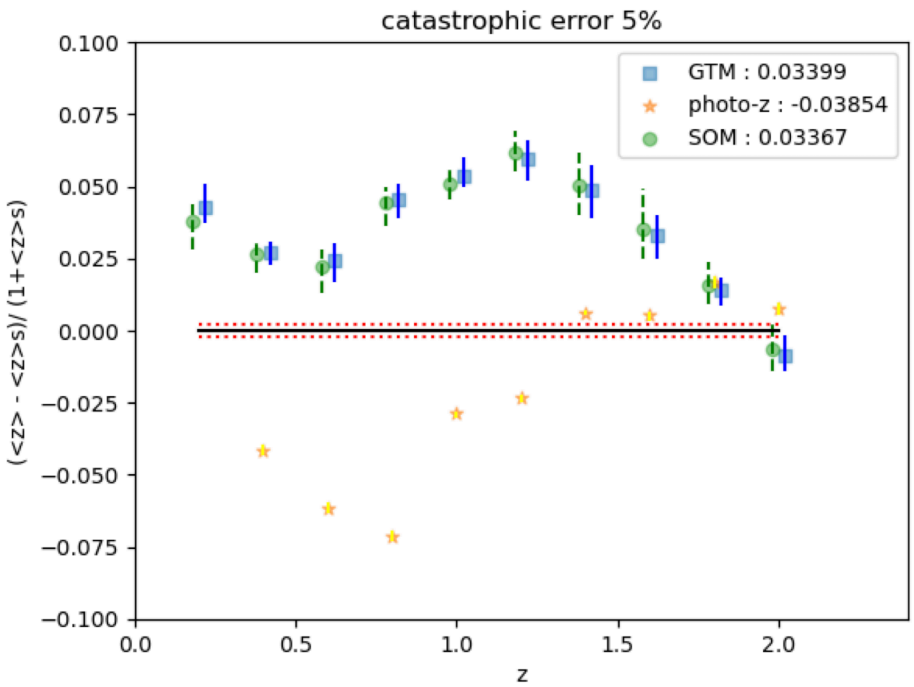


Preliminary Results : N(z)



Preliminary Results : Catastrophic spec-z errors





Conclusions

We propose the use principled alternative to the SOM for photometric redshift estimates

- The preliminary results are (almost) on par with the SOM
- We don't have to throw anything away

What's planned ?

- Bridge the gap between the two approaches