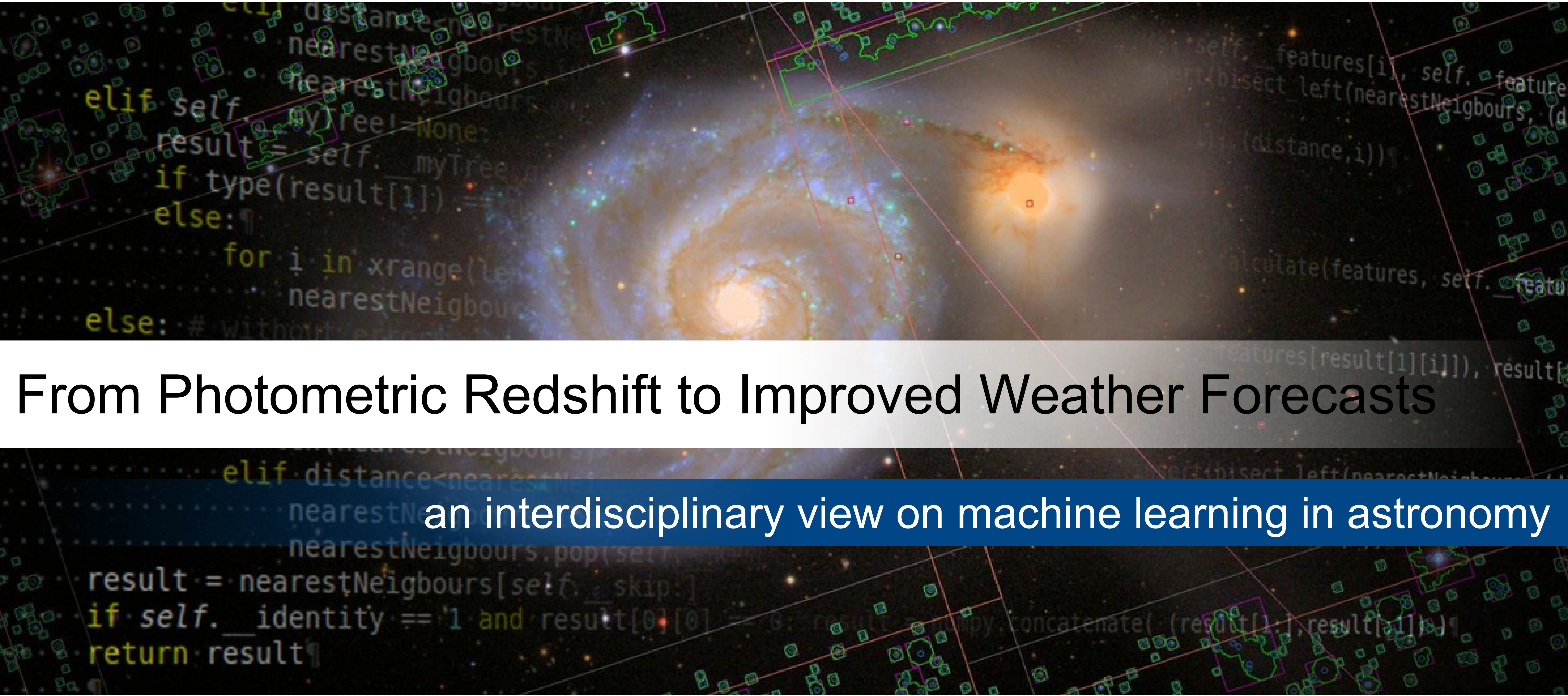
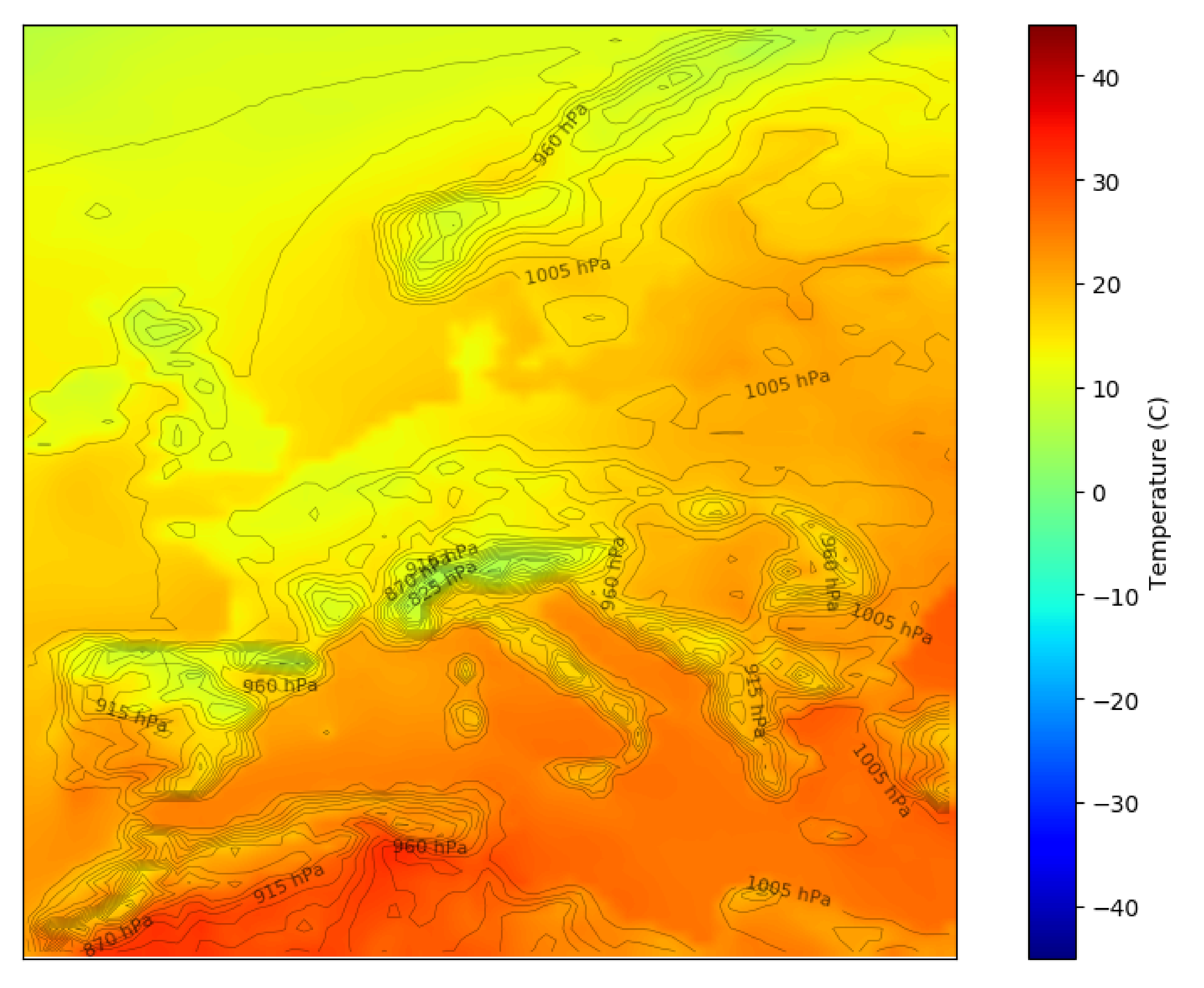
Heidelberg Institute for
Theoretical Studies



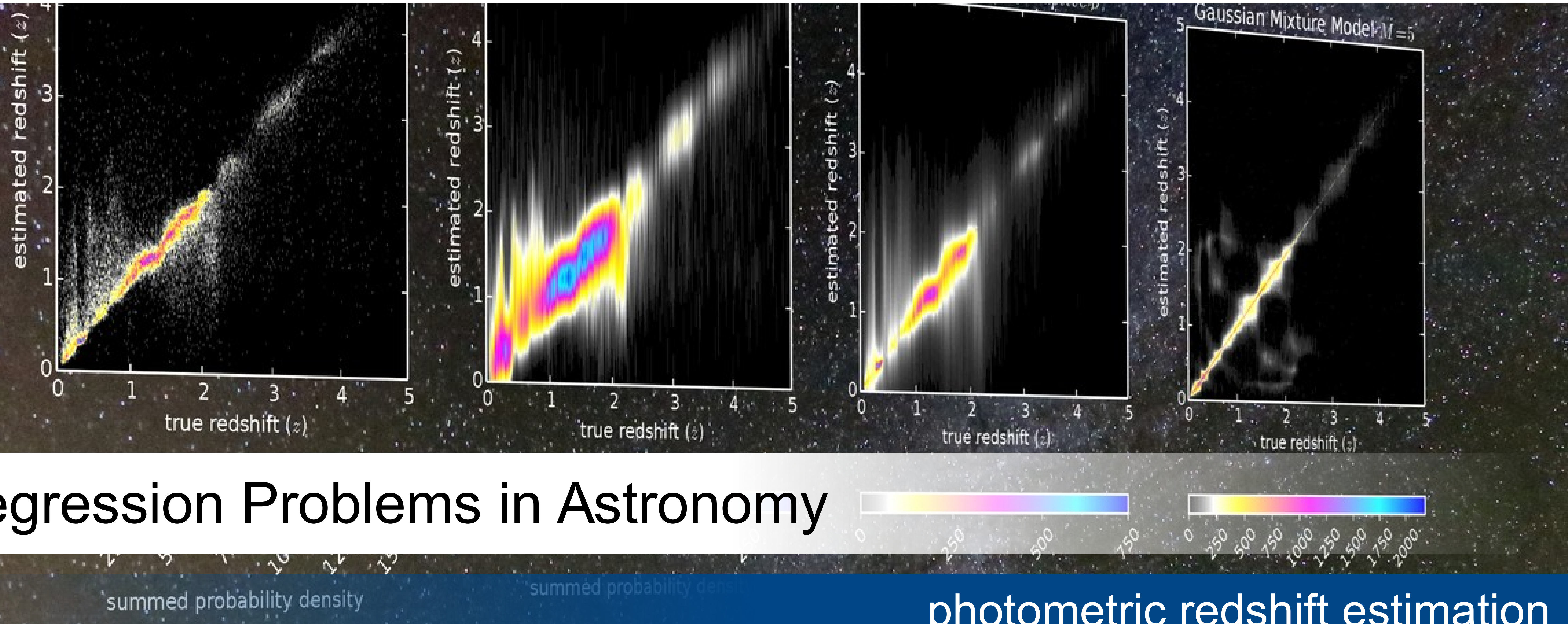
From Photometric Redshift to Improved Weather Forecasts

an interdisciplinary view on machine learning in astronomy


Event Horizon Telescope 2019



960 hPa
1005 hPa
1005 hPa
910 hPa
870 hPa
825 hPa
960 hPa
960 hPa
1005 hPa
960 hPa
915 hPa
915 hPa
1005 hPa
960 hPa
915 hPa
870 hPa
1005 hPa
40
30
20
10
0
−10
−20
−30
−40
Temperature (C)



Regression Problems in Astronomy

photometric redshift estimation

model

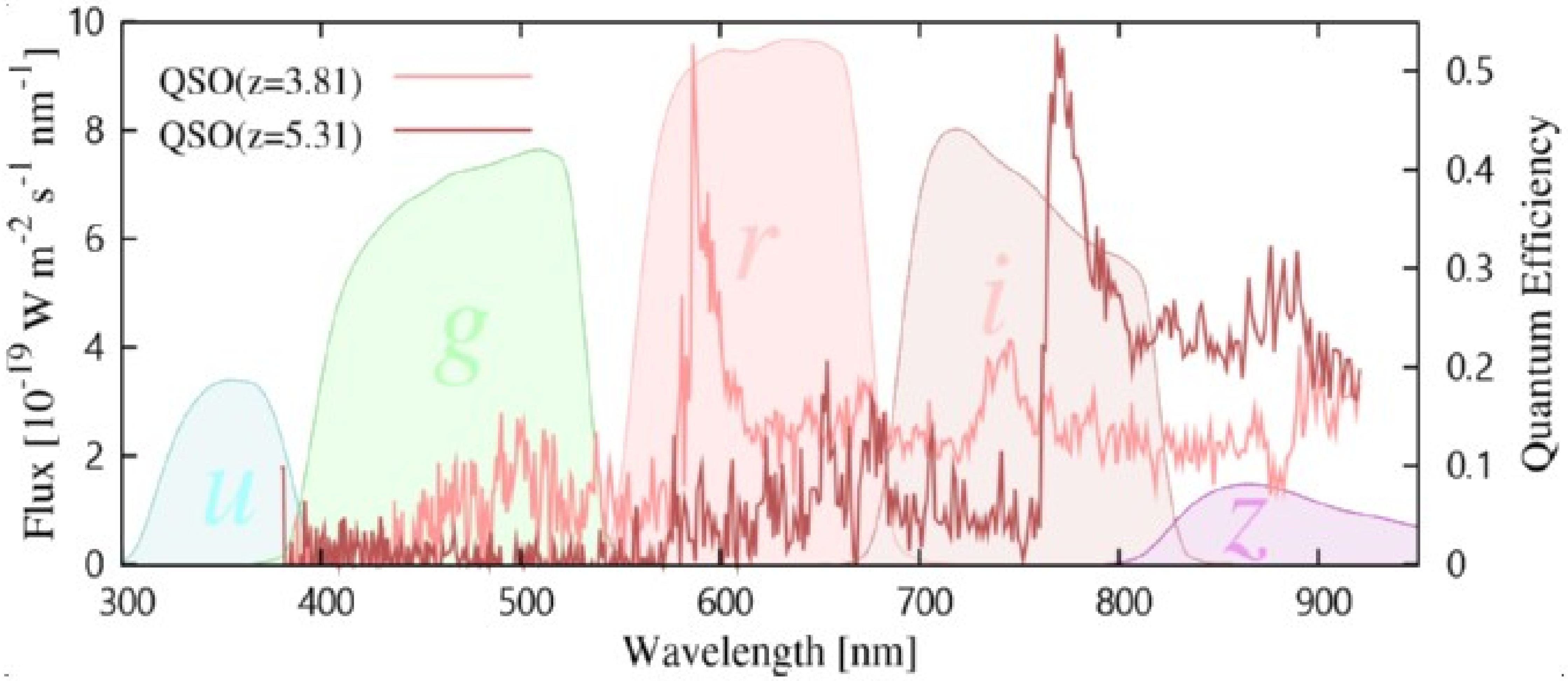
estimate

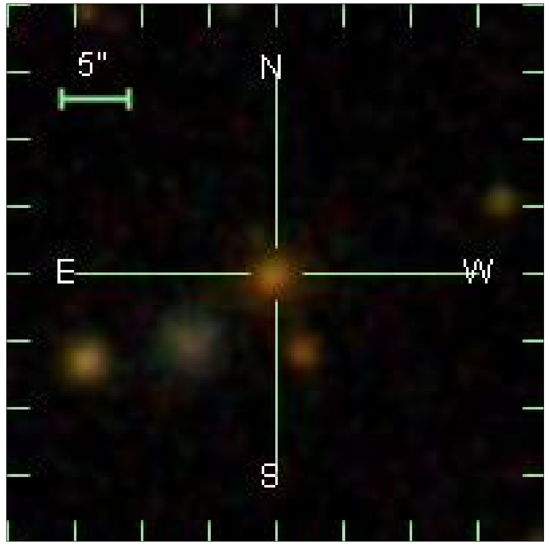
$$f(\vec{x}) \rightarrow y\,,\; where\; \vec{x} \in \mathbb{R}^n\,,\; y \in \mathbb{R}$$

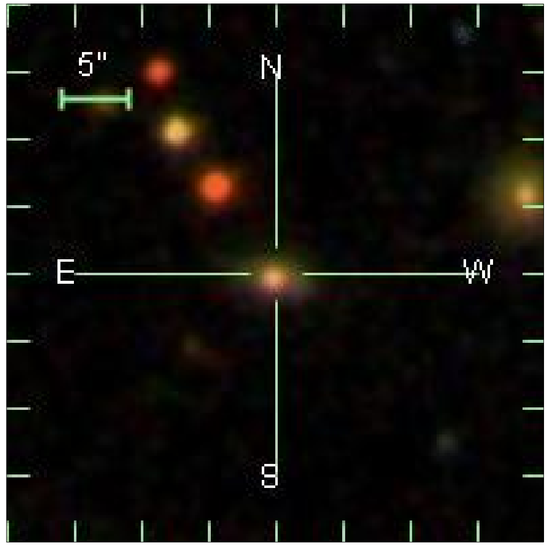
features

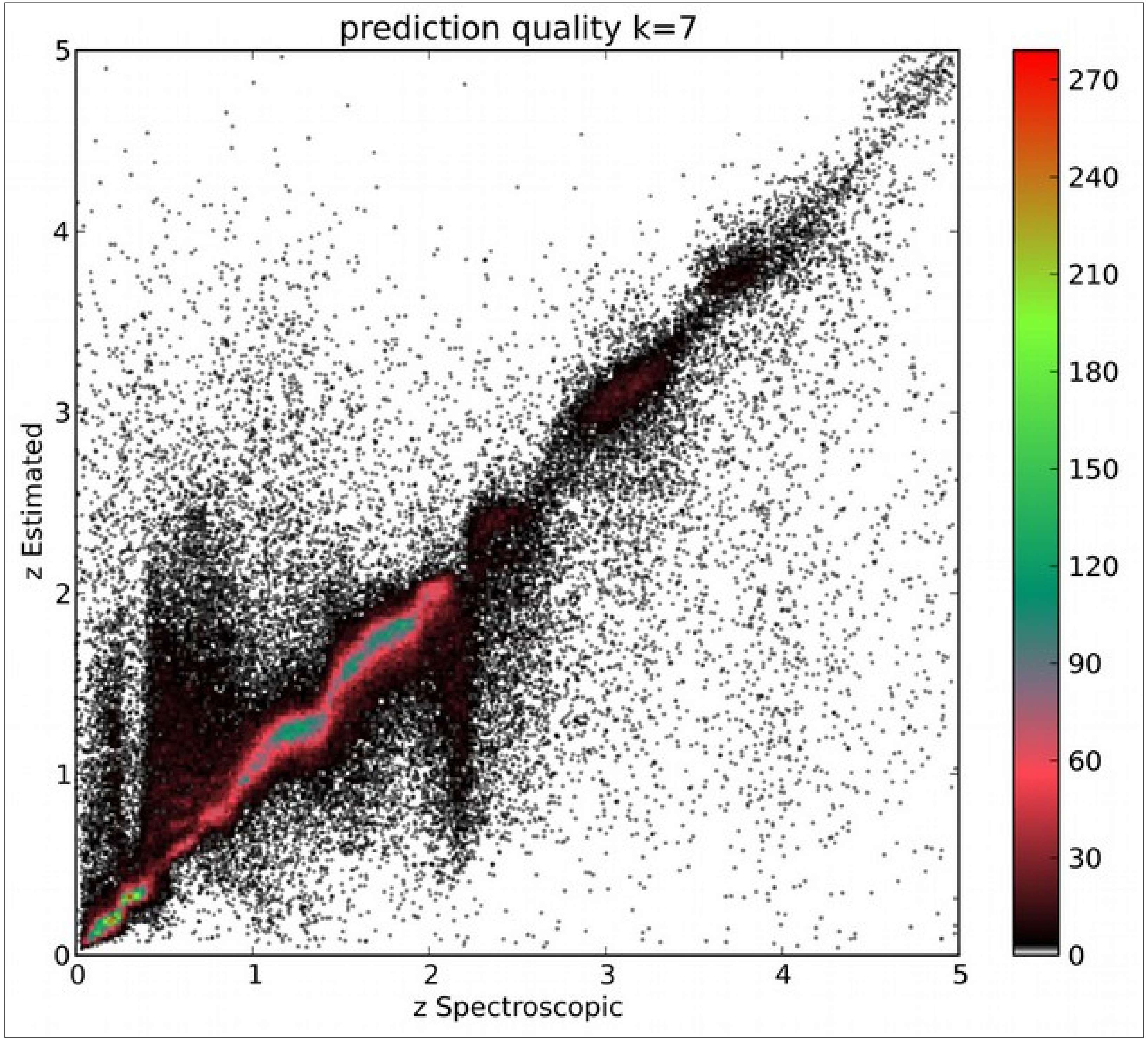
Estimating Redshift Photometrically

$$1 + z = \frac{\lambda}{\lambda_0}$$









Experiment with all quasars in SDSS DR7

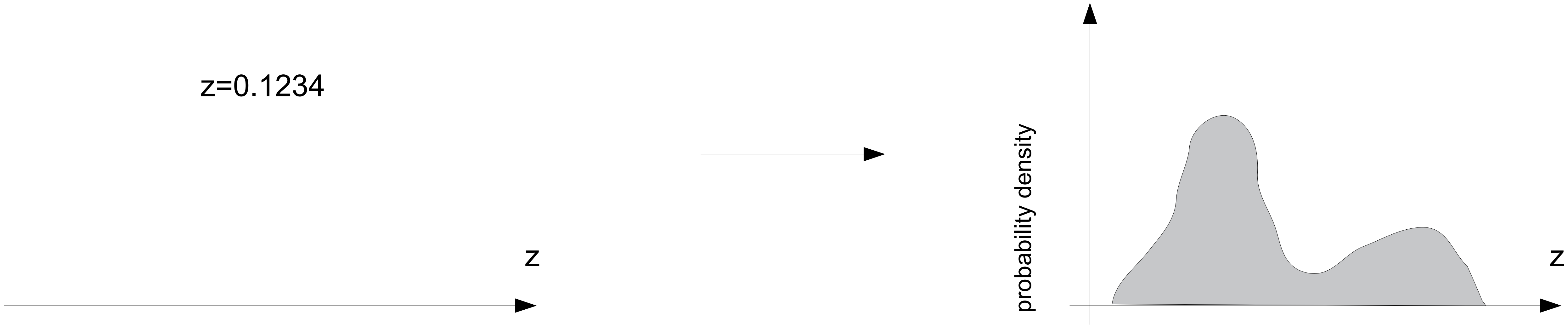
Uncertainties

$$f(\vec{x}) \rightarrow y, \text{ where } \vec{x} \in \mathbb{R}^n, y \in \mathbb{R}$$

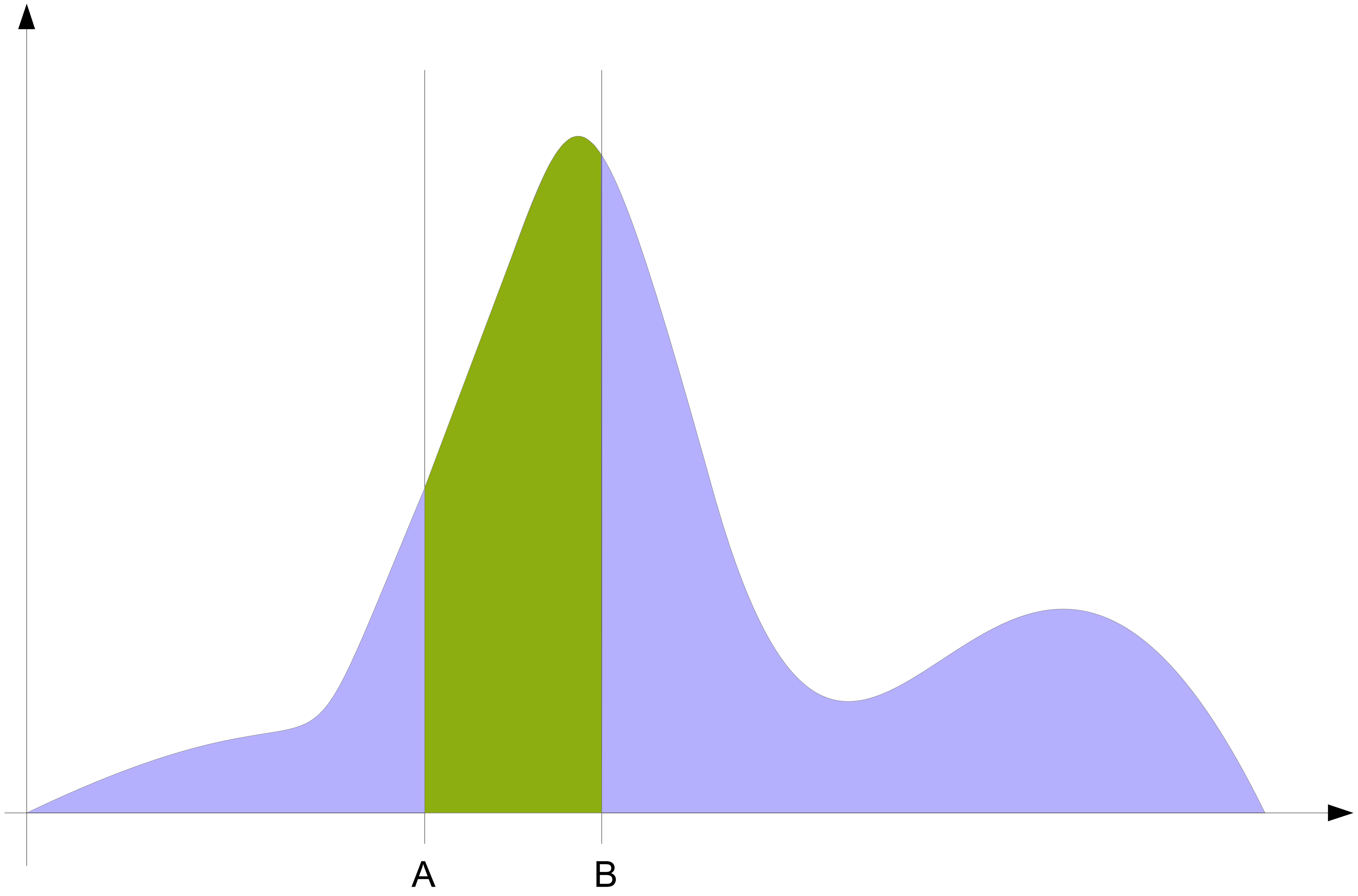
model uncertainty

input uncertainty

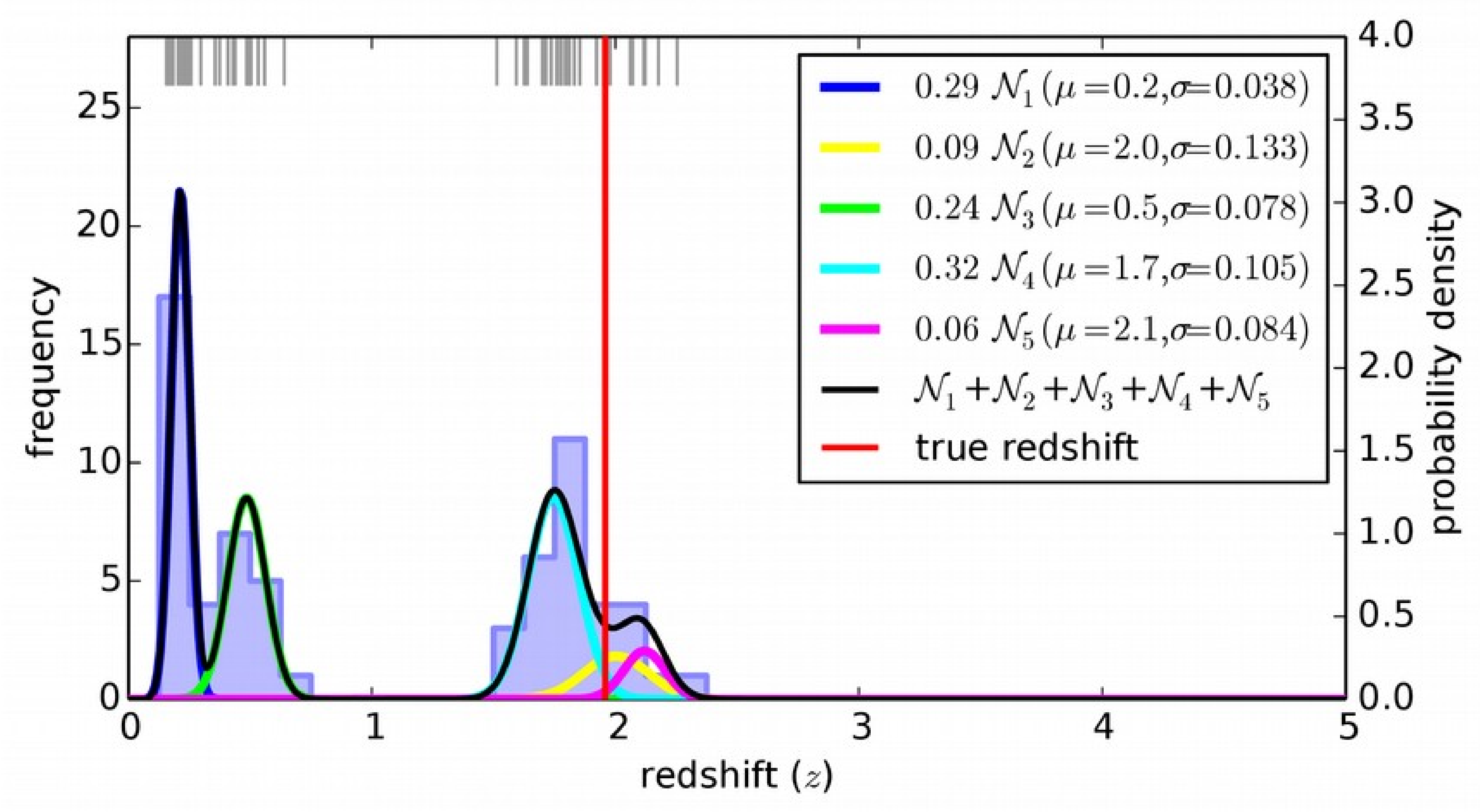
uncertain estimate



Probability Density Function

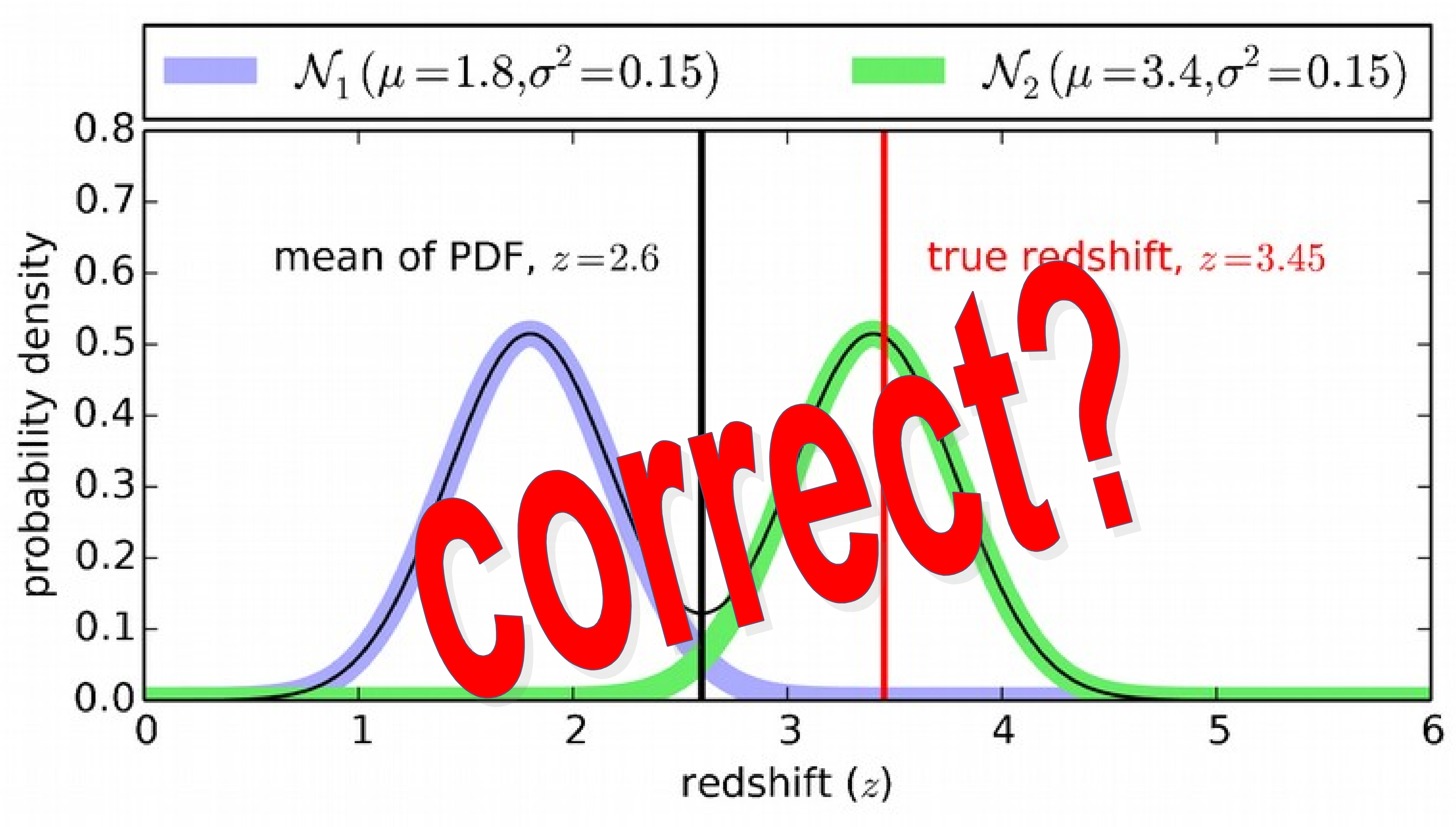


Multi-Modalities



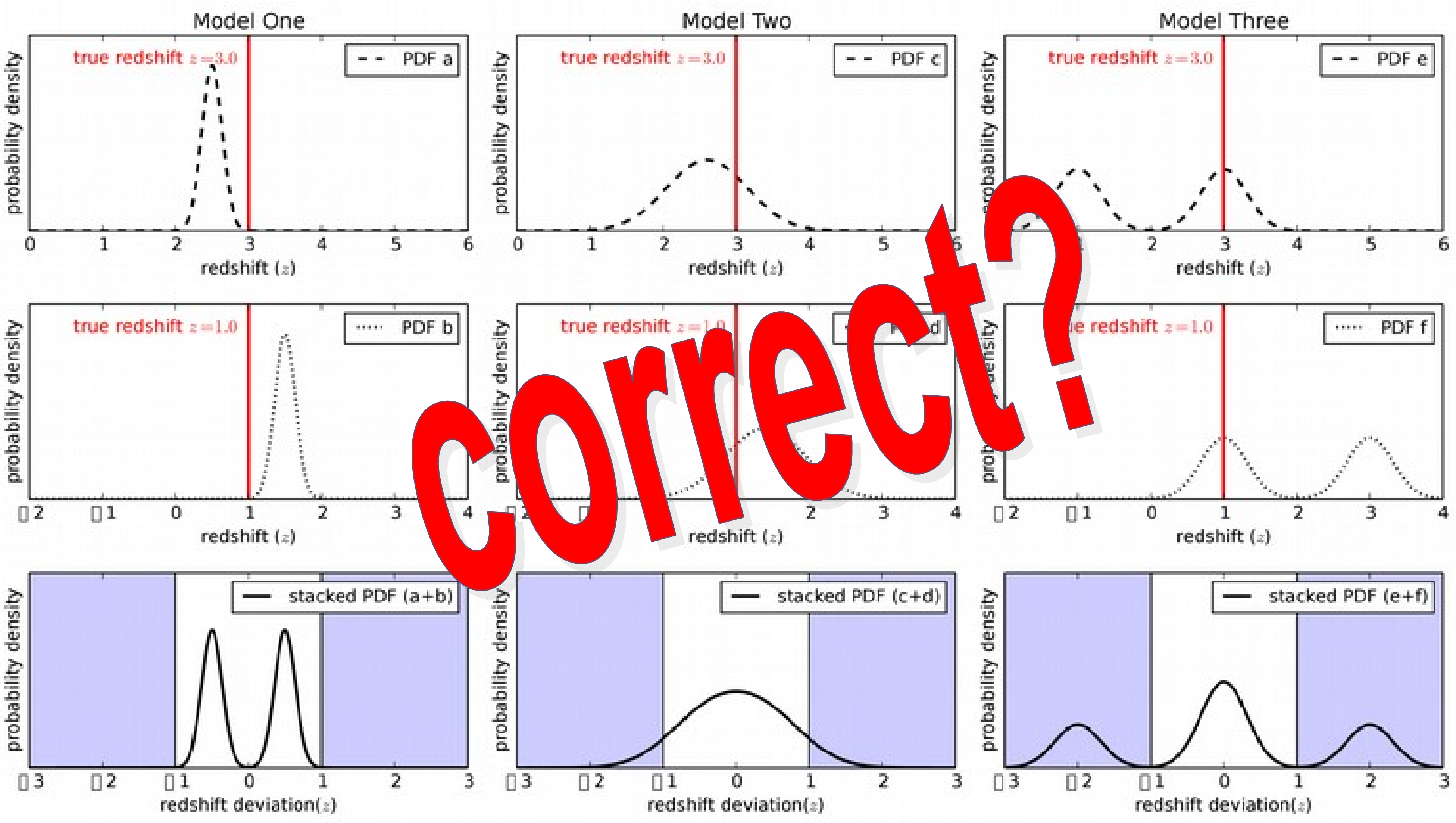
Evaluation Tools

Simplification → RMSE / just use the mean



Evaluation Tools

stacking PDFs

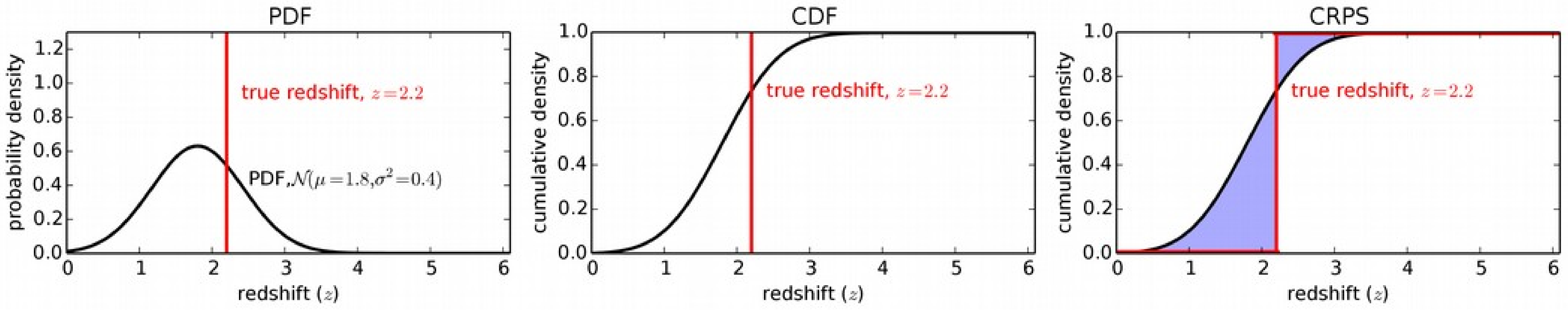


Proper Evaluation Tools / CRPS

continuous rank probability score

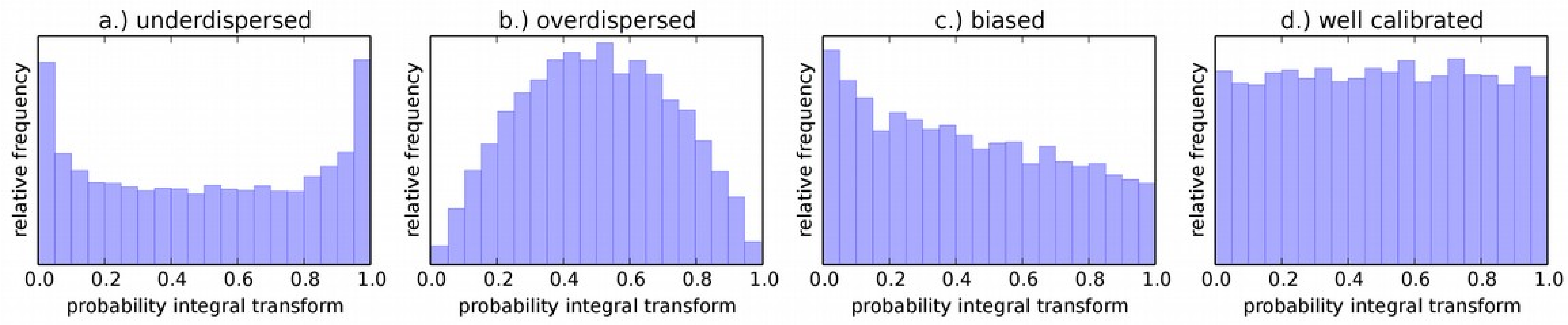
$$CRPS = \frac{1}{N}\sum_{t=1}^{N} crps(CDF_t, z_t),$$

$$\text{with } crps(CDF_t, z_t) = \int_{-\infty}^{+\infty} [CDF_t(z) - CDF_{z_t}(z)]^2 \, dz$$

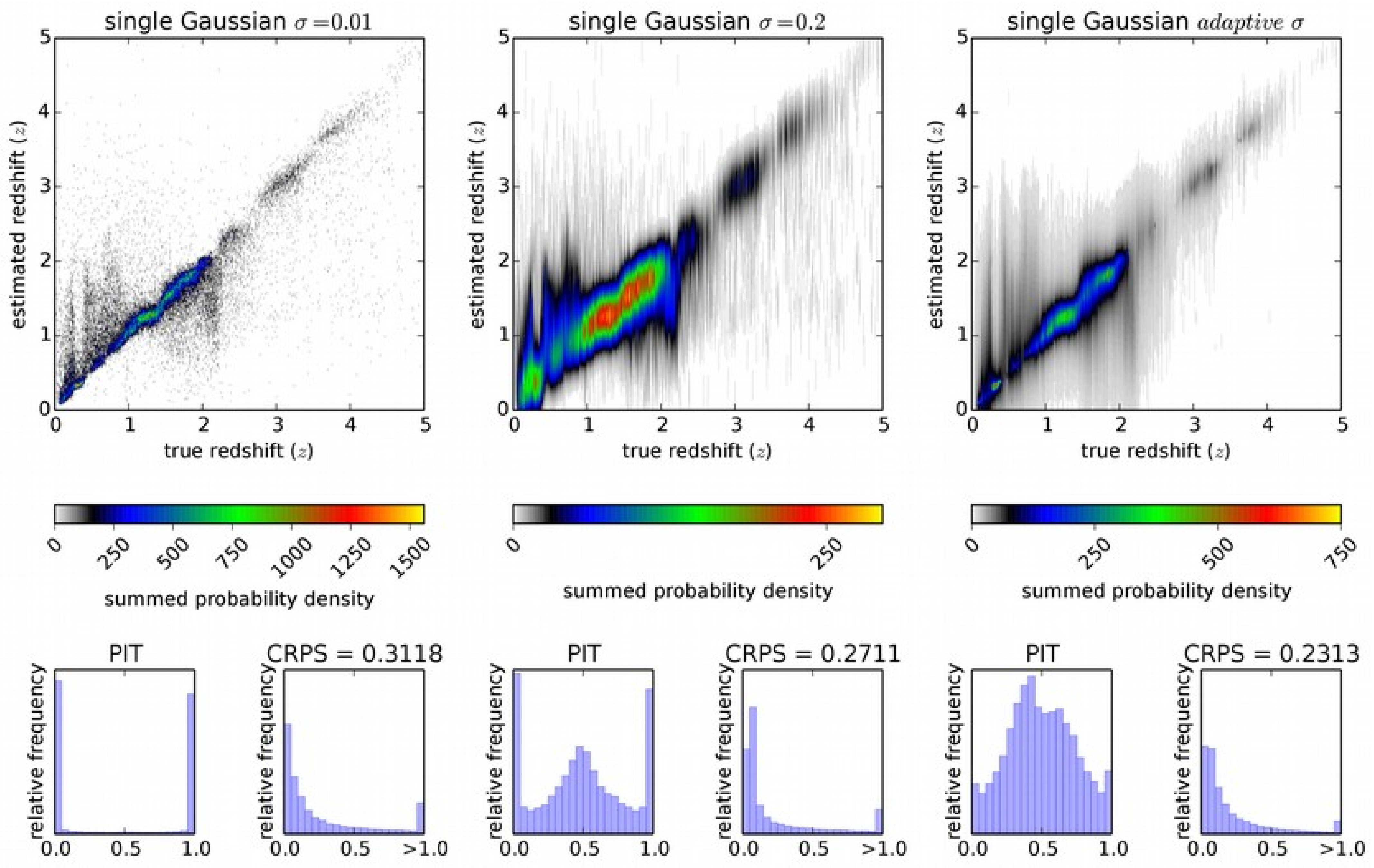


Proper Evaluation Tools / PIT

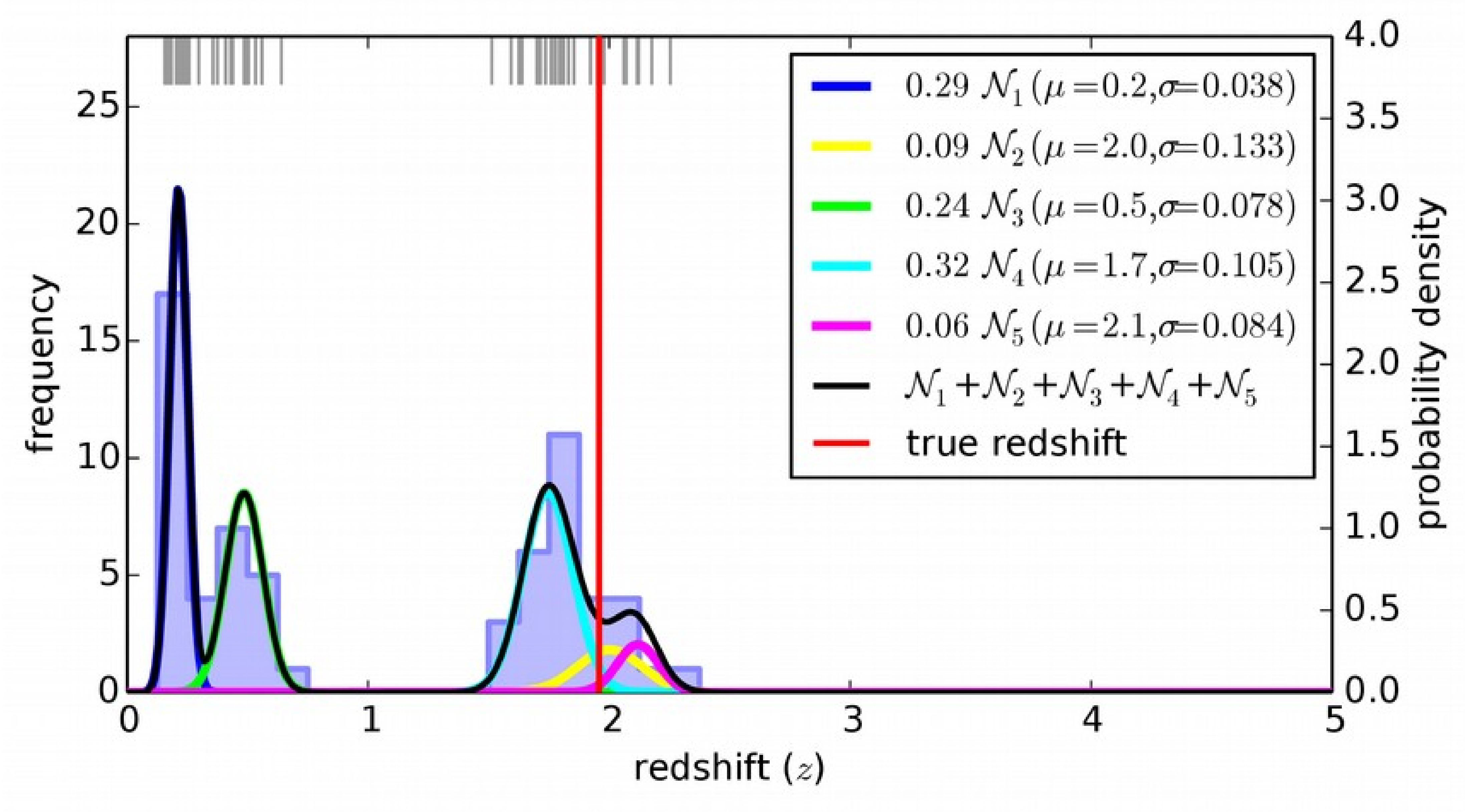
probability integral transform



Uncertain Results / kNN



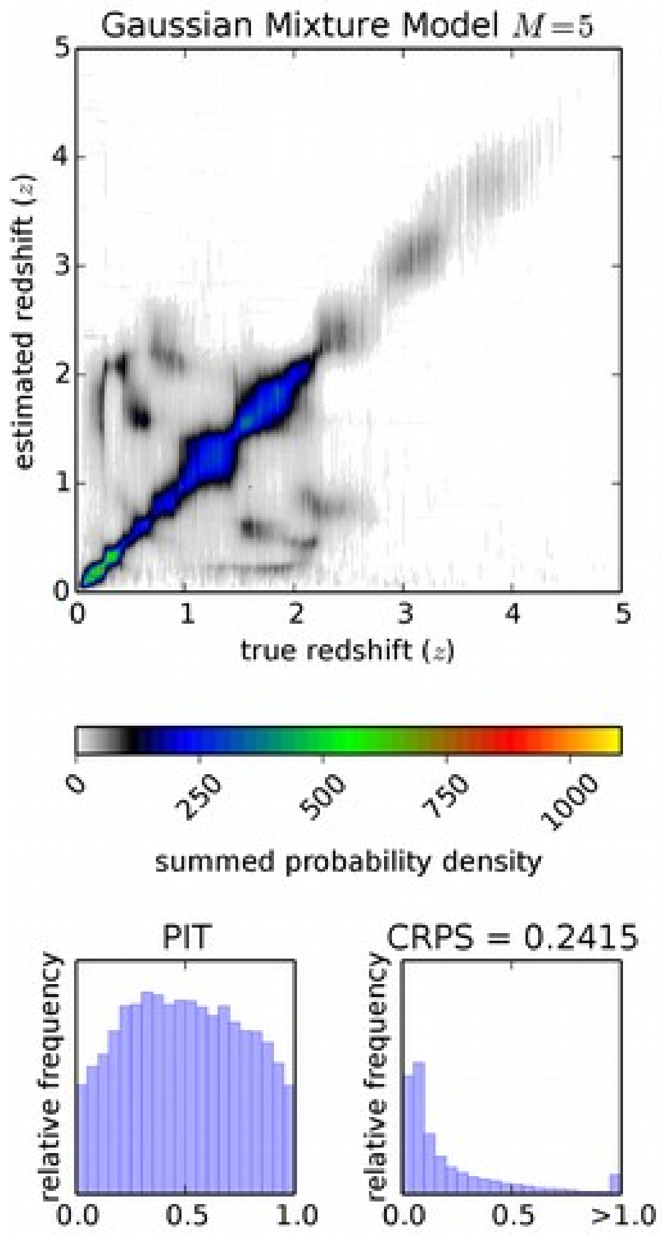
Multi-Modalities



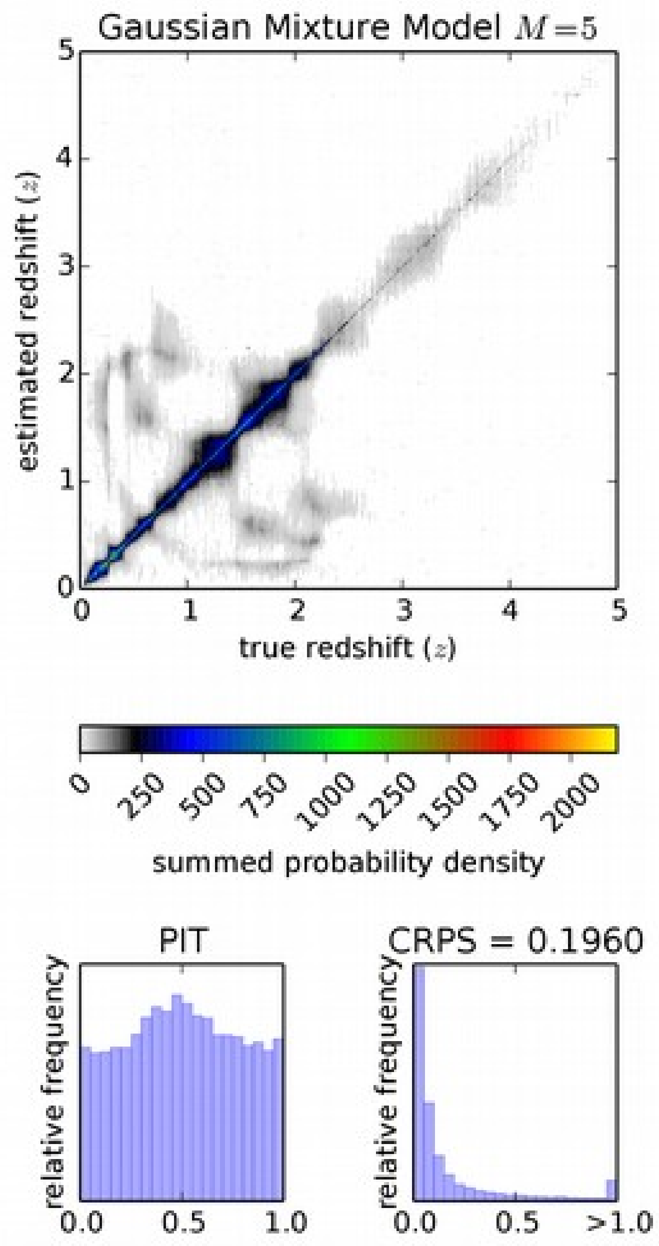
Results

HITS

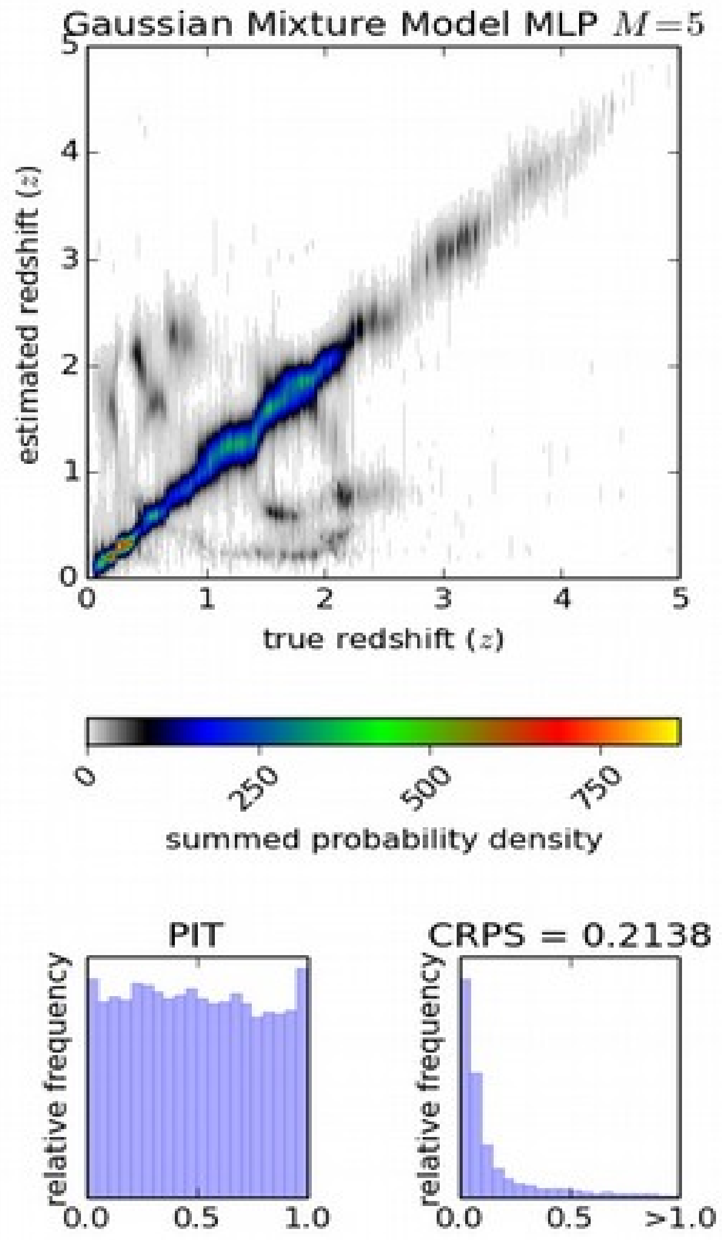
Nearest Neighbors



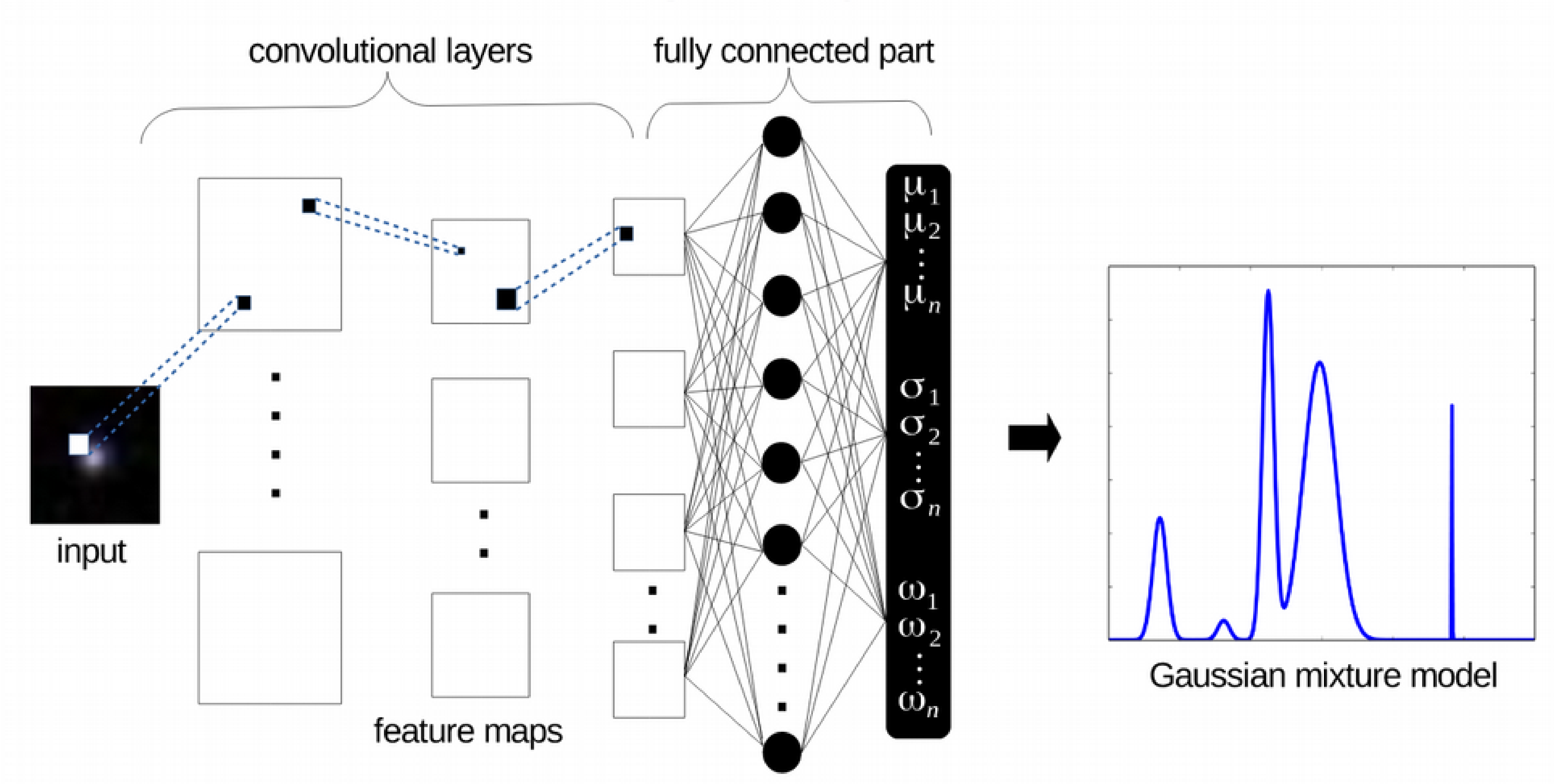
Random Forest



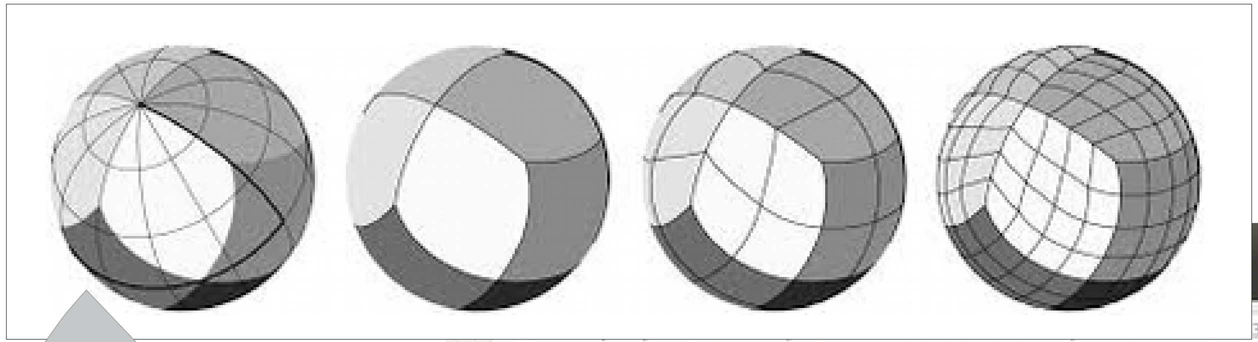
Mixture Density Network

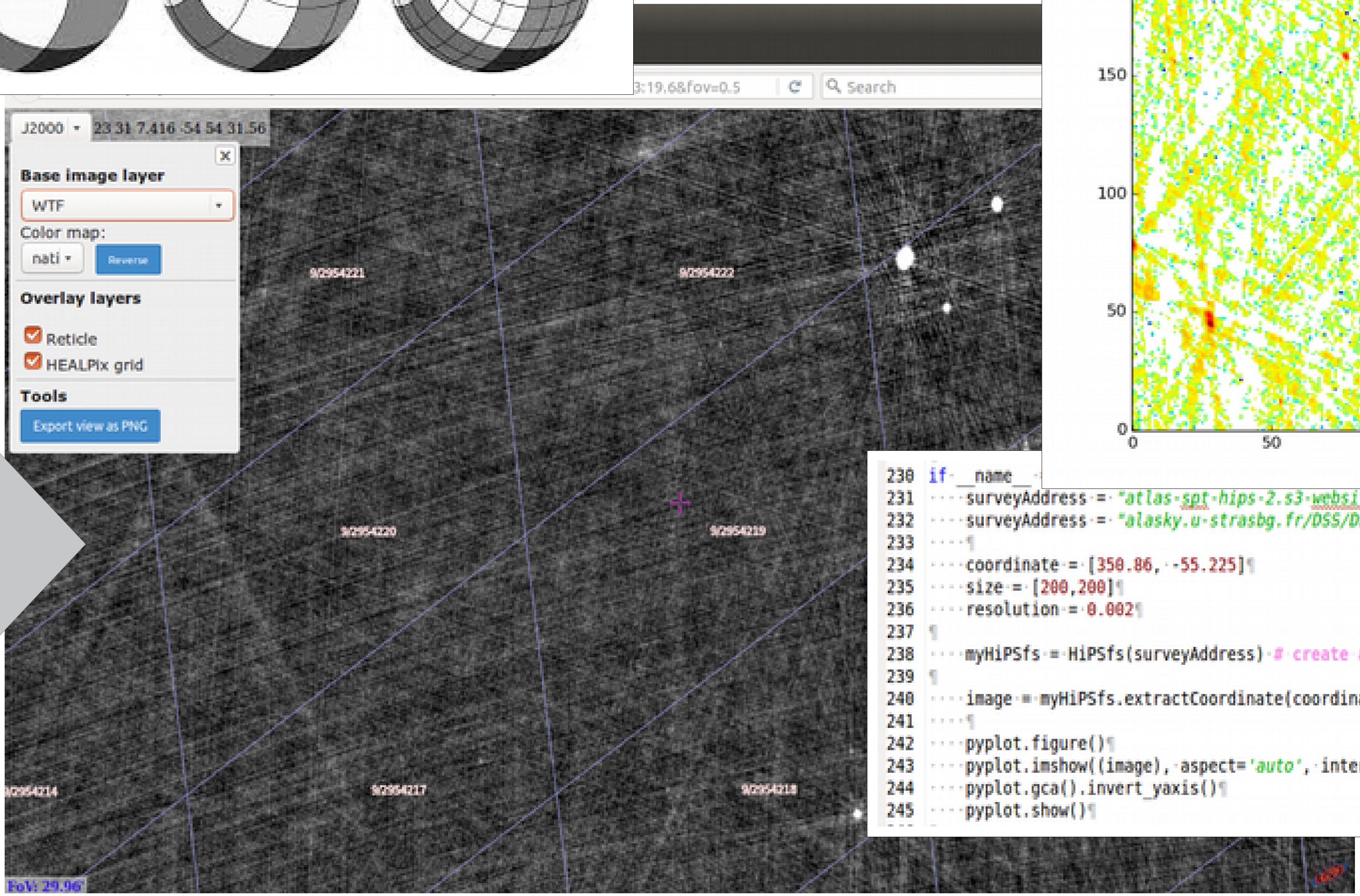


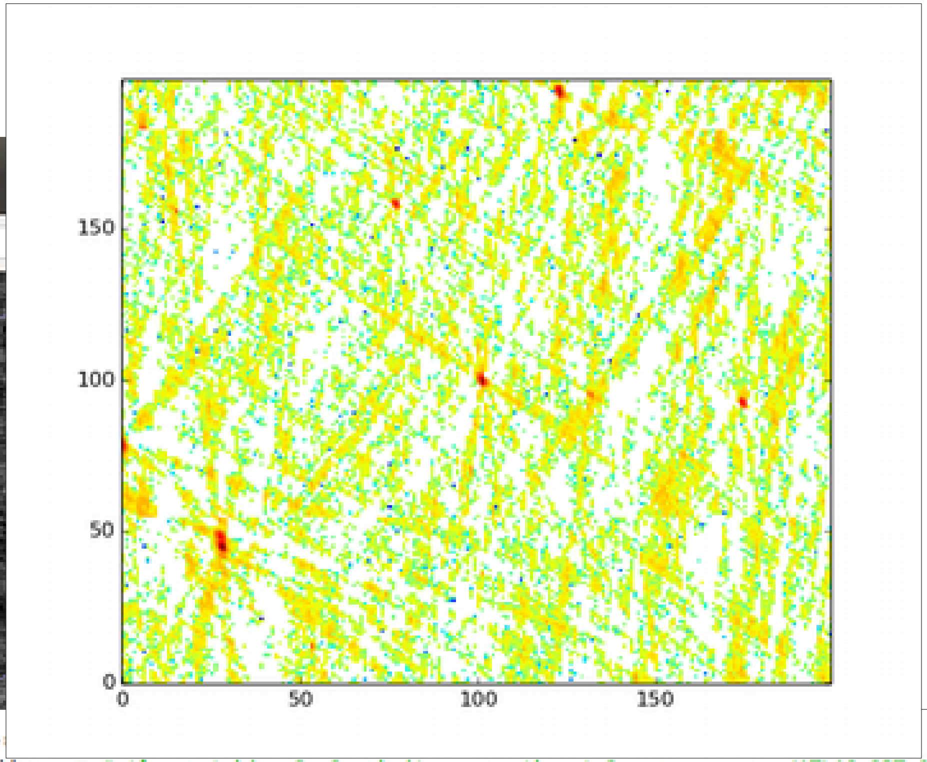
DCN meets MDN



Healpix / HiPS / IVOA

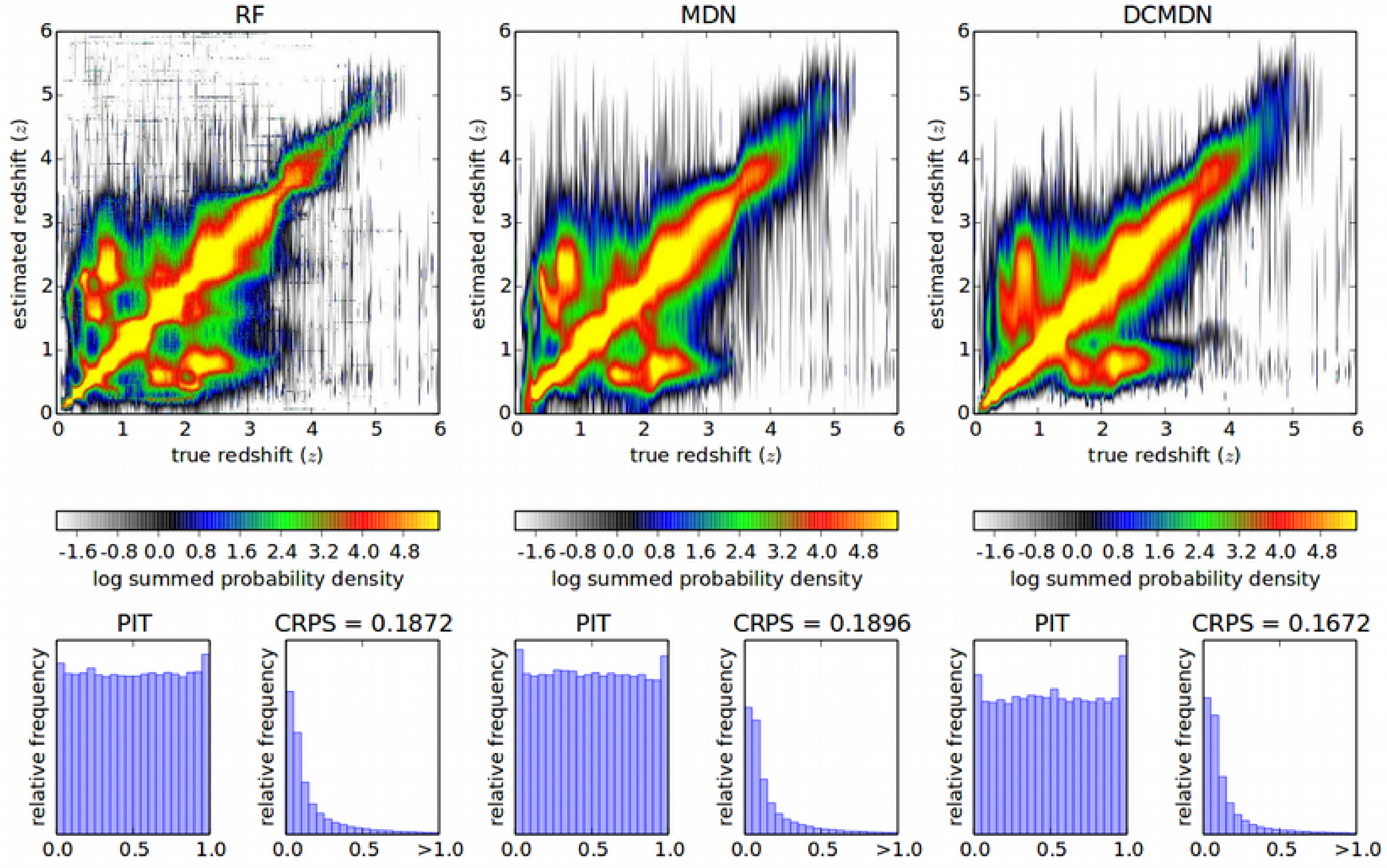






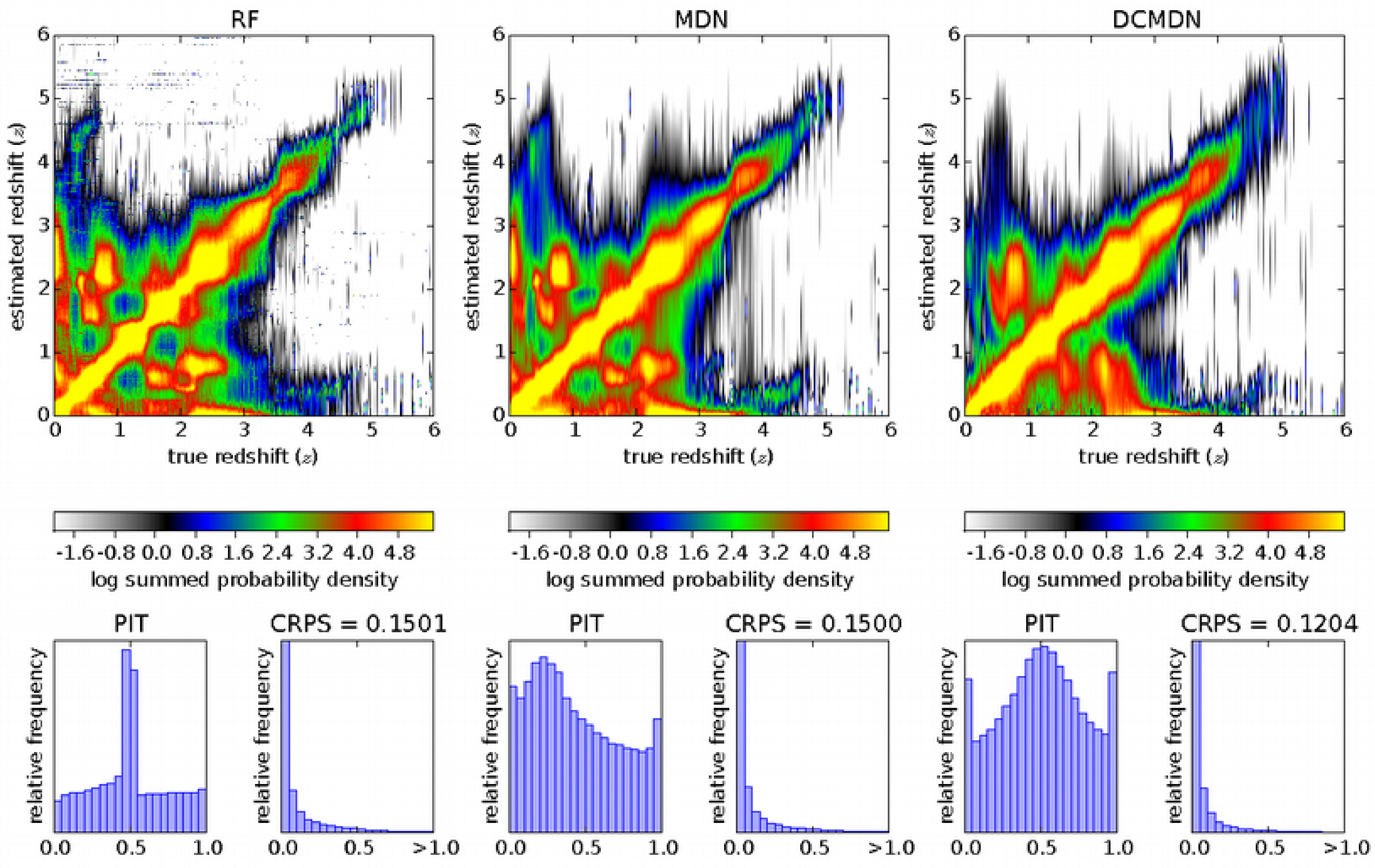

```python
if __name__
    surveyAddress = "atlas-spt-hips-2.s3-website-ap-southeast-2.amazonaws.com/ATLAS-SPT-64x64"
    surveyAddress = "alasky.u-strasbg.fr/DSS/DSS2Merged" # DSS red

    coordinate = [350.86, -55.225]
    size = [200,200]
    resolution = 0.002

    myHiPSfs = HiPSfs(surveyAddress) # create access

    image = myHiPSfs.extractCoordinate(coordinate, size, resolution, nested=True) # extract data array

    pyplot.figure()
    pyplot.imshow((image), aspect='auto', interpolation="nearest")
    pyplot.gca().invert_yaxis()
    pyplot.show()
```

Results



D'Isanto 2018

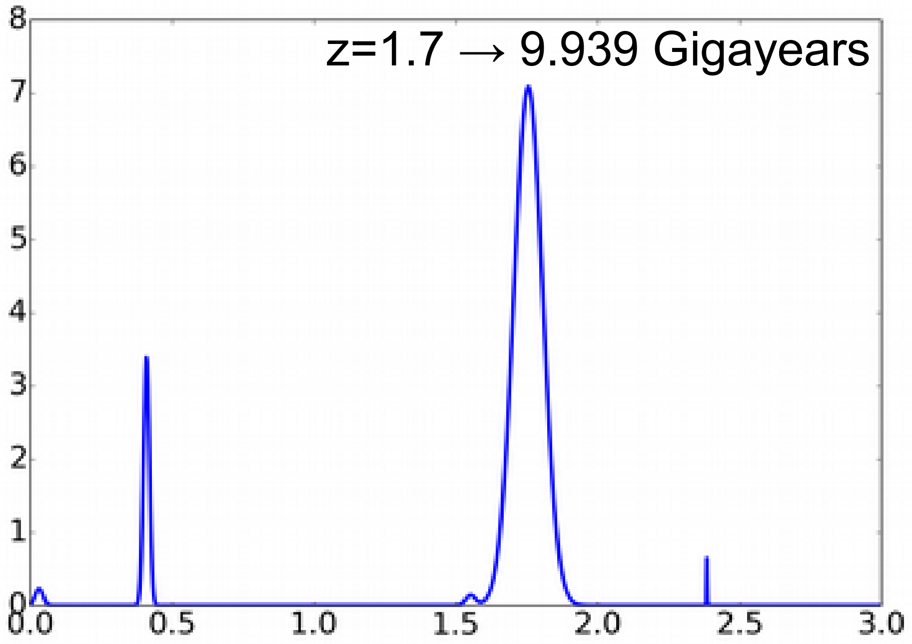
Complex Input Catalog



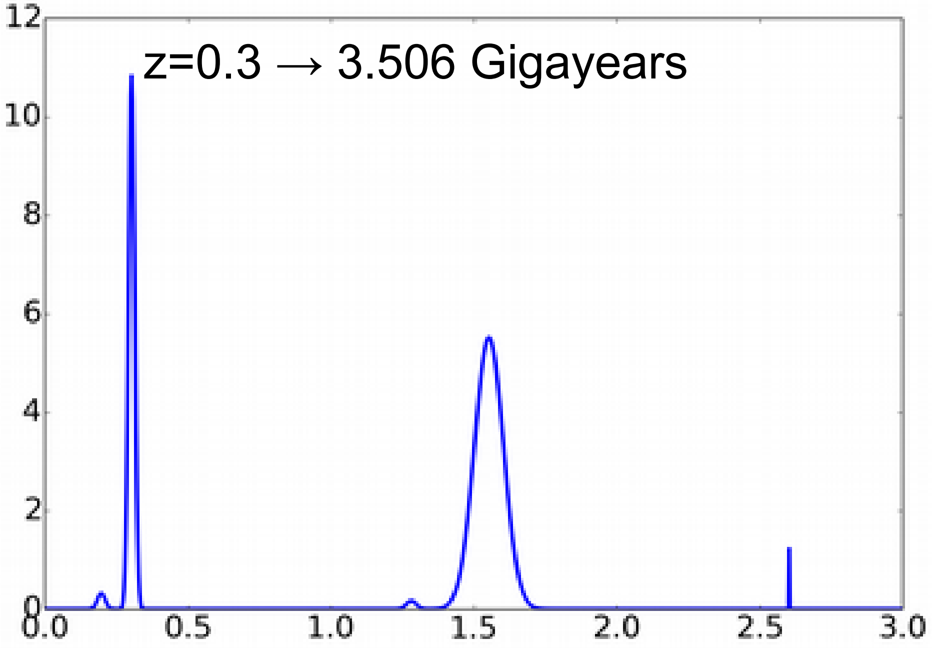
D'Isanto 2018

Challenges / Limitations



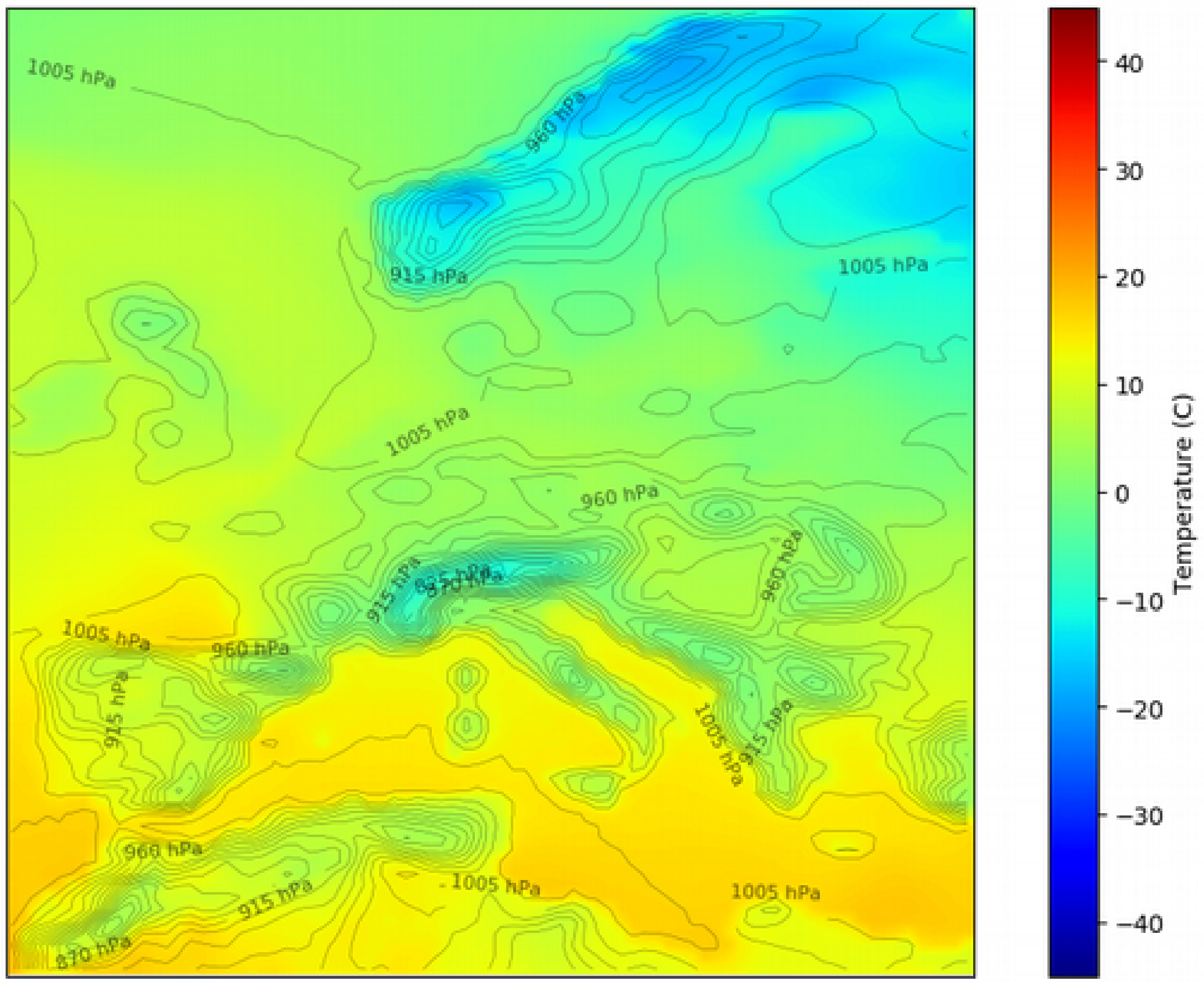


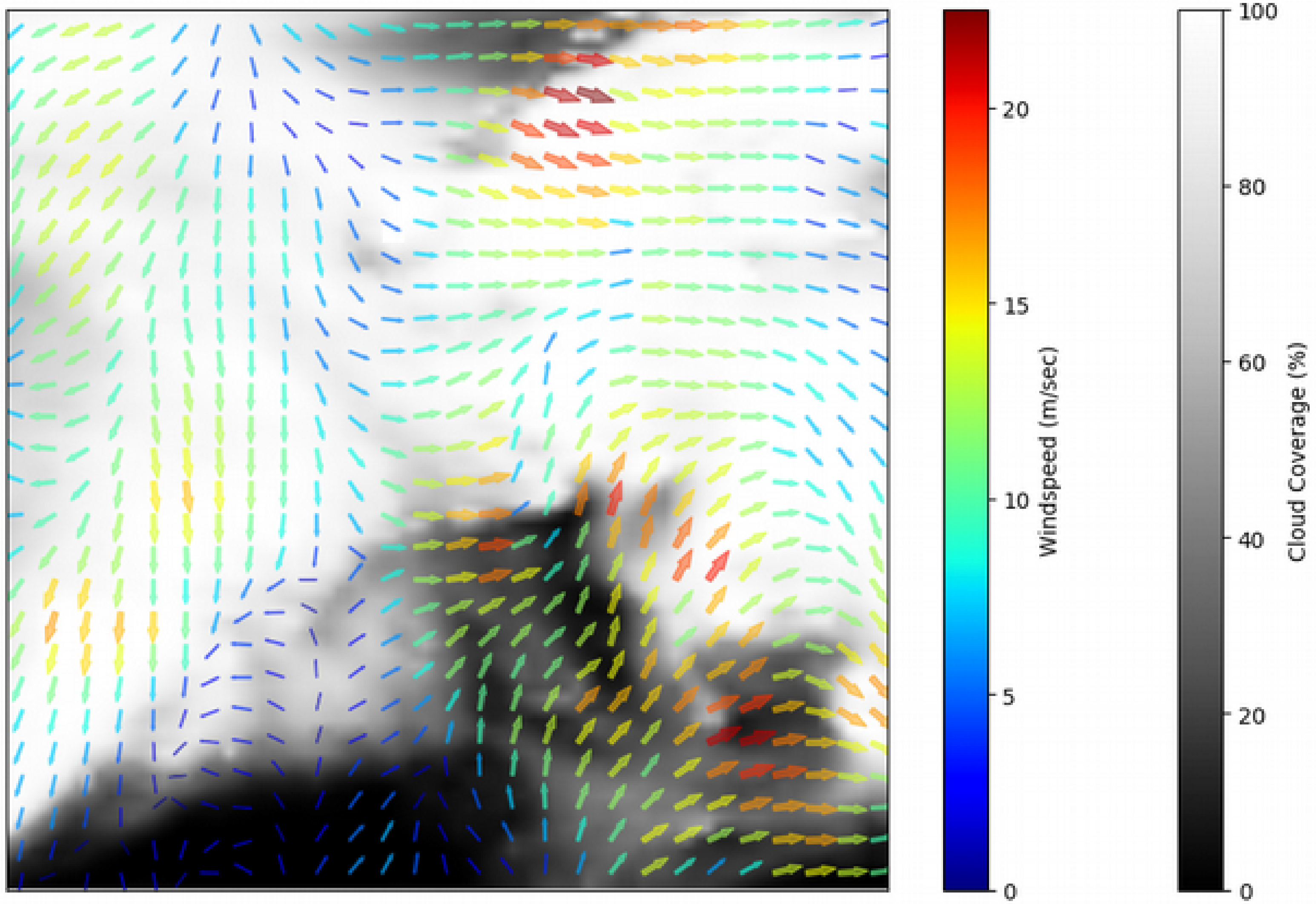




Weather Forecast Simulations

European Centre for Medium-Range Weather Forecasts (ECMWF)

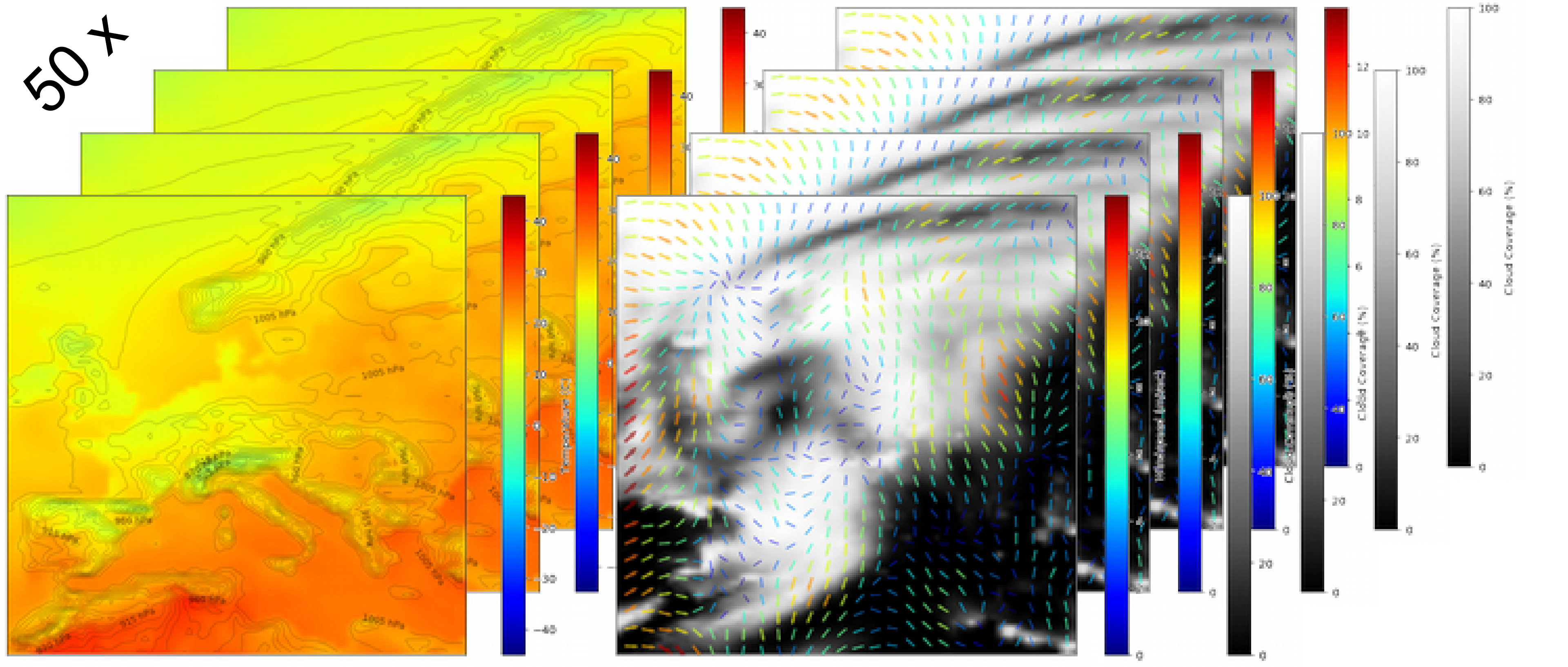




Weather Forecast Simulations

In this example 18 parameters on 81x81 grid for Europe (0.5°, 0.5°)

- t2m: air temperature 2m above ground
- cape: convective available potential energy
- sp: surface pressure
- tcc: total cloud cover
- sshf: sensible heat flux
- slhf: latent heat flux
- u10: 10-meter U-wind
- v10: 10-meter V-wind
- d2m: 2-meter dew point temperature
- ssr: short wave radiation flux
- str: long wave radiation flux
- sm: soil moisture
- u pl500: u-wind at 500 hPa
- v pl500: v-wind at 500 hPa
- u pl850: u-wind at 850 hPa
- v pl850: v-wind at 850 hPa
- gh pl500: Geopotential at 500 hPa
- q pl850: specific humidity at 850 hPa

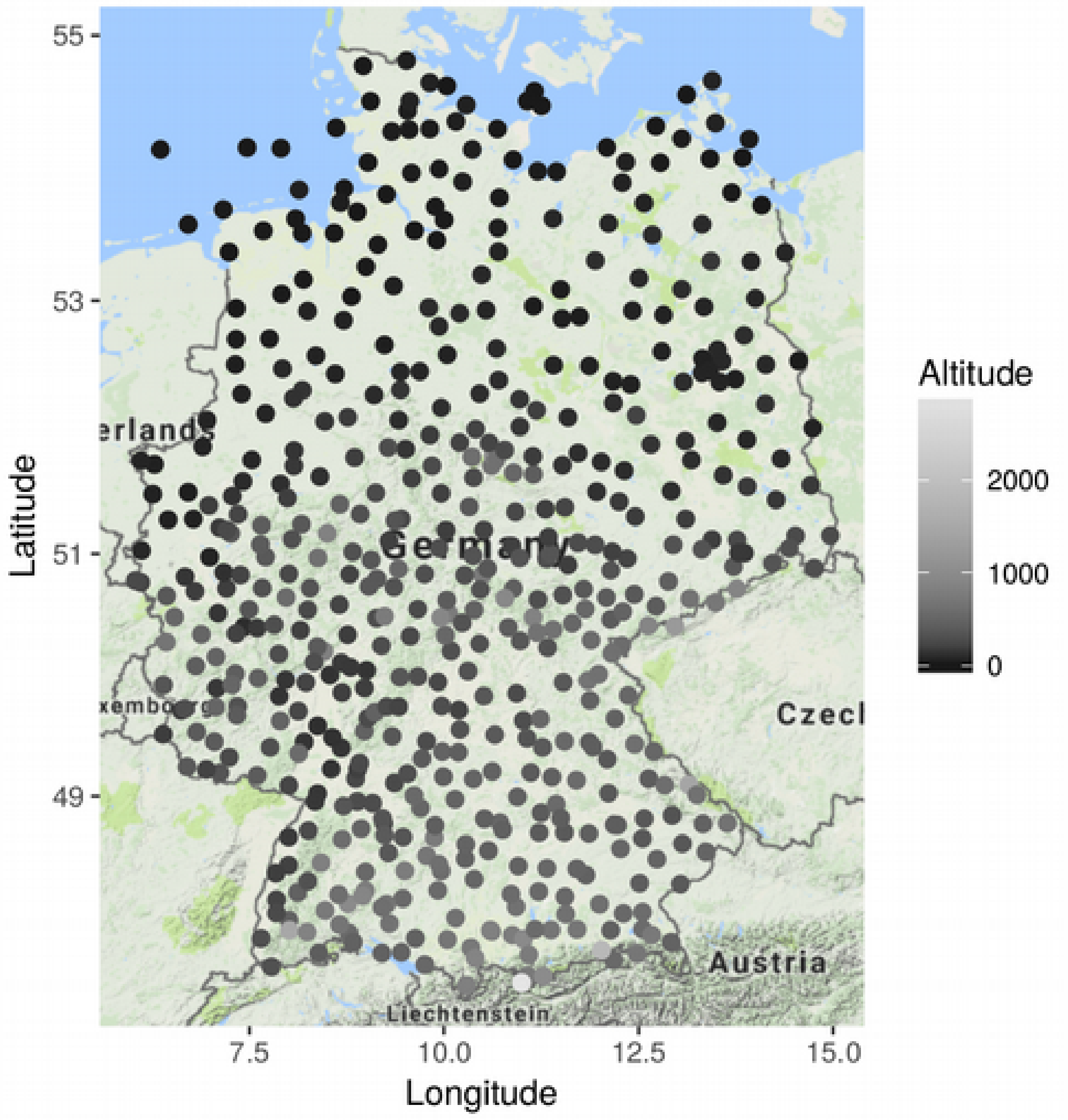


Statistical Post-Processing

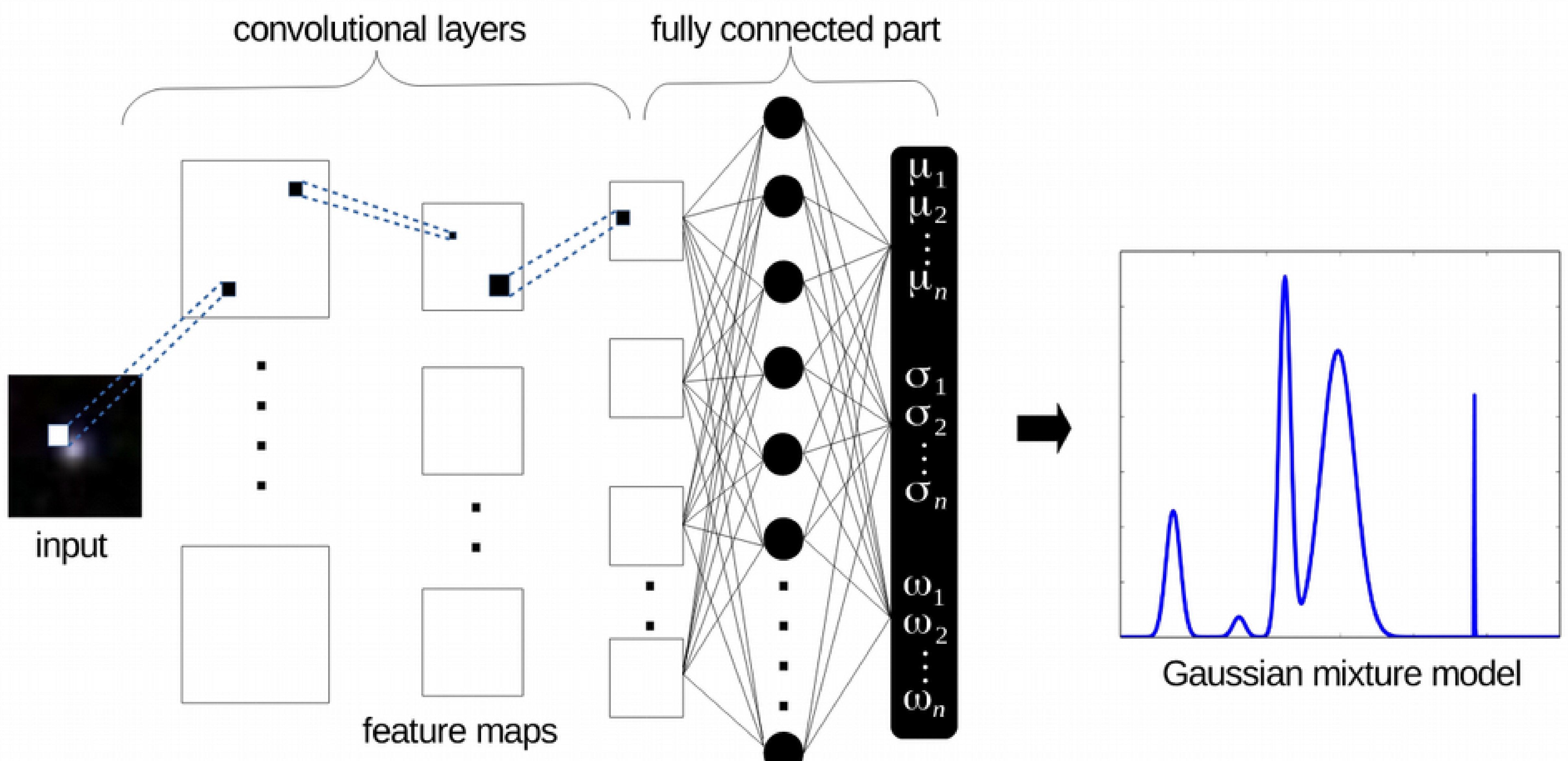
to predict for single stations

- 537 stations with measurements
  - (lat, long, alt, orog., land/sea)
- 48h forecast lead time
- 18 x 2 parameters (mean, stddev)
- 2007-2015 for training
- 2016 for testing

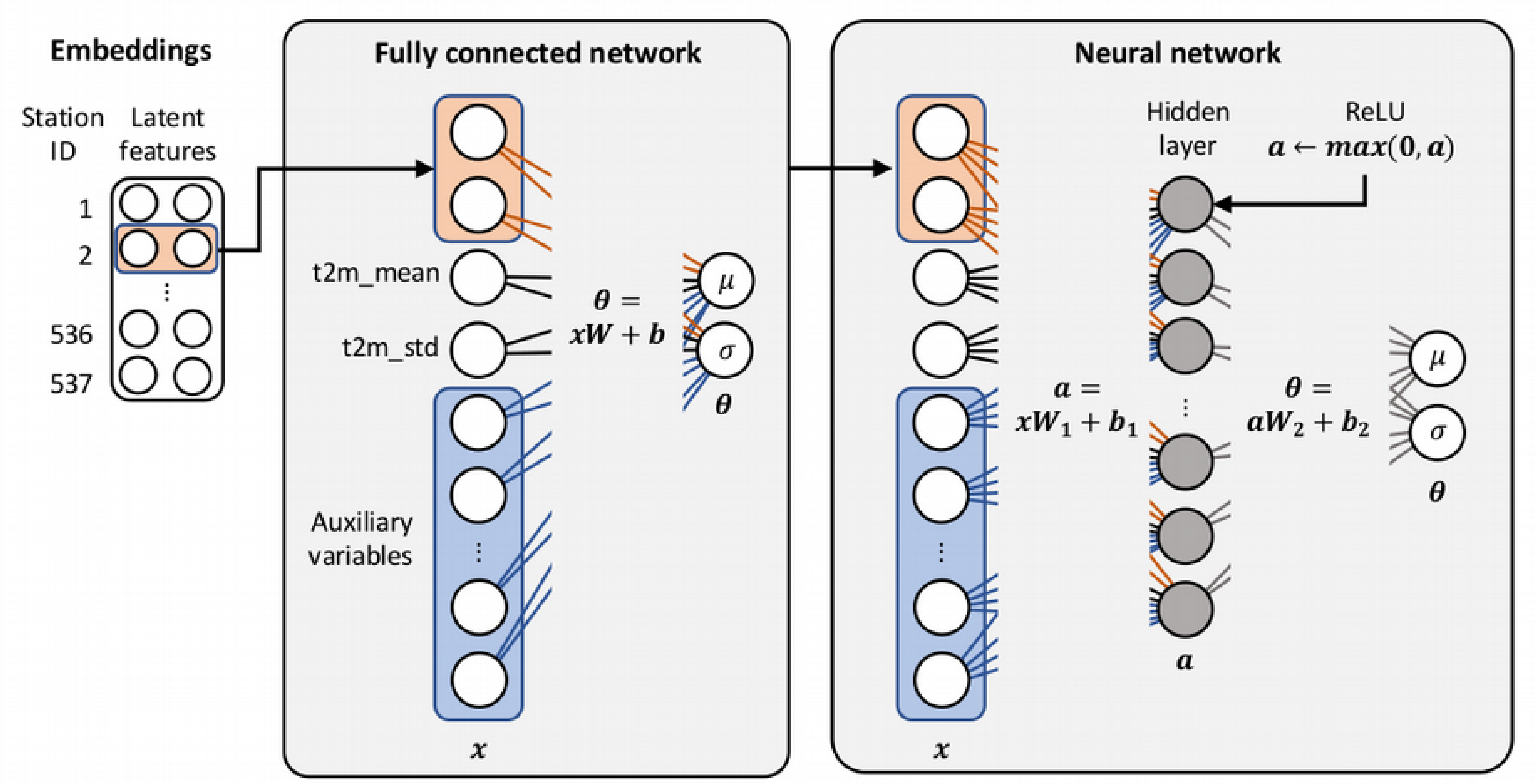
with best method CRPS=0.81



DCN meets MDN



Using ~~DC~~MDN

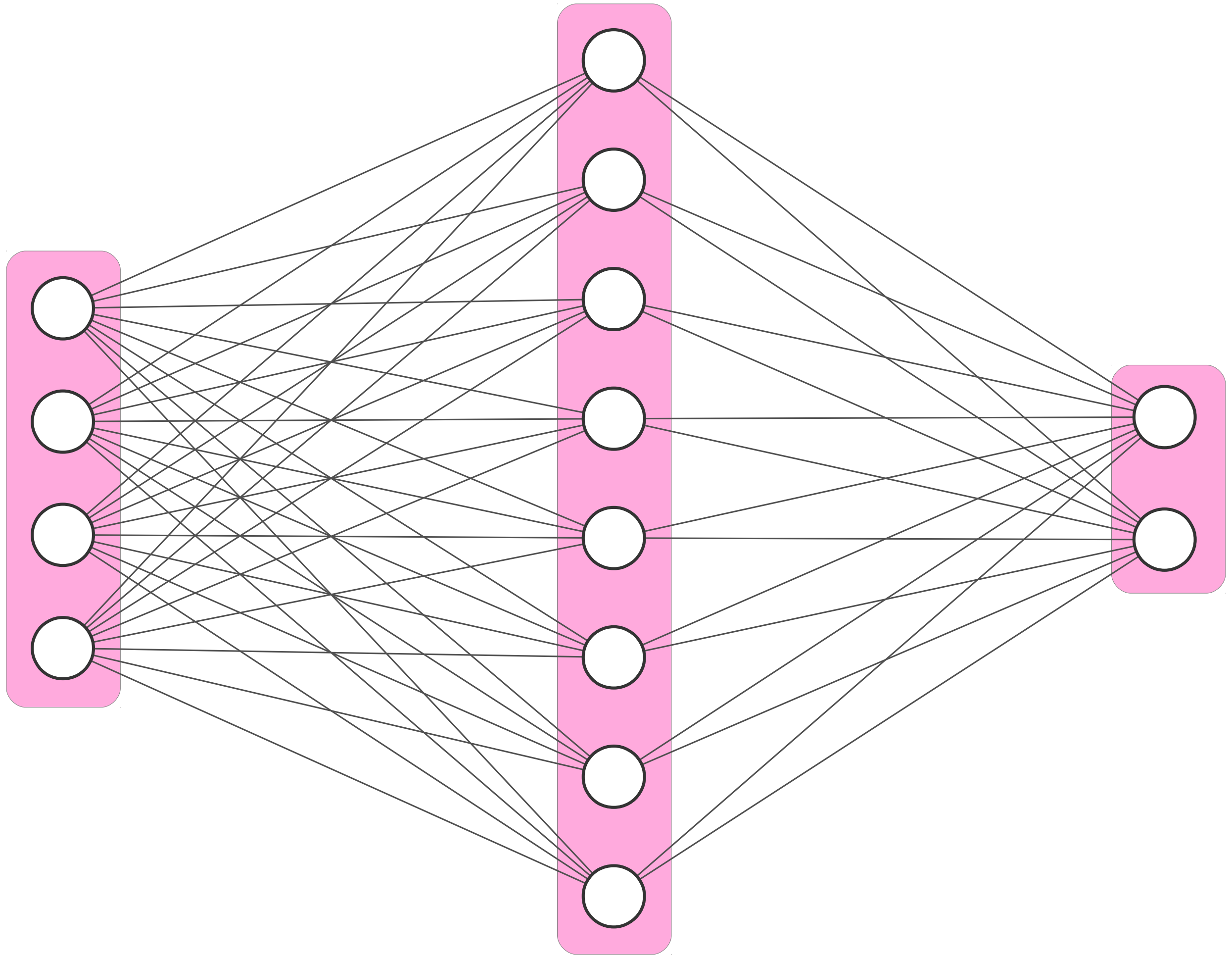


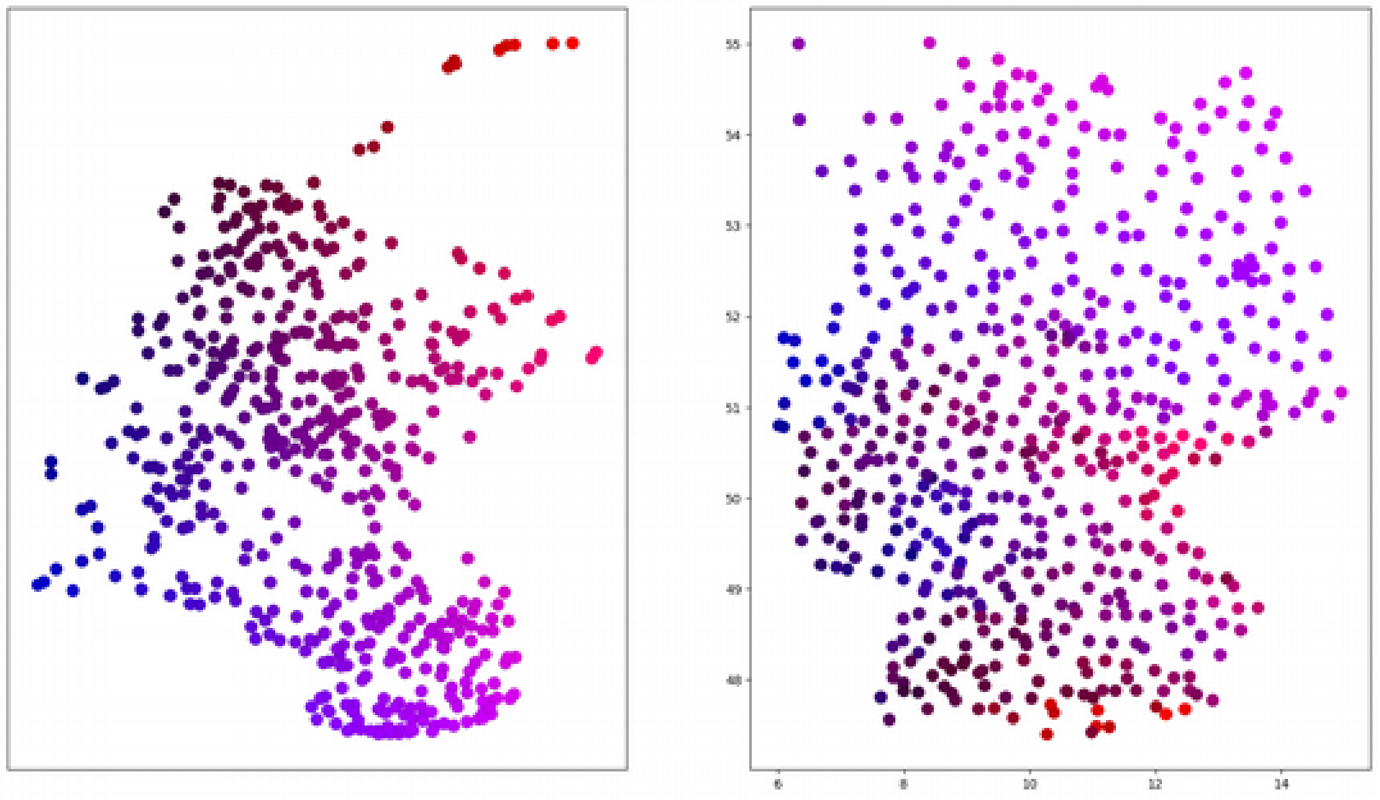
Rasp and Lerch 2018

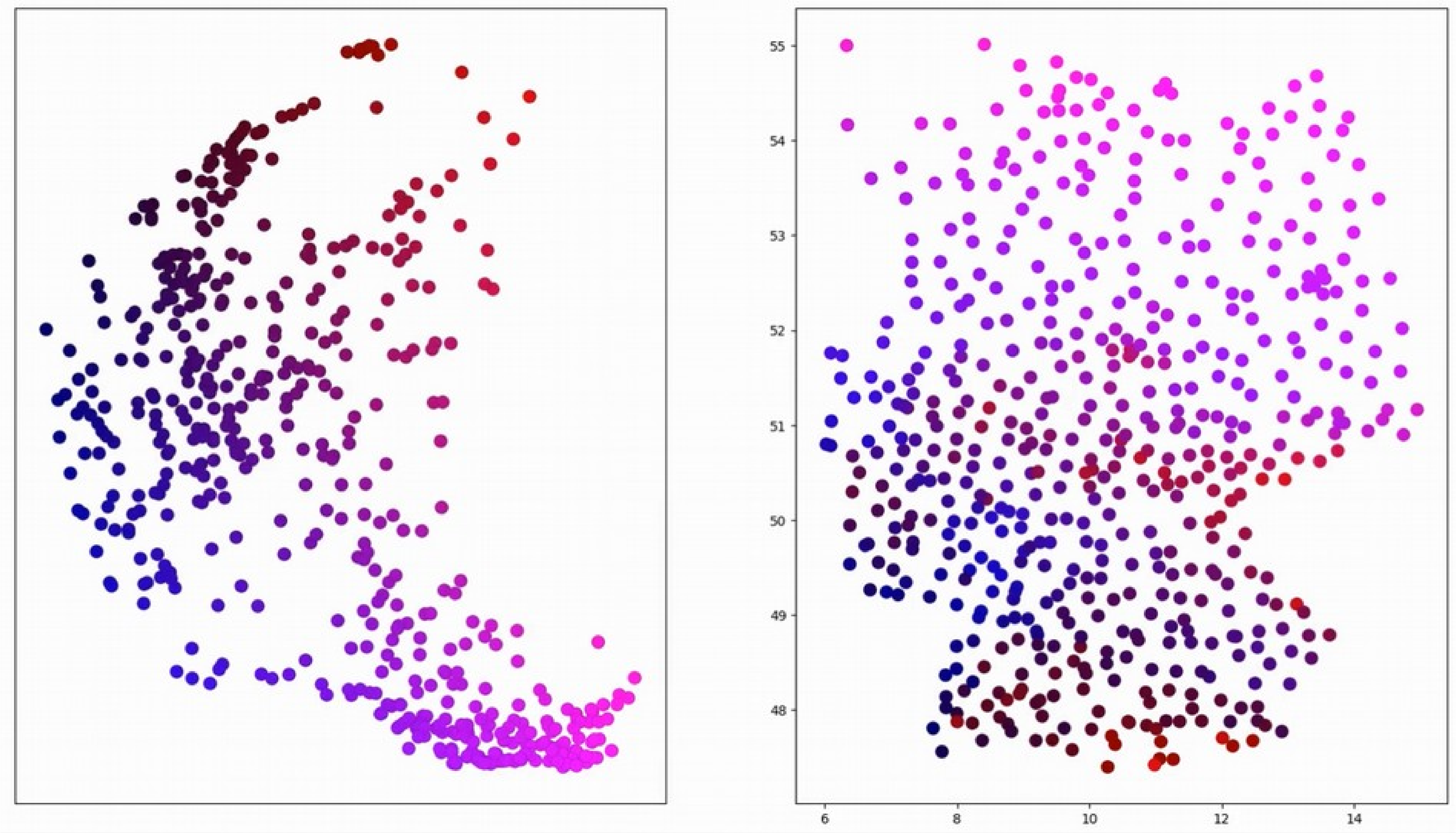
CRPS=0.78

Projecting Station Parameters

latitude, longitude, altitude, orography







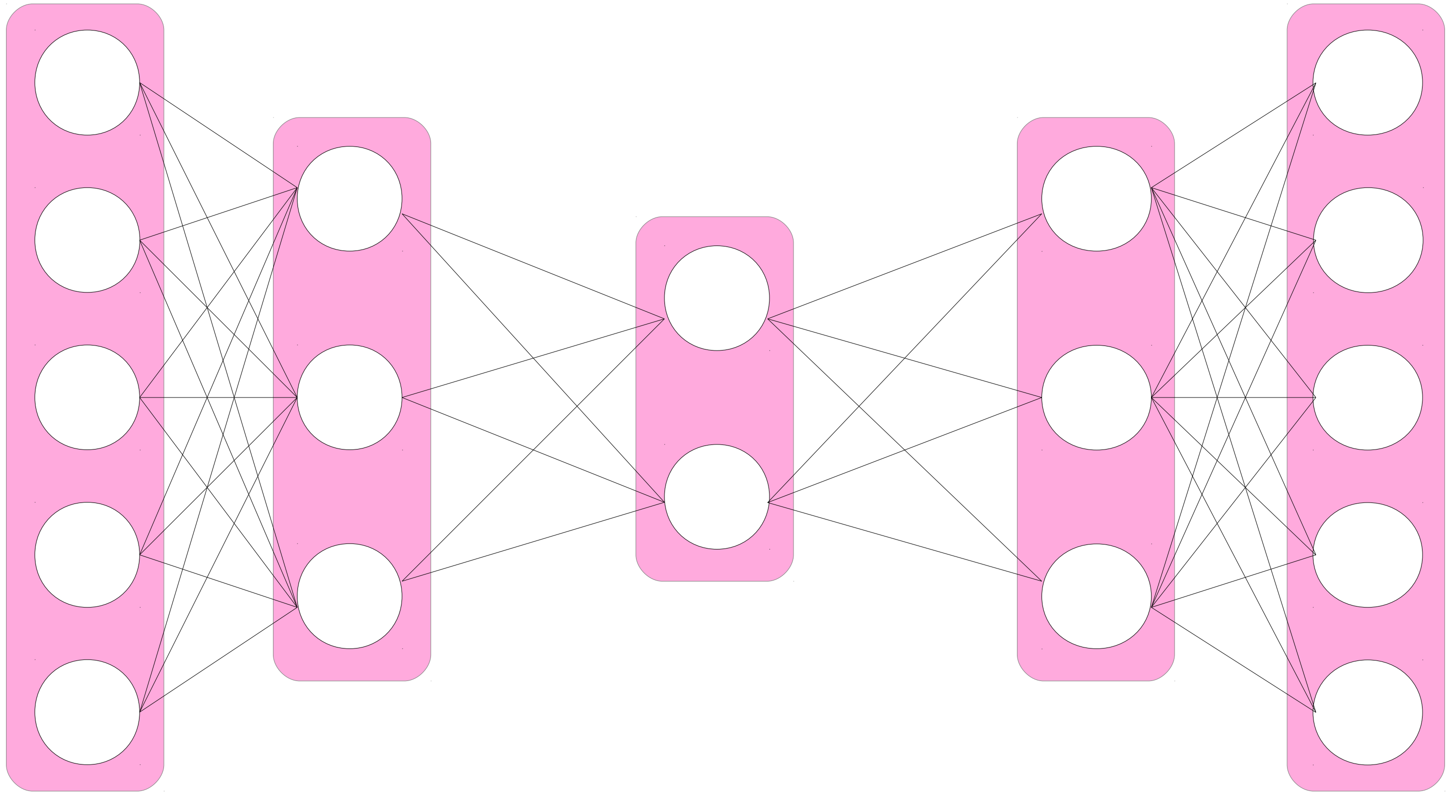
55
54
53
52
51
50
49
48
6
8
10
12
14

More Complex Network

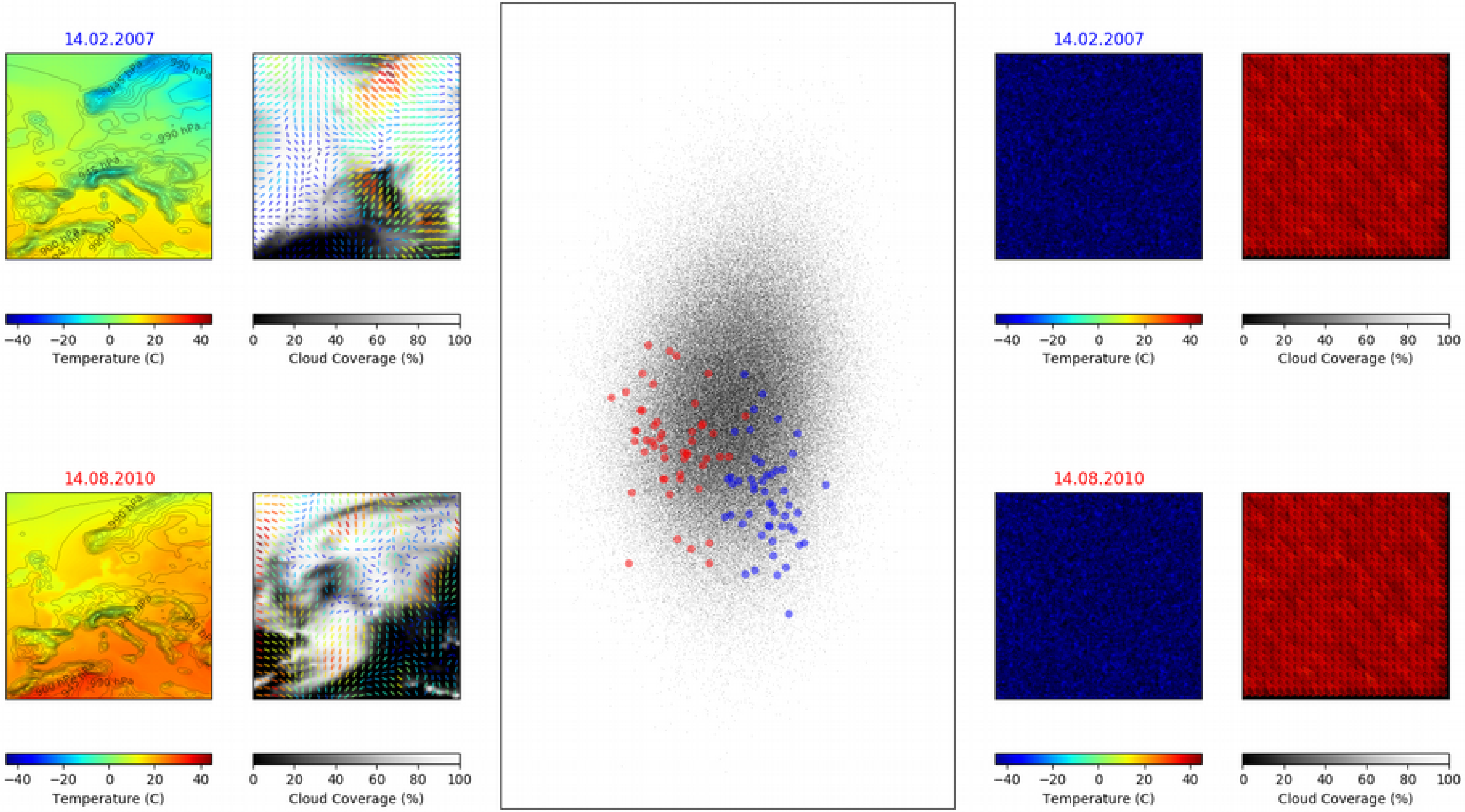
DCMDN → whole ensemble 50*81*81*17 * 535 * 3667

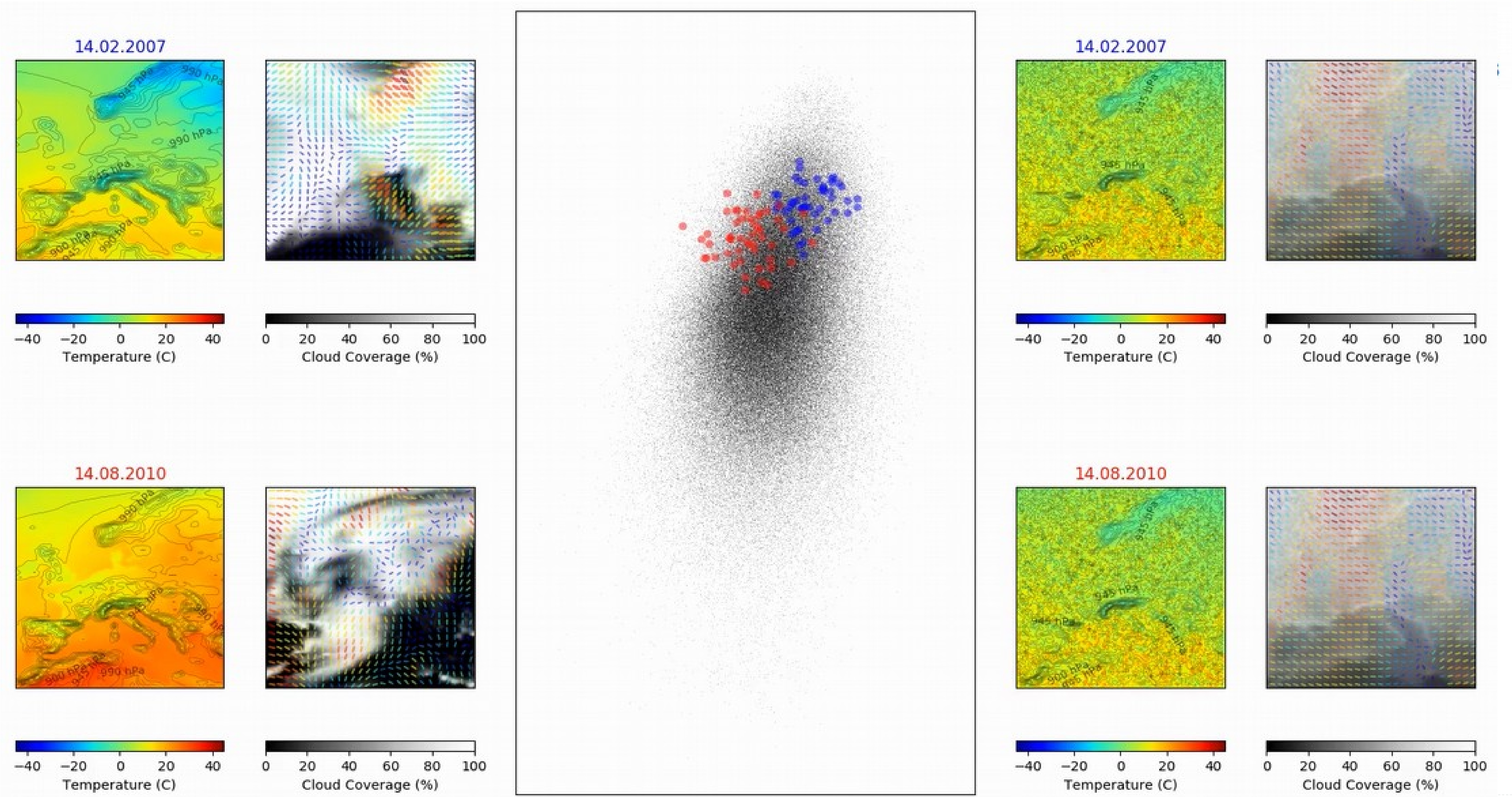
- not enough data for training

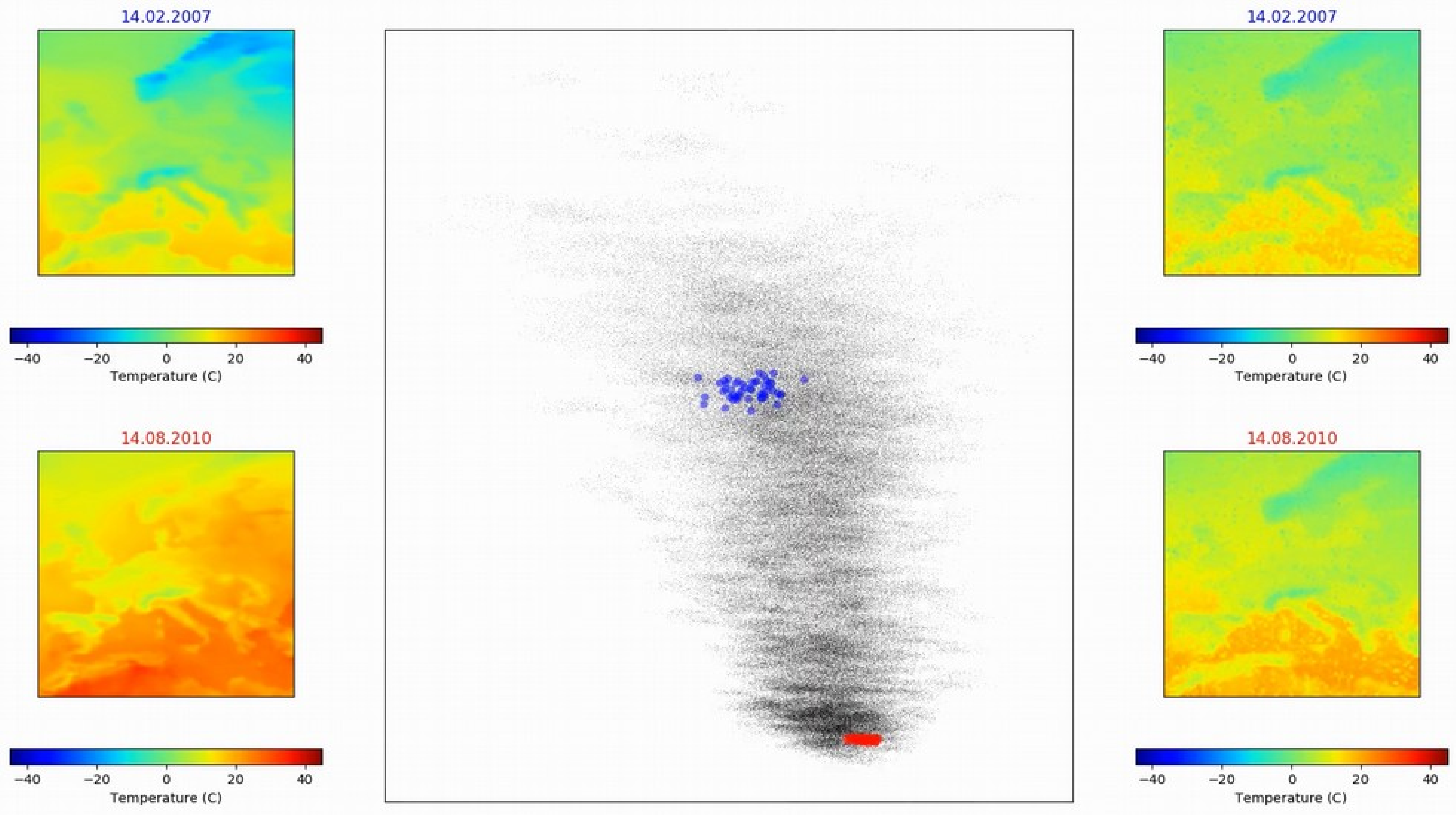
Use different strategy with autoencoders

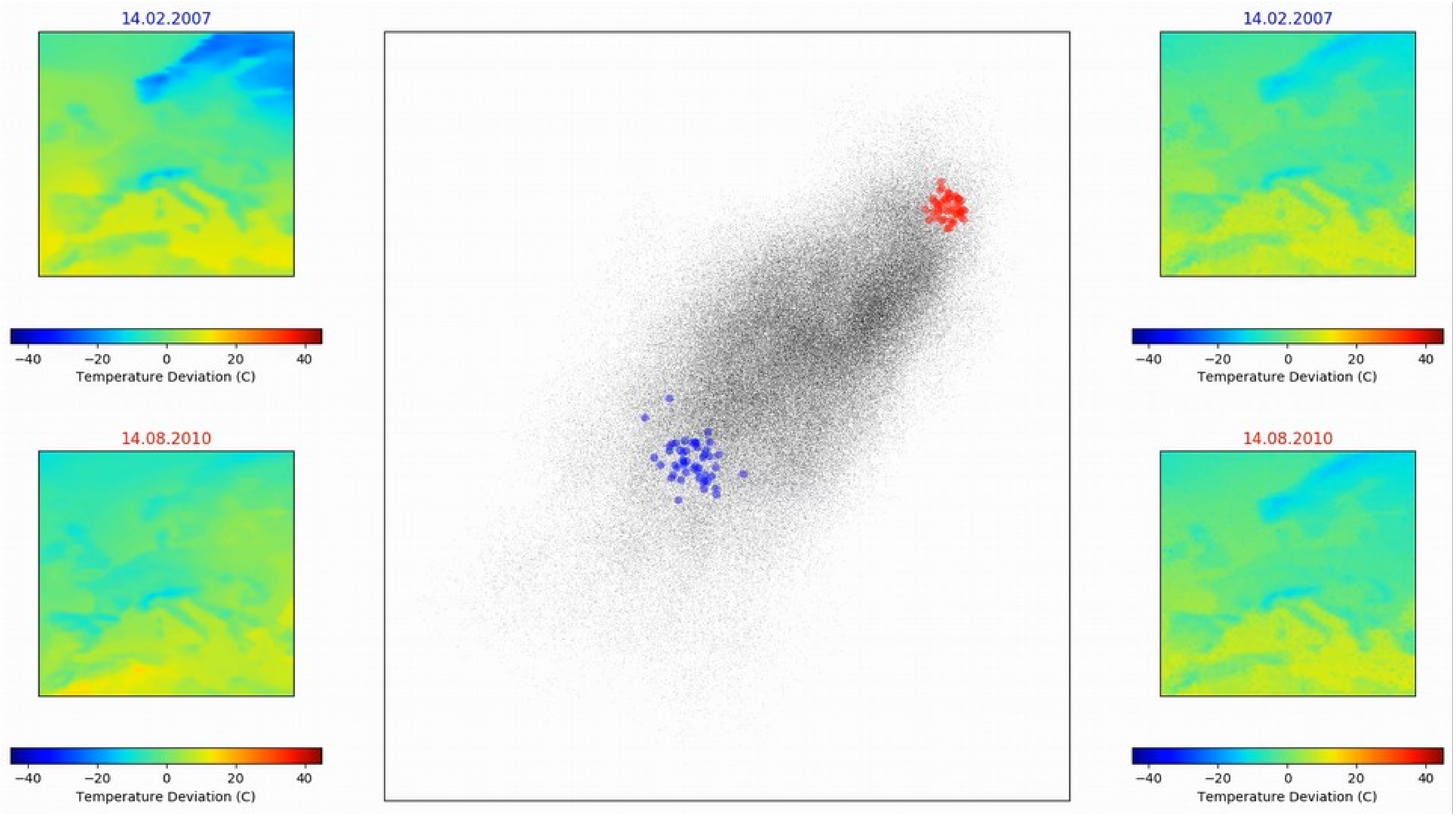


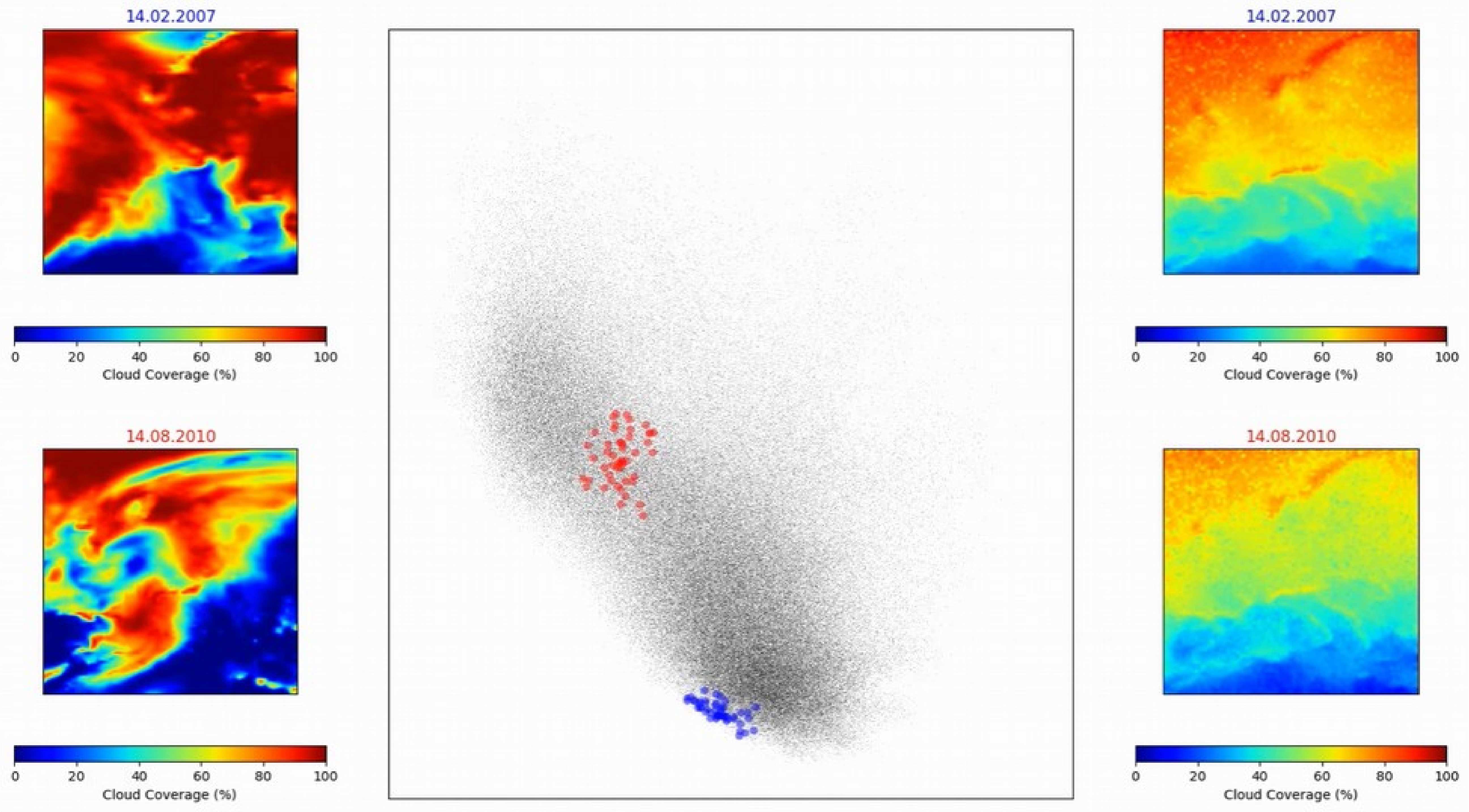
Projecting Forecast Ensembles



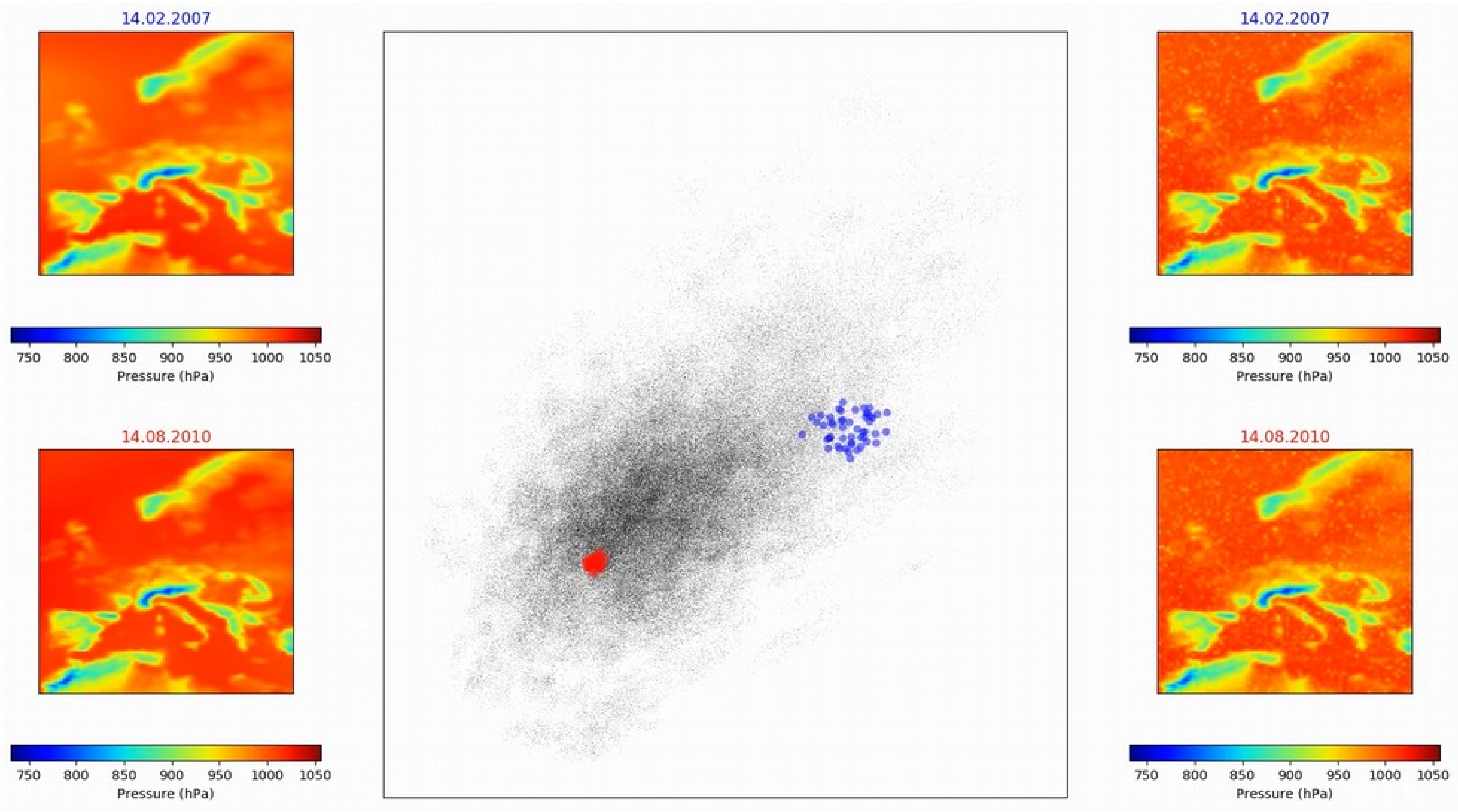

14.02.2007
Temperature (C)
Cloud Coverage (%)
14.08.2010
Temperature (C)
Cloud Coverage (%)
14.02.2007
Temperature (C)
Cloud Coverage (%)
14.08.2010
Temperature (C)
Cloud Coverage (%)

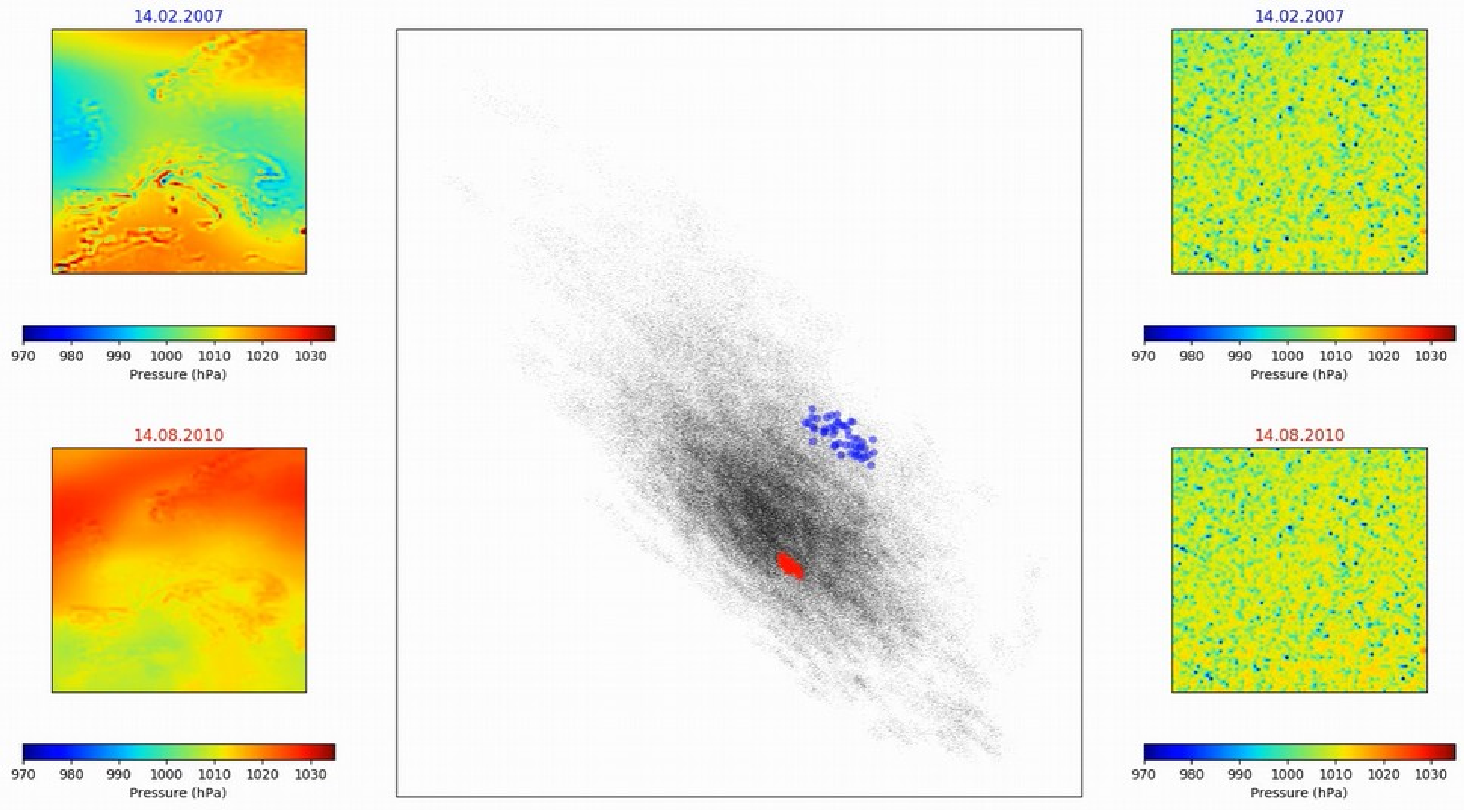

14.02.2007
−40 −20 0 20 40
Temperature (C)
14.08.2010
−40 −20 0 20 40
Temperature (C)
14.02.2007
−40 −20 0 20 40
Temperature (C)
14.08.2010
−40 −20 0 20 40
Temperature (C)


14.02.2007
−40
−20
0
20
40
Temperature Deviation (C)
14.08.2010
−40
−20
0
20
40
Temperature Deviation (C)
14.02.2007
−40
−20
0
20
40
Temperature Deviation (C)
14.08.2010
−40
−20
0
20
40
Temperature Deviation (C)



14.02.2007
0 20 40 60 80 100
Cloud Coverage (%)
14.08.2010
0 20 40 60 80 100
Cloud Coverage (%)
14.02.2007
0 20 40 60 80 100
Cloud Coverage (%)
14.08.2010
0 20 40 60 80 100
Cloud Coverage (%)


14.02.2007
750 800 850 900 950 1000 1050
Pressure (hPa)
14.08.2010
750 800 850 900 950 1000 1050
Pressure (hPa)
14.02.2007
750 800 850 900 950 1000 1050
Pressure (hPa)
14.08.2010
750 800 850 900 950 1000 1050
Pressure (hPa)


14.02.2007
970 980 990 1000 1010 1020 1030
Pressure (hPa)
14.08.2010
970 980 990 1000 1010 1020 1030
Pressure (hPa)
14.02.2007
970 980 990 1000 1010 1020 1030
Pressure (hPa)
14.08.2010
970 980 990 1000 1010 1020 1030
Pressure (hPa)

Projections

Temperature: OK

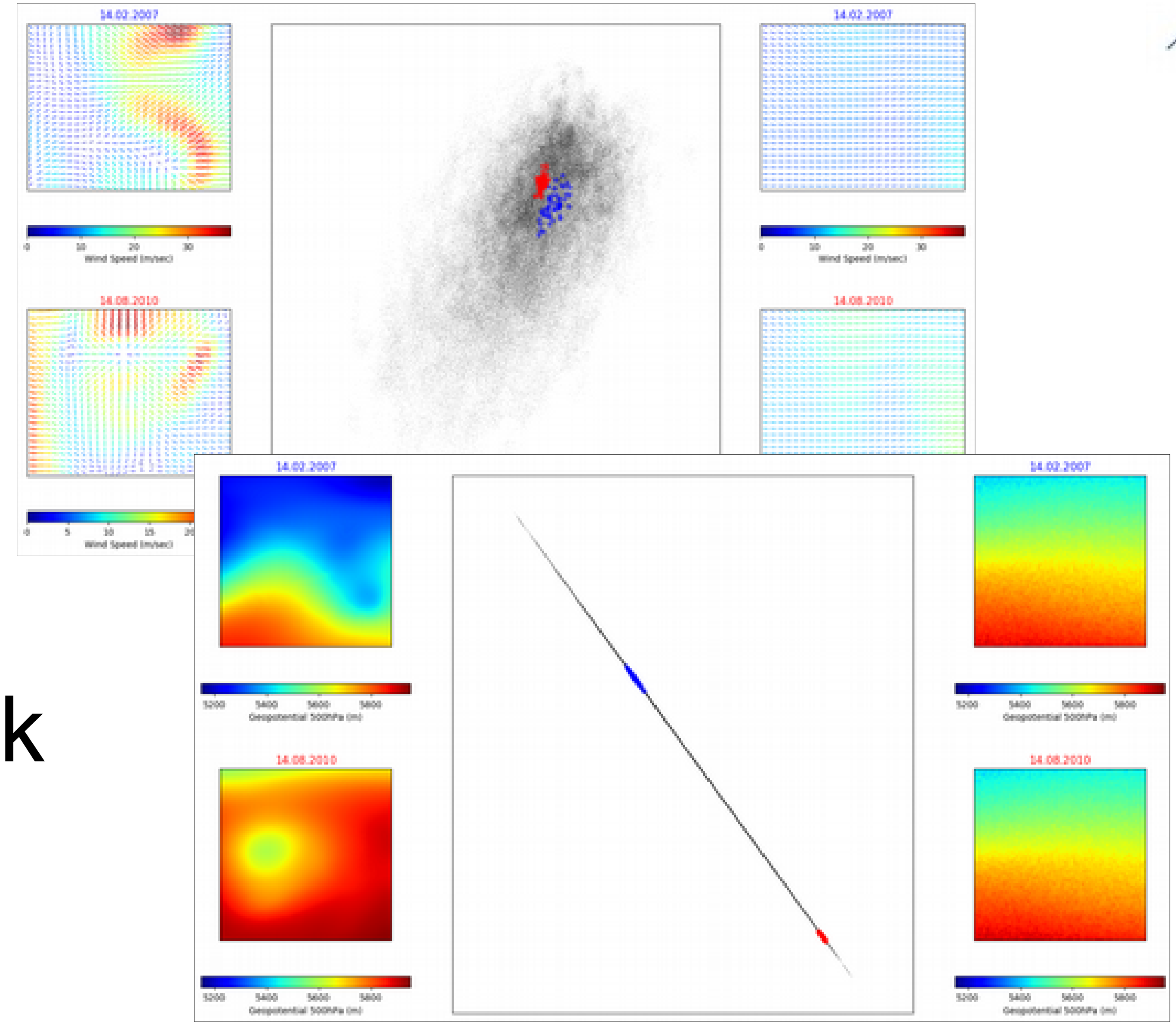
Clouds: sort of OK

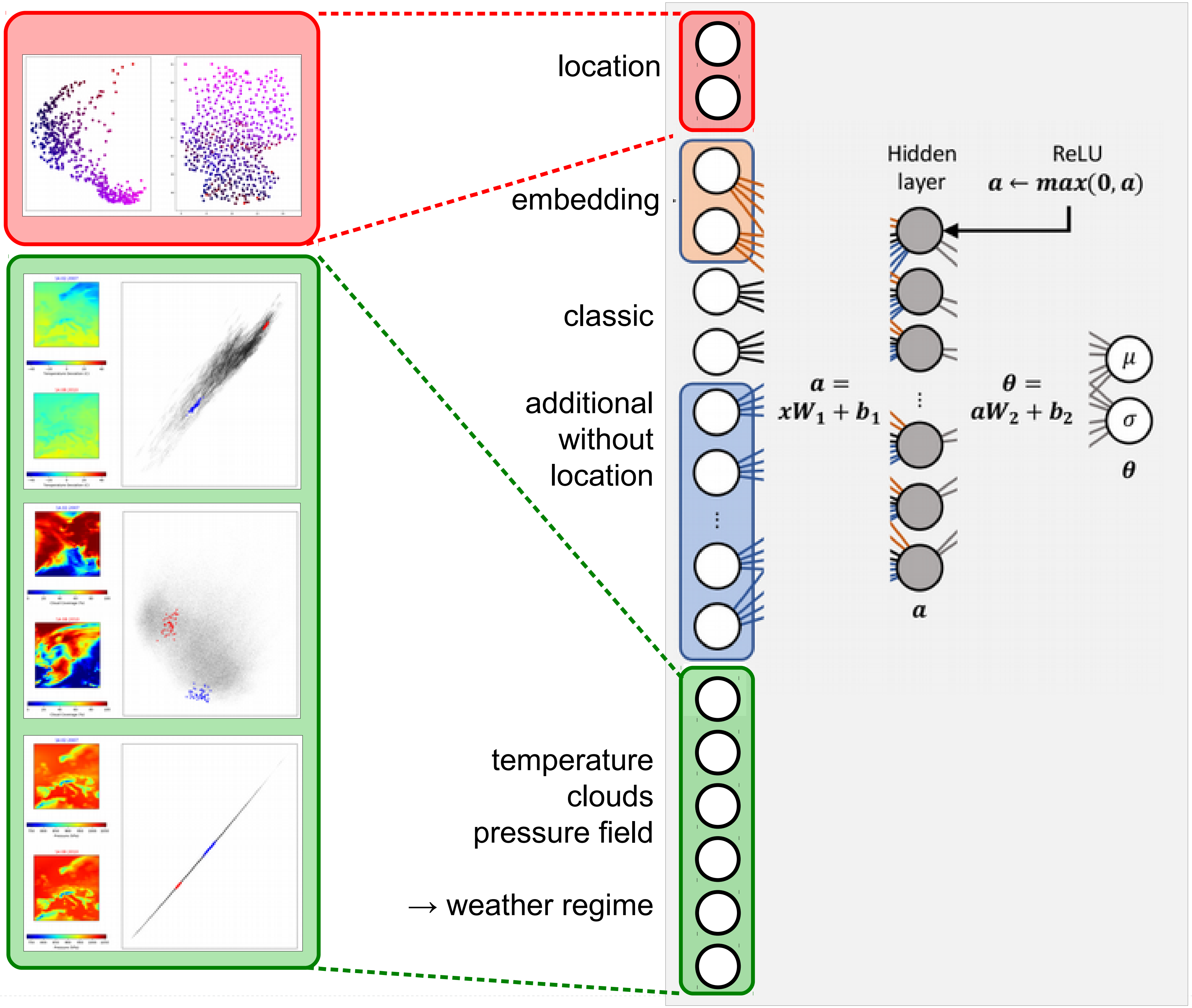
Pressure: forget about it

Wind: what?

Geopotential: better don’t ask

All the others: don’t know




location
embedding
classic
additional
without
location
temperature
clouds
pressure field
→ weather regime
Hidden
layer
ReLU
$a \leftarrow max(0, a)$
$a = xW_1 + b_1$
$\theta = aW_2 + b_2$
μ
σ
θ
a

CRPS=0.76

Conclusion

Compressing complex data might add interpretability

Use of proper scores for training helps a lot → CRPS

Next:

- better dimensionality reduction techniques
- time-series analysis / interpolation for stations
- revisit photometric redshifts