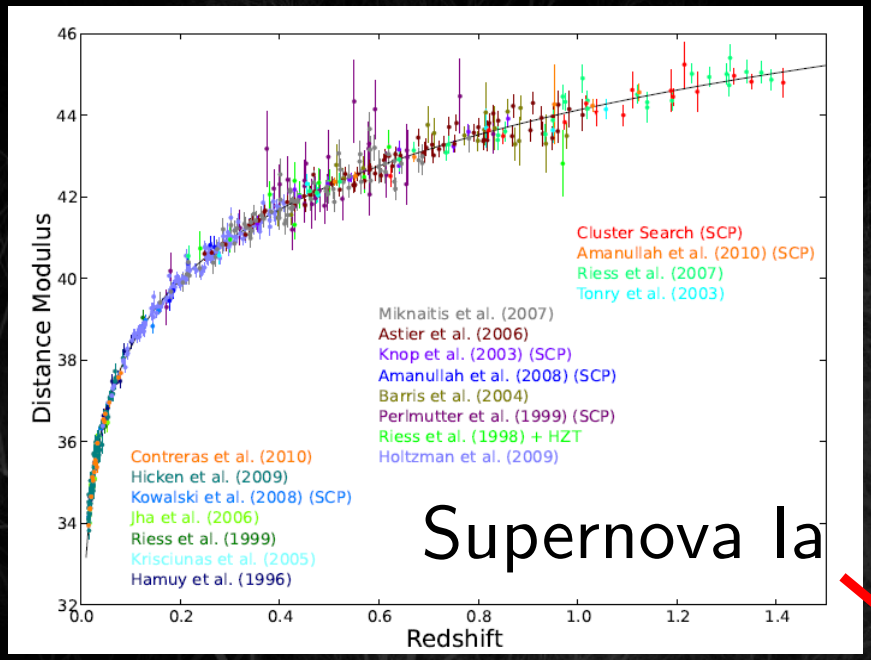
ILLINOIS
UNIVERSITY OF ILLINOIS AT URBANA-CHAMPAIGN

Photo-z, DESDM and DB

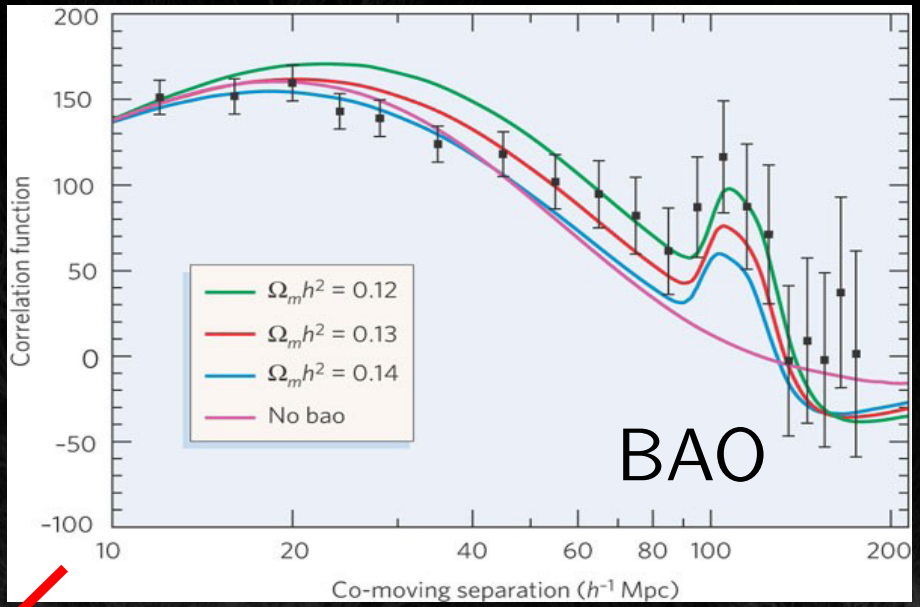
Matías Carrasco Kind
NCSA/Department of Astronomy
University of Illinois at Urbana-Champaign

DECam Community Science Workshop
March 11^{th} - 13^{th}, 2015

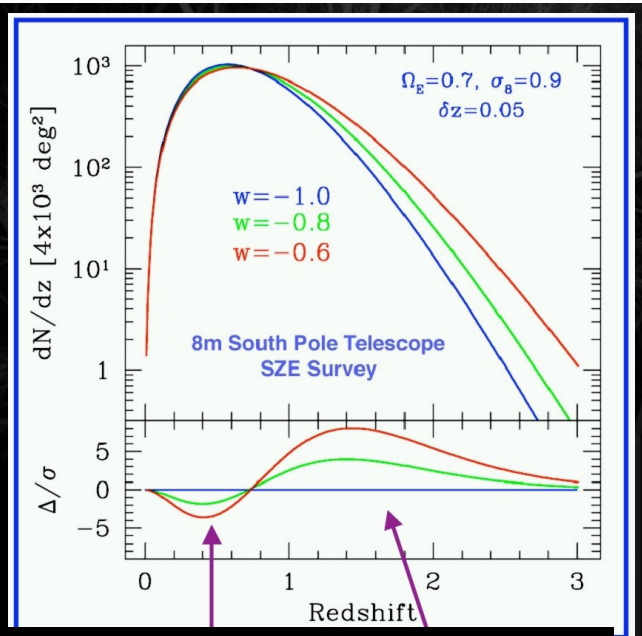
Observational Probes of Dark Energy



Supernova Ia



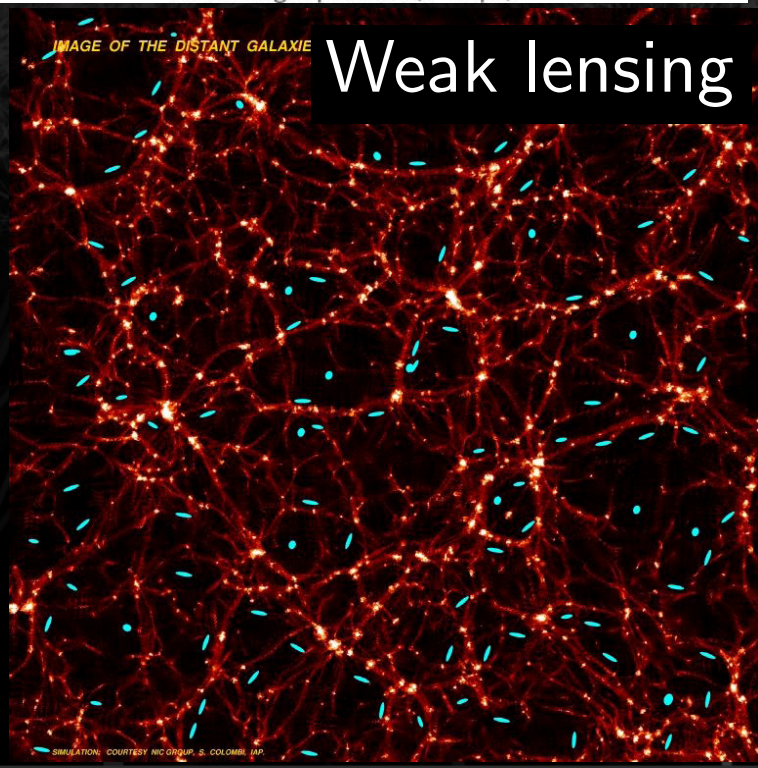
BAO



Cluster abundance

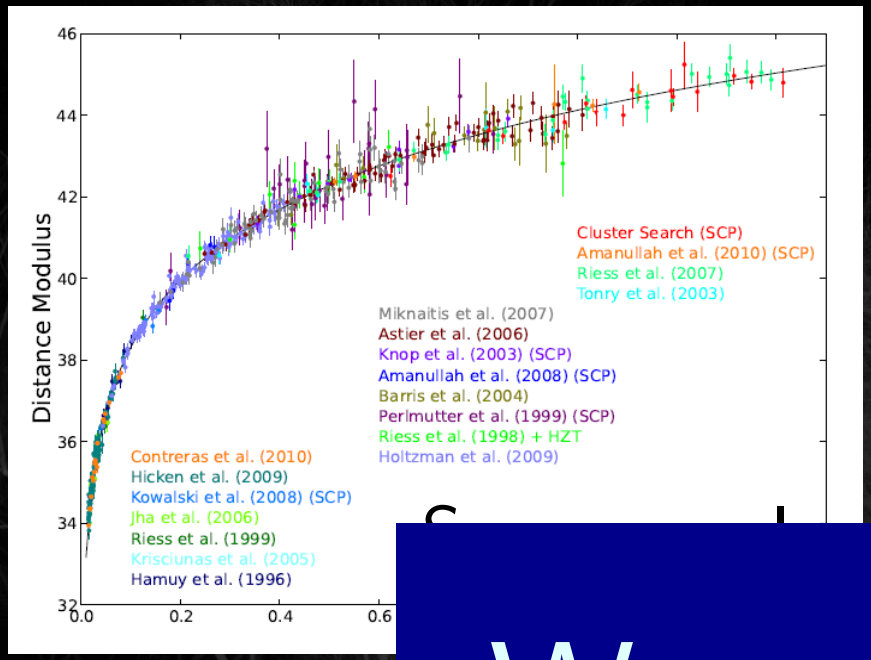
Geometry

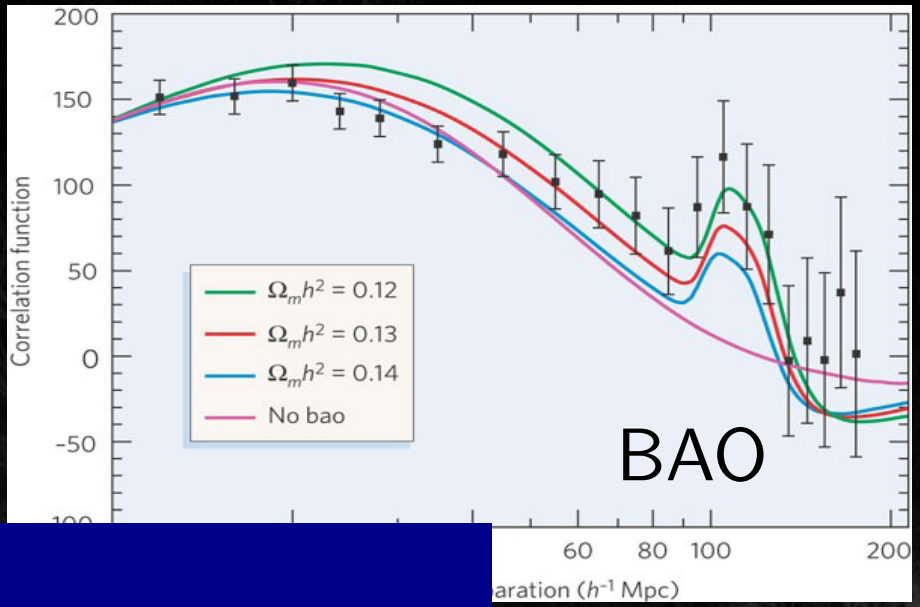
Growth of structure



Weak lensing

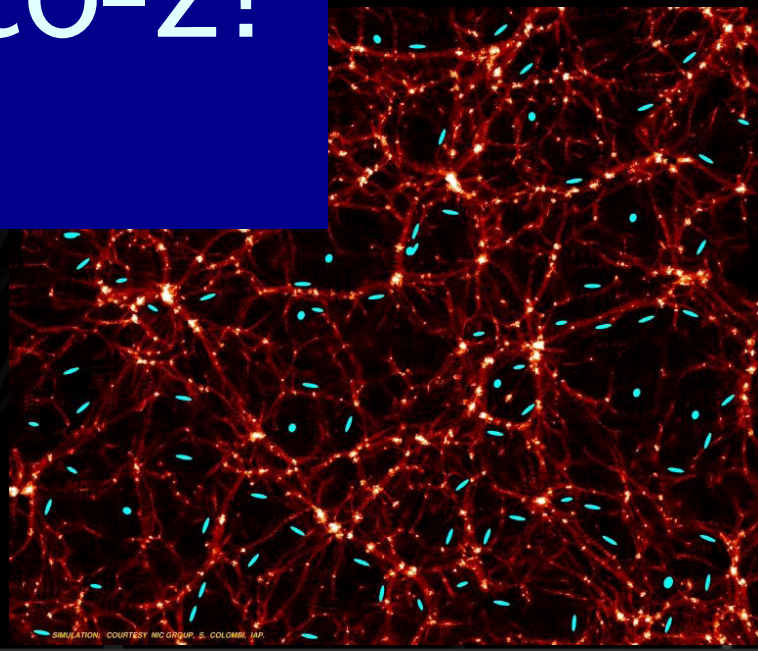
Observational Probes of Dark Energy

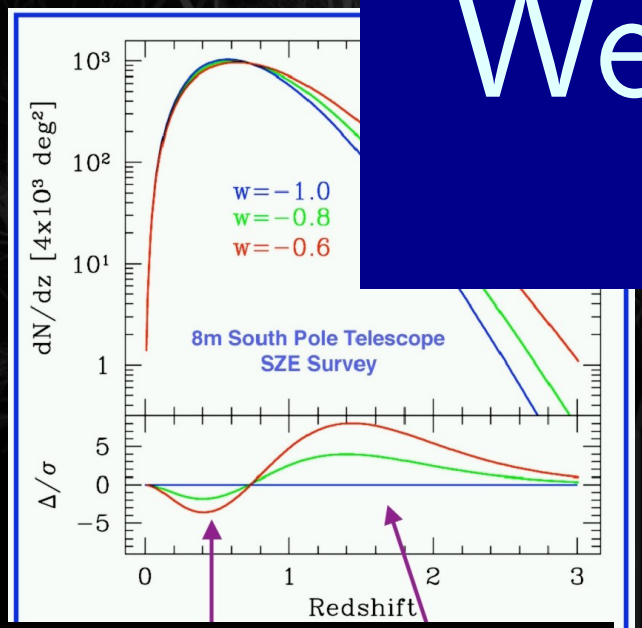




We need photo-z!

Weak lensing





Cluster abundance

Growth of structure

Photo-z for DES SV data

Monthly Notices
of the
ROYAL ASTRONOMICAL SOCIETY

MNRAS **445,** 1482–1506 (2014) doi:10.1093/mnras/stu1836

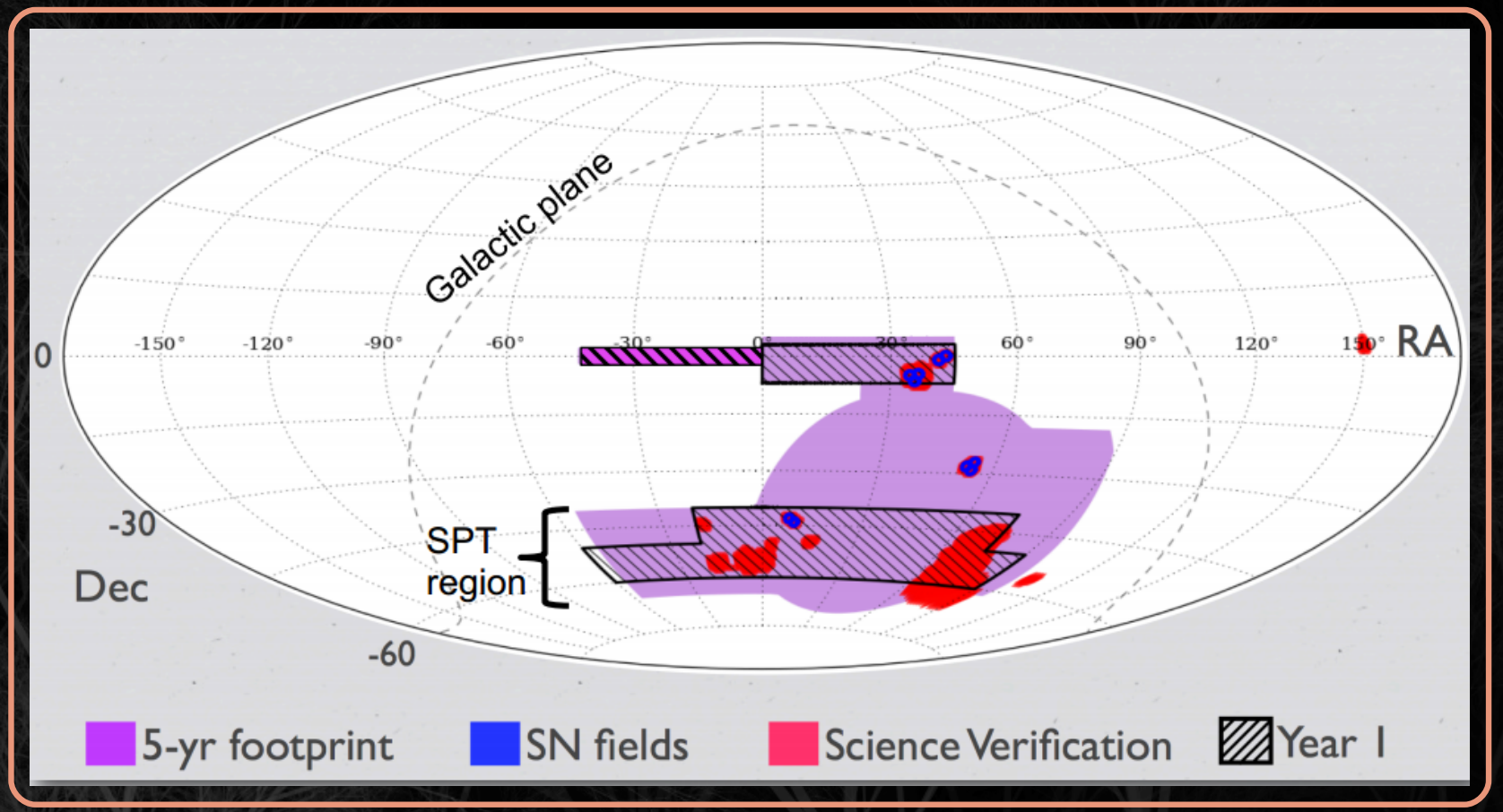
Photometric redshift analysis in the Dark Energy Survey Science Verification data

C. Sánchez,[1⋆] M. Carrasco Kind,[2,3] H. Lin,[4] R. Miquel,[1,5] F. B. Abdalla,[6] A. Amara,[7] M. Banerji,[6] C. Bonnett,[1] R. Brunner,[2] D. Capozzi,[8] A. Carnero,[9,10] F. J. Castander,[11] L. A. N. da Costa,[9,10] C. Cunha,[12] A. Fausti,[10] D. Gerdes,[13] N. Greisel,[14,15] J. Gschwend,[9,10] W. Hartley,[7,16] S. Jouvel,[6] O. Lahav,[6] M. Lima,[10,17] M. A. G. Maia,[9,10] P. Martí,[1] R. L. C. Ogando,[9,10] F. Ostrovski,[9,10] P. Pellegrini,[9] M. M. Rau,[14,15] I. Sadeh,[6] S. Seitz,[14,15] I. Sevilla-Noarbe,[18] A. Sypniewski,[13] J. de Vicente,[18] T. Abbot,[19] S. S. Allam,[4,20] D. Atlee,[21] G. Bernstein,[22] J. P. Bernstein,[23] E. Buckley-Geer,[4] D. Burke,[12,24] M. J. Childress,[25,26] T. Davis,[26,27] D. L. DePoy,[28,29] A. Dey,[21,30] S. Desai,[31,32] H. T. Diehl,[4] P. Doel,[6] J. Estrada,[4] A. Evrard,[13,33,34] E. Fernández,[1] D. Finley,[4] B. Flaugher,[4] J. Frieman,[4] E. Gaztanaga,[11] K. Glazebrook,[35] K. Honscheid,[36] A. Kim,[37] K. Kuehn,[38] N. Kuropatkin,[4] C. Lidman,[38] M. Makler,[39] J. L. Marshall,[28,29] R. C. Nichol,[8] A. Roodman,[12,24] E. Sánchez,[18] B. X. Santiago,[10,40] M. Sako,[22] R. Scalzo,[25] R. C. Smith,[19] M. E. C. Swanson,[3] G. Tarle,[13] D. Thomas,[8,41] D. L. Tucker,[4] S. A. Uddin,[26,35] F. Valdés,[21] A. Walker,[19] F. Yuan[25,26] and J. Zuntz[42]

Affiliations are listed at the end of the paper

Sanchez et al. (2015). First published paper using DES data!

Photo-z for DES SV data



- DES Science Verification data
- Photo-z code comparison and analysis
- Good benchmark results for future releases

Spectroscopic data

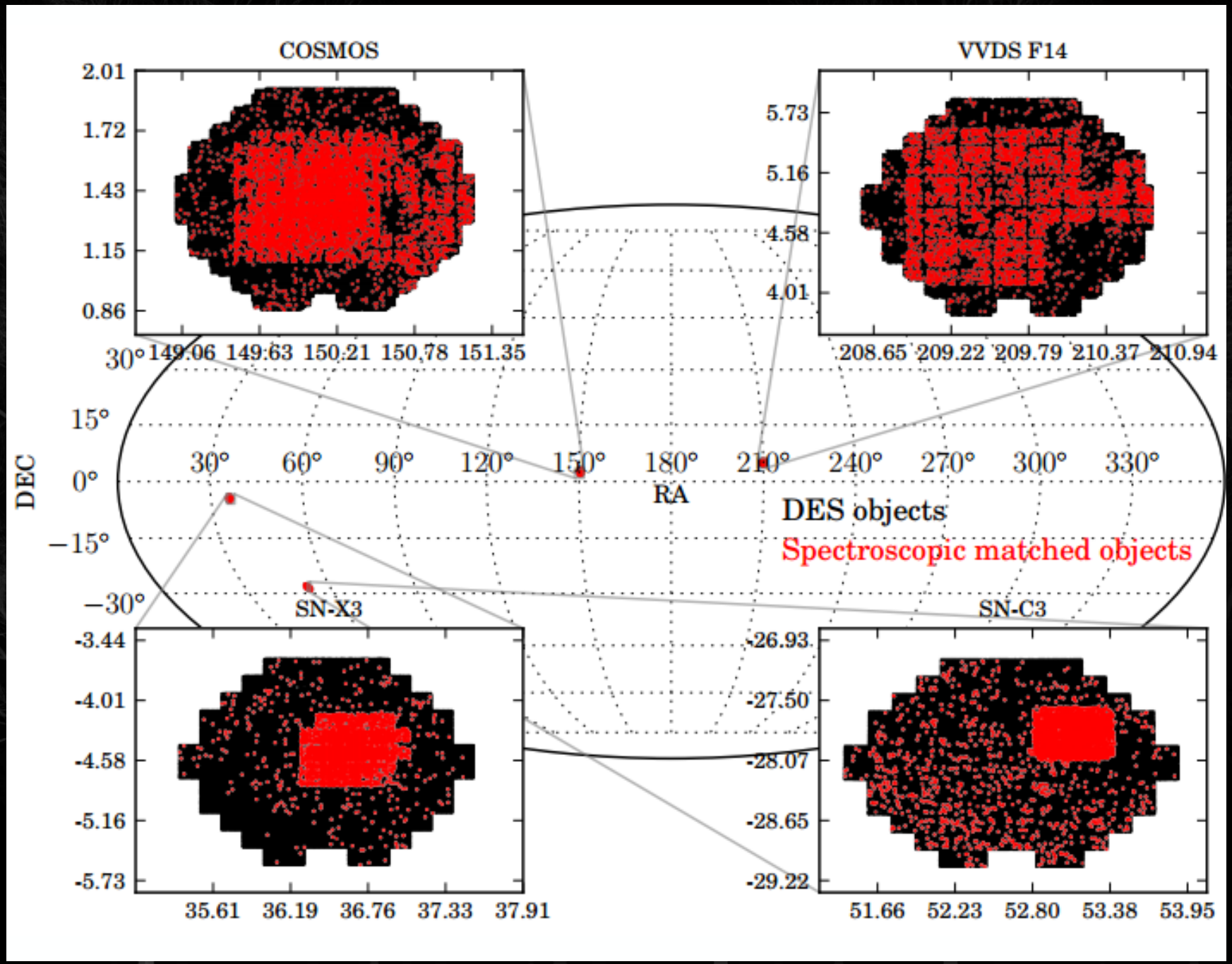
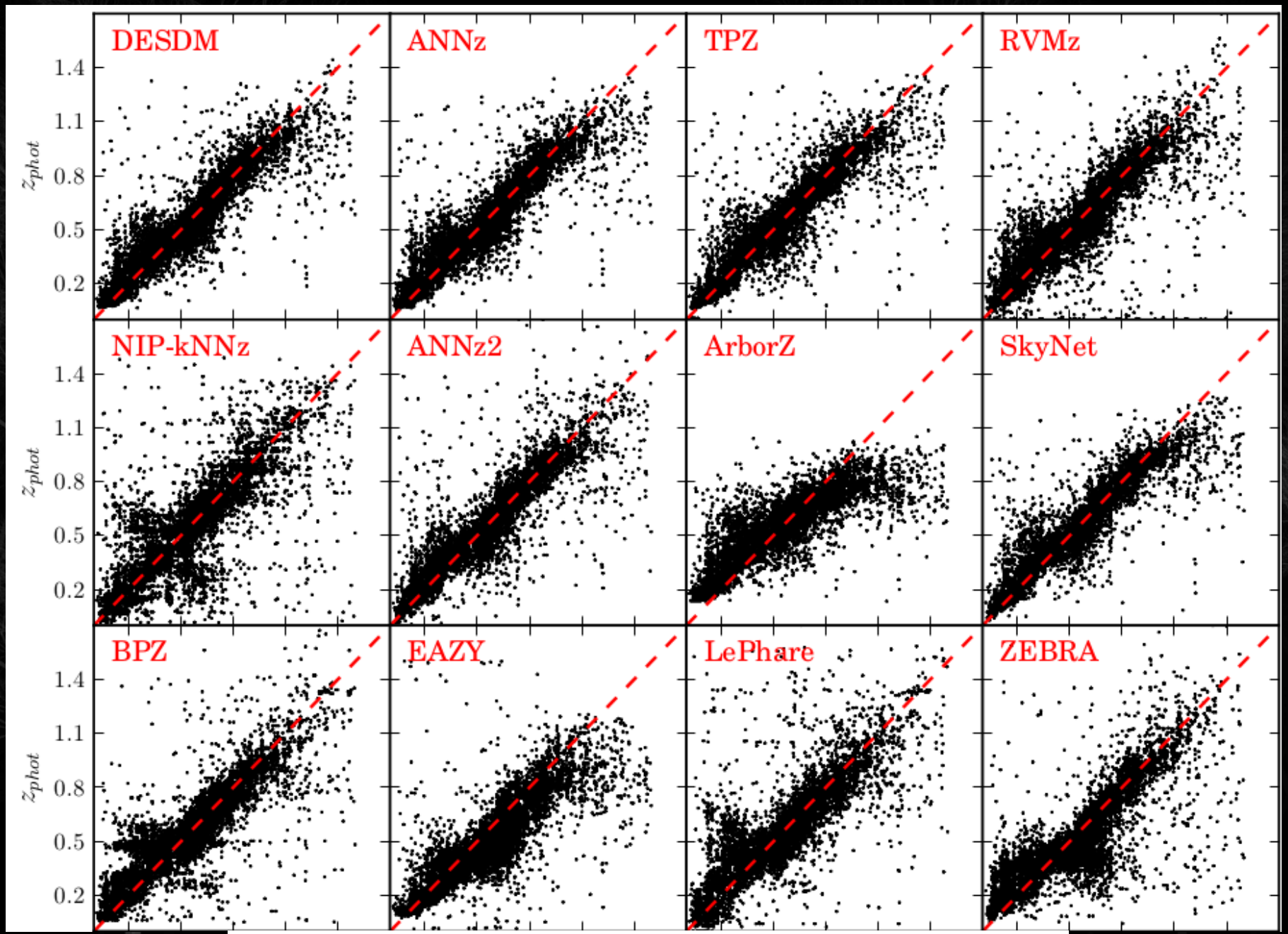


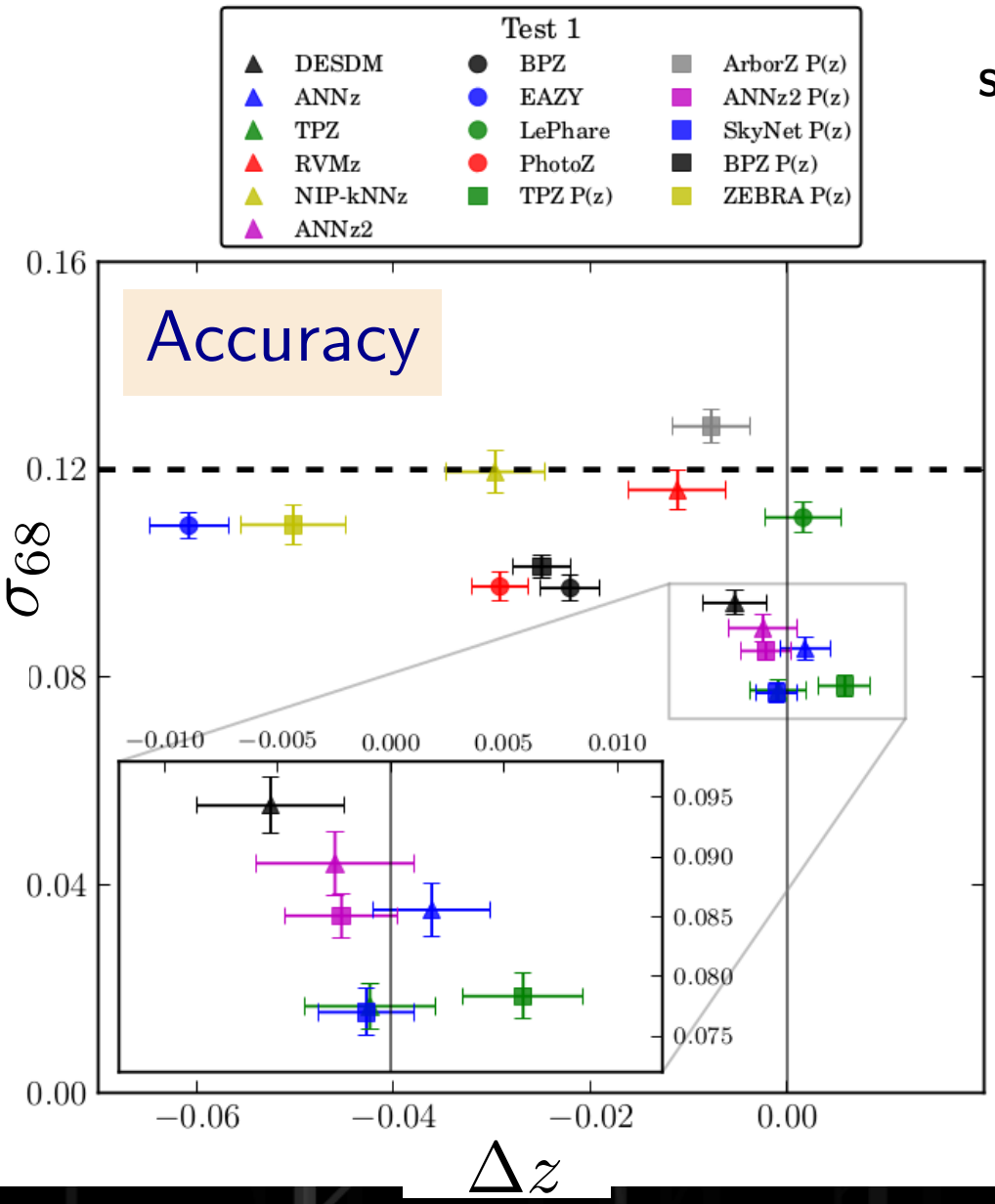
Photo-z PDF application: DES SV data

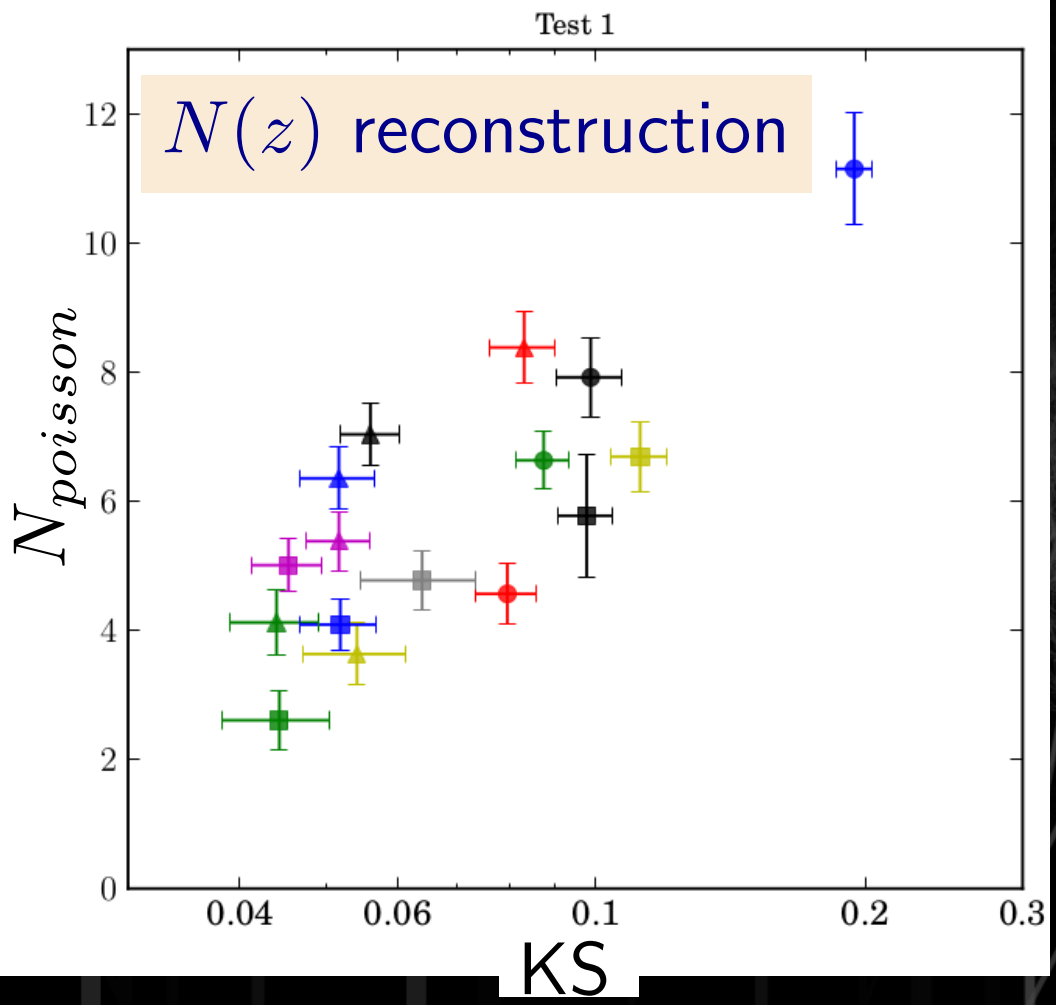


Sanchéz, Carrasco Kind, et al. 2014 (MNRAS, 445, 1482)

Photo-z for DES SV data

Sanchéz, Carrasco Kind, et al. 2014 (MNRAS, 445, 1482)

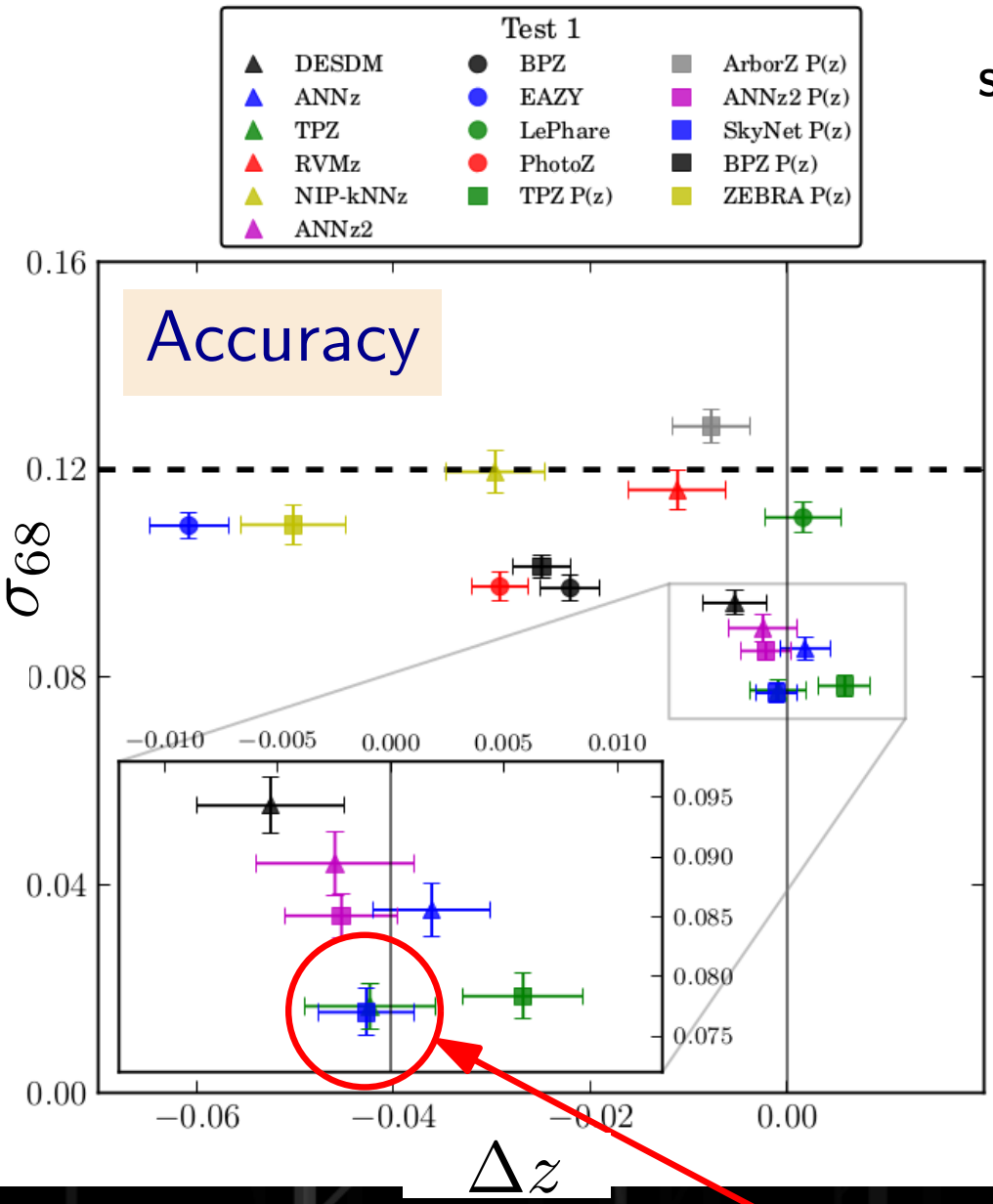


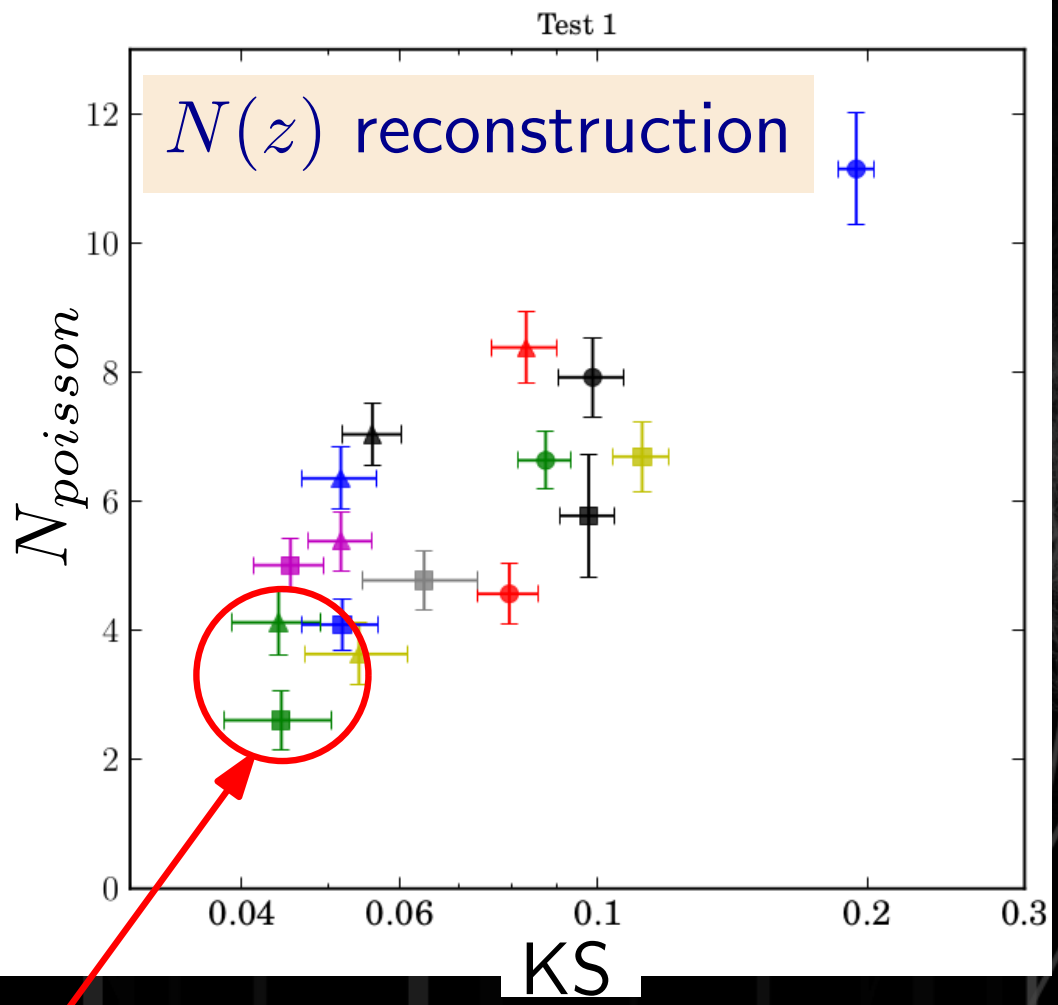


13 photo-z codes comparison

Photo-z for DES SV data

Sanchéz, Carrasco Kind, et al. 2014 (MNRAS, 445, 1482)





TPZ

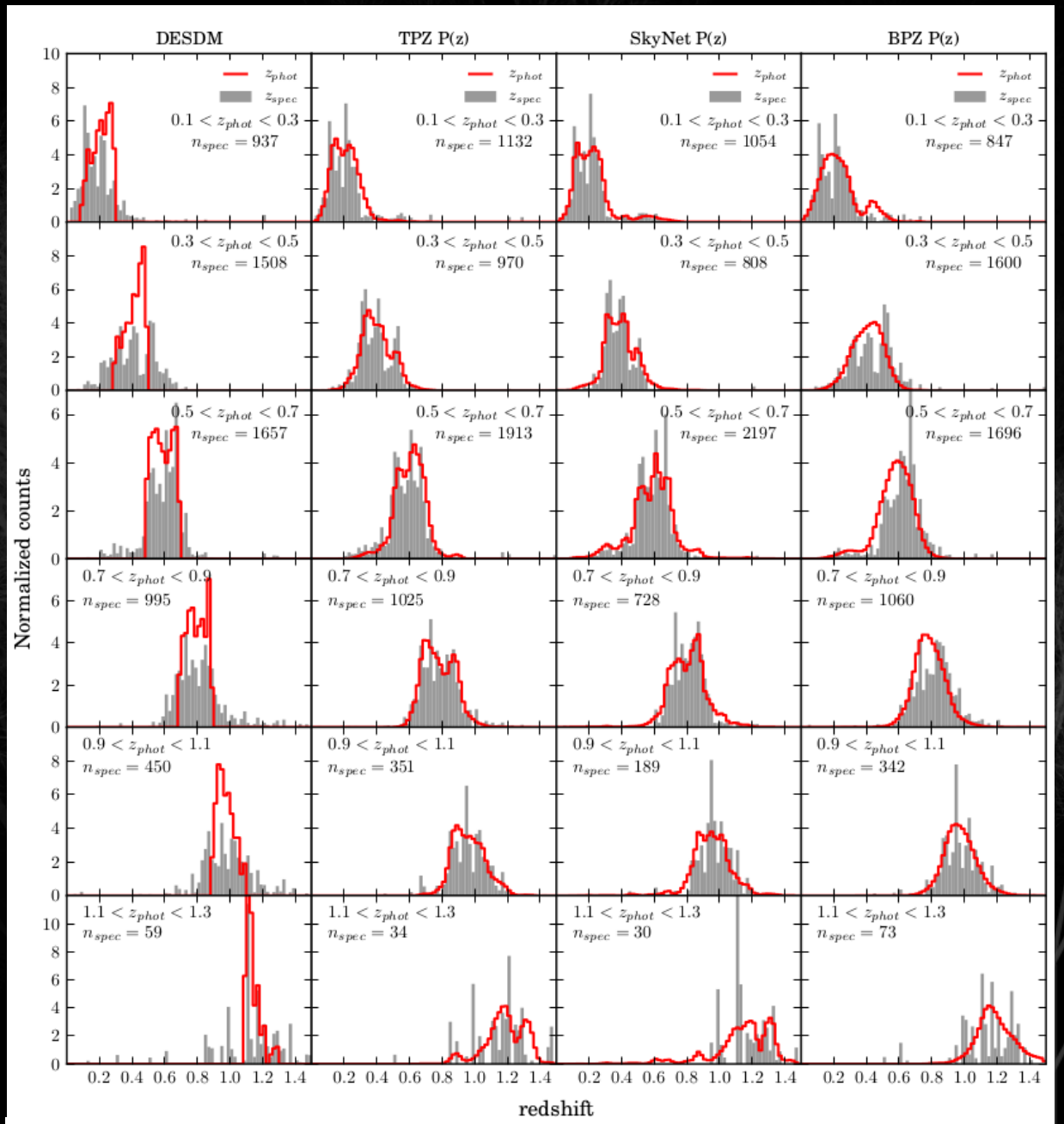
Photo-z for DES SV data

4 codes recommendation

Default, 2 training and 1 template

PDF methods are better for $N(z)$

Combination methods! (not used here)



Sanchéz, Carrasco Kind, et al. 2014 (MNRAS, submitted)

Main results

- Most methods meet DES requirements
- ML and template methods working together
- OzDES spectra and other surveys will reduce error
- Good benchmark results

Photo-z PDF representation and storage in DES DB

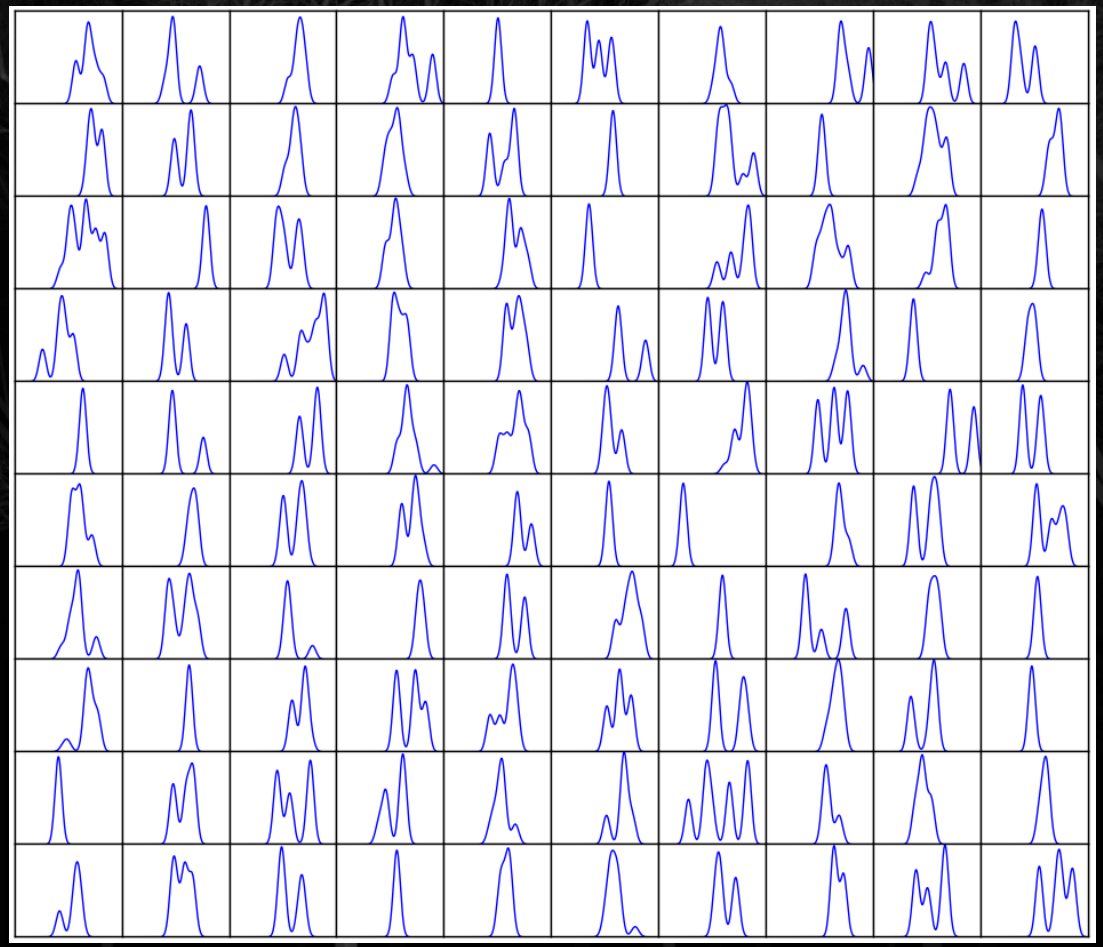


Photo-z PDF storage: Strategies

Single Gaussian fit

Multi-Gaussan fit

Monte Carlo sampling

Sparse representation techniques

Reduce number of points while increasing accuracy

Photo-z PDF storage: Strategies

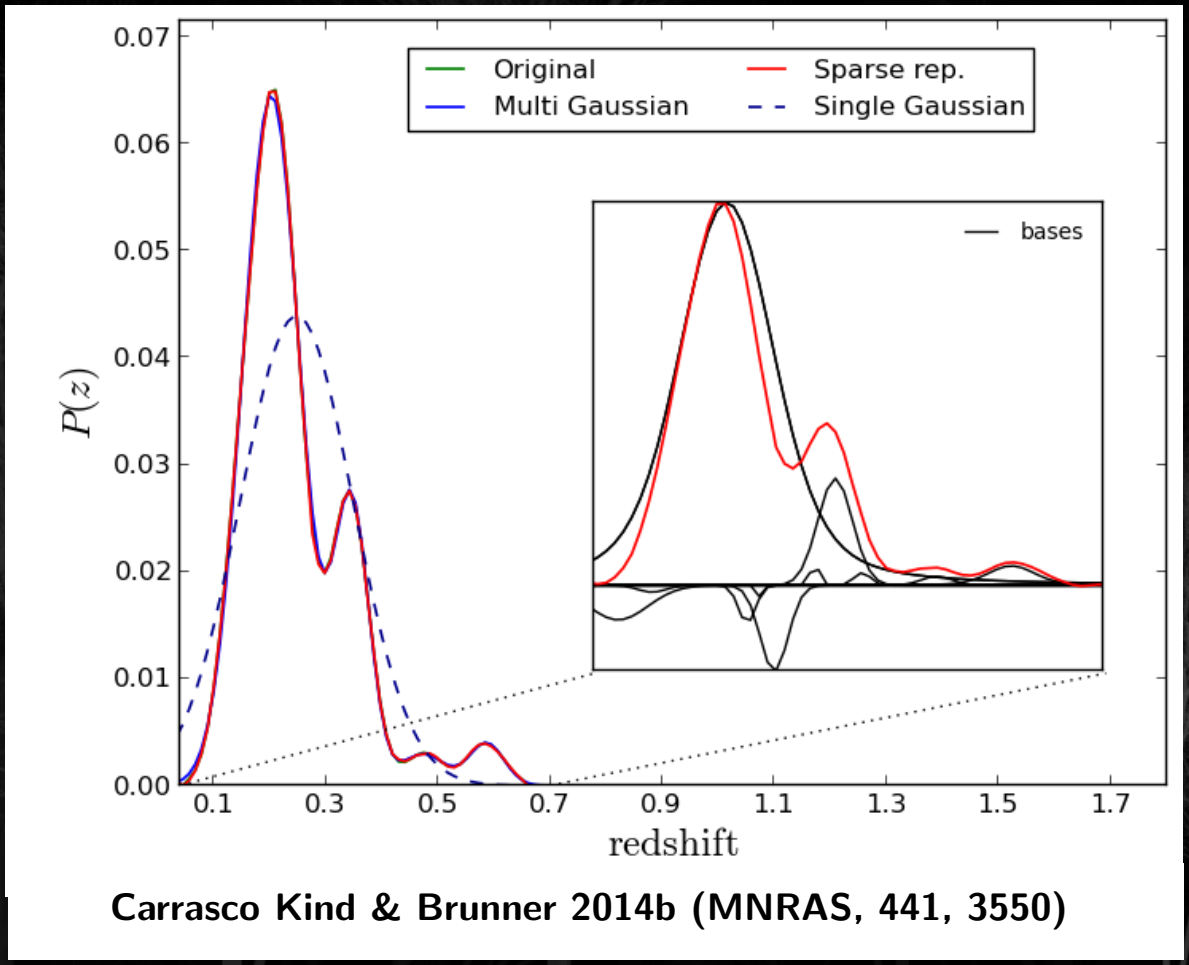
Single Gaussian fit

Multi-Gaussan fit

Monte Carlo sampling

Sparse representation techniques

Reduce number of points while increasing accuracy



Carrasco Kind & Brunner 2014b (MNRAS, 441, 3550)

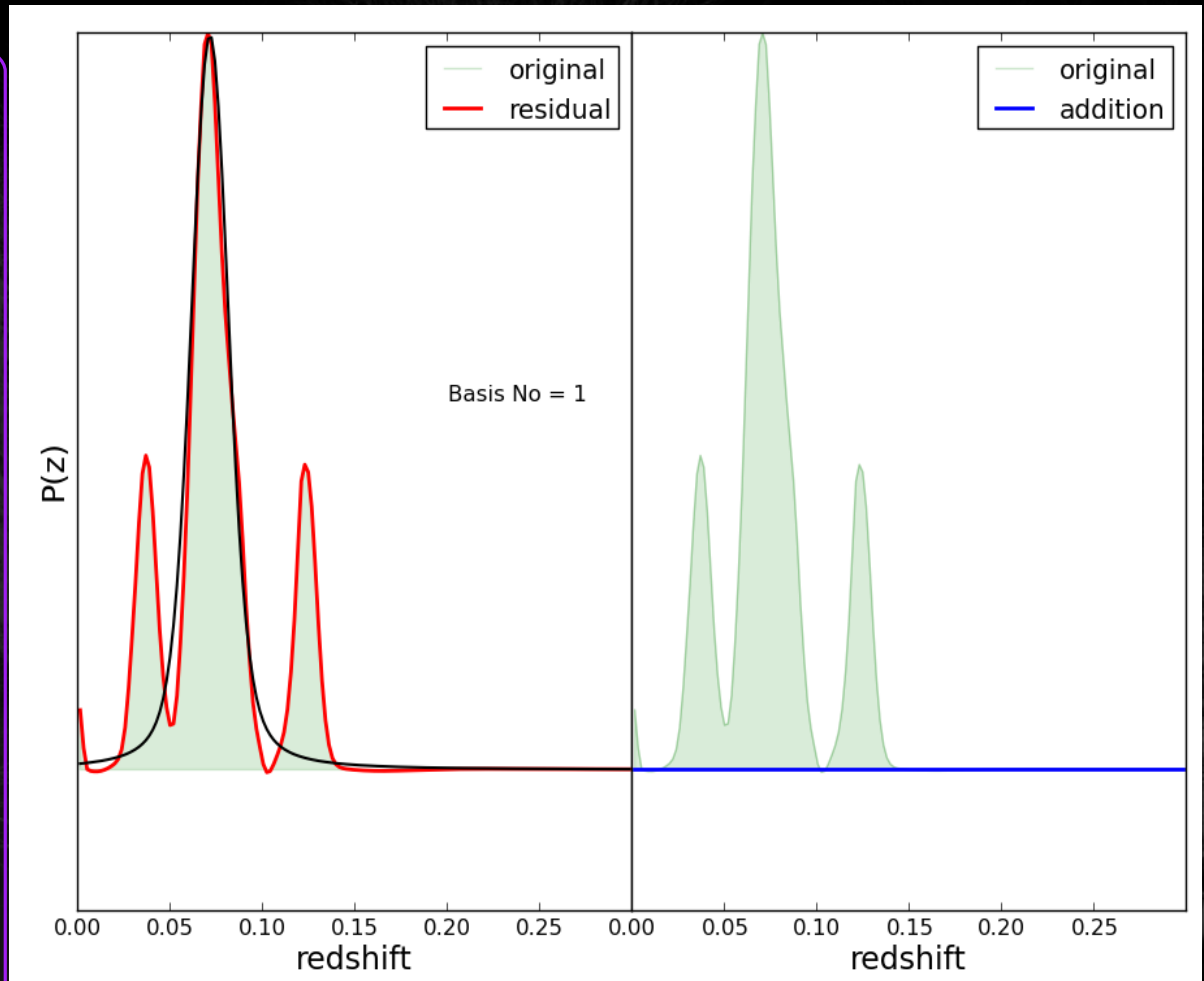
Photo-z PDF storage: Sparse representation

Use Gaussian and Voigt profiles as bases, need $N^2_{\rm original}$ bases

With only 10-20 bases achieve 99.9 % accuracy

Use 32-bits integer per basis, compression

Store Multiple PDFs



Carrasco Kind & Brunner 2014b (MNRAS, 441, 3550)

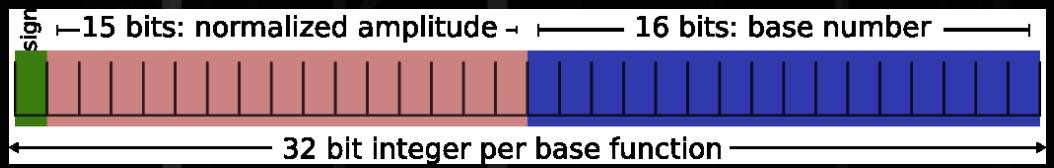


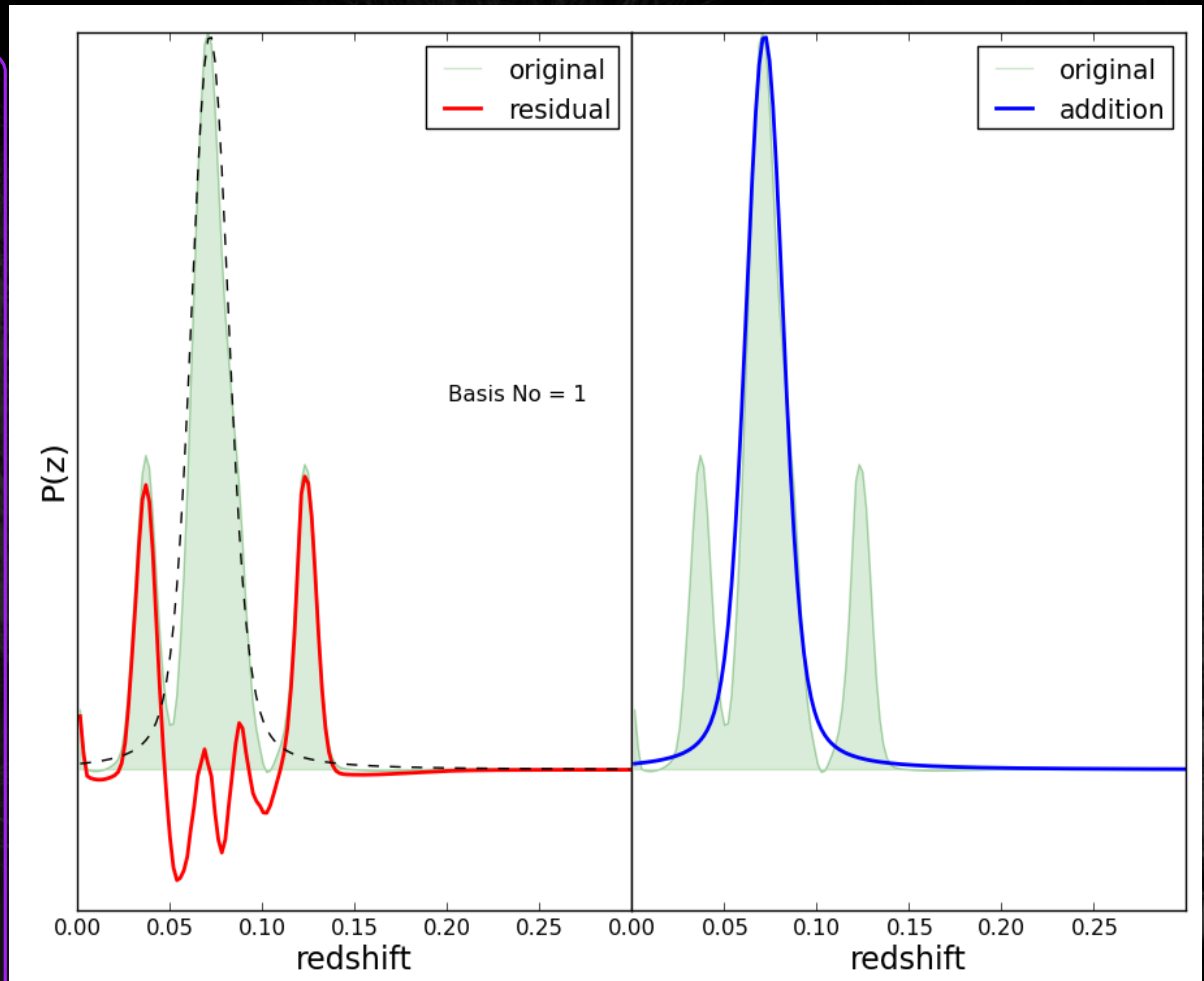
Photo-z PDF storage: Sparse representation

Use Gaussian and Voigt profiles as bases, need $N^2_{\rm original}$ bases

With only 10-20 bases achieve 99.9 % accuracy

Use 32-bits integer per basis, compression

Store Multiple PDFs



Carrasco Kind & Brunner 2014b (MNRAS, 441, 3550)

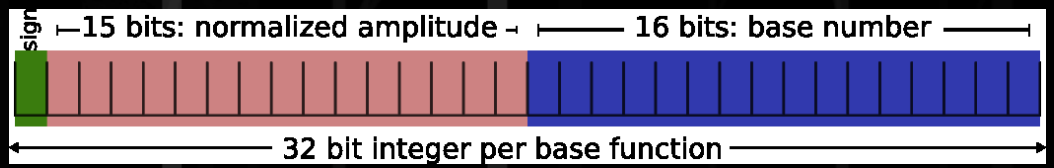


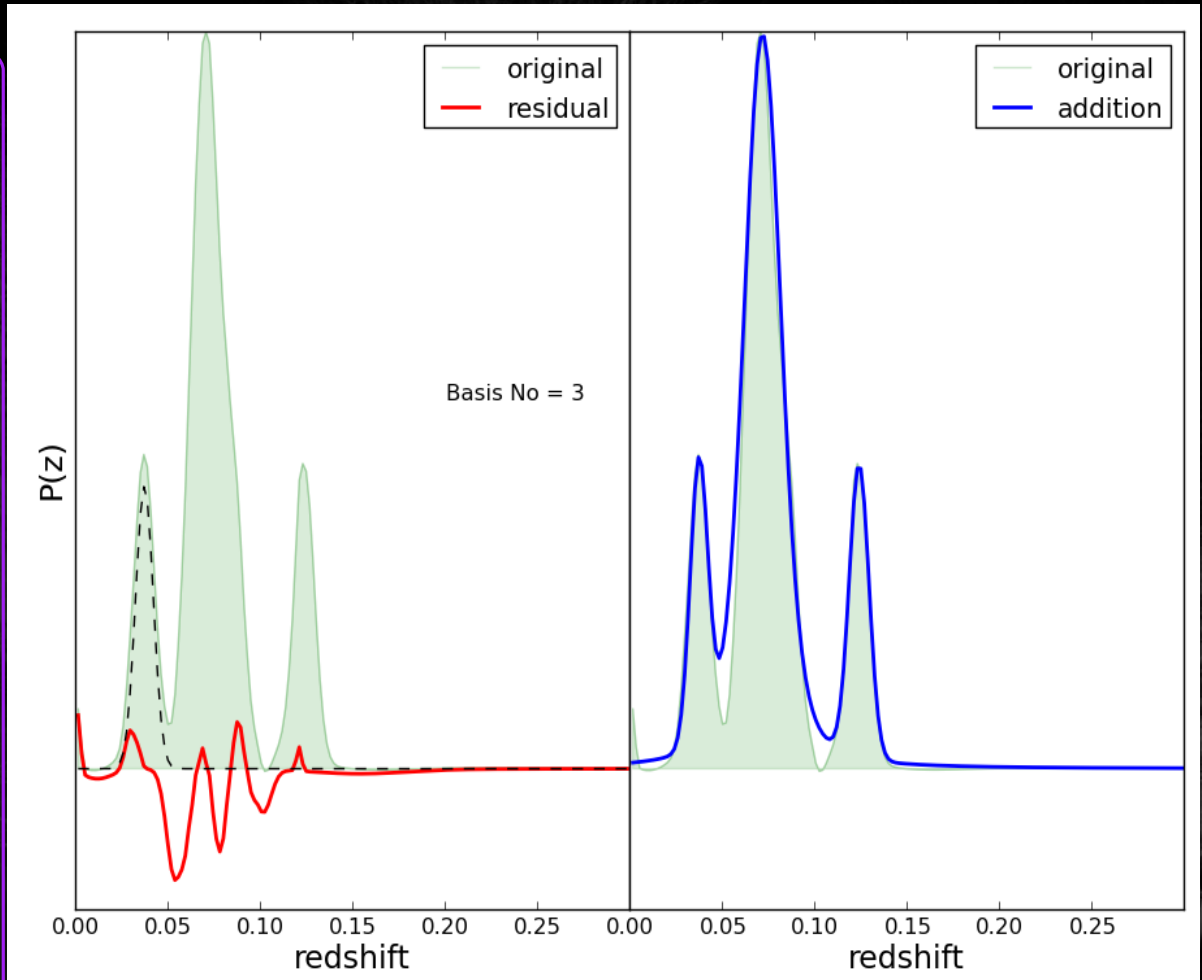
Photo-z PDF storage: Sparse representation

Use Gaussian and Voigt profiles as bases, need $N^2_{\rm original}$ bases

With only 10-20 bases achieve 99.9 % accuracy

Use 32-bits integer per basis, compression

Store Multiple PDFs



Carrasco Kind & Brunner 2014b (MNRAS, 441, 3550)

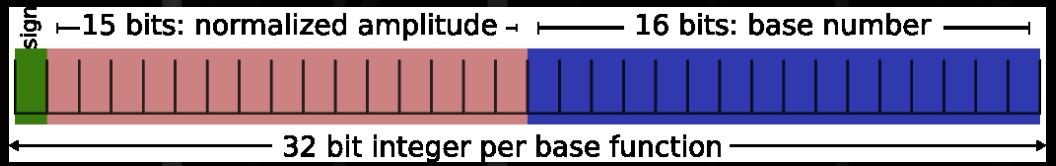


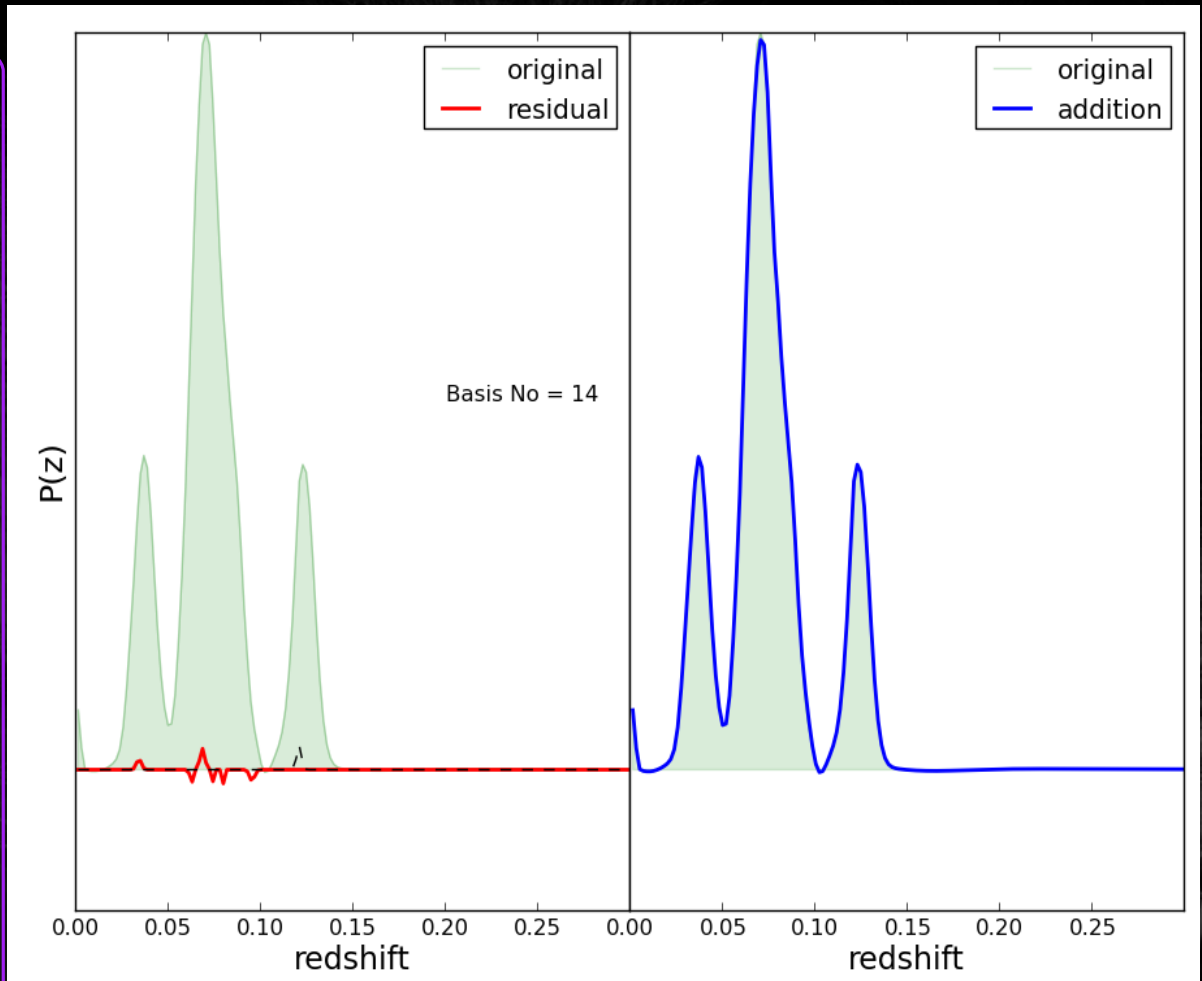
Photo-z PDF storage: Sparse representation

Use Gaussian and Voigt profiles as bases, need $N^2_{\rm original}$ bases

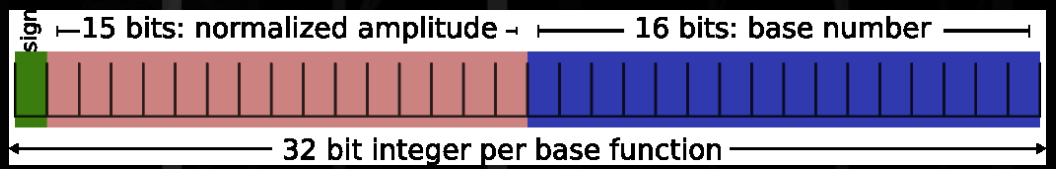
With only 10-20 bases achieve 99.9 % accuracy

Use 32-bits integer per basis, compression

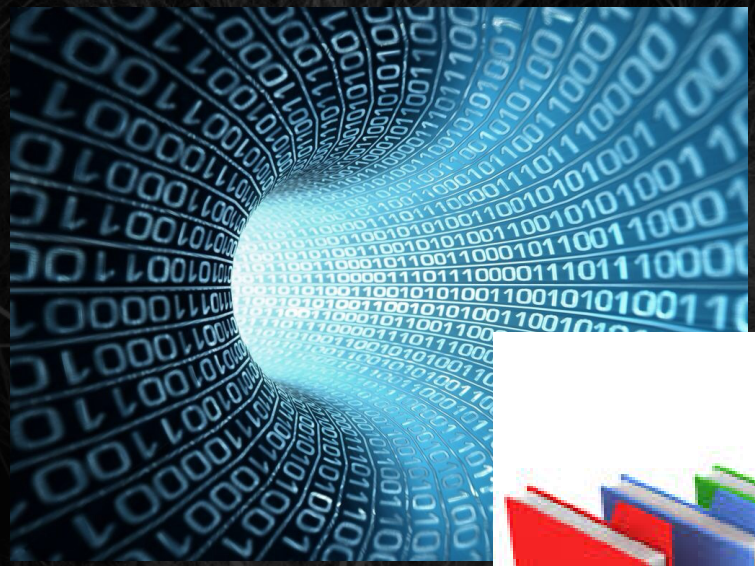
Store Multiple PDFs

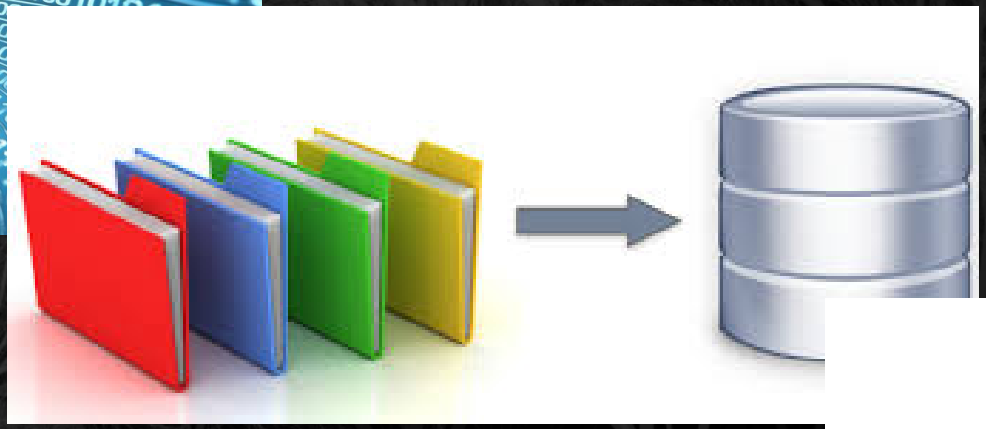


Carrasco Kind & Brunner 2014b (MNRAS, 441, 3550)



How can we do it?



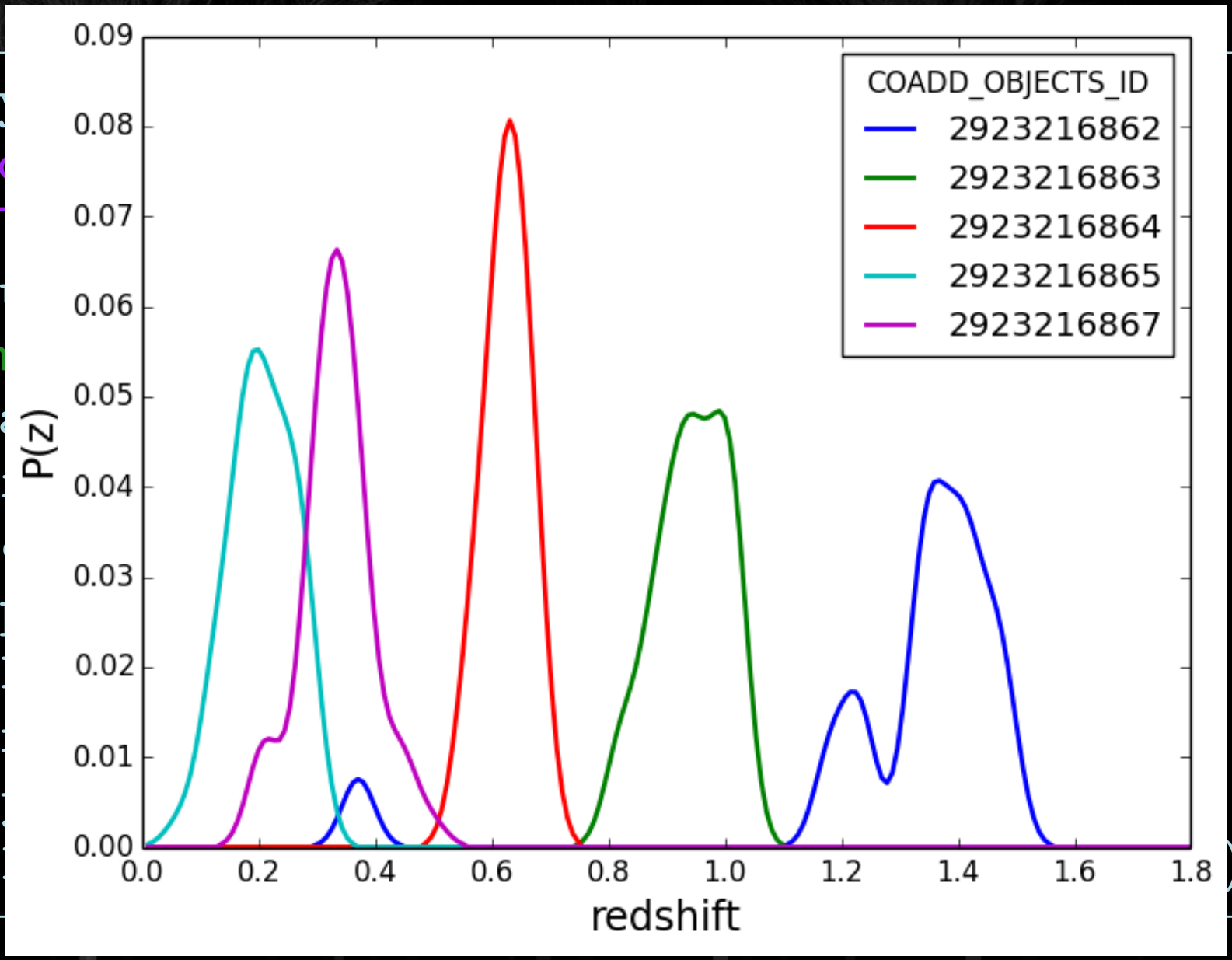


Getting some PDFs from DB

```
query="""
select COADD_OBJECTS_ID,TPZ from
PHOTOZ_PDF_SVA1_GOLD where rownum < 6"""
cc=cursor.execute(query)
#Handling and plot
df=ea.to_pandas(cc)
for i in xrange(5):
    cid=df.COADD_OBJECTS_ID.values[i]
    plt.plot(zbins,df.TPZ.values[i],
    lw=2,label=cid)
plt.xlabel('redshift',fontsize=17)
plt.ylabel('P(z)',fontsize=17)
plt.legend(loc=0, title='COADD_OBJECTS_ID')
```

Getting some PDFs from DB

```
query
sele
PHOT
cc=c
#Ha
df=e
for
plt.
plt.
plt.
```



Now using Sparse rep.

```
query="""
select COADD_OBJECTS_ID,PHZ.GET_PDF(TPZ) as
TPZ from PHOTOZ_SPARSE_SVA1_GOLD
where rownum < 6"""
cc=cursor.execute(query)
#Handling and plot
df=ea.to_pandas(cc)
for i in xrange(5):
    cid=df.COADD_OBJECTS_ID.values[i]
    plt.plot(zbins,df.TPZ.values[i],
    lw=2,label=cid)
plt.xlabel('redshift',fontsize=17)
plt.ylabel('P(z)',fontsize=17)
plt.legend(loc=0, title='COADD_OBJECTS_ID')
```

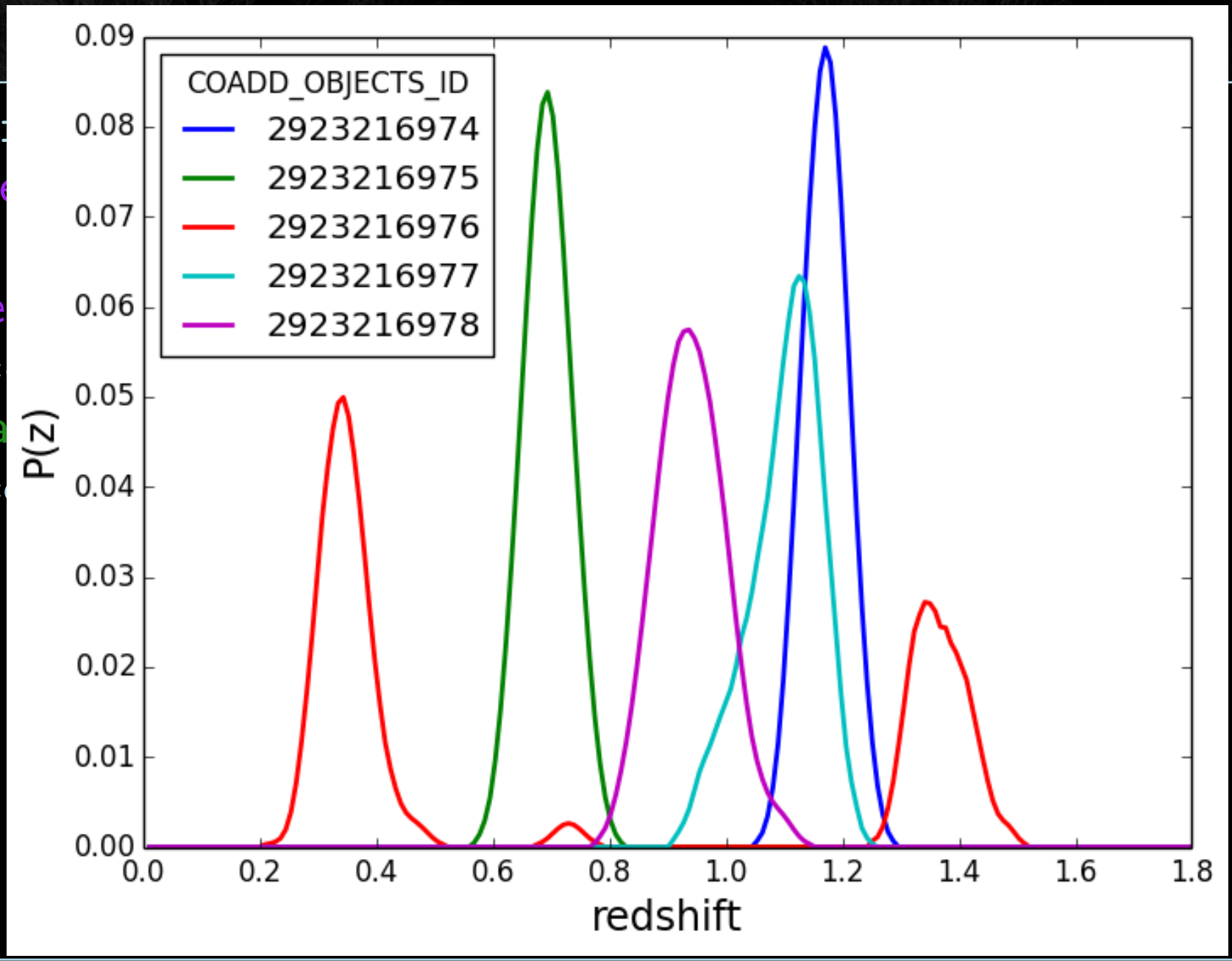
Now using Sparse rep.

```
query="""
select COADD_OBJECTS_ID,PHZ.GET_PDF(TPZ) as
TPZ from PHOTOZ_SPARSE_SVA1_GOLD
where rownum < 6"""
cc=cursor.execute(query)
#Handling and plot
df=ea.to_pandas(cc)
for i in xrange(5):
    cid=df.COADD_OBJECTS_ID.values[i]
    plt.plot(zbins,df.TPZ.values[i],
    lw=2,label=cid)
plt.xlabel('redshift',fontsize=17)
plt.ylabel('P(z)',fontsize=17)
plt.legend(loc=0, title='COADD_OBJECTS_ID')
```

Now using Sparse rep.

```
quer
sele
TPZ
whe
cc=
#Ha
df=
for

plt
plt
plt
```

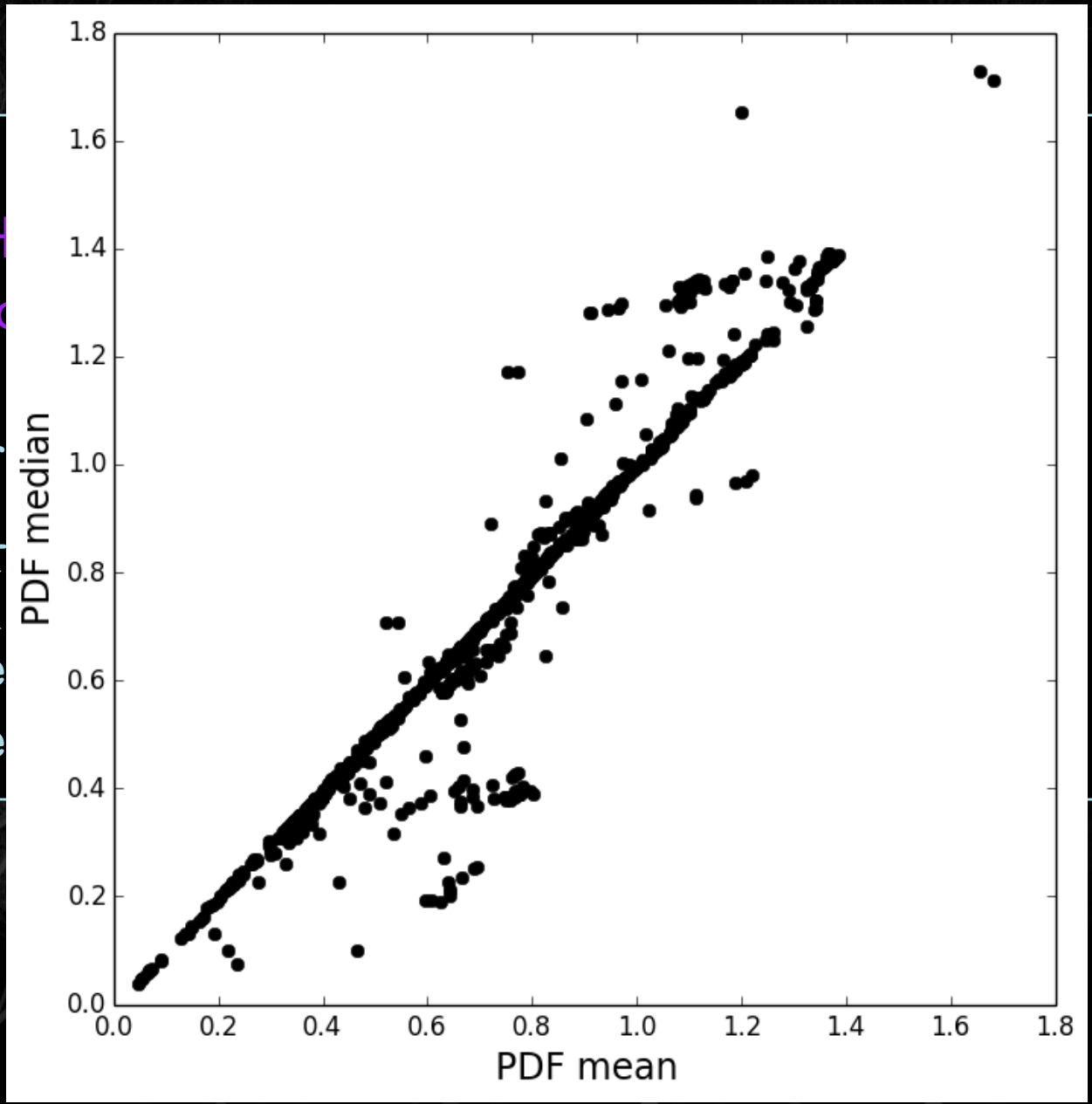


Getting metrics on the fly

```
query="""
Select PHZ.MEAN(tpz) mean, PHZ.MEDIAN(tpz)
median from PHOTOZ_PDF_SVA1_GOLD
where rownum < 1000"""
cc=cursor.execute(query)
df=ea.to_pandas(cc)
plt.plot(df.MEAN,df.MEDIAN,'ko')
plt.xlabel('PDF mean',fontsize=17)
plt.ylabel('PDF median',fontsize=17)
```

Getting metrics on the fly

```
query="""
Select PH                                    pz)
median fr
where row
cc=cursor
df=ea.to_
plt.plot(
plt.xlabe
plt.ylabe
```



Stacking PDFs in DB cluster!

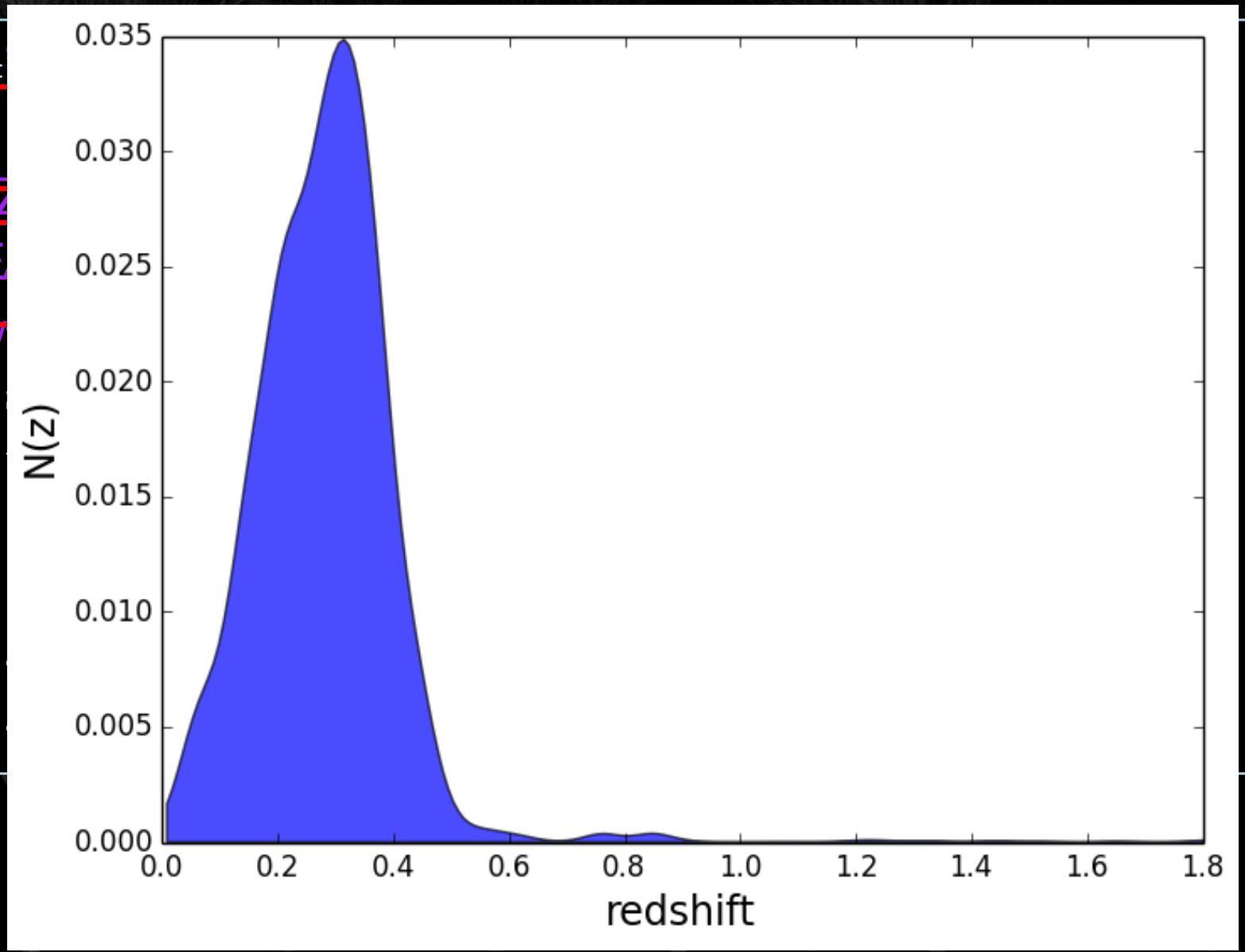
```
query="""
Select NZ(PHZ.TOTABLE(tpz)) as NZ from
PHOTOZ_PDF_SVA1_GOLD where
PHZ.MEAN(tpz) BETWEEN 0.1 and 0.4
and rownum < 100000"""
cc=cursor.execute(query)
df=ea.to_pandas(cc)
plt.fill_between(zbins,df.NZ.values[0],
     facecolor='blue',alpha=0.7)
plt.xlabel('redshift',fontsize=17)
plt.ylabel('N(z)',fontsize=17)
```

Stacking PDFs in DB cluster!

```
query="""
Select NZ(PHZ.TOTABLE(tpz)) as NZ from
PHOTOZ_PDF_SVA1_GOLD where
PHZ.MEAN(tpz) BETWEEN 0.1 and 0.4
and rownum < 100000"""
cc=cursor.execute(query)
df=ea.to_pandas(cc)
plt.fill_between(zbins,df.NZ.values[0],
     facecolor='blue',alpha=0.7)
plt.xlabel('redshift',fontsize=17)
plt.ylabel('N(z)',fontsize=17)
```

Stacking PDFs in DB cluster!

```
query=
Select
PHOTOZ
PHZ.ME
and rov
cc=cur
df=ea.
plt.fi
      f
plt.xl
plt.yl
```



Conclusions

- Photo-z in DECam look good!
- Stay tuned for DES papers
- Photo-z PDF just one example!
- First time doing science wit queries
- Bring analysis (software) to DB!

THANKS!

Questions?

Matias Carrasco Kind
NCSA/UIUC
mcarras2@ncsa.illinois.edu
http://matias-ck.com/
https://github.com/mgckind

EXTRA SLIDES