CAASTRO
ARC CENTRE OF EXCELLENCE FOR ALL-SKY ASTROPHYSICS

Australian National University

Everything you need to know to use the Terabytes

RSAA Mt Stromlo 8 April 2014

Christian Wolf

PHOTOMETRIC REDSHIFTS

What For?

- Just need redshifts
  - Have photometry but can't wait for or won't get spectra
- Probe into new regime
  - Beyond spectroscopic limits: higher-z or fainter
  - E.g. deep fields, early Universe, lensing sources
- All-sky / large samples
  - Spectroscopy infeasible
  - E.g. lensing tomography, LSS/BAO
- Find rare follow-up candidates
  - Spectra only for interesting sample
  - E.g. QSOs, early galaxies, clusters etc.

Photo-Z "Quality"

① Want z value for each object

- Want low rms error and bias
- Want to find & bin objects by z
- Sample dominated by shot noise
- E.g. galaxy evolution studies

➢ Hence need

- Map from flux observables to z
  - Calibrated templates & priors
  - Empirical training set
- Constraining data
  - Many filters, good spectral coverage and resolution

• Quality drivers

- Bias and rms $\Delta z \approx \Delta\lambda/\sqrt{N_{fil}}$
- Wide λ coverage widens z range and reduces outliers (as does non-SED data)

② Want n(z) distribution

- For objects given fluxes, sizes,…
- Large enough sample, analysis dominated by systematics
- E.g. lensing, correlation func's, cosmological parameters

➢ Hence need

- Complete model representing z-frequency given observables
- No particularly constraining data unless high z resolution required

• Trivial propagation of issues

- Incomplete? Missing facts
- Size? Poisson noise
- Unrepresenative? LSS imprint

• Algorithm affects quality

Strong Opinions…

- Template photo-z's are biased and often wrong
  - But as a deep/faint probe they're all we've got
- Empirical training sets are incomplete
  - But they are highly precise on the part they don't miss
- Machine-learning is a compression algorithm
  - But cost-effective solution for commercial environment
  - Off-the-shelf algorithms
  - For overwhelming data volumes
- Do proper statistics!
  - If computer is fast enough
- Complete the training sets!
  - TACs turn down such proposals ('too expensive, no immediate science')
- Careful error propagation
  - TACs are right… sometimes ;-)
  - What do you do with training set-based photo-z's?
    - Luminosity functions *pointless*
    - Photometric sample needs extra information not available in training sample, e.g. xy position

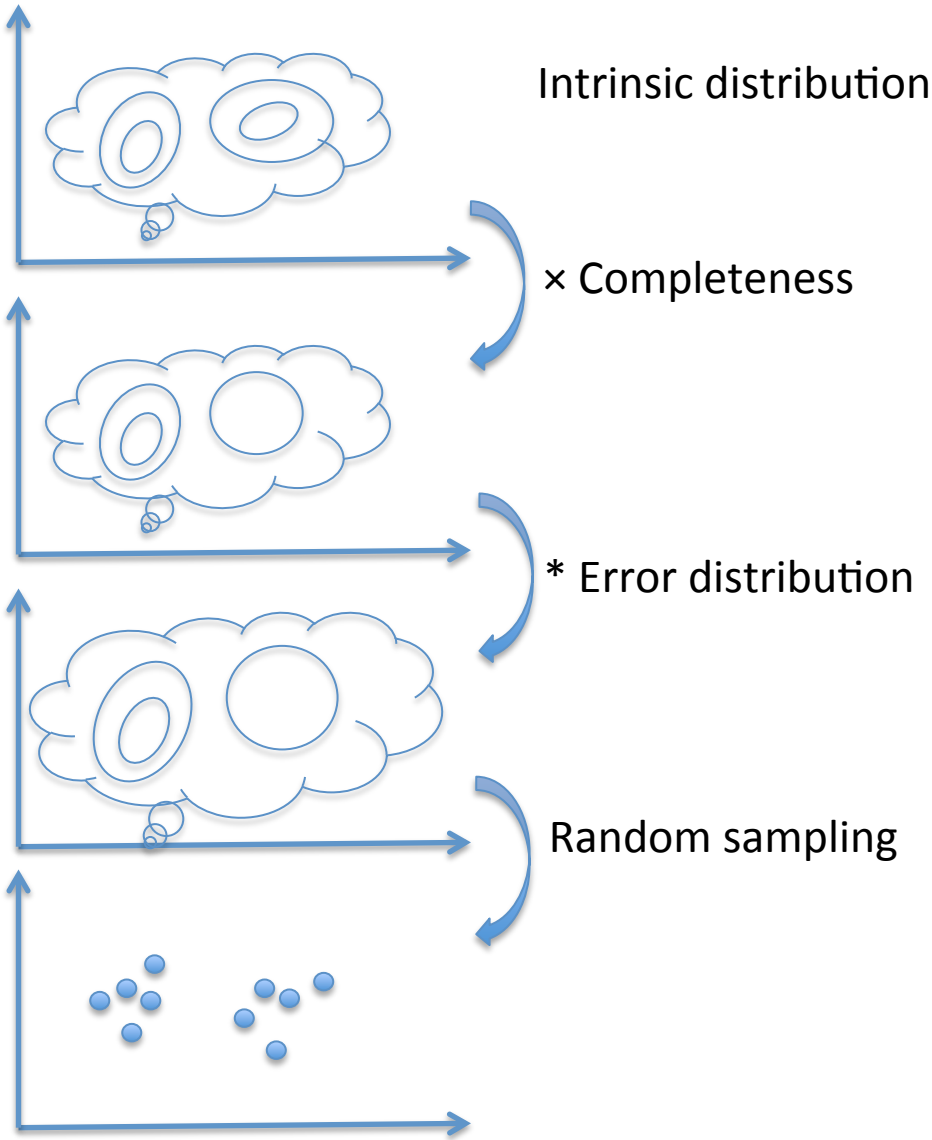
Training Sets Status

- Random samples
  - SDSS, GAMA, DEEP2a, VVDS
- Larger colour-selected samples
  - DEEP2, WiggleZ, VIPERS, …
- Tackle cosmic variance
  - More fields
  - Spectra in fields with typical photometric distribution
- Explore incompleteness
  - More of the same? Weak lines or different redshifts?
  - What are we missing from galaxy populations?

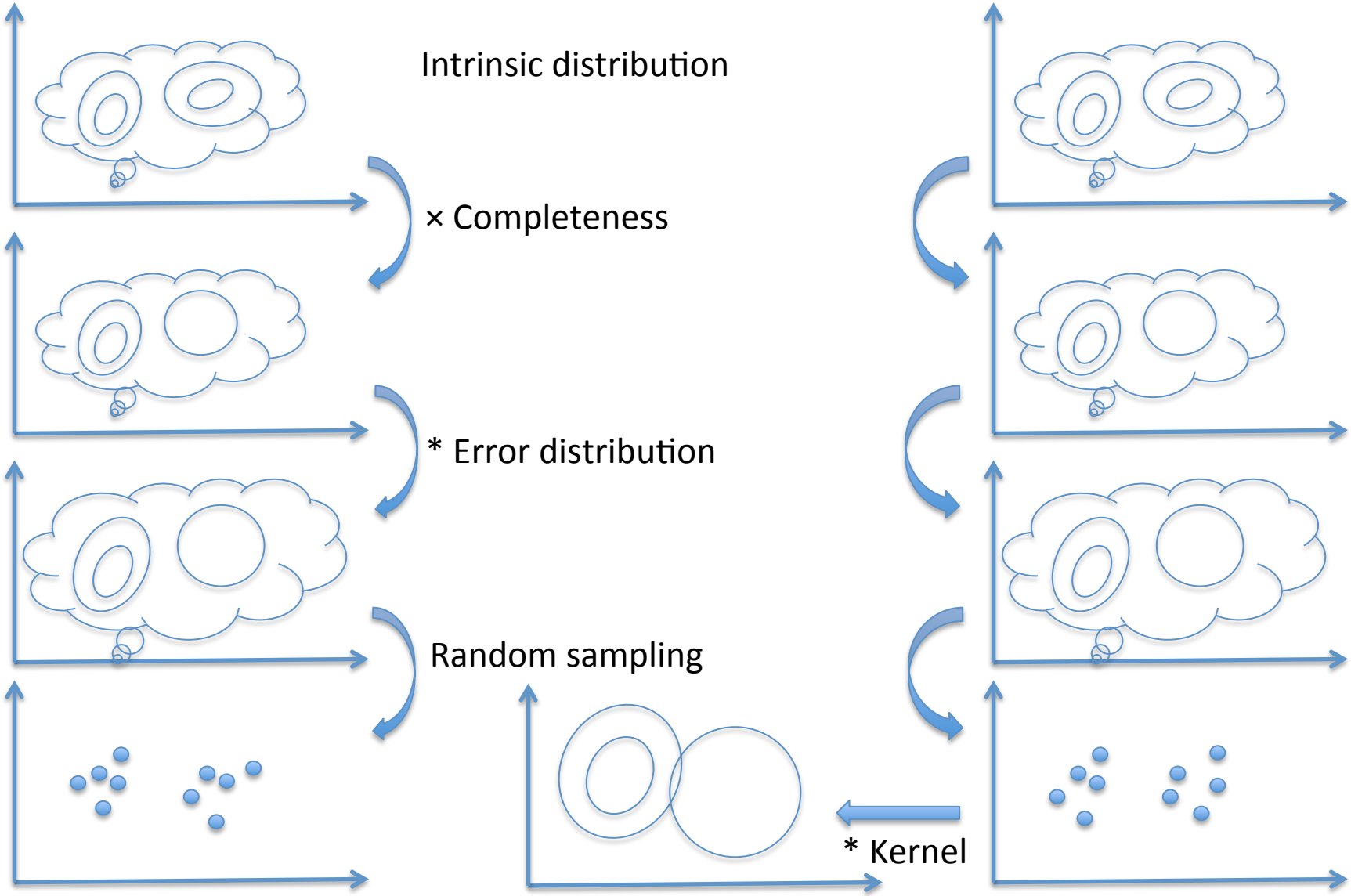
R-Mag	# spec	incompleteness
<17	1,000,000	3-5% ?
<19.5	250,000	3-5% ?
20	X,000	15%
22	10,000	25%
24	X,000	50%

- Next few years
  - OzDES-deep
  - 2dFLENS ?
  - MOSFIRE @ Keck ?
  - Make GAMA 99.x% complete?

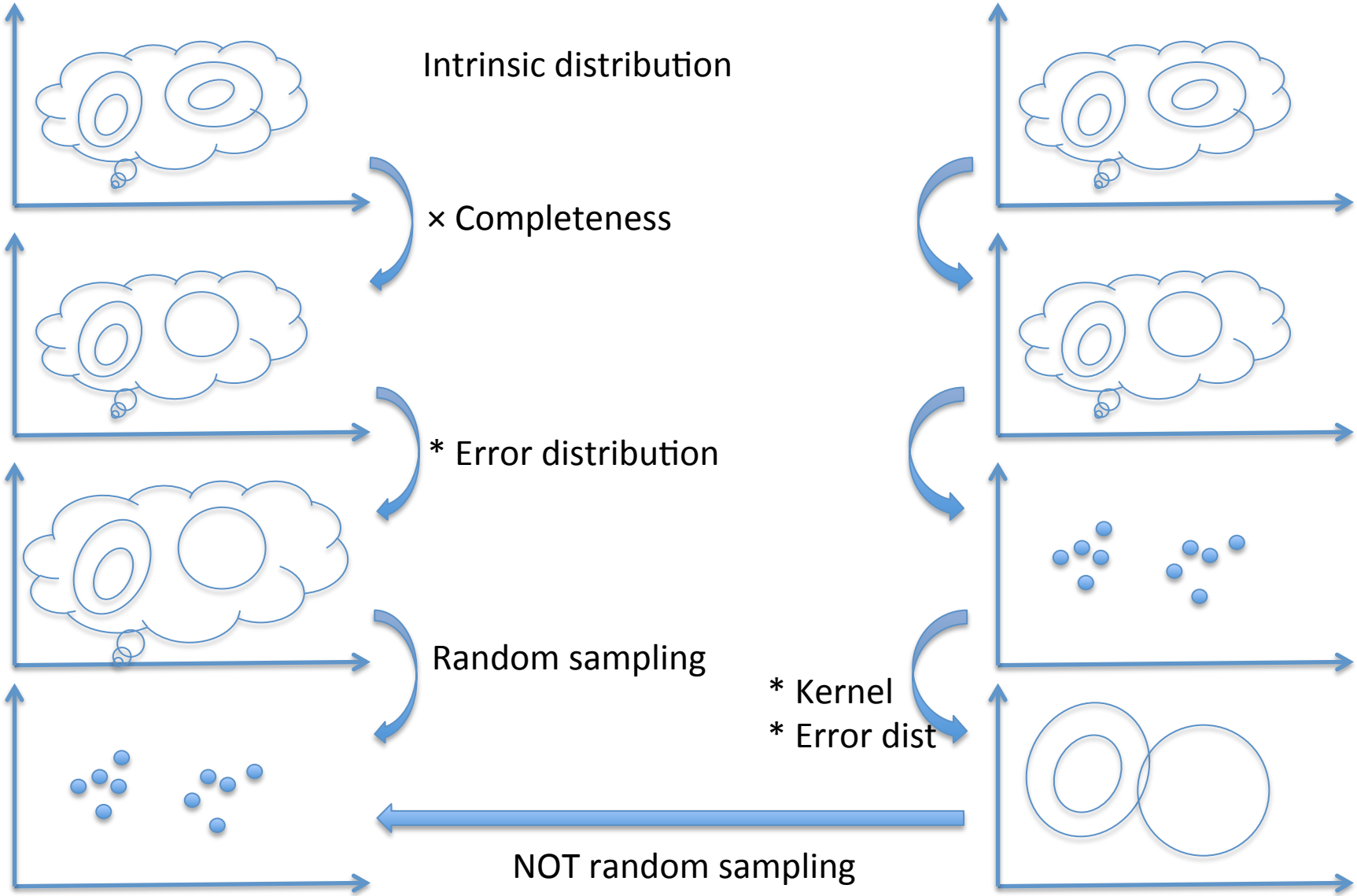
Making a Sample



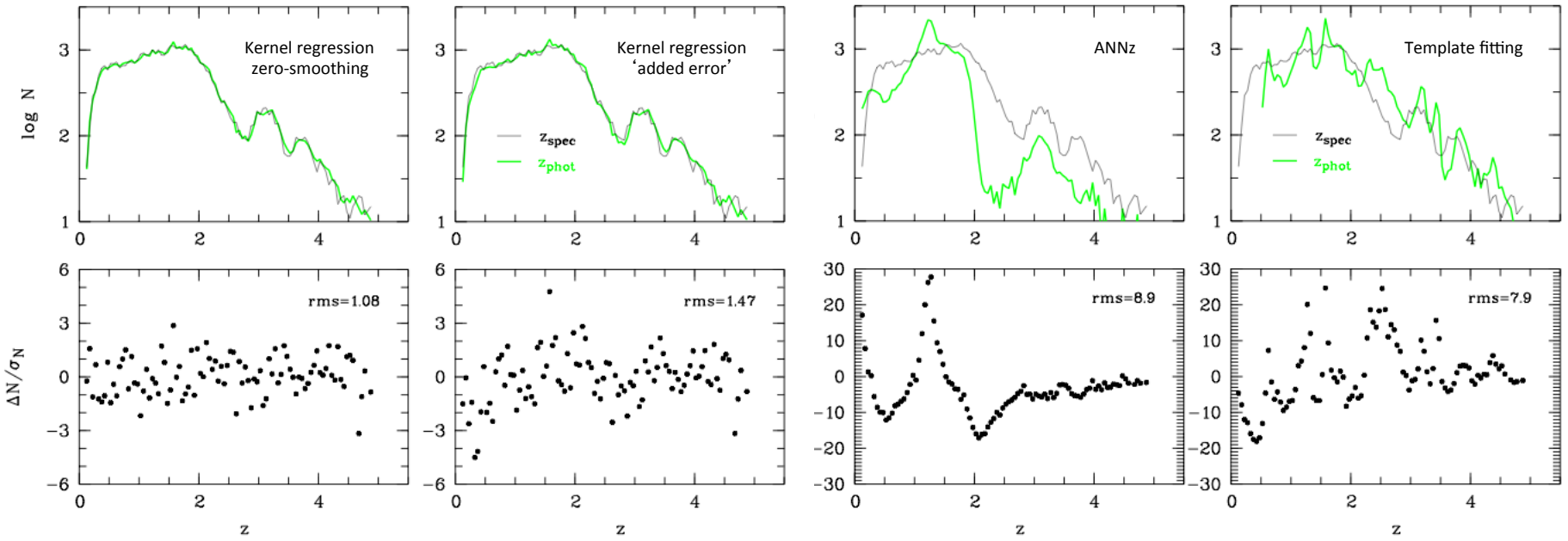
Query Sample vs. Training Sample



Query Sample vs. Training Sample



Zero-Smoothing KDE



75,000 SDSS QSOs, MNRAS, Wolf 2009

Recover n(z) of any subsample within Poisson errors
Zero bias
Estimate of residual error risk
due to incompleteness and limited size

Redshift Error Regimes

For individual z_{phot}, irrelevant for n(z)

- Saturation
  - Model-data calibration offsets, intrinsic scatter, LSS in training set
- Transition
  - Local colour(z) grid linear

$$\sigma_z \propto \sigma_{colour} \propto \frac{N}{S} \propto \frac{1}{10^{0.4m}} \Rightarrow \log\sigma_z \propto m$$

- Breakdown
  - Global colour(z) grid nonlinear