More on systematics: Intrinsic alignments and PSF interpolation with PCA

Tim Schrabback[1]

[1]Leiden Observatory

IPM School and Workshop on Weak Lensing and Photoz Techniques (ISWLP09)

Outline

Intrinsic alignments

e.g. Heymans & Heavens 2003; King & Schneider 2003; Heymans et al. 2004; Takada & White 2004; Hirata & Seljak 2004; Mandelbaum et al. 2006; Bridle & King 2007; Hirata et al. 2007; Kitching et al. 2008; Zhang 2008; Brainerd et al. 2009, astro-ph/0904.3095 + references

Correlate galaxy ellipticities with separation θ, $z_a \leq z_b$:
$\langle e_a, e_b \rangle = \langle (e_a^s + \gamma_a), (e_b^s + \gamma_b) \rangle = \langle e_a^s, e_b^s \rangle + \langle e_a^s, \gamma_b \rangle + \langle \gamma_a, e_b^s \rangle + \langle \gamma_a, \gamma_b \rangle$

- GG: $\langle \gamma_a, \gamma_b \rangle$: gravitational shear: OK!
- II: $\langle e_a^s, e_b^s \rangle$
  - Caused by physically close, tidally aligned galaxies.
  - Dominant at low $z \sim 0.1$, $\lesssim 10\%$ for deep cosmic shear surveys and large scales
  - Can be removed using redshift information (avoid auto-correlation)
  - Possibly creates B-modes

Intrinsic alignments

e.g. Heymans & Heavens 2003; King & Schneider 2003; Heymans et al. 2004; Takada & White 2004; Hirata & Seljak 2004; Mandelbaum et al. 2006; Bridle & King 2007; Hirata et al. 2007; Kitching et al. 2008; Zhang 2008; Brainerd et al. 2009, astro-ph/0904.3095 + references

Correlate galaxy ellipticities with separation θ, $z_a \leq z_b$:
$\langle e_a, e_b \rangle = \langle (e_a^s + \gamma_a), (e_b^s + \gamma_b) \rangle = \langle e_a^s, e_b^s \rangle + \langle e_a^s, \gamma_b \rangle + \langle \gamma_a, e_b^s \rangle + \langle \gamma_a, \gamma_b \rangle$

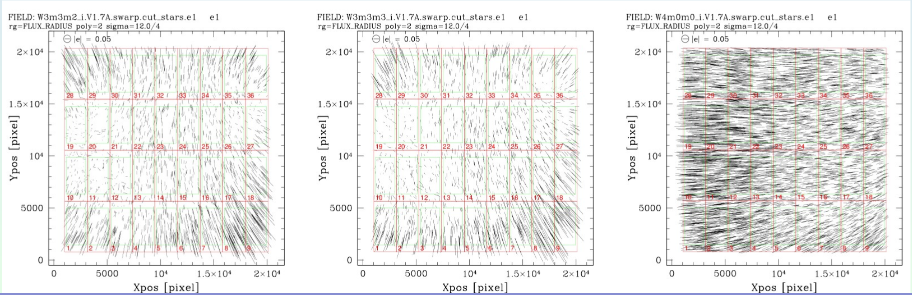
- GI: $\langle e_a^s, \gamma_b \rangle$
  - Alignment of foreground galaxy ellipticity with it's surrounding tidal gravitational field.
  - Causes systematic under-estimation of the shear of order -6% (Hirata et al. 2007)
  - Important to quantify for CS redshift range, especially blue galaxies
  - Different redshift dependence than CS $\Rightarrow$ can be removed with accurate photo-zs ("Nulling", Joachimi & Schneider 2008)
- $\langle \gamma_a, e_b^s \rangle = 0$ as far as we can see...

PSF interpolation

- Always trade-off: capture small-scale features but avoid over-fitting
- For WFI-like data: polynomial: 4^{th}- or 5^{th}-order for full mosaic or 2^{nd}-order chipwise
- If fast variations at chip-edges: maybe rational functions (Hoekstra 2004)
- Good test: determine PSF model with random star subset and check correction with other stars
- Check remaining small-scale power: $\langle e^*, e^* \rangle$
- Test for overfitting: Rowe 2009
- For large survey: PCA

Motivation

- **Wide-field cameras:** Chips contain 30 – 80 stars
  - $\sim$ 2nd-order polynomials: miss small-scale variation
  - Higher-order: too poorly constrained, esp. near chip boundaries
- **HST/ACS:** Total 10 – 20 stars:
  - Too few for any direct polynomial interpolation
- But: Often exposures have similar PSF properties:



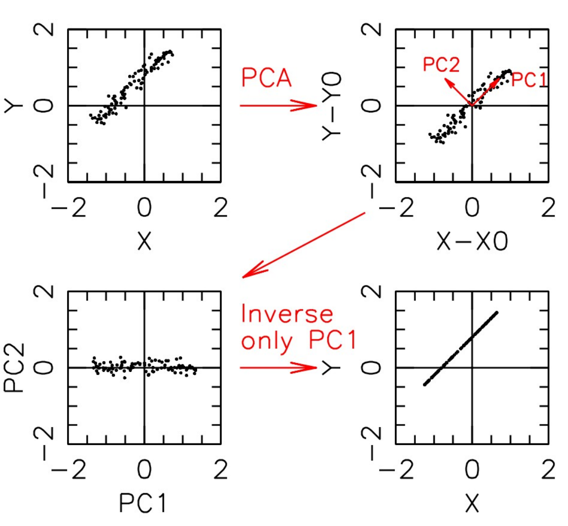
Aim of the PCA:

Make proper use of the “similarity” to combine information from many exposures (each with stars at different positions) in order to obtain PSF models with better spatial interpolation. The similarity is caused by similar physical effects influencing the PSF.

PCA in general: What is a Principal Component Analysis?

also known as Karhunen-Loève transform

Find a **linear coordinate transformation** to an orthogonal system which describes (most of) the data variation with fewer parameters.



PCA in general: A bit of algebra

- N data vectors $\mathbf{d_j}$ in M-dim. parameter space, components d_{ij}
- $a_{ij} = \frac{d_{ij}-m_i}{\sigma_i}$, $m_i = \frac{1}{N}\sum_{j=1}^{j=N} d_{ij}$, $\sigma_i = \sqrt{\frac{1}{N}\sum_{j=1}^{j=N}\left(d_{ij}-m_i\right)^2}$
- Arrange in a $M \times N$ dim. data matrix: $\mathbf{A} = \{\mathbf{a_1}, ..., \mathbf{a_j}, ..., \mathbf{a_N}\}$
- Singular value decomposition: $\mathbf{A} = \mathbf{W\Sigma V}^T$
  - $\mathbf{W}$ orthonormal, consists of the **singular vectors** (SVec) of $\mathbf{A}$
  - $\mathbf{\Sigma} = \{\mathbf{s_{kk}}\}$ **diagonal**, contains the **ordered singular values** (SVal)
  - kth largest SVal $\Rightarrow$ kth SVec = kth **principal component** (PC)
- Correlation matrix $\mathbf{C}$ diagonal in the coordinates spanned by the SVecs: $\mathbf{C} = \mathbf{AA}^T = \mathbf{W\Sigma V}^T\mathbf{V\Sigma}^T\mathbf{W}^T = \mathbf{W\Sigma\Sigma}^T\mathbf{W}^T = \mathbf{W\Lambda W}^T$
  - SVecs of $\mathbf{A}$ are **eigenvectors** of $\mathbf{C}$
  - **Eigenvalue** $\lambda_k = s_{kk}^2$ = **Variance** of $\mathbf{a_j}$ along kth PC
- Data point coefficients in space spanned by SVecs: $\mathbf{B} = \mathbf{W}^T\mathbf{A}$, with components b_{kj}

PCA in general: A bit of algebra

- N data vectors $\mathbf{d_j}$ in M-dim. parameter space, components d_{ij}
- $a_{ij} = \frac{d_{ij} - m_i}{\sigma_i}$, $m_i = \frac{1}{N} \sum_{j=1}^{j=N} d_{ij}$, $\sigma_i = \sqrt{\frac{1}{N} \sum_{j=1}^{j=N} (d_{ij} - m_i)^2}$
- Arrange in a $M \times N$ dim. data matrix: $\mathbf{A} = \{\mathbf{a_1}, ..., \mathbf{a_j}, ..., \mathbf{a_N}\}$
- Singular value decomposition: $\mathbf{A} = \mathbf{W \Sigma V}^T$
  - $\mathbf{W}$ orthonormal, consists of the **singular vectors** (SVec) of $\mathbf{A}$
  - $\mathbf{\Sigma} = \{\mathbf{s_{kk}}\}$ **diagonal**, contains the **ordered singular values** (SVal)
  - kth largest SVal $\Rightarrow$ kth SVec = kth **principal component** (PC)
- Correlation matrix $\mathbf{C}$ diagonal in the coordinates spanned by the SVecs: $\mathbf{C} = \mathbf{AA}^T = \mathbf{W\Sigma V}^T\mathbf{V\Sigma}^T\mathbf{W}^T = \mathbf{W\Sigma\Sigma}^T\mathbf{W}^T = \mathbf{W\Lambda W}^T$
  - SVecs of $\mathbf{A}$ are **eigenvectors** of $\mathbf{C}$
  - **Eigenvalue** $\lambda_k = s_{kk}^2$ = **Variance** of $\mathbf{a_j}$ along kth PC
- Data point coefficients in space spanned by SVecs: $\mathbf{B} = \mathbf{W}^T\mathbf{A}$, with components b_{kj}

PCA in general: A bit of algebra

- N data vectors $\mathbf{d_j}$ in M-dim. parameter space, components d_{ij}
- $a_{ij} = \frac{d_{ij}-m_i}{\sigma_i}$, $m_i = \frac{1}{N}\sum_{j=1}^{j=N} d_{ij}$, $\sigma_i = \sqrt{\frac{1}{N}\sum_{j=1}^{j=N}(d_{ij}-m_i)^2}$
- Arrange in a $M \times N$ dim. data matrix: $\mathbf{A} = \{\mathbf{a_1}, ..., \mathbf{a_j}, ..., \mathbf{a_N}\}$
- Singular value decomposition: $\mathbf{A} = \mathbf{W\Sigma V}^T$
  - $\mathbf{W}$ orthonormal, consists of the **singular vectors** (SVec) of $\mathbf{A}$
  - $\mathbf{\Sigma} = \{\mathbf{s_{kk}}\}$ **diagonal**, contains the **ordered singular values** (SVal)
  - kth largest SVal $\Rightarrow$ kth SVec = kth **principal component** (PC)
- Correlation matrix $\mathbf{C}$ diagonal in the coordinates spanned by the SVecs: $\mathbf{C} = \mathbf{AA}^T = \mathbf{W\Sigma V}^T\mathbf{V\Sigma}^T\mathbf{W}^T = \mathbf{W\Sigma\Sigma}^T\mathbf{W}^T = \mathbf{W\Lambda W}^T$
  - SVecs of $\mathbf{A}$ are **eigenvectors** of $\mathbf{C}$
  - **Eigenvalue** $\lambda_k = s_{kk}^2$ = **Variance** of $\mathbf{a_j}$ along kth PC
- Data point coefficients in space spanned by SVecs: $\mathbf{B} = \mathbf{W}^T\mathbf{A}$, with components b_{kj}

PCA in general: A bit of algebra

- N data vectors $\mathbf{d_j}$ in M-dim. parameter space, components d_{ij}
- $a_{ij} = \frac{d_{ij} - m_i}{\sigma_i}$, $m_i = \frac{1}{N}\sum_{j=1}^{j=N} d_{ij}$, $\sigma_i = \sqrt{\frac{1}{N}\sum_{j=1}^{j=N}(d_{ij} - m_i)^2}$
- Arrange in a $M \times N$ dim. data matrix: $\mathbf{A} = \{\mathbf{a_1}, ..., \mathbf{a_j}, ..., \mathbf{a_N}\}$
- Singular value decomposition: $\mathbf{A} = \mathbf{W\Sigma V}^T$
  - $\mathbf{W}$ orthonormal, consists of the **singular vectors** (SVec) of $\mathbf{A}$
  - $\mathbf{\Sigma} = \{\mathbf{s_{kk}}\}$ **diagonal**, contains the **ordered singular values** (SVal)
  - kth largest SVal $\Rightarrow$ kth SVec $=$ kth **principal component** (PC)
- Correlation matrix $\mathbf{C}$ diagonal in the coordinates spanned by the SVecs: $\mathbf{C} = \mathbf{AA}^T = \mathbf{W\Sigma V}^T\mathbf{V\Sigma}^T\mathbf{W}^T = \mathbf{W\Sigma\Sigma}^T\mathbf{W}^T = \mathbf{W\Lambda W}^T$
  - SVecs of $\mathbf{A}$ are **eigenvectors** of $\mathbf{C}$
  - **Eigenvalue** $\lambda_k = s_{kk}^2$ = **Variance** of $\mathbf{a_j}$ along kth PC
- Data point coefficients in space spanned by SVecs: $\mathbf{B} = \mathbf{W}^T\mathbf{A}$, with components b_{kj}

PCA in general: A bit of algebra

- N data vectors $\mathbf{d_j}$ in M-dim. parameter space, components d_{ij}
- $a_{ij} = \frac{d_{ij} - m_i}{\sigma_i}$, $m_i = \frac{1}{N} \sum_{j=1}^{j=N} d_{ij}$, $\sigma_i = \sqrt{\frac{1}{N} \sum_{j=1}^{j=N} (d_{ij} - m_i)^2}$
- Arrange in a $M \times N$ dim. data matrix: $\mathbf{A} = \{\mathbf{a_1}, ..., \mathbf{a_j}, ..., \mathbf{a_N}\}$
- Singular value decomposition: $\mathbf{A} = \mathbf{W \Sigma V}^T$
  - $\mathbf{W}$ orthonormal, consists of the **singular vectors** (SVec) of $\mathbf{A}$
  - $\mathbf{\Sigma} = \{\mathbf{s_{kk}}\}$ **diagonal**, contains the **ordered singular values** (SVal)
  - kth largest SVal $\Rightarrow$ kth SVec $=$ kth **principal component** (PC)
- Correlation matrix $\mathbf{C}$ diagonal in the coordinates spanned by the SVecs: $\mathbf{C} = \mathbf{AA}^T = \mathbf{W\Sigma V}^T\mathbf{V\Sigma}^T\mathbf{W}^T = \mathbf{W\Sigma\Sigma}^T\mathbf{W}^T = \mathbf{W\Lambda W}^T$
  - SVecs of $\mathbf{A}$ are **eigenvectors** of $\mathbf{C}$
  - **Eigenvalue** $\lambda_k = s_{kk}^2 =$ **Variance** of $\mathbf{a_j}$ along kth PC
- Data point coefficients in space spanned by SVecs: $\mathbf{B} = \mathbf{W}^T\mathbf{A}$, with components b_{kj}

PCA in general: A bit of algebra

- N data vectors $\mathbf{d_j}$ in M-dim. parameter space, components d_{ij}
- $a_{ij} = \frac{d_{ij} - m_i}{\sigma_i}$, $m_i = \frac{1}{N} \sum_{j=1}^{j=N} d_{ij}$, $\sigma_i = \sqrt{\frac{1}{N} \sum_{j=1}^{j=N} (d_{ij} - m_i)^2}$
- Arrange in a $M \times N$ dim. data matrix: $\mathbf{A} = \{\mathbf{a_1}, ..., \mathbf{a_j}, ..., \mathbf{a_N}\}$
- Singular value decomposition: $\mathbf{A} = \mathbf{W \Sigma V}^T$
  - $\mathbf{W}$ orthonormal, consists of the **singular vectors** (SVec) of $\mathbf{A}$
  - $\mathbf{\Sigma} = \{\mathbf{s_{kk}}\}$ **diagonal**, contains the **ordered singular values** (SVal)
  - kth largest SVal $\Rightarrow$ kth SVec = kth **principal component** (PC)
- Correlation matrix $\mathbf{C}$ diagonal in the coordinates spanned by the SVecs: $\mathbf{C} = \mathbf{A A}^T = \mathbf{W \Sigma V}^T \mathbf{V \Sigma}^T \mathbf{W}^T = \mathbf{W \Sigma \Sigma}^T \mathbf{W}^T = \mathbf{W \Lambda W}^T$
  - SVecs of $\mathbf{A}$ are **eigenvectors** of $\mathbf{C}$
  - **Eigenvalue** $\lambda_k = s_{kk}^2$ = **Variance** of $\mathbf{a_j}$ along kth PC
- Data point coefficients in space spanned by SVecs: $\mathbf{B} = \mathbf{W}^T \mathbf{A}$, with components b_{kj}

Using PCA for PSF interpolation

adapted from Jarvis & Jain 2004

Input (separately for N exposures):

- **Star catalog for PCA "similarity" analysis:**
  $X_{\rm mos}$ $Y_{\rm mos}$ chip mag Q_1 Q_2 ... Q_u
  - e.g. $Q_i : e_1, e_2, r_{\rm h}$
- **Star catalog for lensing PSF model:**
  $X'_{\rm mos}$ $Y'_{\rm mos}$ chip mag Q'_1 Q'_2 ... Q'_v
  - e.g. $Q'_i : q_1(r_{g1}), q_2(r_{g1}), {\rm Tr}(P_{\rm sh})/{\rm Tr}(P_{\rm sm})(r_{g1}), q_1(r_{g2}), ...$
  - OR: Shapelet coefficients
  - OR: Normalised pixel values
- **Galaxy catalog:** $X'_{\rm mos}$ $Y'_{\rm mos}$ chip

PCA PSF interpolation steps (Wide-field data):

1. **Large-scale "similarity" fit**
   - Perform a $3^{\mathrm{rd}} - 5^{\mathrm{th}}$-order polynomial fit for each Q_l **for all chips together**, but separate exposures $\Rightarrow$ Describe global behaviour
2. **PCA**
   - For each exposure form **combined data vector $\mathbf{d_j}$ from all fit coefficients**, compute $\mathbf{a_j}$ with $a_{ij} = \frac{d_{ij} - m_i}{\sigma_i}$, and arrange in matrix $\mathbf{A}$
   - Perform a PCA analysis for $\mathbf{A}$ and store b_{kj}
3. **All-scales fit to obtain detailed lensing PSF model**
   - **For each chip** combine the stars **of all exposures** and fit Q'_l with $Q'^{\mathrm{fit}}_{l,\mathrm{chip}}(X'_{\mathrm{mos}}, Y'_{\mathrm{mos}}, j) = \sum_{k=0}^{k_{\max}} b_{kj} P^{(n)}_{k,l,\mathrm{chip}}(X'_{\mathrm{mos}}, Y'_{\mathrm{mos}})$
   - $P^{(n)}_{k,l,\mathrm{chip}}$: n^{th}-order polynomial (e.g. $n = 5$), $b_{0j} = 1$.
   - Choose $k_{\max} \ll N$ such that e.g. 99% of the large-scale PSF variation is taken into account.
   - Total: $(k_{\max} + 1)(n + 1)(n + 2)/2$ coef. per Q'_l+chip, $t \propto N_* k^2_{\max} n^4$
4. **Application to galaxies**

Practical advantages:

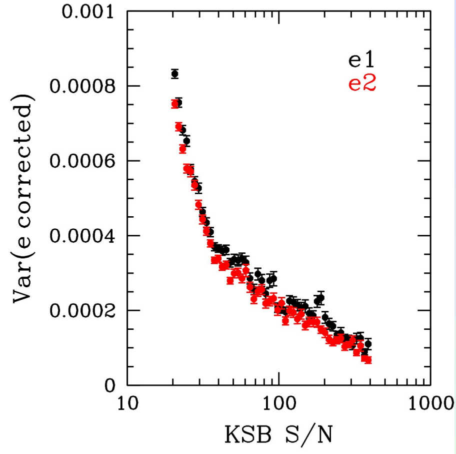
- Total number of stars in all exposures limits accuracy, not number in individual exposure/chip
- No problem at image boundaries, or near masks

Interpretation of principal components:

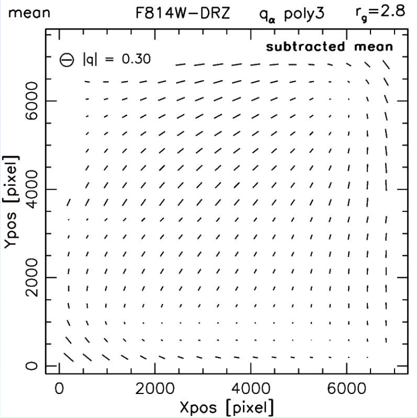
They will be related to the main physical effects affecting the PSF (focus, seeing, elevation, wind, etc.) in some arbitrary metric, but in the end we do not really care

Modifications for HST/ACS:

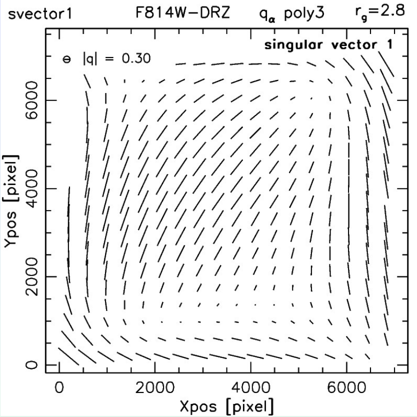
- In galaxy fields stars are too rare for the "similarity" polynomial fit
- Use stellar field exposures with high stellar density to derive the PC models
- Fit galaxy field stars to these components, possibly using the non-drizzled images (no extra-aliasing)
- Stars are rare: apply mag or S/N-dependent weighting



HST/ACS PCA analysis

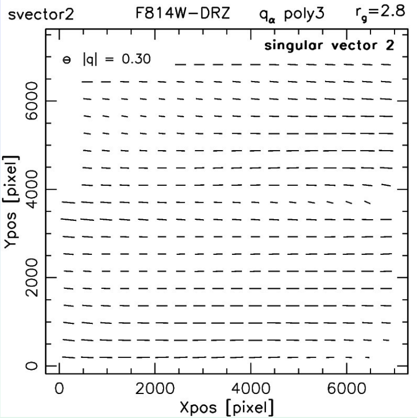


Subtracted mean PSF model (F814W).

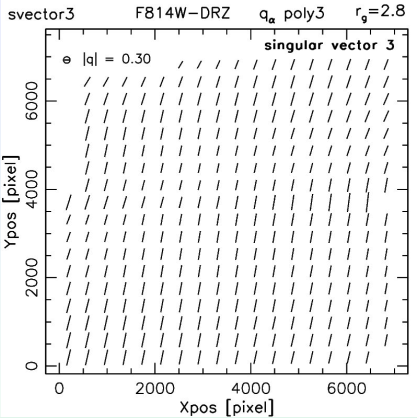


Singular vector of the 1st principal component (F814W).

HST/ACS PCA analysis

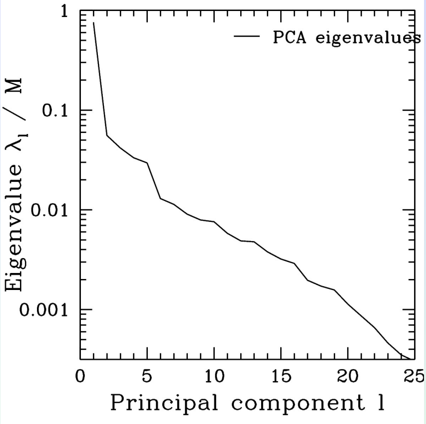


Singular vector of the 2nd principal component (F814W).

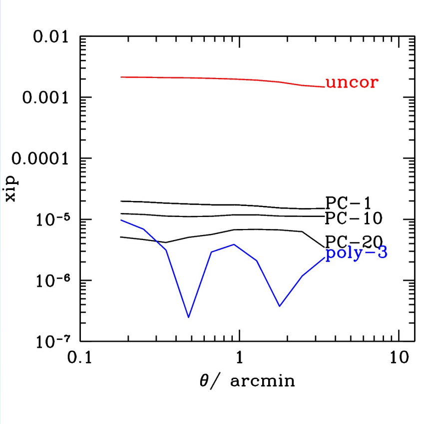


Singular vector of the 3rd principal component (F814W).

HST/ACS PCA analysis



PCA eigenvalues for 3rd-order polynomial fits in the non-drizzled frames considering $Q_i = e_1, e_2, r_\mathrm{h}$ (see also Jee et al. 2007).



Correlation ξ_+ from un-corrected, polynomial-corrected, and PCA-corrected stellar ellipticities measured in non-drizzled stellar field exposures.

Summary and Conclusions

- The PCA technique combines PSF information from many exposures.
- The final spatial resolution of the PSF interpolation is limited by the total number of stars in all fields together, not by the number in each exposure.
- In order to work properly, several 100 exposures are needed.
- For most HST/ACS weak lensing analyses a correction using only the first principal component (1-parameter fit) will be sufficient, moderate improvement can be obtained by fitting $\sim 5 - 10$ principal components.

Some ideas on how to prepare for a PhD in observational astronomy

- Keep your motivation
- Finish your studies well
- Be fluent in English
- Have a good background in statistics (also other maths can be helpful)
- Computer skills are very useful: Linux, C, Python, R, Bash, ...
- Get data analysis skills people need, especially to reduce different kinds of data (imaging, spectroscopy), e.g. with Theli, Multidrizzle, ESO pipelines
- Learn to use data archives (ESO, STScI, SUBARU, CADC, SDSS, ...)
- Try to have a broad background in astronomy
- Find a good masters project you find interesting
- Start reading professional research papers in the field you are interested
- Science is largely about 1. having good ideas and 2. skills to solve all the technical problems efficiently
- Many astronomers stay in the field of their PhD: if you can choose, choose wisely