Lecture 3: Ensemble photometry and transit detection

Ensemble differential photometry
Systematics and decorrelation methods
Transit detection by box least-squares
Noise and completeness
Detrending – the PyKE way

http://keplerscience.arc.nasa.gov/PyKE.shtml
Kepler SAP and PDC-SAP

Simple Aperture Photometry (SAP)

Pre-search Data Conditioning (PDC)

http://keplerscience.arc.nasa.gov/DataAnalysisProducts.shtml
WASP data reduction pipeline

1. Raw
2. Flatfield
3. Bias
4. Dark Current
5. Exposures Map
6. Pre-processed

WASP data reduction pipeline:

- Field recognition, astrometry, aperture photometry, calibration/de-trending

Star 561 P = 1.57155 Δt = -2662.9

- WASP V magnitude
- RMS scatter (mag)

16h43+31d26 field 2004 May-Aug

Vietri Sul Mare 2015
Catalogue-driven photometry

• **Step 1: Object detection**
  - e.g. (S)EXTRACTOR
  - Catalogue all objects on frame to detection threshold

• **Step 2: Establish astrometric solution**
  - Many standard algorithms available, e.g. astrometry.net
  - Cross-match with astrometric catalogue
  - Write world coordinate system to image.

• **Step 3: Aperture photometry**
  - Optimise aperture – many ways to do this!
  - Use frame catalogue to exclude faint objects in sky annulus
Ensemble differential photometry

- **Mean magnitude of star j:**
  \[ \hat{m}_j = \frac{\sum_i m_{ij} w_{ij}}{\sum_i w_{ij}} \]

- **Zero-point correction for frame i:**
  \[ \hat{z}_i = \frac{\sum_j (m_{ij} - \hat{m}_j) u_{ij}}{\sum_j u_{ij}} \]

  
  \[ w_{ij} = \frac{1}{\sigma_{ij}^2 + \sigma_{t(i)}^2} \]

  Additional intra-frame variance (downgrades poor images)

  \[ u_{ij} = \frac{1}{\sigma_{ij}^2 + \sigma_{s(j)}^2} \]

  Additional stellar variance (downgrades variable objects)
Frame quality and stellar variability I

- Define data vector $X$ and model $\mu$
  $$\mathbf{X} = \{m_{ij}, i = 1...n\} \quad \mu = \{\hat{m}_j + \hat{z}_i, i = 1...n\}$$

- Assuming Gaussian errors:
  $$P(X_i|\mu_i) = \frac{1}{\sqrt{2\pi}\sqrt{\sigma_{ij}^2 + \sigma_{t(i)}^2 + \sigma_{s(j)}^2}}$$
  $$\times \exp \left\{ -\frac{(m_{ij} - \hat{m}_j - \hat{z}_i)^2}{2\left[\sigma_{ij}^2 + \sigma_{t(i)}^2 + \sigma_{s(j)}^2\right]} \right\}$$

- Likelihood of entire data vector for star $j$:
  $$L(\mu) = (2\pi)^{-n/2} \prod_i \frac{1}{\sqrt{\sigma_{ij}^2 + \sigma_{t(i)}^2 + \sigma_{s(j)}^2}} \exp \left( -\frac{1}{2} \chi^2 \right)$$

- Where:
  $$\chi^2 = \sum_i \frac{(m_{ij} - \hat{m}_j - \hat{z}_i)^2}{\sigma_{ij}^2 + \sigma_{t(i)}^2 + \sigma_{s(j)}^2}$$
Frame quality and stellar variability II

• Maximise likelihood w.r.t. $\sigma^2_{s(j)}$ and $\sigma^2_{t(i)}$:

  – Solve iteratively for $\sigma^2_{s(j)}$ holding $\sigma^2_{t(i)}$ constant:

  $$
  \sum_{i} \frac{1}{\sigma_{ij}^2 + \sigma_{t(i)}^2 + \sigma_{s(j)}^2} - \sum_{i} \frac{(m_{ij} - \hat{m}_j - \hat{z}_i)^2}{\left[\sigma_{ij}^2 + \sigma_{t(i)}^2 + \sigma_{s(j)}^2\right]^2} = 0
  $$

  – Solve iteratively for $\sigma^2_{t(i)}$ holding $\sigma^2_{s(j)}$ constant:

  $$
  \sum_{j} \frac{1}{\sigma_{ij}^2 + \sigma_{t(i)}^2 + \sigma_{s(j)}^2} - \sum_{j} \frac{(m_{ij} - \hat{m}_j - \hat{z}_i)^2}{\left[\sigma_{ij}^2 + \sigma_{t(i)}^2 + \sigma_{s(j)}^2\right]^2} = 0
  $$

• Iterate $\hat{m}_j$, $\hat{z}_i$, $\sigma^2_{s(j)}$ and $\sigma^2_{t(i)}$ to convergence.
Decorrelation/ systematics removal

Tamuz et al 2005

16h30+28 field
300 stars
2549 observations
100 days
Systematics and correlated noise

• References:
SysRem

• Many patterns of systematics are common to all stars
  – Secondary extinction: dependence on stellar colour
  – Sky brightness: Dependence on target brightness
  – Ambient temperature and focus drift: position dependent?

• SysRem: PCA with error bars!
  – Construct common temporal basis functions
  – Compute optimal scaling coefficient per star
  – Remove and repeat.
SysRem algorithm

- SysRem produces a corrected magnitude given by:

$$\tilde{x}_{i,j} = x_{i,j} - \sum_{k=1}^{M} (k) c_j (k) a_i$$

- Search for the best $c_j$ that minimises:

$$ (k-1) S_i^2 = \sum_{i,j} (k-1) \tilde{x}_{i,j} - (k) c_i (k) a_j)^2 w_{i,j} $$

$$ (k) c_j = \sum_{i} \frac{(k-1) \tilde{x}_{i,j} (k-1) a_i w_{i,j}}{(k-1) a_i^2 w_{i,j}} $$

- Similarly

$$ (k) a_i = \sum_{j} \frac{(k-1) \tilde{x}_{i,j} (k-1) c_j w_{i,j}}{(k-1) c_j^2 w_{i,j}} $$
SysRem basis functions vs HA
Decorrelation/ systematics removal

Tamuz et al 2005

16h30+28 field
300 stars
2549 observations
100 days
Decorrelation/ systematics removal

Tamuz et al 2005

16h30+28 field
300 stars
2549 observations
100 days
Decorrelation & Detection

- **Decorrelation:**
  - 4-stage SysRem
  - TFA with reconstruction

- **Detection:** Cameron et al 2006 MNRAS 373, 799
  - Accelerated BLS algorithm
  - Coarse grid search
  - Newton-Raphson peak-up on 5 strongest peaks
  - Box width varies with period

- **Example:** WASP-1
Box Least-Squares transit search

- **References**
  - Collier Cameron et al 2006, MNRAS 373, 799

- **Step 1: Compute inverse variance-weighted mean flux, subtract and precompute \( \chi^2_0, t \):**
  
  \[
  \hat{x} = \frac{\sum_i \tilde{x}_i w_i}{\sum_i w_i}, \quad x_i = \tilde{x}_i - \hat{x} \]
  
  \[
  \chi^2_0 = \sum_i x_i^2 w_i, \quad t = \sum_i w_i.
  \]

- **Step 2: set up frequency grid**
  - phase of each observation must change by less than transit duration between adjacent frequencies

- **Step 3: estimate range of transit duration**
  - Use stellar density
Box Least-Squares transit search

- **Step 3: Phase-fold and partition:**

  - Mean light level in-transit (L):
    
    \[ s = \sum_{i \in \ell} x_i w_i, \quad r = \sum_{i \in \ell} w_i, \quad q = \sum_{i \in \ell} x_i^2 w_i. \]

    \[ L = \frac{s}{r}, \quad \text{Var}(L) = \frac{1}{r} \]

  - Mean light level out of transit (H):
    
    \[ H = \frac{-s}{t - r}, \quad \text{Var}(H) = \frac{1}{t - r} \]

  - Fitted transit depth (\( \delta \)):
    
    \[ \delta = L - H = \frac{st}{r(t - r)}, \quad \text{Var}(\delta) = \frac{t}{r(t - r)} \]
Box Least-Squares transit search

- **Step 4: Badness-of-fit:**
  - Signal-to-noise ratio of transit depth
    \[
    \frac{S}{N} = s \sqrt{\frac{t}{r(t - r)}}
    \]
  - Improvement in fit relative to constant model is \((S/N)^2\):
    \[
    \Delta \chi^2 = \frac{s^2 t}{r(t - r)}
    \]
  - Goodness of fit outside transit:
    \[
    \chi_h^2 = \chi_0^2 - \frac{s^2}{(t - r)} - q
    \]
Box Least-Squares transit search

- Step 5: BLS periodogram
Box Least-Squares transit search

- Step 5: BLS periodogram
Planet search

- **SysRem+BLS**
  - Tamuz, Mazeh & Zucker 2005
  - Kovacs, Zucker & Mazeh 2002

- Skype-linked network of human neuro-visual processors.

- Each processor allocated several thousand candidate light curves + associated statistics from main database.

- Motivated by promises of food, postdoc employment, fame, glory, etc.
Systematics and red noise

- **Systematics:**
  - Secondary extinction
  - Temperature-dependent focus
  - Sky brightness-dependent bias in background subtraction
  - SysRem: Tamuz et al 2005
  - TFA: Kovacs et al 2005

- **Red noise:**
  - Pont et al 2006
  - Smith et al 2006
Red noise and detection threshold

Smith, A M S et al, 2006
80 nights

Smith, A M S et al, 2006
130 nights
Red noise and planet catch

Season 1

Season 2
Summary

- Systematics and astrophysical variability mask transits
- Not all systematics are understood
- PCA-like methods (e.g. SysRem, PDC-MAP) remove systematics while preserving astrophysical variability
- More aggressive detrending (e.g. TFA) removes astrophysical variability too
- Box least-squares method is efficient
- Correlated noise must be taken into account when assessing detection thresholds.