Theory of Image Subtraction

Robert Lupton, Princeton University
LSST Pipeline/Calibration Scientist

Difference Image Analysis 1
2019-08-13

Measuring Variability

Given two images I_1 and I_2 with PSFs ϕ_1 and ϕ_2 how should I measure a source's light curve (*i.e.* the fluxes A_1 and A_2)?

Measuring Variability

Given two images I_1 and I_2 with PSFs ϕ_1 and ϕ_2 how should I measure a source's light curve (*i.e.* the fluxes A_1 and A_2)? The first thing to try is to measure the flux in each image and subtract. If I know the PSFs I can make an optimal measurement of each flux

$$A_r = \frac{\sum_i I_{r,i}\,\phi_{r,i}/\sigma^2_{r,i}}{\sum_i \phi^2_{r,i}/\sigma^2_{r,i}}$$

where $r = 1, 2$ and i runs over the pixels. For faint sources the noise is dominated by the sky noise, and we find

$$A_1 - A_2 = \frac{\sum_i I_{1,i}\,\phi_{1,i}}{\sum_i \phi^2_{1,i}} - \frac{\sum_i I_{2,i}\,\phi_{2,i}}{\sum_i \phi^2_{2,i}}$$

Measuring Variability

Given two images I_1 and I_2 with PSFs ϕ_1 and ϕ_2 how should I measure a source's light curve (*i.e.* the fluxes A_1 and A_2)? The first thing to try is to measure the flux in each image and subtract. If I know the PSFs I can make an optimal measurement of each flux

$$A_r = \frac{\sum_i I_{r,i}\,\phi_{r,i}/\sigma^2_{r,i}}{\sum_i \phi^2_{r,i}/\sigma^2_{r,i}}$$

where $r = 1, 2$ and i runs over the pixels. For faint sources the noise is dominated by the sky noise, and we find

$$A_1 - A_2 = \frac{\sum_i I_{1,i}\,\phi_{1,i}}{\sum_i \phi^2_{1,i}} - \frac{\sum_i I_{2,i}\,\phi_{2,i}}{\sum_i \phi^2_{2,i}}$$

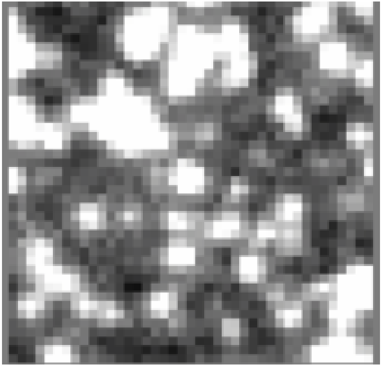
If the images are complicated (*e.g.* the Galactic centre) this measurement may not be very good.

M31

Takada Masahiro and Niikura Hiroko used HSC to image M31 every 2 minutes for an entire night

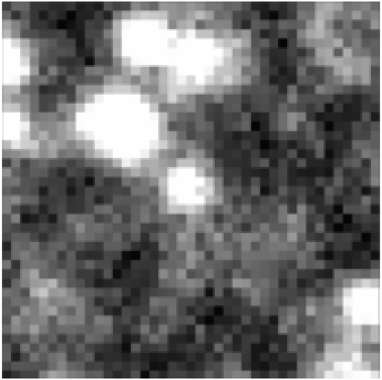
M31

Takada Masahiro and Niikura Hiroko used HSC to image M31 every 2 minutes for an entire night, *e.g.*



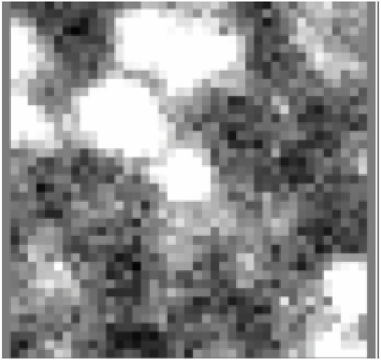
M31

Takada Masahiro and Niikura Hiroko used HSC to image M31 every 2 minutes for an entire night, *e.g.*



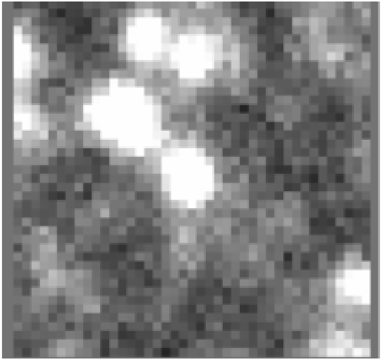
M31

Takada Masahiro and Niikura Hiroko used HSC to image M31 every 2 minutes for an entire night, *e.g.*



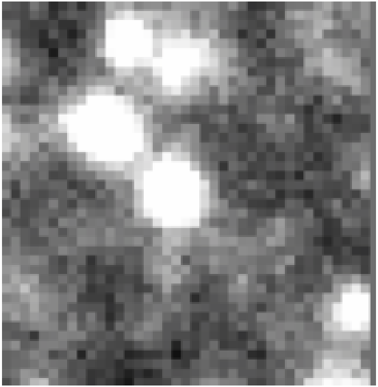
M31

Takada Masahiro and Niikura Hiroko used HSC to image M31 every 2 minutes for an entire night, *e.g.*



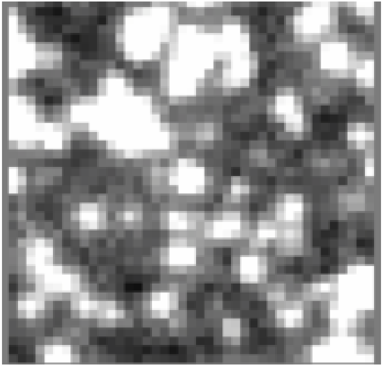
M31

Takada Masahiro and Niikura Hiroko used HSC to image M31 every 2 minutes for an entire night, *e.g.*



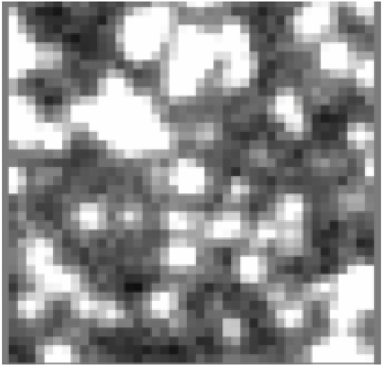
M31

Takada Masahiro and Niikura Hiroko used HSC to image M31 every 2 minutes for an entire night, *e.g.*



M31

Takada Masahiro and Niikura Hiroko used HSC to image M31 every 2 minutes for an entire night, *e.g.*



The "simple image" condition isn't satisfied very well.

Normalising the Seeing

One solution is to match the seeing; if

$$I'_2(k) = I_2(k)\frac{\phi_1(k)}{\phi_2(k)}$$

then

$$A_1 - A_2 = \frac{\sum_i I_{1,i}\phi_{1,i} - \sum_i I'_{2,i}\phi_{1,i}}{\sum_i \phi_{1,i}^2}$$

measures only the part that's changed; neighbouring stars should make the same contribution in the two images and cancelled out.

Normalising the Seeing

One solution is to match the seeing; if

$$I'_2(k) = I_2(k)\frac{\phi_1(k)}{\phi_2(k)}$$

then

$$A_1 - A_2 = \frac{\sum_i I_{1,i}\phi_{1,i} - \sum_i I'_{2,i}\phi_{1,i}}{\sum_i \phi_{1,i}^2}$$

measures only the part that's changed; neighbouring stars should make the same contribution in the two images and cancelled out.

This doesn't help much if you're looking for faint unknown objects that have varied.

Classical Image Subtraction

We can solve this problem by calculating $I_1 - I_2'$ directly to remove the non-variable sources:

$$A_1 - A_2 = \frac{\sum_i \left(I_{1,i} - I'_{2,i}\right) \phi_i}{\sum_i \phi_i^2}$$

Classical Image Subtraction

We can solve this problem by calculating $I_1 - I'_2$ directly to remove the non-variable sources:

$$A_1 - A_2 = \frac{\sum_i \left(I_{1,i} - I'_{2,i} \right) \phi_i}{\sum_i \phi_i^2}$$

Many people (*e.g.* Ciardullo, Tamblyn, and Phillips 1990) have proposed calculating I'_2 naïvely in Fourier space, as I did on the previous slide.

Classical Image Subtraction

We can solve this problem by calculating $I_1 - I'_2$ directly to remove the non-variable sources:

$$A_1 - A_2 = \frac{\sum_i \left(I_{1,i} - I'_{2,i}\right) \phi_i}{\sum_i \phi_i^2}$$

Many people (*e.g.* Ciardullo, Tamblyn, and Phillips 1990) have proposed calculating I'_2 naïvely in Fourier space, as I did on the previous slide.

This turns out to be problematic as you need to know the PSFs' wings well.

Classical Image Subtraction

We can solve this problem by calculating $I_1 - I'_2$ directly to remove the non-variable sources:

$$A_1 - A_2 = \frac{\sum_i \left(I_{1,i} - I'_{2,i}\right) \phi_i}{\sum_i \phi_i^2}$$

Many people (*e.g.* Ciardullo, Tamblyn, and Phillips 1990) have proposed calculating I'_2 naïvely in Fourier space, as I did on the previous slide.

This turns out to be problematic as you need to know the PSFs' wings well. Problems are revealed in the residual image $I_1 - I'_2$.

Cross-Convolution

Another obvious approach (Gal-Yam) is to construct the difference image as

$$\phi_2 \otimes I_1 - \phi_1 \otimes I_2$$

but this sacrifices resolution, and still needs to know the PSF.

Alard and Lupton

Christophe Alard and I proposed a way to circumvent the need to know ϕ. If we write

$$I_2' = \kappa \otimes I_2$$

and expand

$$\kappa = \sum_r a_r B^r$$

we may minimise

$$\left| \frac{I_1 - \sum_r a_r \left(B^r \otimes I_2 \right)}{\sigma} \right|^2$$

by solving linear equations for a_r (and also for the difference in the sky level).

Alard and Lupton

Christophe Alard and I proposed a way to circumvent the need to know ϕ. If we write

$$I_2' = \kappa \otimes I_2$$

and expand

$$\kappa = \sum_r a_r B^r$$

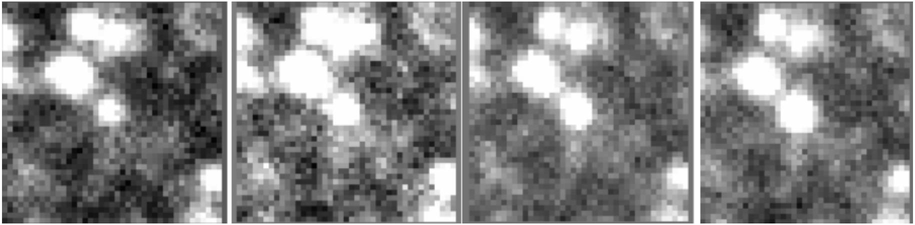
we may minimise

$$\left| \frac{I_1 - \sum_r a_r \left(B^r \otimes I_2 \right)}{\sigma} \right|^2$$

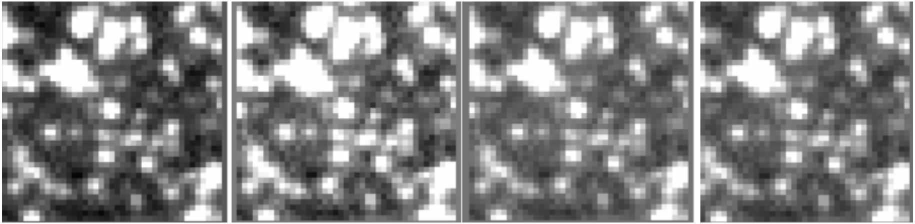
by solving linear equations for a_r (and also for the difference in the sky level).

The choice of B^r is arbitrary. We used Gauss-Hermite functions, but you can also use δ-functions (*i.e.* a pixel basis).

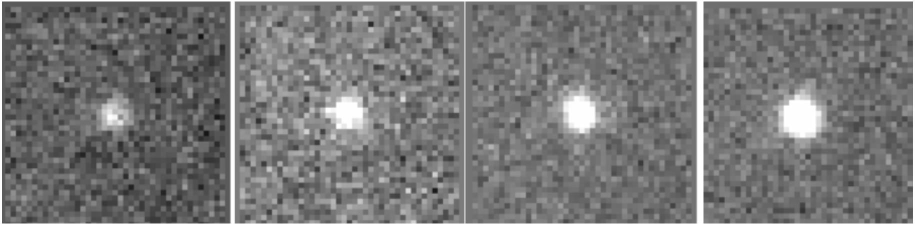
M31

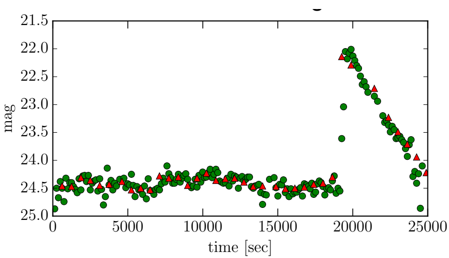


M31



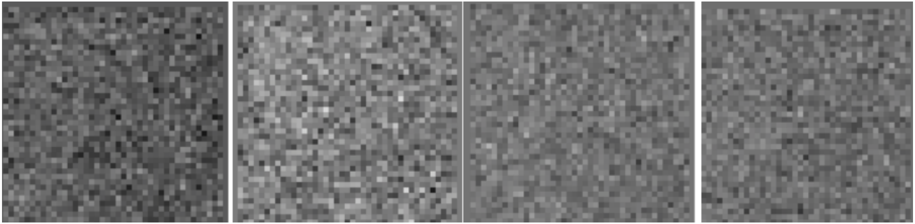
M31

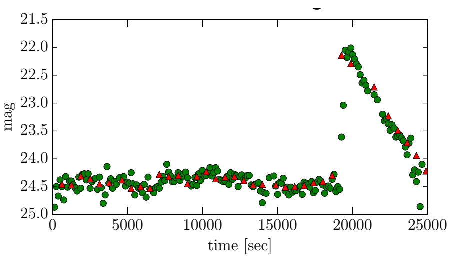




mag
time [sec]
21.5
22.0
22.5
23.0
23.5
24.0
24.5
25.0
0
5000
10000
15000
20000
25000

M31




21.5
22.0
22.5
23.0
23.5
24.0
24.5
25.0
mag
0
5000
10000
15000
20000
25000
time [sec]

Problems with Alard/Lupton

Problems with the A/L Algorithm

A&L is optimal for the problem it was designed to solve, but. . .

Problems with the A/L Algorithm

A&L is optimal for the problem it was designed to solve, but. . .

- Where does the template I_2 come from?

Problems with the A/L Algorithm

A&L is optimal for the problem it was designed to solve, but. . .

- Where does the template I_2 come from?
- What should we do if the template is sharper than the data?

Problems with the A/L Algorithm

A&L is optimal for the problem it was designed to solve, but. . .

- Where does the template I_2 come from?
- What should we do if the template is sharper than the data?
- What is the consequence of noise in the template?

Whence Came the Template?

We didn't attempt to solve this problem, but fortunately Nick Kaiser did 15 years ago.

Whence Came the Template?

We didn't attempt to solve this problem, but fortunately Nick Kaiser did 15 years ago. Unfortunately he didn't publish his result except as "PSDC-002-01[01]-00".

Whence Came the Template?

We didn’t attempt to solve this problem, but fortunately Nick Kaiser did 15 years ago. Unfortunately he didn’t publish his result except as "PSDC-002-01[01]-00".

If I have a set of n realisations of an image I_r with known PSFs ϕ_r, what is the best estimate for the true image above the atmosphere, T?

Whence Came the Template?

We know that

$$I_{r,i} = (T \otimes \phi_r)_i + \epsilon_{r,i}$$

and let's assume that ϵ_r is an $N(0, \sigma_r^2)$ variable (*i.e.* we only care about faint objects)

Whence Came the Template?

We know that

$$I_{r,i} = (T \otimes \phi_r)_i + \epsilon_{r,i}$$

and let's assume that ϵ_r is an $N(0, \sigma_r^2)$ variable (*i.e.* we only care about faint objects) We may estimate each Fourier mode independently using an ML estimator:

$$\ln \mathcal{L} \propto \sum_r \frac{(I_r(k) - T(k)\phi_r(k))^2}{\sigma_r^2}$$

i.e.

$$\hat{T}(k) = \frac{\sum_r I_r(k)\phi_r(k)/\sigma_r^2}{\sum_r \phi_r(k)^2/\sigma_r^2}$$

with variance

$$\mathrm{Var}(\hat{T}(k)) = \frac{1}{\sum_r \phi_r(k)^2/\sigma_r^2}$$

An Optimal Template

If we'd like a template with uncorrelated noise we need to flatten the noise, resulting in

$$\hat{T}'(k) = \frac{\sum_r I_r(k)\phi_r(k)/\sigma_r^2}{\sqrt{\sum_r \phi_r(k)^2/\sigma_r^2}\sqrt{\sum_r 1/\sigma_r^2}}$$

An Optimal Template

If we'd like a template with uncorrelated noise we need to flatten the noise, resulting in

$$\hat{T}'(k) = \frac{\sum_r I_r(k)\phi_r(k)/\sigma_r^2}{\sqrt{\sum_r \phi_r(k)^2/\sigma_r^2}\sqrt{\sum_r 1/\sigma_r^2}}$$

In reality templates are trickier:

An Optimal Template

If we'd like a template with uncorrelated noise we need to flatten the noise, resulting in

$$\hat{T}'(k) = \frac{\sum_r I_r(k)\phi_r(k)/\sigma_r^2}{\sqrt{\sum_r \phi_r(k)^2/\sigma_r^2}\sqrt{\sum_r 1/\sigma_r^2}}$$

In reality templates are trickier:

- There are too few exposures,

An Optimal Template

If we'd like a template with uncorrelated noise we need to flatten the noise, resulting in

$$\hat{T}'(k) = \frac{\sum_r I_r(k)\phi_r(k)/\sigma_r^2}{\sqrt{\sum_r \phi_r(k)^2/\sigma_r^2}\sqrt{\sum_r 1/\sigma_r^2}}$$

In reality templates are trickier:

- There are too few exposures,
- which are taken in poor seeing,

An Optimal Template

If we'd like a template with uncorrelated noise we need to flatten the noise, resulting in

$$\hat{T}'(k) = \frac{\sum_r I_r(k)\phi_r(k)/\sigma_r^2}{\sqrt{\sum_r \phi_r(k)^2/\sigma_r^2}\sqrt{\sum_r 1/\sigma_r^2}}$$

In reality templates are trickier:

- There are too few exposures,
- which are taken in poor seeing,
- at a range of air masses,

An Optimal Template

If we'd like a template with uncorrelated noise we need to flatten the noise, resulting in

$$\hat{T}'(k) = \frac{\sum_r I_r(k)\phi_r(k)/\sigma_r^2}{\sqrt{\sum_r \phi_r(k)^2/\sigma_r^2}\sqrt{\sum_r 1/\sigma_r^2}}$$

In reality templates are trickier:

- There are too few exposures,
- which are taken in poor seeing,
- at a range of air masses,
- contaminated by signal,

An Optimal Template

If we'd like a template with uncorrelated noise we need to flatten the noise, resulting in

$$\hat{T}'(k) = \frac{\sum_r I_r(k)\phi_r(k)/\sigma_r^2}{\sqrt{\sum_r \phi_r(k)^2/\sigma_r^2}\sqrt{\sum_r 1/\sigma_r^2}}$$

In reality templates are trickier:

- There are too few exposures,
- which are taken in poor seeing,
- at a range of air masses,
- contaminated by signal,
- and Balkanised by edges and defects.

An Optimal Template

If we'd like a template with uncorrelated noise we need to flatten the noise, resulting in

$$\hat{T}'(k) = \frac{\sum_r I_r(k)\phi_r(k)/\sigma_r^2}{\sqrt{\sum_r \phi_r(k)^2/\sigma_r^2}\sqrt{\sum_r 1/\sigma_r^2}}$$

In reality templates are trickier:

- There are too few exposures,
- which are taken in poor seeing,
- at a range of air masses,
- contaminated by signal,
- and Balkanised by edges and defects.

If your image is simple enough (*e.g.* a SNe and you know the PSF) you can simultaneously derive the template and light curve: "Scene Modelling" (Holtzman 2008, Guy 2010).

Noise in The Template

This one is harder for the naïve.

Noise in The Template

This one is harder for the naïve. If the template has noise, then $\kappa \otimes I_2$ will have correlated noise which complicates the analysis (for example, those optimal estimates of fluxes were written assuming diagonal covariance estimates). This is not fatal, but it is a nuisance.

Noise in The Template

This one is harder for the naïve. If the template has noise, then $\kappa \otimes I_2$ will have correlated noise which complicates the analysis (for example, those optimal estimates of fluxes were written assuming diagonal covariance estimates). This is not fatal, but it is a nuisance. I knew this.

Noise in The Template

This one is harder for the naïve. If the template has noise, then $\kappa \otimes I_2$ will have correlated noise which complicates the analysis (for example, those optimal estimates of fluxes were written assuming diagonal covariance estimates). This is not fatal, but it is a nuisance. I knew this.

More interestingly if the template is noisy, A&L is no longer optimal.

Noise in The Template

This one is harder for the naïve. If the template has noise, then $\kappa \otimes I_2$ will have correlated noise which complicates the analysis (for example, those optimal estimates of fluxes were written assuming diagonal covariance estimates). This is not fatal, but it is a nuisance. I knew this.

More interestingly if the template is noisy, A&L is no longer optimal. I hadn't realised this until reading a paper by Barak Zackay, Eran Ofek and Avishay Gal-Yam.

Subtracting Two Noisy Images

Let us adopt an A&L approach and write

$$D = I_1 - \kappa \otimes I_2$$

with the Gaussian homoschedastic (faint-object) assumption.

Subtracting Two Noisy Images

Let us adopt an A&L approach and write

$$D = I_1 - \kappa \otimes I_2$$

with the Gaussian homoschedastic (faint-object) assumption. Our model is that the difference image is D convolved with the PSF ϕ_1, so taking a Fourier transform and constructing the log-likelihood gives

$$\ln \mathcal{L} \sim \sum_k \frac{(I_1(k) - \kappa(k) I_2(k) - D(k)\phi_1(k))^2}{\sigma_1^2 + \kappa^2(k)\sigma_2^2}$$

Subtracting Two Noisy Images

The MLE for $D(k)$ is

$$\hat{D}(k) = \frac{I_1(k) - \kappa(k) I_2(k)}{\phi_1(k)}$$

with variance

$$\mathrm{Var}(\hat{D}(k)) = \frac{\sigma_1^2 + \kappa^2(k)\sigma_2^2}{\phi_1^2(k)}$$

Subtracting Two Noisy Images

That variance diverges at large k – not surprising, as we're estimating a deconvolved scene D. As in Kaiser's analysis (and as emphasised by Zackay *et al.*) we can construct an uncorrelated image by whitening the noise, resulting in

$$\hat{D}(k) = (I_1(k) - \kappa(k) I_2(k)) \sqrt{\frac{\sigma_1^2 + \sigma_2^2}{\sigma_1^2 + \kappa^2(k)\sigma_2^2}}$$

(no ϕ! Just the kernel κ) with PSF

$$\phi_D(k) = \phi_1(k) \sqrt{\frac{\sigma_1^2 + \sigma_2^2}{\sigma_1^2 + \kappa^2(k)\sigma_2^2}}$$

Subtracting Two Noisy Images

That variance diverges at large k – not surprising, as we're estimating a deconvolved scene D. As in Kaiser's analysis (and as emphasised by Zackay *et al.*) we can construct an uncorrelated image by whitening the noise, resulting in

$$\hat{D}(k) = (I_1(k) - \kappa(k) I_2(k)) \sqrt{\frac{\sigma_1^2 + \sigma_2^2}{\sigma_1^2 + \kappa^2(k)\sigma_2^2}}$$

(no ϕ! Just the kernel κ) with PSF

$$\phi_D(k) = \phi_1(k) \sqrt{\frac{\sigma_1^2 + \sigma_2^2}{\sigma_1^2 + \kappa^2(k)\sigma_2^2}}$$

i.e. we can estimate κ by standard methods, and then correct it for the noise in the template. You might need to iterate.

‘Proper Image Subtraction’

Zackay *et al.* carry out what amounts to this calculation, assuming that ϕ_1 and ϕ_2 are known and that therefore $\kappa(k) = \phi_1(k)/\phi_2(k)$. If we substitute this into $\hat{D}$ and ϕ_D we find

$$D(k) = (\phi_2(k)I_1(k) - \phi_1(k)I_2(k))\sqrt{\frac{\sigma_1^2 + \sigma_2^2}{\sigma_1^2\phi_2^2(k) + \sigma_2^2\phi_1^2(k)}}$$

$$\phi_D(k) = \phi_1(k)\phi_2(k)\sqrt{\frac{\sigma_1^2 + \sigma_2^2}{\sigma_1^2\phi_2^2(k) + \sigma_2^2\phi_1^2(k)}}$$

which are Zackay *et al.*’s equations 13 and 14, except that I’ve multiplied D by $(\sigma_1^2 + \sigma_2^2)^{1/2}$

‘Proper Image Subtraction’

Zackay *et al.* carry out what amounts to this calculation, assuming that ϕ_1 and ϕ_2 are known and that therefore $\kappa(k) = \phi_1(k)/\phi_2(k)$. If we substitute this into $\hat{D}$ and ϕ_D we find

$$D(k) = (\phi_2(k) I_1(k) - \phi_1(k) I_2(k)) \sqrt{\frac{\sigma_1^2 + \sigma_2^2}{\sigma_1^2 \phi_2^2(k) + \sigma_2^2 \phi_1^2(k)}}$$

$$\phi_D(k) = \phi_1(k)\phi_2(k) \sqrt{\frac{\sigma_1^2 + \sigma_2^2}{\sigma_1^2 \phi_2^2(k) + \sigma_2^2 \phi_1^2(k)}}$$

which are Zackay *et al.*’s equations 13 and 14, except that I’ve multiplied D by $(\sigma_1^2 + \sigma_2^2)^{1/2}$

One interesting feature of these equations is that they are symmetric in I_1 and I_2 and are thus able to handle better seeing in the science image than in the template.

Sensitivity To Noise Levels

If the template is noise free ($\sigma_2 = 0$), we recover

$$\hat{D}(k) = I_1(k) - \kappa(k) I_2(k)$$
$$\phi_D(k) = \phi_1(k)$$

which are just the standard equations for difference imaging.

Sensitivity To Noise Levels

If the template is noise free ($\sigma_2 = 0$), we recover

$$\hat{D}(k) = I_1(k) - \kappa(k)I_2(k)$$
$$\phi_D(k) = \phi_1(k)$$

which are just the standard equations for difference imaging. Numerically, once the S/N in the template is more than *c.* twice the science image the results are similar to the noise-free case