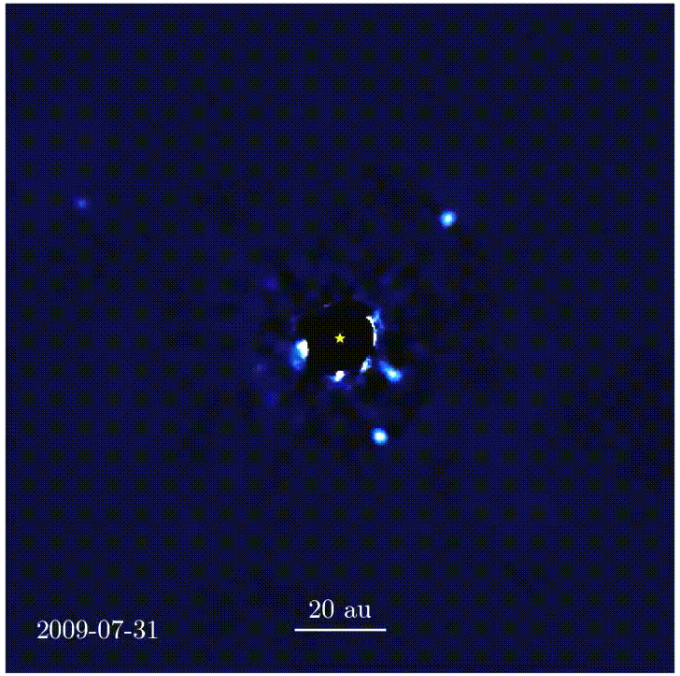
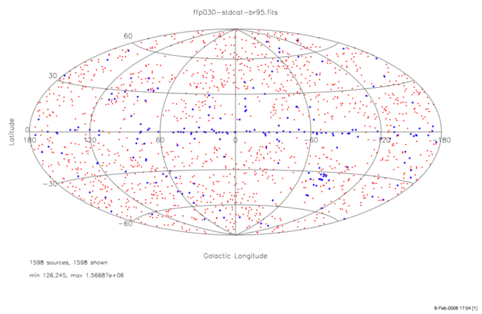
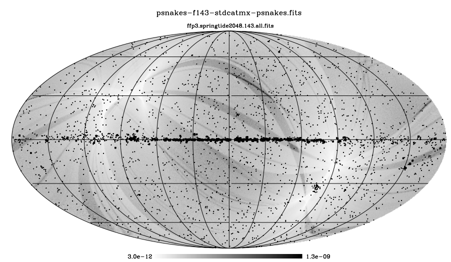
Bayesian High Contrast Imaging Algorithms

Graca Rocha,
Dimitri Mawet,
Bertrand Mennesson,
Tiffany Meshkat,
Gautam Vasisht
Michael Bottom
Jeff Jewell

JPL, Caltech & IPAC







ESI symposium, JPL, 26th of March 2018

Motivation

- **Motivation:**
  - To help enhance the detectability of faint exoplanets at small orbital separations from the host star
    - Both ground-based and space-based instruments have not yet achieved the contrast gain needed to detect mature planets with masses lower than 1 Jupiter mass at separations smaller than 0.5”.
    - The difficulty arises from the residual glare of starlight at small orbital separations due to diffraction, scattered light, and speckles caused by defects in the optical system
  - New approach -> unify source detection and characterization (Position, Flux or Intensity and hence accurate spectrum extraction) into one single rigorous mathematical framework, the Bayesian framework, enabling an adequate hypothesis testing given the S/N of the data.
  - The method will be applied in combination with other post-processing techniques (best suited for this approach), for example KLIP, but now recast in a Bayesian perspective.

- To extend PowellSakes, PwS, a Bayesian approach, to direct imaging data analysis
  - PwS has successfully been applied to detect compact sources immersed in a diffuse background in Planck maps - *Carvalho, Rocha & Hobson, MNRAS, 393, 681C, 2009; Carvalho, Rocha, Hobson & Lasenby, MNRAS, 427, 2011; Bayesian Methods in Cosmology' – CUP, December 2009: chapter on 'Bayesian Source Extraction' by Hobson, Rocha & Savage; Planck catalog (of compact sources and SZ clusters) papers*

Bayesian – what does it really mean?

❑ What does it mean to recast the problem of planet detection into a Bayesian perspective?

A. The Bayesian framework entails defining the following key ingredients:

a data model + a Likelihood shape + model parameter priors

B. Next apply Bayes Theorem – to retrieve the distribution, pdf, of the model parameters:

Bayes Theorem $\longrightarrow$ Posterior distributions of the model + Best Fit models

Bayesian – what does it really mean?

A. The Bayesian framework entails defining the following key ingredients:

a data model + a Likelihood shape + model parameter priors

- **The problem:**
  - Suppose we want to extract the amplitude A of a signal with a known spatial distribution t(x) or τ(x) from a measured signal d (x) which is contaminated by noise n(x)
  - Start by defining your data model, for example:

Amplitude of the signal – what we want to know

Data Model →
$$\underbrace{d(x)=s(x)}_{\text{Signal}}+\underbrace{n(x)}_{\text{Noise}}=\overbrace{A}^{\text{Amplitude of the signal – what we want to know}}\underbrace{t(x)}_{\text{Signal spatial template}}+n(x)$$

Signal Noise Signal spatial template

Bayesian – what does it really mean?

B. Next apply Bayes Theorem – to retrieve the distribution, pdf, of the model parameters:

Bayes Theorem $\longrightarrow$ Posterior distributions of the model + Best Fit models

- Bayes' theorem states that:

Likelihood: L(Θ)

Prior: π(Θ)

$$\Pr(\Theta|d, H) = \frac{\Pr(d|\,\Theta, H)\,\Pr(\Theta|H)}{\Pr(d|H)},$$

Posterior probability distribution of the parameters: P(Θ)

Bayesian evidence: E

- In parameter estimation, the normalizing evidence factor is usually ignored, since it is independent of the parameters - This (unnormalized) posterior constitutes the complete Bayesian inference of the parameter values.

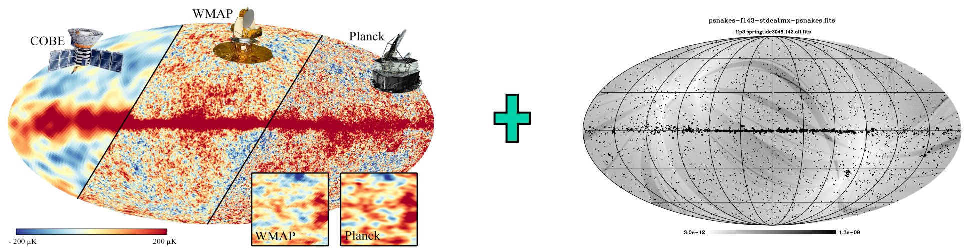
- Inferences are usually obtained either by taking samples from the (unnormalized) posterior using MCMC methods, or by locating its maximum (or maxima) and approximating the shape around the peak(s) by a multivariate Gaussian.

Example
Point sources in Planck mission

- Detection of discrete objects (eg point sources) immersed in a diffuse background - the background can be white noise, correlated noise, non-gaussian emission, etc.

$$d(x) = s(x) + n(x) = At(x) + n(x)$$

  - It is often assumed that the background is smoothly varying and has a characteristic length-scale much larger than the scale of the discrete objects being sought
    - Eg Sextractor (Bertin & Arnouts 1996) – run into problems, when the diffuse background varies on length-scales and with amplitudes similar to those of the discrete objects of interest.



- **The extra complication:**
  - The Cosmic Microwave Background, CMB, emission fluctuations varies on a characteristic scale of order ~10 arcmin, similar to that of extragalactic ‘point’ (i.e. beam-shaped) sources or the Sunyaev–Zel’dovich (SZ) effect in galaxy clusters, the objects we are interested in; the noise is anisotropic and can be corrrelated, destriping residuals, etc..

Standard approach

- Apply a linear filter Ψ(x) to the original image d(x) and analyse the filtered field:

$$d_f(x) = \int \psi(x-y)\, d(y)\, d^2 y.$$

- the filtering process as 'optimally boosting' (in a linear sense) the signal from discrete objects, with a given spatial template, and simultaneously suppressing emission from the background.
- If the original image contains N_{obj} objects at positions X_i with amplitudes A_i :

$$\underbrace{d(x) \equiv s(x)}_{\text{signal}} + \underbrace{n(x)}_{\text{generalised noise}} = \sum_{i=1}^{N_{obj}} A_i t(x - X_i) + n(x),$$

- It is straightforward to design an optimal filter function ψ(x) such that the filtered field (1) has the following properties:
  - (i) $d_f(X_k)$ is an unbiased estimator of A_i ;
  - (ii) the variance of the filtered noise field $n_f(x)$ is minimized;

- the corresponding function ψ(x) is the standard matched filter (eg. Haehnelt & Tegmark 1996).

Standard approach
Matched Filter

550 *M. G. Haehnelt and M. Tegmark*

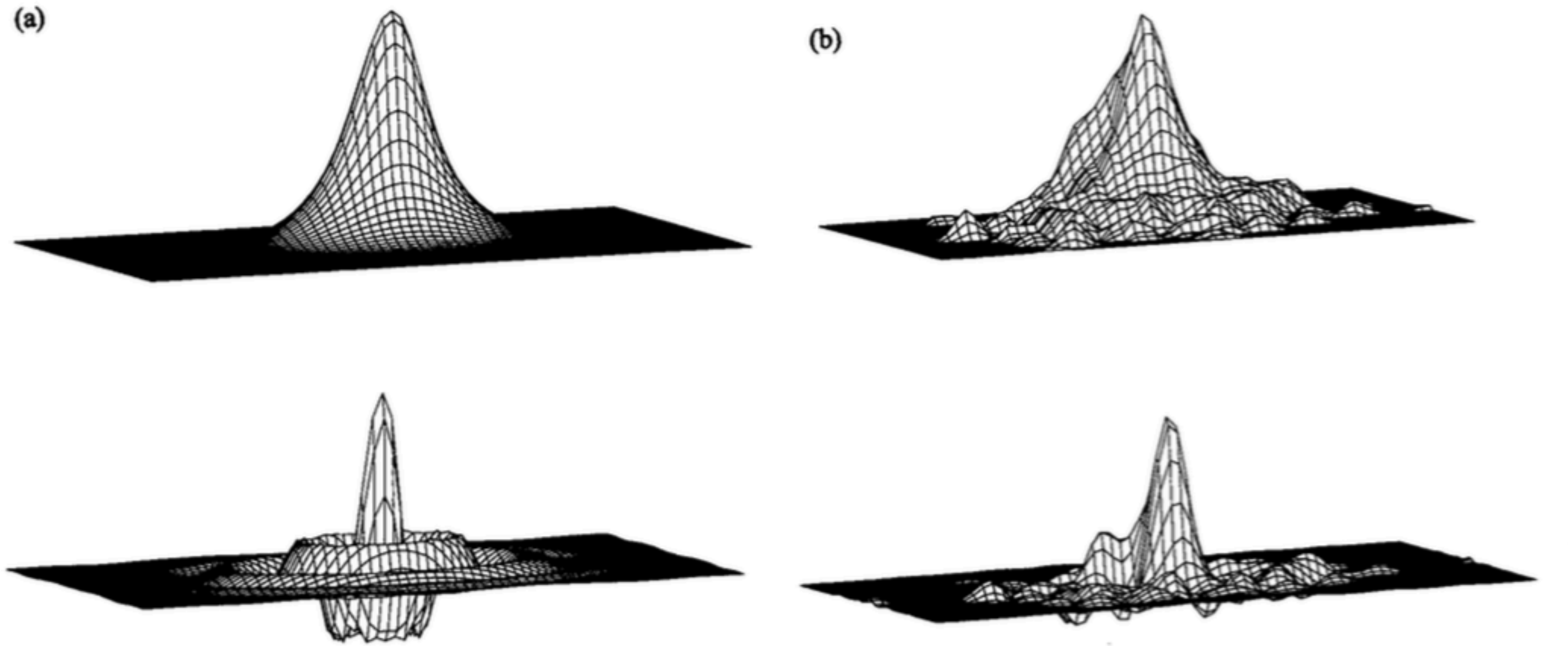


Figure 2. (a) Typical filter function for an axisymmetric cluster in arbitrary units. The cluster profile is shown for comparison. The angular scales are the same. (b) Same as (a) for a non-axisymmetric cluster. The cluster is shown at the top, the corresponding filter function at the bottom.

Filters - examples

- **Matched filters:**

(Herranz et al. 2002, Melin et al. 2006)

beam

$$M(k) = AB(k) + N(k)$$

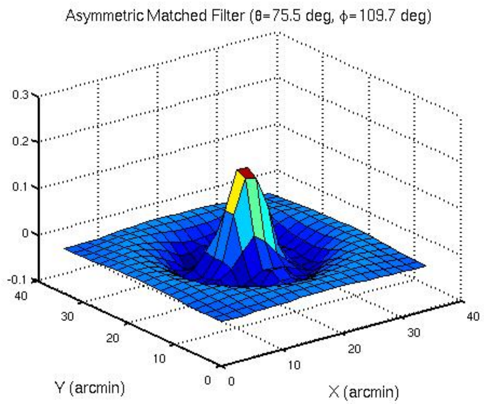
map

noise

AMPLITUDE ?

A is given by the filter: $\psi(k) \propto \dfrac{B(k)}{P(k)}$, where B(k) (beam or PSF) is known

$P(k) =< N(k)N(k)$ is the power spectrum of the generalized noise estimated on maps M(k)



- **Mexican Hat Wavelets:**

Mother wavelet in real space: $\psi_n(x) \propto \nabla^{2n} \exp[-|x|^2/2]$

Mother wavelet in Fourier space: $\widetilde{\psi}_n(k) \propto |k|^{2n} \exp[-|k|^2/2]$

Bayesian vs frequentist

- Some points to consider:

- The approaches outlined have been shown to produce good results, BUT
  - The filtering process is only optimal among the rather limited class of linear filters and is logically separated from the subsequent object detection step performed on the filtered map(s).
  - The detection threshold is empirically established - while the threshold is a logical byproduct of the framework in the Bayesian approach (*Carvalho, Rocha & Hobson, MNRAS, 393, 681C, 2009; Carvalho, Rocha, Hobson & Lasenby, MNRAS, 427, 201*)
  - It is well known that MFs are excellent at finding and locating sources, but not as good at estimating fluxes
  - Do not capitalize on previous knowledge both theoretical (modeling) and observational - for example using prior information we can enhance the probability of detecting very faint sources reliably (see references above)

- Bayesian approach
  - Hobson & McLachlan (2003) first introduced this approach: as in the filtering techniques, the method assumed a parameterized form for the objects of interest, but the optimal values of these parameters, and their associated errors, were obtained in a single step by evaluating their full posterior distribution. It was too slow and slower than the traditional approaches.
  - New efficient approach with PowellSnakes

Bayesian Inference
basic tools

- Bayesian inference methods provide a consistent approach to the estimation of a set parameters Θ in a model (or hypothesis) H for the data d.

- Bayes' theorem states that:

Likelihood - L(Θ)

Prior - π(Θ)

$$\Pr(\Theta|d, H) = \frac{\Pr(d|\Theta, H)\Pr(\Theta|H)}{\Pr(d|H)},$$

Bayesian evidence - E

Posterior probability distribution of the parameters P(Θ)

- In parameter estimation, the normalizing evidence factor is usually ignored, since it is independent of the parameters - This (unnormalized) posterior constitutes the complete Bayesian inference of the parameter values.
- Inferences are usually obtained either by taking samples from the (unnormalized) posterior using MCMC methods, or by locating its maximum (or maxima) and approximating the shape around the peak(s) by a multivariate Gaussian.

Bayesian Inference
basic tools

- In contrast to parameter estimation problems → in model selection the evidence takes the central role and is simply the factor required to normalize the posterior:

evidence

$$\mathcal{Z} = \int L(\Theta)\pi(\Theta)d^D\Theta,$$

Evaluation of this multidimensional Integral is a challenging numerical task – resort to sampling techniques: MCMC, Multinest, (Sivia &Skilling 2006; Feroz et al. 2009), etc. or model the posterior as a multivariate Gaussian centered at its peak(s) and apply the Laplace formula (Hobson, Bridle & Lahav 2002).

- The evidence is the expectation of the likelihood over the prior, and hence is central to Bayesian model selection between different hypothesis H_i

- The evidence automatically implements Occam's razor:

A simpler theory with compact parameter space will have a larger evidence than a more complicated one, unless the latter is significantly better at explaining the data.

Bayesian Inference
basic tools

$$\mathcal{Z} = \int L(\Theta)\pi(\Theta)d^D\Theta,$$

The evidence is the expectation of the likelihood over the prior, and hence is central to Bayesian model selection between different hypothesis H_i

- The question of model selection between two models H_0 and H_1 can then be decided by comparing their respective posterior probabilities given the observed data set d

$$\frac{\Pr(H_1|d)}{\Pr(H_0|d)} = \frac{\Pr(d|H_1)\Pr(H_1)}{\Pr(d|H_0)\Pr(H_0)} = \frac{\mathcal{Z}_1}{\mathcal{Z}_0}\frac{\Pr(H_1)}{\Pr(H_0)},$$

where $\Pr(H_1)/\Pr(H_0)$ is the a priori probability ratio for the models

Bayesian Inference
decision theory

- Probability theory defines only a state of knowledge: the posterior probabilities.

There is nothing in probability theory per se that determines how to make decisions based on these probabilities.

- To deal with such difficulties, apply decision theory - one must first define the loss/cost function L(D, E) for the problem at hand,
  - where D is the set of possible decisions and E is the set of true values of the entities one is attempting to infer.
  - DT can be applied equally well to both parameter estimation and model selection

- Loss function - maps the 'mistakes' in our estimations/selections, D, into positive real values L(D, E), thereby defining the penalty one incurs when making wrong judgments.

The Bayesian approach to DT is simply to minimize, with respect to D, the expected loss:

$$\langle L(\boldsymbol{D}, \boldsymbol{E}) \rangle = \iint L(\boldsymbol{D}, \boldsymbol{E}) \Pr(\boldsymbol{D}, \boldsymbol{E}) \, \mathrm{d}\boldsymbol{D} \, \mathrm{d}\boldsymbol{E}.$$

'decisions' D = parameter estimates $\Theta^{\wedge}$; 'entities' E = true values Θ^{*} of the parameters

Bayesian Object Detection

Ingredients: Data Model, Likelihood and Priors

- Data Model
  - consider our data vector d (pixel values or Fourier coefficients of the image) In each frequency channel. With x=position vector in pixel space; N_s = number of sources

data → $$\boldsymbol{d}(\boldsymbol{x}) = \underbrace{\sum_{j=1}^{N_s} \boldsymbol{s}_j(\boldsymbol{x};\boldsymbol{\Theta}_j)}_{\text{signal from the sources}} + \boldsymbol{b}(\boldsymbol{x}) + \boldsymbol{n}(\boldsymbol{x}),$$

$$\boldsymbol{s}(\boldsymbol{x};\boldsymbol{\Theta}) \equiv \sum_{j=1}^{N_s} \boldsymbol{s}_j(\boldsymbol{x};\boldsymbol{\Theta}_j).$$

signal from the sources

Background sky emission
1) Foregrounds, CMB,…
2) Speckles, systematics,..

\+ Instrumental noise

Generalised noise (correlated and white noise)

$$\boldsymbol{s}_j(\boldsymbol{x};\boldsymbol{\Theta}_j) = A_j \boldsymbol{f}(\boldsymbol{\phi}_j)\underbrace{\boldsymbol{\tau}(\boldsymbol{x}-\boldsymbol{X}_j;\boldsymbol{a}_j)}_{},$$

→ convolved spatial template at each frequency of a source centered at the position Xj and characterized by the shape parameter vector aj

  - vector f = emission coefficients at each frequency, which depend on the emission law parameter vector φj of the source; Aj is an overall amplitude for the source at some chosen reference frequency
  - In this example the j_{th} source parameters are: $\boldsymbol{\Theta}_j = \{A_j, \boldsymbol{X}_j, \boldsymbol{a}_j, \boldsymbol{\phi}_j\}$, amplitude, position, shape parameters, emission law parameters

Bayesian Object Detection

Ingredients: Data Model, Likelihood and Priors

- Likelihoods

  The form of the likelihood is determined by the statistical properties of the generalized noise (background sky emission plus instrumental noise) in each frequency channel

  - if the background 'noise' n is a statistically homogeneous Gaussian random field with covariance matrix $N = \langle nn^T \rangle$ - Multivariate Gaussian Likelihood

$$L(\boldsymbol{a}) = \frac{\exp\left\{-(1/2)\left[\boldsymbol{d} - \boldsymbol{s}(\boldsymbol{a})\right]^{\mathrm{T}} \boldsymbol{N}^{-1} \left[\boldsymbol{d} - \boldsymbol{s}(\boldsymbol{a})\right]\right\}}{(2\pi)^{N_{\mathrm{pix}}/2} \left|\boldsymbol{N}\right|^{1/2}}.$$

$$s(\boldsymbol{x}; \boldsymbol{a}) = A \exp\left[-\frac{(x-X)^2 + (y-Y)^2}{2R^2}\right]$$

  - We are interested in the likelihood ratio between the hypothesis H_s that objects (of a given source type s) are present and the null hypothesis H_0 that there are no such objects (= corresponds to setting the sources signal $s(x;\Theta)$ to zero):

$$\ln\left[\frac{\mathcal{L}_{H_s}(\boldsymbol{\Theta})}{\mathcal{L}_{H_0}(\boldsymbol{\Theta})}\right] = \sum_{\eta} \tilde{\boldsymbol{d}}^t(\eta) \boldsymbol{\mathcal{N}}^{-1}(\eta) \tilde{\boldsymbol{s}}(\eta; \boldsymbol{\Theta}) - \frac{1}{2} \sum_{\eta} \tilde{\boldsymbol{s}}^t(\eta; \boldsymbol{\Theta}) \boldsymbol{\mathcal{N}}^{-1}(\eta) \tilde{\boldsymbol{s}}(\eta; \boldsymbol{\Theta}),$$

tilde denotes a Fourier transform
$K=2\pi\eta$ mode wavenumber ;
$N(\eta)$= generalized noise cross-power spectra

Bayesian Object Detection

Ingredients: Data Model, Likelihood and Priors

- How frequentist approach naturally emerges within the Bayesian framework

  - Maximizing the likelihood ratio, with respect to the source amplitudes A_j

We recover the expression for the MMF

$$\widehat{A}_j = \frac{\mathcal{F}^{-1}\left[\mathcal{P}_j(\boldsymbol{\eta})\widetilde{\tau}(-\boldsymbol{\eta};\widehat{\boldsymbol{a}}_k)\right]_{\widehat{X}_j}}{\sum_{\eta} \mathcal{Q}_{jj}(\boldsymbol{\eta})\,|\,\widetilde{\tau}(\boldsymbol{\eta};\widehat{\boldsymbol{a}}_j)\,|^2}$$

$$\hat{A}(\underline{X}, R) = \frac{\boldsymbol{t}^{\mathrm{T}}(\underline{X}, R)N^{-1}\boldsymbol{d}}{\boldsymbol{t}^{\mathrm{T}}(\underline{X}, R)N^{-1}\boldsymbol{t}(\underline{X}, R)}. \qquad \psi(k) \propto \frac{B(k)}{P(k)}$$

  - Substituting this maximum-likelihood (ML) estimate onto the Likelihood ratio expression we get for the jth object:

$$\max\left[\ln\left(\frac{\mathcal{L}_{H_s}}{\mathcal{L}_{H_0}}\right)\right] = \frac{1}{2}\sum_{\eta}\mathcal{Q}_{jj}(\boldsymbol{\eta})\,|\,\widetilde{\tau}(\boldsymbol{\eta};\widehat{\boldsymbol{a}}_j)\,|^2\widehat{A}_j^2 = \frac{1}{2}\widehat{\mathrm{SNR}}_j^2$$

$\widehat{\mathrm{SNR}}_j$

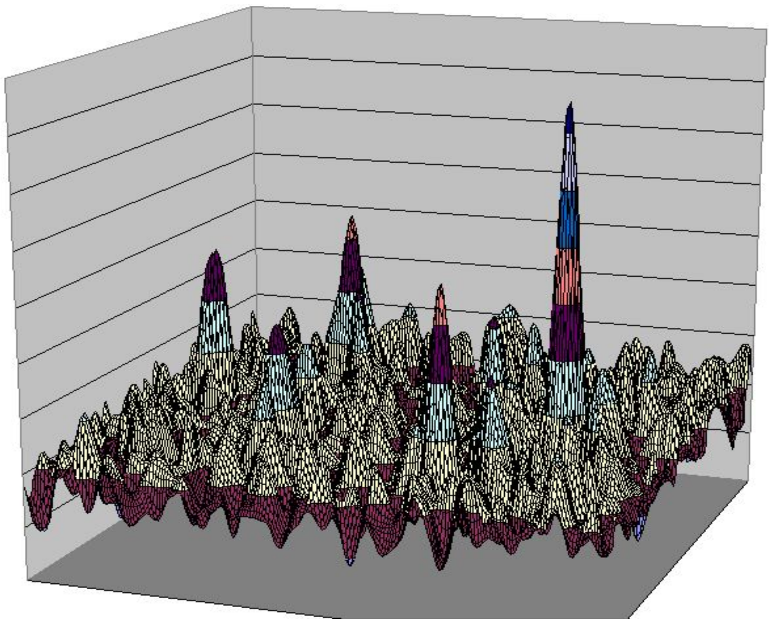
SNR (at the peak) of the jth source

Thus, one sees that in the traditional approach to catalogue making, in which one compares the maximum SNR of the putative detections to some threshold, one is really performing a generalized likelihood ratio test (GLRT)

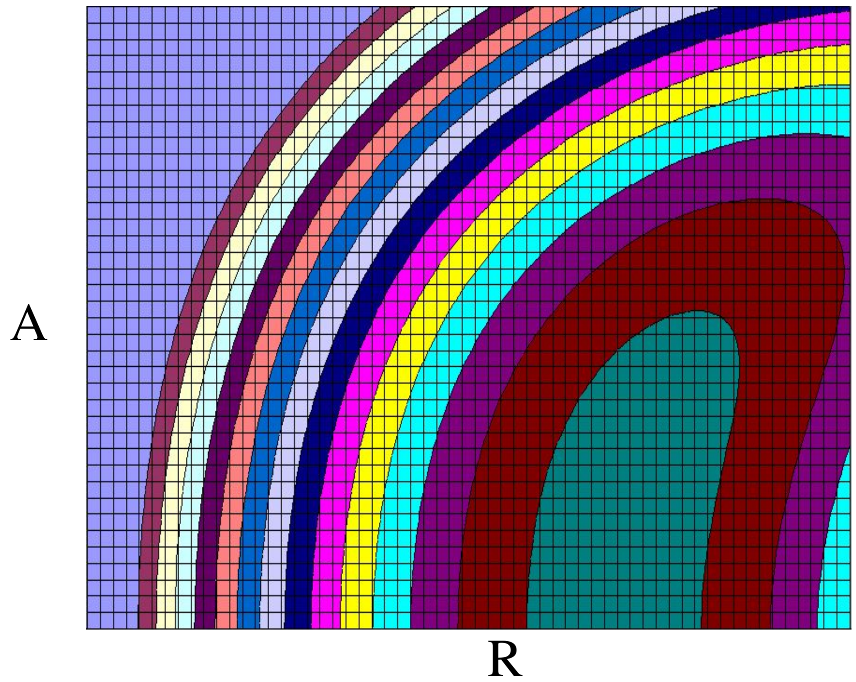
Likelihood manifold

- The filtered field is the projection of the likelihood manifold onto the sub-space of position parameters X_j

$$L(a) = \frac{\exp\left\{-\frac{1}{2}\left[d - s(a)\right]^{\mathrm{T}} \mathrm{N}^{-1}\left[d - s(a)\right]\right\}}{(2\pi)^{N_{\mathrm{pix}}/2}\,|\mathrm{N}|^{1/2}}.$$



Position subspace (X,Y) ; High res antenna



(A,R) subspace ; (X_0,Y_0) of a maximum

Bayesian Object detection

Ingredients: Data Model, Likelihood and Priors

- Priors
  - The Jeffreys (Jeffreys 1961) rule for constructing ignorance priors for the one-dimensional case read:

$$\pi(\theta) \propto \mathcal{J}^{1/2}(\theta), \qquad \text{where} \qquad \mathcal{J}(\theta) \equiv -\left\langle \frac{\partial^2 \ln \mathcal{L}(\theta)}{\partial \theta^2} \right\rangle$$

Fisher information

- Examples of priors used:

- Prior on positions: if the sky patches used are sufficiently small, our locally uniform model can easily cope with clustering when the gradient of the density of sources is small across the patch boundaries.
  - The correctly normalized positions prior for the complete ensemble of sources in a patch is simply

$$\Pr\left(X^{N_s} \mid N_s,\ N_{pix}\right) = \frac{1}{{N_{pix}}^{N_s}},$$

  - Npix is the number of pixels in each patch and Ns is the number of sources in that patch

Bayesian Object detection

Ingredients: Data Model, Likelihood and Priors

- Prior on the models: The prior ratio $Pr(H_1)/ Pr(H_0)$ on the models is often neglected (i.e. assumed to equal unity), but plays a very important role in the PwS detection criterion

    - let us imagine we know in advance all the true values of the parameters that define an object, which translates into delta-function priors, then we obtain the inequality:

$$\mathrm{SNR} \underset{H_0}{\overset{H_1}{\gtrless}} \sqrt{2\left[\xi + \ln\left(\frac{\Pr(H_0)}{\Pr(H_1)}\right)\right]}$$

    - interpret the term $\ln(Pr(H_0) /Pr(H_1))$ as an extra 'barrier' added to the detection threshold
        - because we are expecting more fake objects than the objects of interest, due to background fluctuations
    - Assuming Poisson statistics for the number of sources and the number of likelihood maxima resulting from the background fluctuations:

$$\frac{\Pr(H_1 \mid N_s)}{\Pr(H_0 \mid N_s)} = \left(\frac{\lambda_1}{\lambda_0}\right)^{N_s}$$

λ_0 =expected number of maxima per unit area resulting from background fluctuations above the minimum limit of detection of the experiment
λ_1 = number density of sources above the same limit - derived from their differential counts

PWSIII

- At the Post-Processing stage there are a number of Image Processing techniques that aim at modeling and subtracting the stellar Point Spread Function, PSF, to allow the planet to become detectable, in effect increasing the contrast achievable next to a bright star:

- Angular Differential Imaging, ADI, (Marois et al. 2008)
- LOCI, (Lafreniere et al. 2007);
- Reference Differential Imaging RDI
- Principal Component Analysis, PCA, (Amara & Quanz 2012, Meshkat et al. 2014)
- KLIP (Soummer et al. 2012) which uses the Karhunen-Loeve, KL, transform to model the PSF
- Stochastic speckle discrimination, SSD, (Gladysz & Christou 2008)
- Enhanced faint companion photometry and astrometry using wavelength diversity (Burke & Devaney 2010)
- KLIP-FM (Pueyo 2016)
- …..

PWSIII

(1) Construct the Likelihood,. L ((x, y), t, λ, I):

- As the subspace are independent we can recast it as: L ((x, y), t, λ, I)= L ((x, y), I) x L (t) x L (λ);
  - Spatial likelihood - Multivariate Gaussian
  - Temporal Likelihood - mild Non- Gaussian likelihood specified by the first 3 moments of a PDF *(Rocha et al. 2001, Rocha et al. 2005)*
- The priors for the model parameters will be constructed based on previous observations and any other relevant information that helps distinguishing the signal from the noise
- Construct an optimal adaptive matched filter (MF): based on the spatial estimation of the noise (using KLIP for example) and a spatial model for the planet (e.g. a Airy function) and/or – current study
- Use multi-wavelength data to estimate the covariance of the data, estimate the PSF and construct a new Multi-Matched filter, MMF, a whitening filter (akin to the Hotelling observer),

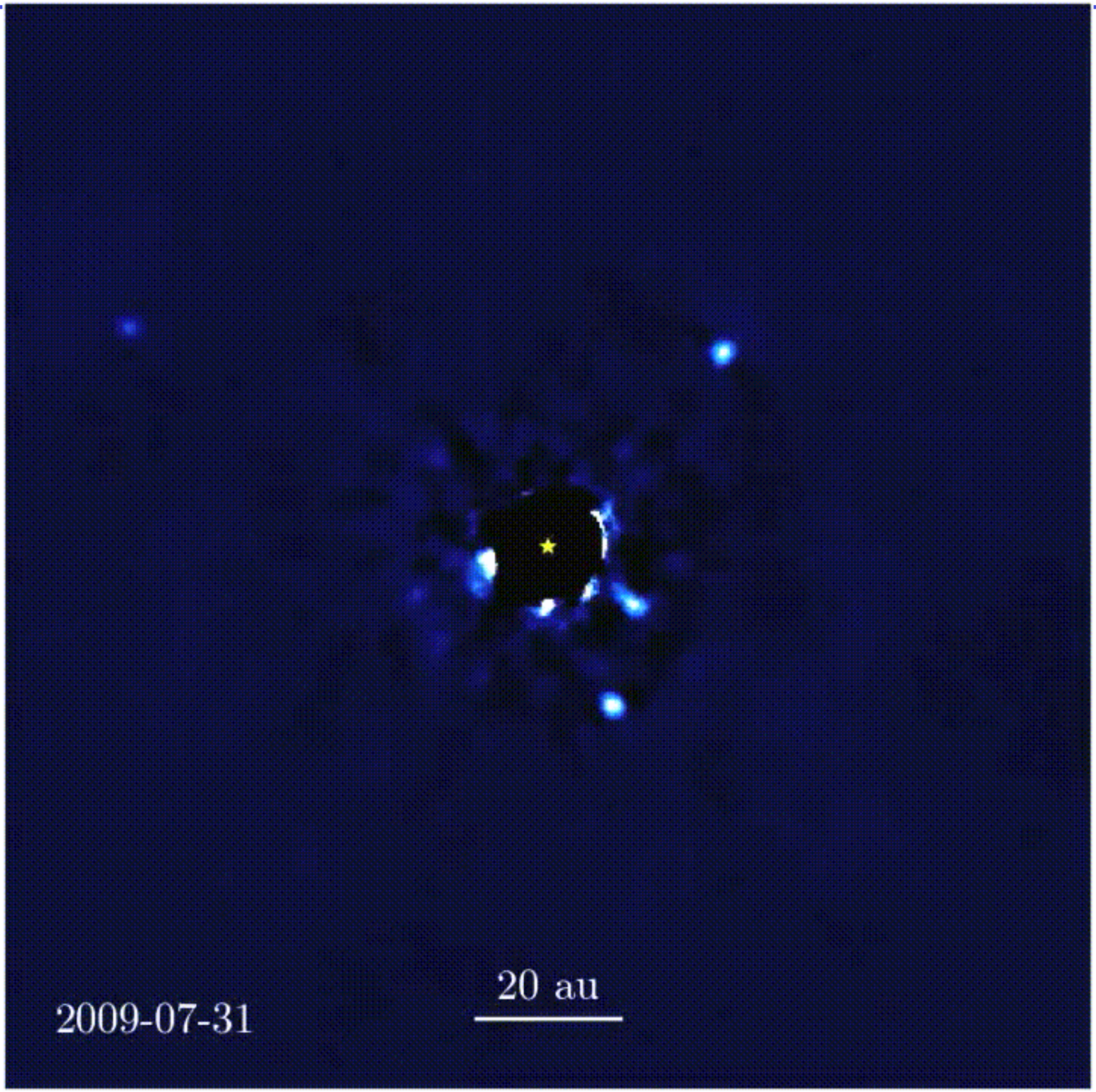
(2) Estimate the posterior distributions of the model parameters + the evidence ratios of the competing models

(3) To improve detection characterization - repeat the previous step - this time as a temporal analysis of the peaks in the previous Likelihood manifold (filtered map in the positional subspace):

- (a) Construction of a potentially Non-Gaussian Likelihood based and construction of priors for the moments of the distribution, (b) estimating the posterior distributions of these moments and (c) the evidence ratios of the competing probability distributions

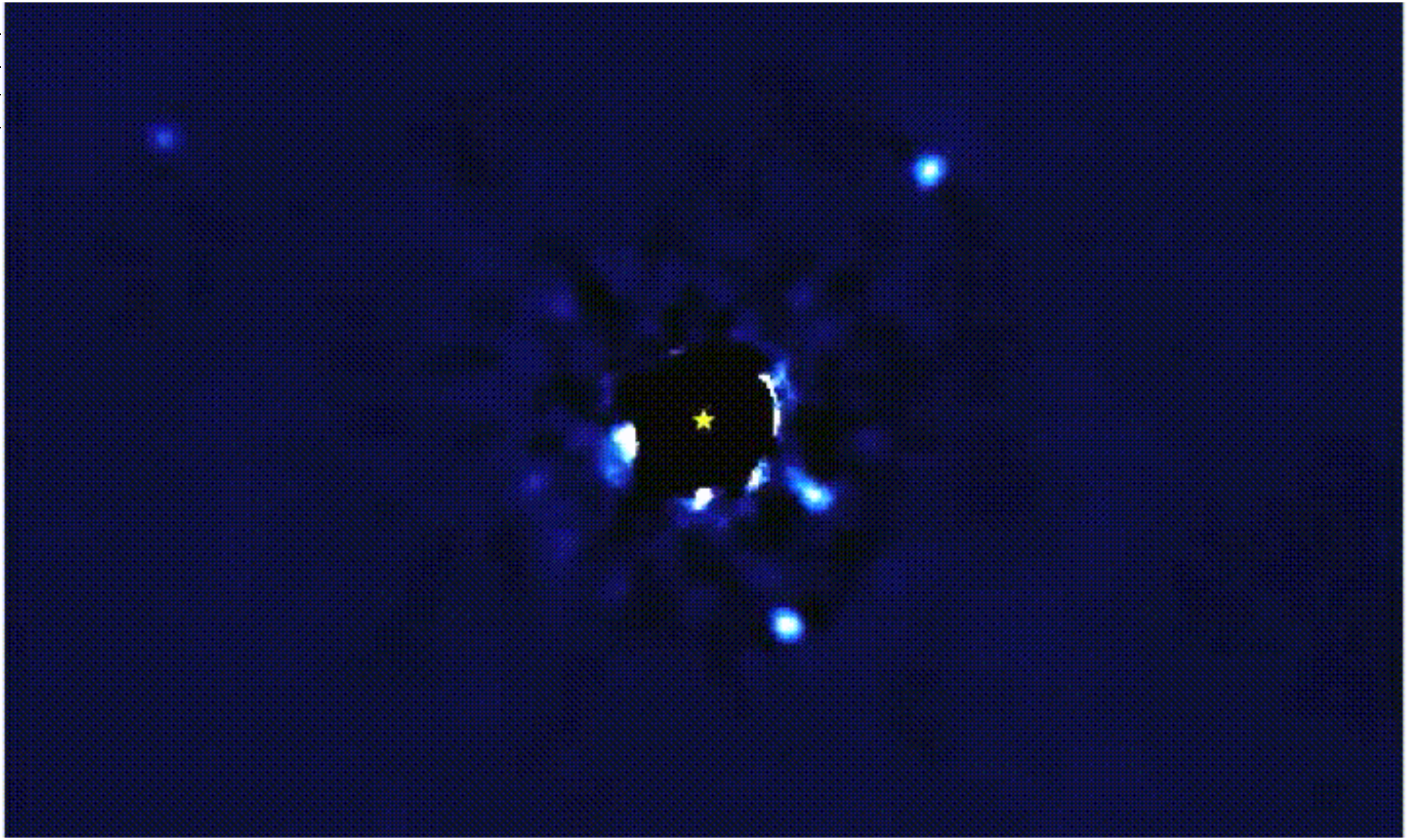
- Quantify performance of PWSIII using simulations with injected planets

PWSIII: current - Keck telescope
future - JWST(MIRI), WFIRST, HabEX, LUVOIR



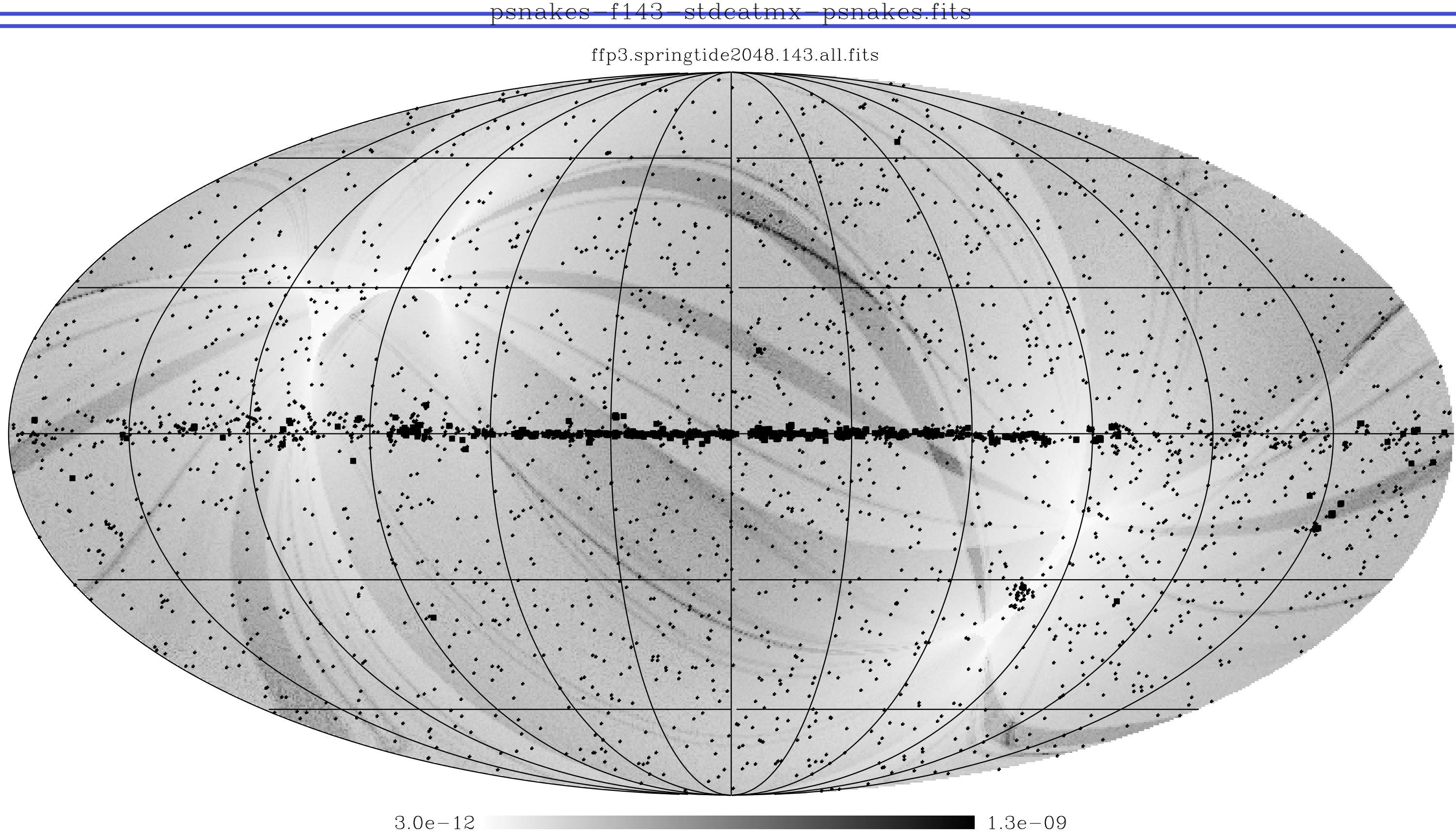
PWSIII: current - Keck telescope
future - JWST(MIRI), WFIRST, HabEX, LUVOIR

HR8799



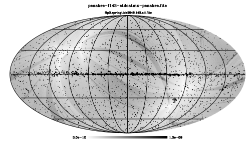
Appendix

PowellSnakes - *A fast Bayesian approach for detection of compact localised objects immersed in a diffuse background on large datasets*



'Bayesian Methods in Cosmology' – CUP, December 2009
chapter on 'Bayesian Source Extraction' by Hobson, Rocha & Savage

Carvalho, Rocha & Hobson, MNRAS, 393, 681C, 2009
Carvalho, Rocha, Hobson & Lasenby, MNRAS, 427, 2011
Rocha, in preparation



A Bayesian approach for detection and characterization of extrasolar planets

Graca Rocha,

Dimitri Mawett,

Bertrand Mendelshon,

Tiffany Meshkat,

Gautam Vasisht

JPL, Caltech & IPAC

ESI at JPL, 26th of March 2018

Motivation

- To extend PowellSakes, PwS, a Bayesian approach, to direct imaging data analysis
  - PwS has successfully been applied to detect compact sources immersed in a diffuse background in Planck maps - *Carvalho, Rocha & Hobson, MNRAS, 393, 681C, 2009; Carvalho, Rocha, Hobson & Lasenby, MNRAS, 427, 2011; Bayesian Methods in Cosmology' – CUP, December 2009: chapter on 'Bayesian Source Extraction' by Hobson, Rocha & Savage.*

- **Motivation:**
  - To help enhance the detectability of faint exoplanets at small orbital separations from the host star
    - Both ground-based and space-based instruments have not yet achieved the contrast gain needed to detect mature planets with masses lower than 1 Jupiter mass at separations smaller than 0.5".
    - The difficulty arises from the residual glare of starlight at small orbital separations due to diffraction, scattered light, and speckles caused by defects in the optical system
  - New approach -> unify source detection and characterization (Position, Flux or Intensity and hence accurate spectrum extraction) into one single rigorous mathematical framework, the Bayesian framework, enabling an adequate hypothesis testing given the S/N of the data.
  - The method will be applied in combination with other post-processing techniques (best suited for this approach), for example KLIP, but now recast in a Bayesian perspective.

Bayesian – what does it really mean?

- Bayes' theorem states that:

Likelihood: L(Θ)

Prior: π(Θ)

$$\Pr(\Theta|d, H) = \frac{\Pr(d|\Theta, H)\Pr(\Theta|H)}{\Pr(d|H)},$$

Posterior probability distribution of the parameters: P(Θ)

Bayesian evidence: E

- In parameter estimation, the normalizing evidence factor is usually ignored, since it is independent of the parameters - This (unnormalized) posterior constitutes the complete Bayesian inference of the parameter values.

- Inferences are usually obtained either by taking samples from the (unnormalized) posterior using MCMC methods, or by locating its maximum (or maxima) and approximating the shape around the peak(s) by a multivariate Gaussian.

Object detection strategy (briefly)

- Evaluation of the odds ratio
  - 'Brute force' evaluation of the evidence integrals is still not feasible
  - Use MCMC methods and thermodynamic integration - can fail when the posterior distribution is very complex
  - use 'Nested Sampling' (Sivia &Skiling 2006),which is much more efficient, although not without its difficulties; MultiNest' (Feroz et al. 2009) efficient implementation of the nested sampling algorithm, which is capable of exploring high-dimensional multi-modal posteriors; other simpler nested sampling scheme (Mukherjee, Parkinson & Liddle 2006) perform well.

- Or use another approach (as in PwsI):
  - PwS I started a Powell minimization chain (hence the name 'PowellSnakes') in many different locations of the manifold in an attempt to find all the maxima - where the Brent line minimizer was 'enhanced' with an ancillary step to allow it to 'tunnel' from one minimum to the next.
  - Explore the fact that we can separate the position variables from all others- so first locate maxima in position space, then start a four-dimensional PwS optimization at each such location to find the ML parameters for that particular peak

Object detection strategy (briefly)

- Exploring the posterior distribution
  - Our initial step provides the ML estimates and the SNR of each detection candidates
  - Only a much smaller sub-set is chosen based on an SNR threshold.
  - This shorter list is then sorted in descending order of SNR and one-by-one the maxima are sent to the nested sampler,
  - The nested sampler returns an evidence estimate and a set of weighted samples that we use to model the full joint posterior distribution
  - The final catalogue is almost completely independent of the SNR threshold if this is not too high
  - From these samples we can compute any parameter estimate, draw joint distribution surfaces, predict HPD intervals of any content over the marginalized distributions to infer the parameter uncertainties

PWSIII

(1) To improve detection characterization repeat the previous step - this time as a temporal analysis of the peaks in the previous Likelihood manifold (filtered map in the positional subspace):

- (a) Construction of a potentially Non-Gaussian Likelihood based and construction of priors for the moments of the distribution, (b) estimating the posterior distributions of these moments and (c) the evidence ratios of the competing probability distributions

- Quantify performance of PWSIII using simulations with injected planets

- Ongoing work with Keck data for HR 8799

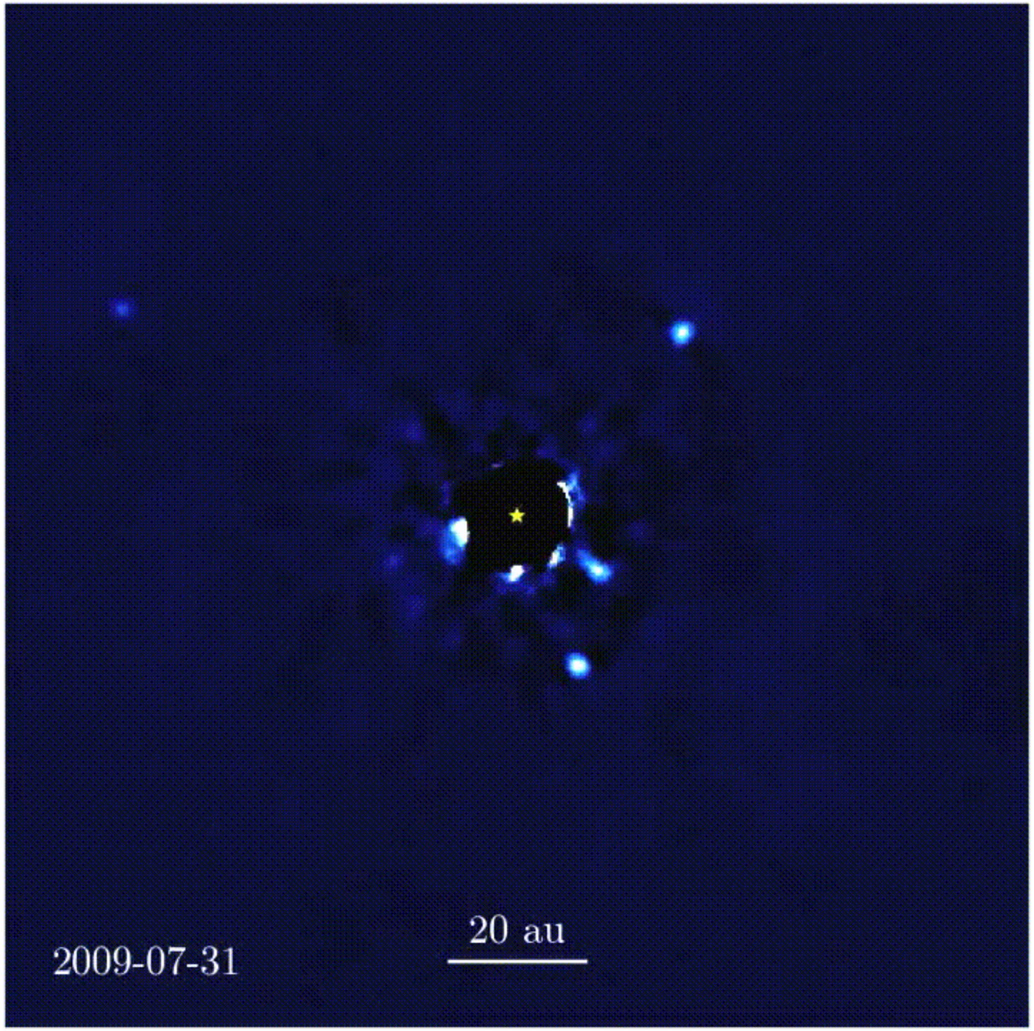
PWSIII

- At the Post-Processing stage there are a number of Image Processing techniques that aim at modeling and subtracting the stellar Point Spread Function, PSF, to allow the planet to become detectable, in effect increasing the contrast achievable next to a bright star:

- Angular Differential Imaging, ADI, (Marois et al. 2008); LOCI, (Lafreniere et al. 2007); Reference Differential Imaging, RDI; Principal Component Analysis, PCA, (Amara & Quanz 2012, Meshkat et al. 2014); KLIP (Soummer et al. 2012) which uses the Karhunen-Loeve, KL, transform to model the PSF; Stochastic speckle discrimination, SSD, (Gladysz & Christou 2008) Enhanced faint companion photometry and astrometry using wavelength diversity (Burke & Devaney 2010); KLIP-FM (Pueyo 2016)

PWSIII

HR8799
Keck data



Bayesian Object detection

Ingredients: Data Model, Likelihood and Priors

- Priors

- If the data model provides a good description of the observed data and the SNR is high –
  - prior will have little or no influence on the posterior distribution.
- At the faint end of the source population, when we are getting close to the instrument sensitivity limit, however, priors will inevitably play an important role.
  - The selection of the priors becomes important and has to be addressed very carefully.

- PwSII separates the tasks of source detection (deciding whether a certain signal is due to a source) and source estimation (determining the parameters of the source).
  - This separation has the advantage of allowing the use of different sets of priors at each stage.
    - First perform the source detection step using 'informative' priors, which encompass all the available information, since they provide the optimal selection criterion and the optimal estimators.
    - After the set of detections has been decided, PwS proceeds to the estimation pass, in which 'non-informative' priors may be used instead.

- majority of applications, the parameters may be assumed independent, so that the prior factorizes

$$\pi(\theta_1, \theta_2 \ldots, \theta_n) = \pi_1(\theta_1)\pi_2(\theta_2)\cdots\pi_n(\theta_n).$$