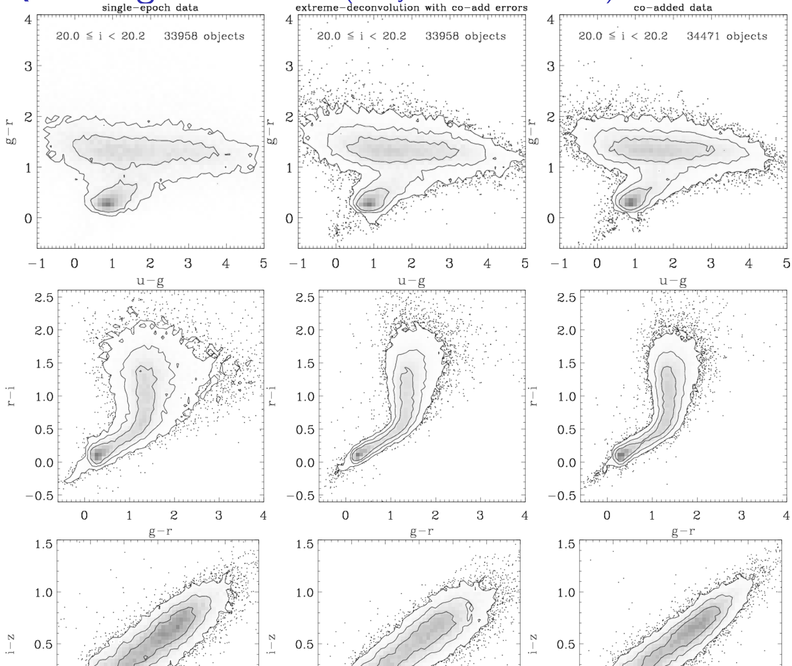
The Astronomer's Theory of Everything

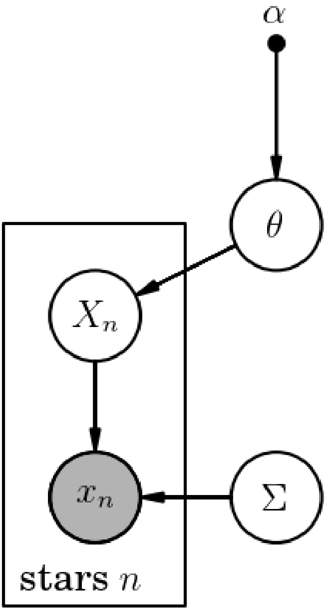
David W. Hogg
Center for Cosmology and Particle Physics,
New York University
and
Max-Planck-Institut für Astronomie, Heidelberg

2013 March 5

XDQSO target selection (Bovy *et al.*, 1011.6392)



lots of bad data is equal to a bit of good data



α
θ
X_n
x_n
Σ
stars n

Hogg's Decadal Survey

- Software is going to be much more productive, per dollar, than anything else in astrophysics!

the new reality

- all our goals are getting substantially more ambitious every ten years
- funding is flat
- the era of under-budgeting and over-designing (and getting away with it) is over
- many observational programs will lead to upper limits or bare detections
  - example: *JWST* will take low s/n spectra of a **couple** of Earth-like exoplanets (in the most optimistic scenario) **over its entire mission lifetime**.
- existing (and aging) data will become far more valuable
  - example: The sky has been imaged thousands of times already in many bands; these data have never seriously been assembled and analyzed coherently.

we need new tools

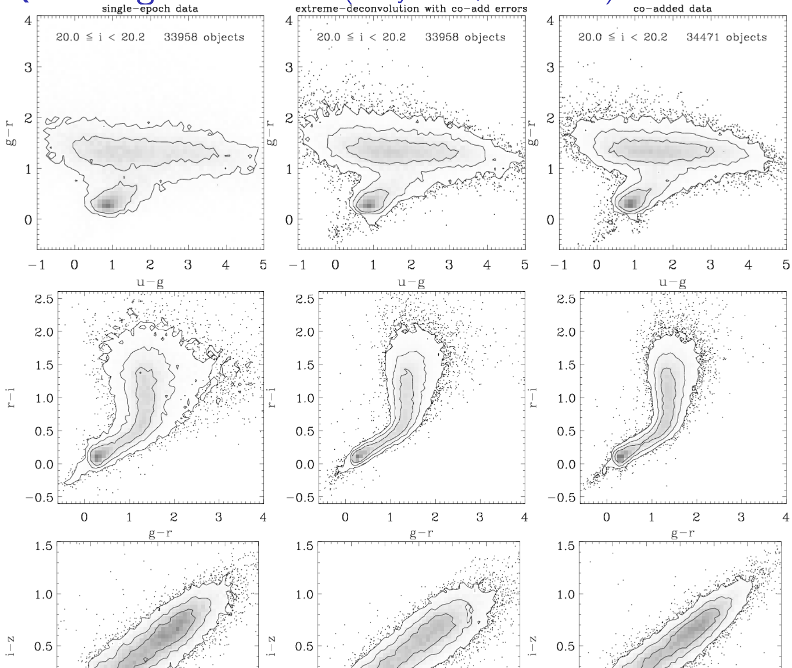
- How do we move **all** of the information from all of the data (ever taken) to the quantities of interest?
- How do we get high s/n information about what we **want** when every individual datum is very noisy?
- It's all about **tools.**

conclusions

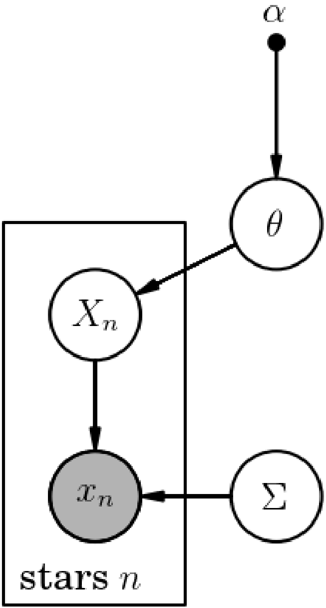
- software, software, software, and applied math
- lots of bad data is equal to a bit of good data
- modeling beats supervised classification
- point estimates are bad, models are good
- heirarchical modeling will require MCMC
- we can find populations, no member of which is individually detectable

modeling beats supervised classification

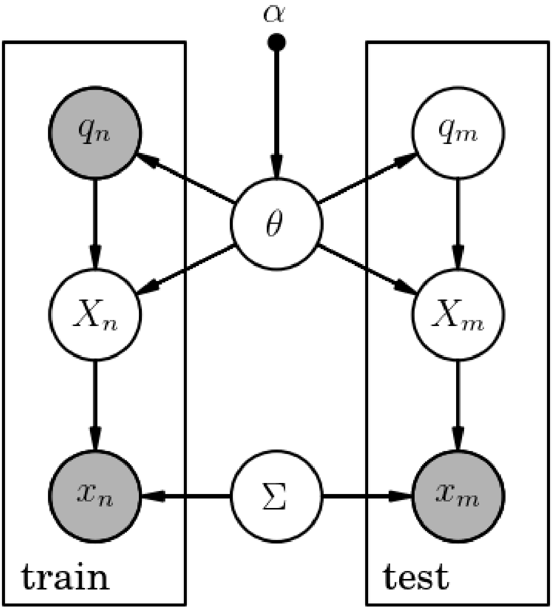
XDQSO target selection (Bovy *et al.*, 1011.6392)



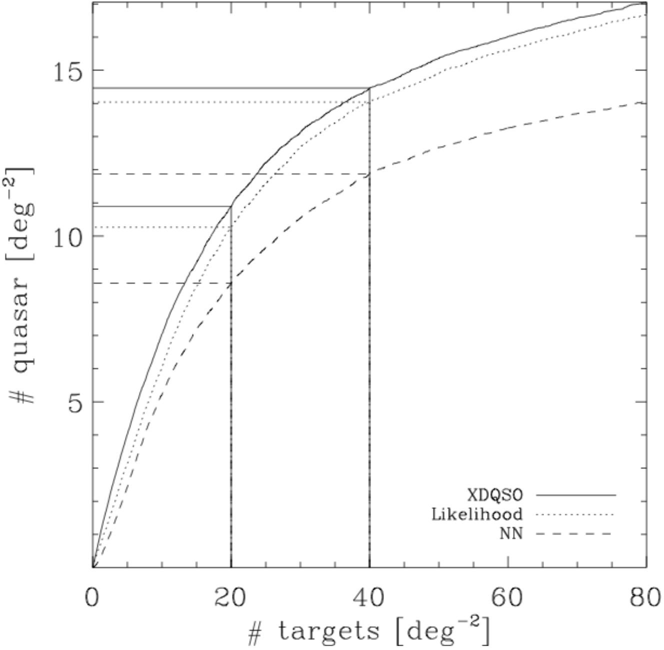
XDQSO target selection (Bovy *et al.*, 1011.6392)



XDQSO target selection (Bovy *et al.*, 1011.6392)



XDQSO target selection (Bovy *et al.*, 1011.6392)

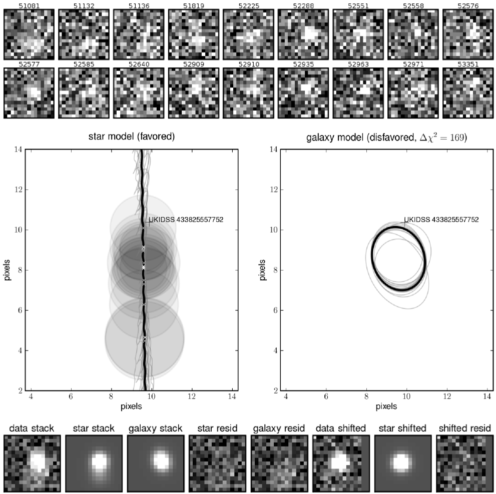


what's wrong with supervised classification?

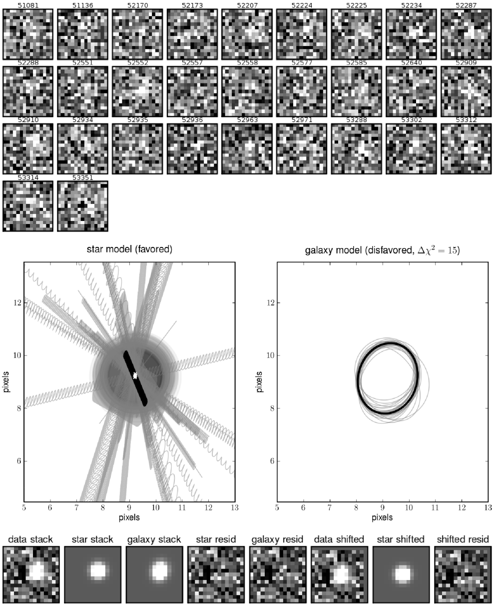
- support vector machines, boosting, deep learning
- these are all **awesome**
- they require that **test data** have the same statistical and error properties as **training data**
  - **never true!**
- they require that all features be measured for all data points
  - **never true!**
  - (If you know enough about your data to fix this problem, then **just write down a likelihood!**)

point estimates are bad, models are good

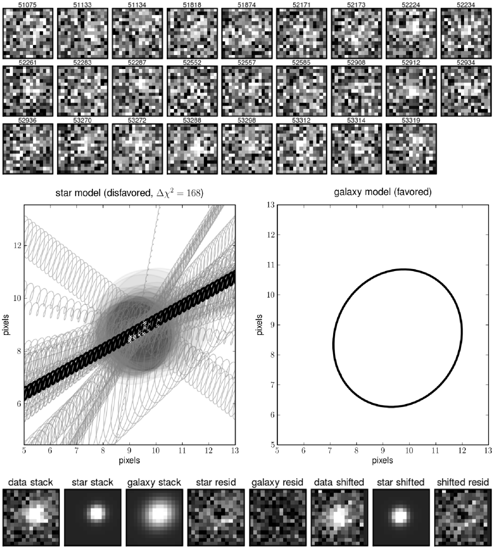
faint proper motions (Lang *et al.* 0808.4004)



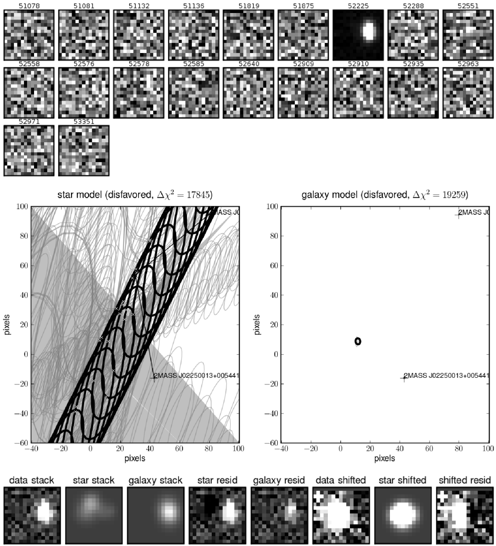
faint proper motions (Lang *et al.* 0808.4004)



faint proper motions (Lang *et al.* 0808.4004)



faint proper motions (Lang *et al.* 0808.4004)



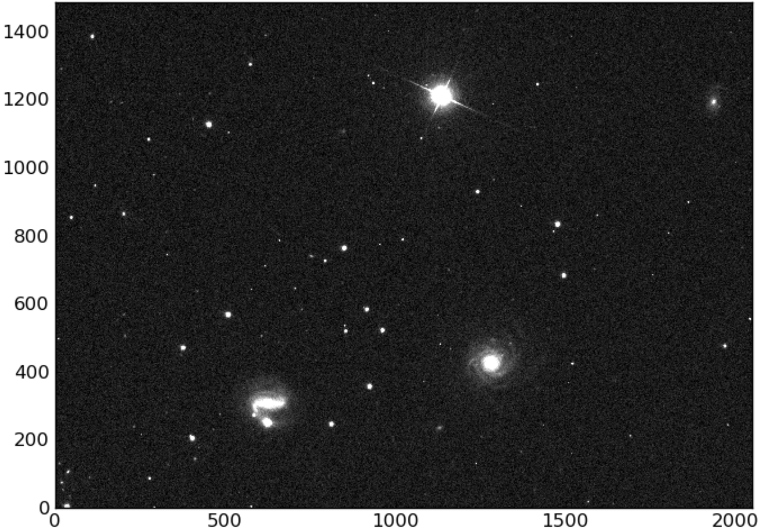
faint proper motions (Lang *et al.* 0808.4004)

- If we had only a catalog, we would have **failed**.
- If we had only a coadd, we would have **failed**.

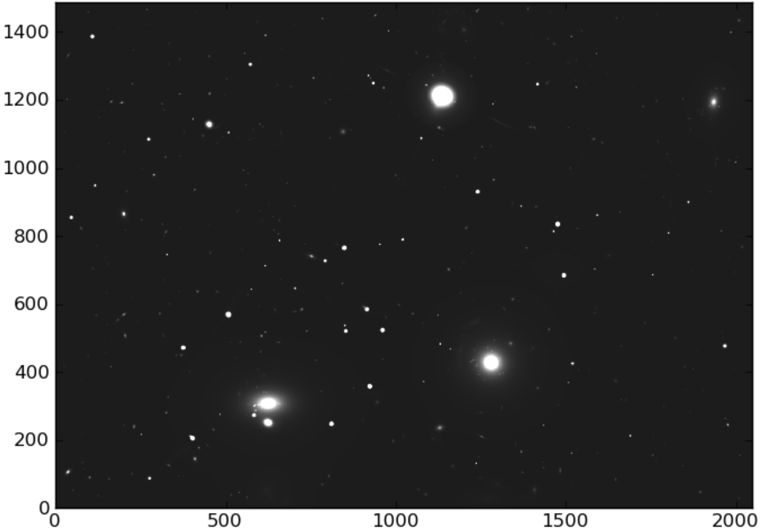
what's wrong with *LSST* and *PanSTARRS*?

- reducing data with point estimates
- building catalogs from "co-adds" with point estimates
- catalog matching
- **All of these throw away information. Does it matter?**
  - Lang and I are betting it does: *theTractor.org*

The *Tractor* (Lang *et al.*)



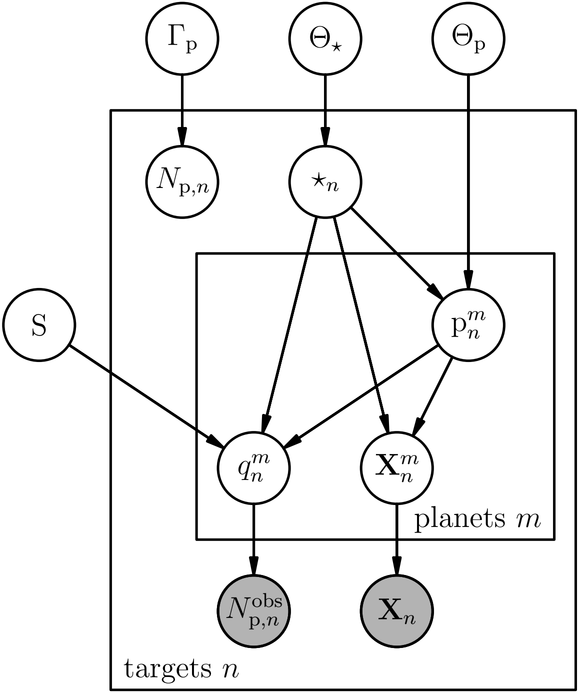
The *Tractor* (Lang *et al.*)



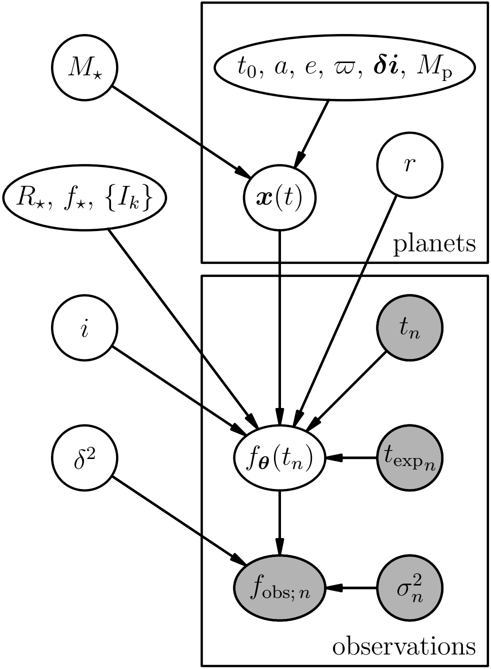
heirarchical modeling will require MCMC

the Exoplanet Theory of Everything

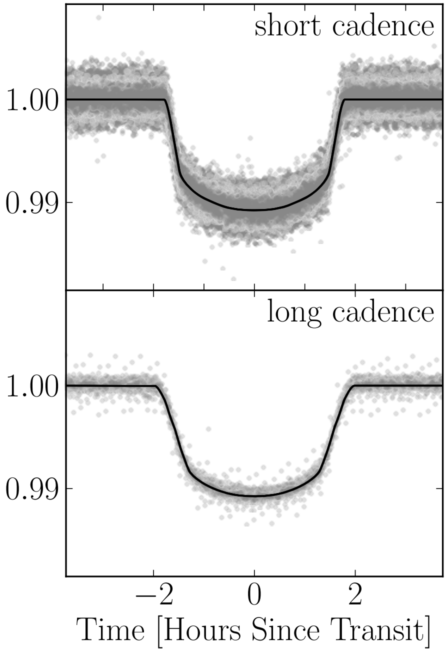
- different techniques find different planets
  - radial velocity, transit, direct detection, astrometry, microlensing
- completeness or selection functions are smooth functions of exoplanet and host-star properties
- most observations show **no clearly detectable** planet
- Earth-like planets are in a "bitter spot" for all observational techniques
- how do we take the (literally) billions of data points and obtain the best possible picture of the full exoplanet population?
  - distributions in orbital and planetary (composition, size) parameters
  - distributions for multiplicity and "architecture"
  - all as a function of host-star properties
  - plus all the individual systems measured as well as possible

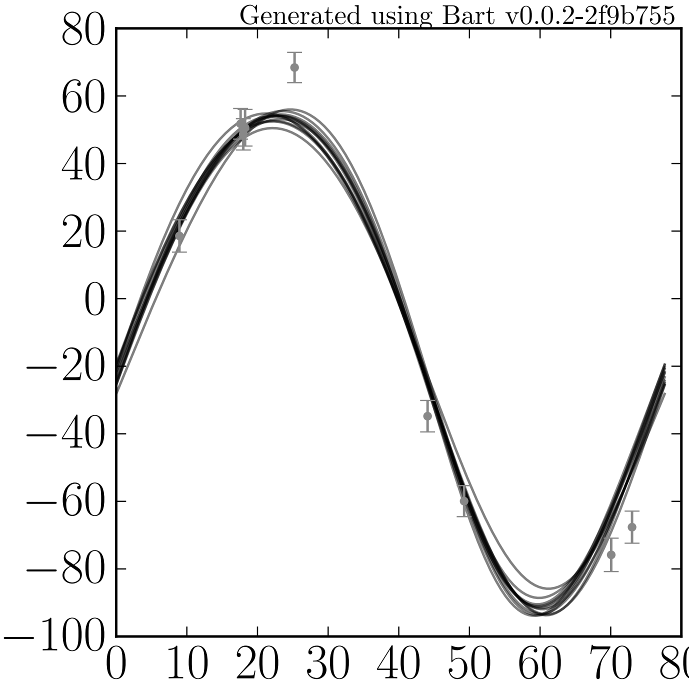


$\Gamma_{\rm p}$
$\Theta_\star$
$\Theta_{\rm p}$
$N_{{\rm p},n}$
$\star_n$
S
p_n^m
q_n^m
$\mathbf{X}_n^m$
planets m
$N_{{\rm p},n}^{\rm obs}$
$\mathbf{X}_n$
targets n

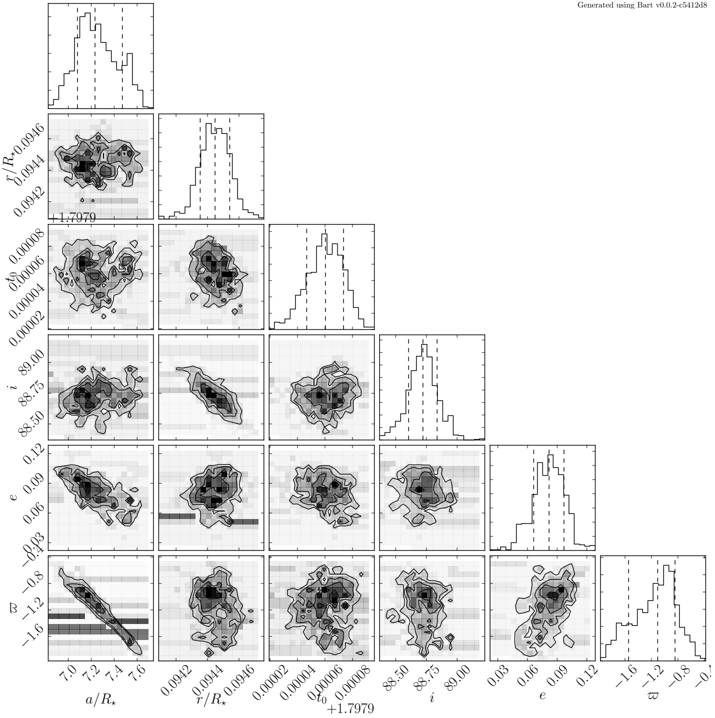

$M_\star$
t_0, a, e, ϖ, $\boldsymbol{\delta i}$, M_p
r
$R_\star$, $f_\star$, $\{I_k\}$
$\boldsymbol{x}(t)$
planets
i
t_n
δ^2
$f_{\boldsymbol{\theta}}(t_n)$
t_{exp_n}
$f_{\mathrm{obs};\,n}$
σ_n^2
observations

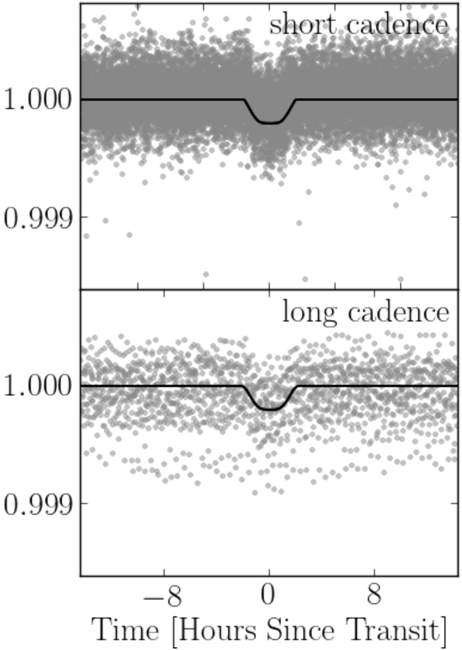
Generated using Bart v0.0.2-2f9b755

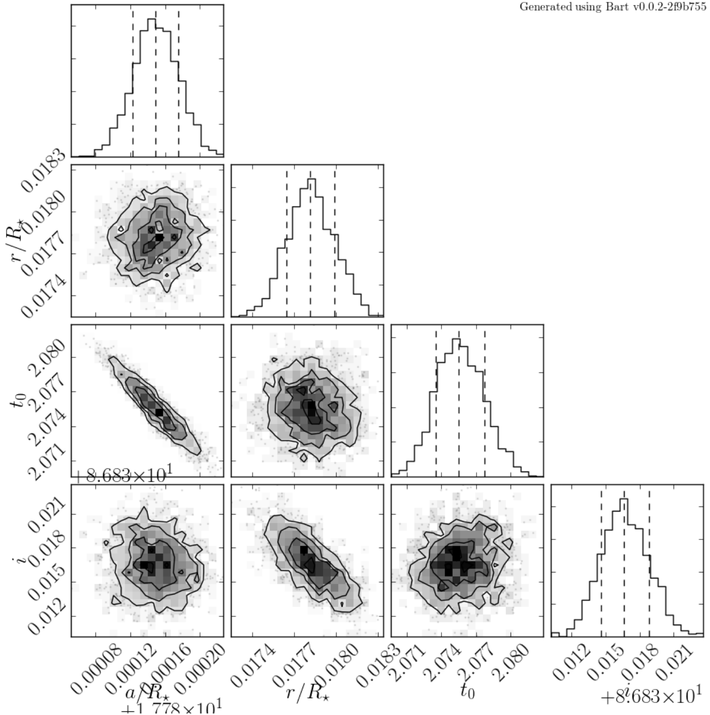




Generated using Bart v0.0.2-2f9b755
80
60
40
20
0
−20
−40
−60
−80
−100
0
10
20
30
40
50
60
70
80

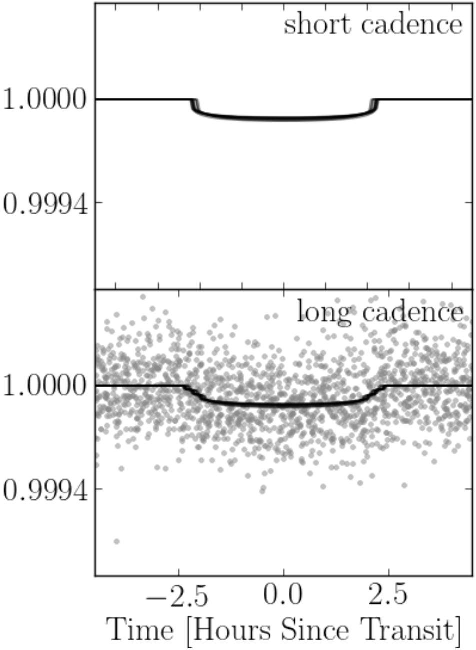

Generated using Bart v0.0.2-c5412d8
$a/R_\star$
$r/R_\star$
t_0
+1.7979
i
e
ϖ

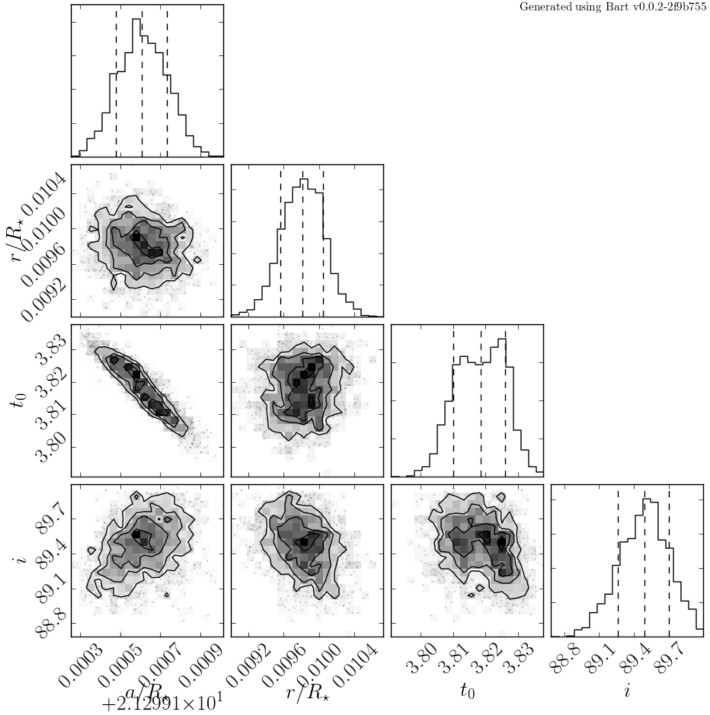

short cadence
long cadence
1.000
0.999
1.000
0.999
−8
0
8
Time [Hours Since Transit]

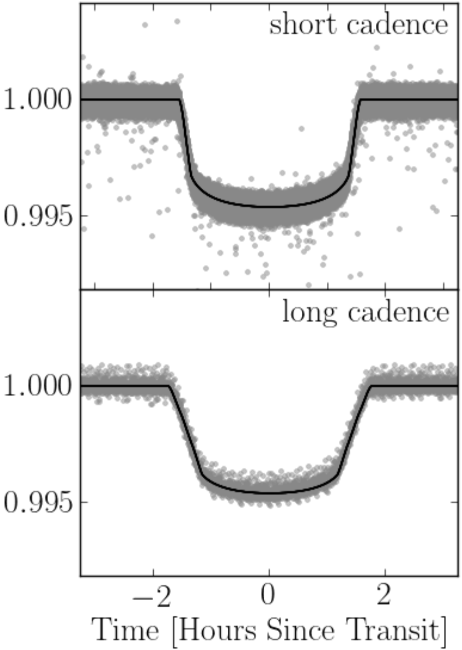


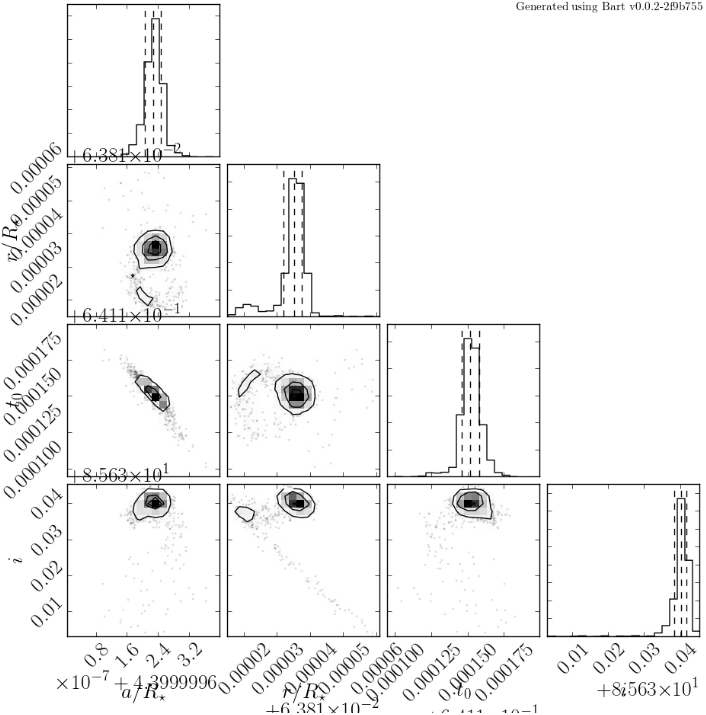
Generated using Bart v0.0.2-2f9b755
$r/R_\star$
t_0
i
$a/R_\star$
$+1.778\times10^1$
$+8.683\times10^1$
0.0183
0.0180
0.0177
0.0174
2.080
2.077
2.074
2.071
0.021
0.018
0.015
0.012
0.00008
0.00012
0.00016
0.00020

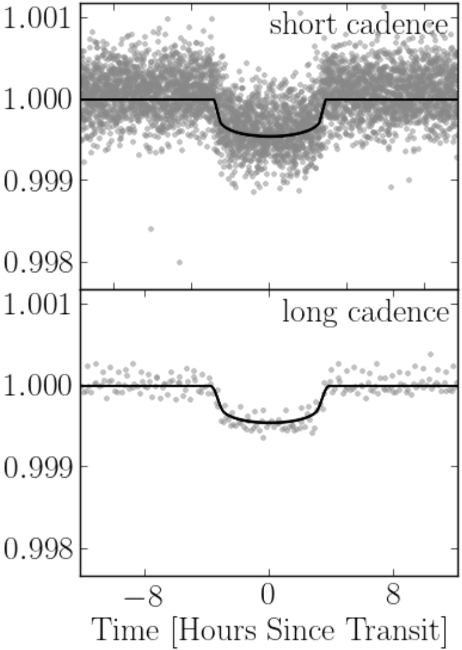
Generated using Bart v0.0.2-2f9b755

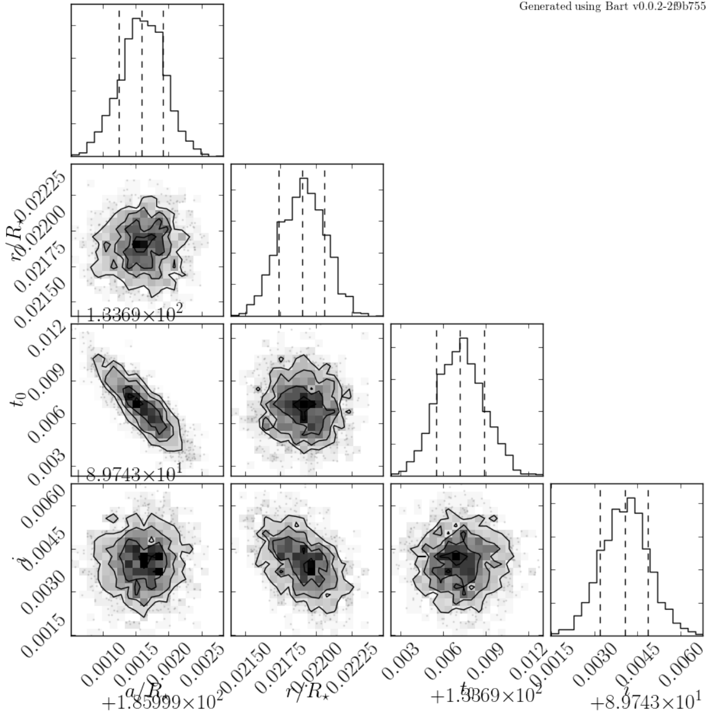



Generated using Bart v0.0.2-2f9b755
$r/R_\star$
0.0092
0.0096
0.0100
0.0104
t_0
3.80
3.81
3.82
3.83
i
88.8
89.1
89.4
89.7
0.0003
0.0005
0.0007
0.0009
$a/R_\star$
$+2.12991\times10^1$
0.0092
0.0096
0.0100
0.0104
$r/R_\star$
3.80
3.81
3.82
3.83
t_0
88.8
89.1
89.4
89.7
i


short cadence
long cadence
1.000
0.995
1.000
0.995
−2
0
2
Time [Hours Since Transit]


Generated using Bart v0.0.2-2f9b755
$r/R_\star$
t_0
i
$a/R_\star$
$\times10^{-7} + 4.3999996$
$+6.381\times10^{-2}$
$+6.411\times10^{-1}$
$+8.563\times10^{1}$


1.001
1.000
0.999
0.998
short cadence
1.001
1.000
0.999
0.998
long cadence
−8
0
8
Time [Hours Since Transit]



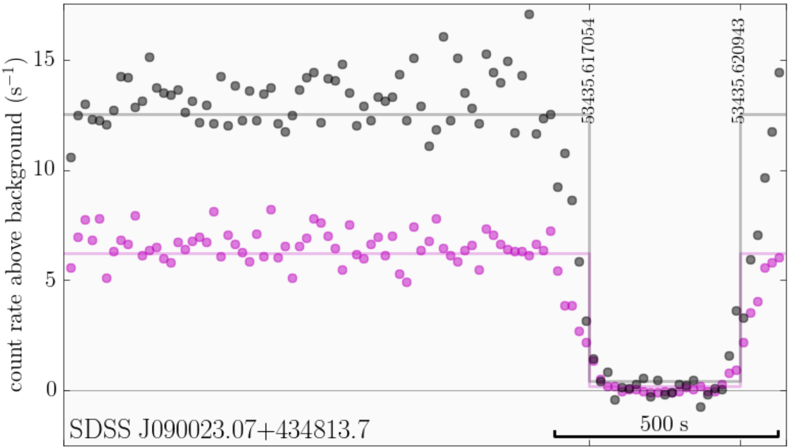
$r/R_\star$
0.02225
0.02200
0.02175
0.02150
$+1.3369\times10^2$
t_0
0.012
0.009
0.006
0.003
$+8.9743\times10^1$
i
0.0060
0.0045
0.0030
0.0015
0.0010
0.0015
0.0020
0.0025
$a/R_\star$
$+1.85999\times10^2$
0.02150
0.02175
0.02200
0.02225
$r/R_\star$
0.003
0.006
0.009
0.012
t_0
$+1.3369\times10^2$
0.0015
0.0030
0.0045
0.0060
i
$+8.9743\times10^1$

Bart (Foreman-Mackey *et al.*, forthcoming)

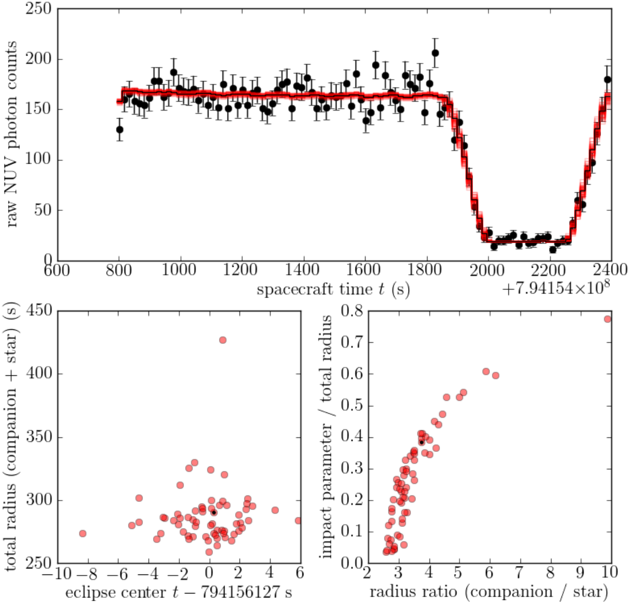
- built on very successful *emcee* package (Foreman-Mackey *et al.*, 1202.3665)
- designed for exoplanet measurement and discovery of false positives
- very easy to use

```
import bart

# Initialize a planet.
planet = bart.Planet(r=0.01, a=21.3, t0=3.85)
planet.parameters += [bart.parameters.Parameter(r"$r$", "r"),
                      bart.parameters.LogParameter(r"$a$", "a")]

# Initialize the star.
ldp = bart.kepler.fiducial_ldp(teff=6438, logg=4.28, feh=0.0)
star = bart.Star(mass=planet.get_mstar(12.4138), ldp=ldp)

# Set up the system.
system = bart.PlanetarySystem(star)
system.parameters.append(bart.parameters.CosParameter(r"$i$", "iobs"))
system.add_planet(planet)

# Add data and fit.
system.add_dataset(bart.KeplerDataset("path/to/kepler/data/lc.fits"))
system.fit(2000)
```

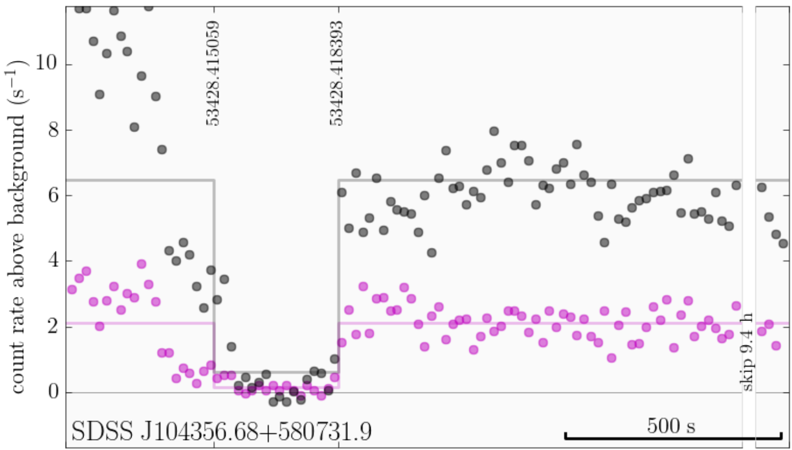
exoplanets around white dwarfs (Schiminovich, Lang, Hogg)



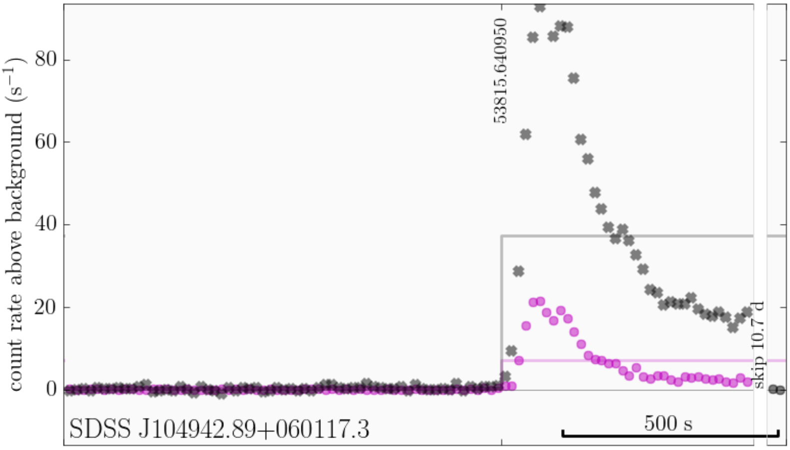
exoplanets around white dwarfs (Schiminovich, Lang, Hogg)



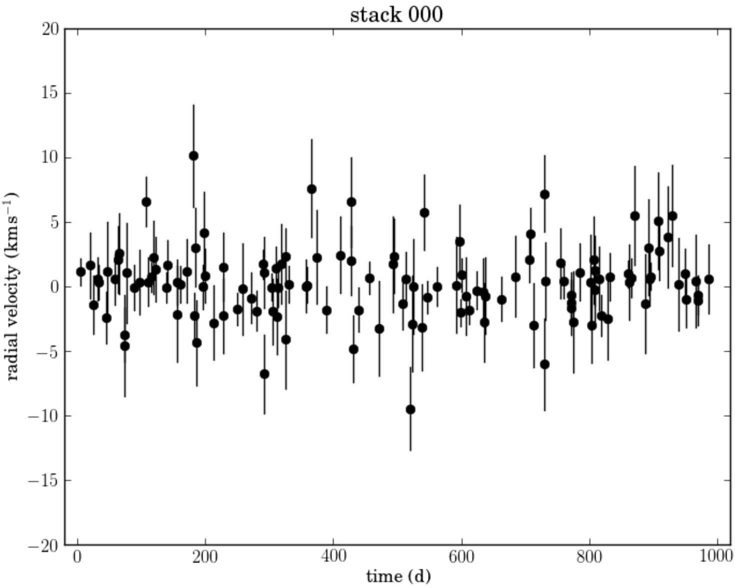
exoplanets around white dwarfs (Schiminovich, Lang, Hogg)



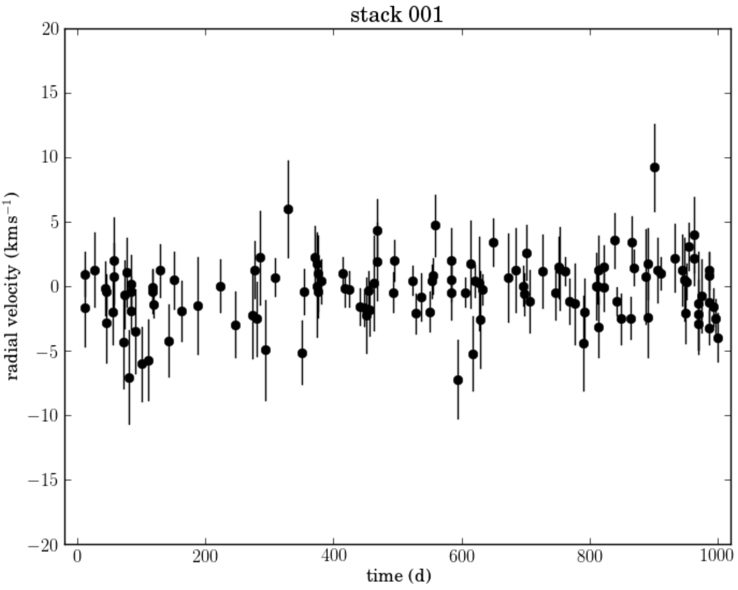
exoplanets around white dwarfs (Schiminovich, Lang, Hogg)

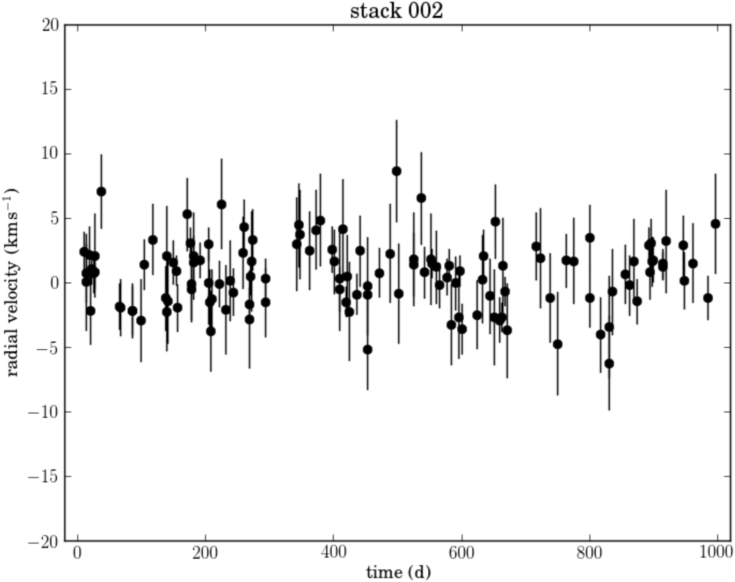


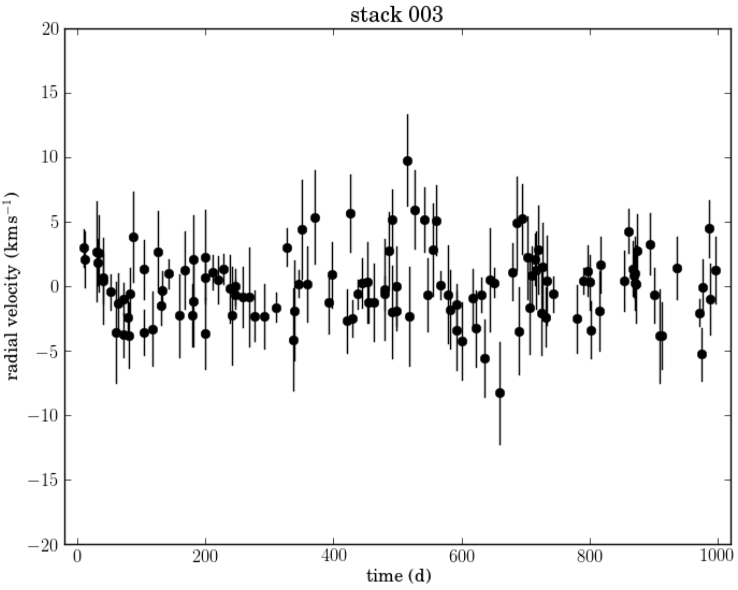
the undetectable can be measured



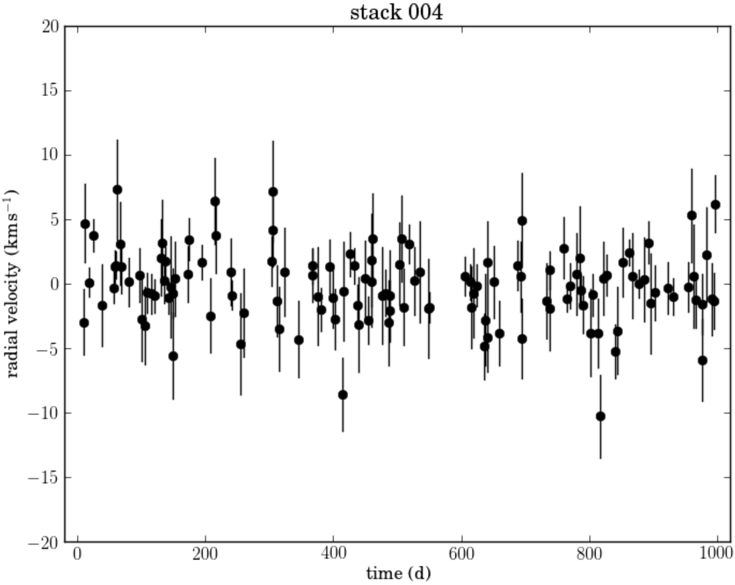
stack 000
radial velocity (kms^{-1})
time (d)
20
15
10
5
0
−5
−10
−15
−20
0
200
400
600
800
1000

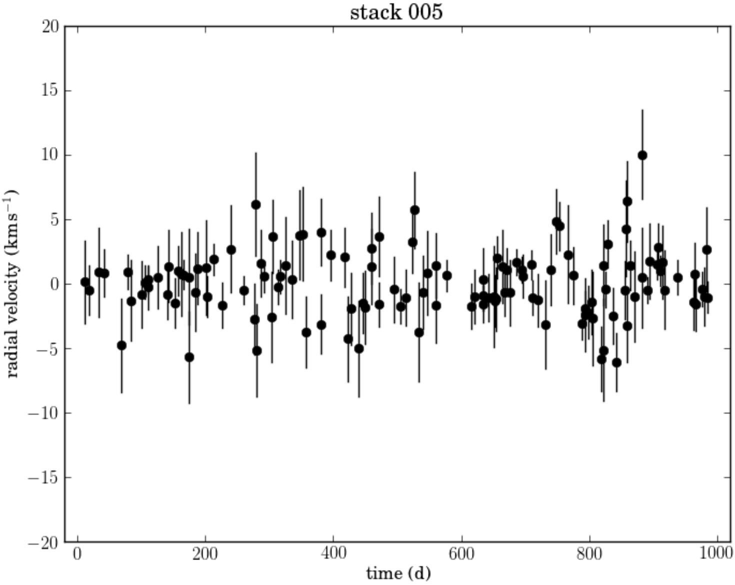

stack 001
radial velocity (kms^{-1})
time (d)
20
15
10
5
0
−5
−10
−15
−20
0
200
400
600
800
1000

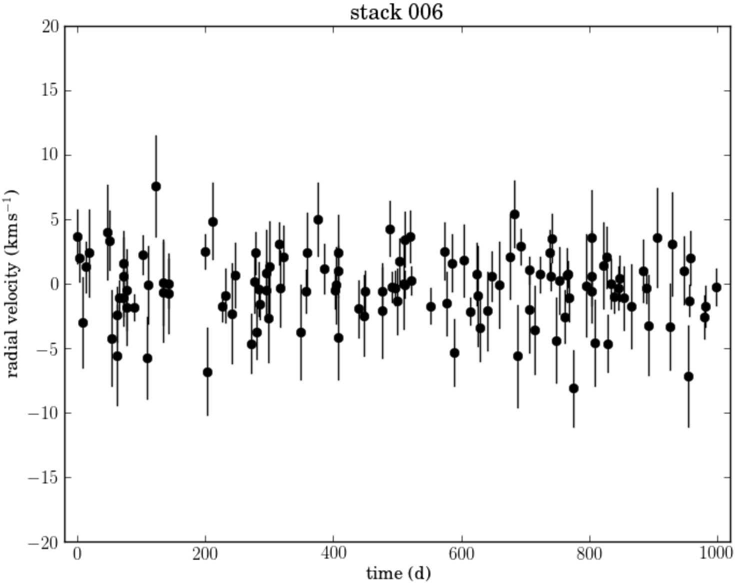

stack 002
radial velocity (kms^{-1})
time (d)
20
15
10
5
0
−5
−10
−15
−20
0
200
400
600
800
1000

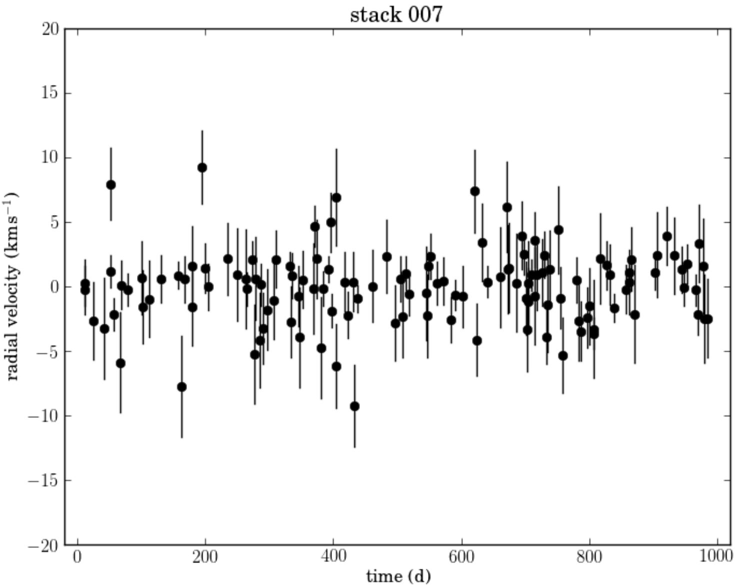


stack 003
radial velocity (kms^{-1})
time (d)
20
15
10
5
0
−5
−10
−15
−20
0
200
400
600
800
1000

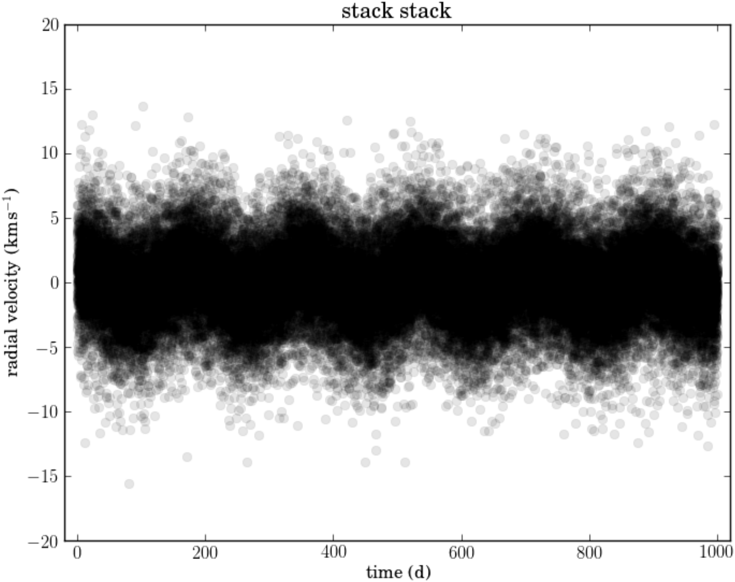

stack 004
radial velocity (kms^{-1})
time (d)
20
15
10
5
0
−5
−10
−15
−20
0
200
400
600
800
1000


stack 005
radial velocity (kms^{-1})
time (d)
20
15
10
5
0
−5
−10
−15
−20
0
200
400
600
800
1000

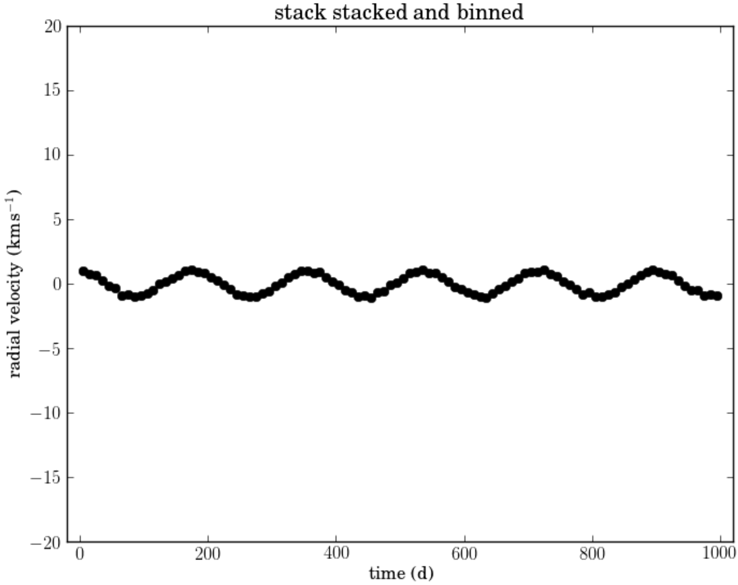

stack 006
radial velocity (kms^{-1})
time (d)
20
15
10
5
0
−5
−10
−15
−20
0
200
400
600
800
1000



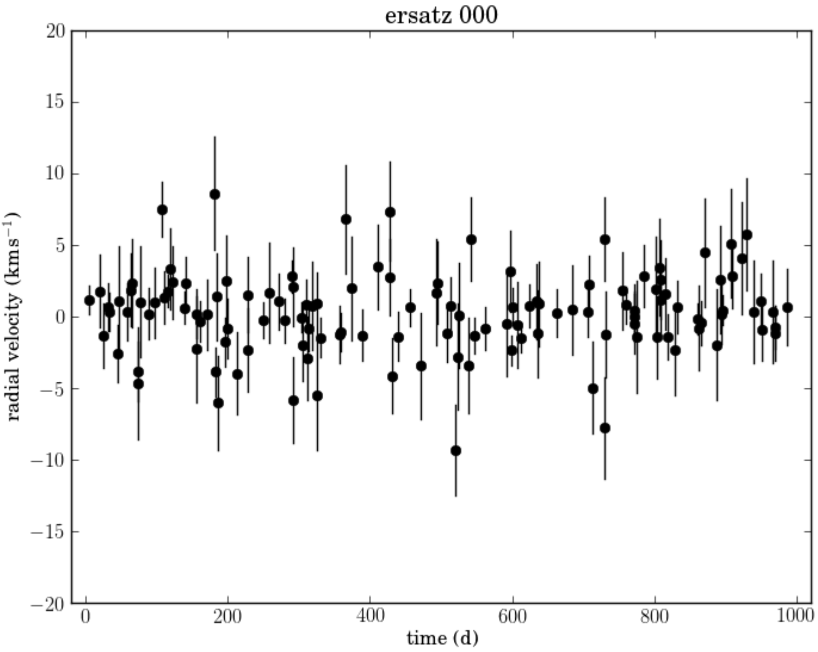
stack 007
20
15
10
5
0
−5
−10
−15
−20
radial velocity (kms^{-1})
0
200
400
600
800
1000
time (d)



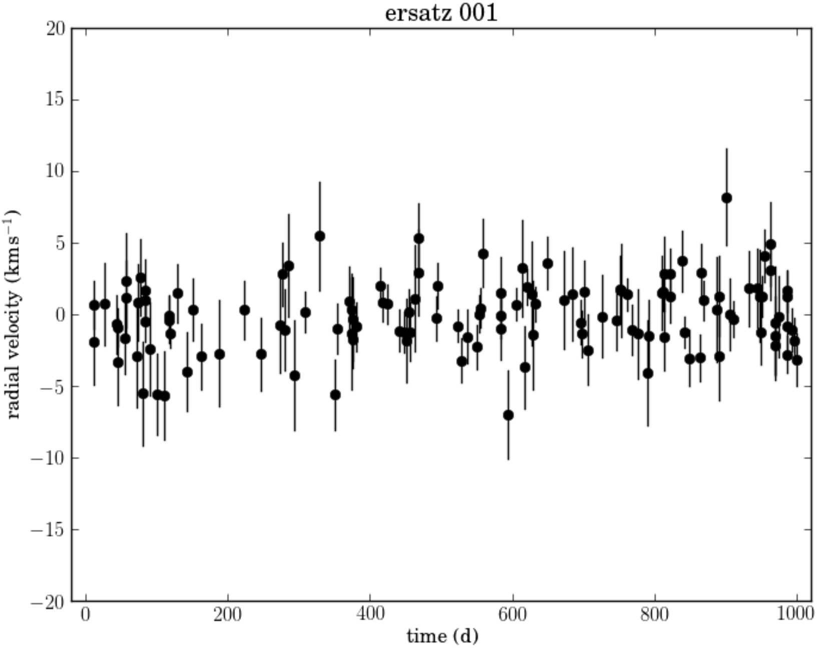
stack stack
radial velocity (kms^{-1})
time (d)
20
15
10
5
0
−5
−10
−15
−20
0
200
400
600
800
1000



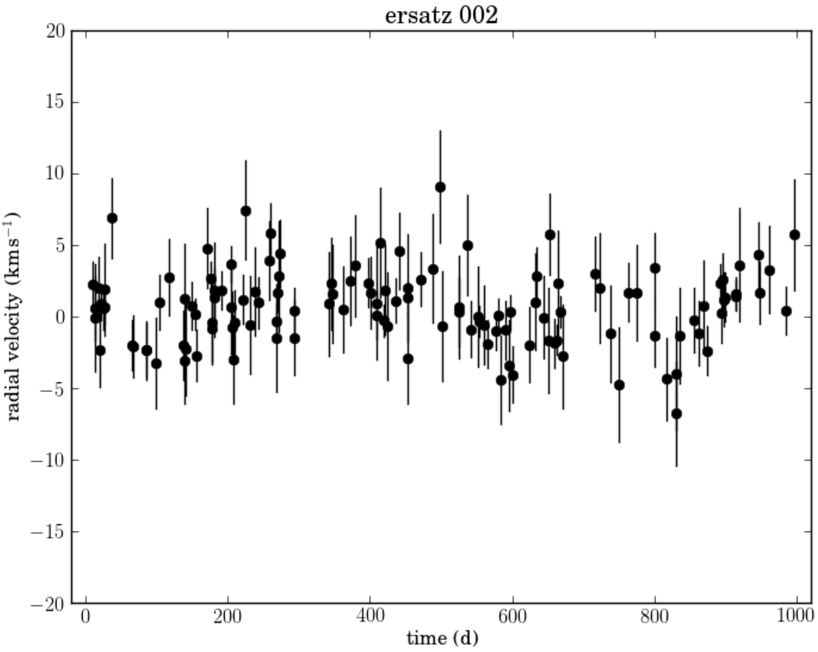
stack stacked and binned
20
15
10
5
0
−5
−10
−15
−20
radial velocity (kms^{-1})
0
200
400
600
800
1000
time (d)



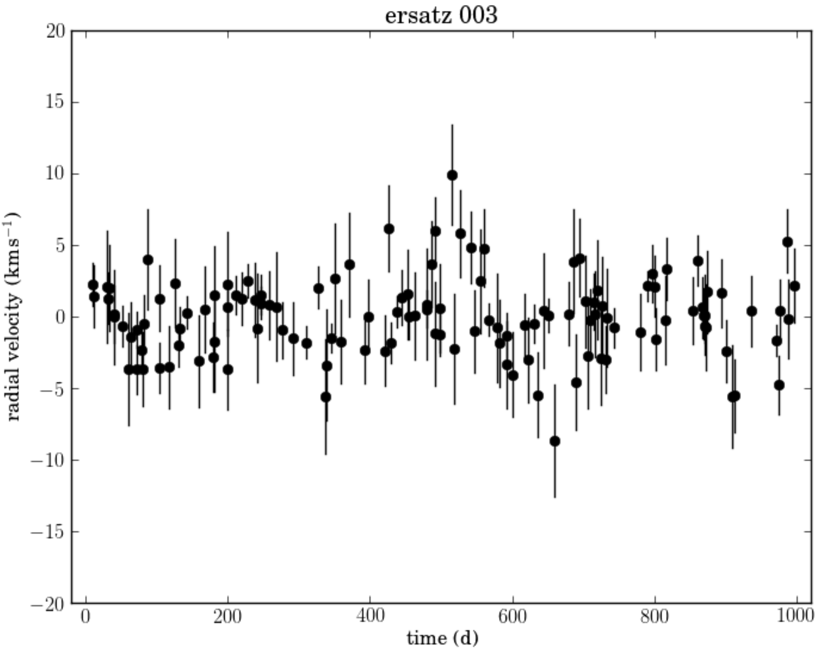
ersatz 000
radial velocity (kms^{-1})
time (d)
20
15
10
5
0
−5
−10
−15
−20
0
200
400
600
800
1000

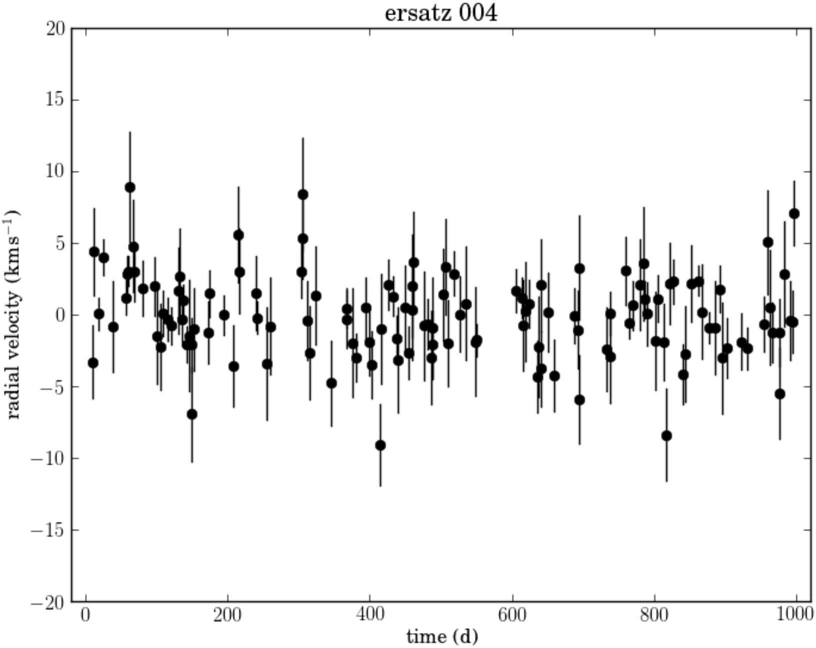


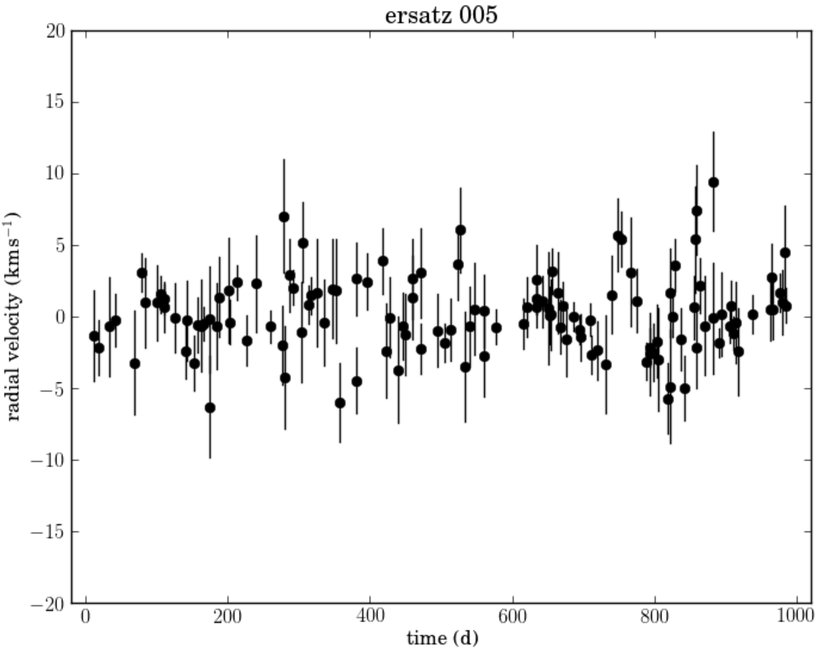
ersatz 001
radial velocity (kms^{-1})
time (d)
20
15
10
5
0
−5
−10
−15
−20
0
200
400
600
800
1000



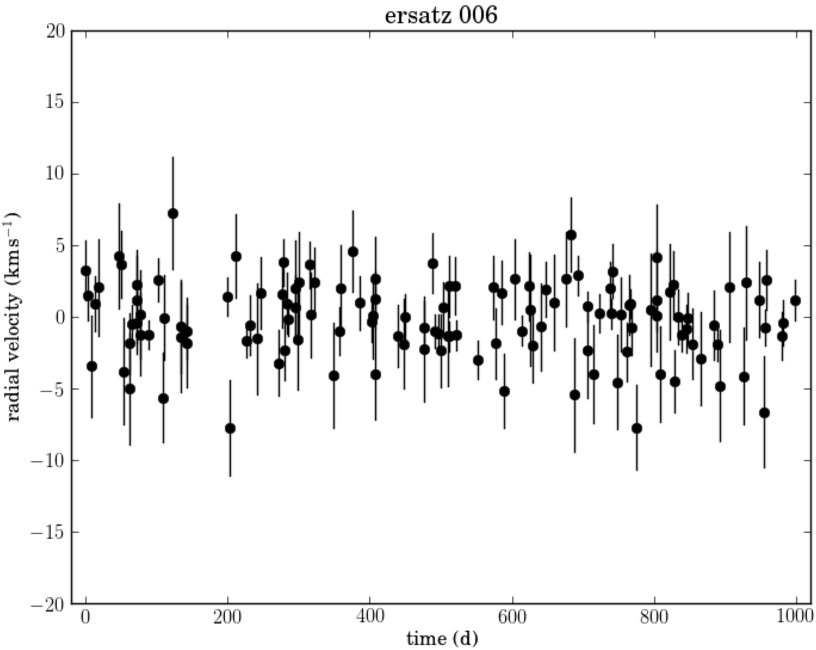
ersatz 002
radial velocity (kms^{-1})
time (d)
20
15
10
5
0
−5
−10
−15
−20
0
200
400
600
800
1000


ersatz 003
radial velocity (kms^{-1})
time (d)
20
15
10
5
0
−5
−10
−15
−20
0
200
400
600
800
1000

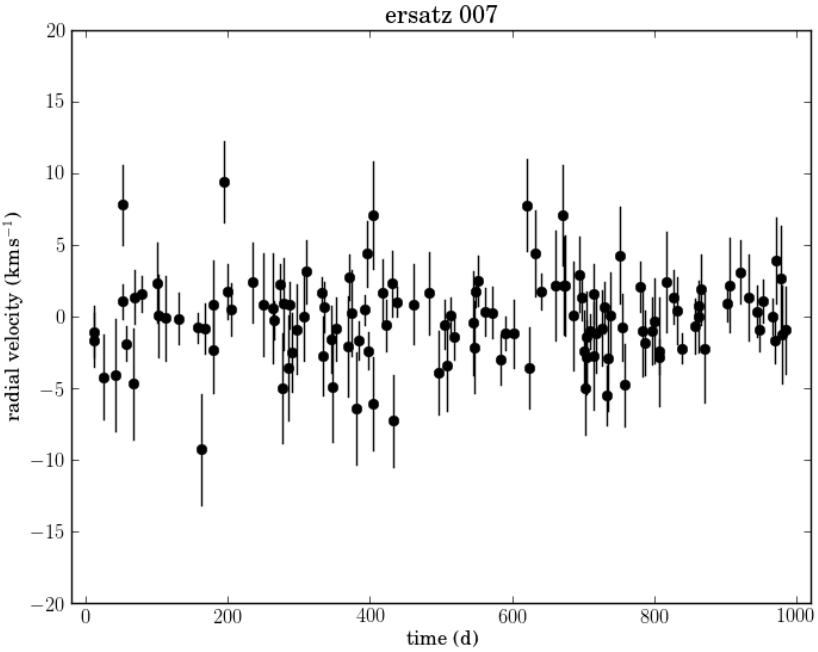

ersatz 004
radial velocity ($\mathrm{km s^{-1}}$)
time (d)
20
15
10
5
0
−5
−10
−15
−20
0
200
400
600
800
1000

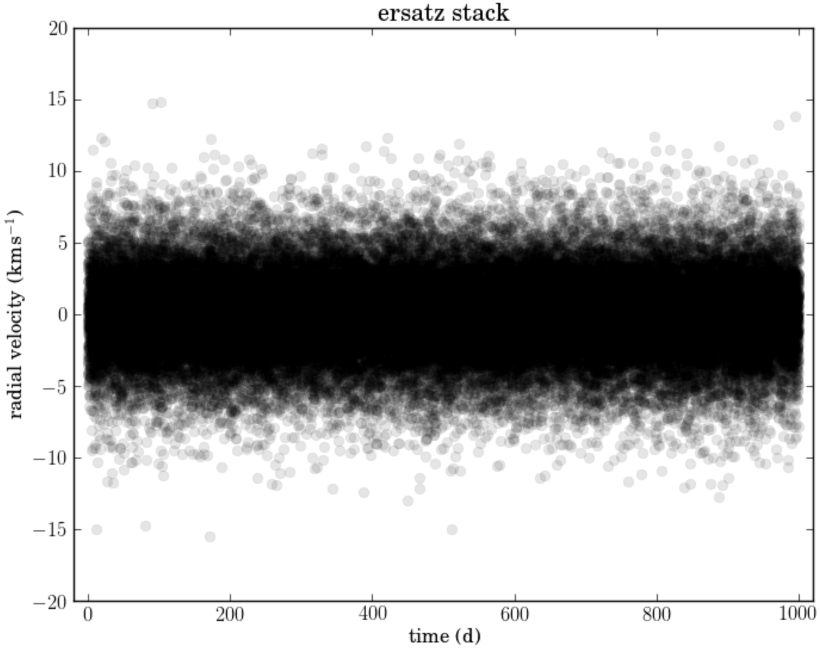


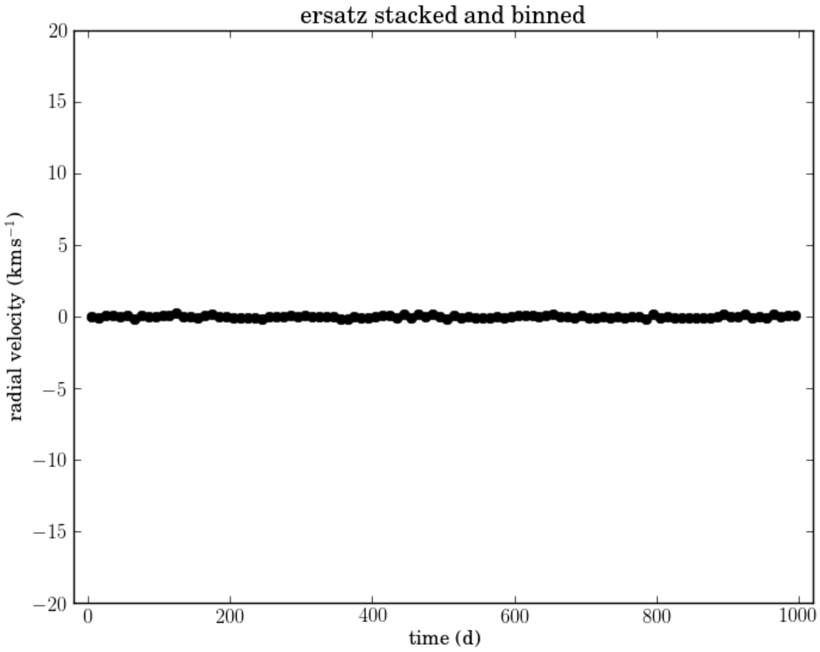
ersatz 005
radial velocity (kms^{-1})
time (d)
20
15
10
5
0
−5
−10
−15
−20
0
200
400
600
800
1000



ersatz 006
radial velocity (kms^{-1})
time (d)
20
15
10
5
0
−5
−10
−15
−20
0
200
400
600
800
1000


ersatz 007
radial velocity (kms^{-1})
time (d)
20
15
10
5
0
−5
−10
−15
−20
0
200
400
600
800
1000


ersatz stack
radial velocity (kms^{-1})
time (d)
20
15
10
5
0
−5
−10
−15
−20
0
200
400
600
800
1000

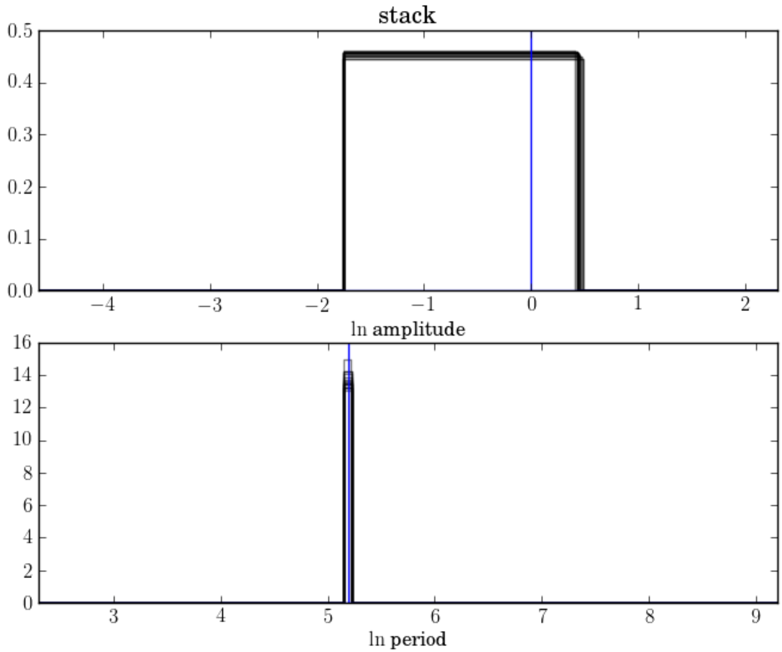

ersatz stacked and binned
radial velocity (kms^{-1})
time (d)
20
15
10
5
0
−5
−10
−15
−20
0
200
400
600
800
1000

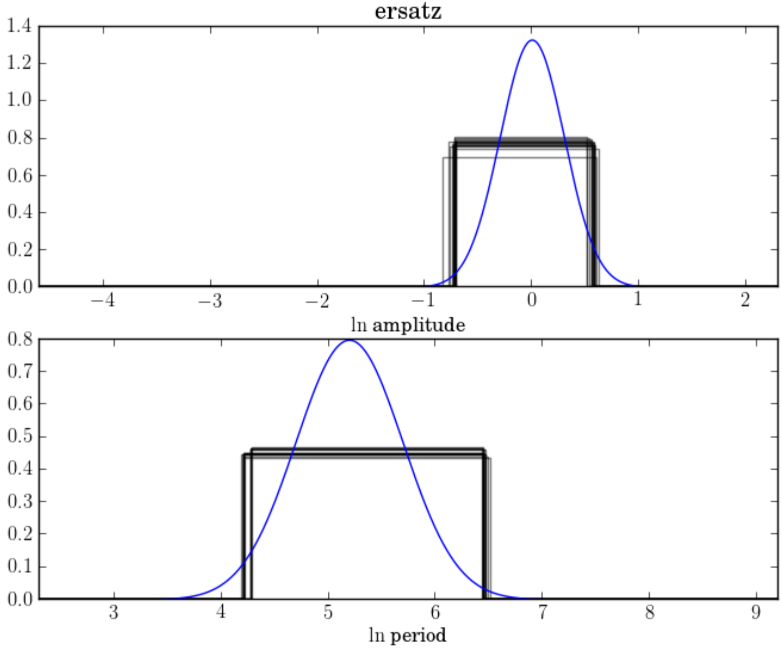
hierarchical population detection

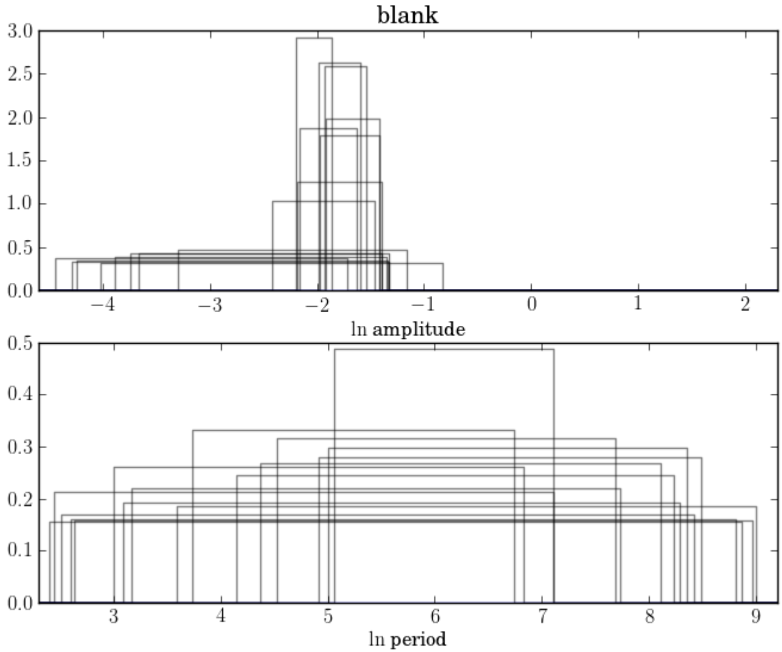
- family of priors $p(\boldsymbol{\omega}_n|\boldsymbol{\alpha})$, parameterized by some $\boldsymbol{\alpha}$
-

$$p(\{\mathbf{D}_n\}_{n=1}^{N}|\boldsymbol{\alpha}) = \prod_{n=1}^{N}\int \mathrm{d}\boldsymbol{\omega}_n\, p(\mathbf{D}_n|\boldsymbol{\omega}_n)\, p(\boldsymbol{\omega}_n|\boldsymbol{\alpha}) \quad (1)$$

- if you believe there can be a **likelihood**, then you believe there can be a **marginalized likelihood**
- the fact that each internal $p(\mathbf{D}_n|\boldsymbol{\omega}_n)$ contains no clear peak (no clear object detection at all) doesn't change anything!


stack
0.5
0.4
0.3
0.2
0.1
0.0
−4
−3
−2
−1
0
1
2
ln amplitude
16
14
12
10
8
6
4
2
0
3
4
5
6
7
8
9
ln period


ersatz
1.4
1.2
1.0
0.8
0.6
0.4
0.2
0.0
−4
−3
−2
−1
0
1
2
ln amplitude
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
3
4
5
6
7
8
9
ln period


blank
3.0
2.5
2.0
1.5
1.0
0.5
0.0
−4
−3
−2
−1
0
1
2
ln amplitude
0.5
0.4
0.3
0.2
0.1
0.0
3
4
5
6
7
8
9
ln period

hierarchical inference: What does it require?

- accurate likelihood functions
  - accurate noise models, or **parameterized** noise models
- fast inference
  - self-tuning MCMC (like *emcee*; Foreman-Mackey *et al.*, 1202.3665)
  - robustness to multimodal likelihood functions
- concept of self-calibration
  - calibration and noise parameters are not different from astrophysical parameters
- racks and racks of metal
  - (it can't be done in "map–reduce" framework)

hierarchical inference: Why does it work?

- The marginalized likelihood is large when there is high prior probability in locations where there is high likelihood.
- When likelihoods are broad, the best prior is the most concentrated prior that is "consistent with" **all** individual-object likelihood functions.
- The operation is a **heteroskedastic deconvolution**.
  - (in modern parlance, a "deconvolution" is always the result of fitting a generative or forward model)

conclusions

- software, software, software, and applied math
- lots of bad data is equal to a bit of good data
- modeling beats supervised classification
- point estimates are bad, models are good
- heirarchical modeling will require MCMC
- we can find populations, no member of which is individually detectable