Experience from Double-Blind Test

Wesley A. Traub
Jet Propulsion Laboratory
California Institute of Technology

NEAT 2013 Simulation Workshop
Grenoble
27-28 May 2013
Context

In 2008-2009, the NASA Exoplanet Exploration Program ran a simulation in support of the Space Interferometry Mission to answer these questions:

(1) Can an Earth-like planet (terrestrial mass, habitable zone) be detected in a multi-planet system, using astrometry and RV?

(2) How does the presence of other planets affect the detectability of an Earth-like planet?

Short answers:
(1) Yes.
(2) Very little.

References:
Organization of the 1-Year Study

- **Planet Modeling**: Five teams each generated 100 candidate planetary systems (total 500 systems).
- **Data Simulation**: NASA Exoplanet Science Institute (NExScI) team generated simulated astrometry & RV data.
- **Analysis Contestants**: 5 analysis teams, 4 competitively-selected, analyzed simulated data to find the planets.
- **Scoring**: Exoplanet Program Chief Scientist led a synthesis & scoring team. Scores based on each team’s results compared to Cramer-Rao uncertainty of minimum uncertainty in each parameter.
- **Oversight**: A NASA-appointed External Independent Readiness Board (EIRB) monitored the process, performed independent checks, and advised NASA HQ.
Expected error bars

- Expected variance (Cramer-Rao minimum variance bound) for each parameter calculated, using Fischer information matrix.

- All correlations between parameters automatically included in each error bar.

- Astro and RV data all included together.

Criteria for correct solution

- Rule: parameter error = true minus fitted parameter value should be within a factor of 3 times the Cramer-Rao (minimum-variance) bound.

- Rule is modified as follows:
  - Low SNR: factor of 3 above is replaced with 3*SNR/5.8, to avoid allowing random false detections to count as good
  - High SNR: period error need not be less than 0.5% and mass errors need not be less than 1%

- Criteria are applied to mass and period errors
Teams for SIM DB Exercise

**Planets: Team A**
- A-1: Eric Ford, U. Florida,
- A-2: Greg Laughlin, UCSC
- A-3: Hal Levison, SRI
- A-4: Doug Lin, UCSC

**Data Simulation: Team B**
- Andy Boden, MSC
- Valeri Makarov, MSC

**Analysis: Team C**
- C-1: Stefano Casertano, STScI
- C-2: Debra Fischer, SFSU
- C-3: Jeremy Kasdin, Princeton
- C-4: Matt Muterspaugh, Berkeley
- C-5: Mike Shao, JPL

**Scoring: Team D**
- Wes Traub, JPL
- Alan Boss, Carnegie
- Andy Gould, Ohio State
- Angelle Tanner, MSC
- Chas Beichman, MSC

One each from Team A, B, C.
Exoplanet Mass vs Period

Discovery space is above each curve.

- Planets ~ Tidally Locked
- Terrestrial Planets
- Habitable-Zone Planets around Sun-like stars
- Range of SIM Target Stars

GAIA (70 μas, 50pc, 100epochs/5yrs)
Discovery space

Results of Phase II double-blind test, 60 best SIM-Lite targets

- Detected planet
- Missed or bad detection

Planet mass (Earths)

Orbital period (years)
Planet Multiplicity

70 planets should have been detectable by SNR and P/T criteria.
Period errors

True period vs. fitted period

![Graph showing the relationship between true period and fitted period.](image)
Period errors vs. SNR
Mass errors

True mass vs. fitted mass

The graph shows a scatter plot of true mass against fitted mass on a log-log scale. The data points are closely aligned along a diagonal line, indicating a strong linear relationship between the true mass and the fitted mass.
Mass errors vs. SNR

Mass fit errors vs SNR

Fractional mass error = \frac{(m_{\text{true}} - m_{\text{fitted}})}{m_{\text{true}}}

SNR

Fractional error
Inclination Errors

True inc vs. fitted inc

Fitted inc

True inc
Inclination errors vs. SNR
Outliers generally have near–zero eccentricity. Near-edge-on inclination can be confused with high eccentricity.
 Completeness vs SNR

- Completeness = detected / detectable planets.
- Curve is theoretical for 1% FAP (Catanzarite et al. 2006).
- At SNR > 5.8, measured completeness is excellent, as predicted.
- SNR is the RSS of RV & Astro SNRs.

50% completeness at SNR ~ 5.8, by theory as well as experiment.
### Summary Statistics for SIM

<table>
<thead>
<tr>
<th>Scoring Category</th>
<th>Part 1</th>
<th>Part 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Completeness: Terrestrial</td>
<td>18/20 = 90%</td>
<td>37/43 = 86%</td>
</tr>
<tr>
<td>Completeness: HZ</td>
<td>13/13 = 100%</td>
<td>21/22 = 95%</td>
</tr>
<tr>
<td>Completeness: Terrestrial HZ</td>
<td>9*/9 = 100%</td>
<td>17**/18 = 94%</td>
</tr>
<tr>
<td>Completeness: All planets</td>
<td>51/54 = 94%</td>
<td>63/70 = 90%</td>
</tr>
<tr>
<td>Reliability: Terrestrial</td>
<td>25/27 = 93%</td>
<td>38/39 = 97%</td>
</tr>
<tr>
<td>Reliability: HZ</td>
<td>16/16 = 100%</td>
<td>20/20 = 100%</td>
</tr>
<tr>
<td>Reliability: Terrestrial HZ</td>
<td>12/12 = 100%</td>
<td>16/16 = 100%</td>
</tr>
<tr>
<td>Reliability: All planets</td>
<td>64/67 = 96%</td>
<td>66/68 = 97%</td>
</tr>
</tbody>
</table>

- Analysts were asked to be aggressive in Part-1 and conservative in Part-2.
- This is reflected in the denominators of the Completeness and Reliability sections.
* All 9 T/HZ Part-1 detected planets were in multiple-planet systems.
** 10 of the 17 T/HZ Part-2 detected planets were in multiple-planet systems
Potential noise models

- Gaussian noise: random values, normal distribution
  - Wiener process (Brownian motion): Gaussian, but not stationary
  - Ornstein-Uhlenbeck process: Gaussian, stationary
  - Brownian bridge: Gaussian, but correlated increments
  - fractional Brownian motion: Gaussian, generalized Wiener

**DBS Planetary-Systems**

- Two phases:
  
  - **Phase-1**: 48 planetary systems, all 1-Sun @ 10 pc
    - 32 from Modeled systems; 8 Solar-system-analogs;
      - 4 single terrestrial in HZ; 4 no-planets,
    - Analysis teams competed with each other.

  - **Phase-2**: 60 planetary systems around candidate nearby FGK target stars (real potential target stars).
    - All 60 scaled for stellar luminosity.
    - Systems contained 1 to 11 planets.
    - Analysis teams jointly selected each best solution.
Planet Detection Criteria

- **Before “answers” revealed:**
  - Each Team C used their own criteria to generate their own solutions; targeted < 1% FAP.
  - Team C’s used a “judgment based” approach to combine their results into a single recommendation.

- **After “answers” revealed:**
  - There were 12 trend planets
  - 11 of these corresponded to actual planets and would have been counted as ‘correctly detected’ had they been scored.

- **6 missed detections**
  - One had $e = 0.93$
  - One planet was at twice the frequency of another planet in the system which was detected; the missed planet may have been absorbed in the $2^{nd}$ harmonic of the detected planet.
Conclusion

• We set up a double-blind test for Earths in multi-planet systems.

• Results are excellent
  – average completeness 92%
  – average reliability 96%

• We find that planets in multiple planet systems are in most cases no harder to detect than if they were isolated.

• RV is an important adjunct to the astro data set.
Lessons learned

• Cramer-Rao minimum variance methodology is straight-forward
• Accuracy of fitted results can be easily compared to C-R predictions

• Noise model is a critical assumption
• Noise model should be validated by lab experiments

• Bias mechanisms lurk everywhere
• Biases should be listed early on, and controls developed
• Biases should be addressed in the lab
Thank you!