Artificial Intelligence Applications in Eclipsing Binaries

E J. Devinney, A. Prsa, E. F. Guinan, M. Degeorge
Villanova University, Villanova, PA

Supported by NSF/RUI Grant No. AST-05-07542f to Villanova University

Talk Topics - 1

1. **A very short intro to Artificial Intelligence (AI)**

2. The EBAI Project: Solving light curves from massive EB datasets with Artificial Neural Networks, and

3. Exploring EBAI results using Cluster Analysis

EB dataset: OGLE II LMC data.

AI - the idea

- Originally – to create machines that think like humans, but better.

  John McCarthy, Stanford University, 1959.

- Today – to attack diverse problems of high complexity.

- AI has developed along two branches:
  a) Knowledge Representation/Manipulation
  b) Algorithmic Methods (“machine learning”) <= this talk

(a) Knowledge Representation

- Capturing knowledge* - various approaches:

  - *RULES* — *Expert Systems*
    IF/THEN
  - *RELATIONSHIPS* — *Semantic Networks*
    JOE is-a BOY
  - *STRUCTURES* — *Frames (structs, objects)*
    Properties list, or slots

– * "knowledge" is explicit, narrow domain, captured by AI expert.
– representation includes methods for manipulating the captured knowledge

(b) Algorithmic Methods
= this talk =

- Algorithmic *machine learning* methods:

  - *Artificial Neural Networks (ANNs)*
  - *Clustering Methods*
  - *Self-Organizing Networks*
  - *…others…*

  – Knowledge is implicit, learning is “unsupervised” – no guidance from a person.

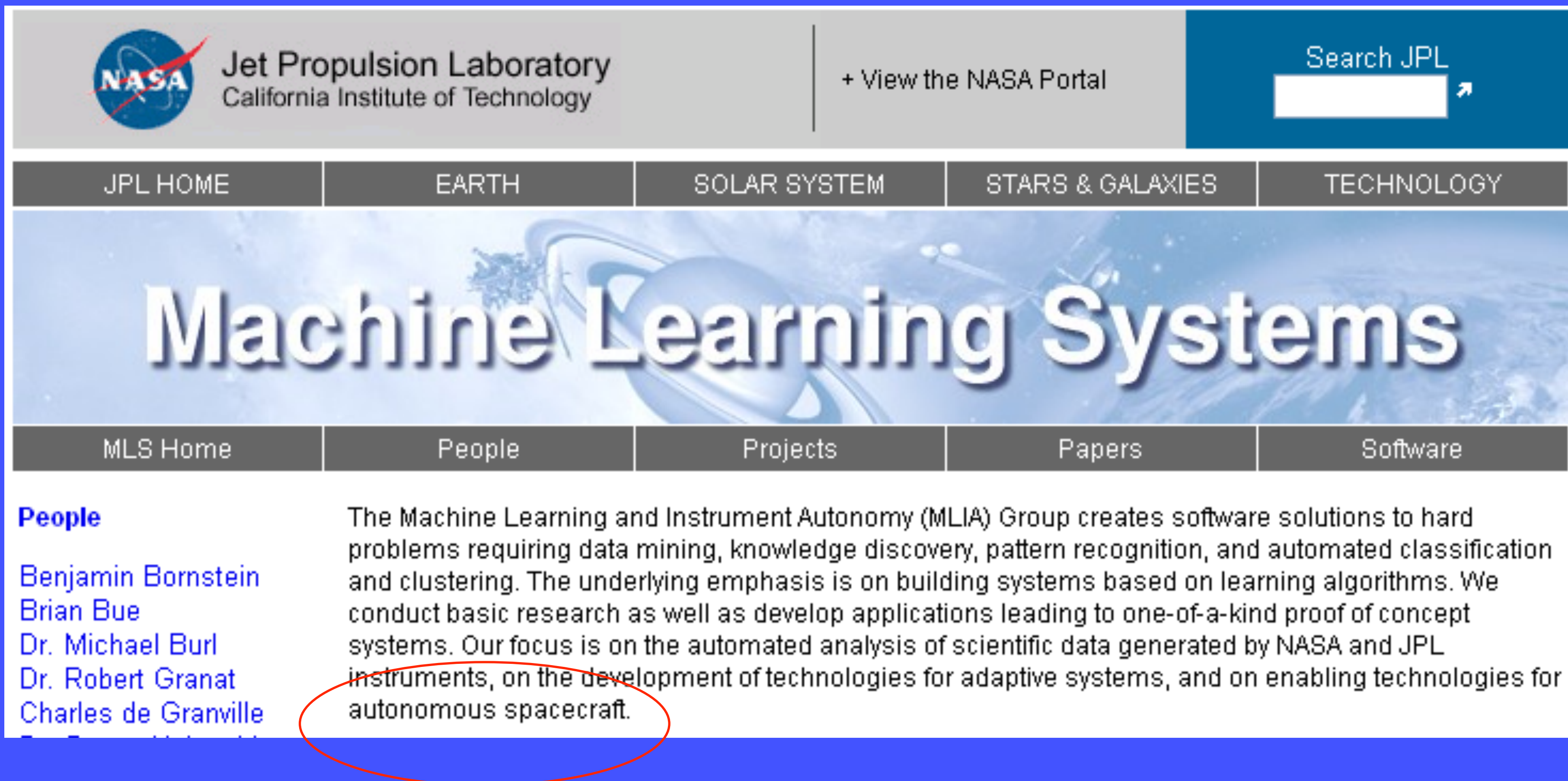
AI in Astronomy

A few examples

- Hubble: optimizing scheduling of observing programs, resolving conflicting needs.
- Autonomous docking in space demonstrated.
  New Scientist, May 2007.
- Planet hunt via robotic telescopes linked with advanced software. Telescopes “bid” to carry out follow-up observations; they are “scolded” if they don’t deliver. *arXiv: 0802.0431*
- NASA Rovers will get a software upgrade to allow them to make 'intelligent' (autonomous) decisions in the study of Martian clouds and dust devils. *BBC News (May 28, 2006).*
- Automated spectral classification via expert systems and neural networks. A. Rodriguez et al. *IEEE, (2004)*

AI in Astronomy

NASA JPL Machine Learning



Talk Topics - 2

1. A very short intro to Artificial Intelligence (AI)
2. **The EBAI Project: Solving light curves from massive EB datasets with Artificial Neural Networks**, and
3. Exploring EBAI results using Cluster Analysis

EB dataset: OGLE II LMC data.

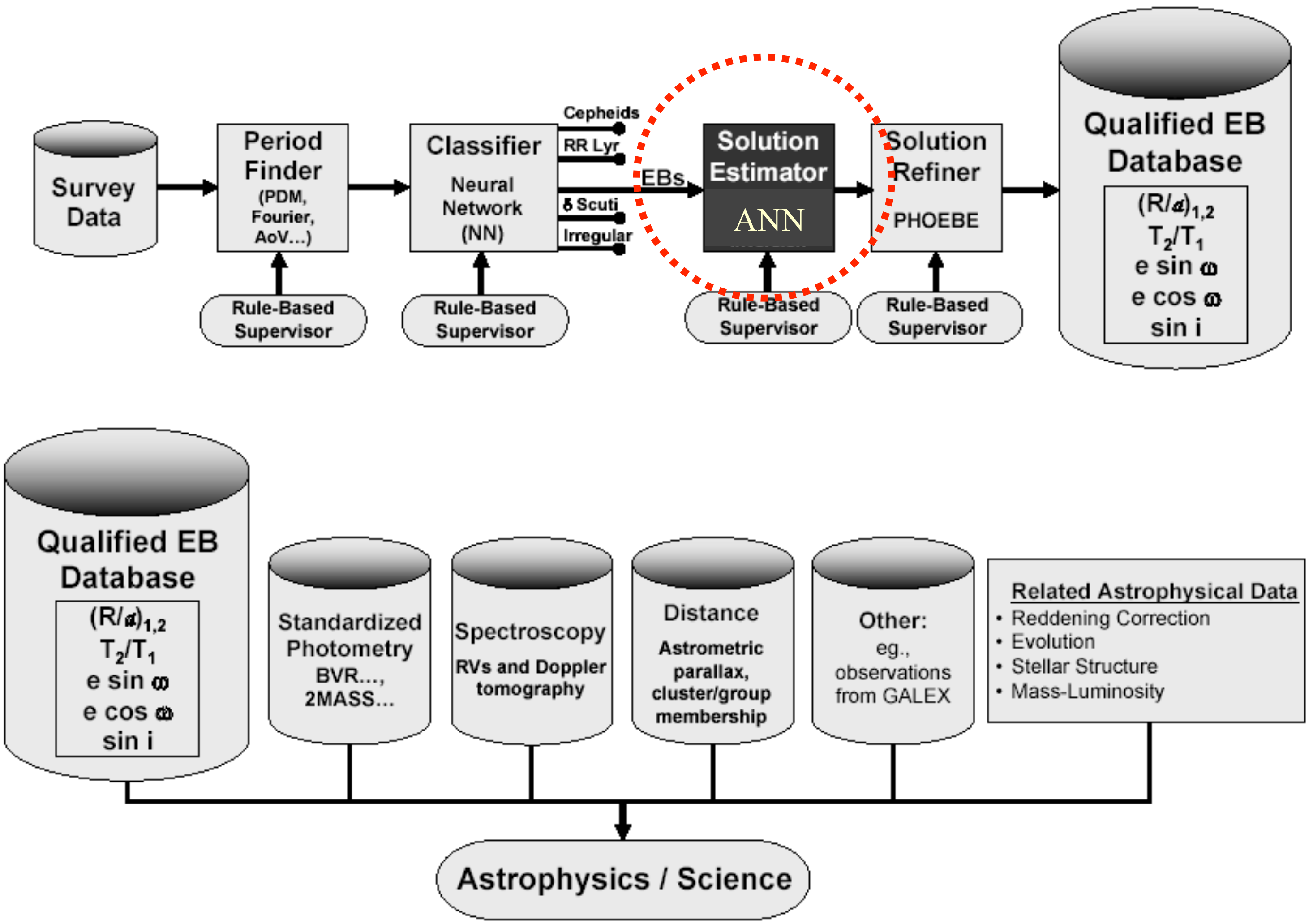
AI for Light Curve Solution Estimation

Problem context and goal

- Ground- and space-based survey observatories will provide a "firehose" of variable star data in the coming decade.
- How to maximize the scientific return from this data for Eclipsing Binaries?
- Humans can't interact with each EB to obtain a starting light curve solution; an automated approach for light curve solution (solution estimation) is required.

Binaries – reduction pipeline

The Solution Estimation Task



The EBAI Project

Artificial Intelligence Approach to the Determination of Physical Properties of Eclipsing Binaries: the EBAI* Project

Prsa, E. F. Guinan, E. J. Devinney, M. DeGeorge, D. H. Bradstreet, J. M. Giammarco, C. R. Alcock, S. G. Engle

* ***E****clipsing* ***B****inaries* *using* ***A****rtificial* ***I****ntelligence*

Artificial Neural Networks (ANNs)

Basic idea

- A computational entity biologically inspired by brain studies which can *learn*.
- ANNs have an INPUT layer, an OUTPUT layer, and a selected number of intermediate "HIDDEN" layers.
- WEIGHTS, or connection strengths, link the layers. What the network does – how it relates a given input to an output – is determined by the values of the weights.

Artificial neural networks (ANNs)

- In general, ANNs map a multidimensional INPUT vector to multidimensional OUTPUT vector.

  - **For EB application:**
    INPUT light curve data; OUTPUT "WD" model parameters

- ANNs are also good interpolators and extrapolators.
  - Provide sensible outputs for data not previously seen.

ANN topology

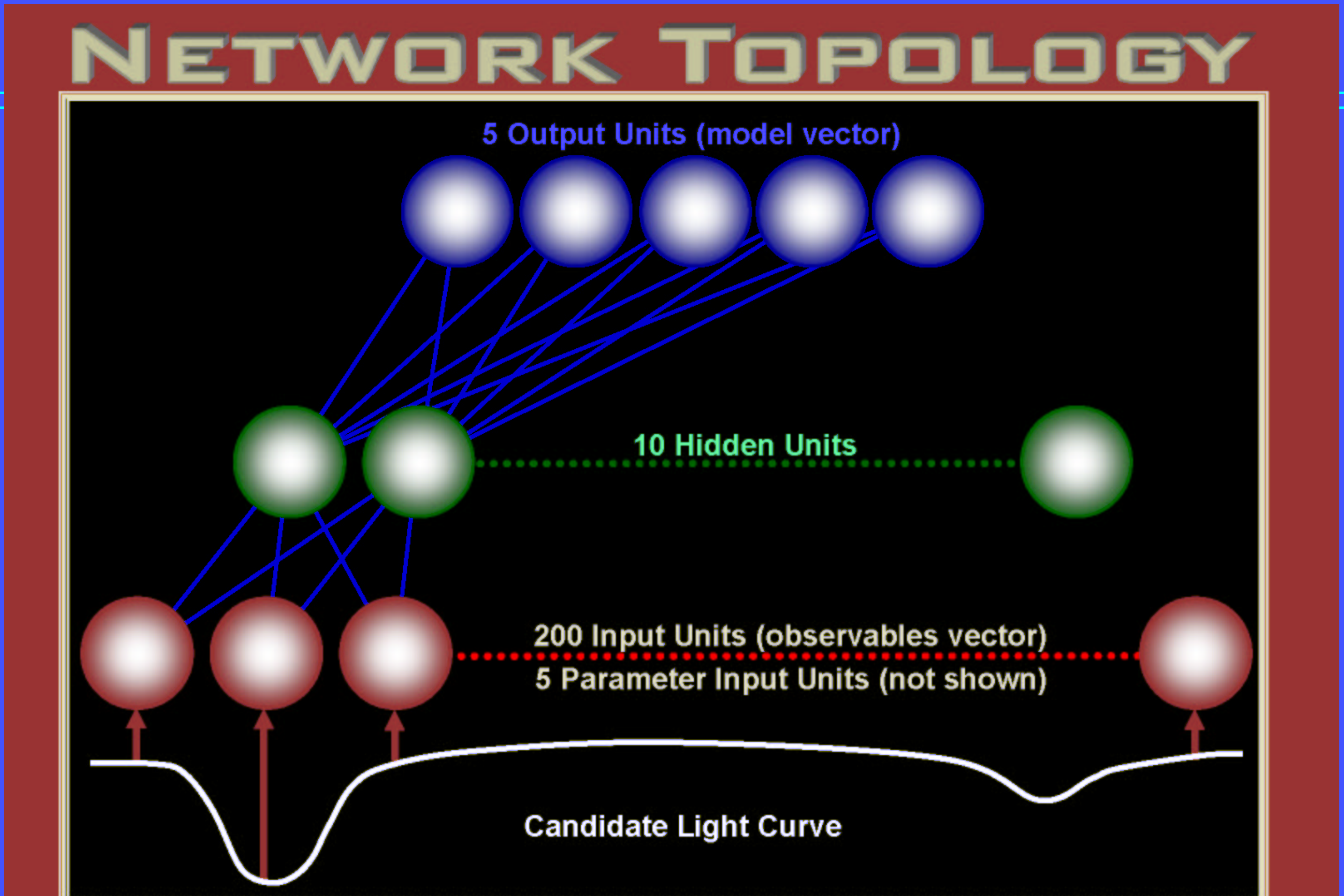
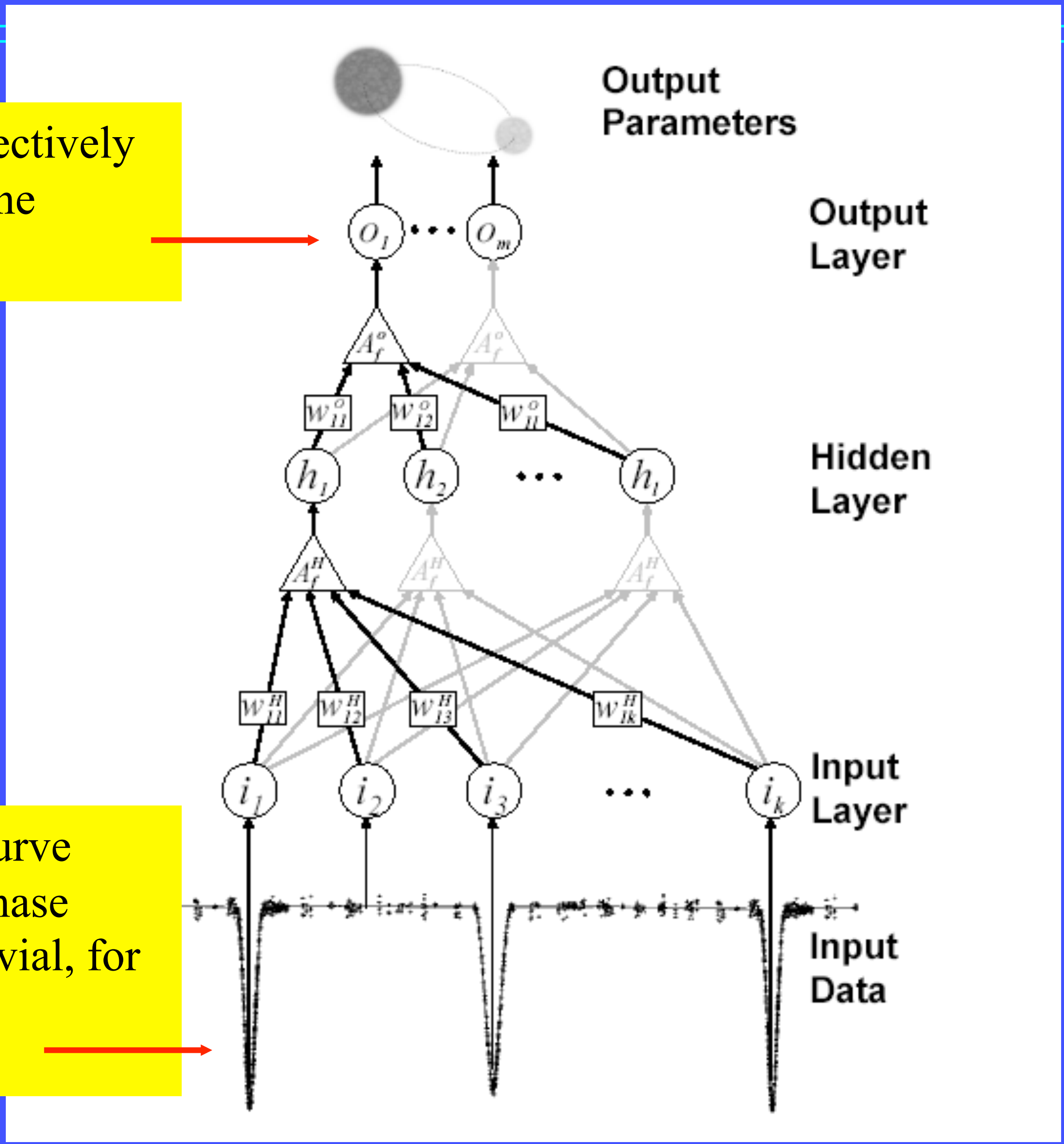


Figure 1: Neural Network indicating nodes, bottom, for input vector, hidden units (nodes) middle, and output parameter nodes (top). When properly trained, the Network will output model parameters corresponding to the input light curve data.

ANNs: Eclipsing Binaries Application

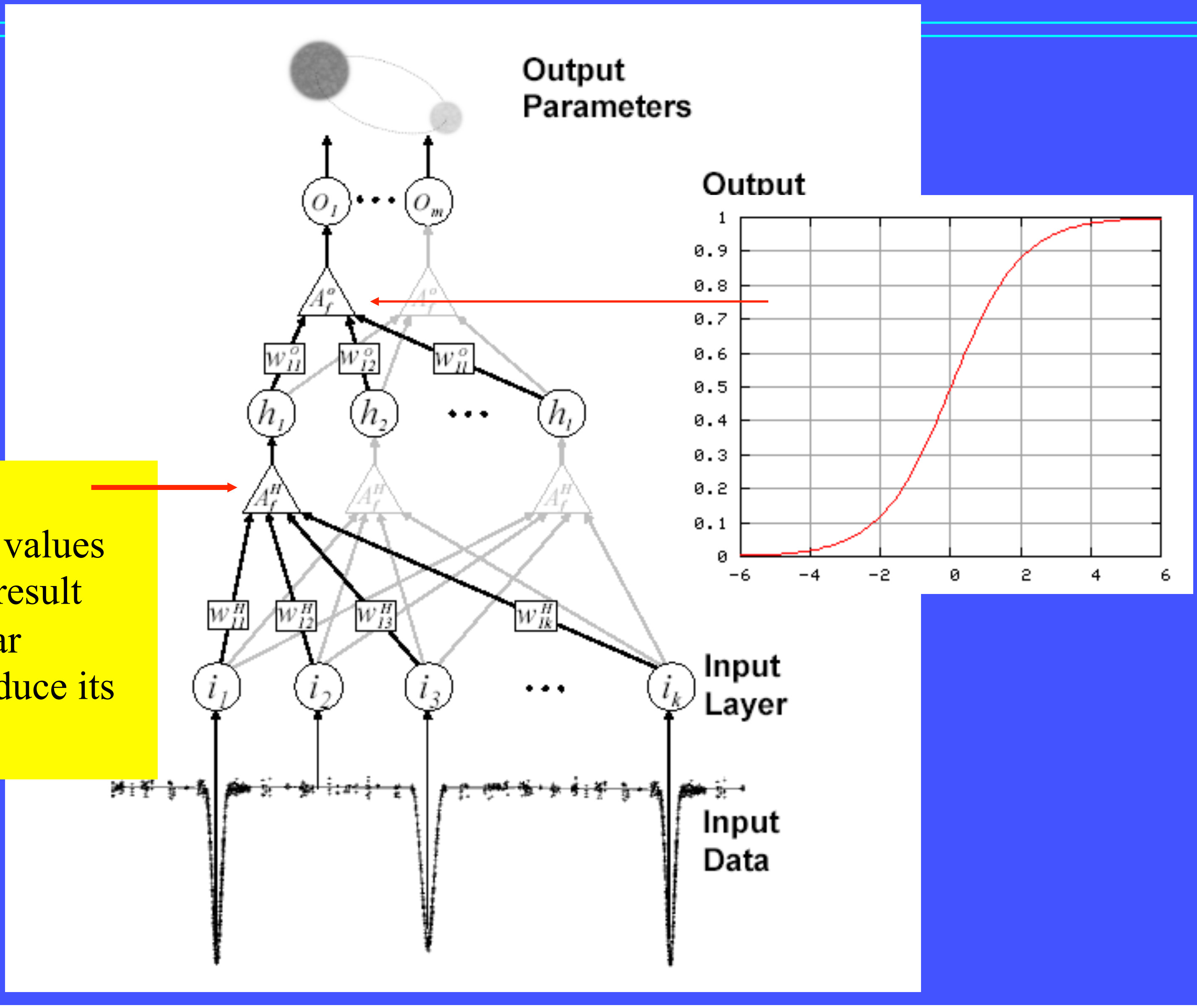
An Input-to-Output Mapping Problem

The 5 outputs respectively are parameters of the model.



Inputs are light curve values at equal phase intervals – not trivial, for real light curves!

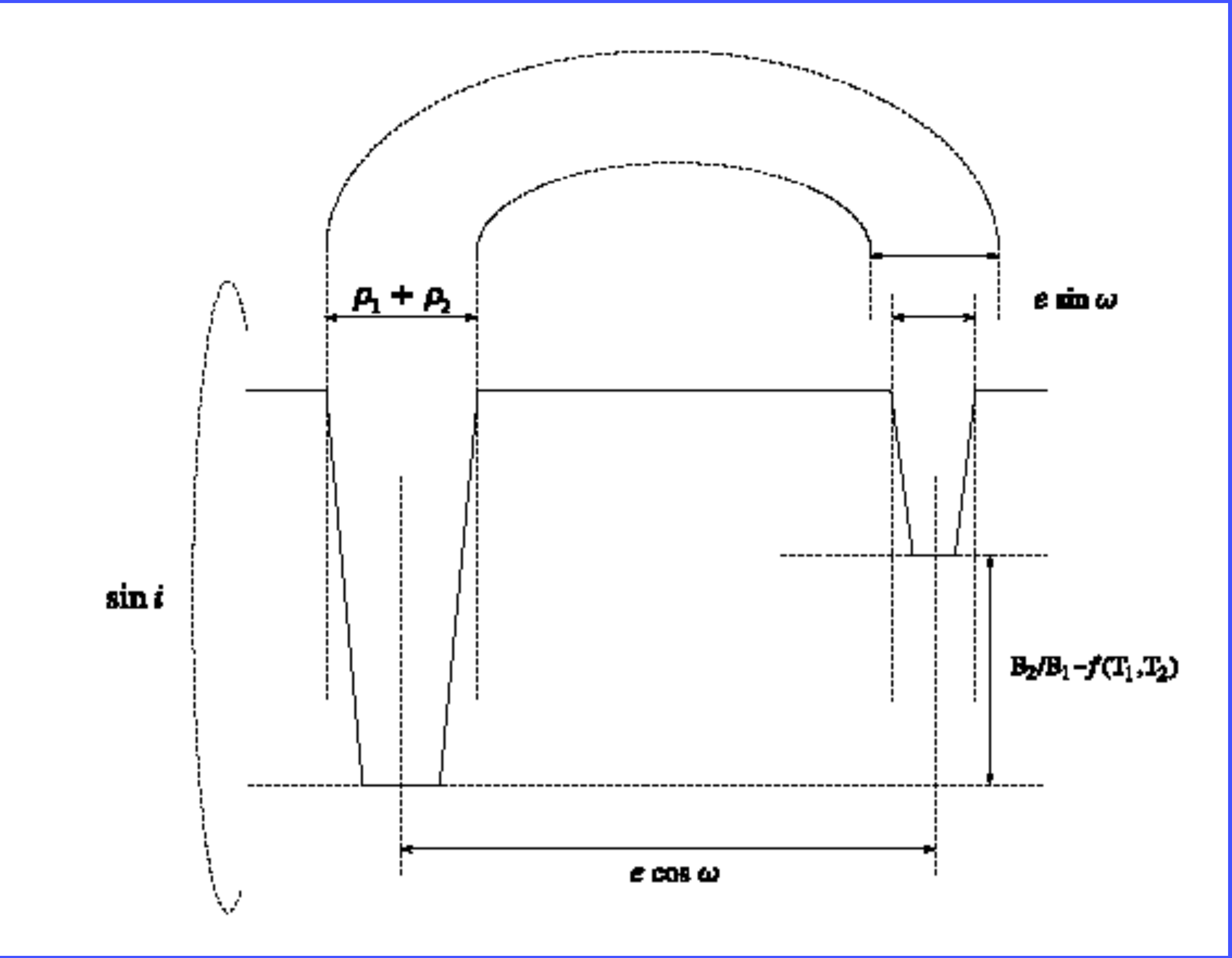
ANN Structure



A "node" sums the weighted input values and passes the result thru a non-linear function to produce its output.

LC Model Parameters for the ANN

Detached Systems



Artificial neural networks for LCs

Training the ANN via model LCs

- For training, exemplar (model) light curves with *known parameters* covering the complete range of parameter space are sequentially input to the ANN, and the output error “back-propagated” to adjust the network weights.
- Each parameter’s space must be covered by a fine enough grid.
- For 5 parameters, a training set of <u>33,000</u> light curves is required.
  - E.g: #p1 x #p2 x #p3 x #p4 x #p5

Artificial neural networks

- *Training* is compute-intensive:

  - Input **33,000** WD detached model light curves
  - Compute parameter errors at ANN output
  - Backpropagate the errors into the network to adjust weights
  - Repeat **500,000** times until errors are sufficiently minimized

- A *trained* network then outputs a solution estimation for an input light curve, and this is very fast.

Training progress vs. iteration

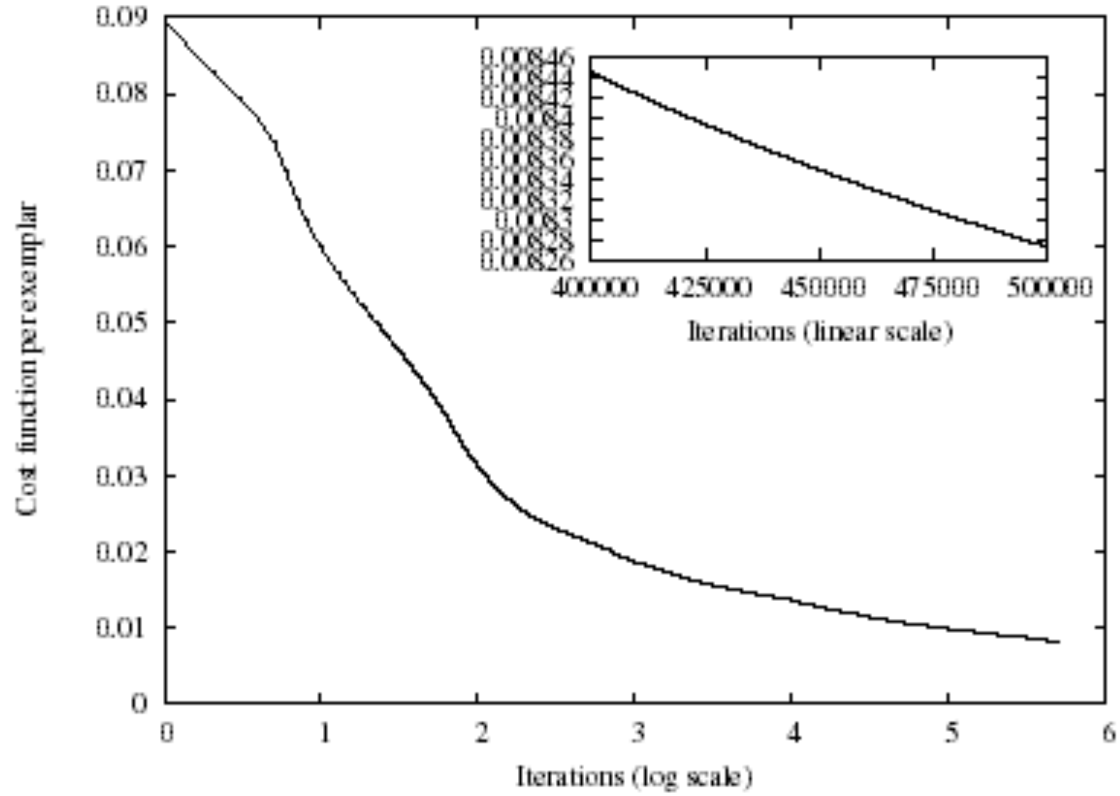


Fig. 11.— Learning curve. The number of iterations on the main plot is depicted in log scale. Cost function is normalized per exemplar, i.e. $\frac{1}{N}\sum_i\sum_{p=1}^{5}(o_p - c_p)^2$, where N is the number of exemplars, o_p and c_p are input and output values of parameters, respectively, index i goes over all exemplars and index p goes over all output parameters. The inset depicts the final 100,000 iterations on a linear scale.

Selected ANN topology

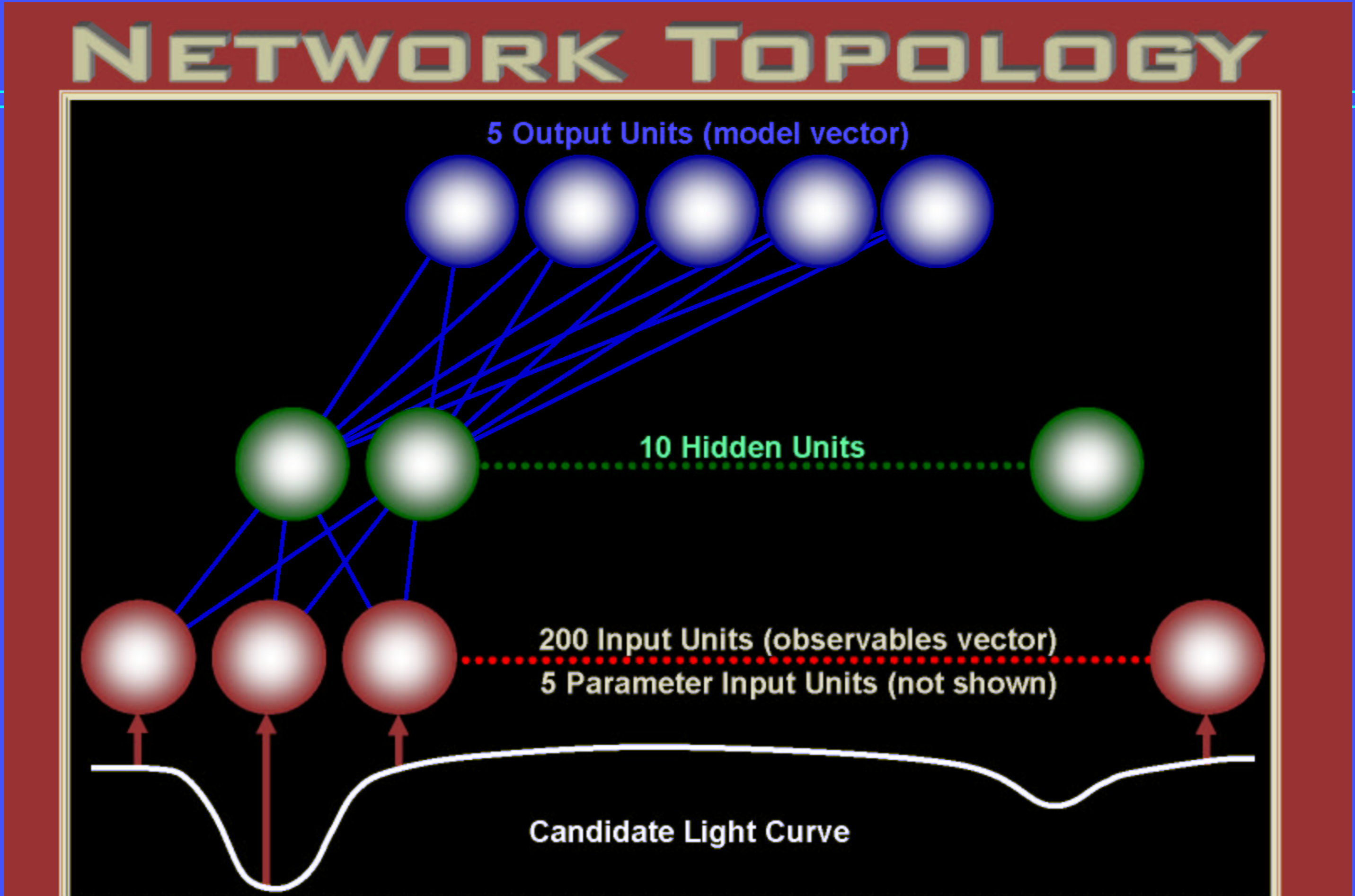


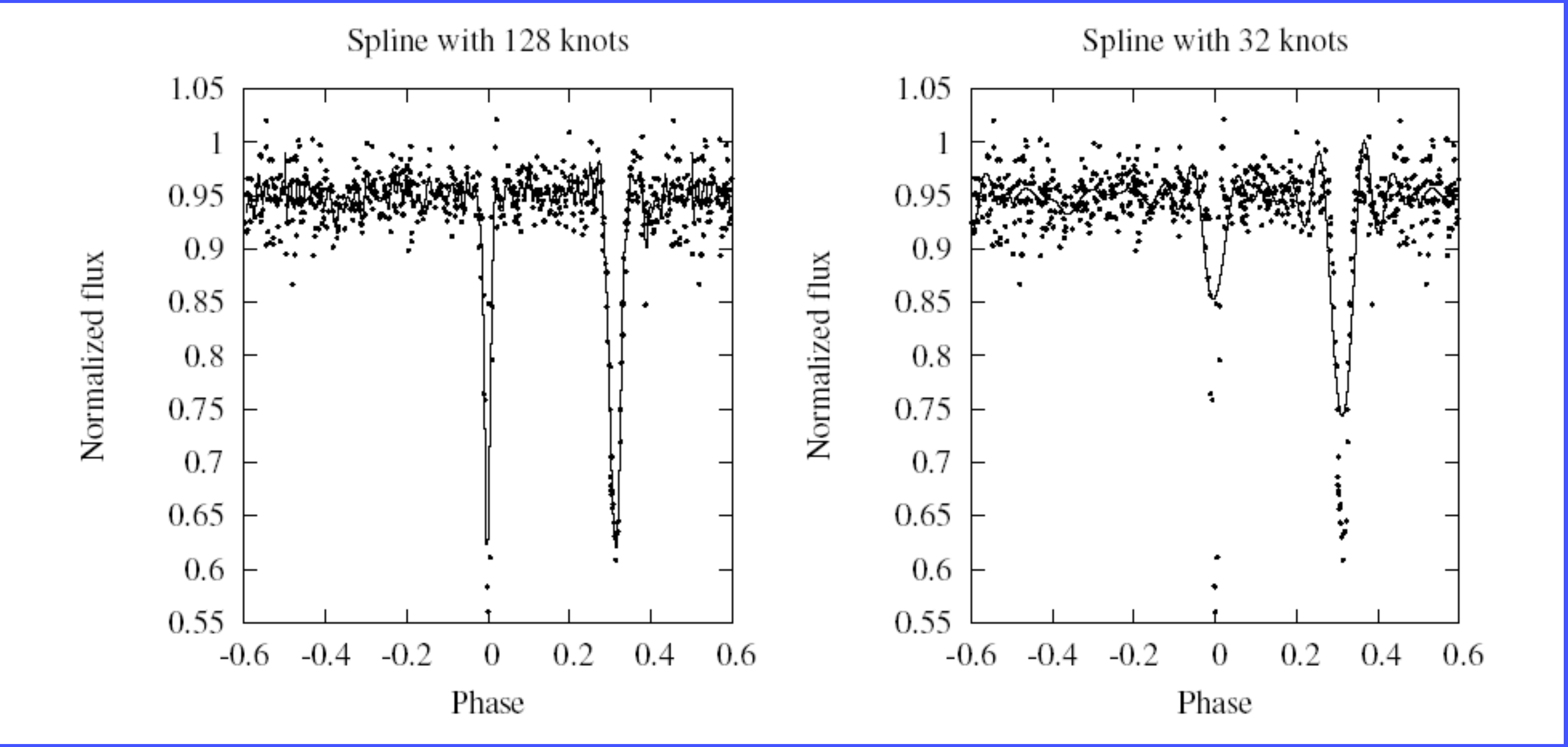
Figure 1: Neural Network indicating nodes, bottom, for input vector, hidden units (nodes) middle, and output parameter nodes (top). When properly trained, the Network will output model parameters corresponding to the input light curve data.

But there was just one problem!

- We want the INPUTs to our network to be light curve data at equal phase intervals.
- But - real light curves are NOT equiphase.
- No problem – we just interpolate the light curve data to equal phase intervals!
- Problem …!

Equiphase light curves

Splines don't work!



Equiphase light curves

POLYFIT

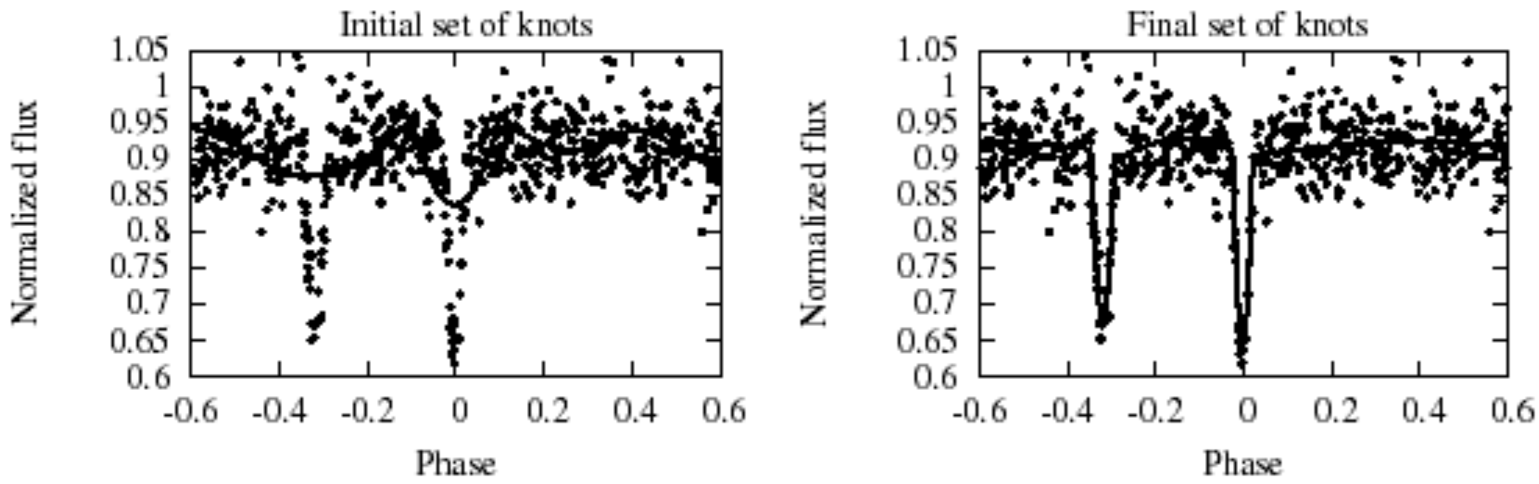
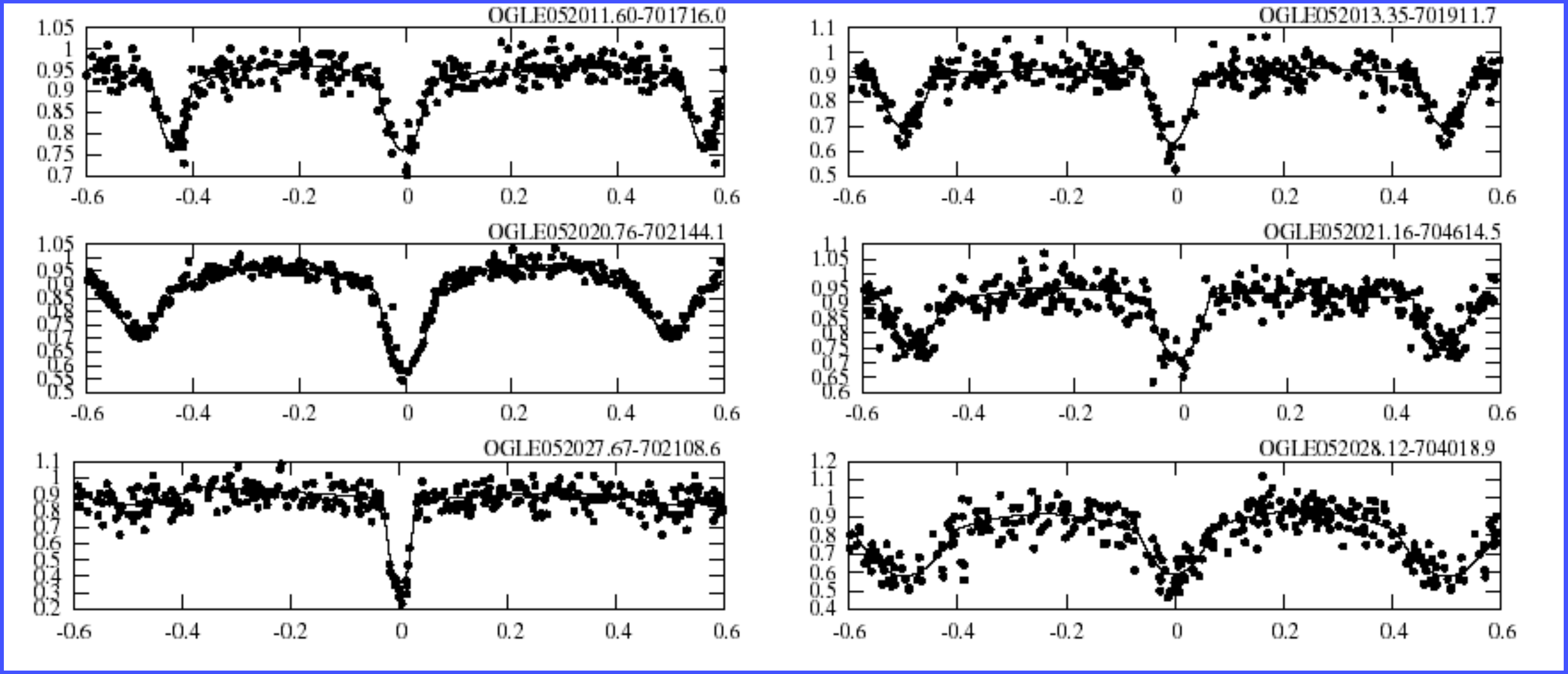


Fig. 18.— Performance of the quadratic 4-knot *polyfit* algorithm on an eccentric binary light curve. Left: the solution of a polynomial chain fit with the default set of initial knots, $\{-0.4, -0.1, 0.1, 0.4\}$. Right: the solution of the polynomial chain fit after 5000 iterations.

POLYFIT

example results



Performance on real data

Test 1 50 detached EBs from Dave Bradstreet's CALEB database

Test 2 2580 OGLE LMC EBs

Wyrzykowski, A&A, 2003

Star name:
ADBoo
NNCep
ARAur
AYCam
BPVul
BSDra
CDTau
CVBoo
CWCMa
DIHer
EKCep
EOVel
EWOri
FLLyr
FSMon
GGOri
ADAnd
KPAql
MNCas
MUCas
RTCrB
HSAur
AOMon
ZZBoo
ZHer
YYSgr
ASCam
WZOph
WYHya
WXCep
WCru
ASCam
IQPer
V909Cyg
V624Her
V526Sgr
V523Sgr
V541Cyg
V499Sco
V478Cyg
V477Cyg
V459Cas
V442Cyg
V396Cas
V364Lac
V364Cas
UZDra
SWCMa
STCen
NYCep

Fig. 20.— EBAI model light curves derived for the CALEB data.

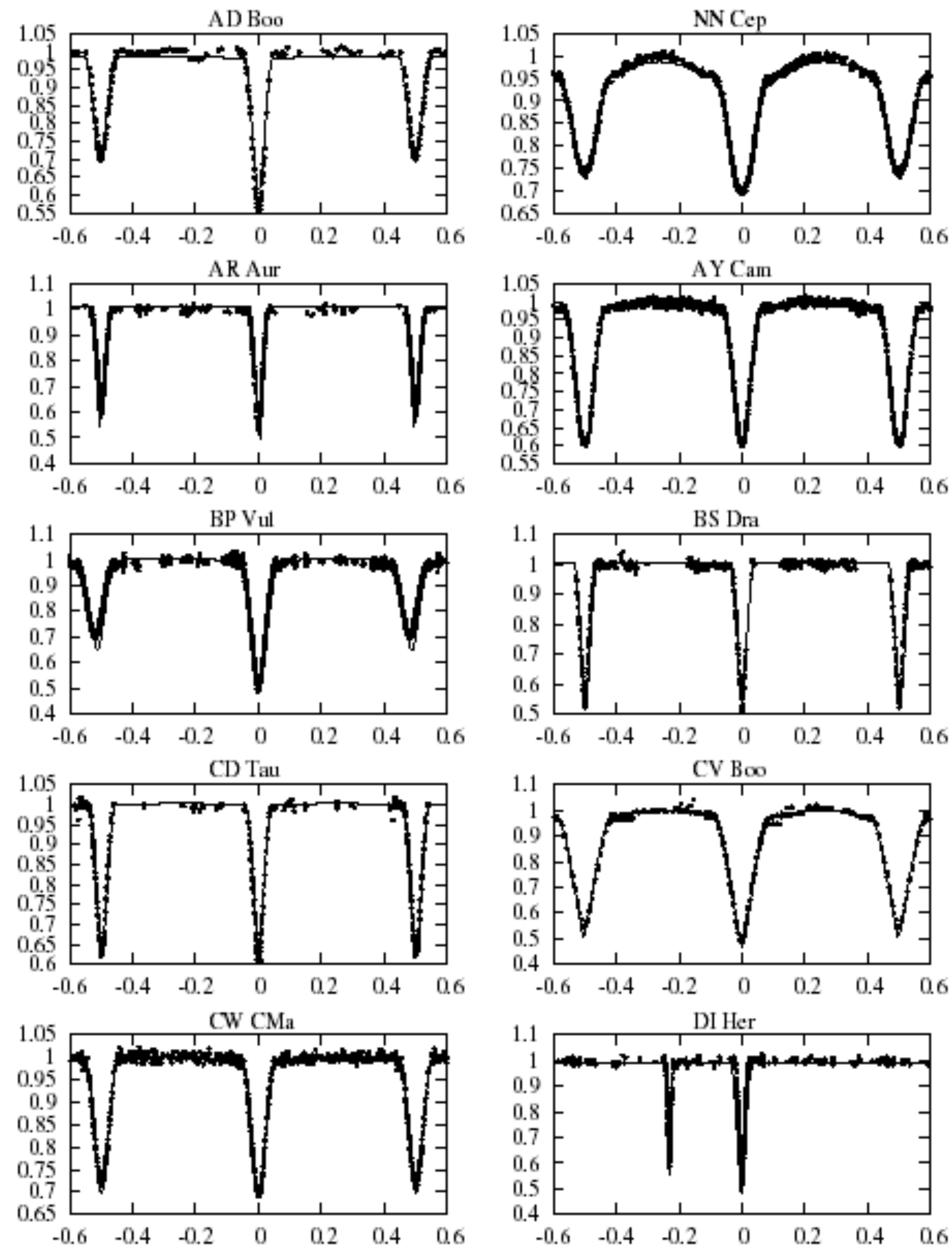


Fig. 21.— Continued: EBAI model light curves derived for the CALEB data.

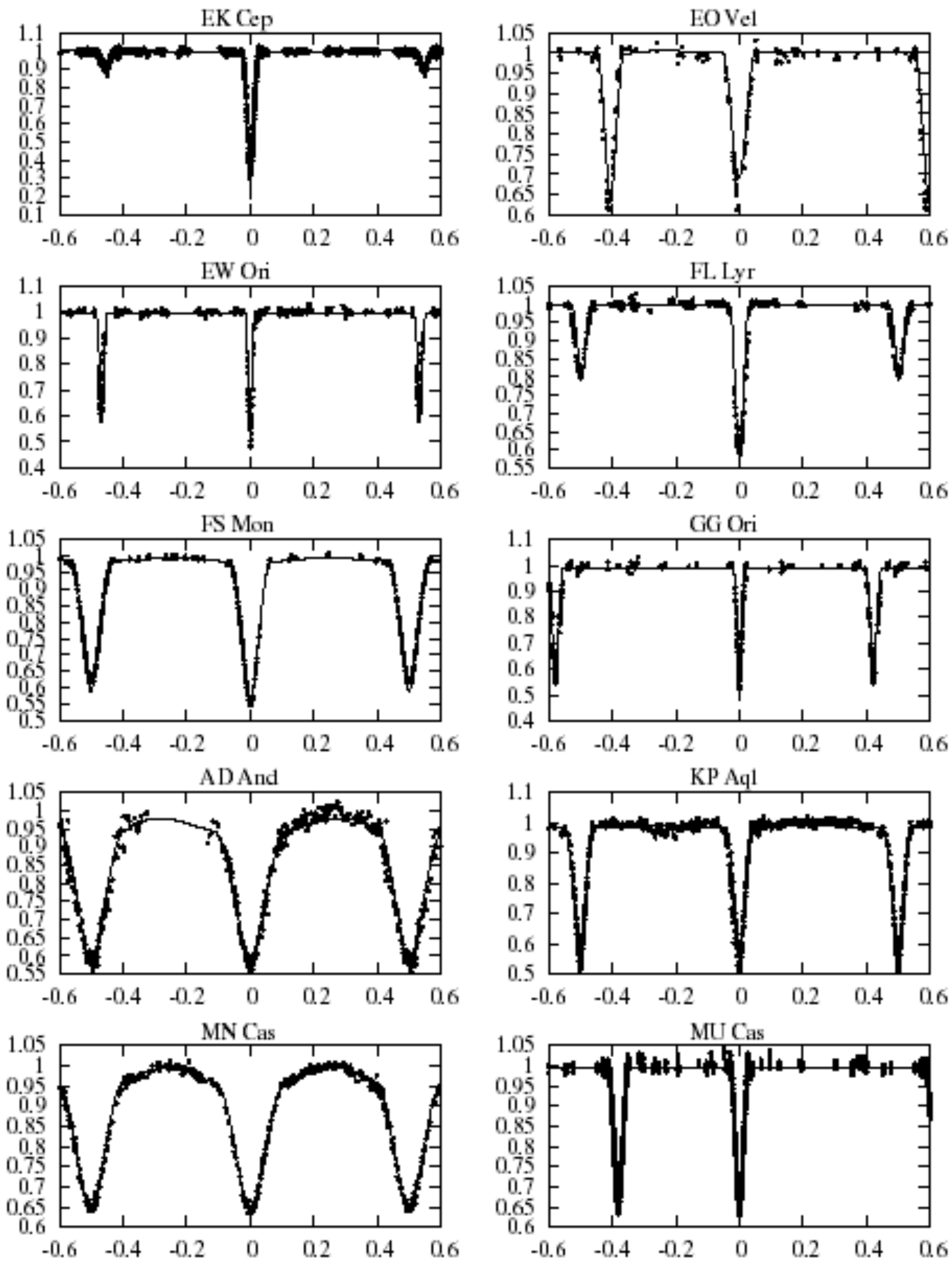


Fig. 22.— Continued: EBAI model light curves derived for the CALEB data.

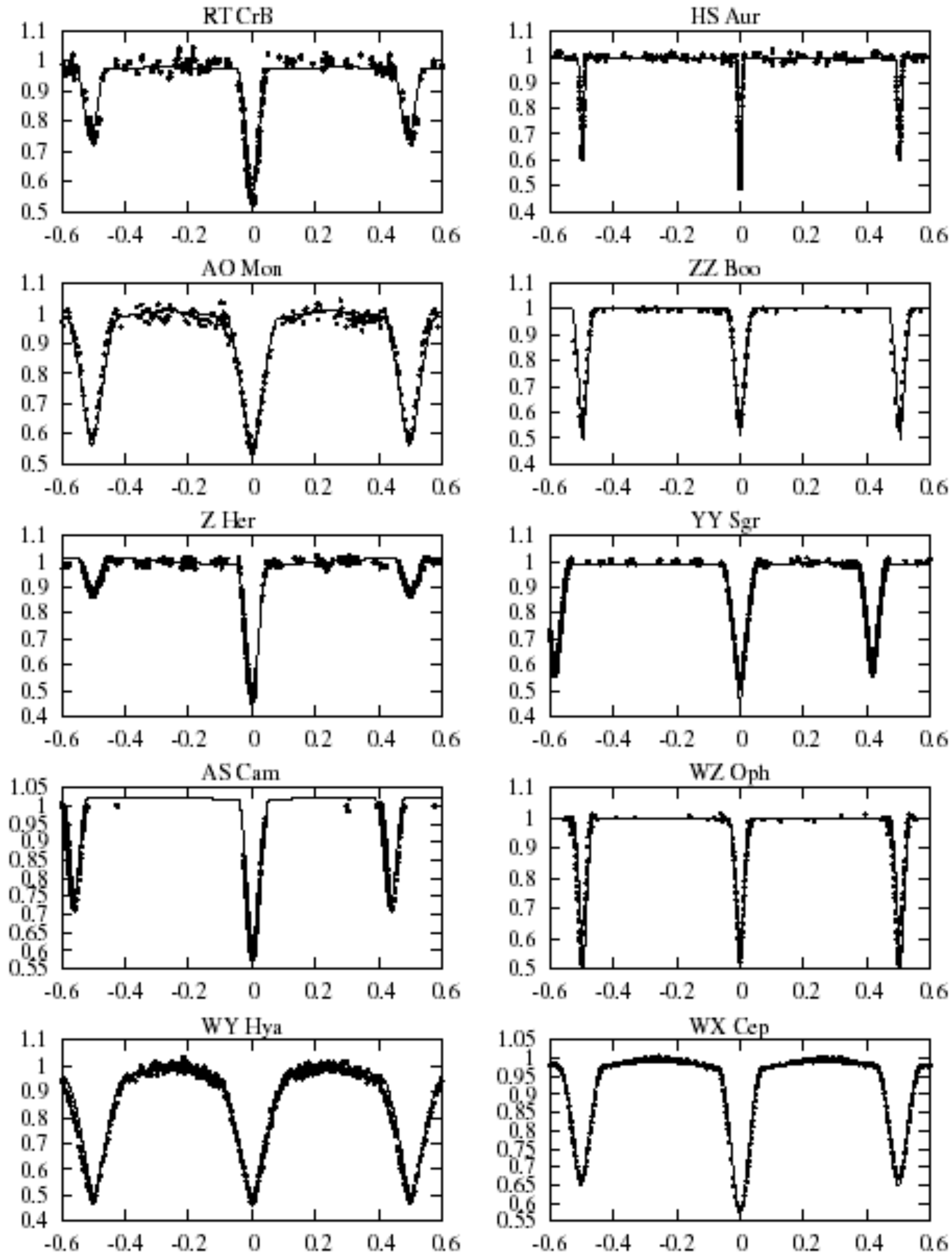


Fig. 23.— Continued: EBAI model light curves derived for the CALEB data.

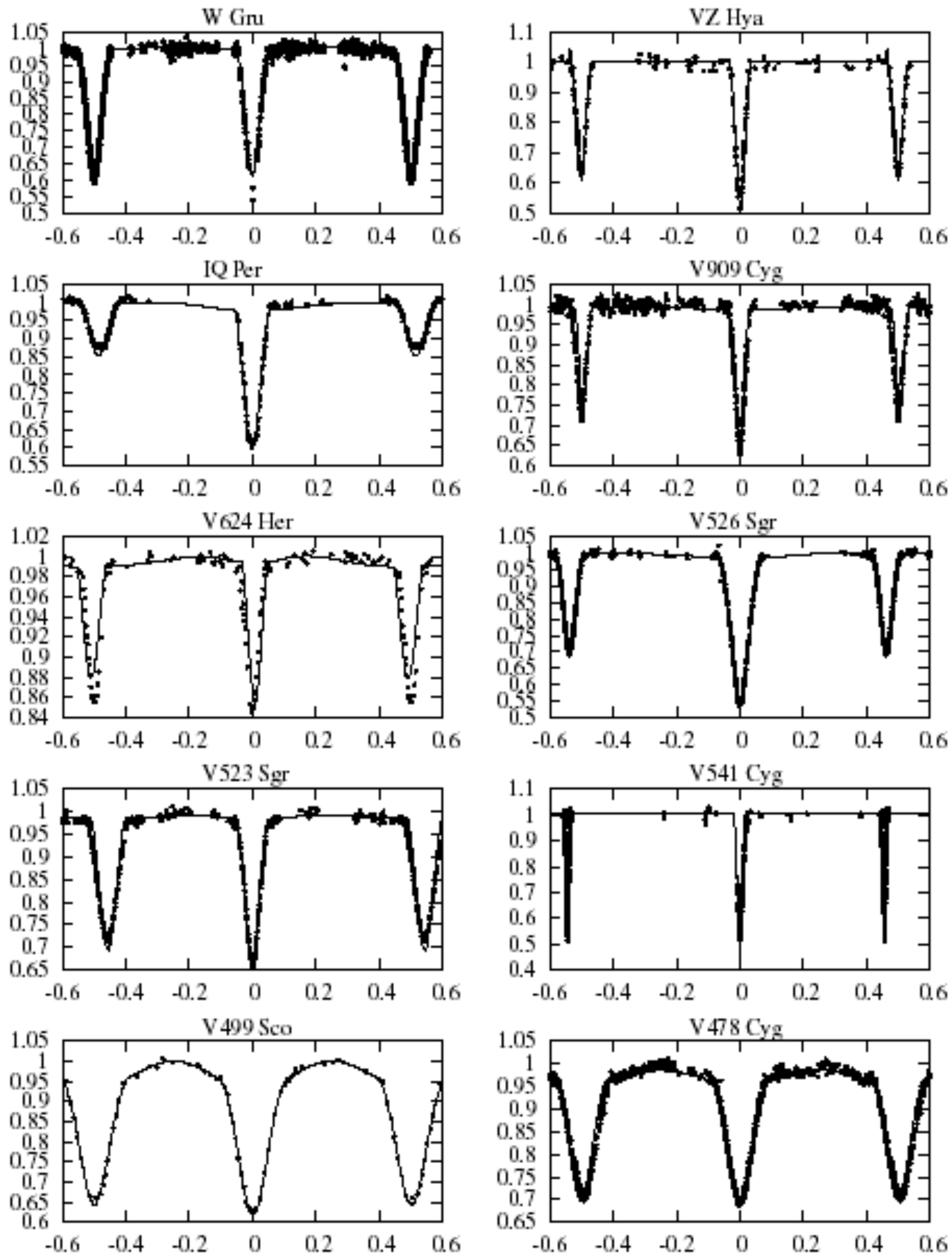
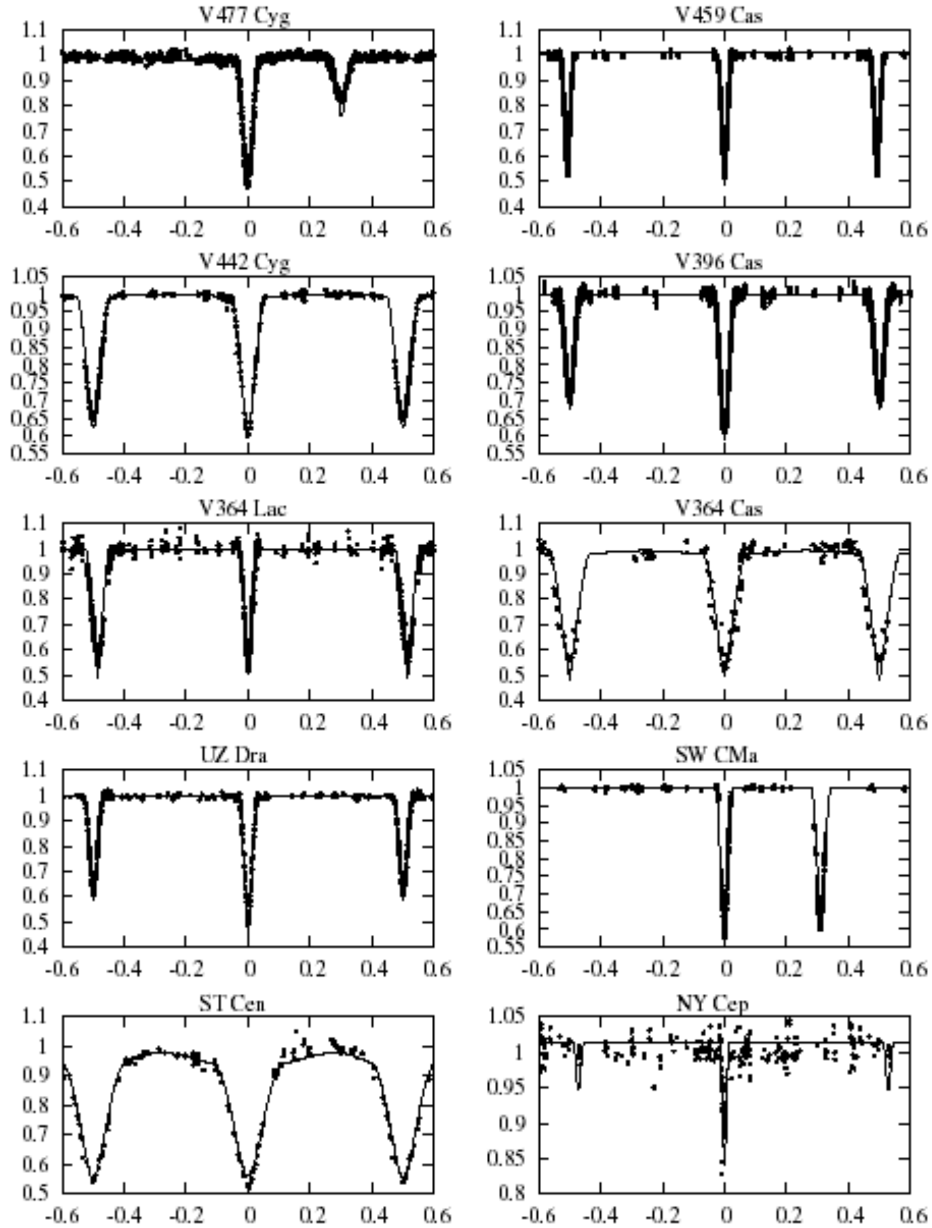


Fig. 24.— Continued: EBAI model light curves derived for the CALEB data.



CALEB Results Summary

- 50 Detached Systems
  - χ-square determined if WD differential correction (DC) is needed

- Results
  - 22/50 systems required no DC correction
  - 16/50 required 1 DC iteration
  - 5/5 required 2 DC iterations
  - 7/50 required 3 DC iterations.

2580 OGLE LMC stars (2 seconds)

Statistical Results

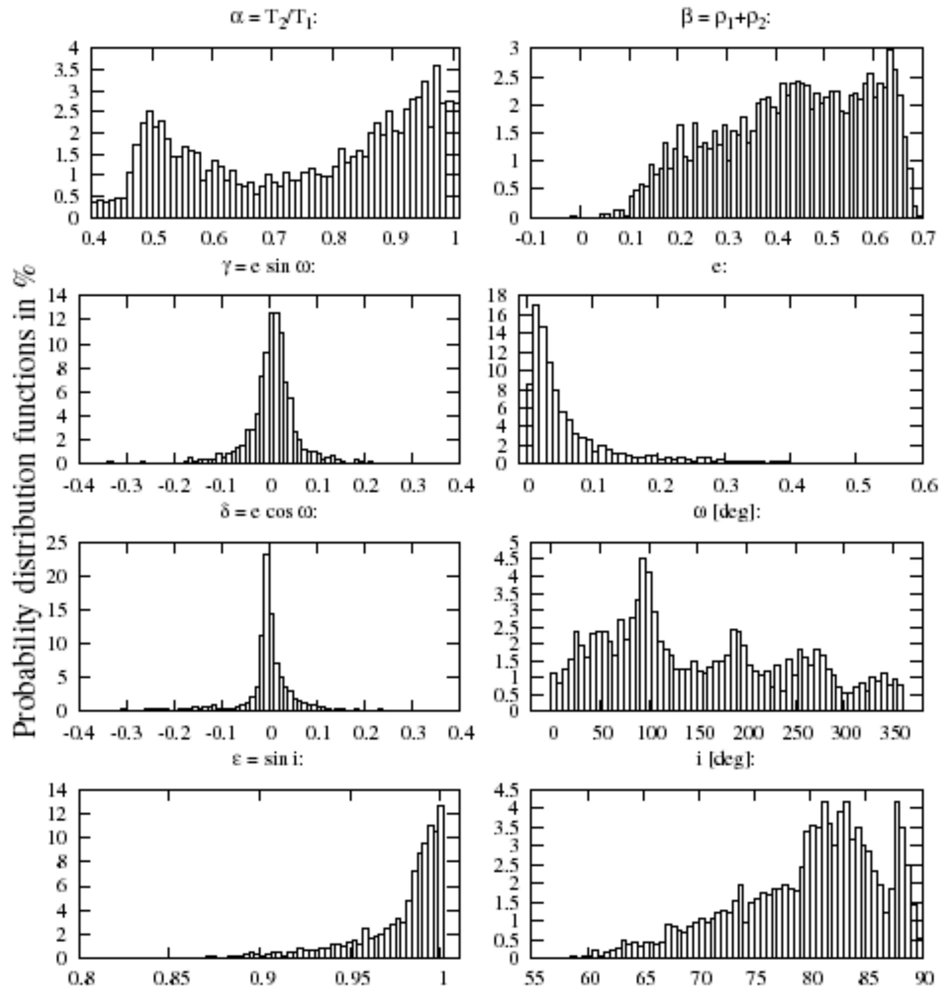
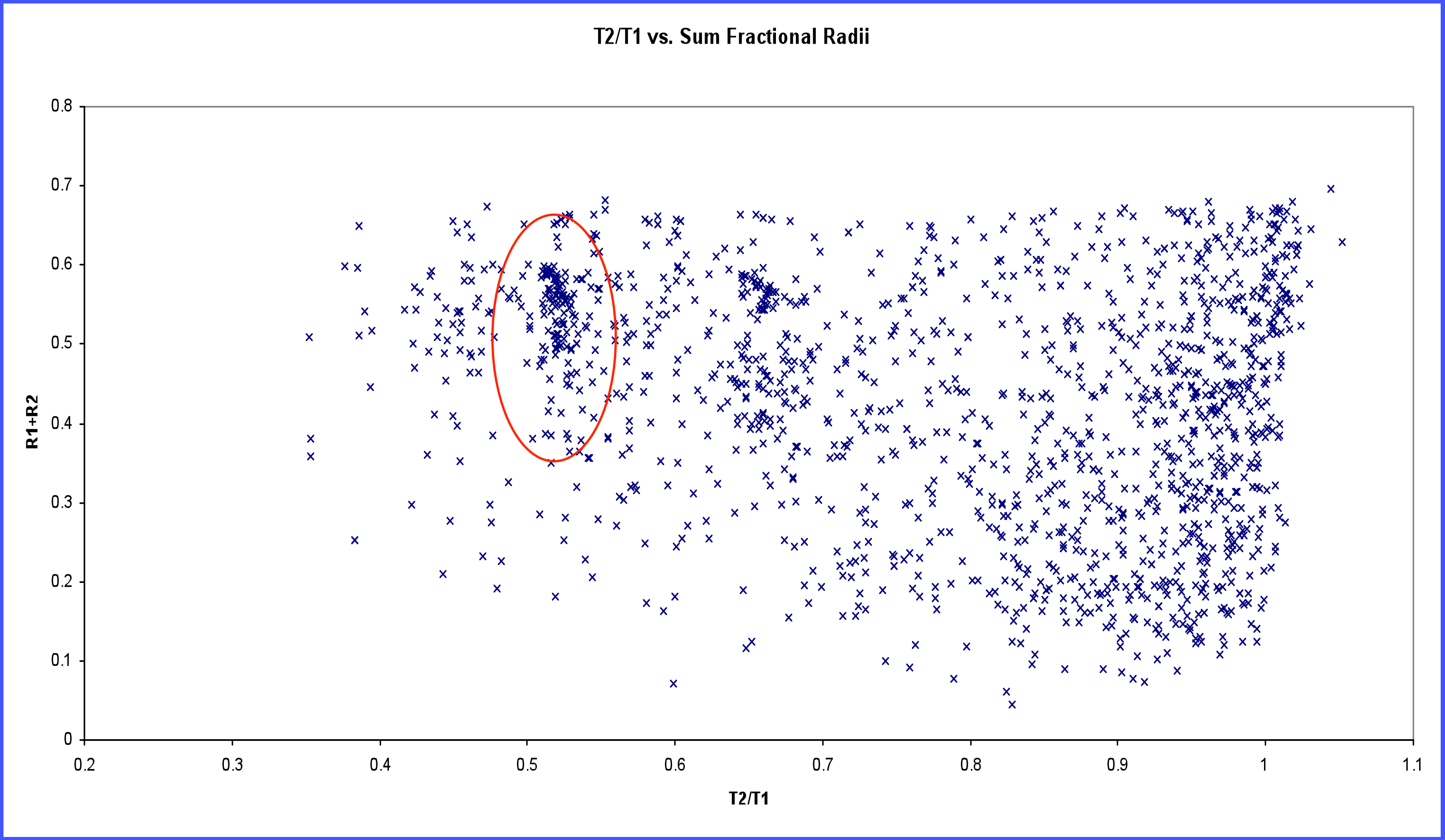


Fig. 16.— The distributions of parameters obtained by the ANN from the OGLE sample of 2580 LMC binaries classified as detached.

We are trying to find out!

Data for OGLE2 LMC"EA" stars



AI for light curve solution estimation
SUMMARY

- ANNs have been shown to be useful for light curve solution estimation.
- Success required additional “value-added” developments: e.g., POLYFIT, and a new Parallelization approach for training.
- The approach shows promise for other applications, including light curve classification.
- ApJ 687, 542-565, 1 Nov 2008.

Talk Topics - 3

1. A very short intro to Artificial Intelligence (AI)
2. The EBAI Project: Solving light curves from massive EB datasets with Artificial Neural Networks, and
3. **Exploring EBAI results using Cluster Analysis**

EB dataset: OGLE II LMC data.

Exploring Large (EB) Datasets
via
Cluster Analysis

Fun with Multidimensional Data

- With OGLE, CoRoT, Kepler and others to come, large high-dimensional datasets are here to stay!
- The flood of eclipsing binary (EB) data has already prompted imaginative new approaches to light curve solutions automation.
- Analysis (mining) of joint observations & solutions data [e.g., OGLE data + EBAI solutions] can offer further insights.
- Two technologies for exploring such high-dimensional datasets:
    - *Advanced Visualization*
    - *Clustering*

Advanced Visualization

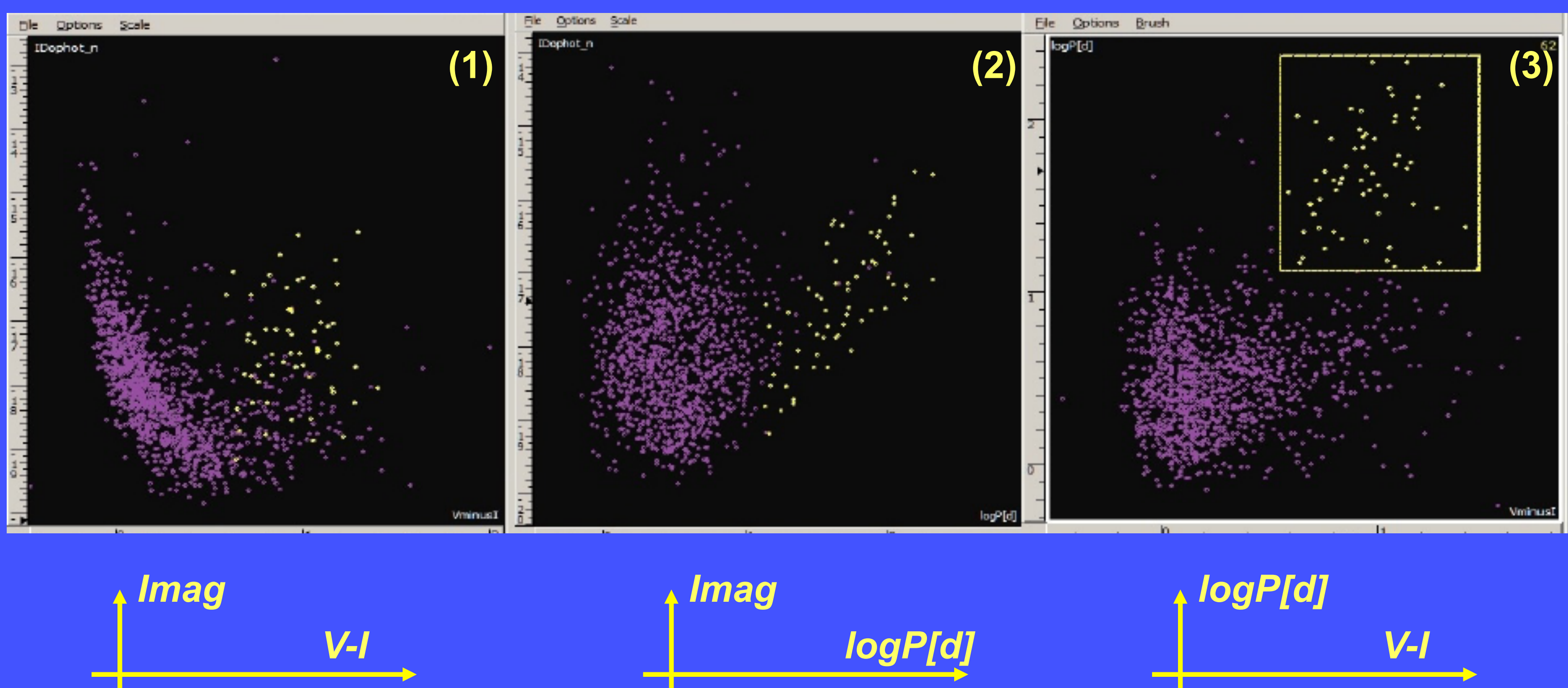
A graphical-interactive process

- **Graphical, interactive exploration of high-dimensional data:**
  - employs multiple, linked 2-D plots.
  - supports “brushing”-- highlighting a data point/region in one plot highlights these respective points immediately in all open complementary plots.
  - other tools: rotating 3-D graphs, tours, PCA and more.
- **Example uses:**
  - to discover/test new relationships between variables.
  - to identify and understand outlier properties.
  - to manually identify physically-related objects.
- **AV toolset:** **www_ggobi_org**

Advanced Visualization EB Example

Dataset: OGLE II LMC EA data with EBAI solutions

- **“Brushing” (links points across open plots)**
  - A rectangular region is brushed in figure (3) below.
    Points selected in (3) are immediately highlighted in plots (1) and (2).
  - Note a Period-Imag relation in (2).



Data Clustering

An automatic process

- **Finds clusters of points in *n*-dimensional space, which are candidate physically-related groups.**

- **Setting up clustering:**
  - choose number of clusters (or let clusterer decide);
  - select distance metric between points;
    e.g., Euclidean distance, Malanobis distance, city-block distance
  - choose the parameters of the dataset to be clustered; i.e., the number of dimensions, n.

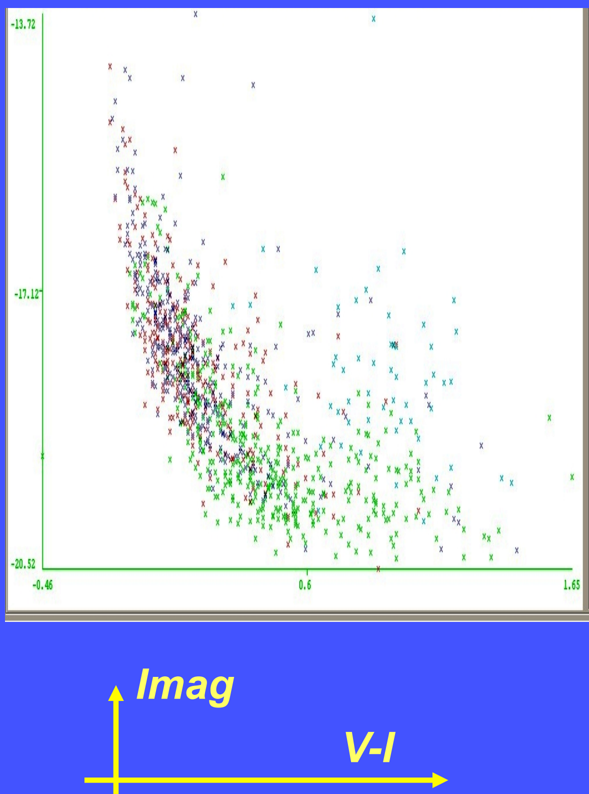
- **Data Clustering toolset: www_cs_waikato_ac_nz/ml/weka**

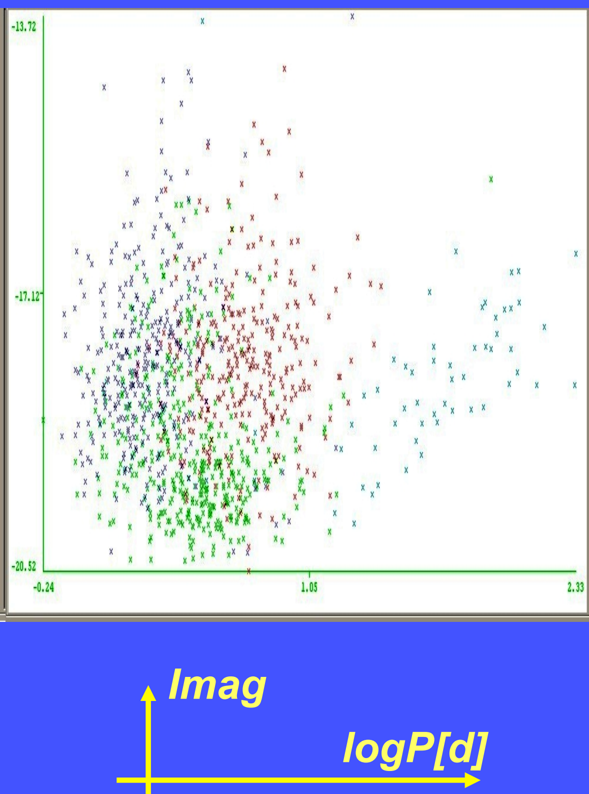
Data Clustering EB Example

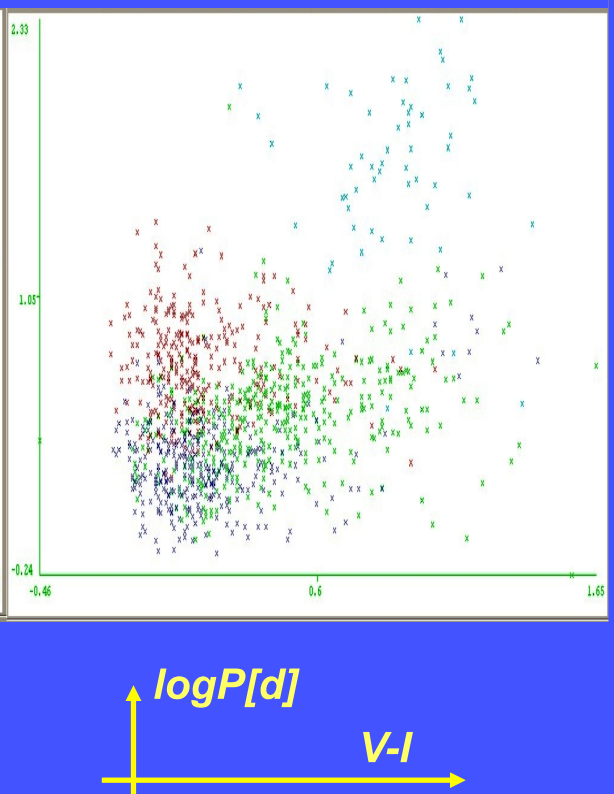
Dataset: OGLE II LMC EA data with EBAI solutions

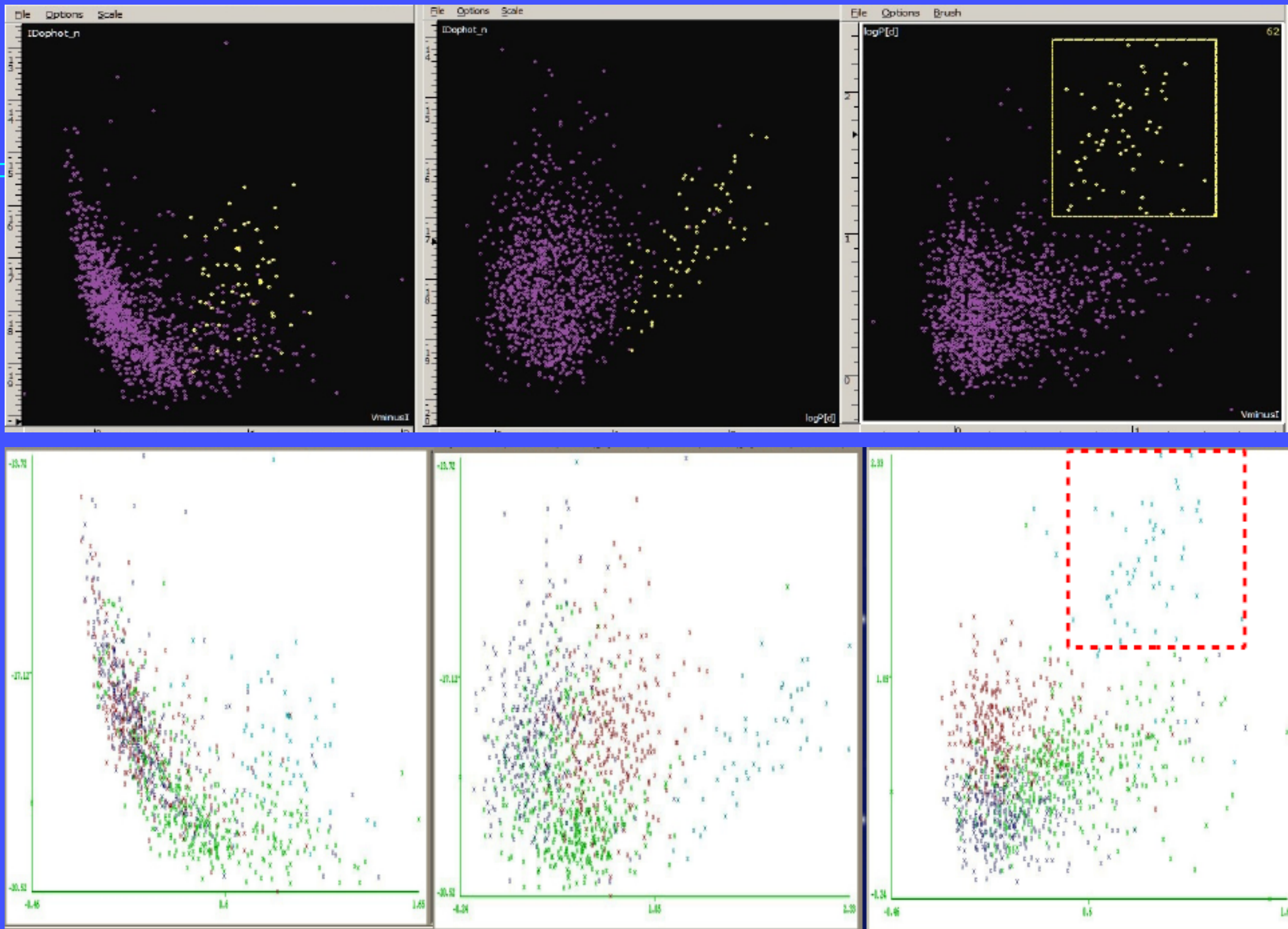
Number of clusters	4
Distance Metric	Euclidean

EB parameters clustered (# dimensions)	● Imag ● V-I ● logP[d] ● primary eclipse depth ● secondary/primary depth ● sum of fractional radii ● temperature ratio T2/T1









- One of four clusters found (red outline), compare to panel above.

Data Clustering EB Example

Dataset: OGLE II LMC EA data with EBAI solutions