Rise of the Machines

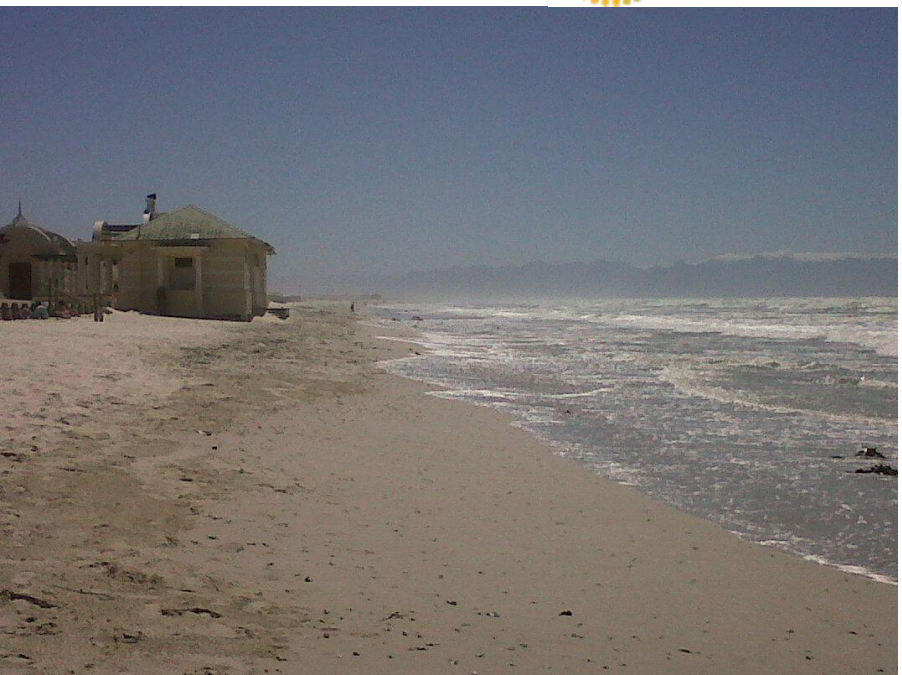
Bruce Bassett

AIMS – SAAO – UCT

Meudon

IAP – 4 October 2013

AIMS African Institute for Mathematical Sciences



www.aims.ac.za

Why am I interested in this topic?

1

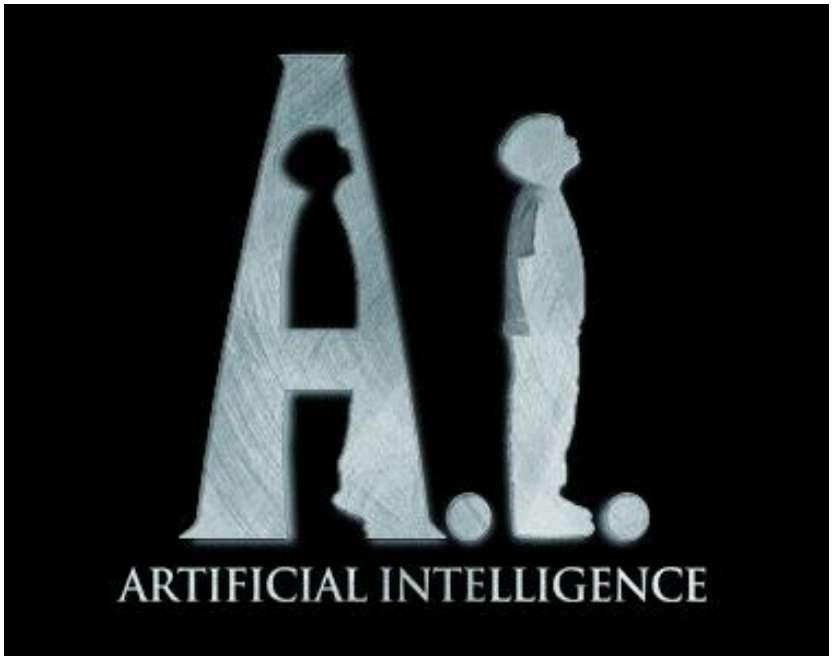
Drinking from the Fire Hydrant

- SKA, LSST, CTA, EUCLID, ALMA, LOFAR, BigBOSS, DESI, etc…
- Will p
  Huma)0
- This is pare our st ne (espe


2
Google
self-driving car
6UEJ089

3

It raises fascinating fundamental issues in articifical intelligence, human cognition, the nature of science etc…



A typical Astronomy example…

Given repeat imaging of the sky:

(a) identify plausible supernova candidates and

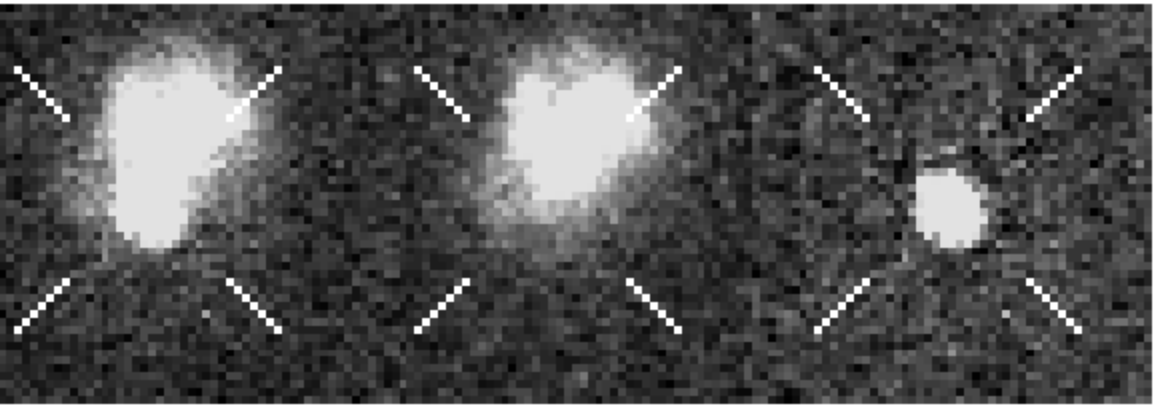
(b) distinguish Type Ia's from non-Ia's using multi-band light-curves…

Edit View Go Bookmarks Tools Help

http://sdssdp47.fnal.gov/sdsssn_data/handscan/manual_scan.php

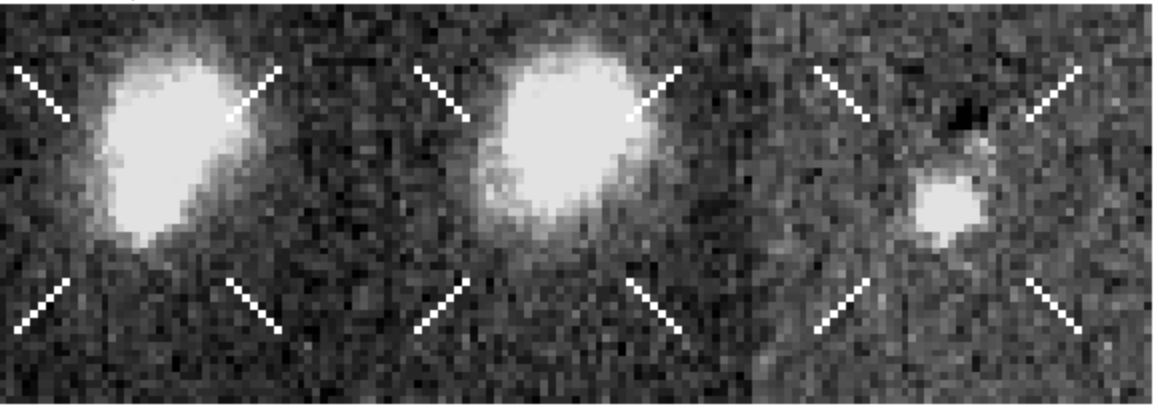
Go

stro-ph authors search for 'heyden' all ye... | Manual Scan | 0509039.pdf (application/pdf Object) | 0509041.pdf (application/pdf Object)



g (srch,

olt, subtr)



r (srch,

Scanner	Bassett	Updating	U
Obj Id	418059		
srun	5566	sfield	3
trun	2662	tfield	3
rr	10	cc	4
ra	14.121100	decl	0
gmag	18.744	g_delta	0
rmag	19.019	r_delta	0
imag	19.215	i_delta	0

[2] Moving
[3] Sat. Star
[4] Dipole
[5] Variable
[6] Transient
[9] Cosmic Ray
[103] SN GOLD
[102] SN SILVER
[101] SN BRONZE
[100] SN OTHER

story for Object Id 418059 By Position

und 0 previous objects.

tart

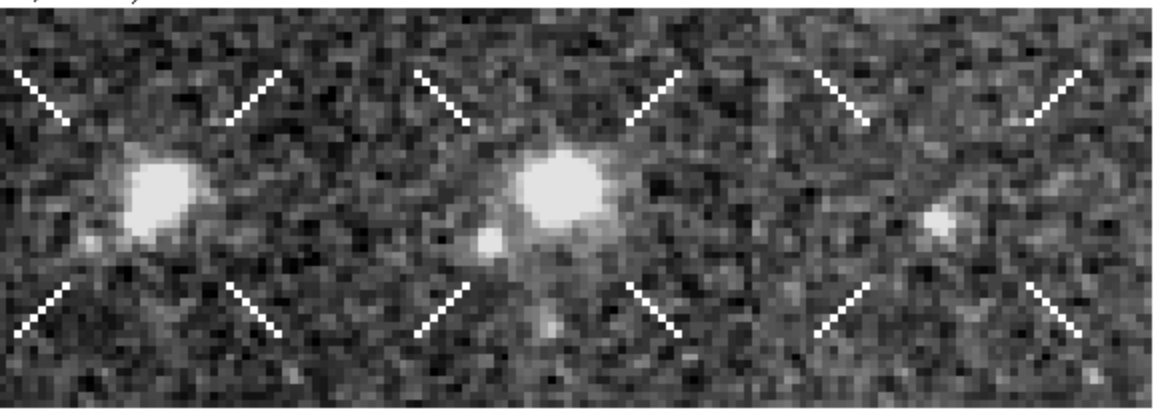
2:31

Edit View Go Bookmarks Tools Help

http://sdssdp47.fnal.gov/sdsssn_data/handscan/manual_scan.php

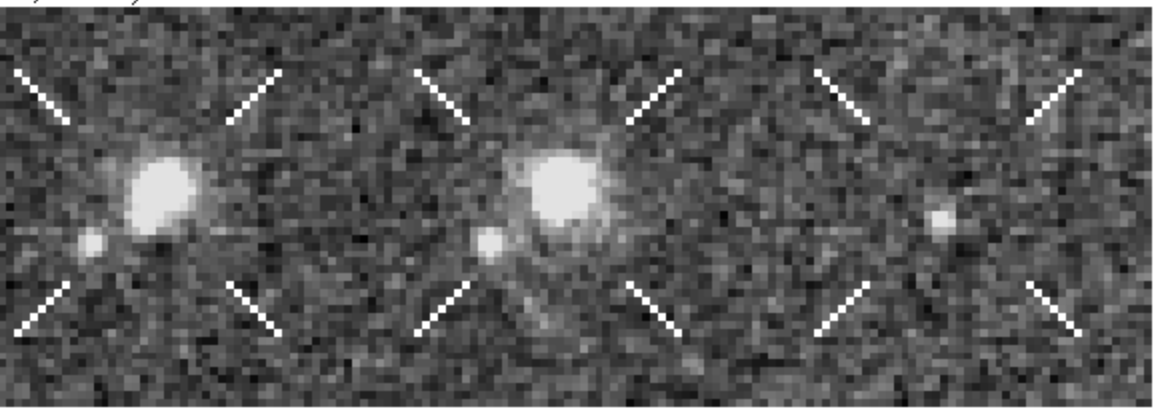
Go

Manual Scan | Images History



r (srch,

olt, subtr)



i (srch,

Scanner	Bassett	Updating	
Obj Id	492576		
srun	5603	sfield	
trun	2659	tfield	
rr	10	cc	
ra	-41.070187	decl	
gmag	21.030	g_delta	
rmag	20.986	r_delta	
imag	20.844	i_delta	

- [3] Sat. Star
- [4] Dipole
- [5] Variable
- [6] Transient
- [9] Cosmic Ray
- [103] SN GOLD
- [102] SN SILVER
- [101] SN BRONZE
- [100] SN OTHER

story for Object Id 492576 By Position

und 0 previous objects.

tart

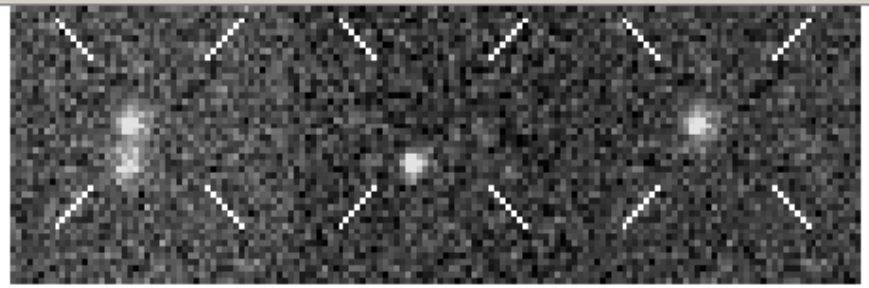
4:05

Manual Scan - Mozilla Firefox

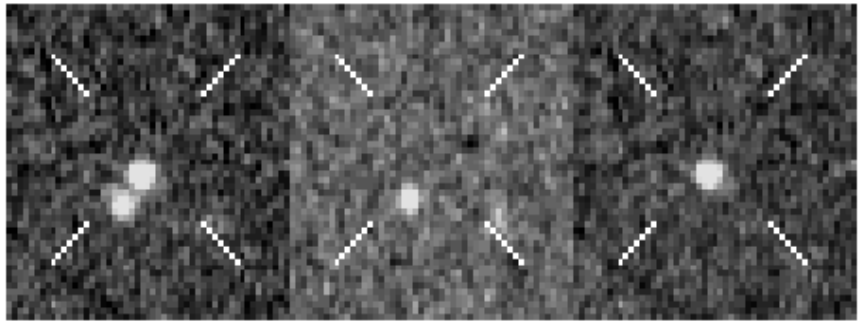
File Edit View Go Bookmarks Tools Help

http://sdssdp47.fnal.gov/sdsssn_data/handscan/manual_scan.php

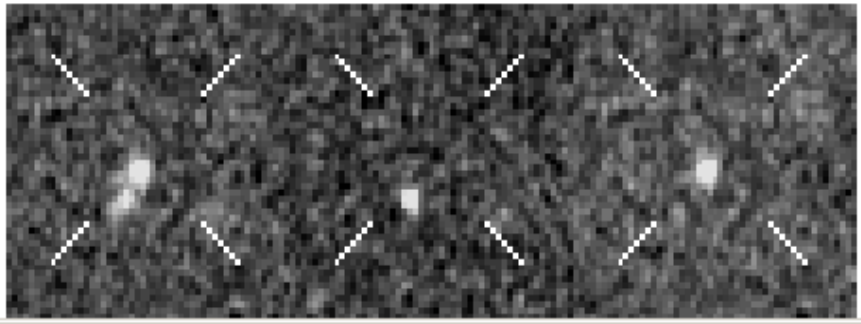
arXiv.org Search | Loading... | Untitled Document | Manual Scan



g (srch, tmplt, subtr)



r (srch, tmplt, subtr)



i (srch, tmplt, subtr)

Obj Id	394494		
srun	4203	sfield	135
trun	2583	tfield	111
rr	57	cc	2
ra	-40.319427	decl	-0.7
gmag		g_delta	0.00
rmag	0.000	r_delta	0.24
imag	0.000	i_delta	0.24
Flags	Moving		

- (•) [0] None
- ( ) [1] Artefact
- ( ) [3] Sat. Star
- ( ) [4] Dipole
- ( ) [5] Variable
- ( ) [6] Transient
- ( ) [9] Cosmic Ray
- ( ) [103] SN GOLD
- ( ) [102] SN SILVER
- ( ) [101] SN BRONZE

History for Object Id 394494 By Position

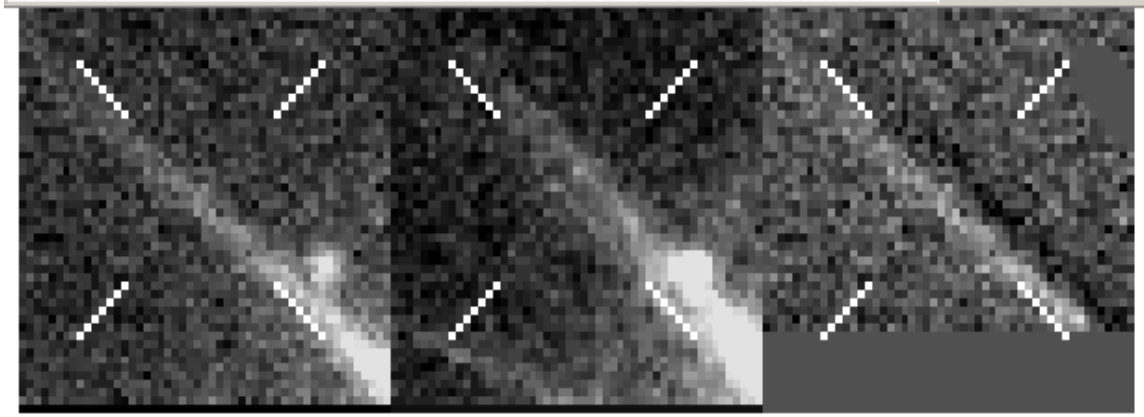
Found 1 previous objects.

Manual Scan - Mozilla Firefox

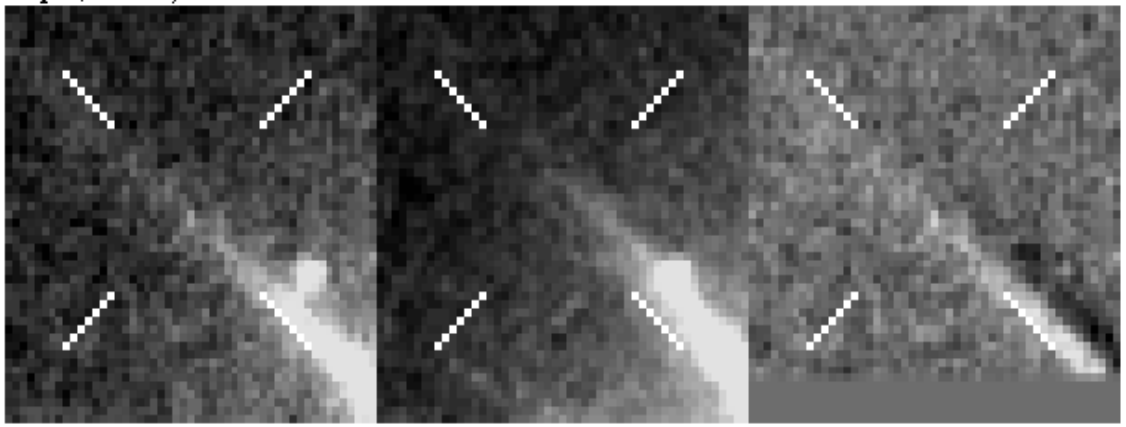
File Edit View Go Bookmarks Tools Help

http://sdssdp47.fnal.gov/sdsssn_data/handscan/manual_scan.php

[hep-th/0505042] Anti-D-brane as Dark Ma... | 0508572.pdf (application/pdf Object) | Manual Scan



r (srch, tmplt, subtr)



i (srch, tmplt, subtr)

Obj Id	398422		
srun	4874	sfield	95
trun	829001	tfield	14
rr	99	cc	1
ra	-48.854362	decl	-0.9
gmag		g_delta	
rmag	0.000	r_delta	1.10
imag	0.000	i_delta	1.10
Flags			

- (•) [0] None
- ( ) [1] Artefact
- ( ) [2] Moving
- ( ) [3] Sat. Star
- ( ) [4] Dipole
- ( ) [5] Variable
- ( ) [6] Transient
- ( ) [9] Cosmic Ray
- ( ) [103] SN GOLD
- ( ) [102] SN SILVER

History for Object Id 398422 By Position

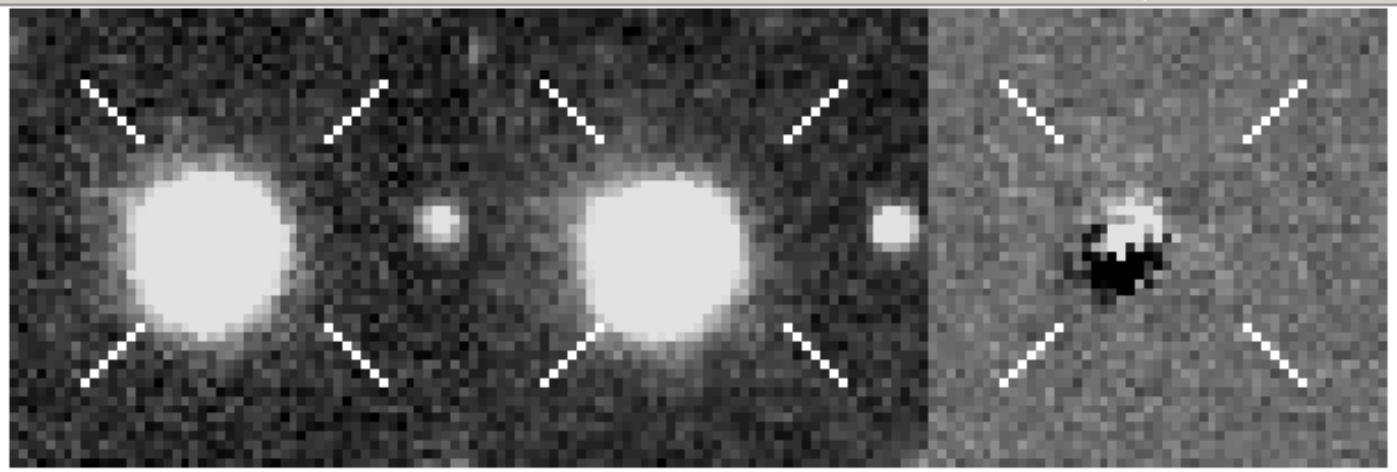
Found 0 previous objects.

Manual Scan - Mozilla Firefox

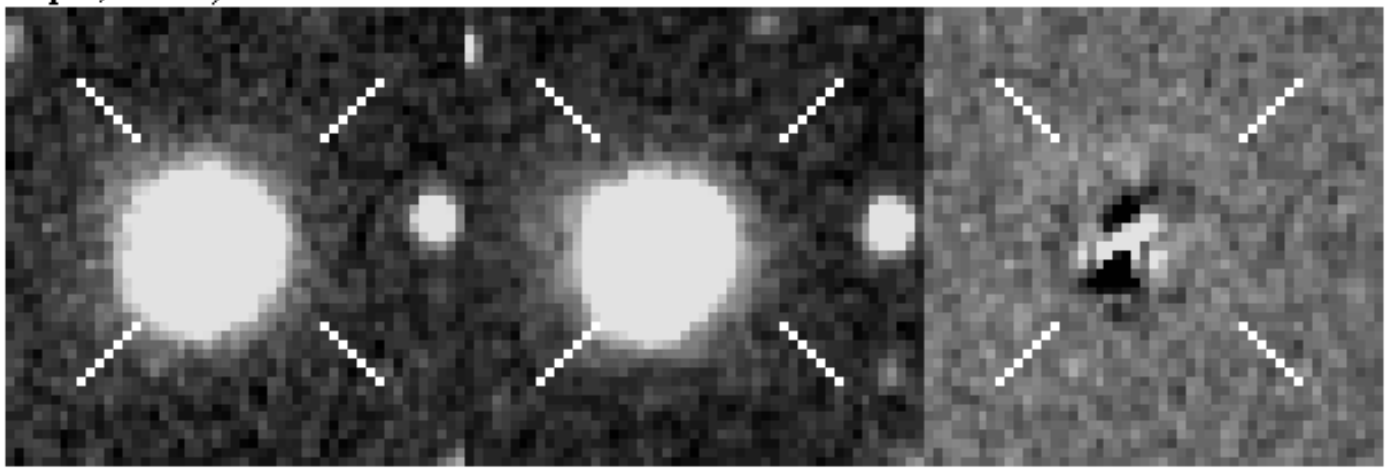
File Edit View Go Bookmarks Tools Help

http://sdssdp47.fnal.gov/sdsssn_data/handscan/manual_scan.php

[hep-th/0505042] Anti-D-brane as Dark Ma... | 0508572.pdf (application/pdf Object) | Manual Scan



r (srch, tmplt, subtr)



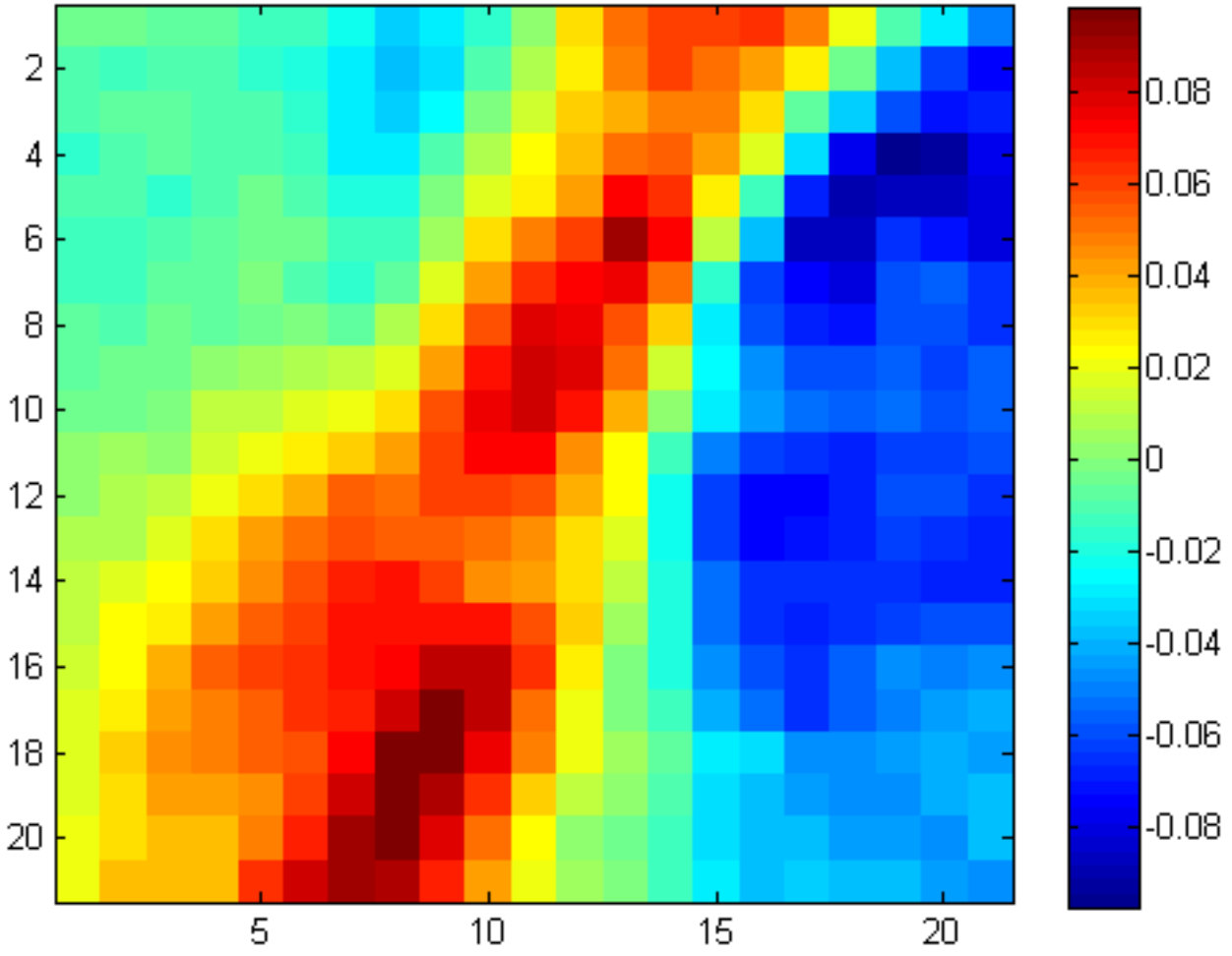
i (srch,

Obj Id	398
srun	487
trun	829
rr	99
ra	-48
gmag	0.0
rmag	0.0
imag	0.0
Flags	

- (•) [0] None
- ( ) [1] Artefac
- ( ) [2] Moving
- ( ) [3] Sat. Sta
- ( ) [4] Dipole
- ( ) [5] Variabl
- ( ) [6] Transie
- ( ) [9] Cosmic

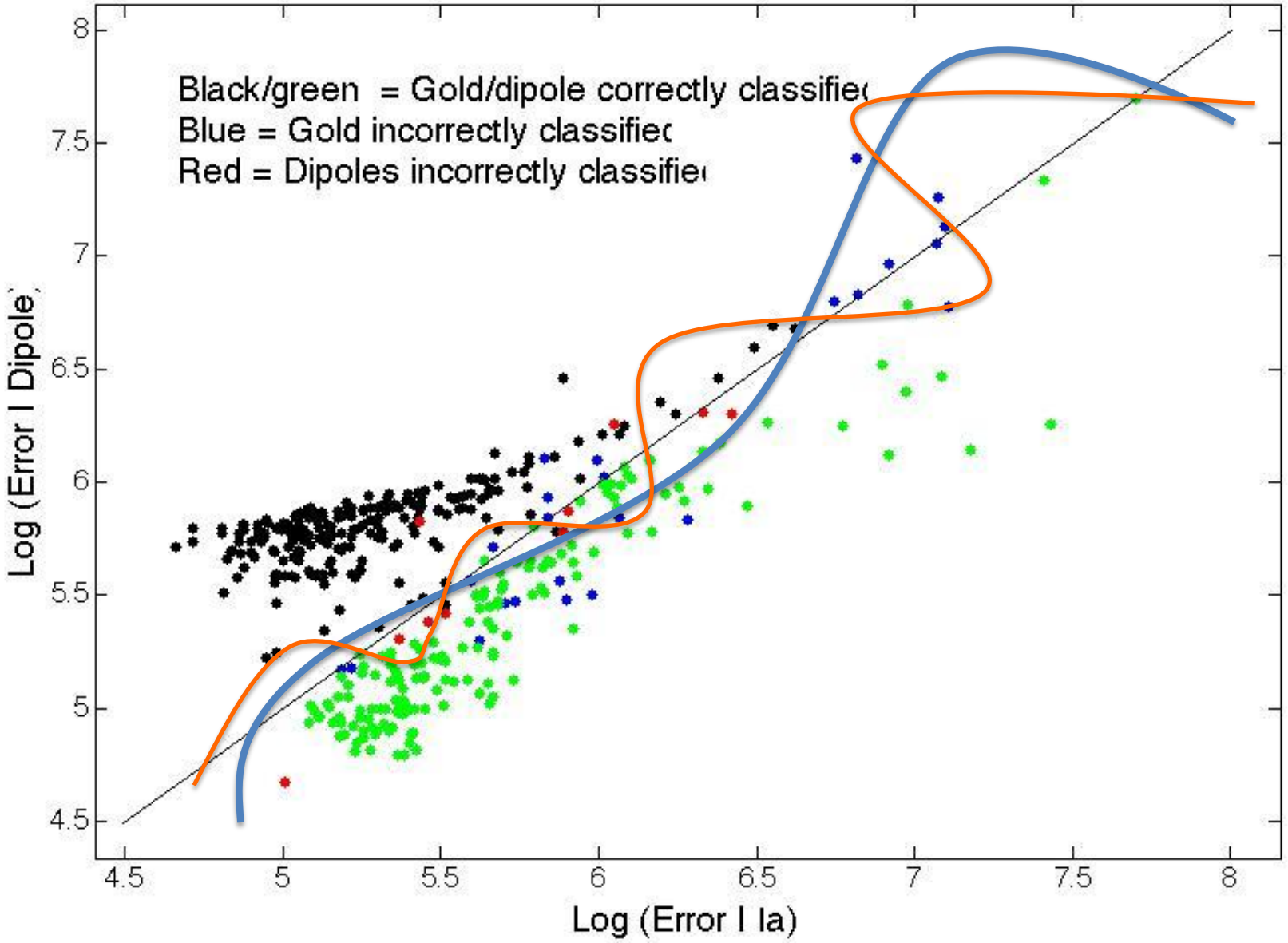
History for Object Id 398424 By Position

Feature extraction using Eigenimages



du Buisson, BB et al 2013

Classification…

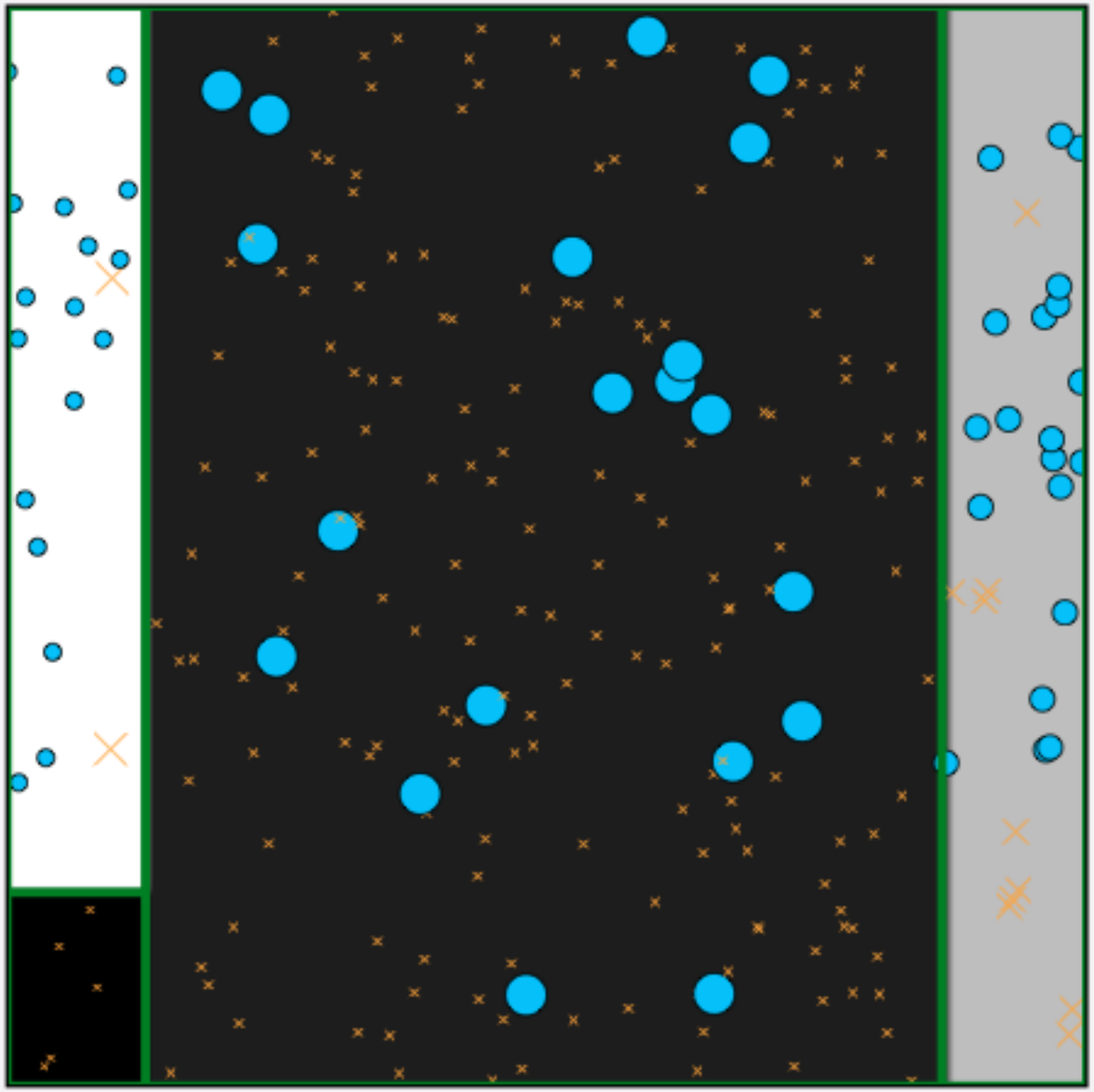


Example: Boosting

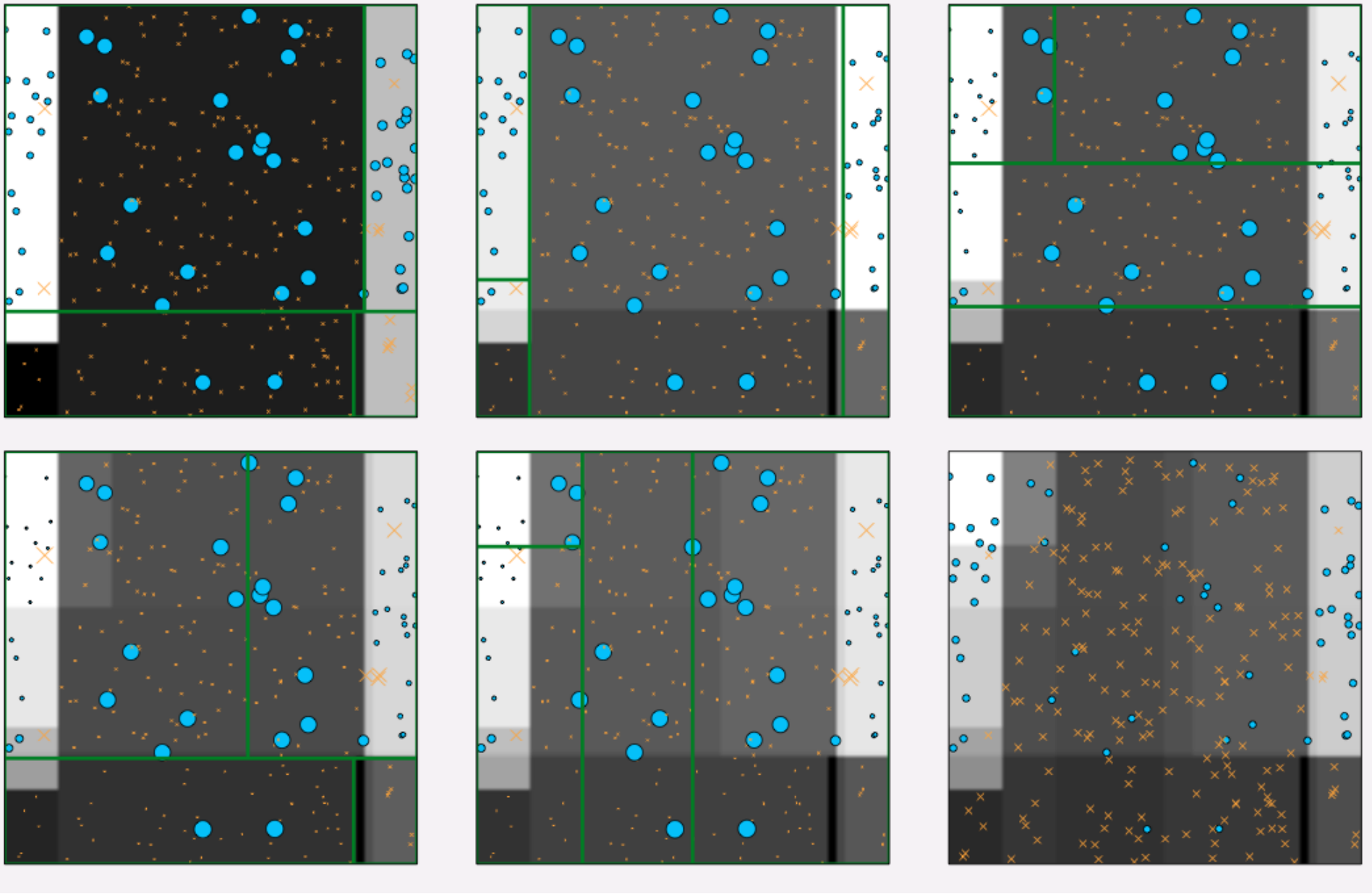
An ensemble classifier based on trees…

Combining many quasi-independent classifiers gives a better classifier

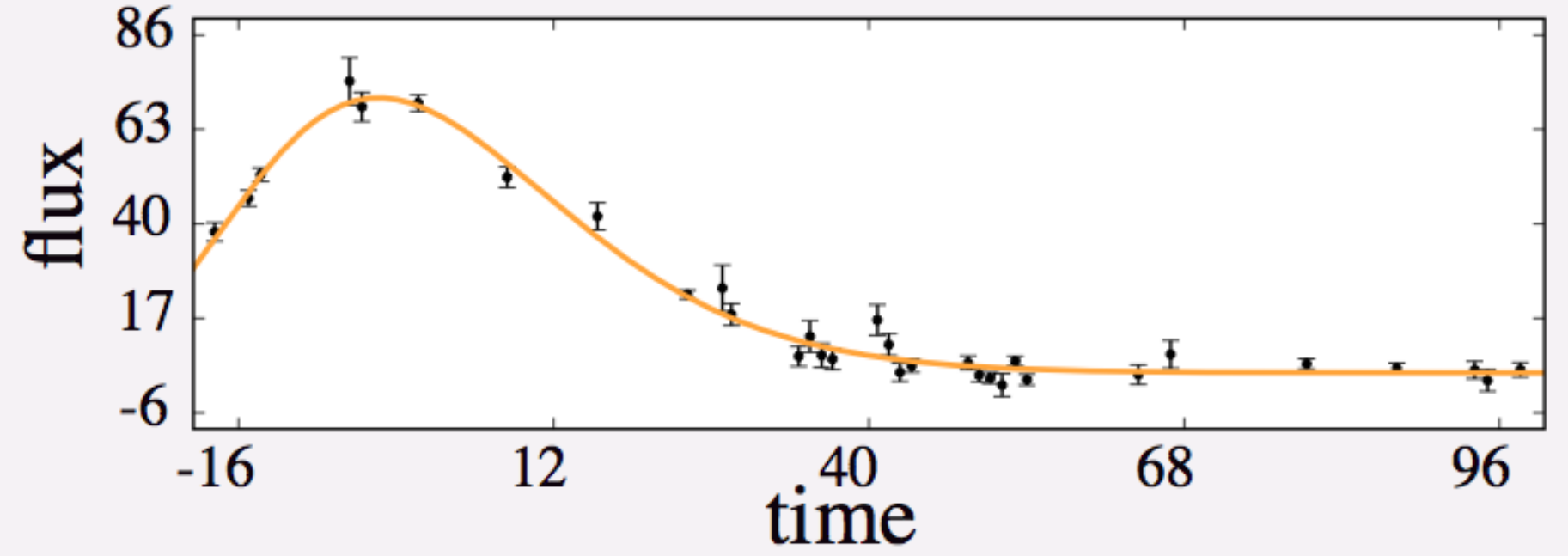
(democracy should allow better decision making)

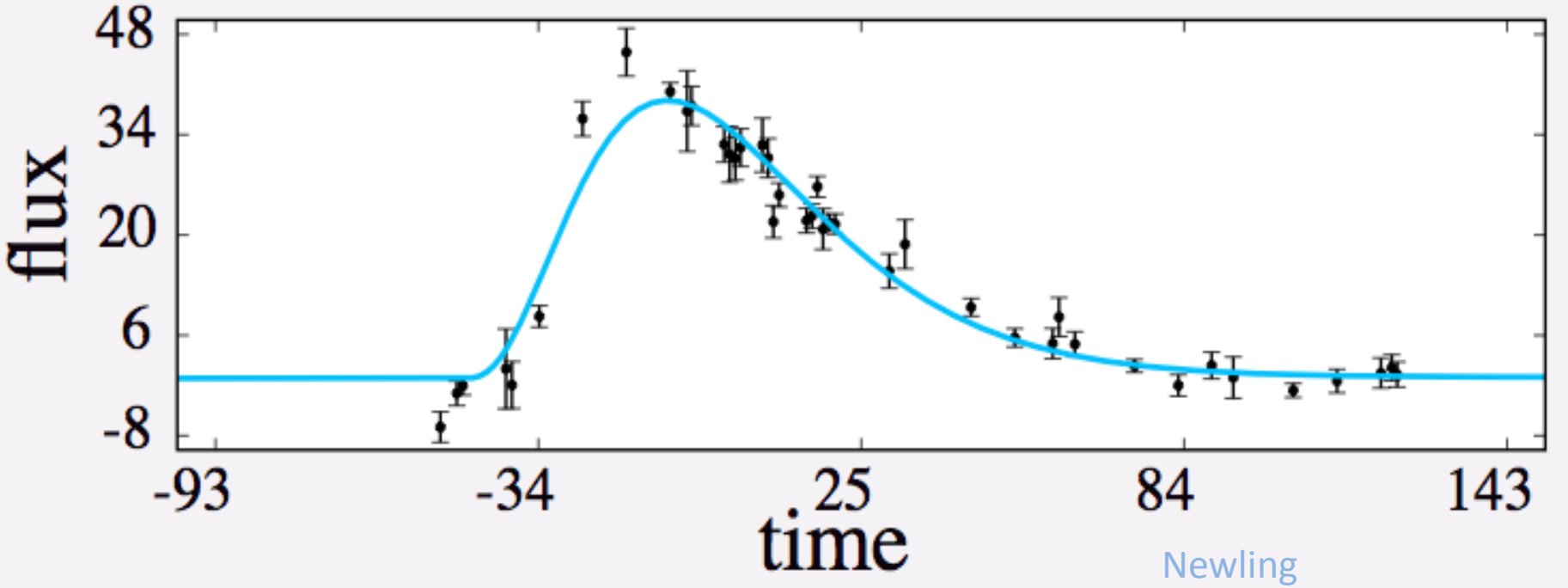


Newling



Newling


86
63
40
17
-6
flux
-16
12
40
68
96
time

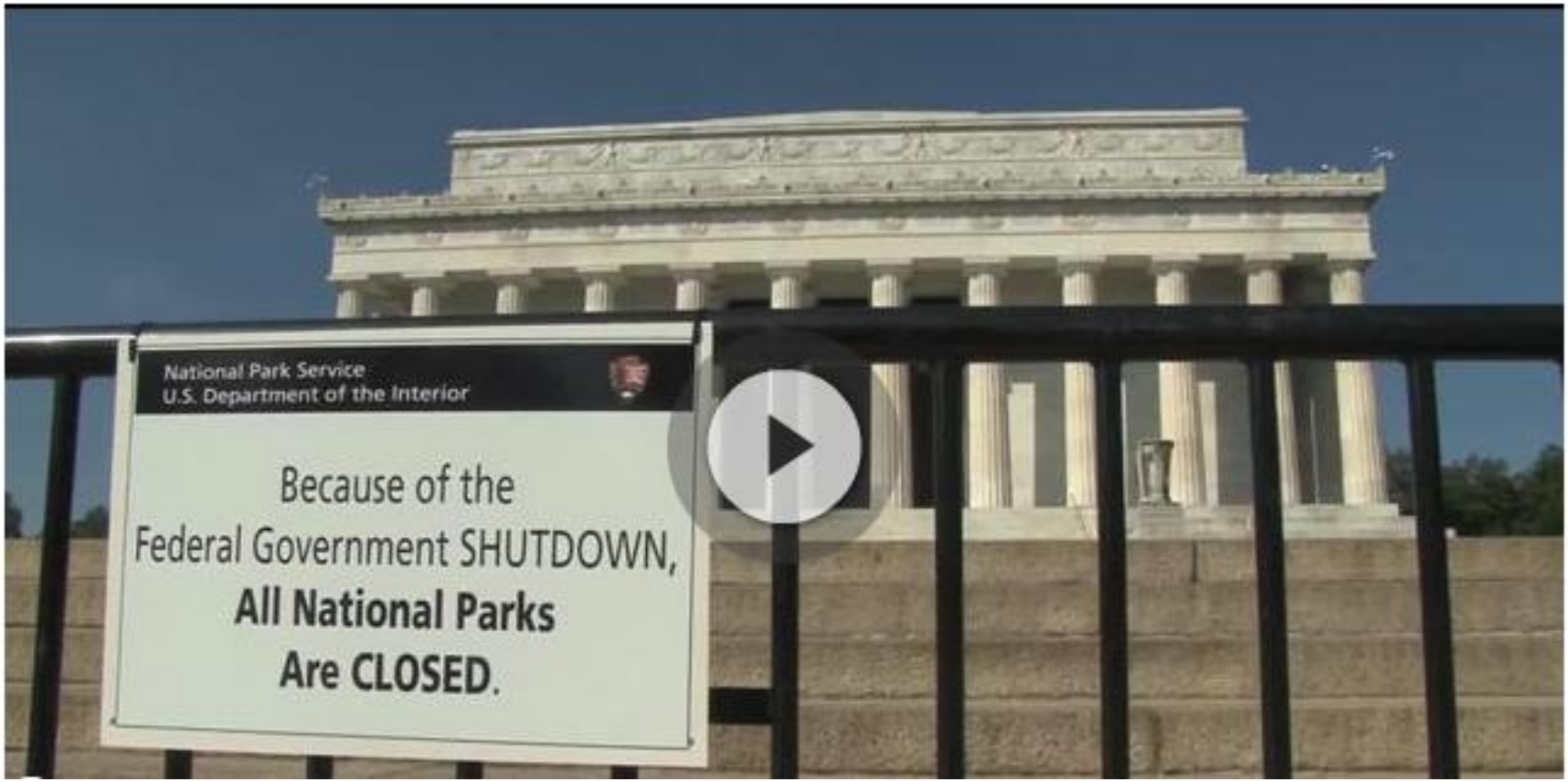

48
34
20
6
-8
flux
-93
-34
25
84
143
time

Newling

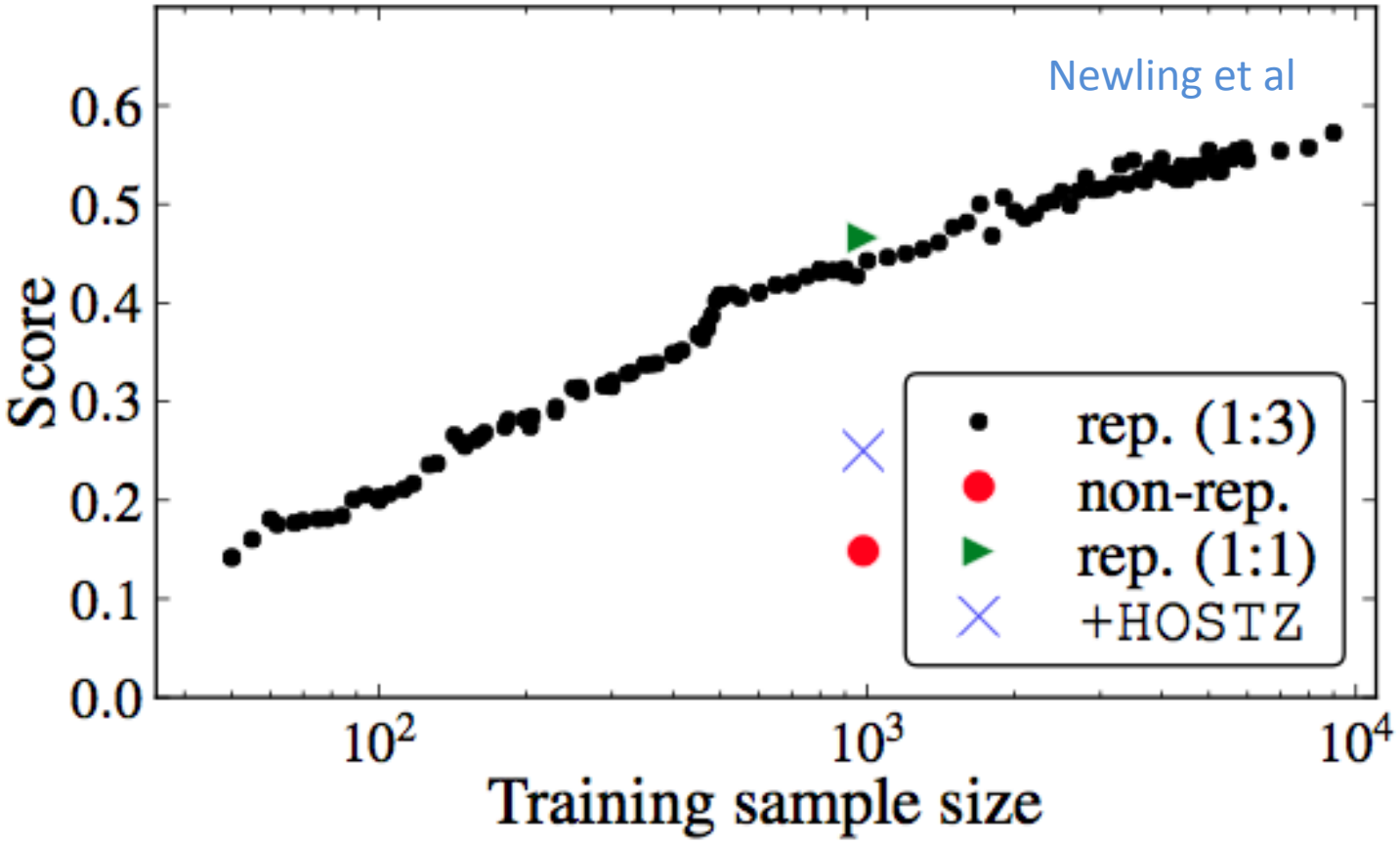
News / USA

Obama: GOP Holding Americans, Economy Hostage in 'Ideological' Shutdown

Print Comment (19) Share:



Representative?



How do we trust machine learning results?

WIRED MAGAZINE: 16.07

SCIENCE : DISCOVERIES

The End of Theory: The Data Deluge Makes the Scientific Method Obsolete

By Chris Anderson 06.23.08

Illustration: Marian Bantjes

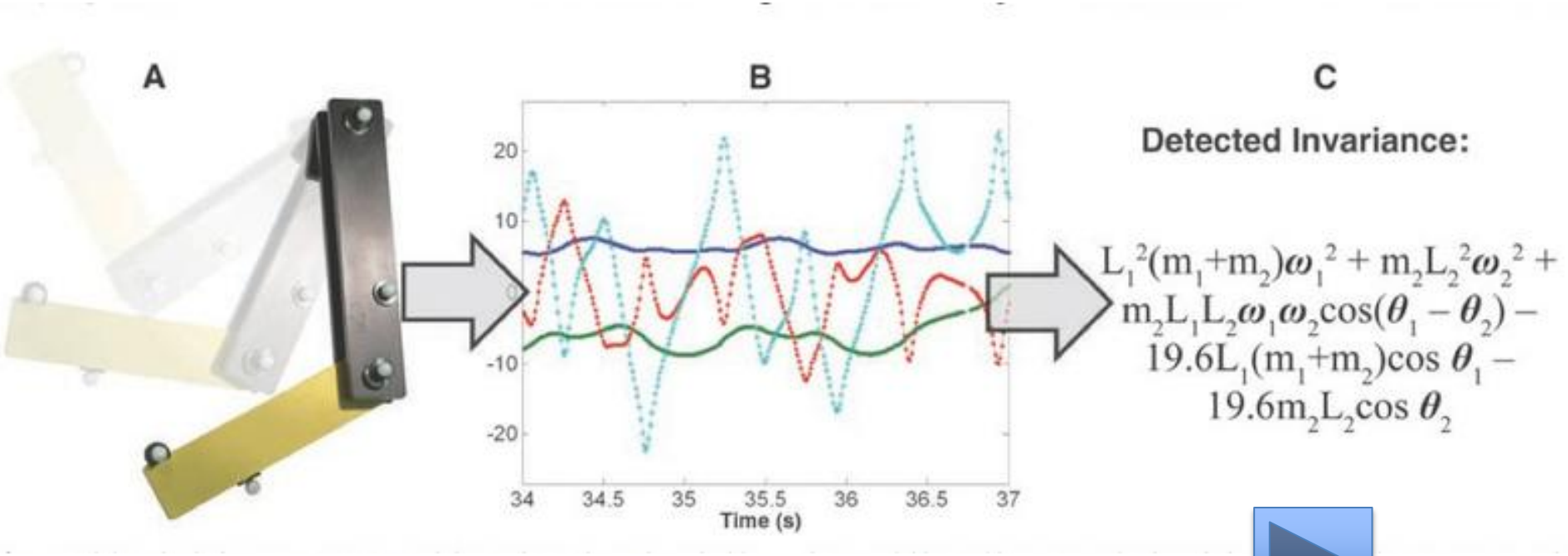
“

There is now a better way. Petabytes allow us to say: "Correlation is enough." We can stop looking for models. We can analyze the data without hypotheses about what it might show. We can throw the numbers into the biggest computing clusters the world has ever seen and let statistical algorithms find patterns where science cannot.

”

Some examples…

Discovering Newton's Laws with symbolic regression…



"Distilling Free-Form Natural Laws from Experimental Data." Schmidt and Lipson. Science,

Letters to Nature

Nature **427**, 247-252 (15 January 2004) | doi:10.1038/nature02236; November 2003

Functional genomic hypothesis generation and experimentation by a robot scientist

ADAM

"The first non-human contribution to human Knowledge"

But is this really what we mean by (great) science?

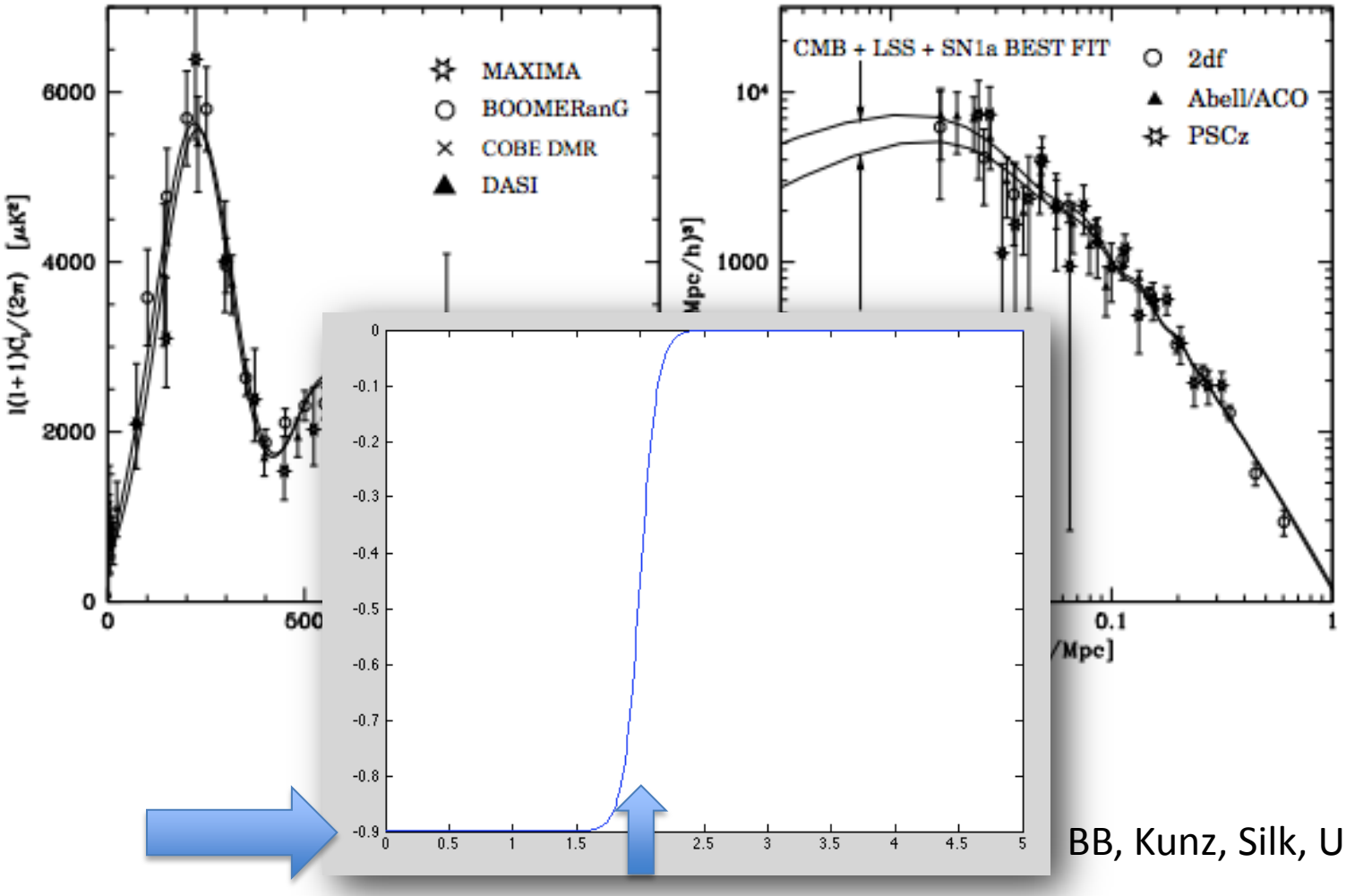
Not really

- Because great science involves two things:
  - Finding the right representation/parameterisation (the “genius” part)
  - Finding the right parameters within the representation (the “algorithmic” part)

Example: Discovering GR from the perehelion shift of mercury…already known to be a problem by 1859.

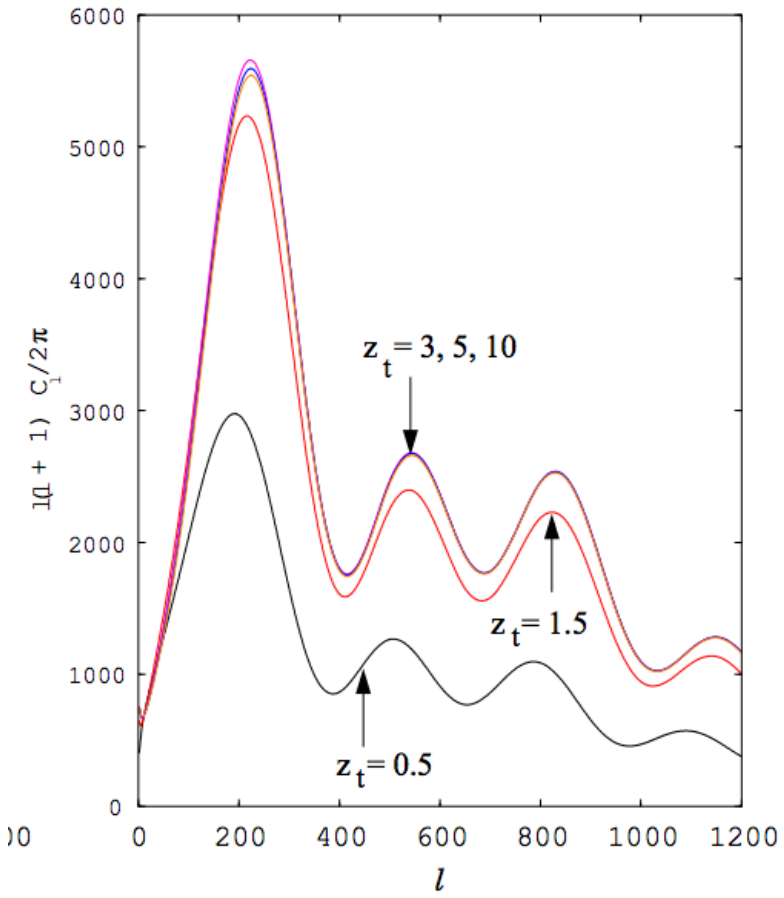
Non-representative Training Set!

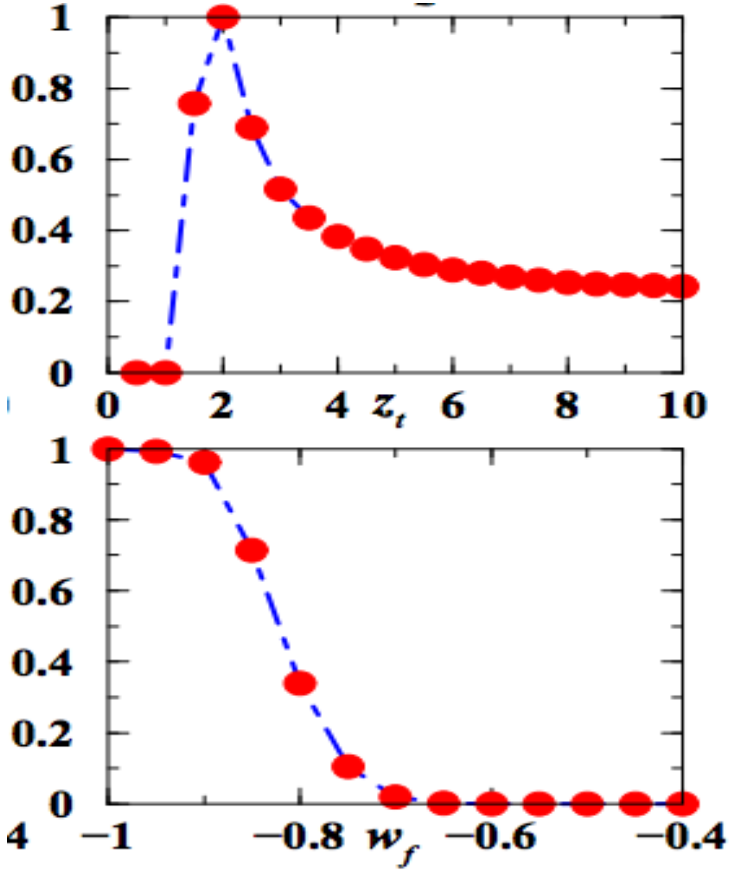
The first claim of a sudden transition in w(z) - 2002



BB, Kunz, Silk, Ungarelli 2002

The first claim of a sudden transition in w(z) - 2002





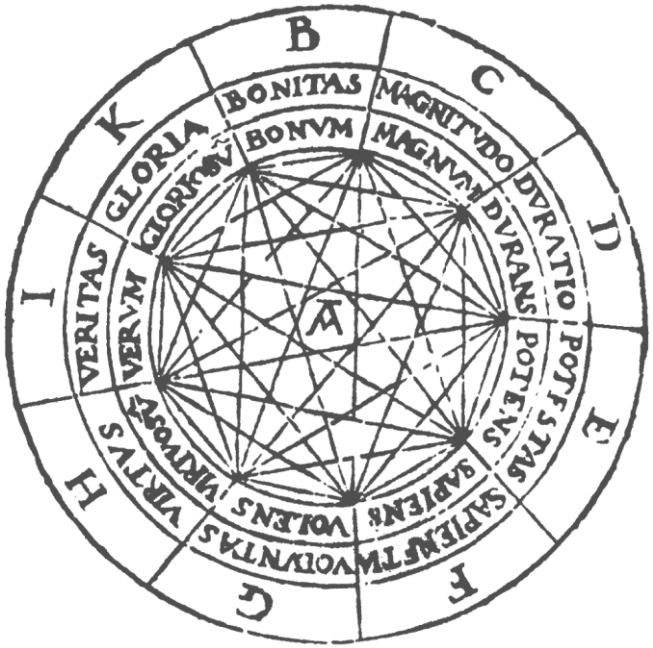
We assumed $n_s = 1$ and $\tau = 0$

Why did we do it?

- Because everyone did it at the time (bad training set!)
- Because we did grid computations, even though MCMC had been published a couple of years before.
- We made the mistake of not looking at what might cause the same effect…
- A machine learning algorithm might not have made the same mistake…

To do “real” science with machine learning one would need a universal space for describing general theories that one could then search through algorithmically, removing the need for “genius”...

Historical ideas about the perfect language



- 1200's – Raymond Lull

- 1600's Leibniz and the *Characterisitca Universalis*

  which lead to calculus. Unpopular at the time because it seemed to remove the need for creativity…

“

We have spoken of the art of complication of the sciences, i.e., of inventive logic... But when the tables of categories of our art of complication have been formed, something greater will emerge. For let the first terms, of the combination of which all others consist, be designated by signs; these signs will be a kind of alphabet.

”

- Leibniz 1666 “The art of combination.”

Hilbert c.1900: formal axiomatic theory

Goedel/Turing: Not possible – incompleteness, uncomputability.

"This statement is unprovable"

The Halting Problem

Can physics be fundamentally described only in terms of computable objects?

Enumeration of 4d topologies…

Despite this, today there is very active work on automated theorem generation and automated reasoning...

Combine this with numerical methods and machine learning and it might soon be possible to have fully automated workflows.

We joke about computers writing scientific papers for us, but perhaps it is us who need to learn to write papers “properly”?

- Arguably the best-written papers are the ones that most closely follow the ideal of

{axioms, assumptions} + {data}
+ {derived propositions}
→ results

- We try to emulate a digital/algorithmic ideal in our papers. i.e. we try to emulate computers (modulo the explanations and analogies used to explain our thinking)

In 1900 this is what it meant to be a “computer” in astronomy:

http://blogs.smithsonianmag.com/

Perhaps by 2030 the meaning of “computer” will have changed dramatically again…