# Using Visualization in high volume data reduction pipelines for quality control purposes

Patricio F. Ortiz

GAIA project,

University of Leicester

# Visualization in data reduction pipelines

Patricio F. Ortiz

GAIA project,

University of Leicester

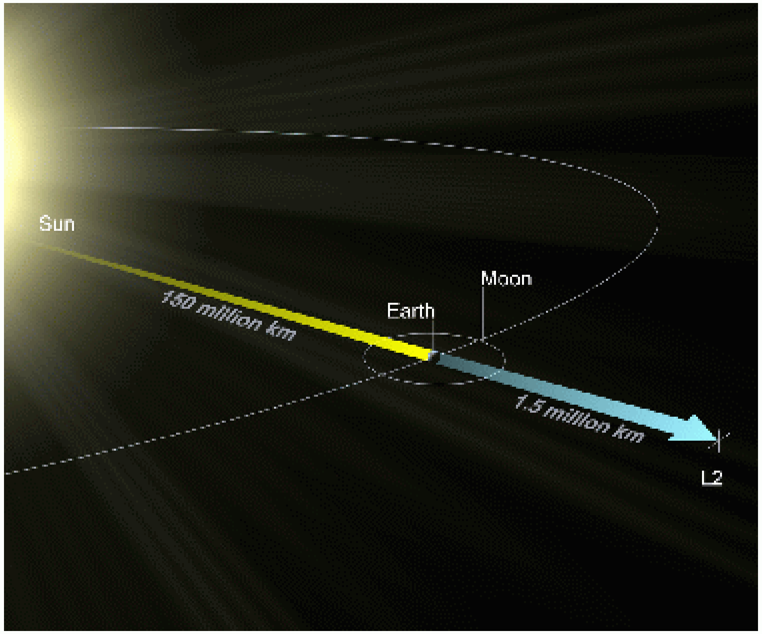
# Talk overview

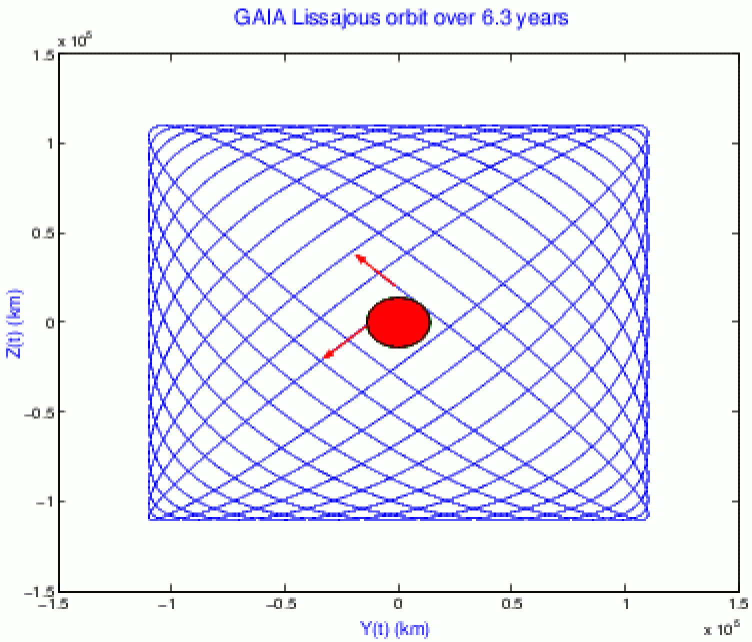
- Motivation: GAIA snapshot
- Quality control: why?
- Quality control: how?
- Visual perception is in the mind…

# GAIA snapshot

- Space mission, E.T.L: 2011
- Aim: astrometry of ~1billion stars, precision 10 micro arcsecs or better
- Length of mission: 5 years
- Each object will be observed 100 times
- 100 billion single observations

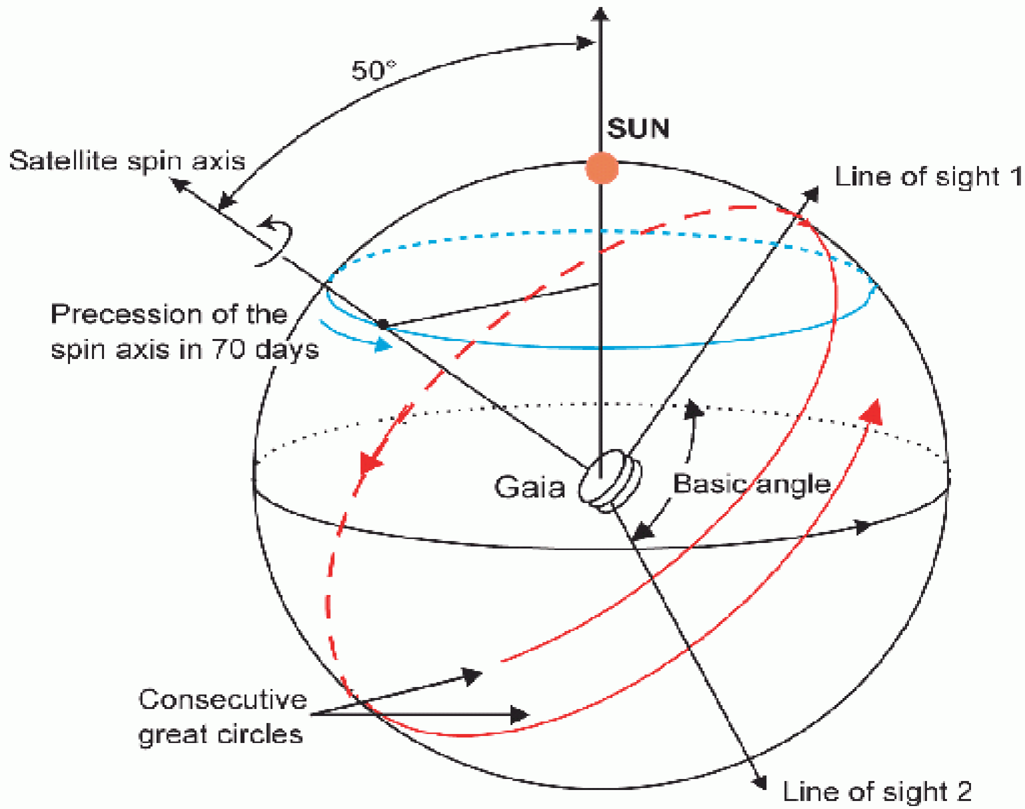
# GAIA orbital location



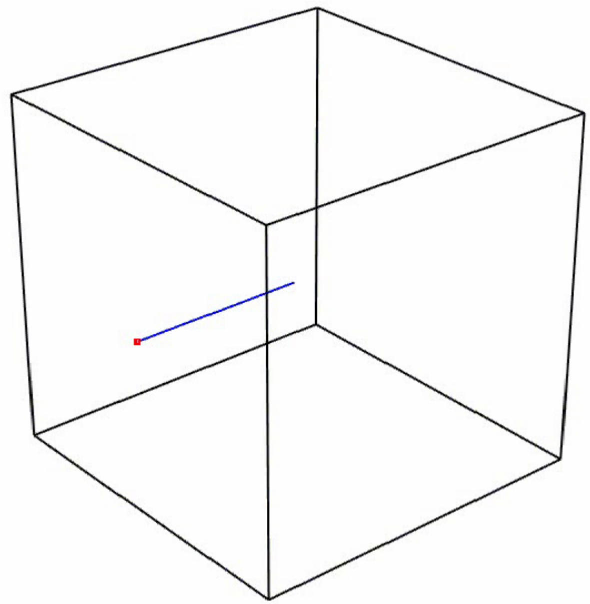

Left: The second Lagrange point lies on the Sun-Earth line, in the direction opposite to the Sun, at a distance of 1.5 million km from the Earth. L2 is a semi-stable region of gravity where spacecraft can be maintained for several years with cheap orbit manoeuvres. Right: Example of a Lissajous orbit projected on the plane perpendicular to the Earth-L2 line, as seen from the Earth. The initial conditions are chosen such that the orbit wanders outside the Earth shadow (red circle at the centre) until the occurrence of the next eclipse more than six years later.

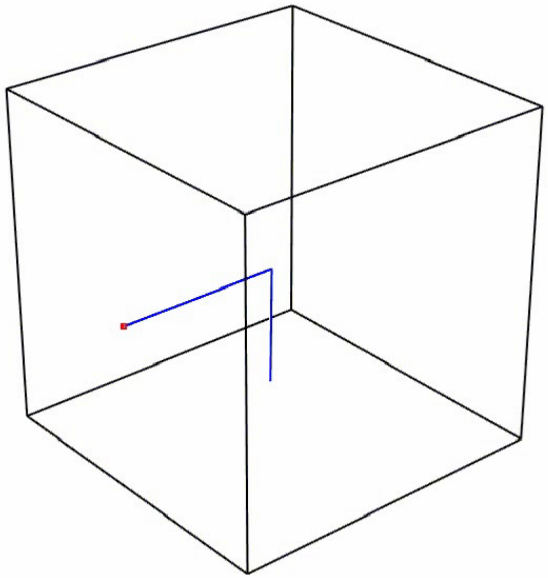
# GAIA Scanning law

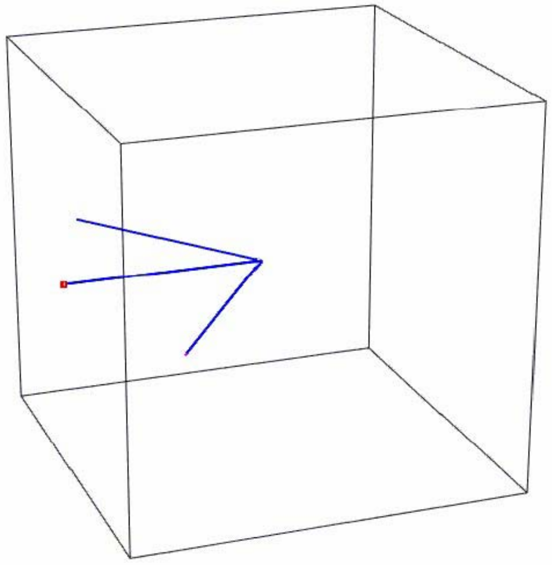



Gaia's two astrometric fields of view scan the sky according to a carefully prescribed 'revolving scanning law'. The constant spin rate of 60 arcsec/s corresponds to 6-hour great-circle scans. The angle between the slowly precessing spin axis and the Sun is maintained at 50°. The basic angle is 99.4°.

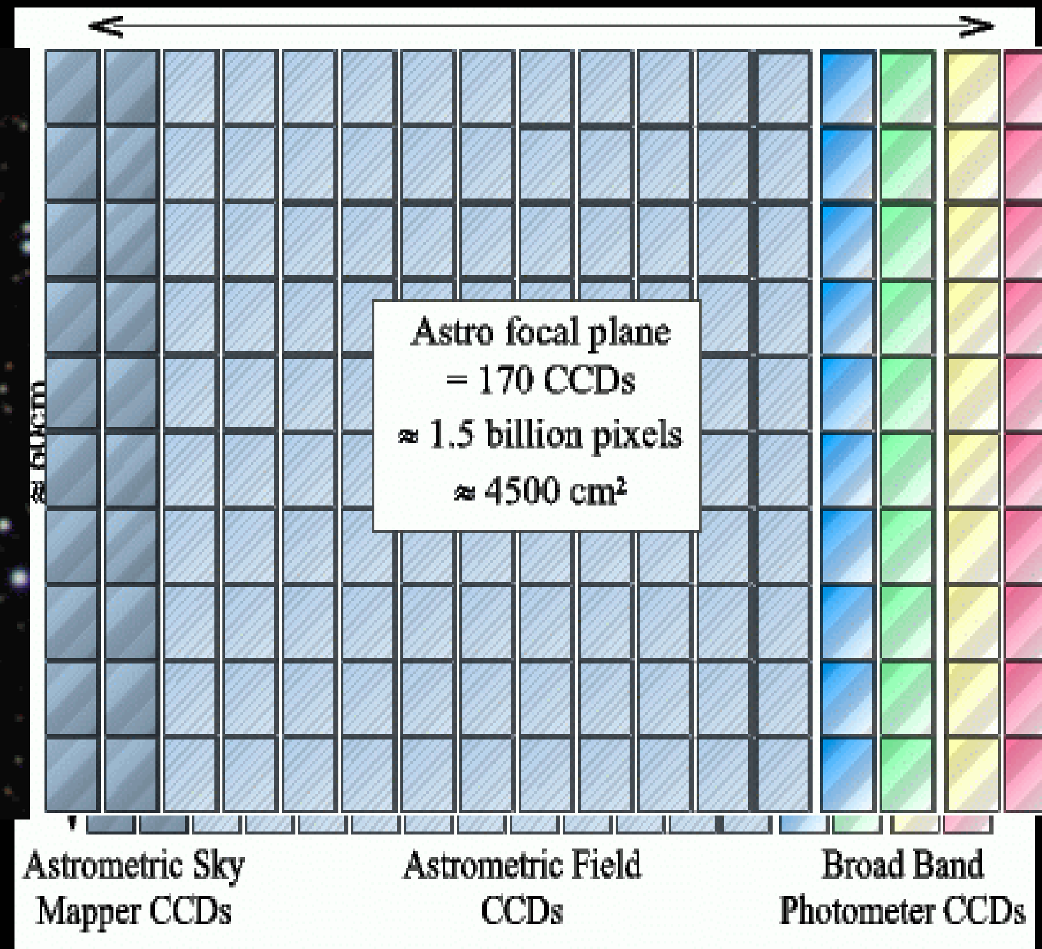
# GAIA: scanning the sky







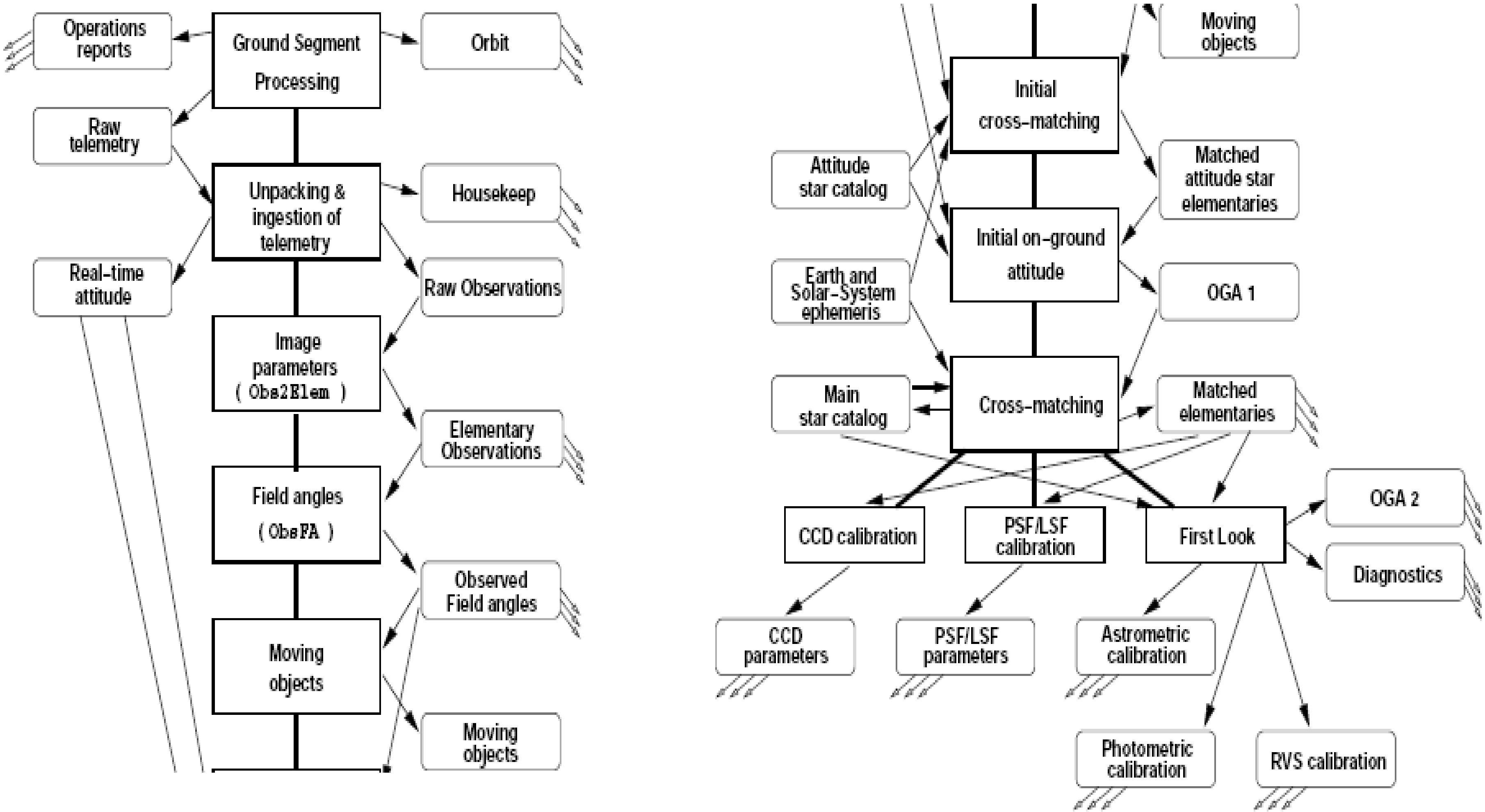
# Gaia observing mode: 24/7



# The challenge of reducing the data

- Volume: continuous observations, no stop for calibration frames, or anything (except perhaps satellite going into safe-mode)
- Integrated solution: solving for positions of one object needs a global solution (GIS)
- Solutions will be improved as new observations come in.
- Distributed nature of the project

# Ground data movements



# Quality control: why?

- *“if something can go wrong, it will…”* Murphy
- Some things that can go wrong:
- Satellite scanning parameters may change
- Instrumental behaviour
- Data processing on-board
- Data download / acquisition
- Data reduction: metadata errors, DB problems, Cross-matching errors, calibration errors, etc.

# Quality Control: how? I

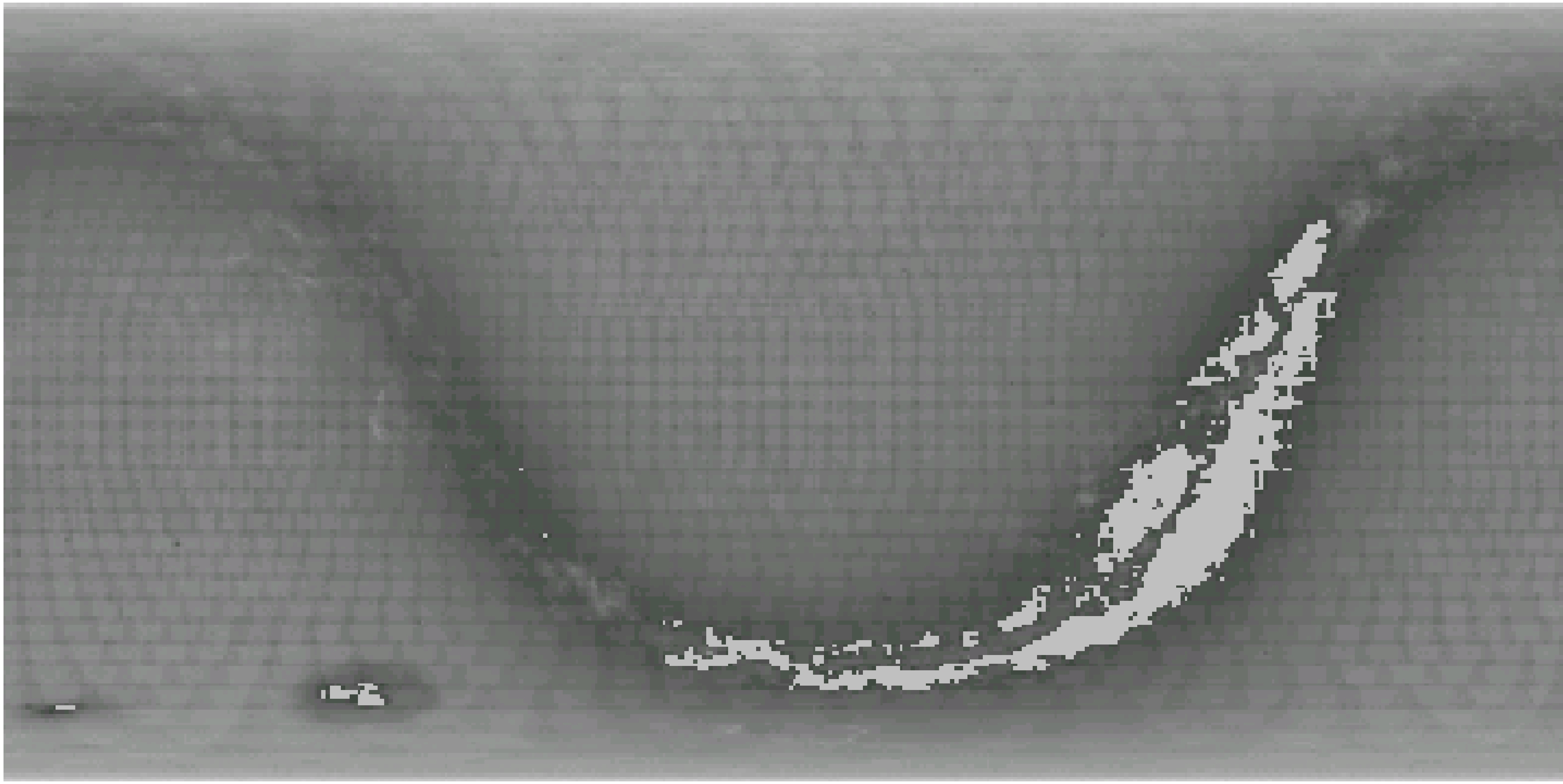
- *“We see what we are prepared to see…”*

- Build dedicated software monitors for each possible failure point, and each possible situation.
- Make these results available to other monitors with enough metadata to make sense.
- Store the monitoring information in a simple to access system (DBMS or flat files).
- Time stamping is paramount.

# Quality control, how? II

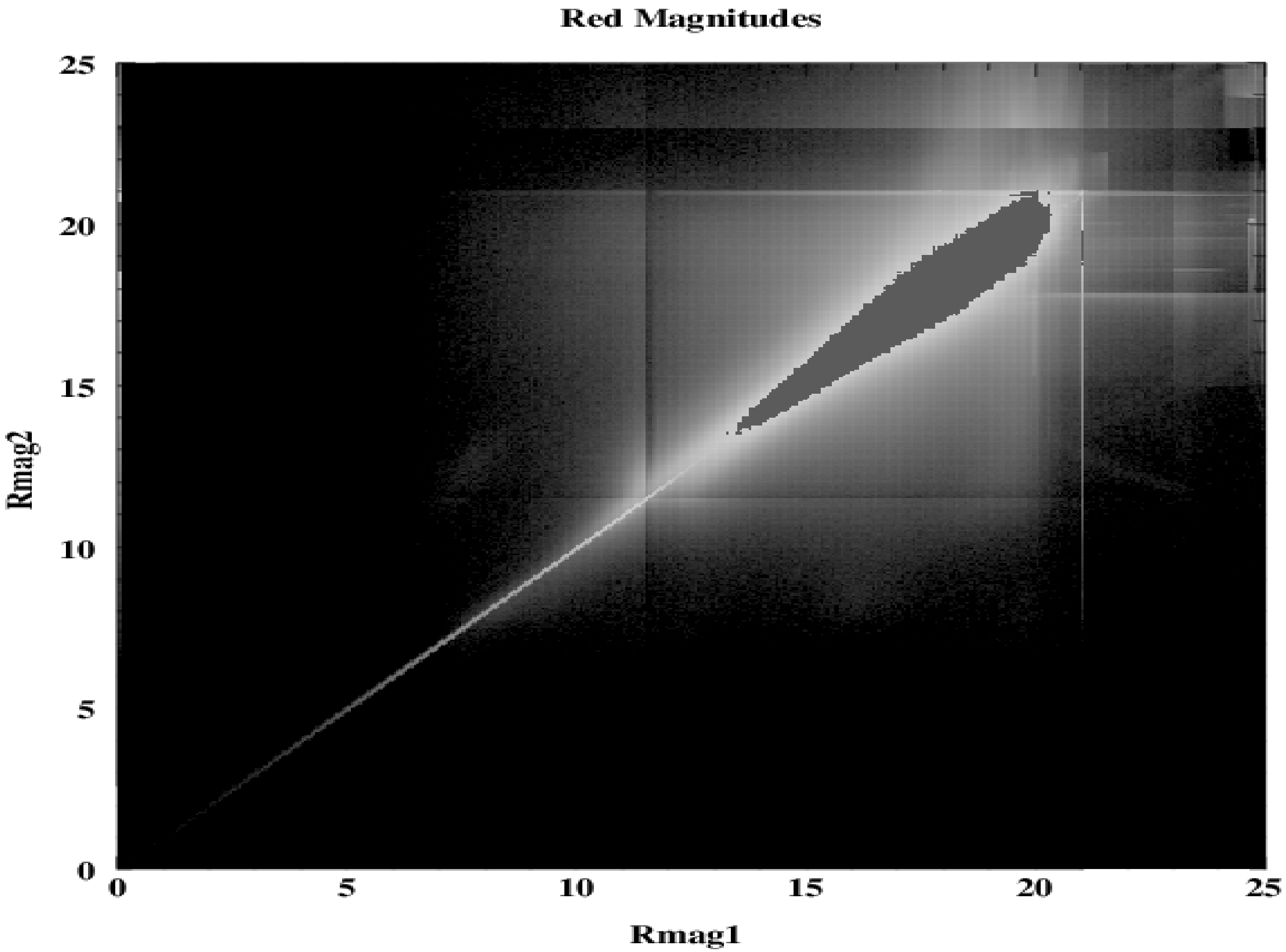
- “Expect the unexpected…”
- Need for flexible ways to inspect results or intermediate stages at random
- Need to discover trends, at any timescale. Averages don’t always tell the whole story
- Vision is a fundamental discovery / interpreting tool, therefore, visualization can be used to a big advantage.

# Real examples I



Sky density of USNOB, 10^9 stars

# Real examples II




USNOB 10^9 stars, some “systematic effects”

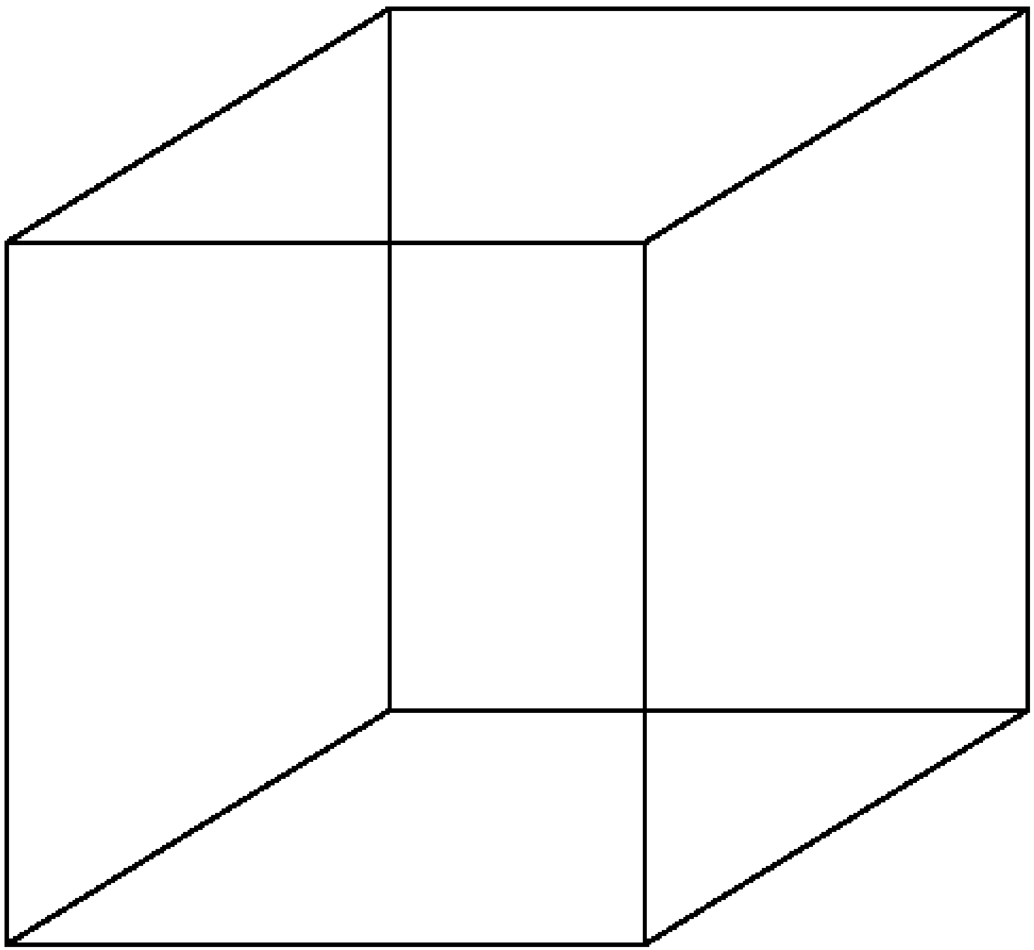
# Real examples III


Location of outliers in 2, 3 or N dimensions is important in detecting deviations or other artifacts

Animations provide a good way to "explore an N-d space", particularly, flybys

Tomography (slicing the space) is also key

# Visual perception is in the mind



- Reality
- Training
- Perception

# Mental features we need to be aware of

- Great perceiving changes (ergo, mpeg)
- Perception thresholds exist
- Log scale is natural in our sight and vision
- We have limited visual scope (displays subtending a small solid angle make sense)
- Clustering (and statistical analysis) are part of our brains since early age.
- Attention to certain features takes away “awareness” about other features occurring at the same time (the Houdini effect)

# Visualization in the XXI century I

- We still view things from 2-D projections (visual caves did not pick up)
- 3D projections for the PlayStation and Xbox are well alive.
- If we have a finite amount of pixels to use, let’s do most out of them, enable zooming, change of POV, transparency, etc in the data.
- Need to handle & show information for zillions of points at a time… drop scatter plots!!
- Adopt density plots instead, split the space in boxes, cubes, hyper-cubes and then count the number of points. Show that number. Allow weights

# Visualization in the XXI century II

- Density plots are “lighter” for transmission, printing, and even storage.
- They convey much more information in areas of high density than scatter plots
- They can be represented in linear or log scale to enhance areas of low density
- Free manipulation of colour scales helps feature discovery greatly
- Just like we manipulate astronomical images in FITS, we might reach the point in which we want to manipulate representations.
- We might want to manipulate layers, like in photography manipulation software, or like in photography, burning or dodging

# Requirements for successful visualization in a high volume environment

- Effective access to all aspects of the data, so that exploration of initially unforeseen spaces is possible.
- A generalized data storage/access approach would simplify putting together data which could be considered unmergeable (e.g., .science - engineering data)
- Integration of Data Access and visualization systems to be able to handle the data volume avoiding tremendous I/O overheads.
- Use of optimized high volume visualizing techniques, including animations, flybys and tomographic views