The LOFAR Transients Pipeline

John Swinbank
swinbank@transientskp.org
The “transients pipeline”

➡ Ingests image cubes (position, frequency, polarization)

➡ Identifies and classifies transients & variable sources within the images

➡ Results in:

➡ Alerts, either within LOFAR or to the community

➡ Archive database of classified lightcurves
The Radio Sky Monitor

- Multiple LOFAR beams tile out the sky
- Individual beams on specific targets
- Imaging at 1, 2, 5, 10, ... second cadence
- Observation strategy under development
...but also

- Monitoring of specific fields
- Piggybacking
- Trawling through the LOFAR archive
- Trawling through other archives (VLA...)
- AARTFAAC
- ...etc
OUTLINE LOFAR TOPOLOGY

BG/P → Cluster → Archive

“ONLINE” → “OFFLINE”
LOFAR TRANSIENT SYSTEM OVERVIEW

- LOFAR Online Systems
- Imaging Pipeline
- Image Cube
- Source Finding

- MonetDB Database
  - Lightcurve Storage
  - Source Association
  - Transient & Variability Analysis
  - Archive Database

- Pipeline Processing
  - Off-line & External Systems

- Classification & Analysis
- Response Scheduling

- Send External Alert
- Schedule New Observation
- Re-run Image Analysis

- Other Observatories
THE LOFAR PIPELINE FRAMEWORK
LOFAR TRANSIENTS SYSTEM OVERVIEW

LOFAR Online Systems -> Imaging Pipeline -> Image Cube -> Source Finding

Source Finding -> Lightcurve Storage

Lightcurve Storage -> Classification & Analysis

Classification & Analysis -> Response Scheduling

Response Scheduling -> Re-run Image Analysis

Re-run Image Analysis -> Schedule New Observation

Schedule New Observation -> Send External Alert

Send External Alert -> Other Observatories

Other Observatories -> Archive Database

Archive Database -> Transient & Variability Analysis

Transient & Variability Analysis -> Source Association

Source Association -> MonetDB Database

MonetDB Database

Pipeline Processing

Off-line & External Systems
Source-finding

- Custom-developed source-finding code
- (Almost) pure Python (+ NumPy, SciPy, etc)
- Available both in a pipeline form, and for interactive use
- Largely developed by Hanno Spreeuw as part of his PhD work at the University of Amsterdam
- New maintainer: John Sanders (Univ. Portsmouth); gearing up for a proper release
Another sourcefinder...

- Fitting all detected sources with elliptical Gaussians (or other shapes)
- Deblending composite sources
- False detection rate algorithm (Benjamini & Hochberg, 1995)
- Proper treatment of errors in the presence of correlated noise (after Condon, 1997)
- Formidable battery of statistical tests (see Spreeuw’s thesis, 2010)
Database

★ Two databases: “pipeline” and “archive”

★ Both based on the same architecture

★ Pipeline database supports real-time pipeline processing

★ Archive database provides long term storage and data mining of up to 100 TB/year.

★ Subject of PhD thesis by Bart Scheers (2011); ongoing development by Scheers
Source Association

- Sourcefinder results are uploaded directly into the database.
- Lightcurves are built automatically as new results arrive.
- Association is done by position, taking into account measurement errors, background source counts, etc.
- In progress development to associate across frequencies per image cube.
Transient detection

➡ Also in the database, automatically as lightcurves are grown

➡ Measuring magnitude and significance of flux variation; easy to add more measures

➡ Query for significant objects directly into pipeline

\[ V_\nu = \frac{1}{I_\nu} \sqrt{\frac{N}{N-1} \left( \overline{I_\nu^2} - \overline{I_\nu}^2 \right)} \]

\[ \eta_\nu = \frac{1}{N-1} \sum_{i=1}^{N} \frac{(I_{\nu,i} - \overline{I_\nu}^*)^2}{\sigma_{I_{\nu,i}}^2} \]

\[ \overline{I_\nu}^* = \frac{\sum_{i=1}^{N} \omega_{\nu,i} I_{\nu,i}}{\sum_{i=1}^{N} \omega_{\nu,i}} \]

\[ \omega_{\nu,i} = \frac{1}{\sigma_{I_{\nu,i}}^2} \]
MonetDB: performance

- ~10000 insertions/second for "full rate" radio sky monitor
- MonetDB makes this practical
- "Column store" architecture
- High-perf numeric kernel
- etc etc
- Killer feature: collaboration with CWI developers
LOFAR TRANSIENTS SYSTEM OVERVIEW
Feature measurement

- Simple Python code that operates on a lightcurve extracted from the database

- Arbitrary properties can be defined, from the simple (average flux, ...) to the complex (fitting parameters, ...)

- The lightcurve is annotated with the features, then passed to the classifier.
Classification

- Manual
  - Astronomer-designed “plugins” identify certain combinations of features
  - Easy to extract your “favourite” sources from the data... if you can describe how they behave

- Automatic
  - Speculative!
  - See Masters thesis by Thijs Coenen (2008)
  - Re-use of feature-extraction code
LOFAR TRANSIENTS SYSTEM OVERVIEW

MonetDB Database
Imaging
Pipeline
Source Finding
Lightcurve Storage
Classification & Analysis
Response Scheduling
Classification & Analysis
Source Association
Receive External Alert
Re-run Image Analysis
Schedule New Observation
Send External Alert
Archive Database
Other Observatories

Pipeline Processing
Off-line & External Systems

LOFAR On-line Systems
Imaging Pipeline
Image Cube
Source Finding

Transient & Variability Analysis

Send External Alert
LOFAR TRANSIENTS SYSTEM OVERVIEW

LOFAR Online Systems → Imaging Pipeline → Image Cube → Source Finding → MonetDB Database → Lightcurve Storage

Source Association → Transient & Variability Analysis

Lightcurve Storage → Classification & Analysis → Response Scheduling

Receive External Alert → Send External Alert → Schedule New Observation

Re-run Image Analysis

Other Observatories

Pipeline Processing

Off-line & External Systems
What
Param
name, unit, UCD, 
dataType, utype, value
Value, D, R
Group
name, type
Param, D, R
Table
name, type
Param, Field, Data, D, R
Field
name, unit, UCD, 
dataType, utype, value
D, R
Data
TR
TD
D, R
Reference
uri, meaning, mimetype
Description
Elements in black
Attributes in green
D = Description
R = Reference

VOEvent
version, ivorn,
role = test, observation,
prediction, utility

Who
AuthorIVORN or 
Author
title, shortName, logoURL,
contactName, contactEmail,
contactPhone, contributor

Why
importance, expires
Name
Concept
Inference
probability, relation
Name, Concept, D, R

WhereWhen
longitude, latitude, positionalError,
time, timeError
observatory, coord_system *
* equivalent information

How
D, R

Citations
EventIVORN
cite = followup,
supersedes, retraction
D

VOEVENT
XML. MACHINE READABLE. FAST TRANSMISSION. AUTOMATIC PROCESSING. THE FUTURE...
Initial versions of the database and sourcefinding systems have been tested, and produced real science:

- Bell, M.E. et al, *An automated archival VLA transients survey*, accepted by MNRAS
- See Martin’s talk on Thursday
LOFAR is now regularly observing fields regarded as “interesting” from a transients point of view.

“Offline” imaging pipeline being run for commissioning & science purposes.

All this data to be fed into the transients pipeline system.
“On-line” LOFAR imaging pipeline + AARTFAAC imaging pipeline coming in the next ~year

Transients pipeline development progressing apace; we must be ready

VOEvent trigger system tested: we have received and acted on triggers from LIGO
More information

- LOFAR Transients
  - http://www.transientskp.org/

- TKP Pipeline Docs
  - http://docs.transientskp.org/

- LOFAR Pipeline Framework
  - http://usg.lofar.org/documentation/pipeline/

- MonetDB
  - http://monetdb.cwi.nl/MonetDB/
  - http://www.scilens.org/

- PhD Theses (Spreeuw & Scheers)
  - http://www.astro.uva.nl/research/theses_phd/