Markus Neteler & GRASS Development Team
Fondazione E. Mach – CRI, Italy
http://gis.cri.fmach.it
http://grass.osgeo.org

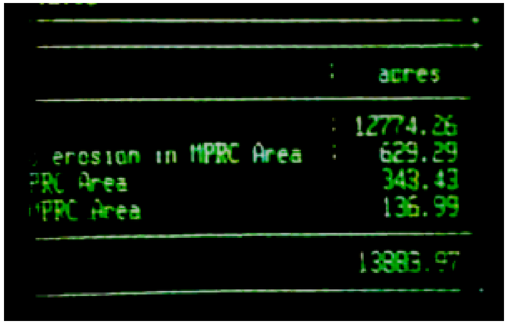
GRASS GIS 7: Efficiently processing big geospatial data

FOSDEM 2015, Brussels
31 Jan & 1 Feb 2015

```
4.8T /grassdata/eu_laea/modis_lst_reconstructed
3.6T /grassdata/eu_laea/modis_lst_reconstructed_europe_daily
2.0T /grassdata/eu_laea/modis_lst_reconstructed_europe_GDD
1.1T /grassdata/eu_laea/modis_lst_reconstructed_europe_weekly
 ...
 48G /grassdata/eu_laea/modis_lst_koeppen
 22G /grassdata/eu_laea/modis_lst_reconstructed_europe_annual
 40G /grassdata/eu_laea/modis_lst_reconstructed_europe_bioclim
275G /grassdata/eu_laea/modis_lst_reconstructed_europe_monthly
 38G /grassdata/eu_laea/modis_lst_reconstructed_europe_monthly_averages
 15G /grassdata/eu_laea/modis_lst_reconstructed_europe_winkler
 55G /grassdata/eu_laea/modis_lst_validation_europe
```

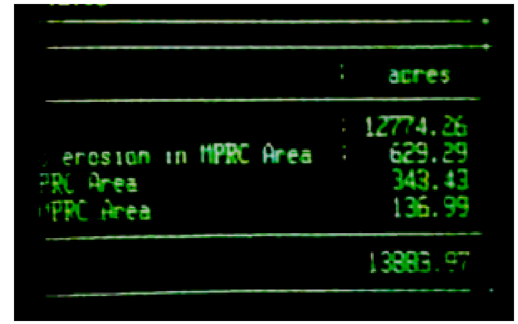
The new release: GRASS GIS 7 User interface

What you think it is...

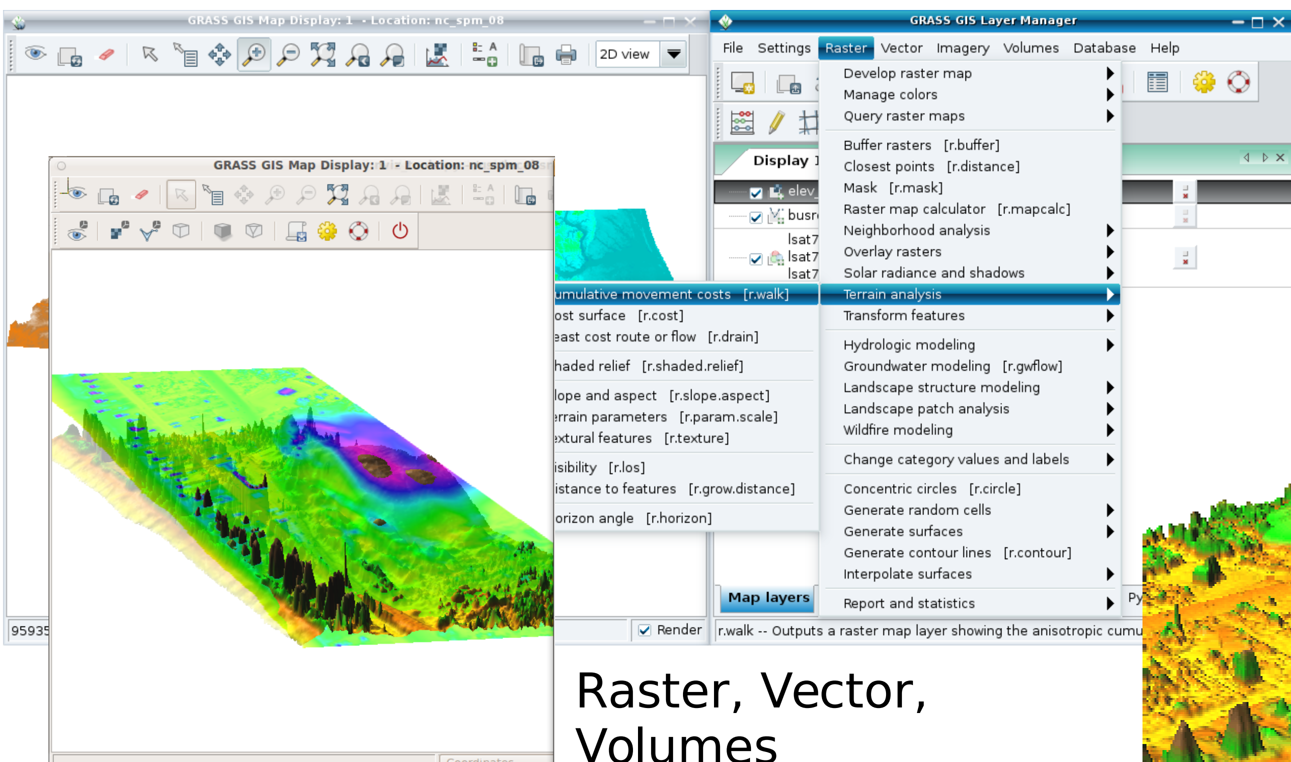



The new release: GRASS GIS 7 User interface

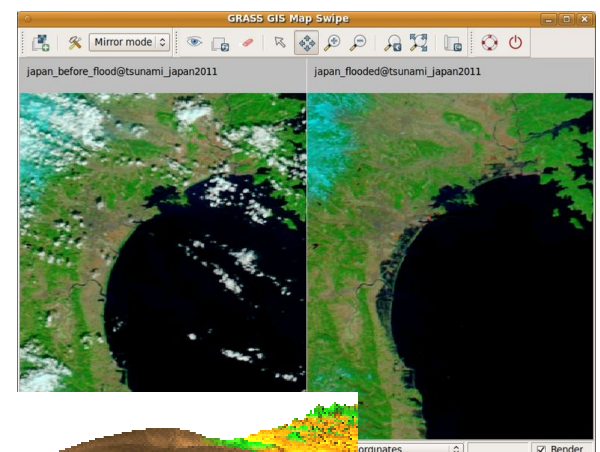
What you think it is...

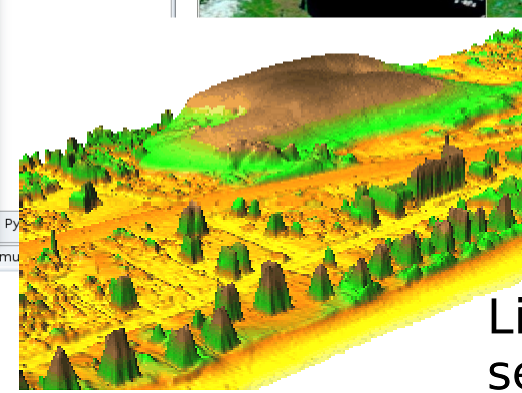


What it really is...



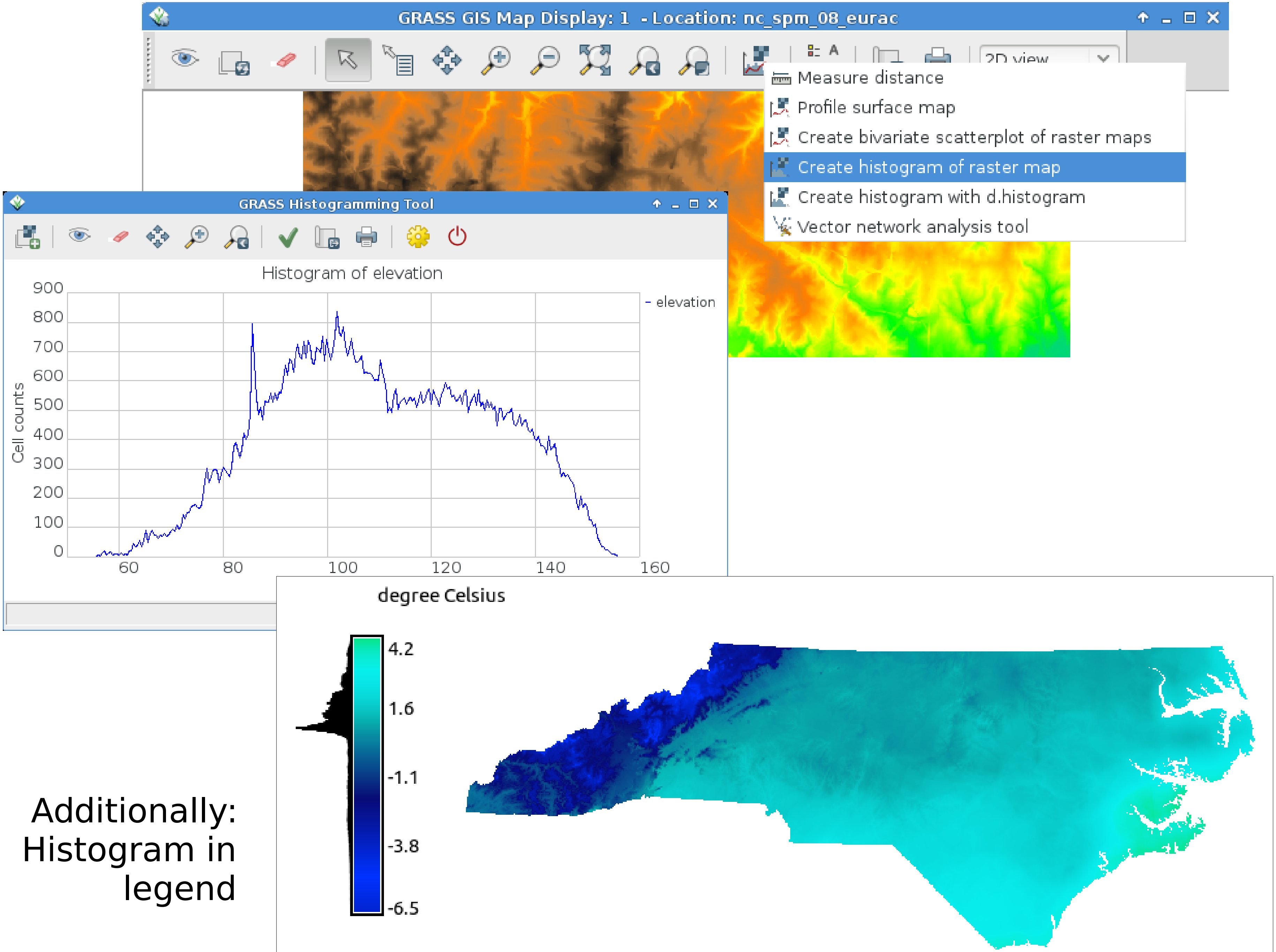
Raster, Vector, Volumes





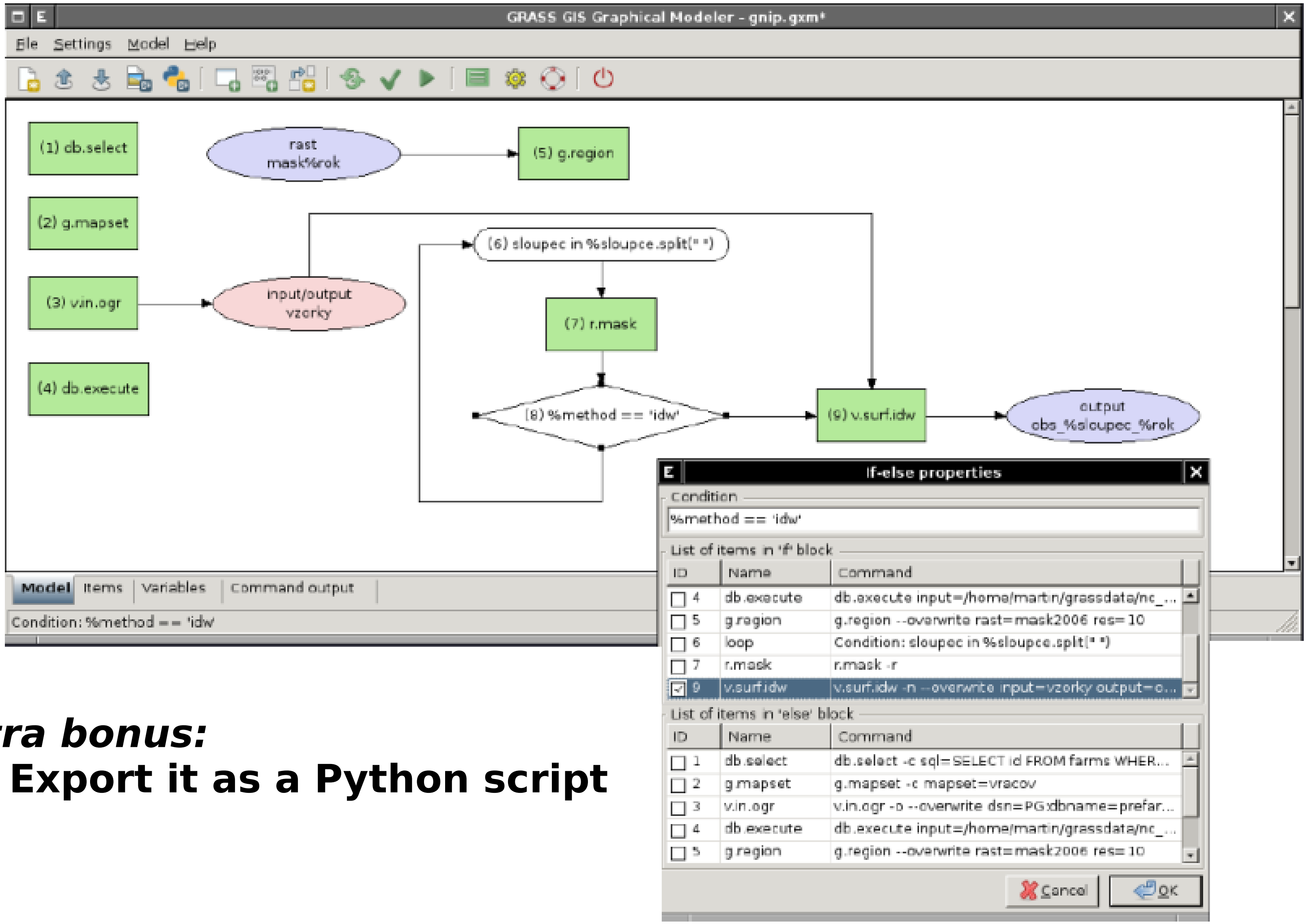
LiDAR time series

GRASS GIS 7: Map histogram tool



Additionally: Histogram in legend

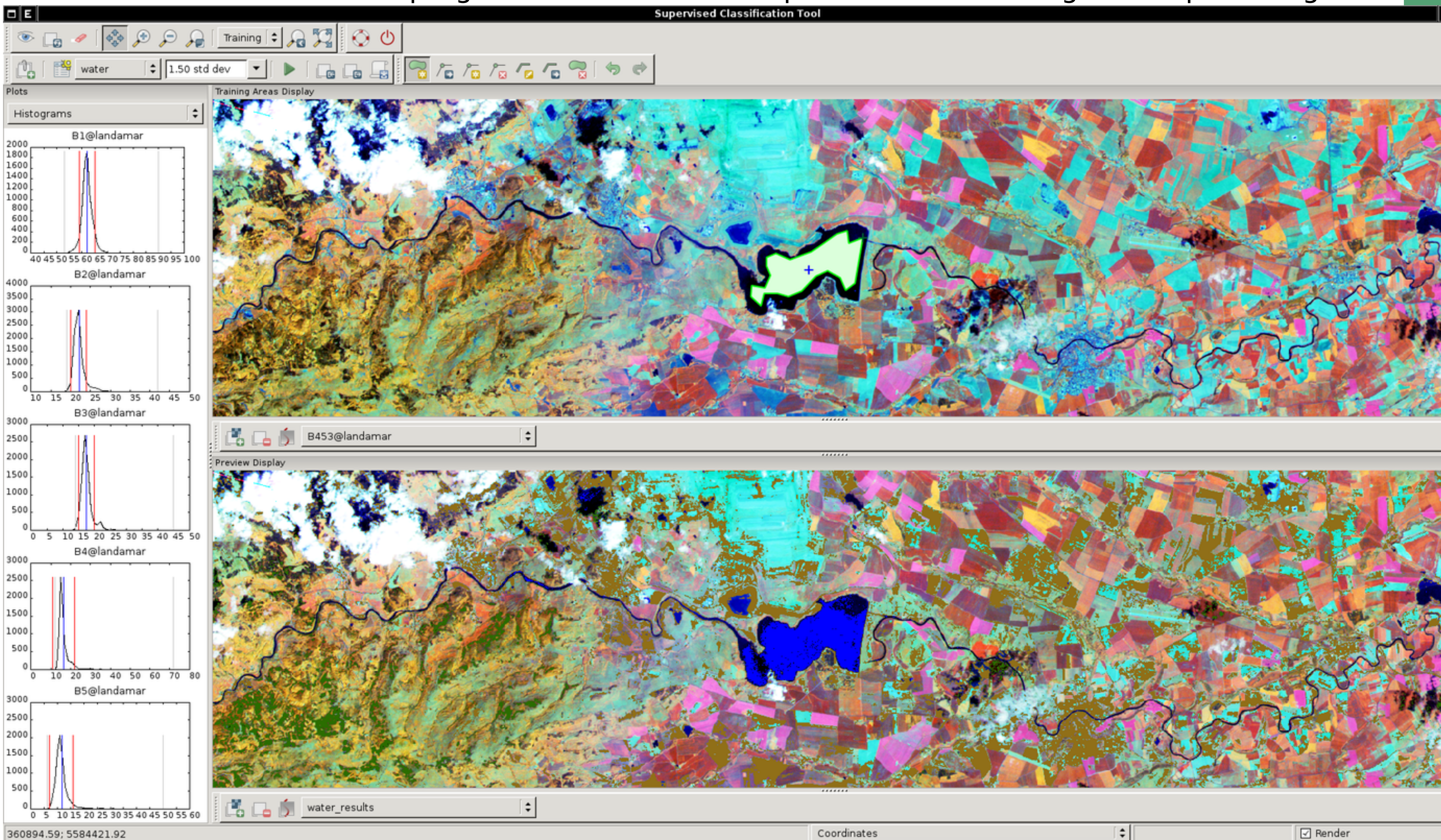
GRASS GIS 7: New Geospatial Modeller



Extra bonus:
Export it as a Python script

GRASS GIS 7: Supervised image classification

http://geo.fsv.cvut.cz/~landa/publications/2012/ogrs2012/poster/figures/

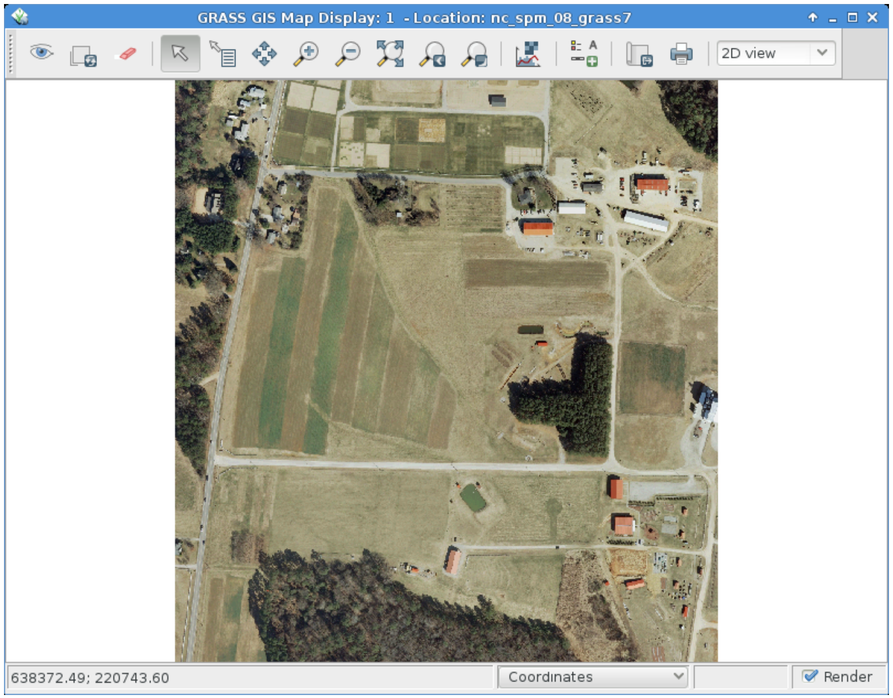


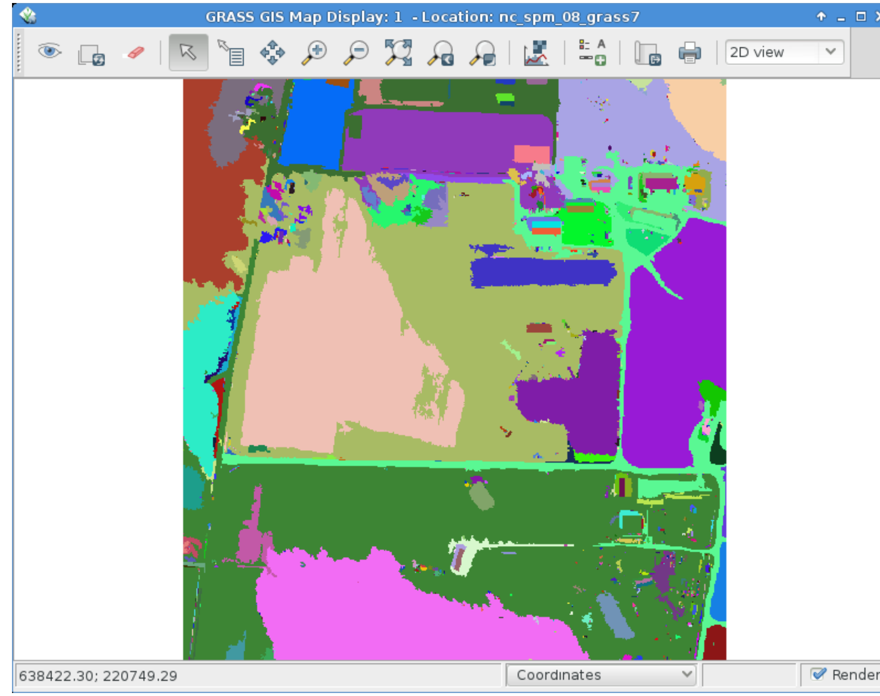
Tool for supervised classification of imagery data.

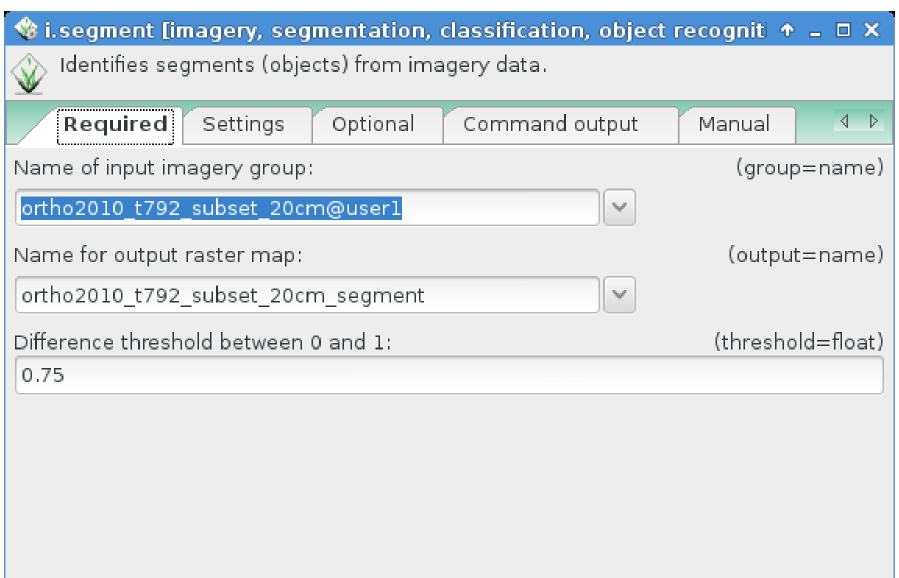
Generates spectral signatures for an image by allowing the user to outline regions of interest.

GRASS GIS 7: Unsupervised image classification

i.segment - Identifies segments (objects) from imagery data.

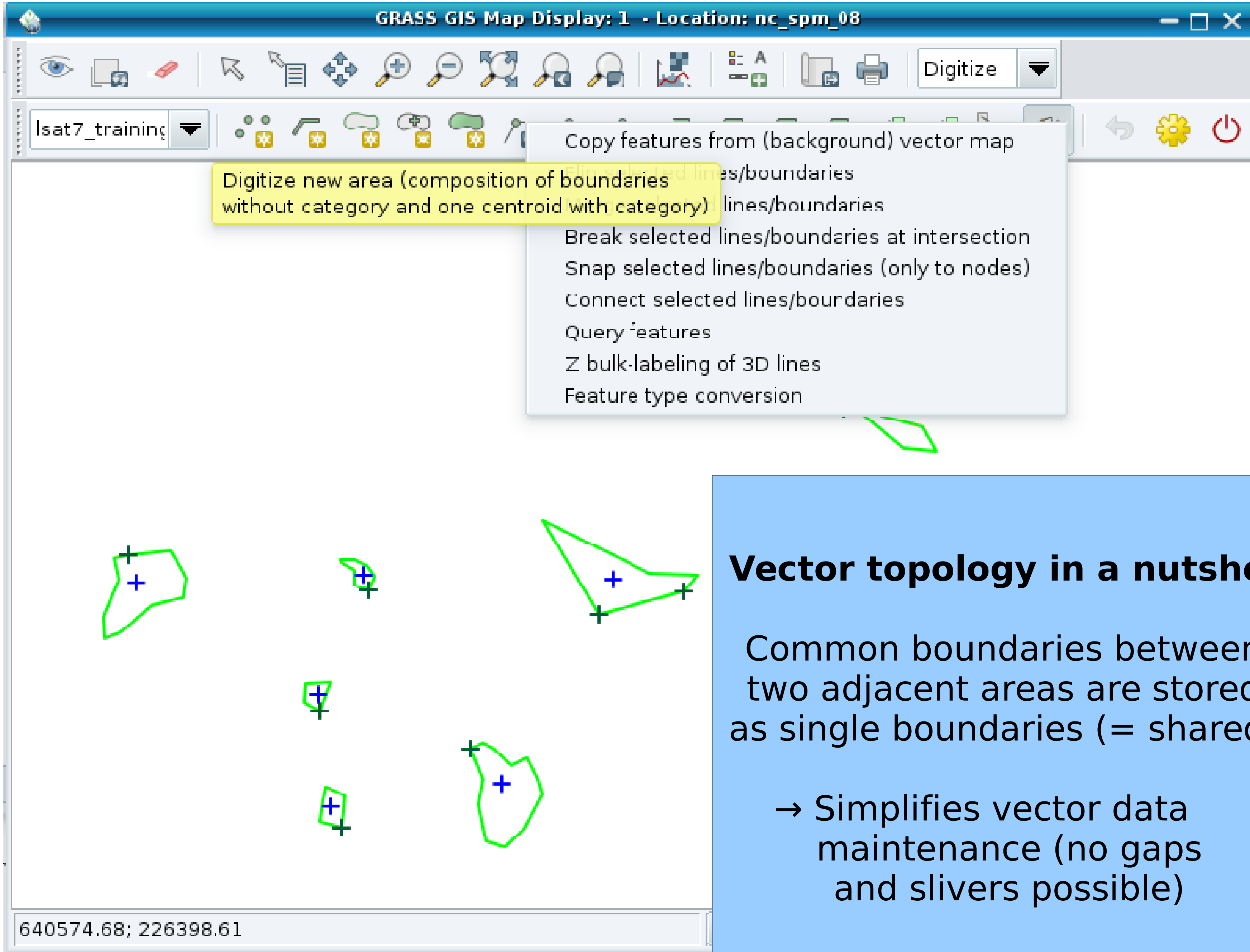






Vector data processing

GRASS GIS 7: Topological Vector Digitizer



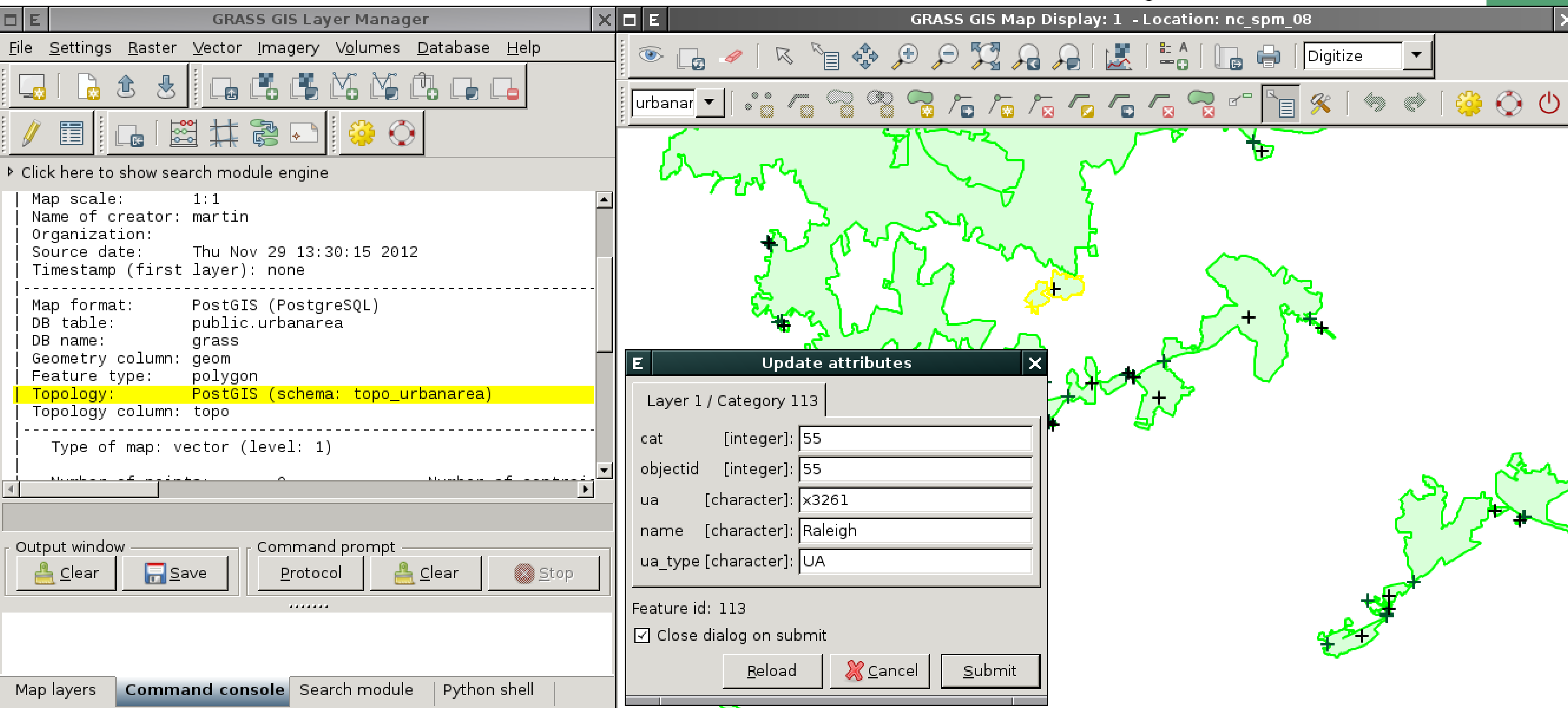
Vector topology in a nutshell

Common boundaries between two adjacent areas are stored as single boundaries (= shared).

→ Simplifies vector data maintenance (no gaps and slivers possible)

GRASS GIS 7: Topological Vector Digitizer in PostGIS 2 (under development)

Programmer: Martin Landa

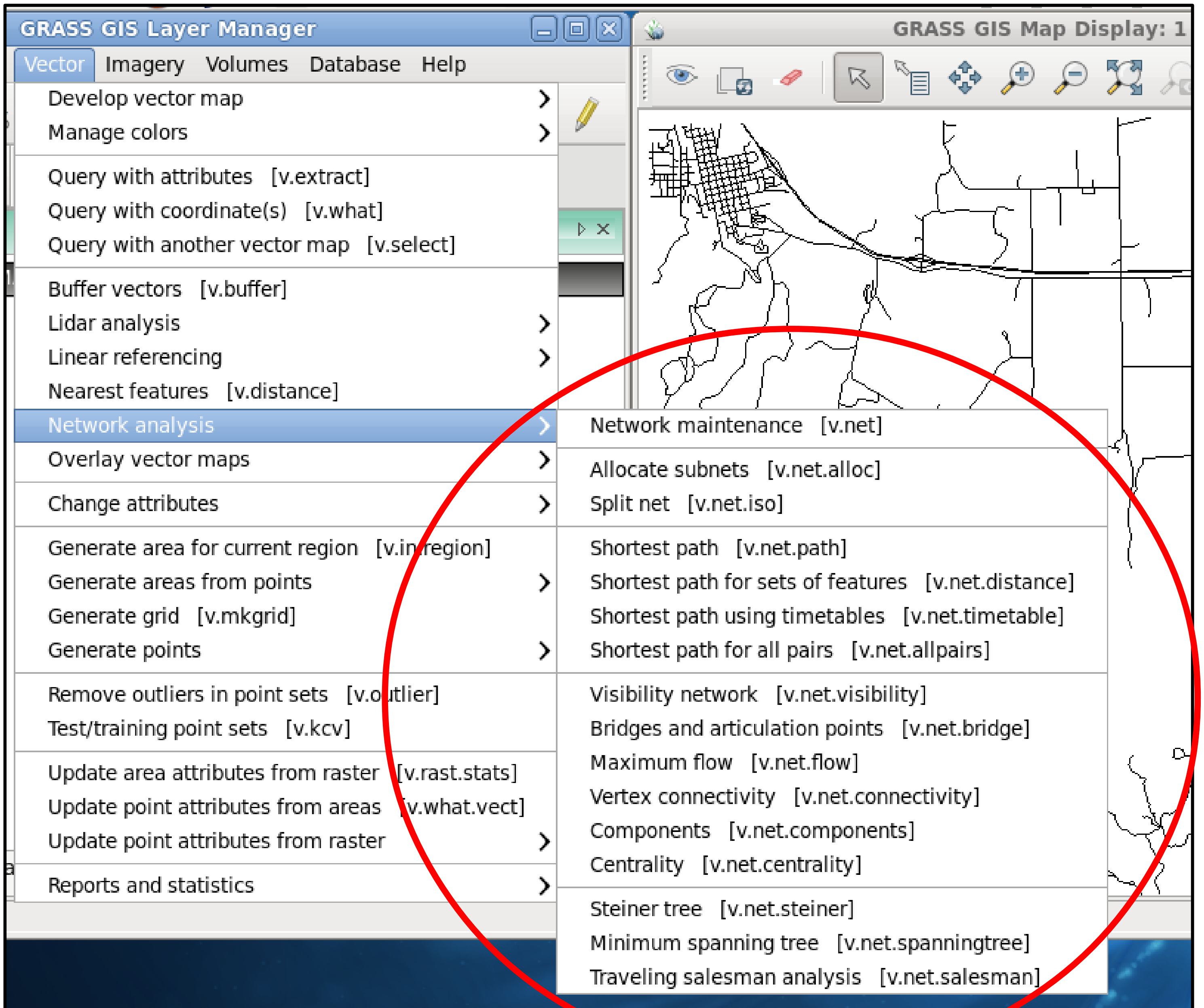


http://grass.osgeo.org/grass70/manuals/v.out.postgis.html

http://grasswiki.osgeo.org/wiki/PostGIS_Topology

Cofunded by Municipality of Trento, Italy

Vector network analysis in GRASS GIS 7



Programmer: Stepan Turek

GRASS GIS 7 Temporal Framework: Time-series support

New Space-Time functionality in GRASS GIS 7

Temporal data processing in GRASS GIS

The temporal GIS framework in GRASS introduces three new datatypes that are designed to handle time series data:

- *Space time raster datasets* (strds) are designed to manage raster map time series. Modules that process strds have the naming prefix *t.rast*.
- *Space time 3D raster datasets* (str3ds) are designed to manage 3D raster map time series. Modules that process str3ds have the naming prefix *t.rast3d*.
- *Space time vector datasets* (stvds) are designed to manage vector map time series. Modules that process stvds have the naming prefix *t.vect*.

Temporal data management in general

List of general management modules:

- t.connect
- t.create
- t.remove
- t.register
- t.unregister
- t.info
- t.list
- t.rast3d.list
- t.vect.list
- t.vect.db.select
- t.sample
- t.support
- t.topology

Export/import conversion

- t.rast.export
- t.rast.import
- t.rast.out.vtk
- t.rast.to.rast3
- r3.out.netcdf
- t.vect.export

Statistics and gap filling

- t.rast.gapfill
- t.rast.univar

Querying and map calculation

- t.rast.list
- t.rast.extract
- t.rast.gapfill
- t.rast.mapcalc
- t.rast3d.extract
- t.rast3d.mapcalc
- t.rast3d.univar
- t.vect.extract
- t.vect.import
- t.vect.observe.strds
- t.vect.univar
- t.vect.what.strds

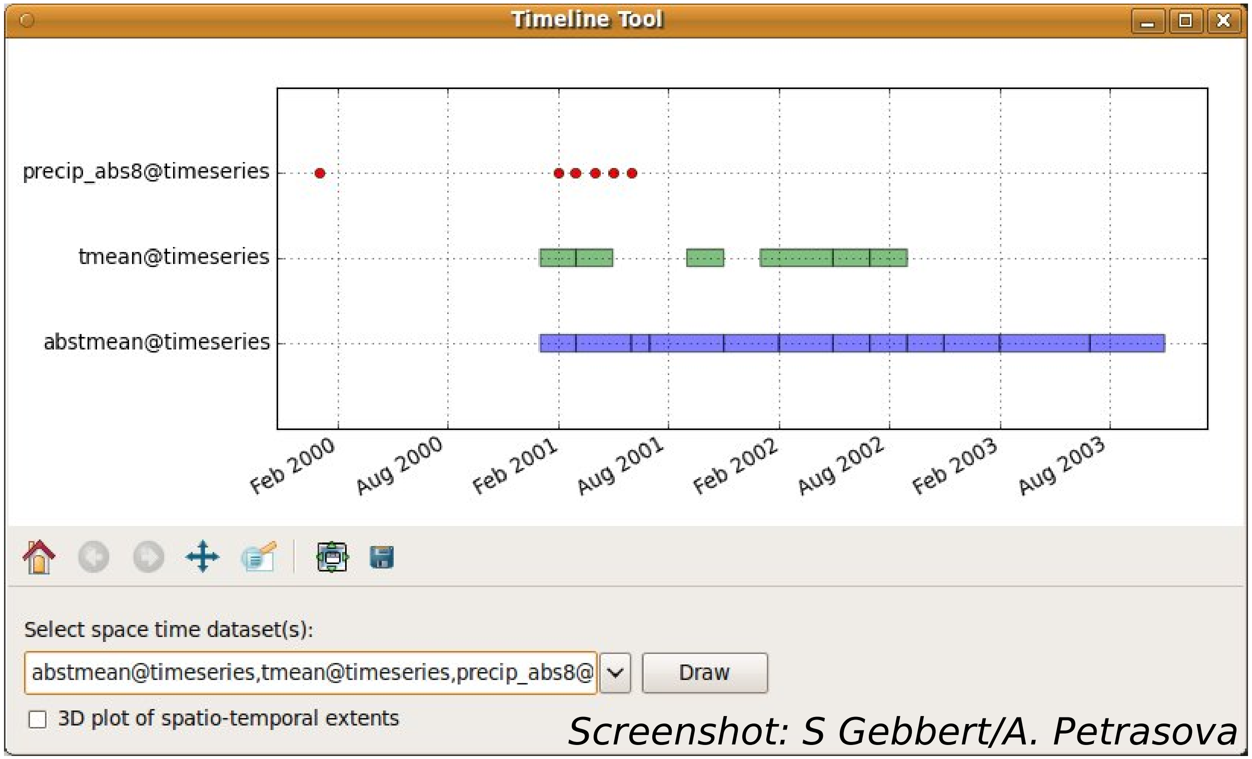
Aggregation

- t.rast.aggregate.ds
- t.rast.aggregate
- t.rast.series

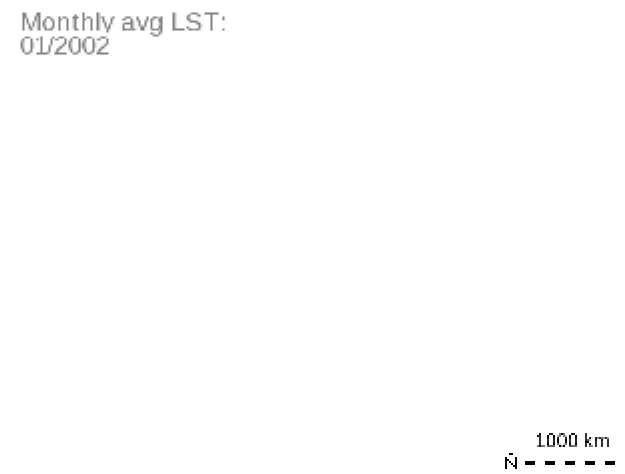
Space time datasets are stored in a temporal database. SQLite3 or PostgreSQL are supported as SQL database back end. Connection settings are performed with t.connect. As default a sqlite3 database will be created in the PERMANENT mapset that stores all space time datasets and registered time series maps from all mapsets in the location.

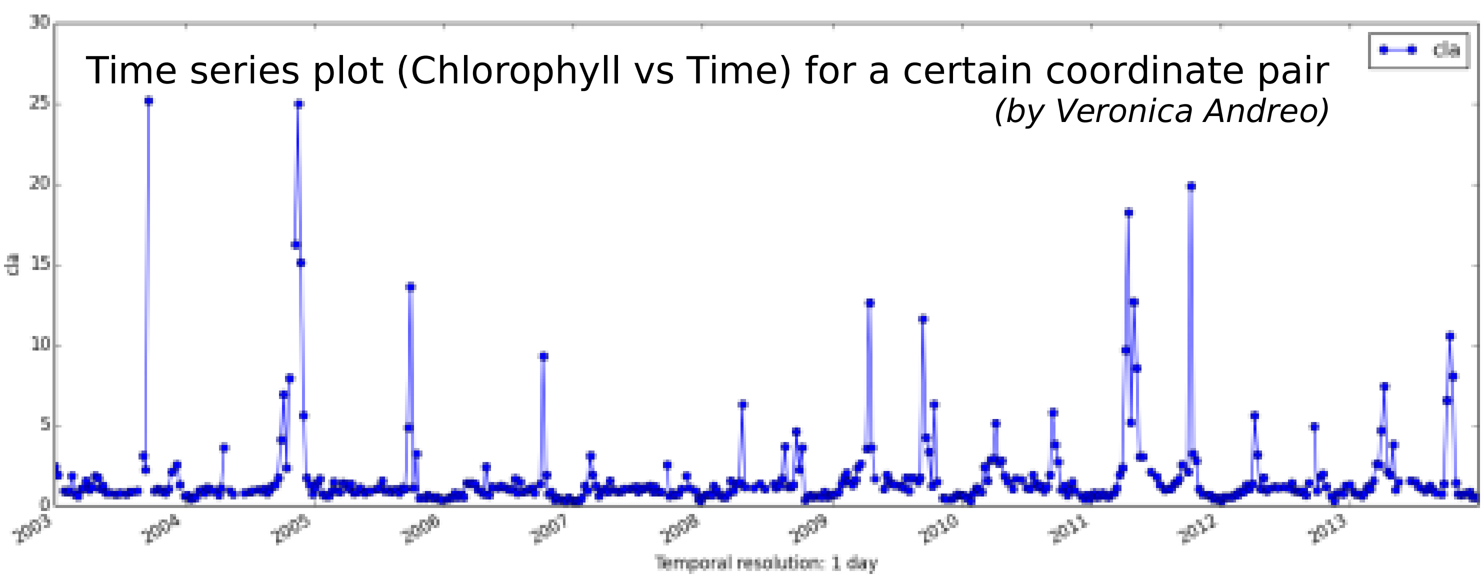
Gebbert, S., Pebesma, E., 2014. *TGRASS: A temporal GIS for field based environmental modeling*. Environmental Modelling & Software 53, 1-12. (DOI)

New Space-Time functionality in GRASS GIS 7



Screenshot: S Gebbert/A. Petrasova





Time series plot (Chlorophyll vs Time) for a certain coordinate pair
(by Veronica Andreo)

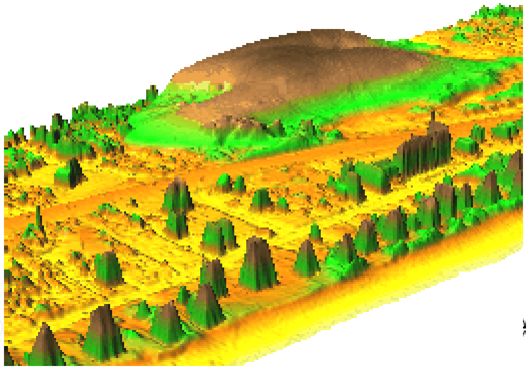
Visualization

GRASS 7: New animation tool for time series

The **Animation Tool** is a *wxGUI* component for animating a series of GRASS raster maps or a space time raster dataset (created by t.* modules).

Animation Tool allows you to:

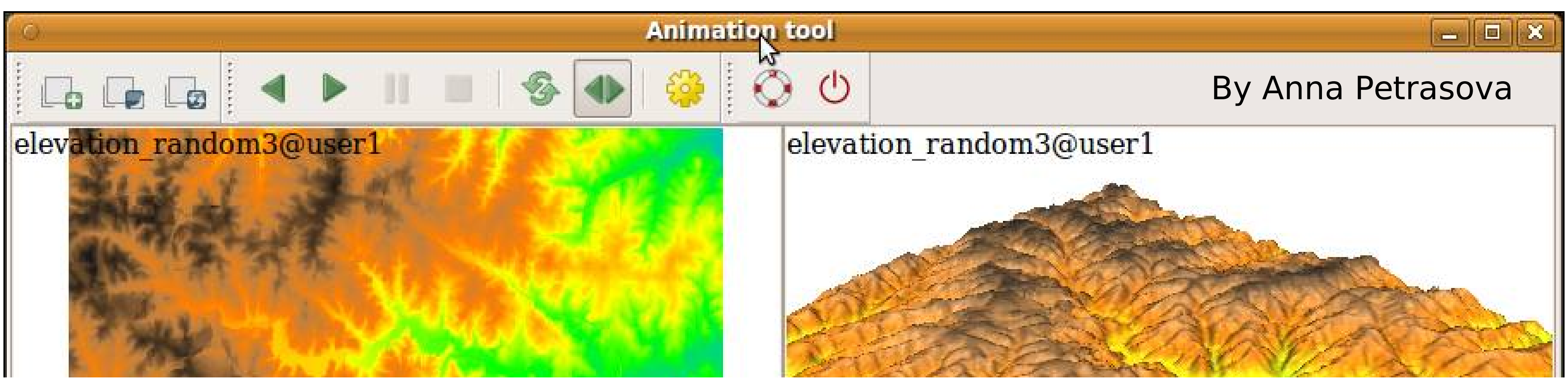
- display up to 4 synchronized animations
- control the animation speed
- interactively change active frame using a slider
- visualize space time datasets with unequally spaced intervals
- animate 3d view (partially implemented)



Nagshead LiDAR time series: dune moving over 9 years (NC, USA)

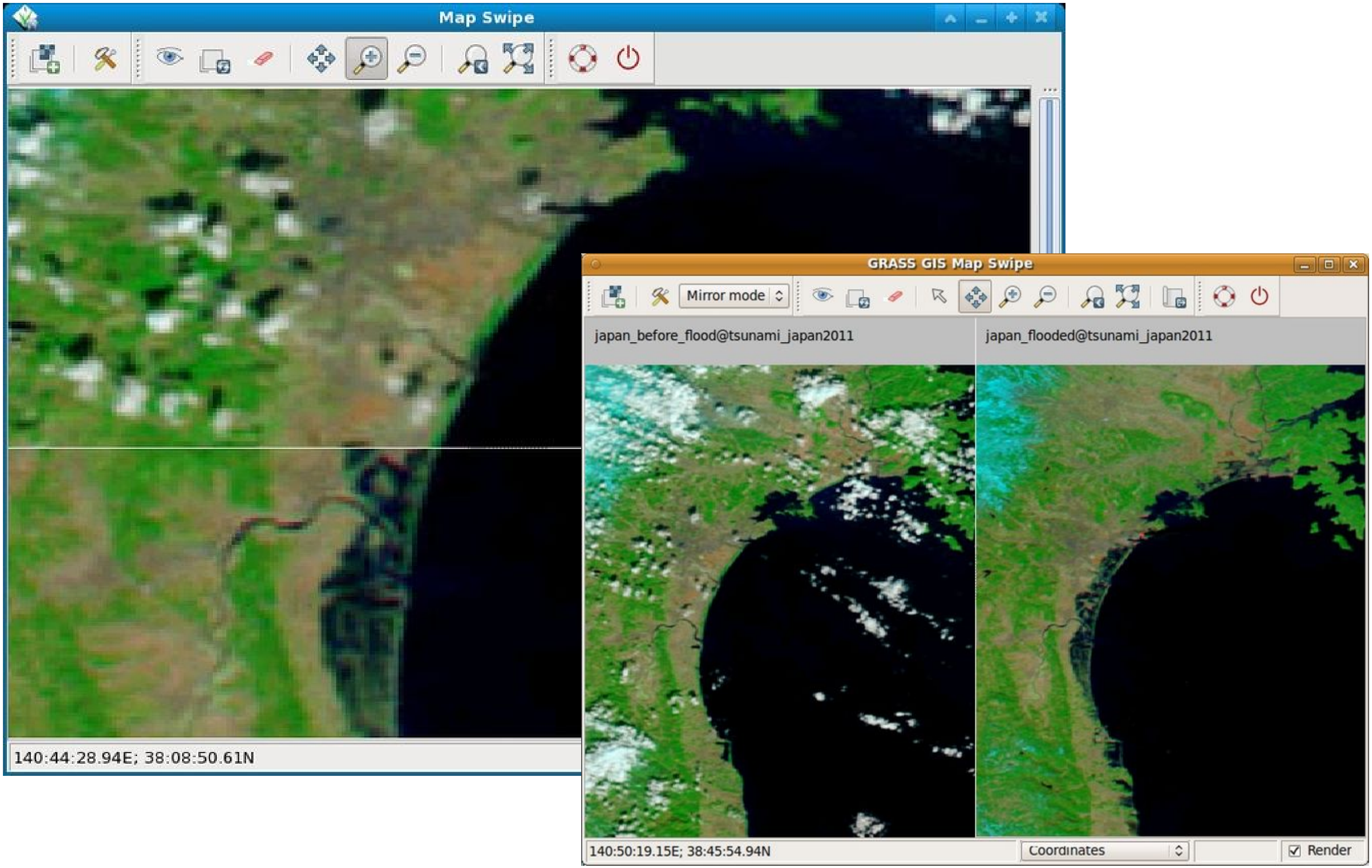
3D view animation enables to animate raster (as an elevation map or a color map) or vecto
m.nviz.image is used. To display 3D view animation follow these steps:

- open GRASS GUI, load maps and start 3D view
- set view, light and other parameters as you like
- save workspace file
- add new animation in Animation Tool, choose 3D view mode
- choose data (series of maps or space time dataset) used for animation
- set workspace file
- choose parameter (parameter of m.nviz.image) to animate (e.g. color_map)



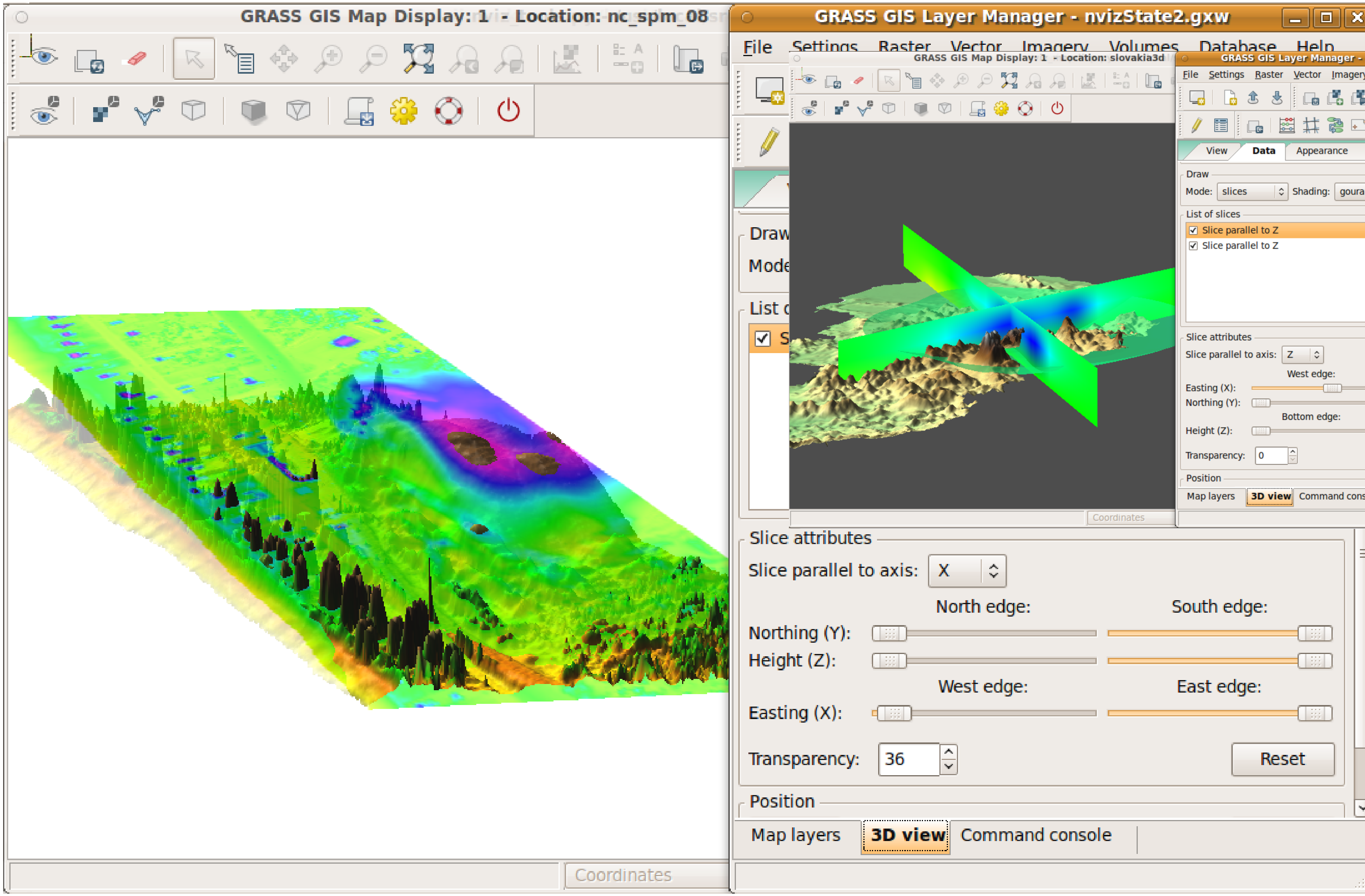
http://grass.osgeo.org/grass70/manuals/g.gui.animation.html

New Map swiping tool for multitemporal maps



Pre and post disaster images of the tsunami in Japan in 2011
(MODIS images taken on February 26 and March 13, 2011)

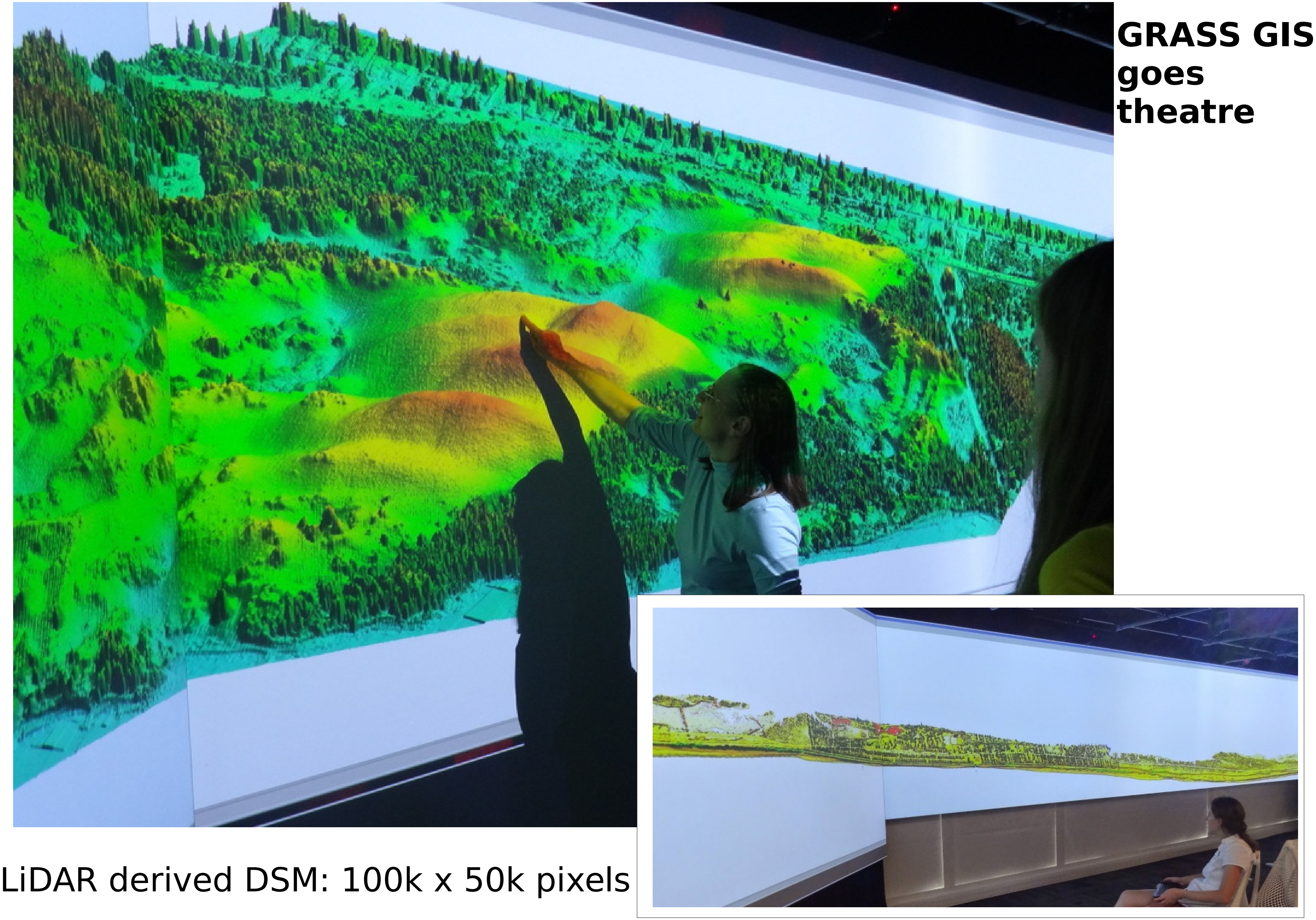
GRASS 7: New visualization tool: wxNVIZ



Programming/screenshot:
Anna Petrasova

http://grasswiki.osgeo.org/wiki/WxNVIZ

New vizualization methods(NC state university)



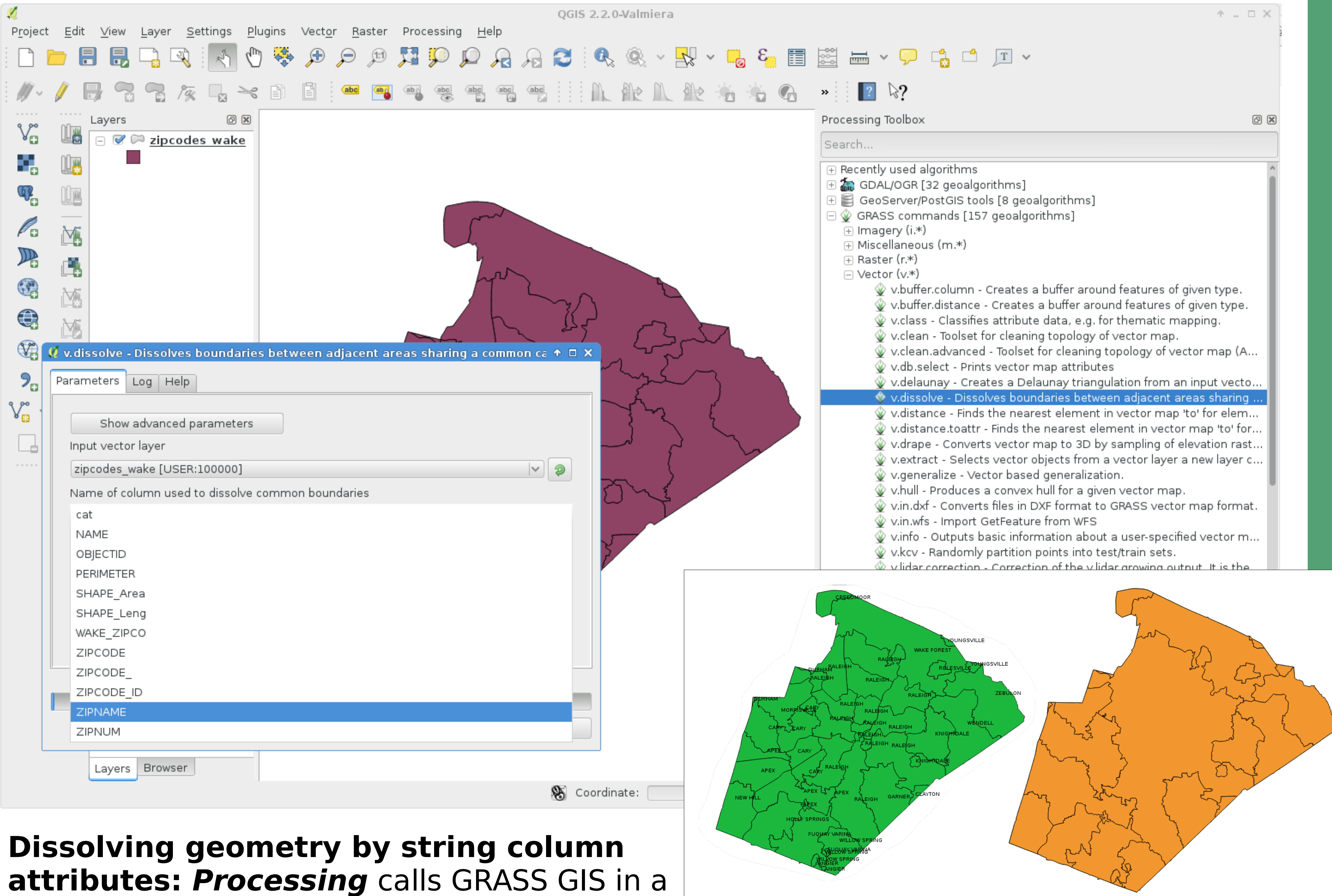
GRASS GIS goes theatre

LiDAR derived DSM: 100k x 50k pixels

GRASS GIS as Open Source GIS backbone:

Connecting to other software packages

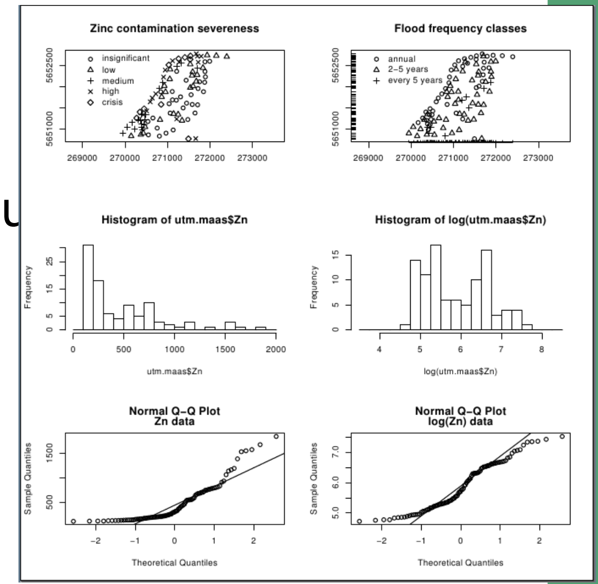
GRASS GIS 7 and QGIS Integration: Processing



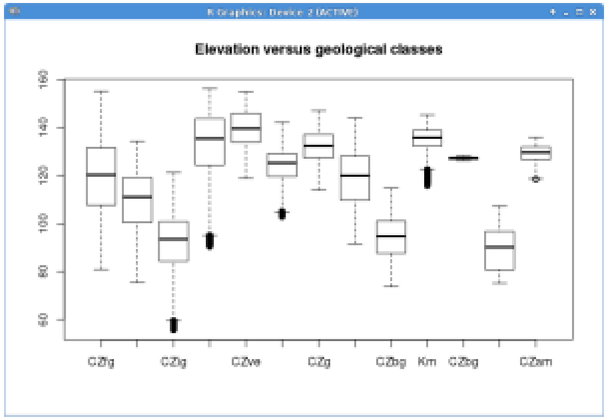
Dissolving geometry by string column attributes: *Processing* calls GRASS GIS in a virtual session which delivers the result back (here: SHAPE file)

GRASS GIS 7 and R Integration

http://grass.osgeo.org/wiki/R_statistics




```
GRASS 7.0.0svn (nc_spm_08_grass7):~ > g.region raster=elevation -p
GRASS 7.0.0svn (nc_spm_08_grass7):~ > R

R version 3.1.2 (2014-10-31) -- "Pumpkin Helmet"
Copyright (C) 2014 The R Foundation for Statistical Compu
Platform: x86_64-redhat-linux-gnu (64-bit)
[...]

> library(spgrass7)
Loading required package: sp
Loading required package: XML
GRASS GIS interface loaded with GRASS version: GRASS 7.0.0svn (2015)
and location: nc_spm_08_grass7
> myrast <- readRAST(c("geology", "elevation"), cat=c(TRUE, FALSE))
> myvect <- readVECT("roadsmajor")

> str(myvect)
> boxplot(myrast$elevation ~ myrast$geology)
> title("Elevation versus geological classes")
> ...
> writeRAST(myrast, "elev_filt", zcol="elev")
```



GRASS GIS 7: Native OGC WPS Support

http://grasswiki.osgeo.org/wiki/WPS

```
r.grow --wps-process-description
```

Web Processing Service

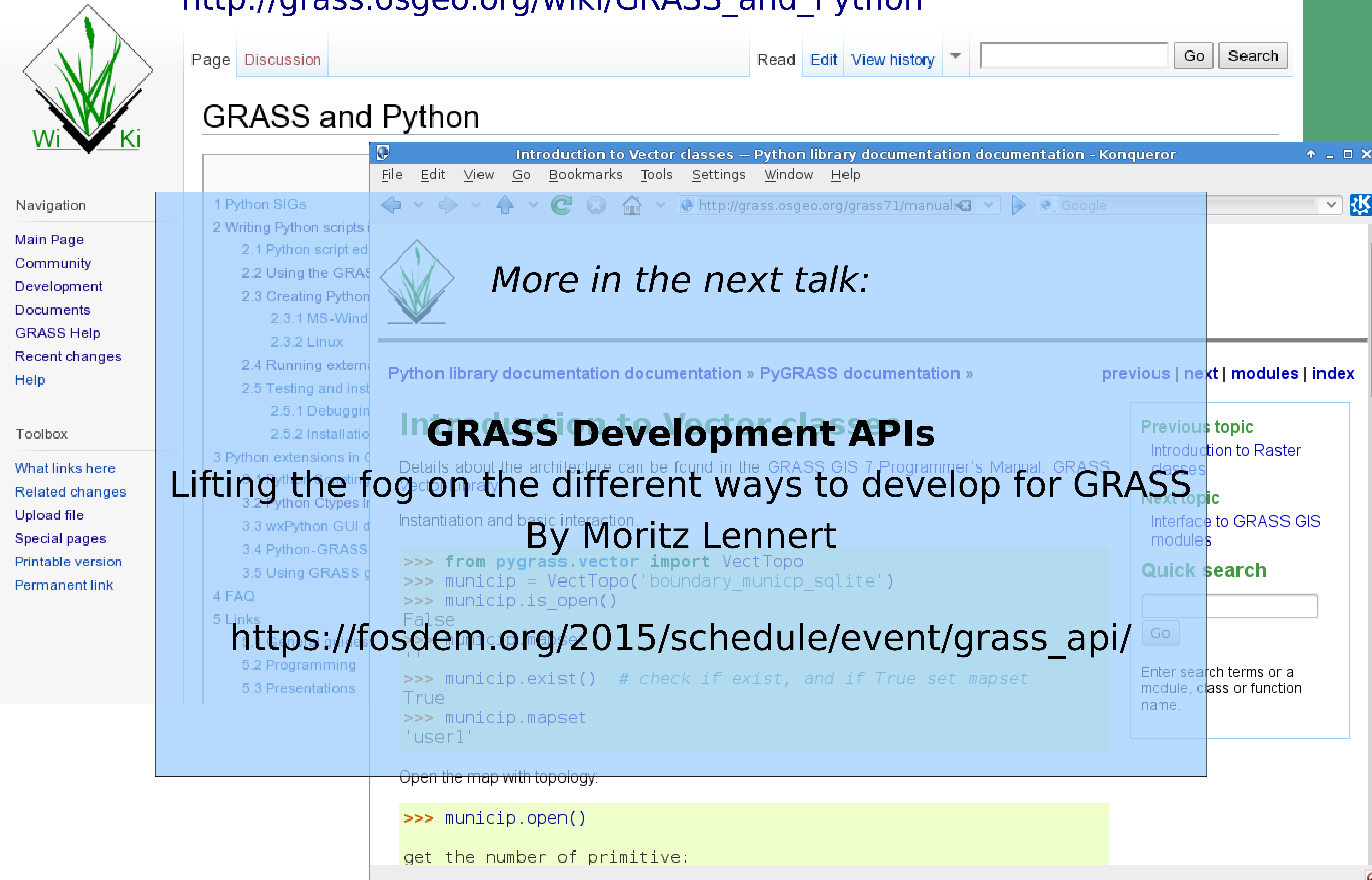

```
<?xml version="1.0" encoding="UTF-8"?>
<wps:ProcessDescriptions xmlns:wps="http://www.opengis.net/wps/1.0.0"
xmlns:ows="http://www.opengis.net/ows/1.1"
xmlns:xlink="http://www.w3.org/1999/xlink"
xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
xsi:schemaLocation="http://www.opengis.net/wps/1.0.0
 http://schemas.opengis.net/wps/1.0.0/wpsDescribeProcess_response.xsd"
 service="WPS" version="1.0.0" xml:lang="en-US">
        <ProcessDescription wps:processVersion="1" storeSupported="true" statusSupported="true">
                <ows:Identifier>r.grow</ows:Identifier>
                <ows:Title>Generates a raster map layer with contiguous areas grown by one cell.</ows:Title>
                <ow
                <ow
                <D
```


```
# register GeoTIFF file in GRASS database:
r.external input=tempmap1.tif output=modis_celsius

# define output directory for files resulting from GRASS calculation:
r.external.out directory=$HOME/gisoutput/ format="GTiff"

# perform GRASS calculation (here: extract pixels > 20 deg C)
# write output directly as GeoTIFF:
r.mapcalc "warm.tif = if(modis_celsius > 20.0, modis_celsius, null() )"

# cease GDAL output connection and turn back to write GRASS raster files:
r.external.out -r

# use the result elsewhere
qgis $HOME/gisoutput/warm.tif
```

Programming own applications with GRASS GIS 7

New GRASS 7 Python API

http://grass.osgeo.org/wiki/GRASS_and_Python



More in the next talk:

GRASS Development APIs

Lifting the fog on the different ways to develop for GRASS

By Moritz Lennert

https://fosdem.org/2015/schedule/event/grass_api/

Support for massive spatial datasets in GRASS GIS

GRASS GIS 7: Support for massive datasets

What is massive?

Massive is relative to

- Hardware resources
- Software capabilities
- Operating system capabilities

Limiting factors

- ϟ RAM
- ϟ Processing time
- ✔ Disk space
- ✔ Largest supported file size

ϟ expensive
✔ cheap
... to solve issue

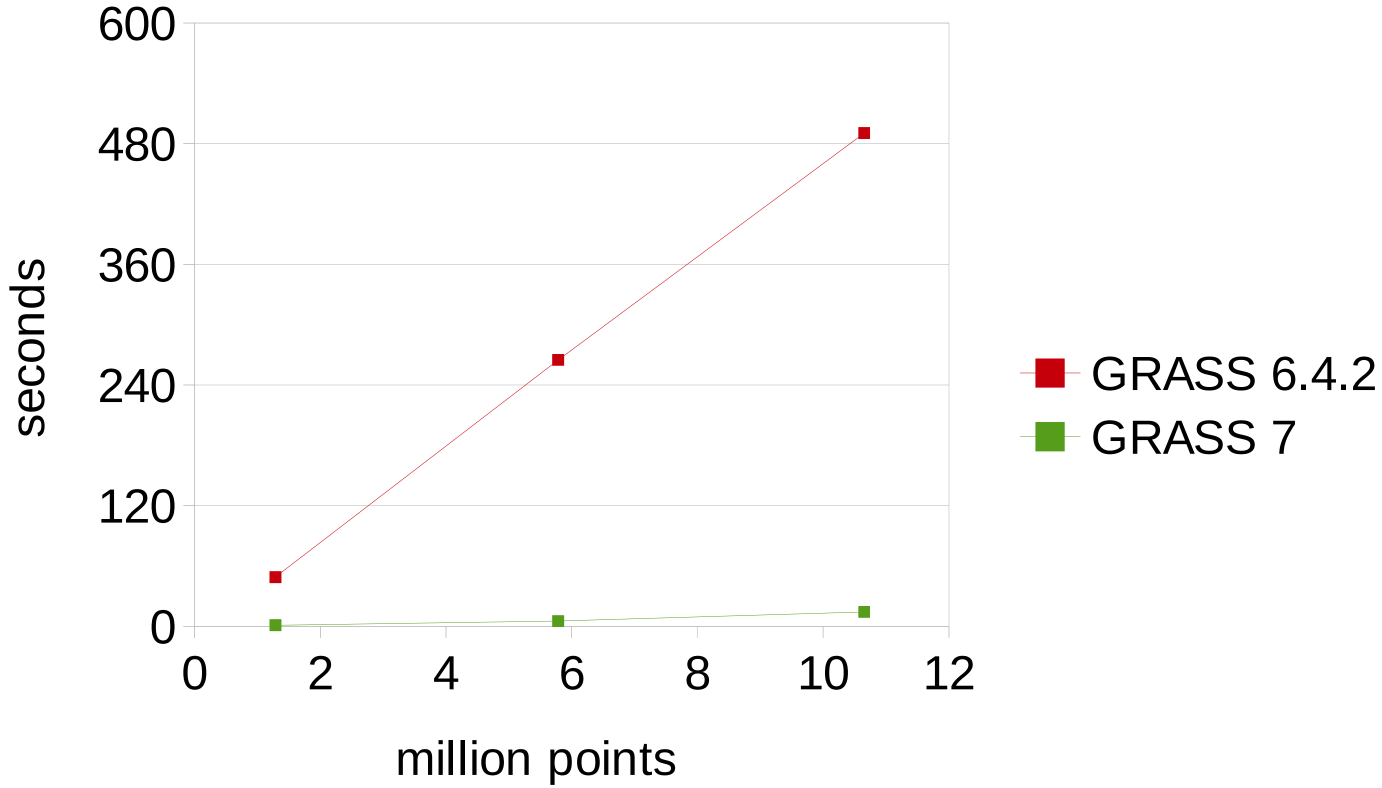
Speed improvements in the vector engine

Spatial query example

Query of vector point maps

GUI: click on vector map, what is there?

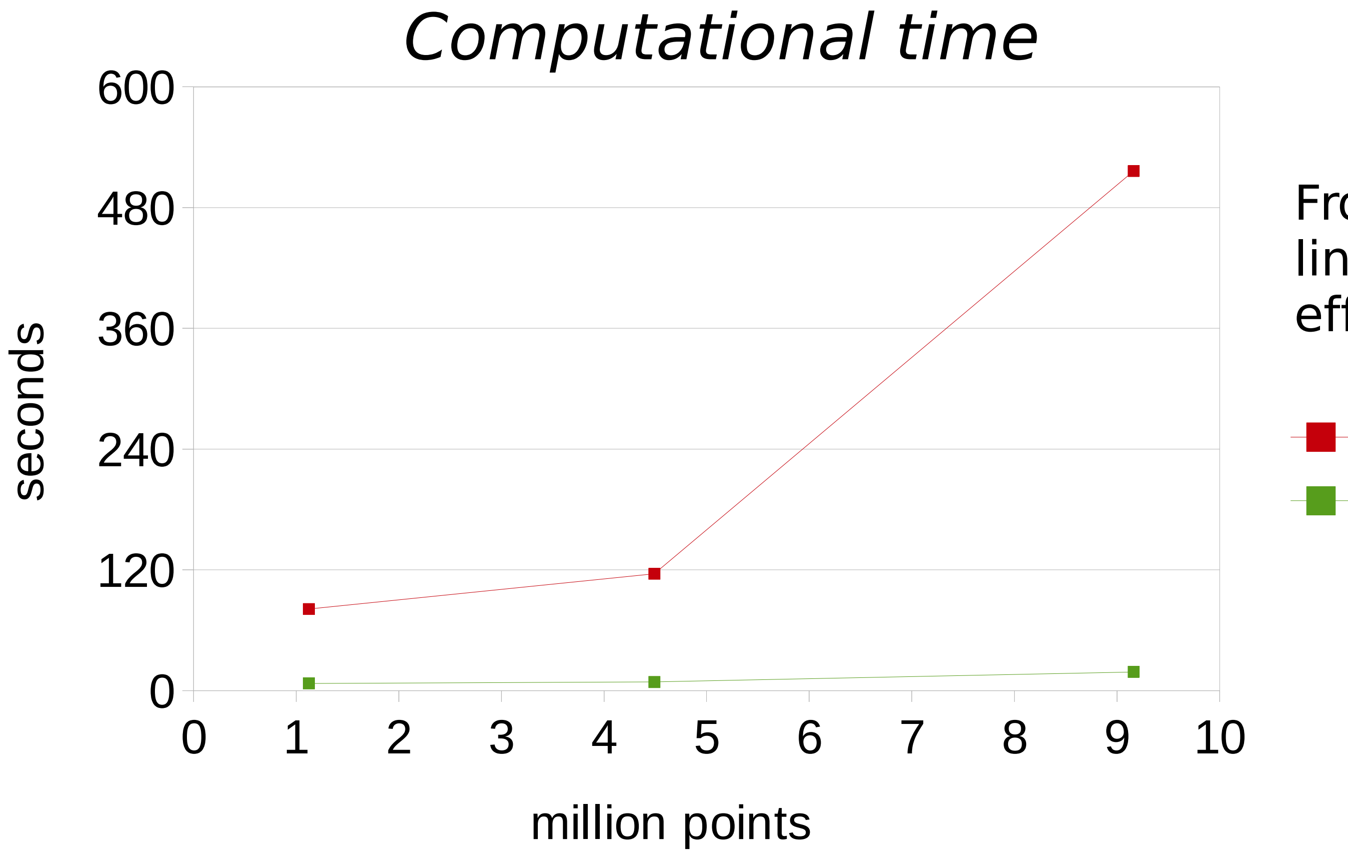
CLI: `v.what east_north=east,north`



Programmer: Markus Metz

GRASS GIS 7: Support for massive datasets

Example cost surfaces: *r.cost*



From non-linear to linear computational effort

Other speed figure:
PCA of 30 million pixels in 6 seconds on this small presentation laptop...

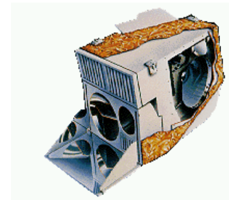
GRASS GIS 7 goes supercomputer

- Since **2005** (10 years) GRASS GIS is running **natively** on 64bit CPUs
- GRASS GIS 7 also offers Large File Support on 32bit Windows

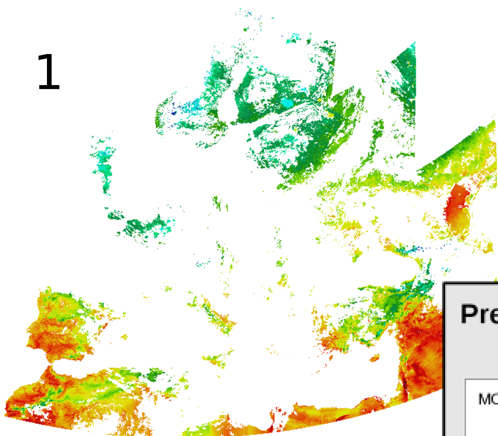
- Installed on Grids and TOP500 supercomputers (AKKA Umeå, ENEA Frascati, Aurel Bratislava, ...)
- Runs on Linux, AIX, Solaris, freeBSD, netBSD, (MS-Windows)...
- Various ways of parallelization

Hints: http://grasswiki.osgeo.org/wiki/Compile_and_Install

EuroLST: MODIS LST daily time series

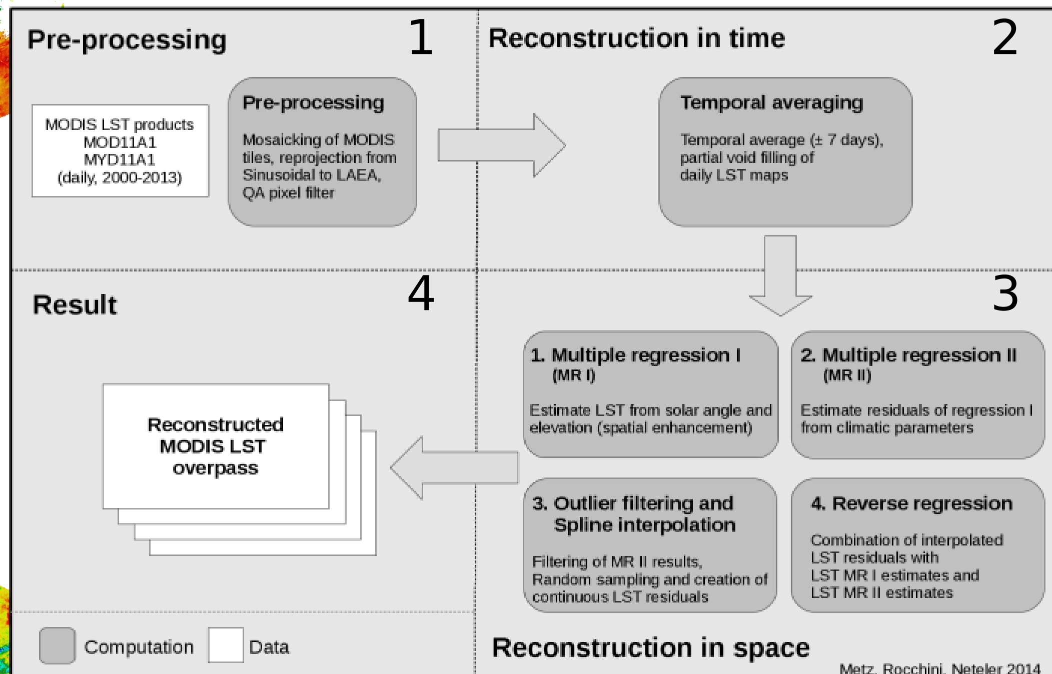


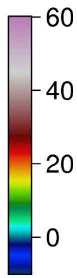
1



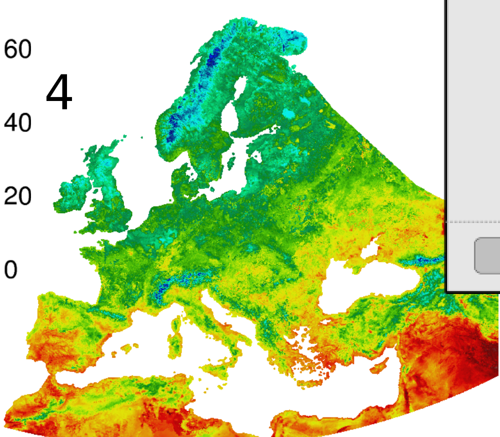
2

Summary workflow of
daily MODIS LST reconstruction at continental scale

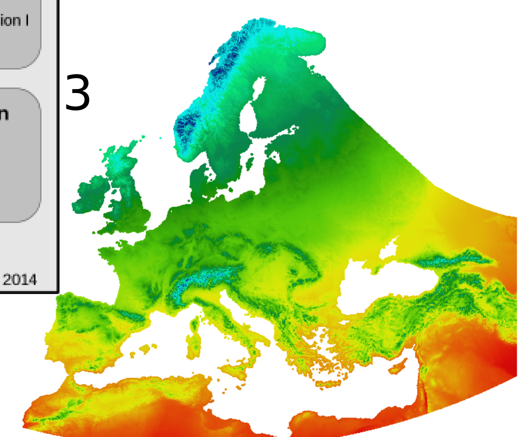




4



3



Data from 2000-today
Each pixel = virtual meteo station

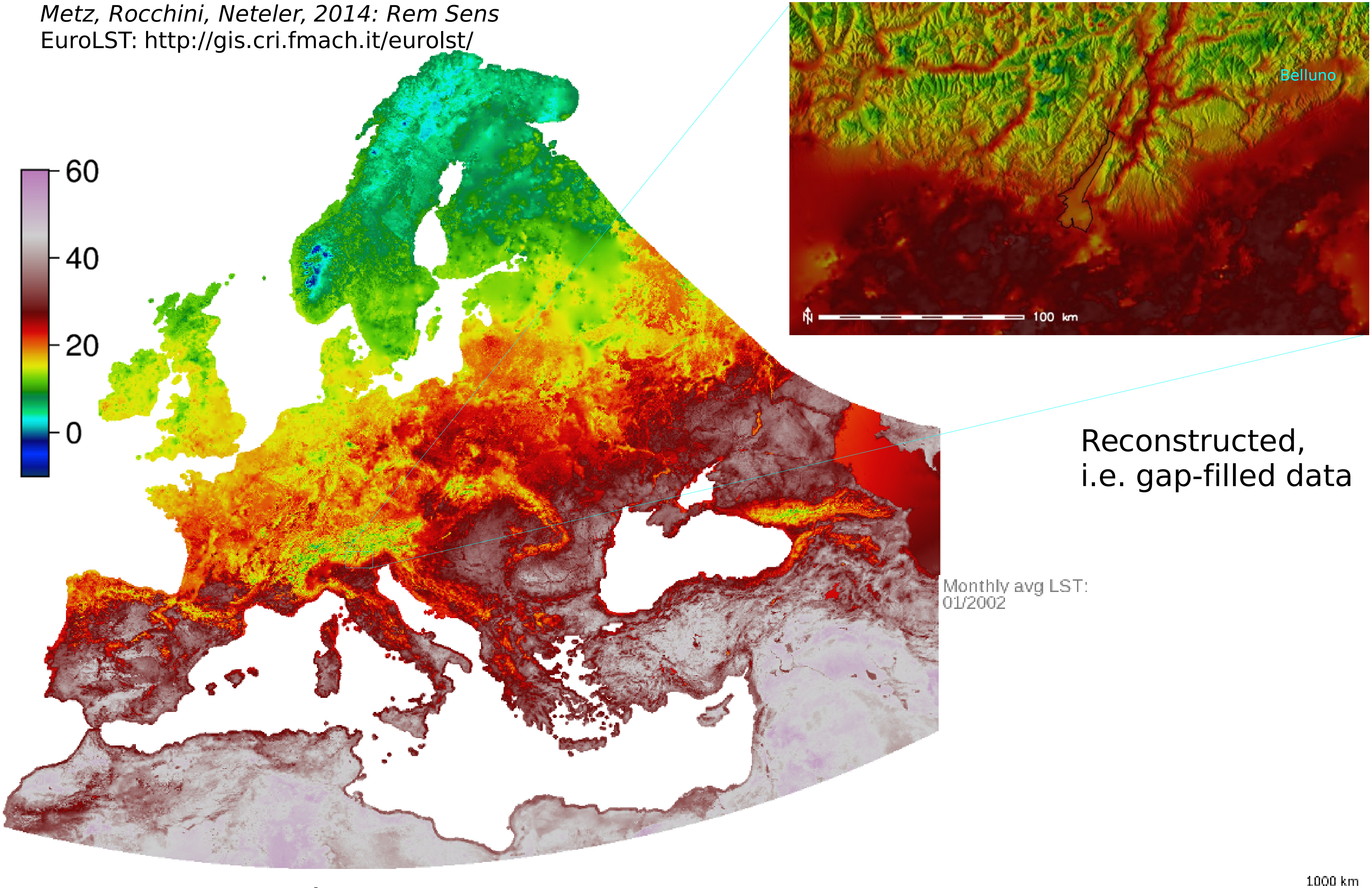
EuroLST: http://gis.cri.fmach.it/eurolst/

Metz, Rocchini, Neteler, 2014: Remote Sens 6, DOI: 10.3390/rs6053822

EuroLST: MODIS LST daily time series

Example: Land surface temperature for Sep 26 2012, 1:30 pm

Metz, Rocchini, Neteler, 2014: Rem Sens
EuroLST: http://gis.cri.fmach.it/eurolst/



Reconstructed,
i.e. gap-filled data

17,000 maps à 415 million pixels

MODIS Land Surface Temperature LST reconstruction

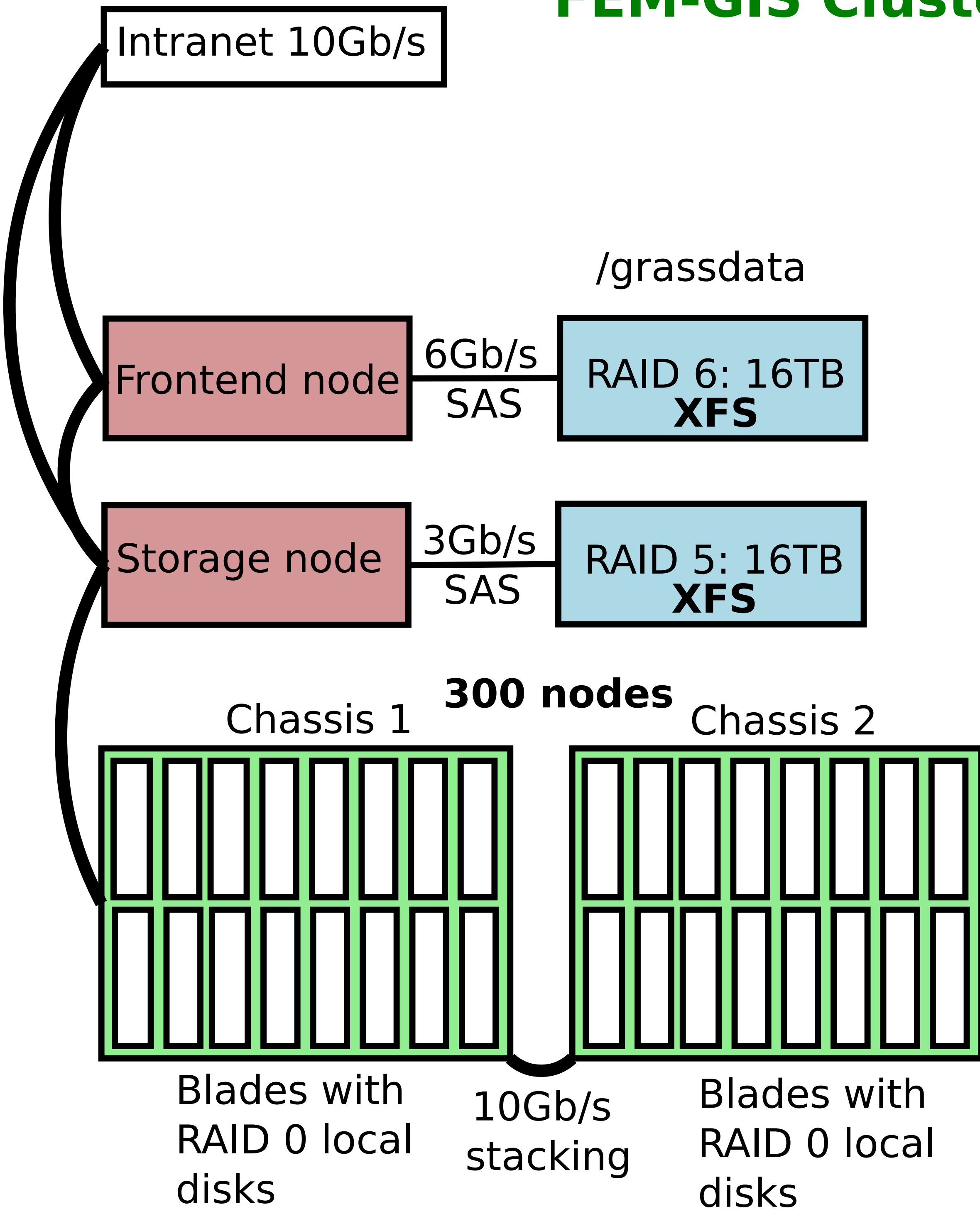
... on a cluster computer

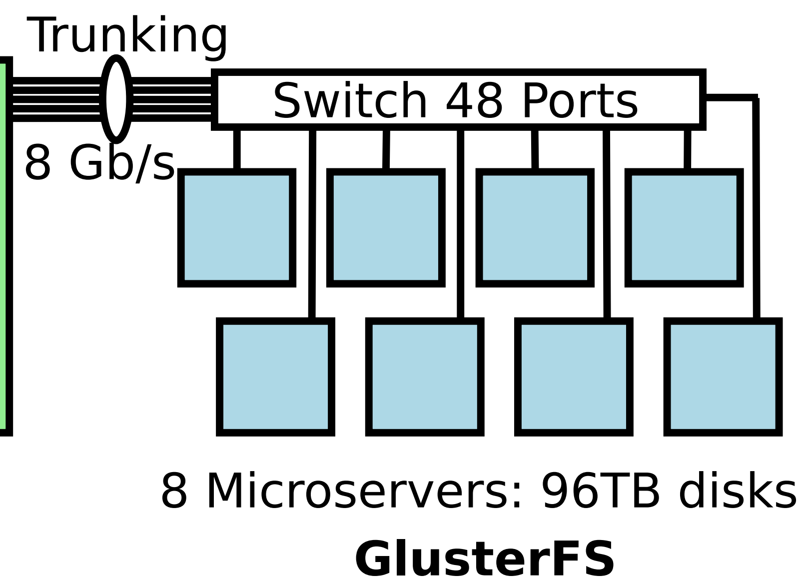
FEM-GIS Cluster

- In total 300 nodes with 600 Gb RAM
- 132 TB raw disk space, XFS, GlusterFS
- Circa 2 Tflops/s
- Scientific Linux operating system, blades headless
- Queue system for job management (Grid Engine), used for GRASS jobs

- Computational time for all data:
  1 month with LST-algorithm V2.0
- Computational time for one LST day:
  3 hours on 2 nodes

FEM-GIS Cluster





Scientific Linux 7

oneSIS diskless blades

NFS used for sharing of partitions

Grid Engine for job management

“Big data” challenges on a cluster

GRASS GIS – LST data processing “evolution”:

- 2008: **internal 10Gb network** connection way to **slow**...

  Solution: TCP jumbo frames enabled (MTU > 8000) to speed up the internal NFS transfer

- 2009: hitting an **ext3 filesystem limitation** (not more than 32k subdirectories but more files in cell_misc/ – each raster maps consists of multiple files)

  Solution: adopting **XFS filesystem** [err, reformat everything]

- 2012: Free inodes on **XFS exceeded**

  Solution: Update XFS version [err, reformat everything again]

- 2013: **I/O saturation** in NFS connection between chassis and blades

  Solution: reduction to one job per blade (queue management), 21 blades * 2.5 billion input pixels + 415 million output pixels

- **GlusterFS saturation**

  Solution: New 48 port switch, 8-channel trunking (= 8 Gb/s)

Where is the stuff?

GRASS GIS 7 Software:

Free download for MS Windows, MacOSX, Linux and source code:

http://grass.osgeo.org/download/

Addons (user contributed extensions):

http://grasswiki.osgeo.org/wiki/GRASS_AddOns

Free sample data:

Rich data set of North Carolina (NC)

… available as GRASS GIS location and in common GIS formats

http://grass.osgeo.org/download/sample-data/

User Help:

Mailing lists (also in different languages):

http://grass.osgeo.org/support/

Wiki:

http://grasswiki.osgeo.org/wiki/

Manuals:

http://grass.osgeo.org/documentation/manuals/

http://grass.osgeo.org

http://trac.osgeo.org/grass/wiki/Grass7/NewFeatures

Markus Neteler

Fondazione E. Mach (FEM)
Centro Ricerca e Innovazione
GIS and Remote Sensing Unit
38010 S. Michele all'Adige (Trento), Italy
http://gis.cri.fmach.it
http://www.osgeo.org

markus.neteler@fmach.it
neteler@osgeo.org

Coming soon:
GRASS GIS 7.0.0!

THANKS

OSGeo
Your Open Source Compass