Markus Neteler

Joint work with
D. Rocchini, L. Delucchi, M. Metz

Massive data processing in GRASS GIS 7: A new gap-filled MODIS Land Surface Temperature time series data set

FOSS4G-Europe 2014, Bremen
July 15 - 17, 2014

```
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
```

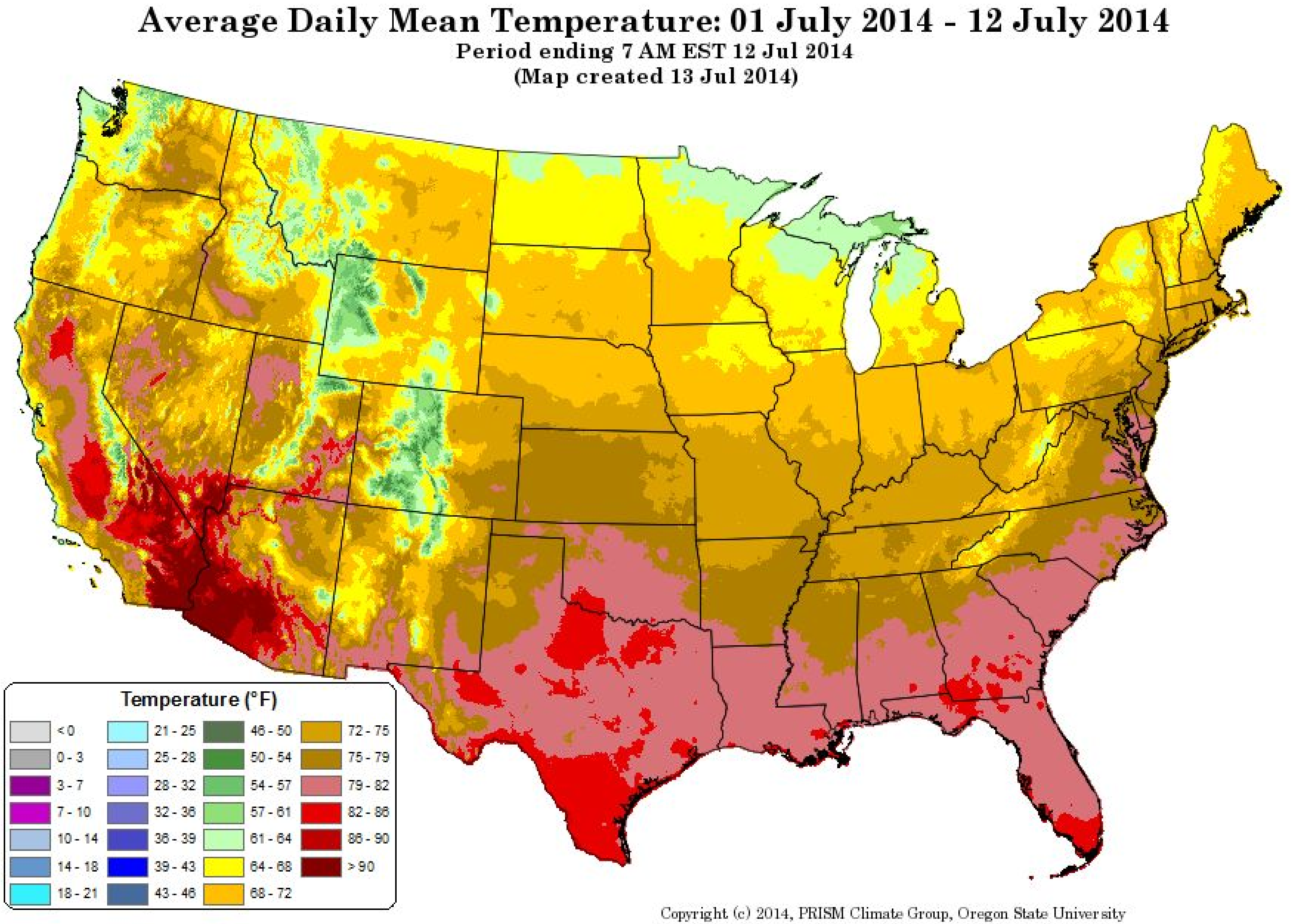
Fondazione Edmund Mach, Trento, Italy

- **Founded 1874** as IASMA - Istituto Agrario San Michele all'Adige (north of Trento, Italy)
- Research Centre + Tech. Transfer Center + highschool, ~ 800 staff
- … of those **350 staff in research** (Environmental research, Agro-Genetic research, Food safety)

http://cri.fmach.eu/

Temperature - available data sets: US

United States: **PRISM data** (30-Year Normals, anomalies, selected monthly data, 800m pixels) http://www.prism.oregonstate.edu/



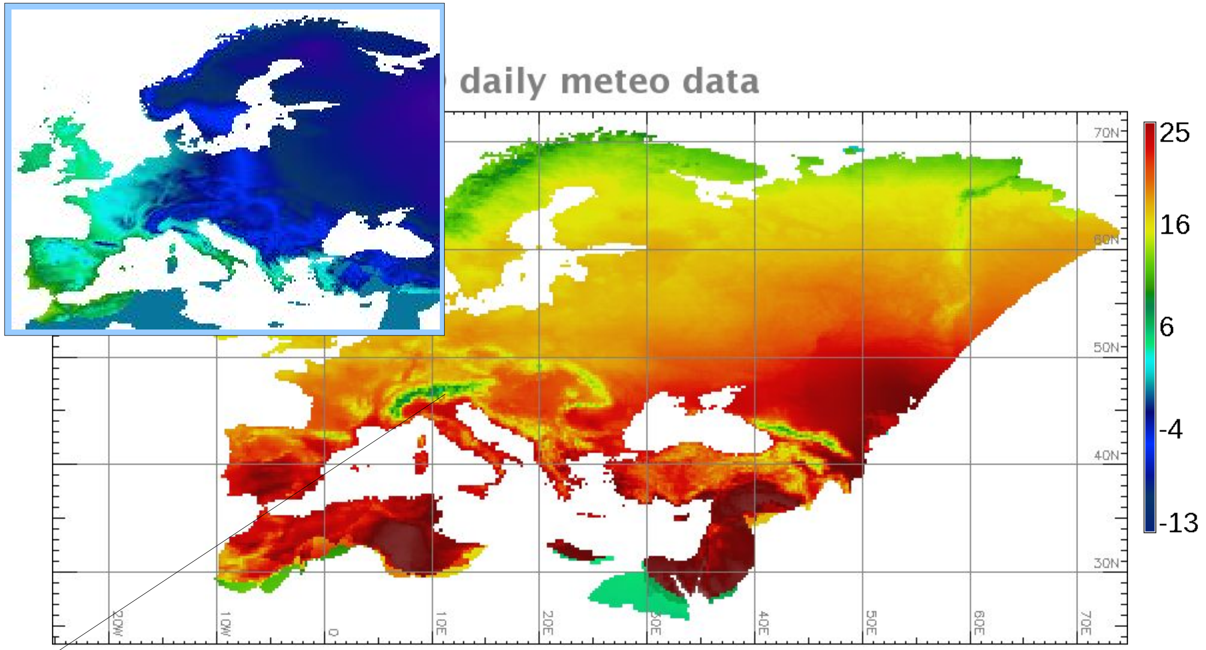
Temperature - available data sets: Europe

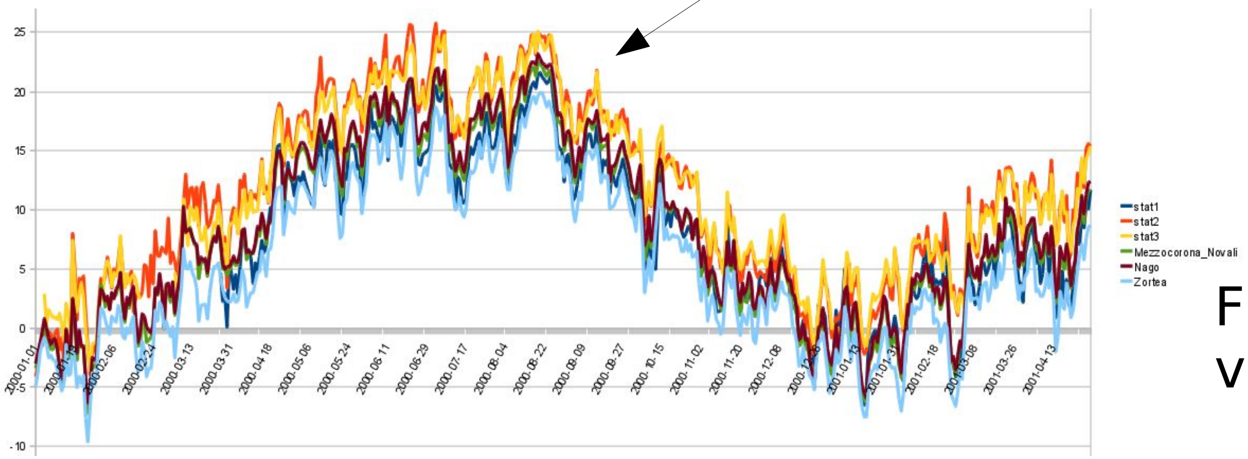
ECA&D - http://www.ecad.eu/
~ 25 km pixels, 60 years, daily

Monthly Tmean: 1950-2013
(derived from daily ECA&D time series)

Climatic variable	Coefficients[§]	Value	Std. Error	t value	Pr (>|t|)
Annual total precipitation	All	−3.814	1.226	−3.112	**
	Pos	−1.477	2.027	−0.729	0.467
	Neg	−2.129	1.600	−1.331	0.185
	Diff.Pos.Neg	0.652	2.568	0.254	0.800
Annual min temperature	All	0.019	0.006	3.424	***
	Pos	0.005	0.010	0.535	0.593
	Neg	0.018	0.006	3.032	**
	Diff.Pos.Neg	−0.013	0.011	−1.134	0.257
Annual max temperature	All	0.034	0.005	7.524	***
	Pos	0.037	0.008	4.448	***
	Neg	0.028	0.005	5.499	***
	Diff.Pos.Neg	0.009	0.009	0.938	0.349

[§]All, slope for all provinces pooling data; Pos, slope for positive provinces; Neg, slope for negative provinces; Diff.Pos.Neg, difference in slopes between positive and negative provinces.
**P≤0.01.
***P≤0.001.
doi:10.1371/journal.pone.0004336.t003



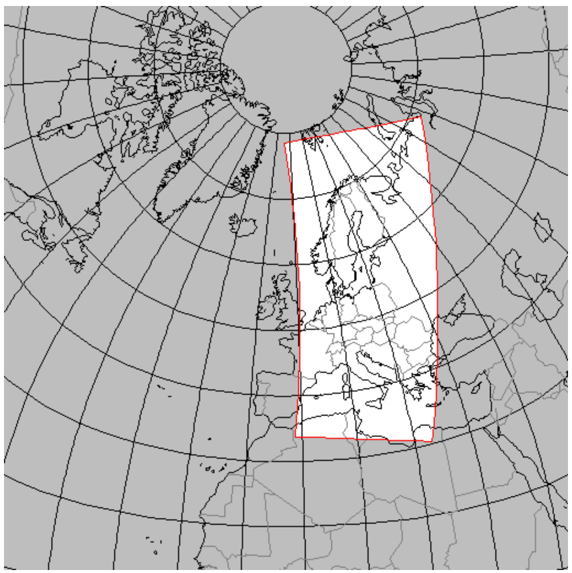


FEM meteo station data
versus ECAD time series

The MODIS Sensor: > 13 years of data

FONDAZIONE EDMUND MACH

The MODIS sensor on board of NASA's Terra and Aqua satellites



Typical MODIS overpass and data coverage (map tiles)

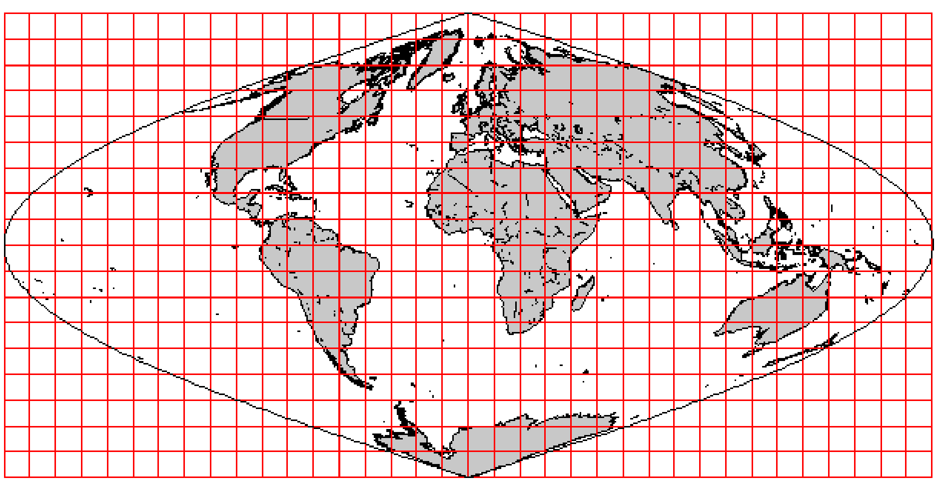
- Sensor with 36 channels in the range of optical light, near and thermal infrared: **Vegetation state, snow, temperature, fire detection ...**
- Delivers data at 250 m, 500 m and 1000 m pixel resolution
- LST error rate: < 1 K ± 0.7 K

MODIS/Terra satellite (EOS-AM):

- startet in Dec. 1999
- overpasses at circa 10:30 + 22:30 solar local time

MODIS/Aqua satellite (EOS-PM):

- startet in May 2002
- overpasses at circa 13:30 + 01:30 solar local time



→ **4 overpasses in 24h**
→ **data availability after ~72h**

EuroLST: MODIS LST daily time series

Efforts of gap-filling the MODIS products MOD11A1 + MYD11A1

Reasons for missing pixels: clouds and aerosols

United States:

Crosson et al. 2012 (Rem Sens Env 119) created a daily merged MODIS LST dataset for conterminous US (1000-m resolution):
16 million grid cells

Europe:

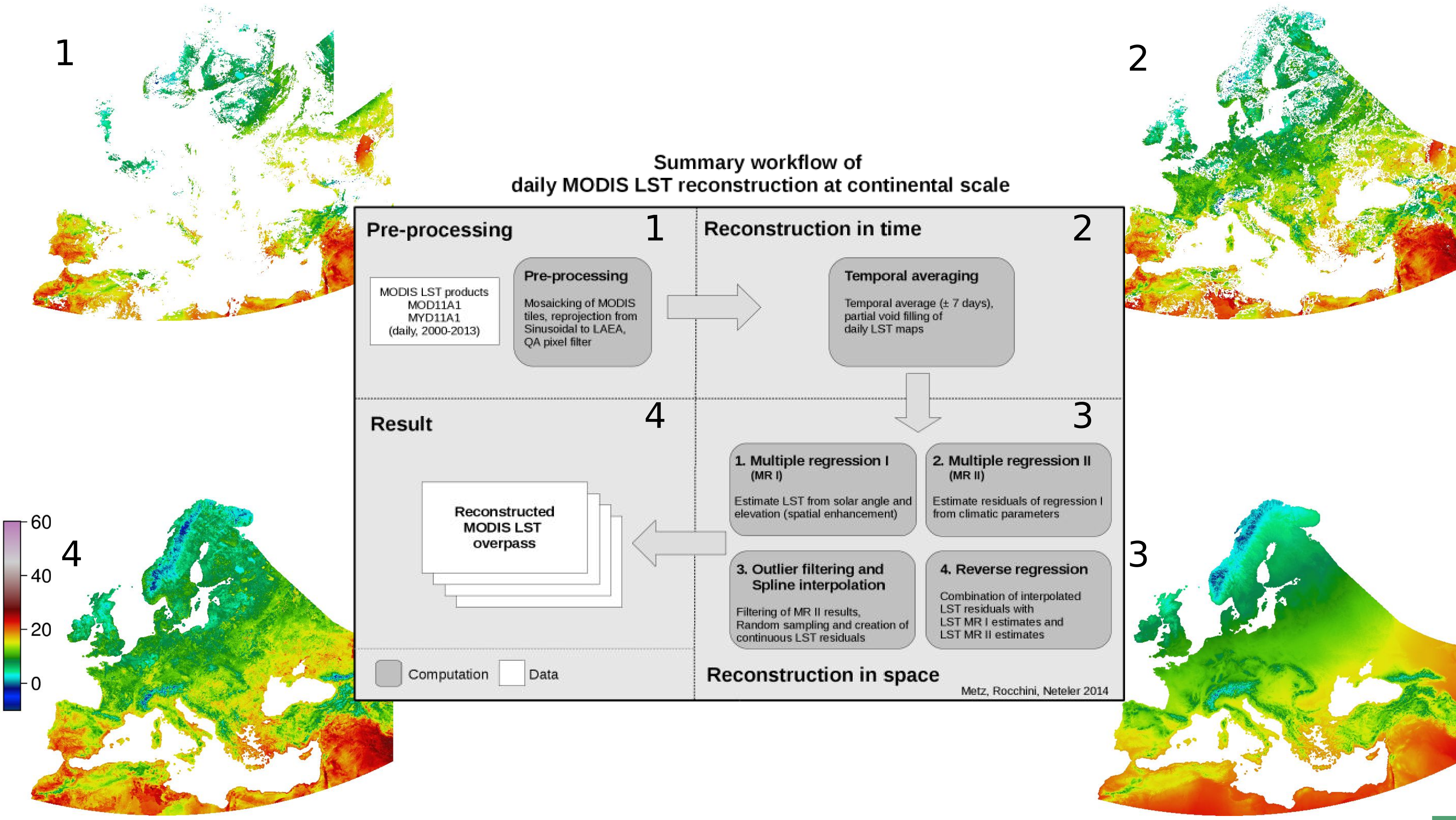
The **new EuroLST** (Metz et al. 2014) covers Europe and Northern Africa, and each overpass (250-m resolution), i.e. 4 maps per day:
415 million grid cells

Processed data:

- PCA and multiple regression of the six input grids (LST, altitude, solar angle, two principal components, ocean mask) with 415 million grid cells each = **2.5 billion pixels per map** reconstruction). Enhancements implemented in GRASS GIS 7.
- In total about 17,000 LST maps processed (each 20 MODIS tiles)

FONDAZIONE EDMUND MACH

EuroLST: MODIS LST daily time series



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

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

EuroLST: MODIS LST daily time series

Software used for LST reconstruction

~~[MODIS Reprojection Tool (MRT 4.1)]~~

(gdalwarp V1.11/VRT is much faster than MRT) [1]

GDAL 1.x

PROJ.4

GRASS GIS 7

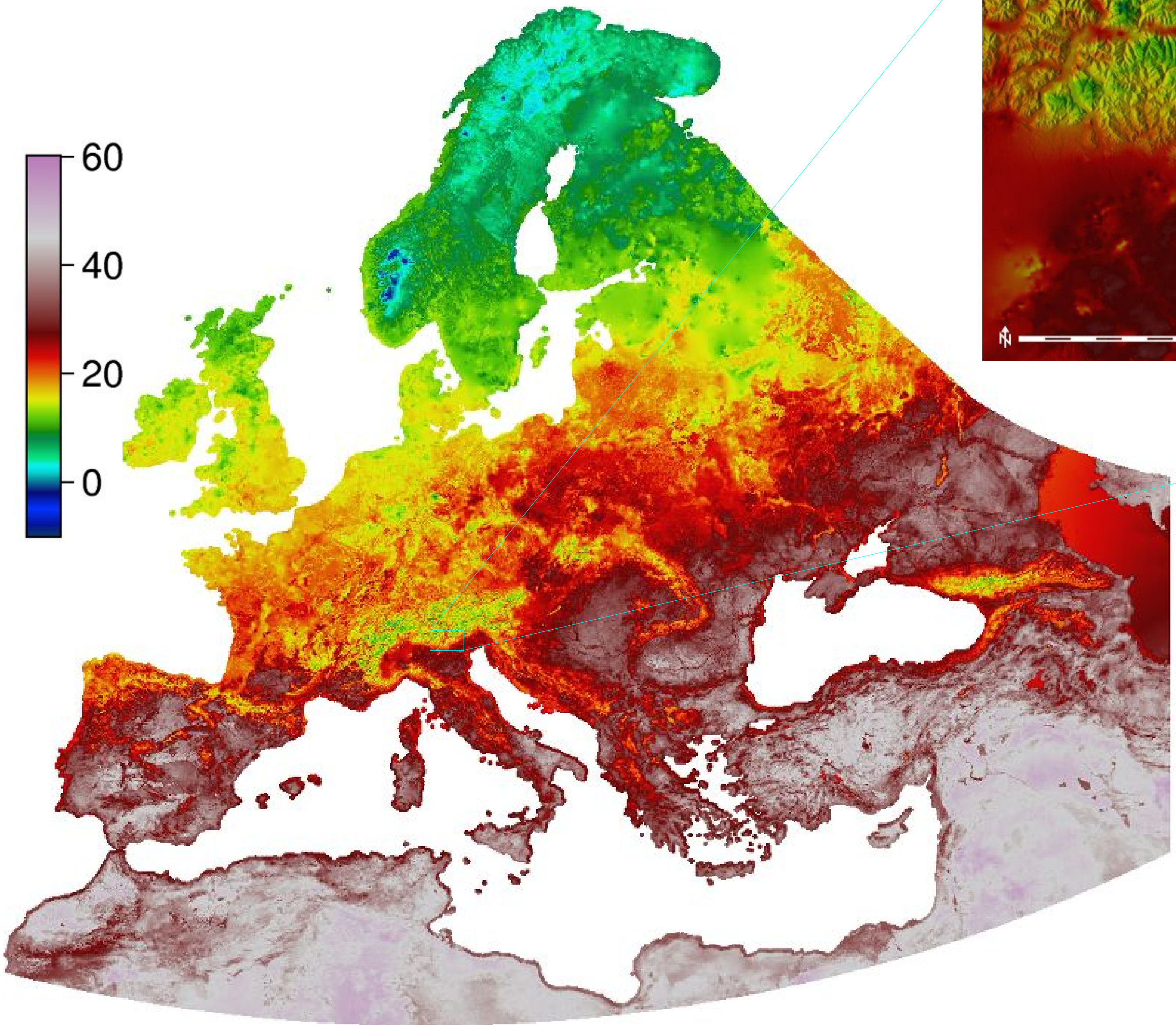
Grid Engine

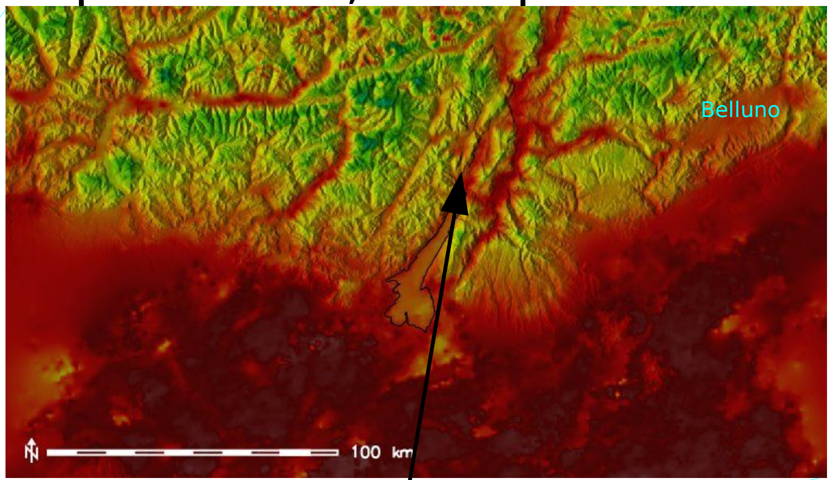
Scientific Linux 6.x

[1] Mosaiking of 20 MODIS LST tiles
** 17,000 overpasses reduced from*
1 week to 1 day

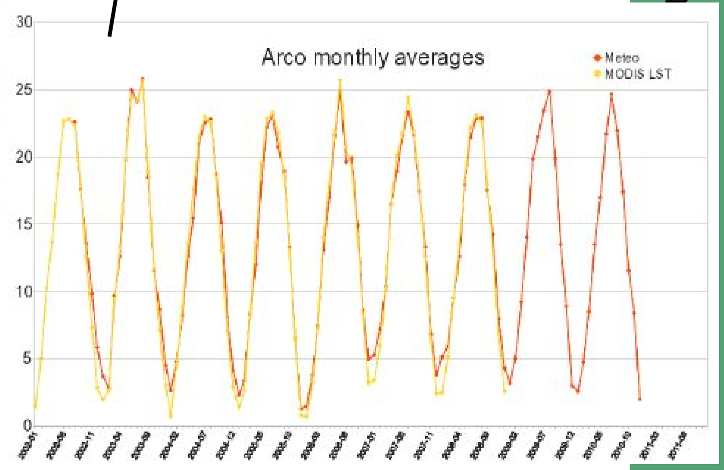
EuroLST: MODIS LST daily time series

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





Pixel-wise time series (meteo versus MODIS LST): ... virtual meteo stations

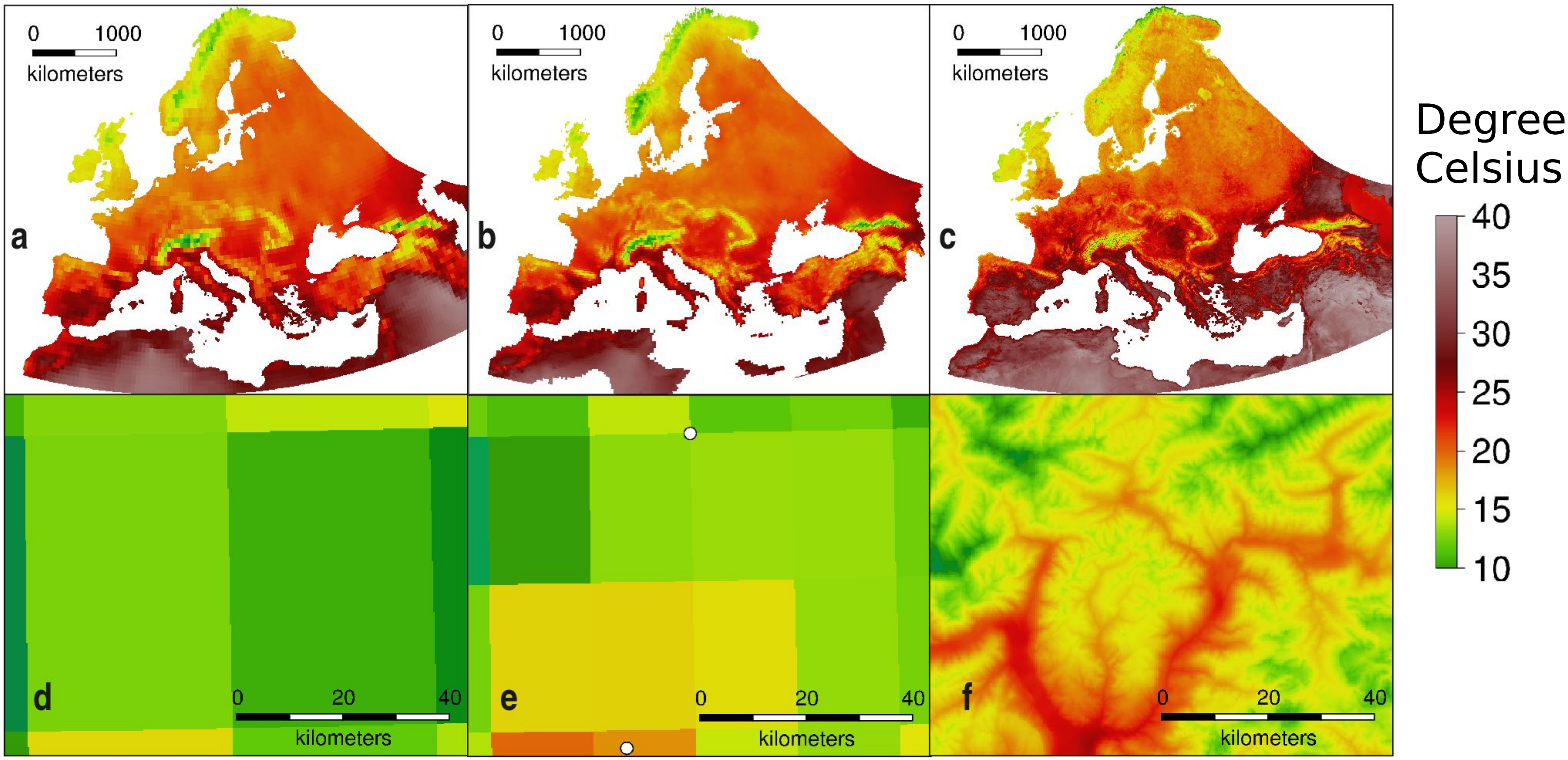


Metz, Rocchini, Neteler, 2014: Rem Sens, 6(5): 3822-3840
EuroLST: http://gis.cri.fmach.it/eurolst/

FONDAZIONE EDMUND MACH

New EuroLST dataset: Comparison to other datasets

(and advantages of using remote sensing time series)



CRU TS 3.1

ECA&D 5.0
(white dots: meteo stations as base for ECA&D interpolation)

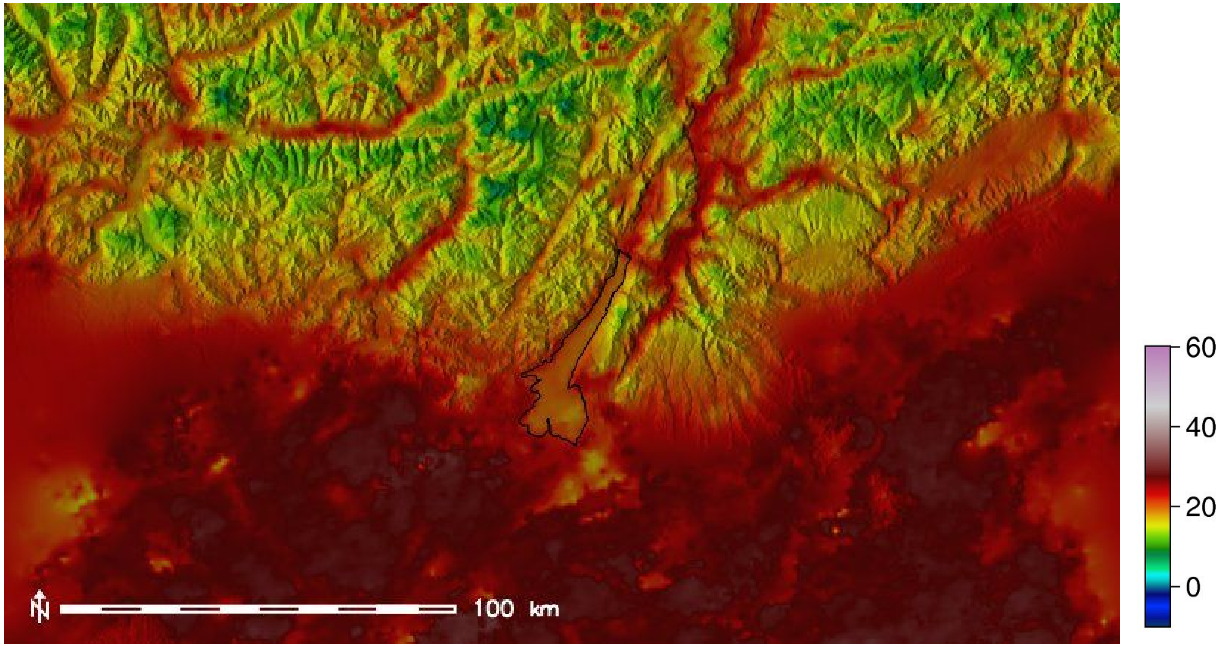
MODIS v005
(reconstructed EuroLST)

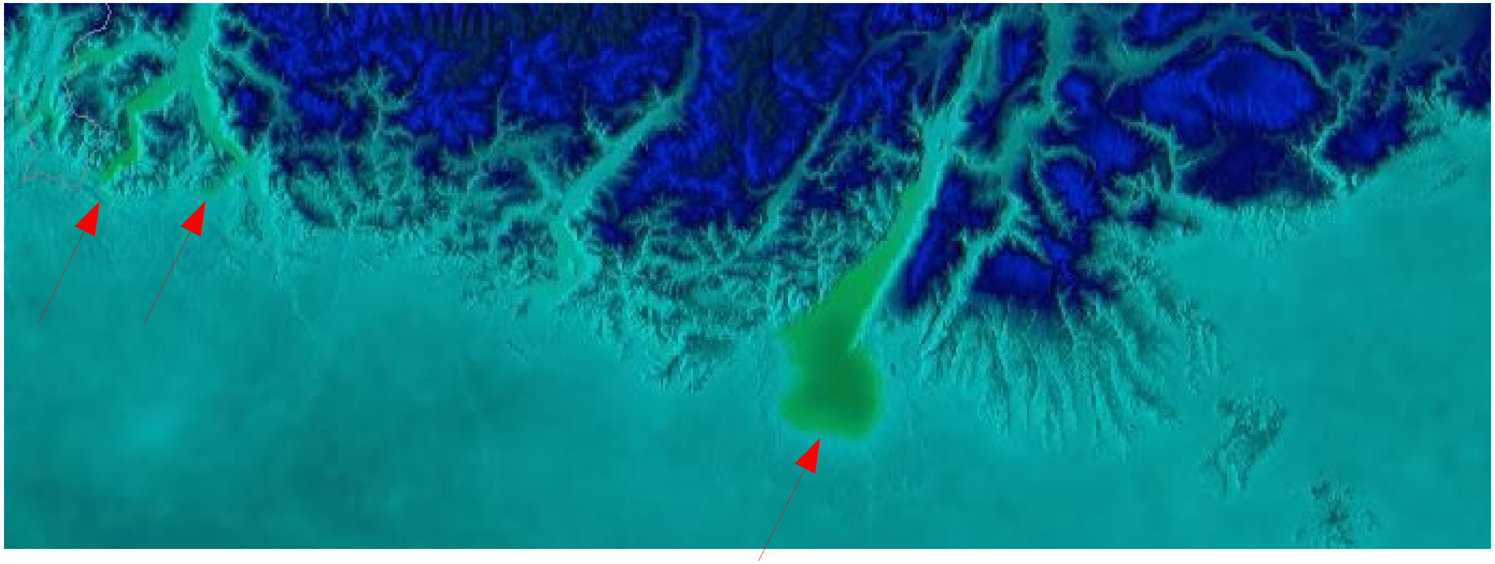
Metz, M.; Rocchini, D.; Neteler, M. 2014: *Surface temperatures at the continental scale: Tracking changes with remote sensing at unprecedented detail*. Remote Sensing. 2014, 6(5): 3822-3840 (DOI | HTML | PDF)

MODIS Land Surface Temperature

Examples:

"Hot" year 2003 and effects in January 2004



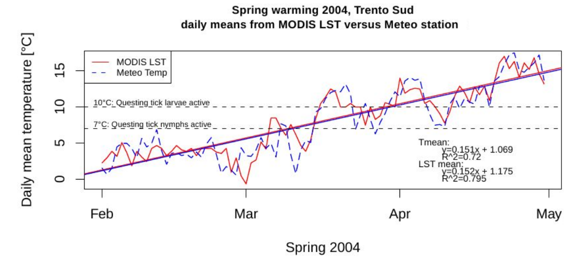


January 2004: Lake Garda still "warm" after hot 2003 summer
--> **local heating effect** = insect overwintering facilitated

Neteler et al. (2013). PLoS ONE, 8(12): e82090 [DOI]

FONDAZIONE EDMUND MACH

Temperature in space and time



Temperature time series

- Average
- Minimum
- Maximum
- Seasonal temperatures: Winter, spring, summer, autumn
- Gradients: Spring warming, Autumnal cooling
- Anomalies, Cool Night Index
- Growing Degree Days (GDD)
- Late frost periods

Monthly avg LST: 01/2002

1000 km
N

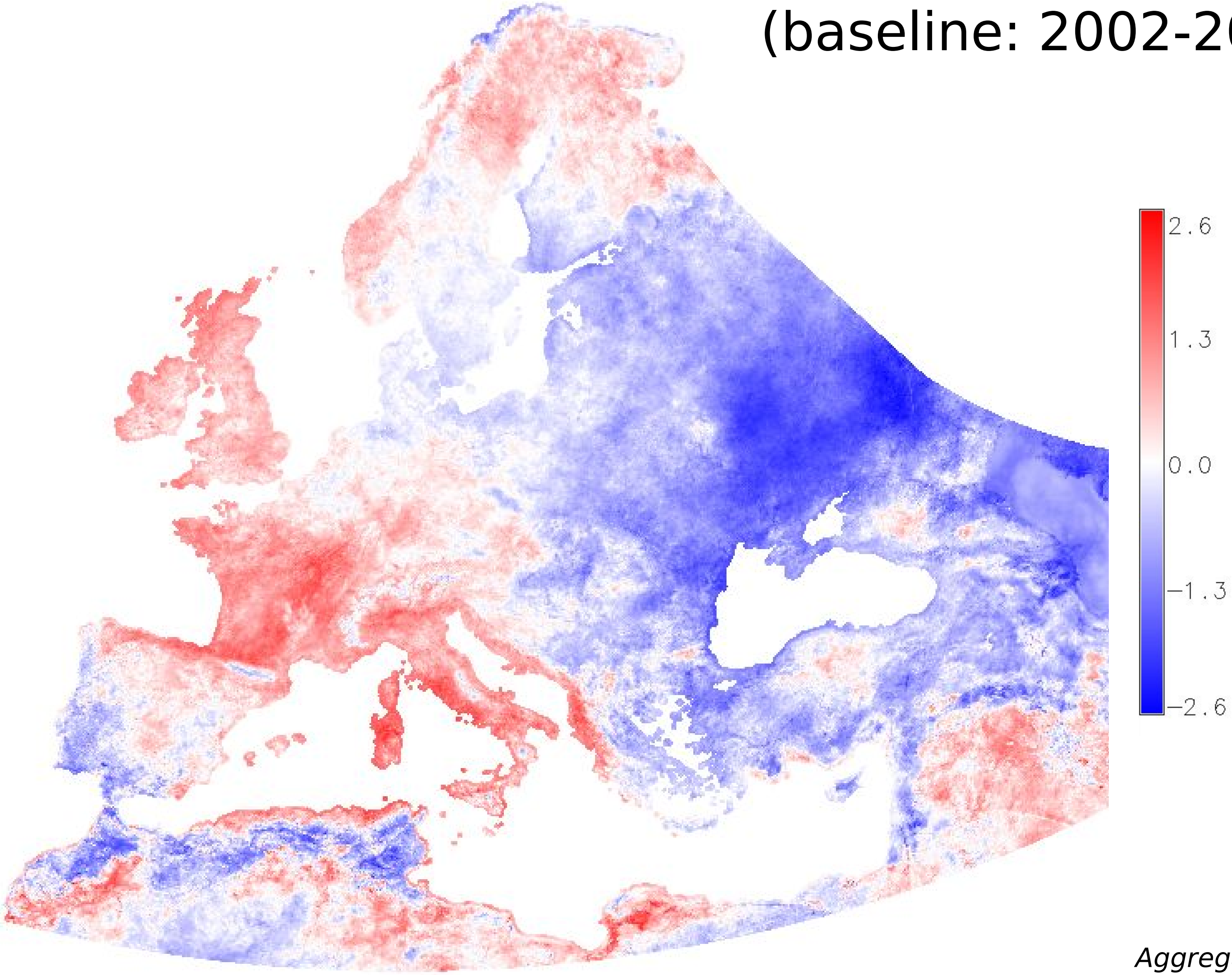
New aggregation tools in GRASS GIS 7

- r.series: pixel-wise aggregation with univariate statistics;
- r.series.accumulate: calculates (accumulated) raster value means (GDD etc);
- r.series.interp: temporal interpolation of missing maps in a time series;
- r.hants (Addon): Fourier based harmonics analysis;
- t.rast.accdetect, t.rast.accumulate, t.rast.aggregate: temporal framework

MODIS LST daily time series

Example of aggregated data

2003: Deviation from baseline average temperature (baseline: 2002-2012)

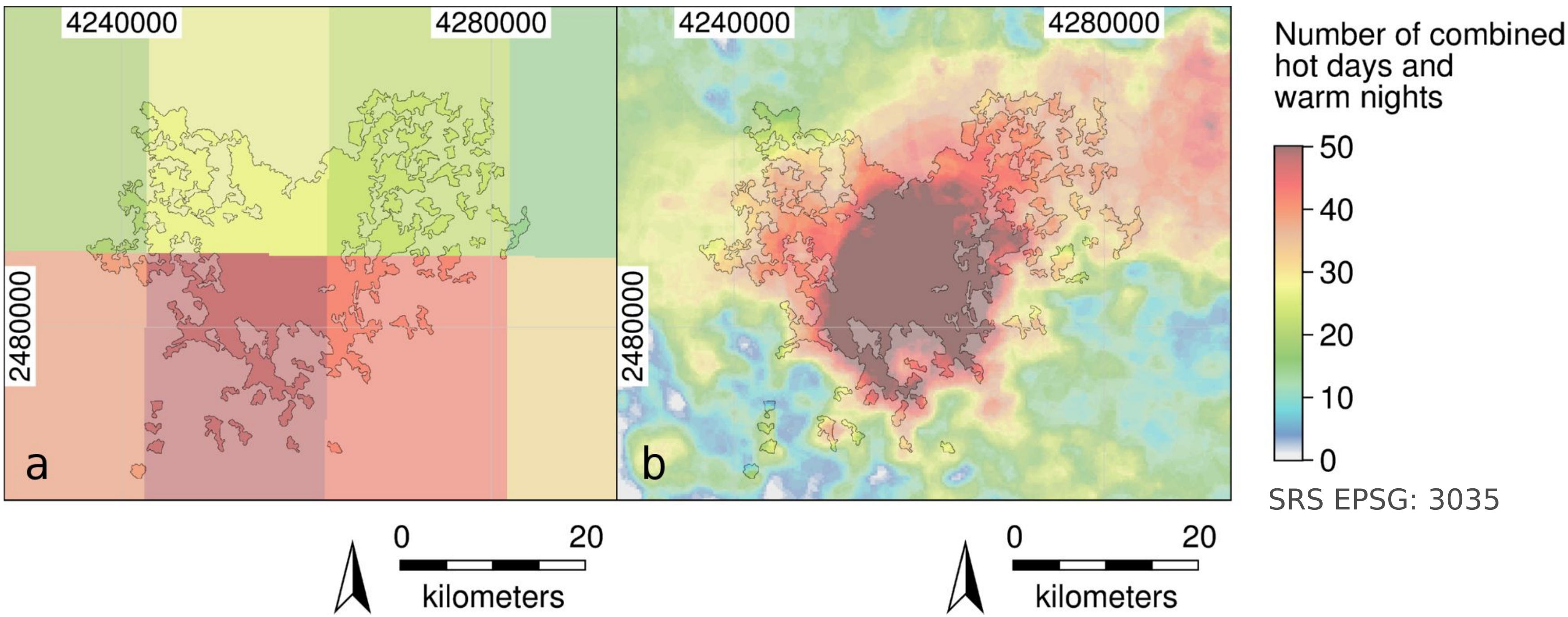


Aggregated by FEM PGIS

Assessing urban heat islands

The example of Milan city, Italy in 2003

Number of **combined hot days** (>35 °C) and **warm nights** (>20 °C)



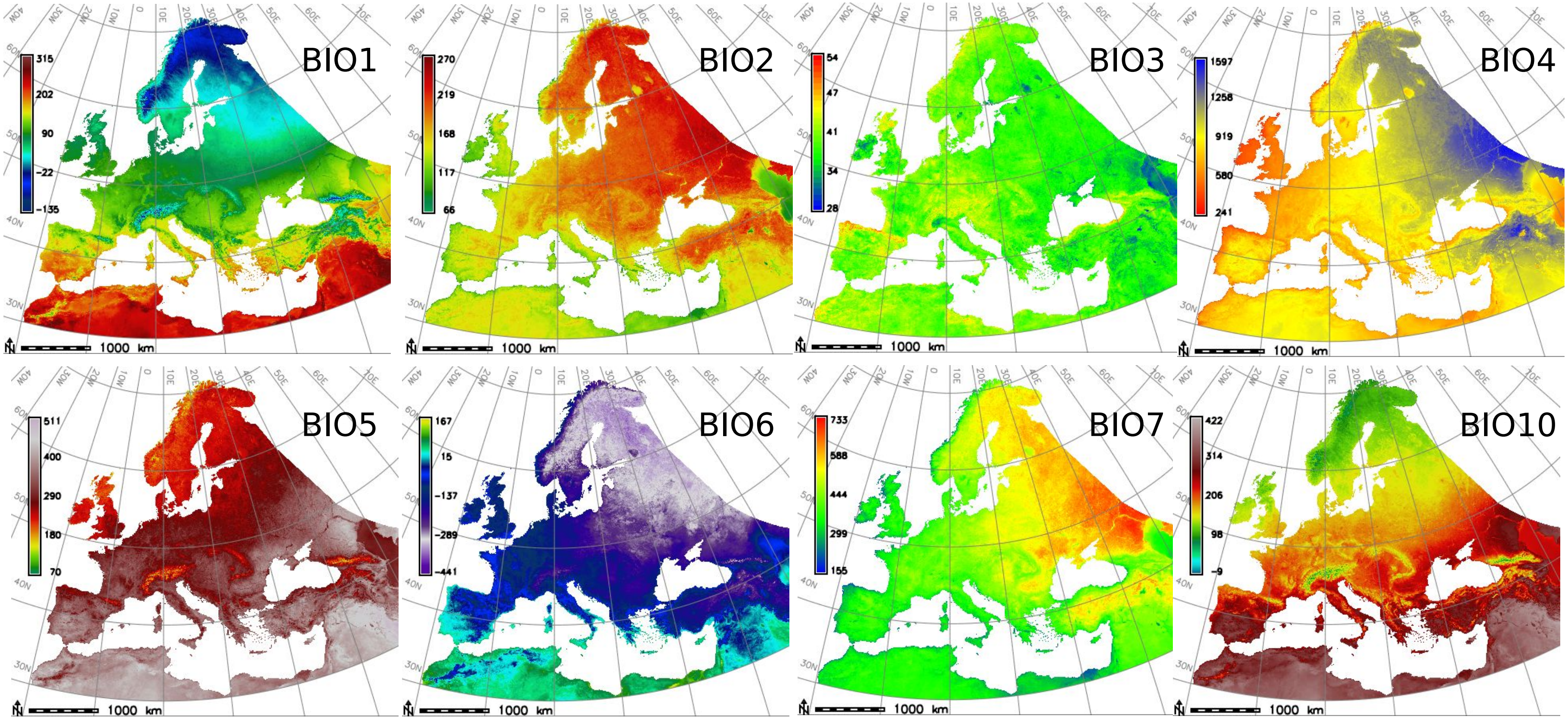
ECA&D data set
~ 25km resolution

EuroLST (MODIS LST)
250m resolution

Refs:

- EEA, http://www.eea.europa.eu/data-and-maps/explore-interactive-maps/heat-wave-risk-of-european-cities-1
- Metz, Rocchini, Neteler, 2014: Rem Sens, 6(5): 3822-3840 [DOI]

BIOCLIM from reconstructed MODIS LST at 250m pixel resolution



BIO1: Annual mean temperature (°C*10)

BIO2: Mean diurnal range (Mean monthly (max - min tem))

BIO3: Isothermality ((bio2/bio7)*100)

BIO4: Temperature seasonality (standard deviation * 100)

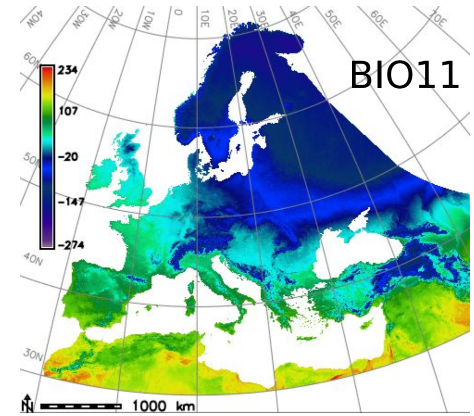
BIO5: Maximum temperature of the warmest month (°C*10)

BIO6: Minimum temperature of the coldest month (°C*10)

BIO7: Temperature annual range (bio5 - bio6) (°C*10)

BIO10: Mean temperature of the warmest quarter (°C*10)

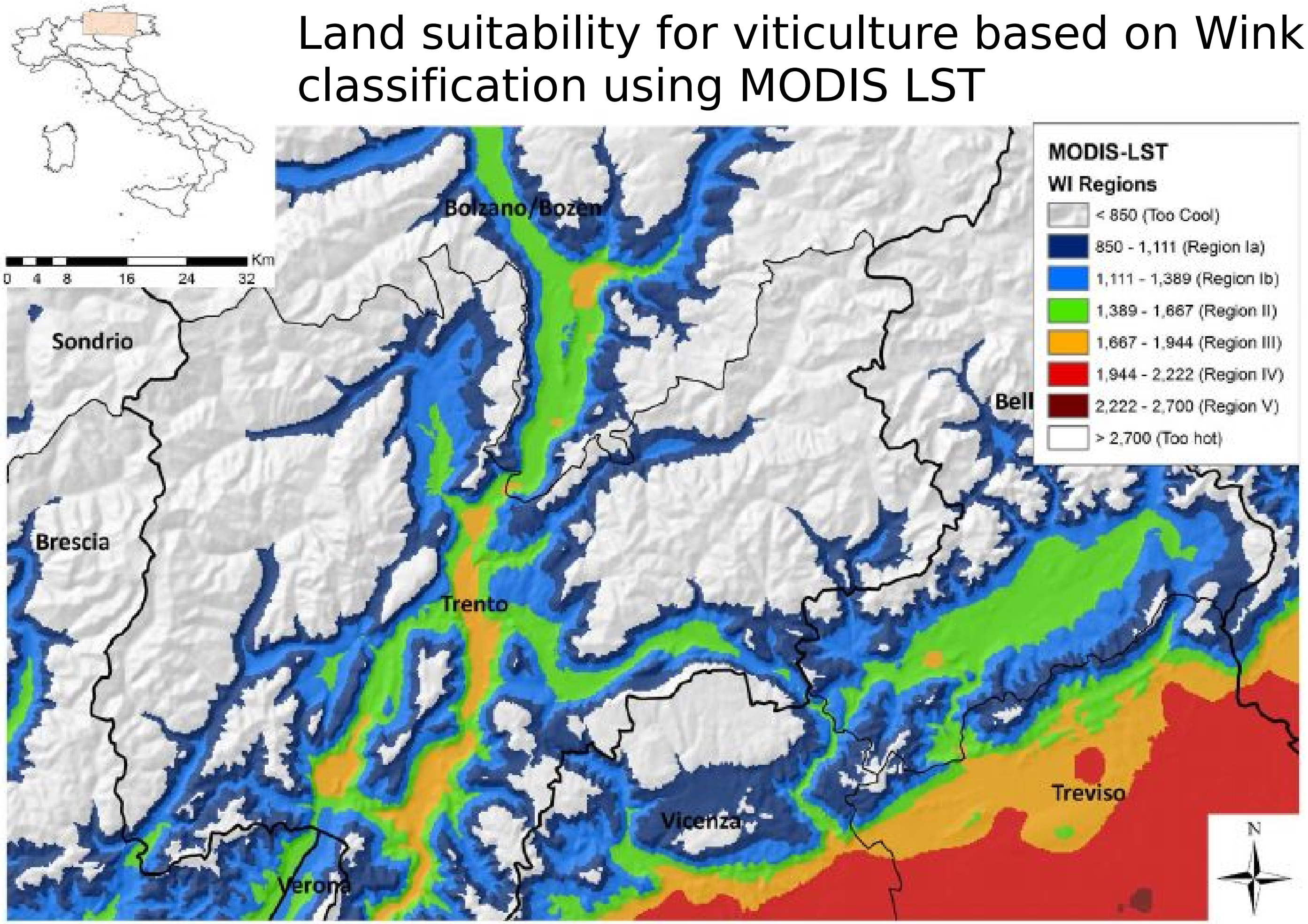
BIO11: Mean temperature of the coldest quarter (°C*10)



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

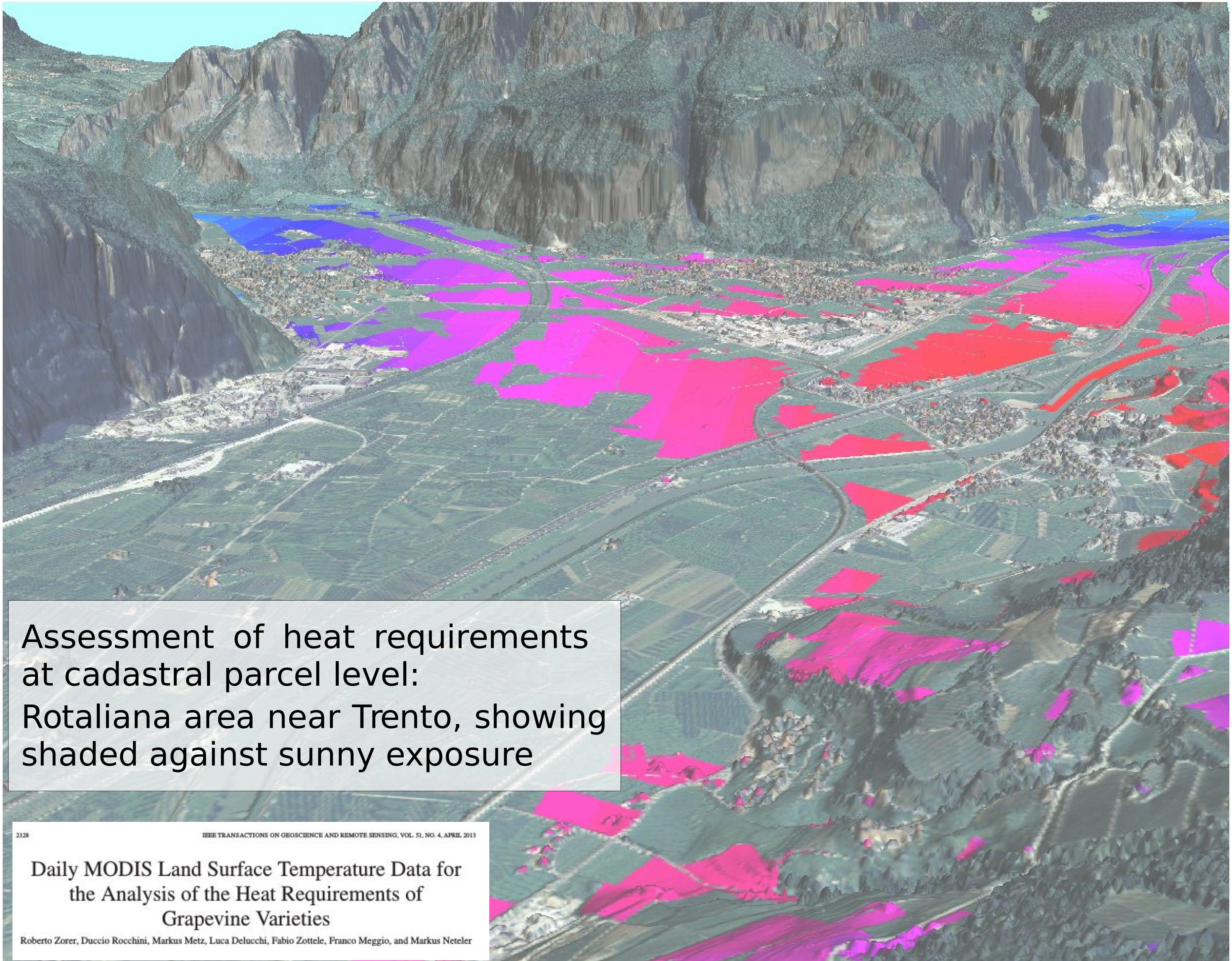
MODIS LST and viticulture

Land suitability for viticulture based on Winkler classification using MODIS LST



- Mountainous areas are **unsuitable** (**too cool**, above 1200 m a.s.l., gray);
- **Regions Ia** (dark blue) and **Ib** (blue): **very cool climate** regions, suitable for hybrid and very early ripening varieties;
- **Region II** (green) is **cool** and suitable for sparkling wine and Müller Thurgau;
- **Region III** (orange) is **warmer** and allows growing of red varieties (Merlot, Cabernet Sauvignon, and the local red varieties Teroldego and Marzemino);
- **Region IV** (red pixels) is **hot** and suitable for late ripening red grape varieties such as Cabernet Franc.

MODIS LST and viticulture



Assessment of heat requirements at cadastral parcel level:
Rotaliana area near Trento, showing shaded against sunny exposure

2128 IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING, VOL. 51, NO. 4, APRIL 2013

Daily MODIS Land Surface Temperature Data for the Analysis of the Heat Requirements of Grapevine Varieties

Roberto Zorer, Duccio Rocchini, Markus Metz, Luca Delucchi, Fabio Zottele, Franco Meggio, and Markus Neteler

Support for massive spatial datasets in GRASS GIS 7

GRASS 7: Support for massive datasets

FONDAZIONE EDMUND MACH

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

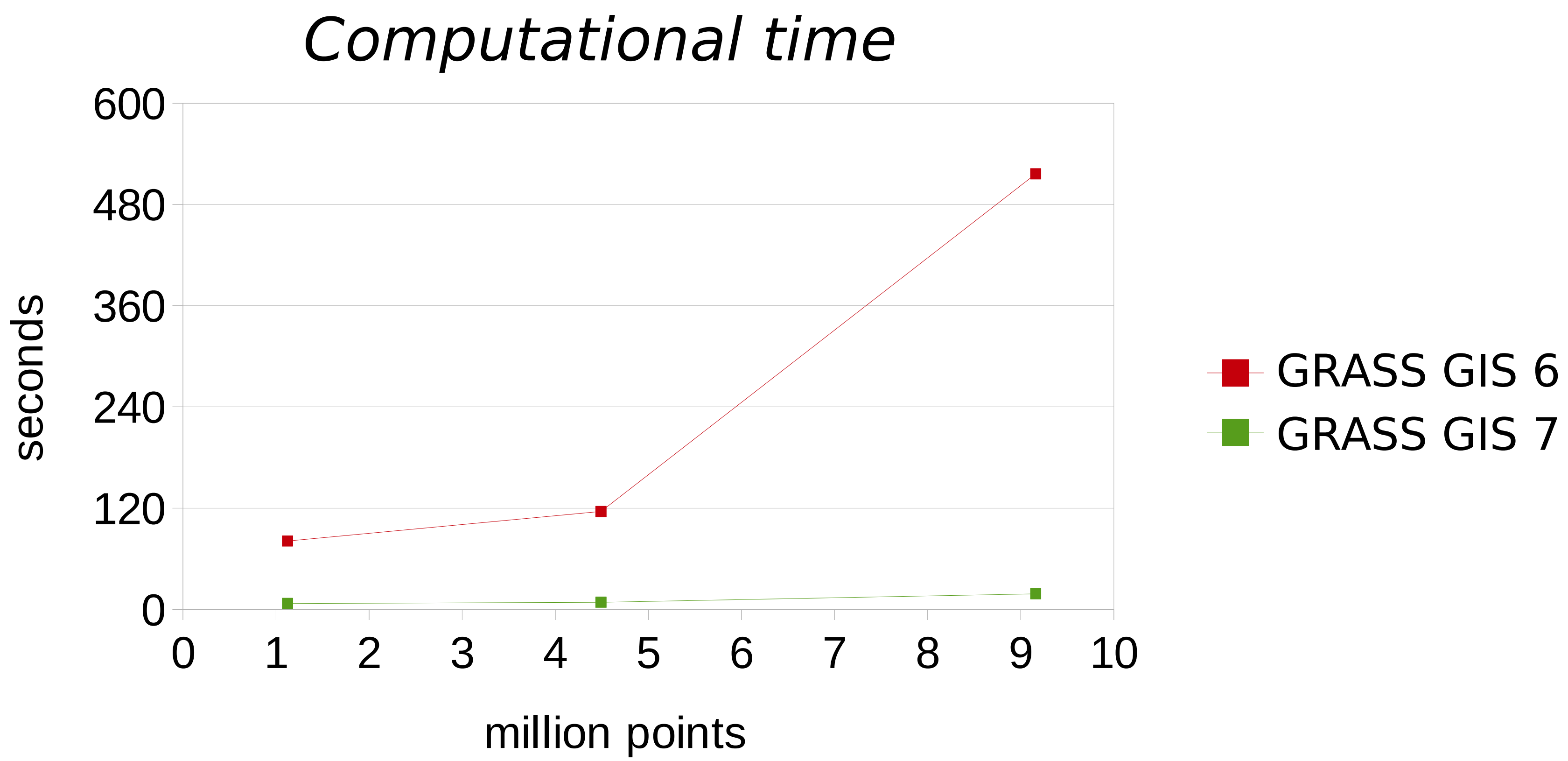
ϟ expensive

✔ cheap
... to solve issue

GRASS 7: Support for massive datasets

FONDAZIONE EDMUND MACH

Cost surfaces: *r.cost*



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

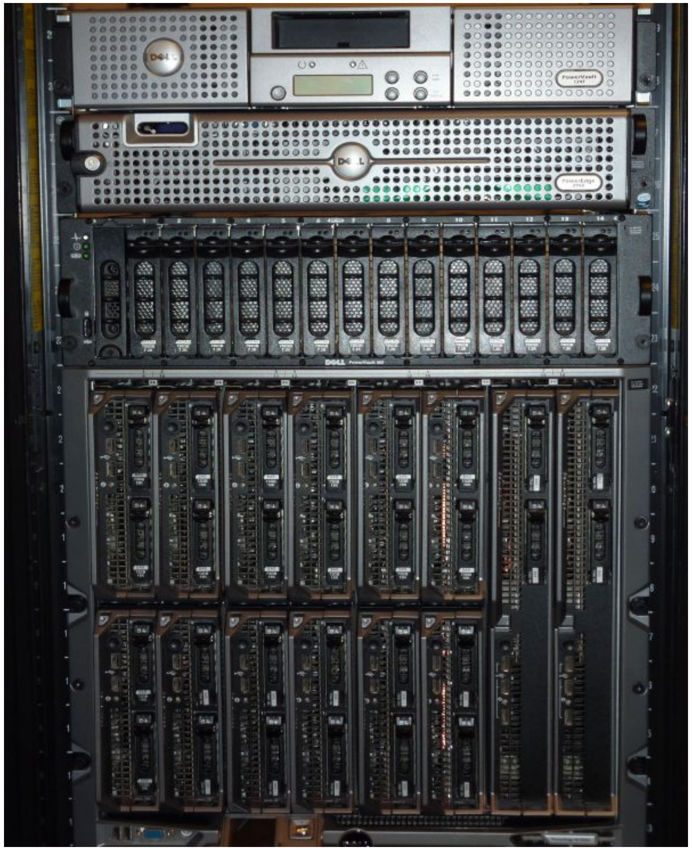
MODIS Land Surface Temperature LST reconstruction

... 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



FONDAZIONE EDMUND MACH

“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)

Environmental Modelling & Software 53 (2014) 1–12

Contents lists available at ScienceDirect

Environmental Modelling & Software

journal homepage: www.elsevier.com/locate/envsoft

TGRASS: A temporal GIS for field based environmental modeling

Sören Gebbert[a,*], Edzer Pebesma[b]

[a] Thünen Institute of Climate-Smart Agriculture, Bundesallee 50, D-38116 Braunschweig, Germany
[b] Institute for Geoinformatics, University of Muenster, Weseler Strasse 253, 48151 Muenster, Germany

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

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

Where is the stuff?

GRASS GIS 7:

Free download for MS Windows, Mac OSX, 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/

Conclusions & Thanks

- **Massive data processing in GRASS GIS 7: most “homework” has been done**
  - **Large file support for raster and vector data**
  - **Temporal data processing framework available**
  - **New Python API integrated (PyGRASS)**
- **New reconstructed MODIS LST dataset available**
- **Next steps:**
  - **Add new big data interfaces to analyse data remotely (rasdaman, sciDB interfaces?)**

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

Thanks to **NASA** Land Processes Distributed Active Archive Center (**LP DAAC**) for making the MODIS LST data available.