

Xth Meeting of the Italian GRASS and GFOSS users
25-27 Feb 2009, Cagliari, Italy

**Ecological indicators extracted from satellite
data: high performance computing with
GRASS GIS**

M. Neteler

Fondazione Mach - Centro Ricerca e Innovazione
38100 Viote del Monte Bondone (Trento), Italy

<http://gis.fem-environment.eu/>
<http://www.grassbook.org>

neteler * cealp.it

The problem: Emerging infectious diseases in Europe



Increasing number of emerging infectious diseases in Europe for ecological and socio-economical reasons



- **Tick-borne** diseases: Tick-borne Encephalitis, Anaplasmosis, Lyme Borreliosis, ...
- **Rodent-borne** diseases: Hantavirus, Arenavirus, ...
- **Mosquito-borne** diseases: West Nile virus, Leishmaniasis, Bluetongue, Chikungunya, ...

Surveillance needed!

Two research projects at FEM-CRI:

1) EDEN (Emerging Diseases in a changing European eNvironment) is a FP6 Integrated Project (2004-2009) with 48 research institutes from 24 countries
<http://www.eden-fp6project.net>

2) RISK TIGER: Risk assessment of the emergence of new arboviruses diseases transmitted by the **tiger mosquito** *Aedes albopictus* (Diptera: *Culicidae*) in the Autonomous **Province of Trento**

<http://risktiger.fem-environment.eu>



CDC

Which data are available?



Sensors/data of ecological relevance and low access costs

Sensor	Period	Spatial resol.	Temporal resol.	Format (GDAL)
AVHRR: <ul style="list-style-type: none"> • Land Surface Temperature (LST) • Vegetation Index (NDVI) 	1978-today	~ 1km	Daily	L1B
MODIS: <ul style="list-style-type: none"> • Land Surface Temperature (LST) • Snow extent • Vegetation Index (NDVI) • LAI/FPAR • ... 	2000-today	1km 500m 250m	Daily	HDF
SPOT Vegetation VGT (NDVI)	1998-today	1km	10 days	HDF
LANDSAT TM1-7 (VIS, NIR, TIR)	1972-today	15/30/ 60m	16 days	GeoTIFF
ASTER (VIS, NIR, TIR)	2000-today	15/30/ 90m	16 days	HDF
Digital elevation model (DEM)	-	meter(s)	-	Various



Needed ecological indicators



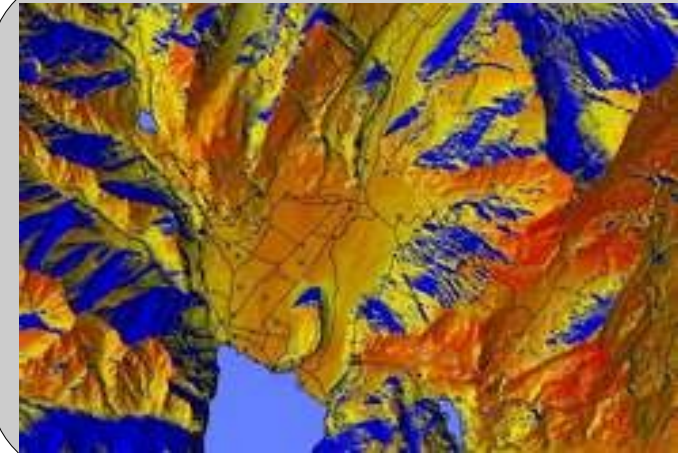
Example: indicators for ticks and mosquitoes surveillance

Specific...



Ixodes ricinus

- Length of growing period
- seasonal gaps between 7°C (tick nymphs active) and 10°C (tick larvae active)
- higher/colder than average summer temperatures
- ...



In common...

- Insolation time (photoperiod)
- Mild winters (> 7°C)

Specific...

- Mean January temperatures
- Degree days
- Onset of greening
- ...

Aedes albopictus



CDC

GIS Processing requirements



<http://grass.osgeo.org>

Software: PROJ4 + GDAL + GRASS GIS

Data:

- **MODIS LST** - reconstruction: **v.vol.rst**, **slow** with several hours per map (size: 3 Italian provinces)
- aggregation: **r.series**, fast
- **MODIS EVI**: **r.mapcalc**, **r.series**, fast
- **MODIS SNOW**: **r.mapcalc**, **r.series**, fast

- **DEM**: **r.horizon** + **r.sun**, **slow**, much RAM needed for high-res DEM of 5m



Refs:

Neteler, 2005. Time series proc. MODIS..., Intl J Geoinformatics
Rizzoli et. al., 2007, TBE. Geospatial Health
Carpi et al., 2008, TBE. Epidem. & Infect.

Indeed, a stress test for a cluster!

High Performance Computing GIS (HPC-GIS)



**Hardware: FEM-CRI GIS Cluster:
128 CPUs with 400 Gb RAM**

Front-end computer + Storage

- 4-Core Xeon with 8Gb RAM
- 14x 750 GB disks in RAID5

Blades

- Chassis
- 7 blades: 2x Quad-Core Xeon with RAM 16GB
- 5 blades: 2x Quad-Core Xeon with RAM 32GB
- 2 double-blades: 4x Quad Core Opteron, RAM 64GB

Network switches

- 10 Gb ethernet

Energy consumption

- Typically: 2700W

(compare: 64-dualcore-PCs * 250W = 16000W)



<http://gis.fem-environment.eu>

High Performance Computing GIS (HPC-GIS)



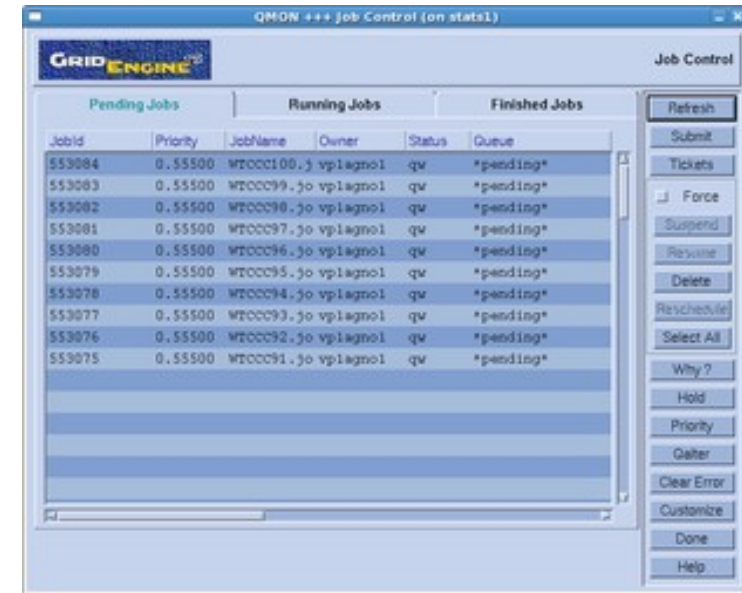
Software: FEM-CRI GIS Cluster

Operating System

- Scientific Linux 5.2 (front-end machine)
- Scientific Linux 5.2 LiveDVD (blades)

Blades are booted diskless via PXE/DHCP/TFTP
the SL LiveDVD with local modifications

<http://www.livecd.ethz.ch/diskless.html>

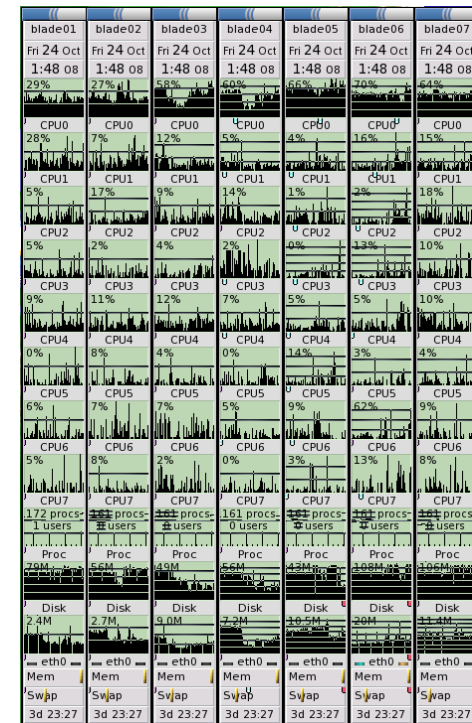


Job scheduler

- Grid engine: <http://gridengine.sunsource.net/>

Data access

- HOME and data directories shared across blades via NFS
- Applications like GRASS installed in and run from HOME



HPC-GIS: Implementing GRASS jobs



Note: GRASS does not (yet) support MPI, no relevant parallelized code

Solution: parallelize by per-map calculations

A) First step: Serial batch job

GRASS_BATCH_JOB: shell script to be processed as batch job

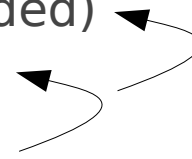
```
export GRASS_BATCH_JOB=/path/to/myjob.sh  
grass64
```

The job will be executed in GRASS without user interaction.
Once this works...

B) Second step: Launch parallel batch jobs

- GRASS_BATCH_JOB as above (use env. variables as needed)
- Launch-the-batch-job script with scheduler parameters
- Submit the launch-the-batch-job script with “qsub”

- Check status with
 - qstats
 - qmon



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

<http://gis.fem-environment.eu/grid-engine-howto/>

Planning GIS computations



GRASS' r.sun and RAM consumption

PAT Lidar 1m/2m DEM downsampled to 5m with gdalwarp (bilinear)

```
> rows=19509
> cols=23522
> IR=4   # elev + asp + slo + turbid
> OR=5   # insol + beam + diff + refl + global
> horizonsteps=30

> numpartitions=1
> rows*cols*((IR*4+horizonsteps)/numpartitions + OR*4) /1024 /1024 /1024
[1] 28.20677

> numpartitions=64
> rows*cols*((IR*4+horizonsteps)/numpartitions + OR*4) /1024 /1024 /1024
[1] 8.854681

      (results in GB)
```

This calculation for each day of the year...



Signs of NFS overload: the r.sun case

Blade

```
top - 22:11:04 up 5:28, 1 user, load average: 3.10, 2.02, 0.90
Tasks: 247 total, 1 running, 246 sleeping, 0 stopped, 0 zombie
Cpu(s): 5.3%us, 0.6%sy, 0.0%ni, 83.0%id, 10.7%wa, 0.0%hi, 0.5%si, 0.0%st
Mem: 66070020k total, 3829264k used, 62240756k free, 129484k buffers
Swap: 31249992k total, 0k used, 31249992k free, 1187296k cached
```

PID	USER	PR	NI	VIRT	RES	SHR	S	%CPU	%MEM	TIME+	COMMAND
21139	neteler	16	0	775m	739m	2876	D	32	1.1	4:42.00	r.sun
21140	neteler	16	0	775m	739m	2872	D	29	1.1	4:41.20	r.sun
21138	neteler	16	0	775m	739m	2876	D	28	1.1	4:39.91	r.sun
14161	partimag	15	0	112m	1196	960	S	6	0.0	20:22.59	gkrellmd
13959	sgadmin	15	0	70188	2600	1548	S	3	0.0	0:09.47	sg_execd

Waiting almost forever...

Front-end

```
top - 22:39:40 up 6:24, 3 users, load average: 32.84, 27.66, 16.83
Tasks: 157 total, 1 running, 156 sleeping, 0 stopped, 0 zombie
Cpu(s): 0.0%us, 1.5%sy, 0.0%ni, 0.0%id, 98.0%wa, 0.1%hi, 0.4%si, 0.0%st
Mem: 8167044k total, 8119292k used, 47752k free, 23588k buffers
Swap: 2031608k total, 140k used, 2031468k free, 7499840k cached
```

...solved with NFS tuning and GE resource management.

Optimization

- NFS settings: hard, intr
- Network MTU: jumbo frames (to better deal with large GIS datasets)



Indicators from MODIS sensor 2/5

Enhanced Vegetation Index (EVI)

EVI tends to perform better than Norm. Differences Veg. Index (NDVI):

- less prone to saturation
- less sensitive to haze

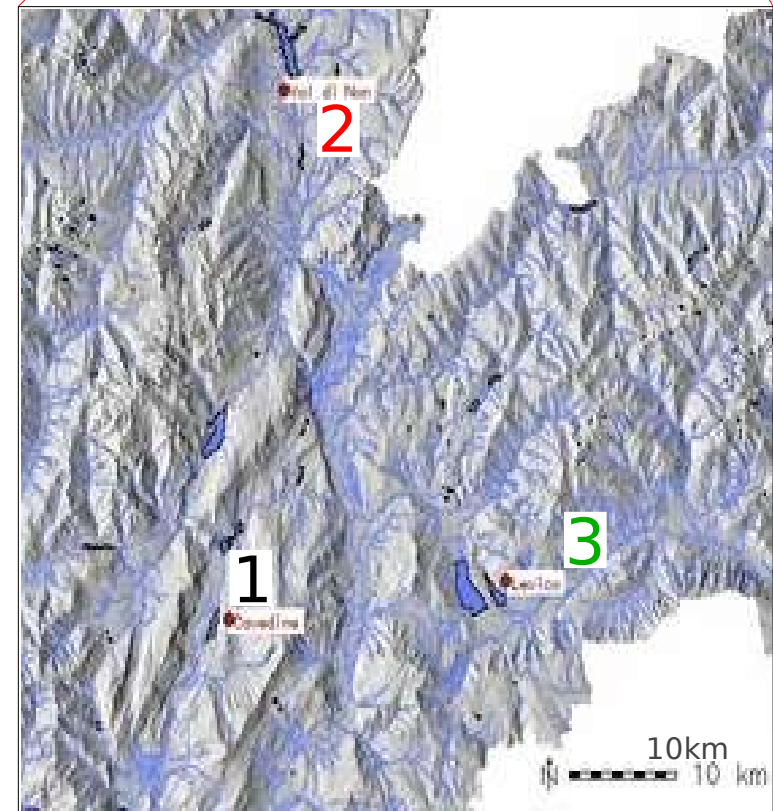
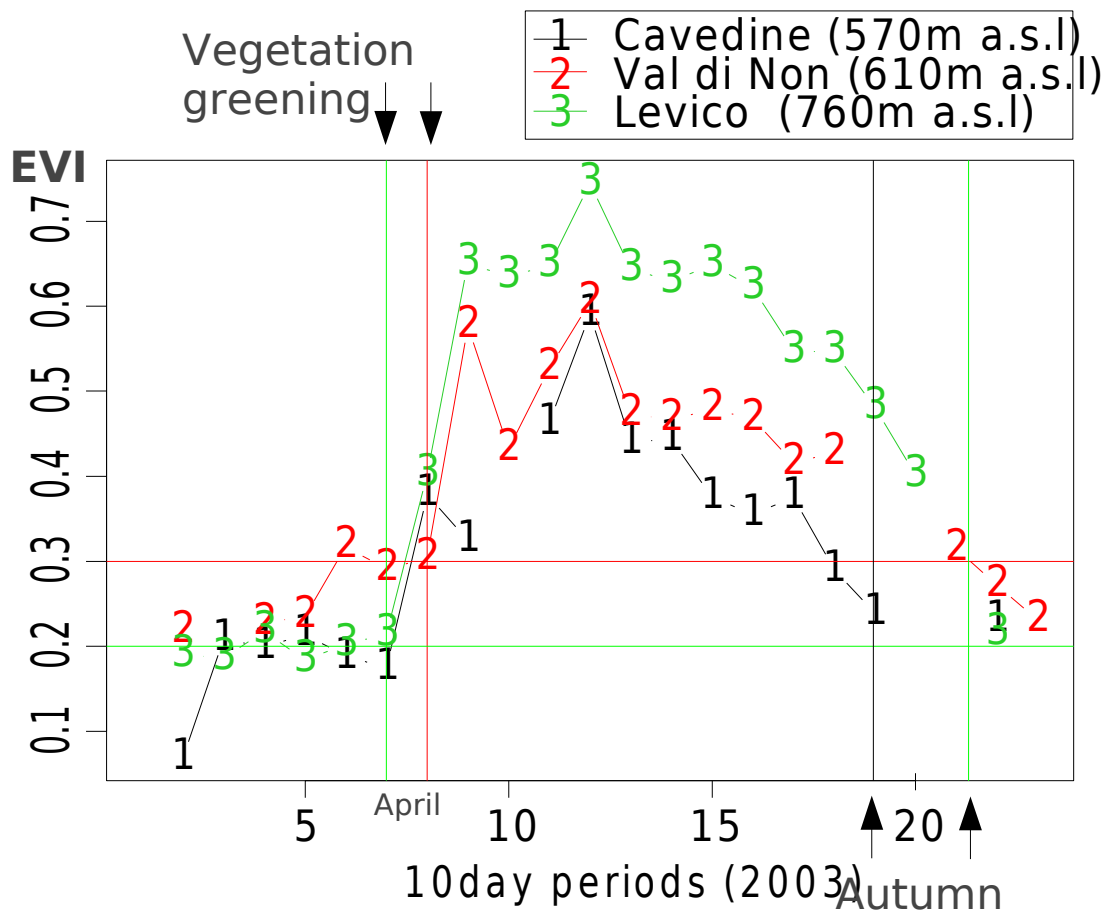




Vegetation Indicators from MODIS sensor

Enhanced Vegetation Index (EVI)

- ◆ “Spring/autumn detection”: Trentino
→ Effects of valley orientation and exposition
- ◆ Growing period length





Snow indicators from MODIS sensor

Maximum snow extent map: accumulated over 8 days

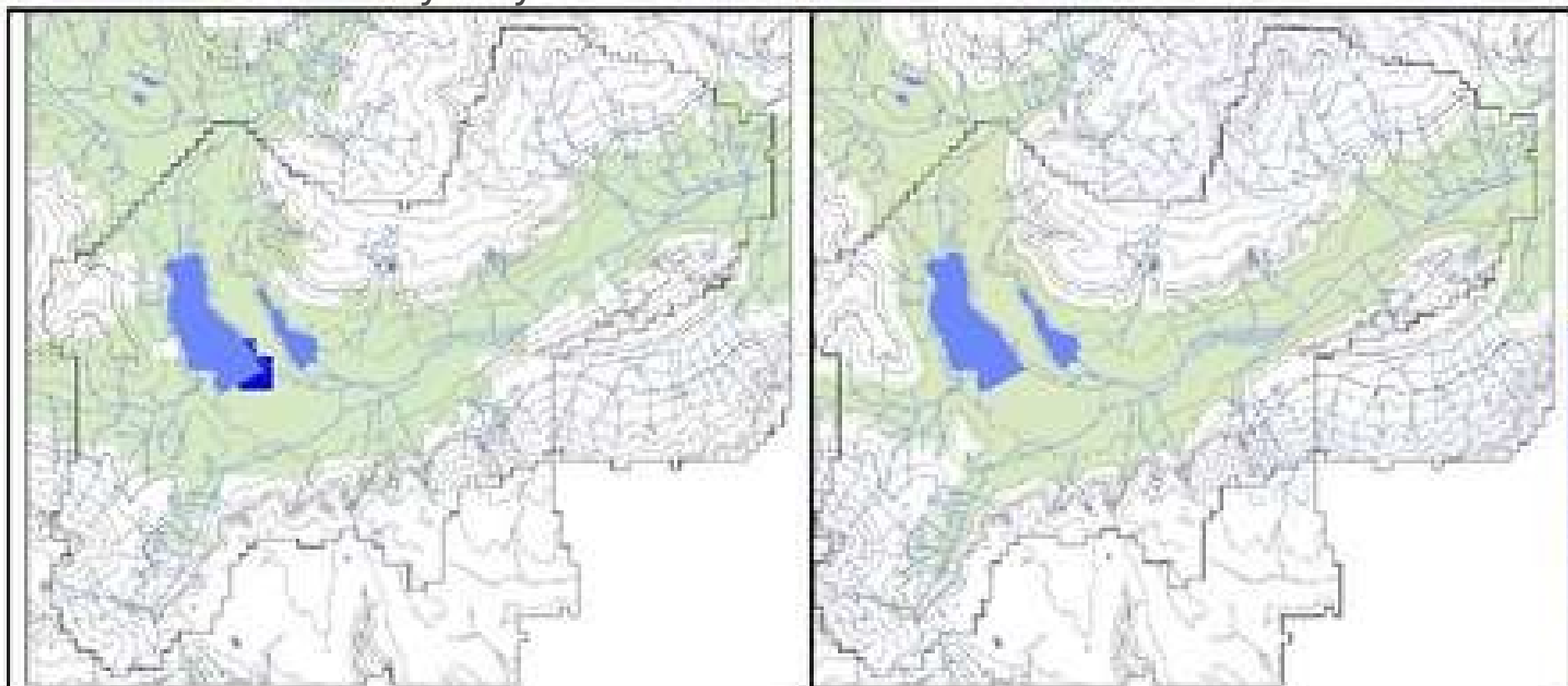
Example: Early snow event in October 2004



MODIS sensor based map
(satellite, every 8 days)
=> the "easy way"



GEOtop snow-model based map
(using climate data)



Situation 24th Oct. 2004
Pergine, Valsugana (Trentino), Italy

Endrizzi, Bertoldi, Neteler,
Rigon, 2005. EGU





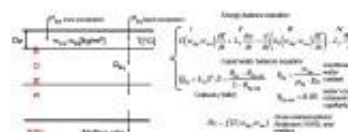
Snow indicators from MODIS sensor

Maximum snow extent map: accumulated over 8 days

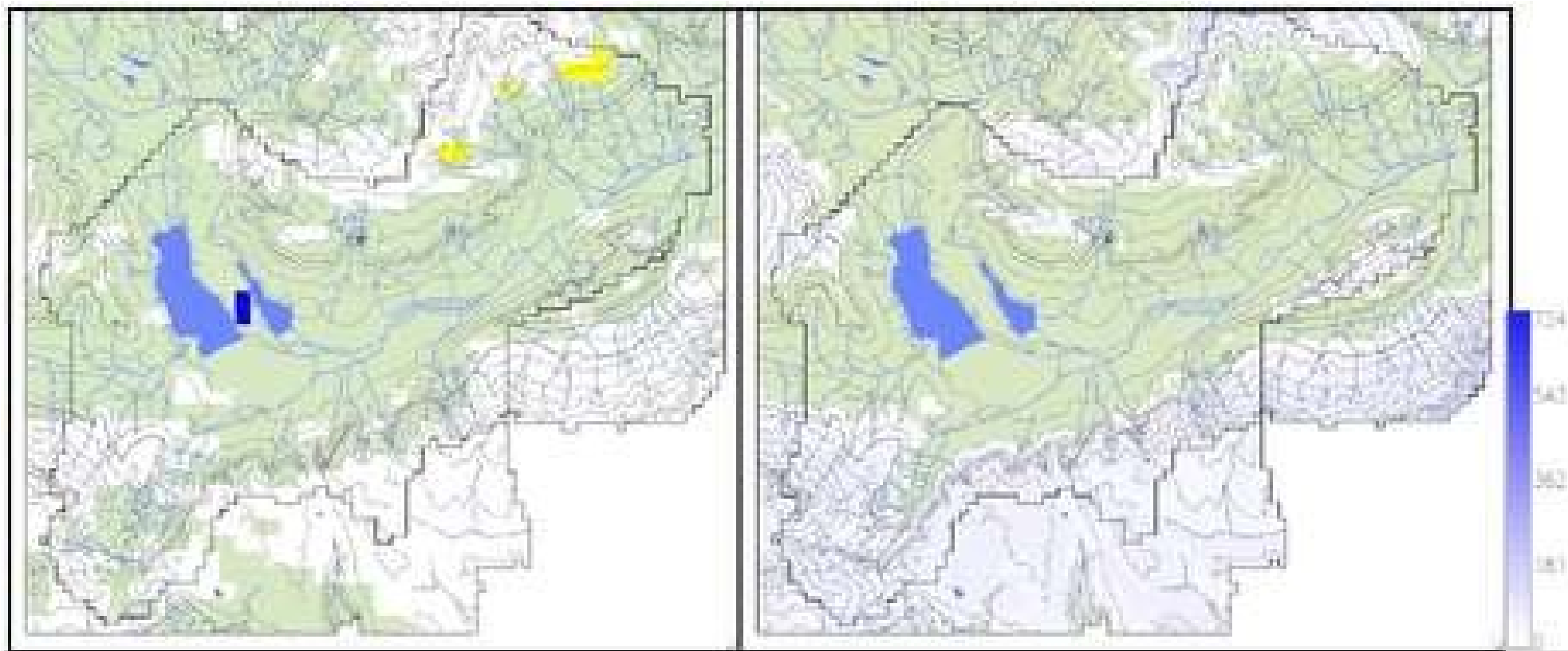
Example: Early snow event in October 2004



MODIS sensor based map
(satellite, every 8 days)
=> the "easy way"



GEOtop snow-model based map
(using climate data)



Situation 17th Nov. 2004
Pergine, Valsugana (Trentino), Italy

Endrizzi, Bertoldi, Neteler,
Rigon, 2005. EGU

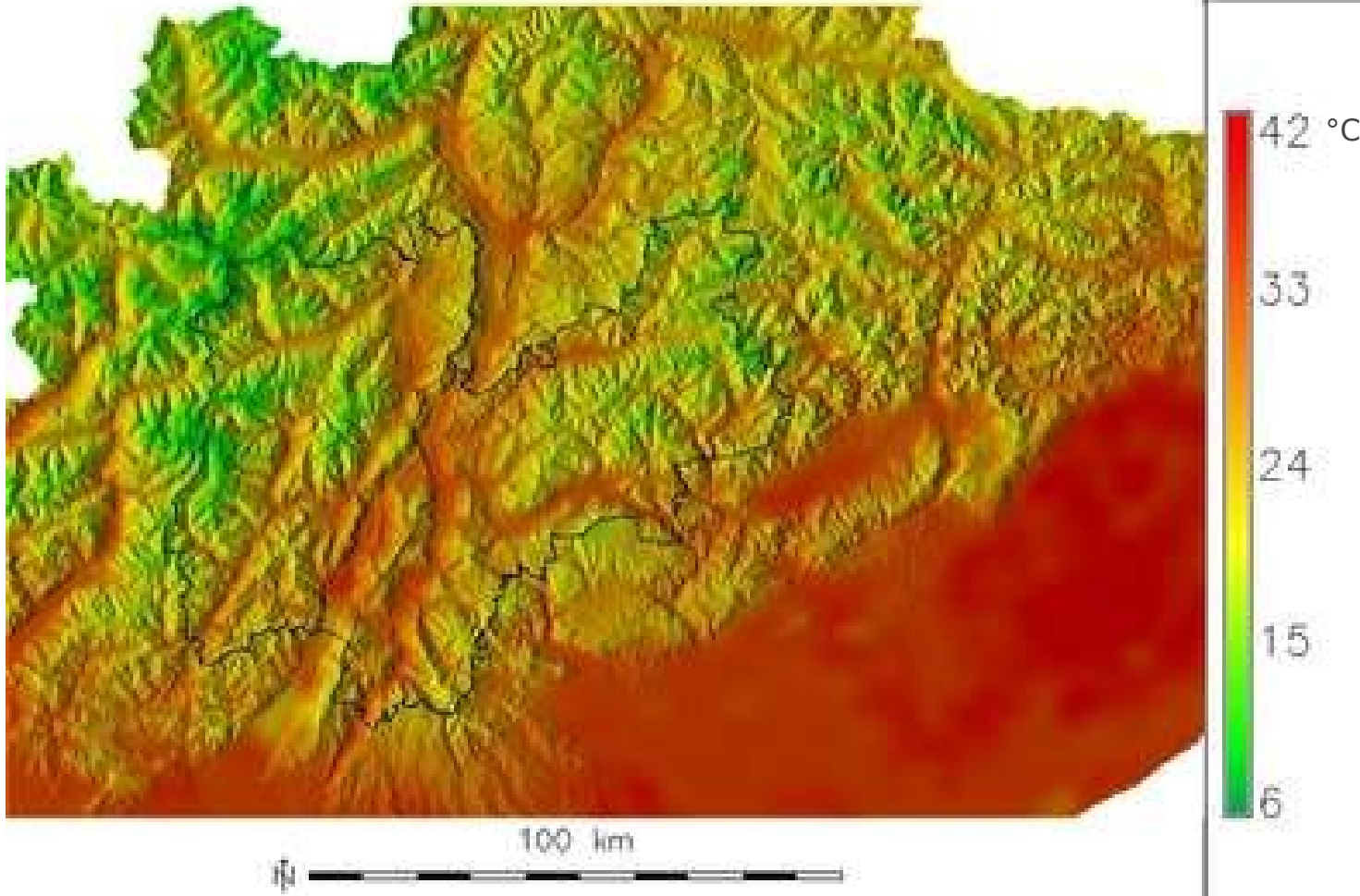


MODIS LST reconstruction 2/5

Approach (simplified)

- Temperature gradient from MODIS LST image statistics
- If too few pixel, use seasonal gradient
- Interpolate with Volume Splines in GRASS using elevation as auxiliary variable
- Correction for south/north exposed slopes

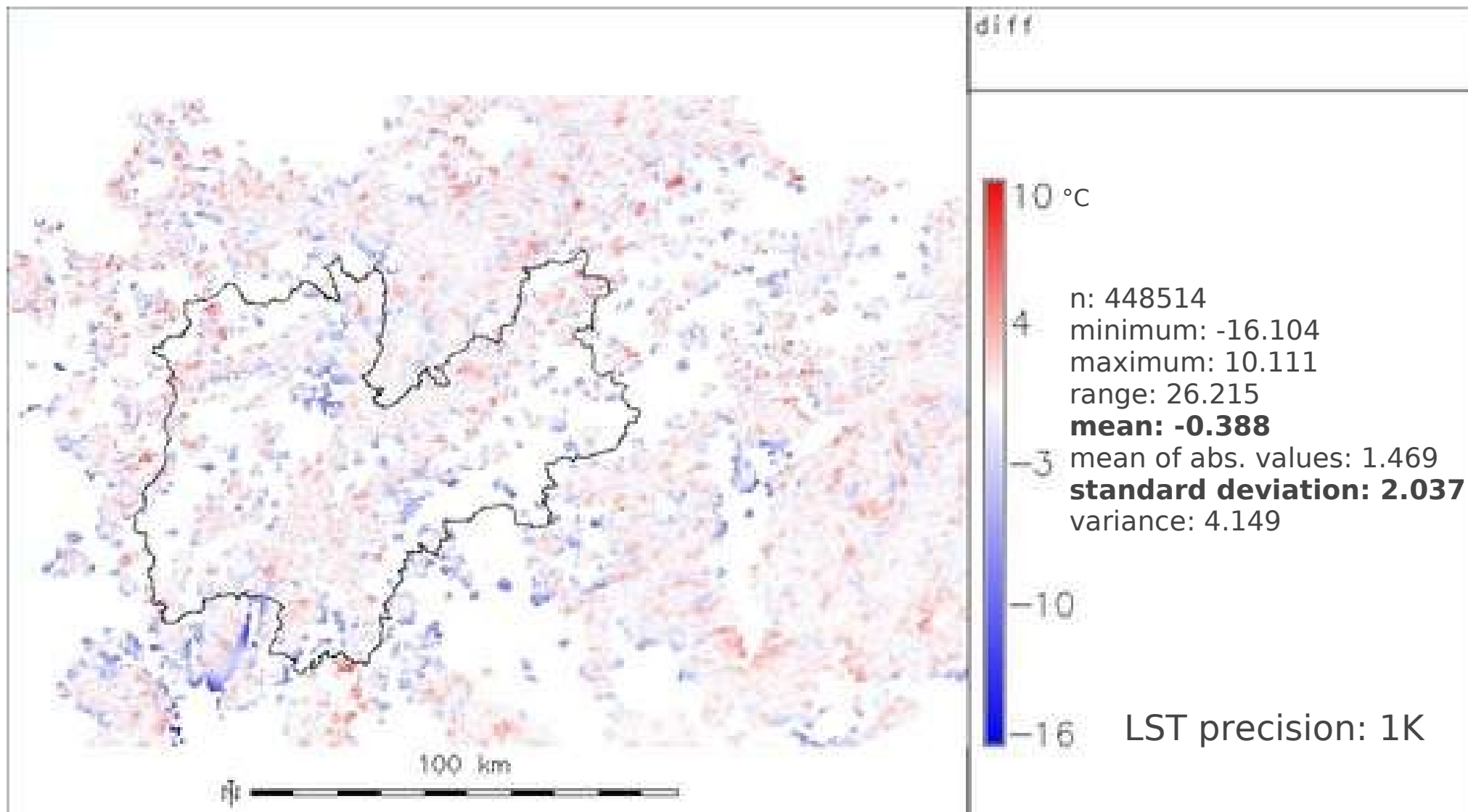
aqua_l1st1km20060628.LST_Day_1km, r
Reconstructed MODIS LST map



MODIS LST reconstruction 3/5



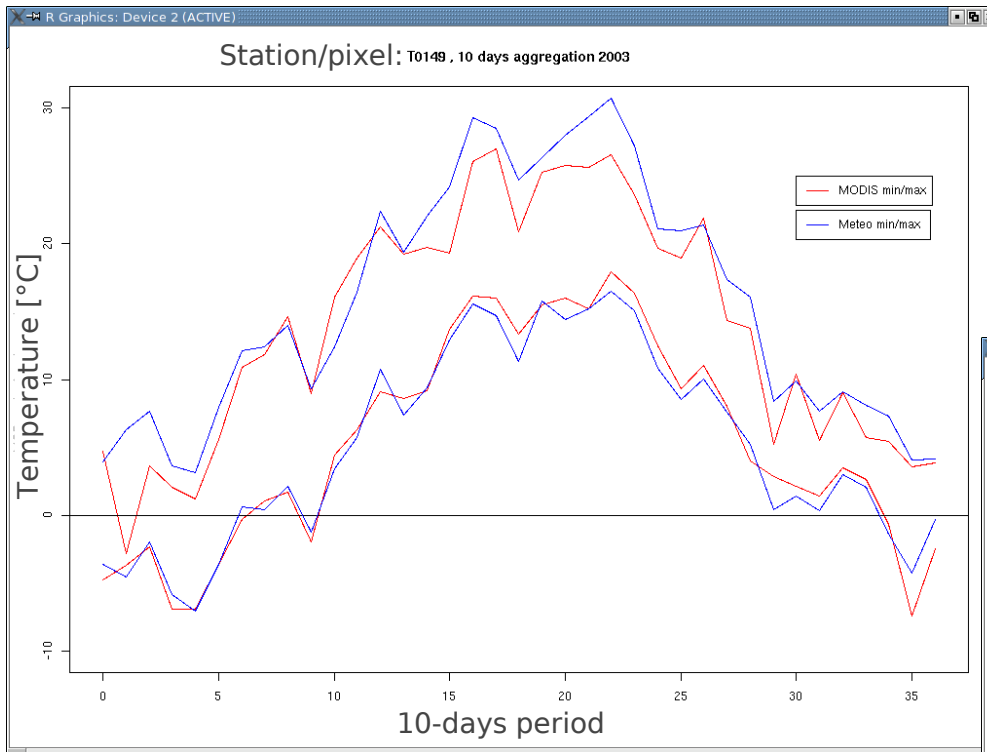
Difference map: filtered MODIS LST - RST3D interpolated MODIS LST



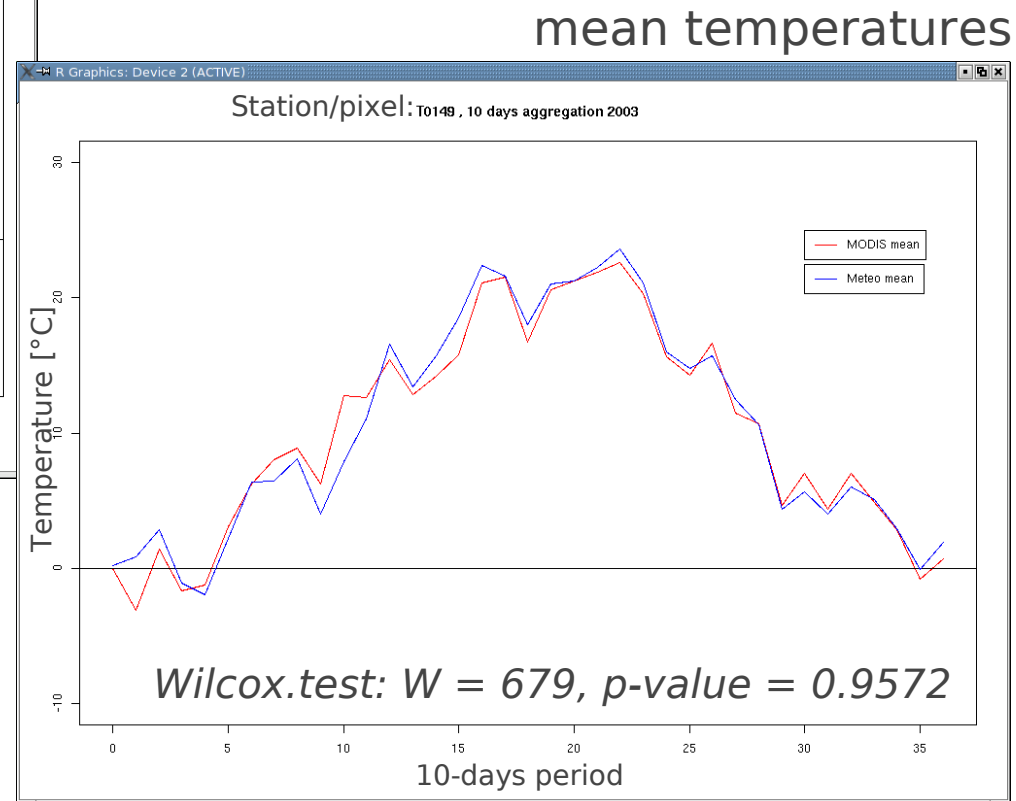
Comparing MODIS LST and meteorological data



10 days aggregates: time series processing (GRASS GIS) Comparison of **meteo station** and **MODIS**



minimum/maximum temperatures



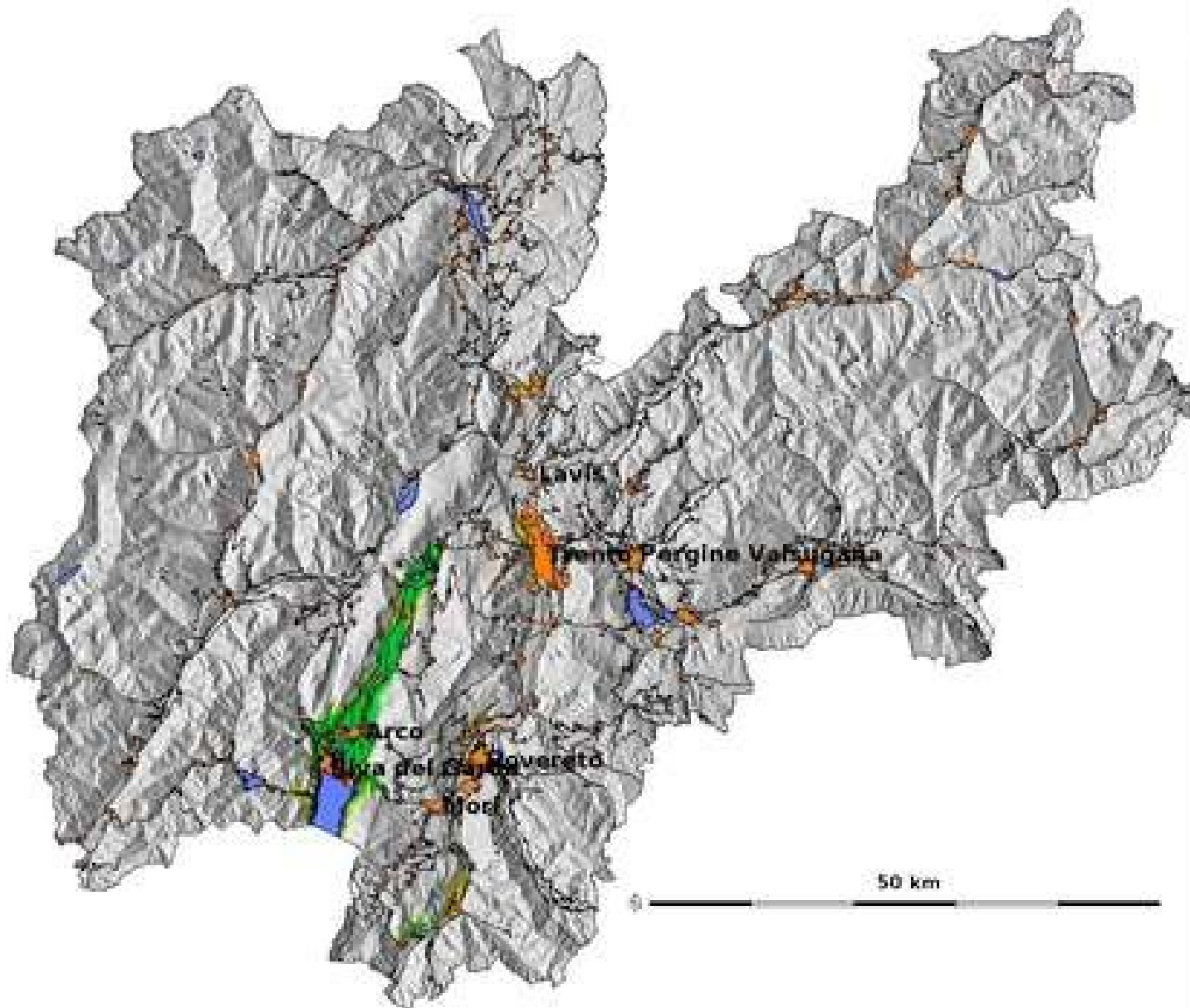
Note: Land surface temperature != air temp.

Station Speccheri (860m;
GB 1666033E 5070563N)

Accumulated MODIS LST map for winter survival



Accumulated numbers of Januaries with mean temp $\geq 0^{\circ}\text{C}$



threshold_gt0_sum

Number of year over Threshold (2000-2007)

1

2

3

4

5



Urban areas

Note: preliminary results based on MODIS V004

Conclusions + Outlook



- Rich archive of remote sensing data available (thanks to the US legislation)
- Data processing is completely based on FOSS4G software
- Time series permit for extraction of ecological indicators
-> seasonality patterns
- New satellite systems provide a wealth of data from which **epidemiologically relevant indicators** can be derived
- TODO: continue to implement SQL based raster time series support in GRASS
- Integrate into risk modelling



Markus Neteler

Fondazione Mach - Centro Ricerca e Innovazione
GIS and Remote Sensing Unit
38100 Viote del Monte Bondone (Trento), Italy
<http://gis.fem-environment.eu> - neteler AT cealp.it