



mundialis

The real Big Data

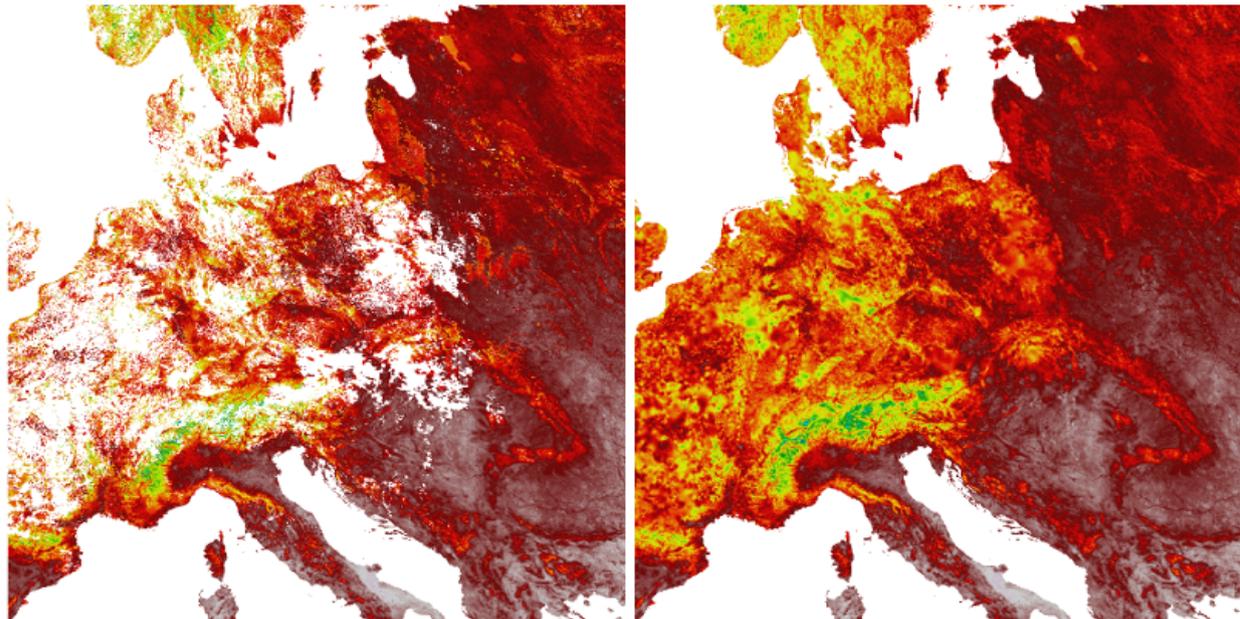
Potentiale eines satellitenbildgestützten Temperaturdatenarchivs

Till Adams & Markus Neteler & Markus Metz

mundialis GmbH & Co. KG

www.mundialis.de

FOSSGIS 2017, Passau

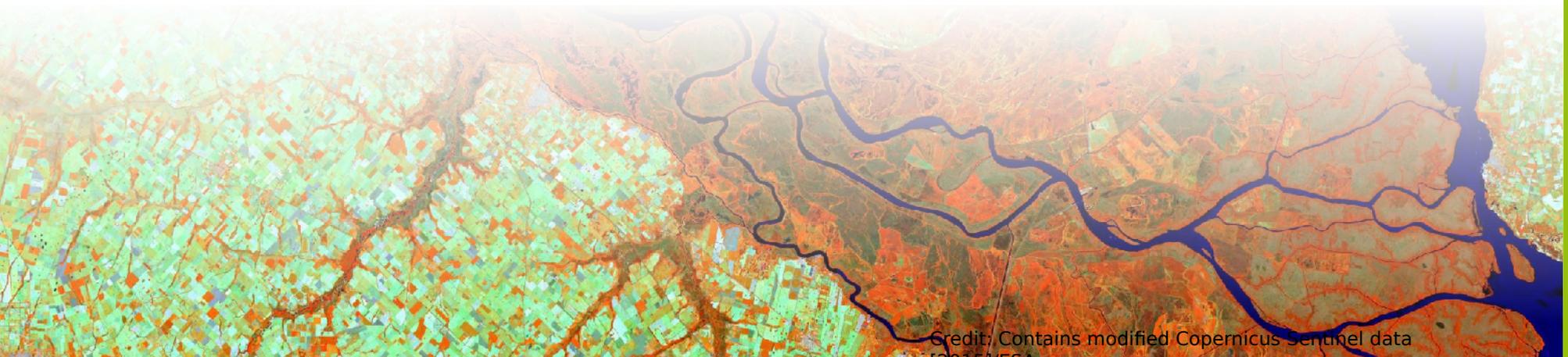




Wer sind wir?

mundialis GmbH & Co. KG

- gegründet in 2015 in Bonn durch T. Adams, H. Paulsen und M. Neteler
- 7 Mitarbeiter
- Massive GIS Daten Prozessierung und Fernerkundung
- Wir bieten jahrelange Erfahrung in Open Source GIS (insbesondere GRASS GIS Entwicklung)
- Fundierte HPC Erfahrung durch Prozessierung von MODIS Land Surface Temperature : "EuroLST"
 - 15 Jahre lückenlose Tageswerte mit 250m Auflösung



Agenda

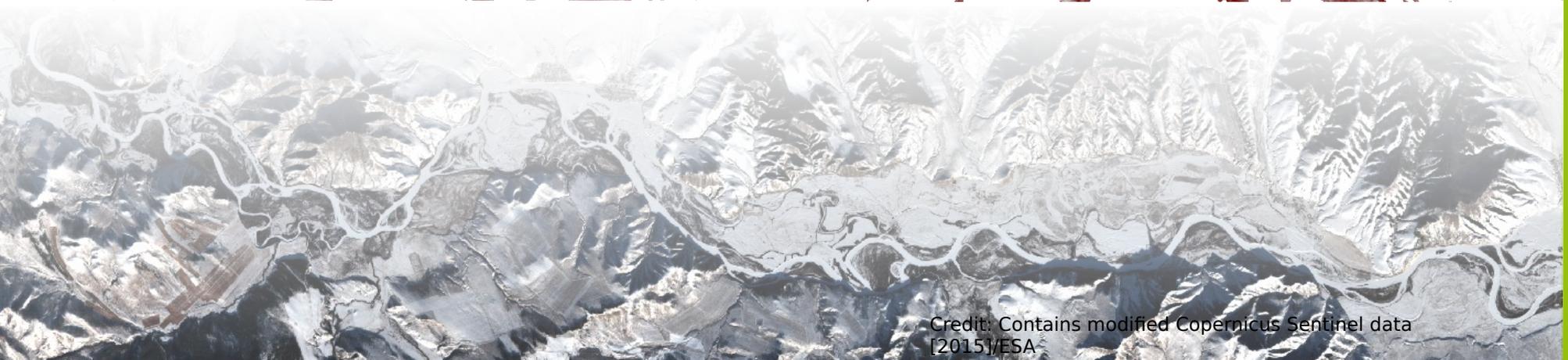
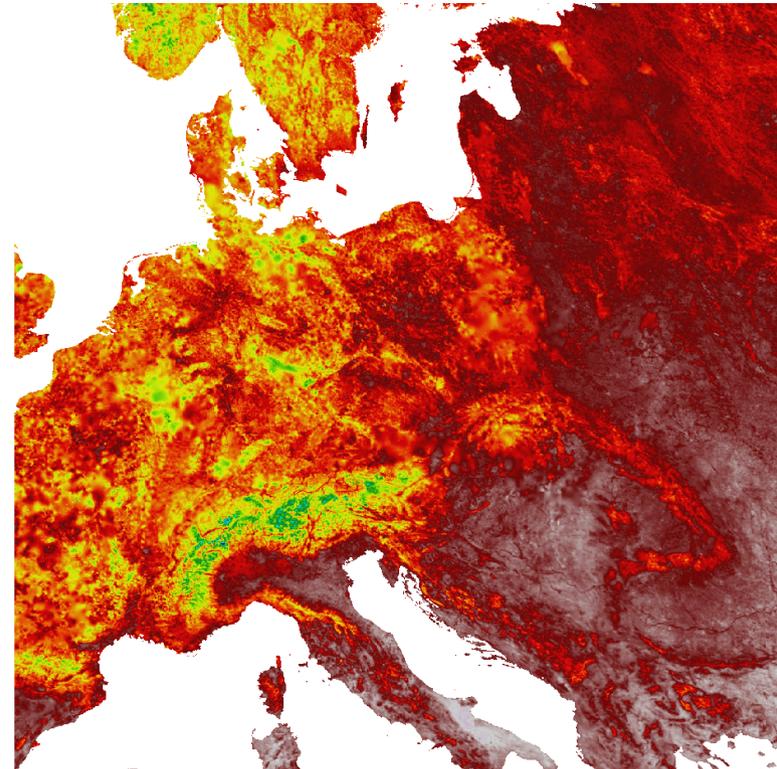
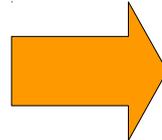
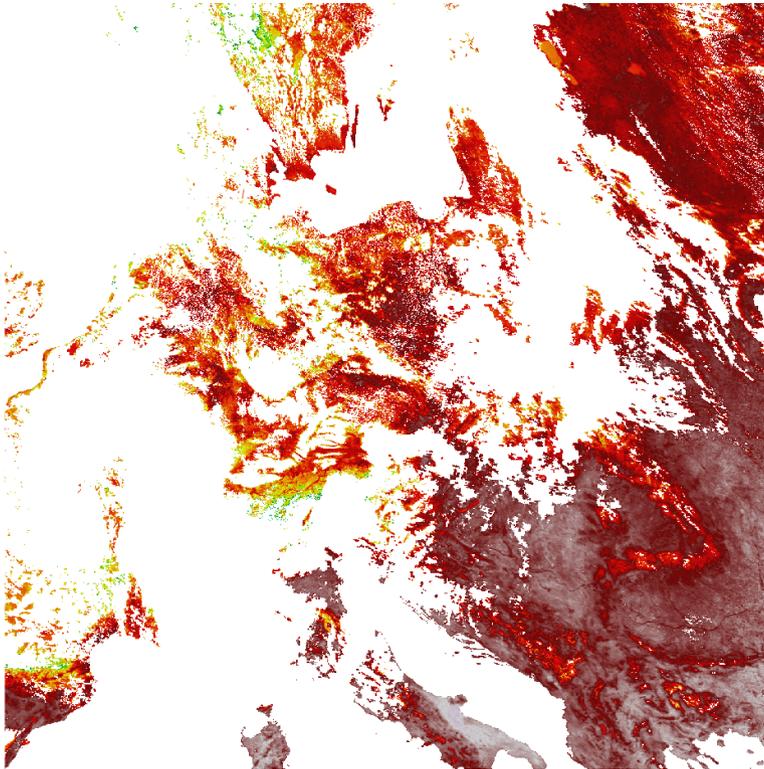
- Worum geht es?
- The “real Big Data”
- Wie verarbeitet man diese Datenmengen?
 - Infrastruktur
 - Software-Setup
 - Verarbeitungsschritte
- Anwendungsfelder für diese Daten
 - Beispiel: Detektion städtische Hitze-Inseln
 - Beispiel: Verbreitungsgebiete Krankheitsüberträger
 - Beispiel: Landwirtschaft & Weinbau (Gradzahltag, Spätfrost)
 - Beispiel: Infrastruktur

Worum geht es?



mundialis

Vervollständigung von Temperaturdatensätzen



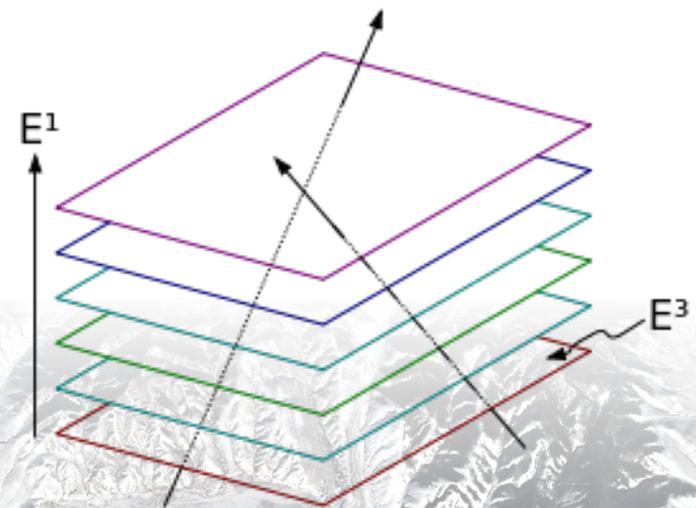
Credit: Contains modified Copernicus Sentinel data [2015]/ESA



Worum geht es?

Raum / Zeit - Datenarchiv

- Prozessierung von MODIS Land Surface Temperature Daten (MODIS LST) um ein
- 15 Jahre zurückreichendes, lückenloses Raum-Zeit-Temperaturdatenarchiv aufzubauen
- Es wird 4 Tageswerte mit 250m Auflösung geben
- permanente Fortschreibung mit aktuellen Daten



Von MovGP0 - Eigenes Werk, CC BY-SA 3.0 de,
<https://commons.wikimedia.org/w/index.php?curid=17734656>

Worum geht es?

MODIS LST

- **M**oderate Resolution **I**maging **S**ensor (MODIS) ist ein Instrument an Bord der Terra und Aqua Satelliten der NASA
- MODIS LST Produkte sind in verschiedenen Auflösungen als Freie Daten erhältlich
- Die Daten beinhalten immer Tag & Nachttemperaturwerte von 2 Satelliten ($2 \times 2 = 4$)
- Und wo ist das Problem?
→ die Wolken!



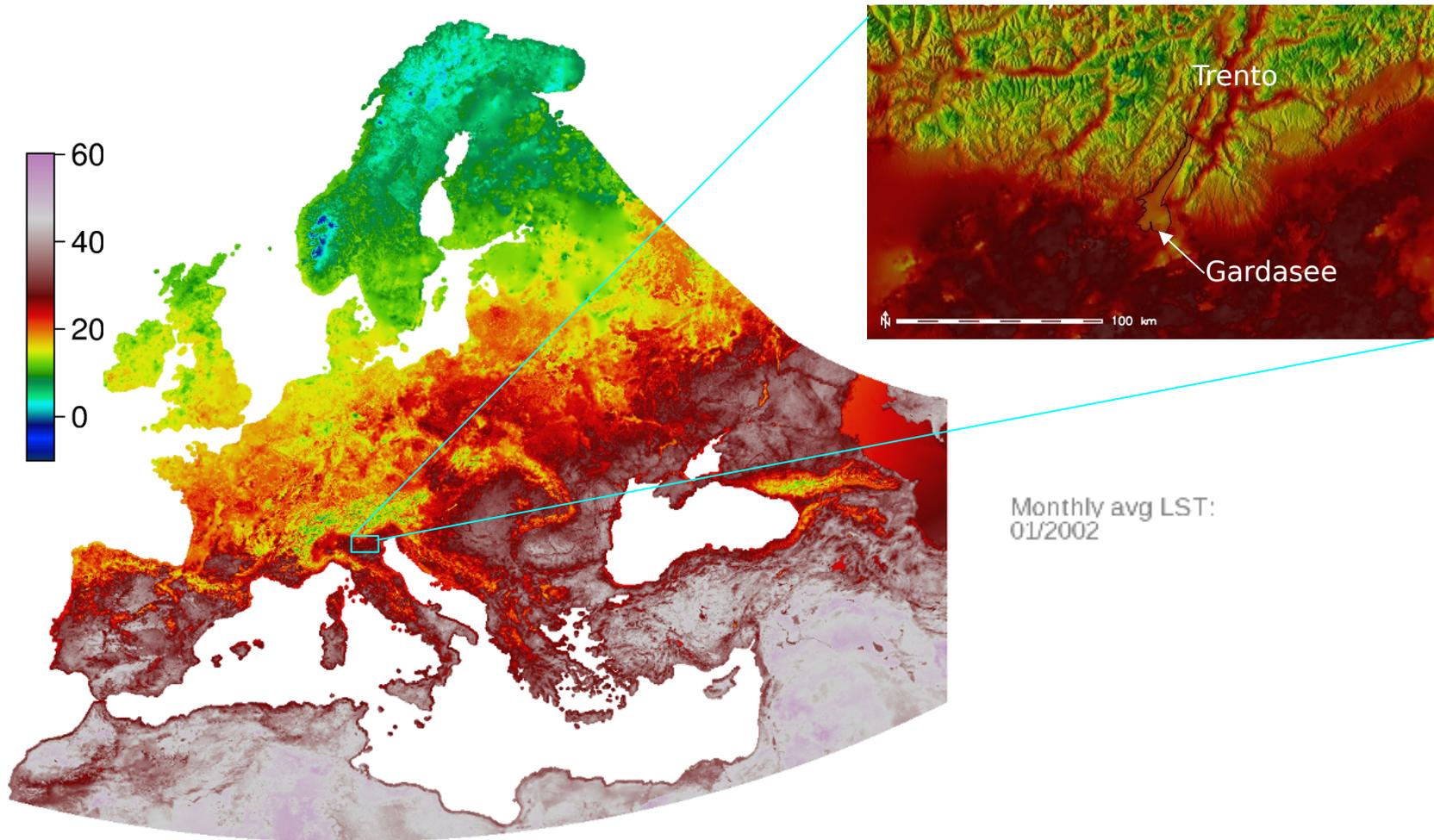
By Simon Eugster --Simon 14:02, 5 July 2006 (UTC) (Own work) [GFDL (<http://www.gnu.org/copyleft/fdl.html>), CC-BY-SA-3.0 (<http://creativecommons.org/licenses/by-sa/3.0/>) or CC BY-SA 2.5-2.0-1.0 (<http://creativecommons.org/licenses/by-sa/2.5-2.0-1.0/>)], via Wikimedia Commons



Worum geht es?

Raum / Zeit - Datenarchiv

“EuroLST”: MODIS LST tägliche Zeitreihen



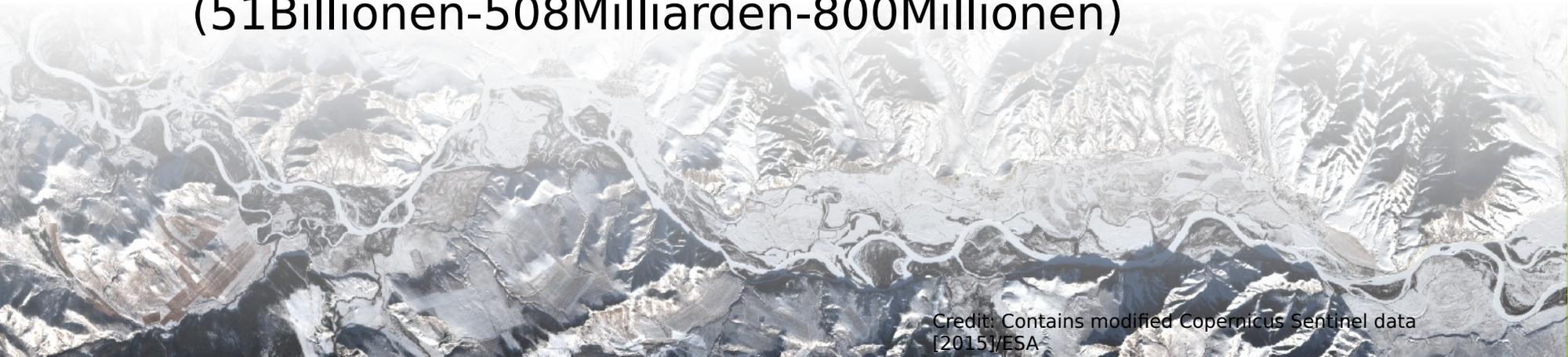
Monthly avg LST:
01/2002

1000 km
N - - - - -



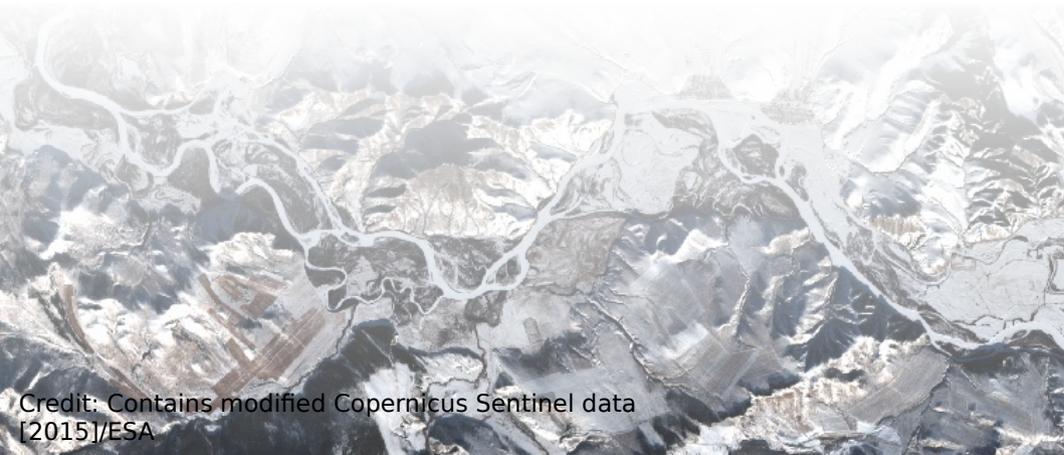
The real Big Data

- $15a * 365t * 4 = 21.900$ Datensätze
- Beispiel: BRD
 - das bedeutet am Beispiel der Fläche der Bundesrepublik (357.375 km^2) $5.718.016$ Pixel pro Datensatz
 - $21.900 \times 16 \times 357.375 = 1.252225 * 10^{11}$ Px
(= $125.224.550.400$ Pixel)
- Abdeckung anvisiert: weltweit
 - $147.000.000 \text{ km}^2 \times 16 \times 21.900 = 5.15088 * 10^{13}$ Px
(= $51.508.800.000.000$ Pixel)
 - (51 Billionen-508 Milliarden-800 Millionen)



The real Big Data

- In Byte & Terabyte
 - Zentraleuropa (4 MODIS Tiles) → 5 GB Rohdaten/Jahr
 - $15 \times 5 \text{ GB} = 75 \text{ GB}$ Rohdaten
 - Import für 1 Jahr dauert ca. 6H
 - Die Welt hat 286 MODIS Tiles → 357,5 GB Rohdaten/Jahr
 - Finales Produkt
 - Zentraleuropa für 15a ~ 795 GB
 - Welt für 15a ~ 57 TB



By George Chernilevsky (Own work) [Public domain], via Wikimedia Commons

Wie verarbeitet man diese Datenmengen?



mundialis

- Infrastruktur

es wird kein Rechenzentrum (mehr) benötigt:

- 2 Server mit je 8 Kernen und 32 GB RAM bilden ein HPC Cluster
- Abwägung zwischen Anzahl Server und Kosten (Beschränkung liegt in der I/O Kapazität eines Servers)
- Es wird ein Software-Job-Manager eingesetzt (Grid Engine)
- Jede Karte (Datensatz) wird als ein Job betrachtet, der Durchsatz beträgt aktuell (4 Tiles auf 1 CPU & 400 MB Ram) 9 Minuten Rechnerzeit pro Datensatz



Credit: Contains modified Copernicus Sentinel data [2015]/ESA

Wie verarbeitet man diese Datenmengen?

- Jede Karte (Datensatz) wird als ein Job betrachtet, der Durchsatz beträgt aktuell (4 Tiles auf 1 CPU & 400 MB Ram)
~9 Minuten pro Datensatz
- 21.900 x 9 Minuten = 3285h (136 Tage und ~21 Stunden)
- Parallelisierung durch 2 x 8 Cores → $3285h / 16 = \sim 205 h$
→ Idealvorstellung, in der Realität vermutlich 20-30% längere Rechenzeit

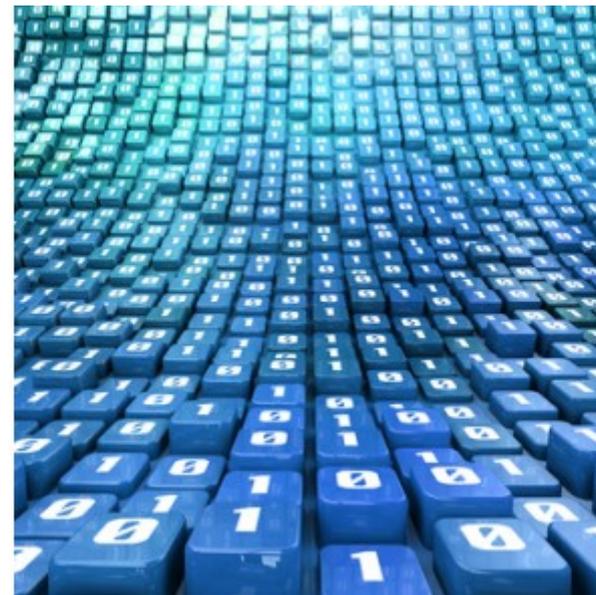
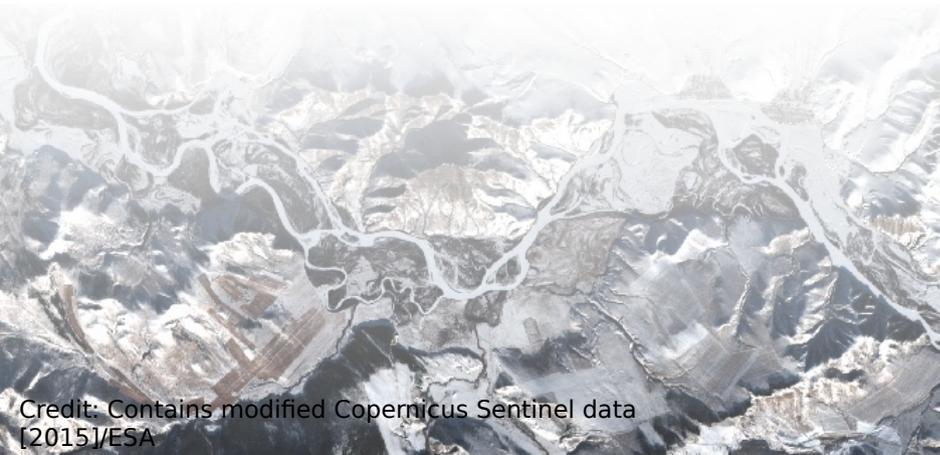
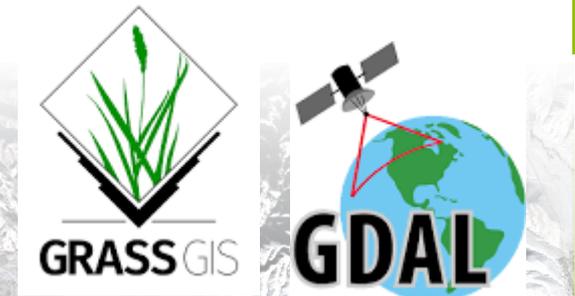


Foto: Dreaming Andy - Fotolia.com

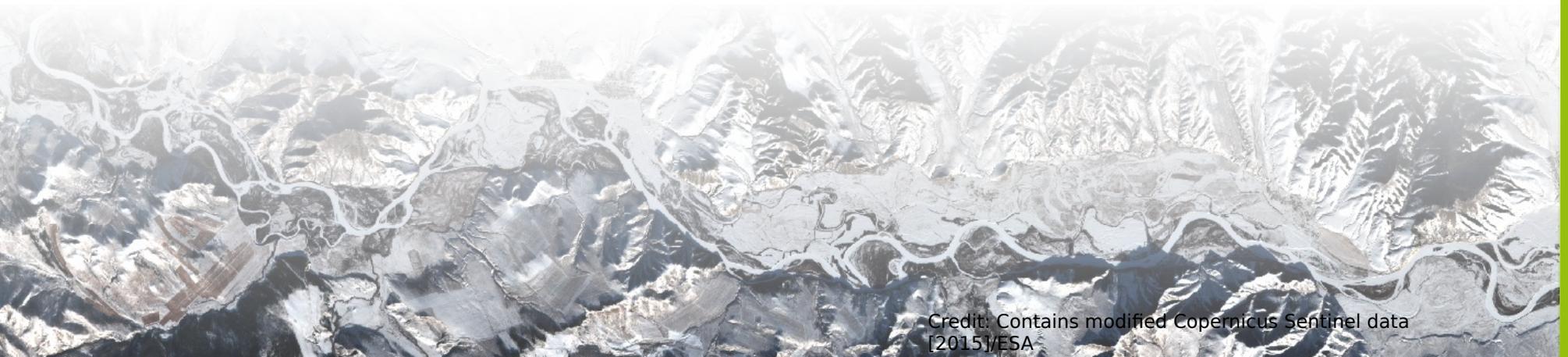
Wie verarbeitet man diese Datenmengen?

- Software-Setup
 - Import Rohdaten in HDF mit GDAL nach GRASS
 - Berechnung in GRASS GIS 7.2 mit Addons
 - Verwendet wird eine Kombination aus gewichtetem zeitlicher temporal Mittelwertberechnung mit statistischer Modellierung und räumlicher Interpolation.
 - derzeit Ablage als GeoTiff, Verwaltung in Datenbank (PostgreSQL/PostGIS) mit TGRASS (temporal framework)

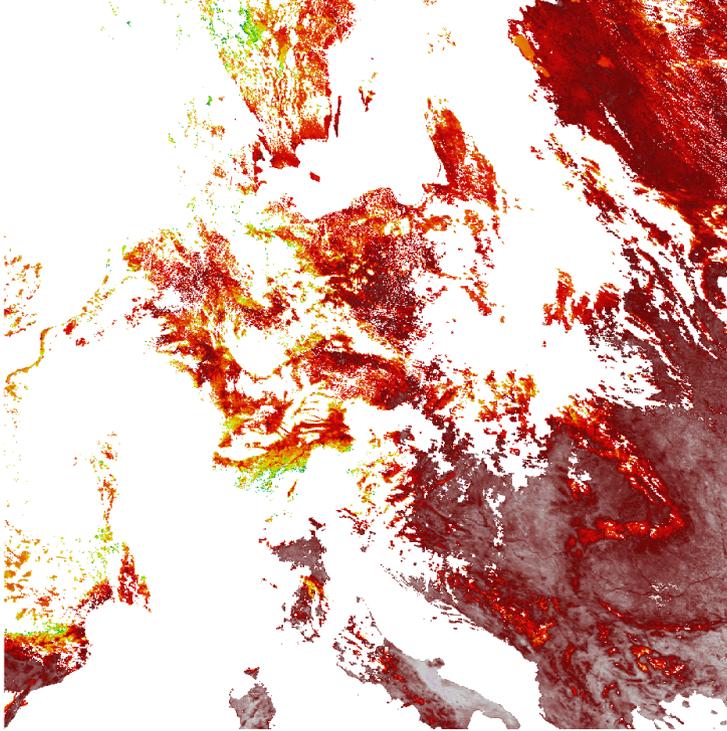


Wie verarbeitet man diese Datenmengen?

- Vereinfachte Arbeitsschritte:
 - zeitlich gewichtete (bis zu einer Woche) Wertezuweisung durch Betrachtung vorangegangener / nachfolgender Szenen
 - mit geostatistischen Methoden (Spline-Interpolation mit Co-Variable, hier: Höheninformation) werden noch vorhandene Lücken gefüllt
 - Ausreißerererkennung und -korrektur sowie Korrektur für Inlandsgewässer und urbane Flächen mit beim Überflug des Satelliten gemessenen Hilfsdaten



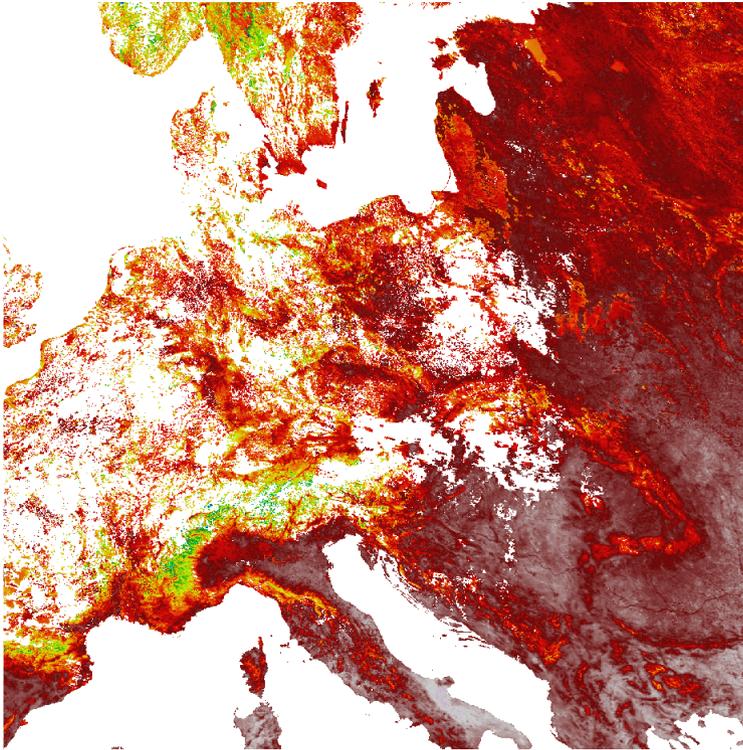
Grob vereinfachte Verarbeitungsschritte:



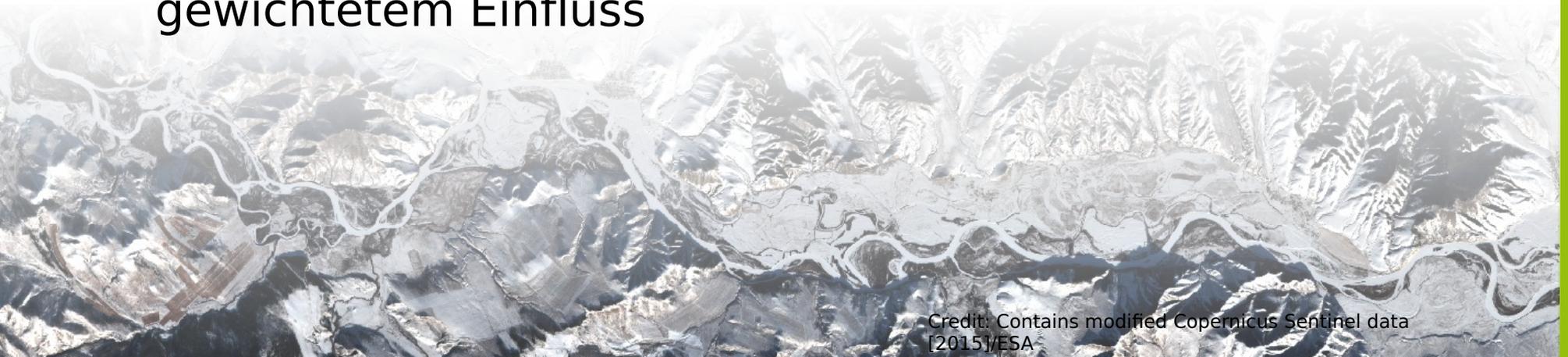
- Visualisierter Rohdatensatz (MODIS LST)
(Satelliten-Überflug um 13:30 am 2011-07-19)



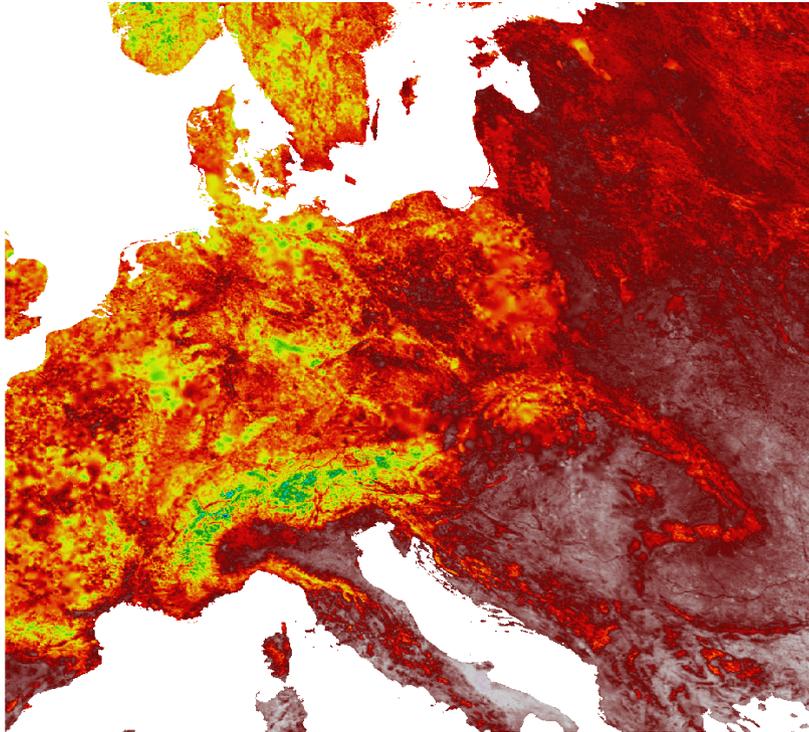
Grob vereinfachte Verarbeitungsschritte:



- Zwischenschritt: Daten-“Verdichtung” durch Verwendung zeitnaher Bilddaten aus gleicher Überflugsphase mit gewichtetem Einfluss



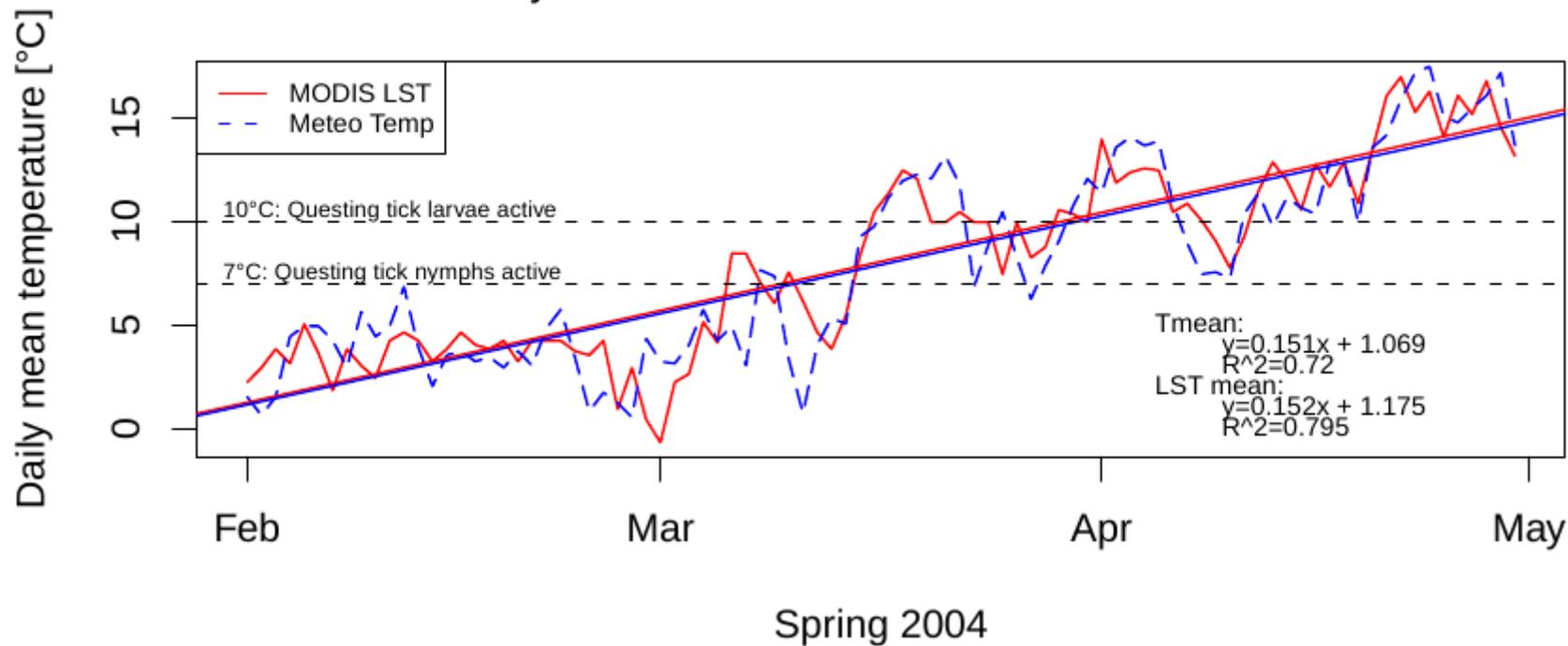
Grob vereinfachte Verarbeitungsschritte:



- Finaler Datensatz nach raum-zeitlicher Interpolation (Satelliten-Überflug um 13:30 am 2011-07-19)



Spring warming 2004, Trento Sud
daily means from MODIS LST versus Meteo stations



- Kein absoluter Vergleich zwischen Meteo-Station und MODIS LST (Messung 2m Höhe ↔ Bodentemperatur)
- Mittelungen sind fast identisch

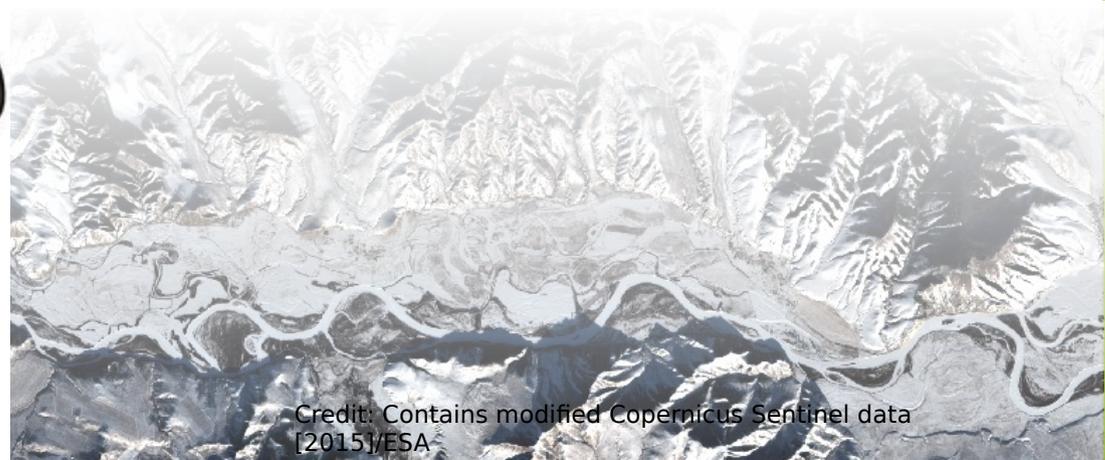


Anwendungsfelder für diese Daten

- Detektion städtischer Hitze-Inseln
 - “Hitzeinsel” = Orte, an denen im innerstädtischen Bereich die Temperatur erheblich über der der Umgebung liegt
 - Beeinträchtigung Lebensqualität
 - Gesundheitliche Folgen
 - Planerische Verminderung dieses Effektes durch Grünflächen, Frischluftschneisen, Freiflächen etc.
 - setzt Lokalisierung der Hitzeinseln voraus
 - globale Verschärfung mit zunehmender Verstädterung



k11714596 fotosearch.com



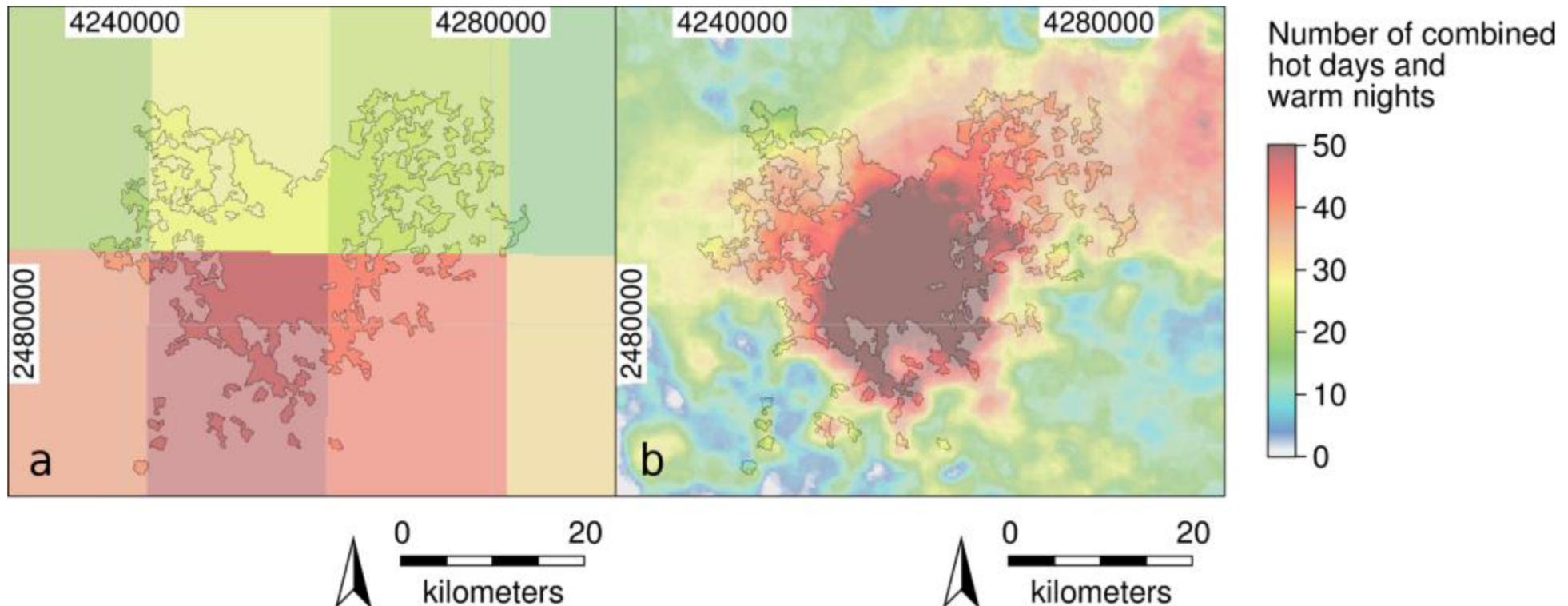
Credit: Contains modified Copernicus Sentinel data [2015]/ESA



Anwendungsfelder für diese Daten

Anzahl kombinierter heißen Tage ($> 35^\circ$) und warmen Nächten ($> 20^\circ$) in Mailand 2003 (Fläche: $\sim 1300 \text{ km}^2$)

- Links (a) auf Basis von Daten der ECAD (25km Auflösung), rechts (b) Equivalent basierend auf MODIS LST Daten



Anwendungsfelder für diese Daten

- Verbreitungsgebiete Krankheitserregern/Schädlingen
 - Verbreitung oft (auch) an Temperatur, insbesondere Frost gebunden
 - Beeinflussung direkt oder indirekt, da Temperatur die Verbreitung von Wirtstieren beeinflusst (z.B. Zecken)
 - Ausbreitungsursache ist meistens Globalisierung (Transport), aber Klimawandel ermöglicht es, das Schädlinge Fuß fassen



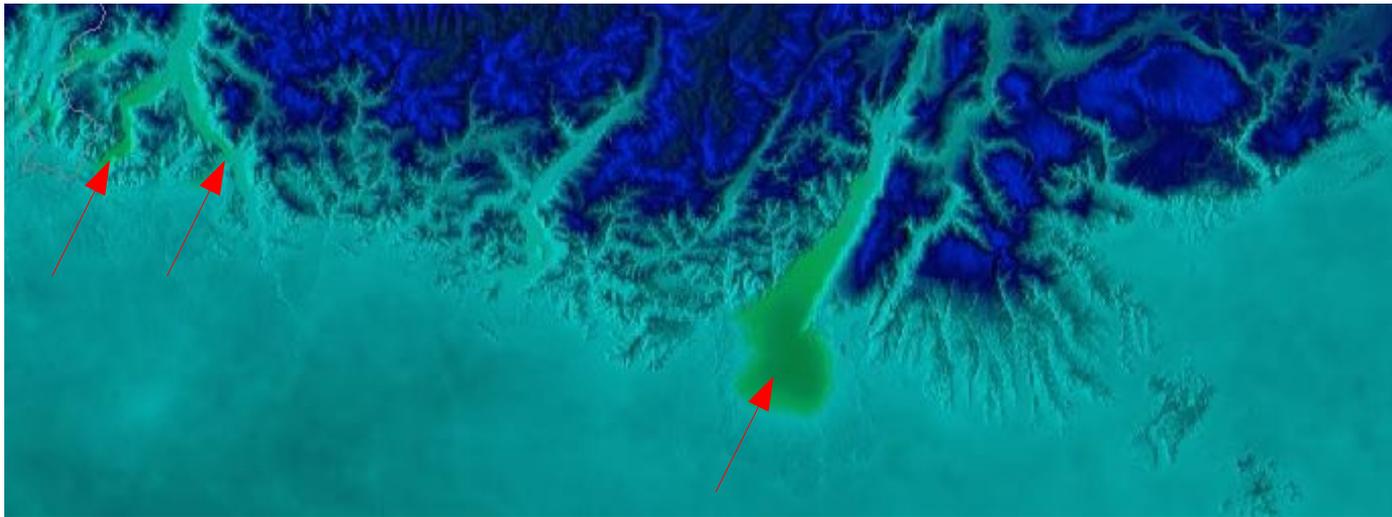
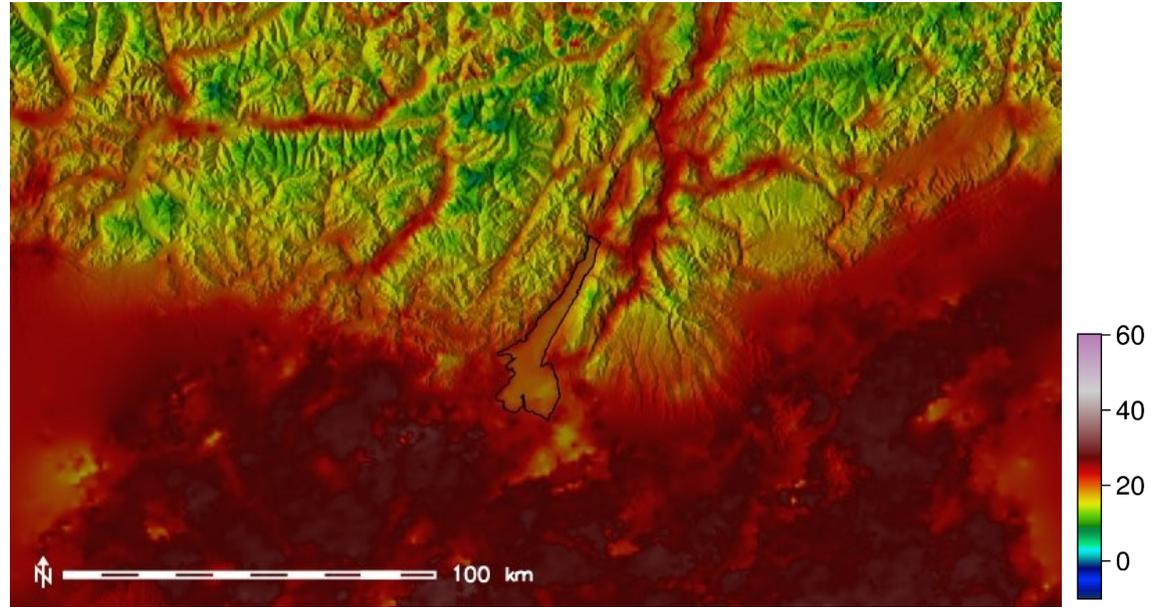
Von Sarefo - Eigenes Werk, CC BY-SA 3.0,
<https://commons.wikimedia.org/w/index.php?curid=2900877>



By James Gathany, CDC [Public domain], via Wikimedia Commons

Anwendungsfelder für diese Daten

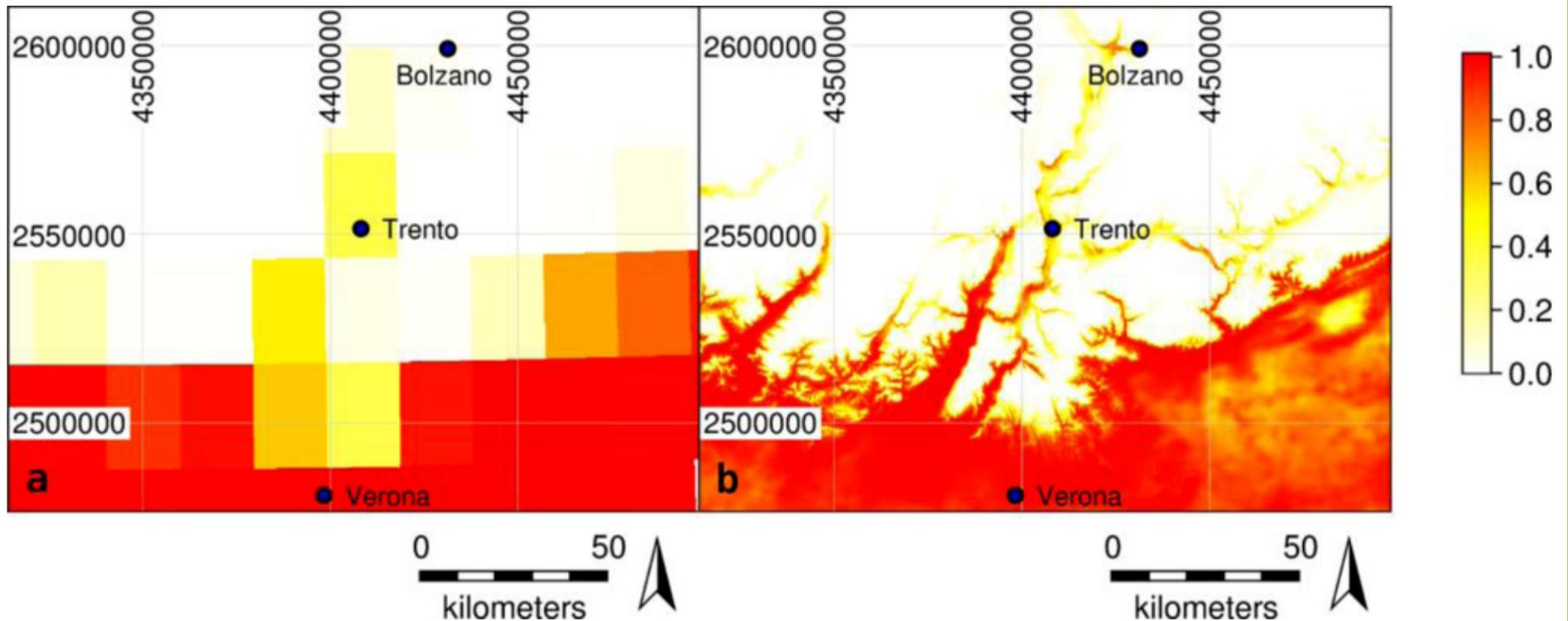
Das “heiße” Jahr 2003 (rechts) und die Auswirkungen im Januar 2004 (unten)



Gardasee noch “warm” → lokaler Heizeffekt
=> ermöglicht das Überwintern von Schädlingen

Anwendungsfelder für diese Daten

Potentielles Überleben von Tigermückenlarven nach dem Winter 2010/2011 in Norditalien



Links meist genutzte ECAD Daten,
rechts rekonstruierte MODIS LST Daten



Anwendungsfelder für diese Daten

- Landwirtschaft & Weinbau
 - Zeitreihenanalyse zeigt Temperaturveränderung als Trend in hoher räumlicher Auflösung
 - Veränderung Frost/Hitzetage
 - Detektion von Kaltluftseen
- Landwirtschaft
 - insbesondere Obstanbau (langfristig)
- Weinbau:
 - Rebsortenwahl
 - da wo möglich, oft Rebsorten an z.B. AOC gebunden



Quelle: <https://de.wikipedia.org/wiki/Weinbau>

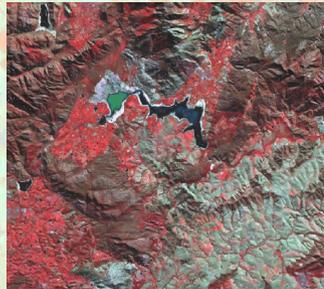
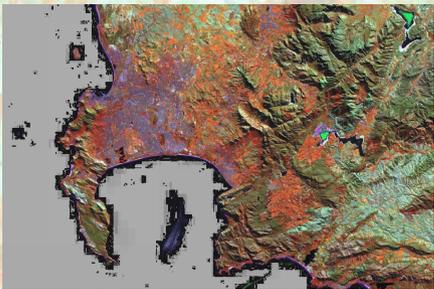


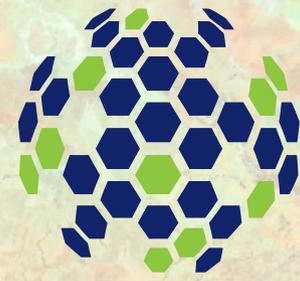
Anwendungsfelder für diese Daten

- langlebige Infrastruktur
 - Bei Verbau langlebiger Infrastruktur z.B. Straßenbelag
 - Straßenbelag muß auf künftig zu erwartende höhere Temperaturen ausgelegt sein
 - hohe Auflösung der Daten erlaubt genauere Analysen der Vergangenheit & Betrachtung von Trends
- Vorteil MODIS LST: Oberflächentemperatur statt 2m Höhe!



LBNL: mit Satelliten malen





mundialis

...Danke!

**Dr. Markus Neteler / Till Adams
mundialis GmbH & Co. KG
Kölnstraße 99
53111 Bonn, Germany**

**Email: neteler@mundialis.de
Web: <http://www.mundialis.de>**

