Gravity, Topography, Magnetics

Geoscience Data Analysis

in Spherical and Planar Geometry

Frederik J Simons

Princeton University

Christopher Harig
Princeton University

Kevin W. Lewis
Johns Hopkins University

Alain M. Plattner
Cal State Fresno

Theme

*“If we are to make progress in data analysis, as it is important that we should, we need to pay attention to our **tools** and our **attitudes**. If these are adequate, our **goals** will take care of themselves.”*

Tukey, 1962

Theme

"The greatest value of a ***picture*** *is when it forces us to notice what we never expected to see."*

Tukey, 1962

Goals

- Fast and flexible programming for **exploratory data analysis**

Goals

- Fast and flexible programming for **exploratory data analysis**
- **Partially** and **noisily** observed **scalar** and **vector** fields

Goals

- Fast and flexible programming for **exploratory data analysis**
- **Partially** and **noisily** observed **scalar** and **vector** fields
- Suitable for **time series**, in the **plane**, on the **sphere**, and in the **ball**

Goals

- Fast and flexible programming for **exploratory data analysis**
- **Partially** and **noisily** observed **scalar** and **vector** fields
- Suitable for **time series**, in the **plane**, on the **sphere**, and in the **ball**

Inverse problems, and **spectral-estimation** problems

Goals

- Fast and flexible programming for **exploratory data analysis**
- **Partially** and **noisily** observed **scalar** and **vector** fields
- Suitable for **time series**, in the **plane**, on the **sphere**, and in the **ball**

Inverse problems, and **spectral-estimation** problems

1. *Linear problems*: given the data, what is their **source**?

Goals

- Fast and flexible programming for **exploratory data analysis**
- **Partially** and **noisily** observed **scalar** and **vector** fields
- Suitable for **time series**, in the **plane**, on the **sphere**, and in the **ball**

Inverse problems, and **spectral-estimation** problems

1. *Linear problems*: given the data, what is their **source**?

$$\mathbf{d} = \mathcal{G}(\mathbf{m}) + \mathbf{n} \qquad \rightarrow \qquad \text{what is } \mathbf{m}?$$

Goals

- Fast and flexible programming for **exploratory data analysis**
- **Partially** and **noisily** observed **scalar** and **vector** fields
- Suitable for **time series**, in the **plane**, on the **sphere**, and in the **ball**

Inverse problems, and **spectral-estimation** problems

1. *Linear problems*: given the data, what is their **source**?

$$\mathbf{d} = \mathcal{G}(\mathbf{m}) + \mathbf{n} \qquad \rightarrow \qquad \text{what is } \mathbf{m}?$$

2. *Quadratic problems*: what is the **correlation structure** of the source?

Goals

- Fast and flexible programming for **exploratory data analysis**
- **Partially** and **noisily** observed **scalar** and **vector** fields
- Suitable for **time series**, in the **plane**, on the **sphere**, and in the **ball**

Inverse problems, and **spectral-estimation** problems

1. *Linear problems*: given the data, what is their **source**?

$$\mathbf{d} = \mathcal{G}(\mathbf{m}) + \mathbf{n} \qquad \rightarrow \qquad \text{what is } \mathbf{m}?$$

2. *Quadratic problems*: what is the **correlation structure** of the source?

$$\mathbf{d} = \mathcal{G}(\mathbf{m}) + \mathbf{n} \qquad \rightarrow \qquad \text{what is } \langle \mathbf{m}^{\mathrm{T}}\mathbf{m} \rangle?$$

Attitudes

- *Researchers*, from students to experts, demand and deserve **total control**

Attitudes

- *Researchers*, from students to experts, demand and deserve **total control**
- *Readers*, from colleagues to the layperson, require **reproducible research**

Attitudes

- *Researchers*, from students to experts, demand and deserve **total control**
- *Readers*, from colleagues to the layperson, require **reproducible research**
- *Skeptics* need the ability to question any and all **assumptions**

Attitudes

- *Researchers*, from students to experts, demand and deserve **total control**
- *Readers*, from colleagues to the layperson, require **reproducible research**
- *Skeptics* need the ability to question any and all **assumptions**
- *Coders* need **modularity**, **flexibility**, **scalability**, and **transparency**

Attitudes

- *Researchers*, from students to experts, demand and deserve **total control**
- *Readers*, from colleagues to the layperson, require **reproducible research**
- *Skeptics* need the ability to question any and all **assumptions**
- *Coders* need **modularity**, **flexibility**, **scalability**, and **transparency**
- When **data** change, so must the **model**, and the **interpretation** — *fast!*

Attitudes

- *Researchers*, from students to experts, demand and deserve **total control**
- *Readers*, from colleagues to the layperson, require **reproducible research**
- *Skeptics* need the ability to question any and all **assumptions**
- *Coders* need **modularity**, **flexibility**, **scalability**, and **transparency**
- When **data** change, so must the **model**, and the **interpretation** — *fast!*

An article about [...] science in a scientific publication is not the scholarship itself, it is merely advertising of the scholarship. The actual scholarship is the complete software development environment and the complete set of instructions which generated the figures.

Sussman & Wisdom, 2001

Tools

Computational algorithms are used to communicate precisely [...]. Expressing the methods [...] in a computer language forces them to be unambiguous and computationally effective. [...] Once formalized as a procedure, a mathematical idea becomes a tool that can be used directly to compute results.

Tools

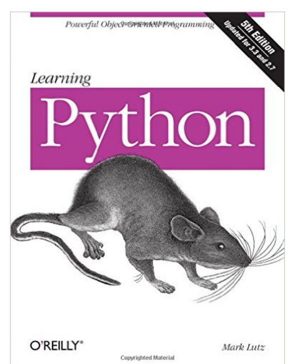
> *Computational algorithms are used to communicate precisely [...]. Expressing the methods [...] in a computer language forces them to be unambiguous and computationally effective. [...] Once formalized as a procedure, a mathematical idea becomes a tool that can be used directly to compute results.*

A growing number of tools lends itself to our goals.

Tools

Computational algorithms are used to communicate precisely [...]. Expressing the methods [...] in a computer language forces them to be unambiguous and computationally effective. [...] Once formalized as a procedure, a mathematical idea becomes a tool that can be used directly to compute results.

A growing number of tools lends itself to our goals.

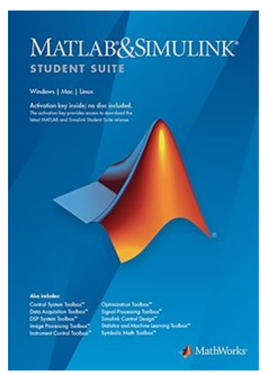


$50 book, free code

Tools

Computational algorithms are used to communicate precisely [...]. Expressing the methods [...] in a computer language forces them to be unambiguous and computationally effective. [...] Once formalized as a procedure, a mathematical idea becomes a tool that can be used directly to compute results.

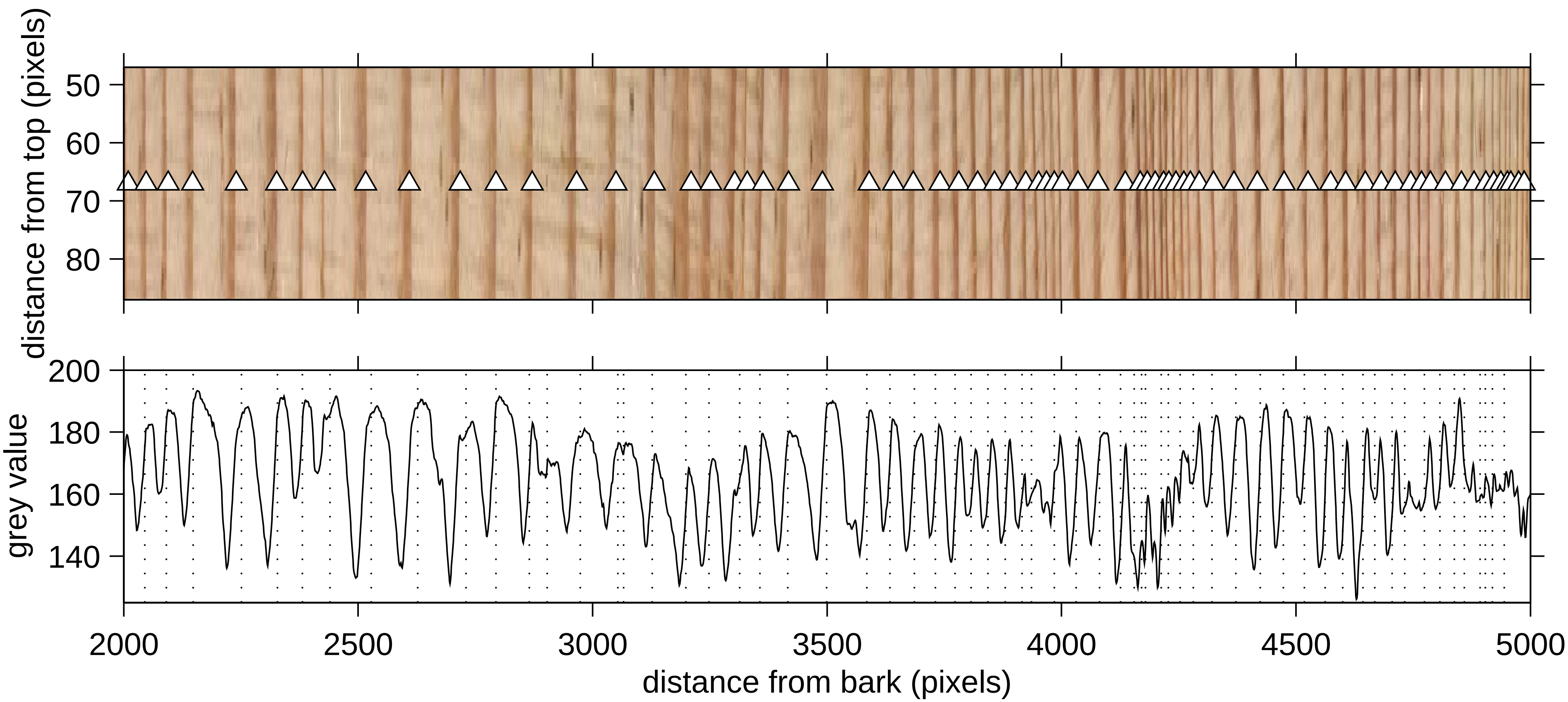
A growing number of tools lends itself to our goals.



$50 code, free book

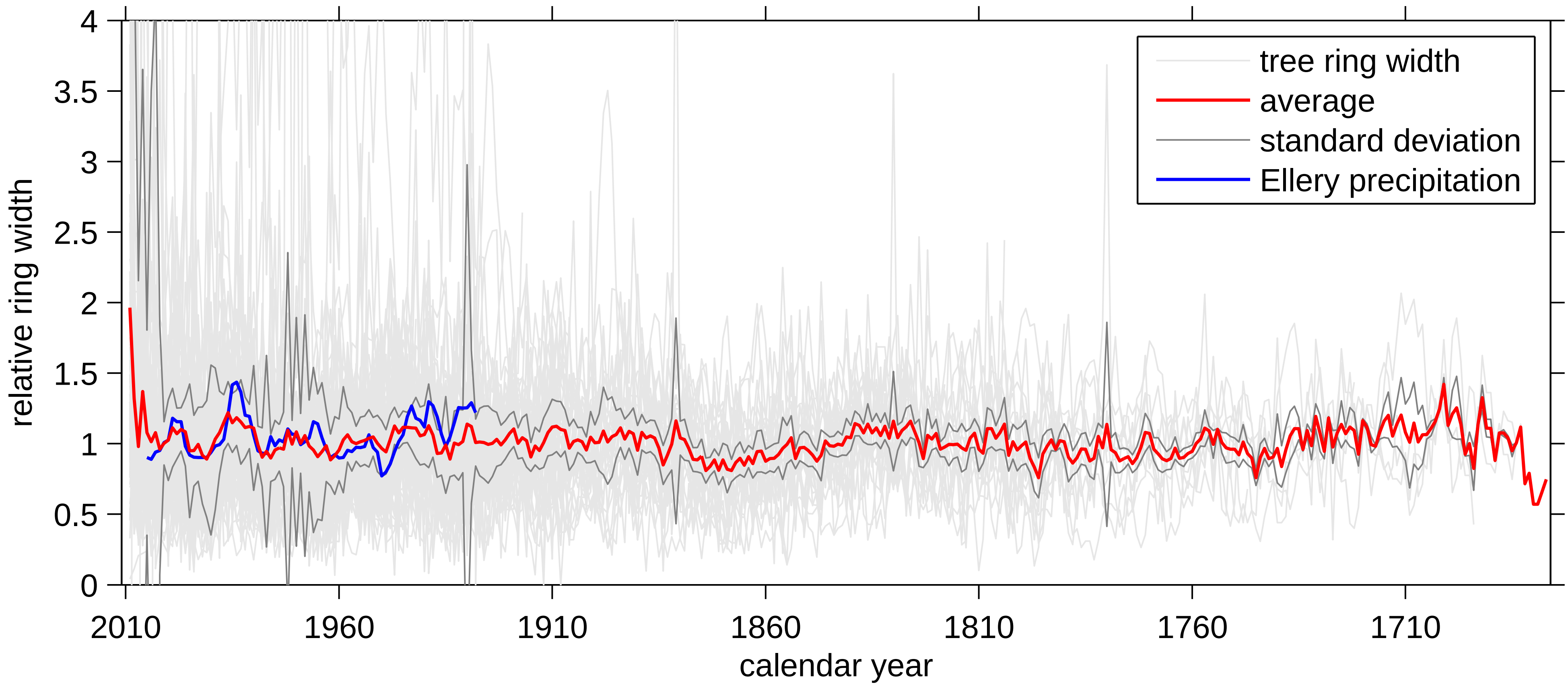
Fomel & Claerbout, 2009

Pictures — 1a

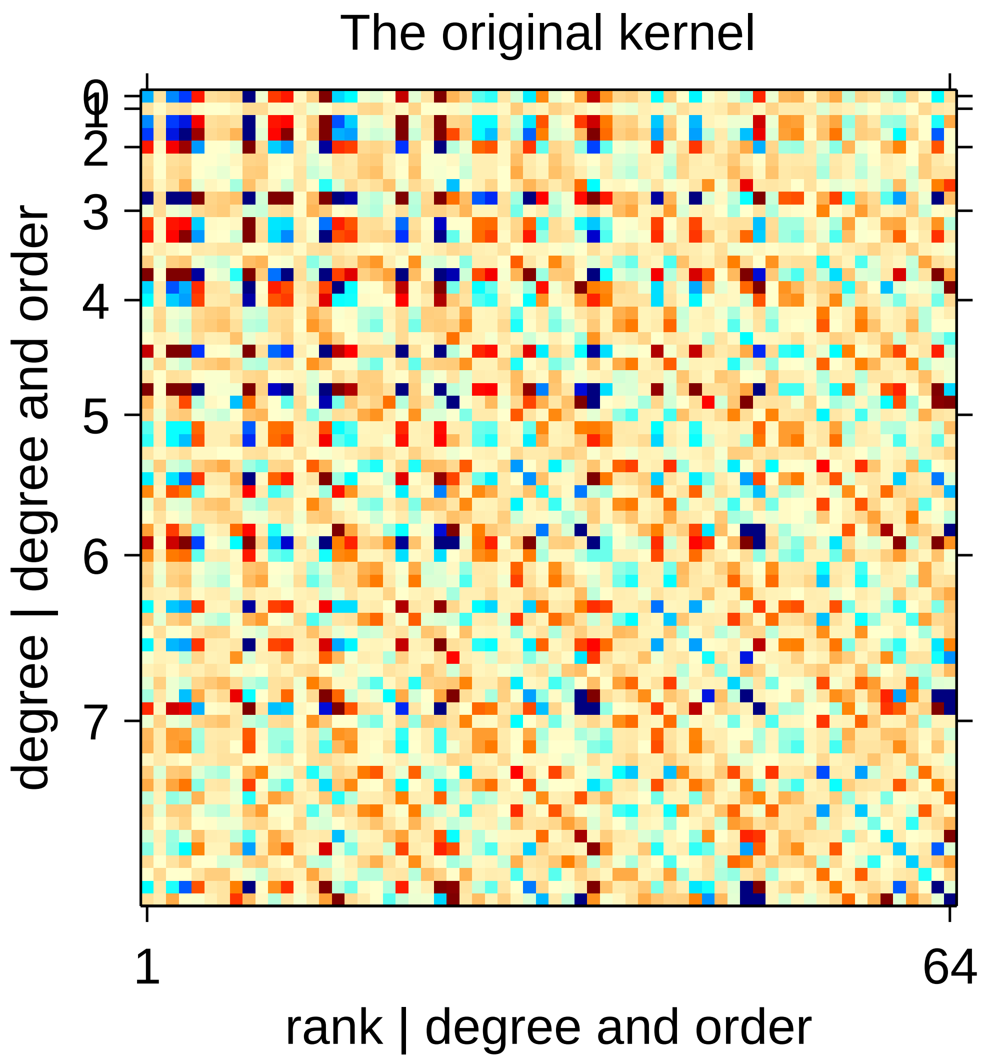


distance from top (pixels)
50
60
70
80
grey value
200
180
160
140
2000
2500
3000
3500
4000
4500
5000
distance from bark (pixels)

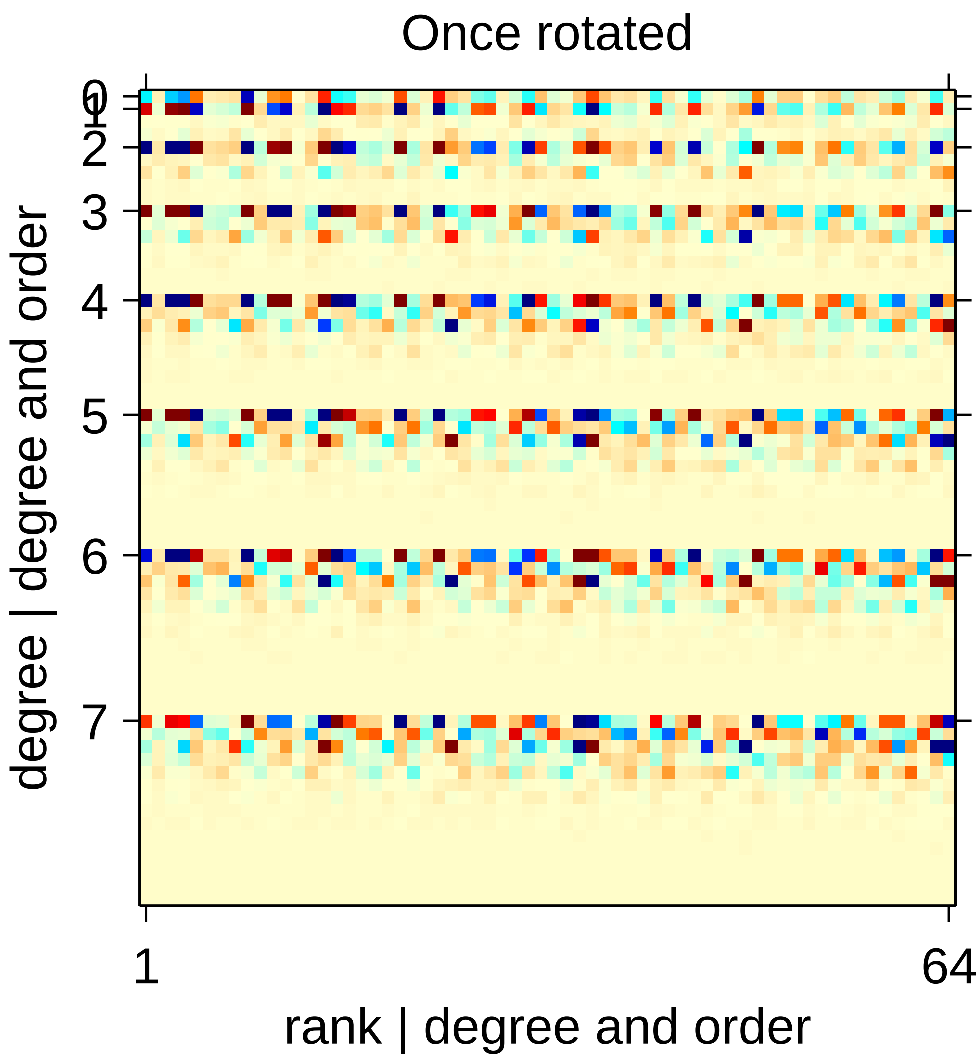
Pictures — 1b

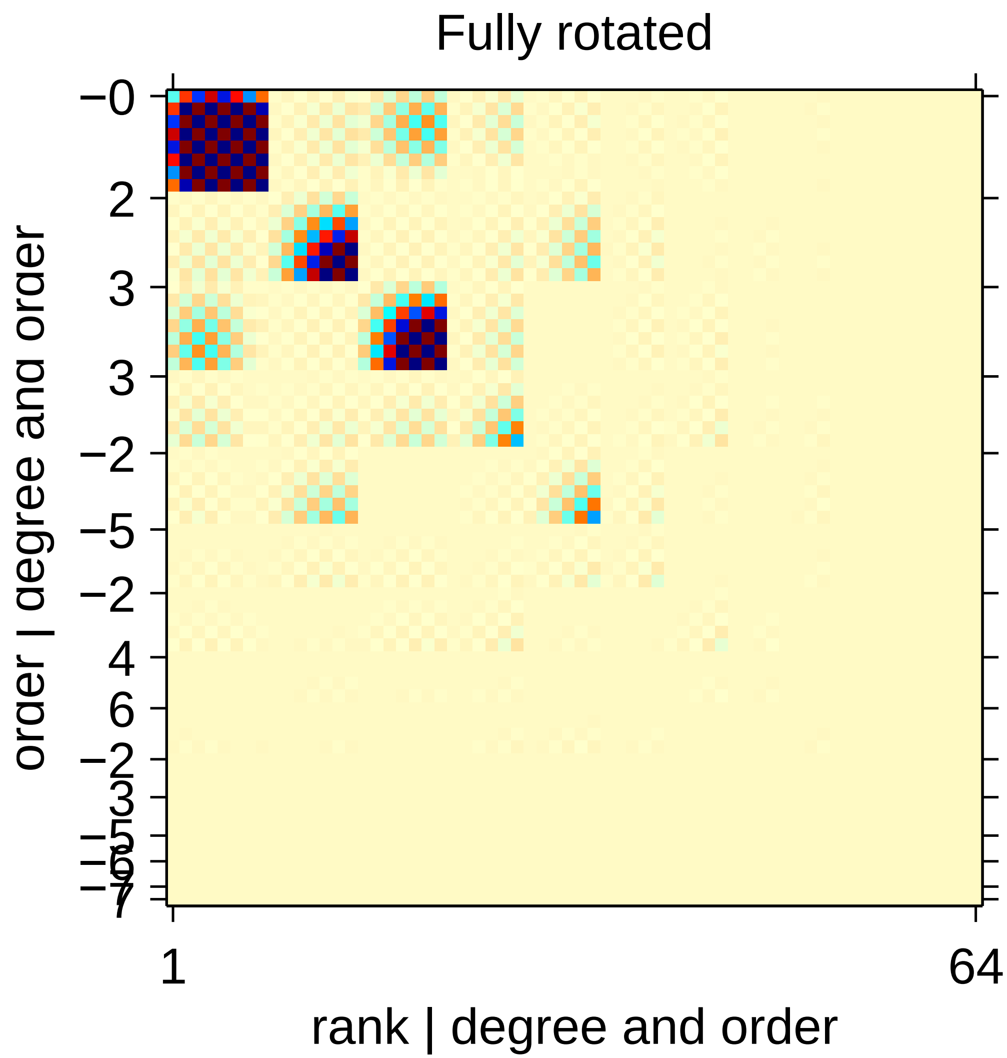


tree ring width
average
standard deviation
Ellery precipitation
relative ring width
0
0.5
1
1.5
2
2.5
3
3.5
4
2010
1960
1910
1860
1810
1760
1710
calendar year

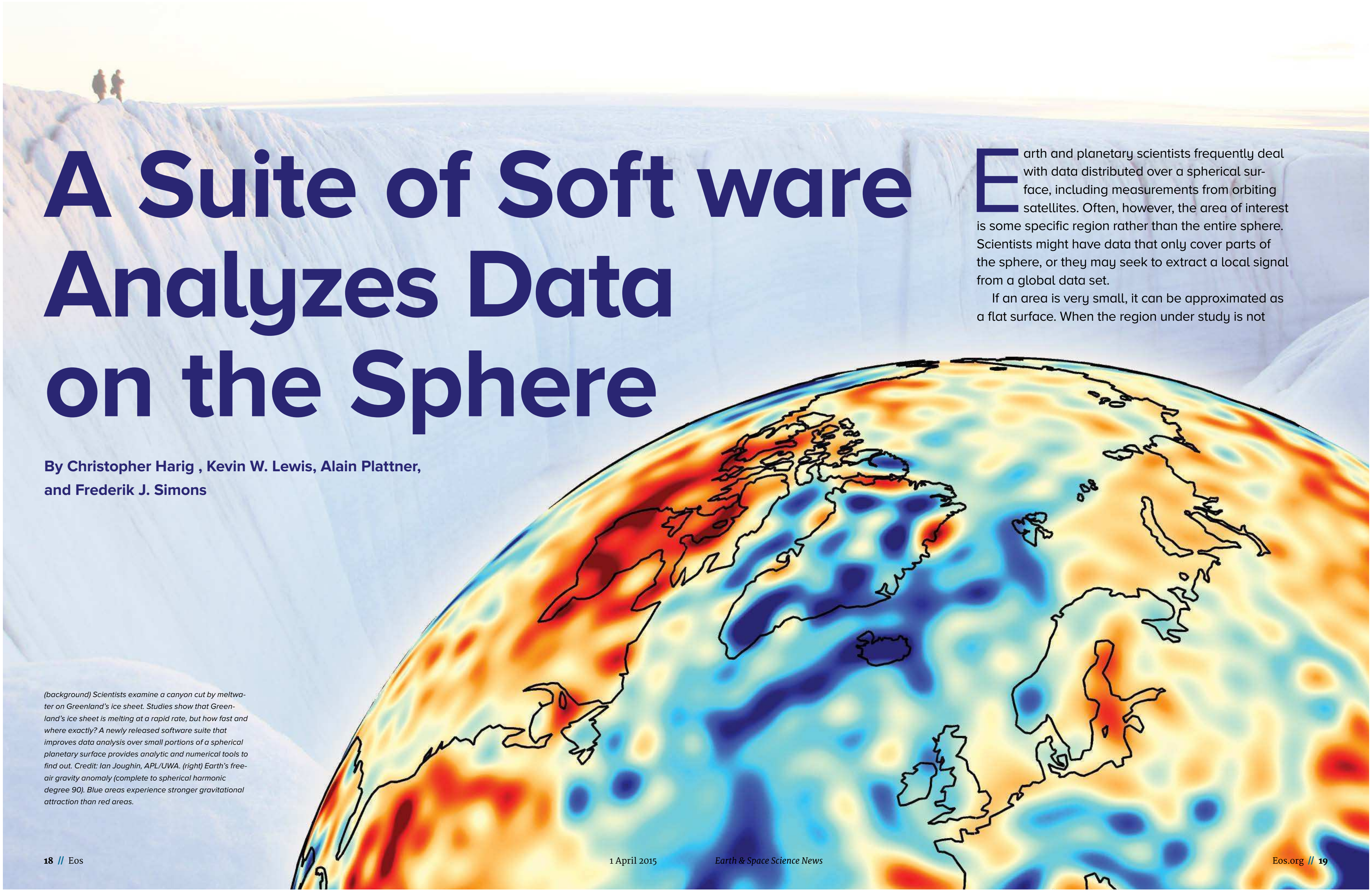

The original kernel
0
1
2
3
4
5
6
7
degree | degree and order
1
64
rank | degree and order

Pictures — 2b


Once rotated
0
1
2
3
4
5
6
7
degree | degree and order
1
64
rank | degree and order


Fully rotated
−0
2
3
3
−2
−5
−2
4
6
−2
1
64
rank | degree and order
order | degree and order

A suite of software: on `GitHub` and CSDMS



A Suite of Soft ware Analyzes Data on the Sphere

By Christopher Harig , Kevin W. Lewis, Alain Plattner, and Frederik J. Simons

Earth and planetary scientists frequently deal with data distributed over a spherical surface, including measurements from orbiting satellites. Often, however, the area of interest is some specific region rather than the entire sphere. Scientists might have data that only cover parts of the sphere, or they may seek to extract a local signal from a global data set.

If an area is very small, it can be approximated as a flat surface. When the region under study is not

(background) Scientists examine a canyon cut by meltwater on Greenland's ice sheet. Studies show that Greenland's ice sheet is melting at a rapid rate, but how fast and where exactly? A newly released software suite that improves data analysis over small portions of a spherical planetary surface provides analytic and numerical tools to find out. Credit: Ian Joughin, APL/UWA. (right) Earth's free-air gravity anomaly (complete to spherical harmonic degree 90). Blue areas experience stronger gravitational attraction than red areas.

Harig et al., EOS 2015

Spherical-harmonic **synthesis** and **analysis**: *(inversion)*

Spherical-harmonic **synthesis** and **analysis**: *(inversion)*

$$d(\mathbf{r}) = \sum_{lm} \left[s_{lm} + n_{lm} \right] Y_{lm}(\mathbf{r})$$

Spherical-harmonic **synthesis** and **analysis**: *(inversion)*

$$d(\mathbf{r}) = \sum_{lm} \left[s_{lm} + n_{lm} \right] Y_{lm}(\mathbf{r}) \qquad \leftarrow \qquad \text{estimate } s_{lm}$$

Spherical-harmonic **synthesis** and **analysis**: *(inversion)*

$$d(\mathbf{r}) = \sum_{lm} [s_{lm} + n_{lm}] \, Y_{lm}(\mathbf{r}) \qquad \leftarrow \qquad \text{estimate } s_{lm}$$

Here, $\mathbf{r}$ can be a discrete collection of scattered points. Estimate is *linear*.

Spherical-harmonic **synthesis** and **analysis**: *(inversion)*

$$d(\mathbf{r}) = \sum_{lm} \left[s_{lm} + n_{lm} \right] Y_{lm}(\mathbf{r}) \qquad \leftarrow \qquad \text{estimate } s_{lm}$$

Here, $\mathbf{r}$ can be a discrete collection of scattered points. Estimate is *linear*.

Linear combinations of spherical harmonics that maximize energy in a region: **Slepian functions**: *(optimization)*

Spherical-harmonic **synthesis** and **analysis**: *(inversion)*

$$d(\mathbf{r}) = \sum_{lm} [s_{lm} + n_{lm}]\, Y_{lm}(\mathbf{r}) \qquad \longleftarrow \qquad \text{estimate } s_{lm}$$

Here, $\mathbf{r}$ can be a discrete collection of scattered points. Estimate is *linear*.

Linear combinations of spherical harmonics that maximize energy in a region: **Slepian functions**: *(optimization)*

$$g_\alpha(\mathbf{r}) = \sum_{lm} g_{\alpha\, lm} Y_{lm}(\mathbf{r})$$

Spherical-harmonic **synthesis** and **analysis**: *(inversion)*

$$d(\mathbf{r}) = \sum_{lm} \left[s_{lm} + n_{lm} \right] Y_{lm}(\mathbf{r}) \qquad \longleftarrow \qquad \text{estimate } s_{lm}$$

Here, $\mathbf{r}$ can be a discrete collection of scattered points. Estimate is *linear*.

Linear combinations of spherical harmonics that maximize energy in a region: **Slepian functions**: *(optimization)*

$$g_\alpha(\mathbf{r}) = \sum_{lm} g_{\alpha\, lm} Y_{lm}(\mathbf{r}) \qquad \longleftarrow \qquad \text{find } g_{\alpha\, lm}, \text{ maximize } \frac{\int_R g_\alpha^2 \, d\Omega}{\int_\Omega g_\alpha^2 \, d\Omega}$$

Spherical-harmonic **synthesis** and **analysis**: *(inversion)*

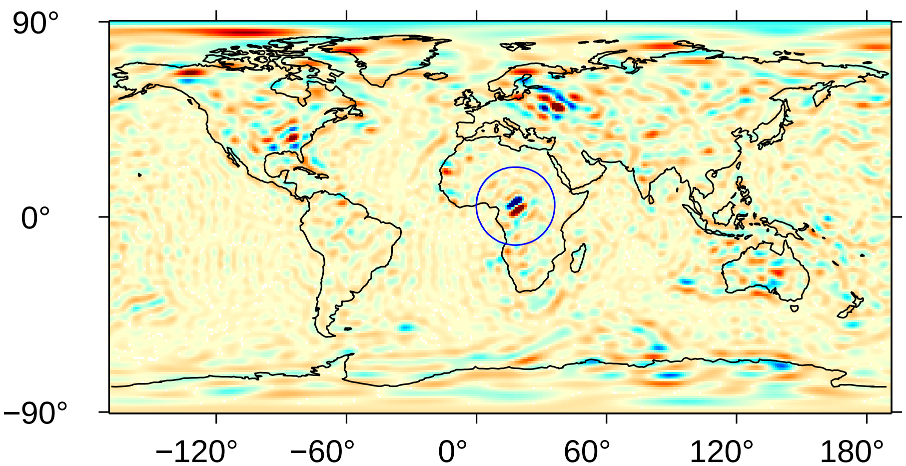
$$d(\mathbf{r}) = \sum_{lm} [s_{lm} + n_{lm}]\, Y_{lm}(\mathbf{r}) \qquad \leftarrow \qquad \text{estimate } s_{lm}$$

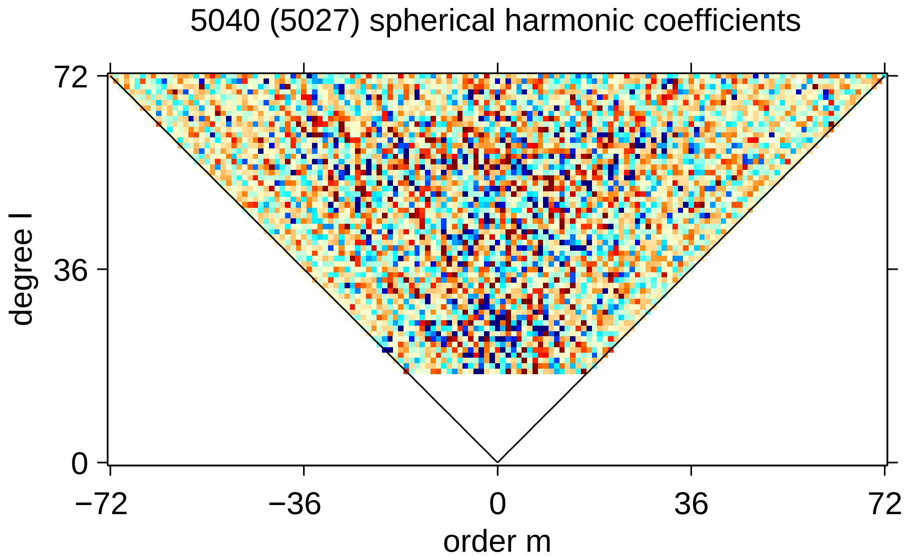
Here, $\mathbf{r}$ can be a discrete collection of scattered points. Estimate is *linear*.

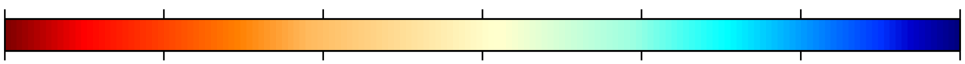
Linear combinations of spherical harmonics that maximize energy in a region: **Slepian functions**: *(optimization)*

$$g_\alpha(\mathbf{r}) = \sum_{lm} g_{\alpha\, lm} Y_{lm}(\mathbf{r}) \qquad \leftarrow \qquad \text{find } g_{\alpha\, lm}, \text{ maximize } \frac{\int_R g_\alpha^2 \, d\Omega}{\int_\Omega g_\alpha^2 \, d\Omega}$$

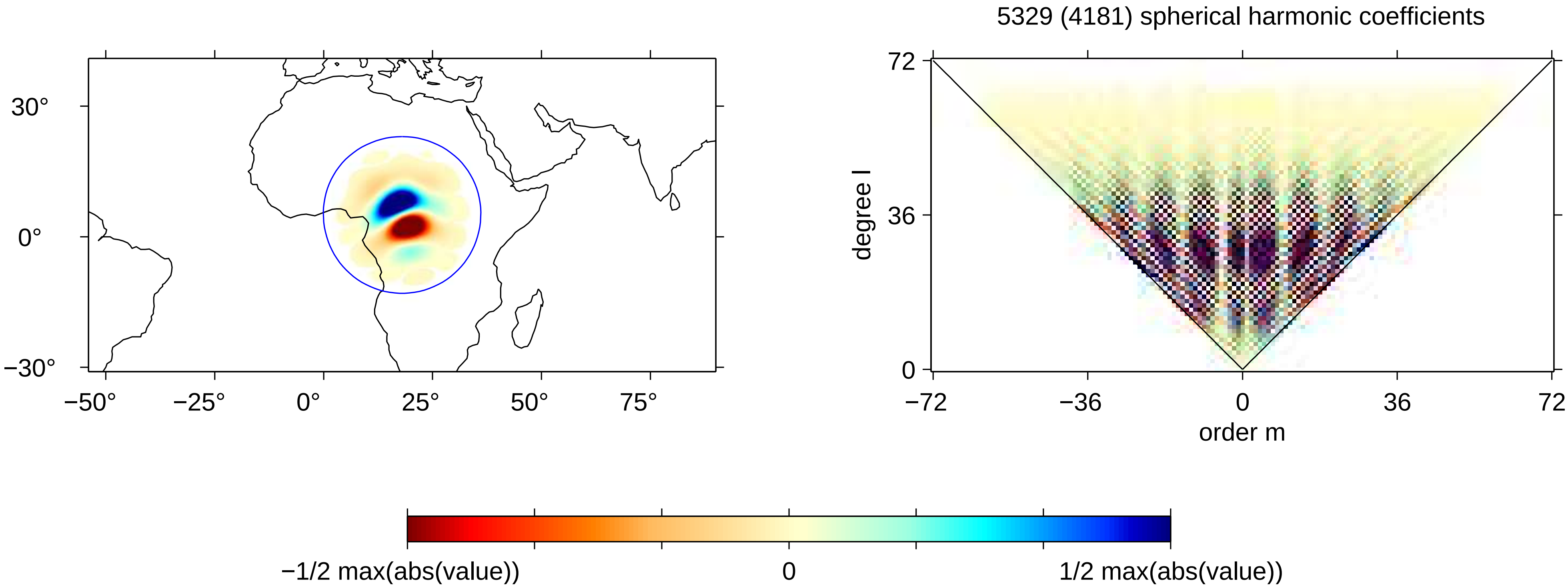
Here, R can be a region of arbitrary geographical description.



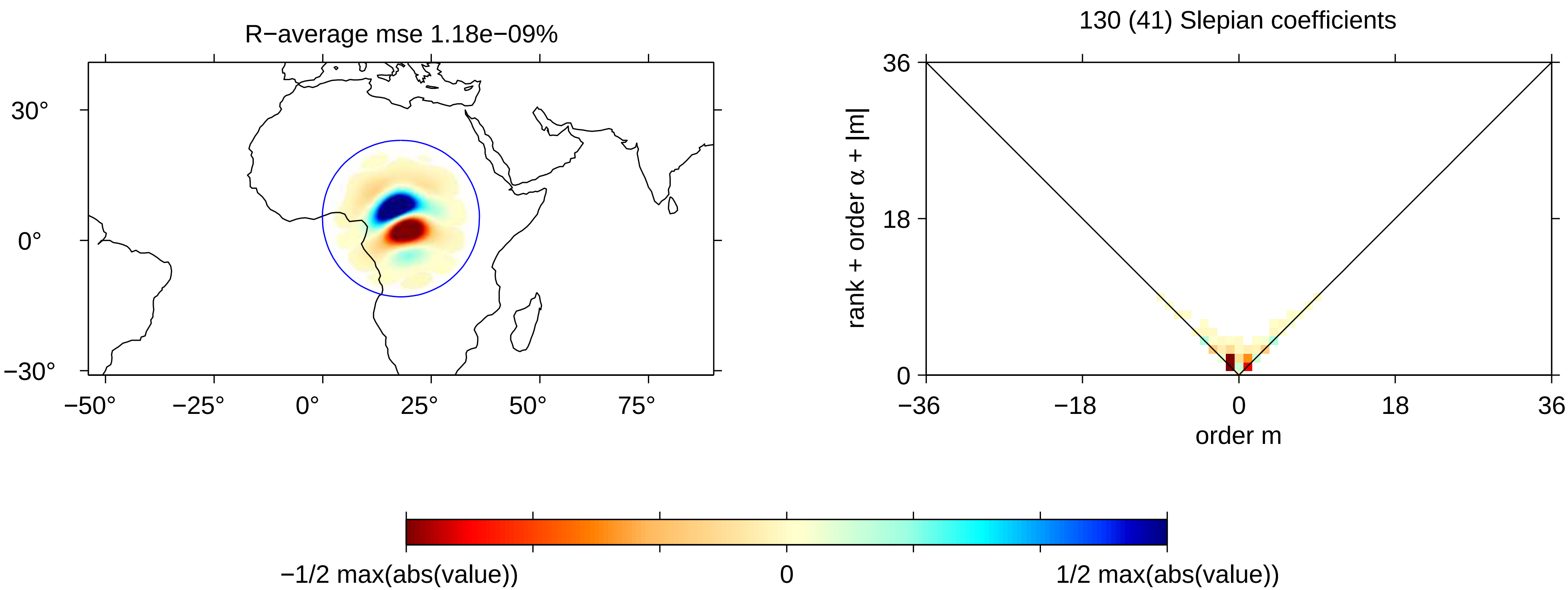




A *global* basis, **good** for *global* problems.



A *global* basis, **bad** for *local* problems.



A *local* basis, **good** for *local* problems. Sparsity!

Spherical-harmonic **power-spectral density estimation**:

Spherical-harmonic **power-spectral density estimation**:

$$d(\mathbf{r}) = \sum_{lm} \left[s_{lm} + n_{lm} \right] Y_{lm}(\mathbf{r})$$

Spherical-harmonic **power-spectral density estimation**:

$$d(\mathbf{r}) = \sum_{lm} [s_{lm} + n_{lm}]\, Y_{lm}(\mathbf{r}) \qquad \leftarrow \qquad \text{estimate } \mathrm{var}\{s_{lm}\} = S_l$$

Spherical-harmonic **power-spectral density estimation**:

$$d(\mathbf{r}) = \sum_{lm} [s_{lm} + n_{lm}]\, Y_{lm}(\mathbf{r}) \qquad \longleftarrow \qquad \text{estimate } \mathrm{var}\{s_{lm}\} = S_l$$

Here, $\mathbf{r}$ is a densely sampled set of points *locally available* within R.

The low-variance, low-bias *quadratic* **spectral estimate**
is of the **Thomson-Slepian-multitaper** variety:

Spherical-harmonic **power-spectral density estimation**:

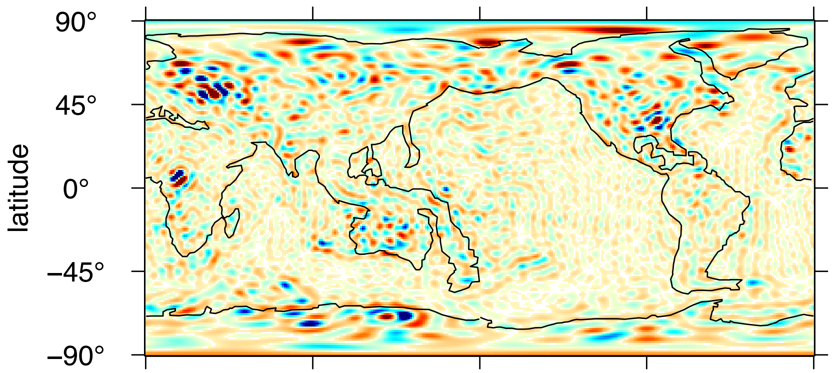
$$d(\mathbf{r}) = \sum_{lm} [s_{lm} + n_{lm}]\, Y_{lm}(\mathbf{r}) \qquad \leftarrow \qquad \text{estimate } \mathrm{var}\{s_{lm}\} = S_l$$

Here, $\mathbf{r}$ is a densely sampled set of points *locally available* within R.

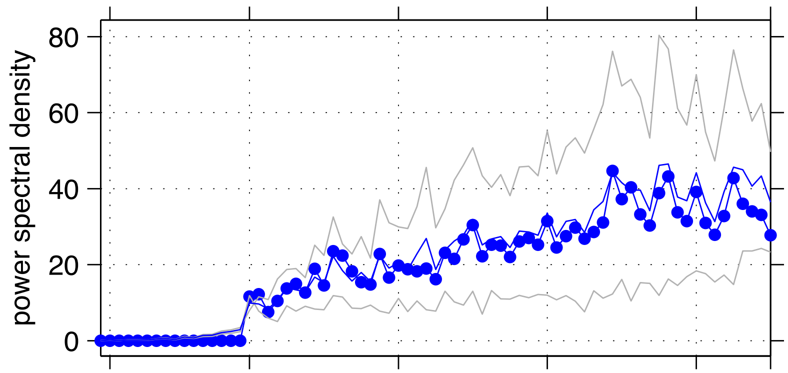
The low-variance, low-bias *quadratic* **spectral estimate**
is of the **Thomson-Slepian-multitaper** variety:

$$\hat{S}_l^{\mathrm{MT}} = \sum_{\alpha} \lambda_\alpha \left(\frac{1}{2l+1} \sum_m \left| \int_\Omega g_\alpha(\mathbf{r})\, d(\mathbf{r})\, Y_{lm}(\mathbf{r})\, d\Omega \right|^2 \right).$$

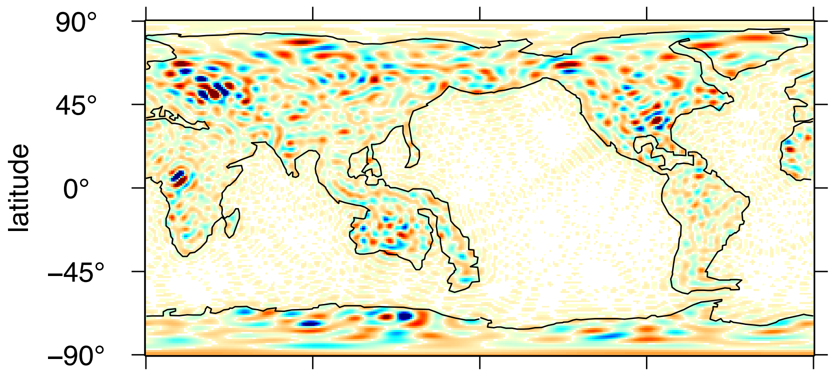
SLEPIAN_Charlie



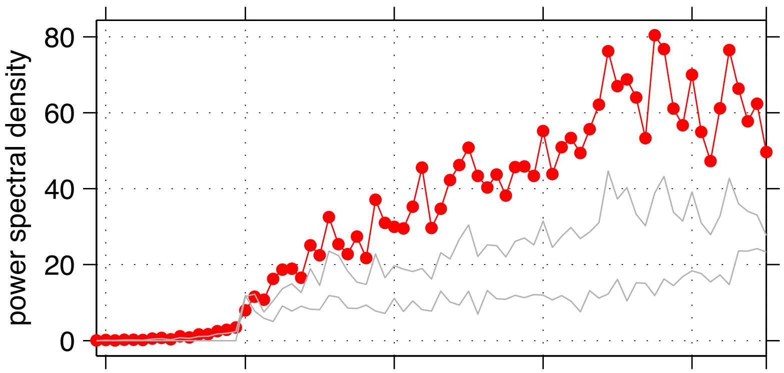
“Whole-sphere”



power spectral density

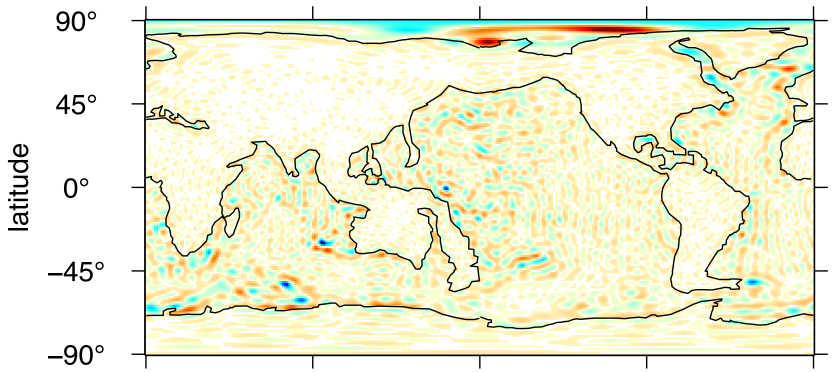


“Continents-only”

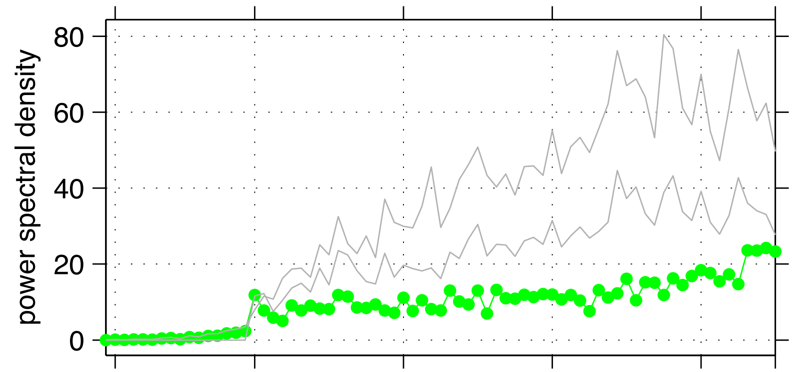


power spectral density

Beggan et al., GJI 2013



“Oceans-only”



power spectral density

Beggan et al., GJI 2013

SLEPIAN_Delta

Time-dependent potential-field estimation
from **noisy** and **incomplete** data:

Time-dependent potential-field estimation from **noisy** and **incomplete** data:

$$s(\mathbf{r},t) = \sum_{lm} [s_{lm} + n_{lm}](t)\, Y_{lm}(\mathbf{r})$$

Time-dependent potential-field estimation
from **noisy** and **incomplete** data:

$$s(\mathbf{r},t)=\sum_{lm}[s_{lm}+n_{lm}](t)\,Y_{lm}(\mathbf{r})\qquad\leftarrow\qquad\text{estimate } s(\mathbf{r},t)$$

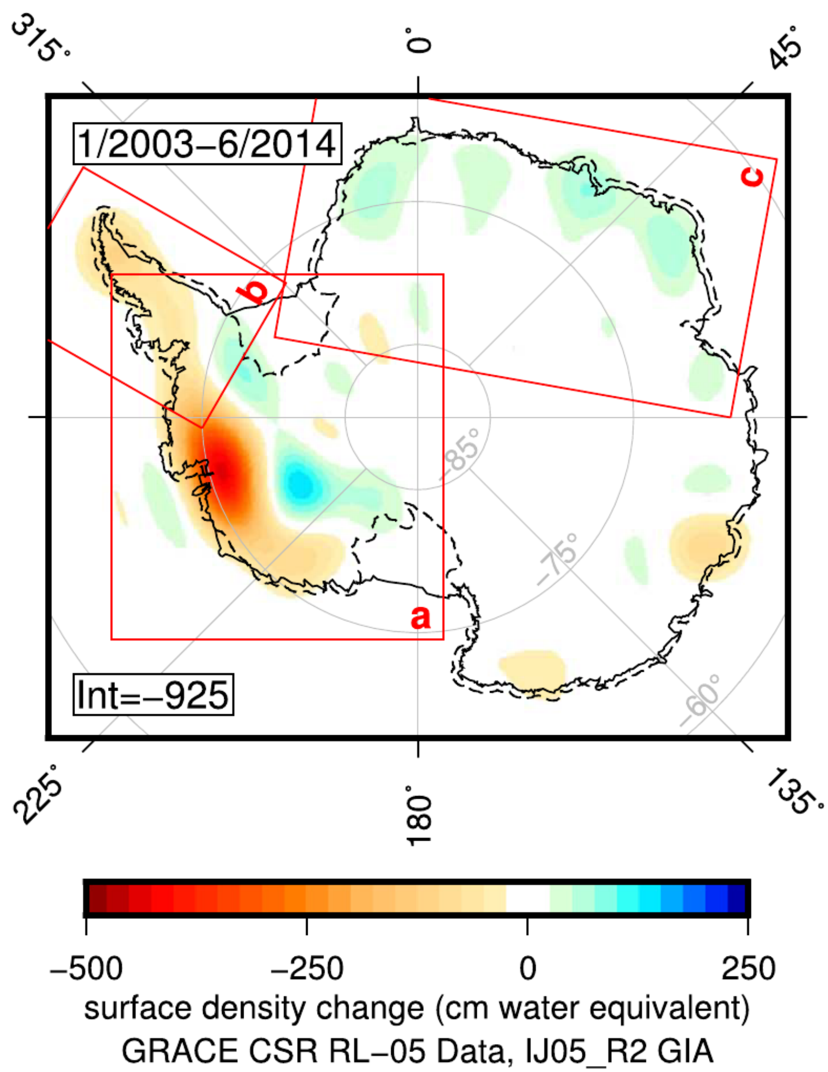
SLEPIAN_Delta

Time-dependent potential-field estimation
from **noisy** and **incomplete** data:

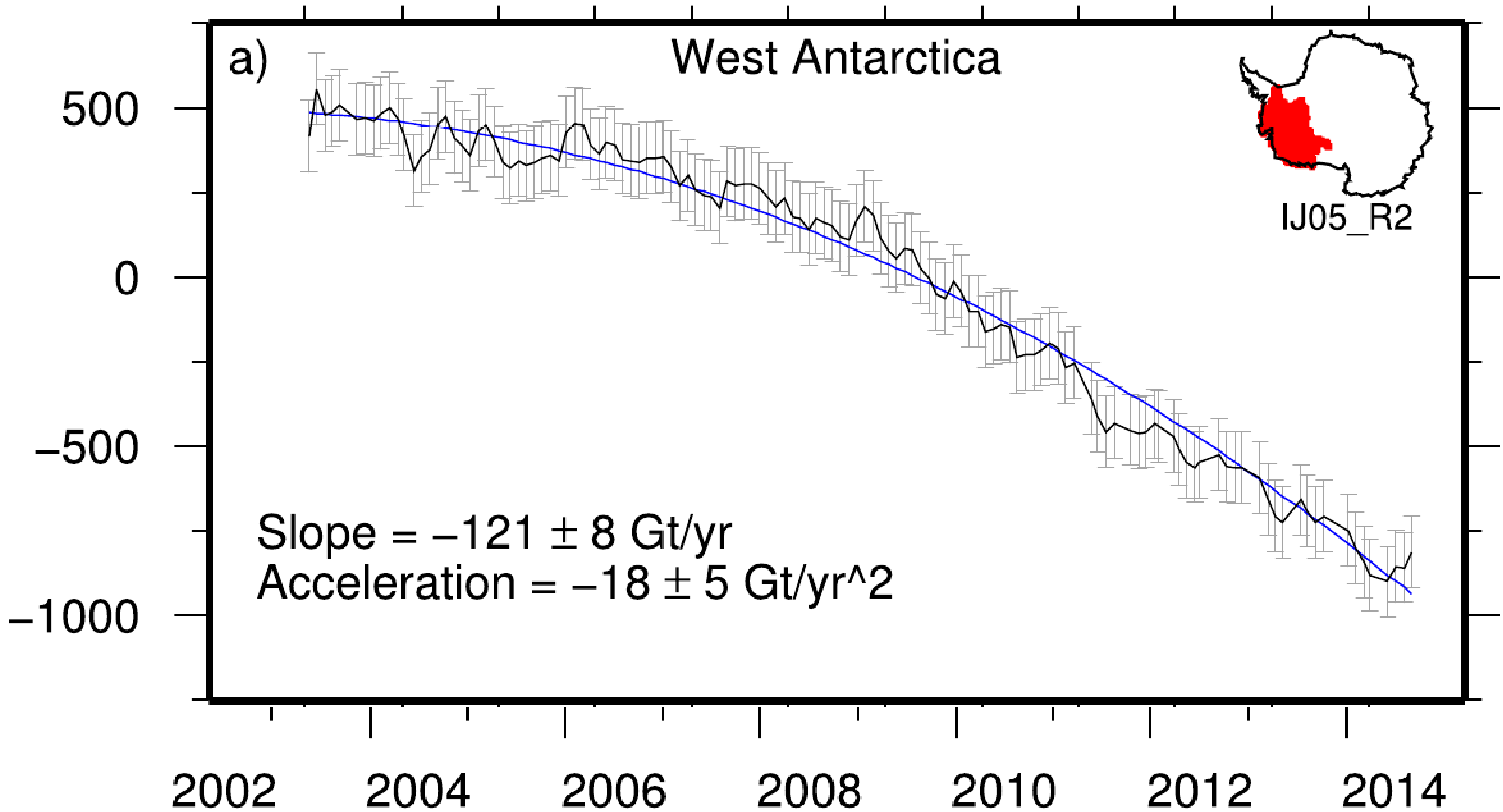
$$s(\mathbf{r},t) = \sum_{lm} [s_{lm} + n_{lm}](t)\, Y_{lm}(\mathbf{r}) \qquad \leftarrow \qquad \text{estimate } s(\mathbf{r},t)$$

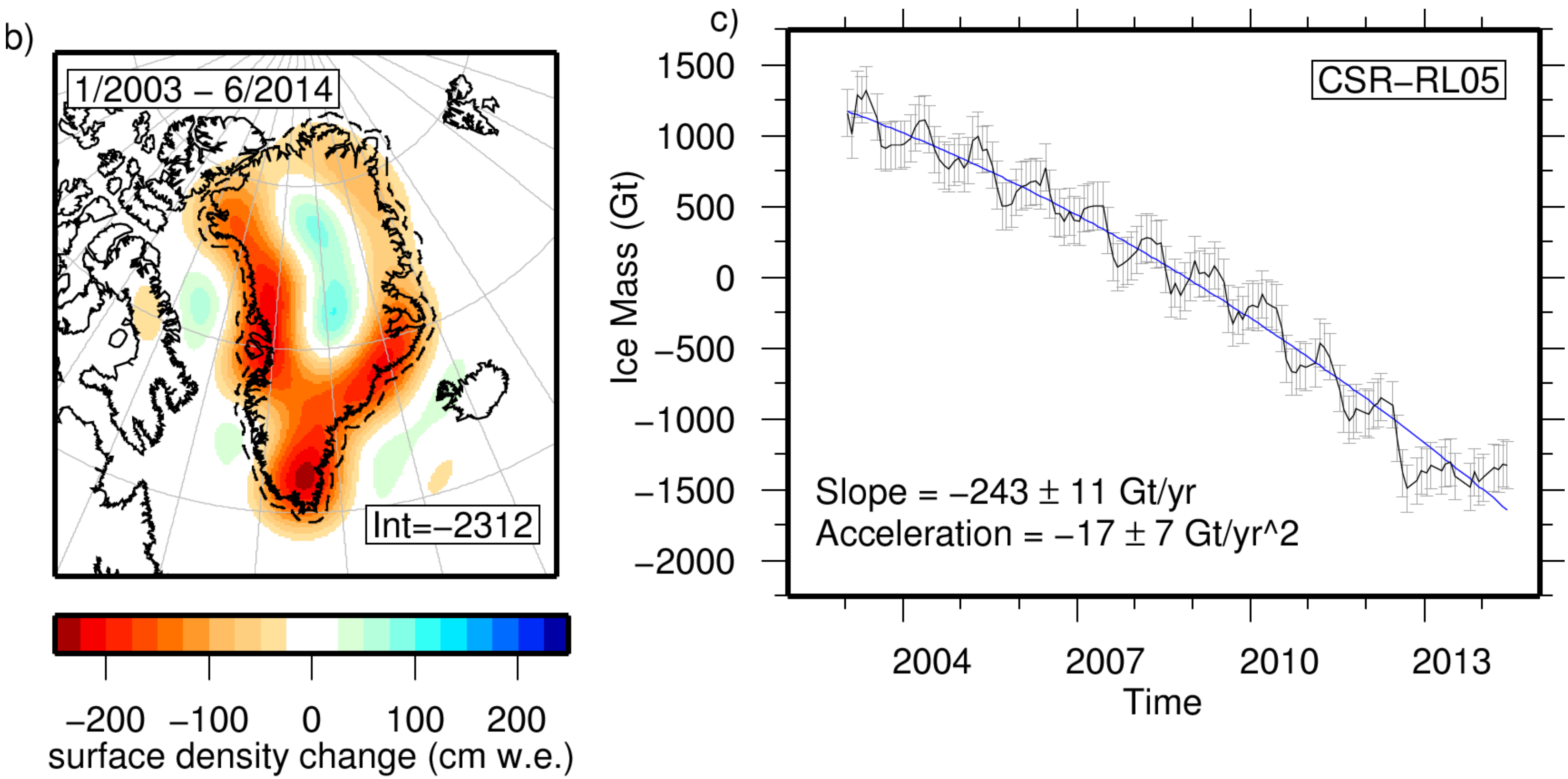
The "data" from the GRACE satellites (at *altitude*), are a non-local mixture $[s_{lm} + n_{lm}](t)$ from which, using the **Slepian basis** we extract the **spatio-temporal signature** of the source —
the time-dependent **ice mass loss** function (at the *surface*).

Harig et al., EOS 2015



Harig & Simons, EPSL 2015


a)
West Antarctica
IJ05_R2
500
0
–500
–1000
Slope = –121 ± 8 Gt/yr
Acceleration = –18 ± 5 Gt/yr^2
2002
2004
2006
2008
2010
2012
2014


b)
1/2003 – 6/2014
Int=–2312
–200 –100 0 100 200
surface density change (cm w.e.)
c)
CSR–RL05
1500
1000
500
0
–500
–1000
–1500
–2000
Ice Mass (Gt)
Slope = –243 ± 11 Gt/yr
Acceleration = –17 ± 7 Gt/yr^2
2004 2007 2010 2013
Time

Conclusions

- `Hello Earth`! We are raising the next generation of **geoscientists** for whom **reproducible research** and **computer programming** are a *must*
- **Matlab** *is* a tool to grow with students from their freshmen days to their professional academic environment — not the *only* language, but an *excellent* one
- We have embarked on a series of official `GitHub` releases of packages to perform **spatio-(temporal)-spectral estimation** of spherical data:

Conclusions

- `Hello Earth!` We are raising the next generation of **geoscientists** for whom **reproducible research** and **computer programming** are a *must*
- **Matlab** *is* a tool to grow with students from their freshmen days to their professional academic environment — not the *only* language, but an *excellent* one
- We have embarked on a series of official `GitHub` releases of packages to perform **spatio-(temporal)-spectral estimation** of spherical data:
- `SLEPIAN_Alpha`: Spherical harmonics, Slepian functions

Conclusions

- `Hello Earth!` We are raising the next generation of **geoscientists** for whom **reproducible research** and **computer programming** are a *must*
- **Matlab** *is* a tool to grow with students from their freshmen days to their professional academic environment — not the *only* language, but an *excellent* one
- We have embarked on a series of official `GitHub` releases of packages to perform **spatio-(temporal)-spectral estimation** of spherical data:
- `SLEPIAN_Alpha`: Spherical harmonics, Slepian functions
- `SLEPIAN_Bravo`: Linear inverse problems

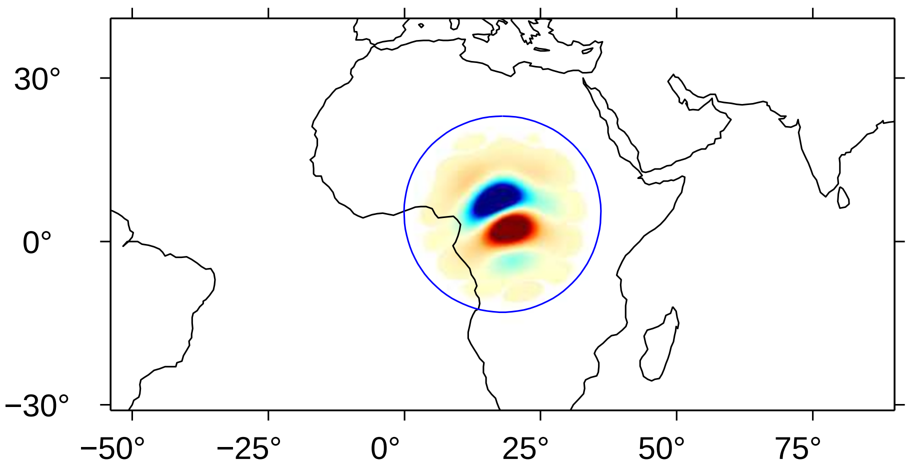
Conclusions

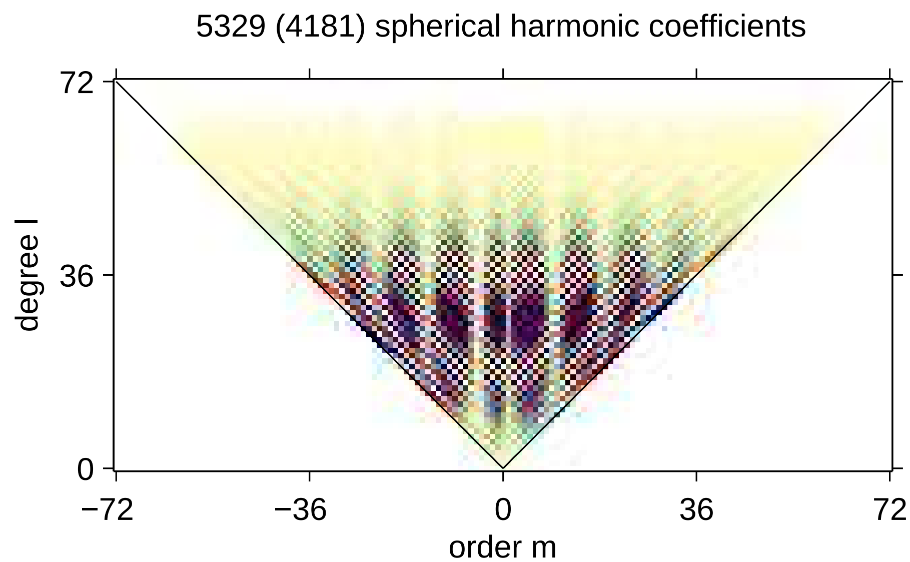
- `Hello Earth!` We are raising the next generation of **geoscientists** for whom **reproducible research** and **computer programming** are a *must*
- **Matlab** *is* a tool to grow with students from their freshmen days to their professional academic environment — not the *only* language, but an *excellent* one
- We have embarked on a series of official `GitHub` releases of packages to perform **spatio-(temporal)-spectral estimation** of spherical data:
- `SLEPIAN_Alpha`: Spherical harmonics, Slepian functions
- `SLEPIAN_Bravo`: Linear inverse problems
- `SLEPIAN_Charlie`: Quadratic spectral estimation

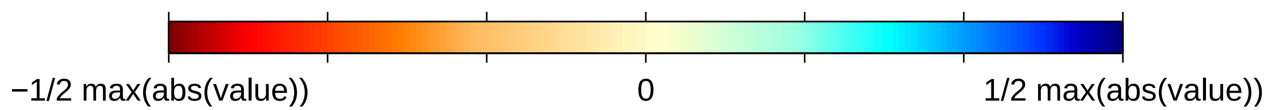
Conclusions

- `Hello Earth!` We are raising the next generation of **geoscientists** for whom **reproducible research** and **computer programming** are a *must*
- **Matlab** *is* a tool to grow with students from their freshmen days to their professional academic environment — not the *only* language, but an *excellent* one
- We have embarked on a series of official `GitHub` releases of packages to perform **spatio-(temporal)-spectral estimation** of spherical data:
- `SLEPIAN_Alpha`: Spherical harmonics, Slepian functions
- `SLEPIAN_Bravo`: Linear inverse problems
- `SLEPIAN_Charlie`: Quadratic spectral estimation
- `SLEPIAN_Delta`: Time-dependent estimation for GRACE satellite data

Basis I: spherical harmonics

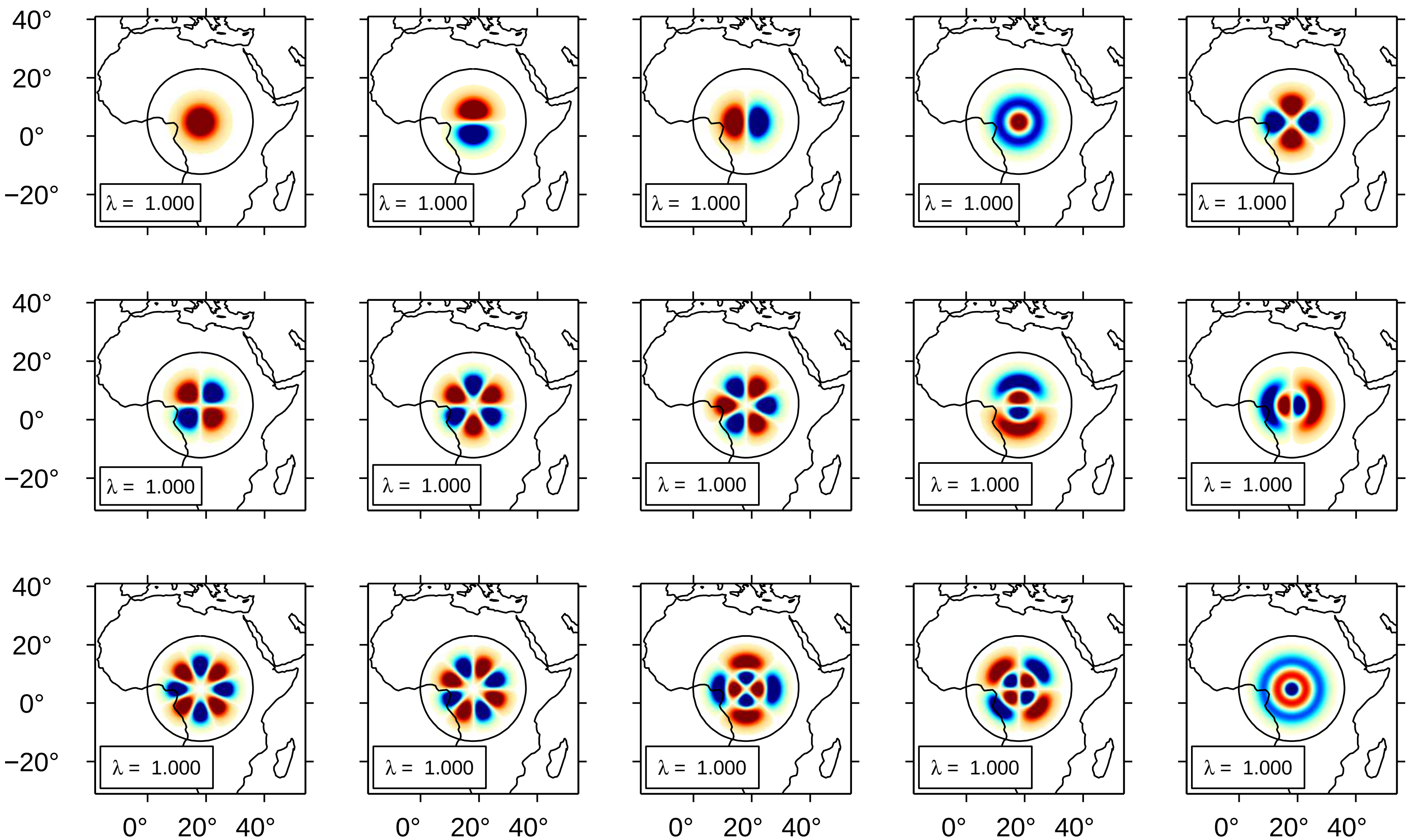




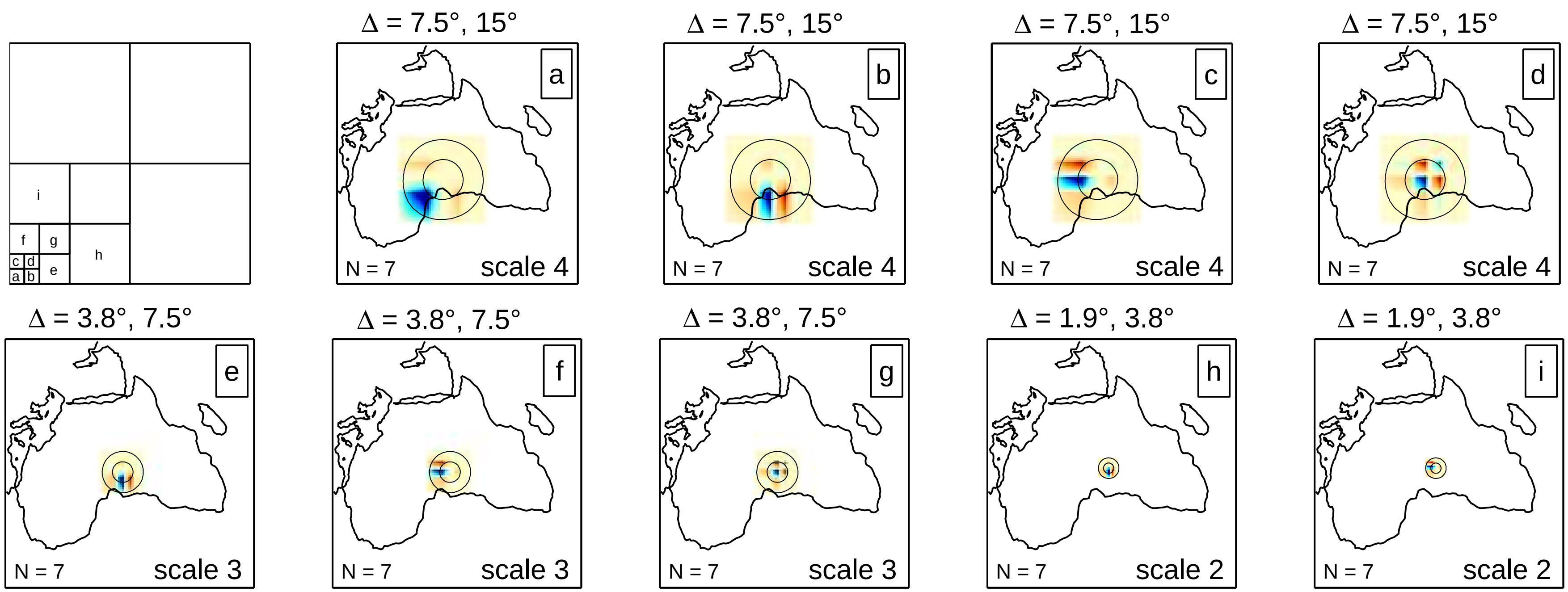


A *global* basis, **bad** for *local* problems.

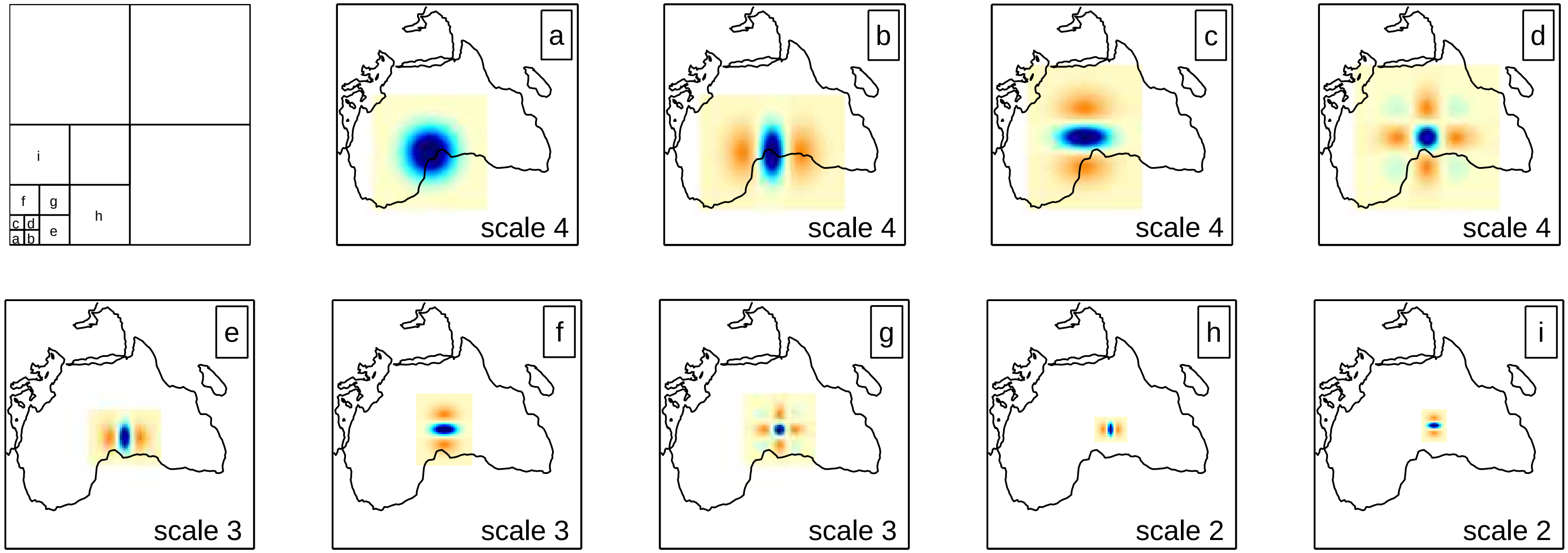
Basis II: Slepian functions



Basis III: cubed-spherical orthogonal wavelets



Loris, Simons et al., *GJI*, 2011



Simons, Loris *et al.*, *GJI*, 2011