Recovering data from seismic images by colourmap estimation

Matt Hall
Agile*

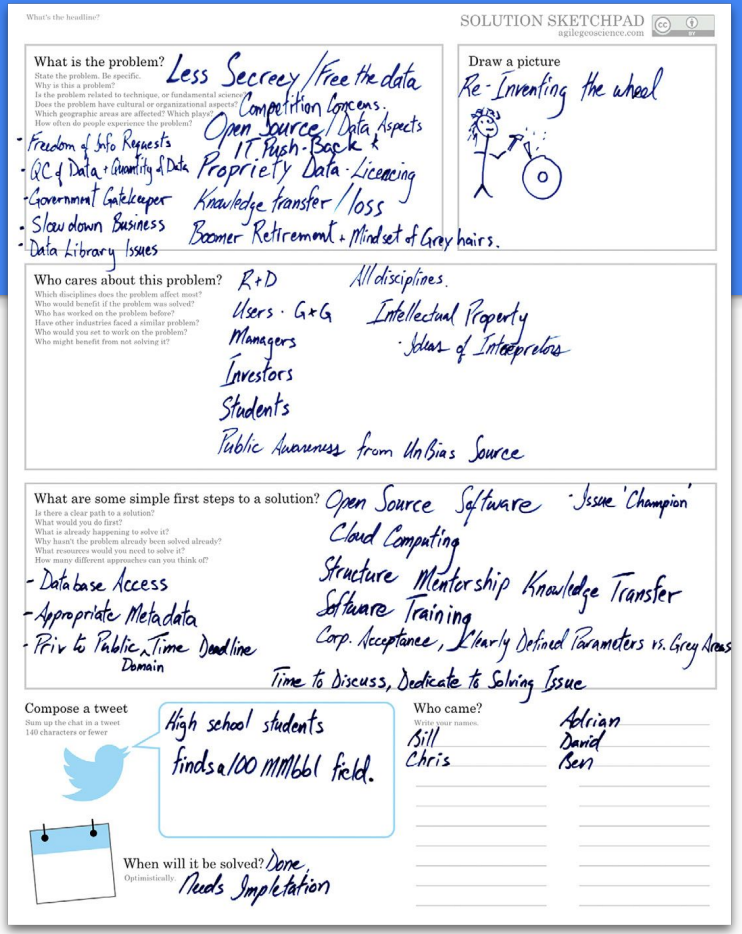
Matteo Niccoli
My Carta

ageo.co/geocon2017

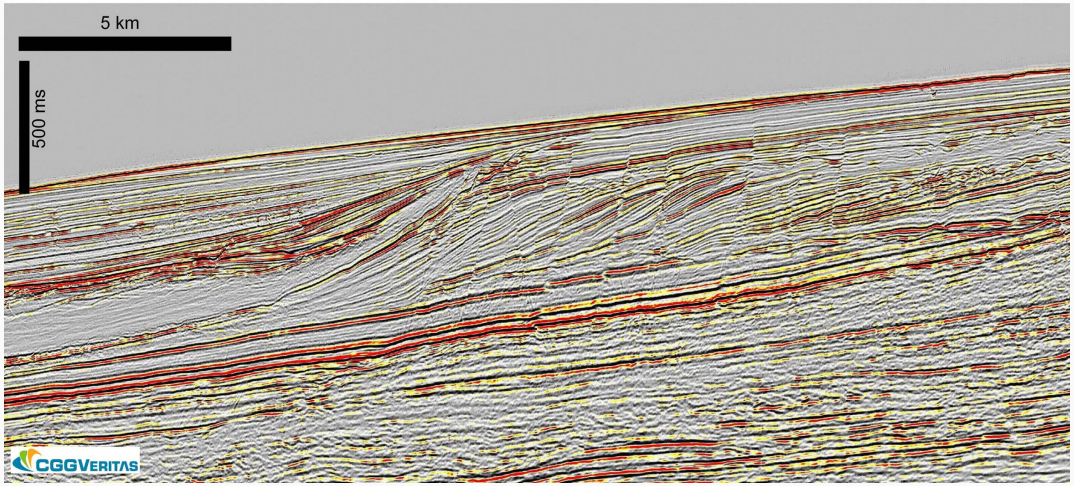
CC–BY 4.0 • Please share this work

csegrecorder.com/articles/view/proceedings-of-an-unsession

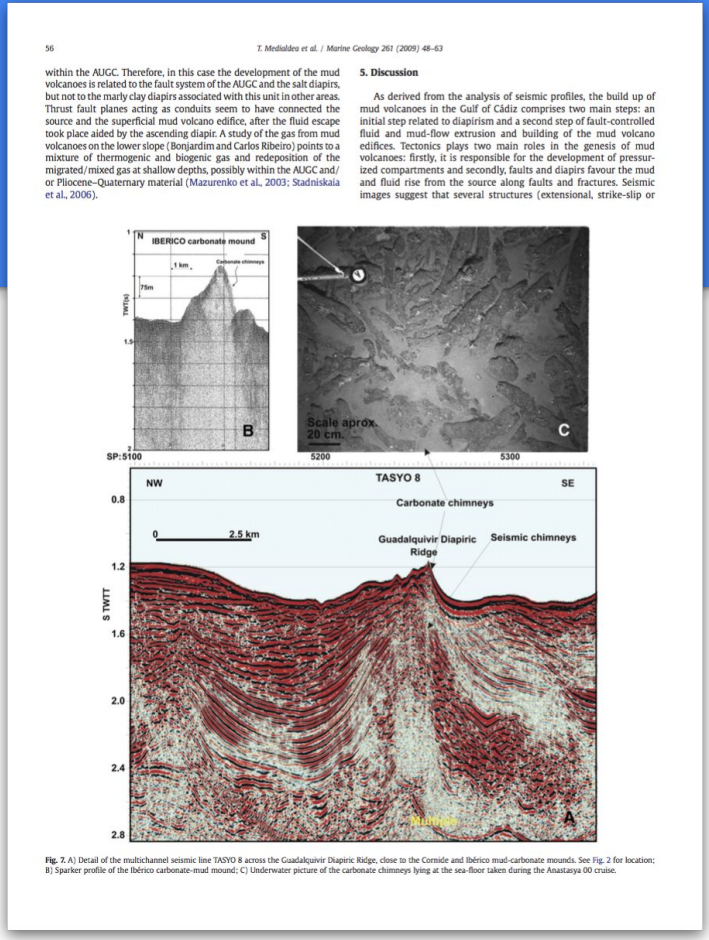
Free the data!



Free the data!



Copyright of Virtual seismic Atlas and CGG; used according to terms.
http://www.seismicatlas.org/entity?id=cba7ba45-6c87-418d-8295-c8180b3b0c38



56 T. Medialdea et al. / Marine Geology 261 (2009) 48–63

within the AUGC. Therefore, in this case the development of the mud volcanoes is related to the fault system of the AUGC and the salt diapirs, but not to the marly clay diapirs associated with this unit in other areas. Thrust fault planes acting as conduits seem to have connected the source and the superficial mud volcano edifice, after the fluid escape took place aided by the ascending diapir. A study of the gas from mud volcanoes on the lower slope (Bonjardim and Carlos Ribeiro) points to a mixture of thermogenic and biogenic gas and redeposition of the migrated/mixed gas at shallow depths, possibly within the AUGC and/or Pliocene–Quaternary material (Mazurenko et al., 2003; Stadniskaia et al., 2006).

5. Discussion

As derived from the analysis of seismic profiles, the build up of mud volcanoes in the Gulf of Cádiz comprises two main steps: an initial step related to diapirism and a second step of fault-controlled fluid and mud-flow extrusion and building of the mud volcano edifices. Tectonics plays two main roles in the genesis of mud volcanoes: firstly, it is responsible for the development of pressurized compartments and secondly, faults and diapirs favour the mud and fluid rise from the source along faults and fractures. Seismic images suggest that several structures (extensional, strike-slip or

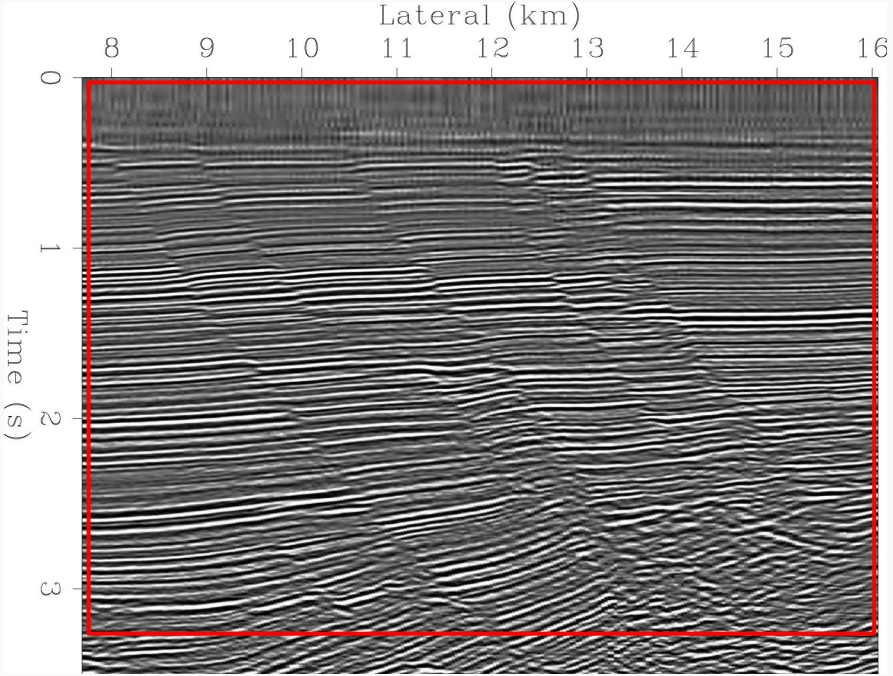
Fig. 7. A) Detail of the multichannel seismic line TASYO 8 across the Guadalquivir Diapiric Ridge, close to the Cornide and Ibérico mud-carbonate mounds. See Fig. 2 for location; B) Sparker profile of the Ibérico carbonate-mud mound; C) Underwater picture of the carbonate chimneys lying at the sea-floor taken during the Anastasya 00 cruise.

Medialdea et al. (2009), Marine Geology ***261***.

ageobot.elasticbeanstalk.com

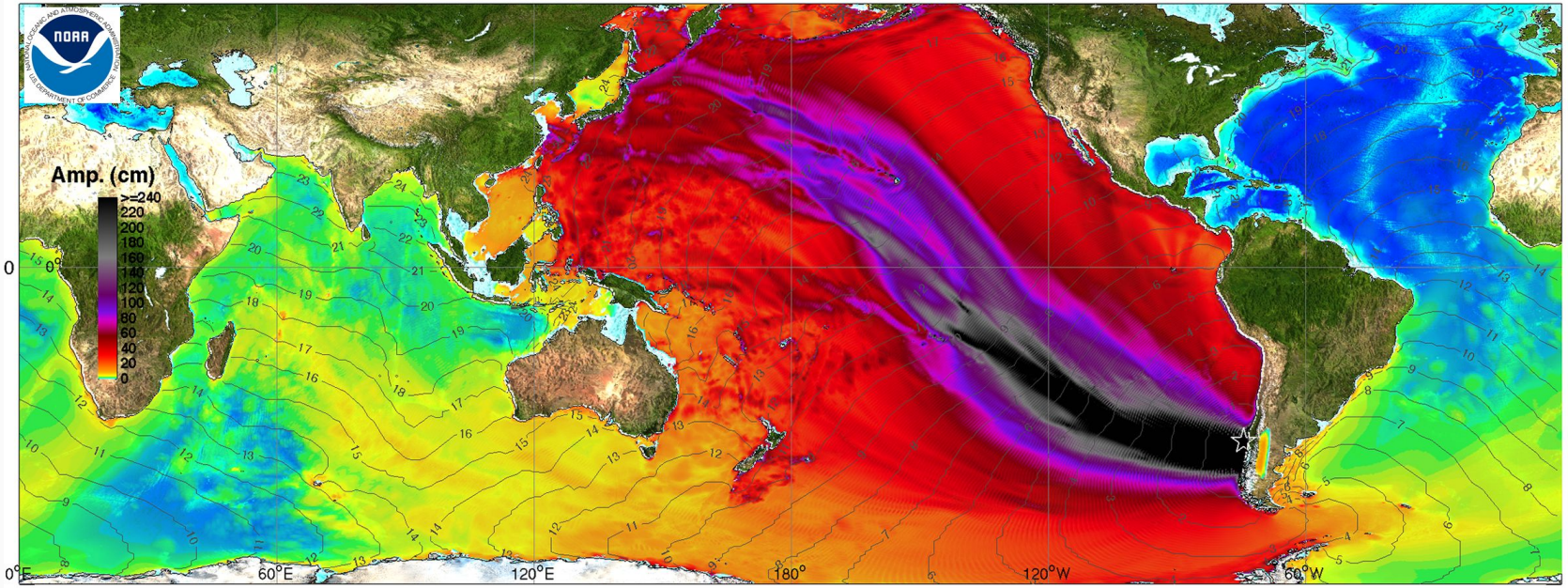
ageobot / freqbot

- Give the URL to an **image**.
- Define a region.
- Provide the **timing**, 0 to 3.2 s.
- **Get metadata**: spectrum, dominant frequency, S:N, etc.
- **Get a SEG-Y file.**



http://nctr.pmel.noaa.gov/chile20100227/

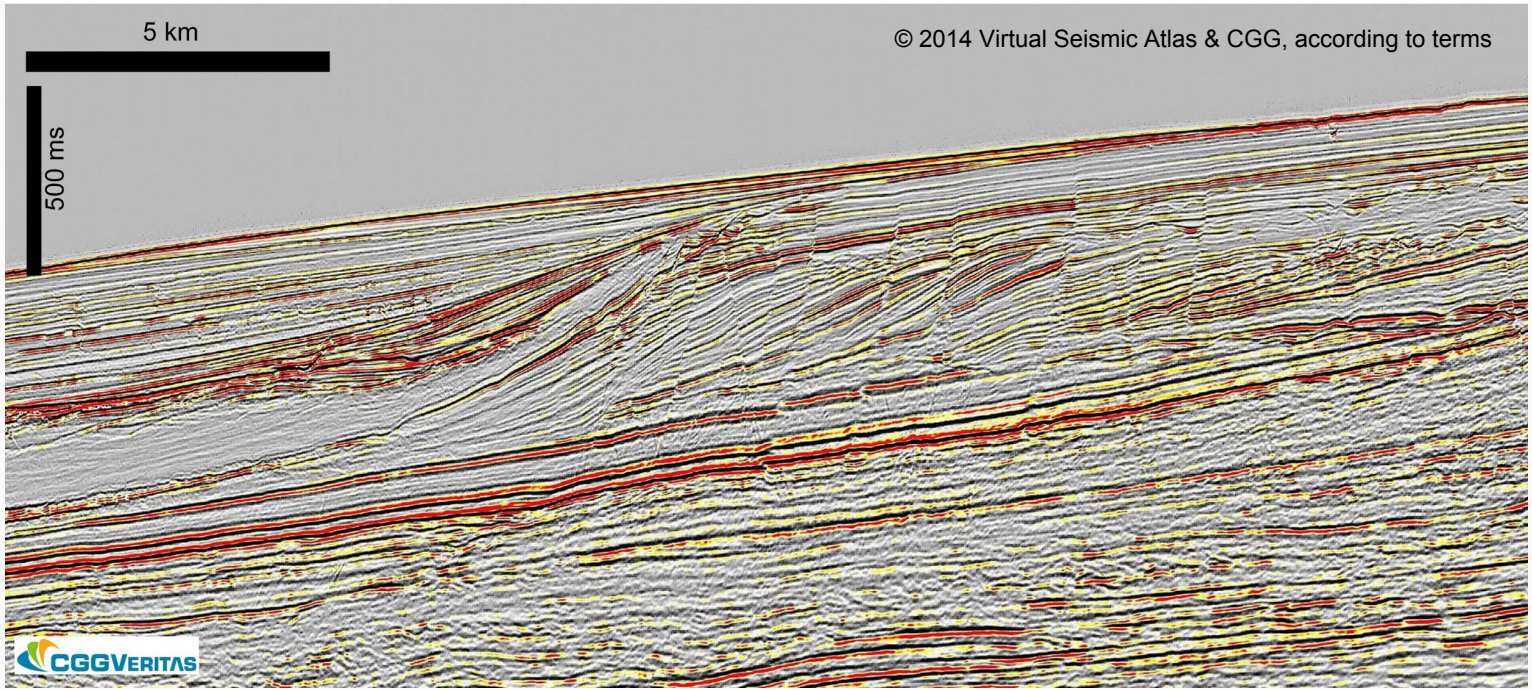
Maps usually have a codebook



1960 Chile tsunami (earthquake magnitude Mw 9.5) Maximum amplitude plot -- for comparison with Feb 27, 2010 Chile tsunami event

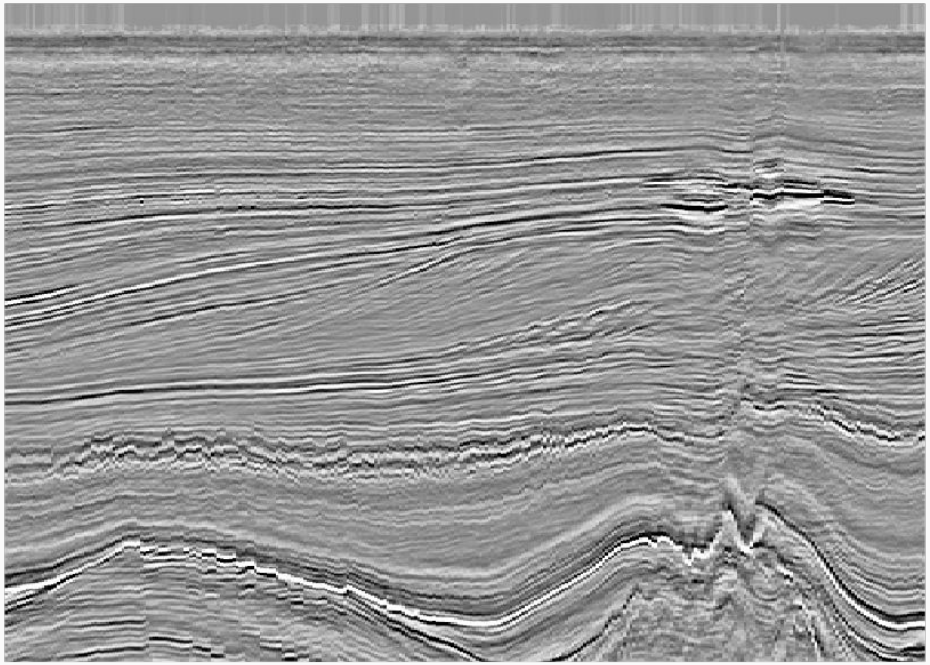
http://www.seismicatlas.org/entity?id=cba7ba45-6c87-418d-8295-c8180b3b0c38

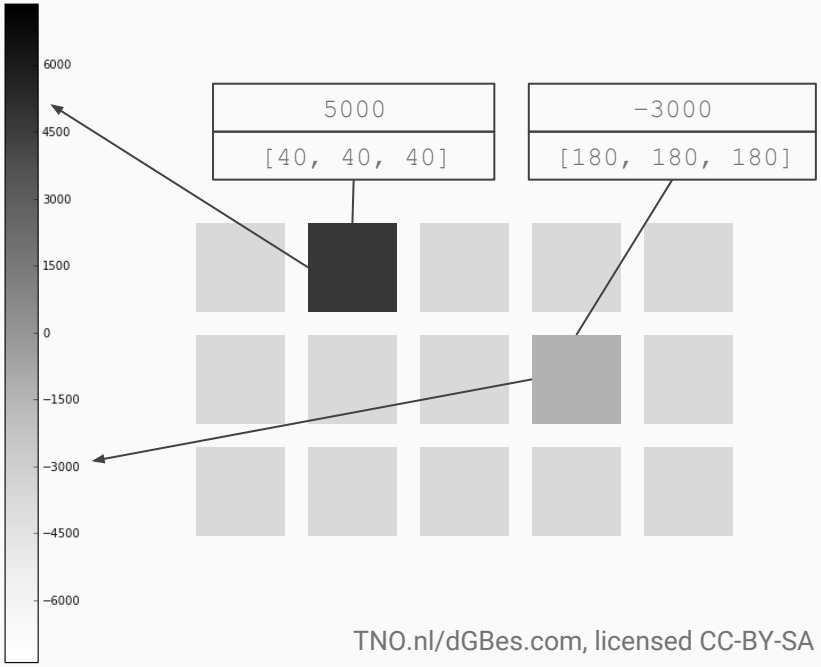
Seismic rarely does



F3 seismic data from Open Seismic Repository

Monochrome images

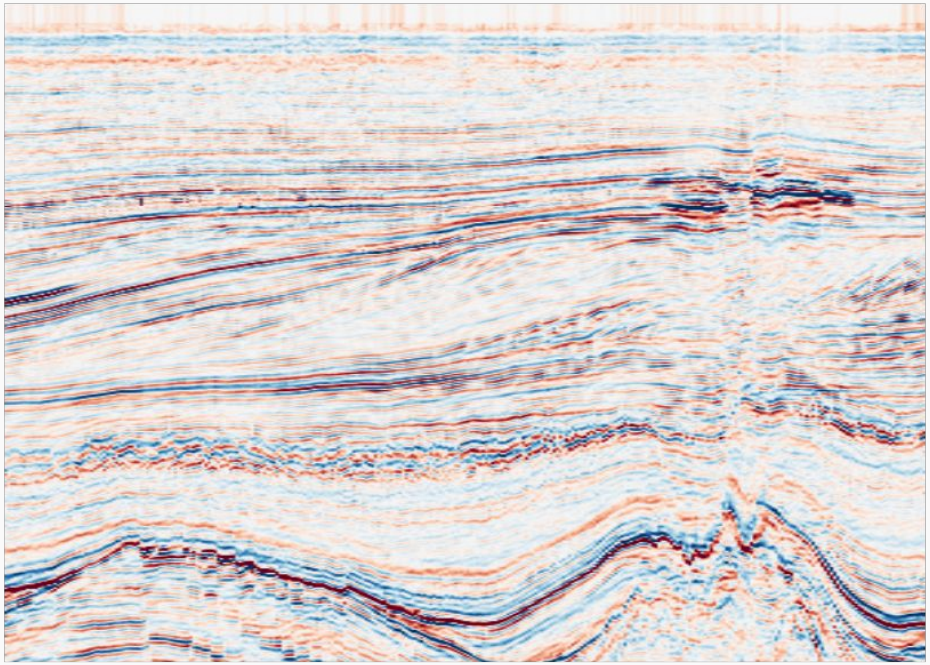


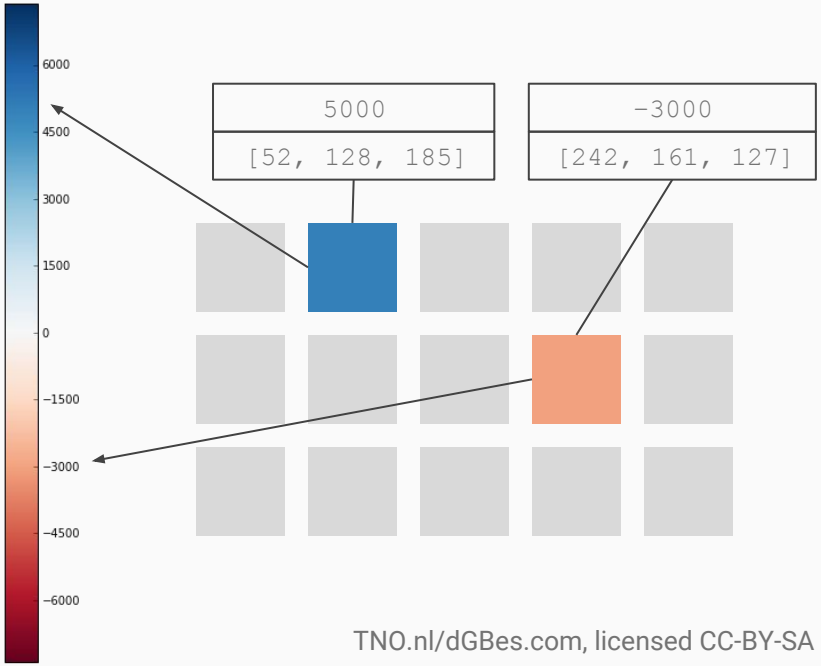


TNO.nl/dGBes.com, licensed CC-BY-SA

F3 seismic data from Open Seismic Repository

Pseudocolour images

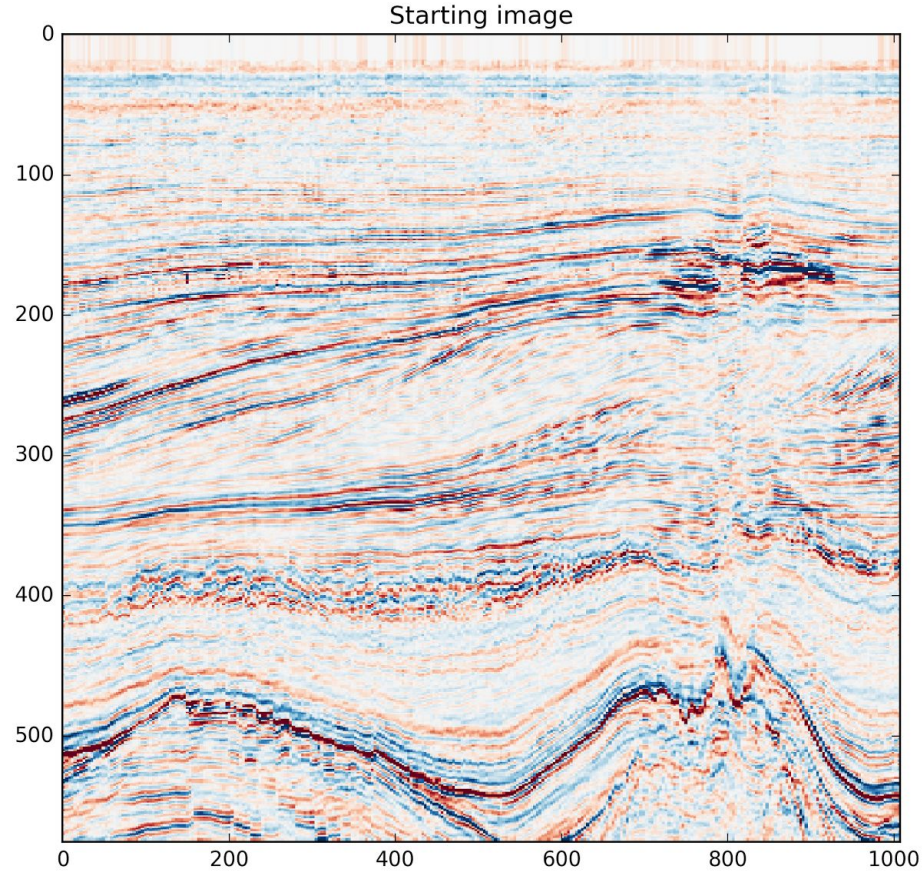


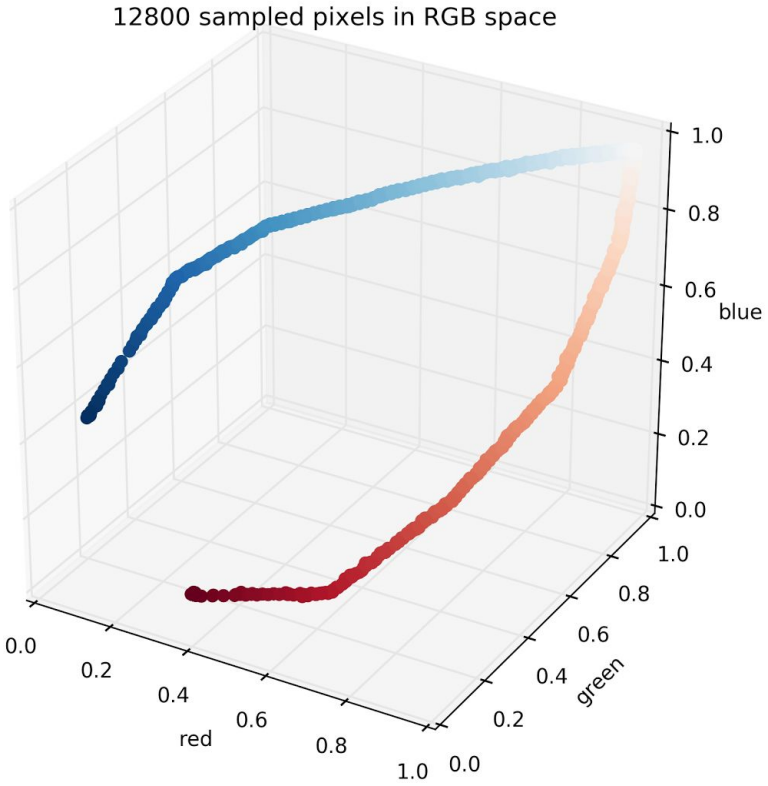


TNO.nl/dGBes.com, licensed CC-BY-SA

Move to RGB space

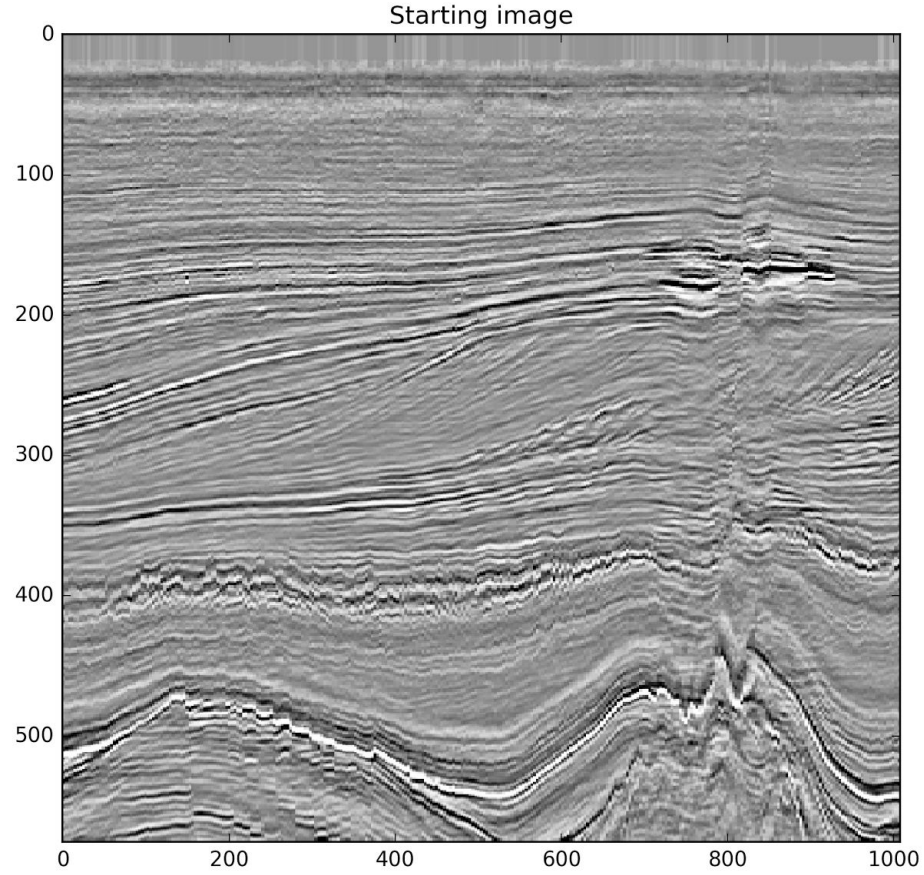
TNO.nl/dGBes.com, licensed CC-BY-SA

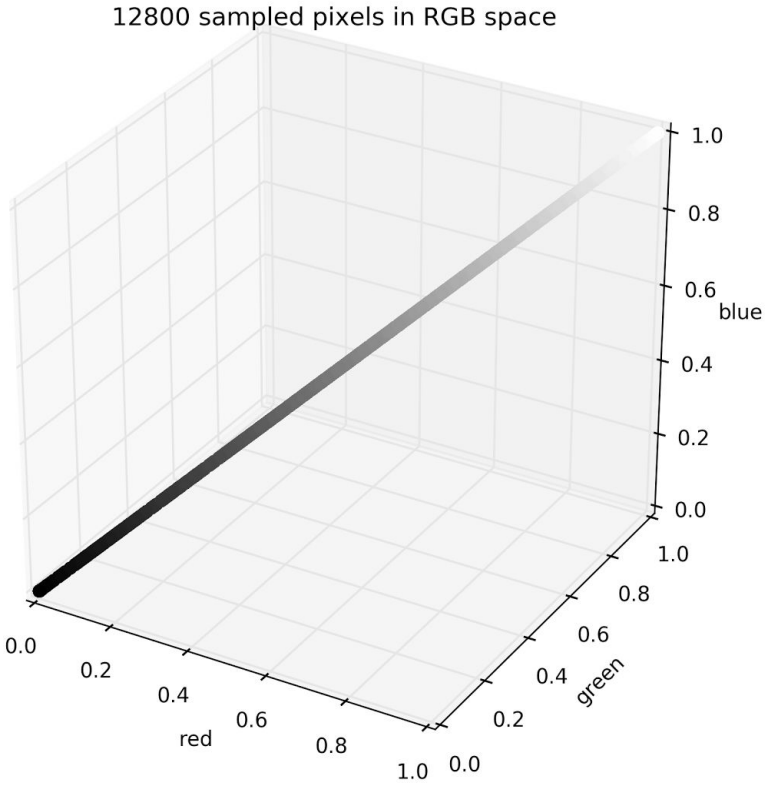




Perceptually monotonic colourmaps

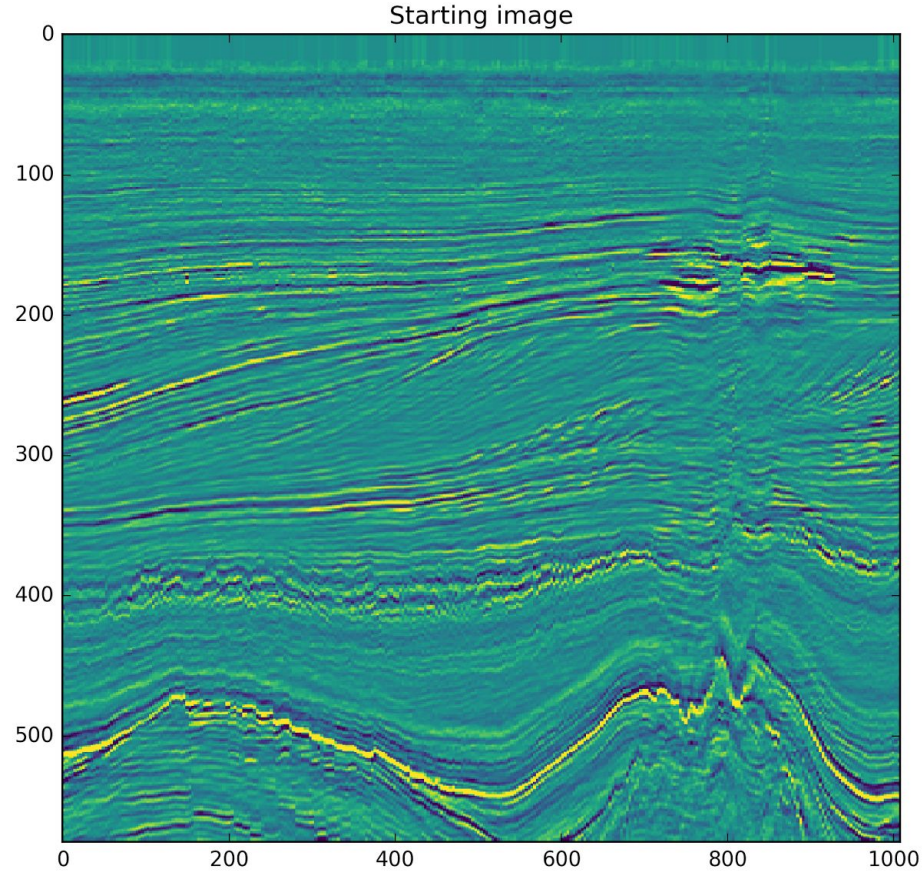
TNO.nl/dGBes.com, licensed CC-BY-SA

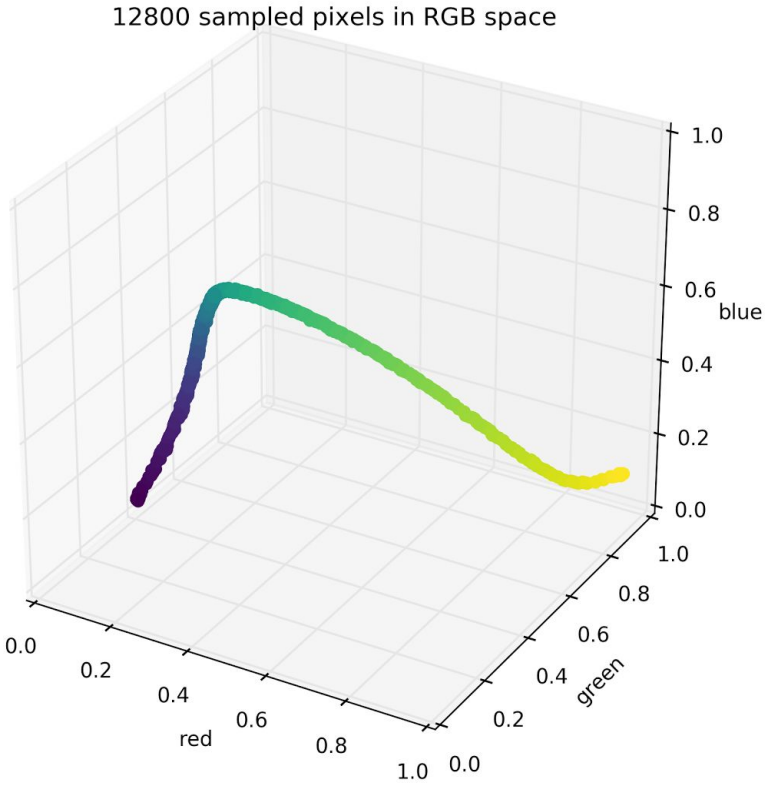




Perceptually monotonic colourmaps

TNO.nl/dGBes.com, licensed CC-BY-SA



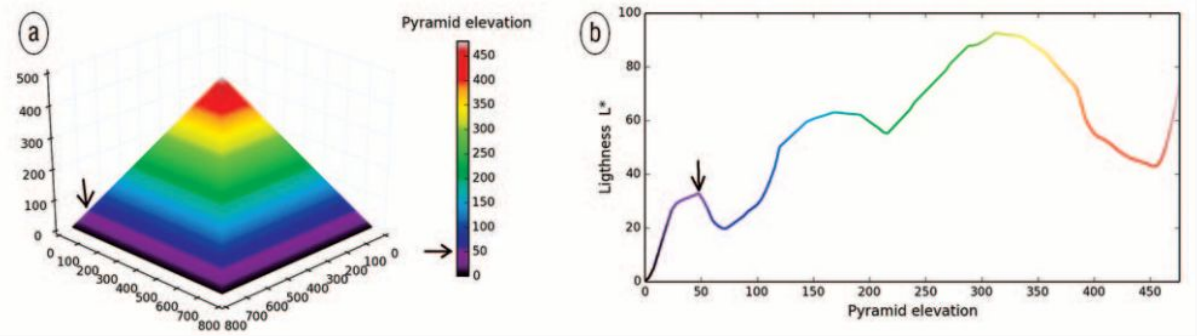


http://library.seg.org/doi/full/10.1190/tle33080910.1

Perceptually monotonic test

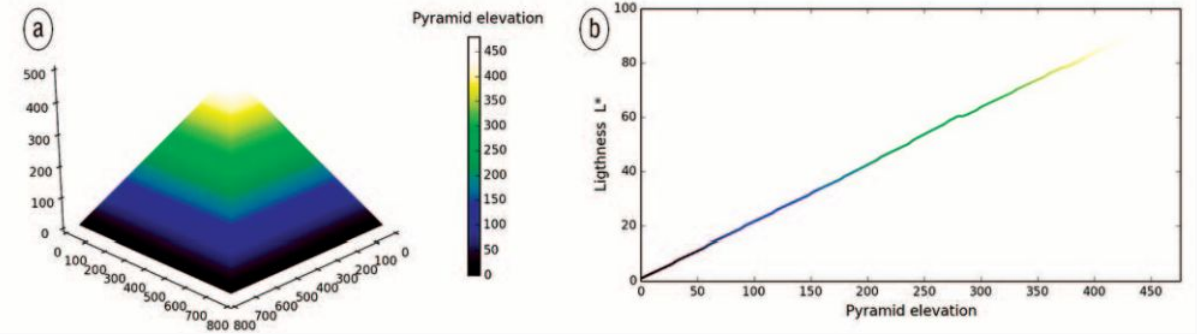
Non-perceptual colourbar

- Non-linear lightness
- Bad for your health

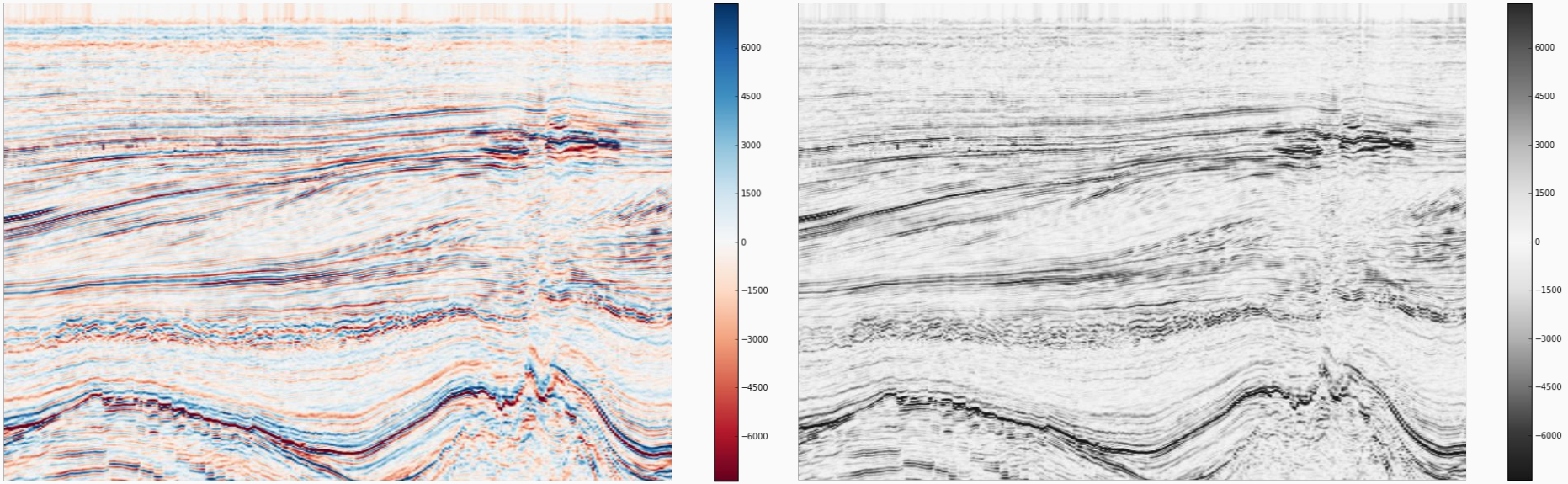


Perceptual colourbar

- Linear lightness
- Good for you!

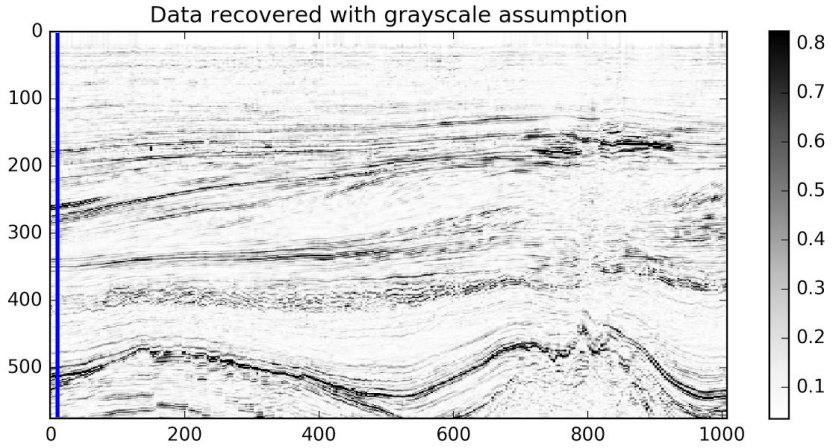


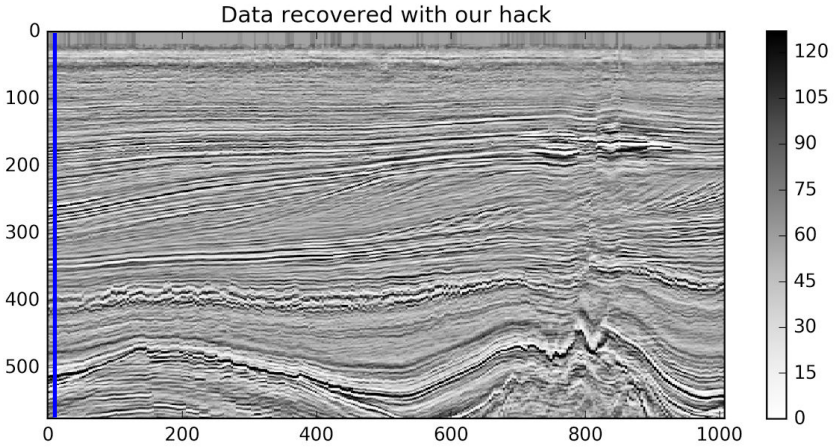
Cannot treat all colourmaps as greyscale

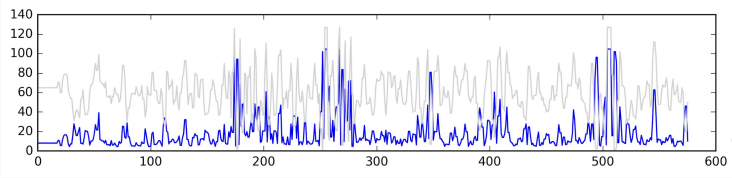


TNO.nl/dGBes.com, licensed CC-BY-SA

Cannot treat all colourmaps as greyscale

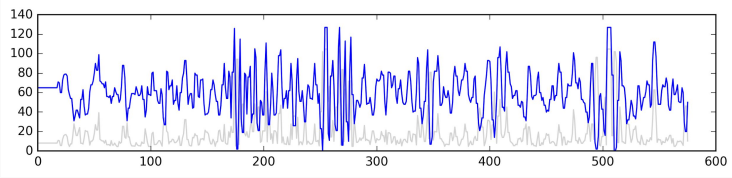






< not seismicky

seismicky>

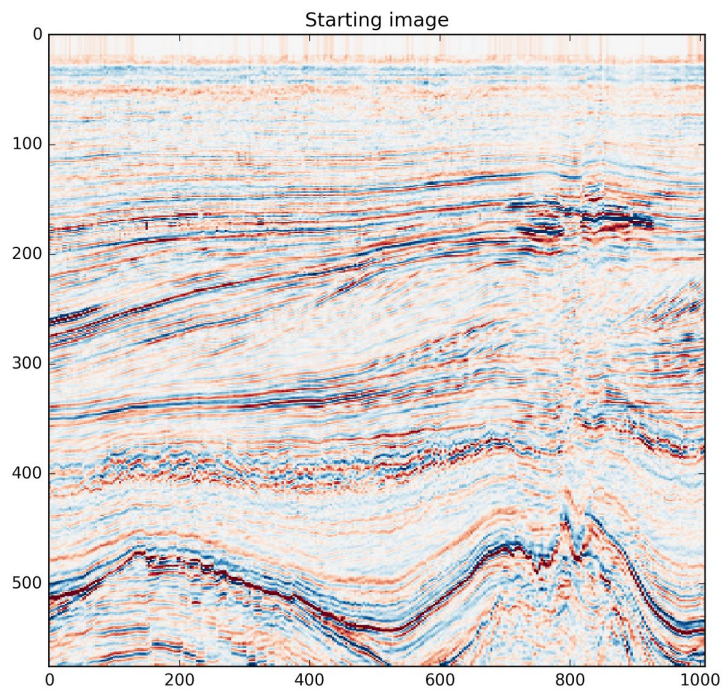


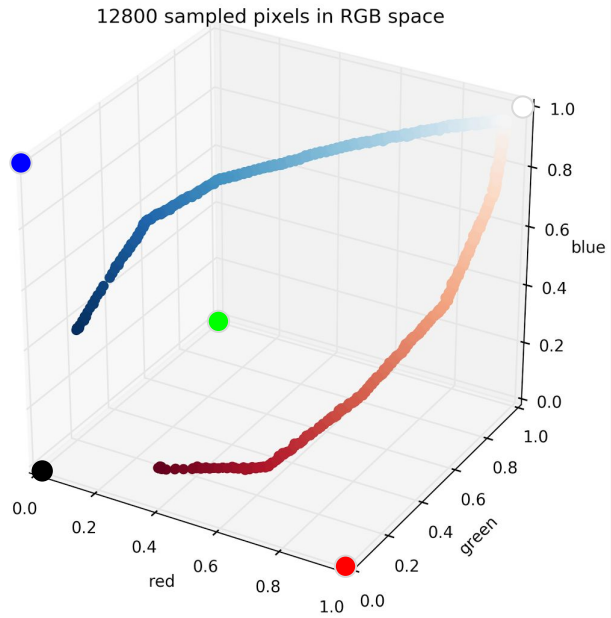
TNO.nl/dGBes.com, licensed CC-BY-SA

```
ax.scatter(*pixels.T, c=pixels)
```

Move to RGB space

TNO.nl/dGBes.com, licensed CC-BY-SA

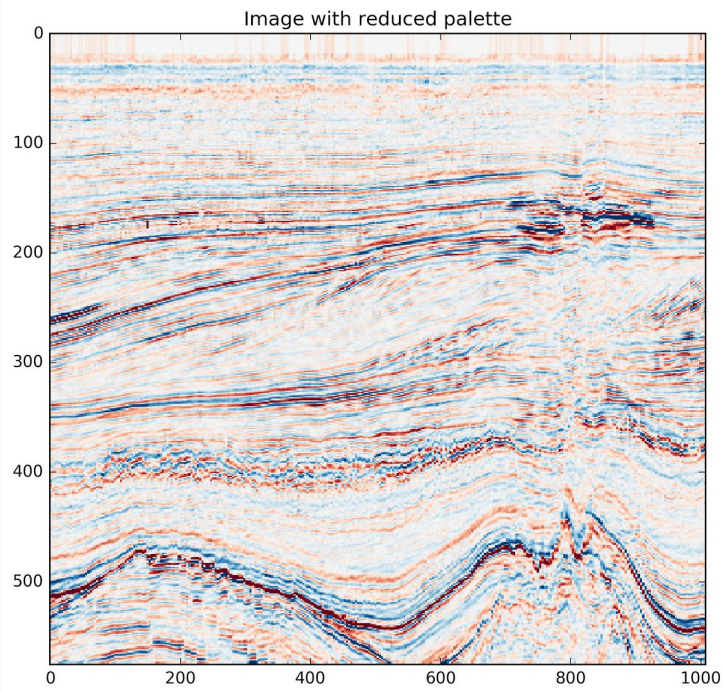


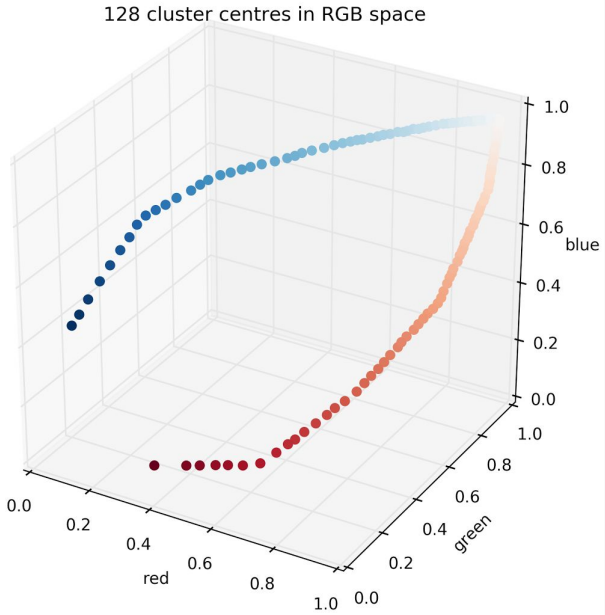


```
p = KMeans(n_clusters=128).fit(pixels).cluster_centers_
```

k-means palette reduction

TNO.nl/dGBes.com, licensed CC-BY-SA

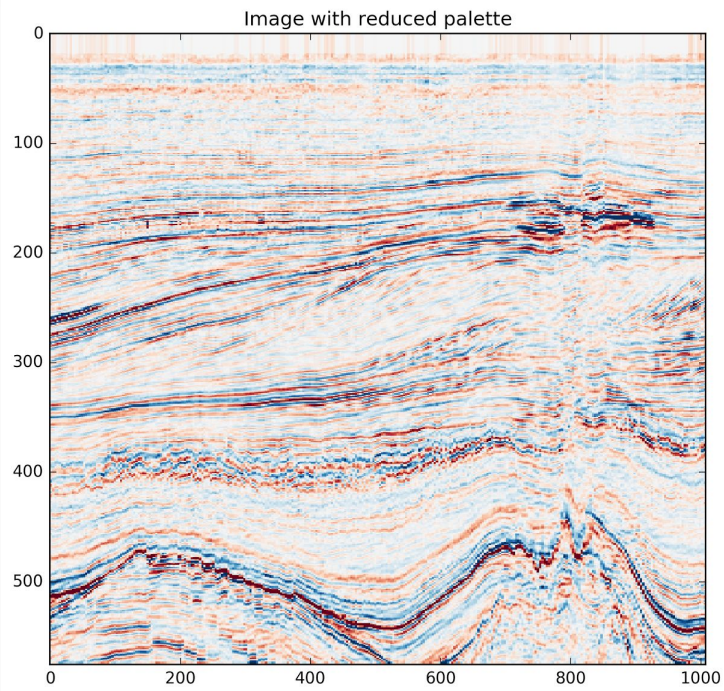


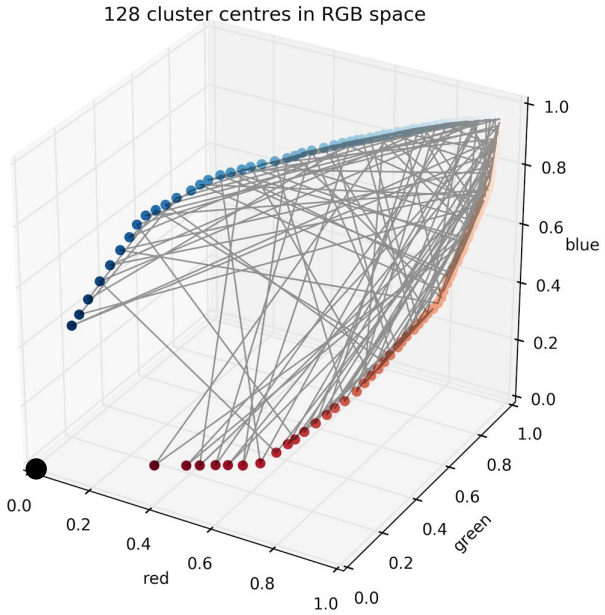


```
p = KMeans(n_clusters=128).fit(pixels).cluster_centers_
```

k-means palette reduction

TNO.nl/dGBes.com, licensed CC-BY-SA





The shortest Hamiltonian path

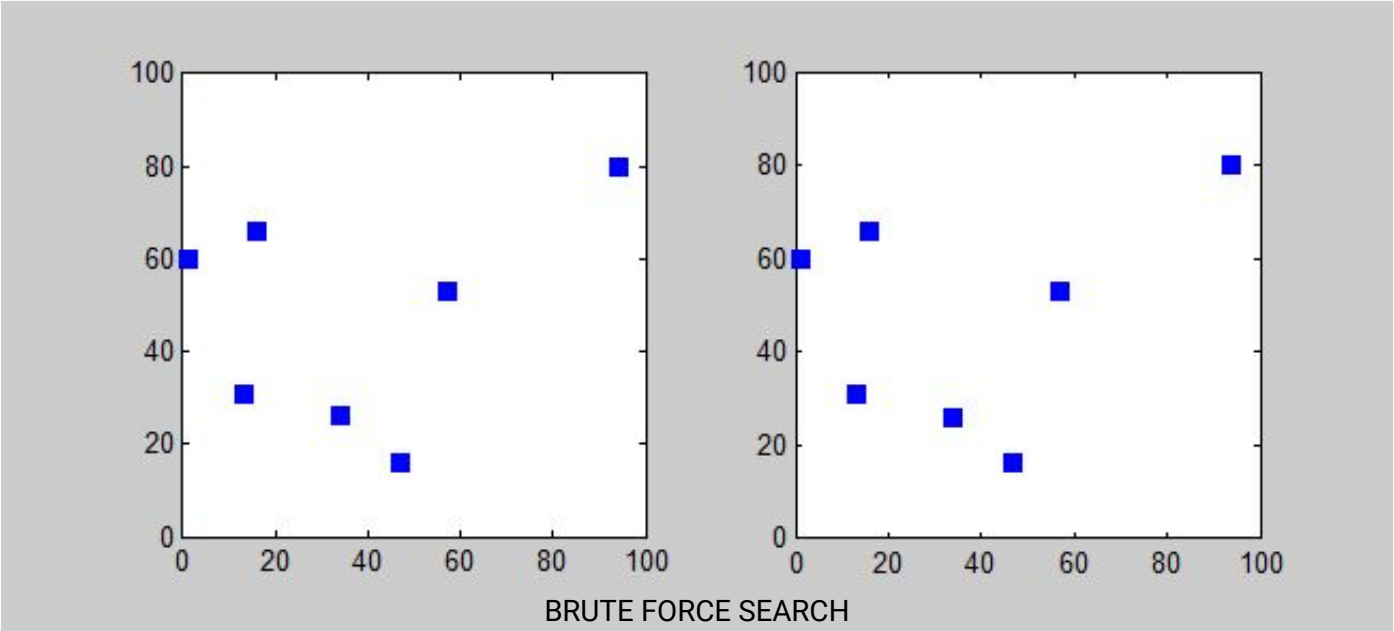
Visit each vertex of a dodecahedron exactly once.

For a cycle, return to the same point.

William Rowan Hamilton's Icosian Game
© RIA Library / Fair use via Twitter

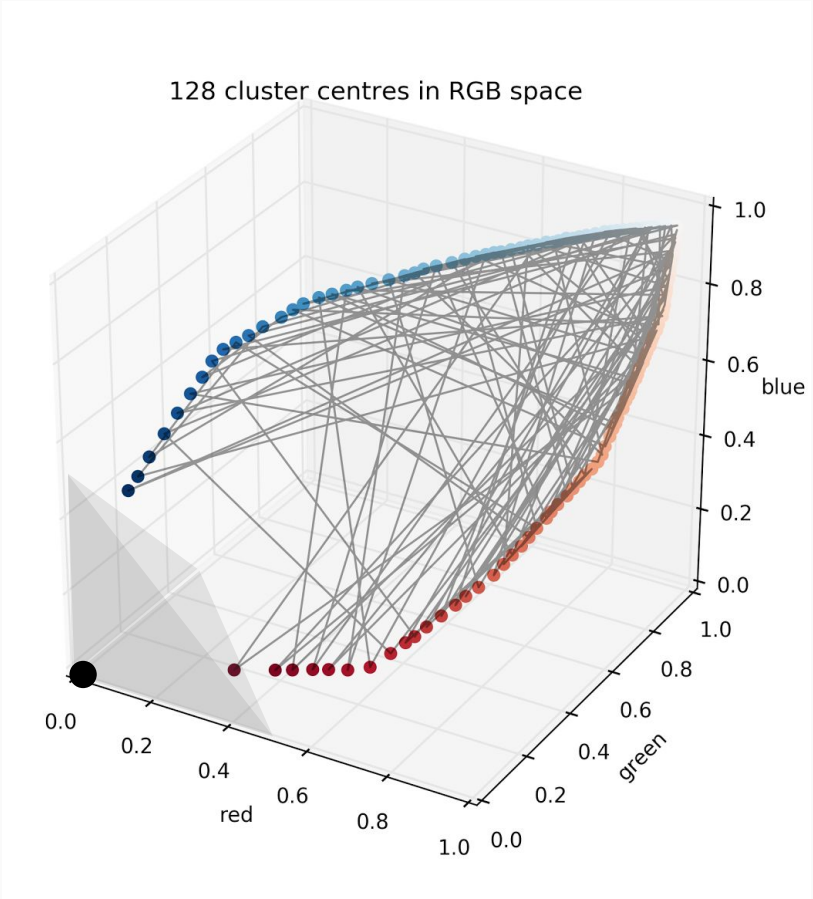
Brute force solution for seven cities: (7 – 1)!/2 = 360 iterations

Travelling salesman problem



Solving shortest Hamiltonian as TSP

- Add a 'zero point' somewhere at or near the black point [0, 0, 0]. Pass this to the solver as the starting point.
- Add an 'everywhere point' that is at a distance zero from every other point. This will be visited last by the solver.



E. L. Lawler, Jan Karel Lenstra, A. H. G. Rinnooy Kan, D. B. Shmoys (1985). The Traveling Salesman Problem: A Guided Tour of Combinatorial Optimization, 1st Edition. Wiley. 476 pp.

github.com/kwinkunks/pytsp

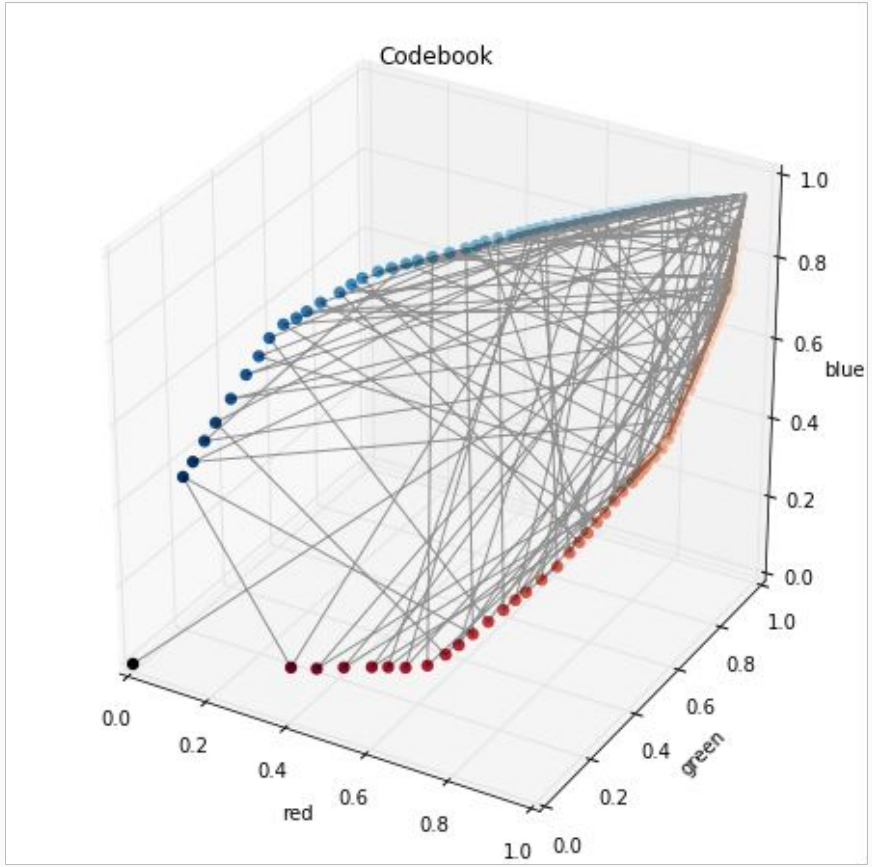
Lin–Kernigan–Helsgaun solver

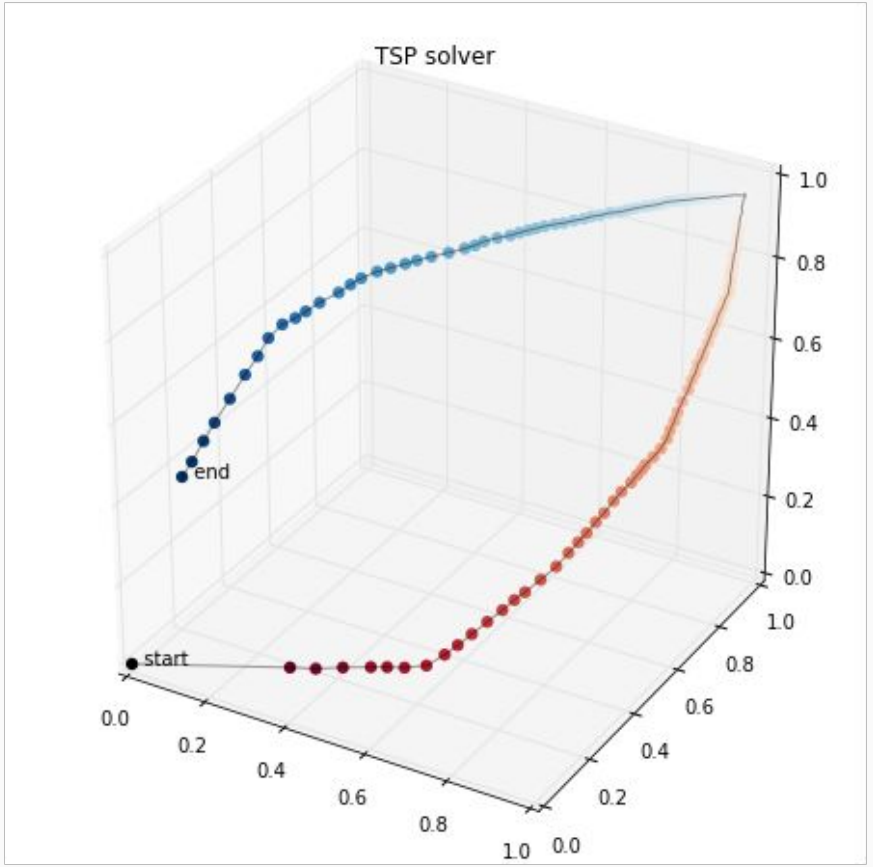
- Best solvers are LKH, Concorde, and Google's O-R Tools.
- LKH was implemented via modified **`pytsp`** wrapper.
- Only input is matrix of distances between all node pairs

```
from scipy.spatial.distance import pdist, squareform
dists = squareform(pdist(p, 'euclidean'))
```

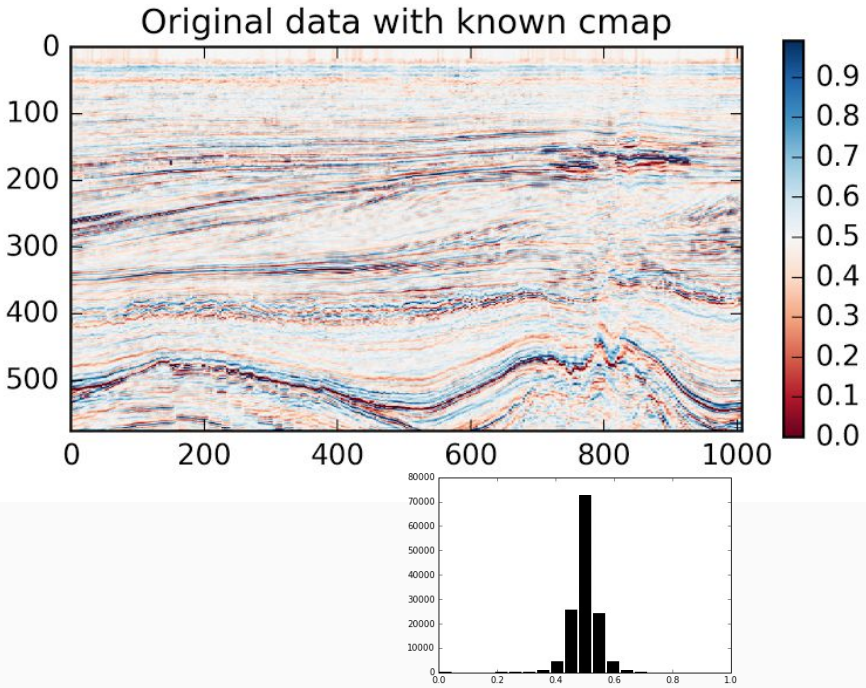
Helsgaun, K (2009). General k-opt submoves for the Lin–Kernighan TSP heuristic. Mathematical Programming Computation. DOI: 10.1007/s12532-009-0004-6.

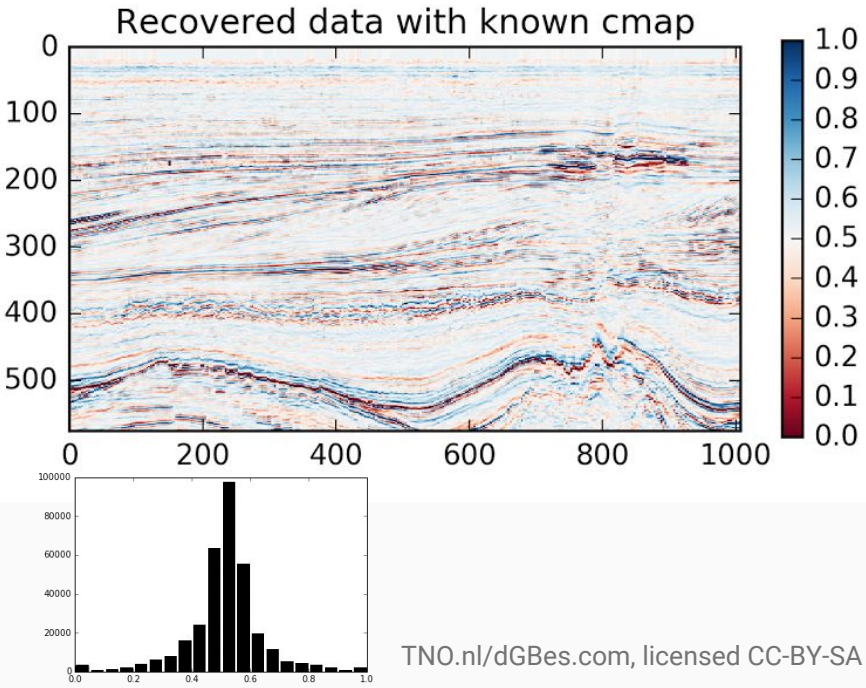
Unordered graph → shortest Hamiltonian path





Result

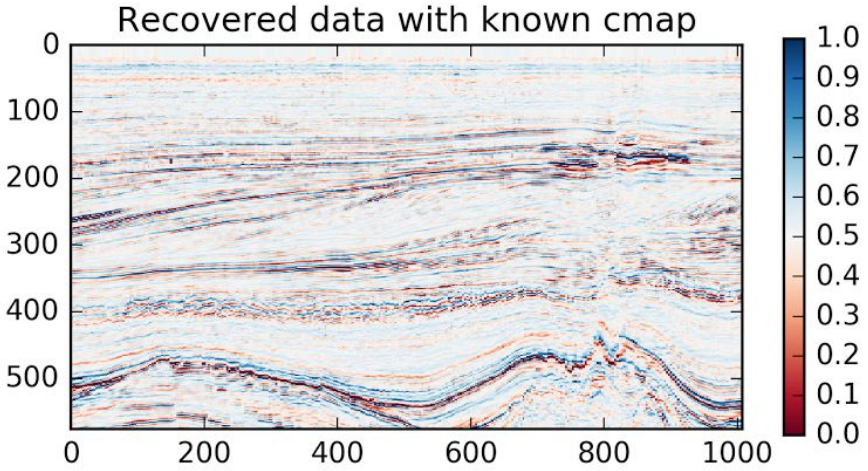


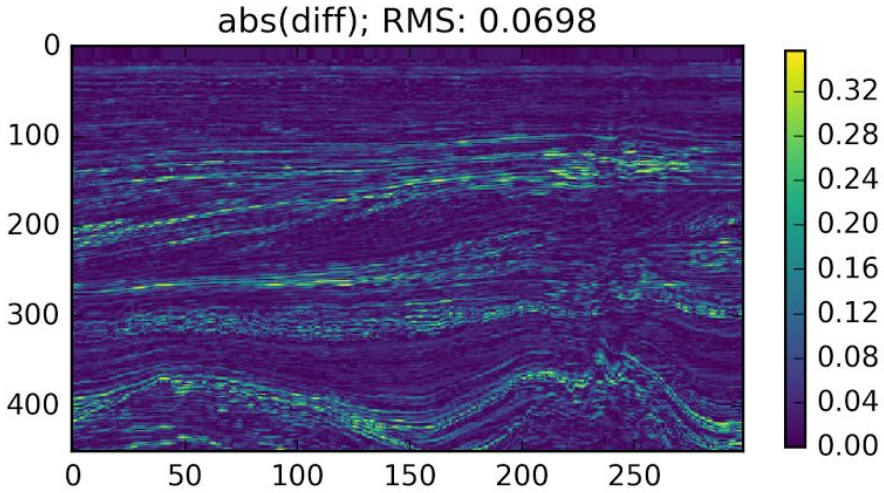


TNO.nl/dGBes.com, licensed CC-BY-SA

Absolute difference and RMS error

Error

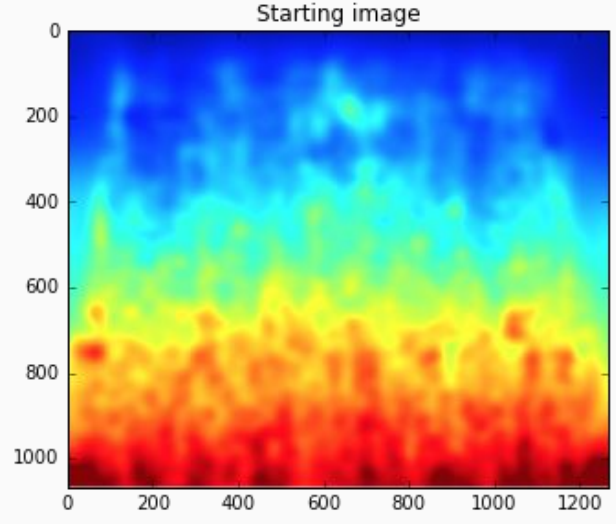


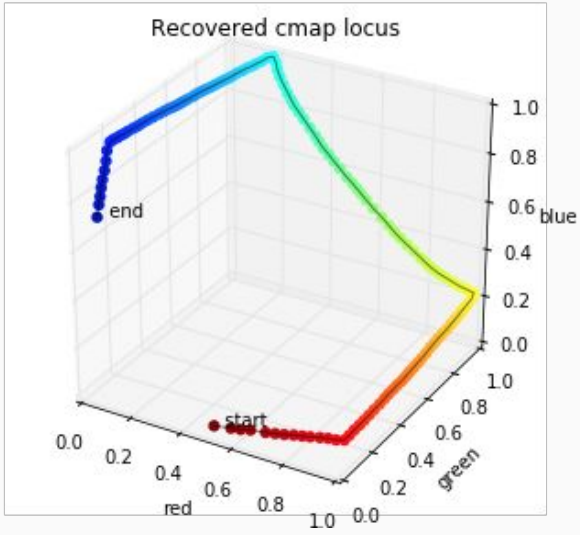


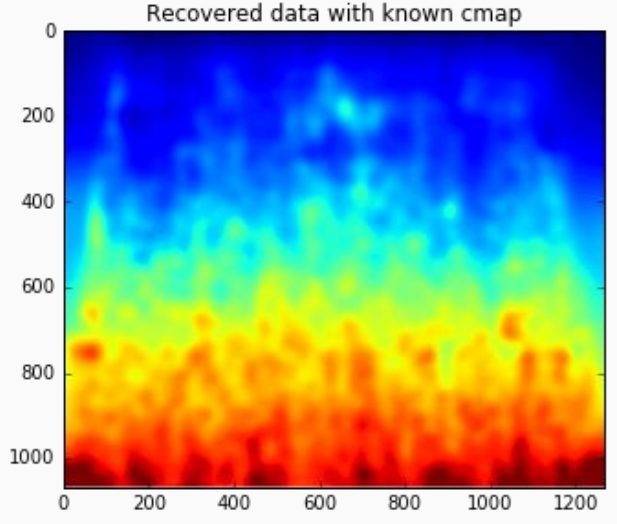
TNO.nl/dGBes.com, licensed CC-BY-SA

`cmap = 'jet'` (most similar)

Seismic velocity field

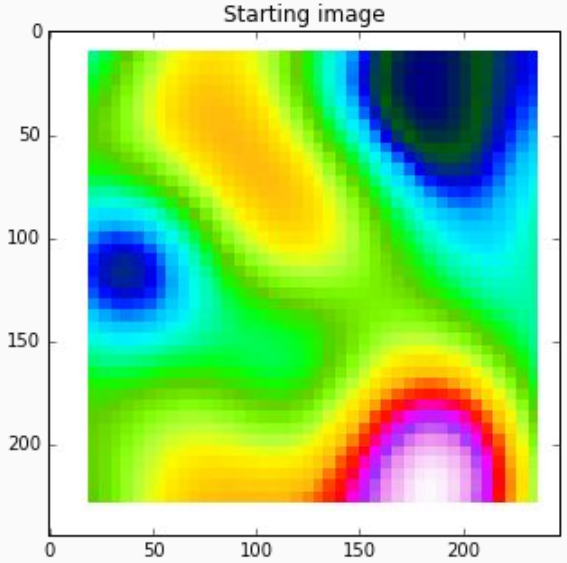


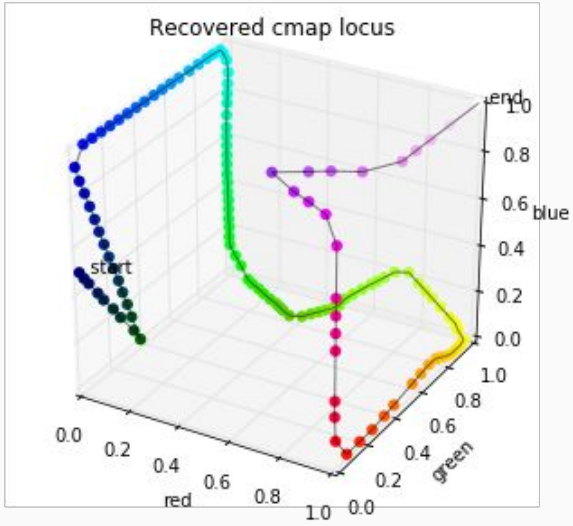


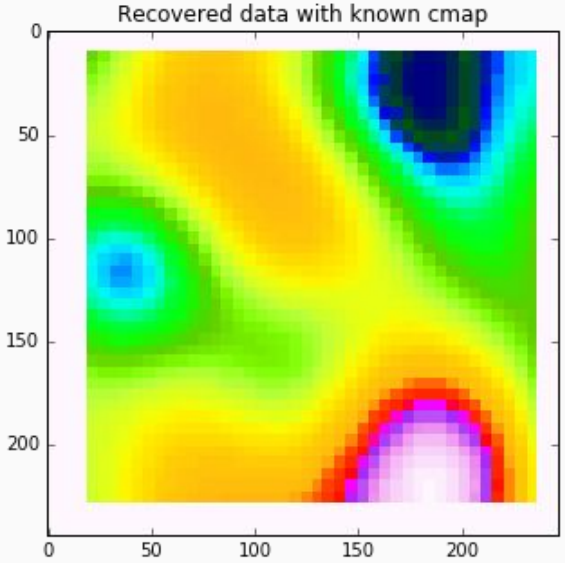


```
cmap = 'gist_ncar'
```

Synthetic data







Conquer all mysteries by rule and line,
Empty the haunted air, and gnomed mine—
Unweave a rainbow, as it erewhile made
The tender-person'd Lamia melt into a shade.

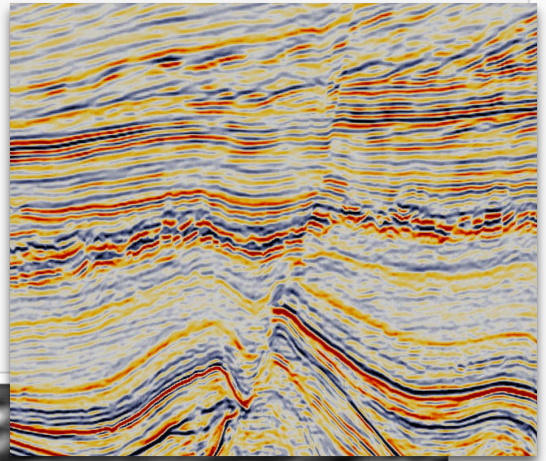
John Keats (1820), *Lamia*.

keats.geosci.ai

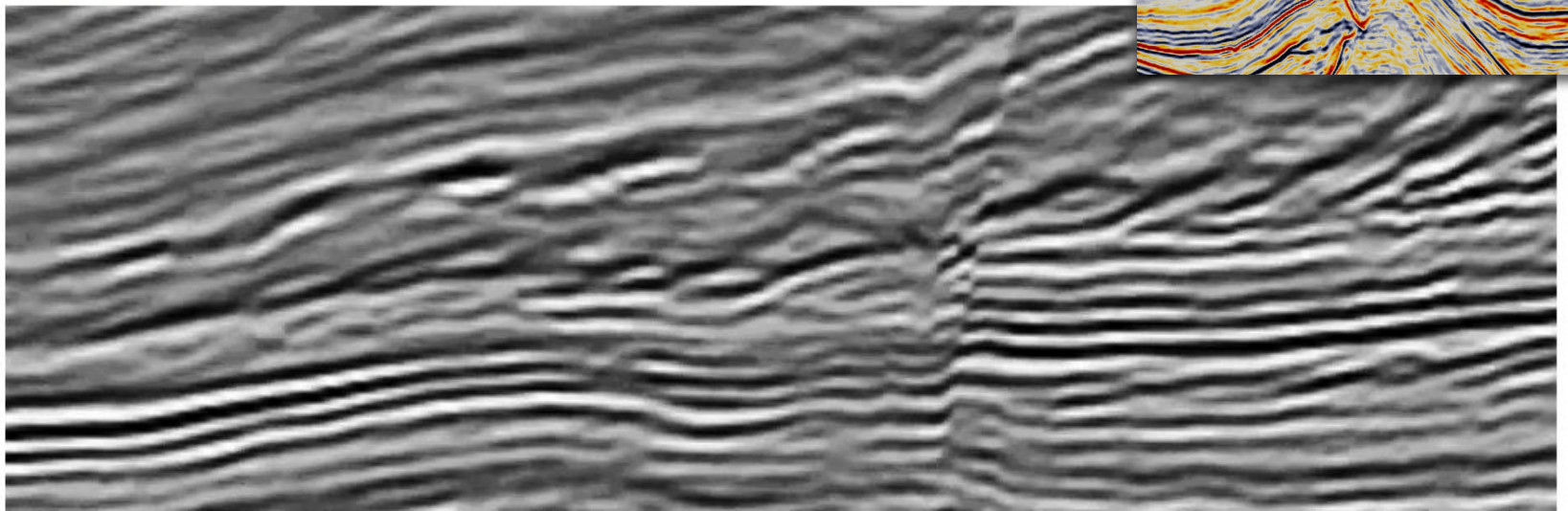
amazon web services EC2

In [17]:
```
import requests
from io import BytesIO
from PIL import Image

url = "http://keats.geosci.ai/api"
query = {"url": "http://www.signatureseismic.com/user_files/images/fault.jpg",
         "format": "png",
         "coolpoint": [0.0, 0.0, 0.1],
        }
r = requests.get(url, query)
r = requests.get(r.json()['result']['image'])
Image.open(BytesIO(r.content))
```



Out[17]:



http://github.com/cseg/2017-hall-niccoli

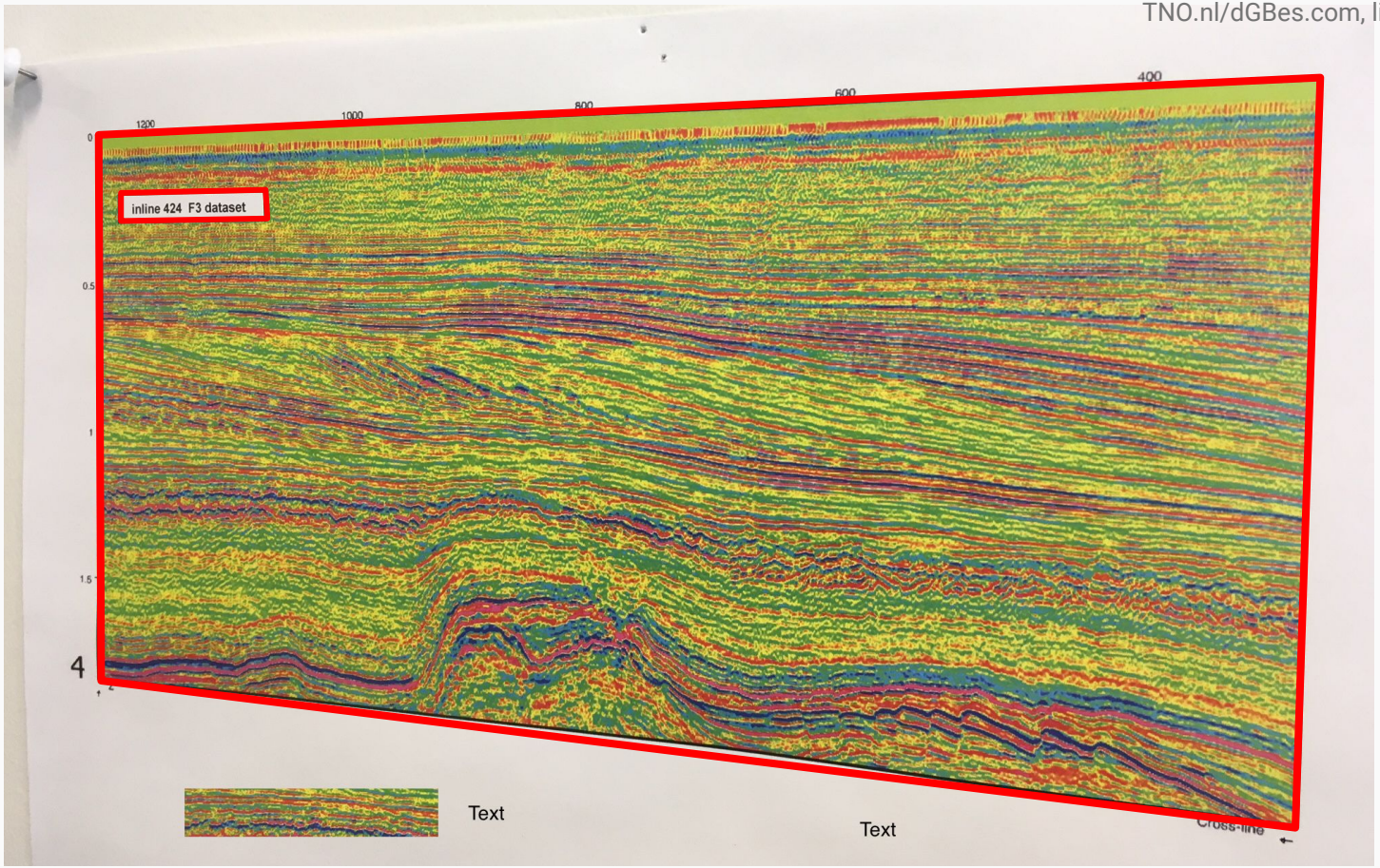
Algorithm

STEP ⇐ SOFTWARE

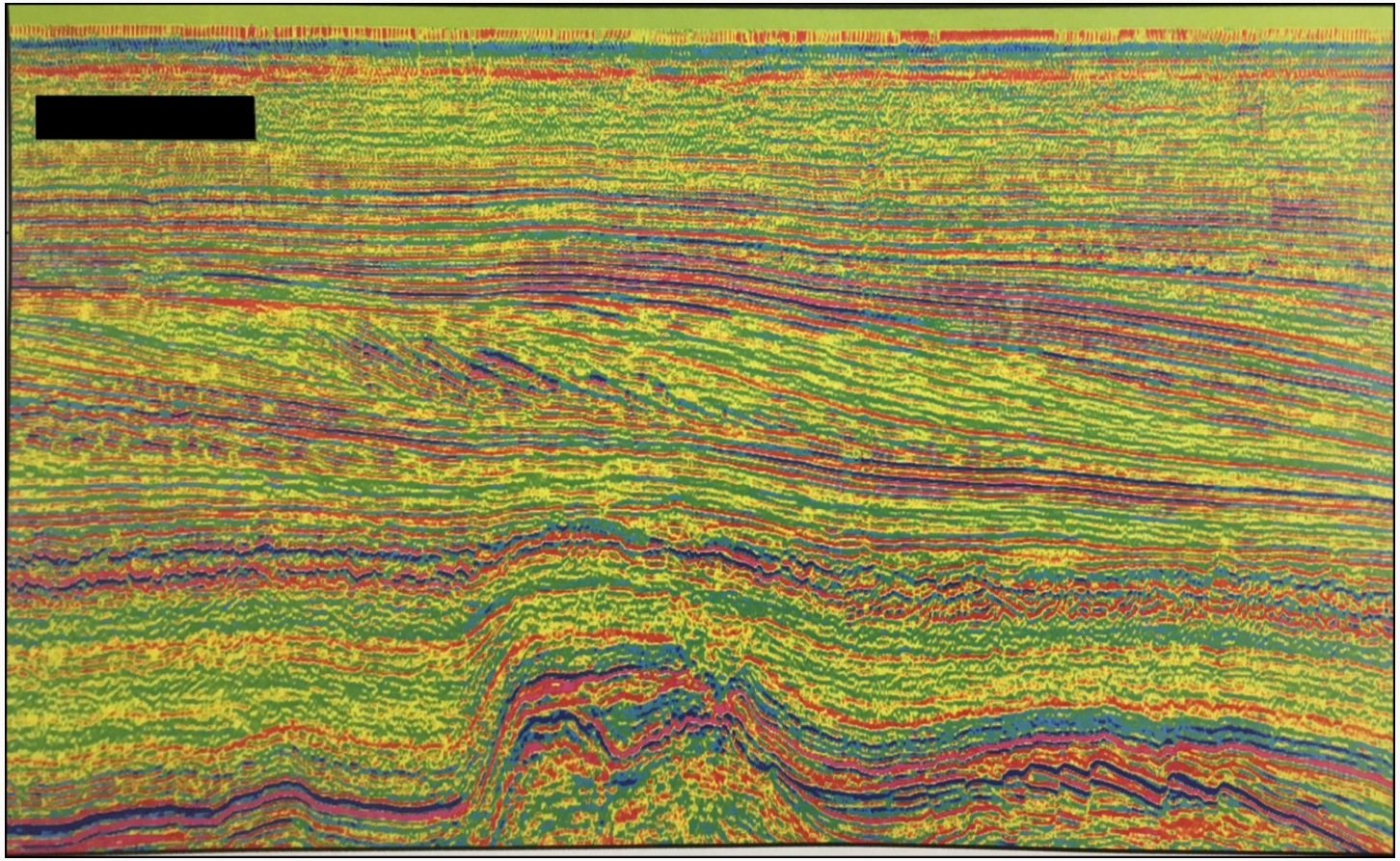
1. Preprocess, e.g. remove marginalia and annotations ⇐ `scikit-image`
2. Reduce colours by unsupervised *k*-means clustering ⇐ `scikit-learn`
3. **Sort the colour palette graph into order** ⇐ `pytsp` & `scipy.spatial`
4. Get index of each image pixel from sorted palette ⇐ `numpy`
5. Save a SEG-Y file of the recovered data & upload to S3 ⇐ `obspy` & `boto3`

Preprocessing in `scikit-image`

TNO.nl/dGBes.com, licensed CC-BY-SA



Preprocessing in `scikit-image`



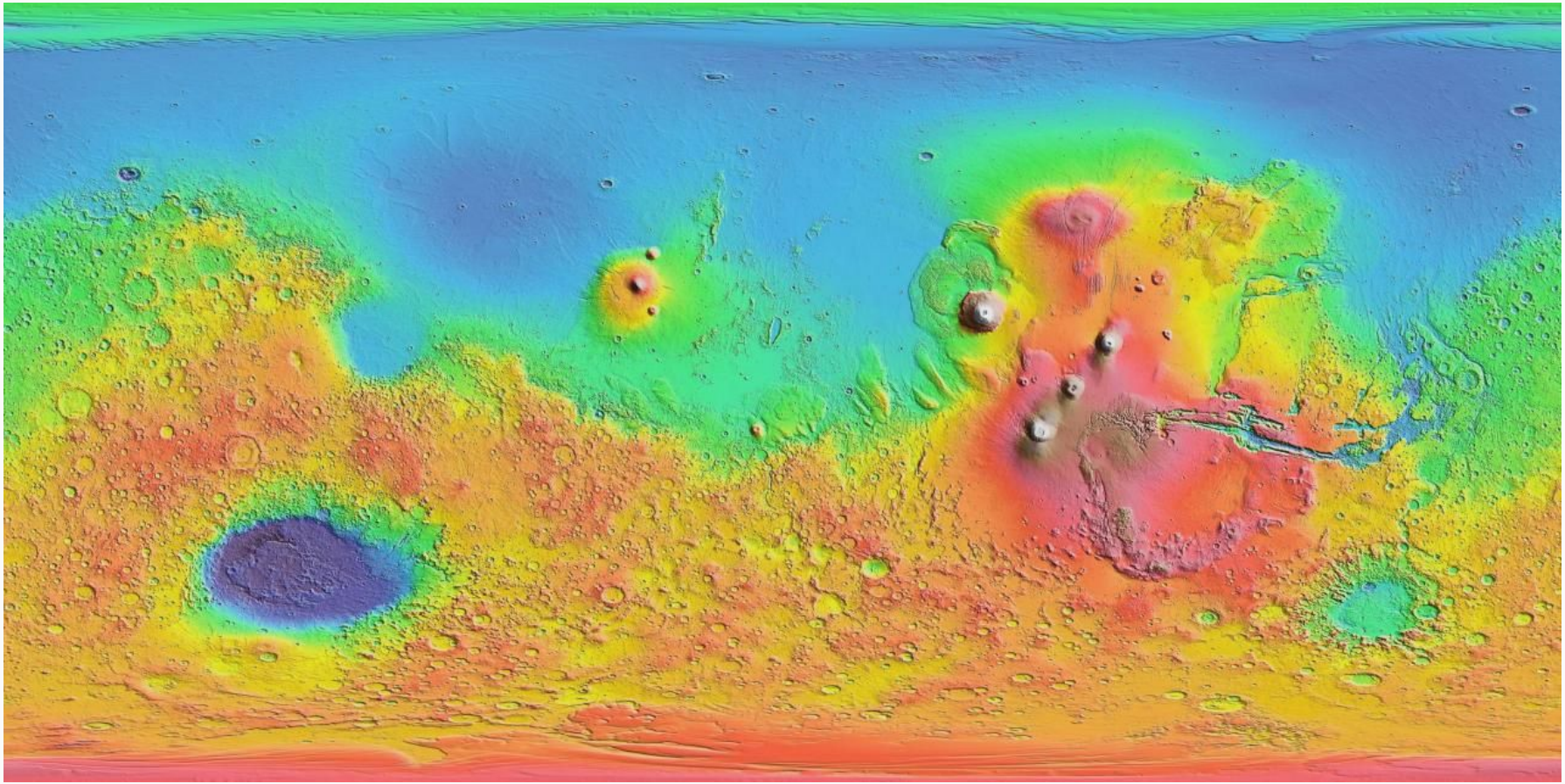
TNO.nl/dGBes.com, licensed CC-BY-SA

Limitations and difficulties

- Cannot recover original **magnitudes**.
- Cannot recover **polarity**: needs human input.
- Small and/or **lossy-compressed** images don't do well.
- Colourmap **interpolation** can cause problems.
- **Hillshading** causes problems.
- **Annotations** inside the data area can be problematic.

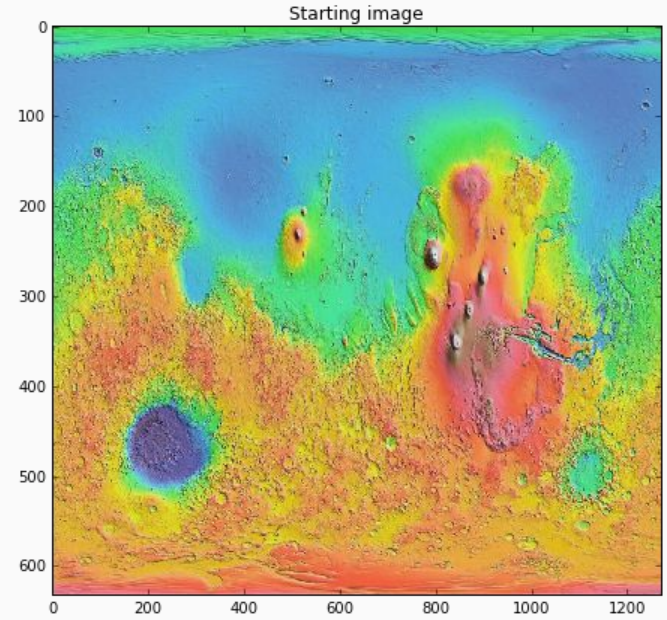
Hill-shaded topography

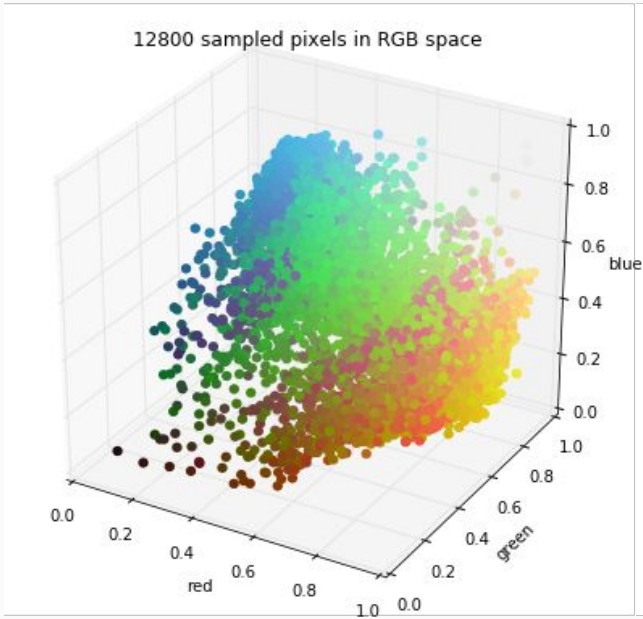
NASA-JPL / Mars Global Surveyor /MOLA

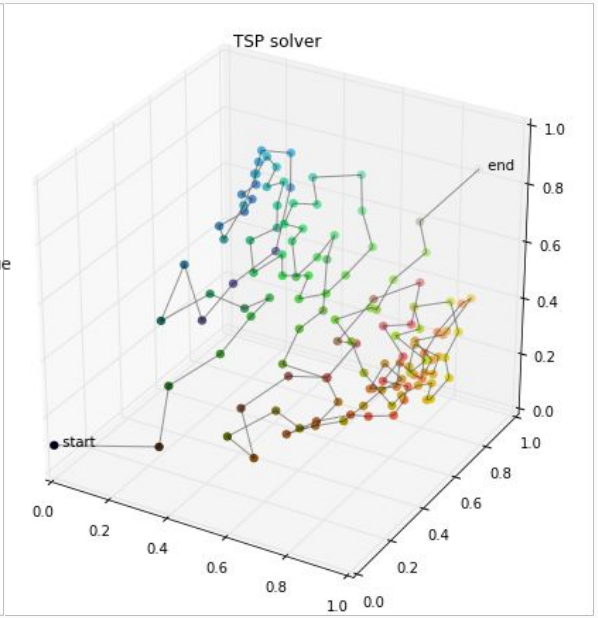


NASA-JPL / Mars Global Surveyor / MOLA

Hill-shaded topography







Summary

- This is a really hard problem to solve in a general way.
- I have something that works sometimes but fails on most real images.
- Next stop: deep learning.
- Code is at http://github.com/cseg/2017-hall-niccoli
- I am at matt@agilescientific.com

CC–BY 4.0 • Please share this work

References

dGB and TNO (1987). F3 seismic data, offshore Netherlands. Open Seismic Repository. http://www.opendtect.org/osr/. Licensed CC-BY-SA.

Helsgaun, K (2009). General k-opt submoves for the Lin–Kernighan TSP heuristic. *Mathematical Programming Computation*, 2009, doi: 10.1007/s12532-009-0004-6.

Pedregosa, F, et al. (2011). Scikit-learn: Machine Learning in Python, JMLR 12, pp. 2825-2830, 2011.

van der Walt, S, JL Schönberger, J Nunez-Iglesias, F Boulogne, JD Warner, N Yager, E Gouillart, T Yu and the scikit-image contributors (2014). Scikit-image: Image processing in Python. PeerJ 2:e453 doi: 10.7717/peerj.453

Appendix

International data rescue award

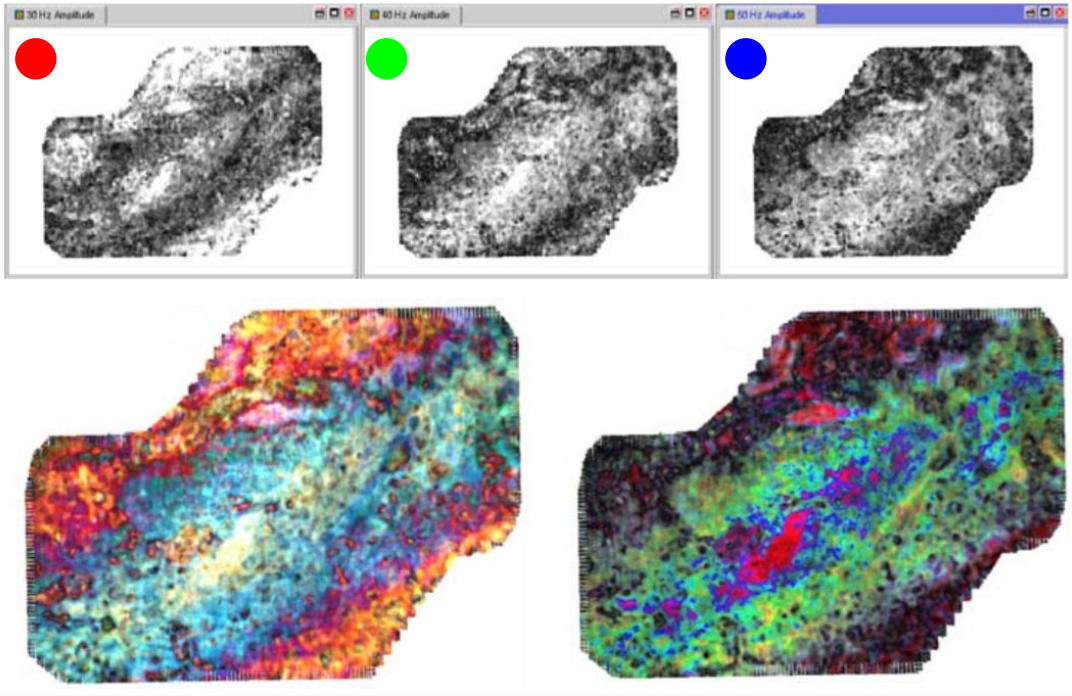
Interdisciplinary Earth Data Alliance

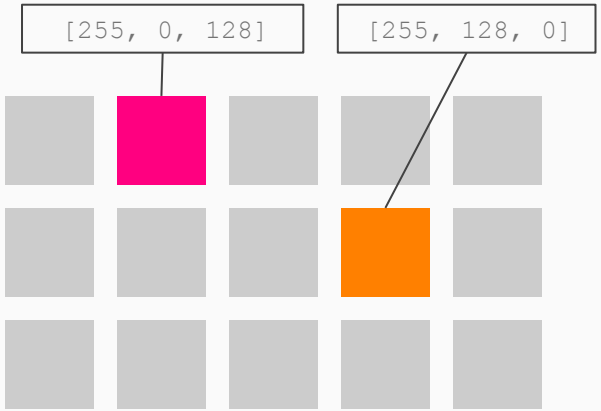
http://www.iedadata.org/

Winner in 2016 is digitizing the NSIDC's Roger G Barry photographic collection, to then extract historical glacier data.

Images are public domain

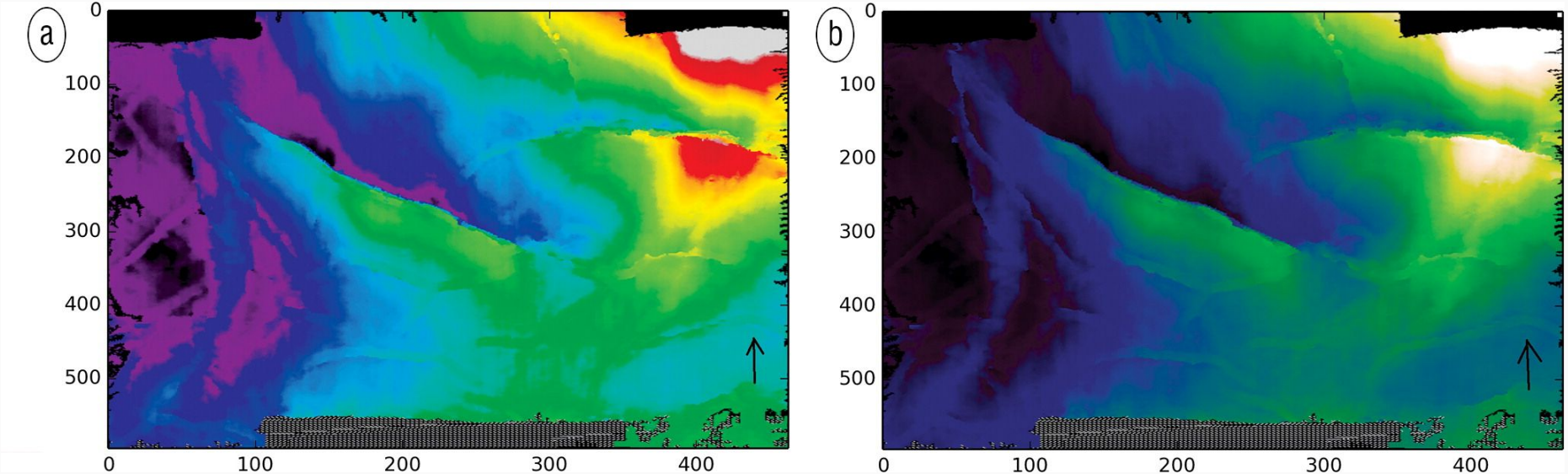
False colour images





Hall, M and E Trouillot (2004), CSEG Convention, Calgary, Canada.

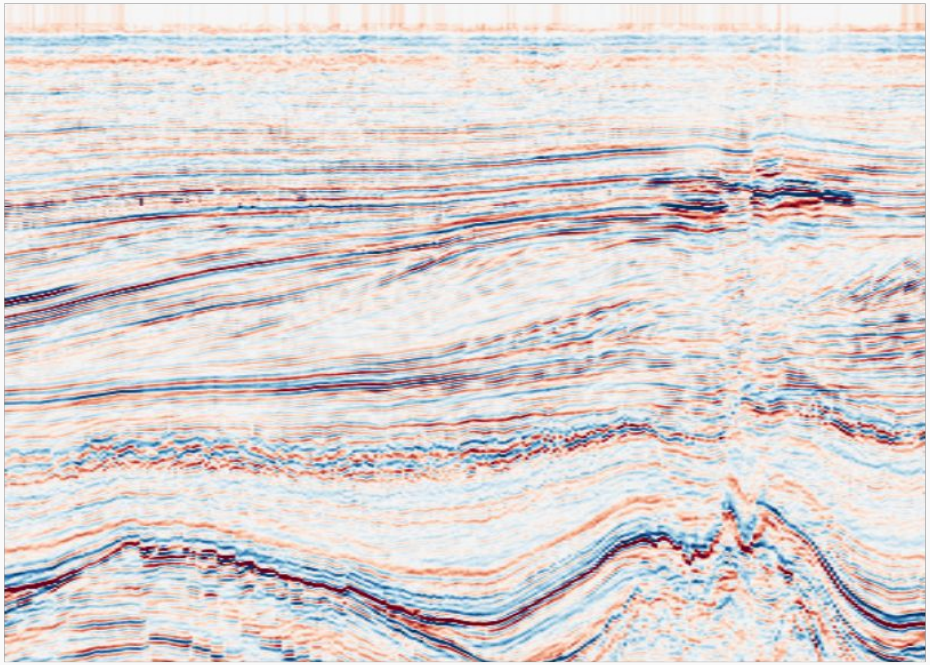
Pseudocolour images

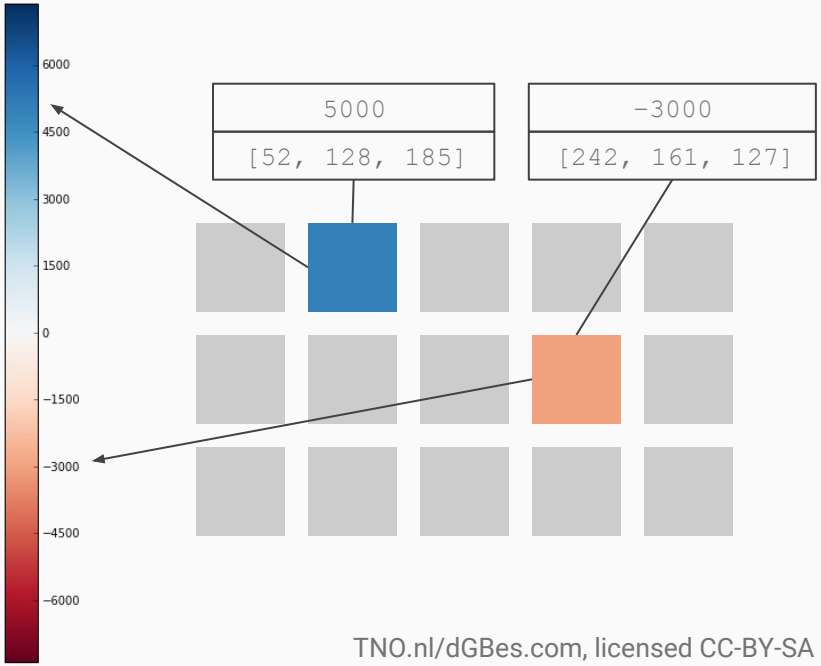


Niccoli, M (2014). The Leading Edge ***33*** *(8). CC-BY-SA.*

F3 seismic data from Open Seismic Repository

Pseudocolour images

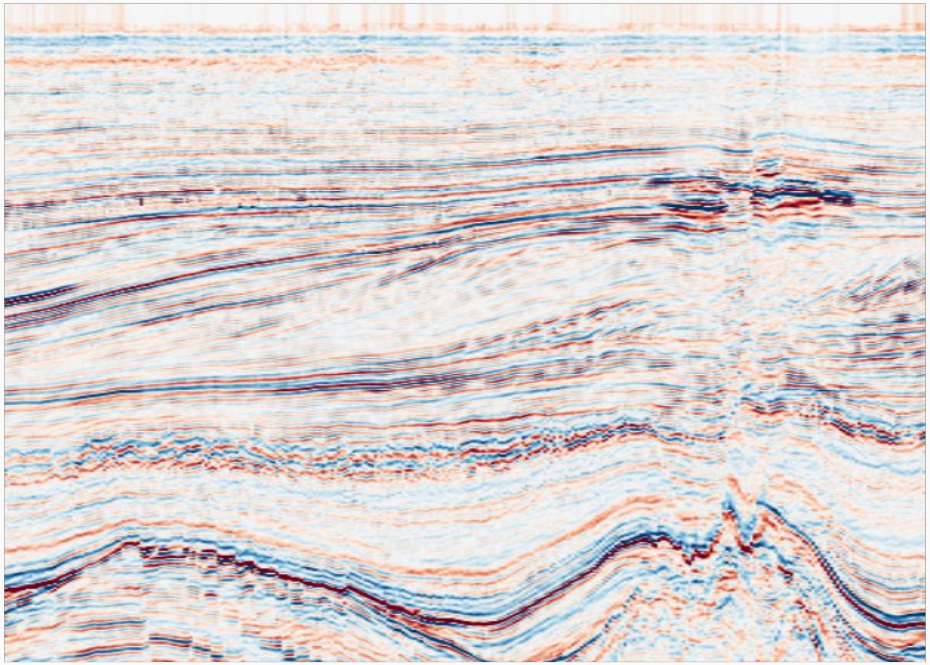


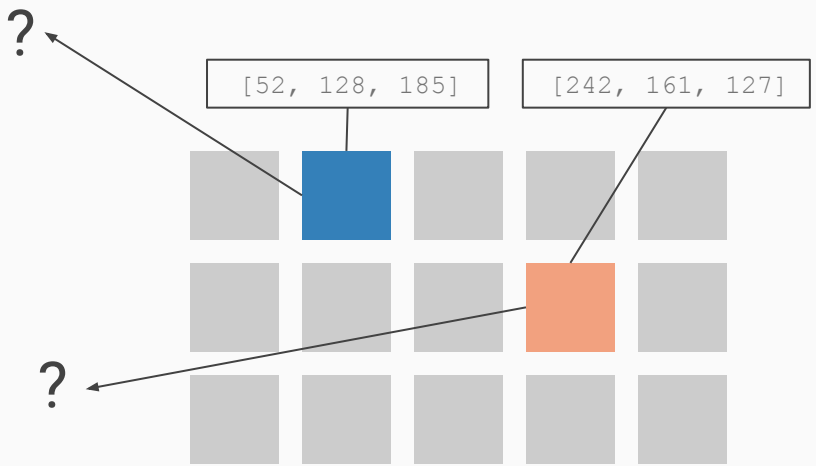


TNO.nl/dGBes.com, licensed CC-BY-SA

F3 seismic data from Open Seismic Repository

Pseudocolour images





TNO.nl/dGBes.com, licensed CC-BY-SA

Interpreting the distribution

Are nonlinearities in the RGB distribution due to the data, or the codebook?

DATA

```
[ 0.        ,  0.01879101,  0.03850932,  0.05892748,  0.07078996,
  0.08340478,  0.09904645,  0.1138833 ,  0.12537836,  0.13820313,
  0.15227014,  0.16596973,  0.17721984,  0.18673782,  0.20108477,
  0.21578165,  0.22876389,  0.23901671,  0.24772986,  0.25800017,
  0.26678331,  0.27747354,  0.28625667,  0.29573966,  0.30235325,
  0.30765462,  0.3119937 ,  0.31997201,  0.32793281,  0.33615607,
  0.34280465,  0.34961071,  0.35641676,  0.3646925 ,  0.3741055 ,
  0.38122649,  0.38719272,  0.39321144,  0.39819788,  0.40246698,
  0.40993789,  0.41649899,  0.42020821,  0.42505468,  0.42946374,
  0.43306797,  0.43779197,  0.44253346,  0.44841221,  0.45213892,
  ... etc.
```

Natural data, eg maps

CODEBOOK

```
[  0,   1,   2,   3,   4,   5,   6,   7,   8,   9,  10,  11,  12,
  13,  14,  15,  16,  17,  18,  19,  20,  21,  22,  23,  24,  25,
  26,  27,  28,  29,  30,  31,  32,  33,  34,  35,  36,  37,  38,
  39,  40,  41,  42,  43,  44,  45,  46,  47,  48,  49,  50,  51,
  52,  53,  54,  55,  56,  57,  58,  59,  60,  61,  62,  63,  64,
  65,  66,  67,  68,  69,  70,  71,  72,  73,  74,  75,  76,  77,
  78,  79,  80,  81,  82,  83,  84,  85,  86,  87,  88,  89,  90,
  91,  92,  93,  94,  95,  96,  97,  98,  99, 100, 101, 102, 103,
 104, 105, 106, 107, 108, 109, 110, 111, 112,113, 114, 115, 116,
 117, 118, 119, 120, 121, 122, 123, 124, 125,126, 127]
```

Bandlimited data, eg seismic

```
dx, ix = scipy.spatial.cKDTree(c).query(im)
```

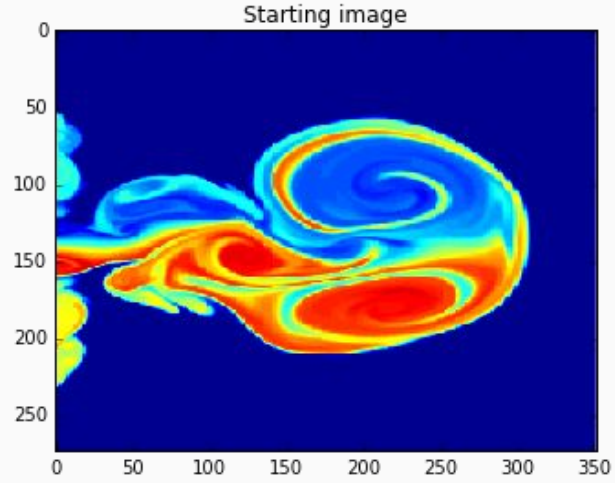
Recover data from KDTree

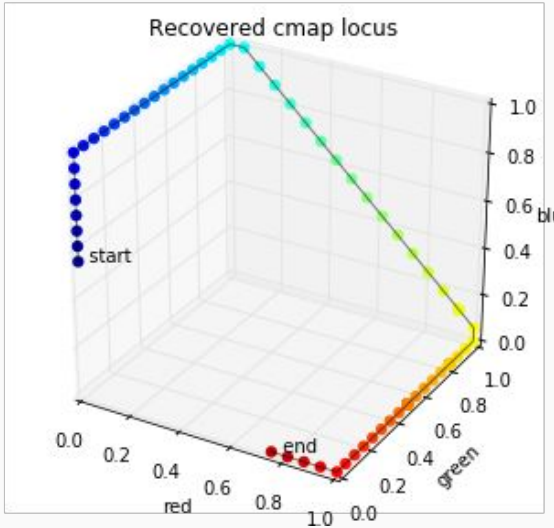
- A k-D tree is a data structure for fast spatial indexing based on norms.
- We make a k-D tree for the codebook.
- For each pixel in the image, we find the nearest entry in the codebook.

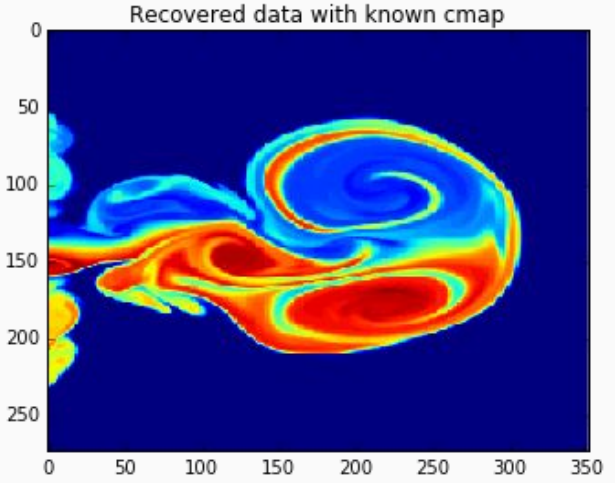
`scipy.spatial.cKDTree()` is 260 times faster than `scipy.spatial.KDTree()`

`cmap = 'jet'` (most similar)

Fluid dynamics





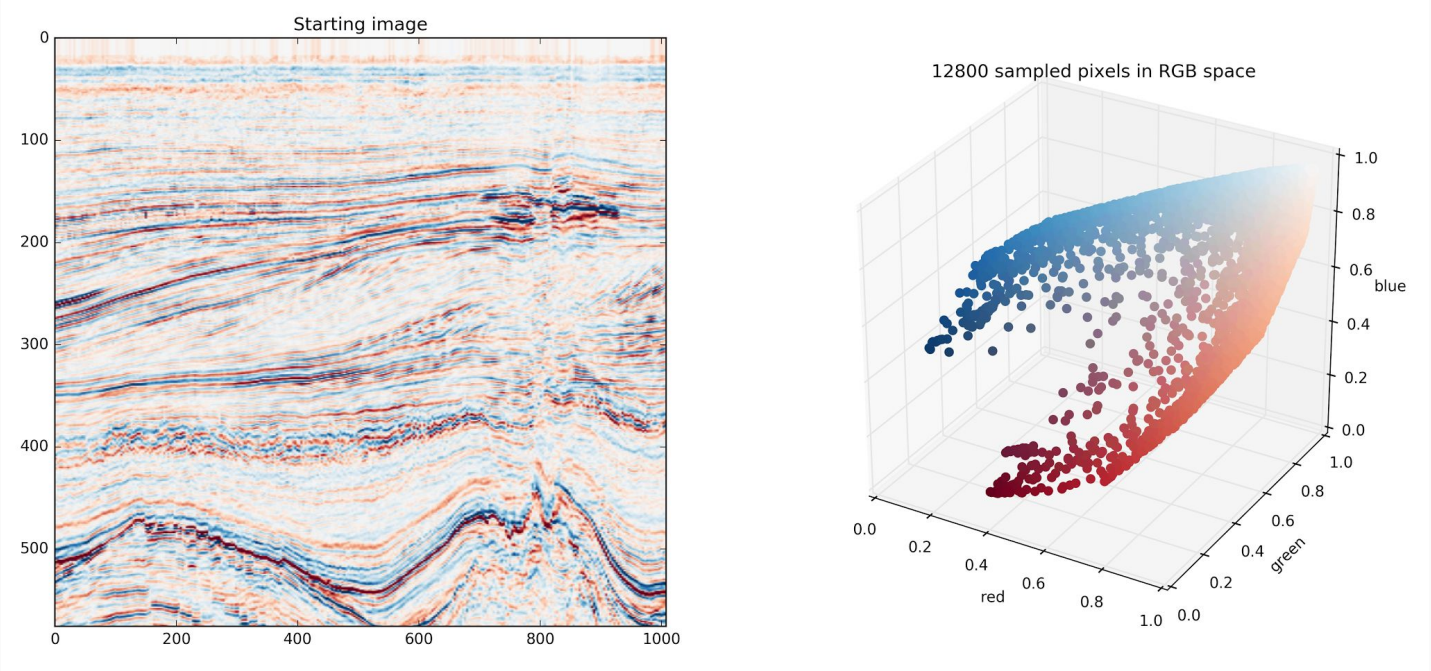


Things that failed

- Detecting the locus with 3D convex hulls.
- Multivariate adaptive regression splines (MARS), via `py-earth`.
- Colour quantization with PIL, NeuQuant, and in CIELab space.
- Colinearity constraint for self-crossing colourmaps.

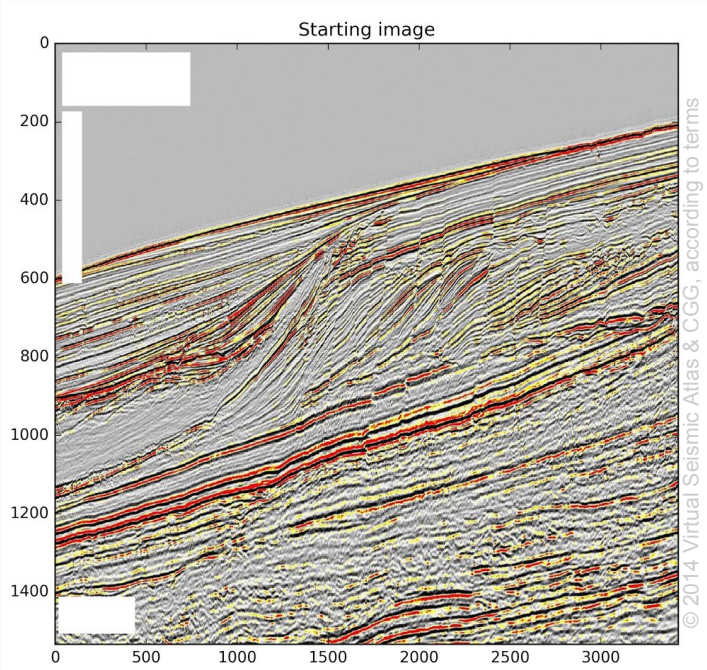
Failure cases

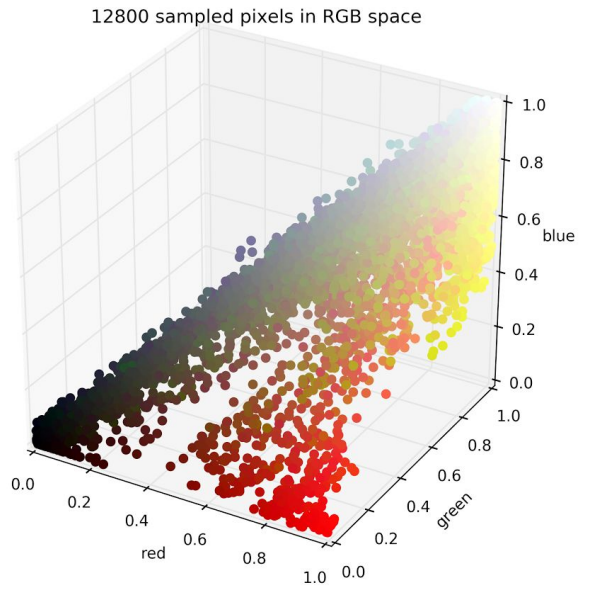
Interpolated display



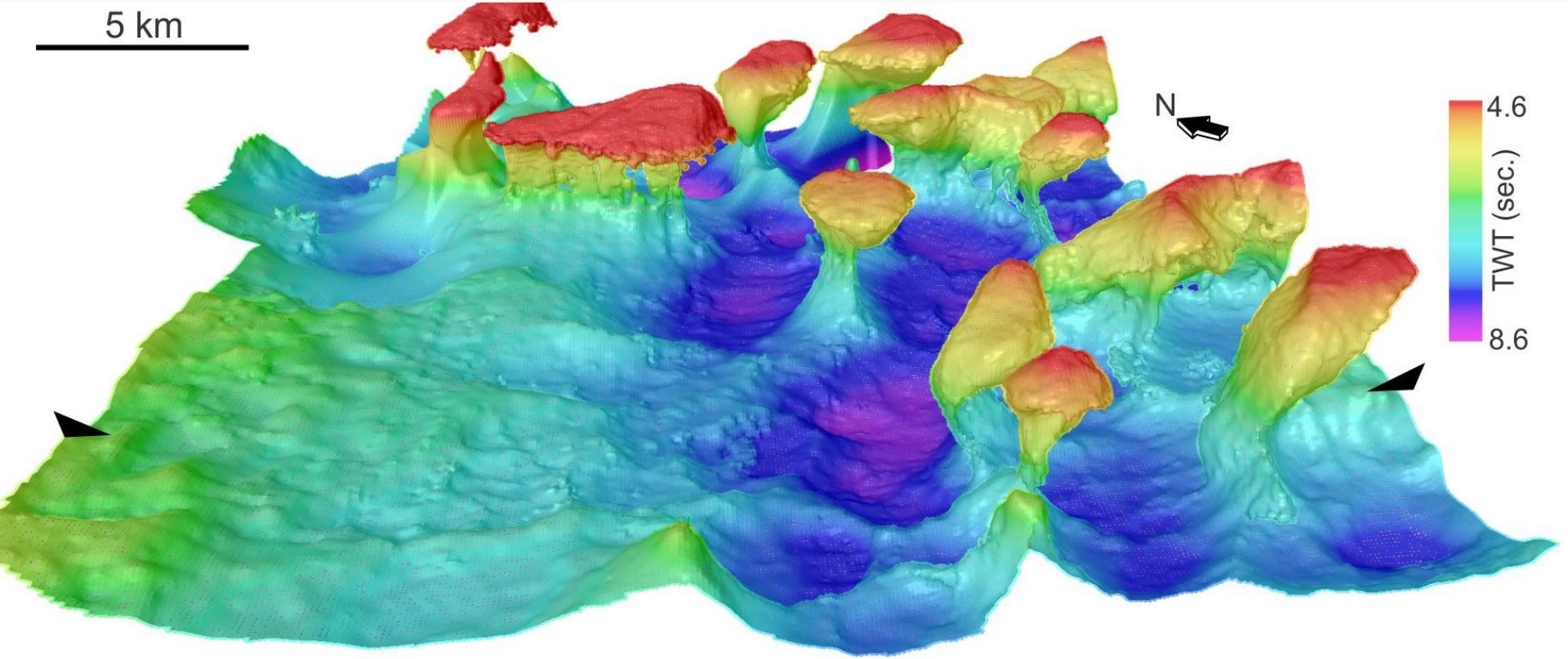
Failure cases

Lossy compression (e.g. JPEG)





3D shaded perspective



Mark Deptuck / Twitter

Thrumcap 3D (CNSOPB Program number NS24-S6-1E/2E)

Hill-shaded topography