

IQmulus (2012-2016)

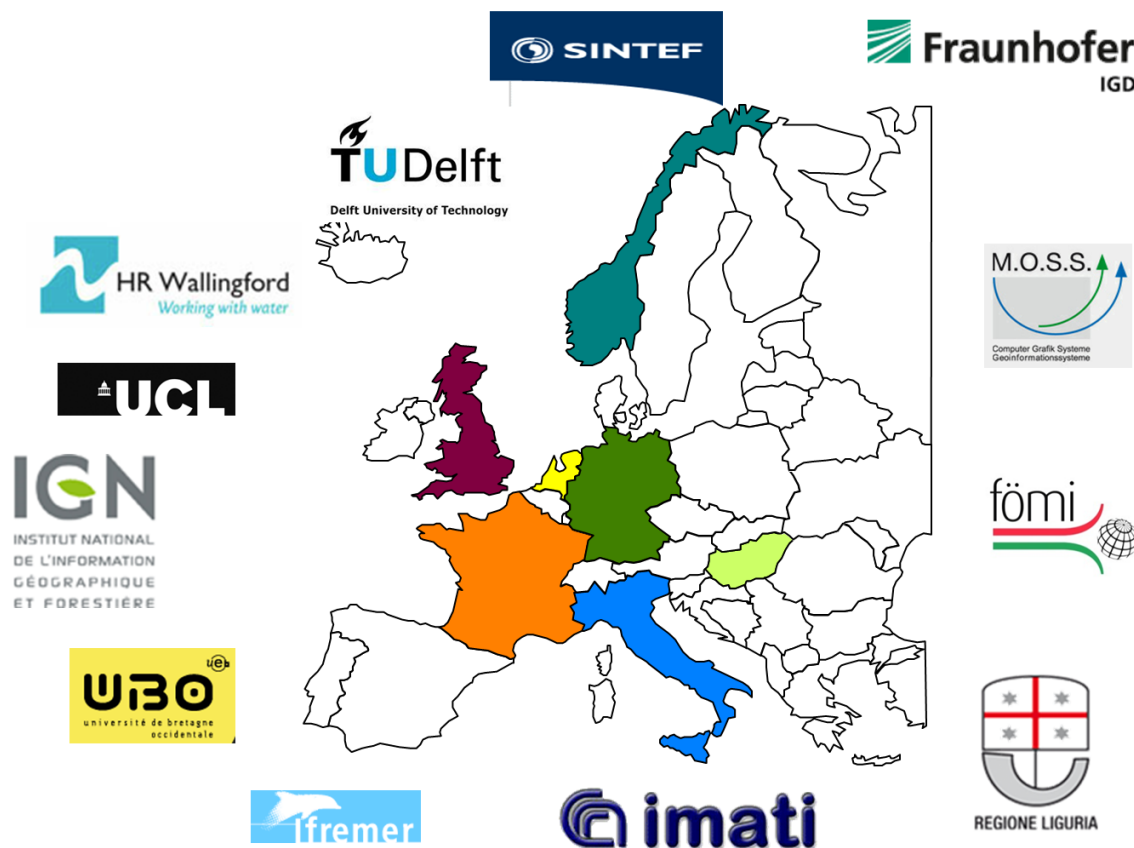
A high-volume fusion and analysis platform for geospatial point clouds

Tor Dokken, SINTEF Digital
IQmulus Coordinator

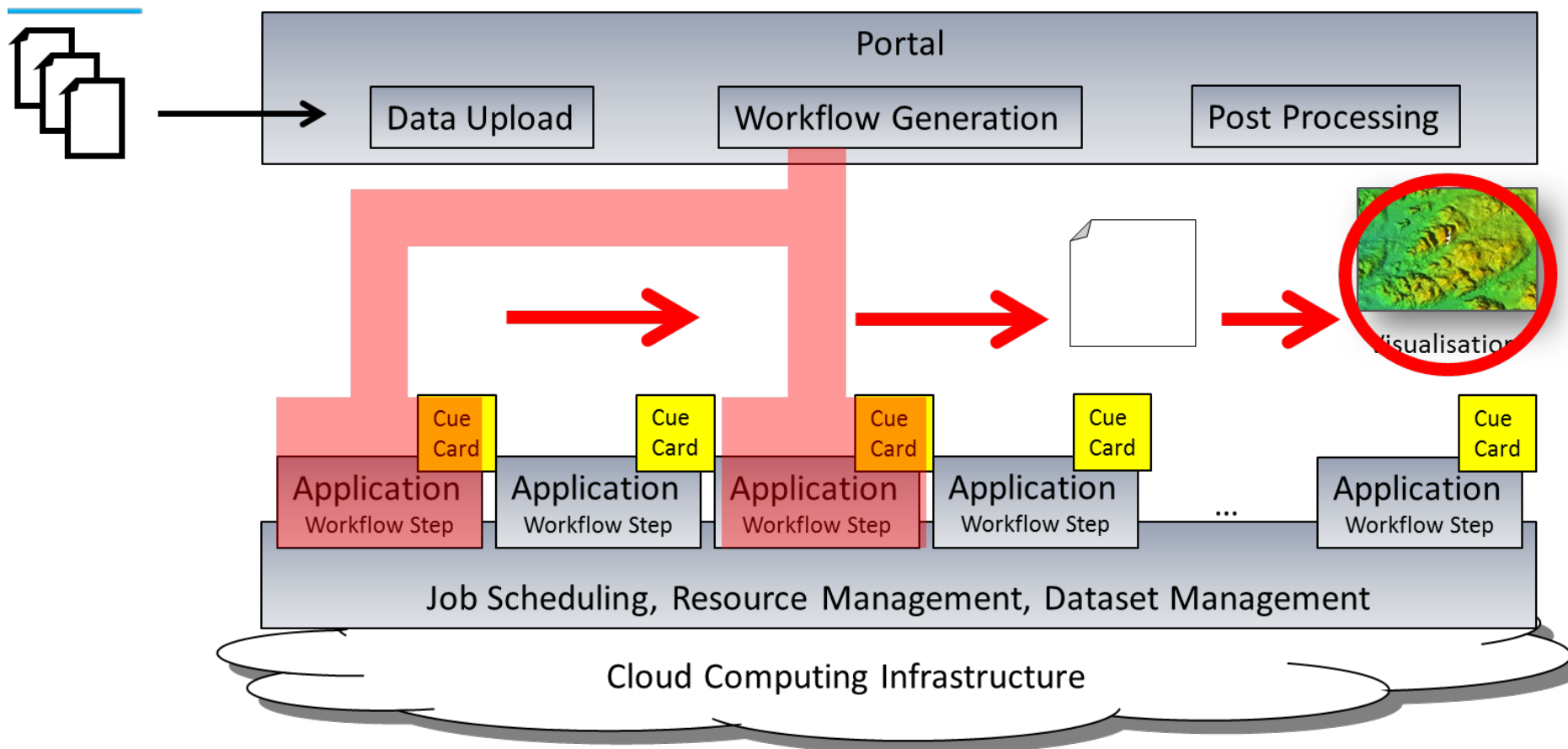


IQmulus (2012-2016, FP7 IP)

- 12 partner from 7 countries
- Funding 7.1 M€
- November 2012- October 2016
- From geospatial big data to smart data
 - Analytics
 - Storage
 - Cloud
 - Workflows

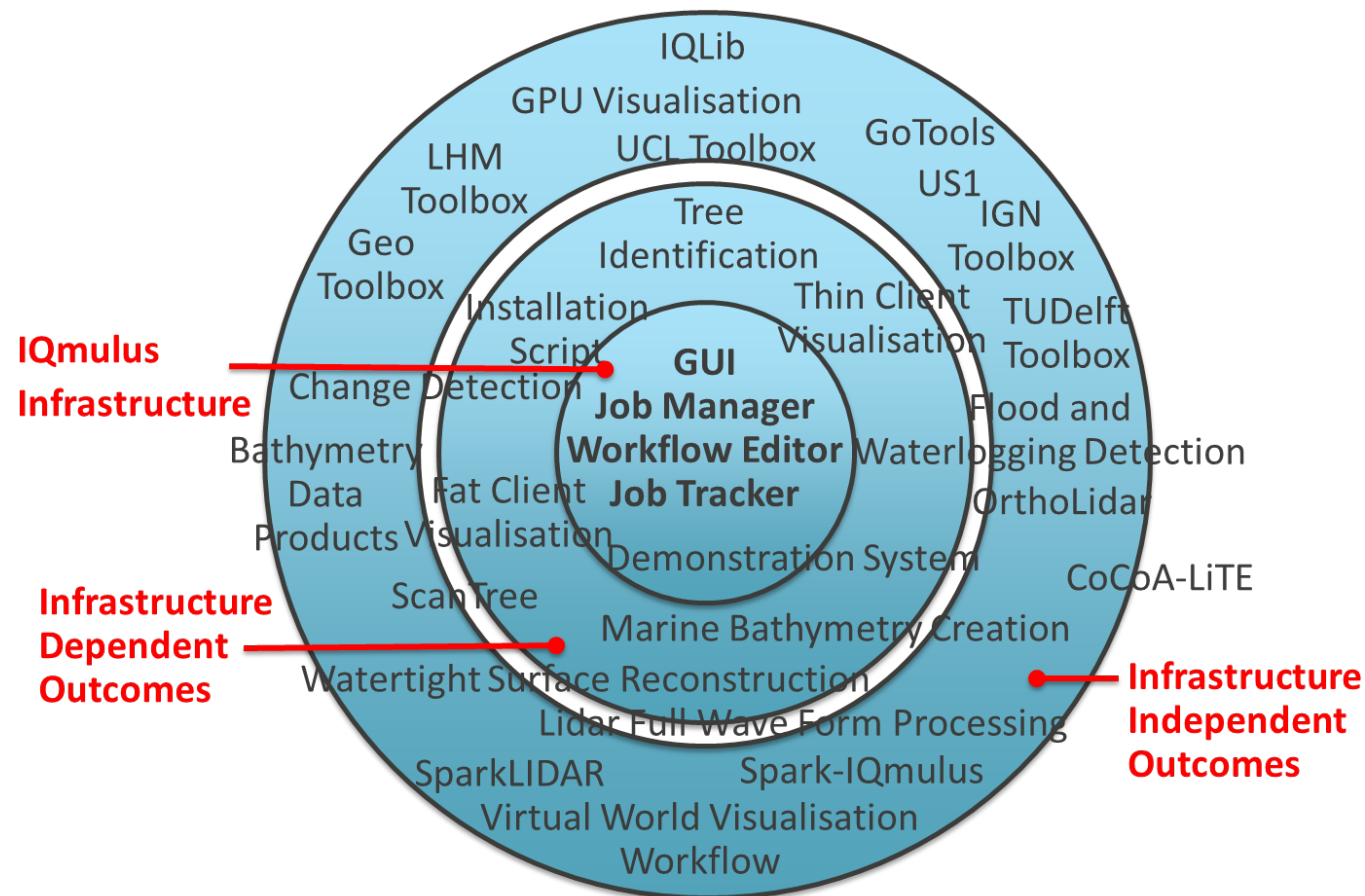


IQmulus infrastructure, algorithms implemented in applications



IQmulus exploitable outcomes

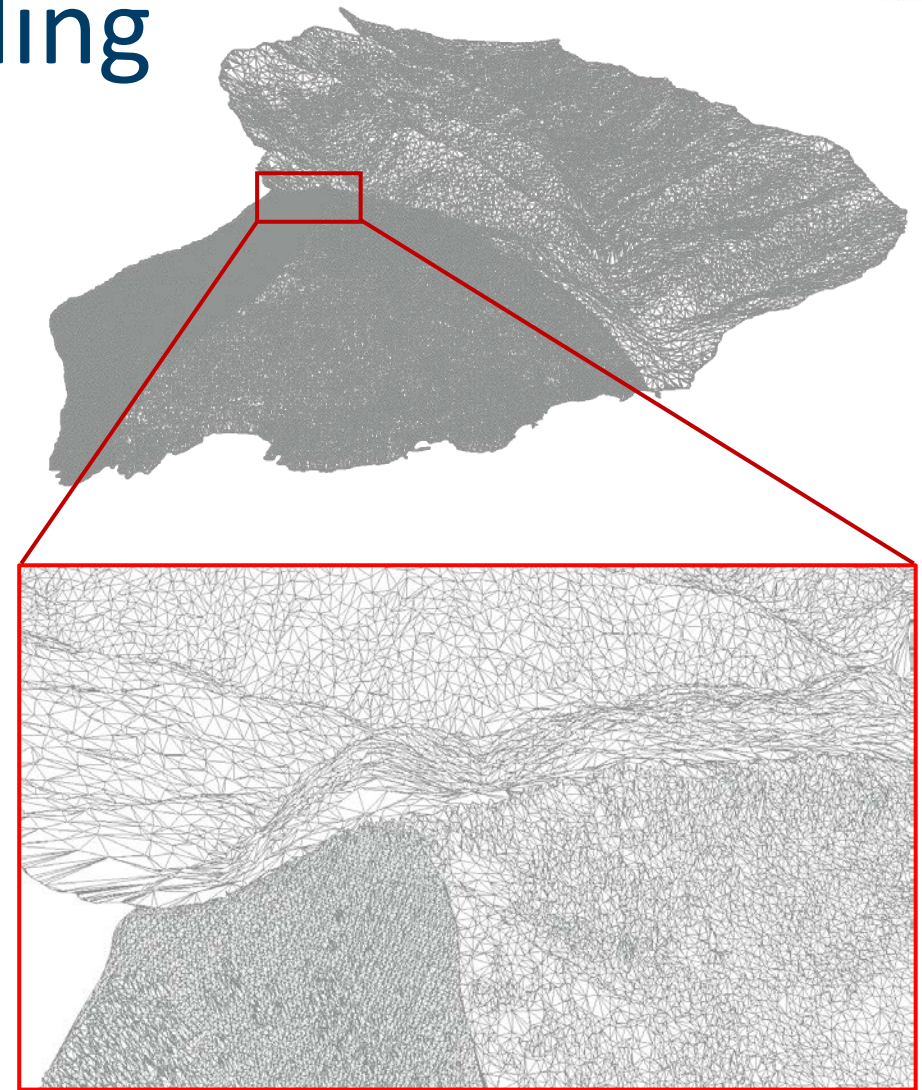
- I will focus on infrastructure independent outcomes related to showcases
 - Land showcases (LS1, LS2, LS3)
 - Urban show cases (US2, US3)
 - Marine showcases (MS1, MS2)
- I will also address how parts of the work continues in the Norwegian ANALYST project (2017-2021)



LS1: Multi-resolution modelling

(CNR-IMATI, IGN, FOMI)

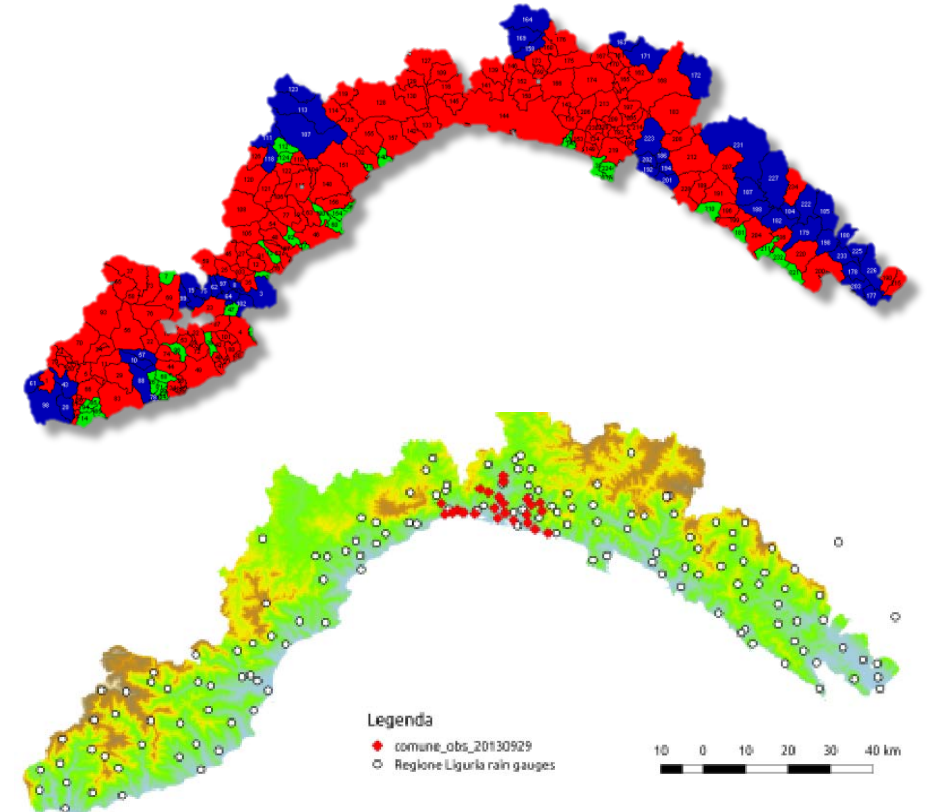
- **Sori area** (Drainage basins)
 - Basin 82 at LOD1
 - Basins 77-81 at LOD5
 - Basin 76 at LOD10
- **Beyond beyond the state-of-the-art**
 - Indexing, pre-processing: filtering, outlier, simplification
 - Preparation of the level of detail (LOD) structure
 - Implemented for a distributed execution framework
- **IQlib** [//github.com/posseidon/IQLib](https://github.com/posseidon/IQLib)
 - Generalize the support to tiling and stitching to *local*, *focal*, and *zonal* data partitioning of large data sets



LS2: Rainfall interpolation & analysis

(CNR-IMATI, Regione Liguria, Italy)

- Aims
 - definition of a rainfall map at small catchments
 - integration of rainfall measures with data acquired by remote sensors (radar, satellite)
 - visualization of the temporal evolution of precipitation fields and the track of its maxima
- Main issue: integration of heterogeneous & time-varying data
- Results: computation of a continuous approximation
 - from time-varying, sparse, and heterogeneous rainfall data from rain gauge networks
 - at different scales (urban/regional areas), time steps, and spatial distribution of the rainfall stations.

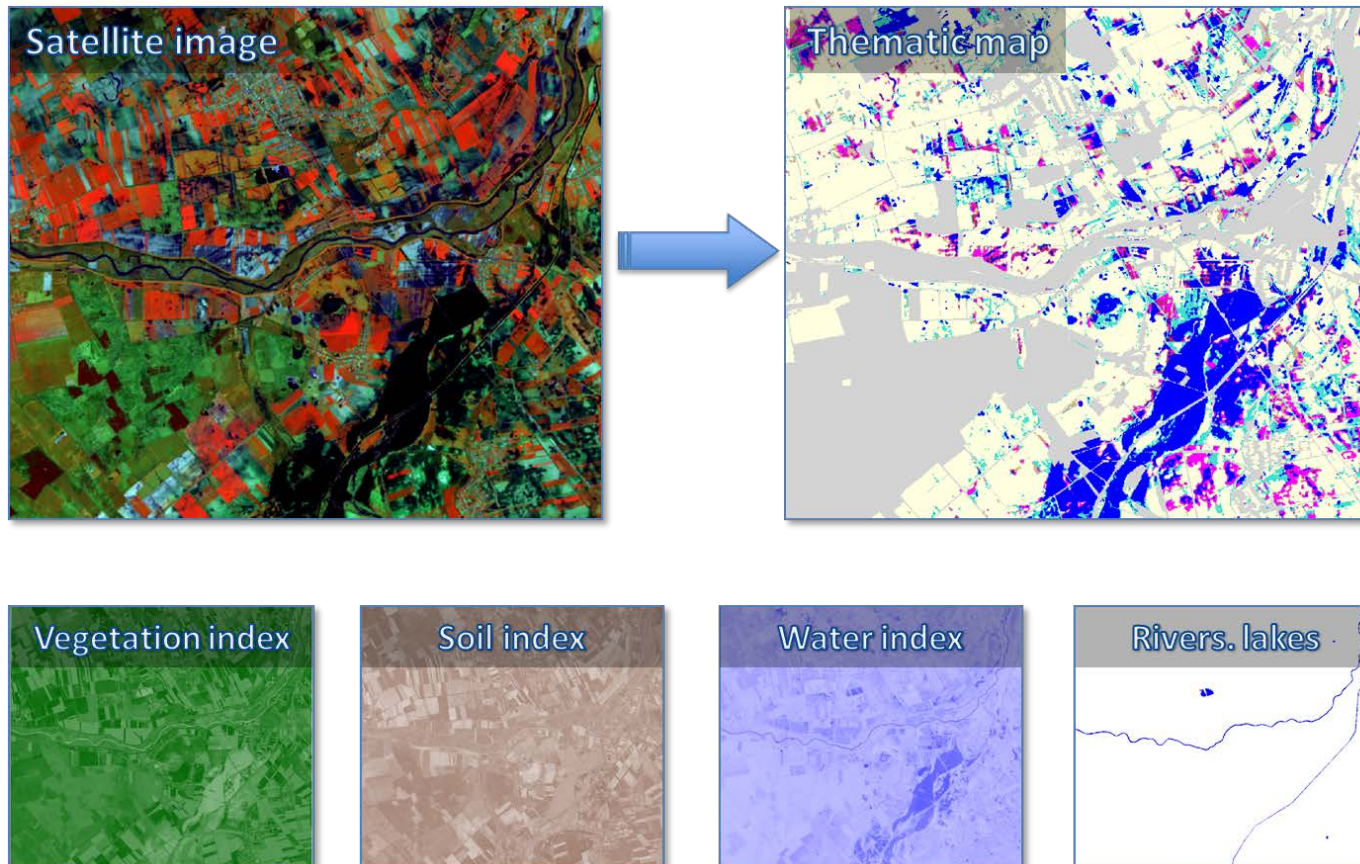


LS 3: Flood and waterlogging detection

(FÖMI, Hungary)

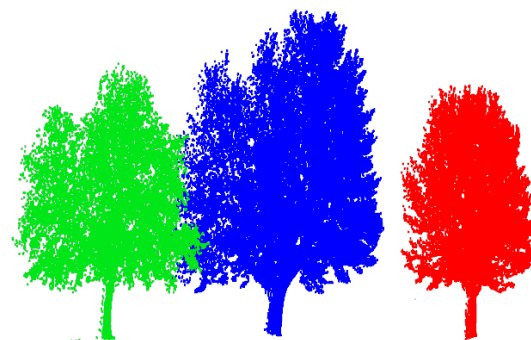
- Distributed and automated processing of satellite imagery for mapping flood and waterlogging is implemented
- Automatic preprocessing implemented based on image metadata parsing
- Works for *Sentinel-2*, *Landsat* and *SPOT* satellite imagery – can be extended to other sensors
- Algorithms for image analysis are implemented on a uniform platform – quicker, more efficient than the previous solution (*~8 mins vs. 25 mins for an LS8 scene*)
- Important exploitation path for the post-project period: replacement of the old solution with IQmulus LS3 workflow

LS 3: Flood and waterlogging detection



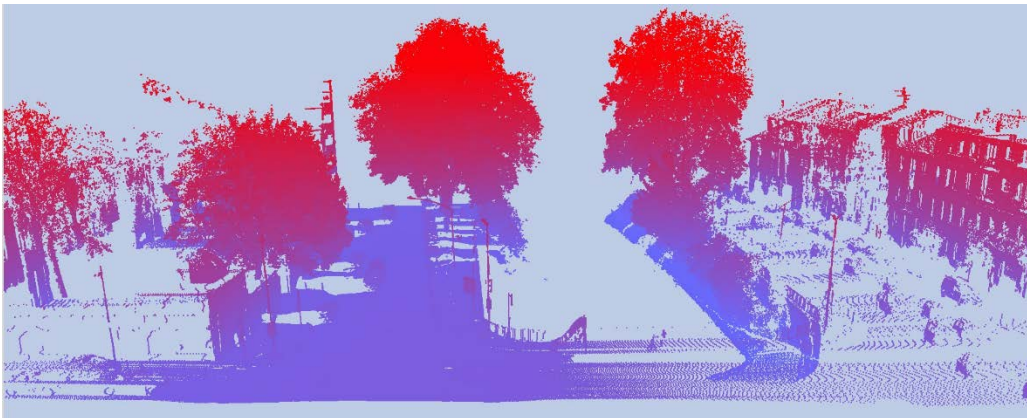
US2: Individual tree extraction from urban LMMS data (IGN, UCL, TUDelft)

- **Objective:** Identify single trees in LMMS point clouds sampling urban environments by:
 - reporting their locations
 - outputting all points sampling trees, organized at individual tree level
- **Need:** Individual tree parameters are high in demand by local and regional authorities for:
 - Street inventory management and city climate assessment
 - Tree cadaster and biomass monitoring requirements
 - Clearance regulations and hazard protection



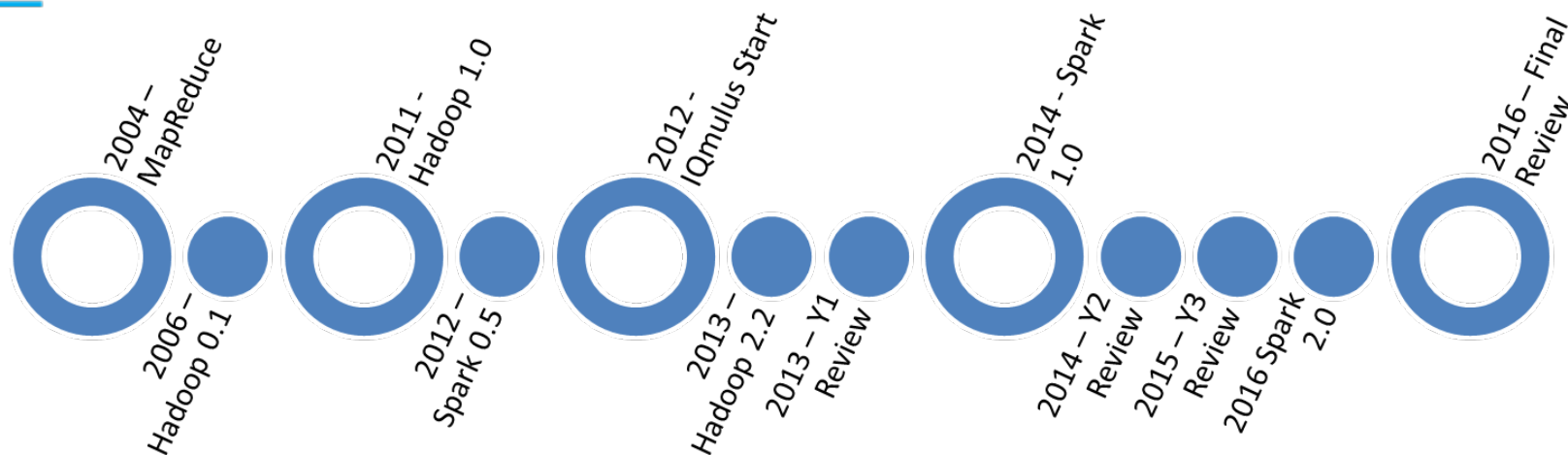
10 km of Mobile Mapping data sampling Toulouse (IGN, Stereopolis system)

- Acquisition time: 121 GB in 2 hours
- Number of tiles: 517 tiles of ~ 3 million points each



4000+ tree locations automatically identified by the US2 workflow.

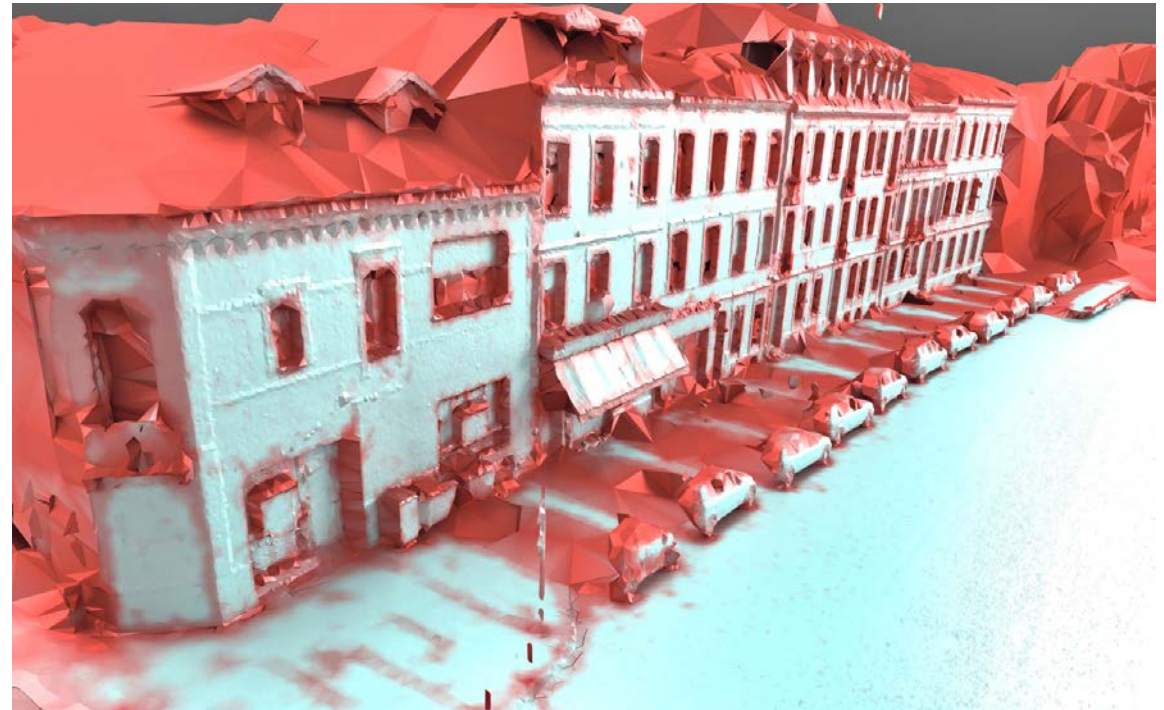
Apache Spark for US2 (UCL, IGN)



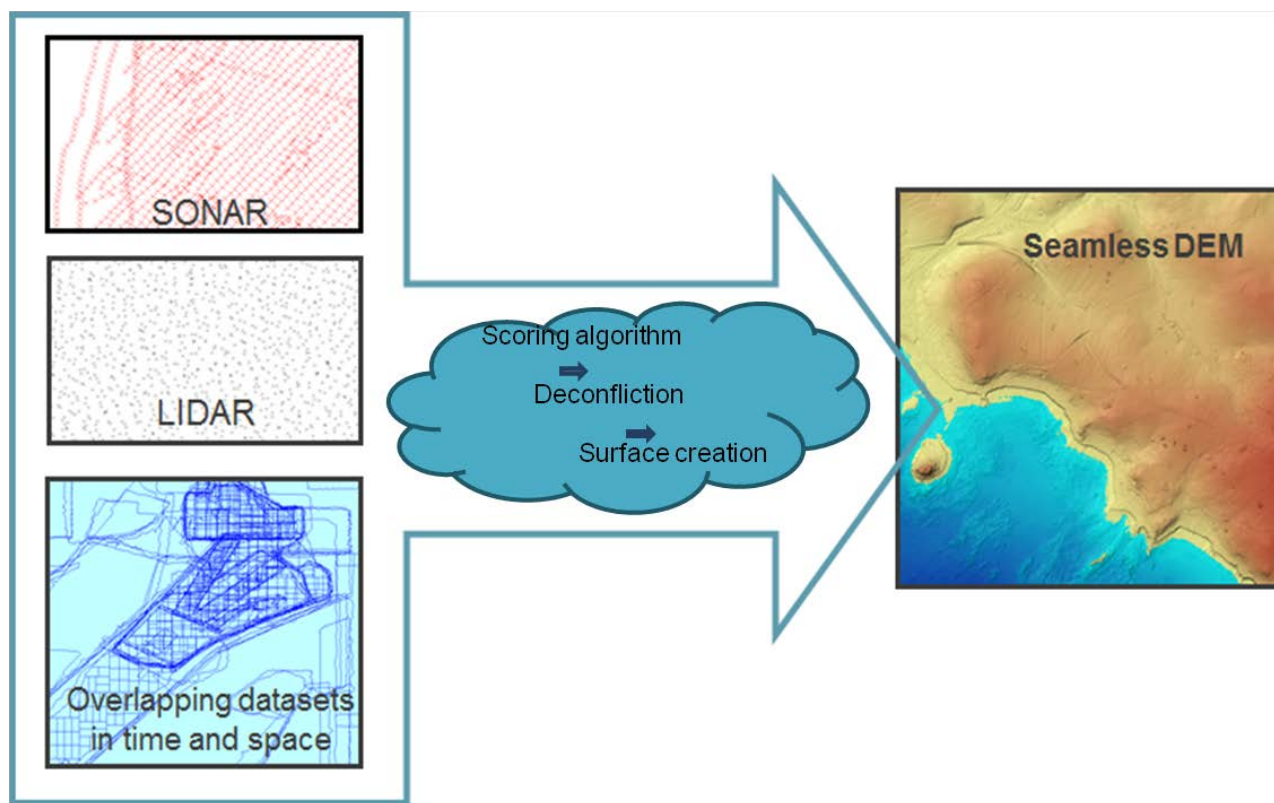
- “Apache Spark - Lightning-fast cluster computing”
 - 100x faster than Hadoop MapReduce in memory, or 10x faster on disk
- Spark-IQmulus Library (IGN)
 - Open source library (Apache 2.0 license) to read and write PLY, LAS and XYZ LIDAR point clouds
- SparkLIDAR (UCL)
 - Point cloud processing framework based on Apache Spark
 - Interfaces to point cloud processing libraries & machine learning libraries
 - Includes functionality for data ingestion, indexing & partitioning

US3: 3D watertight mesh generation with uncertainties (IGN)

- Input
 - I Aerial Lidar
 - I Mobile Mapping Lidar
 - I Photogrammetry
- Output : Watertight surface
 - Flooding simulation
 - 3D City modeling and updating
 - I Indoor / outdoor /underground modelling (security, trafic...)
 - ... first step to higher level
 - processing !

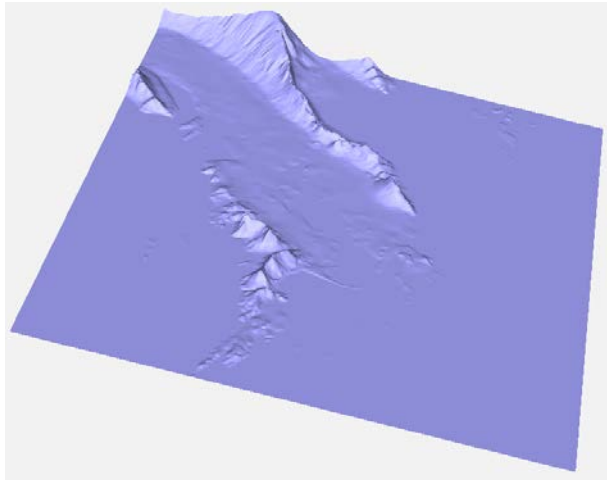


MS1(2) – Surface generation from bathymetry data (SINTEF, HRW)

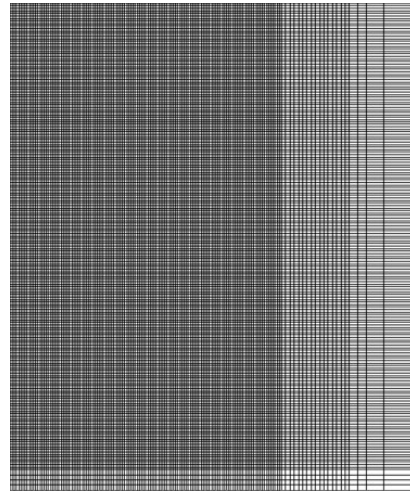


New technology: Locally refined B-splines

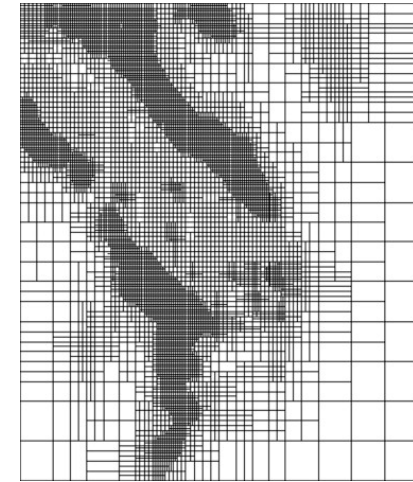
- Piecewise polynomial surfaces
- Locally refinable ➡ adaptive surface approximation
- Increase data size only where needed ➡ huge potential for data reduction
- Generally smooth surfaces with flexibility to represent local detail



Surface (topography)



Polynomial patches, tensor-product spline surface



Polynomial patches, LR B-spline surface

Traditional spline based methods

- Cubic spline interpolation is an efficient method for interpolating points to create a curve composed of a sequence of polynomial pieces (1960s)
- NonUniform Rational B-splines (NURBS) is used for curve and surface representation in Computer Aided Design System (1970s and 1980s)
- B-spline related research bloomed in the period 1970-1995, then CAD-standardization allow less innovation

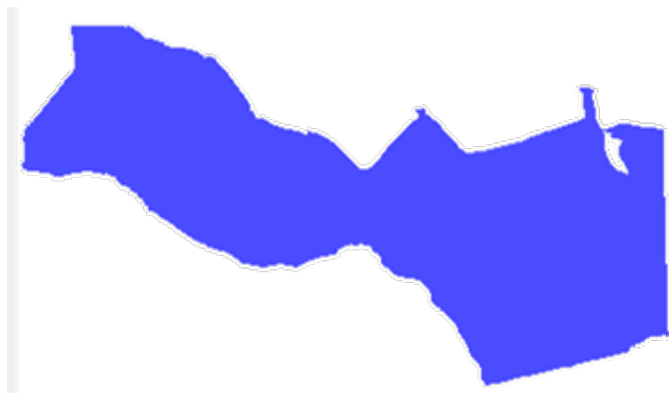
Locally refined splines

- Generalization of tensor product B-splines used in Computer Aided Design
- T-splines (2004) targeting creative design
- Locally refined B-splines (2013),
 - First targeting Isogeometric Analysis (IGA) (Finite element Analysis based on B-splines)
 - Then used for addressing analysis of big-data: Locally refined multi-variable spline regression
- A patchwork of polynomial pieces refined to address the complexity of the data to be approximated/represented (in 2, or more variables)
 - Avoids the explosion of representation bulk inherit in tensor product B-splines
 - A new locally refined spline regression method
- Locally refined splines open up new application areas for B-spline technology
 - Multi-variate data
 - New approaches for data analysis and AI

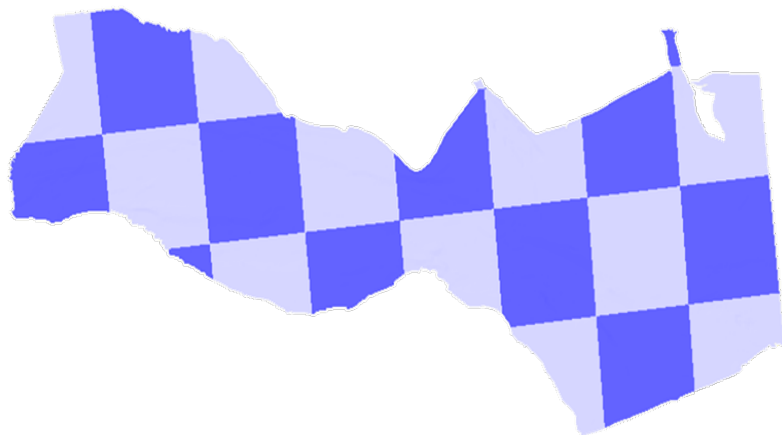
Spline based methods for analytics

- Locally refined splines are very efficient for representing the smooth component of large data sets
- Analysis uses
 - Extract the non-smooth data-points by subtracting the spline representation from the data set and threshold
 - Analyse the smooth component by looking for mathematical features (inflections, min, max,...)
 - The refinement structure provides information on the distribution of variation in the data set
 - Efficiently analyses differences between new old and new data
 - Create regularized data for AI and deep learning

Tiling



One data survey: 131 million points

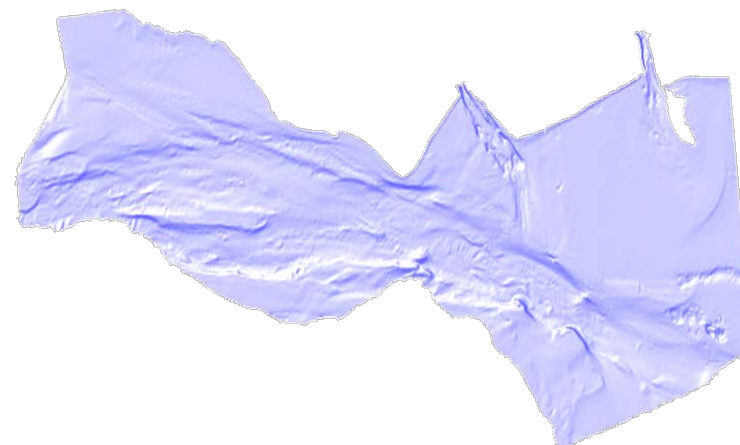


Surface set approximating data set after being split into regular tiles

Many data sets are too large to approximate by a single surface

- Many operations for each single point (execution time)
- Memory limitations (risk of crash)
- Very large surfaces are cumbersome to handle (execution time of subsequent operations)

Tiling enables parallelization on computation nodes



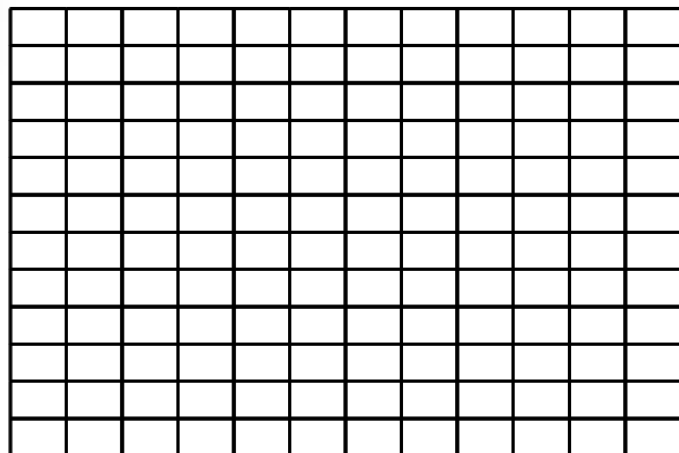
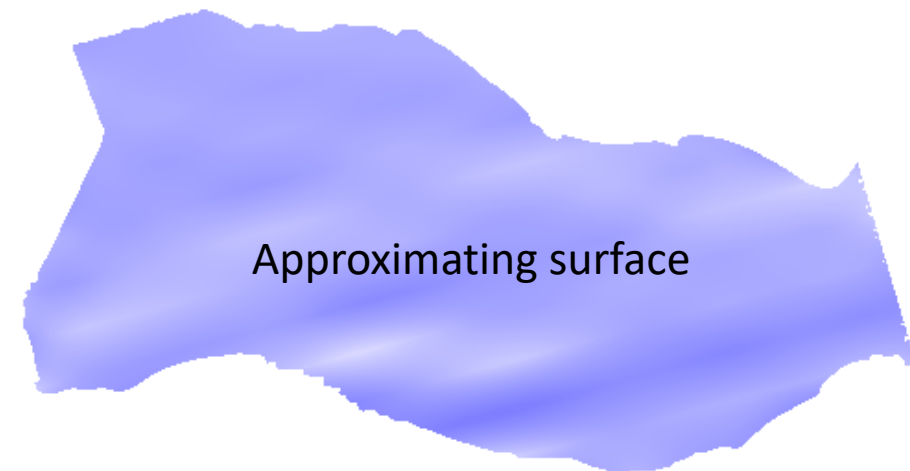
C^1 seamless super surface

Initial approximation of 14.6 mill points (280 Mbyte)

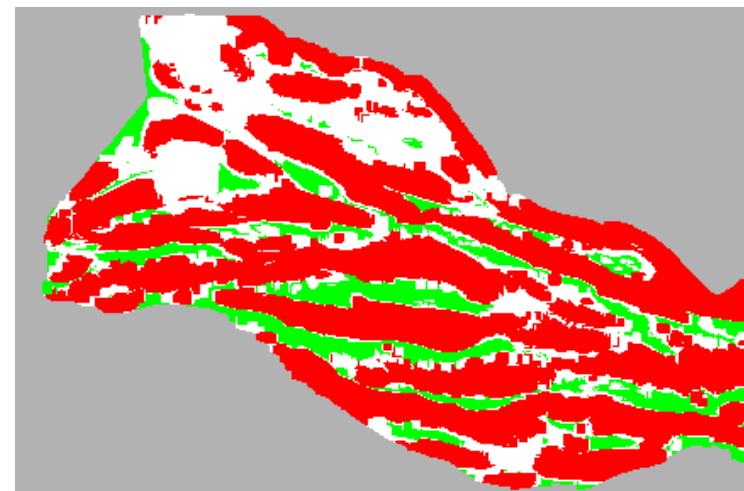
Number of points	14.6 mill
No. of coefs.	196
Surface file size	26 KB
Max. dist	12.8 m.
Average dist	1.42 m.
# points, dist > 0.5 m	9.9 mill

Data courtesy HR
Wallingford, SeaZone

Elevation interval: ~50 m.



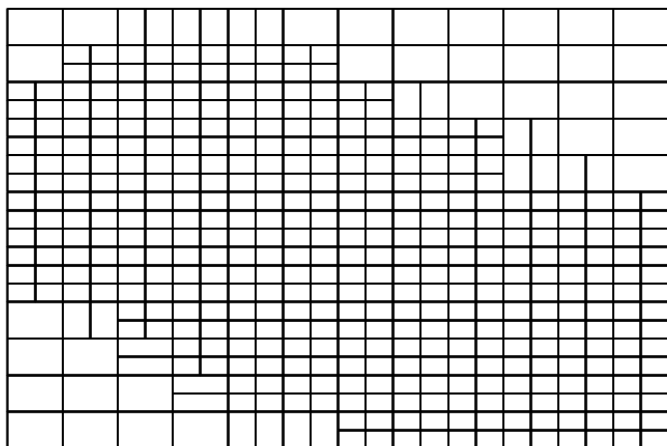
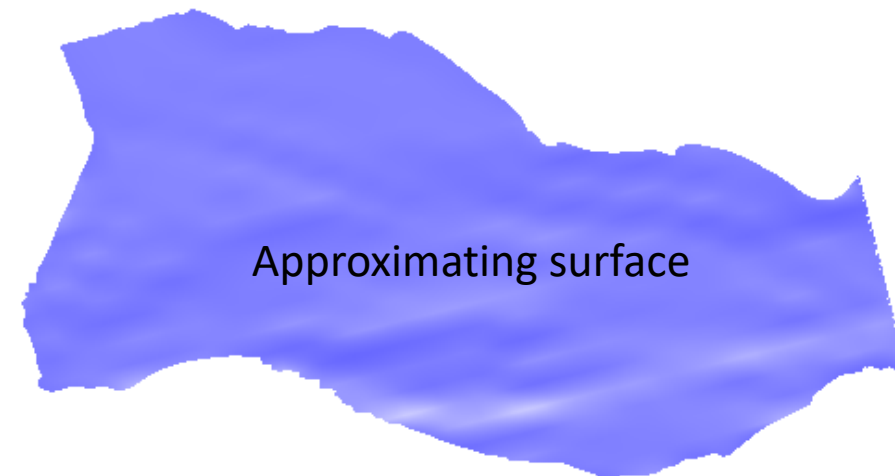
Green points at least 0.5m below surface
White points within 0.5 m of surface
Red points at least 0.5 m above surface



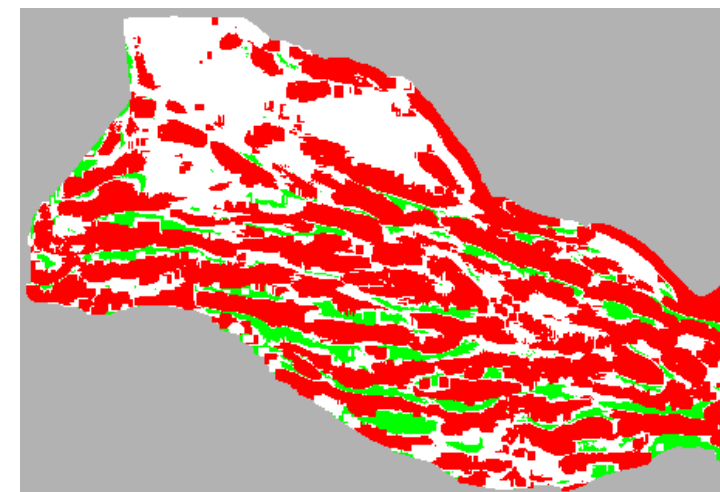
Number of points	14.6 mill
No. of coefs.	507
Surface file size	46 KB
Max. dist	10.5 m.
Average dist	0.83 m.
# points, dist > 0.5 m	7.3 mill

Data courtesy HR
Wallingford, SeaZone

Elevation interval: ~50 m.



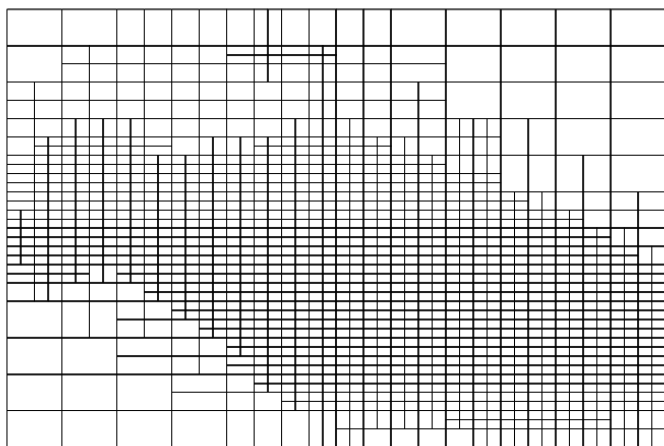
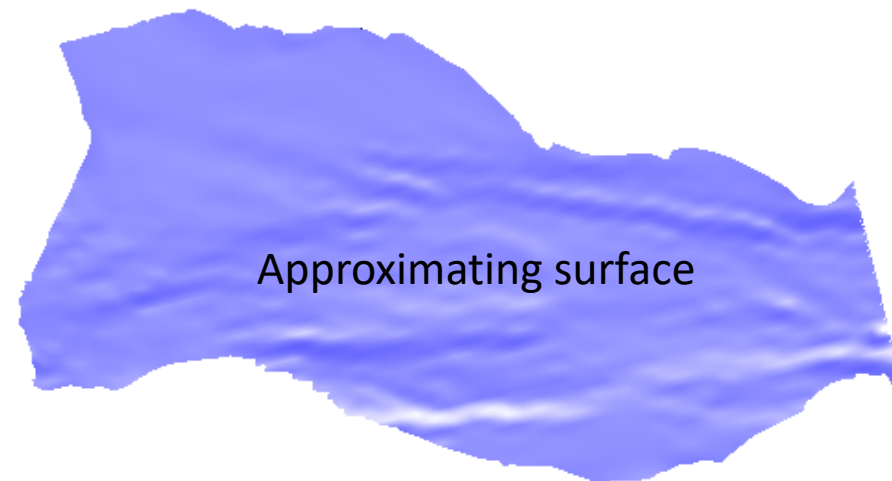
Green points at least 0.5m below surface
White points within 0.5 m of surface
Red points at least 0.5 m above surface



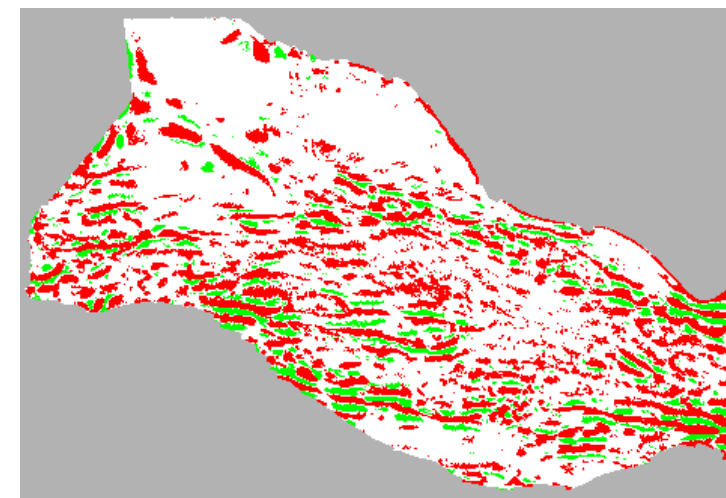
Number of points	14.6 mill
No. of coefs.	1336
Surface file size	99 KB
Max. dist	8.13 m.
Average dist	0.41 m.
# points, dist > 0.5 m	3.9 mill

Data courtesy HR
Wallingford, SeaZone

Elevation interval: ~50 m.



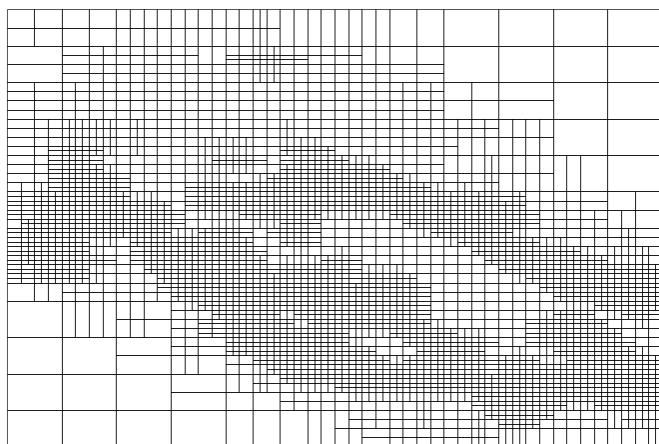
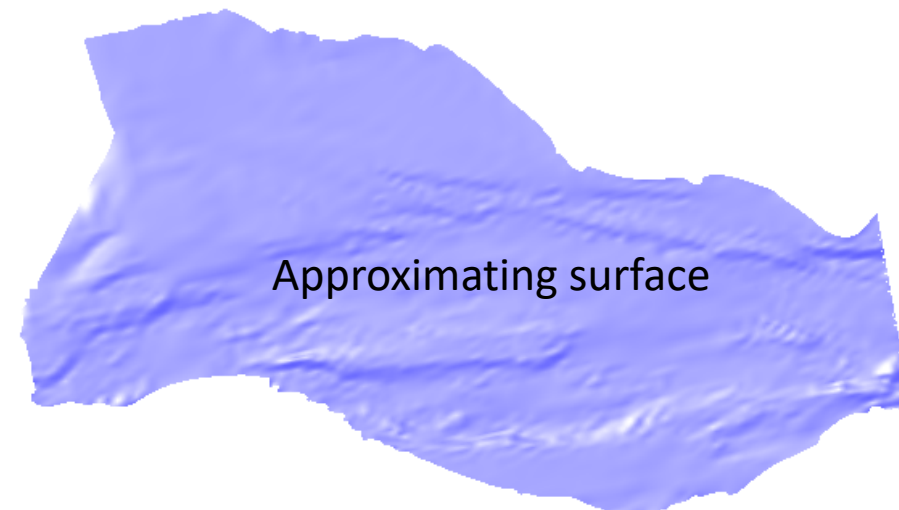
Green points at least 0.5m below surface
White points within 0.5 m of surface
Red points at least 0.5 m above surface



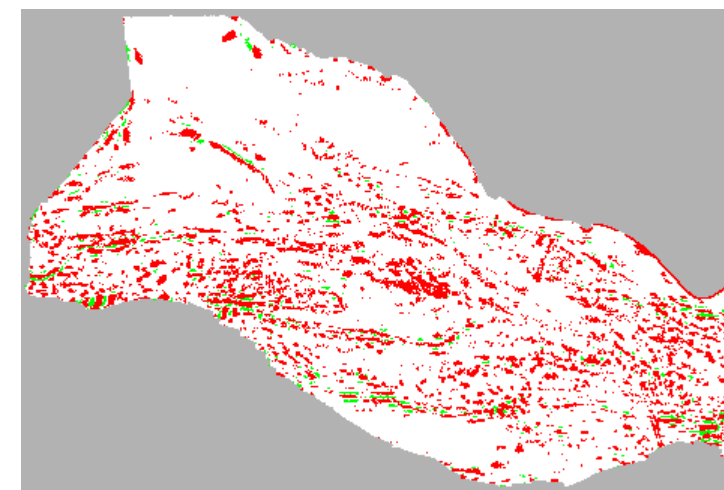
Number of points	14.6 mill
No. of coefs.	3563
Surface file size	241 KB
Max. dist	6.1 m.
Average dist	0.22 m.
# points, dist > 0.5 m	1.4 mill

Data courtesy HR
Wallingford, SeaZone

Elevation interval: ~50 m.



Green points at least 0.5m below surface
White points within 0.5 m of surface
Red points at least 0.5 m above surface

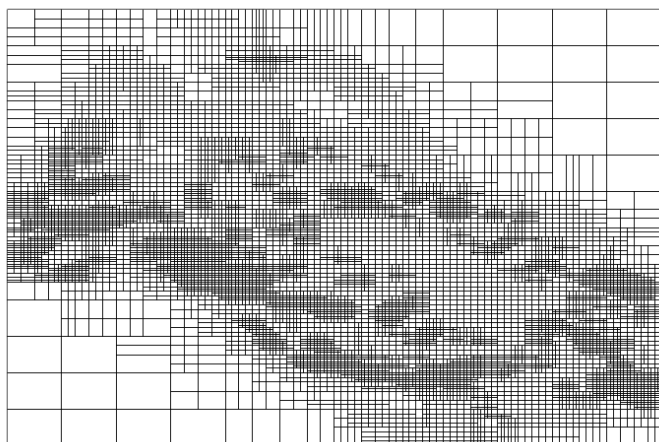
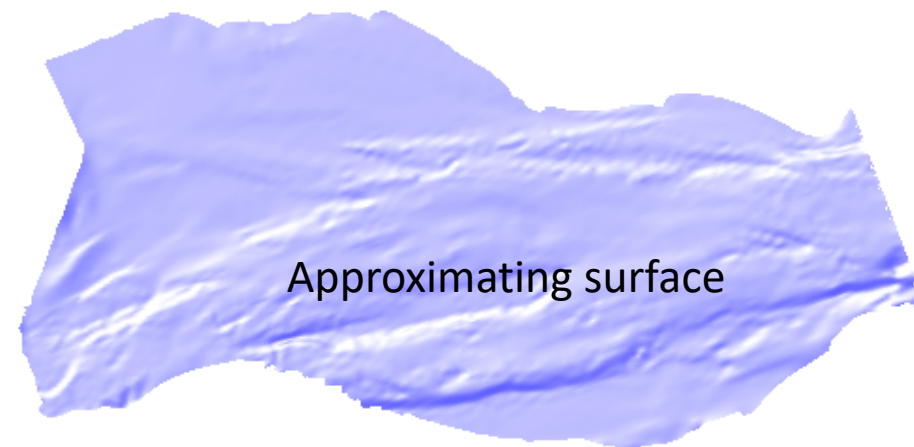


Fourth iteration

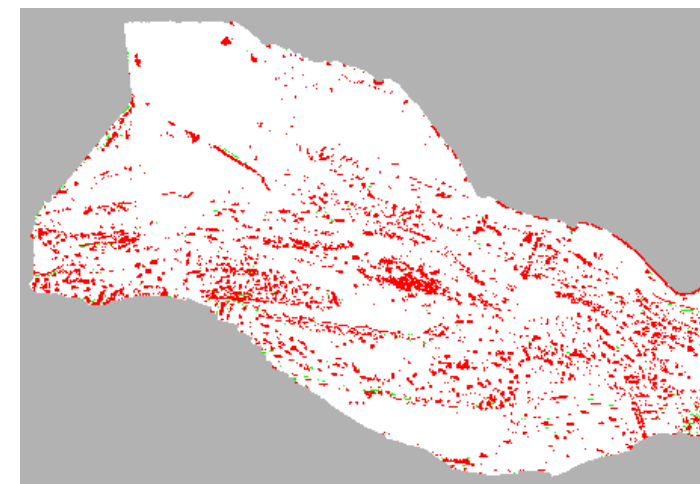
Number of points	14.6 mill
No. of coefs.	9273
Surface file size	630 KB
Max. dist	6.0 m.
Average dist	0.17 m.
# points, dist > 0.5 m	0.68 mill

Data courtesy HR
Wallingford, SeaZone

Elevation interval: ~50 m.



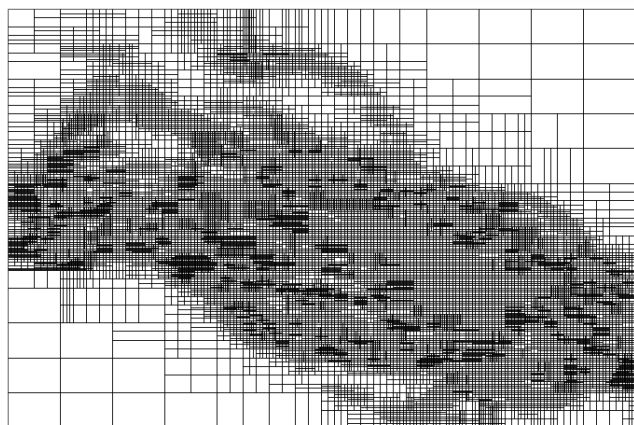
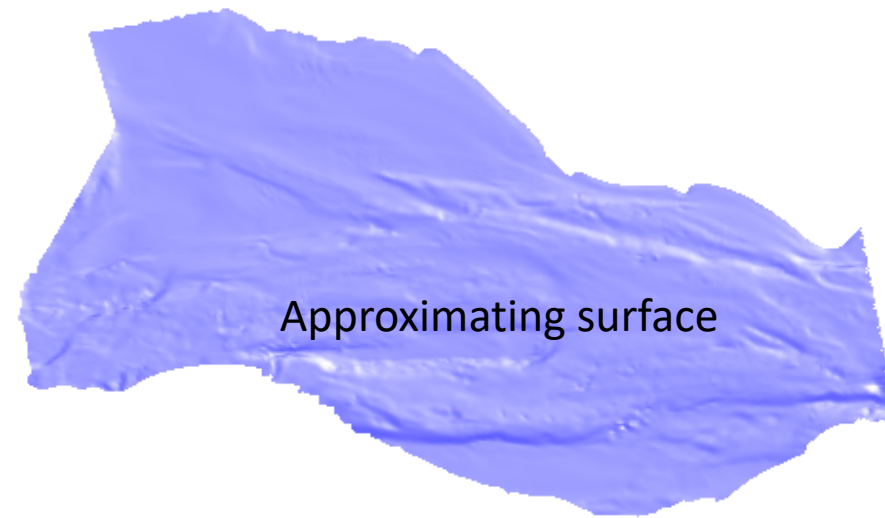
Green points at least 0.5m below surface
White points within 0.5 m of surface
Red points at least 0.5 m above surface



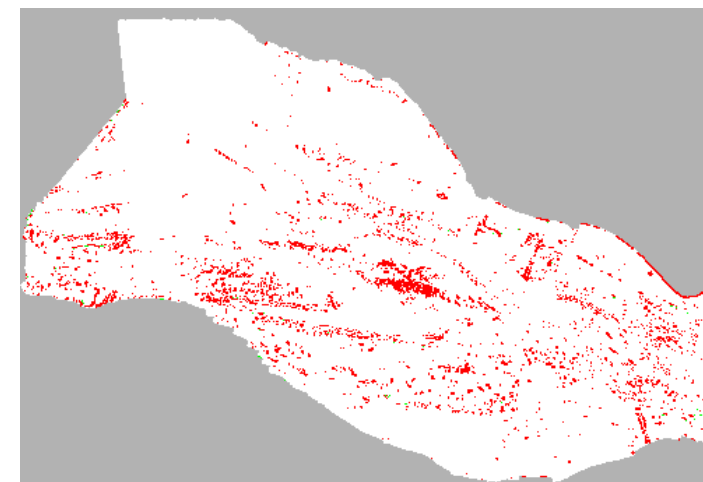
Number of points	14.6 mill
No. of coefs.	23002
Surface file size	1.6 MB
Max. dist	5.3 m.
Average dist	0.12 m.
# points, dist > 0.5 m	244 850

Data courtesy HR
Wallingford, SeaZone

Elevation interval: ~50 m.



Green points at least 0.5m below surface
White points within 0.5 m of surface
Red points at least 0.5 m above surface

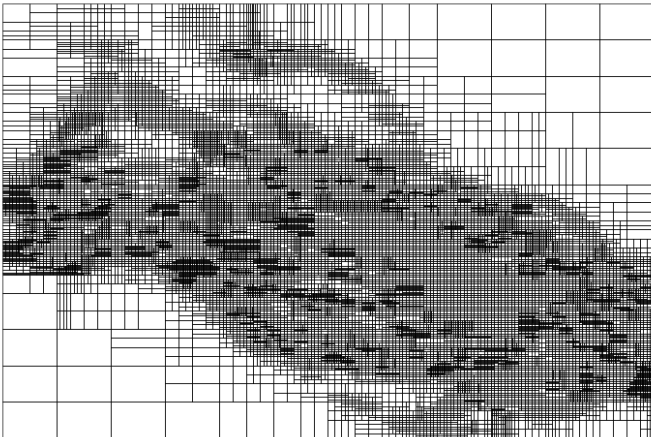
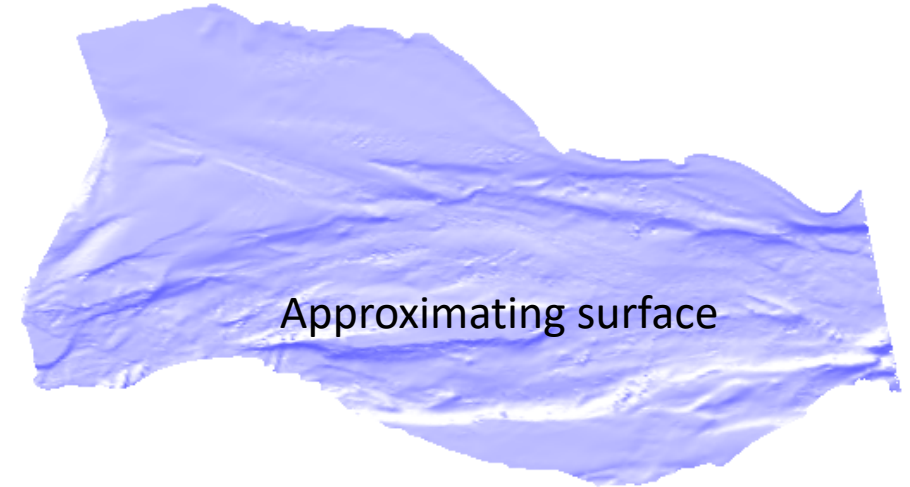


Sixth iteration

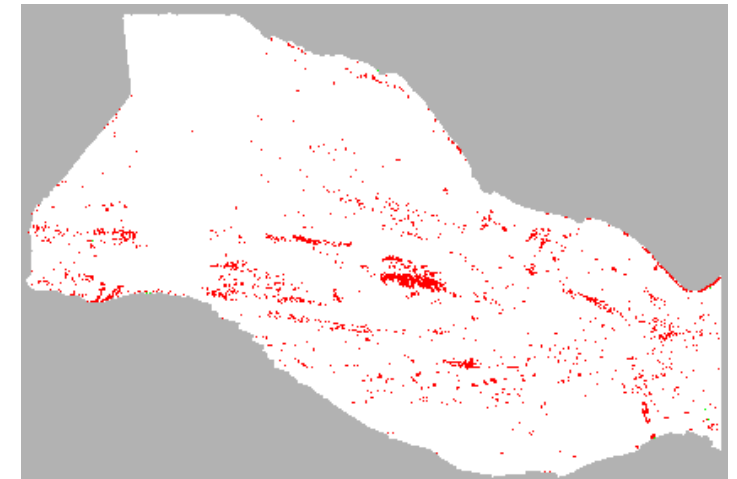
Number of points	14.6 mill
No. of coefs.	52595
Surface file size	3.7 MB
Max. dist	5.4 m.
Average dist	0.09 m.
# points, dist > 0.5 m	75 832

Data courtesy HR
Wallingford, SeaZone

Elevation interval: ~50 m.

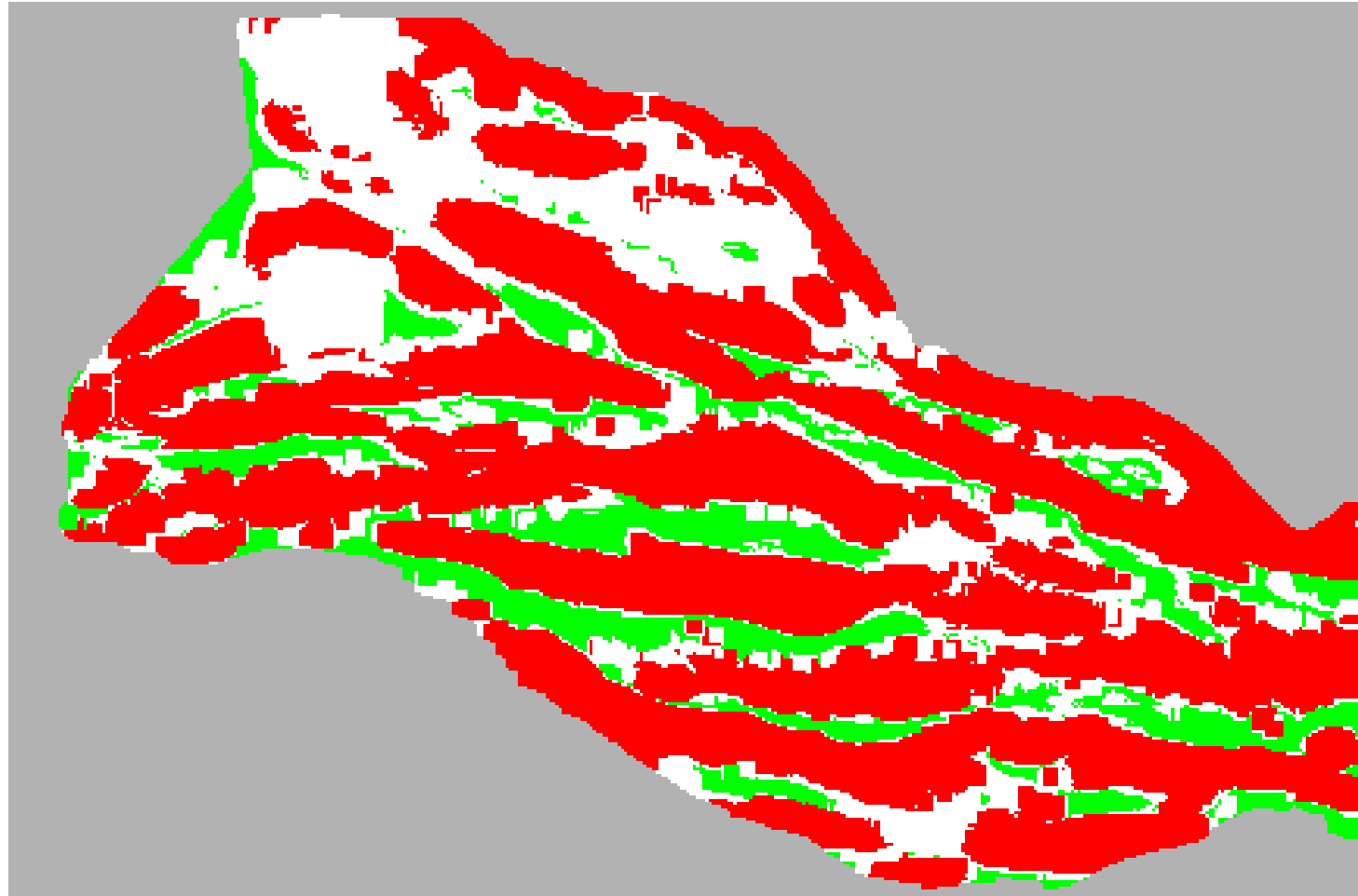


Green points at least 0.5m below surface
White points within 0.5 m of surface
Red points at least 0.5 m above surface



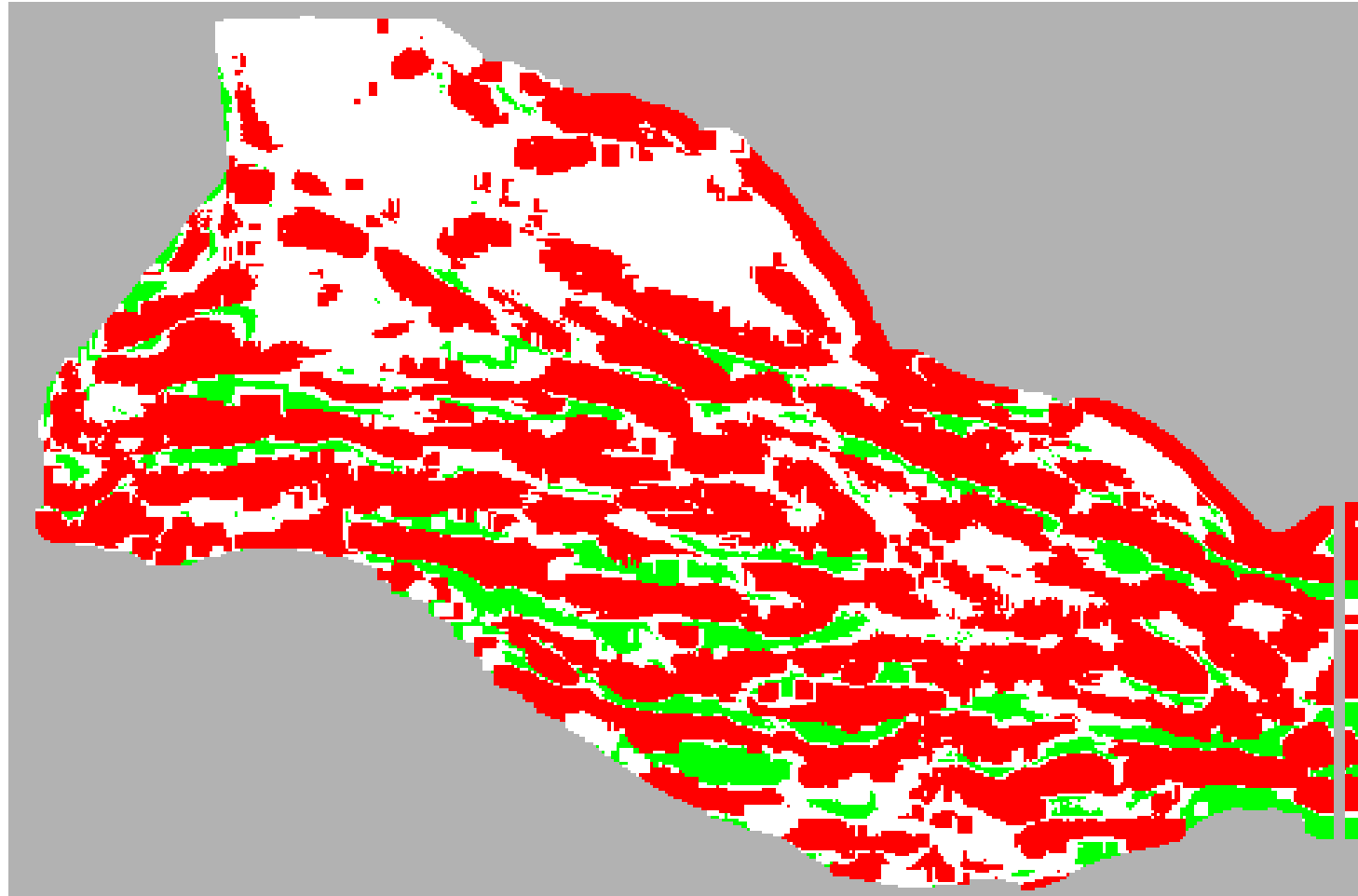
Running through the iterations

Data courtesy HR
Wallingford, SeaZone



196 coefficients

Running through the iterations

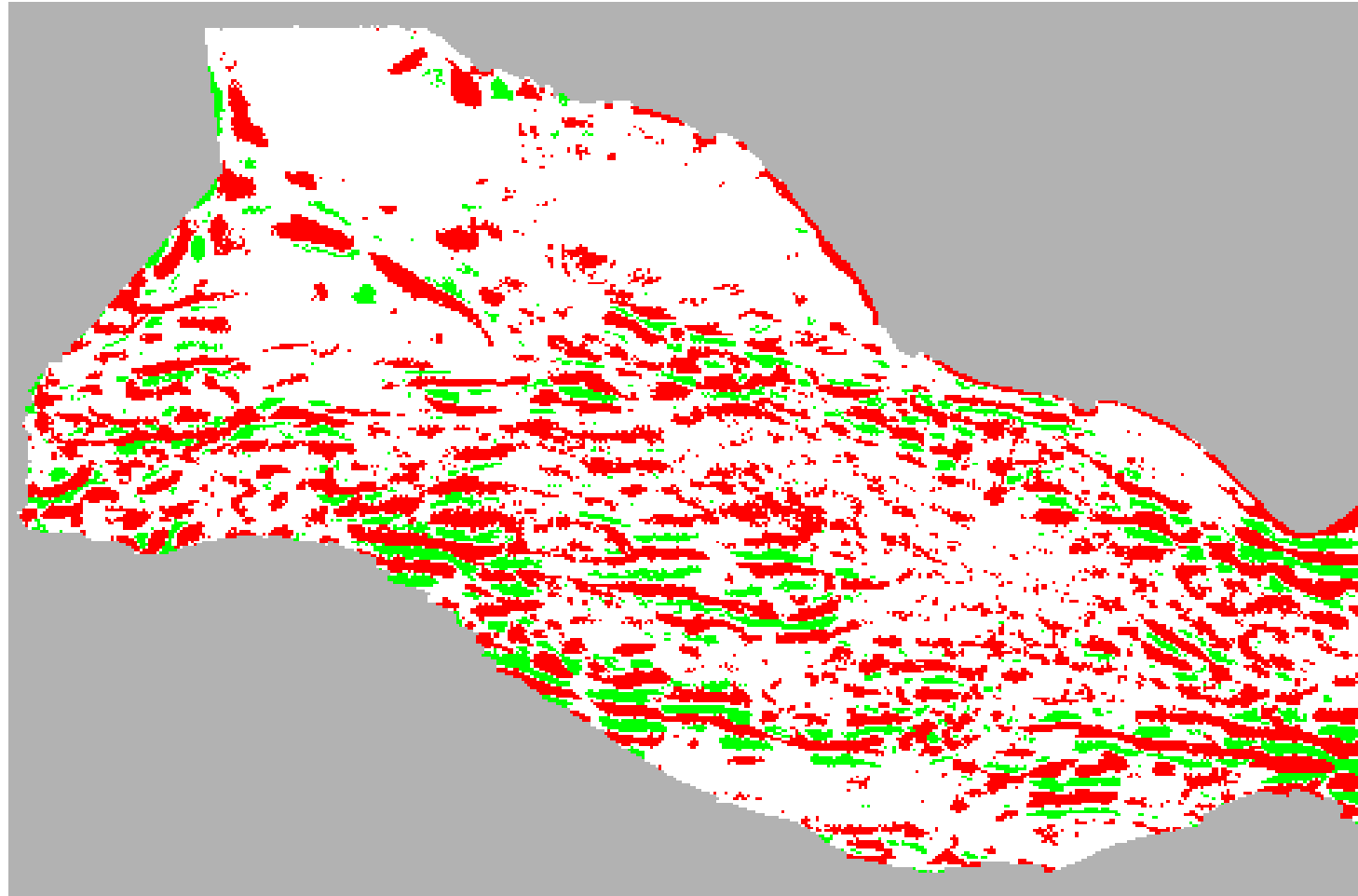


196 coefficients

507 coefficients

Data courtesy HR
Wallingford, SeaZone

Running through the iterations



Data courtesy HR
Wallingford, SeaZone

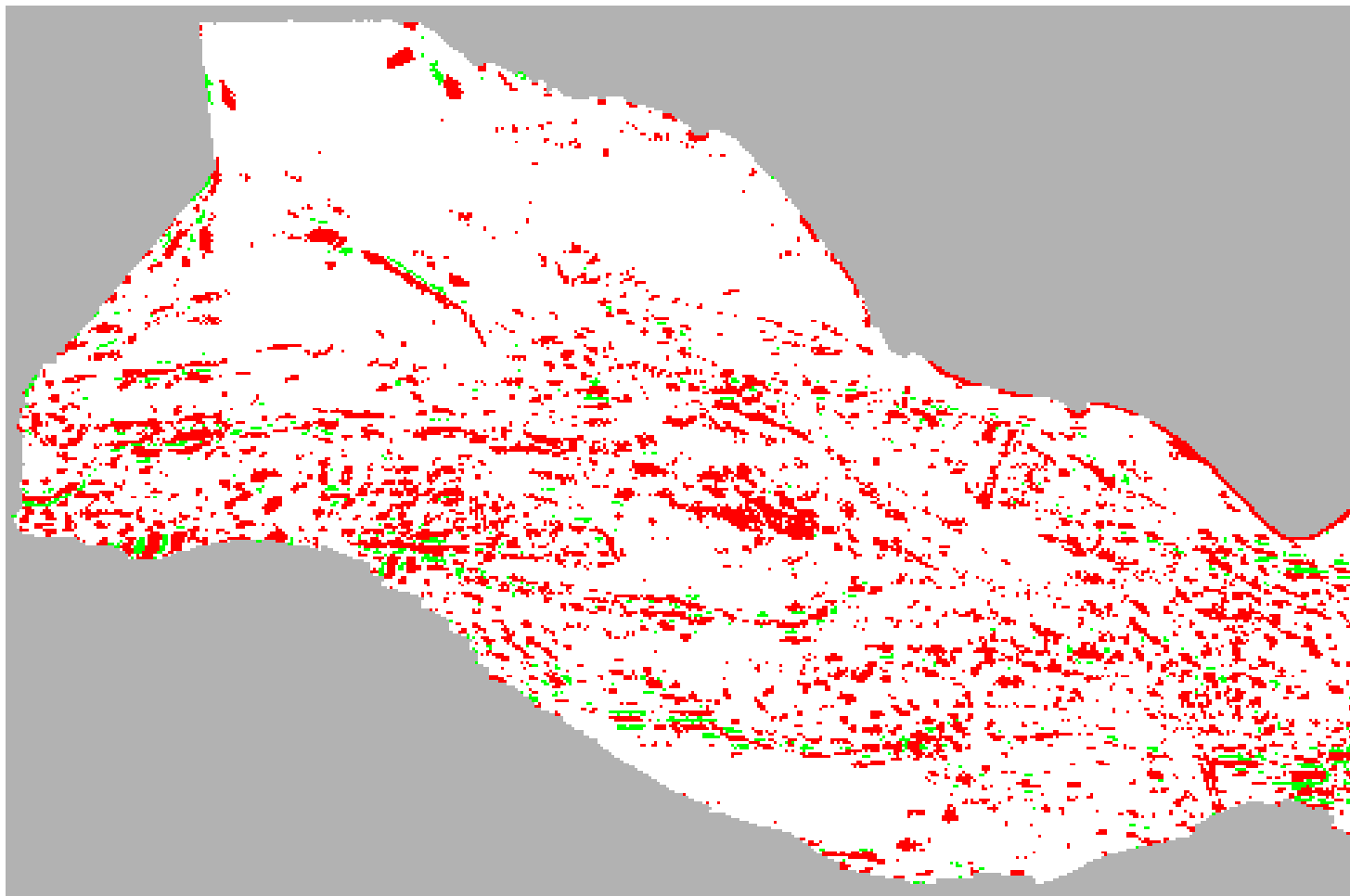
196 coefficients

507 coefficients

1 336 coefficients

Running through the iterations

Data courtesy HR
Wallingford, SeaZone



196 coefficients

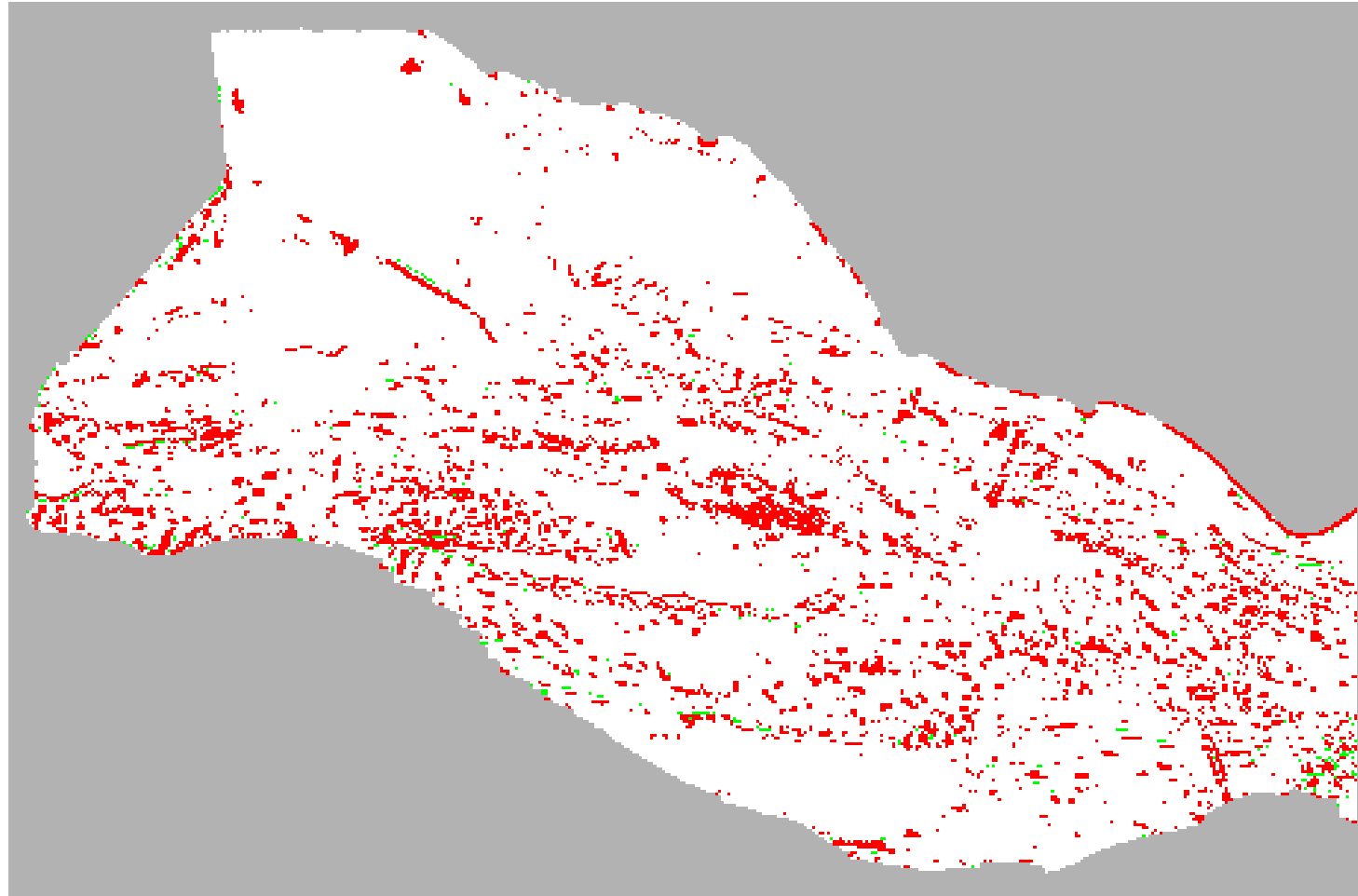
507 coefficients

1 336 coefficients

3 563 coefficients

Running through the iterations

Data courtesy HR
Wallingford, SeaZone



196 coefficients

507 coefficients

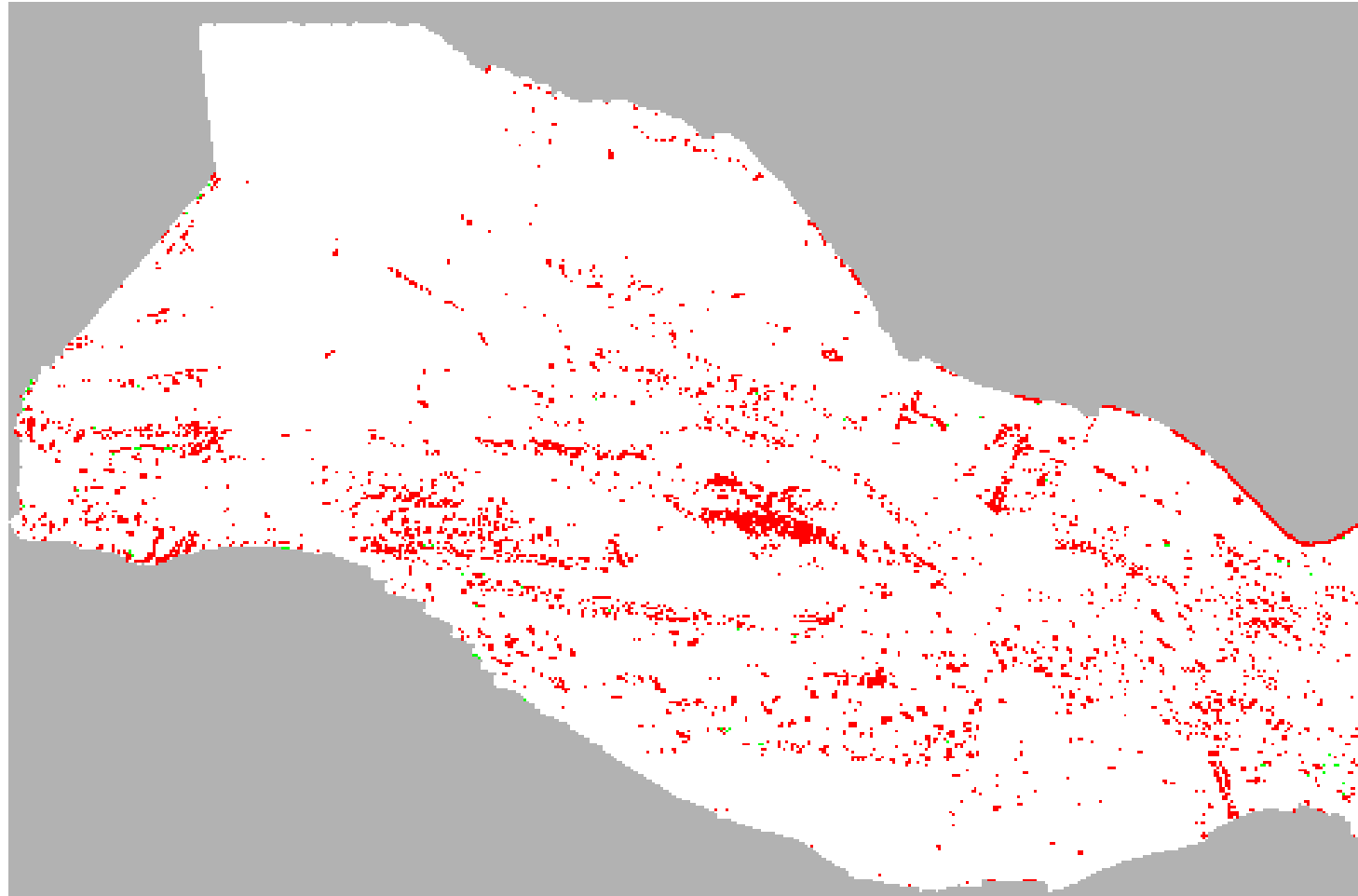
1 336 coefficients

3 563 coefficients

9 273 coefficients

Running through the iterations

Data courtesy HR
Wallingford, SeaZone



196 coefficients

507 coefficients

1 336 coefficients

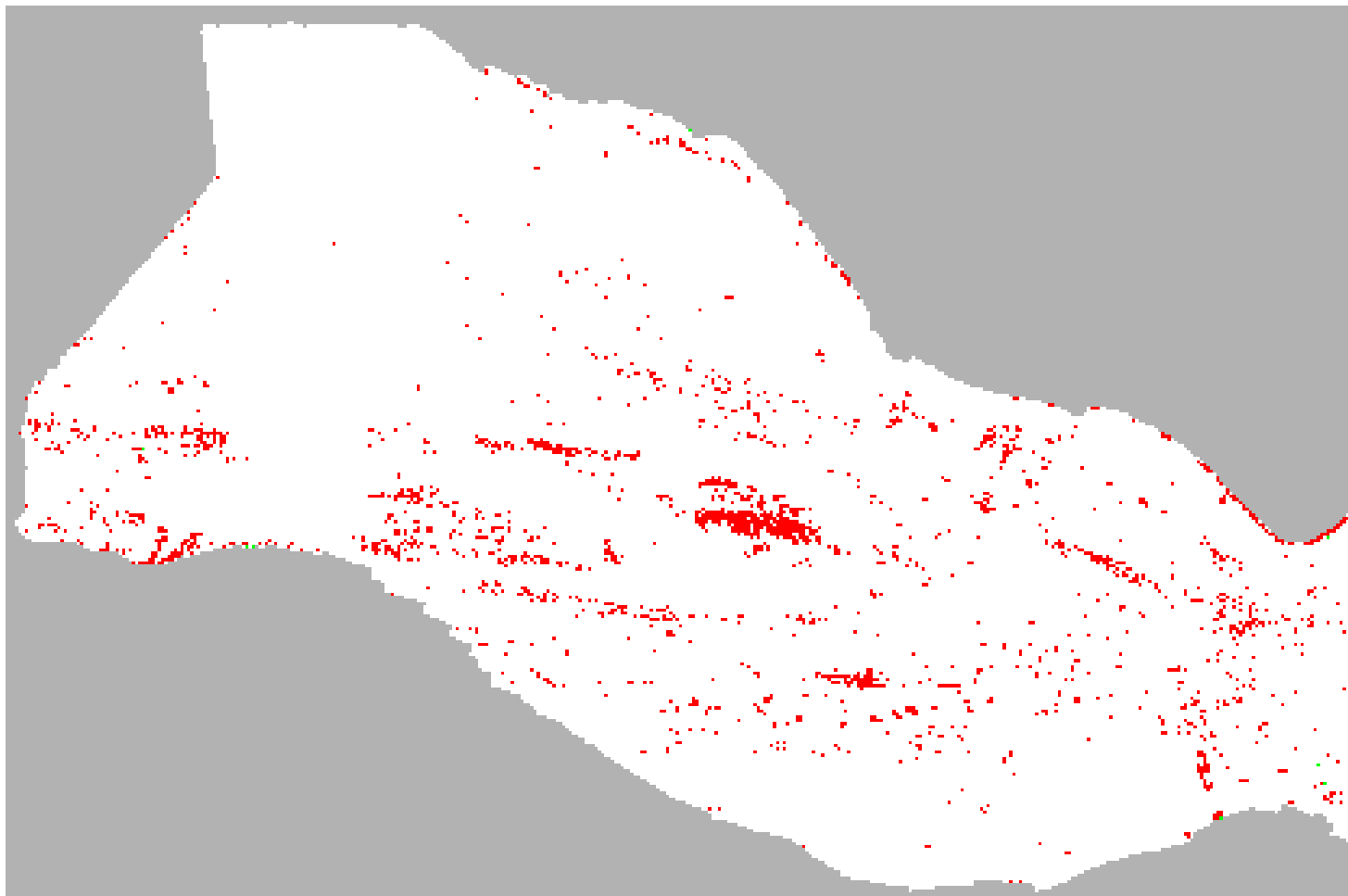
3 563 coefficients

9 273 coefficients

23 002 coefficients

Running through the iterations

Data courtesy HR
Wallingford, SeaZone



196 coefficients

507 coefficients

1 336 coefficients

3 563 coefficients

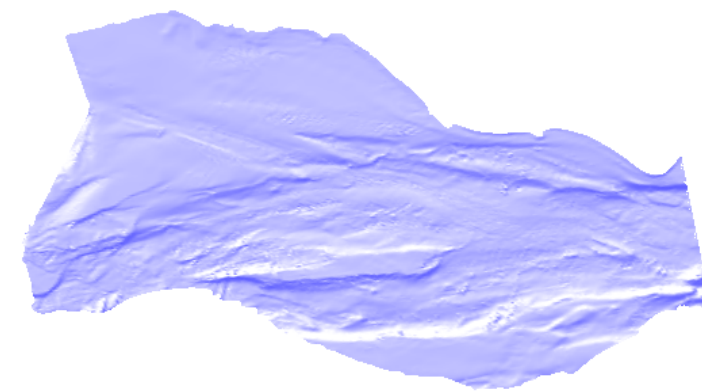
9 273 coefficients

23 002 coefficients

52 595 coefficients

Surface size versus point cloud size

- Original data set contains approx. 58 million points
- We perform successive thinning of the point cloud and approximate with fixed parameters:
- 0.5 m threshold, 6 iteration levels
- Results are very stable showing that the resulting LR B-spline grid is more dependent on the features of the terrain than the number of points in a scan.



Surface size $\approx$ 3.7 MB

No. points	File size	No. coefs.	Max. error	Average error	Average outside	Prop. OOT points
58 578 420	1.1 GB	53 454	5.55	0.092	0.66	0.56%
29 289 210	559 MB	52 709	5.39	0.092	0.66	0.55 %
14 644 406	280 MB	52 595	5.39	0.093	0.65	0.52 %
7 322 302	140 MB	52 611	5.33	0.093	0.65	0.47 %
3 661 151	70 MB	53 628	5.25	0.093	0.65	0.41%
1 830 575	35 MB	51 124	3.24	0.094	0.65	0.40 %

4th iteration

Uses of the spline approximation in the Norwegian ANALYST-project (2018-2021)

- LR B-spline approximation represents the smooth component of the data
 - Store the point outside the tolerance to keep the information on the none smooth component
- Feature detection
 - The points outside the approximation tolerance represents possible feature or outliers
 - The smooth component can be improved by rerunning the approximation without the points outside tolerance
 - This improved approximation will make features stand out more clearly
 - The smooth detection can be used for finding ridges, valleys, summits and hollows using partial derivatives
- Change detection comparing new points sets with smooth representation
- Deconfliction Removing points that don't fit in when combining multiple surveys
- Adaptive tolerances according to use (e.g., increasing tolerances in deep water)

Deconfliction

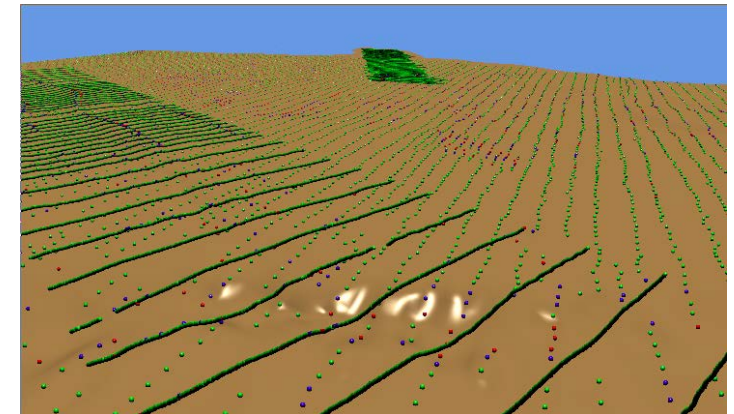
Removing points that don't fit in (outlier detection)

Given

- A number of overlapping data surveys
- A priority score for each survey
- Can not expect complete consistency between the various data surveys
- The combined point cloud will typically have a very heterogenous pattern

Want

- A consistent point cloud for surface generation
- Avoid large areas without points
- Avoid large jumps in height



Deconfliction algorithm

- Create reference surface: Approximate combined point cloud by a low resolution LR B-spline surface
- For each data survey
 - Distribute data points with respect to the polynomial patches of the LR B-spline surface
 - Compute the distances between the data points and the surface
- For each polynomial patch
 - Compute statistics for each survey: distance interval, average distance, standard deviation, number of points above and below surface, ...
 - Use the distance statistics and the priority score of the data surveys to decide whether to keep or remove a particular group of points
- If in doubt: Use decision on adjacent polynomial patches to make a decision
- Post processing: Update reference surface with respect to cleaned (deconflicted) point set for a more accurate surface representation

Final remarks

- IQmulus addressed big geospatial data targeting marine, urban and land usecases using workflows in the cloud
- The work on the marine usecases continues in the national Norwegian IKT Pluss Project ANALYST (2017-2021) combines the LR B-spline ideas with AI and deep learning
 - The Norwegian Mapping Authority Hydrographic Service is partner in ANALYST



Teknologi for et bedre samfunn