

Big changes coming to ECMWF Product Generation system

European Working Group on Operational meteorological Workstations (EGOWS):

15-17 October 2018

Marta Gutierrez

ECMWF Forecast Department

Marta.Gutierrez@ecmwf.int

Summary

- Overview of the Product Generation system
- The new Product Generation
- The new interpolation library : MIR
- The new requirements web interface
- Timelines

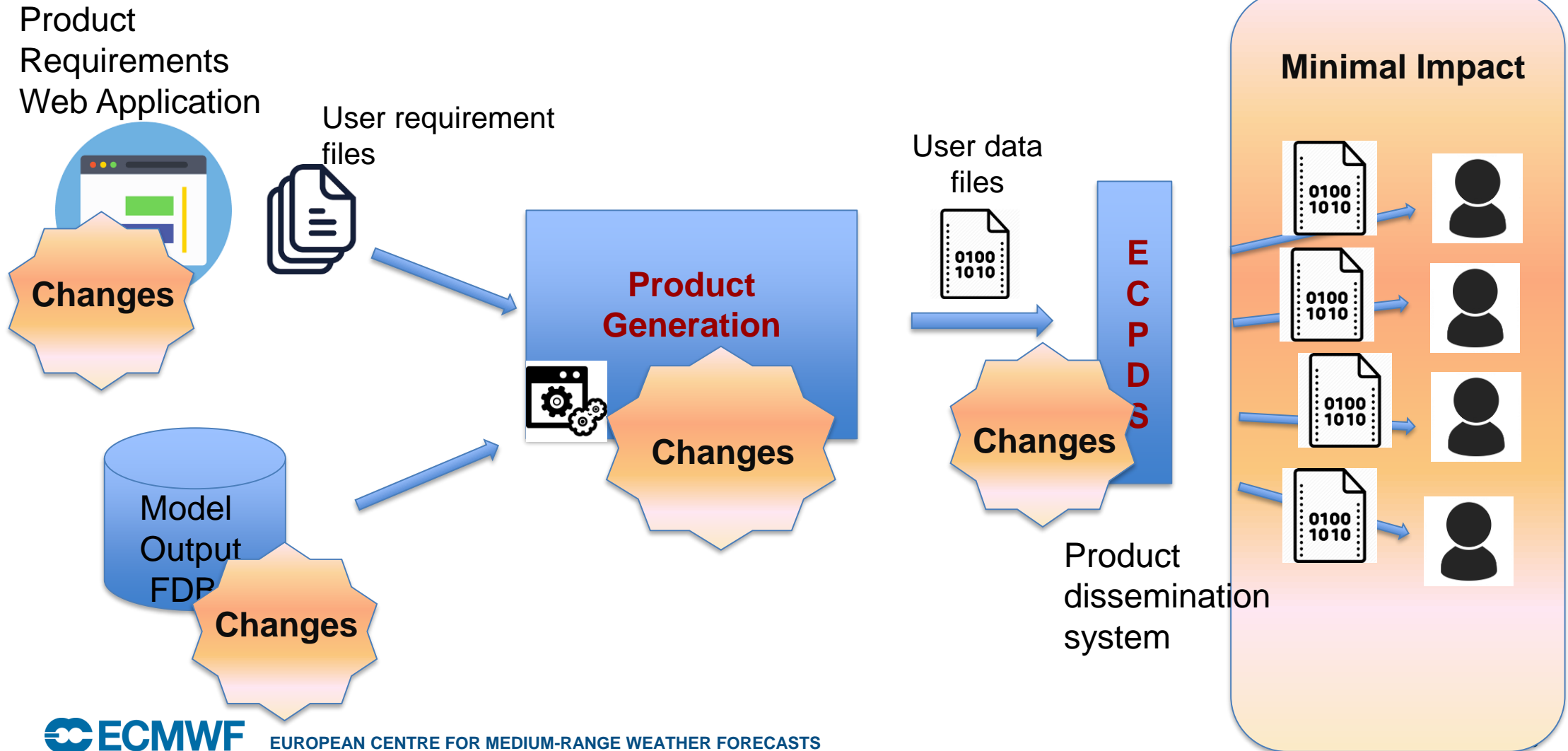
Products Generation team

- Umberto Modigliani, User Support
- Baudouin Raoult, Software Architect
- Dragan Jokic, Products Team
- Manuel Fuentes, Products Team
- Paul Dando, User Support
- Matthias Zink, Products Team
- Axel Bonet, Integration Team
- Marta Gutierrez, Data Services
- Pedro Maciel, Development
- Tiago Quintino, Development

Product Generation and Data delivery

- ECMWF model (IFS) produces raw model output for global fields in the spectral space (spherical harmonic fields) and physical space (reduced Gaussian grid, reduced lat/lon)
- Users interested in obtaining real time data for:
 - individual fields : e.g. 2T/TP
 - Other data representations: lat/lon, regular Gaussian grids
 - Subareas, points, frames
- Product Generation:
 - Model output + User requirements = User data files
- Data delivery system
 - How the data gets to the user

Overview Product Generation and data delivery system



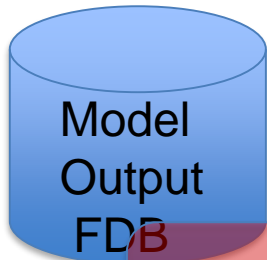
How big is it?

Product Requirements Web Application

User requirement files



230 million products

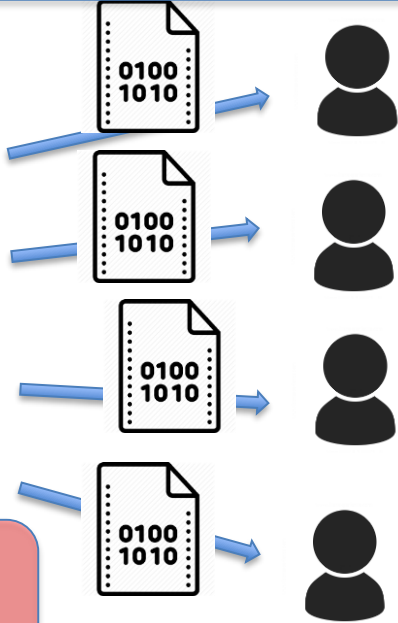


288 million fields
90 TB/day

29 TB/day, 166739 files
54 TB/day, 328012 files (Mon/Thu)



E
C
P
D
S



341 destinations

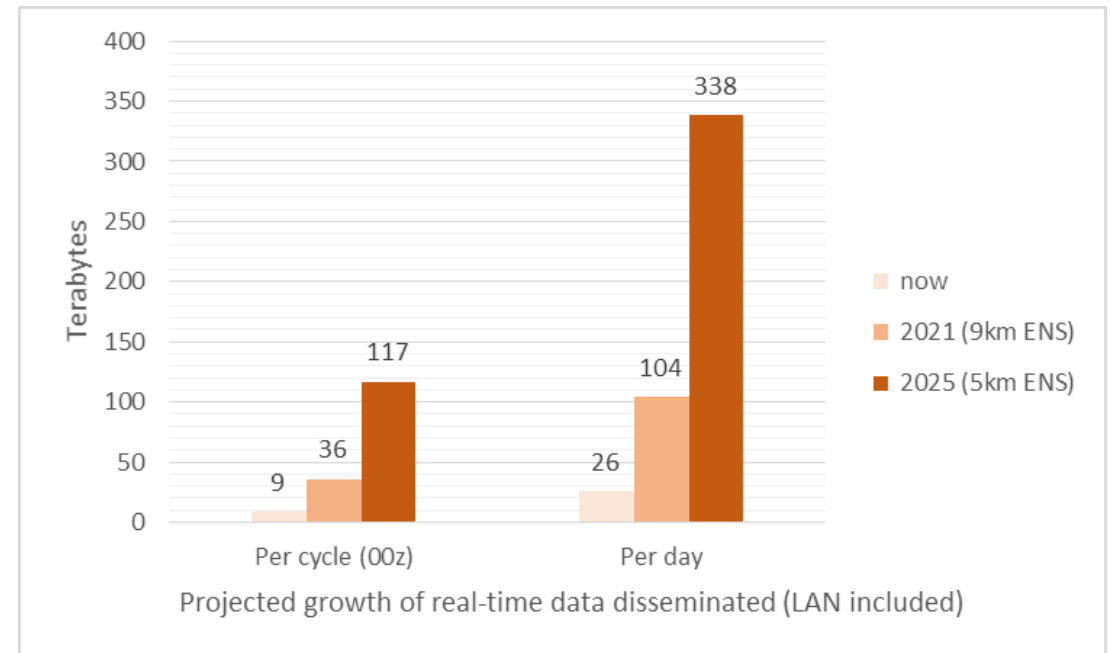
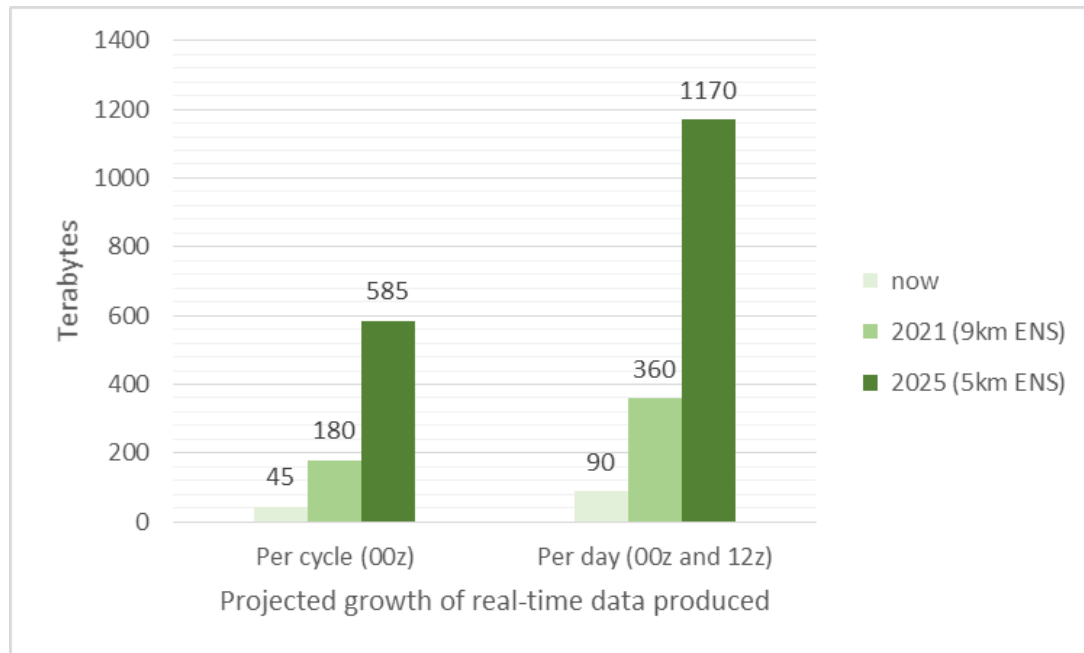
What is Product Generation?

- It's an HPC application , runs on the supercomputer
- It's in the time critical path: run IFS model, run Product Generation and start dissemination within 1 hour.
- It's a large scale interpolation exercise (230 million products)
- It's an intensive I/O application
- Computationally:~300 nodes, 10% of the cluster
- It's parallelised to sort out requests, interpolate , write output data
- Run product generation with all our base time forecasts:
 - 00Z / 06Z / 12Z / 18Z

Main drivers for new Product Generation system

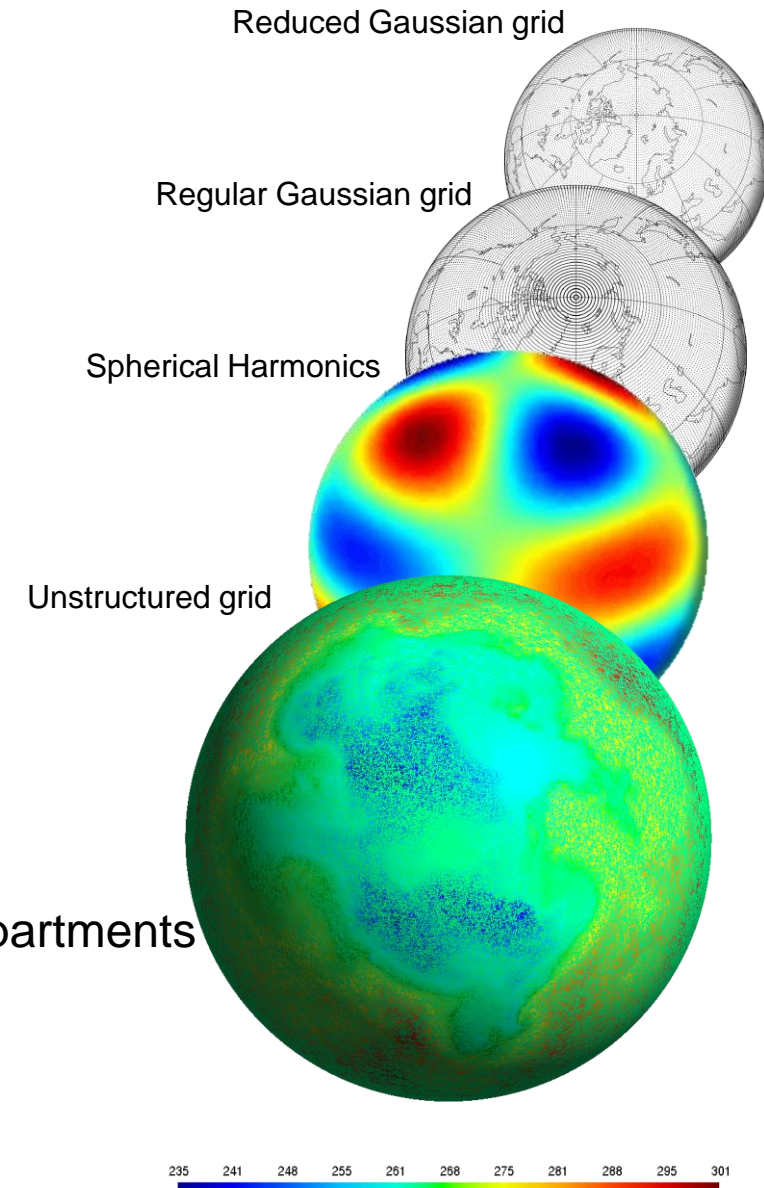
- To address exponential growth of data production
- Migrate from an old interpolation software (EMOSLIB) which is very costly to maintain (in manpower and bugs)

Real Time data volumes growth projections



MIR: The new interpolation library

- Meteorological Interpolation and Regridding
 - Application-driven
 - MARS, Product Generation, Metview (...)
- Architecture
 - Multiple grid formats support (*any-to-any*)
 - Configurable operations and defaults
 - Linear algebra (CPU/GPU...)
 - Cacheable (memory-mapped, shared memory, file I/O)
- Collaboration between ECMWF Research and Forecast Departments



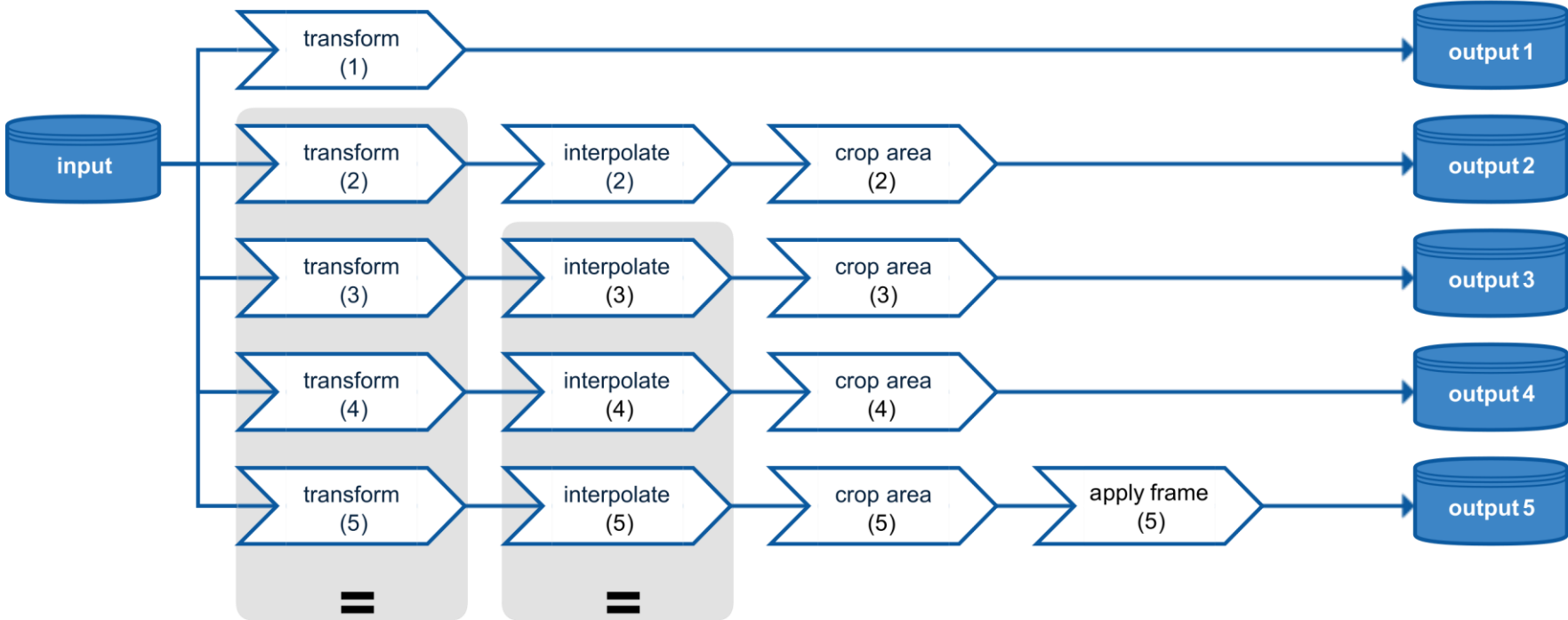
MIR action plan



Spectral to spectral
Spectral to grid

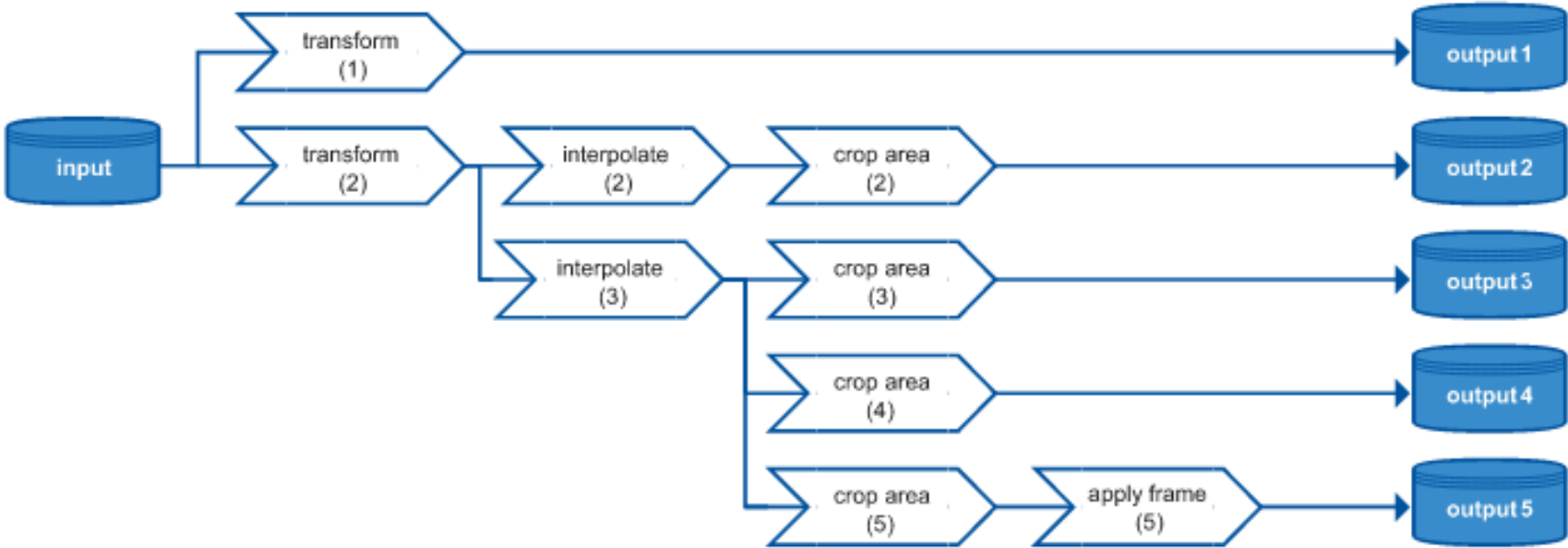
Reduced Gaussian
to Lat/lon

MIR: Product Generation action tree



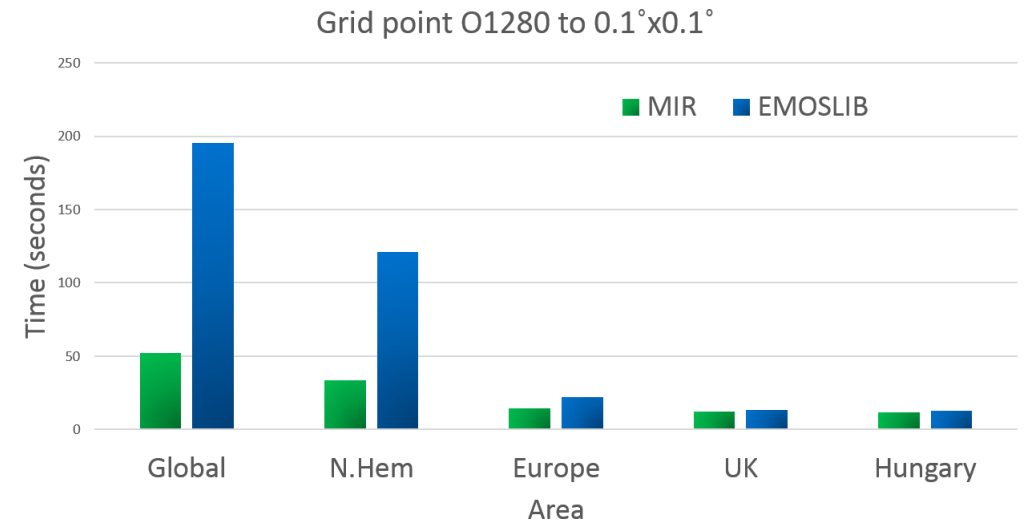
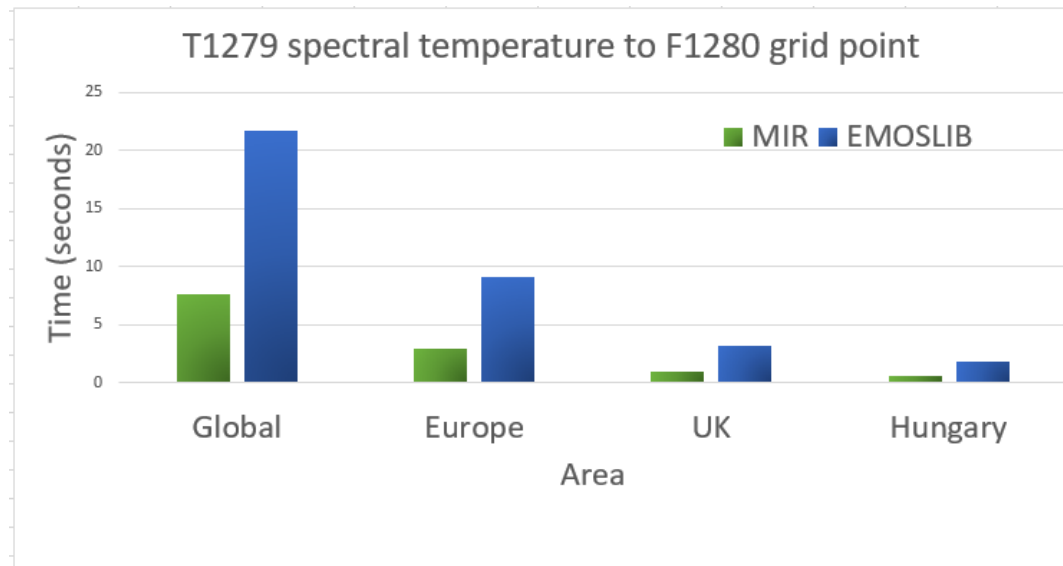
Action comparison
(after first action runs, the others share the results)

MIR: Product Generation simplified tree



Performance gains MIR

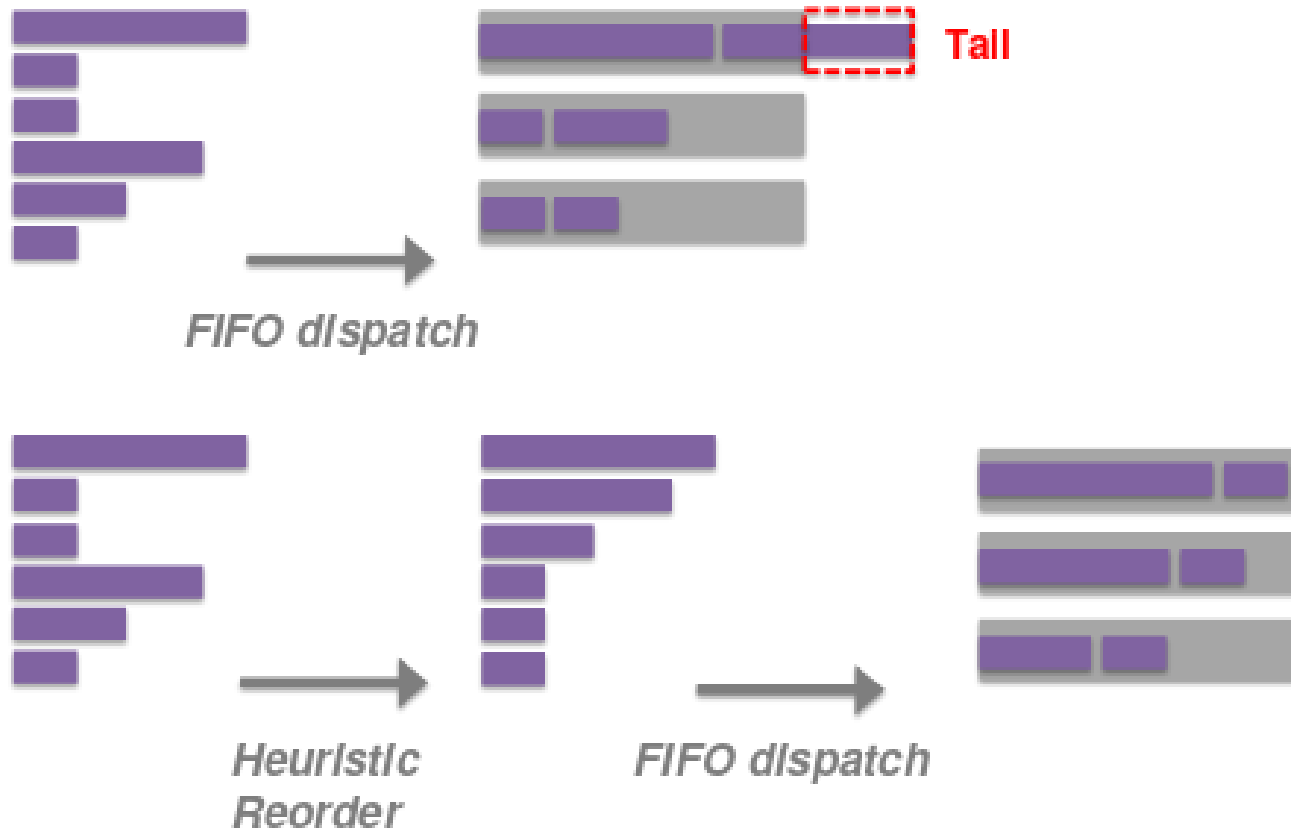
- MIR spectral transforms are faster (up to x3)
- MIR global to global interpolations are faster (x2 ~x3)



Performance gains Product Generation

- I/O processing has been optimised to write the output files.
- Load balancing processing tasks

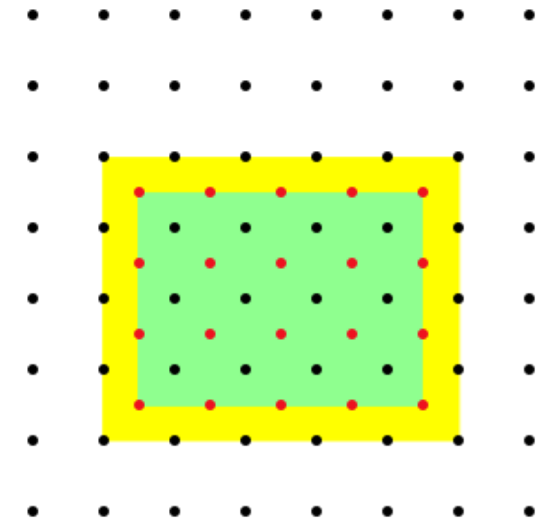
Example: 6 tasks, 3 workers



- Run times for new Product Generation ~ 3 - 5 times faster
- Reduce computational load from 300 to 100 nodes

User impact and known differences: MIR

- Values will be different (better)
- Sub-areas inwards (MARS = Dissemination)
- Sub-areas of reduced grids supported (MARS = Dissemination)
- Support for 'staggered' grids
- No land-sea mask processing by default
- No special treatment of precipitation
- Different distance computation (3D)
 - affects nearest neighbour method
 - No singularities at poles
- More user control over spectral-to-grid transformations

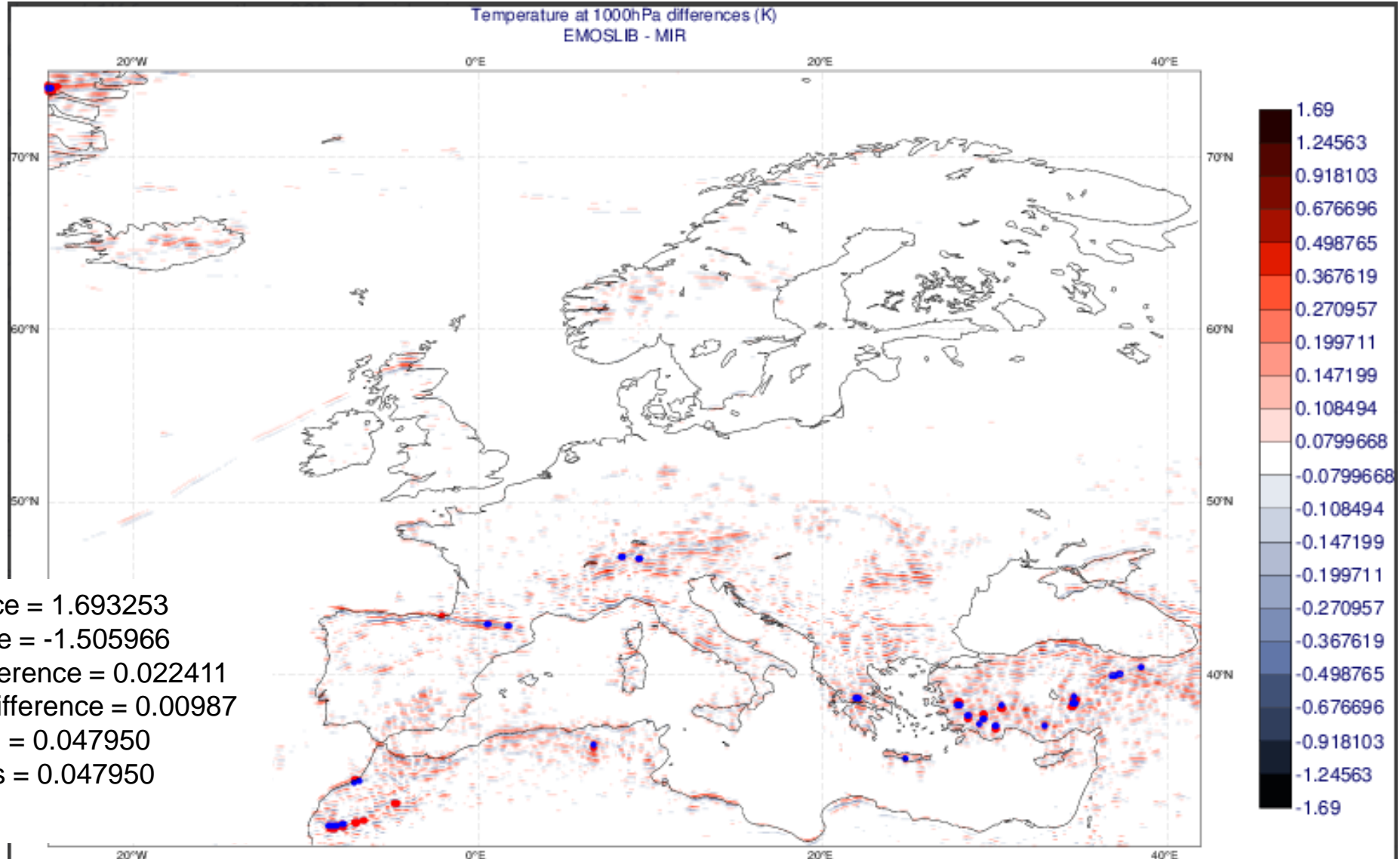


AREA=89.5/0.5/-89.5/359.5,
GRID=1.0/1.0

For full details see:

<https://confluence.ecmwf.int/display/UDOC/MARS+interpolation+with+MIR>

Differences in Temperature at 1000 hPa



Maximum difference = 1.693253
Minimum difference = -1.505966
Mean absolute difference = 0.022411
Median absolute difference = 0.00987
Standard deviation = 0.047950
RMS of differences = 0.047950

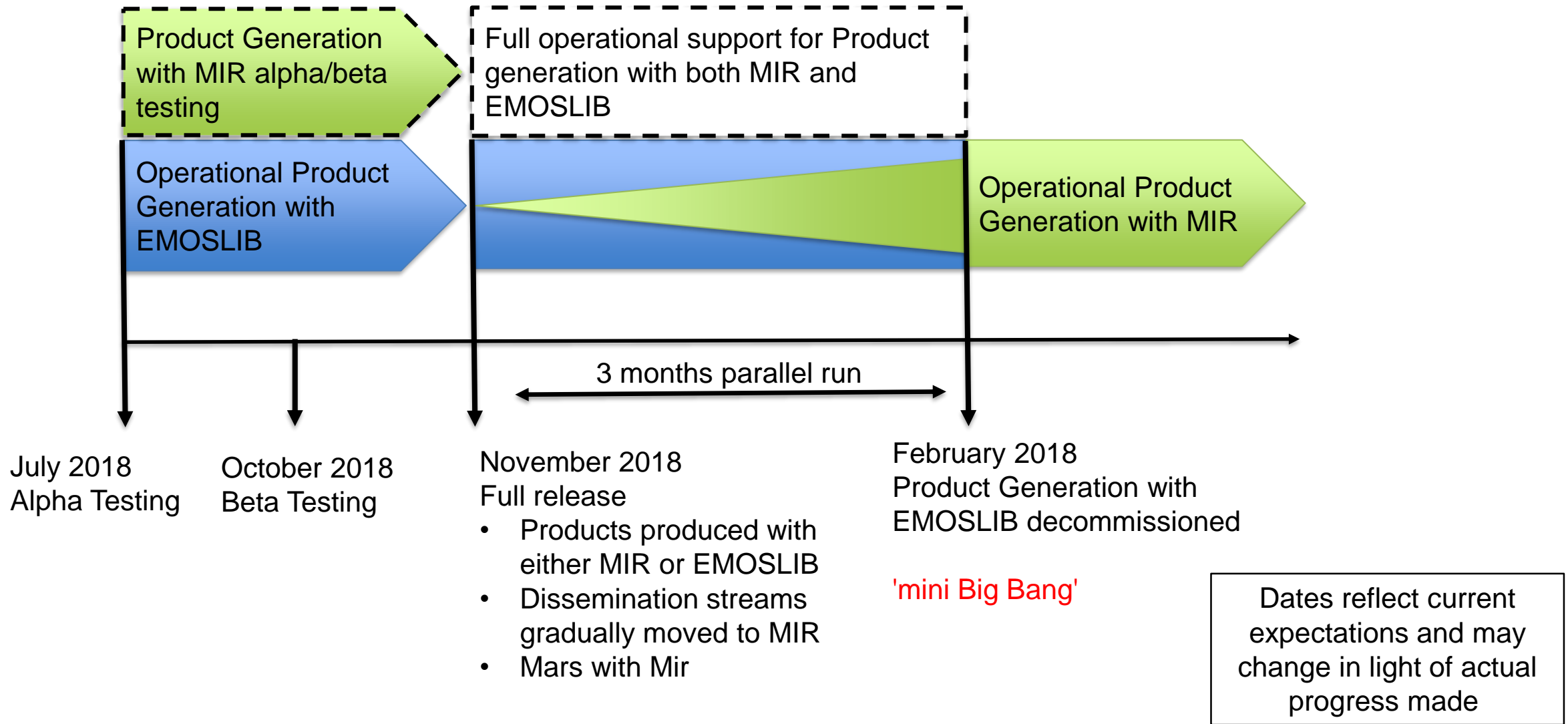
User impact and known differences: Product Generation

- Different interpolation library:
 - Different point count for specific cases of reduced Gaussian grid
- Different ways in which the areas are encoded in GRIB headers
 - Different precision, rounding , different longitude encoding in GRIB1
- Constant fields (will be encoded with bitsPerValue = 0)
- Local definitions retained in GRIB1 for HRES (00/12)

For full details see:

<https://confluence.ecmwf.int/display/UDOC/PGEN++PRODGEN+differences>

Implementation plan: Product Generation schedule



Changes to the Product Requirements Web interface

The screenshot shows the ECMWF Product Requirements Web interface. At the top, there is a navigation bar with the ECMWF logo and links for 'About', 'Forecasts', 'Computing', 'Research', and 'Learning'. Below this, the breadcrumb 'Member State Test > L1' is visible. A left-hand navigation menu lists various options including 'Home', 'Member State Test', 'LIT', and 'Publication requests (4)'. The main content area displays 'LIT:L1' with a 'Disseminated' status. A table summarizes product statistics:

Type	N. Products	Volume
OPER	10200	3.1 GB
SCDA	0	n/a
Total produced	10200	3.1 GB
Total disseminated	10200	3.1 GB

Below the table is a code editor for the 'LIT:L1' configuration. The code defines parameters for dissemination, including stream, domain, expver, type, levtype, param, time, step, format, padding, grid, area, priority, and levelist.

```

1  disseminate,
2  stream      = oper,
3  domain      = g,
4  expver      = 01,
5  type        = an,
6  levtype     = sfc,
7  param       = ms1/2t/2d/10u/10v/sstk/ci/tcc/lcc/mcc/hcc/lsm/sd,
8  time        = 0000/1200,
9  step        = 00,
10 format     = grib,
11 padding     = n,
12 grid        = 0.125/0.125,
13 area        = 73.5/-29.0/31.0/47.0,
14 priority    = 80,
15 levelist    = off
16
17
18  disseminate,
19  type        = an,
20  levtype     = sfc,
21  param       = z,
22  step        = 00
23
24
25  disseminate,
26  type        = fc,
  
```

For the future ...

- Much more data is on the way
- More fields (parameters, levels, etc)
- Improve current I/O bottlenecks
 - Rely on specialised storage systems (object-stores, SSD's, etc)
 - Stream model output from memory to Products Generation
- Possibility to get the products as soon as they are produced
 - Requires flexibility on client
 - Still possible to use fixed schedule.
- European Weather Cloud (EWC)
 - Alternative to pushing data to users
 - Bring user's workflows close to ECMWF data

Thanks!