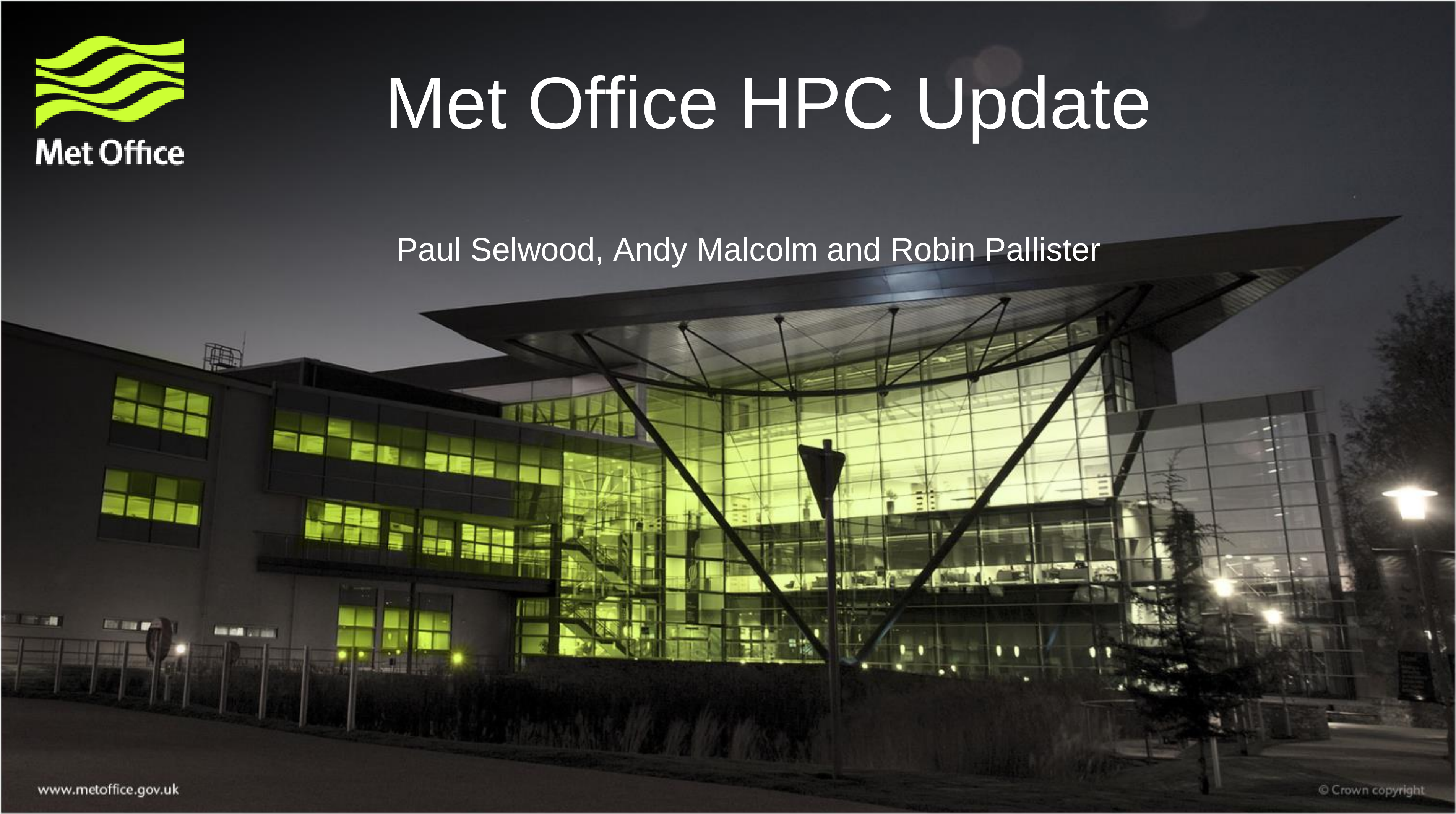


Met Office HPC Update

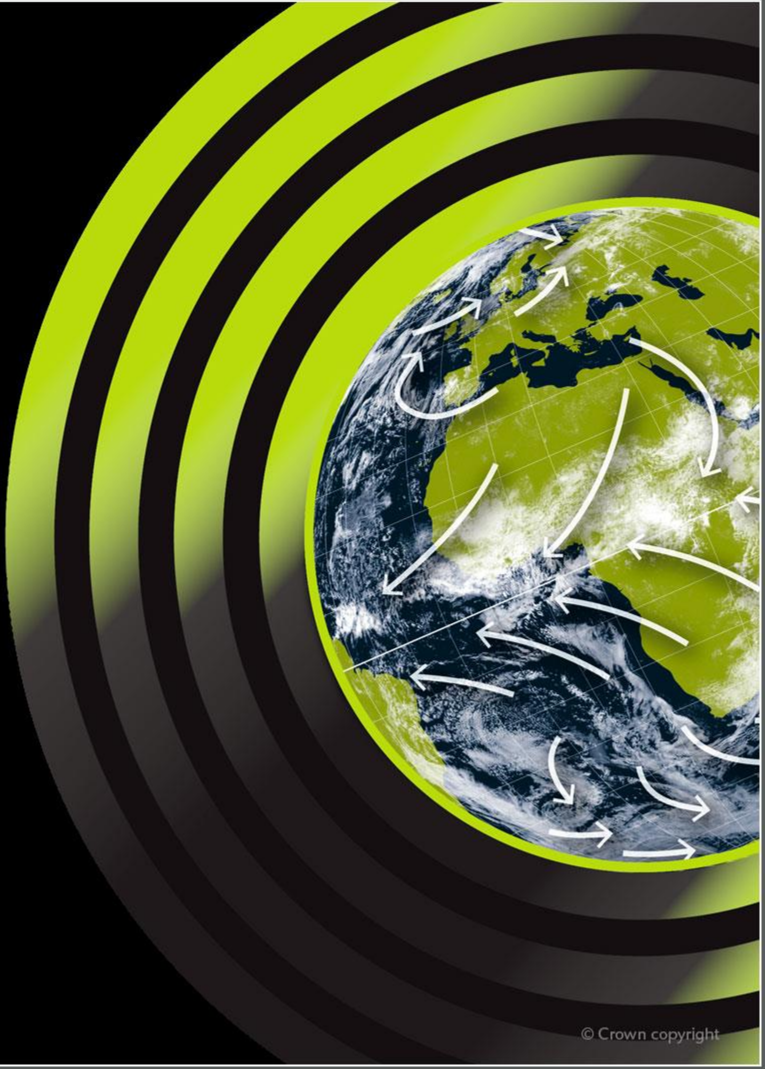
Paul Selwood, Andy Malcolm and Robin Pallister

Overview

HPC Systems

Model Plans

Scaling Results

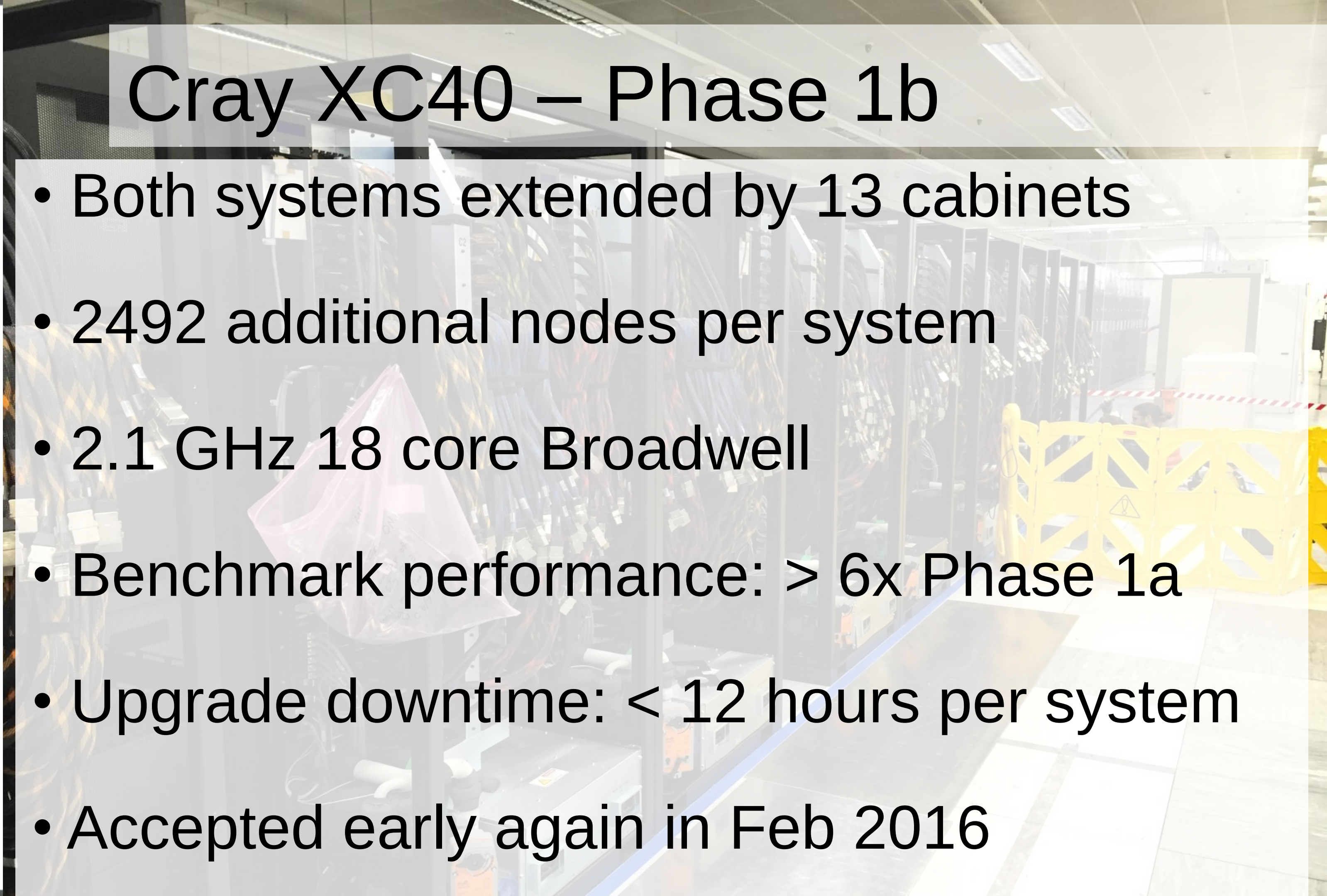


HPC Systems

Cray XC40 – Phase 1a

- Two systems of 4 cabinets – 560 nodes each
- Same capacity as previous IBM systems
  - Available power for churn a problem
- 2.3 GHz 16 core Haswell
- 128 GB/node
- 2x 3PB and 1x 6PB Lustre storage - Sonexion
- Live in Aug 15 – ahead of schedule

Cray XC40 – Phase 1b

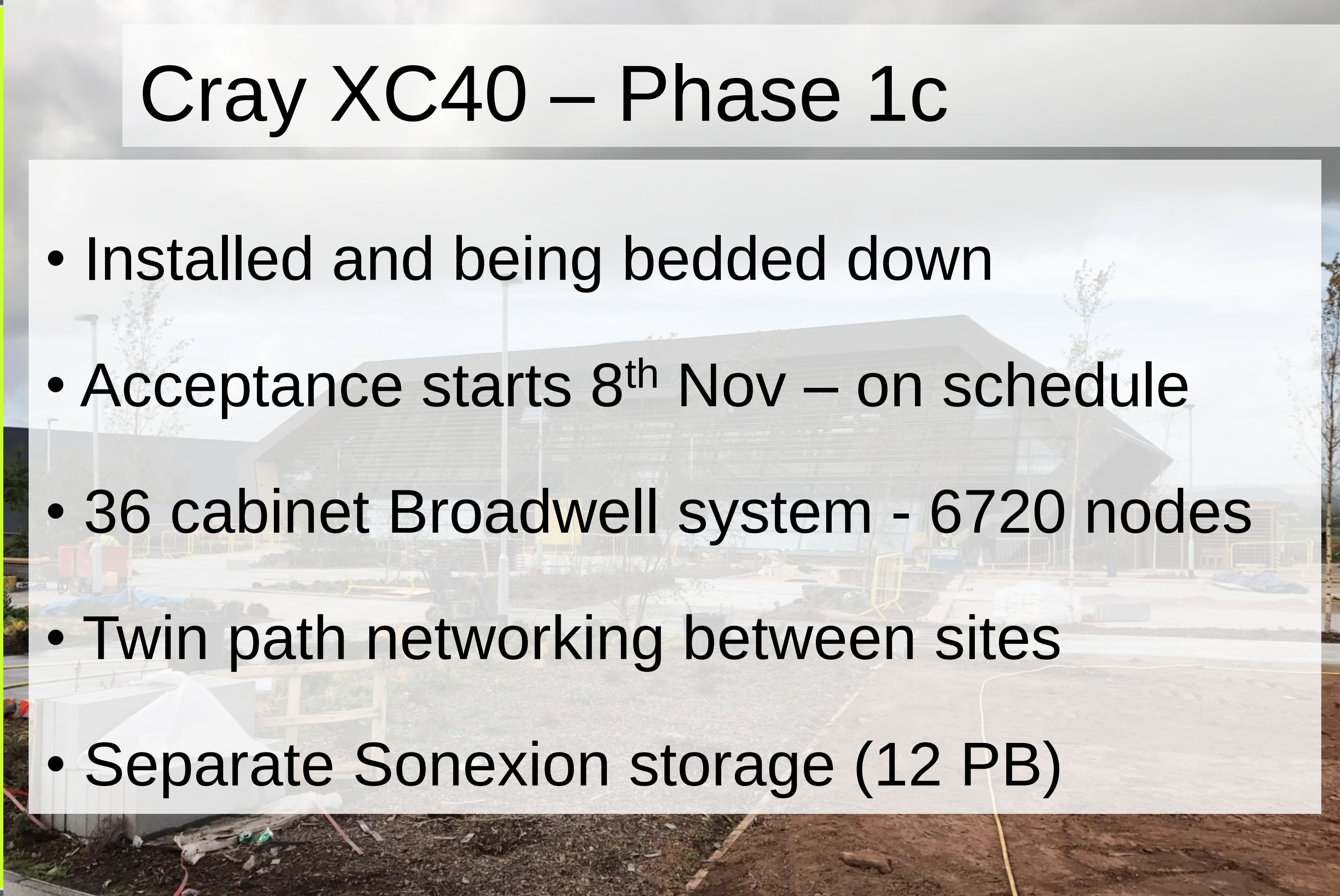


- Both systems extended by 13 cabinets
- 2492 additional nodes per system
- 2.1 GHz 18 core Broadwell
- Benchmark performance: > 6x Phase 1a
- Upgrade downtime: < 12 hours per system
- Accepted early again in Feb 2016

Cray XC40 – Phase 1c

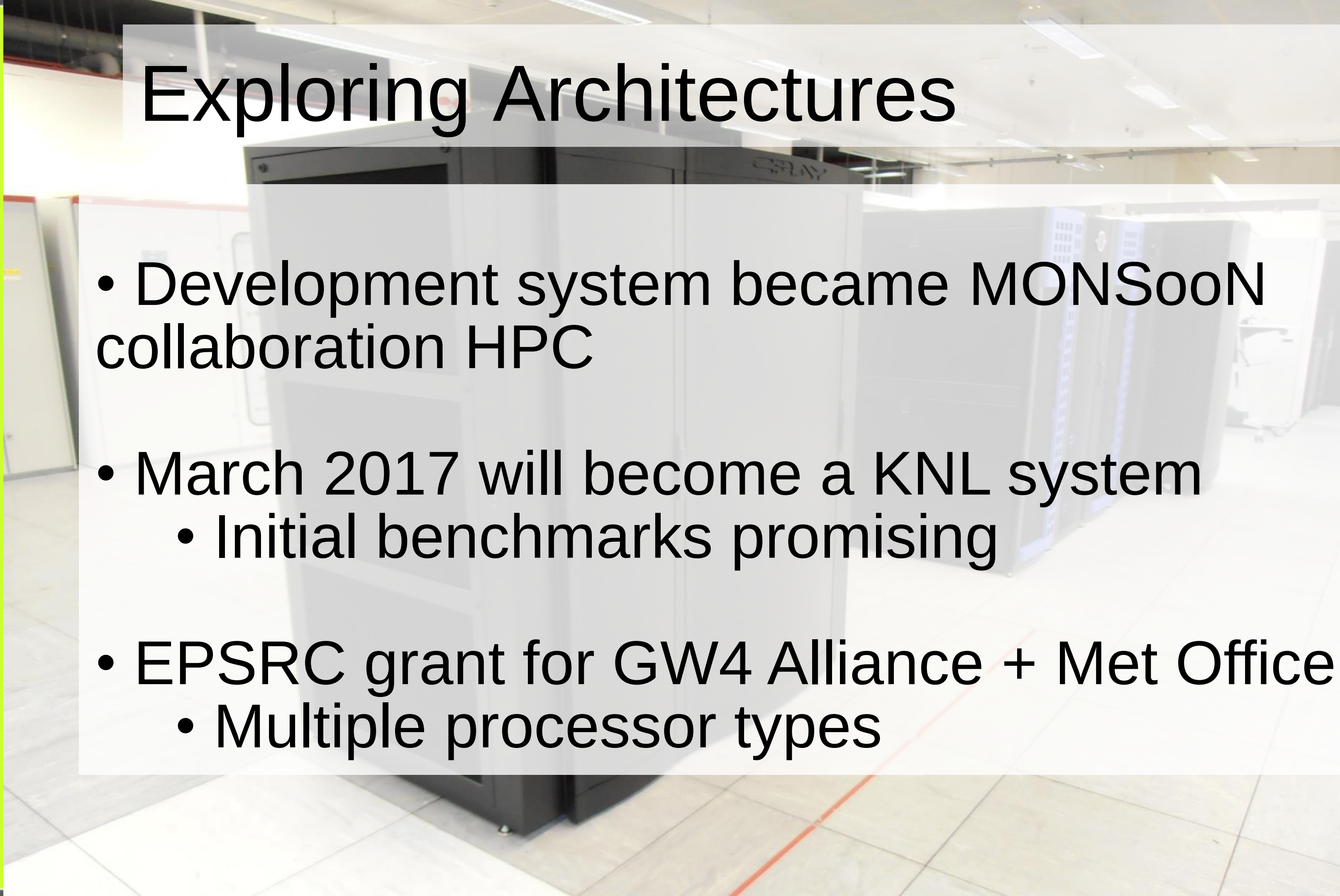
Cray XC40 – Phase 1c

Cray XC40 – Phase 1c



- Installed and being bedded down
- Acceptance starts 8th Nov – on schedule
- 36 cabinet Broadwell system - 6720 nodes
- Twin path networking between sites
- Separate Sonexion storage (12 PB)

Exploring Architectures



- Development system became MONSooN collaboration HPC

- March 2017 will become a KNL system
  - Initial benchmarks promising

- EPSRC grant for GW4 Alliance + Met Office
  - Multiple processor types

Porting process

- November 2014 – August 2015
- > 40 software systems
- 99 Science / IT staff
- Two parallel suites for operations
- > 13 climate configurations ported / validated

Met Office

Problems encountered

- Mostly straightforward
- Thanks to ECMWF, DWD, KMA, ...
- Preparation to reduce metadata accesses

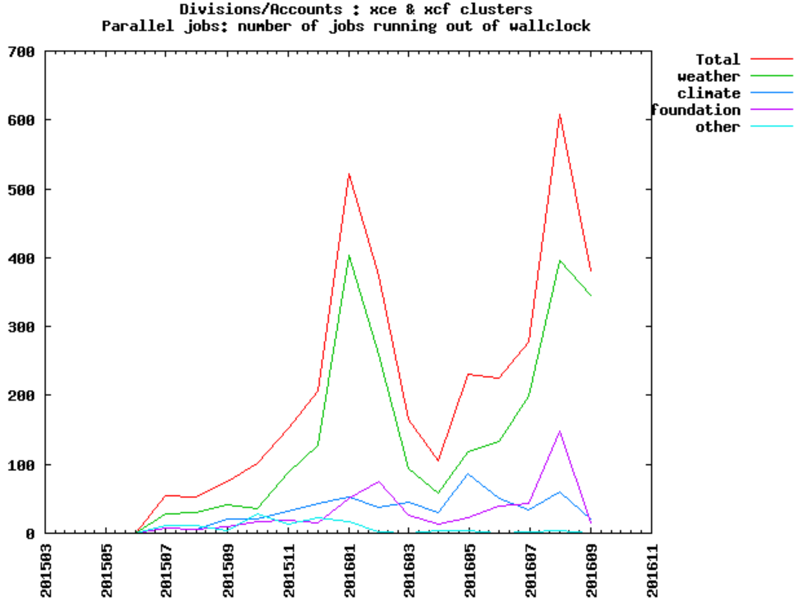
- Scheduling
- Lustre – RAID check
- Lustre – python bytecode

PBS 13 works at scale, PBS 12 doesn’t

- No problems with scheduling on Phase 1a
- On Phase 1b
  - High priority work was OK
  - PBS 12 didn’t schedule lots of research work
  - Job evaluation too slow
  - Machine underutilisation
  - Frustrated users
  - All fixed by PBS 13

RAID checks hurt performance

- Observed poor I/O performance
- User jobs hit wallclock limits
- Regular and exceptional
- Can take days...



Python bytecode

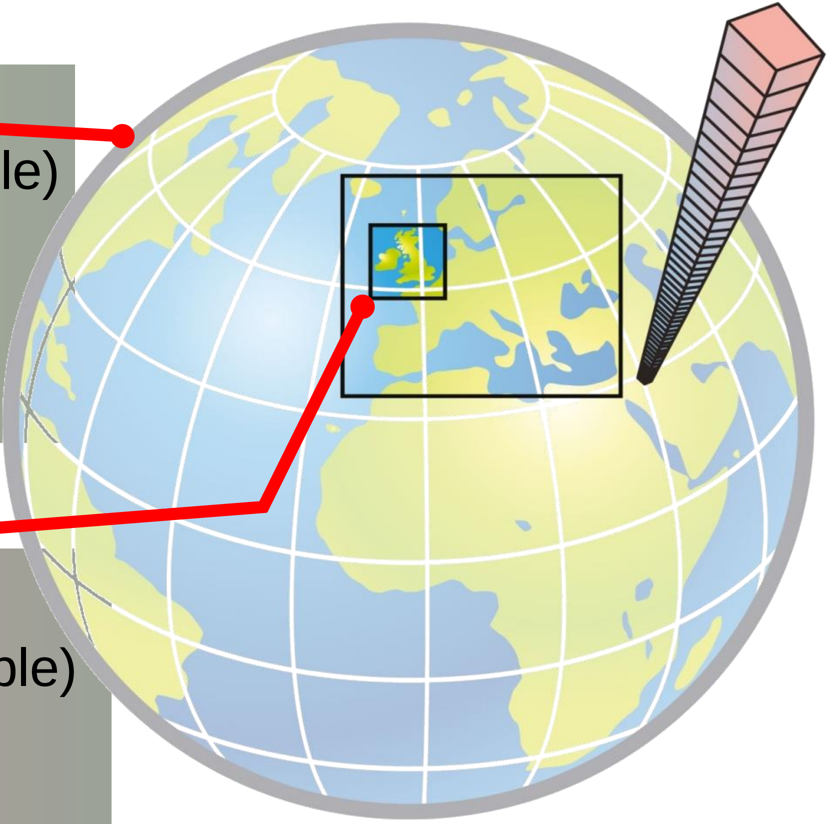
- Met Office suites run via *cylc*, written in python
- A sysadmin used *cylc* as root and different python version
- All user jobs needed to recompile `.pyc` files, but couldn’t
- Metadata load on login nodes made all workflows stall
- `PYTHONDONTWRITEBYTECODE=True` is a good thing on Lustre!

Model Plans

Current NWP Configuration Details

Global

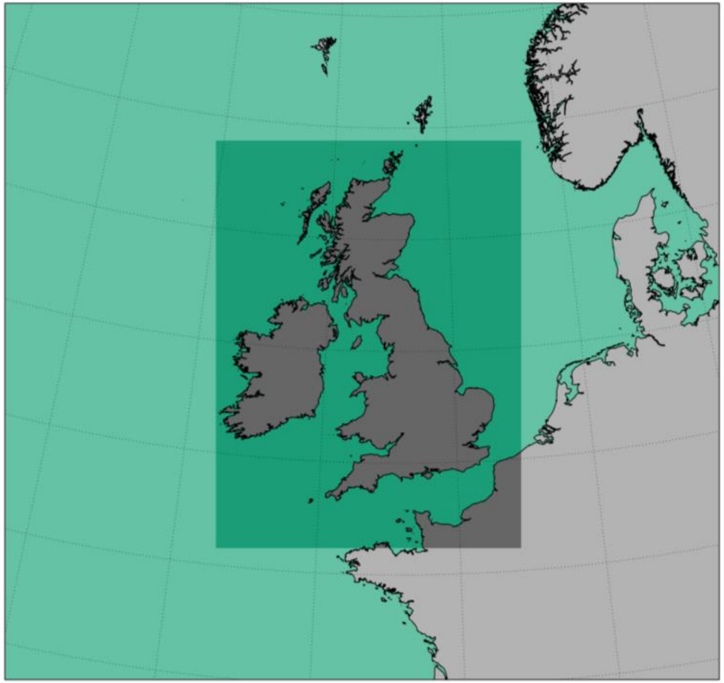
- 17km resolution (33km ensemble)
- 70 vertical levels (80km top)
- 48 hour forecast twice/day
- 6 day forecast twice/day
- Hybrid 4DVar DA



UKV

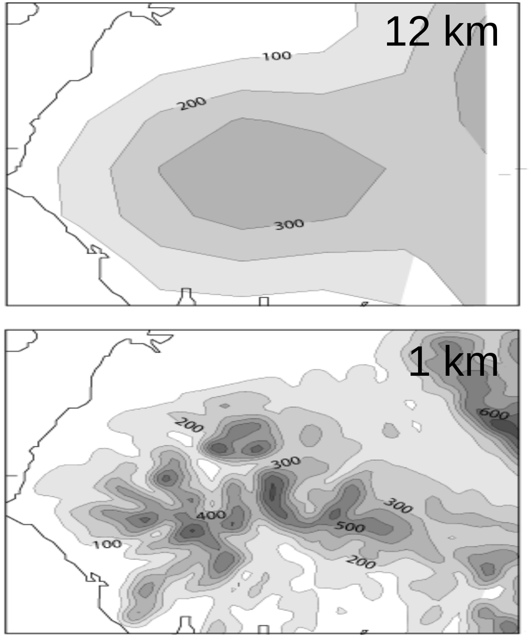
- 1.5km UK model (2.2km ensemble)
- 70 vertical levels (40km top)
- 36 hour forecast eight times/day
- 3DVar DA

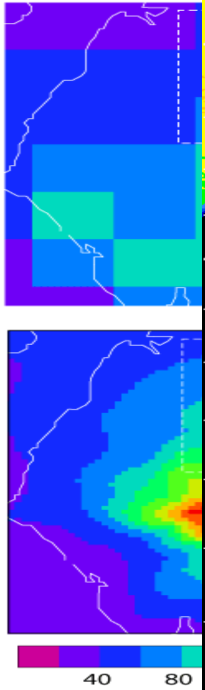
New UK model configuration (Autumn 2016)

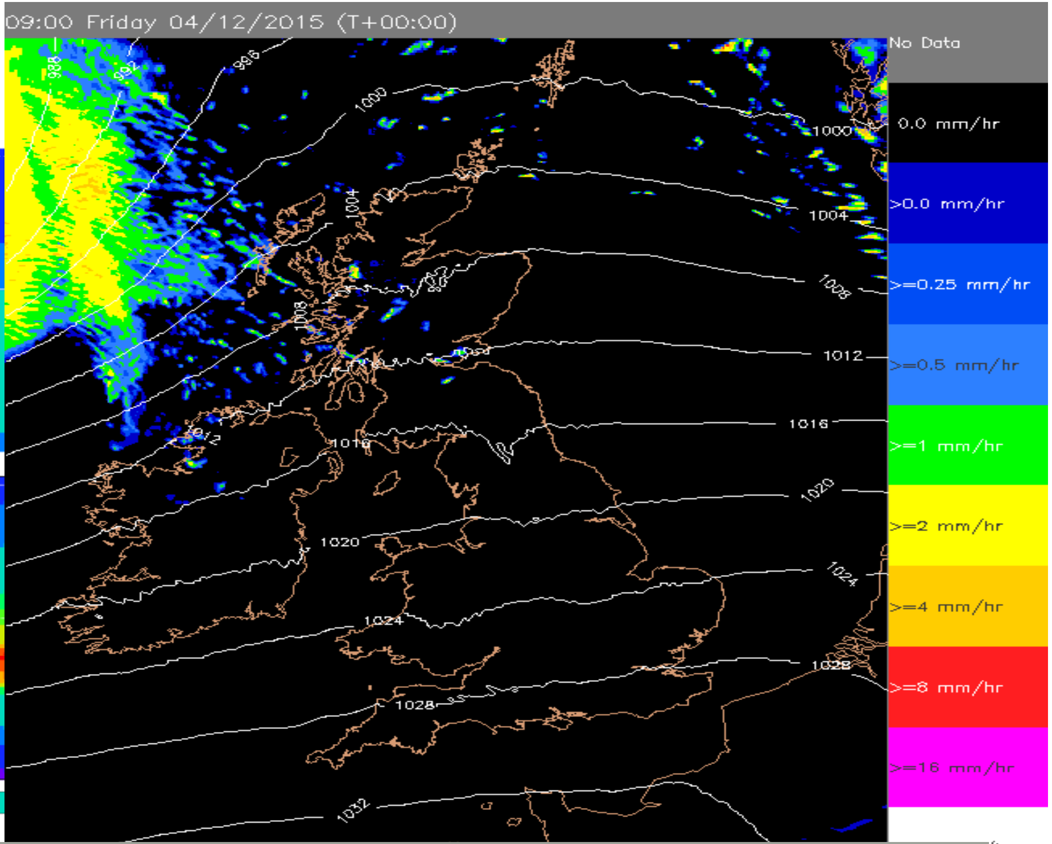


- Expanded domain size
- Exploit high-resolution skill further into forecast period:
  - T+36 extended to T+120 (03/15Z).
  - T+36 to T+54 (00/06/09/12/18/21Z).
  - All MOGREPS-UK ensemble members also extended to T+54
- Improved physics (e.g. convection initiation)
- Additional satellite data.

Storm Desmond (4-6/12/15)



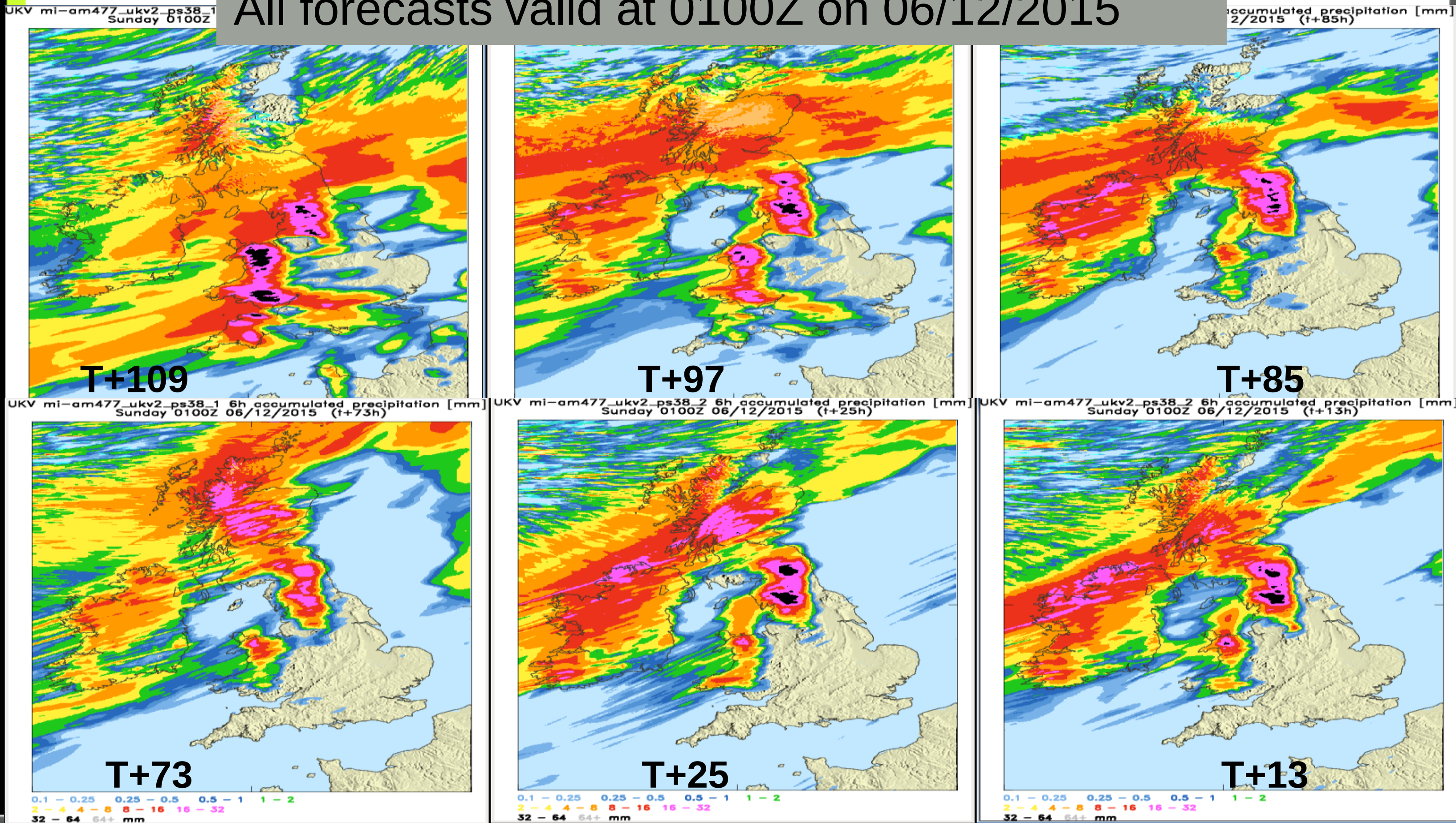




Note: UK model accumulations up to 250mm; global all < 100mm.

Storm Desmond Re-run To 5 Days

All forecasts valid at 0100Z on 06/12/2015


UKV mi-am477_ukv2_ps38_1
Sunday 0100Z
accumulated precipitation [mm]
2/2015 (t+85h)
T+109
T+97
T+85
UKV mi-am477_ukv2_ps38_1 6h accumulated precipitation [mm]
Sunday 0100Z 06/12/2015 (t+73h)
UKV mi-am477_ukv2_ps38_2 6h accumulated precipitation [mm]
Sunday 0100Z 06/12/2015 (t+25h)
UKV mi-am477_ukv2_ps38_2 6h accumulated precipitation [mm]
Sunday 0100Z 06/12/2015 (t+13h)
T+73
T+25
T+13
0.1 – 0.25 0.25 – 0.5 0.5 – 1 1 – 2
2 – 4 4 – 8 8 – 16 16 – 32
32 – 64 64+ mm

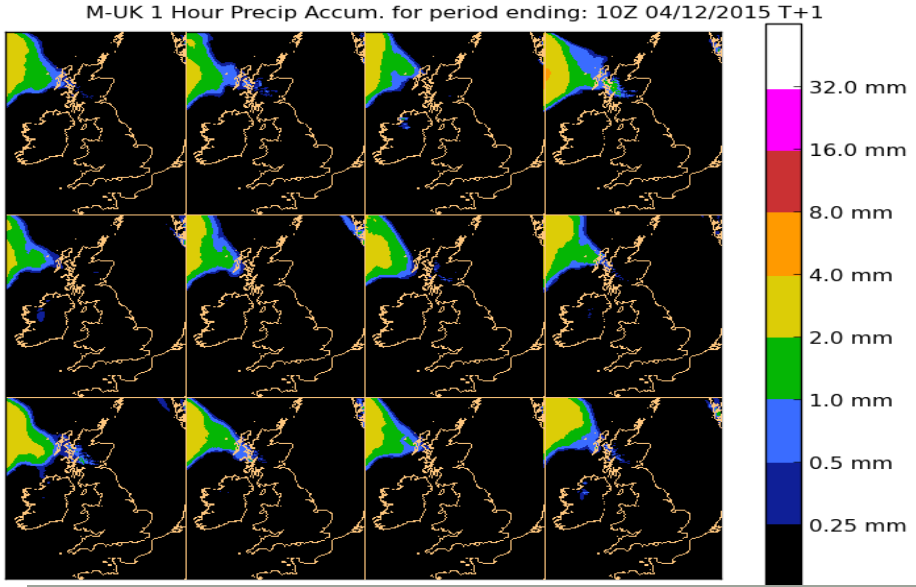
www.metoffice.gov.uk

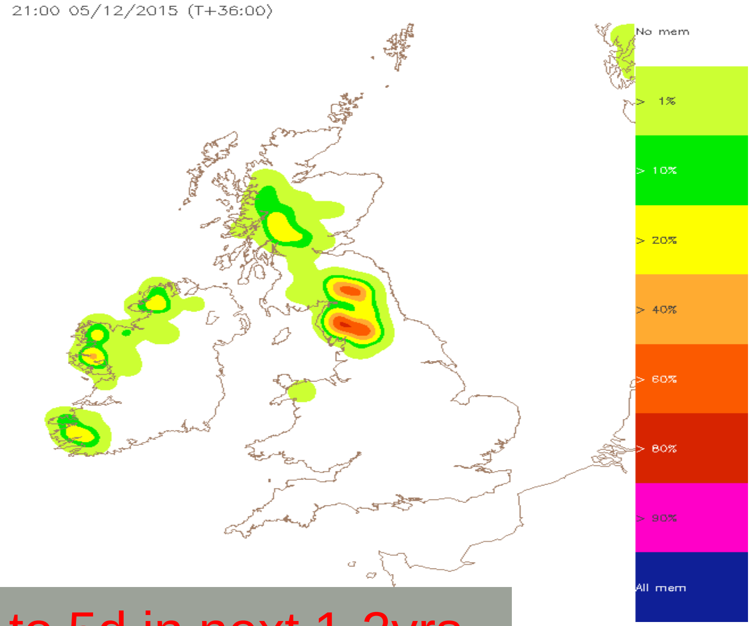
Ensemble forecasting at 2.2km resolution

09Z run on 4th December 2015

12 × 2.2km resolution hourly rainfall accumulation forecasts from MOGREPS-UK

Probability 24 hour rainfall > 100mm. Valid for the period 2100 4th December to 2100 5th December

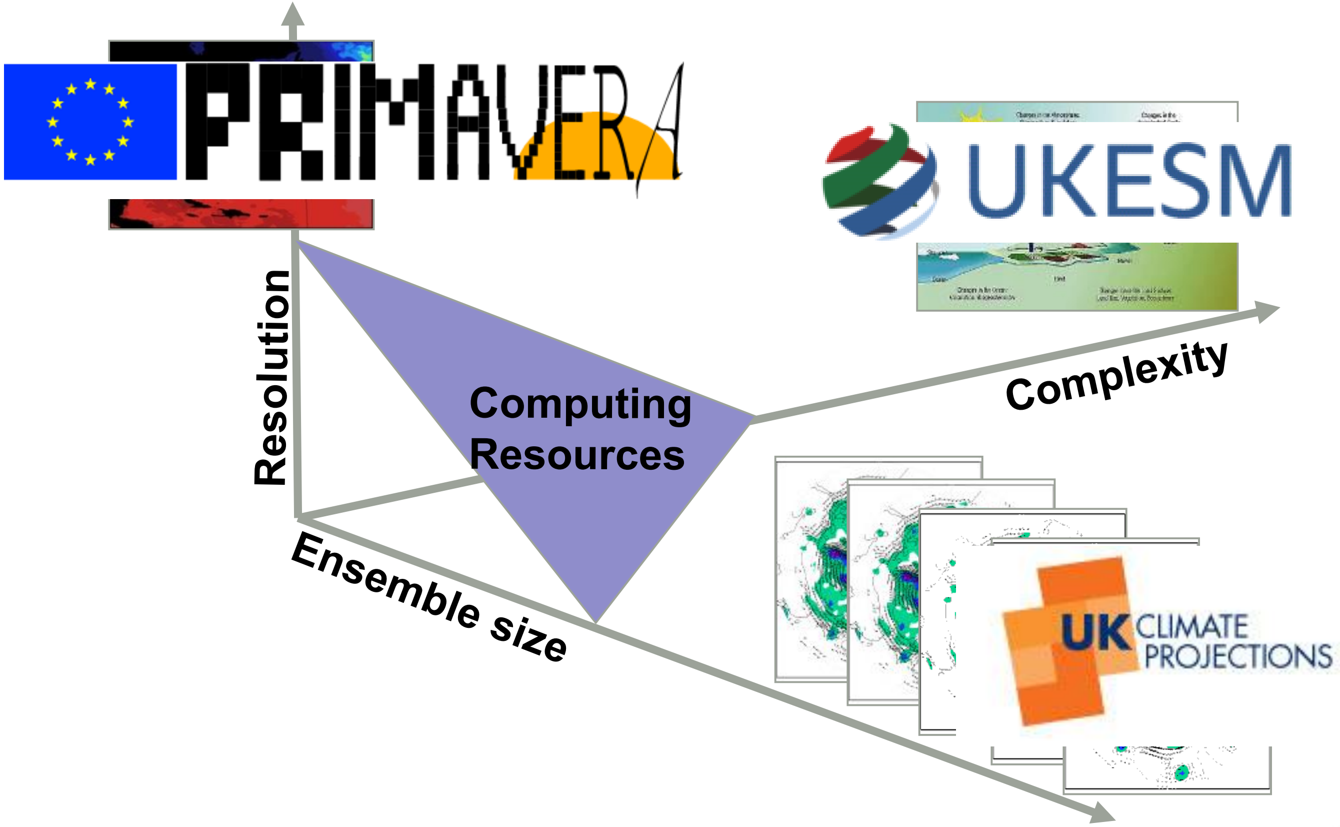

M-UK 1 Hour Precip Accum. for period ending: 10Z 04/12/2015 T+1
32.0 mm
16.0 mm
8.0 mm
4.0 mm
2.0 mm
1.0 mm
0.5 mm
0.25 mm



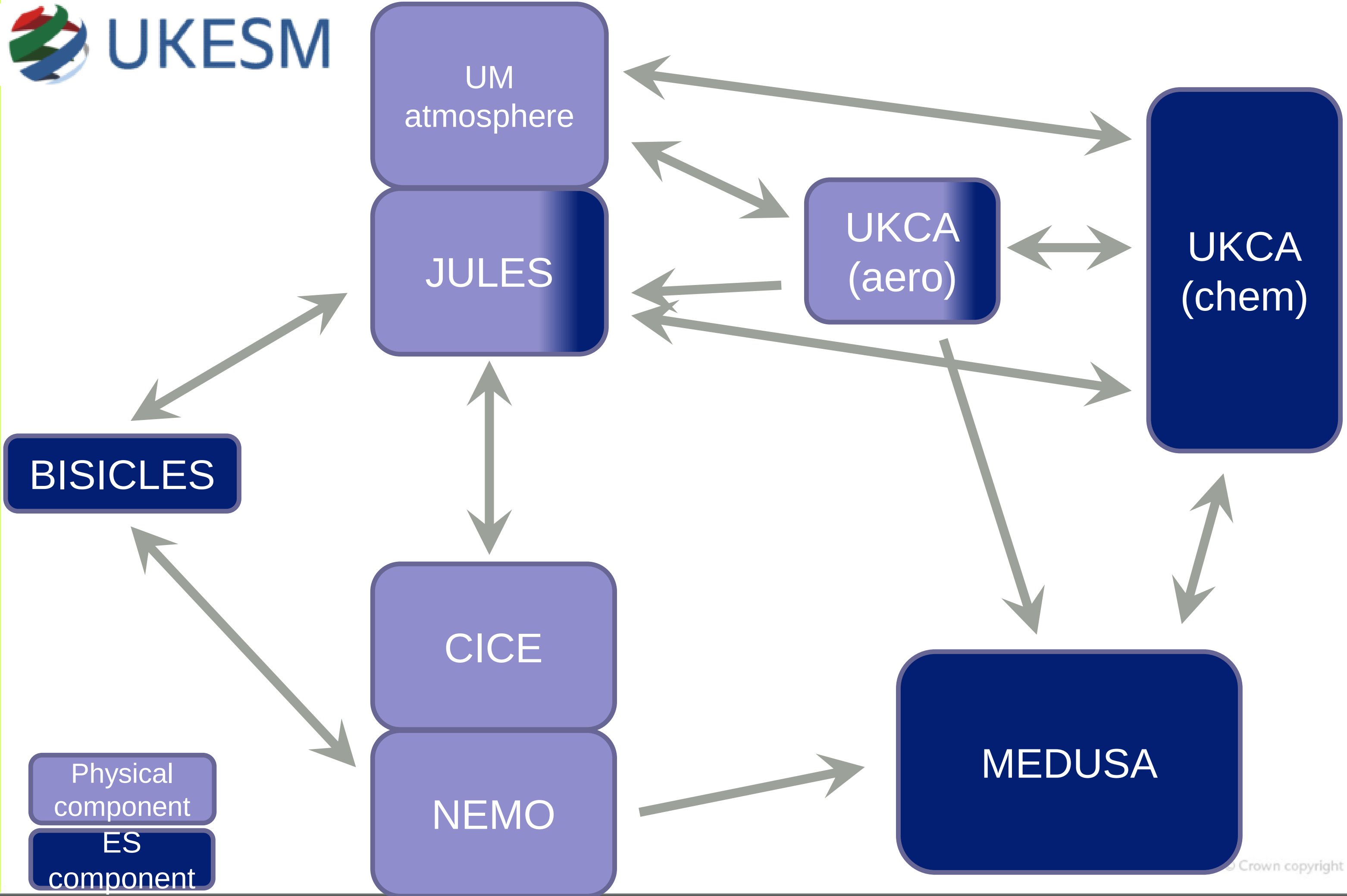
Plan to extend range of MOGREPS-UK to 5d in next 1-2yrs

Met Office

Improving Climate Models



UKESM


UM atmosphere
JULES
UKCA (aero)
UKCA (chem)
BISICLES
CICE
NEMO
MEDUSA
Physical component
ES component

Scaling

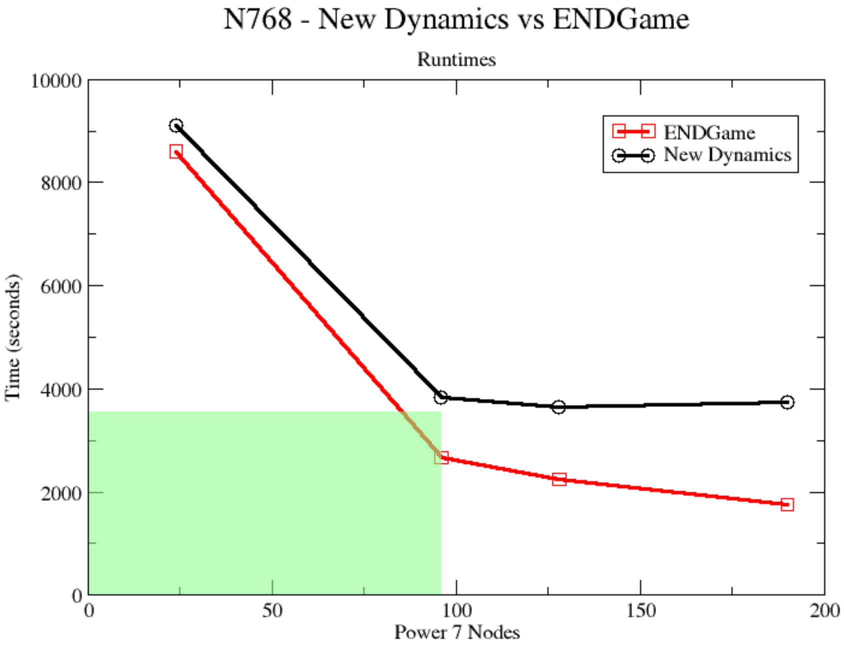
Motivation

- ENDGame enabled 17km global forecasts in 2014

- Comfortable with 10-12km

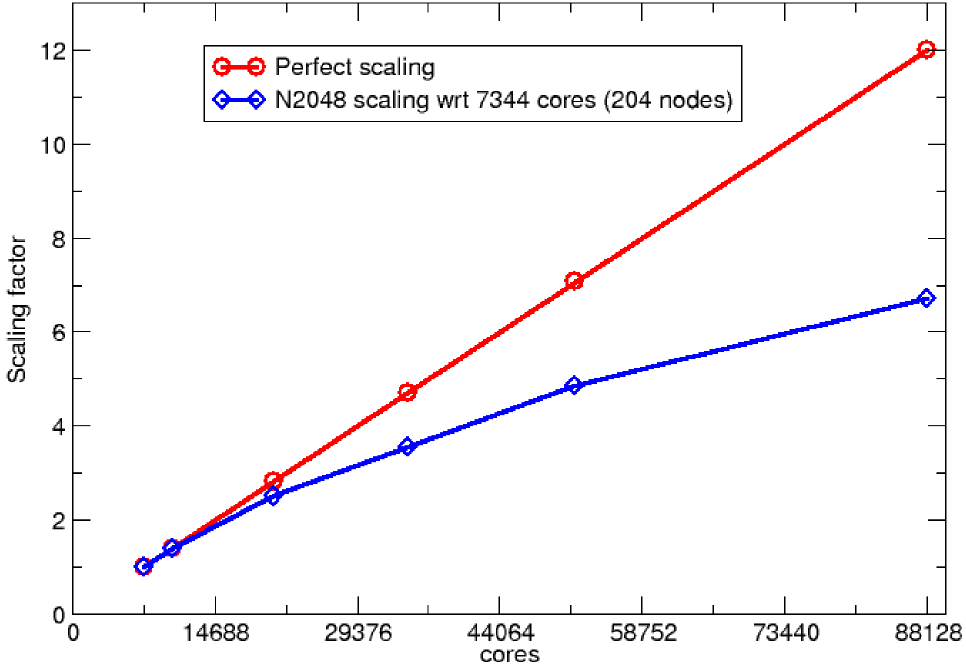
- LFRIC due in 2020s

- How far can ENDGame take us?



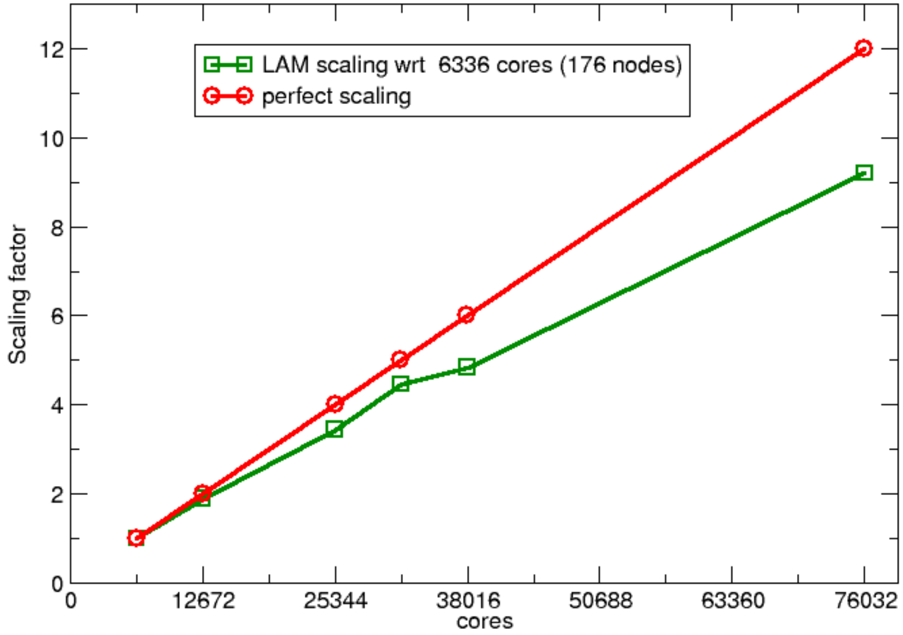
Global Model N2048 scaling

- Core numerics
- No diagnostics
- Not counting input dump read
- Cray XC40 with 36 core Broadwell nodes



Regional Model scaling

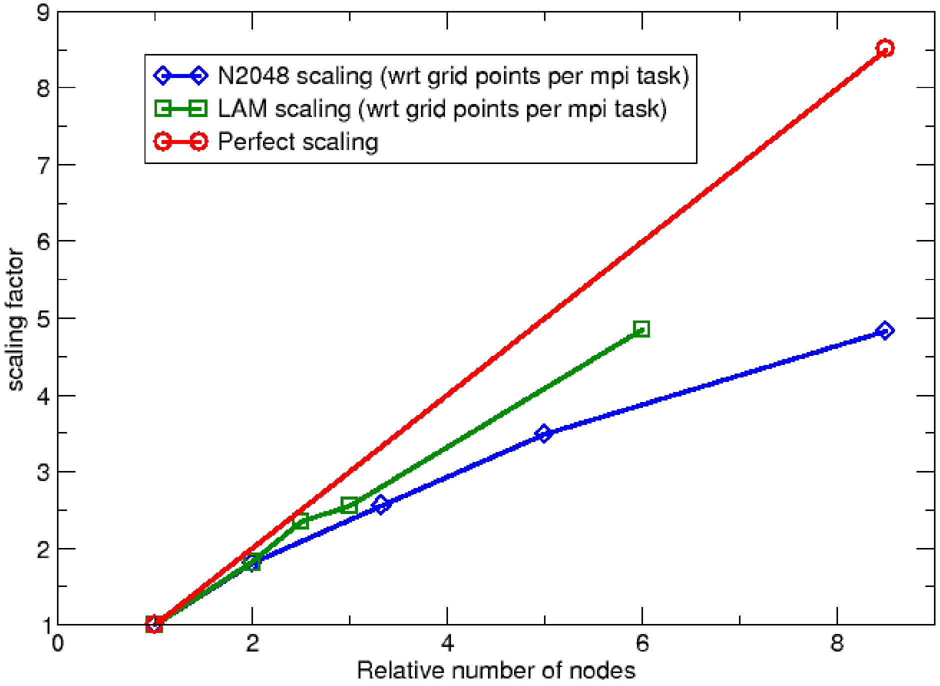
- Setup as for global model
- Large Europe model equivalent to 300m UK
- Similar behaviour for variety of timestep lengths



Scaling comparison

- Comparison for number of gridpoints per MPI task

- Regional model better load-balanced and no poles



Thank You!

Questions?