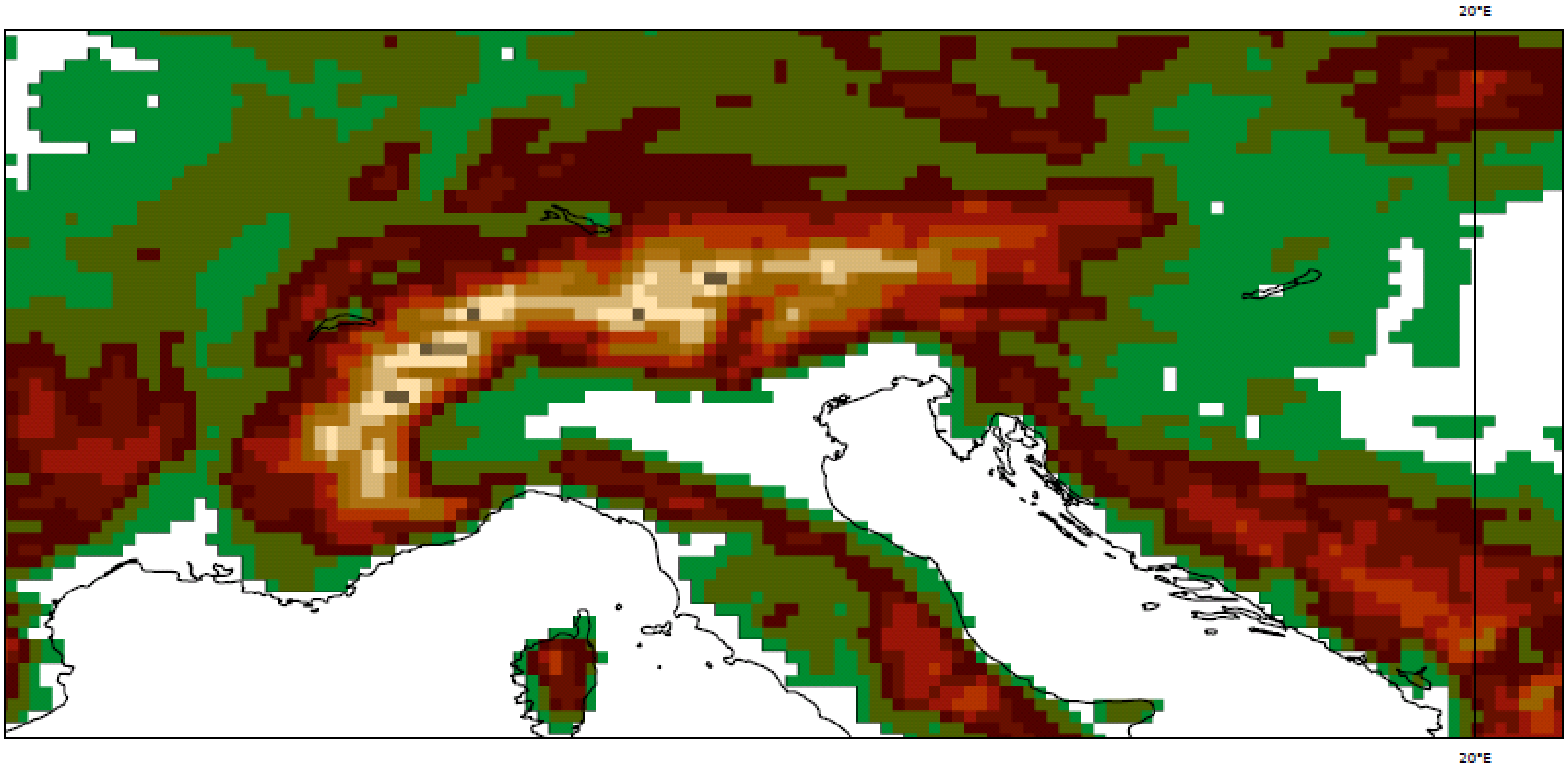
The ECMWF forecast model, quo vadis ?

by Nils Wedi
European Centre for Medium-Range Weather Forecasts
wedi@ecmwf.int

contributors: Piotr Smolarkiewicz, Mats Hamrud, George Mozdzynski,
Sylvie Malardel, Christian Kuehnlein, Willem Deconinck, Jean Bidlot, Linus Magnusson

ECMWF

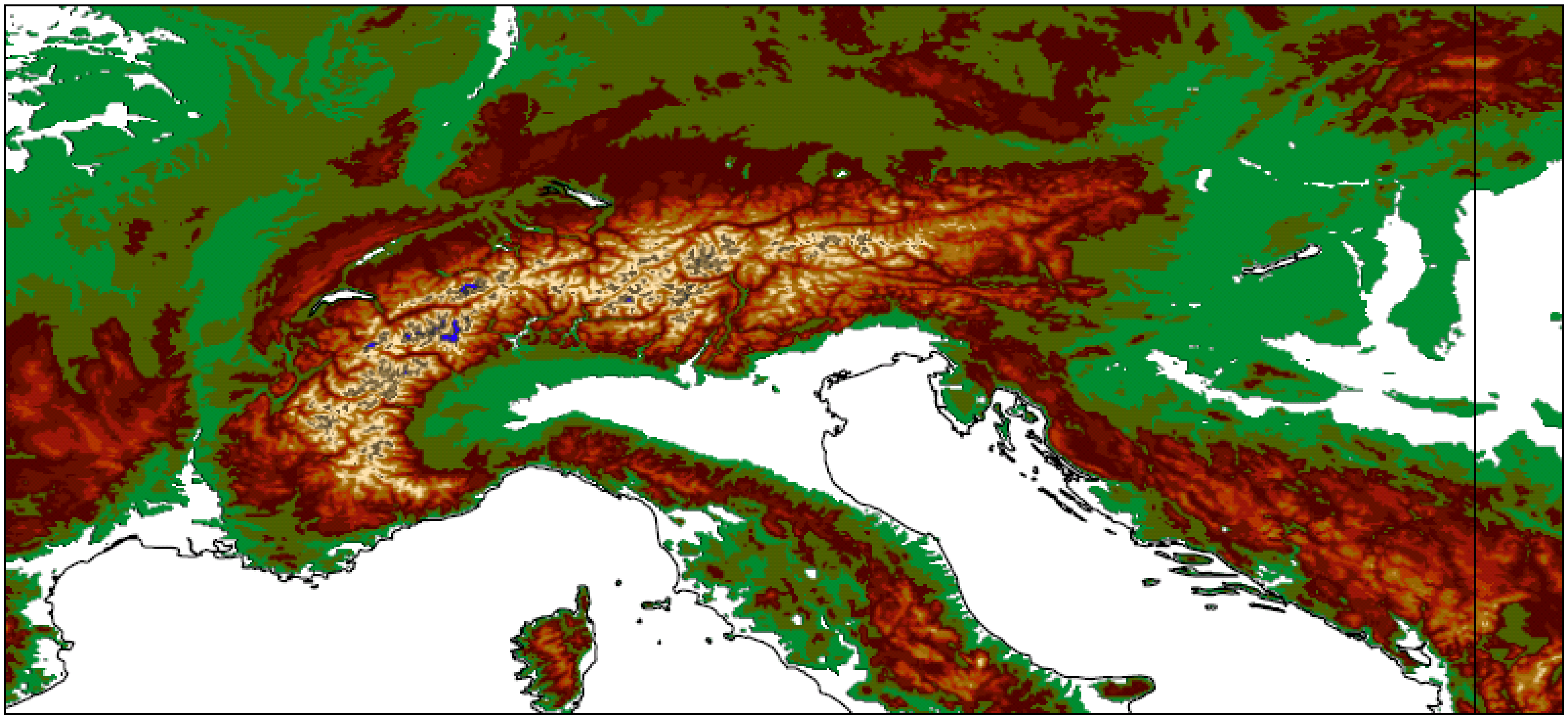
Orography – T1279 (16km)



Alps

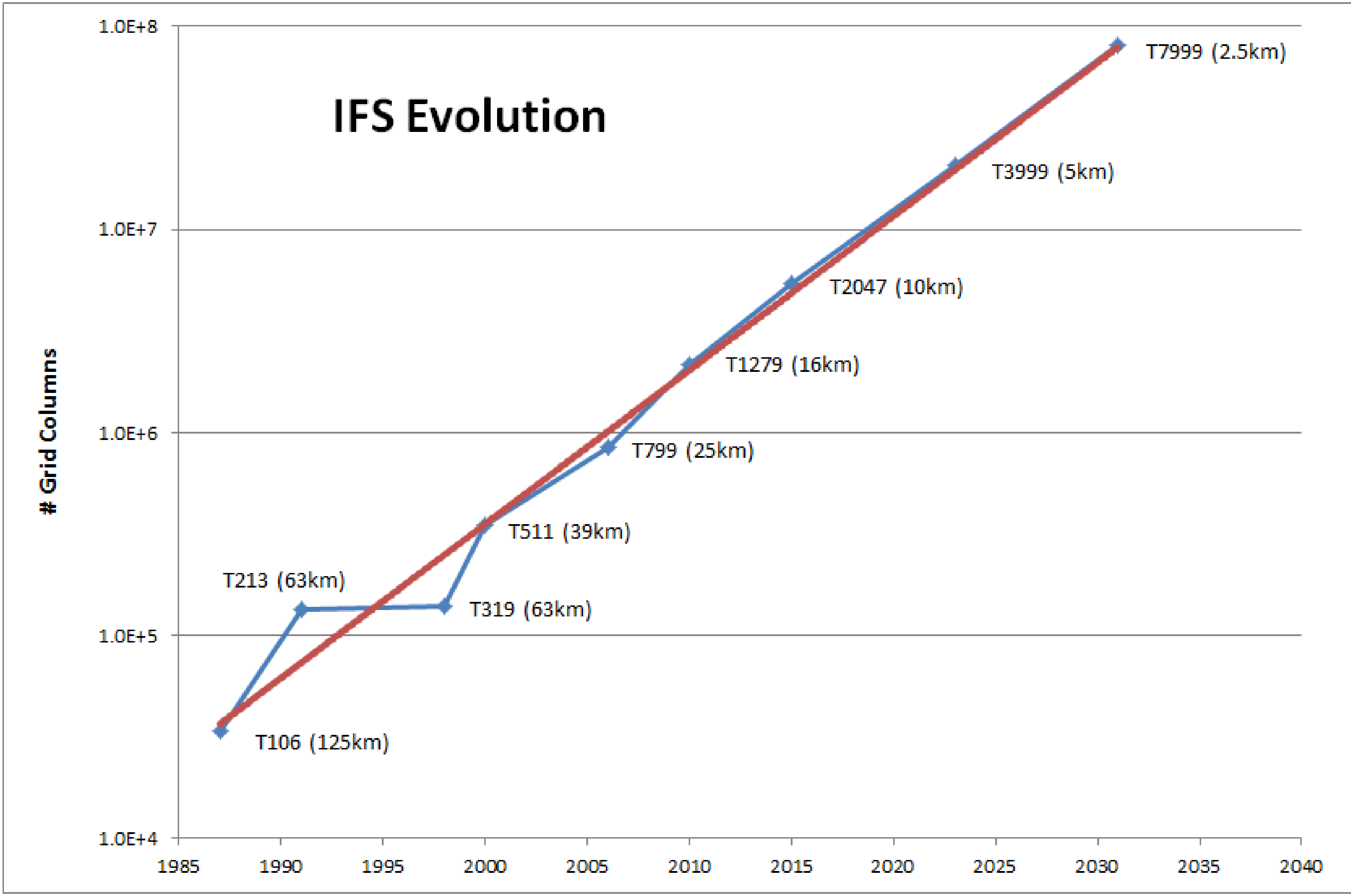
ECMWF

Orography - T7999 (2.5km)



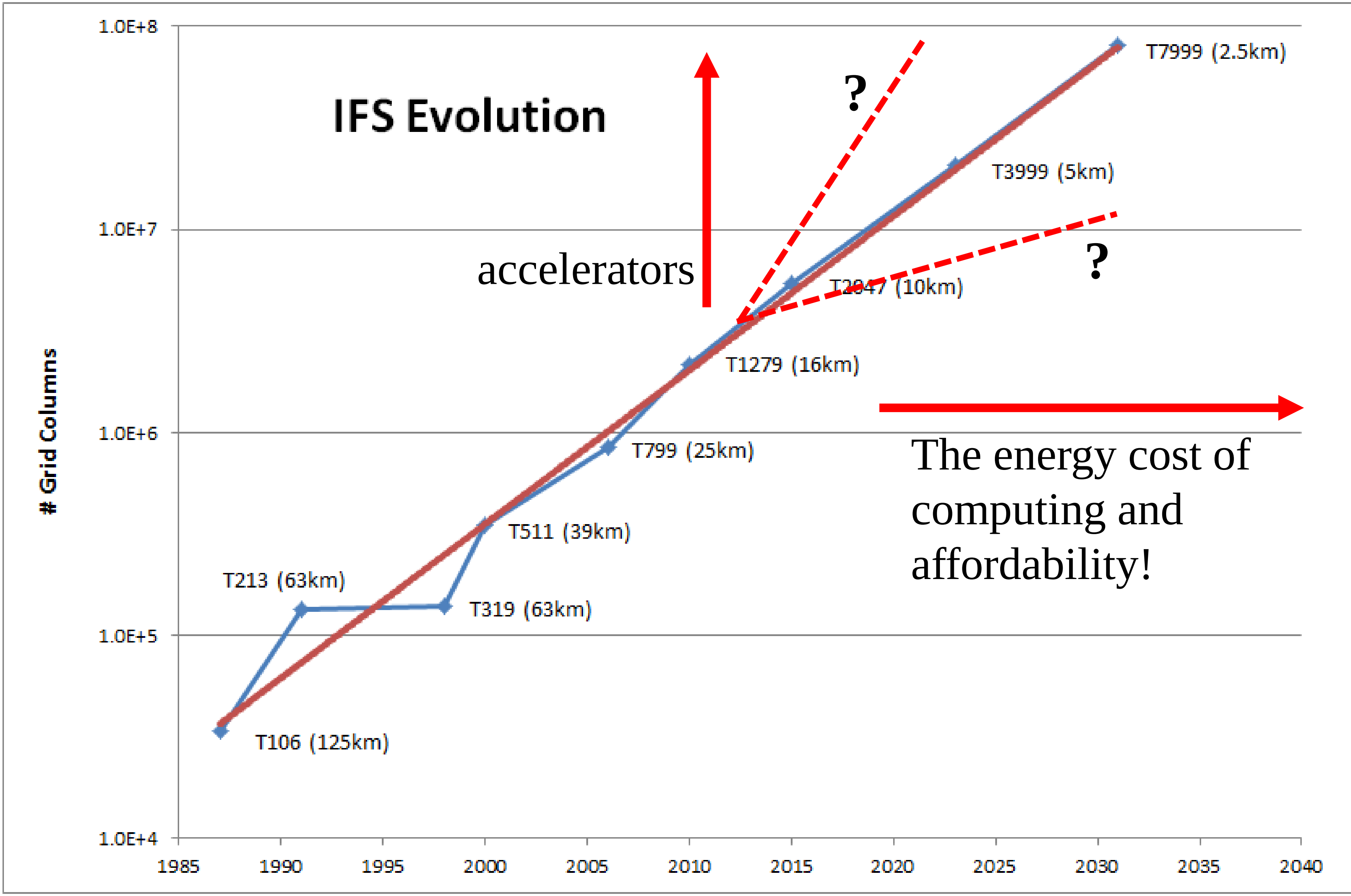
Alps

ECMWF

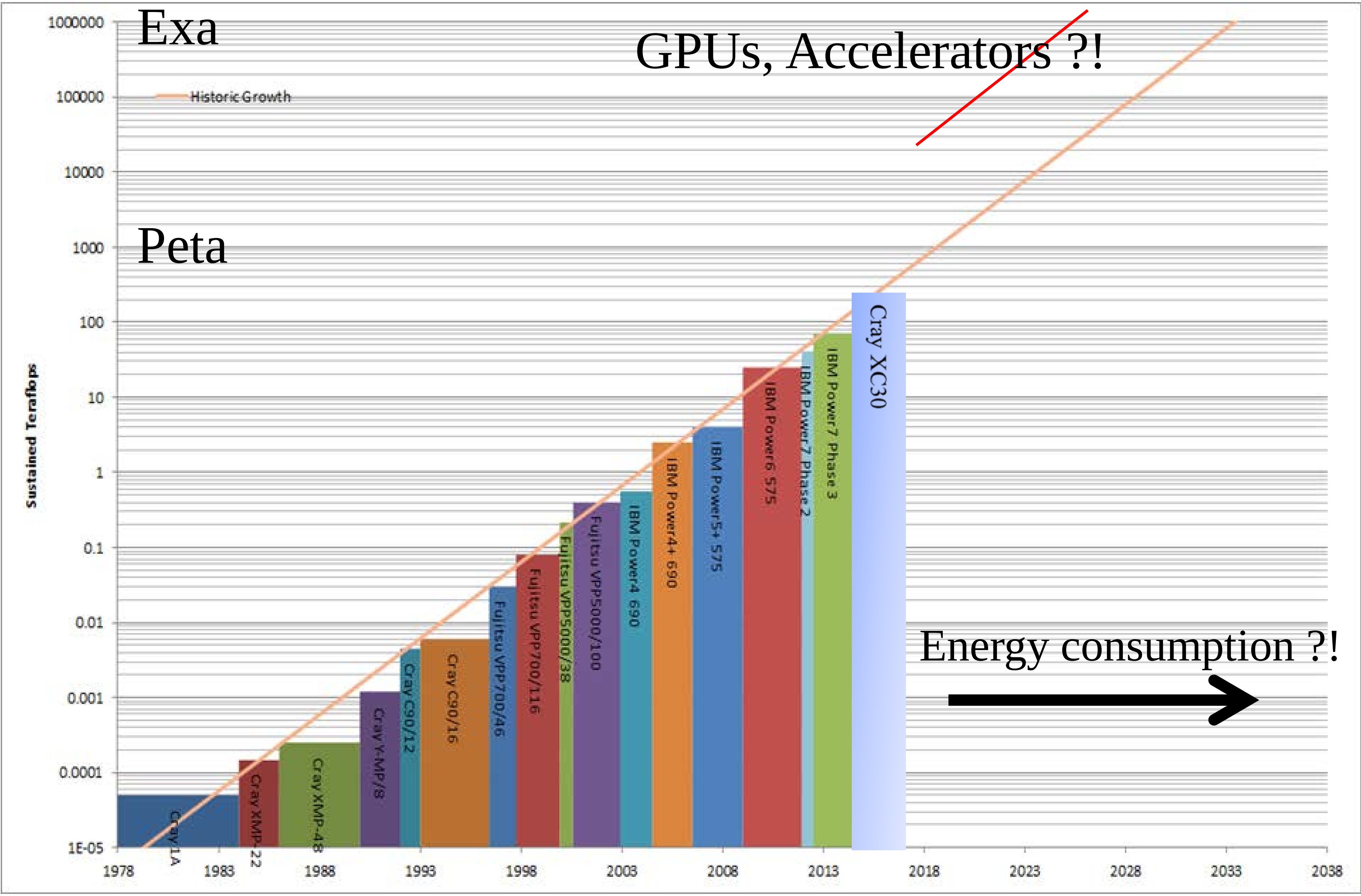


IFS Evolution
Grid Columns
1.0E+8
1.0E+7
1.0E+6
1.0E+5
1.0E+4
1985
1990
1995
2000
2005
2010
2015
2020
2025
2030
2035
2040
T7999 (2.5km)
T3999 (5km)
T2047 (10km)
T1279 (16km)
T799 (25km)
T511 (39km)
T213 (63km)
T319 (63km)
T106 (125km)

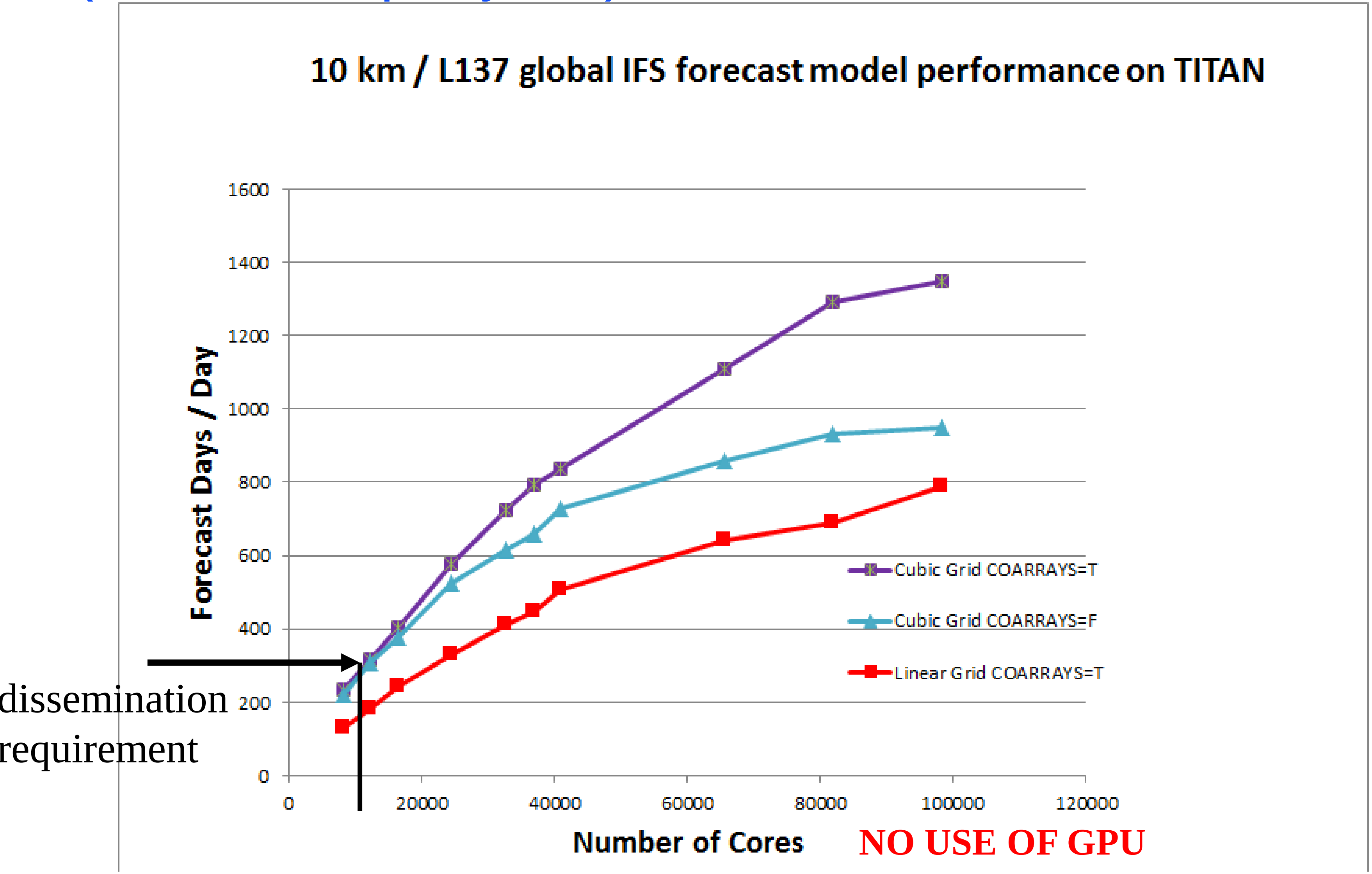
ECMWF


IFS Evolution
accelerators
?
?
The energy cost of computing and affordability!
Grid Columns
1.0E+8
1.0E+7
1.0E+6
1.0E+5
1.0E+4
T7999 (2.5km)
T3999 (5km)
T2047 (10km)
T1279 (16km)
T799 (25km)
T511 (39km)
T213 (63km)
T319 (63km)
T106 (125km)
1985
1990
1995
2000
2005
2010
2015
2020
2025
2030
2035
2040

ECMWF

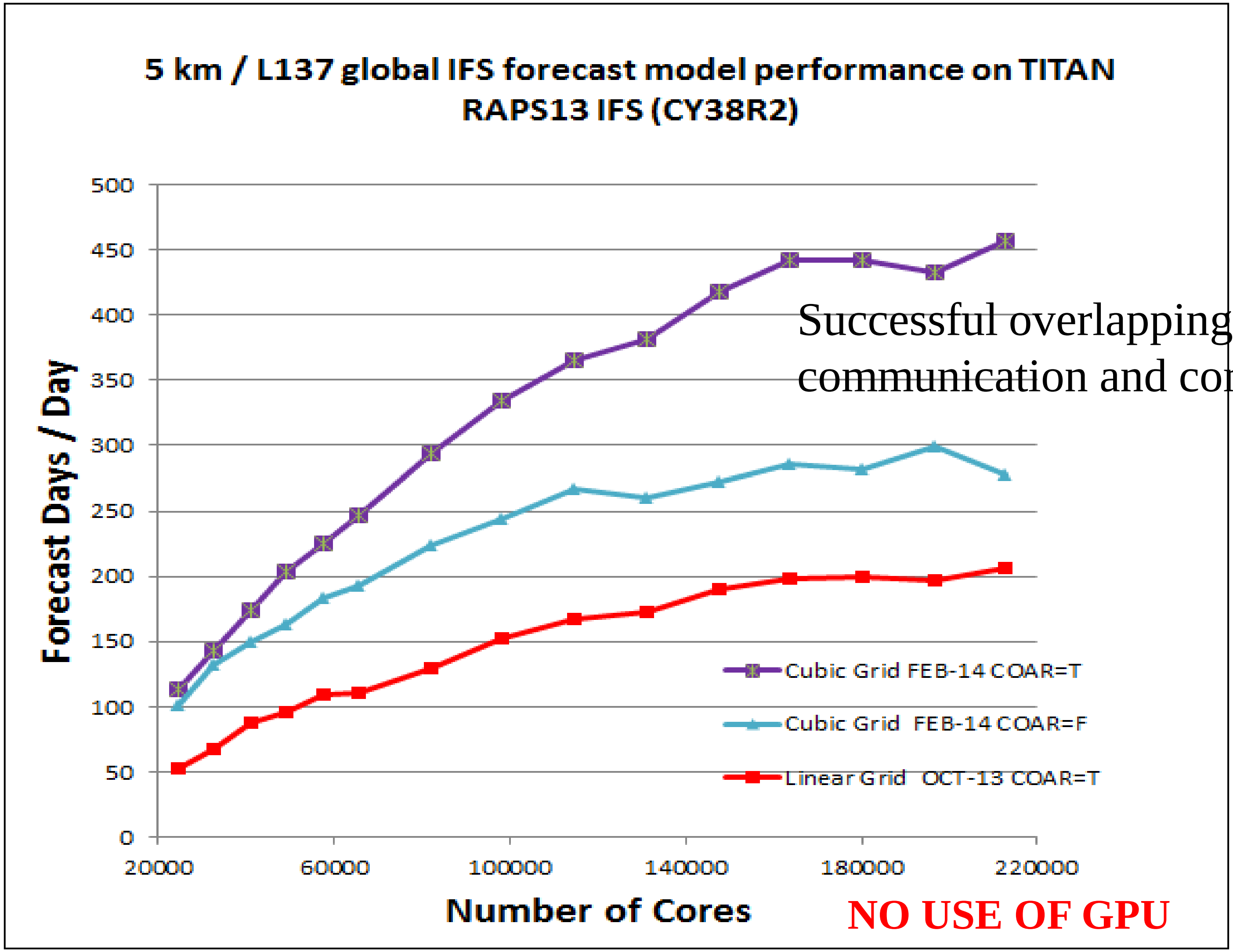

Exa
Peta
GPUs, Accelerators ?!
Energy consumption ?!
Historic Growth
Sustained Teraflops
1000000
100000
10000
1000
100
10
1
0.1
0.01
0.001
0.0001
1E-05
1978
1983
1988
1993
1998
2003
2008
2013
2018
2023
2028
2033
2038
Cray 1A
Cray XMP-22
Cray XMP-48
Cray Y-MP/8
Cray C90/12
Cray C90/16
Fujitsu VPP700/46
Fujitsu VPP700/116
Fujitsu VPP5000/38
Fujitsu VPP5000/100
IBM Power4 690
IBM Power4+ 690
IBM Power5+ 575
IBM Power6 575
IBM Power7 Phase 2
IBM Power7 Phase 3
Cray XC30

10 km IFS model scaling on TITAN (CRESTA project)



ECMWF

5 km IFS model scaling on TITAN (CRESTA project)



Successful overlapping of communication and computation!

NO USE OF GPU

ECMWF

The "cost" of communication:

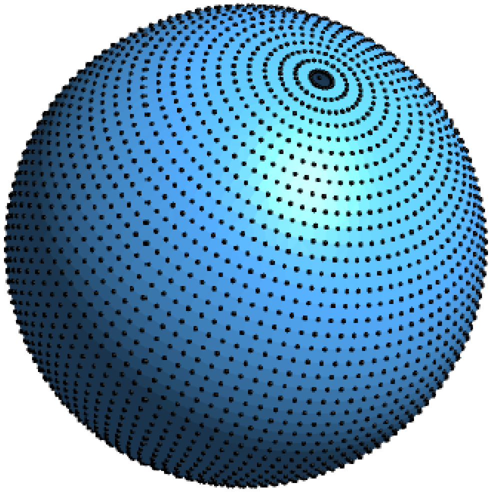
- Assume the following:
  - model time step of 30 seconds
  - 10 day forecast
  - model on 4M cores
  - max 1 hour wall clock
- 1 step needs to run in under 0.125 seconds
- Using 32 threads per task with 128.000 MPI tasks
- A simple MPI_SEND from 1 task to all other 128K tasks will take an estimated 128.000 x 1 microsec = 0.128 seconds

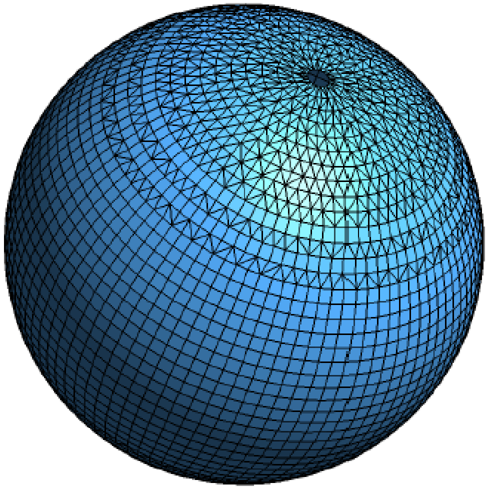
- Implies global communications cannot be used, likely
- each task needs to run with 100's or 1000's of threads or equivalent

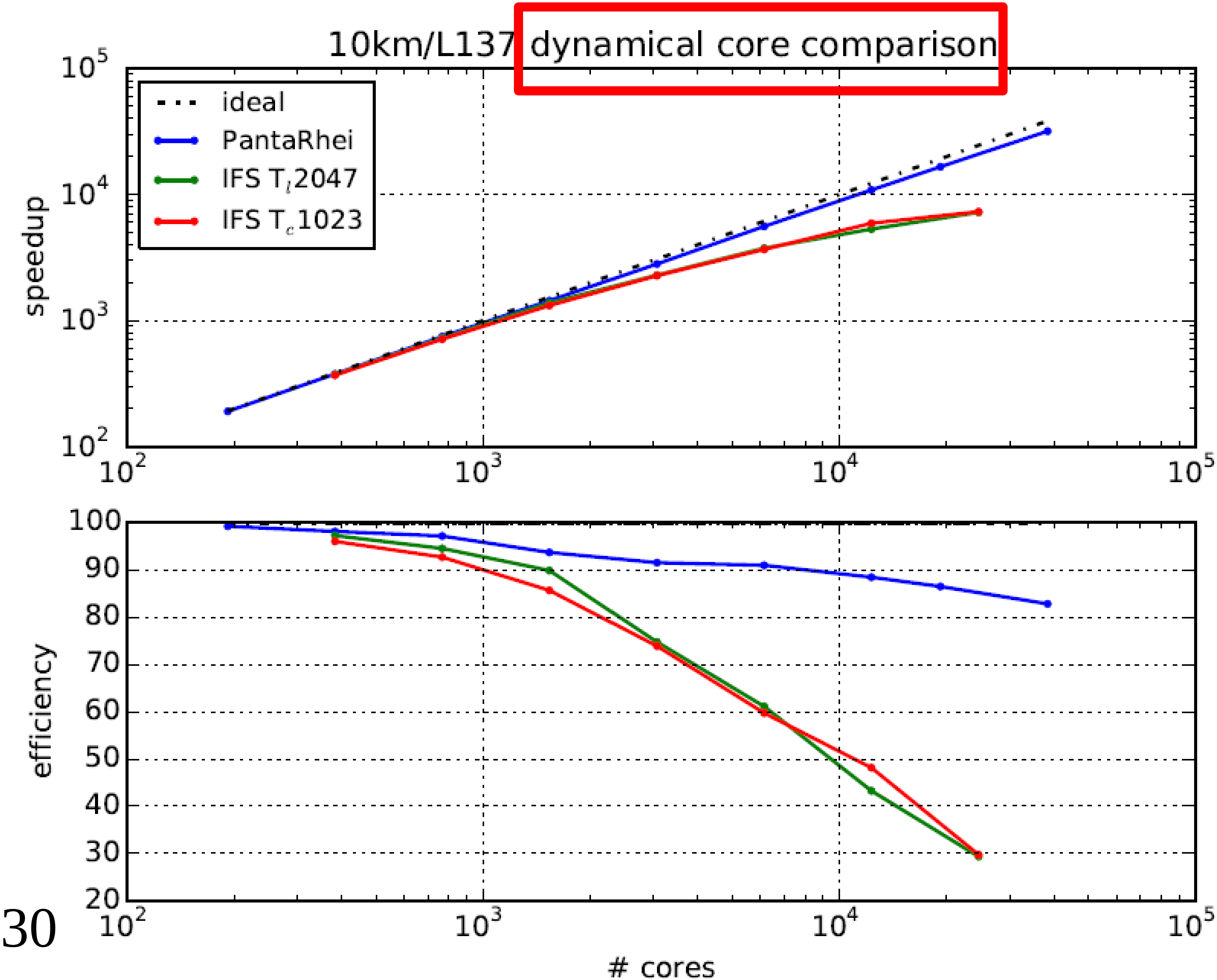
  => aim for max O(10.000) MPI tasks

PantaRhei + CRESTA + Loughborough University

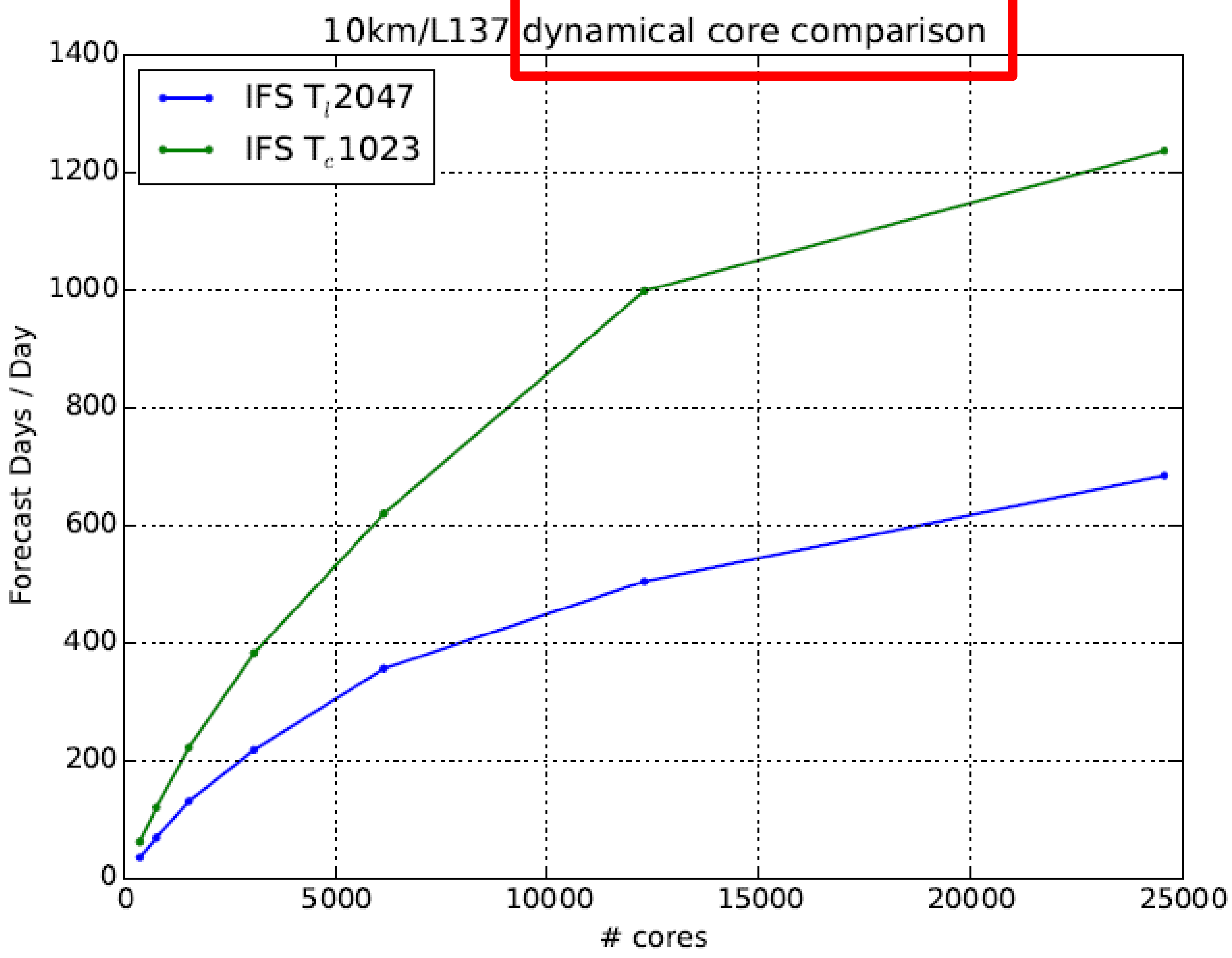
- Flexible unstructured mesh data structure with options to retain reduced Gaussian grid nodes
- Change of equations
- Compact stencil, minimizing communication and data movement
- Fast route to competitive real weather simulations






10km/L137 dynamical core comparison
speedup
efficiency
cores
ideal
PantaRhei
IFS T_l2047
IFS T_c1023
10^2
10^3
10^4
10^5
100
90
80
70
60
50
40
30
20

XC30


10km/L137 dynamical core comparison
IFS T_l2047
IFS T_c1023
Forecast Days / Day
cores
0
200
400
600
800
1000
1200
1400
0
5000
10000
15000
20000
25000

XC30

ECMWF

Underlying physics

- ***Multiscale nature*** **of the atmosphere may be exploited to extract further parallelism**
- **Further resolution increases coincide with two major changes in how we model the physical system, with far reaching consequences for numerical algorithms and equations used:**
  - ***Resolved convection***
  - ***Non-hydrostatic effects***
- **This furthers the need to assess and *quantify forecast uncertainty* but also provides opportunities for new algorithms and/or hardware design**

  → E.g. single precision or low-accuracy computations

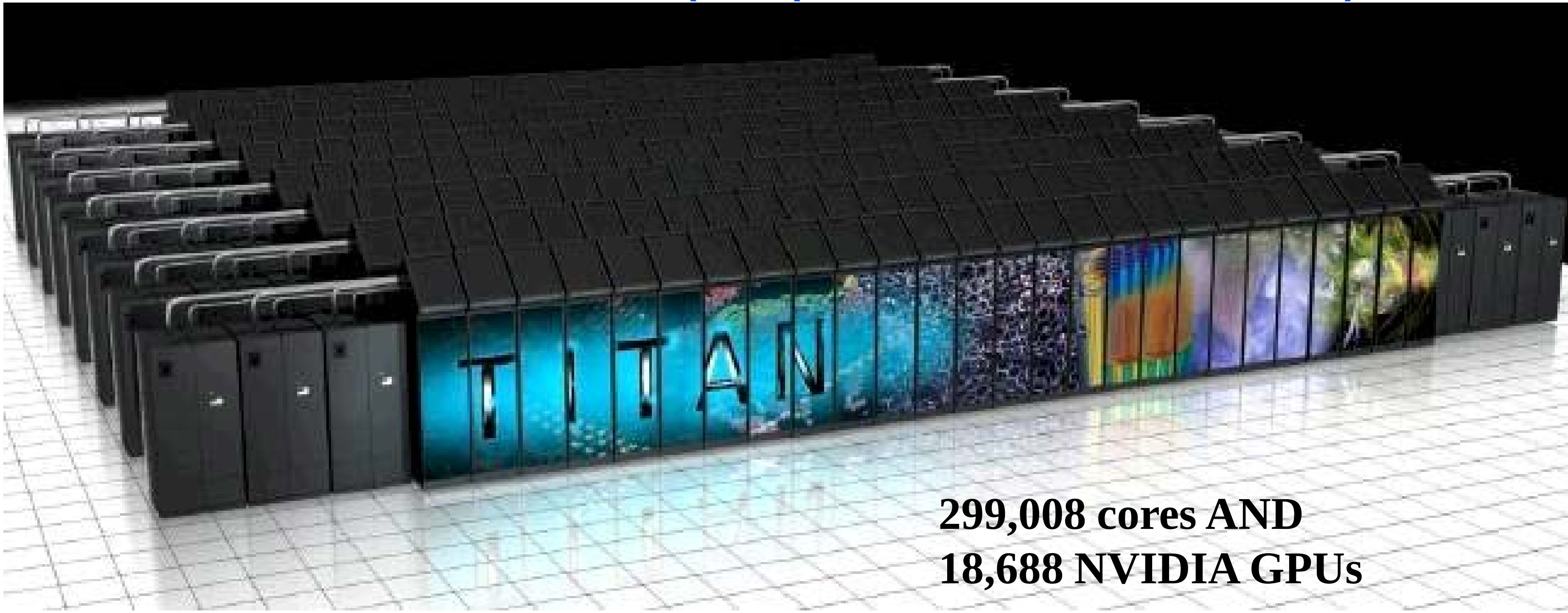
Berkeley Dwarfs and ECMWF's IFS

Recognizing patterns of computation and communication

- **Dense Linear Algebra (matrix/matrix multiply)**
- **Sparse Linear Algebra (Newton – Krylov solver)**
- **Spectral methods**
- **N-body methods (FMM)**
- **Structured grids (data layout)**
- **Unstructured grids (data layout)**
- **MapReduce (I/O, data processing)**
- **New: combinational logic, graph traversal (indirect addressing), dynamic programming (parallel-in-time algorithms), branch-and-bound, graphical models, finite-state machines (event driven) …..**

Titan (Cray), Oakridge, US, No 2 in the world (Top 500, Nov 2013)

~8.2MW

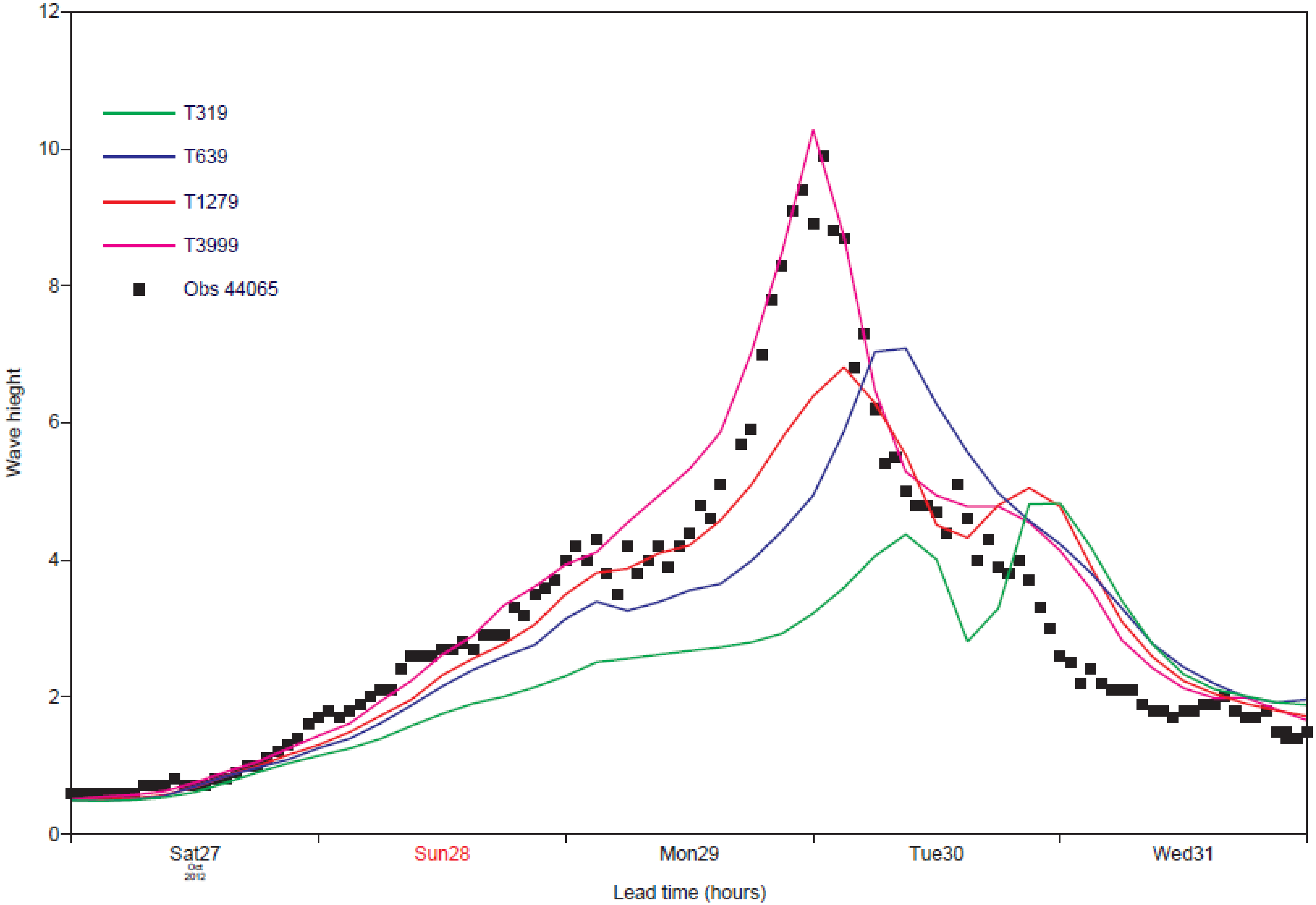


- **Need to explore hybrid architectures**
- **Accelerator adaptation of IFS key components**

Warning the public and saving lives

ECMWF

Wave height 72h forecast, T3999 (~5km)



ECMWF

Efficient coupling to surface, boundary layer, ocean, waves, …

Tianhe-2 (MilkyWay-2)
No 1 in the world (Top 500, Nov 2013)

~17.8MW

!!!!

This computer has 3,120,000 cores,
with Intel Xeon Phi co-processor technology

A 50 member ensemble at ~5km may need this number of cores in order to run in 1 hour.

Energy-aware computing

- **ECMWF uses the equivalent energy comparable to the annual consumption of ~8000 4-bedroom houses!**
- **51 ENS members consume about 330KWh, approximately the same as a single (~5km) global 10-day forecast**
- **Today the energy consumption of *one ENS member* is equivalent to leaving the Kettle on for 3 hours !**

Preparing for the future means for us …

- **Flexibility on the equations solved**
- **Flexibility on the numerical algorithms used**
- **Flexibility on the horizontal and vertical discretization used**
- **Options for the data layout to adapt to massively-parallel, heterogeneous computing architectures**
- **Reduce communication requirements**
- **Develop strategies for resilience**
- **Limit the Mega Watts used per forecast produced!**