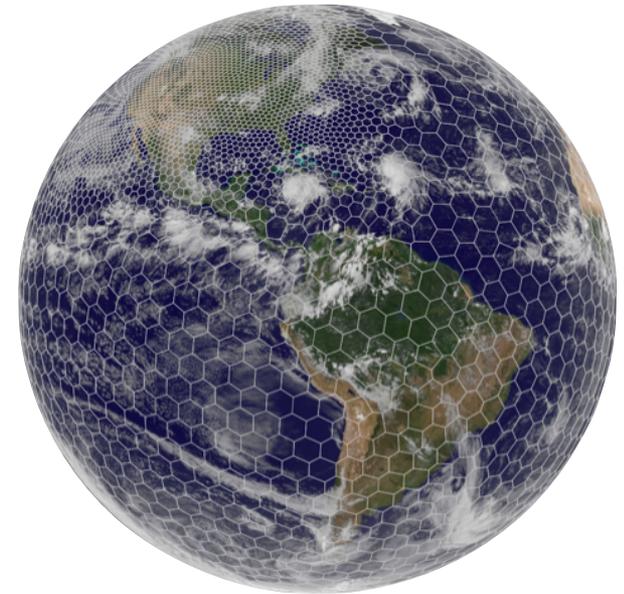


MPAS Extreme Scaling Experiment

Towards convection-resolving, global climate simulations



¹ KARLSRUHE INSTITUTE OF TECHNOLOGY, INSTITUTE OF METEOROLOGY AND CLIMATE RESEARCH

² NATIONAL CENTER FOR ATMOSPHERIC RESEARCH, MESOSCALE AND MICROSCALE METEOROLOGY LABORATORY

² AUGSBURG UNIVERSITY, DEPARTMENT OF GEOGRAPHY, CHAIR OF REGIONAL CLIMATE AND HYDROLOGY

16th Annual WRF Users' Workshop, 17/06/2015

Dominikus Heinzeller¹, Michael Duda² and Harald Kunstmann^{1,3}



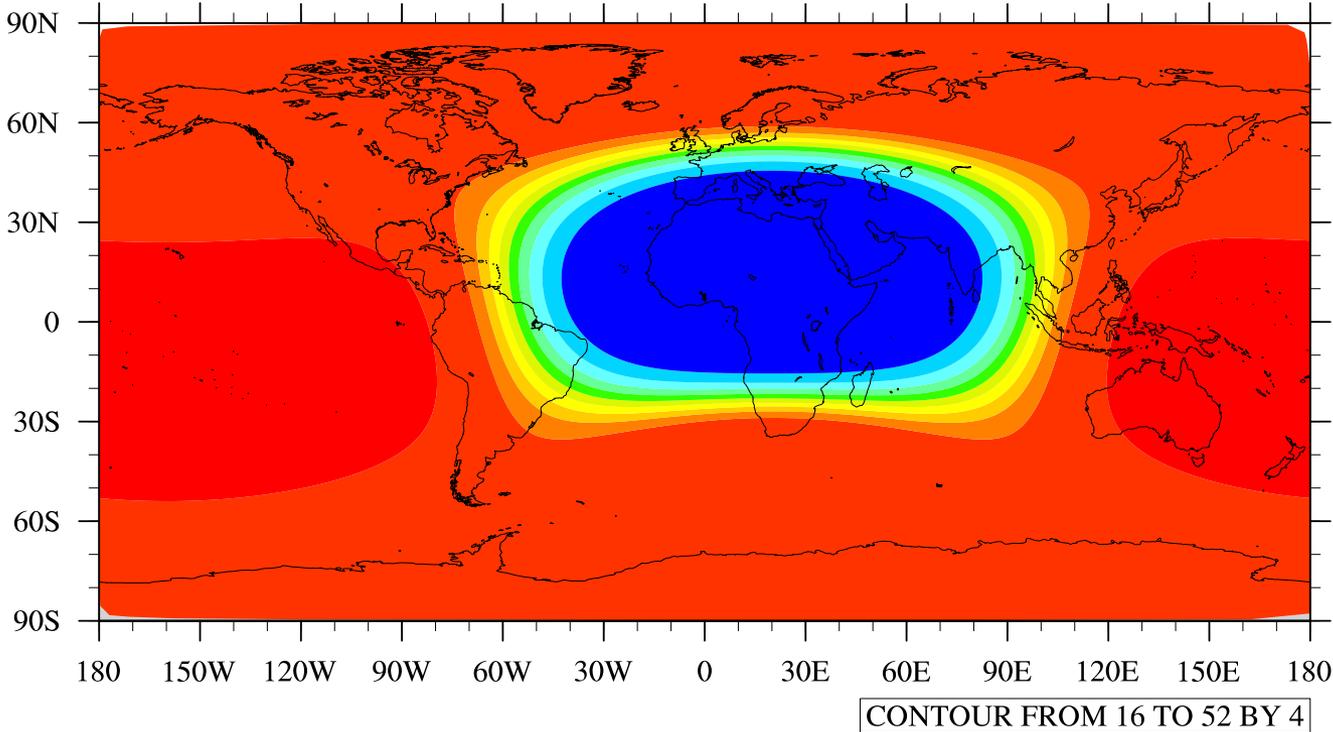
IMK-IFU: Atmospheric Environmental Research



NATIONAL CENTER FOR ATMOSPHERIC RESEARCH

MPAS variable-resolution mesh with 535554 cells

Approximate mesh resolution (km)



Model consistency:

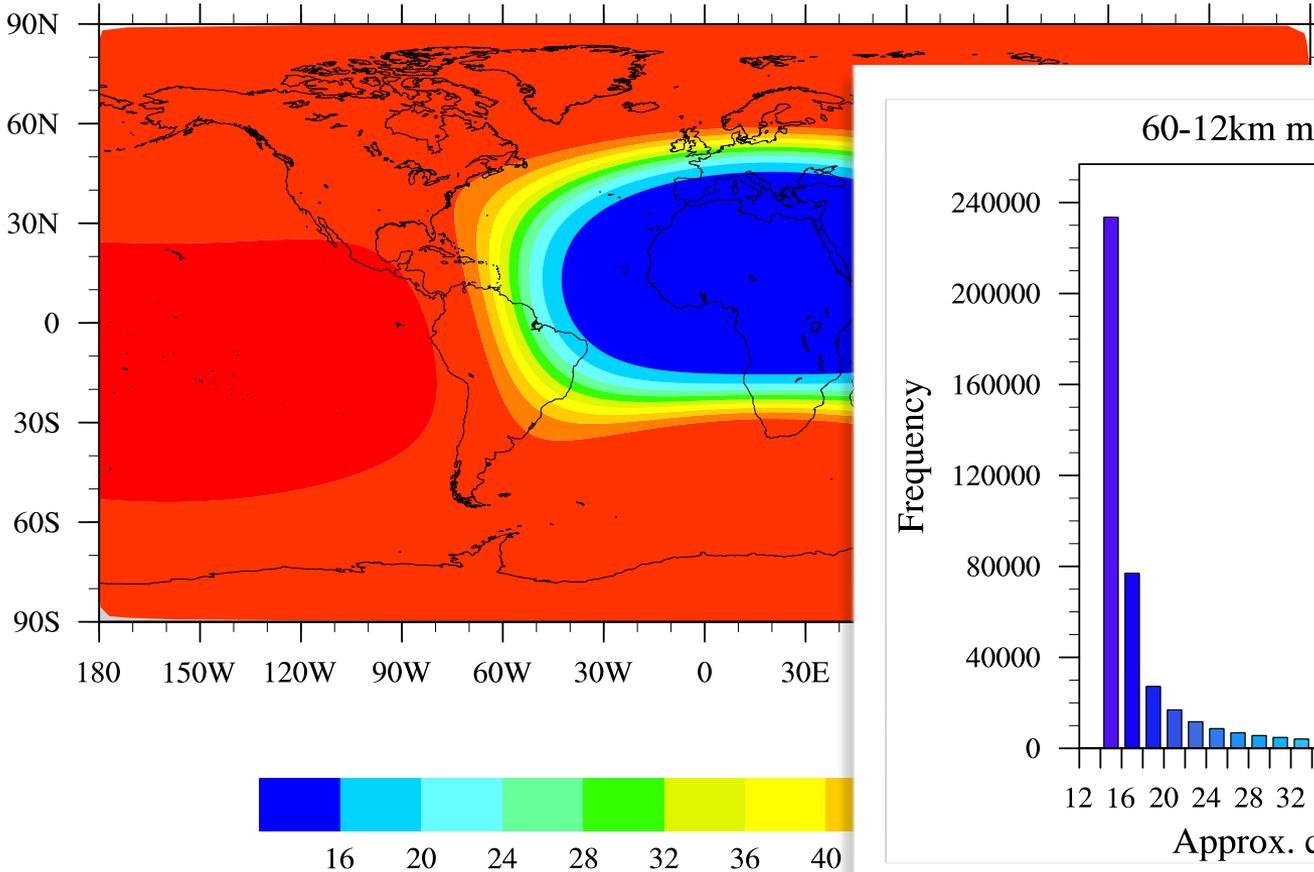
land use / soil type
classification

flow distortions of
jets and waves

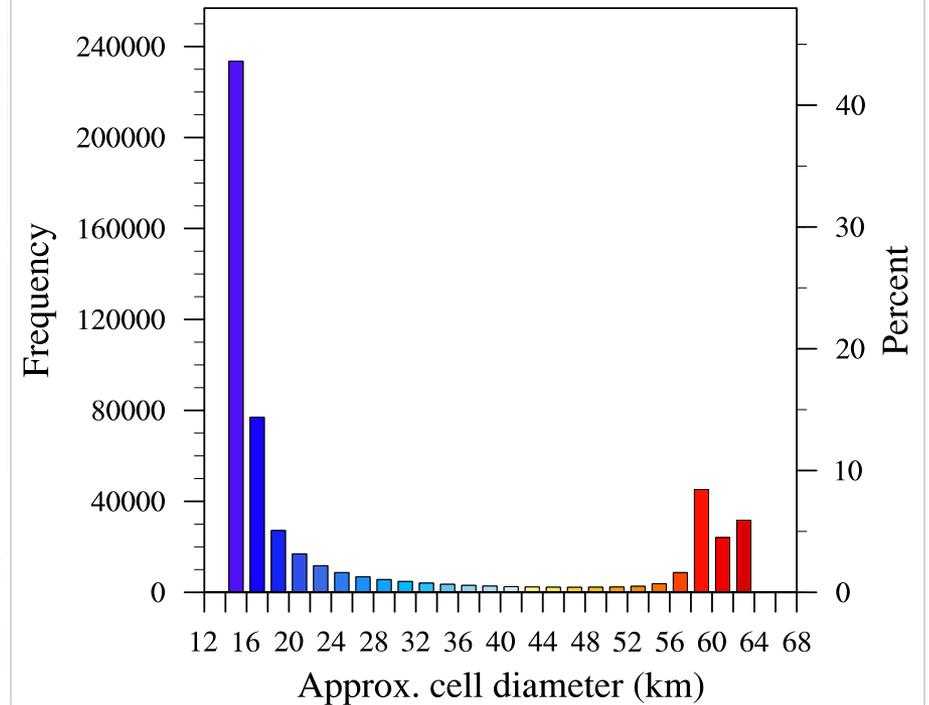
Research question: teleconnection between African monsoon systems and Atlantic/Indian ocean

MPAS variable-resolution mesh with 535554 cells

Approximate mesh resolution (km)



60-12km mesh cell distribution

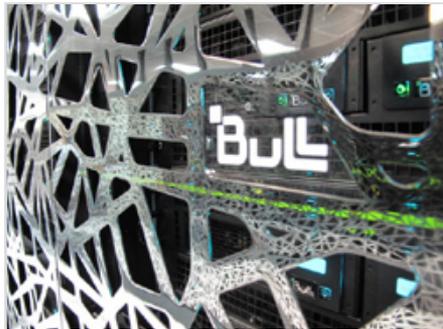


Research question: teleconnection between African monsoon systems and Atlantic/Indian ocean

Variable-resolution mesh with 535554 grid cells

Tested systems

TGCC Curie



FZJ Juropatest



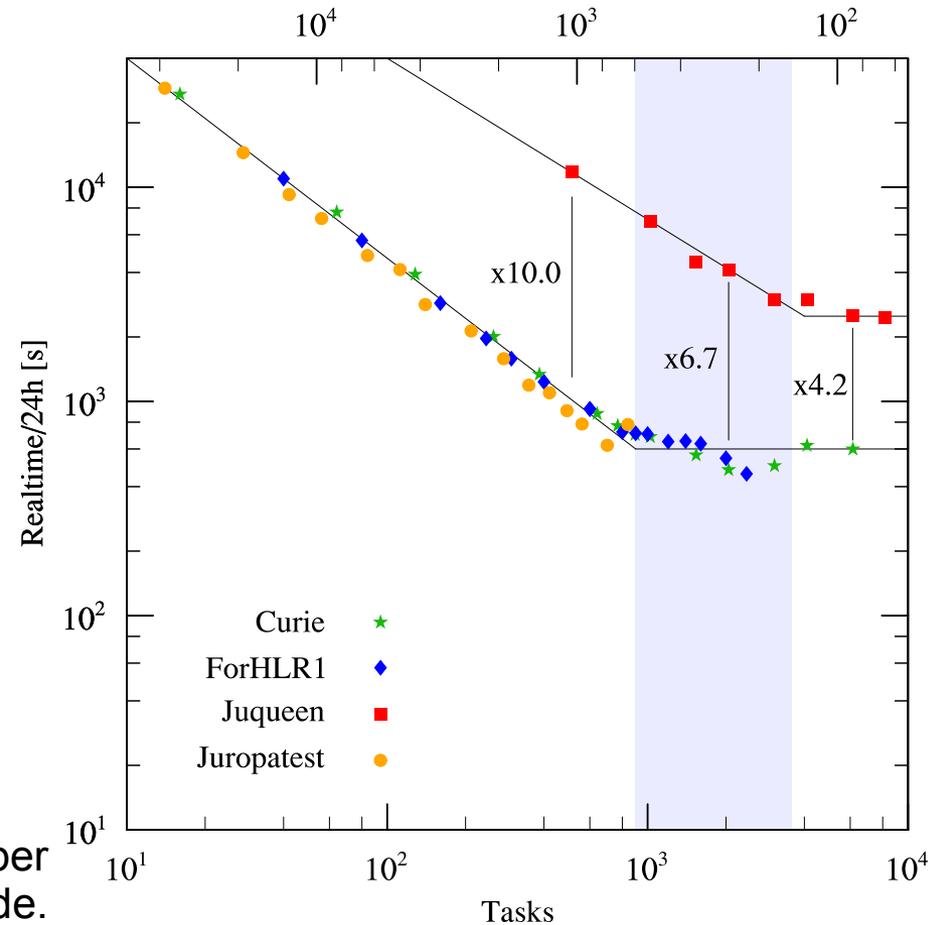
FZJ Juqueen



SCC ForHLR1

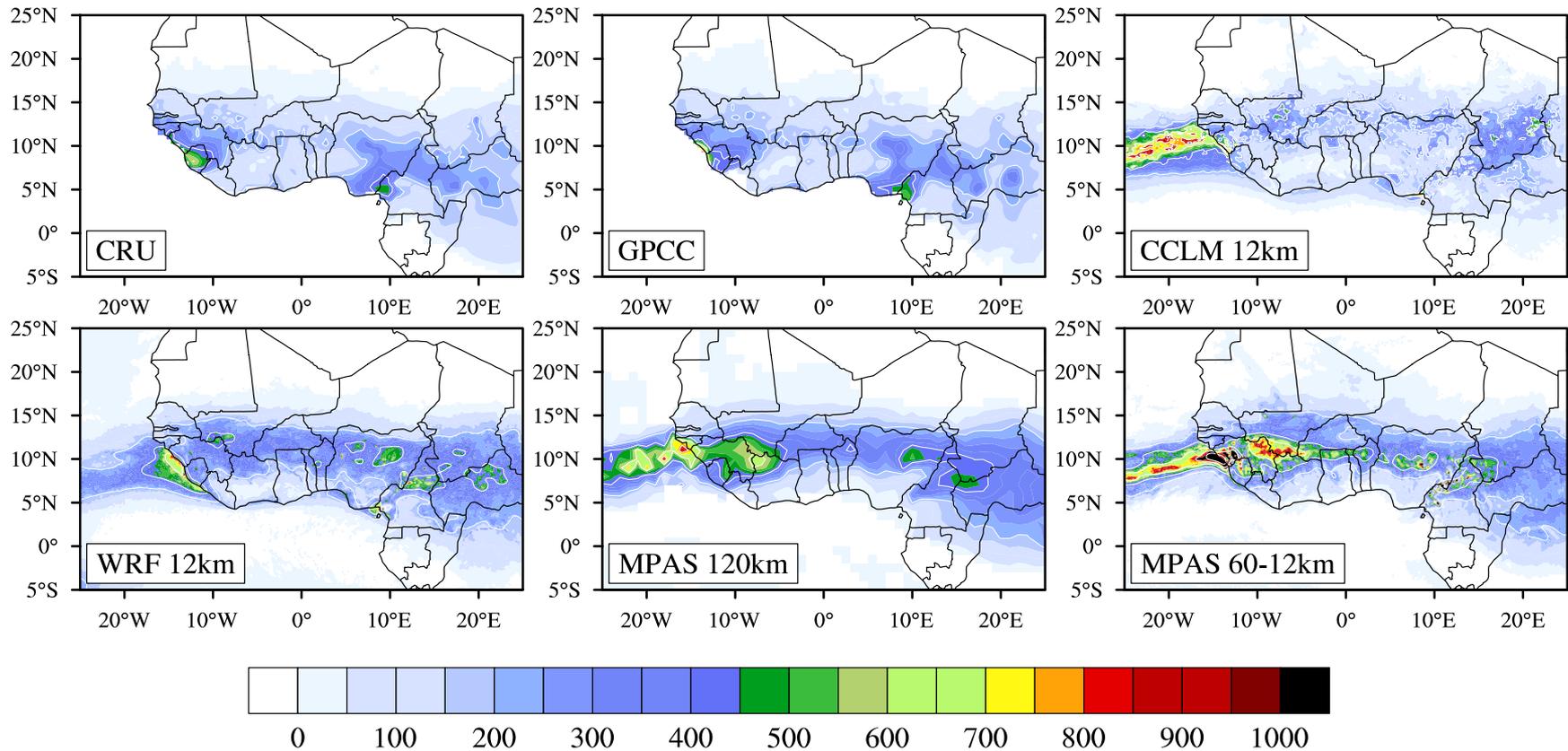
IBM Bluegene /Q: 28 racks, 1024 nodes per rack, 16 cores and 16GB memory per node.

Cells owned per task



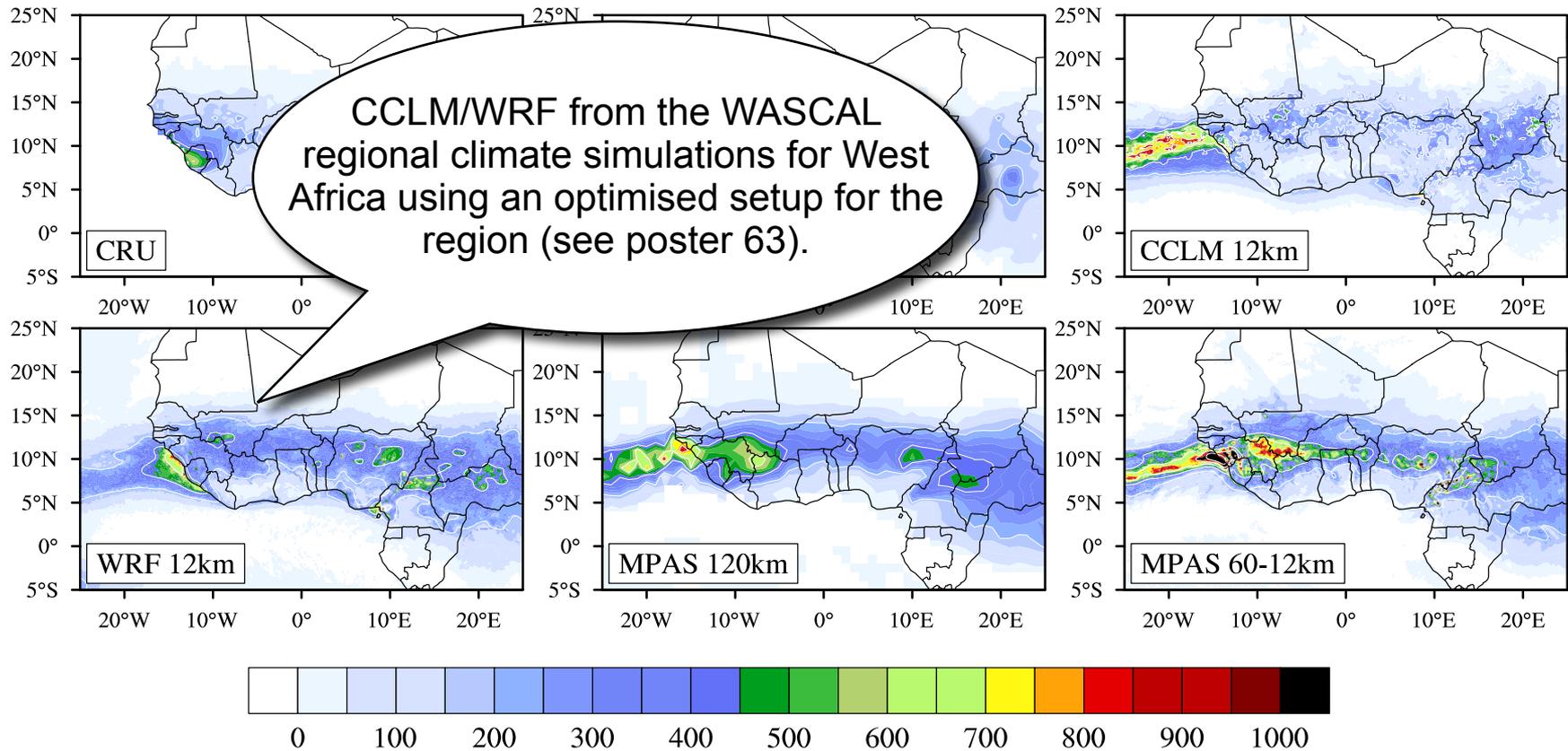
West African Monsoon in MPAS, WRF, CCLM

Total monthly precipitation [mm] - July 1982
 (initialisation: CCLM/WRF 1979-01-01, MPAS 1981-09-01)



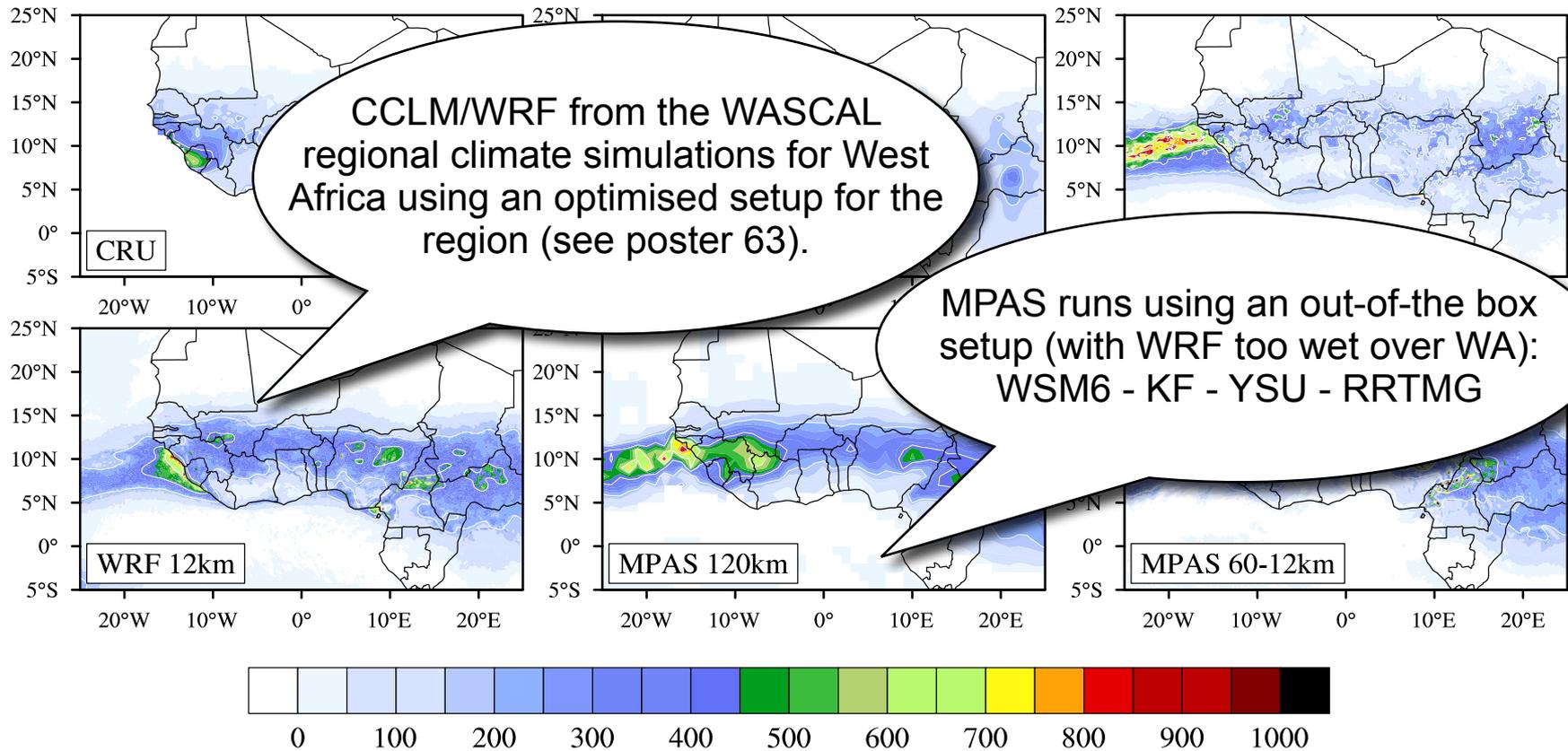
West African Monsoon in MPAS, WRF, CCLM

Total monthly precipitation [mm] - July 1982
 (initialisation: CCLM/WRF 1979-01-01, MPAS 1981-09-01)



West African Monsoon in MPAS, WRF, CCLM

Total monthly precipitation [mm] - July 1982
 (initialisation: CCLM/WRF 1979-01-01, MPAS 1981-09-01)



MPAS Extreme Scaling Experiment as part of the 3rd Juqueen Porting & Tuning Workshop

Extreme scaling on Jukeen: Pandora's box

regular 3km mesh x1.65536002
ncell = 65536002, nvert = 41

expect $\geq 70\%$ parallel efficiency
up to 400000 threads (150 cells per task)

double precision, physics+dynamics:
memory requirement 0.175MB per cell

11.5TB of memory required for 3km mesh,
minimum number of nodes is approx. 750

I/O sizes: initial condition 1.1TB, restart 2.1TB, diagnostics 15GB, history 250GB



Extreme scaling on Juqueen: Pandora's box

regular 3km mesh x1.65536002
ncell = 65536002, nvert = 41

expect $\geq 70\%$ parallel efficiency
up to 400000 threads (150 cells per task)

double precision, physics+dynamics:
memory requirement 0.175MB per cell

11.5TB of memory required for 3km mesh,
minimum number of nodes is approx. 750

I/O sizes: initial condition 1.1TB, restart 2.1TB, diagnostics 15GB, history 250GB



Pitfalls:

Test runs on 1 and 2 racks (1024, 2048 nodes)
failed - memory errors. Defragmentation issue?

Model initialisation took way to long (1hr+),
tuning of I/O and other parameters required

A moment of excitement (for geeks)

```

ssh
config_radt_clscheme = cld_incidence
config_radt_lw_scheme = rrtng_lw
config_radt_sw_scheme = rrtng_sw
config_sfclayer_scheme = monin_obukhov
DT_RADTLW = 1800.000000000
DT_RADTSW = 1800.000000000
DT_CU = 18.000000000
DT_PBL = 18.000000000

IMS = 1 IME = 141
JMS = 1 JME = 1
KMS = 1 KME = 42

IDS = 1 IDE = 141
JDS = 1 JDE = 1
KDS = 1 KDE = 42

ITS = 1 ITE = 141
JTS = 1 JTE = 1
KTS = 1 KTE = 41

--- config_frac_seaice = F
--- xice_threshold = 0.50000000000000000000
skipping over lutype = OLD
landuse type = USGS found 33 categories 2 seasons
--- isice = 24
--- iswater = 16
--- isurban = 1
Initialize NOAA LSM tables
landuse type = USGS found 27 categories
end read VEGPARAM.TBL
input soil texture classification = STAS
soil texture classification = STAS found 19 categories
end read SOILPARAM.TBL
end initialize NOAA LSM tables
min/max of meshScalingDel2 = 1.00000000000000000000 1.00000000000000000000
min/max of meshScalingDel4 = 1.00000000000000000000 1.00000000000000000000
        
```

llview: BLUE GENEO/JUELICH source: www.uid=~/^bgldiag/

Search Filter

Last Update **02/06/15 12:29:01** next in **01 s** Source [WWW](#) Help

Nodes **Waiting**

JUQUEEN

Machine: blue gene/juelich
 Memory: 440 GB, Ropas: 450TB,
 speed: 1.6 Gbit type = PPE A2,
 #Process = 20

used: 100% 486752/486752,
 free: 0.0 mds (0 mds)
 #procs (rank/wait): 3/3255

CPUS	userid	class	books	nodes	tasks	node	status	cpuh	wall	rend	
1	455752	slmet030	reserv	556	20672	20672	?	1	0.0	2:10	13:51

Job Scheduling Prediction

```

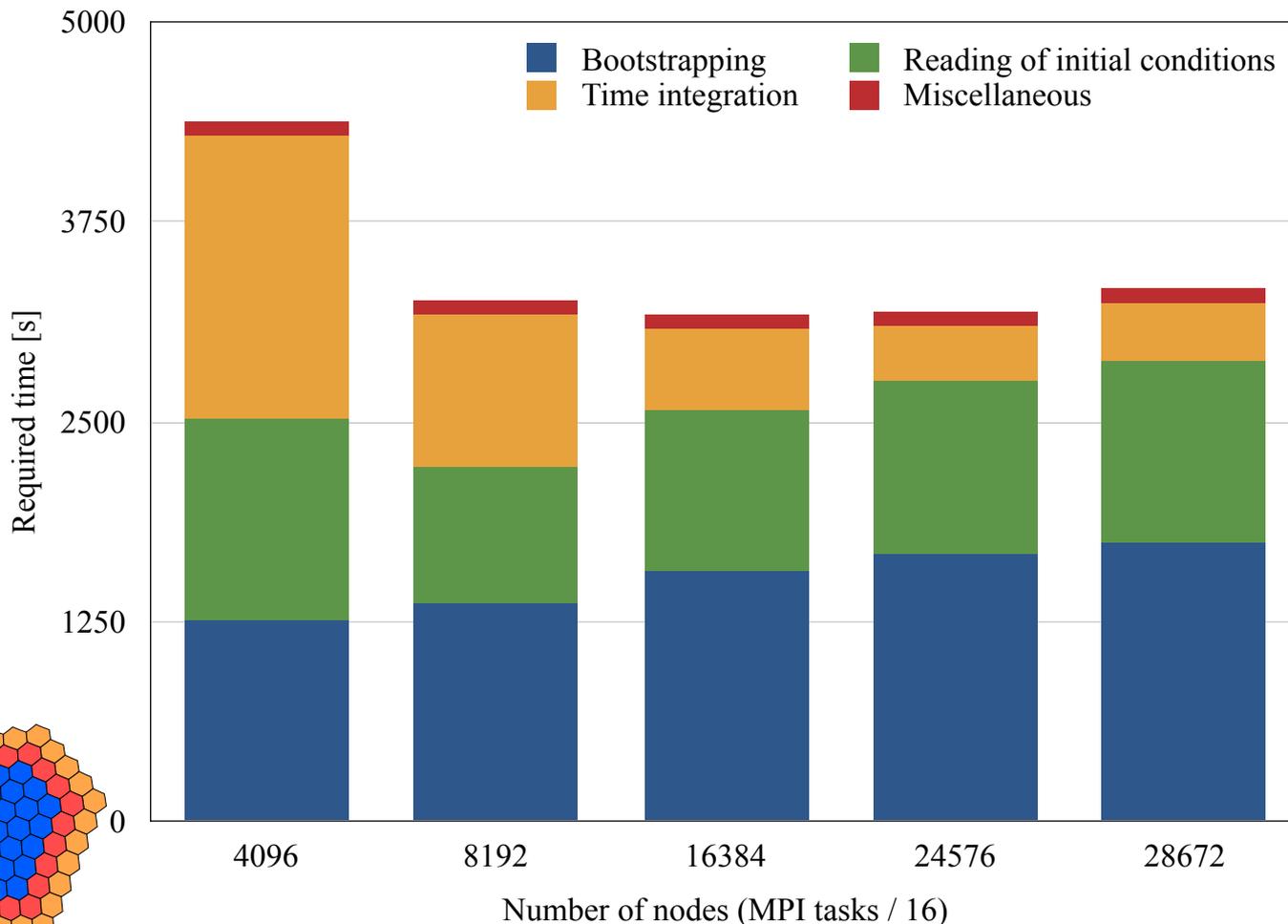
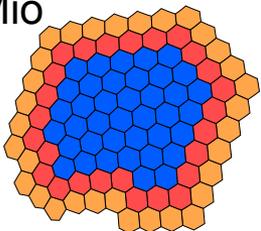
ssh
Every 5.0s: llq          Fri Feb 6 12:30:20 2015

Id          Owner      Submitted ST PRI Class      Running On
-----
juqueenc1.102036.0 hch05slmet030 2/6 10:19 R 50 reserv      juqueenc2c1
juqueenc1c1.220228.0 jhpc0hch053 /6 10:00:50 reserv
juqueenc3c1.102034.0 slmetjhc0902/6 6 10:19 I 000 reserv
juqueenc1c1.220231.0 hch053 2/6 10:21 I 50 reserv
juqueenc1c1.220234.0 hhv00hch053 2/6 11:25 I 50 reserv
juqueenc1c1.220235.0 hfg01hch053 2/6 11:26 I 50 reserv
juqueenc1c1.220236.0 hfg01jhc1301 2/6 11:29 I 50 reserv
juqueenc3c1.102039.0 slmetjhc0902/6 6 11:37 reserv
juqueenc4c1.108977.0 hfg01hv002 2/5 13:09 I 50 reserv
juqueenc4c1.109027.0 hfg010 1 2/6 01:38 I 50 reserv
juqueenc3c1.102035.0 hfg01hfg010 2/6 10:13 I 50 reserv
juqueenc2c1.216445.0 jzom1slmet030/3 116 10:190 m032 reserv
juqueenc3c1.101990.0 jikp0hfg010 /4 105 17:400 m032 48
juqueenc3c1.101991.0 jzom1hfg010 /5 075 17:400 m032 48
juqueenc3c1.101994.0 jhpc1hfg010 /5 145 17:420 m032 48
juqueenc3c1.101336.0 jhpc1jzom1101/5 143 11:010 m032
juqueenc4c1.108775.0 jhpc181kp0300/5 144 10:360 m032
juqueenc3c1.101844.0 jhpc1jzom1101/5 14:360:550 m032
juqueenc3c1.101946.0 rohedjhc1801 2/5 14:35 I 50 m032
juqueenc3c1.101947.0 hhv00hch053 2/5 14:36 I 50 m032
juqueenc3c1.101948.0 jhpc1jhc1801/6 115 14:360 reserm032
juqueenc3c1.101949.0 hch05jhc1801 2/5 14:36 I 50 m032
juqueenc1c1.220188.0 jhpc1rohed /6 115 15:300 reserm032
juqueenc3c1.101873.0 hfg00hhv002 /6 115 10:000 reserm032
juqueenc3c1.102040.0 hfg011 2/6 12:26 I 50 reserv
juqueenc1c1.219841.0 hhh150 2/3 20:06 I 50 m016
juqueenc3c1.101944.0 jhpc1801 2/5 14:35 I 50 m016
juqueenc3c1.101945.0 jhpc1801 2/5 14:35 I 50 m016
        
```

Extreme scaling results

Key facts

- 1hr integration, cold start (init.nc)
- 12s integration time step ($\Delta x \times 4$) to avoid NaNs
- no output except one diagnostic file (15Gb)
- 1 I/O task per 128 tasks, i.e., 128 per rack
- hash table size increased from 27183 to 6 Mio



Extreme scaling results

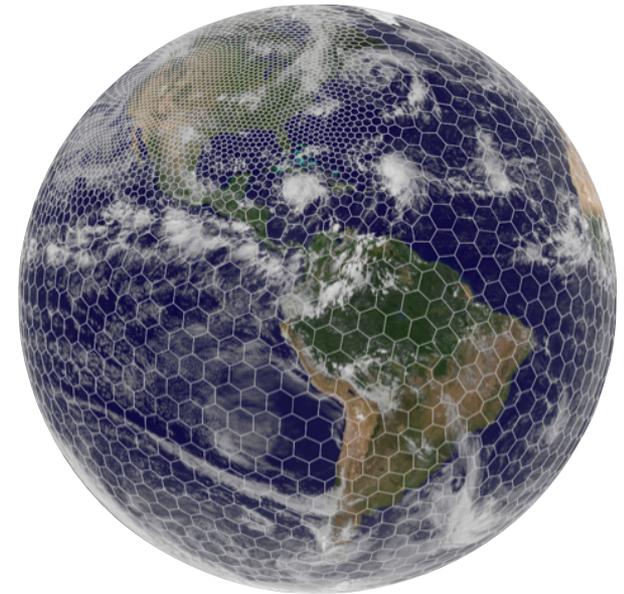
BG size (# nodes)	Parallel efficiency integration only	Integration in 24h walltime (no output)	Estimated integration in 24h walltime (with output)
4096	100.0%	48h	29h
8192	91.2%	87h	53h
16384	90.1%	172h	104h
24576	87.7%	250h	152h
28672	69.5%	231h	141h

BG size (# nodes)	Number of tasks	CPUh per 24h model run time (with output)	Speedup wrt. real time (with output to disk)
4096	65536	1.30 Mio	1.21
8192	131072	1.42 Mio	2.21
16384	262144	1.45 Mio	4.33
24576	393216	1.49 Mio	6.33
28672	458752	1.87 Mio	5.88

Take home messages

MPAS can reproduce the dynamics of the West African Summer Monsoon using an out-of-the box setup and is a promising tool for climate modelling.

Global, convection-resolving atmospheric simulations with MPAS are within reach of current/next generation HPC facilities. Open problems are model initialisation, disk I/O and post-processing of the data.

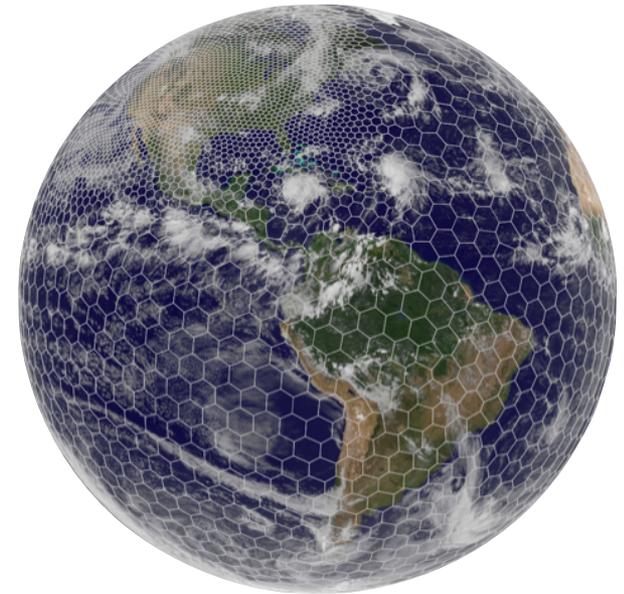


Take home messages

MPAS can reproduce the dynamics of the West African Summer Monsoon using an out-of-the box setup and is a promising tool for climate modelling.

Global, convection-resolving atmospheric simulations with MPAS are within reach of current/next generation HPC facilities. Open problems are model initialisation, disk I/O and post-processing of the data.

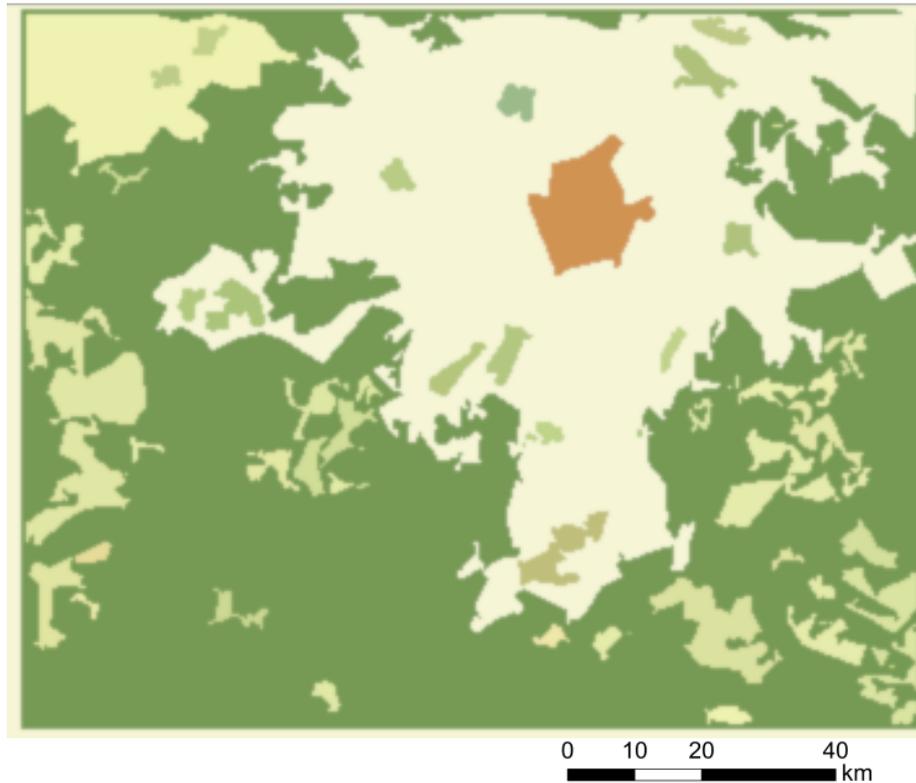
Model developers/users have to tailor and optimise models to run on modern massively parallel systems, which requires in-depth knowledge and time.



Bonus material

Limitations of limited area modelling

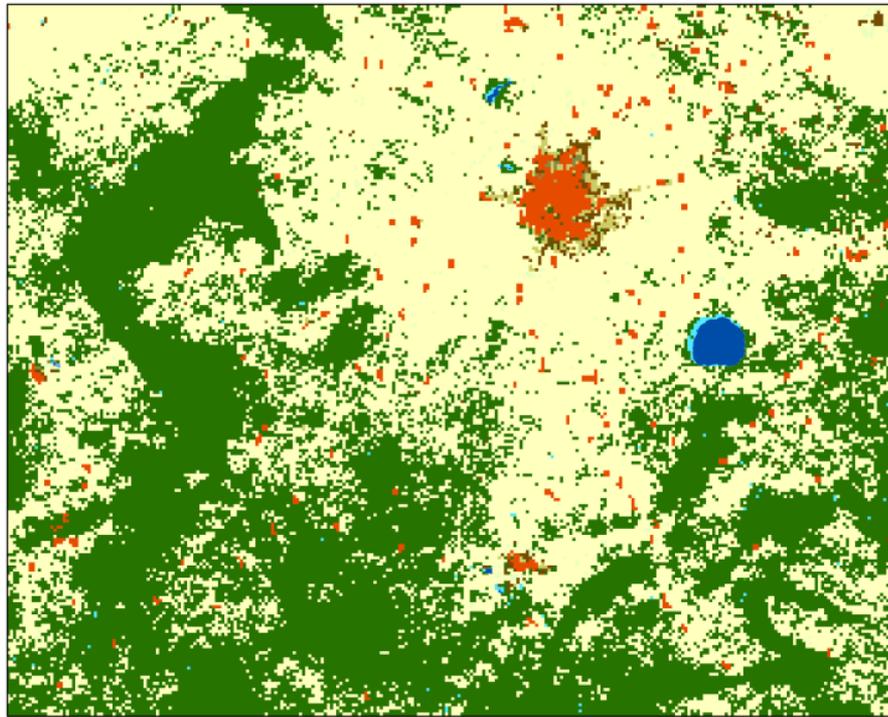
How to include new land surface and soil data or dynamic changes thereof in a consistent way?



GCM standard, Volta region (West Africa)

Limitations of limited area modelling

How to include new land surface and soil data or dynamic changes thereof in a consistent way?



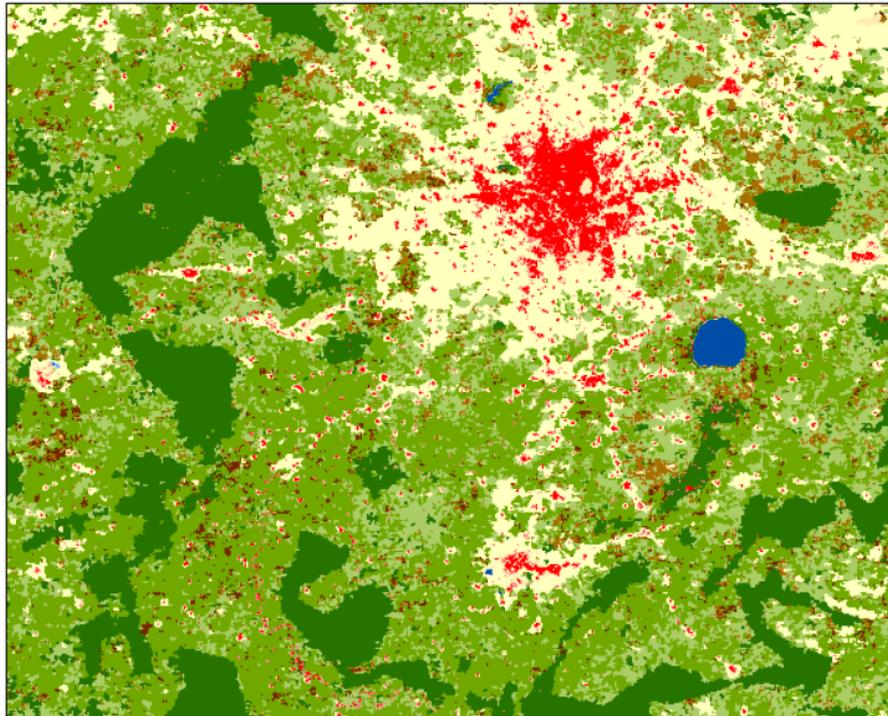
0 10 20 40 km

GCM standard, Volta region (West Africa)

RCM standard land cover (MODIS)

Limitations of limited area modelling

How to include new land surface and soil data or dynamic changes thereof in a consistent way?

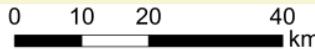


GCM standard, Volta region (West Africa)

RCM standard land cover (MODIS)

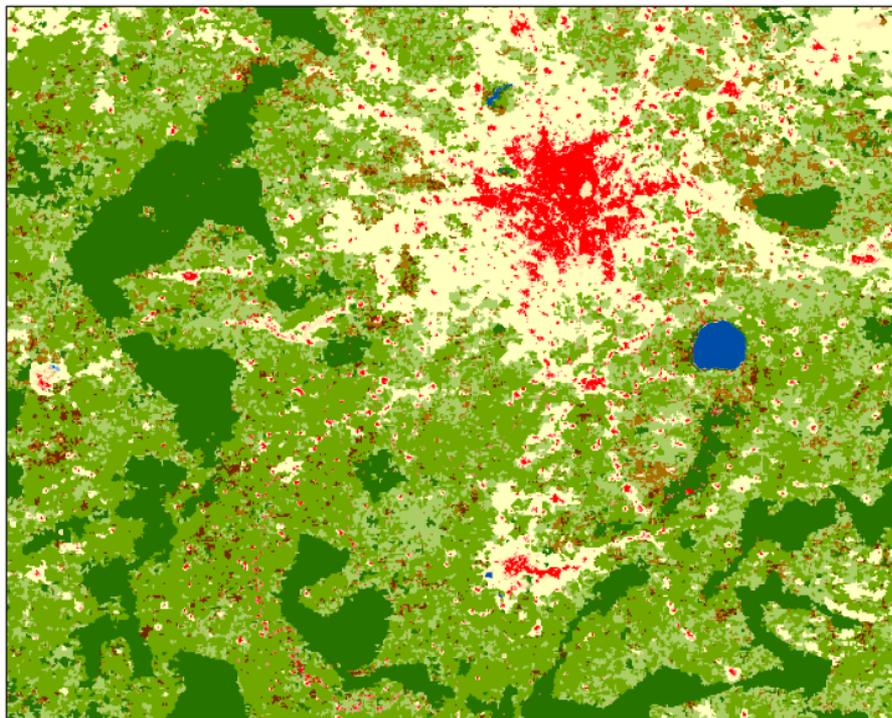
RCM high-resolution land cover
(DLR: MODIS, ASAR, TanDEM-X)

Credits: Ursula Gessner, DLR



Limitations of limited area modelling

How to include new land surface and soil data or dynamic changes thereof in a consistent way?



Credits: Ursula Gessner, DLR 

GCM standard, Volta region (West Africa)

RCM standard land cover (MODIS)

RCM high-resolution land cover
(DLR: MODIS, ASAR, TanDEM-X)

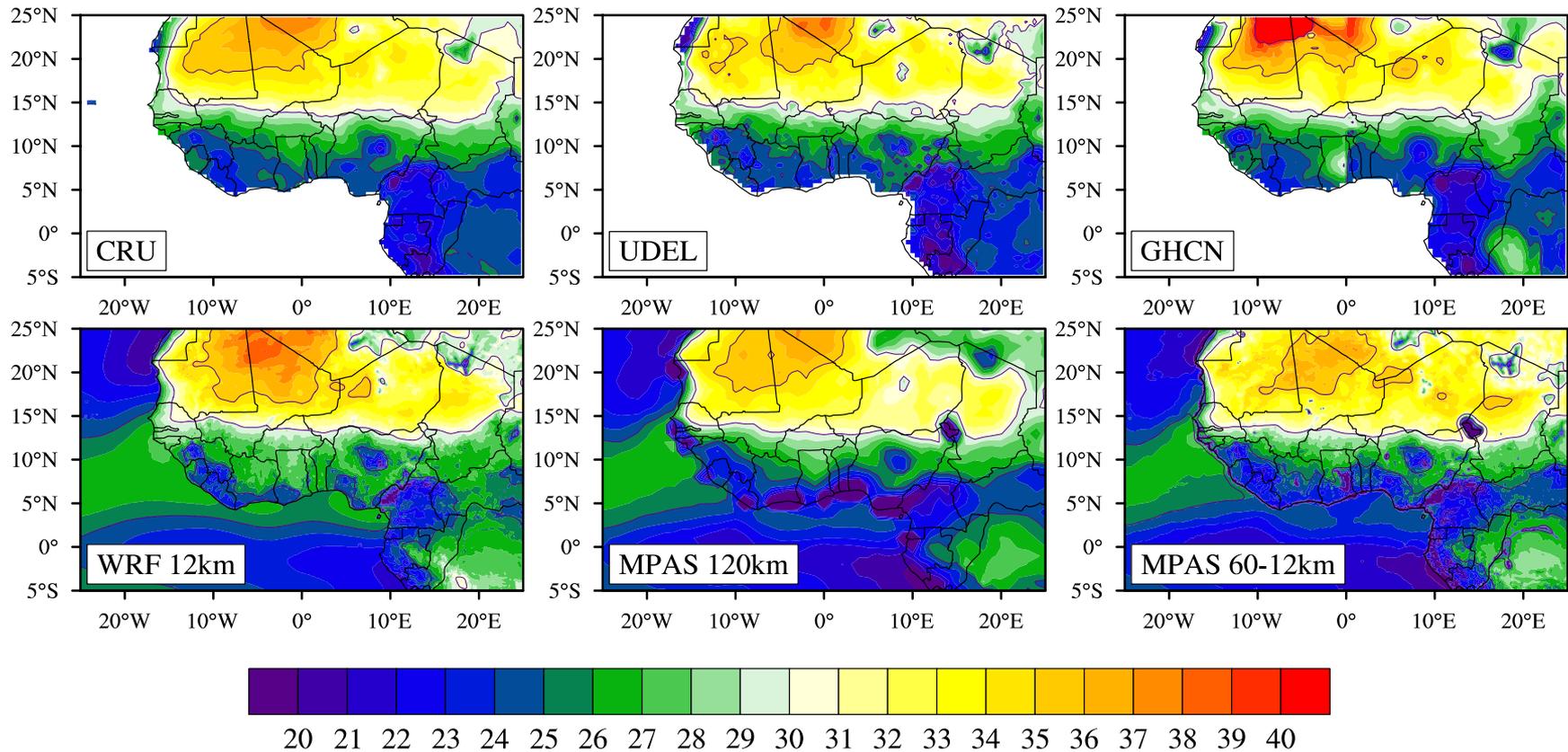
Consistency? Feedback processes?

Dynamic changes of land use?

CMIP5 vs CORDEX issue:
land use classification
and land use change in
ESMs not reflected in RCMs

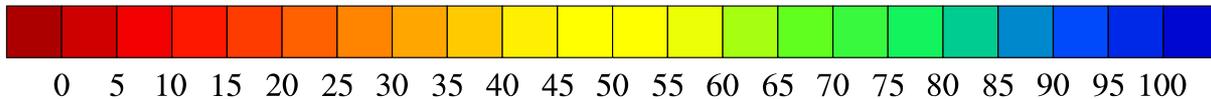
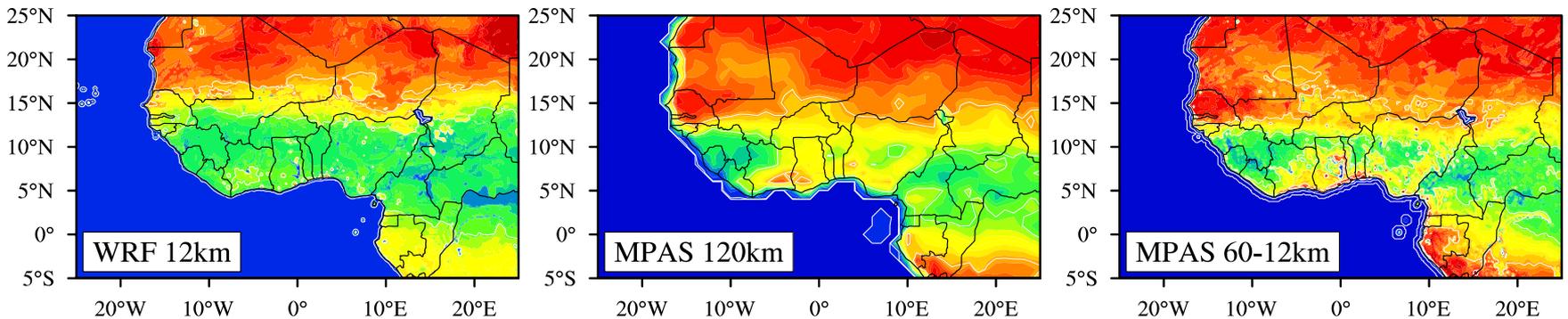
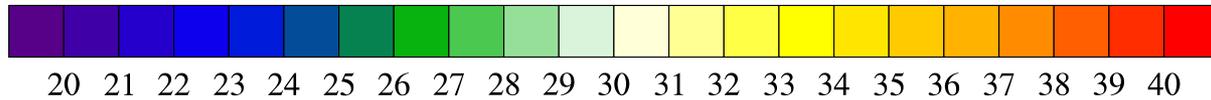
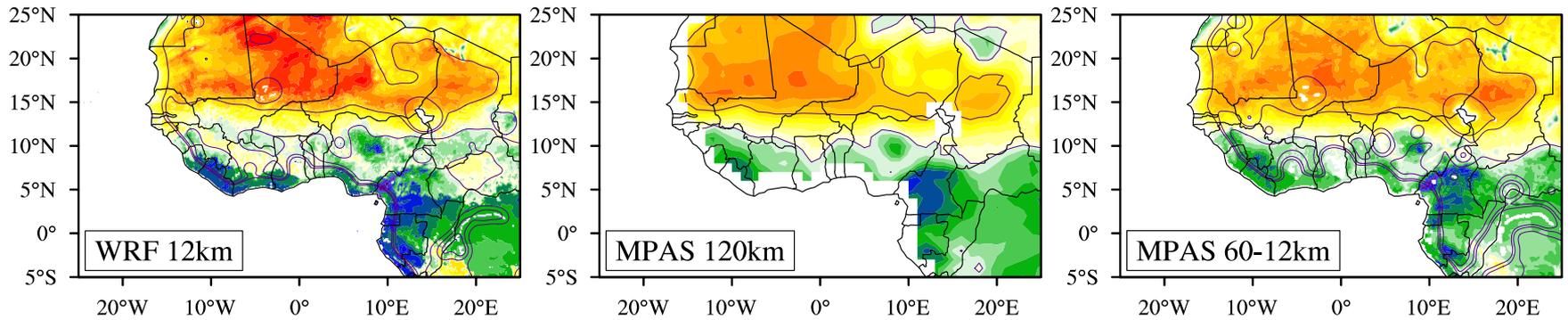
MPAS, WRF, CCLM vs. observations - any good?

Near surface temperature [$^{\circ}\text{C}$] - July 1982 monthly mean
(initialisation: WRF 1979-01-01, MPAS 1981-09-01)



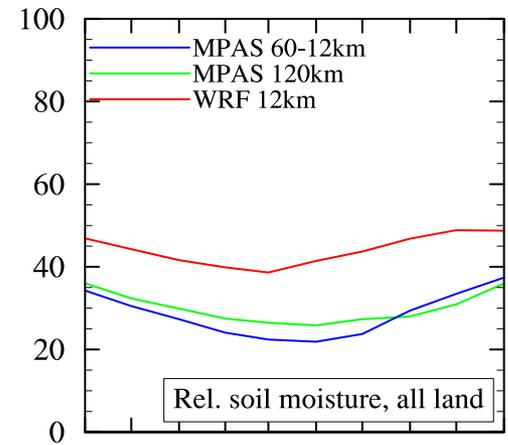
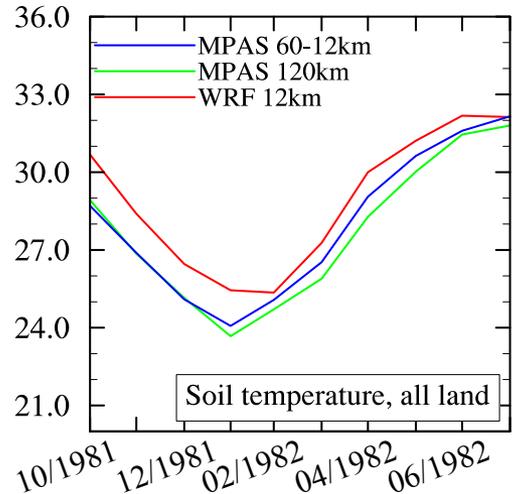
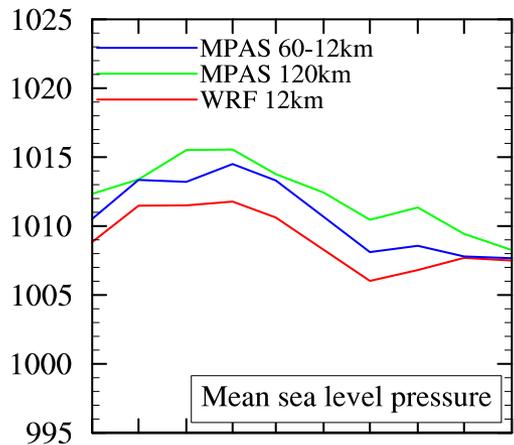
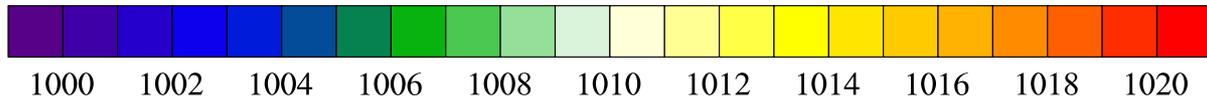
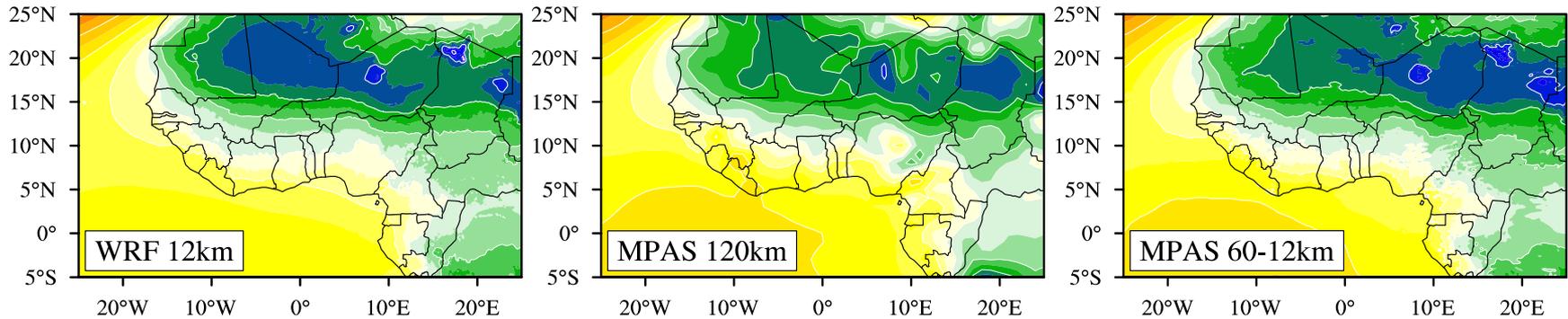
At least one monsoon period needed for spinup

Soil temperature [$^{\circ}\text{C}$] (top) and relative soil moisture [%] (bottom) - July 1982 mean



At least one monsoon period needed for spinup

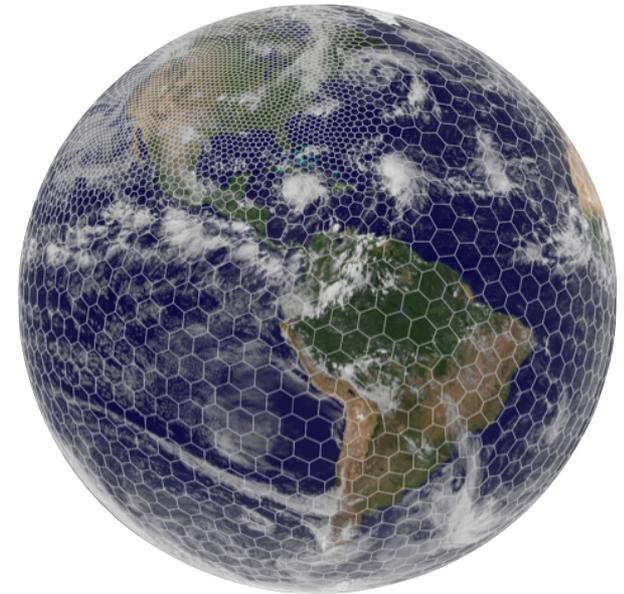
Mean sea level pressure [hPa] - July 1982 monthly mean



Take home messages

MPAS can reproduce the dynamics of the West African Summer Monsoon using an out-of-the box setup and is a promising tool for climate modelling.

Global, convection-resolving atmospheric simulations with MPAS are within reach of current/next generation HPC facilities. Open problems are model initialisation, disk I/O and post-processing of the data.



Take home messages

MPAS can reproduce the dynamics of the West African Summer Monsoon using an out-of-the box setup and is a promising tool for climate modelling.

Global, convection-resolving atmospheric simulations with MPAS are within reach of current/next generation HPC facilities. Open problems are model initialisation, disk I/O and post-processing of the data.

Model developers/users have to tailor and optimise models to run on modern massively parallel systems, which requires in-depth knowledge and time.

