Single Precision IFS

Filip Váňa

P. Düben, S. Lang, G. Carver, M. Leutbecher, D. Salmond, N. Wedi, T. Wilhelmsson, T. Palmer,

S. Saarinen, R. Hogan, P. Marguinaud, P. Towers, M. Hamrud, K. Mogensen, P. Janssen,...

filip.vana@ecmwf.int

ECMWF

Single precision at ECMWF

- Single precision?
  - More optimal way of using computing resources - higher precision used only where unavoidable.
  - Target is to achieve better computing performance through memory access reduction without compromising the model skill.
  - Testing tool (potential to identify precision sensitive or bad codes).

Single precision at ECMWF

- Single precision?
  - More optimal way of using computing resources - higher precision used only where unavoidable.
  - Target is to achieve better computing performance through memory access reduction without compromising the model skill.
  - Testing tool (potential to identify precision sensitive or bad codes).
- Pioneered by Oxford university group with OpenIFS (since 2013).

Single precision at ECMWF

- Single precision?
  - More optimal way of using computing resources - higher precision used only where unavoidable.
  - Target is to achieve better computing performance through memory access reduction without compromising the model skill.
  - Testing tool (potential to identify precision sensitive or bad codes).
- Pioneered by Oxford university group with OpenIFS (since 2013).
- Handed over to ECMWF (2014)
  - Early EPS tests with single precision OpenIFS under prepIFS.
  - Implemented to full IFS CY41R2 allowing both precisions with the unique code (Wam, Nemo, ...).
  - More code protection & additional scientific issues (VFE, SW radiation, PBL height & mass flux, surface scheme), technical development (Dr.Hook) $\Rightarrow$ safe performance up to T_L399 guaranteed.
  - Explored in EPS (uncoupled T_L399) - additional technical complexity (mixed DP & SP jobs), more fixes (stochastic physics).
  - Recognized as serious activity for an operational implementation & new staff position (2016)
  - Seamless compilation of SP libraries (CY43R1).
  - *SP option available under prepIFS for FC & EPS experiments (CY45R1).*

Single precision crash course

- Compiling with new CPP directive "SINGLE" and no automatic real promotions like: *-fdefault-real-8 ...*
  Note: the CPP directive only affects *xrd/module/parkind1.F90* where it redefines the **JPRB** parameter.
- Additional kind created **JPRD** for those parameters which should stay always as **REAL=8**:
  system calls (USER_CLOCK), binary files handling, calls to exclusively DP libraries...
- **Once the executable is compiled the precision is fully determined.**

Compliant coding with single precision

- Avoid hard coded thresholds or constants:

  ZEPS = 1.E-14_JPRB, **IF (ZLBDA(JROF,JLEV) $<$ 1.0E120_JPRB) THEN**

  GCD4=.3608E-69_JPRB.

Compliant coding with single precision

- Avoid hard coded thresholds or constants:
  ZEPS = 1.E-14_JPRB, **IF (ZLBDA(JROF,JLEV) $<$ 1.0E120_JPRB) THEN**
  GCD4=.3608E-69_JPRB.
- Avoid hard coded **real*8** types (observation operators, I/O)

Compliant coding with single precision

- Avoid hard coded thresholds or constants:
  ZEPS = 1.E-14_JPRB, **IF (ZLBDA(JROF,JLEV)** $<$ **1.0E120_JPRB) THEN**
  GCD4=.3608E-69_JPRB.
- Avoid hard coded **real*8** types (observation operators, I/O)
- Avoid binary format whenever possible (Wam passes the same bathymetry field to Nemo (DP) and IFS (SP))

Compliant coding with single precision

- Avoid hard coded thresholds or constants:
  ZEPS = 1.E-14_JPRB, **IF (ZLBDA(JROF,JLEV) $<$ 1.0E120_JPRB) THEN**
  GCD4=.3608E-69_JPRB.
- Avoid hard coded **real*8** types (observation operators, I/O)
- Avoid binary format whenever possible (Wam passes the same bathymetry field to Nemo (DP) and IFS (SP))
- Avoid intrinsic function for specific type:
  DLOG $\rightarrow$ **LOG, DCOS** $\rightarrow$ **COS, DATAN** $\rightarrow$ **ATAN, ...**

Compliant coding with single precision

- Avoid hard coded thresholds or constants:
  ZEPS = 1.E-14_JPRB, **IF (ZLBDA(JROF,JLEV)** $<$ **1.0E120_JPRB) THEN**
  GCD4=.3608E-69_JPRB.
- Avoid hard coded **real*8** types (observation operators, I/O)
- Avoid binary format whenever possible (Wam passes the same bathymetry field to Nemo (DP) and IFS (SP))
- Avoid intrinsic function for specific type:
  DLOG $\rightarrow$ **LOG, DCOS** $\rightarrow$ **COS, DATAN** $\rightarrow$ **ATAN, ...**
- Stay with your computation close to 1.
  arp/setup/sucst.F90 & surf/module/suscst_mod.F90:
  RSIGMA=2.0_JPRB * RPI5 * RKBOL**4 /(15._JPRB* RCLUM**2 * RHPLA**3)**
  replaced by:
  RSIGMA=2.0_JPRB * RPI5 * (RKBOL/RHPLA)**3 * RKBOL/(15._JPRB* RCLUM**2)**

New code to accommodate single precision

- Interface routines to MPI (**MPL_SEND_REAL4_SCALAR,...**)
- Reintroduced interface to both single and double precision Blas/Lapack routines

```
IF (JPRB == JPRD) THEN
  CALL DGEMM(...
ELSE
  CALL SGEMM(...
ENDIF
```

- Interface to binary profiles in radiation
- Interface to ODB and Nemo (so far exclusively double precision)

All tests based on **JPRB==JPRD** check.

Scientific issues

- Enhanced security for variables outside ($\pm 2^{-126} \approx 1.1755E-38$, $\pm 2^{127} \approx 1.7014E+38$):

  ZEPSILON = 100._JPRB * EPSILON(ZEPSILON)

  A/B $\rightarrow$ A/SIGN(MAX(ABS(B),ZEPSILON),B)

  Always better to find alternatives without any "artificial" security!

Scientific issues

- Enhanced security for variables outside ($\pm 2^{-126} \approx 1.1755E-38$, $\pm 2^{127} \approx 1.7014E+38$):

  ZEPSILON = 100._JPRB * EPSILON(ZEPSILON)

  A/B $\rightarrow$ A/SIGN(MAX(ABS(B),ZEPSILON),B)

  Always better to find alternatives without any "artificial" security!

- Badly conditioned code: shallow convection, SW radiation up-scaling, surface scheme, FLT, ...

Scientific issues

- Enhanced security for variables outside ($\pm 2^{-126} \approx 1.1755E-38$, $\pm 2^{127} \approx 1.7014E+38$):

  ZEPSILON = 100._JPRB * EPSILON(ZEPSILON)

  A/B $\rightarrow$ A/SIGN(MAX(ABS(B),ZEPSILON),B)

  Always better to find alternatives without any "artificial" security!

- Badly conditioned code: shallow convection, SW radiation up-scaling, surface scheme, FLT, ...

- Galerkin methods: Legendre transforms and VFE integral operator setup

Scientific issues

- Enhanced security for variables outside ($\pm 2^{-126} \approx 1.1755E-38$, $\pm 2^{127} \approx 1.7014E+38$):

  ZEPSILON = 100._JPRB * EPSILON(ZEPSILON)

  A/B $\rightarrow$ A/SIGN(MAX(ABS(B),ZEPSILON),B)

  Always better to find alternatives without any "artificial" security!

- Badly conditioned code: shallow convection, SW radiation up-scaling, surface scheme, FLT, ...
- Galerkin methods: Legendre transforms and VFE integral operator setup
- Testing tool for TL code.

Scientific issues

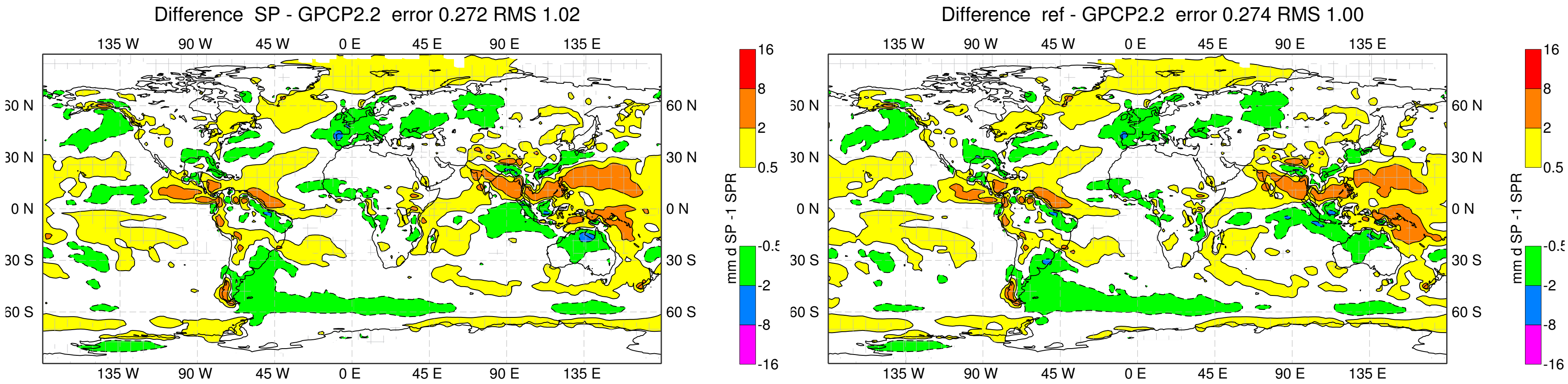
- Enhanced security for variables outside ($\pm 2^{-126} \approx 1.1755E-38$, $\pm 2^{127} \approx 1.7014E+38$):

  ZEPSILON = 100._JPRB * EPSILON(ZEPSILON)

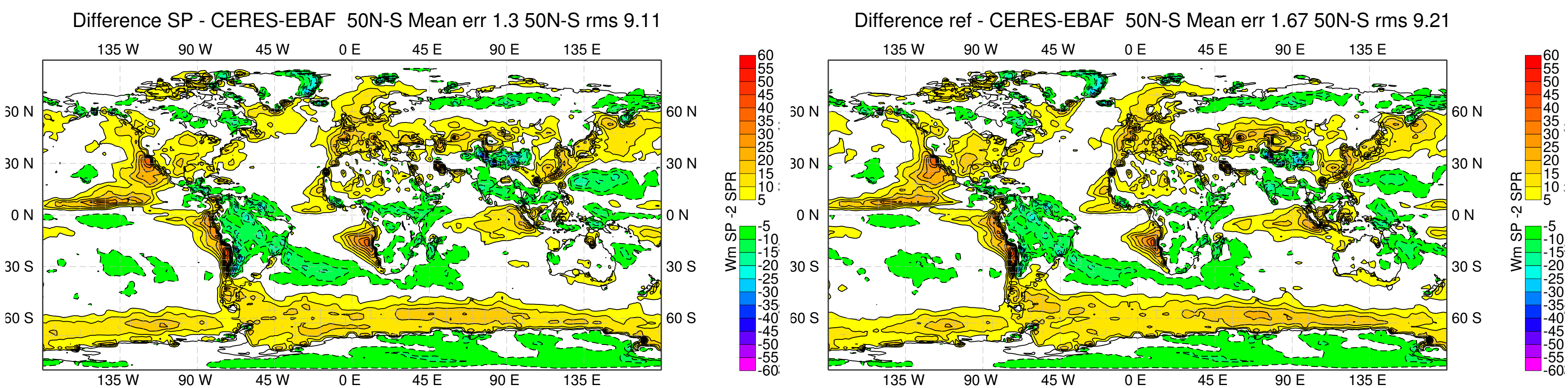
  A/B $\rightarrow$ A/SIGN(MAX(ABS(B),ZEPSILON),B)

  Always better to find alternatives without any "artificial" security!
- Badly conditioned code: shallow convection, SW radiation up-scaling, surface scheme, FLT, ...
- Galerkin methods: Legendre transforms and VFE integral operator setup
- Testing tool for TL code.
- Ongoing process: identify potential and precision sensitive problems (massive data averaging, matrix inversion, quadratic equations,...) $\Rightarrow$ ideally to be rewritten in a more robust way.

Climate runs - I.

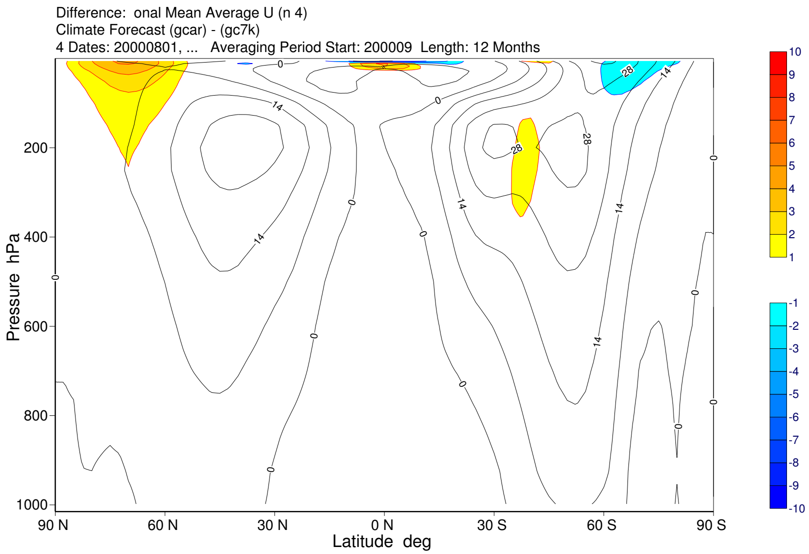


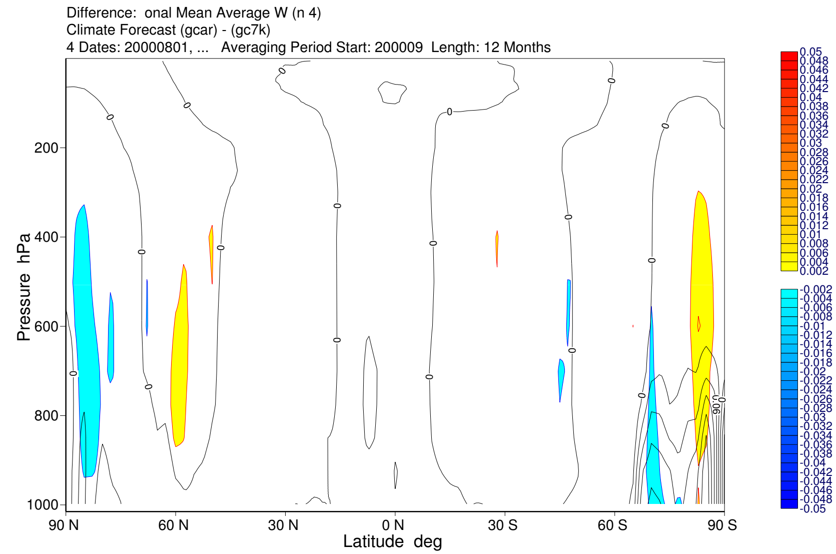
precipitation

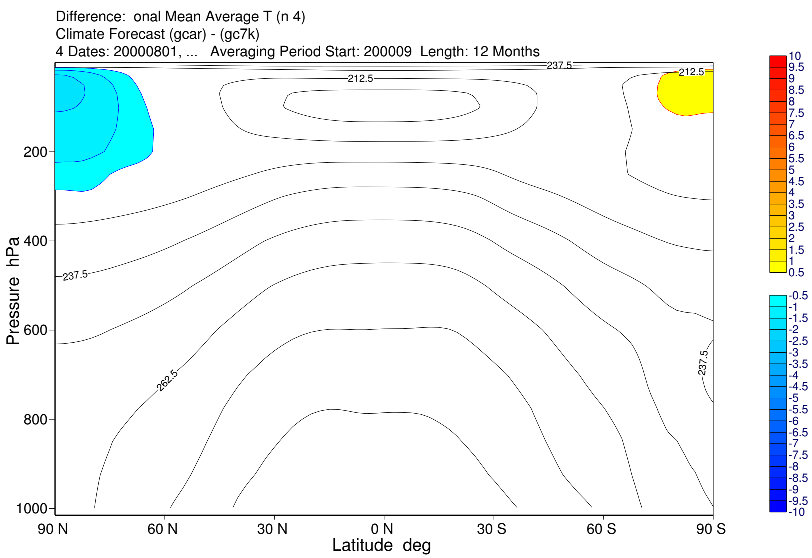


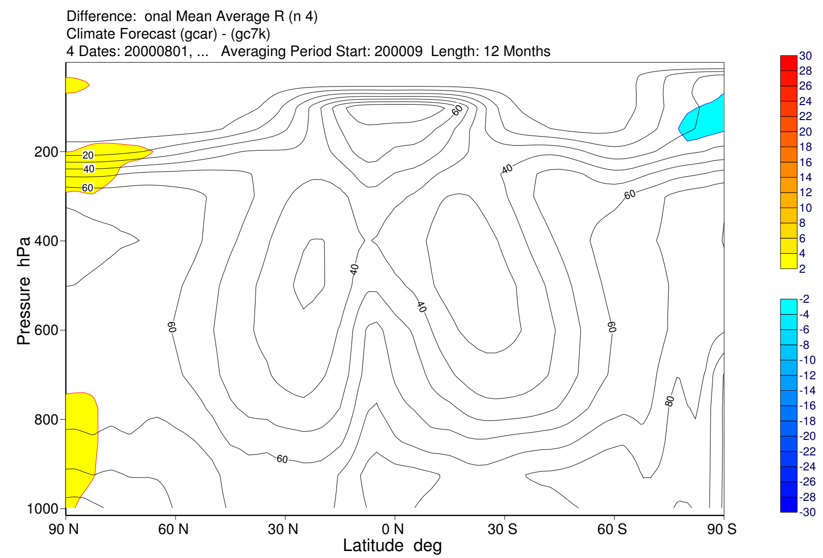
TOA sw

Climate runs - II.

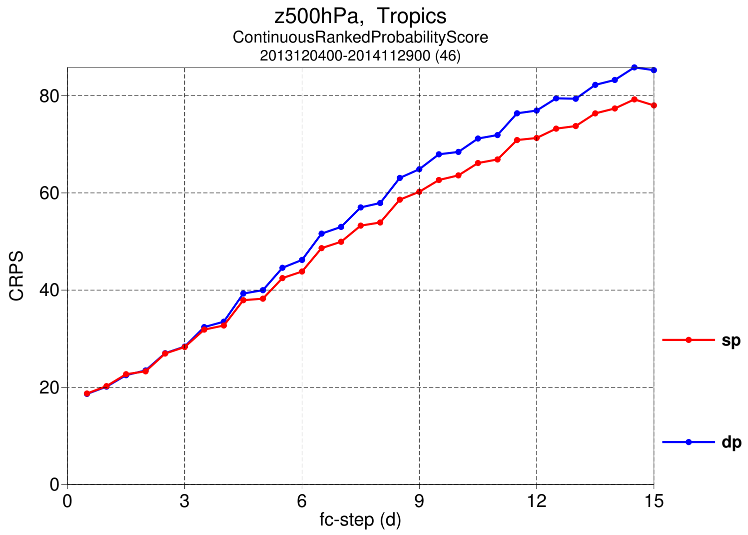


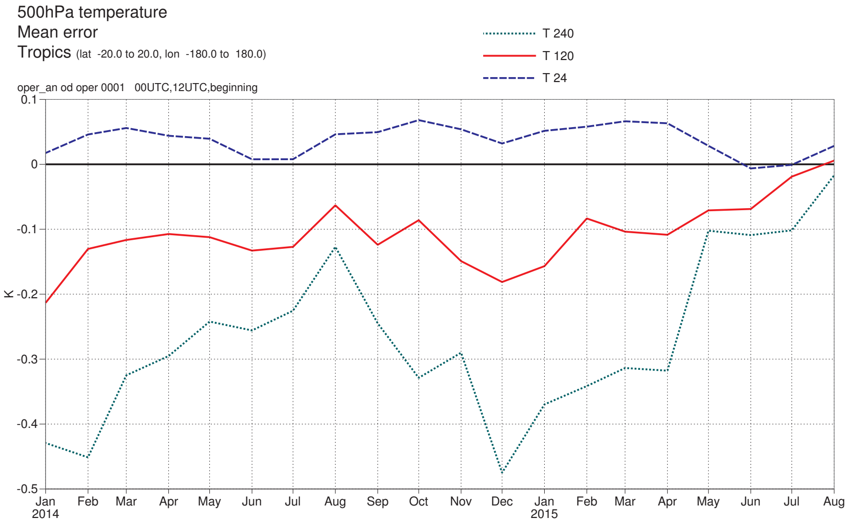




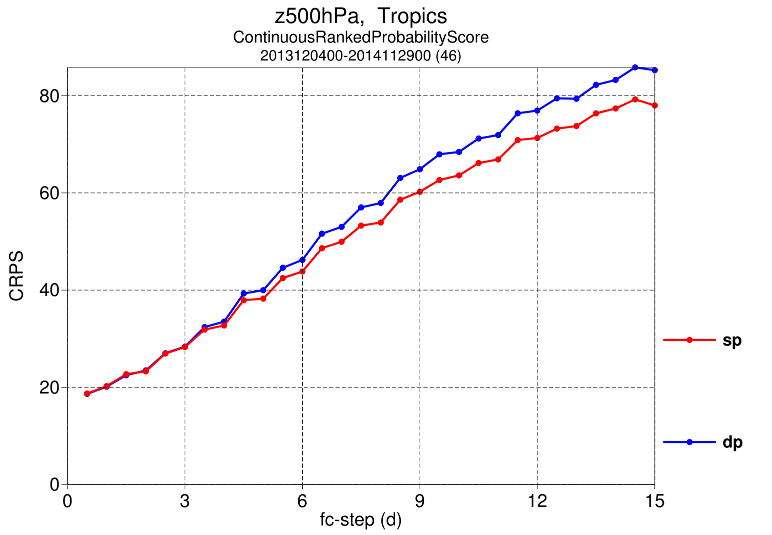


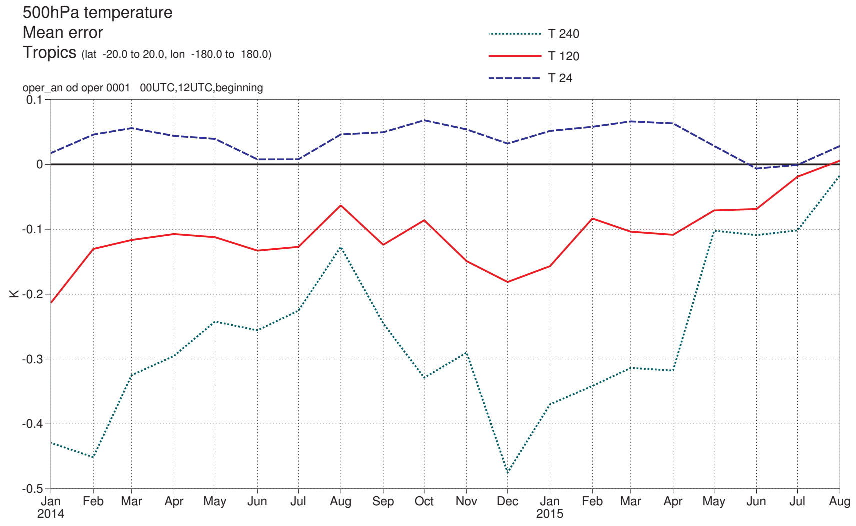
Uncoupled T399 EPS system

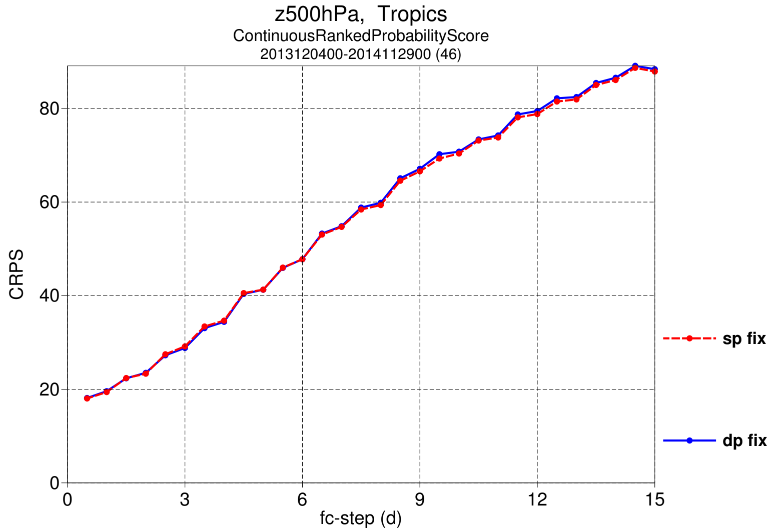


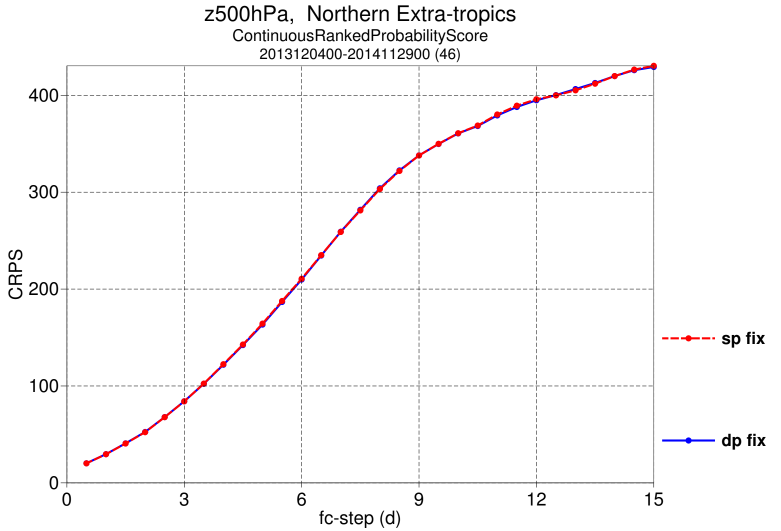


Uncoupled T399 EPS system



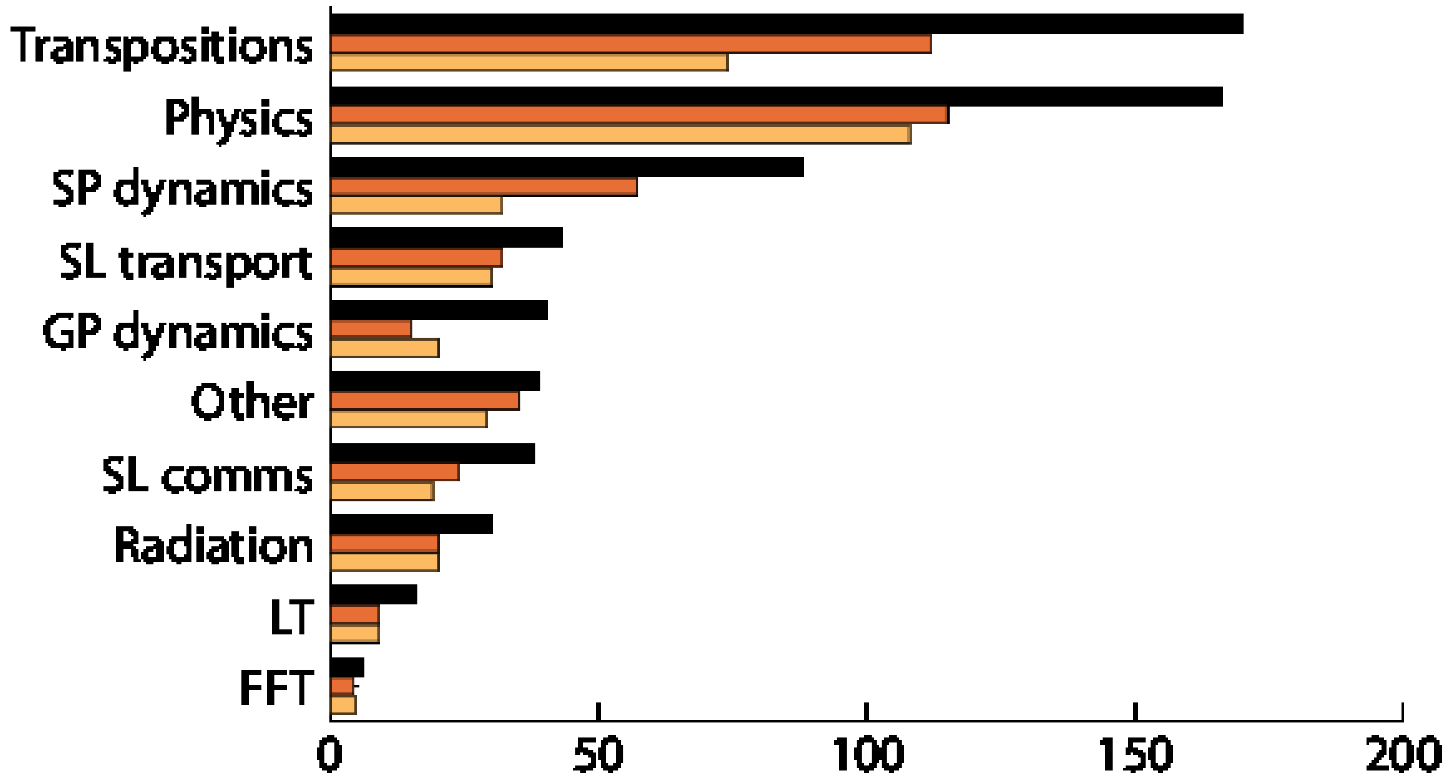






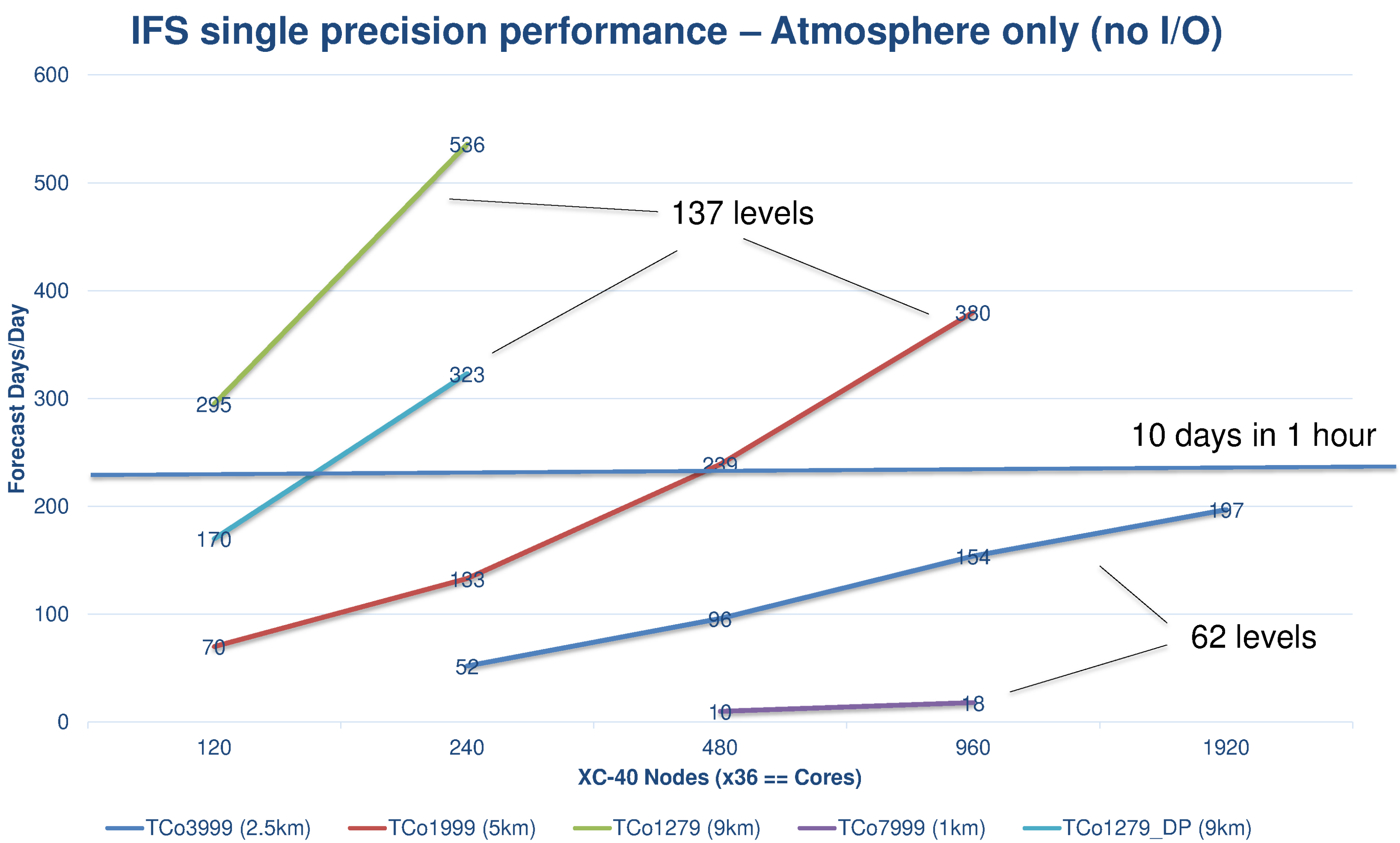
Computational cost - EPS

Single precision runs 37% faster (per standard timestep).
Optimizing cache utilization through NPROMA the gain becomes over 40% w.r.t. the double precision reference.



Cray XC30, 96 MPI x 8 OpenMP

Computational cost - atmospheric model



Single precision plans at ECMWF

- From CY45R1 single precision becomes available for RD testing.
- Extensive validation required. (Offline surface scheme already by default in SP!)
- Conservativeness issue.
- Review of TL code.
- Plan to make it default option for all FC & EPS jobs in RD.
- SP for data assimilation (i.e. mixed SP and DP code).
- Single precision Nemo.
- Investigate use of reduced precision.
- Reduced accuracy vs. more sophistication.