Environment Canada Environnement Canada

Canada

Running operational Canadian NWP models and assimilation systems on next-generation supercomputers

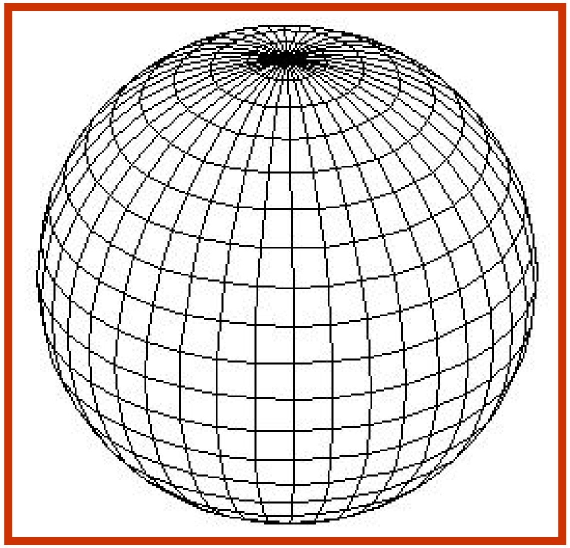
Mark Buehner, Michel Desgagné, Abdessamad Qaddouri, Janusz Pudykiewicz, Michel Valin, Martin Charron

Meteorological Research Division

Scalability Workshop
ECMWF, Reading, UK
April 14-15, 2014

The GEM model

1. Grid point lat/lon model
2. Finite differences on an Arakawa-C grid
3. Semi-Lagrangian (poles are an issue)
4. Implicit time discretization
    1. Direct solver (Nk 2D horizontal elliptic problems)
    2. Full 3D iterative solver based on FGMRES
5. Global uniform, Yin-Yang and LAM configurations
6. Hybrid MPI/OpenMP
    1. Halo exchanges
    2. Array transposes for elliptic problems
7. PE block partitioning for I/O

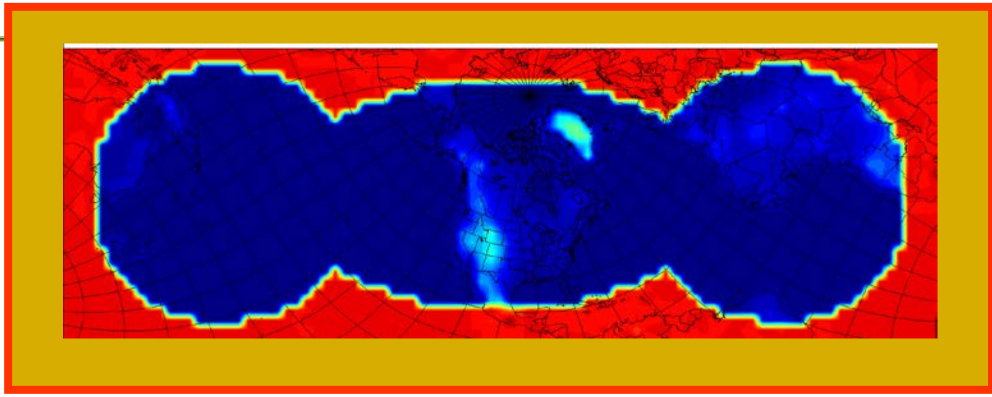


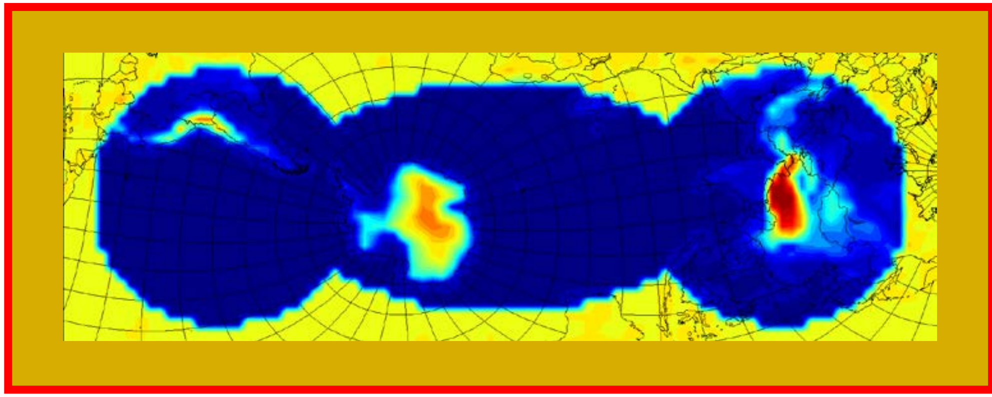
Global uniform grid:

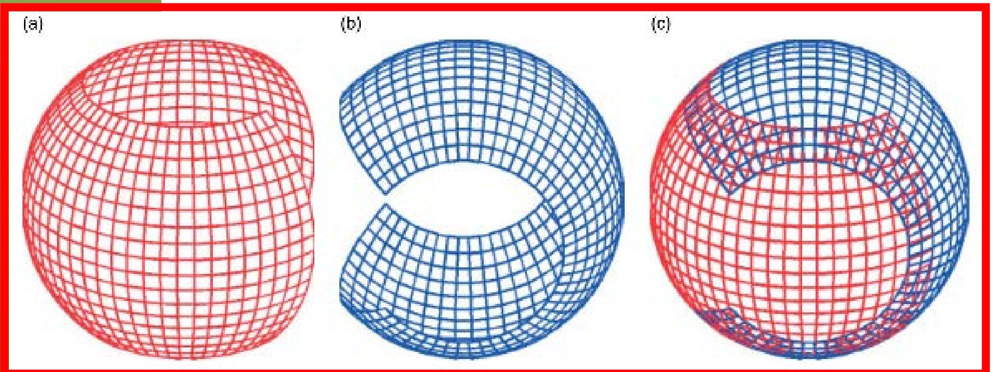
1) a challenge for DM implementation
2) many more elliptic problems to solve due to implicit horizontal diffusion (transposes)
3) semi-Lagrangian near the poles
4) current DM implementation will not scale

Yin-Yang grid configuration

- Implemented as 2 LAMs communicating at boundaries
- Optimized Schwarz iterative method for solving the elliptic problem.
- Scales a whole lot better
- Operational implementation due
- in spring 2015
- Communications are an issue
- Exchanging a Global Uniform scalability problem (poles) by another scalability problem



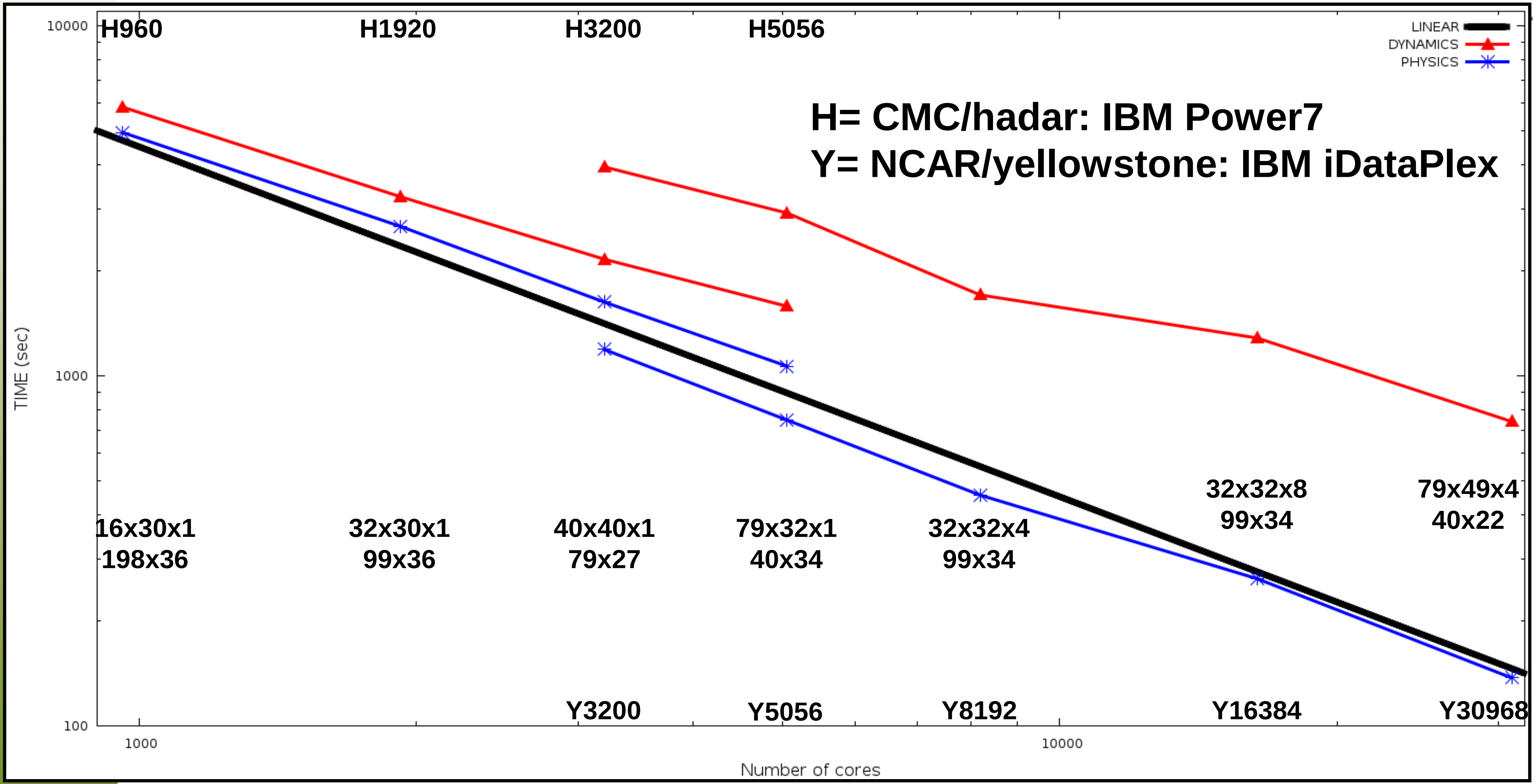




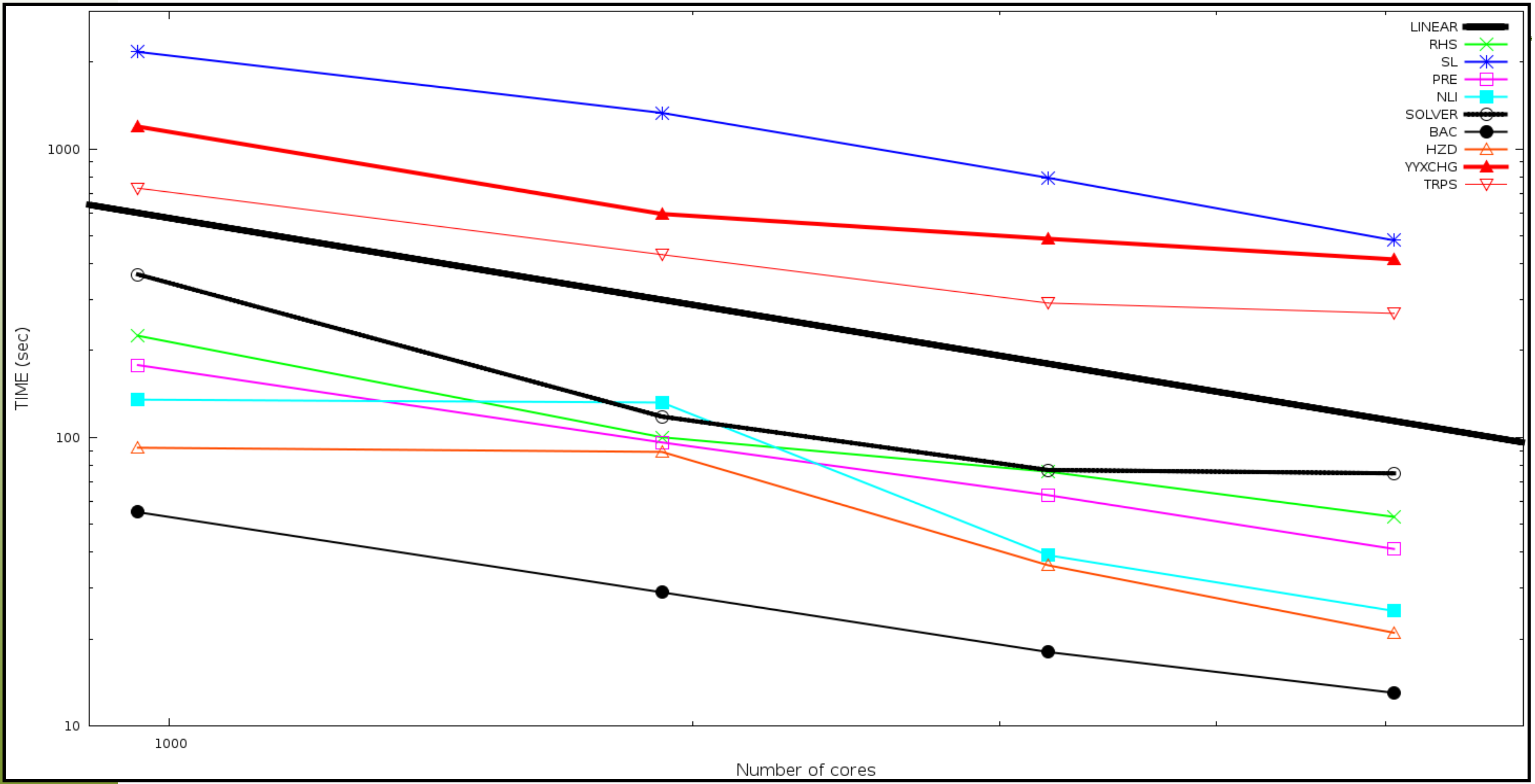
Canada

Yin-Yang 10 km scalability

Ni=3160, Nj=1078, Nk=158



Yin-Yang 10 km scalability
Dynamics components



Environment Canada Environnement Canada

Canada

The future of GEM

- Yin-Yang 2km on order 100K cores is already feasible on P7 processors or similar
- Yin-Yang exchanges will need work
- Using GPUs capabilities is on the table
- Improve Omp scalability
- Re-partition MPI sub-domain (bni,bnj,ni,nj,nk)
- Export SL interpolations to reduce halo size
- Processor mapping to reduce the need to communicate through the switch
- Partition NK
- MIMD approach for I/O

Investigating scalability and accuracy on an icosahedral geodesic grid

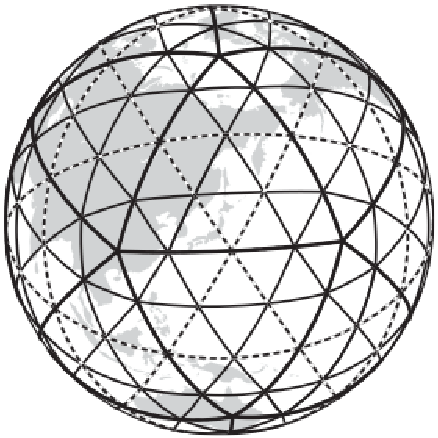
Spatial discretization: finite volume method on icosahedral geodesic grid

Time discretization: exponential integration methods which resolve high frequencies to the required level of tolerance without severe time step restriction

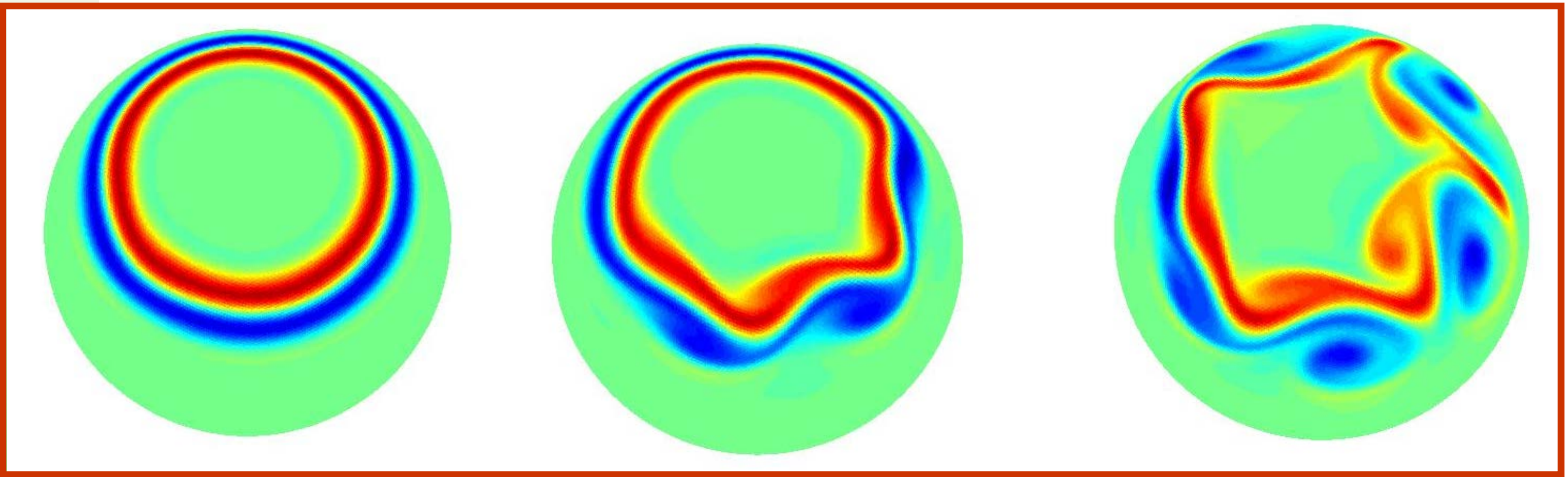
Shallow water implementation already shows great scalability

Vertical coordinates:
Generalized quasi-Lagrangian with conservative monotonic remapping

SPHERICAL GEODESIC
OR ICOSAHEDRAL GRID



Unstable jet, icosahedral grid number 6, dx= 112 km, dt=7200 sec (typically 30 sec)



Pudykiewicz (2006), J. Comp. Phys., 213, pp 358-390
Pudykiewicz J. (2011), J. Comp. Phys., 230, pp 1956--1991
Qaddouri A., J. et al. (2012), Q. J. Roy. Met. Soc., 138, pp 989--1003
Clancy C., Pudykiewicz J. (2013), Tellus A, vol. 65

Ensemble-Variational Assimilation: EnVar

- EnVar will replace 4D-Var at Environment Canada in 2014 for both global and regional deterministic prediction systems
- EnVar uses a **variational assimilation approach** in combination with the already available **4D ensemble covariances** from the EnKF
- By using 4D ensembles, EnVar performs a 4D analysis without need of tangent-linear/adjoint of forecast model
- Consequently, it is **more computationally efficient** and **easier to maintain/adapt** than 4D-Var:
  - EnVar: ~10 min, 320 cores, 50km grid spacing
  - 4D-Var: ~1hr, 640 cores, 100km grid spacing
- Future improvements to EnKF will benefit both ensemble and deterministic forecasts→ **incentive to increase overall effort on EnKF development**
- EnKF scalability examined by Houtekamer et al. (2014, MWR)

EnVar formulation

- In 4D-Var the 3D analysis increment is evolved in time using the TL/AD forecast model (here included in $\mathbf{H}_{4D}$):

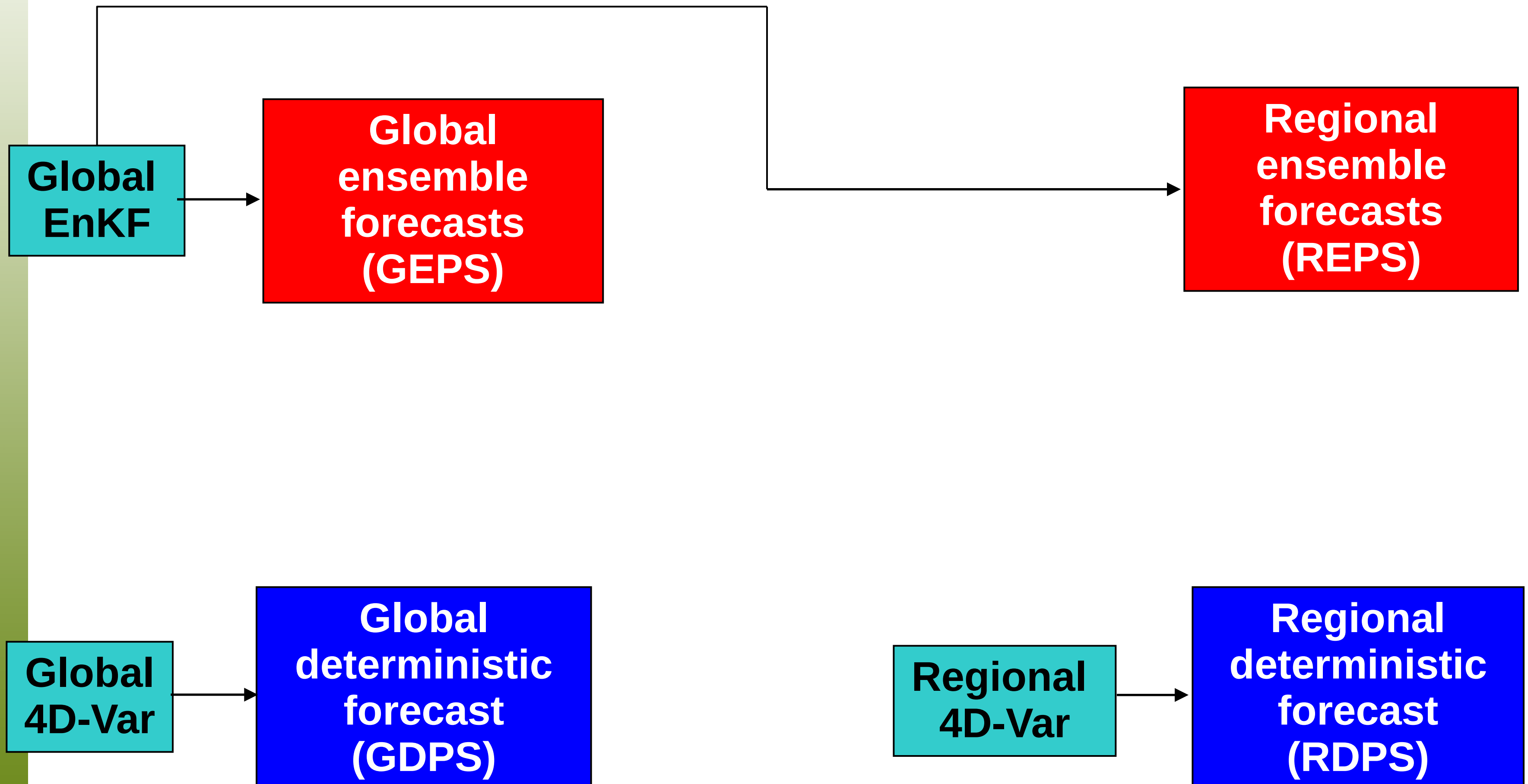
$$J(\Delta\mathbf{x}) = \frac{1}{2}(H_{4D}[\mathbf{x}_b] + \mathbf{H}_{4D}\Delta\mathbf{x} - \mathbf{y})^T \mathbf{R}^{-1}(H_{4D}[\mathbf{x}_b] + \mathbf{H}_{4D}\Delta\mathbf{x} - \mathbf{y}) + \frac{1}{2}\Delta\mathbf{x}^T \mathbf{B}^{-1}\Delta\mathbf{x}$$

- In EnVar the background-error covariances and increment are explicitly 4-dimensional, resulting in cost function:

$$J(\Delta\mathbf{x}_{4D}) = \frac{1}{2}(H_{4D}[\mathbf{x}_b] + \mathbf{H}\Delta\mathbf{x}_{4D} - \mathbf{y})^T \mathbf{R}^{-1}(H_{4D}[\mathbf{x}_b] + \mathbf{H}\Delta\mathbf{x}_{4D} - \mathbf{y}) + \frac{1}{2}\Delta\mathbf{x}_{4D}^{\ T} \mathbf{B}_{4D}^{-1}\Delta\mathbf{x}_{4D}$$

2013-2017: Toward a Reorganization of the NWP Suites at Environment Canada

Current operational systems

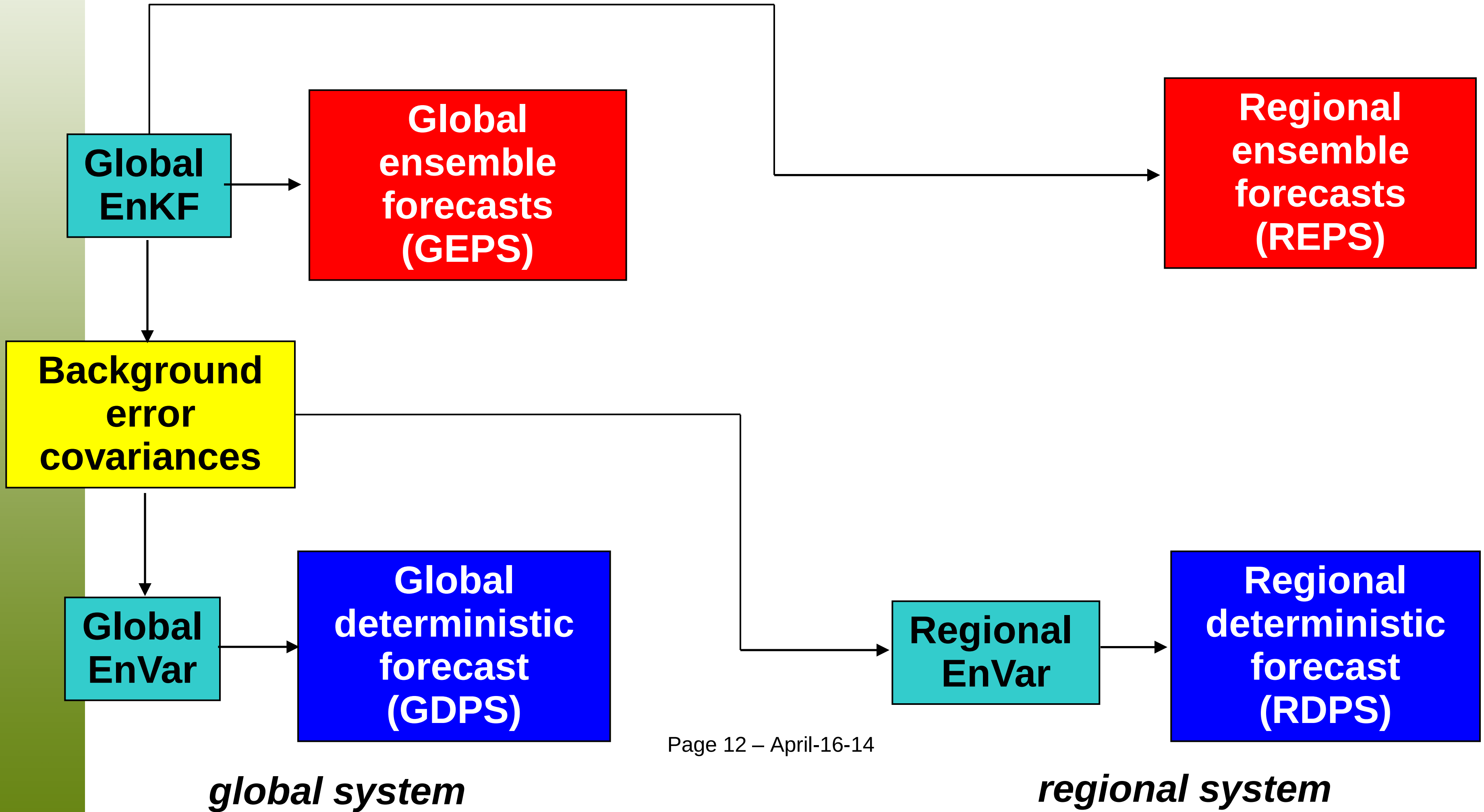


global system

regional system

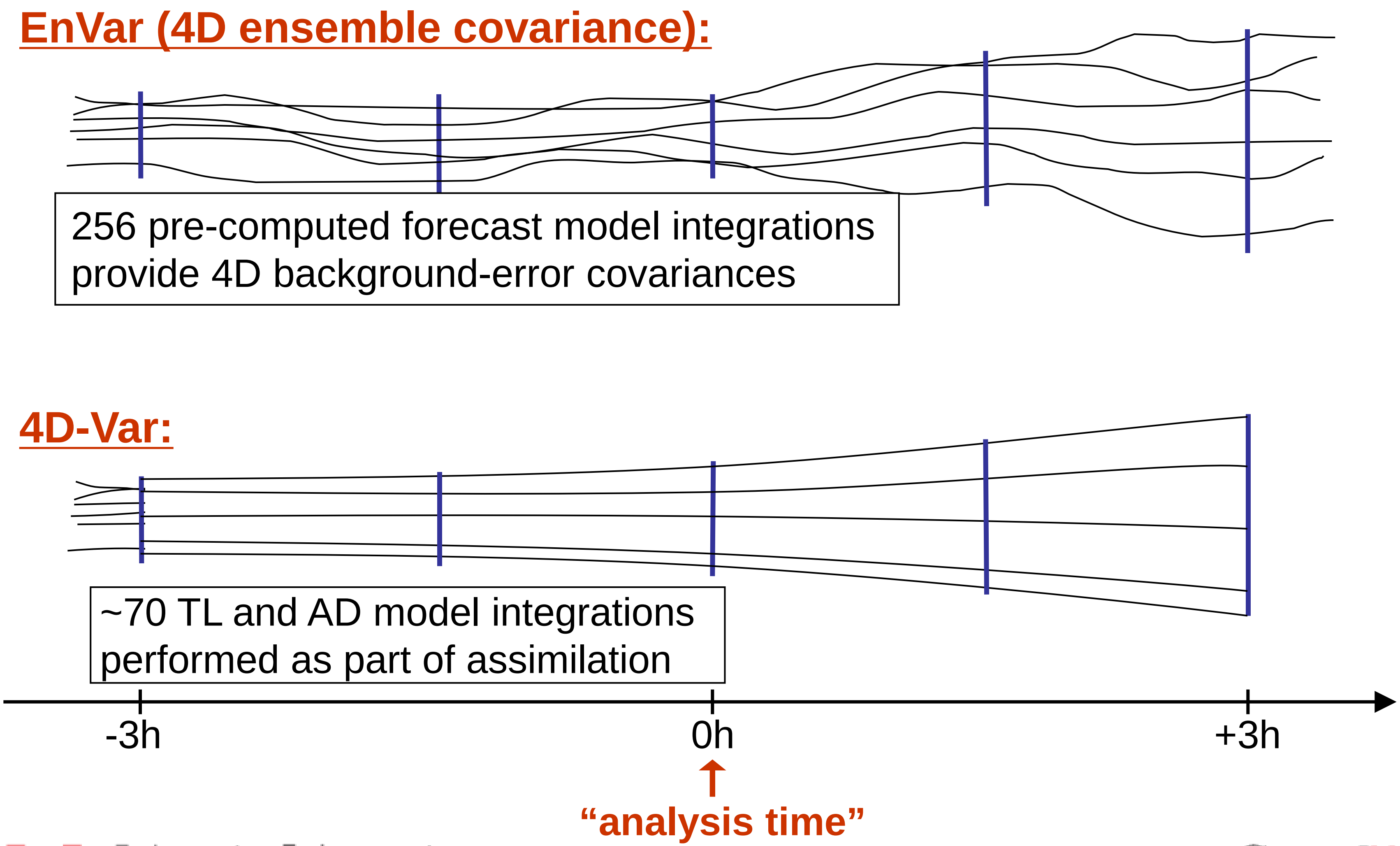
2013-2017: Toward a Reorganization of the NWP Suites at Environment Canada

2014 implementation: 4D-Var replaced by EnVar



4D error covariances

Temporal covariance evolution



Spatial covariance localization in EnVar
(Buehner 2005, Lorenc 2003)

- Complexity of 4D-Var is replaced by relatively simple use of 4D ensemble covariances to compute analysis increment and gradient during each iteration of the cost function minimization
- If the spatial localization function is horizontally homogeneous and isotropic, then spectral approach is efficient (for global lat-lon system):

$$\Delta \mathbf{x}_{4D} = \sum_k \mathbf{e}^k{}_{4D} \circ (\mathbf{L}^{1/2}\, \boldsymbol{\xi}^k) = \sum_k \mathbf{e}^k{}_{4D} \circ (\mathbf{S}^{-1}\, \mathbf{L}_v{}^{1/2}\, \mathcal{L}_h{}^{1/2}\, \boldsymbol{\xi}^k)$$

where:

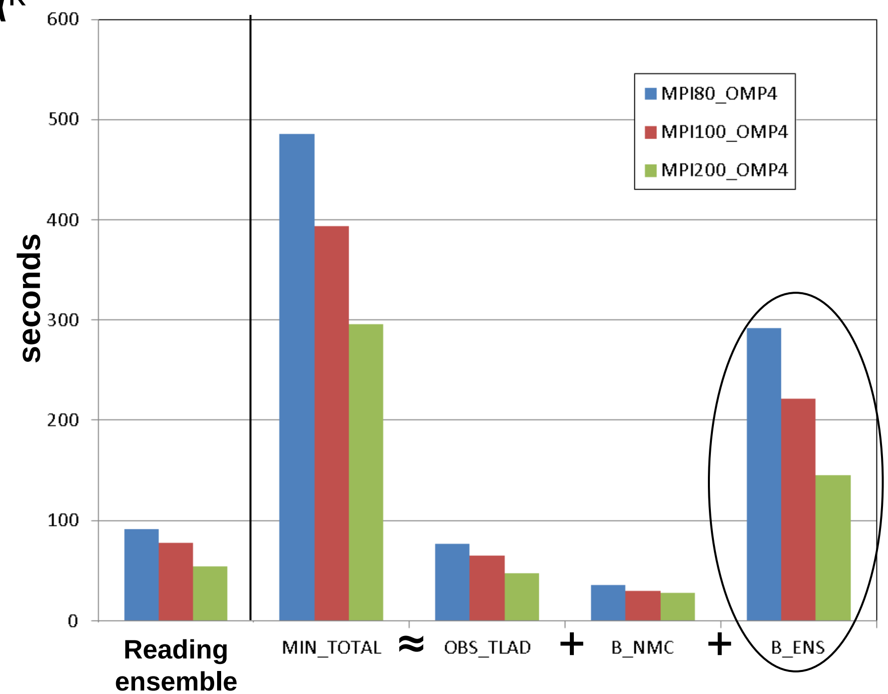
- $\mathbf{e}^k{}_{4D}$ is the kth 4D ensemble deviation
- $\boldsymbol{\xi}^k$ is corresponding control vector
- $\mathbf{S}^{-1}$ is the inverse spectral transform
- $\mathcal{L}_h = \mathbf{S}\mathbf{L}_h\mathbf{S}^T$ is the diagonal spectral horizontal localization matrix

- Equivalent to using sqrt of localized sample covariance matrix:

$$\mathbf{L} \circ [\textstyle\sum_k \mathbf{e}^k\, (\mathbf{e}^k)^T]$$

Scalability of EnVar

- Computation and I/O must be highly parallel, $\mathbf{e}^k_{4D}$ is large:
  - 256 members, 7 times, 4 vars, 800x400x75L → ~640GB
- Calculation of analysis increment (and adjoint) currently about 1/2 of overall cost of minimization: $\Delta\mathbf{x}_{4D} = \sum_k \mathbf{e}^k_{4D} \circ (\mathbf{S}^{-1}\mathbf{L}_v^{1/2}\mathcal{L}_h^{1/2}\boldsymbol{\xi}^k)$
- Currently, Schur product is the most expensive (and simplest) step, but scales perfectly: $\Delta\mathbf{x}_{4D}=\sum_k \mathbf{e}^k_{4D} \circ \alpha^k$
- Spectral transform is next most expensive, involves a global transpose: $\alpha^k=\mathbf{S}^{-1}\mathbf{L}_v^{1/2}\mathcal{L}_h^{1/2}\boldsymbol{\xi}^k$
- Currently only 1D decomposition:
  - $\boldsymbol{\xi}^k$: split by ens member (256)
  - $\Delta\mathbf{x}$: split by latitude (400)
- Improved scalability requires 2D decomposition (3D transposition strategy) → under development



END