Advancing U.S. Operational Weather Prediction Capabilities in the Next Decade (with Exascale HPC)

Mark Govett

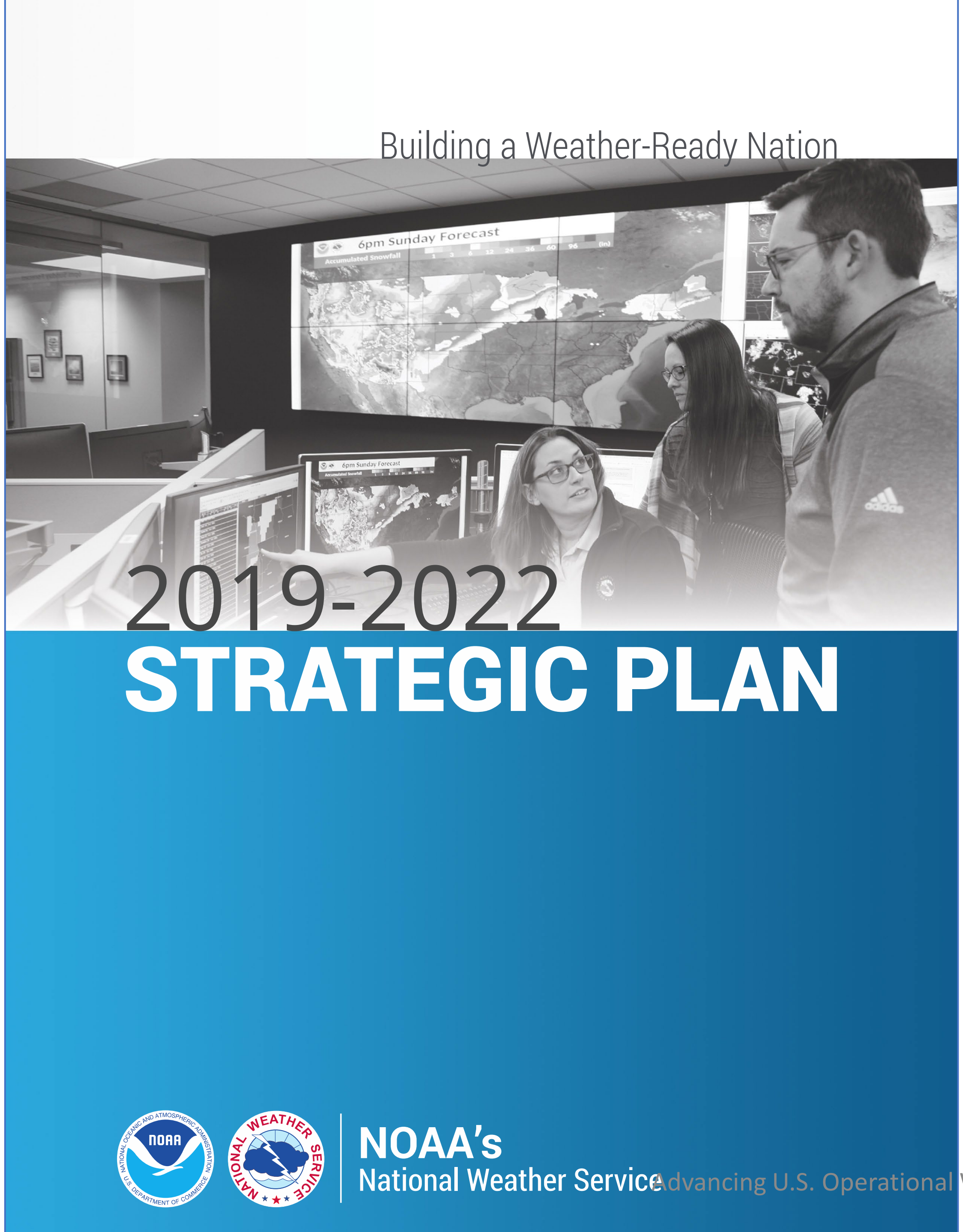
NOAA Earth System Research Laboratory

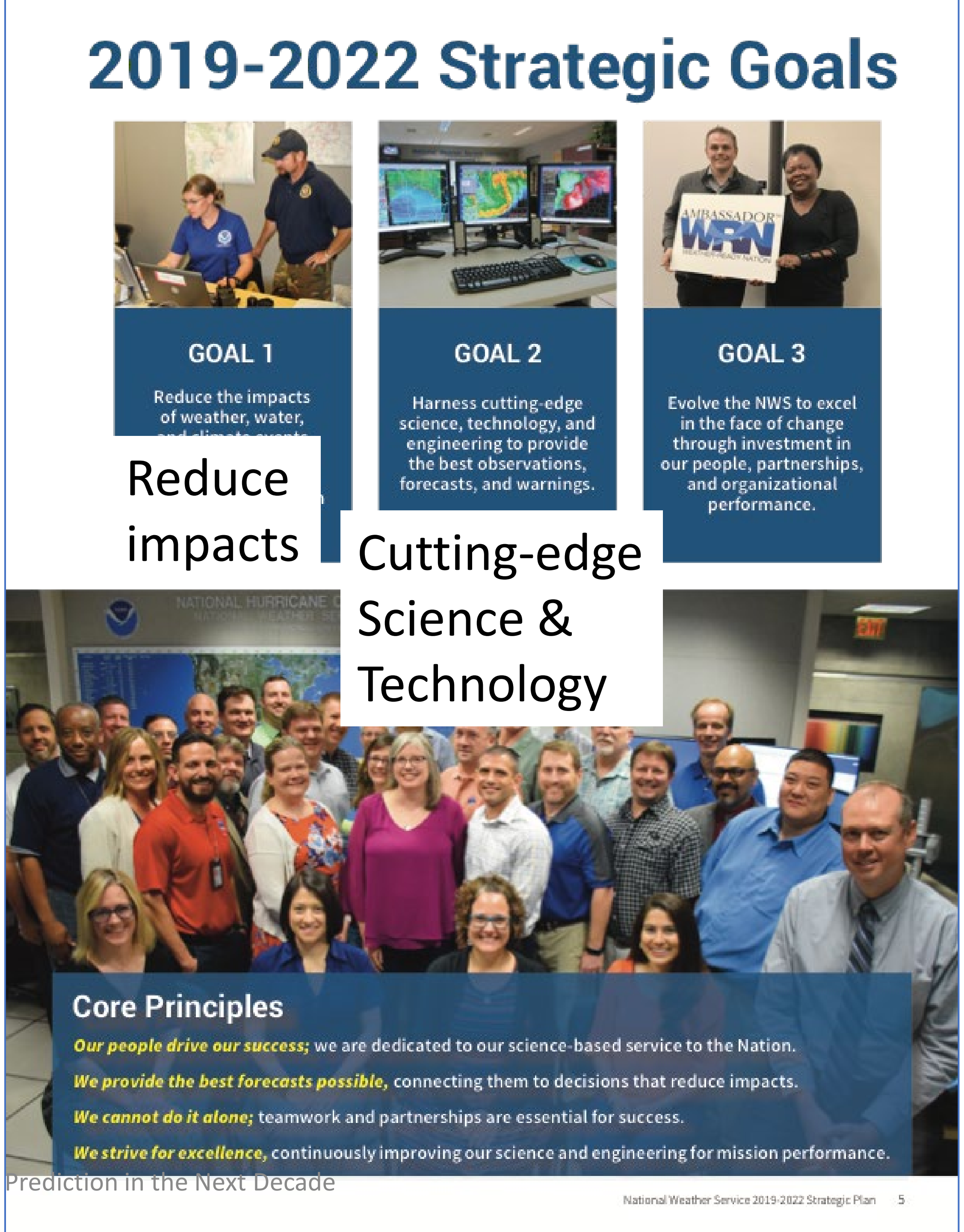
Global Systems Division

Boulder Colorado

NCEP Seminar

May 21, 2019



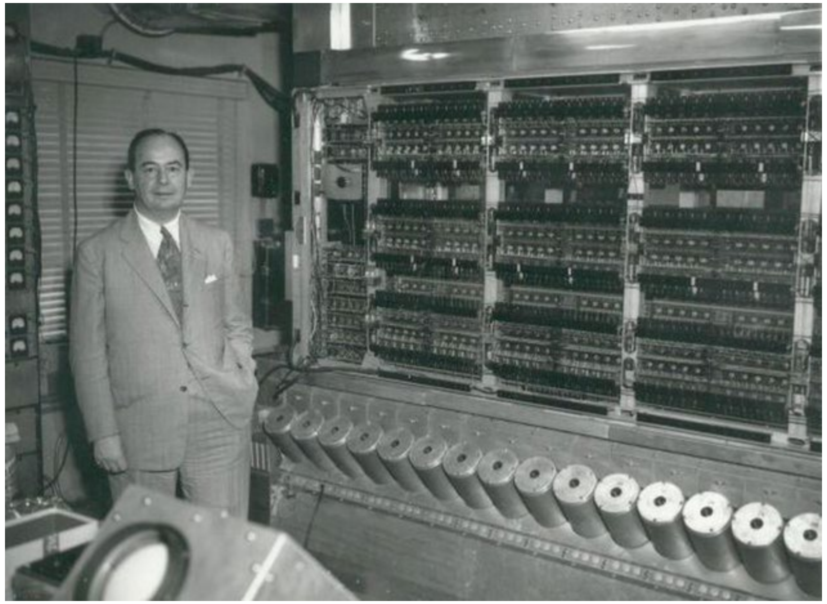


Reduce impacts

Cutting-edge Science & Technology

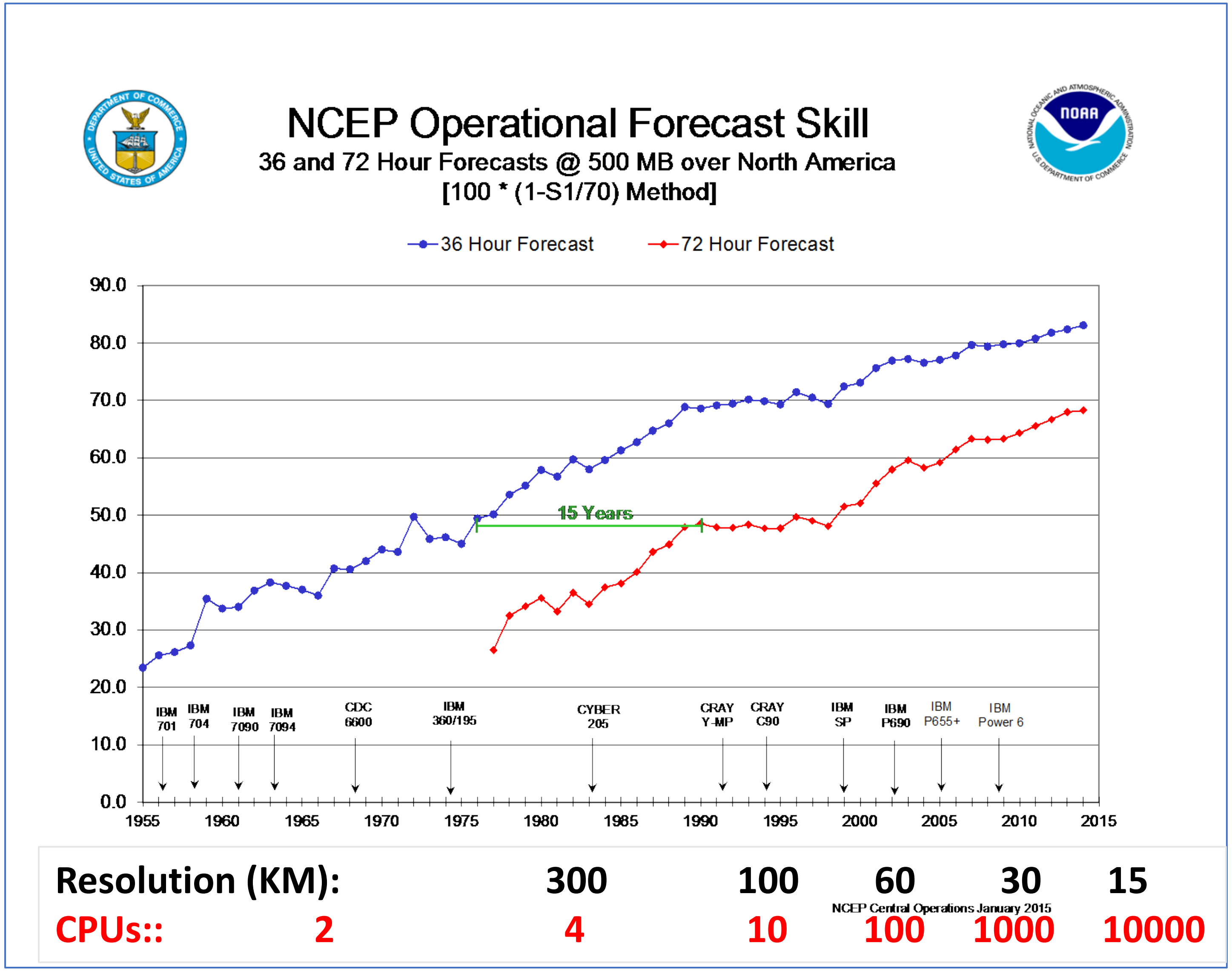
Advancing U.S. Operational Weather Prediction in the Next Decade

HPC & NWP



John von Neumann posing with the ENIAC computer, 1946

photo courtesy of NOAA

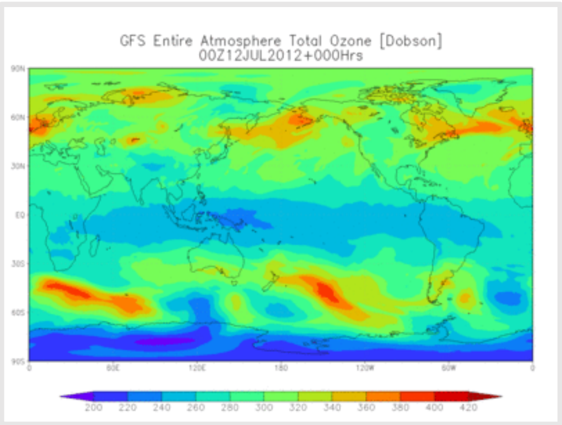


NWS Weather Forecast Models (2019)

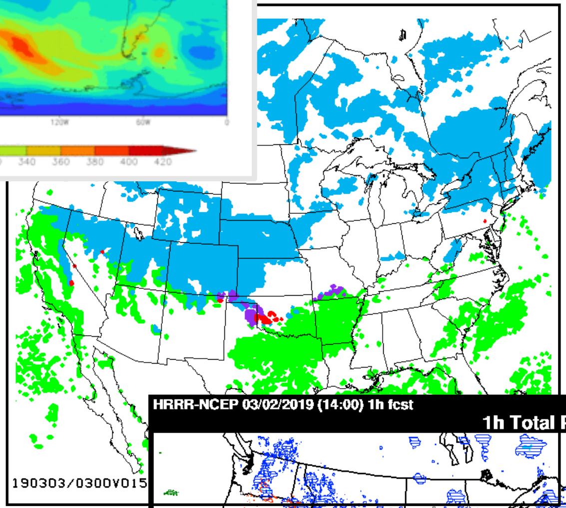
constrained by HPC

Higher resolution *means*
smaller domain *and*
shorter forecasts

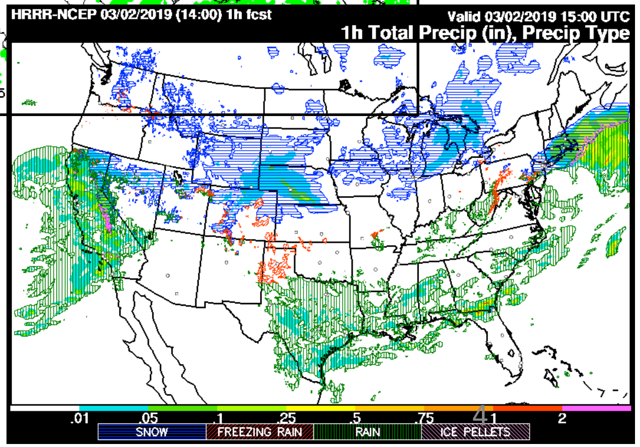
GFS



NAM



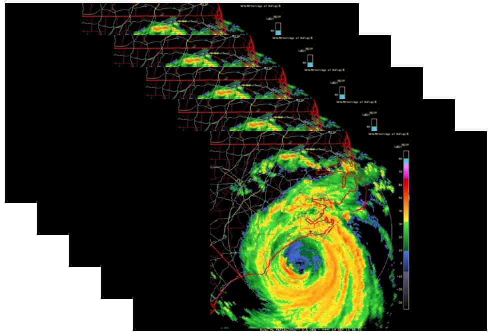
HRRR



- <u>Global</u>: Global Forecast System (GFS) (28 KM)
  - Weeks: 0 - 16 day forecasts, 4x / day
- <u>Regional</u>: North American Model (NAM) (12KM)
  - Days: 84 hours, 4x/day
- <u>Regional</u>: High Resolution Rapid Refresh (3KM)
  - Hours: 36 hours, 24x/day

Improved Weather Prediction

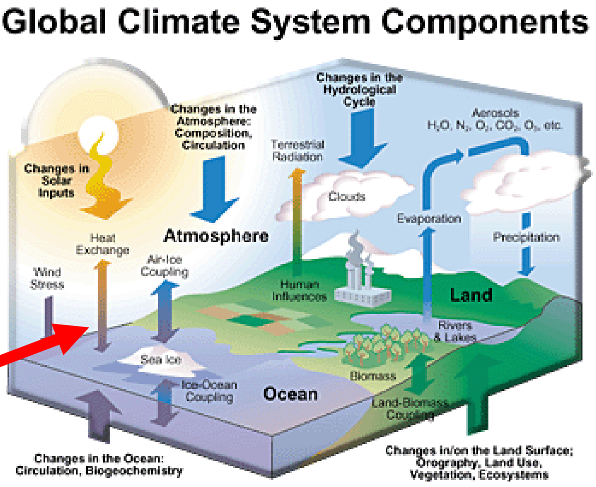
10-100s of members



Ensembles

Model complexity

Global Weather System Components



is a tradeoff between

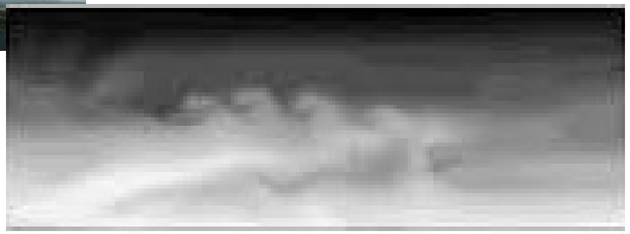
- Computing
- Scientific Accuracy
- Time-to-solution

Model resolution

13 KM

3 KM

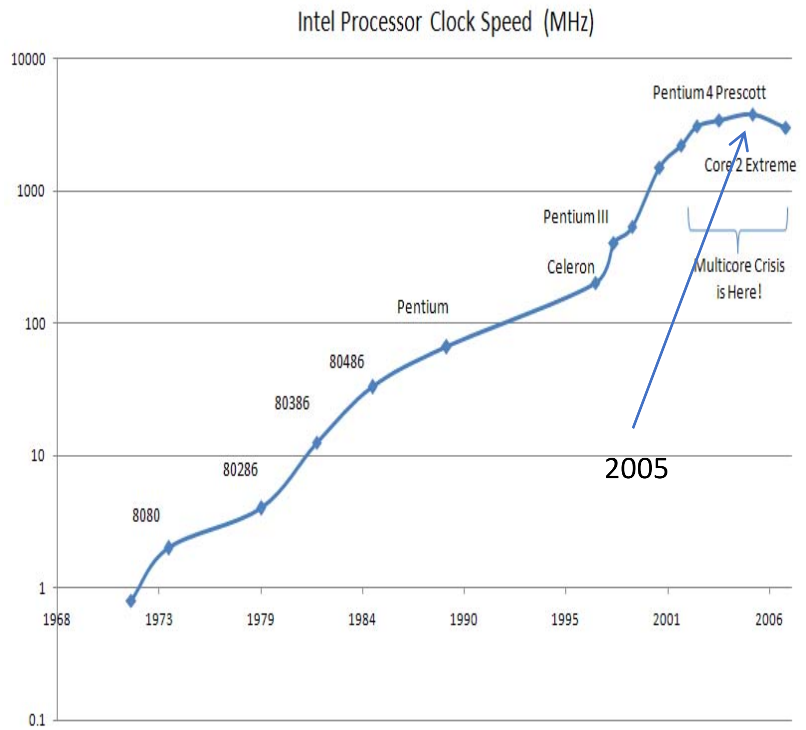
1 KM



Computational Challenges

- Processors are not getting faster
  - Doubling of resolution requires 8X more processors

- ESPC HPC Working Group: 2016 -
  - NOAA, NASA, DoE, DoD Navy, NCAR
  - Discuss HPC challenges, limitations for weather & climate applications



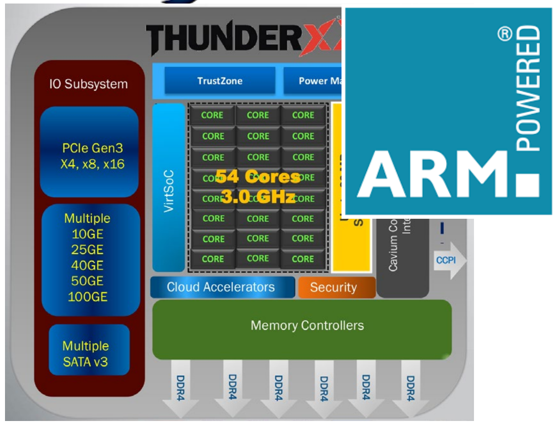
"**HPC architectures are developing in the wrong direction** for state-heavy, low computational intensity (CI) Earth system applications."

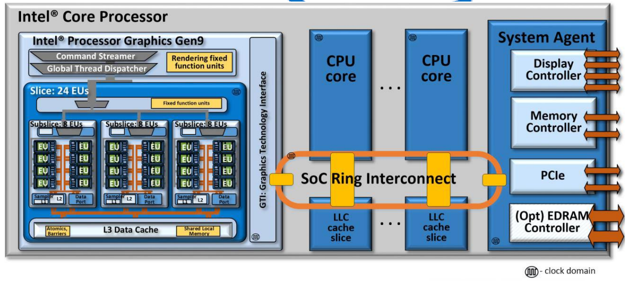
"**NWP applications average less than 2% of peak performance**, constrained by their ability to perform sufficient calculations for each expensive access to memory."

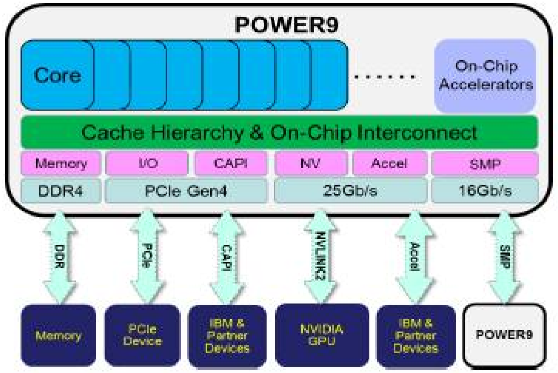
Carman, et al. "**Position Paper on High Performance Computing Needs in Earth System Prediction**." National Earth System Prediction Capability (ESPC) program. April 2017. https://doi.org/10.7289/V5862DH3

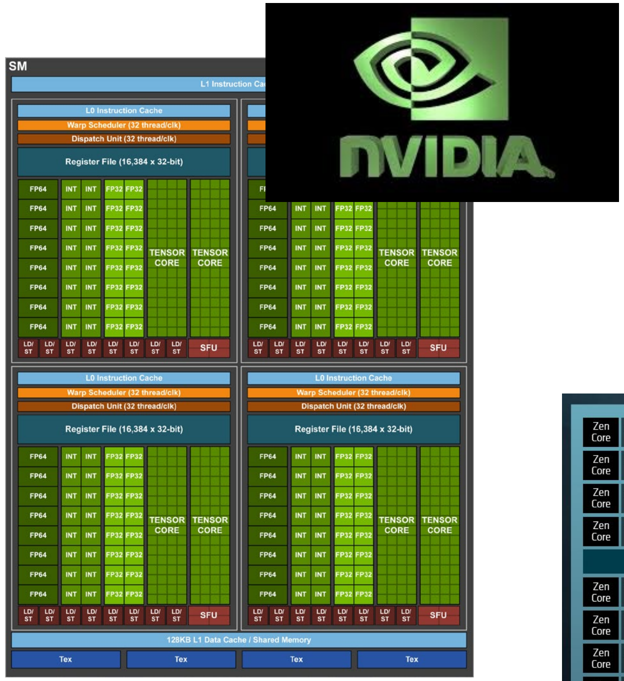
Processor Technologies

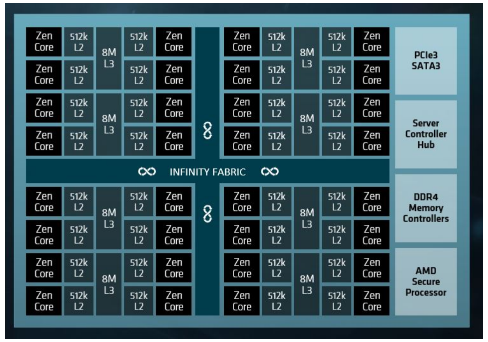
- CPU, GPU TPU, FPGA, ARM
- Diversity
  - Processor
    - Clock speed, energy consumption
    - 10’s to 1000’s of cores
    - Single, double, half precision
  - Memory
    - Size, speed, type
- Burden on compilers, libraries
  - Portability
  - Performance
  - Interoperability





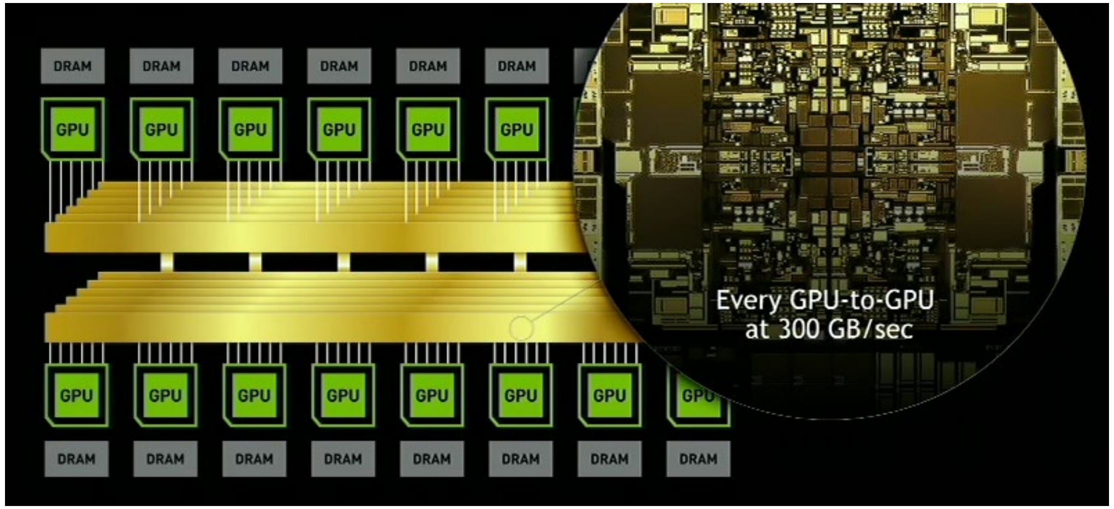






Compute Nodes

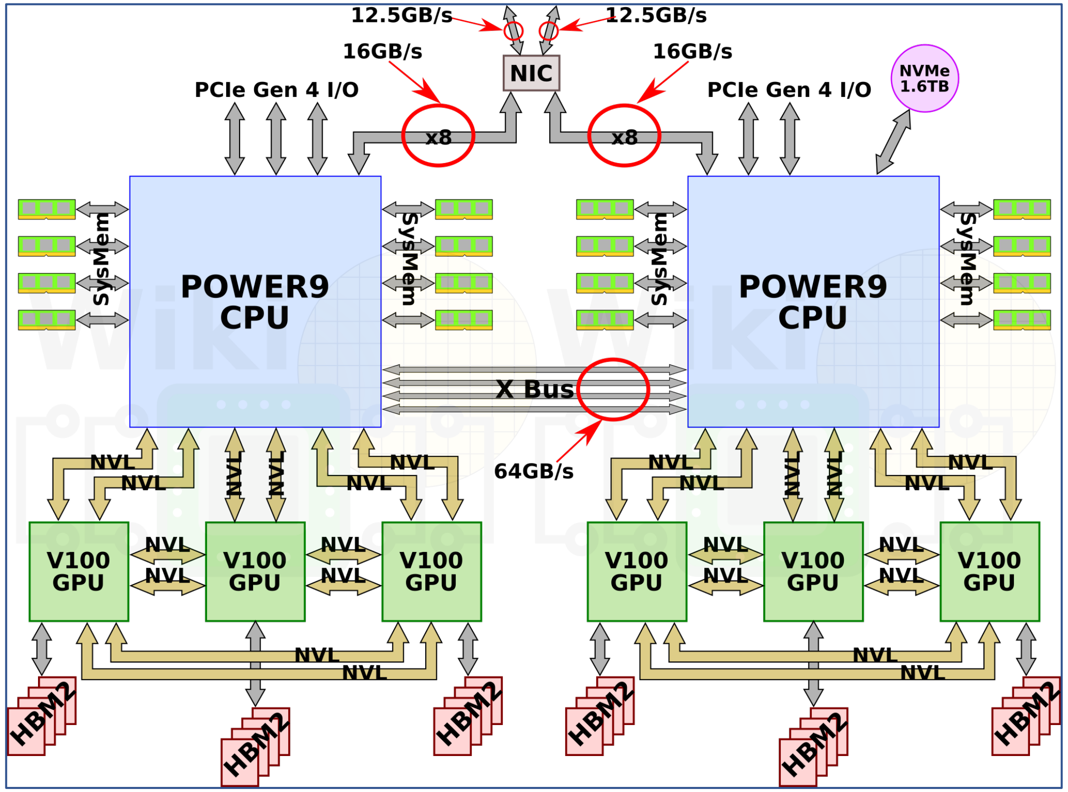
- Standard is a dual-socket CPU
- Increasing complexity, more processors
  - CPU: 100's of cores
  - GPU: 10000 – 80000 cores



NVIDIA DGX-2: 16 Tesla V100 GPUs, (81K GPU, 10K Tensor cores)

- 1.5 TB DDR4 RAM, 500 GB HBM2,   10kW power
- 300 GB/s NVLINK
- PCIe Gen3, 8x EDR IB / 100 Gigabit Ethernet

Summit Node (DoE / ORNL)



DOE Summit node:

- IBM Power9 CPU, 6 V100 GPUs, 30K GPU cores
- 512 GB DDR4 RAM, 96 GB HBM2
- NVLINK, 50GB/s bandwidth per link
- PCIe Gen 4 (16GB/s) for inter-node, I/O

Summit System: 4600 nodes, 27K GPUs

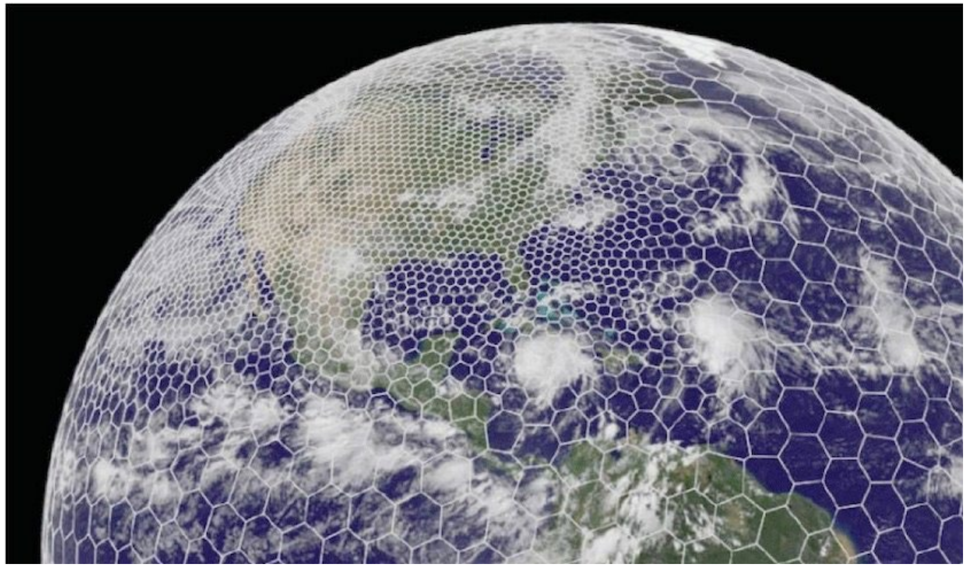
Application Performance – Single Node

MPAS
Model for Prediction Across Scales

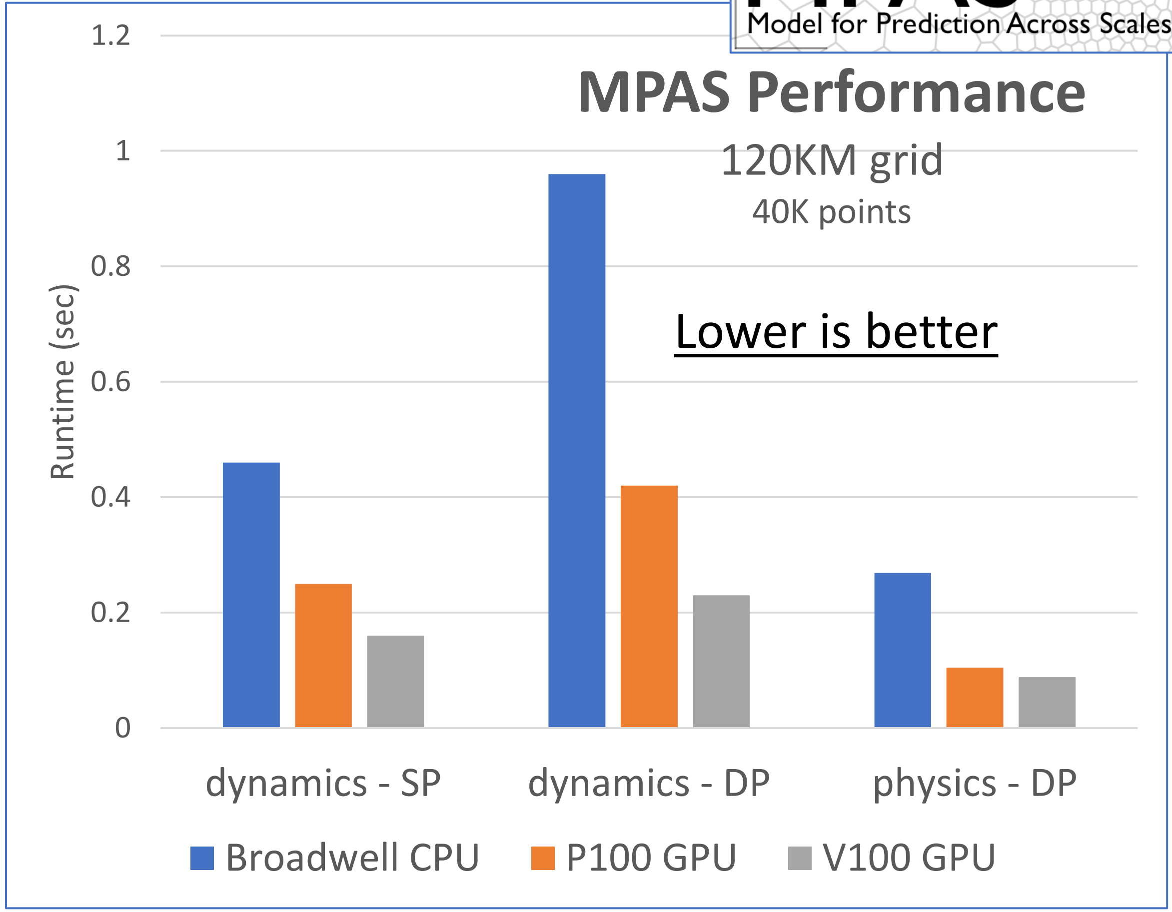
MPAS model developed at NCAR

adopted by IBM Weather Company

- GPU is 3X faster than CPU (Volta versus Broadwell)
- Directive-based, performance portable



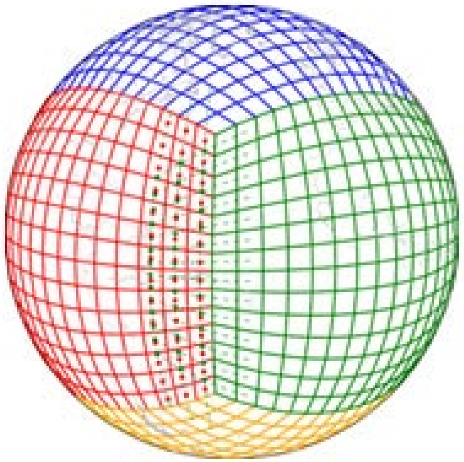
Non-uniform Icosahedral Grid



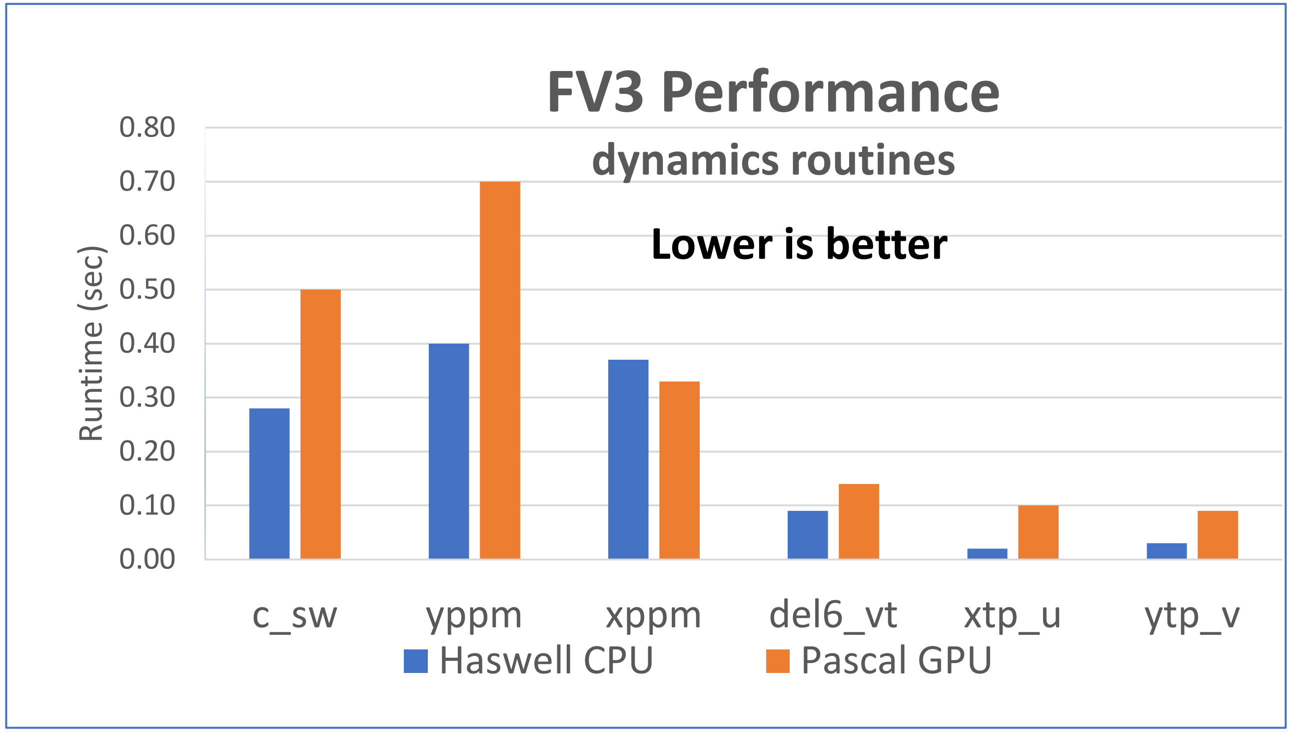
R.Loft, Sept 2018, ECMWF HPC Workshop

FV3 Performance – Single Node

- Designed for CPU,
  - **10% of CPU peak**
  - Efficient use of cache memory
- Slower on GPU
  - Not performance portable
  - Code changes slowed down CPU
- Inefficiencies
  - Limited fine-grain parallelism
  - Non-uniform cube-sphere grid
  - Pervasive edge & corner calculations
- Ongoing efforts to address GPU performance challenges



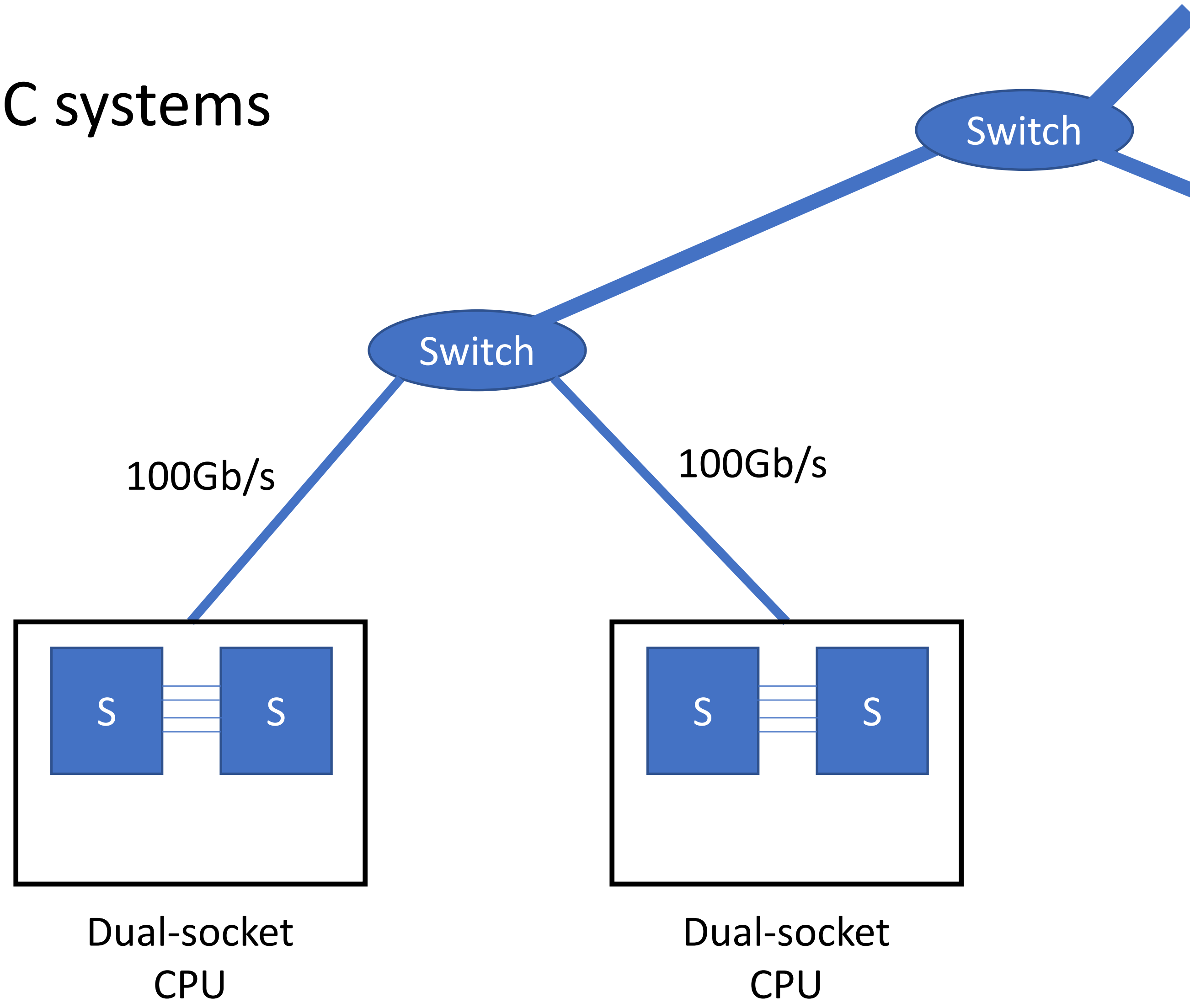
Cube-sphere grid



M. Govett, June 2018, PASC Symposium

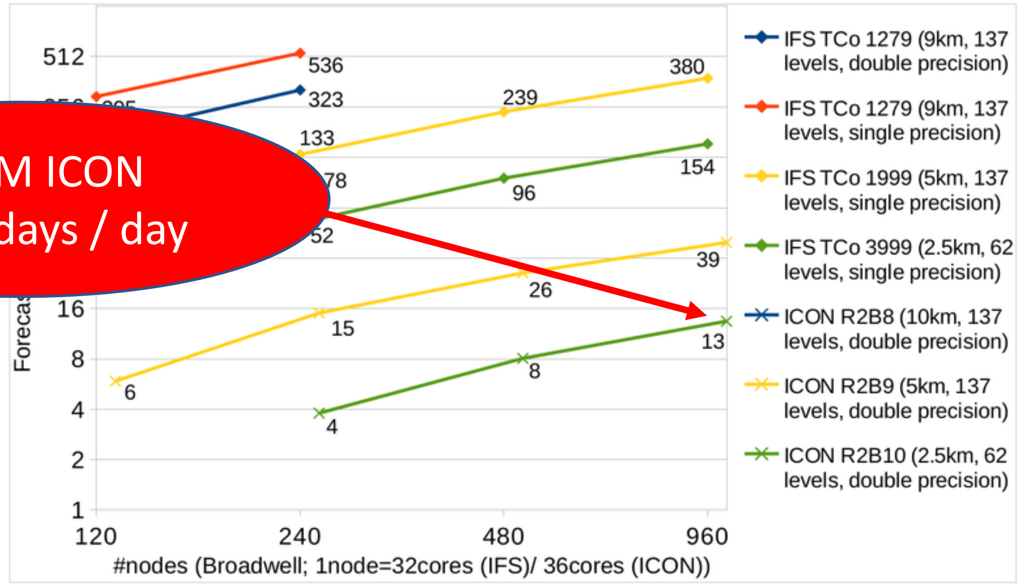
Inter-node Communications

- Inter-connect required for large HPC systems
  - Weakness in system deployments
- MPI communications
  - Pack message buffer
  - Inter-process communications
  - Unpack message buffer
- Scalability a big challenge for application performance



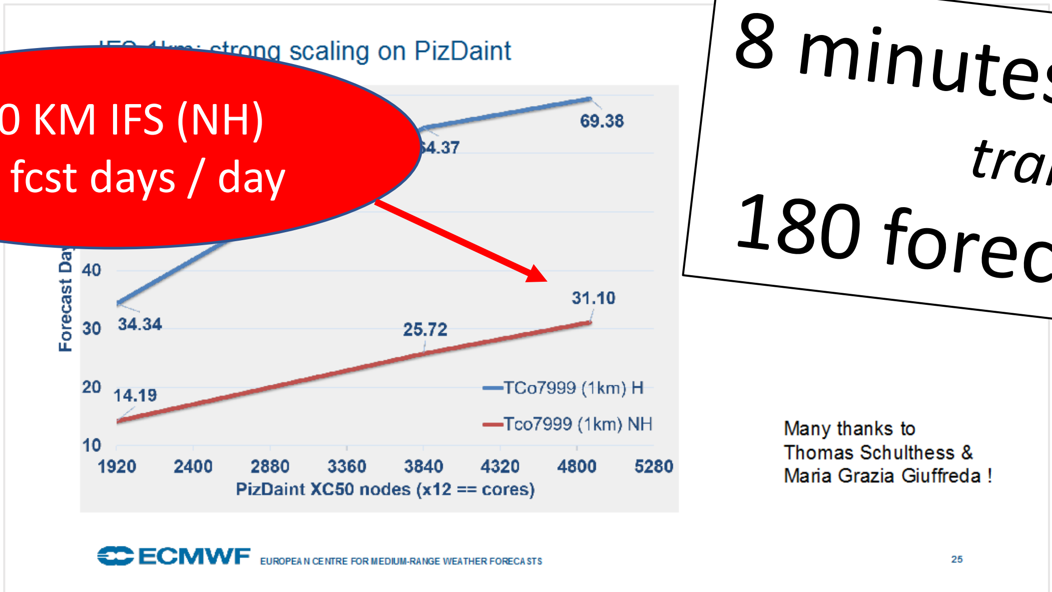
Application Scalability

- Targeting global 1-3 KM resolution
  - NICAM, ICON, MPAS, IFS, FV3, …
- ECMWF Scalability Programme (2014 -)
  - Scaling, I/O, compilers, algorithms



2.5 KM ICON
13 fcst days / day

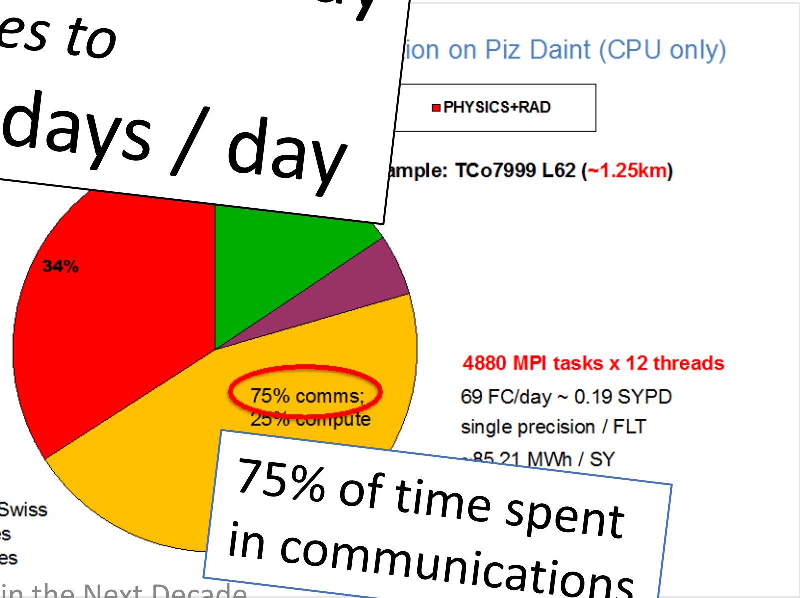
ESiWACE, DYAMOND project, 2017



1.0 KM IFS (NH)
31 fcst days / day

8 minutes / forecast day
translates to
180 forecast days / day

Nils Wedi, ESCAPE Project Presentation
ECMWF HPC Workshop, Sep 2018



75% of time spent
in communications

FV3 Scalability Projections

Perfect scaling: 2X increase in resolution requires 8X more compute cores

DYAMOND model configurations (32-bit, Cray XC40)

	Δx (km)	deep Conv	big_Δt (sec) (Slow physics)	L2E (sec) (intermediate physics)	Acoustic (sec) (Fast-physics)	Cores needed to meet NWP requirement* (estimated, minimal I/O)
C768_L63*	13	ON	225	225	18.75	3,000
C768_L63	13	OFF	225	225	18.75	3,000
C1536_L91	6.5	OFF	225	112.5	9.375	30,000
C3072_L91	3.25	OFF	225	56.25	4.5	240,000

*Assumed NWP requirements: 10 days forecast in less than 100 min.

Slide, S.J. Lin, DYAMOND presentation

FV3GFS Scaling - Estimate

3 KM resolution, 5 day forecast

Weak Scaling to increase resolution

Operational Requirement: 10 day forecast in 1.25 hours (5 days in 2250 seconds)

	Actual Performance		Estimated Performance	
Resolution	**28 KM**	**13 KM**	**6.50 KM**	**3.25KM**
Time Step	225 sec	112.5 sec	56 sec	28 sec
CPU Nodes	64	256	1024	4096
CPU cores	**1536**	**6144**	**24576**	**98304**
Total Time	1094	1916	3357	5880
Dynamics	560	792	1120	1584
Communications	440	710	1146	1851

Runtimes in seconds for a 5 day forecast, *NOAA theia system with 24 core Haswell nodes*

FV3GFS Scaling - Estimate

3 KM resolution, 5 day forecast

Strong Scaling to reduce runtime

Operational Requirement: 10 day forecast in 1.25 hours (5 days in 2250 seconds)

Tile Size / MPI	48 x 48	24 x 48	24 x 24
CPU Cores	98,304	196,608	393,216
Total Time	**5880**	**3962**	**2095**
Dynamics	1584	1275	643
Communications	1851	1390	801

Estimated performance, *NOAA theia system: 27,000 cores, 24 Haswell cores / node*

- 393,216 cores = 16,384 CPU nodes
- 30% of runtime is for inter-process communications

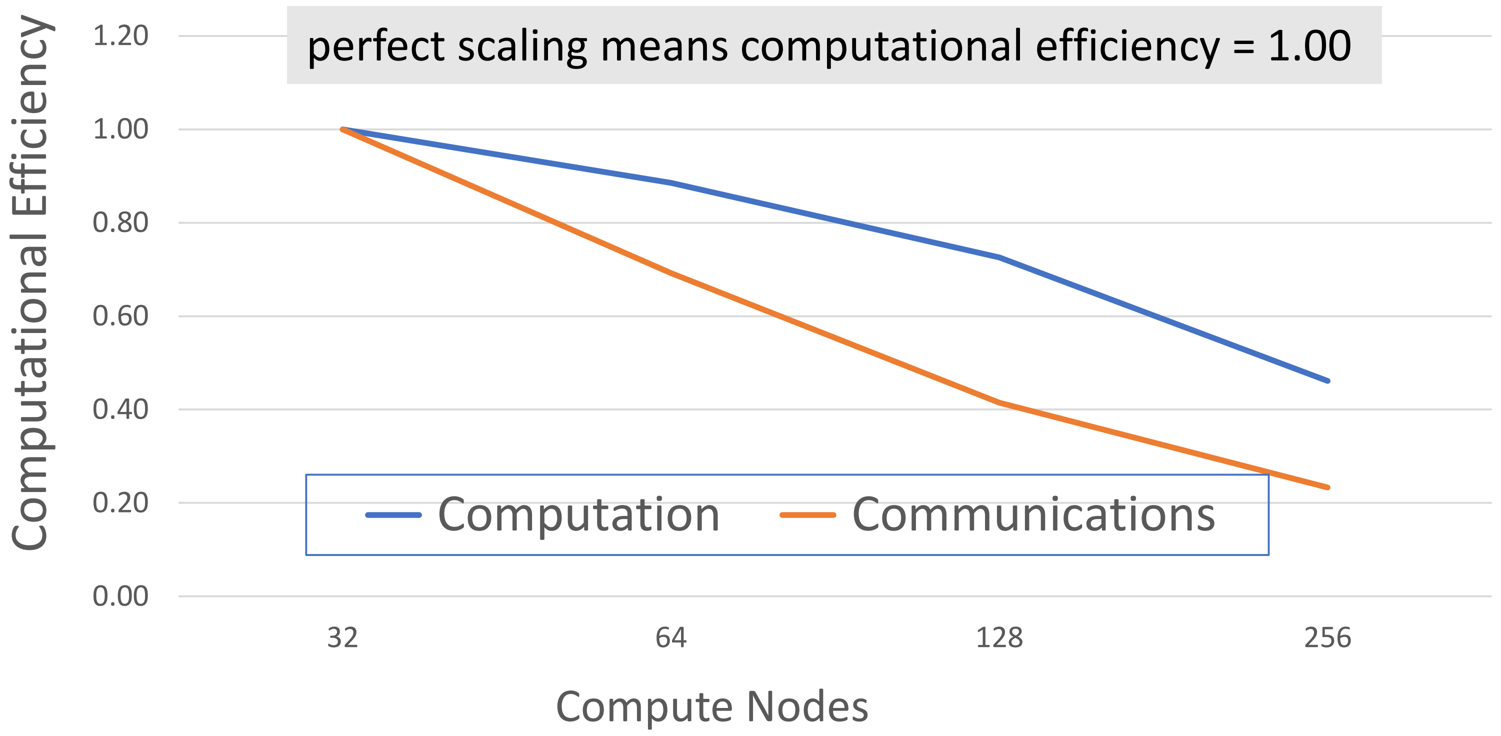
Scaling Factors

- Computation
  - Parallelism
  - Algorithms
  - Model grid
- Communications
  - Frequency
  - Data volume
  - Overlapping

FV3GFS Strong Scaling Efficiency

physics + dynamics

14 KM resolution



Summary on Computational Issues

- Traditional computing is not sufficient to run existing global operational models (ICON, IFS, FV3) at cloud-permitting (3KM) or finer scales
- GPU processors can help
- Scalability remains a big concern

Data Challenges

Data is only useful if it's used

Observations

Data Assimilation

Prediction

Output

Distribution

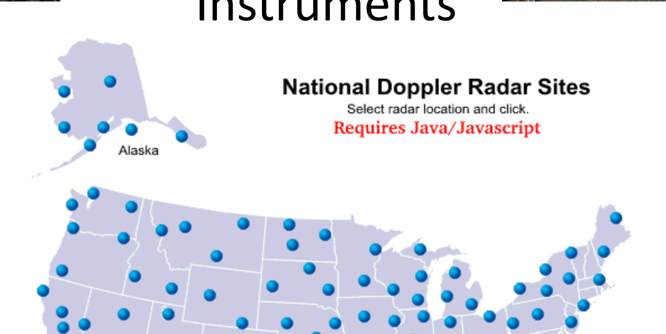
Dissemination

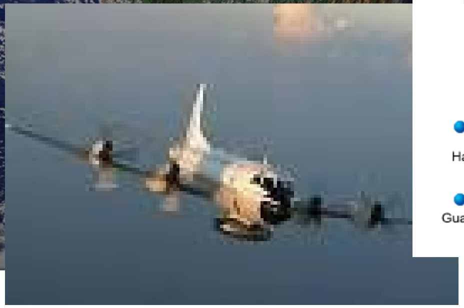
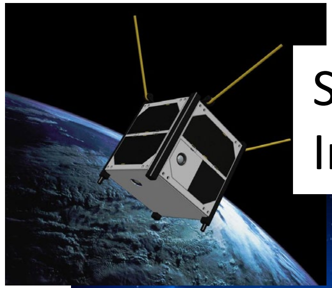
Observations

- We have more data than we can use
  - GOES, JPSS, cubeSats, nanoSats
  - Radar, balloons, ships, planes
- Tremendous potential
  - Autos, cell phones, sensors, …

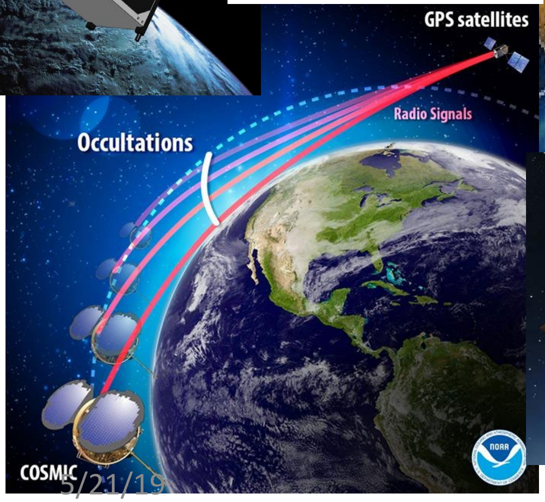



Ground-Based Instruments



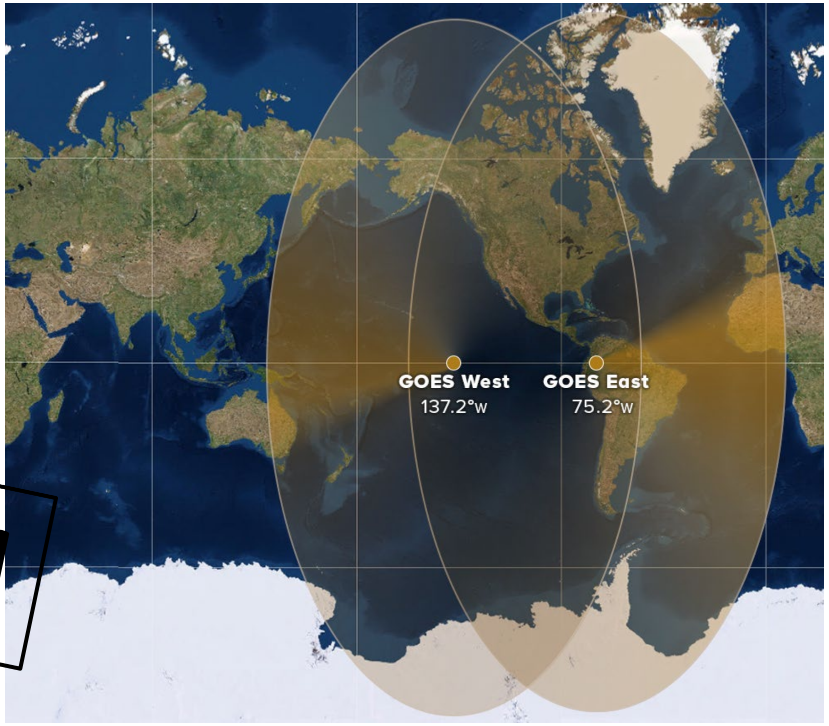



Space-Based Instruments

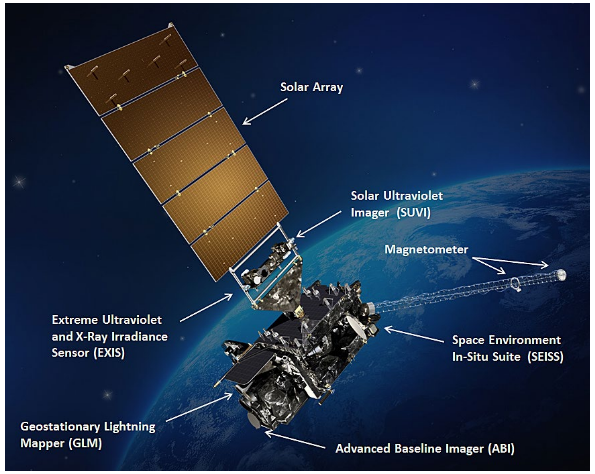



Geostationary Operational Environmental Satellite (GOES)

Only 1% is being used



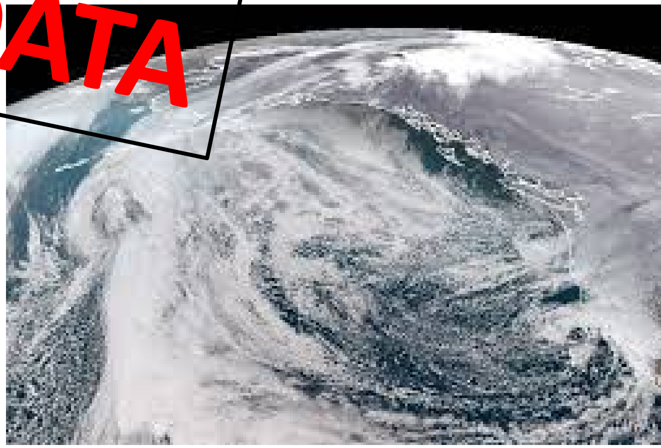
- 2012-2017: GOES-13, GOES-14, GOES-15
  - Scans every 3 hours, 10 bit precision
  - 4 spectral bands @ 4KM
  - 1 visible band @ 1KM



100x MORE DATA

- 2017 - ~2027: GOES-16, GOES-17
  - Scan every 15 minutes, 14 bit precision
  - 14 spectral bands @ 2KM resolution
  - 2 visible bands @ 0.5KM resolution
  - High-res nest every 30-60 seconds

water vapor image



Data Assimilation

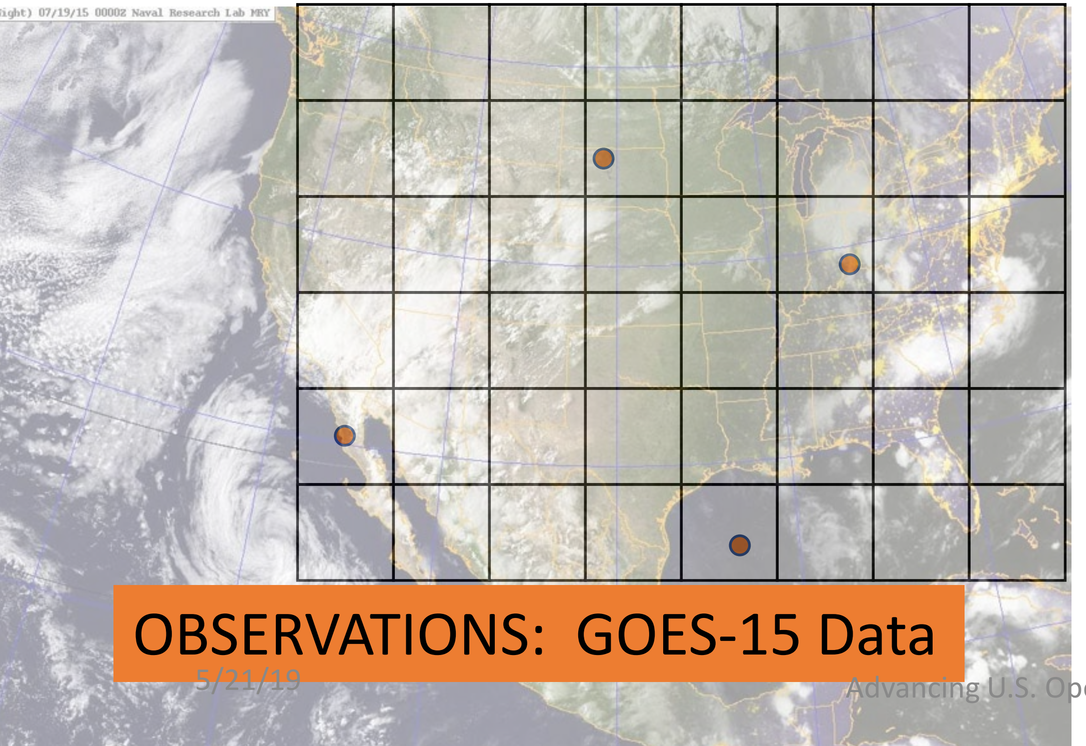
- Improve initial state of the forecast model
  - Variational, ensemble, hybrid approaches
- Complex, computationally expensive
  - ~3X lower resolution than prediction model

Calculations

- Estimate model error, observation error
- Interpolate model to observation
- Adjust nearby grid points, other model fields (winds, temp, …)

GOES-15: 4 KM resolution IR, 1 KM visible

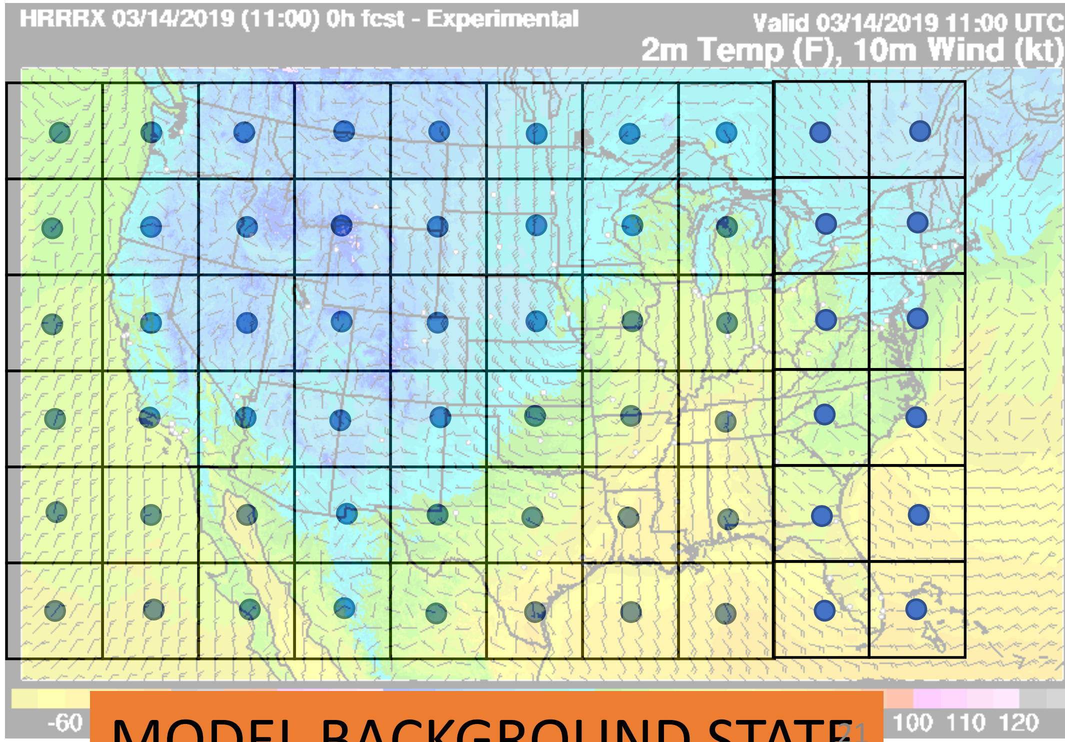
Data Thinning: currently use 1% of data



OBSERVATIONS: GOES-15 Data

update

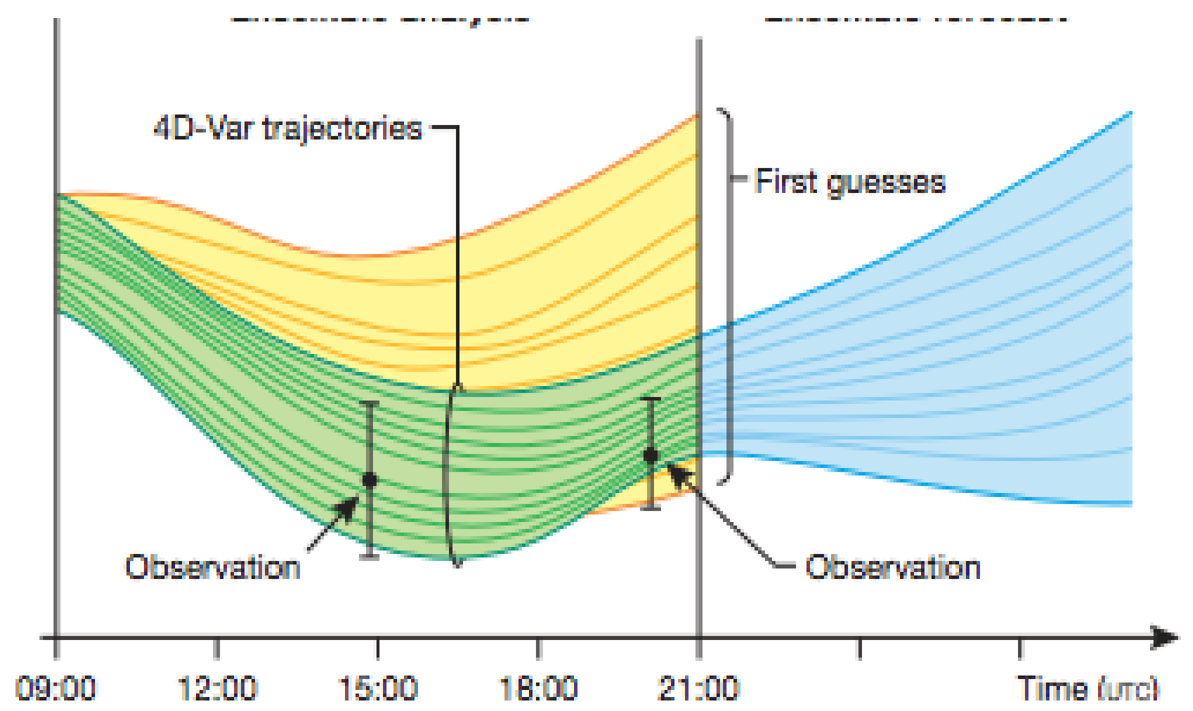
HRRR: 3 KM resolution, 2M temperature



MODEL BACKGROUND STATE

Data Assimilation: Computational Issues

- 3D Ensemble Based Assimilation
  - **Computational & I/O limitations**
  - Only afford 10’s of members
  - 3-10X lower resolution than model
- 4D Variational Assimilation
  - Higher accuracy
  - TL & ADJ are required
  - **~3X slower than 3DVAR methods**
- Hybrid EnKF & 4DVAR solutions
  - IFS, UK-Met, …



Ensemble analysis and forecast cycle

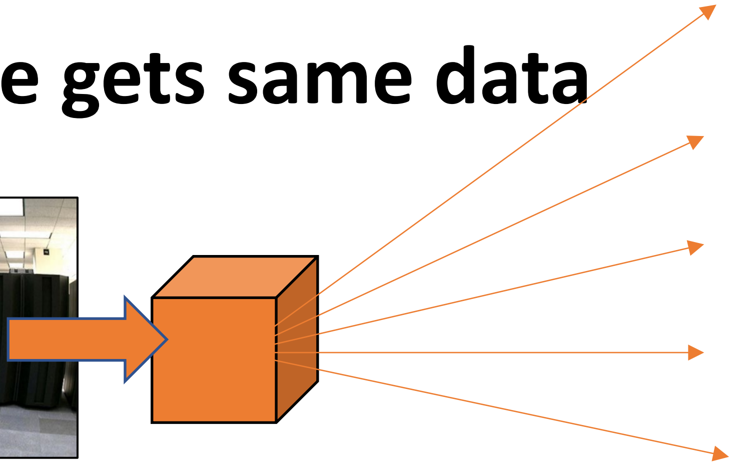
Image courtesy of M.Bonavita (ECMWF)

Data Distribution

- Diverse user requirements
  - Global, regional, local
  - Observations & products
- NWS AWIPS
  - NOAA network is saturated
- **Everyone gets same data**

AWIPS Workstation

NWS office



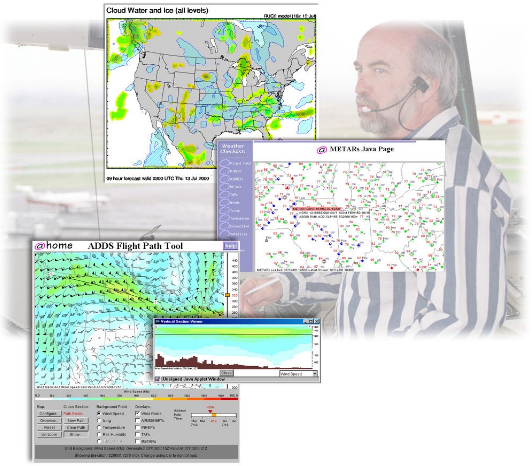
data center

model output

users

NWS Forecast Offices
Hurricane Prediction Center
Storm Prediction Center
National Water Center
Aviation Weather Center
Fire Weather Centers

State, Local, Public
- Floods, fire, winds, hail, …



FAA Air Traffic Control

Model Output: 14KM to 3KM resolution

- Each 3D variable: pressure, temperature, moisture, winds, ….

Resolution (KM)	Vertical Levels	Number of Grid Cells (Millions)	Total Cells (Billions)	Increase in Cells	Per field storage (SP)
14 (1x)	64 (1x)	3.5 (1)	0.25	1x	1 GB
3.5 (4x)	128 (2x)	56.6 (16)	5.4	21x	21 GB

- Model output: per run

14KM - 10 model fields, 6 hourly output, 10 day forecast 400 GB

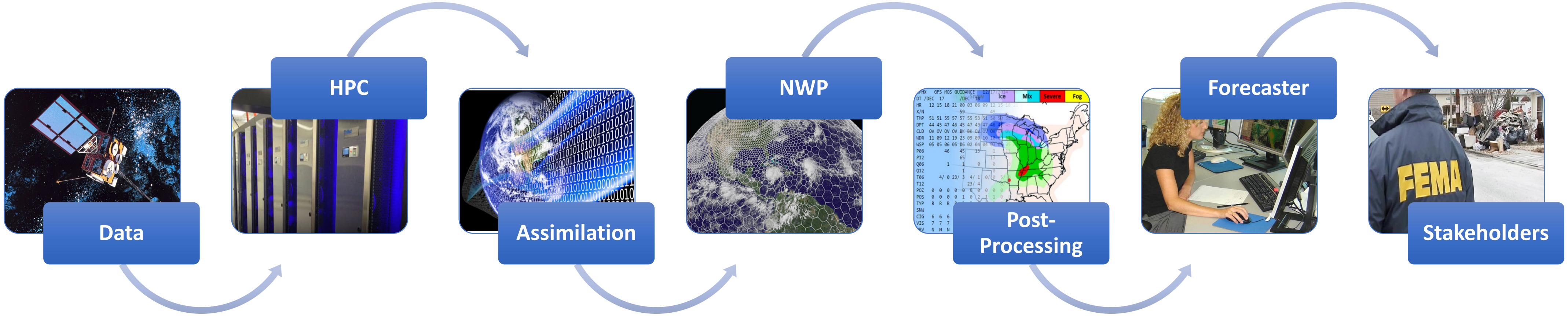
3KM - 10 model fields, 3 hourly output, 10 day forecast 21.8 TB (52X)

State of Operational NWP (2019)

- Exceedingly difficult to run operational 3KM
- HPC
  - No expected increase in processing speed
  - Limited increases in memory speed
  - Parallelism & scalability limitations
  - Operational time-to-solution constraints
- Data
  - Too much data to process
  - Too many observations to use
  - Too large to distribute

15 KM

3 KM



Advancing Weather Prediction

in the next decade

Utilize new technologies

Improved models

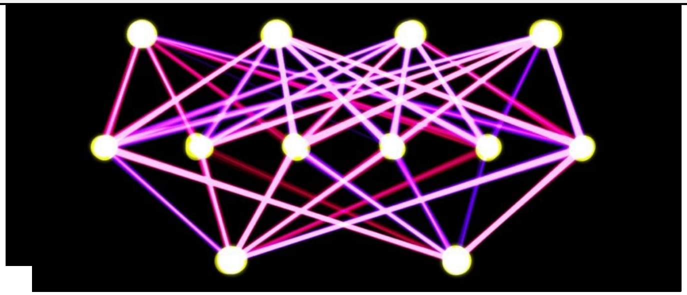
Better data handling

Manage software complexity

Technology Convergence

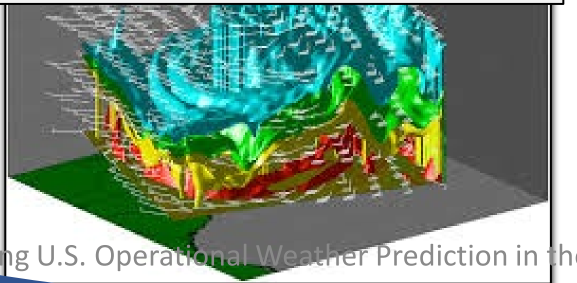
SuperComputing

Machine Learning



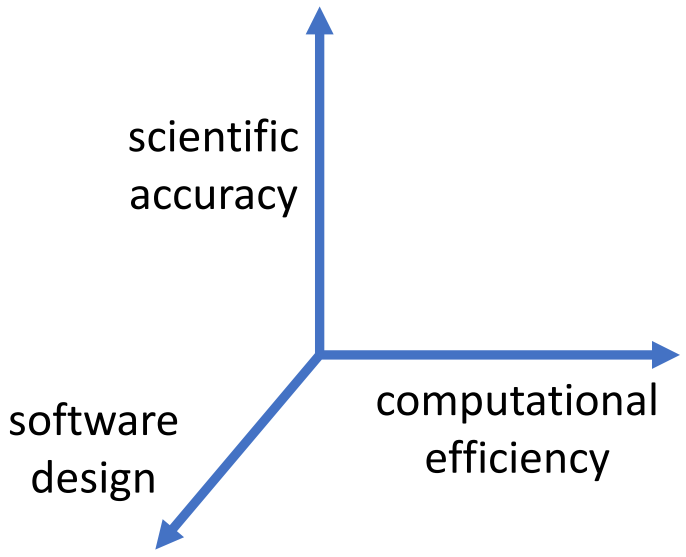
Science

Big Data



#1 Improve Model Performance

- Evaluate models (algorithms, grids, integration) for scientific accuracy AND computational efficiency
- Incorporate parallelism at all levels
- Minimize inter-process communications
- Improve I/O capabilities
- Re-architect, rewrite model



Weather Prediction Models

- dynamics -

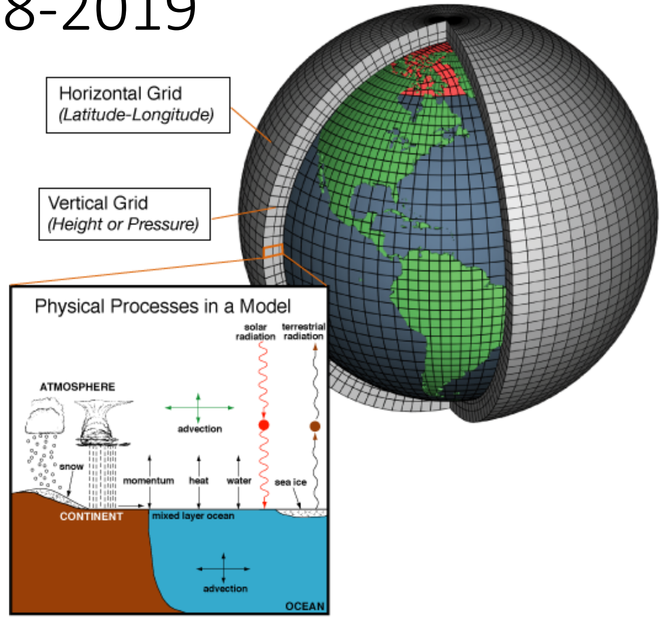
- What are the best models, approaches for emerging HPC
  - Algorithms, grids, time-step, physics, etc.
  - Computational efficiency, scalability, portability

Model Type	Horizontal Grid	Time-Step	Staggering	Models
Finite-volume	Cube-sphere	SISL	A-grid, C-grid, D-grid	FV3GFS
Finite-volume	Icosahedral	HEVI	A-grid	NICAM
Finite-volume	Icosahedral	HEVI	C-grid	MPAS, ICON
Finite-element	Cube-sphere	SISL	C-grid	LFRiC
Spectral-element	Cube-sphere	HEVI	No staggering	NUMA, Neptune, KIM
Spectral	Polar	HEVI	No staggering	IFS, GFS

G.Mengaldo, et.al.,Current and Emerging Time-integration Strategies in Global Numerical Weather and Climate Prediction, https://doi.org/10.1007/s11831-018-9261-8

Dwarf Development with GeoFLOW

Duane Rosenberg, Bryan Flynt, NOAA ESRL, 2018-2019



- GeoFLOW is an application framework to simplify dwarf development in order to evaluate **computational efficiency vs scientific accuracy** of various approaches
- C++ objects to define communications, grid, discretization & time-stepping operators
- Evaluate for 1-3KM global resolution on CPU, GPU, ARM, …

<u>Icosahedral Finite Volume (IFV)</u>	<u>Spectral Element (CG, DG)</u>
• Low order/low accuracy	• High order/high accuracy
• 2D, 3D control volumes	• 2D, 3D elements
• Icosahedral grid	• Cube-sphere grid
• Deep communication	• Shallow communication
• staggered (Arakawa) centering	• Un-staggered centering
• Explicit time step	• Explicit & semi-implicit time step

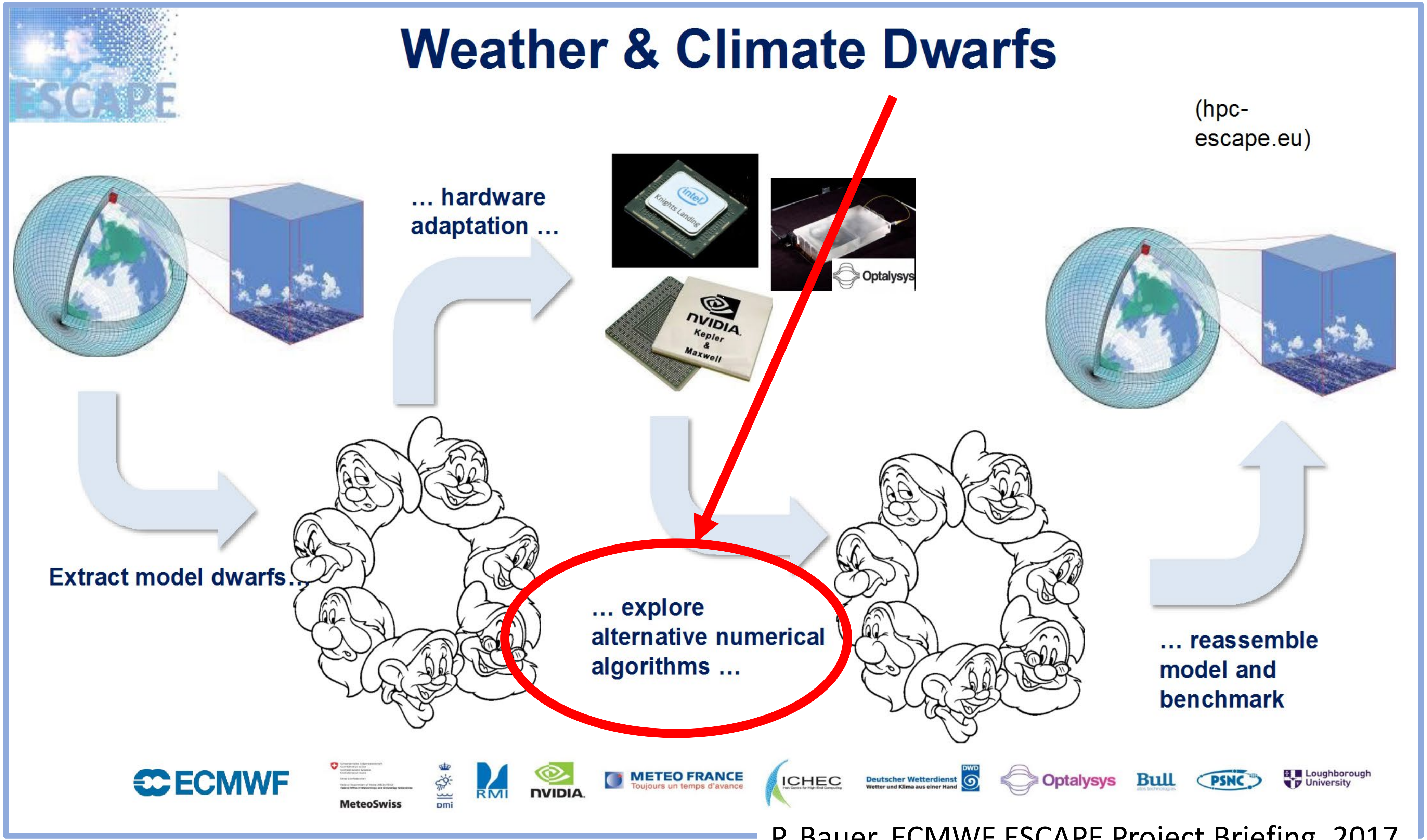
<u>Focus Areas</u>

Advection

+ Convection

+ Radiation

+ …



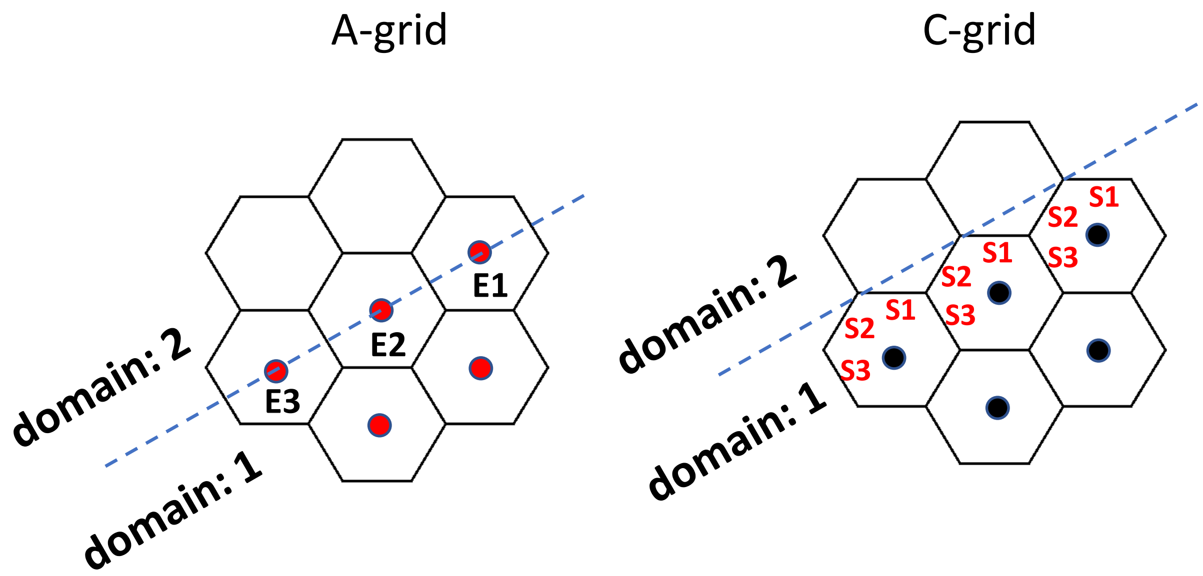
P. Bauer, ECMWF ESCAPE Project Briefing, 2017

Shallow Water Dwarf: A-grid versus C-grid staggering

Yonggang Yu, Ning Wang, Jacques Middlecoff, NOAA ESRL, 2018-2019

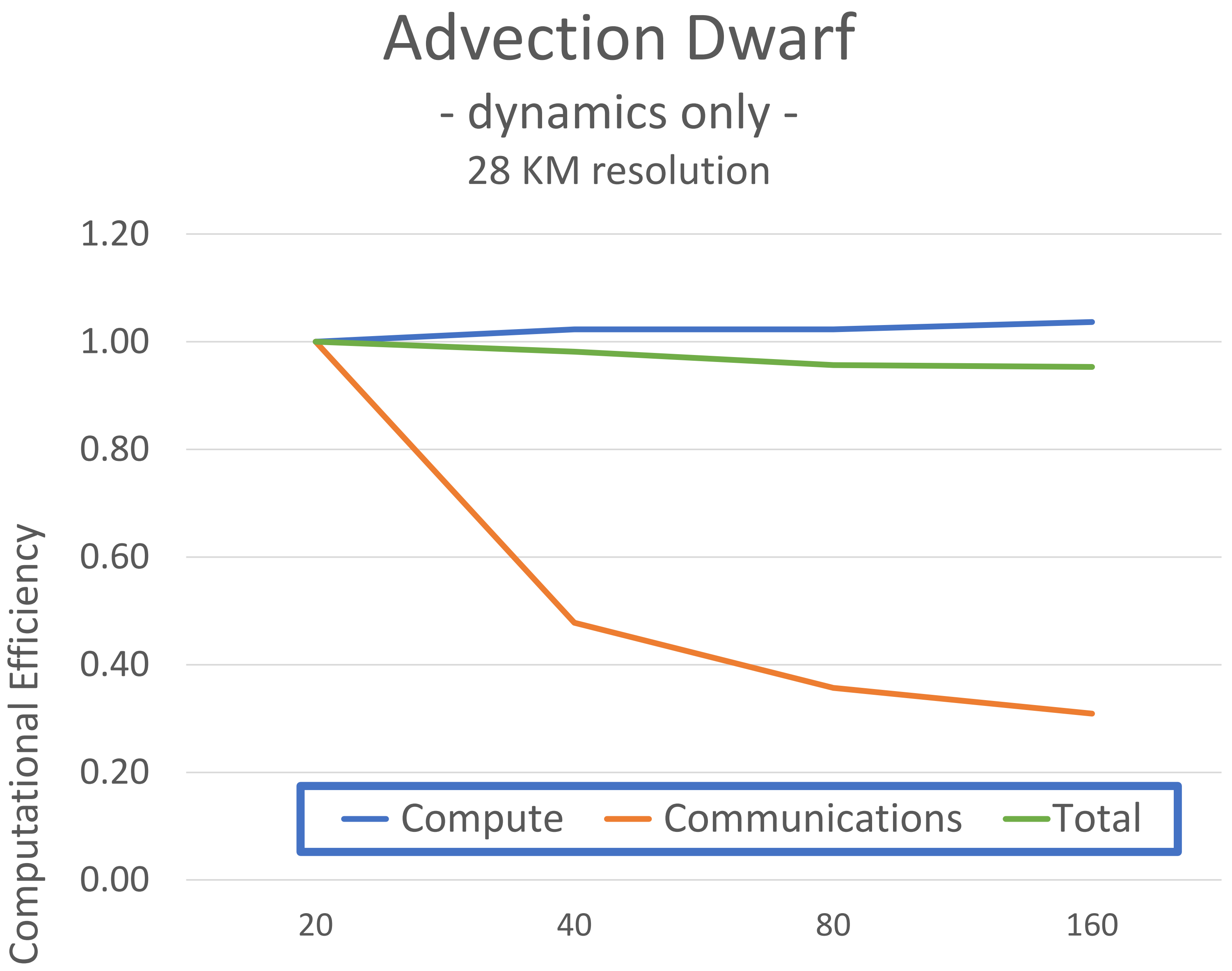
Evaluate performance, scaling and scientific accuracy

- Develop shallow water model for A-grid and C-grid with identical design, grid construction, optimizations, …
- Replicate published dynamical core idealized test results for A-grid (NICAM), C-grid (MPAS)
- OpenMP, OpenACC, MPI parallelization
- Performance & scaling comparison at 1-3 KM resolution
    - CPU, GPU, …



Scaling Patterns

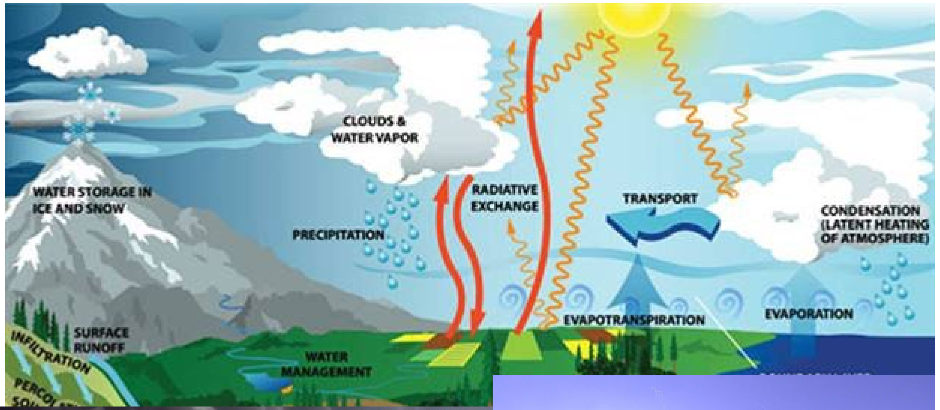
- Computation
  - Good parallelism
  - Icosahedral grid
  - Efficient algorithm
- Communications
  - Minimal frequency
  - Low data volume
  - Some overlapping



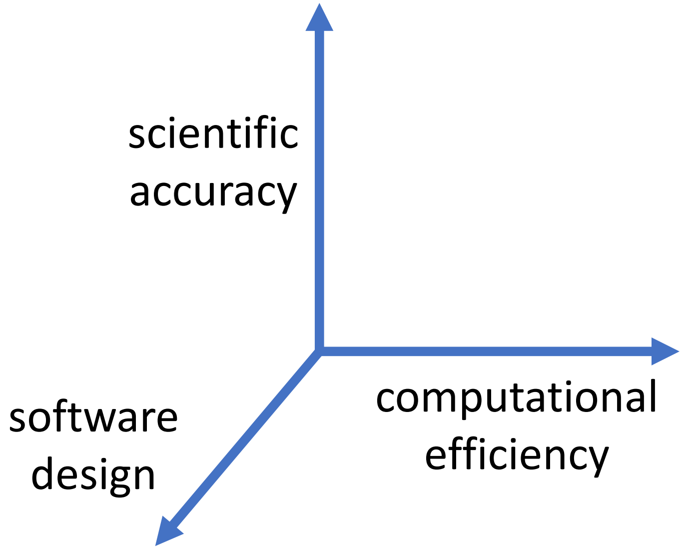
Weather Prediction Models

- physics -

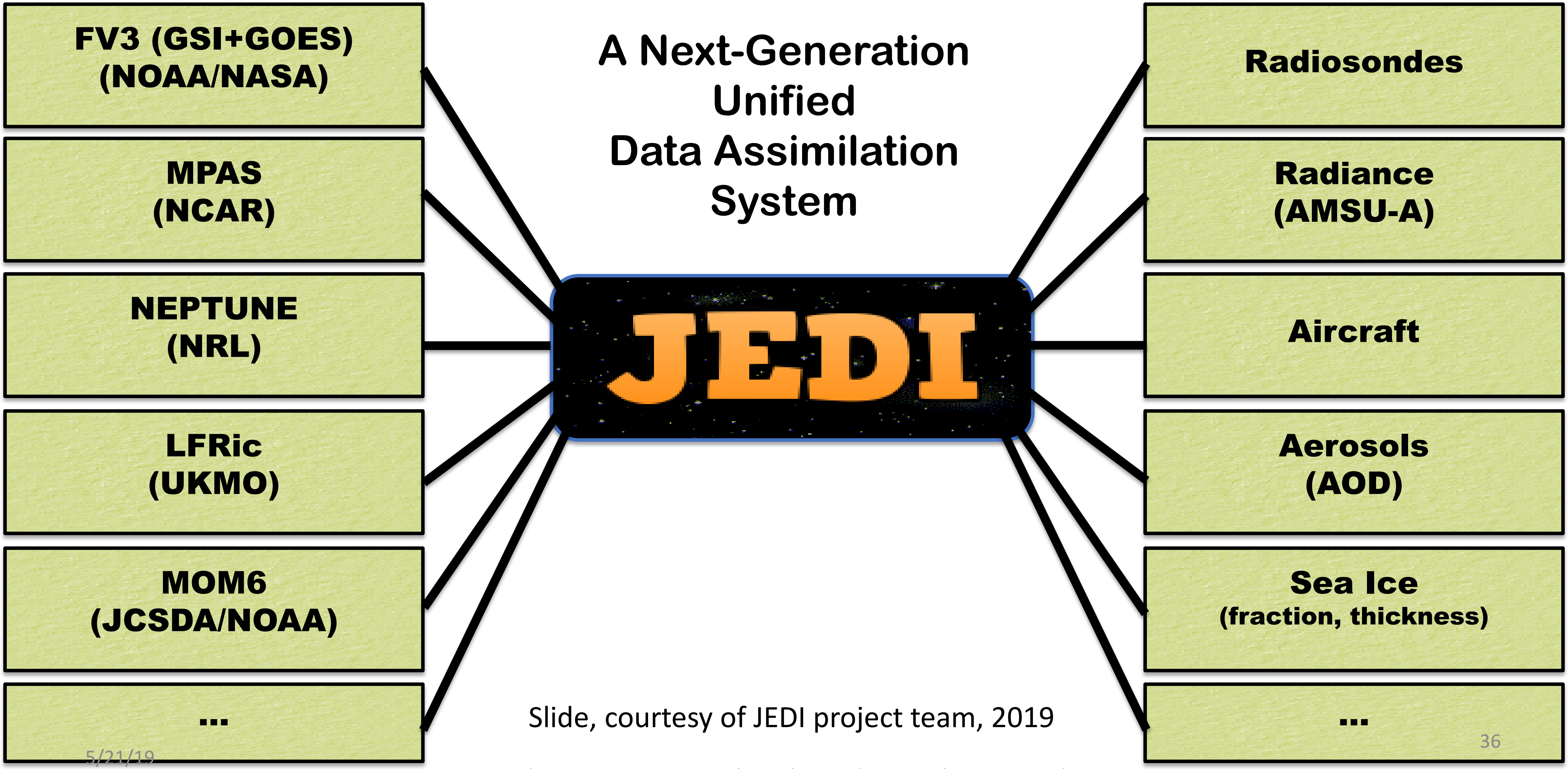
- 2/3 of model source code
  - Convection, radiation micro-physics, surface & boundary layers, gravity & orographic wave drag
- Computationally expensive, complex interactions
  - less parallelism than dynamics
  - **Good potential for ML / DL (~100X faster)**
- Combine physics & dynamics
  - Radiation + dynamics
  - Convection + dynamics



#2 Improve Data Assimilation Performance



Joint Effort for Data Assimilation Integration (JEDI)

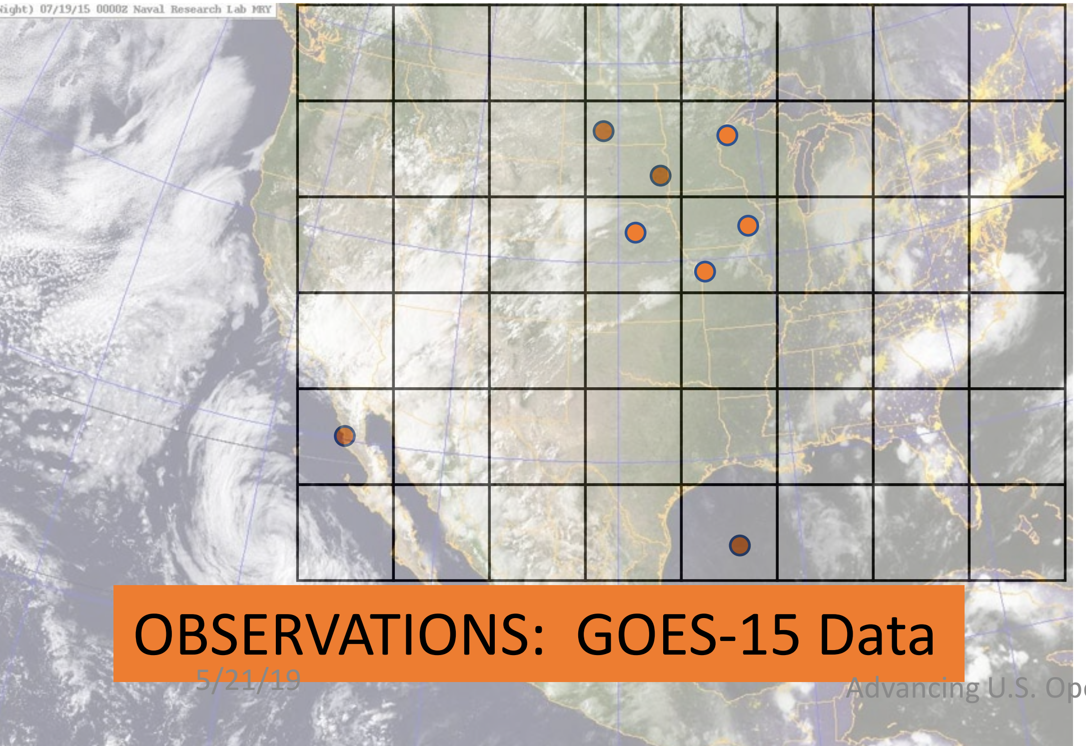


Slide, courtesy of JEDI project team, 2019

Assimilation: intelligent thinning

- Select more observations in one area (severe weather) and less in another (clear sky) determined by:
  - Ensemble uncertainty
  - ML feature recognition

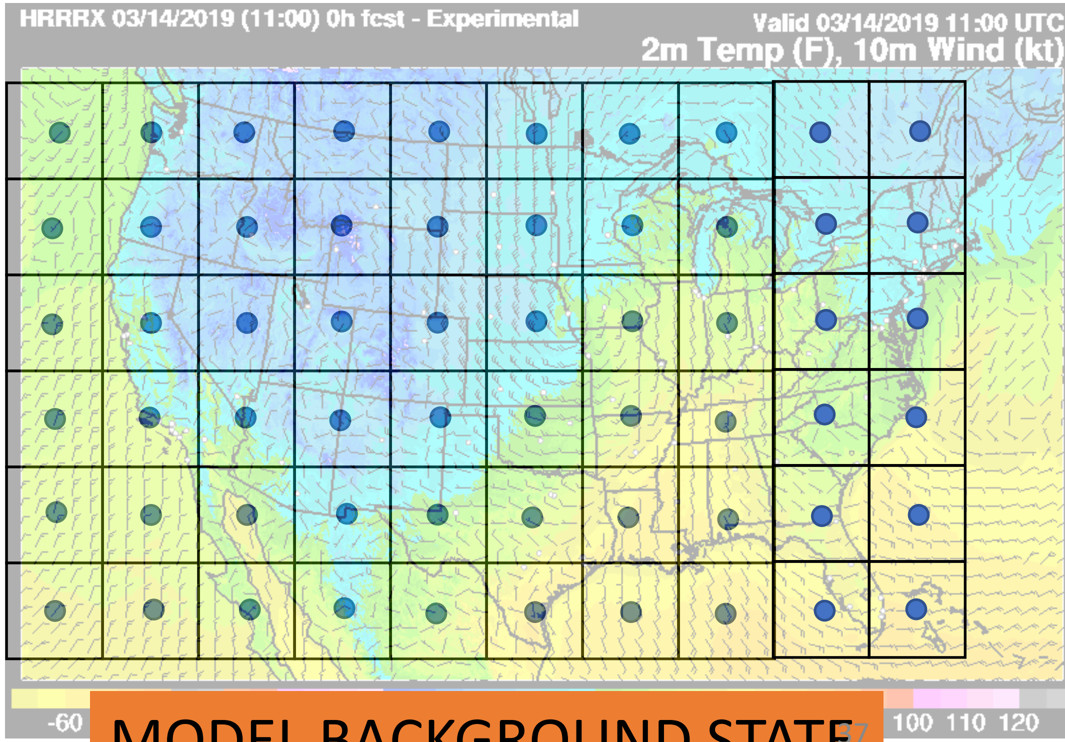
GOES-15: 4 KM resolution IR, 1 KM visible



OBSERVATIONS: GOES-15 Data

update

HRRR: 3 KM resolution, 2M temperature



MODEL BACKGROUND STATE

Feature Detection – Typhoons

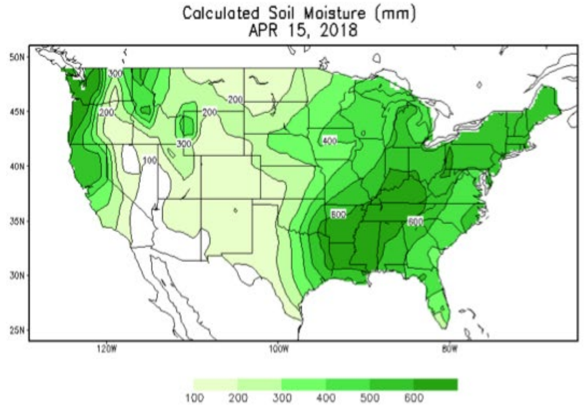
Christina Kumler, Jebb Stewart, NOAA ESRL/GSD, 2018-2019

- Identify typhoons from satellite data
  - Accurate Identification
  - Early detection – prior to formation
- Training - 6 years of data
  - Model output, satellite
  - 11.5 hours (CPU) - 3 minutes (GPU)
  - 5 weeks (CPU) - 3 hours (GPU)
- Inference
  - 1 second (CPU) - 0.04 seconds (GPU)

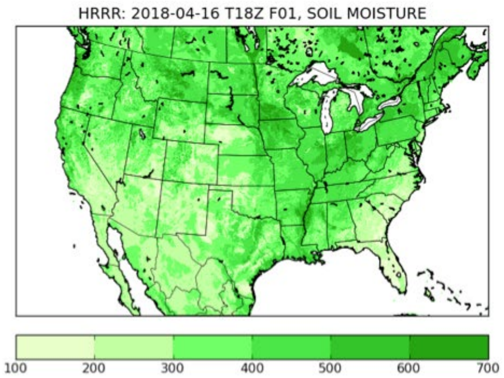
Use of Machine Learning for Improved Initial Soil Moisture State in RAP/HRRR

Isidora Jankov, Jebb Stewart, Lidia Trailovic, NOAA ESRL/GSD, 2018-2019

CPC



HRRR



- soil moisture field from CPC and HRRR for April 15, 2018
- similar features in the two data sets
- over Southeast U.S., CPC has higher values with a spatial pattern not present in HRRR
- potential room for improvement in HRRR representation of soil moisture.

Improvement of RAP/HRRR initial soil state field by using ML will be performed in two steps:

1) improve correlation between observed surface variables and soil state (currently used correlation is empirical and based on limited number of case studies)
2) 2) “nudge” the estimated soil moisture state by utilizing 10.3 micron channel from GOES-16/17 for the CONUS with a spatial resolution of 2 km and temporal resolution of 5 minutes

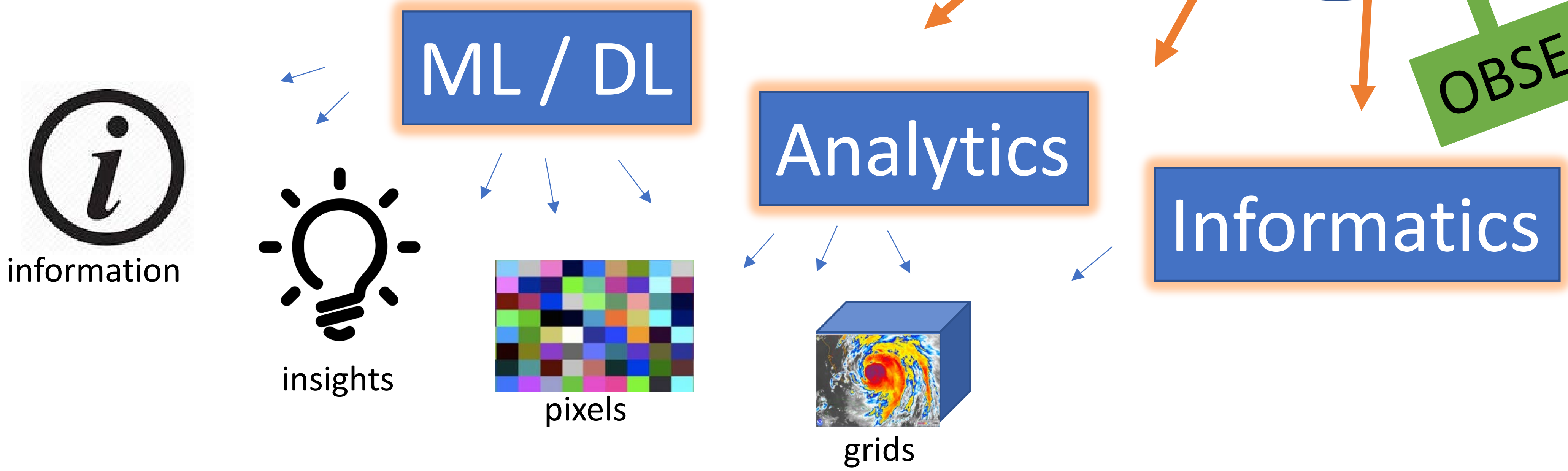
The effort will facilitate:

- more general use of the high-resolution GOES-16/17 ABI data set in data assimilation
- expansion of ML use in areas of Numerical Weather Prediction (NWP) and data assimilation.

#3 Get Data to End-Users

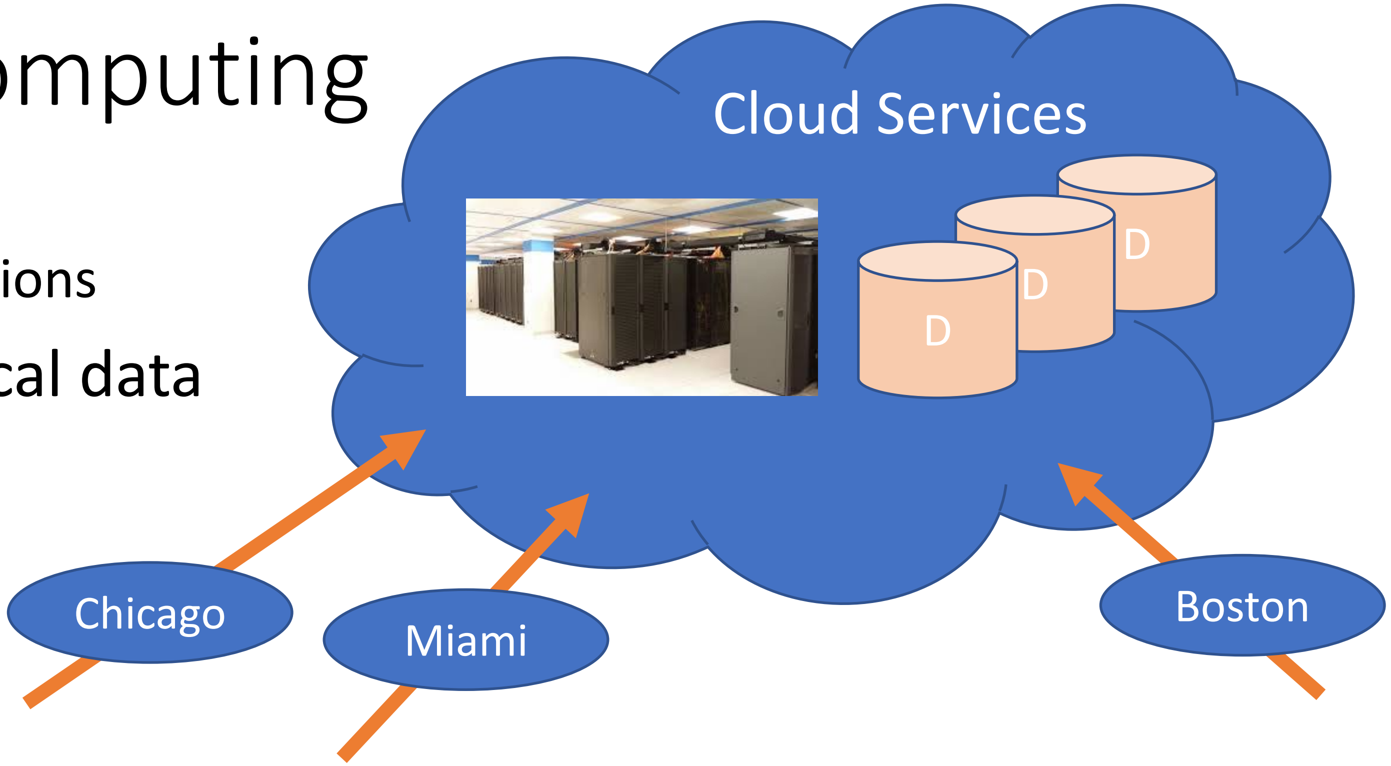
Big Data Handling

- Data is too big to move
  - Co-locate HPC & data
  - On-demand access
  - ML/DL driven analytics



5G Wireless & Edge Computing

- Local compute on local data
  - Sensor networks, local observations
- Global compute on merged local data



Wireless Network	4G	5G
Average Data Rate	25 Mb/s	100 Mb/s
Peak Data Rate	150 Mb/s	10,000 Mb/s
Latency	50 ms	1 ms
Connection Density	2000 cu km	20,000 cu km

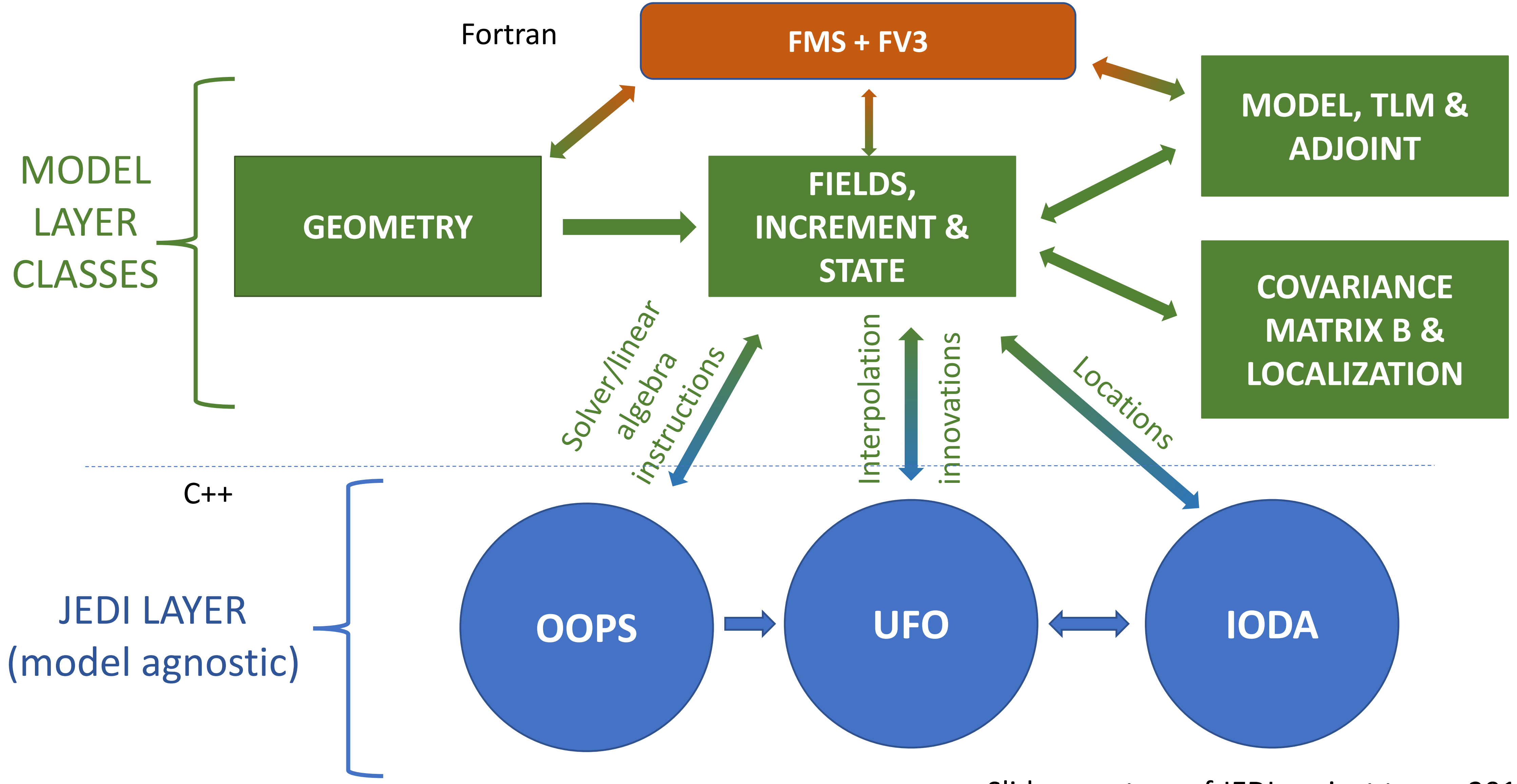
#4 Improve Software Architecture and Development Process

Source code
Scripts
Validation
Documentation

Portability
Performance
Maintainability
Extensibility

Design
Develop
Test
Commit
Refactor

JEDI System Software Architecture

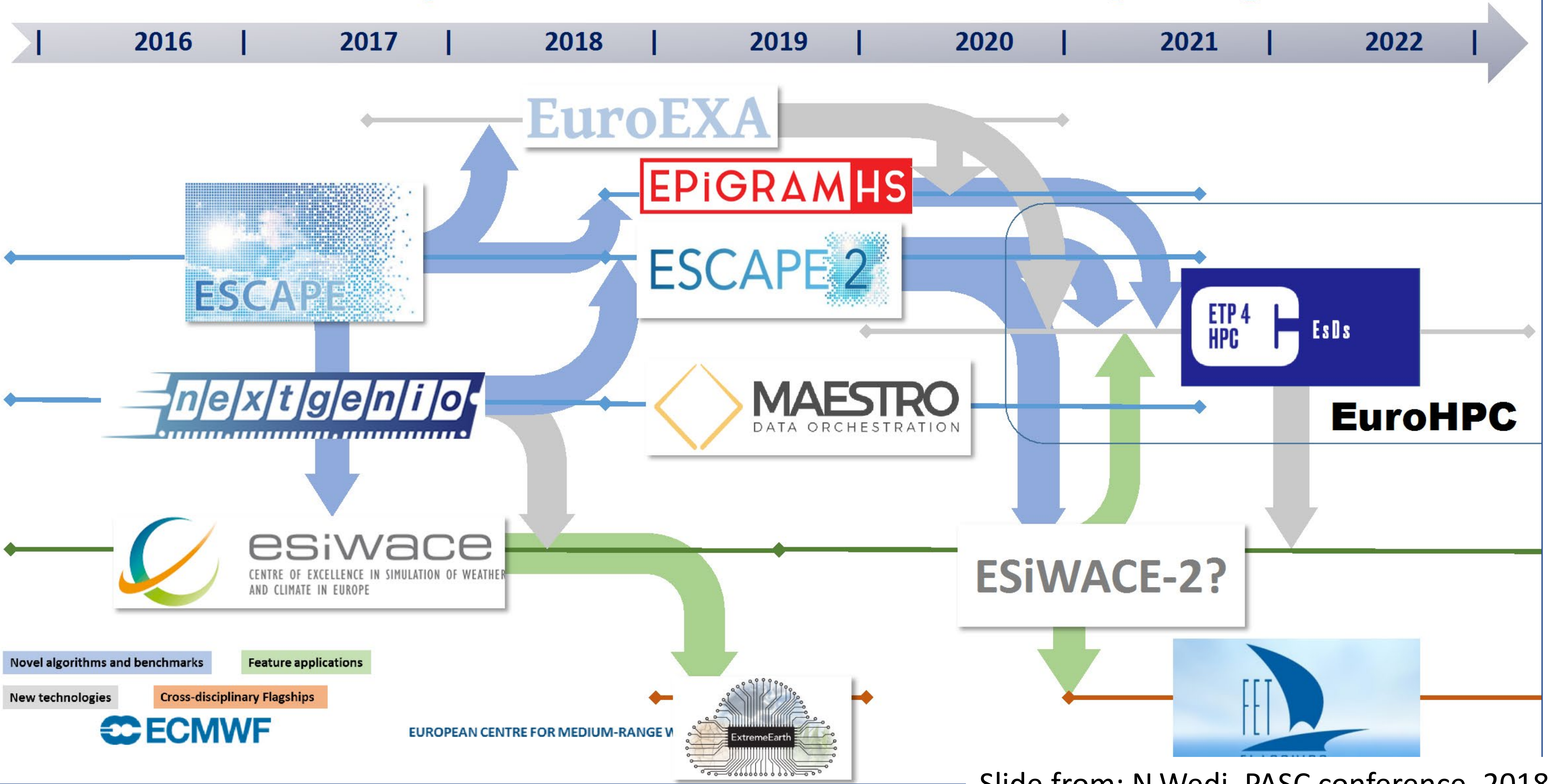


Slide, courtesy of JEDI project team, 2019

Conclusions

- Major challenges in advancing weather prediction capabilities
  - Modeling, computing, data handling, software
- New technologies, approaches needed to move beyond current capabilities
  - Traditional HPC is not sufficient anymore
  - Machine learning, cloud computing, analytics, new HPC
- GSD has begun exploratory development
  - Machine Learning, Cloud computing, GPU computing
  - Quantify scientific accuracy and computational efficiency
  - Tools to deliver information, insights, pixels, grids

(ECMWF) Roadmap for weather & climate computing



Slide from: N.Wedi, PASC conference, 2018

Final Thoughts – A Common Goal

- Strong commitment to FV3GFS
  - Analyze & improve CPU performance
  - Further work on adaption for GPU
  - Containerize, utilize cloud computing
  - Improve netCDF I/O
- Sustained commitment to longer-term research that enables future operational prediction capabilities
  - Modeling, Assimilation, Big Data, Cloud, AI, new HPC