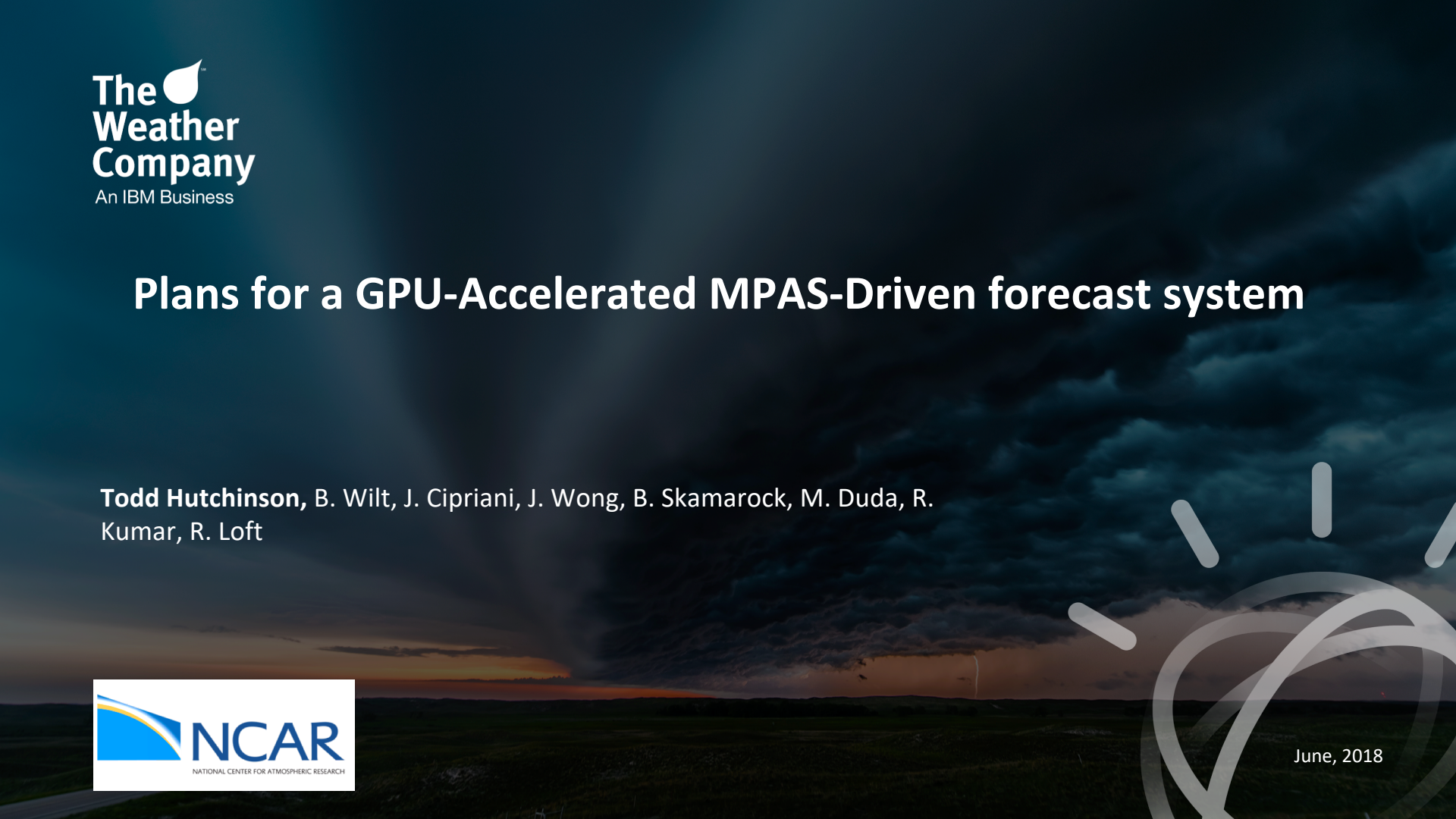


Plans for a GPU-Accelerated MPAS-Driven forecast system

Todd Hutchinson, B. Wilt, J. Cipriani, J. Wong, B. Skamarock, M. Duda, R. Kumar, R. Loft

June, 2018

Topics

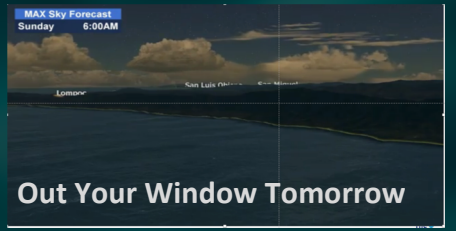
TWC NWP Program Today

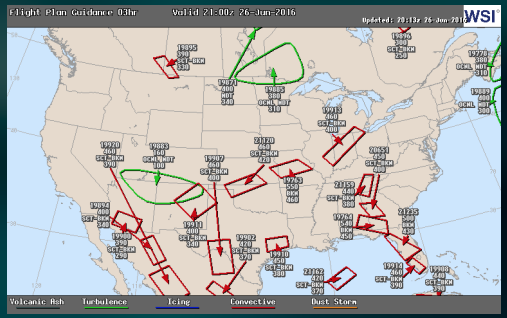
2019 Plans

GPU Acceleration

TWC NWP: Program Now

- Utilize latest NWP technology to aid in delivering accurate weather forecasts
- Provide input to TWC Forecast-On-Demand system, a multi-model blend that provides forecasts to billions of users
- Deliver custom products derived from the large suite of data available from locally run NWP system
  - e.g. aviation turbulence products, simulated “weather” for TV
- Continual Development since 2000



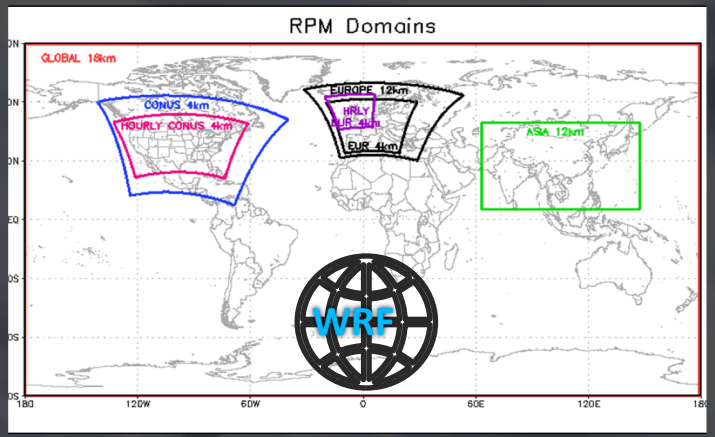


TWC NWP: From WRF to MPAS

Weather Company RPM → **Deep Thunder**

WRF Domains

12-72 hr forecasts



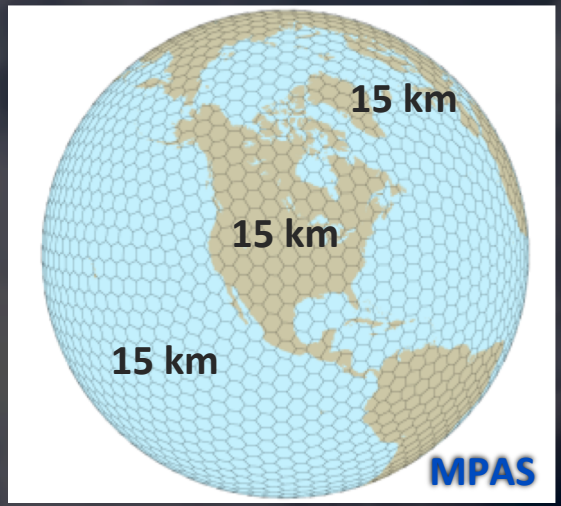
Operational for many years

Global WRF retirement 2018

Regional retirement 2019

MPAS Uniform Mesh

72 hr 4x per day



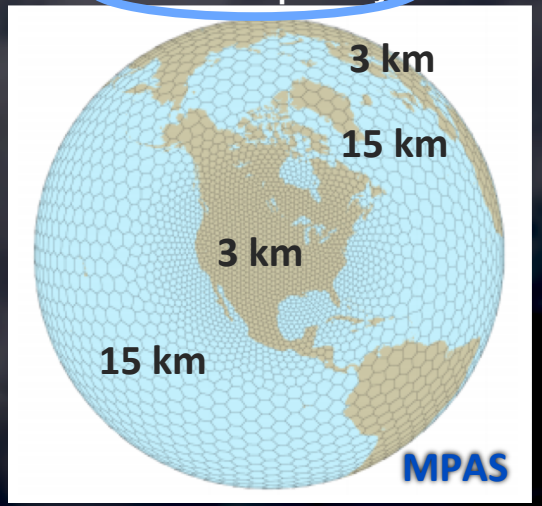
Operational Now

Product Integration Ongoing

MPAS Variable Mesh

72 hr 4x per day

18 hr/24x per day

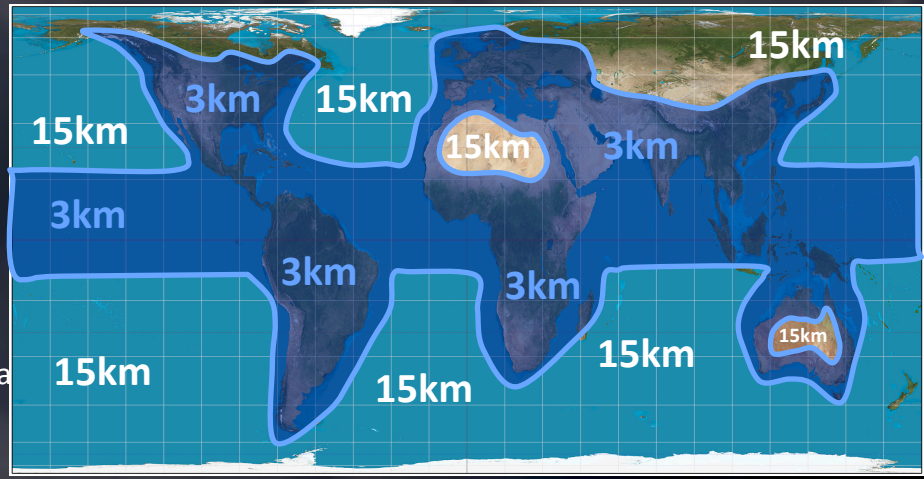


Operational 2019

2019 Plans: Rapidly-Updating System

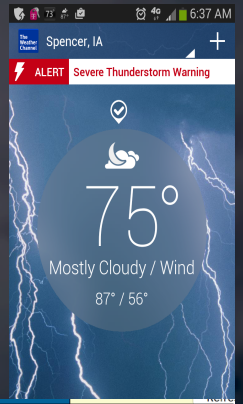
- 0-20 hr forecasts
- Convective-Permitting, where it's *important*
  - Areas of Business Interest: Generally Populated areas
  - Where convection influences global circulations (tropics)
- Variable Resolution Mesh enables running at convective scales, but still control cost
- Plan:
  - 15km over much of world
  - 3km over populated areas, and where convection influences globa circulations (e.g., tropics)
- Challenges:
  - Spanning convective "no-man's land" with scale-aware Tiedtke

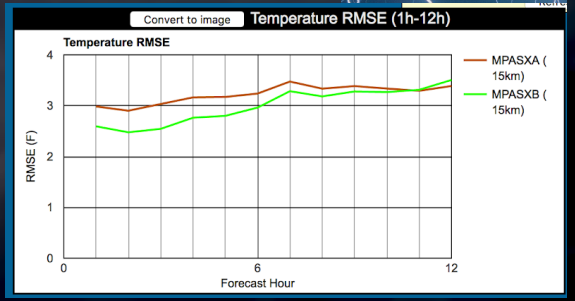
MPAS *Notional*
Variable Mesh
20 hr/24x per day



2019 Plans: Data Assimilation

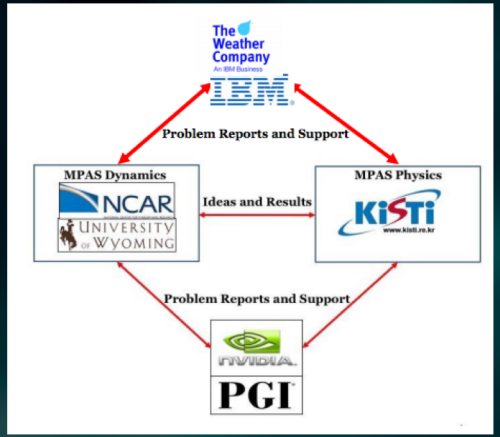
- Global, Rapidly-Updating requires robust Data Assimilation system
  - We can no longer depend on NCEP analyses for initialization
- GSI data assimilation, interfaced to MPAS, will serve as the framework (see Cipriani, 8.3, Thursday 9:00)
- System will be cycled, hybrid EnVar
- Exploring use of novel observations:
  - 100's millions of daily cell phone pressure obs
    - University of Washington is learning
  - Very high-frequency aircraft data (temperature, pressure, wind)

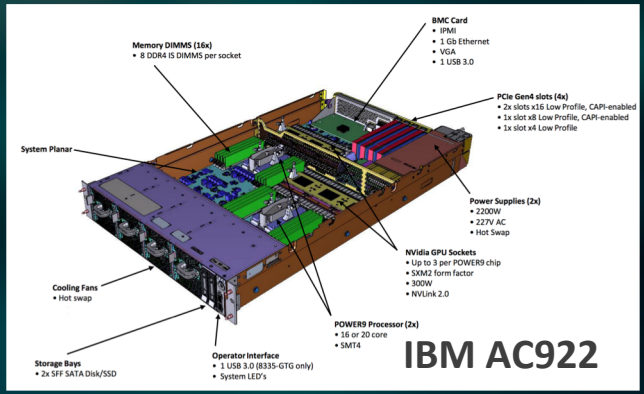




GPU: Accelerating MPAS using GPU

- Weather and Climate Alliance
- Partnership with
  - NCAR
  - Korea Institude of Science and Technology Information
  - nVidia
  - The Weather Company
  - IBM Research
- WACA Objective: Accelerate MPAS using GPU
- TWC Goal: Accelerated Weather Forecasts





GPU: Using OpenACC

- nVidia's PGI compiler supports OpenACC directives
- Directives/kernels can be generated automatically (see https://github.com/NCAR/Kgen)
- But, best performance requires manual refinement on Kgen implementation

Simple Port: Can be done with automated tools

```
subroutine foo
!$acc kernels
do k = k_begin, k_end
    do i = i_begin, i_end
        A(i,k) = B(i,k) + C(i,k)
    end do
enddo
!$acc end kernels
end subroutine
```

Complex Port: Requires Manual Intervention

Up to 50 % faster!

```
subroutine foo
!$acc parallel loop gang vector collapse(2)
do k = k_begin, k_end
    do i = i_begin, i_end
        A(i,k) = B(i,k) + C(i,k)
    end do
enddo
!$acc end parallel
end subroutine
```

GPU: Current Porting Status

- Phase 1: Dry Dynamics Port (complete)
- Phase 2a: Physics Port to GPU(in-progress)
  - Cloud Fraction (complete)
  - YSU PBL (in-progress)
  - WSM6 Microphysics (in-progress)
  - Scale-Insensitive Tiedtke (not started)
  - Monin-Obukhov surface layer (not started)
  - New Tiedtke scale-insensitve Convection (not started)
- Phase 2b: CPU-based Processing
  - Radiation (RRTMG SW + LW) to runs time-lagged on CPU (in-progress)
  - NOAH Land Surface Model runs on CPU (in-progress)
  - SIONlib I/O subsystem runs asynchronously on CPU (in-progress)
- Phase 2c: Integration/Scaling/Other

GPU: MPAS GPU Performance Results

- "Dry Dynamical Core" Performance (No physics)
- Comparison: 1 GPU vs 2 Broadwell CPU
- MPAS 5.2, 56 Vertical Levels

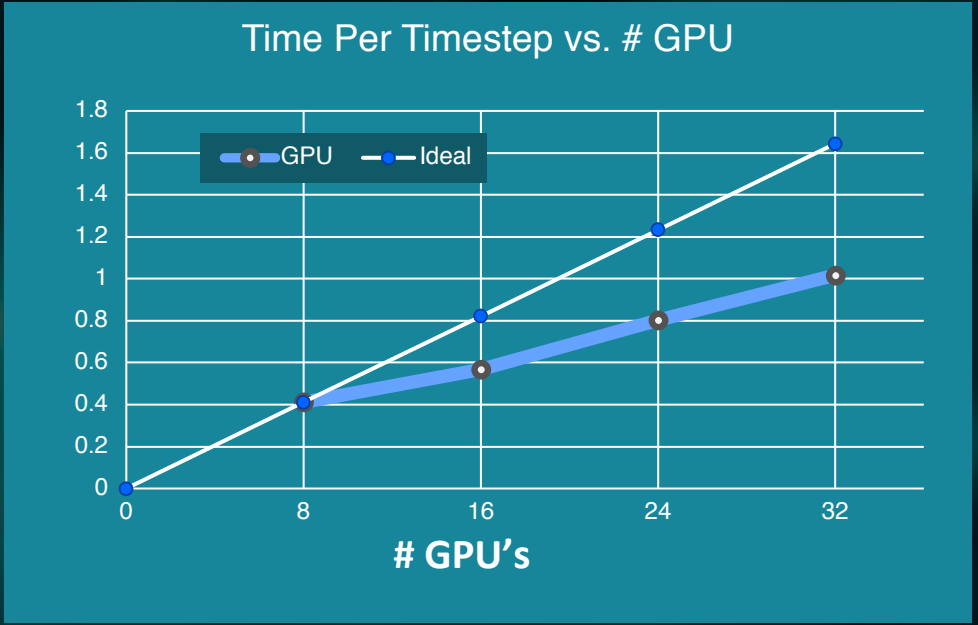
Early Results are Promising!

Dataset		Broadwell (Fully Subscribed, OpenMP Enabled, Intel compiled, Base code)	P100 with Haswell(1 GPU, PGI compiled, OpenACC code)	V100 with Haswell (1 GPU, PGI compiled, OpenACC code)	P100 with Power8(1 GPU, PGI compiled, OpenACC code)	Speedup Broadwell vs P100	Speedup Broadwell vs V100
120 Km (40K)	SP	0.40	0.28	0.19	0.26	1.54	2.16
	DP	0.88	0.40	0.29	0.35	2.51	2.99
60 Km (163K)	SP	1.90	1.02	0.69	1.01	1.88	2.74
	DP	3.80	1.54	1.12	1.41	2.70	3.40

Taking 40k data points per node for SP, 32.8M grid points (15 Km & 3 Km Locally refined grid) need ~800 Volta GPUs

GPU: Scalability

- MPAS scaling, up to at least 32 GPUs
- Early results—Currently a 40% performance hit due to MPI comms
- Optimization will decrease this performance penalty
- Courtesy Raghu Raj Kumar, Rich Loft NCAR: More info:



http://on-demand.gputechconf.com/gtc/2018/presentation/s8812-an-approach-to-developing-mpas-on-gpus.pdf

Thank you!

Questions