

Random Thoughts on

Data Needs for Post Processing

Peter Neilley
The Weather Company

Post Processing Types

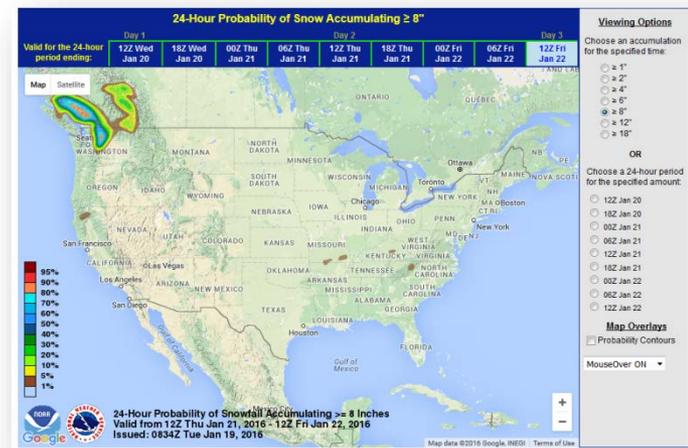
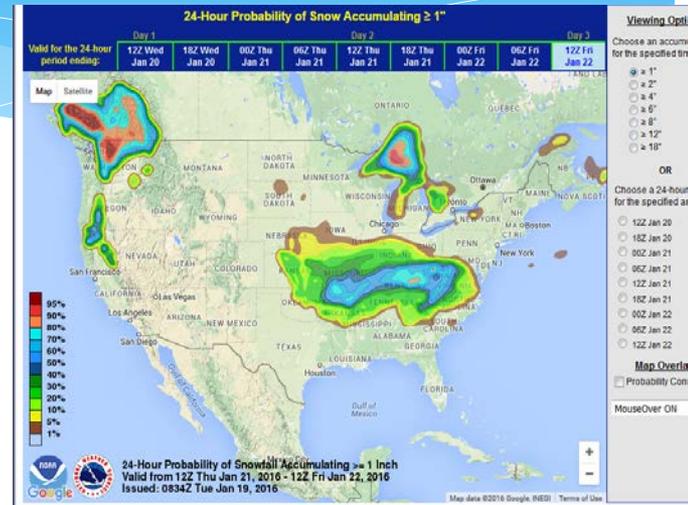
- * **Coupled (dependent) physical models**
 - * Derive new forecast quantities from base model fields
 - * Hydrology, sea state, turbulence, etc.
 - Generally desire native resolution data
- * **Statistical Methods**
 - * New/Improved variables via manipulations
 - * Needs best inputs, but form/resolution application dependent
 - * Generally require lots of data, and that's not enough
 - * Requires “truth”
- * **Hybrid models**
 - * GTG, surge/inundation systems, TWC forecasting

TWC Specific Data Requirements

- * **We are a post processing shop. Hence, most of our (near-term) data needs are upstream of post-processing**
- * **Global (literally) model and truth data**
 - * We (and much of the private sector) serves the world, not just US
 - * We may be data saturated domestically (at least for mean forecasting), considerably short internationally.
- * **4-D native resolution model data, in 4-D**
 - * Most immediate need is improved/consistent time resolution
 - * Hourly (at least) output from every model, member, run
 - * Native vertical resolution desired
- * **Looking forwards, calibrated models and ensemble distributions are of interest**

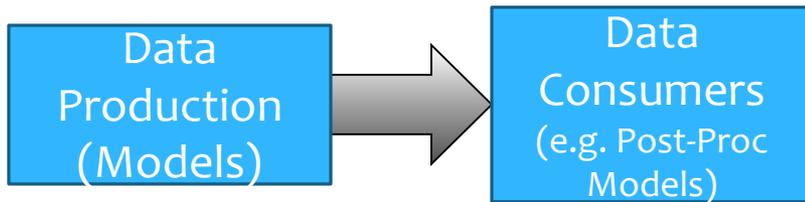
Diversity and Probabilities

- * Current model and ensemble systems are biased and under dispersive
- * Emerging techniques to correct these flaws
- * Data sufficiency is unclear for bias calibration using reforecast data or distribution calibration/reliability based on (multi-model) ensembles
- * **But it certainly is more (and probably much more) than we have today particularly for probabilistic forecasts.** We don't know the answer, as this remains a research initiative for the community.
- * Particularly true for
 - * Rare, high-impact events.
 - * Low-probability, high-impact events (the tail of the distribution)
 - * Nearly any precipitation event.



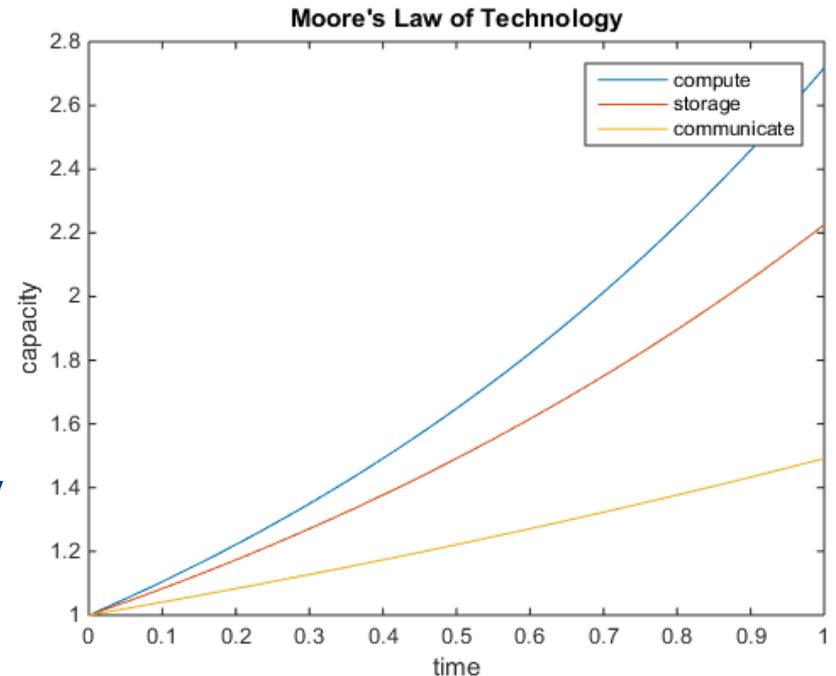
Data Exchange Paradigm

Our national data flow model

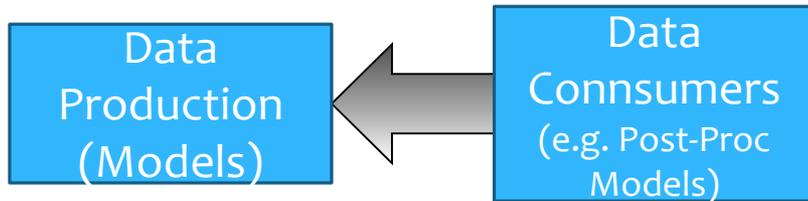


Bring the data to the end uses

Is not scalable and is fundamentally broken



A New Data Flow Paradigm



Bring the end (or middle) users to the data

A new HPC model to serve national interests

High Performance Computer

Models

HPC Internal
Post Proc
Cloud
(any private use)

Some Opportunities and Issues

- * Data and Data Sufficiency
- * Conditional Calibrations
- * Tails of the distributions, not just rare events
- * Role of the human forecasting
- * Off-grid and off-site methods
 - * Downscale the data and then post process, or
 - * Downscale the post-processing and apply locally
- * Community post-processing test bed