An Architecture for an User-Centric Global Interactive Forecast System Testbed

Lloyd Treinish
IBM Thomas J. Watson Research Center
Yorktown Heights, NY USA
lloydt@us.ibm.com
http://www.research.ibm.com/people/l/lloydt

Andrew Barnes
IBM Public Sector
London, UK
andrew.barnes@uk.ibm.com

http://www.research.ibm.com/weather

An Architecture for an User-Centric Global Interactive Forecast System Testbed

- Background and Motivation
- Approach
- Examples
- Issues

Global Interactive Forecasting System

Vision:

- Use the increasing understanding of environmental systems to predict natural disasters and, by mobilising preventative and palliative measures in a timely fashion, reduce loss of life and property
- Consider the guidelines for THORPEX Societal and Economic Applications (SEA) Research

Approach:

Attack the problem from two inter-dependent viewpoints -

- Prediction: improving our ability to model the environment and forecast likely disasters

  … taking requirements from and providing input to …
- Mitigation: running societal and economic models to understand what measures can be taken, and what forecasting information is needed to plan successfully

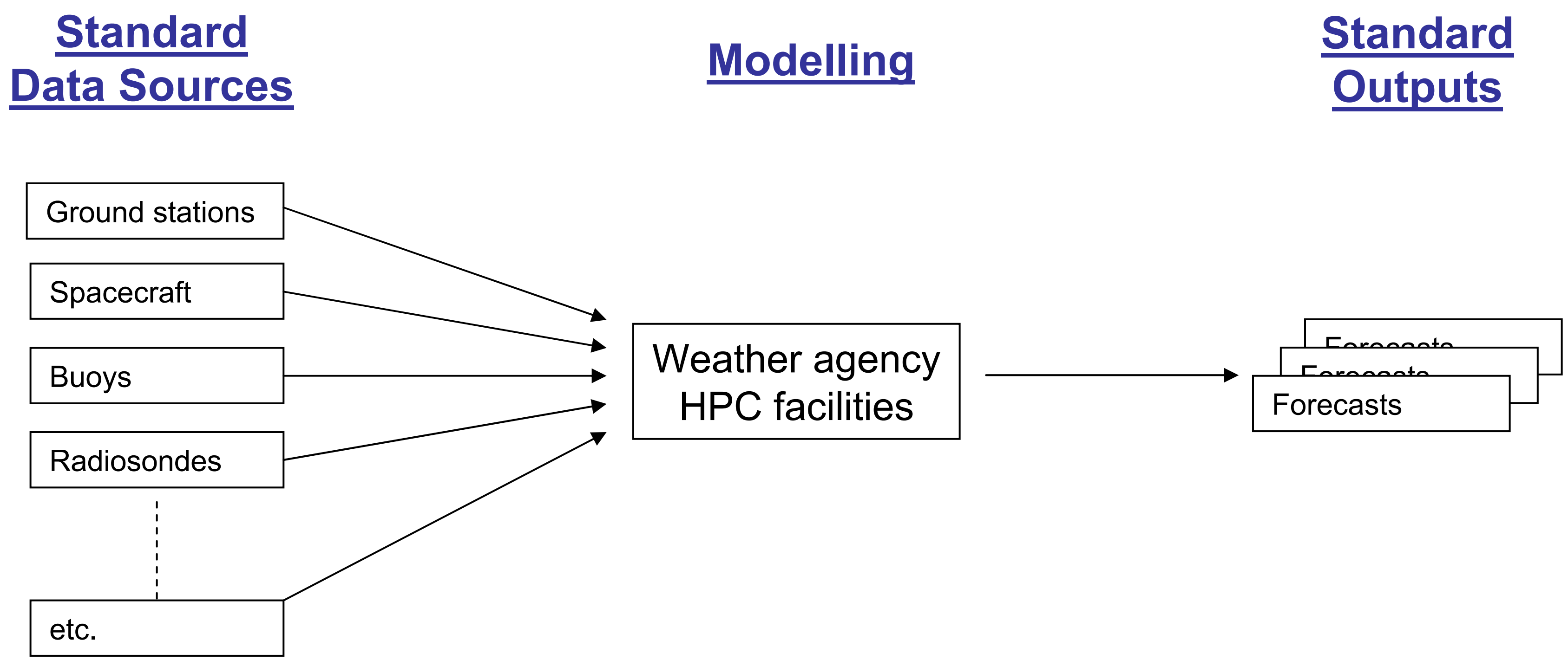
The current exercise has been prompted by an interest amongst the weather forecasting agencies in exploiting their current capabilities to improve the ability to predict disasters by combining and refining forecasts.

Interactive Grand Global Ensemble - Prediction

“Working forwards”

- Starting point is the forecasts produced by the world’s weather agencies
- Need to investigate how these can be combined / refined to produce output needed by the relief (societal and economic) models
- Can the approach be generalised to non-weather-related disasters (e.g., earthquakes/tsunamis)?

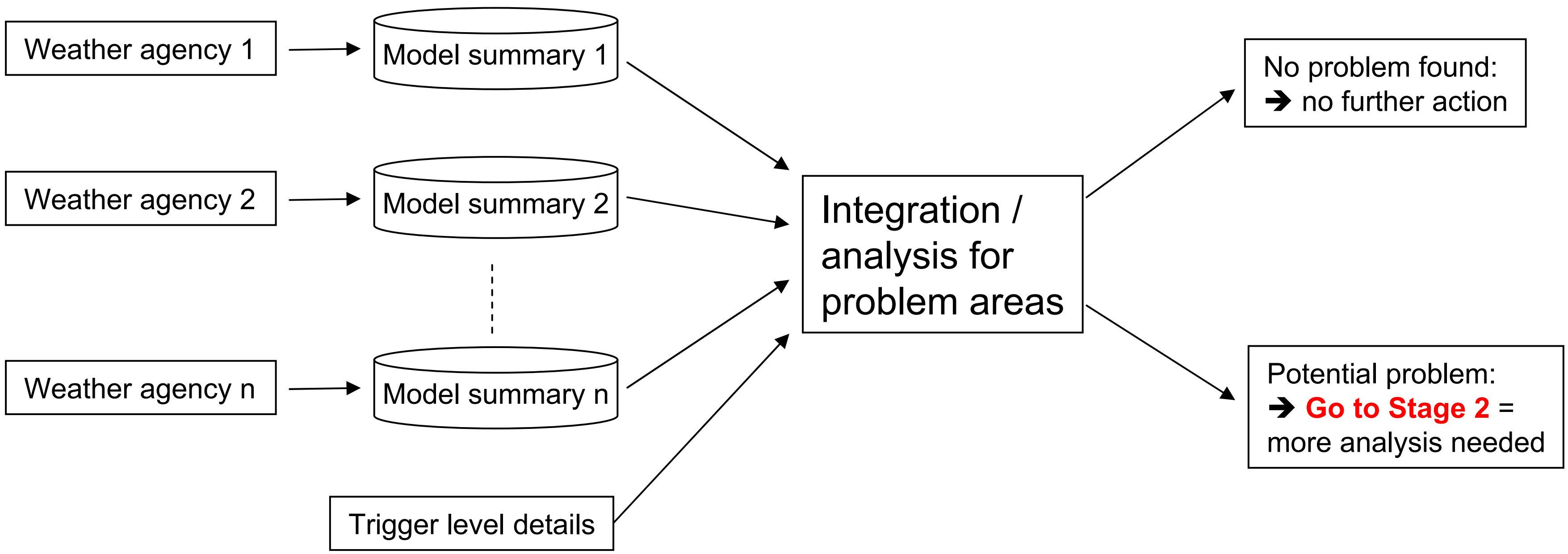
Typical (Traditional) Forecast Production



… carried out on a regular daily schedule by (participating) agencies around the world, but without any integration or aggregation into a multi-model ensemble

Prediction Stage 1 – Problem Identification

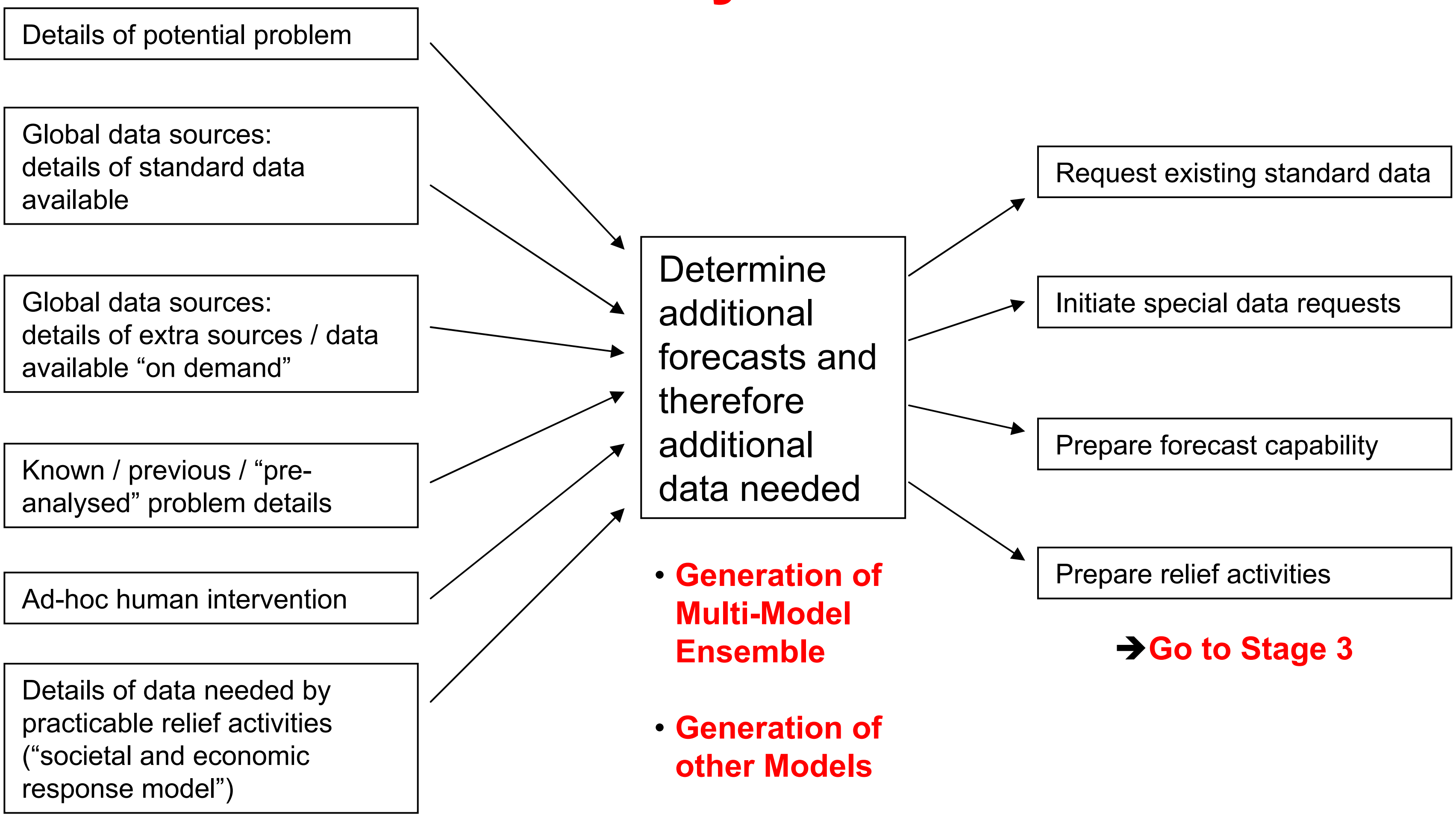
Assumption: **As part of each forecast run, all weather agencies also produce model summaries in a standard format for global integration and analysis**



Issues:

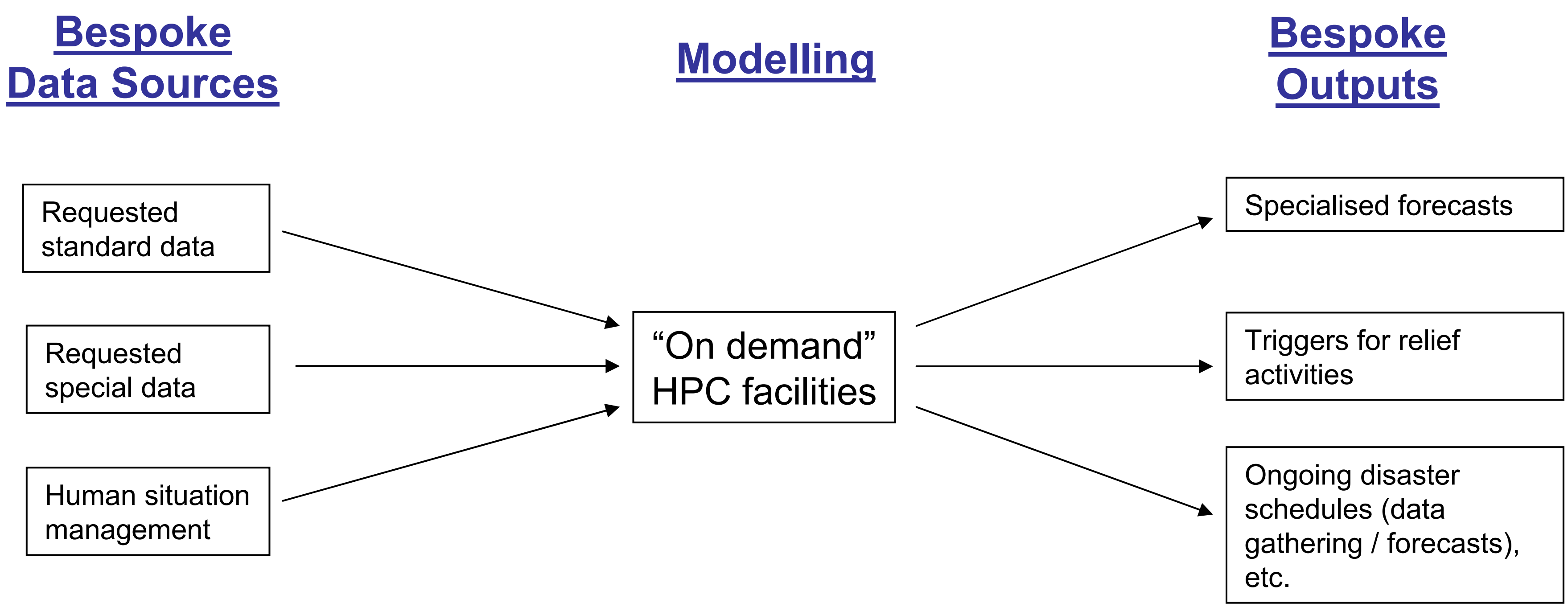
- Can this idea serve as a "prototype" for what a multi-model ensemble could provide
- What is the appropriate format and semantics for the model summary (i.e., not simply providing standard products in GriB) ?
- All participating agencies must agree on model summaries
- Suggests the use of data-level grid among the agencies
- Who provides the compute power needed for integration analysis and eventually ensembling (lead forecast centre or central "on demand" / grid facility) ?

Prediction Stage 2 – Identify Additional Analysis



Who provides the compute power needed for problem analysis (lead forecast centre or central "on demand" / grid facility) ?

Prediction Stage 3 – Analyse Data and Trigger Actions



Who provides the compute power needed for additional forecasts (lead forecast centre or central "on demand" / grid facility) ?

Considerations

- We are not suggesting that the recommended approaches on the ensemble modelling and creation of appropriate multi-model (super)-ensembles are incorrect
- Configuration of the models and their dissemination must avoid an impedance mismatch at the application level, in order to realize a useful and usable solution
- End-user needs must be the driver instead of the requirements for direct dissemination of model data

Interactive Grand Global Ensemble - Mitigation

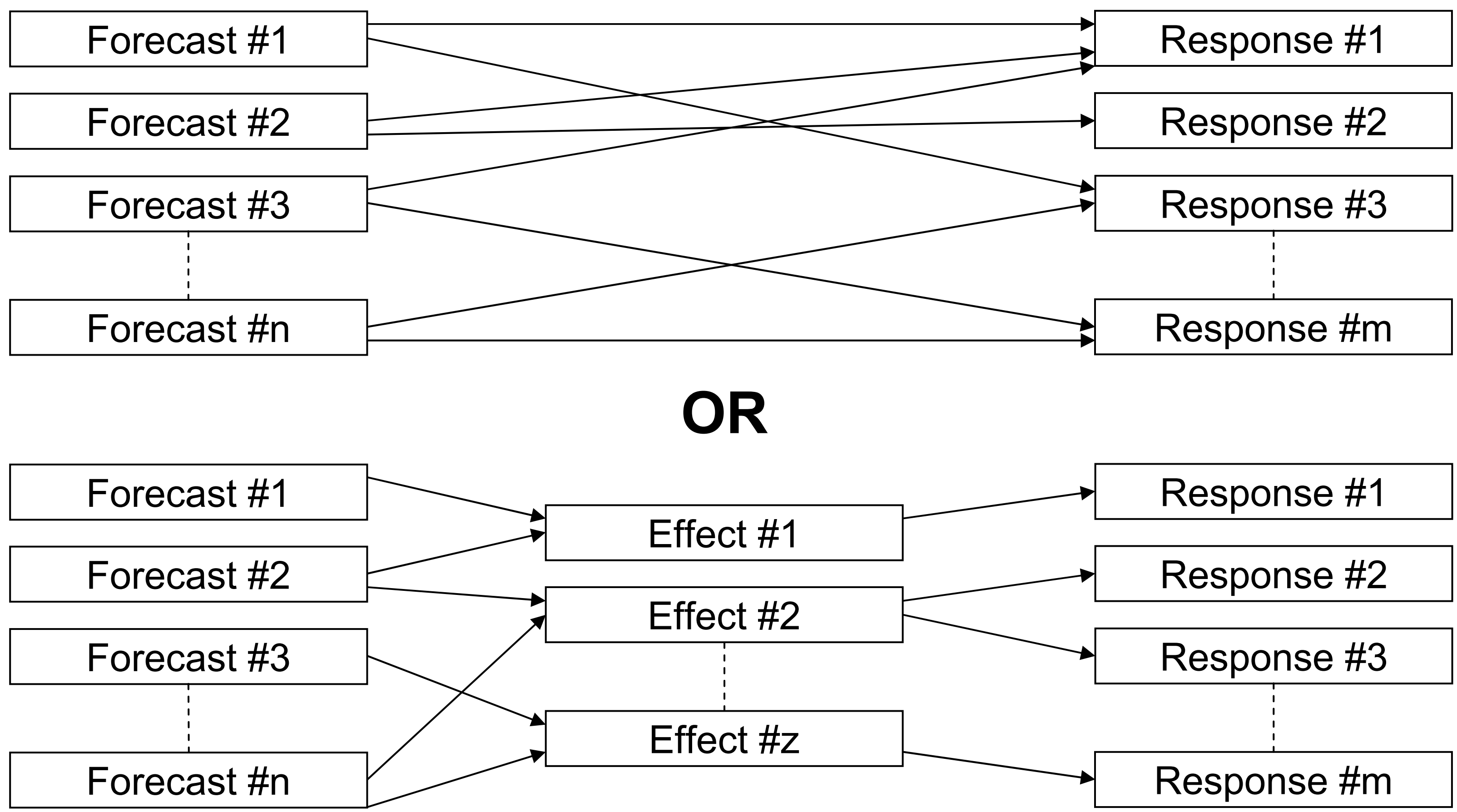
"Working backwards"

- Starting point is information from societal and economic model
- Need to determine what data are needed from bespoke forecast model in order to influence detailed analysis
- Do we need an intermediate "effects" layer, or should it be integrated with the appropriate societal and economic model?
- Do we need a corresponding intermediate "requests" layer, to enable a societal or economic model to request the appropriate forecast information?
- If yes to either question, then these layers also incorporate information about the required temporal and spatial resolution

Do We Need an "Effects" Layer?

Forecast models (Multi-model Ensemble?)

Societal / economic models



… the latter may simplify the effect-response relationship and would more easily allow effects to be handled whether "forecast x" was a weather or an earthquake or any other kind of forecast

Challenges

1. Usability: Are the data relevant and usable for decision makers in weather- climate- and environmentally-sensitive applications?

2. Scalability: Are the data processed and managed data in a timely and cost-effective manner?

- We need to avoid a mismatch between the compelling sophistication of the observing technology and the science/modelling vs. how the data should be utilized

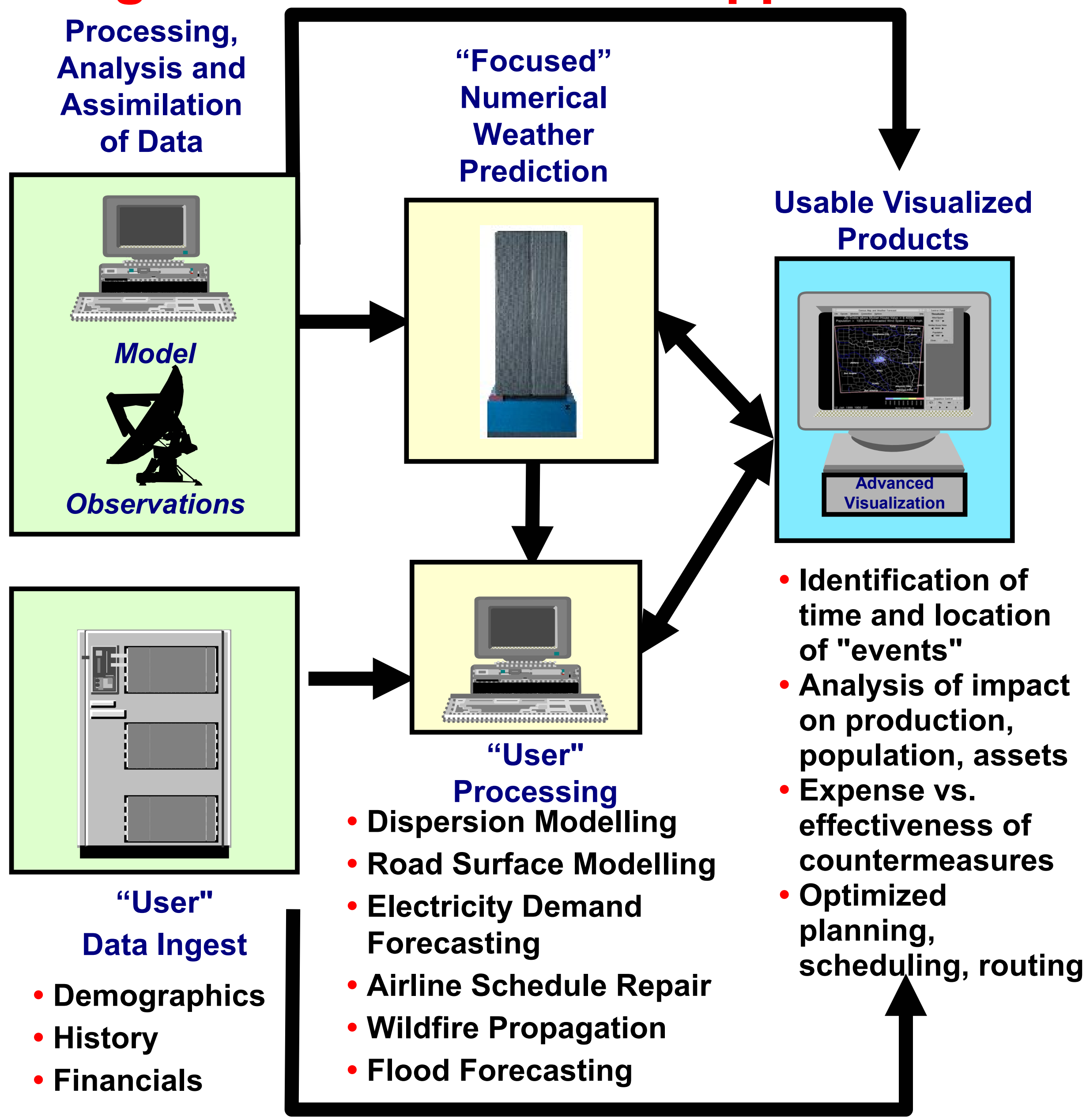
- We need to consider the end-to-end system integration so that the data being generated answer the right questions

Challenges

- Methods are needed to properly validate the utility of forecasting systems for specific societal and economic applications
  - Traditional meteorological (statistical) verification is often inappropriate
  - True end-user metrics are often ill-defined or simply qualitative at present
- What level of capability is “good enough” for an effective deployment from the user’s perspective ?
  - Drive for improvement by meteorological metrics (e.g., 1^{o} C. improvement in temperature forecast) may answer the wrong question
  - Economic impact may be the driver (e.g., O($\$10^5$) to deploy people and equipment for snow removal for major city or airport, even if there is no snow)
  - Need for a “heads-up” for severe weather event, even with opportunity for errors from a meteorological perspective
  - More focused dissemination since standard meteorological products are a clear mismatch with end-user decision makers, whose expertise is in applications and understanding the impact of weather as opposed to meteorology (EPS clearly compounds the situation)

Coupling Model-Based Forecasts for Weather-Sensitive Operations – An "Integrated" User-Centric Approach

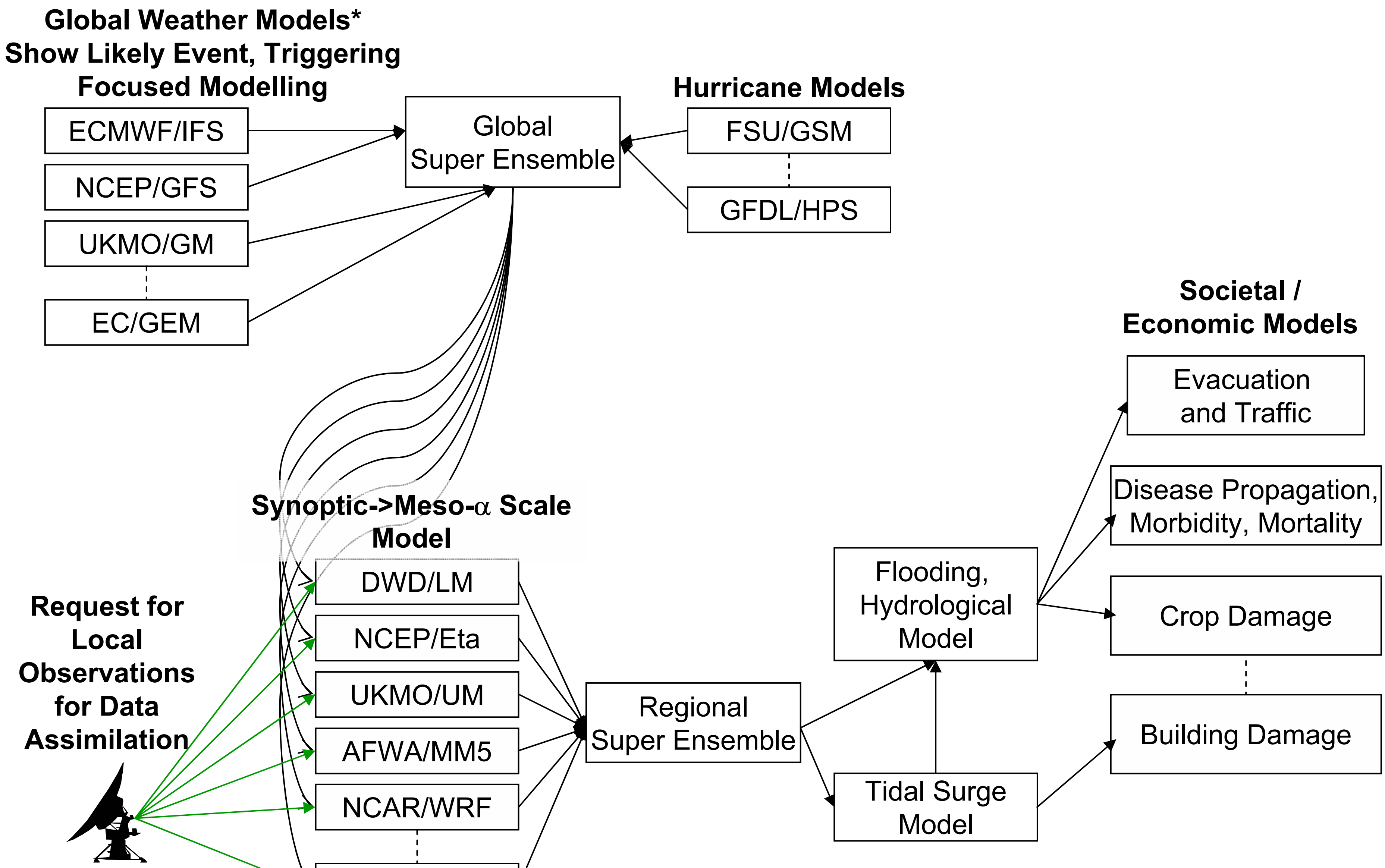
- **Enable proactive decision making affected by weather**
- **Customize & integrate for different users**
- **Provide usable forecast products fast enough to enable timely decisions**
- **Couple to "user" and response processes & models**
- **Past forecasts useful for scenario planning**



An Integrated User-Centric Approach Requires

- "Appropriate" visualization to enable acquired and derived data to be usable
  - Some methods are available, but not widely used nor adequately scaled
  - Understanding of how the relevant data need to be used and why (e.g., human factors concerning how users work and interact)
  - Understanding of how users perceive and interpret visualizations
  - Data must be made relevant for different classes of users using their terminology, and thus, expressible in terms that can be readily understood in real-time without expert interpretation and used with confidence
- Effective coupling to derived modelling and analysis that maps to the end-user problem
  - How to couple and overcome an inherent impedance mismatch is unclear
  - May be a direct match to relevant physical problems (e.g., flooding, dispersion)
  - May be a direct match to relevant operational problems (e.g., crew and equipment optimization [scheduling and routing] that is impacted by environmental factors)

Example – Flood Forecasting in the Indian Subcontinent from a Tropical Cyclone



* Current Operational Forecasting (ensemble and deterministic)

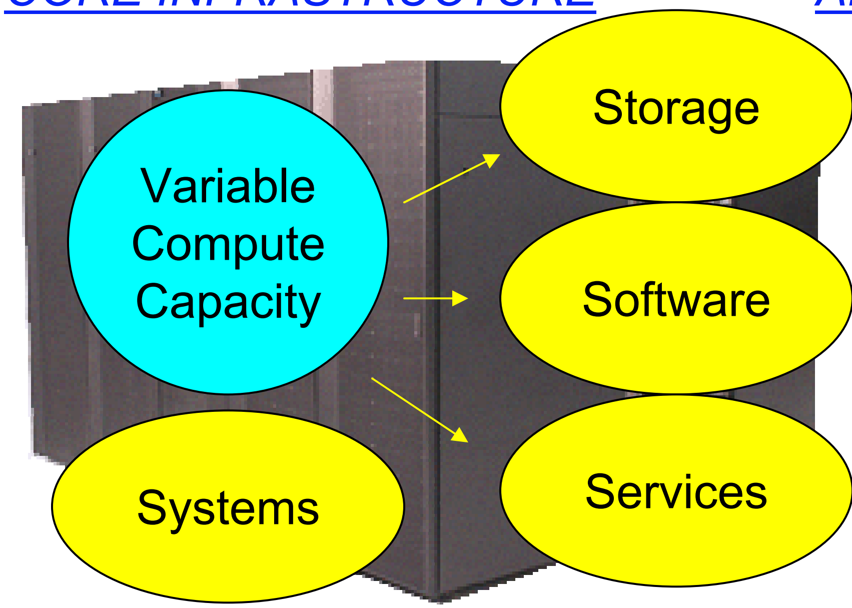
Initial Approach (Testbed/Demonstration)

- Identify one or two simple, yet compelling example problems
  - Proof of concept for science and technology with relatively modest investment
  - Illustration of utility for decision makers
  - Demonstration of potential for a developing country without significant weather forecasting infrastructure
- Leverage extant operational forecasting centers and expertise
- Availability of global, super-ensemble not required for initial demonstration
- Implement supplemental capabilities within IBM On-Demand HPC Centres, utilizing complementary IBM experience

IBM On-Demand HPC Centres

CORE INFRASTRUCTURE

ADVANCED FEATURES AND SERVICES



- IBM @server Storage Nodes
- IBM TotalStorage® Fiber Channel and SCSI Disk

- IBM General Parallel File System (GPFS)
- Job scheduling and workload management (*)
- Grid-enablement capability (*)

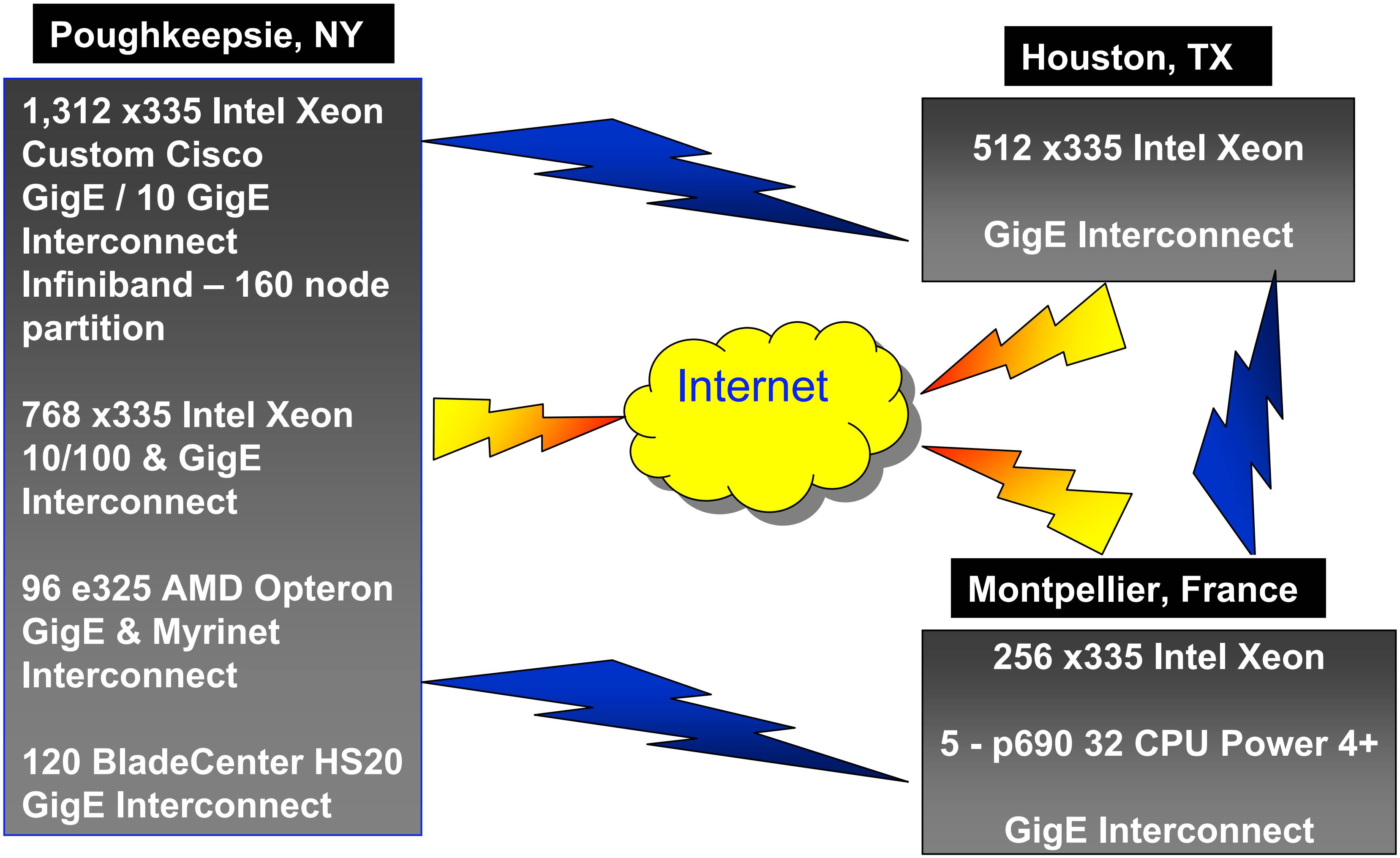
- Advanced monitoring & management
- Software customization & deployment
- Storage/data management (e.g., backup)

- Compute Nodes
  - IBM @server™ Cluster 1350
    - xSeries® Intel® Xeon™ 32-bit technology
    - AMD Opteron™ processor 32/64-bit technology
  - IBM pSeries® POWER4+™ 64-bit technology
- Interconnect: 10/100 & Gigabit Ethernet
- Management Node
- Cluster Systems Management (CSM)
- Virtual Private Network (VPN)

BASE SERVICES

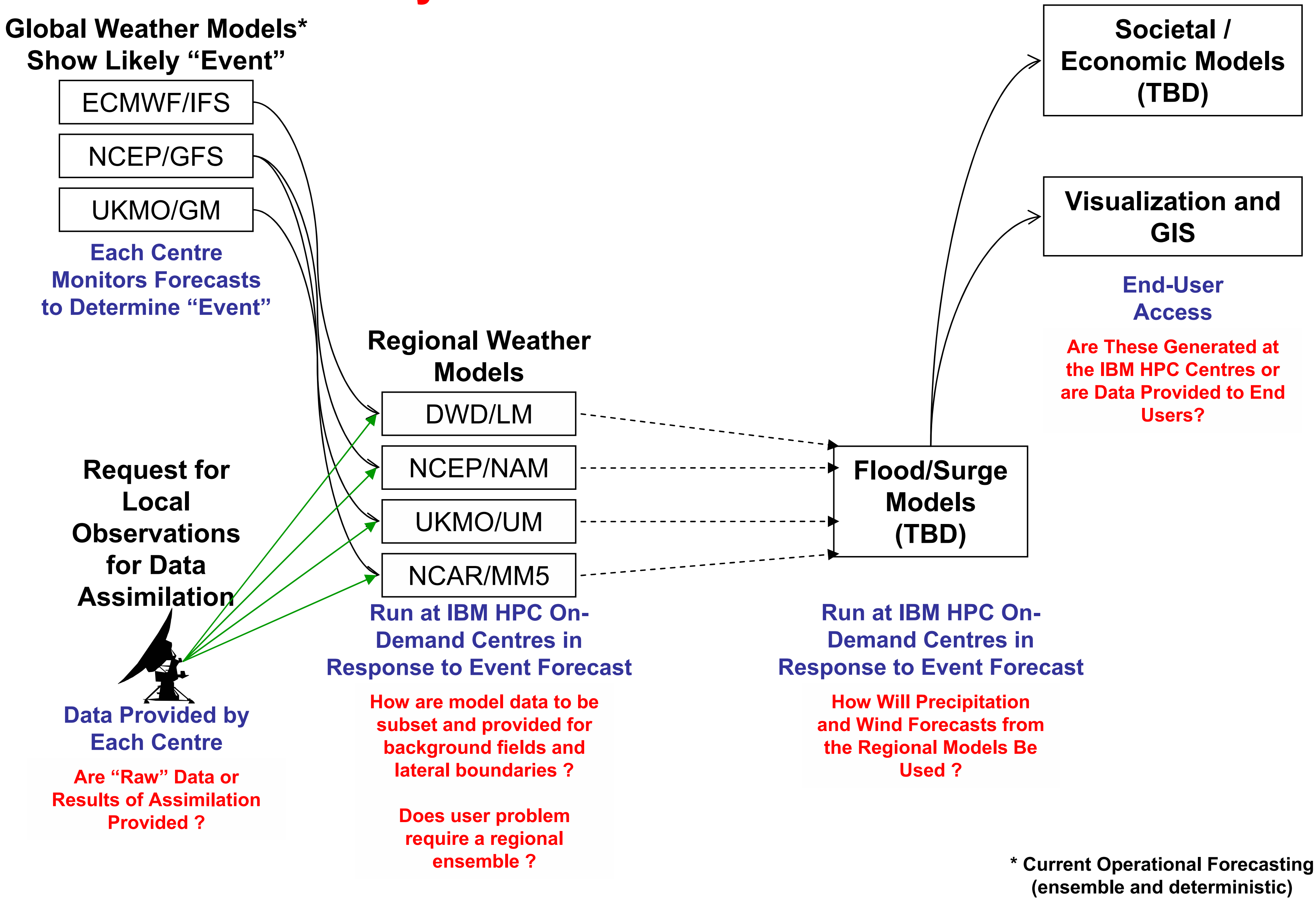
- Facility
- Provisioning
- Deployment
- Virtual Private Network
- Security
- Monitoring
- Maintenance
- Help Desk
- IBM Representative

IBM On-Demand HPC Centres



All statements regarding IBM's future direction and intent are subject to change or withdrawal without notice, and represent goals and objectives only.

Flood Forecasting in the Indian Subcontinent from a Tropical Cyclone: *Demonstration*



Other Issues from Demonstration Scenario

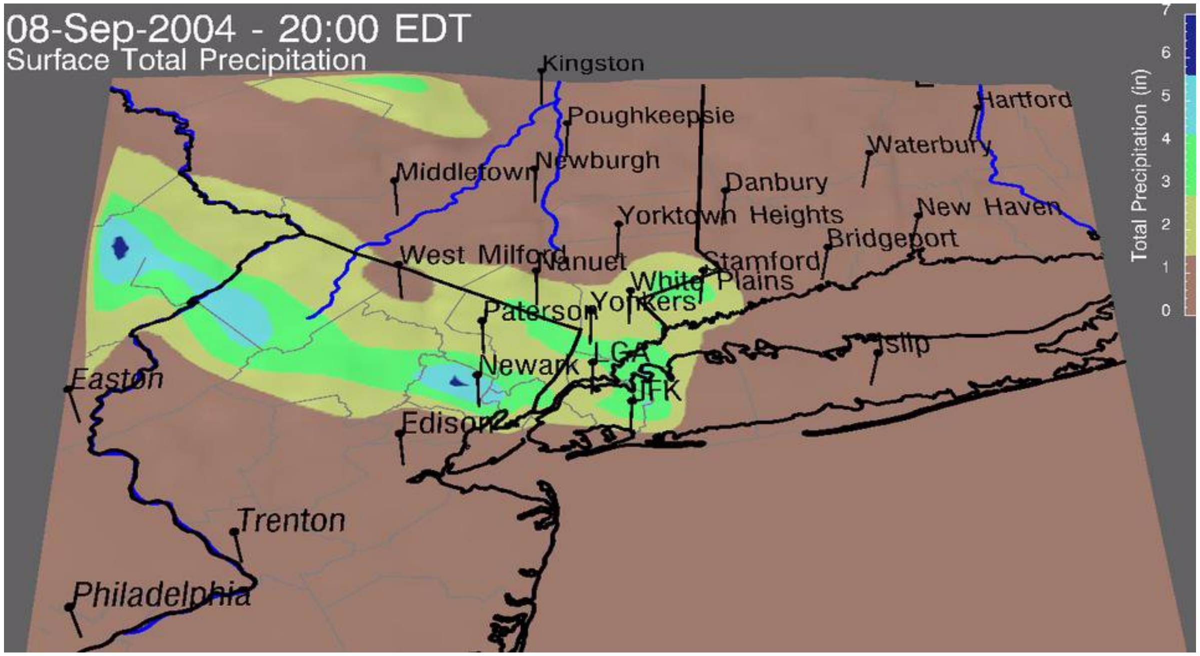
- Where and how is Grid integrated (data/information) ?
  - Since higher-level protocols are needed, the testbed can utilize a well-defined set of model outputs as they are produced today, but that will not scale
  - Implement a cache service to ensure availability of useful and timely data, to compensate for potential of data interruption and discard stale data
  - Can a Services Oriented Architecture (SOA) be layered on top of physical files (e.g., GriB, netCDF COARDS) to simplify use of Grid?
- How much compute and bandwidth is required ?
- Develop a DBMS-like approach to define user-centric views for visualization and dissemination ?

An Example Flooding Event – New York City Metropolitan Area -- 8 September 2004

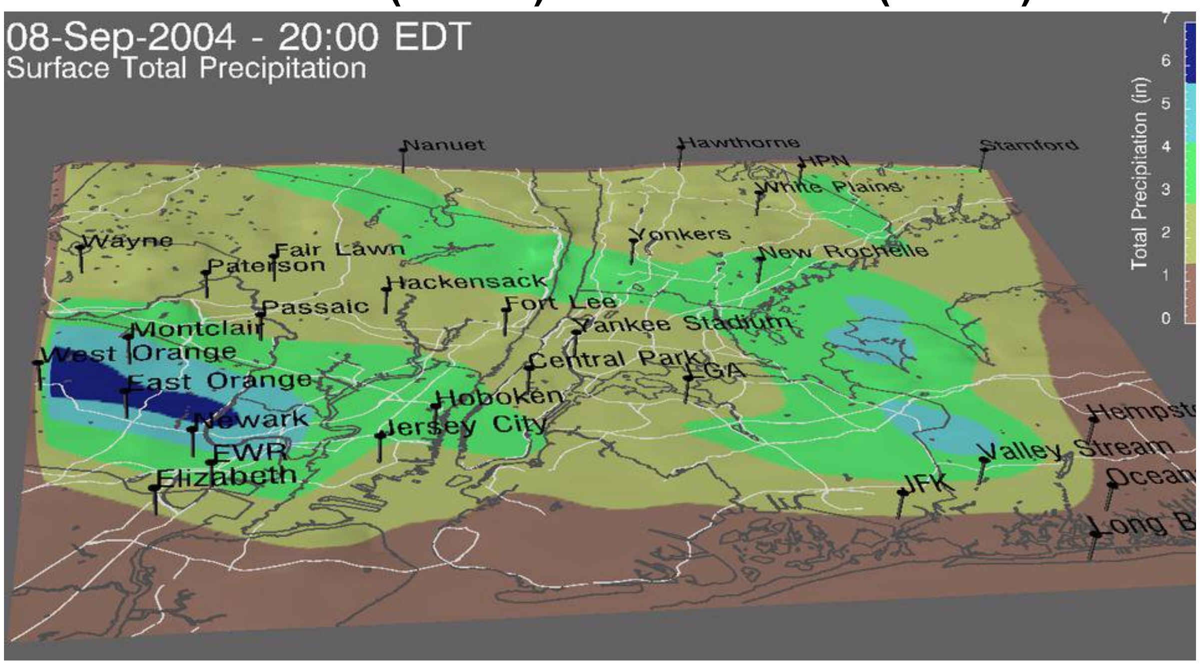
Consider Potential Role of Targeted NWP

- Remnants of Hurricane Frances moved into the New York City metropolitan area early in the morning of September 8
- The heaviest rainfall occurred in an area stretching from northeastern New Jersey through central Westchester County, NY with amounts in excess of 130 mm in some areas (as high as 50 mm/hr)
- There was widespread disruption of transportation systems (e.g., road closures, flooded subways, airport delays) and significant flooding in several regions between 0600 and 1000 local time
- Evening National Weather Service zone forecast (2130 local time, 7 September) for the next day: "showers and a slight chance of thunderstorms, rain may be heavy at time in the morning"
- Revised National Weather Service zone forecast (0440 local time, 8 September), adding: "locally heavy rain possible"
- National Weather Service issued a flash flood watch at 0748 (LT)

Focused Meso-γ-NWP Results: 07 September 2004 Late Evening



Forecasted rainfall totals through 24 hours: 4 km nest (above) and 1 km nest (below)



- Heavy rainfall predicted for the morning with similar distribution to reported rainfall, although some differences in totals
- Forecast initiated with data from 2000 local time (0Z) with results available about midnight
- Deterministic nested, non-hydrostatic model with full bulk microphysics provides results relevant to user problem
- Despite some phase errors, significant "heads-up" for event

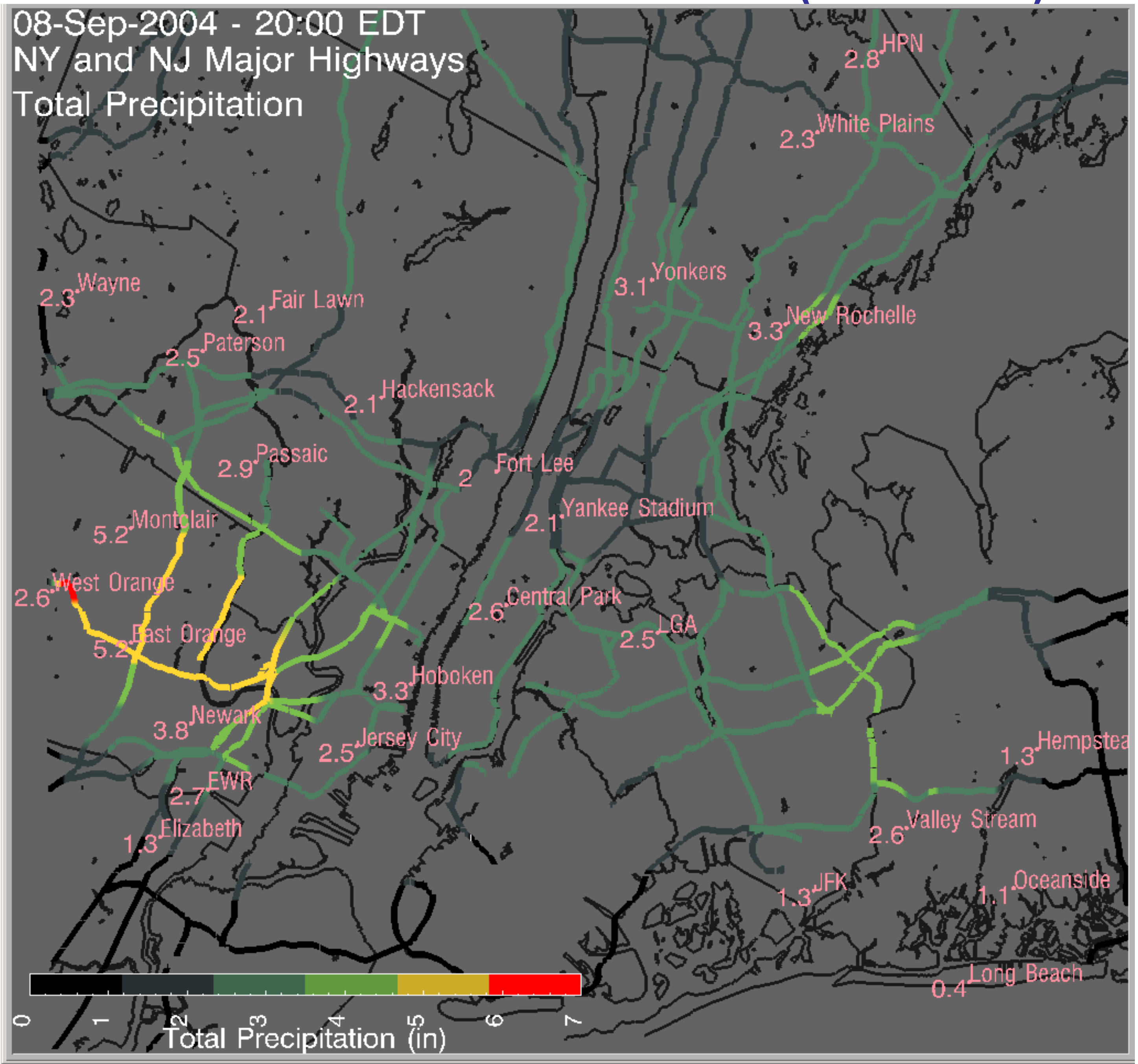
A More Focused Presentation of the Forecast May Be Appropriate for Specific Applications

(e.g., road maintenance and operations, traffic management)

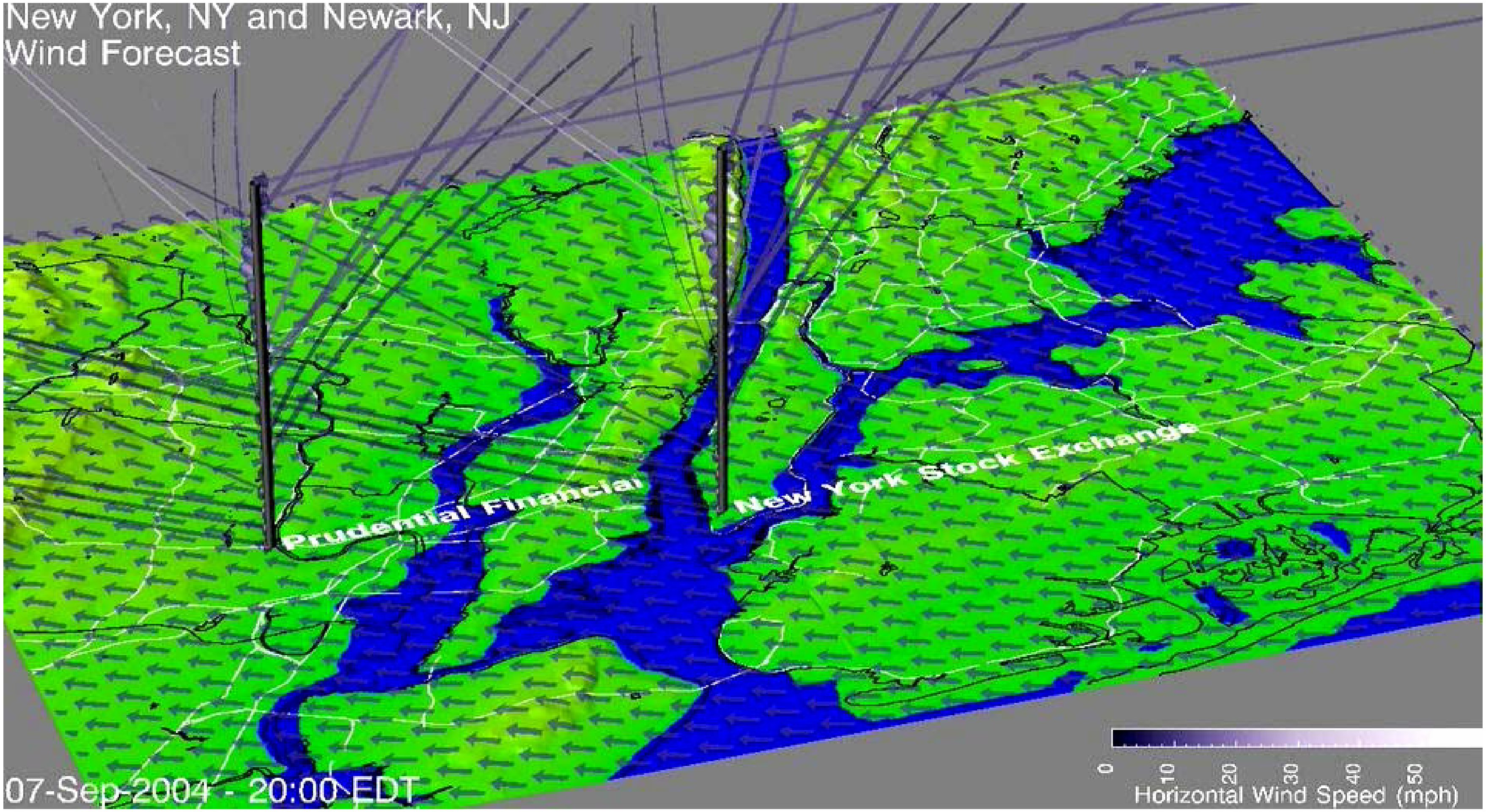
Measured Rainfall (Inches)

JFK	2.76
Mamaroneck	3.73
LGA	3.83
Central Park	3.75
Norwalk	4.25
White Plains	5.85
Fair Lawn	1.50
Bethpage	5.20
Orange	2.30
EWR	2.07
Hoboken	3.87

Road Forecast of Rainfall Totals (1 km Nest)



Forecast Results for Other Applications



- Interest in surface and upper air winds dictates entirely different presentation
- “Virtual wind profilers” at two locations within 1 km nest enhanced with trajectories to show forecasted propagation