Precision Wind Power Forecasting via Coupling of Turbulent-Scale Atmospheric Modeling with Machine Learning Methods

L. Treinish, J. Cipriani and A. Praino
IBM Thomas J. Watson Research Center
Yorktown Heights, NY, USA
{lloydt,jpcipria,apraino}@us.ibm.com

J. Díaz and J. R. Dorronsoro
Instituto de Ingeniería del Conocimiento
Madrid, Spain
{julia.diaz,jose.r.dorronsoro}@iic.uam.es

Other Presentations of Related Work

Conference on US 2012 Weather Impacts
- 2.4 Hindcast Analysis of the June 2012 Derecho and Its Impact on the Baltimore-Washington Metropolitan Area using High-resolution WRF-ARW

Symposium on the Coastal Environment:
- 2.2 High-resolution Simulations of a High-impact Rainfall Event for the Montpellier Region using WRF-ARW
- 2.3 December 2010 Northeast Blizzard: Event Analysis using High-resolution WRF for the New York City Metropolitan Area
- 8.3 Forecast Performance of an Operational Mesoscale Modeling System for Tropical Storm Irene and Post-Tropical Cyclone Sandy in the New York City Metropolitan Region

Symposium on the Next Level of Predictions in Tropical Meteorology:
- 1.5 A Numerical Weather Prediction-Based Infrastructure for Tropical Meteorology Research and Operations in Brunei
- TJ36.3 The DOTSTAR Observations in Improving Tropical Cyclones Forecast using Ensemble-based Data Assimilation

Conference on Transition of Research to Operations:
- 3.1 Enabling a High-Resolution, Coupled Hydro-Meteorological System for Operational Forecasting of Severe Weather and Flooding Events in Rio de Janeiro

Conference on Hydrology:
- 533 A Dynamic River Network Model for Regional-Scale Simulation

Conference on Climate Variability and Change:
- 551 Seasonal Climatology Studies for Tropical Region - Borneo Island Case Study

Conference on Weather, Climate, and the New Energy Economy:
- 1.2 On-going Utilization and Evaluation of a Coupled Weather and Outage Prediction Service for Electric Distribution Operations
- 800 Utilization of a High Resolution Weather and Impact Model to Predict Hurricane Irene
- 409 Advanced Data Assimilation for Short-term Renewable Power Prediction: a Complex Terrain Case

Precision Wind Power Forecasting via Coupling of Turbulent-Scale Atmospheric Modeling with Machine Learning Methods

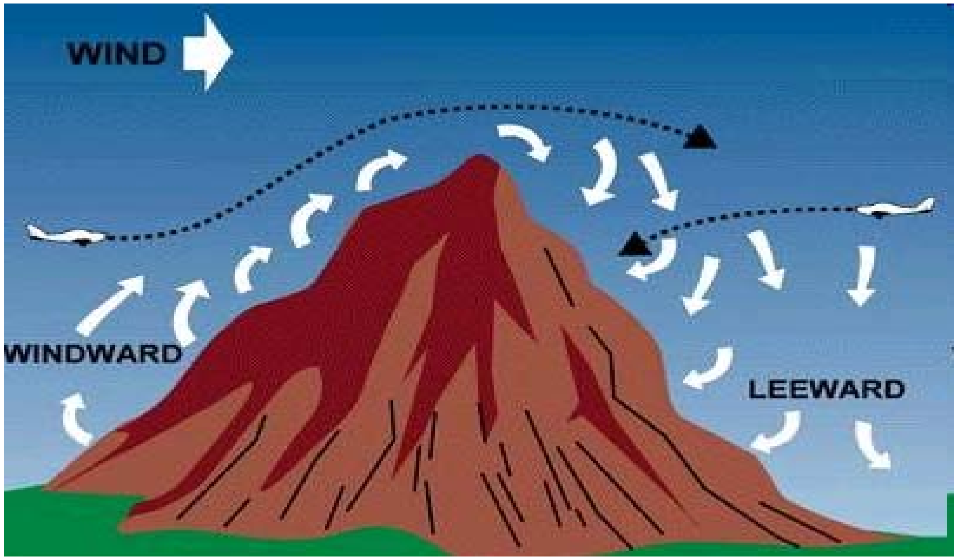
- **Motivation and background**
- **Approach**
- **Preliminary results**
- **Project plans and status**

Motivation and Background

- **Wind power intermittency creates significant barriers to expanding utilization**
  - **Ramp events**
  - **Spinning reserve**
- **Better forecasting and optimized economic dispatch can alleviate these barriers**
  - **Ensemble forecasts**
  - **Stochastic programming**
  - **Dynamic reserves**
- **Challenges are greater for isolated systems such as on islands**
  - **No grid interconnection with larger systems**

Canary Islands

- Isolated system across the archipelago (7,493 square km of land area)
- Red Eléctrica de España: 45 wind farms, located on 5 of the 7 islands with ~142 MW aggregated capacity with a wide variety of equipment (327 turbines)
- Complex topography leads to turbulent flow, especially along the coastlines
  - Peaks up to 3500m, inducing vortices
- Large power output variability as a result of ramp events
  - For example, 7 November 2010: 61% variability in a four-hour time span on Gran Canaria
  - Impacting reliability, electricity generation
- Ramp events are poorly predicted
  - NWP-based forecasts do not capture flow
  - Machine learning and statistical methods are brute force and lack good training sets

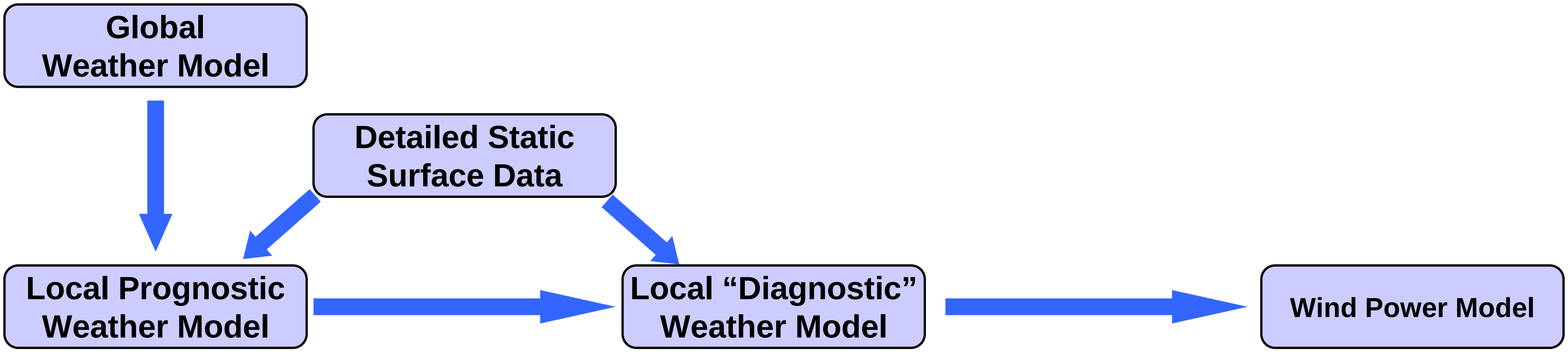


Approach

- Given the geography of the archipelago, and the spatial distribution of the individual turbines, turbulence-scale modeling becomes essential to capture the flow
- Introduce large eddy simulations (LES)
  - –Capture planetary boundary layer (PBL) effects
  - –Momentum, heat, moisture flux terms become critical
  - –High-temporal resolution required to capture transients (output every five minutes)
- Retrospective analysis of critical ramp events
  - –Many numerical experiments to enable effective model configuration
- Balance detail vs. performance
  - –Horizontal/vertical resolution vs. time step, subject to CFL stability criteria

    $|U|*dt/dx <= 1$

Approach to Coupled Weather and Power Modelling



Local Prognostic Weather Model

- 1-2 km horizontal resolution
- 20-1000m vertical resolution
- Full dynamics and 3d physics

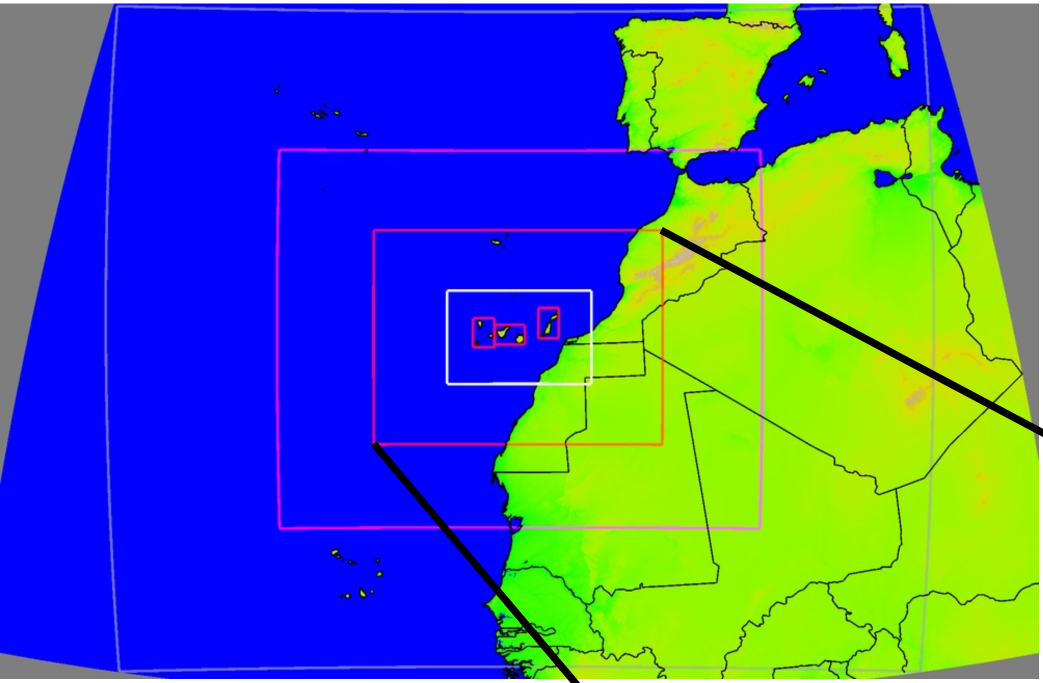
Local "Diagnostic" Weather Model

- 400-800m horizontal resolution
- 20-1000m vertical resolution
- Turbulent-scale dynamics and more limited physics

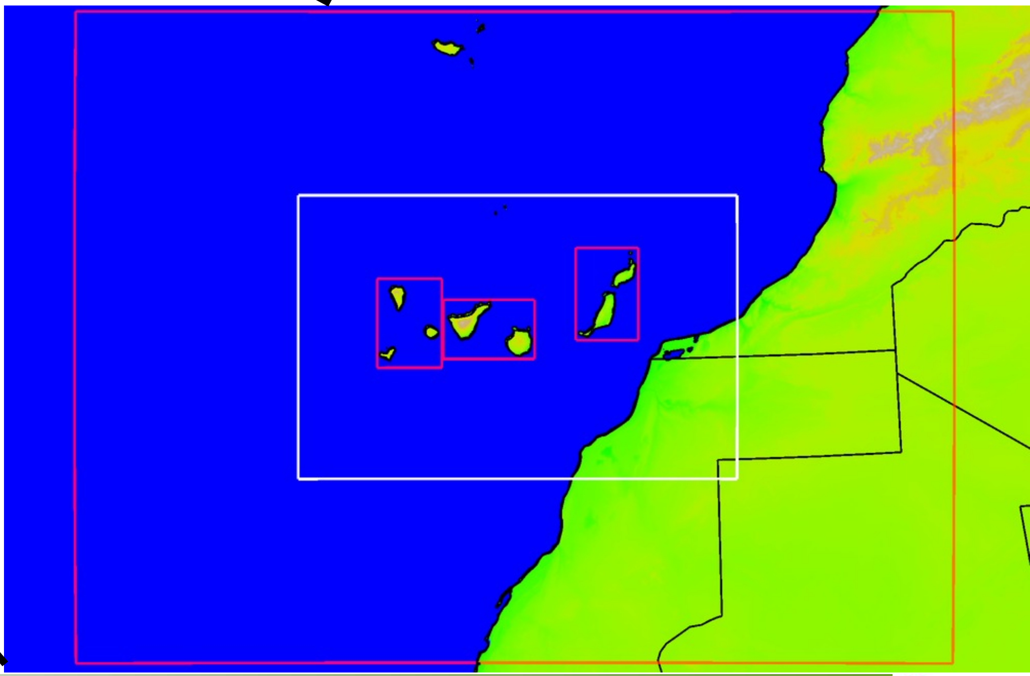
Weather Model Configuration: WRF-ARW 3.3.1

- 50 vertical levels with ~10-15 in the planetary boundary layer to ensure capturing of orographic effects
- 24 hour runs initialized at 0 UTC
- NOAA GFS for background and lateral boundary conditions
- SRTM-based model orography (90m)
- MODIS-based land use data
- 1km-resolution JPL SSTs
- Four 2-way nests at 54-km (87x70), 18-km (151x118), 6-km (268x199), 2-km (358x244) focused on the Canary Islands
  - WSM 5-class single moment microphysics, RRTM long wave radiation, GSFC short wave radiation, YSU PBL, NOAH LSM, Kain-Fritsch cumulus physics
- Three one-way LES domains embedded within domain four at 666.67m resolution (178x244, 250x163, 172x253)
  - WSM 5-class single moment microphysics, new GSFC long and short wave radiation, LES PBL, NOAH LSM, explicit cumulus physics
- Data assimilation is not feasible given the lack of a comprehensive observing system

Weather Model Configuration

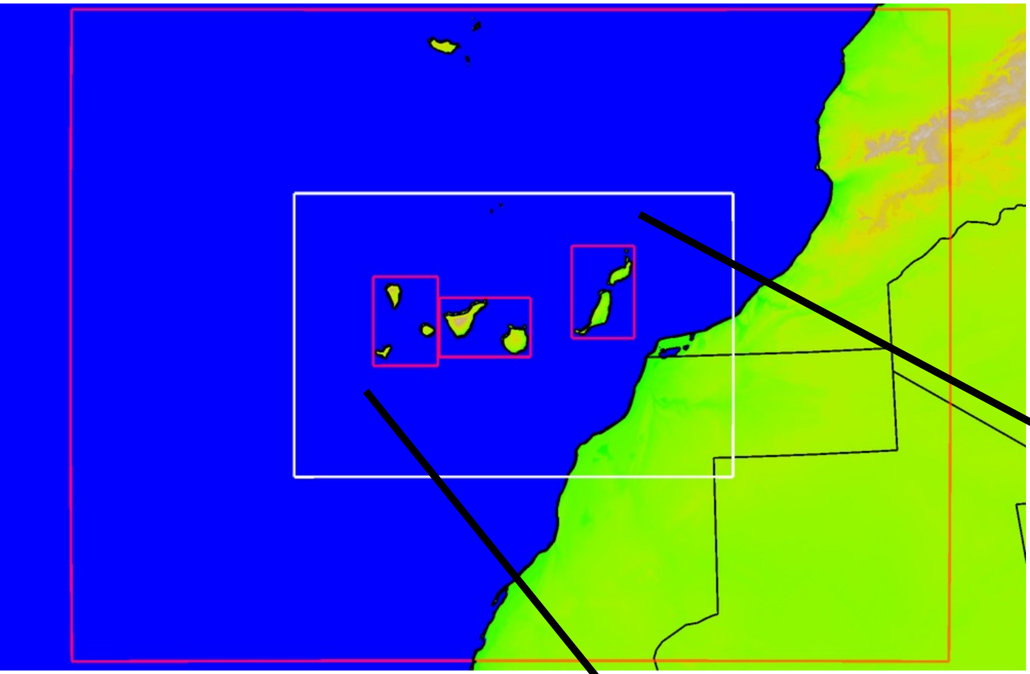


Four 2-way telescoping nests at 54, 18, 6 and 2 km horizontal resolution driving three, independent one-way LES nests at 667m resolution focused on the Canary Islands

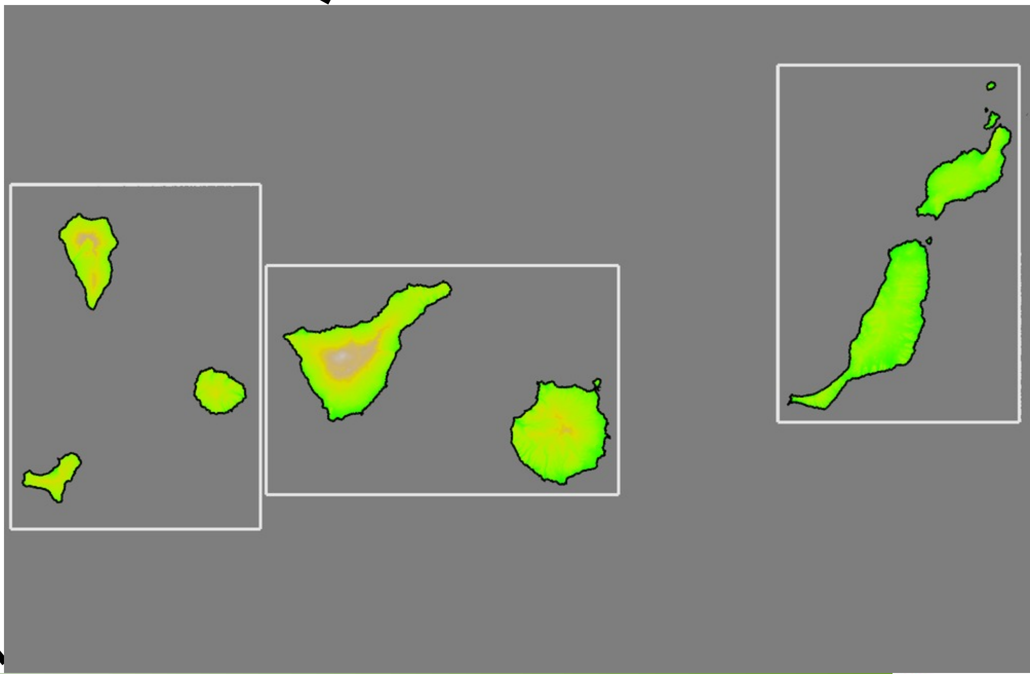


50 vertical levels with 10 to 15 in the planetary boundary layer

Weather Model Configuration



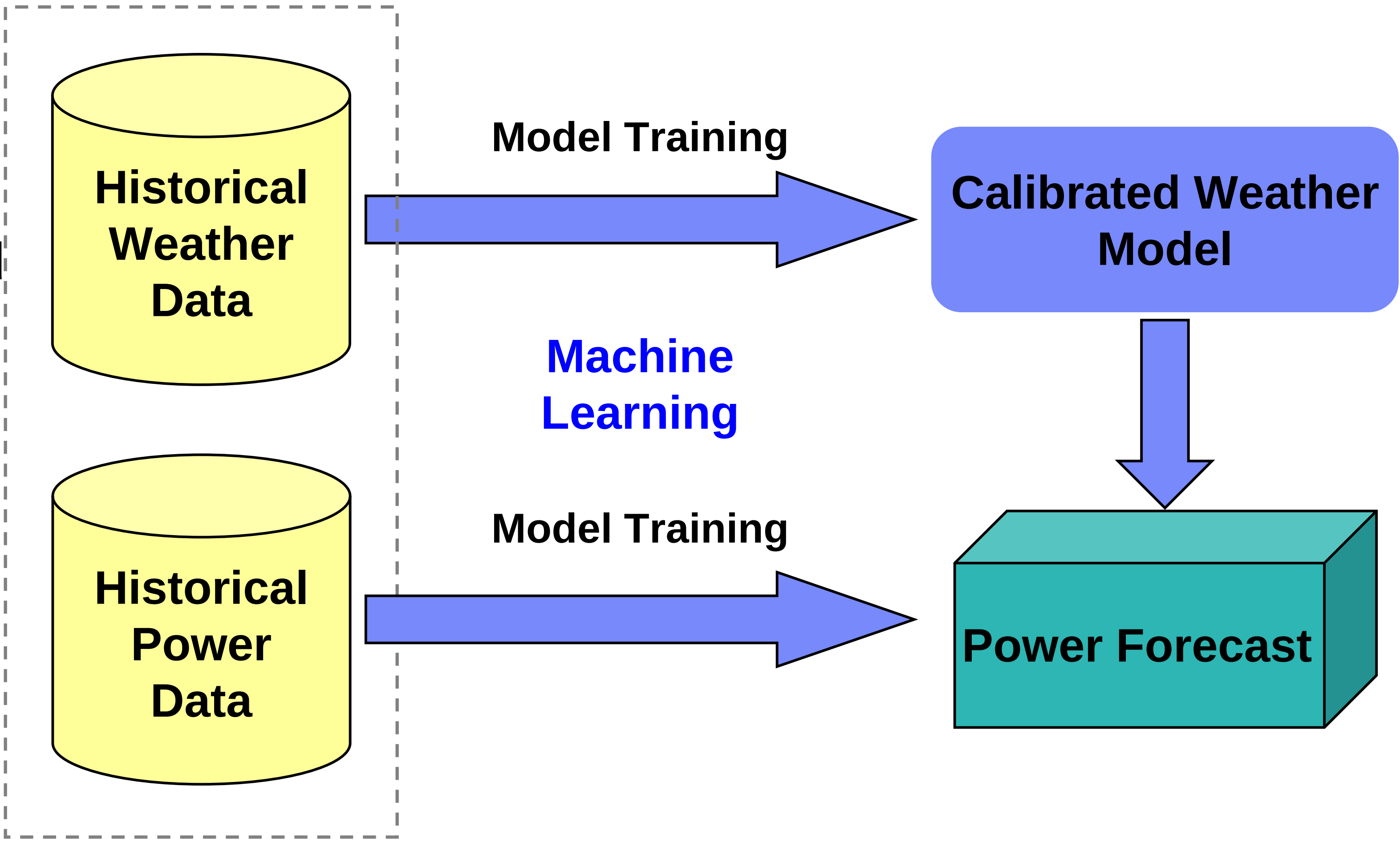
Four 2-way telescoping nests at 54, 18, 6 and 2 km horizontal resolution driving three, independent one-way LES nests at 667m resolution focused on the Canary Islands



50 vertical levels with 10 to 15 in the planetary boundary layer

Coupled Weather and Power Modelling

- **Build predictive model from historical weather forecasts and power, and related data**
  - **Wind farm power**
  - **Turbulent flow**
  - **Wind farm locations and characteristics**



Approach

- **Create a targeted hindcast database as a training set for the machine learning algorithms, based on the atmospheric physics**
  - Starting with 19 ramp events throughout 2010 and 2011
  - Include a six-month continuous period (daily)
- **Avoid brute-force machine learning approach**
  - Use only data that relate to energy extraction process
  - Volumetric Turbulent Kinetic Energy (TKE), absolute vorticity, 3-vector wind fields
  - Derived surface gusts and Clear-air Turbulence (CAT) index (two-dimensional)

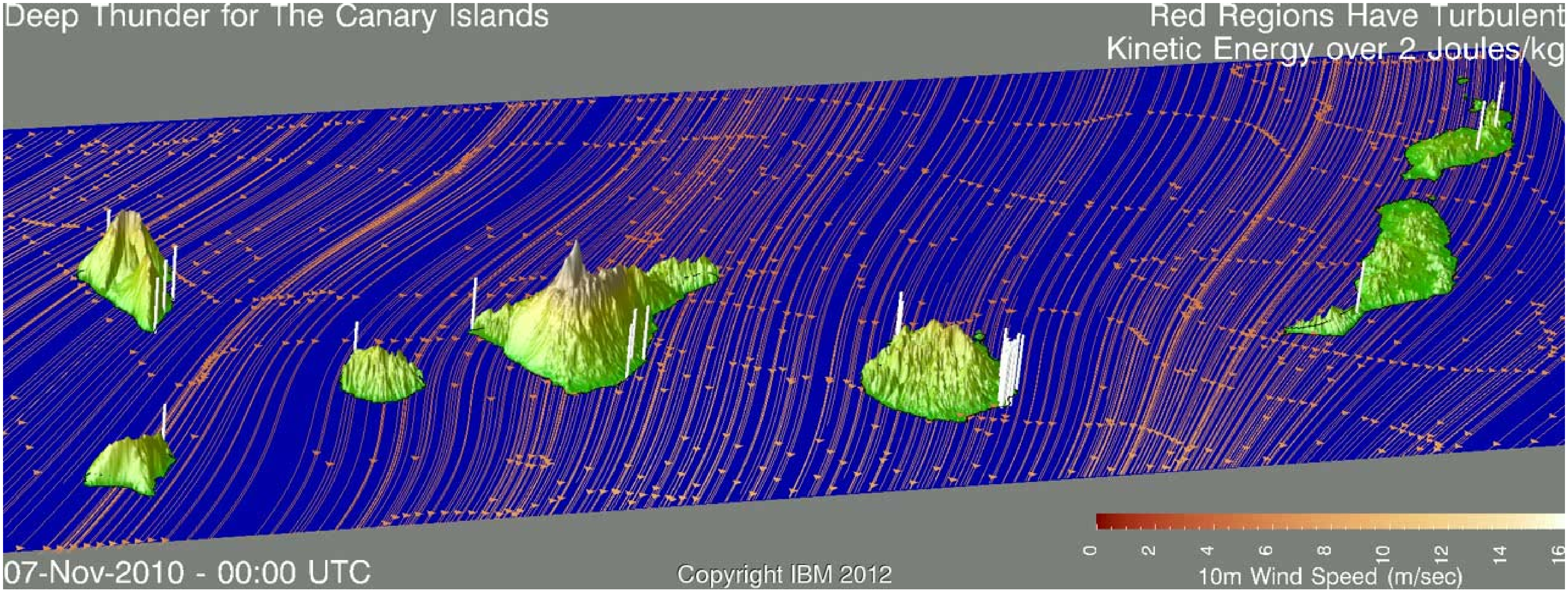
Computational Issues

- **Ordinarily, significant resources required to enabling LES forecasts in a production environment over a large domain**
  - Some effort to address the practicality using a modest HPC system
  - Effort to build training set and potential for operational use
- **Optimized for a cluster of ten 32-way Power7 nodes, each with 256GB memory with a DDR Infiniband interconnect**
  - Four 2-way nests (to 2km) run 24 hours in 50 minutes on six nodes
  - Three 667m LES nodes run in parallel, each using three nodes requires about 100 minutes as an NWP post-process (1-way nests)
  - End-to-end processing is about 3.5 hours per 24-hour simulation
  - Six-month climatology requires ~630 hours of compute time
- **Each run generates 180 GB of data (uncompressed)**
  - Most of the data are not relevant to drive machine learning
  - Four 2-way nests are cheaper to recompute than store
  - Only store fields related to turbulent flow in lower part of the boundary layer, resulting in a six-month climatology ~1 TB in size

Additional Challenges: Verification

- **Insufficient weather observations: 7 stations across the archipelago with only hourly reporting**
  - **May miss the transient events**
  - **5-minute interval data from hindcasts unverifiable**
- **Power data are "limited"**
  - **Hourly percentage increase/decrease, aggregated over each island**
  - **No power curve information for each turbine or farm**
  - **Only two years available: 2010-2011**

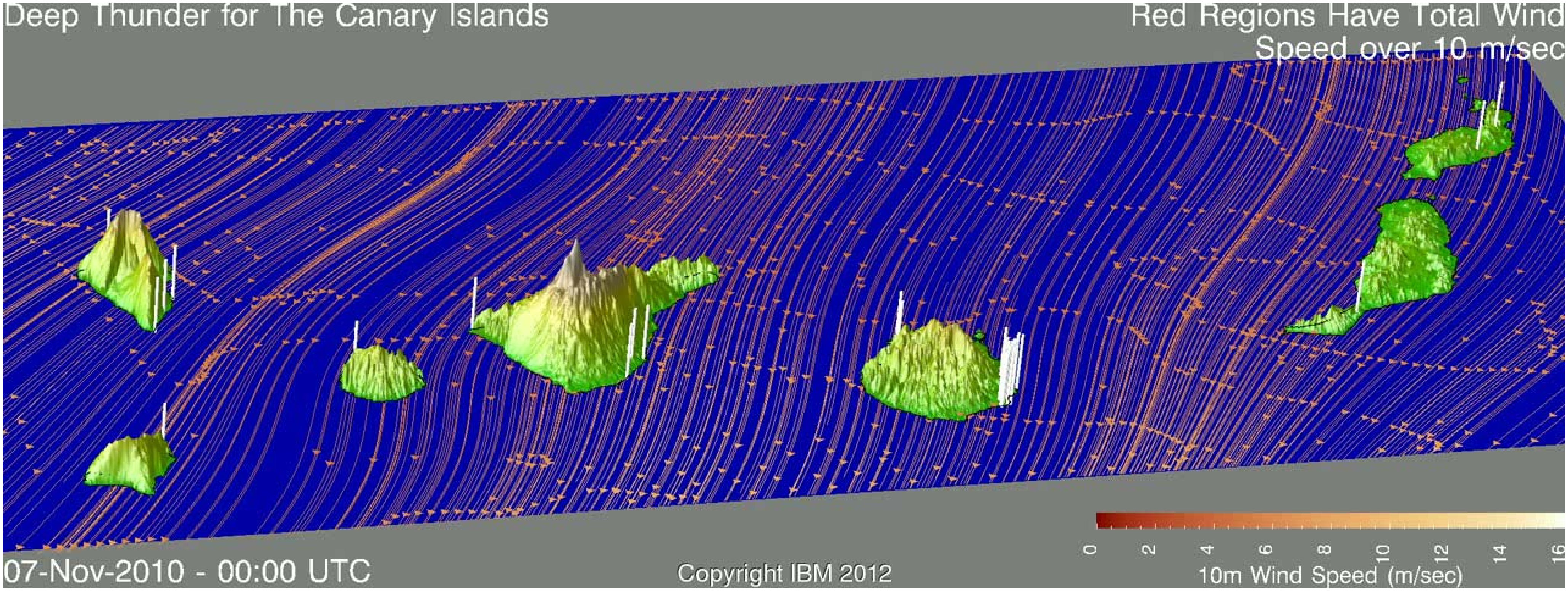
Preliminary Results: 7 November 2010 Ramp Event 0300 - 0700 UTC



Animation of 10m winds and TKE (red isosurfaces at 2J/kg) indicative of turbulent flow

- **The 10m wind streamlines above the water are derived from the 2km nest (4)**
- **The isosurfaces are derived from the three 666.67m nests (5)**
- **The terrain of each island is shown**
- **The location of each of the 45 wind farms are marked with a white pole whose height corresponds to the blade extent for the deployed turbines**

Preliminary Results: 7 November 2010 Ramp Event 0300 - 0700 UTC



Animation of 10m winds and [u,v,w] (red isosurfaces at 10m/sec) indicative of turbulent flow

- **The 10m wind streamlines above the water are derived from the 2km nest (4)**
- **The isosurfaces are derived from the three 666.67m nests (5)**
- **The terrain of each island is shown**
- **The location of each of the 45 wind farms are marked with a white pole whose height corresponds to the blade extent for the deployed turbines**

Machine Learning (ML) for Power Forecasts

- **Transform NWP output into energy forecasts**
- **Currently ML-based energy forecasts are derived from numerical patterns obtained from NWP outputs derived at the synoptic scale**
- **This approach implies both a large increase of the NWP pattern dimension and of sample size, which goes from eight to at least 24 patterns per day**
- **Sample sizes and dimension will have the same order of magnitude, contrary to the ML rule of thumb of sample size being an order of magnitude greater**
- **The ML algorithms to be used must be able to cope with this situation**
  - **Support Vector Machines (SVMs) are, therefore, a natural choice, as SVM models do not rely on individual pattern features, but rather on the overall pattern distribution**
- **Given the emphasis on wind and turbulence data, large correlations between NWP features and that effective NWP pattern dimension will be smaller**
  - **This suggests applying dimension-reduction techniques before model construction, such as Principal Component Analysis (PCA)**

Status and Future Work

- Model development and optimization was more complex than originally estimated
- Training set production now underway
  - Being evaluated with ML methods, but no useful results yet
- When training set is complete, apply model to other days in power data base to validate
- Additional ML methods to be considered
  - Modelling systems with a built-in capability for dimensionality reduction, particularly linear regression models coupled with sparsity-enforcing regularization, such as Lasso, Group Lasso and Elastic Net
- (Limited) verification of weather hindcasts
- Experiment with model configuration in forecast model