From HPC Performance to Climate Modeling: Transforming Methods for HPC Predictions Into Models of Extreme Climate Conditions

Ryan McKenna, Vivek K. Pallipuram, Rodrigo Vargas, and Michela Taufer

University of Delaware
“Predict the occurrence of idle resources (e.g., CPUs) by collecting and benchmarking data from cluster nodes and using a prediction window on the order of seconds/minutes“
From Cluster Performance ....

“Predict the occurrence of idle resources (e.g., CPUs) by collecting and benchmarking data from cluster nodes and using a prediction window on the order of seconds/minutes”

... to Weather Prediction

“Predict the occurrence of extreme weather conditions by collecting and benchmarking data from geographical locations and using a prediction window on the order of hours/months”
Our Contributions

• Transform a **method for predicting cluster performance** into a **tool to predict likelihood** of a region undergoing **extreme climate events**

• Use the tool to predict the proportion of the **matorral tropical region in Mexico** likely to experience extremely high or low temperature, and high precipitation as well as the event frequency
Outline

• Analogies between performance modeling and climate modeling
• Overview of our tool
  ▪ Learning phase and prediction phase
• Framework in action
  ▪ Data characterization
  ▪ Case studies: occupancy of extremely high or low temperatures
• Conclusions
Patterns in Cluster Performance Traces

Traces of a 8-node cluster executing 4000-task DAG-like workflow
Patterns in Cluster Performance Traces (II)

Traces of a 32-node cluster executing 4000-task DAG-like workflow
Patterns in Cluster Performance Traces (III)

Traces of a 256-node cluster executing 4000-task DAG-like workflow
Empirical Cumulative Frequency Analysis

- Use **empirical cumulative distribution functions (ECDF)** to model **event occurrences** in DAG-like workflows
  - E.g., computation, communication, and idle resource instances

- Use model to predict **optimal window size** for which the occurrence likelihood of idle resources is within a given confidence

Historical resource data \([X_1, X_2, X_3, \ldots, X_N]\)

Freq. of Exceedance

Extreme Climate Events

• We are interested in modeling the occurrence of extreme climate events and the optimal prediction time windows within a specified region

• Examples of extreme climate events includes:
  ▪ High and low temperatures
  ▪ High rain precipitations
Patterns in Climate Date

Times

Longitudes and latitudes

min: 0

max: 3600

0  500  1000  1500  2000  2500  3000  3500  4000
Performance Prediction

Step 1: Collect trace files for cluster nodes

Step 2: Define a window of resource event occurrences

Step 3: Sample resource events in the window

Step 4: Frequency analysis

Step 5: Obtain likelihood of resource event occurrences

Climate Prediction

TO

Gather climate data for geographical region

Define a window of extreme climate events

Sample extreme climate events in the window

Frequency analysis

Obtain likelihood of extreme climate event occurrences
Learning and Prediction Phases

Learning phase

Optimal modeling and forecasting windows

Prediction phase

Historical climate data

Shape file

Current climate data

GeoTiff files

GeoTiff files
Modeling and Forecasting Windows

• Modeling window: number of past years for which the climate events should be sampled in order to construct an ECDF model
• Forecasting window: the number of future years that can be studied by using the ECDF model given by the modeling window.
From Geospatial Climate Traces to Models

Spatial: Region of interest
- Shape file

Temporal: Climate data
- GeoTiff files

Feature Extraction

Cumulative Distribution Function

Candidate model windows

Candidate forecast windows

Model and forecast windows

Frequency
From Climate Data to Frequencies

Spatial:
Region of interest
Shape file

GeoTiff files

Temporal:
Climate data

Feature extraction mask
Events
Thresholds
Climate event occurrence generator

Extremes

Extreme Event Classifier

Frequencies
From Frequencies to CDF Models

1. Candidate model windows (M)
2. Candidate forecast windows (F)
3. Frequencies
4. Cumulative distribution function generator
5. ECDF\(_M\) uses modeling window
6. ECDF\(_F\) uses forecasting window
7. Kolmogorov-Smirnov (KS) test
8. ECDF\(_M\) \sim ECDF\(_F\)
9. HIT
10. Select ECDF\(_M\) and ECDF\(_F\) with highest hit rates across historical data for the prediction phase.
Data Characterization

- Use optimal modeling and forecasting windows from the learning phase to perform the ECDF-based frequency analysis on separate climate data
- Monthly temperature data over 24 years
- Prediction phase: 2001-2013

*TThe data was obtained from http://daymet.ornl.gov.*
Modeling and Forecasting Windows

- Extreme high temperature
  - Modeling window: 4 years
  - Forecasting window: 6 years
- Extreme low temperature
  - Modeling window: 4 years
  - Forecasting window: 6 years
- Extreme high precipitation
  - Modeling window: 4 years
  - Forecasting window: 6 years
Modeling Validation

- Extreme high temperature:
  - Modeling window: 4 years
  - Forecasting window: 6 years

- Extreme low temperature
  - Modeling window: 4 years
  - Forecasting window: 6 years

- Extreme high precipitation
  - Modeling window: 4 years
  - Forecasting window: 6 years

HIT RATE: ECDF models built using modeling and forecasting windows are statistically similar
Modeling Validation

- Extreme high temperature:
  - Modeling window: 4 years
  - Forecasting window: 6 years
- Extreme low temperature
  - Modeling window: 4 years
  - Forecasting window: 6 years
- Extreme high precipitation
  - Modeling window: 4 years
  - Forecasting window: 6 years

HIT RATE: ECDF models built using modeling and forecasting windows are statistically similar
Modeling Validation

- Extreme high temperature:
  - Modeling window: 4 years
  - Forecasting window: 6 years

- Extreme low temperature
  - Modeling window: 4 years
  - Forecasting window: 6 years

- Extreme high precipitation
  - Modeling window: 4 years
  - Forecasting window: 6 years

HIT RATE: ECDF models built using modeling and forecasting windows are statistically similar

HIT RATE: 93.5%
Case Study: Extreme High Temperatures

- **Summer (July):**
  - Probability of Exceedence: 23.1%

- **Winter (January):**
  - Probability of Exceedence: 6.8%

Forecasting window: 6 years

- Probability of Exceedence
- Frequency of Recurrence

~2 years
Case Study: Extreme Low Temperatures

Forecasting window: 6 years

- **Summer (July)**: 1.9%
  - Probability of Exceedence
  - Proportion of Region

- **Winter (January)**: 15.6%
  - Probability of Exceedence
  - Proportion of Region

- **Frequency of Recurrence**
  - ~1.5 years
  - ~2 years

---

23
Lesson Learned

• We extend HPC performance modeling technique to study extreme climate events
  ▪ First of its kind
  ▪ Models extremely high/low temperature and high precipitation events

• We demonstrate the tool using real climate data-set
  ▪ Build window models using monthly climate data from 1980 to 2000
  ▪ Validate using a 13 year window (2001 – 2013)

• With over 85% confidence, we answer critical questions such as “What is the expected proportion of the region that will observe extreme climate events? What is the likelihood to exceed this proportion multiple times?”
Acknowledgments

Global Computing Lab @ UDel

Sponsors:

Contact: taufer@udel.edu
gcl.cis.udel.edu