

Automatic Unstructured Grid Refinement Using Machine Learning for the Analog Ensemble of Numeric Weather Prediction

Weiming Hu*, Guido Cervone*, Shantenu Jha†, Vivek Balasubramanian†, Matteo Turilli†

*Geoinformatics and Earth Observation Laboratory, weiming@psu.edu

Dept. of Geography and Institute for CyberScience, The Pennsylvania State University

†Research in Advanced Distributed Cyberinfrastructure and Applications, The State University of New Jersey

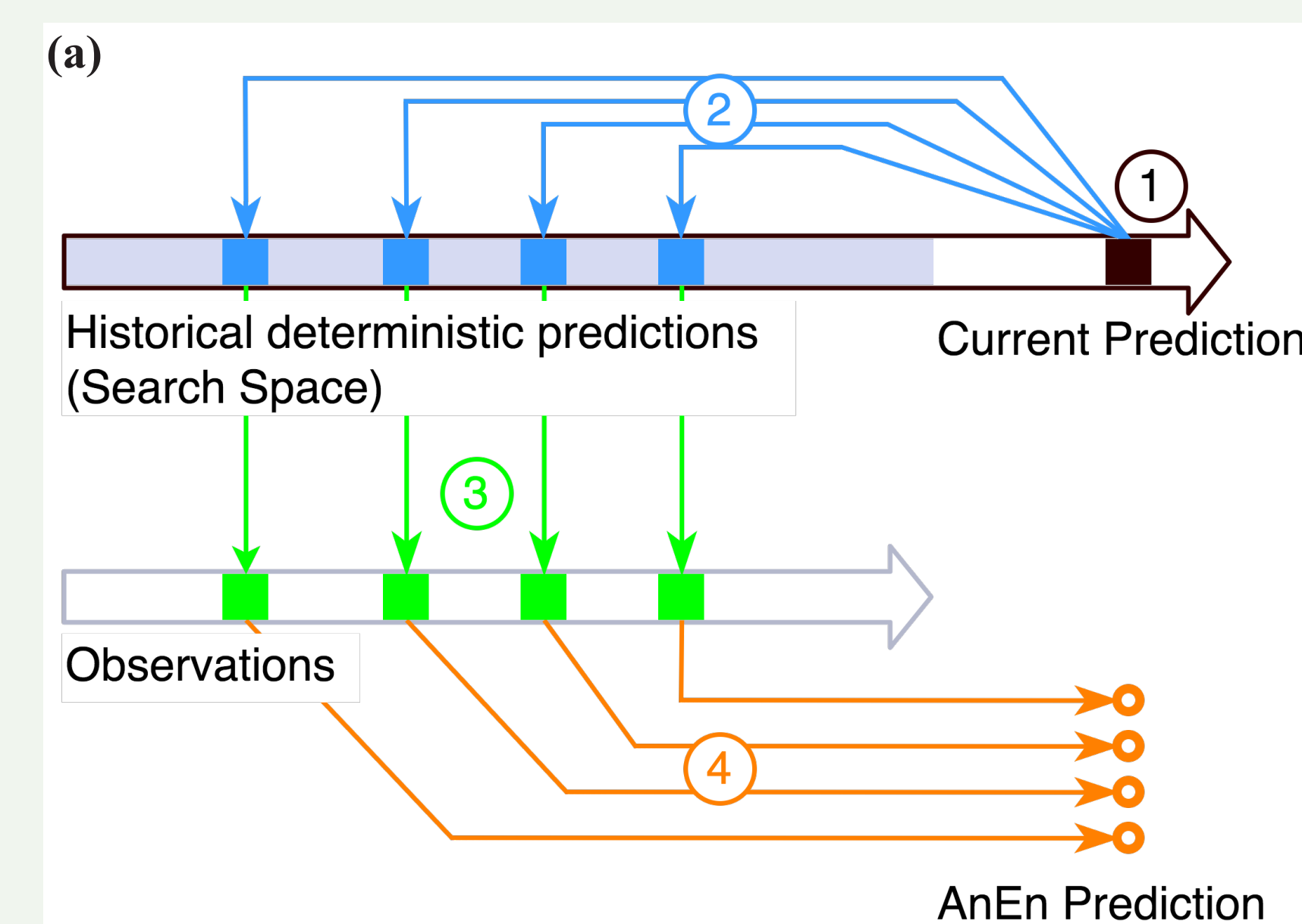
Introduction

Despite the fact that certain physical changes occur non-uniformly across space and time, Numerical Weather Prediction models generally use a structured grid to represent the domain. For example, temperatures tend to change more rapidly in mountains than plateaus but the model output assumes a gridded structure and therefore does not take the spatiotemporal features into consideration. However, it is often not necessary to generate unified resolution predictions across the domain. With an increasing spatial model resolution, the required computation can potentially grow very large and exceed the state-of-art computing capabilities.

An Analog Ensemble (AnEn) technique driven by a machine learning algorithm is proposed to automatically learn the optimal unstructured grid topology. There are two primary benefits from the optimal unstructured grid topology:

- Computing resource management is optimized as fewer grids are constructed;
- Operational temperature predictions can reuse this optimized grid topology in rapid high-resolution predictions.

Methodology



Analog Ensemble

The AnEn technique generates probabilistic forecasts.

1. Starts with a current deterministic prediction, a set of historical predictions, and the corresponding observations;
2. Finds the most similar historical predictions to the current prediction based on a multi-variate metric;
3. Takes the observations associated with the historical predictions;
4. Generates the prediction ensemble for the current time.

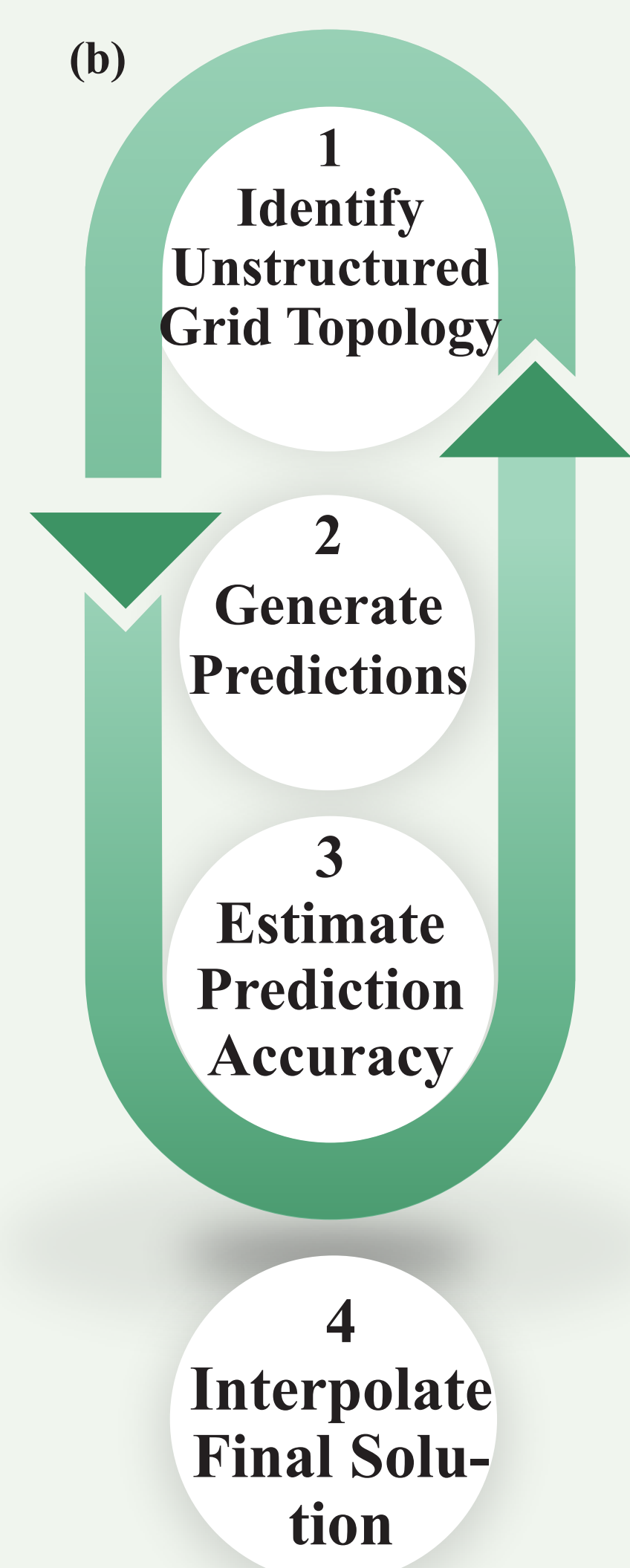
Unstructured Grid Refinement

An unstructured grid is constructed to characterize weather dynamics. The method to automatically identify the grid is referred to as the adaptive algorithm, shown in Figure (b).

1. Identifies an unstructured grid topology and builds Delaunay triangles using the identified vertices. For initialization, the grid is randomly generated given a preset amount of vertices;
2. Performs the AnEn computation only on vertices. The computation can be distributed and parallelized;
3. Estimates prediction accuracy and identifies regions with higher errors which would be marked for refinement in the next iteration;

The above 3 steps iterate until a termination condition is met. For example, the accuracy improvement is negligible, which indicates convergence of the algorithm, or computing resources are exhausted.

An interpolation step is carried out to generate the final solution.



Data and Initialization

North American Mesoscale (NAM) model forecasts and analysis are collected from 1 Jan 2015 to 30 Apr 2017 at a 12 km resolution. The domain consists of 262,792 grids. Figure (c) shows one year of data for 2016 for a selection of weather variables, together with an example output from the AnEn method.

- Line charts are the current predictions for 4 variables on January 15, 2017, with 4 forecast lead times starting from 06h.
- Scatter plots visualize historical predictions for the 4 variables across the year of 2016.
- Predictions marked red on the right are selected to form the prediction ensemble for the forecast lead time 18h which is marked red on the left.

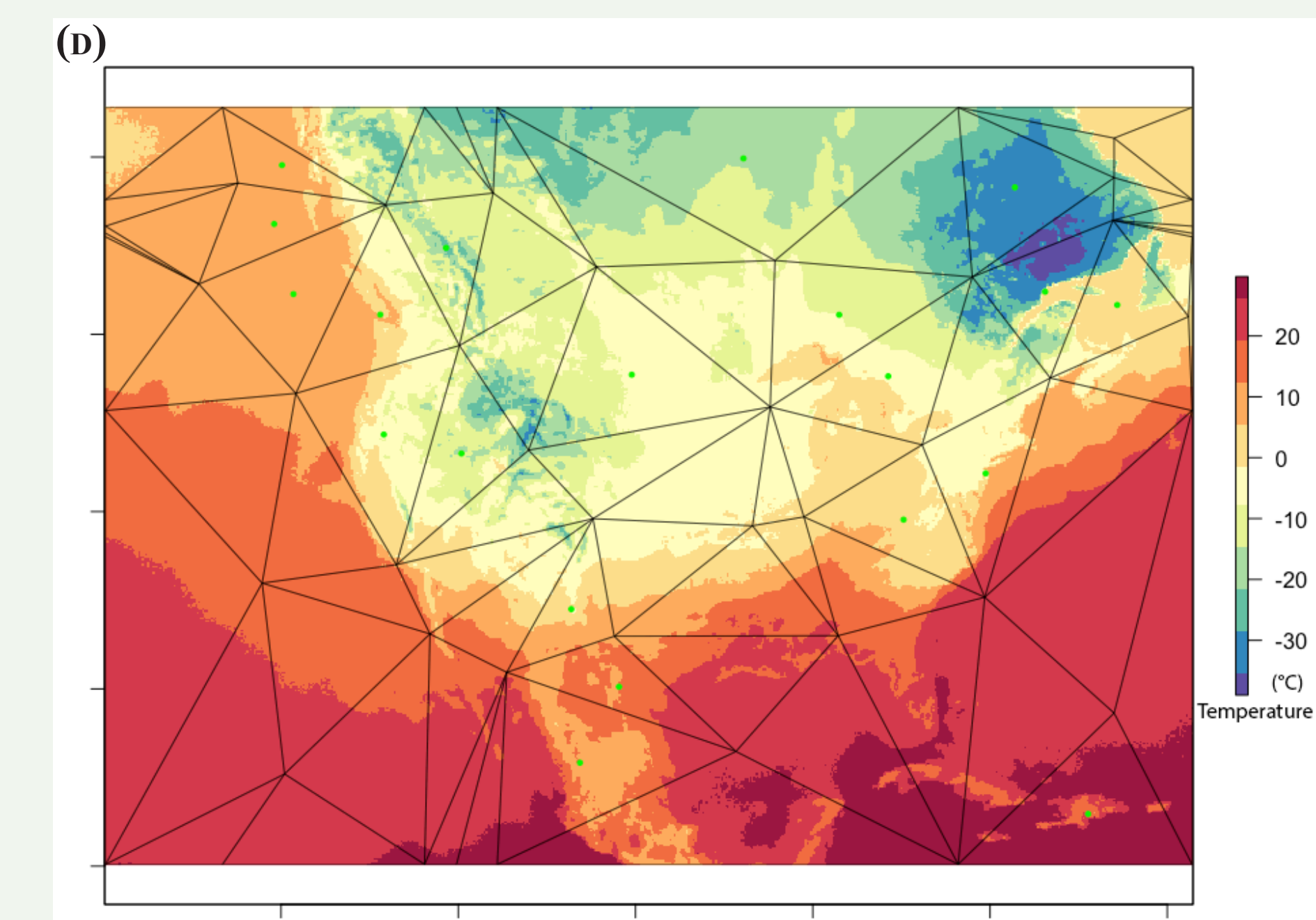
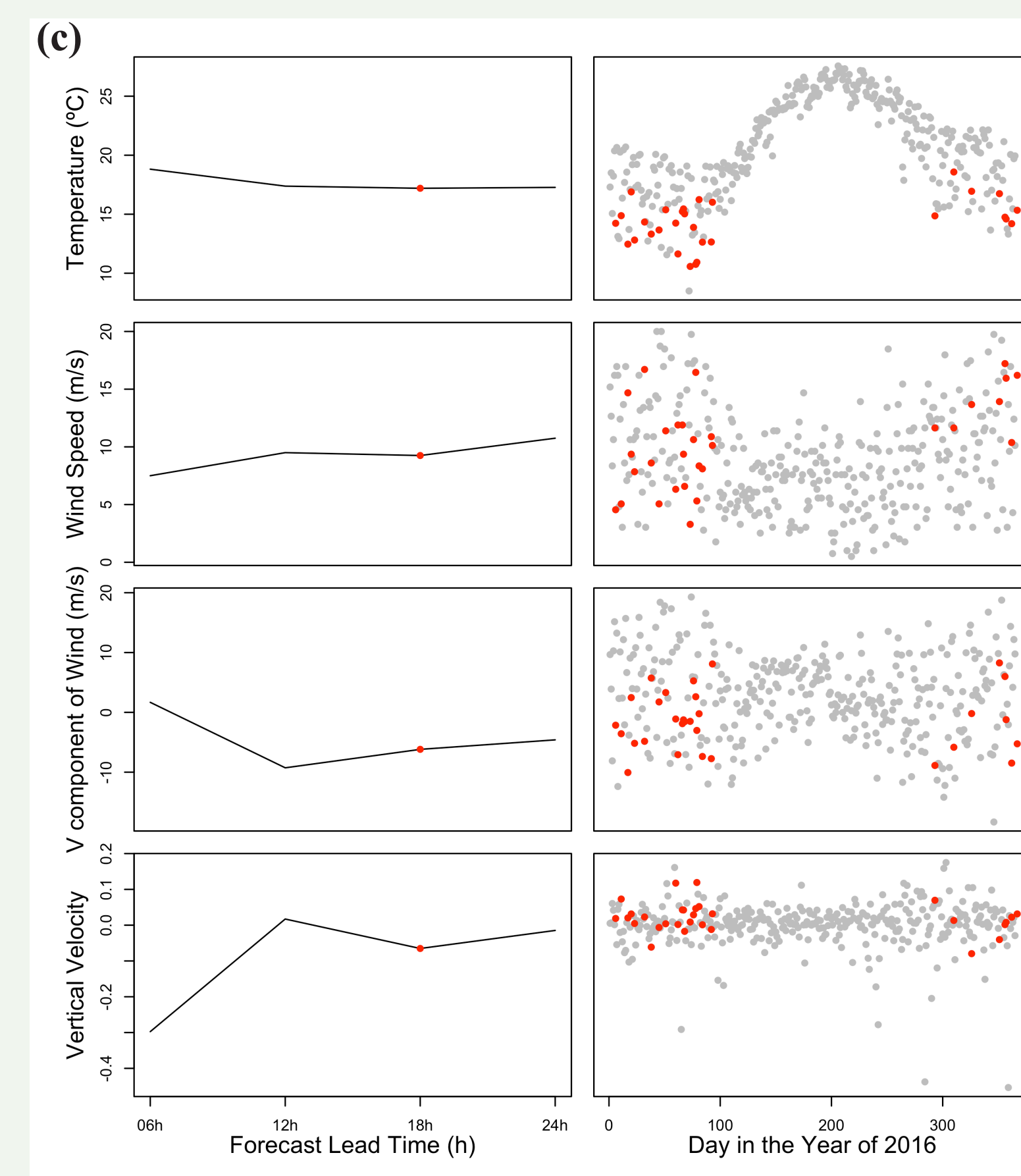
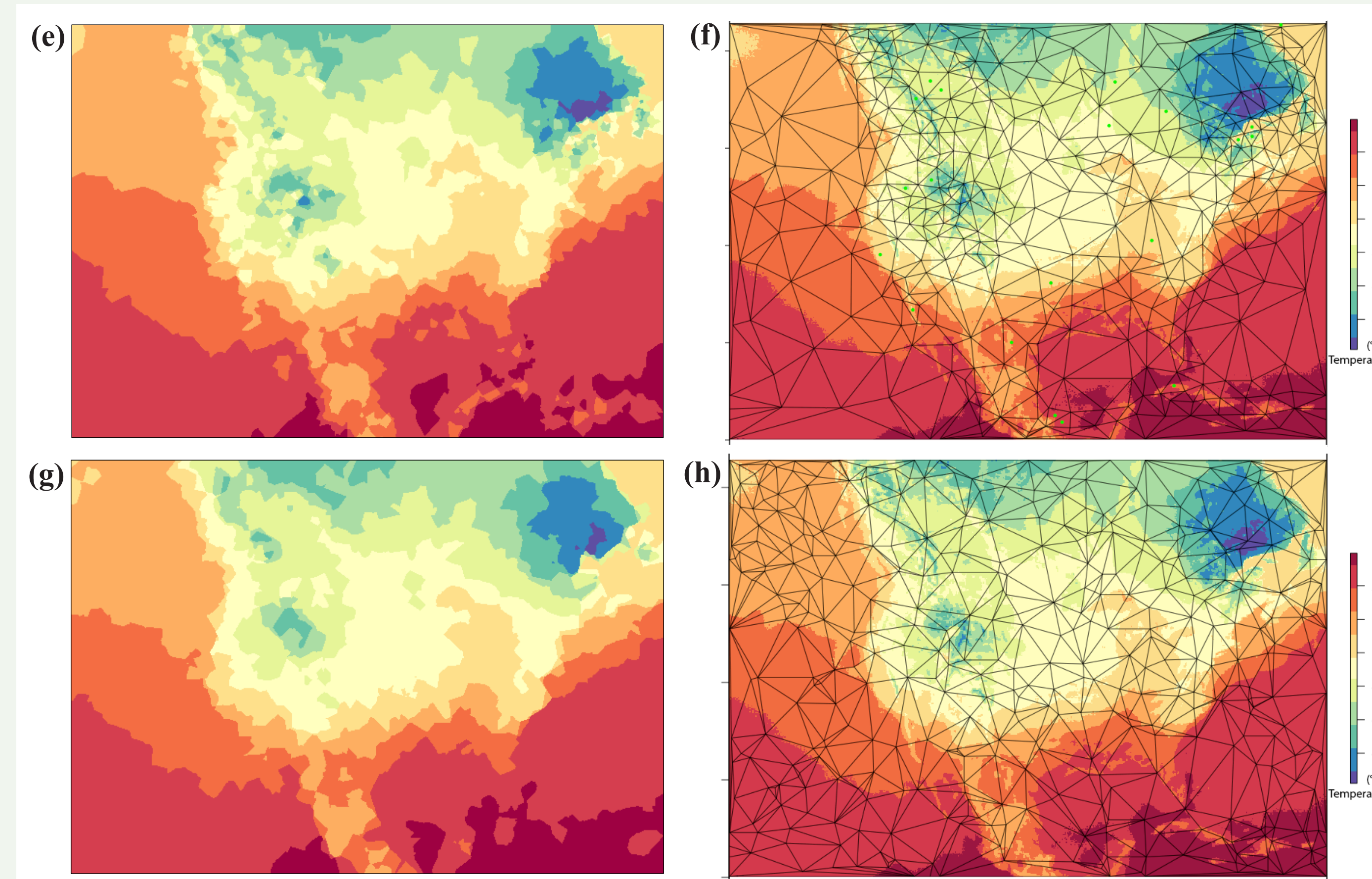


Figure (d) shows the temperature prediction map for January 15, 2017. The triangular mesh denotes the initialization for the grid refinement randomly created from 50 vertices. The green points suggest triangles that might need to refine in the next iteration.

Results and Analysis



Figures (e~h) show temperature predictions on January 15, 2017, at 06h. The AnEn computation is carried out on vertices and additional predictions are interpolated from vertices to each available grid on the original NAM output for a grid-by-grid comparison.

Figures (e) and (g) show the interpolated prediction maps from the adaptive and the random methods. Figures (f) and (h) to the right corresponds to the unstructured meshes used to generate the predictions to its left. The base maps for figures (f) and (h) are the temperature maps for January 15, 2017, at 06h which show the variability of temperature at the certain time stamp. Where the temperature has a high variability should be covered with a finder grid topology, and therefore leads to more detailed predictions.

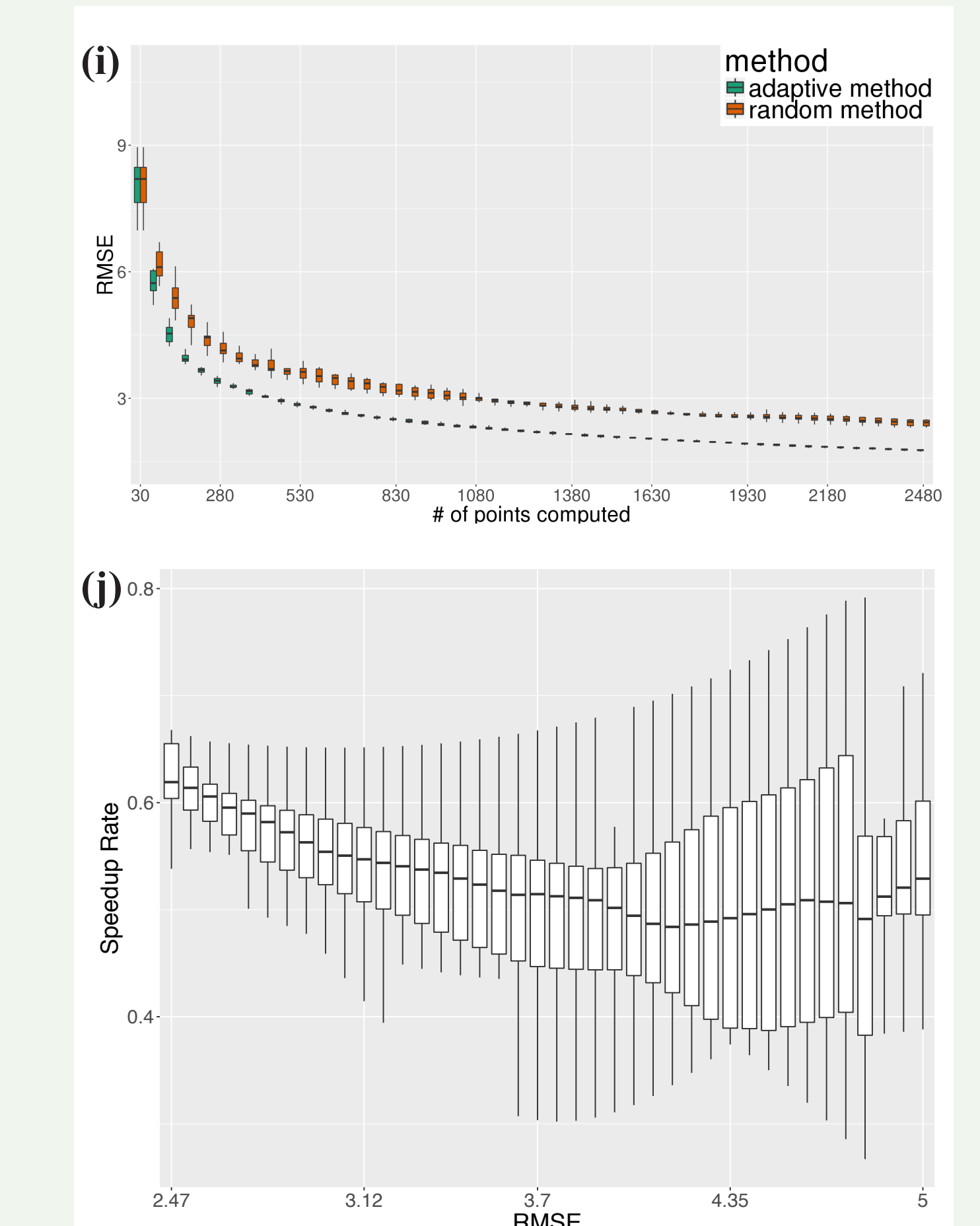


Figure (i) shows the distribution of Root Mean Square Errors (RMSE) of the adaptive and the random methods; figure (j) shows the speedup rate of the adaptive method compared defined below with the random method. Experiments are repeated 30 times for statistical accuracy.

$$\text{SpeedupRate} = \frac{\# \text{ of RandomPoints} - \# \text{ of DirectedPoints}}{\# \text{ of RandomPoints}}$$

Use the QR code to watch the full video on how two unstructured grids evolve differently.

