School of Engineering and Applied Science (SEAS), Ahmedabad University

CSE 400: Fundamentals of Probability in Computing

Milestone-1 Scribe

S2_G11_CLI

Sr.No	Enrollment No	Name	Email Id
1.	AU2440097	Chaitanya Jammula	jammula.m@ahduni.edu.in
2.	AU2440247	Jenil Modi	jenil.m1@ahduni.edu.in

Scribe Question 1: Project System and Objective

Probabilistic Problem: The project deals with the uncertainty in quantitative precipitation forecasting (very-short-range forecasting, 0 -6 hours). The phenomenon of precipitation is known to be inconsistent across a broad spectrum of space-time scales and this, in conjunction with the chaotic nature of the atmosphere, restricts the deterministic predictability of precipitation.

System Objective: This is to generate a group of precipitation forecasts (nowcasts) that is representative of the probability distribution of future rainfall intensity. The system is to be implemented as the `pysteps` library, which is supposed to measure predictive uncertainty to aid in decision-making regarding severe weather hazards (e.g., floods).

Primary Sources of Uncertainty:

- **Initial State Errors:** Error in prediction of the present rainfall field (R) and the motion (advection) field (V) of the radar data.
- **Model Errors:** Limitations of the Lagrangian persistence assumption (the assumption that rainfall patterns move without changing intensity) and the growth of perturbations due to chaotic dynamics.

Scribe Question 2: Key Random Variables and Uncertainty Modeling

Key Random Variables:

- $R(\mathbf{x}, t)$: The precipitation intensity field at location $\mathbf{x}$ and time t.
- $\mathbf{V}(\mathbf{x}, t)$: The advection (velocity) field describing the motion of precipitation.
- $N(\mathbf{x}, t)$: The stochastic perturbation field (noise) added to the model to simulate uncertainty.

Uncertainty Modeling: Uncertainty is modeled using the Short-Term Ensemble Prediction System (STEPS) approach. The precipitation field is decomposed into a multiplicative cascade of spatial scales (using Fast Fourier Transforms).

- The temporal evolution of each scale is assumed to be a stochastic process,namely an Auto-AR(1) or AR(2). The process may at mostbe an autoregressive order one (or two) process.

  • Noise Injection: Gaussian noise that isspatially correlated, summed at each scale for every timestep.

  The noise variance is tuned so that the lossof predictability due to lead time is correct.

  Probabilistic Assumptions:

  • Log-Normalnity: The rate of rain is modeled as alog-normal distribution. The model operates

  onstructured data dBR = 10 log10 R.

  • Errors are Gaussian: In transformed space, the predictionerrors and random noise terms n i are normal ran-

  tions are assumed to be gaussian.

Scribe Question 3: Probabilistic Reasoning and Dependencies

Probabilistic Relationships:

- **Temporal Dependence (Lagrangian Persistence):** The system presumes the state of the future to be. of a given feature of perception relies on its present condition and its displacement. This is modeled via the AR process:

$$\eta_k(t+1) = \rho_k \eta_k(t) + \sqrt{1-\rho_k^2}\epsilon_k(t)$$

  where η_k is the component at scale k, ρ_k is the Lagrangian autocorrelation (dependence), and ϵ_k is the stochastic noise (innovation).

- **Spatial Dependence:** Filtering of the noise term ek is done to maintain spatial correlations. that spectral density (S) of power of the observed field. This guarantees the produced ensembles. look realistic.

- **Independence:** The cascade levels (different spatial scales) are often treated as independent processes during the evolution step to simplify computation, though they are recombined to form the final field

Scribe Question 4: Model–Implementation Alignment

Alignment: The probabilistic model is in accordance with the implementation in terms of pysteps Python.

- **Decomposition:** FFT -based bandpass filters are used to implement the theoretical cascade model.

- **Motion Estimation:** The advection field V is approximated with the help of algorithms of Optical Flow (e.g., where Lucas-Kanade), which is consistent with the Lagrangian transport assumption.

- **Ensemble Generation:** The stochastic implementation produces N realizations (mem-ensembling). bers) by employing various random seeds of the noise generator, in a direct expression of the probability.

Assumptions Influencing Design:

- **Stationarity:**The model presupposes the statistical stationarity of the precipitation field in the short. nowcasting window (0-2 hours) to forecast the autoregressive parameters.

- **Frozen Field Hypothesis:** When the forecast is being done, the advection field is usually held constant (frozen). or extrapolated in a straight line, which has an impact on longer lead time evaluation.

Scribe Question 5: Cross-Milestone Consistency and Change

Current State (Well-Defined):

- The framework for **optical flow** motion estimation and **stochastic noise generation** (STEPS method) is well-defined and robust for short lead times (0–60 mins).

- The **scale decomposition** methodology (isolating large vs. small features) is a stable component of the probabilistic model.

Aspects to Evolve/Refine:

- **Blending with NWP:** Pure extrapolation fails after ≈ 2 hours. Future milestones (or advanced modules) involve blending these probabilistic nowcasts with Numerical Weather Prediction (NWP) ensembles to extend consistency up to 6 hours.
- **Non-Stationary Uncertainty**: The present global autocorrelation parameter assumption may develop into estimates that are spatially variable (local) in order to better manage intricate convective systems.

Scribe Question 6: Open Issues and Responsibility Attribution

Unresolved Probabilistic Questions:

- **Extreme Values:** The current log-normal and AR assumptions sometimes struggle to capture the rapid inception (growth) and decay of extreme convective cells (heavy rain), leading to "regression to the mean."
- **Non-Gaussian Features:** Precipitation is intermittent (mixed discrete-continuous distribution with many zeros). The current masking and transformation approach is an approximation that may introduce bias.

Responsibility:

- **Next Milestone Tasks:** It is the role of the Verification Module to measure these biases in form of probablistic scores (e.g., ROC curves, Reliability diagrams).
- **Future Development:** It is pointed out that the addition of Machine Learning (e.g., GANs) is a task that the next generation of research will need to undertake in order to take care of non-linear growth/decay, which the basic stochastic model cannot cover.