Milestone-2 Scribe

Scribe Question 1: Project System and Objective

In Milestone-1, our paper has firmly established that the deterministic Numerical Weather Prediction (NWP) rainfall forecasts are a failure to provide uncertainty representation in a direct way. We have also brought in the Bayesian Joint Probability (BJP) framework to express the probabilistic dependence between a forecast rainfall (X) and an observed rainfall (Y).

In Milestone-2, we extend the notion of modeling to the actual derivation of the mathematical expression of this relationship.

The probabilistic problem being addressed is the estimation of the conditional exceedance probability :-

$$P(Y > t \mid X = x)$$

where t is a predefined rainfall threshold relevant for agricultural decision-making.

The work being done here is to develop a joint probability model for the (X, Y) variables at a formally specified level, figure out the conditional distribution analytically and then to quantify the risk of the exceedance in presence of the uncertainty. This transformation takes the system from a qualitative uncertainty representation (Milestone-1) to a mathematically rigorous predictive framework.

The main sources of uncertainty still stay the same :-

- **Atmospheric variability:** the intrinsic stochasticity of monsoon systems.
- **Forecast uncertainty:** the variability of the ensemble in NWP outputs.
- **Measurement uncertainty:** the noise of rainfall observation.

- **Parameter uncertainty:** the uncertainty in estimated means, variances, and correlation.

In this milestone, all these uncertainties are made clear orally and incorporated into a bivariate Gaussian joint model, which through conditional inference, are taken forward.

Scribe Question 2: Key Random Variables and Uncertainty Modeling

The main random variables are still the same as in Milestone-1 and now they have been given formal parameterization.

1. Forecast Rainfall (X)

Forecast rainfall is modeled as :-

$$X \sim \mathcal{N}(\mu_X, \sigma_X^2)$$

where the μ_X and the σ_X^2 respectively represent the mean and variance of transformed forecast rainfall. A monotonic transformation (e.g., log or log-sinh) is applied here to approximate Gaussian behavior. The variance σ_X^2 captures predictive spread and forecast uncertainty.

Assumption

- The transformed predictions are approximately normally distributed, i.e.,

$$X \sim \mathcal{N}(\mu_X, \sigma_X^2).$$

- The relationship between forecasts and observations remains stable during the calibration period, implying constant parameters

$$(\mu_X, \mu_Y, \sigma_X^2, \sigma_Y^2, \rho).$$

2. Observed Rainfall (Y)

Observed rainfall is modeled as :-

$$Y \sim \mathcal{N}(\mu_Y, \sigma_Y^2)$$

where the μ_Y and the σ_Y^2 capture intrinsic climatic variability and measurement uncertainty after transformation.

Assumption

- After transformation, the rainfall distribution is sufficiently symmetric to justify the use of Gaussian modeling for both X and Y.

3. Joint Structure of (X, Y)

We assume a bivariate normal distribution :-

$$(X, Y) \sim \mathcal{N}_2\left(\begin{pmatrix}\mu_X \\ \mu_Y\end{pmatrix}, \begin{pmatrix}\sigma_X^2 & \rho\sigma_X\sigma_Y \\ \rho\sigma_X\sigma_Y & \sigma_Y^2\end{pmatrix}\right)$$

where ρ represents the correlation between forecast and observed rainfall and quantifies forecast skill.

4. Threshold Exceedance Indicator (Z)

Extreme rainfall events are defined as following :-

$$Z = \mathbf{1}(Y > t)$$

Therefore ,

$$Z \sim \text{Bernoulli}(p(x))$$

where ,

$$p(x) = P(Y > t \mid X = x)$$

This connects continuous probabilistic modeling to agricultural decision-making.

5. Parameter Uncertainty

Let

$$\theta = (\mu_X, \mu_Y, \sigma_X^2, \sigma_Y^2, \rho)$$

Within a Bayesian framework, inference is based on :-

$$\pi(\theta \mid \text{data}) \propto L(\theta)\pi(\theta)$$

This will make sure that exceedance probabilities take into account both climatic variability and parameter estimation uncertainty.

Scribe Question 3: Probabilistic Reasoning and Dependencies

Milestone-1 conceptually described dependence. In Milestone-2, this dependence is formalized mathematically.

1. Dependence Structure

Forecast and observed rainfall are modeled as dependent random variables with :-

$$\mathrm{Corr}(X, Y) = \rho$$

If the correlation coefficient $\rho = 0$, the forecast simply does not have any predictive information. As the magnitude of the correlation $|\rho|$ increases, the skill of the forecast improves and the predictive uncertainty becomes less.

2. Conditional Distribution

From the properties of the bivariate normal distribution :-

$$Y \mid X = x \sim \mathcal{N}(\mu_{Y|x}, \sigma^2_{Y|x})$$

where,

$$\mu_{Y|x} = \mu_Y + \rho \frac{\sigma_Y}{\sigma_X}(x - \mu_X)$$

$$\sigma^2_{Y|x} = \sigma^2_Y(1 - \rho^2)$$

The factor $(1, \rho^2)$ reduces the conditional variance, thus mathematically expressing the extent to which forecast information tightens predictive uncertainty.

3. Exceedance Probability

The exceedance probability is given by :-

$$P(Y > t \mid X = x) = 1 - \Phi\left(\frac{t - \mu_{Y|x}}{\sigma_{Y|x}}\right)$$

where $\Phi(\cdot)$ denotes the standard normal cumulative distribution of the function.
This shows the following :-

- Forecast deviation affects the conditional mean.
- Correlation reduces conditional variance.
- The conditional distribution determines exceedance risk.

Thus, the chain of probabilistic reasoning is :-

Joint Model $\rightarrow$ Conditional Distribution $\rightarrow$ Exceedance Probability $\rightarrow$ Decision Support

4. Simulation-Based Reasoning

In order to put the model into practice, one may generate correlated variables from independent standard normal variables $Z_1, Z_2 \sim \mathcal{N}(0, 1)$:-

$$X = \mu_X + \sigma_X Z_1$$

$$Y = \mu_Y + \sigma_Y \left(\rho Z_1 + \sqrt{1 - \rho^2} Z_2\right)$$

This enables Monte Carlo estimation of exceedance probabilities and showcase how uncertainty is transferred through the system.

Scribe Question 4: Model–Implementation Alignment

The current probabilistic model is based on a bivariate normal joint distribution between transformed forecast rainfall X and transformed observed rainfall Y:

$$\begin{pmatrix} X \\ Y \end{pmatrix} \sim \mathcal{N}_2\left(\begin{pmatrix} \mu_X \\ \mu_Y \end{pmatrix}, \begin{pmatrix} \sigma_X^2 & \rho\sigma_X\sigma_Y \\ \rho\sigma_X\sigma_Y & \sigma_Y^2 \end{pmatrix}\right)$$

Implementation Alignment

The implementation aligns with the model in the following ways:

1. Historical paired data $\{(x_i, y_i)\}_{i=1}^n$ are used to estimate parameters

   $$\theta = (\mu_X, \mu_Y, \sigma_X^2, \sigma_Y^2, \rho)$$

   via maximum likelihood estimation:

   $$\ell(\theta) = \sum_{i=1}^{n} \log f_{X,Y}(x_i, y_i; \theta)$$

2. Conditional prediction is computed analytically using multivariate normal theory:

   $$Y \mid X = x \sim \mathcal{N}(\mu_{Y|x}, \sigma_{Y|x}^2)$$

   where

   $$\mu_{Y|x} = \mu_Y + \rho \frac{\sigma_Y}{\sigma_X}(x - \mu_X)$$

   $$\sigma_{Y|x}^2 = \sigma_Y^2(1 - \rho^2)$$

3. Extreme rainfall risk for agriculture is evaluated using exceedance probability:

   $$P(Y > t \mid X = x) = 1 - \Phi\left(\frac{t - \mu_{Y|x}}{\sigma_{Y|x}}\right)$$

Key Assumptions Affecting Implementation

- Joint normality assumption after rainfall transformation.
- Linear dependence between forecast and observed rainfall.
- Constant correlation parameter ρ across time.
- Stationarity of rainfall distribution.

These assumptions influence model evaluation, simulation design, and forecast calibration decisions.

Scribe Question 5: Cross-Milestone Consistency and Change

Current State of the Probabilistic Model

In Milestone 1, the project conceptually aimed to quantify conditional extreme rainfall probability for agricultural planning.

In Milestone 2, this objective has been formalized into a mathematically defined joint probability model:

$$P(Y > t \mid X = x) = 1 - \Phi\left(\frac{t - \mu_{Y|x}}{\sigma_{Y|x}}\right)$$

where

$$\mu_{Y|x} = \mu_Y + \rho\frac{\sigma_Y}{\sigma_X}(x - \mu_X), \quad \sigma^2_{Y|x} = \sigma^2_Y(1 - \rho^2)$$

Well-Defined Components

The following components are now mathematically specified:

- Random variables X and Y.
- Joint distribution structure.
- Dependence parameter ρ as forecast skill measure.
- Analytical exceedance probability.
- Likelihood-based parameter estimation.

Expected Refinements in Future Milestones

The following aspects may evolve:

1. Non-normal rainfall modeling (e.g., heavy tails).
2. Extreme value theory for high thresholds.
3. Time-varying correlation ρ_t.
4. Bayesian parameter uncertainty quantification:

$$p(\theta \mid \text{data}) \propto L(\theta)\pi(\theta)$$

5. Crop-specific loss functions

$$\mathbb{E}[L(Y)]$$

for decision optimization.

Justification

Milestone 2 improves consistency by transforming the conceptual forecast-risk idea into a fully specified probabilistic system. Future milestones are expected to refine distributional assumptions, extreme-event modeling, and decision-theoretic integration.

Scribe Question 6: Open Issues and Responsibility Attribution

At the current phase of Milestone, 2, the probabilistic framework has been formally defined through a bivariate Gaussian joint model and conditional exceedance derivation. Nevertheless, a number of probabilistic questions and modeling ambiguities have not yet been resolved.

1. Tail Behavior and Extreme Rainfall Modeling

Indeed, while transformation makes data more symmetric, the assumption of a Gaussian distribution may downplay the presence of a heavy upper tail during extreme monsoon events. Given that extreme rainfall events greatly influence agricultural risk decisions, it is still uncertain whether the normal assumption is adequate in the tails of the distribution.

Responsible Role: Modeling team to explore heavy, tailed alternatives (e. g. , t, distributions, extreme value extensions) and evaluate tail calibration performance.

2. Stationarity Assumption

The current model assumes that parameters $(\mu_X, \mu_Y, \sigma_X, \sigma_Y, \rho)$ remain constant during the calibration period. The point is that monsoon systems can be undergoing regime changes or long, term climatic variations. In case the relationship between the forecast and the observation changes, the premise of a constant correlation () may be invalid.

Responsible Role: Data and modeling teams to assess temporal stability and develop non, stationary or seasonally varying parameter models.

3. Parameter Uncertainty Propagation

Although a Bayesian likelihood framework has been introduced, full integration of posterior uncertainty into predictive exceedance intervals requires further computational validation. In particular, the sensitivity of

$$P(Y > t \mid X = x)$$

to prior choices and limited data must be examined.

Responsible Role: Inference and validation team to perform posterior diagnostics and uncertainty quantification experiments.

4. Dependence Structure Adequacy

The model presently presumes that the linear dependence is sufficiently characterised by the correlation . Nevertheless, it is possible that rainfall relations include nonlinear or asymmetric dependence features.

This would mean that one might ask whether the Gaussian covariance structure really captures forecastobservation dependence without the need for additional features.

Responsible Role: Modeling team to explore alternative dependence structures (e. g. , copula, based models) if empirical diagnostics suggest deviations.

5. Monte Carlo Convergence and Simulation Stability

The Monte Carlo exceedance estimator relies on the number of simulations (N) used. The stability and rate of convergence of the estimates of exceedance probabilities should be systematically evaluated.

Responsible Role: Implementation and validation team to conduct convergence analysis and document computational reliability.