https://doi.org/10.5194/egusphere-2026-3482
Preprint. Discussion started: 26 June 2026

Uncertainty of Rainfall Forecasts for Impact-Based Flood Warning in the Cagayan River Basin, Philippines

Yuta Kurihara[1,2,3] and Mamoru Miyamoto[1,2]

[1]National Graduate Institute for Policy Studies (GRIPS), Tokyo 106-0032, Japan
[2]International Centre for Water Hazard and Risk Management (ICHARM), PWRI, Tsukuba 300-2621, Japan
[3]Oriental Consultants Global Co., Ltd., Tokyo 163-1409, Japan

Correspondence: Yuta Kurihara (kurihara.yuta@gmail.com)

Abstract. Effective impact-based flood early warning requires not only information on when heavy rainfall may occur, but also a calibrated estimate of the uncertainty in the resulting impacts. This study develops a framework for translating multi-centre sub-seasonal to seasonal (S2S) rainfall forecasts into probabilistic, municipality-level impact-based early-warning information for staged flood preparedness in the Cagayan River Basin, Philippines. Daily Gamma-kernel Bayesian model averaging is first applied to ensemble forecasts from ECMWF, NCEP, and UKMO to generate continuous predictive distributions of basin-mean rainfall. Leave-one-year-out verification for 2015–2025 identifies the ECMWF+NCEP+UKMO combination as the most robust tested input, with useful daily rainfall information mainly retained up to approximately lead days 5–6. The daily predictive samples are then accumulated into rolling seven-day rainfall distributions, because flood impacts in the basin are more closely related to multi-day rainfall than to isolated daily totals. Threshold-based probability recalibration improves the reliability of seven-day exceedance probabilities, raising Brier Skill Scores from -0.02 to $+0.08$ at 100 mm per seven days and from -0.07 to $+0.02$ at 150 mm per seven days relative to a monthly climatological baseline. The recalibrated seven-day rainfall distributions are subsequently propagated through municipality-level rainfall–damage functions to estimate probabilistic impacts on affected population, building damage, rice damage, and maize damage. An application to Typhoon Ulysses in November 2020 demonstrates how forecast-state-dependent impact intervals evolve as the event approaches and how municipalities with potentially large impacts can be prioritised. The results show that calibrated S2S rainfall probabilities can support uncertainty-aware, impact-based flood preparedness, while also highlighting limitations related to lead-time skill, basin-mean rainfall representation, upper-tail rainfall coverage, and the validation of rainfall–damage functions.

1 Introduction

Flood early-warning systems have helped reduce losses of lives and assets. Information at different lead times supports different preparedness decisions: weeks-ahead information can support monitoring, inter-agency coordination, and precautionary planning; days-ahead information can support the protection of exposed assets and preparation of local response measures; and hours-ahead information mainly supports emergency response and evacuation. However, unnecessary or false-alarm preparations can impose financial, operational, and psychological burdens on communities and agencies. Therefore, rainfall informa-

https://doi.org/10.5194/egusphere-2026-3482
Preprint. Discussion started: 26 June 2026

tion at multi-day to two-week lead times should be expressed probabilistically and evaluated in terms of uncertainty, reliability, and calibration, so that it can be translated into impact-relevant information for staged preparedness.

The Cagayan River Basin in northern Luzon—the largest river basin in the Philippines, with an aggregate basin area of 27,281 km^2—is repeatedly affected by typhoon- and monsoon-driven flooding. Flood impacts in the basin affect people, buildings, agricultural production, and public infrastructure, and are shaped not only by daily rainfall peaks but also by multi-day rainfall accumulation, catchment response, and the spatial distribution of exposed assets. Early-warning information is therefore most useful when it provides both lead time and an interpretable estimate of possible consequences (Coughlan de Perez et al., 2015; White et al., 2017; Kurihara and Miyamoto, 2026).

Sub-seasonal to seasonal (S2S) ensemble forecasts provide information at the multi-day to two-week range through the S2S prediction database (Vitart et al., 2017, 2025). Raw S2S precipitation forecasts typically show systematic differences from observed precipitation, so statistical post-processing is commonly applied to obtain calibrated probabilistic rainfall information suitable for preparedness use. Relevant approaches include Bayesian joint-probability methods (Li et al., 2021), probabilistic machine learning for extremes (Zhan et al., 2025), and Bayesian model averaging (BMA; Raftery et al., 2005), which has been widely used for multi-model and precipitation ensembles (Sloughter et al., 2007; Fraley et al., 2010). These probabilities also need to be interpreted with respect to preparedness actions whose cost, reversibility, and avoided loss differ by lead time (Lala et al., 2022; White et al., 2022).

The contribution of this study is neither a new hydrological–hydraulic flood model nor a new BMA formulation. Instead, we develop a diagnostic framework that carries calibrated multi-centre S2S rainfall uncertainty forward into probabilistic, municipality-level impact-based early-warning information for staged flood preparedness. Daily Gamma-kernel BMA is used as a probabilistic rainfall model that makes members of S2S ensemble forecasts into continuous daily predictive distributions. These distributions are accumulated into rolling seven-day rainfall totals, recalibrated as exceedance probabilities, and propagated through municipality-level rainfall–damage functions to estimate uncertain impacts on affected population, buildings, rice, and maize. The goal is to support phased preparations by estimating the areas expected to be affected, together with the associated uncertainty.

Three questions guide the analysis:

RQ1. Does combining multiple S2S forecast models provide more robust post-processed daily probabilistic rainfall forecasts than using a single S2S forecast model?

RQ2. Does recalibration improve rolling seven-day exceedance probabilities, and can the resulting predictive distribution express rainfall uncertainty at the preparedness-relevant time scale?

RQ3. Can rainfall-forecast uncertainty be propagated into damage uncertainty to support municipality-level, impact-based flood early-warning information?

The framework is used to examine whether post-processed outputs from forecast models can represent damage uncertainty and provide municipality-level guidance for staged preparedness decisions. These outputs are interpreted in relation to

https://doi.org/10.5194/egusphere-2026-3482
Preprint. Discussion started: 26 June 2026

lead-time skill, upper-tail ensemble signal, the basin-mean rainfall assumption, and the simulation-derived rainfall–damage functions, which also define the main pathways for improving practical usefulness.

2 Methods

2.1 Overall framework

The framework consists of four steps. First, candidate combinations of S2S forecast models are screened after applying the same daily Bayesian model averaging (BMA) post-processing model to each combination. This step identifies the multi-centre input that provides the most skilful post-processed daily probabilistic rainfall forecasts. Second, the selected ensemble is converted into lead-day-specific daily rainfall predictive distributions using Gamma-kernel BMA. Third, daily predictive samples are accumulated into rolling seven-day rainfall distributions and recalibrated, because flood preparedness is more closely related to multi-day accumulated rainfall than to isolated daily rainfall. Fourth, the recalibrated seven-day rainfall distributions are propagated through municipality-level rainfall–damage functions to obtain probabilistic impact information for preparedness planning.

2.2 Study area and dataset

The observational rainfall reference is basin-mean daily rainfall ($\mathrm{mm\,day^{-1}}$) derived from a bias-corrected GSMaP v8_MVK precipitation dataset prepared following Kurihara et al. (2025). This basin-mean reference is used because the study focuses on large-scale flood preparedness in the Cagayan River Basin and because the S2S forecasts are available at relatively coarse spatial resolution. The gridded 0.1° daily rainfall fields were corrected using ground rain-gauge observations and spatially averaged over the Cagayan River Basin. Rolling seven-day observed rainfall is computed by summing this daily reference over the same lead-day windows as the forecast.

Real-time ensemble forecasts were obtained from the S2S database at 0.5° resolution. Three forecast centres were used as candidates: ECMWF, NCEP, and UKMO, representing the European, United States, and United Kingdom operational extended-range ensemble prediction systems, respectively. These centres were selected because they provide sufficiently long and comparable real-time forecast archives over the study period and are widely used in S2S prediction studies. The period January 2015–December 2025 was used. ECMWF and NCEP provided 1,148 Monday/Thursday issue dates each, while UKMO was available from December 2015, giving 1,047 issue dates. JMA was not included because the available real-time issue sequence over 2015–2025 was not sufficiently consistent with the Monday/Thursday verification sample used here; including it would have changed the verification sample and reduced the comparability of the centre-combination screening. Lead days 1–14 were evaluated throughout, and forecasts were converted to member-specific basin-mean daily rainfall over the Cagayan River Basin.

Figure 1 shows the Region II study area, with the Cagayan River Basin, the major river network, and selected municipalities (panel a), together with the active S2S 0.5° grid cells and the corresponding basin-averaging weights (panel b).

https://doi.org/10.5194/egusphere-2026-3482
Preprint. Discussion started: 26 June 2026

Table 1. Candidate S2S forecast centres used in the forecast-source screening.

Centre	Institution	Members	Issue dates	Real-time period
ECMWF	ECMWF (EU)	51	1,148	Jan 2015–Dec 2025
NCEP	NOAA (USA)	16	1,148	Jan 2015–Dec 2025
UKMO	Met Office (UK)	4	1,047	Dec 2015–Dec 2025

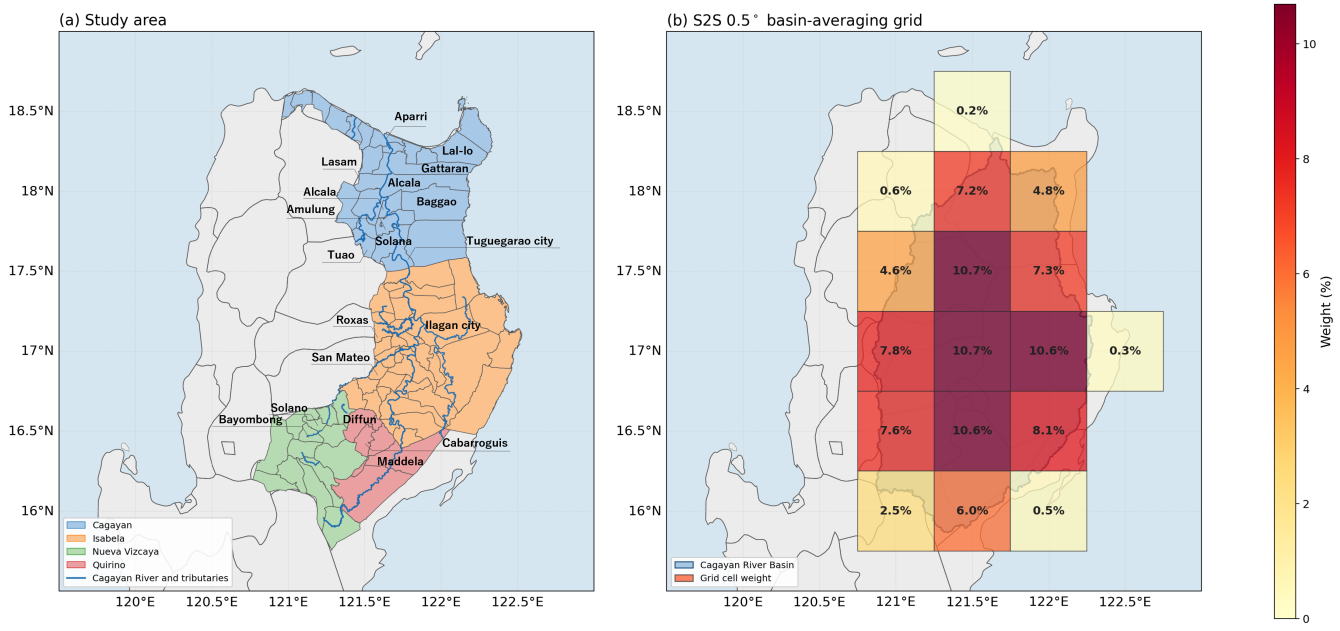

Figure 1. (a) Study area: Region II of the Philippines, municipality boundaries coloured by province, the Cagayan River and its tributaries, and some municipalities labelled for reference. (b) S2S 0.5° grid cells overlying the basin, coloured by their basin-averaging weight.

2.3 Forecast-source screening and verification scores

The forecast-source screening identifies the centre combination that gives the most skilful post-processed daily probabilistic rainfall forecasts for the subsequent seven-day aggregation and impact translation. All combinations of ECMWF, NCEP, and UKMO were tested. For each candidate combination, the same daily Gamma-kernel BMA post-processing model was refitted under leave-one-year-out cross-validation (LOOCV), and the resulting daily predictive distributions were verified against the observed basin-mean rainfall. In LOOCV, each verification year is held out in turn while the remaining years are used to fit the BMA parameters, so the verification scores reflect out-of-sample performance and avoid inflation from in-sample reuse. The detailed BMA formulation is given in Sect. 2.4; this subsection defines the verification scores used for screening.

The primary screening metric is the continuous ranked probability score (CRPS; Hersbach, 2000; Gneiting and Raftery, 2007), which compares the predictive rainfall distribution with the corresponding observed rainfall amount. CRPS is defined

https://doi.org/10.5194/egusphere-2026-3482
Preprint. Discussion started: 26 June 2026

as

$$\mathrm{CRPS}(F_{t,\ell}, y_{t,\ell}) = \int_{-\infty}^{\infty} \left(F_{t,\ell}(x) - \mathbf{1}\{x \geq y_{t,\ell}\}\right)^2 \mathrm{d}x. \tag{1}$$

Because the BMA forecasts are represented by predictive samples, CRPS is computed using the empirical form:

$$\mathrm{CRPS}_{t,\ell} = \frac{1}{N}\sum_{i=1}^{N} |X_{t,\ell}^{(i)} - y_{t,\ell}| - \frac{1}{2N^2}\sum_{i=1}^{N}\sum_{j=1}^{N} |X_{t,\ell}^{(i)} - X_{t,\ell}^{(j)}|, \tag{2}$$

For forecast issue date t and lead day ℓ, $F_{t,\ell}$ is the predictive cumulative distribution function, $y_{t,\ell}$ the observed basin-mean rainfall, and $X_{t,\ell}^{(i)}$ and $X_{t,\ell}^{(j)}$ are predictive rainfall samples from the same distribution; N is the number of samples. The first term measures the average distance between the prediction samples and the observation, while the second term accounts for the spread of the predictive distribution by averaging all $N \times N$ sample-pair distances. Lower CRPS indicates a better probabilistic forecast.

As a threshold-based diagnostic, we use the Brier Skill Score (BSS) for daily rainfall exceeding $20\,\mathrm{mm\,day^{-1}}$. This diagnostic evaluates whether the forecast assigns higher probabilities to heavy-rainfall days and lower probabilities to non-event days, relative to a monthly climatological baseline.

For a rainfall threshold T, the forecast exceedance probability for case i is $p_i = \hat{P}(R_i > T)$, and the observed outcome is $o_i = \mathbf{1}\{y_i > T\}$, where $o_i = 1$ if the observed rainfall exceeds T and $o_i = 0$ otherwise. The Brier score is

$$\mathrm{BS} = \frac{1}{N}\sum_{i=1}^{N}(p_i - o_i)^2. \tag{3}$$

Because BS is the squared error between the forecast probability and the observed binary outcome, smaller values indicate better agreement and larger values indicate poorer probabilistic forecasts.

The climatological reference probability is computed within each leave-one-year-out fold from the training years only. For a target case in calendar month $m(i)$, $p_{\mathrm{clim},m(i)}$ is the fraction of training-year observations in the same month that exceed T. The corresponding climatological Brier score is

$$\mathrm{BS}_{\mathrm{clim}} = \frac{1}{N}\sum_{i=1}^{N}(p_{\mathrm{clim},m(i)} - o_i)^2. \tag{4}$$

The Brier Skill Score is then

$$\mathrm{BSS} = 1 - \frac{\mathrm{BS}}{\mathrm{BS}_{\mathrm{clim}}}. \tag{5}$$

A positive BSS indicates skill relative to the monthly climatology. For the daily source screening, $T = 20\,\mathrm{mm\,day^{-1}}$, and all forecast–observation pairs are used, not only days on which the observed rainfall exceeds the threshold. No parametric distribution is fitted to the climatological reference.

https://doi.org/10.5194/egusphere-2026-3482
Preprint. Discussion started: 26 June 2026

2.4 Daily BMA probabilistic rainfall model

Daily BMA is used to convert the selected S2S ensemble forecasts into continuous, lead-day-resolved rainfall predictive distributions. Statistical post-processing is used to improve the reliability and probabilistic skill of S2S rainfall forecasts, as shown in previous precipitation post-processing studies (Sloughter et al., 2007; Li et al., 2021). We adopt BMA because it provides a flexible and established framework for combining multi-centre ensemble members into continuous predictive rainfall distributions (Raftery et al., 2005).

The BMA model is fitted separately for each lead day ℓ. This allows the centre weights, rainfall spread, and zero-rainfall probability to vary with lead time, reflecting the fact that forecast errors and ensemble reliability change across the forecast horizon. For a candidate set of C forecast centres, the lead-day-specific parameter vector is

$$\theta_\ell = \left(\{w_c^{(\ell)}\}_{c=1}^{C}, \{\sigma_c^{2(\ell)}, d_{0,c}^{(\ell)}, d_{1,c}^{(\ell)}\}_{c=1}^{C}\right), \tag{6}$$

where $w_c^{(\ell)}$ are non-negative centre weights summing to one, $\sigma_c^{2(\ell)}$ controls the spread of the positive-rainfall kernel for centre c, and $d_{0,c}^{(\ell)}$ and $d_{1,c}^{(\ell)}$ control its zero-rainfall probability.

For each ensemble member forecast, positive rainfall is represented by a Gamma kernel centred on the raw member forecast. The Gamma kernel is adopted because positive rainfall is non-negative and right-skewed. This choice does not assume that the overall observed rainfall climatology follows a Gamma distribution; rather, the Gamma kernel represents the conditional uncertainty of observed rainfall around each ensemble member forecast. Because the Gamma distribution requires a positive mean, near-zero forecasts are replaced by a small numerical floor, $x^* = \max(x_{c,m}, \varepsilon)$ with $\varepsilon = 0.1\,\mathrm{mm}$:

$$g(y \mid x_{c,m}, \sigma_c^{2(\ell)}) = \mathrm{Gamma}\left(y; \mathrm{shape} = \frac{(x^*)^2}{\sigma_c^{2(\ell)}}, \mathrm{rate} = \frac{x^*}{\sigma_c^{2(\ell)}}\right). \tag{7}$$

Here, $x_{c,m}$ denotes the basin-mean daily rainfall forecast from ensemble member m of forecast centre c at the target lead day. This parameterisation gives the Gamma kernel mean x^* and variance $\sigma_c^{2(\ell)}$. Thus, $\sigma_c^{2(\ell)}$ controls how widely observed rainfall can vary around the member forecast for centre c at lead day ℓ.

Zero rainfall is treated separately from positive rainfall. For each centre, the probability of zero observed rainfall is modelled as a logistic function of the member forecast:

$$p_{0,c}(x_{c,m}) = \frac{1}{1 + \exp[-(d_{0,c}^{(\ell)} + d_{1,c}^{(\ell)} x_{c,m})]}. \tag{8}$$

The coefficients $d_{0,c}^{(\ell)}$ and $d_{1,c}^{(\ell)}$ control how the probability of no observed rainfall changes with the forecast rainfall amount.

Combining these components, the predictive density for observed basin-mean daily rainfall y at lead day ℓ is

$$p(y \mid \{x_{c,m}\}, \theta_\ell) = \sum_{c=1}^{C} w_c^{(\ell)} \frac{1}{M_c} \sum_{m=1}^{M_c} \left[\{1 - p_{0,c}(x_{c,m})\} g(y \mid x_{c,m}, \sigma_c^{2(\ell)}) + p_{0,c}(x_{c,m}) \delta_0(y)\right], \tag{9}$$

where M_c is the number of members in centre c and $\delta_0(y)$ is a point mass at zero rainfall. Members within each centre are weighted equally, while the centres receive estimated BMA weights. During the period before UKMO forecasts became

https://doi.org/10.5194/egusphere-2026-3482
Preprint. Discussion started: 26 June 2026

available, the BMA predictive distribution is formed from ECMWF and NCEP only, with their relative centre weights preserved and rescaled to sum to one.

Given the training pairs at lead day ℓ, the BMA parameters are estimated by maximum a posteriori (MAP) estimation. MAP estimation is used to obtain stable point estimates while applying weakly informative priors to avoid unrealistic parameter values under the limited heavy-rainfall sample:

$$\log \sigma_c^{2(\ell)} \sim \mathcal{N}(\log 100, 1.5^2), \qquad d_{0,c}^{(\ell)} \sim \mathcal{N}(0, 3^2), \qquad d_{1,c}^{(\ell)} \sim \mathcal{N}(-0.05, 0.5^2). \tag{10}$$

With a symmetric prior on the centre weights, the estimated parameter vector is

$$\hat{\theta}_\ell = \arg\max_{\theta_\ell} \left[\sum_{i=1}^{N_\ell} \log p(y_i \mid \{x_{i,c,m}\}, \theta_\ell) + \log p(\theta_\ell) \right], \tag{11}$$

where N_ℓ is the number of training forecast–observation pairs at lead day ℓ. The first term is the log-likelihood of the observed rainfall under the BMA predictive density, and the second term is the log-prior contribution.

The objective is maximised using L-BFGS-B (Byrd et al., 1995), and θ_ℓ is re-estimated within each leave-one-year-out cross-validation fold. The full per-lead-day kernel parameters and centre weights are reported in Appendix A.

2.5 Rolling seven-day accumulated rainfall distributions

Rolling seven-day accumulated rainfall is used because flood preparedness and damage are more closely related to multi-day rainfall than to a single daily total. For each forecast issue date t, daily BMA predictive samples for lead days 1–14 are summed into eight rolling seven-day windows:

$$R_{7,t,s} = \sum_{\ell=s}^{s+6} R_{t,\ell}, \qquad s = 1, \ldots, 8, \tag{12}$$

corresponding to D 1–7, D 2–8, …, D 8–14.

The summation is performed sample by sample, producing an empirical predictive distribution of seven-day rainfall for each window. This distribution is used as the main input for recalibration and impact translation because it preserves the daily evolution of forecast risk while expressing uncertainty at the preparedness-relevant seven-day scale.

Two sensitivity checks are also performed. First, an independence variant samples each lead day's BMA distribution independently before summation. Second, a direct seven-day BMA is fitted to rolling seven-day totals:

$$x_{c,m,t,s}^{(7)} = \sum_{\ell=s}^{s+6} x_{c,m,t,\ell}, \qquad y_{t,s}^{(7)} = \sum_{\ell=s}^{s+6} y_{t+\ell}. \tag{13}$$

These checks are used only to assess robustness.

For verification, the monthly climatology is constructed from observed rolling seven-day accumulations in the same calendar month of the training years. The raw-ensemble benchmark is computed by accumulating each ensemble member over the same seven-day window.

https://doi.org/10.5194/egusphere-2026-3482
Preprint. Discussion started: 26 June 2026

2.6 Threshold recalibration of seven-day rainfall probabilities

Recalibration is applied to seven-day accumulated rainfall probabilities because the raw accumulated BMA probabilities may still be miscalibrated at preparedness-relevant thresholds. For a seven-day threshold T_7, the raw exceedance probability is estimated from the accumulated BMA samples as

$$\hat{P}_{\mathrm{raw}}(R_{7,t,s} > T_7) = \frac{1}{N_s} \sum_{j=1}^{N_s} \mathbf{1}(R_{7,t,s}^{(j)} > T_7). \tag{14}$$

Let $x \in [0,1]$ denote this raw seven-day exceedance probability. The recalibrated probability is

$$p_{\mathrm{cal}} = 1 - (1 - x^a)^b, \tag{15}$$

where (a, b) are estimated from binned reliability pairs within the LOOCV training folds. The function is monotone and preserves 0 and 1, so it adjusts the probability scale without changing the ranking of forecasts within each fitted fold.

For distributional impact calculations, the same mapping is applied to the empirical survival function of the seven-day rainfall distribution:

$$S_{\mathrm{cal}}(r) = 1 - \{1 - S_{\mathrm{raw}}(r)^a\}^b, \qquad F_{\mathrm{cal}}(r) = 1 - S_{\mathrm{cal}}(r), \tag{16}$$

where $S_{\mathrm{raw}}(r) = P_{\mathrm{raw}}(R_7 > r)$. This yields a threshold-recalibrated empirical seven-day rainfall distribution that can be sampled and passed through the rainfall–damage functions. Because the recalibration parameters are estimated from threshold exceedance reliability, the survival-function adjustment should be interpreted as threshold-informed recalibration rather than full distributional calibration. The recalibration is fitted separately by rolling-window group (W1–W3, W4–W6, W7–W8), pooling the reliability pairs across the seven-day thresholds (100, 150, and 200 mm per seven days) for stability; where a group holds too few events the mapping defaults to the identity rather than an unstable curve.

2.7 Rainfall–damage functions and impact translation

Rainfall–damage functions are used to propagate recalibrated seven-day rainfall uncertainty into impact uncertainty. Following the uncertainty-propagation concept of Kurihara and Miyamoto (2026), recalibrated rainfall samples are passed through municipality-level damage functions rather than converting a single deterministic rainfall value into damage.

Municipality-level rainfall–damage functions are fitted from flood-damage simulations that report damage for each municipality i, impact category k, and rainfall scenario from 50 to 800 mm. Because many municipalities show no simulated damage below a rainfall threshold, we use a threshold-piecewise quadratic function:

$$D_{i,k}(R) = \begin{cases} 0, & R < R_{i,k}^{\mathrm{thr}}, \\ a_{i,k}R^2 + b_{i,k}R + c_{i,k}, & R_{i,k}^{\mathrm{thr}} \le R \le R_{i,k}^{\max}, \\ D_{i,k}^{\max}, & R > R_{i,k}^{\max}, \end{cases} \tag{17}$$

https://doi.org/10.5194/egusphere-2026-3482
Preprint. Discussion started: 26 June 2026

where R is seven-day accumulated rainfall, $R_{i,k}^{\mathrm{thr}}$ is the damage-onset threshold, $R_{i,k}^{\max}$ is the largest simulated rainfall scenario, and $D_{i,k}^{\max}$ is the corresponding maximum simulated damage. The quadratic part is fitted by ordinary least squares to the above-threshold simulation points and is constrained to start from zero damage at $R_{i,k}^{\mathrm{thr}}$. The four impact categories are affected population, building damage, rice damage, and maize damage. Functions are fitted only where at least three above-threshold rainfall–damage points are available; fitted coefficients, R^2, and rainfall ranges are reported in Appendix B.

Each recalibrated seven-day rainfall sample $R_7^{(j)}$ is then mapped to a damage sample:

$$D_{i,k}^{(j)} = \mathrm{clip}\left(D_{i,k}\left(R_7^{(j)}\right), 0, D_{i,k}^{\max}\right). \tag{18}$$

The resulting Monte Carlo damage samples form an impact distribution for each municipality and category, summarised by the expected impact, median, predictive intervals, and high-impact quantiles. Basin-level impact distributions are obtained by summing municipality-level samples. These impact distributions are interpreted as preparedness-oriented information, not as independently validated loss forecasts.

3 Results

3.1 Daily BMA as the probabilistic rainfall model

To answer RQ1, we first evaluate whether combining multiple S2S forecast centres improves the skill of post-processed daily probabilistic rainfall forecasts. The seven candidate inputs are the three single-centre forecasts, the three two-centre combinations, and the three-centre ECMWF+NCEP+UKMO combination. For each candidate input, the same daily Gamma-kernel BMA model is fitted under leave-one-year-out cross-validation and evaluated using CRPS and BSS at $20\,\mathrm{mm\,day^{-1}}$.

Table 2 shows the pooled verification scores over all lead days. The monthly climatology benchmark has a CRPS of $3.06\,\mathrm{mm\,day^{-1}}$. Among the tested candidates, the three-centre ECMWF+NCEP+UKMO combination gives the lowest overall CRPS, $3.01\,\mathrm{mm\,day^{-1}}$, and the highest, or least-negative, BSS at $20\,\mathrm{mm\,day^{-1}}$, -0.013. Although the BSS remains slightly negative, the three-centre combination performs best among the tested forecast-source combinations and gives the most balanced daily predictive distribution. We therefore use the ECMWF+NCEP+UKMO combination as the daily BMA input for the seven-day rainfall and impact analyses. The lead-day-averaged centre weights are 0.34 for ECMWF, 0.47 for NCEP, and 0.19 for UKMO; the full per-lead-day weights and kernel parameters are reported in Appendix A.

Figure 2 shows the same source-selection scores by lead day, together with the raw multi-centre ensemble and the monthly climatology as references.

Individual lead days can favour different centre combinations, but the three-centre combination is the most robust overall. Daily BMA skill is concentrated in the early part of the forecast horizon: the BMA CRPS increases from $2.36\,\mathrm{mm\,day^{-1}}$ at lead day 1 to about $3.5\,\mathrm{mm\,day^{-1}}$ at lead day 14, and becomes similar to or worse than the monthly climatological benchmark around lead days 5–6. This indicates that useful daily rainfall information is mainly retained at shorter leads, whereas longer-lead daily S2S rainfall prediction remains difficult. The comparison is also demanding because the monthly climatology is constructed directly from observations in the training years, whereas the BMA distribution must extract event-specific

https://doi.org/10.5194/egusphere-2026-3482
Preprint. Discussion started: 26 June 2026

Table 2. Averaged CRPS and BSS scores over lead days 1–14 for candidate centre combinations in the daily BMA probabilistic rainfall model. BSS is computed at 20 $\mathrm{mm\,day^{-1}}$ against the monthly climatology, and the selected combination is shown in bold.

Combination	CRPS (all)	$\mathrm{BSS_{20}}$ (all)
ECMWF (51)	3.54	−0.356
NCEP (16)	3.38	−0.166
UKMO (4)	3.19	−0.065
ECMWF+NCEP (67)	3.13	−0.073
ECMWF+UKMO (55)	3.18	−0.105
NCEP+UKMO (20)	3.08	−0.024
All three centres (71)	**3.01**	**−0.013**

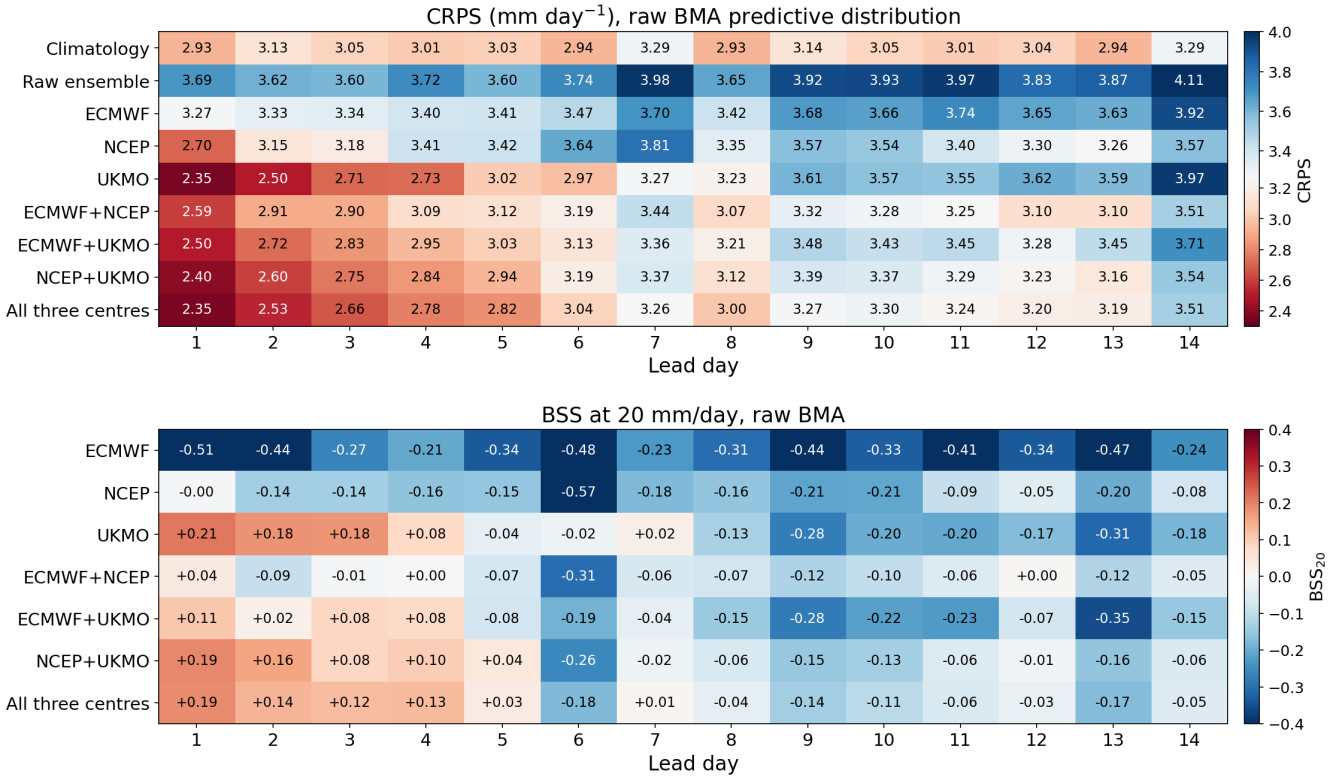

Figure 2. Lead-day heatmaps of forecast-source selection scores for the seven candidate centre combinations. Top: CRPS; bottom: BSS at 20 $\mathrm{mm\,day^{-1}}$ against the monthly climatology. Pooled scores are reported in Table 2.

information from model forecasts. The result therefore supports using the BMA output as a short-lead probabilistic rainfall component, while also highlighting the limited skill available at longer S2S lead times.

https://doi.org/10.5194/egusphere-2026-3482
Preprint. Discussion started: 26 June 2026

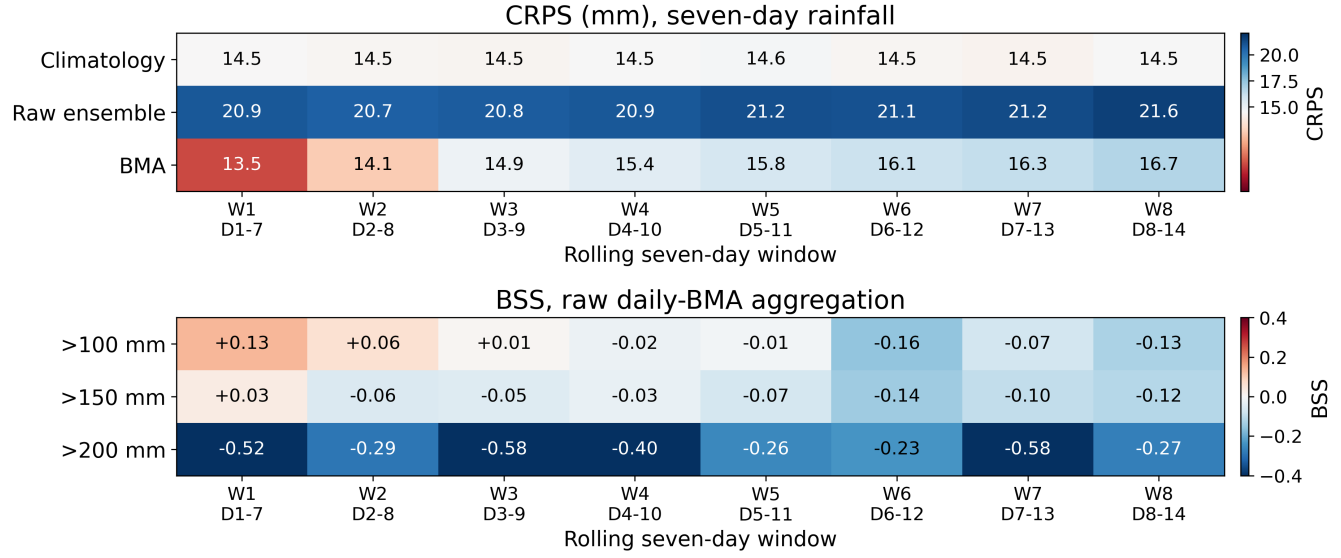

Figure 3. Lead-window heatmaps of seven-day rainfall scores. Top: CRPS by rolling seven-day window for the climatology, the raw ensemble accumulation, and the daily-BMA aggregation. Bottom: raw seven-day BSS of the daily-BMA aggregation by rolling window and threshold (100/150/200 mm per seven days) against the monthly climatology.

3.2 Rolling seven-day rainfall probabilities for preparedness

The daily BMA samples are accumulated into eight rolling seven-day windows (D 1–7, D 2–8, …, D 8–14) to obtain the preparedness-oriented rainfall variable R_7 (RQ2). This converts lead-day-specific daily forecasts into multi-day rainfall distributions while retaining the evolution of forecast risk as the event approaches.

Figure 3 evaluates the rolling seven-day forecasts using CRPS and BSS. The daily-BMA aggregation gives the best performance among the post-processed seven-day variants, with an overall CRPS of 15.4 mm, and is much better than the raw ensemble accumulation (21.0 mm). However, it only matches or outperforms the monthly climatology in the shortest windows and falls behind from around W3 onward. This is consistent with the daily lead-day evaluation in Sect. 3.1: daily BMA skill is mainly retained up to about lead day 5, so rolling windows that depend increasingly on longer leads become less competitive with the observation-based climatological benchmark. The BSS heatmap shows the same limitation for threshold exceedance probabilities before recalibration, motivating the seven-day probability recalibration in Sect. 3.3.

3.3 Recalibration of seven-day exceedance probabilities

The seven-day recalibration (Sect. 2.6) was applied before impact translation, with no identity fallback triggered in any rolling-window group. The fitted curves correct a consistent overconfidence in the raw probabilities ($a < 1$, $b < 1$; e.g. $a \approx 0.85$, $b \approx 0.41$ for W1–W3). Recalibration improved reliability and the Brier Skill Score (computed against the monthly climatological base rate): pooled over all windows it raised BSS from -0.02 to $+0.08$ at $R_7 > 100$ mm and from -0.07 to $+0.02$ at $R_7 >$

https://doi.org/10.5194/egusphere-2026-3482
Preprint. Discussion started: 26 June 2026

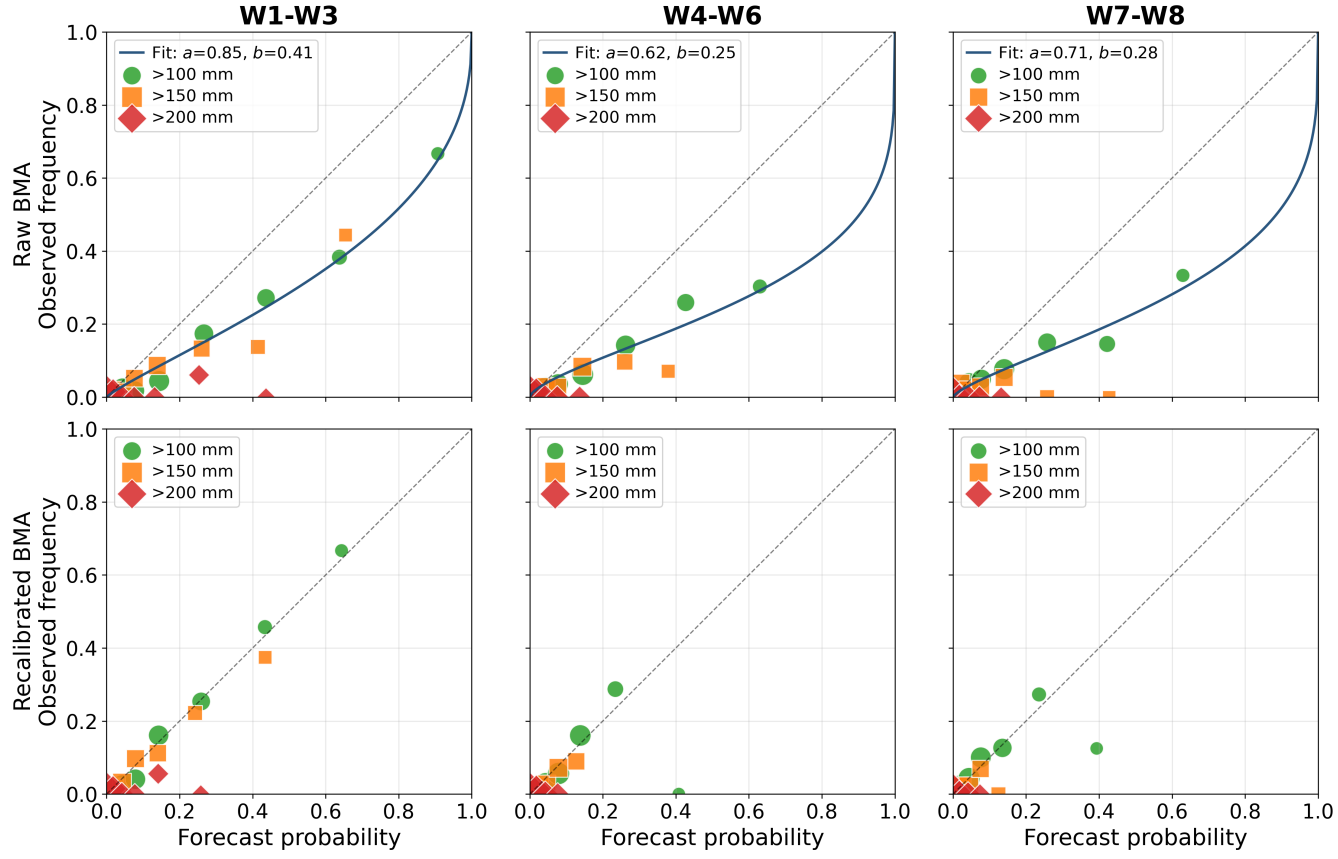

Figure 4. Seven-day exceedance reliability by rolling-window group, raw (top) and recalibrated (bottom). Markers are thresholds (100/150/200 mm/7 d); the curve on the raw row is the fitted recalibration. Closer to the 1:1 line is better calibrated.

150 mm, turning the raw negative scores positive (Fig. 5). At $R_7 > 200$ mm the events are too rare (18 over 2015–2025) for a reliable score, and BSS stays negative (raw -0.36, recalibrated -0.11).

The reliability diagrams (Fig. 4) confirm that the recalibrated probabilities lie closer to the diagonal than the raw ones. Because the mapping is fitted to threshold exceedance reliability, it is best read as a threshold-informed adjustment of the probability scale rather than as full distributional calibration.

3.4 Translation into municipality-level impact early-warning information

We illustrate the impact translation with Typhoon Ulysses (international name Vamco), which struck Luzon on 11–12 November 2020 and caused one of the worst recent floods in the Cagayan River Basin (National Disaster Risk Reduction and Management Council (NDRRMC), 2021). The observed basin-mean seven-day accumulation peaked near 195 mm. Three forecast issues, 2, 5, and 9 November, bracket the 9–15 November flood window and show how the seven-day rainfall forecast evolved as the event approached (Fig. 6). For the strongest seven-day window, the recalibrated median increased from about 79 mm

https://doi.org/10.5194/egusphere-2026-3482
Preprint. Discussion started: 26 June 2026

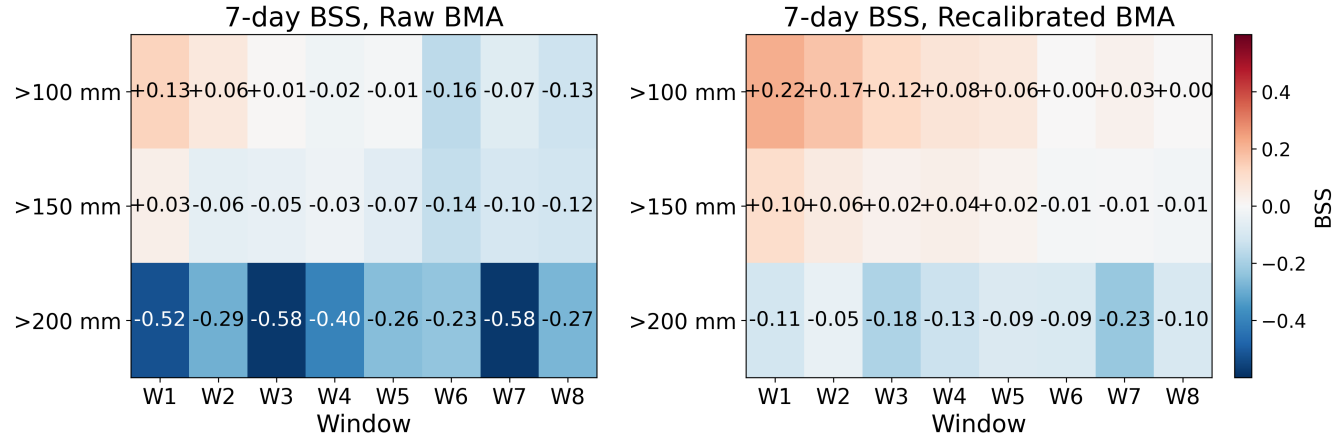

Figure 5. Brier Skill Score of seven-day exceedance probabilities before and after recalibration. BSS is computed relative to the monthly climatological baseline. Positive values indicate skill above climatology.

in the 2 November issue to about 101 mm in the 5 November issue and about 182 mm in the 9 November issue, becoming progressively closer to the observed accumulation.

The 9 November issue is used for the municipality-level impact example because it provides the clearest pre-event signal for the 9–15 November flood window. Propagating its recalibrated seven-day rainfall distribution through the municipality-level rainfall–damage functions gives basin-total expected impacts of about 3.7×10^5 affected people, with a median of 3.5×10^5 and a 95% predictive interval of approximately 2.2–6.0×10^5. The corresponding expected economic impacts are approximately PHP 0.13 billion for buildings, PHP 0.62 billion for rice, and PHP 1.27 billion for maize.

Figure 7 shows the municipality-level impact distributions from the 9 November issue. The dumbbell plot ranks municipalities by forecast affected population and shows the median together with nested 50%, 85%, and 90% predictive intervals. These intervals show how rainfall uncertainty is propagated into impact uncertainty and how the resulting distributions can provide municipality-level guidance for staged preparedness. The black stars indicate RRI-based reference impacts from a calibrated flood simulation of the same event, constrained by observed Magat dam releases and water levels at Buntun Bridge.

Figure 8 shows the affected-population distribution for Tuguegarao City across the three forecast issues. The predictive interval changes with the evolving rainfall forecast distribution rather than remaining fixed across issue dates. In the 2 November issue, the rainfall median remains low and the affected-population distribution stays close to zero. In the 5 November issue, the rainfall median increases near the event window and the impact interval widens. By the 9 November issue, the rainfall median is close to the observed accumulation and the impact distribution becomes more concentrated around the reference impact. This shows that the impact uncertainty range is forecast-state-dependent and can provide dynamic decision-support information as an event approaches.

https://doi.org/10.5194/egusphere-2026-3482
Preprint. Discussion started: 26 June 2026

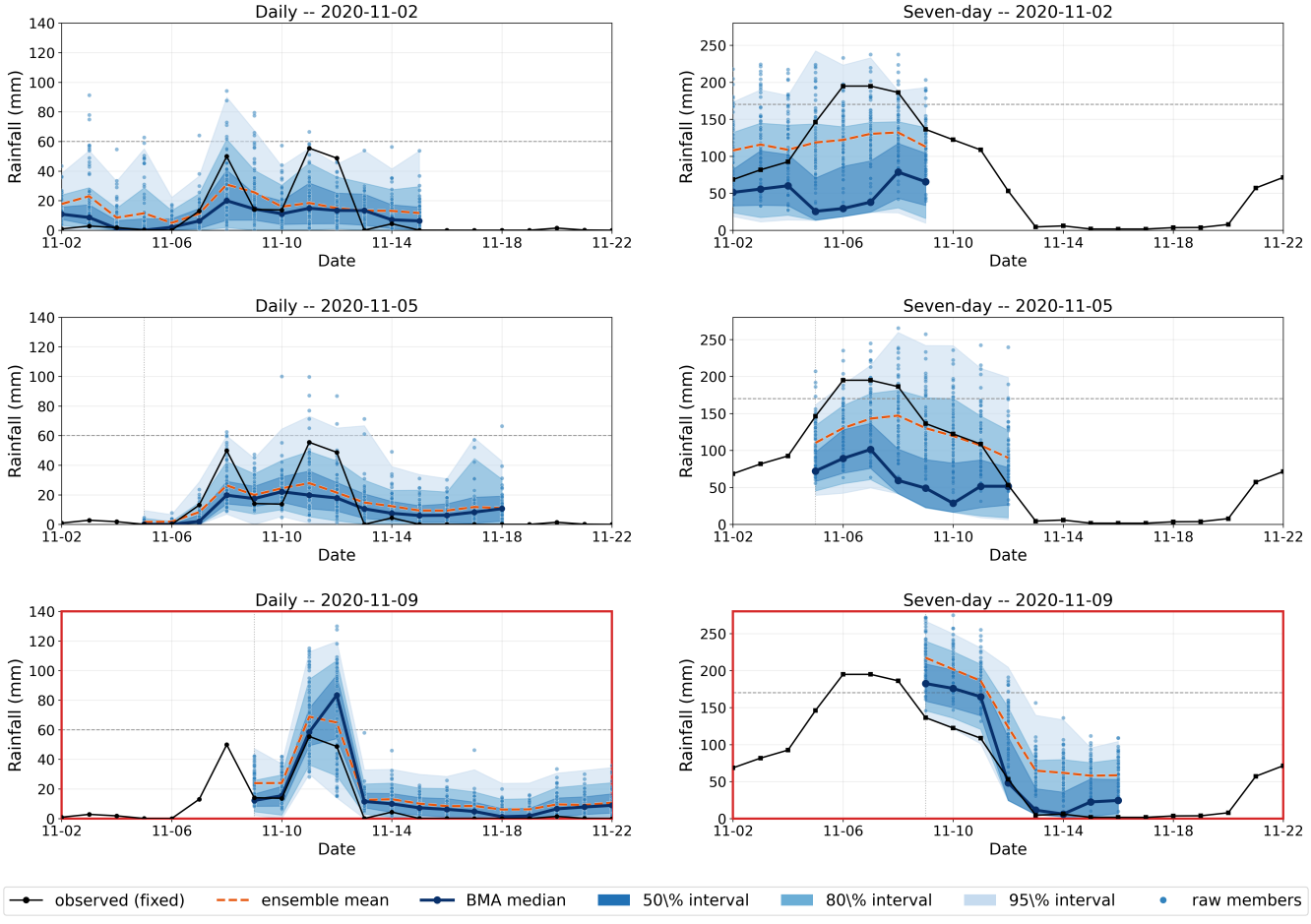

Figure 6. Daily rainfall forecasts (left) and rolling seven-day rainfall forecasts (right) for Typhoon Ulysses, issued on 2, 5, and 9 November 2020. Black lines show observations, blue shading shows the 50/80/95% predictive intervals, and dark-blue lines show the median. Dashed lines indicate the 1-year return-period levels.

4 Discussion

4.1 From calibrated rainfall probabilities to impact distributions

The main result of this study is not that S2S rainfall forecasts can predict a specific flood impact deterministically, but that calibrated rainfall forecast distributions can be transformed into probabilistic impact information at a scale that is more directly relevant to preparedness decisions. The daily BMA model first converts multi-centre S2S ensemble outputs into continuous predictive rainfall distributions. These distributions are then accumulated into rolling seven-day rainfall totals, recalibrated at preparedness-relevant thresholds, and propagated through municipality-level rainfall–damage functions. This sequence

https://doi.org/10.5194/egusphere-2026-3482
Preprint. Discussion started: 26 June 2026

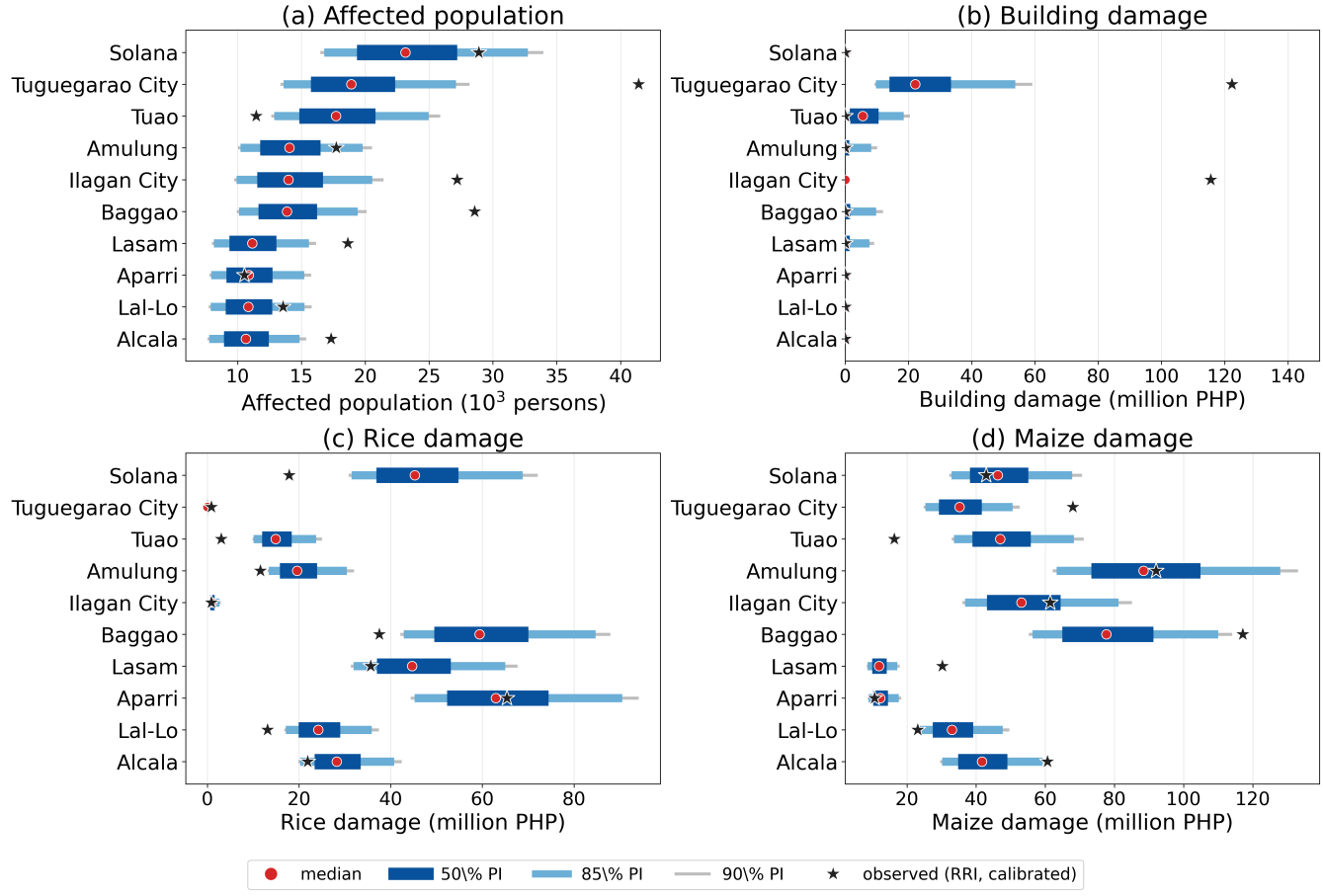

Figure 7. Municipality-level forecast impacts from the 9 November 2020 issue of Typhoon Ulysses for the D 1–7 window. Panels show affected population, building damage, rice damage, and maize damage for the same ten municipalities. Red dots show medians, horizontal bars show nested 50%, 85%, and 90% predictive intervals, and black stars show RRI-based reference impacts.

changes the forecast product from "how much rain may fall" to "what range of impacts may occur, where, and with what uncertainty."

This distinction is important for impact-based early warning. In probabilistic forecasting, predictive distributions should be interpreted in terms of both calibration and sharpness (Gneiting et al., 2007; Gneiting and Raftery, 2007). A single deterministic rainfall or damage estimate can hide the range of plausible outcomes, whereas a predictive interval explicitly shows whether the forecast signal is concentrated or uncertain. BMA is well suited to this purpose because it converts ensemble forecasts into predictive probability distributions while accounting for the relative information content of ensemble members or centres (Raftery et al., 2005; Sloughter et al., 2007).

For preparedness, forecast uncertainty should not be treated only as an error source to be removed. It is also information that can be interpreted. A wide upper interval can indicate that high-impact outcomes remain plausible even when the median impact

https://doi.org/10.5194/egusphere-2026-3482
Preprint. Discussion started: 26 June 2026

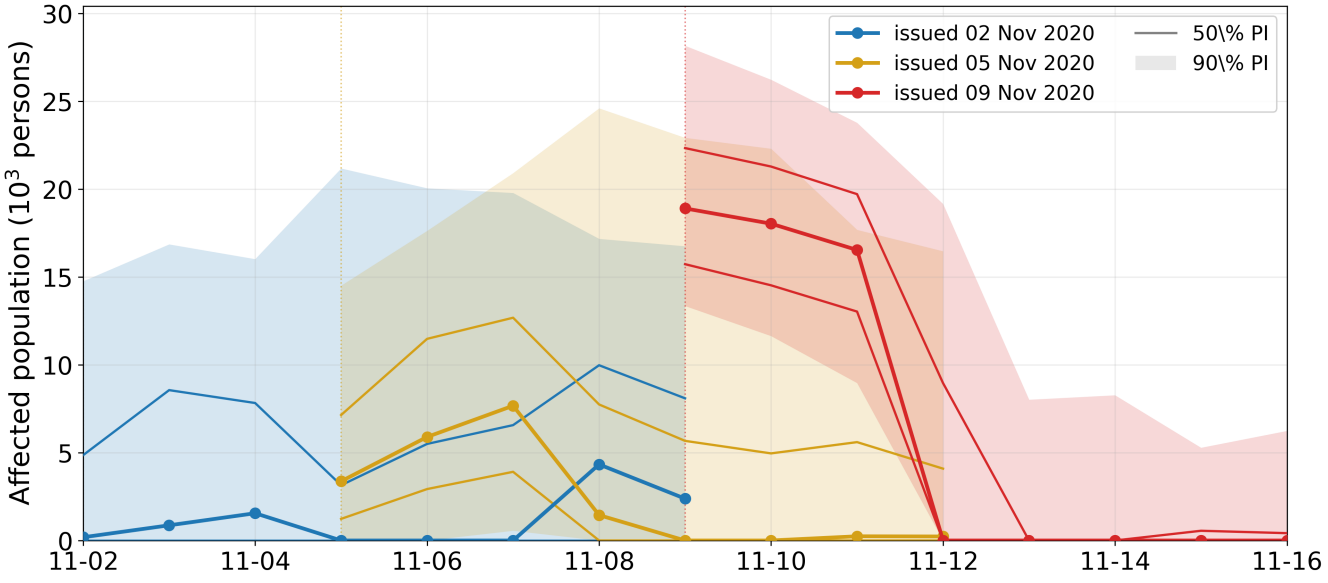

Figure 8. Forecast affected-population distributions for Tuguegarao City from the 2, 5, and 9 November 2020 issues. Lines show medians and predictive-interval bounds obtained by propagating recalibrated seven-day rainfall samples through the affected-population damage function.

is moderate. Conversely, a narrow interval around high rainfall can provide a stronger signal for more specific and resource-intensive preparedness actions. The proposed framework therefore makes the predictive spread operationally interpretable by translating it into uncertainty ranges for affected population, building damage, rice damage, and maize damage.

4.2 Skill horizon and the role of seven-day recalibration

The results clarify the practical skill horizon of the proposed framework. The daily verification shows that the post-processed S2S rainfall signal is mainly retained at shorter lead times, approximately up to lead days 5–6. Beyond this range, the daily BMA forecasts become similar to, or worse than, the monthly climatological benchmark. Therefore, the framework should not be interpreted as providing reliable deterministic impact guidance two weeks in advance. Its more realistic role is to support early monitoring and staged preparedness, with increasing specificity as the event approaches.

The use of rolling seven-day rainfall is important because flood impacts in the Cagayan River Basin are more closely related to multi-day rainfall accumulation than to isolated daily rainfall peaks. However, accumulation alone is not sufficient. The raw accumulated BMA probabilities still show miscalibration, and recalibration is needed before the probabilities are translated into impacts. The improvement in BSS from negative to positive values at 100 and 150 mm per seven days indicates that the recalibrated probabilities contain useful information beyond the monthly climatological baseline at moderate-to-high preparedness thresholds. At the same time, the continued negative BSS at 200 mm per seven days shows that the framework remains weak for the rarest and most severe rainfall outcomes.

https://doi.org/10.5194/egusphere-2026-3482
Preprint. Discussion started: 26 June 2026

This result has an important operational implication. The framework is most defensible for identifying elevated preparedness conditions and municipality-level prioritisation under moderate-to-high rainfall risk, rather than for confidently predicting the most extreme flood scenarios. In practical use, the 100 and 150 mm per seven-day thresholds may support early preparedness triggers, whereas the 200 mm threshold should be treated more cautiously and supplemented by hydrological monitoring, local rainfall observations, dam operation information, and shorter-range forecasts.

4.3 Implications for staged flood preparedness

The impact distributions provide a basis for staged decision-making because different preparedness actions require different levels of forecast confidence. This interpretation is consistent with the broader development of ensemble and probabilistic flood forecasting, in which forecast uncertainty is used to support risk-based interpretation and decision-making (Cloke and Pappenberger, 2009; Dale et al., 2014). Low-regret actions, such as enhanced monitoring, inter-agency information sharing, checking evacuation centres, and preparing agricultural advisories, may be justified when the upper predictive interval begins to increase, even if the median impact remains low. More resource-intensive actions, such as pre-positioning relief items, mobilising field teams, or preparing larger-scale evacuation support, may require stronger evidence, such as an increasing median impact together with a high upper predictive interval or repeated forecast signals across issue dates.

The municipality-level impact distributions are particularly useful for prioritisation. Preparedness resources are often limited, and agencies need to decide where to focus field checks, agricultural advisories, evacuation readiness, and emergency supplies. Ranking municipalities by expected impact, while also showing predictive intervals, gives decision-makers two complementary pieces of information: where impacts are most likely to be large, and where uncertainty remains large enough to justify precautionary attention. This is especially relevant in the Cagayan River Basin, where flood preparedness involves both populated urban areas and extensive agricultural exposure.

The Typhoon Ulysses application illustrates this interpretation. The forecast impact distribution did not remain fixed; it evolved across the 2, 5, and 9 November forecast issues as the rainfall distribution changed. Earlier forecasts can support awareness and coordination, while later forecasts can provide more specific guidance on priority municipalities and impact categories. In this sense, the framework is not intended to replace local warning systems or hydrological forecasts, but to add an uncertainty-aware impact layer that can support anticipatory planning.

4.4 Decision value and interpretation of probabilistic impacts

A key advantage of the proposed approach is that it expresses potential impacts as distributions rather than as single values. This allows preparedness decisions to consider different parts of the impact distribution according to the cost, timing, and reversibility of possible actions. The median impact can be used as an indication of the central forecast scenario, while upper predictive intervals can support precautionary planning when decision-makers wish to avoid missing high-impact outcomes.

This distinction is important because preparedness actions do not all require the same probability threshold. Low-cost and reversible actions may be justified by a relatively small probability of a severe outcome, whereas costly or disruptive actions may require stronger evidence. This interpretation is consistent with impact-based warning and anticipatory-action approaches,

https://doi.org/10.5194/egusphere-2026-3482
Preprint. Discussion started: 26 June 2026

in which hazard information, forecast probabilities, exposure, vulnerability, and possible actions are considered together (World Meteorological Organization, 2015; Coughlan de Perez et al., 2015; Lala et al., 2022; White et al., 2022).

The framework also provides a common quantitative basis for comparing different impact categories. Affected population, building damage, rice damage, and maize damage are expressed as separate impact distributions, allowing users to identify whether the main concern is humanitarian, residential, or agricultural. By presenting both central estimates and uncertainty ranges, the framework can help decision-makers identify where impacts are likely to be large, where uncertainty remains substantial, and where precautionary action may be justified.

4.5 Limitations and pathways for improvement

Several limitations should be considered when interpreting the results. First, the framework uses basin-mean rainfall as the meteorological input. This is appropriate for basin-scale preparedness in a large river basin and is consistent with the coarse spatial resolution of S2S forecasts. However, it cannot fully represent spatial rainfall gradients, tributary-scale floods, or local convective rainfall. As a result, municipality-level impact estimates may be less reliable where flood impacts depend strongly on local rainfall distribution rather than basin-wide accumulation. Future work should therefore test sub-basin rainfall indices or spatially distributed post-processed rainfall fields.

Second, the upper tail of the rainfall distribution remains insufficiently represented. The negative BSS at the 200 mm per seven-day threshold suggests that the ensemble and post-processing framework still have difficulty distinguishing the rarest high-rainfall events from climatology. This limitation matters because disaster management is often most concerned with low-probability, high-impact outcomes. Additional forecast centres, larger hindcast samples, extreme-oriented post-processing, or hybrid approaches that combine S2S forecasts with event-based hydrological monitoring may improve the representation of such upper-tail risks.

Third, the rainfall–damage functions are derived from flood-damage simulations rather than independent observed loss records. They provide an internally consistent way to translate rainfall uncertainty into impact uncertainty, but the resulting impact distributions should be interpreted as preparedness-oriented estimates rather than verified damage forecasts. Further validation using observed inundation maps, reported damages, crop-loss records, and additional historical flood events would strengthen the operational credibility of the impact scale.

Fourth, the current framework propagates rainfall uncertainty but does not fully propagate hydrological, hydraulic, exposure, or vulnerability uncertainty. In reality, flood impacts also depend on antecedent soil moisture, reservoir operations, river water levels, flood duration, flow velocity, building vulnerability, crop growth stage, warning response, and local protection measures. Including these factors would likely widen the impact intervals and make them more realistic for operational use.

The framework is transferable to other large river basins where three conditions are met: a sufficiently long archive of ensemble rainfall forecasts, a reliable observational rainfall reference for post-processing, and impact functions or damage models that link rainfall or flood hazard to exposed assets. The method is particularly relevant for basins where local agencies need preparedness information several days before a flood but where deterministic forecasts remain uncertain. However, trans-

https://doi.org/10.5194/egusphere-2026-3482
Preprint. Discussion started: 26 June 2026

fer to smaller basins, flash-flood-prone catchments, or highly urbanised drainage systems would require finer spatial rainfall representation and shorter-range hydrometeorological forecasting.

Overall, the study shows that calibrated S2S rainfall probabilities can be converted into probabilistic, municipality-level impact information, but only within diagnosed limits of lead-time skill, threshold reliability, upper-tail representation, and impact-function validity. The value of the framework lies in making these limits explicit while still providing actionable uncertainty information for staged flood preparedness.

5 Conclusions

This study developed a framework for translating outputs from multi-centre S2S forecast models into probabilistic, municipality-level impact-based flood early-warning information for staged preparedness in the Cagayan River Basin, Philippines. The framework links daily probabilistic rainfall post-processing, rolling seven-day rainfall accumulation, probability recalibration, and rainfall–damage translation into a single workflow. The main conclusions are as follows.

(1) **RQ1.** Combining multiple S2S forecast models in the daily Gamma-kernel BMA gives more robust post-processed probabilistic rainfall forecasts than using any single S2S forecast model. Among the tested combinations, the ECMWF+NCEP+UKMO ensemble provides the most robust overall performance, with the clearest daily rainfall signal concentrated at shorter leads, mainly through approximately lead days 5–6.

(2) **RQ2.** Accumulating the daily BMA predictive samples into rolling seven-day rainfall distributions, and applying probability recalibration to their exceedance probabilities, improves the usefulness of the forecasts at preparedness-relevant thresholds. Brier Skill Scores improve from negative to positive values at 100 and 150 $\mathrm{mm/7\,d}$, indicating that the recalibrated seven-day distributions can express rainfall uncertainty at the preparedness-relevant time scale beyond the monthly climatological baseline.

(3) **RQ3.** Propagating the recalibrated seven-day rainfall distributions through municipality-level rainfall–damage functions converts rainfall-forecast uncertainty into municipality-level impact uncertainty, supporting impact-based early-warning information for staged flood preparedness. The Typhoon Ulysses case study shows that the resulting impact predictive intervals evolve with the forecast state across issue dates, identifying the plausible range of impacts, the municipalities where impacts are likely to be large, and cases where uncertainty remains large enough to justify precautionary attention for affected population, building damage, rice damage, and maize damage.

These results show that post-processed outputs from S2S forecast models can be transformed into impact-relevant information for staged flood preparedness. Continued improvements in longer-lead forecast skill, spatial rainfall representation, upper-tail rainfall coverage, and validation of rainfall–damage functions would further strengthen the practical value of this framework as a decision-support tool for uncertainty-aware impact-based flood early warning.

https://doi.org/10.5194/egusphere-2026-3482
Preprint. Discussion started: 26 June 2026

Code and data availability. The S2S forecast data used in this study are publicly available through the S2S prediction database (https://apps.ecmwf.int/datasets/data/s2s/). The GSMaP v8_MVK precipitation dataset is distributed by JAXA (https://sharaku.eorc.jaxa.jp/GSMaP/). The processed datasets, the daily BMA implementation, the rolling seven-day recalibration code, and the rainfall–damage translation code that support the findings of this study are available from the corresponding author upon reasonable request.

https://doi.org/10.5194/egusphere-2026-3482
Preprint. Discussion started: 26 June 2026

Table A1. Per-centre Gamma-kernel parameters (kernel variance σ^2 in mm^2; zero-rainfall coefficients d_0, d_1) and BMA centre weights w by lead day for the selected ECMWF+NCEP+UKMO model. The per-centre parameters are weakly identified (see text).

	ECMWF			NCEP			UKMO			Centre weights		
Lead day	σ^2	d_0	d_1	σ^2	d_0	d_1	σ^2	d_0	d_1	w_E	w_N	w_U
1	298.3	+5.53	−0.702	42.0	+3.71	−1.522	5.8	+0.03	−0.568	0.17	0.57	0.26
2	1.8	−2.17	−0.002	77.2	+4.34	−1.180	17.0	+2.59	−1.843	0.18	0.43	0.39
3	104.1	+4.04	−0.916	7.1	+1.76	−1.822	35.2	+3.68	−1.546	0.33	0.33	0.34
4	105.0	+5.26	−1.206	4.7	+1.15	−1.676	29.6	+3.42	−1.409	0.40	0.27	0.33
5	50.7	+5.09	−1.685	0.4	−3.00	+0.162	33.8	+4.13	−1.668	0.50	0.17	0.33
6	52.2	+4.93	−1.616	14.0	+2.91	−1.785	43.3	+3.77	−1.215	0.41	0.37	0.22
7	106.2	+5.37	−1.343	8.8	+2.24	−1.797	24.9	+3.20	−1.300	0.40	0.37	0.23
8	69.8	+5.34	−1.420	23.2	+3.91	−1.837	0.3	−2.09	+0.162	0.44	0.46	0.10
9	100.8	+5.62	−1.432	25.8	+4.14	−2.002	0.5	−3.07	+0.153	0.40	0.50	0.11
10	135.6	+6.08	−1.308	13.6	+3.14	−2.026	32.8	+2.96	−1.208	0.38	0.50	0.12
11	108.3	+5.33	−1.280	25.1	+4.42	−2.033	0.5	−3.82	+0.472	0.31	0.59	0.09
12	174.5	+5.89	−1.052	20.8	+4.01	−2.201	0.4	−3.35	+0.363	0.25	0.68	0.07
13	145.3	+5.90	−1.169	30.6	+4.73	−2.018	0.6	−3.28	−0.265	0.27	0.64	0.09
14	144.7	+5.92	−1.241	17.6	+3.92	−2.256	99.6	+0.01	−0.047	0.31	0.69	0.00

Appendix A: Full BMA parameters of the selected three-centre model

Table A1 lists the full-sample estimated per-centre Gamma-kernel parameters of the final ECMWF+NCEP+UKMO BMA by lead day: the kernel variance σ_c^2 and the logistic zero-rainfall coefficients $d_{0,c}$ and $d_{1,c}$ for each centre, together with the BMA centre weights w_c on the right (only the lead-day-averaged weights are quoted in the main text). These per-centre kernel parameters are only weakly identified—the centres trade off against one another—so they should be read as descriptive rather than as physically interpretable values; the centre weights and the out-of-sample predictive scores are the robust quantities.

Appendix B: Rainfall–damage functions

The same threshold-piecewise quadratic function (Eq. (17)) is used for all four impact indicators—affected population, building damage, rice damage, and maize damage. The quadratic is anchored at $D(R^{\mathrm{thr}}) = 0$, so the curve starts smoothly from the

https://doi.org/10.5194/egusphere-2026-3482
Preprint. Discussion started: 26 June 2026

Table B1. Fitted municipality-level threshold-piecewise rainfall–damage functions (Eq. (17)) by province and category: n_{mun} municipalities fitted, n_{flat} with a flat (rainfall-independent zero) response. The reported R^2 refers to the above-threshold quadratic fit. Full per-municipality coefficients are provided as supplementary data.

Province	Category	n_{mun}	n_{flat}	median R^2	min R^2	max R^2
Cagayan	Affected population	28	3	0.981	0.944	0.996
Cagayan	Building damage	28	9	0.972	0.790	0.998
Cagayan	Rice damage	28	4	0.975	0.899	0.999
Cagayan	Maize damage	28	4	0.990	0.973	0.998
Isabela	Affected population	37	1	0.973	0.519	0.994
Isabela	Building damage	37	10	0.961	0.796	1.000
Isabela	Rice damage	37	4	0.994	0.969	1.000
Isabela	Maize damage	37	4	0.989	0.976	0.998
Nueva Vizcaya	Affected population	15	0	0.966	0.912	0.989
Nueva Vizcaya	Building damage	15	7	0.985	0.776	1.000
Nueva Vizcaya	Rice damage	15	2	0.991	0.932	0.999
Nueva Vizcaya	Maize damage	15	3	0.988	0.959	0.998
Quirino	Affected population	6	0	0.985	0.971	0.994
Quirino	Building damage	6	1	0.964	0.860	0.983
Quirino	Rice damage	6	0	0.987	0.941	0.995
Quirino	Maize damage	6	0	0.994	0.983	0.997

breakpoint, and the no-intercept fit in $u = R - R^{\mathrm{thr}}$ is solved by ordinary least squares on the above-threshold simulation points. The breakpoint R^{thr} is defined for every category as the largest simulated rainfall at which the flood-damage model still produces zero damage. The simulation provides damage at rainfall scenarios of 50–800 mm for 86 municipalities of the Cagayan River Basin; rice and maize are treated as two distinct crops (maize is the crop also called corn). The fits are close above the breakpoint (median R^2 about 0.97–0.99 across categories); these R^2 values describe how well the quadratic reproduces the simulation-derived rainfall–damage relationships in the active range and do not imply that the impact forecasts have been independently validated against observed disaster losses. The scatter plots in Figs. B1 (affected population), B2 (building), B3 (rice), and B4 (maize) are overview diagnostics; a province/category summary is given in Table B1, and the complete per-municipality coefficient table (municipality, category, a, b, c, R^{thr}, R^2, n, and rainfall range) is provided as supplementary data. Municipalities with a rainfall-independent zero response (no damage at any simulated scenario) are listed in n_{flat}.

https://doi.org/10.5194/egusphere-2026-3482
Preprint. Discussion started: 26 June 2026

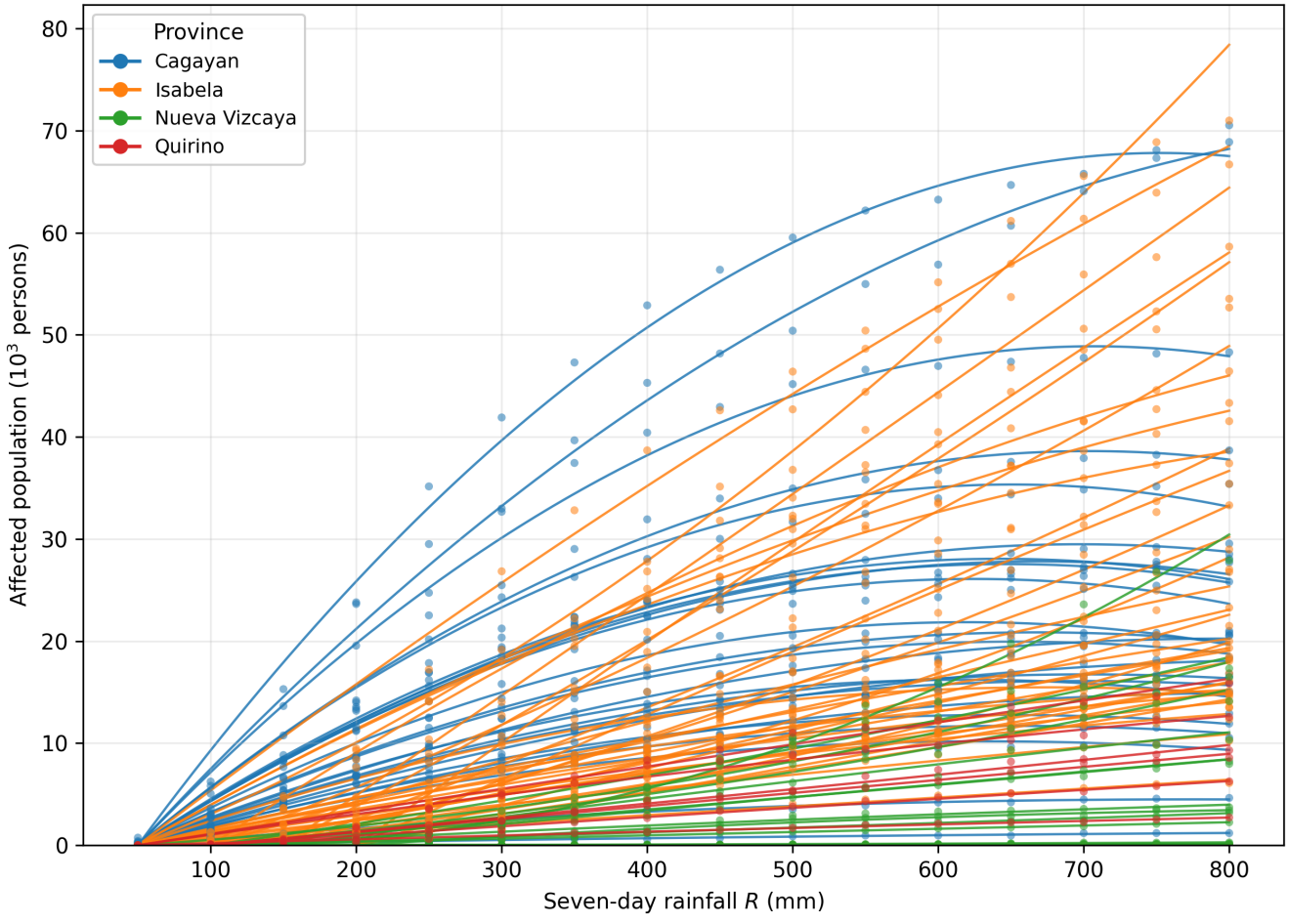

Figure B1. Threshold-piecewise rainfall–damage functions for affected population (Eq. (17)): points are simulation results and curves the fitted functions.

https://doi.org/10.5194/egusphere-2026-3482
Preprint. Discussion started: 26 June 2026

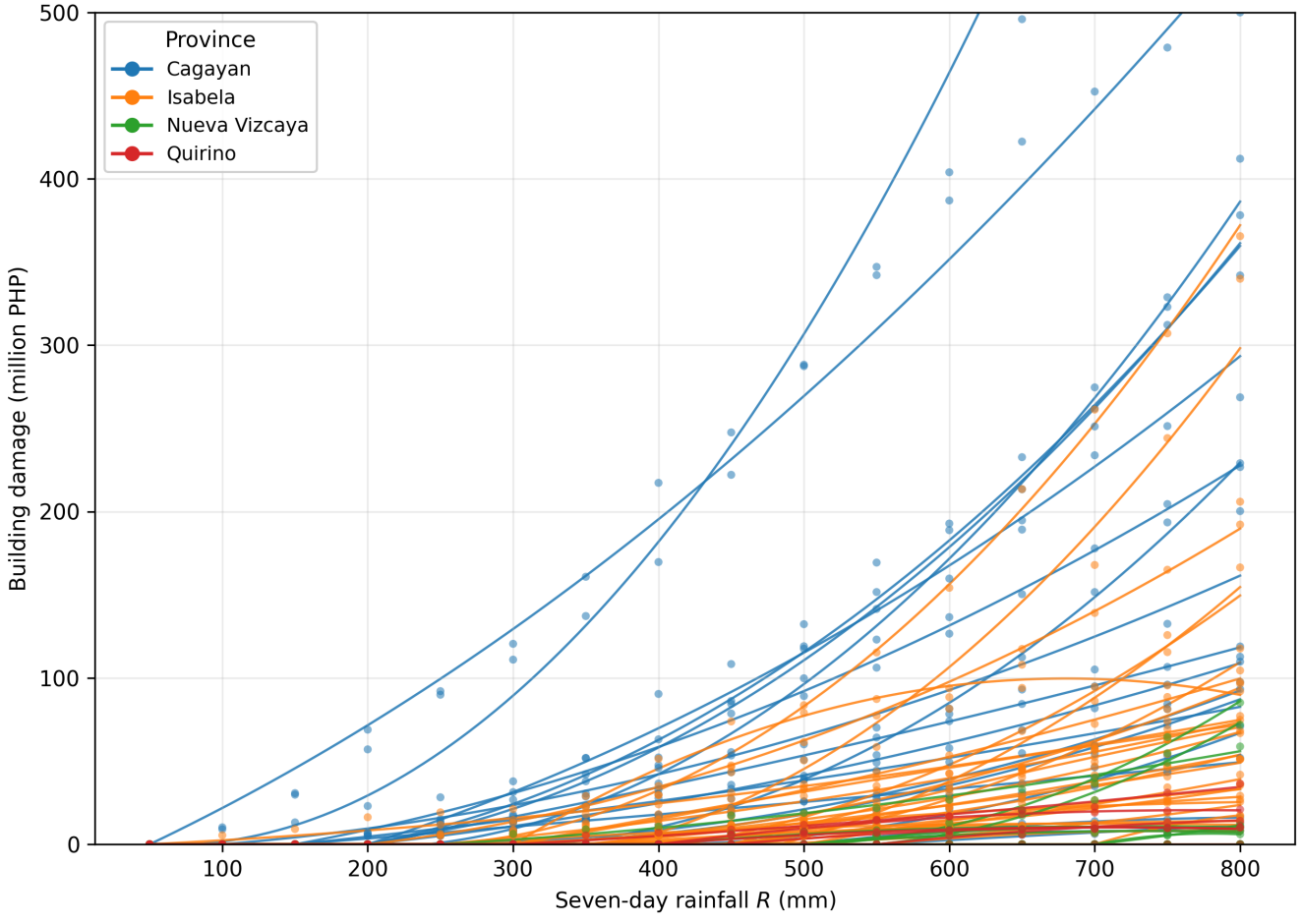

Figure B2. Threshold-piecewise rainfall–damage functions for building damage (Eq. (17)). Points are simulation results and curves the fitted functions; y-axis capped at 500 million PHP.

https://doi.org/10.5194/egusphere-2026-3482
Preprint. Discussion started: 26 June 2026

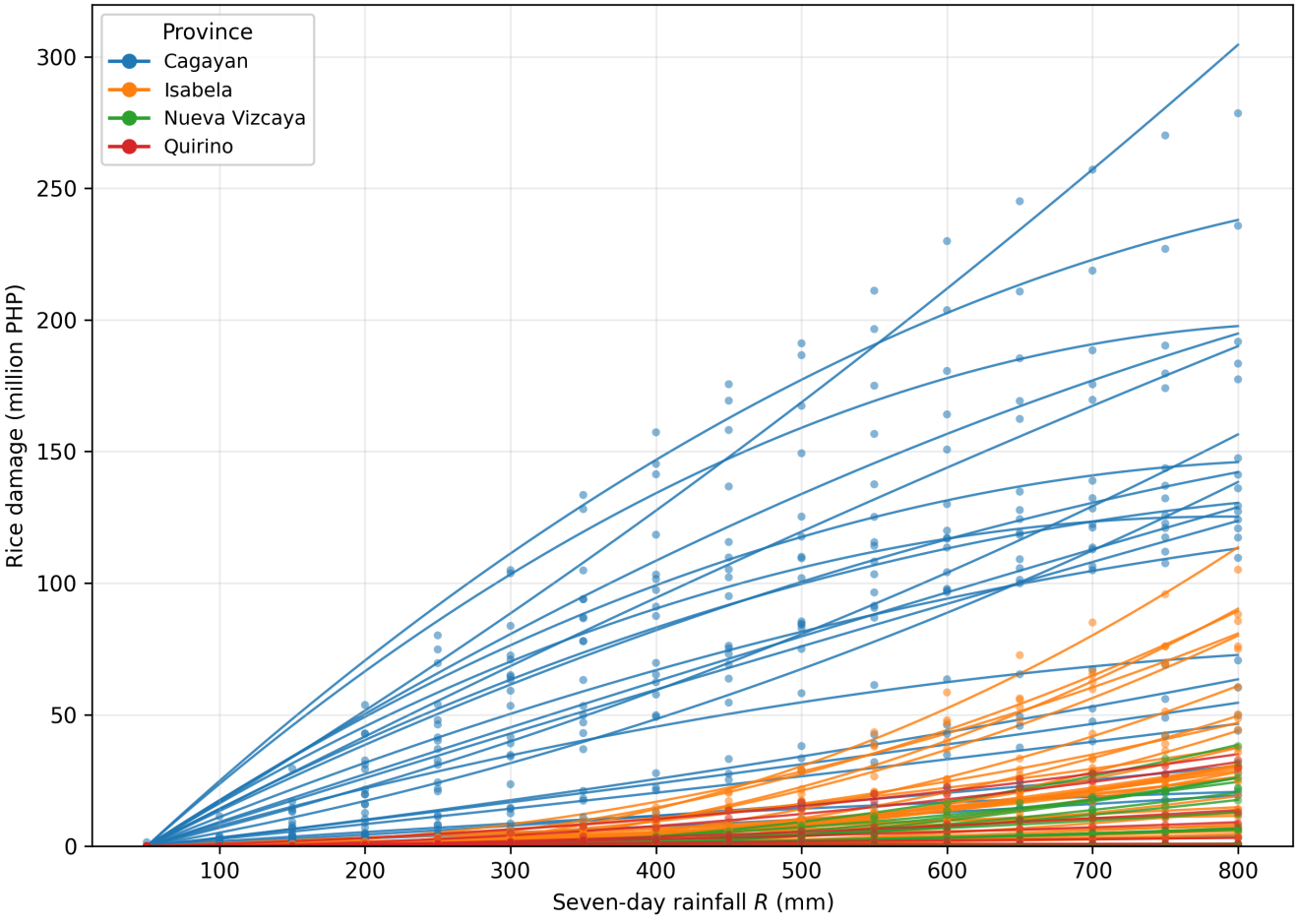

Figure B3. Threshold-piecewise rainfall–damage functions for rice damage (Eq. (17)): points are simulation results and curves the fitted functions.

https://doi.org/10.5194/egusphere-2026-3482
Preprint. Discussion started: 26 June 2026

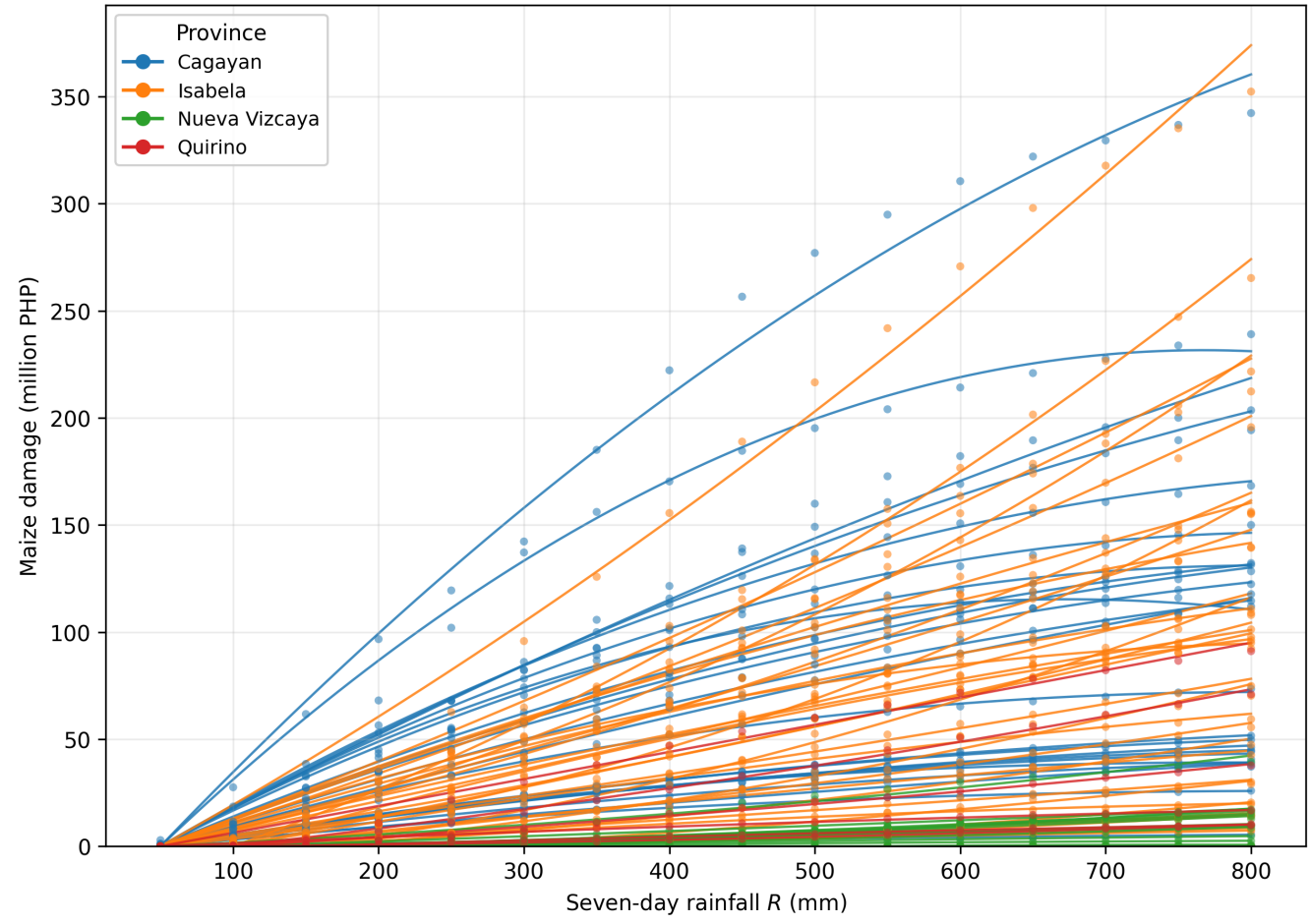

Figure B4. Threshold-piecewise rainfall–damage functions for maize (corn) damage (Eq. (17)): points are simulation results and curves the fitted functions.

https://doi.org/10.5194/egusphere-2026-3482
Preprint. Discussion started: 26 June 2026

Author contributions. YK designed the methodology, performed the analysis, and wrote the manuscript. MM supervised the research.

Competing interests. The authors declare that they have no conflict of interest.

Acknowledgements. The first author is deeply grateful to FASID (Foundation for Advanced Studies on International Development) and Oriental Consultants Global Co., Ltd, for their generous support, which made this research possible.

https://doi.org/10.5194/egusphere-2026-3482
Preprint. Discussion started: 26 June 2026

References

Byrd, R. H., Lu, P., Nocedal, J., and Zhu, C.: A Limited Memory Algorithm for Bound Constrained Optimization, SIAM Journal on Scientific Computing, 16, 1190–1208, https://doi.org/10.1137/0916069, 1995.

Cloke, H. L. and Pappenberger, F.: Ensemble flood forecasting: A review, Journal of Hydrology, 375, 613–626, https://doi.org/10.1016/j.jhydrol.2009.06.005, 2009.

Coughlan de Perez, E., van den Hurk, B., van Aalst, M. K., Jongman, B., Klose, T., and Suarez, P.: Forecast-based financing: an approach for catalyzing humanitarian action based on extreme weather and climate forecasts, Natural Hazards and Earth System Sciences, 15, 895–904, https://doi.org/10.5194/nhess-15-895-2015, 2015.

Dale, M., Wicks, J., Mylne, K., Pappenberger, F., Laeger, S., and Taylor, S.: Probabilistic flood forecasting and decision-making: An innovative risk-based approach, Natural Hazards, 70, 159–172, https://doi.org/10.1007/s11069-012-0483-z, 2014.

Fraley, C., Raftery, A. E., and Gneiting, T.: Calibrating Multimodel Forecast Ensembles with Exchangeable and Missing Members Using Bayesian Model Averaging, Monthly Weather Review, 138, 190–202, https://doi.org/10.1175/2009MWR3046.1, 2010.

Gneiting, T. and Raftery, A. E.: Strictly Proper Scoring Rules, Prediction, and Estimation, Journal of the American Statistical Association, 102, 359–378, https://doi.org/10.1198/016214506000001437, 2007.

Gneiting, T., Balabdaoui, F., and Raftery, A. E.: Probabilistic forecasts, calibration and sharpness, Journal of the Royal Statistical Society: Series B (Statistical Methodology), 69, 243–268, https://doi.org/10.1111/j.1467-9868.2007.00587.x, 2007.

Hersbach, H.: Decomposition of the Continuous Ranked Probability Score for Ensemble Prediction Systems, Weather and Forecasting, 15, 559–570, https://doi.org/10.1175/1520-0434(2000)015<0559:DOTCRP>2.0.CO;2, 2000.

Kurihara, Y. and Miyamoto, M.: Quantifying uncertainty in climate change impacts: A Bayesian approach for the Cagayan River Basin, International Journal of Disaster Risk Reduction, 141, 106 178, https://doi.org/10.1016/j.ijdrr.2026.106178, 2026.

Kurihara, Y., Miyamoto, M., and Kyuzaki, R.: An integrated AI–statistical framework for future flood risk assessment under climate and land use change in the Cagayan Valley region, Philippines, SSRN preprint, https://doi.org/10.2139/ssrn.5391526, posted August 2025, 2025.

Lala, J., Lee, D., Bazo, J., and Block, P.: Evaluating prospects for subseasonal-to-seasonal forecast-based anticipatory action from a global perspective, Weather and Climate Extremes, 38, 100 510, https://doi.org/10.1016/j.wace.2022.100510, 2022.

Li, Y., Wu, Z., He, H., Wang, Q. J., Xu, H., and Lu, G.: Post-processing sub-seasonal precipitation forecasts at various spatiotemporal scales across China during boreal summer monsoon, Journal of Hydrology, 598, 125 742, https://doi.org/10.1016/j.jhydrol.2020.125742, 2021.

National Disaster Risk Reduction and Management Council (NDRRMC): Situational Report No. 29 re Preparedness Measures and Effects of Typhoon "ULYSSES" (I.N. VAMCO), 2021.

Raftery, A. E., Gneiting, T., Balabdaoui, F., and Polakowski, M.: Using Bayesian Model Averaging to Calibrate Forecast Ensembles, Monthly Weather Review, 133, 1155–1174, https://doi.org/10.1175/MWR2906.1, 2005.

Sloughter, J. M., Raftery, A. E., Gneiting, T., and Fraley, C.: Probabilistic Quantitative Precipitation Forecasting Using Bayesian Model Averaging, Monthly Weather Review, 135, 3209–3220, https://doi.org/10.1175/MWR3441.1, 2007.

Vitart, F., Ardilouze, C., Bonet, A., Brookshaw, A., Chen, M., Codorean, C., Déqué, M., Ferranti, L., Fucile, E., Fuentes, M., Hendon, H., Hodgson, J., Kang, H.-S., Kumar, A., Lin, H., Liu, G., Liu, X., Malguzzi, P., Mallas, I., Manoussakis, M., Mastrangelo, D., MacLachlan, C., McLean, P., Minami, A., Mladek, R., Nakazawa, T., Najm, S., Nie, Y., Rixen, M., Robertson, A. W., Ruti, P., Sun, C., Takaya, Y., Tolstykh, M., Venuti, F., Waliser, D., Woolnough, S., Wu, T., Won, D.-J., Xiao, H., Zaripov, R., and Zhang, L.: The Subseasonal to Seasonal (S2S)

https://doi.org/10.5194/egusphere-2026-3482
Preprint. Discussion started: 26 June 2026

Prediction Project Database, Bulletin of the American Meteorological Society, 98, 163–173, https://doi.org/10.1175/BAMS-D-16-0017.1, 2017.

Vitart, F. et al.: The WWRP/WCRP S2S Project and Its Achievements, Bulletin of the American Meteorological Society, 106, https://doi.org/10.1175/BAMS-D-24-0047.1, 2025.

White, C. J., Carlsen, H., Robertson, A. W., Klein, R. J. T., Lazo, J. K., Kumar, A., Vitart, F., Coughlan de Perez, E., Ray, A. J., Murray, V., Bharwani, S., MacLeod, D., James, R., Fleming, L., Morse, A. P., Eggen, B., Graham, R., Kjellström, E., Becker, E., Pegion, K. V., Holbrook, N. J., McEvoy, D., Depledge, M., Perkins-Kirkpatrick, S., Brown, T. J., Street, R., Jones, L., Remenyi, T. A., Yerokhin, O., and Coughlan de Perez, E.: Potential applications of subseasonal-to-seasonal (S2S) predictions, Meteorological Applications, 24, 315–325, https://doi.org/10.1002/met.1654, 2017.

White, C. J., Domeisen, D. I. V., Acharya, N., Adefisan, E. A., Anderson, M. L., et al.: Advances in the application and utility of subseasonal-to-seasonal predictions, Bulletin of the American Meteorological Society, 103, E1448–E1472, https://doi.org/10.1175/BAMS-D-20-0224.1, 2022.

World Meteorological Organization: WMO Guidelines on Multi-hazard Impact-based Forecast and Warning Services, Tech. Rep. WMO-No. 1150, World Meteorological Organization, Geneva, Switzerland, 2015.

Zhan, S., Ye, A., et al.: Ensemble post-processing of sub-seasonal to seasonal precipitation forecasts based on a novel probabilistic double machine learning method, Journal of Hydrology, 660, 133 484, 2025.