Plain Language Summary: This study is the first to use large-ensemble weather forecasts with JMA's flood forecasting system to probabilistically predict flood risk levels. It successfully predicted extreme flooding with 60% probability, 11 hours in advance. We examined the number of ensemble members needed for reliable forecasts and found that both 500-member samples and the full 1000-member ensemble yielded similar accuracy across 36 small and medium-sized rivers.
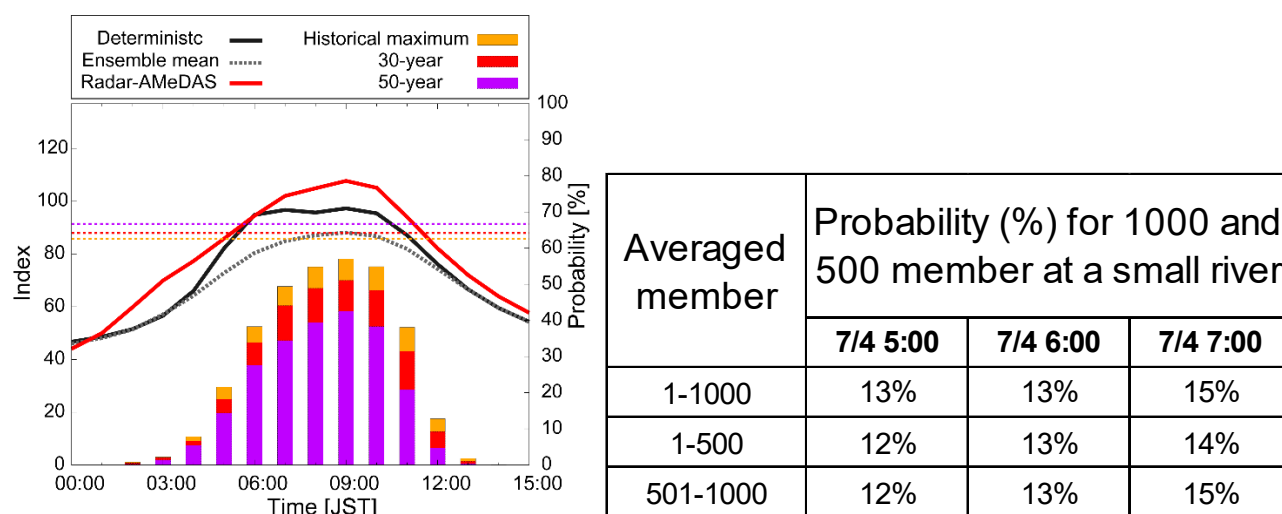
Averaged member	Probability (%) for 1000 and 500 member at a small river		
	7/4 5:00	**7/4 6:00**	**7/4 7:00**
1-1000	13%	13%	15%
1-500	12%	13%	14%
501-1000	12%	13%	15%

Fig. 1. Probability of flood risk based on 1,000 ensemble members. The red line shows RA; the black, the deterministic forecast. The orange dashed line marks the threshold of the historical maximum. Orange bars indicate the probability (right axis) of exceeding the historical maximum. A record-breaking flood is predicted between 05:00 and 11:00, with a probability ranging from 20% to 60%.

Table 1. Probability of exceeding the historical maximum at the Ogawa River, calculated using all 1,000 ensemble members and a subset of 500 members sampled from 100 members.

- The 1000-member ensemble predicted both the risk and timing of record-breaking flooding, while the 100-member ensemble predicted the risk but was less accurate in timing. The 21-member ensemble only detected the occurrence of flooding.
- When we compared the probabilities from the full 1000-member ensemble with those from sampled sets of 100, 250, and 500 members across 36 rivers, the range of deviation was ±10% for each 100-member sample, ±5% for each 250-member sample, and ±2% for each 500-member sample.
- These results suggest that it is not always necessary to compute all 1000 members; selecting 500 members from the initial ensemble yields results nearly identical to those obtained with the full 1000-member set.