



Figure S1. (Left) Experimental cultures of *Limnospira platensis* in 250 mL (Right) Erlenmeyer flasks with different media and aeration conditions, during the inoculation (A) and growth phases (B).

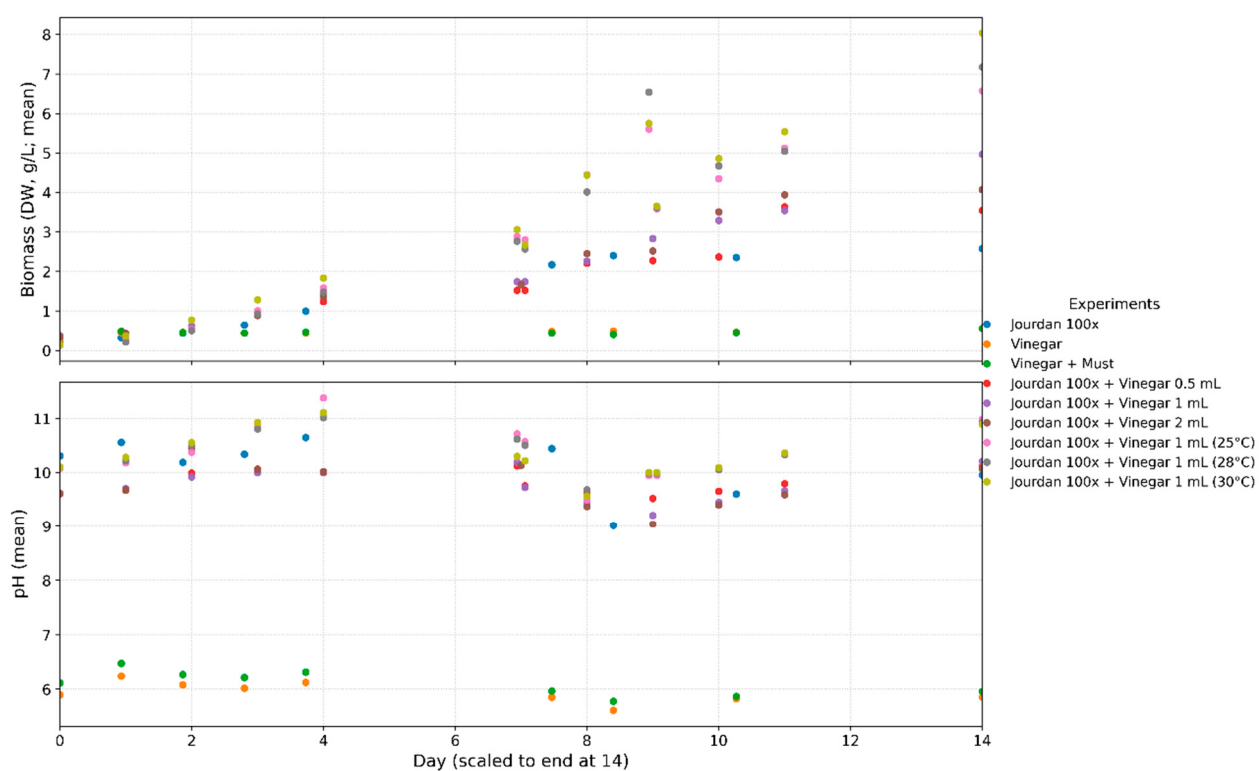


Figure S2. Biomass (DW, g L⁻¹) and pH time courses for Jourdan 100× cultures supplemented with vinegar (0.5–2 mL) and for temperature-controlled trials (25–30 °C), including vinegar-only and vinegar+must controls.

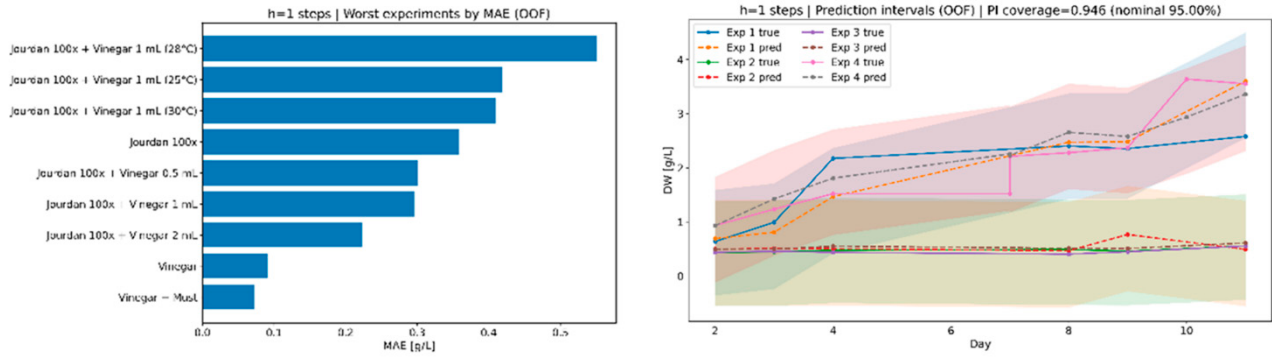


Figure S3. (Left) worst-performing experiments ranked by experiment-level MAE, and (right) representative out-of-fold prediction intervals (nominal 95%) for selected experiments (shaded bands), with empirical OOF coverage reported in the panel title.

Experiment-level MAE ranking (Figure S3 left) further indicates that a limited subset of experiments contributes disproportionately to the overall error, suggesting that specific cultivation regimes or operational perturbations dominate short-term prediction uncertainty.

Uncertainty quantification via residual-based prediction intervals (Figure S3 right) demonstrates well-calibrated uncertainty at this horizon. The empirical out-of-fold coverage is 0.946, closely matching the nominal 95% target, with a median prediction interval width of approximately 1.94 g L^{-1} . These results indicate that both point predictions and associated uncertainty estimates are reliable for short-term forecasting under grouped validation.

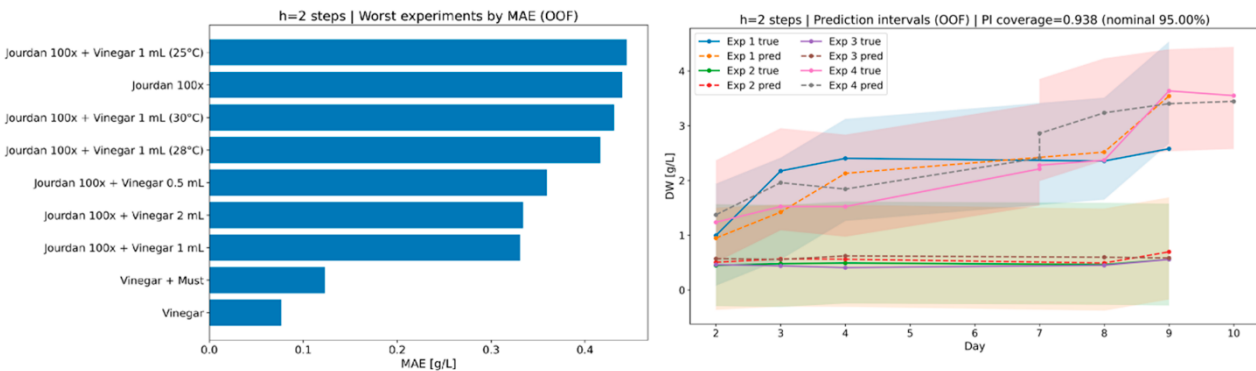


Figure S4. (Left) worst-performing experiments ranked by experiment-level MAE, and (right) representative out-of-fold prediction intervals (nominal 95%) for selected experiments (shaded bands), with empirical OOF coverage reported in the panel title.

Experiment-level rankings (Figure S4 left) identify many of the same experiments among the worst performers, suggesting that sources of difficulty are persistent across horizons rather than horizon-specific artifacts. Such consistency points to structured drivers of prediction error, potentially linked to temperature-dependent kinetics, delayed responses to substrate addition, or phase transitions within the cultivation cycle. Prediction intervals at this horizon (Figure S4 right) remain well calibrated, with empirical coverage of 0.938 and a median width of approximately 1.86 g L^{-1} . Despite the longer forecast step, uncertainty remains controlled, indicating that the residual-based bootstrap approach continues to provide meaningful uncertainty estimates under grouped validation.

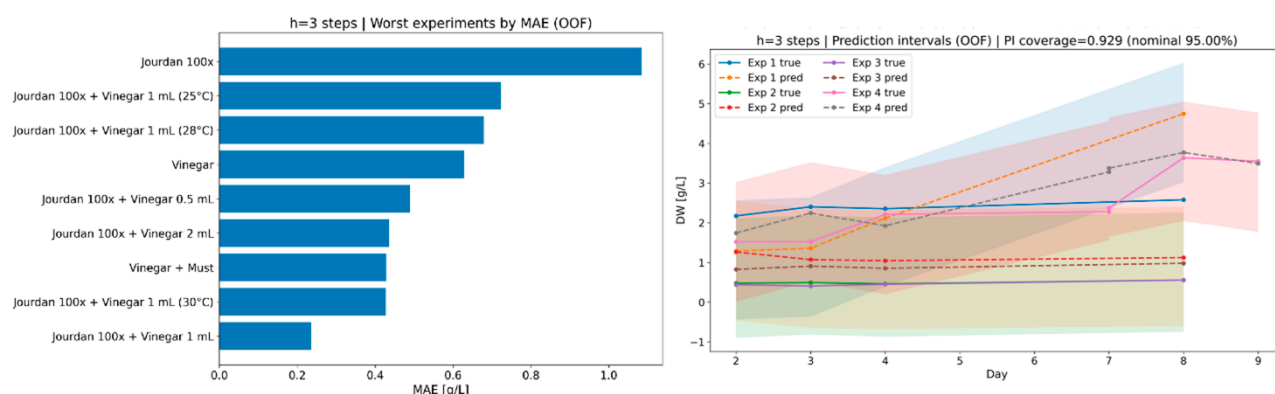


Figure S5. (Left) worst-performing experiments ranked by experiment-level MAE, and (right) representative out-of-fold prediction intervals (nominal 95%) for selected experiments (shaded bands), with empirical OOF coverage reported in the panel title.

Experiment-level MAE rankings (Figure S5 left) emphasize that a recurring subset of experiments dominates long-horizon error. The persistence of these “hard” experiments across horizons reinforces the conclusion that predictability is structured by experimental regime and operational conditions rather than by random noise. Prediction intervals (Figure S5 right) appropriately widen at this horizon, with empirical coverage of 0.929 and a median width increasing to approximately 3.00 g L⁻¹, capturing the expected growth in uncertainty associated with longer forecasting steps.