Lewis, A. S. L., W. M. Woelmer, H. L. Wander, D. W. Howard, J. W. Smith, R. P. McClure, M. E. Lofton, N. W. Hammond, R. S. Corrigan, R. Q. Thomas, and C. C. Carey. 2021. Increased adoption of best practices in ecological forecasting enables comparisons of forecastability. *Ecological Applications.*

**Appendix S2: List of standardized criteria used to assess each paper**

1. Paper title
2. Digital object identifier (doi)
3. Author list
4. Year of publication
5. Journal or conference in which the paper was published
6. Forecast spatial scale, classified into five categories
7. Geographic coordinates of the forecast site using decimal degrees. Locations for regional and national forecasts are approximately the center of the forecast area
8. Forecast ecosystem: forest, grassland, freshwater, marine, desert, tundra, atmosphere, agricultural, urban, global, other
9. Forecast class: biogeochemical or organismal (population or community)
10. Identity of forecast variables
11. Model dimension: 0D, 1D, 2D, 3D
12. Model type: empirical (dependent on correlative or statistical relationships) or process-based (explicitly simulating ecological processes). For forecasting workflows that involve a pipeline of multiple models, this refers to the “final” model that forecasts the forecast variable of interest
13. If specified: more detailed description of model: for example, Bayesian hierarchical, machine learning, named model (e.g., PROTECH), etc.
14. Are meteorological covariates used in this forecast? 1 = yes, 0 = no
15. Are physical covariates (e.g., streamflow) used in this forecast? 1 = yes, 0 = no
16. Are biological covariates used in this forecast? 1 = yes, 0 = no
17. Are chemical covariates used in this forecast? 1 = yes, 0 = no
18. Does the paper include an ensemble forecast (ensemble within model)? 1 = yes, 0 = no
19. Number of ensemble members
20. Does the paper use an ensemble of models to produce one output? 1 = yes, 0 = no
21. How many models in the ensemble model
22. Are multiple models with different model structures compared (NOT including null models)? 1 = yes, 0 = no
23. How many models with different structures are compared?
24. Was a forecast null model (persistence or climatology) included? 1 = yes, 0 = no
25. How many null models?
26. What type of null model (climatology or persistence)?
27. Maximum time into the future that the forecast predicts in this paper, described in days
28. Time step of forecast output. For example, a forecast that gives predictions for the next 16 days but was only run once a week would have a time step of one day (not one week)
29. Are the forecasts described in the papers iterative (i.e., data updating forecasts iteratively)? Any form of iteration counts here: updating initial conditions with new data, refitting the model to incorporate new dta, updating parameter values, etc. State updating via the autoregressive term counts as data assimilation for autoregressive models
30. What technique of data assimilation was used? For example, KF, enKF, refit, update IC, etc.
31. Extent to which uncertainty is included in the forecast, classified within 5 categories:
    1. no (this model does not contain uncertainty)
    2. contains (the model contains uncertainty, but uncertainty is not derived from data; e.g. uncertainty comes from spin-up initial conditions or hand-tuned parameters)
    3. data\_driven (the model contains data-driven uncertainty; e.g. uncertainty in meteorological drivers)
    4. propagates (the model propagates some source of uncertainty)
    5. assimilates (the model iteratively updates uncertainty through data assimilation)
    6. NOTE: this is assumed to be a hierarchy (e.g. if the forecast contains data driven uncertainty and propagates that uncertainty, it would be marked "propagates")
32. What sources of uncertainty were incorporated?
33. Was observation uncertainty included? 1 = yes, 0 = no
34. Are at least two different sources of uncertainty quantified and compared? 1 = yes, 0 = no. NOTE: the two sources may be in the same category of uncertainty—e.g. two forms of driver data)
35. Initial condition uncertainty partitioned? 1 = yes, 0 = no
36. Driver uncertainty partitioned? 1 = yes, 0 = no
37. Parameter uncertainty partitioned? 1 = yes, 0 = no
38. Process uncertainty partitioned? 1 = yes, 0 = no
39. Other partitioned sources of uncertainty? 1 = yes, 0 = no
40. If at least two categories of uncertainty were partitioned, what was the dominant source of uncertainty?
41. If the dominant source varies by forecast horizon, season, etc. please describe here
42. Paper states that forecast was evaluated? 1 = yes, 0 = no
43. Forecast evaluation results reported in paper? 1 = yes, 0 = no
44. List all skill metrics used (e.g. R2, RMSE, bias, MAE). SD and Bayesian credible intervals are not skill metrics
45. Is forecast performance assessed at multiple forecast horizons (results must be reported in paper/supplemental info)? 1 = yes, 0 = no
46. Maximum forecast horizon such that the forecast was better than the null model (out of any models used)
47. Temporal coverage of data used to create this forecasting paper
48. Was new data (driver and/or observations) available to the model in real time (<24 hours from collection) without any manual effort when the system was working as intended? 1 = yes, 0 = no
49. Forecast archiving described in text? 1 = yes, 0 = no
50. Repository in which forecasts are archived
51. Archiving website is still accessible via the link in the paper as of 14 Jun 2021? 1 = yes, 0 = no
52. Text specifies that driver data are publicly available to reproduce the forecasts? 1 = yes, 0 = no
53. Specific end user identified (proper noun)? 1 = yes, 0 = no
54. Partnership with the end user in forecast development mentioned in paper? 1 = yes, 0 = no
55. Forecast being used by the end user according to paper? 1 = yes, 0 = no
56. Forecast delivery method identified? 1 = yes, 0 = no
57. Forecast delivery method?
58. Any ethical considerations mentioned? 1 = yes, 0 = no