

Keeping a “None of the Above” Hypothesis Matters More Than Planning With It: A Controlled Test for Transient Follow-Up

JOYCE MALIK¹

¹[Department], [Institution], [City], India

ABSTRACT

Information-driven follow-up planners for astronomical transients usually assume the event belongs to one of a fixed set of classes. We test, in a controlled toy problem, what happens when it does not. Across 3000 simulated fading transients (3 seeds), a closed-world expected-information-gain planner was confidently wrong on 99.9% of events from a withheld class. Adding an explicit unmodeled hypothesis h_0 , represented by a Gaussian process, to the posterior removes most of this failure. Planning with h_0 helped only modestly: the open-world planner flagged 72.3% of withheld events, against 66.6% for the closed-world planner’s measurements rescored with h_0 . Randomly timed measurements rescored the same way flagged 89.0%. Information-gain planning, even with h_0 , chose measurements that exposed model inadequacy less well than random timing, so inadequacy detection has to be designed into the planning objective explicitly.

1. MOTIVATION

Wide-field surveys (Bellm et al. 2019; Ivezić et al. 2019) discover more transients than follow-up resources can observe, so which measurement to take next is a planning problem. Expected information gain (EIG) about the class label (Lindley 1956; Chaloner & Verdinelli 1995) is a natural objective, but it presumes the true class is among those modeled. Rare and new phenomena are precisely the ones that are not. We isolate this assumption in a deliberately small problem.

2. SETUP

Each event is drawn uniformly from three classes: two *modeled* classes with flux $Ae^{-t/\tau}$ ($\tau = 1$ and 4 d), and one *withheld* class, a Gaussian rise peaking at 1.5 d followed by exponential decay, which no planner’s model contains. Amplitudes are log-uniform on $[5, 20]$ with Gaussian noise $\sigma = 1$. Each strategy takes 5 measurements chosen from 20 log-spaced epochs between 0.2 and 10 d. Noise is pre-drawn per epoch, so all strategies see identical events and noise.

The closed-world belief is an exact grid posterior over class and amplitude (25 log-spaced values). The open-world posterior adds h_0 , scored by the marginal likelihood of a zero-mean Gaussian process (squared-exponential kernel, length scale 1.2 d, variance 100; fixed, not fitted), with a uniform prior over {modeled classes, h_0 }. EIG is computed by exact quadrature. Three strategies choose epochs: random; closed-world EIG; and open-world EIG, whose predictive mixture includes the h_0 Gaussian process. Because every run records both posteriors at every step, we also score the random and closed-world EIG measurement sequences with the open-world posterior after the fact. This separates the effect of *having* h_0 from that of *planning* with it. A decision is “confident” when its posterior exceeds 0.9. Code and per-event outputs are public (see below).

3. RESULTS

Table 1 gives pooled results with 95% Wilson intervals over 1974 modeled-class and 1026 withheld-class events. Modeled-class accuracy was between 97.4% and 99.5% for all strategies. On withheld events, every closed-world posterior must name a modeled class, and closed-world EIG did so with high confidence on 99.9% of them. With the open-world posterior, confident errors fell to 28.4% for the same closed-world EIG measurements, 5.9% for random measurements and 22.9% for the open-world planner. On paired withheld events, the open-world planner flagged h_0 where post-hoc closed-world EIG did not in 282 cases, and the reverse occurred in 223 (exact McNemar $p = 0.0098$). The ordering of h_0 -flag rates (random above open-world EIG above closed-world EIG) held in each of the three

seeds. The area under the ROC curve of $P(h_0)$ separating withheld from modeled events was 0.971 (random), 0.899 (closed-world EIG) and 0.936 (open-world EIG).

Table 1. Pooled outcomes (percent, 95% Wilson interval). “Post hoc”: measurements chosen without h_0 , scored with the open-world posterior.

Strategy	Acc. (modeled)	Conf. wrong (withheld)	Flagged h_0 (withheld)	AUROC $P(h_0)$
Random times, closed-world posterior	97.6 [96.8, 98.2]	91.2 [89.3, 92.8]	0.0 [0.0, 0.4]	–
Random times, open-world posterior (post hoc)	97.4 [96.6, 98.0]	5.9 [4.7, 7.6]	89.0 [86.9, 90.8]	0.971
Closed-world EIG, closed-world posterior	99.5 [99.1, 99.7]	99.9 [99.4, 100.0]	0.0 [0.0, 0.4]	–
Closed-world EIG, open-world posterior (post hoc)	99.2 [98.8, 99.5]	28.4 [25.7, 31.2]	66.6 [63.6, 69.4]	0.899
Open-world EIG, open-world posterior	99.3 [98.8, 99.6]	22.9 [20.4, 25.6]	72.3 [69.5, 75.0]	0.936

4. INTERPRETATION AND LIMITATIONS

In this setting, most of the benefit of an open-world treatment comes from the *posterior*. Planning with h_0 in the predictive mixture adds a small, significant improvement over closed-world planning, but both information-gain planners detected inadequacy less often, and separated it less well, than random timing. We have not isolated why. A plausible reading is that EIG favours epochs that discriminate among hypotheses, which is not the same as epochs that expose misfit. The practical implication is that follow-up planners should carry an explicit adequacy term in their utility, such as the mutual information between a measurement and a misspecification indicator, rather than relying on class-level information gain.

This is a toy problem: two exponential classes, one withheld shape, fixed noise, fixed Gaussian process hyperparameters, a uniform class prior, no observability constraints and a single threshold. The results establish a failure mode and a design requirement, not performance in realistic follow-up. A study with published light-curve templates, real orbital visibility and a preregistered evaluation is in progress.

CODE AVAILABILITY

Simulator, planners, raw outputs and the script that generates every number here: <https://github.com/joycemalik/orbitalnet-dashboard> (branch `research/open-world-planner`).

REFERENCES

- | | |
|---|--|
| Bellm, E. C., et al. 2019, PASP, 131, 018002 | Ivezić, Ž., et al. 2019, ApJ, 873, 111 |
| Chaloner, K., & Verdinelli, I. 1995, StaSc, 10, 273 | Lindley, D. V. 1956, AnMS, 27, 986 |