Forecasting a flying-fox fly-out from nine years of weather radar

The Yarra Bend grey-headed flying-fox camp, Melbourne

Methods, version 1, November 2026. BoEG (boeg.com.au), non-commercial.

> **Status: initial build.** This is the first version of the fly-out forecast, and it will change. Each part has already been tested on data it was not fitted to: the radar filter on 95 dusks labelled blind before it was run (97.1% of fly-outs found, section 3.2), its sizes against 27 counts at the camp, and every forecast on held-out years against climatology (section 7). What is not yet known, or not yet good enough, is in sections 10 and 11.

Summary

Each evening thousands of grey-headed flying-foxes (*Pteropus poliocephalus*) leave the Yarra Bend camp, 4 km north-east of Melbourne's centre, and cross the inner northern suburbs to feed. Counts at the camp ranged from 1,425 to 19,900 between November 2022 and May 2026; the largest count on record is 54,210, in February 2022. The Bureau of Meteorology's weather radars record the fly-out as echo, as Meade et al. (2019) showed at this camp.

We built an archive of every dusk from both Melbourne radars, October 2017 to November 2026: 3,287 dusks from radar 02 and 3,129 from radar 01. From it we trained and back-tested a daily forecast of the fly-out's timing, size and direction, and of whether, when and how strongly the stream passes any point within 9 km. The forecast runs each morning from ECMWF IFS weather forecasts and, where available, the radars' measured state two dusks earlier. It is published at boeg.com.au/bats.

On held-out years, the forecast times the peak at the camp with a mean absolute error of 4.0 min, against 5.1 min for the month's climatology. It forecasts whether the stream passes a given point with an AUC of 0.91, against 0.81 for that point's own climatology. It forecasts the fly-out's size, as an index against the month's usual, with a log10 error of 0.22 against 0.33. Its 80% ranges hold their coverage on held-out years, including on rainy, cloudy and hot dusks. To our knowledge, no daily forecast of a bat emergence, at any spatial scale, has been published before.

Every number below is marked **measured** (from the archive, out of sample where it is a skill), **sourced** (from the literature or an agency) or **assumed** (a stated modelling choice).

1. Background and precedent

Weather radar and bats. Radar ornithology and aeroecology are established fields (Chilson et al. 2012), and the polarimetric radar products used for animals aloft are well described (Stepanian et al. 2016). Bat emergences have been measured from weather radar for Brazilian free-tailed bats in Texas (Horn and Kunz 2008; Frick et al. 2012, who related emergence timing to weather).

For pteropodids, the method literature is one paper. Meade et al. (2019) used ten years (2008 to 2017) of the Melbourne radar (the same radar 02), with horizontal reflectivity and radial velocity. They showed that emergence timing follows sunset with a seasonal amplitude, and that numbers and directions can be read from the echo. Meade et al. (2025) used Bureau radar, after the 2019 method, to time an Adelaide fly-out.

Forecasting animal movement from radar. Machine-learning forecasts of nocturnal bird migration, trained on weather radar and driven by weather forecasts, are operational in the United States (Van Doren and Horton 2018). We found no forecast of a bat emergence, and none of a roost departure's path at street scale.

Flying-fox behaviour used here. - Emergence timing responds to light and cloud (Welbergen 2006). - GPS tracking of the species gives ground speeds of 8.5 to 9.5 m/s and flight heights of about 21 to 36 m (Meade et al. 2025). - The mean nightly foraging distance from a colony is 13 km or less (Meade et al. 2021, as summarised by Meade et al. 2025). - Extreme heat (42 C or more) suppresses or delays emergence at this camp (Meade et al. 2019). - Colony counts come from CSIRO's National Flying-fox Monitoring Program to 2022 and, after it, the Australian Flying-Fox Monitor, which carries Victoria's counts by the Department of Energy, Environment and Climate Action (DEECA).

2. Data

Source	Use	Licence
Bureau of Meteorology radar 02 (Laverton, S-band, dual-polarisation, 24 km from the camp) and radar 01 (Broadmeadows, C-band, 13 km), via the AURA archive at NCI Australia	the fly-out: 3,287 and 3,129 dusks, 11 October 2017 to 23 November 2026	CC BY-NC-SA 4.0; derived tables only, no radar imagery
BARRA-C2 (Bureau of Meteorology reanalysis), ERA5 (Hersbach et al. 2020)	training weather; the 100 m wind	Bureau and Copernicus terms
ECMWF IFS HRES (about 9 km), from Open-Meteo's open-data archive	the forecast's weather, 2017 to now, at the camp's grid point	CC BY 4.0
CSIRO National Flying-fox Monitoring Program, camp 201 (November 2017 to August 2022, survey month only)	counts to calibrate size; the month's median colony	CC BY 4.0

Source	Use	Licence
Australian Flying-Fox Monitor (J. Welbergen and J. Meade, Western Sydney University), with DEECA's counts, November 2022 to May 2026 (survey dates)	counts to calibrate size (27 counts with CSIRO's)	CC BY-NC-SA 4.0
Vicmap Elevation, Vicmap tree_urban, Vicmap Admin localities (State of Victoria); City of Yarra and City of Melbourne 3D buildings; OpenStreetMap	the 3D map, its places and suburbs	CC BY 4.0; ODbL

3. Measuring the fly-out from radar

3.1 Filter v3 (frozen November 2026)

The filter reads each radar 02 gate within 5 km of the camp, from 20 min before sunset to 72 min after:

1. **Frame.** Azimuth labels are corrected. Radar 02 read 0.52 degrees high before 7 December 2023; this was measured from sun spikes and clutter steps, and confirmed by radar 01.
2. **Biology.** The correlation coefficient (RHOHV) must be below 0.85, and the radial speed 1 to 17 m/s. A limit on differential reflectivity was tested and dropped: against radar 01, 29% of bats exceed 3 dB, and the value rises as the radar looks along the flight.
3. **Outbound.** Where radar 02's beam lies along the line from the camp, the gate needs an outward airspeed of 2.5 m/s or more after removing radar 02's own wind (ERA5's 100 m wind when that fails). Elsewhere, radar 01's nearest gate takes the same test, because radar 02 has blind wedges north and south of the camp.
4. **Size.** Linear reflectivity times area over the outbound gates, at the peak (+5 to +72 min), less the median before sunset. Background subtraction is used because an airspeed cut showed no gap between bats and drifting insects.
5. **Detection.** The peak must exceed the pre-sunset scatter. A **drift flag** marks net inflow at the camp over +5 to +60 min.
6. **Direction.** The outbound profile 2 to 5 km out, in 10-degree bins, within 10 min of the peak. Every core at half the maximum or more that persists in more than half the scans counts.

3.2 Validation

- **Blind pass test.** On 95 dusks, with the random seed and the labels committed before any of the filter's output was read, the filter found 68 of 70 fly-outs: 97.1% recall (Wilson 95% interval 90.2 to 99.2%). The drift flag agreed with the labeller on 17 of 25 dusks labelled none or unclear, including all 8 labelled none. An earlier filter failed its own test (it found no southward fly-outs) and was replaced. *Measured.*
- **Size against counts.** Against 27 counts, Spearman 0.73 and Pearson 0.79 on log scales; the slope is within the counts' own error (Deming 0.85 to 1.10); the 80% prediction interval is x/÷2.5.

Ground counts themselves vary 26 to 55% between counters and run well below thermal-drone counts (McCarthy et al. 2023), so the fit cannot show the radar is better than the counts. *Measured; size is an index, never a headcount.*

- **Speed.** Dual-Doppler from the two radars (250 dusks) shows the dense stream moving outward at 10.7 to 11.3 m/s within 3 min of the peak in the warm season, and 10.0 to 10.3 m/s in the cool. GPS on the species gives 8.5 to 9.5 m/s over whole flights, turns included (sourced), which is consistent. *Measured.*
- **Height.** The echo sits at 100 to 200 m, but a forward model of the beams shows they weight the high fliers. With GPS heights of 21 to 36 m, about 90% of bats are modelled near 30 m and 10% near 225 m. *The low mode sourced; the shares fitted; the two-mode form assumed.*

3.3 Where the stream thins

Bats flying straight out spread over a wider arc with distance, so the echo summed over the dusk, times distance, stays flat along a bearing unless bats leave the stream. Its outward loss rate, less the median at the same distance (so range effects cancel), maps where the stream thins.

Across 3,286 dusks the map is reproducible: even years against odd years correlate 0.86 (warm season) and 0.78 (cool). It is not explained by radar geometry, the hand-over between radars, clutter or beam blockage. It does **not** follow food trees in flower, night sightings or late echo. We therefore describe it as where the stream thins at the radars' heights, not as landing sites. *Measured; the interpretation stated.*

3.4 Before and after the fly-out

A long window (sunset to +3 h, 120 dusks) shows echo milling over the camp before the stream rises on 93% of dusks. Few bats return in the first three hours (2% of the outbound peak). Lingering in parks after the fly-out cannot be measured with these radars: late echo sits on the city and the freeways, consistent with beam bending and traffic. *Measured.*

---

4. Ecological and physical findings, 2017 to 2026

- **The fly-out follows the season.** The measured peak comes a median 28.6 min after sunset in December and January, and 53.5 min after it in June and July: about 25 min later in midwinter (3,139 detected dusks). Darker and cloudier dusks bring it earlier, warmer dusks too. *Measured.*
- **Two corridors.** The main core leaves near 296 degrees (west-north-west) from October to March: 72% of main cores lie between 270 and 300 degrees, and a second core between 30 and 60 degrees (north-east) appears on 11%. From April to September the main core is near 287 degrees. On a typical warm-season dusk, one 22.5-degree sector carries about a third of the stream 4 km out, over Princes Park. *Measured.*

- **Wind lowers the measured size, not the bats' headings.** On 1,960 detected clean dusks, the windiest quarter (10 m wind of 14.9 km/h or more) shows 0.47 times the median size of the calmest quarter (7.9 km/h or less): about 0.55 times per 10 km/h. One radar cannot separate fewer bats leaving from the same bats flying lower, under the beam. Dual-Doppler shows the bats' air headings are independent of the wind, and their ground speed takes 0.2 to 0.4 of a tailwind. *Measured.*
- **Stable across the years.** Against the counts, there is no shift between the years before and after 2022: post-2022 residuals are 0.14 log10 lower on all nights (p = 0.37), and 0.01 lower on calm nights (p = 0.97). With 24 counts, a shift smaller than about a factor of two could not be detected. A radar reading at every dusk complements ground counts made about four times a year: it shows the night-to-night variation the counts cannot. *Measured.*

---

5. Forecast targets

At the camp: the peak minute; the size (log10, against the month's median); the main direction (16 sectors) and the stream's expected share by sector; the onset.

At any point within 9 km: whether the stream passes within 0.75 km, its share there, and its first, peak and thinning minutes.

The point targets use a matched filter. The camp's launch profile, delayed by the flight time to the point, is regressed on the point's echo. A pass is a t-value at or above the 95th percentile of the same statistic with the profile shifted 25 min early or late (t = 2.50, over 1.4 million null fits), so false passes are 5% by construction. A simpler threshold rule passed noise on 75% of dusks at Abbotsford Convent. The matched filter passes 26% there, and 82% at Edi Gardens. *Measured.*

---

6. Models

6.1 Inputs

- Day of year and sunset.
- ECMWF IFS weather at the camp: temperature, cloud by layer, light in the hour before sunset, dew-point depression, rain, the 10 m and 100 m wind, gusts and the pressure change.
- The moon.
- In the "with radar" variant, the measured state of the 30 days before, at the archive's two-day posting lag.

A second batch of inputs (cooling, antecedent rain, heat runs) was pre-registered and tested; it gave no gain and was not adopted. Inputs available only from 2024 were excluded, because to a tree model their absence in earlier years marks the date.

6.2 The comparison

Ten model families were run on the same rolling-origin folds. Each test year from 2020 to 2026 was forecast by a model trained on every earlier dusk (2,022 to 2,557 dusks scored). The families:

- climatology, ridge regression and a generalised additive model;
- LightGBM (Ke et al. 2017) tuned with Optuna (Akiba et al. 2019), XGBoost (Chen and Guestrin 2016), extra trees, a Gaussian process and a multilayer perceptron;
- two hybrids that read the 30 days before each dusk through a gated recurrent unit (GRU; Cho et al. 2014) or a temporal convolutional network (TCN; Bai et al. 2018), joined with tonight's inputs;
- a rolling non-negative stack, and equal-weight pairs.

Every model was tested against the best by the Diebold-Mariano test (Diebold and Mariano 1995), with Newey-West errors (Newey and West 1987) at seven-day lags. A model counts as within noise of the best at p >= 0.05.

6.3 Adopted

- **Peak:** the mean of LightGBM and the GRU hybrid.
- **Size:** the mean of a generalised additive model or LightGBM with the GRU hybrid (with radar); the TCN (without).
- **Direction:** the mean of the TCN, LightGBM and a multinomial logistic model, blended with the month's climatology.
- **Points:** a hybrid of the camp's forecast plus a learned delay (LightGBM on distance, bearing, position and the point's own history). Each value is averaged over a 3 x 3 stencil at ±0.25 km, so a tapped point and a named place 0.2 km apart agree.

The sequence and tree models together beat either alone, because they make different errors. The onset cannot be forecast beyond its seasonal range, and is shown only as that range.

6.4 Uncertainty

The 80% ranges are cross-fitted, locally weighted split-conformal intervals (Vovk et al. 2005; Lei et al. 2018). Each night's expected error is modelled from its inputs, then scaled to 80% using earlier years only.

Ranges by season alone held 80% overall, but covered only 58% of rainy dusks. The adopted ranges cover: - 79 to 82% of all dusks; - 79 to 85% of rainy dusks; - 76 to 83% of cloudy dusks; - 70 to 80% of hot dusks (only 20 of them).

LightGBM's own quantile intervals under-covered (61 to 72%). *Measured.*

6.5 Forecast weather instead of reanalysis

Models trained on reanalysis and run on ECMWF IFS weather lose nothing measurable: the peak with radar, 4.225 against 4.227 min (p = 0.95); the size, 0.230 against 0.232 (p = 0.40). Training on IFS weather gains nothing. These figures are for the models before the dusk-light, cloud-layer, humidity, gust and pressure inputs were added; the adopted models, with those inputs, score 4.00 min (the results below).

The effect of forecast lead time (24-hour-old forecasts) has so far been assessed only by a sensitivity analysis with assumed error sizes: the peak's skill falls from 16% to 12% at typical errors. *Measured; the lead-time effect assumed until tested.*

---

7. Results (out of sample, held-out years)

Target	With radar	Without radar	Climatology
Peak at the camp (MAE)	4.00 min	4.36 min	5.14 min
Peak at a point (MAE)	3.99 min	4.28 min	5.05 min
Size (log10 MAE)	0.220 to 0.223	0.288 to 0.293	0.326
Main direction (log loss, 16 sectors)	1.36	1.46	1.88
The stream passes a point (AUC)	0.906	0.890	0.81
The same, at points never seen in training	0.900	0.878	
Onset	the month's range only		

Held-out lanes. With each 45-degree block withheld in turn, the point model beats climatology in all eight.

Named places (peak MAE, against their climatology): Edi Gardens 4.12 (4.61) min, Princes Park 4.20 (4.81), Melbourne Zoo 4.00 (4.72).

---

8. From the forecast to the page

A daily process at 7:00 am, repeated through the day, writes tonight's and the next six evenings' forecasts. They cover the camp, 23 named places and a grid of 1,554 points: every 0.25 km within 3 km of the camp and over the inner north, every 0.5 km out to 9 km. An afternoon process (3:20 pm, once the radar archive posts the previous day) measures the last dusk and updates the state.

- **The 3D stream** is drawn as particles. The generator's parameters were fitted so that the particles, sampled at the radar's 5-minute interval and read as the radar targets are read, reproduce the measured rise, peak and tail at 1 to 9 km: RMSE 0.30 min (warm season), 0.34 (cool). The fitted speed is about 15% above dual-Doppler's, because near the camp bats are low and climbing

under the beams; it is used as fitted, so the picture matches the point forecast. *Measured fit; form assumed.*

- **Where the stream runs dense.** The share at a point grows with distance as the radars see more of the stream. The map therefore shows the concentration: the share over the median share at the same distance, times the chance. It is drawn in four classes, at 1, 2, 4 and 8 times.
- **Streamlines.** These are equal-flux streamlines of a stream function. Around each ring from the camp, the density is accumulated by bearing, and each of 40 lines keeps one fortieth of the stream. They bunch where the stream is dense, and bend only where the lane moves from ring to ring.

Lines along the gradient of the peak-time field were considered and rejected. Their implied front speed agrees (12.8 m/s, against the particles' 12.9), but between lanes the front moves sideways while the bats do not: a fifth of the stream's area points more than 34 degrees off radial. - **Still flying.** This uses the particles' peel-off: an exponential in distance flown, with the sourced mean (13 km or less, so it gives an upper bound on the bats still flying), shaped by the map of where the stream thins. A place's bats overhead are the stream's division around its ring (measured) times the share still flying. The page shows this as "at most N in 1,000".

Automated verification checks confirm that: - the rendered density integrates to the underlying forecast grid at every node; - each streamline carries 1/40 of the flux, within 0.25%; - the particles reproduce the forecast's lane shares and peel-off; - every figure displayed matches the daily forecast output.

9. What is new, and why it is justified

Element	Closest precedent	Justification here
Two radars and dual-Doppler on one bat colony over nine years	one radar (Meade et al. 2019); a reflectivity mosaic for free-tailed bats (Frick et al. 2012)	a blind pass test; dual-Doppler speeds agree with GPS
Radar airspeed, heading and tailwind response of a fruit bat	radar methods standard for birds and insects (Stepanian et al. 2016); fruit bats by GPS only (Sapir et al. 2014)	standard method, new result, checked against GPS
A daily forecast of a bat emergence	bird migration forecasts (Van Doren and Horton 2018)	rolling-origin back-test, Diebold-Mariano tests, conformal coverage
A forecast at any point (whether the stream passes, and when)	none found	matched-filter targets with 5% false passes by construction; points and lanes withheld from training
Night-specific conformal ranges for an ecological forecast	conformal prediction is standard in statistics	coverage checked by weather stratum
The stream function as a public map of an animal flow	none found	equal flux verified

Our literature search did not systematically cover conference proceedings, agency technical reports or theses, so "none found" means none found, not that none exists.

10. Limits

- Size is an index calibrated to counts that are themselves noisy; it is never shown as a headcount.
- The radars see the upper part of the stream; most bats fly below the beams.
- Radar 02 loses about half the outbound echo behind the camp to the east-north-east, to towers and clutter; the clutter filter takes 3 to 8 dB from bats.
- Where bats land is not known; the thinning map shows where the stream thins at the radars' heights.
- The drop-off with distance rests on a sourced bound (13 km or less) and an assumed exponential form.
- One labeller, who knew the filter's design, labelled the pass test.
- **Radar calibration over 2017 to 2026:** radar 02's calibration was checked against fixed clutter targets (steady at 91.5 to 92.7 dBZ). Its August 2020 scan change lowered the weak background by 5 dB but not the bat measures. Radar 01's long-term calibration has not been verified separately.
- **Forecast lead time:** the test with archived 24-hour-old forecasts is pending; its effect is estimated by sensitivity analysis only.

11. Future work

- The archived-forecast test of lead time.
- GPS tracks from this camp, if they become available, to test the peel-off model and where bats land.
- Radar 01's long-term calibration.
- Regional flowering data, for the influxes that change the colony's size from season to season.
- An independent review of the radar processing and the forecast, invited.

References

- Akiba T, Sano S, Yanase T, Ohta T, Koyama M (2019). Optuna: a next-generation hyperparameter optimization framework. *Proceedings of KDD 2019*.
- Australian Flying-Fox Monitor. J. Welbergen and J. Meade, Western Sydney University (monitoring.flyingfoxes.info). Yarra Bend, camp 201: counts November 2022 to May 2026 from

the Victorian Department of Energy, Environment and Climate Action (DEECA). CC BY-NC-SA 4.0.

- Bai S, Kolter JZ, Koltun V (2018). An empirical evaluation of generic convolutional and recurrent networks for sequence modeling. arXiv:1803.01271.
- Chen T, Guestrin C (2016). XGBoost: a scalable tree boosting system. *Proceedings of KDD 2016*.
- Chilson PB, Frick WF, Kelly JF, et al. (2012). Partly cloudy with a chance of migration: weather, radars, and aeroecology. *Bulletin of the American Meteorological Society* 93:669-686.
- Cho K, van Merrienboer B, Gulcehre C, et al. (2014). Learning phrase representations using RNN encoder-decoder for statistical machine translation. *Proceedings of EMNLP 2014*.
- CSIRO. National Flying-fox Monitoring Program, v4. doi:10.25919/tfhg-ej16. CC BY 4.0.
- Diebold FX, Mariano RS (1995). Comparing predictive accuracy. *Journal of Business and Economic Statistics* 13:253-263.
- Frick WF, Stepanian PM, Kelly JF, et al. (2012). Climate and weather impact timing of emergence of bats. *PLoS ONE* 7:e42737.
- Hersbach H, Bell B, Berrisford P, et al. (2020). The ERA5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society* 146:1999-2049.
- Horn JW, Kunz TH (2008). Analyzing NEXRAD doppler radar images to assess nightly dispersal patterns and population trends in Brazilian free-tailed bats (*Tadarida brasiliensis*). *Integrative and Comparative Biology* 48:24-39.
- Ke G, Meng Q, Finley T, et al. (2017). LightGBM: a highly efficient gradient boosting decision tree. *Advances in Neural Information Processing Systems* 30.
- Lei J, G'Sell M, Rinaldo A, Tibshirani RJ, Wasserman L (2018). Distribution-free predictive inference for regression. *Journal of the American Statistical Association* 113:1094-1111.
- McCarthy ED, Martin JM, Boer MM, Welbergen JA (2023). Ground-based counting methods underestimate true numbers of a threatened colonial mammal: an evaluation using drone-based thermal surveys as a reference. *Wildlife Research* 50(6):484-493. doi:10.1071/WR21120.
- Meade J, van der Ree R, Stepanian PM, Westcott DA, Welbergen JA (2019). Using weather radar to monitor the number, timing and directions of flying-foxes emerging from their roosts. *Scientific Reports* 9:10222.
- Meade J, Martin JM, Welbergen JA (2021). Fast food in the city? Nomadic flying-foxes commute less and hang around for longer in urban areas. *Behavioral Ecology* 32:1151-1162.
- Meade J, Martin JM, McKeown A, Turbill C, Walker MJ, Boardman WSJ, Welbergen JA (2025). Flight behaviour and short-distance homing by nomadic grey-headed flying-foxes: a pilot study. *Movement Ecology* 13:22. doi:10.1186/s40462-025-00532-x.
- Newey WK, West KD (1987). A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica* 55:703-708.
- Sapir N, Horvitz N, Dechmann DKN, Fahr J, Wikelski M (2014). Commuting fruit bats beneficially modulate their flight in relation to wind. *Proceedings of the Royal Society B* 281:20140018.

- Stepanian PM, Horton KG, Melnikov VM, Zrnić DS, Gauthreaux SA (2016). Dual-polarization radar products for biological applications. *Ecosphere* 7(11):e01539. doi:10.1002/ecs2.1539.
- Van Doren BM, Horton KG (2018). A continental system for forecasting bird migration. *Science* 361:1115-1118.
- Vovk V, Gammerman A, Shafer G (2005). *Algorithmic Learning in a Random World*. Springer.
- Welbergen JA (2006). Timing of the evening emergence from day roosts of the grey-headed flying fox, *Pteropus poliocephalus*: the effects of predation risk, foraging needs, and social context. *Behavioral Ecology and Sociobiology* 60:311-322.

Data. Bureau of Meteorology radar via AURA at NCI Australia (CC BY-NC-SA 4.0). ECMWF IFS via Open-Meteo (CC BY 4.0). BARRA-C2 (Bureau of Meteorology). ERA5 (Copernicus Climate Change Service). Flying-fox counts: CSIRO (CC BY 4.0); the Australian Flying-Fox Monitor with DEECA (CC BY-NC-SA 4.0). Vicmap (State of Victoria, CC BY 4.0). OpenStreetMap contributors (ODbL). The derived radar tables are shared under CC BY-NC-SA 4.0.