INSTITUTE FOR DEFENSE ANALYSES

Prediction Uncertainty for Autocorrelated Lognormal Data with Random Effects

Laura J. Freeman, *Project Leader*
Matthew R. Avery

July 2017

IDA NS D-8629

Log: H 2017-0000425

INSTITUTE FOR DEFENSE ANALYSES
4850 Mark Center Drive
Alexandria, Virginia 22311-1882

About This Publication

A common use case in operational testing is estimating system performance across future operating conditions. When test data is generated through complex processes, it can become difficult to generate an estimate of uncertainty that is based on sound statistical theory yet remains intuitive to understand. The motivating data come from radar tracks of aircraft. Track accuracy is correlated both within each flight generally and from one second to the next. This presentation discusses multiple approaches for uncertainty quantification for this challenging use case.

Acknowledgments

Technical review was performed by Laura J. Freeman, Colin E. Anderson, and Kelly M. Avery from the Operational Evaluation Division.

For more information:
Laura J. Freeman, Project Leader
lfreeman@ida.org • (703) 845-2084

Robert R. Soule, Director, Operational Evaluation Division
rsoule@ida.org • (703) 845-2482

INSTITUTE FOR DEFENSE ANALYSES

IDA NS D-8629

Prediction Uncertainty for Autocorrelated Lognormal Data with Random Effects

Laura J. Freeman, *Project Leader*
Matthew R. Avery

Executive Summary

Accurately presenting model estimates with appropriate uncertainties is critical to the credibility and defensibility of any piece of statistical analysis. When dealing with complex data that require hierarchical covariance structures, many of the standard approaches for visualizing uncertainty are insufficient. One such case is data fit with log-linear autoregressive mixed effects models. Data requiring such an approach have three exceptional characteristics:

1. The data are sampled in "groups" that exhibit variation unexplained by other model factors.
2. The data are sampled over time and exhibit autocorrelation.
3. The data originate from a skewed distribution.

These data are addressed using a log-linear autoregressive mixed model (LLARMM), which accounts for each of these characteristics.

A. Stream-Fire Data

One example of this type of data in the realm of Defense Department testing comes from operational testing of electronic countermeasures designed to improve survivability of aircraft against command-guided surface-to-air missiles. Data from these tests can include so-called "stream-fire" data, where an aircraft equipped with the system under test flies through the weapon engagement zone of a simulated threat radar system. Throughout the aircraft's flight, simulated missiles are "stream-fired" at regular intervals. The closest point of approach for each simulated missile is calculated via digital simulation, resulting in a series of miss distances over the course of the aircraft's flight. The process is repeated over multiple sorties, with the aircraft performing maneuvers according to a test plan.

These miss distances exhibit the three characteristics mentioned above. Empirical results show that sorties exhibit variation unexplained by physical factors like aircraft type, maneuver, and countermeasures. Miss distances observed sequentially within a single sortie are highly correlated. The distribution of simulated miss distances has a long right-side tail, consistent with a lognormal distribution.

These data are used for evaluating critical countermeasures systems on U.S. strike aircraft, and reporting estimates with accurate uncertainties is crucial for informing decision makers.

B. Quantifying Uncertainty

There are many questions that a LLARMM can address, but we will focus on three here.

1. How does average system performance change with covariates?

This is a broad question that decision makers interested in comparisons (such as "wet" vs. "dry" scenarios, or comparisons between two countermeasures systems) may ask. One approach for answering this question is by reporting average performance across the operational envelope. For comparisons to be useful, uncertainties about these estimates should be included, and the simplest way to do this is by including a confidence interval for mean performance.

2. What is the probability that a missile closes within its lethal radius?

This is a question typically asked when evaluating a system's requirement. To answer this question, it is necessary to look not at the average miss distance but at the probability that the miss distance is low enough for the missile to damage the aircraft. These probabilities can be visualized using prediction intervals. Plotting prediction bounds across the operational space can give a quick visual indicator of the conditions where the system meets its threshold and where it fails to meet its threshold.

3. What does miss distance look like over the course of a flight when the pilot executes a "real" maneuver?

This is the most operationally-direct and mission-oriented question. Answering it leverages the unique characteristics of these data, which may include information about the maneuvers executed during sorties and how the simulated miss distances change with those maneuvers. Once again, prediction intervals are useful, but in this case we estimate miss distance across a particular sortie rather than across the operational envelope. The interval width is estimated differently because of the autocorrelation within sorties. Plots of these prediction bands show performance not across a hypothetical battlespace but for a realistic maneuver. The goal of these countermeasures systems is to protect U.S. aircraft and pilots throughout a sortie. Considering the full sortie as a unit may be more useful for decision makers and warfighters.

C. Conclusions and Future Work

Uncertainty estimates remain vital for understanding system performance, even when complex models are used. As data and their accompanying statistical models become more sophisticated, it is important that we consider the right approaches for quantifying uncertainty around our estimates. The approaches identified here are useful for quantifying uncertainty for various applications of LLARMMs.

The interval estimates presented above are centered on system performance for an average sortie. Empirical evidence from operational test suggests that miss distance exhibits substantial variation from sortie to sortie, even after accounting for other factors that affect performance. Future work on this topic will include generalizing these approaches for cases where we wish to look at differences in performance across sortie that are not attributable to the aircraft's maneuver. Additionally, in order to make these methods more accessible, we will be working to make simple implementations of these methods available via an R package.

Prediction Uncertainty for Autocorrelated Lognormal Data with Random Effects

Matthew Avery
Institute for Defense Analyses

JSM 2017

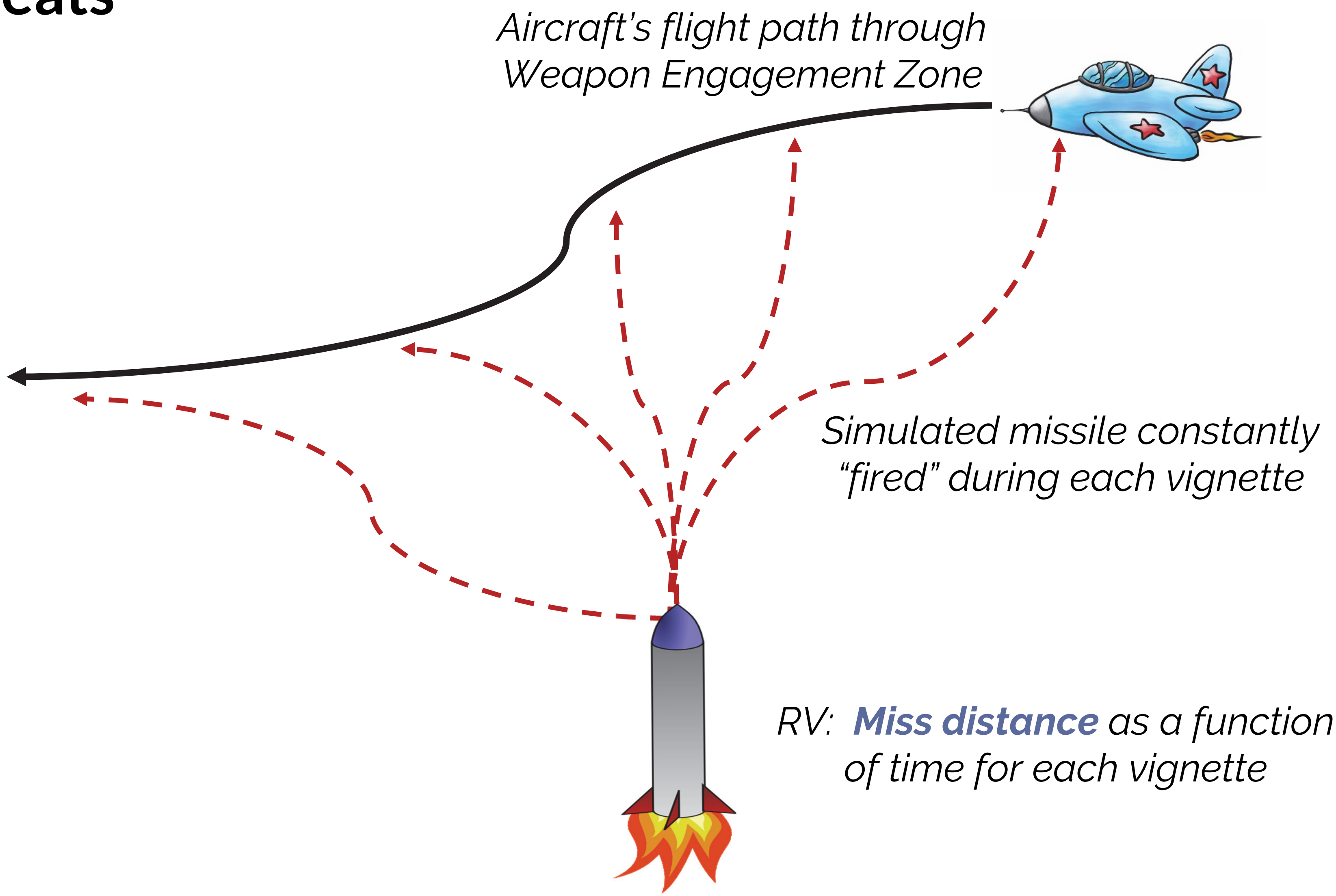
For mixed models with special covariance structures, there are many approaches to quantifying uncertainty

Special case: Characterizing uncertainty for log-linear autoregressive mixed models

Goal of Operational Testing: Estimate system performance across operational space

In context, that means model predictions (rather than parameter estimates) across all relevant combinations of factors

Motivating example: Flight test data from fighter aircraft against simulated missile flyouts from surface threats

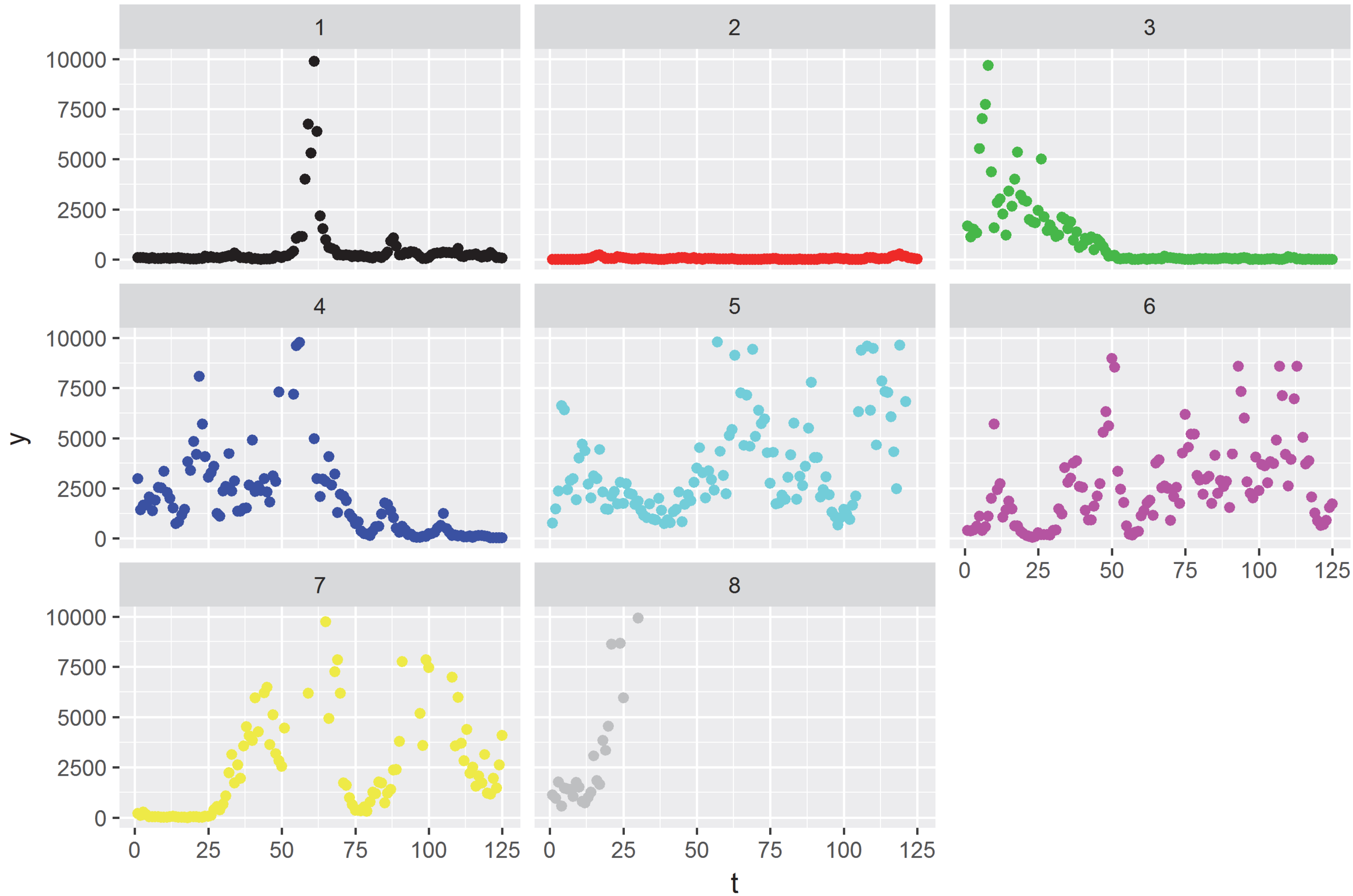


Track error at time t impact miss distances at time $t+1$, $t+2$, …, $t+m$.

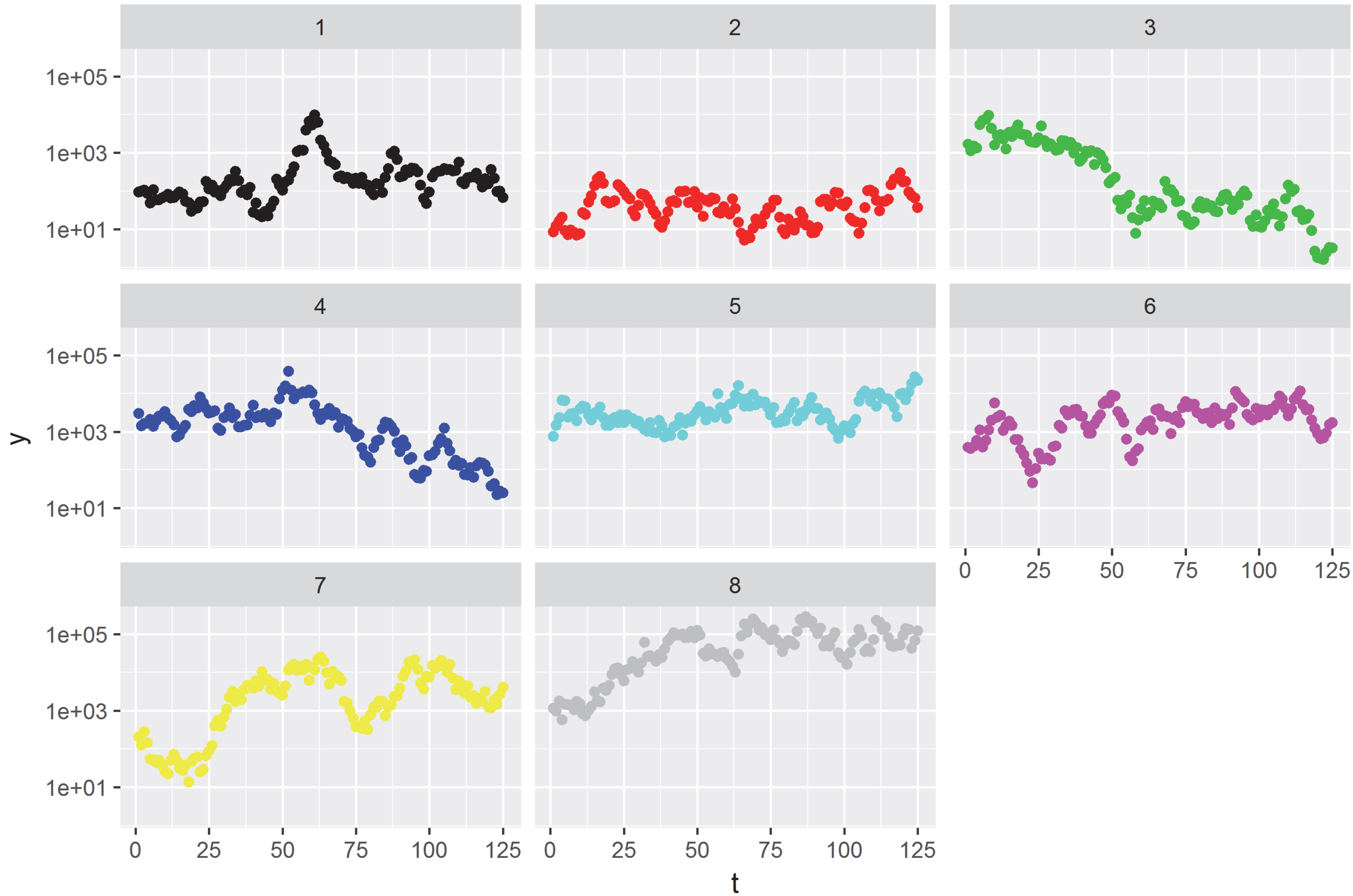
What types of questions are we trying to answer?

1. How does average system performance change with the covariates?
2. What is the probability that a missile closes within its lethal radius?
3. What does miss distance look like over the course of a flight when the pilot executes a “real” maneuver?

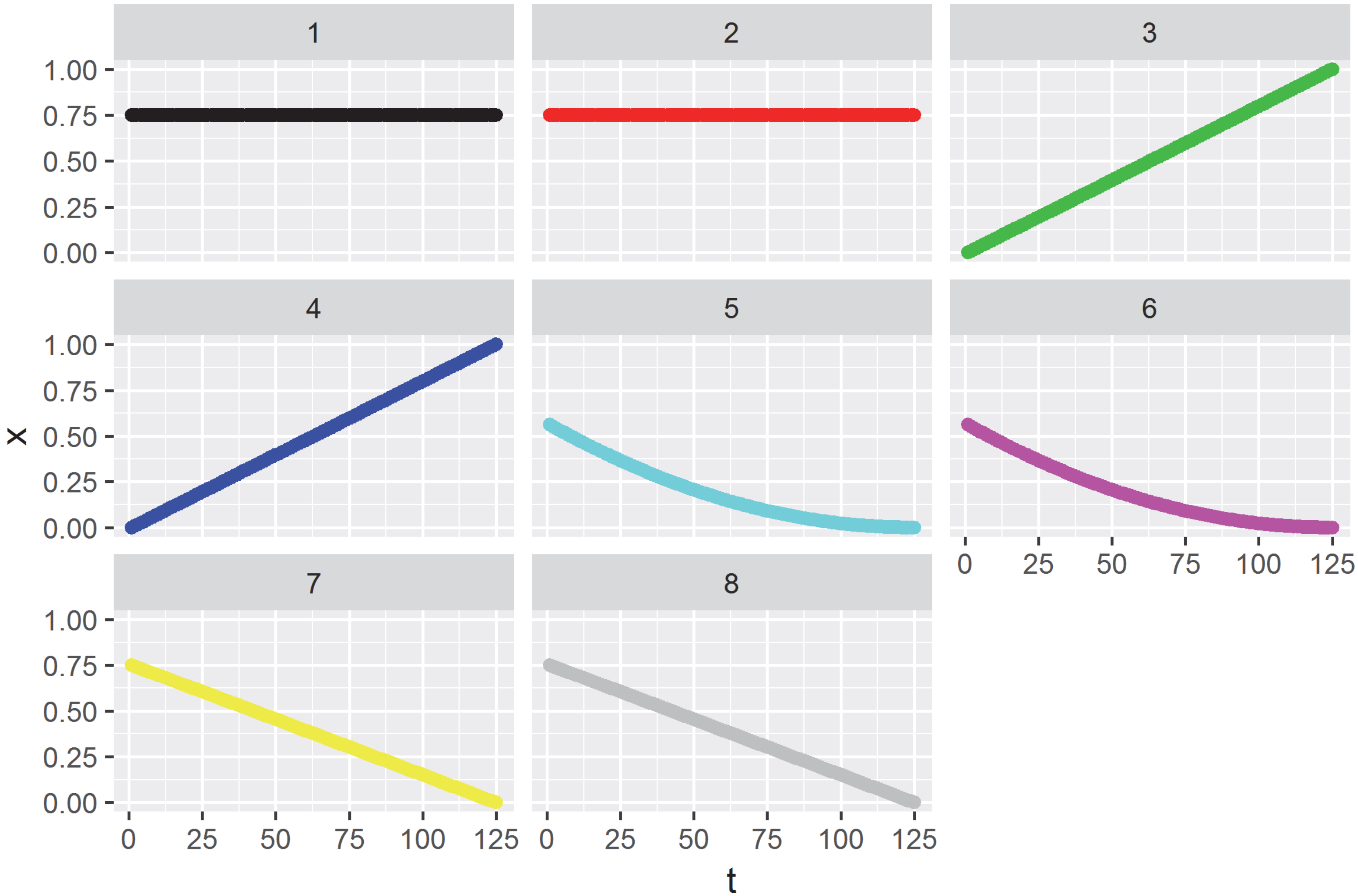
Miss distance over time (simulated; sorry!)



It looks a lot better on the log scale!



Pilots may take evasive action while in range of a threat



Loglinear autoregressive mixed models (LLARMMs)

For the *i*th run,

$$\log(\boldsymbol{Y}_i)|G_i = g_i \sim N(\boldsymbol{x}_i\beta + g_i, \Sigma_\mathrm{i})$$

$$G_i \sim N(0, \sigma_G^2)$$

where there are n_i observations for the *i*th run.

AR(1) variance-covariance matrix

$$\Sigma_i = \sigma^2 \begin{pmatrix} 1 & \rho & \rho^2 & \cdots & \rho^{n_i-1} \\ \rho & 1 & \rho & \cdots & \rho^{n_i-2} \\ \rho^2 & \rho & 1 & \cdots & \rho^{n_i-3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho^{n_i-1} & \rho^{n_i-2} & \rho^{n_i-3} & \cdots & 1 \end{pmatrix}$$

Hence, given a particular group,

$$var(Y_t|Y_{t-1}) = (1 - \rho^2) * \sigma^2$$

Each component of this model is included to address specific characteristics of the data

Long tail <- lognormal distribution

Run-to-run variability <- random effect

Autocorrelation with runs <- AR(1) covariance structure

There are many drivers of uncertainty for these models

1. σ^2
2. $SE(\hat{\beta})$
3. σ_G^2
4. $SE(\hat{g})$
5. ρ

Recall:

$$\log(\boldsymbol{Y}_i)|G_i = g_i \sim N(\boldsymbol{x}_i\beta + g_i, \Sigma_\mathrm{i})$$

$$G_i \sim N(0, \sigma_G^2)$$

Note: The standard errors for the variance parameters are only relevant to us if we wish to look at the uncertainty of our CI/PIs.

How does average system performance change with the covariates?

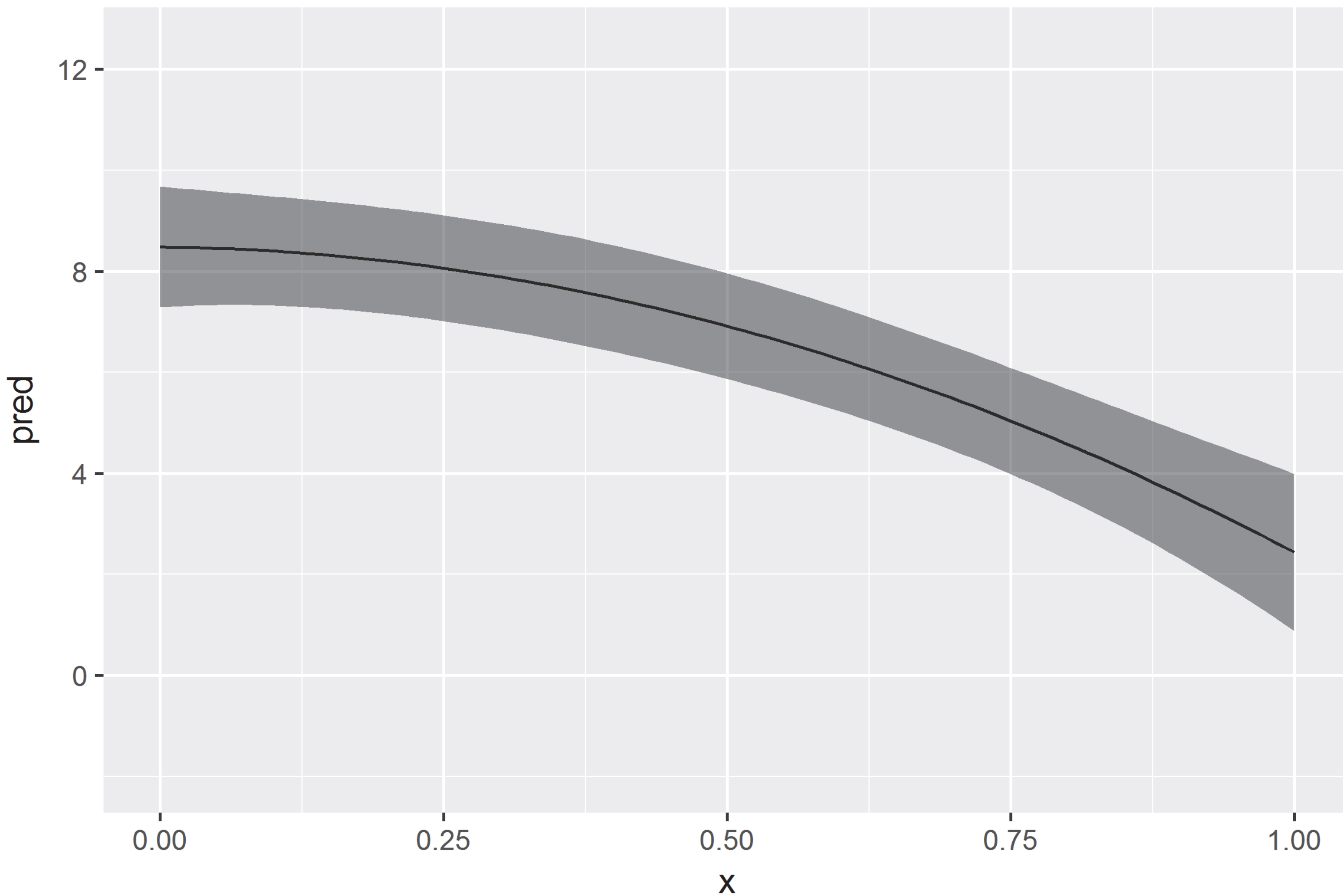
$$E\left(\hat{Y}_{+} | x_{+}\right) = x_{+}\hat{\beta}$$

$Var\left(\hat{Y}_{+}\right)$ depends on $SE\left(x_{+}\hat{\beta}\right)$

Wald-type interval:

$$x_{+}\beta \pm SE\left(x_{+}\hat{\beta}\right) * c_{\alpha}$$

Confidence band interval for Y(x)



What is the probability that a missile closes within its lethal radius?

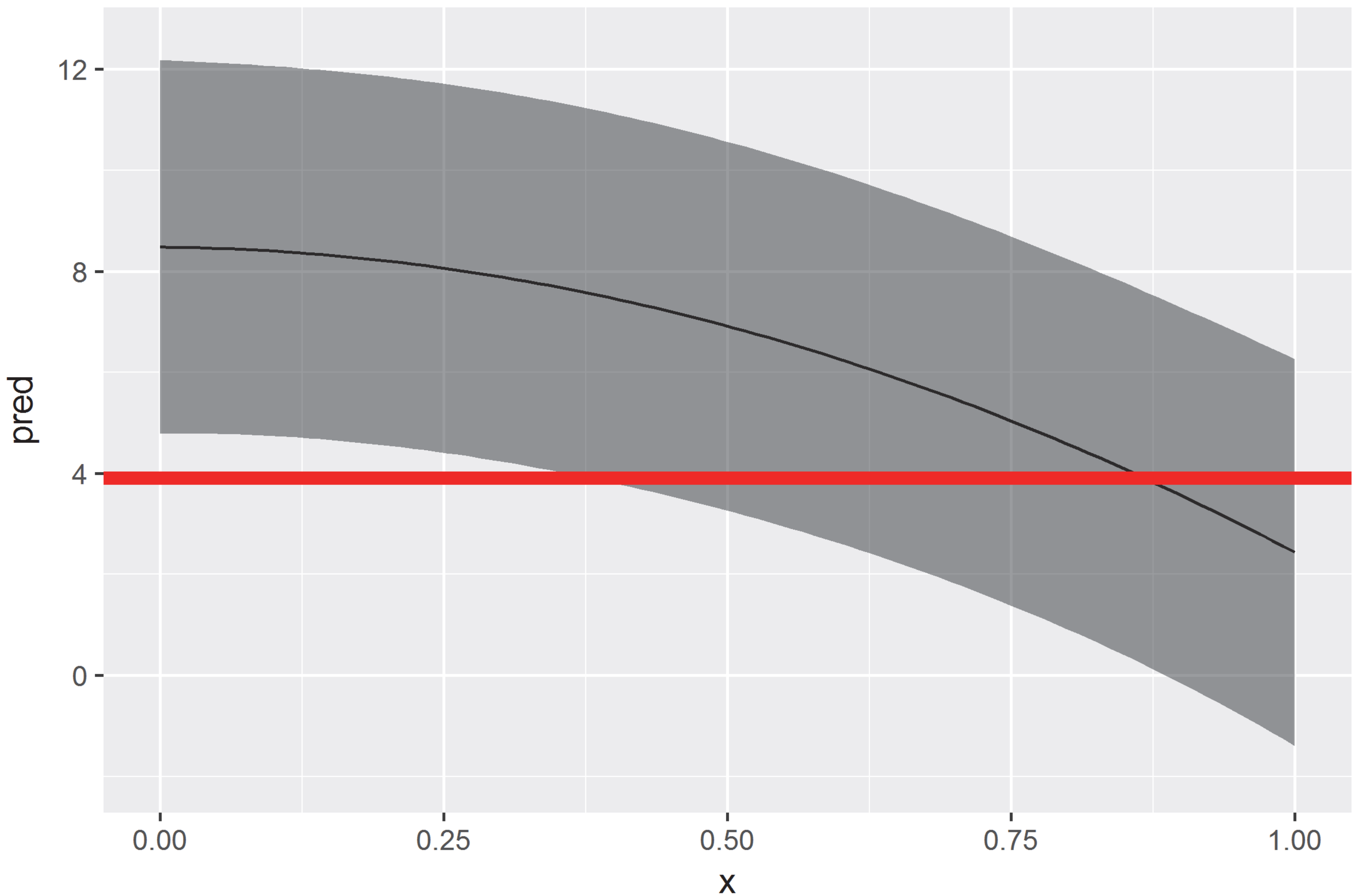
$$P\left(\hat{Y}_+ < R | x_+\right) = P(x_+\hat{\beta} + G_+ + \epsilon_+ < R)$$

$Var\left(\hat{Y}_+\right)$ depends on $SE\left(\hat{\beta}\right)$, σ^2 and σ_G^2.

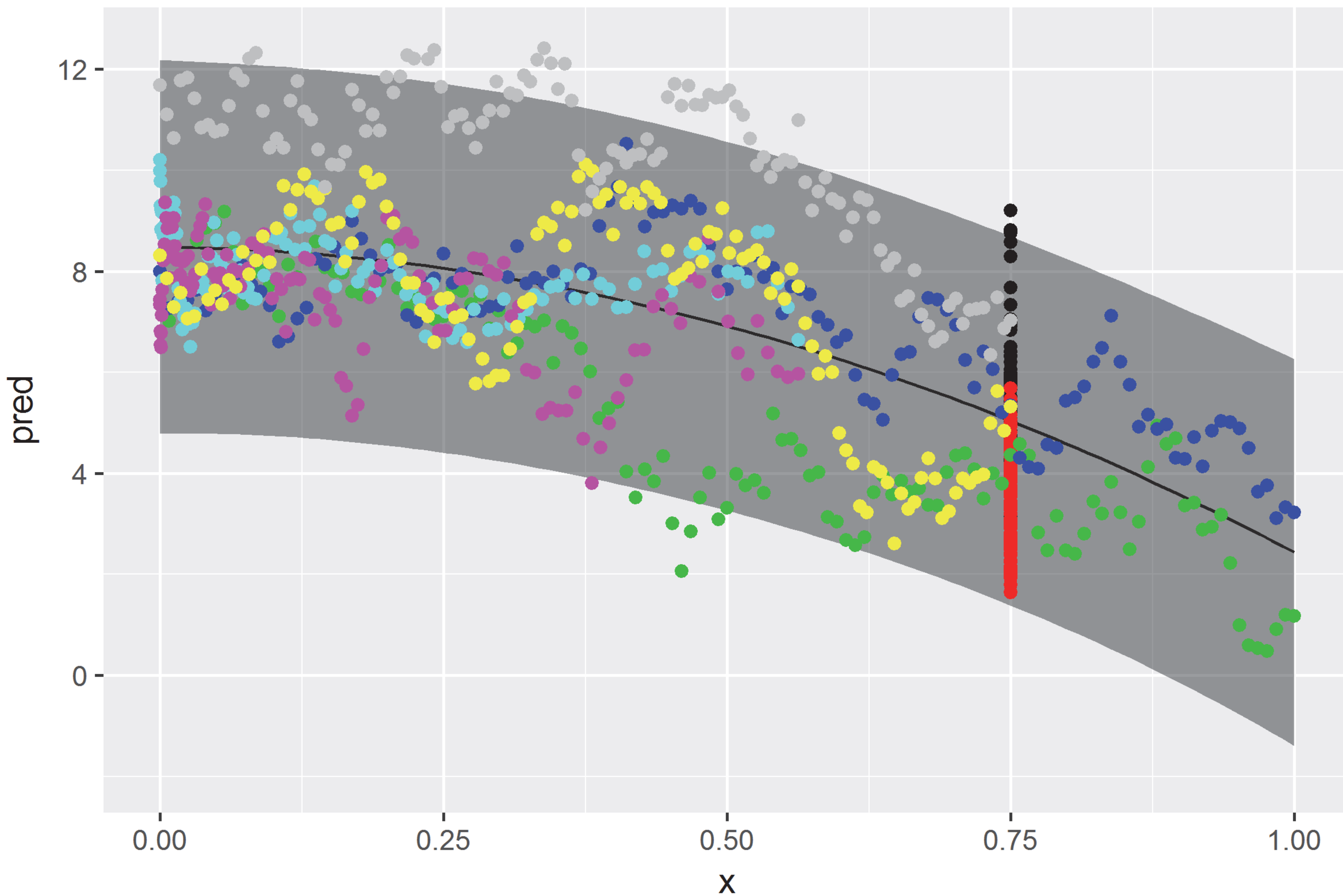
Wald-type prediction interval:

$$x_+\beta \pm \left(SE\left(x_+\hat{\beta}\right)^2 + \sigma^2 + \sigma_G^2\right)^{\frac{1}{2}} * c_\alpha$$

Prediction interval for Y(x)



Probability band with data



What does miss distance look like over the course of a flight when the pilot executes a "real" maneuver?

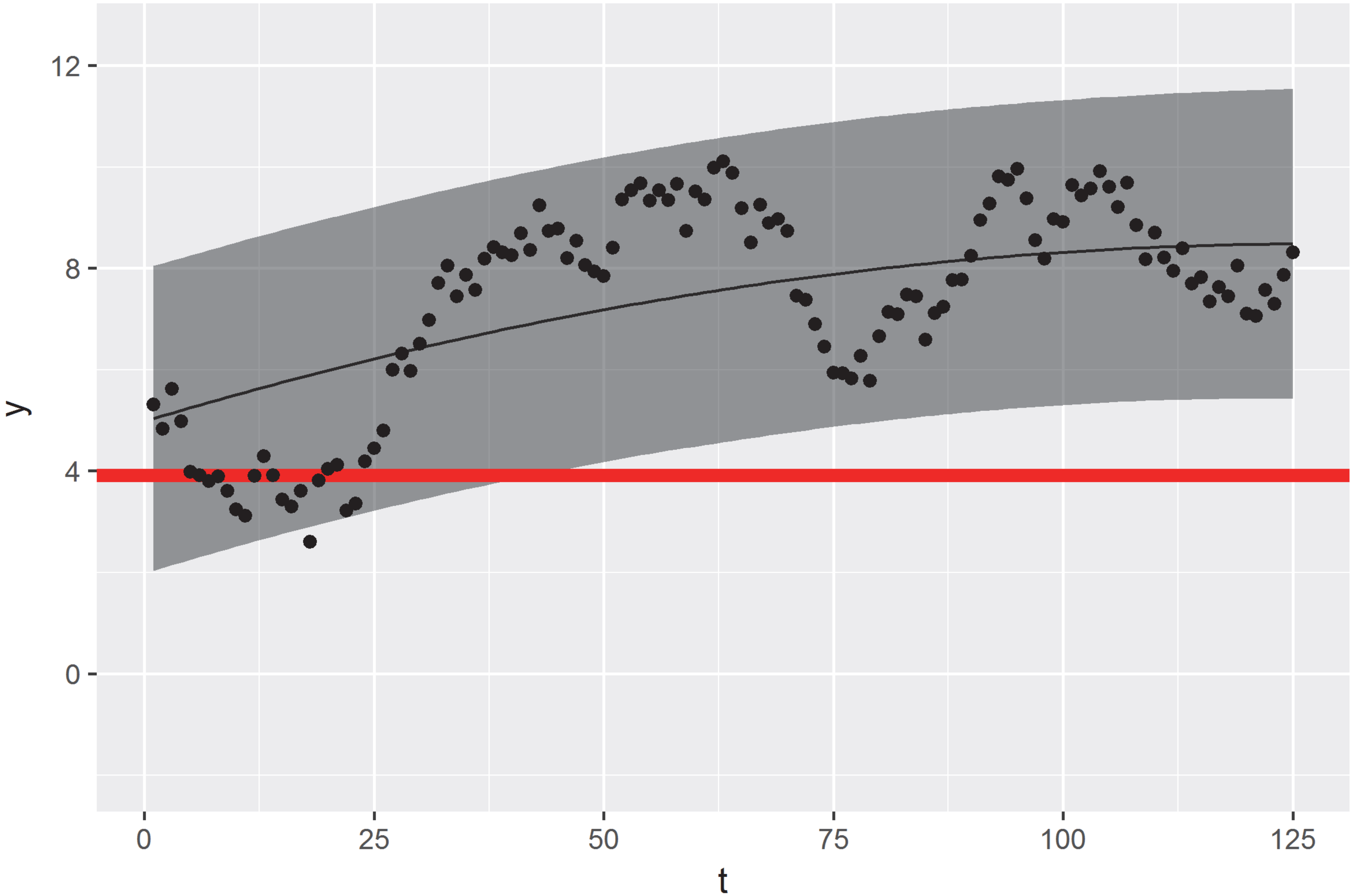
$$P\left(\widehat{\boldsymbol{Y}}_{+} < R \middle| \boldsymbol{x}_i\right) = P(\boldsymbol{x}_i\hat{\beta} + G_{+} + \boldsymbol{\epsilon}_{+} < R)$$

$Cov\left(\widehat{\boldsymbol{Y}}_{+}\right)$ depends on $SE\left(\hat{\beta}\right)$, σ^2 and σ_G^2, and ρ.

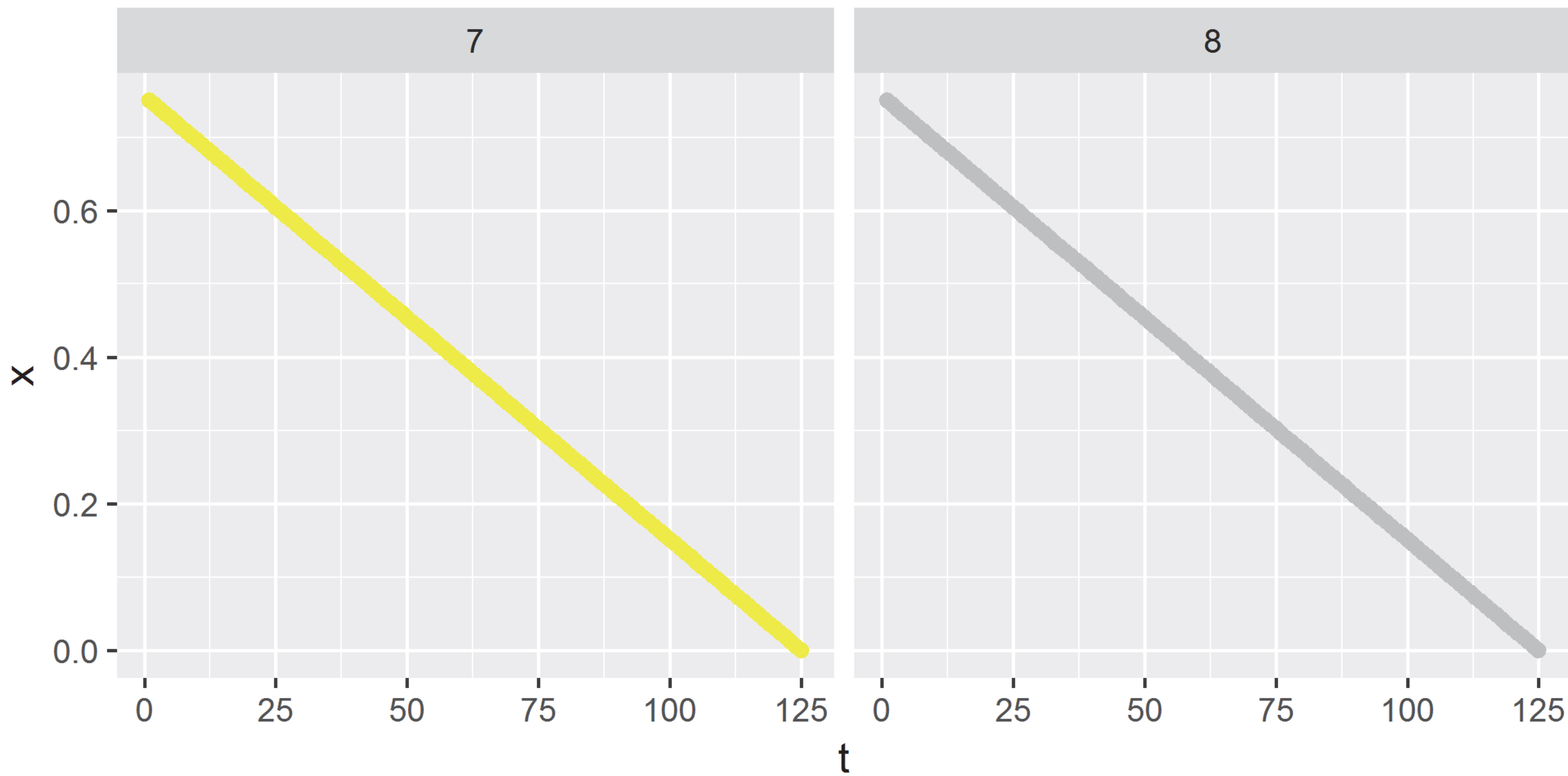
Wald-type prediction interval:

$$\boldsymbol{x}_i\beta \pm \left(SE\left(\boldsymbol{x}_i\hat{\beta}\right)^2 + (1-\rho^2) * \sigma^2 + \sigma_G^2\right)^{\frac{1}{2}} * c_\alpha$$

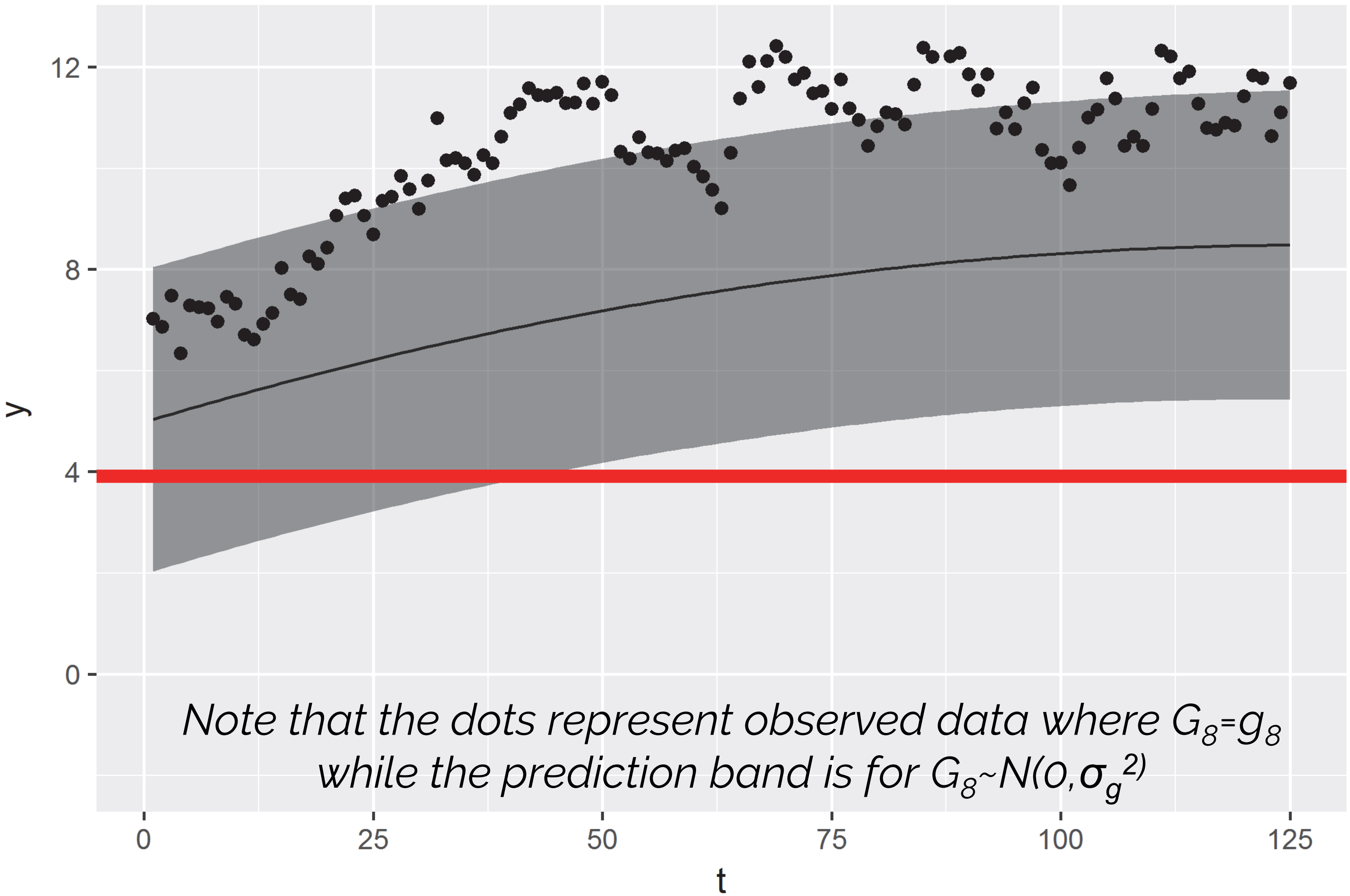
$Y|x_7(t)$



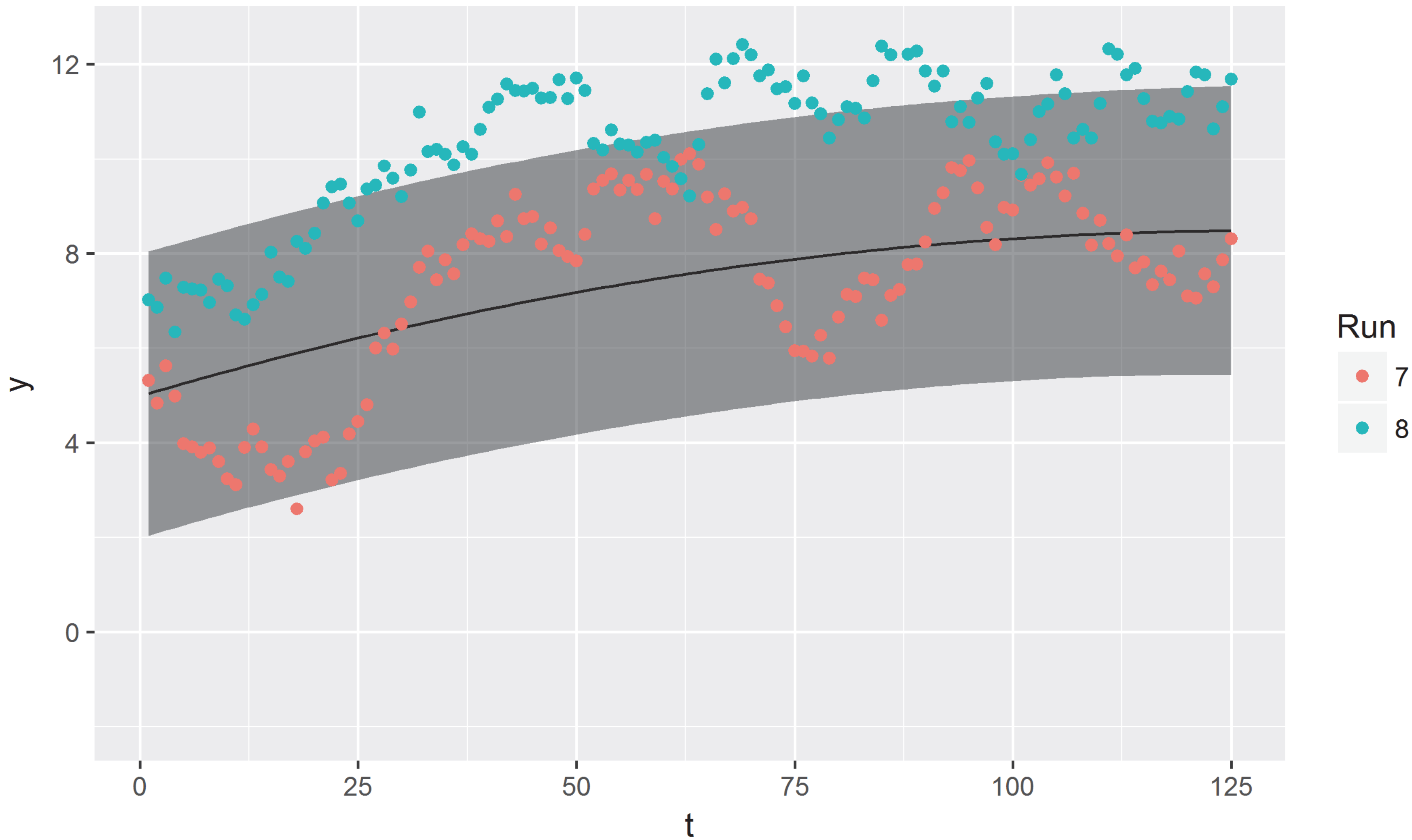
Runs that have the same maneuver will have the same prediction band


7
8
x
0.6
0.4
0.2
0.0
0
25
50
75
100
125
t

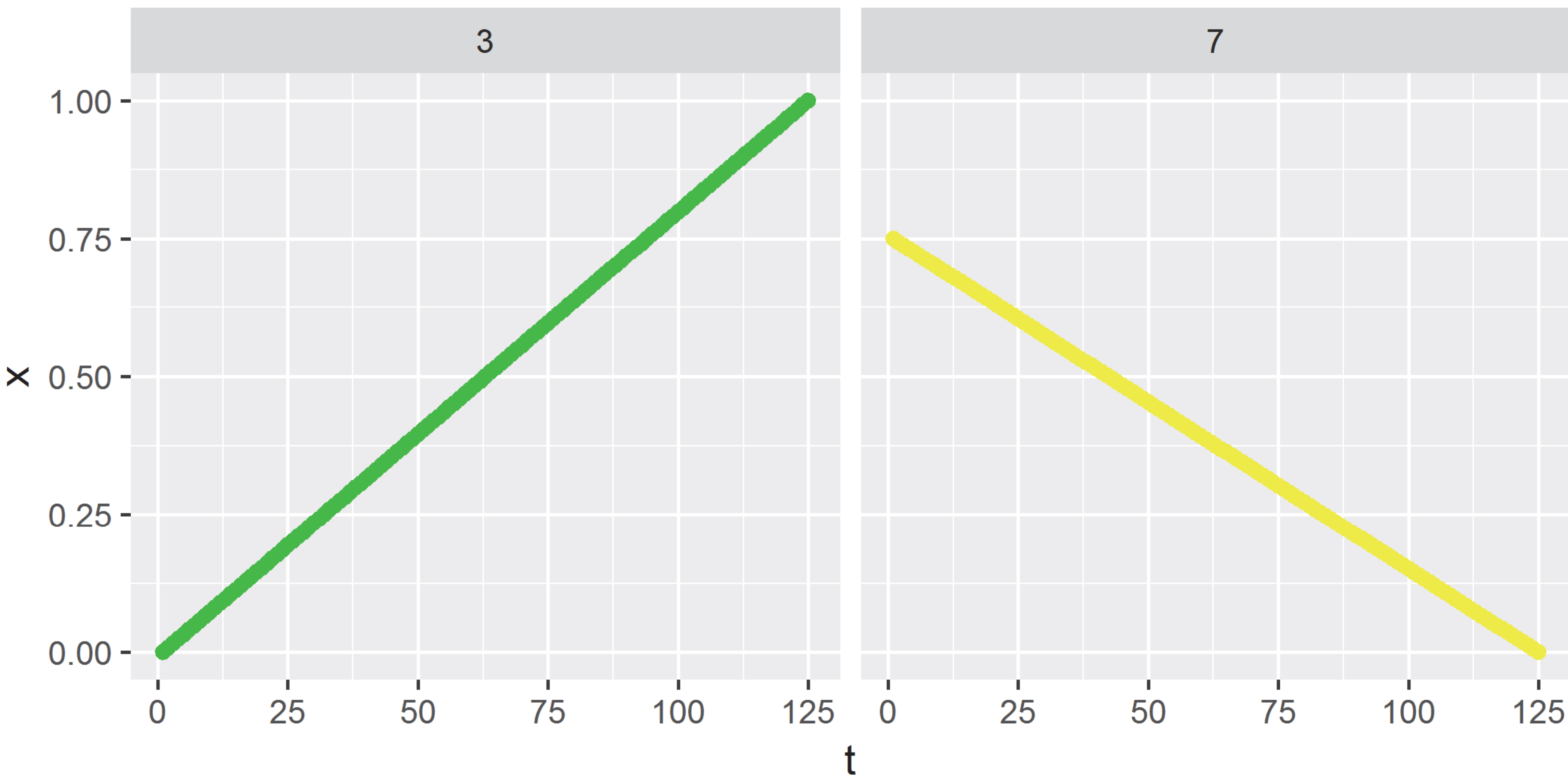
$Y|\mathbf{x}_8(t)$


12
8
4
0
y
0
25
50
75
100
125
t
Note that the dots represent observed data where G_8=g_8
while the prediction band is for G_8~$N(0,\sigma_g^2)$

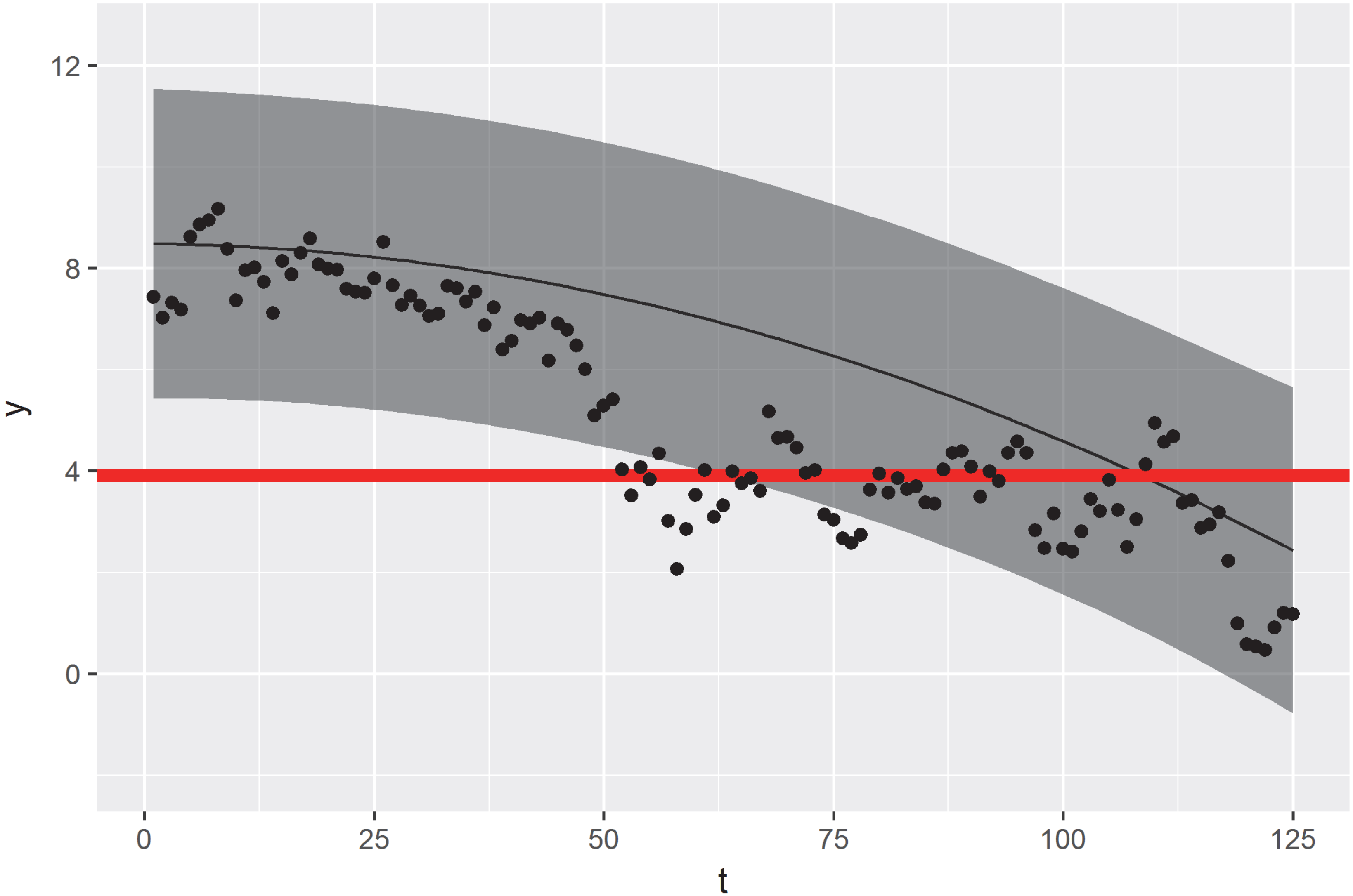
Probability band for runs 7 and 8



Runs that have the different maneuvers will have different prediction bands


3
7
1.00
0.75
0.50
0.25
0.00
x
0
25
50
75
100
125
t

$y|x_3(t)$



Extensions and future work

Conditional CIs and PIs:

$$E\left(\hat{Y}_+ | x_+, g_i\right) = x_+\hat{\beta} + \hat{g}_i$$
$$P\left(\hat{Y}_+ < R | x_+, g_i\right) = P(x_+\hat{\beta} + \hat{g}_i + \epsilon_+ < R)$$
$$P\left(\widehat{\boldsymbol{Y}}_i < R | \boldsymbol{x}_i, g_i\right) = P(\boldsymbol{x}_i\hat{\beta} + \hat{g}_i + \boldsymbol{\epsilon}_+ < R)$$

Implementation in R through developmental package, `ciTools`.