Ch 11. Propagating analyzing, and reducing uncertainty

Amy Hurford

Memorial University

Sensitivity analysis

- How the change in a parameter, translates into a change in a quantity, i.e., an eigenvalue, an equilibrium, a variable at time, T.

- Sensitivities are slopes

- Elasticities standardize by the mean of the variable.

Dietz 2017. Ecological forecasting.

Variables and parameters

$$\frac{dS(t)}{dt} = \theta - \beta S(t)I(t) - \mu S(t),$$

$$\frac{dI(t)}{dt} = \beta S(t)I(t) - \gamma I(t) - \mu I(t),$$

$$\frac{dR(t)}{dt} = \gamma I(t) - \mu R(t).$$

Sensitivity

$$R_0 = \frac{\beta S(t)}{\gamma + \mu}, \qquad T_D = \frac{\log(2)}{\beta S(t) - \gamma - \mu}, \qquad I^* = \frac{\theta - \mu \frac{\gamma + \mu}{\beta}}{\gamma + \mu}.$$

$$\frac{\partial R_0}{\partial \beta} = \frac{S(t)}{\gamma + \mu}, \qquad \frac{\partial T_D}{\partial \beta} = -\frac{S(t) \log(2)}{(\beta S(t) - \gamma - \mu)^2},$$

$$\frac{\partial R_0}{\partial \gamma} = -\frac{S(t)}{(\gamma + \mu)^2}. \qquad \frac{\partial T_D}{\partial \gamma} = \frac{\log(2)}{(\beta S(t) - \gamma - \mu)^2}.$$

Sensitivity

$$I^* = \frac{\theta - \mu\frac{\gamma+\mu}{\beta}}{\gamma + \mu}$$

$$\frac{\partial I^*}{\partial \beta} = \frac{\theta\beta(\gamma + \mu) - \beta\theta + \mu(\gamma + \mu)}{(\beta(\gamma + \mu))^2},$$

$$\frac{\partial I^*}{\partial \gamma} = \frac{\mu\beta(\gamma + \mu) - \beta^2\theta + \mu\beta(\gamma + \mu)}{(\beta(\gamma + \mu))^2}.$$

Previous models

$$y_t \sim N(x_t, \tau_{obs})$$ Data model

$$x_{t+1} \sim N(x_t, \tau_{add})$$ Process model

$$1/\tau_{obs} \sim Gamma(a_{obs}, r_{obs})$$ Observation prior

$$1/\tau_{add} \sim Gamma(a_{add}, r_{add})$$ Process prior
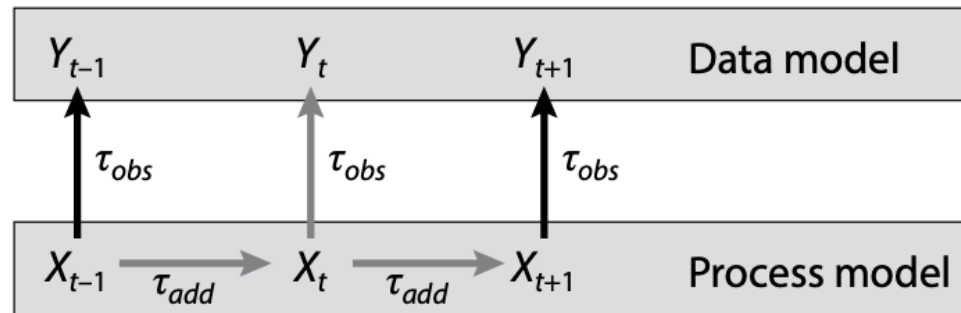
$$x_0 \sim N(x_{IC}, \tau_{IC})$$ Initial condition prior

FIGURE 8.1. State-space model describing the evolution of the latent state variable, X, conditional on the observations, Y. In this random-walk example the only components are observation error, τ_{obs}, and process error, τ_{add}. The gray arrows indicate the connections relevant for estimating the posterior distribution for X_t.

Dietz 2017. Ecological forecasting.
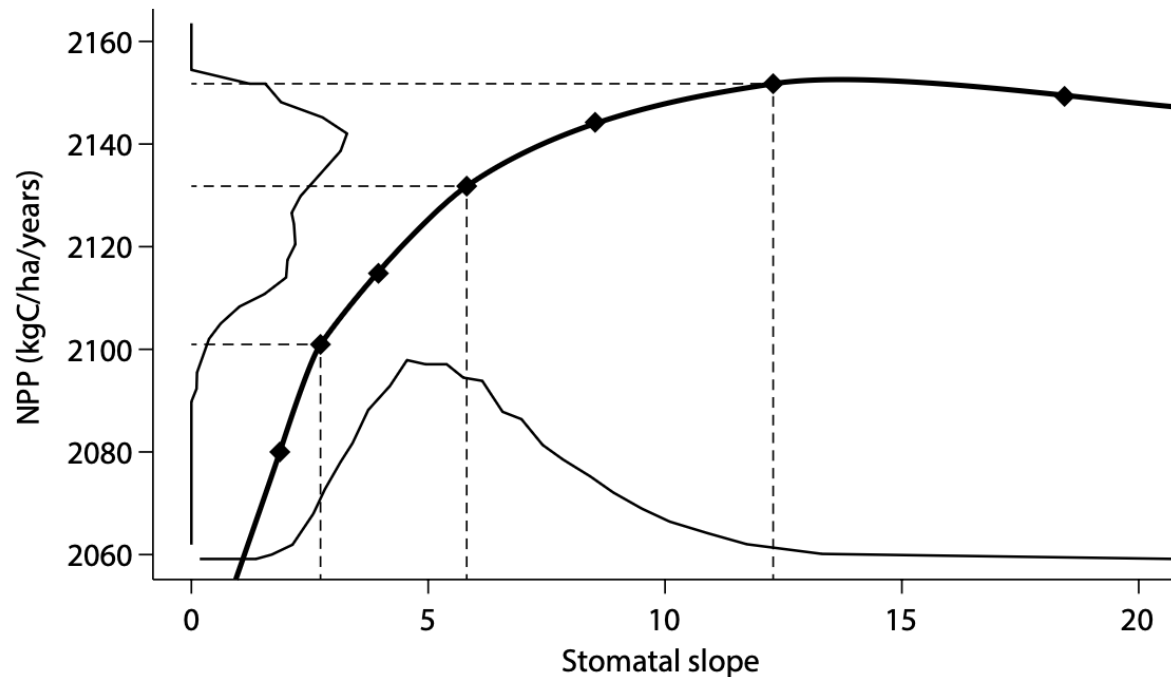
Local sensitivity - OAT

FIGURE 11.1. Translation of uncertainty in an input parameter into uncertainty in an output. The bold black curve illustrates the sensitivity of NPP to stomatal slope, with the black diamonds denoting the median and quantile equivalents of ±1,2,3 standard deviations for stomatal slope based on a one-at-a-time sensitivity analysis. The probability density on the x-axis represents the uncertainty in one model input—in this case, the stomatal slope (Leuning 1995) parameter in the Ecosystem Demography terrestrial biosphere model (Medvigy et al. 2009). The probability density on the y-axis is the uncertainty in NPP attributable to the uncertainty in stomatal slope, with the dashed lines indicating the translation of specific values. Reproduced from Dietze et al. (2014).

Dietz 2017. Ecological forecasting.

Definitions

- Local sensitivity – applies at one location in parameter space.

- Global sensitivity – considers all parameter values.

- One-at-a-time sensitivity (OAT) – all parameters equal to their mean and vary one;

Dietz 2017. Ecological forecasting.

Global sensitivity - Monte Carlo

- Brute force – computationally demanding

- Sample from parameter space and run model. Uniform is common, but not necessarily recommended (?)

- Never truly global - sensitive to the domain chosen. Recommends setting the range to 95% or 99% CI.

- Estimate the sensitivity by fitting a multiple regression to the output as a function of the randomly sampled inputs
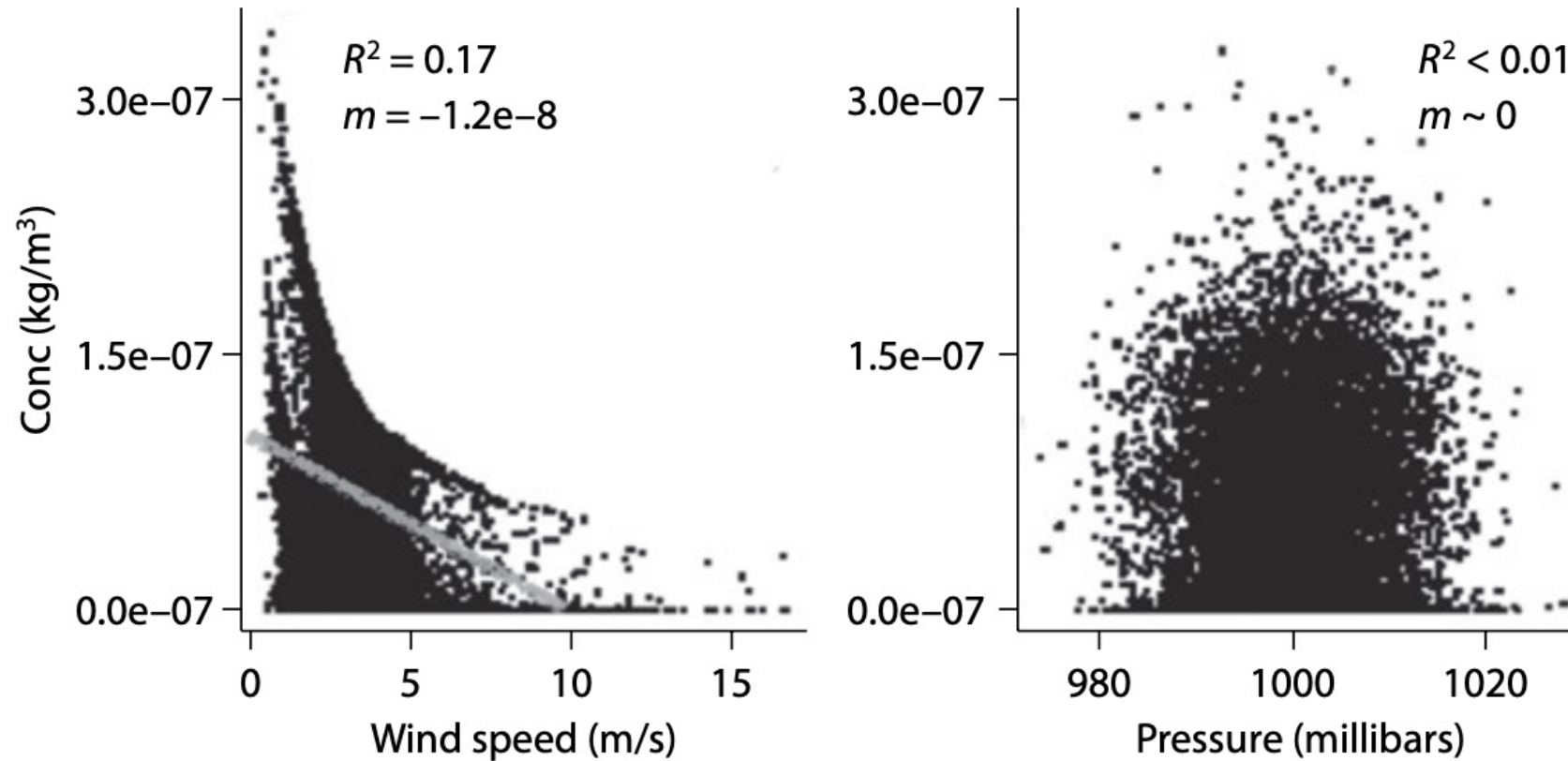
Global sensitivity - Monte Carlo

FIGURE 11.2. Monte Carlo sensitivity analysis of Gaussian Plume atmospheric dispersion model to variability in wind speed and atmospheric pressure. Sensitivity is approximated by the slope, m, while uncertainty partitioning is approximated by the R^2.

Dietz 2017. Ecological forecasting.

Global sensitivity – other approaches

- Entire books on this topic, i.e., Saltelli et al. 2008

- For computational efficiency:
  - Halton sequence
  - Sobol's LP sequence
  - <u>Latin Hypercube Sampling</u>
  - Use an emulator (Box 11.1)

Latin Hypercube Sampling

Sensitivity and Uncertainty Analysis of Complex Models of Disease Transmission: an HIV Model, as an Example

S.M. Blower and H. Dowlatabadi†

School of Public Health (Epidemiology) 140, Earl Warren Hall, University of California, Berkeley, California 94720, USA. †Department of Engineering & Public Policy, Carnegie Mellon University, Pittsburgh, PA 15213 USA.

Summary

HIV transmission models have become very complex. The behavior of some of these models may only be explored by uncertainty and sensitivity analyses, because the structural complexity of the model are coupled with a high degree of uncertainty in estimating the values of the input parameters. Uncertainty analysis may be used to assess the variability (prediction imprecision) in the outcome variable that is due to the uncertainty in estimating the input values. A sensitivity analysis can extend an uncertainty analysis by identifying which parameters are important in contributing to the prediction imprecision (i.e., how do changes in the values of the input parameters alter the value of the outcome variable). In this paper an uncertainty and a sensitivity analysis are described and applied; both analyses are based upon the Latin Hypercube Sampling (LHS) scheme, which is an extremely efficient sampling design proposed by McKay, Conover & Beckman (1979). The methods described in this paper have not previously been applied to deterministic models of disease transmission, although these models have many characteristics in common with the risk assessment models that the strategies were designed to investigate. The utility of the LHS uncertainty and the LHS/PRC (Latin Hypercube Sampling/Partial Rank Correlation) sensitivity analysis techniques are illustrated by analyzing a complex deterministic model of HIV transmission.

Latin Hypercube Sampling

1. Define the probability distribution functions for the K parameters and the state variable initial values

2. Calculate the number of simulations

3. Divide the range of each of the K parameters into N equi-probable intervals

4. Create the LHS table

5. Sample the values of the input parameters and perform the N simulations

6. Analysis of the outcomes: uncertainty analysis

7. Analysis of the outcomes: sensitivity analysis

LHS – HIV model

Table 1

Parameter definitions for the HIV model. All of the transmission efficiencies are conditional on the fact that the partner or needle is infected.

β_{db}	HIV transmission efficiency per buddy partnership
β_{dn}	HIV transmission efficiency per needle injection
β_{fm}	heterosexual transmission efficiency per partnership (female to male)
β_{mf}	heterosexual transmission efficiency per partnership (male to female)
$c_{fb}(t)$	rate of change of sex partners per year (female buddy-users) at time t
$c_{fs}(t)$	rate of change of sex partners per year (female stranger-users) at time t
$c_{fn}(t)$	rate of change of sex partners per year (female non-IVDUs) at time t
$c_{mb}(t)$	rate of change of sex partners per year (male buddy-users) at time t
$c_{ms}(t)$	rate of change of sex partners per year (male stranger-users) at time t
$c_{mn}(t)$	rate of change of sex partners per year (male non-IVDUs) at time t
i_f	rate of sharing needles per year (for female stranger-users)
i_m	rate of sharing needles per year (for male stranger-users)
j_f	rate of change of buddy partners per year (for female buddy-users)
j_m	rate of change of buddy partners per year (for male buddy-users)
q_1	vertical transmission efficiency (seropositive mother, without AIDS)
q_2	vertical transmission efficiency (AIDS mother)
s_a	average adult survival time (years)
s_b	average pediatric survival time (years)
v_a	average adult incubation time (years)
v_b	average pediatric incubation time (years)

Blower & Dowlatabadi (1994)

1. Define the PDFs for the parameters

Table 2

Parameter distribution functions

Parameter	Min	Max	Median	Standard deviation	Function shape
β_{db}	β_{dn}	1	0·56	0·23	triangular (peak at β_{dn})
β_{dn}	0	1	0·28	0·23	triangular (peak at 0·0)
β_{fm}	0	0·5	0·25	0·15	uniform
β_{mf}	0	0·5	0·25	0·15	uniform
$c_{fb}(t)$	1	11	1	1·74	left skewed
$c_{fn}(t)$	1	20	2·19	2·46	left skewed
$c_{fs}(t)$	1	100	2	20.99	left skewed
$c_{mb}(t)$	1	20	1	3·02	left skewed
$c_{mn}(t)$	1	38	2	4·98	left skewed
$c_{ms}(t)$	1	15	1	2·94	left skewed
i_f	13	5,265	299	1,201	left skewed
i_m	13	3,120	228	738	left skewed
j_f	0	4	1·8	0·77	triangular (peak at 1·0)
j_m	0	4	1·8	0·76	triangular (peak at 1·0)
q_1	0	1	0·28	0·23	triangular (peak at 0·0)
q_2	q_1	1	0·56	0·23	triangular (peak at q_1)
s_a	1·0	5·0	1·0	0·85	left skewed
s_b	0·21	4·8	1·04	1·09	left skewed
v_a	1·36	20	8	3·71	Weibull
v_b	0·1	20	0·33 & 5·5	4·99	mixture of two Weibulls

Blower & Dowlatabadi (1994)

2. Calculate the number of simulations

- Necessary $N > (4/3)K$

- When N is large, better ability to detect effects in statistical analysis (PRCC)

- Blower & Dowlatabadi (1994): $K = 20 + 7$, $N = 100$

Blower & Dowlatabadi (1994)

3. Divide each parameter range into N equi-probable regions

Parameter A

1 Expected Value N

Equi-probable slices of A and B

Parameter B

1 Expected Value N

First Sample Second Sample Third Sample

Figure 2. *Creating and sampling the equi-probable intervals. In the Latin Hypercube Sampling design, each parameter is defined in terms of a probability density function (pdf). These pdfs are sliced into N equi-probable intervals—where N is the number of simulations. For each simulation a value for each parameter is selected from one of these intervals at random, and without replacement.*

Blower & Dowlatabadi (1994)

4. Create the LHS table

- N x K table, random sampling of PDF indices, i, without replacement

	β_{db}	β_{dn}	β_{fm}	...	male-strangers t=0
1	i=73	i=22	i=10	...	i=12
2	i=22	i=4	i=22	...	i=17
3	i=1	i=77	i=65	...	i=91
...	...	...	...	...	...
N	determined	determined	determined	determined	determined

5. Run the N simulations

```{r}
## calculate mean of all inputs
ppt.mean <- matrix(apply(ppt_ensemble,2,mean),1,NT) ## driver
## parameters
params <- as.matrix(out$params)
param.mean <- apply(params,2,mean)
## initial conditions
IC <- as.matrix(out$predict)

N.det <- forecastN(IC=mean(IC[,"N[6,30]"]),
                   ppt=ppt.mean,
                   r=param.mean["r_global"],
                   Kg=param.mean["K_global"],
                   beta=param.mean["beta"],
                   alpha=param.mean["alpha_site[6]"],
                   Q=0,   ## process error off
                   n=1)


## Plot run
plot.run()
lines(time2,N.det,col="purple",lwd=3)
```

6. Outcomes: Uncertainty analysis

Table 3.

Descriptive statistics from the uncertainty analysis

	Cumulative number of aids cases in 30 years	
	Adult cases	Pediatric cases
Minimum	49,134	246
Maximum	347,420	161,615
Mean	238,571	37,330
Median	257,085	22,663
Variance	$4\cdot7 * 10^9$	$1\cdot2 * 10^9$
5th percentile	116,422	1,780
95th percentile	333,932	108,173

Prediction precision is low

7. Outcomes: sensitivity analysis

Table 4

Partial rank correlation coefficients

Adult cases		Pediatric cases	
Parameter	PRCC	Parameter	PRCC
β_{mf} & β_{fm}	0·84***	q_1	0·77***
v_a	−0·72***	β_{mf} & β_{fm}	0·77***
β_{db}	0·35***	v_a	0·51***
$c_{fn}(0)$	0·29**	X_{4f}	0·36***
$c_{mn}(0)$	0·29**	$c_{mb}(0)$	0·36***
$c_{mb}(0)$	0·25*	s_a	0·35***
$c_{ms}(0)$	0·23*	v_b	−0·30**
i_m	0·22*	$c_{ms}(0)$	0·28**
$c_{fs}(0)$	0·21*	$X_{3f}(0)$	0·20*
β_{dn}	0·20*		

The PRCCs are between the input values of the biological-behavioural transmission parameters and the output values (the cumulative number of adult and pediatric AIDS cases in 30 years). The results are significant at the 0·05 level (*), the 0·01 level (**) or the 0·001 level (***).
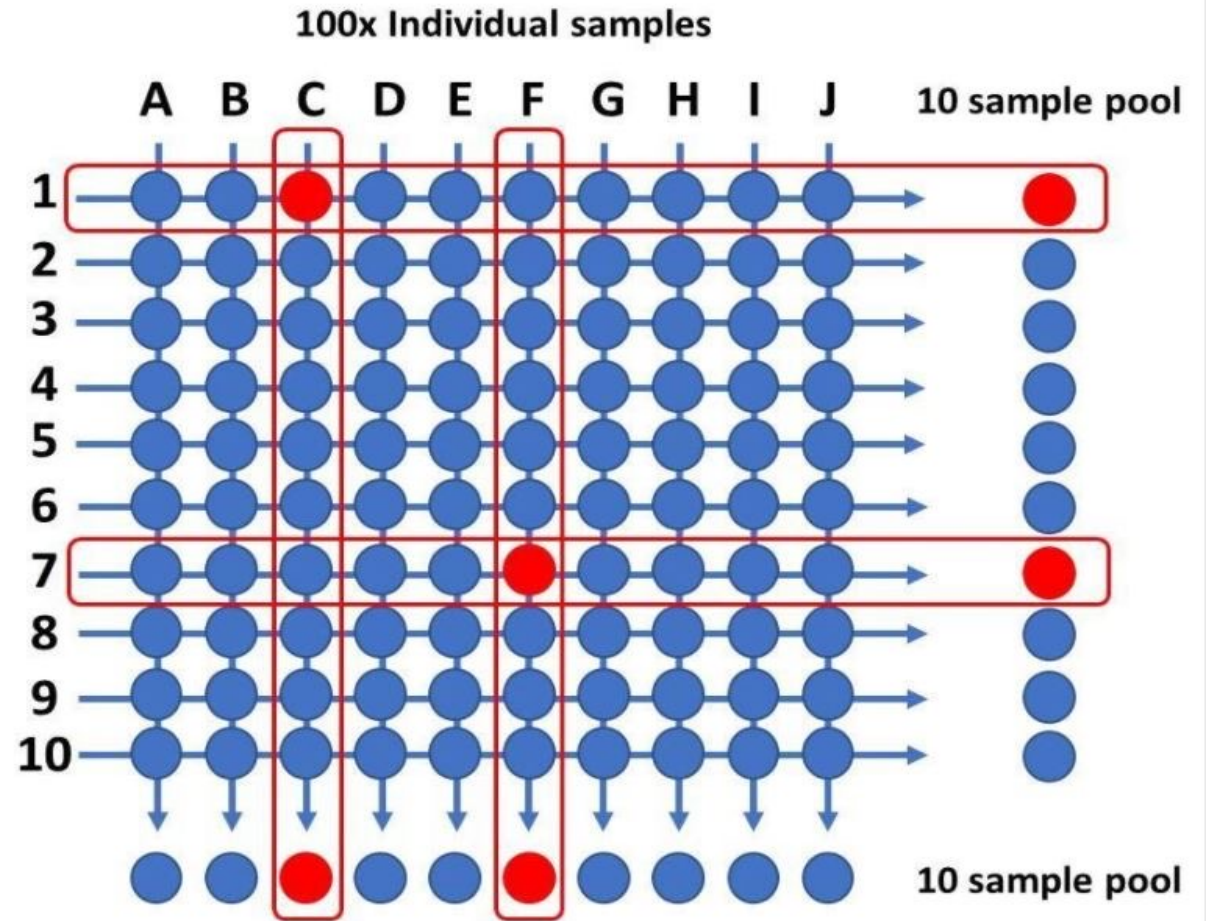
- PRCC determines statistical relationship between each input variable and each outcome variable while keeping all other inputs at their mean value

- Assumes monotonicity between input and outcome – assess with scatter plot

- Sign of the PRCC indicates the qualitative effect

- Magnitude of the PRCC importance of uncertainty of the input quantity on the imprecision of the output variable

HIV model conclusion

- The two heterosexual transmission efficiencies and the average adult incubation period) are the most critical in affecting the prediction precision of the future number of adult AIDS cases.

- The estimation uncertainty of these three parameters are also critical in contributing to the prediction precision of the number of pediatric AIDS cases, however in this case the vertical transmission efficiency is also of great importance (see Table 4).

- Note: HIV model has stable dynamics; monotonicity assumption may not be satisfied for models with chaos, i.e. measles

Global sensitivity – sampling methods

- Group sampling

- Super-saturated design

Pooled testing for SARS-CoV-2 could provide the solution to UK's testing strategy

Key concepts - sensitivity

1. Sensitivity analysis assess how a change in an input translates into a change in output

2. The curse of dimensionality: parameter space increases exponentially with the number of parameters

3. The range of parameters explored should be tied to the probability distribution of the inputs

• Monte Carlo is a brute force approach, there are other more computationally efficient sampling strategies

Ch 11. Propagating analyzing, and reducing uncertainty

Amy Hurford

Memorial University

Uncertainty propagation

- A forecast without a statement of uncertainty is of limited value

- Uncertainty propagation refers to inputting the uncertainty into the model to determine the uncertainty in the output

  o Approach: analytical or numerical

  o Output: full probability distribution or statistical moments

Dietz 2017. Ecological forecasting.

Variable transformation

- Approach: Analytical; Output: full probability density

$$p_Y[y] = p_X[f^{-1}(y)] \left| \frac{df^{-1}(y)}{dy} \right|$$

- Not possible except in simplest of cases

Dietz 2017. Ecological forecasting.

Analytical moments

- Approach: Analytical; Output: Statistical moments

$$y = \beta_0 + \beta_1 x + \epsilon$$

$$\epsilon \sim N(0, \sigma^2)$$

TABLE 11.2. Properties of Means

$E[c] = c$	Mean of a constant.
$E[X + c] = E[X] + c$	Mean plus a constant.
$E[cX] = c \cdot E[X]$	Mean times a constant.
$E[X + Y] = E[X] + E[Y]$	Means are additive.
$E[a + bX + cY] = a + bE[X] + cE[y]$	Application of preceding rules.
$E[E[X|Y]] = E[X]$	Iterated expectation.
$E[g(X)] \neq g(E[X])$	Jensen's inequality, weak form (unless $g(x)$ is linear).
$E[XY] = E[X]E[Y] + Cov[X, Y]$	Multiplication.

Analytical moments

- Approach: Analytical; Output: Statistical moments

$$y = \beta_0 + \beta_1 x + \epsilon$$

$$\epsilon \sim N(0, \sigma^2)$$

TABLE 11.3. Properties of Variances

$Var[c] = 0$	Variance of a constant
$Var[X + c] = Var[X]$	Variance plus a constant
$Var[cX] = c^2 Var[X]$	Variance times a constant
$Var[X + Y] = Var[X] + Var[Y] + 2Cov[X, Y]$	Sum of variances
$Var[a + bX + cY] = b^2 Var[X] + c^2 Var[Y] + 2bc Cov[X, Y]$	Application of preceding rules
$Var[\sum a_i X_i] = \sum\sum a_i a_j Cov[X_i, X_j]$	Generalizes sum of variances
$Var[X] = Var[E[X|Y]] + E[Var[X|Y]]$	Variance decomposition

Dietz 2017. Ecological forecasting.

Analytical moments – Taylor series

- Approach: Analytical; Output: Statistical moments

$$E[f(x)] \approx E\left[f(a) + \frac{f'(a)}{1!} (x - a) + \frac{f''(a)}{2!} (x - a)^2 + \cdots \right]$$

TABLE 11.2. Properties of Means

$E[c] = c$	Mean of a constant.
$E[X + c] = E[X] + c$	Mean plus a constant.
$E[cX] = c \cdot E[X]$	Mean times a constant.
$E[X + Y] = E[X] + E[Y]$	Means are additive.
$E[a + bX + cY] = a + bE[X] + cE[y]$	Application of preceding rules.
$E[E[X|Y]] = E[X]$	Iterated expectation.
$E[g(X)] \neq g(E[X])$	Jensen's inequality, weak form (unless $g(x)$ is linear).
$E[XY] = E[X]E[Y] + Cov[X, Y]$	Multiplication.

Dietz 2017. Ecological forecasting.

Monte Carlo (distribution)

- Approach: Numerical, Output: full probability density

1. Sample random values, x_i, from the (joint) probability distribution of X.
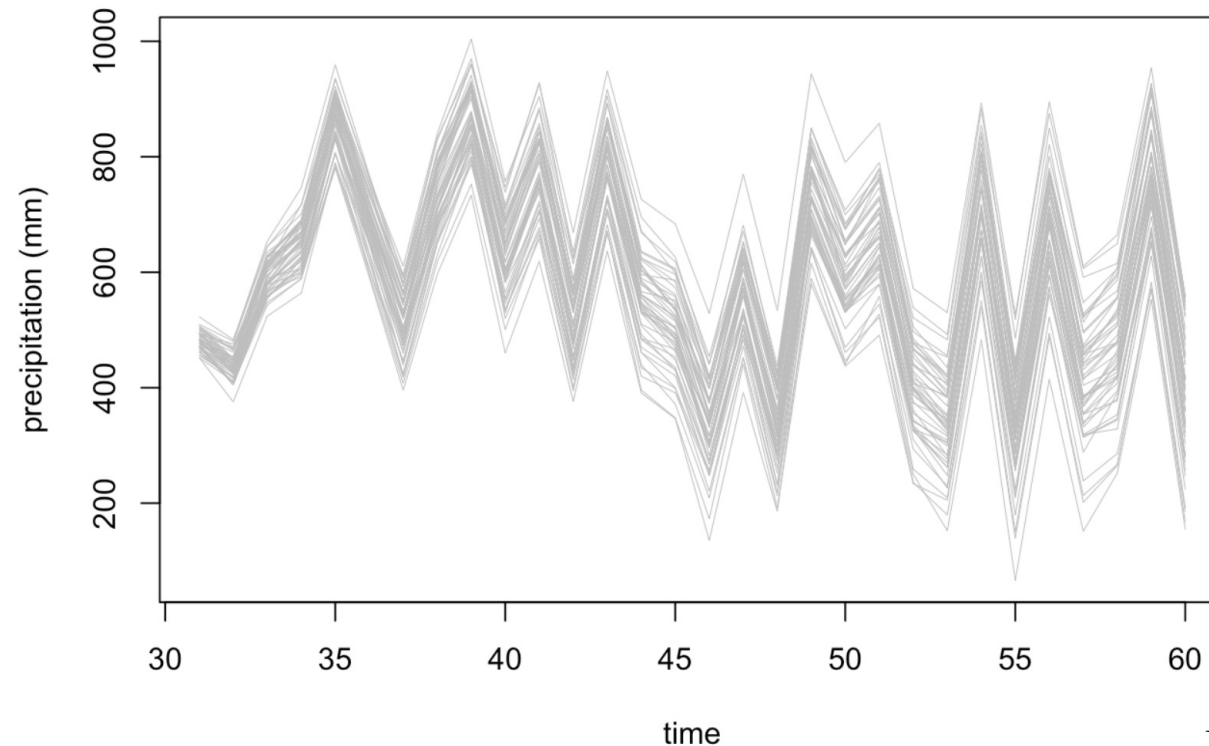2. Calculate $y_i = f(x_i)$ for all samples of X.
3. Use the sample of y's to approximate the PDF of Y.

Dietz 2017. Ecological forecasting.

Monte Carlo (distribution)

- Approach: Numerical, Output: full probability density

Dietz 2017. Ecological forecasting.

Ensemble (Stats)

- Approach: Numerical, Output: full probability density

- Closely related to Monte Carlo, but smaller sample size: 10-100

Dietz 2017. Ecological forecasting.

Uncertainty analysis

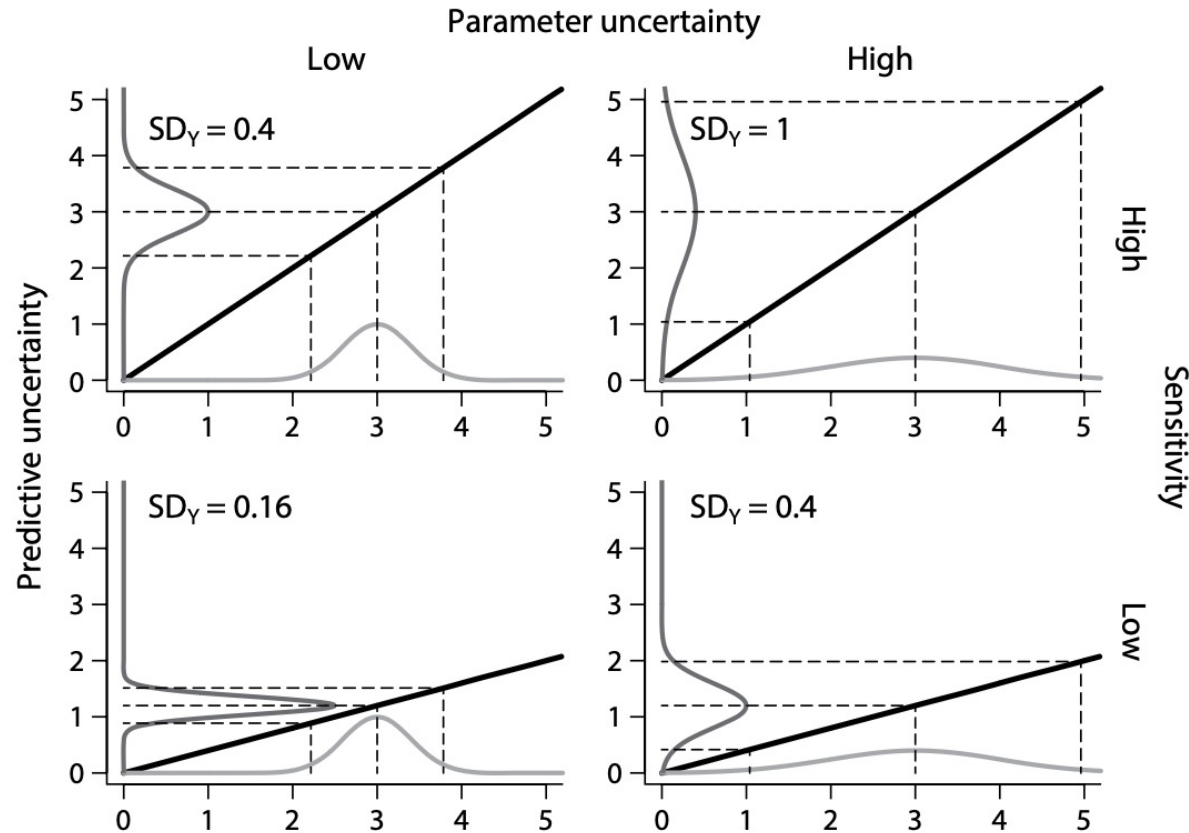- Attribute the uncertainty in some output, to different inputs

FIGURE 11.8. Predictive uncertainty (distribution on the y-axis and reported standard deviation, SD_Y) is controlled by parameter uncertainty (distribution on the x-axis, varies by column) and model sensitivity (slope of diagonal line, varies by row).
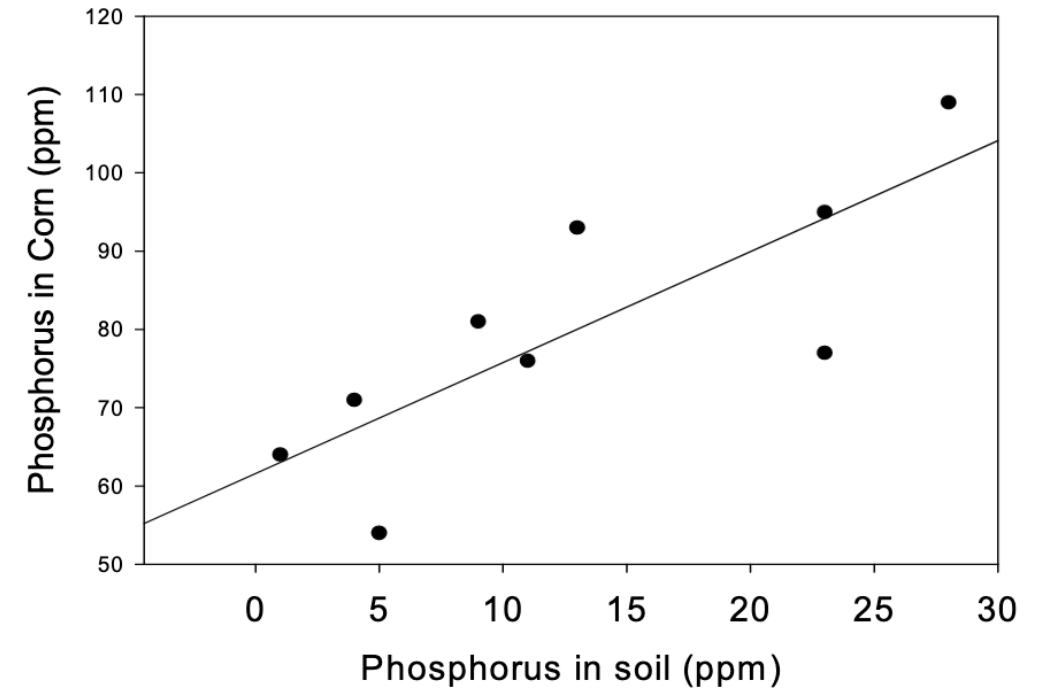
Dietz 2017. Ecological forecasting.

Partitioning uncertainty

- General linear models ($R^2 = 1 - SS/SS_{Total}$)

Source	df	SS	MS	F	---->	p
PSoil	1	1473.6	1473.6	12.89		
Res	7	800.4	114.34			
Total	8	2274.0				

Here are some additional statistics.

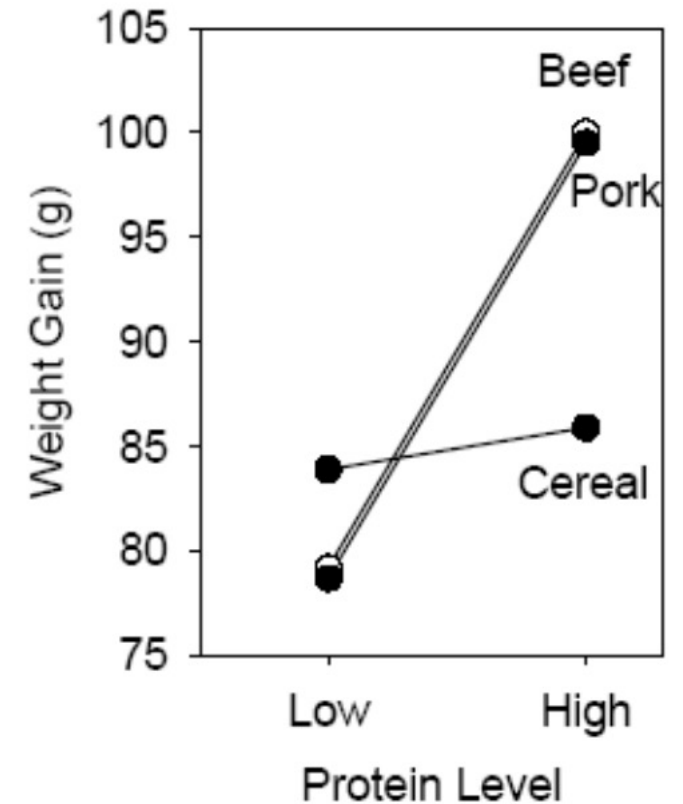r^2 = explained variance = SS_{model} / SS_{tot}

By Dave Schneider

Partitioning uncertainty

- General linear models ($R^2 = 1-SS/SS_{Total}$)

7. **ANOVA table – Frequentist with fixed Type I error.**
   Calculate MS, F, and p

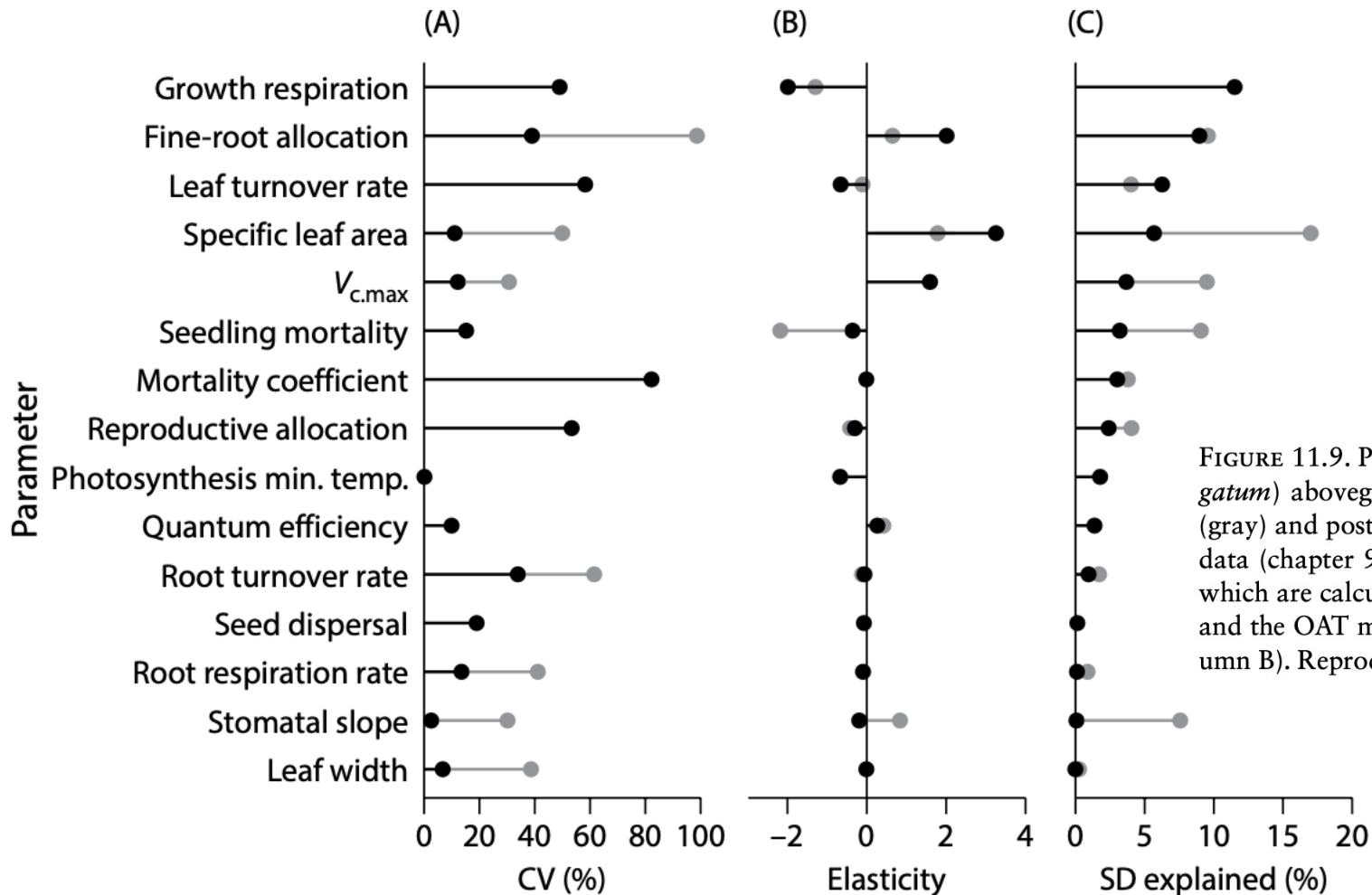| Source	df	SS	MS	F	p	
Source	2	266.5333	133.2667	0.62	0.541132	Can't reject H_o
Level	1	3168.267	3168.267	14.77	0.000322	Reject H_o
S * L	2	1178.133	589.0667	2.75	0.073188	Can't reject H_o
Error	54	11586.00	214.5556			
Total	59					

By Dave Schneider

Partitioning uncertainty

FIGURE 11.9. Partitioning of the uncertainty in predictions of switchgrass (*Panicum virgatum*) aboveground biomass by the Ecosystem Demography model under both prior (gray) and posterior (black) parameter estimates generated from a meta-analysis of trait data (chapter 9). Parameters are ranked by their component uncertainties (column C), which are calculated from the parameter uncertainties (visualized as the CV, column A), and the OAT model sensitivity to that parameter (visualized as parameter elasticity, column B). Reproduced from LeBauer et al. (2013).

Dietz 2017. Ecological forecasting.

Tools for model-data feedback

- How much effort should be invested in estimating particular model inputs, given limited resources?

- Standard error in a parameter estimate decreases with sample size

$$SE(n) = \frac{\sigma}{\sqrt{n}}$$

- A general approach to power analysis is Monte Carlo simulation of different sampling schemes or sizes

Dietz 2017. Ecological forecasting.

Power analysis

$$SE(n) = \frac{\sigma}{\sqrt{n}}$$

Observational design
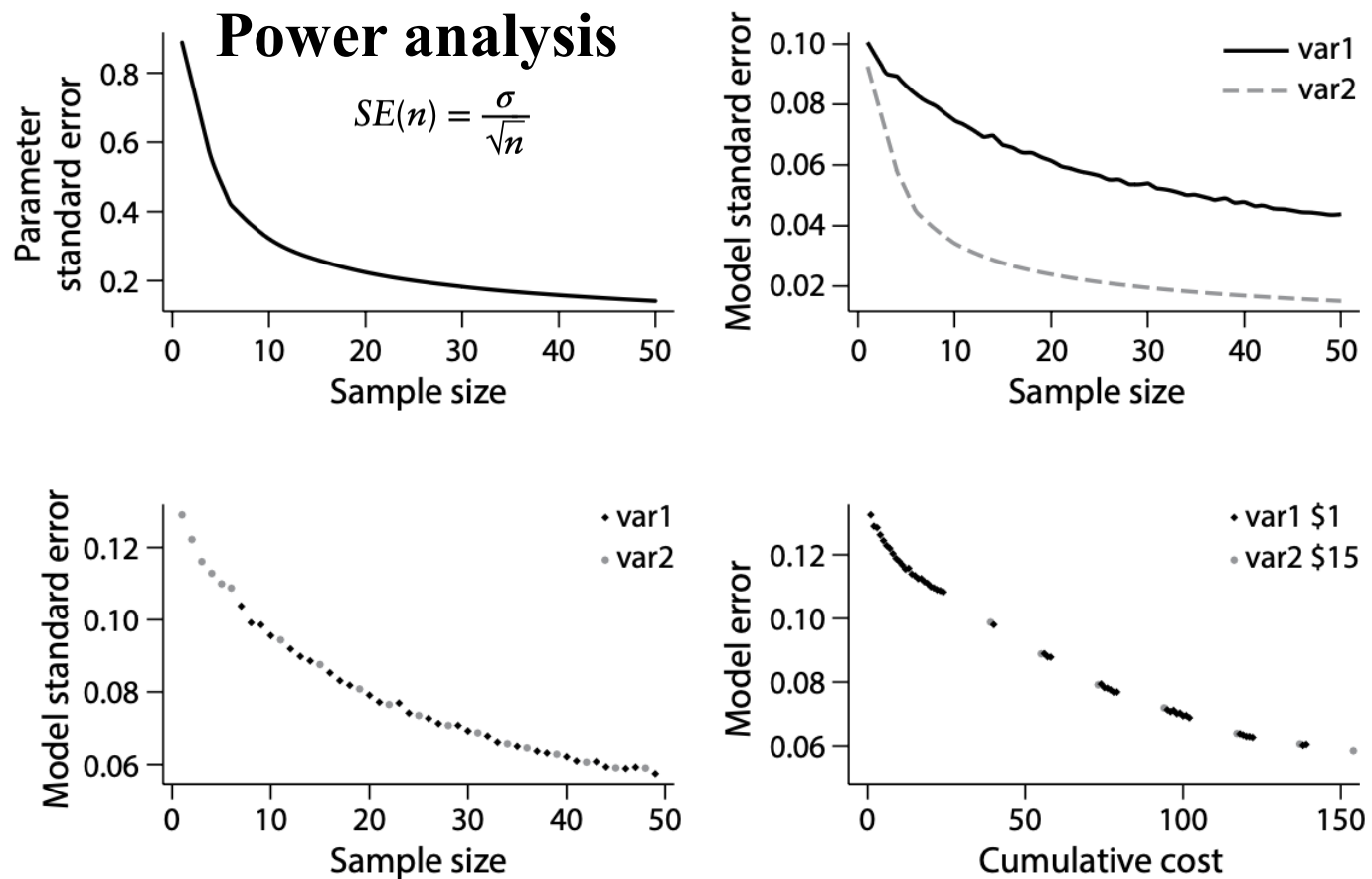
Dietz 2017. Ecological forecasting.

FIGURE 11.10. Observational design. Top left: Power analysis estimating reduction in parameter uncertainty as a function of sample size. Top right: Translation of parameter uncertainty into model output uncertainty for two different variables. Bottom left: Combined output uncertainty as a function of sample size with the optimal design indicated by color and shading. For example, at a sample size of 8 the optimal design is six samples of variable 2 and two of variable 1. Bottom right: Output uncertainty as a function of overall project cost assuming different marginal costs for the two data types. Compared to the equal-cost design (bottom left), there is considerably greater sampling of variable 1.

Key concepts

4. Uncertainty propagation is the process of translating uncertainty in our model inputs into uncertainty in model outputs

7. The uncertainty in a deterministic forecast with known parameters and boundary conditions is dependent upon three components: the uncertainty about the state, $Var[x_t]$, the sensitivity of the system, $f'(\bar{x}_t)^2$, and the process error, q. In ecological systems with stabilizing feedbacks, $f'(\bar{x}_t) < 0$, process error dominates forecasts.

9. A parameter can be important either because it is sensitive, or because it's unknown.

Dietz 2017. Ecological forecasting.

Key concepts

9. A parameter can be important either because it is sensitive, or because it's unknown.

10. More precise answers require more computation, while approximations require stronger assumptions.