There is more to ensemble models than model weights

Michael A. Spence

30th November 2022

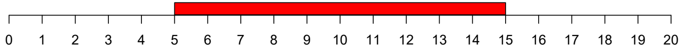
Think of a number

There is an unknown integer between 0 and 100. What is it?

Think of a number

There is an unknown integer between 0 and 100. What is it?

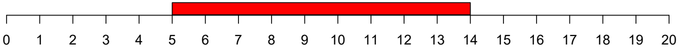
- Kelly told me the number was between 5 and 15



Think of a number

There is an unknown integer between 0 and 100. What is it?

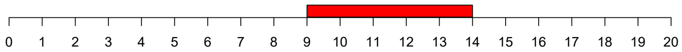
- Kelly told me the number was between 5 and 15
- Jim told me the number was between 3 and 14



Think of a number

There is an unknown integer between 0 and 100. What is it?

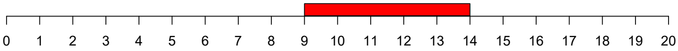
- Kelly told me the number was between 5 and 15
- Jim told me the number was between 3 and 14
- Angela told me the number was between 9 and 17



Think of a number

There is an unknown integer between 0 and 100. What is it?

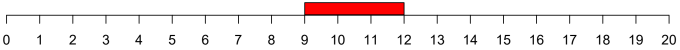
- Kelly told me the number was between 5 and 15
- Jim told me the number was between 3 and 14
- Angela told me the number was between 9 and 17
- Kevin told me the number was between 8 and 14



Think of a number

There is an unknown integer between 0 and 100. What is it?

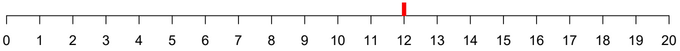
- Kelly told me the number was between 5 and 15
- Jim told me the number was between 3 and 14
- Angela told me the number was between 9 and 17
- Kevin told me the number was between 8 and 14
- Dwight told me the number was between 3 and 12



Think of a number

There is an unknown integer between 0 and 100. What is it?

- Kelly told me the number was between 5 and 15
- Jim told me the number was between 3 and 14
- Angela told me the number was between 9 and 17
- Kevin told me the number was between 8 and 14
- Dwight told me the number was between 3 and 12
- Pam told me the number was between 12 and 18



Think of a number example

We are interested in the some variable y that is informed by m models with x_i being the ith model.

Think of a number example

We are interested in the some variable y that is informed by m models with x_i being the ith model. Initially we know y is between 0 and 100, i.e,

$$p(y) \sim \text{unif}\{0, 100\}$$

Think of a number example

We are interested in the some variable y that is informed by m models with x_i being the ith model. Initially we know y is between 0 and 100, i.e,

$$p(y) \sim \text{unif}\{0, 100\}$$

and after Kelly's information

$$p(y|x_{\text{Kelly}}) \sim \text{unif}\{5, 15\}$$

Think of a number example

We are interested in the some variable y that is informed by m models with x_i being the ith model. Initially we know y is between 0 and 100, i.e,

$$p(y) \sim \text{unif}\{0, 100\}$$

and after Kelly's information

$$p(y|x_{\text{Kelly}}) \sim \text{unif}\{5, 15\}$$

and after Jim's information

$$p(y|x_{\text{Kelly}}, x_{\text{Jim}}) \sim \text{unif}\{5, 14\}$$

and so on...

Average (possibly weighted)

The estimate of the truth $\tilde{y}$ is

$$\tilde{y} = \sum_{i=1}^{m} w_i x_i,$$

where $\sum_{i=1}^{m} w_i = 1$.

Average (possibly weighted) – example

Unweighted example: $w_i = \frac{1}{6}$ for all i,

Average (possibly weighted) – example

Unweighted example: $w_i = \frac{1}{6}$ for all i, Taking each persons expectation

$$\frac{10 + 8.5 + 13 + 11 + 7.5 + 15}{6} = 10\frac{5}{6}$$

Average (possibly weighted) – example

Unweighted example: $w_i = \frac{1}{6}$ for all i, Taking each persons expectation

$$\frac{10 + 8.5 + 13 + 11 + 7.5 + 15}{6} = 10\frac{5}{6}$$

Weighted example: weighted by the precision of the estimate, $w_i \propto \frac{12}{r_i^2 - 1}$ for all i,

Average (possibly weighted) – example

Unweighted example: $w_i = \frac{1}{6}$ for all i, Taking each persons expectation

$$\frac{10 + 8.5 + 13 + 11 + 7.5 + 15}{6} = 10\frac{5}{6}$$

Weighted example: weighted by the precision of the estimate, $w_i \propto \frac{12}{r_i^2 - 1}$ for all i, Taking each persons expectation and r_i to be the range then the estimate is 11.68636.

Average (possibly weighted)

The estimate of the truth $\tilde{y}$ is

$$\tilde{y} = \sum_{i=1}^{m} w_i x_i,$$

where $\sum_{i=1}^{m} w_i = 1$.

Average (possibly weighted)

The estimate of the truth $\tilde{y}$ is

$$\tilde{y} = \sum_{i=1}^{m} w_i x_i,$$

where $\sum_{i=1}^{m} w_i = 1$. If

$$E(x_i) = y$$

then

$$\lim_{m \to \infty} \tilde{y} = y.$$

Average (possibly weighted)

Why would this be true?

Average (possibly weighted)

Why would this be true? Model can be

- fitted to the same data

Average (possibly weighted)

Why would this be true? Model can be

- fitted to the same data
- built with same knowledge

Average (possibly weighted)

Why would this be true? Model can be

- fitted to the same data
- built with same knowledge
- have similar processes

Average (possibly weighted)

Why would this be true? Model can be

- fitted to the same data
- built with same knowledge
- have similar processes
- fitted by similar or sometimes the same people

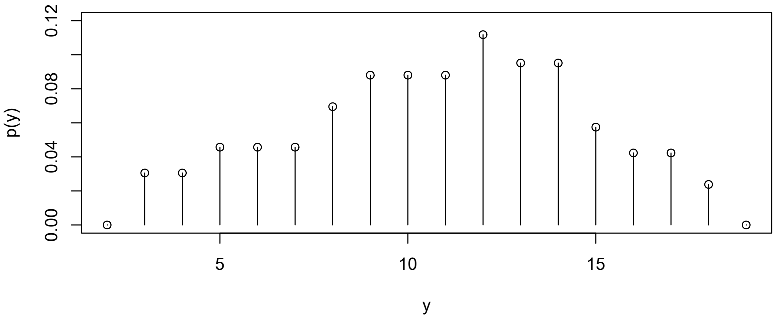
Weighted model

One of the models is right (e.g. Bayesian model averaging), implies

$$p(y) = \sum_{i=1}^{m} w_i p(y|x_i),$$

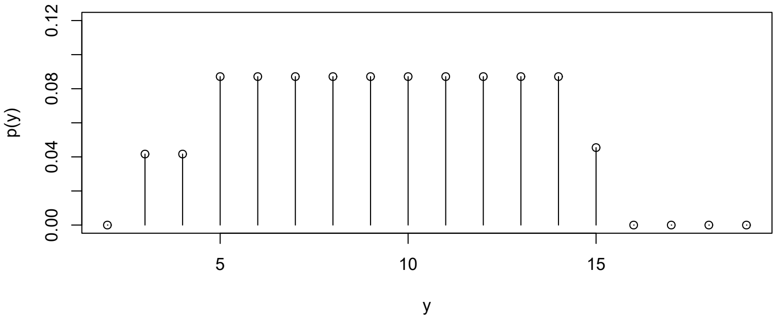
where $\sum_{i=1}^{m} w_i = 1$.

Weighted model – example equally weighted

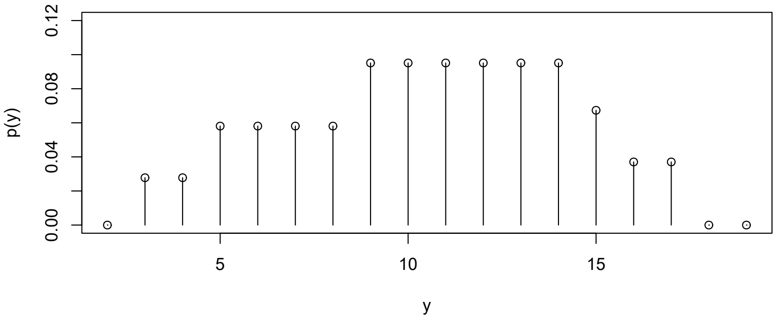


Weighted model – example equally weighted – before Angela

Weighted model – example equally weighted – before Angela



Weighted model – example equally weighted – after Angela



Weighted models – adding an $m + 1$ model

With m models we asked,

One of these m models is the truth, which one?

Weighted models – adding an $m+1$ model

With m models we asked,

One of these m models is the truth, which one?

and then when we have $m+1$ models we ask,

One of these $m+1$ models is the truth, which one?

Adding to our knowledge

Ensemble modelling is using all of the information at once.

Adding to our knowledge

Ensemble modelling is using all of the information at once. That is

$$p(y|x_1,\ldots,x_m) = \frac{p(y)p(x_1,\ldots,x_m|y)}{p(x_1,\ldots,x_m)}$$

Adding to our knowledge

Ensemble modelling is using all of the information at once. That is

$$
\begin{aligned}
p(y|x_1,\ldots,x_m) &= \frac{p(y)p(x_1,\ldots,x_m|y)}{p(x_1,\ldots,x_m)} \\
&= \frac{p(y)p(x_1|y)p(x_2|y,x_1)\ldots p(x_m|y,x_1,\ldots,x_{m-1})}{p(x_1)p(x_2|x_1)\ldots p(x_m|x_1,\ldots,x_{m-1})}.
\end{aligned}
$$

Adding to our knowledge

Ensemble modelling is using all of the information at once. That is

$$\begin{aligned} p(y|x_1,\ldots,x_m) &= \frac{p(y)p(x_1,\ldots,x_m|y)}{p(x_1,\ldots,x_m)} \\ &= \frac{p(y)p(x_1|y)p(x_2|y,x_1)\ldots p(x_m|y,x_1,\ldots,x_{m-1})}{p(x_1)p(x_2|x_1)\ldots p(x_m|x_1,\ldots,x_{m-1})}. \end{aligned}$$

What if we added an $m+1$ model?

Adding to our knowledge

Ensemble modelling is using all of the information at once. That is

$$\begin{aligned} p(y|x_1,\ldots,x_m) &= \frac{p(y)p(x_1,\ldots,x_m|y)}{p(x_1,\ldots,x_m)} \\ &= \frac{p(y)p(x_1|y)p(x_2|y,x_1)\ldots p(x_m|y,x_1,\ldots,x_{m-1})}{p(x_1)p(x_2|x_1)\ldots p(x_m|x_1,\ldots,x_{m-1})}. \end{aligned}$$

What if we added an $m+1$ model? Then,

$$p(y|x_1,\ldots,x_{m+1}) = \frac{p(y|x_1,\ldots,x_m)p(x_{m+1}|y,x_1,\ldots,x_m)}{p(x_{m+1}|x_1,\ldots,x_m)}.$$

Adding to our knowledge

Ensemble modelling is using all of the information at once. That is

$$p(y|x_1,\dots,x_m) = \frac{p(y)p(x_1,\dots,x_m|y)}{p(x_1,\dots,x_m)} = \frac{p(y)p(x_1|y)p(x_2|y,x_1)\dots p(x_m|y,x_1,\dots,x_{m-1})}{p(x_1)p(x_2|x_1)\dots p(x_m|x_1,\dots,x_{m-1})}.$$

What if we added an $m+1$ model? Then,

$$p(y|x_1,\dots,x_{m+1}) = \frac{p(y|x_1,\dots,x_m)p(x_{m+1}|y,x_1,\dots,x_m)}{p(x_{m+1}|x_1,\dots,x_m)}.$$

Hence we need to know

$$p(x_{m+1}|y,x_1,\dots,x_m).$$

What does a model say?

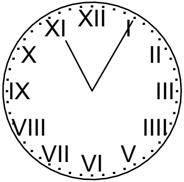
What does a model say?

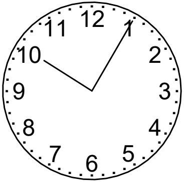
We know that

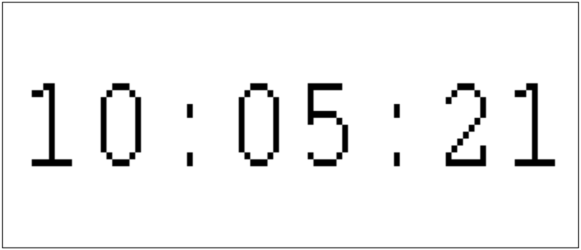
$$x_i = y + \delta_i,$$

where δ_i is the discrepancy.

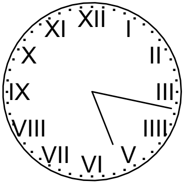
Broken clocks

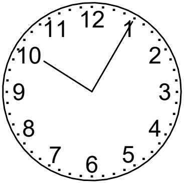


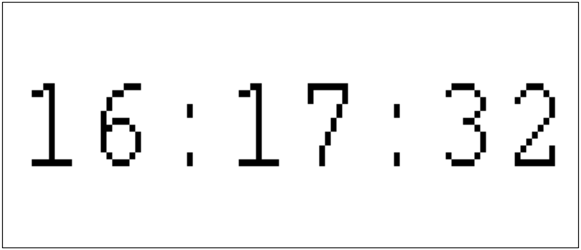




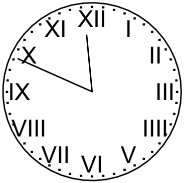
Broken clocks

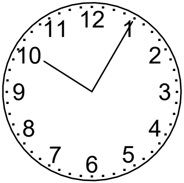


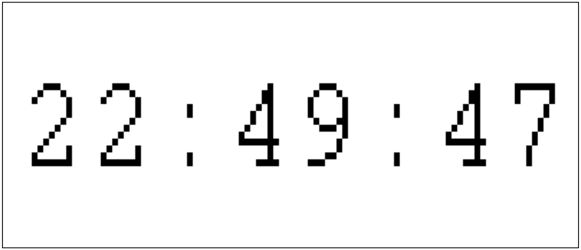




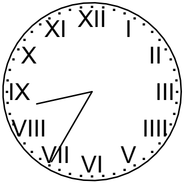
Broken clocks

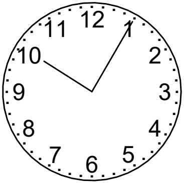




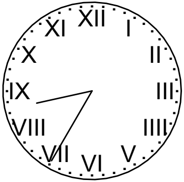


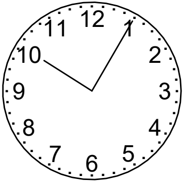
Broken clocks

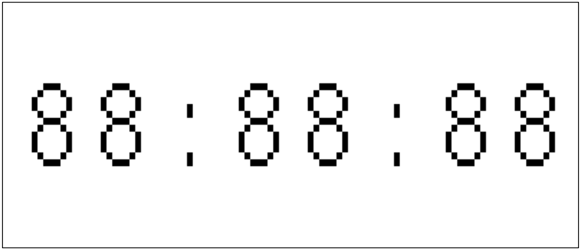




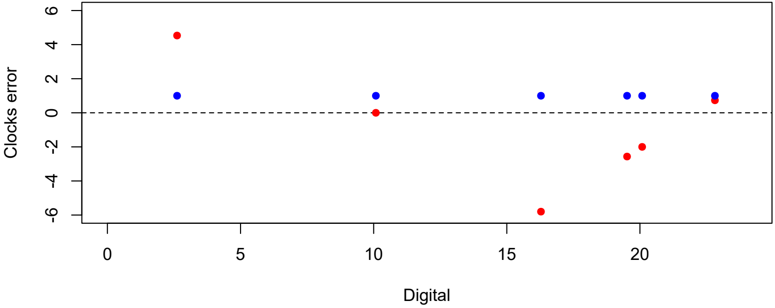
Broken clocks



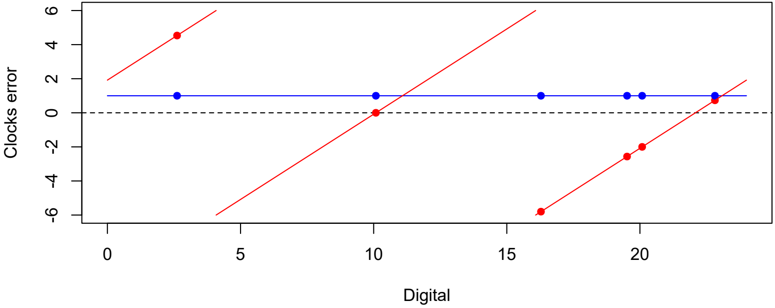




What time is it?



What time is it?



What does a model say?

We know that

$$x_i = y + \delta_i,$$

where δ_i is the discrepancy.

What does a model say?

We know that

$$x_i = y + \delta_i,$$

where δ_i is the discrepancy. The Roman numeral clock has

$$E(\delta_1) = 1 \text{ hour}$$

and

$$\text{var}(\delta_1) = 0.$$

What does a model say?

We know that

$$x_i = y + \delta_i,$$

where δ_i is the discrepancy. The Roman numeral clock has

$$E(\delta_1) = 1 \text{ hour}$$

and

$$\text{var}(\delta_1) = 0.$$

The decimal clock has

$$E(\delta_2) = 0$$

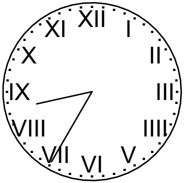
and

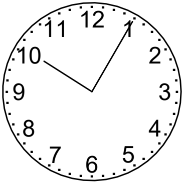
$$\text{var}(\delta_2) = 12,$$

with

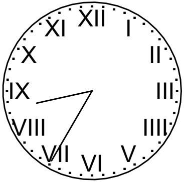
$$\delta_2 \sim \text{U}_{[-6,6]}.$$

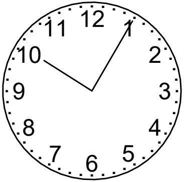
Broken clocks

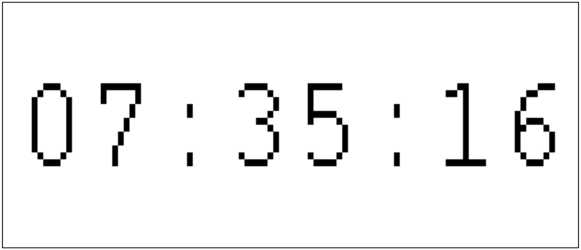




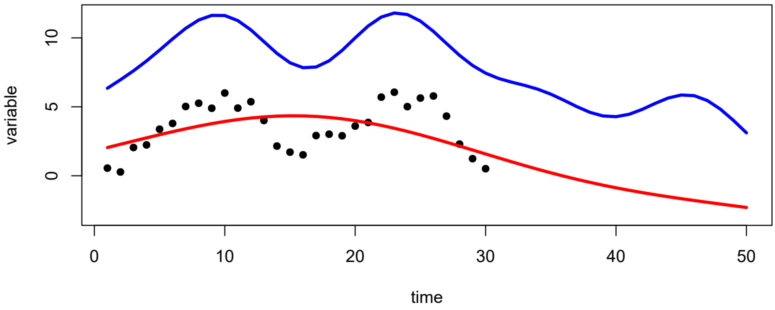
Broken clocks







Discrepancy example

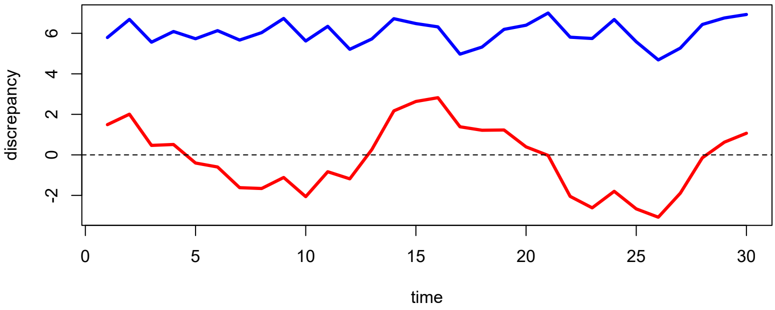


Modelling discrepancy

We know

$$x_{i,t} = y_t + \delta_{i,t}.$$

Modelling discrepancy



Modelling discrepancy

We know

$$x_{i,t} = y_t + \delta_{i,t}.$$

Modelling discrepancy

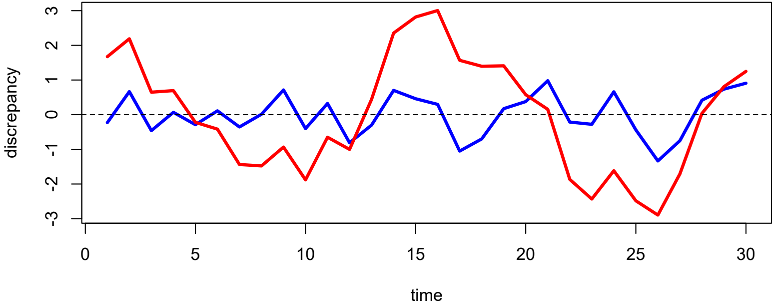
We know

$$x_{i,t} = y_t + \delta_{i,t}.$$

Lets say

$$x_{i,t} = y_t + \nu_i + \zeta_{i,t}$$

Modelling discrepancy



Modelling discrepancy

We know

$$x_{i,t} = y_t + \delta_{i,t}.$$

Lets say

$$x_{i,t} = y_t + \nu_i + \zeta_{i,t}$$

Modelling discrepancy

We know

$$x_{i,t} = y_t + \delta_{i,t}.$$

Lets say

$$x_{i,t} = y_t + \nu_i + \zeta_{i,t}$$

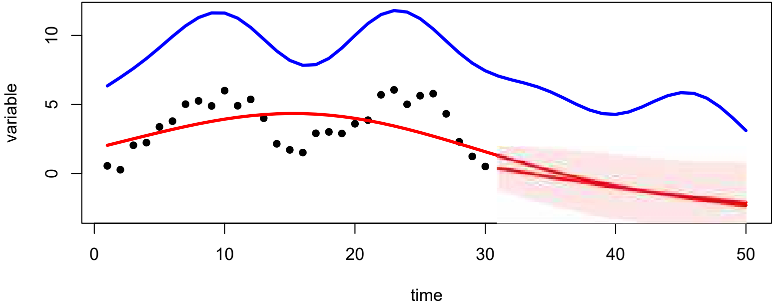
with

$$\zeta_{i,t} = \rho_i \zeta_{i,t-1} + \epsilon_{i,t}$$

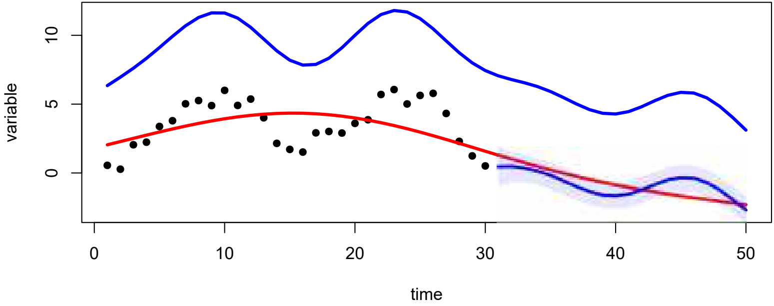
and

$$\epsilon_{i,t} \sim N(0, \sigma_i^2).$$

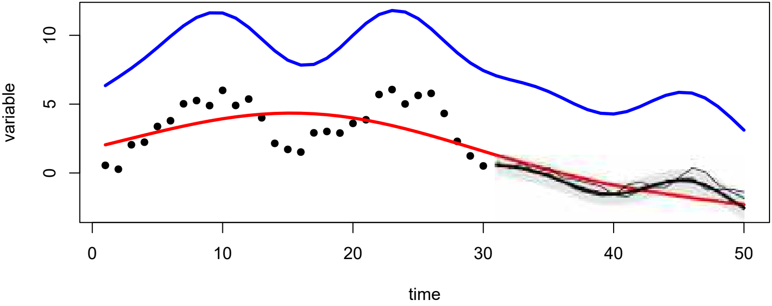
Modelling discrepancy



Modelling discrepancy



Modelling discrepancy



Modelling discrepancy

Look again at the equation

$$x_i = y + \delta_i$$

Modelling discrepancy

Look again at the equation

$$x_i = y + \delta_i + \epsilon_i,$$

where ϵ_i is model specific error (e.g. parameter uncertainty).

Modelling discrepancy

Look again at the equation

$$x_i = y + \delta_i + \epsilon_i,$$

where ϵ_i is model specific error (e.g. parameter uncertainty). We could say

$$x_i = y + \nu + \eta_i + \epsilon_i$$

with $E(\eta_i) = 0$ and $\text{var}(\eta_i) = \sigma_\eta^2$.

Modelling discrepancy

Look again at the equation

$$x_i = y + \delta_i + \epsilon_i,$$

where ϵ_i is model specific error (e.g. parameter uncertainty). We could say

$$x_i = y + \nu + \eta_i + \epsilon_i$$

with $E(\eta_i) = 0$ and $\text{var}(\eta_i) = \sigma_\eta^2$.

Ensemble modelling can be mixed effects modelling

Summary

Summary

- Averages and weighting models comes with some strong assumptions

Summary

- Averages and weighting models comes with some strong assumptions
- These incorrect assuptions can lead to increase uncertainty with more knowledge

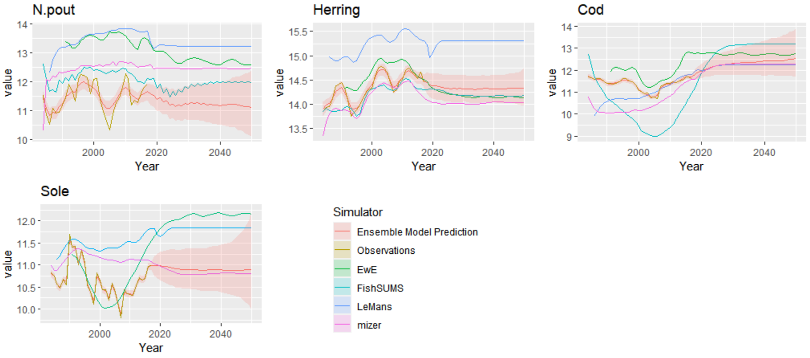
Summary

- Averages and weighting models comes with some strong assumptions
- These incorrect assuptions can lead to increase uncertainty with more knowledge
- Examining discrepancies can utilise information

Summary

- Averages and weighting models comes with some strong assumptions
- These incorrect assuptions can lead to increase uncertainty with more knowledge
- Examining discrepancies can utilise information
- Ensemble modelling is a regression problem e.g. random effects model

EcoEnsemble



References

- Chandler, R.E. (2013) Exploiting strength, discounting weakness: combining information from multiple climate simulators. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, **371**, 20120388.
- Spence et al. (2018), A general framework for combining ecosystem models. *Fish and Fisheries*, 19(6):1031–1042.
- Spence et al. (2021) Sustainable fishing can lead to improvements in marine ecosystem status: an ensemble-model forecast of the North Sea ecosystem. *Mar Ecol Prog Ser* 680:207-221.
- Spence M. A, Martindale J. A & Thomson M. J (2022) EcoEnsemble: A General Framework for Combining Ecosystem Models. *https://CRAN.R-project.org/package=EcoEnsemble*

Subjective beliefs - example prePam

The beliefs after Dwight were

$$p(y|x_{\text{Kelly}}, \ldots, x_{\text{Dwight}}) = \begin{cases} \frac{1}{4} & \text{if } y = 9, 10, 11 \text{ or } 12 \\ 0 & \text{otherwise.} \end{cases}$$

with Pam being

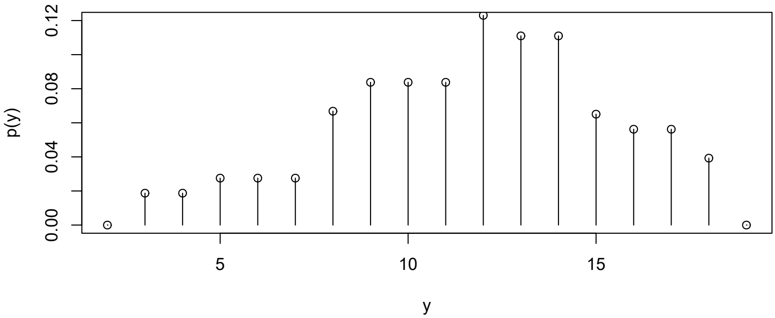
$$\begin{aligned} p(y|x_{\text{Pam}}) &= \frac{p(y)p(x_{\text{Pam}}|y)}{\sum_{z=0}^{100} p(z)p(x_{\text{Pam}}|z)} \\ &= \begin{cases} \frac{1}{7} & \text{if } y = 12, 13, 14, 15, 16, 17 \text{ or } 18 \\ 0 & \text{otherwise.} \end{cases} \\ &= p(x_{Pam}|y) \end{aligned}$$

Subjective beliefs - example including Pam

Adding Pam's information, we get

$$
\begin{aligned}
p(y|x_{\text{Kelly}},\ldots,x_{\text{Dwight}},x_{Pam}) &= \frac{p(y|x_{\text{Kelly}},\ldots,x_{\text{Dwight}})p(x_{Pam}|y)}{\sum_{z=0}^{100} p(z|x_{\text{Kelly}},\ldots,x_{\text{Dwight}})p(x_{\text{Pam}}|z)} \\
&= \begin{cases} 1 & \text{if } y = 12 \\ 0 & \text{otherwise.} \end{cases}
\end{aligned}
$$

Weighted model – example weighted



Adding to our knowledge

What if the $m+1$ model was the same as the first, and we knew that?

Adding to our knowledge

What if the $m+1$ model was the same as the first, and we knew that? Then,

$$p(x_{m+1}|y, x_1, \ldots, x_m) = p(x_{m+1}|x_1).$$

Adding to our knowledge

What if the $m+1$ model was the same as the first, and we knew that? Then,

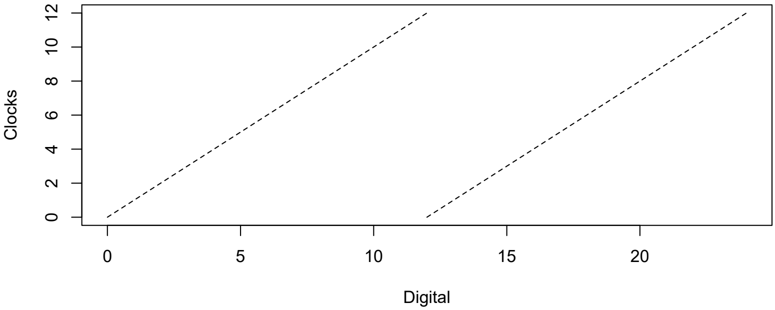
$$p(x_{m+1}|y, x_1, \ldots, x_m) = p(x_{m+1}|x_1).$$

and so

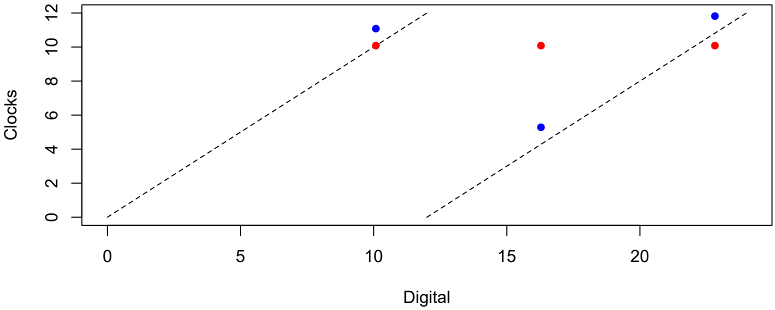
$$p(y|x_1, \ldots, x_{m+1}) = p(y|x_1, \ldots, x_m).$$

We learnt nothing.

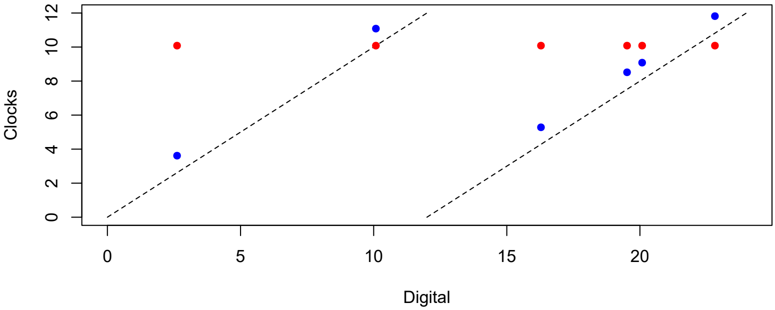
What time is it?



What time is it?



What time is it?



Temporary page!

LaTeX was unable to guess the total number of pages correctly. A
there was some unprocessed data that should have been added t
the final page this extra page has been added to receive it.
If you rerun the document (without altering it) this surplus page
will go away, because LaTeX now knows how many pages to expe
for this document.