

Peer Review for Shannon Gallagher

Amanda Luby

1 General Comments

Overall, I find your writing very easy to read and understand. However, it is easy to spot the passages and sections that you haven't worked on quite as much. The sentences get a lot choppier than the rest of your writing, and some of your thoughts don't seem fully expressed. This seems like it would be a hard section to write, since you have to go back and forth between expressing intuitive ideas and using the model notation in the text, but overall you do a good job of making it readable and understandable. I think you also did a good job of picking out the most helpful figures to understand the points that you're trying to make. A lot of people either have too few or too many figures, but I think you have the right balance in this section.

For an ADA paper, when you're not working under length constraints, I think you could take more time to explain why certain decisions, especially with regards to parameters and distributions in the model. I also mentioned it in the specific comments, but I also think it would be helpful to use an abbreviation for the EB with regional effects model (maybe EBRE). I found it a little hard to distinguish between the two models when you were discussing them.

2 Specific Comments, suggestions, and edits

- Section 2.1, sentence 2,3: These sentences seemed a little more blunt than most of your writing, maybe combine them to make it flow better. For instance,

Finding which modifications to make is difficult in practice, we thus add simplifying assumptions to establish a path for fitting a model:

- Page 7, last sentence: seems very repetitive and similar to a previous sentence, since you've already mentioned relaxing the assumptions, the 'new' information seems to be the model. The wording in the section makes it a little hard to figure out what you've done versus what's already been done, so maybe be more explicit about 'the model'. I would suggest:

By relaxing the assumption of regional independence, we build upon Brooks' model through a regional effects extension.

- Section 2.2, first sentence: I think you can make it more clear that 'regional effects extension of EB' is a NEW model and that it's *your* name for it? Maybe

Relaxation of the assumption that the regions are independent from one another leads to an extension to the EB model, which we call EB with regional effects.

- Section 2.2 sentence 3: Putting the reference to the previous section before the statement about non-independence might make this read a little better.
- Section 2.2, paragraph 2: This is the first time that the terminologies ‘EB’ and ‘EB with regional effects’ started to trip me up. It might be helpful to refer to ‘EB with regional effects’ as ‘EBRE’ or something so that the abbreviations for the two models are more distinct?
- Section 2.2, paragraph 3, sentence 3: I’m not a huge fan of ‘... flu threshold, *meaning* we do not adjust these values’. I’d do something like ‘... flu threshold, and do not adjust those values’.
- Section 2.2, paragraph 3, sentence 4: This sentence reads really choppy and is a little hard to follow. Since you’ve already mentioned that you do not adjust the values under the flu threshold, maybe start with that information and then mention that flu curves are not linear:
Scaling values that are already close to zero would distort the analysis, since flu curves are not linear functions.
- Section 2.2, paragraph 4, sentence 1: should $s = 1, 2, \dots, 10$ be $s = 1, 2, \dots, 15$ instead?
- Section 2.2, paragraph after introducing the priors: I think it would be helpful to talk about the priors in an intentional order? Starting with \hat{F} seemed a little weird since it was in the middle of the list. I think it would help to do it systematically and talk about them in the order that they were introduced. It would also be helpful to specify why the distributions have specific numbers?
- Section 2.2, Figure 4: I found this very helpful and think it’s a great graphic! Even though the notation got a little dense, having the picture there to refer to made things easier to follow.
- Section 2.2, after likelihood equation: T_{s_0} doesn’t appear in the preceding equation, so I’m a little confused as to why it immediately follows the equation?
Also, in the sentence ‘Finally, as our prior consists of uniforms (sp), our posterior $\lambda \propto \mathcal{L}$, I think you should specify what the prior is for, and say ‘uniform distributions’ instead of ‘uniforms’. I’d also write out that ‘our posterior, λ , is proportional to the likelihood.
- Page 10, first full sentence: This is just really messy and I’m a little unsure of what you’re trying to say?
- Section 2.2.1, sentence 3/4: I’d say:
As described in Equation 1, the normalizing constant of the posterior distribution is unknown. Through importance sampling, instead of sampling from λ , ...

to organize the new/old information a little more, and make it clear what importance sampling gives you.

- section 2.3 header: should this be EB with Regional Effects results? Maybe EB and EBRE results?
- section 2.3, sentence 1/2/3: I think you could combine these sentences to make it flow a little better:

We work with 16 seasons of flu data for each of the ten regions, leaving out the 2009 Swine Flu season because we believe it to be an outlier. In addition, we leave out Region 9 from our analyses for the entirety of the paper, because it seems not to follow the same trends as the other regions.
- section 2.3 paragraph 2: I think you should either expand this or take it out altogether - if it's just one sentence of information, it doesn't seem worth it to include, but if you expand on it and talk more about the difference it makes in runtime, I think it could be interesting.
- Section 2.3 paragraph 3: I think you can remove the sentence 'In this figure, we see how the EB with regional effects method works'. Since you introduce the figure in the previous two sentences and explain what it is we see in the following sentences. This sentence seems unnecessarily sandwiched between sentences with more information.
- Page 11, sentence 1: mistyping errors with 'have' and 'nature' :)
- Page 11, sentence 2: I'm a little confused about what exactly 'just does not compare to EB at predicting peak height and week'
- Page 12, first full paragraph: I think you can remove the phrase 'which is required for the likelihood'. I'd also use 'maximum' instead of 'max', and for the second to last sentence, I'd try to make it a little more explicit what 'this' is. for instance,

This is an infeasible number of calculations, and we instead approximate by fixing regional variables as described above.
- Page 12, paragraph 2, last sentence: I'd make it less about 'we' and more about the method -

EB performs consistently better than EB with regional effects until the peak, when EB with regional effects begins to perform slightly better than the original EB model.
- Page 12, paragraph 3, first sentence: Should this be *Figure 6*?
- Page 12, paragraph 3, last 2 sentences: I'm not entirely sure what you're trying to say here. If you're talking about specific measurable quantities, I find it helpful to use single quotations: 'weeks known' and 'season'.