Reviewer 1:

The paper introduces a Bayesian graphical causal inference framework to investigate solar radiation and temperature effects on dissolved gaseous mercury (DGM) concentrations. This is an exciting contribution with clear potential to advance environmental data analysis.

- *We thank the reviewer for this kind comment. We agree that it is our intention with this technical note to advance environmental data analysis in line with other fields that use Causal and Bayesian methods.*

However, major revisions are required to ensure that the method is applied following best practices and clearly communicated to a broader audience in environmental sciences who may not have a statistical background.

- *We agree with the reviewer that we must ensure that our proposed framework follows best practices and is communicated in a way that is applicable, understandable, and useful for a broader audience in environmental science,*

Major Comments

1. Justification for Bayesian Approach

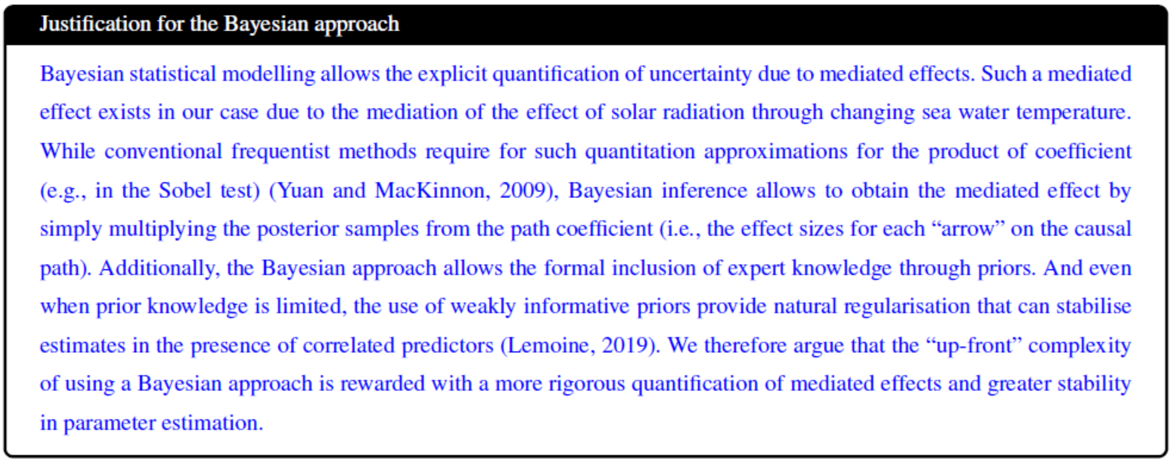
The study does not explicitly demonstrate that frequentist methods fail or that Bayesian inference provides a clear empirical advantage. No comparison is made (e.g., between regression or structural equation models and their Bayesian alternatives) to show instability or bias under a frequentist framework. Since Bayesian methods are technically more complex, the manuscript should clarify when and why they are preferable and under what conditions their use provides meaningful benefits.

- *We thank the reviewer for the opportunity to justify our methodological choice of using a Bayesian approach. We agree that frequentist methods would not necessarily fail. However, we chose a Bayesian Data Analysis approach for three specific advantages that align with our framework:*
    - *Mediation analysis: A core component of our study is estimating indirect effects (Section 5). In a frequentist approach, estimating indirect effects, which involves the product of regression coefficients typically requires strong approximations (Delta method in Sobel test for example), or bootstrapping. The Bayesian approach allows the simple multiplication of posterior samples of the path coefficients to derive the posterior distribution of the indirect effect. We argue that the "up-front" complexity of the Bayesian setup is rewarded with a rigorous and straightforward quantifaction of mediation.*
    - *Formalising expert knowledge: Our framework aims to offer ways to formalise the use of expert knowledge. Bayesian approaches provide a mathematically consistent mechanism to encode physical constraints and domain knowledge for example in the choice of priors.*

- *Regularisation: Even where prior knowledge is limited, the use of weakly informative priors provide regularization. This ensures stability in parameter estimation in situations with strong correlation of predictor variables, such as the relationship between solar radiation and temperature.*

- *We have added a textbox in Section 4.5 summarising these justifications:*

  Section 4.5 Textbox at #386:



Justification for the Bayesian approach

Bayesian statistical modelling allows the explicit quantification of uncertainty due to mediated effects. Such a mediated effect exists in our case due to the mediation of the effect of solar radiation through changing sea water temperature. While conventional frequentist methods require for such quantitation approximations for the product of coefficient (e.g., in the Sobel test) (Yuan and MacKinnon, 2009), Bayesian inference allows to obtain the mediated effect by simply multiplying the posterior samples from the path coefficient (i.e., the effect sizes for each "arrow" on the causal path). Additionally, the Bayesian approach allows the formal inclusion of expert knowledge through priors. And even when prior knowledge is limited, the use of weakly informative priors provide natural regularisation that can stabilise estimates in the presence of correlated predictors (Lemoine, 2019). We therefore argue that the "up-front" complexity of using a Bayesian approach is rewarded with a more rigorous quantification of mediated effects and greater stability in parameter estimation.

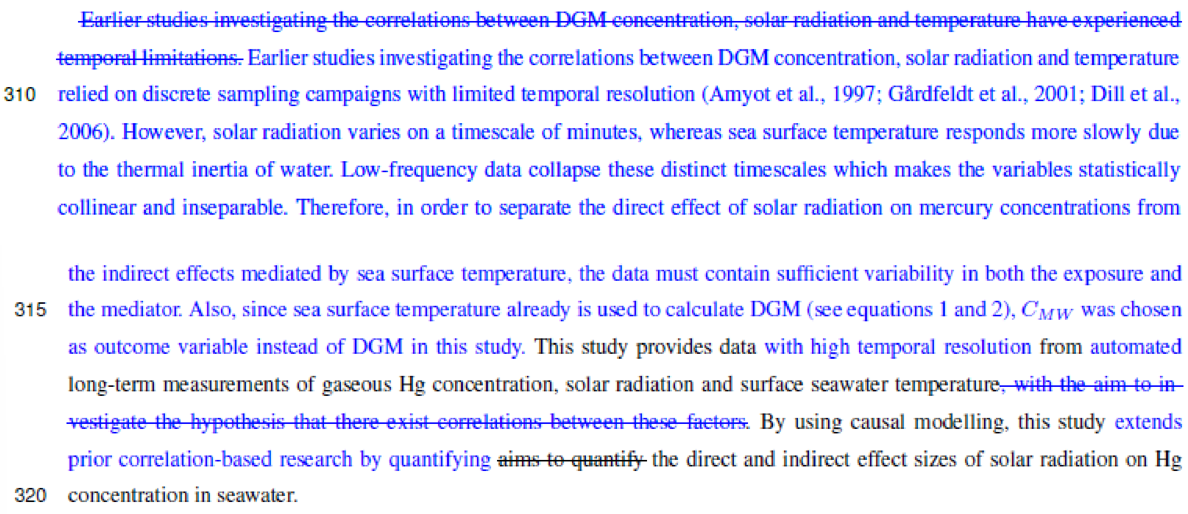
2. Temporal Novelty and Model Structure (#255)

The authors claim that previous studies suffered from temporal limitations. While this study uses high-frequency data, the model itself does not incorporate time as a structural or dynamic dimension—it treats each time step as an independent observation. The manuscript should clearly explain how this approach differs from earlier studies and whether the higher temporal resolution truly enhances inference or simply provides finer data granularity.

- *We thank the reviewer for this observation. We agree that our models treat time steps as independent observations and that they do not explicitly model temporal dynamics for example through autoregressive terms. We have revised the manuscript in Section 4.1 to clarify that with "temporal limitations" of previous studies, which deployed discrete sampling strategies, we referred to their low sampling frequency rather than any limitations in their modelling strategy.*
- *The high temporal resolution of the automated sampling deployed in our study is not only aiming for a finer granularity of the data, but it is a prerequisite for being able to separate direct and indirect effects for two specific reasons:*
  - *Solar radiation varies on a timescale of minutes, wheres sea surface temperature, as also highlighted by the second reviewer, responds more slowly due to the thermal inertia. We need a high time resolution to distinguish the immediate photochemical effects of sun radiation from the indirect and slower thermal effects. Low-frequency data would collapse the distinct timescales which makes the effects inseparable.*

    - *A second, yet more secondary reason is that large sample sizes lead to a more robust converge on posterior distributions, even if the prior assumptions are non-informative or weak.*

- *We reflect these changes in Section 4.1 #309-319:*



~~Earlier studies investigating the correlations between DGM concentration, solar radiation and temperature have experienced temporal limitations.~~ Earlier studies investigating the correlations between DGM concentration, solar radiation and temperature relied on discrete sampling campaigns with limited temporal resolution (Amyot et al., 1997; Gårdfeldt et al., 2001; Dill et al., 2006). However, solar radiation varies on a timescale of minutes, whereas sea surface temperature responds more slowly due to the thermal inertia of water. Low-frequency data collapse these distinct timescales which makes the variables statistically collinear and inseparable. Therefore, in order to separate the direct effect of solar radiation on mercury concentrations from the indirect effects mediated by sea surface temperature, the data must contain sufficient variability in both the exposure and the mediator. Also, since sea surface temperature already is used to calculate DGM (see equations 1 and 2), C_{MW} was chosen as outcome variable instead of DGM in this study. This study provides data with high temporal resolution from automated long-term measurements of gaseous Hg concentration, solar radiation and surface seawater temperature~~, with the aim to investigate the hypothesis that there exist correlations between these factors~~. By using causal modelling, this study extends prior correlation-based research by quantifying ~~aims to quantify~~ the direct and indirect effect sizes of solar radiation on Hg concentration in seawater.

3. Distributional Assumption for C_{MW} (#355)

The assumption of a Normal likelihood for C_{MW}is weakly justified. While the Normal distribution is commonly used, its prevalence does not imply appropriateness; the appeal to the Central Limit Theorem oversimplifies environmental concentration data, which are typically multiplicative and right-skewed -- Figure 11(e) shows a long-tailed distribution. The authors could either demonstrate that residuals are approximately normal (supported by residual–fitted value plots) or acknowledge this limitation and discuss whether a log-normal likelihood would be more appropriate.

- *We thank the reviewer for this observation and suggestion regarding the choice of likelihood for C_{MW} . We agree that environmental data are often multiplicative and that the Normal distribution. To address the alternative of a log-normal likelihood, we have added the Appendix "Discussion on the distributional assumption for C_{MW}. In this appendix, we plotted, as suggested, the residuals of the Normal model against fitted values (Figure G1) and concluded that the plot suggests an increasing variance with the mean of C_{MW}.*

  As a consequence, we implemented a modified model m_4^{log} that uses a log-normal likelihood. We then compared the parameter estimates of the Normal and Log-Normal models by calculating the implied effects on C_{MW} , listed in Table G1. The results of this comparison show that all effect sizes differ by less than 1%, which is why we decided to accept the original Normal likelihood assumption. However, we have updated the main text at Section 4.5 to acknowledge the limitation of the Normal assumption, and we refer to the Appendix for the detailed analysis. Section 4.5 #429-430:

Normal distribution and environmental phenomena typically involve the aggregation of a large number of underlying processes. Furthermore, we had no reason to assume another distribution for the outcome C_{MW}. However, Appendix G discusses the alternative choice of a log-normal likelihood for C_{MW} that can often be appropriate for environmental data. As part of the

- *New Appendix G #932-962:*

Appendix G: Discussion on the distributional assumption for C_{MW}

While assuming that the outcome data is normally distributed can be sensible in many cases, environmental data may show a multiplicative and right-skewed character which may also be indicated in the long-tail distribution of the observed C_{MW} data shown in Figure 11 (e). In order to check if the normal likelihood assumption is appropriate for C_{MW}, we plotted the residuals of model m_4 against its fitted values. The resulting *residual plot* visualises the predication error between each observation and the model's estimate. If the assumption of normal distributed data for C_{MW} holds, the residuals will be symmetrically distributed around zero with a more or less constant spread. However, the plot in Figure G1 suggests that the spread of the residuals is not constant but instead widens as the predicted C_{MW} increases, indicating that the model's error scales with the magnitude of C_{MW}. This pattern would justify the adoption of a Log-Normal likelihood which, unlike the Normal likelihood, models a multiplicative and long-tailed distribution nature of C_{MW}.

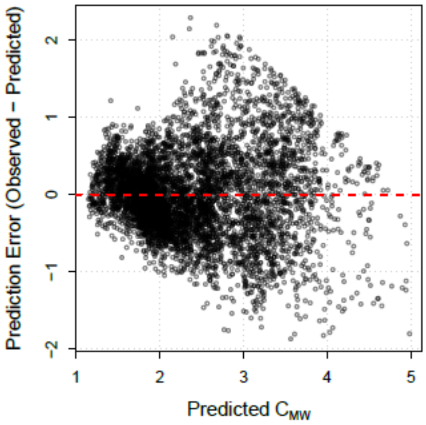
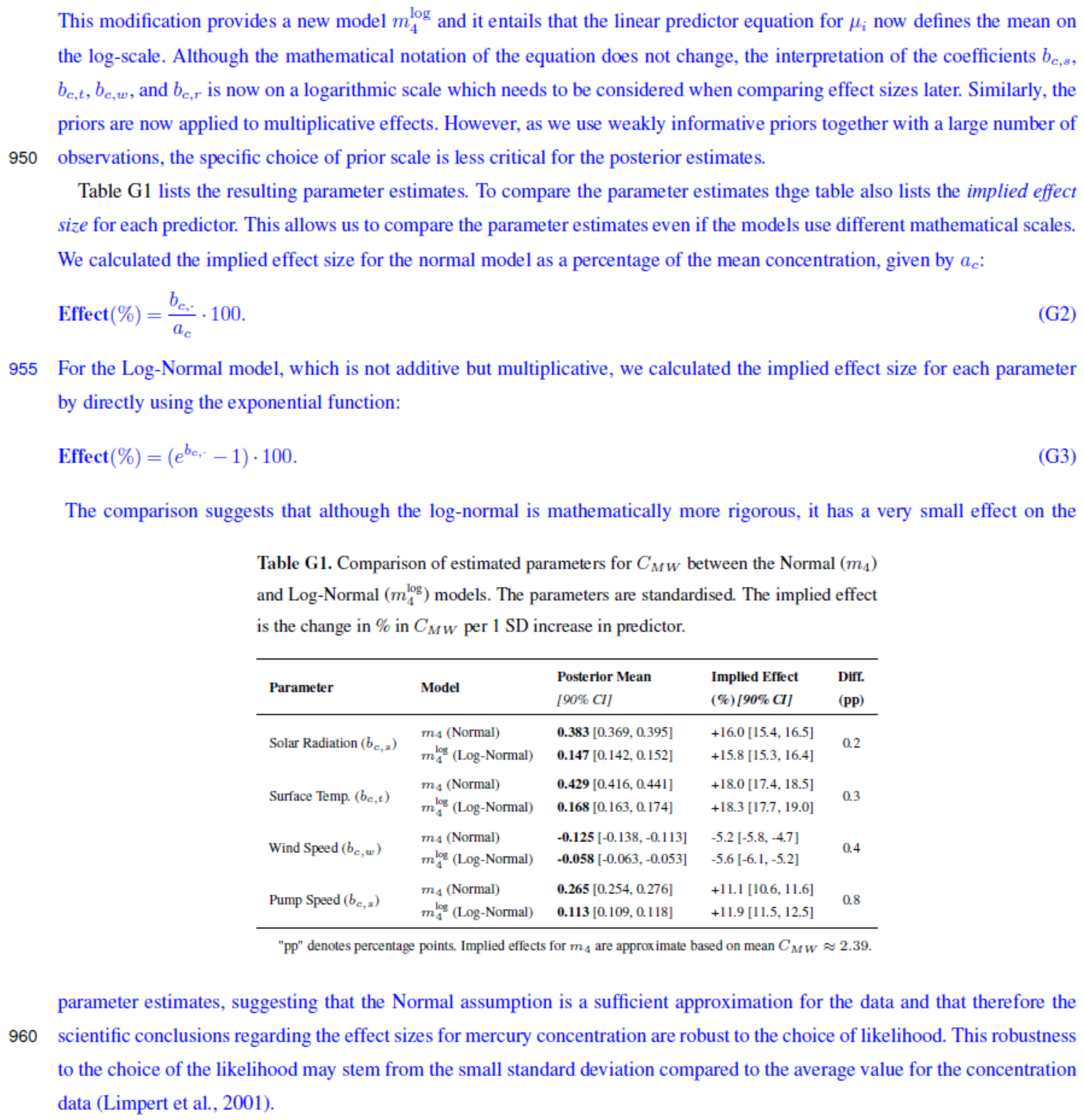


Figure G1. Residual plot which shows the prediction error against the predicted C_{MW}

Modified model m_4 with log-normal likelihood

As we cannot conclusively exclude the appropriateness of a Log-Normal likelihood for C_{MW}, we modified the likelihood of model m_4 listed in Table (2) Equation (9) to

$$C_{MW_i} \sim \text{Log-Normal}(\mu_i, \sigma). \tag{G1}$$



This modification provides a new model $m_4^{\log}$ and it entails that the linear predictor equation for μ_i now defines the mean on the log-scale. Although the mathematical notation of the equation does not change, the interpretation of the coefficients $b_{c,s}$, $b_{c,t}$, $b_{c,w}$, and $b_{c,r}$ is now on a logarithmic scale which needs to be considered when comparing effect sizes later. Similarly, the priors are now applied to multiplicative effects. However, as we use weakly informative priors together with a large number of observations, the specific choice of prior scale is less critical for the posterior estimates.

Table G1 lists the resulting parameter estimates. To compare the parameter estimates thge table also lists the *implied effect size* for each predictor. This allows us to compare the parameter estimates even if the models use different mathematical scales. We calculated the implied effect size for the normal model as a percentage of the mean concentration, given by a_c:

$$\textbf{Effect}(\%) = \frac{b_{c,\cdot}}{a_c} \cdot 100. \tag{G2}$$

For the Log-Normal model, which is not additive but multiplicative, we calculated the implied effect size for each parameter by directly using the exponential function:

$$\textbf{Effect}(\%) = (e^{b_{c,\cdot}} - 1) \cdot 100. \tag{G3}$$

The comparison suggests that although the log-normal is mathematically more rigorous, it has a very small effect on the parameter estimates, suggesting that the Normal assumption is a sufficient approximation for the data and that therefore the scientific conclusions regarding the effect sizes for mercury concentration are robust to the choice of likelihood. This robustness to the choice of the likelihood may stem from the small standard deviation compared to the average value for the concentration data (Limpert et al., 2001).

Table G1. Comparison of estimated parameters for C_{MW} between the Normal (m_4) and Log-Normal ($m_4^{\log}$) models. The parameters are standardised. The implied effect is the change in % in C_{MW} per 1 SD increase in predictor.

Parameter	**Model**	**Posterior Mean** *[90% CI]*	**Implied Effect (%)** ***[90% CI]***	**Diff. (pp)**
Solar Radiation ($b_{c,s}$)	m_4 (Normal) $m_4^{\log}$ (Log-Normal)	**0.383** [0.369, 0.395] **0.147** [0.142, 0.152]	+16.0 [15.4, 16.5] +15.8 [15.3, 16.4]	0.2
Surface Temp. ($b_{c,t}$)	m_4 (Normal) $m_4^{\log}$ (Log-Normal)	**0.429** [0.416, 0.441] **0.168** [0.163, 0.174]	+18.0 [17.4, 18.5] +18.3 [17.7, 19.0]	0.3
Wind Speed ($b_{c,w}$)	m_4 (Normal) $m_4^{\log}$ (Log-Normal)	**-0.125** [-0.138, -0.113] **-0.058** [-0.063, -0.053]	-5.2 [-5.8, -4.7] -5.6 [-6.1, -5.2]	0.4
Pump Speed ($b_{c,s}$)	m_4 (Normal) $m_4^{\log}$ (Log-Normal)	**0.265** [0.254, 0.276] **0.113** [0.109, 0.118]	+11.1 [10.6, 11.6] +11.9 [11.5, 12.5]	0.8

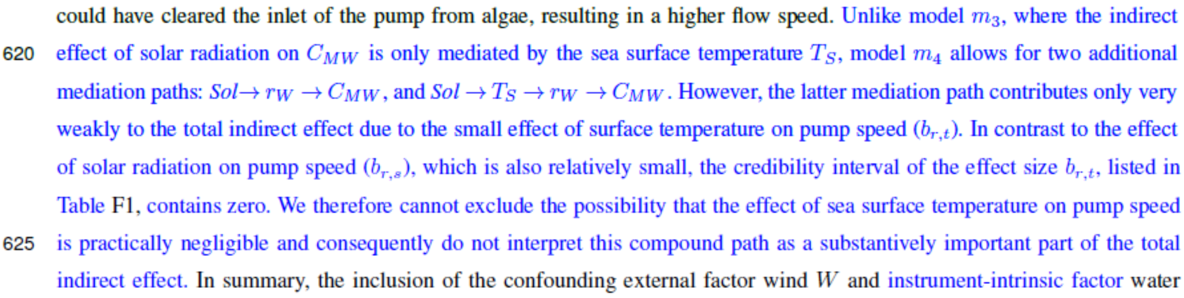
"pp" denotes percentage points. Implied effects for m_4 are approximate based on mean $C_{MW} \approx 2.39$.

4. Indirect Effects and DAG Interpretation (#520)

For model m4, the paper discusses indirect effects through Sol → T_S → C_{MW} and Sol → W → C_{MW} but omits the valid multi-step path Sol → T_S → r_W → C_{MW}. The authors should clarify whether such compound mediation effects are included in the total indirect effect and provide clearer guidance on interpreting direct, indirect, and total effects from the DAG.

- *We thank the reviewer for pointing to the compound mediation path Sol → T_S → r_W → C_{MW}. We have revised Section 5.2 (just before Table 4) and added a note in the table to acknowledge this additional path. We clarify that the multi-step path contributes only weakly to the total indirect effect due to the small estimated effect of sea surface temperature on pump speed b_{r,t}, and which, unlike the estimated effect of sol on pump speed b_{r,s}, contains zero in its 90% credible interval (Table F1). Consequently, we do not interpret the compound path Sol → T_S → r_W → C_{MW} as a substantively important mediation mechanism in model m_4.*

- *Changes in Section 5.2 #626-633:*



could have cleared the inlet of the pump from algae, resulting in a higher flow speed. Unlike model m_3, where the indirect effect of solar radiation on C_{MW} is only mediated by the sea surface temperature T_S, model m_4 allows for two additional mediation paths: $Sol \rightarrow r_W \rightarrow C_{MW}$, and $Sol \rightarrow T_S \rightarrow r_W \rightarrow C_{MW}$. However, the latter mediation path contributes only very weakly to the total indirect effect due to the small effect of surface temperature on pump speed ($b_{r,t}$). In contrast to the effect of solar radiation on pump speed ($b_{r,s}$), which is also relatively small, the credibility interval of the effect size $b_{r,t}$, listed in Table F1, contains zero. We therefore cannot exclude the possibility that the effect of sea surface temperature on pump speed is practically negligible and consequently do not interpret this compound path as a substantively important part of the total indirect effect. In summary, the inclusion of the confounding external factor wind W and instrument-intrinsic factor water

- *Change in Table 4:*

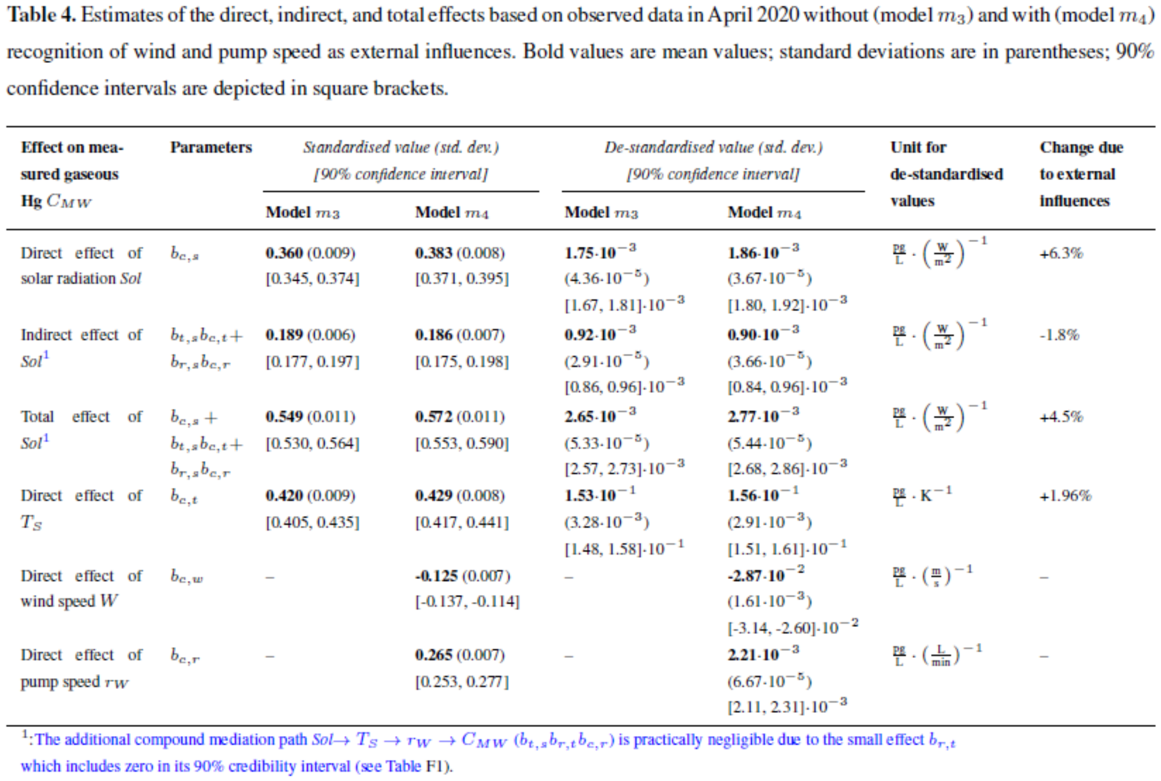


Table 4. Estimates of the direct, indirect, and total effects based on observed data in April 2020 without (model m_3) and with (model m_4) recognition of wind and pump speed as external influences. Bold values are mean values; standard deviations are in parentheses; 90% confidence intervals are depicted in square brackets.

Effect on measured gaseous Hg C_{MW}	**Parameters**	*Standardised value (std. dev.) [90% confidence interval]*		*De-standardised value (std. dev.) [90% confidence interval]*		**Unit for de-standardised values**	**Change due to external influences**
		Model m_3	**Model m_4**	**Model m_3**	**Model m_4**		
Direct effect of solar radiation *Sol*	$b_{c,s}$	**0.360** (0.009) [0.345, 0.374]	**0.383** (0.008) [0.371, 0.395]	$\mathbf{1.75 \cdot 10^{-3}}$ $(4.36 \cdot 10^{-5})$ $[1.67, 1.81] \cdot 10^{-3}$	$\mathbf{1.86 \cdot 10^{-3}}$ $(3.67 \cdot 10^{-5})$ $[1.80, 1.92] \cdot 10^{-3}$	$\frac{\text{pg}}{\text{L}} \cdot \left(\frac{\text{W}}{\text{m}^2}\right)^{-1}$	+6.3%
Indirect effect of *Sol*[1]	$b_{t,s}b_{c,t}+$ $b_{r,s}b_{c,r}$	**0.189** (0.006) [0.177, 0.197]	**0.186** (0.007) [0.175, 0.198]	$\mathbf{0.92 \cdot 10^{-3}}$ $(2.91 \cdot 10^{-5})$ $[0.86, 0.96] \cdot 10^{-3}$	$\mathbf{0.90 \cdot 10^{-3}}$ $(3.66 \cdot 10^{-5})$ $[0.84, 0.96] \cdot 10^{-3}$	$\frac{\text{pg}}{\text{L}} \cdot \left(\frac{\text{W}}{\text{m}^2}\right)^{-1}$	-1.8%
Total effect of *Sol*[1]	$b_{c,s}+$ $b_{t,s}b_{c,t}+$ $b_{r,s}b_{c,r}$	**0.549** (0.011) [0.530, 0.564]	**0.572** (0.011) [0.553, 0.590]	$\mathbf{2.65 \cdot 10^{-3}}$ $(5.33 \cdot 10^{-5})$ $[2.57, 2.73] \cdot 10^{-3}$	$\mathbf{2.77 \cdot 10^{-3}}$ $(5.44 \cdot 10^{-5})$ $[2.68, 2.86] \cdot 10^{-3}$	$\frac{\text{pg}}{\text{L}} \cdot \left(\frac{\text{W}}{\text{m}^2}\right)^{-1}$	+4.5%
Direct effect of T_S	$b_{c,t}$	**0.420** (0.009) [0.405, 0.435]	**0.429** (0.008) [0.417, 0.441]	$\mathbf{1.53 \cdot 10^{-1}}$ $(3.28 \cdot 10^{-3})$ $[1.48, 1.58] \cdot 10^{-1}$	$\mathbf{1.56 \cdot 10^{-1}}$ $(2.91 \cdot 10^{-3})$ $[1.51, 1.61] \cdot 10^{-1}$	$\frac{\text{pg}}{\text{L}} \cdot \text{K}^{-1}$	+1.96%
Direct effect of wind speed W	$b_{c,w}$	–	**-0.125** (0.007) [-0.137, -0.114]	–	$\mathbf{-2.87 \cdot 10^{-2}}$ $(1.61 \cdot 10^{-3})$ $[-3.14, -2.60] \cdot 10^{-2}$	$\frac{\text{pg}}{\text{L}} \cdot \left(\frac{\text{m}}{\text{s}}\right)^{-1}$	–
Direct effect of pump speed r_W	$b_{c,r}$	–	**0.265** (0.007) [0.253, 0.277]	–	$\mathbf{2.21 \cdot 10^{-3}}$ $(6.67 \cdot 10^{-5})$ $[2.11, 2.31] \cdot 10^{-3}$	$\frac{\text{pg}}{\text{L}} \cdot \left(\frac{\text{L}}{\text{min}}\right)^{-1}$	–

[1]:The additional compound mediation path $Sol \rightarrow T_S \rightarrow r_W \rightarrow C_{MW}$ ($b_{t,s}b_{r,t}b_{c,r}$) is practically negligible due to the small effect $b_{r,t}$ which includes zero in its 90% credibility interval (see Table F1).

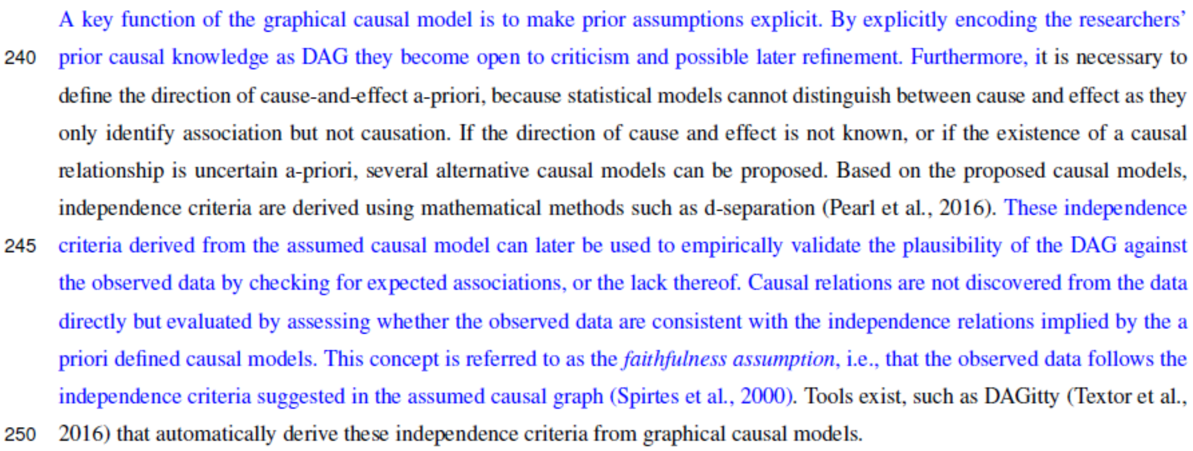
5. Limitation of dependence on DAG specification (#665)

The causal conclusions rely on the correctness of the assumed DAG structure in many aspects, in addition to independence, mis-specified relationships or omitted variables - such as unmodeled nonlinear effects or unobserved confounders - could lead to misleading causal inferences. The authors should discuss the potential impact of those DAG misspecification.

- *We thank the reviewer for suggestion to include a discussion on the potential misspecification of causal models. We agree that the causal conclusion derived from observational data depends on the assumed causal structure and that DAGs, as representation for assumed causal structures, can be misspecified through omitted variables, incorrect directions of cause-and-effect, or inadequate functional assumptions which can affect the causal interpretation of the results. We have therefore revised the manuscript to explicitly reflect and discuss these limitations.*

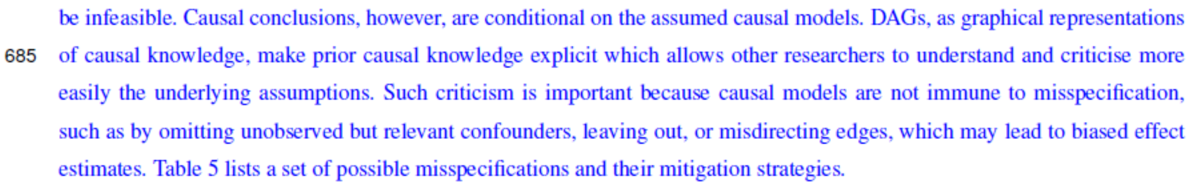
In Section 3, when introducing the framework for causal inference, we clarify that a key function of graphical causal models is to make the researchers' prior causal assumptions explicit which opens these assumptions to criticism and possible refinement. Furthermore, as part of the discussion in Section 6, we explicitly state that the causal conclusions are conditional on the assumed causal models and that DAGs are not immune to misspecification. We introduced Table 5, which summarises a set of possible DAG misspecifications such as omitted confounders, unmodelled nonlinearities and missing or misdirected edges, discusses their potential impact on the results, and provides general mitigation strategies.

- *Changes in Section 3 #239-249:*



A key function of the graphical causal model is to make prior assumptions explicit. By explicitly encoding the researchers' prior causal knowledge as DAG they become open to criticism and possible later refinement. Furthermore, it is necessary to define the direction of cause-and-effect a-priori, because statistical models cannot distinguish between cause and effect as they only identify association but not causation. If the direction of cause and effect is not known, or if the existence of a causal relationship is uncertain a-priori, several alternative causal models can be proposed. Based on the proposed causal models, independence criteria are derived using mathematical methods such as d-separation (Pearl et al., 2016). These independence criteria derived from the assumed causal model can later be used to empirically validate the plausibility of the DAG against the observed data by checking for expected associations, or the lack thereof. Causal relations are not discovered from the data directly but evaluated by assessing whether the observed data are consistent with the independence relations implied by the a priori defined causal models. This concept is referred to as the *faithfulness assumption*, i.e., that the observed data follows the independence criteria suggested in the assumed causal graph (Spirtes et al., 2000). Tools exist, such as DAGitty (Textor et al., 2016) that automatically derive these independence criteria from graphical causal models.

- *Section 6.1 #684-688 and Table 5:*



be infeasible. Causal conclusions, however, are conditional on the assumed causal models. DAGs, as graphical representations of causal knowledge, make prior causal knowledge explicit which allows other researchers to understand and criticise more easily the underlying assumptions. Such criticism is important because causal models are not immune to misspecification, such as by omitting unobserved but relevant confounders, leaving out, or misdirecting edges, which may lead to biased effect estimates. Table 5 lists a set of possible misspecifications and their mitigation strategies.

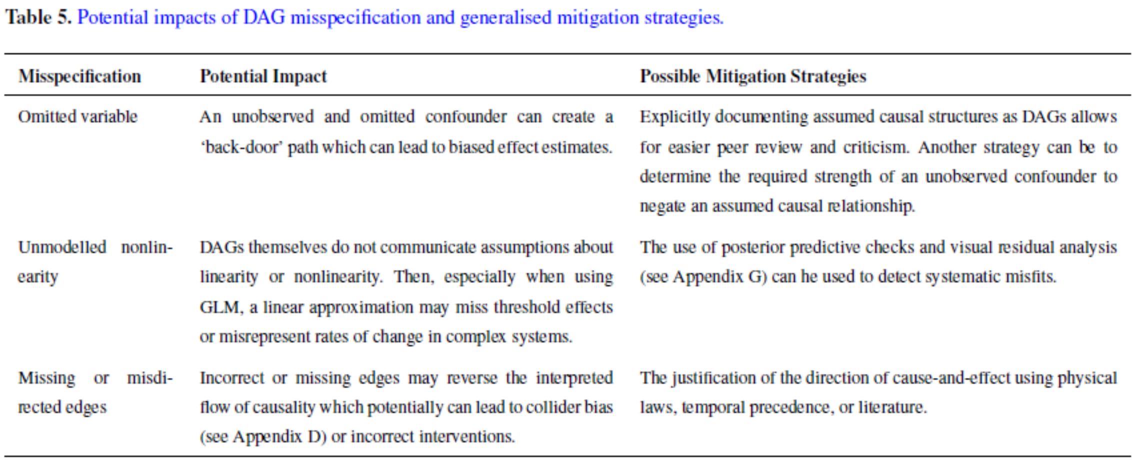


Table 5. Potential impacts of DAG misspecification and generalised mitigation strategies.

Misspecification	Potential Impact	Possible Mitigation Strategies
Omitted variable	An unobserved and omitted confounder can create a 'back-door' path which can lead to biased effect estimates.	Explicitly documenting assumed causal structures as DAGs allows for easier peer review and criticism. Another strategy can be to determine the required strength of an unobserved confounder to negate an assumed causal relationship.
Unmodelled nonlinearity	DAGs themselves do not communicate assumptions about linearity or nonlinearity. Then, especially when using GLM, a linear approximation may miss threshold effects or misrepresent rates of change in complex systems.	The use of posterior predictive checks and visual residual analysis (see Appendix G) can he used to detect systematic misfits.
Missing or misdirected edges	Incorrect or missing edges may reverse the interpreted flow of causality which potentially can lead to collider bias (see Appendix D) or incorrect interventions.	The justification of the direction of cause-and-effect using physical laws, temporal precedence, or literature.

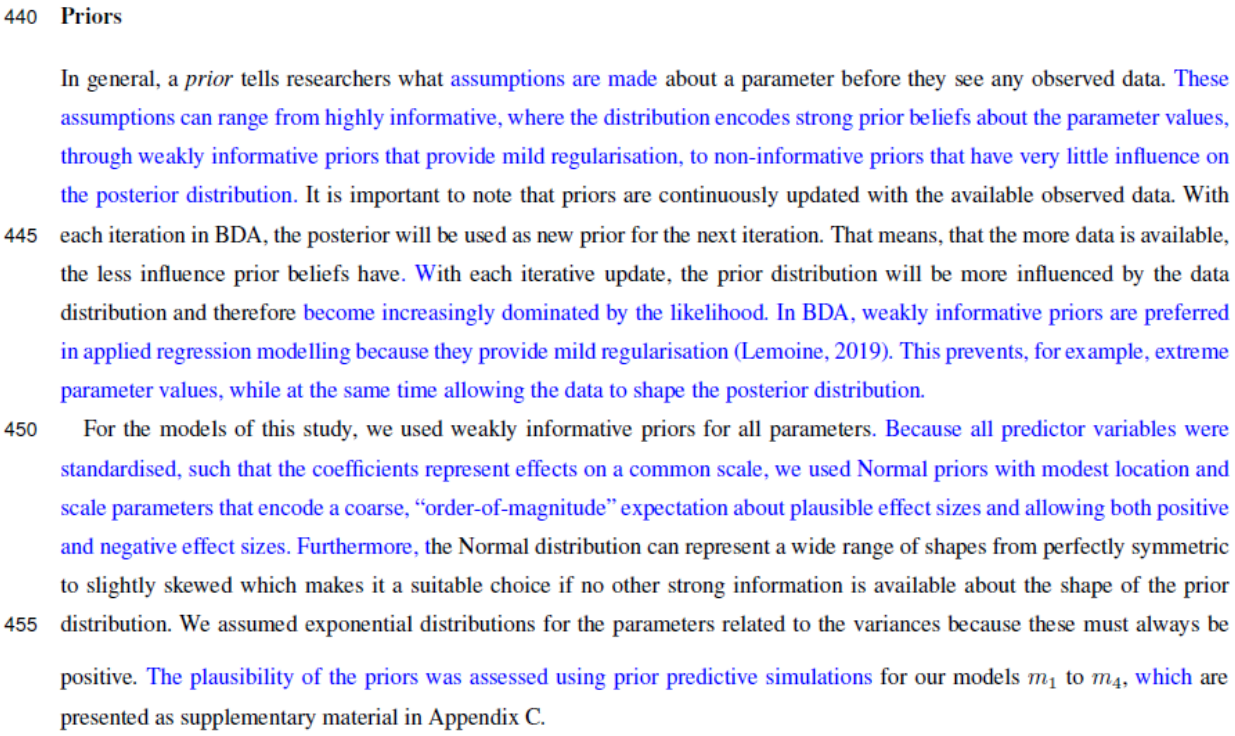
Minor Comments

1. #330

   The priors (e.g., Normal(0.5, 1), Normal(0.5, 0.5)) appear somewhat arbitrary and not elicited from domain experts. **The study would be strengthened by (a) justifying these priors through expert input or empirical reasoning, or (b) using uninformative priors.**

- *We thank the reviewer for raising the point regarding the justification for the choice of priors. We have revised the paragraph that provides the rationale and the role of the used priors in Section 4.5.*
  Specifically, we now explicitly state that the priors are weakly informative rather than expert-elicted or non-informative, and we explain why this choice is appropriate for our analysis. We further added a clarification that uninformative priors are not generally preferable in applied regression models, and we refer to recent methodological work that recommends weakly informative prior as a principled default in BDA (Lemoine, 2019). Finally, we also emphasise that the plausibility of the priors was asses using prior predictive simulations.

- *Section 4.5 #4481-454:*



Priors

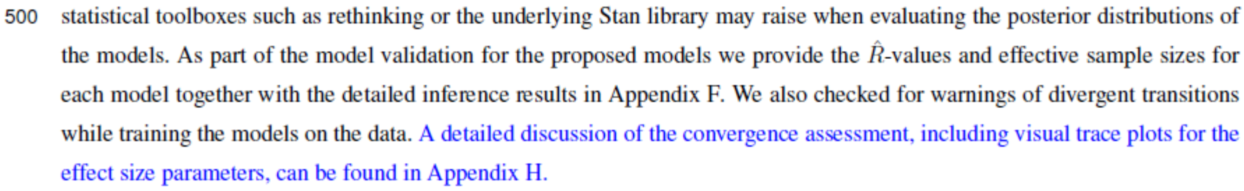
In general, a *prior* tells researchers what assumptions are made about a parameter before they see any observed data. These assumptions can range from highly informative, where the distribution encodes strong prior beliefs about the parameter values, through weakly informative priors that provide mild regularisation, to non-informative priors that have very little influence on the posterior distribution. It is important to note that priors are continuously updated with the available observed data. With each iteration in BDA, the posterior will be used as new prior for the next iteration. That means, that the more data is available, the less influence prior beliefs have. With each iterative update, the prior distribution will be more influenced by the data distribution and therefore become increasingly dominated by the likelihood. In BDA, weakly informative priors are preferred in applied regression modelling because they provide mild regularisation (Lemoine, 2019). This prevents, for example, extreme parameter values, while at the same time allowing the data to shape the posterior distribution.

For the models of this study, we used weakly informative priors for all parameters. Because all predictor variables were standardised, such that the coefficients represent effects on a common scale, we used Normal priors with modest location and scale parameters that encode a coarse, "order-of-magnitude" expectation about plausible effect sizes and allowing both positive and negative effect sizes. Furthermore, the Normal distribution can represent a wide range of shapes from perfectly symmetric to slightly skewed which makes it a suitable choice if no other strong information is available about the shape of the prior distribution. We assumed exponential distributions for the parameters related to the variances because these must always be positive. The plausibility of the priors was assessed using prior predictive simulations for our models m_1 to m_4, which are presented as supplementary material in Appendix C.

2. #445

   Please clarify how model convergence was assessed under the Bayesian MCMC framework. Including trace plots or diagnostics is important for verifying convergence. A useful reference is: *Reich, Brian J., and Sujit K. Ghosh. Bayesian Statistical Methods. Chapman and Hall/CRC, 2019.*

- *We thank the reviewer for suggesting to improve the documentation of the model convergence under the Bayesian MCMC framework in the manuscript. We have revised the manuscript accordingly to explicitly describe how we assessed convergence, including visual trace plots. In Section 4.9 (Paragraph Workability) we refer to a new appendix section (Appendix H) that presents trace plots and provides a detailed discussion of the convergence assessment.*
- *Section 4.9 #503-504:*



statistical toolboxes such as rethinking or the underlying Stan library may raise when evaluating the posterior distributions of the models. As part of the model validation for the proposed models we provide the $\hat{R}$-values and effective sample sizes for each model together with the detailed inference results in Appendix F. We also checked for warnings of divergent transitions while training the models on the data. A detailed discussion of the convergence assessment, including visual trace plots for the effect size parameters, can be found in Appendix H.

- *New Appendix H #964-975:*

Appendix H: Workability: Assessing the convergence under Bayesian MCMC

We conducted Bayesian inference using a Markov Chain Monte Carlo (MCMC) sampling approach with Hamiltonian Monte Carlo implemented in Stan (Stan Development Team) using the rethinking interface by McElreath (2020). We assessed convergence using both quantitative diagnostics, including $\hat{R}$ and the effective samples size (ESS/n_{eff}) as well as visual diagnostics, following standard recommendation for Bayesian workflows (Vehtari et al., 2021; Reich and Ghosh, 2019). First, the $\hat{R}$ values for all parameters were close to 1 and < 1.01 as reported in Table F1. Second, all parameters have effective sample sizes (ESS/n_{eff}) exceeding 10% of the total sample sizes which we assume sufficiently large (see Vehtari et al. (2021) and Furia et al. (2022) for a discussion on the sufficient ESS for BDA). Finally, we visually inspected the trace plots to verify adequate mixing, absence of strange divergent behaviour and stationarity. The trace plots are provided in Figure H1 and show no indication of non-convergence such as slow trends, chain separation or autocorrelation. Together, these diagnostics provide evidence that the MCMC chains converged.

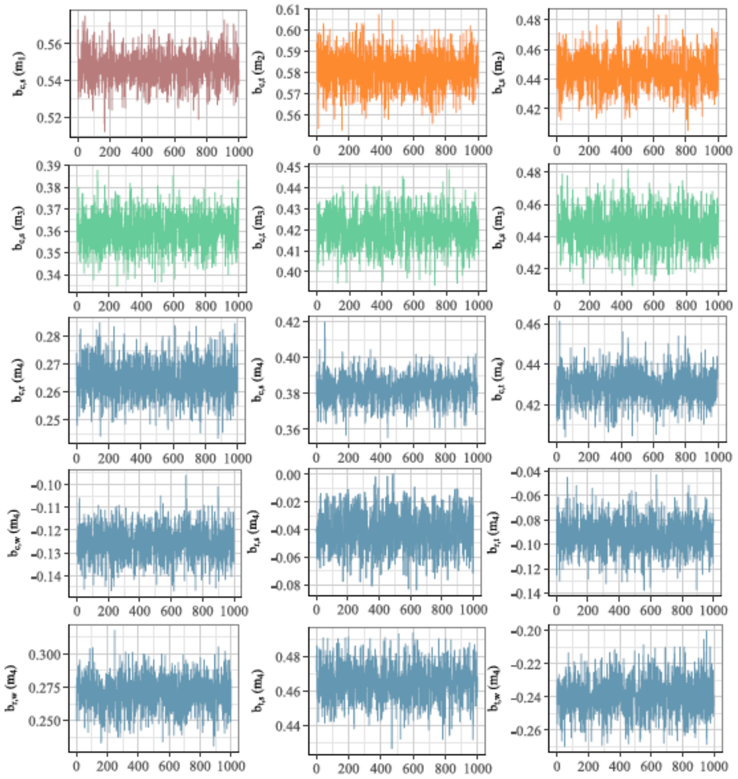


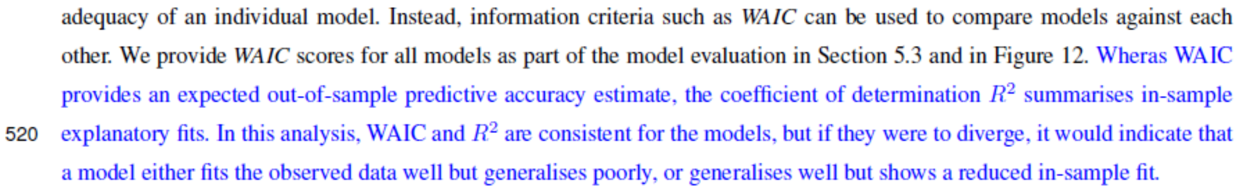
Figure H1. Trace plots for all effect size parameters of models m_1 (red), m_2 (orange), m_3 (green), and m_4 (blue).

3. #445

Both R2 and WAIC are reported and appear consistent. However, if they diverged, how should this be interpreted? A short explanation of their conceptual difference would improve clarity.

- *We thank the reviewer for this suggestion. We added in Section 4.9 a brief clarification on the difference between R^2 (in-sample explanatory fit) and WAIC (expected out-of-sample predictive accuracy estimate). We also discuss how a potential divergence can be interpreted.*

- *Addition made to Section 4.9 #518-521:*



adequacy of an individual model. Instead, information criteria such as $WAIC$ can be used to compare models against each other. We provide $WAIC$ scores for all models as part of the model evaluation in Section 5.3 and in Figure 12. Wheras WAIC provides an expected out-of-sample predictive accuracy estimate, the coefficient of determination R^2 summarises in-sample explanatory fits. In this analysis, WAIC and R^2 are consistent for the models, but if they were to diverge, it would indicate that a model either fits the observed data well but generalises poorly, or generalises well but shows a reduced in-sample fit.

4. #605

Figure 13(b) seems to show narrower confidence intervals than (a), but this is hard to discern. The figure could be redesigned for better contrast. Also, revise the phrasing “noisier but also more reliable,” as “noisier” typically suggests lower precision.

- *We thank the reviewer for suggesting to improve Figure 13 (b). We have revised the Figure to improve contrast and interpretability by plotting now the posterior mean regression functions and their associated 95% posterior credible intervals instead of the earlier posterior predictive simulations which included observational noise. This change allows for a better direct visual comparison of the effect sizes. We have also revised the corresponding Section 6.1 to remove the “noiser” term and to clarify why model m_4 provides a less biased and more causally interpretable estimate by accounting for mediating and confounding processes.*

- *Revised Figure 13:*

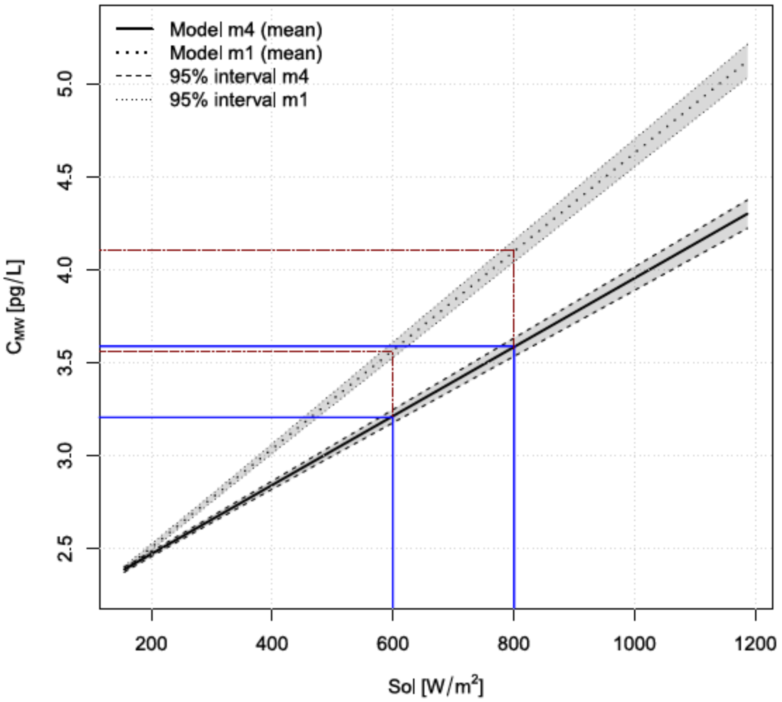
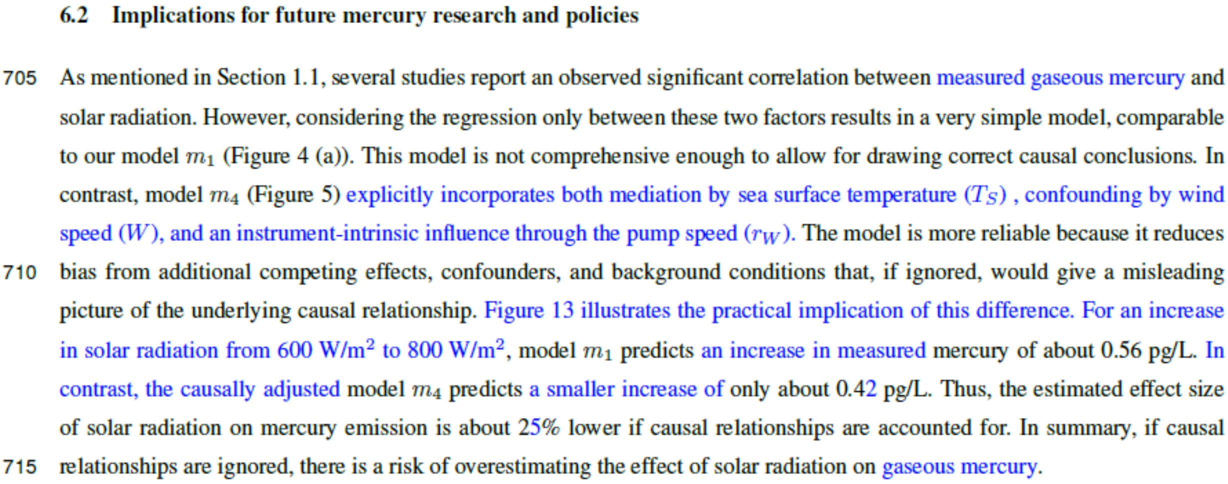


Figure 13. Comparison of posterior mean regressions of the effect of solar radiation Sol on mercury concentration C_{MW} using model m_1 (dotted line) and model m_4 (solid line). Grey shading indicates the 95% confidence interval. Vertical reference lines at Sol= 600 and 800 W/m^2 are the solar radiation levels used to compare effect sizes between the models.

- *Section 6.2 (former 6.1) #711-721:*



6.2 Implications for future mercury research and policies

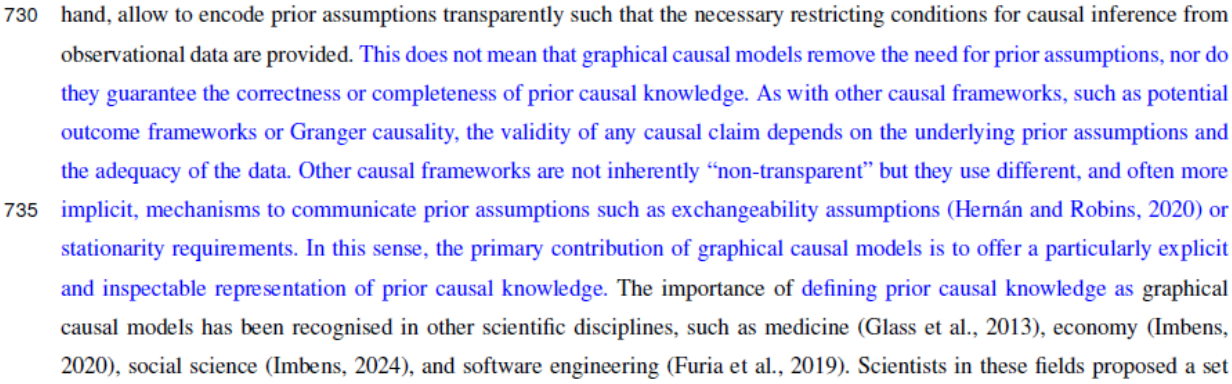
As mentioned in Section 1.1, several studies report an observed significant correlation between measured gaseous mercury and solar radiation. However, considering the regression only between these two factors results in a very simple model, comparable to our model m_1 (Figure 4 (a)). This model is not comprehensive enough to allow for drawing correct causal conclusions. In contrast, model m_4 (Figure 5) explicitly incorporates both mediation by sea surface temperature (T_S) , confounding by wind speed (W), and an instrument-intrinsic influence through the pump speed (r_W). The model is more reliable because it reduces bias from additional competing effects, confounders, and background conditions that, if ignored, would give a misleading picture of the underlying causal relationship. Figure 13 illustrates the practical implication of this difference. For an increase in solar radiation from 600 W/m^2 to 800 W/m^2, model m_1 predicts an increase in measured mercury of about 0.56 pg/L. In contrast, the causally adjusted model m_4 predicts a smaller increase of only about 0.42 pg/L. Thus, the estimated effect size of solar radiation on mercury emission is about 25% lower if causal relationships are accounted for. In summary, if causal relationships are ignored, there is a risk of overestimating the effect of solar radiation on gaseous mercury.

5. #615

The rationale for preferring graphical causal models over alternatives (e.g., Granger causality, potential outcomes) is generally sound. Graphical models do enhance transparency and facilitate the integration of mechanistic knowledge. However, they do not eliminate assumptions or guarantee correctness. Traditional causal frameworks are not inherently "non-transparent" but rely on different theoretical foundations. Acknowledging this nuance would make the argument more balanced.

- *We thank the reviewer for suggesting a more balanced comparison between graphical causal models and other causal frameworks. We have therefore revised Section 6.2 to clarify that alternative causal frameworks are not inherently non-transparent but instead formalise assumptions using different constructs such as exchangeability assumptions. We further emphasise that the primary contribution of graphical causal models lies in making prior causal assumptions explicit and inspectable rather than removing prior assumptions altogether.*

- *Section 6.3 (former 6.2) #731-737:*



hand, allow to encode prior assumptions transparently such that the necessary restricting conditions for causal inference from observational data are provided. This does not mean that graphical causal models remove the need for prior assumptions, nor do they guarantee the correctness or completeness of prior causal knowledge. As with other causal frameworks, such as potential outcome frameworks or Granger causality, the validity of any causal claim depends on the underlying prior assumptions and the adequacy of the data. Other causal frameworks are not inherently "non-transparent" but they use different, and often more implicit, mechanisms to communicate prior assumptions such as exchangeability assumptions (Hernán and Robins, 2020) or stationarity requirements. In this sense, the primary contribution of graphical causal models is to offer a particularly explicit and inspectable representation of prior causal knowledge. The importance of defining prior causal knowledge as graphical causal models has been recognised in other scientific disciplines, such as medicine (Glass et al., 2013), economy (Imbens, 2020), social science (Imbens, 2024), and software engineering (Furia et al., 2019). Scientists in these fields proposed a set

6. #805

Appendix E Figure E1, used to validate statistical independence, could be clearer. Adding fitted lines with distinct colors for different temperature levels would improve readability and interpretation.

- *We thank the reviewer for suggesting adding fitted lines with distinct colours to the scatter plots. We have revised Figure E1 accordingly.*

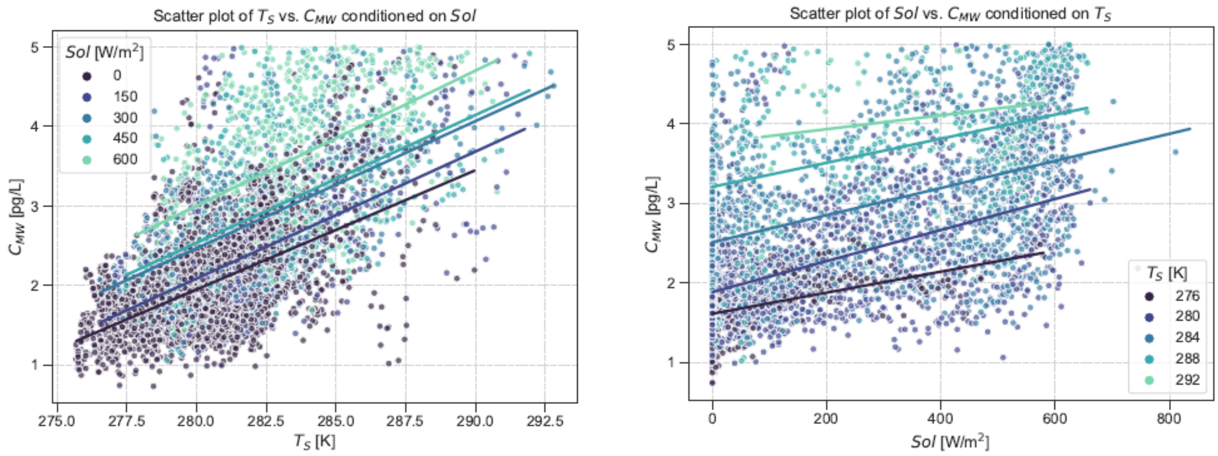


Figure E1. Scatter plots for inspecting suggested independence between variables for (a) m_1: $T_S \perp C_{MW} \mid Sol$ and (b) m_2: $Sol \perp C_{MW} \mid T_S$.