Replication of Hoffmann et al. (2020)

Jack Murphy

05/05/2021

Overview of the Original Paper

"A meta-analysis of country-level studies on environmental change and migration" by Roman Hoffmann, Anna Dimitrova, Raya Muttarak, Jesus Crespo Cuaresma, and Jonas Peisker, published in Nature in October 2020, conducted, as the title suggests, a meta-analysis of 30 selected studies published between 2006 and 2019 examining environmental change and out-migration.[1] The authors extracted estimates of the relationship between environmental change factors and migration from these 30 papers, generating 1803 such estimates each of which became a separate observation in their own models.

The authors focus on country-level studies that consider a variety of environmental factors like changing temperatures, heavy storms, and droughts. They standardized the effects of these factors by finding the mean and standard deviation for effects used across multiple studies, and expressing those values in terms of standard deviations from their respective means.

The authors then review the studies generally, categorizing the environmental changes researched into rapid-onset or gradual changes, noting that a large majority of the used estimates were for gradual environmental changes rather than rapid-onset events such as floods or landslides. They note that 88.0% of the estimates focus on international migration, and that 27 of the studies find a notable positive relationship between environmental changes and migration.

After weighting the study results for their precision, the paper finds a statistically significant positive relationship between such environmental changes and migration, though the strength and direction of the relationship vary between studies. The authors found that "on average, a one standard deviation change in the environmental conditions leads to an increase in migration by 0.021 standard deviations."

Their first figure explores the heterogeneity of the studies analysed: regressing characteristics of the models the original studies used. They found that rapid-onset changes in the environment have large impacts on migration, as do rainfall anomalies (like droughts), but changes in temperatures by contrast have a small - yet statistically significant - impact. Unfortunately the publicly available code did not include the lines needed to replicate this figure, so it is not included here.

The below table is the first table included in the paper, the replicated results completely match the original published results. It displays the resulting estimates for the impact of increasing numbers of variables on migration, as calculated by five increasingly complex fixed-effect linear models.

The first model shows the estimated impact of temperature-level changes on migration in terms of standard deviations of migration, relative to a comparable change in precipitation levels.

The model presented in the second column takes into account broader timeframes for the relationship between environmental changes and migration, looking at 5- and 10-year timeframes as opposed to 1-year time frames examined in the first column.

The third model estimates take into account both these extended time frames as well as global migration flows. However as most of the original studies did not measure internal migration, the differences are small

[1] The data and code used in this replication report can be obtained at https://github.com/johnpgmurphy/migration_env_replication

(albeit robust).

The fourth and fifth models dive into the international migration flows, finding larger effects tying environmental changes to migration towards non-OECD and low- or middle-income countries. The difference between the internal and global migration effects and these effects corroborates findings in the literature "that environmental migration is often short-distance, regional and temporary."

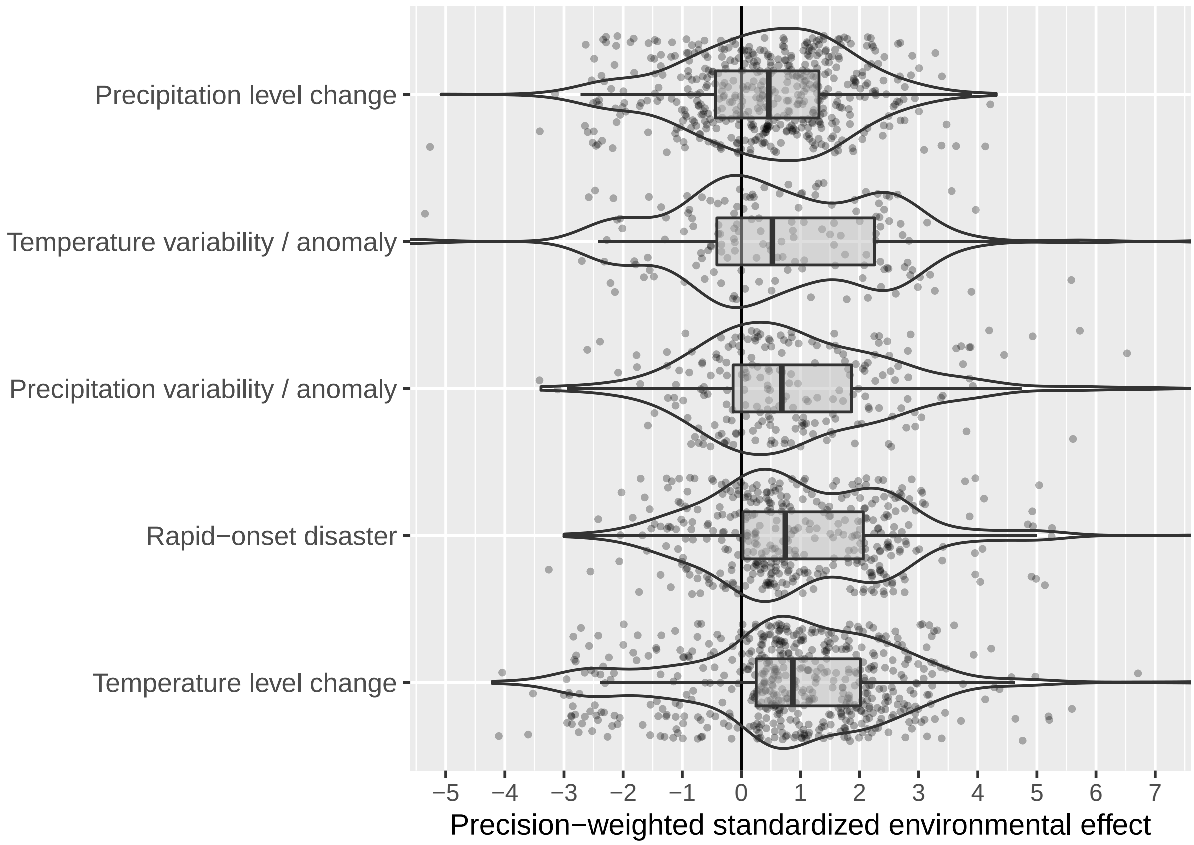
These last two models take various samples of the countries examined in the original studies, sorting them into groups such as non-OECD, low-income, low-middle-income, upper-middle-income, and agriculturally dependent countries. This grouping is intended to evaluate the difference in impact of environmental changes on migration depending on the economic characteristics of the country of origin. These estimates indicate that the economic constraints faced by residents of low-income countries restrict migration following environmental changes as contrasted with the movements of residents of higher-income countries.

Table 1:

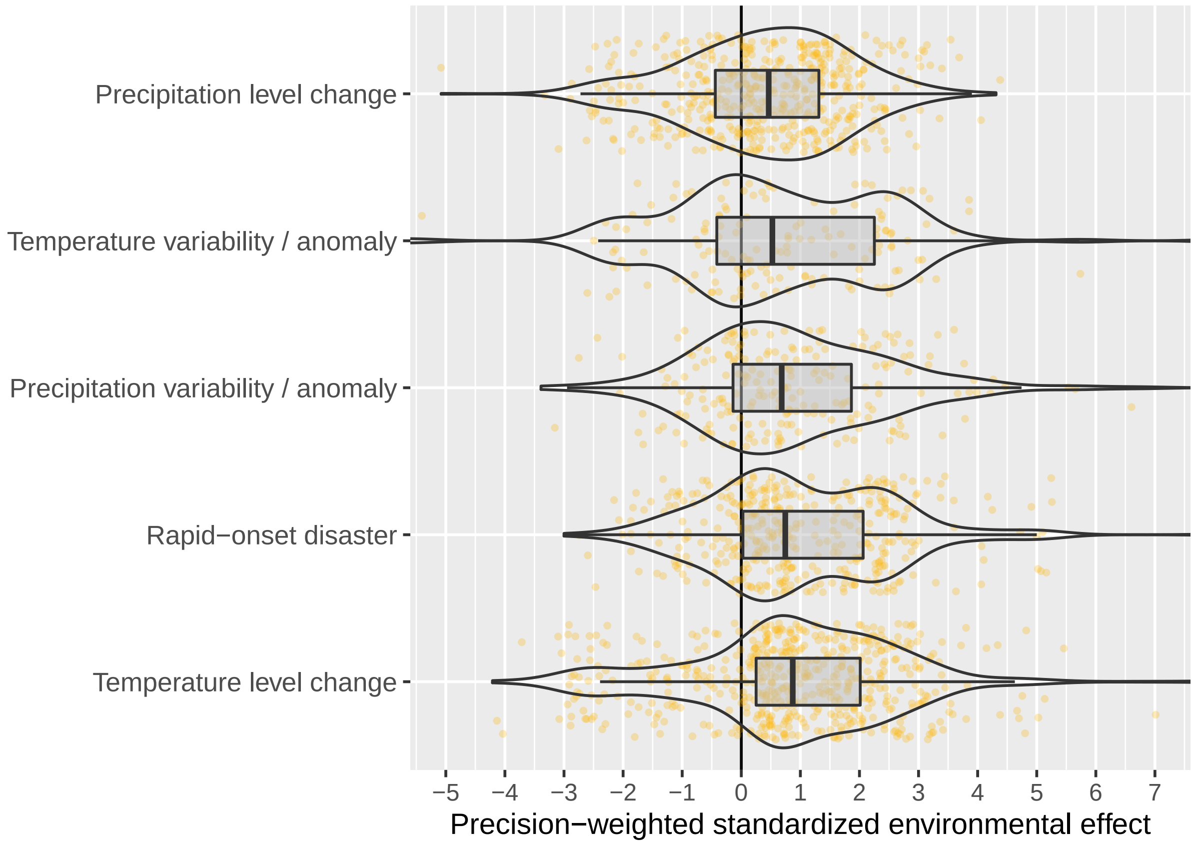
	Dependent variable:				
	stancoeff				
	felm				
	(1)	(2)	(3)	(4)	(5)
pre_var	0.017^{**} (0.008)	0.016^{*} (0.008)	0.016^{**} (0.007)	0.016^{**} (0.007)	0.015^{*} (0.008)
rapidonset	0.015^{*} (0.008)	0.015^{*} (0.008)	0.015^{*} (0.008)	0.015^{*} (0.008)	0.014^{*} (0.008)
tem_lev	0.018^{**} (0.008)	0.017^{**} (0.008)	0.018^{**} (0.008)	0.018^{**} (0.008)	0.017^{**} (0.008)
tem_var	0.016^{*} (0.008)	0.014^{*} (0.008)	0.014^{*} (0.007)	0.014^{*} (0.007)	0.013^{*} (0.008)
env_lag_dum		−0.001 (0.001)	−0.0004 (0.0005)	−0.0004 (0.001)	−0.001 (0.001)
env_timespan_dum		-0.016^{***} (0.0004)	-0.016^{***} (0.0003)	-0.017^{***} (0.0003)	-0.010^{***} (0.002)
env_other_dum		-0.002^{**} (0.001)	-0.003^{***} (0.0005)	-0.003^{***} (0.0005)	-0.003^{***} (0.001)
internal			0.006^{**} (0.002)	0.006^{**} (0.002)	0.006^{**} (0.003)
dest_low			0.069^{**} (0.030)	0.068^{**} (0.029)	0.063^{**} (0.027)
dest_high			0.005 (0.003)	0.005^{*} (0.003)	0.003 (0.003)
dest_ambi			-0.008^{***} (0.003)	-0.007^{**} (0.003)	-0.010^{***} (0.003)
nonoecd				0.006^{**} (0.002)	
L					-0.074^{***} (0.021)
LM					0.014^{**} (0.006)
UM					0.044^{***} (0.008)
agr					0.104^{***} (0.022)
conflict_mepv_5					-0.022^{***} (0.003)
Observations	1,803	1,803	1,803	1,803	1,803
R^2	0.272	0.284	0.305	0.309	0.340
Adjusted R^2	0.255	0.267	0.287	0.291	0.321
Residual Std. Error	1.576 (df = 1763)	1.564 (df = 1760)	1.543 (df = 1756)	1.539 (df = 1755)	1.506 (df = 1751)
Note:					*p<0.1; **p<0.05; ***p<0.01
					Test

In extending this paper I would amend three figures, although as the figures themselves do convey information adequately, the proposed extensions are more aesthetic than substantive.

The first is figure 2, which visualizes the distribution of the effects of give types of environmental hazards on migration using violin plots. The original is replicated below.

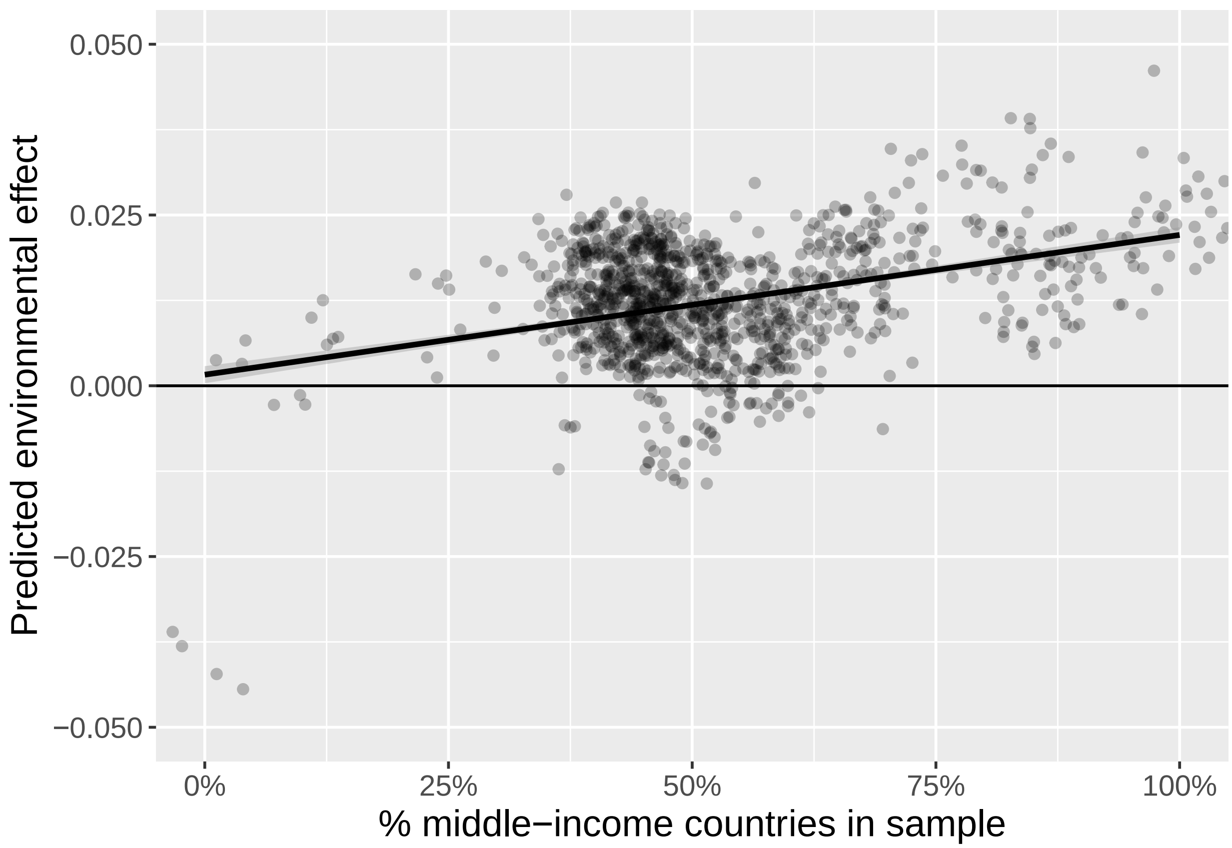


The proposed extension to this figure would make it more visually intuitive, adding varying point shapes which highlight the outliers of the graph, and colors that

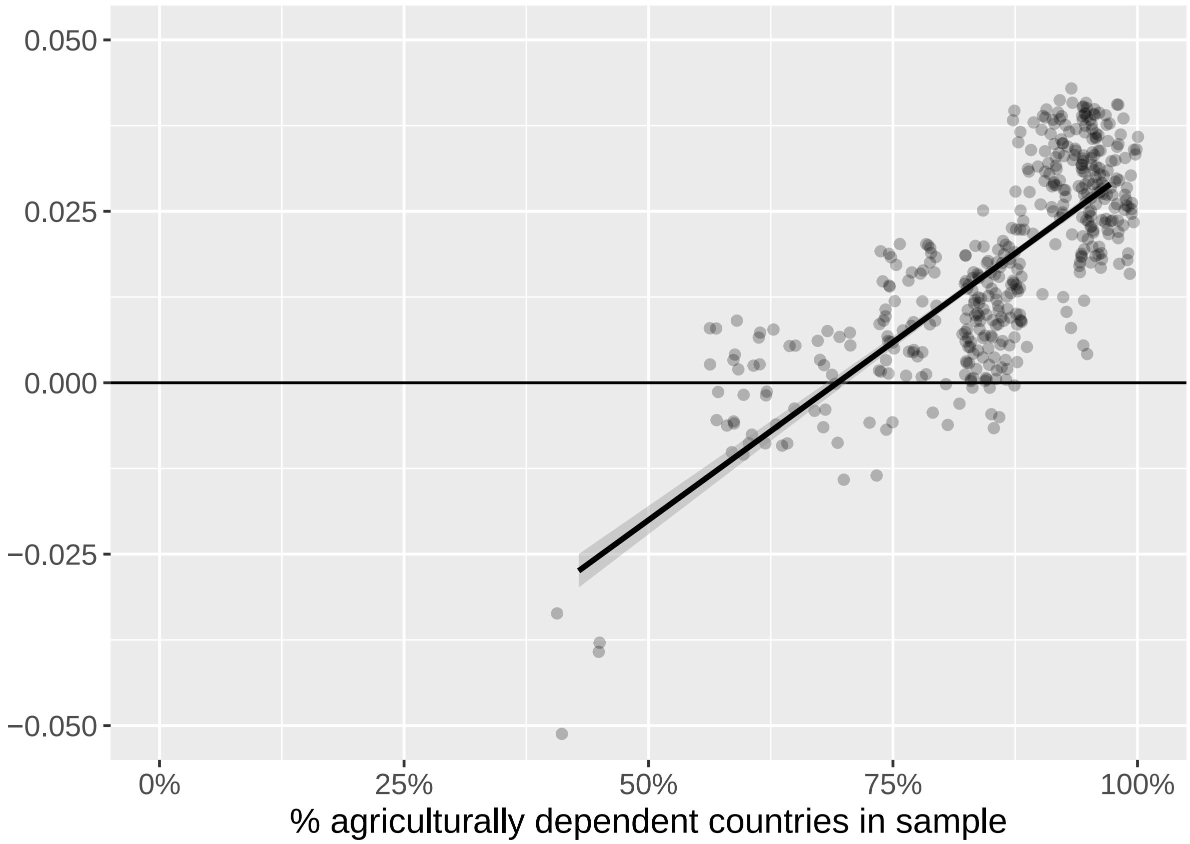


The second would be figure 3, similarly adding color to the plotted data and regression lines to make the visualisations more appealing.

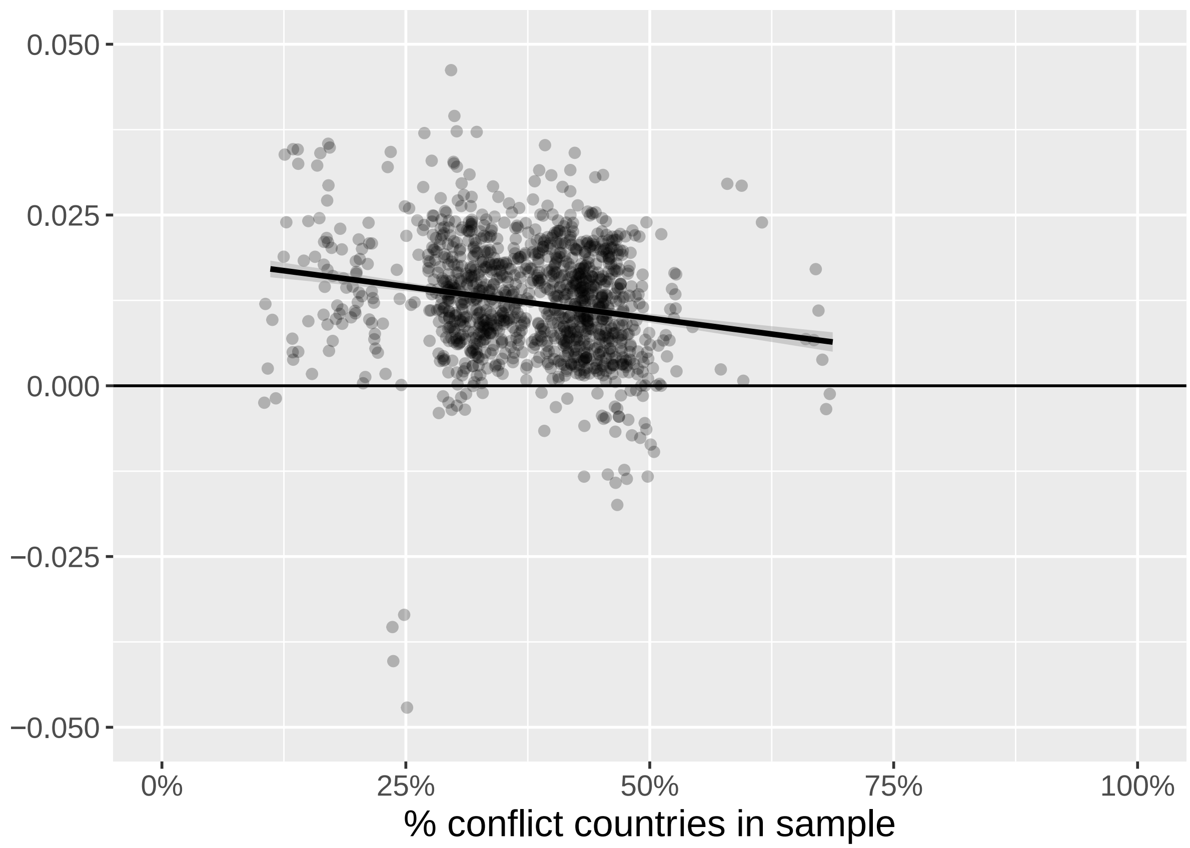
The original visualisations are made below:



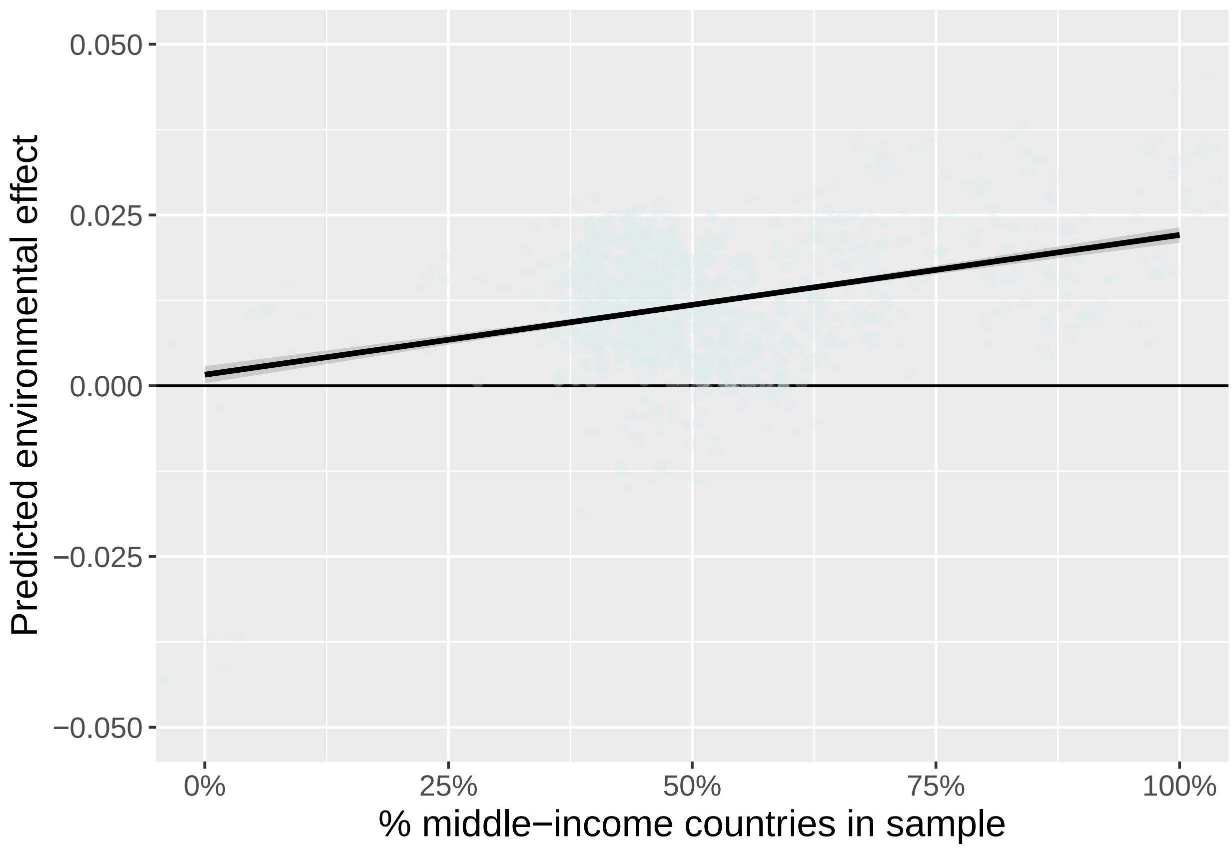
Predicted environmental effect
0.050
0.025
0.000
−0.025
−0.050
0%
25%
50%
75%
100%
% middle−income countries in sample



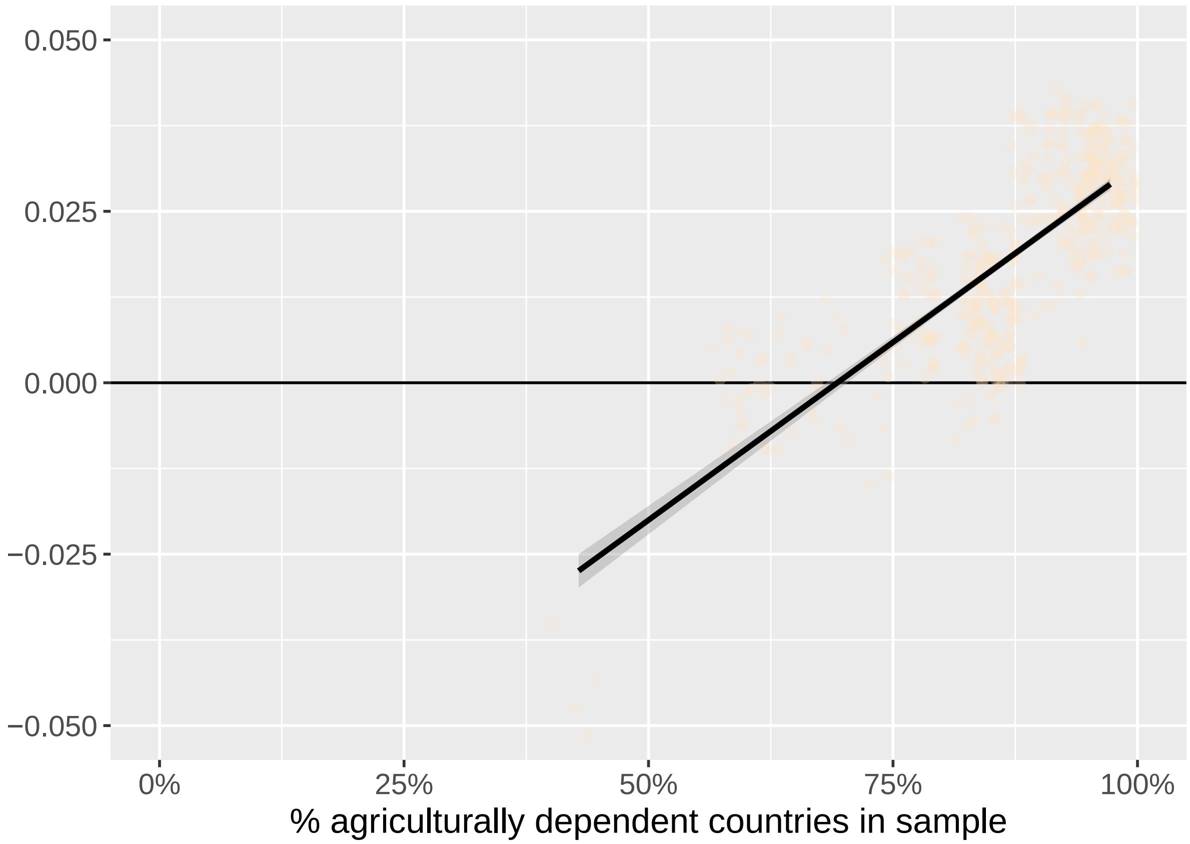
0.050
0.025
0.000
−0.025
−0.050
0%
25%
50%
75%
100%
% agriculturally dependent countries in sample

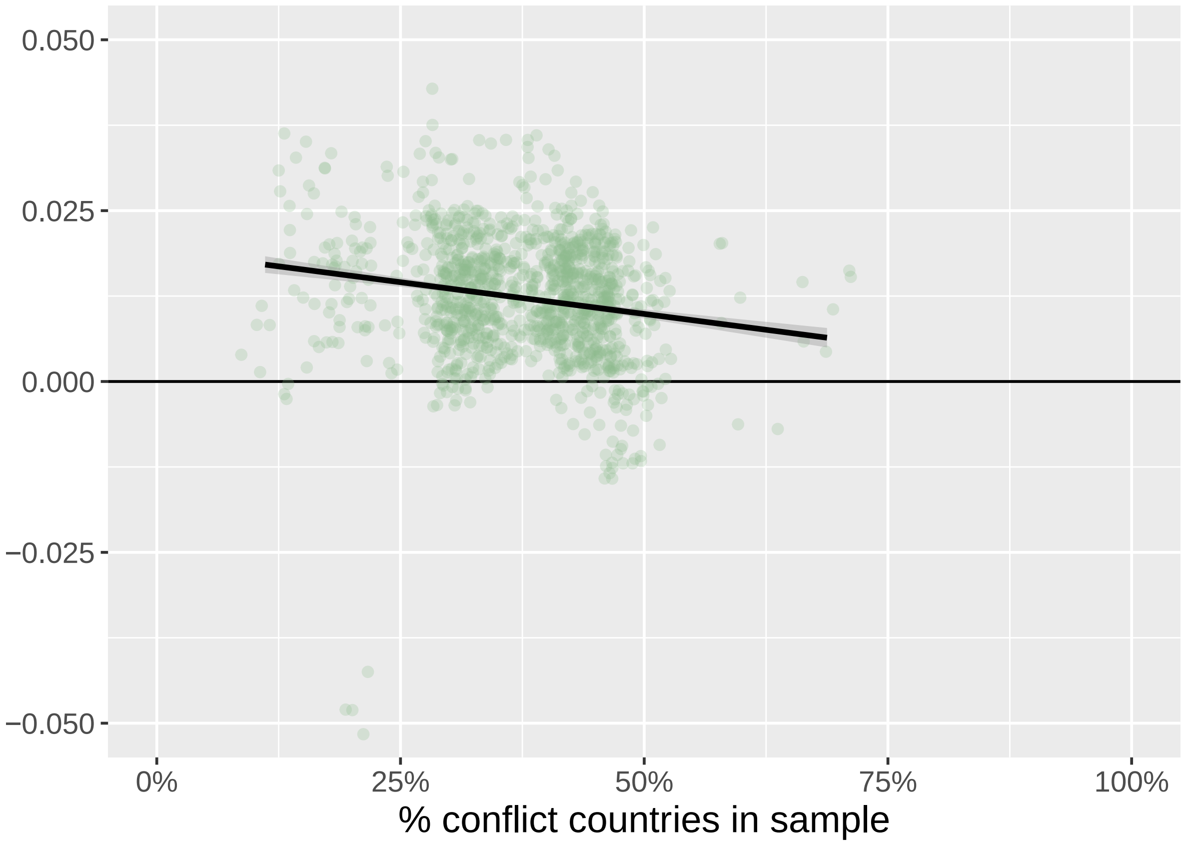


The proposed amended plots would look something like this:

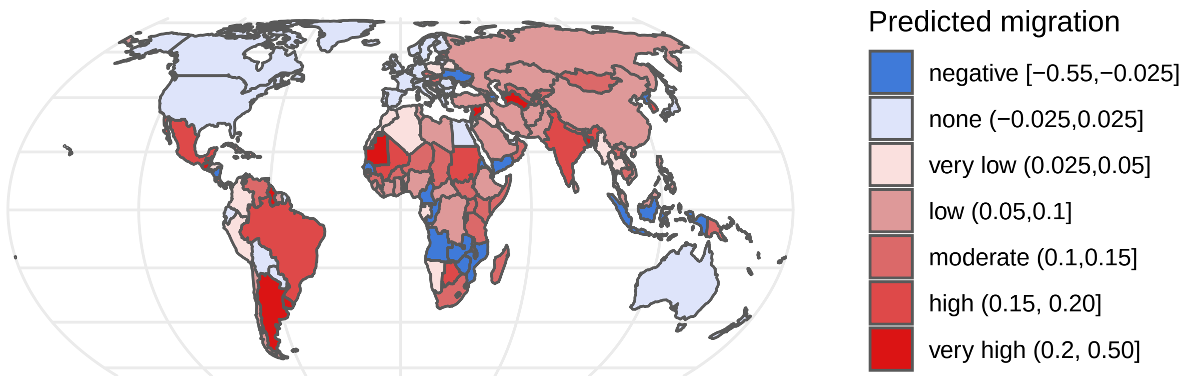


0.050
0.025
0.000
−0.025
−0.050
Predicted environmental effect
0%
25%
50%
75%
100%
% middle−income countries in sample


0.050
0.025
0.000
−0.025
−0.050
0%
25%
50%
75%
100%
% agriculturally dependent countries in sample



The last would be figure 4, which color-codes the countries of the world according to their predicted migration based on their historical exposure to environmental change from 1960 to 2000. This data is measured in standard deviations of the world distribution and their economic and sociopolitical characteristics measured in the year 2000. The latter is measured using income level, agricultural dependency, and conflict as in model 5 from Table 1.



My proposed extension to this figure would be converting this map which identifies hotspots of environmental migration based on data from 2000 into an animation that displays global out-migration trends over a longer time period. More than half of the studies incorporated in this meta-analysis contain data from years after 2000, with 16 of the 30 total studies including data up to 2010. In addition, half of the studies included data from before 1975. However, only 4 studies of the 30 had data from both prior to 1975 and later than 2000.

The selected country-level data, in terms of income-status identification for each country, whether the country was involved in a conflict, and whether a country is more of an agricultural exporter or importer, and their environmental change score in a particular year, could be calculated for any year.

An animated progression through the years based on the available data for each country in each year would be a fascinating extension of this project, providing not only an understanding of what regions' migration patterns are being affected by environmental changes and how, but also of how those patterns are developing year by year. This information is currently not conveyed by the original visualisation.

Unfortunately, the original authors of the paper did not include the data used to construct their countrydata dataframe (which is the 2000 snapshot). However, with the same data used to build that dataframe, similar ones could certainly be built for other years and allow for a time-lapse animation using the same mapping functions as well as a package like gganimate.

Conclusion

To replicate this project I consulted the code and data provided by the original authors at https://doi.org/10.7910/DVN/HYRXVV. I am grateful to the original authors for making this data and code publicly available.

Their data was well-organised and accessible, and their variables used in the model were correctly coded. The code provided ran easily with the specified packages and reproduced practically identical results, with very minor variations due to the chosen sampling method. The first figure I was unable to replicate, as the authors did not publish its code. However, the remaining tables and figures were replicable and thus I have high confidence that the first figure's data and code are reliable.

The authors created vectors of several combinations of key variables to plug into the custom functions func.felm and func.lmer that they created. The models used in the main section of the paper were all run using the func.felm function, producing fixed-effect linear models.

The results make sense: based on the data provided by the authors, the environmental changes, and in particular rapid-onset events, increase out-migration from the affected countries. This is corroborated elsewhere in literature not included in this meta-analysis, such as in:

Gori Maia, Alexandre, and Schons, Stella Zucchetti. "The Effect of Environmental Change on Out-migration in the Brazilian Amazon Rainforest." Population and Environment 42, no. 2 (2020): 183-218.

and

De Lellis, Pietro, Ruiz Marín, Manuel, and Porfiri, Maurizio. "Modeling Human Migration Under Environmental Change: A Case Study of the Effect of Sea Level Rise in Bangladesh." Earth's Future 9, no. 4 (2021): N/a.

To extend the original models I recommend creating an animated version of figure 4, a world map showing migration hot spots predicted based on factors including conflict, GDP per capita, and environmental changes. Such a map, if animated as a time-lapse, would provide a further dimension to understand the implications of the findings of this meta-analysis.

With the data for such an animation, the paper could also be extended to include a dichotomous outcome model such as a logit model by setting a threshold for when a country is deemed a "migration hotspot" as is used in the current map. The model could then be used to predict, based on the selected variables, whether a given country with a set of characteristics is likely to be an out-migration hotspot. This could even be converted into a multi-level model using annual progression as its level, as each migration score is related to its temporal neighbours.

As the output variable, stancoeff, represents the number of standard deviations from the global average out-migration likelihood a given country is expected to shift as a result of numerous variables including environmental changes and conflict and is as a consequence non-dichotomous and not counting, the chosen model (linear) suits the data better than alternatives like a Poisson or a logit model. The data, unfortunately, is almost all at the country level, so multilevel models don't match the data. This was tested by running several models using, separately, the region focused on in each study as levels, the model type used by the studies as levels, the type of environmental change as levels, and the individual studies as levels. None of the resulting standard deviations for the selected level-variable in the produced multilevel models were large enough relative to the residual standard deviation to justify their use as levels.

As such, continuing with the approach of the original authors, I produced a sixth linear model to extend the analysis. The env_type variable was inconsistently coded in the data, so I condensed the groupings (e.g. placing all types of environmental disasters under the umbrella label 'disaster', joining 'excess' and 'extreme' versions of the same phenomena, and ensuring that 'precipitation' and 'rainfall' were considered equal). I then added both the env_type variable which nuances the results by showing the effect of specific environmental changes or hazards on migration and the period_end variable to show how studies of these changes which take into account more recent events show varying average effects. The results of this model are below.

The model shows that, based on their t-values, that more recent data included in a particular study is likely to reduce the average estimated likelihood of migration from a country due to environmental changes by 0.001 for each year after 1990 included in the study. In addition, the results show that, disaster-level events aside, floods are the environmental change most likely to cause an increase in migration, increasing the predicted migration by an average of 0.014. Extreme temperature events also had a statistically significant effect on migration, interestingly decreasing the likelihood of migration.

Adding in the environmental types into the model also produces a curious effect for the results of the

Table 2:

	Dependent variable:
	stancoeff
	normal
pre_var	0.004* (−0.0005, 0.008)
rapidonset	−0.209** (−0.411, −0.008)
tem_lev	−0.191* (−0.392, 0.011)
tem_var	−0.210** (−0.411, −0.008)
env_lag_dum	0.0002 (−0.002, 0.003)
env_timespan_dum	−0.006*** (−0.008, −0.004)
env_other_dum	0.0001 (−0.002, 0.002)
internal	−0.003* (−0.006, 0.0003)
dest_low	0.031*** (0.018, 0.043)
dest_high	0.010*** (0.008, 0.013)
dest_ambi	0.010*** (0.005, 0.015)
L	−0.039*** (−0.054, −0.024)
LM	0.005 (−0.003, 0.014)
UM	0.043*** (0.030, 0.056)
agr	0.071*** (0.051, 0.090)
conflict_mepv_5	−0.026*** (−0.039, −0.014)
period_end_c	−0.001*** (−0.001, −0.001)
env_typedrought	−0.212** (−0.413, −0.010)
env_typeearthquake	0.004* (−0.0004, 0.009)
env_typeextreme rainfall	−0.196* (−0.399, 0.006)
env_typeextreme temperature	−0.017*** (−0.022, −0.011)
env_typeflood	0.015*** (0.008, 0.021)
env_typehurricane	−0.009 (−0.021, 0.003)
env_typelandslide	0.002 (−0.001, 0.006)
env_typemoderate temperature	0.001 (−0.017, 0.019)
env_typemoisture	
env_typestorm	0.0004 (−0.006, 0.007)
env_typetemperature shortage	
Constant	0.193* (−0.009, 0.394)
Observations	1,803
Log Likelihood	2,423.026
Akaike Inf. Crit.	−4,780.052
Note:	*p<0.1; **p<0.05; ***p<0.01
	Test

rapid-onset variable. Instead of being positive and having an average effect of 0.014 on the migration score, its effect in this model is now negative.

Bibliography

De Lellis, Pietro, Ruiz Marín, Manuel, and Porfiri, Maurizio. “Modeling Human Migration Under Environmental Change: A Case Study of the Effect of Sea Level Rise in Bangladesh.” Earth’s Future 9, no. 4 (2021): N/a.

Gori Maia, Alexandre, and Schons, Stella Zucchetti. “The Effect of Environmental Change on Out-migration in the Brazilian Amazon Rainforest.” Population and Environment 42, no. 2 (2020): 183-218.

Hoffmann, Roman, Dimitrova, Anna, Muttarak, Raya, Crespo Cuaresma, Jesus, and Peisker, Jonas. “A Meta-analysis of Country-level Studies on Environmental Change and Migration.” Nature Climate Change 10, no. 10 (2020): 904.

Meager, Rachel, “Understanding the Average Impact of Microcredit Expansions: A Bayesian Hierarchical Analysis of Seven Randomized Experiments” The American Economic Journal: Applied Economics, January 2019.