Evaluating the computational (“Big Data”) turn in studies of media coverage of climate change

Myanna Lahsen

Abstract

Machine-assisted big data (MABD) research is enabling quantitative studies of large-scale social phenomena, including societal responses to climate change. The rise of MABD science is causing both enthusiasm and concerns. Reviewing prominent criticisms of MABD and their relevance for MABD explorations of macro-structural factors shaping media coverage of climate change, this article finds that the quality and contributions of such studies depend on avoiding common pitfalls. The review focuses specifically on MABD studies’ attempts to identify and make sense of correlations—or lack thereof—between climate vulnerability and climate coverage in different countries. The review draws on insights from a single, nationally focused, context-attentive, and relatively more qualitative “small data” study in the Global South (Brazil) to shed critical light on assumptions, claims, and policy recommendations made based on the computer-assisted macro-studies. The review illustrates why more narrowly focused and qualitative small data studies are complementary and indispensable. Besides providing vital understanding of causal relationships that elude MABD studies, more narrowly focused and context-sensitive qualitative studies can foster understanding of the consequential mediating roles of place-specific meaning-making and political strategizing in how climate and weather phenomena are framed by social actors and mass media in particular places. These are dimensions that escape the Big Data quantitative methods, but that are vital to sound policy advice, as illustrated by the Small Data research from Brazil.

This article is categorized under:
Social Status of Climate Change Knowledge > Knowledge and Practice

KEYWORDS
big data, meaning-making, media coverage, small data, vulnerability

1 | INTRODUCTION

The new ubiquity of massive amounts of digital data has enabled a rapid rise in large-scale, quantitative “Big Data” (BD)1 research on social phenomena, assisted by computerized means of data analysis, including algorithms and
Machine Learning. The extension of human cognitive capacity offered by machine-assisted big data (MABD) analysis is widely expected to improve understanding of complex social phenomena (Abbasi et al., 2016; Graham & Shelton, 2013) and to help identification of solutions to environmental problems (Jensen & Campbell, 2019; Khakurel, 2018; Lahsen et al., 2020). High expectations of MABD-based science are bolstered by broader historical, cultural, and political valuations in science and policy which privilege quantitative approaches over other ways of analyzing reality (Daston & Galison, 1992; Porter, 1995; Shapin & Schaffer, 1985). Quantitative approaches also tend to endow social science with more prestige (Alexander, 2019). This further spurs the “computational turn,” even in fields which traditionally have centered on less comprehensive “Small Data” (SD) methods, such as ethnographic observation and close reading of texts (Bruns, 2013). Concerns are also being expressed about the quality and practical contributions of MABD-based science, however. In the social sciences, it is criticized and feared for stimulating a further scientistic, positivist, and quantitative turn, at the detriment of critical, qualitative, and postpositivist research (Graham & Shelton, 2013, p. 257), which arguably is more capable of accessing the inherently subjective, contested, and meaning-laden nature of social worlds (Alexander, 2019).

The growing use of MABD in social analysis begs critical assessment. This article reviews prominent critiques of MABD science and examines their relevance to MABD research exploring variations in national mass media coverage of climate change. Enabled by the availability of massive amounts of digital newspaper content, these MABD studies use computerized methods such as complex string-searches, algorithms, and machine learning to analyze tens and hundreds of thousands of articles communicating about climate change, searching for relationships between patterns in these communications and macro-level factors, such as levels of development, governance features, and vulnerability to climate change. This review centers on assumptions and claims made about vulnerability to climate change, evaluating their empirical basis in light of relevant insights from a single Brazil-focused, context-sensitive SD study of how purposeful political strategizing shapes framing choices about vulnerability, a meaning-making aspect that escapes the globally oriented MABD studies of media coverage of climate change (MCCC). Attention to correlations between patterns in MCCC and climate vulnerability reflects a wider interest in encouraging climate policy-supporting news coverage that adequately and repetitively identifies the human role where warranted, approaching extreme weather-related events and other expressions of climate change as opportunities to strengthen public awareness and climate policies (Lahsen et al., 2020; McNutt, 2019).

Considering the high expectations and rapid rise of MABD research in the social sciences, a guiding concern of this review is to evaluate contributions of such research to understanding of MCCC, and to do so considering (1) prominent criticisms of pitfalls in BD research and (2) the normative decisions and politics that shape Brazilian media coverage of climate change, as established by more qualitative SD methods.

2 COMPETING EVALUATIONS AND WORRIES ABOUT BD SCIENCE

Seeking to foster critical engagements and understanding of “big data,” scholars urge critical research to establish the relative strengths of BD versus SD approaches to knowledge production, including their respective abilities to access and consider contextual factors relevant to what they describe. In seminal work, and Crawford (2011) argue against the widespread acceptance of large datasets as unproblematic research resources. They raise a series of critical questions, including how an emphasis on BD might change what counts as knowledge (see also Bruns, 2013). boyd and Crawford highlight MABD-associated temptations, including tendencies toward “apophenia”—that is, seeing patterns where none actually exist, simply because massive quantities of data can offer connections that radiate in all directions”—and to wrongly trust that it is possible to see and understand “everything at a 30,000-foot view” (p. 663). Rather than drown out noise introduced by biases in any given media platform’s population or data, as commonly believed (Ruths & Pfeffer, 2014), large datasets can obscure more than reveal the complexity of social processes (Graham & Shelton, 2013).

Prominent science policy analyst, Daniel Sarewitz, adds to the above warnings. He worries that the BD trend in science yields large volumes of false results and uninsightful science, driving “science into a sea of data with few constraints on where it might drift,” (Sarewitz, 2016, p. 33). He writes: “If mouse models are like looking for your keys under the streetlamp, BD is like looking all over the world for your keys because you can—even if you don’t know what they look like or where you might have dropped them or whether they actually fit your lock” (Sarewitz, 2016, p.32). The number of possible causal links between variables is hugely greater than the few that a scientist might conceive of and be able to test. Moreover, scientists can sift through vast volumes of peer-reviewed published papers and “cherry-
pick” cooperating findings that lend credibility even to erroneous findings (Sarewitz, 2016, p. 31). Misinterpreting correlations, researchers risk claiming connections where none exist, or where other variables are more important but not identified:

Researchers have identified more than a hundred variables that may influence obesity, from genes to education to job stress to how fast you eat to whether you were breastfed. Exploring the relations between even some small number of combined variables would give you billions of possible hypotheses to test. The likelihood that you will happen upon one that reveals important causal relations is thus extremely small, while the opportunities for introducing bias or discovering meaningless correlations abound. So even if you get a positive result, it will likely be spurious. [... ] Perhaps obese people with less education, more job stress, and a specific genetic marker do eat more quickly than others, but the reason they are obese may be due to something else entirely, say, not having time to exercise because they live far from work (Sarewitz, 2016, pp.31-32).

Ioannidis’ (2005, p. 0696) influential meta-scientific analysis confirms that drawing conclusions on the basis of formal statistical significance (typically for a p-value less than 0.05) in a single study is a fallacy, since “almost everything,” including minuscule effects, can seem significant when searching very large datasets (Abbasi et al., 2016, p. xxiii; George et al., 2014, p.323). Ioannidis concludes that most published research does not meet good scientific standards of evidence and can be proven false (Ioannidis, 2005).

Arguing that insightful science requires starting from robust understanding of causal relationships, Ioannidis, Sarewitz, and other critics propose more rigorous causal analysis in the initial phases, to select for deeper investigation variables that are more likely to be relevant and significant, both quantitatively and practically; the true test of significance is practical significance, they argue (Abbasi et al., 2016, p. xiii; Chatfield, 1995, p.70; Lin et al., 2013, p. 906). Sarewitz provocatively proposes that science needs to be “saved” from the spread of false and practically insignificant science encouraged by BD (Sarewitz, 2016). Backing his concerns, a study found that nearly half of 100 scientific BD-based studies supported hypotheses exclusively on the basis of low p-values and the sign of a regression, and nearly half of them failed to discuss practical significance (Lin et al., 2013).

3 | COMPARATIVE BD STUDIES’ FINDINGS ABOUT CLIMATE VULNERABILITY-NEWS COVERAGE CORRELATIONS

This section reviews findings from three exceptionally comprehensive internationally comparative MABD “global” studies of the relationship between national climate vulnerability and national climate coverage: Schmidt et al. (2013), Barkemeyer et al. (2017), and Vu et al. (2019). After briefly summarizing some of their broader features and findings, I specifically evaluate assumptions and claims made in the studies about the relationship between MCCC and vulnerability. In doing so, I draw on the above alerts about the minefields of overemphasis on correlations over causal understanding, and on statistical significance over practical significance.

3.1 | Selection criteria

The three studies were selected for the following shared features:

1. They involve BD in the form of over 10,000 newspaper articles from dozens (27 and more) countries, obtained from searches via databases such as LexisNexis or Factiva. Only three such comprehensive studies were found. While not literally global, Tindall et al. (2018) refer to them as such due to their exceptional comprehensiveness among studies of MCCC.

2. They deploy computerized methods to identify and make sense of relevant content.

3. They explore the content for correlations between macro socio-economic factors and MCCC.

Schmidt et al. and Barkemeyer et al. only attended to issue attention (based on overall volume of climate coverage), whereas Vu et al. also probed media content, assisted by machine learning.
Table 1 below compares key features of the three macro-studies and summarizes their findings about the correlation between vulnerability and MCCC.

### 3.2 The common correlation hypothesis

The three macro-studies’ authors all hypothesize that more vulnerable countries—most of which are in the global South—focus relatively more media coverage on climate change. Schmidt et al. (2013, p.1238–1239) articulate the reasoning: Vulnerability to negative consequences of climate change likely increases climate related activities as well as media coverage. They posit that more direct and higher personal and material damage increases the urgency of societal responses and, thus, interest in climate change and its impacts, thereby pushing up coverage of the topic as societal actors and media seize on these “opportunities for climate change awareness raising activities” and seek knowledge about outcomes of climate negotiations and options for enhancing societal resilience. Reflecting similar logic, Vu et al. (p. 2) expect that news coverage should increase in volume in step with the strengthening of scientific evidence of human-induced climate change. As for Barkemeyer et al., they note the absence of any “simple relation between the amount of media coverage and the legitimation or implementation of ambitious climate policies.” However, they assume—and judge, normatively—that news media should show higher levels of climate coverage to facilitate climate action, arguing that levels of media attention “serve as a precondition for any type of widespread support” for “measures that effectively address climate change” (p. 1047). Similarly, Schmidt et al. consider evidence of high levels of coverage in countries with great responsibility for climate action to be “a positive sign” for international climate policy (Schmidt et al., 2013, pp. 1245–1246).

### 3.3 Key findings

The three sets of authors did not find the hypothesized correlations between climate vulnerability and media coverage with any consistency. Analyzing 152,125 news articles from 27 countries for the years 1996–2010, Schmidt et al. (2013) found that, on average, media attention to climate change was even slightly lower in countries marked by high, severe, or acute vulnerability according to formal risk and vulnerability indices. Media attention in particularly vulnerable countries was clearly higher, by contrast, confirming their hypothesis. The largest of the studies reviewed here, Barkemeyer et al. (2017) analyzed 2.6 million articles in elite newspapers from 41 countries during 2008. They found national exposure to climate change to be an unreliable determinant of levels of news coverage. Vu et al. (2019) analyzed 37,670 news articles from 45 nations over a five-year period (2011–2015). They found climate severity to be a statistically significant predictor of a natural impact focus in the MCCC (p. 6); richer countries’ coverage focused

| Study                        | Number of countries | Unit of analysis | Number of articles | Years in focus | Positive correlation between levels of vulnerability and climate coverage? |
|------------------------------|---------------------|------------------|--------------------|----------------|-------------------------------------------------------------------------|
| Schmidt et al., 2013         | 27                  | Volume of MCCC   | 152,125            | 1996–2010      | Slightly lower correlation in more vulnerable countries; correlation found within non-Annex B countries (non-Annex B countries are countries that were not included in Annex B of the Kyoto Protocol and thus did not have binding commitments.) |
| Barkemeyer et al., 2017      | 41                  | Volume of MCCC   | 2.6 M              | 2008           | No; insignificant correlation                                              |
| Vu et al., 2019              | 45                  | Volume of MCCC   | 37,670             | 2011–2015      | Yes: climate impact-focused coverage correlates positively with higher climate vulnerability |

Table 1

| Study                        | Number of countries | Unit of analysis | Number of articles | Years in focus | Positive correlation between levels of vulnerability and climate coverage? |
|------------------------------|---------------------|------------------|--------------------|----------------|-------------------------------------------------------------------------|
| Schmidt et al., 2013         | 27                  | Volume of MCCC   | 152,125            | 1996–2010      | Slightly lower correlation in more vulnerable countries; correlation found within non-Annex B countries (non-Annex B countries are countries that were not included in Annex B of the Kyoto Protocol and thus did not have binding commitments.) |
| Barkemeyer et al., 2017      | 41                  | Volume of MCCC   | 2.6 M              | 2008           | No; insignificant correlation                                              |
| Vu et al., 2019              | 45                  | Volume of MCCC   | 37,670             | 2011–2015      | Yes: climate impact-focused coverage correlates positively with higher climate vulnerability |
relatively less on natural impacts. Rather than merely relying on overall levels of climate coverage for their analysis—a feature the two other sets of authors also recognize as limited—their finding was based on computer-assisted content analysis of a random and statistically significant subset of the news articles.

Focused on only 15 countries plus Taiwan and using less automated methods, Broadbent et al.’s (2016) study does not meet the first and second selection criteria used in this review. As a fourth “global level” study (Tindall et al., 2018), and because of its findings concerning Brazil and other countries of the Global South, its results about MCCC-vulnerability correlations nevertheless merit description here. Analyzing 18,721 newspaper articles from the years 2007–2008 and applying a standardized, manual coding scheme to characterize thematic foci and other variables, Broadbent et al. found impacts of climate change to be among four factors that most influence variation in media foci and framings across the 16 cases during the years in focus (2007–2008). Like the other three “global” studies, their findings did not yield easy supra-national generalizations, however. Of the five relatively less developed countries in Broadbent et al.’s analysis, three (India, Mexico, and Taiwan) confirmed the hypothesis that developing countries’ greater vulnerability to negative climate impacts intensifies their MCCC. However, China and Brazil – two significantly climate vulnerable countries in their study – did not have higher MCCC relative to less vulnerable nations. The authors posit that Chinese newspapers’ deemphasis might be a function of political party control intended to deter popular protest (p. 10). They could not explain Brazil’s divergence. Brazil is geographically and culturally relatively proximate not only to Mexico but also—and even more so—to Peru, where an additional study using the same methods found a strong focus on climate impacts in the national coverage (Takahashi & Meisner, 2012).

Thus concluding from their explorations that levels of MCCC are not exclusively influenced by vulnerability, the authors of the three macro-studies in focus explore other possible factors, in partially overlapping manner. They all investigate correlations between national levels of MCCC and economic development (GDP), finding significant correlation. Usually associated with higher levels of development and GDP, however, national carbon dependency correlated positively with higher levels of coverage in Schmidt et al.’s study but not in that of Barkemeyer et al., who found stronger correlation with regulatory quality. Barkemeyer et al. also posit the likely role of lower journalistic capacity in poorer countries. They base this assumption on claims to that effect in smaller-scale, more qualitative literature on MCCC (Billett, 2010; Boykoff, 2011, p. 24; Shanahan, 2009). Schmidt et al.’s study, published four years earlier, did not find evidence of any “information gap” (p. 1245) in poorer countries. While both sets of authors thus raise the issue and possible impact of information gaps, it is left unspecified what evidence and methods might confirm the existence of the hypothesized gap and its causes and consequences.

The three sets of authors agree that policy contexts and governance factors are important determinants of MCCC, not least the well-established MCCC-instigating role of international climate negotiations and associated scientific assessment reports (Boykoff, 2011; Broadbent et al., 2016). Schmidt et al. and Barkemeyer et al. find that strength of emission reduction commitments under international policy instruments correlate positively with higher MCCC. Barkemeyer et al. explore the role of regulatory quality by means of a World Bank macro-structural indicator, a single, aggregated indicator supposedly capturing “perceptions of the ability of the government to formulate and implement sound policies and regulations that permit and promote private sector development” (p. 1035). Vu et al. examine correlations between MCCC and governance effectiveness using a World Bank aggregated indicator centered on “aspects of governance including civil service and government independence from political pressures, policy formulation and implementation, and government’s commitment to such policies” (p. 5). Statistically, they found it to correlate significantly with levels of MCCC, but that GDP per capita correlated more strongly, contrasting Barkemeyer et al.’s finding.

Barkemeyer et al. were alone in the admirable but difficult attempt to identify the impact of cultural factors such as norms, ideologies, and beliefs on MCCC. Acknowledging the difficulty of operationalizing culture as factor, they chose to use CIA Fact Book data about national distributions of five major religious denominations. Overall, they did not find significant correlation of these factors (“culture”) with higher levels of climate coverage.

4 | ANALYSIS

How do the three analyses navigate the minefields reviewed in Section 2, particularly overemphasis on correlations and statistical significance over causal understanding and practical significance? What is the quality of the reviewed studies’ findings?

The reviewed “global” studies are commendable as exercises and first stabs at generating knowledge. Their attention to so many country contexts simultaneously, to macro-level factors, mere correlations, and factors of unclear importance and interconnections, limits understanding and confidence in the results, which also are partly contradictory
across the studies. Understanding of how socioeconomic, political, and environmental factors bear on MCCC, and vice versa, remains elusive on the basis of these studies. With regards to the correlation between MCCC and climate vulnerability specifically, innumerable factors might increase MCCC, among them national wealth and economically consequential commitments under international treaties. The authors acknowledge this and other limitations of their studies’ ability to yield understanding of sub-national differences in media framing and the causes of the observed international patterns. They call for studies that can do so, including traditional content analysis. Schmidt et al. are especially explicit about the need for closer studies, including traditional (i.e., non-automated) content analysis of how policy options are presented and perceived, and of blame politics (what institutions or countries might bear responsibility for action, p. 2046).

The various sets of authors do not go to equal lengths to justify their chosen macro-structural variables and the robustness, limitations, and causal relevance of the indicators used to explore them, such as those obtained from the World Bank. Barkemeyer et al. and Schmidt et al. are careful to justify their chosen macro-structural variables and the robustness, limitations, and causal relevance of the indicators used to explore them. Schmidt et al. dedicate two pages (pp. 1237–1239) to specification of their chosen macro-structural variables. As noted, Barkemeyer et al. acknowledge the difficulty of finding data about cultural factors and they are transparent about the tenuous scientific basis for their choice to use data about religious denominations to explore the role of culture and environmentalism in international variations in MCCC. Barkemeyer et al.’s choice of religious denomination to explore the role of culture nevertheless illustrates how data sometimes is used to explore questions that it is not obviously apt to answer.

In a marked example of lack of rigor that sets it apart, Vu et al. do not discuss how they expect media freedom to impact media coverage. They merely posit—in vague terms—that media freedom is “either directly or indirectly related to how climate change may be portrayed in the media” (p. 4). The article that they cite as support (Baettig et al., 2007) bears no relevance to the topic, and in fact does not even mention media.

The scientific and practical value of the statistical explorations of correlations between levels of MCCC and vulnerability also needs to be critically evaluated. Common linking of extreme events to climate change should, logically, increase MCCC in countries where such events are frequent (i.e., vulnerability is high). However, there is little established knowledge about whether, when, and where more frequent discursive linking of weather-related extreme events intensifies climate policy action, despite the authors’ hypotheses. This prevents the solid understanding of causal relations that Sarewitz and others urge should precede MABD explorations of correlations, as well as a sound basis for making normative and policy judgments about the lack of higher levels of MCCC in countries that are highly vulnerable.

Despite their studies’ important limitations, the three sets of authors hypothesize about causal relations based on correlations, and they sometimes appear to attribute considerable ontological status to their findings. For example, Schmidt et al. purport to study “the influence of weather and climate characteristics as well as various social events and feedbacks on issue attention,” rather than just correlations (Schmidt et al., 2013, p. 152). Barkemeyer et al. claim to improve “understanding of the determinants of climate change-related media attention across a diverse cross-country sample” (Barkemeyer et al., 2017, p. 19, emphasis added). Moreover, using the word “impacts” when presenting findings of mere correlations between MCCC and macro factors that they explore, Barkemeyer et al. misleadingly imply having established causal relations and not merely correlations (“Climate Risk Index scores are found to have significantly positive impacts on media attention to climate change” and “No significant impact is found throughout for Climate Change performance” p. 1044). Vu et al. move seamlessly from identification of statistically significant relationships to language describing their findings as if these were demonstrations of actual causal relationships. For example, they argue that their finding of a statistically significant correlation between natural impacts-focused MCCC and the World Bank indicator for government effectiveness “demonstrates” that “in countries with an effective government, climate change is less likely to be framed in the light of natural disasters” (p. 6). These slippages are reminiscent of those observed in how global circulation modelers in the Earth sciences talk and write about their simulated output (Lahsen, 2005).

Furthermore, two of the three sets of authors move from their explorations of correlations—as we have seen, without clear causal understanding—to issuing questionable policy claims and recommendations. Vu et al. claim that “Practically, [their] results could be used as a reference in designing national campaigns to reduce greenhouse gases” (p. 1). They do not specify which results, exactly, nor how these results are practically relevant and useful. Barkemeyer et al. claim that, policy-wise, “important implications emerge from the absence of a significant relationship between a country’s exposure to climate change-related risk and levels of climate change-related media coverage” (Barkemeyer et al., 2017, p.19). They conclude that their findings “appear to confirm” studies which suggest that such disassociation
of risk and climate coverage is caused by cognitive dissonance, defined as a “failure to make the link” between near-term events and the more complex, distant in time and less tangible phenomenon of climate change (Barkemeyer et al., 2017, p. 19). They move from this already tenuous set of causal assumptions to suggest that the (implicitly global) “pivotal challenge” for policymakers is to “educat[e] the media as well as the general public about the link” (ibid.). The section that follows reveals the limited validity and appropriateness of this decontextualized global assumption and the implied policy recommendation.

A close, nonautomated, partly qualitative SD content analysis (Lahsen et al., 2020) of Brazilian MCCC during the years 2007–2008 and 2013 defies assumptions that underpin Barkemeyer et al.’s analysis, as well as their education-centered policy recommendation. The study makes sense of a puzzling contrast: On the one hand, (1) surveys spanning decades have shown Brazilians to be among the most climate concerned populations in the world. On the other hand, (2) the MCCC shows clearly that Brazilians—led by environmental activists, climate experts, and a few climate-concerned politicians—are disinclined to link national disasters from flooding and landslides to climate change. The analysis suggests that their disinclination reflects strategic calculations about policy opportunities and about trade-offs between action on climate change, forest conservation, and disaster prevention (Lahsen et al., 2020). Due to limited opportunity for new national climate policy measures at the time of the study in focus, for reasons the authors explore and explain, even environmentalists and their supporters tended strongly toward nonclimatic framing strategies and policy arenas to advance issues of concern, including climate-relevant disaster preparedness and forest conservation.

Moreover, framing climate change as the cause of tragic, recurring disasters went against desires to foreground national decisionmakers’ responsibility for societal vulnerabilities that transform extreme events into disasters. Lahsen et al.’s close, SD, nonautomated and partly qualitative content analysis of Brazilian MCCC revealed that government officials responsible for disaster preparedness were inclined to emphasize climate as cause of extreme events—and that their framings were seen as attempts to evade responsibility and therefore subject to vigorous push-back from civil society (Lahsen et al., 2020). Extreme weather events and ensuing disasters are a long-standing problem in Brazil independently of human-induced climate change, and poor disaster planning explains a great part of the societal vulnerability that the disasters reveal. Civil society actors contest framings that deflect national responsibility for action by directing attention toward the more distant and loosely connected international climate policy forums, perceiving that government actors tend to seek to deflect responsibility. In these forums, moreover, Brazilian diplomats historically emphasize Northern primary responsibility for causing the problem and for addressing it - a framing that is hegemonic and thus difficult to contest in Brazil, as apparent in Lahsen et al.’s close content analysis of three prominent Brazilian newspapers’ climate coverage in 2007–2008 and 2011 (Lahsen et al., 2020, pp. 221–223).

To the extent that Brazilian government officials greatly dominate policy discussions in Brazilian MCCC, as Lahsen et al. found to be the case, a high correlation between climate vulnerability and MCCC could theoretically reflect the dominance of their discursive attempts to displace responsibility for action to the international level, as opposed to foster proactive climate-relevant action, including disaster preparedness. Such complexities, strategizing, and power dimensions escape the global studies reviewed here, and they are key to sound policy recommendations and disaster framing (see also Lahsen & Ribot (2021) on about the politics of attributing extreme events and disasters to climate). Rather than campaigns to raise public awareness about climate change, for example, as Barkemeyer et al. recommend, Lahsen et al.’s (2020) context-attentive SD content analysis of MCCC suggests that Brazilian environmental leaders—politicians, activists, and scientists alike—prefer to see funds spent in more direct manner to protect national forests, improve disaster-related decision-making, overcome governmental inertia, and reduce systemic inequalities that also heighten vulnerability to weather and climate.

5 | CONCLUSION

New informational and computational capacities are spurring social science research in direction of more comprehensive, machine-enabled analyses. These new capacities are also cause for concerns; their implications for science, and for climate knowledge, are insufficiently charted.

MABD offers many benefits—many more than mentioned here or even yet known. This review does not account for the full range of contributions that the reviewed MABD studies make to climate research; it focuses on only one aspect, namely how they cover the relationship between MCCC and climate vulnerability. In this respect, the reviewed MABD studies’ authors show awareness of their own studies’ limitations, but also some instances of the fallacies to which “meta-science” critics such as Ioannidis and Sarewitz call attention. The authors of the three main BD studies
reviewed acknowledge limitations of their studies. But some of them took greater care than others to ensure rigor and minimize the weaknesses and pitfalls reviewed in Section 2.

The important conclusion is that bigger data are not always better data, and that SD and BD studies are complementary, each offering means to keep the other in check. MABD can extend human analytic capacities, but it is vital to also keep expectations in check, not least given common misconceptions that more data necessarily translates into enhanced rigor and reliability by supposedly drowning out noise and bias. On the presumption that all data is of the same nature, only more and less aggregate, some researchers assume that analyses done with SD can be done better with BD (boyd & Crawford, 2011, p. 8). Taken out of context, data—big or small—lose meaning and value, however; context matters. Merits aside, BD research can be difficult to replicate and compare, insufficiently insightful, besides tempting overclaims and apophenia. Confirming Sarewitz’ observation, this review shows that many more complex combinations of causes shape levels and foci of climate coverage around the world than the reviewed MABD studies access. The flip side of BD international comparisons tends to be decontextualization, oversimplification, and attention to mechanical relationships and mere correlations.

SD studies, in turn, can also be insufficiently insightful and significant. They can be overly myopic and yield overly fragmented knowledge, preventing (sound) generalizations. On the other hand, at least in some respects qualitative SD studies may be more conducive to understanding of human meaning-making and, by force of that, key aspects of social phenomena. Wikipedia suggests that by force of being smaller in volume and format, SD is more conducive to human comprehension and, thereby, more “accessible, informative and actionable.” Whether or not that is the case, at the least geographically more narrowly focused, meaning- and politics-attentive analyses are a vital complement to macro-studies of the sort reviewed here. SD studies can yield needed insight into connections between knowledge-making and cultural aspects, not least qualitative, context-sensitive case studies. Such studies need to inform broad, quantitative analyses because they allow recognition of ambiguities, gaps, and blind spots in understandings of the world’s complexity (Hulme, 2010)—and such understanding is crucial to sound policy recommendations. The Brazilian qualitative case study is limited in scope and methods in its own way, but it illustrates that many other variables need to be studied and understood before sound normative conclusions and policy recommendations can be drawn about statistically significant patterns in MCCC, including conclusions about the role of climate vulnerability and recommendations about what policy makers should do and how media should frame phenomena such as extreme weather events to strengthen climate action in any given context. Meaning- and politics-attentive analysis of contextual particularities is indispensable, whether or not science indicates a statistically significant link between extreme events and climate change, and whether or not a correlation is found between climate vulnerability and MCCC. The authors of the three main BD studies reviewed here indicate awareness of this. For instance, Schmidt et al. call for content analysis in further studies, using both automated and conventional methods (p. 2046).

The reviewed MABD studies did not access signification processes that matter centrally for understanding of climate coverage, and that should inform policy recommendations. The case study of Brazil underscores the importance of sustaining and valuing not only both BD and SD research in more equal manner, but also meaning-focused research, in particular. “Surveillance capitalism” shows the supreme power of MABD for understanding and intervening in cognition (Zuboff, 2019); there is nothing inherent about BD studies that prevents them from accessing cognitive processes, perhaps even to the contrary. Highly context-dependent, implicit political strategies and policy goals that shape framing choices and media coverage may not be as easily accessible through BD, however. At any rate, it of course matters crucially what BD sources are chosen and what—and how—questions are explored in any given study.

This review thus supports boyd and Crawford’s emphasis that “Research insights can be found at any level, including at very modest scales. In some cases, focusing just on a single individual can be extraordinarily valuable... The size of data being sampled should fit the research question being asked: in some cases, small is best” (2011, p. 8). This review also supports the recommendation (e.g., [Sivarajah et al., 2017]) for greater reliance on qualitative case studies and survey-based quantitative studies, where appropriate, to strengthen MABD research.

The values of comprehensiveness and the intensifying quantification capacities must not cause devaluation and defunding of equally needed, deeper cutting small data analyses or theoretical work. As sociologist Jeffrey Alexander (Alexander, 2019) recently commented about contemporary social science, it looks too “longingly to what it imagines as the explanatory perfections and achievements of the natural sciences,” in ill-guided manner increasingly striving to catch up with them. Social scientists should accept that social phenomena often elude the explanatory scope and universal reach possible for macro-studies and matter-focused natural science research. They should resist tendencies to “queue to become a science” (ibid, p. 44) to the extent that this devalues (and thus discourages and limits the space for) vitally needed scholarship into the distinctive political, emotional, and cognitive aspects of social worlds that
fundamentally shape status quo and the possibilities for change - social science contributions focused on signification, on meaning-making, which constitute a distinct and highly valuable contribution of the social sciences. Privileging quantitative approaches in the social sciences obscures the distinctive ontology of the social world, Alexander (ibid.) underscores:

The social world is obdurate but subjective; its structures are fueled by interpretation; its so-called laws are actually norms re-instituted time and again, dramatized every moment of every day. The “realism” of society and its investigation are achieved and performed; they are not naturally there (p. 44).

It is vitally important to teach rigor and ethics in BD research (as in all research), and in uses of machine-learning, to harness their powerful potential to help societies reach progressive political and environmental goals (Lahsen, 2020; Luers et al., 2020; Nowak et al., 2018). Another key role for specialists in the age of MABD research, however, is “to contextualise and offer insights into what our data do, and maybe more importantly, don’t tell us” (Graham, 2012). Comprehensive, quantitative BD studies do not protect against subjective belief. Assumptions that they do, and that qualitative methods are necessarily more subjective and less valuable diminish and discourage scholarship that might improve our ability to grasp and address the decisively important “social heart” (Hackmann et al., 2014) of existential threats facing humanity in the 21st Century.

ACKNOWLEDGMENT
Thanks to dear friends and colleagues who read and commented on versions of this manuscript. Thanks also to reviewers for astute reviews and helpful suggestions for its improvement.

CONFLICT OF INTEREST
The author has declared no conflicts of interest for this article.

DATA AVAILABILITY STATEMENT
Data sharing is not applicable to this article as no new data were created or analyzed in this study.

ORCID
Myanna Lahsen © https://orcid.org/0000-0001-5225-2048

ENDNOTES
1 Inherently relative and undefinable in absolute numeric terms, Big Data is data that is “huge in volume, high in velocity, diverse in variety, exhaustive in scope, fine-grained in resolution, relational in nature, and flexible in trait” (Kitchin, 2014).
2 An algorithm is a set of instructions to be executed at the encounter of a defined trigger. A subset of artificial intelligence, machine learning involves multiple algorithms and their modification, as well as creation of new algorithms, in adaptive response to inputs and data. It involves “learning” from data without relying on rules-based programming. Statistical modeling is rules-based, by contrast; it is a formalization of relationships between variables in the data in the form of mathematical equations.
3 Tindall et al. mention only three “global level” studies. Confirming that few comprehensive (“global”), macro-level studies exist, I found only four such studies, counting the publication of Vu et al., which was published after Tyndall et al.’s review. One of these studies was that by Broadbent et al. (2016), which is discussed in the next section titled “Key findings.” As further confirmation that few “global” studies exist, my search in the references lists of the four mentioned studies also did not reveal additional studies fitting this description, nor did a search in the more recent (2020) three-country analysis of newspaper climate coverage by Shea et al. (2020), nor my examination of articles that cite Barkemeyer et al.’s, 2017 article before the year 2021 (I found these searching for articles that cite their article between January 2018 and 31 December 2020, according to Google Scholar).
4 Schmidt et al. used complex, broadly-defined search strings to perform full-text searches in electronic databases, but performed manual checks of each article to ensure relevance (p. 1240). Barkemeyer et al. used econometrics and perform software-facilitated text mining that uses stemming algorithms to find relevant media coverage of climate change (henceforth, “MCCC”) content via LexisNexis. Vu et al. used a computerized algorithmic technique to perform content-analysis of the initial set of 37,670 articles.
5 See http://daraint.org/climate-vulnerability-monitor/climate-vulnerability-monitor-2012/, accessed 6 May 2016.

6 Vu et al. compiled this factor from four annual figures in each of the affected countries: (1) number of natural disasters, (2) death tolls, (3) number of affected people, and (4) financial loss caused by those disasters.

7 Vu et al. coded articles that approached the topic of “natural impacts,” subsuming in this category a wide range of relevant phenomena. These included disasters ensuing from extreme weather but also phenomena such as ice or glacier melting, temperature, glaciers, sea-level, habitats, species loss.

8 Disclosure: I am co-author of Broadbent et al. (2016) and contributed the Brazilian data analyzed in the article.

9 Vu et al. explore “media system,” but their indicator for media freedom is not obviously linked to capacity issues.

10 https://en.wikipedia.org/wiki/Small_data accessed 22 February 2021

RELATED WIREs ARTICLES

Media(ted)discourses and climate change: a focus on political subjectivity and (dis)engagement
Political economy, media, and climate change: sinews of modern life
Attribution of extreme weather and climate-related events

REFERENCES

Abbasi, A., Sarker, S., & Chiang, R. H. L. (2016). Big data research in information systems: Toward an inclusive research agenda. Journal of the Association for Information Systems, 17(2), 3–XXXII.

Alexander, J. C. (2019). What social science must learn from the humanities. Sociologia & Antropologia, 9(1), 43–54.

Baettig, M. B., Wild, M., & Imboden, D. M. (2007). A climate change index: Where climate change may be most prominent in the 21st century. Geophysical Research Letters, 34(1), L01705.

Barkemeyer, R., Figge, F., Hoepner, A., Holt, D., Kraak, J. M., & Yu, P.-S. (2017). Media coverage of climate change: An international comparison. Environment and Planning C: Politics and Space, 35(6), 1029–1054.

Billett, S. (2010). Dividing climate change: global warming in the Indian mass media. Climatic Change, 99(1–2), 1–16.

Boyd, D. & Crawford, K. (2011). Six provocations for big data. A Decade in Internet Time: Symposium on the Dynamics of the Internet and Society. Oxford Internet Institute Oxford. p. 21.

Boykoff, M. T. (2011). Who speaks for the climate?: Making sense of media reporting on climate change. Cambridge University Press.

Broadbent, J., Sonnett, J., Botetzagias, I., Carson, M., Carvalho, A., Chien, Y.-J., Edling, C., Fisher, D., Giouzepas, G., Haluza-DeLay, R., Hasegawa, K., Hirschi, C., Horta, A., Ikeda, K., Jin, J., Ku, D., Lahsen, M., Lee, H.-C., Lin, T.-L. A., … Zhengyi, S. (2016). Conflicting climate change frames in a global field of media discourse. Socius, 2, 1–17.

Bruns, A. (2013). Faster than the speed of print: Reconciling ‘big data’ social media analysis and academic scholarship. First Monday, 18(10), 1–5.

Chatfield, C. (1995). Problem solving: A statistician’s guide (2nd ed.). CRC Press.

Daston, L., & Galison, P. (1992). The Image of Objectivity. Representations, 40, 81–128.

George, G., Haas, M. R., & Pentland, A. (2014). Big data and management. Academy of Management.

Graham, M. (2012, 9 March). 'Big data and the end of theory?’, The Guardian. https://www.theguardian.com/news/datablog/2012/mar/09/big-data-theory.

Graham, M., & Shelton, T. (2013). Geography and the future of big data, big data and the future of geography. Dialogues in Human Geography, 3(3), 255–261.

Hackmann, H., Moser, S. C., & St. Clair, A. L. (2014). The social heart of global environmental change. Nature Climate Change, 4(8), 653–655.

Hulme, M. (2010). Problems with making and governing global kinds of knowledge. Global Environmental Change, 20(4), 558–564.

Ioannidis, J. P. A. (2005). Why most published research findings are false. PLoS Medicine, 2(8), e124.

Jensen, D. & Campbell, J. (2019). The case for a digital ecosystem for the environment: bringing together data, algorithms, and insights for sustainable development. Discussion Paper. Science Policy Business Forum. UNEP, https://un-spbf.org/wp-content/uploads/2019/03/Digital-Ecosystem-final.pdf.

Khakurel. (2018). The rise of artificial intelligence under the lens of sustainability. Technologies, 6(100), 1–18.

Kitchin, R. (2014). The data revolution: Big data, open data, data infrastructures and their consequences. Sage.

Lahsen, M. (2005). Seductive simulations? Uncertainty distribution around climate models. Social Studies of Science, 35(6), 895–922.

Lahsen, M. (2020). Should AI be Designed to Save Us From Ourselves? Artificial Intelligence for Sustainability. IEEE Technology and Society Magazine, 39(2), 60–67.

Lahsen, M. & Ribot, J. (2021). Politics of attributing extreme events and disasters to climate change. WIREs Climate Change. https://doi.org/10.1002/wcc.750.

Lahsen, M., de Azevedo Couto, G., & Lorenzoni, I. (2020). When climate change is not blamed: the politics of disaster attribution in international perspective. Climatic Change, 158(2), 213–233.
Lin, M., Jr, L., Henry, C., & Shmueli, G. (2013). Research commentary—too big to fail: Large samples and the p-value problem. *Information Systems Research, 24*(4), 906–917.

Luers, A., Garard, J., St. Clair, A. L., Gaffney, O., Hassenboehler, T., Langlois, L., Mougeot, M., & Luccioni, S. (2020). Leveraging digital disruptions for a climate-safe and equitable world: The D2S agenda. *IEEE Technology and Society Magazine, 39*(2), 18–31.

McNutt, M. (2019). Time’s up, CO2. *Science, 365*(6452), 411.

Nowak, A., Lukowicz, P., & Horodecki, P. (2018). Assessing artificial intelligence for humanity: Will AI be the our biggest ever advance? or the biggest threat [opinion]. *IEEE Technology and Society Magazine, 37*(4), 26–34.

Porter, T. (1995). *Trust in numbers*. Princeton University Press.

Ruths, D., & Pfeffer, J. (2014). Social media for large studies of behavior. *Science, 346*(6213), 1063–1064.

Sarewitz, D. (2016). ‘Saving science’. *The New Atlantis, 49* (Spring/Summer ). pp. 4–40.

Schmidt, A., Ivanova, A., & Schäfer, M. S. (2013). Media attention for climate change around the world: A comparative analysis of newspaper coverage in 27 countries. *Global Environmental Change, 23*(5), 1233–1248.

Shanahan, M. (2009). ‘Time to adapt? Media coverage of climate change in nonindustrialised countries. In *Climate change and the media* (pp. 145–157). Peter Lang.

Shapin, S., & Schaffer, S. (1985). *Leviathan and the air pump: Hobbes, Boyle, and the Experimental Life*. Princeton University Press.

Shea, M. M., Painter, J., & Osaka, S. (2020). Representations of Pacific Islands and climate change in US, UK, and Australian newspaper reporting. *Climatic Change, 1-20*, 89–108.

Sivarajah, U., Kamal, M. M., Irani, Z., & Weerakkody, V. (2017). Critical analysis of Big Data challenges and analytical methods. *Journal of Business Research, 70*, 263–286.

Takahashi, B., & Meisner, M. (2012). Climate change in Peruvian newspapers: The role of foreign voices in a context of vulnerability. *Public Understanding of Science, 22*(4), 427–442.

Tindall, D. B., Stoddart, M. C. J., & Callison, C. (2018). The relationships between climate change news coverage, policy debate, and societal decisions. In *Oxford research encyclopedia of climate science*. Oxford University Press.

Vu, H. T., Liu, Y., & Tran, D. V. (2019). Nationalizing a global phenomenon: A study of how the press in 45 countries and territories portrays climate change. *Global Environmental Change, 58*, 101942.

Zuboff, S. (2019). *The age of surveillance capitalism: The fight for a human future at the new frontier of power*. Public Affairs.

---

**How to cite this article:** Lahsen, M. (2022). Evaluating the computational (“Big Data”) turn in studies of media coverage of climate change. *Wiley Interdisciplinary Reviews: Climate Change, 13*(2), e752. [https://doi.org/10.1002/wcc.752](https://doi.org/10.1002/wcc.752)