Herrenhausen Conference October 9-11, 2019 EXTREME EVENTS – BUILDING CLIMATE RESILIENT SOCIETIES.

Report of session 4 "Data Science for human wellbeing"

Synthesis research & action agenda

Title	**Data Science for human wellbeing**
Authors	MIGUEL D. MAHECHA (Max Planck Institute for Biogeochemistry), SALVANO BRICEÑO (ROLAC/ISC), RÉMI FROMENT (Université Catholique de Louvain), FABIAN GANS (Max Planck Institute for Biogeochemistry), ALEXANDRA GARATZOGIANNI (Leibniz Information Center for Science and Technology (TIB)), MARIO HERNANDEZ (ISPRS), MALAN KETCHA ARMAND KABLAN (Université Félix Houphouët-Boigny), AZREEN KARIM (Bangladesh Institute of Development Studies (BIDS)), GUIDO KRAEMER (Max-Planck-Institut für Biogeochemie), SVEN KUNZE (Universität Heidelberg), EMMANUEL LETOUZÉ (Data Pop Alliance), WOOD LINDSAY (Resilienz Ltd), SHUAIB LWASA (Department of Geography Makerere University), BLESSING MASAMHA (Bindura University of Science Education), MELANIE MEIS (CONICET-UBA), ILAN NOY (Victoria University of Wellington, New Zealand), CARL-FRIEDRICH SCHLEUSSNER (Climate Analytics), JEROEN SMITS (Radboud University Nijmegen), SOPHIE VANWAMBEKE (Université Catholique de Louvain), ASTRID WIGIDAL (Universität zu Köln), GREGORY YETMAN (Columbia University) and DEBARATI GUHA SAPIR (Université Catholique de Louvain)

Context, Motivation, Urgency	The consensus expectation of a global increase in average temperatures also implies that we will experience more climate extremes in the near future. Examples are intensified, floods, storms, heat-waves, and droughts in many areas of the world. Understanding how human well-being has been affected by climate extremes in the past and present is therefore of paramount importance. Only if catastrophic consequences chains that occured in the past are well understood, will one be able to anticipate impacts of future events and future climate scenarios. Today we live in a data-rich world. Information from diverse and heterogeneous data streams is available including, for instance, data from satellite remote sensing of the land surface, climate reanalysis, and, crucially, records of social impacts at local level. Increasing data availability comes at a time where not only computational resources are increasingly powerful, but where many modern tools of machine learning and artificial intelligence can be widely used (a phenomenon sometimes also referred to as "democratization of machine learning"). The overarching motivation for this session was therefore to shed light into the questions, what strategies should be promoted to jointly interpret the plethora of information to understand how, where, and why are people affected by climate extremes and what are potential impact chains that may have been overlooked so far. One starting point is certainly to better integrate data, knowledge, and requirements of contributing disciplines.

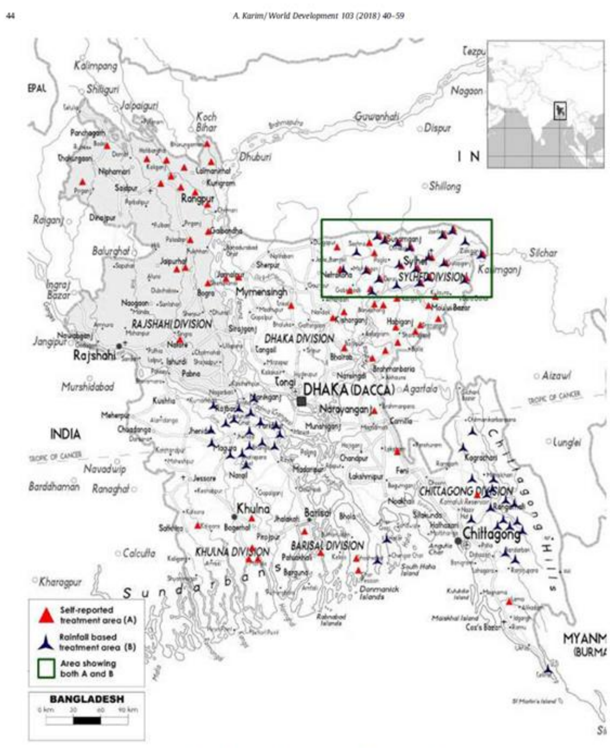

Figure 2. Map showing the treatment areas (sub-districts) in the study.

Source: Karim, Azreen (2018). The household response to persistent natural disasters: Evidence from Bangladesh. *World Development,* 103 (3), pp.40-59. The figure shows the discrepancy between areas that are potentially threatened by flooding if the evidence is only collected from precipitation data, and self-reported and damages. The discrepancy underscores the need to map vulnerabilities explicitly in space and time.

Key research questions incl. expected methodologies and disciplines involved	From a conceptual point of view, “data science” can offer a clear path ahead that needs to be explored. However, there are a wide range of caveats. Key issues and questions that need to be addressed are the following: ● In all phases of disaster risk assessment and management, including forecasting, emergency response and post-disaster reconstruction, the need for interconnected multidisciplinary data for reporting and analysis is always present. **How can data on disaster impact be produced and made available in a timely manner and in a format that is of broad use?** ● In this context, the session concluded that a research question that needs to be addressed with priority is **how to characterize “vulnerability”?** Assuming that hazards and exposure are well characterized by all kinds of data streams, describing “vulnerabilities” implies a much deeper approach that might be supported by, but not addressed completely, by data driven approaches only. However, identification strategies based on interdisciplinary measures of disaster risk exposure needs to be carefully characterized as variations in identifying affected communities might lead to mis-targeted development policy decisions (see also Fig. 1). ● The session emphasizes one question that needs to be addressed in-depth: **What are the ethics of data collection in the aftermath of a catastrophe?** This is a particular challenge in this context, as it may collide with the FAIR principles requiring data to be (Findable, Accessible, Interpretable, and Reproducible). ● Data science is an umbrella term that essentially describes how domain knowledge, classical statistics, and e.g. big-data or AI algorithms can work in tandem to solve a given set of problems. In the context of disaster risk assessment and reduction, the question is: **What domain-knowledge areas need to be integrated in a data-science approach to disaster risk?** Clearly, the interdisciplinary requirements are huge. Not only do we need a clear

	understanding of climate-biosphere feedbacks (e.g. leading to long-term water shortage), but also a clear socioeconomic vision on potential impact chains in different regions. A recurring theme in the working group related to the power of cross-disciplinary work and the need to enhance cross-disciplinary understanding and communication, from very early on in a project, to optimise methodology, data collection etc. This includes a) between disciplines engaged in the project and b) optimising the value of the data and research for subsequent applications. One very practical aspect is formulating interdisciplinary requirements for data and encouraging interaction between data users and producers. There is a clear need for documenting data uncertainties and applying recommendations more explicitly or for broader audiences.

Priorities for action items & implementation	Action priorities need to be placed on the following aspects: 1. *Education on exploiting data resources:* The real value here is to have a "very complete" training of the usage of (geospatial and temporal) data, including a broad understanding of potential interpretation caveats and confounding factors. The broader vision must be to achieve a deep understanding "how information flows" from the first steps of data generation to the final interpretation. This would be the basis for translating data into knowledge and further to action to reduce the risk due to climate extremes in the future. 2. *Understanding future risks:* Another priority is to establish a broad process rethinking the risks of climate extremes in a not too distant future. We need structuring multiple data sources and observations that enable understanding interactions between multiple hazards that are likely to influence future risk. However, we need to consider that not only may the climate baseline has shifted but also societies will have changed in the wake of the digital revolution. 3. *Discussing the "ethics of data collection and exploitation" dilemma:* In times of FAIR data requirement in the scientific domains, we need to initiate a dialogue among practitioners, stakeholders, and scientists to understand and come up with universal principles [including non-intrusive forms of data collection, and the resultant implications for methodologies for data usage.

Participants	Salvano Briceño, Rémi Froment, Fabian Gans, Alexandra Garatzogianni, Lindsay Wood, Debarati Guha-Sapir, Mario Hernandez, Malan Ketcha Armand Kablan, Karim Azreen, Guido Kraemer, Sven Kunze, Emmanuel Letouzé, Miguel D. Mahecha, Blessing Masamha, Melanie Meis, Ilan Noy, Carl-Friedrich Schleussner, Gunther Seckmeyer, Jeroen Smits, Sophie Vanwambeke, Astrid Wigidal, Gregory Yetman, Shuaib Lwasa

How to cite	Mahecha M. D. *et.al.* (2019) Report of session 4 "Data Science for human wellbeing": Synthesis research & action agenda. Herrenhausen Conference October 9-11, 2019 Extreme Events – Building Climate Resilient Societies. doi: 10.17871/HK-report-session-4