AI-Driven Tool for Climate Change Risk Assessment: An Application on Airports Adaptation.

Jacopo Grassi, Politecnico di Torino / WSP in Italia (jacopo.grassi@polito.it / jacopo.grassi@wsp.com)

Abstract

As climate change accelerates, critical infrastructure such as airports are increasingly exposed to a range of climate-related hazards, including extreme heat, flooding, and sea level rise. In this study, we present an AI-powered Multi-Agent System (MAS) designed to assess climate risks to airports through a structured, semi-quantitative framework based on the IPCC risk concept. The system is composed of thematic agents specialized in analyzing Exposure, Vulnerability, and Hazard. Each agent leverages specific data sources and tools, ranging from digital elevation models and land cover datasets to online research capabilities and gridded climate projections from CMIP6 models. Risk scores are synthesized by combining the three dimensions using a logistic normalization function, designed to better reflect medium-to-high risk scenarios. The consistency and validity of the results are evaluated by a Review Agent, which flags questionable outputs for reanalysis. We apply this framework to a case study involving seven Italian airports and four different climate hazards, demonstrating the potential of LLM-driven agents in supporting climate risk assessments for critical infrastructure.

Introduction

In 2023, the global temperature exceeded the 1.5°C increase compared to preindustrial levels for the first time[1]. As this marks a drastic reduction in the intervention window to take actions to reach the Paris Agreements[2], it is clear that Climate Change is already causing massive losses[3] and that this trend is projected to persist with a magnitude that, in the short-medium term, will continue regardless the emissions scenario assumed[4].

In the sixth Assessment Report (AR6), the Intergovernmental Panel on Climate Change (IPCC) has underlined how the concept of risk can serve as framework for understanding Climate Change, and its relationship with impacts and adaptation measures[5].

Risk is defined as the "result of the dynamic interactions between climate-related hazards [potential occurrence of a physical event] with the exposure [the presence of people or infrastructure that could be adversely affected] and vulnerability [the propensity to be adversely affected] of the affected human or ecological system to the hazards"[6].

This definition expands beyond the conventional view of risk as the product of likelihood and consequences of an event. It allows for a more comprehensive assessment of climate risk, which can also account for complex risks[7,8] and include physical hazards that are difficult to define in terms of probability, either because of their chronic nature or due to insufficient scientific understanding.

Concurrently, Climate Services[9], which are meant to provide the decision makers with the appropriate information and knowledge to address climate risks[10], still fall short of their potential and face numerous challenges, including relevance, interdisciplinarity, scalability, and the ability to meet the needs of decision-makers[11–13]. To address these challenges, Climate Services are increasingly integrating emerging technologies, particularly Machine Learning (ML) and Artificial Intelligence (AI), to enhance key components of their value chain, including data assimilation, predictive modeling, uncertainty quantification, and knowledge synthesis[14–17].

As Large Language Models (LLMs) gain prominence, the climate scientific community has commenced to investigate their potential within Climate Services applications[18]. With the rapid advancement of LLMs and the introduction of concepts such as Artificial Intelligence Autonomous Agents and Augmented Language Models (ALMs)[19], it has become evident that LLMs can be effectively instructed to perform highly specialized tasks with remarkable precision. In some instances, their performance surpasses that of highly educated humans.

However, intricate tasks frequently demand multidisciplinary expertise and surpass the capabilities of a single AI agent. Consequently, Multi-Agent Systems (MAS)[20] have emerged, wherein multiple autonomous agents collaborate within a structured framework to solve complex problems. When powered by AI, MAS can replicate human-like collaboration, enabling distributed reasoning, task delegation, and dynamic problem-solving.

These systems are particularly well-suited for domains that require the integration of heterogeneous data sources, diverse analytical approaches, and context-sensitive decision-making attributes that closely align with the demands of Climate Services, and many example are arising in geosciences[21].

Here, we present an AI-driven Multi-Agent System (MAS) designed to perform climate change risk assessments in alignment with the conceptual framework defined by the IPCC AR6. The proposed system leverages the coordinated interaction of specialized AI agents within a modular and extensible architecture.

With this work we aim to demonstrate the feasibility of applying AI-powered MAS to highly specific and multidisciplinary contexts, such as climate change risk assessment, where diverse expertise and heterogeneous data sources must be integrated coherently. Moreover, it lays the groundwork for the operational use of such systems in Climate Services, highlighting their potential to evolve alongside the rapid advancements in Large Language Models (LLMs) and domain-specialized AI agents.

We conducted a case study of seven selected Italian airports to assess the risk associated with four distinct climate hazards: heatwaves, cold waves, extreme rainfall, and droughts. As many other critical infrastructure, airports are expected to face significant material impacts related to Climate Change[22]. For example, an increased frequency of inundation driven by extreme precipitation and sea level rise can lead to extended disruption[23,24], while increase in extreme high temperatures will impact the performances of aircrafts during take-off and landing[25,26].

Methods

Architecture

The proposed tool is based on a Large Language Model (LLM) driven Multi-Agent System. The architecture consists of three main thematic agents, each responsible for conducting an analysis related to Exposure, Vulnerability, and Hazard, respectively. Each agent is equipped with a distinct set of capabilities tailored to its specific task. These agents generate detailed technical reports outlining the analytical procedures and key findings.

Subsequently, the risk is defined using a semi-quantitative approach, which mathematically formalizes the approach of the so-called risk matrices[27,28].

A Risk Synthesis Agent assigns standardized scores for Exposure (0–5), Vulnerability (1–5), and Hazard (1–5), along with a rationale justifying each score. The overall risk score is then computed with the following procedure: i) the three scores are multiplied to obtain a preliminary risk score (r), ii) the final risk score is obtained as:

$$R = 5\left(1 + e^{-k(r-r_0)}\right)^{-1} \qquad Eq. 1$$

This normalization step is based on a logistic function and serves to mitigate the disproportionate compression of intermediate-to-high risk combinations that arises from the direct multiplication of the three scores. We set $k = 0.8$ to assure a moderate slope, and $r_0 = 30$ to assure that a combination of medium hazard, exposure and vulnerability (≈ 3) results in a medium risk. The final risk R is thus a numerical decimal value in the range 0 to 5.

Once this procedure has been executed for each combination of airport and climate hazard defined, a Review Agent assesses the consistency of the entire analysis and identifies potential concerns. The combinations of climate hazard-airport that are flagged are subsequently reviewed by thematic agents, who are specifically instructed to repeat the analysis while incorporating the observations made by the Review Agent.

Thematic agents

Generative AI agents are artificial entities capable to achieve a goal by observing, planning and acting upon the real or virtual world, allowing them to overcome the limitations of tradition LLM models, such as hallucination and a lack of specific competences[19,29]. AI agents are built on three core components: the LLM which serves as centralized decision maker, the tools that empower the agent to interact with the external world and the orchestration layer that describes the general behavior of the agent [30]. In this project we used OpenAI GPT 4o-mini[31] as LLM model, and built the orchestration layer basing on the ReAct framework[32]. Each thematic agent has been augmented with the capability to interact with a set of tools suitable for the given task:

- The Exposure Agent has four core capabilities: i) it can retrieve the elevation of the airport, basing on a Digital Elevation Model[33], ii) it can compute the distance of the airport from rivers and coastlines, basing on the information contained in open source shapefiles[34], iii) it can access information regarding water baseline risk indicators[35], iv) it can retrieve the land cover classification statistics for an area of about $2\ km^2$ in the surrounding of the airport.
- The Vulnerability Agent can conduct online research on the internet. The research must focus on the vulnerabilities of the airport, specifically identifying known vulnerabilities and/or known adaptation efforts implemented by the airport. The Agent can compose a query, retrieve five results, and provide a concise summary of the findings.
- The Hazard Agent has the capability to access a large dataset of gridded climate indicators at high resolution (0.25°x0.25° latitude x longitude), derived from an ensemble of CMIP6 models and distributed by the World Bank Climate Change Knowledge Portal[36]. The dataset contains more than 30 indicators, and spans the period 2015 to 2100 yearly, on three different emission scenarios (SSP 1-2.6, SSP 2-4.5, and SSP 5-8.5). The agent can select the most relevant indicators and perform additional statistical operation to extract relevant insights from the dataset.

The specific prompts used to instruct the agents are reported in

Appendix A – Prompts for Thematic Agents.

Evaluation

Assessing MAS has been a topic of discussion even before the emergence of LLM, which can function as a reasoning hub[37]. Common gaps in developing a clear evaluation framework for AI systems include the lack of understanding of the emergent behavior of AI agents when collaborating and the absence of specific sector-specific benchmarks[20]. Several methodologies have been developed to address this task, mainly based on benchmarks and related to the behavior of single agent performances[38–40].

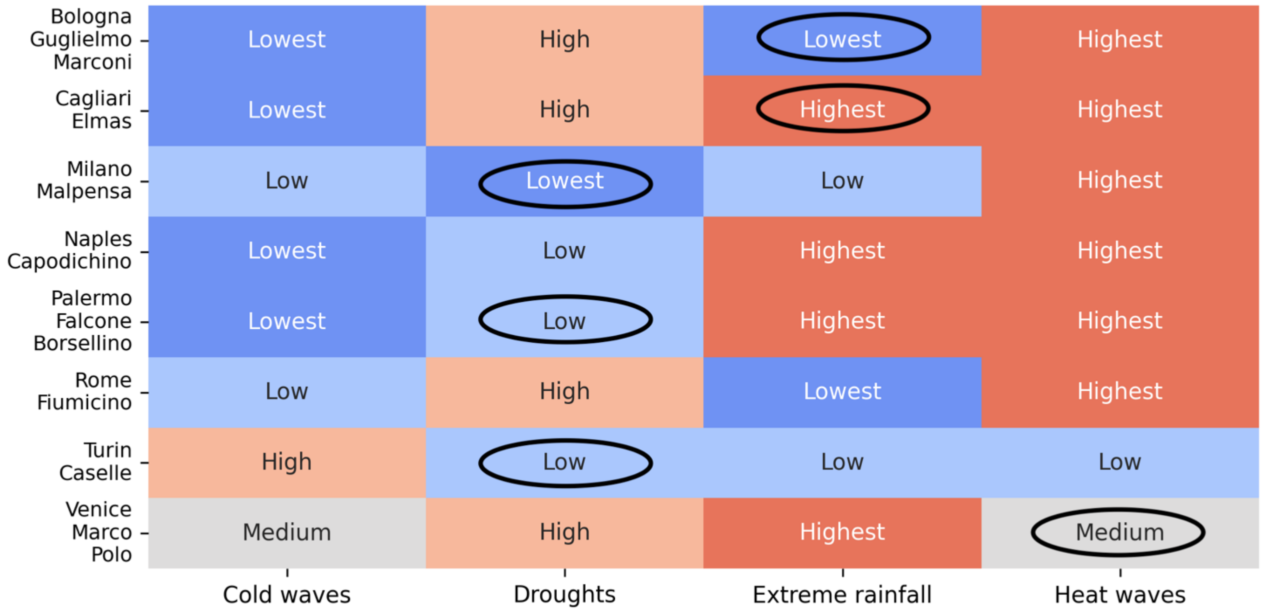

Figure 1 *- Risks Identified in the Initial Analysis. To enhance the representation, the final risk has been categorized into five classes (Lowest, Low, Medium, High, and Highest) based on the numerical risk result, which was previously on a scale of 0 to 5. Black circles indicate the results that were flagged by the Review Agen*

A method for evaluating AI agents is known as 'LLM as a judge'[41,42], based on the evidence that

LLMs have some ability to express their internal confidence level when generating content[43]. Here we defined a Review Agent which is deputy to assess the overall analysis acting as an LLM as a judge. The Review Agent identifies inconsistencies and flags them, following the criteria listed in *Appendix C – Prompt for Review Agent*.

Case study

We applied the tool to a set of seven Italian Airports: Milano Malpensa, Torino Caselle, Venezia Marco Polo, Bologna Marconi, Roma Fiumicino, Cagliari Elmas, Napoli Capodichino, and Palermo Falcone-Borsellino. We extended the analysis on four climate hazards: heat waves, cold waves, drought and extreme rainfall. These hazards can be evaluated by multiple climate indicators, as the ones suggested by the Expert Team on Climate Change Detection and Indices (ETCCDI[44]). We intentionally omitted a definition of the hazards, as this assessment is entrusted to the thematic Hazard Agent. Additionally, we specifically requested the Hazard Agent to focus the analysis on the scenario SSP 2-4.5 for the year 2050.

The initial iteration of the tool engages all thematic agents for each combination of Airport and Climate Hazard. Subsequently, the Risk Synthesis Agent assigns scores to each determinant (Hazard, Exposure, and Vulnerability), and the final risk is computed based on Eq. 1. Finally, the Review Agent identifies critical analysis that necessitates control (***Figure 1***).

For instance, in the context of droughts at Palermo Falcone-Borsellino Airport the Review Agent identifies an inconsistency in the analysis as follows:

"The exposure score is 5, but the hazard score is 1. This discrepancy raises questions about the rationale provided for the drought hazard, which lacks supporting data. What specific environmental data supports the high exposure score for Palermo Falcone Borsellino Airport -

Droughts, and how does this correlate with the low hazard score? "

This review prompts a reevaluation of the thematic agents' assessment of drought in Palermo Airports. The analysis corroborates the scores of Exposure (5) and Vulnerability (4), while the Hazard score has been revised from 1 to 2, resulting in a shift from Low to High risk (**Figure 2**).

In summary, the Review Agent identified six analysis areas that required revision. Five of these areas resulted in an escalation of the overall risk score, while one area led to a reduction.

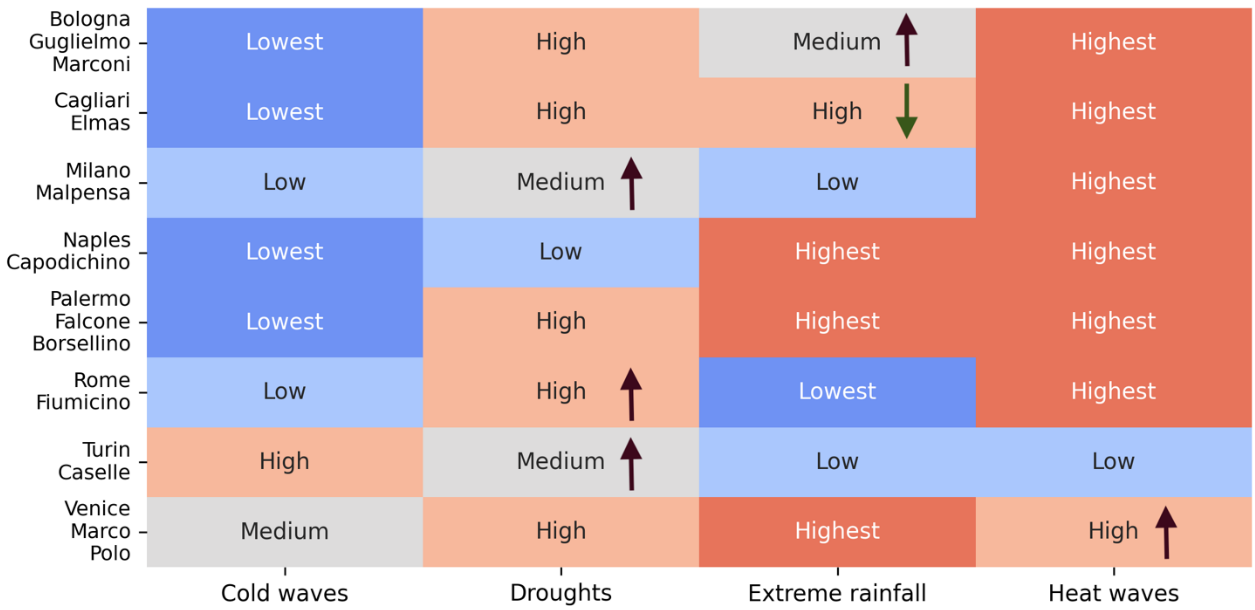

Figure 2 *- Final Risks after the reanalysis triggered by the Review Agent. Arrows quantifies the changes in the Final Risk after the reanalysis.*

Conclusion

Addressing climate-related risks is a multifaceted subject that necessitates an initial analysis, commonly referred to as a Climate Change Risk Assessment. This analysis demands multidisciplinary expertise and the delineation of an analysis context that presents substantial challenges to Climate Services providers.

In this research, we developed an AI-powered MAS based on LLMs to conduct this type of analysis. Our approach is grounded in the risk framework outlined by IPCC-AR6. We have incorporated three distinct thematic agents, each with specialized capabilities for assessing Exposure, Vulnerability, and Hazard. These assessments serve as the basis for semi-quantitative evaluation of the overall risk.

This approach has facilitated the integration of diverse data sources, including a substantial repository of climate indicators' projections and unstructured information obtained through web research, into a seamless workflow. Leveraging the textual capabilities of LLMs, each stage of the process is accessible to a human reviewer, thereby mitigating the black box phenomenon.

It is noteworthy that this aspect is not a secondary consideration. Climate Change Risk, particularly within the context outlined by the IPCC AR6, is a highly subjective and context-dependent process. In such a scenario, the comprehension of the underlying factors that influence the outcomes holds greater significance than the outcomes themselves.

References

1. Copernicus CCS. Global Climate Highlights 2024. (2025).
2. Bevacqua, E., Schleussner, C.-F. & Zscheischler, J. A year above 1.5 °C signals that Earth is most probably within the 20-year period that will reach the Paris Agreement limit. *Nat. Clim. Change* **15**, 262–265 (2025).
3. Newman, R. & Noy, I. The global costs of extreme weather that are attributable to climate change. *Nat. Commun.* **14**, 6103 (2023).
4. Kotz, M., Levermann, A. & Wenz, L. The economic commitment of climate change. *Nature* **628**, 551–557 (2024).
5. Ara Begum, R. *et al.* Point of Departure and Key Concepts. in *Climate Change 2022: Impacts, Adaptation and Vulnerability. Contribution of Working Group II to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change* (eds. Pörtner, H.-O. et al.) 121–196 (Cambridge University Press, Cambridge, UK and New York, NY, USA, 2022). doi:10.1017/9781009325844.003.
6. Reisinger, A. *et al.* Annex I: Glossary. in *Climate Change 2023: Synthesis Report. Contribution of Working Groups I, II and III to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change* (eds. Team, C. W., Lee, H. & Romero, J.) 119–130 (IPCC, Geneva, Switzerland, 2023). doi:10.59327/IPCC/AR6-9789291691647.002.
7. Simpson, N. P. *et al.* A framework for complex climate change risk assessment. *One Earth* **4**, 489–501 (2021).
8. Zscheischler, J. *et al.* Future climate risk from compound events. *Nat. Clim. Change* **8**, 469–477 (2018).
9. Villwock, A. Literature based guiding principles for high-quality climate services. *D41 Clim. Proj.* (2023).
10. Weichselgartner, J. & Arheimer, B. Evolving Climate Services into Knowledge–Action Systems. *Weather Clim. Soc.* **11**, 385–399 (2019).
11. Bessembinder, J. *et al.* Need for a common typology of climate services. *Clim. Serv.* **16**, 100135 (2019).
12. Guentchev, G., Palin, E. J., Lowe, J. A. & Harrison, M. Upscaling of climate services – What is it? A literature review. *Clim. Serv.* **30**, 100352 (2023).
13. Findlater, K., Webber, S., Kandlikar, M. & Donner, S. Climate services promise better decisions but mainly focus on better data. *Nat. Clim. Change* **11**, 731–737 (2021).
14. Huntingford, C. *et al.* Machine learning and artificial intelligence to aid climate change research and preparedness. *Environ. Res. Lett.* **14**, 124007 (2019).
15. Chen, L. *et al.* Artificial intelligence-based solutions for climate change: a review. *Environ. Chem. Lett.* **21**, 2525–2557 (2023).
16. Eyring, V., Gentine, P., Camps-Valls, G., Lawrence, D. M. & Reichstein, M. AI-empowered next-generation multiscale climate modelling for mitigation and adaptation. *Nat. Geosci.* 1–9 (2024) doi:10.1038/s41561-024-01527-w.
17. Chavula, P. & Kayusi, F. Role of Artificial Intelligence in Disseminating Climate Information Services in Africa. *LatIA* 76 (2025).
18. Koldunov, N. & Jung, T. Local climate services for all, courtesy of large language models. *Commun. Earth Environ.* **5**, 1–4 (2024).
19. Mialon, G. *et al.* Augmented Language Models: a Survey. Preprint at https://doi.org/10.48550/arXiv.2302.07842 (2023).
20. Guo, T. *et al.* Large Language Model based Multi-Agents: A Survey of Progress and Challenges. Preprint at https://doi.org/10.48550/arXiv.2402.01680 (2024).
21. Pantiukhin, D., Shapkin, B., Kuznetsov, I., Jost, A. A. & Koldunov, N. Accelerating Earth Science Discovery via Multi-Agent LLM Systems. Preprint at https://doi.org/10.48550/arXiv.2503.05854 (2025).
22. Pek, S. & Caldecott, B. Physical climate-related risks facing airports: an assessment of the world's largest 100 airports. (2020).
23. Fickling, D. Submerged Risks of Climate Change Haunt Low-Level Airports. *Bloomberg* (2018).
24. Yesudian, A. N. & Dawson, R. J. Global analysis of sea level rise risk to airports. *Clim. Risk Manag.* **31**, 100266 (2021).
25. Gratton, G., Padhra, A., Rapsomanikis, S. & Williams, P. D. The impacts of climate change on Greek airports. *Clim. Change* **160**, 219–231 (2020).

26. Zhou, Y., Zhang, N., Li, C., Liu, Y. & Huang, P. Decreased takeoff performance of aircraft due to climate change. *Clim. Change* **151**, 463–472 (2018).
27. Kadkhodaei, M. H. & Ghasemi, E. Development of a Semi-quantitative Framework to Assess Rockburst Risk Using Risk Matrix and Logistic Model Tree. *Geotech. Geol. Eng.* **40**, 3669–3685 (2022).
28. Wood, R. A. *et al.* A Climate Science Toolkit for High Impact-Low Likelihood Climate Risks. *Earths Future* **11**, e2022EF003369 (2023).
29. Xi, Z. *et al.* The rise and potential of large language model based agents: a survey. *Sci. China Inf. Sci.* **68**, 121101 (2025).
30. Wiesinger, J., Marlow, P. & Vuskovic, V. Agents [White paper]. *Google Docs* https://drive.google.com/file/d/1oEjiRCTbd54aSdB_eEe3UShxLBWK9xkt/view?usp=embed_facebook (2024).
31. OpenAI. GPT-4o mini model. (2024).
32. Yao, S. *et al.* ReAct: Synergizing Reasoning and Acting in Language Models. Preprint at https://doi.org/10.48550/arXiv.2210.03629 (2023).
33. European Space Agency (ESA). Copernicus Digital Elevation Model (DEM). (2023) doi:10.5270/ESA-c5d3d65.
34. Kelso, N. V. & Patterson, T. INTRODUCING NATURAL EARTH DATA - NATURALEARTHDATA.COM. (2010).
35. Kuzma, S. *et al.* Aqueduct 4.0: Updated Decision-Relevant Global Water Risk Indicators. (2023).
36. World Bank Group. World Bank Climate Change Knowledge Portal (CCKP). (2025).
37. Di Bitonto, P., Laterza, M., Roselli, T. & Rossano, V. An Evaluation Method for Multi-Agent Systems. in *Agent and Multi-Agent Systems: Technologies and Applications* (eds. Jędrzejowicz, P., Nguyen, N. T., Howlet, R. J. & Jain, L. C.) 32–41 (Springer, Berlin, Heidelberg, 2010). doi:10.1007/978-3-642-13480-7_5.
38. Liu, X. *et al.* AgentBench: Evaluating LLMs as Agents. Preprint at https://doi.org/10.48550/arXiv.2308.03688 (2023).
39. Gioacchini, L. *et al.* AgentQuest: A Modular Benchmark Framework to Measure Progress and Improve LLM Agents. Preprint at https://doi.org/10.48550/arXiv.2404.06411 (2024).
40. Chan, J. S. *et al.* MLE-bench: Evaluating Machine Learning Agents on Machine Learning Engineering. Preprint at https://doi.org/10.48550/arXiv.2410.07095 (2025).
41. Gu, J. *et al.* A Survey on LLM-as-a-Judge. Preprint at https://doi.org/10.48550/arXiv.2411.15594 (2025).
42. Li, D. *et al.* From Generation to Judgment: Opportunities and Challenges of LLM-as-a-judge. Preprint at https://doi.org/10.48550/arXiv.2411.16594 (2025).
43. Lin, S., Hilton, J. & Evans, O. Teaching Models to Express Their Uncertainty in Words. Preprint at https://doi.org/10.48550/arXiv.2205.14334 (2022).
44. Organization (WMO), W. M., Zwiers, F. W. & Zhang, X. Guidelines on Analysis of extremes in a changing climate in support of informed decisions for adaptation. https://library.wmo.int/records/item/48826-guidelines-on-analysis-of-extremes-in-a-changing-climate-in-support-of-informed-decisions-for-adaptation (2009).

Appendix A – Prompts for Thematic Agents

The blue words in {} brackets are iteratively changed during the execution of the tool.

A1 – Exposure Agent Prompt

You are an expert climate analyst tasked with gathering **relevant environmental data** about {location} in relation to the specific climate hazard: {climate_hazard}.

**OBJECTIVE: **

Your goal is to **collect and summarize key environmental factors** that influence the exposure of {location} to {climate_hazard}. Do not assign a score—your role is purely informational.

**METHODOLOGY: **

1. **Analyze Key Environmental Factors: **

  - You can use some tools for gathering punctual data.

  - Don't use all the tools, just the ones that you are sure are useful for the task.

  - Use the data that can be clearly linked with the climate hazard {climate_hazard}.

2. **Use of Aqueduct Risk Data: **

  - If relevant, reference Aqueduct data but **do not** treat it as the primary determinant of exposure.

3. **Logical Assumptions: **

  - Where direct data is unavailable, apply well-reasoned assumptions based on climatic principles.

  - Example: If the location is at high latitude and elevation, infer the likelihood of snow-related exposure.

**OUTPUT: **

- Provide a **structured summary** of the gathered information in bullet points.

- Ensure clarity, relevance, and conciseness.

- Do **not** provide an exposure score—pass the collected information to the next agent for evaluation.

A2 – Vulnerability Agent Prompt

You are an expert climate analyst tasked with gathering **relevant data** on the vulnerability of {location} to the specific climate hazard: {climate_hazard}.

**OBJECTIVE: **

Your goal is to **collect and summarize key vulnerability factors** that influence the ability of {location} to withstand and recover from {climate_hazard}. Do not assign a score—your role is purely informational.

**METHODOLOGY: **

1. **Conduct Targeted Research: **

- Utilize credible sources to gather information about the area's susceptibility to {climate_hazard}.

- Focus on factors that either **increase or decrease vulnerability**.

2. **Assess Key Vulnerability Factors: **

- **Infrastructure: ** Evaluate the resilience of critical infrastructure (e.g., roads, energy, communication).

- **Population Density & Demographics: ** Consider population distribution, socioeconomic conditions, and vulnerable groups.

- **Economic Activities: ** Identify key economic sectors and their susceptibility to {climate_hazard}.

- **Existing Protective Measures: ** Assess current mitigation and adaptation strategies in place.

- **Social & Institutional Capacity: ** Evaluate the community's ability to respond and recover.

- **Environmental Factors: ** Analyze natural buffers (e.g., wetlands, forests) and exacerbating elements (e.g., deforestation, erosion).

3. **Strict Focus on {climate_hazard}: **

- Do **not** infer vulnerability to other climate hazards.

**OUTPUT: **

- Provide a **structured summary** of the gathered information in bullet points.

- Ensure clarity, relevance, and conciseness.

- Do **not** assign a vulnerability score—pass the collected information to the next agent for evaluation.

A3 – Hazard Agent Prompt

I want data to explore the {climate_hazard} hazard in {location}.

1 - Select the data for the location.

3 - Decide a set of indicators (**5 MAXIMUM**) which the best ones to describe the {climate_hazard} hazard. Then select it.

You already have the dataset information, so you don't need to call the tools with dataset information in Json format.

Just call the 5 tools in order.

Appendix C – Prompt for Review Agent

You are tasked with evaluating the results of a climate risk screening for multiple infrastructures.

Your role is to identify and explain any inconsistencies or issues in the data. These inconsistencies may indicate errors that require revision.

Please base your evaluation on the following criteria:

1. Within the same infrastructure, there are two or more similar hazards with very different scores, not justified by the rationales.

2. Across different infrastructures that appear similar, there are two or more similar hazards with very different scores, not justified by the rationales.

3. There are scores that appear contradictory or implausible based on context (e.g., high heatwave risk at very high latitudes).

4. A score is not supported by an adequate rationale. As a reminder:

i. The rationale for Exposure should include environmental numerical data.

ii. The rationale for Vulnerability should include findings from specific research about the airport.

iii. The rationale for Hazard should include numerical data related to atmospheric conditions.

Keep in mind that your evaluation will trigger a follow-up analysis. Therefore, structure your output clearly:

- Start with a summary of your considerations.

- Then, provide a list of combinations in the following format: INFRASTRUCTURE-HAZARD, indicating which ones need to be revised.

**IMPORTANT: Keep the names of the INFRASTRUCTURE and HAZARDS exactly as they appear in the results. DO NOT change their case, spelling, or formatting. **