

INTEGRATING AI AND STATISTICAL METHODS TO STUDY THE IMPACT OF CLIMATE CHANGE ON BIODIVERSITY

Alessia Benevento, Fabrizio Durante**

* Dipartimento di Matematica e Fisica "Ennio De Giorgi", Università del Salento, Lecce, Italy.

Abstract

Climate change and biodiversity loss are two of the most pressing challenges of recent decades. While the former has received significant attention from both the scientific community and the public, the latter often remains in the background. Nevertheless, the two phenomena are deeply interconnected, and Artificial Intelligence (AI), through machine learning and predictive modeling, provides valuable means to help address some of the global challenges they present. In particular, AI can be employed to predict the risks and impacts associated with climate change, providing information that directly contributes to a better understanding of biodiversity dynamics and vulnerabilities. In this work, we present an application of a machine learning method, namely the clustering of compound events, to demonstrate how the joint analysis of two or more interacting factors can provide a more comprehensive understanding of socio-environmental phenomena.

Keywords

Biodiversity, climate change, compound events, machine learning, statistics.

1. Introduction

In October 2024, a devastating flood resulted in the death of 228 individuals in Valencia. Approximately 360000 people were left without access to potable water for many days, while 50000 experienced a prolonged lack of electricity. The estimated economic damage amounted to 16.5 billion euros (Calvo-Sancho et al., 2025). The flood also severely impacted biodiversity, particularly at the Albufera Natural Park, by inundating the area with plastic waste, sewage, and other debris, contaminating the habitat and threatening species like the endangered Valencia toothcarp (see, for instance, Chauvet, 2024).

In Los Angeles, on January 7th, 2025, severe wildfires led to the death of 29 people. Around 150000 individuals were displaced, 10000 houses were destroyed, and the full extent of the damages is still under assessment (Clemens, 2025). Also in this case, the event negatively impacted biodiversity through habitat destruction, displacement of wildlife, and soil erosion that harms aquatic ecosystems (see, for instance, Senzaki and Deehan, 2025).

These events can be interpreted as examples of the direct consequences of ongoing climate change. In fact, recent analyses (Calvo-Sancho et al., 2025) indicate that human-induced global

warming significantly increased both the intensity and likelihood of the Valencia flood. These findings are consistent with broader assessments from the IPCC Sixth Assessment Report (2023), which emphasize the growing attributional link between anthropogenic climate change and the intensification of extreme weather events worldwide. The climate emergency is now both present and pressing.

At the same time, it has been recognized that climate change is also contributing to biodiversity loss and change (Urban, 2015). Extreme climatic events such as cyclones, droughts, floods or heatwaves can generate severe consequences on biodiversity.

What remains less widely acknowledged is that biodiversity is also a powerful ally in confronting climate change and increasing societal resilience. However, this resource is insufficiently valued and utilized.

In 2018, the storm known as Vaia struck Italy, particularly the Triveneto area. The event caused the death of 37 people and inflicted damages amounting to 5 billion euros. Within three days, precipitation reached 715 millimeters, an amount that, under normal conditions, would fall over the course of an entire year. Vast areas of fallen trees in locations such as the Pinè plateau in the Gorai

mountain chain became emblematic of the meteorological disaster (Chirici et al., 2019).

Notably, the storm disproportionately affected monocultures of Norway spruce. This species, highly valued for its rapid growth and timber yield, requires only eighty years before harvesting, compared to approximately twice as many years for oak. For this reason, since the late 1800s, Norway spruce plantations were established widely across the Alpine arc, replacing native broadleaf forests (Candotti et al., 2025).

The reforestation of areas affected by Vaia represents a relevant case study for analyzing adaptive management strategies in forest ecosystems under increasing climatic variability. The extensive damage to monocultural stands of Norway spruces revealed the structural fragility of homogeneous systems when exposed to extreme meteorological stress. This observation aligns with existing evidence that biodiversity enhances the capacity of ecosystems to recover from disasters (e.g., Holling, 1973; Folke et al., 2004).

Post-disturbance management should therefore aim to promote mixed forest structures, rather than replicating pre-existing monocultures. Diversified systems have been shown to reduce the propagation of damage and the economic losses, and stabilize ecological processes over time (Thompson et al., 2009). While such approaches may entail higher management complexity, they contribute to lowering systemic risk across temporal and spatial scales.

The Vaia event hence underscores the need to reinterpret the environmental (extreme) events as a complex socio-ecological system whose quantification and (spatio-temporal) modeling remain essential for informed decision-making. Biodiversity, in this framework, constitutes a functional component of resilience rather than a purely ecological variable.

In fact, biodiversity encompasses the full range of plants, animals, and microorganisms present on Earth, constituting the biological wealth of the planet. All organisms have co-evolved, forming complex interconnections and interdependencies akin to a web. While the rupture of a few connections may not compromise the system entirely, the loss of many would result in collapse.

The mathematical theory of complex systems provides a comprehensive framework for the analysis of systems composed of multiple interacting components whose collective behavior cannot be fully understood by examining

individual elements in isolation. Such systems can be conveniently represented by random variables that, although seemingly governed by underlying rules, may reveal hidden patterns upon closer examination.

Specifically, here we illustrate compound (weather and climate) events (Leonard et al., 2014) that arise from the combination of multiple drivers that together may generate societal or environmental risks. Many major climate-related disasters, such as droughts and storms, result from such compound events, during which interacting drivers exceed the coping capacity of affected systems.

Even though our understanding of climate extremes and associated impacts is continuously improving, events that break the coping capacity of social and environmental systems often surprise us. This is because most current risk estimates tend to overlook the risks associated with correlated compound drivers.

Studying compound events often requires a multidisciplinary approach combining domain knowledge of the underlying processes with, for example, statistical methods, climate model outputs and, more recently, AI (Artificial Intelligence) algorithms. In this work, after examining the link between climate change and biodiversity, we illustrate the application of compound events in this context and highlight some of their potential uses.

2. *Climate changes and biodiversity: two dimensions and a common challenge*

As noted above, climate change and biodiversity are closely interconnected. Nevertheless, a comparison between the two reveals three fundamental differences: *awareness*, *impact*, and *measurement* (as underlined in the TEDx talk of Castellucci, 2025).

In terms of *awareness*, climate change has become central to public debate and widely recognized as the most urgent environmental crisis. Biodiversity loss, despite being one of the first crises acknowledged by the scientific community, has not received equivalent visibility. For instance, search data indicate that “climate change” is searched globally up to three times more often than “biodiversity”. This imbalance is reflected in media coverage and institutional attention.

The institutional response highlights further disparities. The Conference of the Parties (COP) on

climate was established in 1992 during the Rio Earth Summit, and by 2025 will have convened 30 sessions since its inaugural meeting in Berlin in 1995. In contrast, the COP on biodiversity, initiated in Nassau in 1994, has convened only 16 times. This discrepancy underscores how global governance prioritizes climate over biodiversity.

The second difference lies in *impact*. Greenhouse gases disperse uniformly across the atmosphere, producing global consequences irrespective of the continent of origin. In contrast, biodiversity loss manifests primarily at the local scale, such as the disappearance of forests in the Amazon or grasslands in Europe. Although the impacts are initially local, their cumulative effects amount to a global crisis.

The third distinction concerns *measurement*. Climate change can be quantified using carbon dioxide equivalent (CO₂e), a metric that consolidates emissions into a single comparable value. Biodiversity, however, involves multiple interacting dimensions, making the establishment of a synthetic metric challenging (Colwell, 2009; Daly et al., 2018; Pallara, 2024). A possible solution is the Mean Species Abundance (MSA) index, which evaluates how close an ecosystem is to its natural state on a scale from zero (completely degraded) to one hundred (undisturbed).

Technological tools aid in assessing biodiversity through a combination of satellite imagery, in-field sensors, and DNA analysis. For instance, acoustic sensors can monitor pollinators, particularly wild bees (Barlow & O'Neill, 2020), by capturing sound frequencies and species presence. Satellite imagery, processed through Artificial Intelligence, segments the Earth's surface into classified land-use categories, enabling a detailed mapping of natural and anthropogenic areas. By comparing these classifications to reference baselines, the MSA can be estimated and applied across diverse ecosystems, from deserts to rainforests. Crucially, MSA reflects not the quantity but the quality of biodiversity relative to natural conditions, allowing equitable comparisons across ecosystems of varying richness.

Despite this difference in approach, climate change and biodiversity are strongly correlated. According to the Global assessment report of the Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services (IPBES, 2019), climate change is one of the five drivers that have affected biodiversity in the past 50 years. Observational evidence indicates an accelerating

rate of environmental change across marine, terrestrial, and freshwater ecosystems, with direct repercussions on agriculture, aquaculture, fisheries, and ecosystem services. The combined action of multiple anthropogenic drivers produces compound effects that amplify impacts beyond the contribution of individual factors. Such dynamics are increasingly documented in diverse ecosystems, from coral reefs and Arctic environments to savannas, where interacting stressors lead to nonlinear responses and potential regime shifts.

Various forms of pollution and the proliferation of invasive alien species further intensify these pressures. Despite regional variability, atmospheric, terrestrial, and aquatic pollution levels have continued to rise in several areas. Marine plastic contamination, for instance, has increased more than tenfold since 1980, affecting hundreds of species and propagating through trophic networks with potential implications for human health. Concurrently, the number of recorded alien species has grown by approximately 40% since 1980, largely driven by global trade and human mobility (IPBES, 2019). Nearly one fifth of the planet's land surface is now considered vulnerable to biological invasions, with cascading effects on biodiversity, ecosystem functions, and socio-economic stability.

Given the accelerating pace and interconnected nature of these processes, investigating compound effects, that is, the interactions among multiple concurrent environmental drivers, has become a scientific and societal urgency. Traditional single-factor analyses are no longer sufficient to capture the emergent behavior of coupled human-natural systems. As highlighted in Boero (2024), the coupling of atmospheric and oceanographic processes, together with coastal and seabed geomorphology, plays a crucial role in determining physical dynamics, while human activities increasingly modify atmospheric, terrestrial, and aquatic features. Moreover, biotic interactions, in concert with physical drivers, are fundamental in shaping biodiversity composition and ecosystem functioning.

Thus, developing integrative frameworks capable of quantifying, modeling, and predicting such interactions is essential for anticipating critical thresholds, guiding adaptive management, and informing policy responses under increasing climatic and ecological uncertainty.

In the following sections, we present a recent methodology (compound events) that serves to describe those phenomena that are caused by several factors or have multiple effects (e.g., droughts, heat waves, wildfires, and air pollution). In fact, although its use has so far been mainly focused on climate-related events, it can also be valuable for biodiversity studies, as these are likewise characterized by a multitude of drivers and factors.

3. Climate changes and biodiversity: a statistical tale

The study of climate changes and biodiversity requires a rigorous statistical analysis of long-term environmental data. Over the past century, global average surface temperatures have increased by approximately 1.1°C compared to pre-industrial levels, with a marked acceleration observed since the 1970s (IPCC, 2023). This trend is strongly correlated with the rise in atmospheric concentrations of greenhouse gases, particularly carbon dioxide (CO₂), methane (CH₄), and nitrous oxide (N₂O), which have reached levels unprecedented in at least the last 800000 years.

Time-series analyses of meteorological and satellite datasets confirm both the upward trajectory of mean global temperatures and the increased variability of climatic patterns. For example, in Figure 1 we show the data of the monthly maximum temperatures in summer (June-July-August) from 1960 to 2024 in a station located in Southern Germany. The blue line interpolates the actual data, while the red line shows the linear trend of the data. It shows an increasing trend that can be observed in most parts of the world. To build Fig. 1 we have used the data extracted from the ERA5 dataset (Copernicus Climate Change Service, 2025), which collects global climate and weather data of the past 8 decades.

Extreme events, including floods, wildfires, heatwaves, and droughts, exhibit a statistically significant increase in frequency and intensity. For instance, precipitation anomalies demonstrate not only greater deviation from historical averages but also an altered distribution, with episodes of extreme rainfall concentrated in shorter periods. Similarly, the incidence of large-scale wildfires has risen in correlation with prolonged droughts and elevated mean seasonal temperatures.

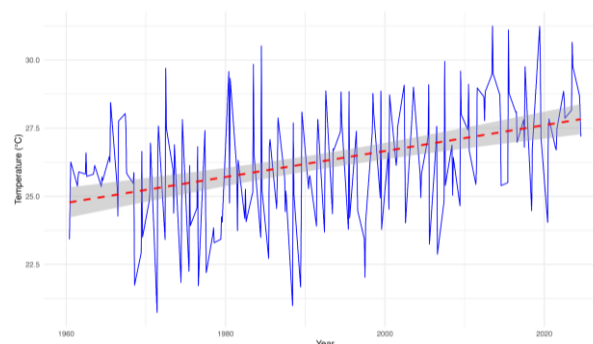


Fig. 1: Temperature time series trend from 1960 to 2024 in a station located in Southern Germany

Statistical modeling, employing regression analyses and probabilistic forecasting, indicates that these observed phenomena cannot be attributed solely to natural variability. Instead, anthropogenic factors are the dominant drivers, as confirmed by the Intergovernmental Panel on Climate Change (IPCC, 2023). The use of representative concentration pathways (RCPs) and shared socioeconomic pathways (SSPs) further enables projections of future climatic conditions under different emissions scenarios.

From a biodiversity perspective, these statistical trends underscore the urgency of integrating climate metrics with ecological indicators. While CO₂ equivalent serves as a synthetic and widely adopted measure of anthropogenic impact, its intersection with biodiversity indices such as Mean Species Abundance (MSA) may provide a more integrated understanding of the cascading consequences of climate change on ecosystems.

Another indicator used to quantify biodiversity is the Living Planet Index (LPI), which provides a complementary perspective on global ecological change. The LPI is a composite indicator that tracks the state of global biodiversity by measuring the average change in abundance of vertebrate species populations over time. Developed by the Zoological Society of London and WWF (Zoological Society of London & WWF, 2025), the index aggregates thousands of population time series for mammals, birds, amphibians, reptiles, and fish across all major biogeographical regions. Values are normalized to 100 in 1970, meaning that any subsequent decrease reflects an overall decline in population abundance relative to that baseline.

Figure 2 shows the temporal evolution of the LPI across different world regions from 1970 to 2020. The data have been downloaded from the World Wildlife Fund and Zoological Society of London (2025), processed by Our World in Data. The index value considers the change in abundance in 34836 populations across 5495 native species both terrestrial and marine.

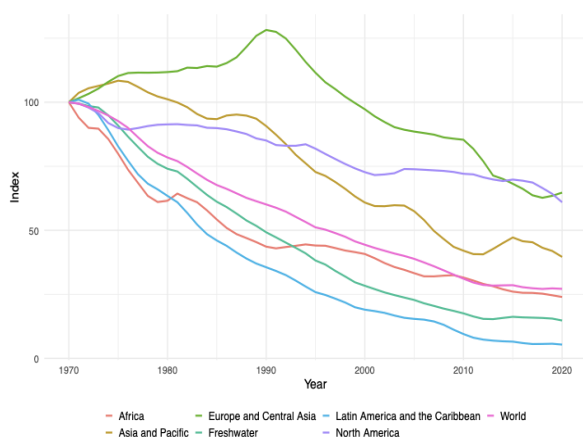


Fig. 2: Living Planet Index by region (1970–2020)

The general trend reveals a substantial decline in biodiversity, particularly pronounced in tropical and developing regions such as Latin America and Africa, whereas some temperate regions (e.g., Europe and North America) show relatively smaller decreases or partial stabilization. The global curve (in pink) indicates an overall reduction of around 60% in average population sizes since 1970.

When considered together with the increasing trends in maximum temperature illustrated above, these results highlight the urgent need to better understand the complex interconnections between climate change and biodiversity loss. Rising temperatures, habitat degradation, and other anthropogenic pressures jointly may have contributed to the observed declines in wildlife populations, underlining the importance of integrating climate–biodiversity conservation strategies.

4. Climate changes and biodiversity: the role of AI

Artificial intelligence (AI) is emerging as a transformative tool in addressing the multifaceted challenges of climate change, offering new pathways for both mitigation and adaptation. Through advanced data analytics, machine

learning, and predictive modeling, AI can enhance the precision of climate forecasting, optimize energy systems, and support the transition toward low-emission technologies (see, e.g., De Padova et al., 2024). In the environmental domain, AI can facilitate the monitoring of land use, water resources, and biodiversity. These applications can help identify habitat changes and contribute to efforts aimed at protecting ecosystems. Beyond technical innovation, AI also plays a crucial role in integrating diverse climate data sources thereby strengthening decision-making frameworks in agriculture, disaster management, and urban planning. However, the deployment of AI in climate research is not without risks. High computational energy consumption, biases in environmental datasets, and unequal access to digital infrastructure may inadvertently reinforce existing vulnerabilities. Ensuring that AI development aligns with principles of transparency, equity, and sustainability is therefore essential to realizing its potential as an ally in climate action.

Recent research initiatives (e.g., Naughtin et al., 2025) emphasize the need for interdisciplinary collaboration to ensure that AI systems contribute not only to emissions reduction and resource optimization but also to the protection of biodiversity and ecosystem services, which remain foundational to global climate resilience.

Another aspect that has been revolutionized in recent years thanks to advances in digital technologies and AI is the approach to biodiversity protection (see, e.g., De Nunzio & Rizzo, 2024; Ullah et al., 2025). Traditionally, monitoring ecosystems and species required significant resources, lengthy timeframes, and manual interventions, often limited by logistical challenges. Today, thanks to machine learning algorithms and advanced data analysis systems, it is possible to process enormous amounts of information from diverse sources quickly, enabling real-time and highly detailed insights.

AI, for example, allows us to monitor animal and plant species through images, sounds, or environmental data collected in the field, identifying their presence, movements, and behaviors without directly interfering with their natural habitat. This data can be cross-referenced with advanced climate models to predict the future effects of climate change on species, thus anticipating risk situations and guiding targeted conservation interventions.

Furthermore, predictive analytics techniques may help to identify the areas most vulnerable to habitat loss or invasion by alien species, allowing for the implementation of preventive policies and effective conservation strategies before the damage becomes irreversible.

Additionally, machine learning is a valuable tool for analyzing and interpreting biological data, particularly genetic data (Greener et al., 2022). Biodiversity conservation also involves protecting the genetic heritage of species, a key element for their ability to adapt and survive.

Through sophisticated genetic databases, learning algorithms can identify genetic diversity within animal and plant populations, monitor the presence of harmful mutations, and design more effective repopulation interventions.

Digital gene banks, integrated with AI systems, also allow for the virtual and physical preservation of DNA, seed, and tissue samples, facilitating the exchange of information between research centers around the world and accelerating the development of customized conservation strategies for endangered species (Marinelli et al., 2022).

Among the artificial intelligence techniques increasingly applied in climate science, clustering has proven particularly valuable (Straus, 2019). As an unsupervised learning approach, it enables the identification of hidden patterns and recurrent configurations within large and heterogeneous climate datasets. When applied to high-resolution reanalysis datasets like ERA5, clustering methods allow the detection of co-occurring anomalies in temperature, precipitation, soil moisture, or wind patterns, providing deeper insights into the mechanisms driving compound extremes. This approach enhances both the understanding and the prediction of multi-hazard climate interactions, ultimately supporting more targeted adaptation and risk management strategies for vulnerable ecosystems and human systems.

5. Clustering of compound events

While clustering techniques are usually adopted in grouping single variables, the analysis of complex systems requires the development of techniques that cope with more variables at the same type. This is particularly needed in the study of compound climate events. These are situations in which two or more climate drivers and/or hazards interact, either simultaneously or sequentially, to produce impacts that are

substantially different and often more severe than the sum of their individual effects. These events arise from statistical dependence among climatic variables (e.g., temperature, precipitation, humidity, wind) and are increasingly recognized as critical drivers of risk in both ecological and socio-economic systems (Bevacqua et al., 2021).

Compound climate *extremes* refer to events in which multiple climatic or environmental variables interact simultaneously, producing outcomes that often generate disproportionately severe impacts on both human societies and ecosystems. The likelihood of such events is directly influenced by the degree of dependence among their underlying drivers. Generally, the stronger the correlation, the higher the probability of co-occurrence. Within the risk framework, compound events primarily relate to the hazard component, where the probability of an extreme climatic condition carries the potential for large-scale impacts. Any change in the likelihood of hazards consequently alters overall risk.

Tab. 1: Examples of compound climate events

Hazard	Main drivers	Example cases
Drought	Precipitation, evapotranspiration, soil moisture, temperature	Europe 2003 (Vicente-Serrano et al., 2010)
Fire risk	Temperature, precipitation, relative humidity, wind, lightning	Australia 2019-20 (Nolan et al., 2020)
Drought+Heat	Temperature, precipitation, evapotranspiration, atmospheric humidity	Russia-Europe 2010 (Barriopedro et al., 2011)
Wind + precipitation extremes	Wind speed, precipitation, orography, large-scale atmospheric circulation	Storm Kyrill 2007 (Pinto et al., 2009)

Importantly, many major climate- and weather-related disasters are inherently compound in nature, as the convergence of two or more factors, individually not necessarily extreme, can quickly exceed the adaptive capacity of ecological or social systems. Under conditions of

global warming, such compound events are expected to occur more frequently, while the likelihood of unprecedented and unexpected combinations will also increase, further amplifying uncertainty in risk assessment and management. In Table 1 we provide some examples of climate events that in literature are already considered *compound*.

As regards a possible analysis of data, a study on compound climate events faces huge and complex datasets: thousands of measurements of temperature, rainfall, humidity, wind, soil moisture, and more, collected across long time periods and wide regions. In Fig. 3, we show an example of pairs of climatological events that are strongly correlated. The figure represents the joint distribution of maximum air temperatures and mean soil moistures, i.e., the statistical relationship between the two variables, collected over a station in South Italy from 1940 till 2024 in the summers (June-July-August, Copernicus Climate Change Service, 2025).

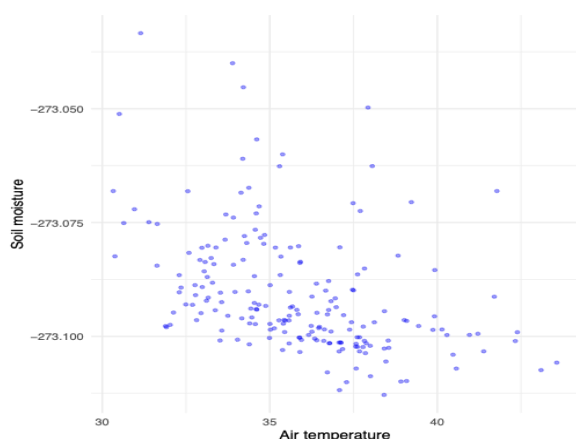


Fig. 3: Joint distribution of maximum air temperature and mean soil moisture observed at a station in Southern Italy during summer seasons from 1940 to 2024

Looking at single variables one by one does not capture how extremes combine. This is where clustering becomes useful. Clustering is a statistical technique that groups together events that are similar, where the similarity is defined based on several characteristics shown by the data at the same time. For example, in Figure 4, we show two possible ways to cluster a set of objects based on two different characteristics: the colours (a) or the shapes (b).

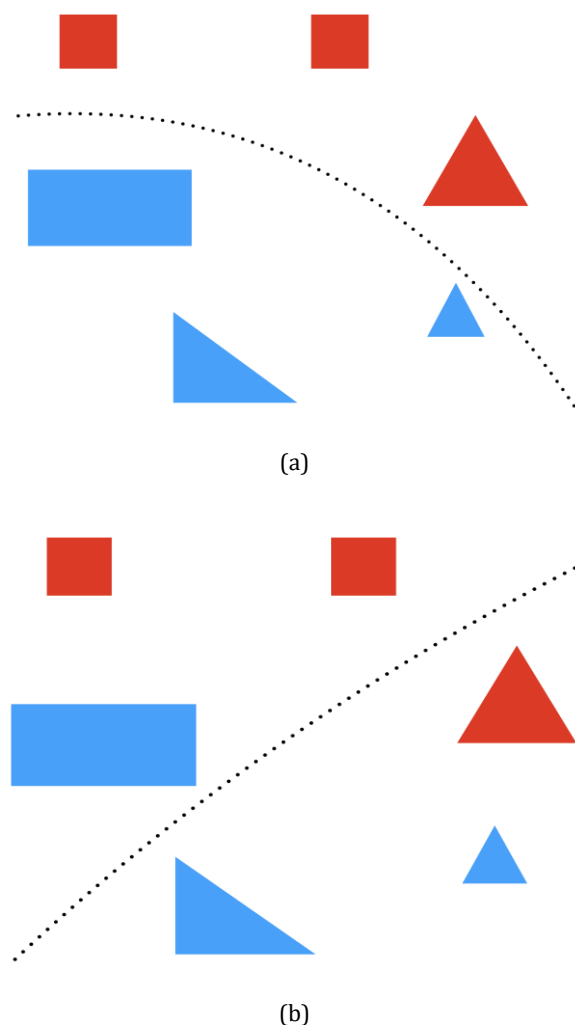


Fig. 4: Example of two clustering approaches applied to the same set of objects: (a) grouping by colour and (b) grouping by shape

From a mathematical point of view, the similarity or dissimilarity between elements is generally measured using a concept called *distance*. A distance quantifies how far apart two objects are. There are different ways to define it depending on what we want to capture.

The most familiar is the Euclidean distance, which corresponds to the straight-line distance between two points in space as shown in Figure 5. The Manhattan distance, instead, indicates how far you'd have to travel if you could only move up/down and left/right - like driving along city blocks in Manhattan (that's where the name comes from!).

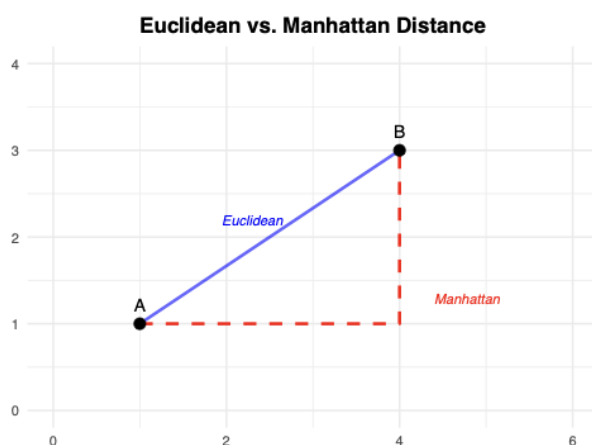


Fig. 5: Comparison between Euclidean distance (straight-line distance between two points) and Manhattan distance (path distance restricted to horizontal and vertical moves)

However, when the objects we compare are more complex, such as probability distributions, rather than single points, more sophisticated measures are involved. One important example is the Wasserstein distance (Fig. 6). It can be understood as the minimum amount of work required to transform one distribution into another, where *work* means shifting probability mass across space.

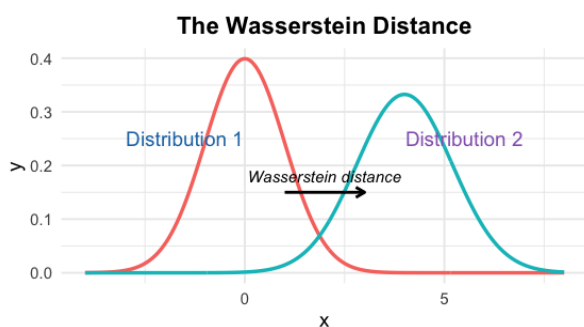


Fig. 6: Example of Wasserstein distance for optimally transporting one distribution into another

Thus, the choice of distance measure depends on the nature of the data: simple geometric distances work well for points, while more advanced notions such as the Wasserstein distance are needed for comparing entire distributions.

By identifying these clusters it is possible, for example, to recognize which combinations of variables lead to extreme impacts, to quantify how often each type of event occurs or to detect emerging event types that were rare in the past but could become common in future climates. Thus,

clustering helps translate raw climate data into meaningful groups of compound events that can be more easily studied, compared, and communicated. Instead of thousands of scattered data points, we get a smaller set of well-defined risk profiles that can be used for further analysis.

6. Clustering of compound events: an application

We want to show an application of clustering techniques to climatological data. The main methodology has been described in detail in Benevento and Durante (2023), Benevento et al. (2024), Castrovilli et al. (2024).

As regards the clustering algorithm, we apply a hierarchical clustering with the average linkage method. The optimal number of clusters was determined in a data-driven way using the Dunn index (Dunn, 1974), an internal validity measure that evaluates the ratio between the smallest inter-cluster distance and the largest intra-cluster distance, thus favoring compact and well-separated clusters.

Specifically, we consider a grid of 527 sites over Italy. At each site (s_i) we collect a random vector (X_1, X_2) representing two time series, each related to a particular feature of the phenomenon under consideration: X_1 represents the time series of hourly maximum temperatures for the months of June, July, and August (JJA) from 2014 to 2024; X_2 represents the hourly potential evapotranspiration (hPet) computed via the Penman-Monteith measure in the same period.

The data have been downloaded from the ERA5 dataset (Copernicus Climate Change Service, 2025). The obtained time series have no seasonal component, but may present a trend due to climate change. Thus, we estimate and remove a linear trend fitted by regression to obtain the detrended time series of seasonal maxima/averages. In this way, we focus on the detection of the anomalies in each time series, since we disentangle the effects of long-term climatological trends.

Given the assertion concerning the co-occurrence of the events, we take advantage of the opportunities offered by the class of bivariate Extreme Value (EV) copulas to formalize this fact into a mathematical model (Durante and Sempi, 2016; Salvadori et al., 2007).

The class of bivariate EV copulas represents a family of dependence models specifically designed to describe the joint behavior of two variables under extreme conditions. These copulas capture how two extreme events are likely to occur

together, even when each follows a different marginal distribution. Unlike standard copulas, EV copulas are characterized by their ability to model the tail dependence, that is, the probability that both variables simultaneously take on very large values. This property makes them particularly suitable for analyzing compound events, where the co-occurrence of extremes plays a critical role in risk assessment.

For example, in Fig. 7 we show two simulated time series representing correlated climate variables generated using a Gumbel copula, which induces strong upper-tail dependence. The red dashed lines indicate the 95th percentile thresholds for each variable, while the red points mark time steps where both series simultaneously exceed their respective thresholds.

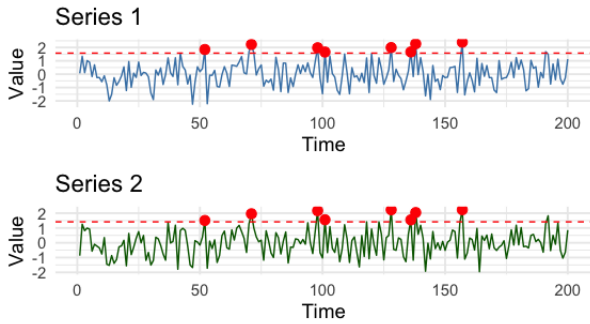


Fig. 7: Illustration of upper-tail dependence in compound events

The first clustering experiment focuses on comparing the probability laws of each attribute X_i with $i=1,2$, across all sites. Specifically, we analyze the dissimilarity among the distributions of the anomalies of temperature maxima and the hPet, respectively. For a given attribute i , the dissimilarity between two sites s_ℓ and $s_{\ell'}$ is quantified through the Wasserstein distance d_W between their empirical probability distributions:

$$\begin{aligned} \delta_{\ell\ell'} &= d_W(X_i(s_\ell), X_i(s_{\ell'})) \\ &= \left(\int_0^1 |F_\ell^{-1}(u) - F_{\ell'}^{-1}(u)|^2 du \right)^{1/2} \end{aligned}$$

where F_ℓ^{-1} and $F_{\ell'}^{-1}$ denote the quantile functions associated with $X_i(s_\ell)$ and $X_i(s_{\ell'})$, respectively. The resulting dissimilarity matrix has dimensions 527×527 , where 527 is the number of geographical locations under consideration. Notice that only land sites are included in the analysis. This choice, which serves merely as an

example of how the proposed method can be applied to cluster a set of objects, was made to improve the clarity and interpretability of the final cluster representation. Nevertheless, the influence of the sea on the climate is still evident, as shown in Figure 8a, where coastal areas display patterns that differ from those observed in more continental regions.

The clustering approach explored to this point allows us to group together locations that share similar marginal distributions of a given climatic variable. Hence, no compound effects are explicitly modeled in this stage: each cluster reflects areas with a comparable “marginal” behavior, that is, a similar level of temperature-related or hPet-related risk.

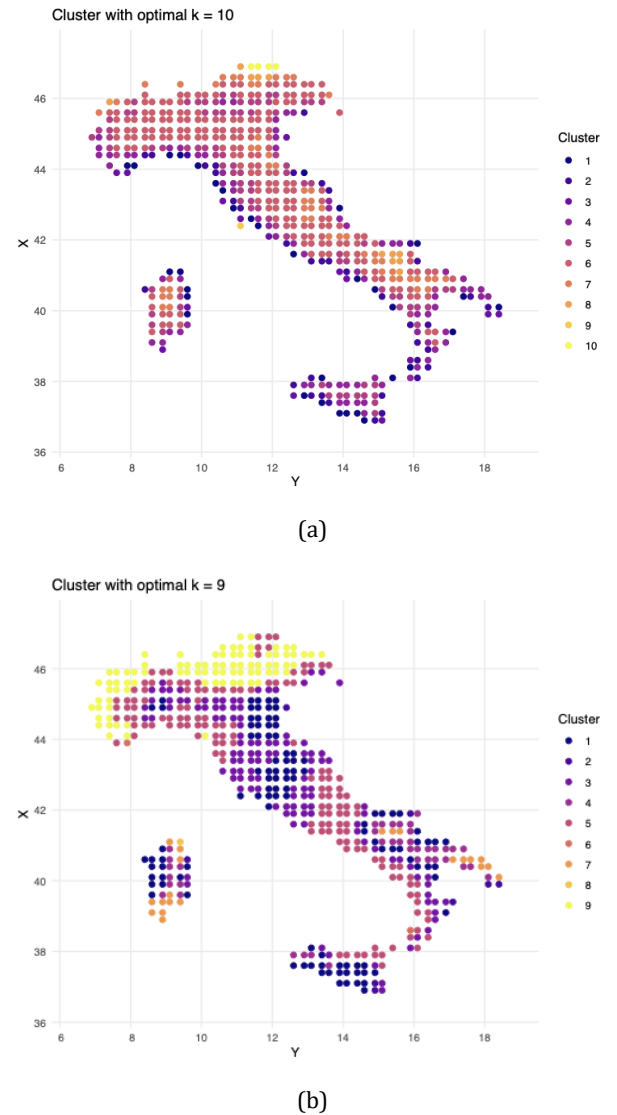


Fig. 8: Regionalization obtained with respect to Temperatures (a) and hPet (b) in the summer period from 2014 to 2024 across the whole Italian territory.

Such clustering provides a preliminary understanding of spatial coherence in single-variable extremes (see Fig. 8) before exploring joint dependencies between attributes.

The second clustering approach focuses on the dependence between attributes at each site rather than on their marginal distributions.

For every location, a copula is estimated from the two variables using pseudo-observations derived from their normalized ranks.

Each site is thus represented by a copula C_ℓ that summarizes the local dependence pattern. The dissimilarity between sites s_ℓ and $s_{\ell'}$ is quantified through the Wasserstein distance between their copulas,

$$\delta_{\ell\ell'} = d_W(C_\ell, C_{\ell'}).$$

To compute this distance, we rely on the class of bivariate EV copulas, each associated with a Kendall distribution function $K_C(u)$, which describes how the dependence encoded by a copula C is distributed over the unit square (Salvadori et al., 2007).

The comparison between sites is therefore expressed as

$$d_W(C_\ell, C_{\ell'}) = d_W(K_{C_\ell}, K_{C_{\ell'}}),$$

which reflects differences in their joint dependence structure.

This clustering procedure groups together sites that share similar cross-attribute behavior, emphasizing areas where compound events tend to manifest in a comparable way, as shown in Fig.9.

By jointly analyzing multiple variables, we can identify hidden patterns and correlations that may remain undetected when studying each factor in isolation.

7. Conclusions

The interdependence between climate change and biodiversity loss requires analytical frameworks capable of capturing the complexity of their interactions. In this work, we have illustrated how the integration of Artificial Intelligence and statistical methods can enhance our understanding of multifactorial environmental phenomena, with a special focus on clustering on compound events.

The proposed methodology not only supports the exploration of large climatological datasets but

also provides a flexible foundation for predictive modeling and scenario analysis.

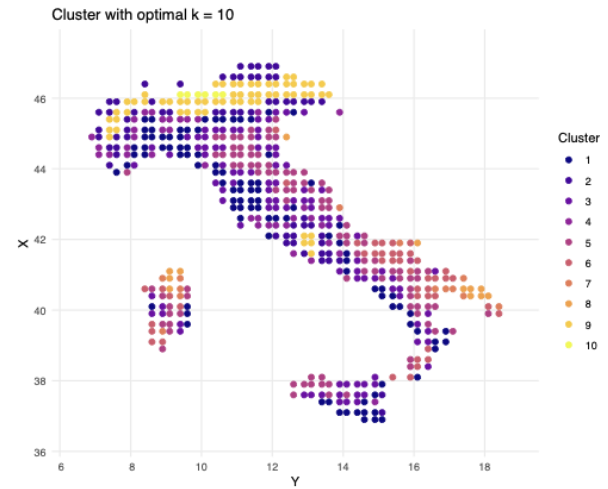


Fig. 9: Regionalization obtained by simultaneously comparing Temperatures and hPet, in the summer period from 2014 to 2024 across Italy. Comparison of the double probability laws observed at each site.

Such tools are essential for informing policy decisions, improving environmental monitoring systems, or proposing new strategies to mitigate the impacts of climate and ecological disruptions.

Moreover, a better understanding of compound events and their associated risk may help in the understanding of biodiversity effects. In fact, the growing risk of compound event occurrence has the potential to erode the ability of biodiversity features to anticipate, absorb or recover from the effects of future environmental changes, representing a growing challenge for global biodiversity conservation (Ameca et al., 2024).

Acknowledgments

A preliminary version of this work was presented by FD at the workshop held in Lecce on May 30th, 2025, in celebration of the International Day for Biodiversity. FD wishes to thank the main organizer, Dr. Virginia Valzano, for creating an inspiring and stimulating environment for discussion. The authors also extend appreciation to the anonymous referees for their constructive comments, insightful suggestions, and for indicating relevant references that have contributed to improving this paper.

AB and FD have been supported by MUR-PRIN 2022 PNRR, Project “Stochastic Modeling of Compound Events” (No. P2022KZJTZ) funded by the European Union - Next Generation EU. The work of FD has been also carried out with partial

financial support from ICSC - Centro Nazionale di Ricerca in High Performance Computing, Big Data and Quantum Computing, funded by EU - Next Generation EU (CUP F83C22000740001).

REFERENCES

- Ameca, E. I., Nie, Y., Wu, R., Mittermeier, R. A., Foden, W., & Wei, F. (2024). Identifying protected areas in biodiversity hotspots at risk from climate and human-induced compound events for conserving threatened species. *Science of The Total Environment*, 938, 173192.
- Barlow, S. E., & O'Neill, M. A. (2020). Technological advances in field studies of pollinator ecology and the future of e-ecology. *Current Opinion in Insect Science*, 38, 15-25.
- Barriopedro, D., Fischer, E. M., Luterbacher, J., Trigo, R. M., & García-Herrera, R. (2011). The hot summer of 2010: redrawing the temperature record map of Europe. *Science*, 332(6026), 220-224.
- Benevento, A., & Durante, F. (2023). Wasserstein dissimilarity for copula-based clustering of time series with spatial information. *Mathematics*, 12(1), 67.
- Benevento, A., Durante, F., Gallo, D., & Gatto, A. (2024). Hierarchical clustering of time series with Wasserstein distance. In *Mathematical and Statistical Methods for Actuarial Sciences and Finance* (pp. 49-54). Cham: Springer Nature Switzerland.
- Bevacqua, E., De Michele, C., Manning, C., Couasnon, A., Ribeiro, A. F. S., Ramos, A. M., Vignotto, E., Bastos, A., Blesić, S., Durante, F., Hillier, J., Oliveira, S. C., Pinto, J. G., Ragno, E., Rivoire, P., Saunders, K., van der Wiel, K., Wu, W., Zhang, T., Zscheischler, J. (2021). Guidelines for studying diverse types of compound weather and climate events. *Earth's Future*, 9, e2021EF002340.
- Boero, F. (2024). A roadmap to knowledge-based maritime spatial planning. In *Advances in Marine Biology* (Vol. 97, pp. 167-189). Academic Press.
- Calvo-Sancho, C., Díaz-Fernández, J., González-Alemán, J. J., Azorín-Molina, C., Halifa-Marín, A., Montoro-Mendoza, A., ... & Martín, M. L. (2025). *Anthropogenic Climate Change Attribution to a Record-breaking Precipitation Event in October 2024 in Valencia, Spain* (No. EGU25-15941). Copernicus Meetings.
- Candotti, A., Ennemoser, M., Seeber, J. & Tomelleri, E. (2025). Norway spruce dominates natural regeneration five years after a large-scale wind disturbance in the higher montane and lower subalpine belts in the eastern Alps. *Forest Ecology and Management*, 595, 123053.
- Castrovilli, R., Durante, F., Gallo, D., & Salvadori, G. (2024). Regionalization Methods for Compound Extremes Based on the Wasserstein Distance. In *Scientific Meeting of the Italian Statistical Society* (pp. 360-365). Cham: Springer Nature Switzerland.
- Castellucci, V. (2025). *La biodiversità: un alleato contro il cambiamento climatico* [TEDx Talk]. TEDxSondrio. YouTube. <https://youtu.be/bdpV3R0hhzI>.
- Chirici, G., Giannetti, F., Travaglini, D., Nocentini, S., Francini, S., D'Amico, G., ... & Marchetti, M. (2019). Stima dei danni della tempesta "Vaia" alle foreste in Italia. *Forest@*, (1), 3-9.
- Chauvet, R. (2024). *Albufera's fragile ecosystem is at risk: Another tragic victim of València's historic floods*. *Euronews*. Retrieved from <https://www.euronews.com/green/2024/11/14/we-really-need-to-adapt-valencias-floods-devastated-this-national-parks-fragile-ecosystem>.
- Clemens, R. (2025). An Evaluation of the Varied Economic Impacts of the January 2025 California Wildfires.
- Colwell, R. K. (2009). Biodiversity: concepts, patterns, and measurement, pp. 257–263. Princeton University Press.

Copernicus Climate Change Service. (2025). *ERA5 hourly data on single levels from 1940 to present* [Data set]. European Centre for Medium-Range Weather Forecasts (ECMWF). Retrieved from <https://cds.climate.copernicus.eu>.

Daly, A. J., Baetens, J. M., & De Baets, B. (2018). Ecological Diversity: Measuring the Unmeasurable. *Mathematics*, 6(7), 119.

De Nunzio, G., & Rizzo, R. (2024). Artificial Intelligence and Biodiversity. *SCIRES-IT – SCientific RESearch and Information Technology*, 14 (Special Issue), 53–70. <http://dx.doi.org/10.2423/i22394303v14Sp53>

De Padova, D., Mossa, M., Chiaia, G., Chimienti, G., Mastrototaro, F., & Adamo, M. (2024). Optimized environmental monitoring: Innovative solutions to combat climate change. *SCIRES-IT – SCientific RESearch and Information Technology*, 14 (Special Issue), 43–52. <http://dx.doi.org/10.2423/i22394303v14Sp43>

Dunn, J. C. (1974). Well-separated clusters and optimal fuzzy partitions. *Journal of Cybernetics*, 4(1), 95–104.

Durante, F., & Sempi, C. (2016). *Principles of Copula Copula Theory*. CRC Press, Boca Raton, FL.

Folke, C., Carpenter, S., Walker, B., Scheffer, M., Elmqvist, T., Gunderson, L., & Holling, C. S. (2004). Regime shifts, resilience, and biodiversity in ecosystem management. *Annual Review of Ecology, Evolution and Systematics*, 35(1), 557–581.

Greener, J. G., Kandathil, S. M., Moffat, L., & Jones, D. T. (2022). A guide to machine learning for biologists. *Nature reviews Molecular cell biology*, 23(1), 40–55.

Holling, C. S. (1973). Resilience and stability of ecological systems. *Annual Review of Ecology and Systematics*, 4, 1–23. <https://doi.org/10.1146/annurev.es.04.110173.000245>

IPBES (2019). Global assessment report on biodiversity and ecosystem services of the Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services. E. S. Brondizio, J. Settele, S. Díaz, and H. T. Ngo (editors). IPBES secretariat, Bonn, Germany. 1148 pages. <https://doi.org/10.5281/zenodo.3831673>

IPCC (2023): Summary for Policymakers. 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* [Core Writing Team, H. Lee and J. Romero (eds.)]. IPCC, Geneva, Switzerland, pp. 1–34, doi:10.59327/IPCC/AR6-9789291691647.001

Leonard, M., Westra, S., Phatak, A., Lambert, M., van den Hurk, B., McInnes, K., ... & Stafford-Smith, M. (2014). A compound event framework for understanding extreme impacts. *Wiley Interdisciplinary Reviews: Climate Change*, 5(1), 113–128.

Marinelli, E., Ghabach, E., Yan, Y., Bolbroe, T., Sella, O., Heinis, T., & Appuswamy, R. (2022). Digital preservation with synthetic DNA. In *Transactions on Large-Scale Data-and Knowledge-Centered Systems LI: Special Issue on Data Management-Principles, Technologies and Applications* (pp. 119–135). Berlin, Heidelberg: Springer Berlin Heidelberg.

Naughtin, C., Evans, D., Mason, C., & Trinh, K. (2025). AI for Climate R&D Roadmap: Consultation paper.

Nolan, R. H., Boer, M. M., Collins, L., Resco de Dios, V., Clarke, H., Jenkins, M., ... & Bradstock, R. A. (2020). Causes and consequences of eastern Australia's 2019–20 season of mega-fires. *Global change biology*, 26(3), 1039–1041.

Pallara, D. (2024). Biodiversity, Sustainability, Growth: Thoughts of a Mathematician. *SCIRES-IT – SCientific RESearch and Information Technology*, 14 (Special Issue), 11–16. <http://dx.doi.org/10.2423/i22394303v14Sp11>

- Pinto, J. G., Zacharias, S., Fink, A. H., Leckebusch, G. C., & Ulbrich, U. (2009). Factors contributing to the development of extreme North Atlantic cyclones and their relationship with the NAO. *Climate dynamics*, 32(5), 711-737.
- Salvadori, G., Michele, C. D., Kottegoda, N. T., & Rosso, R. (2007). *Extremes in nature: an approach using copulas*. Dordrecht: Springer Netherlands.
- Senzaki, M., & Deehan, L. (2025). Wildfires in Los Angeles: Impacts on wildlife and ocean ecosystems. Environment California. Retrieved from <https://environmentamerica.org/california/articles/wildfires-in-los-angeles-impacts-on-wildlife-and-ocean-ecosystems/>.
- Straus, D. (2019). Clustering Techniques in Climate Analysis. *Oxford Research Encyclopedia of Climate Science*. Retrieved 18 Oct. 2025, from <https://oxfordre.com/climatescience/view/10.1093/acrefore/9780190228620.001.0001/acrefore-9780190228620-e-711>.
- Thompson, I., Mackey, B., McNulty, S., & Mosseler, A. (2009). Forest resilience, biodiversity, and climate change. In *Secretariat of the Convention on Biological Diversity, Montreal. Technical Series no. 43*. 1-67. (Vol. 43, pp. 1-67).
- Ullah, F., Saqib, S., & Xiong, Y. C. (2025). Integrating artificial intelligence in biodiversity conservation: bridging classical and modern approaches. *Biodiversity and Conservation*, 34(1), 45-65.
- Urban, M.C. (2015). Accelerating extinction risk from climate change. *Science*, 348, 571-573.
- Vicente-Serrano, S. M., Beguería, S., & López-Moreno, J. I. (2010). A multiscalar drought index sensitive to global warming: the standardized precipitation evapotranspiration index. *Journal of Climate*, 23(7), 1696-1718.
- World Wildlife Fund and Zoological Society of London (2024) – processed by Our World in Data. “Living Planet Index” [dataset]. World Wildlife Fund and Zoological Society of London, “Living Planet Index” [original data].
- Zoological Society of London & WWF. (2025). *Living Planet Index statistics portal*. Retrieved from <https://stats.livingplanetindex.org>.
- Zscheischler, J., Martius, O., Westra, S. *et al.* (2020). A typology of compound weather and climate events. *Nat Rev Earth Environ* 1, 333–347.
- Zscheischler, J., Westra, S., Van Den Hurk, B. J., Seneviratne, S. I., Ward, P. J., Pitman, A., ... & Zhang, X. (2018). Future climate risk from compound events. *Nature Climate Change*, 8(6), 469-477.