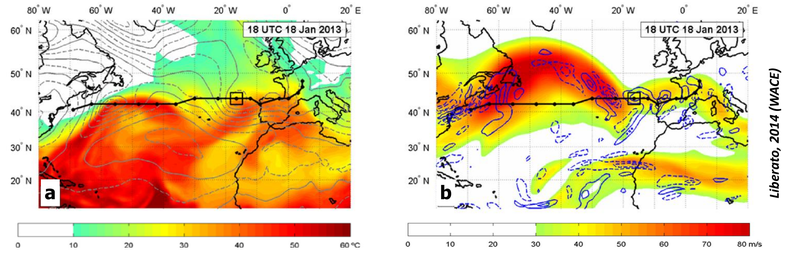
Expert Crowdsourcing for Semantic Annotation of Atmospheric Phenomena

Margarida L. R. Liberato[1,2], Hugo Paredes[2,3], Alexandre Ramos[1], Arsénio Reis [2,3], Riccardo Hénin[1], João Barroso[2,3]

1 – Instituto Dom Luiz (IDL), Faculdade de Ciências, Universidade de Lisboa, Lisboa, Portugal (mlr@utad.pt) 2 – Universidade de Trás-os-Montes e Alto Douro, UTAD, Vila Real, Portugal
3 – INESC Technology and Science (INESC TEC), Porto, Portugal

1. Introduction

Weather and **climate extremes** have gained great attention to the general public and policy makers recently. **Extratropical cyclones** and **frontal systems** are central components of weather over mid latitudes. These phenomena are associated with **compound events**, resulting from the **combination of physical processes leading to an impact**. In fact, dramatic changes in temperature, wind extremes and heavy precipitation events occurring in winter over land in the mid latitudes are mostly associated with extratropical cyclones. It is well known that the **Iberian Peninsula**, due to its location, is **prone to the occurrence** of these **compound events** and **associated hazards** (Liberato *et al.*, 2013; 2014).



2. Background

Large quantities of data gathered through climate and climate change research are difficult to describe, analyse and index, constraining the understanding of atmospheric phenomena and forecasting of future events. Many techniques and methodologies have been used for **annotating data**, enabling **semantic understanding and interoperability**. **Automatic processes** make use of special designed or customized **algorithms** to identify common characteristics in the data. Manual processes require **human computation** to process data and **associate semantics**. **Crowdsourcing** is often used as an engagement solution, allowing an efficient usage of human computation.

On the other hand **algorithms** on data given on **multidimensional structured grids** for the efficient detection and tracking of features in spatiotemporal atmospheric data continue to be developed **at increasing complexity**. Often algorithms fail to identify and to track these systems over consecutive occurrences due to

- the systems' characteristics,
- the diverse data sources,
- the resolution of the data,

requiring the development of specific and **more complex algorithms** to handle these kind of systems.

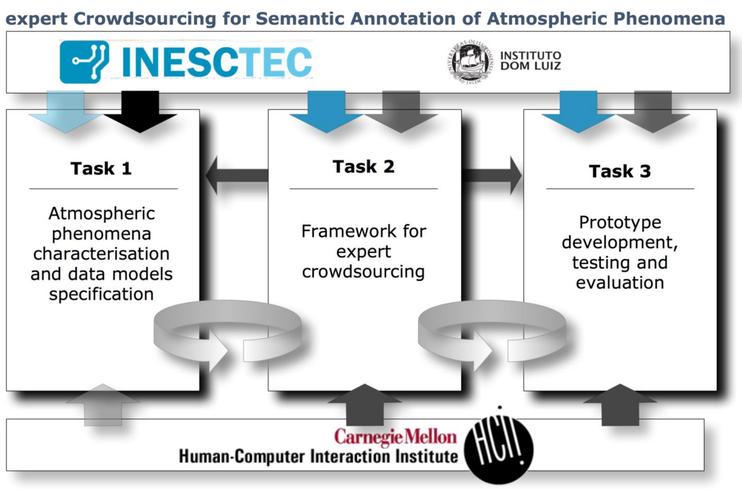
3. Expert crowdsourcing for semantic annotation

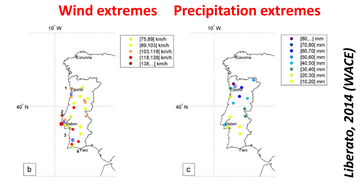
For atmospheric scientists, separating consecutive occurrences of such phenomena may be relative simple task. Therefore it can be outsourced to an expert crowd and current algorithms may still be used, with clear benefits for the computational power required.

4. Approach and research challenges

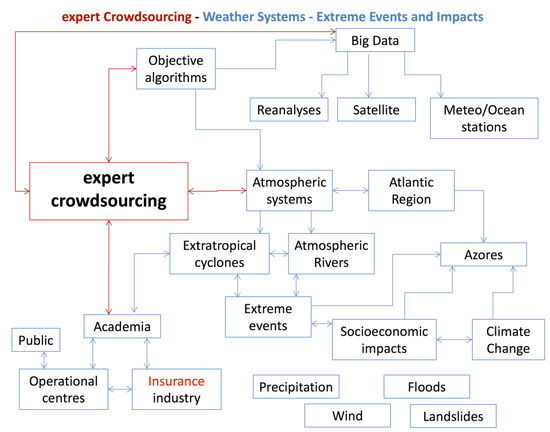
The approach follows a **Design Science Research methodology** organized in the three cycles (Hevner, 2007): relevance, design and rigor. The methodology cycles orchestrate the design science research activities with the environment and the knowledge base. The ultimate goal is to find evidence to answer the **following research challenges**:

- (RQ1) How to represent **atmospheric and climate data** in format that can be **human** and **machine understandable**?
- (RQ2) Can an **expert crowdsourcing framework** address the requirements of **atmospheric** phenomena **semantic annotation**?
- (RQ3) How will the **experts be engaged to be crowd workers**?
- (RQ4) Does **human computation** represent an **added value** for **climate big data analysis**?





eCSAAP research conceptual map



5. Final Remarks

In this project we explore the usage of expert crowdsourcing for annotating compound **hydrometeorological extreme** events over **the Euro-Atlantic region**, so automated methods and computational resources can be optimized in a **future hybrid approach**.

Atmospheric phenomena annotation aims at **bringing new dimensions to current big data problems in climate** and **atmospheric sciences**. Today big data full potential in weather and climate science domain is still restricted by the poor semantic knowledge of data gathered and the inability to correlate data with other domains.

Acknowledgments: This work is supported by the Portuguese Science and Technology Foundation (Fundação para a Ciência e Tecnologia – FCT), under projects UID/GEO/50019/2013 – Instituto Dom Luiz and CMU/CS/0012/2017 – "eCSAAP - expert Crowdsourcing for Semantic Annotation of Atmospheric Phenomena".

References:

Hevner 2007 Scandinavian J. Information Systems http://aisel.aisnet.org/sjis/vol19/iss2/4

Liberato *et al.* 2013 Nat. Hazards Earth Syst. Sci. doi: 10.5194/nhess-13-2239-2013

Liberato 2014 Weather and Climate Extremes doi: 10.1016/j.wace.2014.06.002