Thesis Plan

Detecting and Assessing Pollution Events from Wildfires Using Remote Sensing and Meteorological Data: A Data Science Approach

Green Data Science

Ana Russo - ISA

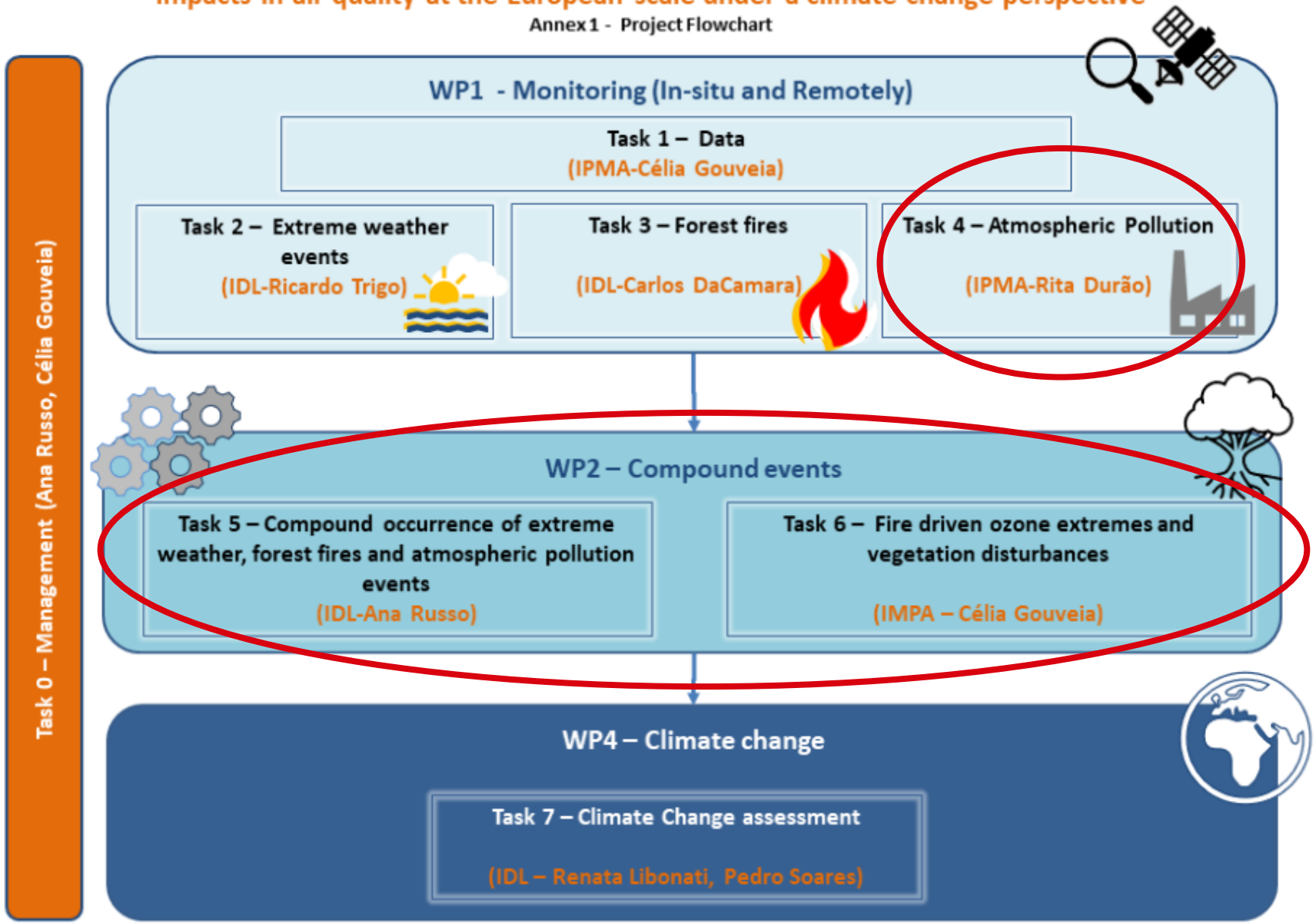
Célia Gouveia - IPMA

Framework

- Branch from an ongoing project called DHEFEUS – research wildfire pollution events
- Wildfires are a major environmental and health threat, worsening air quality as climate change intensifies fire activity
- Atmospheric dispersion models have limitations (computational demands and discrepancies)

DEFHEUS- Droughts, heatwaves, and fires: Exploring compound and cascading hazards and their impacts in air quality at the European scale under a climate change perspective

Annex 1 - Project Flowchart



Data & Sources

- Big Data (Tri and Quadri dimensional) that include coordinates, altitude and meteorological data
- Main sources of data - Copernicus Atmosphere Monitoring Service (CAMS) and Climate Data Store (CDS) - ERA5 hourly data on single levels from 1940 to present

Thesis Objectives

- Identify key indicators correlating pollution events to wildfires
- Improve pollution detection
- Evaluate remote sensing technologies for wildfires
- Analyze wildfire emissions
- Develop predictive model(s)
- Provide insights to improve air quality monitoring

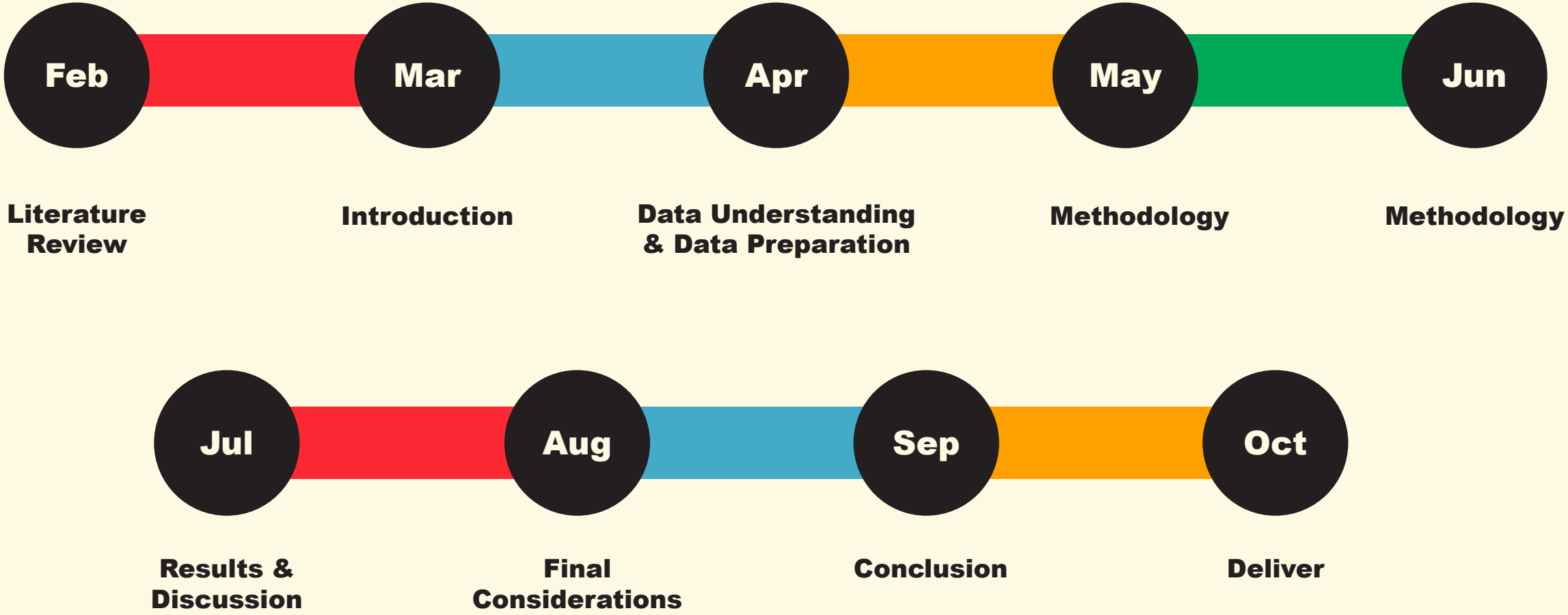
Thesis Tasks

- Extensive reading on past work
- Data collection & preparation
- Atmospheric model analysis
- Train and test models for pollution events
- Model validation & impact assessement
- Results interpretation

Expected results

- Identification of air pollution events linked to wildfires
- Relationship between wildfires and air pollution
- Development of machine learning for monitoring
- Further analysis of how wildfire smoke spreads

Timeline


Feb
Mar
Apr
May
Jun
Literature Review
Introduction
Data Understanding & Data Preparation
Methodology
Methodology
Jul
Aug
Sep
Oct
Results & Discussion
Final Considerations
Conclusion
Deliver

0 Past Literature

- **Goal**: Read and understand work done previously by reading the available contents. Conduct a thorough review of existing work in three main areas: air pollution, remote sensing, and data science applications.
- **Milestones**: Finalize Literature Review
- **Deliverables**: Document with relevant literature and others. Summary of current knowledge on wildfire pollution and transboundary impacts. Review of remote sensing tools (e.g., MODIS, SEVIRI, Sentinel) and data science methods. Identify knowledge gaps to justify your approach.
- **Timeline**: Week 1-3

1 Introduction

- Goal: Finalize the introduction section by outlining the context, significance, and aims of the research.

- Deliverables: Background information - Contextualize the problem. Problem statement - Clearly articulate the gap in current solutions. Objectives and research questions - Refine measurable goals and hypotheses.

- Timeline: **Week 4-5**

2 Data & Methodology

- **Goal**: Define data sources, preprocessing steps, and modeling techniques.
- **Milestones**: Data Understanding, Collection and Processing & Model Implementation
- **Deliverables**:
  - Data Collection: Obtain meteorological data (CAMS, WRF-Chem, station data). Acquire remote sensing data (FRE, FRP from MODIS, Sentinel).
  - Data Processing: Perform cleaning and preprocessing tasks (handling missing values, scaling). Integrate datasets for a unified analysis framework.
  - Model Development: Implement machine learning models (Random Forests, XGBoost, Neural Networks). Develop geospatial tools for spatial and temporal smoke dispersion mapping.
  - Evaluation: Validate models with performance metrics (accuracy, F1-score, AUC). Cross-validate against historical wildfire pollution events (e.g., Portugal's 2017 megafires).
- **Timeline**: Week 6-14

3 Results & Discussion

- **Goal**: Analyze and interpret results from the implemented models and spatial-temporal analyses.
- **Milestones**: Spatial-Temporal Analysis & Results and Discussion
- **Deliverables**:
  - Model Performance: Present tables, graphs, and metrics of model results. Spatial-Temporal Analysis: Create maps showing pollution dispersion, patterns, and affected regions.
  - Discussion: Correlate FRP/FRE data with pollution levels. Evaluate model strengths, weaknesses, and their implications.
- **Timeline**: Week 15-18

4 Conclusion

- **Goal**: Summarize findings, discuss implications, and propose future work.
- **Milestones**: Conclude practical work
- **Deliverables**: Recap main findings and contributions. Discuss potential policy and real-world applications. Suggest areas for future research (e.g., incorporating health impact data).
- **Timeline**: **Week 19-21**

Thank You

Sofia Rodrigues

27899