# FLOOD FORECASTING

Presented by Gabriel Santiago B.

# AGENDA

Introduction

Data

Model

Conclusion

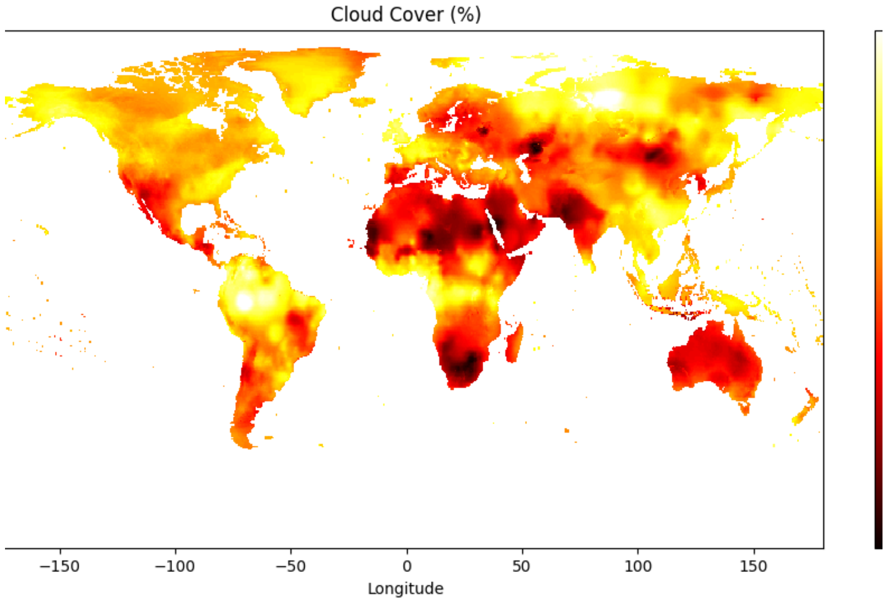
# INTRODUCTION

# PROBLEMATIC

Peru is an Emerging Market Economy with sustained economic growth. However, as many Latin-American countries, it has not been unrelated to climate change.

## -1.5% PIB
## +245k affected families

# DATA




## Gridded Time Series Dataset

Set of month, latitude and longitude information of:

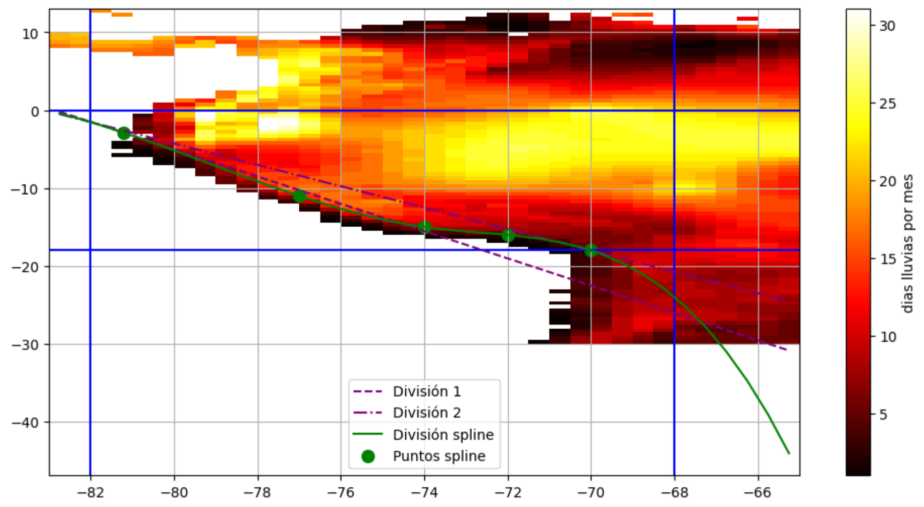
1. Cloud Coverage (%)
2. Temperature Range
3. Frost Days
4. Potential evapo-transpiration
5. Precipitation rate
6. Minimum temperature
7. Maximum temperature
8. Mean temperature
9. Vapour pressure
10. Wet days

## Country Aggregates Dataset

Set of month, and country average of above-mentioned variables.

# DATA SOURCES

# DATA PROCESSING



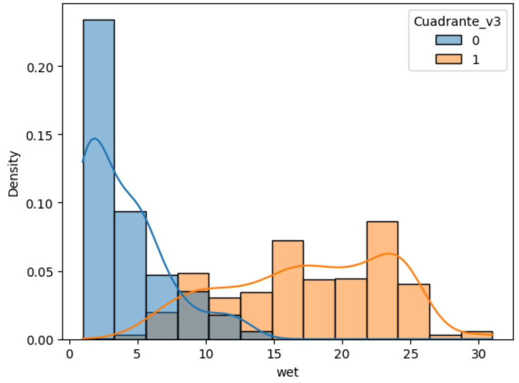

Regions created based on target* and regional knowledge

**Latitude and Longitude Delimitation:**

We filter latitude and longitude to contain Perú.

**Regions Creation:**

We created 2 different Regions based on the target* distribution using cubic splines.




Target Distribution based on created regions

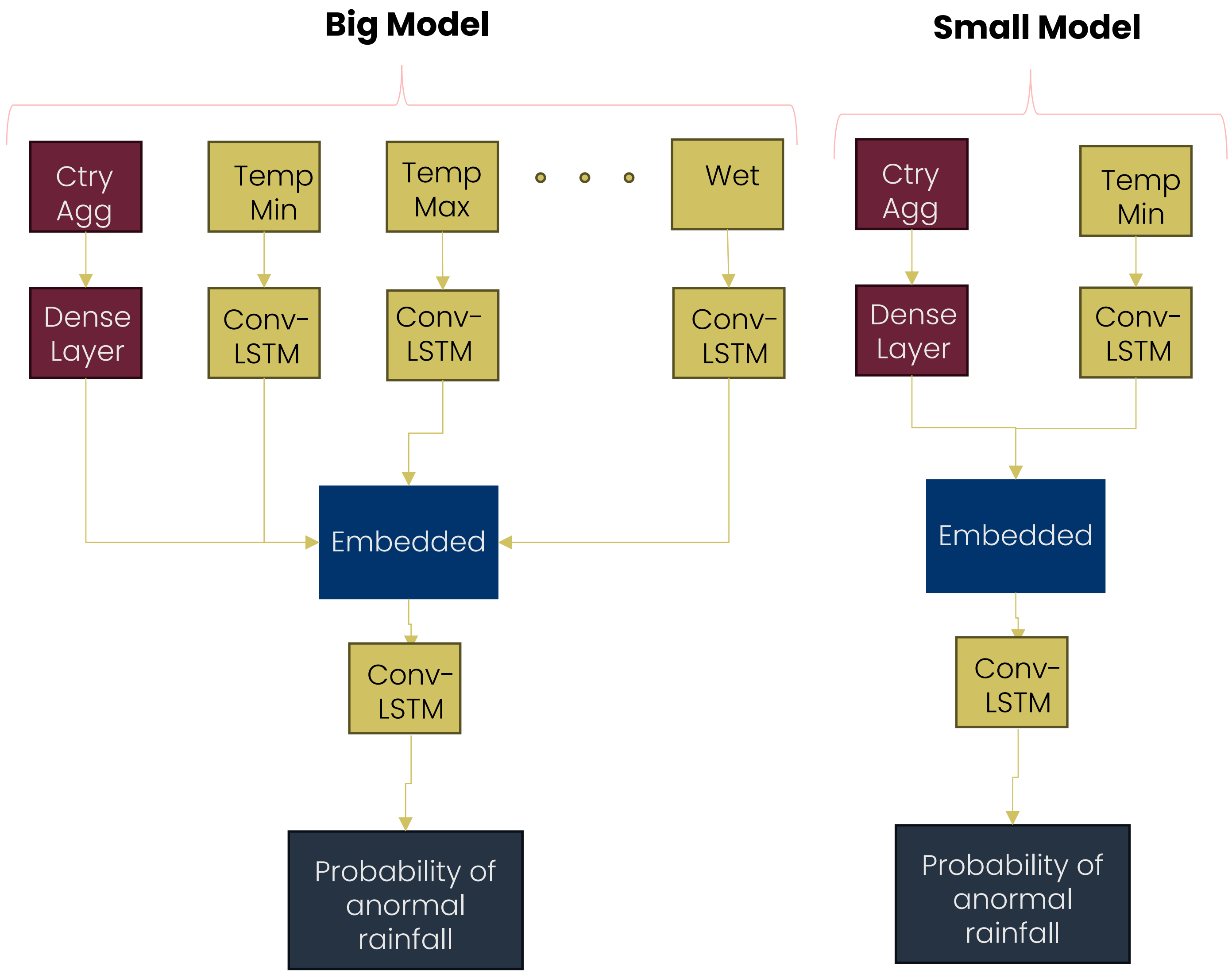
**Target creation:**

We used a 10 years rolling window to generate thresholds of the variable “Wet days” by region and then set to 1 the places where the value was above the calculated threshold.

* To be more precise, we used the distribution of the variable that was latter going to be used as target.

# MODEL

# ARCHITECTURE




**ConvLSTM Layer:**

We account for weather similarities between locations by using a Convolutional Layer. We also account for the autoregressive nature of weather by combining by adding an LSTM to the architecture.

**Dense Layer:**

In other to give weather context of nearby countries without having to incur in more computational complexity by expanding the latitude and longitude thresholds, we use tabular data of country weather aggregates in a simple Dense Layer.

# BENCHMARK

**Logit Model:**

To make sure that our approach is robust against more simple methodologies, we build a Logistic regression on the same target but without considering neither the time or proximity factors.

# CONCLUSIONS & NEXT STEPS

Conclusions:

**Multivariate works**

We showed that giving more context on weather situation yields better performing models.

**Complexity vs efficiency**

Our results suggest that the complex approach led to inaccurate results, probably due to miss specification of the model (too much history for the LSTM).

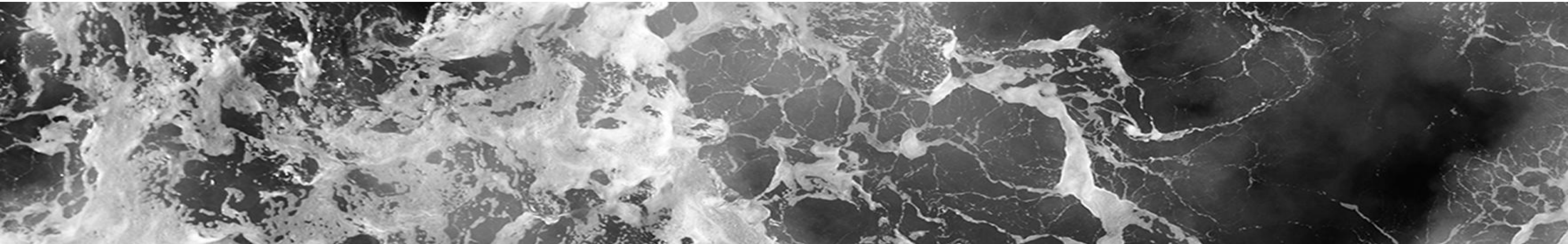
Next Steps:

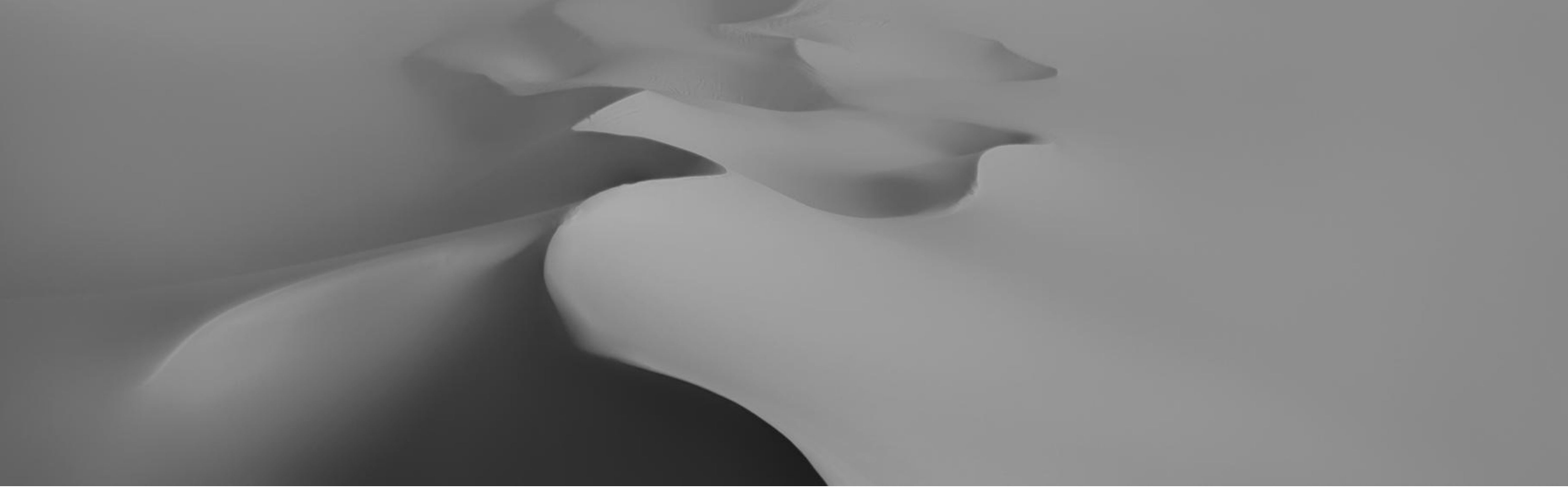
**Change lags used for LSTM**

For this version we used 8 years of data as lags for the LSTM (monthly data). That raised the complexity of the model without giving expected results.

**GNN approach**

Intuitively in our dataset there is a strong relation between records due to both time and proximity factors. This gives us the possibility to try a GNN approach.





# THANK YOU

Presented by Gabriel Santiago