The Disaster Risk Financing
Challenge Fund

SMART

A **S**tatistical, **Ma**chine **L**earning Framework for Parametric **R**isk **T**ransfer

Final Report

June 2021

Table of contents

1 Introduction

1.1 Project Summary

Parametric insurance is a promising disaster risk management strategy whereby prompt payouts are triggered whenever measurable indices exceed predefined thresholds. This promptness is critical in developing countries, as they are particularly exposed to short-term liquidity gaps that may overwhelm their capacity to cope with large disasters. From a machine learning perspective, this is essentially a classification rule for predicting loss or no loss based on the trigger variable. The rule is developed using past training sets of hazard and loss data (supervised learning).

Despite the advantages of simplicity, transparency and rapid payouts, there is a lack of confidence and reluctance to use parametric insurance because of basis risk – the misclassification of events due to false positives and false negatives. Basis risk leads to inefficient transactions and higher product costs, reducing their appeal to end-users and investors. It can also result in a failure to issue payouts when disaster events occur, which can have adverse consequences for insureds counting on rapid post-event funding. The problem of basis risk in parametric insurance is exacerbated by:

1. The use of ad-hoc parametric trigger rules that have not used rigorous statistical modelling to learn from past data and optimally exploit increasing amounts of data;
2. The lack of robust evaluation methods for understanding and quantifying parametric trigger performance.

The SMART project aims to address these issues through the application of appropriate machine learning and statistical concepts to develop a new framework for parametric trigger modelling. The project covers Thematic Area 2 of the Terms of Reference, entitled “Machine Learning and Big Data for Disaster Risk Financing”.

To demonstrate the methodology and its pathway to operationalization, a pilot study is made for multiple hazards in the Dominican Republic. The Dominican Republic currently has little reliance on risk transfer mechanisms despite being highly exposed to natural hazards, such as tropical cyclones, floods, and droughts. In 2015, the average annual loss due to natural hazards was estimated at USD 420 million. Productive sectors, particularly agriculture, tend to be severely affected, leading to socio-economic consequences and food insecurity for the country. As such, SMART pays particular attention to this aspect, covering the Focus Area of the Terms of Reference “Disaster Risk Financing Mechanisms to Manage Food Insecurity”.

1.2 Scope of the Report

The present report describes the main activities and outcomes of the SMART project, which started in mid-2019 and is now concluded. The document is organized as follows. Section 2 briefly describes both the climatological and socio-economic context of the Dominican Republic and the catastrophe insurance market. The research and development activities and results are presented in Section 3. Section 4 addresses the pathway to potential future operationalization of the project. Lastly, Section 5 covers the monitoring and evaluation indicators required to assess the success of the project.

2 Pilot country: the Dominican Republic

2.1 Climatological and socio-economic context

The Dominican Republic is located on the eastern part of the island of Hispaniola, one of the Greater Antilles, in the Caribbean region. Its area is approximatively 48,671 km². The central and western parts of the county are mountainous, while extensive lowlands dominate the southeast. There are four main mountain chains, from Northeast to Southwest: the Cordillera Septentrional, the Cordillera Central, the Sierra de Neyba, and the Sierra de Bahoruco, where the higher Antillean peaks (Duarte Peak, 3,087m and La Pelona, 3,085 m) are located. The mountain chains are separated by three major Northeast-Southwest tectonically controlled valleys. The physiographic configuration has a strong influence on the distribution of wet and dry areas in the country, producing very different environments between the windward and leeward side of the Cordillera Central (Izzo et al., 2010)

The climate of the Dominican Republic is classified as "tropical rainforest". However, due to its topography, the country's climate shows considerable variations over short distances. The average annual rainfall is approximately 1,500 mm. Along the eastern and southern coasts, there is a rainy season between late April and October, while the northern coast, which is exposed to the trade winds, is rainy throughout the year. However, the north-western coast experiences a decrease in rainfall from June to September.

The average annual temperature is about 25 °C, with January being the coldest month (average monthly temperature over the period 1901-2009 of about 22 °C) and August the hottest (average monthly temperature over the period 1901-2009 of about 26 °C) (World Bank, 2019)

Because of the North-eastern trade winds, which blow from November to March, the north facing slopes are usually wetter than those exposed to the south. However, the presence of mountain ranges can increase or decrease the effect of the trade winds, so much so that some areas have a very humid climate and lush vegetation, while others are nearly barren. For example, in Azua, on the coast of the Ocoa Bay, only 640 mm of rainfall per year, and even less at Lake Enriquillo, closed between the two Sierras; the south-west coast of the Pedernales province is quite arid as well.

From December to March, usually the country is affected by some frontal systems from the United States, which could bring clouds and rains, and a cool wind, able to lower the temperature by a few degrees. In these cases, the temperature can drop below 10 °C in the inland hilly areas, and around freezing in the mountains.

The Dominican Republic is exposed to hurricanes and tropical cyclones. Traditionally in the Caribbean the hurricane season starts on June 1st and ends on November 30th; in the Dominican Republic the most active months for tropical cyclones are usually August and September.

The Dominican Republic has a population of 10,6 million people. The country is divided into 31 provinces and 10 regions. Provinces are the first level administrative subdivisions of the country. Provinces are then divided into municipalities. Population is mainly located in the Santo Domingo province, followed by Santiago and Distrito National, where the capital city of Santo Domingo is situated (Figure 1 and Figure 2).

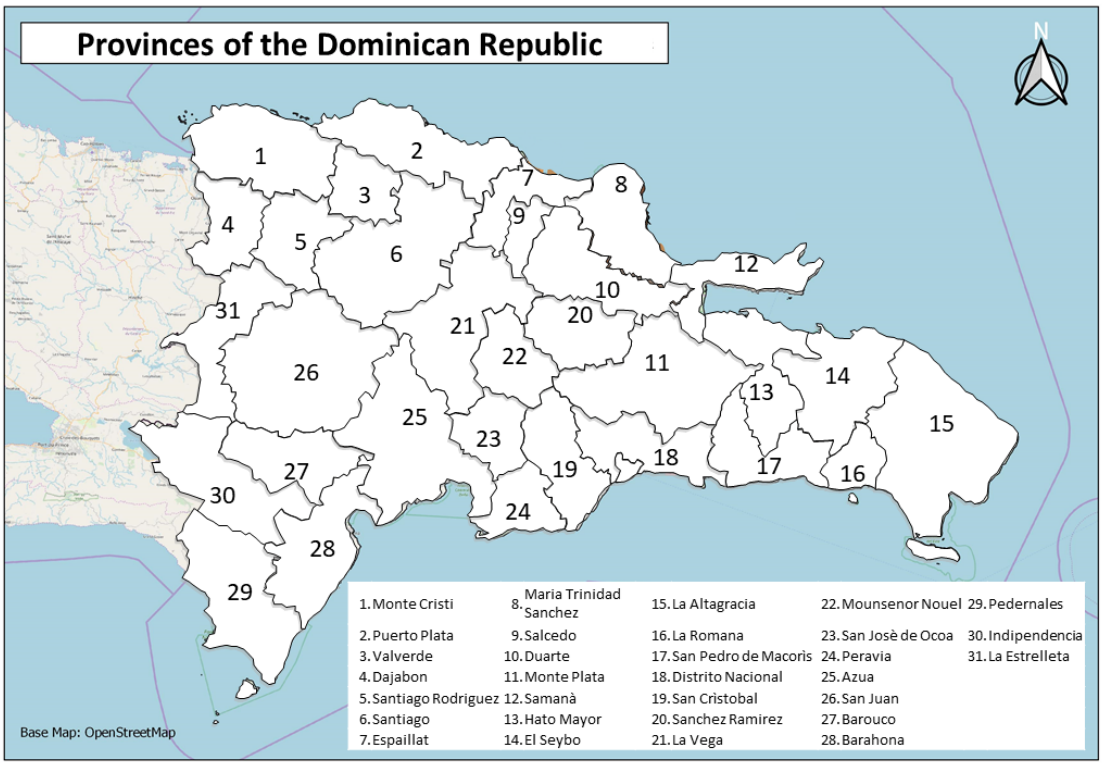

Figure 1: Provinces of the Dominican Republic

Figure 2: Administrative regions of the Dominican Republic

The Dominican Republic is classified as an upper-middle income developing country (World Bank, 2019a). The economy, which is the largest in the Caribbean and Central America region, depends on mining, agriculture, trade, and services. In recent years, the service sector, due to growth in tourism and free-trade zones, has overtaken agriculture as the leading employer of Dominicans. However, agriculture is still the most important sector in terms of domestic consumption and is in second place in terms of export earnings. Agriculture represents 11% of the GDP and employs nearly 15% of the population. Small farmers represent 72% of the total number of farmers, but account only for 28% of cultivated area (World Bank, 2012). The fertile Cibao Valley is the main agricultural centre. The Dominican Republic is the second-largest producer of sugarcane in the Caribbean after Cuba. Sugarcane is the nation's most important crop. Besides sugar, the Dominican Republic is also a major exporter of coffee, cocoa and tobacco. All these crops are cultivated in the Cibao Valley. Bananas and fresh fruit exports have increased in recent years. The Dominican Republic is the largest producer of organic bananas worldwide. The dairy sector is a key source of local employment, although it is not able to meet the demand of dairy products consumed domestically and in the hotels. Dominican milk is produced by 59,000 farmers who own about 1,200,000 cattle distributed mainly in the Northwest, Southwest and Eastern regions of the country.

Weather-related disasters have a significant impact on the economy of the Dominican Republic. Particularly, the agricultural sector is highly vulnerable to these natural hazards. The country is ranked as the 10^{th} most vulnerable in the world and the second in the Caribbean, as per the Climate Risk Index for 1997-2016 report (Eckstein et al., 2017). It has been affected by spatial and temporal changes in precipitation, sea level rise, and increased intensity and frequency of extreme weather events. Climate events such as droughts and floods have had significant impacts on all the sectors of the country's economy, resulting in socio-economic consequences and food insecurity for the country. Over the period from 1960 to present, the most frequent natural disasters were tropical cyclones (45% of the total natural disasters that hit the country), followed by flood (37%) according to the International Disaster Database EM-DAT. Floods, storms and droughts were the disasters that affected the largest amount of people and caused huge economic losses. In one case, hurricane George in 1998, caused the loss of about 55% of the agricultural system, partly from landslides and flash floods. 90% of all banana plantations in the area were destroyed. Rice, bananas and cassava, the most basic foods on the island, were hit hard. Large pastures for animals were destroyed as well as poultry and other necessities. The damage to farmland and agriculture would total out to about $434 million. In case of droughts, DesInventar dataset reports that in 1991 a widespread drought led to about 15 million US $ losses (United Nations Office for Disaster Risk Reduction, 1994). In 2010, during a severe drought, serious losses were reported all over the country. The cocoa and rice sector experienced a lack of water for irrigation due to the decreased storage capacity of dams; rice-producing regions reported a 30% reduction in rice production, while rice production at national level decreased between 10% and 15% (World Bank, 2013). In 2011, hundreds of quintals of plantain, yucca, and sweet potato were lost in the provinces of La Vega, Santiago, Valverde, Montecristi (Payano-Almanzar and Rodriguez, 2018). However, they underline that storms and floods produced more economical damages with respect to droughts over the 1970-2011 period.

2.2 Brief description of catastrophe insurance market

Even though the Dominican Republic is highly exposed to natural hazards, the insurance market is small and the proposed insurance policies are characterized by a low degree of sophistication. Insurance penetration in the country is low with respect to other Caribbean nations; in fact, it is limited to 1.21% of the GDP, while in Jamaica it is 5.03% of GDP and in Puerto Rico it reaches 15.02% of the GDP. The insurance market is characterized by the presence of 33 insurance companies, including two reinsurers. In 2013, the five leading companies had a cumulative market share of 80%, while 16 companies had a market share of less than 1%.
Currently, a traditional indemnity-based crop insurance product is offered by AGRODOSA (Asecuradora Agropecuaria Dominicana S.A.). AGRODOSA is a state-funded agricultural insurance company that provides insurance for about 20% of the loans provided by the state-funded rural bank, Banco Agricola (BAGRICOLA). The AGRODOSA insurance program has experienced many difficulties that affected the implementation of a broader sustainable agricultural insurance program, such as premium subsidy whereby 30-50% of the premiums for individual farmers are covered directly by the state. AGRODOSA insures an area of approximately 4,000 hectares of which 70% is voluntary and not linked to credit or a loan from Banco Agricola. It also requires insured farmers to follow specific guidelines regarding risk management of their crops. In the event of damage or loss of crops, the insured is indemnified based on the amount of the investment made up to the time that the loss occurred. Generally, it offers farmers two type of policies:

1. Yield-based multi-peril crop insurance (MPCI): the product covers damage to crops based on variance from historical recorded yields. However, the product is only available for those crops for which sufficiently detailed historical data are available (rice and green beans). The policy provides coverage for the major part of perils (excess rain, wind, hail, drought, flood, cyclone, hurricane, tornado, accidents and plant diseases and insect infestations). Fire is the only peril excluded. The MPCI coverage payout is determined when, “actual yield obtained by the insured on its insured unit falls below the guaranteed yield (which is usually set at a maximum of 70% of the normal average yield) determined for each county and crop season. The indemnities are subject to the application of deductibles equivalent to 10% of the total sum insured for drought and 5% for the remaining covered perils (USAID and REDDOM, 2013).
2. Standing Crop Insurance: In this case, the value of the plant is covered. This type of policy is used for bananas and plantains.

The body directly responsible for agricultural insurance within the Ministry of Agriculture is the General Directorate of Agricultural Risks (DIGERA). This groups public and private organism such as the Ministry of Agriculture (which is the coordinator), the Ministry of Finance, the Superintendency of Insurance, the Dominican Agroenterprise Board, Inc. (JAD), the National Council of Producers of the Reform Agrarian, the Dominican Chamber of Insurers and Reinsurers, the Dominican Association of Ranchers and Farmers, (ADHA) and the Dominican Agricultural Bank.

DIGERA is the entity in charge of regulating and promoting the agricultural subsidy granted by the Dominican government to small and medium-sized producers in the country to ensure their

crops. According to the law, any insurer can have the contribution of the State. DIGERA is supervised by the Superintendency of Insurance regarding the issuance, control, extension of insurance policies and contracts. Every year, DIGERA issues a resolution indicating the crops that will be supported with insurance subsidies purchased by producers. AGRODOSA absorbs the subsidy once the insurance policy.

DIGERA grants a subsidy of 50% of the cost of the policy, with the objective that the producers ensure their crops, infrastructure, or livestock stocks, and thus minimize losses due to abnormal variations in nature and change climate. Annual budget allocations exceed 165 million pesos Dominicans ($EU 3.6 million). The State contribution until 2017 was 11 million dollars, which constitutes the contribution to premiums. In 2018, an extraordinary contribution of 6 million dollars was requested to be consigned in the budget of 2019. The insurance system administered by AGRODOSA receives state support to cover between 25% and 50% of the premium, thus, the insured agricultural producer pays the remaining 50% to 75%.

In another work, at the end of 2012, USAID in cooperation with REDDOM designed the Climate Resiliency and Index Insurance (CRII) for Small Farmers in the Dominican Republic. The original goal of the program was to improve resilience to climate change, reduce disaster risk, and promote public-private partnerships. As part of this goal, the program supported the establishment of a commercially sustainable index insurance product. Dairy was selected as the most suitable agricultural sector to implement an index-based insurance program. REDDOM consulted 75 associations of dairy producers, including about 7,800 farmers, and found that 94% of the farmers believe that drought is the main risk for cattle operation and milk production in the Northwest region. In November 2014, the index insurance product was presented to insurers, financial institutions and farmers. It was based on satellite images of vegetation conditions (the Normalized Difference Vegetation Index, NDVI, was used to evaluate the availability of forage for cows). It could be considered as a turning point in the Dominican insurance market because: 1. It was the first index insurance product in the country; 2. It was the first private insurance product for the agricultural sector in the country; and 3. It was the first insurance product designed with the participation of all value chain stakeholders including farmers, insurer and delivery channel. However, due to the high cost of the premium, the product has never been operative on the market (USAID and REDDOM, 2013).

3 Research and development

The development of models supporting catastrophe insurance products should be based on a sound scientific foundation. This section presents the three lines of research that were pursued during the SMART project, having in mind their future application in the development of parametric insurance programs. The first line of research takes upon the challenge of using machine learning algorithms to identify extreme weather events with the purpose of reducing basis risk, the biggest challenge in parametric insurance. The ability to objectively detect such events can play a role in parametric insurance, not only from the improvement in their reliability, but also in the promptness of the payouts, which is key for the insured. The second line of research moves towards the predictions of time series with the aim of being able to evaluate reduction in production of milk. In the Dominican Republic, the dairy sector is very vulnerable to drought and most farmers (94%) pointed out to REDDOM that drought is affecting the daily production of milk. In terms of handling losses, almost 70% of producers have had to borrow from intermediaries to replenish lost cattle stocks and/or develop prevention initiatives (USAID and REDDOM, 2013). Therefore, models able to accurately predict milk production with a certain lead time at country level were developed. The two lines of research developed models performing tasks at national scale. This choice was dictated from the data available at the moment. In the final part of the project, we decided to work on a finer spatial resolution by means of crop models. These models are able to reproduce the behavior of crops depending on environmental and meteorological parameters. The advantage of using these models is their ability to produce yield estimates at higher spatial resolution rather than country-wide dependent on the resolution of the input data. These characteristics make the outcomes of this work very interesting for parametric insurance programs granting the possibility of designing a tailor-made product for the farmers.

3.1 Identification of floods and droughts

The initial undertaking of the project was to develop a methodology, which aims to assess the potential of machine learning for weather index insurance. To achieve this, we propose and apply a machine learning methodology that is capable of objectively identifying extreme weather events, namely flood and drought, in near-real time, using quasi-global gridded climate datasets derived from satellite imagery or a combination of observation and satellite imagery.

3.1.1 Datasets

3.1.1.1 *Environmental variables*

The data-driven nature of ML models implies that the results yielded are as good as the data provided. Thus, the effectiveness of the methods depends heavily on the choice of the input variables, which should be able to represent the underlying physical process (Bowden et al., 2005). The data selection (and subsequent transformation) therefore requires a certain amount of a priori knowledge of the physical system under study. For the purpose of this work, precipitation and soil moisture were used as input variables for both flood and drought. An excessive amount of rainfall is the initial trigger to any flood event (Barredo, 2007), while scarcity of precipitation is one of the main reasons that leads to drought periods (Tate and Gustard, 2000). Soil moisture is instead used as a descriptor of the condition of the soil. With the idea to implement a tool that can be exploited in the framework of parametric risk financing, we selected the datasets to retrieve the two variables according to five criteria:

1. Spatial resolution: a fine spatial resolution that takes into account the climatic features of the various areas of the considered country is needed to develop accurate parametric insurance products.
2. Frequency: the selected datasets should be able to match the duration of the extreme event that we need to identify. For example, in the case of floods, which are quick phenomena, daily or hourly frequencies are required.
3. Spatial coverage: global spatial coverage enables the extension of the developed approach to areas different from the case study region.
4. Temporal coverage: since extreme events are rare, a temporal coverage of at least 20 years is considered necessary to allow a correct model calibration.
5. Latency time: a short latency time (i.e., time delay to obtain the most recent data) is necessary to develop tools capable of identifying extreme-events in near-real time.

Based on a comprehensive review of available datasets, we found six rainfall datasets and one soil moisture dataset, comprising 4 layers, matching the above criteria. With respect to the studies analysed in (Fung et al., 2019; Hao et al., 2018; Mosavi et al., 2018) that associated a single dataset to each input variable, here six datasets are associated to a single variable (rainfall).

The use of multiple datasets is able to improve the ability of models in identifying extreme events, as demonstrated for example by (Chiang et al., 2007) in the case of flash floods. In addition, single datasets may not perform well; the combination of various datasets produces higher quality estimates (Chen et al., 2019)Two merged satellite-gauge products (the Climate Hazard Group Infrared precipitation, CHIRPS and the CPC Morphing technique, CMORPH,) and four satellite-only (the Global Satellite Mapping of Precipitation GSMaP, the Integrated Multi-Satellite Retrievals for GPM, IMERG; the Precipitation Estimation from Remotely-Sensed Information using Artificial Neural Networks, PERSIANN; and the Global PERSIANN Cloud Classification System, PERSIANN-CCS) datasets were used. The spatial variability of the annual average precipitation is displayed in **Figure 3**.

Soil moisture was retrieved from the ERA5 reanalysis dataset, produced by the European Centre for Medium Range Weather Forecast (ECMWF). The dataset provides information on 4 soil moisture layers (Layer 1: 0-7 cm, Layer 2: 7-28cm; Layer 3: 28-100cm; Layer 4: 100-289cm). Figure 2 displays the average amount of soil moisture for the four layers.

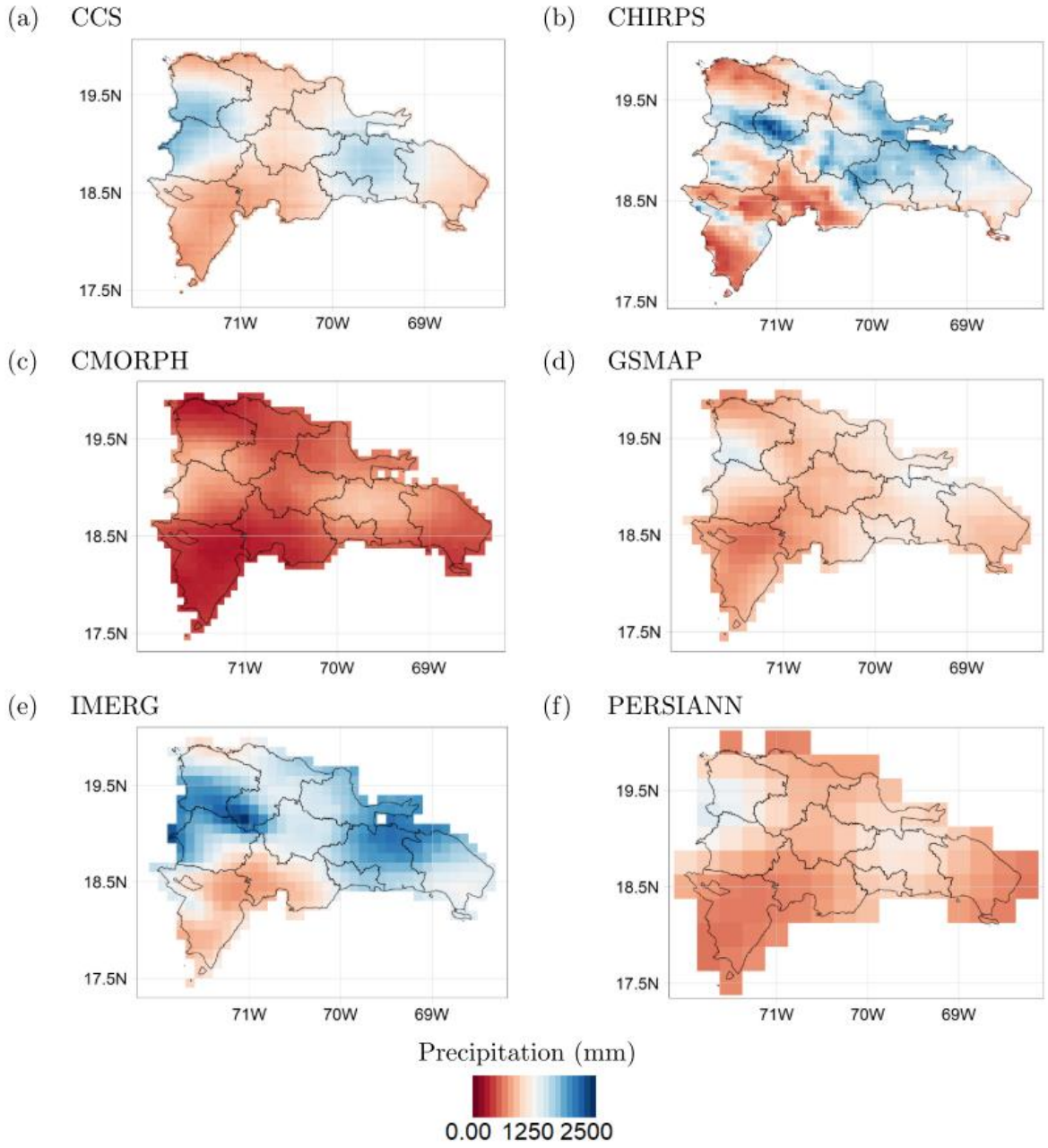

Figure 3: Average annual rainfall over the Dominican Republic according to the six considered datasets. (a) CCS, (b) CHIRPS, (c) CMORPH, (d) IMERG, (e) GSMaP, (f) PERSIANN.

The main features of the selected datasets are reported in **Table 1**.

Table 1: Main features of the selected datasets

Data set	Type	Resolution	Frequency	Coverage	Time span	Latency	Reference
CCS	Satellite	0.04°	1h	60°S–60°N	January 2003–present	Within 6 hours	(Hong et al., 2004)
CHIRPS/ CHIRP	Satellite-Gauge/ Satellite only	0.05°	Daily	50°S–50°N	January 1981-present	Within 3 weeks /Within 3 days	(Funk et al., 2015)
CMORPH	Satellite-Gauge	0.07278°	3h	60°S–60°N	January 1998–present	Within 14 days	(Joyce et al., 2004)
GSMaP	Satellite	0.1°	1h	60°S–60°N	March 2000–present	Within 12 hours	(Ushio and Kachi, 2010; Ushio et al., 2009)
IMERG (Late run)	Satellite	0.1°	30 min	60°S–60°N	June 2000-present	Within 12 hours	(Bolvin et al., 2018)
PERSIANN	Satellite	0.25°	1h	60°S–60°N	March 2000-present	Within 48 hours	(Sorooshian et al., 2000)
ERA-5	Reanalysis	0.25°	1h	Global	1979-present	Within 5 days	(ECMWF et al., 2018)

3.1.1.2 *Historical catalogue of events*

The performances of a ML model are strictly related to the data the algorithm is trained on, hence, the reconstruction of historical events (i.e., the targets), although time-consuming, is paramount to achieve solid results. Therefore, a wide range of text-based documents from multiple sources have been consulted to retrieve information on past floods and droughts that hit the Dominican Republic over the period from 2000 to 2019. International disasters databases, such as the world renowned EMDAT, Desinventar and ReliefWeb have been considered as primary sources.

The events reported by the datasets have been compared with the ones present in hazard-specific datasets (such as FloodList and the Dartmouth Flood Observatory) and in specific literature (Herrera and Ault, 2017; Payano-Almanzar and Rodriguez, 2018) to produce a reliable catalogue of historical events. Only events reported by more than one source were included in the catalogue. shows the past floods and droughts hitting the Dominican Republic over the period from 2000 to 2019. More details on the events can be found in Table 2 (floods) and Table 3 (droughts).

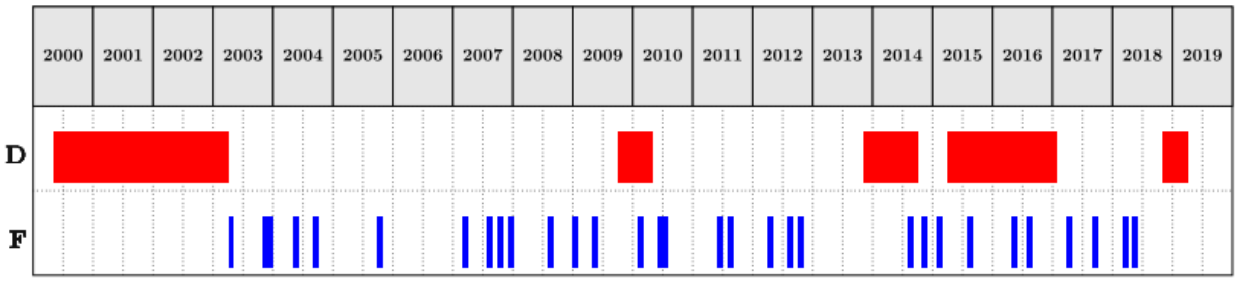

Figure 4: Time series of historical events

Table 2: Flood events in the Dominican Republic (2000-2019)

Event number	Start date	End date	type	Source
1	11/5/1981	12/5/1981	heavy rain	EM-DAT/Desinventar
2	14/09/1985	14/09/1985	heavy rain	DFO
3	30/05/1987	2/6/1987	heavy rain	DFO
4	22/09/1987	23/09/1987	tropical cyclone-Emily	EM-DAT
5	1/2/1988	2/2/1988	heavy rain	DFO
6	24/08/1988	26/08/1988	heavy rain	EM-DAT
7	2/11/1988	2/11/1988	heavy rain	Desinventar
8	18/09/1989	18/09/1989	tropical cyclone-Hugo	EM-DAT
9	22/04/1991	22/04/1991	heavy rain	EM-DAT
10	26/05/1992	27/05/1992	heavy rain	Desinventar
11	21/05/1993	21/05/1993	heavy rain	EM-DAT
12	26/05/1993	27/05/1993	heavy rain	Desinventar
13	18/08/1995	20/08/1995	storm	EM-DAT
14	6/10/1995	7/10/1995	storm	EM-DAT
15	10/9/1996	12/9/1996	tropical cyclone-Hortense	EM-DAT
16	22/10/1996	22/10/1996	heavy rain	EM-DAT
17	17/11/1996	23/11/1996	tropical cyclone	DFO
18	22/09/1998	24/09/1998	tropical cyclone-Georges	Glide/EM-DAT
19	23/10/1999	25/10/1999	tropical cyclone	DFO
20	6/10/2001	6/10/2001	tropical cyclone -Iris	EM-DAT
21	2/6/2002	6/6/2002	heavy rain	Glide/EM-DAT
22	18/04/2003	18/04/2003	heavy rain	Glide
23	14/11/2003	14/11/2003	coastal flood	Glide
24	6/12/2003	7/12/2003	tropical cyclone-Odette	DFO
25	23/05/2004	24/05/2004	heavy rain	DFO
26	1/9/2004	1/9/2004	tropical cyclone - Frances	EM-DAT
27	9/9/2004	10/9/2004	tropical cyclone -Ivan	EM-DAT

28	16/09/2004	19/09/2004	storm -Jeanne	Glide
29	13/05/2005	14/05/2005	heavy rain	EM-DAT
30	23/10/2005	23/10/2005	tropical cyclone -Alpha	EM-DAT
31	26/03/2007	26/03/2007	heavy rain	EM-DAT
32	31/05/2007	31/05/2007	heavy rain	DFO
33	28/10/2007	31/10/2007	tropical cyclone-Noel	DFO
34	12/12/2007	12/12/2007	tropical cyclone-Olga	DFO
35	15/08/2008	16/08/2008	storm-Fay	DFO
36	25/08/2008	28/08/2008	tropical cyclone-Gustav	Glide
37	2/9/2008	3/9/2008	tropical cyclone-Hanna	EM-DAT
38	20/05/2009	24/05/2009	heavy rain	Glide
39	20/06/2010	22/06/2010	heavy rain	Glide
40	17/07/2010	23/07/2010	heavy rain	EM-DAT
41	3/7/2011	5/7/2011	heavy rain	DFO
42	4/8/2011	4/8/2011	storm-Emily	EM-DAT
43	22/08/2011	24/08/2011	tropical cyclone-Irne	Glide/EM-DAT
44	25/04/2012	25/04/2012	heavy rain	EM-DAT
45	23/08/2012	26/08/2012	tropical cyclone-Isac	Glide/EM-DAT
46	23/10/2012	26/10/2012	tropical cyclone-Sandy	EM-DAT
47	22/08/2014	25/08/2014	storm-Cristobal	DFO
48	3/11/2014	3/11/2014	heavy rain	EM-DAT
49	28/08/2015	30/08/2015	tropical cyclone-Erika	Glide
50	23/09/2015	25/09/2015	heavy rain	EM-DAT
51	7/5/2016	9/5/2016	heavy rain	DFO
52	31/07/2016	1/8/2016	heavy rain	floodList
53	3/10/2016	5/10/2016	tropical cyclone-Matthew	floodList
54	5/11/2016	6/11/2016	heavy rain	DFO
55	23/04/2017	1/5/2017	torrential rain	DFO
56	16/05/2017	17/05/2017	heavy rain	floodList
57	21/09/2017	22/09/2017	tropical cyclone -Maria	floodList
58	4/5/2018	6/5/2018	heavy rain	floodList
59	10/7/2018	10/7/2018	tropical cyclone -Beryl	DFO

Table 3: Historical drought events in the Dominican Republic.

Event number	Start date	End date	Duration (days)	Source
1	Oct-1982	Apr-1983	182	Desinventar
2	May-84	May-1984	30	Desinventar
3	Apr-1985	Jul-1985	91	US Department of Agriculture
4	May-1990	Sep-1990	123	Desinventar
5	Jun-1991	May-1992	335	Desinventar
6	Nov-1993	Feb-1994	92	Desinventar

7	Jun-1994	Aug-1995	426	Desinventar
8	Mar-1997	Aug-1998	518	Desinventar
9	May-2000	Mar-2003	1034	Desinventar
10	Oct-2009	Apr-2010	182	Payano-Almanzar
11	Nov-2013	Sep-2014	304	Payano-Almanzar
12	Apr-2015	Jan-2017	641	Payano-Almanzar
13	Nov-2018	Mar-2019	120	GDACS

3.1.2 Methodology

As previously mentioned, in index insurance, payouts are triggered whenever measurable indices exceed predefined thresholds. Using machine learning, the rule that defines the issuance of a payout can be developed using past training sets of hazard and loss data (supervised learning). Conceptually, the development of a parametric trigger should correspond to an informed decision-making process i.e., a process which, based on data, a-priori knowledge and an appropriate modelling framework, can lead to optimal decisions and effective actions. This work aims to leverage the aptitude of machine learning, particularly supervised learning algorithms, to support the decision-making process in the context of parametric risk transfer, applying NN and SVM for the identification of extreme weather events, namely flooding and drought for this particular study. Consider the occurrence of losses caused by a natural hazard on each time unit $t = 1, \dots, T$ over a certain study area G, and let L_t be a binary variable defined as

$$L_t = \begin{cases} 0 & \textit{if loss occurs on day t } \ t \\ 1 & \textit{if loss doesn't occurr on day t} \end{cases} \quad (1)$$

The aim is then to predict the occurrence of losses based on a set of explanatory variables obtained from non-linear transformations of a set of environmental variables. This hybrid approach aims to capture some of the physical processes of how the hazard creates damage by incorporating a priori expert knowledge on environmental processes and damage-inducing mechanisms for different hazards. Raw environmental variables are not always able to fully describe complex dynamics like flood induced damage, therefore, the usage of expert knowledge is important to provide the machine learning model with input data that are able to better characterize the natural hazard events.

The models used for this task are set up such that they produce probabilistic predictions of loss rather than directly classifying events in a binary manner. This allows the parametric trigger to be optimized in a subsequent step, in a metrics-based, objective and transparent manner, by disentangling the construction of the model from the decision making regarding the definition of the payout-triggering threshold. Probabilistic outputs are also able to provide informative predictions of loss occurrence that convey uncertainty information, which can be useful for end users when a parametric model is operational (Figueiredo et al., 2018).

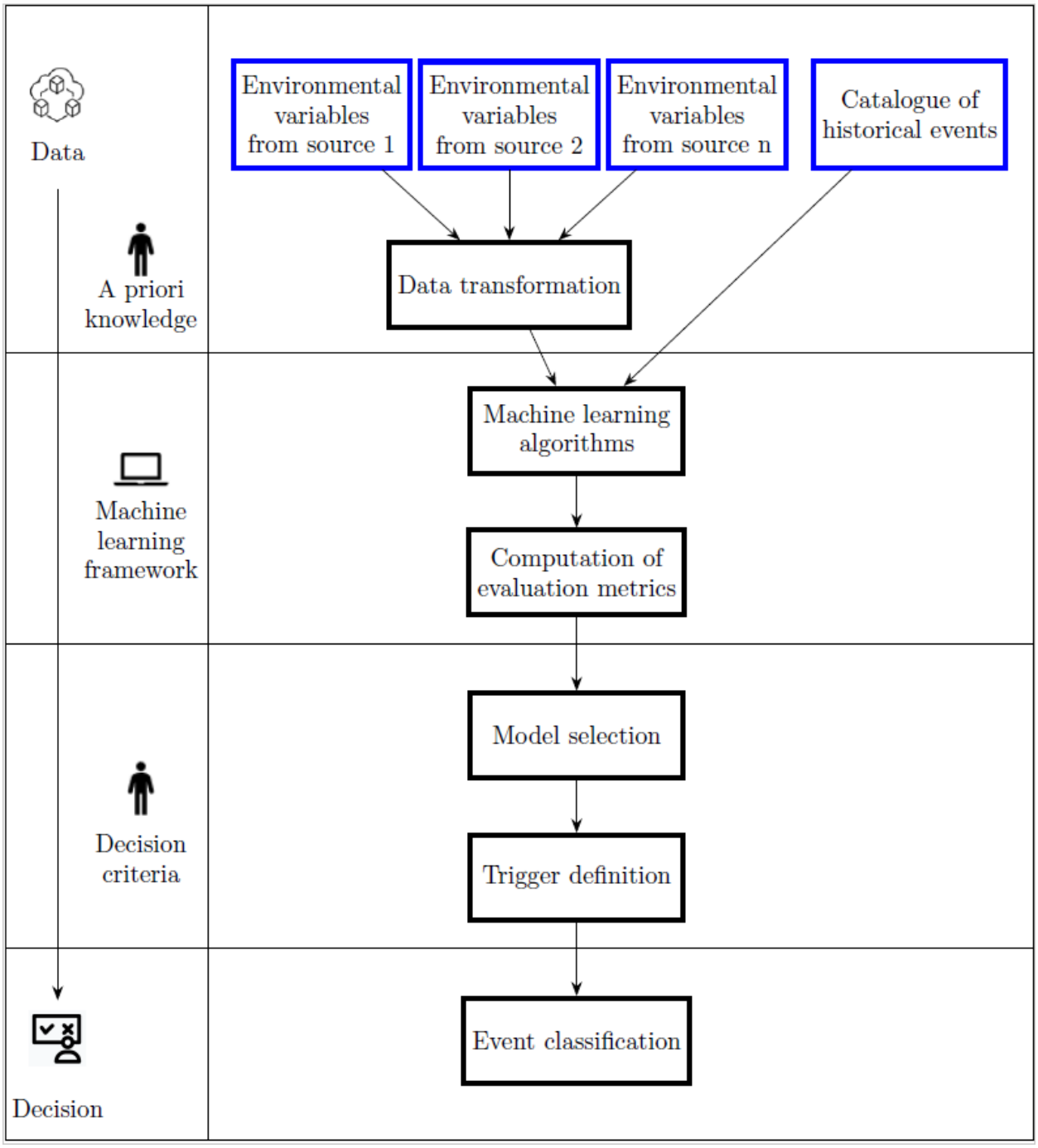

Figure 5: Flowchart summarizing the proposed methodology

3.1.2.1 *Data transformation*

Flood indicator

Flood damage is not directly caused by rainfall, but from physical actions originated by water flowing and submerging assets usually located on land. As a result, even if floods are triggered by rainfall, a better predictor for the intensity of a flood and consequent occurrence of damage is

warranted. To achieve this, we adopt a variable transformation to emulate, in a simplified manner, the physical processes behind the occurrence of flood damage due to rainfall, based on the approach proposed by (Figueiredo et al., 2018), which is now briefly described. Let $X_t(g_j)$ represent the rainfall amount accumulated over grid cell g_j belonging to G on day t. Potential runoff is first estimated from daily rainfall. This corresponds to the amount of rainwater that is assumed to not infiltrate the soil and thus remain over the surface, and is given by

$$R_t(g_j) = \max\{X_t(g_j) - u, 0\} \quad (2)$$

where u is a constant parameter that represents the daily rate of infiltration. Overland flow accumulates the excess of rainfall over the surface of a hydrological catchment. This process is modelled using a weighted moving time average, which preserves the accumulation effect and allows the contribution of rainfall on previous days to be considered.

The moving average is restricted to a three-day period. The potential runoff volume accumulated over cell g_j over days $t, t-1, t-2$ is thus given by:

$$R_t^*(g_j) = \theta_0 R_t(g_j) + \theta_1 R_{t-1}(g_j) + \theta_2 R_{t-2}(g_j) \quad (3)$$

where $\theta_0, \theta_1, \theta_2, > 0$ and $\theta_0, +\theta_1 + \theta_2 = 1$

Finally, let Y_t be an explanatory variable representing potential flood intensity for day *t*, which is defined as:

$$Y_t = \sum_{j=1}^{J} \frac{R_t^*(g_j)^{\lambda} - 1}{\lambda} \quad (4)$$

The Box-Cox transformation provides a flexible, non-linear approach to convert runoff to potential damage for each grid cell, which is summed over all grid cells in a study area to obtain a daily index of flood intensity. In order to obtain the Y_t variable that best describes potential flood losses due to rainfall, the transformation parameters u, θ_1, θ_2 *and* λ are optimised by fitting a logistic regression model to concurrent potential flood intensity and reported occurrences of losses caused by flood events, and maximising the likelihood using a quasi-Newton method:

$$L_t \sim Bernoull(p_t) \quad (5)$$

with

$$\log\left(\frac{p_t}{1-p_t}\right) = \beta_0 + \beta_1 Y_t \quad (6)$$

Drought indicator

Lack of precipitation is the first indicator used to assess drought conditions. Precipitation deficit is quantified through the use of the Standardized Precipitation Index (SPI). The SPI is one of the

most commonly employed drought indices all over the world, since it is recommended by the World Meteorological Organization to characterize meteorological drought, i.e. the drought that originates from a shortage in precipitation (WMO, 2012). The SPI, developed in (McKee et al., 1993) shows many advantages: it is based on precipitation only, is applicable in all climates and, being a standardized index, values of SPI for different climate regimes are comparable. However, SPI doesn't account for temperature, therefore the comparison of events with different temperature scenarios but the same SPI can be misleading (WMO and GWP, 2016). The SPI measures precipitation anomalies at a given location, based on a comparison of observed total precipitation amounts for an accumulation period of interest (e.g. 1, 3, 12, 48 months), with the long-term historic rainfall record for that period. The historic record is fitted to a probability distribution (the "gamma" distribution), which is then transformed into a normal distribution such that the mean SPI value for that location and period is zero (European Drought Observatory, 2020). Table 4 reports the drought classification according to the SPI.

Table 4: Drought Classification Based on SPI according to (McKee et al., 1993)

Category	SPI	Probability (%)
Extremely wet	2 and above	2.3
Severely wet	1.5 to 1.99	4.4
Moderately wet	1 to 1.49	9.2
Near normal	-0.99 to 0.99	68.2
Moderately dry	-1.49 to -1	9.2
Severely dry	-1.5 to -1.99	4.4
Extremely dry	-2 and below	2.3

Because SPI values are in units of standard deviation from the long-term mean, the indicator can be used to compare precipitation anomalies for any geographic location and for any number of timescales. Note that the name of the indicator is usually modified to include the accumulation period. Thus, SPI-3 and SPI-12, for example, refer to accumulation periods of three and twelve months, respectively (European Drought Observatory, 2020).

As was said before, the SPI is computed starting from rainfall. The six rainfall datasets presented in **Table 1** have been used to retrieve SPI.

3.1.2.2 *Machine learning*

Machine learning is a subset of artificial intelligence whose main purpose is to give computers the possibility to learn, throughout a training process, without being explicitly programmed (Samuel, 1959). It is possible to distinguish machine learning models based on the kind of algorithm that they implement and the type of task that they are required to solve. Algorithms may be divided into two broad groups: the ones using labelled data (Maini and Sabri, 2017)also known as supervised learning algorithms, and the ones that during the training receive only input data for

which the output variables are unknown (Ghahramani, 2004)also called unsupervised learning algorithms.

A concern when using machine learning models is overfitting. This phenomenon takes place when a model starts overlearning from a specific set of data that is used to train the model, hindering the ability of such model to generalise to samples that are outside the domain of data the model were trained on.

This section will focus on the machine learning algorithms adopted in this work, starting with a short introduction and description of their basic functioning, and next delving into the metrics used to evaluate the models are introduced and the reasoning behind their selection is highlighted. Lastly, will be explained the steps taken to fight overfitting.

Neural network

Neural networks are a machine learning algorithm composed by nodes (or neurons) that are typically organised into three types of layers: input, hidden and output. Once built, a neural network is used to understand and translate the underlying relationship between a set of input data (represented by the input layer) and the corresponding target (represented by the output layer). In recent years and with the advent of big data, neural networks have been increasingly used to efficiently solve many real-world problems, related for example with pattern recognition and classification of satellite images (Dreyfus, 2005) where the capacity of this algorithm to handle nonlinearity can be put to fruition (Stevens and Antiga, 2019). A key problem when applying neural networks is defining the number of hidden layers and hidden nodes. This must usually be done specifically for each application case, as there is no globally agreed-on procedure to derive the ideal configuration of the network architecture (Mas and Flores, 2008). Figure 6 displays the different parts composing a neural network and their interaction during the learning process. A neural network with multiple layers can be represented as a sequence of equations, where the output of a layer is the input of the following layer. Each equation is a linear transformation of the input data, multiplied by a weight (w) and the addition of a bias (b) to which a fixed nonlinear function is applied (also called activation function).

$$\begin{aligned} x_1 &= f(w_0 x_0 + b_0) \\ x_2 &= f(w_1 x_1 + b_1) \\ &\vdots \\ y &= f(w_n x_n + b_n) \end{aligned} \tag{7}$$

The goal of these equations is to diminish the difference between the predicted output and the real output. This is attained by minimising a so-called loss function through the fine tuning of the parameters of the model, the weights. The latter procedure is carried out by an optimiser, whose job is to update the weights of the network based on the error returned by the loss function.

The iterative learning process can be summarised by the following steps:

1. Start the network with random weights and bias,
2. Pass the input data and obtain a prediction,
3. Compare the prediction with the real output and compute the loss function, which is the function that the learning process is trying to minimise.

4. Backpropagate the error, updating each parameter through an optimiser according to the loss function.
5. Iterate the previous step until the model is trained properly. This is achieved by stopping the training process when either the loss function is not decreasing anymore or when a monitored metric has stopped improving over a set amount of definition.

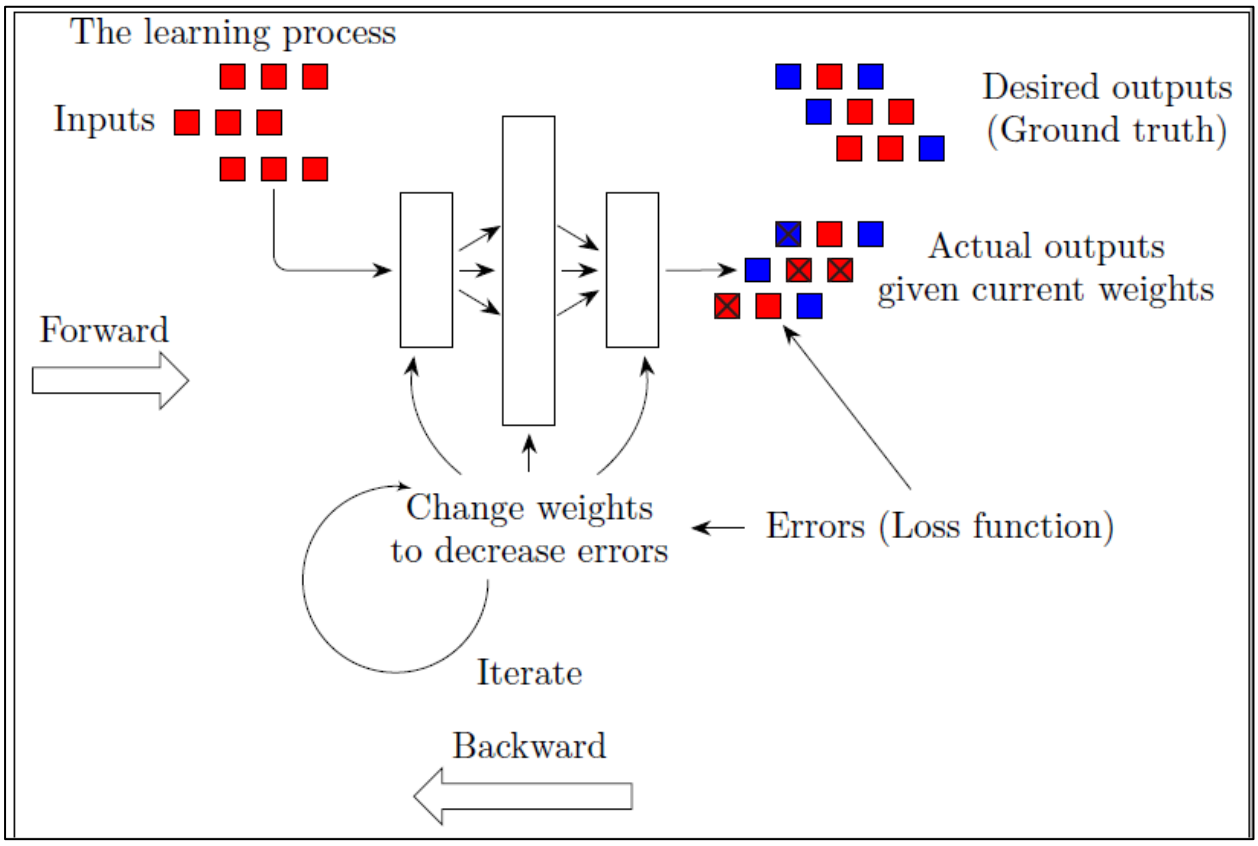

Figure 6: Learning process of a neural network (Stevens and Antiga, 2019)

Support Vector Machine

Support vector machine is a supervised learning algorithm used mainly for classification analysis. It constructs a hyperplane (or set of hyperplanes) defining a decision boundary between various data points representing observations in a multidimensional space. The aim is to create a hyperplane that separates the data on either side as homogeneously as possible. Among all possible hyperplanes, the one that creates the greatest separation between classes is selected. The support vectors are the points from each class that are the closest to the hyperplane (Wang, 2005). In parametric trigger modelling, as in many other real-world applications, the relationships between variables are non-linear. A key feature of this technique is its ability to efficiently map the observations into a higher dimension space by using the so-called kernel trick. As a result, a non-linear relationship may be transformed into a linear one. A support vector machine can also be used to produce probabilistic predictions. This is achieved by using an appropriate method such as Platt scaling (Platt, 1999), which transforms its output into a probability distribution over classes by fitting a logistic regression model to a classifier's scores. In this work, the support vector

machine algorithm was implemented using the C-support vector classification (Boser et al., 1992) formulation implemented with the scikit-learn package in python (Pedregosa et al., 2011). Given training vectors $x_i \in R_i^p = i, \ldots, 1$ and a label vector $y \in \{0,1\}^n$, this specific formulation is aimed at solving the following optimisation problem:

$$\min(w, b, \xi) = \frac{1}{2}\omega^t\omega + C\sum_{i=1}^{l}\xi_i$$
$$subject\ to\ y_1(\omega^t\Psi(x_i) + b) \geq 1 - \xi_i \quad (8)$$
$$\xi_i \geq 0, i = 1, \ldots, l$$

where ω and b are adjustable parameters of the function generating the decision boundary, Ψ_i is a function that projects x_i into a higher dimensional space, ξ_i is the slack variable and C > 0 is a regularisation parameter, which regulates the margin of the decision boundary allowing an increasing number of misclassifications for lower value of C and decreasing number of misclassifications for higher C (Figure 7).

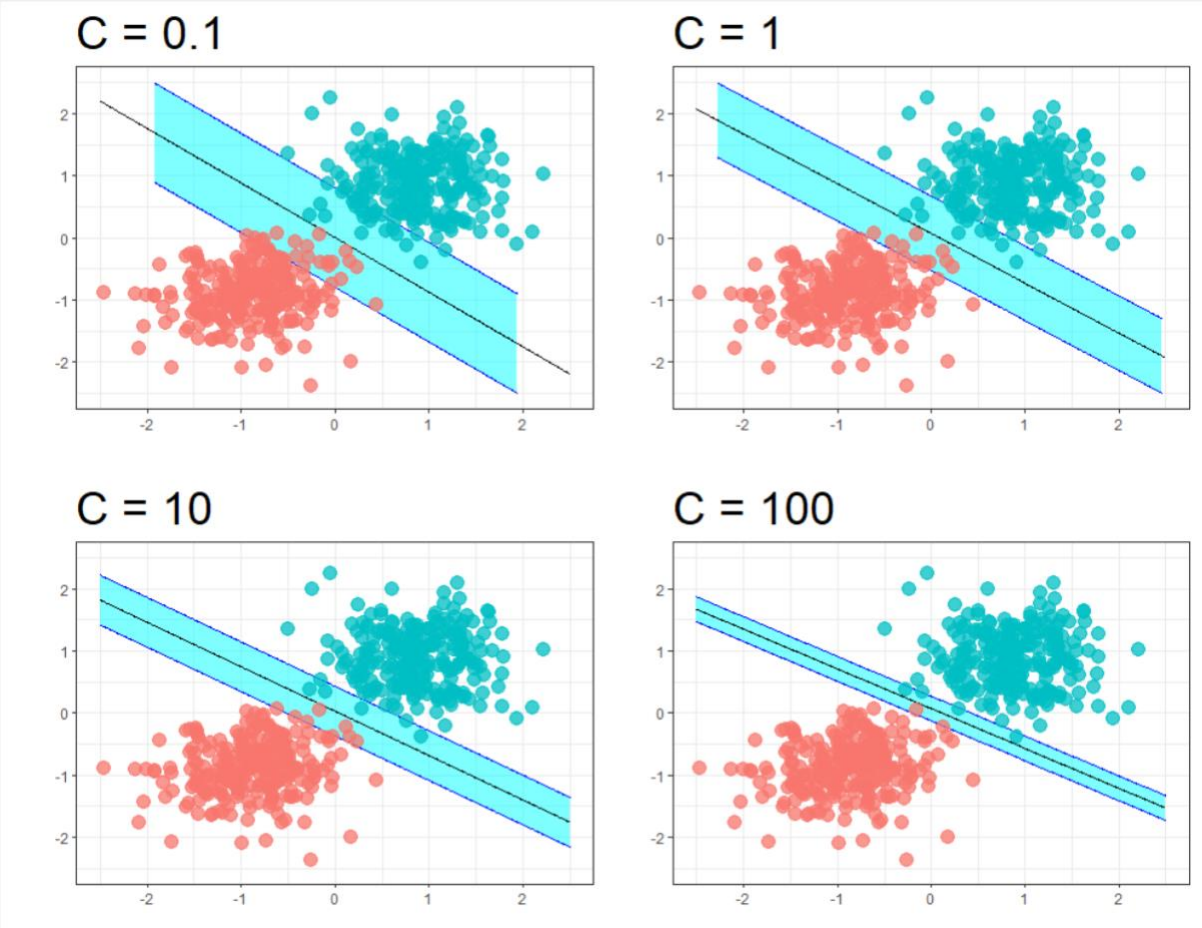

Figure 7: Decision boundary of support vector machine's algorithm, with changing regularization parameter C.

3.1.2.3 *Model construction*

The establishment of a robust chain of model construction is paramount when assembling the machine learning models. The algorithms presented in the previous section are built with several parameters, each with their own range of values. In order to find the parameters achieving the best performing model configuration we proposed a procedure that revolves around the following steps:

1. Pre-processing of the data
2. Selection of appropriate metrics evaluation
3. Search of the optimal parameters for a given model

3.1.2.4 *Pre-processing of the data*

For the case at hand, the preprocessing operations were split into three categories: data partitioning, feature scaling and the adoption of resampling techniques aimed at dealing with class imbalance. The first two are crucial for the development of a valid model, while the latter is required when dealing with the classification of rare events.

The partitioning of the dataset into training, validation and testing portions is fundamental to give the model the ability to learn from the data and avoid a problem often encountered in ML application: overfitting. This phenomenon takes place when a model starts overlearning from the training dataset, picking up patterns that belong solely to the specific set of data it is training on and that are not depictive of the real-world application at hand, making the model unable to generalize to sample outside this specific set of data. To avoid overfitting one should split the data into at least 2 parts (McClure, 2017). The training set, upon which the model will learn, and a validation dataset functioning as a counterpart during the training process of the model, where the losses obtained from the training set and those obtained from the validation set are compared to avoid overfitting. A further step would be to set aside a testing set of data that the model has never seen. Evaluating the performances of the model using data that it has never encountered before, is an excellent indicator of its ability to generalize. Thus, the splitting of the data is key to the validation of the model. In this work, the training of the NN was carried out splitting the dataset in 3 parts: training (60%), validation (15%) and testing (25%) set. During training, the neural networks used only the training set, evaluating the loss on the validation set at each iteration of the training process. After the training, the performance of the model was evaluated on the testing set that the model has never seen. Concerning the SVMs, a k-fold cross validation (Mosteller and Tukey, 1968) was used to validate the model, using 5 folds created by preserving the percentage of sample of each class, the algorithm was therefore trained on 80% of the data and its performances were evaluated on 20% of the remaining data that the model had never seen.

Feature scaling is a procedure aimed at improving the quality of the data by scaling and normalizing numeric values so as to help the ML model in handling varying data in magnitude or unit (Aksoy and Haralick, 2001). The variables are usually rescaled to the [0, 1] range or to the [−1, 1] range or normalized subtracting the mean and dividing by the standard deviation. The scaling is carried on after the splitting of the data and is usually calibrated over the training data, and then, the testing set is scaled with the mean and variance of the training variables (Mueller and Massaron, 2016).

Lastly, when undertaking a classification task, particular attention should be put on addressing class imbalance, which reflects an unequal distribution of classes within a dataset. Imbalance

means that the number of data points available for different classes is significantly different; if there are two classes, a balanced dataset would have approximately 50% points for each of the classes. For most machine learning techniques, little imbalance is not a problem, but when the class imbalance is high, e.g., 85% points for one class and 15% for the other, standard optimization criteria or performance measures may not be as effective as expected (Garcia et al., 2012). Extreme events are by definition rare, hence, the imbalance existing in the dataset should be addressed. One approach to address imbalances is using resampling techniques such as over-sampling (Lee et al., 2008) and SMOTE (Chawla et al., 2002). Over-sampling is the process of up-sampling the minority class by randomly duplicating its elements. SMOTE (Synthetic Minority Over-sampling Technique) involves the synthetic generation of data looking at the feature space for the minority class data points and considering its k nearest neighbor where k is the desired number of synthetic generated data. Another possible approach to address imbalances is weight balancing, which restores balance in the data by altering the way the model "looks" at the under-represented class. Oversampling, SMOTE and class weight balancing were the resampling techniques deemed more appropriate to the scope of this work, namely, identifying events in the minority class.

3.1.2.5 *Evaluation metrics*

A reliable tool to objectively measure the differences between model outputs and observations is the confusion matrix. Table 5 shows a schematic confusion matrix for a binary classification case.

Table 5: Confusion matrix

	Event observed		
Event predicted	Yes	No	Total
Yes	*a* (True Positive)	*b* (False Positive)	$a + b$
No	*c* (False Negative)	*d* (True Negative)	$c + d$
Total	$a + c$	$b + d$	$a + b + c + d = n$

When dealing with thousands of configurations and, for each configuration, with an associated range of possible threshold probabilities, it is impracticable to manually check a table or a graph for each setup of the model. Therefore, a numeric value, also called evaluation metric, is often employed to synthesise the information provided by the confusion matrix and describe the capability of a model (Hossin and Sulaiman, 2015). There are basic measures that are obtained from the predictions of the model for a single threshold value (i.e., value above which an event is considered to occur). These include the precision, sensitivity, specificity, and false alarm rate, which take into consideration only one row or column of the confusion matrix, thus overlooking other elements of the matrix (e.g., high precision may be achieved by a model that is predicting a high value of false negatives). Nonetheless, they are staples in the evaluation of binary classification, providing insightful information depending on the problem addressed. Accuracy and F1 score, on the other hand, are obtained by considering both directions of the confusion matrix, thus giving a score that incorporates both correct predictions and misclassifications. The accuracy is the ratio between the correct prediction over all the instances of the dataset, and is able to tell how often, overall, a model is correct. The F1 score is the harmonic mean of precision and

recall. In its general formulation derived from (Jones and Van Rijsbergen, 1976)'s effectiveness measure, one may define a F_β score for any positive real β (Equation 9):

$$F_\beta = 1 + \beta^2 \frac{precision * sensitivity}{(\beta^2 * precision) + sensitivity} \quad (9)$$

where β denotes the importance assigned to precision and sensitivity. In the F1 score both are considered to have the same weight. For values of β higher than one more significance is given to false negatives, while β lower than one puts attention on the false positive. The goodness of a model may also be assessed in broader terms with the aid of Receiver Operating Characteristic (ROC) and Precision-Sensitivity curves (PS). The ROC curve is widely employed and is obtained plotting the sensitivity against the false alarm rate over the range of possible trigger thresholds (Krzanowski and Hand, 2009). The PS curve, as the name suggests, is obtained plotting the precision against the sensitivity over the range of possible thresholds. For this work, the threshold corresponds to the range of probabilities between 0 and 1. These methods allow evaluating a model in terms of its overall performance over the range of probabilities, by calculating the so-called area under curve (AUC). It should be noted that both ROC curve and the accuracy metric should be used with caution when class imbalance is involved (Saito and Rehmsmeier, 2015) as having a large amount of true negative tends to result in low value of FPR (or 1- specificity). Table 6 summarises the metrics described above used in this paper to evaluate model performances.

Table 6: metrics used in the work

Metric	Equation
Accuracy	$(a+d)/n$
Precision	$a/(a+b)$
Sensitivity (Recall)	$a/(a+c)$
False alarm rate	$b/(b+d)$
F1 score	$2 * \frac{Precision*Sensitivity}{Precision+Sensitivity}$
AUC under the ROC curve	$\int_0^1 ROC(t)dt$
AUC under the PS curve	$\int_0^1 PS(t)dt$

The best settings of the ML algorithms were selected based on the highest values of F1 score and area under the PS curve, the predictive performances of the models were compared to those of logistic regression (LR) models. The logistic regression adopted as a baseline takes as input multiple environmental variables, in line with the procedure followed for the ML methods and used a logit function (Equation 7) as link function, neglecting interaction and nonlinear effects amid predictors. The logistic regression is a more traditional statistical model whose application to index insurance has recently been proposed, and can be said to already represent in itself an improvement over common practice in the field (Calvet et al., 2017; Figueiredo et al., 2018). Thus, this comparison is able to provide an idea about the overall advantages of using a ML method.

3.1.2.6 *Search for optimal model parameters*

In order to explore the domain of possible model configurations, for each ML method multiple key aspects were tested. Both methods shared an initial investigation of the sampling technique and the combination of input datasets to be fed into the models; all the data resampling techniques previously introduced were tested along with the data in their pristine condition where the model tries to overcome the class imbalance by itself. All the possible combinations of input dataset were tested starting from one dataset for SVM and with two datasets for NN up to the maximum number of environmental variables used. The latter procedure can be used to determine whether the addition of new information is beneficial to the predictive skill of the model and also to identify which datasets provide the most relevant information.

Table 7: Summary of parameters used to derive the domain of model configurations

Model	Parameter	Flood	Drought
NN	Input dataset combinations	61 combinations of environmental variables	61 combinations of environmental variables 4 SPI (1,3,6,12)
	Sampling	Unweighted Class Weight Over-sampling SMOTE	Unweighted Class Weight
	Loss	Binary Cross Entropy	Binary Cross Entropy
	Optimizer	ADAM	ADAM
	Number of layers & nodes	Layers: $[1;9]$ Nodes: $2^{nl+1} : 2^{nl+9}(*)$	Layers: $[1;9]$ Nodes $2^{nl+1} : 2^{nl+9}(*)$
	Activations	ReLu Tanh	ReLu Tanh
Number of Configurations		4392	8784
SVM	Input dataset combinations	67 combinations of environmental variables	67 combinations of environmental variables 4 SPI (1,3,6,12)
	Sampling technique	Unweighted Class Weight Over-sampling SMOTE	Unweighted Class Weight
	C-Regularization parameter	$C = (0.1, 1, 10, 100, 500)$	$C = (0.1, 1, 10, 100, 500)$
	Kernel Function	Linear Polynomial Radial Basis	Linear Polynomial Radial Basis
Number of Configurations		4020	8040

$(*)$: nl: number of layers

As previously discussed, these models present a multitude of customisable facets and parameters. For the support vector machine, the regularisation parameter C and the kernel type were the elements chosen as the changing parts of the algorithm. Five different values of C were adopted, starting from a soft margin of the decision boundary moving towards narrower margins, while three kinds of kernel functions were used to find the separating hyperplane: linear, polynomial and radial. The setup for a neural network is more complex and requires the involvement of more parameters, namely, the LF and the optimisers 360 concerning the training process, plus, the number of layers and nodes and the activation functions as key building blocks of the model architecture. Each of the aforementioned parameters can be chosen among a wide range of options; moreover, there is not a clear indication for the number of hidden layers or hidden nodes that should be used for a given problem. Thus, for the purpose of this study, the intention was to start from what was deemed the "standard" for the classification task for each of these parameters, deviating from these standard criteria towards more niche instances of the parameters trying to cover as much as possible of the entire domain.

Table 7 summarise the range of parameters explored for both ML methods and both hazards.

3.1.3 Results

The results are presented in this section separating the two types of extreme events investigated, flood and drought. For each hazard, the results are presented by introducing at first the models achieving the highest value of the F1 score for a given configuration and threshold probability (i.e., a point in the ROC or precision-sensitivity space). Secondly, the best performing model configurations for the whole range of probabilities according to the AUC of the precision-sensitivity curve are presented and discussed. The reasoning behind the selection of these metrics is discussed previously, in Section 3.1.2.5. As described in the same section, the performances of the ML algorithms are evaluated through a comparison with a LR model.

3.1.3.1 *Flood*

The flood case presented a strong challenge from the data point of view. Inspecting the historical catalogue of events, the case study reported 5516 days with no flood events occurring and 156 days of flood, meaning approximately a 35:1 ratio of no event/event. This strong imbalance required the use of the data augmentation techniques presented in Section 3.1.2.4. The neural network settings returning the highest F1 score were given by the model using all ten datasets applying an over-sampling to the input data. The network architecture was made up of 9 hidden layers with the number of nodes for each layer as already described, activated by a ReLu function. The LF adopted was the binary cross entropy and the weights update was regulated by an Adam optimiser. The highest F1 score for the support vector machine was attained by the model configuration using an unweighted model taking advantage of all ten environmental variables with radial basis function as kernel type and a C parameter equal to (i.e., harder margin).

Figure 8 shows the predictions of the three methods over the testing set (i.e., never seen by the model) in chronological order. A translation of **Figure 8** into numeric values can be found in **Table 8**. Overall, the two ML methods outperform decisively the logistic regression with a slightly higher F1 score for the neural network.

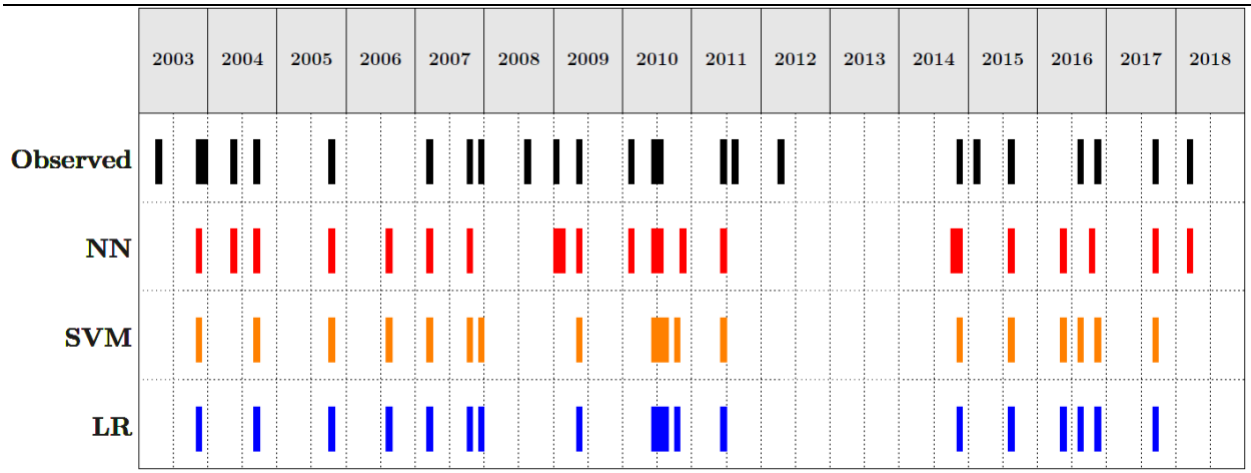

Figure 8: Predictions of the three methods for the flood case

Table 8: Metrics flood case

Method	Precision	Sensitivity	Specificity	F1 score	Accuracy
NN	0.57	0.57	0.99	0.57	0.98
SVM	0.63	0.49	0.99	0.55	0.98
LR	0.46	0.42	0.99	0.43	0.97

In **Figure 9**, panel (a) the highest F1 scores by method are reported in the precision-sensitivity space along with all the points belonging to the top 1% configurations according to F1 score. The separation between the ML methods and the logistic regression can be appreciated, particularly when looking at the emboldened dots in **Figure 9**a representing the highest F1 score for each method. Also, the plot highlights a denser cloud of orange points in the upper left corner and denser cloud of red points in the lower right corner attesting, on average, a higher sensitivity achieved by the NNs and a higher precision by the SVMs. **Figure 9**, panel (b), depicts the goodness of NN and SVM versus the LR model, showing how the F1 scores of the best-performing settings for each of the three methods vary by increasing the number of input datasets. This plot shows that the SVM and LR models have similar performances up to the second layer of soil moisture, while NN performs considerably better overall. The NN and the SVM as opposed to the LR, show an increase in the performances of the models with increasing information provided. The LR seems to plateau after 4 rainfall dataset and the improvements are minimal after the first layer of soil moisture is fed to the model. This would suggest, as expected, that the ML algorithms are better equipped to treat larger amounts of data.

Figure 10 presents the best-performing configurations according to the area under the PS curve. For neural network, this configuration is the one that also contains the highest F1 score, whereas for support vector machine the optimal configuration shares the same feature of the one with the best F1 score with the exception of a softer decision boundary in the form of C equal to 100. The results reported in **Figure 10** (a) and (b) about the best-performing configurations are further confirmation of the importance of picking the right compound measurement to evaluate the

predictive skill of a model. In fact, according to the metrics using the true negative in their computation (i.e.,

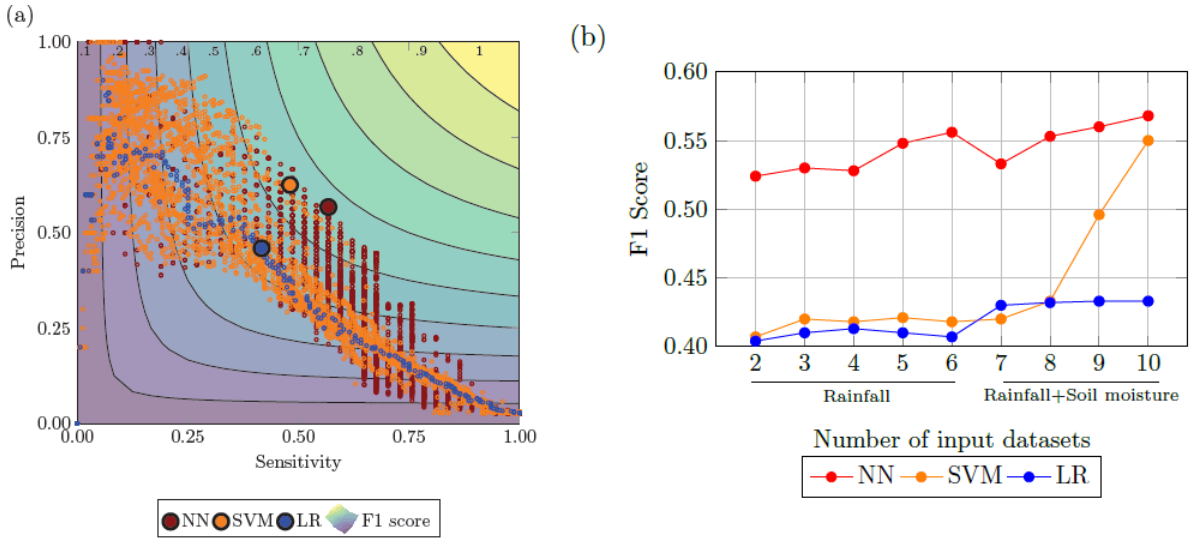

Figure 9: (a): Performance evaluation for the flood case: (a) Performances of the top 1% configurations in the precision-sensitivity space highlighting the highest F1 score, (b): Comparison of ML methods with LR with combination using increasing number of input datasets.

specificity, accuracy and ROC) , one may think that these models are rather good, and this deceitful behaviour is not scaled appropriately for very bad models. The aim of this work is to correctly identify a flood event rather than being correct when none occur, hence, overlooking the correct rejections seems reasonable. Panel (a) and (b) of **Figure 10** shed a light on the inaccuracy of the ROC curve and the relative area under the curve (AUC). On the right are displayed the ROC curves, whilst on the left the PS curves of the ideal configurations for each method according to the highest AUC. The points in both curves represent a 0.01 increment in the trigger probability. The receiving operator curve indicates the NN as the worst model being the closest to the 45° line and having, along with SVM, a lower AUC with respect to the logistic model. This signal is strongly contradicted by other metrics and the precision-sensitivity curve, where the red dots are the closest to the upper-right corner where the perfect model resides. The behaviour of these curves is linked, once again, to the disparity in the classes. Additionally, looking at panel (a), all models are pretty distant from the always-positive classifier (i.e., a baseline independent from class distribution represented by the black hyperbole in bold) more appropriate as a baseline to beat than a random classifier (Flach and Kull, 2015).

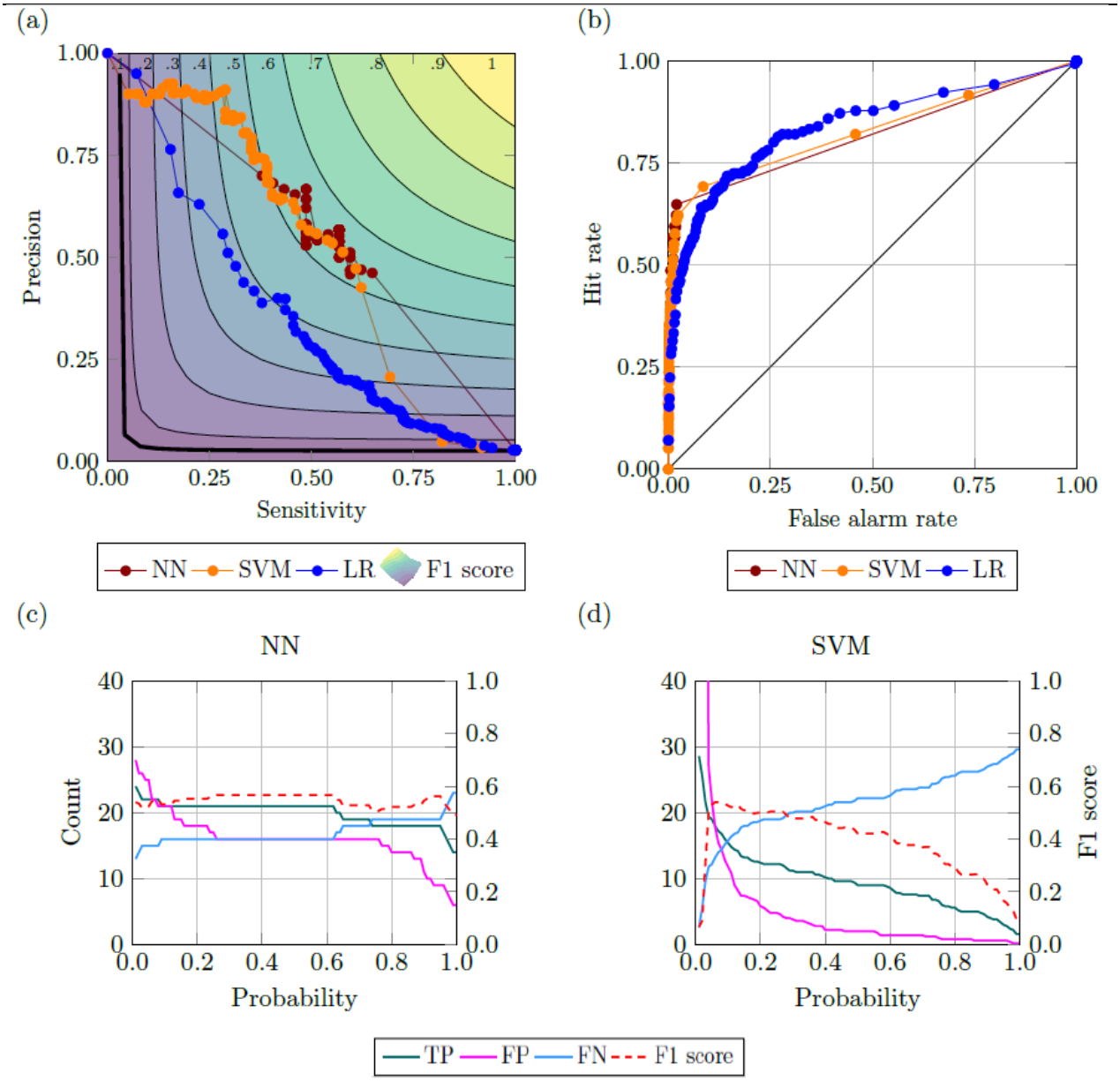

Figure 10: Best-performing configurations for the flood case: (a) PS curve, (b) ROC curve, (c) Variation of true positive, false positive, true negative and F1 score for the range of probability in NN and (d) in SVM.

Figure 11 portrays the properties of the top one-percent model configuration for both methods according to the area under the PS curve. Neural networks prefer the adoption of oversampling to enhance the input data and almost 60% of the configurations use a rectified linear unit function to activate its layer. Relative to the architecture of the network, a double peak can be observed at 8 and 9 layers, where the best-performing configurations can be found but it is noticeable an even larger presence of model configuration with 3 and 4 layers. On the other hand, support vector machines use the highest value of the C-parameter, which is the one used by the configuration attaining the highest F1 score. A bigger divide can be observed amid the sampling technique and

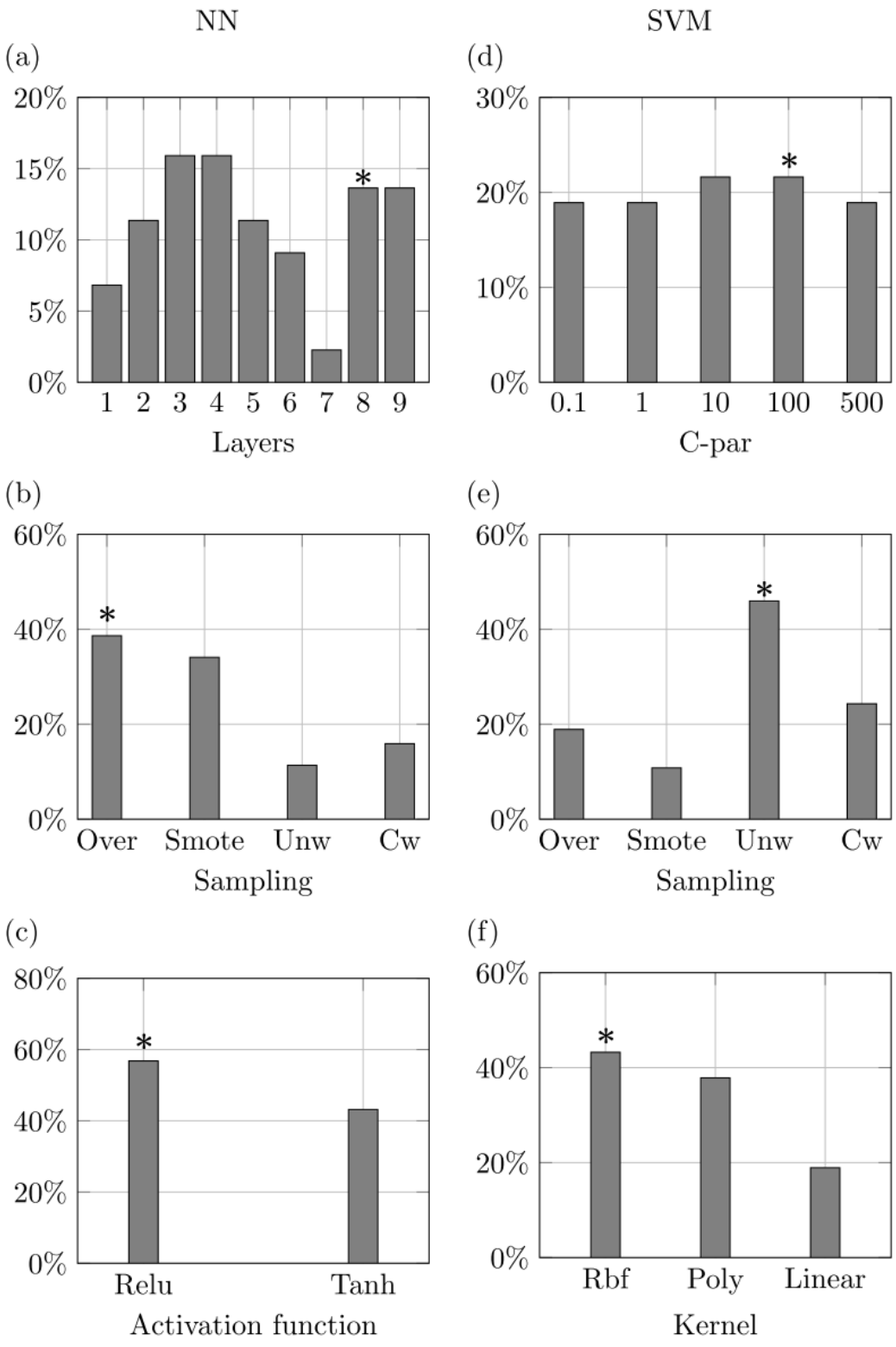

Figure 11: Properties of top 1% model configurations for the flood case. The stars denote the characteristics of the best-performing configurations according to the highest area under the precision-sensitivity curve.

the kernel function, where data input with no manipulation provided (i.e., Unw) is the most recurrent option occurring more than 40% of the time; similar percentage is attained by the radial basis function.

3.1.3.2 *Drought*

The data transformation for drought required the computation of the SPI from the precipitation data. The SPI was computed for different accumulation periods: Shorter accumulation periods (1-3 months) detect immediate impacts of drought (on soil moisture and on agriculture), while longer accumulation periods (12 months) indicate reduced streamflow and reservoir levels. As shown in **Table 9** models using SPI6 and SPI12 showed the best results and the values of the metrics are close to each other, thus, for brevity and in favour of clarity only one of the two is reported, namely, SPI over a six-month accumulation period.

Table 9: NN and SVM metrics median value of the top 5% configuration according to F1 score

	NN F1 Score	**SVM F1 Score**
SPI 1	0.8387	0.7709
SPI 3	0.8454	0.8341
SPI 6	0.9167	0.9317
SPI 12	0.9524	0.9464

Contrary to the flood case, the drought historical catalogue of events reported 1283 weeks with no droughts and 696 weeks of drought, with a ratio around 1.85 : 1 of no event/event. Albeit balanced, models with weights assigned were also investigated. The performances of the neural networks and the support vector machines were evaluated, like before, by a set of evaluation metrics and curves and a comparison against a logistic regression. It is important to point out that SPI is updated at weekly scale, same temporal resolution of the predictions, implying that each week counts as an event. Considering the duration of the drought in our historical catalogue of events (i.e., 17 weeks for the shortest one and 148 weeks for the longest), the temporal resolution adopted is an aspect to keep in mind when analysing the results obtained for these models. The highest F1 scores for the drought case were obtained for the NN model using all the datasets with weights for the classes. The network architecture was made up of 8 hidden layers with the relative number of nodes, activated by a ReLu function. The LF adopted was the binary cross entropy and the parameters update were regulated by an Adam optimiser. Regarding the SVM, the highest F1 score was achieved from the unweighted model using all ten environmental variables with radial basis function as kernel type and a C parameter equal to 100.

As for the flood case, **Figure 12** depicts the predictions over the testing set in a chronological order, and **Table 10** summarises the performances of the models, showing that the ML algorithms are a strong improvement with due respect to the logistic regression, showing high values across all the prediction skill measurements. The NN results as the most accurate model, while the SVM is the more precise overall. The implementation of either model should take into account the job that

these models are required to take on. If the purpose of the model is to balance the occurrences of false alarms and missed events, the NN is preferable. For a task that would require a stronger focus on the minimisation of false positives (i.e. reduce the number of false alarms), the SVM should be used.

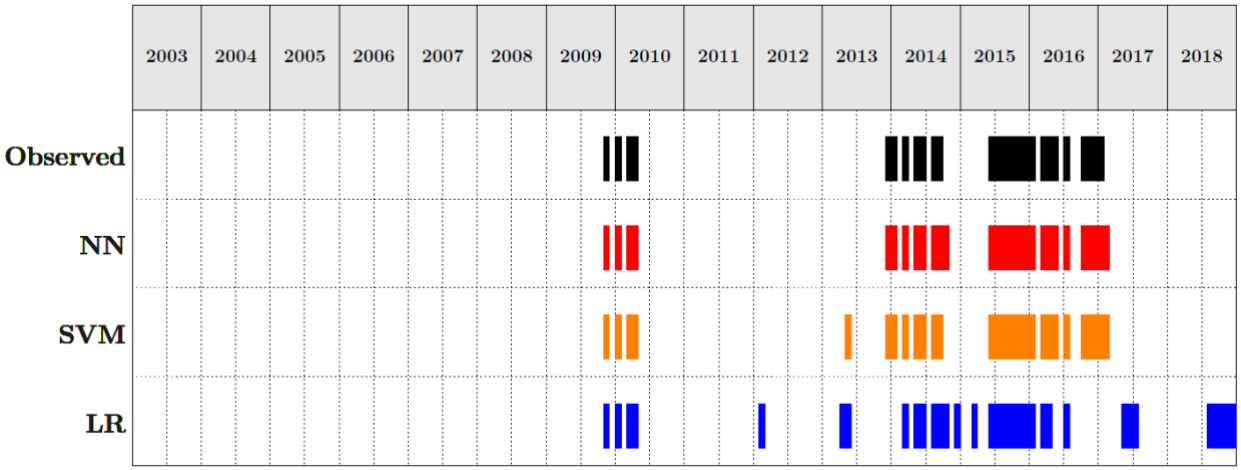

Figure 12: Predictions of the three methods for the drought case

Table 10: Metrics for the drought case

Method	Precision	Sensitivity	Specificity	F1 score	Accuracy
NN	0.95	1	0.99	0.97	0.99
SVM	0.96	0.96	0.99	0.96	0.98
LR	0.63	0.74	0.89	0.68	0.85

Figure 13 remarks the distance between the ML methods and the logistics regression as well as echoes what is observed for flood that the points for NN gravitate towards the area of the plot with higher sensitivity value while the SVM points tend to stay on the precision side of the plot. The addition of further datasets is still beneficial to the performances of the ML methods as displayed by Fig. 12, panel (b). The increasing trend for both ML models starts to slow down from the fourth rainfall datasets onward, which might be due to the redundancy of the rainfall datasets. On the other hand, the addition of the layers of soil moisture improves the performances especially for the support vector machine, which keeps improving steadily reaching the highest value of F1 score when the whole set of information is fed to the model.

Figure 14 refers to the best-performing configurations identified as the one with the highest area under the precision-sensitivity curve. The best configurations for either neural network and support vector machine are the ones containing the point with the highest F1 score, thus having the same features previously listed. The disparities between classes for drought are closer than those for flood, giving the accuracy, and the ROC curve, more reliability from a quality assessment point of

view. Looking at panels (a) and (b) of Fig. 13, both precision-sensitivity and. ROC curves show the ML methods decisively outperforming the no-skill and always-positive classifier. Furthermore, both plots exhibit a tendency of the neural network to group the points closer to each other towards the area containing the ideal model, which may indicate a more dependable prediction of the events as indicated by panel (c). In fact, while the two configurations have a high value of F1 score for a wide range of probabilities, the neural network has steadier prediction of true positive, false positive and false negative.

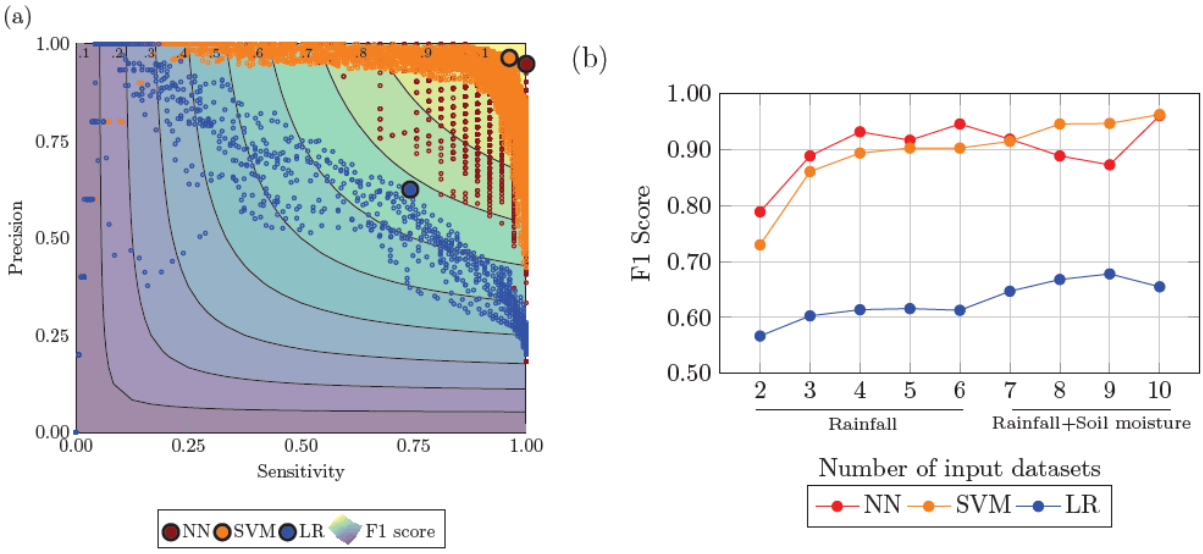

Figure 13: Performance evaluation for the drought case: (a) Performances of the top 1% configurations in the precision-sensitivity space highlighting the highest F1 score, (b): Comparison of ML methods with LR with combination using increasing number of input datasets.

This behaviour of the neural network could also be linked to the miscalibration of the confidence (i.e., distance between the probability returned by the model and the ground truth) associated with the predicted probability (Guo et al., 2017). The phenomenon arose with the advent of modern neural networks that employing several layers (i.e., tens and hundreds) and a multitude of nodes were able to improve the accuracy of their prediction while worsening the confidence of said prediction. Indeed, a miscalibrated neural network would return a probability that would not reflect the likelihood that the event will occur turning into a numeric output produced by the model.

The features breakdown of the model configurations top one percent shown in **Figure 15**, shows that the best NN configurations are predominantly the ones using weight for the two classes, and the ReLu activation function. Also, a large number of models use a high number of layers in accordance with the configuration with the highest area under the PS curve. The fact that most of the configurations obtaining the best performances have deeper layers may be a confirmation of the miscalibration affecting the estimated probabilities. For the SVM models, Fig. 14 denotes a marked component of the models using harder margins (i.e., high values of C-par) and radial basis function as a kernel.

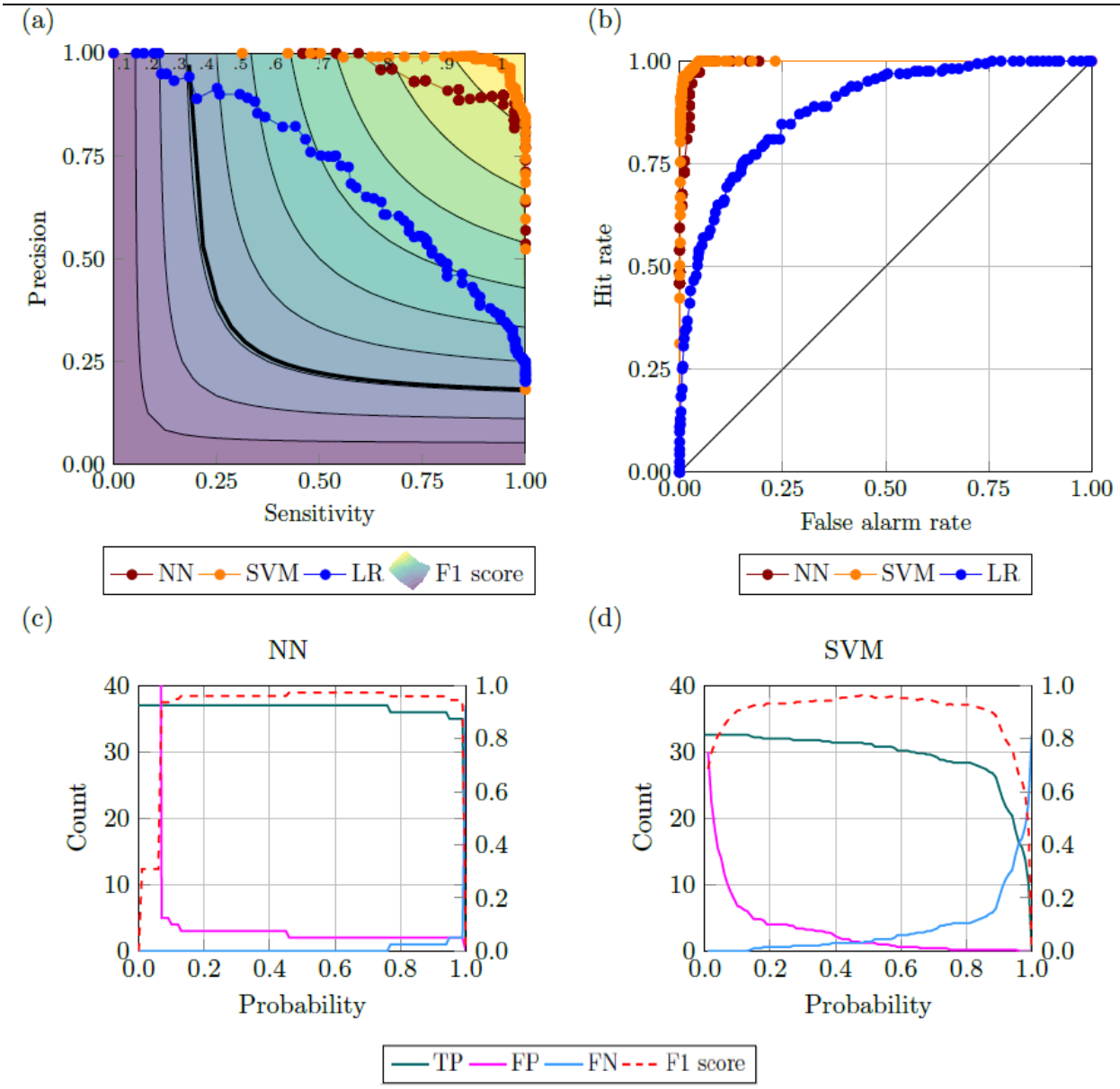

Figure 14: Best-performing configurations for the drought case: (a) PS curve, (b) ROC curve, (c) Variation of true positive, false positive, true negative and F1 score for the range of probability in NN and (d) in SVM.

3.1.4 Discussion

The machine learning framework presented merges a priori knowledge of the underlying physical processes of weather events with the ability of ML methods to efficiently exploit big data and can be used to support informed decision making regarding the selection of a model and the definition of a trigger threshold. Although several issues raised in this article warrant further research, there is clear potential in the application of machine algorithms to take advantage of increasing amounts of available environmental data within the context of weather index insurance.

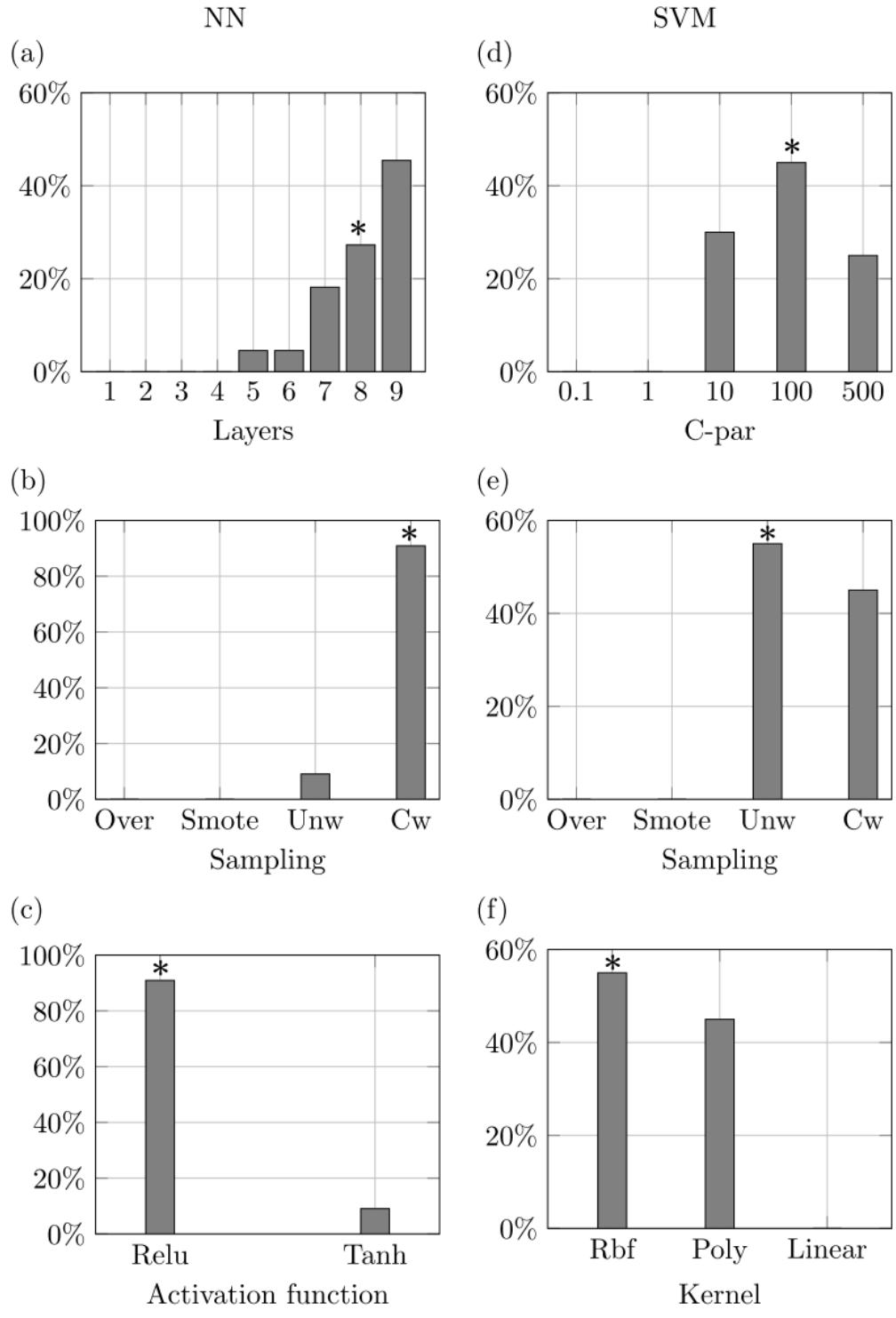

Figure 15: Properties of top 1% model configurations for the drought case. The stars denote the characteristics of the best-performing configurations according to the highest area under the precision-sensitivity curve.

The capability of these algorithms to reduce basis risk with respect to traditional methods could play a key role in the adoption of parametric insurance in the Dominican context and more generally for those countries that detain a low level of information about risk. Indeed, being able to rely on global data that are disentangled from the resources of a given territory, both from the point of view of climate data (e.g., lack of rain-gauge network) and from the point of view of information about past natural disasters, is an appealing feature of the work presented that would make the approach proposed feasible for other countries. The framework presented and topics discussed in this study provide a scientific basis for the development of robust and operationalizable ML-based parametric risk transfer products.

3.2 Prediction of milk production

The dairy industry is prominent in the Dominican Republic, and it is often affected by extreme weather events, impacting the livelihoods of those sustained by the sector. Furthermore, the anticipated increase in temperature and humidity caused by climate change are directly linked to heat stress, which has been proved to be a threat to milk yield and the ability of cattle to reproduce (Deng et al., 2007; Garry et al., 2021). Hence, being able to predict with a certain amount of anticipation the production of milk could be helpful in the development of insurance products and/or early warning systems to aid farmers in developing countries cope with extreme natural events or future adverse climatic conditions.
Recently, machine learning has gained popularity in the agricultural field for a variety of tasks. Several studies focus on crop yield prediction (Filippi et al., 2019; Johnson et al., 2016; Wang et al., 2018) which is dependent on different elements like weather, characteristics of the soil, irrigation schedule, irrigation technique, leveraging the ability of machine learning algorithms to handle large data and multiple input variables. Other research works exploit the ability of ML models to identify and recognize patterns in images to monitor specific crops, aiming to prevent the onset of diseases during the life-cycle of plants (Kim et al., 2006). Artificial intelligence has also provided advances related to livestock management and products, and more generally to the management of the resources usually used in a farm. For example, (Shine et al., 2018) tested several ML algorithms using milk production amid its input variables to predict monthly water and electricity consumption for 58 commercial dairy farms in Ireland. Another application closely related to milk production was conducted by (Caraviello et al., 2006) that used decision trees to identify which factors are the most influencing in the reproductive performances of lactating cows. (Grzesiak et al., 2006) test the ability of artificial neural networks (ANN) to predict daily milk yield from cows in production farms.

This section presents the case study of milk production in the Dominican Republic, where deep neural networks, namely the Long Short-Term Memory (LSTM) and the Convolutional Neural Network (CNN), are employed to predict the monthly production of milk at country level. The selection of milk production as predictand of our models is justified by the interaction with the local partner, which highlighted the importance of the dairy sector both for the livelihood of farmers and the role that such product plays in the economy of the Dominican Republic.
We first briefly introduce the functioning of the algorithms and the metrics used to evaluate the performances of the models. Then, a description of the dataset used to train the model is presented before delving into the results, where the predictive performances of the ML models are compared to an ARIMA model, used as a baseline. Finally, the main findings of this research and the implications they have for the Dominican context are discussed.

3.2.1 Datasets

The production of milk is governed by several factors ranging from climate conditions and types of pastures to the treatment and type of cattle. In the development of our machine learning models, we decided to focus on data related to climate, which is easier to obtain and is deemed appropriate in a country where most of the pastures are wild.

It was possible to obtain reliable Dominican milk data at monthly scale only for a short period of time, spanning from 2009 to 2015. Considering the limited data at disposal, in order to develop

the model and establish a sound training of the ML algorithms we decided to leverage an important capacity of machine learning models, transfer learning. This technique is adopted mostly when developing deep learning models in the field of image classification and computer vision, where models trained on large, general datasets are re-purposed in another context for a second task. In the case of milk production, we trained the models on monthly time series of milk production obtained from the EUROSTAT database and re-purposed such models, taking advantage of the features learned in Europe, to perform the predictions on milk production in the Dominican Republic.

3.2.1.1 *Environmental variables*

The selection of the right input data is an important step towards the realization of reliable ML models. To build models able to predict the milk production, three environmental variables and one vegetation index were selected. The input datasets selected for the training of the ML algorithms were obtained from both reanalysis (precipitation, temperature, and relative humidity from ERA5) and satellite product (NDVI from NOAA) at monthly time scale. A summary of the characteristics of the variables selected can be found in Table 11.

Table 11: Features of the environmental variables

Variable	Product	Type	Start date	Spatial resolution	Temporal resolution	Latency
NDVI	NOAA STAR-VHI	Satellite	1981-present	4km	7 days	Within 7 days
Precipitation	ERA 5	Reanalysis	1979-present	0.25°	1h	Within 5 days
Temperature	ERA 5	Reanalysis	1979-present	0.25°	1h	Within 5 days
Relative Humidity	ERA 5	Reanalysis	1979-present	0.25°	1h	Within 5 days

Precipitation data underwent a transformation as described in Section 3.1.2.1 to derive the standard precipitation index, useful to reproduce the occurrence of droughts. Relative humidity and temperature were used to derive the Temperature Humidity Index (THI) as described in (Deng et al., 2007):

$$THI = TD - \big(0.55 - (0.55 * RH)\big) * (TD - 58) \quad (10)$$

THI is a commonly used indicator of heat stress which, as already mentioned, has a strong influence on the capability of a cow to produce milk.

3.2.1.2 *Milk Production*

Retrieving reliable data regarding milk production is not always an easy task. The local data have a monthly time scale grouped at country level and with several periods of data not validated or missing. The longest period of reliable data related to the Dominican Republic is between 2009 and 2015 and the time series is illustrated in Figure 16. This plot shows what seems to be an

increasing trend in the milk production country-wise as well as an interannual pattern. Both characteristics are good examples of problems that deep learning models used for time series prediction are well-equipped to handle.

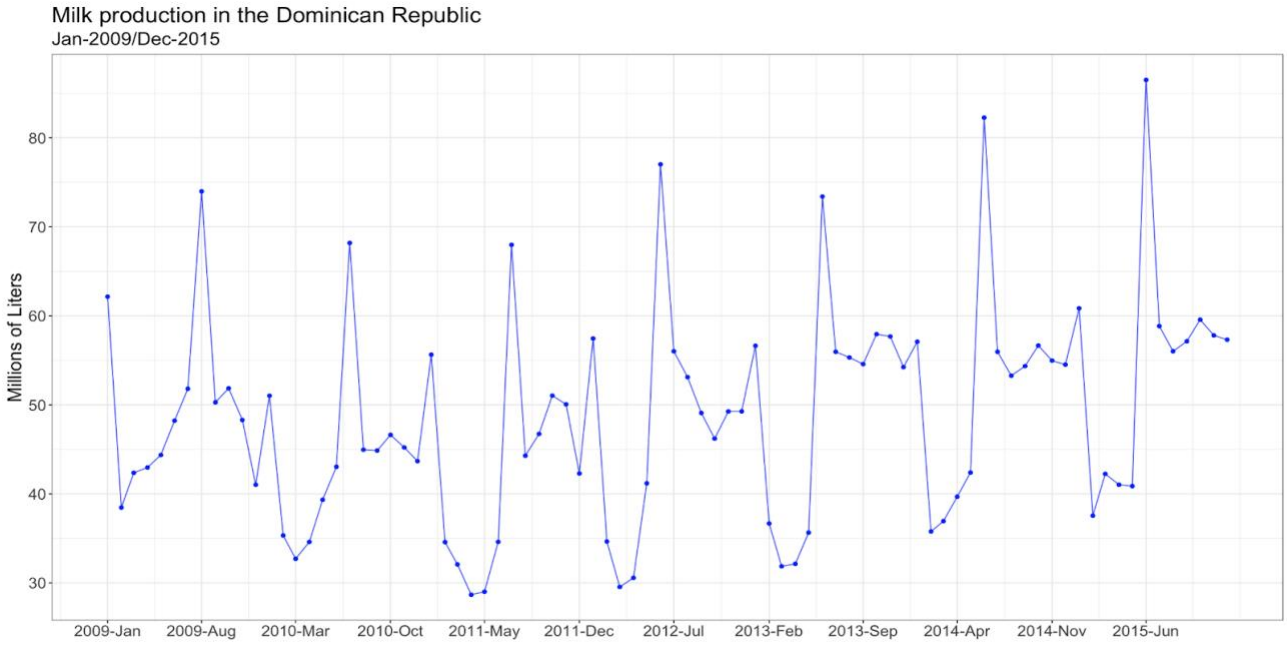

Figure 16: Time series milk production in the Dominican Republic

The size of the dataset (i.e. 12 months x 7 years equal to 84 timesteps) available for the Dominican Republic was not believed appropriate to train a deep learning model. Therefore, we decided to exploit the ability of these models to *transfer learning*. The main idea is to train a ML model in a country and/or area where reliable data are available, so as to have the model learning features underlying the whole milk production process and not the features specifically related to a given country. For this task, we gather monthly data of milk production for countries belonging to the EU, which are openly available and updated monthly on the Eurostat data store website. The time span covered by the data, as shown in Figure 17, goes from 1968 to 2020 providing a sufficiently long and more reliable set of training data.

3.2.2 Methodology

The methodology proposed for the prediction of milk production follows a similar conceptual framework to the one described in Figure 5. The main idea is to use open quasi-global data, applying a priori knowledge by means of a data transformation (e.g., deriving the temperature humidity index as suggested in (Deng et al., 2007)) to augment the data and then train a machine learning model to perform, in this instance, a regression task. In this study we adopted two deep learning algorithms well-suited for regression problems: recurrent neural networks (RNN), specifically long short term memory (LSTM), and 1-dimensional convolutional neural networks (CNN1D).

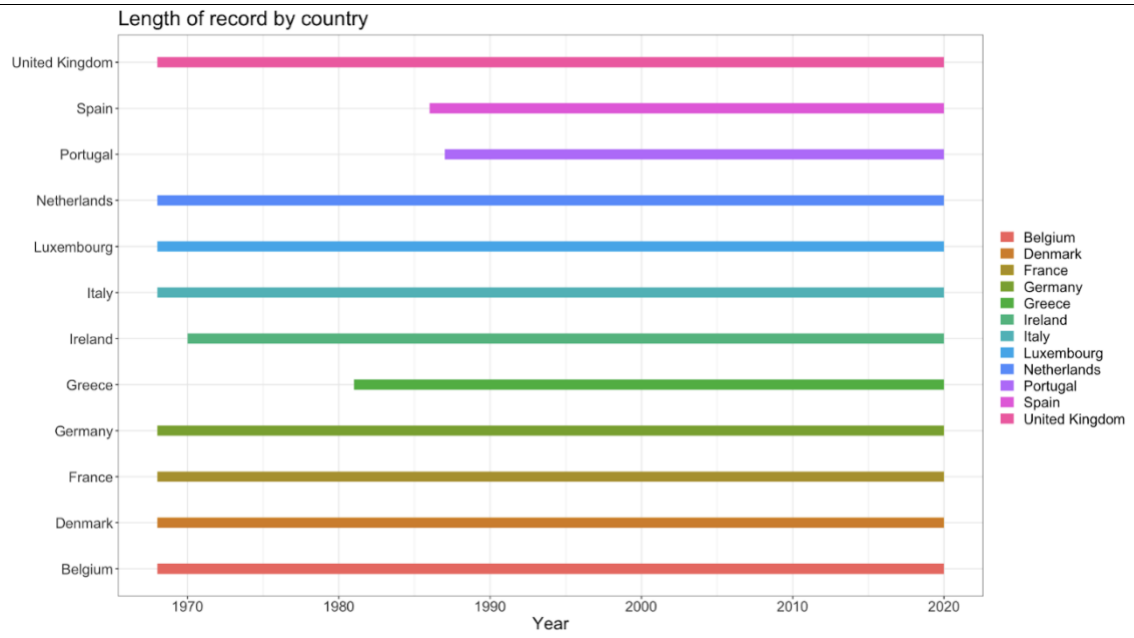

Figure 17: Data availability in EU

For the specific case of milk production in the Dominican Republic, finding reliable local data for an extended period of time was not feasible, therefore, we decided to tackle the problem by taking advantage of the ability of these models to transfer knowledge (Wang et al., 2018) by means of training the ML models with data coming from European countries, and then, using said model to validate the prediction of milk production in the Dominican Republic for the few reliable years of data available.

3.2.2.1 *Machine learning*

Long Short-Term Memory (LSTM) Recurrent Neural Networks

Densely connected neural networks and convolutional neural networks have no memory (Chollet, 2017) meaning that each input is processed independently. While this might not be an issue for tasks such as identification of events or object detection, when dealing with time series data, as in the case of milk production, keeping track of the information stored in the sequential nature of the data is crucial. To overcome the no-memory issue, we introduced recurrent neural networks (RNN). LSTM is a type of network belonging to the family of the RNN that is widely used due to its ability to overcome the problem of the vanishing gradient (Hochreiter and Schmidhuber, 1997). The structure of an LSTM model is based on two main components called states and gates. The former comprises the hidden state, which is responsible for storing the values of previous hidden states functioning as the memory of RNN, and the input states, which are a combination of the input data and previous hidden states. Regarding the latter, there are three types of gates in a LSTM architecture:

- The forget gate, where the decision on whether the information of the current state is worth keeping or not is made.

$$f_t = \sigma\left(W_{f_x x_t} + W_{f_s s_{t-1}} + b_f\right)$$

- The input gate, which is used to decide if the internal state should be used as a memory cell.

$$i_t = \sigma\left(W_{i_x x_t} + W_{i_s s_{t-1}} + b_i\right)$$

$$c_t = tanh\left(W_{c_x x_t} + W_{c_s s_{t-1}} + b_c\right)$$

$$C_t = f_t \cdot C_{t-1} + i_t \cdot c_t$$

- The output gate, which decides what the next hidden state should contain.

$$o_t = \sigma\left(W_{o_x x_t} + W_{o_s s_{t-1}} + b_o\right)$$

$$\tilde{s}_t = o_t \cdot \tanh\left(C_t\right)$$

where, σ denotes the sigmoid function, $W_{f_x}, W_{f_s}, W_{i_x}, W_{i_s}, W_{c_x} W_{c_s}, W_{o_x}, W_{o_s}, b_f, b_i, b_c, b_o$ are the weights and biases used at different layers, respectively and $\tilde{s}_t$ denotes the output of LSTM network at time signal t.

A schematic representation of an LSTM is depicted in Figure 18.

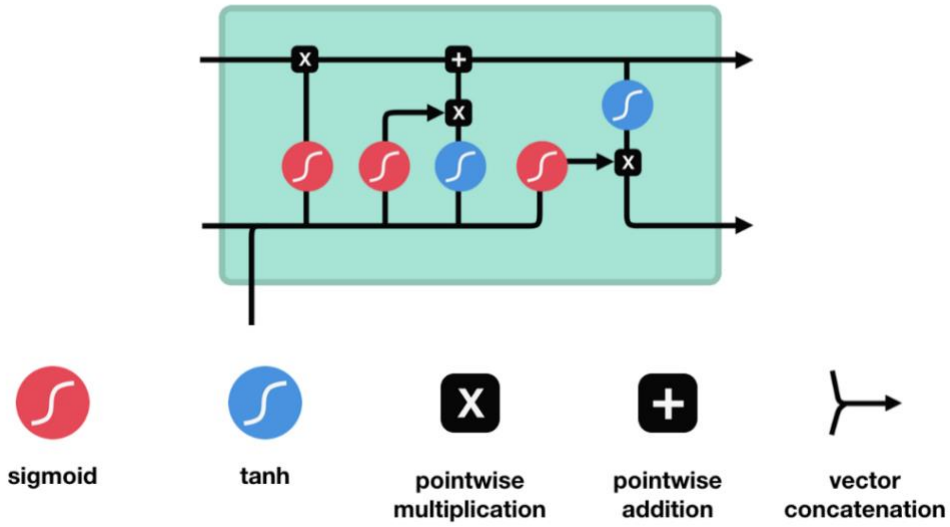

Figure 18: Basic functioning of a LSTM cell

Convolutional Neural Network (CNN1D)

Traditionally, convolutional neural networks are used to process images, moving a 2-dimensional kernel over the images trying to extract features that allow them to recognise objects or reproduce patterns. This concept can be translated to time series using a 1-dimensional CNN, where instead

of a 2D kernel, a 1D window slides through the time series extracting features from the data as shown in Figure 19.

The elements of the kernel are multiplied by the corresponding element of the time series. The sum of the product is then passed through a nonlinear activation function and the resulting value is stored in the new convolutional layer, whose length is equal to the number of convolutional kernels (i.e., the number of features the model is trying to extract from the initial time series). The values in the convolutional layer are later on pooled or flattened into a feature vector that is used as input to a regular fully connected layer that is ultimately in charge of making the prediction.

One-dimensional CNNs are gathering much attention due to the possibility of having a larger kernel with respect to its 2D counterpart; usually, kernels for images use a 3x3 convolutional kernel, which in the 1D case can be easily extend to a convolutional window of size 7 or 9 (Chollet, 2017). Furthermore, CNN are much faster computationally than recurrent neural networks due to the fact that the mechanism responsible for holding information of past states in the LSTM is not present in convolutional neural networks.

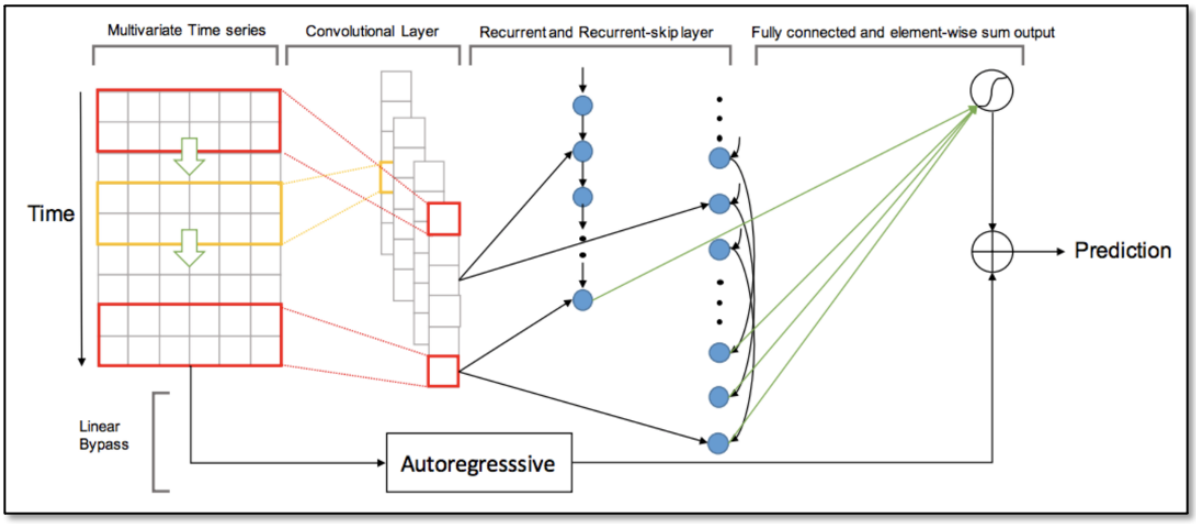

Figure 19: Modelling framework for a 1D Convolutional Neural Network

3.2.2.2 *Data pre-processing*

Deep learning models do not train using the entire dataset at once, but rather splitting the dataset into smaller batches. For LSTM and CNN1D alike, when used with data where sequentiality plays a role (e.g., time series of milk production), these batches are a series of 3D tensors where the three dimensions (also called shape) are: number of samples (also called batch size), number of time steps and number of features. The samples are the number of 2D vectors with dimension (timesteps, features) present in the period of time the dataset covers. The number of timesteps is the time span we want the model to look at for each sample, while the features are the number of input variables used to predict the target variable (i.e., a univariate time series will have width equal to 1). Let's take an example where a dataset is constructed from daily measurements of temperature and precipitation. The daily recordings of the two variables can be stored in a one-dimensional vector of length 2 (i.e., number of features). Instead, an entire week of recording can be stored into a 2D tensor of shape (7,2), while an entire year of recordings could be stored in a

3D tensor of shape (52,7,2), where 52 is the number of weeks (samples) in a year for non-leap years.

In the case of time series predictions, the number of timesteps is usually chosen as the time window that better describes the phenomenon we are trying to predict (e.g. looking at 2 years of hourly data to predict 1 hour in the future the price of Bitcoin might not be very useful). The length of the window can also be a parameter of the model that can be tuned. The division of the dataset into consecutive samples with a certain length is also referred to as data windowing.

Another important aspect that regards the pre-processing of time series data is the creation and usage of the so-called lag features. Lag features are time series of the target variable at previous timesteps, thus, when the target variable is believed to depend on its previous values, it is common practice to leverage past information to create additional features able to improve the predictive capacity of the models.

Lastly, pre-processing entails the splitting of the dataset at disposal into training, validation and testing of the data as already discussed for the classification problem in Section 3.1. This is essential to avoid overfitting of the model and to guarantee a proper training. The models in charge of predicting milk production were trained on 70% of the data available with 20% of the data used for validation during the training to avoid overfitting. The remaining 10% were used to check the ability of the models to generalise to a set of data that has never been seen.

3.2.2.3 *Model evaluation*

An important step in the pipeline of developing a machine learning model is the selection of appropriate metrics to evaluate model performances.

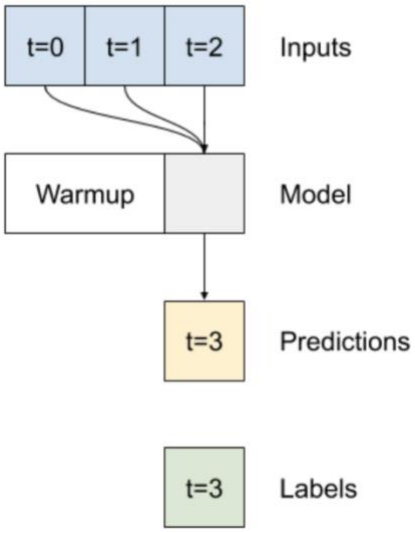

Figure 20: Single-step-ahead prediction scheme.
Source:https://www.tensorflow.org/tutorials/structured_data/time_series?hl=en

For this case, three well known metrics for regression tasks were chosen to evaluate prediction quality: mean square error (MSE), mean absolute error (MAE) and the coefficient of determination (R^2). The mathematical formulations of these three metrics are:

$$MSE = \frac{\sum_{i=1}^{n}\left(\hat{Y}_k - Y_k\right)^2}{n}$$

$$MAE = \frac{\sum_{i=1}^{n}\left|\hat{Y}_k - Y_k\right|}{n}$$

Where:

- $\hat{Y}_k$ is the vector of predicted values
- Y_k is the vector of observed values

$$R^2 = 1 - \frac{\sum_{k=1}^{n}(y_k - \hat{y}_k)^2}{\sum_{k=1}^{n}(y_k - \bar{y}_k)^2}$$

Where:

- y_k are the observation;
- $\bar{y}_k$ is the mean of the observation;
- $\hat{y}_k$ is the prediction of the model.

There are two main types of prediction used in problems of time series forecasting:

- Single-step-ahead (SSP)
- Multi-step-ahead (MSP)

As the names suggest, in the first case the model tries to predict a single time step in the future as displayed in Figure 20 (e.g., using past information to predict milk production for the next month).

In the second case, the model predicts multiple time steps in the future as shown in Figure 21. There are different ways to perform a multi-step prediction (Bontempi et al., 2013); in this work we used the single-shot approach, where the model predicts the entire forecast horizon in one shot (e.g. model predicts the next 12 months all at once).

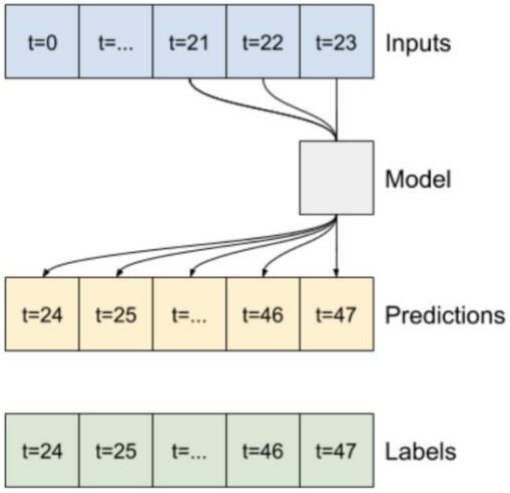

Figure 21: Multi-step-ahead prediction scheme. Source: https://www.tensorflow.org/tutorials/structured_data/time_series?hl=en

3.2.3 Results

The aim of this branch of research was to train deep learning models to predict milk yield. In order to evaluate the goodness of the performances of the ML algorithms we used the ARIMA model as a baseline. The presentation of the results is divided into two parts, first relative to SSP and then relative to MSP. Each algorithm presented was trained for different parameters, and both methods were tested using as input data all the combinations of the environmental variables and lagged milk production. Additionally, for the LSTM model, a range of values between 2 and 24 months was explored for length of the time window. This procedure resulted in a large number of model configurations. Here we will highlight the one showing the best results.

Table 12: Comparison models trained in Europe

Metrics	**Input variables**	**Italy**			**Germany**			**France**		
		ARIMA	LSTM	CNN1D	ARIMA	LSTM	CNN1D	ARIMA	LSTM	CNN1D
MSE	Lag Milk	2.23	0.273	**0.113**	0.09	0.13	**0.052**	0.095	0.083	**0.023**
MAE	Lag Milk	1.33	0.423	**0.261**	0.238	0.28	**0.176**	0.248	0.025	**0.12**
R^2	Lag Milk	-1.59	0.715	**0.807**	0.767	0.583	**0.833**	0.717	0.711	**0.918**

An initial assessment of the performances of the LSTM and CNN1D models was carried out, for European countries, for which more data are available. Several countries were tested to develop the models; for the sake of brevity, Table 12 reports the comparison between the models run in Italy, Germany and France, the countries for which the models presented the best overall performances.

Table 13: Comparison of CNN1D model results model trained in Europe vs model trained in Dominican Republic.

Country of model training		Dom. Republic	Germany		Italy		France	
2nd-step training using local data		n/a	x	✓	x	✓	x	✓
Performance metrics	**MSE**	0.274	0.54	0.296	0.67	0.139	0.58	0.37
	MAE	0.39	0.54	0.371	0.64	0.288	0.63	0.405
	R^2	0.612	0.56	0.76	0.456	0.887	0.53	0.701

The metrics reported in the above table are the ones obtained from the testing set and it is already noticeable that for all three models the CNN1D model outperforms the LSTM and the ARIMA. Also, the best performing models for both ML algorithms are obtained using only the 1-month lagged milk as input, which hints towards the autoregressive nature of milk production.

Once established the CNN1D as the best model for milk production, the models trained in the European countries were used to predict monthly values of milk in the Dominican Republic, using the period 2009-2014 as an additional training period and the year 2015 as the period upon which the performances of the model are evaluated. In addition, the European models were also directly applied to the Dominican data without any further training. A comparison of the results obtained can be found in Table 13.

The CNN1D model trained in the Dominican Republic provides reasonable performances for the Dominican Republic even with the training over the short period of observations. Nonetheless, the models trained in Europe and then receiving additional training in the Dominican Republic provide a strong improvement with the very best being the Italian model, which exhibits small errors and a R^2 close to 0.9. The comparison of single-step-predictions over 2015 between the different models can be appreciated in Figure 22.

Similarly to the SSP, for the multi-step-prediction counterpart, the models were first trained in an European country, and the best performing models were then used to perform a 12 month forecast in the Dominican Republic using the same split of the data as described for the SSP. Table 14 reports the value of the metrics averaged over the 12 month forecast from January 2015 to December 2015 on DR milk production. Averaging the metrics over the entire window of the forecast was used to select the best model.

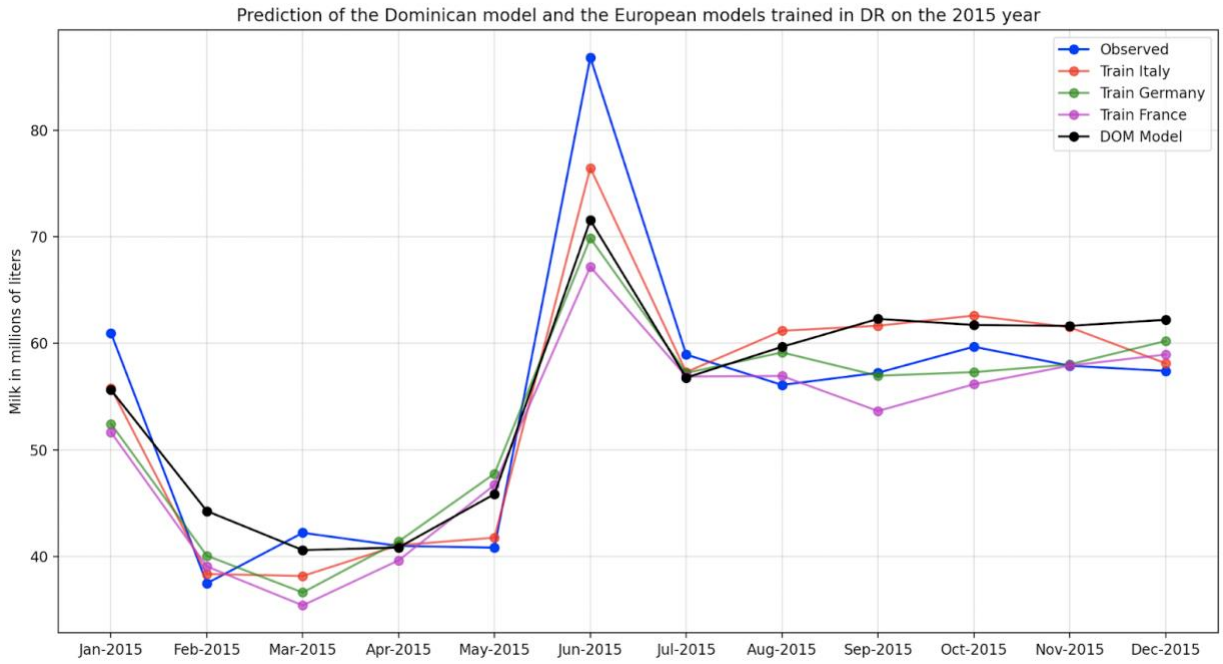

Figure 22: Single-step-prediction over the DR

Table 14: Twelve months forecast over the French milk production time series

Metrics	DOM	Germany	Italy	France
Avg. MSE	1.89	0.625	0.578	0.616
Avg. MAE	1.075	0.612	0.521	0.634

In 2015, the Dominican Republic produced 655 million liters of milk. The error at yearly scale is estimated at about 122 millions liter that means around 18% error. This behavior might be expected since the model is trying to perform a forecast on a period of length equal to 25% of the training set, hence, the model might have difficulties in extracting enough information to make such a long prediction. Table 15 reports the monthly breakdown of the metrics between the Dominican model and the best-performing European model according to the average MSE, the Italian model. The monthly breakdown echoes the low predicting skills of the Dominican model already qualitatively assessed in Figure 23. An interesting outcome of the MSP results is the fact that the model improves its performances using THI alongside the lagged milk as predictors.

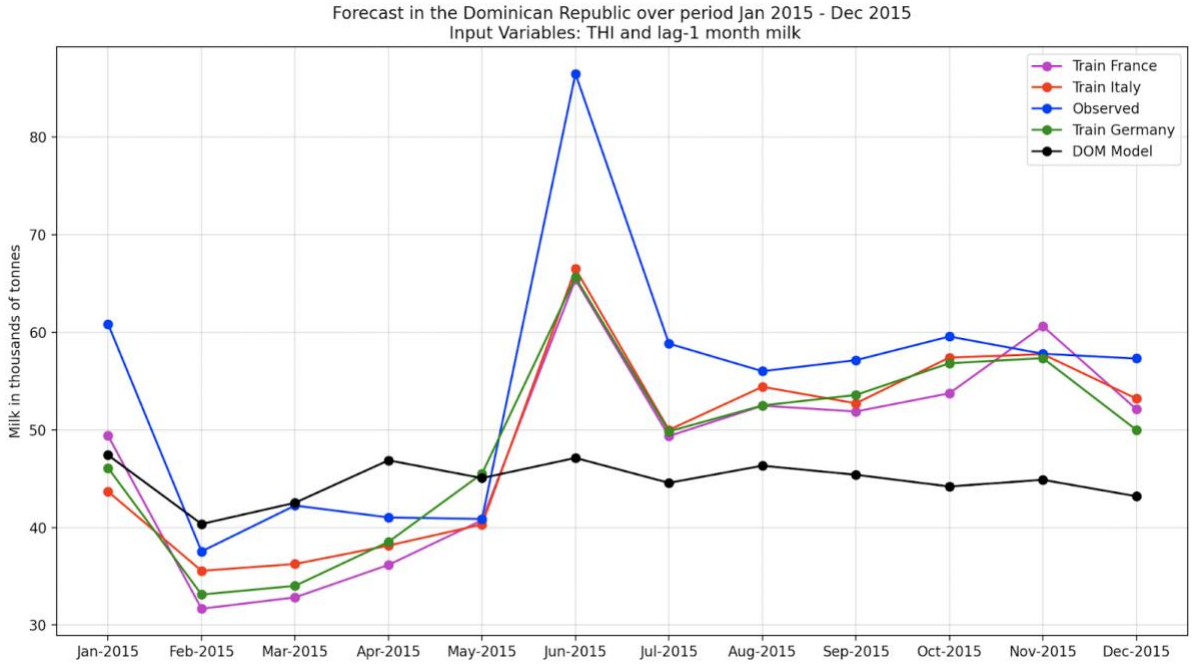

Figure 23: Comparison of the twelve months forecast for DR between the Dominican model and the European trained models.

Table 15: Monthly breakdown of model metrics for the Dominican model and the Italian model (i.e., the best performing according to the averaged MSE and MAE)

Month	Dominican Republic		Italy	
	MSE	MAE	MSE	MAE
1	4.64	2.15	2.36	1.54
2	0.32	0.56	0.03	0.18
3	1.02	1.01	0.29	0.54
4	0.58	0.76	0.07	0.26
5	0.07	0.26	0.003	0.05
6	9.81	3.13	3.21	1.79
7	1.9	1.38	0.64	0.79
8	1.11	1.05	0.02	0.14
9	0.64	0.8	0.16	0.4
10	1.3	1.14	0.04	0.19
11	0.28	0.58	0.001	0.004
12	0.49	0.7	0.135	0.368

3.2.4 Discussion

In conclusion, the research on milk production showed the potentiality of these deep learning models in predicting time series. The results obtained by transferring the model trained in Europe to the Dominican Republic are a promising indicator of its applicability to other regions of the world. Moreover, specific to the Dominican context, an improvement in the data collection that

would lengthen the milk production time series would be beneficial in putting these models into operationalization, particularly for the multi-step predictions. Lastly, the autoregressive nature of the milk production time series along with the results obtained suggested that the prediction capability of the ML models are unaffected by environmental variables, which seems to be in contrast with what is discussed in the literature. The justification for this phenomenon might reside in the minimal information at disposal regarding the local treatment of cattle and farms when dealing with extreme events. In fact, based on local information, the impact of severe weather events like floods and droughts on milk production is mitigated by the government, who provides financial assistance to farmers that allows them to take adequate measures. For example, in case of a heat wave farmers might compensate for the hot weather with an increase in the watering of the cattle or with refrigerating systems in the farms. Naturally, this is not reflected by the climatic variables. Therefore, these activities are not modelled by the current framework, although they are responsible for keeping the milk production from having a drastic drop when an event occurs, justifying the low (or negative) influence of the climatic variables in the predicting performances of the deep learning models when used as input for the training.

3.3 Applicability of crop models in the context of parametric insurance

Crop models simulate the growth process of a specific crop type. They are based on multiple equations that mimic the response of the crop to environmental and meteorological parameters, which are provided as input to the model. The final output of such models is the crop yield. The integration of these types of models into the design of parametric insurance programs is still at an early stage, as underlined in a recent research paper by (Afshar et al., 2021). Nevertheless, the same article highlights the potential that the innovative use of crop models has in reducing basis risk. Therefore, this is a highly relevant topic for the SMART project, which led us to carry out an exploratory study on the applicability of this type of model to support the design of parametric triggers.

As mentioned previously, the construction of robust parametric triggers can help in making parametric insurance more appealing for final beneficiaries, such as farmers and their associations. In this context, crop models are particularly appealing because they are able to provide yield estimates at a finer scale with respect to the national one, depending on the spatial resolution of the input environmental and meteorological parameters. In principle, this allows calibrating parametric insurance policies based on the climatological and environmental conditions of different locations, and therefore tailoring them to specific end-users.

In data sparse regions, such as the Dominican Republic, long time series (around 20 years) of meteorological parameters are sparse; therefore, the option of running crop models with gridded dataset should be carefully evaluated. Today there are different robust gridded datasets that can be exploited, as already shown in the previous sections of this report. An important step in deciding whether to use a specific gridded dataset is its validation. In this work, the performance of a crop model run with gridded data is compared with the ones of the same crop model run with station-recorded data over the period for which such data are available. This step is the basis to evaluate the feasibility of using crop models to improve the design of parametric insurance policies.

Crop models are sets of mathematical equations that provide quantitative knowledge on how a crop grows in interaction with its environment (Nassiri Mahallati, 2020). The models are based on mathematical algorithms that capture the quantitative information of agronomy and physiology experiments in such a way that can explain and predict crop growth and development together with yield, biomass, and water and nutrient uptake (Asseng et al., 2014). The input data include:

1. Daily weather parameters (average, minimum and maximum temperature, rainfall, solar radiation, wind speed, relative humidity),

2. Environmental parameters (soil characteristics and initial soil conditions),

3. Crop parameters (crop type, planting date).

Crop models, given their ability to simulate yield at different growth stages, can be a useful tool in the context of parametric insurance. Their application in designing insurance policies has recently received scientific attention. (Afshar et al., 2021) demonstrated that the accuracy of yield estimation carried on with the APSIM crop model, (described in (Keating et al., 2003)) combined with satellite information significantly outperformed models based solely on satellite vegetation indices and was consistent with existing research using crop models and satellite data for yield estimation in India. However, the accuracy of weather-yield relationships derived from crop

models is highly dependent on the quality of the underlying input weather data used to run them (Parkes et al., 2019).

A major challenge in data sparse countries, such as the Dominican Republic, is the lack of sufficiently long time series (at least 20 years of data are required) of accessible and reliable meteorological datasets. Gridded weather datasets, derived from combinations of in-situ gauges, remote sensing, and climate models, provide a solution to fill this gap, and have been widely used to evaluate climate impacts on agriculture in data-scarce regions worldwide (Parkes et al., 2019). However, the reference datasets may be biased and contain uncertainties. Therefore, they may not be able to reproduce the local climate. Thus, it is necessary to assess the capability of crop models to produce reliable yield estimates when running with gridded data.

This part of the study compares the yield estimates obtained running a crop model with gridded meteorological data as input and the ones obtained when the same crop model runs with station recorded data as input. The analysis is divided into three steps:

1. Identification of crop harvested areas: areas in which a crop is grown are determined using the MAP Spam dataset.
2. Yield computation: calculation of crop yield using a crop model, at first run with gridded data and then with station data.
3. Comparison of the results.

3.3.1 Datasets

3.3.1.1 *Identification of crop harvested area*

The MAP SPAM dataset was used to identify the crop harvested area. Map SPAM, the Spatial Production Allocation Model, uses a cross-entropy approach to make plausible estimates of crop distribution within disaggregated units (Yu et al., 2020). MAP SPAM relies on a collection of relevant spatially explicit input data, such as crop production statistics, cropland data, biophysical crop "suitability" assessments, population density, as well as any prior knowledge about the spatial distribution of specific crops or crop systems. SPAM provides harvested area for 42 crops. The used version is the 2.0, which has 2010 as reference year (International Food Policy Research Institute, 2019). SPAM provides gridded information at a spatial resolution of 10 km and has a global coverage. Therefore, the procedure described here to identify crop harvested area can easily be implemented in other countries.

3.3.1.2 *Meteorological variables*

Two sets of weather input parameters were considered:

1. Station recorded weather parameters;
2. Gridded weather data.

Four automatic weather stations were considered (Figure 24). These stations were installed in the period between the end of 2016 and mid-2017, and three of them (Agua de Luis, Los Montones and Juliana Jaramillo) are still recording. For each station, the recording period reported in Table 16 was considered to run the Aquacrop model. Table 16 also presents the stations' main features,

namely their elevations and geographical coordinates. The considered weather variables were hourly rainfall, temperature, relative humidity, and extra-terrestrial and net solar radiation.

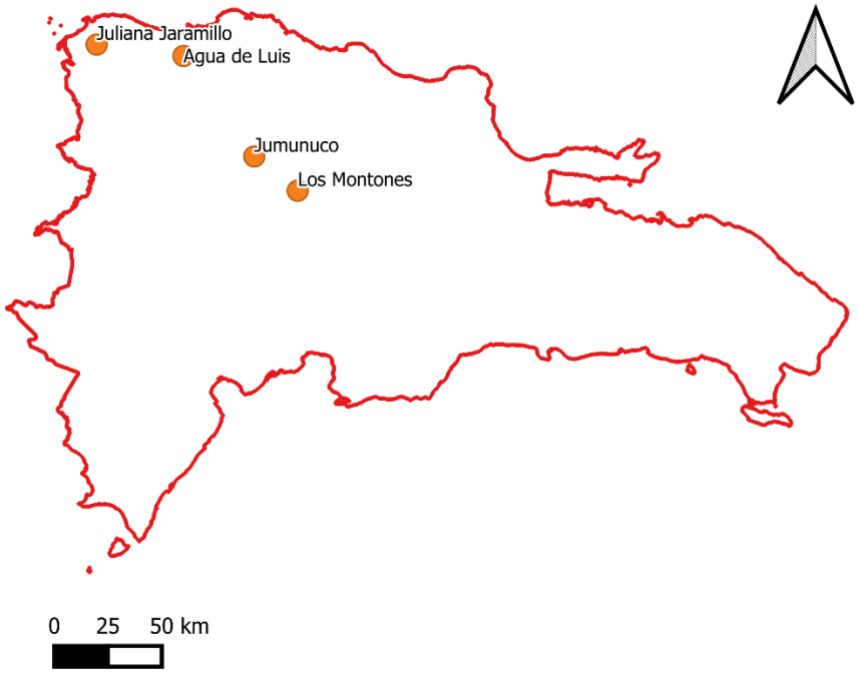

Figure 24: Automatic weather stations included in the analysis.

Table 16: Coordinates and elevation of the four considered weather stations

Station Name	Longitude	Latitude	Elevation (m)	Recording period
Agua de Luis	-71.23	19.73	129	27/12/2016 to 03/05/2020
Jumunuco	-70.92	19.29	690	29/05/2017 to 22/10/2019
Los Montones	-70.73	19.14	717	23/05/2017 to 28/10/2020
Juliana Jaramillo	-71.61	19.78	10	30/08/2018 to 28/10/2020

The same variables were extracted from the ERA-5 reanalysis gridded dataset for the period between 2017 and 2020. ERA-5, already described in **Table 1**, has a 25km spatial resolution and global coverage. The dataset has already been used by (Afshar et al., 2021) to design parametric insurance policies since it has a short latency time (around 5 days). The blue grid cells shown in Figure 25 were considered in the analysis since they correspond to the automatic weather stations' locations.

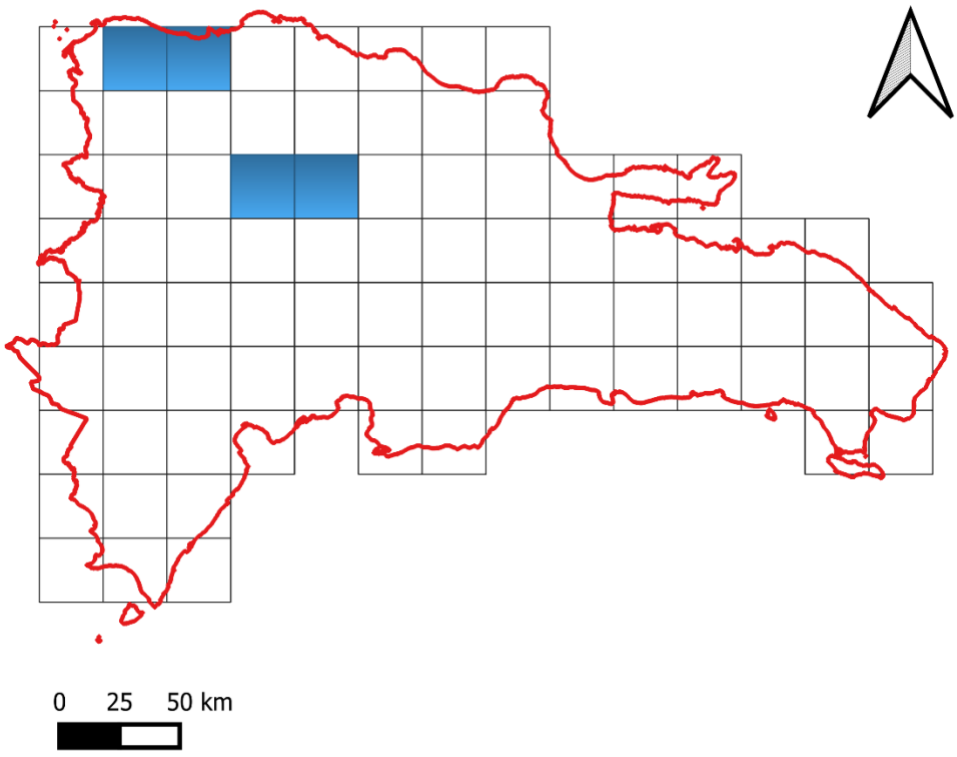

Figure 25: ERA-5 grid cells included in the analysis.

3.3.1.3 *Soil data*

Soil characteristics, such as soil texture and soil bulk density, which are necessary to run the AquaCrop model, were derived from the ISRIC soil data hub (World data center for Soils, 2010). ISRIC is the World Data Centre for Soils which, in collaboration with its partners, has been working for over 50 years on compiling and harmonising data on soils and their properties. The SoilGrids dataset, implemented by ISRIC, is a system for global digital soil mapping that uses state-of-the-art machine learning methods to map the spatial distribution of soil properties across the globe. SoilGrids is fitted using over 230000 soil profile observations. The outputs of SoilGrids are global soil property maps at six standard depth intervals (0-5cm, 5-15cm, 15-30cm, 30-60cm 60-100cm and 100-200cm) at a spatial resolution of 250 meters. In this work the following SoilGrids parameters were used: bulk density, sand content, clay content, silt content, nitrogen and carbon content. The soil type was classified according to the USDA soil texture triangle (Soil Science Division Staff, 2017), based on the percentage of silt, clay and sand retrieved from the SoilGrids datasets.

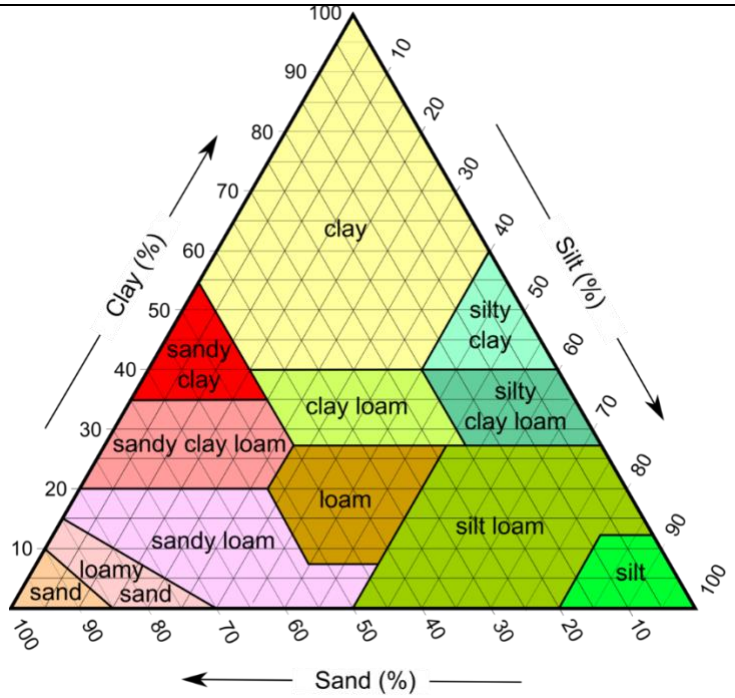

Figure 26: Soil texture triangle (Soil Science Division Staff, 2017)

3.3.2 Methodology

3.3.2.1 *Yield computation*

At first, a review of the available crop models was performed to select the most appropriate model for the study. The selected crop model was AquaCrop, developed by the Food and Agriculture Organization (FAO) to address food security and to assess the effect of environment and management on crop production (Food and Agriculture Organization, 2017). The AquaCrop model has previously been used in the Caribbean context: (Rankine et al., 2015) who applied the model in Jamaica to simulate sweet potato yield, and (Campo et al., 2017) describe how local farmers were trained to use AquaCrop in Trinidad and Tobago. AquaCrop simulates yield response to water of herbaceous crops and is particularly suited to address conditions where water is a key limiting factor in crop production (Food and Agriculture Organization, 2012). AquaCrop was developed in 2009 and has since been used worldwide in different agro-ecological conditions. The model, which is based on the relationships between crop yield and water described in (Steduto et al., 2012), was selected because it requires relatively few input parameters and can be run even in countries where information on soil properties and characteristics and crop variety are not detailed.

Crop specific parameters already included in the Aquacrop model were used, while information on crop planting and harvesting dates were retrieved from crop calendars for the Dominican Republic (United States Department of Agriculture, 2008).

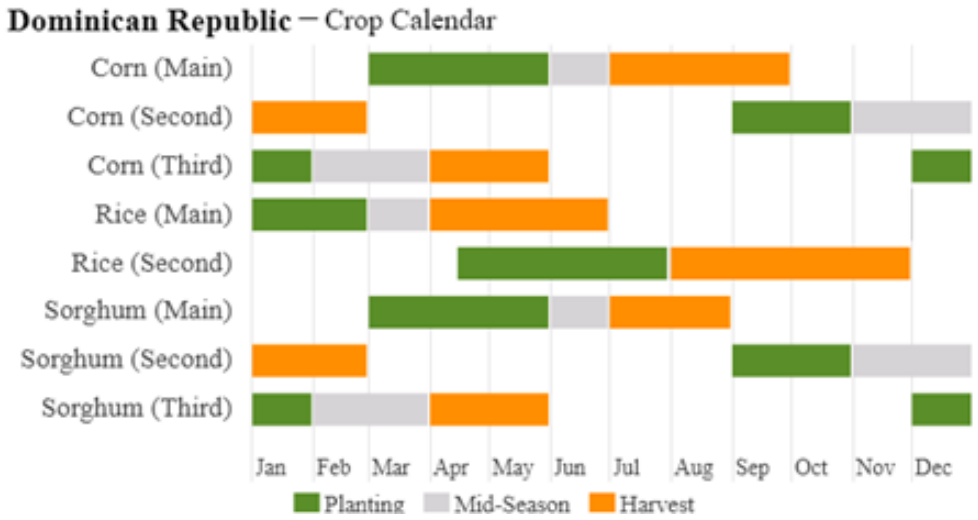

Figure 27: Dominican Republic crop calendar (United States Department of Agriculture, 2008).

3.3.3 Results

At first, a comparison between the weather variables extracted from stations and the ones derived from the ERA-5 dataset was performed. The Los Montones station was considered for illustration purposes, because it's the station that recorded for the longest period. Figure 28 shows the time series of rainfall from ERA-5 and rainfall from the Los Montones automatic station. This station started recording from 23/05/2017; station recorded data are not available from November 2018 to August 2019. It can be observed that ERA-5 tends to overestimate precipitation with respect to the automatic station.

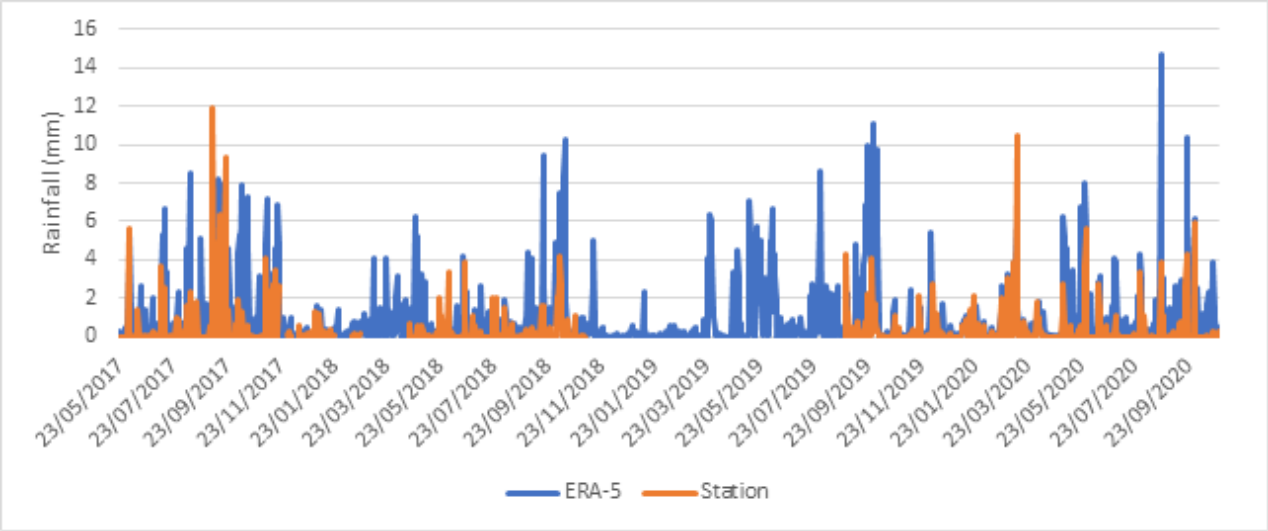

Figure 28: Comparison between station recorded daily rainfall and ERA-5 daily rainfall. Los Montones station

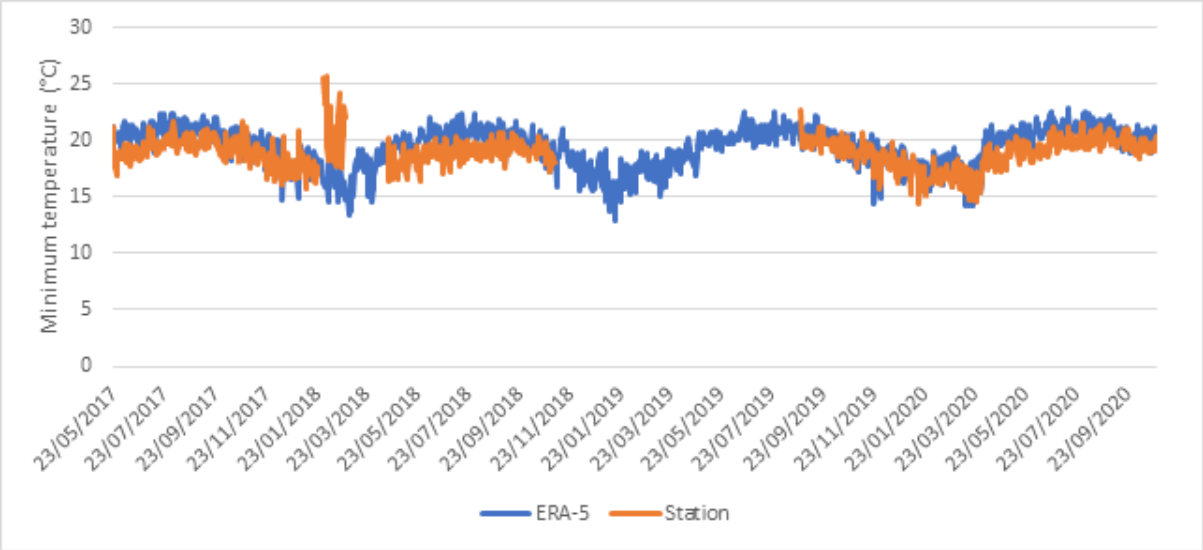

Figure 29: Comparison between station recorded daily minimum temperature and ERA-5 daily minimum temperature. Los Montones station

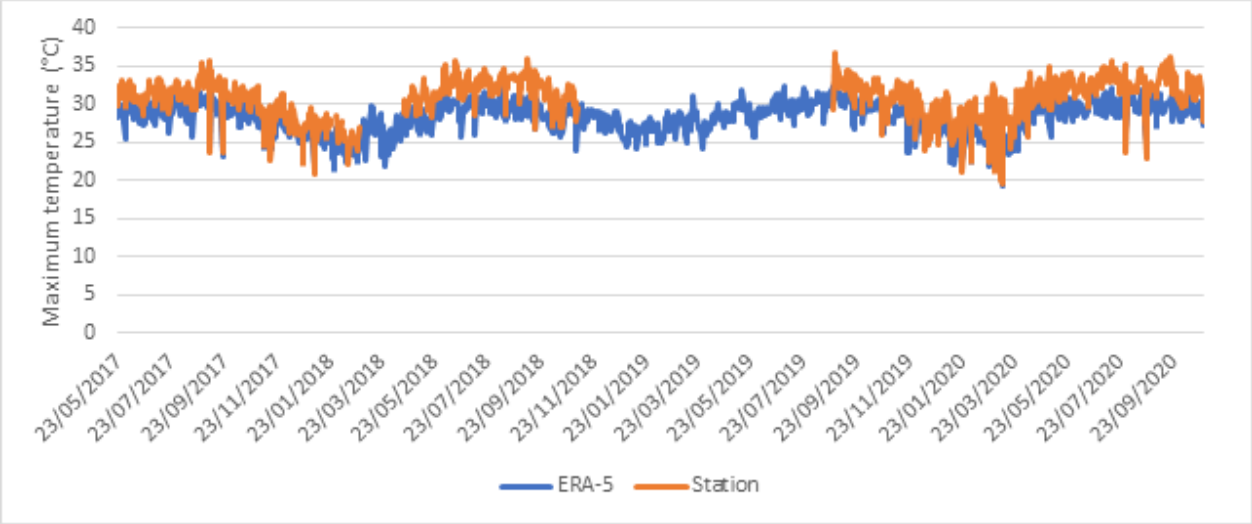

Figure 30: Comparison between station recorded daily maximum temperature and ERA-5 daily maximum temperature. Los Montones station

Figure 29 and Figure 30 show the comparison between minimum and maximum temperature retrieved from the Los Montones weather station and ERA-5. Overall, there is a good agreement between ERA-5 temperature estimates and station recorded temperature values. Minimum temperature recorded from the automatic station in the month of February 2018 significantly differs from the ERA-5 estimates; however, the station stopped recording immediately after this period and the values may be linked with a dysfunction in the sensor.

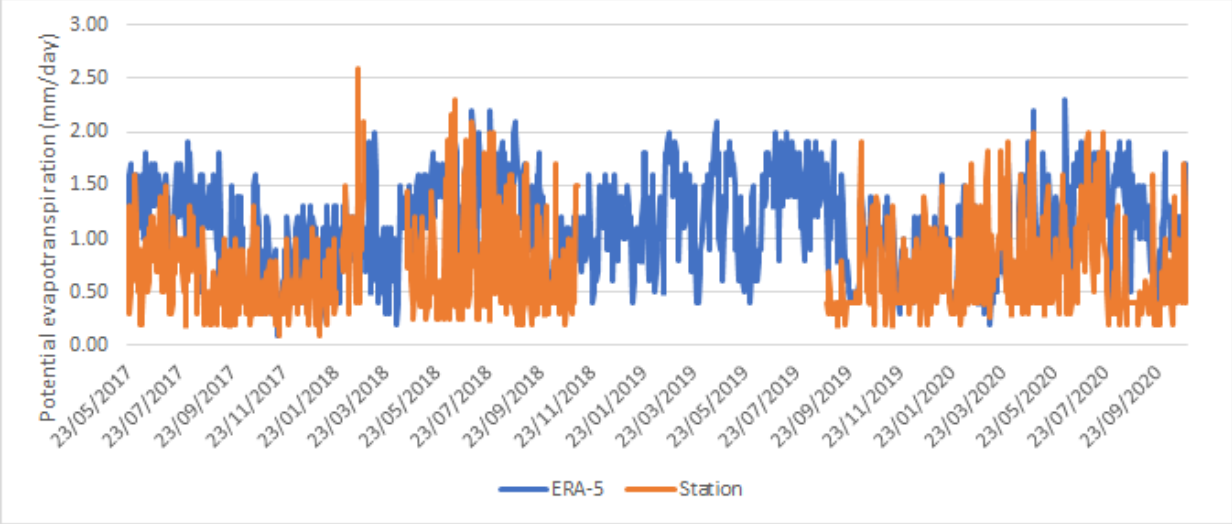

Figure 31: Comparison between potential evapotranspiration computed with station recorded data and ERA-5 data. Los Montones station

Potential evapotranspiration was computed according to the FAO Penman-Monteith method (Allen et al., 1998). Figure 31 shows the comparison between potential evapotranspiration computed with station recorded weather parameters and ERA-5 variables. The differences between the two series (ERA-5 reanalysis and station) are linked to the differences in rainfall estimates. In addition, the anomalous potential evapotranspiration values in February 2018 derive from the peak in the minimum temperature already discussed.

The considerations regarding the comparison between data recorded from weather stations and reanalysis data expressed in the case of Los Montones station hold also for the other three automatic weather stations.

After the comparison between weather variables from stations and ERA-5, the AquaCrop model was run. Rice was selected to perform the investigation since it is one of the major food crops grown in the country and is already implemented in AquaCrop. The attention on rice production in the Latin America context is high; for example, FAO has promoted the application of AquaCrop to estimate yield reduction in a climate change scenario in Colombia (Cortés B. et al., 2013).

Rice harvested area in the Dominican Republic is shown in Figure 32. Rice is mainly cultivated in the Northern part of the country according to the information retrieved from the MAP SPAM dataset.

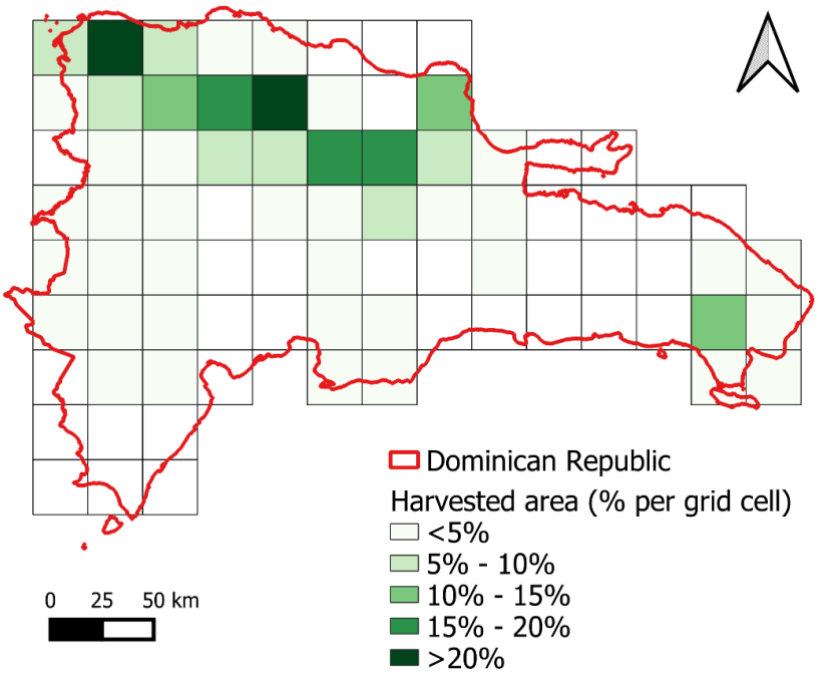

Figure 32: Rice harvested area, expressed as percentage of grid cell cultivated with rice.

Two rice yields per year were considered, based on the crop calendar shown in Figure 4. The main yield is sown in January and harvested between May and June, while the second yield is planted between April and May and harvested around November. The simulations reported in Table 17 were run. The simulations were selected according to the availability of meteorological input parameters from the automatic weather stations.

Table 17: Simulations considered for rice

Station	Yield	Year	Yield Station (ton/ha)	Yield ERA5 (ton/ha)
Los Montones	Second	2017	3.742	6.642
Los Montones	Second	2018	3.845	5.921
Los Montones	Second	2019	4.843	1.528
Los Montones	Second	2020	4.774	3.954
Los Montones	Main	2018	4.523	2.605
Los Montones	Main	2020	3.787	4.215
Jumunuco	Second	2017	2.181	6.676
Jumunuco	Second	2018	1.743	3.433
Jumunuco	Main	2018	9.168	8.814
Agua de Luis	Main	2017	1.883	4.126
Agua de Luis	Main	2018	0.423	0.164
Agua de Luis	Main	2019	0.076	0.136

Agua de Luis	Main	2020	0.701	0.361
Agua de Luis	Second	2017	0.23	0.082
Agua de Luis	Second	2018	0.423	0.065
Agua de Luis	Second	2019	0	0
Juliana Jaramillo	Second	2018	2.613	1.53
Juliana Jaramillo	Main	2019	1.304	0.089
Juliana Jaramillo	Second	2019	0.117	0.063
Juliana Jaramillo	Main	2020	0.117	0.203

The obtained results are promising. Figure 33 shows the relationship between rice yields computed running AquaCrop with station recorded data, which can be considered the most accurate estimate of yields that can be obtained at this resolution, and rice yield computed running AquaCrop with reanalysis data. The Pearson correlation coefficient, a measure of the strength of the relationship, is 0.779. The Pearson correlation coefficient ranges from -1 to 1, with -1 indicating a negative linear relationship among the parameters, 0 indicating the absence of a linear relationship between the parameters, and 1 indicating a positive linear relationship between the parameters. A Pearson correlation coefficient of 0.779 indicates the existence of a significant linear relationship between the yield computed with station data and the one computed with reanalysis data. The existence of the positive relationship shows the possibility to obtain good rice yield estimates even in the absence of station recorded meteorological data. The reanalysis dataset ERA-5 can be used as a proxy for weather variables to compute yield.

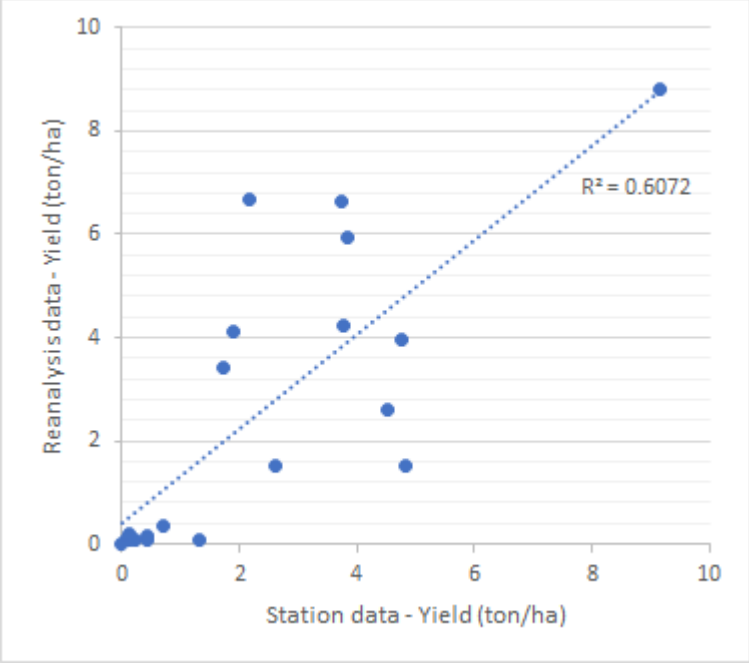

Figure 33: Comparison of results obtained running AquaCrop with station data and reanalysis data.

3.3.4 Discussion

The application of crop models in the design of parametric insurance programs is promising. The AquaCrop model run with station recorded data produced rice yield estimates similar to those obtained when the same model is run with gridded data from the ERA-5 reanalysis. The Pearson correlation coefficient between the rice yield computed running AquaCrop with station recorded data and rice yield computed running AquaCrop with reanalysis data is 0.779, which indicates a significant linear relationship between the two variables. The result is of high importance since it demonstrates the feasibility of using the ERA-5 reanalysis dataset to run Aquacrop. ERA-5 has a global coverage and provides data on a gridded basis from 1979, thus allowing the application of the same methodology adopted in this study in other countries and in different time periods. The use of the AquaCrop, combined with reanalysis data, can enhance the capability of identifying significant yield losses. This suggests that the proposed methodology can be applied to design more robust triggers to determine the program payout.

4 Pathway to operationalization

During the SMART project, significant effort was put into the potential future operationalization of the developed models. This was done along three main lines of action, which are described in this section.

4.1 Outreach end engagement with Dominican stakeholders

In close collaboration with the project's local partner, Fundación REDDOM, several outreach and dissemination actions were carried out with a variety of local stakeholders. Moreover, our team also engaged directly with farmer's associations in order to gather input and data that helped steer the project. Through these activities, significant interest on the SMART project has been garnered with local stakeholders; in addition, the methodological development was carried out according to local needs. This laid what we consider to be a strong foundation on which future operationalization may be achieved. Please refer to the following section (i.e. Monitoring and Evaluation) for a detailed description on these activities.

4.2 Identification of potential improvements in the Dominican insurance institutional framework

Another critical aspect that is required to support future operationalization of the project in the Dominican Republic is a good grasp on the local catastrophe insurance institutional context. This allows understanding the potential role that a new parametric insurance product can have in improving the existing system. As described in Section 2.2, the government supports the current risk management system directly and indirectly.

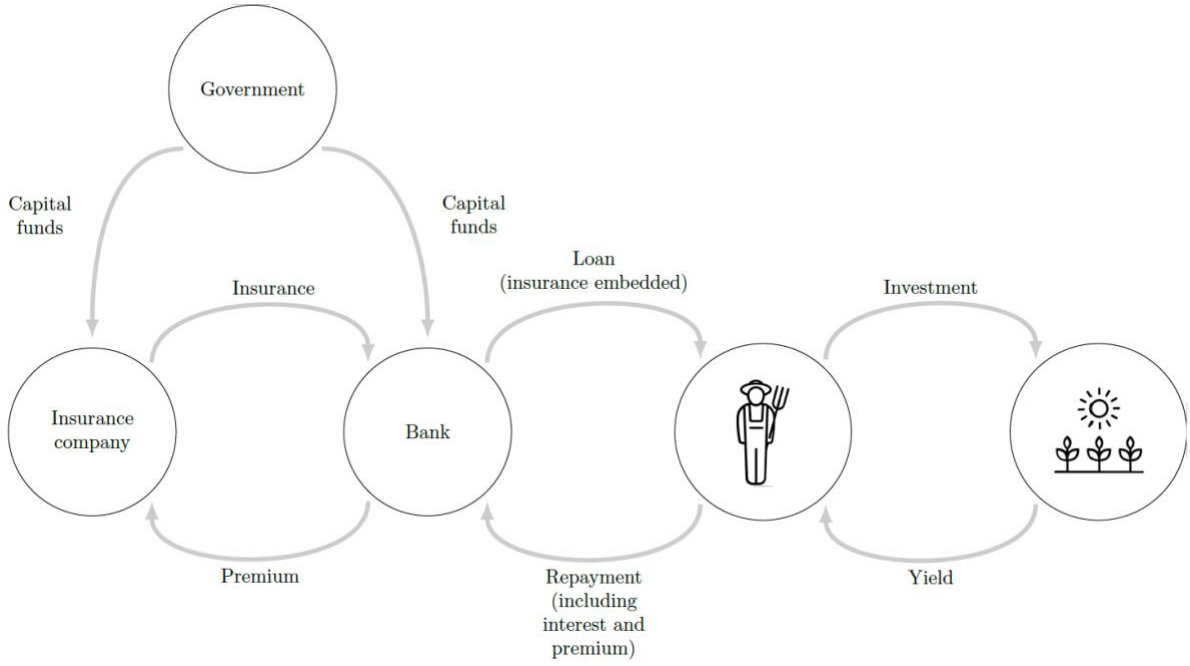

Figure 34: Schematic representation of the current risk management system in place in the Dominican Republic.

Although there may exist other specific cases, for the sake of simplicity here we refer to the situation where farmers are financed by the government. In this situation, reported in Figure 34:

- the farmers do not have all the financial resources to invest in farming and get a loan from the bank (Banco Agricola) at favourable conditions;
- the bank receives insurance coverage for the loan and capital funds from the government;
- the insurance company (Agrodosa) provides coverage for the loan and receives capital funds from the government (Ministry of Agriculture).

In normal conditions, when no loss event occurs, the farmers get the crop yield and repay the loan that includes the interests and the embedded insurance premium. In case of a loss event, the farmers lose the investment and the crop yield. A loss adjuster is appointed to estimate the loss, and a few weeks later, he sends the report to the insurance company. The insurance company pays an indemnity to the bank to compensate for the lost loan.

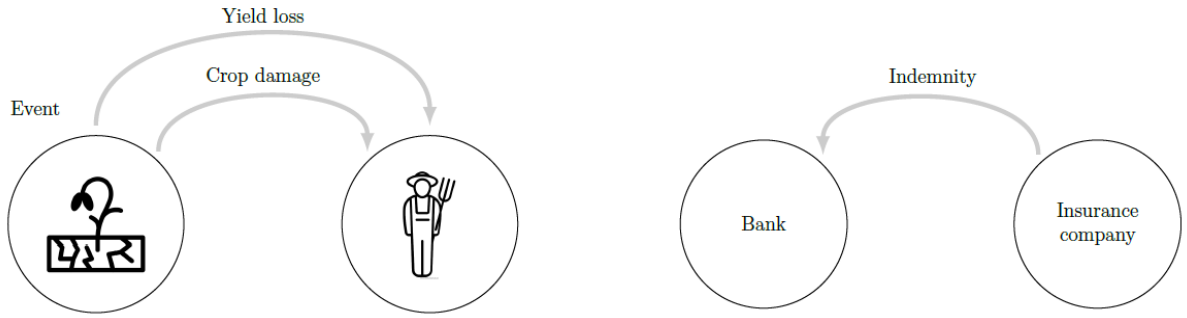

Figure 35: Schematic representation of the current risk management system in place in the Dominican Republic when an event occurs.

This system has several flaws:

- the farmers are those who are taking the risk, but they are not empowered, nor do they get the proper profit from their enterprise;
- the indemnity process is prolonged and frustrates recovery actions in the aftermath of an event;
- the contribution of the government is not directed to the farmers but to the bank and the insurance company.

The parametric insurance program developed in the SMART project enables a new risk management system. Besides the technical issues discussed in Section 3, a set of actions and agreements must be undertaken to operationalize the system.

In the new configuration, each stakeholder should have a different role:

- the farmers get a loan from the bank and a parametric insurance coverage from the insurance company; they invest the money in farming the crop;
- the bank pays 25% - 50% of the insurance premium to the insurance company;
- the government provides capital funds both to the bank and to the insurance company.

In normal conditions, when no loss event occurs, the farmers get the crop yield and repay the loan to the bank and a portion of the premium (50% - 75%) to the insurance company.

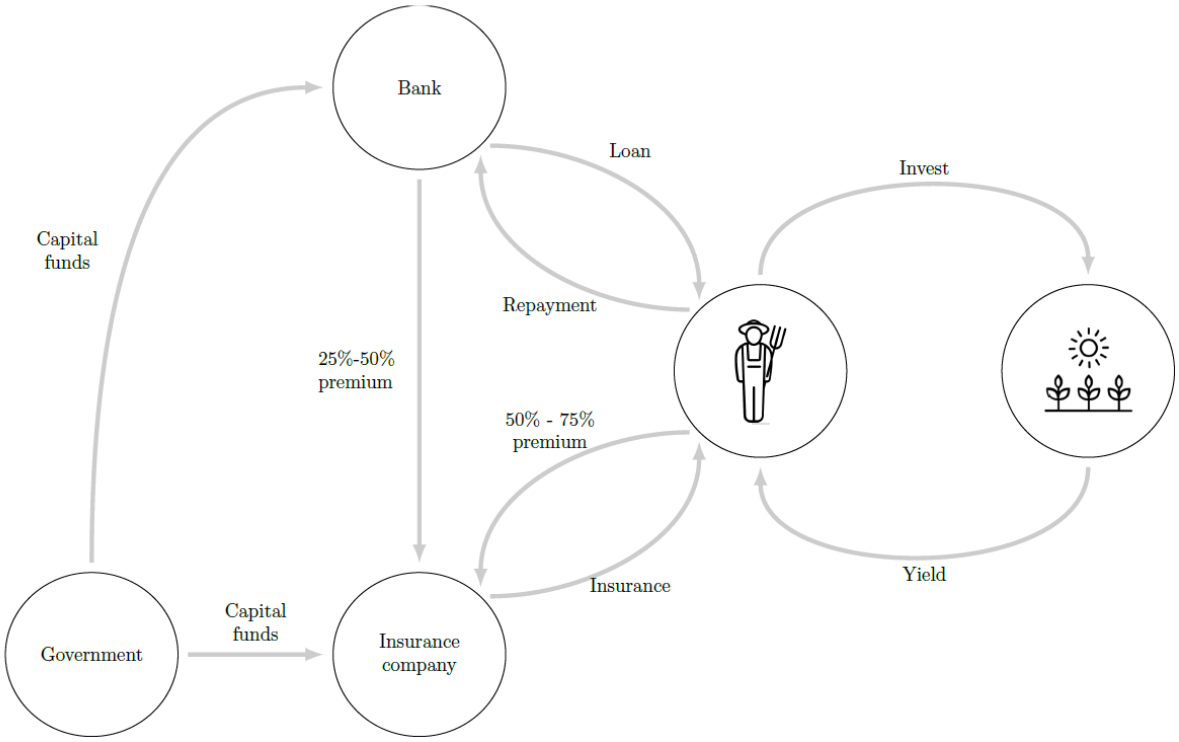

Figure 36: schematic representation of the proposed risk management system.

In case of a loss event, the farmer receives a swift payout from the insurance company based on the parametric model. Thanks to these available financial resources, the farmers put in place actions to recover the crops and reduce the yield loss. Farmers repay the loan to the bank and a portion of the premium (50% - 75%) to the insurance company.

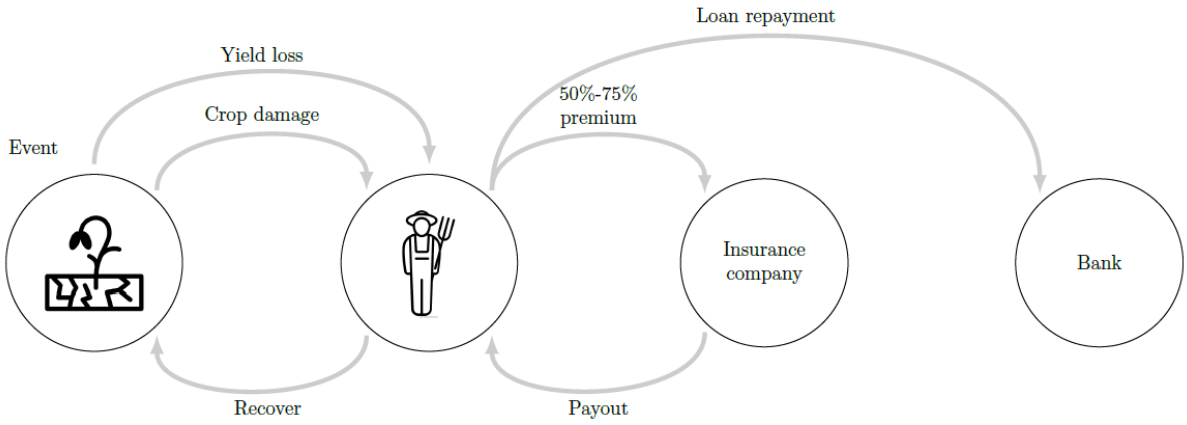

Figure 37: Schematic representation of the proposed risk management system in case of event.

In this system, there are many innovations:

1. Recognize the central role of the farmers: the new organization puts the farmer in the centre of the system;
2. Payout Timing: by providing a payout soon after the event, there are several advantages. (a) Farmers have financial resources to recover the crops and possibly get some yield

(otherwise lost); (b) They can repay the loan; (c) The overall loss for the system is reduced.

3. Raise awareness of the risk from natural hazards: the farmers will directly manage the natural hazard risk and take charges and profits from their actions and risk management strategy.

4.3 Scalability

A key feature of the methodologies proposed in the "Research and Development" section is their ease of scalability, i.e. the possibility to use them with an increasing number of meteorological and environmental data coming from different sources. In fact, the proposed methodologies can handle multiple types of input data and can be easily adjusted to exploit new products that are likely to become available in coming years. In addition, the proposed methodologies can in principle be easily transferable to other regions and/or countries.

Regarding the methodology for the identification of flood and drought events described in Section 3.1, because it is based on meteorological data from satellite and reanalysis products with global coverage, it can be transferred to other countries without the need to change input data. Furthermore, the catalogue of historical floods and droughts described in Section 3.1.1.2 was derived from international databases of natural disasters, which again have global coverage; therefore, historical event catalogues events may be developed to train models for other countries using the same approach. The selected databases also report hazards other than floods and droughts, meaning that they may be used in the development of analogous machine learning models for other hazards. Although different types of hazards naturally require different sets of input data, our model development framework streamlines the process of identifying the optimal set of data for each context, simplifying transferability.

The framework introduced to predict milk production is based on the methodology proposed for the identification of extreme events, therefore, similar conclusions about the scalability of the model can be drawn. Additionally, the model developed for predicting milk production takes advantage of the capability of these algorithms to transfer learning, already providing concrete proof of its applicability to other contexts. Indeed, the model was built using globally available environmental and meteorological variables and has been trained on European milk data before providing excellent results in the Dominican Republic, showcasing once again the possibility of employing the algorithm in any other countries of the world, granted that monthly milk production data (i.e., the targets of our predictions) are available. The algorithm can also be scaled to predict crop production if appropriate input data are employed.

Lastly, the methodology developed to apply crop models in the context of parametric insurance is again applicable in countries other than the Dominican Republic. The meteorological and environmental variables come from the already described ERA-5 reanalysis dataset, which has global coverage, while the datasets used to identify crop growing areas and parametrize soil features have global coverage. The AquaCrop model, developed by the Food and Agriculture Organization, has already been applied in many countries of the world. Thus, the methodology should be easily scalable and transferable to other regions and/or crop types.

4.4 Web app development and open-source code

Our team is currently finalizing an interactive web app that illustrates some of the developed models. The platform aims to facilitate dissemination and understanding of the SMART project outcomes and highlight their potential in the Dominican context. The framework adopted for the development of the web app is Shiny (Chang et al., 2021), an R package that facilitates the build of interactive web apps thanks to its twofold flexibility: the wide range of customizable features aimed at creating the desired layout and the liberty to host the app on webpages and/or embed them into documents.

The interpretation of the results coming from machine learning models is not always trivial and might require some specific knowledge on the topic. The main purpose of this app is to showcase the potential of the models developed during the project and facilitate the understanding of the results obtained by allowing the users to interactively "touch" these models. Depending on the user's final aim, navigating through the app might provide insightful information for decision makers as well as assistance to farmers interested in a parametric insurance product to protect its crop.

The code used to create the web app together with the code used to train and develop the machine learning algorithms are provided as open-source software and can be found on Github at the following link: https://github.com/luigicesarini/SMART

The web app will undergo a first testing period on the RStudio server at the following link: https://luigic.shinyapps.io/SMART/

Once the stability of the app is consolidated, our team is considering its migration to the local partner website (http://climared.com/).

The layout of the app has been designed to have different sections, each related to a different topic addressed during the project. Clicking on the link, the user will be directed to the homepage where the aim of the project and the partners are described (Figure 38).

Figure 38: Homepage of the SMART web app

The other sections are designed to have a static welcoming page where information to the specific topic are provided along with in depth looks at results and references. The welcoming page is then linked to an interactive page where the user, through the graphical user interface, setting some parameters is able to look at different results. For example, Figure 39 shows the prediction of a flood event for the 17th of April 2007 together with the precipitation maps for that day for the four rainfall datasets used as input.

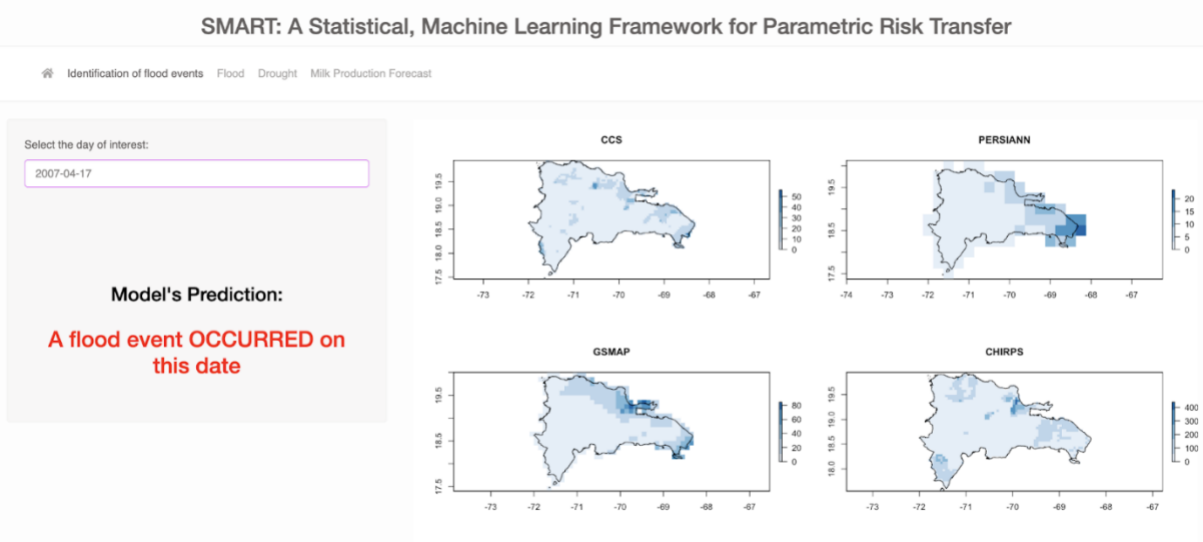

Figure 39: Interactive page of the SMART web app

The interactive app was built mainly to bring Dominican farmers and decision-makers closer to the topics of parametric insurance and machine learning, providing a simple and easy to use tool.

5 Monitoring and evaluation

During the SMART project, in parallel with the research and development activities of state-of-the-art models for parametric insurance, significant effort was put on their potential future operationalization in the Dominican Republic. In order to report back to donors and provide an objective basis to evaluate the success of the project in terms of fostering operationalization, in this document we provide a discussion related with the applicable monitoring and evaluation criteria, as agreed with the GFDRR. These monitoring and evaluation criteria are being delivered ahead of the project's final report, as requested by the GFDRR. The contents of this document will also be present in the final report.

a. Forming or strengthening local partnerships

At the core of the SMART project and its potential future operationalization is the partnership established between the project team members. In particular, Fundación REDDOM, the local project partner, is a Dominican non-governmental, non-profit organization that promotes sustainable development through the introduction of associative, technical and entrepreneurial innovation. Its team has specific expertise in the design, implementation, and evaluation of development projects, including climate resilience and adaptation measures, access to financial markets for farmers, and food security. REDDOM plays a central role in the agricultural sector in the Dominican Republic, promoting sustainable rural development by identifying competitive solutions and managing innovative resources and processes based on dialogue with community actors. REDDOM establishes collaborative relationships between the private sector, international cooperation, local government, universities, non-governmental organizations, and multinational corporations to facilitate economic growth.

The direct involvement of REDDOM is therefore key to providing a bridge between model development, implementation, and on-the-ground operationalization for any potential insurance product for the agricultural sector. REDDOM's wide network of contacts is the optimal basis upon which partnerships between local stakeholders, final beneficiaries, and/or foreign institutions can be established, and it is through such partnerships that the access to and use of disaster risk financing innovations in the Dominican Republic can be promoted. Thus, from its inception, the SMART project aimed to establish a robust pathway for future operationalization leveraging on REDDOM's pivotal role within the Dominican context. In this regard, REDDOM already has numerous well-established and ongoing partnerships with some of the most relevant stakeholders and possible end-users from the agricultural sector for potential weather index insurance products. Therefore, concerning local partnerships that can contribute towards future implementation of the developed models, SMART focused its efforts on 1) strengthening existing partnerships between REDDOM and relevant local stakeholders, and 2) fostering the creation and/or strengthening of partnerships among those stakeholders.

Throughout its duration, REDDOM presented SMART and its objectives to a range of Dominican stakeholders, and there was general interest in the project. Several farmers associations representing different crops were involved in the development of SMART, namely ADOBANANO (bananas), CONACADO (cocoa) and FEDEGANORTE (milk). In addition, other 11 farmers associations were contacted and provided input and/or data that supported the development of the project (Bloque Cacao Hato Mayor, ASOPROCON, BANELINO, FIKSAR/

BANANO, Grupo Banamiel, JAD, EKOBAN, Asociación de Ganaderos Villa Los Almácigos, Asociación de Ganaderos Juancito Fermín, Asociación de Ganaderos Dajabon, Asociación de Ganaderos Partido).

Likewise, REDDOM held meetings with several organizations to discuss the potential use of machine learning tools - as developed by the project - for strengthening both risk transfer mechanisms - i.e. agricultural insurance - and climate-smart decision-making in the agricultural sector. These organizations and Dominican government entities include the following: DIGERA (General Directorate of Agricultural Risks), AGRODOSA (agricultural insurer company), Ministry of Agriculture, Ministry of Environment, ONAMET (National Weather Office), INDHRI (National Instituto of Water Resources –in charge of irrigation channels throughout the country), ISA University (local university focused in agriculture) and MEPyD (Ministry of Economy, Planning and Development). In particular, AGRODOSA and DIGERA see great potential in using these tools to evaluate risk, improve insurance policies and provide a more accurate and cost-efficient service to farmers. Similarly, the Ministry of Agriculture's planning department envisions the use of these tools to integrate climate risks management into their annual plans.

These activities allowed not only better adjusting SMART to the local context through information and data provided directly by local stakeholders, but also disseminating the project among such stakeholders, which is a critical first step towards future operationalization. Through these activities, we consider that existing partnerships between REDDOM and local stakeholders have been reinforced, especially in the context of a possible future implementation of a parametric insurance product.

In addition, the most important milestone of SMART in terms of promoting its outcomes and fostering local partnerships was the project workshop that took place on 5 March 2021. This workshop, which we had envisaged from the beginning of the project, was planned to physically take place in the Dominican Republic but had to instead be organized as a virtual workshop due to the ongoing COVID-19 pandemic. Nevertheless, even in this format, we consider that the workshop was highly productive. Stakeholders from different sectors with whom REDDOM interacted during the SMART project were invited. The workshop was attended by the participants shown in Table 1.

Table 18: List of participants in the SMART workshop

Role	Participant name	Institution
SMART project team	Jeffery Perez	Fundación REDDOM
	Luis Tolentino	
	Emilin Sena	
	Nidia de los Santos	
	Mario Martina	University School for Advanced Studies IUSS Pavia
	Rui Figueiredo	
	Luigi Cesarini	
	Beatrice Monteleone	

Dominican Republic government ministries	Juan Mancebo	Ministry of Agriculture, Directorate of Risk Management and Climate Change
	Kenia Amarilis	Ministry of Environment
	Kirverlin Valera	
	Maria Ramos	
	Mary Galan	
	Zoreydi Medina	
	Miguel Montero	Ministry of Economy, Planning and Development
Dominican Republic governmental institutions	Miriam Matos	National Weather Bureau (ONAMET)
	Quisqueia Perez	Dominican Coffee Institute (INDOCAFE)
	Kioris Alcantara	National Institute for Forest and Agriculture Research (IDIAF)
Farmers association	Judelka Reyes	FEDEGANO (Dairy Farmers Federation of the Northwest Region, in representation of 44 smaller associations)
Insurance	Miguel Marrero	AGRODOSA (Agriculture Insurance Company)

The workshop was organized into three parts. In the first part, each participant presented himself/herself and described the institution he/she represents. In the second part, REDDOM delivered a presentation in Spanish to the audience. This presentation included a description of the Challenge Fund, the SMART consortium, and the project goals. It then went on to provide general information about parametric insurance, artificial intelligence, and how machine learning may be used to improve models and better manage climate risks. Finally, a selection of results was shown. The presentation was delivered in a flexible and interactive manner, which allowed several interesting questions by the audience to be addressed as the presentation progressed. The presentation slides will be provided as an attachment to the final report.

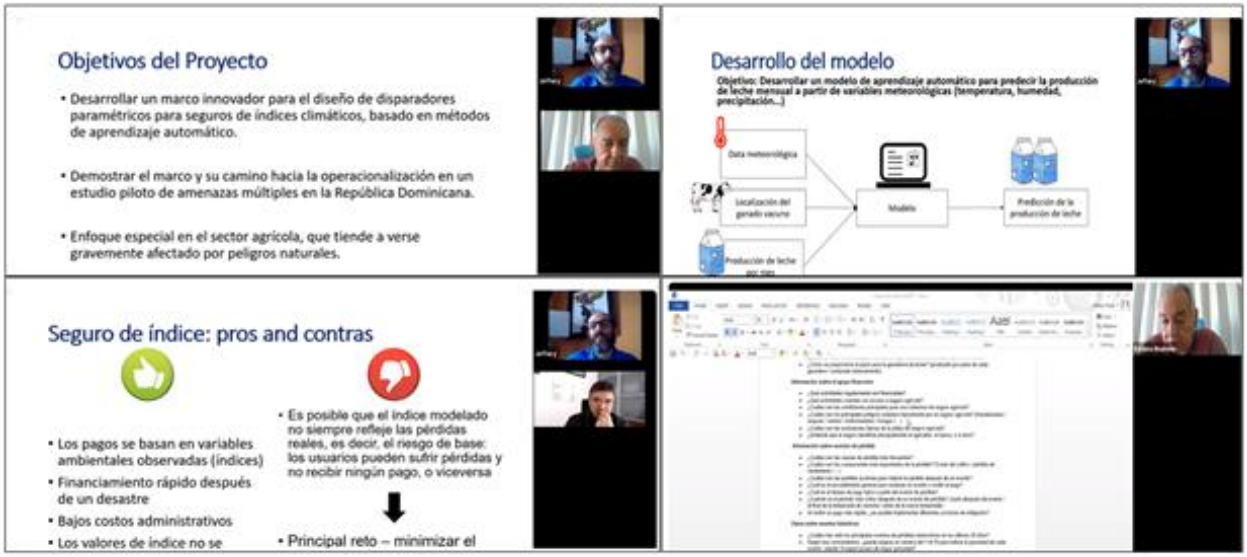

Figure 40: Some snapshots of the project workshop

Lastly, the third part of the workshop consisted in a roundtable discussion among all the participants, with the moderation of REDDOM. During this discussion, several stakeholders showed significant interest in the implementation of index insurance to the Dominican agricultural sector and in the outcomes of the SMART project. Moreover, different aspects of the Dominican agricultural reality were discussed. Overall, we consider that the workshop met its objectives entirely, having exceeded our initial expectations: all participant stakeholders are now well informed about the project, and the groundwork for future partnerships and for operationalization has been laid.

b. Co-development/refinement of the risk financing innovation with beneficiaries

As mentioned in the previous point, different potential beneficiaries of a weather index insurance product (e.g., farmers and farmers associations related with cocoa, bananas, and dairy) were involved in the development of the parametric insurance framework. Their inputs and data provided played a key role in deciding which crops and which regions to focus on during model development. Further refinement of the models together with the target beneficiaries is feasible, but pertinent only within a subsequent stage of model implementation closer to actual operationalization.

c. Developing or supporting local champion(s)

The SMART local partner, Fundación REDDOM, was involved in the development of the machine learning models that can be used to support innovative index insurance products for the Dominican agricultural sector. Given its unique positioning at the intersection of 1) specific technical expertise on the SMART project and on the developed models, and 2) its wide network of contacts and well-established partnerships with many of the relevant stakeholders, REDDOM will naturally be the main local champion that enable further dissemination and access of the innovation to potential users.

In addition, the Dominican farmers' organizations are also expected to play a central role championing SMART in the context of a possible future operationalization, since they are in

principle able to establish a bridge between farmers and institutional stakeholders. For example, FEDEGANO, which represent 44 dairy farmers' associations of the northwest region of the Dominican Republic, participated and was actively engaged in the project workshop. It is now well informed about the project, and in the context of a subsequent stage of operationalization it is expected to play a key role in disseminating and promoting the innovative index-based insurance solution among its associates. Based on the experience of this workshop, we believe that a similar level of engagement with associations representing other crops and regions is possible and likely.

d. Training of local beneficiaries

REDDOM participated in the co-development of the models and its team and is therefore knowledgeable on most of the project's technical components. Regarding potential beneficiaries, a significant amount of information regarding index-based insurance and the SMART project was provided during the project workshop. This workshop served partly as a dissemination action and partly a training action, where participants apprehended a number of core concepts that are necessary to understand the project and how it can be beneficial for users and the Dominican agricultural sector. Additional, more specific training is feasible at a subsequent stage of implementation and operationalization. It is also worth noting that our team is also finalizing an interactive web app that aims to facilitate dissemination and understanding of the SMART project outcomes.

e. Cash or in-kind leverage toward further access or use of the developed risk financing innovation

The SMART team is available and willing to continue supporting the project for a transitory period after the end of the contract towards potential operationalization in the Dominican Republic, as well as to explore the possibility of applying and/scaling the models to other contexts.

f. Using information on gender gaps relevant to your risk financing innovation to try to close those gaps

Not applicable, given the scope of the project.

To conclude, we consider that SMART has established a clear pathway to operationalization from both a technical and a practical viewpoint, and that all the conditions are in place to achieve the intended implementation of project outcomes through local partnerships and collaborations.

6 References

Afshar, M. H., Foster, T., Higginbottom, T. P., Parkes, B., Hufkens, K., Mansabdar, S., … Kramer, B. (2021). Improving the Performance of Index Insurance Using Crop Models and Phenological Monitoring. *Remote Sensing*, *13*(5), 924. doi:10.3390/rs13050924

Aksoy, S., and Haralick, R. M. (2001). Feature normalization and likelihood-based similarity measures for image retrieval. *Pattern Recognition Letters*, *22*(5), 563–582. doi:10.1016/S0167-8655(00)00112-4

Allen, R. G., Pereira, L. S., Raes, D., and Smith, M. (1998). *Crop Evapotranspiration: Guidelines for computing crop water requirements - FAO Irrigation and drainage paper 56*. (Food and Agriculture Organization, Ed.). Retrieved from http://www.fao.org/3/x0490e/x0490e00.htm#Contents

Asseng, S., Zhu, Y., Basso, B., Wilson, T., and Cammarano, D. (2014). Simulation Modeling: Applications in Cropping Systems. In *Encyclopedia of Agriculture and Food Systems* (pp. 102–112). Elsevier. doi:10.1016/B978-0-444-52512-3.00233-3

Barredo, J. I. (2007). Major flood disasters in Europe: 1950-2005. *Natural Hazards*, *42*(1), 125–148. doi:10.1007/s11069-006-9065-2

Bolvin, D. T., Braithwaite, D., Hsu, K., Joyce, R., Kidd, C., Nelkin, E. J., … Xie, P. (2018). *NASA Global Precipitation Measurement (GPM) Integrated Multi-satellitE Retrievals for GPM (IMERG). Algorithm Theoretical Basis Document (ATBD) Version 5.2*. Retrieved from https://pmm.nasa.gov/sites/default/files/document_files/IMERG_ATBD_V5.2_0.pdf

Bontempi, G., Ben Taieb, S., and Le Borgne., Y.-A. ̈el. (2013). Machine Learning Strategies for Time Series Forecasting. In *Business Intelligence*. doi:https://doi.org/10.1007/978-3-642-36318-4_3

Boser, B. E., Vapnik, V. N., and Guyon, I. M. (1992). Training Algorithm Margin for Optimal Classifiers. *Perception*, 144–152.

Bowden, G. J., Dandy, G. C., and Maier, H. R. (2005). Input determination for neural network models in water resources applications. Part 1 - Background and methodology. *Journal of Hydrology*, *301*(1–4), 75–92. doi:10.1016/j.jhydrol.2004.06.021

Calvet, L., Lopeman, M., de Armas, J., Franco, G., and Juan, A. A. (2017). Statistical and machine learning approaches for the minimization of trigger errors in parametric earthquake catastrophe bonds. *SORT*, *41*(2), 373–391. doi:10.2436/20.8080.02.64

Campo, K.-R. K., Robinson, A., Isaac, W.-A. P., and Ganpat, W. (2017). Connecting Small Farmers in the Caribbean to Knowledge, Networks, and Institutions Through ICTs. *Journal of Agricultural & Food Information*, *18*(2), 81–95. doi:10.1080/10496505.2017.1279973

Caraviello, D. Z., Weigel, K. A., Craven, M., Gianola, D., Cook, N. B., Nordlund, K. V., … Wiltbank, M. C. (2006). Analysis of reproductive performance of lactating cows on large dairy farms using machine learning algorithms. *Journal of Dairy Science*, *89*(12), 4703–4722. doi:10.3168/jds.S0022-0302(06)72521-8

Chang, W., Cheng, J., Allaire, J., Sievert, C., Schloerke, B., Xie, Y., … Borges, B. (2021). shiny: Web Application Framework for R.

Chawla, N. V., Bowyer, K. W., Hall, L. O., and Kegelmeyer, W. P. (2002). SMOTE: Synthetic Minority Over-sampling Technique. *Journal of Artificial Intelligence Research*, *16*, 321–357.

doi:10.1613/jair.953

Chen, T., Ren, L., Yuan, F., Tang, T., Yang, X., Jiang, S., … Zhang, L. (2019). Merging ground and satellite-based precipitation data sets for improved hydrological simulations in the Xijiang River basin of China. *Stochastic Environmental Research and Risk Assessment*, *33*(10), 1893–1905. doi:10.1007/s00477-019-01731-w

Chiang, Y. M., Hsu, K. L., Chang, F. J., Hong, Y., and Sorooshian, S. (2007). Merging multiple precipitation sources for flash flood forecasting. *Journal of Hydrology*, *340*(3–4), 183–196. doi:10.1016/j.jhydrol.2007.04.007

Chollet, F. (2017). *Deep Learning with Python*.

Cortés B., C. A., Bernal, J., Díaz A., E., and Méndez, J. (2013). *Uso del modelo Aquacrop para estimar rendimientos para el cultivo de maiz en los departamentos de Córdoba, Meta, Tolima y Valle de Cauca*. *Fao*.

Deng, X., Barnett, B. J., Vedenov, D. V., and West, J. W. (2007). Hedging dairy production losses using weather-based index insurance. *Agricultural Economics*, *36*(2), 271–280. doi:10.1111/j.1574-0862.2007.00204.x

Dreyfus, G. (2005). *Neural Networks: Methodology and Applications*. Springer.

Eckstein, D., Künzel, V., and Schäfer, L. (2017). *Global Climate Risk Index 2018. Who Suffers Most From Extreme Weather Events? Weather-related Loss Events in 2016 and 1997 to 2016*. Retrieved from http://www.germanwatch.org/en/cri

ECMWF, Copernicus, and Climate Change Service. (2018). ERA5 hourly data on single levels from 1979 to present. doi:10.24381/cds.adbb2d47

European Drought Observatory. (2020). Standardized Precipitation Index (SPI). European Commission. Retrieved from https://edo.jrc.ec.europa.eu/documents/factsheets/factsheet_spi.pdf

Figueiredo, R., Martina, M. L. V, Stephenson, D. B., and Youngman, B. D. (2018). A Probabilistic Paradigm for the Parametric Insurance of Natural Hazards. *Risk Analysis*, *38*(11), 2400–2414. doi:10.1111/risa.13122

Filippi, P., Jones, E. J., Wimalathunge, N. S., Somarathna, P. D. S. N., Pozza, L. E., Ugbaje, S. U., … Bishop, T. F. A. (2019). An approach to forecast grain crop yield using multi-layered, multi-farm data sets and machine learning. *Precision Agriculture*, *20*(5), 1015–1029. doi:10.1007/s11119-018-09628-4

Flach, P. A., and Kull, M. (2015). Precision-Recall-Gain curves: PR analysis done right. *Advances in Neural Information Processing Systems*, *2015-Janua*, 838–846.

Food and Agriculture Organization. (2012). Introducing Aquacrop.

Food and Agriculture Organization. (2017). AquaCrop, the crop water productivity model. Retrieved from http://www.fao.org/3/a-i7455e.pdf

Fung, K. F., Huang, Y. F., Koo, C. H., and Soh, Y. W. (2019). Drought forecasting: A review of modelling approaches 2007–2017. *Journal of Water and Climate Change*. doi:10.2166/wcc.2019.236

Funk, C., Peterson, P., Landsfeld, M., Pedreros, D., Verdin, J., Shukla, S., … Michaelsen, J. (2015). The climate hazards infrared precipitation with stations—a new environmental record for monitoring extremes. *Scientific Data*, *2*(1), 150066. doi:10.1038/sdata.2015.66

Garcia, V., Sanchez, J. S., and Mollineda, R. A. (2012). On the effectiveness of preprocessing

methods when dealing with different levels of class imbalance. *Knowledge-Based Systems*, *25*(1), 13–21. doi:10.1016/j.knosys.2011.06.013

Garry, F. K., Bernie, D. J., Davie, J. C. S., and Pope, E. C. D. (2021). Future climate risk to UK agriculture from compound events. *Climate Risk Management*, *32*, 100282. doi:10.1016/j.crm.2021.100282

Ghahramani, Z. (2004). Unsupervised learning. In O. Bousquet, U. von Luxburg, and G. Ratsch (Eds.), *Advanced Lectures on Machine Learning* (I., pp. 72–112). Tubingeb: Springer-Verlag Berlin Heidelberg.

Grzesiak, W., Błaszczyk, P., and Lacroix, R. (2006). Methods of predicting milk yield in dairy cows-Predictive capabilities of Wood's lactation curve and artificial neural networks (ANNs). *Computers and Electronics in Agriculture*, *54*(2), 69–83. doi:10.1016/j.compag.2006.08.004

Hao, Z., Singh, V. P., and Xia, Y. (2018). Seasonal Drought Prediction: Advances, Challenges, and Future Prospects. *Reviews of Geophysics*, *56*(1), 108–141. doi:10.1002/2016RG000549

Herrera, D., and Ault, T. R. (2017). Insights from a New High-Resolution Drought Atlas for the Caribbean spanning 1959 to 2016. *Bulletin of the American Meteorological Society*, *30*(September). doi:10.1175/JCLI-D-16-0838.1

Hochreiter, S., and Schmidhuber, J. (1997). Long Short-Term Memory. *Neural Computation*, *9*(8), 1735–1780. doi:10.1162/neco.1997.9.8.1735

Hong, Y., Hsu, K. L., Sorooshian, S., and Gao, X. (2004). Precipitation estimation from remotely sensed imagery using an artificial neural network cloud classification system. *Journal of Applied Meteorology*, *43*(12), 1834–1852. doi:10.1175/jam2173.1

Hossin, M., and Sulaiman, M. N. (2015). A Review on Evaluation Metrics for Data Classification Evaluations. *International Journal of Data Mining & Knowledge Management Process*, *5*(2), 01–11. doi:10.5121/ijdkp.2015.5201

International Food Policy Research Institute. Global Spatially-Disaggregated Crop Production Statistics Data for 2010 Version 2.0 (2019). doi:https://doi.org/10.7910/DVN/PRFF8V

Izzo, M., Rosskopf, C., Aucelli, P., Maratea, A., Méndez, R., Pérez, C., and Segura, H. (2010). A new climatic map of the dominican republic based on the thornthwaite classification. *Physical Geography*, *31*(5), 455–472. doi:10.2747/0272-3646.31.5.455

Johnson, M. D., Hsieh, W. W., Cannon, A. J., Davidson, A., and Bédard, F. (2016). Crop yield forecasting on the Canadian Prairies by remotely sensed vegetation indices and machine learning methods. *Agricultural and Forest Meteorology*, *218–219*, 74–84. doi:10.1016/j.agrformet.2015.11.003

Jones, K. S., and Van Rijsbergen, C. J. (1976). Progress in documentation: Information retrieval test collections. *Journal of Documentation*, *32*(1), 59–75. doi:10.1108/eb026616

Joyce, R. J., Janowiak, J. E., Arkin, P. A., and Xie, P. (2004). CMORPH: A method that produces global precipitation estimates from passive microwave and infrared data at high spatial and temporal resolution. *Journal of Hydrometeorology*, *5*(3), 487–503. doi:10.1175/1525-7541(2004)005<0487:CAMTPG>2.0.CO;2

Keating, B. ., Carberry, P. ., Hammer, G. ., Probert, M. ., Robertson, M. ., Holzworth, D., … Smith, C. . (2003). An overview of APSIM, a model designed for farming systems simulation. *European Journal of Agronomy*, *18*(3–4), 267–288. doi:10.1016/S1161-0301(02)00108-9

Kim, K. S., Gleason, M. L., and Taylor, S. E. (2006). Forecasting site-specific leaf wetness duration for input to disease-warning systems. *Plant Disease*, *90*(5), 650–656. doi:10.1094/PD-90-0650

Krzanowski, W., and Hand, D. J. (2009). *ROC Curves for Continuous Data* (Vol. 111). Boca Raton, FL: Chapman and Hall/CRC. doi:10.1201/9781439800225

Lee, W., Stolfo, S. J., Ling, C. X., Schultz, M. G., Eskin, E., Zadok, E., … Mitra, P. (2008). Data Mining for Direct Marketing: Problems andLing, Charles X. *Proceedings of the 7th USENIX Security Symposium*, *98*(1), 38–49. doi:10.1109/72.977258

Maini, V., and Sabri, S. (2017). *Machine Learning for Humans*. (S. Maini, Ed.)*Medium*. Retrieved from https://everythingcomputerscience.com/books/Machine Learning for Humans.pdf

Mas, J. F., and Flores, J. J. (2008). The application of artificial neural networks to the analysis of remotely sensed data. *International Journal of Remote Sensing*, *29*(3), 617–663. doi:10.1080/01431160701352154

McClure, N. (2017). *TensorFlow Machine Learning*. Retrieved from https://www.tensorflow.org/federated

McKee, T. B., Doesken, N. J., and Kleist, J. (1993). The relationship of drought frequency and duration to time scales. *AMS 8th Conference on Applied Climatology*, 179–184.

Mosavi, A., Ozturk, P., and Chau, K. W. (2018). Flood prediction using machine learning models: Literature review. *Water (Switzerland)*, *10*(11), 1536. doi:10.3390/w10111536

Mosteller, F., and Tukey, J. W. (1968). Data analysis, including statistics. *Handbook of Social Psychology*, *2*, 80–203.

Mueller, J., and Massaron, L. (2016). *Machine learning for dummies*. John Wiley & Sons, Inc.

Nassiri Mahallati, M. (2020). Advances in modeling saffron growth and development at different scales. In *Saffron* (pp. 139–167). Elsevier. doi:10.1016/B978-0-12-818638-1.00009-5

Parkes, B., Higginbottom, T. P., Hufkens, K., Ceballos, F., Kramer, B., and Foster, T. (2019). Weather dataset choice introduces uncertainty to estimates of crop yield responses to climate variability and change. *Environmental Research Letters*, *14*(12), 124089. doi:10.1088/1748-9326/ab5ebb

Payano-Almanzar, R., and Rodriguez, J. (2018). Meteorological, Agricultural and Hydrological Drought in the Dominican Republic: A review. *Current World Environment*, *13*(1), 124–143. doi:10.12944/CWE.13.1.12

Pedregosa, F., Varoquaux, G., Buitinck, L., Louppe, G., Grisel, O., and Mueller, A. (2011). Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, *12*, 2825–2830.

Platt, J. C. (1999). Probabilistic Outputs for Support Vector Machines and Comparisons to Regularized Likelihood Methods. In *Advances in Large Margin Classifiers* (pp. 61–74). MIT Press.

Rankine, D. R., Cohen, J. E., Taylor, M. A., Coy, A. D., Simpson, L. A., Stephenson, T., and Lawrence, J. L. (2015). Parameterizing the FAO AquaCrop Model for Rainfed and Irrigated Field-Grown Sweet Potato. *Agronomy Journal*, *107*(1), 375–387. doi:10.2134/agronj14.0287

Saito, T., and Rehmsmeier, M. (2015). The precision-recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets. *PLoS ONE*, *10*(3), 1–21. doi:10.1371/journal.pone.0118432

Samuel, A. L. (1959). Eight-move opening utilizing generalization learning. (See Appendix B,

Game G-43.1 Some Studies in Machine Learning Using the Game of Checkers. *IBM Journal*, 210–229.

Shine, P., Murphy, M. D., Upton, J., and Scully, T. (2018). Machine-learning algorithms for predicting on-farm direct water and electricity consumption on pasture based dairy farms. *Computers and Electronics in Agriculture*, *150*(April 2018), 74–87. doi:10.1016/j.compag.2018.03.023

Soil Science Division Staff. (2017). *Soil Survey Manual, USDA*.

Sorooshian, S., Hsu, K. L., Gao, X., Gupta, H. V., Imam, B., and Braithwaite, D. (2000). Evaluation of PERSIANN system satellite-based estimates of tropical rainfall. *Bulletin of the American Meteorological Society*, *81*(9), 2035–2046. doi:10.1175/1520-0477(2000)081<2035:EOPSSE>2.3.CO;2

Steduto, P., Hsiao, T. C., Fereres, E., and Raes, D. (2012). *Crop yield response to water*. (Food and Agriculture Organization, Ed.). Rome.

Stevens, E., and Antiga, L. (2019). *Deep Learning with PyTorch Essential Excerpts*. Shelter Island, New York: Manning Publications Co.

Tate, E. L., and Gustard, A. (2000). Drought Definition: A Hydrological Perspective, 23–48. doi:10.1007/978-94-015-9472-1_3

United Nations Office for Disaster Risk Reduction. (1994). DesInventar. Retrieved from https://www.desinventar.net

United States Department of Agriculture. (2008). Crop Calendars for Mexico, Central America and the Caribbean. Retrieved May 6, 2021, from https://ipad.fas.usda.gov/rssiws/al/crop_calendar/ca.aspx

USAID, and REDDOM. (2013). *Climate Resiliency and Index Insurance for Small Farmers in the Dominican Republic: Market Research Assessment*.

Ushio, T., and Kachi, M. (2010). Kalman Filtering Applications for Global Satellite Mapping of Precipitation (GSMaP). In M. Gebremichael and F. Hossain (Eds.), *Satellite Rainfall Applications for Surface Hydrology* (pp. 105–123). Dordrecht: Springer Netherlands. doi:10.1007/978-90-481-2915-7_7

Ushio, T., Sasashige, K., Kubota, T., Shige, S., Okamoto, K., Aonashi, K., … Kawasaki, Z.-I. (2009). A Kalman Filter Approach to the Global Satellite Mapping of Precipitation (GSMaP) from Combined Passive Microwave and Infrared Radiometric Data. *Journal of the Meteorological Society of Japan. Ser. II*, *87A*, 137–151. doi:10.2151/jmsj.87A.137

Wang, A. X., Tran, C., Desai, N., Lobell, D., and Ermon, S. (2018). Deep transfer learning for crop yield prediction with remote sensing data. *Proceedings of the 1st ACM SIGCAS Conference on Computing and Sustainable Societies, COMPASS 2018*. doi:10.1145/3209811.3212707

Wang, L. (2005). *Support Vector Machines: Theory and Applications*. (L. Wang, Ed.)*Studies in Fuzziness and Soft Computing* (First., Vol. 177). Springer.

WMO. (2012). *Standardized Precipitation Index User Guide*. Geneva: World Meteorological Organization.

WMO, and GWP. (2016). *Handbook of Drought Indicators and Indices*. Geneva: World Meteorological Organization and Global Water Partnership.

World Bank. (2012). *Agricultural Risk Management in the Caribbean: Lessons and Experiences*.

Washington, DC.

World Bank. (2013). Agriculture in the Dominican Republic: highly vulnerable, mostly uninsured. Retrieved April 1, 2020, from https://www.worldbank.org/en/news/feature/2013/04/26/Agricultura-Republica-Dominicana-desastres-naturales

World Bank. (2019). Data: World Bank Country and Lending Groups. Retrieved April 1, 2020, from https://datahelpdesk.worldbank.org/knowledgebase/articles/906519-world-bank-country-and-lending-groups

World data center for Soils. (2010). ISRIC World Soil Information. Retrieved May 6, 2021, from https://data.isric.org/geonetwork/srv/eng/catalog.search#/home

Yu, Q., You, L., Wood-Sichra, U., Ru, Y., Joglekar, A. K. B., Fritz, S., … Yang, P. (2020). A cultivated planet in 2010 – Part 2: The global gridded agricultural-production maps. *Earth System Science Data*, *12*(4), 3545–3572. doi:10.5194/essd-12-3545-2020