SEBASTIEN

M9 – Scientific report on final indicators

Milestone Lead	UNITUS
Milestone due date	2023/10/31
Status	FINAL
Version	V1.0
Project	SEBASTIEN

Co-financed by the Connecting Europe Facility of the European Union

Agreement: INEA/CEF/ICT/A2020/2373580 **Action**: 2020-IT-IA-0234

DOCUMENT INFORMATION

Title	Milestone 9 – Scientific report on final indicators
Agreement	INEA/CEF/ICT/A2020/2373580
Action	2020-IT-IA-0234
Creator	Daniele Pietrucci (UNITUS)
Milestone Description	Brief reports on final indicators and indices
Means of verification	Brief report containing the catalogue of final indicators/indices is shared with the Agency
Contributors	Giovanni Chillemi, Marco Milanesi, Giovanni Vignali (UNITUS), Alessandro D'Anca, Alfredo Reder (CMCC), Paolo Ajmone Marsan, Mario Barbato, Elisa Somenzi (UCSC)
Requested deadline	M22
Reviewer	Sergio Noce (CMCC) Jim Yates (Nature4)

SEBASTIEN

Summary

1. Executive summary

The purpose of this milestone is to provide a brief report on the key indices and indicators that have been defined during the development of the four services generated within the project Sebastien. Through the integrated analysis of several data sources, for each service, indices and indicators have been developed to be efficiently utilized by stakeholders. The document is structured to provide: 1) a definition and description of the indicators/indices and then 2) a description of the indices/indicators for each Service.

2. Introduction

Within the European project Smarter livEstock Breeding through Advanced Services Tailoring Innovative and multi-sourcE data to users' Needs (SEBASTIEN), four Services have been developed to monitor, evaluate and detect the effect of climate/environmental changes on the livestock systems. These Services have allowed the identification of a set of qualitative and quantitative indices and indicators that can be used by stakeholders to predict and mitigate climate change effects on productions and animal health. The identification of the indices required a multifaceted approach, starting with a thorough scan of the scientific literature, followed by the application of artificial intelligence techniques to integrate heterogeneous data types and sources (e.g., climatic and spatial data). The detailed description of the methodologies used for the development of the models has been targeted by the Milestone 4 "Report on data synthesis".

This document reports the indices and indicators developed for each Service and provides information on the temporal resolution that such indices can give to stakeholders. Predictions can be developed both in the short and long term. After providing the brief list of indices and indicators, we also include a description of each with detail on the associated service.

3. List and definition of indices and indicators

An *index* or *indicator* refers to any parameter that may be of interest to stakeholders to counter the effects of climate change in the livestock sector. An indicator is a parameter that provides information on a specific (environmental or health) phenomenon. The information is quantified and can be easier to interpret, and easier to communicate to others [1]. An index, on the other hand, is calculated by considering a series of indicators, in order to provide a multidimensional, albeit simplified, view on a system [2]. The main indicators/indices described in this document are reported in Table 1.

Indices and indicators can be divided according to the temporal aspect, i.e. whether they can provide information in the long or short term.

In the case of the long term, indices and indicators are predicted using long-term climate projection data (VHR-PRO - D2.2, WC 11) as input [3]. In this case, the indices provide a ’static’ prediction, with indication referring to thirty-years in the future: from 2020 to 2050. The indicators/indices are compared with those of the reference thirty-year period (1989-2019) to detect anomalies and provide information about the long-term effects on farm activities that can be useful for the stakeholders. This prediction is calculated only once, and is therefore defined as ’static’.

Conversely, ‘dynamic’ predictions are meant to provide short-term information. Dynamic indices and indicators are predicted using short term weather forecast data as input, such as the two-day predictions of COSMO-2I [4]. These indices are constantly recalculated on a daily basis and provide stakeholders with precise information at the time of querying the services. Since these values are calculated frequently (daily, for most of the indicators/indices), the data are defined as ‘dynamic’.

The main characteristics of the indices and indicators are reported in **Table 1**. Several indices, albeit provided as static information for the long term effects, are also recalculated with high frequency to generate valuable predictions on the short term, hence acting both as static and dynamic. Finally, indices and indicators can be divided according to the type of sector to which they can provide indications, namely: agriculture or animal health. A brief list of indices and indicators is provided in **Table 1**.

Table 1. List of indicators and indices obtained by the Sebastien services. For each index, its name is provided, whether it is qualitative or quantitative, and whether it can be used statically or dynamically. Finally, the service that led to the generation of the index is indicated.

Indices/Indicators/ Datasets	**Dynamic/Static**	**Sector**	**Service**
Percentage variation of milk yield	Dynamic / Static	Agriculture	1
Percentage variation of protein content	Dynamic / Static	Agriculture	1
Percentage variation of fat content	Dynamic / Static	Agriculture	1
THI (Temperature-Humidity index)	Dynamic	Agriculture / Health	2
Pasture data	Dynamic	Agriculture	3
Probability of developing blue-tongue disease	Static	Health	4a
Somatic cell count (SCC) variation	Dynamic / Static	Health	4b
Concentration of gasses and particles in the air collected by sensors (environmental information)	Dynamic	Health	not related to services
IoT Animal Sensor data	Dynamic	Health	not related to services

4 Description of indices and indicators

4.1 Percentage variation of milk yield, protein and fat content - Service 1

This Service aims to study how climate change will affect the production capacity of farms, by evaluating its effects on milk production and fat and protein content. The services will provide indices that may help to counteract losses and drops in milk quality. A Machine Learning (ML) model was created using a dataset of milk production regarding the 'Pezzata Rossa Italiana' from 1990 to 2020 from the Friuli-Venezia Giulia region as input. This dataset was made available by ANAPRI

("Pezzata Rossa Italiana" Breeders' Association) and the LEO project (AIA, https://www.leo-italy.eu/). Milk yield is expressed in kilograms (kg), whereas fat and protein content are reported as percentage (%w/w).

Production data are functional production controls that are performed approximately every month. An analysis of the production data is performed to obtain a linear mixed model (LMM). This model allows us to calculate the residues, a value that represents the estimation error plus the environmental effect. Consequently, it provides an estimation of the environmental effect on milk yield and production.

The production residues were associated with climatic variables provided by CMCC (Fondazione Centro Euro-Mediterraneo sui Cambiamenti Climatici). The two types of data used in our experiment consist of a dataset of historical climate data, namely the VHR_REA dataset, which has a resolution of 2.2 km x 2.2 km and was obtained through a scaling of the ERA5 dataset [3]. This dataset collects information from 1981 to 2020. Residues calculated in this way are used to create a ML model. This model allows the prediction of the amount of milk produced, protein and fat content using climate data as input.

The main outcomes of this Service are three indices that represent the variation of milk yield, fat content, and protein content in the future, both for long (static) and short (dynamic) predictions. The result of the prediction is the residual of the expected value. The value assumed by the residual indicates the effect that the climate has on animal production and product quality. For example, negative values indicate a reduction in the quantity of milk produced or in its protein and fat content. Conversely, positive values indicate an increase in milk production and quality.

4.2 THI - Service 2

The Service 2 has the objective to evaluate the near future Temperature-Humidity index inside the stables. THI is a bioclimatic index that evaluates the stress in livestock combining together the effect of temperature and relative humidity. The temporal range of this prediction is the following two days with hourly resolution. Providing the system with information about the geographic positioning (latitude, longitude, altitude) of a stable on the Italian territory, the service is able to compute the THI index for the following two days. THI is a bioclimatic index that evaluates the stress in livestock combining together the effect of temperature and relative humidity. The formula used in SEBASTIEN is:

$$THI = (1.8 \times T + 32) - (0.55 - 0.55 \times RH) \times [(1.8 \times T + 32) - 58]$$

where T is the temperature (°C) and RH is the relative humidity. THI is expressed in °F degrees.

In order to evaluate the THI inside the stables for the following 2 days (at hourly resolution) a ML approach has been chosen starting from the ground truth provided by observed measurements. To train the neural network sensors data coming from 677 different stables spread across Italy has been used: internal temperature, relative humidity and internal THI are the variables measured.

Moreover, external THI is also used, computed using hourly ERA5-Land reanalysis datasets which provide temperature and relative humidity. Other data used for the training of the network are the latitude and the altitude of the stable.

The trained ML model has learnt the relationship between the external THI, the latitude, and the altitude of the stable and the internal THI index. During the inference phase the user provides the position on the map (so latitude, longitude, while altitude can be automatically computed). Starting from the location (latitude and longitude) the forecasts provided by the COSMO-2I model are used to extract the external THI for the next 48 hours. The ML model is so able to compute the related internal THI at hourly resolution for the same time period.

The Service 2b has the objective to assess the variations of THI inside a stable related to near- and long-time horizons under IPCC-RCP4.5 and RCP8.5 scenarios. A specific thirty-years baseline (1981-2010) has been used to compare the future projections with historical values.

The prediction of the internal THI follows the same approach used for the previous case. By providing geographic variables (latitude, longitude) the system is able to extract the corresponding external THI from the future climate forecasts in the aforementioned climate scenarios and compute the internal THI in terms of possible anomalies compared with the identified baseline.

4.3 Pasture data - Service 3

The service 3 aims to predict the quantity, expressed in kg, of fresh and dry matter biomass in an user defined area (i.e. pasture).
To obtain these predictions, models were developed using on field data (quantity of biomass and its characteristics) as targets. The satellite data (Sentinel2), as single band or derived indexes (eg. NDVI), in combination with climatic and topography data, were used in the linear regression models. The model returns the total quantity of fresh and dry matter biomass in the selected area. In the Service, the user can visualize the number of animals or the number of days to use the area. To do this, the estimated total quantity will be divided by the expected fresh and dry matter biomass consumption for young beef cattle. The user could modify these parameters in order to obtain more reliable results. This calculation leads to an index of the service, which can be obtained dynamically, providing short-term predictions. These provide useful information supporting breeders in real-time herding their animals.
The field data (quantity of biomass and its characteristics) are also made available as dynamical indicators.

4.4 Probability of developing the blue-tongue - Service 4.a

The Service 4.a predicts the risk probability of blue-tongue infection, transmitted by the *Culicoides midges* vector, for goats in the Sardinia region. A model was developed taking farm management data as input. The data on farm management enabled us to determine the day on which a clinical case was confirmed. As a result, we collected climatic information from the days preceding the clinical case to investigate which weather conditions may have influenced the survival of the vector and, therefore, the development of the infection. These data also allowed us to understand if the company had vaccinated animals or not. Climatic information was obtained on the days preceding each infection case and spatial information on the territory surrounding the farm were also collected. The integration of climatic (numeric) and spatial (categorical) data allows the calculation of the probability that blue tongue will develop in a particular area at a particular time of year using a 'Gradient Boosting Machine' algorithm. The output is a quantitative index, ranging from 0 to 100, where 0 indicates "no animals in the stable develop the disease" while 100 indicates "100% of the animals present in the stable develop the disease". This index provides information on animal health and can be used to make static (long-term) and dynamic (short-term) predictions.

4.5 Somatic Cell Count variation - Service 4.b

The Service 4.b aims to study the somatic cells (somatic cell counts - SCCs) which are normally secreted during milking. The SCCs are used to estimate mammary health and milk quality of dairy animals worldwide. Different factors cause the presence of SCCs in milk (e.g., the animal's health, lactation stage, breed). Other factors associated with the number of SCC are the environmental conditions and stress conditions. The SCCs are also a general indicator of the hygienic conditions of milk production on farms [5]. The SCC is a marker to evaluate the prevalence of mastitis in dairy farms [5]. Due to these reasons, this Service is strictly related in the field of animal health and can be used as a proxy for the animal health status.

The objective in this Service is to evaluate how the environment affects the presence of the SCC in the milk (e.g heat stress). The ML pipeline used in the Service 4.1 has been used to evaluate the residue of the SCC. This will be used to predict a short term (dynamic) effect and a long term (static) effect, as in the case of service 1.a. The output is a number that represents the probability of developing an increasing trend in SCCs. The number represents the residual of the expected effect calculated with the ML, and it is the index of the service and can take negative and positive values. If the values are positive, an increase in SCCs occurs, with possible consequences for the animal's health. Conversely, negative values indicate a reduction in the number of somatic cells in the sample. This value can be calculated using static and dynamic data.

4.6 Sensors data

Sensors collect two types of real-time data:

- environment sensors These sensors are installed in the barns and collect information about the gas concentration in the air. The gasses analyzed are: CO_2, H_2S, NH_3 and CH_4 (measured in part-per millions). Furthermore, the quality of the air will be evaluated measuring the concentration of particulate matter, such as PM_1, $PM_{2.5}$ and PM_{10} (measured in µg/m^3). Moreover, temperature and relative humidity are collected and used to compute the THI.

The THI can be used by stakeholders to monitor the health and stress levels of the animals. This dataset can be categorized as "dynamic" as it is collected in real time by the sensor and can provide stakeholders with short-term information.

- animal sensors. These have been implemented as collars. In this way, a series of indices and indicators related to animal health can be calculated and provided to the breeder. Examples of indicators in this case are movements, inside stable temperature and relative humidity, GNSS position, and heart rate. In this case also, temperature and humidity are used to estimate the THI index, allowing to evaluate the animal' stress due to climatic conditions. Biometric measurements, such as heart rate and movement, can be used to monitor the health and well-being of animals that may be due to climate events. For example, changes in animal movements can provide information about the physiological state of animals (i.e., the estrus period in females).

This information can be collected individually for each animal in real time. For this reason, they can be considered dynamic data.

5 Bibliography

1. Hammond AL, & World Resources Institute. Environmental indicators: a systematic approach to measuring and reporting on environmental policy performance in the context of sustainable development. Washington, DC: World Resources Institute, 1995. Vol. 36.
2. Kutay N, & Tektüfekçi, F. A New Era for Sustainable Development: A Comparison for Sustainability Indices. Financ. Audit. Stud, 2016; 2(*2)*:70–95
3. Raffa M, Reder A, Marras GF, Mancini M, Scipione G, Santini M, et al. VHR-REA_IT Dataset: Very High Resolution Dynamical Downscaling of ERA5 Reanalysis over Italy by COSMO-CLM. Data. 2021; 6(8):88. https://doi.org/10.3390/data6080088.
4. Schaettler U., Doms G., Schraff C., A description of the nonhydrostatic regional COSMO-model part VII: user's guide, Consortium for Small-Scale Modelling 2021, https://www.cosmo-model.org/content/model/cosmo/coreDocumentation/cosmo_userguide_6.00.pdf
5. Alhussien, M. N., & Dang, A. K. Milk somatic cells, factors influencing their release, future prospects, and practical utility in dairy animals: An overview. Vet World. 2018; 11(5):562. 10.14202/vetworld.2018.562-577