Project Acronym:	**LIFE GAIA Sense**
Grant Agreement number:	**LIFE17 ENV/GR/000220**
Project Title:	**LIFE GAIA Sense: Innovative Smart Farming services supporting Circular Economy in Agriculture**

DELIVERABLE

Data Report

Type of Document	Deliverable
Contractual date of delivery	12/2021
Deliverable Leader	NP
Status – version, date	Summary – v.1.0.S , 30/12/2021
Action	B2

Project co-funded by the European Commission within the LIFE 2014-2020 programme		
Dissemination Level		
P	**Public**	
C	**Confidential, only for members of the consortium and the Commission Services**	**X**

As this report is confidential, the uploaded document does not contain all the information/ content and all the chapters that were included at the "original" report.

This deliverable was produced under the co-finance of the European financial instrument for the Environment (LIFE) programme during the implementation of the project "LIFE GAIA Sense" (LIFE17 ENV/GR/000220).

Executive Summary

The aim of the deliverable is to present the overall data lifecycle management approach designed and implemented in the scope of the LIFE GAIA Sense project. As it is analysed in this document the core sources of data are: The installed network of telemetric stations (fixed IoT sensors), remote sensing platforms providing satellite data (Sentinel 1&2, Landsat), manually imported data from farmers and advisors referring to applied cultivation practices, number/type of insects captured in traps, and farm properties. Data collections are periodically transmitted to a cloud computing data repository where data are homogenized, stored and processed. The overall objective is to support a holistic view of the conditions that are taking place at the field level and to generate Smart Farming advice with regards to applied cultivation practices. Overall the implemented data collection process proved to be sound and robust mitigating successfully the various disruptions that can occur given also the harsh conditions at rural areas.

Role	Name (Organisation)
Deliverable Leader:	N. Kalatzis (NP)
Contributors:	N. Kalatzis, V.Protonotarios, D. Tsesmelis (NP)
Reviewers:	Efstratios Kefalas (GAIA)
Approved by:	N. Kalatzis (NP)

Version	Date	Contributor(s)	Description
0.1	1/06/2021	N. Kalatzis (NP)	Initial version
0.2-0.3	15/09/2021	N. Kalatzis, V.Protonotarios, D. Tsesmelis (NP)	Draft version for per review
0.4	1/12/2021	Efstratios Kefalas (GAIA)	Reviewed version
1.0	31/12/2021	N. Kalatzis (NP)	Final version
1.0.S.	31/12/2021	V. Pyrgiotis (NP)	Summary

Table of Contents

List of Figures

Definitions, Acronyms and Abbreviations

Acronym	Title
CE	Circular Economy
GAIA	GAIA EPICHEIREIN ANONYMI ETAIREIA PSIFIAKON YPIRESION
NP	NEUROPUBLIC AE PLIROFORIKIS & EPIKOINONION
SF	Smart Farming
EO	Earth Observation

1. Introduction

1.1. Project Summary

The main objective of the LIFE GAIA Sense project is to demonstrate gaiasense, an innovative "Smart Farming" (SF) solution that aims at reducing the consumption of natural resources, as a way to protect the environment and support Circular Economy (CE) models.

More specifically, this project will launch 18 demonstrators across Greece, Spain and Portugal covering 9 crops (olives, peaches, cotton, pistachio, potatoes, table tomatoes, industrial tomatoes, grapes, kiwi, walnut) in various terrain and microclimatic conditions. They will demonstrate an innovative method, based on high-end technology, which is suitable for being replicated and will be accessible and affordable to Farmers either as individuals or collectively through Agricultural Cooperatives.

Moreover, LIFE GAIA Sense aims to promote resource efficiency practices in SMEs of the agricultural sector and eventually, contribute to the implementation of the Roadmap to a Resource Efficient Europe. This project will demonstrate a method on how the farmer will be able to decide either to use or avoid inputs (irrigation, fertilizers, pesticides etc.) in a most efficient way, without risking the annual production. The focus is on the resource consumption reduction side of CE, and the results will be both qualitatively and quantitatively, considering the resources' efficiency in agricultural sector.

1.2. Document Scope

This document reports the main data sources and the nature of the data that are collected towards the realization of the gaiasense Smart Farming solution. The deliverable initially presents the core design principles dictating the lifecycle of data that are captured, processed, stored and modeled and the respective data products and knowledge that are generated. As it will be analyzed the data are collected from the installed network of telemetric stations (fixed IoT sensors) and traps, from remote sensing platforms (satellites), as well as from data sources with information about input applications in the field and other parameters that characterize the selected farms.

2. Data lifecycle management

Among the core objectives of the LIFE GAIA Sense project is to establish the appropriate technological means that will collect and analyse data from various sources, in order to generate Smart Farming advices. This approach is demonstrated through the deployed pilots covering a large number of cultivations and geographical areas of Greece, but also indicative pilot fields in Portugal and Spain.

Data capturing, processing, storing, modeling and sharing is among the core mechanisms towards the realization of the LIFE GAIA Sense objectives. To this end, a review on best practices on data management has been realized aiming to adopt best practices on this domain and avoiding to reinvent the wheel. In general, data lifecycle models provide a high-level framework for representing all stages of data throughout their life, from the creation to the consumption/reuse/destruction stages. As there is no widely accepted framework regarding the data lifecycle, various frameworks and models have been deployed and adapted according to specific needs of business and academic communities (Sinaeepourfard et al., 2015). According to a recent review of the state of the art concerning Data Lifecycle Management models for the ICT supported Agricultural domain (Demestichas et al., 2020), one of the dominant approaches to data lifecycle modeling has been introduced by the United States Geological Survey (USGS) in (Faundeen et al, 2013.). The USGS identified the following Data Lifecycle Management stages:

1. Plan: This stage refers to the evaluation and documentation of all the other elements of the data lifecycle model.
2. Acquire: This stage refers to the collection, generation, and evaluation for re-use of data.
3. Process: This stage includes activities for preparing datasets (e.g., through extraction, transformation, and load operations) for their subsequent analysis and/or integration.
4. Analyze: This stage includes the processes of exploring, interpreting, and transforming date to extract knowledge and/or new data.
5. Preserve: This stage includes actions for storing data and ensuring that they can be accessed and used in the future.
6. Publish/Share: This stage involves activities for distribution and sharing of information.

The following three cross-cutting elements are also introduced:

1. Describe: This element stresses the importance of metadata based on standards as well as use of data.
2. Manage Quality: The second cross-cutting element refers to quality assurance (QA) and quality control (QC) measures for all lifecycle stages.
3. Backup and Secure: The third cross-cutting element underlines the importance of preventing physical data losses and promotes secure data management methods.

The aforementioned stages on data management have be chosen, among other parameters, both from an administrative and technical requirements specification point of view but also for the needs of the LIFE GAIA Sense project.

3. Data collection sources

This section elaborates on the data sources that are utilized for the need of the "gaiasense" smart farming solution. As it will be analysed "gaiasense" collects data from various sources in order to monitor in a full extend the conditions at farm level. The sources of information can be categorized within the following groups: "Telemetric stations", "Satellite data", "Traps", "Farmer provided data"

3.1. Background information of targeted field

One of the most challenging objectives when the gaiasense Smart Farming solution is about to be deployed at a new field, is to develop and calibrate models supporting the generation of a Smart Farming advice on pest management, irrigation, and fertilization to the context and needs of the parcels' soil–climate zone. To this end, the first action is to collect historic information related to the cultivar adjusted at the region, the cultivation conditions, and common practices and also the weather conditions existed at the candidate fields. Moreover, information concerning regular infestations from pest and diseases, the adjacent cultivations that could have hosted potential pathogens and pest enemies, etc., are completing the historic information of the pilot fields. The information is collected with the use of questionnaires that are completed by the administrative entity of each pilot site or through accessing open data repositories. The next step is the determination of soil and climatic zones and the selection of representative fields where the telemetric stations will be installed. At this stage, the use of Earth Observation data from satellite missions are considered as a valuable tool. The overall objective is that data monitored from the installed agro-environmental telemetric stations is to be representative for the entire soil and climatic zone, aiming to avoid the installation of a large number of sensing equipment.

3.2. Telemetric Stations

The telemetric stations also called "gaiatrons" is an IoT "Deploy-and-Forget" platform incorporating a wide variety of sensors intended for the continuous surveillance of cultivation environment variables in the selected agricultural areas. It is designed by NEUROPUBLIC (https://www.neuropublic.gr), manufactured by qualified contractors and installed by specially trained company personnel. The gaiatron is designed to be installed next or inside cultivated fields and left there for years with minimal or, if possible, no maintenance whatsoever. Main communication of the deployed gaiatrons with the NEUROPUBLIC servers, in a Cloud environment, is achieved through the utilisation of either one of the following protocols: GPRS/3G, UHF. Both protocols are integrated in all gaiatrons data types with the implementation of equivalent modules. The typical gaiatrons' s data types monitored by sensors are described in table 1. It should be noted that these communication protocols (GPRS/3G) have been selected over more modern approaches (e.g., NB-IoT, Sigfox, LoRa) because of their service availability by telecom operators and coverage capabilities over the southeast Europe region. Throughout the South Europe, high geographical variability is demonstrated where mountainous and lowland fields usually coexist in the same area, reducing the robustness and performance of other telecommunication protocols.

At a first stage collected measurements from the sensors of the gaiatron are locally stored within the data logger that is included in the IoT station. Each gaiatron communicates with the back-end cloud data repository and sends the collected data at predefined time intervals (e.g. every 1 hour). In order not to lose any recordings when connectivity of the gaiatron with the back-end cloud data repository fails for any reason, the data can remain locally stored and be retransmitted in bulk when connection is restored. Data sanitization mechanism operating at the cloud, ensure that missing data are correctly integrated on the main repository.

In case of a major failure of a specific IoT station additional mechanisms are activated in order to perform data collection from nearby stations. Triangulation methods are utilized so as to automatically identify and select the most appropriate IoT station and integrate the respective recordings.

3.3. Satellite

The gaiasense Smart Farming solution incorporates Earth Observation data derived from satellites. The main source of satellite data are openly provided (with no cost) from EU's Copernicus Sentinel I & II missions but also NASA's Landsat. However the provided data are raw hence there is a need to be processed in order to be fused within the gaiasense data repository and extract useful outcomes. For the processing and preparation of satellite imagery, a fully automated image processing chain developed by NP is utilised, which includes all necessary processing steps from sensor-raw image up to web-delivered map-ready products via standardised intermediate processing levels. All available satellite images are automatically downloaded and stored on data infrastructure at NP. The pre-processing steps include the latest algorithms for cloud masking, atmospheric correction, sensor cross-calibration and the generation of spectral indices and value-added products. Furthermore, the developed approach fully supports integration of Copernicus core services to leverage information from additional sources. Satellite imagery available for a particular region of interest and time frame from Copernicus and Landsat missions are being downloaded and processed within 24 hours from acquisition resulting in a pool of new, unprocessed products. Then the collected images are feeding a processing pipeline in order to produce the desired outcome. As a result, the initial data are upgraded to higher-level products, such as NDVI (Normalized Difference Vegetation Index) or LAI (Leaf Area Index) that allowing farmers and advisors to understand their crops growth and vitality. The platform automatically assigns the extracted information to the agricultural parcels or the management units belonging to the database of gaiasense. Figure 1 illustrates the processing steps for satellite images.

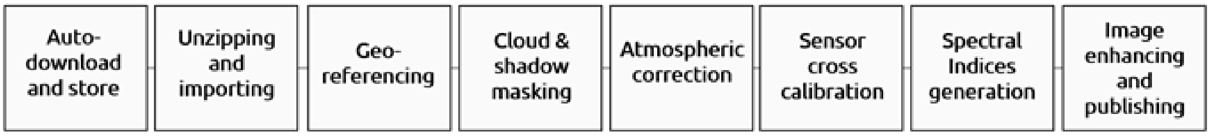

Figure 1. Typical processing of satellite images.

For reasons of complicity a short description of the main sources of satellite images is presented hereafter:

Sentinel-2 is a constellation with two twin satellites, Sentinel-2A and Sentinel-2B. They acquire optical imagery at a spatial resolution of 10, 20 and 60 m. The revisit frequency of the combined constellation is 5 days (2-3 days at mid-latitudes). Sentinel-2 satellites give new perspectives of land and vegetation. Their multispectral imager (13 spectral bands) provides high-quality image layers that can be used to derive various spectral indices and ratios, useful for instance for crop monitoring. More specifically, the spectral bands in red-edge and infrared regions of light spectra are used to calculate various spectral indices (SI) identifying the signal of certain vegetation, crop and soil characteristics. However in practice there may be clouds in the sky making the images non exploitable. The cloud issue depends on geographic location; in Europe, clouds are generally more frequent in northern and central EU countries or in mountainous areas. More detailed information about S2 may be found on: https://sentinel.esa.int/documents/247904/685211/Sentinel-2_User_Handbook

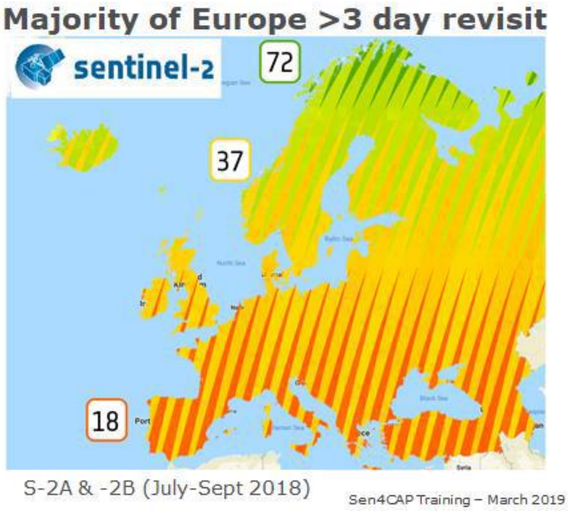

Figure 2. Coverage of Europe by Sentinel 2

Sentinel-1 The Sentinel-1 radar mission comprises a constellation of two polar-orbiting satellites (Sentinel-1A and Sentinel-1B,) which share the same orbital plane, operating day and night and acquiring imagery regardless of the weather. Sentinel-1 is performing Synthetic Aperture Radar (SAR) imaging, i.e. the sensor aperture is managed by digital means. SAR is a type of active data collection which means that the sensor emits its own energy, in the form of a signal (in a given band - interval of frequencies) and then records the amount of that energy reflected back (reflected backscatter) after interacting with the Earth. SAR data require a different way of thinking compared to the optical data because the signal is instead responsive to surface characteristics like structure and moisture. Radar images contain information divided into the phase and the amplitude of the wave. Radar can penetrate clouds and because of this, these data provide an advantage over optical imagery. Radar signatures from the VV, VH and VH/VV backscatter ratio (in decibels) from Sentinel-1 can be used for vegetation monitoring as polarisation temporal profiles can be created for inspecting the vegetation signals. More info on Sentinel-1 is available here: https://sentinel.esa.int/web/sentinel/missions/sentinel-1/overview?inheritRedirect=true

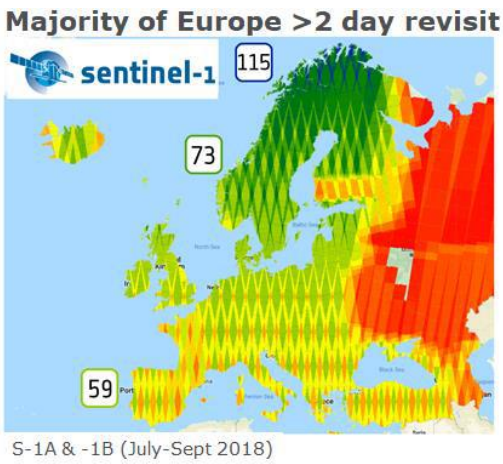

Figure 3. Coverage of Europe by Sentinel 1

NASAs Landsat-8 satellite sensor is an American Earth observation satellite developed by NASA and the U.S. Geological Survey (USGS). It has two main sensors: the Operational Land Imager (OLI) and the Thermal Infrared Sensor (TIRS) acquiring images in various wavelengths consisting of nine spectral bands with a spatial resolution of 30m, one panchromatic band at 15m and two thermal bands collected at 100m. Its revisit cycle (temporal resolution) is 16 days and the data is freely available. Landsat data is considered of high quality and very stable products as the Landsat satellite program is the longest continuous Earth imaging program in history. Since 1972, Landsat satellites have collected huge amounts of consistent spectral imagery (Hemati et al., 2021). Landsat 8 has a coarser spatial resolution (and so, largest tiles) than Sentinel-2 and that makes the images less suitable for EO monitoring. However, Landsat data is often used as supplementary data to Sentinel-2 images for constructing denser satellite time-series (combining Landsat and Sentinel the revisit time can drop below 3 days) or to mitigate the cloud issue, which is also the case for the gaiasense system.

For applying crop monitoring to 18 pilot sites in Greece, Spain and Portugal, we have used 1500 cloud free Copernicus Sentinel-2 images for the growing season of 2020 & 2021.

3.4. Traps

Traps are an essential source of information for developing scientific models for pests. Depending on the crop type and the type of enemies different types of traps are used and different types of recordings are made. More information about the detailed study on the type and location of the traps for each use case can be found in the deliverable “Traps placement study, Action B2”.

The use of traps facilitates the monitoring of insect/pests or enemies population in different crops. In some cases, traps are also used as an environmental friendly alternative to pesticides when it comes to pest management. Their main use though is to examine the distribution of enemies’ occurrences during the cultivation period. Data related with insects populations are manually entered with gaiasense system through the use of “Intelligent Management Crop – (iCM)” and/or “Field Collect”. Both of these services are presented in “Final Smart Farming Application” (section 2.2. Integration with existing services).

The traps are manually monitored every 3 days in the beginning of the cultivation while in the middle of the cultivation period they are monitored every 5 days.

3.5. Farmers data

The gaiasense smart farming solution considers the recording of cultivation activities performed by the producers of high importance. The gaiasense offers the proper ICT tools to record all information that is related to the daily cultivation work of the producer such as fertilization application, plant protection, time and duration of irrigation. This information provides the full and detailed picture of the exploitation, which contributes significantly to the decision-making process. Gaiasense maintains a digital farmer’s calendar which contains the respective recording of the actions that the farmer/advisor performs at the field. This information provides the full and detailed picture of the cultivation, which contributes significantly to the decision-making process. Provision of this information is realized through various means (mobile applications, web services, phone calls with the farmer). The central information system to view and manage farmers calendar is the “Intelligent Management Crop – (iCM)” figure 4. It is expected that data entries are performed every 2-4 days during the time periods that the cultivation is active (e.g. Spring to Autumn).

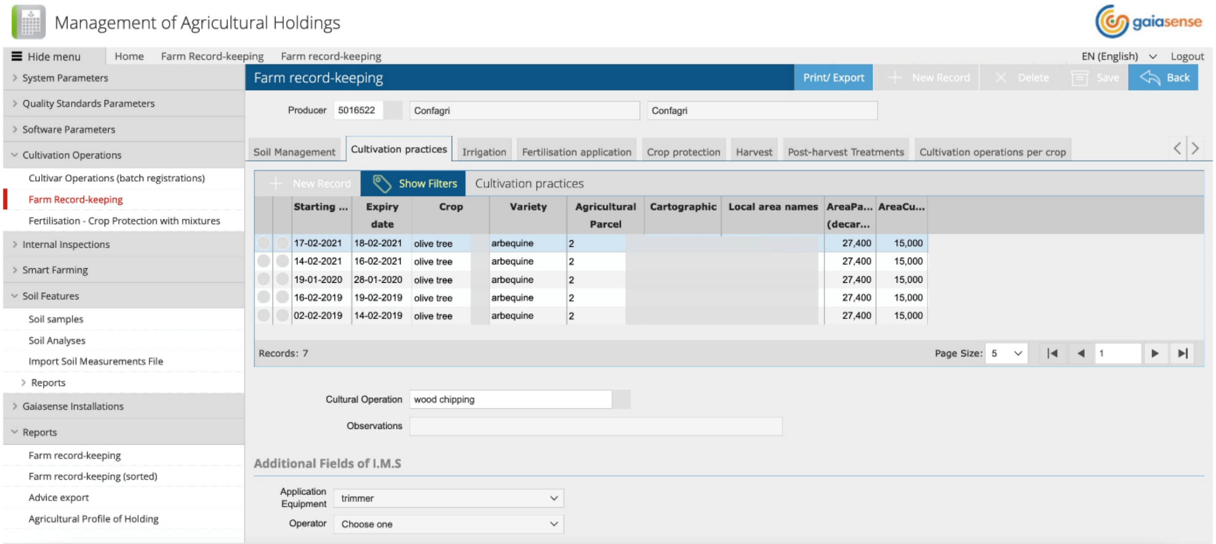

Figure 4. View of the digital Farm Record-Keeping page.

4. Data storage, processing & visualization

All the aforementioned data are collected to a central cloud computing repository maintained by NEUROPUBLIC. Collected data are combined with data from other sources (i.e. external weather forecast services, Earth Observation sources) and converted into facts using advanced data analytic and scientific knowledge. (Figure 5).

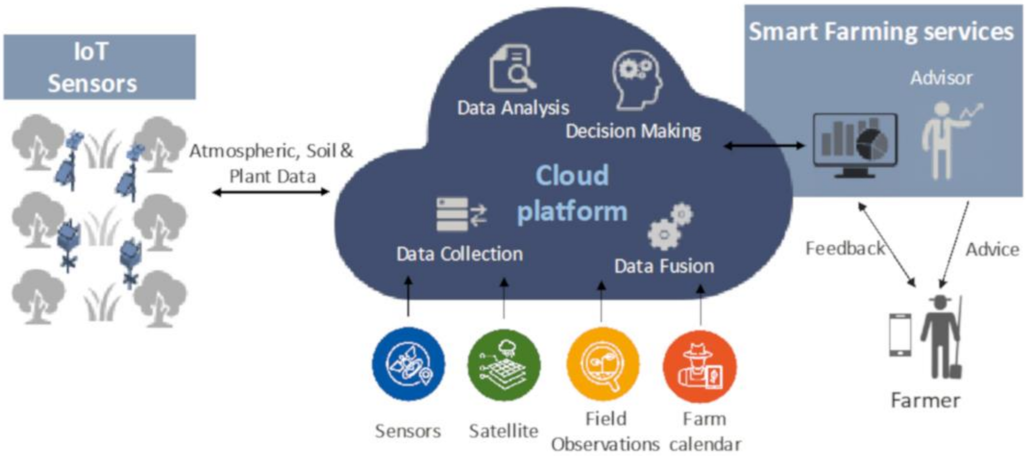

Figure 5. Information flow for the Smart Farming as a Service approach.

The outcomes of the processing are analyzed by experts (e.g.agronomists and data experts) in order to generate Smart Farming advices towards the optimization of irrigation, pest management, and fertilization, tailored to the context of the specific parcel. The advice along with selected agro-environmental measurements are then mediated to the farmers.

It should be noted that the role of the advisor/expert remains significant once the generated Smart Farming advice is issued since it needs to be explained to the farmer and the respective farmer's cultivation practices need to be supported and monitored. The necessary feedback, related to the actual farming practice that was applied as a response to the advice, is then registered to the gaiasense system (e.g., through the farmer's digital calendar) in order to be further analyzed and incorporated, supporting the generation of any future advices. (Adamides et al, 2020)

As it was stated data collections are maintained in a cloud computing repository provided by NP. In general, the cloud-computing based solution allows for greater scalability, immediate availability and security policies to prevent data access and transferring risks (e.g.: firewall, password protection, encryption, database backup mechanisms). Within this context, NP is responsible for implementing all necessary security and information management features.

It should be noted that farmers' personal data (e.g. name, gender, age, location) maintained by the gaiasense smart farming system are not shared with any remote entities. NP which is the entity responsible for applying the data protection policies, including GDPR directive, is an EN ISO 9001:2015 & ISO/IEC 27001:2013 certified company (https://www.neuropublic.gr/en/certifications-iso/)

Role base access control is enforced to all stages of the data management lifecycle stages.

5. References

- Sinaeepourfard, A.; Masip-Bruin, X.; Garcia, J.; Marín-Tordera, E. A Survey on Data Lifecycle Models: Discussions toward the 6Vs Challenges; Technical report (UPC-DACRR-2015-18); Universitat Politècnica de Catalunya (UPC, BarcelonaTech): Barcelona, Spain, 2015; pp. 4–16.
- Demestichas, K.; Daskalakis, E. Data Lifecycle Management in Precision Agriculture Supported by Information and Communication Technology. Agronomy 2020, 10, 1648. https://doi.org/10.3390/agronomy10111648
- Hemati, M.; Hasanlou, M.; Mahdianpari, M.; Mohammadimanesh, F. A Systematic Review of Landsat Data for Change Detection Applications: 50 Years of Monitoring the Earth. Remote Sens. 2021, 13, 2869. https://doi.org/10.3390/rs13152869
- Adamides, G.; Kalatzis, N.; Stylianou, A.; Marianos, N.; Chatzipapadopoulos, F.; Giannakopoulou, M.; Papadavid, G.; Vassiliou, V.; Neocleous, D. Smart Farming Techniques for Climate Change Adaptation in Cyprus. Atmosphere 2020, 11, 557.