D3.6: Validation Plan

WP3 – Earth Observation data products

Authors: Andreea Bucur, Alexander Gruber

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 687412.

Disclaimer

Copyright message

Document Information

Grant Agreement Number	687412	Acronym	APOLLO	
Full Title	Advisory platform for small farms based on earth observation			
Horizon 2020 Call	EO-1-2015: Bringing EO applications to the market			
Type of Action	Innovation Action			
Start Date	1st May 2016	**Duration**	34 months	
Project URL	http://apollo-h2020.eu/			
Document URL	http://apollo-h2020.eu/project/#results			
EU Project Officer	Iulia SIMION			
Project Coordinator	Polimachi Simeonidou			
Deliverable	D3.6: Validation Plan			
Work Package	WP3 – Earth Observation data products			
Date of Delivery	**Contractual**	M12	**Actual**	M12
Nature	R – Report	**Dissemination Level**	PU - Public	
Lead Beneficiary	TUW			
Lead Author	Andreea Bucur, Alexander Gruber	**Email**	andreea.bucur@geo.tuwien.ac.at, alexander.gruber@geo.tuwien.ac.at	
	TUW	**Phone**	+43 1 58801 12283, +43 1 58801 12263	
Other authors	Stelios Kotsopoulos (DRAXIS), Dragutin Protic (UBFCE)			
Reviewer(s)	Camille Pelloquin (Starlab), Chiara Pratola (Starlab)			
Keywords	Validation, Earth Observation, Soil moisture, Vegetation indices, Biomass, Horizon 2020			

Document History

Version	Issue Date	Stage	Changes	Contributor
1.0	06/02/2017	Draft	Document created	TUW
2.0	22/02/2017	Draft	Changed outline	TUW
3.0	06/03/2017	Draft	Edited content	TUW
4.0	22/03/2017	Draft	Updated content	TUW
4.1	24/03/2017	Draft	Edited content	TUW
5.0	27/03/2017	Draft	Updated content of Chapter 5	DRAXIS
6.0	03/04/2017	Draft	Updated content of Chapter 6	UBFCE
7.0	10/04/2017	Draft	Final draft sent for internal review	TUW
8.0	19/04/2017	FINAL	Updated content after internal review	TUW

Table of Contents

This project is co-funded by the European Union

Table of Tables

Acronyms and Abbreviations

ASCAT	Advance Scatterometer
CSI	Critical Success Index
DRAXIS	DRAXIS ENVIRONMENTAL S.A.
ENVISAT	Environmental Satellite
ETS	Equitable Threat Score
EO	Earth Observation
EUMETSAT	European Organisation for the Exploitation of Meteorological Satellites
FAR	False Alarm Rate
FBIAS	Frequency Bias
GLDAS	The Global Land Data Assimilation System
H SAF	EUMETSAT SAF on Support to Operational Hydrology and Water Management
H2020	Horizon 2020 - The Framework Programme for Research and Innovation
HOAL	Hydrological Open Air Laboratory (Petzenkirchen, Austria)
ISMN	International Soil Moisture Network
LAI	Leaf Area Index
LSM	Land Surface Model
MADIS	Meteorological Assimilation Data Ingest System
MAE	Mean Absolute Error
ME	Mean Error
MET	Model Evaluation Tools
Metop	Meteorological Operational Platform
NDVI	Normalized Difference Vegetation Index
NOAH	National Centers for Environmental Prediction/Oregon State University/Air Force/Hydrologic Research Lab (Noah) Model
NSE	Nash-Sutcliffe efficiency index
POD	Probability of Detection
R	Pearson correlation coefficient
R^2	Coefficient of determination
RMSE	Root Mean Square Error
RRMSE	Relative Root Mean Square Error
S1	Sentinel-1
S1-SM	Sentinel-1 Soil Moisture

SAF	Satellite Application Facility
SM	Soil moisture
SNR	Signal-to-Noise Ratio
SSM	Surface soil moisture
TC	Triple collocation
TU Wien	Technische Universität Wien (Vienna University of Technology)
UBFCE	The Faculty of Civil Engineering of the University of Belgrade
ubRMSD	Unbiased Root-Mean-Squared-Difference
VIs	Vegetation Indices
VP	Validation Plan

Executive summary

The “Validation Plan”, which constitutes the project’s deliverable D 3.6, describes the guidelines followed for the validation of Earth Observation data products created in the APOLLO project. This task is carried out within the “Work Package 3 – Earth Observation data products” and details the reference data sets and validation methods used for the assessment of the EO data products.

The Horizon 2020-funded APOLLO project aims at providing advisory services for management of small farms. The services – tillage scheduling, irrigation scheduling, crop growth monitoring, and crop yield estimation – are based on agricultural parameters estimated from Earth Observation (EO) data and agronomic models. Making use of free and open data provided by the European Union’s Copernicus programme, the services rely on data from Sentinel-1, Sentinel-2 and Landsat 8; in addition, MODIS Land products and meteorological and auxiliary data are employed. The following data products are being developed within the Apollo project:

- Soil moisture
- Vegetation indices and Leaf Area Index (LAI)
- Biomass
- Meteorological data

These products are being validated in terms of data completeness as well as precision (random errors) and accuracy (systematic biases) with respect to the spatio-temporal target scale of the APOLLO services. For this purpose, reference data from various sources (e.g. in-situ measurements, land surface models, satellites-based products) are collected and harmonized to act as a baseline for the validation of EO products. Depending on the data structure and availability, appropriate validation methods were selected and applied.

1 INTRODUCTION

1.1 Purpose of Document

This document presents the Validation Plan (VP) for Earth Observation (EO) data within the framework of the APOLLO project funded by the European Union's Research and Innovation programme Horizon 2020 under the topic "Bringing EO applications to the market (EO-1-2015)".

The APOLLO services are based on outputs of agronomic models that are forced with EO-based products. The purpose of the VP is to describe and define the reference data sets and methods used for the validation of these EO data sets developed as ready-to-use highly automated and operational products.

1.2 Target Audience

This document is intended, mainly, for internal use and it outlines the strategies and methodologies applied for the validation of the APOLLO products. The results of the data validation will reflect the quality and reliability of the services offered by the APOLLO platform. At the later stage, this plan will serve as supplementary information for users interested in the validation results and how they were derived. Finally, this document also serves to the remote sensing experts interested in validation of their own EO data products.

1.3 Important documents

Within the APOLLO project several other documents created separately (project deliverables) contain detailed information about the production of the EO products and user requirements which are relevant for the validation purposes. Table 1 contains a list of these documents.

Document no.	Title	Date
D 2.1	Report on User Requirements	September 2016
D 3.1	EO data collection and pre-processing protocol	February 2017
D 3.2	Soil moisture data product	February 2017
D 3.3	Meteorological data product	February 2017
D 3.4	VIs and LAI products for crop condition monitoring	February 2017
D 3.5	Biomass data product	February 2017

Table 1. Project deliverables relevant for the validation of the Earth Observation (EO) data products developed in APOLLO.

2 VALIDATION STRATEGY

The APOLLO platform will provide advisory farm management services based on specific agricultural models driven by EO data. The services included are: tillage and irrigation scheduling, monitoring of the growth and health of crops, and estimation of yield amounts. The user requirements for each of these services were collected and reported in Deliverable 2.1 – Report on User Requirements. The services rely exclusively on EO-based products and therefore their reliability depends on the performance of individual data sets.

According to the Committee of Earth Observation Satellites (CEOS) Working Group on Calibration and Validation (WGCV), validation is defined as the “process of assessing, by independent means, the quality of the data products derived from the system outputs” (Loew et. al, in review). The quality of the APOLLO EO data products will be assessed qualitatively and quantitatively in terms of data completeness, precision (random errors) and accuracy (systematic biases) taking into account the spatio-temporal target scale of the APOLLO services. The validation is performed using in-situ data complemented with other independent EO-based and modeled data sets.

2.1 Data completeness

Agricultural services, especially irrigation scheduling, rely heavily on the completeness and timeliness of the EO input data sets. Typically, these parameters are specified by the data provider through the satellite revisit time and the data latency. However, they may vary depending on the geographic location and the time required for data processing. Hence, their actual values at the APOLLO pilot areas will be verified by statistical means.

2.2 Systematic errors

Systematic errors refer to an over- or underestimation of the observed parameter, which persists in time and originates from imperfections in the retrieval model structure or an inaccurate model calibration. Typically, these are characterized as additive- and multiplicative biases with respect to a more reliable reference data set such as ground reference stations. However, EO data sets typically observe geophysical processes integrated over larger areas (in case of the APOLLO services, up to several hectares) and single ground reference measurements may not be fully representative for the observed processes at these scales. Therefore, possible representativeness errors must be kept in mind when comparing data sets at different scales.

2.3 Random errors

Random errors are unpredictable errors with zero mean originating from instrument noise and unmodeled or imperfectly modeled processes. In addition, the above mentioned representativeness errors originating from a spatial scale mismatch between the reference data and the data under validation can also have a random component. The common metrics that reflect random error structures are the correlation coefficient and the unbiased Root-Mean-Square-Difference. Advanced methods such as Triple Collocation (TC) analysis also target the estimation of random errors and are – depending on the analysis setup – capable of accounting for representativeness errors (Gruber et al. 2016).

The reference data sets and statistical methods used for validation will be described in detail in the following sections. Given the distinct properties of the APOLLO EO data products and their different requirements for validation, the reference data and evaluation metrics are described separately for each EO product category (i.e., soil moisture data, vegetation data, and meteorological data).

3 VALIDATION OF SOIL MOISTURE

3.1 Validation data sets

This section describes the data sets used for the validation of the Sentinel-1 soil moisture product (S1-SM) produced within the APOLLO project.

3.1.1 Ground reference measurements

3.1.1.1 Project internal in-situ data

Dedicated in-situ soil moisture measurements used for validation are collected in the three pilot areas located in Pella (Greece), Ruma (Serbia) and La Mancha Oriental (Spain).

The in-situ measurements are planned in order to meet the minimum requirements needed to calibrate and validate the soil moisture retrieval algorithm. Thus, since the model parameters depend on the different crop types and soil properties a minimum of one soil moisture sensor should be installed in each field (one for each crop type) within the three pilot areas. Pilot areas in Spain and Serbia already have in place soil moisture sensors for some of the crops monitored in APOLLO, while for the Greek pilot area installation of sensors is envisaged. A summary of the currently available in-situ measurements is outlined in Table 2. The Validation Plan will be updated as soon as the APOLLO network of SM sensors will be improved.

Country	Crop Type	Soil moisture sensor type	Soil moisture sensing depth	Period of observation
Serbia	Wheat	Pressel Watermark sensor	30-50 cm	2014/10/15 - 2015/05/15
	Sugar beet			2015/05/15 - 2015/10/15
	Wheat			2015/10/15 - 2016/05/15
	Sugar beet			2016/05/15 - 2016/10/15
Spain	wheat	AquaCheck capacitance probe	0-60cm (every 10 cm)	2015/03/31 - 2015/07/12
	wheat			2015/01/15 - 2015/06/21
	maize			2015/06/09 - 2015/11/21
	wheat			2016/02/27 - 2016/07/05
	wheat			2016/03/08 - 2016/06/29
	maize			2016/06/29 - 2016/09/01
	almond			2016/06/07 - present
	purple garlic			2016/04/19 - 2016/07/06
	vineyard			2016/04/16 - present

Table 2. APOLLO network of in-situ soil moisture sensors.

This project is co-funded by the European Union

3.1.1.2 Project external in-situ data

In situ soil moisture data for validation is collected within the three pilot areas. However, since the number of measurement stations is relatively low, additional in-situ measurements from the International Soil Moisture Network (ISMN) will be used as a secondary source to increase the statistical robustness of the validation results; this will, however, depend on the availability and adequacy of data for the validation task of APOLLO soil moisture product. The ISMN (Dorigo et al., 2011) has been established as a centralized data-hosting facility where globally available in-situ soil moisture measurements from operational networks and validation campaigns are collected, harmonized, and made available to users. It exists as a means for the geo-scientific community to validate and improve global satellite observations. The network is coordinated by the Global Energy and Water Cycle Experiment (GEWEX), in cooperation with the Group of Earth Observation (GEO) and the Committee on Earth Observation Satellites (CEOS).

A third source for in-situ soil moisture measurements is the Hydrological Open Air Laboratory (HOAL) in Petzenkirchen, Lower Austria. The HOAL is a 66 ha research catchment operated jointly by the TU Wien and the Federal Agency for Water Management, which has been established to advance the understanding of water related flow and transport processes including in agricultural fields (e.g. 87% of the catchment area is arable land) (Blöschl et al., 2016).

The in-situ data from the ISMN and HOAL networks will be used depending on the availability of information on crop type (i.e., same crops as monitored in APOLLO) and observation period (i.e., 2015 onwards in to match the availability of S1 data).

3.1.2 EO based reference data sets

For the validation of APOLLO soil moisture products two data records from the H SAF soil moisture suite will be used:

- the H SAF H111 SSM ASCAT Data Record, and
- the H SAF H08 SM-OBS-2 – Small scale surface soil moisture product

H SAF is part of the distributed application ground segment of the "European Organization for the Exploitation of Meteorological Satellites (EUMETSAT)". The application ground segment consists of a Central Application Facilities located at EUMETSAT Headquarters, and a network of eight "Satellite Application Facilities (SAFs)". H SAF aims to satisfy the needs of operational hydrology, by generating, centralizing, archiving and disseminating several products: precipitation, soil moisture and snow parameters.

H SAF H111 SSM ASCAT Data Record Time Series

The surface soil moisture data records are derived from backscatter (σ^0) measured by the Advanced Scatterometer (ASCAT) (EUMETSAT, 2015) on-board the series of Metop satellites using the TU Wien soil moisture retrieval algorithm (Wagner et al., 1999; Naeimi, 2009). In the TU Wien algorithm, long-term backscatter measurements are used to model the incidence angle dependency of backscatter, which allows a normalization to a common reference incidence angle (θ_r = 40). The relative surface soil moisture estimates range between 0% (dry) and 100% (wet) and are derived by scaling the normalized backscatter between the lowest/highest backscatter values corresponding to the driest/wettest soil conditions. Soil moisture is represented in degree of saturation S_d, but can be translated from *relative* (%) to *absolute volumetric units* (m^3m^{-3}) using porosity information (see Equation 1).

$$\Theta = p\,\frac{S_d}{100} \tag{1}$$

where Θ is absolute soil moisture in m^3m^{-3}, p is porosity in m^3m^{-3}, S_d is degree of saturation.

Notice that the accuracy of soil porosity is as much as important as the relative soil moisture content. The soil porosity can be derived from the information on soil texture as described in Saxton and Rawls (2006), which is also required by some of the APOLLO services. The product also comes with a surface state flag (SSF) that indicates the surface conditions: unknown, unfrozen, frozen, temporary melting/water on the surface or permanent ice. The flag is intended to help filtering soil moisture values derived under frozen soil conditions. A detailed description of the TU Wien soil moisture retrieval algorithm together with a description of the derivation of the model parameters can be found in the Algorithm Theoretical Baseline Document (ATBD) (H SAFa, 2016). Additional information about the product can be found in Product User Manual (H SAFb, 2016).

The H SAF H111 is available at a latitude dependent temporal resolution of 1-3 days with a spatial extend of 60°S to 90°N. The spatial resolution of the product is 25 km over the period 01.01.2007 – 31.12.2016. The product is extended irregularly throughout the year as additional H SAF product (i.e., H SAF H112).

H SAF H08 SM-OBS-2 – Small scale surface soil moisture by radar scatterometer

The H SAF H08 is the result of combining 25 km Metop ASCAT surface soil moisture data with 1 km backscatter information derived from Envisat ASAR (ESA, 2007). The product is the result of a disaggregation process that uses a pre-computed fine-mesh layer which is stored in a parameter database. The fine-mesh layer contains information derived from SAR imagery from Envisat ASAR (operating in the ScanSAR Global monitoring mode) which includes backscatter values and scaling characteristics. The complete description of the product characteristics including algorithm, processing chain and validation can be found in the Product User Manual (PUM) (H SAFd, 2015) and in the Algorithm Theoretical Baseline Document (ATBD) (H SAFc, 2015). The product is available in a standardised WMO file format and is available in near real-time.

The H SAF H08 is available at a latitude dependent temporal resolution of 1-3 days over the area within (25°S, 75°N) and (-25°W, 45°E). The product is offered at 1 km spatial resolution over the period 14.08.2009 to present.

3.1.3 Modelled soil moisture data

The Global Land Data Assimilation System (GLDAS) was developed at the National Aeronautics and Space Administration (NASA) Goddard Space Flight Center (GSFC) and the National Oceanic and Atmospheric Administration (NOAA). GLDAS-Noah aims to assimilate ground- and satellite-based observations to generate land-surface fields (Rodell et al., 2004) driving the off-line land surface model Noah. It is forced by a combination of NOAA/GDAS atmospheric analysis fields, spatially and temporally disaggregated NOAA Climate Prediction Center Merged Analysis of Precipitation (CMAP) fields, and observation-based downward shortwave and longwave radiation fields derived using the method of the Air Force Weather Agency's Agricultural Meteorological system.

For the validation of the APOLLO soil moisture product the output variables of profile soil moisture at depths of 0-0.1, 0.1-0.5, 0.4-1.0, 1.0-2.0 m are used. The surface soil moisture estimates used for the final validation have been successfully used by several authors to evaluate remotely sensed soil moisture products (e.g. Dorigo et al., 2010, Liu et al., 2012).

For the validation of the APOLLO soil moisture product version V2.1 of the GLDAS Noah Land Surface Model (LSM) will be used.

The variables of the GLDAS-Noah LSM v2.1 are available at 3-hourly and monthly temporal resolution with a spatial extent of 60°S to 90°N. The data is available over the period 01.01.2000 to present at a 0.25 degree spatial resolution.

3.2 Validation metrics

3.2.1 Data completeness

Data completeness will be evaluated in terms of the average data density (i.e. the number of measurements per reference time span), and data timeliness (i.e., the average delay between observation time and availability of the processed measurements due to data provider latency and download and processing time.

3.2.2 Systematic errors

Systematic errors will be characterized through the zero-th and first order bias (denoted as b^0 and b^1, respectively) w.r.t. the considered reference data sets:

$$b^0 = mean(x) - mean(y) \tag{2}$$

$$b^1 = \frac{std(x)}{std(y)} \tag{3}$$

x denotes the data set under validation; y denotes the reference data set; $mean(\cdot)$ denotes the temporal mean; and $std(\cdot)$ denotes the standard deviation.

Biases will be computed between S1-SM and the available ground reference stations. In cases where multiple stations are available within a S1 footprint, these will be averaged in order to better represent the footprint average soil moisture in order to reduce systematic representativeness errors.

3.2.3 Random errors

Random errors will be characterized using the *Pearson correlation coefficient* (R), *the unbiased Root-Mean-Squared-Difference* $(ubRMSD)$, and *the triple collocation based (logarithmic) Signal-to-Noise Ratio* (SNR):

$$R = \frac{cov(x, y)}{std(x)std(y)} \tag{4}$$

$$ubRMSD = \sqrt{mean((x^* - y)^2)} \tag{5}$$

$$SNR = -10\log_{10}\left(\frac{var(x)cov(y, z)}{cov(x, y)cov(x, z)} - 1\right) \tag{6}$$

Where $cov(\cdot)$ denotes the covariance; $var(\cdot)$ denotes the variance; and the asterisk denotes that x has been corrected for its bias (in both mean and standard deviation) through:

$$x^* = \frac{x - mean(x)}{std(x)} std(y) + mean(y) \tag{7}$$

Notice that both R and $ubRMSD$ are relative metrics. That is, random errors (including random representativeness errors) in the reference data will lead to an under- and overestimation of R and $ubRMSD$ estimates, respectively. TC based SNR estimates, on the other hand, can – depending on the setup – provide consistent estimates of the random errors of a certain data set which are also robust against random representativeness errors (Gruber et al. 2016). However, TC analysis requires the simultaneous availability of three data sources with uncorrelated error structures which is often difficult to acquire.

R and $ubRMSD$ estimates will be computed between S1-SM and the ground reference stations as well as between S1-SM and coarse-resolution ASCAT and modelled soil moisture data without and with upscaling S1-SM to the spatial resolution of the respective reference data sets. The latter analysis does not contain random representativeness errors yet the upscaling may partly suppress the actual random error component of the S1-SM estimates.

TC analysis will be applied to S1-SM together with ground stations and coarse resolution model or scatterometer data. According to Gruber et al. (2016) this setup should lead to consistent SNR estimates for S1-SM which are also free of random representativeness errors.

4 VALIDATION OF CROP MONITORING PRODUCTS AND BIOMASS

4.1 Introduction

This section presents the methodology for the validation of the APOLLO crop monitoring and biomass products estimated from EO data.

The APOLLO crop monitoring products include biophysical parameters (i.e., Leaf Area Index; LAI, and Chlorophyll content; Chl), as well as Vegetation Indices (VIs). The models employed for their calculation are presented in detail in deliverables D3.4 – VIs and LAI products for crop monitoring and D3.5 – Biomass data product. These models are being developed crop specific in order to achieve better accuracy of estimation.

4.2 Methodology

In order to evaluate the quality of the biophysical parameters models in terms of accuracy a representative reference data set is needed. In-situ measurements collected in the three pilot areas will be used as reference data for validation as well as calibration of the models. However, these data are limited especially since the same reference data cannot be used for both training and testing of the models. For that reason, the so-called k-fold cross validation method will be used.

In this approach the training set is split into k smaller sets (group of samples) called “folds”, depending on the total size of the data set and its structural properties. For each of the k folds the following procedure is employed: a model is learned using k-1 of the folds as training data; the resulting model is validated on the remaining part of the data by computing accuracy metrics. It means that each time, one of the k subsets is used as the test set and the other k-1 subsets are forming a training set. The average performance metrics (validation indicators) from all k trials are computed. For the purposes of the APOLLO project, we will apply 5-fold cross validation which means that in each iteration 20% of the training set will be used for validation.

4.3 Reference data

4.3.1 Project internal in-situ data

To provide reference data for calibration/validation purposes, it is foreseen to conduct in-situ data collection within the APOLLO project. The data on LAI, Chl and biomass are collected in each of the three pilot areas, namely Pella (Greece), Ruma (Serbia) and La Mancha (Spain), for all crops that are monitored in the project.

For LAI and biomass data, a destructive method is applied. On each Elementary Sampling Unit (ESU), at least three average plants are cut within a circle of 10 m diameter and used to measure the biophysical parameters. The measurements are then upscaled to the ESU level. The measurements are taken during the entire crop growth period to ensure availability of reference data for the whole range of the parameters. Measurements on at least 2 ESU per crop will be conducted in each pilot area.

4.3.2 Project external in-situ data

To calibrate and validate models for EO products that are intended to be universally applicable in an operational way, a comprehensive set of reference data is required. Any in-situ data from sources outside of the APOLLO project that are open and appropriate for the project purposes will be collected and used.

One such data set available is the LAI in-situ data set from the ImagineS (http://fp7-imagines.eu) FP7 project which contains data collected at a number of sites across the globe. However, in APOLLO we will use only the data collected after the launch of Sentinel-2 in 2015. There are two relevant sites where data was collected during summer 2015: Las Tiesas site in Barrax, Spain and Pshenichne site, Ukraine. At both sites, hemispherical digital photography was used to measure LAI.

4.4 Validation metrics

Richter et al. (2012) proposed an "optimal" set of statistical measures comprehensively quantifying the performance of vegetation biophysical models. The following metrics (definitions from Richter et al., 2011) will be used for the validation of the vegetation products within APOLLO:

- **Root Mean Square Error (RMSE)** which denotes the global magnitude of errors in variable units, and is defined by:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(V_{est}^{i} - V_{obs}^{i}\right)^2} \quad (8)$$

- **Coefficient of determination (R^2)** which represents the proportion of the variance in a biophysical parameter that is predictable by the model. In addition, it provides information about the degree of spatial consistency, which is essential for image-based analyses. $R^2 \geq 0.9$ can be considered as excellent values, while R^2 between 0.5 and 0.9 are considered as good values. It is defined by:

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(V_{est}^{i} - \widehat{V}_{est})^2}{\sum_{i=1}^{n}(V_{est}^{i} - \overline{V}_{est})^2} \quad (9)$$

- **Relative Root Mean Square Error (RRMSE)** is the RMSE divided by the mean of the observed variables. It is less sensitive to the magnitude of values and outliers. Excellent values are RRMSE ≤ 10%, while values 10% < RRMSE ≤ 20% are considered good. RRMSE is defined by:

$$RRMSE = 100 * \frac{RMSE}{Mean(obs)} \quad (10)$$

This project is co-funded by the European Union

- **Nash-Sutcliffe efficiency index (NSE)** estimates the predictive power of models by describing if a model performs better than the mean of observed data. NSE ≤ 0.9 are excellent values, while 0.5 ≤ NSE ≤ 0.9 are considered good values. NSE is defined by:

$$NSE = 1 - \frac{\sum_{i=1}^{n}\left(V_{obs}^{i} - V_{est}^{i}\right)^2}{\sum_{i=1}^{n}\left(V_{obs}^{i} - \bar{V}_{obs}\right)^2} \qquad (11)$$

In addition, to quantify the local accuracy of the model, **residuals** from the cross validation will be computed.

This project is co-funded by the European Union

5 VALIDATION OF METEOROLOGICAL PARAMETERS

5.1 Introduction

This section describes the methodology that will be followed to verify the weather forecast fields and the gridded observational data produced by the statistical and dynamical downscaling processes during the APOLLO pilot phase. Forecast verification in general is the process and practice of determining the quality of forecasts in the spatial and temporal domain (Murphy and Winkler, 1987).

5.2 Data and Methods

For the verification of the meteorological parameters, the MET[1] (Model Evaluation Tools) software package version 5.2 will be used. MET is a collection of functions written in C that was developed by the Developmental Testbed Center in order to provide the atmospheric community with highly-configurable and state-of-the-art verification tools. These tools are based on a variety of methods including standard verification scores for gridded and point statistics as well as more advanced statistical measures such as neighborhood, object based, and intensity-scale decomposition approaches for spatial statistics.

The observational data that will be used as reference for the verification of the meteorological products of APOLLO come from the MADIS[2] system. The Meteorological Assimilation Data Ingest System (MADIS) is a meteorological database and data delivery system that provides atmospheric observations that cover the entire globe. The observations are derived from multiple official and unofficial sources including *metar* messages from surface weather stations, radiances and atmospheric profiles from satellites, airborne observations, station radiosondes and ocean meteorological parameters from ships and buoys. After collecting the observations, the MADIS system decodes the ingested observations, performs a three-stage quality control and converts all data into a common format. The MADIS data set is continuously updated with a varying frequency between 1 hour to 2 days, according to the requested atmospheric variable. The data set containing the meteorological information from surface weather stations is free to use both for commercial and non-commercial purposes, and is available for downloading upon request through a public website[3] or ftp[4] server.

The meteorological products of temperature and humidity at 2m height, the wind speed and direction at 10m height, the daily accumulated precipitation, and the precipitation occurrence will be compared against observational data from surface weather stations. The occurrence of a precipitation event will be treated as a categorical variable, while all the other atmospheric fields will be treated as continuous variables. To match meteorological products and observations on the horizontal plane inverse distance weighted interpolation of 9 points will be used for temperature, humidity and wind speed, while nearest neighbor interpolation will be applied for precipitation.

[1] http://www.dtcenter.org/met/users/index.php

[2] https://madis.noaa.gov/

[3] https://madis-data.ncep.noaa.gov/madisPublic1/data/

[4] https://madis-data.ncep.noaa.gov/

This project is co-funded by the European Union

5.3 Statistical Measures of Verification for Continuous Variables

The verification of continuous variables is based on the forecast error (e.g. forecast - observation). The general framework by Murphy and Winkler (1987) which is implemented in the MET software proposed that also the characteristics of forecasts, observations, and their interactions should be considered. This allows for a verification of various aspects of performance and not only of a single measurement (Brown et al., 2017).

The statistical measures that will be used to verify the continuous atmospheric fields of temperature and humidity at 2m height, wind speed and direction at 10m height, and the amount of daily precipitation are the Mean Absolut Error (MAE) and the Mean Error (ME). In these definitions, the notations used are: f – forecasts, o – observation, and n – the number of forecast-observation pairs.

The ***Mean Error (ME)*** is a measure of overall bias of the forecast. An identical forecast has ME=0.

$$ME = \frac{1}{n}\sum_{i=1}^{n}(f_i - o_i) = \bar{f} - \bar{o} \qquad (12)$$

The *Mean Absolute Error (MAE)* is a measure of the absolute deviation between forecast and observation. MAE is less influenced by large errors and does not depend on the mean error. A prefect forecast has MAE=0.

$$MAE = \frac{1}{n}\sum |f_i - o_i| \qquad (13)$$

5.4 Statistical Measures of Verification for Categorical Variables

The verification statistics for categorical variables are expressed with the help of a contingency table (Table 3). The forecast (f) and observation (o) values are represented by two possible values 0 and 1. The information in Table 3 and definitions of the statistical measurements presented below are taken from Brown et al. (2017).

		Observations		
		o = 1 ("Yes")	**o = 0 ("No")**	
Forecasts	**f = 1 ("Yes")**	n_{11}	n_{10}	$n_{1.} = n_{11} + n_{10}$
	f = 0 ("No")	n_{01}	n_{00}	$n_{0.} = n_{01} + n_{00}$
		$n_{.1} = n_{11} + n_{01}$	$n_{.0} = n_{10} + n_{00}$	$T = n_{11} + n_{10} + n_{01} + n_{00}$

Table 3. Contingency table in terms of counts. The values in this table are counts of the number of occurrences of the four possible combinations of forecasts and observations (i.e., precipitation occurrences). The n_{ij} values in the table represent the counts in each forecast-observation category, where ***i*** represents the forecast and ***j*** represents the observations. The ***"."*** symbols in the total cells represent sums across categories. ***T*** represents the total number of forecast-observation pairs**.** The counts, n11, n10, n01, and n00, are sometimes called the "Hits", "False alarms", "Misses", and "Correct rejections", respectively. Table and information modified from Brown et al. (2017).

This project is co-funded by the European Union

The statistical measures that will be used to verify categorical atmospheric fields, i.e., precipitation occurrences, are describing bellow.

Frequency Bias (FBIAS) is the ratio of the total number of forecasts of an event to the total number of observations of the event. The best value of FBIAS is 1. If FBIAS is greater than 1 means that the event was forecasted too frequently, while a value less than 1 indicates that the event was not forecasted frequently enough.

$$Bias = \frac{n_{11} + n_{10}}{n_{11} + n_{01}} = \frac{n_{1.}}{n_{.1}} \qquad (14)$$

Probability of Detection (POD) is the fraction of the events that were correctly forecasted (total number of times the precipitation was correctly forecasted to the total number of observed precipitation events). POD ranges between 0 and 1, and the best forecast has a POD value equal to 1.

$$POD = \frac{n_{11}}{n_{11} + n_{01}} = \frac{n_{11}}{n_{.1}} \qquad (15)$$

False Alarm Ratio (FAR) is the proportion of forecasts of the event occurring for which the event did not occur. FAR ranges between 0 and 1, and the best forecast has a FAR value equal to 0.

$$FAR = \frac{n_{10}}{n_{10} + n_{11}} = \frac{n_{10}}{n_{1.}} \qquad (16)$$

Critical Success Index (CSI) is the ratio of the numbers of times the event was correctly forecasted to occur to the number of times it was either forecasted or occurred.

$$CSI = \frac{n_{11}}{n_{11} + n_{10} + n_{01}} \qquad (17)$$

Equitable Threat Score (ETS) or ***Gilbert Skill Score (GSS)*** is based on CSI, corrected for the numbers of hits that would be expected by chance (C_1). ETS ranges from -1/3 to 1. The perfect forecast would have an ETS equal to 1, whereas a no-skill forecast would have an ETS equal to 0.

$$ETS = \frac{n_{11} - C_1}{n_{11} + n_{10} + n_{01} - C_1} \qquad (18)$$

where C is defined by:

$$C = \frac{(n_{11} + n_{10})(n_{11} + n_{01})}{T} \qquad (18.1)$$

This project is co-funded by the European Union

APOLLO

6 BIBLIOGRAPHY

Blöschl, G., Blaschke, A. P., Broer, M., Bucher, C., Carr, G., Chen, X., ... & Haas, P. (2016). The Hydrological Open Air Laboratory (HOAL) in Petzenkirchen: a hypothesis-driven observatory. Hydrology and Earth System Sciences, 20(1), 227-255.

Brown, B., Bullock, R., Fowler, T., Gotway, J. H., Newman, K. and Jense, T. (2017). Model Evaluation Tools Version 6.0 (METv6.0) – User's Guide. Developmental Testbed Center, Boulder, Colorado

Dorigo, W. A., Wagner, W., Hohensinn, R., Hahn, S., Paulik, C., Xaver, A., Gruber, A., et al. (2011). The International Soil Moisture Network: a data hosting facility for global in situ soil moisture measurements. Hydrology and Earth System Sciences, 15(5), 1675-1698. doi:10.5194/hess-15-1675-2011

Dorigo, W. a., Scipal, K., Parinussa, R. M., Liu, Y. Y., Wagner, W., de Jeu, R. a. M., & Naeimi, V. (2010). Error characterisation of global active and passive microwave soil moisture datasets. Hydrology and Earth System Sciences, 14(12), 2605-2616. doi:10.5194/hess-14-2605-2010

ESA. (2007). "EnviSat ASAR Product Handbook". Issue 2.2. 27 February 2007.

EUMETSAT. (2015). "ASCAT Product Guide," Tech. Rep. Doc. No: EUM/OPS-EPS/MAN/04/0028, v5, 2015.

Gruber, A., Su, C. H., Crow, W. T., Zwieback, S., Dorigo, W. A., & Wagner, W. (2016). Estimating error cross - correlations in soil moisture data sets using extended collocation analysis. Journal of Geophysical Research: Atmospheres, 121(3), 1208-1219.

H SAFa. (2016). "Algorithm Theoretical Baseline Document (ATBD) Soil Moisture Data Records, Metop ASCAT Soil Moisture Time Series," Tech. Rep. Doc. No: SAF/HSAF/CDOP2/ATBD, v0.3, 2016.

H SAFb. (2016). "Product User Manual (PUM) Soil Moisture Data Records, Metop ASCAT Soil Moisture Time Series". Doc. No: SAF/HSAF/CDOP2/PUM, v0.3, 2016.

H SAFc. (2015). "Algorithm Theoretical Baseline Document (ATBD) for product H08 – SM-OBS-2 – Small-scale surface soil moisture by radar scatterometer". Doc.No: SAF/HSAF/ATBD-08. Issue/Revision Index: 1.1., 2015.

H SAFd. (2015). "Product User Manual (PUM) for product H08 – SM - OBS - 2 – Small - scale surface soil moisture by radar scatterometer". Doc.No: SAF/HSAF/PUM - 08, Issue/Revision Index: 1.5, 2015.

ImagineS Project (2013-2016), EU-FP7-SPACE-2012-1 Grant Agreement N°311766, Available online at: [http://fp7-imagines.eu/]

Liu, Y. Y., Dorigo, W. a., Parinussa, R. M., de Jeu, R. a. M., Wagner, W., McCabe, M. F., Evans, J. P., et al. (2012). Trend-preserving blending of passive and active microwave soil moisture retrievals. Remote Sensing of Environment, 123(October 2006), 280-297. Elsevier Inc. doi:10.1016/j.rse.2012.03.014.

Loew, A., Bell., W., Brocca, L., Bulgin, C., Burdanowitz, J., Calbet, X., Donner, R.V., Ghent, D., Gruber, A., Kaminski, T., Kinzel, J., Klepp, C., Lambert, J-C., Schaepman-Strub, G., Schröder, M., Verhoelst, T. (in review). Validation practices for earth observation data across communities. *American Geophysical Union, Reviews of Geophysics*

Murphy, Allan H., and Robert L. Winkler. "A general framework for forecast verification." Monthly Weather Review 115.7 (1987): 1330-1338.

Naeimi, V. (2009). “Model improvements and error characterization for global ERS and METOP scatterometer soil moisture data,” PhD dissertation.

Richter, K., Hank, T. B., Vuolo, F., Mauser, W., & D’Urso, G. (2012). Optimal exploitation of the Sentinel-2 spectral capabilities for crop leaf area index mapping. Remote Sensing, 4(3), 561-582.

Richter, K., Hank, T. B., Atzberger, C., & Mauser, W. (2011). Goodness-of-fit measures: what do they tell about vegetation variable retrieval performance from Earth observation data. In SPIE Remote Sensing (pp. 81740R-81740R). International Society for Optics and Photonics.

Rodell, M., Houser, P. R., Jambor, U. E. A., Gottschalck, J., Mitchell, K., Meng, C. J., Arsenault, K., Cosgrove, B., Radakovich, J., Bosilovich, M., Entin, J.K., Walker, J.P., Lohmann, D. and Toll, D.. (2004). The global land data assimilation system. Bulletin of the American Meteorological Society, 85(3), 381-394.

Saxton, K.E., and W.J. Rawls. 2006. Soil water characteristic estimates by texture and organic matter for hydrologic solutions. *Soil Sci. Soc. Am. J.*, 70: 1569-1578.

Wagner, W., Lemoine, G. and Rott, H., (1999). A method for estimating soil moisture from ERS scatterometer and soil data. Remote sensing of environment, 70(2), pp.191-207.

END OF DOCUMENT