GROUND TRUTH

D4.3

MAY 31, 2022

VERSION 1.2

D4.3 – CLUG-D4.3_GROUND_TRUTH

DOCUMENT'S ADMINISTRATION	
Title	CLUG-D4.3_Ground_Truth
Deliverable number	D4.3
Leader/Responsible of this Deliverable	Siemens Mobility (SMO)
Author(s)	Bernhard Stamm (SMO), Niko Harnge (SMO), Samantha Ndibu (SNCF), Paul Fritsche (DB), Tobias Schlegel (DB)
Reviewer(s)	Andreas Wenz (SBB), Samantha Ndibu (SNCF), Valentin Barreau (SNCF), Paul Fritsche (DB), Bernhard Stamm (SMO), Niko Harnge (SMO)
Validator(s)	Muthu-Kumar Kumar (DBN)
Due date of deliverable:	November 29, 2021
Submission date	May 31, 2022
Version	1.2
Status	final

DISSEMINATION LEVEL		
PU	Public	
CO	Confidential, restricted under conditions set out in Model Grant Agreement	X

CONFIDENTIAL

VERSIONS OF THE DOCUMENTS

Revision	Date	Description
0.1	18/10/21	Initial version for review of structure
0.5	29/11/21	First version with input from SNCF and DB for review
1.0	23/12/21	Release after review for submission
1.1	22/02/22	Release for external review
1.2	31/05/22	Update according to comments from external reviewers

CONFIDENTIAL

EXECUTIVE SUMMARY

This document contains the description of how Ground Truth data is being produced as part of WP4 of the CLUG project. Together with the actual Ground Truth data supplied with each set of raw data It forms the Deliverable D4.3.

This document is based on the terms and conditions established in the Grant Agreement (GA) and its Annexes, as well as in the Consortium Agreement (CA).

CONFIDENTIAL

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

APPLICABLE DOCUMENTS

The following documents define the contractual requirements that all project partners are required to comply with:

- Grant Agreement N°687384 (which includes DOW, Grant Preparation Forms and annexes): This is the contract with the European Commission which defines what has to be done, how and the relevant efforts.
- Consortium Agreement: This defines our obligations towards each other.

Each of the above documents was established at the start of the project, and copies were supplied to each partner. Each document could potentially be updated independently of the others during the course of the project following a prescribed process. In the event of any such update, the latest formal issued version shall apply.

In the event of a conflict between this document and any of the contractual documents referenced above, the contractual document(s) shall take precedence.

REFERENCES

CLUG. «Consortium Agreement.» 2020.

—. «D1.1 - Project Management Plan.» February 2020.

—. «D1.2 - PROJECT QUALITY PLAN.» February 2020.

—. «D1.4 - Data Management Plan .» March 2020.

—. «D5.1 - Communication Strategy and Action Plan (CSAP).» 2020.

—. «GRANT AGREEMENT NUMBER 870276.» 2019.

ERA. «ERTMS Longer Term Perspective.» 2015.

GSA. «GNSS Market Report issue 5 .» 2017.

Institute, Project Management. *A Guide to the PROJECT MANAGEMENT BODY OF KNOWLEDGE.* 2017.

CONFIDENTIAL

ACRONYMS

A

AEF *Agence d'essai ferroviaire / Railway Test Agency of SNCF*

APPS *advanced post processing software*

aTL *advanced Train Lab, test train used by DB*

C

CA *Consortium Agreement*

D

DB *Deutsche Bahn*

Domino *Regional train type used as test train in Switzerland*

E

ETCS *European Train Control System*

EUREF *Reference Frame Sub-commission for Europe of the International Association of Geodesy (IAG)*

G

GA *Grant Agreement*

GNSS *Global Navigation Satellite Services*

I

IMU *Inertial Measurement Unit*

INS *Inertial Navigation System*

M

Martine *Passenger coach used as test vehicle by SNCF*

P

PL *Protection Level*

PVT *Position Velocity Time*

PVT+ *Position Velocity Time plus Auxiliary Data*

R

RTK *Real Time Kinematics*

S

SBB *Schweizerische Bundesbahnen, Swiss Federal Railways*

SNCF *Société nationale des chemins de fer français*

STARS *Satellite Technology for Advanced Railway Signalling*

1 INTRODUCTION

1.1 IDENTIFICATION AND PURPOSE

1.1.1 General Information

In order to analyse the performance of any positioning solution which produces position, speed, time and potentially other data, a reference which can be assumed to be true is required. This true data is called Ground Truth and needs to be generated for every set of data which shall be analysed.

This document contains a description of the methods and procedures applied and the tools used to produce Ground Truth in the CLUG project. It contains three separate sections for the measurement campaigns performed by Siemens and SBB in Switzerland, by SNCF in France and by DB in Germany, as the conditions regarding available track maps, sensor data and absolute position references are different in the three countries.

The aim of this document is to allow users of Raw and Ground Truth data, which will be published by the CLUG project as Deliverable D4.4, to understand the underlying methodologies used to generate Ground Truth.

1.1.2 Basic Concept of WP4

The basic concept of WP4 is based on a test platform, which emulates a positioning system, and in which raw data from GNSS and from other sensors collected in the rail environment are combined to generate corresponding PVT and PL streams. The main advantage of this method is that different fusion algorithms can be applied, and their performance compared using raw data from the same trips, without the need for additional data collection for each algorithm.

Additional data beyond the standard PVT, such as acceleration, attitude, heading etc. is also being generated, as such data might be applied in future railway applications. The term PVT+ is therefore used in this document to indicate the presence of such additional information.

As part of WP4, raw data from GNSS and many other sensors is collected on a commercially operated train in Switzerland, and on two test trains operated in France and Germany. Data is being collected in different environments and during operational scenarios, which represent train operation on the Trans European railway network.

Auxiliary data required for Ground Truth generation is also being collected on these trains, this data is however different for each train, respectively for the method applied for each train to generate Ground Truth.

Once PVT+ and Ground Truth have been produced for a journey, they are compared with each other using a data analysis tool specifically adapted in WP4 for the CLUG project. This allows evaluating the performance of different sensor combinations and fusion algorithms. Performance analysis will also be done for other aspects than positioning accuracy, such as the calculated protection level vs. the true error, the calculated speed and the behaviour of individual sensors.

CONFIDENTIAL

2 PRODUCING GROUND TRUTH DATA FOR SWITZERLAND

2.1 INTRODUCTION

In Switzerland, Ground Truth is based on absolute reference points (Eurobalises) and on odometry data from a high-resolution sensor, which is independent from slip and slide. This data is then combined with GNSS referenced track topography and track topology data provided by SBB to generate the true position of the train at any time (Ground Truth). This true position is therefore generated independently from GNSS and IMUs, which are the main sensors used by the CLUG positioning solution, to avoid impacts on the test results from common mode effects.

2.2 TEST TRAINS / TEST TRACKS / DATA COLLECTION

2.2.1 Test Train

In Switzerland, raw data, as well as data for generating Ground Truth is being collected on a Domino power car. Domino trains are used in local and regional services. They consist of a power car (RBDe 560 216 on the test train), between one and three coaches and a control cab car.

The Domino power car was selected as it has already been used in the STARS project for similar work, most of the equipment required for CLUG had therefore already been installed. While some equipment has been upgraded for the CLUG project, and some additional sensors installed, the most time consuming and complex elements, such as installing roof antennas, a cabinet for equipment, providing power and certifying the installation could be reused. This made it possible to start collecting data much quicker than if a new vehicle had to be equipped, and also reduced the cost significantly. Using another train would not have been possible considering the time required to obtain permission to use a train, to do train specific engineering, to procure and install the necessary equipment and to certify the installation.

Figure 1: Domino Test Train used in Switzerland

A number of aspects have led to the selection of the Domino power car as test vehicle already in the STARS project. These reasons also apply to the CLUG project. For the case of generating Ground Truth, the equipment of the train with ETCS is the most relevant aspect, as this provides the test system with access to Eurobalise data

to obtain absolute position references. Also, the train has a baggage compartment, which is not used anymore operationally, providing the necessary space for the equipment. The baggage compartment also permits performing attended measurements during commercial operation, without interfering with passengers.

2.2.2 Test Tracks

The Domino test train is part of a larger fleet of identical trains which rotate between different services. As they are equipped with ETCS, they can operate on any standard gauge line in Switzerland. Thanks to this, data can be collected in many different environments without the need of performing dedicated test runs. The following graph shows all lines, on which data can potentially be collected within the frame of the CLUG project.

Figure 2: Test Tracks in Switzerland

Collecting data over such a vast network however provides some challenges to generate Ground Truth, as no specific measures can be taken trackside. If such measures would be taken, measurement would have to be limited to a small number of dedicated tracks.

2.2.3 Eurobalise Data

Fortunately, the entire standard gauge rail network in Switzerland has been converted to ETCS Level 1 over the last few years, or is equipped with ETCS Level 2, providing Eurobalises which can be used as absolute position references. The Swiss standard gauge network consists of approximately 3'000 km of lines with approximately 7'500 km of track, on which around 30'000 Eurobalises are currently installed.

2.2.4   Track Map

SBB also maintains a very accurate, geo-referenced database of its assets, which is being used for many purposes, including civil works. This database not only includes track centre lines in cm accuracy, but also many other assets, such as the locations of signals, points, Eurobalises and also tunnels, bridges, buildings, platforms etc. In addition, the track data is maintained as a node and spoke model, from which also the track topology can be derived. While no formal confirmation could be obtained on the accuracy of the map, its accuracy is considered by SBB to be in the range of a few mm, the actual data values are provided with a resolution of 0.1 mm. The data extracted from the database forms the track map used in the CLUG project.

Note that the track centreline included in the map can deviate from the true centre line due to e.g. expansion or shrinking caused by changing temperatures.

The database is however not referenced to the WGS 84 coordinate system, which is being used by GNSS, but to the local Swiss Coordinate System, which is sometimes also called Swiss Grid. Swiss Grid is the standard system used in Switzerland for surveying and mapping by the Swiss Federal Office of Topography swisstopo, as well as by cantonal and municipal authorities. Fortunately, swisstopo provides a conversion algorithm between Swiss Grid and WGS84, which is being used to convert the track map for usage in the CLUG project.

The following figure shows an extract of the database, which has been converted into the WGS84 referenced track map as used in the CLUG project. The map data in this example is overlayed onto OpenStreetMap data for visualisation purpose:
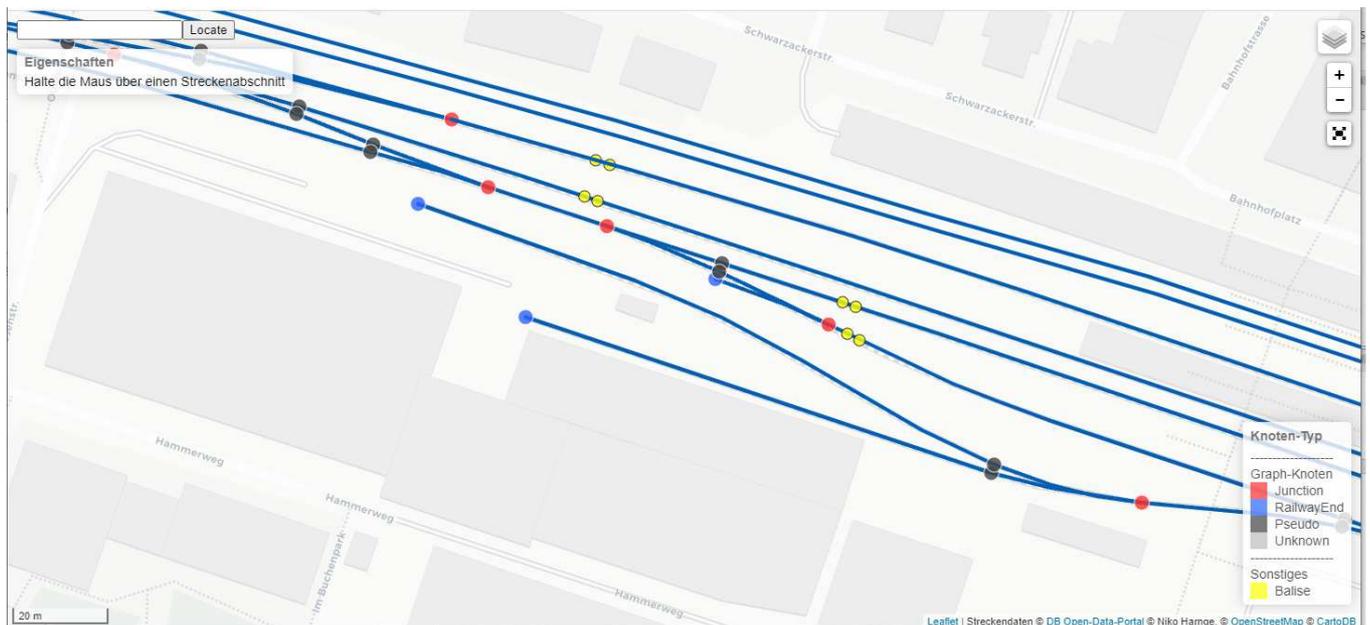
Figure 3: Example of Data from Track Database

The blue lines are the track centrelines, representing the spokes of the topology model, the red, blue and black dots are the nodes of different types (one, two or three spokes attached). Eurobalises are shown as yellow dots in the above figure.

Note that balises and nodes are contained in the database with their GNSS referenced position, track centrelines with their length, as well as a sequence of GNSS referenced positions every approximately 10 m. As the curve radius is typically not lower than 300 m, this results in a lateral error of only a few cm even if the track between subsequent GNSS referenced positions is considered a straight line.

CONFIDENTIAL

Note also that Swiss regulation require that infrastructure managers ensure that the true position of the track is less than 2.5 cm from the position specified in the map (AB-EBV).

2.2.5   Odometry Data

In order to generate Ground Truth data also in between absolute positions marked by Eurobalises, odometry data is being used. Fortunately, a wide array of odometry sensors is installed on the Domino train, including two tachos (32 pulses per rotation, 9.2 cm per pulse), as well as two radars (speed every 180 ms with 0.2% accuracy) and a Corrail sensor (0.2 km/h up to 100 km/h, 0.2% above), the latter ones not being impacted by slip and slide of the axles. While each of these sensors generates some erroneous data under certain environmental or operational conditions, accurate odometry data (speed and distance travelled) can be generated by combining the data from multiple sensors.

2.3   GROUND TRUTH GENERATION

2.3.1   High Level Method of generating Ground Truth

Once Eurobalise, odometry and map data has been collected, Ground Truth is generated as follows:

- By reading a Eurobalise and retrieving its GNSS referenced position from the track map, an absolute position can be generated.
- By reading a subsequent Eurobalise, the next absolute position can be generated.
- Once both balises are identified in the track database, the route between the balises can be identified using the topology of the track derived from the node and spoke model.
- Using odometry data, intermediate longitudinal positions can be calculated along the route.
- These longitudinal positions can then be converted into GNSS referenced positions using the respective track data from the track map.

This method is described in more detail in the following sections.

2.3.2   Reading Eurobalises

The Eurobalise reader of the ETCS on-board system generates the following low-level signals:

- An Attention Signal, which indicates the presence of a carrier signal in the Eurobalise frequency band
- A Clock Signal, which indicates the presence of a data stream transmitted by the Eurobalise
- A Data Signal, which contains the actual data transmitted by the Eurobalise

By using the two flanks of the attention signal, the balise centre position can be estimated. Experience from using Eurbalises as reference points for the accurate stopping of trains in metro applications shows that the resulting longitudinal position error is typically in the range of a few cm.

By decoding the data transmitted by the balise, the individual balise can be identified and its position extracted from the track map.

Note that some balises might not be listed in the track map, or their position might not be accurate. When such balises are identified, they can be marked as "do not use" in the Ground Truth tool.

CONFIDENTIAL

2.3.3   Route Determination

As a next step, the balises recorded during each trip are being used to identify the exact route the train has taken during that specific trip. Note that even if a train runs back and forth over a line, different routes might be taken in some locations due to traffic situations.

For cases where multiple routes exist between absolute reference locations, an algorithm has been developed to identify the most likely route. The algorithm essentially compares the length of the possible routes with the distance travelled between the balises, which is calculated from the odometry data. Correct route identification however needs to be verified manually for each trip, using either the pictures taken by the on-board camera, the speed with which the train has operated or also GNSS and IMU raw data. To support this manual verification, the tool identifies the sections of each route where alternative paths could have been taken.

When an incorrect route has been selected, this can be corrected in the Ground Truth tool by marking track segments as mandatory for the route taken during that trip.

2.3.4   Interpolating Position between Balises

Odometer information from the various on-board speed and distance sensors is used to generate longitudinal positions between the absolute reference points with a selectable time interval (typically 100 ms). Multiple sensors are being used for this, depending on a number of different criteria. Despite the usage of multiple sensors, both linear and non-linear odometer errors still result as a number of factors impact the accuracy of the various sensors (e.g. incorrect wheel diameter for tachos, misalignment of radar sensors, calibration factors etc.). The linear portion of that error is being corrected by comparing the distance recorded by the odometer sensors between absolute reference locations with the true length of the track derived from the track database between these absolute references, which is then used to backwards correct the calculated relative reference points.

2.3.5   Mapping of intermediate positions to the track map

Once intermediate, longitudinal positions have been generated, they are being mapped to the GNSS referenced track location data contained in the track map. As the route has already been identified, the map matching algorithm knows which of the track spines it must use.

As a result of the map matching, true positions are being produced with a very small lateral error, which is caused by errors in the track database, as well as by the curvature of the track between subsequent GNSS referenced track location points. A slightly larger longitudinal error results from errors of the reference balise locations, the recording of the balises as well as non-linear odometer error.

The following figure shows a sample of the resulting Ground Truth data visualised on a map, with the balises shown in yellow, those balises used as absolute references in pink, and the string of relative references interpolated from odometer data between balises marked in green. Note that the offset between the balise antenna and the GNSS antenna is corrected in this step, which is why the balises used (pink) are shown with an offset from their true location (yellow, adjacent on the same track).

CONFIDENTIAL

Figure 4: Example of Ground Truth Data
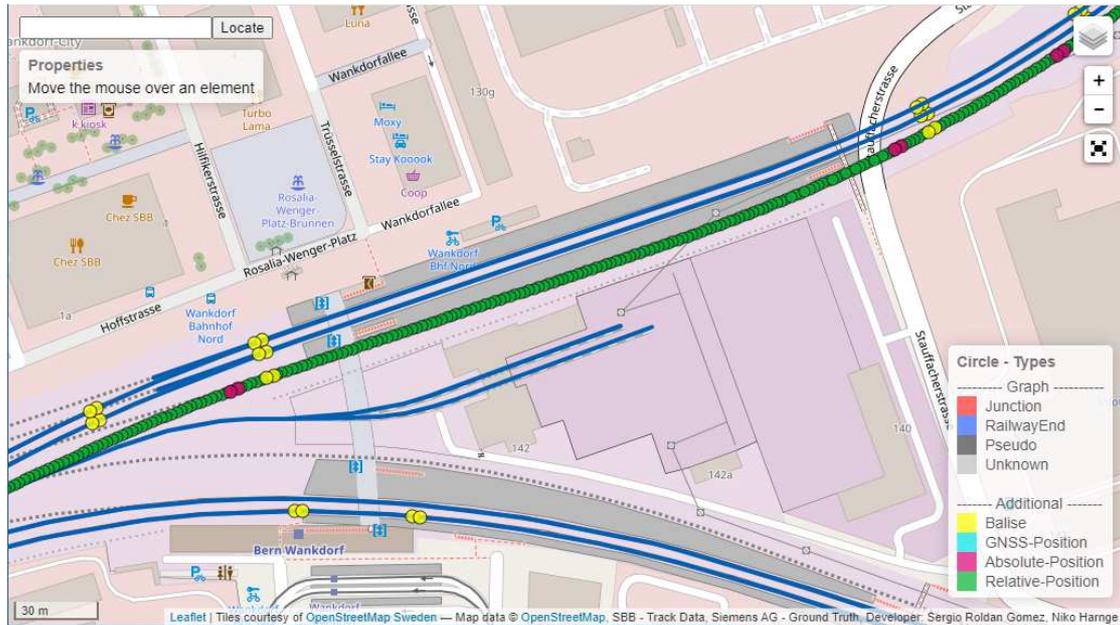
2.4   GROUND TRUTH GENERATION WORKFLOW

2.4.1   Introduction

The following figure shows the high-level workflow for generating Ground Truth data in Switzerland. Note that the conversion of track data received from SBB to .json files (SbbTrack2Json in green box), which also includes the conversion from the Swiss Grid reference system to WGS 84, only needs to be done once per data set, while the other steps (Balisedecoder, Ranalyzer2.0 and GroundTruthApp) need to be performed for each analysed trip.

CONFIDENTIAL
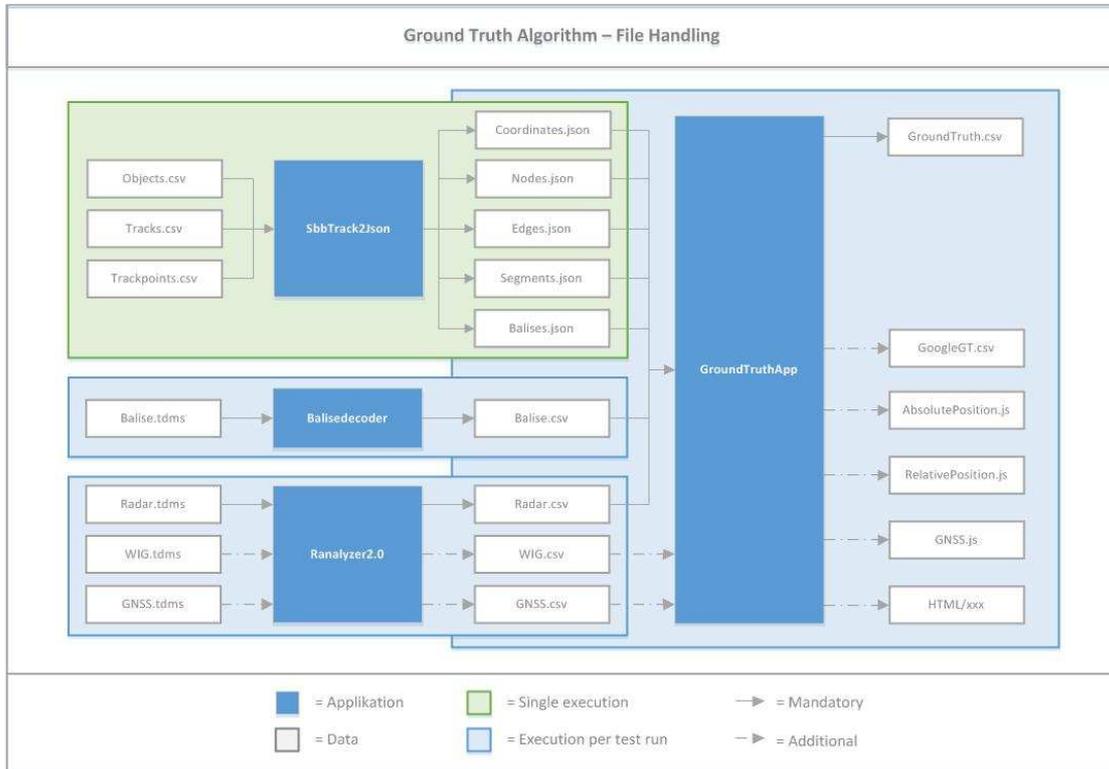
Figure 5: Ground Truth Data Generation Workflow

2.4.2   Processing of SBB Track Data (SbbTrack2Json)

Data from the SBB database is received in the form of a node-edge-model, with separate .csv files for Objects, Tracks and Trackpoints. The data in these files is referenced to Swiss Grid coordinates, the Swiss national geo referencing coordinate system.

The SbbTrack2Json application converts this .csv data to .json files, one each for Coordinates, Nodes, Edges, Segments and Balises, as the subsequent tool requires this file format. At the same time, it also converts the data from Swiss Grid coordinates to WGS 84.

Note: As a default, the SbbTrack2Json produces .json files for the entire network of SBB, covering 7'500 km of track and 50'000 balises. To speed up data handling, the tool permits limiting the data to a rectangular geographical area defined by the north-western and south-eastern corners in WGS 84 format (e.g. 46.85 / 8.55 – 46.15 / 9.05 for the Gotthard – Tessin area). Multiple areas are therefore being generated for each version of the SBB database, representing the areas in which the train typically operates. Only the respective area data is then used relevant for each trip.

2.4.2.1   Input Data

The input data received from SBB is stored in a directory whose name indicates the date on which the data has been produced. Each time a new version of the track data is received it is stored in a new directory.

Note: It is necessary to keep the .json files generated from older data releases, if raw data from the respective time shall be reprocessed in the data analysis tool chain, as map data representing the status of the network at the time the raw data has been collected needs to be used.

2.4.2.2   Execution

The SbbTrack2Json is run as a separate application. Input parameters / output parameters are read from a config file, which includes input and output directory paths as well as the coordinates within which data shall be generated.

2.4.2.3   Output Data

Output data is placed in a subdirectory to the one where the input data is placed, with the directory name representing the geographical area to which the data is limited. The output data consists of the .json files used by the tool chain.

2.4.3   Generation of Balise Data from TMDS Files (Balisedecoder)

During each trip where GNSS data is collected also data received from balises, as well as from odometry sensors is being recorded with the National Instruments data recorder installed on the test train. This data is stored in a proprietary .tdms file format.

Due to processing time required to decode balise data, only the raw bit stream of the balise (at 565 kB/s) can be recorded in real-time. As the bit stream contains a continuous repetition of the balise telegram, whose length is either 1023 or 341 bits, 2'000 bits per balise are recorded as they should contain at least one valid telegram.

For the actual generation of Ground Truth, also the balise ID and the position of the balise within a balise group are required. To generate this data the received bit stream needs to be decoded to the point, where the telegram header can be read. This step is performed by the Balisedecoder function.

2.4.3.1   Input Data

Input data is the balise data recorded by the National Instruments recorder in a .tdms file format.

2.4.3.2   Execution

The Balisedecoder application is not run as a separate application, it is executed from within the GroundTruthApp application. The Balisedecoder application does not require any configuration parameters.

2.4.3.3   Output Data

The output data is stored as a .csv file in a subdirectory of the directory where the Ground Truth generated for a specific trip respectively time slot is stored.

2.4.4   Generation of Speed Information from TDMS Files (Ranalyzer2.0)

During each trip, data from a number of speed sensors (the two tachos, as well as the two radar sensors and an optical sensor) is recorded by the National Instruments data recorder installed on the test train and stored in a proprietary .tdms file format.

For the actual generation of Ground Truth Data is necessary to produce train speed data by combining this raw sensor data. This step is done by the Ranalyzer2.0 application.

2.4.4.1   Input Data

The input data recorded by the National Instruments recorder stored in a subdirectory of the directory where the entire, trip related data is stored.

2.4.4.2   Execution

The Ranalyzer2.0 application is not run as a separate application but is executed from within the GroundTruthApp application.

It is possible to select the sensors used by the Ranalyzer2.0 application by parameters in the configuration file of the GroundTruthApp.

2.4.4.3   Output Data

The output data is stored as a .csv file in a subdirectory of the directory where the Ground Truth generated for a specific trip respectively time slot is stored.

2.4.5   Generation of Ground Truth Data (GroundTruthApp)

The GroundTruthApp is producing the actual Ground Truth data from the outputs of the three previous steps SbbTrack2Json, BaliseDecoder and Ranalyzer2.0.

- It checks whether balises recorded during a trip are listed in the track data,
- searches the track data for a possible route between subsequent recorded balises listed in the track data (which might not always be unambiguous)
- calculates the distance between recorded balises from the generated speed data
- calculates the distance between recorded balises from the track data and the possible route
- calculates the difference between the two distances (odometry error)
- interpolates the position of the train along the found route in definable time intervals, correcting it by the calculated odometry error
- maps the distance along the route from the referenced start balise at each time interval to the GNSS coordinates of the track listed in the track data

2.4.5.1   Input Data

The input data is read from the .csv files generated by the three previous steps.

2.4.5.2   Configuration

A number of options exist to configure the generation of ground truth data. A configuration file is used to store the required parameters. It is called config.properties, and is stored in the same directory where the .tdms files with the odometry and balise information are stored.

A sample configuration file looks as follows:

```
================================================================================
GROUND TRUTH GENERATOR CONFIGURATION FILE                                      |
================================================================================
This files contains the configuration properties and the values to be used in
the Ground Truth generator application. The properties and the format of the values
are explained below:

* trackdataPath:        Directory that contains the json files with the track data
                            for a concrete line.
* outputPath:           Directory where the results of the application will be stored.
* sensorType:           Sensor input data to interpolate relative positions.
                            Valid values: radar, wig or blank (means both)
* resolution:           Number of ms between interpolated relative positions.
* offset:                 Number of ms to adjust the time offset between the NI time and
                            the actual GNSS adquisition time.
* forcedRoutes:         It forces one or more paths to contain an edge id.
                            Example: ReferenceId1,ForcedId1;ReferenceId2,ForcedId2
                            The desired edge Id will be chosen as forced path once the
                            reference edge id is detected. Note the separators used in the
                            example.
* balisesSkipped:       The application will skip the balises defined in this field,
                            with the following format.
                            Example: 453-3314-1;453-3314-2...
* csvFilesCreated:      True if process of tdms files conversion shall be omitted,
                            false otherwise.
================================================================================
trackDataPath=E:\\Streckendaten\\Gotthard-Tessin\\json
outputPath=E:\\GT\\Data\\171215\\Trip3\\GT
sensorType=radar
resolution=100
offset=1
forcedRoutes=
balisesSkipped=453-3314-1;453-3314-2;453-2313-1;453-2313-2;
csvFilesCreated=true
```

Figure 6: Ground Truth Generation Configuration File

Some information on the selectable options is listed in the first part of the file, while the actual parameters are stored at the end of the file.

2.4.5.3   Execution

To start the generation of ground truth data the following conditions have to be met:

- The config file needs to be stored in the location mentioned above.
- It needs to contain the correct path to the input / output directory.
- The input data must include Radar, Tacho, Corrail, Balise and GNSS data.

One that has been ensured the ground truth generation is started from a command window as follows:

E:\GT\Tools\java –jar GroundTruth.jar "E:\GT\Data\171215\Trip3\groundtruth_!"

Where the Path "E:\GT\Data\171215\Trip3\ groundtruth _!" has to point at the directory where the TDMS files required to produce the ground truth as well as the config.properties file are stored.

The application then automatically launched the individual applications in sequence, requiring no further user input.

2.4.5.4   Output Data

The output data is stored in a subdirectory of the directory where the Ground Truth generated for a specific trip respectively time slot is stored.

The following output files are generated:

Output	Description
Ground Truth.csv	Ground truth data as specified in the respective CLUG specification.

CONFIDENTIAL

GT Test Coordinates.csv	For internal use
Importable Ground Truth.csv	For internal use
Log.txt	Contains the information which is displayed during the ground truth generation in the CMD window, most notably the path which has been generated from balise group to balise group, with the balise group ID's, the true distance calculated from the track data and the distance measured by the odometry. The difference between the true distance and the odometry distance is corrected in a linear fashion between the two used balise groups.
\html\ GT-Leaflet.html	The GT-Leaflet.html file in a newly created \html sub-directory is a browser viewable visualisation of the ground truth, which can be used for visual verification.

Table 1: Ground Truth Output

2.4.6   Verification of Ground Truth Data (GroundTruthApp)

A manual verification of the Ground Truth is then performed for the following purpose:

- to check whether the correct path has been selected
- to verify whether all balises are recorded in the track map with their correct position

If either a wrong path has been selected at a certain location, or some balises are not place correctly then this can be corrected in the configuration file shown in Figure 6 shown above. A wrong path can be corrected by marking a track segment with its ID number under "forcedRoute =", incorrectly placed balises can be suppressed under "balisesSkipped =". Once these corrections have been made the GroundTruthApp has to be executed again.

2.5   VALIDITY AND ACCURACY OF THE GROUND TRUTH

2.5.1   General Information

The accuracy of the Ground Truth generated in Switzerland is impacted by the following factors:

- The accuracy of the balise detection
- The reliability of the route determination
- The quality of the odometry
- The accuracy and correctness of the track map

Each of these aspects is described in more detail in the following paragraphs.

2.5.2   Balise reading

The accuracy of the balise reading depends largely on the time resolution, with which the attention signal is being recorded. As this is done with the full time-resolution of the NI system, which is synchronised to the same time reference as GNSS, the resulting error can be assumed to be negligible.

CONFIDENTIAL

Another source of error is the distortion of the magnetic field of the balises due to nearby metallic objects, which would impact the calculation of the balise centre location. This error is also assumed to be negligible, as balises are installed according to the installation rules defines as part of the respective standard.

2.5.3   Route determination

As described in section 2.3.3 and section 2.4.6, route determination needs to be manually verified, as multiple routes might exist between two balises. It is assumed that the correct route is being used, once the manual verification has been done, so that this source of error is excluded.

If an incorrect route is being used none-the-less, then this would result in a significant lateral error over a longer distance, and thus be visible in the data analysis.

2.5.4   Odometry

The odometry sensors used are impacted by various environmental factors. Due to that, multiple sensors are used to generate the odometry data used by the Ground Truth. Remaining errors can be both linear, e.g. by incorrect wheel diameters used with the tachos, or by mounting inaccuracies of the radars.

Errors can also contain a non-linear component, such as from slip and slide for tachos, or from spurious errors in the radar and the optical sensor data. Finally, radars only produce reliable and accurate odometry information above a minimum speed, as the doppler shift becomes too small to be reliably detected at low speeds.

2.5.5   Track Map

The track map provided by SBB contains data with surveyor quality, as the data is used for civil works. While no guarantee is given on the accuracy, it can be assumed that this is within a few cm. For the purpose of the CLUG project this is assumed to be sufficient, without requiring verification.

Note: Also, the track inspection vehicles of SBB are relying on the track map, any significant errors would therefore also be exposed by the analysis of that data.

Track map data can however be outdated, as multiple projects are continuously being executed across the SBB network, resulting in both relaying of tracks, as well as of balises. Once projects have been completed the track map is being updated, this might however take some time.

Using an out-of-date map could result in multiple errors:

- If an incorrect along the track balise location would be read from the track map, this would be detected by the algorithm which compares the distance between balises determined from the track map data and the distance travelled calculated from odometry data. The percentage of error is indicated by the tool and is typically quite constant, any significant difference in accuracy, if happening repetitively for the same balise, can then be investigated.
- If an incorrect lateral balise location would be read from the track map (balise on wrong track) then this would be detected by the algorithm which determines the route, as no route might be found, or by the route verification, as a wrong route would be determined.
- If newly installed points are missing in the track map, then this would be detected by the algorithm which determines the route, as no route might be found, or by the route verification, as a wrong route would be determines.

- If a track has been re-positioned, but the new position not yet contained in the track map, then this would result in a lateral position error over a longer distance, which can be detected in the analysed data.

2.5.6 Conclusion

From the above-described mechanisms it can be concluded that any error in the Ground Truth will likely be exposed by either the tool chain, the manual route verification or be visible in the output data generated by the data analysis tool.

For verifying the quality of the Ground Truth, two methods will none-the-less be tested:

- By a statistical analysis of the odometry error resulting from comparing the true distance derived from the track map with the distance measured by the odometry, incorrect balise location data in the track map can be exposed. Such an analysis can also verify the quality of the odometry.
- By excluding some intermediate balises on a line from the generation of the Ground Truth and comparison of the Ground Truth position interpolated at the location of the not used balises, the accuracy of the interpolation can be determined.

3   PRODUCING GROUND TRUTH DATA FOR FRANCE

3.1   INTRODUCTION

In France, Ground Truth is based on high grade INS data which are post processed with the APPS software. Thereby, IMU and GNSS raw measurement post processed with RTK are processed by APPS postprocessing INS engine will then run the data in forward and backward time directions. This double processing will allow improving the output reference solution.

Ground truth generation can be resumed by the following figure which shows inputs, methods and outputs involved in the ground truth process.
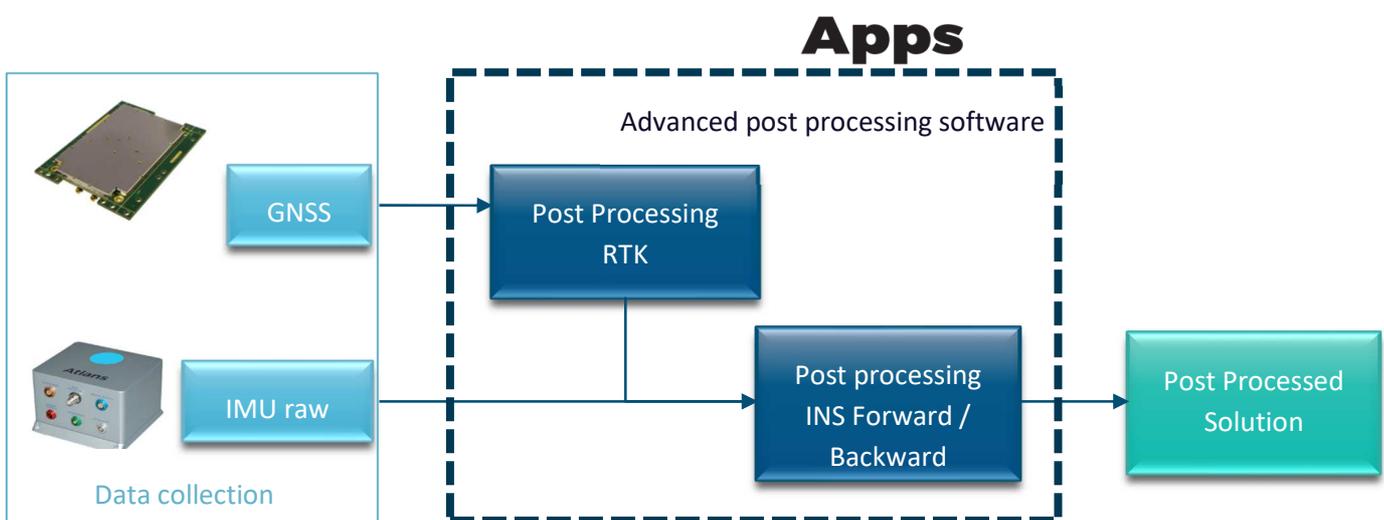
Figure 7: APPS: 8 Ground truth architecture through APPS software

3.2   TEST TRAINS / TEST TRACKS / DATA COLLECTION

3.2.1   Test Train

The French test train Martine, shown in the previous figure, is composed of:

- a freight locomotive BB 60 000. This is a diesel locomotive that can be considered as an intermediate power machine, and
- a Corail coach (Martine), which is a type of passenger car from SNCF which have been put into service between 1975 and 1989. This car is equipped to run at speeds up to 200km/h but commonly runs at 90km/h when testing.

CONFIDENTIAL

Figure 9 : Martine Test Train used in France

Martine is a dedicated test train which collects raw and ground truth data. It was chosen because this train is used by the localisation team in France to make field trials with AEF. Many test and installation have been done since 2019 and allowed SNCF to rise in competences in data acquisition.

To prepare the CLUG project, data recorder system was developed to record time synchronized data and all the calculated and raw data from localization sensors such as IMU, GNSS, odometer or radio frequency receivers. To do so, it required to develop a data acquisition chain and to install all the sensors in a dedicated area which allow them to be fixed in the train. This has been done during the end of 2020 and all of them have been fixed in the middle of the vehicle with their own basements.

Now, during every test run, SNCF records, among others, high grade INS data that can be post processed to generate the ground truth associated to the measurements. Thereby, Martine is fully equipped to record at the same time raw and ground truth data.

3.2.2 Test Tracks

The Martine test train can run on many tracks due to its adaptability. As a matter of fact, the train can be used on all commercial and service tracks because it has a universal gauge specified as G1 gauge.

In fact, independently of each other, the freight locomotive and the coral coach can circulate on all the French railway network. Very few exceptions exist and can be neglected because the coach can be carried by other locomotives that have the right circulation authorizations.

However, Martine allows SNCF to cover all scenarios, environments, and tracks in CLUG framework. It can run on tracks such as:

- Regional tracks,
- Freight tracks,
- Main tracks.

Meanwhile, if Martine needs to go beyond 100km/h, the BB 60000 locomotive needs to be changed for a:

- BB 22200: travellers electrical or diesel locomotive (to run up to 200km/h),
- BB 26000: travellers electrical or diesel locomotive (to run beyond 200km/h).

CONFIDENTIAL

Thanks to its adaptability, Martine can also be operated on high-speed lines if the locomotive is equipped with the TVM (Machine Track Transmission) system.

At this stage, Martine has already run on the following tracks during dedicated runs:

- Regional lines (example: field trials on RERC, regional train in Ile de France or bi-weekly runs from Vitry to Montereau),
- Freight lines (example: run from Paris to La Tour De Carol through Limoges and Toulouse),
- Main lines (example: run from Paris to La Tour De Carol through Limoges and Toulouse).

Thus, SNCF test train can be used on all the French railway networks if we changed its locomotive and can therefore covers all the type of tracks required for the CLUG project through dedicated test runs. As a matter of fact, French railway network represents 30.000km of lines. It is more reasonable to investigate and compute ground truth data only on railway environments of interest.

3.2.3   INS data

Ground truth is generated with INS data that are post processed. To post process this data, INS measurements are recorded during test runs with the Atlans A7 INS from IXblue.

The data acquisition tool chain allows to save these data in CLUG defined formats (CSV files). Meanwhile, to post process with APPS the GNSS and IMU measurements, it is necessary to save these data into their preoperatory format. Thereby, during every test run, two Atlans protocols are recorded in the INS internal disks:

- ATLANS-YYYYMMDD-HHMMSS_OutC_POSTPROCESSING.log : unfiltered IMU data, PVT (and associated standard deviations), attitudes and GNSS positions.
- ATLANS- YYYYMMDD-HHMMSS _OutD_GPS_RAW.log : NMEA (GGA) data and raw data (pseudo range, carrier phase, doppler and SNR)

Both of them are time synchronized and referenced and can be loaded on APPS software in order to be processed. They are recorded from the initialization of the INS to the end of the journey.

3.2.4   RTK data

Ground truth is generated through INS data that are post processed. To post process the GNSS raw measurements and to gain precision, RTK corrections are required and need to be collected. The RTK data collection is done after the runs.

RTK corrections data might come from two different sources: RTK Orpheon web interface or APPS (IXblue software). Both allows to increase precision by providing two main services:

- RTK (Real Time Kinematic): provides in real time corrections data from an observation base to GNSS receivers,
- NRTK (Network Real Time Kinematic): uses several observation bases to optimize corrections information.

CONFIDENTIAL
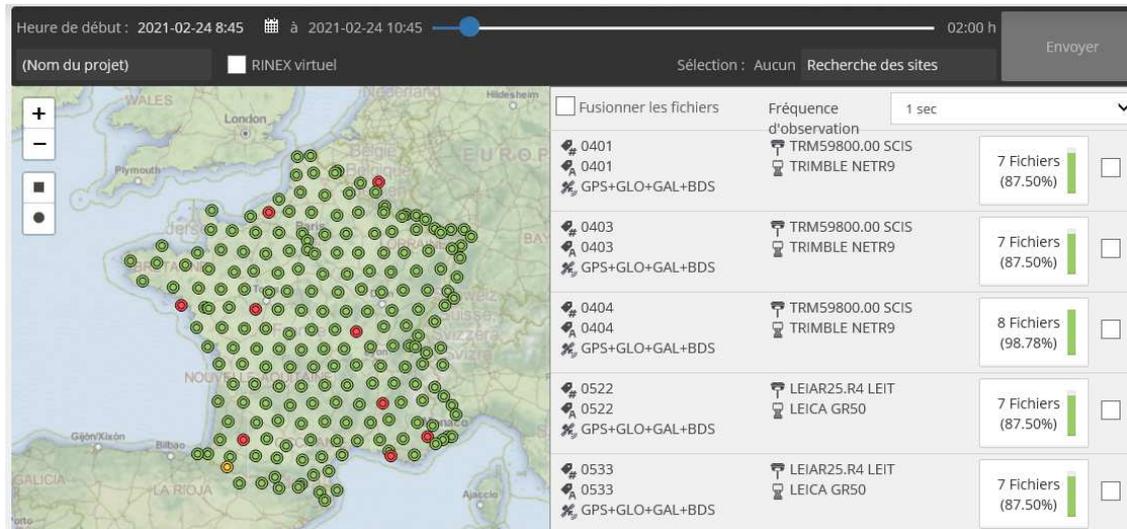
Figure 10: RTK Orpheon web interface

On RTK Orpheon, data of these observations' bases are stored in a centralized server and are saved for 30 days. Thereby, they can be download after experimentations. For that, users must select the starting date on the top of the interface (see Figure 10) to see available stations from the experimentation day. Then, stations can be selected in the list and RINEX observation files from them can be generated and downloaded.

This RINEX files can then be interpreted by APPS to post processed GNSS raw measurement from the INS sensor.

Figure 11: APP RTK database extractions interface

On APPS, a tool allows to download RINEX databases from the IGN stations as shown of Figure 11. Thanks to the GNSS raw data recorded by the INS, it identifies the train journey (dates, times and ways) and all the surrounded stations. Then, it is possible to download the RINEX data of each base station and to save it in the same folder as the GNSS raw and IMU measurements one.
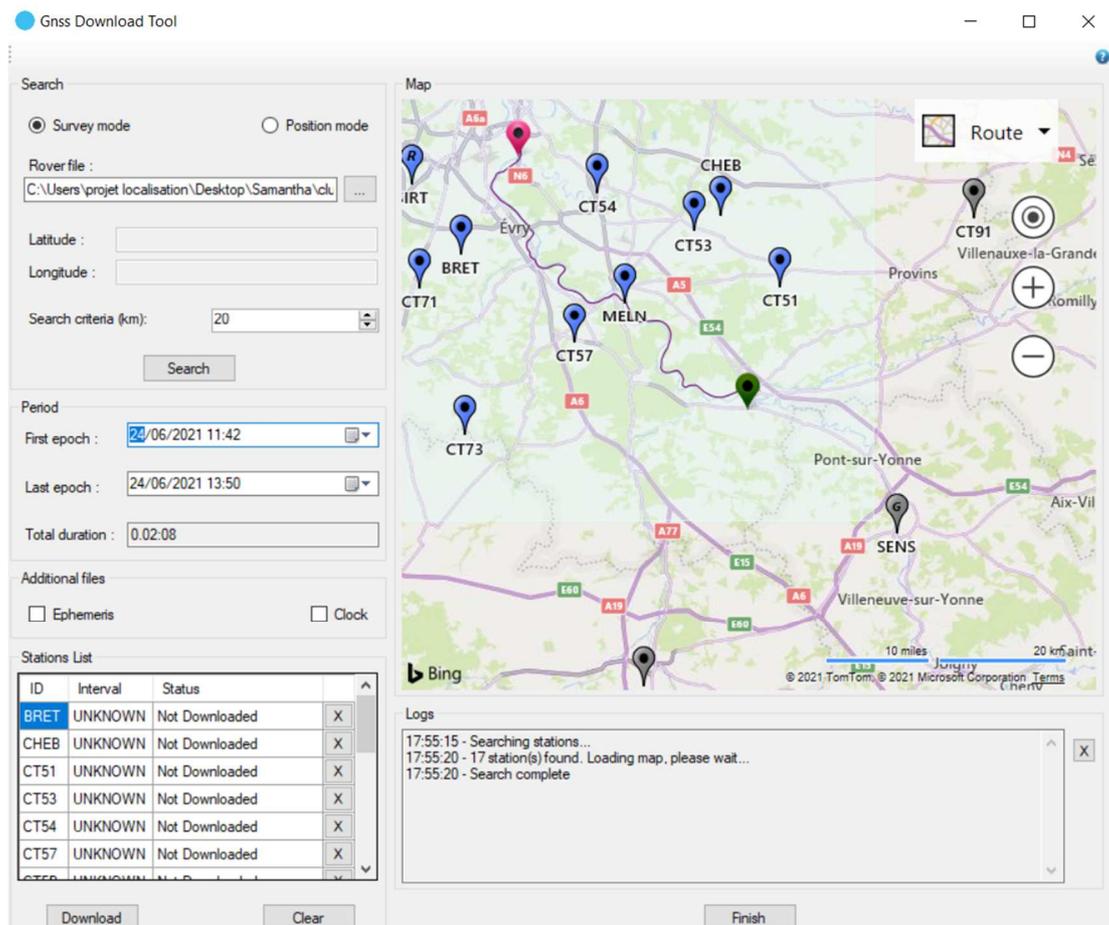
3.3 GROUND TRUTH GENERATION

3.3.1 High Level Method of generating Ground Truth

Once INS data has been collected during test runs, Ground Truth is generated as follows:

- By uploading IMU and GNSS data in their proprietary format in an APPS project
- By reading GNSS raw data measurements to retrieve the route and the surrounding RTK base station (20km around measurements)
- By downloading RINEX data from the RTK base stations identified
- By applying lever arms to INS data to project measurements at the train front end
- Once data have been projected at the train frond end, forward and backward processing can be applied to data
- Ground truth is then generated with the INS data that has been processed
- Conversion to ground truth format is done by performing a Python script on processing resulting data

CONFIDENTIAL

This Method is described in more detail in the following sections.

3.3.2   Data importation to APPS' project

APPS (Advanced Post Processing Software) is a software that allows SNCF to process INS data to create the ground truth. It is a very powerful software as it can process both GNSS and IMU data with techniques such as: RTK corrections and forward-backward processing.

That is why SNCF decided to use this software to generate the ground truth, and so that data might be imported into it. Thereby, data that has been stored in proprietary format is imported in APPS. IMU and GNSS raw measurements are imported as both are used to generate the reference trajectory. As already mentioned, all data are already time stamp and time synchronized. Their sampling frequency is 200Hz.

3.3.3   RTK corrections

Once data has been imported, RTK corrections can be realized. SNCF disposes of two methods to extract the RINEX data to apply RTK corrections: with RTK Orpheon service and APPS software. Meanwhile, only one method will be presented in this section: RTK correction with APPS.

Thereby, APPS has a "GNSS Download Tool" which permits to identifies RTK base stations and to download RINEX data from the good epochs all along the route. In fact, GNSS raw measurements are used to identify the route of the journey (cf Figure 11). Then it can retrieve: the 1st and the last epoch of the journey and so the time of all the measurements. Thus, for every measurement:

- The closer base stations in a perimeter of 20km are investigated.
- Active base stations at the equivalent epoch are investigated.
- RINEX data respecting both perimeter and epoch criteria are download.

RTK corrections can be then perform on GNSS data during forward backward step as it is included in the global function of processing in HMI. This function regroups INS and GNSS processing.

3.3.4   Data projection to train front end

As it has been defined in the CLUG project, the ground truth should be computed at the train front end. This transformation is done in APPS before the processing. By default, the reference solution output by APPS will be related to the Atlans A7 IMU location. Therefore, we need to compute the reference train position at the train front-end wheel axis location. The lever arm from IMU reference point to front-end wheel axis will be physically estimated in/on the train and the value will be parametrized in APPS before to run the postprocessing software. Besides, lever arms have been measured by AEF with a centimetre resolution.

▼ **Output**	
Label	primary
▼ **Lever arm**	
XV1 (positive f...	14.025 m
XV2 (positive l...	0.545 m
XV3 (positive t...	-0.217 m
► **Misalignment**	
User speed	Yes
Intermediate traj...	No
► **Rate**	

Figure 12: Screenshot of lever arm set up in APPS

Thus, before processing the data, APPS proposes to set the output lever arms as shown in Figure 12.

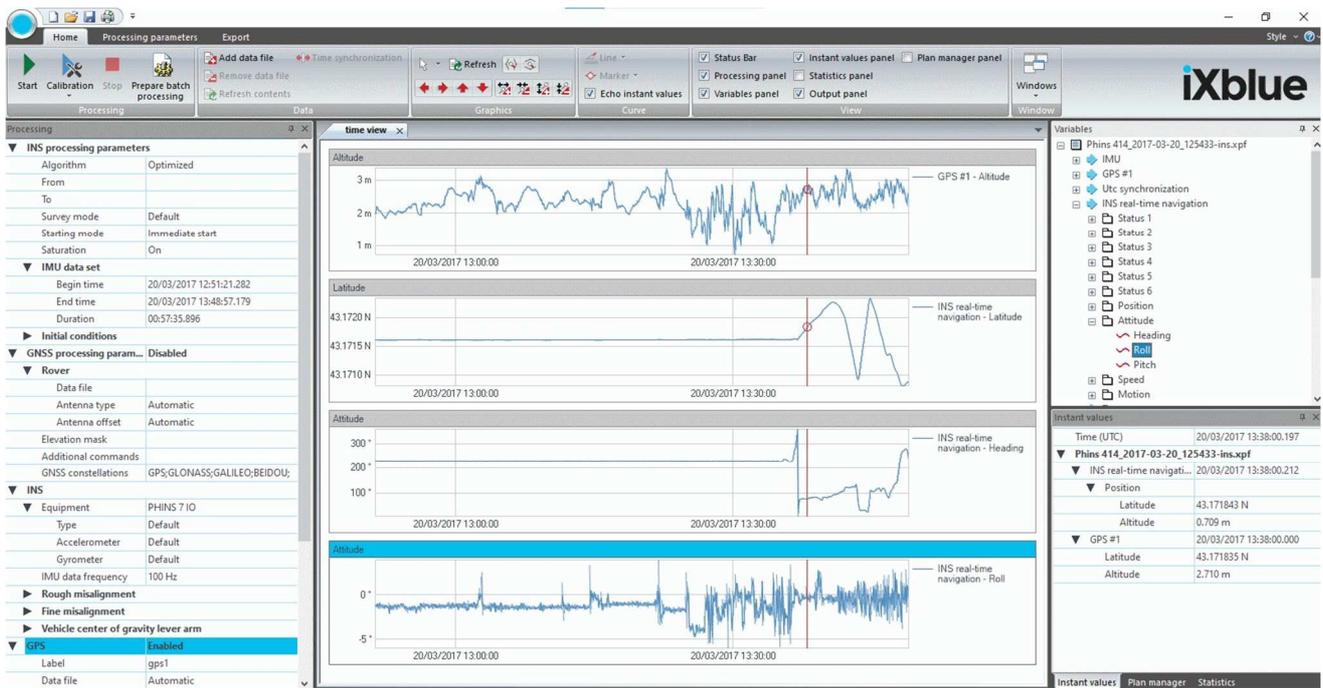CONFIDENTIAL

3.3.5 Forward backward processing

Figure 13: APPS global user interface

Figure 13 shows the APPS user interface. As it is shown, there is a dedicated tab to define processing parameters for INS, IMU and GNSS. On the right, a tab is showing the available variables that can be displayed and exported. The main tab, in the center of the window, is used to display variables.

Under INS processing parameters tab, algorithm is configured on 'Optimized'. It corresponds to the forward backward post processing method. Thereby, the post processing can be processed on INS data by pressing start button on the upper left side of the window.

3.3.6 Ground truth conversion to the CLUG format

Tools have been developed to convert the ground truth output from APPS to the CSV file required for the project. It consists in a Python script which can convert the exported txt files of the computed solution to a CSV file into the appropriated structure. The conversion method will be developed in the next section as it is linked to data workflow.

3.4 GROUND TRUTH GENERATION WORKFLOW

3.4.1 Introduction

The Figure 14 shows the high-level workflow for generating Ground Truth data in France. All data are in the WGS84 reference system. There is no change of reference system for the data except in the post processing process where data are projected at the train front end.
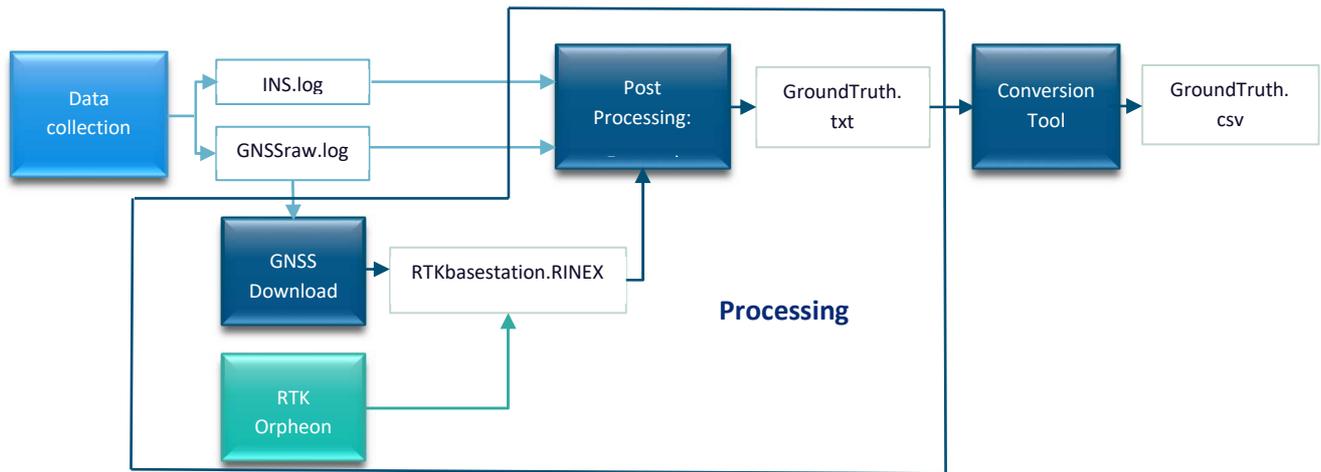
CONFIDENTIAL

Figure 14: SNCF Ground Truth Data Generation Workflow

3.4.2 Processing of INS data

3.4.2.1 Input Data

Data collection provides IXblue proprietary data with log extension which is binary data. This data is stored in two files through two proprietary protocols which have already been presented (OutC_POSTPROCESSING and OutD_GPS_RAW). It consists in INS and GNSS raw data which are coming from Atlans A7 IXblue.

The content of these input files is not presented in this section as it is only binary that can only be interpreted by APPS. The two files are stored in the folder link to the actual field trial, named and timestamped according the date and the starting time (UTC time).

3.4.2.2 Execution

APPS processing can be run with the starting button if:

- GNSS RINEX data of RTK base stations has been download for all of the journeys.
- GNSS RINEX data of RTK base stations has been added as rover data in processing parameters.
- Output lever arms have been set in processing parameters.
- Optimized processing has been set in processing parameters (required to perform both forward and backward processing

Once all these parameters are set, the processing can be started, and it will provide the output data presented in the next sections.

3.4.2.3 Output Data

When the processing is successful, post processed data can be exported in a text file as defined in Table 2:

Variable	Description
Date	Date of validity (yyyy/mm/dd)
Time	Time of validity (hh:mm:ss:ssss)
Userlabel	Associated sensor (for synchronized output)

CONFIDENTIAL

UserData	Associated sensor data (for synchronized output)
Latitude	Latitude, decimal degree
Longitude	Longitude, decimal degree
EllipsoidHeight	Ellipsoid height, meter positive upward
OrthometricHeight	Orthometric height, meter positive upward
NorthingSd	Standard deviation on latitude, meter
EastingSd	Standard deviation on longitude, meter
AltitudeSd	Standard deviation on altitude, meter
NorthingEastingCv	Crossvariance between latitude and longitude, meter^2
Heading	Heading, in degree
Roll	Roll, in degree
Pitch	Pitch, in degree
HeadingSd	Standard deviation on heading, degree
RollSd	Standard deviation on roll, degree
PitchSd	Standard deviation on pitch, degree
SpeedNorth	Speed in north direction, meter/second
SpeedEast	Speed in east direction, meter/second
SpeedVertical	Speed in vertical direction, meter/second
SpeedNorthingSd	Standard deviation on speed in north direction, meter/second
SpeedEastSd	Standard deviation on speed in east direction, meter/second
SpeedVerticalSd	Standard deviation on speed in vertical direction, meter/second

Table 2: Output data from data processing with APPS

3.4.3   Conversion of Ground Truth Data

3.4.3.1   Input Data

Input data of ground truth data conversion is the output data from INS data processing through APPS step.

3.4.3.2   Execution

To convert data from APPS text format to CLUG CSV format of file, a Python tool has been developed internally. Thanks to it, the text file is converted regarding the ground truth defined format and structure of file.

CONFIDENTIAL

Thus, the output file from APPS is defined as the input file of the Python tool. It only keeps variables of interests for the output file and if necessary, it converts the variables in the indicated units or directions. This is mandatory for the following variables:

- Altitude: changed from positive upward to positive downward.
- Attitude (heading, roll and pitch): changed from degree to radian.
- Standard deviations of attitude: changed from degree to radian.

Some random checking has been performed to ensure that the conversion between units is correct.

3.4.3.3   Output Data

Output file of the ground truth conversion is the final file as defined in CLUG D4.1 document. It is stored in the same folder as the raw measurements in internal server and it can be exported to AWS cloud respecting the naming.

3.5   VALIDITY AND ACCURACY OF THE GROUND TRUTH

The accuracy of the Ground Truth as generated in France is impacted by the assemblies' parameters and IMU performances:

- Lever arms between aiding and the IMU,
- IMU performances

3.5.1   Lever arms accuracy

In France, ground truth is generated through IMU measurements which are post processed. Meanwhile, some of the measurements must be projected in the IMU reference frame. To project these measures, lever arms between aiding systems such as lever arms between antenna and the IMU have been measured.

Thereby, lever arms have been measured and computed by combining laser rangefinder and decimetre ruler measurements with vehicle plan information. As a matter of fact, antenna is located on the roof of the vehicle. To project the measurements into the vehicle, Martine's plan and laser rangefinder have been used. A reference point in the vehicle has been defined. Then, distance between the projected reference point in the train body and the IMU is measured with the decimetre ruler.

Laser rangefinder accuracy can be neglected in front of decimetre ruler measurements. As a matter of fact, decimetre ruler errors have more impact on lever arms accuracy has it involves human in measurements. However, decimetre ruler has an accuracy of 0,5mm subject to human errors such as:

- Reading errors,
- Un-aligned measurements (regarding the three directions),
- Wrong or unprecise reference point used to realize the next measurements.

Considering the way these lever arms have been measured, it can be assumed that the related error and so accuracy is around 15cm.

3.5.2   IMU performances

Regarding the way ground truth is generated in France, IMU performances are clearly one key factor that impacts ground truth accuracy.

CONFIDENTIAL

However, three "areas" might be considered when we evaluate IMU performances. As a matter of fact, IMU performances are linked to GNSS availability. Therefore, it is preferable to split performances regarding these areas:

- IMU performances with GNSS only (Standard Positioning Service),
- IMU performances with GNSS and RTK up to 40km from base stations,
- IMU performances during GNSS outages and RTK up to 40km from base stations.

Only horizontal accuracy will be considered regarding how IMU performances are defined by manufacturers.

To be noticed that ground truth accuracy can only be evaluated regarding sensor accuracy. As a post processing is realized on IMU data, all the following accuracies can be considered as maximum accuracies. Actually, it should be ten times smaller due to processing and filtering.

3.5.2.1 IMU under natural GNSS

Atlans A7 for land applications considering DMI usage is defined with a precision of 1.200m in horizontal position. However, considering RTK base stations in a 20km circle ensure a high level of availability of RTK corrections and limits the case under natural GNSS. As a matter of fact, many data bases are available to access to RTK corrections (IGN, Orpheon etc…)

In this case, lever arms measurements can be neglected against IMU accuracy, and the ground truth accuracy can be bounded to 1.200m.

3.5.2.2 IMU performances with GNSS

When GNSS RTK corrections are available, the IMU accuracy rises to 0.008m + 0.5ppm. In this case, ground truth accuracy can be bounded to the lever arms precision since the IMU performance with GNSS is clearly unsignificant. As a matter of fact, $Precision_{leveramrs} = 15cm \gg IMU_{accuracywithGNSS}$

3.5.2.3 IMU performances without GNSS

When GNSS RTK corrections are not available, the IMU horizontal position drift is between 0.40m and 0.80m.

In this case, ground truth accuracy can be approximated to the sum of the lever arms precision and IMU position drift. Thereby, the accuracy may be bounded to 1m.

4 PRODUCING GROUND TRUTH DATA FOR GERMANY

4.1 INTRODUCTION

In Germany, Ground Truth is based on a high accuracy GNSS/INS system, which consists of three high precision ring laser gyroscopes. In order to obtain the most accurate transformation between antenna, IMU and train front, a 3D laserscan measurement with mm precision had been performed of the test train. Furthermore, the data is enhanced via post involving correction data from nearby base stations.

4.2 TEST TRAINS / TEST TRACKS / DATA COLLECTION

4.2.1 Test Train

The test train operated in Germany is Diesel driven ICE - called advanced TrainLab (aTL- which was specifically modified for data acquisition on the track side, see Figure 15. Due to the Diesel engine this ICE class 605 is independent from the traction power network and therefore can be used in versatile ways. The train consists of four segments, each containing one engine, to permit a maximum speed of 200 km/h. Additionally, the train is equipped with the national systems PZB (Punktförmige Zugbeeinflussung) and LZB (Linienförmige Zugbeeinflussung), which perform various functions in the area of railway signalling and train protection system. In the current expansion stage, the aTL is not equipped with ETCS.

Figure 15: advanced TrainLab

During the project time, the aTL has been equipped with a rack, which contains needed hardware, as for example GNSS/INS devices, PTP-Timer for synchronization, switches and a Linux based recording system. Furthermore, a GNSS-antenna and three odometry sensors have been installed.

4.2.2 Test Tracks

The aTL only operates on tracks in the German rail network. An advantage of the Diesel-powered engines is the wider area of operations, more specific to run on tracks with no overhead lines.

Figure 16: Map of the Germany railway network, url: https://geovdbn.deutschebahn.com/isr (23.12.2021)

The aTL is not allowed to drive on ETCS only tracks, like on the VDE 8 (German Unity Transport Project 8), due to the non-existing ETCS configuration of the train. Other than that, the train is permitted to access LZB tracks. In Figure 16 the accessible railway tracks are shown in black.

4.2.3   IMU, GNSS and odometry Data

The basis of the Ground Truth are the data from a wheel encoder, IMU and GNSS receiver. The data of this data sources are getting recorded in their raw format.

CONFIDENTIAL

To increase the accuracy of the position afterwards, among others when the GNSS signal is disturbed by passing tunnels, bridges etc. the raw odometry and IMU data are recorded. These data are post processed together with the GNSS data and thus the accuracy of the position is improved.

Herby the IMU from the iNAT-RQT-4001-3 for generating the Ground Truth consists of three high precision ring laser gyroscopes, three servo-based accelerometers and a multi-frequency GNSS receiver. Additionally, interface to a wheel encoder is used. The data is getting recorded with a data rate of 400 Hz. For the exact location of the IMU with respect to the vehicle body frame, train front and the mounted antenna high precision 3D laser scan was conducted.

4.2.4   RTK Data

The accuracy of the recorded GNSS data is increased through the integration of GNSS augmentation data. Therefore, the free accessible base station data of the EUREFPermanent GNSS Network are taken for the relevant period. A section of the EUREF Permanent GNSS Network is shown in the map in Figure 17.
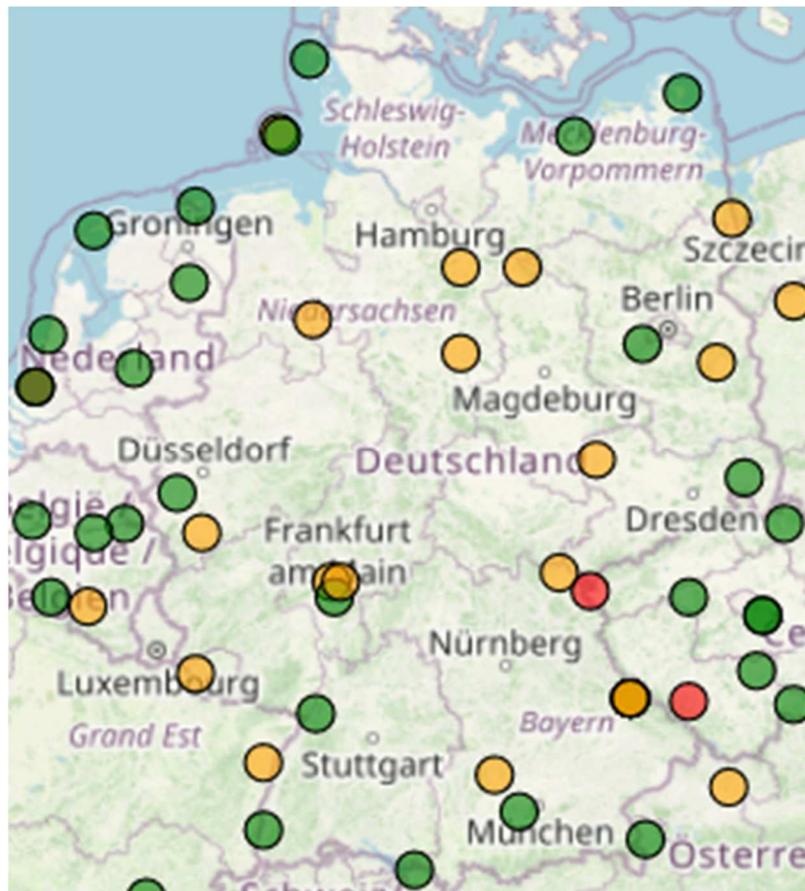
Figure 17: EUREF Permanent GNSS Network, url: https://geovdbn.deutschebahn.com/isr (26.11.2021)

4.3   GROUND TRUTH GENERATION

4.3.1   High Level Method of generating Ground Truth

Once the data from the multifrequency GNSS receiver and high accuracy IMU (iMAR iNAT-RQT-4001) has been collected, Ground Truth is generated as follows:

- Preprocessing the IMU and odometry data

CONFIDENTIAL

- Creating an Inertial Explorer project and import the GNSS and IMU data
- Obtaining GNSS correction data from base stations near to the test run.
- Performing forward and backward-propagation post processing involving correction data
- Transforming the Ground Truth data from the IMU to the train front with information from the 3D laser scan measurement
- Converting the data to CLUG specific data format

This Method is described in more detail in the following sections.

4.3.2   Preprocessing the IMU and odometry data

As a first step the binary file from GNSS/INS device collected during the test ride needs to be interpreted by the Waypoint Inertial Explorer tool that is used for post-processing, improving the quality of the Ground Truth and processing to an understandable format for the next program of the tool chain.

Hereby the iXCOM-CMD tool, the data processing and visualization tool of iMAR, is used. A result of this binary interpreter are two different files, one containing the IMU data and the other the odometry data.

4.3.3   Data importation to project

Waypoint Inertial Explorer is the software to create the needed data for the Ground Truth. The data sources are the time synchronized IMU, GNSS and odometry data which needs to be imported when starting a new project.

4.3.4   RTK corrections

The RTK engine from the Waypoint Inertial Explorer enables precise positioning and further improves the position accuracy of the Ground Truth. The nearby base stations of the taken route from the train need be identified. For convenience, the nearby base stations can be identified and downloaded manually according to the necessary correction services. A maximum of 32 base stations can be added to the project. To get proper correction data from the base stations, the distance from the base station to the corrected positions does not exceed 70 km. The tropospheric and ionospheric corrections are not accurate enough over large distances and therefore are not used if the distance to the base station is greater than 70 km. Another important factor is the number of available satellites, they should be at least 5 satellites.

4.3.5   Forward and backward processing

For a precise positioning, the IMU, odometry and GNSS data are getting processed including the usage of the RTK engine. This is done in the forward and backward direction, meaning the data are getting processed in the sequence in which they were recorded and their reverse recording sequence. Resulting in a more precise positioning. More specific the tightly coupled method is used to combine the GNSS and INS data and processing them in a further step through a Kalman Filter.

4.3.6   Data projection to train front end

After the processing is done the reference point from the body frame of the IMU is moved to the train front position. This take the different lever arms from the x, y and z transition with respect to the body frame into account. Finally, this leads to the position information of the required Ground Truth.

4.3.7   Ground truth conversion to the CLUG format

With the data gathered from the previously done steps this data needs to be transfer into the CLUG format. The CLUG format represents a reduction of the available data from Inertial Explorer. Additionally, the header and order to the columns is adjusted using a Python script.

4.4 GROUND TRUTH GENERATION WORKFLOW

4.4.1 Introduction

The high-level workflow for generating Ground Truth data in Germany is shown in Figure 18. The recorded raw GNNS data are in the WGS84 reference system. Other than the base station data which are based on the ETRS89 reference system and transformed into WGS84. As the RINEX header doesn't contain the related datum of the coordinate, we have to specify the coordinate datums of the base station. The resulting transformation is calculated automatically by the program. We here rely on Hexagon | NovAtel as a global leader in the positioning solution who programmed the software tool which is used to generate the Ground Truth.

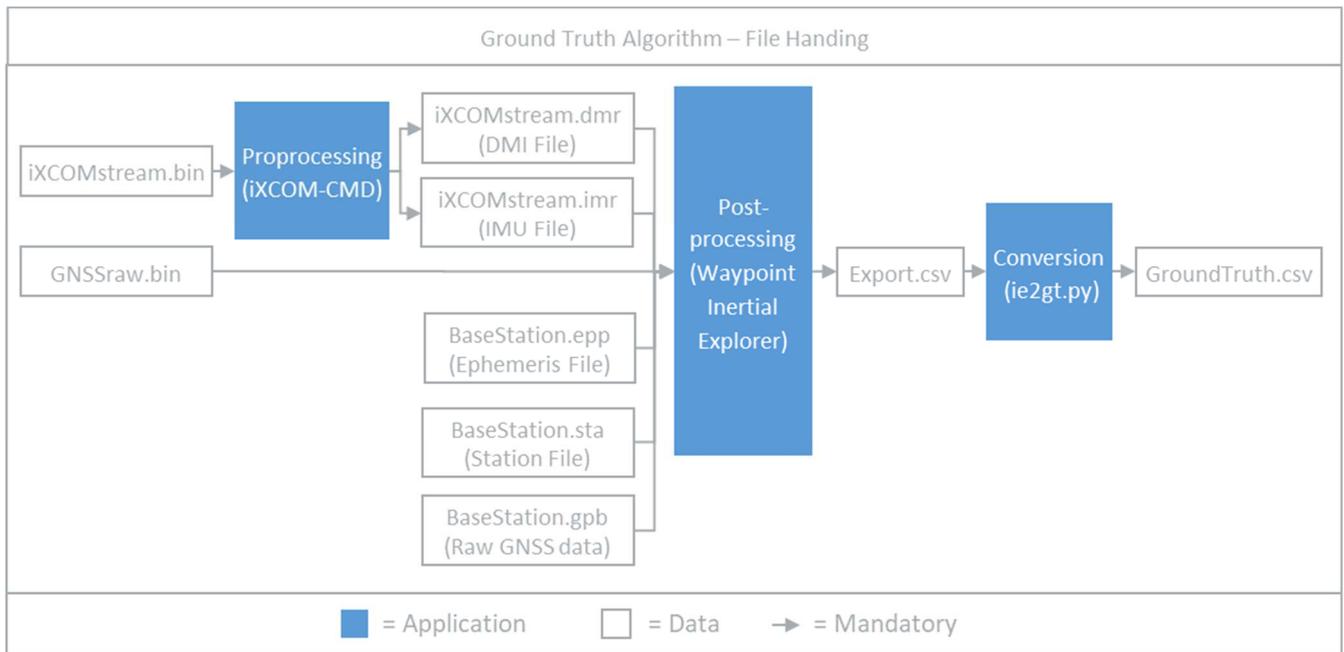All the data from one data set are achieved in the same unique directory corresponding to the timestep.

Figure 18: Ground Truth Data Generation Workflow

4.4.2 Converting of data of the reference unit (iXCOM-CMD)

4.4.2.1 Input Data

The collected data of the sensors devoid of the GNSS data is imported to the iXCOM-CMD tool. The format of the sensor data is a binary file format (iXCOMStream.bin).

4.4.2.2 Execution

The conversion process of the iXCOM-CMD tool is interpreting the binary file. Before the execution the output format is selected as "Inertial Explorer", so that it generates a readable format for the Waypoint Inertial Explorer.

4.4.2.3 Output Data

The result contains two files which are in same directory. While the .dmr file contains the sensor data of the odometry unit, the .imr file contains the data of the IMU.

CONFIDENTIAL

4.4.3   Postprocessing (Waypoint Inertial Explorer)

4.4.3.1   Input Data

Different data sources are needed for the postprocessing. The initialisation of the project is done by the import of the GNSS raw data in a binary format (GNSSraw.bin). In addition, the .imr file and .dmr file are imported. The nearby base stations are be selected directly in the tool or by importing them in the .gpb file format. The .sta file contains the position data of the station if present in the same directory. Same goes with the ephemeris data in the .epp file format. The number of base stations is limited to a maximum of 32, according to the user manual of Inertial Explorer.

Note: It is important to select the appropriate coordinate datum and the related epoch of each imported base station.

4.4.3.2   Execution

With the input of the different data the processing configuration need to be manually selected. Through the selection of the tightly coupled method the basis is set. The differential GNSS uses the GNSS data is also essential. Regarding the body frame the profile for ground vehicles and its related datum is chosen. Additionally, the configuration data is disposed in the Inertial Explorer based on the datasheet of the IMU, which contains among other the drift and biases of the gyroscope and accelerometer. The lever arms and IMU data are read from the .imr file. One example result of the processing is shown in Figure 19 and the corresponding accuracy values are shown in Table 4.

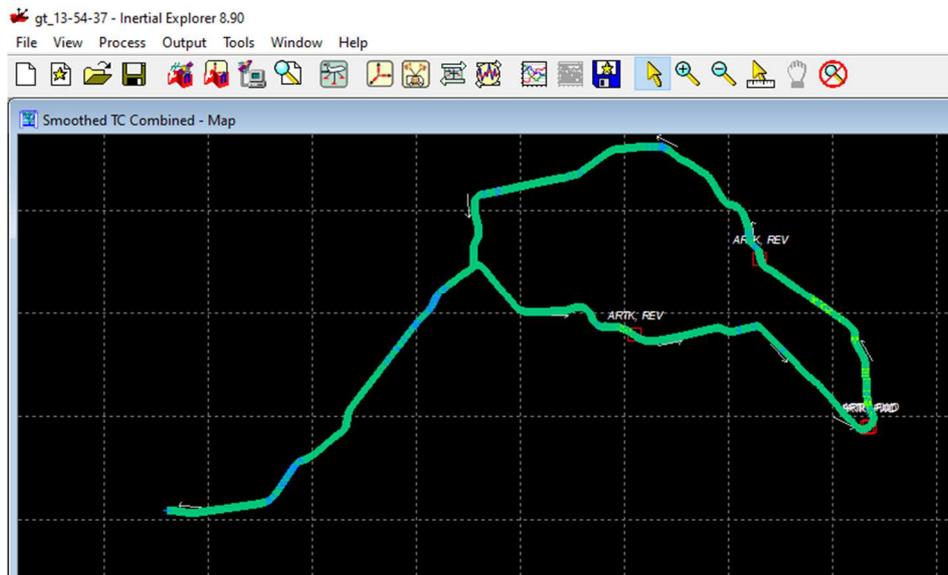*Figure 19: Processing after backward and forward processing of a dataset around Berlin*

Quality	Color	Description	3D Accuracy (m)
1	Green	Fixed integer	0.00 – 0.15
2	Cyan	Converged float or noisy fixed integer	0.05 – 0.40
3	Blue	Converging float	0.20 – 1.00

CONFIDENTIAL

4	Purple	Converging float	0.50 – 2.00
5	Magenta	DGPS	1.00 – 5.00
6	Red	DGPS	2.00 – 10.00

Table 3: Colour description for Figure 19

Before the export of the processed data, the transformation to the train front according to the body frame of the IMU is done.

4.4.3.3   Output Data

The output of the processed data is a .csv file and the position data based on the WGS84 reference system. A Section of this output data and which are relevant for the Ground Truth is shown in Table 4.

Variable	Description
UTCDate	Date according to UTC time, DD/MM/YYYY
UTC-Corr	UTC time corrected, hh:mm:ss.ss
Latitude	Latitude, decimal degree
Longitude	Longitude, decimal degree
H-MSL	Height of the sea level, meter positive upward
H-Ell	Ellipsoid height, meter positive upward
X-ECEF	ECEF-X position, meter
Y-ECEF	ECEF-Y position, meter
Z-ECEF	ECEF-Z position, meter
VNorth	Velocity in the north direction, meter/second
VEast	Velocity in the east direction, meter/second
VUp	Velocity in the upwards direction, meter/second
VelBdyX	Velocity component of the vehicle body X-axis, meter/second
VelBdyY	Velocity component of the vehicle body Y-axis, meter/second
Roll	Roll, degree
Pitch	Pitch, degree
Heading	Heading, degree
SDNorth	Standard deviation of the northing position, meter
SDEast	Standard deviation of the easting position, meter
SDHeight	Standard deviation of the height position, meter
SDX-ECEF	Standard deviation of the estimated ECEF-X position, meter
SDY-ECEF	Standard deviation of the estimated ECEF-Y position, meter

CONFIDENTIAL

SDZ-ECEF	Standard deviation of the estimated ECEF-Z position, meter
SD-VN	Standard deviation of the northing velocity, meter/second
SD-VE	Standard deviation of the easting velocity, meter/second
SD-VH	Standard deviation of the vertical velocity, meter/second
PitchSD	Standard deviation of the pitch, degree
RollSD	Standard deviation of the roll, degree
HdngSD	Standard deviation of the heading, degree
NS	Number of satellites
Q	Quality information regarding the position

Table 4: Relevant data for the Ground Truth

4.4.4   Conversion of Ground Truth Data

4.4.4.1   Input Data

The input data is a .csv file with the processed position data. The first few rows contain project information, e. g. occurred errors and warnings, IMU to GNSS antenna lever arms, UTC time offset, used station data and used datums. The information section is followed by the column names and the actual data, shown in Table 4.

4.4.4.2   Execution

During the execution of this python script ie2gt.py the section with the project information is removed. Also, the column names are formatted and the units are converted according to CLUG D4.1 document.

4.4.4.3   Output Data

The output file is the .csv file following to the naming convention and the header of the CLUG D4.1 document.

4.5   VALIDITY AND ACCURACY OF THE GROUND TRUTH

The accuracy of the Ground Truth as generated in Germany is impacted by the following factors:

- Accuracy of calibration parameters
- Quality of the Odometry/IMU
- Accuracy of the Position

Each of these aspects is described in the following sections.

4.5.1   Calibration parameters

The parameters for the calibration are having uncertainties which affects the measurements indirectly. One effect on the dimensional parameters is reduced by 3D laser scan with an accuracy of around 1 cm. Which means the important dimensions, like the lever arms with respect to the train front position can be minimalized. An example of the point cloud with its coordinates is shown in Figure 20.
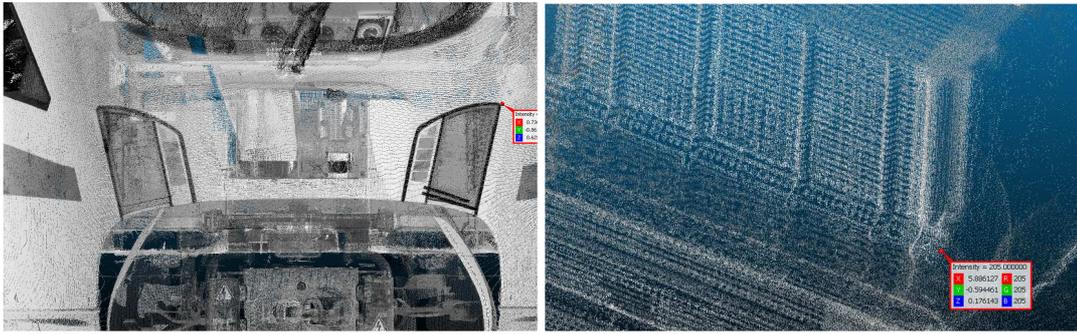
CONFIDENTIAL

Figure 20: Point cloud of the aTL, left image shows the train front, right image shows the reference coordinates of the IMU

The distance values regarding the boogies were taken from the vehicle specification. As a needed value for the IMU configuration, they also affect the accuracy. Here is the accuracy unknown, the tolerance deviation is assumed to be around 10 cm. This also considers transposing an area to an exact point coordinate.

4.5.2   Odometry/IMU

The data quality is highly dependent on the accuracy of the sensor measurements.

Considering the odometry sensor which only calculates the travelled distance based on the circumference. And is thereby dependent on the diameter of the wheel, which was obtained from a regular routine. Due to abrasion of the material the wheel diameter decreases over time. Additionally, the odometry is affected by the slip and slide effect. The sinus movement in combination with the wheel profile lead to a non-constant diameter.

The accuracy of the IMU is affected by the noise, drift and bias of the gyroscope and the accelerometer.

4.5.3   Position

The absolute position is highly dependent on the satellites. More specific it is affected – as described in chapter 4.3.4 – by the number of available satellites. To fulfil the requirements needed for a RTK corrected position the baseline to the base station should not be exceed by 70 km. In the case of corrected positions, the accuracy is according to the software manufacturer in the level of centimetre. This will be verified - for the across track error - in a further step by comparing track map data against the corrected position.

5   COMPARISON OF GROUND TRUTH METHODS

5.1   GENERAL ASPECTS

The following table contains a comparison of the main characteristics used to produce ground truth in the three countries:

Aspect	Switzerland	France	Germany
Independence of GNSS	Yes	No	No
Usability in area with no GNSS coverage	Yes	Time limited	Time limited
Initial position at startup	Currently only after first balise passage	Yes, where GNSS is available	Yes, where GNSS is available
Accuracy	Depending largely on the track map	Depending on the quality of IMU and time of GNSS outage	Depending on the quality of IMU and time of GNSS outage

Table 5: Comparison of Ground Truths

5.2   ACCURACY

A detailed analysis of the accuracy of each ground truth could not be performed in the frame of the CLUG project. It is however assumed that significant inaccuracies can be identified in the frame of the data analysis.