D4.1 Description of methodology for data record sorting and saving

Project acronym:	STARS
Project full title:	Satellite Technology for Advanced Railway Signalling
EC Contract No.:	(H2020) 687414

Version of the document:	05
Protocol code:	STR-WP4-D-AZD-053-03
Responsible partner:	AZD
Reviewing status:	Issued
Delivery date:	27/02/17
Dissemination level:	PUBLIC

European Global Navigation Satellite Systems Agency

This project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No. 687414

CHANGE RECORDS

Version	Date	Changes	Authors
01	08.12.16	First draft	L. Bazant (AZD), H. Mocek (AZD), P. Kacmarik (AZD)
02	30.12.16	Sections from responsible partners were merged in this document.	L. Bazant (AZD), H. Mocek (AZD), P. Kacmarik (AZD), M. Bohac (AZD), O. Straka (ZCU), Adriano Eliseo (ASTS), D. Spiegel (TU BS), A. Geffert (TU BS), J. Marais (IFSTTAR), T. Klaiq (BT), V. Giener (INECO), A. Pazos (SIE), J. Poumailloux (TAS-F)
03	27.01.17	Correction according to decisison described in review sheet of D4.1, completion of section concernig RF measurements	L. Bazant (AZD), P. Kačmařík (AZD)
04	14.02.17	Correction after TMT review process	L. Bazant (AZD)
05	27.02.17	Final version after TMT approval	A. Toma (DAPP)

TABLE OF CONTENTS

LIST OF FIGURES

1 INTRODUCTION

1.1 EXECUTIVE SUMMARY

The purpose of this document is to present proposed methodology for recorded data sorting and recommendation for data storage. This is related to the Task 4.1 defined at the WP4 project planning [1].

Data is expected to be progressively collected during the measurement campaign in WP3. The objective of Task 4.1 is to provide summary of tools, methods and algorithms which can be used for data post-processing in Task 4.2 and subsequently for data analysis resulting in railway environment characterisation in Task 4.3.

Section 2.1 identifies possible techniques for GT estimation and validation which include the design/ description of methods proposed in WP2 for synchronization of individual sensors/ devices used for GT. Section 2.2 describes methods for GT transformation into the antenna reference position. Section 2.3 deals with recommendations of SW tools, methods and algorithms to be used for validation, merging, sorting, clearing and post-processing of field data before putting them into reference dataset as stated in Task 4.2. Section 2.4 gives an overview of SW tools and methods suitable for analysis and evaluation of data from reference dataset according to Task 4.3 – Railway environment characterisation. Section 2.5 shows proposal of synchronization methods for the measured data that are not part of the GT calculation. Next sections depict the database concept proposal and assessment of data storage capacity requirement for the database. Finally, a description of identification of the data storage server for the database is presented.

1.2 DEFINITIONS AND ACRONYMS

Acronym	Meaning
1-D	One Dimensional
3-D	Three Dimensional position
ADC	Analog-toDigital Convertor
AGC	Automatic Gain Control
ATR	Antenna of Transponder Reader
ARP	Absolute Reference Point
ASTS	Ansaldo STS
AVV	Automatic train control
AZD	AZD Praha s.r.o.
BIPM	Bureau International des Poids et Mesures
BSD	Berkeley Software Distribution (a kind of permissive free software licence)
BTM	Balise Transmission Module
CA	Consortium Agreement

CELENEC	European Committee for Electrotechnical Standardisation (Comité Européen de Normalisation Electrotechnique)
CEN	European Committee for Standardisation
CORS	Continuously Operating Reference Stations
CS	Coordinate System
CW	Continuous Wave
DCF77	German Longwave Time Signal
DOP	Dilution of precision
EC	European Commission
ECEF XYZ	x, y, z coordinates in Earth Centred Earth Fixed frame
EDAS	EGNOS Data Access Service
EGNOS	European Geostationary Navigation Overlay Service
EKF	Extended Kalman Filter
ENU	East North Up
EOA	End of Authority
ERTMS	European Rail Traffic Management System
ESA	European Space Agency
ETCS	European Train Control System
EVC	European Vital Computer
FD	Fault Detection
FDE	Fault Detection and Exclusion
GA	Grant Agreement
GBAS	Ground Based Augmentation System
GLONASS	Russian Global Navigation Satellite System
GNSS	Global Navigation Satellite System
GPS	Global Positioning System
GPST	GPS Time

GSSF	Galileo System Simulation Facility
GST	Galileo System Time
GT	Ground Truth
IERS	International Earth Rotation and Reference Frames Service
IGS	International GNSS Service
IMU	Inertial Measurement Unit
IGN	Institut Geographique National
I/Q	In-Phase, Quadrature (Used to denote the complex format on which the RF data is stored)
LAT	Latitude
LDS	Location Determination System
LEU	Lineside Electronic Unit
LHCP	Left Hand Circular Polarization
LON	Longitude
LOS	Line-of-Sight
LRBG	Last Relevant Balise Group
MA	Movement Authority
MET	Meteorological data file designation
MMA	Map Matching Algorithm
MIB	Magnetic Identification Balise
NAV	Navigation Message or Navigation data file designation
NLOS	Non-Line-of-Sight
NTP	Network Time Protocol
OBS	Observation data file designation
OBU	On Board Unit
PMO	Project Management Office
PPK	Post Processing Kinematic
PPP	Precise Point Positioning

PPS	Pulse Per Second
PPSDK	Septentrio Post Processing Suite Development Kit
PRC	Pseudo Range Correction
PTB	The National Metrology Institute of Germany
PTP	Precision Time Protocol
PVT	Position Velocity Time
QM	Quality Manager
QZSS	Quasi-Zenith Satellite System
RF	Radio Frequency
RBC	Radio Block Center
RFID	Radio Frequency Identification
RHCP	Right Hand Circular Polarisation
RMS	Root Mean Square
RPS	Record Playback System
RTP	Real-time Transfer Protocol
SA	Selective Availability
SC	Steering Committee (SC)
SBAS	Satellite Based Augmentation System
SBB	Swiss Fedaral Railways (Schweizerische Bundesbahnen)
SBF	Septentrio Binary Format
SIE	Siemens
SIL	Safety Integrity Level
SSE	Error Sum of Squares
STARS	Satellite Technology for Advanced Railway Signalling
SV	Space Vehicle
TAI	International Atomic Time (Temps Atomique International)
TALS	Trackside LDS Server

TAM	Track Axis Map
THR	Tolerable Hazard Rate
TMT	Technical Management Team
TSI	Technical Specification for Interoperability
TU BS	Technische Universität Braunschweig
UBX	Ublox Binary Format
USNO	United States Naval Observatory
UT	Universal Time (UT0, UT1, UT2)
UTC	Coordinated Universal Time
UTC(PTB)	Realization of UTC by PTB
UTM	Universal Transverse Mercator
VB	Virtual Balise
VBR	Virtual Balise Reader
WGS	World Geodetic System
WIG	Wheel Impuls Generator
WP	Work Package
WPL	Work Package Leader (WP Leader)
WT	Wheel Tachometer

2 METHODOLOGY FOR RECORDED DATA SORTING, DATA STORAGE AND REFERENCE POSITION DETERMINATION

This section gives an overview of methods and recommendations that should be implemented on the measured data coming from measurement campaign performed in WP3. Moreover, the first two sections include a description of methods used for satellite navigation independent determination of the reference position of the GNSS antenna installed on the railway vehicle.

2.1 GROUND TRUTH

The purpose of the independence of GT on GNSS was widely described in documents of WP2. Different sensors and components of the systems used by all three partners (AZD, ASTS and SIE) for GT determination within measurement campaign were also presented in these documents.

In this section, the processing methods proposed for GT determination by each of the partners responsible for measurement campaign are outlined.

GT position is obtained as result of travelled distance in given direction from absolute reference point situated on the track, known coordinates of the absolute reference points and 3-D map of track axis.

In section 2.1.1, the processing methods for Ground Truth (GT) computation are proposed by TU BS. Next sections describe GT data processing proposed by each partner responsible for the measurement campaign (AZD, ASTS, and SIE).

2.1.1 Methods for GT data processing

Information on the format of track data, accuracy and probability of track axis data survey serves as input for the implementation of GT processing methods. Furthermore, the description of the sensor platform for GT measured odometry data and the data format are provided as well.

The first step needed in order to fulfill the requirements for GT computation is the establishment of the available sensors. The list is provided by [2]. The list shows the different sensors that provide the measurement of the vehicle dynamics. Two types of sensors are available. There are sensors which provide absolute and others which provide relative information. Among the described sensors in [2], it is noticeable that the SBB railway vehicle RBDe 560 possesses two relative velocity sensors based on different measurement principles, the wheel tachometer and the radar sensor. Nevertheless, the velocity of each sensor is inaccurate due to measurement noise. A granted method to cope with this situation is to merge the redundant velocity data of the two sensors in a sensor fusion filter. Another advantage of using a fusion filter is the possibility of using the position data provided by the balise in cooperation with the velocity sensors.

Based on those criteria, a suitable fusion filter will be presented in the following sections.

FILTER FOR DATA FUSION

Independent from sensors or the measurement principle, no 100% precise measurements of the system states exists. The noise in the sensors introduces some deviations (often described as errors) to the measurements and the result is less accurate and less trustworthy. Another source of deviations has its origin in the system description. Unexpected and hardly definable disturbances, such as sudden acceleration changes due to wind or track irregularities, may cause a less trustworthy result as well. This kind of disturbance is called process noise. Both described distortions have a negative influence on accuracy and are mostly represented by stochastic processes. Despite the mentioned difficulties, it is possible to achieve trustworthy results by implementing a suitable filter. There exist many different filters to approach the required solution, but the most common one is the Kalman filter.

The Kalman filter not only allows the observation of unmeasurable states, it also provides the expected estimated value of system states under the influence of stochastic disturbances. Further information on the Kalman filter theory is described in [3]. Depending on the system behaviour, there exist many variants of the Kalman filter [4]. In order to select the required filter, it is necessary to describe the motion of the train. In the following, the required equations to describe the train motion will be presented and based on that, a suitable filter will be selected.

Dynamic model

The required model must be able to describe the motion of the train during straight trips and curves as well. Thus it is a model that requires working in both scenarios.

One approach to describe the vehicle is documented in [5]. The equations of motion have to be adjusted in order to describe the motion of the train correctly. Figure 1 shows the dynamics of the train when performing a turn into the right direction.

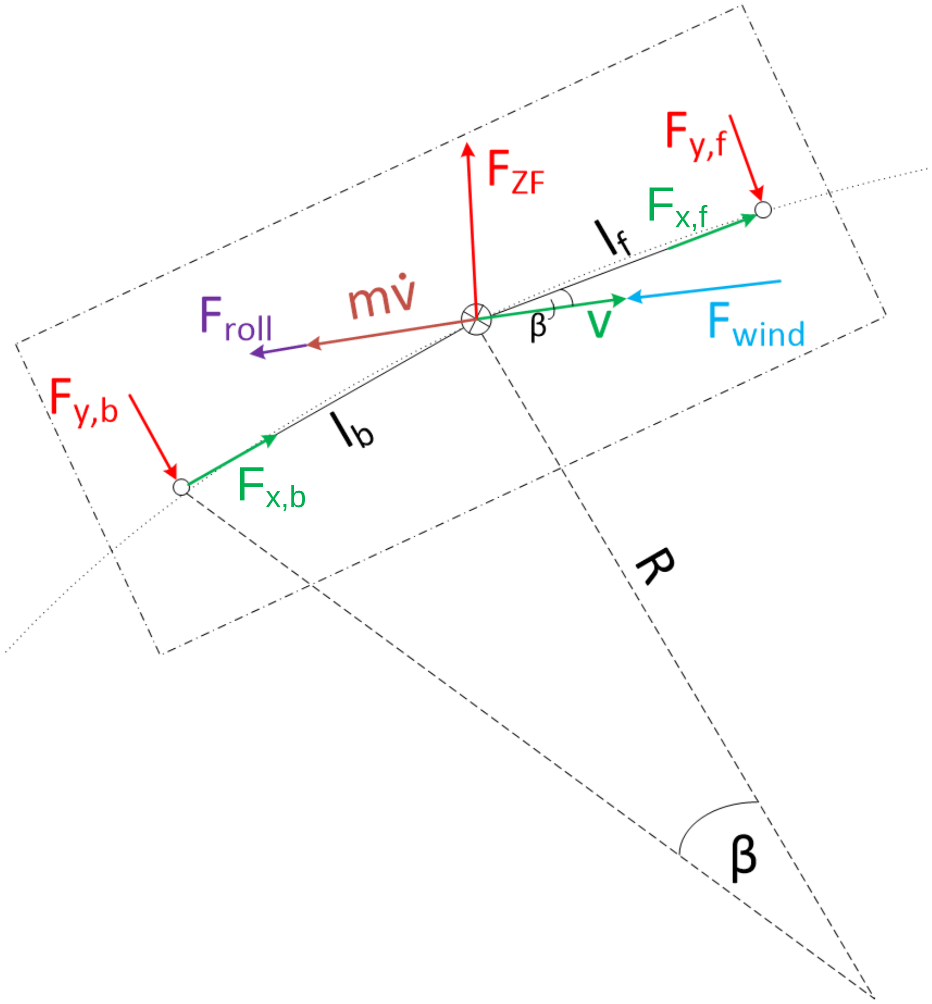


Figure 1: Dynamics of train in right curve

The definition of the utilized forces can be looked up in [5]. The assumption of a small and constant side slip angle β allows the simplification of the presented equations:

$$0 \leq \beta \leq 1$$
$$\sin(\beta) \approx \beta$$
$$\cos(\beta) \approx 1$$

By establishing equilibrium equations and by solving these equations respectively to their derivative of highest order, one gets the following:

$$\dot{v} = \frac{1}{m}\left(-F_{roll} - F_{wind} + F_{x,b} + F_{x,f} + F_{y,b}\beta + F_{y,f}\beta\right) \tag{1.1}$$

$$\dot{\psi} = \frac{1}{mv}\left(-\beta\left(F_{x,b} + F_{x,f}\right) + F_{y,b} + F_{y,f}\right) \tag{1.2}$$

$$\ddot{\psi} = \frac{1}{J_T}\left(F_{y,f}l_f - F_{y,b}l_b\right) \tag{1.3}$$

Here ψ is the yaw angle (cf. Figure 5), v is the velocity, F_{wind} is a force due to air friction/wind (making the model nonlinear) [6], F_{roll} is the rolling resistance, $F_{x,b}$ and $F_{x,f}$ are the drive forces due to the input momentum M_{in} (assumption of equal distribution in both axes):

$$F_{x,b} = F_{x,f} = \frac{1}{2}\frac{M_{in}}{r_{wheel}} \tag{1.4}$$

$F_{y,b}$ and $F_{y,f}$ are the forces of the rails holding the wheels on track, m is the mass of the train, r_{wheel} is the radius of the wheel, and J_T is the momentum of inertia.

For this purpose, the differential equations of the system will be transformed into state space in order to apply the Kalman filter algorithm.

STATE SPACE MODEL

The state space model describes a differential equation of n^{th} order with n differential equations of first order. Firstly, an adequate state vector (containing all required variables that completely describe the system) has to be defined. The following state vector has been chosen:

$$\underline{x} = \left(x, y, \psi, \dot{\psi}, v\right)^T = \left(x_1, x_2, x_3, x_4, x_5\right)^T$$

The momentum on the wheels u = M_{in} is the input of the system.

With the differential equations for position x and y – described in [5] and in eqn. (1.1) to (1.3) – the state space model is:

$$\dot{\underline{x}} = \begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \\ \dot{x}_3 \\ \dot{x}_4 \\ \dot{x}_5 \end{bmatrix} = \begin{bmatrix} x_5 \cos\left(x_3 + \beta\right) \\ x_5 \sin\left(x_3 + \beta\right) \\ \frac{1}{mx_5}\left[F_{y,b} + F_{y,f} - \frac{\beta}{r_{wheel}}u\right] \\ \frac{1}{J_T}\left(F_{y,f}l_f - F_{y,b}l_b\right) \\ \frac{1}{m}\left(-F_{roll} - F_{wind} + \beta\left(F_{y,b} + F_{y,f}\right) + \frac{1}{r_{wheel}}u\right) \end{bmatrix} \tag{1.5}$$

A simple look reveals the nonlinearity of the model. The system with a movement in two axes and F_{wind} [6] is considered as nonlinear in velocity. The model that describes the motion of the train is a nonlinear model that requires the implementation of nonlinear filters, such as the Extended Kalman Filter (EKF) presented below. The mentioned values in [7] serve for specification of the forces $F_{y,b}$ and $F_{y,f}$.

In straight lines, the equation $F_{y,b} = F_{y,f} = \beta = 0$ holds and as a consequence, the equations for $\dot{x}_3 = \dot{x}_4 = 0$ hold, which means that the yaw angle ψ (cf. Figure 5) remains constant. For curves into other direction, the sign of $F_{y,b}$ and $F_{y,f}$ changes. In order to implement this state space model in the Extended Kalman filter, a model discretization has to be performed. Applying the differential quotient to approximate the time derivations (with time step Δt), the state space model can be transformed into:

$$\underline{x}_{k+1} = \begin{bmatrix} x_{1(k+1)} \\ x_{2(k+1)} \\ x_{3(k+1)} \\ x_{4(k+1)} \\ x_{5(k+1)} \end{bmatrix} = \begin{bmatrix} x_{1(k)} + x_{5(k)} \cos\left(x_{3(k)} + \beta\right)\Delta t \\ x_{2(k)} + x_{5(k)} \sin\left(x_{3(k)} + \beta\right)\Delta t \\ x_{3(k)} + \dfrac{1}{m x_{5(k)}} \left[F_{y,b} + F_{y,f} - \dfrac{\beta}{r_{wheel}} u_k \right] \Delta t \\ x_{4(k)} + \dfrac{1}{J_T} \left(F_{y,f} l_f - F_{y,b} l_b \right) \Delta t \\ x_{5(k)} + \dfrac{1}{m} \left(-F_{roll} - F_{wind} + \beta\left(F_{y,b} + F_{y,f}\right) + \dfrac{1}{r_{wheel}} u \right) \Delta t \end{bmatrix} \tag{1.6}$$

For this approach, the frequency of the EKF has to be synchronized with the sampling frequency of the available sensors. It is also possible to use non-equidistant sampling periods of irregular measurements without having to adjust the EKF frequency to the sensor sampling frequency (e.g. pre-/retrodiction of measurement data to the EKF frequency, using buffers).

Extended Kalman Filter

The EKF is the derivation of the standard Kalman filter for nonlinear models. The detailed derivation of the EKF is presented in [4]. Figure 2 shows the basic algorithms of the EKF.

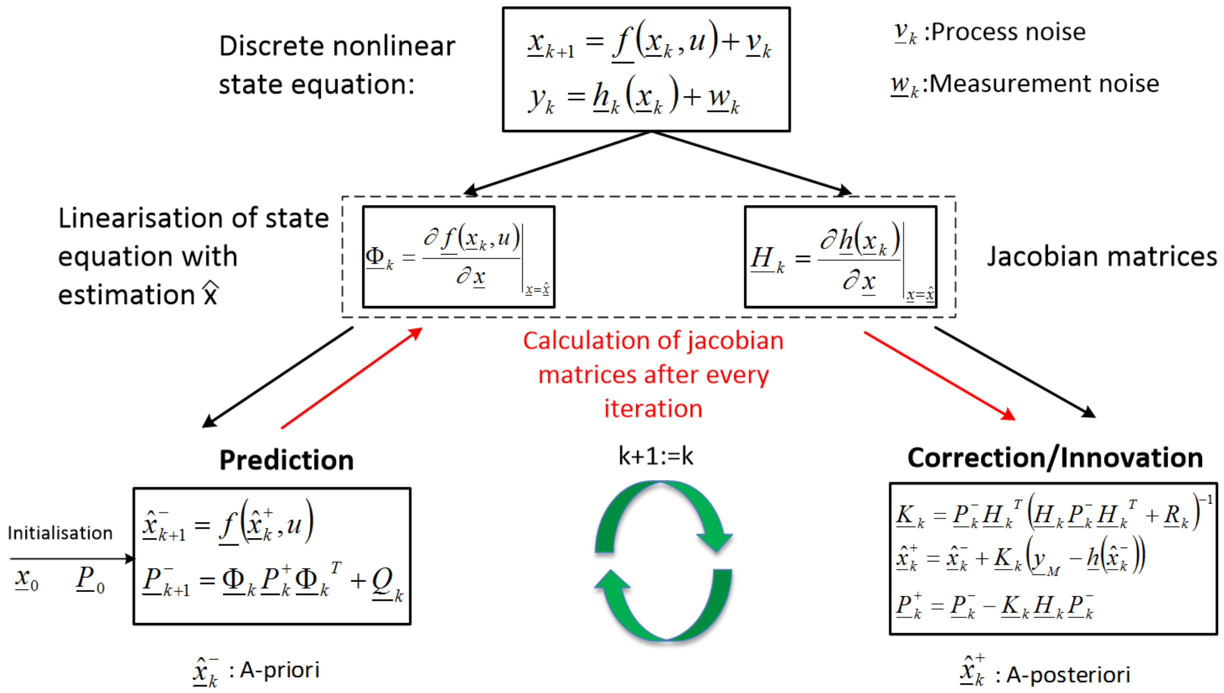


Figure 2: Algorithms of EKF

In the *prediction* step, with the knowledge of the system behaviour, the estimation of the system states takes place. In order to initialise the filter, first assumptions for the initial conditions are required.

In the first equation of the *correction* step, the Kalman Gain K is calculated. With this value, the correction of the estimation (having been done in the prediction step) is realised. The Kalman gain makes a statement of trust based on the covariance matrices of estimation and measurement. If the uncertainty of the prediction is larger than the uncertainty of the measurement, the filter will give a higher weight to the measurement value. The opposite occurs if the uncertainty of the prediction is smaller than the uncertainty of the measurement. This correction is realised in the second equation of the *correction* step. The product of Kalman gain with the difference of measured $\underline{y}_M$ and predicted $\underline{\hat{x}}_k^-$ values realizes the correction. Lastly, a new covariance matrix for the corrected estimation uncertainty is calculated.

Further analysis of nonlinear behaviour should show whether the selected EKF provides exact solutions. It has been proven that the Sigma-Point-Kalman-Filter can achieve better results for nonlinear models under respective effort [8].

DATA FUSION IN EKF

In this section, the methods of data fusion will be discussed. The available data for fusion consists of the two velocity sensors and the position measurement with the balise. In spite of the discrete information of the position provided by the balises (distribution of balise reference points along the track), the measurement at the absolute position has a high accuracy ([2]: +/- 1*m*).

Every sensor has a certain inaccuracy, which can be described by measurement noise. Assuming the measurement noise is a white noise with a normal distribution, the inaccuracy can be described by the standard deviation σ or its variance σ^2.

For the EKF algorithm, the covariance matrix R for the measurement uncertainty has to be defined. This matrix contains the variances of the measurements and describes the measurement noise. The covariance matrix of the process noise Q contains the variance of the states caused by stochastic disturbances in the system behaviour. The matrix Q contains in its main diagonal the variances of the respective state. With the assumption of noise independence, all other elements can be set to 0.

Fusion methods for velocity sensors

To fuse the velocity data provided by the two different sensors, several approaches exist. In [9], two different methods are discussed. The first method consists of merging the two velocity data sets by increasing the observation matrix. That means that the fusion is realised in the filter itself. The second method consists of merging the two measurements before entering the EKF. This second method obtains the fused measurement information by weighted observation, which means that the multisensor data is combined based on a minimum-square-error estimation. The observation matrix remains unchanged when using the second method. In [9], it is also shown that both methods are functionally equivalent if the observation matrices of the sensors are identical. Since both sensors provide velocity data, it can be assumed that the observation matrices of the sensors are equal. Based on that fact, the second method is chosen and the equations presented:

$$y_{fusion} = \left[\sum_{j=1}^{N} R_j^{-1} \right]^{-1} \sum_{j=1}^{N} R_j^{-1} y_j \tag{1.7}$$

$$\underline{h}_{fusion} = \left[\sum_{j=1}^{N} R_j^{-1} \right]^{-1} \sum_{j=1}^{N} R_j^{-1} \underline{h}_j \tag{1.8}$$

$$R_{fusion} = \left[\sum_{j=1}^{N} R_j^{-1}\right]^{-1} \tag{1.9}$$

A simple calculation of the new observation vector based on the respective observation vector of the sensors shows the following outcome. For

$$y_1 = \underbrace{\begin{bmatrix} 0 & 0 & 0 & 0 & 1 \end{bmatrix}}_{\underline{h}_1} \begin{bmatrix} x \\ y \\ \psi \\ \dot{\psi} \\ v \end{bmatrix} + w_1$$

$$y_2 = \underbrace{\begin{bmatrix} 0 & 0 & 0 & 0 & 1 \end{bmatrix}}_{\underline{h}_2} \begin{bmatrix} x \\ y \\ \psi \\ \dot{\psi} \\ v \end{bmatrix} + w_2$$

the merged observation vector is $\underline{h}_{fusion}$:

$$\underline{h}_{fusion} = \left[\sum_{j=1}^{N} R_j^{-1}\right]^{-1} \sum_{j=1}^{N} R_j^{-1} \underline{h}_j = \left[\frac{R_1 R_2}{R_1 + R_2}\right]\left(\frac{1}{R_1}\underline{h}_1 + \frac{1}{R_2}\underline{h}_2\right)$$
$$\Leftrightarrow \underline{h}_{fusion} = \underline{h}_1 = \underline{h}_2 = \begin{bmatrix} 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

The observation vector does not change its shape after the fusion. Figure 3 shows the second fusion method for the velocity sensors.

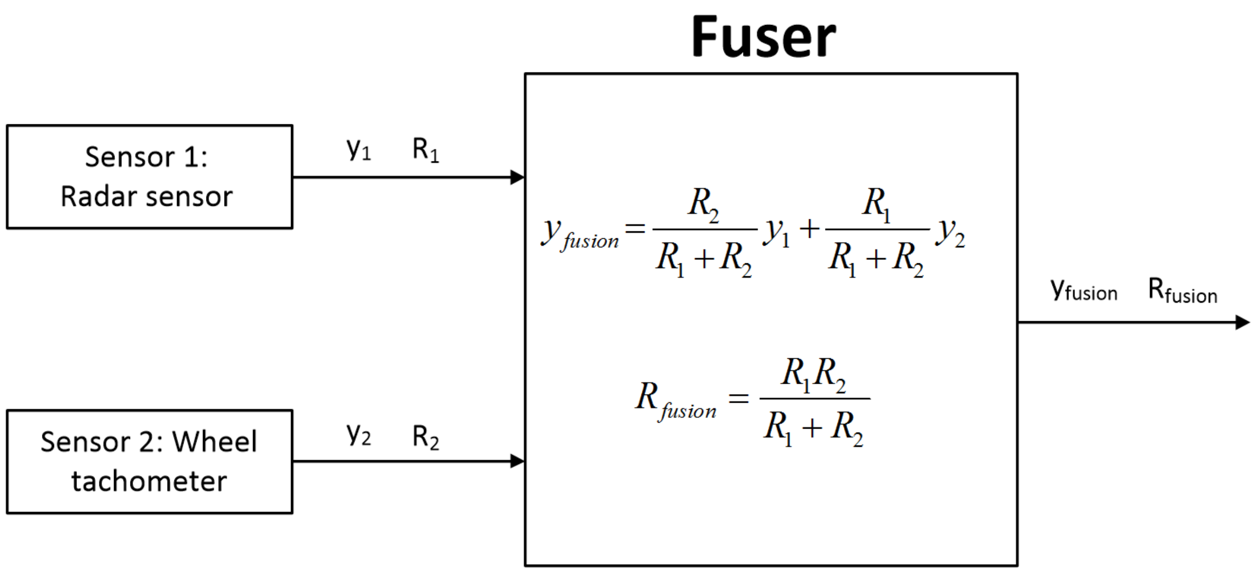


Figure 3: Measurement fusion (Method II)

The merged value y_{fusion} can now be used for the innovation step of the EKF. The merged data is always better than the data of each sensor, as noticeable in the equation 1.9.

Fusion of balise and velocity sensors

The system balise (Absolut Reference Point: ARP) and ARP reader, which is installed in the RBDe 560, provide the information of the current absolute position of the train. Despite the balises do not provide that information constantly – that is to say at discrete points on the track – the measurement is highly accurate [2]. For that reason, the balise data will be part of the data fusion. The last section showed the fusion of the velocity sensors in a previous step. In this step, the fusion of the velocity data and balise data will be implemented in the Kalman filter. Figure 4 shows the principles of the data fusion.

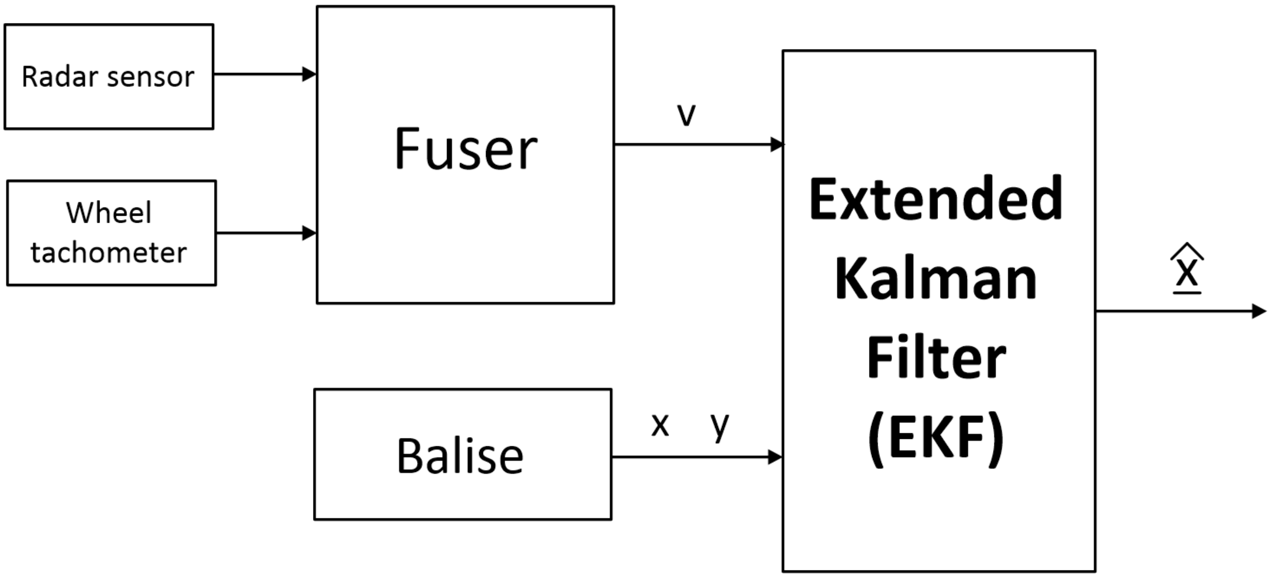


Figure 4: Data fusion of balise and fuser data

In order to implement the measurements of three sensors (two after the fusion of velocity), a new observation matrix has to be defined. The data fusion takes place in the „innovation step“, where the difference of measurement and estimation is calculated and afterwards weighted by the Kalman gain.

Fusion implementation: $\underline{K}_k \left[\underline{y}_k - \underline{H}_{fusion}\hat{\underline{x}}_k^-\right]$

Noticeable is the fact that $\underline{y}_k$ is now a row vector containing the measurement of the position (balise) and velocity (fuser). The matrix $\underline{H}_{fusion}$ consequently contains the dimensions (n x n), where n is the number of states. Based on the fact that the balise does not provide continuous position data, three different observation matrices will be defined.

Observation matrix 1:

The first observation matrix includes the case that only the balise measurement is available at a certain time

$$\underline{H}_{balise} = \underline{H}_{fusion} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

This observation matrix allows to calculate the difference between measured x and y value of the balise and the estimated states for mentioned values.

Observation matrix 2:

The second observation matrix includes the case that only the velocity measurement is available at a certain time

$$\underline{H}_{velocity} = \underline{H}_{fusion} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

This observation matrix allows calculating the difference between the measured (and merged) velocity value v of the fuser and the estimated state for mentioned value.

Observation matrix 3:

The third observation matrix includes the case that the velocity and position measurement are available at a certain time

$$\underline{H}_{velocity} = \underline{H}_{fusion} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

This observation matrix allows calculating the difference between the measured (and merged) velocity value v of the fuser and the estimated state for mentioned value. In the algorithm, there must be a flow control dealing with the choice of the correct observation matrix.

With the observation matrices presented above, it is possible to realize the data fusion. In the innovation step of the EKF, the estimated states will be refreshed with the measurement data and propagated into the next prediction step.

After the introduction to the data fusion of the available sensors, the question still remains how to detect the yaw angle during curves in order to describe or estimate the states with the EKF. In the following section, a method for angle determination is introduced by means of map matching.

MAP MATCHING

A digital map contains the information of at least one track segment. In the digital track map, all track information is stored, such as the balise positions, the positions of curves and straight lines. It also contains the data for each coordinate on the tracks and the angle of the tracks relative to an inertial coordinate system. In order to describe the motion of the train in 2-dimensional space, the definition of some coordinate systems is required. Figure 5 shows the RBDe 560 driving along a segment of a theoretical track. The coordinate systems are also shown.

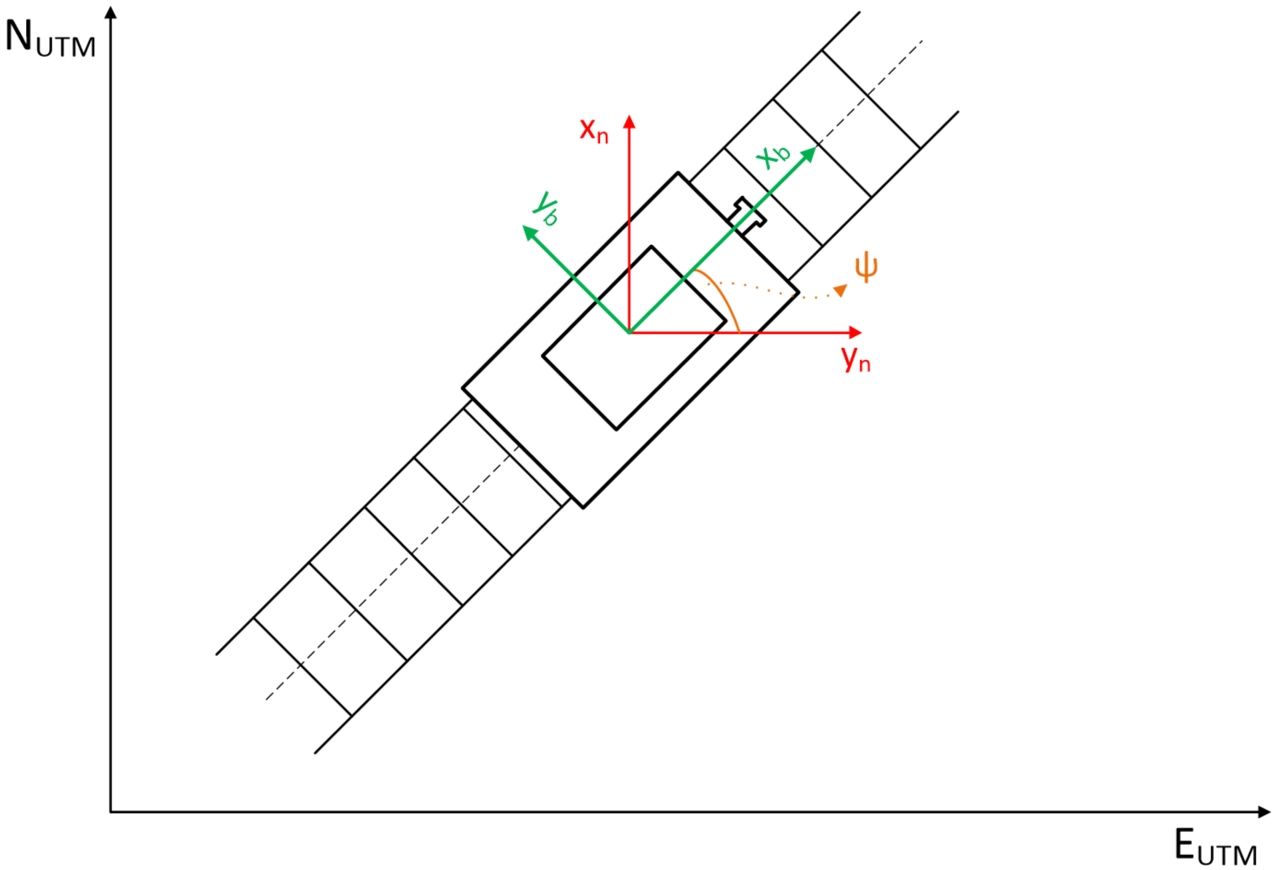


Figure 5: Definition of coordinate systems for RBDe 560

The inertial coordinate system can be for example a UTM projection of a certain place on earth. This coordinate system is 2-dimensional. Its y-axis (N_{UTM}) points to north and the x-axis (E_{UTM}) to east. In this coordinate system, the motion of the train is described. The vehicle has two different coordinate systems attached to its point of gravity. The first one is the body frame (b-frame), whose x-axis (x_b) points towards the longitudinal axis of the train. The other coordinate system is the navigation frame (n-frame), which points always in north x_n, east y_n and down (z_n not displayed) direction. Regardless of the down component in the n-frame, the UTM axes are always parallel to the n-frame axes. As a matter of fact, every yaw angle displacement of the b-frame relative to the n-frame corresponds to a displacement in the inertial coordinate system as well.

With those assumptions, it is also possible to determine the yaw angle ψ of the tracks or rather of the train. This yaw angle may be used to estimate the states of the system in the EKF, without necessarily having to make any measurement with an expensive IMU.

It remains the question of how to use the digital map to display the motion of the train. As shown in the sections before, it is only possible to measure the velocity of the train and the position with the balise. The solution can be achieved by using a suitable Map Matching Algorithm (MMA). In the following, the idea of the MMA will be described and followed by the implementation of this MMA in the EKF. Figure 6 shows the MMA applied to a straight line.

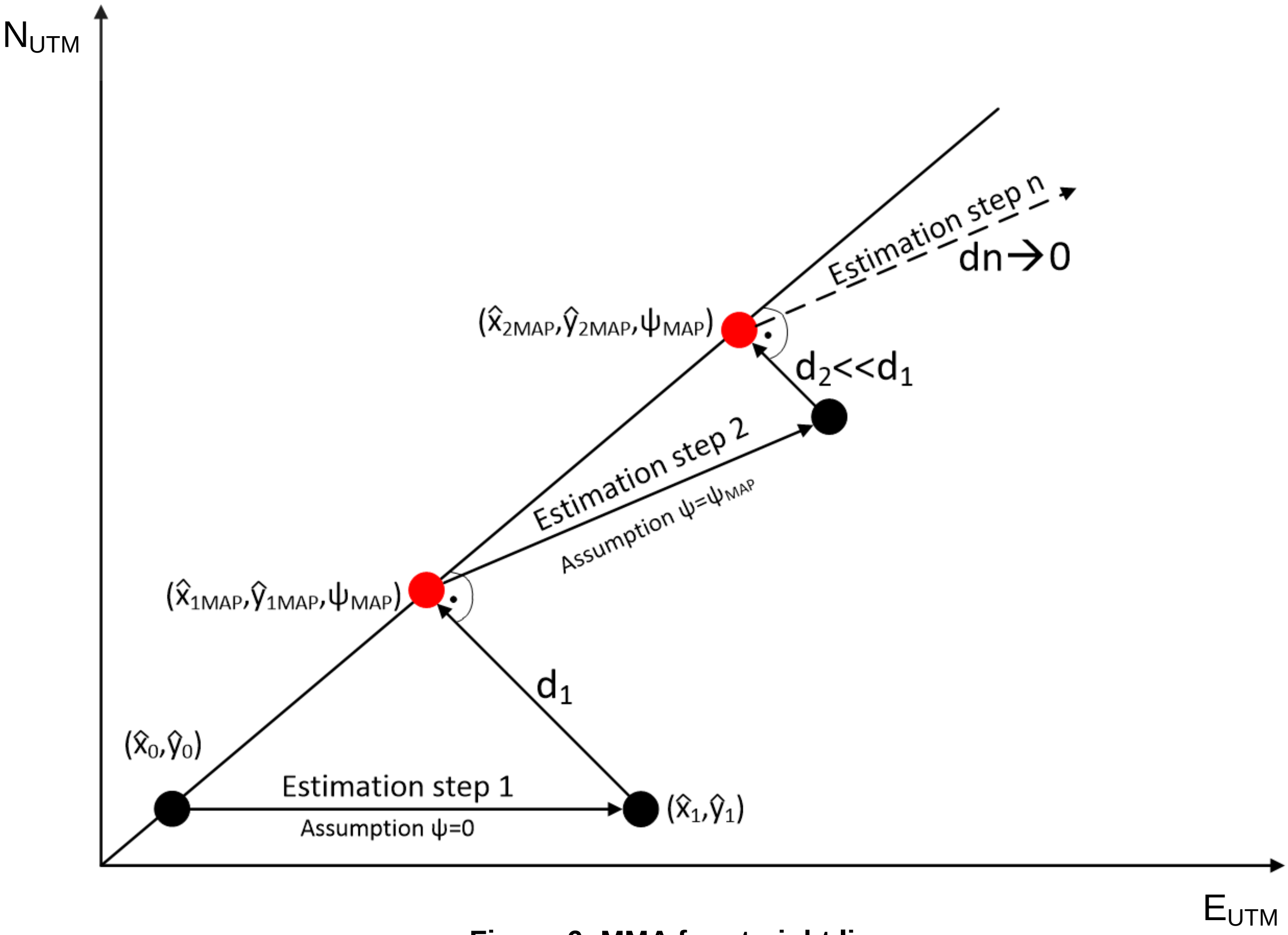


Figure 6: MMA for straight line

The MMA is based on the principles of variational calculus, that is to say the least distance between a given point and a function is searched. In this case, the given point will be the estimation of the x and y values and the function will be $f(x) = mx + n$, which describes a straight line. The MMA practically projects the estimated position $(\hat{x}_1, \hat{y}_1)$ onto a point on the track $(\hat{x}_{1MAP}, \hat{y}_{1MAP})$, where the yaw angle ψ_{MAP} is known as well. This point and angle become the new estimated values and can be used in the next propagation step. For a straight line, the yaw angle remains constant, facilitating the estimation with the EKF. In the best scenario, the distance d_n equals zero.

The MMA operates in the same way for curves as well. Figure 7 shows the application of the MMA in curves.

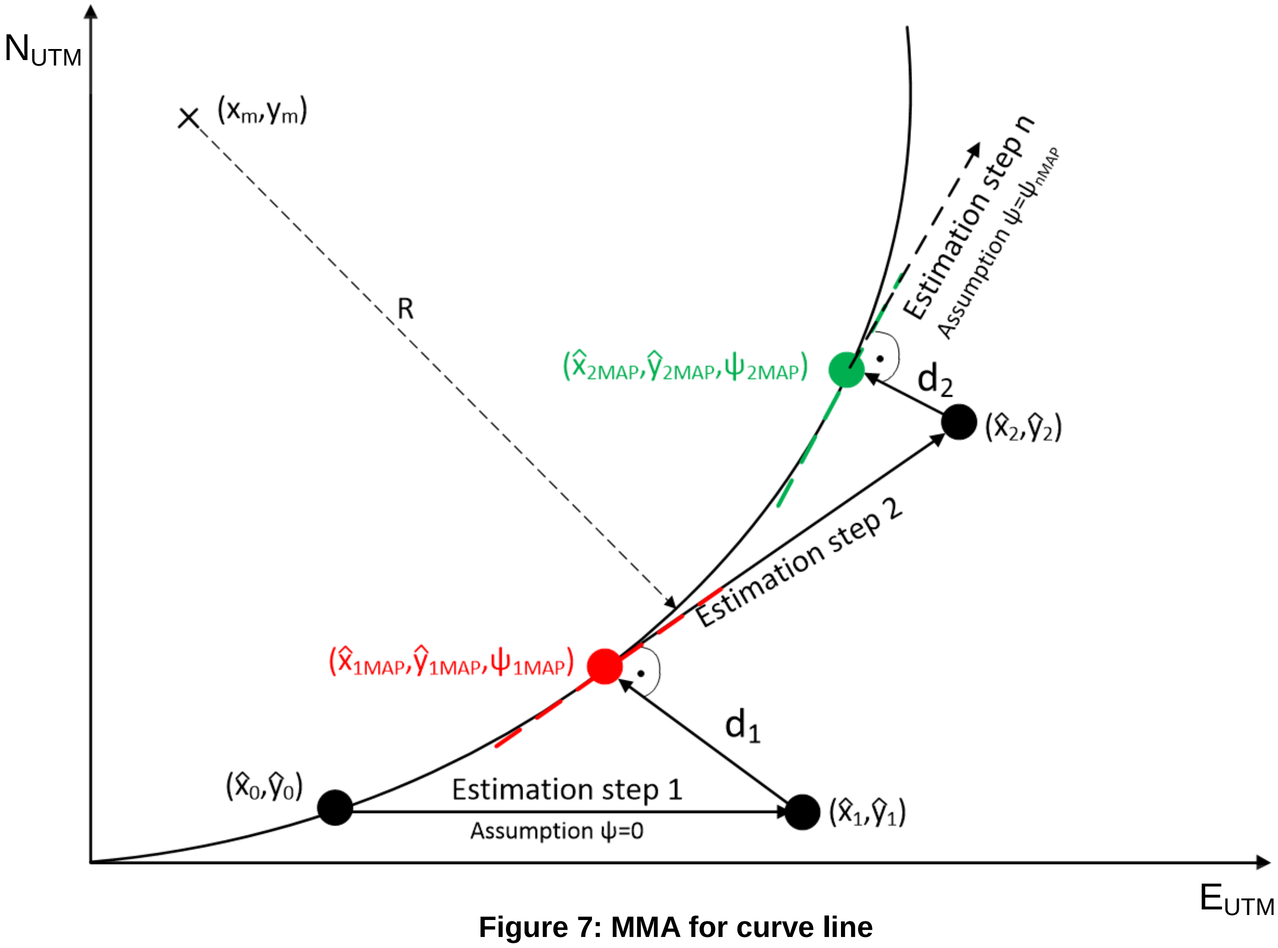


Figure 7: MMA for curve line

For a curve, the shape of a circle sector may be assumed. The function of a curve may be

$$f(x) = y_m + \sqrt{R^2 - (x - x_m)^2} .$$

In a curve, the yaw angle of the tracks varies depending on the motion. At each projected point, the current angle has to be calculated for the next propagation step. In order to minimize position deviations, the calculation frequency of the EKF has to be relatively high. In the EKF, the projected positions are taken into account for the next propagation step. Further analysis of the MMA has to be done, once real measurement data is available. If the position deviations are too vast, the use of an IMU can be helpful.

After presenting the methods of how to obtain the current position of the train based on MMA and EKF, the next step consists in transforming the results to WGS 84 coordinates and the calculation of the confidence interval. The coordinate transformation to WGS 84 is not further detailed due to its extensive availability in literature.

CONFIDENCE INTERVAL

In this section, the calculation of the confidence interval is detailed. In this case, the population mean is assumed to be unknown and the confidence interval calculation is based on the sample mean.

Before the sample mean can be calculated, the data has to be transferred by means of interpolation from equidistant values in time to equidistant values in space. Typically these have a spacing of 1 or multiple meters. Due to the slow dynamics of trains and high frequency of ground truth calculation, the uncertainty due to interpolation is negligible. The following sample mean and confidence interval calculation are based on the interpolated values.

Sample mean:

$$\bar{X} = \frac{\sum_{n=1}^{N} x_n}{n} \tag{1.10}$$

Here x_n is the value of the random variable and “n” the number of the random variable values.

Sample variance:

$$\hat{\sigma}^2 = \frac{\sum_n \left(\bar{X} - x_n\right)^2}{n-1} \tag{1.11}$$

Sample standard deviation:

$$\hat{\sigma} = \sqrt{\hat{\sigma}^2} = \sqrt{\frac{\sum_n \left(\bar{X} - x_n\right)^2}{n-1}} \tag{1.12}$$

The following assumptions are made:

- The population is assumed to be normally (Gaussian) distributed. For the sake of simplicity, deviant effects (leading to fat tails, skewness etc.) are neglected.
- Measurement data of the same track route is taken multiple times (several trips). Thus mean and variance can be calculated for each position according to the formulae given above.
- The confidence interval will describe the distribution of mean position values. With the assumptions made above and due to the fact that the variance is estimated based on the samples, the mean position has to be described with Student’s t-distribution. For growing amount of samples n, the t-distribution converges to the normal distribution.

Based on the sample mean and the sample standard deviation, limits of the confidence interval can be calculated. The equation for the confidence interval follows [8].

Confidence interval:

$$\bar{X} - t_{\left(1-\frac{\alpha}{2};n-1\right)} \frac{\hat{\sigma}}{\sqrt{n}} \le \mu \le \bar{X} + t_{\left(1-\frac{\alpha}{2};n-1\right)} \frac{\hat{\sigma}}{\sqrt{n}} \tag{1.13}$$

Here t represents the value for a given confidence level, which can be looked up in the literature about t-distributions. The confidence level is $\gamma = 1 - \alpha$ and has to be defined. For instance for a confidence level (confidence niveau) of 95%, the relation

$$\gamma = 1 - \alpha = 95\% = 0{,}95$$
$$\Rightarrow \alpha = 0{,}05$$

can be determined. Here α is the error probability that the population mean does not fall in the limits of the confidence interval.

With given x_n and confidence level, it is also possible to calculate the confidence interval. In order to apply these equations to calculate the confidence interval of the position estimation made by the EKF, many measurements are required. This means that the measurement data of many trips has to be gathered and post processed in the EKF. In this way, it is possible to calculate the confidence interval for each position estimation. Figure 8 shows the application of these equations.

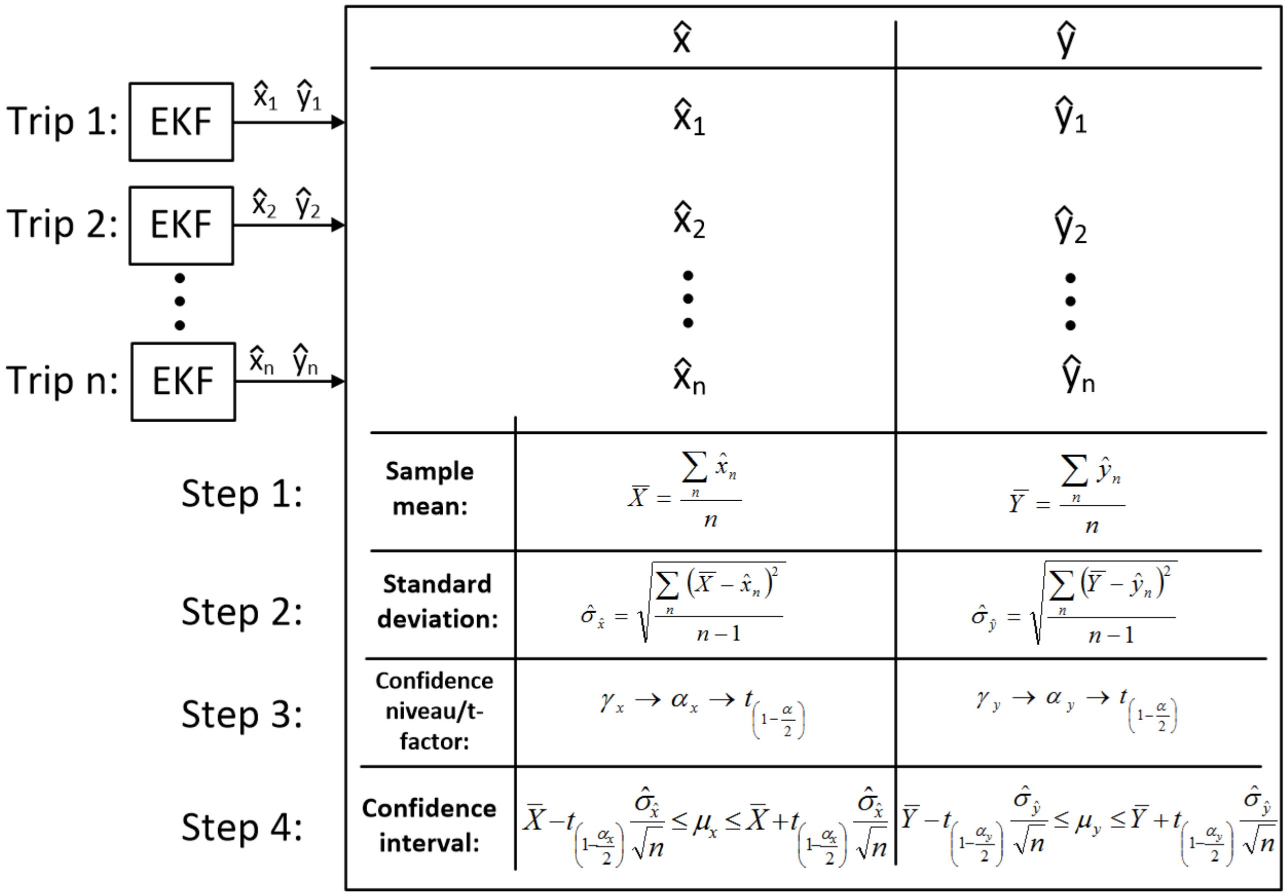


Figure 8: Confidence interval calculation

Finally, the question remains of how many trips are required in order to obtain the correct length of the confidence interval. In [9] the least number n is described and calculated by

$$n \geq \frac{4t^2\hat{\sigma}^2}{L^2} \tag{1.14}$$

Here, the same parameters are used as for the confidence interval described above. The new parameter L describes the desired length of the confidence interval.

2.1.2 GT data processing from measurement campaign in the Czech Republic

The computation of the GT and timestamped GNSS antenna reference positions will make use of the following coordinate systems (CS):

- ECEF CS – a right handed Cartesian CS (x, y, z) that is fixed to the Earth.
- WGS 84 CS – a geodetic CS (latitude, longitude, height above reference ellipsoid) based on the reference ellipsoid of WGS 84 that is fixed to the Earth.
- Navigation CS – a right handed Cartesian East-North-Up (ENU) CS that defines local tangent plane.
- Track CS – a right handed Cartesian ENU CS that defines local tangent plane of the track.
- Bogie body CS – a right handed Cartesian ENU CS that defines bogies of a train.

- Vehicle body CS – a right handed Cartesian ENU CS that defines vehicle body of a train.

The GT is to be determined using the following means:

- Absolute position tags placed along a track (ETCS balises (Eurobalises), AVV MIBs, RFID tags),
- Antenna enabling tag's readings (ETCS antenna, AVV antenna, RFID antenna). This antenna is installed on the vehicle body.
- Train on-board absolute position reader (balise reader, RFID tag reader),
- Train on-board relative sensors (odometer, speed sensor, IMU),
- Map database (Track Axis Map (TAM) with superelevation).

The GT is defined as a sequence of timestamped ECEF coordinates of train representative points on the track axis map. At each time instant the train representative point is defined as the point of the track axis map that is closest to a reference point of Antenna of Transponder Reader (ATR) mounted on the train. Note that the ECEF coordinates contained in the GT determine the origins of the navigation CSs that are subsequently used to compute the reference antenna positions.

Note that AZD selected two test railway lines in the Czech Republic that differ in absolute position tags and available information [10]. In case of the test railway line 1 (Čížčenice-Volary), RFID tags with known ECEF coordinates are used. For the test railway line 2 (Česká Třebová-Brno hl.n.), MIBs with information about distances to the neighboring MIBs are used. The data sources and their processing to obtain the GT are depicted in Figure 9. Note that the illustration is for the railway line 1 (Čížčenice-Volary). For the railway line 2 with the MIBs the data processing flow chart is almost identical.

First, the RFID database containing the RFID tag ID and its position in ECEF CS is combined with the map database containing TAM and superelevation to obtain position of the RFID tags on TAM. Then, the data measured by the RFID on-board system, i.e. the ID of the tag and the time of its reading, are combined (PROCESSING OF ABSOLUTE POSITION READINGS) with the position of the RFID tags on TAM to obtain absolute positions along TAM. Subsequently, these absolute positions along TAM are combined (PROCESSING OF ABSOLUTE AND RELATIVE POSITION MEASUREMENTS) with the measurements obtained from relative sensors (IMU, odometer, Doppler radar) to obtain position of the train representative points along TAM and corresponding time. Finally, this position is transformed (POSITION TRANSFORMATION FROM TAM TO ECEF) using the map database to the ECEF CS to obtain the GT, i.e. the position of the train representative points in the ECEF CS and corresponding GPS time.

Once the GT is obtained, for the purpose of analysis of the position obtained by means of the GNSS, the GT is transformed (COMPUTATION OF GNSS ANTENNA REFERENCE POSITION) to the reference antenna position as is described in Section 2.2.2.

The detailed description of the processing blocks and data structures is given in [11].

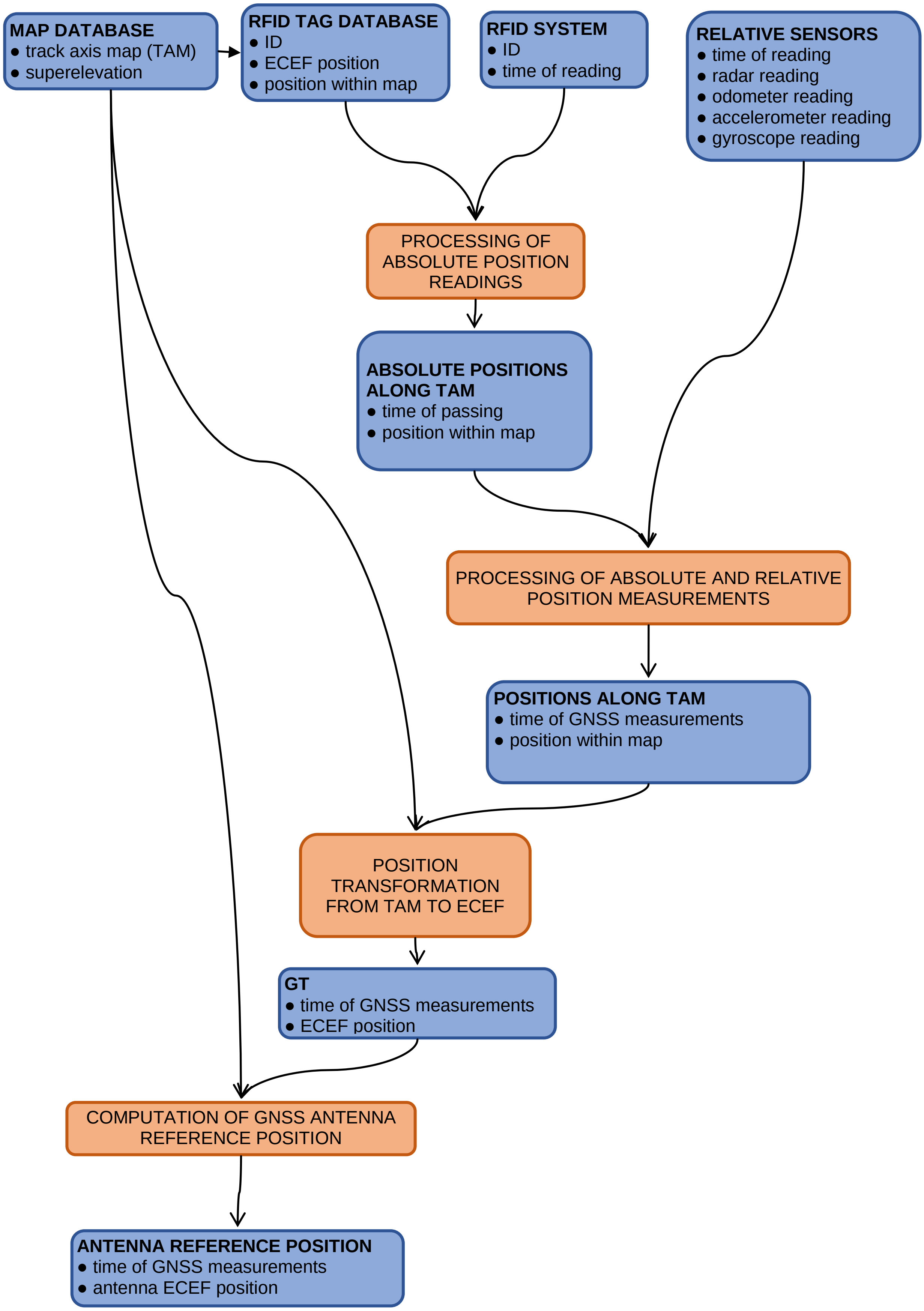


Figure 9: Data processing flow chart for GT computation

2.1.3 GT data processing from measurement campaign in Italy

METHOD FOR GT DATA PROCESSING BASED ON ETCS BALISES AND ODOMETER

Overview of the odometry and GPS systems and data preparation

The odometry system provides an accurate estimation of time, position and speed of the train running on a railroad. Taking into account the starting point of the train, odometry estimates the travelled distance meant as an offset from the starting point. Meanwhile, odometry system also measures the elapsed time (meant as a progressive monotonic time counter, usually with a high frequency, from the power-on of the train) and the speed (the first order derivative estimation of travelled distance within the time interval).

Because of the relationship between travelled distance and movement of the wheels on the rail, phenomenon such as wheels slipping or sliding could mislead the travelled distance estimation (and the related speed), so proprietary algorithms allow the detection and correction (or anyway the reduction) of misleading information.

GPS system allows the estimation of time, position and speed of any point on the earth's surface. In the special case of the train position estimation within the *Satellite-Based Localization Determination System* (hereafter LDS) application, the position (as well as the speed) has to be understood as constrained to the railway and the estimated value consists of a progressive monotonic offset from a starting point.

Therefore, with the odometry system as reference, GPS system in the LDS application provide homogeneous quantities that we could compare each other.

To properly managing the two homogeneous datasets, we must perform three steps:

1. Timescales alignment
2. Gathering of the balises detection events
3. Travelled distance linear interpolation

Once obtained the time-distance alignment, we can extrapolate, by using the Track Database, the ECEF coordinates and build up the ground truth.

Timescales alignment

GPS system and Odometry system use two different and independent timescales to measure the elapsed time.

To allowing the alignment between them, we use a common time scale that is the tick counter of the acquiring system. This proprietary real-time acquiring system located on board, collects both the interrupt coming from GPS receiver (that sends each second the PPS - Pulse Per Seconds - signal) and the interrupt coming from the Odometry system (that sends its time pulse at its working rate).

For each PPS signal (that carries the current GPS time value), by taking into account the system ticks to compensate the interrupt delays, we can estimate the correct time value (reference value) of the odometry system. Just for sake of clarity, the PPS signal is a binary signal that raises an high priority interrupt in the real-time acquiring system. This interrupt allows to "taking a snapshot" of the system status and its input data, such as odometry time and GPS time. The matching foresees a resolution of less than two milliseconds.

Because of the sampling rate of a GPS receiver that could be higher than 1 Hz (that is the nominal rate of PPS signal), the odometry time should be linearly interpolated for each GPS data sample.

Between two PPS samples we can consider negligible clock steering; therefore we can perform a linear interpolation of the odometry time without lack of precision.

For each GPS time, we can implement the following formula to perform the interpolation:

$$t_{Odometry}(GPStime) = ODO_{REF_{i-1}} + \frac{GPSTime - GPSTimePPS_{i-1}}{GPSTimePPS_i - GPSTimePPS_{i-1}} * (ODO_{REF_i} - ODO_{REF_{i-1}})$$

Where:

$GPSTimePPS_i$ is the GPS time linked to the ith PPS signal

ODO_{REF_i} is the reference odometry time for the ith $GPSTimePPS$

$i : MIN\{GPSTime < GPSTimePPS_i\} \, \forall \, GPSTimePPS$

Gathering of the balises detection events

As previously described, the odometry system estimates the distance travelled by the train with related uncertainty.

Proprietary algorithms ensure the accuracy level by managing the misleading information due to wheel slipe and slide events. Anyway, the accuracy of odometry system is at most 5% of the measured values. To keep errors low, the idea is to consider the travelled distance measurement acquired between two balises.

The acquiring system logs the balises detection events in the time scale of the odometry system as well as the travelled distance (and the speed) measurement.

The second step foresees that the balises detection events will be listed and taken into account for the third last step.

Distance linear interpolation

The timescales alignment allows to having for each record GPS (therefore for each PVT result) a correspondent odometer time. Without lack of accuracy, we can estimate the interpolated space for each interpolated odometer time.

The basic formula could be:

$$\overline{ODO}distance(\bar{t}_{Odometry}) = ODOdistance(i-1) + \frac{\bar{t}_{Odometry} - t_{Odometry_{i-1}}}{t_{Odometry_i} - t_{Odometry_{i-1}}} * (ODOdistance_i - ODOdistance_{i-1})$$

Where:

$t_{Odometry_i}$ is the ith odometer time

$\bar{t}_{Odometry}$ is the interpolated odometer time with reference the time GPS

$ODOdistance_i$ is the ith distance measured by the odometer

$i : MIN\{\bar{t}_{Odometry} < t_{Odometry_i}\} \, \forall \, t_{Odometry}$

The linear interpolation of travelled distance approximates an uniform motion of train within a time interval very short (that is the time span between two odometer time sampled at a frequency usually of about 6-20 Hz). The distance drift, with respect to the real motion of train, within the short time interval is negligible.

Data Elaboration

By knowing the exact position of balises (in term of ECEF coordinates and offset from the starting point of the track database) through the line survey campaign, and by collecting the data prepared following the guidelines of the previous section, we can elaborate the ground truth.

Ground truth format foresees a mapping between GPS time and offset from the starting point of the track database (hereafter *Mileage*) as well as the speed value and ECEF position coordinates.

The ground truth exchange file is a CSV file where seven semicolon-separated fields follow the following structure:

Field 0	Field 1	Field 2	Field 3	Field 4	Field 5	Field 6
GPSTime [ms]	Mileage [cm]	Speed [cm/s]	ID = 1	ECEF X [cm]	ECEF Y [cm]	ECEF Z [cm]

And where:

- GPS time is the time provided by the GPS receiver, expressed in milliseconds;
- Mileage is the travelled distance (meant as offset expressed in centimetres) from the starting point of railroad;
- Speed of train expressed in centimetres/seconds;
- ID is a constant fixed value (set to “1”);
- ECEF X,Y,Z are the Cartesian coordinates of a point constrained to the track which refer to the mileage value

Next subsections will explain the methodology to estimate those characteristics.

Mileage calculation

By following the previous section, the generic ith sample of GPS time has got an interpolated value of $\overline{ODO}distance_i$ as well as for each balises detection event we can identify both the *ODOdistance* measured by odometer and the *Mileage* value known by survey.

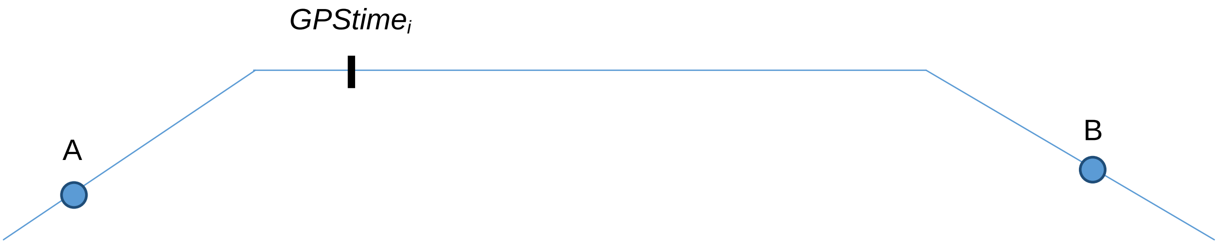


Figure 10: Mileage calculation

By following the Figure 10, we have:

- A: Balise$_A$ position;
- B: Balise$_B$ position;
- *GPStime$_i$* : ith sample of GPStime.

Furthermore:

- M$_A$: the *Mileage* value related to the Balise$_A$ position;

- M_B : the *Mileage* value related to the $Balise_B$ position;
- $ODOdistance_A$: the space value measured by the odometer when the $Balise_A$ detection event occurred;
- $ODOdistance_B$: the space value measured by the odometer when the $Balise_B$ detection event occurred;
- $\overline{ODO}distance_i$: the space value interpolated for the i^{th} sample of GPStime;
- $\overline{M}_i$: the *Mileage* value, we have to calculate, related to the i^{th} sample of GPStime.

Given the following relation:

$$\overline{M}^A{}_i = M_A + \left(\overline{ODO}distance_i - ODOdistance_A\right)$$

$$\overline{M}^B{}_i = M_B - \left(ODOdistance_B - \overline{ODO}distance_i\right)$$

If we had a perfect accuracy, we would have that $\overline{M}^A{}_i = \overline{M}^B{}_i$ and the GroundTruth could be filled by using or $\overline{M}^A{}_i$or $\overline{M}^B{}_i$ with no difference, for each GPStime sample.

However, because of the not perfect accuracy, we have to associate to each GPStime sample the result of following expression:

$$\overline{M}_i = \frac{\overline{ODO}distance_i - ODOdistance_A}{ODOdistance_B - ODOdistance_A} \cdot \overline{M}^B{}_i + \frac{ODOdistance_B - \overline{ODO}distance_i}{ODOdistance_B - ODOdistance_A} \cdot \overline{M}^A{}_i$$

That acts as a weighted mean of the calculated value of *Mileage* for each GPStime sample.

Speed Calculation

Because of the accuracy of the *Mileage* estimation, we can estimate the speed value related to each sample of GPS time by calculating the first order derivative of travelled distance estimation.

ECEF extrapolation

We can model the track database as a set of adjacent segments spatially ordered, where each segment is a couple of Cartesian coordinates in the ECEF reference frame (hereafter *Nodes*).

The cumulative geometric distance of each *Node,* represent formally the *Mileage* value of each *Node.* For each *segment* we can identify the gradient that is the unitary vector that represents the direction of the track in the space in a given point.

To convert a generic *Mileage value* in ECEF coordinates, we have to identify the segment which nodes include the given *Mileage* value. Once identified the segment, ECEF result in a geometric application of the proper offset to the track gradient.

2.1.4 GT data processing from measurement campaign in Switzerland

For the Ground Truth Calculation there are several sensors available at the Domino train, Vehicle Type RBDe 560, as shown in Figure 11 and Figure 12.

The type of installed sensors can be classified, according to the provided type of reference as:

- Absolute reference:
  - Eurobalises
  - Camera
- Relative reference:
  - Wheel Tachometers
  - Radar sensors
  - Inertial measurement unit (IMU)

- Optical correlation sensors

Geographic track data is the basis for the calculation and should be converted and supplemented as digital track data.

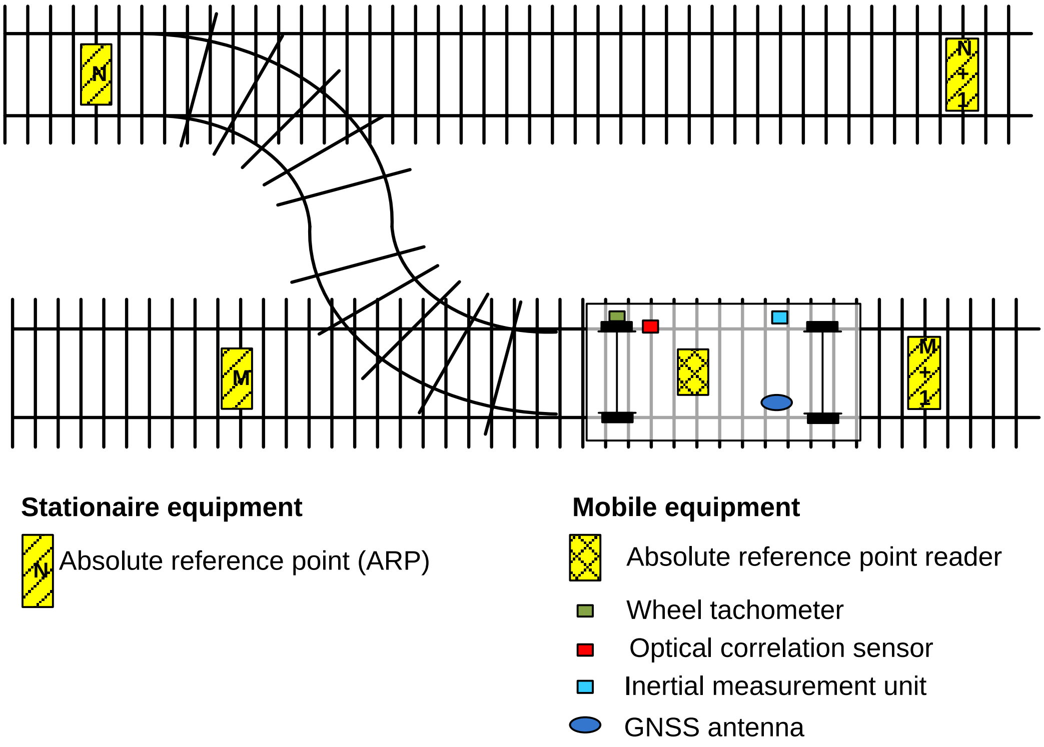


Figure 11: Available sensors at Domino train

BASIC PROCEDURE

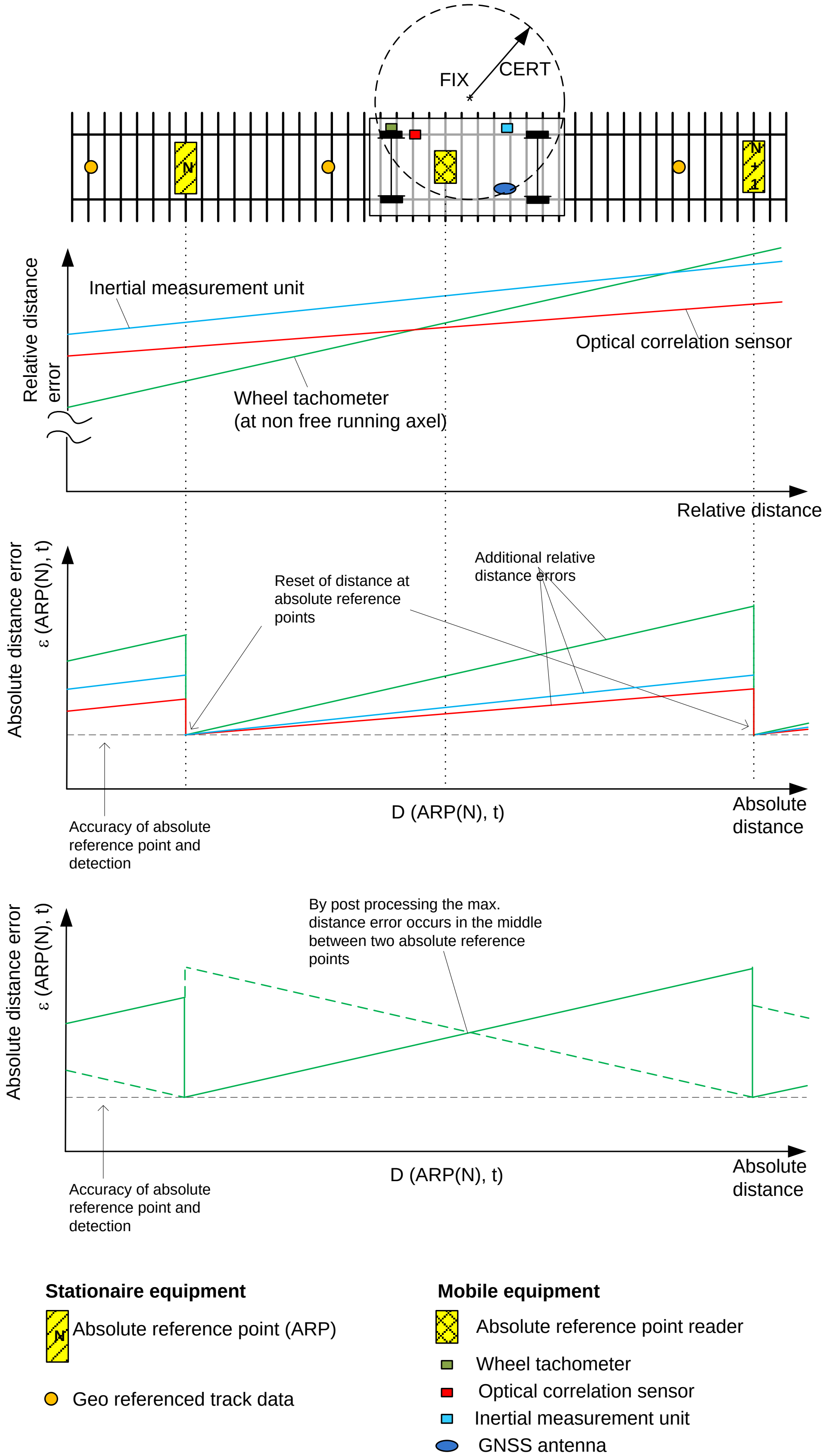


Figure 12: Basic idea of the development of the absolute distance error

The main idea is to calculate the GT depending on the given track data with the help of the absolute and relative reference points named in previous section.

The Eurobalises should be used to calibrate the relative reference points, e.g. the wheel tachometer, to get a minimal distance error. The development of the error rate over time is shown in Figure 12. In the first diagram is depicted that each sensor has a different relative error rate development but all have in common that the distance error increases with distance travelled. This problem could be improved by resetting the errors at absolute reference points to the absolute distance error, as shown in the second diagram. So the relative errors would only increase between two reference points. This could be further improved by adding post processing as outlined in the last diagram. With merging the pre and post processing results, a minimal distance error should be calculated.

With the distance and the available track data the Ground Truth should be calculated and added to the track data. In the next step of the project it can be used for the comparison with the navigation solution out of GPS data.

A brief overview of the whole procedure is shown in Figure 13 and is elucidated in Table 1.

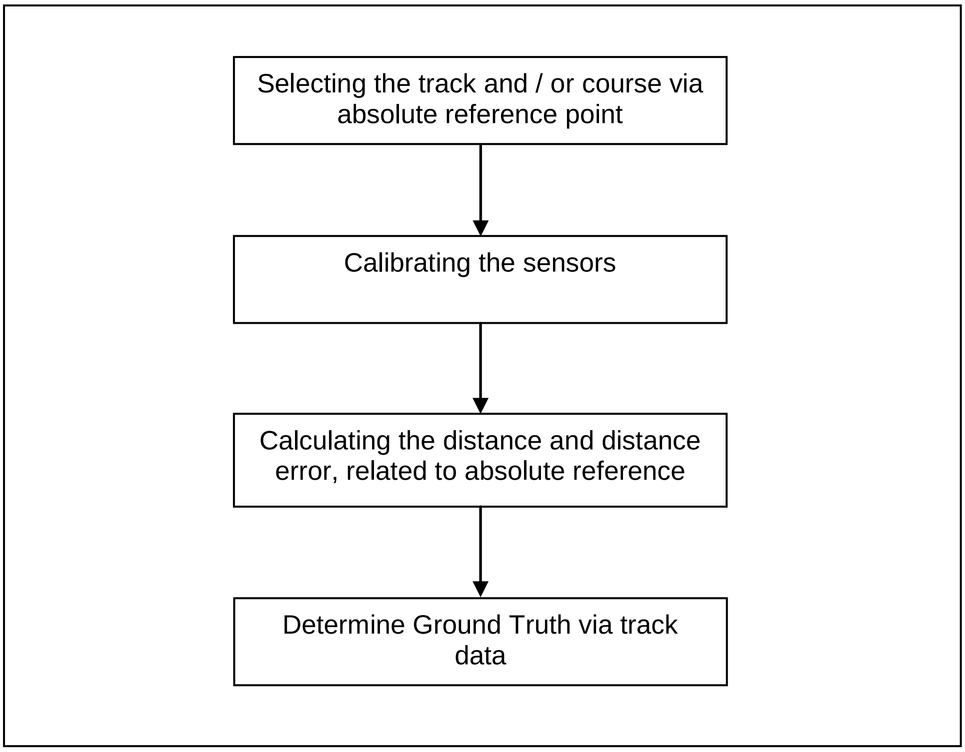


Figure 13: Basic procedure for GT determination

Work package	Description
Selecting the track and / or course via absolute reference point	The measurement system doesn't have a direct information e.g. from interlocking or disposition system on which track the train is running. It also doesn't know which course is set by the switching of points. In the measurement data the absolute reference points provide the link to the used track. And two consecutive detected absolute reference points give information about the settled course.
Calibrating of the sensors	By knowing the track and the course, the distance between the two absolute reference points can be calculated which could be used to adjust slightly the calibration of the relative distance sensors as well as for plausibility checks. Algorithms from robotics could be interesting for this issue, as they also rely on accurate (position) measurements (e.g. [13]).
Calculating the distance and distance error, related to absolute reference point	The distance measurement will take place sectional in relation to the next closest absolute reference point (D(ARP(N, t)) by means of the relative distance measurement sensors. At the absolute reference point the position errors is smallest and results from the accuracy of positioning as wells as of the detection of the absolute reference points. Beside it the sectional distance uncertainty of the relative distance sensor will be added.
Determine Ground Truth via track data	By using the track data the transformation from the sectional distance measurement related to the absolute reference point (D(ARP(N, t)) to the geo reference could be done, taking the distance error into account. D(ARP(N, t)) +/- ε (ARP(N, t)) ➔ LAT+/- ε ‘, LON +/-ε’’

Table 1: Basic procedure for GT determination

ALGORITHMS

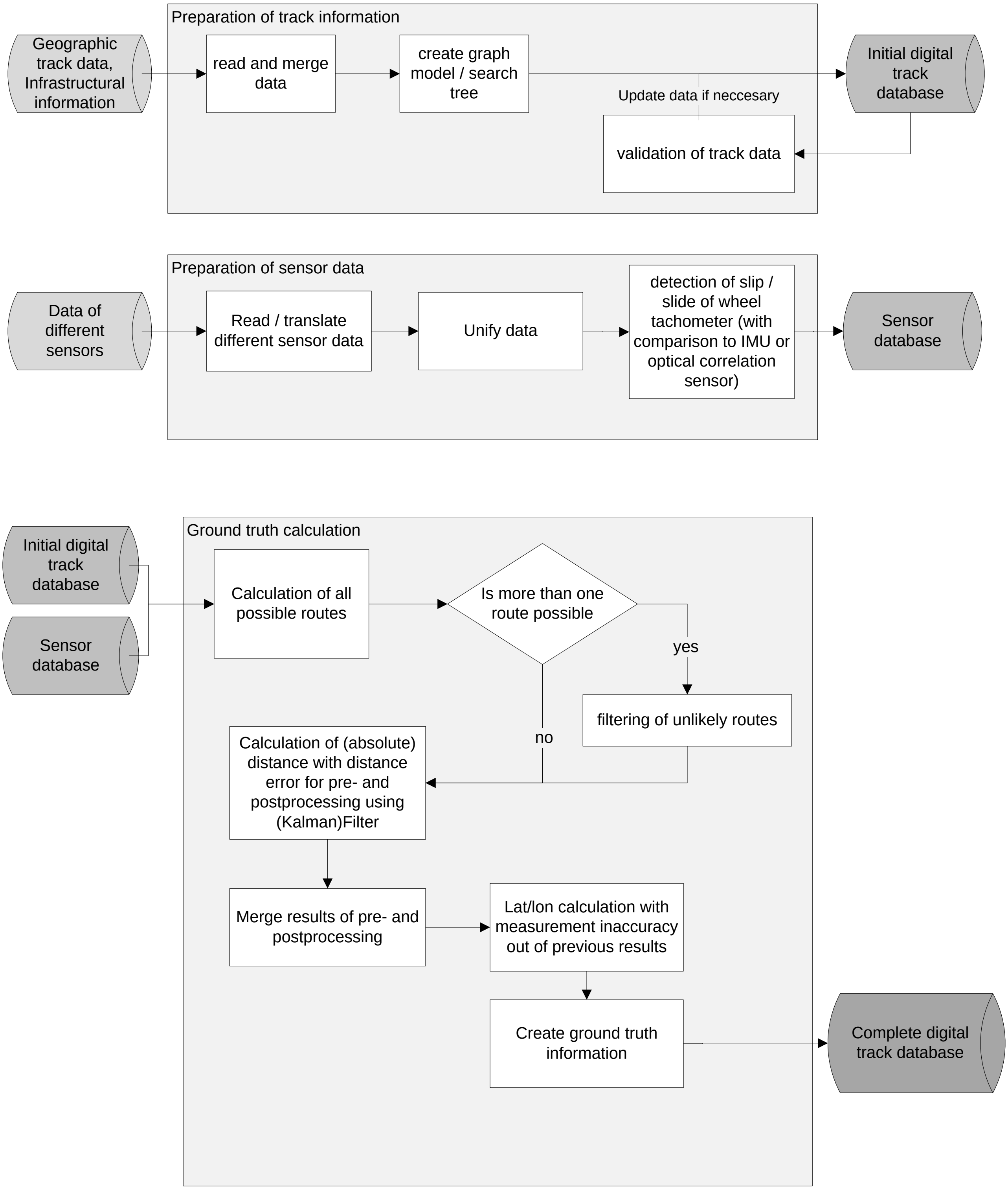


Figure 14: Overview of algorithms

The description of Figure 14 is outlined in Table 2 below.

	Work package	Description
Track Information	Read and merge / unify data & Create graph model / search tree	- If necessary convert geographical data to data structure for database / model-creation - Match nodes and segments to track sections - Match Eurobalises to track sections - Create graph model and/or search tree out of those information → railway switches as nodes
	Validation of track data	- Check data for ○ Integrity ○ Consistency ○ Accuracy - When fulfilled, create and fill initial track database
Sensor data	Translate different sensor data	- Get measurements of sensor telegrams
	Plausibility checks	- Check for incorrect time bases (e.g. if there are two different measurements for one timestamp) - Check for anomalies in sensor values
	Unify data	- Unify data by their time base (e.g. to know what wheel tachometer measurements belong to passed balises) → also consider the distances between sensors on vehicle and the resulting time shift
	Detection of slip/slide of wheel tachometer (see Annex - Slip / Slide Detection and calculation of relative speed and distance)	- Use IMU and/or optical correlation sensor (if available) for slip/slide detection of wheel tachometer - If necessary correct this data
GT Calculation	Calculation of all possible routes	- Get all possible routes for a measurement set depending on passed balises
	Filtering of unlikely routes	- If more than one possible route is found, try to filter unlikely routes - For example: ○ use IMU (if available) to detect railway switch ○ use relative distance of wheel tachometer and compare if it's close to distance (d) between two balises each → if $\Delta d > \varepsilon$ then rank route option as unlikely and discard ○ ...

	Work package	Description
	Calculation of absolute distance + distance error pre and post processing	For each successive balises: - at each balise: set absolute distance to zero and absolute distance error to minimum - pre-processing: get absolute distance and distance error of associated WT to first balise (till second balise) post-processing: get absolute distance and distance error of associated WT to second balise (till first balise) → procedure from [14] (presented below in section Slip/slide Detection and calculation of relative speed and distance) could be used

	Merging of results	- Merge pre and post processing results for each WT measurement → also note here the measurements of the different WIG, which have to be merged as well
		- Influence of pre/post processing results on merged result decreases with distance to its start point, e.g.
		    o at first Balise pre result has 100% and post result 0% influence
		    o at second Balise pre result has 0% and post result 100% influence
		    o in middle of both Balises both have influence of 50%

	Lat/Lon calculation with measurement inaccuracy	- Depending on absolute reference GPS-coordinates of Eurobalises get the Lat/Lon coordinates for each WT measurement point from calculated absolute distance and confidence intervals from absolute distance error - Get speed and acceleration from calculated absolute distance and timestamp - Ground Truth

	Create Ground Truth information	- match Ground Truth to appropriate edges/track sections and add information to database
		- information provided by ground truth:
		    o UTC timeframe
		    o LAT/LON coordinates
		    o Confidence intervals
		    o Track IDs
		    o Speed
		    o Acceleration
		- Desired precision: TBD

Table 2: Short description of Figure 14

SLIP / SLIDE DETECTION AND CALCULATION OF RELATIVE SPEED AND DISTANCE

In [15] a simply attempt for the slip/slide detection and hence calculation of the relative speeds and distances from odometer measurements is presented. The detection is realized by comparing the speed and acclearations of two odometers. If the difference exceeds a defined threshold, the wheel slips/slides.

This approach was further developed by [14]. The idea there is to also take an (monoaxial) accelerometer measurement to compare with WT measurements (beside the comparison of both WT). Again slip/slide is detected, when the difference of both sensors exceeds a defined threshold.

Out of this rating the speed and travelled distance can be calculated. There are three different calculation for the three possible cases: (per wheel 1,2)

- No slip/slide:
  - $v_i = \max(v_{1i}, v_{2i})$
- Sliding + train is accelerating
  - $v_i = \min(v_{1i}, v_{2i}, v_{i-1} + a_{est} * \Delta T)$ with
  - $a_{est} = \max\left(min\left(\frac{a_1+a_2}{2}, a_m + \Delta a\right), a_m - \Delta a\right)$
- Sliding + train is breaking
  - $v_i = \min(v_{1i}, v_{2i}, v_{i-1} + d_{est} * \Delta T)$
  - $d_{est} = \max\left(min\left(-\frac{a_1+a_2}{2}, -a_m - \Delta d\right), -a_m - \Delta d\right)$
- Legend:

  $i - index\ of\ sample$; $v - Speed$; $\Delta T - Sampling\ time$;
  $a_m -\ acceleration\ (measured\ by\ accelerometer)$; $a_{est} -\ estimated\ acceleration$;
  $d - decelation$

For all cases the travelled distance estimation is calculated as: $s_i = s_{i-1} + v_i * \Delta T$ and will always be reseted when a Balise is passed.

2.2 REFERENCE POSITION OF GNSS ANTENNA

The reference position of GNSS antenna is necessary for a comparison with GNSS position solution derived from data provided by GNSS receivers. The independence of reference position determination of GNSS antenna on the satellite navigation must be guaranteed.

Determination of the reference positionof GNSS antenna is based on 3-D GT position. Transformation between GT position and reference antenna position considers the knowledge of: 1) the offset between GNSS antenna position and the position of RFID antenna (absolute reference) resulting from installation, 2) the influence of vehicle-track interaction.

Parameters such as the superelevation and vehicle frame tilting have been investigated before the design of transformation. The aim was to provide information on real overall tilt of the railway vehicle and its impact on transversal and height deviations related to virtual antenna position, where virtual antenna position is result of antenna installation offset and GT position.

The first section presents the theoretical basis for determination of reference position of GNSS antenna elaborated by TU BS. Next sections shortly describe basic procedure taken by each partner responsible for measurement campaign (AZD, ASTS, SIE).

2.2.1 Methods for assessment of antenna reference

In this section, the method for GT transformation into antenna position is proposed by TU BS. The main purpose of this transformation is to calculate the antenna position offset caused by a tilt of railway vehicle body, superelevation, gradient of the track and installation antenna offset (out of bogie vertical axis and placement on the roof).

In order to propose a general transformation equation from the reference to the GNSS antenna, different scenarios of the train motion have to be analysed. Basically, the major influence that causes the antenna displacements is given during curves. As a result, additional antenna offsets have to be taken into account. Track parameters, such as superelevation, causes a tilt of the vehicle, which leads to a displacement of the GNSS antenna in cross track direction. For straight lines, the transformation is reduced to a simple constant offset. Figure 15 shows the SBB railway vehicle RBDe 560 with the corresponding dimensions.

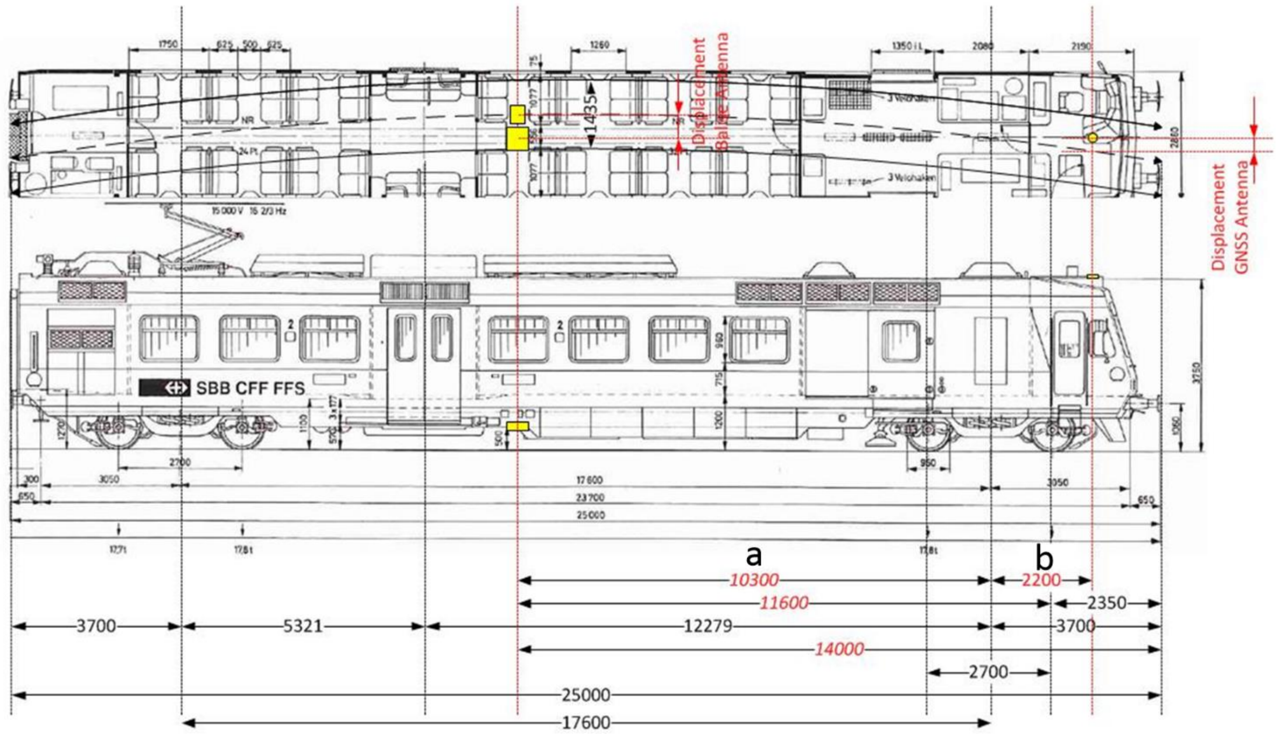


Figure 15: RBDe 560

The displacement between the GT reference (balise antenna) and GNSS antenna is caused mainly by two different factors.

The first one is a displacement caused by the rigidity of the vehicle body. Figure 15 shows the displacement of the GNSS antenna into the outside curve direction. The balise antenna, however, is displaced into the inner direction of the curve.

The second displacement is caused by the superelevation of the curve, where both antennas are displaced into the inner direction of the curve. For this reason, the calculus of the resulting displacement will be explained stepwise. In order to describe the first displacement due to the rigidity of the coach, a simple sketch of the train dimensions is depicted in Figure 16.

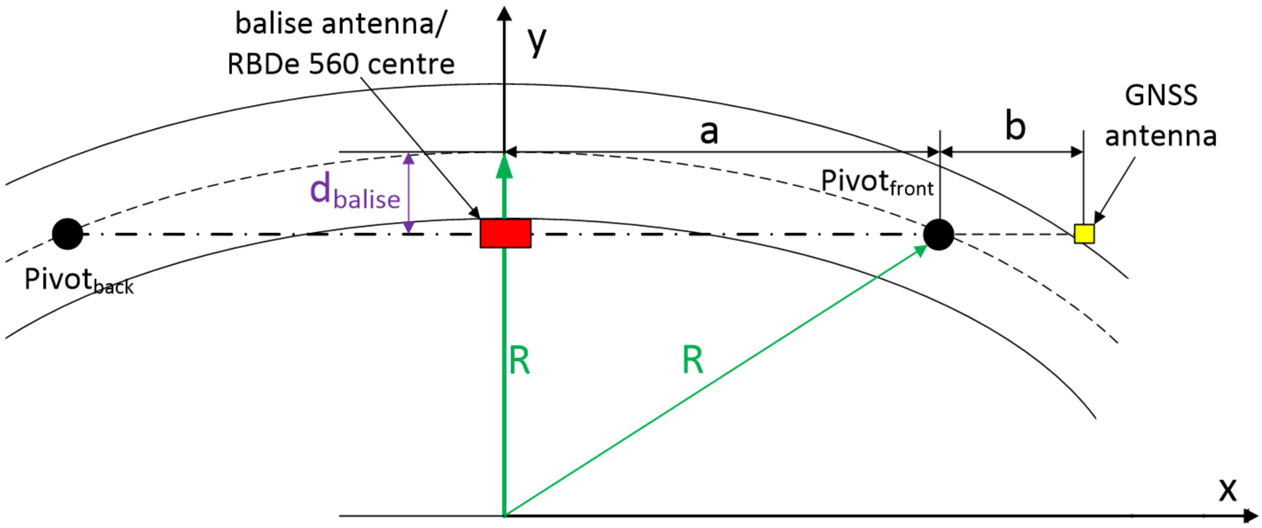


Figure 16: Displacement during curves

The centre between the pivots, where the balise antenna is assembled, is assumed to be the centre of the local coordinate system on the train. With this assumption, the coordinates x_{GNSS} and y_{GNSS} can be described with the following equations:

$$\begin{aligned} x_{GNSS} &= a + b \\ y_{GNSS} &= R - d_{balise} \end{aligned} \tag{2.1}$$

The parameters a = 10300 *mm* and b = 2200 *mm* are shown in Figure 16. These coordinates describe the offset between GT reference position (ARP in track middle) and GNSS antenna.

The remaining unknown variable d_{balise} in equation 2.1 describes the displacement of the balise antenna/vehicle centre during curves. Noticeable in Figure 16 is that the calculus of d_{balise} is possible with the following equation (Pythagoras):

$$d_{balise} = R - \sqrt{R^2 - a^2}\ . \tag{2.2}$$

Since the GT reference position is determined to be in the track middle where the ARP is assembled, the variable d_{balise} describes in this case the offset between GT reference and GNSS antenna in cross track direction.

The second displacement of the GNSS antenna is caused by the superelevation during the curves, as shown in Figure 17.

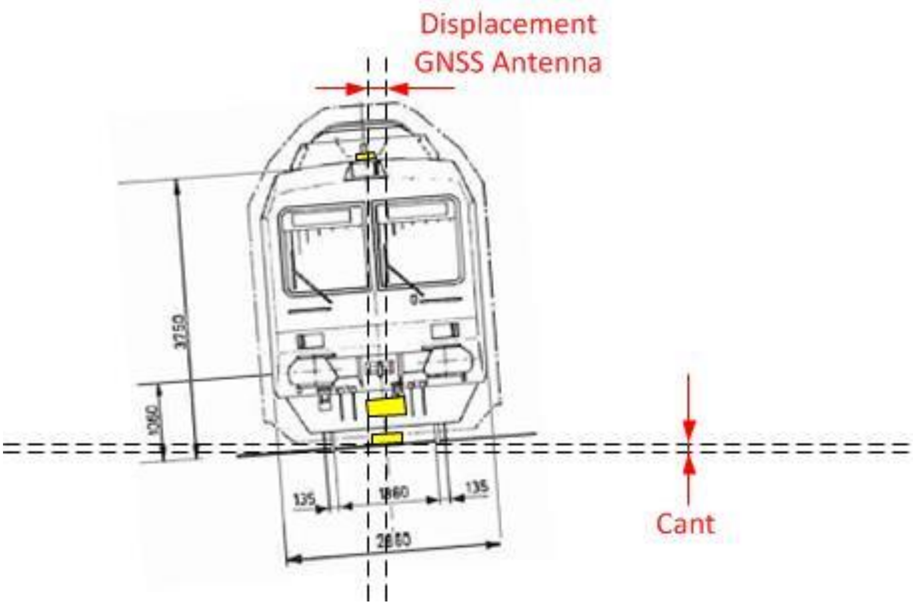


Figure 17: Displacement due to superelevation

As mentioned above, the GNSS and balise antennas are both displaced into the inner direction of the curve due to superelevation. This means for the total GNSS antenna displacement that the displacement takes place into the opposite direction as the displacement due to position of the GNSS antenna not inbetween the pivots. In order to describe the displacement, another simple sketch (Figure 18) is depicted.

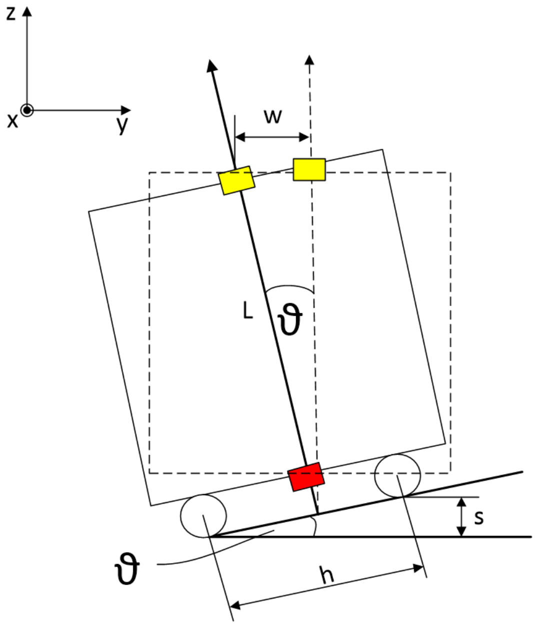


Figure 18: Displacement due to superelevation

The displacement of the balise antenna due to superelevation is neglected. According to Figure 18, only the displacement of the GNSS antenna is relevant. As shown in Figure 18, the displacement w can be calculated using the following equation:

$$w = L\sin(\vartheta) \quad (2.3)$$

The unknown angle ϑ can be computed with the geometric relationships given by the track and superelevation dimensions h and s.

$$\vartheta = \sin^{-1}\left(\frac{s}{h}\right) \tag{2.4}$$

This angle inserted in equation 2.3 provides the displacement w:

$$w = L\sin\left(\sin^{-1}\left(\frac{s}{h}\right)\right) = L\frac{s}{h} \tag{2.5}$$

With this equation the final position y_{GNSS} can be calculated:

$$y_{GNSS/Total} = -d_{balise} - w \tag{2.6}$$

At this stage, it is possible to calculate the total displacement of the GNSS antenna to the track middle. This $y_{GNSS/Total}$ replaces the y_{GNSS} value in the equation 2.1.
For straight lines, there are no complex displacements of the antennas. A transformation from balise antenna to GNSS antenna can be achieved with the dimensions given in Figure 15.
Another factor that may cause a displacement in along the track direction is the track gradient. For tracks with large gradients, the GNSS antenna will be displaced as shown in Figure 19.

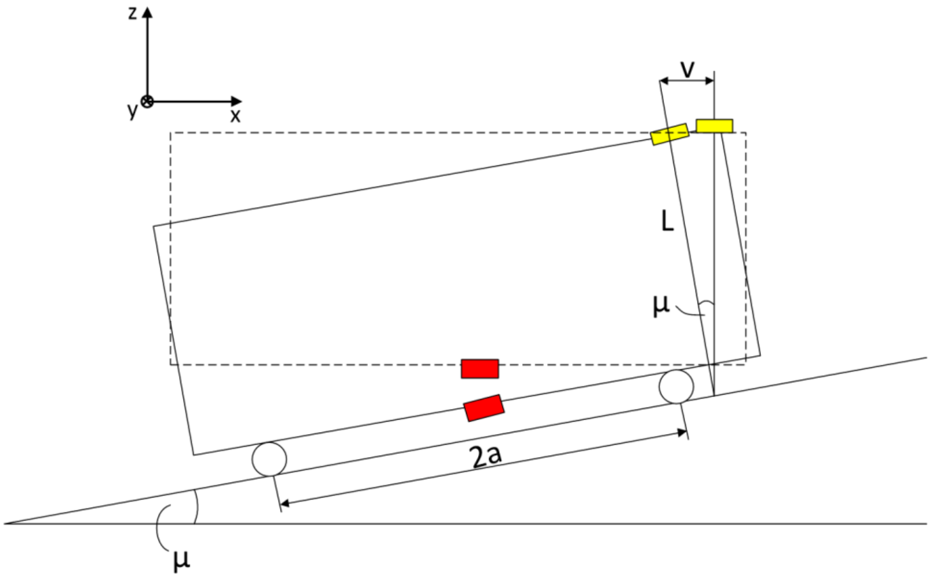


Figure 19: Displacement due to track gradient

The computation can be achieved analogously to the superelevation calculation. Therefore no equations will be presented here. The displacement of the balise antenna can also be neglected in this case.
The transformation from the balise antenna reference system to the GNSS antenna, as shown previously, depends on the track geometry. For this reason, a general transformation equation will be presented, which allows the transformation from the balise antenna reference position (GT reference) to the GNSS antenna. The balise reference position is assumed to be in the middle of the tracks, where the ARP are assembled. From this point on the track, the transformation to the GNSS antenna is realised. With the application of the geometrical relationships shown in the Figure 16 the following general transformation equation can be defined:

$$\begin{bmatrix} x_{offset} \\ y_{offset} \\ z_{offset} \end{bmatrix} = \begin{bmatrix} a + b - L \cdot \sin(\mu) \\ -d_{balise} - L \cdot \frac{s}{h} \\ L \cdot \cos(\vartheta) \cdot \cos(\mu) \end{bmatrix} \tag{2.7}$$

In case of a straight line with R = ∞ *m* and d_{balise} zero as well, no superelevation s and with zero track gradients, the equation 2.7 is reduced to the following form:

$$\begin{bmatrix} x_{offset} \\ y_{offset} \\ z_{offset} \end{bmatrix} = \begin{bmatrix} a+b \\ 0 \\ L \end{bmatrix} \quad (2.8)$$

Equation 2.8 represents the standard geometrical transformation from the balise antenna (GT reference) to the GNSS antenna. There is no offset in the y direction due to the assembly position of both antennas in the vehicle longitudinal centre.

Furthermore, the displacement due to dampers has to be investigated. According to Siemens, the maximum displacement for the RBDe 560 will be less than 10 *cm*.

2.2.2 Reference antenna position determination within measurement campaign in the Czech Republic

The GT is represented by time stamped ECEF coordinates of a representative point on the track axis map. As the processing of the GNSS signals results into an estimate of the GNSS antenna position, a reference antenna position needs to be determined from the GT to compute the error of the GNSS antenna position estimation. A worst-case analysis of the relative antenna motion with respect to the track was done in [12]. The analysis considers the superelevation, gradient, tilting and axis misalignment in curves.

The computation of the antenna reference position requires the knowledge of position offset with respect to the GT. A train vehicle usually consists of two bogies and one body. To simplify the mathematical model, it is assumed that the bogies can be considered as one rigid body. Therefore, only two rigid bodies, a bogie body (BB) and a vehicle body (VB), are considered for the mathematical model. Both rigid bodies have their own local right-handed coordinate frames assigned to them.

For each GT point the computation of the reference antenna position in ECEF CS is performed in the following steps:

1. The GT point is used to construct the navigation CS
2. The gradient and superelevation are used to define the track CS with respect to navigation CS.
3. The typical height of the ATR above the track and orientation of the train on the track are used to construct the bogie CS in the track CS.
4. The information about train tilt is used to construct the vehicle CS in the bogie CS.
5. The reference antenna position is estimated.

A detailed description of the antenna position determination is given in [11].

2.2.3 Reference antenna position determination within measurement campaign in Italy

TRANSLATION OF COORDINATES

The Ground Truth, as described in previous section, foresees a mileage value meant as offset along the curvilinear abscissa that shapes the railroad, and the related extrapolated coordinates belonging to that curvilinear abscissa.

Such typology of ground truth fits well with location applications that provide GNSS solutions constrained to the railroad.

Anyway, it could be useful to compare GNSS solutions provided by high-grade reference GNSS receivers with constrained GNSS solutions or constrained Ground Truth reference coordinates.

To allow this scenario, an operation of coordinates translation is needed. To "move" a point from the track to the space, we have to perform two basic steps:

1. Preparing the train coordinate system
2. Applying the displacement vector

Preparing the Train coordinate system

The track database represents the sequence of points crossed by a reference point of the train during its movement along the track. Possible multiple GNSS antennas (e. g. two antennas as ilustrated in Figure 20) cross the same positions, only with a relative displacement between them.

Anyway, the above-mentioned reference point is usually referred as "zero train", and can be physically located in any point of the vehicle (for example in a point coincident with the BTM antenna).

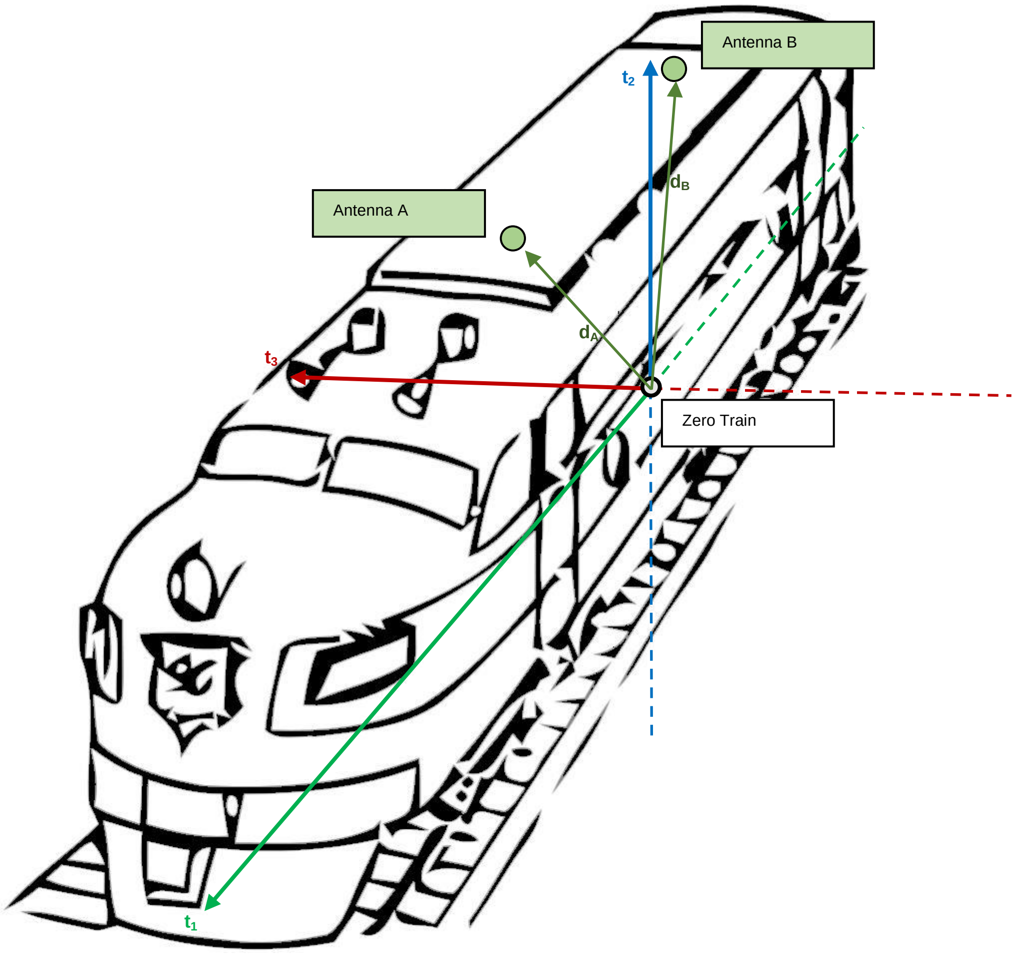


Figure 20: Train Coordinate System

The GNSS antennas will be installed likely on the roof of the train. The displacement from the *zero-train* can be of considerable amount, and it must be properly estimated.

By looking the

Figure *20*, the easiest way to represent the displacement is in the "*train*" coordinate system, i.e. a coordinate system:

- Centred in the *zero-train*;

- With axis t_1 along the longitudinal axis of the train, conventionally with the same orientation of the vector that links the cabin B to the cabin A;
- With axis t_2 along the vertical axis of the train;
- With axis t_3 along the latitudinal axis of the train, with the orientation given by the right-hand rule applied to the vector product of $t_1 \times t_2$.

In the *"train"* coordinate system the displacements for the two antennas are d_A and d_B.

In order to apply the proper displacement compensation, the transformation of displacement vectors from train coordinate system to ECEF must be defined.

The train coordinate system is a local system that changes its orientation w.r.t. the ECEF systems when the train moves along the track. It forms an orthonormal basis of $\mathbf{R}^3$, where each axis of this local reference system is represented by a vector (centred in the zero train) in the ECEF reference system.

Applying the displacement vector

With the above defined reference system, each antenna can be identified with a relative displacement w.r.t. the zero train.

Formally, if T is the following matrix:

$$T = \begin{bmatrix} t_1^x & t_1^y & t_1^z \\ t_2^x & t_2^y & t_2^z \\ t_3^x & t_3^y & t_3^z \end{bmatrix}$$

and $d_{zero\text{-}i}$ is the vector linking the zero train with the ith antenna on the train's roof, (d_i is the geometric distance of the ith antenna with the zero train) the relative lateral, longitudinal and vertical displacements can be computed as:

$$d_{zero-i} = T * [d_i]^T$$

By taking in reference the GroundTruth format, assuming the train is moving in the conventional direction, the displacements should be applied to the ECEF coordinates related to the zero train to obtain a spatial position shift. In this way, the constrained point (zero train point) is physically moved in the space both in lateral, longitudinal and vertical directions.

Formally:

$$P_{antenna_i}^{ECEF} = P_{constrained}^{ECEF} + d_{zero-i}$$

2.2.4 Reference antenna position determination within measurement campaign in Switzerland

Transformation of GT position into reference position of antenna is crucial for comparison of position of GNSS antenna computed from GNSS receiver's data and reference position of GNSS antenna determined independently on GNSS. As it's evident from diagram in

Figure *21* firstly the alignment of the train must be determined from Ground Truth position of the train, track data corresponding this position and optionally tilting sensors, consequently position offset between GNSS antenna position and Eurobalise antenna must be applied in order to obtain reference position of GNSS antenna.

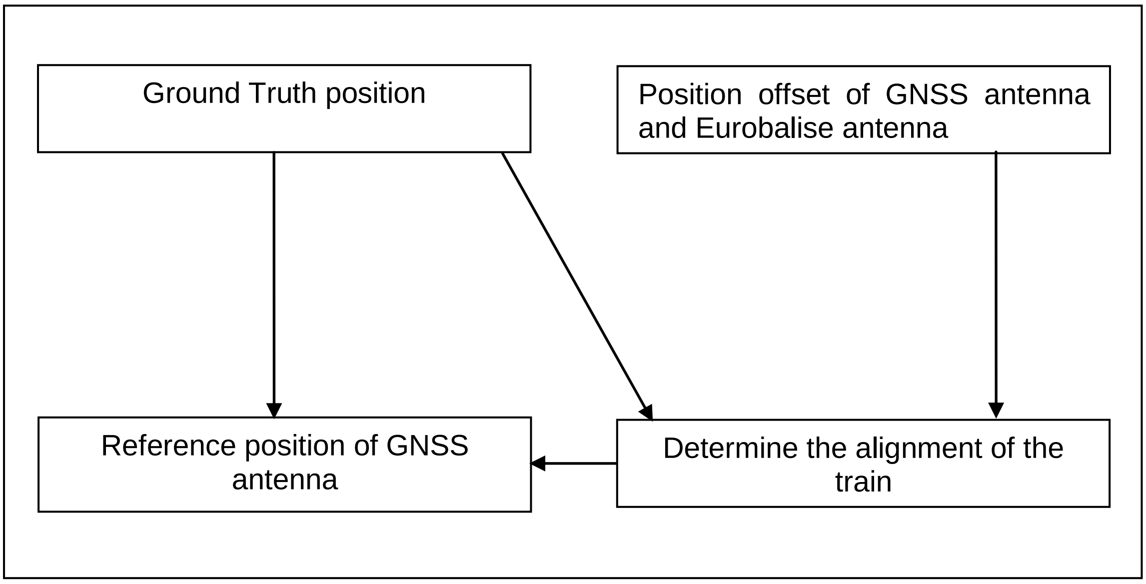


Figure 21: Basic procedure for reference position of GNSS antenna determination

Work package	Description
Determine the alignment of the train	The alignment respectively the position of the antenna towards the ground reference point at the train must be determined. Influence factors are: - Position of the antenna towards the bogies - Lateral displacement of the antenna towards axis of the train - Tilt of train in super-elevated curves - Displacement in curves and points ...
Convert absolute reference point reader into reference position of GNSS antenna	The navigation solution is related to the mounting position of the GNSS antenna and for comparison the ground reference point at the train has to be transformed to mounting position of the GNSS antenna.

Table 3: Basic procedure for reference position of GNSS antenna determination

2.3 OVERVIEW OF TOOLS AND METHODS APPROPRIATED FOR MEASURED DATA PROCESSING

The processing of measured data is an important step which has to be done before storage of the measured data in the database and employment of these data in techniques considered for evaluation of phenomena which negatively affect GNSS performance. In addition this step also includes some calculation according to the subtasks defined within Task 4.2 [1].

This section provides an overview on tools and methods appropriated for processing of different type of measured data.

2.3.1 Outputs of Devices Included in Measurement Set

Different sources of different data are supposed in the STARS measurement campaign. The list of all devices of measurement sets used in the campaign and devices' corresponding outputs is provided in Table 4.

Some devices (e.g. GNSS receivers) can provide more than one type of output data format. This ability arises due the need to meet standardized data format. Thus the standardized ASCII data format and particular binary data format corresponding to given manufacturer and device can be available at device's output. The preference is given to store data in binary data format. The binary data format mostly comprises more information from particular device than standardized format can bear and also it is much more economic from data volume perspective.

Other reason for more type of device's output is a nature of device's output as obvious in case of e.g. a fisheye camera.

Device used in campaign	Output formats	Selected output format
GNSS receiver Septentrio AsteRx4[1]	NMEA (ASCII) SBF (binary)	SBF
GNSS receiver Septentrio AsteRx3	NMEA (ASCII) SBF (binary)	SBF
GNSS receiver Javad TRE-G3T	NMEA (ASCII) JPS (binary)	JPS
GNSS receiver Javad Alpha2 (TR-G3T)	NMEA (ASCII) JPS (binary)	JPS
GNSS receiver u-blox NEO M8T	NMEA (ASCII) UBX (binary)	UBX
Spectrum analyzer Aaronia Spectran HF-8060 V5 RSA[1]	DAT (RF spectrum), TAG (time scale)	DAT, TAG
Record and playback system Spirent GSS 6450	GNS, SCN, COM1, MPG	Matlab file
Record and playback system Spirent GSS 6425	TBD (by ASTS)	TBD (by ASTS)
TeleOrbit	TBD (by SIE)	TBD (by SIE)
Fish eye camera Vivotek FE8174V	H.264, MPEG-4, MJPG, JPG	JPG
Basler ace GigE Vision acA640-300gc	Mono (8, 10, 10 Packed), Bayer BG (8, 10, 10 Packed), YUV 4:2:2 (Packed, YUYV Packed)	TBD (by SIE)
Go Pro camera (type TBD by ASTS)	TBD (by ASTS)	TBD (by ASTS)

[1] In Table 4 the devices marked with footnote *no.1* belong to the common measurement set. It is recommended to prefer processing the data of these devices.

Table 4: List of devices and output formats

2.3.2 Overview on tools and methods appropriated for data processing

The data processing consists of several steps which ensure that the database will contain only undamaged data in proper format easily readable for next processing and evaluation.

Firstly, the measured data should be somehow checked for consistency. Next, the measured data should be split and merged into files with reasonable size with respect to next processing and evaluation (if the device does not produce files of appropriate size already). Finally, the correct names should be appointed to new files according to the naming convention introduced in WP3 in [16] and updated in section 2.6.2 within this document.

The tables included in this section present suitable tools and method concerning the process described above. Each table corresponds to particular sub-task of Task 4.2 as defined in [1].

In the first column of the table the corresponding device is specified. The second column gives information on source data. The third column gives information on output data resulting from operation on the source data. The forth column sets a symptom if this output is intended to be uploaded into the database. The fifth column sets the purpose of method/SW (Ca – calculation, Co-conversion, X - checking, S - splitting, M - merging, R - renaming, F - filtering, E - extraction). The sixth and seventh columns contain a proposal of tools and methods/SW for data processing. The eighth column provides brief description of tools or methods outlined in the Table 2. The last column provides information on status of tools or methods/SW (R - ready to use, Rm - ready to use after modification, W - code has to be written, D - method has to be developed).

All files before uploading into a database should be signed by checksum and named according to the convention described in section 2.6.2 within this document.

2.3.2.1 Methods and tools related to GNSS data processing

In this section there are presented appropriated tools and methods for solution of the task 4.2.2. The data from different receivers should be processed within this subtask. It supposes data collected for different configuration of each receiver if possible and desirable.

The table of suitable tools and methods for data processing is provided for each receiver manufacturer.

Septentrio receivers

Task 4.2.2 GNSS Data Processing – GNSS Septentrio receivers							
Source Data	Output Data	Data for Database Upload	Purpose (Type of Operation)	Tools	Method/SW	Description	Status of method/sw
All SBF data of the receiver	SBF split or merged data	No	S,M	Septentrio RxTools	1. RxLogger (RxTools) or 2. SW for recording SBF data	Splitting data to files after 1 hour was defined in WP3 in configuration proposal. Splitting is ensured by 1. Proper configuration of RxLogger or by 2. A SW written by partner responsible for measurement.	1. R 2. W
SBF split or merged data	SBF validated data	No	X	Septentrio RxTools	SBF Converter	Conversion to RINEX can check consistency	R
SBF validated data	SBF data in renamed file	Yes	R	-	Script (e.g. Bash) for renaming files	The script will be developed according to naming convention described in section 2.6.2 within this document	W

Task 4.2.2 GNSS Data Processing – GNSS Septentrio receivers (continuation)							
Source Data	Output Data	Data for Database Upload	Purpose (Type of Operation)	Tools	Method/SW	Description	Status of method/sw
SBF data in renamed file	RINEX files	No	Co, X, S, M	TEQC - The Toolkit for GPS/GLONASS/Galileo/SBAS/Beidou/QZSS Data https://www.unavco.org/software/data-processing/teqc/teqc.html	Septentrio SBF proprietary format conversion to RINEX.	Teqc (pronouced "tek") is a simple yet powerful and unified approach to solving many pre-processing problems with GPS, GLONASS, Galileo, SBAS, Beidou-2/Compass, and QZSS data, especially in RINEX or BINEX format: translation: binary data reading/translation of native binary formats (optional RINEX file creation for OBS, NAV, and/or MET files or optional creation of BINEX) editing: including time windowing; file splicing; SV or other filtering; metadata extraction, editing, and/or correction of RINEX header metadata or BINEX metadata records quality check: quality checking of GPS and/or GLONASS data (native binary, BINEX, or RINEX observation files; with or without ephemerides) These three main functions (from which teqc gets its name: translation, editing, and quality check) can be performed altogether, in pairs, or separately. Translators (with varying limitations*1) are included for: Trimble RT17- and RT27-structured DAT, MES, ION, EPH download fileset Trimble RT17- and RT27-structured RS-232 real-time stream format Trimble 4700, R7, 5700, NetRS, NetR9 BINEX Topcon TPS/Javad JPS format *2 Ashtech "Version 3" B/E/S/D download fileset Ashtech MBEN/PBEN and DBEN RS-232 real-time stream format Ashtech Z-12 CGRS R-file format Ashtech micro-Z BINEX Ashtech micro-Z CGRS U-file format AOA ConanBinary AOA TurboBinary Leica System 500 and 1200 MDB binary format Leica System 500 and 1200 LB2 binary format Leica MC1000/SR9600 LB2 binary format Leica DS fileset format Septentrio Binary Format (SBF) Navcom binary format JPL Soc format IGS RTigs format u-blox UBX format Canadian Marconi binary Rockwell Zodiac binary Motorola Oncore binary (limited: no phase) Texas Instruments TI-4100 GESAR & BEPP/CORE formats Texas Instruments TI-4100 ROM format ARGO .dat and .orb format	R Download: https://www.unavco.org/software/data-processing/teqc/teqc.html

Task 4.2.2 GNSS Data Processing – GNSS Septentrio receivers (continuation)							
Source Data	Output Data	Data for Database Upload	Purpose (Type of Operation)	Tools	Method/SW	Description	Status of method/sw
SBF data in renamed file	RINEX	No	Co, X, S, M	Septentrio PPSDK Tools	Recalculation of a PVT Solution given an SBF input files and optionally an SBF files from a Base Station; • Conversion of an SBF file to RINEX 2.x or RINEX 3.x format; • Conversion of a set of RINEX files in either 2.x or 3.x format to an SBF File; • Listing of the contents of an SBF File in ASCII format; • Listing of the contents of an SBF File in Septentrio Text Format (STF); • Listing of commands contained in an SBF File; • Manipulation of an SBF files such as block filtering and cropping; • Conversion of an SBF file to gpx format; • Conversion of an SBF file to kml format for Google Earth visualization; • Listing of the block types contained in an SBF File.	Septentrio PPSDK offers these functions, which can be invoked from the PPSDK Console.	R

Table 5: Methods and tools suitable for data from Septentrio receivers

Javad receivers

Task 4.2.2 GNSS Data Processing – GNSS Javad receivers							
Source Data	Output Data	Data for Database Upload	Purpose (Type of Operation)	Tools	Method/SW	Description	Status of method/sw
All JPS data	JPS validated data	No	X	Justin	jps2rin	Javad tool for wide range of geodetic and surveying tasks, converter jps2rin	R (tool is commercial, jps2rin is free of charge)
JPS validated data	JPS data in renamed file	Yes	R	-	Script (e.g. Bash) for renaming files	The script will be developed according to naming convention described in section 2.6.2 within this document	W
JPS validated data	RINEX data	No	Co	Justin	jps2rin	Javad tool for wide range of geodetic and surveying tasks, converter jps2rin	R (tool is commercial, jps2rin is free of charge)

Table 6: Methods and tools suitable for data from Javad receivers

U-blox receivers

Task 4.2.2 GNSS Data Processing – GNSS U-blox receivers							
Source Data	Output Data	Data for Database Upload	Purpose (Type of Operation)	Tools	Method/SW	Description	Status of method/sw
All UBX data of the receiver	UBX validated data	No	X	TEQC - The Toolkit for GPS/GLONASS/Galileo/SBAS/Beidou/QZSS Data https://www.unavco.org/software/data-processing/teqc/teqc.html	U-blox proprietary format conversion to RINEX.	TEQC - The Toolkit for GPS/GLONASS/Galileo/SBAS/Beidou/QZSS Data https://www.unavco.org/software/data-processing/teqc/teqc.html	R Download: https://www.unavco.org/software/data-processing/teqc/teqc.html
UBX validated data	UBX data in renamed file	Yes	R	-	Script (e.g. Bash) for renaming files	The script will be developed according to naming convention described in section 2.6.2 within this document	W
UBX validated data	RINEX	No	Co	TEQC - The Toolkit for GPS/GLONASS/Galileo/SBAS/Beidou/QZSS Data https://www.unavco.org/software/data-processing/teqc/teqc.html	U-blox proprietary format conversion to RINEX.	TEQC - The Toolkit for GPS/GLONASS/Galileo/SBAS/Beidou/QZSS Data https://www.unavco.org/software/data-processing/teqc/teqc.html	R Download: https://www.unavco.org/software/data-processing/teqc/teqc.html

Table 7: Methods and tools suitable for data from U-blox receivers

2.3.2.2 METHODS AND TOOLS RELATED TO POST-PROCESSING COMPUTATION OF PVT

The inputs to this section are divided in the following groups:

- Methods and tools for GNSS signals and Raw Data Postprocessing
- Methods and tools for PVT computation
- Methods and tools for GNSS supporting data analysis and use

Methods and tools for GNSS signals and Raw Data Post-processing

This analysis considers two analysis domains depending on the input data files:

- Analyses of Septentrio SBF files
- Analyses of RINEX files exported from Septentrio/Javad/U-Blox receivers' output files

a) GNSS Raw Data Analyses from Septentrio SBF files:

The direct processing of these files has to be done with Septentrio proprietary SW tools, namely:

- Septentrio Post Processing SDK (PPSDK)
- Septentrio RxTools

The Table 8 takes as input WP2 measurement parameters and identifies the SW functions available in Septentrio tools to produce the required outputs.

b) GNSS Raw Data Analyses from RINEX files

Methods and tools for PVT computation

a) PVT computation from Septentrio receiver and tools
b) PVT computation from RINEX files

Methods and tools for GNSS supporting data analysis and use (*see section 2.3.2.6*)

a) Conversion of a set of RINEX files in either 2.x or 3.x format to an SBF File
b) Recalculation of a PVT Solution given an SBF input files and optionally an SBF files from a Base Station; Use of PPSDK pvtcalc tool

Sources:

- IGS: http://www.igs.org/rts
- Local geodetic institute networks
- Local CORS stations
- EGNOS EDAS
- Local Base Station

In Table 8 there are presented appropriated tools and methods for solution of Task 4.2.3. The data from different receivers should be processed within this task. It supposes data collected under different configuration of each receiver if possible and desirable.

Task 4.2.3 Post-processing computation of PVT								
Device	Source Data	Output Data	Data for Database Upload	Purpose (Type of Operation)	Tools	Method/SW	Description	Status of method/sw
STARS developed software	All receivers' output files converted to RINEX.	All parameters included in the measurement campaign as per D2.1 Table 4	YES	Ca	C++ Compiler; Matlab Octave	GNSS DATA PROCESSING Volumes I & II: ESA TM-23/2 May 2013	In case a specific in house PVT processing is required. Algorithms and code is included in the reference. Endorsed by ESA	R http://m.esa.int/About_Us/ESA_Publications/ESA_TM-23_GNSS_DATA_PROCESSING
GNSS Receiver Septentrio	All receivers' output files converted to SBF or RINEX.	PVT solutions	YES	Ca	Septentrio PPSDK Tools	Internal Algorithms	Recalculation of a PVT Solution given an SBF input files and optionally an SBF files from a Base Station	R
fastgps	All receivers' output files converted to RINEX.	All parameters included in the measurement campaign as per D2.1 Table 4	YES	Ca	fastgps	fastgps internal algoritms	**fastgps** is a GPS software receiver written in c++, which aims to be fast. Like all software receivers, fastgps performs all signal processing in software, which allows for easy tweaking at all stages: correlation, acquisition, and navigation Endorsed by GSA: https://egnos-portal.gsa.europa.eu/developer-platform/developer-toolkit/tools	R https://sourceforge.net/projects/fastgps/

Task 4.2.3 Post-processing computation of PVT								
Device	Source Data	Output Data	Data for Database Upload	Purpose (Type of Operation)	Tools	Method/SW	Description	Status of method/sw
UNAVCO-RTKLIB	All receivers' output files converted to RINEX.	All parameters included in the measurement campaign as per D2.1 Table 4	YES	Ca	UNAVCO-RTKLIB	RTKLIB internal algoritms	RTKLIB is an open source program package for standard and precise positioning with GNSS (global navigation satellite system). RTKLIB consists of a portable program library and several APs (application programs) utilizing the library. The features of RTKLIB are: It supports standard and precise positioning algorithms with: GPS, GLONASS, Galileo, QZSS, BeiDou and SBAS (2) It supports various positioning modes with GNSS for both real-time and post-processing: Single, DGPS/DGNSS, Kinematic, Static, Moving-Baseline, Fixed, PPP-Kinematic, PPP-Staticand PPP-Fixed (3) It supports many standard formats and protocols for GNSS: RINEX 2.10, 2.11, 2.12 OBS/NAV/GNAV/HNAV/LNAV/QNAV, RINEX 3.00, 3.01, 3.02 OBS/NAV, RINEX 3.02 CLK, RTCM ver.2.3, RTCM ver.3.1 (with amendment 1-5), ver.3.2, BINEX, NTRIP 1.0, RTCA/DO-229C, NMEA 0183, SP3-c, ANTEX 1.4, IONEX 1.0, NGS PCV and EMS 2.0 (refer the Manual for details) (4) It supports several GNSS receivers' proprietary messages: NovAtel: OEM4/V/6, OEM3, OEMStar, Superstar II, Hemisphere: Eclipse, Crescent, u-blox: LEA-4T/5T/6T, SkyTraq: S1315F, JAVAD: GRIL/GREIS, Furuno: GW-10 II/III and NVS NV08C BINR (refer the Manual for details) (6) It supports external communication via: Serial, TCP/IP, NTRIP, local log file (record and playback) and FTP/HTTP (automatic download) (7) It provides many library functions and APIs for GNSS data processing: Satellite and navigation system functions, matrix and vector functions, time and string functions, coordinates transformation, input and output functions, debug trace functions, platform dependent functions, positioning models, atmosphere models, antenna models, earth tides models, geoid models, datum transformation, RINEX functions, ephemeris and clock functions, precise ephemeris and clock functions, receiver raw data functions, RTCM functions, solution functions, Google Earth KML converter, SBAS functions, options functions, stream data input and output functions, integer ambiguity resolution, standard positioning, precise positioning, post-processing positioning, stream server functions, RTK server functions, downloader functions	R http://www.rtklib.com/ See Licence conditions https://opensource.org/licenses/BSD-2-Clause

Task 4.2.3 Post-processing computation of PVT								
Device	Source Data	Output Data	Data for Database Upload	Purpose (Type of Operation)	Tools	Method/SW	Description	Status of method/sw
UNAVCO-Bernese	All receivers' output files converted to RINEX.	All parameters included in the measurement campaign as per D2.1 Table 4	YES	Ca	UNAVCO-Bernese http://www.bernese.unibe.ch/	UNAVCO-Bernese internal algorithms	The Bernese GPS software is a high performance, high accuracy, and highly flexible reference GPS/GLONASS post-processing package. State of the art modeling, detailed control over all relevant processing options, performant automatization tools, the adherence to up to date internationally adopted standards, and the inherent flexibility due to a highly modular design are characteristics of the Bernese GPS Software. Bernese is a commercial software product. Typical users: Scientists for research and education Survey agencies responsible for high-accuracy GPS surveys (e.g. first order networks) Agencies responsible to maintain arrays of permanent GPS receivers Commercial users with complex applications demanding high accuracy, reliability, and high productivity Bernese is particularly well suited for: Rapid processing of small-size single and dual frequency surveys Automatic processing of permanent networks Processing of data from a large number of receivers Combination of different receiver types, taking antenna phase center variations into account Combined processing of GPS and GLONASS observations Ambiguity resolution on long baselines (2000 km and longer) Ionosphere and troposphere monitoring Clock estimation and time transfer Generation of minimum constraint network solutions Orbit determination and estimation of Earth orientation parameters	R http://www.bernese.unibe.ch/

Task 4.2.3 Post-processing computation of PVT								
Device	Source Data	Output Data	Data for Database Upload	Purpose (Type of Operation)	Tools	Method/SW	Description	Status of method/sw
UNAVCO-Gamit	All receivers' output files converted to RINEX.	All parameters included in the measurement campaign as per D2.1 Table 4	YES	Ca	UNAVCO-Gamit https://www.unavco.org/software/data-processing/postprocessing/gamit/gamit.html	UNAVCO-Gamit internal algorithms	The Department of Earth, Atmospheric, and Planetary Sciences at the Massachusetts Institute of Technology (MIT) is funded by the National Science Foundation (NSF) to support investigators using the GAMIT and GLOBK processing software. MIT's efforts include training for investigators and their overseas collaborators, and improving the capability, reliability, user interfaces, and documentation of the software. GAMIT, GLOBK, and TRACK form a comprehensive suite of programs for analyzing GPS measurements primarily to study crustal deformation. The software has been developed by MIT, Scripps Institution of Oceanography, and Harvard University with support from the National Science Foundation. GAMIT is a collection of programs used for the analysis of GPS data. It uses the GPS broadcast carrier phase and pseudorange observables to estimate three-dimensional relative positions of ground stations and satellite orbits, atmospheric zenith delays, and earth orientation parameters. The software is designed to run under any UNIX operating system. GLOBK is a Kalman filter whose primary purpose is to combine various geodetic solutions such as GPS, VLBI, and SLR experiments. It accepts as data or "quasi-observations" the estimates and covariance matricies for station coordinates, earth-orientation parameters, orbital parameters, and source positions generated from the analysis of the primary observations. The input solutions are generally performed with loose a priori uncertainties assigned to all global parameters, so that constraints can be uniformly applied in the combined solution. TRACK is a GPS differential phase kinematic positioning program. TrackRT is a Realtime GPS kinematic processing program.	R GAMIT/GLOBK/TRACK may be obtained without written agreement or royalty fee by individuals, universities, and government agencies for any non-commerical purpose. To obtain the download password and be added to the mail list for future updates, please send e-mail to Dr. Robert W King (rwkchandler.mit.edu). You must include in the e-mail the full name, address, and telephone and fax numbers of your institution. For processing support, contact Bob King (rwkchandler.mit.edu) or Tom Herring (tahmtglas.mit.edu) by e-mail or see the GAMIT/GLOBK/TRACK Documentation.

Task 4.2.3 Post-processing computation of PVT								
Device	Source Data	Output Data	Data for Database Upload	Purpose (Type of Operation)	Tools	Method/SW	Description	Status of method/sw
NASA JPL APPS	RINEX Files	•All coordinates are in ITRF2008 •Summary file: Run status (success/error), key output values, list of output. •XYZ file: Time, X, Y, Z coordianates, together with their formal errors. •LLH file: Time, Lat, Lon, Height coordinates, together with their formal errors. •GIPSY STACOV file: X, Y, Z coordinates with full covariance (for static only) •GIPSY TDP file: Time, X, Y, Z coordinates in GIPSY's TDP format. •Log file: GIPSY run time messages and statistics •Google Earth URL: with the plotted trajectory/site •Instant positioning for web users, where you will see the X, Y, Z coordinates and the Lat, Lon, Height on this web site •Access information: URL of output files and password	IF required	Ca	NASA JPL APPS	NASA JPL internal algoritms	APPS accepts GPS measurement files, and applies the most advanced GPS positioning technology from NASA's Jet Propulsion Laboratory to estimate the position of your GPS receivers, whether they are static, in motion, on the ground, or in the air. APPS employs: •Real-time GPS orbit and clock products from JPL's GDGPS System •JPL's daily and weekly precise GPS orbit and clock products •JPL's GIPSY-OASIS software for processing the GPS measurements APPS continues to provide JPL's venerable AutoGIPSY (AG) service - for free, for static post-processing (e.g. measurement latency of a week or more), but also offers new and unique services: • APPS will generate a time series of positions if your receiver was in motion • APPS has access to real-time GPS orbit and clock products so you never have to wait • APPS is fast. Positioning is available in seconds Fly-Land-Send - APPS will estimate your flight trajectory within minute s after submission. APPS support a diverse customer base, from occasional users to heavy-duty industrial users. Occasional users may upload their measurement files manually through this web site. Heavy-duty users will receive Secure FTP access to our servers, to which they can upload their measurement files, and more easily automate their processing. It is also possible to interface with APPS through e-Mail. APPS supports input in RINEX 2, RINEX 2.11 input files, GIPSY TDP files	R http://apps.gdgps.net/index.php

Task 4.2.3 Post-processing computation of PVT								
Device	Source Data	Output Data	Data for Database Upload	Purpose (Type of Operation)	Tools	Method/SW	Description	Status of method/sw
GIPSY-OASIS II	RINEX Files	TBD	IF required	Ca	GIPSY-OASIS II	GIPSY-OASIS II Internal Algorithms	GIPSY-OASIS, or GIPSY, is the GNSS-Inferred Positioning System and Orbit Analysis Simulation Software package. GIPSY is developed by the Jet Propulsion Laboratory (JPL), and maintained by the Near Earth Tracking Applications and Systems groups. Features: Analysis of data from Global Navigation and Satellite Systems (GNSS): Global Positioning System (GPS) Russian GLONASS French DORIS Satellite Laser Ranging (SLR) system Precise centimeter-level GNSS-based positioning and timing: Space platforms, including low-Earth orbiters and GNSS constellations Airplanes Terrestrial (ground) stations, static or moving Time transfer Single high-level user interface supports majority of precise positioning applications Single-receiver ambiguity resolution using JPL's GPS orbit and clock products Documentation accompanying software modules, and online help Training at annual GIPSY user group classes, and online Hundreds of research and educational users in more than 20 countries Builds on more than 25 years of JPL experience with GPS data analysis Applications: Terrestrial positioning for geophysical research: Earth deformation, plate tectonics GNSS Networks (e.g. PBO and SCIGN) Ice flow Climate studies through observation of troposphere and ionosphere Reference frame (geocenter and scale) and Earth rotation parameters Airplane positioning Precise orbit determination Low-Earth orbiters: Topex/Poseidon, Jason-1, Jason-	R https://gipsy-oasis.jpl.nasa.gov/index.php?page=software https://gipsy-oasis.jpl.nasa.gov/index.php?page=faq TOOL MANUAL available after registration. TWO JPL presentations have been downloaded for preliminary assessment.

Task 4.2.3 Post-processing computation of PVT								
Device	Source Data	Output Data	Data for Database Upload	Purpose (Type of Operation)	Tools	Method/SW	Description	Status of method/sw
							2/OSTM, GRACE, Champ, SAC-C, COSMIC, Space Shuttle (e.g. SRTM), Commercial Satellites GNSS constellations including GPS and GLONASS OUTPUT data are provided in QM Proprietary Files. GIPSY Oasis provides a tool to Convert QM files to ASCII files. OUTPUT parameters can be cofigured.	

Task 4.2.3 Post-processing computation of PVT								
Device	Source Data	Output Data	Data for Database Upload	Purpose (Type of Operation)	Tools	Method/SW	Description	Status of method/sw
PEGASUS	RINEX Files	PVT solutions	IF required	Ca	PEGASUS	PEGASUS internal algorithms	PEGASUS is a toolset which allows analysis of GNSS data collected from different SBAS and GBAS systém. The tool provides several functionalities such as computation of position simulating MOPS-compliant receivers and determination of GNSS augmentation attributes like accuracy, integrity, computation of trajectory errors, prediction of continuity and availability and simulation of GBAS Ground Station processing algorithm. Pegasus provides the following features: Data plausibility checking; SBAS/GBAS GAST-C/D message content distribution; Static and dynamic performance evaluation; Data integrity analysis; Air data calculations; Statistics accumulation; Multipath investigation; Spatial extrapolation of augmentation system performance and performance prediction; Aircraft Dynamics corrections and translation of data to a common navigation reference point; Graphical CSV data file viewer; Generic data viewers and predetermined GNSS-specific performance analysis graphs	R (available upon request) http://www.eurocontrol.int/pegasus

Table 8: Methods and tools suitable for sub-task 4.2.3

2.3.2.3 METHODS AND TOOLS RELATED TO PROCESSING OF GNSS SUPPORTING DATA

The data from meteorological servers or GNSS directly supporting servers (EDAS, IGS) have quite different format.

Regarding EDAS and IGS the type of available data and its format are described in [17], [18]. The data will be stored in original format which is suitable for next processing. The methods and tools are listed in Table 9.

Information on meteorological servers and data provided by them is given in [19]. National providers have to be contacted to provide data format and historical data, this service is mostly paid service.

Task 4.2.7 Processing of GNSS supporting data								
Device	Source Data	Output Data	Data for Database Upload	Purpose (Type of Operation)	Tools	Method/SW	Description	Status of method/sw
ESA GSSF GALILEO System Simulation Facility (and GPS) https://www.gssf.info/default.htm	Measurement Scenario Geographical Data RINEX files for Analysis	- **Visibility** Analyses - **Coverage** Analyses - **DOP** Analyses - **Navigation** System - **Precision** Analyses - **Integrity** Analyses **- RAW Data** Generation	IF required	Ca, X	Service Volume Simulator - **GSSF-SVS** which supports navigation and integrity performance analyses over long time periods and over large geographical areas Raw Data Generator - GSSF-RDG which supports early validation of ground segment algorithms by Galileo and GPS raw data generation (RINEX observation, IGS SP3)	GNSS performance simulation algorithms (developed under ESA contract) - Set up and configure Simulation Scenarios. - Run a simulation. Analyse Data. Visualise Data. Import and Export Data.	GPS and Gaileo services volume simulator certified by ESA. To be used in campaign preparation and to check measured data.	R Licence available https://www.gssf.info/GSSF%20licence%20request.htm
RIMS2RINEX 2.0	EGNOS RIMS output files	RINEX files	Yes	Co	RIMS2RINEX 2.0	Format conversion	Tool to convert EGNOS RIMS native files to RINEX format	R Licence available http://www.egnos-pro.esa.int/software.html

Systems Tool Kit (STK)	Scenario Terrain and GNSS constellation Data	Line-of-sight calculations Access tool, satisfaction intervals and gaps Geometric limitation constraints Azimuth, elevation, range, altitude, angular rate and sensor field of view Multi-constraint satisfaction intervals Combine any number of access constraints Thousands of pre-defined data providers Lat, long, alt, yaw, pitch, roll, lighting times Model Time Model, analyze and visualize in real or simulated time	IF required	Ca, X	Systems Tool Kit (STK) http://www.agi.com/products/stk/	Proprietary Algorithms	Systems Tool Kit (STK) is the foundation of AGI's product line. This highly capable, free modeling environment is used by thousands of engineers, mission analysts, and software developers to model complex systems - such as aircraft, missiles, satellites and their sensors - analyze mission simulations and visualize dynamic datasets in 4D (X,Y,Z,Time). STK's open API, file interoperability and add-on modules can extend these core capabilities, and solutions can be deployed using our software development kits and server product. Finally, STK users can share results such as reports, graphs and HD videos with colleagues, managers and decision makers.	R Licence available https://www.agi.com/products/stk/license/

Table 9: Methods and tools suitable for sub-task 4.2.7

2.3.2.4 Methods and tools related to RF I/Q samples processing

The RF I/Q sample data can be provided by RPS or GNSS receiver Septentrio AsteRx4.

In case of the RPS Spirent GSS6450 and GSS6425 the data management is described in [20]. Simply summarized a new file is created after switch on recording and closed after switch off recording. The data will be stored in the original binary format. A description of the protocol is proprietary to Spirent, but should be released for AZD and ASTS. Consequently, AZD and ASTS will convert data from original files into file with known data format readable in e.g. Matlab or Octave.

The RPS TeleOrbit... TBD by SIE

Concerning the Septentrio receiver the RxTools enables to split and convert data stored in SBF file into the text format.

2.3.2.5 Methods and tools related to processing from video data from cameras

In order to gain a visibility mask, a situation of the pantograph on a train and information about the ambient weather, an IP camera will be used (AZD's measurement system will employ FE8174V manufactured by company VIVOTEK). The camera has the ability to capture surroundings around 360° (ie. “Fisheye”). Scanned records from this camera will be processed with image processing algorithms. Results of the processing will be the information above.

Tools for acquiring data from IP camera Vivotek FE8174V

Video from the used camera will be saved as a set of JPEG images to the PC MOXA hard disk. The free cross-platform software “avconv”, that is capable of streaming, converting and recording the video, will be used for this purpose.

Quality, resolution and frame rate of the video stream can be set directly on the camera for three independent streams (hereinafter referred to as stream1, stream2, stream3). Some options for each stream are presented in Table 10. There are indicative sizes of JPEG images depending on quality of the stream in Table 11. Two independent streams with parameters presented in Table 12 will be used in measurement campaign, but proposed setting is not definitive. It can be changed according to the necessary requirements for the resolution and image quality during development of our image processing algorithms.

Images for the mask visibility will be gained from stream1 once a month. Then they will be processed with image processing algorithms and based on the result, the visibility mask will be updated. When creating the mask, we have to take inclination of the train into the account. It will be determined by biaxial inclinometer installed on the train.

Images from stream2 will be used for information about the ambient weather and the situation of the pantograph. This stream will be recorded continuously.

Attribute	Possibility
Compression	MPEG-4, H.264, MJPEG
Resolution	192x192, 256x256, 384x384, 512x512, 768x768, 1056x1056, 1280x1280, 1536x1536, 1920x1920
Framerate	1fps - 15fps
Video quality	Medium, Standard, Good, Detailed, Excellent

Table 10: Optional parameters of streams

Resolution	192x192		768x768		1920x1920	
Video quality	Standard	Excellent	Standard	Excellent	Standard	Excellent
Indicative size of JPEG image [KB]	11	13	28	31	108	115

Table 11: Indicative size of JPEG images

Stream1	Purpose	Visibility mask
	Compression	MJPEG
	Resolution	768x768
	Framerate	1fps
	Video quality	Good
	Period of image storing	1s
	Indicative size of the image	29KB
Stream2	Purpose	Situation of pantograph; ambient weather
	Compression	MJPEG
	Resolution	768x768
	Framerate	1fps
	Video quality	Standard
	Period of image storing	10s
	Indicative size of the image	28KB

Table 12: Setting of stream parameters for campaign

2.4 OVERVIEW OF TOOLS AND METHODS APPROPRIATED FOR DATA ANALYSIS AND EVALUATION OF NEGATIVE PHENOMENA

Before the evaluation can start another important step has to be taken. This step comprises an analysis of different pre-processed or measured data. Based on the results of these analyses an evaluation can be carried out.

Following sections present suitable tools and method concerning to the step described above. Each section corresponds to particular sub-task of Task 4.3 as defined at [1].

2.4.1 Identification of factors disturbing GNSS signal and position solution based on comparison in position domain (Task 4.3.2)

This section includes a set of proposals for methodologies, algorithms and tools to be used in order to carried out analyses activities in tasks 4.3.2.x (x:1-5). Some assumptions are made:

a) The proposals consider the Septentrio receivers features and functionalities.

b) GT, reference position of GNSS antenna (derived from GT) and methodologies are available.

c) Terrain and Track digital high accuracy georeferenced (GNSS compatible) data is available. Track geometry can also be obtained by track points determination based reference antenna position, and static PPK application (interpolation could be used depending on the points spacing and the positions comparing technique used). This track modelling technique could be costly in time and track occupancy, and requires that antenna is located at the same relative height as the measuring antenna mounted on the train.

d) The difference of magnitudes between cross-track and along track errors of computed are negligible for analyses purposes, specially for those cases in which only a geometrical analysis (no time reference) is carried out by comparing the PVT computed point and orthogonally projected point on the track.

It is important to keep in mind that the main objective of these analyses is to identify factor disturbing the GNSS signal, for which GNSS performance data is just a means.

2.4.1.1 Analysis based on comparison of antenna position from pseudorange solution and reference antenna position based on ground truth (task 4.3.2.1)

For this analysis two different types of GT could be considered:

- Terrain and Track digital high accuracy map georeferenced (GNSS compatible) is available
- Specific points - GT based reference antenna positions - are available and known with high accuracy, and a time frame translation reference has been envisaged such as to synchronize GT element detection event with the corresponding GNSS epoch and positioning. This approach is expected to be employed as the synchronization of GT based reference position and GNSS solution is supposed.

Figure 22 presents the geometric layout as a reference for comparison of GNSS and GT based positioning.

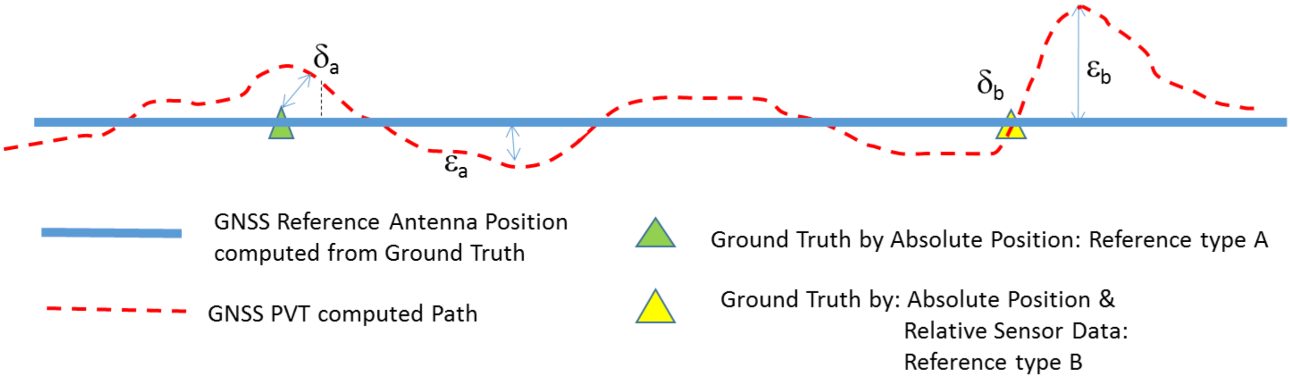


Figure 22: Geometric layout: Reference position derived from georeferenced digital track data (blue), Path calculated from GNSS PVT (dashed red), GT based reference position data (green and yellow triangles)

In Figure 22 the trajectory of reference position of GNSS antenna has been presented as a straight segment just to simplify representation, the same approach is suitable for curved segments, considering that GNSS path point correspondence is determined by the orthogonal projection of the track curve on the GNSS path.

According to this approach the following analysis methodology is proposed:

INPUTS:

- GT based reference position 1 Data: Derived from digital track data georeferenced
- GT based reference position 2 Data: Two types for reference position determination can be considered, Type A – real balise, i.e. absolute GT reference where the highest accuracy of the reference position of GNSS antenna is obtained, Type B – GT position based on

relative sensor measurement of travelled distance from last balise. In this case high accuracy in position (georeferenced) and timing are considered.

- Path calculated by GNSS PVT computation.
- GNSS performance forecast are available so that GNSS average errors simulated estimations are known for each specific test campaign (forecast tools presented in section 2.3).

WORKFLOW:

- The Error $\varepsilon(t_i)$ will be computed for each epoch t_i as length of the track orthogonal segment that crosses the GNSS path curve. Note that this is a statistical approach and no error decoupling: cross-track / along the track is intended. Besides, it is strictly a geometrical estimation of error between real track and computed path, which should provide a good estimation of the best possible performances for a certain segment.

  Statistic parameters for the $\varepsilon(t_i)$ series are calculated: average value, mean deviation.

  $\varepsilon(t_i)$ error values can be classified according to their deviation for example: ε_a low error (acceptable), ε_b high error (see Figure 22). For those high error areas a threshold error value ε_{th} should be set, so that in an cross correlation analysis between the scenario descriptive parameters and errors could be carried out. For these points, receiver data and signal data information will also be used.

- For those points in which GT elements (time reference enabled) are available an additional calculation based on errors $\delta_i(t_i)$ could be carried out. For this case (time referenced), the error could be decoupled in the cross-track / along-track components. In a similar fashion, δ errors can be classified δ_a larger errors, δ_b very small errors. For the set of values of high errors a statistical correlation with the occurrence of scenario descriptive parameter could be made.

Task 4.3.2 / 4.3.2.1 Comparison of antenna position from pseudorange solution and reference antenna solution based on GT								
Device	Source Data	Output Data	Data for Database Upload	Purpose (Type of Operation)	Tools	Method/SW	Description	Status of method/ sw
GNSS Receiver Septentrio AsteRx4	standalone position computed by receiver, reference position of GNSS antenna based on GT	symptom of exceeding defined level, computed difference	Yes	Ca	Matlab, Octave etc. GIS tool and terrain modelling tool (e.g AGI STK) Statistical Analysis SW Package	SW for comparison See section 2.3	SW for comparison of both antenna positions SW for comparing TRACK digital data based position of GNSS antenna and PVT computed by GNSS receiver.	W

Table 13: Methods and tools suitable for sub-task 4.3.2.1

2.4.1.2 Analysis based on comparison of antenna position from pseudorange solution and reference antenna position from PPK (task 4.3.2.2)

This analysis is conceptually similar to analysis in section 2.4.1.1, the main difference being that for this case we will have two paths based on two sets of PVT solutions, a first one without augmentation data applied and a second one based on PPK post-processing techniques that will improve significantly the accuracy.

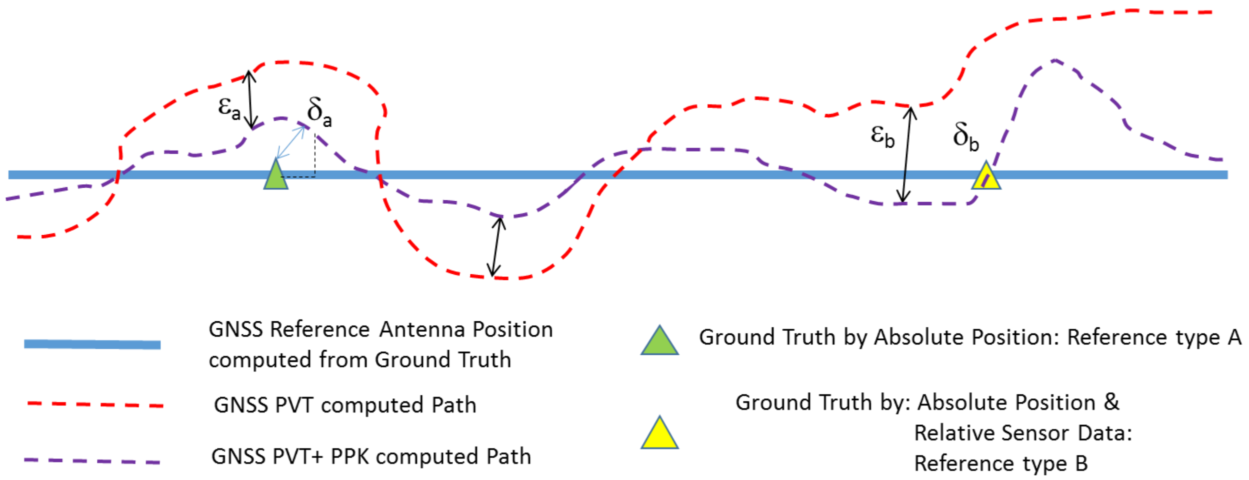


Figure 23: Geometric layout: Reference position derived from georeferenced digital track data (blue), Path calculated from GNSS PVT (dashed red), Path calculated from PPK (dashed magenta), GT based reference position data (green and yellow triangles)

The statistical analyses will compare:

- PVT and PVT-PPK paths. In this case the error comparison $\varepsilon(t_i)$ will be done on time basis, i.e. for each epoch $\varepsilon(t_i)$.
- PVT-PPK vs Track digital data based reference. For this analysis track digital data error is fundamental.

Special attention will be paid to those interference factors that are corrected to PPK, so that divergence between PVT and PVT+PPK above average are identified as signs of interference elements in the scenario.

Task 4.3.2 / 4.3.2.2 Comparison of antenna position from pseudorange solution and reference antenna solution based on PPK								
Device	Source Data	Output Data	Data for Database Upload	Purpose (Type of Operation)	Tools	Method/SW	Description	Status of method/ sw
GNSS Receiver Septentrio AsteRx4	standalone position by receiver, reference antenna position from PPSDK	symptom of exceeding defined level, computed difference	Yes	Ca	Matlab, Octave etc. GIS tool Statistical Analysis SW Package AGI STK	SW for comparison	SW for comparison of both antenna positions	W

Table 14: Methods and tools suitable for sub-task 4.3.2.2

2.4.1.3 Analysis based on comparison of antenna position based on corrected pseudorange solution and reference antenna position based on GT (Task 4.3.2.3)

This analysis follows a methodology analogous to section 2.4.1.1, being the main difference that pseudoranges are corrected by augmentation data. The following augmentation data will be considered:

- EGNOS data.
- IGN or CORS ground stations data
- PPP orbital and clock corrections provided by the identified sources.

Task 4.3.2 / 4.3.2.3 Comparison of antenna position from corrected pseudorange solution and reference antenna solution based on GT								
Device	Source Data	Output Data	Data for Database Upload	Purpose(Type of Operation)	Tools	Method/SW	Description	Status of method/sw
GNSS Receiver Septentrio AsteRx4	augmented position computed by receiver, reference antenna position from PPSDK	symptom of exceeding defined level, computed difference	Yes	Ca	Matlab, Octave etc.	SW for comparison	SW for comparison of both antenna positions	W

Table 15: Methods and tools suitable for sub-task 4.3.2.3

2.4.1.4 Analysis based on comparison of antenna position from corrected pseudorange solution and reference antenna position from PPK (Task 4.3.2.4)

This analysis follows a methodology analogous to sections 2.4.1.2, being the main difference that pseudoranges are corrected by augmentation data. The following augmentation data will be considered:

- EGNOS data.
- IGS or CORS ground stations data
- PPP orbital and clock corrections provided by the identified sources.

Task 4.3.2 / 4.3.2.4	Analysis based on comparison of antenna position from corrected pseudorange solution and reference antenna position from PPK							
Device	Source Data	Output Data	Data for Database Upload	Purpose (Type of Operation)	Tools	Method/SW	Description	Status of method/sw
GNSS Receiver Septentrio AsteRx4	augmented position computed by receiver, reference antenna position from PPSDK	symptom of exceeding defined level, computed difference	Yes	Ca	Matlab, Octave etc PPSDK RAW Data generator (ESA GSSF)	SW for comparison	SW for comparison of both antenna positions	W

Table 16: Methods and tools suitable for sub-task 4.3.2.4

2.4.1.5 Analysis based on comparison of position solution from different satellite subsets (Task 4.3.2.5)

For this analysis the following methodology will be followed:

INPUTS:

- Selection of locations and instants of interest to carry out the analysis.

- Extraction of RAW data for those selected points.
- Arrangement of RAW data to present different constellation configuration (satellite subsets) for each location and instant, i.e. extraction and storage of the corresponding pseudoranges.
- Constellation configurations: Mission planner.

METHODOLOGY:

- Geometrical DOP factors and position difference between reference antenna position and GNSS estimated position will be computed for each constellation geometrical layout (satellite subset).
- Geometrical impact will be ranked for different mask angles.
- Position difference together with DOP factor will be ranked to enable identification of multipath occurance.

Task 4.3.2 / 4.3.2.5	Analysis based on comparison of position solution from different satellite subsets							
Device	Source Data	Output Data	Data for Database Upload	Purpose (Type of Operation)	Tools	Method/SW	Description	Status of method/ sw
GNSS Receiver Septentrio AsteRx4	standalone position computed by receiver, reference antenna position from PPSDK	symptom of exceeding defined level, DOP factors, computed difference	Yes	Ca	Matlab, Octave etc. Mission Planner PPSDK	SW for comparison	SW for comparison of both antenna positions	W

Table 17: Methods and tools suitable for sub-task 4.3.2.5

2.4.2 Identification of factors disturbing GNSS signal and position solution based on raw data and specific features of receivers

2.4.2.1 Analysis based on deviation of pseudoranges in time (Task 4.3.3.1)

This analysis will compare an expected deviation of pseudoranges in time for a given location and a given time-window with the real measured pseudoranges evolution together. If differences between computed and real values are larger than a given threshold for certain time periods within the time windows. Then scenarios characteristics and other receiver parameters will be used to determine if interference elements were present (e.g. obstacles or reflecting surfaces originating multipath effects).

Please note that this analysis is based on the evolution of pseudorange values, rather than on absolute values or errors, therefore no high accuracy orbitography is required for this analysis.

The following methodology will be used

- Track points of special interest will be selected.
- For different SV a RAW Data Generator (e.g. ESA GSSF) will be used to generate a range of pseudorange values.
- The theoretically estimated pseudorange values will be compared with measured pseudoranges.
- For those cases in which deviations are larger than expected, the causes will be investigated by:
    - Cross correlating with other receiver measured parameters: (multipath, noise)

- Cross correlating with scenario elements (obstacles detected by optical means)

- The different results from each selected points will be compared in order to find coherency in the causes for the deviations.
- Additionally the measured pseudorange is investigated from perspective of its noise and jumps caused by disturbances on the signal should be detected. (PPSDK and RxControl SW can be used for noise and cycle-slip detection).

Task 4.3.3 / 4.3.3.1 Analysis based on deviation of pseudoranges in time								
Device	Source Data	Output Data	Data for Database Upload	Purpose (Type of Operation)	Tools	Method/SW	Description	Status of method/ sw
GNSS Receiver Septentrio AsteRx4	RAW Data Generator computed pseudo ranges. Real measured pseudoranges	Symptom of exceeding defined level, correlation coefficients between pseudorange deviations and local interference elements	Yes	Ca	RAW DAT Genrator ESA GSSF Matlab, Octave etc. PPSDK (for noise and cycle slips detection)	SW for comparison and statistics computation	See section 2.4.2.1	R, W

Table 18: Methods and tools suitable for sub-task 4.3.3.1

2.4.2.2 Analysis based on comparison of measured pseudorange and distance between SV and reference antenna position (Task 4.3.3.2)

This analysis will determine the absolute error affecting pseudorange values by comparing the measured pseudorange values and the true distances between the reference antenna position and satellite antennas` centre of phases.

In order to make high accuracy computation, both the reference antenna and the SV position will have to be known with high accuracy.

INPUT:

- High accuracy PPP SV orbitography.
- High accuracy reference antenna position.
- Measured pseudoranges.
- Theoretically estimated errors in the pseudoranges domain (performance forecast simulations) (different error models can be used, e.g. different ionospheric models.

Methodology:

- Geometrical and measured psedoranges are compared to determine the absolute value in the pseudorange domain.
- Errors can be translated to the position domain by means of the Geometrical Matrix.
- Errors can be compared with theoretical performance simulation data, for different error modelling parameters, in order to estimate the goodness of the different error models (e.g. ionospheric models, multipath models) for the terrestrial environments, and the need to develop new error budgets specific for railway scenarios.

Note that the analysis realized in this task exploits reference position of GNSS antenna. This reference position is derived from GT data, antenna position offset caused by installation on a vehicle and a tilt of the vehicle (superelevation, vehicle frame tilting). GT data and also reference antenna position data should be represented by points with given ECEF XYZ coordinates, reference time (e.g. GPS time) and confidence interval/ max error (described in Task 4.1.2). Each of these points is related to the corresponding GNSS measurement since it is created in the same reference time. The accuracy of estimate of cross track error by projection of position estimate from GNSS measurement to reference position of the antenna will be dependent on the accuracy of reference position determination (i.e. confidence interval/ max error).

Task 4.3.3 / 4.3.3.2	Analysis based on comparison of measured pseudorange and distance between SV and reference antenna position							
Device	Source Data	Output Data	Data for Database Upload	Purpose (Type of Operation)	Tools	Method/SW	Description	Status of method/ sw
GNSS Receiver Septentrio AsteRx4	High Accuracy SV orbitography High accuracy reference antenna position Rx4 Raw Data	Symptom of exceeding defined level, absolute error in the pseudorange domain.	Yes	Ca	RAW data Generator ESA GSSF Matlab, Octave etc.	SW for comparison and statistics computation	See section 2.4.2.2	W

Table 19: Methods and tools suitable for sub-task 4.3.3.2

2.4.2.3 C/N_0 DATA BASED ANALYSIS (TASK 4.3.3.3)

The Carrier-to-Noise density ratio characterizes the quality of the different received satellite signals for each available antenna and is expressed in dB-Hz.
Alternately second quantity Signal-to-Noise Ratio (SNR) can be used for an analysis of the quality of received satellite signals. The relationship between both quantities is given by formula:

$$C/N_0 = SNR + BW$$

Where BW is the bandwidth of observation, which is usually the noise equivalent bandwidth of the last filter stage in a receiver's RF front-end.
This analysis will identify track locations or time periods in which the C/N_0 figure of merit, presents small values. These track areas will be investigated in order to determine the occurrence of either carrier power attenuation elements (e.g. vegetation, multipath) or EM noise sources.

C/N_0 values will be obtained from the corresponding SBF files or Rinex files:

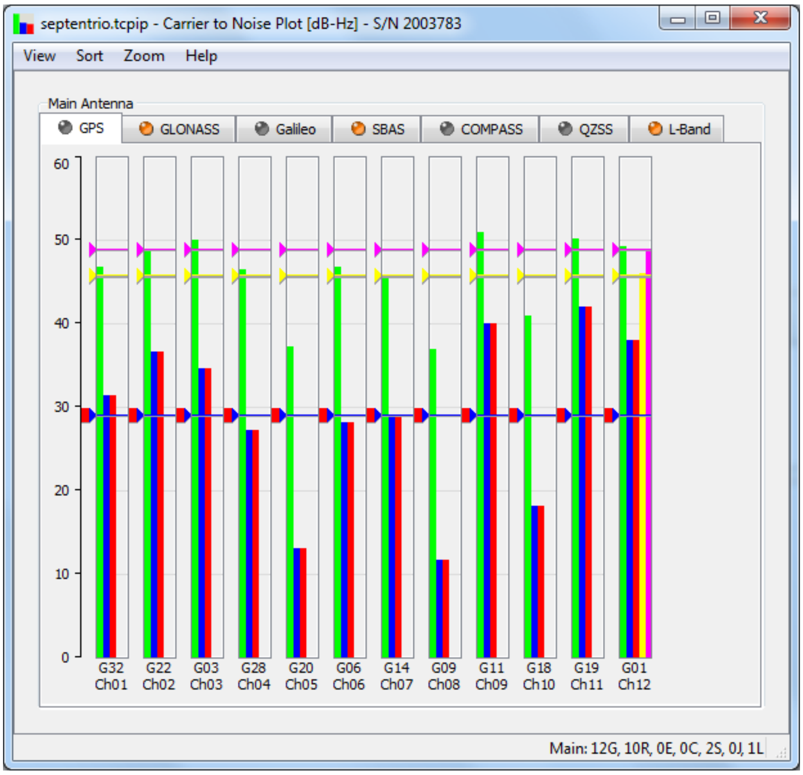


Figure 24: Carrier to Noise density ratio plot example in RxControl

The RxControl SW can be used for a visual inspection of data as shown in Figure 24.

INPUTS:

- SBF or RINEX files

METHODOLOGY:

- Analysis of carrier-to-noise density ratio to identify abnormally small values.
- Mapping of C/N_0 small values with measuring location and times.
- Cross correlation with scenario interference elements occurrence.
- C/N_0 low values for specific SV can be used for LOS analyses.

  Analysis based on carrier-to-noise density ratio comparison from measurement with dual polarization antenna (RHCP, LHCP) for multipath detection

Task 4.3.3 / 4.3.3.3				C/N0 data based analysis				
Device	Source Data	Output Data	Data for Database Upload	Purpose (Type of Operation)	Tools	Method/SW	Description	Status of method/ sw
GNSS Receiver Septentrio AsteRx4	SBF and RINEX files	Symptom(s) of exceeding defined level(s), difference of C/N0 between RHCP and LHCP measurements	Yes	Ca	Septentrio Rx Control SW PPSDK Matlab, Octave etc.	SW for comparison and statistics computation	See section 2.4.2.3	R,W

Table 20: Methods and tools suitable for sub-task 4.3.3.3

2.4.2.4 CODE MINUS CARRIER BASED ANALYSIS (TASK 4.3.3.4)

The code multipath can be seen by plotting the difference of code and carrier ionosphere-free combinations. The evolution of this difference can be followed with a sampling rate of 1 Hz. Due to its geometric nature, the effect of multipath repeats with the receiver satellite geometry.

The Septentrio Rx SW offers Measurement combination plots: These plots offer a means for advanced users to monitor several conditions such as ionospheric activity and multipath.
– **Lx - Ly** plot displays the carrier phase range difference between two signals per satellite (iono).
– **Px - Py** plot displays the code range difference between two signals per satellite (iono/-multipath).
– **Px - Lx** plot displays the difference between the code range and the carrier phase range per satellite (iono/multipath).
– **MPx** plot displays a computed indication for multipath.

INPUTS:

- SBF files recorded in selected locations which are candidates to be hostile railway scenarios.

Methodology

- SBF are analysed using RxControl measurement combination plots to determine multipath occurrence, together with other Rx parameters (e.g. multipath mitigation coefficient).
- For those locations where strong multipath occurrence has been identified for certain constellation positions (SV positions), presence of scenario elements will be cross correlated.

Task 4.3.3 / 4.3.3.4				Code minus carrier based analysis				
Device	Source Data	Output Data	Data for Database Upload	Purpose (Type of Operation)	Tools	Method/SW	Description	Status of method/ sw
GNSS Receiver Septentrio AsteRx4	SBF and RINEX files	Points of strong multipath occurrence , other Rx parameters (multipath mitigation coefficient, etc.)	Yes	Ca	Septentrio Rx Control SW PPSDK Matlab, Octave etc	SW for comparison and statistics computation	See section 2.4.2.4	R, W

Table 21: Methods and tools suitable for sub-task 4.3.3.4

2.4.2.5 SSE BASED ANALYSIS (TASK 4.3.3.5)

Sum of Square Errors are consistency checking algorithms and techniques that can be also used for error multipath detection and mitigation as in classical RAIM algorithm.

INPUTS:

- SBF files
- RAIM prediction tools (e.g Eurocontrol augur http://augur2.ecacnav.com/)

Methodology:

- Check RAIM (FD and FDE) and marked data as faulty as required.

Use RxControl:

The RAIM Statistics is a group of columns that shows the detailed results of the RAIM algorithm which ensures the integrity of the computed position solution, provided that sufficient satellites are available. The next values can be displayed:

- i Code: The Code a-posteriori measurement residual
- w i Code: The absolute value of the w-test statistic for the Code
- MDB Code: The Minimal Detectable Bias for the Code
- e i Phase: The Phase a-posteriori measurement residual
- w i Phase: The absolute value of the w-test statistic for the Phase
- MDB Phase: The Minimal Detectable Bias for the Phase
- i Doppler: The Doppler a-posteriori measurement residual

Additionally these algorithms (different approaches can be found in the GNSS literature) can be programmed using development tools as required (Matlab, C, etc).

Points where a lack of integrity has been detected will be marked as faulty. For those points a correlation analysis with enviromental factors will be carried out.

Task 4.3.3 / 4.3.3.5				SSE based analysis				
Device	Source Data	Output Data	Data for Database Upload	Purpose (Type of Operation)	Tools	Method/SW	Description	Status of method/ sw
GNSS Receiver Septentrio AsteRx4	SBF and RINEX files	Points where defined level has been exceeded, SSE absolute values.	Yes	Ca	Septentrio RxControl SW PPSDK Matlab, Octave, C, etc. RTKLib modification	SW for comparison and statistics computation	See section 2.4.2.5	W

Table 22: Methods and tools suitable for sub-task 4.3.3.5

2.4.2.6 Analysis based on multipath detection and mitigation algorithm built-in receivers (Task 4.3.3.6)

Most professional receivers have built-in multipath detection and mitigation functionalities.

INPUT:

- SBF files

Methodology

- Use Septentrio RxControl SW for MPx plot displays as computed indication for multipath.

SBF parameters indicating multipath will be extracted and presented as explained in section 2.3.2.1.

For those track locations where multipath indicator reaches high values a correlation analysis with environmental factors will be carried out.

Task 4.3.3 / 4.3.3.6			Analysis based on multipath detection and mitigation algorithm built-in receivers					
Device	Source Data	Output Data	Data for Database Upload	Purpose (Type of Operation)	Tools	Method/SW	Description	Status of method/ sw
GNSS Receiver Septentrio AsteRx4	SBF and RINEX files	Locations where multipath indicator reach high values.	Yes	Ca	Septentrio RxControl SW PPSDK Matlab, Octave etc.	SW for comparison and statistics computation DSP SW	See section 2.4.2.6	R, W

Table 23: Methods and tools suitable for sub-task 4.3.3.6

2.4.2.7 Analysis based on RF interference detection and mitigation algorithm built-in receiver (Task 4.3.3.7)

The RxControl Spectrum View enables spectral analysis of the RF signal received from the antenna. The user can choose between visualizing the RF spectrum, the raw IF samples or a histogram distribution of the samples. The represented frequency band can be selected at the top of the window. The user can also determine whether the values must be averaged before visualization or not, and the number of values that must be included in the average. For the histogram distribution the user can choose to visualize the I and Q samples or I2 + Q2. The user can also choose to normalize the histogram.
The user can see in the RF spectrum at which frequency interference is present (the peaks in the

plot). From the raw samples the user can see pulsed interference. If the Septentrio Receiver is heavily jammed the I and Q samples histograms will not show a gaussian distribution and the I2 + Q2 histogram will not show a chi-square shape distribution.

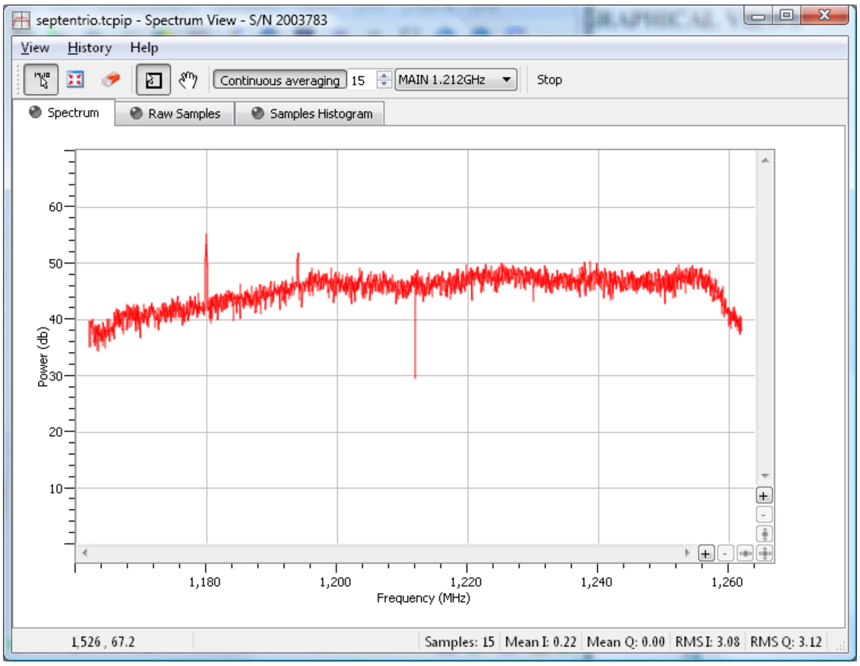


Figure 25: RF spectrum plot example in RXControl

Additionally DSP (Digital Signal Processing) SW Packages can be used for further analysis.

Task 4.3.3 / 4.3.3.7		Analysis based on RF interference detection and mitigation algorithm built-in receiver						
Device	Source Data	Output Data	Data for Database Upload	Purpose (Type of Operation)	Tools	Method/SW	Description	Status of method/ sw
GNSS Receiver Septentrio AsteRx4	SBF and RINEX files	Symptom(s) that RF spectrum is disrupted.	Yes	Ca	Septentrio RxControl SW PPSDK Matlab, Octave etc.	SW for comparison and statistics computation DSP SW	See section 2.4.2.7	R, W

Table 24: Methods and tools suitable for sub-task 4.3.3.7

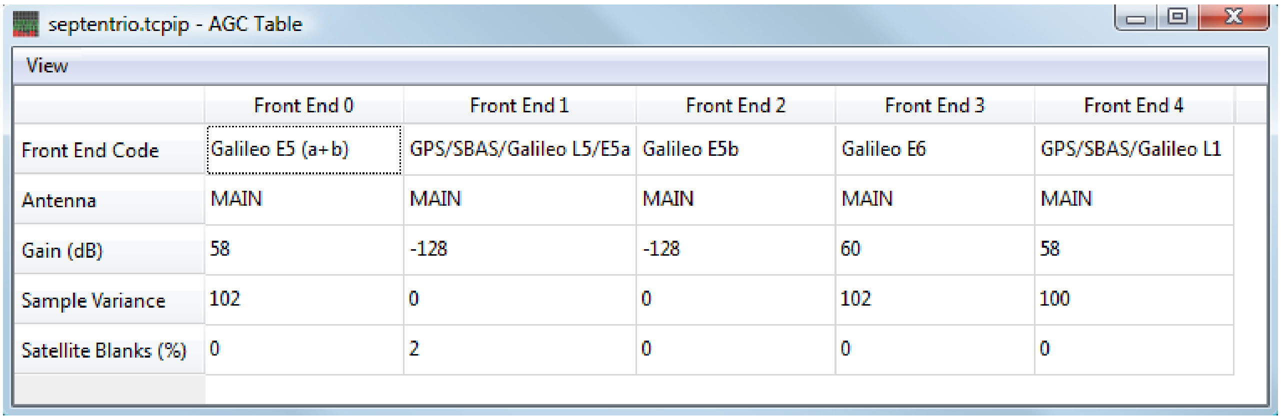
2.4.2.8 Analysis based on AGC level evaluation (Task 4.3.3.8)

The AGC section is an adaptive system, implemented as a feedback loop, to increase the dynamic range to control the quantization levels, and optimize the ratio between the quantization threshold and RMS noise (as described in the quantization section):

For automatic gain control, different implementations can be used. The most common implementation is to adjust the signal gain depending on incoming signal levels (i.e. an estimate of the noise standard deviation). This AGC mode operates completely in the analog domain. An alternative approach is to use the ADC output levels to adjust the AGC gain by preserving the Gaussian shape of ADC output samples, and mapping the ADC output power to ADC input power.

However, classical implementations of automatic gain control do not work well in the presence of continuous wave (CW) interference, and the SNR degradation is about 10 dB for a Jammer-to-Noise ratio (J/N) of 20 dB. To mitigate this effect, the quantization interval can be dynamically adjusted, or overquantization can be used to increase the AGC dynamic range. In the presence of pulsed interference, using additional quantization bits allows for techniques such as digital blanking to be implemented: the quantized value is compared to a threshold (which is often dynamic), and the samples that exceed the threshold are set to zero.

The columns of the AGC table depicted in Figure 26 contain the real-time values of the Septentrio Receiver. Each column represents an analog front end part. The first row of every column is the front end code, which gives the signals that can be tracked with the front end part. The next row gives the antenna to which the front end is connected to.
The third row gives the actual front end gain in dB. The fourth row indicates the normalized variance of the IF samples. The nominal value for this variance is 100. The last row gives the percentage of samples being blanked by the pulse blanking unit. This field is always 0 for receivers without pulse blanking unit.



	Front End 0	Front End 1	Front End 2	Front End 3	Front End 4
Front End Code	Galileo E5 (a+b)	GPS/SBAS/Galileo L5/E5a	Galileo E5b	Galileo E6	GPS/SBAS/Galileo L1
Antenna	MAIN	MAIN	MAIN	MAIN	MAIN
Gain (dB)	58	-128	-128	60	58
Sample Variance	102	0	0	102	100
Satellite Blanks (%)	0	2	0	0	0

Figure 26: AGC Table example in RXControl

Task 4.3.3 / 4.3.3.8				Analysis based on AGC level evaluation				
Device	Source Data	Output Data	Data for Database Upload	Purpose (Type of Operation)	Tools	Method/SW	Description	Status of method/ sw
GNSS Receiver Septentrio AsteRx4	SBF files	Consistency check and consistency parameters	Yes	Ca	Septentrio RxControl SW PPSDK RAIM prediction tool	SW for comparison and statistics computation	See section 2.4.2.8	R, W

Table 25: Methods and tools suitable for sub-task 4.3.3.8

2.4.2.9 Analysis based on correlator output data (Task 4.3.3.9)

Some GNSS receiver provides also correlator outputs which can reveal some additional (detailed) information of received GNSS signal quality and thus of RF interference in measured spot.
In case of Septentrio AsteRx4, the Prompt/Punctual branch correlator outputs are available through *IQCorr* packets.
The list of appropriate methods for utilization of correlator output is not finished now, however following ones are supposed at least:

- Estimation of GNSS signal power and its analysis in time
- Visualization of constellation diagram and its analysis in time

- Analysis of bit error rate of received navigation message

INPUT: SBF file with *IQCorr* packets

Methodology:
The Prompt/Punctual branch correlator outputs will be extracted from SBF file and analyzed. Matlab/Octave is supposed for next correlator output processing and visualization.

Task 4.3.3 / 4.3.3.9					Analysis based on correlator output data			
Device	Source Data	Output Data	Data for Database Upload	Purpose (Type of Operation)	Tools	Method/SW	Description	Status of method/ sw
GNSS Receiver Septentrio AsteRx4	SBF file with IQCorr package	Consistency check and consistency parameters	Yes	Ca	Septentrio Rx Control SW Septentrio PPSDK Matlab/Octave	See section 2.4.2.9	See section 2.4.2.9	R, W

Table 26: Methods and tools suitable for sub-task 4.3.3.9

2.4.3 Identification of factors disturbing GNSS signal and position solution based on analysis of recorded RF I/Q samples and RF spectrum

2.4.3.1 Analysis based on GNSS SW receiver implementation (Task 4.3.4.1)

It is supposed that in the frame of task 4.3.4.1 a GNSS SW receiver will be used primarily for analysis of multipath phenomena. The method is based on the fact that a mutlipath mitigation algorithm (which is an inherent part of tracking loop detectors in advanced receivers) can estimate current mutlipath situation, thus the output of such estimation can be used for multipath phenomena suppression. However, multipath information is not always available on receiver output. By employing the SW receiver not only various mutlipath mitigation algorithms can be utilized but also the receiver can be modified to be multipath related information available as a part of receiver output.

The analysis will be done in post processing mode on recorded RF samples in one of the GNSS bands (L1, L2, L5). The receiver will be processing a recorded signal and will be configured to provide tracking loops outputs together with estimated parameters coming from the multipath mitigation algorithm. On the other hand, it is not necessary to have complete observables or even PVT solution at receiver output. Thus, this SW receiver implementation can be simpler here if compared with ordinary GNSS receivers.

The weak point of this analysis consists in the need of modifications (implementation) of SW receiver. The good knowledge not only in GNSS signal processing but also in the particular SW receiver project (in case of SW receiver adoption) is necessary for the success.

There is wide range of GNSS software receiver projects or GNSS libraries which can be used as a basis for this task analysis, someones are even with an open licence. As an example, it is worth to mention GNSS-SDR (http://gnss-sdr.org/) written in C++ and SoftGNSS (https://github.com/kristianpaul/SoftGNSS) written on Octave (Matlab).

Basic description of GNSS signal tracking processing

The *I*, *Q* outputs are used to create detector outputs for carrier tracking and code tracking.

Assuming that I_D and Q_D represent parameters obtained thanks to the data channel and I_P and Q_P represent parameters obtained from the pilot channel. Several discriminators can be created by the combination of those parameters.

Assuming that *A* is the amplitude associated with the total (data and pilot) signal power *P* ($A = \sqrt{2P}$), C_D and C_P are the spreading codes for the data and pilot channels respectively, f_{IF} is the Intermediate Frequency, τ and ϕ are the signal code delay and carrier phase and n is the filtered Gaussian noise.

Above mentioned parameters can be described as

$$I_D(t) = \frac{A}{2\sqrt{2}} R_D(\epsilon_\tau) D(t-\tau) sinc(\pi \Delta f T) \cos(\epsilon_\phi) + n_{I_D}(t)$$

$$Q_D(t) = \frac{A}{2\sqrt{2}} R_D(\epsilon_\tau) D(t-\tau) sinc(\pi \Delta f T) \sin(\epsilon_\phi) + n_{Q_D}(t)$$

$$I_P(t) = \frac{A}{2\sqrt{2}} R_P(\epsilon_\tau) sinc(\pi \Delta f T) \cos(\epsilon_\phi) + n_{I_P}(t)$$

$$Q_P(t) = \frac{A}{2\sqrt{2}} R_P(\epsilon_\tau) sinc(\pi \Delta f T) \sin(\epsilon_\phi) + n_{Q_P}(t)$$

The traditional carrier tracking discriminator, called Costas discriminator, is obtained by multiplying *I* and *Q* for data channel and normalizing it by $I_D^2 + Q_D^2$:

$$D_{Data\_Costas} = I_D Q_D = \frac{A^2}{16} R_D^2(\epsilon_\tau) \sin(2\epsilon_\phi).$$

Another classical discriminator, the arctangent one, is defined as follow, with an operational range of $[-\frac{\pi}{2}; \frac{\pi}{2}]$:

$$D_{Data\_arc} = \arctan(\frac{Q_D}{I_D}) = \epsilon_\phi$$

More advanced discriminators, coherent discriminator, are created thanks to pilot channel and normalized by $\sqrt{I_D^2 + Q_D^2}$:

$$D_{Pilot} = Q_P = \frac{A}{2\sqrt{2}} R_P(\epsilon_\tau) \sin(\epsilon_\phi)$$

And

$$D_{Pilot\_arc} = arctan2(Q_D, I_D) = \epsilon_\phi$$

With an operational range of $[-\pi; \pi]$, twice as large as the traditional arctangent discriminator.

Those four discriminators may be used on data/pilot receivers.

Algorithms to use these parameters already exist but further development can be done to improve performance and robustness.

Multipath mitigation algorithm

Following sections briefly describe multipath mitigation (detection) algorithms with pros and cons discussion.

Narrow correlator

A narrow correlator is a basic algorithm for suppression of multipath phenomena which utilizes just two correlator branches, Early and Late. It is based on the fact that the bias as a consequence of correlation peak distortion due to multipath is smaller if Early-Late spacing is also small (compared to ordinary Early-Late spacing of 1 chip). Typical Early-Late spacing of narrow correlator is about 0.1 chip.

For purpose of this analysis the utilization of Narrow correlator is a suitable method due to its simplicity. The way how to indicate the signal with multipath can be based on comparison of outputs of two correlators (discriminators), first, the ordinary correlator with Early-Late spacing of 1 chip, second, the Narrow correlator with spacing of 0.1 chip. In case of no multipath environment, both correlators have to have identical outputs (neglecting the noise). In case of multipath environment, the outputs of ordinary correlator will be more impacted than narrow correlator output, i.e. more intensive multipath bigger difference in both outputs.

Multipath Elimination Technology (MET)

This method modifies the tracking loop to limit the effect of multipaths by playing on the correlation of the loop. The method, also called early-late technical slope, consists of placing two pairs of correlators closely spaced on either side of the correlation peak. These correlators are used to model the shape and the two slopes close to the peak of correlation of the direct path and thus to determine it. The fast correlation channels are then synchronized with the point where the slopes intersect which is the correlation peak of the direct path.

This technique directly reduces the effects of multiple paths and eliminates the need for hardware reduction technologies (such as a choke ring antenna that is not always adapted to the mission or positioning of the receiver)

Conversely, this method has the disadvantage of complicating the architecture of the receiver (and therefore of increasing its cost). Moreover, this technique is not very effective when the difference between the reflected signal and the direct signal is small, as well as in the particular case of the "alternate" path, where the direct path is masked and only a reflected path is received. In this case, the correlation peak of the direct path does not exist, it is only a correlated peak shifted from the reflected path. Correlators in such case cannot correct the induced delay and the pseudo-range measurement remains erroneous.

More information can be found in [21], [22].

Multipath Estimating Delay-Lock-Loop (MEDLL)

This method belongs to the family of methods using the estimation of the maximum likelihood on the direct and reflected signals in order to distinguish them and to attenuate them.

At the input of a GNSS receiver, the signal composed of a direct path and *M*–1 reflected paths can be modelled as follows

$$r(t) = \sum_{i=1}^{M} A_i c(t-\tau_i)e^{j\varphi_i} + n(t)$$

where (A_i, τ_i, φ_i) the amplitude, the delay and the phase of the i^{th} path and n the Gaussian noise.

The aim of the method is to estimate

$$\sum_{i=2}^{M} A_i c(t-\tau_i)e^{j\varphi_i}$$

That represents the contribution of multi-paths in order to reduce the impact on position measurement.

For this, during the estimation of the signal parameter, the reconstructed components of the multi-paths are subtracted from the received signal in order to obtain a good estimate of the direct path.

The problem is formulated as a problem of statistical estimation of the unknown parameter

$$\theta = (A_1, A_2, \dots, A_M, \tau_1, \tau_2, \dots, \tau_M, \phi_1, \phi_2, \dots, \phi_M)$$

The best estimates are those that maximize the likelihood function:

$$P(r;\theta) = c\,\exp\left\{-\frac{1}{2\sigma^2}\int_{T_0} |r(t) - \hat{A}_1 c(t-\hat{\tau}_1)e^{j\widehat{\phi_1}} - \sum_{i=2}^{M} \hat{A}_i c(t-\widehat{\tau_i})e^{j\widehat{\phi_i}}\ |^2 dt\right\}$$

In practice, when the signal is sampled, this integral may be replaced by the sum on all the samples in the time interval T_0 considered. It is equivalent to minimize the mean error squared between the signal received and its estimated version thanks to the logarithmic function of likelihood:

$$\Gamma(\theta) = \int_{T_0} |r(t) - A_1 c(t-\tau_1)e^{j\phi_1} - \sum_{i=2}^{M} \hat{A}_i c(t-\widehat{\tau_i})e^{j\widehat{\phi_i}}\ |^2 dt$$

In the case of the MEDLL, the partial derivatives are recovered with respect to the parameters of the signal set to zero of the preceding equation.

This gives a set of non-linear equations. To solve them, MEDLL approximates the global cross correlation function by using a set of reference correlation functions with certain delays, phases and amplitudes

$$R_{rc}(\tau) = \sum_{i=0}^{M} R_i(\tau)$$

where $R_i(\tau)$ is the component of $R_{rc}(\tau)$ corresponding to the i^{th} path.

$R_{rc}(\tau)$ is calculated at the delay $\tau = k\Delta\tau$ in a parallel bank of correlators. The cross-correlation values $R_{rc}(k\Delta\tau)$ are the input of the DSP which solves the MEDLL equations.

Steps are:

- Step 1: Initialization : Calculation of the correlation function $R_{rc}(\tau)$, find the maximum (called peak 1) and the corresponding delay, amplitude and phase.
- Step 2: Successive cancellation of the multi-paths: Subtract the contribution of the calculated peak 1 to obtain a new approximation of the correlation function:

  $$R_{rc}^{(1)}(\tau) = R_{rc}(\tau) - \widehat{A_1} R_c(\tau - \widehat{\tau_1})e^{j\widehat{\phi_i}}$$

  Find the new peak of the residual correlation function $R_{rc}^{(1)}(\tau)$ and the corresponding delay, amplitude and phase. Subtract the contribution of this peak 2 to $R_{rc}(\tau)$ in order to find a new estimate of the peak 1.

- Step 3: Convergence: Repeat step 2 until a convergence criterion is defined.

The advantages of MEDLL are to improve performance over other methods using correlation such as the use of a narrow-correlator receiver (a simplified version of MET).

Its disadvantages are similar to those of the MET method: complexification of the receiver requires significant algorithmic calculations and does not correct the alternate paths.

Additional information can be found in [23], [24], [25].

Multipath Mitigation Technology (MMT)

This technique was developed for the case where there is only one reflected path. It uses a nonlinear transformation on multi-path parameters to allow a fast calculation of the log-likelihood function for 2 paths. Moreover, this function is maximized according to 4 new parameters instead of the 6 conventional parameters (delay, phase and amplitude of the 2 paths).

There is therefore only one multi-path in addition to the direct signal, this baseband signal is separated between its real part *x*(*t*) and its imaginary part *y*(*t*). The logarithmic likelihood function with the 6 classical parameters is written as follows:

$$\Gamma = \int_{T_0} [x(t) - A_1 c(t-\tau_1)\cos(\phi_1) - A_2 c(t-\tau_2)\cos(\phi_2)]^2 dt + \int_{T_0} [y(t) - A_1 c(t-\tau_1)\sin(\phi_1) - A_2 c(t-\tau_2)\sin(\phi_2)]^2 dt$$

To minimize the above function, the following simplification is used:

$$a = A_1\cos(\phi_1)\,, b = A_2\cos(\phi_2)\,, c = A_1\sin(\phi_1)\,, d = A_2\sin(\phi_2)$$

Applying this transformation, the following minimization problem is obtained:

$$\Gamma = \int_{T_0} [x^2(t) + y^2(t)]dt + (a^2 + b^2 + c^2)R_c(0) - 2aR_{xc}(\tau_1) - 2bR_{xc}(\tau_2) + 2abR_c(\tau_1 - \tau_2) - 2cR_{yc}(\tau_1) - 2dR_{yc}(\tau_2) + 2cdR_c(\tau_1 - \tau_2)$$

This new function is quadratic in *a*, *b*, *c* and *d*. By setting the partial derivatives with respect to all these parameters to zero, we obtain a linear system. For each pair of values τ_1 and τ_2, this system can be solved explicitly for the minimizing values of *a*, *b*, *c* and *d*.

The MMT algorithm can be summarized as follows:

- Step 1: Search in the domain(τ_1, τ_2): For each point of the domain, look for the values of *a*, *b*, *c* and *d* which can minimize Γ.
- Step 2: Identify the point $(\tau_1, \tau_2)_{ML}$ where the smallest of these minima can be found, as well as the corresponding values of *a*, *b*, *c* and *d*.
- Step 3: Compute the estimates of $\hat{A}_{1,ML}$, $\hat{A}_{2,ML}, \hat{\phi}_{1,ML}$ at $\hat{\phi}_{2,ML}$ using the inverse transformations of *a*, *b*, *c* and *d*.

The advantages of MMT are the simplification of calculations for the log-likelihood function and the use of fewer parameters for them.

The major disadvantage of this technique is that it does not make it possible to detect or reduce more than one reflected path. It cannot therefore be implemented to respond to the case of a difficult environment where several reflected paths will reach the receiver.

Additional information is avalable in [24].

Weighted Multipath Estimating FFT (WMEFFT)

This technique is carried out in two stages:

1. Use the Fourier transform FFT to estimate the parameters of the reflected signals.
2. Apply a weighting to correct the error induced by multi-paths.

The first step is carried out as follows:

A number of samples of a received code chip are correlated with a code replica that varies circularly

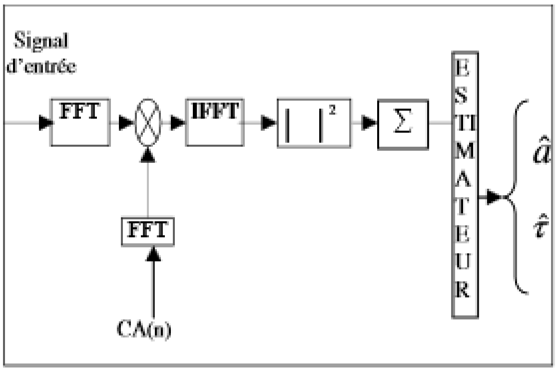


Figure 27: Schematic diagram of WMEFFT

This operation can be expressed as follows:

$$R(m) = \sum_{n=0}^{L-1} x(n) \cdot CA((n+m)_L)$$

This operation is a circular convolution, and this operation becomes a product once transposed in the frequency domain. We can therefore write the previous operation as follows:

$$R(m) = x(n) \otimes CA(-n) = F^{-1}[F(x(n)) \cdot F(CA(n))^*]$$

Where F and F^{-1} represent respectively the discrete Fourier transform and its inverse.

Thanks to this transformation, the parameters of the signal can then be estimated.

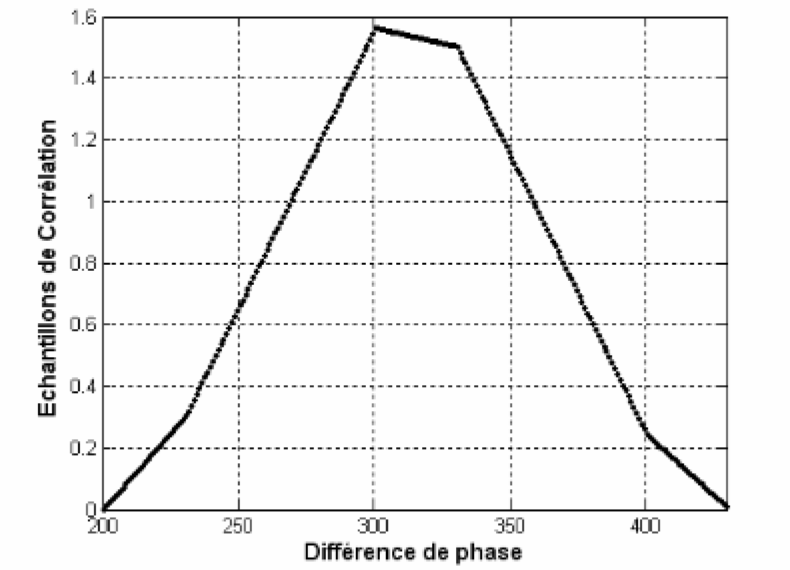


Figure 28: Correlation samples for the WMEFFT technique (fft 256 points)

Once this estimate has been made, a weighting process is used, with the code being controlled by 4 correlators.

The advantages of this technique are the use of a single branch for multi-path detection, which reduces complexity compared to other detection methods. It can also detect more multi-paths than the MEDLL technique.

The disadvantages are the addition of several correlators to correct these multi-paths. Additional information can be found in [26].

2.4.3.2 Analysis based on evaluation of RF sample histogram (Task 4.3.4.2)

See section 2.4.2.7

Task 4.3.3 / 4.3.3.2			Analysis based on evaluation of RF sample histogram					
Device	Source Data	Output Data	Data for Database Upload	Purpose (Type of Operation)	Tools	Method/SW	Description	Status of method/ sw
GNSS Receiver Septentrio AsteRx4	SBF files with *BBSamples* packets	Consistency check and consistency parameters	Yes	Ca	Matlab/Octave	SW for histogram computation and its comparison with Gauss distribution	Since GNSS signal is very weak on GNSS receiver input the received GNSS signal is buried in noise. In case of no additional RF interference, the noise is white and has Gaussian distribution. The method does a check if in specified interval the measured RF signal has Gaussian distribution	W
RPS Spirent/TeleOrbit	MAT files with RF samples (original RPS file will be transformed into Matlab MAT files)	Consistency check and consistency parameters	Yes	Ca	Matlab/Octave	SW for histogram computation and its comparison with Gauss distribution	Same as above.	W

Table 27: Methods and tools suitable for sub-task 4.3.4.2

2.4.3.3 Analysis based on power spectral density evaluation (Task 4.3.4.3)

See section 2.4.2.7

Task 4.3.3 / 4.3.3.3			Analysis based on power spectral density evaluation					
Device	Source Data	Output Data	Data for Database Upload	Purpose (Type of Operation)	Tools	Method/SW	Description	Status of method/ sw
GNSS Receiver Septentrio AsteRx4	SBF files with *BBSamples* packets	Consistency check and consistency parameters	Yes	Ca	Rx SW PPSDK	Septentrio Rx SW can compute and visualize Power Spectral Density from samples in *BBSamples* packets	Since GNSS signal is very weak on GNSS receiver input the received GNSS signal is buried in noise. In case of no additional RF interference, the noise is white and has Gaussian distribution. The method does a check if in specified interval the noise is white, i.e. no extra RF interference is apparent in spectral domain. See section 2.4.2.7	R
RPS Spirent/Te	MAT files with RF	Consistency check	Yes	Ca	Matlab/Oct	Power Spectral	Same as above.	W

Task 4.3.3 / 4.3.3.3			Analysis based on power spectral density evaluation					
Device	Source Data	Output Data	Data for Database Upload	Purpose (Type of Operation)	Tools	Method/SW	Description	Status of method/ sw
leOrbit	samples (original RPS file will be transformed into Matlab MAT files)	and consistency parameters			ave	Density has to be computed from the RF samples of given interval.		
Aaronia spectrum analyzer Spectran V5	output files .dat (RF spectrum) and .tag (time scale) or format based on TCP/IP communication on port 65002 (description is available)	Consistency check and consistency parameters	Yes	Ca	Matlab/Octave	Visualization of RF spectrum, interference detection through applied RF spectrum mask.	Same as above.	W

Table 28: Methods and tools suitable for sub-task 4.3.4.3

2.4.4 Tools and methods for evaluation data from IP camera

Currently, there are several software libraries and tools aimed to implementation of image processing algorithms. MATLAB and openCV (software library written in C++) belongs to the most common cross-platform solutions. Unlike Matlab, openCV solution is free (BSD license) for both academic and commercial use, so it will be used in this project. In addition, OpenCV solution is faster, more powerful and implements more algorithms for image processing and object classification. This library can be used on Windows, Linux, Mac OS, iOS and Android.

It is assumed that we will use some of the following image processing algorithms:

- brightness correction based on shape of histogram
- image filtering (smoothing) by convolution kernel
- image segmentation using edge detection, thresholding, region growing
- dilation and erosion due to removing area with insignificant size (small objects, degradation)
- object classification
- using algorithms for image comparison

Creation of the visibility mask is one of the tasks in this project. Currently, there are several works, that are focused this theme. The following paragraphs provide a brief overview.

IFSTTAR has developed the PREDISSAT tool that provides captured data from real railway environments, specifically mask angles of embankments and other obstacles. Originally developed for the use of linear camera [25], the more recent algorithms developed are devoted to fish-eye images acquired with a camera oriented towards the sky in order to present the same “view” as the GNSS antenna. The image processing algorithms developed in previous projects, by the laboratory, detect automatically the horizon line, i.e. the pixels separating the sky area and the masking obstacles [28]. The pixels separating sky and non-sky areas are drawn in red on the right image in Figure 29 and can be translated into elevations all around the antenna after characterization of the distortion caused by the lens of the camera. For each azimuth around the antenna, the height of obstacles is then quantified (in terms of elevation degrees).

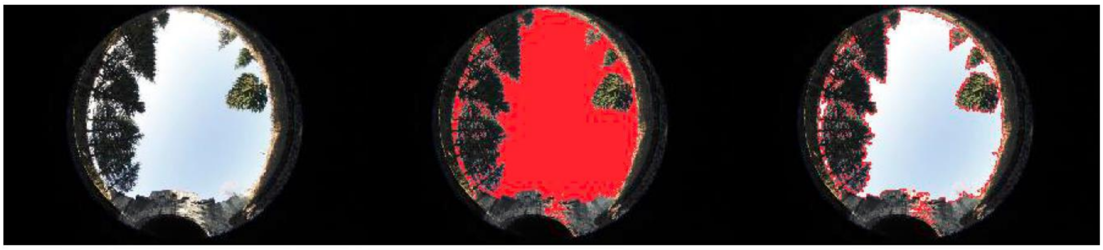


Figure 29: Fish-eye images and processing: acquired image (left), image with sky detection in red (middle), image with horizon line in red separating sky and non-sky areas (right). Processing performed in the Satloc project [29].

When replacing satellite elevation and azimuth in the image, satellite state of reception (LOS/NLOS) can be defined for each satellite. However, it is important to notice that either the placement of satellites in the image or the translation of the skyline into elevation angles require a calibration step in order to take the distortion of the lens into account.

This sky detection algorithm is capable of real-time processing.

Other teams have investigated the topic in the last years and implemented other segmentation methods for the same objective in road applications: the PLAN team of Calgary University [30], the Tokyo University of Marine Science and Technology [31], the ENAC [32] and the University of Nottingham [33].
All of these works are performed by academic teams and, as far as we know, none of them provides a commercialized solution. The use of the processing sometimes differs.

Next step are automatic objects recognition in order to distinguish the nature of the obstacles.

2.5 PROPOSAL OF MEASURED DATA SYNCHRONIZATION

2.5.1 Introduction to synchronization

The STARS measurement campaign aims at gathering data relevant to characterization of railway environment with respect to use of GNSS. As a wide range of dissimilar measuring equipment is considered to be used during this measurement campaign, measured or post-processed data need to be synchronized before they can be analyzed. The aim of this section is to discuss the problem of data synchronization at a general level.

DATA SYNCHRONIZATION

This section provides a general discussion about data synchronization and two basic steps that need to be done before the data from various measuring equipment can be compared. During the first step, data are timestamped using a common time scale and accuracy of timestamp is evaluated. The second step focuses on using the timestamped data to provide data that pertains to almost the same time instant.

It is assumed that measuring equipment generates data at discrete time instants that are not necessarily periodic and synchronous. It should be stressed that although some pieces of measuring equipment could be triggered periodically to make a measurement, there are measuring equipment that are triggered by irregular events and cannot be made synchronous in any case. An example of two pieces of measuring equipment is illustrated in Figure 30. The time instants $t_{\mathrm{A},k}$ denote physical times when the data were generated by the measuring equipment A and time instants $t_{\mathrm{B},k}$ denote the same for the measuring equipment B. The index k is just a counter. In order to compare data from the measuring equipment A with data from the measuring equipment B that was generated at approximately the same time instants (i.e. approximately the same position of the train), the timestamps $\hat{t}_{\mathrm{A},k}$ and $\hat{t}_{\mathrm{B},k}$ need to be assigned to the data.

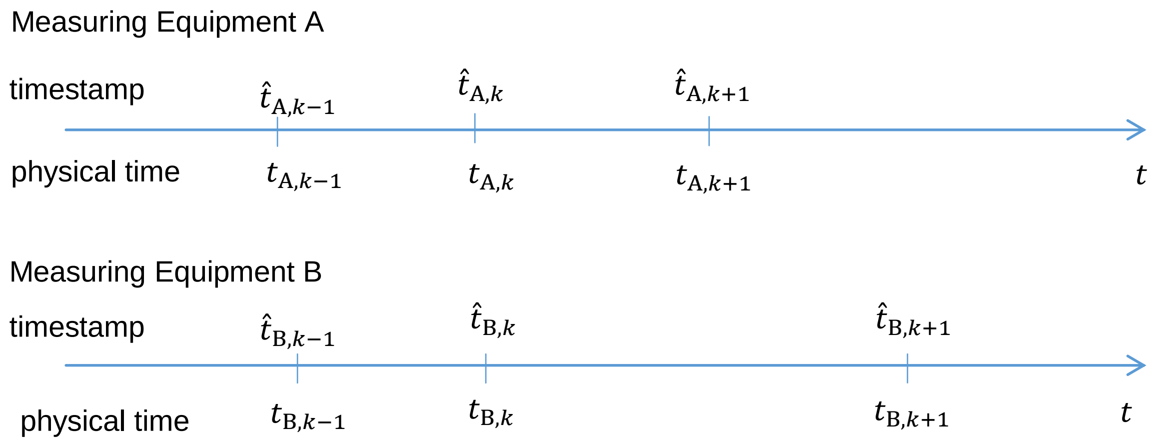


Figure 30: An illustration of data generated by two pieces of measuring equipment

The timestamps can be assigned either in on-line or off-line mode:

1. In on-line mode, the timestamps are assigned directly when the data is generated by the measuring equipment during the measurement campaign. Note that it might be necessary to determine different priority levels of assigning timestamps to data in this mode.
2. In off-line mode, the timestamps are assigned after the measurement campaign during post-processing. This mode can be used only for a complex data that inherently contains timestamps inside them, e.g. raw data from GNSS receivers or RF samples of GNSS signals.

The assignment of timestamps in on-line mode is tightly connected to the arrangement of the measuring equipment discussed in [34]. In centralized arrangement, the timestamps are assigned by a real-time processor. In this case, the accuracy of timestamps depends heavily on processing and transmission delays. Although the exact values of these delays are not usually known, an upper bounds need to be determined at least. On the other hand, each piece of measuring equipment assigns timestamps to its data in distributed arrangement. The problem with processing and transmission delays is not so pronounced, but local clocks of measuring equipment need to be synchronized. The clock synchronization ensures that the timestamps derived from different clocks can be meaningfully ordered. Due to various reasons, the timestamps are never perfect and can differ from the actual time the data were generated. Certain common timestamp accuracy ϵ_{ts}, i.e. $|t_k - \hat{t}_k| \leq \epsilon_{\mathrm{ts}}$ is assumed for each piece of measuring equipment. Note that when timestamping is performed off-line it might be also necessary to convert timestamps to a common time scale. After this step, all timestamps should be given in one common time scale and should have a known accuracy.

As the measured or post-processed data are possibly sampled at different time instants by different measuring equipment, the second step of data synchronization aims at identifying data that belongs to the same time instant or rather the same narrow time interval. Three cases can be distinguished:

1. Data with exactly the same timestamps are available for all considered measuring equipment. Such data set can be analyzed and compared directly. This could be the case for GNSS receivers synchronized by 1PPS pulse but otherwise it is a rare case.

2. Data with timestamps that differ only up to the timestamp accuracy are available for all considered measuring equipment. As differences in timestamps are below the accuracy, these data set can be also considered to be generated at almost the same time instant and can be analyzed and compared directly as well.
3. No complete data set containing all considered measuring equipment can be simply formed as the timestamps do not satisfy any of the two previous cases. This can be solved either by excluding this time instant from analysis or by interpolating missing data using neighboring data in the time series.

As the distributed arrangement of measuring equipment is slightly preferred in [34], the following two subsections focuses on clock synchronization and data interpolation.

Clock synchronization

If the measuring system derives timestamps from more than one clock, the need for clock synchronization shall be analyzed as every clock suffers from a drift. A relationship between times of a reference clock C1 and another clock C2 at the same time instant can be generally described as

$$t_2 = f(t_1),$$

where t_1 is the reading of the reference clock C1, t_2 is the reading of clock C2, and $f: \mathbb{R} \mapsto \mathbb{R}$ is a function that describes evolution of time t_2 in dependence on t_1. If C1 and C2 are ideal and synchronized clocks, it holds $f(t_1) = t_1$. The quality of clock is usually expressed using its accuracy and stability. The accuracy measures a time drift in [s/s] and stability denotes the variation in the drift itself.

The need for clock synchronization should be evaluated based on:

- the required maximum error of clock synchronization,
- the drift of clocks,
- and the time span over which the clocks should be kept synchronized to the prescribed precision.

According to [34] the maximum clock synchronization error of 1 [ms] is considered to be acceptable. Although the maximum train speed on the identified representative lines is 140 [km/h] [10], the proposed maximum clock synchronization error is evaluated using overestimated maximum train speeds of 300 [km/h]=83.3 [m/s] or 250 [km/h]=69.4 [m/s]. In such cases 1 [ms] clock synchronization error corresponds to longitudinal position error of 8.33 [cm] or 6.94 [cm], respectively. If the clocks were synchronized only at the beginning of the measurement campaign of approximately two month duration, their drift would have to be less than 2e-10 [s/s] to stay within the required precision of clock synchronization. Since such small drift is hard to achieve, the clock synchronization needs to be performed continuously throughout the measurement campaign.

The synchronization of several clocks can be basically done in the following two ways:

1. internally by synchronization of several clocks with each other,
2. or externally by synchronizing several clocks to one reference clock.

The internal clock synchronization is usually used in higher level of timekeeping with very precise clocks of a similar quality. The external clock synchronization assumes a reference clock of a higher quality and it is definitely more suitable for the measuring system.

There are several clock synchronization algorithms available. In the following there are listed some examples of clock synchronization algorithms that suppose the time sources are not accurate or noisy:

- Cristian's algorithm [35],

- the Berkley algorithm [36],
- Marzullo's algorithm [37].

As most measuring equipment has Ethernet networking capabilities, NTP and PTP are foreseen to be used for clock synchronization [34]. In both protocols some of the above mentioned algorithms are internally used in client-server architecture to synchronize several slave clocks to a master clock. A typical structure of the NTP or PTP network is given in Figure 31. A grandmaster clock is externally synchronized using a GPS receiver or an NTP server and represents the time source for the devices on a local network. The time is provided to slaves through a boundary clock that resides usually in a switch that allows several slaves to be connected.

The NTP is used worldwide to synchronize devices over the Internet. It can be also used locally to synchronize the clocks of measuring equipment to the UTC time and a particular time zone. The synchronization actions are performed automatically in a gradual way such that the time is monotonically increasing and significant step changes are prevented. The quality of clock synchronization can be negatively impacted by network congestion and asymmetric routes. The expected accuracy of clock synchronization over local area network is in order of milliseconds.

A better clock synchronization can be achieved using PTP that is meant for clock synchronization of servers and devices on a local network with accuracy of synchronization in order of microsecond. Although PTP provides a better accuracy, it is not clear whether measuring equipment will support it on the hardware level.

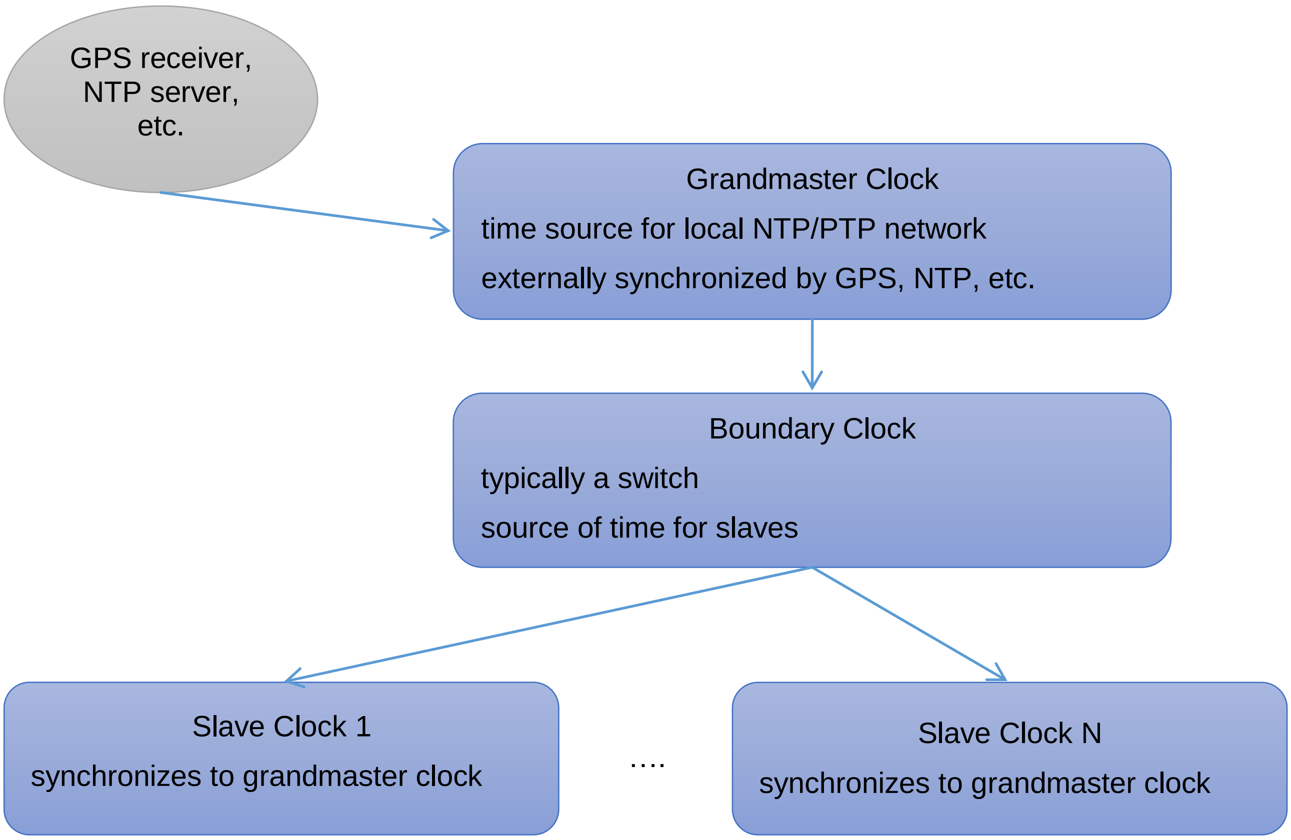


Figure 31: A typical structure of a NTP or PTP clock synchronization of devices on a local network

Data interpolation

Even if the measuring system consists of one clock or several synchronized clocks, measuring equipment may produce data asynchronously and at different rates (e.g. reading of absolute reference position tag is an asynchronous event). It can also happen that measured data are missing at some time instants for one reason or another.

A strategy for data interpolation between two timestamps needs to be chosen. In order to compute any interpolation a model for possible evolution of data is need. The availability and complexity of such model should be the main factor when considering the strategy of interpolation. For example, the data for the ground truth can be generated virtually for any timestamp if absolute reference position and relative measurements are available, because constant speed model of train motion is simple to use and reasonable for sufficiently short time intervals. On the other hand, a model that would allow two images to be interpolated is much harder to develop and use.

Communication with Measuring Equipment

Communication with the measuring equipment for purpose of timestamp synchronization or data retrieval can be divided as

1. Active– a cooperation of control unit and sensor occurs.
2. Passive – a control unit cannot modify a sensor

It should be known whether it is possible to issue a command to the measuring equipment to measure at specific predefined reference time instants and the sensors are synchronized in advance to measure data exactly at the given reference time instants. In such case the sensor has to take into account the period necessary to acquire the measurement (e.g., time to focus and open the camera shutter). If the sensor is unable of such behavior, the time to get the measurement is delayed from the command by the speed of sensor processing.

SELECTION OF SYNCHRONIZATION METHOD

The selection of synchronization methods depends on the following factors:

- the arrangement of the measuring equipment,
- the capabilities of individual measuring equipment.

These factors will be discussed in the following sections.

2.5.2 Characteristics of measuring equipment

This section deals with arrangement and properties of measuring equipment with a special focus on their synchronization.

ARRANGEMENT OF MEASURING EQUIPMENT

The following two possible arrangements of measuring equipment are presented in [34]:

- Centralized arrangement of measuring equipment – This arrangement assumes that all measuring devices are connected to a real-time processor that assigns timestamps to received measured data. Since there is only one source of timestamps, the main problem to be addressed is the latency of measuring device processing and data transmission from device to the real-time processor.
- Decentralized arrangement of measuring equipment – This arrangement supposes that each measuring device is a more or less independent subsystem that performs data acquisition and time stamping on its own. Individual subsystems are connected to a supervising processor that provides control and remote access functions but no synchronization. Although latency seems to be of no problem with this arrangement, there are several time sources that need to be kept synchronized.

These arrangements represent two opposite approaches. In [34] the decentralized arrangement of measuring equipment is preferred and justification is provided. However, a hybrid arrangement could be used as well due to diverse nature of measuring equipment and synchronization needs.

PROPERTIES OF MEASURING EQUIPMENT

A list of measuring equipment that could be used during the STARS measurement campaign is given in [34]. As not all the equipment will be used at all test sites, a common set of equipment to be used at all test sites is identified in [38]. The basic synchronization capabilities of measuring equipment are reviewed in [34]. Some additional notes are provided in the following sections.

Record and Playback system (Spirent GSS6450/GSS6425)

The measured data are saved using a proprietary format that is accessible and documented. The timestamps can be recovered by replaying the recorded signal to a receiver.

Spectrum analyzer (Aaronia Spectran HF-8060 V5 RSA)

The device consists of RF measurement unit with DSP part and PC which runs Linux operating system. The measured data (spectra) are saved into proprietary file format. The file names contain time stamps of measurement beginning; smoother time information is further included into file content together with spectra. The time stamps are obtained from operation system time. Thus, it is worth keeping the time quality through NTP (Network Time Protocol) client. Linux operation system enables to be set to arbitrary time-zone. For our purpose UTC time is the most convenient. Measured data with spectra can be replayed and analyzed with Aaronia software (RTSA suit). Further, since proprietary data format are known to STARS consortium, the measured data can be also parsed and loaded into generic mathematic program (e.g. Matlab) and visualized and analyzed there.

Camera

Since several cameras of different synchronization capabilities are considered for use in [34], a general discussion is provided only. It is necessary to consider the internal processes running inside the camera while taking an image. It includes autofocus, shutter opening/closing, and others. To get a sufficiently sharp image the shutter speed must be sufficiently fast. In any case taking an image always takes some finite time and a reference point needs to be chosen to assign a timestamp. This reference point can be considered, e.g., as

- a time the shutter begins to open,
- a time the shutter begins to close,
- time-weighted average over the duration of the exposure.

Fisheye fixed dome network camera Vivotek FE8174

This camera can use NTP client to synchronize its internal clock to UTC. The possible image format is JPEG and video formats MPEG 4, H264, and MJPEG are supported. MPEG4 and H264 can be streamed using RTP. A timestamp with a second resolution can be embedded into video or image. This timestamps can be extracted later during post-processing using image recognition software. By counting images within a second, a better than one second resolution can be achieved. However, as the internal latency of the camera is unknown, the accuracy of timestamps cannot be expected better than tenths of second.

If RTP is used to stream video in MPEG4 or H.264 format, the timestamp included in RTP packets can be used to obtain a better synchronization. Although the timestamps have NTP format, the latency between taking an image and creating a packet is unknown. Another difficulty is that these timestamps cannot be used directly. They are assumed to be used within the protocol only and thus they are not referenced to UTC. To achieve synchronization of RTP to other events, the sender and receiver reports that provide links between RTP timestamps and UTC should be combined.

Although the camera also supports a response to an event triggered by external digital input, it does not seem to be suitable for events with high rates as snapshots are sent to an FTP server or an email.

Basler ace GigE Vision acA640-300gc

This camera is capable of clock synchronization to UTC using a PTP client. Instead of using a common image of video format, the camera is sending pixel information by Ethernet. Then using a Synchronous Free Run Mode, images can be obtained at specified frame rate starting at a specified UTC time.

GoPro like camera

The synchronization capabilities of the camera will be evaluated when a camera model is chosen.

GNSS receivers

In order to get the most information possible from GNSS receiver it is assumed that a set of proprietary messages will be saved in a native GNSS receiver format such as SBF, UBX, etc. The GNSS receivers are synchronized with GPS time frame after first PVT fix and thus the receiver messages are synchronized to GPS time frame.

Absolute position reference tag readers

The absolute position reference tag events are asynchronous measurements that provide a timestamp and an ID of the reference tag.

Eurobalise and MIB readers

These readers are part of an on-board train system with its own clock. As this system should not be influenced by the measuring equipment used during measurement campaign, the clock synchronization is not possible. The time scale and accuracy of the timestamps provided by these readers have to be known to analyze the synchronization capabilities of the readers.

RFID tag reader Harting RF-R300

This reader is independent of the on-board train systems and will probably allow its internal clock to be synchronized to UTC through NTP client-server architecture.

Relative sensors (odometer, Doppler radar, IMU)

If these sensors are part of on-board train system, the same comments as for the Eurobalise and MIB readers apply. The relative sensors outside the on-board train system could be synchronized off-line by 1 PPS signal or to UTC through NTP that uses internal clock synchronization algorithms.

IDENTIFICATION OF RISKS

The arrangement of measuring equipment and its synchronization capabilities might pose a risk of not achieving the required synchronization accuracy. The potential risks can include the following:

- lack of information about an internal processing performed by measuring equipment that prevents to estimate processing delay (e.g. camera),
- a measuring equipment that use its own clock and external synchronization is not possible (e.g. train on-board unit)
- start and stop latency of hardware/software,
- timestamp conversion and precision – timestamps of each subsystems may have different possibilities of storing the timestamps, i.e., some information can be lost even when synchronizing data from several sensors, e.g. 1,2 [s] or 1,149512 [s].

This list is not complete by any means and other risks can be uncovered in future.

2.5.3 Identification of measured data to be synchronized

This section focuses on identification of measured and post-processed data that need to be synchronized. The identification was done based on deliverable [34] and discussion with STARS partners performing the measurement campaign.

DATA SYNCHRONIZATION NEED BASED ON WP2 DELIVERABLES

Deliverables D2.1 and D2.2 specify the following synchronization needs:

- Synchronization of absolute position tag (Eurobalises, MIB, RFID/optical Tags) events to GNSS measurements.
- Synchronization of a combination of absolute position tag events and relative measurements to GNSS measurements
- Synchronization of video frames or snapshots from camera to GNSS measurements

DATA SYNCHRONIZATION NEED BASED ON PARTICULAR ARRANGEMENT OF MEASURING EQUIPMENT

Each STARS partner performing the measurement campaign will use its own measuring system. It gives rise to a particular need of measured data synchronization.

AZD

AZD has proposed a hybrid arrangement of measuring equipment that includes independent measuring equipment and equipment that has internal clock synchronized using a NTP client-server architecture. The NTP server runs on a supervision processor and is synchronized using a GNSS receiver. This server provides UTC time to other selected measuring equipment connected through Ethernet.

Among two representative railways identified for performing the measurement campaign, the odometry information is obtained from train on-board system that has clock synchronized using DCF77 and external synchronization is not possible. Therefore, the clock synchronization realized using DCF77 needs to be analyzed.

SIE

Siemens has integrated a centralized measurement system responsible of synchronizing the ground truth data to the GNSS timestamp. This is a National Instrument based real-time system which collects GPS and sensor data. Independent GNSS receivers are also used (Septentrio, JavaD) and are synchronized to the GPS timestamp. It has to be analyzed the time synchronization of TeleOrbit RPS and Aaronia Spectran spectrum analyzer.

ASTS

ASTS has proposed a hybrid arrangement of measuring equipment that includes independent measuring equipment from one side (Aaronia Spectran, Spirent GSS6425, and GNSS receivers) all of which are characterized by off-line time stamping, and equipment that has its own on-line time stamping synchronized by the 1PPS signal for synchronization of odometric information with GPS time.

2.5.4 Selection of common time scale

There are several time scales in use. Some are based on Earth rotation and others are artificially constructed time scales. When collecting and evaluating data from measuring equipment, it is essential to relate the data to one common, accurate, well-defined, and reliable time scale. In this part, selected important time scales are presented. The stress is laid on the GPST scale.

GLOBAL TIME SCALES

The time scales listed in this section are based on two periodic processes: the Earth's rotation (UT, UTC) and atomic oscillations (TAI, UTC, GST, GPST).

UT

The UT framework is derived from the Earth's rotation and orbit dynamics. It consists of several individual time scales due to the fact that the Earth's rotation is not definite at all times. The rotation can be influenced by many factors, e.g., the Earth's tides, mass transport due to seasonal changes, geologic shifts, or earthquakes. A monitoring of the Earth's rotation and determination of the time scale is executed by the IERS. The changes in spin are measured by a network of observatories and radio telescopes [39] around the world.

- UT0, so called mean solar time, is derived from the instantaneous Earth's rotation determined from the direct astronomical observations.
- UT1 is one of the most used UT time scale derived from UT0 by correcting effects of movements of the Earth relative to the axis of rotation. Unequal mass distribution of the Earth causes the Earth to rotate not exactly around the axis of symmetry, but around its center of mass known as a polar motion. In UT1, this rotation drift is cancelled out.
- UT2 is a version of UT1 where periodic factors influencing the Earth's rotation, e.g. seasonal Earth's tides, are considered and removed.

TAI

TAI is a highly stable and precision time scale maintained by the BIPM. TAI is calculated based on the world's primary frequency standards while the calibrations is done based on about 250 caesium clocks and hydrogen masers in more than 65 different laboratories worldwide. Each of the caesium clocks has assigned a weighting factor according to its performance. TAI is then computed as a weighted average of current time. TAI is not generated in real time because it takes about a month to collect all data and to perform the necessary calculations. The instability of TAI compared to UT1 is about 6 orders of magnitude smaller [40].

UTC

To serve public and practice, UTC time scale that is maintained by the BIPM was defined in 1972 to compromise between the stable atomic time and inconstancy of the Earth's rotation. Due to the time drift with TAI, UTC time is adjusted by integral seconds, so called leap seconds [41], to keep the difference between UT1 and UTC less than 0.9 [s]. This means that at any time and neglecting accuracy errors, UTC and TAI differ only by an integer number of seconds. UTC is calculated using data from that same caesium clocks as TAI and applying the weighting algorithms [42]. The time computations are demanding, nevertheless, there are several centers worldwide including USNO that provide real-time estimates and can be found in [43]. Most of the UTC estimates provided by these centers are kept within 1 [µs] of UTC and some of the centers provide even better accuracy. UTC is very consistent, e.g., time prediction error of UTC is about 60 [ns] compared to 60 [ms] in case of UT2 [43].

As mentioned in [43], the concerns are rather about the instabilities in the time and frequency transfer techniques than about the accuracy of the source of UTC since accuracy of atomic clock is constantly improving.

GST

GNSS Galileo aims to provide autonomous precision navigation system. Thus, GST was created to serve as a reference time scale. GST is not corrected to match the Earth's rotation, therefore, it is continuous and no leap seconds are introduced. GST started its function on August 22nd, 1999 while aligned with UTC time. GST is maintained by Galileo Mission Segment and it is synchronized to TAI with a nominal offset below 50 [ns] [42]. There are other GNSS time scales such as GLONASS time or BeiDou time.

GPS TIME (GPST)

GPST, similarly to GST, is a continuous-time scale that is computed based on a set of terrestrial atomic clocks and atomic clocks on board the satellites. The GPST is maintained by USNO. Besides GPST, the USNO computes the real-time estimate of UTC, called USNO Master Clock or UTC (USNO). Except the leap second difference, there is a steered drift between GPST and UTC.

Specifically, GPST is steered towards the UTC (USNO) to maintain the maximum deviation of 1 [µs] modulo 1 [s]. In practice, the accuracy is much better within few hundred [ns], typically within 40 [ns] [43], [44]. Note that the difference between GPST and UTC (USNO) from the past 30 or 180 days can be found in [45]. In the NAV, the following information is included:

- the total accumulated leap seconds between GPST and UTC,
- the date of the next planned introduction of leap second(s),
- the bias between GPST and UTC (UNSO).

The offset data between GPST and UTC (UNSO) in the NAV are intended to be employed by the user segment, e.g. GPS receiver. The accuracy of the data may be decreased if the GPS Control Segment is unable to upload actual data to the satellites. A degradation of GPS accuracy to civil users directed by the Department of Defence is possible, e.g., using SA.

The beginning of GPST epoch occurred at midnight of January 5th to January 6th 1980 and GPST was set to match zero difference from UTC. At that time, the difference between the TAI and UTC was 19 seconds. Compared to other day formats, e.g. year-month-day in Gregorian calendar, GPST keeps track of week number, so called GPS weeks. It is obvious that the GPS week started at 0h 0m 0s UTC on January 6th, 1980 by week number zero (0); it is incremented by one (1) at the beginning of a new GPS week epoch, and included in the NAV. One GPS week is defined as 604800 seconds.

LEAP SECONDS

The leap seconds are determined by IERS and can be positive or negative. In 1972 for the first time, 10 seconds were added to UTC. The adjustment can be done at the end of a UTC month, preferably in June and December, or as a second preference in March and September. In case of the positive leap second, the start is at 23h 59m 60s and the end at 0h 0m 0s of the first day of the following month. The negative leap second starts at 23h 59m 58s and is followed one second later by 0h 0m 0s of the first day of the following month. The IERS should decide about introduction of a leap second at least eight weeks in advance [46].

The last leap second was added on July 1st, 2015 and the next leap second will occur on December 31st, 2016. Considering the past evolution of the leap second management, it can be supposed that from January 1st 2017 no leaps seconds will be introduced until the end of the measurement campaign, which is considered to be December 31st 2017. Nevertheless, the expected introduction of a leap second would be notified through Bulletin C of IERS, which is mailed every six months. Its last issue was sent on July 6th 2016.

The time relations of selected time scales at the present time and after December 31st, 2016 are shown in Table 29.

	From July 1st, 2015 – January 1st, 2017	From January 1st, 2017 – further notice
TAI - UTC	+36	+37
GPST - UTC	+17	+18
TAI - GPST	+19	+19

Table 29: TAI, UTC, and GPST relations due to introduction of a leap second

ANALYSIS OF DCF77 SYNCHRONIZATION PRECISION

DCF77 is a German radio service that provides time and frequency signal over longwave. The timing signal is aligned to UTC (PTB) realization of UTC time at the transmission site with error less than approximately 0.1 [ms]. The error of timing signal at the receiver is mainly influenced by transmission and processing delays. The transmission delay is given by the distance between the receiver and the transmitting antenna. The transmission delay of 3.3 [ms] can be expected for receiver 1000 [km] away from the transmitting antenna. The processing delay results from a trade-off between interference rejection and the bandwidth of the receiver. A receiver with a good interference rejection ensured by rather small bandwidth of 10 [Hz] can have time uncertainty of 0.1 [s]. Therefore, the timing precision provided by DFC77 depends also on the receiver and expected synchronization error can be significant if not corrected for the delays.

RECOMMENDATION FOR GLOBAL TIME SCALE

The problem of selecting a global time scale should be looked upon from two perspectives: the internal representation and representation for the user. In terms of the internal representations, there is no distinction between using the UTC and GPST scales as a simple conversion between those two is available. Based on the recommendations of AZD and ASTS, the GPST should be preferred. As far as the representation for the user is concerned, it should be human readable. With respect to the naming conventions of the filenames for data, the format YYYYDDDHHMM is recommended [34].

2.6 DATABASE CONCEPT PROPOSAL

The goal of this section is to present proposed database concept for storing the measured data of the STARS project. This is going to be used for uploading measured data from the three railway lines where the measurement campaigns are performed, for downloading measured data by partners responsible for their processing and analysis and for uploading results based on processed data.

2.6.1 Overview of stored data

The data stored at the STARS measurement campaign is quite diverse as it has different source of origin in terms of measurement equipment. In addition there are three different locations where the data is going to be measured. Between the three installations minimum common measurement equipment has been defined. This gives us an overview of the type and approximate size of the data to be stored; however some differences have to be also foreseen.

Based on the common minimal set of equipment [47] the different type of data has been identified.

DATA FROM MEASUREMENTS ON A DAILY BASIS

Onboard the train, the following data are intended to acquire:

- Images from cameras on the roof or in the cockpit, between every 5s to every 1min
- Raw data from different GNSS receivers
- Data from the spectrum analyzer[2]

Stationary, data from GNSS supporting servers are supposed to be saved:

- Weather data
- EGNOS data
- Precise ephemeris data

DATA FROM MEASUREMENTS ON A ONCE PER MONTH BASIS

- Data from the spectrum analyzer[2]

[2] Use of the spectrum analyser depends on mode of data recording. It's supposed to use the spectrum analyser in recording mode employing signal level trigger in daily operation and in the mode of continues recording once per month.

- RF I/Q Samples of GNSS signals from Reply Playback Systems

REFERENCE POSITION DATA

- GT
- Reference position of GNSS antenna

DATA RESULTING FROM PROCESSING OF MEASURED DATA

- virtual sensor data[3]

2.6.2 Database model

This section describes selected database model and its structure. Based on consideration carried out in [48], expected volume of measured data and different nature of all saving data as described in previous section, the decision to use directory model of database has been taken. To clarify this decision, the reasons are shortly summarized as follows.

Binary raw data files from GNSS receivers comprise many parameters bearing special and useful information for next analysis e. g. multipath residuum, interference level, AGC level etc. For this reason it seems to be advantageous to store original files with raw data. The storage of original files in directory based database compared to decoding and putting single figures (measured parameters) into relational database is much less time consuming and also more effective from perspective of stored data volume.

From perspective of stored data volume the similar situation occurs for Replay Playback System.

It could be also difficult to download such volume of decoded data originated from all measurements by any project partner.

The last but not least reason is data processing. There was not enough time even opportunity to test remote data processing and also experience of most WP4 partners involved in Task 4.3 is probably missing.

The proposed directory model should enable to download data in compressed data format, decode and processed them locally without any troubles.

The following structure for directory model of database has been proposed:

- separate directory will be created for each partner responsible for measurement; directory name will correspond to company name i.e. AZD, ASTS, SIE,

- separate subdirectory will be created for each day of measurement campaign; subdirectory name will be in format YYYYMMDD,

- separate sub-subdirectory will be created for each device; name of sub-subdirectory will be in format XX_{dec} , i.e. two-digit code specifying device according to Table 30.

These sub-subdirectories will include

- separate file (or continuous sequence of files) will be created for data from each data source (measuring device with given configuration, virtual sensor, GT data, data of Reference position of GNSS antenna, etc.),

- separate metafile with measurement description will be created for each measurement,

[3] It was proposed that data obtained from processing according to sub-tasks of Task 4.3 will be labeled as data from another virtual sensor

- Separate file (or continuous sequence of files) with signature of data file; md5 sum is supposed to be used.

To identify source of measured data, its configuration (if there is any), test site, measurement period and other important parameters the naming convention derived from RINEX version 3.02 and later [18] has been proposed and adapted to this project.

The file name consists of following fields:

Measurement site/line identification CCC: three-character code specifying the particular measurement site or line

CTB	Česká Třebová – Brno
CVO	Číčenice – Volary

Device/equipment source XXX: three-digit code specifying the particular device, which produce content of the file; 000 to 099 are devices from agreed common measurement set, 100 to 199 are ASTS specific devices, 200 to 299 are AZD specific devices and 300 to 399 are SIE specific devices. Codes 900 to 999 are dedicated for data from external sources (web servers).

004	Fish eye camera Vivotek FE8174/74V
005	GNSS receiver Septentrio AsteRx4
006	RPS system Spirent GSS6450
007	RTSA Aaronia Spectran HF-8060 V5 RSA
210	RFID reader Harting RF-R300
220	GNSS receiver JAVAD TRE G3T
221	GNSS receiver uBlox EVK M8T
910	data from CDDIS – NASA's Archive of Space Geodesy Data

Start time YYYYDDDHHMM: as a time frame is proposed GPS System Time (GPST) due to absence of leap seconds. It is required that the first measurement item in the file is aligned to one minute boundary according GPST.

Duration/period of file XXU: where XX is numerical value and U is a code of unit. E.g. 15M represents 15 minutes, 01H an hour, etc.

Frequency of measurement items XXU (or 3-char code): where XX is numerical value and U is a code of unit. E.g. 10Z represents 10Hz, 01S represents 1-seconds interval, NOA indicates, that the frequency is not applicable to these data (data frequency could be irregular and internally coded).

File format CCC: three-character code represents a file format (usually corresponds to file extensions). E.g. SBF for Septentrio binary file, UBX for uBlox binary file, RNN for RINEX navigation file, RNO for RINEX observation file, etc.

Compression CC or CCC: two or three-character code represents possible file compression. E.g. GZ, TGZ, ZIP, etc.

All file name elements are fixed length (expect the last one, compression) and are separated by an underscore "_" (except the last two items, file format and compression, which are separated with a period ".").

Here is a file name example:

CVO_005_20170291100_01H_NOA.SBF.GZ

for one hour GNSS receiver measurement on Cicenice Volary with Septentrio AsteRx4 receiver. The file contains binary Septentrio data and are compressed using G-ZIP. Data would be obtained on 29.1.2017 since 11:00AM to 12:00PM. The item frequency is indicated as NOA (NOt Applicable) since the file contains different measurement items with different frequencies.

Following tables just describe a way how the hardware + configuration are reflected in file names.

Legend		**file version: 02, 2016-11-09**
configuration groups (3bits)		
bin	***dec***	
000	0	common meas. (common to two meas. systems at least)
001	1	ASTS meas.
010	2	AZD meas.
011	3	SIE meas.
100	4	computed data in WP4.3 (computed data from database)
101	5	reserved
110	6	reserved
111	7	sources outside STARS (web servers, etc.)
configurations (5bits)		
bin	***dec***	
00000	0	conf. 0.
00001	1	conf. 1.
00010	2	conf. 2.
…	…	…
11111	31	conf. 31.
device sources (4bits)		
bin	***dec***	
0000	0	data computed from various devices in WP4.2 (computed GT data)
0001	1	not used
0010	2	not used
0011	3	not used
0100	4	camera (Vivotek or GoPro)
0101	5	GNSS rec 1: Septentrio AsteRx4
0110	6	Record and Playback System (GSS6425, GSS6450, TeleOrbit)
0111	7	RTSA Aaronia Spectran HF-8060 V5 RSA
1000	8	GNSS rec 2: JAVAD TRE-G3T
1001	9	GNSS rec 3: uBlox NEO-M8T
1010	10	reserved
1011	11	reserved
1100	12	reserved
1101	13	reserved
1110	14	reserved
1111	15	reserved

Table 30: Codes used in proposed naming convention

Configuration codes of records								file version: 02, 2016-11-09
code-com			code	c.m.	meas.sys.			
grp	conf	dev			ASTS	AZD	SIE	description
0	0	5	005	Y	Y	Y	Y	AsteRx4+Antonic+Conf00
0	1	5	015	Y	Y	Y	Y	AsteRx4+Antonic+Conf01
0	2	5	025	Y	Y	Y	Y	AsteRx4+Antonic+Conf02
2	30	5	5e5			Y		AsteRx4+Antcom+AzdConf01
2	31	5	5f5			Y		AsteRx4+Antcom+AzdConf00
2	0	8	408			Y		TRE-G3T+Antonic+Conf00
2	1	8	418			Y		TRE-G3T+Antonic+Conf01
2	30	8	5e8			Y		TRE-G3T+AntCom+Conf01
2	31	8	5f8			Y		TRE-G3T+AntCom+Conf00
2	0	9	409			Y		NEO-M8T+Antonic+Conf00
2	1	9	419			Y		NEO-M8T+Antonic+Conf01
2	30	9	5e9			Y		NEO-M8T+Antcom+Conf01
2	31	9	5f9			Y		NEO-M8T+Antcom+Conf00
0	0	6	006	Y	Y	Y	Y	RPS+2bits+30MHz+GPSL1+GPSL5
2	0	6	406			Y		RPS+4bits+30MHz+GPSL1+GPSL5
2	1	6	416			Y		RPS+4bits+30MHz+GPSL1
2	2	6	426			Y		RPS+4bits+30MHz+GPSL5
2	0	4	404			Y		FE8174V+AZDconf
1	0	4	204		Y			ASTSCam+ASTSConf
3	0	4	604				Y	SIECam+SIEConf
0	0	7	007	Y		Y	Y	HF8060+HaS+GPSL1
0	1	7	017	Y		Y	Y	HF8060+HaS+GPSL2
0	2	7	027	Y		Y	Y	HF8060+HaS+GPSL5
1	0	7	207		Y			HF8060+ASTSAntenna+GPSL1
1	1	7	217		Y			HF8060+ASTSAntenna+GPSL2
1	2	7	227		Y			HF8060+ASTSAntenna+GPSL5
2	0	0	400			Y		GT-AZD+RFID
2	1	0	410			Y		GT-AZD+MIB
3	0	0	600				Y	GT-SIE+EB
1	0	0	200		Y			GT-ASTS+EB

Table 31: Coding table for file names in database

2.7 ASSESSMENT OF DATA STORAGE CAPACITY REQUIREMENT FOR DATA STORAGE

Assessment of data storage capacity comprises of activity T4.1.7 of the STARS project. An excel based calculator was chosen In order to accommodate the flexible configuration of sensors and GNSS receivers on-board the train as described in the documents, Common minimal set of equipment and Identification of the representative Railway lines. This facilitates a configurable method of calculation that is adaptable to changes without negating any work performed prior to implementing changes. This choice of tool also facilities a dynamic estimation capability whereby the metrics of the GNSS performance measurement process can be changed and the effect on storage requirements directly observed which will assist in determining an optimum balance between the quantity of performance measurement in the field and value of particular types of measurement while fulfilling the constraints of data storage such as cost, availability and performance.

Factors which have an impact upon storage capacity are

- Number and configuration of GNSS receivers
- Number and configuration of sensors and the nature of data output by those sensors, continuous or irregular.
- Information generated from amassed data such as ground truth and PVT.
- Geographical reference data.
- Kinetic information of the vehicle to which the GNSS performance measurement is applied.
- Quantity of performance measurement in terms of duration (time) and distance.
- Overhead for total overhead for each individual record of data created.

The organisation of the calculator is shown in the diagram below, Figure 32

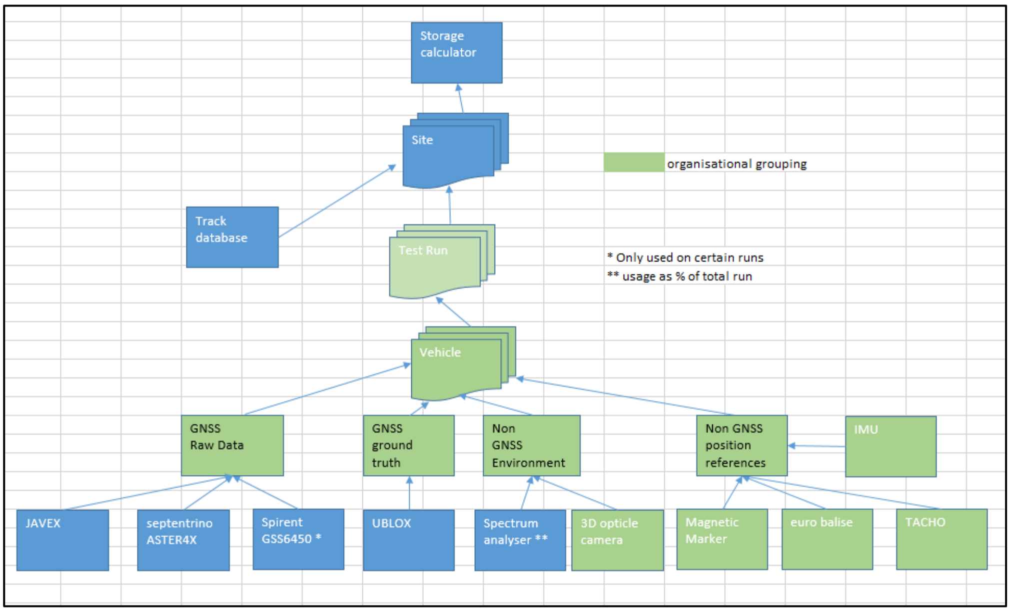


Figure 32: Structure of Storage Calculator

The First and main page is the data storage calculator page which shows a dynamic total of the storage requirements which are calculated from underlying sheets in the excel file. As shown in the diagram above the blue entities represent individual excel sheets and the green objects depict their organisation relationship however, the information that they represent has been integrated into the closet excel sheet object above it in the diagram.

The individual sheets comprising the calculator are covered individually in the following sections.

2.7.1 Data Storage Calculation

FOR ADVANCED
RAILWAY SIGNALLING

STARS WP4 Task 4.1.7 Assesment of data storage capacity requirement for the database

Vers 3

Byte overhead per individual record/sample for sensor data: 12

Calculate

Site	Sensor Data	data overhead for sample/indiv records of sensor data KB	Data in file format (GNSS receivers, Camera etc) KB	GNSS RF KB	spectrum analysis KB	Track data KB	Ground Truth KB	Total DB storage GB
Pontremolese Ale642 Le682	1269214	77104848	1803068	765000000	7602060	7080730,85	3000	859,86
Cagliari ALN668	15	6420,168	150	0	0	1250083,85	0,25	1,26
Swiss RDBe 560	39943008	2595197616	60718303	540000000	5636010	4500003750	101025	7741,60
BRNO 362WTB	295463	92493024	2163681	1125000000	10747740	5202031,35	3600	1235,91
Volary 814-914	2350424	781118136	18586622	1890000000	188714586	1568009,75	30925	2882,37
								12720,99

The cell “Byte overhead per individual record/sample” is an input field which determines the overhead per individual record created and stored by the measurement system. It impacts several other calculations and is placed here even though considered from a data hierarchy perspective it shown be placed elsewhere.

The cells in the Sensor data column show per site/train the sum of data outputs from all types of on-board equipment excluding the Spectran spectrum analyser, the Spirent GNSS RF recorder and units that produce a continuous stream of data which is file based such as some cameras, GNSS raw data etc.

The Sensor data units (which have a specific configuration per Site/train) can be split into devices which provide data continuously or are event driven/triggered but they are both summed here.

The data overhead for sample/individual records column is the sum of the overhead that the sample data will incur based on the byte overhead (yellow field) and the a) sample frequency and per packet of sensors providing continuous data over the elapsed time of performance measurement and b) the distance travelled and estimated frequency of occurrence of irregular events (balises, tags etc) based on kilometres travelled. Lots of small data packets at high sample freq (tacho) will incur a higher overhead that large packets that occur less frequently (balise).

The data in file format column is a sum of the data produced by equipment in files i.e. GNSS raw data, cameras etc. this data has been separated as it will be stored as file reference data’s as described in [48], data with links and as such the data storage capacity is calculated differently.

GNSS RF data is assumed to be stored in data file format but is shown separately as it has a high impact on storage and as such it is envisaged that there will be specific train runs for RF recording purposes and as its own column it is valuable to be able to assess the RF contribution to storage requirements independently.

Spectrum analysis is for the Spectrum analyser and it is also shown as in a separate column for the same reasons as for FR data but also because it is calculated form a percentage of the run time.

The Track data is the estimated size of the initial digital track database.

2.7.2 Track Data

Track data from track databases						
All table data in bytes except totals						
No of reference objects and km taken from D2.3						
Site	km	ref objects	data overhead for representation of an object	data overhead per km representation in cm	Data overhead for track representation between objects	Total GB
Pontremolese	119	3078	50	59500000	50	7,08073085
Cagliari	50	1118	50	25000000	50	1,25008385
Swiss	3000	50000	50	1500000000	50	4500,00375
BRNO	102	418	50	51000000	50	5,20203135
Volary	56	130	50	28000000	50	1,56800975

Figure 33: Track data

The track data sheet, Figure 33, is used to calculate the storage requirements for the initial digital track database. It is based on the track databases for each of the sites along with the number of references objects and the number of kilometres covered in by the input track databases. The calculations are based on cm level representation of the track, the reference objects and data overheads for reference object relationships i.e. one track splitting into two after a point. The data representation in this table is in bytes. The totals from this page are carried over to the main calculator page.

Km and number of reference objects are taken from the document D2.3 identification of the representative Railway/Lines Sites [49].

Data overhead for representation of an object is an estimated value of the overhead per object.

Data overhead per km rep in cm is the data overhead for representation of the track into cm. =B5*5*100000, where B is the km column and 5 is the byte overhead per cm, 100000 is for km to cm conversion.

Data overhead for track representation between objects is the overhead for linking information between objects in order to form a relational representation.

2.7.3 Site/train data

Site/train data sheets are similar for the five sites each with its own specific train. These are shown in Figure 34.

Pontremolese Ale642 Le682	Cagliari ALN668	Swiss RDBe 560	BRNO 362WTB	Volary 814-914

Figure 34: Site/train

For each of these sites/Trains there is a log where duration and distance covered by GNSS performance measurements are estimated. It is possible to enter individual runs or subtotals of runs of a specific character for all estimated sample train runs with Spirent GNSS RF recording could be summed together by marking the “GNSS RF” column.

	total Km	total Time			total KB continous data	total KB for fixed references	total KB overhead for individual records	total KB RF data	total KB Spectrum analysis
	386	12000			1070424	198790	77104848	765000000	7602060
Run ID	Km	time/s	GNSS RF	Spectrum Analysis % of trip	cont. Data subtotal KB	discrete/irregular event KB	Sample/record overhead @6 bytes	Spirent RPS	Spectran Kb
Parma - la spezia 1	112	3600	1	5	321127,2	57680	23130816	540000000	4718520
La specia - palma 2	112	3500		1	312207	57680	22488816	0	917490

Figure 35: Site/Train log

The cells in the “GNSS RF” column shall be set to indicate an estimated duration and distance for RF measurement.

The cells in the “spectrum Analysis % of trip” should be used as an estimated % spectral analysis recording during the estimated trip or run. As the exact usage criterion of the spectrum analysis is unclear it is considered that the Spectrum analyser is triggered when the environment RF interference has reached a threshold trigger. There are a lot of estimates here but the idea is to provide flexibility so as to optimise the schedule of measurements in order to be able to meet a database storage estimate or vice-versa.

The cell “continuous data subtotal” is the sum data output from units in a continuous stream, that will be sampled and timestamped in to records at a given frequency i.e. wheel tachometers. The cell formula is the sum of the duration of performance measurement * sum of continuous data/second subtotal from the “Train configurator” which is a group of rows hidden at the top of the excel sheet. This “Train Configurator” section will be addressed in the next section.

The column discrete/irregular events is the data (Bytes) sum of the events per km estimate for irregular events i.e. Balises, RFID etc. the data sum is also taken from the Train configurator.

The column “Sample/record overhead” is a data “bytes” overhead for every individual record created for both fixed and continuous data. The number of samples is taken from the train configurator based on either estimated irregular events per km AND from the sensor inputs divided by the sample record frequency on-board the train multiplied by the performance measurement duration OR time respectively.

The Spirent and Spectran columns are for the data output of these units multiplying the output bandwidth by the measurement duration.

2.7.4 Train configurator

Each Site/Train sheet is adapted for the particular equipment on the train by the Train configurator, see Figure 36. The units available are fixed and are the same for all sites/train but by individually enabling/disabling a unique configuration can be attained. Column D shows the output bandwidth of the sensor unit and is used to calculate the total bandwidth/s. Column C shows the on-board sampling rate for that particular unit that as given in [49]. These values can be changed but it is important to reflect any change in the other sites/trains otherwise it will be difficult to maintain discipline as far as calculator organisation is concerned.

Column F gives the data output per event for discrete/irregular events and an estimate of frequency of occurrence is given by [49].

Units on-board the train that output data in a file format, which in the final database will be identified by file references as these files inherently contain accurate timing information are shown at the top of the configurator.

The totals in cells 24 for columns D and G (G being the subtotal for F * estimated frequency per km) are then used in the site/train log calculations.

Units which have a file output and as such are not subject to being sampled and timestamped on-board (with the exception of the cabin camera) are list in the top rows 1-7. The configuration for these units is given on separate sheets for those units where the setup is more complex.

	B	C	D	E	F	G	H
1	enable	samples/s	kBytes/sec	events/km	Kbyte/event	total/km	Info
2	1		43,11				File
3	1		20,18				File
4	1		24,96				File
5	1	15	27				Fisheye camera, fi
6	1			5	100	500	file
7	1	25	35				File Camera model
8	0			2	0,004		1 per 0,5km, more
9	1	100	0,002				Start - stop + nur
10		100	3				Hasler CoRRail
11	1	100	9				32mB/hr
12	0	100	8				
13	0	5	0,01				czech train estima
14	1	0.01	0,5				Guess
15	1	100	1				
16	1			3	5	15	
17	0			2	0,019	0,038	
18	0			2	0,004		
19	1	10	78				Speed, acc/dec, re
20	1	10	0,2				Dynamic +static
21	1	10	0,50				
22							
23							
24		535	89,20	14		515	

Figure 36: Train configurator

2.7.5 GNSS Receivers, RFI spectrum analysis units

Certain devices are given their own sheet which is then used in the on-board configurator in the site/data sheet. This is for units that have a relatively high level of configurability or where configuration flexibility may be important. This is shown in Figure 37.

Figure 37: File based output devices

The AsteRx4 device is the most complex as it is highly configurable as well as having a data output that is variable based on changes in data. The yellow input fields, see Figure 38, are used to enter an estimate of the bands that will be tracked along with number satellites per constellation. These, of course, can be changed and the effects of the changes will be reflected through the calculator and shown on the main page subject to the duration of measurement given on the Site/train pages. Note that the configuration of the unit will be the same for all sites.

For "Blocks output when new data is available..", the "Update Time" fields have been conservatively estimated based on general information regarding the different SIS/Navigation messages.

Receiver channels		GPS	GLO	Galileo	SBAS
# of satellites		7	7	7	2
Tracked L-band	L1/E1/G1	1	1	1	1
Tracked L-band	L5/E5				
Tracked L-band	L2/G2				
# of tracked freq		1	1	1	1
Total # of channels	23	7	7	7	2

SBF Blocks

Blocks output at a defined rate:

Measurement blocks:		Enable	Header + Time Stamp [Bytes]	Data [Bytes]	Padding [Bytes]	Total [Bytes]	OnChange [ms]	Max update Frequency	Used update Frequency	Average Data kBytes/s	Comment
Measure	MeasEpoch	1	14	66	0	80	10	100	10	0,80	Required for creation
Measure	MeasExtra	1	14	321	1	336	10	100	10	3,36	
Measure	IQCorr	1	14	174	0	188	10	100	10	1,88	
Measure	ISMR	1	14	174	0	188	1000	1	1	0,19	
Measure	EnodOfMeas	1	14	0	2	16	10	100	10	0,16	
Measurement blocks, Total:						808				6,39	

Figure 38: Configuration sheet for Septentrio AsteRx4 receiver

2.8 IDENTIFICATION OF DATA STORAGE SERVER

This section presents possible solution for hosting database described in section 2.6. With respect of expected data volume, company security, time and project budget the only cloud solution was evaluated to be acceptable. Following sections provide basic information on cloud solution and description of services of one selected provider.

2.8.1 Cloud Computing Based Solution

This is an internet-based computing solution which offers on-demand resources as data storage, computer power, etc. with a minimal administration effort and accessible through the Internet.

It offers a clear set of advantages for storing big amounts of data with the peculiarities of the STARS project:

- Easy and centralized administration
- Security system implemented to allow remote access, valid for the different partners
- It avoids the necessity of buying, installing and maintaining specific data servers because all the data is stored in the cloud infrastructure
- Cost reductions for storing the data. Also the costs could be calculated according of the data load of the servers
- It is very flexible and scalable, it allows increasing the data storage capacity in the same installation
- The maintenance of the system is easier
- It is an actual standard
- Productivity is also increased because allows multiple users accessing the same data simultaneously
- It provides already pre-installed packages for OS and for databases

2.8.2 Cloud service providers

There are different solutions for cloud computing services offered by the main software providers as Microsoft (https://azure.microsoft.com), Google (https://cloud.google.com/), Amazon (https://aws.amazon.com/) and others. An evaluation of the main cloud providers (Amazon and Google) has been performed in order to select the best provider. The selection process has been based in the requirements of the project, the price of the solution and the offered support.

AMAZON WEB SERVICES (AWS)

Amazon Web Services (AWS) is a secure cloud service platform working on-demand. It offers a flexible compute power and database storage allowing building modular services stored in the cloud.

For file storage, database and processing services within STARS the AWS cloud will be used.

- File storage
  - Simple Storage Solution S3
- Database
  - as SQL database: MariaDB
  - as NoSQL database: MongoDB / OrientDB / Neo4j
- Processing service
  - Elastic Compute Cloud (EC2) to manage Windows/Linux server instances

The connection between local servers and cloud services will be realized via ssh tunnel, secured by ssh key pairs provided by each member.

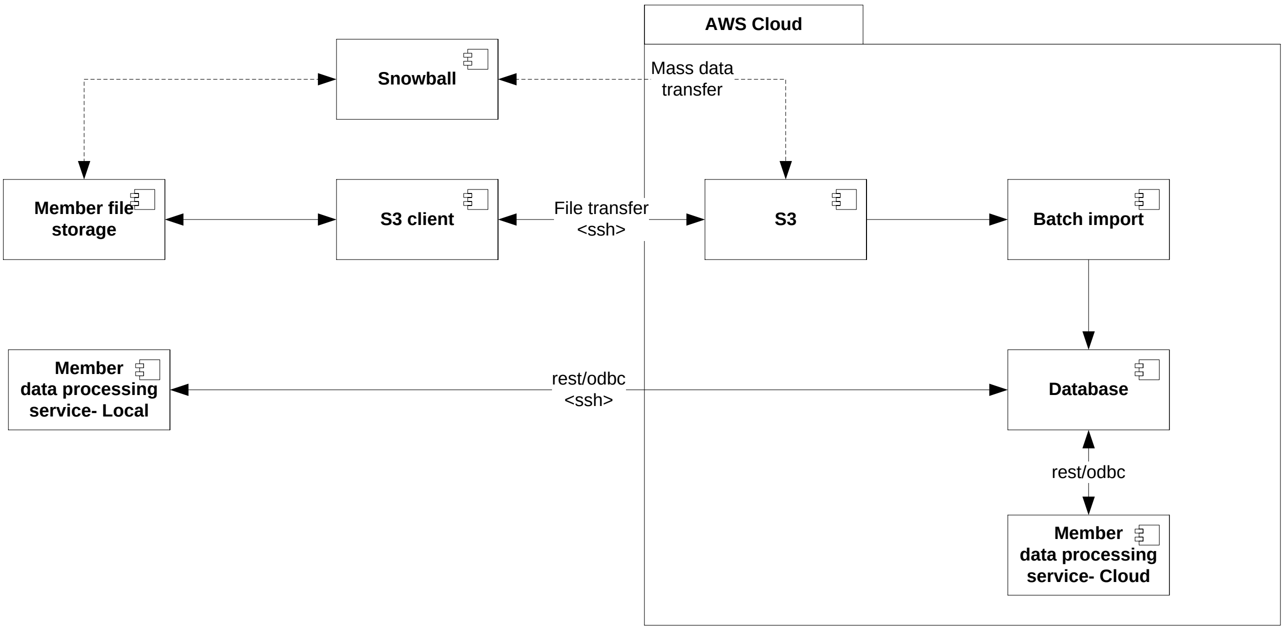


Figure 39: Overview Stars Data Server

The storage solution for STARS could be implemented by using the AWS Simple Storage Service (S3). The Amazon S3 allows the users to store and retrieve any amount of data at any time, from anywhere on the web.

The data will be stored in buckets. Structure and size of buckets can be defined by each member.

The file transfer from and to S3 is possible by using a Cloudberry client [50]. For huge amounts of data the Snowball transfer described in the next section can be used.

The access to each bucket can be configured via access policies. It is possible to provide access for all members or only for selected members.

The AWS Import/Export via AWS Snowball is a service that accelerates transferring large amounts of data into and out of AWS using physical storage appliances, bypassing the Internet.

Shipping in Snowballs allows the transfer of large amounts of data between a local data centers and Amazon Simple Storage Service (Amazon S3) at a significantly faster rate than any transfer via internet. Snowball is available as an 80TB model.

Snowball uses Snowball appliances and provides simple interfaces that can be used to create jobs, transfer data, and track the job status.

Google Cloud Platform

Google offers a cloud computing service using the Google infrastructure providing a set of modular cloud-based services with a host of development tools. For the purpose of the STARS project, the Google Cloud Storage has been analyzed.

The Google Cloud Storage is a unified object storage which allows data serving, data analytics and data archiving. It provides multi-regional access with redundant capabilities as well as regional access with higher availability. It offers two different types of storage depending of the type of data to be stored.

- Nearline is a fast, low-cost, and highly durable storage for data accessed less than once a month
- Coldline is a fast, low-cost durable storage for data accessed less than once a year

Based of the estimated data generated by the measurement campaign, a detailed solution has been defined. The advantage of using a cloud based service is that this can be extended without any influence in the system architecture. This is the monthly expected traffic:

Network Bandwidth	Egress - Americas and EMEA	15369 GB
Cloud Storage	Nearline storage	10240 GB
	Coldline storage	25600 GB
	Class A operations (millions)	100 1000

Table 32: Google Cloud Storage - monthly expected traffic

A test account has been setup by Siemens together with Google in order to test the cloud solution. Initial data has been uploaded and downloaded using a web interface (see Figure 40).

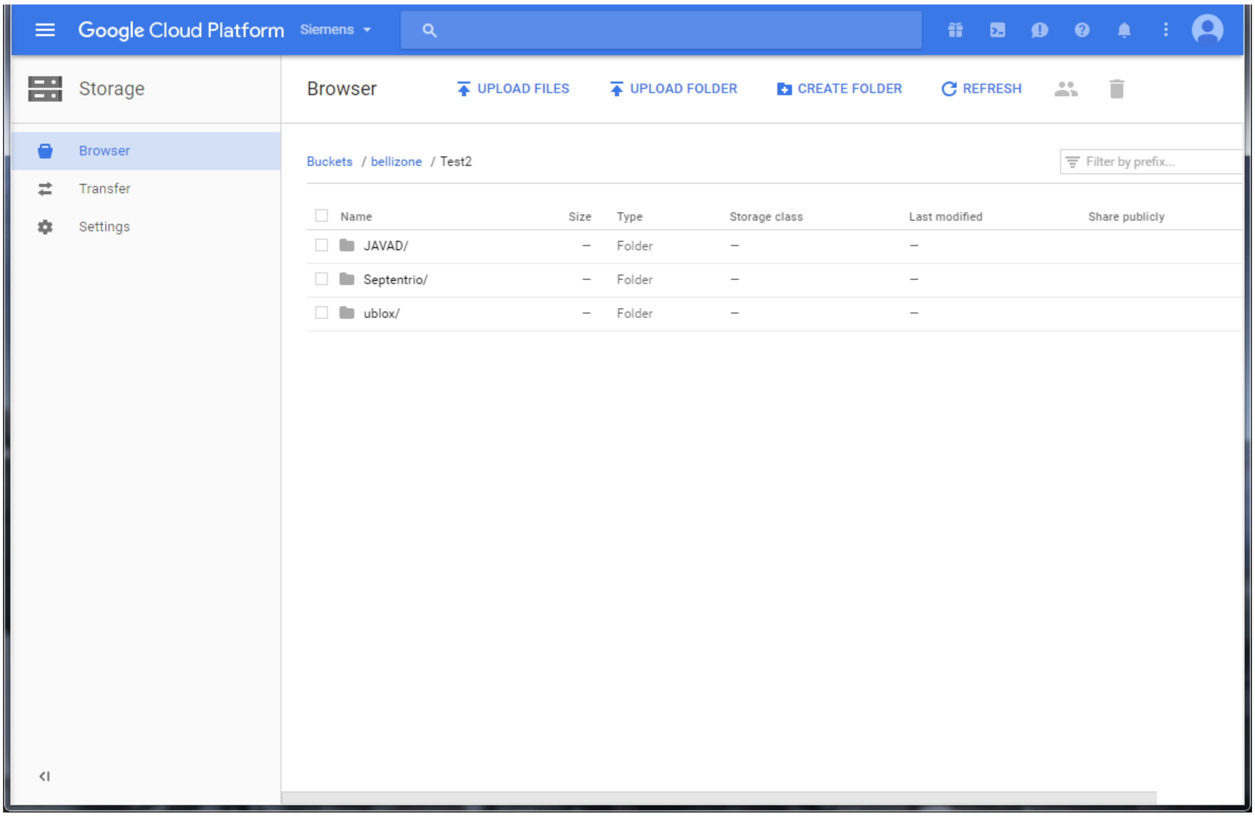


Figure 40: Google Cloud Storage web interface

It is also possible to use the Cloud SDK in order to automatize some of the actions of data synchronization. Further information can be found (https://cloud.google.com/sdk/docs/).

3 REFERENCES

[1] STARS “WP4 Project Planning, Definition of Tasks”, STR-WP4-G-AZD-016-01, October 2016.

[2] STARS “Development of methods for GT data processing”, STR-WP4-T-SIR-01, August 2016.

[3] K. Brammer, G. Siffling, “Deterministische Beobachtung und stochastische Filterung”, Munich 1985.

[4] F. Böhringer, “Gleisselektive Ortung von Schienenfahrzeugen mit bordautonomer Sensorik”, Karlsruhe 2008.

[5] A. Himmel, V. Wiedmeyer, “Interdisziplinäres Teamprojekt: oTToCAR”, Magdeburg 2013. Available: http://ottocar.cs.ovgu.de/wp-content/uploads/2013/02/OTTocarBericht_A_Himmel_V_Wiedmeyer.pdf.

[6] P. Spiess, DB Systemtechnik GmbH, “Fahrdynamik des Schienenverkehrs”, Minden 2005. Available: https://www.ids.uni-hannover.de/fileadmin/IDS/ids_lehre/SFZ/Fahrdynamik_2006.pdf.

[7] DB Systemtechnik GmbH, “Hochgeschwindigkeitsverkehr in Deutschland”, Minden 2016. Available: https://www.db-systemtechnik.de/file/dbst-de/11420656/MdTgXvSdPRRneYFejXAKi5JfQ5A/12136062/data/flyer_schulprojekt_2016_eichstaett.pdf.

[8] J. Wendel, “Integrierte Navigationssysteme: Sensordatenfusion, GPS und inertiale Navigation”, Karlsruhe 2006.

[9] Q. Gan, C. J. Harris, “Comparison of two Measurement Fusion Methods for Kalman-Filter-Based Multisensor Data Fusion”, *IEEE Transactions on Aerospace and Electronic Systems*, vol. 37, no. 1, pp. 273-279, January 2001.

[10] STARS D2.3 Identification of the Representative Railway Lines/Sites, STR-WP2-D-ANS-048

[11] STARS Methods for Ground Truth data processing (AŽD), technical note, 2016.

[12] Punčochář, Straka, Král, Škach, Ajgl: Design of the selected way of track axis map incorporation into ZZ RB1. Internal document of the project BETA, 2015.

[13] A. Censi, L. Marchionni and G. Oriolo, "Simultaneous maximum-likelihood calibration of odometry and sensor parameters," *2008 IEEE International Conference on Robotics and Automation,* pp. 2098-2103, 19-23 May 2008.

[14] M. Malvezzi, B. Allotta and M. Rinchi, "Odometric estimation for automatic train protection and control systems," *Vehicle System Dynamics,* vol. 49, no. 5, pp. 723-739, May 2011.

[15] M. Malvezzi, P. Toni, B. Allotta and V. Colla, "Train Speed and Position Evaluation using Wheel Velocity Measurements," *2001 IEEE/ASME International Conference on,* vol. 1, pp. 220-224, 8-12 July 2001.

[16] STARS WP2 RecordConfCodes-02 (Excel sheet), 9th of November 2015.

[17] EDAS Service Definition Document, version 2.1, 19th of December 2014.

[18] IGS Formats, Accessed on Novemeber 10, 2016. Available: ftp://ftp.igs.org/pub/data/format/

[19] STARS Data from Servers, STR-WP3-T-CET-016-001, Novemeber 2016.

[20] GSS6450 USER MANUAL, HARDWARE AND SOFTWARE USER GUIDE FOR SPIRENT'S GNSS SIGNAL RECORD AND PLAYBACK SYSTEMS DGP01444AAA, Issue: 1.31, Date of Issue: 09 November 2016.

[21] Paul D. Groves, Principles of GNSS, Inertial, and Multisensor Integrated Navigation Systems, Artech House.

[22] Discussions on RF Signal Propagation and Multipath, http://www.novatel.com/assets/Documents/Bulletins/apn008.pdf, Accessed on February 3rd, 2003.

[23] Sahmoudi, M., Landry, R. Jr., Multipath Mitigation Techniques Using Maximum-Likelihood Principle, http://www.insidegnss.com/auto/novdec08-sahmoudi-v1.pdf, InsideGNSS 11/12 2008.

[24] Tonwsend, B. R., Richard van Nee, D. J., Fenton, P. C., Van Dierendonck, J. V., Novatel Communications Ltd., Faculty of Electrical Engineering TU Delf, Performance Evaluation of the Multipath Estimating Delay Lock Loop, http://www.novatel.com/assets/Documents/Papers/File25.pdf, ION National Technical Meeting, Anaheim, California, January 18-20, 1995.

[25] Richard van Nee, D. J., Siereveld, J., Fenton, P. C., Tonwsend, B. R., The Multipath Estimating Delay Lock Loop: Approaching Theoretical Accuracy Limits, http://www.novatel.com/assets/Documents/Papers/File21.pdf, IEEE Position, Location and Navigation Symposium Las Vegas, Nevada, April U-15, 1994.

[26] Rouabah, K., Chikouche, D., Utilisation de la Transformée de Fourier FFT pour la Détection des Multitrajets dans les Récepteurs de Navigation C/A GPS, http://iceeac-dz.org/docs/articles/Theme_C/Articles_C/C15.pdf, 4th International Conference on Computer Integrated Manufacturing CIP'2007, 03-04 November 2007.

[27] Marais, J., Berbineau, M., & Heddebaut, M. (2005). Land mobile GNSS availability and multipath evaluation tool. *IEEE transactions on vehicular technology*, *54*(5), 1697-1704.

[28] Marais, J., Meurie, C., Quantify and improve GNSS quality of service in land transportation by using image processing, 1st CNES-ONERA Workshop on Earth-Space Propagation, ONERA, Toulouse, France, 21-23 January 2013.

[29] Marais, J., Meurie, C., Flancquart, A., Lithgow, S., & Barbu, G. (2014, July). Innovative simulations of GNSS performances in a realistic railway environment. In CERGAL, International Symposium on Certification of GNSS Systems & Services (p. 5p).

[30] Gakne, P. V., & Petovello, M. (2015, October). Assessing image segmentation algorithms for sky identification in GNSS. In Indoor Positioning and Indoor Navigation (IPIN), 2015 International Conference on (pp. 1-7).

[31] Suzuki, T., & Kubo, N. (2014). N-LOS GNSS signal detection using fish-eye camera for vehicle navigation in urban environments. In Institute of Navigation, the 27th International Technical Meeting of The Satellite Division of the Institute of Navigation (ION GNSS+ 2014).

[32] Sanromà Sánchez, J., Gerhmann, A., Thevenon, P., Brocard, P., Ben Afia, A., Julien, O., (2016). Use of a FishEye Camera for GNSS NLOS Exclusion and Characterization in Urban Environments. ION International Technical Meeting, January 25-28, 2016, Monterey, California.

[33] Roberts, S., (2016). Railway environment modelling. RHINOS project workshop, Stanford, November 29th-December 1st 2016.

[34] STARS D2.2 Specification of the Field Measurement Procedures, STR-WP2-D-ANS-047-06, 15 December 2016.

[35] Cristian F., “Probabilistic clock synchronization,” Distributed Computing, vol. 3, no. 3, pp. 146-158, 1989.

[36] Gusella R. and Zatti S., “The Accuracy of the Clock Synchronization Achieved by TEMPO in Berkley UNIX 4.3BSD,” IEEE Transactions on Software Engineering, vol. 15, no. 7, pp. 847-853, 1989.

[37] Marzullo K. A., “Maintaining the time in a distributed system: an example of a loosely-coupled distributed service (synchronization, fault-tolerance, debugging),” Ph.D. dissertation, Stanford, CA, USA, 1984.

[38] STARS Technical note “Common Minimal Set of Equipment” STR-WP2-T-ANS-049-01, 10 October 2016.

[39] Very Long Baseline Interferometry, http://space-geodesy.nasa.gov/multimedia/VLBIHistoryVideo.html , Accessed on October 3rd, 2016.

[40] McCarthy D. D.; Astronomical Time, Proceedings of the IEEE, Volume 79, Number 7, July 1991.

[41] Leap seconds, http://tycho.usno.navy.mil/leapsec.html , Accessed on October 3rd, 2016.

[42] Subirana J. S., Zornoza J. M. J., Hernández-Pajares M.; GNSS Data Processing, Vol. 1: Fundamentals and Algorithms (ESA TM-23/1), ESA Communications, Noordwijk, the Netherlands, May, 2013.

[43] Allan, D. W., Ashby N., Hodge C. C.; Hewlett Packard Application Note 1289, 1997.

[44] Global Positioning System Standard Positioning Service Performance Standard, 4th Edition, September 2008.

[45] GPS Timing Data and Information, http://www.usno.navy.mil/USNO/time/gps/gps-timing-data-and-information , Accessed on October 4th, 2016.

[46] RECOMMENDATION ITU-R TF.460-6*, International Telecommunication Union, February 2002.

[47] Common minimal set of equipment, STR-WP2-T-ANS-049-01, October 2016.

[48] Database Concept Proposal, STR-WP4-T-SIR-012-01, September 2016.

[49] STARS D2.3 Identification of the Representative Railway Lines / Sites, STR-WP2-D-ANS-048, 31st October, 2016.

[50] CloudBerry Explorer Client, http://www.cloudberrylab.com/free-amazon-s3-explorer-cloudfront-IAM.aspx.