GEO1003 - Shared Notes

Master Geomatics Students

January 24, 2025

Contents

1 Introduction

These notes were created by students from the MSc Geomatics for the Built Environment at TU Delft.

They are based on the lectures and literature provided by the course GEO1003 (Positioning and Location Awareness) by Edward Verbree.

2 How does GNSS work?

2.1 Introduction

The Global Positioning System (**GPS**), also known as the NAVigation Satellite Time And Ranging (**NAVSTAR**) system had its first satellite launched back in February 1978. GPS is a *one-way* radio ranging system which provides realtime knowledge of one's Position and Velocity, and a very accurate Time reference as well (all together referred to as **PVT**).

Below are the 4 constellations of GNSS systems:

- **Global Positioning System (GPS)** - United States
  - *Standard Positioning Service (SPS)*
  - *Precise Positioning Service (PPS)*
- **GLONASS** - Russia
- **Galileo** - Europe
  - *Open service (OS)*
  - *Public regulated service (PRS)*
- **Beidou(/Compass)** - China

Their properties are:

Table 8.14 Properties of GNSS Mid-Earth Orbits

Constellation	*Number of Planes*	*Radius (km)*	*Height (km)*	*Period*	*Orbits per Sidereal Day*	*Ground-Track Repeat Period (Sidereal Days)*	*Inclination Angle*
GPS	6	26,580	20,180	11 hr, 58 min	2	1	55°
GLONASS	3	25,500	19,100	11 hr, 15 min	2.125	8	64.8°
Galileo	3	29,620	23,220	14 hr, 5 min	1.7	10	56°
Beidou	3	27,840	21,440	12 hr, 52 min	1.857	7	55°

Figure 1: Properties of GNSS systems

2.2 Introduction

Notes from the chapter Introduction of the book "Principles of GNSS, Inertial, and Multisensor Integrated Navigation Systems" - Paul D. Groves

2.2.1 1.1 Fundamental Concepts

Navigation involves two main concepts: determining the position and velocity of a moving body relative to a reference point, and planning a course to move from one place to another while avoiding obstacles. **Positioning** is the determination of the position of a body and is thus a subset of navigation.

Positioning techniques can be categorized into:

1. **Real-time vs. Postprocessed**: Real-time positioning calculates positions immediately, while postprocessed techniques do so later. Real-time positioning is further divided into continuous (for navigation) and instantaneous (for single-point applications).

2. **Fixed vs. Movable Objects**: Positioning for fixed objects is static, while for movable objects (like in navigation), it is dynamic. Navigation requires velocity, which may be derived from updated positions or other methods.

3. **Self-positioning vs. Remote Positioning**: Most navigation applications use self-positioning, where the object calculates its position. Remote positioning involves calculating the position elsewhere without the object's cooperation.

A **navigation system** is a device that determines position and velocity automatically. Similarly, a positioning system determines position. An integrated navigation system determines position and velocity using more than one technology.

Navigation sensors like accelerometers and gyroscopes measure properties to compute outputs, and the **navigation solution** provides position, velocity, and sometimes additional data like heading or acceleration. Depending on the application, this can be 2D (for cars, trains, etc.) or 3D (for air, space, underwater, or indoor navigation).

The **user** in navigation refers to the person or software that receives the position and velocity solution, and **user equipment** refers to the system located on the object being positioned.

Position is expressed quantitatively (e.g., coordinates), while **location** is expressed qualitatively (e.g., city, street). A map or GIS system can convert between the two.

Some authors use the term **localization** instead of positioning, particularly for short-range applications.

Positioning and navigation techniques rely on two fundamental methods:

- **Position Fixing**: Uses external information like signals (radio, acoustic, ultrasound, optical, or infrared) or environmental features (e.g., landmarks, terrain, signs, roads, rivers, terrain height, sounds, smells, and even variations in the magnetic and gravitational fields) to determine position. A navigation aid (AtoN) is a landmark used specifically for navigation.

- **Dead Reckoning**: Measures distance and direction traveled using a self-contained system, like an INS, without relying on external infrastructure. Environmental features can also assist dead reckoning by comparing repeated measurements.

2.2.2 1.2 Dead reckoning

Dead reckoning, possibly derived from "deduced reckoning," involves calculating the current position by measuring changes in position or velocity and integrating this data. This information is added to a previous position to determine the current one. Measurements are taken in body-aligned axes, requiring an attitude solution to establish the travel direction relative to the environment.

- 2D Navigation: A heading measurement suffices.
- 3D Navigation: Requires a full three-component attitude measurement.
- Accuracy: Smaller step sizes in calculations improve accuracy, especially when attitudes change.

Modern methods include automated pace counting with pedometers and advanced pedestrian dead reckoning (PDR) using accelerometers to estimate step length.

An odometer measures distance via wheel rotations and is standard in vehicles, dating back to Roman times. Marine equivalents include electromagnetic speed logs or sonar, while aircraft use Doppler radar. Environmental feature tracking (e.g., camera or radar images) is also employed.

Heading Measurement:

- Magnetic compasses (ancient but now with electronic readouts) are common.
- Marine heading: Determined with gyrocompasses.
- Land heading: Derived from vehicle trajectory.
- Roll and pitch: Measured using accelerometers, tilt sensors, or celestial observations (sun, moon, stars).
- Gyroscopes and differential odometry help measure changes in attitude and heading.

Inertial Navigation Systems (INS):

- A complete 3D dead-reckoning system includes inertial sensors (IMU: accelerometers and gyroscopes) and a navigation processor.
- Accelerometers measure specific force, and gyroscopes maintain the attitude solution. The processor integrates these to compute position, velocity, and attitude.

Performance:

- Navigation accuracy depends on sensor quality
- The principal advantages of inertial navigation and other dead-reckoning techniques, compared to position fixing, are continuous operation, a high update rate, low short-term noise, and the provision of attitude, angular rate, and acceleration as well as position and velocity.
- The main drawbacks are that the position solution must be initialized and the position error grows with time because the errors in successive distance and direction measurements accumulate.
- In an integrated navigation system, position-fixing measurements may be used to correct the dead-reckoning navigation solution and also calibrate the dead-reckoning sensor errors.

2.2.3 1.3 Position Fixing

2.2.3.1 1.3.1 Position-Fixing Methods There are five main methods: proximity, ranging, angular positioning, pattern matching, and Doppler positioning.

- **Proximity**:
  Simplest method, assumes the receiver's position is the transmitter's or a nearby feature's. Accuracy improves with proximity to landmarks. Short-range signals like Bluetooth, RFID, or indoor features are ideal. Using multiple landmarks allows averaging of their positions.
  Containment Intersection: A refined proximity method where containment zones around landmarks are defined. Observed landmarks localize the position to the intersection of these zones, with the center of the intersection as the position fix.

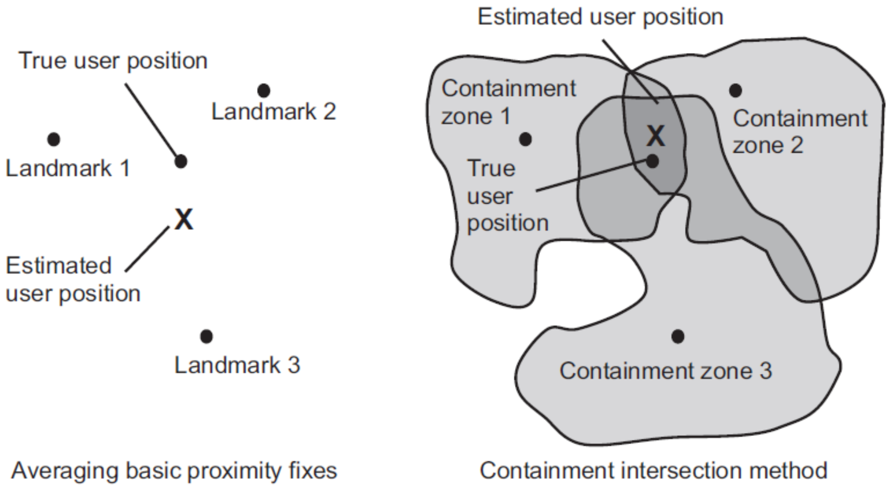

Figure 1.7 Basic and advanced proximity positioning using multiple landmarks.

- **Ranging**:
  Measures distances to landmarks and defines circular lines of position (LOPs) based on these ranges. The user's position lies at the intersection of LOPs. At least two LOPs are needed; three resolve ambiguities. (Or prior information)
  In three-dimensional positioning, each range measurement defines a spherical surface of position (SOP) centered at a landmark. The intersection of:
  - Two SOPs forms a circular line of position (LOP).
  - Three SOPs intersects at two points, requiring additional information or a fourth SOP to determine a unique position fix.

  If the user and landmarks are coplanar, only planar position components can be determined, limiting vertical accuracy in terrestrial systems.
  Range Measurement: based on signal time of flight (TOF), multiplied by the speed of light or sound. Accurate TOF measurement depends on transmitter-receiver time synchronization.
  - Two-way ranging (e.g., distance measuring equipment - DME) minimizes synchronization errors via bidirectional signal exchange.
  - One-way ranging (e.g., GNSS) requires synchronized transmitter clocks, treating receiver clock offset as an unknown. This technique is known as passive ranging and is how GNSS works. Additional transmitters or reference receivers can correct for clock offsets.
  - Where the landmark is an environmental feature, active sensors like radar, sonar, or laser transmit signals to a landmark, measure the reflected signal's round-trip time, and calculate the range.
- **Bearing**:
  A bearing is the angle in the horizontal plane between the line of sight to an object and a reference direction (e.g., true or magnetic north).
  Position Fixing with Bearings:
  - Two-Dimensional Positioning: Measuring the bearing to two known landmarks defines straight LOPs. Their intersection gives the user's position. Without a reference direction, measuring the angle difference between two landmarks defines a curved LOP, requiring three landmarks for positioning.
  - Three-Dimensional Positioning: Extends by measuring the elevation angle (line of sight vs. horizontal plane) to one landmark. However, accuracy decreases with distance, and Earth's curvature affects measurements.
- **Angle of Arrival (AOA)**:
  Direction Finding: Uses steerable directional antennas to determine signal bearing.

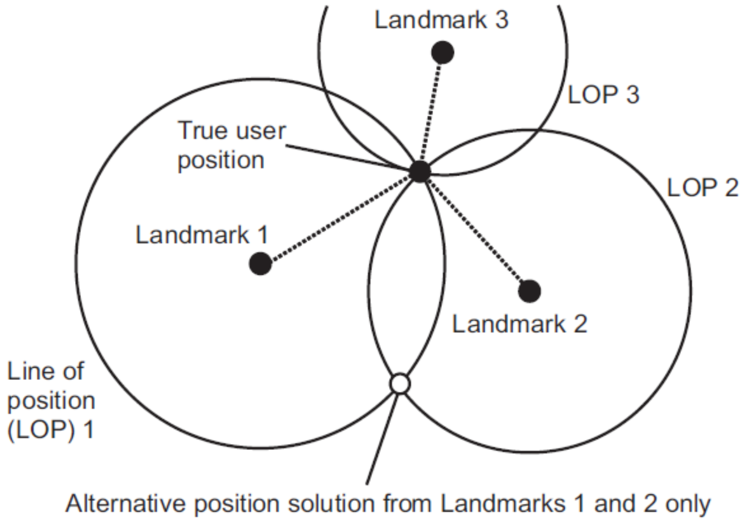

Figure 1.8 Positioning by ranging in two dimensions.

Figure 2: Positioning by ranging

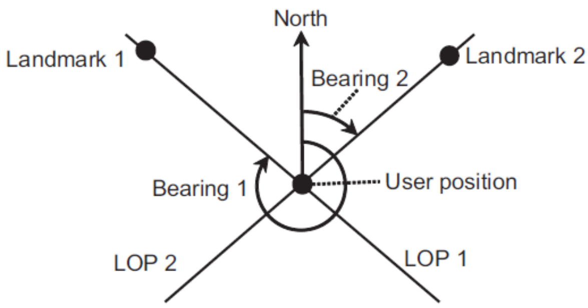

Figure 1.9 Angular positioning in two dimensions.

Figure 3: Angular positioning

- Nonisotropic Transmissions: Broadcast signals with modulation varying by direction enable bearing and/or elevation determination. Examples include VHF omnidirectional radiorange (VOR) and Nokia HAIP (high-accuracy indoor positioning).
- Environmental features can be measured using cameras, laser scanners, imaging radar, or multibeam sonar. The position of a feature in the sensor's image, combined with sensor orientation, determines its bearing and elevation.

- **Integrated Navigation System**:
  Full position fixes are unnecessary; single measurements (e.g., range, bearing, or elevation of a landmark) can contribute to the navigation solution. For example, a two-dimensional position fix may be obtained by measuring the range and bearing of a single landmark as shown in Figure 1.10. Adding elevation provides a three-dimensional fix.
- Landmark Identification:
  Signals are identified via demodulation, transmitter IDs, or frequency. Environmental features are matched with stored data, requiring distinctive features and approximate position inputs to limit the size of database that must be searched to obtain a match. Even so, positioning using environmental features is normally more processor intensive than signal-based positioning.
- **Pattern Matching**:
  A database contains parameters varying by location (e.g., terrain height, signal strength, magnetic fields, GNSS obstructions). Measured values at the user's position are compared with database values at candidate grid points. The best match determines the position. Interpolation refines results if several candidates match well. As with feature matching for landmark identification, the input of an approximate position solution limits the size of the database to be searched.
  Combining multiple parameters into a **location signature** enhances uniqueness and accuracy. In some cases, such as terrain height, there is insufficient information to obtain an unambiguous position fix. However, if the navigation system is moving, measurements may be made at multiple positions, collectively known as a transect. Position fixing can be enhanced by this. Using dead reckoning to relate transect points allows measurements to form a location signature, compared with a database for positioning, requiring interpolation if spacing differs.
- **Doppler Positioning**:
  Measures Doppler shift from relative transmitter-receiver motion to determine relative velocity and a conical surface of position (e.g., used in Iridium positioning).
- **Height Measurement**:
  Barometric Altimeter: Uses pressure to estimate height or underwater depth. Radar Altimeter: Measures height above terrain for aircraft where terrain height is known.
- **Data and Databases**:
  Position-fixing methods rely on data like landmark positions, feature info, and pattern-matching parameters. Databases may be preloaded but require updates and significant storage for large areas.
- **SLAM (simultaneous localization and mapping)**:
  Allows systems to create and update its own landmark databases by exploring environments, observing features several times and and using dead reckoning to track traveled distances.

Many signal-based positioning systems include transmitter positions in their signals, but this can delay position computation after signal reception. A separate data link, known as assistance, can provide the necessary information on demand to reduce delays.

Position fixing is essential for determining absolute position, with errors independent of the distance traveled, but it depends on the availability of suitable signals or environmental features. To enhance position accuracy and availability, multiple position-fixing technologies or dead reckoning can be combined to bridge gaps.

2.2.3.2 1.3.2 Signal-Based Positioning Early Radio Navigation: Radio-based navigation started in the 1920s, with systems like Omega providing global coverage in the 1970s.

Terrestrial Systems: Older systems like DME, VOR, and Loran offer long-range navigation (hundreds to 3,000 km), while modern techniques use existing signals from Mobile phone signals, WLANs or Wi-Fi, wireless personal area networks (WPANs), such as Bluetooth and Zigbee, RFID, ultrawideband (UWB) communications, television signals, and broadcast radio. These newer systems typically provide short-range positioning.

Signal of Opportunity (SOOP): Some systems use signals without operator cooperation, such as mobile phone or broadcast signals, relying on calibration or reference stations for timing.

Acoustic signals are used for underwater ranging over a few kilometers. Ultrasound, infrared, and optical signals may be used for short-range positioning, typically within a single room.

Satellite Navigation: GPS, GLONASS, Galileo, and other systems provide global positioning, with a minimum of four satellites required for 3D position fixes and clock offset calibration by passive ranging. GNSS (Global Navigation Satellite Systems) offers high accuracy, but is susceptible to interference and obstructions. Each global GNSS constellation is designed to incorporate 24 or more satellites.

GNSS offers a basic positioning accuracy of a few meters. Differential techniques can improve this by making use of base stations at known locations to calibrate some of the errors. Carrier-phase positioning techniques can give centimeter accuracy for real-time navigation and can also be used to measure attitude.

- GNSS provides three-dimensional positioning, whereas other terrestrial solutions are limited to horizontal positioning because of their signal geometry
- GNSS also provides higher accuracy than the terrestrial systems, except for UWB
- only current position fixing technology to offer global coverage.
- vulnerable to incidental interference, deliberate jamming, and attenuation by obstacles such as buildings, foliage, and mountains.
- DME and enhanced Loran, provide a backup to outdoor GNSS and short-range systems provide coverage of indoor and dense urban environments where GNSS signals struggle

Combining multiple signal types maximizes position availability and robustness, offering a more reliable navigation solution. Summary of accuracy:

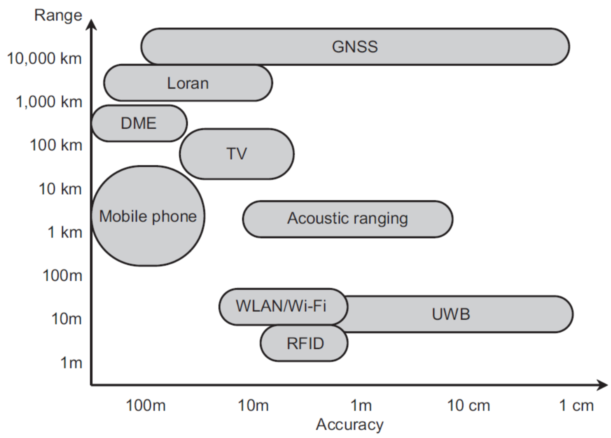

Figure 1.13 Range and accuracy of signal-based positioning technologies.

Figure 4: Range and accuracy

2.2.3.3 1.3.3 Environmental Feature Matching

- Natural Navigation: Humans and animals navigate using environmental features, often comparing them with maps, memory, or written directions.
- Static or Predictable Features: Features used for navigation must be either stationary or predictable in their movement.
- Historical Positioning: Historically, position fixes (LOPs) were made by manually identifying distant landmarks and using tools like a theodolite and compass.
- Modern Techniques: Image-based positioning techniques now automate the process, and devices like stereo cameras, radar, laser scanners, or sonar enable both angular positioning and ranging.
- Pattern Matching: Position can also be inferred directly from images through pattern-matching techniques.

Celestial Navigation:

- Using Sun, Moon, and Stars: Celestial bodies like the sun and stars have historically served as landmarks. For example, the sun's highest angle at the equinox equals the latitude.
- Star Navigation: The position can be determined by measuring the elevation angles of stars, and accurate time is required to calculate longitude. This has been practical since the 1760s, after advances in timekeeping by John Harrison.
- Modern Tools: Today, star imagers automate this celestial navigation process.

Terrain-Referenced Navigation (TRN):

- User Position from Terrain Height: TRN determines position based on the terrain height below the vehicle.
- Aircraft: Uses radar altimeters (radalt) or laser scanners to measure height above terrain, then compares it with the vehicle's height to determine the terrain height.
- Ship/Submarine: Sonar measures the depth of the terrain below.
- Land Vehicles: Inference of terrain height is made directly from their height solution.
- Pattern Matching: The vehicle's terrain measurements are compared with a terrain database to determine position.
- Accuracy and Limitations: Radalt-based navigation is accurate to about 50m, and works best over hilly or mountainous terrain. It is not effective over flat terrain.

Map-Matching Techniques:

- Constraints on Dead Reckoning: Land vehicles follow roads or rails, so map-matching uses this to constrain dead reckoning and correct positioning errors.
- Car Navigation: Map matching is crucial in car navigation, combining proximity and pattern-matching methods.
- Height Inference: Maps can also help infer terrain height from horizontal position.

Other Environmental Features:

- Magnetic/Gravity Anomalies & Pulsars: These can also be used for position fixing.
- Heterogeneous Feature Mix: A mix of different environmental features may be used to improve position determination.

Pattern Matching & Fault Detection:

- False Matches: Pattern matching may occasionally result in false or ambiguous position fixes.
- Fault Detection: It's important to implement fault detection and recovery techniques to address these potential errors.

2.2.4 1.4 The Navigation System

- Context-Specific Requirements: The design of a navigation system depends on the application and its operating context, which may provide additional information for the navigation solution.
- Integration of Multiple Technologies: Different positioning technologies should be combined to create an optimal navigation solution, with one technology potentially aiding another.
- Communications Link: A communications link can provide extra information to assist the navigation system, and navigation systems at different locations can communicate to cooperate or collaborate.
- Fault Detection and Correction: To ensure a reliable navigation solution, faults must be detected and corrected when possible.

2.2.4.1 1.4.1 Requirements Requirements for Navigation Systems:

Diverse Needs Across Applications: Navigation requirements differ significantly depending on the application. Factors include accuracy, update rate, reliability, budget, size, and mass. Some applications also require attitude solutions in addition to position and velocity.

Examples:

- High-Value, Safety-Critical Applications: Systems like those for airliners and ships require high integrity (ensuring navigation error bounds are reliable) and availability, with moderate accuracy needs and substantial budgets.
- Military Applications: Must function in electronic warfare, be stealthy, and accept some risk, with varying accuracy needs.
- Personal and Road Vehicle Navigation: Cost, size, weight, and power consumption are primary considerations.

Positioning Philosophies:

- High-Value Applications: Systems are custom-designed to meet strict requirements, with dedicated infrastructure and equipment.
- Lower-Value Applications: Utilize available sensors, radios, and environmental data originally intended for other purposes. Performance varies based on the context and available resources.

2.2.4.2 1.4.2 Context Context refers to the environment and behavior of the navigation system's host vehicle or user. It provides additional information to refine the navigation solution.

Examples:

Physical Constraints:

- Land vehicles stay near terrain; ships remain on water, constraining one dimension of position.
- Cars use roads, trains follow rails, and pedestrians avoid walls, providing positional constraints.

Dynamic Behavior:

- Vehicles and individuals have maximum speed, acceleration, and angular rates, which influence navigation calculations.
- Relationships between speed and turn rate help balance noise reduction with responsiveness.

- Motion constraints (e.g., wheeled vehicles' limited vertical movement) reduce the need of sensors required for dead reckoning or constrain error growth of INS.

Environmental Context:

- Indoor, urban, and open environments offer varying radio signals with different error characteristics.
- Behavior and navigation aids differ by environment (e.g., cars slow down and turn more in urban areas).
- Aircraft navigation varies based on height and whether flying over land or sea, while underwater environments lack most radio signals.

Adaptation to Changing Contexts:

- Navigation systems, especially mobile devices like smartphones, often switch between contexts (e.g., indoor to outdoor, stationary to moving).
- To maintain optimal performance, systems should detect and adapt to their context dynamically—a concept known as **context-adaptive** or **cognitive positioning**.

2.2.4.3 1.4.3 Integration An integrated navigation system combines two or more navigation technologies, either position-fixing methods or a mix of position-fixing and dead reckoning.

Integration Types:

Measurement-Domain Integration:

Combines raw measurements (e.g., ranges, bearings) into a common position estimation algorithm. Advantages:

- Allows systems to contribute even without enough data for an independent solution.
- Easier error characterization ensures optimal weighting.

Position-Domain Integration:

- Inputs position solutions directly from different systems.

Typical Architecture:

- Combines position-fixing (e.g., GNSS) with dead-reckoning systems (e.g., INS).
- Dead-reckoning provides a continuous navigation solution, while position-fixing corrects errors and calibrates dead-reckoning sensors.
- Corrections and calibration are managed by an estimation algorithm, typically a Kalman filter.

Benefits:

- Exploits complementary error characteristics of position-fixing and dead reckoning.
- Enhances robustness and accuracy of the integrated solution.

2.2.4.4 1.4.4 Aiding Position-fixing and dead-reckoning systems can be aided using the integrated navigation solution or other positioning technologies.

Dead Reckoning Aiding:

- Requires initialization of position, velocity, and sometimes attitude.
- Integration algorithms provide periodic corrections to navigation solutions and sensor outputs.
- Estimated sensor errors can be fed back to improve accuracy.

Position-Fixing Aiding:

- Pattern-matching systems need an approximate position input to limit the search area and improve efficiency.
- Signal-based systems can use this approximate position to search for signals effectively.
- Tiered positioning: One system provides a coarse position, aiding another system in achieving higher precision.

Velocity Aiding:

- Used in transect-based techniques like Terrain-Referenced Navigation (TRN) to create a location signature by combining measurements from different positions.
- Enhances sensitivity in radio positioning systems and compensates for motion effects in two-way ranging systems.

2.2.4.5 1.4.5 Assistance and Cooperation Assistance and Cooperation in Navigation Systems

Assistance:

- Utilizes a separate communication link to provide navigation systems with information about signals and environmental features for positioning.
- Includes data on transmitter locations, signal characteristics, landmark identification, and pattern-matching information.
- Benefits include access to up-to-date data, reduced onboard storage needs, faster position fixes, and support in areas with poor signal reception.
- Assistance data is often provided by commercial services, like mobile operators, with potential subscription fees.

Cooperative Positioning:

- Involves direct data exchange between nearby users over short-range links (e.g., **collaborative** or **peer-to-peer** positioning).
- Participants can include personnel, vehicle fleets, or public users, enabling cross-device cooperation (e.g., smartphone with a car or train).
- Shared data can include:
  - Signal quality and availability.
  - Transmitter clock offset and positions for signals of opportunity.
  - Positions and identification of environmental features.
  - Terrain height and calibration parameters.

Relative Positioning: Participants measure their positions relative to each other using proximity, ranging, or angular data. This is especially useful when there is insufficient information available to determine a stand-alone position solution.

2.2.4.6 1.4.6 Fault Detection Ensuring a reliable navigation solution involves detecting and addressing faults in user equipment, software, databases, or external components like radio signals. **Integrity monitoring** operates at three levels: **fault detection** (alerting the user to a fault), **fault isolation** (identifying and excluding faulty components), and **fault exclusion** (verifying the corrected solution). User-based monitoring uses redundant information to detect faults, while infrastructure-based monitoring relies on base stations to detect and alert users about signal faults. Safety-critical applications, such as civil aviation, require formally certified integrity monitoring systems to meet strict performance standards.

2.3 GPS segments

The GPS system consists of three segments:

1. The **space segment**, consisting of 24 or more satellites, with accurate atomic clocks on board, continuously transmitting ranging signals to Earth.
2. The **control segment**, consisting of a number of ground stations, which monitors the satellites, computes their orbits and clock offsets, and uploads this information to the satellites, which in turn encode this information on the ranging signal (the so-called navigation data).
3. The **user segment**, simply consisting of many GPS receivers, which each track four or more GPS satellites, and compute their own position.

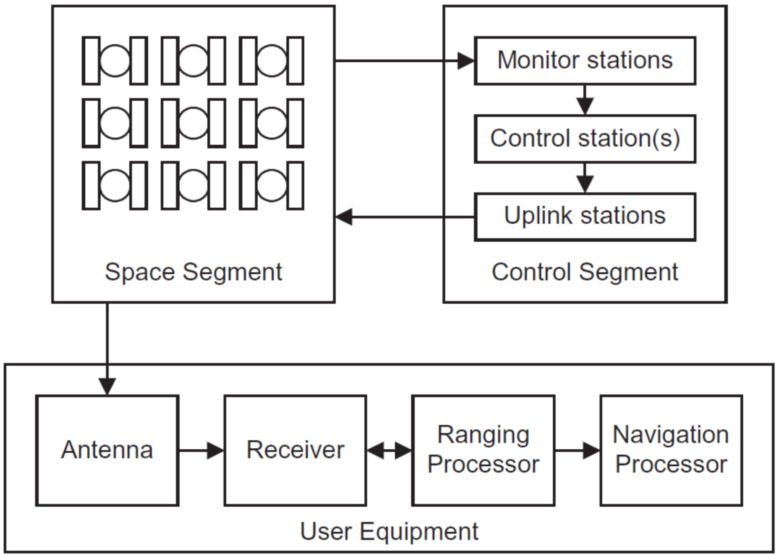

Figure 5: GPS segments

GNSS orbital planes are **inclined** with respect to the equator (at 55° for GPS). All the satellites form a **constellation**.

2.4 Radio Signal

2.4.1 GPS L1 CA-signal

The GPS radio signal is modulated through **Biphase shift key (BPSK) modulation**, with the amplitude of the signal given by:

$$s(t) = \sqrt{2P}C(t)D(t)\cos(2\pi f_{\text{ca}}t + \phi_0)$$

with:

- P: the signal power.
- $C(t)$: the **spreading code** (± 1). It is also called the **Pseudo Random Noise** (PRN) and is unique to each satellite, publicly available.
- $D(t)$: the **navigation data** (± 1). It contains the satellite orbit and clock information.
- f_{ca}: the **carrier frequency**. It is in the L-band between 1 and 2 GHz.
- t: time.
- ϕ_0: phase offset.

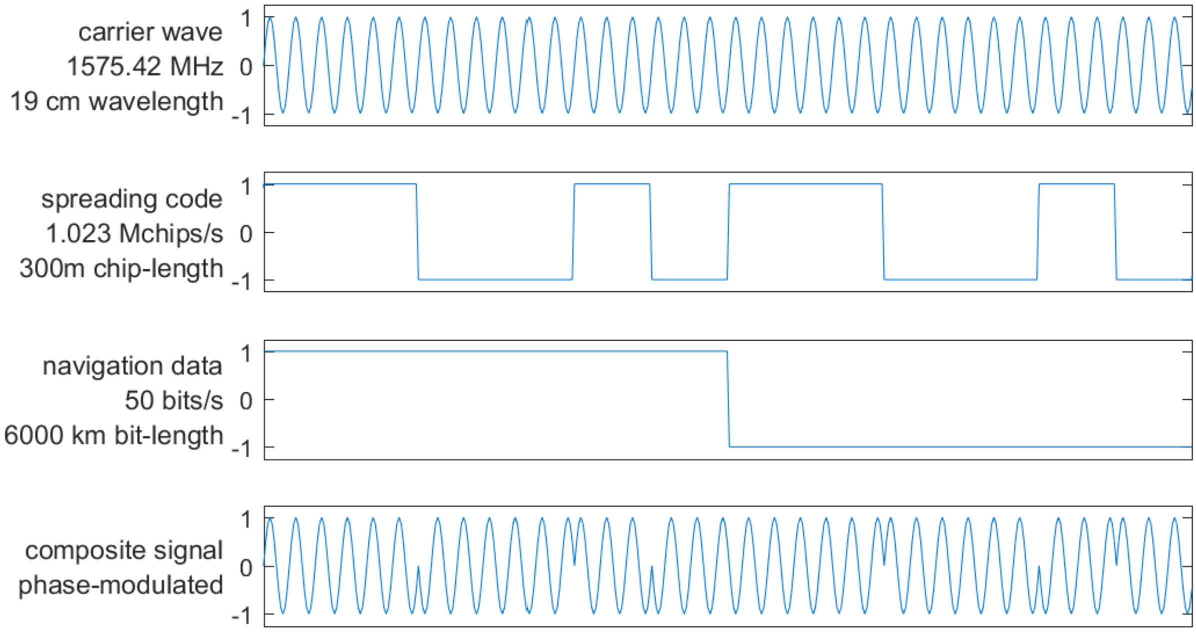

Figure 6: GPS L1 CA-signal (scale is not accurate)

2.4.2 Message Format

There are two main formats of **navigation data message**:

- **Fixed frame**: data always transmitted in same order with same repetition intervals
- **Variable frame**: a serie of fixed-length messages may be transmitted in any order

2.5 Initialisation

A GPS receiver typically consists of tens to hundreds of so-called **channels**, and will allocate each of these to a specific GPS (GNSS) satellite. When a GPS receiver first starts up, it will begin to *search for a particular GPS satellite on each of its channels*, by trying to scan for the corresponding **spreading codes** at different *Doppler offsets* and *time delays*. This is done by overlaying the received signal with a *local copy or replica of the same code* and then (time) shifting it until correlation shows a maximum (best fit, or match).

2.6 Pseudorange Measurement

Once the receiver is locked on to a satellite's spreading code, it will continue to track it by *continuously adjusting the time delay and Doppler offset* to keep the correlation at a maximum. The time delay is then used to calculate the **pseudorange** $p_{r,s}$ to the satellite, which is (theoretically) the travel time $\tau_{r,s}$ multiplied by the speed of light c:

$$p_{r,s} = c \cdot \tau_{r,s} \text{ where } \tau_{r,s} = t_r - t_s$$

with:

- t_s: signal transmission time (from satellite s)
- t_r: time of signal arrival (determined by receiver clock)

There are then two situations:

- **Signal acquisition**: pseudo-range prediction unknown, receiver-generated spreading code searched until correlation peak is found
- **Signal tracking**: pseudo-range prediction known, only vary the receiver-generated code phase slightly

Perceived carrier frequency varies due to: **Doppler effect** and **receiver clock drift**.

A **GNSS navigation solution** is *4D* with three position dimensions and one time dimension. For any satellite, the pseudo-range measurement, corrected for satellite clock error (and other known errors):

$$\rho(t_{s,a}) = \sqrt{(r_s(t_{s,t}) - r_a(t_{s,a}))^T \cdot (r_s(t_{s,t}) - r_a(t_{s,a}))} + \delta\rho(t_{s,a})$$

with:

- $r_s(t_{s,t})$: satellite position at time of signal transmission
- $r_a(t_{s,a})$: user antenna position at time of signal arrival
- $\delta\rho(t_{s,a})$: receiver clock offset

See Error Sources for more information on the errors in the pseudorange measurement.

2.7 Carrier Phase Measurement

A GPS receiver may measure the **fractional phase difference** between the received *carrier wave* from the satellite and a locally generated copy (replica). The carrier wave measurement is a **very precise measure** of the distance between the satellite and the receiver, but the initial number of carrier wave cycles is *unknown*, and needs to be estimated before the carrier phase measurements can be effectively used.

The much better precision of the carrier phase measurement with respect to the pseudorange code measurement can be explained by **much smaller period** of the carrier compared to the code chip duration (for the L1 CAcode signal, 1540 periods of the carrier fit in one chip of the Pseudo Random Noise (PRN) spreading code).

2.8 Jamming and Spoofing

There are multiple mays a GNSS signal may be threatened, jamming and spoofing being intentional attacks.

2.8.1 Jamming

By the time GNSS signals arrive at the antennas of a GNSS positioning system, the power level of these signals is very low. This low power level makes the signals susceptible to interference from other signals transmitted in the GNSS frequency range.

Jamming is a special case of signal interference where an attacker tries to block the incoming GNSS signal to a specific person/area.

GNSS receivers can use several methods to protect against interference and jamming:

- Signal filtering
- Multiple navigation sensors. For short-term interference, other sensors can help the receiver bridge brief periods of GNSS outage.
- Multi-frequency/multi-constellation GNSS makes it much harder to jam a signal on multiple different frequencies at once.
- Anti-jam antennas use multiple antenna elements to control the amount of signal received from a particular direction. When an anti-jam system senses interference from a direction, it turns down the antenna gain for it.

2.8.2 Spoofing

Unlike interference where GNSS is denied by overpowering the satellite signal, spoofing tricks the receiver into reporting an incorrect position. Spoofing is done by first jamming the GNSS receiver and then providing a false satellite signal that is either created by a signal generator or is a rebroadcast of a pre-recorded GNSS signal. Unlike interference, spoofing is always an intentional attack.

To protect against spoofing the same methods apply as against interference. Additionally, one of the most effective ways to protect against spoofing is to track encrypted signals that are broadcast by several of the GNSS constellations. Access to the encrypted signals is restricted and not available to all users.

2.8.3 Signal blockage

The GNSS signal can be blocked by many objects like trees or buildings, especially in urban areas. The main protection is again using multiple constellations and using additional sensors like an IMU.

2.8.4 Constellation failure

Although it is extremely unlikely that an entire constellation will fail, receivers that can track more than one constellation protect against this unlikely scenario.

3 GNSS performance

3.1 Error Sources

3.1.1 Pseudorange Calculation

Multiple issues affect the calculation of the pseudorange:

- the **satellite clock** has an **offset**, which is *known* (part of the navigation message)
- the **receiver clock** has an **offset**, which is *unknown*
- the **ionosphere** causes a delay (due to a lower speed of light), which is *unknown*
- there might be other errors, such as multipath, which are unknown

Any of these issues will cause the calculated pseudorange to be **inaccurate**. The calculation is very sensible since $c \approx 3 \times 10^8$ m/s, and a **1 µs** error will cause a **300 m** error in the calculated distance, since we have:

$$p_{r,s} = r_{r,s} + c \cdot (\delta t_s - \delta t_r)$$

where:

- $p_{r,s}$: pseudorange
- $r_{r,s}$: actual range
- δt_s: satellite clock offset
- δt_r: receiver clock offset

3.1.2 Ionosphere Delay

One of the *major error sources* in GPS is due to the **ionosphere**, which contains *free electrons* that cause the speed of light to be lower than in vacuum.

The ionospheric delay may be **highly variable**, as a function of both **time** and **space**. In terms of distance ranging, it can go from *a few meter to hundreds of meters*, and is maximum round the geomagnetic equator around local noon, and during solar maxima.

The ionosphere delay scales, to a very good approximation, with the **inverse of the square of the radio frequency** of the signal, so using two different frequencies allows to create the so-called *ionosphere-free range measurements*. This is why GPS satellites were originally designed to transmit ranging signals on both the **L1** (1575.42 MHz) and **L2** (1227.60 MHz) frequency.

3.1.3 Masking Angle

GNSS receivers **ignore signals** from below a certain elevation, making them prone to errors (typically between 5° and 15°).

3.1.4 GNSS Augmentation Systems

GNSS augmentation systems supply differential corrections and integrity alerts that meet the needs of safety-critical applications. There are two types:

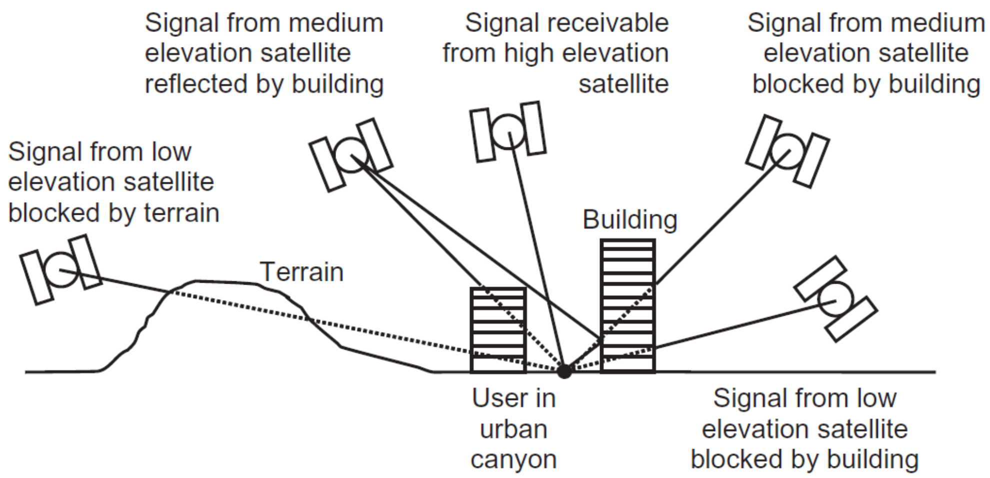

Figure 7: Errors due to terrain, buildings, and elevation angle

Criteria	Space-based augmentation systems (SBAS)	Ground-based augmentation systems (GBAS)
Coverage	Large country or small continent	Local area (e.g. an airfield)
Broadcast	Geostationary satellites	Ground-based transmitters
Precision	Lower than GBAS	Higher than SBAS

3.2 Accuracy and Precision

Accurate results may not be precise and precise results may not be accurate

- Accuracy = how close is the measurement to the actual value, can only be calculated if the ground truth is known
- Precision = how close is the measurement to other measurements, might be high even though some systematic error cause all measurements to be off

The receiver can measure the received **signal strength**, through the so-called carrier-to-noise-density ratio C/N_0, which gives an indication of the **quality of the measurement** (larger signal strength yields more precise measurement).

The **pseudorange measurement** precision is typically at the *one or few meter* level for low--cost, mass-market equipment, and can get down to the *few decimeter level* for professional highend equipment.

The **carrier phase measurement** precision ranges from the *few centimeter to the millimeter level*. The carrier phase is an ambiguous measurement of distance, but it is more precise than the pseudorange, typically by **two orders of magnitude**.

3.3 Dilution of Precision

The geometry of the visible satellites affects accuracy. Poor satellite geometry (e.g. satellites clustered together) increases DOP, leading to less precise positioning. Good geometry (satellites spread out) reduces DOP and improves accuracy.

$$\text{Position Error} = \text{DOP} * \text{Range Error (UERE)}$$

If DOP is very high, the inaccuracy of the computed position will be much larger than the inaccuracy of the range measurement.

DOP Formula:

$$\sigma_G = \sqrt{\sigma_E^2 + \sigma_N^2 + \sigma_U^2 + \sigma_T^2}$$

Where

- σ_G is the GDOP (Geometric Dilution of Precision) value
- σ_E^2, σ_N^2, σ_U^2 are the East, North and Up variance components of receiver position estimate
- σ_T^2 is Variance of receiver clock offset estimate

If DOP is very high, the inaccuracy of the computed position will be much larger than the inaccuracy of the range measurement

Types of DOP:

- GDOP (Geometric Dilution of Precision)
  - Amplifies pseudorange error of the position determination
  - $GDOP^2 = PDOP^2 + TDOP^2$
- PDOP (Position Dilution of Precision) - 3D uncertainty representation
  - $PDOP^2 = HDOP^2 + VDOP^2$
    $\sigma_P = \sqrt{\sigma_E^2 + \sigma_N^2 + \sigma_U^2}$
- HDOP (Horizontal Dilution of Precision) - 2D uncertainty representation $\sigma_H = \sqrt{\sigma_E^2 + \sigma_N^2}$
- VDOP (Vertical Dilution of Precision)
  - Typically, VDOP > HDOP because satellites 'below' the receiver cannot be acquired
  - This disparity increases with latitude due to less available satellites high in the sky
    * GPS constellation's orbit inclination is 55^o
    * There will never be a GPS satellite directly overhead at latitudes $> 55^o$ N/S
- TDOP (Time Dilution of Precision)

Assume only 4 satellites are being observed by a receiver. Maximising the tetrahedron formed by the 4 satellite points tends to minimize GDOP.

Conversely, DOP can be infinitely large (eg: 999.0 in assignment) when receiver-satellite vectors lie in the same plane. When this happens, the solution cannot distinguish between error in receiver clock and error in receiver position.

DOP values represent accuracy, with high DOP leading to poor precision. Impact of satellite position. DOP values change with time of day and location. For example, in high-latitude areas or urban environments, poor satellite geometry can lead to high DOP and lower positioning accuracy.

3.4 Availability, Continuity and Integrity

- Availability: the percentage of time that a sufficient amount of satellites have unblocked direct lines of sight (LOSs).
- Continuity: the ability of the total navigation system to continue to perform its function during the intended operation. Continuity is critical whenever reliance on a particular system is high. For a pilot during an instrument approach procedure, continuity and integrity are vital.

- Integrity: how much the information supplied by the system can be trusted to be correct. This requires the system to provide timely warnings to the user when the equipment is unreliable for navigation purposes—due to obstructions, jamming, multipath, or any other event that degrades accuracy. Almanac: contains information about which satellite is where at which time

3.5 PPP & RTK

3.5.1 Abbreviations

- **SV**: space vehicles or orbiting space vehicles
- **RTK**: Real-Time Kinematic
- **PPP**: Precise Point Positioning
- **PPP-RTK**: Hybrid of PPP and RTK
- **CORS**: Continuously Operating Reference Station
- **NRTK**: Network RTK
- **OSR**: Observation State Representation
- **SSR**: State Space Representation

3.5.2 PPP

Precise Point Positioning (PPP) is a GNSS signal augmentation technique that offers high accuracy positioning using a single receiver. The errors which PPP corrected:

- satellite errors
  - SV Orbit error
  - SV Clock error
  - SV Bias

3.5.2.1 Key features of PPP

- **Does not resolve carrier phase ambiguities**
  Instead, it uses an estimation, leading to a longer initialization time and requiring full re-initialization if the signal is lost. This is a key difference from RTK, which uses carrier phase measurements for precise positioning.
- **Eliminates GNSS system errors**
  PPP uses GNSS satellite clock and orbit corrections to achieve high-accuracy positioning without needing a local base station. PPP typically offers a decimetre accuracy (10 cm).
- **Relies on a global network of Continuously Operating Reference Station (CORS)**
  These stations generate the corrections needed to eliminate system errors.
- **Delivers corrections via satellite or internet**
  This allows for global coverage and removes the need for local infrastructure.
- **Provides dm-level or better real-time positioning**
  This accuracy surpasses standalone GNSS capabilities, making it suitable for applications requiring higher precision.
- **Requires a convergence period**
  Typically ranging from 5 to 30 minutes, this time is needed to resolve local biases such as atmospheric conditions, multipath, and satellite geometry.
- **Uses the State Space Representation (SSR) message format**
  This format separates and corrects individual error components, unlike OSR used in RTK.

- **Suitable for applications with no local infrastructure**
  Ex: sparsely populated areas and marine applications where setting up a network of base stations is challenging.

While PPP offers global coverage and good accuracy, its long convergence times can be a drawback for applications that require rapid positioning. Nonetheless, it's a valuable technique for situations where high accuracy is needed without relying on local base stations.

3.5.3 RTK

RTK, which utilises the OSR approach, corrects for location-dependent errors like Ionospheric and Tropospheric delays by providing a localized solution based on a network of base stations (CORS). The errors which RTK corrected:

- satellite errors
  - SV Orbit error
  - SV Clock error
  - SV Bias
- location-dependent errors
  - Ionospheric delay
  - Tropospheric delay

3.5.3.1 Key features of RTK

- **RTK enables the rover to resolve the ambiguities of the differenced carrier phase data and estimate the coordinates of the rover position.**
- **A CORS transmits its raw measurements or observation corrections to a rover receiver.**
  This is done via a direct (two-way) communication channel. The rover is a potentially moving receiver whose position is being determined.
- **Very high accuracy positioning over a short range (30–50 km).**
  This is due to the degradation of distance-dependent biases, such as orbit error and ionospheric and tropospheric signal refraction.
- **Within close proximities of the base station (10–20 km), RTK provides near-instant high accuracy positioning of up to 1 cm + 1 ppm.**
- **A direct communication channel is required between the rover and the base station.**
  Bandwidth limitations prevent large numbers of users utilising the same base station, making RTK ill-suited to mass-market applications.
- **RTK is the most popular GNSS signal augmentation technology.**
  It is used in industries such as surveying and agriculture, and is especially common in regions with well-developed CORS networks.
- **RTK uses the Observation Space Representation (OSR) approach.**
  This groups the errors together and provides the total correction measurements, rather than for the individual parameters. All parameters are updated at the same frequency regardless of their time sensitivity.

3.5.3.2 OSR Approach The OSR approach (and thus RTK) has high bandwidth requirements because:

- **It requires a two-way communication channel for each user.**
  Both the base station and the rover need to transmit data back and forth, increasing the amount of data being transferred.
- **OSR groups errors together and provides total correction measurements.**
  Instead of sending corrections for individual parameters separately, OSR sends all the corrections together. This leads to a larger data packet size compared to SSR, which separates individual error components.
- **All parameters are updated at the same frequency.**
  Regardless of the time sensitivity of each parameter, they are all updated at the same rate (the most time-sensitive one), leading to more frequent data transmissions and increased bandwidth usage.

The high bandwidth requirement of OSR is a major limitation, particularly for mass-market applications. If a large number of users were to utilise the OSR approach, current mobile networks would likely be overwhelmed. This is why OSR is not well-suited for applications like smartphones, IoT, and the automotive industry.

3.5.4 PPP-RTK

PPP-RTK is a hybrid GNSS signal augmentation technology that combines the strengths of both Precise Point Positioning (PPP) and Real-Time Kinematic (RTK). Here's a breakdown of its key features:

- **Utilises a network of CORS stations**
  Similar to RTK, it relies on a network of Continuously Operating Reference Stations (CORS) to generate corrections.
- **Provides atmospheric error corrections**
  PPP-RTK utilises a "un-differenced" map of atmospheric errors generated by a network of CORS, specifically for ionospheric and tropospheric delays, which are calculated using the CORS network. Achieving fast ambiguity resolution and high accuracy.
- **Enables fast convergence times**
  Thanks to the atmospheric error corrections, convergence times are significantly reduced, typically in the range of 1-10 minutes and potentially within seconds under ideal conditions.
- **Delivers cm-level accuracy**
  Comparable to traditional RTK techniques, PPP-RTK can achieve centimetre-level accuracy, exceeding the performance of standalone PPP.
- **Employs the State Space Representation (SSR) message format**
  Unlike RTK, which uses OSR, PPP-RTK uses SSR to broadcast corrections. This allows for efficient data transmission and enables an unlimited number of users to connect without overloading the system.
- **Has lower bandwidth requirements than RTK**
  The use of SSR and the efficient transmission of corrections result in significantly lower bandwidth requirements compared to RTK, making it suitable for mass-market applications.
- **Offers global coverage with graceful degradation**
  While it requires a regional CORS network, if a user moves beyond its range, the service seamlessly transitions to standard PPP, ensuring continuous positioning capability.

Overall, PPP-RTK offers a promising solution for mass-market applications by providing high accuracy, fast convergence, global coverage, and efficient bandwidth usage.
It bridges the gap between traditional PPP and RTK, offering a more versatile and scalable approach to high-accuracy positioning.

Solution	Benefits	Drawbacks
PPP	Has no local ground infrastructure requirements Global	Long convergence times Lower accuracy
RTK	High accuracy (2cm) Near-instant convergence times	Highly reliant upon local ground infrastructure Short range of transmissions
PPP-RTK	Fast convergence times High accuracy Lower density CORS network than NRTK Degrades to standard PPP	Reliant upon local ground infrastructure

Figure 8: PPP vs RTK

3.5.4.1 Exhibit 7: High-Level View of Main Benefits and Drawbacks of PPP-RTK Compared to PPP and RTK Only

3.5.5 Comparing RTK, PPP, and PPP-RTK

Feature	RTK	PPP	PPP-RTK
Accuracy	**cm-level** (up to 1 cm + 1 ppm)	**dm-level or better** (less than 10 cm)	**cm-level**, similar to RTK
Coverage Area	**Limited range** (typically 30-50 km from the base station)	**Global**	**Global** with graceful degradation to standard PPP outside the range of the CORS network
Message Format	**OSR** (Observation Space Representation)	**SSR** (State Space Representation)	**SSR** (State Space Representation)
Transmission Channel	**Two-way communication** between base station and rover	Corrections delivered via **satellite or the internet**	Corrections **broadcast to users**, enabling a large number of users to connect simultaneously
Convergence Time	**Near-instantaneous** (typically less than 5 seconds)	**Relatively long** (typically 5-30 minutes)	**Fast** (typically 1-10 minutes, potentially within seconds under ideal conditions)
Errors Solved	Orbit errors, clock errors, bias, **ionospheric delay, tropospheric delay**	Orbit errors, clock errors, bias	Orbit errors, clock errors, bias, **ionospheric delay, tropospheric delay**, enabling **integer ambiguity resolution**
Key Strengths	High accuracy, very fast convergence time	Global coverage, no reliance on local base stations	High accuracy, fast convergence time, global coverage, lower bandwidth requirements compared to RTK, graceful degradation outside CORS range
Key Limitations	Limited range, high bandwidth requirements, reliance on local base stations	Long convergence time, lower accuracy compared to RTK	Still requires a CORS network (though less dense than RTK) and may degrade to standard PPP with increasing distance from CORS station

OSR vs. SSR:

- **OSR:** Groups errors together, requires **two-way communication**, higher bandwidth requirements.
- **SSR:** Separates individual error components, enables efficient data transmission, **one-way communication (broadcast)**, lower bandwidth requirements.

The choice depends on the application's needs and available infrastructure.

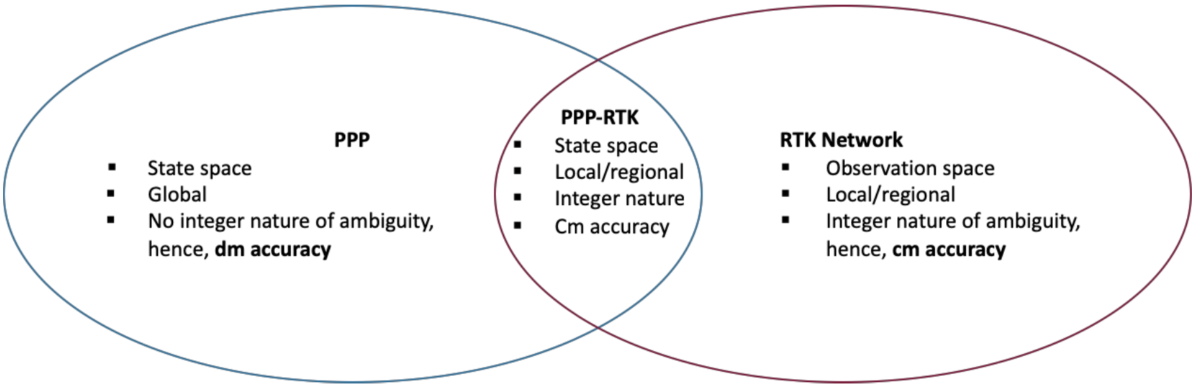

Figure 9: PPP vs RTK vs PPP-RTK

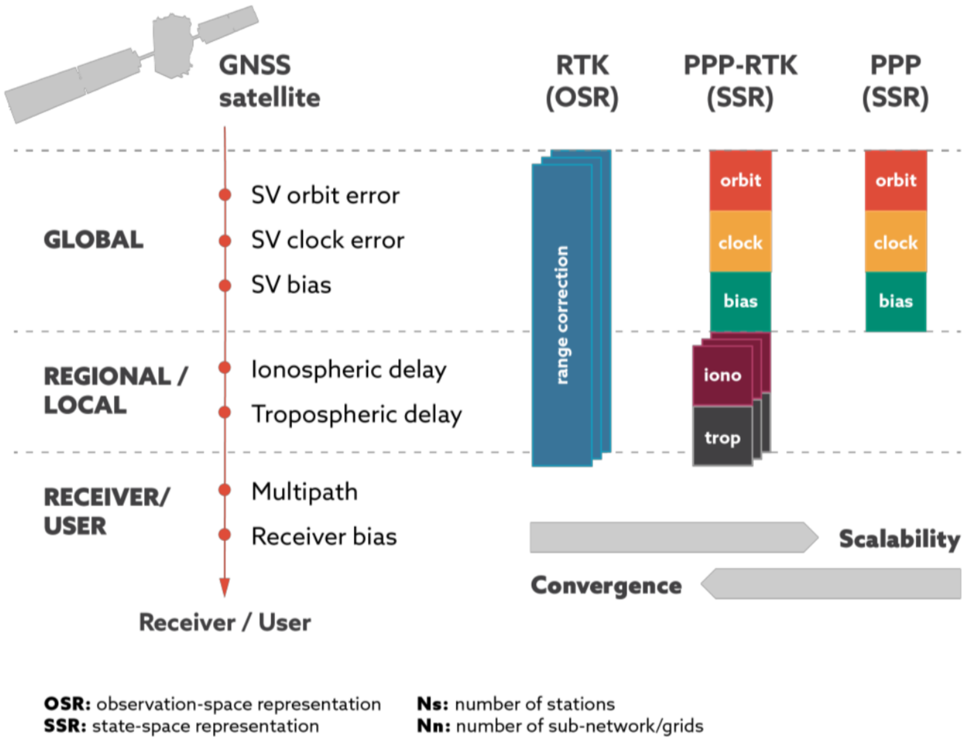

Figure 10: Difference in message format and resolved errors

3.6 Differential GNSS (DGNSS)

By comparing pseudo-range measurements with those made by equipment at a **pre-surveyed location**, known as a **reference station or base station**, the correlated range errors may be calibrated out.

- Improves the navigation solution accuracy
- The tracking, multipath, and Non-Line-of-Sight (NLOS) reception errors still exist (as it is uncorrelated between users at different locations).

3.6.1 How does it work?

DGPS uses a data link to a **nearby base** or **reference station** (GPS receiver at an accurately known position). The errors affect both receivers **almost identically** if the distance between them is small enough (5-10km).

1. From the differenced observations, the **baseline** (vector) between the two receivers can be computed through least-squares estimation.
2. The position of the rover is then obtained by adding the baseline vector to the accurately known coordinates of the reference station.

3.6.2 Local and Regional Area DGNSS

In a local area DGNSS(LADGNSS) system, corrections are transmitted from a single reference station to mobile users, sometimes known as rovers, within the range of its transmitter.

- **Range corrections** are transmitted, allowing the user to select any combination of the satellites tracked by the reference station.
- The user's navigation processor simply solves for the relative clock offset and **drift between the user and reference**, instead of the user receiver clock errors.

Regional area DGNSS (RADGNSS) enables LADGNSS users to obtain greater accuracy by using corrections from **multiple reference stations.**

3.6.3 Wide Area DGNSS

A wide area DGNSS (WADGNSS) system aims to provide positioning to meter accuracy over a continent, such as Europe, or a large country, such as the United States.

- Using much fewer reference stations than LADGNSS or RADGNSS would require.
- The key difference is that **corrections** for the different error sources are **transmitted separately**.
- Uses **satellite broadcasts** to disseminate correction data as compared to terrestrial radio or network transmissions.

Corrections:

- 10 or more reference stations at known locations send pseudo-range and dual-frequency ionosphere delay measurements to a master control station (MCS).
- The MCS then computes corrections to the GNSS system broadcast ephemeris and satellite clock parameters, together with ionosphere data, which are transmitted to the users.

3.6.4 Relative GNSS

Relative GNSS (RGNSS) is used where the user position must be known accurately with respect to the reference station, but the position accuracy with respect to the Earth is less important.

This relative position is known as a **baseline**. In RGNSS, the reference station transmits absolute pseudo-range measurements, which are then differenced with the user's pseudo-range measurements.

4 CRS

4.1 Coordinate Systems

4.1.1 Coordinate Reference Systems

According to the EPSG and ISO guidelines, four groups of CRSs can be distinguished:

- **Geographic CRSs** give geographic coordinates in degrees (2D latitude and longitude, and for 3D also height) relative to an ellipsoidal model of the Earth's surface. Examples include the European ETRS89, the global ITRF2020, and WGS 84. Geographic CRSs, such as ETRS89 and WGS 84, cannot be visualized on a flat plane without a map projection.
- **Projected CRSs** present geo-information on a flat surface in Cartesian (x and y) coordinates. A projected CRS is a derivative of a geographic CRS, where a map projection is used for the depiction on the flat surface. An example of this is the Dutch RD with the geographic CRS RD-Bessel as its basis.
- **Vertical CRSs** for recording height and depth relative to a reference plane. This reference plane is often based on the direction of gravity and normally does not coincide with the surface of an ellipsoid. With a reference plane based on the direction of gravity, no water flows between two points of equal height, but water usually does flow between two points of equal ellipsoidal height. Examples of vertical CRSs are the NAP and the LAT.
- **Compound CRSs** are composite CRSs, for example: RDNAP which is composed of the projected CRS RD and the vertical CRS NAP.

In this chapter of the summary, we will highlight two examples of Coordinate Reference Systems. Before doing so, here is an overview of some terminology and their examples:

Geodetic terminology:

- Terrestrial Reference System (definition) (e.g. ETRS89, WGS84)
- Terrestrial Reference Frame (realisation) (e.g. ETRF2000, ITRF2020)
- Ellipsoid (e.g. GRS80, Bessel 1841)
- 'Coordinate notation' (e.g. axis, order, units)
- Map projection (optional) (e.g. UTM, LCC)
- Epoch (for time-dependent coordinates) (e.g. 2024.91 (27 November 2024)

NB: be aware of the difference between epoch and the year in names!

Geo-spatial terminology:

- Datum ensemble != Terrestrial Reference System
- Datum = Terrestrial Reference Frame + ellipsoid
- Coordinate system = Coordinate notation (+ map projection)
- CRS (= projection in some software) = datum + coordinate system

4.1.2 Geographic Coordinate Reference Systems

The use of Geographic Coordinate Reference Systems is very common. They use degrees of latitude and longitude and sometimes also a height value to describe a location on the earth's surface.

Lines of latitude run parallel to the equator and divide the earth into 180 equally spaced sections from North to South. The reference line for latitude is the equator and each **hemisphere** is divided into ninety sections, each representing one degree of latitude. Wherever you are on the earth's surface, the distance between the lines of latitude is the same (60 nautical miles).

Lines of longitude, on the other hand, do not stand up so well to the standard of uniformity. Lines of longitude run perpendicular to the equator and converge at the poles. The reference line for longitude (the prime meridian) runs from the North Pole to the South Pole through Greenwich, England. At the equator, and only at the equator, the distance represented by one line of longitude is equal to the distance represented by one degree of latitude.

Using the geographic coordinate system, we have a grid of lines dividing the earth into squares that cover approximately 12363.365 square kilometres at the equator — a good start, but not very useful for determining the location of anything within that square. To be truly useful, a map grid must be divided into small enough sections so that they can be used to describe (with an acceptable level of accuracy) the location of a point on the map. To accomplish this, degrees are divided into **minutes** (') and **seconds** ("). There are sixty minutes in a degree, and sixty seconds in a minute (3600 seconds in a degree). So, at the equator, one second of latitude or longitude = 30.87624 meters.

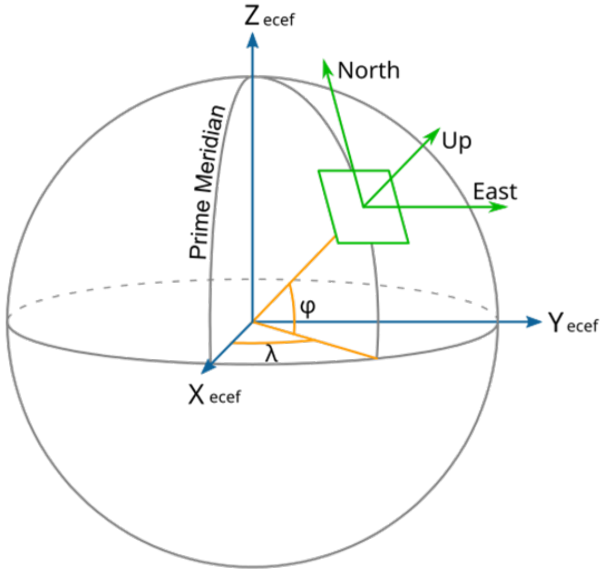

Figure 11: Fundamentals of a Geographic coordinate system

4.1.3 Projected Coordinate Reference Systems

There is a difference between a geographic coordinate system (GCS) and a projected coordinate system (PCS). In short, a GCS defines **where** the data is located on the earth's surface; a PCS tells the data **how** to draw on a flat surface. A GCS is round, and so records locations in angular units (usually degrees). A PCS is flat, so it records locations in linear units (usually meters).

The GCS is what ties your coordinate values to real locations on the earth. Only knowing the latitude and longitude of a location is thus not good enough, as it only tells you where a location is within a GCS. To draw a graticule, you need a model of the earth that is at least a regular spheroid, if not a perfect sphere. There are many different models of the earth's surface, and therefore many different GCS!

Once you know where to draw something, you need to know how. The earth's surface — and your GCS — are round, but the map is flat. This is where the map projections come in. They tell you how to distort the earth so the parts that are most important to your map get the least distorted and are displayed best on the flat surface of the map. A **projected coordinate system** (PCS) is a GCS that has been flattened using a map projection.

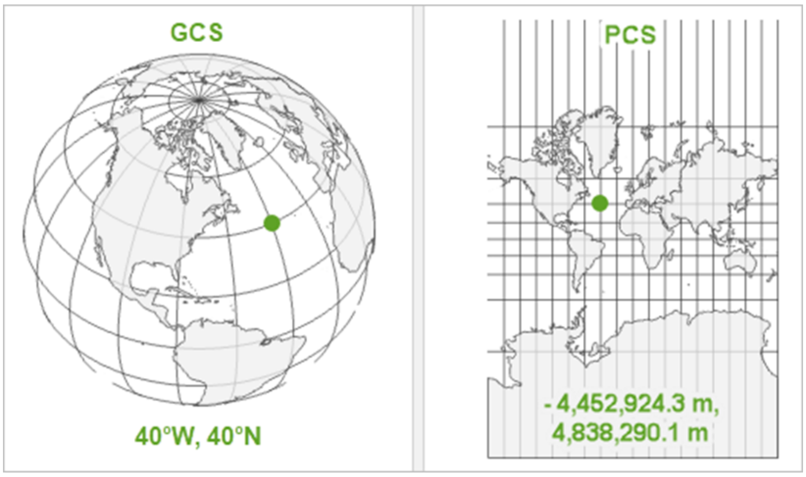

Figure 12: Geographic coordinate systems vs Projected coordinate system

Your data must have a GCS before it knows where it is on Earth. Projecting your data is optional, but projecting your map is not. Maps are flat, so your map must have a PCS to know how to draw.

Coordinates in a PCS are recorded in **a Linear Unit**, often meters. A PCS also contains a Geographic Coordinate System (e.g. WGS 1984)! Remember that a PCS is just a GCS that has been projected.

Coordinates in a GCS are recorded in an Angular Unit, usually degrees. The **Prime Meridian** is an arbitrary line of longitude that is defined as 0°. The **Datum** defines which model is used to represent the earth's surface and where that model is positioned relative to the surface. The **Spheroid** is the regular model of the irregular earth. It's part of the datum. **Semimajor Axis, Semiminor Axis**, and **Inverse Flattening** define the size of the spheroid.

4.1.4 Linear Reference Systems

Linear referencing is the method to store and geographically locate data using relative positions along a measured line feature without the need to explicitly use x,y coordinates or an address. When data is linearly referenced, measure values are used to measure the distance along a line feature, allowing multiple sets of dynamically changing attribute data to be associated with any portion of an existing linear feature, independent of its beginning and end. Linear referencing is used for many reasons. The following are the two primary reasons:

- Many locations are recorded as events along linear features.

  For example, locations of traffic accidents are recorded using a convention such as "27 meters east of reference mile marker 35 along State Highway 287." Many sensors record conditions using measures of distance or time along the lines—along pipelines, along roads, along streams, and so forth.

- Linear referencing is also used to associate multiple sets of attributes to portions of linear features without requiring that underlying lines be segmented (split) each time that attribute values change.

  For example, most road centreline feature classes are segmented where three or more road segments intersect and where the road names change.

Figure 13: Hectometre marker

Great examples of Linear Reference Systems are the Dutch **hectometre markers**. As the name suggests, Dutch hectometre markers are spaced at 100-metre intervals. In addition to showing the motorway number and location, they also bear a **carriageway identifier** – Li for Links (Left) and Re for Rechts (Right). The carriageways are identified as being left-hand and right-hand as viewed by somebody looking in the direction of increasing location numbers. By and large, Dutch location numbers increase as one moves away from Amsterdam, or in the case of roads that do not originate in Amsterdam, location numbers increase as one moves eastwards away from the North Sea.

Dynamic segmentation is the process of computing the map locations of events stored and managed in an event table and displaying them on a map using route features. The term dynamic segmentation is derived from the concept that line features need not be split (in other words, segmented) each time an attribute value changes; you can dynamically locate the segment.

Using dynamic segmentation, multiple sets of attributes can be associated with any portion of an existing linear feature independently of where it begins or ends. These attributes can be displayed, queried, edited, and analyzed without affecting the underlying linear feature's geometry.

4.2 Terrestrial Reference Systems and Frames

In this chapter of the summary, we will highlight some examples of Coordinate Reference Systems and Frames:

ITRS: International Terrestrial Reference System uses the GRS80 ellipsoid, with as its origin the centre of mass of the Earth. Greenwich (UK) is used as a meridian. When using time-dependent coordinates, you always need to specify the epoch (e.g. @2024.91).

ITRF: International Terrestrial Reference Frames are realisations of the ITRS. They publish updates every ~1-6 years. Due to the movement of tectonic plates, the differences between each iteration can be multiple centimetres to decimetres. Two examples are ITRF2014 and ITRF2020 (NB: Frame != epoch, e.g. ITRF2014@2022.90). These updates reflect:

- Improved precision of the station positions r(t0) and velocities r due to the availability of a longer period of observations, which is particularly important for the velocities,
- Improved datum definition due to the availability of more observations and better models,
- Discontinuities in the time series due to earthquakes and other geophysical events,
- Newly added and discontinued stations,
- Occasionally a new reference epoch t0.

WGS 84 is aligned with ITRS (WGS 84-G2296 (2024) = ITRF2020). For time-dependent coordinates, however, it has limited precision. The realisations (frames) often have differences

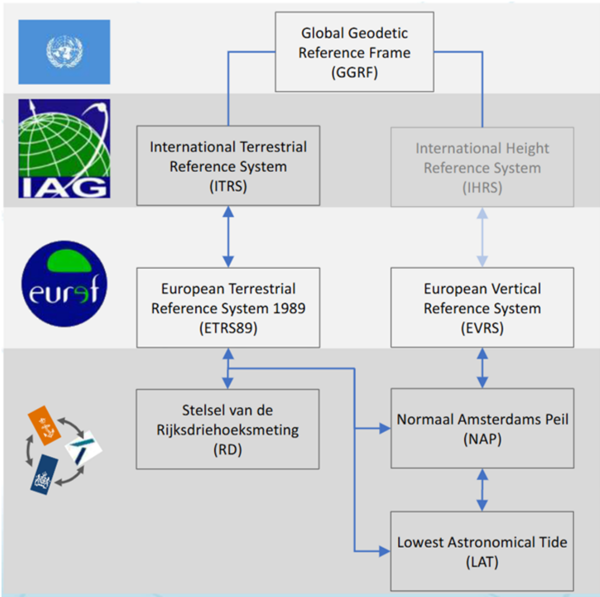

Figure 14: Terrestrial Reference Systems

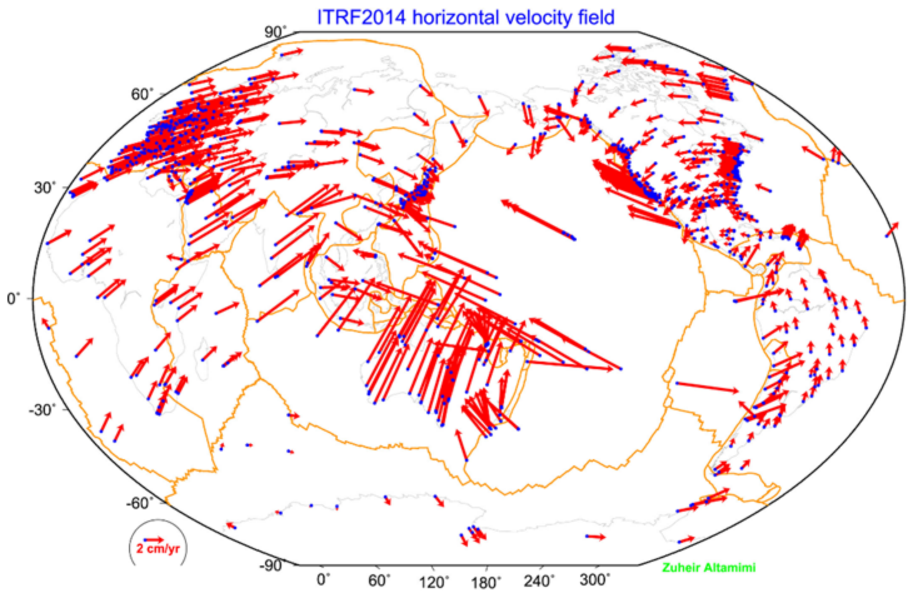

Figure 15: Changes in ITRF2014

between m – cm. The ensemble code (+- 2m): EPSG:4326 for 2D (often used as unknown latlon).

ETRS89: European Terrestrial Reference System 1989 is ITRF89 at epoch 1989.00. It moves with a stable part of Europe, "no" time-dependency. A new realisation (frame) comes with every ITRF. ETRF2000 is recommended for mapping and the ensemble code has an accuracy of +- 0.1m.

A transformation from ETRS89 to ITRS (and WGS84) is an example of a **Time-dependent transformation** (2.4 cm/year), but a specification of realisation and epoch is needed and most software does not include this transformation. A **Null transformation** is possible, simple and the current practice, but it is not future-proof, since the difference is growing (now 0.9m).

4.3 Datum and Transformations

4.3.1 Transformations and conversions

The International Association of Oil and Gas Producers (EPSG) used a *de facto* standard instead of an ISO standard. The EPSG collects all the different reference systems and their transformations. There are two different steps when working with 3D data. First, the coordinates need to be converted to a new coordinate system, after which the height values are transformed.

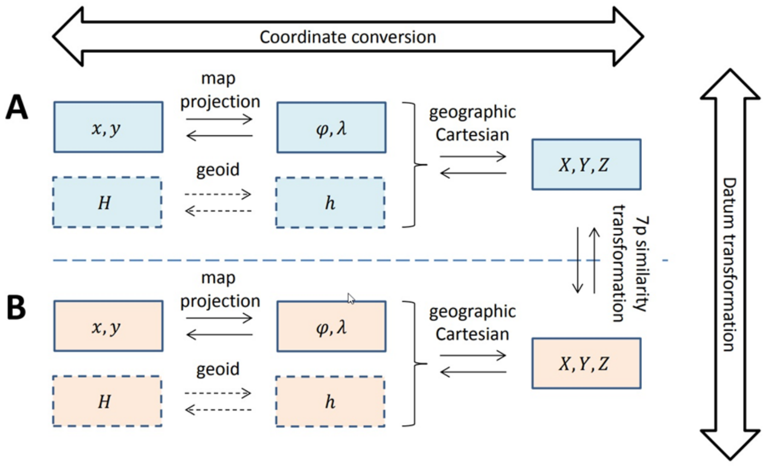

Figure 16: Map of coordinate conversion and datum transformation

Note, that a geoid is a stochastic height transformation and not a conversion. It uses in total 14 parameters:

- 7 parameters;
  - 3 Translations [m];
  - 3 Rotations ["];
  - 1 Scale factor [ppm];
- 7 rates and reference epoch (t0).

These parameters can be given (by an official source), user-estimated (empirical) and/or conventional (by definition). A correction grid can be used as an alternative transformation or an additional conversion.

A big difference between datum transformations and coordinate conversions is that the parameters for the datum transformation are often empirically determined and thus subject to measurement errors, whereas coordinate conversions are fully deterministic. More specific, three possibilities need to be distinguished for the datum transformation parameters:

1. **The datum transformation parameters are conventional**. This means they are chosen and therefore not stochastic. The datum transformation is then just some sort of coordinate conversion (which is also not stochastic).
2. **The datum transformation parameters are given but have been derived by a third party through measurements**. This third party often does new measurements and updates the transformation parameters occasionally or at regular intervals. This is also related to the concepts of reference systems and reference frames. Reference frames are considered (different) realisations of the same reference system, with different numerical values assigned to the coordinates of the points in the reference frame, and often with different realisations of the transformation parameters. The station coordinates and transformation parameters are stochastic, so new measurements, mean new estimates that are different from previous estimates.
3. **There is no third party that has determined the transformation parameters, and you as a user, have to estimate them using at least three common points in both systems**. In this case you will need coordinates from the other reference system. Keep in mind that the coordinates from the external reference system should all come from the same realization, or, reference frame.

4.4 Datums

When people in the field of Geomatics are talking about height, they can reference multiple different definitions of height:

- **Geometric**: Ellipsoidal height (max. MSL deviation ±150m)
- **Physical**: Height above the geoid ≈ Mean Sea Level (MSL)
- **Relative**: Height above ground level (DTM)
- **Water depth**: Lowest Astronomical Tide (LAT)

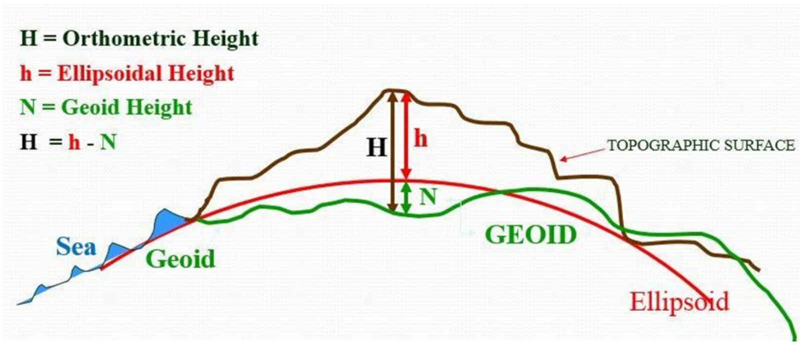

Figure 17: Physical heights in Geomatics

There are multiple physical height standards depending on where you want to know the height:

- **International Height Reference Systems** (IHRS) – No realisation yet
- **European Vertical Reference System** (EVRS) – Realisations available, but not widely used yet (dm – cm)
- **Earth Gravitation Model** (EGM84, 96, 2008, 2020) – By USA like WGS84 (m – dm)
- National Height Systems – Based on local MSL

4.5 Map Projections

Map projections try to portray the surface of the earth or a portion of the earth, on a flat piece of paper or computer screen.They try to transform the earth from its spherical shape (3D) to a planar shape (2D).

A coordinate reference system (CRS) then defines how the two-dimensional, projected map relates to real places on Earth. The decision of which map projection and CRS to use depends on the regional extent of the area you want to work in, on the analysis you want to do, and often on the availability of data.

When viewed at close range the earth appears to be relatively flat. Maps are representations of reality: they are designed to not only represent features but also their shape and spatial arrangement. Each map projection has **advantages** and **disadvantages**.

The best projection for a map depends on the **scale** of the map, and on the purposes for which it will be used. For example, a projection may have unacceptable distortions if used to map the entire African continent but may be an excellent choice for a **large-scale (detailed) map** of your country.

Different projection methods can be produced by surrounding the globe in a **cylindrical** fashion, as a **cone**, or even as a **flat surface**. Each of these methods produces what is called a **map projection family**. Therefore, there is a family of **planar/azimuthal projections (c)**, a family of **cylindrical projections (a)**, and another called **conical projections (b).**

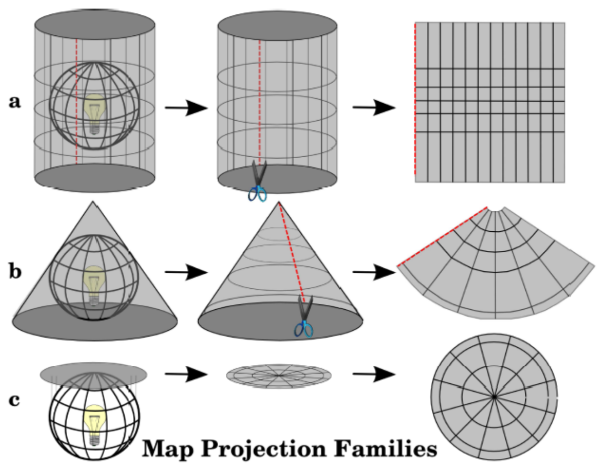

Figure 18: Map Projection Families

Map projections are never absolutely accurate representations of the spherical Earth. As a result of the map projection process, every map shows **distortions of angular conformity, distance and/or area**. It is usually impossible to preserve all characteristics at the same time in a map projection. This means that when you want to carry out accurate analytical operations, you need to use a map projection that provides the best characteristics for your analyses.

When working with a globe, the main directions of the compass (North, East, South and West) will always occur at 90 degrees to one another. In other words, the East will always occur at a 90-degree angle to the North. Maintaining correct **angular properties** can be preserved on a map projection as well. A map projection that retains this property of angular conformity is called a **conformal** or **orthomorphic projection**. These projections are used when the **preservation of angular relationships** is important. They are commonly used for navigational or meteorological tasks.

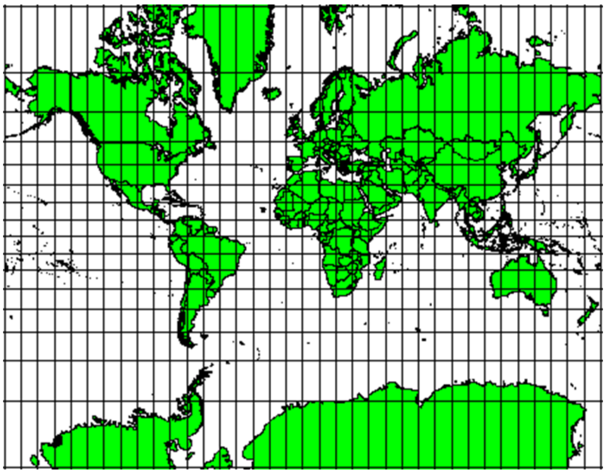

Figure 19: Conformal map projection

If your goal in projecting a map is to accurately measure distances, you should select a projection that is designed to preserve distances well. Such projections, called **equidistant projections**, require that the **scale** of the map is **kept constant**. A map is equidistant when it correctly represents distances from the centre of the projection to any other place on the map.**Equidistant projections** maintain accurate distances from the centre of the projection or along given lines. These projections are used for radio and seismic mapping, and navigation.

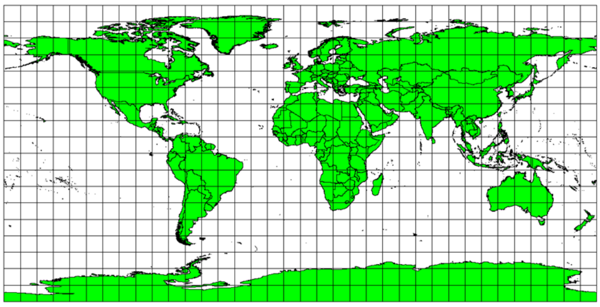

Figure 20: Equidistant map projection

When a map portrays areas over the entire map, so that all mapped areas have the same proportional relationship to the areas on the Earth that they represent, the map is an **equal area map**. As the name implies, these maps are best used when calculations of area are the dominant calculations you will perform. If, for example, you are trying to analyse a particular area in your town to find out whether it is large enough for a new shopping mall, **equal area projections** are the best choice. On the one hand, the larger the area you are analysing, the more precise your area measures will be, if you use an equal area projection rather than another type. On the other hand, an equal area projection results in **distortions of angular conformity** when dealing with large areas.

4.6 RDNAP

4.6.1 0. Coordinate Systems

Official 3D coordinate system of the Netherlands and Europe: European Terrestrial Reference System 1989 (ETRS89). ETRS89 is linked to the International Terrestrial Reference System (ITRS) by a time-dependant coordinate transformation. National coordinate systems in Europe are linked to ETRS89.

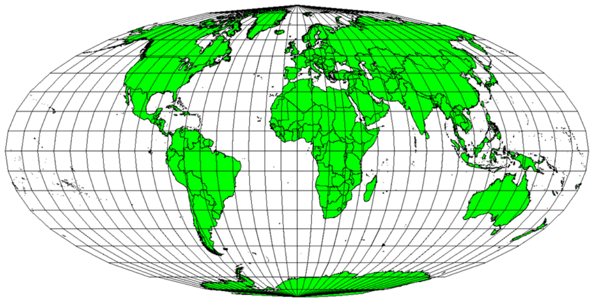

Figure 21: Equal area map projection

4.6.1.1 Rijksdriehoeksmeting (RD) Coordinates in the Dutch **Stelsel van de Rijksdriehoeksmeting (RD)** are the most-frequently used 2D coordinates on land and internal waters. RD coordinates are defined by the official transformation from ETRS89 coordinates. Maintaining reference points for ETRS89 and the transformation to RD coordinates are legal responsibilities of Kadaster.

The **Stelsel van de Rijksdriehoeksmeting (RD)** was created using triangulation from church spires and stone markers (historical). It has no time dependence; the differences are <1cm since 2000. It uses one EPSG-code (EPSG:28992 (RD New) for 2D and EPSG:7415 for compound CRS with NAP). The RD projection has its origin in Amersfoort in the middle of the Netherlands, it uses conformal stereographic projection (angles are preserved) and the north of the map is not equal to the true north. Unlike what you might think, the highest accuracy is not in Amersfoort itself, but rather in a circle ~100km around it.

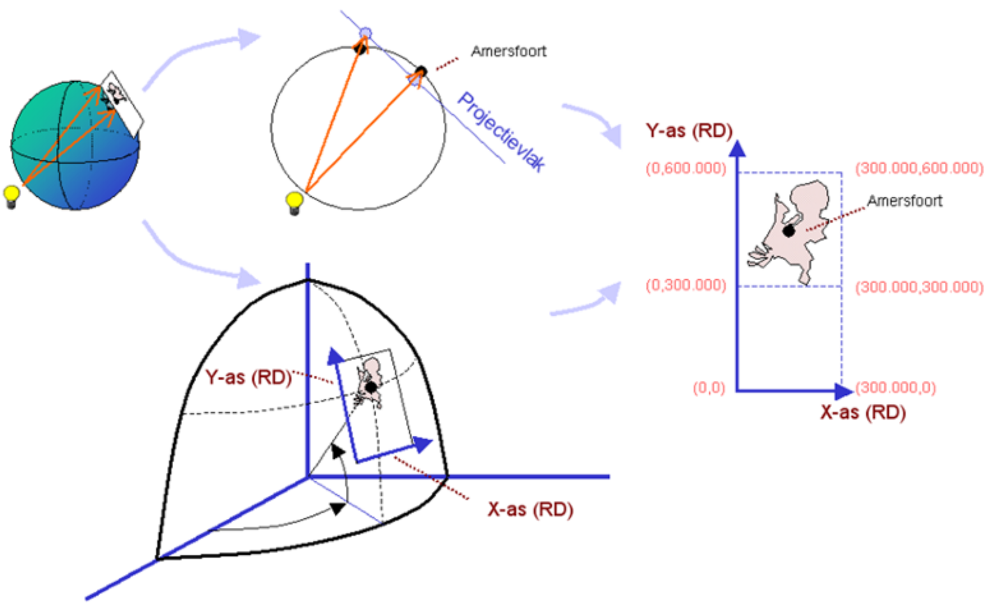

Figure 22: RDNAP diagram

To prevent confusion between the x-coordinates and y-coordinates, and to obtain always positive coordinates, the origin of the coordinates was shifted 155 km to the West and 463 km to the South (**False Easting and Northing**). This resulted in only positive x- and y-coordinates, where the y-coordinates are always larger than the x-coordinates.

4.6.1.2 Normaal Amsterdams Peil (NAP) The national height system on land and internal waters in the Netherlands is called **Normaal Amsterdams Peil (NAP)** and is based on the average summer flood in 1683-1684. Maintenance is based on point stability and it uses the NLGEO2018 geoid for GNSS measurements. Ellipsoidal heights in ETRS89 can be transformed with the quasi-geoid model to NAP with a precision higher than ETRS89 coordinates obtained with most GNSS measurements. Values range between 39.1 – 48.7m. The NAP is a legal responsibility of Rijkswaterstaat

For parts of the Noordzee however, the NAP cannot be used. Instead, they use the **Lowest Astronomical Tide** (LAT), which is the water depth in Worst-case astronomical conditions and average meteorological conditions.

4.6.2 1. Coordinate transformation

The official coordinate transformation between European ETRS89 coordinates and Dutch coordinates in RD and NAP is called RDNAPTRANS™. It uses a Datum transformation and a correction grid (see fig.) in combination with the map projection to transform the values. The height of the transformation is determined by a quasi-geoid.

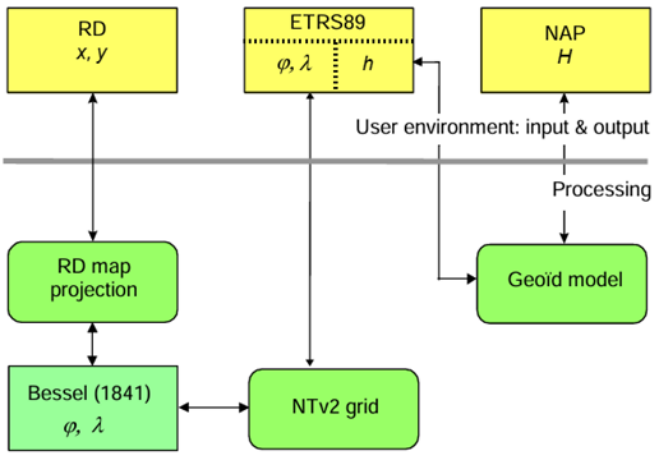

Figure 10.5: NTv2 transformation procedure used by RDNAPTRANS™2018. The figure outlines the relationships and transformations between ETRS89, RD2000 and NAP using the proposed NTv2 procedure, in variant 2 of RDNAPTRANS™2018 where the datum transformation is included in the correction grid. The coordinates below the line are used only for computational purposes and should never be published or distributed to other users.

Figure 23: NTv2 transformation procedure used by RDNAPTRANS™2018

Below are the errors of the RDNAPTRANS:

The recommended ETRS89 realisation is ETRF2000 at epoch 2010.50 (AGRS2010). When using RDNAPTRANS™2018 it is important to use this realisation and epoch, especially for the height. For applications demanding high accuracy, it is recommended to obtain the NAP height of the point of interest by levelling to nearby NAP benchmarks.

There are two variants for the implementation of the horizontal component of RDNAPTRANS™2018 and two variants for the vertical component (figure called Figure 1.2.2).

Implementation variant 1 applies the datum transformation as a separate step using a 3D similarity transformation. The advantage of implementation variant 1 is that it has no strict bounds for the area where horizontal coordinates can be transformed correctly. The disadvantage is that many software packages do not support implementation variant 1 for the horizontal component.

Figure 24: Errors of RDNAPTRANS™

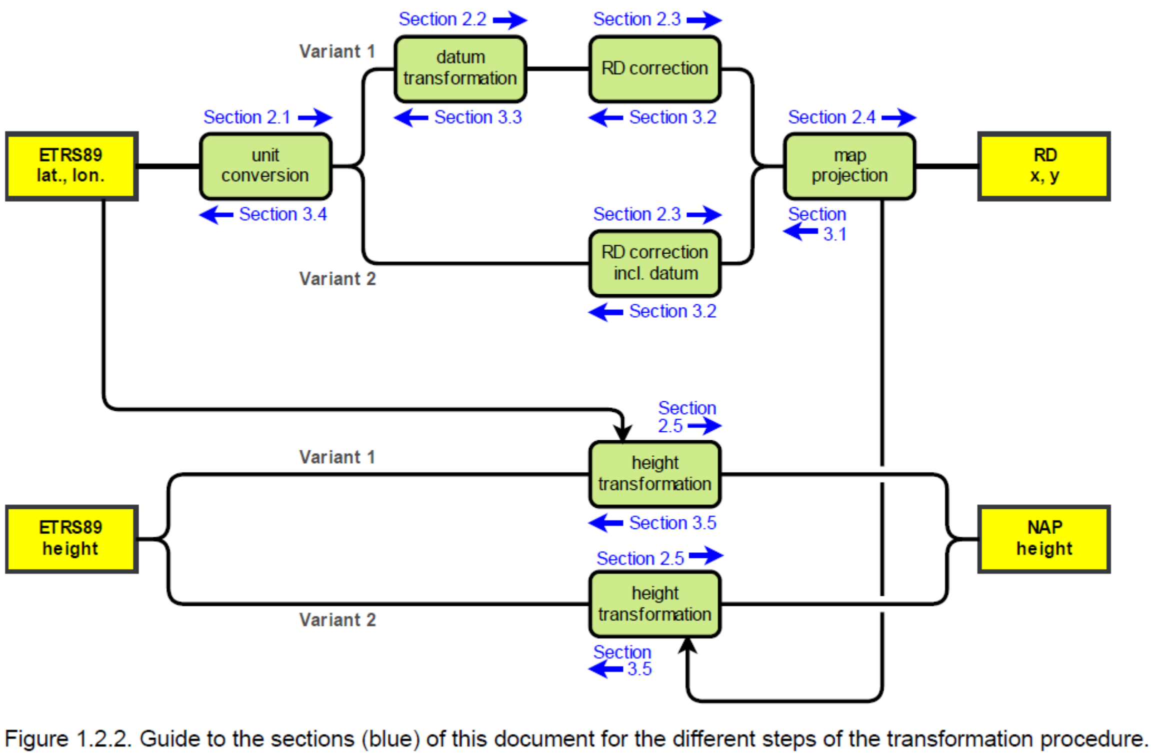

Figure 25: Figure 1.2.2

Implementation variant 2 includes the datum transformation in the correction grid and uses a different quasi-geoid grid for the height transformation. Implementation variant 2 for the horizontal component is supported by more software but can only be used within the bounds of the correction grid (Figure 1.1.1). The difference in the resulting coordinates between the two variants is well below 0.0010 m within the bounds of the RDNAPTRANS™2018 grids.

4.6.3 2. Transformation from ETRS89 to RD and NAP

4.6.3.1 2.1 Notation in degrees, minutes and seconds ETRS89 coordinates are commonly expressed in ellipsoidal geographic coordinates latitude, longitude and ellipsoidal height.

4.6.3.2 2.2 Datum transformation

4.6.3.2.1 2.2.1 Conversion to geocentric Cartesian coordinates Variant 1 - The ellipsoidal geographic ETRS89 coordinates of a point of interest must be converted to geocentric Cartesian ETRS89 coordinates to be able to apply a 3D similarity transformation. (Formula 2.2.1) Variant 2 - the datum transformation (Section 2.2.1, 2.2.2 and 2.2.3) is included in the correction grid (Section 2.3).

A fixed ellipsoidal height is used instead of the actual height of the point of interest. As a result, points with the same latitude and longitude in ETRS89 that differ in height get exactly the same RD coordinates. This enables 2D transformation between ETRS89 and RD and straightforward implementation in software like GIS packages. However, it introduces small differences between back and forth transformation.

4.6.3.2.2 2.2.2 3D similarity transformation The formula for a 3D similarity transformation must be applied to the geocentric Cartesian ETRS89 coordinates of the point of interest (Formula 2.2.2). The obtained geocentric Cartesian coordinates are in the geodetic datum of RD. Since the name RD is often used for projected coordinates only, the geodetic datum is often referred to as RD Bessel or just Bessel.

4.6.3.2.3 2.2.3 Conversion from geocentric Cartesian coordinates After the 3D similarity transformation, the geocentric Cartesian Bessel coordinates of the point of interest must be converted back to ellipsoidal geographic Bessel coordinates (Formula 2.2.3).

4.6.3.3 2.3 RD correction

4.6.3.3.1 2.3.1 Bilinear correction grid interpolation The ellipsoidal geographic coordinates of a point of interest obtained by datum transformation of implementation variant 1 are pseudo Bessel coordinates. Due to the error propagation of measurement noise of the original (1888–1928) measurements of RD, the pseudo Bessel coordinates must be corrected up to 0.25 m to obtain real Bessel coordinates. For implementation variant 2, the datum transformation is included in the correction grid (Section 2.3.4). The corrections are obtained from a regular grid of values for latitude correction and a regular grid of values for longitude correction, using bilinear interpolation (Formula 2.3.1).

4.6.3.3.2 2.3.2 Determine nearest grid points To transform the point of interest, the nearest NW, NE, SW and SE grid values are required. Grid values can be read one by one from the binary grid file by direct access or the entire grid of the binary or ASCII text file can be assigned to an array variable first.

4.6.3.3.3 2.3.3 Iterative correction The horizontal ellipsoidal geographic pseudo Bessel coordinates of the point of interest must be corrected to real Bessel coordinates (Formula 2.3.3) using the interpolated correction grid value of the point of interest. The horizontal ellipsoidal geographic coordinates of the correction grid points are in real Bessel. Therefore, also the coordinates of the point of interest are needed in real Bessel to determine the right correction. To solve this, the real Bessel coordinates are computed iteratively, until the difference between subsequent iterations becomes smaller than the precision threshold.

4.6.3.3.4 2.3.4 Datum transformation in the correction grid It is possible to include the datum transformation in the correction grid. The alternative grid for this implementation variant 2 contains the latitude and longitude corrections up to 0.25 m, but also the datum difference (about 0.1 km in the central part of the Netherlands). In that way the 3D similarity transformation (Section 2.2) is not needed.

4.6.3.4 2.4 Map projection

4.6.3.4.1 2.4.1 Projection from ellipsoid to sphere The corrected ellipsoidal geographic Bessel coordinates of a point of interest must be projected to obtain RD coordinates. The used RD map projection is a double projection. The first step is a Gauss conformal projection from the ellipsoid to a sphere (Formula 2.4.1).

4.6.3.4.2 2.4.2 Projection from sphere to plane The second step of the RD map projection of the point of interest is an oblique stereographic conformal projection from sphere to a plane to obtain RD coordinates (Formula 2.4.2).

4.6.3.5 2.5 Height transformation

4.6.3.5.1 2.5.1 Bilinear quasi-geoid grid interpolation The ellipsoidal height is not used with RD coordinates as it is purely geometrical and has no physical meaning. The height transformation from ellipsoidal ETRS89 height of a point of interest to NAP height is based on the quasi-geoid model NLGEO2018. The quasi-geoid height at the point of interest is obtained by bilinear interpolation of a regular grid of quasi-geoid height values (Formula 2.5.1).

To transform the point of interest, the nearest NW, NE, SW and SE grid values are required.

The horizontal coordinates of the grid points for which the quasi-geoid height is given are in ETRS89 (variant 1) or in Bessel (variant 2), but the quasi-geoid height is relative to the ETRS89 ellipsoid in both cases.

Implementation variant 1 uses the ETRS89 grid for transformation in both transformation directions, for ETRS89 to RD and NAP as well as RD and NAP to ETRS89. Using a different grid for the transformation back is not recommended, as it can result in too large differences after repeatedly transforming back and forth.

The quasi-geoid heights are in metres, spacing and coordinates of grid bounds are given in decimal degrees with conventional sign, thus east of the Greenwich meridian is positive.

4.6.3.5.2 2.5.2 Transformation to NAP The ellipsoidal ETRS89 height of the point of interest must be transformed to NAP height (Formula 2.5.2) using the interpolated quasi-geoid height of the point of interest.

4.6.4 3. Transformation from RD and NAP to ETRS89

4.6.4.1 3.1 Inverse map projection

4.6.4.1.1 3.1.1 Projection from plane to sphere RD coordinates of a point of interest must be converted to Bessel coordinates before the other steps of the transformation can be performed. The RD map projection is a double projection. The first step of the inverse map projection is an inverse oblique stereographic conformal projection from the RD projection plane to a sphere (Formula 3.1.1).

4.6.4.1.2 3.1.2 Projection from sphere to ellipsoid The second step of the inverse RD map projection is an inverse Gauss conformal projection from the sphere to the Bessel ellipsoid to obtain Bessel coordinates of the point of interest (Formula 3.1.2).

4.6.4.2 3.2 RD correction

4.6.4.2.1 3.2.1 Direct correction The ellipsoidal geographic coordinates of a point of interest obtained by the inverse map projection, are real Bessel coordinates. Due to the error propagation of measurement noise of the original (1888–1928) measurements of RD, the real Bessel coordinates must be corrected up to 0.25 m to obtain pseudo Bessel coordinates. For implementation variant 2, the datum transformation is included in the correction grid (Section 3.2.2). The corrections are obtained from a regular grid of values for latitude correction and a regular grid of values for longitude correction, using bilinear interpolation (Formula 2.3.1).

To transform the point of interest, the nearest NW, NE, SW and SE grid values are required The horizontal ellipsoidal geographic real Bessel coordinates of the point of interest must be corrected to pseudo Bessel coordinates (Formula 3.2.1) using the interpolated correction grid value of the point of interest. No iteration is needed for the transformation from RD to ETRS89 coordinates as the grid is given in real Bessel coordinates.

4.6.4.2.2 3.2.2 Datum transformation in the correction grid It is possible to include the datum transformation in the correction grid. In that way the 3D similarity transformation (Section 3.3) is not needed. With this alternative grid a bilinear interpolation of the latitude and longitude corrections (Formula 2.3.1) at the nearest grid points (Formula 2.3.2) and correction of real Bessel coordinates (Formula 3.2.1) can be applied as for a correction grid without the datum transformation, but in this case the output are ETRS89 coordinates of the point of interest instead of pseudo Bessel coordinates.

4.6.4.3 3.3 Datum transformation The corrected ellipsoidal geographic Bessel coordinates of a point of interest must be transformed to ellipsoidal geographic ETRS89 coordinates. This is only needed for implementation variant 1, for variant 2 the datum transformation (Section 3.3) is included in the correction grid (Section 3.2). First, the ellipsoidal geographic Bessel coordinates of a point of interest must be converted to geocentric Cartesian Bessel coordinates (Formula 2.2.1) to be able to apply a 3D similarity transformation.

Unlike conventional use in transformations, including RDNAPTRANS™2008 and earlier versions of RDNAPTRANS™, a fixed ellipsoidal height is used instead of the actual height of the point of interest. Points with the same latitude and longitude in RD that differ in height get exactly the same horizontal ETRS89 coordinates. This enables 2D transformation between RD and ETRS89 and straightforward implementation in software like GIS packages. However, it introduces small differences between back and forth transformation. These differences are below 0.0010 m up to 500 km outside the bounds of the RDNAPTRANS™2018 grids.

The 3D similarity transformation must be applied to the geocentric Cartesian Bessel coordinates of the point of interest to obtain geocentric Cartesian ETRS89 coordinates. The datum transformation uses the same formulas for the 3D similarity transformation from RD to ETRS89 as from ETRS89 to RD (Formula 2.2.2). So the order is first rotate and then translate, instead of inversing the order to translating and rotating. As a result, the inverse transformation parameters do not only have opposite sign, but the parameters also have slightly different numbers (except for the scale difference parameter). Sometimes the differences are only apparent in the last digit. The parameters could be computed from the inverse parameters too (Section 2.2.2).

After the 3D similarity transformation, the geocentric Cartesian ETRS89 coordinates of the point of interest must be converted back to ellipsoidal geographic ETRS89 coordinates (Formula 2.2.3). The latitude is computed iteratively. The parameters of the GRS80 ellipsoid (Section 2.2.1) are needed for the conversion to ellipsoidal geographic coordinates for the transformation from RD to ETRS89.

4.6.4.4 3.4 Notation in degrees, minutes and seconds The ETRS89 coordinates of a point of interest obtained after transformation are in radians or decimal degrees, depending on the type of goniometry functions used. These coordinates must be converted to decimal degrees and optionally to degrees, minutes and seconds (Formula 3.4).

4.6.4.5 3.5 Height transformation The physical NAP height of a point of interest can be transformed to the purely geometrical ellipsoidal ETRS89 height. The height transformation from NAP to ETRS89 is based on the quasi-geoid model NLGEO2018. The quasi-geoid height at the point of interest is obtained by bilinear interpolation of a regular grid of quasi-geoid height values (Formula 2.5.1). A unit conversion (Formula 2.1) is only needed when working with radians.

To transform the point of interest, the nearest NW, NE, SW and SE grid values are required. Grid values can be read one by one from the binary grid file by direct access or the entire grid of the binary or ASCII text file can be assigned to an array variable first. In both cases, the indices of the required grid values need to be determined. The horizontal coordinates of the grid points for which the quasi-geoid height is given are in ETRS89 (variant 1) or in Bessel (variant 2), but the quasi-geoid height is relative to the ETRS89 ellipsoid in both cases. Implementation variant 1 uses the ETRS89 grid for transformation in both transformation directions, for RD and NAP to ETRS89 as well as ETRS89 to RD and NAP. Using a different grid for the transformation back is not recommended, as it can result in too large differences after repeatedly transforming back and forth.

The NAP height of the point of interest must be transformed to ellipsoidal ETRS89 height (Formula 3.5) using the interpolated quasi-geoid height of the point of interest.

5 Indoor Positioning

5.1 Spaces

Perceiving and describing space:

- Empty or containing things
- Unlimited or bounded
- Physical or imaginary

Cell: is a bounded portion of space (a space unit)

Space in Positioning and Localization: partitioning space from the **sensor reception perspective**

Space classification according to reception of GPS signal:

- **Open outdoors**: outside building, open sky condition, enough satellites for positioning
- **Semi-outdoors**: outside building, slight coverage (e.g. wooded area), some satellites availability
- **Light indoors**: inside building, slight coverage (e.g. areas around windows), some satellites availability
- **Deep indoors**: inside building, no satellite coverage

Environment	Open Outdoors	Semi-Outdoors	Light Indoors	Deep Indoors
Definition	Outside a building	Near a building	In a room with windows	In a room without windows
Example				

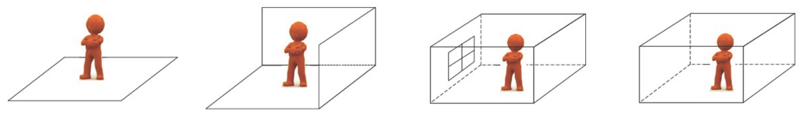

Figure 26: Types of spaces for GPS

Spaces are abstracted and represented using:

- Boundary Representation (BRep)
- Constructive Solid Geometry (CSG)
- Spatial Occupancy Enumeration

Navigation network (supported by Poincaré duality theory):

- **Nodes**: associated with space units, can contain semantic information about location
- **Edges**: represent connectivity between spaces
- **Costs** (of edges): indicate distance or travel time between nodes

Space partition in 3D:

- Bottom enclosure
- Side enclosure
- Top enclosure

Field	Classification	Physical Boundary
Navigation	Functional space, Object space, Remaining space & Indoor, Semi-indoor, Semi-outdoor, Outdoor	Architectural (e.g. wall, floor, roof, fence)
Positioning and Localization	Indoor, Semi-outdoor, Outdoor	Building, waterbody, bridge, tunnel

5.2 Wi-Fi-Based Approaches

Wi-Fi monitoring and fingerprinting are techniques used to gather information about wireless networks, but they differ in purpose, methodology, and application.

5.2.1 Wi-Fi Monitoring

- Main Idea: Wi-Fi monitoring involves passively observing and capturing Wi-Fi traffic (data packets) in the surrounding environment. This includes analyzing signals from access points (APs) and devices, such as SSIDs, signal strength, channel usage, and even packet contents (if not encrypted).
- Purpose: It's typically used for network troubleshooting, performance optimization, and security auditing.
- How It Works: A Wi-Fi adapter is set to monitor mode, allowing it to capture all wireless traffic in range, even if not destined for the monitoring device.
- Applications:
  - Analyzing traffic patterns and identifying potential interference.

5.2.2 Wi-Fi Fingerprinting

- Main Idea: Wi-Fi fingerprinting involves mapping and storing unique characteristics (or "fingerprints") of Wi-Fi signals at different locations to determine a device's location or context later.
- Purpose: It's primarily used for location-based services and indoor positioning systems (IPS).
- How It Works:
  - Offline phase: Wi-Fi signals (like Received Signal Strength Indicator, or RSSI) are measured and recorded at various locations to create a "radio map."
  - Online phase: The current Wi-Fi signal characteristics are compared to the radio map to estimate the device's location.
- Applications:
  - Indoor navigation and wayfinding (e.g., in malls, airports).
  - Asset tracking in warehouses.
  - Context-aware services like smart lighting or targeted advertisements.

5.2.3 Key Differences

Aspect	Wi-Fi Monitoring	Wi-Fi Fingerprinting
Objective	Traffic analysis, security, and troubleshooting	Location determination
Methodology	Passive traffic capture and analysis	Signal characteristic mapping and matching
Scope	Focuses on network behavior and devices	Focuses on spatial signal patterns
Output	Data about devices, networks, and traffic	Estimated location or spatial context

In summary, Wi-Fi monitoring observes and analyzes Wi-Fi traffic for network insights, while Wi-Fi fingerprinting leverages signal characteristics to provide location-based information.

5.3 Radio-based Localisation Techniques

5.3.1 Received Signal Strength (RSS)

RSSI values are Received Signal Strength (RSS) values averaged over a certain sampling period and are measured as a perceived power value P_r with the units of decibels. Using the physical property of **attenuation**, the gradual loss of intensity over a spatial propagation, RSS can be used to determine a distance d to an emitting source, as the output power value P_T is known. RSSI localisation uses trilateration with at least 3 transmitters.

$$P_r \propto P_T \frac{G_T G_R}{4\pi d^p}$$

Attenuation is characterised by a path loss exponent p which in free space is represented by a literature value $p = 2$, it is however heterogeneous in an indoor space. P_T is the transmitted power by the at the emitter. In theory, by calculating multiple perceived distances d from emitters to various receivers, position can be calculated via multi-lateration. This however this is subject to multi-path influences, introducing error.

Advantages:

- Simple and cost-efficient due to low hardware requirements

Disadvantages:

- Poor localization accuracy, especially in non-Line-of-Sight situations
- Additional signal attenuation from transmission through walls, obstacles, and multipath
- Potentially high fluctuation over time

5.3.2 Channel State Information (CSI)

- Captures amplitude and phase response across different frequencies
- Higher granularity than RSSI, gives more information to reduce multipath and provide more consistent measurements
- Overall higher localisation accuracy than RSSI

5.3.3 Fingerprinting/Scene Analysis

- Discrete estimation of user location based on grid
  - Delicate balance between size of grid (and corresponding resolution), and the signal variation error
  - Eg: impossible to estimate the correct point if the difference in RSSI values between each grid cell is less than the uncertainty range of the signal strength
- Offline phase: collect RSSI/CSI measurements to form a fingerprint map
- Online phase: compare real-time measurements with fingerprint map using algorithms:
  - Probabilistic method: use histogram/kernel to calculate probability of receiver being in x grid of fingerprint map
  - Artificial Neural Network: train a neural network (eg: Multi-Layer Perceptron) and provide real-time input with weights
  - k-Nearest Neighbours: average nearest measurement patches to estimate the location
  - Support Vector Machine (SVM): machine learning method applicable to indoor localisation

5.3.4 Angle of Arrival (AoA)

- Receiver's antennae array estimates the angle of transmitted signal by the time difference of arrival at individual parts of the array
- Advantages:
  - Receiver location can be estimated with just 3 transmitters in a 3D environment
  - Accurate estimation when transmitter-receiver distances are small
- Disadvantages:
  - Slight error in angle calculation translates to huge error in position determination
    * Thus not accurate at larger transmitter-receiver distances
  - Line of Sight for angle calculation is hard to obtain due to indoor multipath effects

5.3.5 Time of Flight (ToF)

- Calculate physical transmitter-receiver distance using signal propagation time and speed of light
- Trilateration is used similar to RSSI to estimate position
- Requires strict time synchronisation between transmitter and receiver, such that timestamps may also be included in the signal
- Advantages:
  - High sampling rate and large signal bandwidth increases the resolution of the position estimation
- Disadvantages:
  - Highly sensitive to obstacles as they deflect the signals, resulting in a longer propagation time

5.3.6 Time of Arrival (ToA)

This principle measures the absolute travel time of a signal from transmitter to receiver. A **euclidian distance** can be derived using the wave speed.

5.3.7 Time Difference of Arrival (TDoA)

- Create hyperbolas between each pair of transmitters
- Intersection point of all hyperbolas represents the position estimate
  - Ie: solve the system of hyperbola equations for X,Y,Z
- Difference between TDoA and ToF
  - ToF uses absolute signal propagation time between each transmitter and receiver
  - TDoA uses the difference between these propagation times
    * Synchronisation is thus only necessary between transmitters (receiver error is common between transmitter pairs)

5.3.8 Return Time of Flight (RToF)

- Measures the round-trip signal propagation time from transmitter to receiver, and response signal from receiver to transmitter
- Difference between RToF and ToF
  - Only moderate synchronisation between transmitter and receiver is required
  - All factors affecting ToF are worse in RToF due to twice the distance
  - Time delay in transmitting response signal (depending on the receiver's electronics) adds to position estimation error

5.3.9 Phase of Arrival (PoA)

- Measures phase difference from transmission to reception
- Typically used in conjunction with other methods (RSSI, ToF, TDoA) to enhance localisation accuracy
- High accuracy is only achievable with direct line-of-sight

5.3.10 Angle of Arrival (AOA)

This methodology aims to obtain the direction of an incident incoming wave signal. For this, directionally sensitive antennas are needed. By having sensors that are responsive to a specific direction in space, vector addition can be used to evaluate an angle of incidence (arrival).

5.3.11 Path-Loss

The ITU Model for Indoor Attenuation takes into account how radio waves propagate indoors. It models the path loss L, relative to the distance d between emitter and receiver.

$$L = 20 \log f + p \log d + c(k, f) - 28$$

Here f is the radio frequency and c is an empirical floor loss penetration factor. k is the number of floors between transmitter and receiver.

5.3.12 Fine Timing Measurement (FTM)

5.3.13 Radio Frequency Identification (RFID)

A Radio Frequency Identification system consists of an antenna which reads nearby active transceivers (transmit & receive) and passive tags on the radio wave part of the EM-spectrum. These transmit a unique tag ID. To position these tags in space the Cell of Origin (CoO) principle is used. A combination of ToA and AoA is used to pinpoint a location of an RFID tag.

Active RFID sensors are equipped with batteries which interrogate active radio transceivers. They are more costly than their passive opponent but offer long detection ranges of 30m or more. Location detection accuracies on the meter level have been proven possible.

Passive receivers rely on inductive coupling of EM waves and do not require batteries. These rely on receiving enough information by a scanner in order to transmit codes back. They are less costly, but only function in ranges smaller than 2 meters as signals need to travel two ways.

Applications of passive scanners include the identification of buried pipes as well as inventory control.

Using RFID scanners for indoor locations implies either having permanent active RFID sensors casting information or scanners suitable to detect passive ones. Both require large capital cost of installation.

5.3.14 Comparison

Technology	Range	Advantages	Disadvantages
Wifi 802.11 n/ac/ad	35m	Widely availableHigh accuracyExisting infrastructure	Sensitive to noiseComplex processing algorithms needed
Wifi 802.11 ah (new)	1km	High coverage rangeLow power consumption	New technology, not yet tested for indoor environments
Ultrawide Band (UWB)	10-20m	Immune to interferenceHigh accuracy	Short RangeAdditional hardwareHigh cost
Radio Frequency Indentification Device (RFID)	200m	Low power consumptionHigh coverage range	Low accuracy
Bluetooth	100m	High throughput High coverage rangeLow power consumption	Low accuracySensitive to noise
Ultrasound	<20m	Less absorption from obstacles	High dependence on sensor placement
Acoustics	<2m	High accuracy	Niche applicationsSensitive to sound pollutionAdditional hardware/anchor points
Visible Light (LEDs)	1.4km	Widely availableHigh potential (with AoA technique)Multipath-free	Requires Line-of-SightHigher power consumption

5.4 Hybrid and Other Techniques

5.4.1 Meshlium

Meshlium is a device that uses WiFi and Bluetooth scanners to detect other devices, which can be used for a range of applications/research (e.g. Vehicle Traffic Detection).

WiFi and Bluetooth radios (of devices) periodically send out messages, containing:

- MAC address of wireless interface
- Strength of the signal (RSSI)
- Vendor of the smartphone
- WiFi Access Point and Bluetooth friendly name
- Class of Device (CoD) (only when Bluetooth)

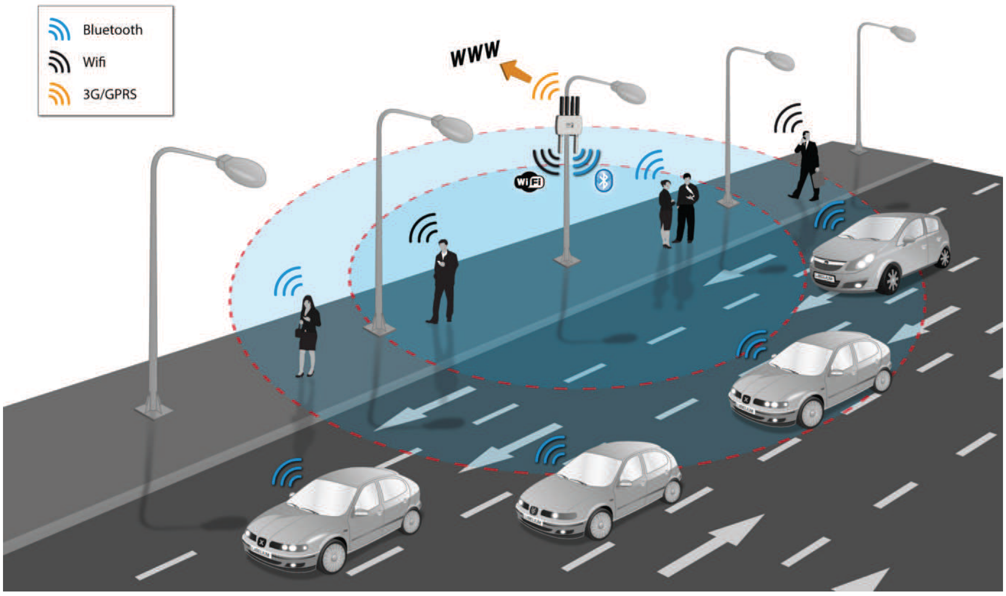

Figure 27: Meshlium Summary

MAC address randomization: for privacy randomized MAC address, reverts to **"factory" MAC address** when connected to WiFi

Adaptive Frequency Hopping (AFH): algorithm that enables Bluetooth radio to dynamically identify channels already in use and avoid them

5.4.2 Visual Based Indoor Localisation

- Image processing module: interpret visual data
  - Object recognition in images also help to contextualise the location
    - Eg: stove, fridge and sink indicate the place is likely a kitchen
  - Local feature descriptors: detect local interest points in an image, describe and store them as words
    - Bag-of-words model then compares the collection of word descriptors with the map (trained with places also described by bag-of-words) to match a place
    - Local features do not take the overall geometry into account
      - Thus pose-invariant: place is recognizable regardless of the position/orientation of the source image

        - However, adding geometric information improves robustness of place matching
    - Global feature descriptors: create a fingerprint of a location based on detected features
        - Uses color histograms, feature detection (edges, corners, color patches)
        - These features are ordered from 0^o to 360^o into a fingerprint, using omnidirectional cameras during training phase
        - Assumes the input live data is at similar height/location of the training data
    - Generally, combining both local and global descriptors provides the best results

- Map: maintains a representation of knowledge of the world
    - Usually a relational (topological/cognitive) map rather than absolute/geometric positioning
    - Consists of bounded places
        - Place signature: set of visual information that distinguishes it from other places
        - Gateway: physical boundaries of a place, where the physical appearances changes significantly
    - Methods:
        - Pure image retrieval: matching based only on image similarities, no position information is required/given
        - Pure topological map: stores relative positions of places, no metric information stored
            - Speeds up searching as indexing is possible
        - Topological-metric / Topometric map: enhance topological maps with direction and/or distance
            - Appearance-based option: metric information only between places, not within places
            - Sparse landmark option: metric information extracted from depth values between key landmarks inside the image
            - Dense occupancy grid option: same as sparse landmark but for more feature points, more GPU/memory-intensive

- Belief generation: combines information from above components to make decision on place familiarity
    - Bag-of-words model: TF-IDF scoring (term frequency - inverse document frequency)
        - Each visual word in image is scored by frequency of it appearing in image, against how common the word is across all images
    - Voting scheme: use multiple data streams to vote confidence of matching
        - Eg: multiple color bands that give unanimous voting and confidence value > threshold
    - Artifical neural network: Continuous attractor network (CAN)
        - Mimic neural network of a rat hippocampus using local excitation and global inhibition layers

Changing Environments

- Image processing module:
    - Invariant methods: focus on features that are invariant despite changing environments
        - Eg: edges and corners remain prominent despite lighting changes

        * This is also true for convolutional neural networks: mid-level features are robust to changes in the environment
        * Alternatively, use training images that are as 'change-invariant' as possible, or pre-process live data to reduce changes
    - Learning methods: define the relationship between how a place can appear across different times
        * Eg: use pairs of images between two different seasons, or day and night

- Map: how to deal with different representations of the same place
    - Remember and forget: Balance between new observations (that may be fleeting/inconsequential) and overwriting obsolete information
    - Multiple representations of the same environment: to capture cyclic/regular changes in environment
        * Eg: seasonal changes are cyclic
        * Store information of the same place (or whole-map level) at different instances of the required timescale

5.5 IndoorGML

Indoor location technologies are a diverse set of technologies that use different physical phenomena and processing techniques to gather and present information. All information has to come together for visualisation purposes, usually in the shape of a map. Compared to a visualisation of the outdoor environment, the specifications for indoor maps are not as easily defined. Therefore a standard, the OGC's IndoorGML was developed, serving as a standard data model.

Requirements
The required features of a map arise from the constraints the environments present. Other than for an outdoor space, the euclidean distance does not represent the actual distance from point a to b as doors, walls and ceilings might be in the way. Another difference is the purpose of the map. One might divide it up into rooms and hallways, however other representations like public and private areas is possible as well. This requires contextual information of the space, which might be more interesting than the 3D spatial extent of a room.

Rooms, also called cells, are therefore along with their 2- or 3D spatial extent given a set of attributes, such as a classification, usage etc.

5.5.1 Basic Concepts

The OGC published a set of requirements for indoor maps that include but are not limited to:

- Reflecting properties of indoor space
- Cellular space model
- Minimal set of specifications
- Interoperability with other standards
- Extensibility

The OGC's model is based on the cellular space model. A cellular space is defined as a set of non-overlapping cells, where each cell has an identifier and the union of cells is a subset of the entire indoor space. This is integrated by four main concepts:

- Cell geometry
- Topology between cells
- Cell semantics
- Multilayered space model

5.5.2 Cell Geometry

A 2D surface or 3D solid, defined by one of 3 options:

- No spatial information, only topology
- Spatial and topology
- Only topology, with reference to spatial data sets

5.5.3 Topology between cells

Important for navigation is the spatial relationship between cells. The connection of cells via doors is given via a dual topological graph. The nodes represent rooms, the edges represent doors. Several other attributes may be derived from these like accessibility or travel time between cells.

5.5.4 Cell Semantics

Given that all spaces have different functions, the semantics of cells need to be specified. This includes different applications like indoor facility management. It may also be useful to include a cell boundary as a semantic.

5.5.5 Multilayered Space Model

A space can and needs to be interpreted by different specifications of usage. Therefore a mechanism to include multiple layers for the same space must exist, called interpretations. Each one corresponds to a cellular space layer with its own geometric and topological properties. For example, the topology of a space might be different from the walking and wheelchair interpretation as they are partitioned differently.

6 Location awareness and privacy

6.1 Position, Location, Place and Area

Aspects of addressing space:

- **Reference**: relative (with reference to space or other objects) or absolute (agreed to by general consensus)
- **Specificity** and **Uncertainty**: the extent of the addressable space
- **Scope**: placement at different scales
- **Context**: with or without context

	Position	Location	Place	Area
Reference	Absolute (e.g. coordinate system)	Absolute (e.g. room number)	Relative, placement in a room (inside)	Relative, placement in an aggregation of rooms
Specificity/ Uncertainty	Depends on the device providing the position	Certain, defined by the physical borders (walls)	Uncertain, defined by the functional space of an object (e.g. desk)	Uncertain, defined by a more general notations (floors, parts of building)
Scope	Defined by a reference frame	Contains places	Contained in locations	Contains locations
Context	No context	Context	Context	Context
Example	"I am at 28.2314° - 33.4577°"	"I am in the living room"	"I am at the photocopier"	"I am on the second floor"

Four concepts of placement:

- **Position:** pin-point placements
- **Location:** smallest physically defined space in a building
- **Place:** placement of particular object and the uncertain (functional) space around it
- **Area:** generalised space or sub-space, containing multiple addressable locations

Framework modelling indoor space composed of:

- **Agents**: entities that navigate space, access resources and perform activities
- **Resources**
- **Space**: entirety of the enclosed environment to be navigated
- **Sub-spaces**
  - ***Inert spaces***: inaccessible by agents
  - ***Free spaces***:
    - Allow agents to move through them
    - Contain resources
    - Host activities
- **Modifiers**:
  - Can be applied to sub-spaces, agents and resources
  - Define the environment of a sub-space, a sub-space can be encumbered by multiple modifiers

- **Activities**

Network models: graph structure G(V, E) representing indoor space

- Nodes V: subdivisions
- Edges E: topological relationship between nodes

6.2 Personal Data Protection in the European Union

6.2.1 Data Processing Terminology

6.2.1.1 Personal Data Under EU law, personal data is defined in the Article 4 of the General Data Protection Regulation (GDPR) as:

> [A]ny information relating to an **identified** or **identifiable natural person** ("data subject"); an identifiable natural person is one who can be identified, **directly or indirectly**, in particular by reference to an identifier such as a *name*, an *identification number*, *location* data, an *online identifier*, or to *one or more factors* specific to the physical, physiological, genetic, mental, economic, cultural, or social identity of that natural person
>
> *GDPR, Article 4(1)*

A person who's data is being processed is a '**data subject**'.

If information relates to an identified or identifiable person it is personal data. This includes opinions of a person by someone else, for example the results of a workplace assessment.

Information that can be used to identify a person includes:

- Name
- Identification number
- Location data
- Online identifier
- Vehicle registration number
- Physical characteristics
- Genetic data
- Cultural identity

Name, date of birth and place of birth are rarely enough to identify a single person, but if used in conjunction with each other they can single someone out. The **metadata** should also be considered, as it sometimes contains even more information. For example, the **metadata of a picture** taken with a smartphone can contain the GPS coordinates, the date and time, the author, the camera model and the settings of the camera.

The concept of identifiability is explained by the Recital 26 of the GDPR:

> To determine whether a natural person is **identifiable**, account should be taken of all the means that are **reasonably likely** to be used, such as detection, by the controller or another person, to identify the natural person directly or indirectly. To determine whether the means are reasonably likely to be used to identify the natural person, account should be taken of all objective factors, such as the **cost and time required** for identification, taking into account the **technology available at the time of processing** and **technological developments**
>
> *GDPR, Recital 26*

This approach is called the **risk-based approach**. Data should be considered personal data if there is a high enough risk that it can be used to identify a person with reasonable effort. It also means that non-personal data:

- Could theoretically be used to identify a person, but the risk is low enough that it is not considered personal data
- Could become personal data with technological advancements

The form that personal data takes is not relevant to the laws that govern its usage. CCTV footage, recorded audio, pictures, DNA samples and digital communications are all examples of personal data.

Different countries had *different interpretations* of the relevance of context:

- Countries adopting the **context irrelevance** could create classes of data that are always or never personal data, regardless of the context
- Countries adopting the **context relevance** instead classify (almost) all data as potentially personal data under the right circumstances

All in all, this problem of distinguishing personal data from non-personal data must be handled as a **dynamic problem**. The controller must continuously **monitor the technological advancements and the capabilities of other actors** to adopt the right measures in due time.

6.2.1.2 Anonymisation Data can be kept in a form that allows for identification **no longer than is necessary for the purposes** for which the data is being processed. After personal data has served its purpose it needs to either be **erased** or **anonymised**.

Data is anonymised when all identifying elements are removed. There is no one-size-fits-all solution for anonymisation. The optimal solution needs to be determined on a case-by-case basis. Data is anonymised if no parts of that data can, with reasonable effort, lead to the identification of a natural person. What constitutes reasonable effort depends on the nature of the anonymised data, the context of their use, the available technologies and related costs.

Data that has been anonymised properly is no longer considered personal data and therefore data protection legislation no longer applies.

However, it was shown by many studies that it is possible to identify an individual through the combination of various anonymised datasets. This process is called **re-identification**. For example, a study conducted in 2014 by the MIT showed that it was possible to identify individuals in an anonymised dataset of credit card transactions by combining it with other datasets. Even though the dataset only contained only metadata (amounts spent, type of store, code per person), the researchers managed to extract patterns and track the spending of 1.1 million people.

Therefore, pretending to achieve **anonymisation that is permanent as erasure is utopic**. But it is still better than leaving the data in its initial state, reducing the risk to its lowest possible level.

6.2.1.3 Pseudonymisation In Article 4 of the GDPR, the concept of **pseudonymisation** is defined as:

> [P]rocessing of personal data in such a manner that the personal data can no longer be attributed to a **specific data subject** without the use of **additional information**, provided that such additional information is **kept separately** and is subject to **technical and organisational measures** to ensure that the personal data are not attributed to an identified or identifiable natural person

GDPR, Article 4(5)

Encrypted data is an example of pseudonymized data, as an additional encryption key is required to make sense of the information. Unlike anonymized data, pseudonymized data is **still considered personal data** under EU law.

6.2.1.4 Data Processing **Data processing** covers a large number of possible actions. Examples include:

- collection
- organisation
- structuring
- storage
- alteration
- retrieval
- usage
- disclosure
- restriction
- erasure

Automated and non-automated processes both count as data processing.

6.2.1.5 Users of Personal Data There are two types of entities that handle personal data: **controllers** and **processors**. A controller is a natural or legal person that determines the purpose and means of processing. A processor is a natural or legal person who processes the data on behalf of the controller. A controller oversees and controls the processing, as well as being responsible and legally liable.

A processor becomes a controller if they determine the means and purposes of data processing themselves.

Any person to whom personal data are disclosed is a 'recipient'.

6.2.2 Lawfulness, Fairness and Transparency of Processing Principles

6.2.2.1 Lawfulness of Processing **Lawful processing** of personal data requires the **consent of the data subject** or **another legitimate reason**. The other five reasons are:

1. When processing personal data is necessary for performance of a **contract**.
2. For the performance of a task by a **public authority**.
3. For compliance with a **legal obligation**.
4. For the purpose of the **legitimate interests** of the controller or third parties.
5. Or if necessary to protect the **vital interests of the data subject**.

6.2.2.1.1 Consent Controllers have a duty to keep a verifiable record of any consent received. Consent can be **withdrawn at any time**. The four characteristics of consent are:

1. **Free**: Consent must be freely given. Consent is only valid if the data subject was given a real choice with no risk of intimidation, deception or coercion.
2. **Informed**: The data subject must have sufficient information before making a decision. The purposes for which the personal data is necessary should be precise and easily understandable.
3. **Specific**: For consent to be valid it must also be specific to the processing purpose. If the processing is changed or expanded the data subject must give consent again.

4. **Unambiguous**: There should be no reasonable doubt that the data subject wanted to express their agreement to the processing of their data.

6.2.2.2 Fairness of Processing

Data subjects should **be notified** by controllers that they are processing their data in a lawful and transparent manner, and should be able to demonstrate that they are doing so. Processing should not happen in secret and data subjects should be aware of any potential risks.

6.2.2.3 Transparency of Processing

Controllers are obligated to take appropriate measures to ensure that data subjects remain **informed** about how their data is being used. Transparency can mean:

- The information given to a data subject before processing.
- The information that should be available to the subject during processing.
- The information given when a data subject requests to access their own data.

Processing operation must be explained to the data subject and a way that is easily accessible.

6.2.3 Data Processing Principles

6.2.3.1 The Principle of Purpose Limitation

Data subjects must know what the processing of their personal data will entail before processing is started. Data cannot be processed further in a way that is **not compatible with the original purpose**, although exceptions are possible if the new purpose is either:

- **Archiving** purposes in the public interest.
- **Scientific** or **historical** research.
- **Statistical** purposes.

Every new purpose for processing data which is not compatible with the original one must have its own particular legal basis.

6.2.3.2 The Data Minimisation Principle

Processing of personal data must be **limited** to what is **necessary** to fulfil a legitimate purpose. The processing of personal data should only take place if the purpose of processing said data cannot be achieved through other means. Data processing is not allowed to disproportionately interfere with the interests, rights and freedoms at stake.

6.2.3.3 The Data Accuracy Principle

A controller holding personal data is not allowed to process said data without ensuring with reasonable certainty that the data are **correct** and **up to date**. Inaccurate data must be **erased** or **rectified** without delay. It is possible that data may need to be checked regularly so as to ensure its accuracy. In some cases it may be legally prohibited to update stored data, as the purpose of storing the data is to provide a historical ‘snap-shot’.

6.2.3.4 The Storage Limitation Principle

Data must be **deleted or anonymised** as soon as they are **no longer needed** for the purposes for which they were collected. The controller of the data should establish a time limit for erasure or periodic review.

6.2.3.5 The Data Security Principle Controllers of personal data are required to implement appropriate **technical or organisational measures** when processing data. How appropriate a security measure is depends on the context and is determined on a **case-by-case basis** and should be regularly reviewed. Pseudoanonymizing data is a way in which to secure personal data. Another example of a security measure is if controllers offer the option of two-factor authentication, is this will make it more difficult for unlawful access of personal data. Data controllers are required to communicate with supervisory authorities if a breach of personal data occurs. If the breach is likely to result in a risk to a data subjects rights or freedoms, they should also be informed.

6.2.3.6 The Accountability Principle Controllers and processors are required to **actively and continuously** implement measures to promote and safeguard **data protection** in their processing activities. Controllers must be able to prove compliance with data protection laws at any time. Processors also have obligations, such as keeping a record of processing operations.

6.2.4 Specific to Location Data

6.2.4.1 Sources of Location Data Location data comes from a variety of sources, including:

- **GNSS**, using satellites to get an accurate position
- **Wi-Fi**, using only the detection of the MAC address of close-by access points
- **Cell Phone Tracking**, using the cell phone towers of telecommunication operators to get an approximate position
- **Bluetooth Beacons transmitters**, using the detection of the MAC address of close-by beacons

Modern devices use a **combination** of these methods to get a more accurate position.

The use of **mobile apps** is one of the main reasons of the increase in the amount of location data collected, by a wide range of actors. The **diversity of sensors** inside mobile devices (microphone, camera, infrared, GPS, Bluetooth, accelerometer, Wi-Fi, fingerprint sensor, etc.) makes it easy to collect and combine a wide range of data. This wide range of data can then be combined with other data sources to **infer private information** about the user.

All this data is accessed by apps through APIs provided by the **operating system** (OS), which also exploits the data for its own purposes.

6.2.4.2 European Framework For location data, besides the GDPR, the European legal framework also encompasses the **e-Privacy Directive**, which establishes rules to ensure privacy and personal data protection in the electronic communications sector, making the framework more complex.

The Article 2(c) of the e-Privacy Directive defines **location data** as:

> [A]ny data processed in an electronic communications network or by an electronic communications service, indicating the **geographic position** of the terminal equipment of a user of a publicly available electronic communications service
>
> *e-Privacy Directive, Article 2(c)*

The Recital 14 the specifies that such data:

> [M]ay refer to the *latitude, longitude and altitude* to the user's terminal equipment, to the *direction of travel*, to the *level of accuracy* of the location information, to the *identification* of the network cell in which the terminal equipment is located at a certain point in time and to the *time* the location information was recorded
>
> *e-Privacy Directive, Recital 14*

In the context of processing location data, the **data controller** — who determines the purposes and means of the processing — could be:

- The **OS developer** if the data is collected by the OS
- The **provider of the app** which processes the data (whether the app is installed on the device or accessed through a web browser)
- The data controllers of the **geolocation infrastructure**
- Any **other party** that processes the data

The Article 9 of the GDPR also establishes **special categories of personal data** which are particularly sensitive, such as racial or ethnic origin, political opinions, religious or philosophical beliefs, genetic biometric and health data or data concerning a natural person's sex life or sexual orientation. Location data may help infer such data, making it particularly sensitive.

6.2.4.3 Some Selected Cases

6.2.4.3.1 Location of Employees According to WP29, the use of **geolocation of employees** can find legal basis in the **legitimate interest** of the employer, who is the data controller. However, the employer must be able to demonstrate the **necessity** of the processing and the **balance of the interests** of the employer and the employees. The employer must also inform the employees about the processing of their location data. In it Opinion 8/2001, WP29 states that consent can hardly be a legal basis for the processing of location data when one of these situations arises:

- There is a real or potential prejudice to the worker in case of refusal
- The consent is a condition of employment

In theses situations, due to the dependency of the employee, the consent is not freely given and is therefore not valid.

In Italy, remote control of employees must respect employee **freedom** and **dignity** and **avoids excessive, prolonged, and indiscriminate surveillance** (e.g., allowing employees to turn off trackers). It must be negotiated with union representatives first and can only be used for:

- Organisational and production needs
- Workplace safety
- Protection of company assets

In any case, the tracking must comply with the GDPR, ensuring transparency, proportionality and privacy by design.

In France, the CNIL stated that the use of geolocation of employees is limited to:

- Control services related to the vehicle usage
- Ensure the security of employees and goods
- Check working hours

It is considered an **intrusive measure** that requires a prior **Data Protection Impact Assessment** (DPIA), and is forbidden in other cases, such as speed limits of vehicles or collection of data outside working hours.

6.2.4.3.2 Smart Vehicles Smart vehicles are equipped with a wide range of sensors and communication systems, which can collect a wide range of data, including location data. This location data is particularly sensitive, as it can reveal the **habits and preferences** of the driver. The **data controller** — which can be the vehicle producer, the equipment manufacturers or the service providers — shall **make the data subject aware** of how the data is processed, i.e. the frequency of collection, the possibility to shut down the tracking system and the third parties that can access the data.

The collection of location data shall be **proportionate** to the purposes by modulating the *frequency* and the *precision*. The purpose also influences the length of the data retention (data minimisation principle). For security reasons, personal data should also be **processed internally** as far as possible, and only sent to third parties when absolutely necessary.

6.2.4.3.3 Contact Tracing The COVID-19 pandemic has led to the development of **Digital Contact Tracing** (DCT) apps. These apps use tracking technologies to monitor the simultaneous presence of individuals in the same place. There are two main ways to implement DCT:

- Using **proximity data**, usually with *Bluetooth Low Energy* (BLE) beacons. The absolute position is unknown and data is stored locally on the device unless a user is tested positive. This approach was used by Trace Together in Singapore.
- Using **location data**, usually with GNSS. The absolute position is known, and the data is stored on a central server. This approach was used by WeChat and Alipay in China.

In Europe, the European Commission and the European Data Protection Board (EDPB) have expressed a preference for BLE for privacy reasons. The EDPB also gave criteria for the adoption of DCT apps:

- **Voluntary** use
- **DPIA** before development
- Predilection for **proximity** data
- Disclosure of information on who the infected has been in close contact with
- Data **minimisation** and data **protection** by design
- **Encrypted identifiers** generated by BLE
- **Anonymity** of third users involved

In the end, the revealing nature of location data makes **proportionality** and **transparency** essential. The main goal should be to find a balance between the **right to privacy** and the **right to health**, ensuring that the processing of location data is **necessary** and **proportionate** to the purpose.