Geodetic Reference Systems

Martin Vermeer

29th November 2010

Contents

1 Celestial and terrestrial reference frames

Every twenty-four hours, the Earth rotates around its axis relative to the heavens at what is very nearly a constant rate, about what it very nearly a fixed axis. The direction of this rotation axis will serve as the z axis of both the celestial and the terrestrial frame. In order to completely define the orientation of our reference frame, we then need to *conventionally fix two longitudes*:

1. On the celestial sphere: we take for this the *vernal equinox*, where the Sun crosses the equator S-N
2. On the Earth: the International Meridian Conference in Washington DC, 1884, chose *Greenwich* as the prime meridian.

   A *bonus* of this choice, which was realized after the conference, was that at the same time was defined a single, unified *global time system*, comprising 15° broad hourly time zones, so – especially in the United States, which was expanding Westward over many time zones – the trains would run on time.

See figure 1.1.

Red denotes an ECEF (Earth-Centred, Earth-Fixed) reference frame, which co-rotates with the solid Earth, so the x axis always lies in the plane of the Greenwich meridian. This is also called a CT (Conventional Terrestrial) System. Locations on the Earth's surface are (almost) constant in this kind of frame, and can be published, e.g., on maps. However, moving vehicles, ocean water and atmospheric air masses will sense "pseudo-forces" (like the Coriolis force) due to the non-uniform motion of this reference frame

blue denotes a (quasi-)inertial system, which does not undergo any (rapid) rotations relative to the fixed stars. Also called a celestial reference frame, as the co-ordinates of the fixed stars are nearly constant in it and may be published. Also the equations of motion of, e.g., satellites or gyroscopes apply strictly, without pseudo-forces induced by non-uniform reference system motion.

A *Conventional Terrestrial System* (CTS) is defined as follows:

- the origin of the frame coincides with the centre of mass of the Earth

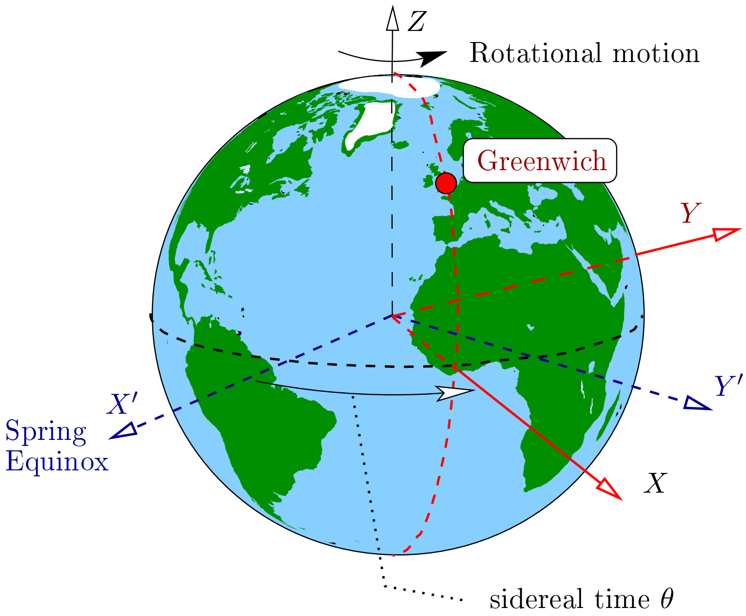


Figure 1.1: Geocentric reference frames

- the Z axis is directed along the rotation axis of the Earth, more precisely the Conventional International Origin (CIO) , i.e., the average direction of the axis over the time span 1900-1905
- the XZ-plane is parallel to the zero meridian as defined by "Greenwich", more precisely by: earlier the BIH (Bureau International de l'Heure, International Time Bureau), today the IERS (International Earth Rotation and Reference Systems Service), based on their precise monitoring of the Earth rotation.

In figure 1.2 we see the Earth orbit or *ecliptic*, the Earth axis tilt relative to the ecliptic plane, and how this tilt causes the most impressive climating variation observable to human beings: the four seasons.

The orientation of the Earth's axis undergoes slow changes. Relative to the stars, i.e., in inertial space, this motion consists of *precession* and *nutation*. It is caused by the gravitational torque exerted by the Sun and the Moon, which are either in (Sun) or close to (Moon) the ecliptic plane. See figure 1.3.

If we study the motion of the Earth's axis, and Earth rotation in general, relative to a reference frame connected to the solid Earth itself, we find different quantitities:

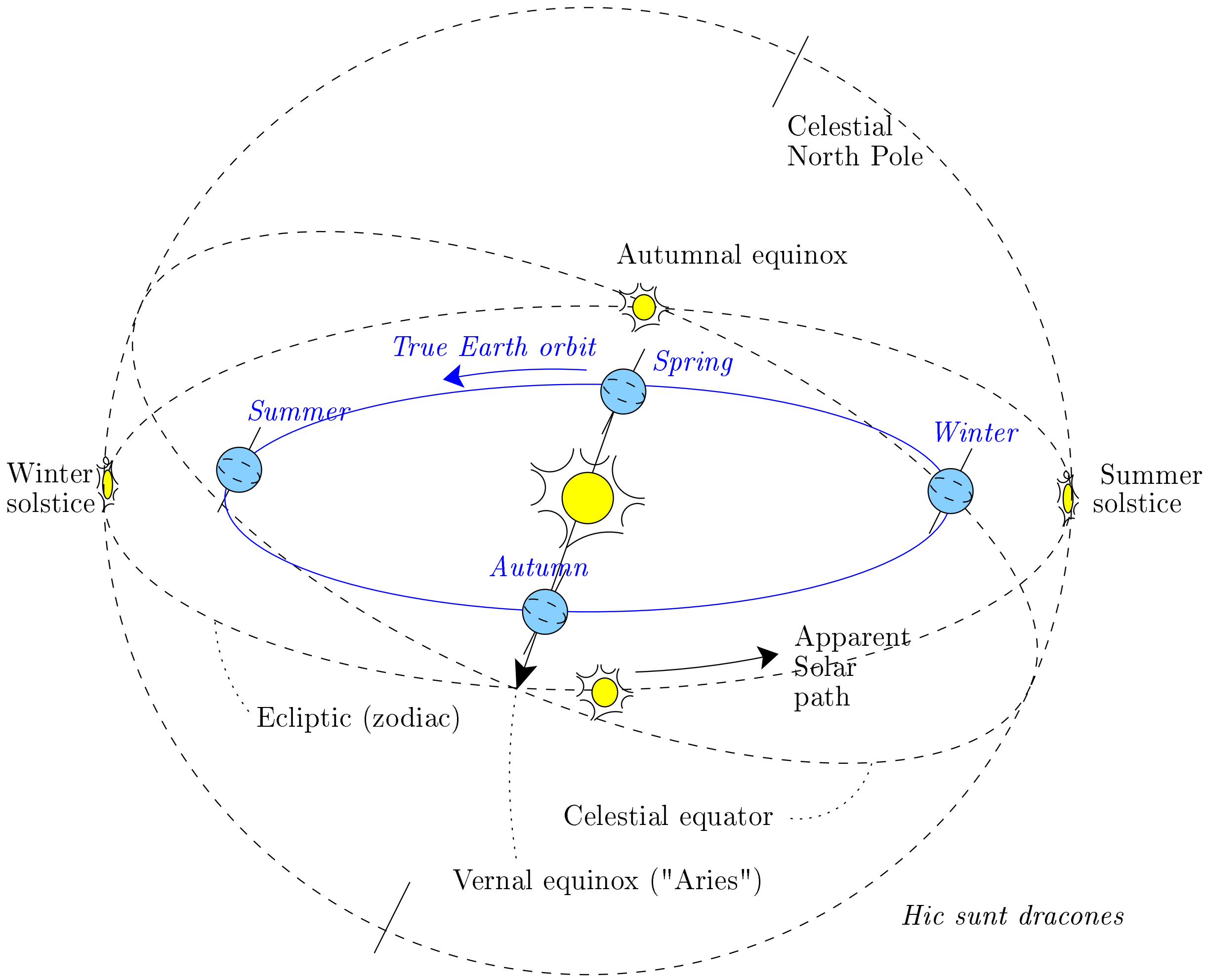


Figure 1.2: Geometry of the Earth's orbit and rotation axis. The seasons indicated are boreal

- Polar motion: this consists of an annual (forced) component and a 14-months component called the *Chandler wobble*.
- Length of Day.

Together these are called *Earth Orientation Parameters* (EOP). They are nowadays monitored routinely, and available after the fact from the International Earth Rotation Service over the Internet. The variation of these parameters is geophysically well understood, e.g., for the Chandler wobble it is mainly the pressure of the Earth's oceans and atmosphere that is responsible ([Gro00]).

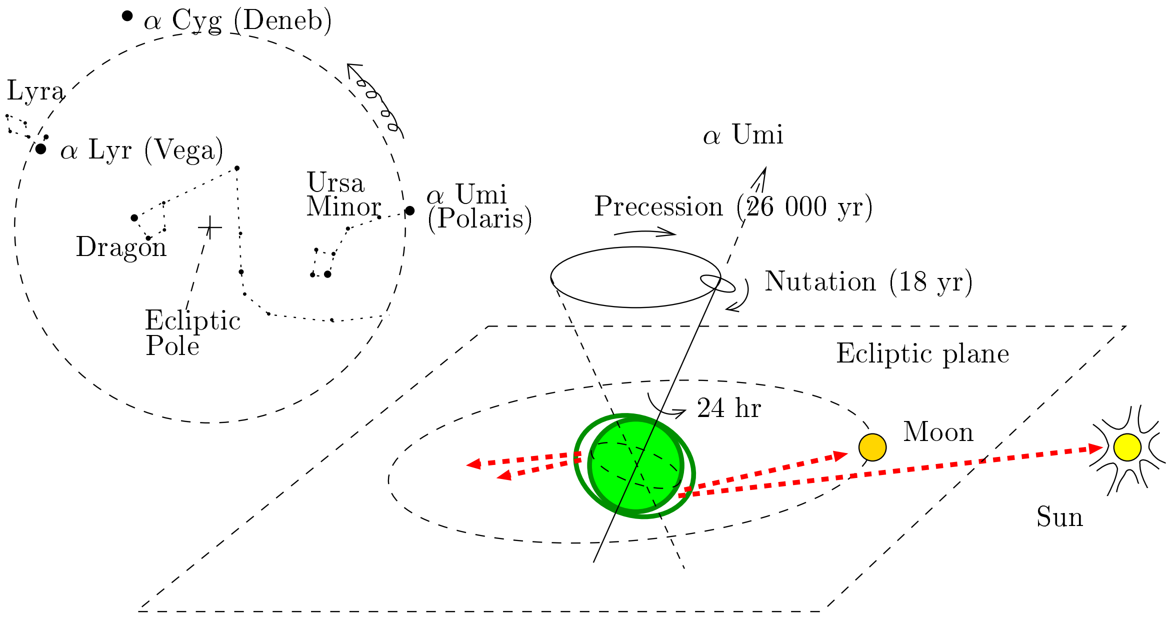


Figure 1.3: Precession, nutation and the torques from Sun and Moon

1.1 Polar motion

The direction of the Earth's rotation axis is slightly varying over time. This *polar motion* has two components called x_P and y_P, the offset of the instantaneous pole from the CIO pole in the direction of Greenwich, and perpendicular to it in the West direction, respectively. The transformation between the instantaneous and conventional terrestrial references is done as follows:

$$\mathbf{x}_{IT} = R_Y\left(x_p\right) R_X\left(y_p\right) \mathbf{x}_{CT}.$$

Here, note that the matrix R_Y denotes a rotation by an amount x_P *about the Y axis*, i.e., the Y axis stays fixed, while the X and Z co-ordinates change. Similarly, the matrix R_X denotes a rotation y_p about the X axis, which changes only the Y and Z co-ordinates. The matrices are

$$R_Y\left(x_p\right) = \begin{bmatrix} \cos x_P & 0 & -\sin x_P \\ 0 & 1 & 0 \\ \sin x_P & 0 & \cos x_P \end{bmatrix} \quad \text{and} \quad R_X\left(y_p\right) \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos y_P & \sin y_P \\ 0 & -\sin y_P & \cos y_P \end{bmatrix}.$$

Because the angles x_p and y_p are very small, order of magnitude second of arc, we may approximate $\sin x_p \approx x_p$ and $\cos x_p \approx 1$ (same for y_p), as well as $x_P y_P \approx 0$, obtaining

$$R_Y\left(x_p\right) R_X\left(y_p\right) \approx \begin{bmatrix} 1 & 0 & -x_p \\ 0 & 1 & 0 \\ x_p & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & y_p \\ 0 & -y_p & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & -x_p \\ 0 & 1 & y_p \\ x_p & -y_p & 1 \end{bmatrix}.$$

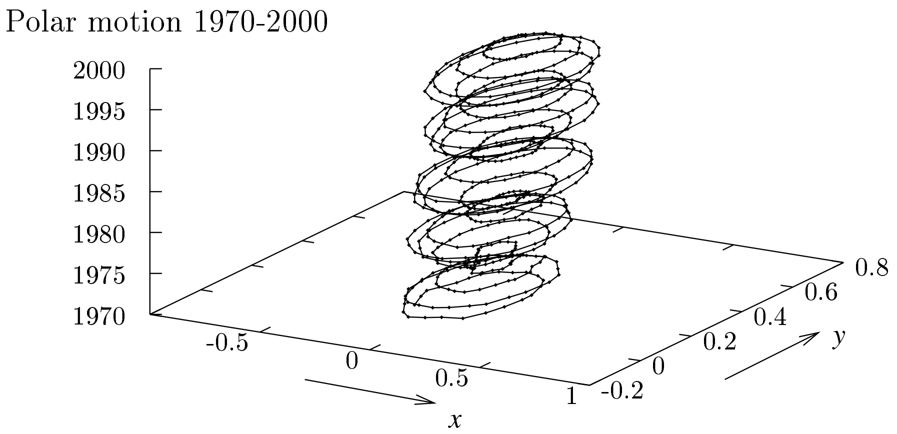


Figure 1.4: Polar motion

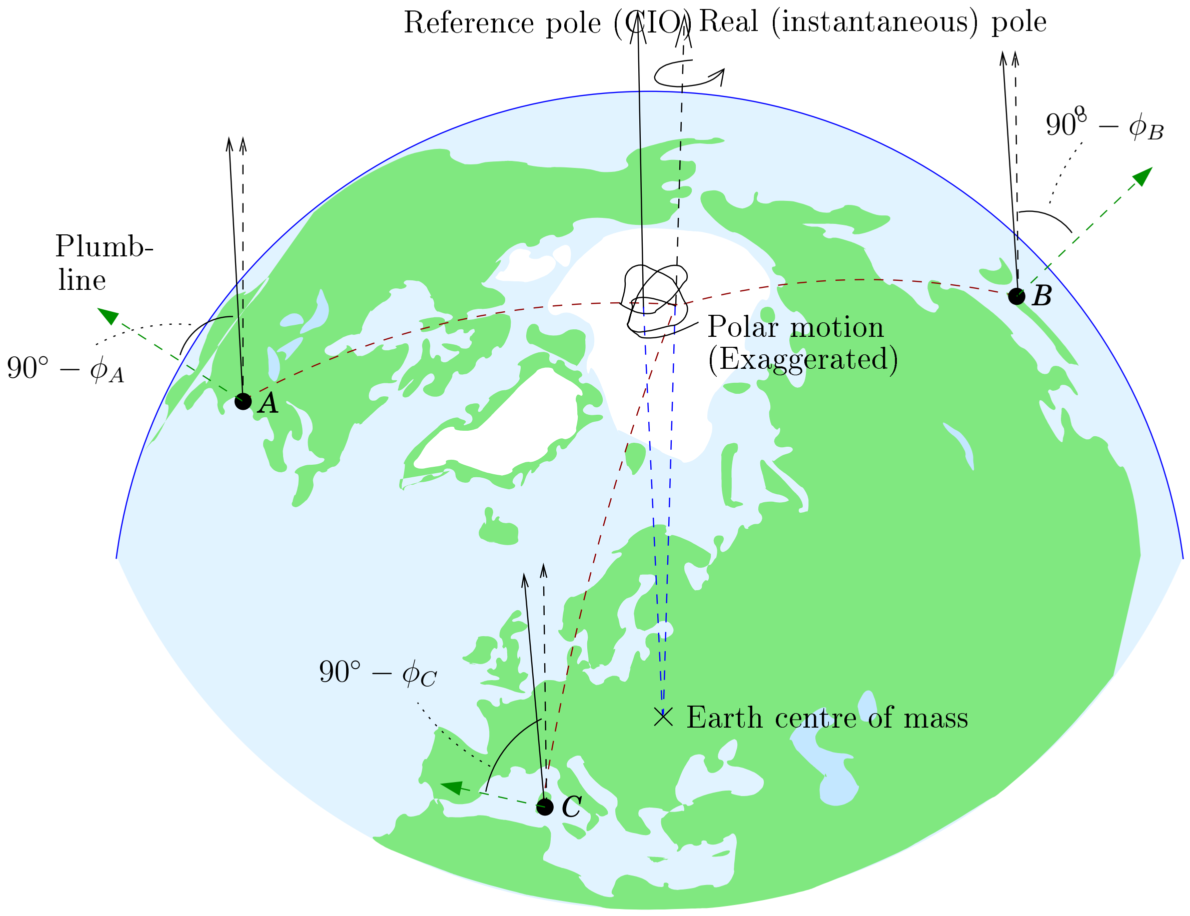


Figure 1.5: How to monitor polar motion using latitude observatories

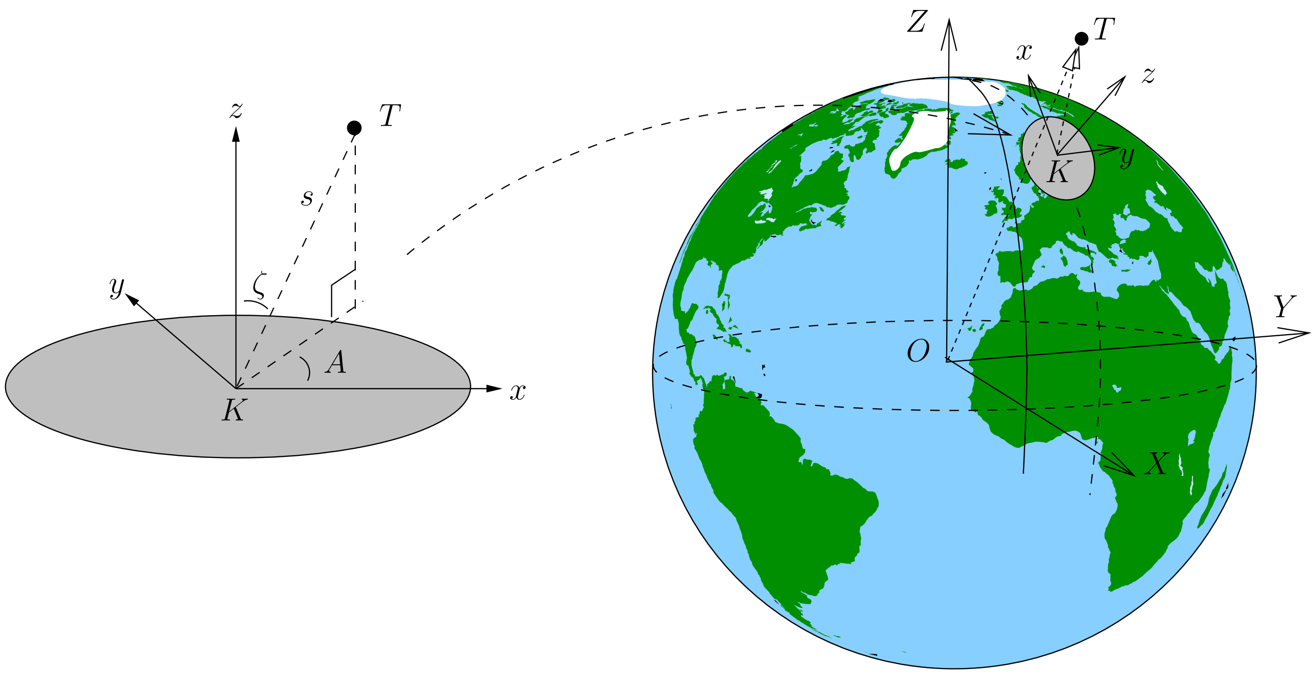


Figure 1.6: Local astronomical co-ordinates

1.2 Topocentric co-ordinates

These also called *local astronomical co-ordinates*. The z-axis points upward along local plumbline (vertical); the xy-plane lies in the local horizon.

The X-axis points due North. A variant of this are *instrumental co-ordinates*, where (for general theodolites) the X-axis points in an arbitrary direction defined by the horizontal circle's orientation.

If we have the observation of a celestial object in the form of azimut A and elevation ζ, we can obtain rectangular local astronomical co-ordinates as follows:

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix}_{LA} = \begin{bmatrix} \cos A \cos \zeta \\ \sin A \cos \zeta \\ \sin \zeta \end{bmatrix}.$$

The reverse transformation is done as follows (note that the rectangular vector is a *unit vector*, also called a direction vector or "direction cosines"):

$$\begin{aligned} \zeta &= \arcsin z, \\ A &= 2 \arctan \frac{y}{x + \sqrt{x^2 + y^2}}. \end{aligned}$$

The latter formula is known as the *half-angle formula* and avoids the problem of finding the correct quadrant for A. The result is in the interval $(-180°, 180°]$ and negative values may be incremented by 360° to make them positive.

The transformation between local astronomical and conventional terrestrial requires knowledge of the *local astronomical position*, i.e., the direction of the local plumbline which defines the vertical of the LA system: Λ, the local astronomical longitude, and Φ, the local astronomical latitude. Then:

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix}_{CT} = \begin{bmatrix} -\sin\Phi\cos\Lambda & -\sin\Lambda & \cos\Phi\cos\Lambda \\ -\sin\Phi\sin\Lambda & \cos\Lambda & \cos\Phi\sin\Lambda \\ \cos\Phi & 0 & \sin\Phi \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix}_{LA},$$

or in symbolic form

$$\mathbf{x}_{CT} = R\mathbf{x}_{LA},$$

where

$$\begin{aligned} R &= R_Z(180° - \Lambda)\, R_Y(90° - \Phi)\, M_Y = \\ &= \begin{bmatrix} -\cos\Lambda & \sin\Lambda & 0 \\ -\sin\Lambda & -\cos\Lambda & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \sin\Phi & 0 & -\cos\Phi \\ 0 & 1 & 0 \\ \cos\Phi & 0 & \sin\Phi \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 1 \end{bmatrix}. \end{aligned}$$

The matrix M_Y simply mirrors the Y axis. The correctness of the R_Y and R_Z rotation matrices is easiest to verify by making a two-dimensional plot on paper.

Note that this transformation formula is only approximate, due to the fact that the conventional terrestial system is not the same as the instantaneous terrestrial reference system (IT), in which the astronomical observations of (Φ, Λ) are being made. So, in the above, instead of CT we actually obtain IT, and further corrections due to polar motion and Earth rotation variations.

1.3 Sidereal time

See the pretty figure 1.7. We have the following quantities:

- GAST = Greenwich Apparent Sidereal Time
- LMST = Local Mean Sidereal Time;
- the equinox varies irregularly with time due to precession and nutation. That's why we distinguish Mean and Apparent. The difference is called the *equation of equinoxes.*
- h is the hour angle
- h_{Gr} is the Greenwich hour angle
- α is the right ascension (of a celestial object)

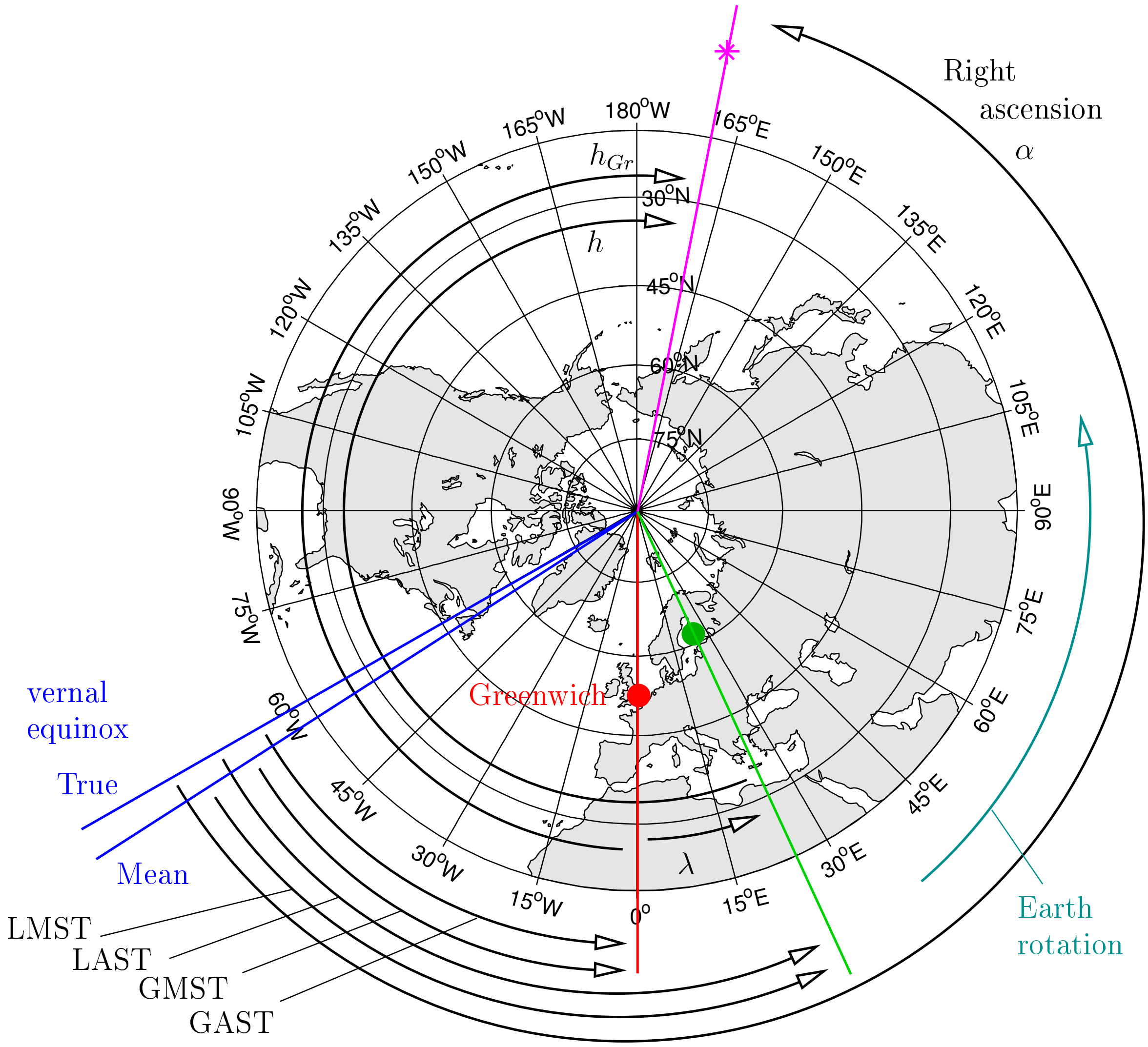


Figure 1.7: Sidereal time

- λ is the longitude (of a terrestrial object).

$$
\begin{aligned}
h &= LAST - \alpha, \\
h_{Gr} &= GAST - \alpha, \\
LAST &= GAST + \lambda, \\
LMST &= GMST + \lambda.
\end{aligned}
$$

More detailed discussions in [Tor01] Section 2.4.

1.4 Quasi-inertial geocentric system

The quasi-inertial, also celestial, or real astronomical (RA) reference frame, drawn in blue in figure 1.1, is a *geocentric* system, like the conventional terrestrial system. It is however celestial in nature and the positions of stars are approximately constant in it. It is defined as follows:

- the origin of the frame coincides with the centre of mass of the Earth
- the Z axis is directed along the *instantaneous* rotation axis of the Earth, and so
- the XY plane is parallel to the *instantaneous* equatorial plane of the Earth
- The X axis points to the *instantaneous* vernal equinox point, the intersection of celestial equator and ecliptic.

Like for the conventional terrestial reference system there was a corresponding "instantaneous" system connected to the instantaneous rotation axis of the Earth, there exists a *conventional celestial* system too in which the places of stars on the celestial sphere are given in stellar catalogues and on star charts. These places of stars are given in some "precession epoch" or "equinox", e.g., 2000.0. To obtain the *apparent* place of a star, i.e., the place in the sky that it "appears" to be at the moment of observation, a reduction computation must be executed accounting for precession and nutation, among other things.

The spherical co-ordinates of this system are right ascension α and declination δ. Considering the celestial sphere again a "direction sphere" only, we may assume all direction vectors to be unit vectors and obtain for rectangular RA co-ordinates:

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix}_{RA} = \begin{bmatrix} \cos\alpha\cos\delta \\ \sin\alpha\cos\delta \\ \sin\delta \end{bmatrix}.$$

Note that, while declination is given in degrees ($^\circ$), right ascension is for traditional convenience reasons given in time units, i.e., hours, minutes and seconds. Going from rectangular to spherical again requires the following formulae:

$$\begin{aligned} \delta &= \arcsin Z \\ \alpha &= 2\arctan\frac{Y}{X+\sqrt{X^2+Y^2}}. \end{aligned}$$

Again, negative values for α can be made positive by adding $24\,^{\mathrm{h}}$.

We can transform between RA and IT (instantaneous terrestrial) co-ordinates as follows:

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix}_{RA} = R_Z\left(-\mathrm{GAST}\right)\begin{bmatrix} X \\ Y \\ Z \end{bmatrix}_{IT} = \begin{bmatrix} \cos\left(\mathrm{GAST}\right) & -\sin\left(\mathrm{GAST}\right) & 0 \\ \sin\left(\mathrm{GAST}\right) & \cos\left(\mathrm{GAST}\right) & 0 \\ 0 & 0 & 1 \end{bmatrix}\begin{bmatrix} X \\ Y \\ Z \end{bmatrix}_{IT}.$$

1.5 Transformations between systems

See the following diagram, which depicts only the *rotations* between the various systems:

$$\begin{array}{ccccccc} & \Phi,\Lambda & & x_p, y_p & & \text{GAST} & \\ \text{Local} & \Longleftrightarrow & \text{Conventional} & \Longleftrightarrow & \text{Instantaneous} & \Longleftrightarrow & \text{Real} \\ \text{Astronomical} & & \text{Terrestrial} & & \text{Terrestrial} & & \text{Astronomical} \end{array}$$

Here, Φ, Λ are local astronomical co-ordinates (direction of the plumbline), while x_p, y_p are the co-ordinates of the pole in the CIO system. GAST is Greenwich Apparent Sidereal Time.

1.6 Compound co-ordinate "systems"

We may represent horizontal location in two dimensions by either (φ, λ) or (x, y). Similarly we may represent height in one dimension by, e.g., orthometric height H.

- In two dimensions we use rectangular *map projection co-ordinates* (x, y), which have a 1-to-1 correspondence with ellipsoidal latitude and longitude (φ, λ), typically computed on a suitably chosen reference ellipsoid.
- In one dimension, we have the height system. See below.

Then we have compound systems, like (φ, λ, H) or (x, y, H). These are often used, but one should remember that they are artificial, being "synthetic". Always remember that

- (x, y) and (φ, λ) are connected by a *map projection*
- H is connected to three-dimensional location (or ellipsoidal height h) by a *geoid model.*

2 The reference ellipsoid

2.1 Geodetic co-ordinates

Geodetic co-ordinates are always given relative to a chosen *reference ellipsoid.* Besides choosing the parameters of this ellipsoid, one has to choose the location of its centre – generally not geocentric, or not precisely so – and the orientation of its axes – generally, the Z-axis will not be parallel, or not precisly so, with the Earth's rotation axis.

The refence ellipsoids used in geodesy are *bi-axial* ellipsoids, with the shorter axis pointing approximately along the Earth's rotation axis. The two, equal, longer axes then point approximately within the equatorial plane.

More rarely, tri-axial reference ellipsoids have been proposed, but not found practical. For other celestial bodies – the Moon, Mars – this may be different.

We can define *rectangular geodetic coordinates* $\begin{bmatrix} X_G & Y_G & Z_G \end{bmatrix}^T$ as having their origin at the centre of the reference ellipsoid chosen – not necessarily, or not precisely, the geocentre or centre of mass of the Earth, though close to it – and having the Z_G-axis aiming along the shorter axis of the ellipsoid. The direction of the X_G-axis is arbitrary in principle, but is chosen in practice to be close to the Greenwich meridian plane. The $X_G Y_G$ plane is the equatorial plane of the ellipsoid.

A geodetic co-ordinate system is meant to be as close to geocentric (the CT system) as possible, but using measurements on finite precision this is not possible exactly. The non-geocentricity of the origin may be called $\begin{bmatrix} \Delta X & \Delta Y & \Delta Z \end{bmatrix}^T$, and the non-alignedness of the axes, $\epsilon_x, \epsilon_y, \epsilon_z$. With these six parameters known, one can derive the *transformation formula* between the two systems, allong one to calculate co-ordinate differences between them. These six parameters together define a *geodetic datum.*

2.2 Ellipsoids and Geocentricity

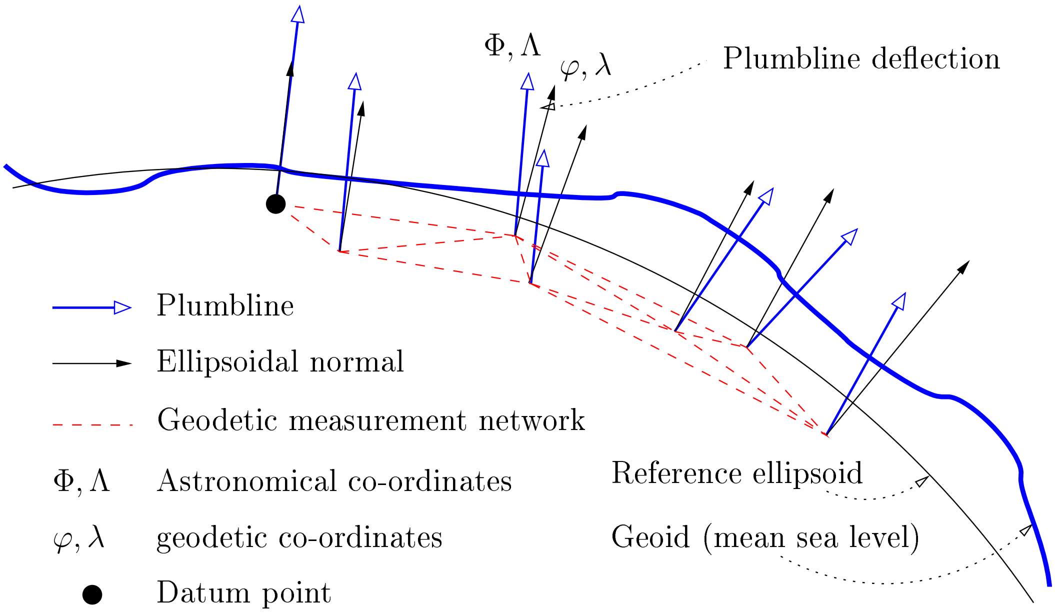
Old reference systems were created using traditional measurement technology tied to the Earth's surface and the local plumbline. The reference ellipsoid was only a computational device, with no claim of global representativeness or geocentricity.

- North direction was fixed with the aid of *Laplace azimuths*

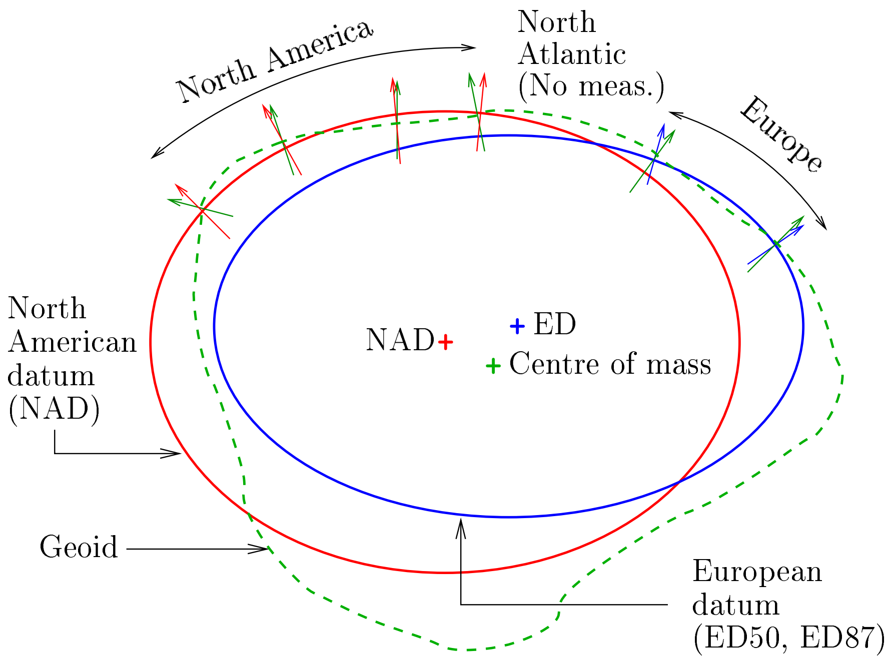
- Origin was fixed *in two dimensions* by choice of “datum point”, e.g., Potsdam, where the astronomical latitude and longitude equalled the (conventionally fixed) geodetic ones.
- Vertical level was fixed by setting “geoid undulation” to zero in datum point
- Thus, the network location and orientation in space was fixed ⇒*not geocentric*
- Non-geocentricity related to (physical) plumbline deflections in datum point: every second of arc produces geocentrically

$$\Delta R = 2\pi R \frac{1''}{360^\circ} \approx 4.8 \cdot 10^{-6} R \approx 31\,\mathrm{m}.$$

See figure:



European Datum 1950 (ED50)



2.3 Local geodetic co-ordinates

For this system, LG, the origin is in the location of observation, the z-axis points along the outward ellipsoidal normal (which differs from the local vertical!), the x-axis points to the geodetic North – i.e., the direction on the ellipsoidal surface that intersects with the ellipsoid's shorter axis Z_G – and the y-axis completes a left-handed system, i.e., points ellipsoidally due East.

The orientation of this system relative to the rectangular geodetic co-ordinates defined above, by two angles: φ and λ, the *geodetic latitude and longitude*. These differ from the astronomical latitude and longitude Φ and Λ forming the basis of the local astronomical (LA) system, by amounts called the *deflection of the vertical*. Otherwise, these two systems are similar.

2.4 Conversion of geodetic latitude and longitude to rectangular co-ordinates

We will often extend the two ellipsoidal or geodetic co-ordinates on the reference ellipsoid, geodetic latitude φ and geodetic longitude λ, with a third co-ordinate, the height h above the reference ellipsoid's surface. This is usually called "ellipsoidal height", not the perhaps more logical "geodetic height". As these three co-ordinates (h, φ, λ) completely specify the location of a point, a unique conversion to rectangular (X, Y, Z) co-ordinates is possible:

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \begin{bmatrix} (N(\varphi) + h) \cos\varphi \cos\lambda \\ (N(\varphi) + h) \cos\varphi \sin\lambda \\ \left(\left(b^2/a^2\right) N(\varphi) + h\right) \sin\varphi \end{bmatrix}.$$

Here, a and b are the ellipsoid's semi-major and minor axes (equatorial and polar radius) and N, or $N(\varphi)$, the normal or transversal (E-W direction) radius of curvature

$$N(\varphi) = \frac{a^2}{\sqrt{a^2 \cos^2\varphi + b^2 \sin^2\varphi}}.$$

2.5 Datum transformations

We derive approximate datum transformation formulas in *spherical approximation.* This is acceptable for datum transformations which are "small", i.e., they cause small changes in the co-ordinates between the two datums.

Geocentrically (ϕ is geocentric latitude):

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = (R + h) \begin{bmatrix} \cos\phi \cos\lambda \\ \cos\phi \sin\lambda \\ \sin\phi \end{bmatrix} + \begin{bmatrix} X_0 \\ Y_0 \\ Z_0 \end{bmatrix}.$$

Write this for two different systems (assuming same axes orientation):

$$\begin{aligned} \begin{bmatrix} X^{(1)} \\ Y^{(1)} \\ Z^{(1)} \end{bmatrix} &= \left(R + h^{(1)}\right) \begin{bmatrix} \cos\phi^{(1)} \cos\lambda^{(1)} \\ \cos\phi^{(1)} \sin\lambda^{(1)} \\ \sin\phi^{(1)} \end{bmatrix} + \begin{bmatrix} X_0^{(1)} \\ Y_0^{(1)} \\ Z_0^{(1)} \end{bmatrix}, \\ \begin{bmatrix} X^{(2)} \\ Y^{(2)} \\ Z^{(2)} \end{bmatrix} &= \left(R + h^{(2)}\right) \begin{bmatrix} \cos\phi^{(2)} \cos\lambda^{(2)} \\ \cos\phi^{(2)} \sin\lambda^{(2)} \\ \sin\phi^{(2)} \end{bmatrix} + \begin{bmatrix} X_0^{(2)} \\ Y_0^{(2)} \\ Z_0^{(2)} \end{bmatrix}. \end{aligned}$$

Demand the left hand sides to be the same:

$$\begin{bmatrix} \Delta X \\ \Delta Y \\ \Delta Z \end{bmatrix} = \left\{ \left(R + h^{(2)}\right) \begin{bmatrix} \cos\phi^{(2)} \cos\lambda^{(2)} \\ \cos\phi^{(2)} \sin\lambda^{(2)} \\ \sin\phi^{(2)} \end{bmatrix} \right\} -$$

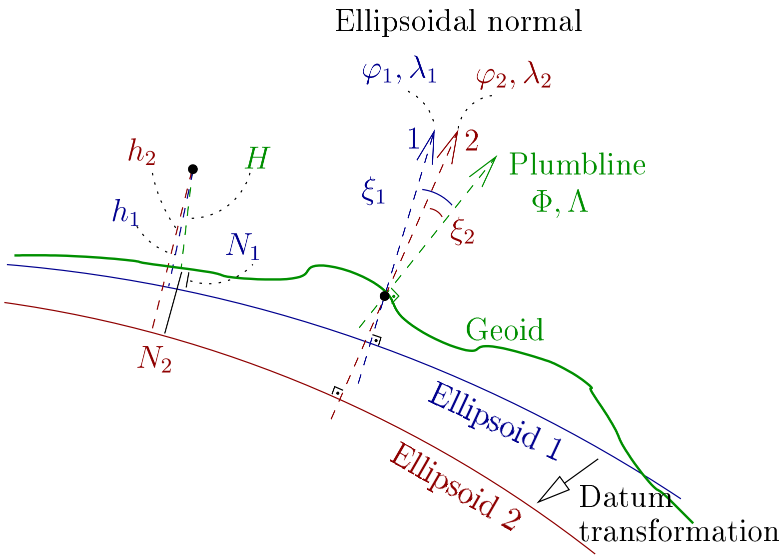


Figure 2.1: A datum transformation as a change of reference ellipsoid

$$
\begin{aligned}
& - \left\{ \left(R + h^{(1)} \right) \begin{bmatrix} \cos\phi^{(1)} \cos\lambda^{(1)} \\ \cos\phi^{(1)} \sin\lambda^{(1)} \\ \sin\phi^{(1)} \end{bmatrix} \right\} + \begin{bmatrix} \Delta X_0 \\ \Delta Y_0 \\ \Delta Z_0 \end{bmatrix} = \\
= & \ \Delta \left\{ (R + h) \begin{bmatrix} \cos\phi\cos\lambda \\ \cos\phi\sin\lambda \\ \sin\phi \end{bmatrix} \right\} + \begin{bmatrix} \Delta X_0 \\ \Delta Y_0 \\ \Delta Z_0 \end{bmatrix} = 0.
\end{aligned}
$$

Taking partial derivatives (linearization) yields:

$$
\begin{aligned}
\begin{bmatrix} \Delta X_0 \\ \Delta Y_0 \\ \Delta Z_0 \end{bmatrix} & = \begin{bmatrix} -\cos\phi\cos\lambda & +R\sin\phi\cos\lambda & +R\cos\phi\sin\lambda \\ -\cos\phi\sin\lambda & +R\sin\phi\sin\lambda & -R\cos\phi\cos\lambda \\ -\sin\phi & -R\cos\phi & 0 \end{bmatrix} \begin{bmatrix} \Delta h \\ \Delta\varphi \\ \Delta\lambda \end{bmatrix} = \\
& = \begin{bmatrix} -\cos\phi\cos\lambda & +R\sin\phi\cos\lambda & +R\cos\phi\sin\lambda \\ -\cos\phi\sin\lambda & +R\sin\phi\sin\lambda & -R\cos\phi\cos\lambda \\ -\sin\phi & -R\cos\phi & 0 \end{bmatrix} \begin{bmatrix} \Delta N \\ -\Delta\xi \\ -\frac{\Delta\eta}{\cos\phi} \end{bmatrix}.
\end{aligned}
$$

Here we see the shifts in *ellipsoidal* co-ordinates h, φ, λ (still in spherical approximation) and in datum quantities N, ξ, η, i.e., geoid undulation and the two components of deflection of the vertical.

2.6 Connection between geodetic and astronomical co-ordinates

We describe the connection between the local astronomical (LA) and local geodetic (LG) systems. These equation hearken back to the old marquis Pierre Simon de Laplace

$$\begin{aligned} \xi &= \Phi - \varphi, \\ \eta &= (\Lambda - \lambda)\cos\varphi, \\ \Delta A &= A - a = (\Lambda - \lambda)\sin\varphi, \end{aligned}$$

where ξ and η are the *deflections of the vertical* in the direction of the meridian (North-South) and the transversal direction (West-East) respectively. A and a again are the directions of a target in the LA and LG systems projected onto their XY planes, i.e., the *astronomical* and *geodetic azimuths*. It is the latter so-called Laplace equation thatg enables a local geodetic network to be absolutely oriented with the help of astronomical observations.

See the following diagram:

$$\begin{array}{ccccc} & & \Delta X, \Delta Y, \Delta Z, \epsilon_x, \epsilon_y, \epsilon_z & & \\ & \text{Conventional Terrestrial} & \Longleftrightarrow & \text{Geodetic} & \\ \Phi, \Lambda & \Updownarrow & & \Updownarrow & \varphi, \lambda \\ & \text{Local Astronomical} & \Longleftrightarrow & \text{Local Geodetic} & \\ & & \Delta A, \xi, \eta & & \end{array}$$

3 Reference systems and realizations

3.1 Old and new reference systems; ED50 vastaan WGS84/GRS80

In Finland like in many European countries, the traditional reference system is non-geocentric and based on an old reference ellipsoid, the International Ellipsoid computed by John Fillmore Hayford, and adopted by the International Union of Geodesy and Geophysics (IUGG) in 1924. European Datum 1950 (ED50) was created in 1950 by unifying the geodetic networks of the countries of Western Europe, and was computed on the Hayford ellipsoid.

The newer systems, both World Geodetic System (WGS84) and Geodetic Reference System 1980 (GRS80) are designed to be geocentric.

So, the differences can be summarized as:

1. Reference ellipsoid used: International Ellipsoid (Hayford) 1924 vs. GRS80
2. Realized by terrestrial measurements vs. based on satellite (and space geodetic) data
3. Non-geocentric (100 m level) vs. geocentric (cm level).

The figure of a *reference ellipsoid* is unambiguously fixed by two quantities: the semi-major exis or equatorial radius a, and the flattening f.

- International ellipsoid 1924: $a = 6378388\,\mathrm{m}$, $f = 1 : 297$
- GRS80: $a = 6378137\,\mathrm{m}$, $f = 1 : 298.257222101$

The reference ellipsoid of GPS's WGS84 system is in principle the same as GRS80, but due to poor numerics it ended up with

- $f = 1 : 298.257223563$. The net result is that the semi-minor axis (polar radius) of WGS84 is longer by 0.1 mm[1] compared to GRS80.

To complicate matters, as the basis of the ITRS family of co-ordinate systems, and their realizations ITRF, was chosen *the GRS80 reference ellipsoid*.

We tabulate the defining and important derived quantities:

[1]Computation:

$$\Delta b = -\frac{\Delta f}{f}(a-b) = \frac{\Delta\left({}^1\!/\!_f\right)}{\left({}^1\!/\!_f\right)}(a-b) = \frac{0.000001462}{298.25722}\cdot(21384\,\mathrm{m}) = 0.000105\,\mathrm{m}.$$

Quantity	Symbol	Value	Units
Equatorial radius	a	6378137	m
Gravitational mass	GM	(WGS84) 3986004.418	$10^8\ {}^{\mathrm{km}^3}/_{\mathrm{s}^2}$
		(GRS80) 3986005.	
Dynamic flattening	J_2	1082.63×10^{-6}	
Rotation rate	ω	7.292115×10^{-5}	s^{-1}
Inverse flattening	$1/f$	(WGS84) 298.257223563	
		(GRS80) 298.257222101	
Polar radius	b	(WGS84) 6356752.314245	m
		(GRS80) 6356752.314140	

3.2 WGS84 and ITRS

Both WGS84 and the International Terrestrial Reference System (ITRS) are *realized* by computing co-ordinates for polyhedra of points (stations) on the Earth's surface. The properties of these systems are:

- *Geocentric*, i.e., the co-ordinate origin and centre of reference ellipsoid is the Earth's centre of mass (and the Earth's mass includes oceans and atmosphere, but not the Moon!). This is realized by using observations to satellites, whose equations of motion are implicitly geocentric.
- The *scale* derives from the SI system. This is realized by using range measurements by propagation of electromagnetic waves. The velocity of these waves in vacuum is conventionally fixed to $299\,792\,458\,\mathrm{m\,s^{-1}}$. Thus, range measurement becomes time measurement by atomic clock, which is very precise.
- *Orientation*: originally the Conventional International Origin (CIO) of the Earth axis, i.e., the mean orientation over the years 1900-1905, and the direction of the Greenwich Meridian, i.e., the plumbline of the Greenwich transit circle. Currently, as the orientation is maintained by the International Earth Rotation and Reference Systems Service (IERS) using VLBI and GPS, this is no longer the formal definition; but continuity is maintained.
- The current *definition* uses the BIH (Bureau International de l'Heure) 1984 definition of the conventional pole, and their 1984 definition of the zero meridian plane. Together, X, Y and Z form a right-handed system.

3.3 Co-ordinate system realizations

Internationally, somewhat varying terminology is in use concerning the realization of co-ordinate systems or *datums*.

- ISO: Co-ordinate reference system / co-ordinate system

- IERS: Reference system / reference frame
- Finnish: koordinaattijärjestelmä / koordinaatisto ([Ano08])

The *latter* of the names is used to describe a system that was implemented in the terrain, using actual measurements, producing co-ordinate values for the stations concerned; i.e., a *realization*. Then also, a *datum* was defined, with one or more *datum points* being kept fixed at their conventional values.

The *former* refers to a more abstract definition of a co-ordinate system, involving the choice of reference ellipsoid, origin (Earth center of mass, e.g.) and axes orientation.

3.4 Realization of WGS84

Because "WGS84" is often referred to as the system in which satellite positioning derived co-ordinates are obtained, we shall elaborate a little on how this system has been actually realized over time. Our source is [KR06]. The first version of WGS84 was released in 1987 by the US Defense Mapping Agency, currently the NGA (National Geospatial-Intelligence Agency). After that, it was updated in 1994 (G730), 1996 (G873) and 2001 (G1150).

As you will see, there are a number of problems even with the latest realization of WGS84. For this reason it is better to consider WGS84 as an approximation at best, of the reference frames of the ITRF/ETRF variety. The precision of this approximation is clearly sub-metre, so using WGS84 for metre precision level applications should be OK. See the following note: `ftp://itrf.ensg.ign.fr/pub/itrf/WGS84.TXT`.

If you want more confusion, read [Ste08].

3.5 Realizations of ITRS/ETRS systems

All these systems are the responsibility of the international geodetic community, specifically the IERS (International Earth Rotation and Reference Systems Service). "I" stands for International, "E" for European. The "S" stands for "system", meaning the principles for creating a reference frame before actual realization. With every ITRS corresponds a number of ITRF's ("Frames"), which are *realizations*, i.e., co-ordinate solutions for networks of ground stations computed from sets of actual measurements. Same for ETRS/ETRF, which are the corresponding things for the European area, where the effect of the slow motion of the rigid part of the Eurasian tectonic plate has been corrected out in order to obtain approximately constant co-ordinates.

Data used for realizing ITRF/ETRF frames: mostly GPS, but also Very long Baseline Interferometry (VLBI) providing a strong orientation; satellite and lunar laser ranging (SLR, LLR) contributing to the right scale, and the French DORIS satellite system. Nowadays also GLONASS is used.

Currently the following realizations exist for ITRS: ITRF88, 89, 90, 91, 92, 93, 94; 96, 97; ITRF2000, ITRF2005 and ITRF2008.

The *definition* of an ITRFyy is as follows [McC96]:

- The mean rotation of the Earth's crust in the reference frame will vanish globally (cf. for ETRF: on the Eurasian plate). Obviously then, co-ordinates of points on the Earth's surface will slowly change due to the motion of the plate that the point is on. Unfortunately at the current level of geodetic precision, it is not possible to define a global co-ordinate frame in which points are fixed.
- The Z-axis corresponds to the IERS Reference pole (IRP) which corresponds to the BIH Conventional terrestrial Pole of 1984, with an uncertainty of 0.005"
- The X-axis, or IERS Reference Meridian, similarly corresponds to the BIH zero meridian of 1984, with the same uncertainty.

Finally, note that the Precise Ephemeris which are computed by IGS (the International GPS Geodynamics Service) and distributed over the Internet, are referred to the current (newest) ITRF, and are computed using these co-ordinates for the tracking stations used.

3.6 The three-dimensional Helmert transformation

The form of the transformation, in the case of small rotation angles, is

$$\begin{bmatrix} X^{(2)} \\ Y^{(2)} \\ Z^{(2)} \end{bmatrix} = (1+m) \begin{bmatrix} 1 & e_z & -e_y \\ -e_z & 1 & e_x \\ e_y & -e_x & 1 \end{bmatrix} \cdot \begin{bmatrix} X^{(1)} \\ Y^{(1)} \\ Z^{(1)} \end{bmatrix} + \begin{bmatrix} t_x \\ t_y \\ t_z \end{bmatrix}, \tag{3.1}$$

where $\begin{bmatrix} t_x & t_y & t_z \end{bmatrix}^T$ is the translation vector of the origin, m is the scale factor correction from unity, and (e_x, e_y, e_z) are the (small) rotation angles about the respective axes. Together we thus have *seven* parameters. The superscripts (1) and (2) refer to the old and new systems, respectively.

Eq. (3.1) can be re-written and linearized as follows, using $me_x = me_y = me_z = 0$, and replacing the vector $\begin{bmatrix} X^{(1)} & Y^{(1)} & Z^{(1)} \end{bmatrix}^T$ by approximate values $\begin{bmatrix} X^0 & Y^0 & Z^0 \end{bmatrix}^T$. This is allowed as m and the e angles are all assumed small.

$$\begin{aligned} \begin{bmatrix} X^{(2)} - X^{(1)} \\ Y^{(2)} - Y^{(1)} \\ Z^{(2)} - Z^{(1)} \end{bmatrix} &\approx \begin{bmatrix} m & e_z & -e_y \\ -e_z & m & e_x \\ e_y & -e_x & m \end{bmatrix} \begin{bmatrix} X^{(1)} \\ Y^{(1)} \\ Z^{(1)} \end{bmatrix} + \begin{bmatrix} t_x \\ t_y \\ t_z \end{bmatrix} \approx \\ &\approx \begin{bmatrix} m & e_z & -e_y \\ -e_z & m & e_x \\ e_y & -e_x & m \end{bmatrix} \begin{bmatrix} X^0 \\ Y^0 \\ Z^0 \end{bmatrix} + \begin{bmatrix} t_x \\ t_y \\ t_z \end{bmatrix}. \end{aligned}$$

An elaborate rearranging yields

$$\begin{bmatrix} X_i^{(2)} - X_i^{(1)} \\ Y_i^{(2)} - Y_i^{(1)} \\ Z_i^{(2)} - Z_i^{(1)} \end{bmatrix} = \left[\begin{array}{c|ccc|ccc} X_i^0 & 0 & -Z_i^0 & +Y_i^0 & 1 & 0 & 0 \\ Y_i^0 & +Z_i^0 & 0 & -X_i^0 & 0 & 1 & 0 \\ Z_i^0 & -Y_i^0 & +X_i^0 & 0 & 0 & 0 & 1 \end{array}\right] \left[\begin{array}{c} m \\ \hline e_x \\ e_y \\ e_z \\ \hline t_x \\ t_y \\ t_z \end{array}\right]. \tag{3.2}$$

Here we have added for generality a point index i, $i = 1, \ldots, n$. The number of points is then n, the number of "observations" (available co-ordinate differences) is $3n$. The full set of these "observation equations" then becomes

$$\left[\begin{array}{c} X_1^{(2)} - X_1^{(1)} \\ Y_1^{(2)} - Y_1^{(1)} \\ Z_1^{(2)} - Z_1^{(1)} \\ \hline \vdots \\ \hline X_i^{(2)} - X_i^{(1)} \\ Y_i^{(2)} - Y_i^{(1)} \\ Z_i^{(2)} - Z_i^{(1)} \\ \hline \vdots \\ \hline X_n^{(2)} - X^{(1)} \\ Y_n^{(2)} - Y_n^{(1)} \\ Z_n^{(2)} - Z_n^{(1)} \end{array}\right] = \left[\begin{array}{c|ccc|ccc} X_1^0 & 0 & -Z_1^0 & +Y_1^0 & 1 & 0 & 0 \\ Y_1^0 & +Z_1^0 & 0 & -X_1^0 & 0 & 1 & 0 \\ Z_1^0 & -Y_1^0 & +X_1^0 & 0 & 0 & 0 & 1 \\ \hline \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \hline X_i^0 & 0 & -Z_i^0 & +Y_i^0 & 1 & 0 & 0 \\ Y_i^0 & +Z_i^0 & 0 & -X_i^0 & 0 & 1 & 0 \\ Z_i^0 & -Y_i^0 & +X_i^0 & 0 & 0 & 0 & 1 \\ \hline \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \hline X_n^0 & 0 & -Z_n^0 & +Y_n^0 & 1 & 0 & 0 \\ Y_n^0 & +Z_i^0 & 0 & -X_n^0 & 0 & 1 & 0 \\ Z_n^0 & -Y_n^0 & +X_n^0 & 0 & 0 & 0 & 1 \end{array}\right] \left[\begin{array}{c} m \\ \hline e_x \\ e_y \\ e_z \\ \hline t_x \\ t_y \\ t_z \end{array}\right]. \tag{3.3}$$

This is a set of observation equations of form $\underline{\ell} + \underline{\mathrm{v}} = A\widehat{\mathrm{x}}$ (but without the residuals vector v). There are seven unknowns $\widehat{\mathrm{x}}$ on the right. They can be solved in the least-squares sense if we have co-ordinates (X, Y, Z) in both the old (1) and the new (2) system for at least three points, i.e., nine "observations" in the observation vector ℓ. In fact, two points and one co-ordinate from a third point would suffice. However, it is always good to have redundancy.

3.7 Transformations between ITRF realizations

For transformation parameters between the various ITRF realizations, see the IERS web page: `http://itrf.ensg.ign.fr/trans_para.php`. As an example, the transformation parameters *from* ITRF2008 *to* ITRF2005, at epoch 2005.0, `http://itrf.ensg.ign.fr/ITRF_solutions/2008/tp_08-05.php`:

	$T1$	$T2$	$T3$	D	$R1$	$R2$	$R3$
	mm	mm	mm	ppb	0.001"	0.001"	0.001"
	-0.5	-0.9	-4.7	0.94	0.000	0.000	0.000
±	0.2	0.2	0.2	0.03	0.008	0.008	0.008
Rate	0.3	0.0	0.0	0.00	0.000	0.000	0.000
±	0.2	0.2	0.2	0.03	0.008	0.008	0.008

These parameters[2] are to be used as follows:

$$\begin{aligned}
\begin{bmatrix} X \\ Y \\ Z \end{bmatrix}_{ITRF2005}(t) &= \left\{1+\begin{bmatrix} D & -R_3 & R_2 \\ R_3 & D & -R_1 \\ -R_2 & R_1 & D \end{bmatrix}\right\}\begin{bmatrix} X \\ Y \\ Z \end{bmatrix}_{ITRF2008}(t)+ \\
&+ (t-t_0)\frac{d}{dt}\begin{bmatrix} D & -R_3 & R_2 \\ R_3 & D & -R_1 \\ -R_2 & R_1 & D \end{bmatrix}\begin{bmatrix} X \\ Y \\ Z \end{bmatrix}_{ITRF2008}(t)+ \\
&+ \begin{bmatrix} T_1 \\ T_2 \\ T_3 \end{bmatrix} + (t-t_0)\frac{d}{dt}\begin{bmatrix} T_1 \\ T_2 \\ T_3 \end{bmatrix} = \\
&= \left\{1+\begin{bmatrix} 0.94 & 0 & 0 \\ 0 & 0.94 & 0 \\ -0 & 0 & 0.94 \end{bmatrix}\cdot 10^{-9}\right\}\cdot\begin{bmatrix} X \\ Y \\ Z \end{bmatrix}_{ITRF2008}(t)+ \\
&+ \begin{bmatrix} -0.5+0.3\,(t-2005.0) \\ -0.9 \\ -4.7 \end{bmatrix}
\end{aligned}$$

with the numbers given, and forgetting about the uncertainties. Here $\frac{d}{dt}$ refers to the rates, of which only that of T_1 is non-vanishing in this example.

3.8 Map projections

Typically map projections used for topographic maps are *conformal*; in the Nordic area are used Gauß-Krüger and Universal Transverse Mercator.

Gauß-Krüger:

- Conformal transverse Mercator
- Central meridian: map scale 1.0

[2] Note the change in parameter names compared to the previous. For (t_x, t_y, t_z) we now have $(T1, T2, T3)$; m is now called D; and (e_x, e_y, e_z) became $(R1, R2, R3)$.

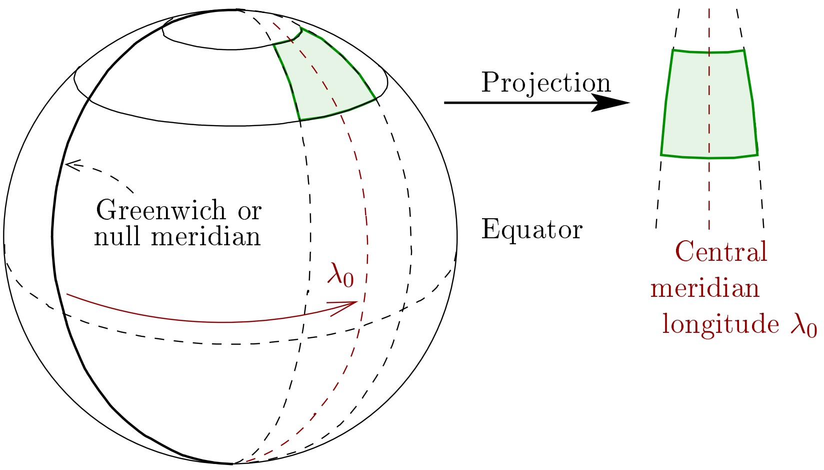


Figure 3.1: Map projection

- Central meridian spacings 3° in longitude

UTM:

- Conformal transverse Mercator
- Central meridian: map scale 0.9996
- Central meridian spacings 6° in longitude
- Zone numbers: starting from date line $\lambda = 180°$ eastward. See table.

UTM zones:

λ (degrees)	Zone Nr	λ (degrees)	Zone Nr
180-176 W	1	0-6 E	31
176-172 W	2	6-12 E	32
...	...	12-18 E	33

λ (degrees)	Zone Nr	λ (degrees)	Zone Nr
18-24 E	34	...	...
24-30 E	35	172-176 E	59
30-36 E	36	176-180 E	60

See also `http://www.dmap.co.uk/utmworld.htm`.

4 What Co-ordinate Frame Do Measurements Give?

4.1 Real time measurement

When doing differential measurement in the form of code-based DGPS or carrier phase based real time kinematic (RTK), it may be assumed that the co-ordinate solution obtained is in the same reference frame as the base station(s) used are.

That would typically be the national realization of the ETRS89 system: in Finland, EUREF-FIN, in Sweden, SWEREF-99, in Norway, EUREF-NOR94/95/96, and in Denmark, EUREF-DK94.

See also [LJ06] and [Lid03].

In practice one may identify this frame with WGS84, as is often done, as it agrees with this on the decimetric level.

4.2 Static measurement

For static precise measurement, the same holds: the reference frame of the base station(s) used transfers to the new position. However, due to the high precision of carrier phase positioning, in large scale (national) campaigns one should compute the network solution originally in the ITRS system in which also the satellite (precise) ephemeris are given, and transform *the result* to the preferred national realization of ETRS89. See [BA].

If the base station co-ordinates are in the national realization, they should be transformed to the ephemeris ITRS before use.

4.3 (Aside) antenna type

Using base station data assumes that the moving receiver ("rover") uses the same antenna type, or the software understands antenna phase centre variation modelling.

Both the RTCM-106 version 2.x and the new RTCM version 3.0 can transmit information on antenna type in Message Type 23, and Message Type 1007-1008. Version 2.x even offers the offsets of the antenna phase centres for L1 and L2 separately, promising "mm precision". Version 3.0 doesn't do that.

Apparently it is up to the receiving software to handle the antenna phace centre effects intelligently, using information from the base station on antenna/radome type, and internal calibration tables. It is not clear to what extent existing software is doing this.

5 Case: Finland

5.1 Traditional map projections

In Finland, the traditional map projection has been Gauß-Krüger with zone width 3°. System name: *kkj* ("National Map Grid Co-ordinate System") created in 1970 [Par88]. Following characteristics:

- Based on International (Hayford) reference ellipsoid of 1924; datum was taken from the European datum of 1950 by keeping fixed the triangulation point Simpsiö (nr. 90), at
  - latitude and longitude values from the ED50 European adjustment, and
  - geoidal undulation from the Bomford astro-geodetic geoid [Bom63].
- Map plane co-ordinates were obtained using Gauß-Krüger for central meridians of $19°, 21°, 24°, 27°, 30°$; for small-scale all-Finland maps, $27°$ is used.
- These co-ordinates (x, y) were further transformed in the plane using a four-parameter similarity ("Helmert") transformation in order to achieve agreement with the pre-existing provisional *vvj* ("Old State System", also "Helsinki System") co-ordinates, cf. [Oll93].

  Equation:

$$\begin{bmatrix} x \\ y \end{bmatrix}_{kkj} = \begin{bmatrix} 1.00000075 & -0.00000439 \\ 0.00000439 & 1.00000075 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}_{ED50} + \begin{bmatrix} -61.571\,\mathrm{m} \\ 95.693\,\mathrm{m} \end{bmatrix}$$

5.2 Modern map projections

The modern Finnish system is quite different:

- Based on the GRS80 reference ellipsoid of 1980; datum is called EUREF-FIN, created by keeping fixed four stations fixed to their ITRF96 values at epoch 1997.0: the permanent GPS stations Metsähovi, Vaasa, Joensuu and Sodankylä [OKP00]. Then, a *transformation* [BA95] was applied to obtain co-ordinates in ETRF89. Thus, the datum is correctly described as ETRF89, but the *epoch* remains 1997.0, as no correction for individual station motion (mostly, glacial isostatic adjustment) was made in the transformation.

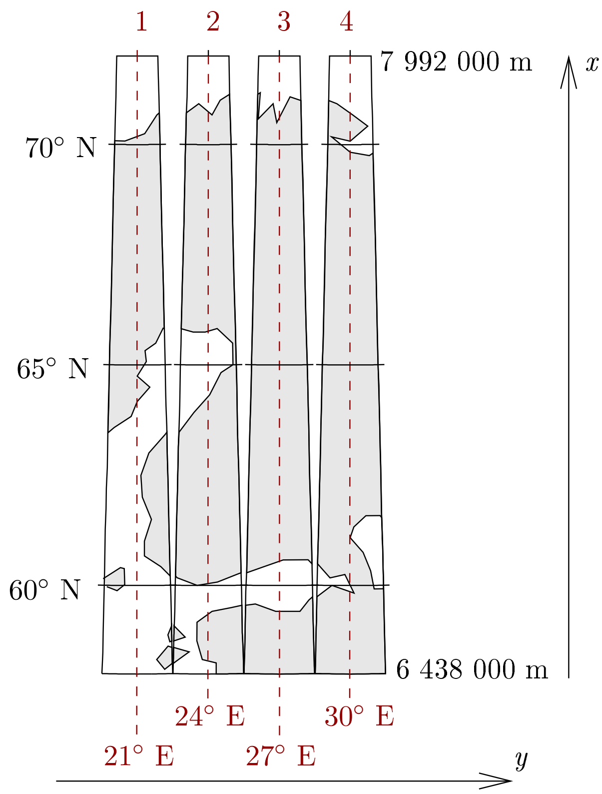


Figure 5.1: Four projection zones for the Finnish Gauss-Krüger projection

Equation:

$$\mathbf{X}^E(t_C) = \mathbf{X}^I_{yy}(t_C) + \mathbf{T}_{yy} + \begin{bmatrix} 0 & -\dot{R}_3 & \dot{R}_2 \\ \dot{R}_3 & 0 & -\dot{R}_1 \\ -\dot{R}_2 & \dot{R}_1 & 0 \end{bmatrix}_{yy} \mathbf{X}^I_{yy}(t_C) \cdot (t_C - 1989.0)$$

with t_C observations central epoch, $yy = (19)96$. The values $\mathbf{T}_{96}$ and $\dot{R}_{i,96}$ are tabulated in [BA] Tables 3 and 4.

- For small-scale and topographic maps, the UTM projection is used with a central meridian of 27° (zone 35) *for the whole country*, producing the ETRS-TM35FIN plane co-ordinate system. This also defines the map sheet division. However, on maps in parts of Finland where another central meridian would be more appropriate (like zone 34, central meridian 21°), the corresponding co-ordinate grid is also printed on the map, in purple [Ano03].

- For large scale maps, such as used for planning and cadastral work, the Gauß-Krüger projection continues to be used (but based on the above reference ellipsoid and datum), with a central meridian interval of only one degree: ETRS-GKn, where n designates the central meridian longitude. This avoids the problem of significant scale distortions.

5.3 The triangulated affine transformation used in Finland

5.3.1 Plane co-ordinates

The National Land Survey offers a facility to convert *kkj* co-ordinates to the new ETRS89-TM35FIN system of projection co-ordinates. The method is described in the publication [Ano03], where it is proposed to use for the plane co-ordinate transformation between the projection co-ordinates of ETRS-89 and the *ykj* co-ordinate system, a *triangle-wise affine transformation.*

Inside each triangle we may write the affine transformation can be written like

$$\begin{aligned} x^{(2)} &= \Delta x + a_1 x^{(1)} + a_2 y^{(1)} \\ y^{(2)} &= \Delta y + b_1 x^{(1)} + b_2 y^{(1)} \end{aligned}$$

where $\left(x^{(1)}, y^{(1)}\right)$ are the point co-ordinates in ETRS-GK27, and $\left(x^{(2)}, y^{(2)}\right)$ are the co-ordinates of the same point in *ykj*. This transformation formula has six parameters : Δx, Δy, a_1, a_2, b_1 ja b_2. If, in the three corners of the triangle, are given both $\left(x^{(1)}, y^{(1)}\right)$ and $\left(x^{(2)}, y^{(2)}\right)$, we can solve for these uniquely .

The transformation formula obtained is inside the triangles linear and continuous across the edges, but not differentiable: the scale is discontinuous across triangle edges. Because the mapping is not conformal either, the scale will also be dependent upon the direction considered.

A useful property of triangulation is, that it can be locally "patched" : if better data is available in the local area – a denser point set, whose co-ordinate pairs $\left(x^{(i)}, y^{(i)}\right)$, $i = 1, 2$ are known – then we can take away only the triangles of that area and replace them by a larger number of smaller triangle, inside which the transformation will become more precise. This is precisely the procedure that local players, like municipalities, can use to advantage.

Write these equations in vector form:

$$\begin{bmatrix} x^{(2)} \\ y^{(2)} \end{bmatrix} = \begin{bmatrix} \Delta x \\ \Delta y \end{bmatrix} + \begin{bmatrix} a_1 & a_2 \\ b_1 & b_2 \end{bmatrix} \begin{bmatrix} x^{(1)} \\ y^{(1)} \end{bmatrix}.$$

Most often the co-ordinates in the (1)and (2)datums are close to each other, i.e., $\begin{bmatrix} \Delta x & \Delta y \end{bmatrix}^{\mathrm{T}}$ are small. In that case we may write the *shifts*

$$\begin{aligned} \delta x &\equiv x^{(2)} - x^{(1)} = \Delta x + (a_1 - 1)\, x^{(1)} + a_2 y^{(1)}, \\ \delta y &\equiv y^{(2)} - y^{(1)} = \Delta y + b_1 x^{(1)} + (b_2 - 1)\, y^{(1)}. \end{aligned}$$

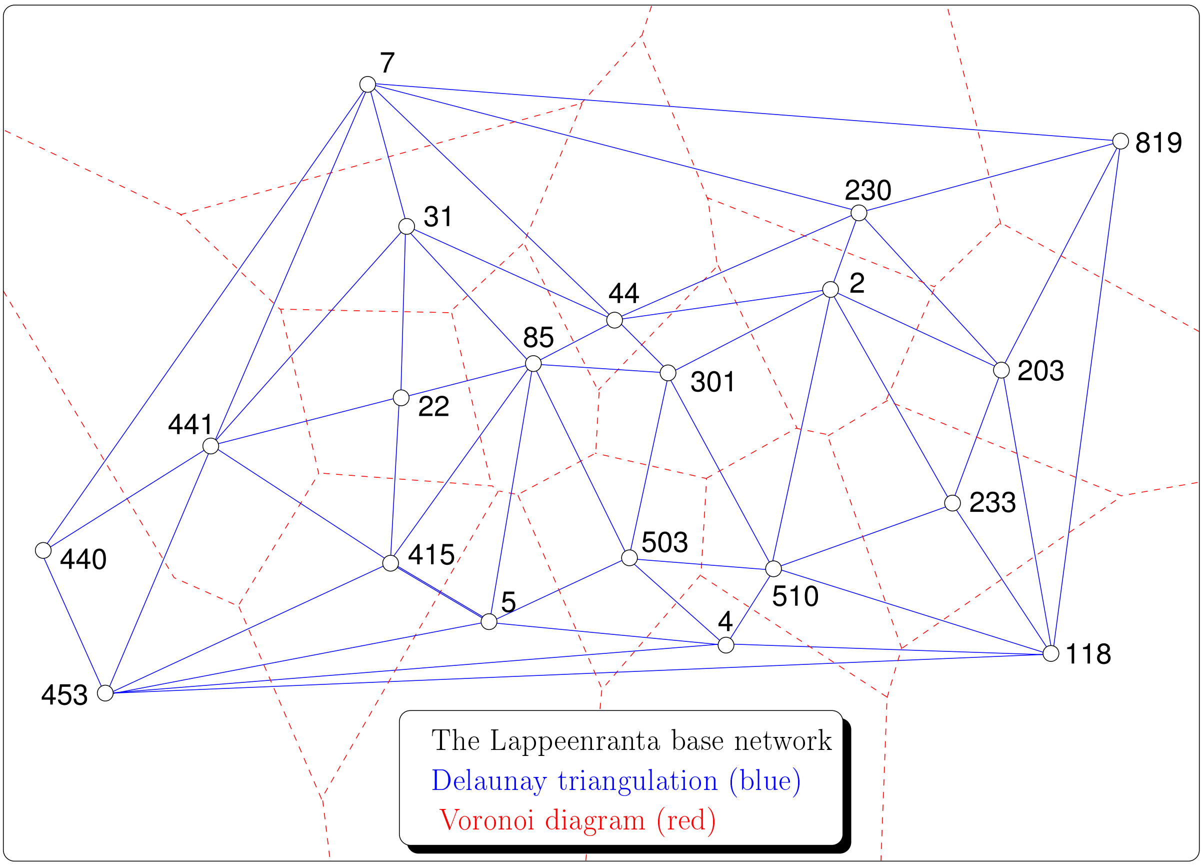


Figure 5.2: Lappeenranta densification of the national triangular grid

If we now define

$$\Delta\mathbf{x} \equiv \begin{bmatrix} \Delta x \\ \Delta y \end{bmatrix}, \; \mathbf{A} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \equiv \begin{bmatrix} a_1 - 1 & a_2 \\ b_1 & b_2 - 1 \end{bmatrix},$$

we obtain the short form

$$\delta\mathbf{x} = \Delta\mathbf{x} + \mathbf{A}\mathbf{x}^{(1)}.$$

Also in this generally, if the co-ordinates are close together, the elements of $\mathbf{A}$ will be numerically small. Let there be a triangle ABC. Then we have given the shift vectors of the corners

$$\begin{aligned} \delta\mathbf{x}_A &= \Delta\mathbf{x} + \mathbf{A}\mathbf{x}_A^{(1)}, \\ \delta\mathbf{x}_B &= \Delta\mathbf{x} + \mathbf{A}\mathbf{x}_B^{(1)}, \\ \delta\mathbf{x}_C &= \Delta\mathbf{x} + \mathbf{A}\mathbf{x}_C^{(1)}. \end{aligned}$$

Write this out in components, with $\Delta\mathbf{x}, \mathbf{A}$ on the right hand side:

$$\begin{aligned}
\delta x_A &= \Delta x + a_{11}x_A^{(1)} + a_{12}y_A^{(1)} \\
\delta y_A &= \Delta y + a_{21}x_A^{(1)} + a_{22}y_A^{(1)} \\
\delta x_B &= \Delta x + a_{11}x_B^{(1)} + a_{12}y_B^{(1)} \\
\delta y_B &= \Delta y + a_{12}x_B^{(1)} + a_{22}y_B^{(1)} \\
\delta x_C &= \Delta x + a_{11}x_C^{(1)} + a_{12}y_C^{(1)} \\
\delta y_C &= \Delta y + a_{21}x_C^{(1)} + a_{22}y_C^{(1)}
\end{aligned}$$

or in matrix form

$$\begin{bmatrix} \delta x_A \\ \delta y_A \\ \delta x_B \\ \delta y_B \\ \delta x_C \\ \delta y_C \end{bmatrix} = \begin{bmatrix} 1 & 0 & x_A^{(1)} & 0 & y_A^{(1)} & 0 \\ 0 & 1 & 0 & x_A^{(1)} & 0 & y_A^{(1)} \\ 1 & 0 & x_B^{(1)} & 0 & y_B^{(1)} & 0 \\ 0 & 1 & 0 & x_B^{(1)} & 0 & y_B^{(1)} \\ 1 & 0 & x_C^{(1)} & 0 & y_C^{(1)} & 0 \\ 0 & 1 & 0 & x_C^{(1)} & 0 & y_C^{(1)} \end{bmatrix} \begin{bmatrix} \Delta x \\ \Delta y \\ a_{11} \\ a_{21} \\ a_{12} \\ a_{22} \end{bmatrix},$$

from which they can all be solved.

Let us write the coordinates (x, y) as follows:

$$\begin{aligned}
x &= p^A x_A + p^B x_B + p^C x_C, \\
y &= p^A y_A + p^B y_B + p^C y_C,
\end{aligned}$$

with the further condition $p^A + p^B + p^C = 1$. Then also

$$\begin{aligned}
\delta x &= p^A \delta x_A + p^B \delta x_B + p^C \delta x_C, && (5.1) \\
\delta y &= p^A \delta y_A + p^B \delta y_B + p^C \delta y_C. && (5.2)
\end{aligned}$$

The set of three numbers $\left(p^A, p^B, p^C\right)$ is called the *barycentric co-ordinates* of point P See figure 5.3.

They can be found as follows (geometrically $p^A = \frac{\omega(\Delta BCP)}{\omega(\Delta ABC)}$ etc., where ω is the surface area of the triangle) using determinants :

$$p^A = \frac{\begin{vmatrix} x_B & x_C & x \\ y_B & y_C & y \\ 1 & 1 & 1 \end{vmatrix}}{\begin{vmatrix} x_A & x_B & x_C \\ y_A & y_B & y_C \\ 1 & 1 & 1 \end{vmatrix}}, \; p^B = \frac{\begin{vmatrix} x_C & x_A & x \\ y_C & y_A & y \\ 1 & 1 & 1 \end{vmatrix}}{\begin{vmatrix} x_A & x_B & x_C \\ y_A & y_B & y_C \\ 1 & 1 & 1 \end{vmatrix}}, \; p^C = \frac{\begin{vmatrix} x_A & x_B & x \\ y_A & y_B & y \\ 1 & 1 & 1 \end{vmatrix}}{\begin{vmatrix} x_A & x_B & x_C \\ y_A & y_B & y_C \\ 1 & 1 & 1 \end{vmatrix}}.$$

These equations are very suitable for coding.

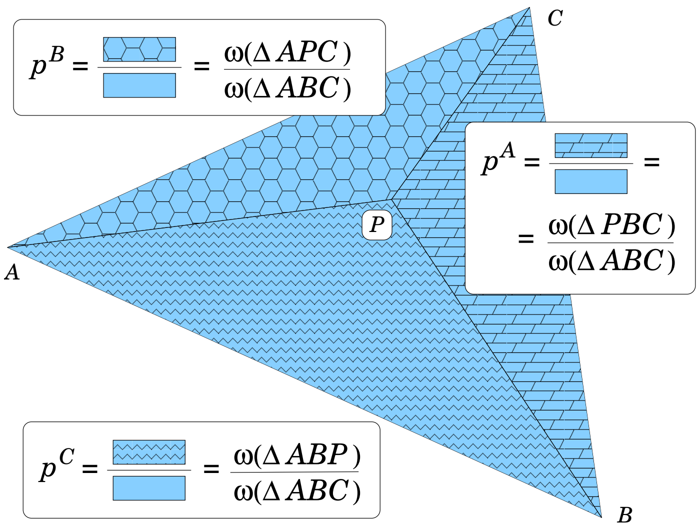


Figure 5.3: Computing barycentric co-ordinates as the ratio of the surface areas of two triangles

5.3.2 Heights

The same method for triangulated linear transformation can also be used for one-dimensional co-ordinates like heights. In this case the model reduces to a TIN (Triangulated Irregular Network) representation.

5.4 Inter-zone transformations

The easy way to do inter-zone transformations between zone 1 and zone 2 is:

$$(x, y)_1 \overset{\text{Proj 1}}{\Longleftrightarrow} (\varphi, \lambda) \overset{\text{Proj 2}}{\Longleftrightarrow} (x, y)_2$$

However, also direct transformations can be constructed. As the projection is conformal for both zones, it is possible to express the transformation as a complex polynomial in the (x, y) map plane. One can solve the coefficients of this polynomial by a least-squares fit to a small number of points in the target area for which both map co-ordinates have been computed, and then hard-wired in the transformation software. This is numerically very efficient, and not many polynomial coefficients are needed for adjacent zones.

6 Height Systems and Vertical Datums

6.1 Height types

There are two dominant height types in the world: *orthometric* heights and *normal* heights. Dynamic heights are much lesser used.

6.2 Orthometric heights

Levelled heights are related to geopotential numbers by local gravity. Heights obtained by summing levelled height differences ΔH: $\sum_{i=1}^{3} \Delta H_i$, are not the same as "real" heights from the geoid: $\sum_{i=1}^{3} \Delta H'_i$ along the plumbline. Note how *equipotential surfaces* are not parallel: thus, water may flow "upward".

Note also that the gravity vector $\mathbf{g}$ is *everywhere perpendicular* to equipotential surfaces, and its length is inversely proportional to distance between the surfaces for "round" values of the geopotential.

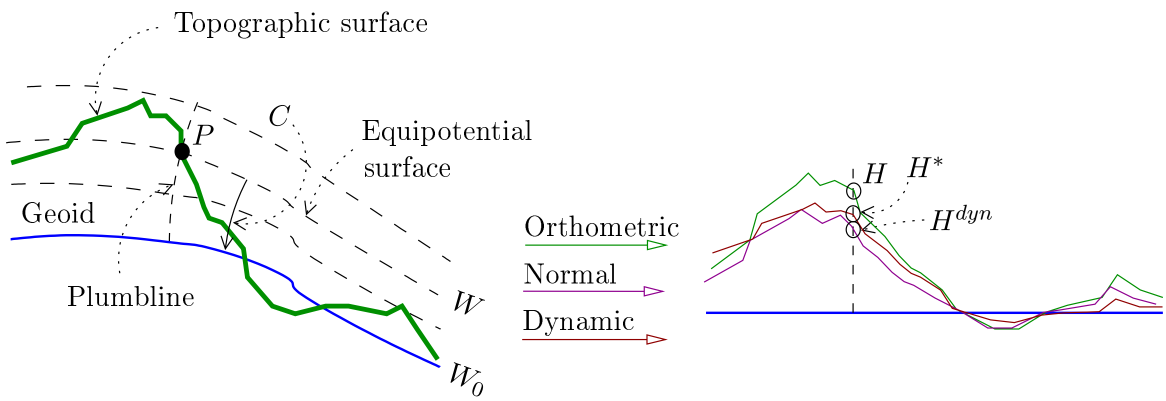


Figure 6.1: The different height types, simply explained. The differences between them are exaggerated

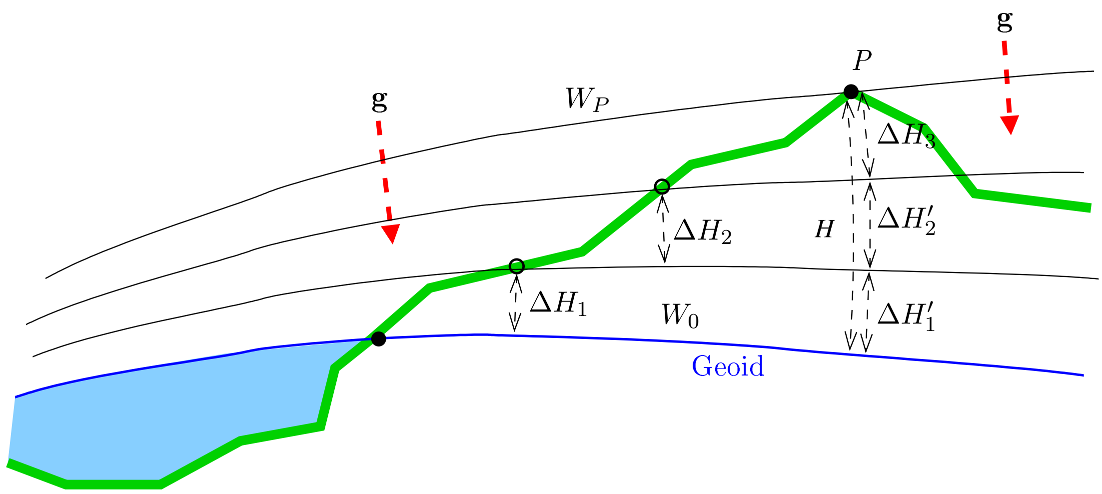


Figure 6.2: The definition of orthometric heights

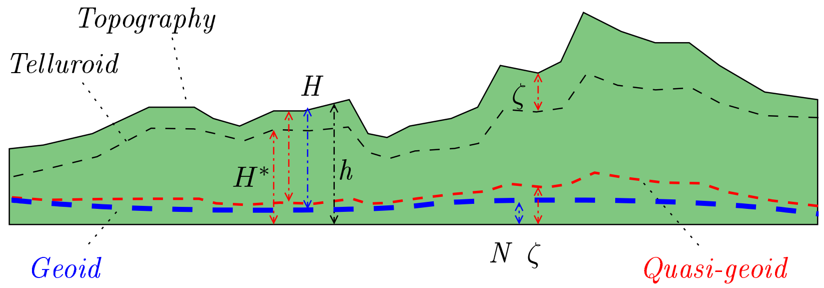


Figure 6.3: Geoid, telluroid and quasi-geoid

Computing orthometric heights:

$$H = \frac{C_P}{\bar{g}},$$

with

$$\bar{g} = \frac{1}{H} \int_0^H g\,(z)\,dz,$$

with z along the plumbline. Solved iteratively. *Problem*: we don't really know $g\,(z) \Rightarrow$ often, we will use the *Helmert height* approximation instead of exact orthometric height.

6.3 Normal heights

Normal heights are based on MOLODENSKY's theory [MEY62]. In this theory, we need a surface called the *telluroid*:

In this theory, we talk about a point Q where the *normal* potential is equal to P:s *true* potential: $U_Q = W_P$. Then, we say that Q is P:s *telluroid mapping.*

Now, normal height is

$$H^* = \frac{C}{\overline{\gamma_{0H}}},$$

with $\overline{\gamma_{0H}}$ the mean *normal* gravity along the *ellipsoidal* normal from 0 (ellipsoid) to H^*(telluroid, i.e., level of Q).

Molodensky also proposed the *height anomaly*:

$$\zeta = \frac{W - U}{\bar{\gamma}_{Hh}}$$

with

$$\bar{\gamma}_{Hh} = \frac{1}{\zeta} \int_{H^*}^{h} \gamma(z) dz,$$

i.e., integrated from the telluroid to the Earth surface... note that z is counted from the reference ellipsoid.

6.4 Geoid, quasi-geoid, geopotential

The geoid is connected with height systems through the geopotential. We have, *at sea level*

$$T = W - U$$

the disturbing potential T, with true geopotential W and normal potential U; and the geoid undulation

$$N = \frac{T(0)}{g(0)}$$

(Bruns) where g is the average gravity at sea level.

We have, with H orthometric height and h ellipsoidal height:

$$h = H + N.$$

When using normal heights H^*, we would like to use a similar equation:

$$h = H^* + \zeta$$

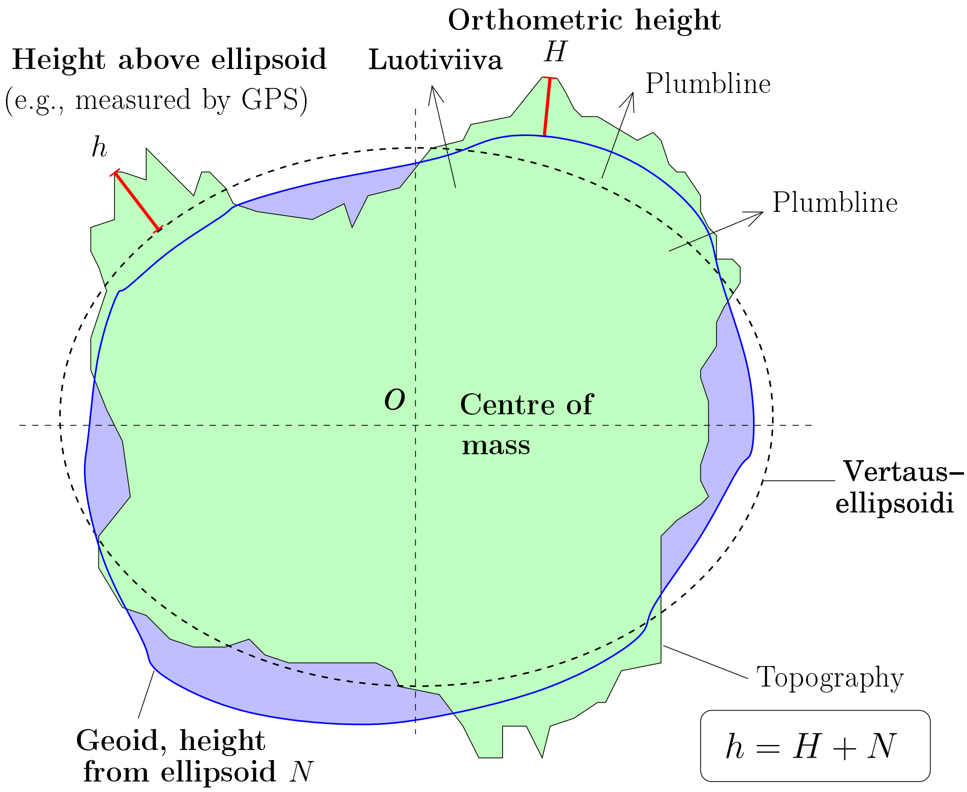


Figure 6.4: Heights above the geoid and above the reference ellipsoid in a global picture

where we call ζ the “quasi-geoid height”. It is also equal to

$$\zeta(H^*) = \frac{T(H^*)}{g(H^*)},$$

the so-called height anomaly *on the terrain*, not at sea level. So, note that ζ is a *spatial* field, not a 2-D map.

Gravity anomalies If we measure gravity g in a point P, its height “above sea level” is H, and its latitude is Φ, we may compute its gravity anomaly as follows:

$$\Delta g_P \equiv g_P - \gamma(H, \Phi),$$

where $\gamma(H, \Phi)$ is normal gravity, computed at height H and ja latitude Φ. Thus are defined *free air anomalies*.

Bouguer anomalies are computed in order to remove the attraction by the masses of the Earth's crust located above sea level, i.e., the geoid. The true topography is *approximated* by a Bouguer plate, see figure 6.6. The difference between the Bouguer plate attraction and the true attraction is called the *terrain correction* (areas I and II).

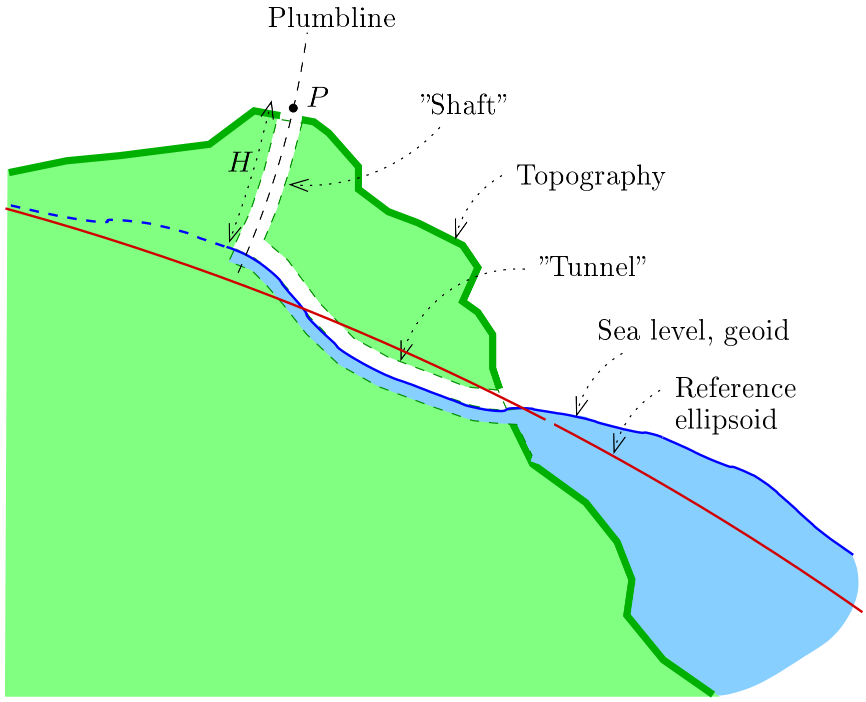


Figure 6.5: An "engineering definition" of the geoid and orthometric height

As the difference between H and H^* is approximately equal to

$$H - H^* = -\frac{\Delta g_B}{\gamma} H,$$

it follows that also

$$\zeta - N = -\frac{\Delta g_B}{\gamma} H.$$

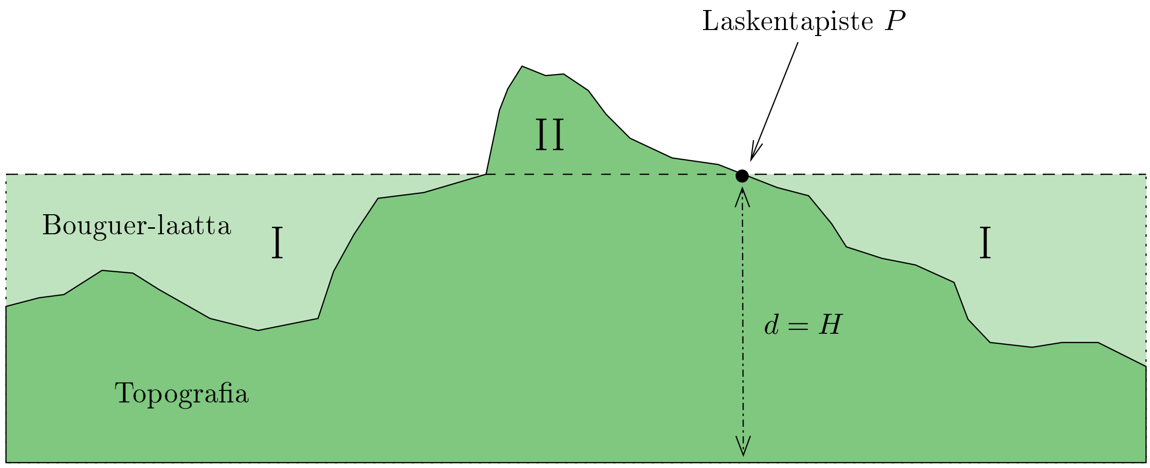
Here Δg_B is BOUGUER gravity anomaly at point considered. In mountainous terrain, approx. (in metres and milligals)

$$\Delta g_B \approx -0.1H,$$

so it follows

$$\zeta - N = H - H^* \approx -10^{-7} H^2,$$

expressing everything in metres. Thus, for 1 km high mountains we get -10 cm, but for 100 m high hills, only -1 mm.



Kuva 6.6: BOUGUER plate as an approximation to the topography

6.5 Geoid tidal types

A complication with height systems and (quasi-)geoids is that there are different ways in use to correct for the permanent part of the tidal potential caused by Sun and Moon:

- The *mean* geoid, where both direct effect of the lunisolar tidal potential, and effect of change in geopotential due to tidal deformation, have been left uncorrected. This geoid solution is best suited to oceanographic use as undisturbed mean sea level will strive toward this geoid.
- The *null* geoid, where direct tidal potential has been removed, but tidal deformation potential left in. Merit of this geoid solution is that it produces a potential associated only with masses internal to geoid (much theory, including the Stokes equation, is based on this assumption), but no questionable deformation hypotheses needed.
- The *tide-free* geoid, where both the direct effect of the lunisolar tidal potential has been computationally removed, including its permanent part, and change in the geopotential due to the permanent tidal deformation.

In computations with geoid heights, orthometric / normal heights and ellipsoidal (GPS) heights, one must have all three reduced according to the same treatment of the permanent tide. E.g., in

$$h = N + H$$

all three quantities must be considered.

Sweden and Denmark have used non-tidal heights; in Finland, heights have been mean. The correction from nontidal to mean is given by [EM96]:

$$\Delta H = 29.6\gamma\left(\sin^2\varphi_N - \sin^2\varphi_S\right),$$

where φ_N and φ_S are the latitudes of a Northern and Southern station, and $\gamma \approx 0.7 \ldots 0.8$ is the elasticity factor.

6.6 Old height datums

A vertical reference system must be *realized*: this means choosing a starting point and its height. Traditionally, every nation has chosen its own height datum. The old (mid to late 20th century) height datums are:

Sweden: RH70, the datum point computationally being Amsterdam (NAP), based on second precise levelling.

Norway: NN1954 for South of country, normal or orthometric (unclear). Formal datum point: Tregde. No unified vertical datum [LPMS06]

Finland: N60, the datum point being a polished surface of a stone pillar in the garden of Helsinki astronomical observatory. The agreed height corresponded closely to mean sea level in Helsinki Harbour for 1960.0. The height type was Helmert heights using crustal densities from a geological map; a good approximation to orthometric heights.

Denmark: DNN (Dansk Normal Nul, Danish Ordnance Datum). Based on ten tide gauges where sea level was monitored 1885-1904, and connected across the country by precise levelling, including hydrostatic transfers
(*). The formal datum point was a metal plate attached to the wall of Århus Cathedral.

*) `http://www.slideboom.com/presentations/37439/DNN-og-andre-kotesystemer`

6.7 New height datums

Modern height datums have been and are being established throughout the Nordic region. The situation is now, that it is desirable that these datums are internationally compatible. This means:

- All are based upon the Amsterdam datum (NAP); the Baltic levelling Ring used to connect them
- All are of the same height type: normal heights
- Important work doen by EUVN (European Vertical Network) EUREF working group
- True global height datum at this point still impossible
- Use of these heights together with satellite techniques (like GPS) should be straightforward.

Sweden: RH2000; Finland: N2000; Denmark: DVR. In Norway, a new height datum has not yet been established, but see [LOPS07].

7 Height Systems and Geophysics

We are entering an age where creating height datums at the precision that is technologically possible, requires understanding of the geophysics involved. This includes

- The time-varying and quasi-permanent sea surface topography
- The ongoing process of glacial isostatic adjustment, due to old (≈15 ka) deglaciation
- The ongoing process of sea level rise, which is partly due also to continental ice sheet melt
- Techniques for integrating sea level data from tide gauges and satellite altimetry
- The *acceleration* of current sea level rise
- The *unevenness* in this process, as ice sheet mass is being redistributed [MTDM01]
- The *isostatic response* to this process

This is an unfinished story during our lifetime.

7.1 Sea Surface Topography

Sea surface topography is the (semi-) permanent deviation of sea level from an equipotential surface, i.e., the geoid.

SST is typically of magnitude few decimetres. Intercontinentally it can be as large as over one metre. One should also be aware that SST may change as climate changes; such changes are suspected for the Baltic Sea, as the pumping effect in the Danish Straits changes due to depressions taking a different track.

When defining a vertical reference frame, obviously these effects should be considered.

Source: [PS00]

SST change can be several decimetres. Considering global warming according to the IPCC A1 scenario, comparing the period 2091-2100 to 1981-2000, [JSS09] found changes for the North Atlantic due to changes in ocean circulation of this magnitude. As they note, “The dynamic SLR is mainly a result of the cessation of the deep convection and deep-water formation in the Labrador Sea, and the slowdown of the subpolar gyre.”

7.2 Glacial isostatic adjustment

This phenomenon, more commonly known as "land uplift" – in spite of this being a three-dimensional phenomenon! – , obviously affects the definition of vertical reference systems. We already saw how, in the definition of the three-dimensional EUREF-FIN datum in Finland, the epoch for crustal motions was kept at 1997.0, although the coordinate reference system is stated as ETRS89. In the definition of European terrestrial reference frames, glacial isostatic adjustment is not accounted for.

A good model for the post-glacial land uplift, usable for correcting heights measured by GPS into heights "above sea level", was constructed using least-squares collocation [Ves06]. One should expect such models to become official in the future, when height system maintenance is moving to GPS, traditional precise levelling being too expensive.

7.3 Tide gauges, altimetry, sea level

The advent of satellite radar altimetry has made possible the *combination* of tide gauge and altimetry data in an optimal way using PCA; see [CW06]. Such studies are tricky and require, e.g., careful corrections for the ongoing glacial isostatic adjustment, both at the tide gauge sites used, and over the ocean at large. The *merit* of the approach is that the altimetry allows the assessment of the behaviour of the sea surface over the open ocean, and applying this to the geographically more limited data from tide gauges. From this study and others, it is clear that sea level rise has increased over the 20th century: the rise from 1870 to 2000 amounts to 20 cm, but, while over the whole 20th century sea level rise has amounted to $1.31 \pm 0.30\,\mathrm{mm/yr}$[WMBA07], over the satellite altimetry period 1993-2008 it has been $3.1 \pm 0.4\,\mathrm{mm/yr}$ [CL10].

7.4 Geodynamics and the vertical reference

See [MTDM01] for how the sea level rise due to the Greenland deglaciation is *uneven* in the Northern hemisphere.

Sea level rise will be very relevant for vertical reference, and vertical reference has immediate relevance for sea level rise impact studies.

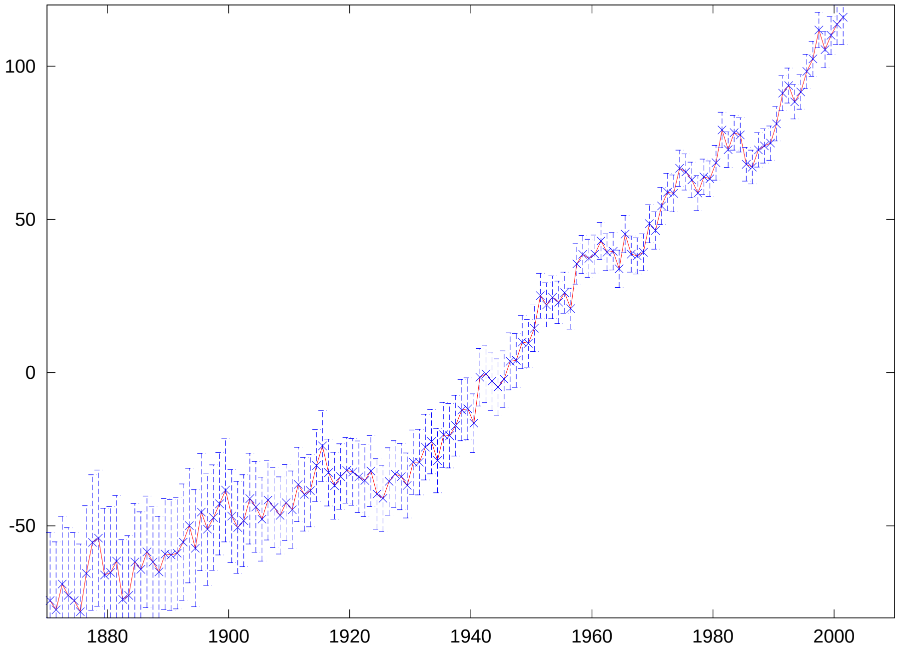


Figure 7.1: Sea level rise according to [CW06]

Bibliography

[Ano03] Anon. JHS154. ETRS89 -järjestelmään liittyvät karttaprojektiot, tasokoordinaatistot ja karttalehtijako (Map projections, plane co-ordinates and map sheet division in relation to the ETRS89 system). Web site, Finnish National Land Survey, 2003. URL: `http://www.jhs-suositukset.fi/suomi/jhs154`, accessed August 30, 2005.

[Ano08] Anon. JHS153. ETRS89-jŗjestelmŭ mukaiset koordinaatit Suomessa. Web site, Finnish National Land Survey, 2008. URL: `http://www.jhs-suositukset.fi/suomi/jhs153`, accessed Oct. 19, 2010.

[BA] Claude Boucher and Zuhair Altamimi. *Specifications for reference frame fixing in the analysis of a EUREF GPS campaign.* Version 4, 24-10-2008, `http://etrs89.ensg.ign.fr/memo-V7.pdf`.

[BA95] Claude Boucher and Zuhair Altamimi. Specifications for Reference Frame Fixing in the Analysis of a EUREF GPS Campaign. In *VI Meeting of the TWG in Bern, March 9 and 10, Veröff. der Bayerischen Kommission für die Internationale Erdmessung*, number 56, pages 265–267, München, 1995.

[Bom63] Guy Bomford. Report of Study Group No. 10. The Geoid in Europe and Connected Countries. In *Traveaux de l'Association Internationale de Géodésie*, number 22, 1963.

[CL10] Anny Cazenave and William Llovel. Contemporary Sea Level Rise. *Annual Review of Marine Science*, 2:145–173, 2010.

[CW06] J. A. Church and N. J. White. A 20th century acceleration in global sea-level rise. *Geophys. Res. Lett.*, 33:L01602, 2006.

[EM96] Martin Ekman and Jaakko Mäkinen. Mean sea surface topography in the Baltic Sea and its transition area to the North Sea: A geodetic solution and comparisons with oceanographic models. *Journal of Geophysical Research*, 101(C5):11,993–11,999, May 15 1996.

[Gro00] Richard S. Gross. The excitation of the Chandler wobble. *Geophys. Res. Lett.*, 27:2329–2332, 2000.

[JSS09] Jianjun Jin, Michael E. Schlesinger, and Ronald J. Stouffer. Model projections of rapid sea-level rise on the northeast coast of the United States. *Nature Geoscience*, 2(4):262–266, 2009.

[KR06] Muneendra Kumar and James P. Reilly. Is definition of WGS 84 correct. URL: http://mycoordinates.org/is-definition-of-wgs-84-correct/, Dec. 2006.

[Lid03] Martin Lidberg. Future height systems in the Nordic countries, their relation to the EVRS2000 and to INSPIRE GIS standards. In *EUREF Technical Working Group meeting, Toledo*. EUREF/IAG, June 3 2003. `http://www.euref-iag.net/symposia/book2003/5-3.pdf`.

[LJ06] Martin Lidberg and Jan M. Johansson. Geodetic reference frames in the presence of crustal deformations. In *EUREF symposium Riga*, June 14-16 2006. `http://www.euref-iag.net/symposia/2006Riga/06-07.pdf`.

[LOPS07] D.I. Lysaker, O.C.D. Omang, B.R. Pettersen, and D. Solheim. Quasi-geoid evaluation with improved levelled height data for Norway. *Journal of Geodesy*, 81(9):617–627, September 2007.

[LPMS06] D.I. Lysaker, B.R. Pettersen, O. Mathisen, and D. Solheim. The Norwegian Height System NN1954 Revisited. *Nordic Journal of Surveying and real Estate Research*, 3(1):7–19, 2006. `http://mts.fgi.fi/njsr/issues/2006/njsrv3n12006_lysaker_et_al.pdf`.

[McC96] Dennis D. McCarthy. IERS Conventions (1996). (IERS Technical Note ; 21). Technical report, Central Bureau of IERS - Observatoire de Paris, 1996. URL: http://www.iers.org/nn_11216/IERS/EN/Publications/TechnicalNotes/tn21.html.

[MEY62] M. S. Molodensky, V. F. Eremeev, and M. I. Yurkina. *Methods for the study of the external gravitational field and figure of the Earth.* Israel program for scientific translations, Jerusalem, 1962. Transl.

[MTDM01] Jerry X. Mitrovica, Mark E. Tamisiea, James L. Davis, and Glenn A. Milne. Recent mass balance of polar ice sheets inferred from patterns of global sea level change. *Nature*, 409:1026–1029, February 2001.

[OKP00] Matti Ollikainen, Hannu Koivula, and Markku Poutanen. The densification of the EUREF network in Finland. Publication 129, Finnish Geodetic Institute, Kirkkonummi, 2000.

[Oll93] Matti Ollikainen. GPS-koordinaattien muuntaminen Kartastokoordinaateiksi (Transforming GPS co-ordinates to the National Grid Co-ordinate System). Tiedote 8, Finnish Geodetic Institute, Helsinki, 1993.

[Par88] Teuvo Parm. Kansallisen koordinaattijärjestelmän luominen Suomessa (The creation of a national co-ordinate system in Finland). *Maanmittaus*, (1), 1988.

[PS00] Markku Poutanen and Tapani Stipa. Temporal and spatial variation of the sea surface topography of the Baltic Sea. Presented at the Gravity, Geoid and Geodynamics meeting in Banff, August 20 2000.

[Ste08] Phil Stevenson. The Truth About WGS84. URL: http://www.ashgps.com/out3/GeneralNovember 2008.

[Tor01] W. Torge. *Geodesy.* Walter de Gruyter, Berlin, New York, 2001.

[Ves06] Olaf Vestøl. Determination of postglacial land uplift in Fennoscandia from leveling, tide-gauges and continuous GPS stations using least squares collocation. *Journal of Geodesy*, 80:248–258, 2006.

[WMBA07] G. Wöppelmann, B. Martin Miguez, M.-N. Bouin, and Z. Altamimi. Geocentric sea-level trend estimates from gps analyses at relevant tide gauges world-wide. *Global and Planetary Change*, 57(3-4):396 – 406, 2007.