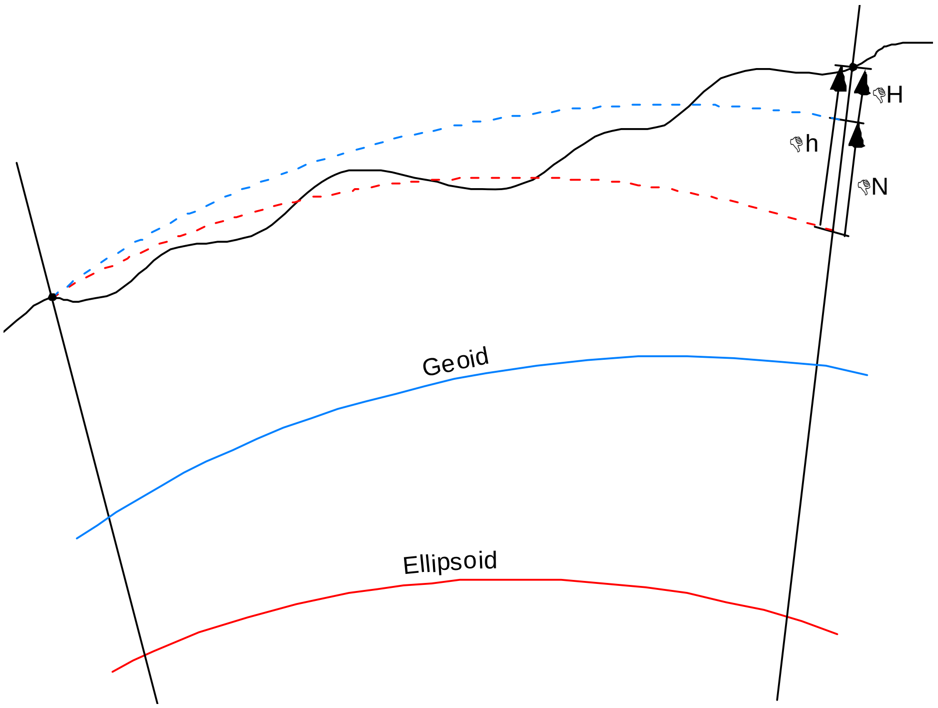
451-337 Satellite Positioning and Geodesy
Exam Solutions 2005

1. (a) GPS is a truly 3D measurement system, providing baselines (vectors) between points. However, the height component of a GPS baseline is not immediately useful for most surveying applications. Explain why this is so and illustrate the problem with an appropriately annotated diagram.

(5 marks)

The baseline that results from two simultaneously operating GPS receivers is expressed in cartesian components (ΔX, ΔY, ΔZ) relative to the WGS84 reference frame. When that baseline is converted into the geodetic coordinate differences, it is expressed as (Δφ, Δλ, Δh), where Δh is the ellipsoidal height difference between the end points. The ellipsoidal height difference is not a useful measure of “height” in conventional terms because ellipsoidal heights (and height differences) do not strictly relate to earth’s gravity field. It is for this reason that national height datums like the AHD attempt to use the geoid rather than the ellipsoid as a reference surface. The geoid is an equipotential of the earth’s gravity field and heights related to the geoid (denoted by H) will indicate directions and rates of fluid flow. The following diagram illustrates the problem :



(b) Two significant advantages emerge if the GPS heighting problem alluded to in Question 1 (a) can be overcome. Provide an explanation of these advantages and, in this context, briefly outline options for solution of the problem.

(5 marks)

The problem of using GPS for heighting purposes hinges on resolving the relative geoid undulation, denoted by ΔN in the above diagram. If this can be done, then ellipsoidal height differences from GPS (Δh) can be converted into orthometric height differences (ΔH). Achieving this transformation from ellipsoidal to orthometric height differences to a sufficient degree of accuracy will allow :

- ☞ GPS to be used for levelling, thus eliminating the need for conventional levelling in all but the most demanding cases.
- ☞ GPS baselines to be incorporated into rigorous 3D adjustments with conventional observations, including levelling.

Solution to the GPS heighting problem means modelling the behaviour of the geoid relative to the relevant geodetic ellipsoid. A number of options exist for this purpose. These include :

- ☞ Geopotential models such as EGM96. This provides a good long wavelength model of the geoid but lacks the ability to resolve short wavelength features.
- ☞ Planar geoid models. Means using GPS and levelling or existing benchmarks to obtain a direct measure of the geoid undulation at discrete points and then using this information to create a planar model of the geoid surface. Higher order surfaces can be used but are not recommended.
- ☞ Gravity based geoid modelling techniques such as Stokes integral and Least squares collocation. These are complex mathematically and heavily dependent on the availability of good quality local gravity data. Strictly for the experts.
- ☞ National geoid models such as AUSGEOID98. Geoscience Australia makes available a national geoid solution based on a combination of EGM96 and gravity data. The quality of the model is generally quite good and it is freely available via the internet.

(c) A grid-based approach has been adopted in Australia for the distribution of a transformation model to move spatial data from AGD66 to GDA94 and vice versa. Describe the main differences between AGD66 and GDA94.

(5 marks)

AGD66 was the national geodetic datum used in Australia from 1966 until the adoption of GDA94 in January 2000 (with the exception of those jurisdictions that made the intermediate transition to AGD84). AGD66 is a regional (non-geocentric) datum whereas GDA94 is geocentric. The origin offset between AGD66 and GDA94 is about 200 metres. On the ground, this imposes an apparent shift of 200 metres in the north-easterly direction when transforming from AGD66 to GDA94.

Definition of AGD66 consists of :

- The defining parameters of the Australian National Spheroid (ANS) – *size* and *shape*
- The coordinates of the Johnston Geodetic Origin - *location*
- The orientation of the ellipsoid axes to be parallel to the Greenwich meridian, the equatorial plane and the Conventional International Origin – *orientation*

Realisation of AGD66 was achieved by :

- Measurement of a national geodetic network including distances, horizontal angles and astronomic observations
- Adjustment of the national geodetic network as it existed at that time (about 2500 stations)
- Publication of the latitude and longitude of each station included in the adjustment

Definition of GDA94 consists of

- The defining parameters of the Geodetic Reference System 1980 (GRS80) – *size* and *shape*
- The coordinates of a number of ITRF tracking sites, constrained to their ITRF92@1994.0 values – *location* (with redundancy)
- Coincidence between the coordinate axes and the Greenwich meridian, the equatorial plane and the CIO – *orientation*

Realisation of GDA94 was achieved by :

- Measurement of the AFN and ANN within the ITRF92 framework of control
- Adjustment of the AFN, ANN and subsidiary networks to obtain coordinates for ground monuments
- Publication of the coordinates for use by the general public

(d) In relation to Question 1(c), outline the rationale and the methodology behind the development of the grid-based approach to coordinate transformation.

(5 marks)

As a datum, AGD66 contains a significant amount of distortion, amounting to several metres in some places. GDA94 on the other hand is a much less distorted datum and thus represents a more *accurate* picture of the spatial location of features and monuments. Because of the distortion in AGD66, it was recognised that to use a conformal transformation to move data from AGD66 to GDA94 would be inappropriate. Using conformal transformation, existing distortion would be retained, thereby defeating one of the significant benefits of the move to GDA94 – *better accuracy*. In response, it was decided that a non-conformal transformation model should be used to allow the distortion to be (at least partially) eliminated as part of the transformation process.

The development of the non-conformal transformation model was along the following lines :

- Coordinates in both datums were supplied for many thousands of points across Australia. In the following discussion, these are denoted as AGD66_A and GDA94_A. The subscript “A” indicates that these coordinates were derived rigorously from the respective national or jurisdictional adjustments.
- National conformal transformation parameters were developed between AGD66 and GDA94 ($\text{C}_{66\text{-}94}$) using these two coordinate sets
- The supplied AGD66 coordinates were conformally transformed to GDA94, thereby retaining any distortion within the AGD66 values. The transformed coordinates are denoted by GDA94_T
- Computing the differences between GDA94_A and GDA94_T allowed the distortion in AGD66 at each point to be determined ($\Delta\varphi$, $\Delta\lambda$).
- These distortion values at discrete (random) was used to create a grid of the distortion
- Conformal transformation components at each grid node were algebraically added to the distortion components at each grid node to give ($\Delta\varphi_{\text{total}}$, $\Delta\lambda_{\text{total}}$)

The rationale for using a grid to distribute the transformation information was purely so as to facilitate widespread adoption of the model and uniform implementation.

2. (a) The observation equation for the GPS carrier phase observable is given below (in units of metres) :

$$\Phi = c(dt - dT) + R - d_{ion} + d_{trop} - \lambda N(t_0)$$

Give a detailed definition of each term in this equation.

(5 marks)

Φ	raw carrier phase observable (metres)
c	speed of light in a vacuum (metres/second)
dt	satellite clock error (seconds)
dT	receiver clock error (seconds)
R	geometric range between receiver and satellite (metres)
d_{ion}	ionospheric delay (negative because the ionosphere speeds up the carrier) (metres)
d_{trop}	tropospheric delay (metres)
λ	wavelength of the relevant carrier (either L1 or L2) (metres)
$N(t_0)$	integer ambiguity, the number of full cycles between receiver and satellite at lock on (cycles)

(b) Using the observation equation given in Question 2(a), develop an expression for the GPS double difference observable. In presenting this derivation, take particular care to explain and justify each step.

(10 marks)

The double difference observable is formed by combining raw carrier phase data from two satellites and two receivers recorded at the same epoch. The justification for performing differencing is that this process will reduce (and in some cases eliminate) the impact of spatially correlated errors such as receiver and satellite clock errors and atmospheric affects.

The raw carrier phase observation between *Receiver i* and *Satellite p* can be written as :

$$\Phi_i^p = c(dt^p - dT_i) + R_i^p - (d_{ion})_i^p + (d_{trop})_i^p - \lambda N(t_0)_i^p$$

Similarly, the observation between *Receiver i* and *Satellite q* can be written as :

$$\Phi_i^q = c(dt^q - dT_i) + R_i^q - (d_{ion})_i^q + (d_{trop})_i^q - \lambda N(t_0)_i^q$$

The between-satellite single difference at *Receiver i* can now be formed by differencing these two equations :

$$\nabla\Phi_i^{pq} = c\nabla dt^{pq} + \nabla R_i^{pq} - (\nabla d_{ion})_i^{pq} + (\nabla d_{trop})_i^{pq} - \lambda\nabla N(t_0)_i^{pq}$$

Note that the clock error at *Receiver i* has been eliminated in this process and the ionospheric and tropospheric errors have been reduced.

We can write a similar equation for the between-satellite single difference at *Receiver j* :

$$\nabla\Phi_j^{pq} = c\nabla dt^{pq} + \nabla R_j^{pq} - (\nabla d_{ion})_j^{pq} + (\nabla d_{trop})_j^{pq} - \lambda\nabla N(t_0)_j^{pq}$$

The double difference is formed by taking the difference between the two between-satellite single differences, one at *Receiver i* and the other at *Receiver j* :

$$\Delta\nabla\Phi_{ij}^{pq} = \Delta\nabla R_{ij}^{pq} - (\Delta\nabla d_{ion})_{ij}^{pq} + (\Delta\nabla d_{trop})_{ij}^{pq} - \lambda\Delta\nabla N(t_0)_{ij}^{pq}$$

Forming the double difference has eliminated the satellite clock errors and further reduced the impact of tropospheric and ionospheric delays.

(c) What advantages does the double difference observable have over raw carrier phase observations?

(5 marks)

Raw carrier phase observations are impacted upon by a number of sources of error, some of which can be very difficult to model mathematically. As mentioned above, the double difference observable offers a means of removing or minimising the influence of those errors which are spatially correlated. As a result, satellite and receiver clock errors do not impact on the double difference observable and the impact of ionospheric and tropospheric refraction is reduced for short baselines. In fact for baselines <15 km in length the impact of the ionosphere and troposphere can often be ignored without significantly biasing the results for the ambiguity parameter and the baseline components.

Thus the main advantage of the double difference observable is the reduction in the number of errors that need to be accounted for. On the negative side, this observable will be more noisy that the raw carrier phase observation due the impact of propagation of variances.

3. (a) In describing the basic structure of the Global Positioning System, it is usual to subdivide the system it into three parts – the space segment, the control segment and the user segment. Describe each segment and their inter-relationships.

(10 marks)

Space segment : The space segment consists of the satellites that make up the GPS constellation and the information that those satellites transmit. The GPS constellation is made up of (at least) 24 satellites placed in six near circular orbital planes. The orbits are separated by 60° in right ascension and are inclined at 55° to the equatorial plane. Each orbit has nominally four satellites in it. Satellites complete their orbits in 11h 58m, meaning that the satellite constellation repeats its geometry every 23h 56m – i.e four minutes earlier each day. The satellites transmit two L-band carrier waves (L1 and L2) and two pseudo-random-noise (PRN) codes (C/A-code and P-code). The C/A-code is modulated onto the L1 carrier and the P-code is modulated onto both L1 and L2 carriers. The L1 carrier has a nominal wavelength of 0.19 metres and the L2 carrier wavelength is 0.24 metres. Wavelength of the C/A-code is about 300 metres and the P-code wavelength is about 3 metres. The C/A-code repeats itself every millisecond (300 km) and the P-code repeats itself every 7 days (on a per satellite basis). The C/A-code and P-code are unique to each satellite, allowing for unique satellite identification and real-time range measurements. Finally, the satellites transmit a 50 bps data stream known as the Satellite Message, which provides essential information such as broadcast ephemeris parameters, clock correction parameters, satellite health, almanac, ionospheric parameters and other information.

Control segment : The control segment provides operational control and maintenance of the GPS satellite constellation, as well as the maintenance of GPS system time. The control segment is made up of five ground-based tracking stations, including the Master Control Station (MCS) at Falcon Air Force Base, Colorado Springs. The tracking stations collect GPS range and meteorological data which is transmitted to the MCS for processing. The computational process at the MCS allows for the prediction of real-time orbit parameters and satellite clock correction coefficients. This information is passed back to the uplink stations within the control segment and these stations then upload the data to the satellites. This information is provided to users in the form of the Satellite Message.

User segment : The user segment simply consists of all those individuals and organisations that rely on GPS for positioning, navigation, time transfer, meteorological monitoring and other applications. Users typically fall into one of two categories : they are either authorised users (military) or non-authorised users (civilian and others). Regardless of the application, every user must be equipped with a GPS receiver to track the signals being transmitted by the GPS satellites.

(b) The C/A-code transmitted by GPS satellites gives rise to the code pseudo-range observable. Develop the pseudo-range observation equation. Show how the pseudo-range observation equation relates to the various Dilution of Precision (DOP) factors. (10 marks)

By definition, the code pseudo-range is the time difference between the satellite clock at the time of signal transmission (t) and the receiver clock at the time of reception (T). In mathematical form, this is given by :

$$\tau = T - t$$

However, both the satellite clock and the receiver clock will be offset from GPS system time. We denote the satellite clock offset by dt and the receiver clock offset by dT. Further, the GPS system time at the time of transmission as t_{GPS} and at the time of reception as T_{GPS}, then we can write :

$$t = t_{GPS} - dt$$
$$T = T_{GPS} - dT$$

Thus the pseudo-range observation equation can be re-written as :

$$\begin{aligned}\tau &= (T_{GPS} - dT) - (t_{GPS} - dt)\\ &= (T_{GPS} - t_{GPS}) - (dT - dt)\\ &= \tau_{GPS} - (dT - dt)\end{aligned}$$

Where τ_{GPS} is the true time difference between transmission time and reception time. We assume that the satellite clock error can be eliminated by using the clock correction parameters contained in the Satellite Message. Thus the equation for the pseudo-range reduces to :

$$\tau = \tau_{GPS} - dT$$

Multiplying by the speed of light (c) converts the equation from units of time to units of length, noting that $c\tau_{GPS}$ gives the geometric range (R) between the receiver and the satellite, which can subsequently be expressed as a function of the 3D geocentric cartesian coordinates of the satellite (X^S, Y^S, Z^S) and the receiver (X_R, Y_R, Z_R).

$$\begin{aligned}\rho &= R - cdT\\ &= ((X^S - X_R)^2 + (Y^S - Y_R)^2 + (Z^S - Z_R)^2)^{1/2} - cdT\end{aligned}$$

This equation is the observation equation for the code pseudo-range observable. When four or more satellites are observed, it is possible to compute a least squares solution to estimate the receiver coordinates and receiver clock offset. The equation above must be linearised and the least squares solution computed to yield estimates for the parameters. The least squares solution will take the following form :

$$\mathbf{m} = \mathbf{A}\mathbf{x} + \mathbf{v}$$

Precision of the least squares estimates will be given by :

$$\mathbf{V}_x = \mathbf{A}^T \mathbf{V}_m^{-1} \mathbf{A}$$

Which can be written as :

$$\mathbf{V}_x = \begin{bmatrix} \sigma_X^2 & \sigma_{XY} & \sigma_{XZ} & \sigma_{XT} \\ \sigma_{YX} & \sigma_Y^2 & \sigma_{YZ} & \sigma_{YT} \\ \sigma_{ZX} & \sigma_{ZY} & \sigma_Z^2 & \sigma_{ZT} \\ \sigma_{TX} & \sigma_{TY} & \sigma_{TZ} & \sigma_T^2 \end{bmatrix}$$

The various DOP factors are all derived from the $\mathbf{V}_x$ matrix. For example :

$$\text{PDOP} = (\sigma_X^2 + \sigma_Y^2 + \sigma_Z^2)^{1/2}$$

$$\text{GDOP} = (\sigma_X^2 + \sigma_Y^2 + \sigma_Z^2 + c^2\sigma_T^2)^{1/2}$$

4. (a) Give complete definitions for each of the following :

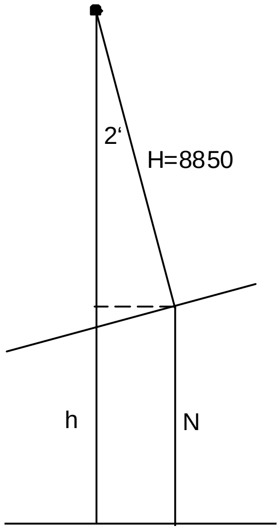
(10 marks)

- *MGA* – Map Grid of Australia. A UTM projection of latitudes and longitudes relative to the Geocentric Datum of Australia.
- *Cycle slip repair* – the process of finding and fixing cycle slips in raw GPS carrier phase observations. Cycle slips occur when the reception of a signal by a receiver from a satellite is temporarily interrupted for some reason (e.g. an obstruction).
- *Triple difference observable* – is formed from eight raw carrier phase observations involving two satellites, two receivers and two epochs. The triple difference observable is primarily used in the identification and repair of cycle slips as these show up as outliers in the triple difference solution.
- *Multipath error* – occurs when GPS signals arrive at an antenna after having been reflected off some nearby surface. Thus the range (and the phase) of the incoming signal will be incorrect. If not detected, the reflected signal can corrupt the solution for position and make ambiguity resolution difficult.
- *Broadcast ephemeris* – is the set of 16 parameters included in the Satellite Message that allows a GPS user to compute satellite location in real-time. Six of the 16 parameters define the pure Keplerian orbit, the remaining 10 parameters model perturbations to the Keplerian orbit.

(b) The summit of Mt Everest is 8850 m above MSL. The deflection of the vertical at this point is 2′. Show that even in this extreme case, the approximation inherent in the following formula is valid :

$h = H + N$

(5 marks)



Looking at the diagram, the rigorous formula for the ellipsoidal height is :

$h = (H \cos\theta) + N$

θ is the deflection of the vertical (2′ in this case). Calculating the value of $(H \cos\theta)$ and comparing it to the orthometric height H, will reveal the level of approximation :

$H \cos\theta = 8850 * 0.9999998 = 8849.999$ metres

Thus the error (approximation) in the formula is only 1 mm even in these extreme case. Such an error is certainly not of significance and so the approximation is valid.

(c) Differential GPS (DGPS) based on the C/A-code offers a means of reducing the effect of spatially correlated errors. Discuss and illustrate this statement. Outline the fundamental requirements of utilising a commercial DGPS service.

(5 marks)

There are a number of spatially correlated errors that impact on the accuracy of positioning that can be achieved with a GPS receiver operating autonomously. Examples include satellite orbit errors, receiver and satellite clock errors and atmospheric biases. In differential GPS (DGPS) the impact of the spatially correlated errors is minimised and in some cases totally eliminated. Thus only the residual (non-common) spatially correlated errors and other non-spatially correlated errors such as multipath impact on DGPS quality. It is for this reason that the pseudo-range correction method of differential GPS is so effective. The errors that impact on the pseudo-ranges at the base station can be determined because the receiver and satellite coordinates are both known. Assuming spatial correlation, the same errors are transmitted to and applied at the pseudo ranges measured at the rover receiver thereby significantly improving the quality of positioning.

A commercial DGPS service offers real-time pseudo-range correction data to DGPS users. Fundamental requirements in utilising such a service include :

- Service coverage
- Service cost
- Service reliability
- Transmission mode (radio, mobile phone, communications satellite)
- Latency
- Physical bulk and weight of receiving equipment
- Network versus single station solution

5. (a) In moving from code-based GPS positioning to the carrier phase, it is possible to make significant improvements in achievable precision. However, to maximise the potential gains, it is necessary to carefully model or eliminate all known sources of error. Of particular note in this regard is the impact of the earth's atmosphere. Give a comprehensive description of how the atmosphere interferes with GPS signals and the options that exist to overcome this source of error.

(15 marks)

The two layers of the atmosphere that impact on the transmission of GPS signals are the ionosphere and the troposphere. It is necessary to consider these two mediums separately as the nature of their interaction with the GPS signals is quite distinct.

The Ionosphere : is that layer of the atmosphere that extends from the upper level of the troposphere (about 50 km above the surface of the earth) to about 1000 km. The ionosphere is an electrically charged medium due to the presence of free electrons released through the impact of solar radiation. It is the presence of these free electrons that dictates the degree of interference that will be experienced by the GPS signals as they pass through the ionosphere. The number of free electrons is quantified by the *Total Electron Content (TEC)*. The higher the TEC count, the more ionospheric refraction will take place. From a user's perspective, ionospheric refraction varies with factors such as time of day, time of year and location. It is typically worse during the early afternoon and at equatorial latitudes. There is also a strong correlation between ionospheric interference and the 11-year solar cycle.

The ionosphere is a *dispersive* medium, meaning that the refraction that takes place is a function of the frequency of the transmitted wave. The GPS codes and carriers are affected differently by the ionosphere. In fact the codes are delayed (*group delay*) and the carriers are advanced (*phase advance*). The amount by which the codes are delayed is equal (but opposite in sign) to the amount by which the carriers are advanced (accelerated).

The ionosphere is the biggest source of error that impacts on GPS positioning quality. It is for this reason that a great deal of research has been focussed and continues to focus on strategies to model or eliminate the ionospheric error. Below is a brief list of some of the modelling options available to users :

- <u>The model of Klobuchar</u>. The satellite message contains eight parameters that may be used to correct C/A-code pseudo-range observations using the model developed by Klobuchar. The model is necessarily coarse and its usefulness is limited to single receiver positioning. At mid-latitudes, it is estimated that this approach eliminates up to 50% of the expected ionospheric error.
- <u>Differential positioning</u>. Because the ionosphere is a spatially correlated error source, performing differential positioning (either code-based or carrier) will assist in reducing the impact of the ionosphere. This is particularly true for receivers only separated by short distances.
- <u>Dual frequency observations</u>. Because the ionosphere is a dispersive medium and the refraction is frequency dependent, using dual frequency observations allows the impact of the ionosphere to be estimated and eliminated. Indeed this is the reason why GPS is a dual frequency system. Residual ionospheric errors after elimination by dual frequency observations are relatively small and can be

neglected for most practical purposes. However, in some high precision applications, further modelling may be required.

The Troposphere : is the so-called *weather zone*. It is that layer of the atmosphere that extends from the surface of the earth to a height of about 40-50 km. Unlike the ionosphere, the troposphere is a non-dispersive medium and so its impact on the GPS signals is not a function of frequency. It is for this reason that having a dual frequency system like GPS is of no real benefit in accounting for the influence of the troposphere. The GPS carriers (L1 and L2) and codes (C/A and P) are affected (slowed down) to the same degree by the troposphere. The degree of tropospheric inteference is a function of typical atmospheric parameters such as temperature, pressure and relative humidity along the wave path. The impact of the troposphere is typically broken down into two components : the wet part accounts for about 10% of the delay and the dry part makes up the remaining 90%. The dry component is relatively easy to account for by standard modelling techniques and the models of Hopfield and Saastamoinen are commonly used for this purpose using a standard atmosphere. The wet component is ignored in most processing software, but can be and is modelled as an additional parameter in more sophisticated GPS processing packages. Finally it should be noted that the troposphere delay can be largely eliminated by relative positioning over short baselines and under the condition that the two receivers are at similar heights. If this latter condition is not satisfied, even over a short line, quite different tropospheric conditions can prevail and will not be eliminated during the measurement differencing process.

(b) In an era well before the advent of GPS, somebody said “time is of the essence”. This quote could aptly be applied to GPS. Explain the role of time in the operation of GPS for positioning and navigation purposes.

(5 marks)

The whole operation of GPS for navigation and positioning is explicitly dependent on the maintenance and measurement of accurate time. In the first instance, the GPS Control Segment is responsible for keeping a consistent atomic time standard known as GPS System Time. The satellites likewise carry atomic clocks which are accurate to 10^{-13} seconds per day and which are synchronised to GPS System Time both physically (by periodic clock adjustments) and mathematically by the computation and dissemination through the Satellite Message of the parameters of a second-order polynomial clock correction model. The satellite clocks are the source of the signals transmitted by the GPS satellites. The codes and carrier phase signals all emanate from the satellite clocks and are therefore intrinsically a function of time. Finally, every GPS receiver contains a clock. Though typically not of the same quality as the satellite clocks, the clock contained in a GPS receiver is the basis of the measurement process, whether measuring codes and/or carriers. Thus pseudo-range and carrier phase observations are in fact measures of time. Time inaccuracies (either satellite or receiver based) result in range errors which propagate into position errors. It is for this reason that time is indeed “of the essence”.