9

Troposphere, Ionosphere, and Tide Corrections

9.1 Contributions to Phase

Most scientists who construct interferograms are trying to measure surface deformation associated with the solid Earth or cryosphere. However, as discussed in Chapter 5, there are many contributions to the phase of an interferogram that may dominate the deformation signal. As a review, the contributions to the phase are:

$$\begin{aligned} \textit{phase} = {} & \textit{earth curvature} \text{ (almost a plane, known)} + \\ & \textit{topographic phase} \text{ (broad spectrum, known)} + \\ & \textit{surface deformation} \text{ (broad spectrum, unknown)} + \\ & \textit{orbit error} \text{ (almost a plane, largely known)} + \\ & \textbf{troposphere delay} \text{ (power law, unknown)} + \\ & \textbf{ionosphere advance} \text{ (often a plane or 20–40 km wavelength waves, unknown)} + \\ & \textbf{tides} \text{ (almost a plane, largely known)} + \\ & \textit{phase noise} \text{ (white spectrum, unknown)} \end{aligned} \tag{9.1}$$

In Chapter 5 we developed the geometry and methods to remove the contributions from the *curvature of the Earth* and *topography* using a precise orbit and a digital elevation model. These topographic phase contributions increase as the interferometric baseline increases. Modern synthetic aperture radar (SAR) missions have well-controlled interferometric baselines so given the accurate, near-global topography available today, these phase contributions can be accurately removed. The last contribution on the list, *phase noise*, is largely random in space. In Chapter 6 we discussed filtering and stacking approaches to reduce the phase noise. This chapter is focused on the three remaining contributions to the phase, including **tropospheric delay**, **ionospheric advance**, and **tides**.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 19 Nov 2026 at 17:36:53, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009606226.010

9.1.1 Quiz – Where Is the Tectonic Signal?

Before moving into the details of what causes these unwanted phase contributions and how they can be removed or mitigated, we start with a quiz with 5 Sentinel-1 C-band interferograms crossing the San Andreas Fault and the Los Angeles basin (Figure 9.1). They all show interesting phase patterns. The question is, which interferogram shows the tectonic signal? At C-band, one fringe of phase corresponds to 28 mm delay or surface deformation away from the radar. We know from global navigation satellite system (GNSS) studies that the total deformation rate across this fault system is ~30 mm/yr so in 1 year we would expect to see ~10 mm of motion when projected into the line of sight (LOS) of the radar. This corresponds to about one third of one fringe per year. Since each of these interferograms has multiple fringes, the answer to the quiz is that none of the interferograms show any significant tectonic signal.

We can speculate on what causes the phase in each interferogram:

(a) This case has the smallest number of fringes amounting to a phase ramp of about –60 mm from the southwest to the northeast (i.e., the range is decreasing toward the northeast). We could attribute this to orbit error since it is almost a plane, but Sentinel-1 has orbit accuracy usually less than 30 mm so this could be a phase delay due to atmospheric pressure changes in the dry part of the troposphere or the ionosphere.
(b) This case has many wave-like signals with amplitudes of 10–20 mm. These are most likely due to gravity waves in the troposphere.
(c) This case has many fringes that correlate with the topography (4–5 fringes or up to 140 mm). These signals are probably due to a change in the water vapor in the atmosphere at levels less than the highest topography. This layered water vapor effect is discussed later.
(d) This case has a very flat phase except in the mountain areas inside the red box. The phase associated with the highest topography is decorrelated. This could be due to snow or ground freezing at the highest elevation. In addition, this case shows a lower phase (~28 mm) in the Los Angeles basin that could be due to differences in the layered water vapor.
(e) The final case has multiple fringes that are not correlated with topography and have mainly large spatial scales. This could be due to changes in the total electron content (TEC) in the ionosphere, which affect the speed of the radar waves. The satellite orbiting at 700 km is just above the maximum electron density in the ionosphere so small spatial scales in TEC will be smoothed during the construction of the synthetic aperture (~5 km).

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 19 Nov 2026 at 17:36:53, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009606226.010

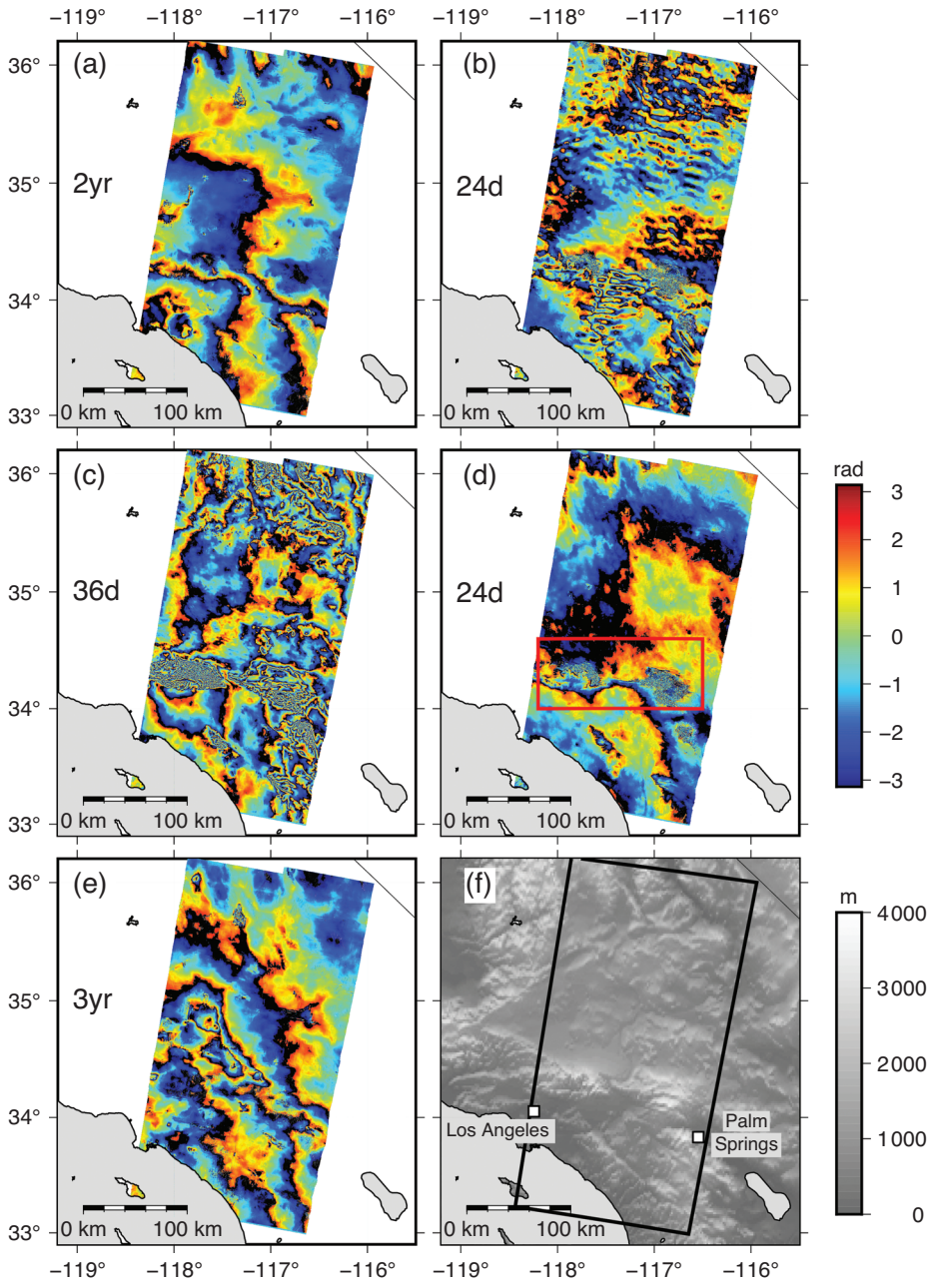


Figure 9.1 Example interferograms from Sentinel-1 wide-swath data across the Mojave Desert to the northeast and the Los Angeles basin to the southwest. The San Andreas Fault system crosses this track in a northwest direction passing through Palm Springs. Red box marks the location of the high mountains of Southern California.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 19 Nov 2026 at 17:36:53, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009606226.010

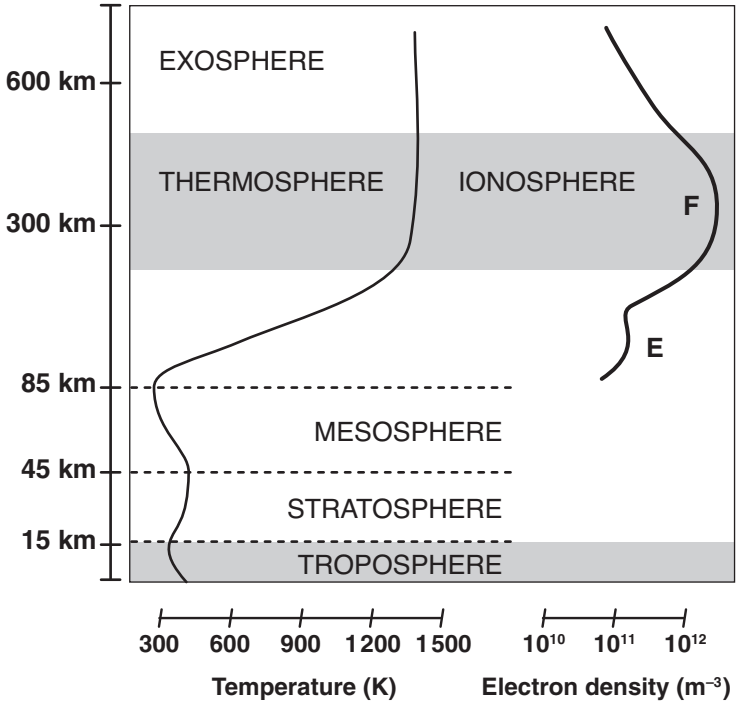


Figure 9.2 Layers of the atmosphere and ionosphere. The main layers affecting the propagation of microwaves (gray boxes) are the troposphere and the F-layer of the ionosphere. Adapted from Liang et al., 2019. Reprinted with permission from *IEEE Transactions on Geoscience and Remote Sensing*.

In the remainder of this chapter we will discuss the basic physics and properties of the troposphere and ionosphere (Figure 9.2) as well as large-scale surface deformation due to solid Earth tides. In addition, we discuss ways to correct for, or at least mitigate, the unwanted noise sources.

9.2 Refractive Index of the Troposphere and Ionosphere

So far, we have assumed that the electromagnetic waves travel at the speed of light c as they propagate from the antenna to the ground and back. However, the speed varies because of the variations in pressure, temperature, and water vapor in the troposphere as well as the free electrons in the ionosphere as shown in Figure 9.3. The relevant parameter is the refractive index n, which is the ratio of the speed of light to the phase velocity.

$$v_p = \frac{c}{n}. \tag{9.2}$$

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 19 Nov 2026 at 17:36:53, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009606226.010

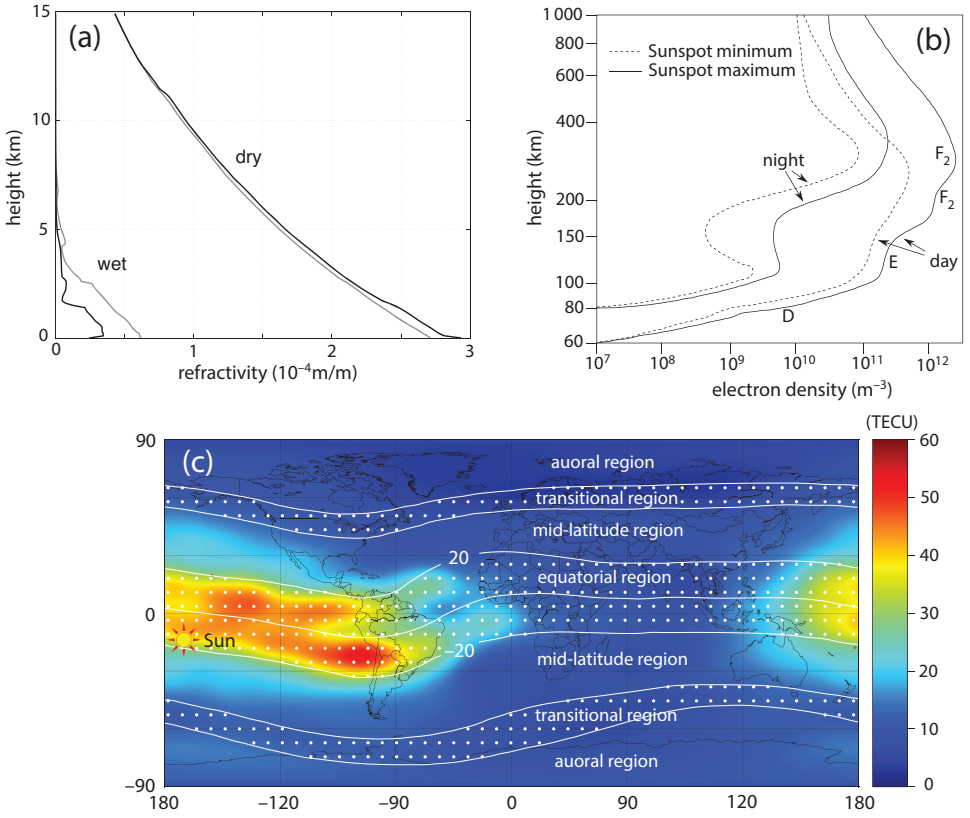


Figure 9.3 (a) The troposphere of the Earth extends from the surface to an elevation of about 15 km. Dry and wet refractivity versus height for two scenarios illustrates temporal variations (Pany, 2003). (b) Typical number density of free electrons in the D, E, and F layers of the ionosphere versus height for day and night, as well as solar maximum and minimum (from Liang et al., 2019) (c) Data from over 100 continuously operating GPS receivers in a global network are being used to produce global maps of the ionosphere's TEC (https://iono.jpl.nasa.gov/) (from Liang et al., 2019). One TEC Unit (TECU) is 10^{16} electrons m^{-2}. These global ionosphere maps (GIM) provide instantaneous "snapshots" of the global TEC distribution that can be used for a first-order correction for the largest spatial scale variations in the ionosphere. Reprinted with permission from *IEEE Transactions on Geoscience and Remote Sensing*.

The refractive index in a vacuum is 1. We are interested in the delay of the radar wave along a one-way vertical path between the surface of the Earth z_o and the radar at an altitude of H. The travel time increase is

$$\Delta t = \int_{z_o}^{H} \frac{1}{v_p\,(z)} dz - \int_{z_o}^{H} \frac{1}{c} dz = \frac{1}{c} \int_{z_o}^{H} [n\,(z) - 1]\, dz. \tag{9.3}$$

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 19 Nov 2026 at 17:36:53, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009606226.010

This corresponds to an excess path increase L of

$$L = \int_{z_o}^{H} [n(z) - 1]\, dz. \tag{9.4}$$

Given this excess path in the vertical, and assuming the lateral variation in the refractivity is smooth, one can compute the excess phase in the interferogram as a two-way phase increase of

$$\Delta\phi = L\frac{4\pi}{\lambda \cos\theta} \tag{9.5}$$

where θ is the local incidence angle of the radar. Incidence angle is used rather than look angle (see Section 5.4) because most of the troposphere is close to the ground rather than the satellite. We will use look angle for the ionosphere correction because the F-layer of the ionosphere is closer to the satellite.

Since the refractive index is usually slightly greater than 1, it is customary to introduce the refractivity as $N = 10^6 (n - 1)$. The refractivity of the troposphere and ionosphere is given by the following equation with four terms (Bevis et al., 1994; Doin et al., 2009)

$$N = k_1\frac{P_d}{T} + k_2\frac{P_v}{T} + k_3\frac{P_v}{T^2} - \alpha\frac{N_e}{f^2} \tag{9.6}$$

where
P_d – partial pressure of dry air (Pa)
P_v – partial pressure of water vapor (Pa)
T – absolute temperature (K)
N_e – number density of free electrons (m^{-3})
f – radar frequency (Hz)
$k_1 = 0.776$ (K Pa^{-1})
$k_2 = 0.716$ (K Pa^{-1})
$k_3 = 3.75 \times 10^3$ (K^2 Pa^{-1})
$\alpha = 40.28$ (Hz^2 m^{-3}).

The first term corresponds to the slowing of the waves due to the dry component of the troposphere and the combined second and third terms represent the slowing of the wave due to the water vapor (wet) component of the troposphere. The last term represents the increase in the phase velocity of the wave as it travels through the ionosphere containing free electrons as discussed in Section 9.4.1. Figure 9.3 (a) shows examples of the relative contributions of the wet and dry terms to the refractivity (Pany, 2003).

Our objective is to correct the two-way travel time (i.e., phase) along the non-vertical LOS path of the radar for the effects of pressure, temperature, and water

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 19 Nov 2026 at 17:36:53, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009606226.010

vapor in the neutral troposphere as well as the electron density in the ionosphere. For interferometry, we seek the change in the phase between the reference and repeat SAR images and how it varies spatially over the area of the interferogram.

The path-integrated tropospheric delays require a 3-D, time-dependent model of the troposphere. In Section 9.3 we discuss corrections using numerical weather models and/or GNSS measurements of the troposphere. These models/measurements are generally sparse and inaccurate, so in Section 9.3.1, we discuss a variety of approximate methods to correct for the tropospheric phase delay, including GNSS time series at known locations and common scene stacking, assuming the temporal variations in the surface deformation are slower than the cadence of the SAR acquisitions.

The ionospheric effects and correction methods are discussed in Section 9.4. The phase and group velocity of radar waves depend on the integrated electron density along the path in the ionosphere; this is called the total electron content (TEC) (Figure 9.3(c)). As seen in Equation 9.6, and derived in Section 9.4.1, the velocity perturbation is proportional to the inverse of the radar frequency squared. Therefore, the path delays are large for a low-frequency L-band radar, moderate for an intermediate frequency C-band radar, and small for a higher frequency X-band radar. Moreover, because the ionosphere is dispersive, the path delay can be accurately corrected if the radar has two widely separated frequencies. GNSS satellites broadcast timing signals at two frequencies. The receiver on the ground measures the travel-time difference between the two signals to estimate the TEC along the ray path and uses that to correct the travel time of the higher frequency signal to achieve a subcentimeter correction accuracy. The split-spectrum approach can be used to correct SAR interferograms for the ionospheric distortions, although current systems do not have sufficient frequency bandwidth to provide accurate corrections (Meyer et al. (2006); Liang et al. (2019)). The dual-frequency approach will be more accurate with the next generation of dual-frequency radars such as NISAR and ALOS-4.

9.3 Characteristics and Mitigation of the Troposphere Delay

Here we consider the zenith excess path L due to just the neutral atmosphere (mostly troposphere). The satellite altitude is well above the top of the atmosphere so the integral Equation 9.4 can extend to infinity. The zenith path delay is given by Doin et al. (2009)

$$L = 10^{-6} \int_{z_0}^{\infty} [k_1 \frac{(P_d + P_v)}{T} + (k_2 - k_1) \frac{P_v}{T} + k_3 \frac{P_v}{T^2}] dz \tag{9.7}$$

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 19 Nov 2026 at 17:36:53, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009606226.010

where the total pressure $P = (P_d + P_v)$ has been combined in the first term assuming hydrostatic equilibrium. This first term can be integrated analytically using $dP = -\rho g dz$ where g is the acceleration of gravity averaged over the troposphere (see Problem 9.1). The density is found using the ideal gas law such that

$$\rho = \frac{P_d}{R_d T} + \frac{P_v}{R_v T} = \frac{P}{R_d T} + \left(\frac{1}{R_d} - \frac{1}{R_v}\right)\frac{P_v}{T} \tag{9.8}$$

where R_d (287.05 J kg^{-1}K^{-1}) and R_v (461.495 J kg^{-1}K^{-1}) are the specific gas constants for dry air and water vapor, respectively.

After integration, one finds

$$10^6 L = \frac{k_1 R_d}{g} P(z_o) + \int_{z_0}^{\infty} \left[\left(k_2 - \frac{R_d}{R_v} k_1\right)\frac{P_v}{T} + k_3 \frac{P_v}{T^2}\right] dz \tag{9.9}$$

where $P(z_o)$ is the surface pressure. This equation has two important terms. The first term on the right is called the *zenith hydrostatic delay* (ZHD). Note it only depends on the measurement of the pressure at the surface at an elevation of z_o. The second term on the right is called the *zenith wet delay* (ZWD) and it depends on the integration of water vapor over the height of the troposphere. Measuring the ZWD is much more challenging because it requires a 3-D model of the pressure of the water vapor. Moreover, for InSAR, the integral should be performed along the nonvertical ray path of the radar, which could be different from the ZWD mapped using the cosine of the look angle (Equation 9.5) if there are significant horizontal variations in water vapor with lengths scales less than about 5 km.

At this point it is important to remember that an interferogram measures the phase or path difference between two acquisition times. Therefore, we only care about the difference in the ZHD and ZWD between the two times. For example, consider the first term, ZHD, in Equation 9.9. The long-term mean absolute pressure at a given position/elevation on the surface of the Earth cancels in the difference. The GNSS community places atmospheric pressure sensors at the sites so the temporal variations can be measured. However, for InSAR, spatially dense surface pressure data are not available everywhere so one needs a weather model, or surface interpolation among the GNSS sites, to estimate ZHD over the entire SAR image. Similarly, there are no spatially dense sensors to measure the vertically integrated water vapor, although this can be done at GNSS sites since the path delays to multiple GPS satellites, having many incidence angles, can be analyzed simultaneously. The conclusion is that a 3-D model of pressure, temperature, and water vapor is needed to correct the interferogram. This can be obtained from high-resolution weather models combined with sparse GPS measurements, although there are significant errors in these models (Yu et al., 2018).

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 19 Nov 2026 at 17:36:53, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009606226.010

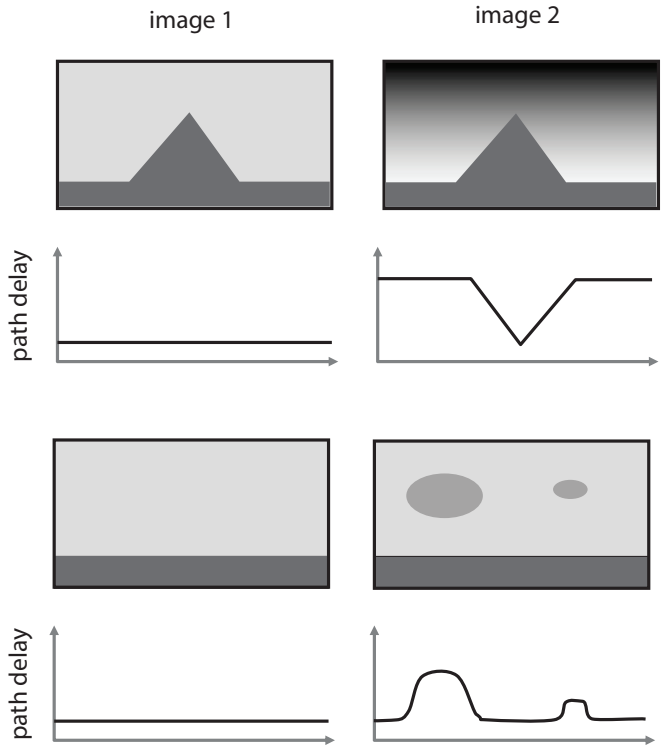


Figure 9.4 The top row illustrates the change in the atmospheric delay of image 2 with respect to image 1 for a stratified troposphere. The second row illustrates the corresponding path delay seen in the interferogram. The delay is proportional to the local topography and is commonly assumed to be linear with topography. The third row illustrates the change in atmospheric delay due to random variations (turbulence and waves) in mostly water vapor. The fourth row illustrates the path delay seen in the interferogram.

The practical approach to correcting interferograms for tropospheric delay is to first decompose the delays (refractivity) into changes caused by stratification and changes caused by turbulence (Hanssen, 2001).

$$N(\mathbf{x}, z) = \bar{N}(z) + \delta N(\mathbf{x}, z) \tag{9.10}$$

where $\bar{N}(z)$ is the vertical stratification assumed uniform across the image and $\delta N(\mathbf{x}, z)$ is the 3-D deviation from the average stratification. Figure 9.4 provides a schematic of these two terms.

The most straightforward method of correcting for both types of delays is to use a 3-D atmospheric weather model that provides the surface pressure $P[z_o(\mathbf{x})]$ needed for the ZHD as well as the partial pressure of water vapor $P_v(\mathbf{x}, z)$ needed

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 19 Nov 2026 at 17:36:53, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009606226.010

for the ZWD. These models must have adequate spatial and temporal resolution to perform the correction.

There are numerous studies on the spatial and temporal characteristics of the troposphere and the effects on InSAR phase (e.g., Hanssen (2001); Emardson et al. (2003); Doin et al. (2009); Bekaert et al. (2015); Yu et al. (2018)). This is an active area of research, and a full discussion is beyond the scope of this book. However, there are several important observations that inform the analysis of InSAR time series as well as the mitigation schemes. In particular, Emardson et al. (2003) examined the spatial and temporal characteristics of the atmospheric delay using continuous GNSS data to provide an overview of the issues. Their study was based on 126 GNSS stations in Southern California spanning the period January 1998 to March 2000. Over a time span of 106 seconds (12 days) or shorter, the surface deformation from tectonic or hydrological processes is assumed to be zero. The relative locations of the GNSS stations in this array are known to millimeter accuracy. The ionospheric delays are corrected using the dual-frequency approach discussed in Section 9.4 and tidal deformation is corrected using accurate models. Therefore, the position differences between any two stations primarily reflect variations in the delay of the EM waves as they traverse the neutral atmosphere (mostly troposphere). The delays include both ZHD and ZWD.

The characteristics of the atmospheric delay for each panel in Figure 9.5 are:

(a) This panel shows the variation in the atmospheric delay as a function of time offset for two spatial baselines of 20 km and 244 km (thin line). There are two important conclusions from this analysis. (1) The structure function (variance) flattens for time increments greater than one day. Most repeat-pass InSAR systems have a cadence of greater than 6 days so one can assume that the atmospheric effects of all the interferograms in a large analysis are temporally uncorrelated. (2) If one plans to use a 3-D weather model for correcting individual interferograms, ideally the model output should occur at the same times as the SAR acquisitions. However, weather models such as ECMWF provide outputs only every 6 hours. A typical mismatch in the sampling times will lead to large sampling errors. A sampling rate of 5 minutes would essentially eliminate the error due to non-coincidence sampling. In some regions, the GACOS model correction scheme uses available continuous GNSS data to perform a temporal interpolation of the poorly sampled weather models.

(b) This panel illustrates the issues related to the spatial characteristics of the tropospheric delay. There are three important conclusions. (1) The atmospheric variance, as measured by pairs of GNSS receivers, increases roughly as the square root of the distance between the receivers. For example, two points separated by 1 km will have a relative atmospheric error of less than 8 mm. (2) In

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 19 Nov 2026 at 17:36:53, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009606226.010

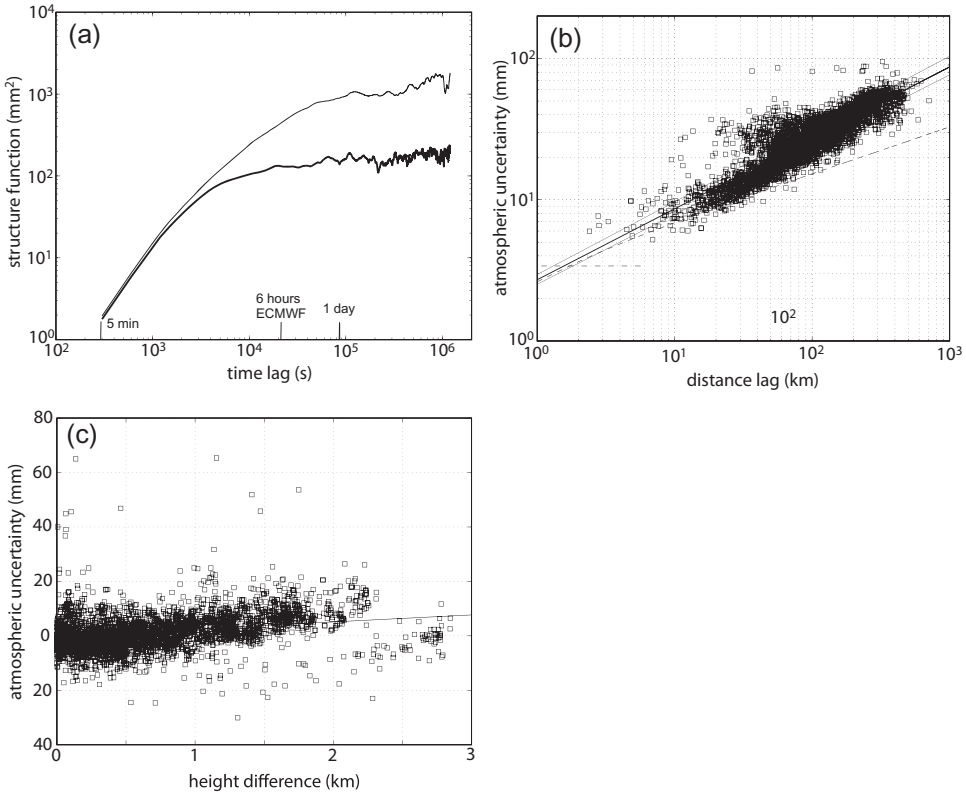


Figure 9.5 Temporal and spatial characteristics of the atmospheric delay from (Emardson et al., 2003). The details of each panel are discussed in the text.

contrast, at length scales of 100 km, the variance due to the atmospheric error is commonly 30–50 mm. Thus, to achieve say 10 mm accuracy over length scales greater than 100 km could require averaging interferograms from many SAR images (perhaps 16). (3) In a region that has a high density of continuously operating GNSS stations, one can dramatically improve the accuracy of InSAR time series by correcting each interferogram using the large-scale atmospheric estimated from the GNSS array (e.g., Neely et al., 2019). For example, the typical GNSS spacing is 20 km in Western North America so point-wide GNSS delays should be used to optimize the high spatial resolution provided by the InSAR time series.

(c) This panel illustrates the effects of the stratified tropospheric delay ZHD on the relative accuracy of two GPS measurements at differing altitudes. This effect is highly dependent on the dates of the measurements, so there is significant scatter in the uncertainty versus height difference, yet we will see next that this effect can dominate the tropospheric error and should be corrected if possible.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 19 Nov 2026 at 17:36:53, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009606226.010

9.3.1 Mitigation of the Tropospheric Delay

There are three basic approaches to mitigation of the tropospheric delay.

(1) Stacking – The first approach is to assume that the deformation of the surface is steady with time and so one averages images to reduce the atmospheric noise. As shown in Figure 9.5(a), the atmospheric delay decorrelates after a time interval greater than about one day. Most InSAR systems have repeat intervals of 6 days or longer so simply averaging (stacking) many interferograms, each scaled by their temporal baseline, will reduce the error by the square root of the number of independent SAR images in the stack. This approach works well under the steady deformation rate assumption. However, many GNSS and InSAR time series display seasonal deformation signals due to annual variations in hydrology – either elastic deformation from water loading or compaction and swelling of unconfined aquifers. In the cases where the seasonal signals are large, the temporal averaging should be confined to a single season, or a multiyear average of interferograms from SAR images within the same season, assuming the seasonal amplitude is uniform.
(2) Atmospheric modeling – The second approach is to obtain the best possible atmospheric delay model using a combination of global or local atmospheric weather models along with zenith tropospheric delay measurements at GNSS sites. For example, the generic atmospheric correction online service (GACOS) correction facility (Yu et al., 2018) "integrates operational high-resolution European Centre for Medium-Range Weather Forecasts (ECMWF) data (0.125° grid, 137 vertical levels, 6-hour interval) and continuous GNSS tropospheric delay estimates (every 5 min in some regions) using an iterative tropospheric decomposition model." Their approach uses an exponential function to model the stratified delays, and then the turbulent part of the delay is constructed from interpolation of the residual path difference between the sparse GNSS sites and the ECMWF model. An example of the correction applied to a 12-day Sentinel-1 interferogram from the UK is shown in Figure 9.6. The raw interferogram (d) has strong atmospheric delay errors with a standard deviation of 27.5 mm. The correction has a component from the change in the stratified troposphere (a) and a larger scale, mostly trend from the water vapor and pressure across the scene (b). These two corrections (c) reduce the rms to 6.9 mm. The approach works best when there is a high density of GNSS stations to provide the high temporal sampling (5 minutes) as well as adequate spatial sampling. In areas of lower GNSS density or large-amplitude atmospheric perturbations, the atmospheric correction is often less effective.
(3) Parametric Modeling – The third approach is to use the InSAR data along with a digital elevation model to first estimate and remove the tropospheric signal

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 19 Nov 2026 at 17:36:53, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009606226.010

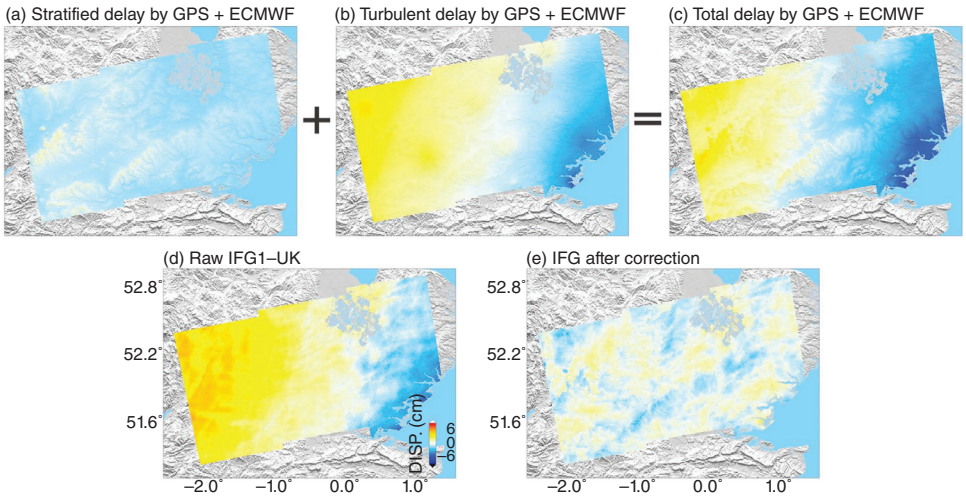


Figure 9.6 Sentinel-1, 12-day interferogram of a region in the UK has large atmospheric noise modeled by the GACOS method (from Yu et al. (2018)). (a) Model of the LOS displacement due to changes in the stratified component of the delay are correlated with topography. (b) Model of LOS displacement due to the turbulent component of the delay has a prominent linear trend with an amplitude of 40 mm. (c) Total delay is applied to the original interferogram (d) to form the corrected interferogram (e).

due to changes in atmospheric stratification and then use a method such as common scene stacking to "filter" the atmospheric signals over a prescribed timescale where the surface deformations are presumed to be steady.

An example of the effects of the changes in atmospheric stratification on an interferogram for the Southern California area is shown in Figure 9.7. This stratification error often dominates the tropospheric signal and should be modeled and removed from all the interferograms used in a stack or time series. Prior to estimating the linear regression between residual phase and elevation, one should remove the overall trend in the residual phase since this may be related to residual orbit error or other larger-scale errors such as ionosphere delay or a planar pressure variation. This linear fit with height is often very effective but can fail when the deformation signals are correlated with topography or the relationship between phase and height is not constant throughout the interferogram. After correcting for the delay from the changes in the stratified component, a second estimation method must be used to remove the turbulent component. One can use a common scene stacking/time series approach to remove the remaining small-scale residual noise as discussed in Chapter 10.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 19 Nov 2026 at 17:36:53, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009606226.010

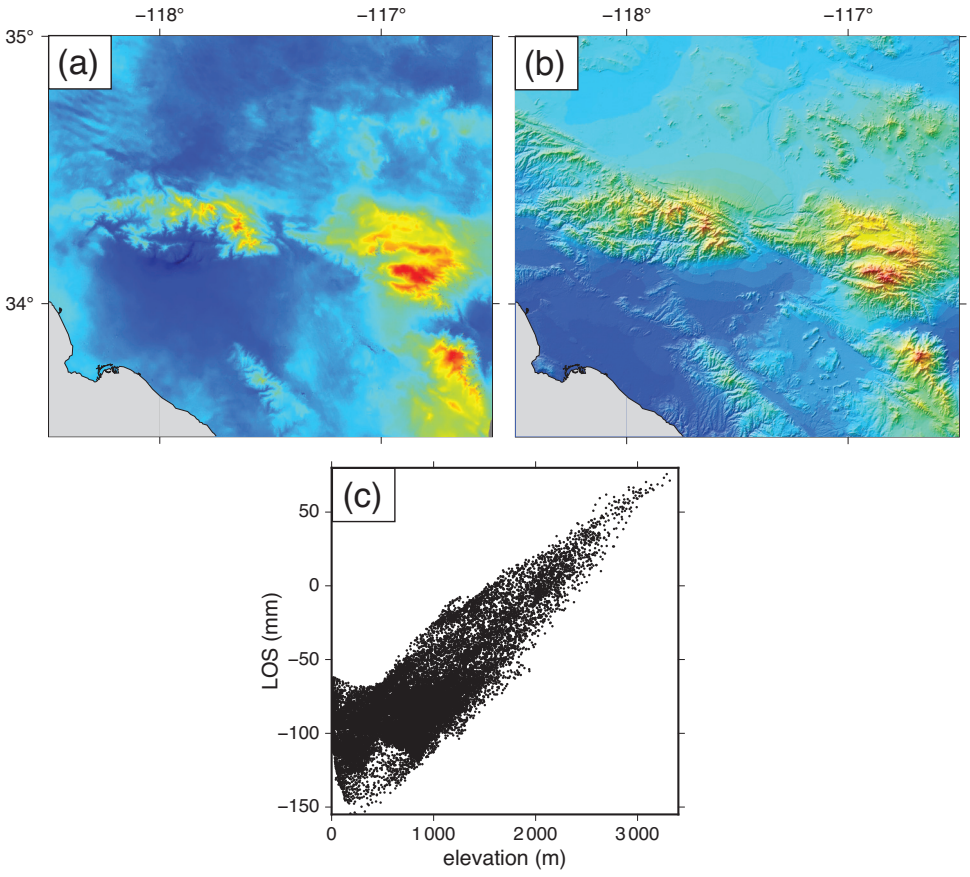


Figure 9.7 Example of correlation between interferometric phase (a) and elevation (SRTM DEM) (b) due to temporal changes in atmospheric stratification. This is the same 36-day Sentinel-1 TOPS mode interferogram shown in Figure 9.1(c). A plot of LOS versus elevation shows a high correlation with 140 mm of phase delay due to the stratified troposphere draped on the topography of 3 000 m.

In summary, atmospheric delay is a dominant source of error in the InSAR technique. Typical errors are 20 – 50 mm and can be as large as 100 mm in extreme cases of atmospheric fronts. The GNSS community removes most of this error because the stations are fixed in space, they have a high temporal sampling rate, and they observe the LOS atmospheric distortions from a multitude of satellite overpasses, and thus can develop a highly accurate map of the atmospheric water vapor for the ZWD. In addition, pressure sensors at each site provide accurate measurement of the ZHD. InSAR has much poorer temporal sampling and only one LOS look through the troposphere, so separating the ZHD from a small surface deformation signal (10 mm or less) is problematic. Moreover, since the surface pressure

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 19 Nov 2026 at 17:36:53, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009606226.010

is only known at a few widely spaced GNSS sites, changes in the stratified component of the lower troposphere can dominate the deformation signal, especially in areas of high topographic relief. This topographically correlated delay is especially problematic in areas where the volcanic or fault deformation is correlated with topography.

9.4 Characteristics and Mitigation of the Ionosphere Advance

The ionosphere is a region of the Earth's upper atmosphere, extending from approximately 50–1 000 km in altitude (Figure 9.3(b)). Ultraviolet radiation from the sun ionizes atoms and molecules of oxygen and nitrogen, resulting in a plasma of free electrons. The ionosphere is characterized by its dynamic nature, exhibiting variations with altitude, latitude, time of day, and phase of the 11-year solar cycle. Understanding and mapping these complex variations is essential for assessing and mitigating its impact on microwave geodetic systems, including GNSS and InSAR.

Altitude is a primary factor in ionosphere structure (Figure 9.3(b)). The ionosphere is conventionally divided into several layers, including the D, E, and F layers (Figure 9.3(b)). The D region, closest to the Earth's surface, is most prominent during the daytime and absorbs/reflects lower-frequency radio waves. The E region is prominent during daylight hours and reflects medium-frequency radio waves. The F region, with sublayers F1 and F2, has the highest electron density between 200 and 500 km altitude and delays and distorts the propagation of microwaves. SAR satellites have typical orbital altitudes of 500–800 km and thus travel within the upper part of the F-region. In Section 9.4.1, we develop the theory for the phase velocity of an EM wave as it propagates through a plasma. The most important parameter for microwave signal propagation is the integral of the density of free electrons along the ray path from the satellite to the ground; this is called the total electron content and has units of electrons per area. Therefore, the radial variations in ionospheric properties are not as important as the temporal and spatial variations in TEC.

Time of day significantly affects the ionosphere's behavior (Figure 9.3(c)). During the day, solar radiation increases ionization and thus TEC, while at local sunset the electrons recombine with their associated atom/molecule, and there is a dramatic and sometimes turbulent decrease in density that affects microwave signals. The TEC is minimum just before sunrise.

Latitudinal variations (geomagnetic latitude) introduce additional complexity to the ionosphere's behavior (Figure 9.2(c)). At low latitudes (0°–20°), ionization is highest due to maximum solar illumination. There are local maxima on both sides of the geomagnetic equator between 10° and 15°. The ionosphere is quietest at mid-latitudes (20°–60°), with low spatial variations except for occasional geomagnetic

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 19 Nov 2026 at 17:36:53, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009606226.010

storms. At high latitudes (>60°), the ionosphere is strongly affected by the converging lines of the geomagnetic field and charged particles from the solar wind entering the ionosphere. There are many dynamic processes that delay and distort the GNSS and InSAR signals at high latitudes so mitigating these effects is important even at C-band (Liang et al., 2019). Ionospheric disturbances often occur as wave-like perturbations in the electron density and thus TEC within the area of an interferogram.

The 11-year solar cycle, characterized by fluctuations in solar activity, significantly influences the ionosphere. During periods of high solar activity (Figure 9.3(b), sunspot maximum), increased ionization results in higher electron density and greater signal delays and distortions. Conversely, solar minimum has lower ionization and weaker signal delay.

9.4.1 Propagation of EM Waves through a Plasma

The theory of the propagation of electromagnetic (EM) waves in a plasma, such as the Earth's ionosphere, is well developed (e.g., Jackson (2012)). First, consider an EM wave traveling in the z-direction having a single x-component of electric field of strength E_o.

$$E_x(z,t) = E_o e^{i(\omega t - kz)} \tag{9.11}$$

where $\omega = 2\pi/T$, T – is the wave period, $k = 2\pi/\lambda$, and λ – is the wavelength. Note the frequency in radians per second is related to the frequency in Hz by $\omega = 2\pi f$.

The phase of this wave has a constant value when $\omega t = kz$, so the velocity of the wave, corresponding to constant phase, is $v_p = \frac{z}{t} = \frac{\omega}{k}$. EM waves traveling in a medium free of charges and currents obey the wave equation

$$\frac{\partial^2 E_x}{\partial z^2} - \frac{\varepsilon}{c^2}\frac{\partial^2 E_x}{\partial t^2} = 0 \tag{9.12}$$

where c is the speed of light and ε is the relative permittivity, also called the dielectric constant. Note by definition, $c = (\varepsilon_o \mu_o)^{-1/2}$ where $\varepsilon_o = 8.854 \times 10^{-12}$ F m^{-1} is the permittivity of free space and $\mu_o = 4\pi \times 10^{-7}$ H m^{-1} is the magnetic permeability of free space. Equation 9.11 is a solution to the wave Equation 9.12 so by taking the derivatives, one finds:

$$(-ik)^2 e^{i(\omega t - kz)} - \frac{\varepsilon}{c^2}(i\omega)^2 e^{i(\omega t - kz)} = 0 \quad \text{or} \quad k^2 + \frac{\varepsilon}{c^2}\omega^2 = 0 \tag{9.13}$$

Therefore, the phase velocity of the wave is $v_p = \frac{\omega}{k} = c\varepsilon^{-1/2}$. Since the phase velocity is also defined as $v_p = cn^{-1}$, we have the usual relationship between the index of refraction and the dielectric constant $n = \varepsilon^{1/2}$.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 19 Nov 2026 at 17:36:53, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009606226.010

Next, consider this wave propagating through a low-density electron plasma (e.g., ionosphere). The plasma consists of free electrons of density N_e (electrons m^{-3}) having charge $e = 1.602 \times 10^{-19}$ Coulombs and mass $m = 9.109 \times 10^{-31}$ kg and an equal number of positively charged molecules of oxygen/nitrogen, and so on. The positive ions are too massive to move in response to the variations in the electric field. The x-component of the differential equation for the forces acting on the free electron is:

$$m\ddot{x} + m\omega^2 x = -eE_x\,(x,t) \tag{9.14}$$

In the high-frequency limit, the solution to this equation has a natural resonant frequency (Jackson, 2012) given by:

$$\omega_p^2 = \frac{N_e e^2}{\varepsilon_o m}. \tag{9.15}$$

For a typical electron density of the F-layer of the ionosphere of 10^{12} electrons m^{-3}, the plasma frequency is 9 MHz. The lowest frequency L-band SAR satellites operate at 1.3 GHz so the high frequency approximation $\omega \gg \omega_p$ is valid. Under this approximation, the dielectric constant and index of refraction are:

$$\varepsilon = n^2 = 1 - \frac{\omega_p^2}{\omega^2}. \tag{9.16}$$

The phase velocity of the wave is:

$$v_p = c\left(1 - \frac{\omega_p^2}{\omega^2}\right)^{-1/2}. \tag{9.17}$$

Note the phase velocity exceeds the speed of light since the denominator is always less than 1. However, this does not violate the law that no information can travel faster than c, because a single frequency wave does not carry any information. Information can only be carried by a wave group having a finite band of frequencies. The group velocity of a wave is given by $v_g = \frac{\partial \omega}{\partial k}$. It is left as an exercise (Problem 9.2) to show that the group velocity of a plasma, in this high-frequency limit, is:

$$v_g = c\left(1 - \frac{\omega_p^2}{\omega^2}\right)^{1/2}. \tag{9.18}$$

Note the product of the phase and group velocities is c^2. It is interesting to note that the ionospheric effects on both GNSS and InSAR methods involve both the phase velocity $> c$ and the group velocity $< c$. In the case of GNSS, the phase of the carrier wave between the satellite and receiver on the ground is advanced after propagating through the ionosphere with respect to propagation through a vacuum.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 19 Nov 2026 at 17:36:53, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009606226.010

In contrast, the range signal that carries the timing information is delayed after propagating through the ionosphere. Similarly, for InSAR, the phase difference between a reference and repeat interferogram will be advanced by an increase in the electron density of the ionosphere between the reference and repeat times. However, the radar backscatter signal, which consists of multiple wavelengths, will be delayed by an increase in electron density between the reference and repeat times. This dispersive nature of the ionosphere leads to an elegant way to correct for both the phase advance and range delay.

Now we have almost all the tools needed to calculate these advances and delays. The last term in Equation 9.6 for the refractivity of the ionosphere depends on the electron density and inversely on the frequency squared:

$$N_{ion} = -\alpha \frac{N_e}{f^2}. \tag{9.19}$$

Note there is a constant α that is still undefined. Our equation for the index of refraction has a square root, so it does not match the form of Equation 9.19

$$n = \left(1 - \frac{\omega_p^2}{\omega^2}\right)^{1/2}. \tag{9.20}$$

However, since we are in the limit where the wave frequency is much greater than the plasma frequency, we can use the binomial expansion to rewrite Equation 9.20 as:

$$n \cong 1 - \frac{\omega_p^2}{2\omega^2} = 1 - \frac{e^2 N_e}{2\varepsilon_o m \omega^2}. \tag{9.21}$$

Therefore, the refractivity of the ionosphere in terms of frequency in Hz is

$$N_{ion} = -\frac{e^2}{8\pi^2 \varepsilon_o m} \frac{N_e}{f^2}. \tag{9.22}$$

We see that the constant α in Equation 9.6 is based on well-known physical parameters

$$\alpha = \frac{e^2}{8\pi^2 \varepsilon_o m} = 40.28 \text{ Hz}^2\text{m}^3. \tag{9.23}$$

Now using Equation 9.4, we can calculate the 1-way phase advance due to the ionosphere

$$L = \int_{z_o}^{H} [n(z) - 1]\, dz = -\frac{\alpha}{f^2} \int_{z_o}^{H} N_e(z) dz = -\frac{\alpha}{f^2} TEC. \tag{9.24}$$

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 19 Nov 2026 at 17:36:53, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009606226.010

Table 9.1 *Typical ionosphere phase advance.*

Frequency GHz	Wavelength (mm)	Band	Two-way phase advance (m)
1	300	P	20
1.3	230	L	12
5	60	C	0.8
10	30	X	0.2
13	23	Ku	0.1

The integral on the right side of the equation of the electron density over the path is the TEC, which has units of electrons per unit area. It is interesting to calculate the two-way phase advance for a typical TEC of 25 TECU or 2.5×10^{17} electrons m^{-2} as provided in Table 9.1.

Of course, this theory also provides the range delay of a microwave pulse as it travels through the ionosphere. The group velocity is $v_g = c^2/v_p$, so the range delay is the negative of phase advance.

9.4.2 Ionospheric Corrections for GNSS and InSAR

Precise GNSS positioning systems exploit the dispersive nature of the ionosphere to correct for the phase advance and range delay of the signals. GPS broadcasts its signals at two frequencies, f_1 (1.575 GHz) and f_2 (1.227 GHz). Consider a pseudo range measurement between the satellite and the ground receiver $P = c\Delta t$ where Δt is the time of flight of the pulse as measured by the receiver. After a bit of algebra (Problem 9.3), one can show that the range correction is given by:

$$\Delta\rho = \frac{f_2^2}{\left(f_1^2 - f_2^2\right)}\,(P_1 - P_2). \tag{9.25}$$

where $\Delta\rho$ is the excess range to be subtracted from pseudo range P_1. Similarly, one should add $\Delta\rho$ to correct to the advance of the phase of the carrier at frequency f_1. Since one can only count cycles of wavelengths of distance between the satellite and the receiver, the $\Delta\rho$ is used to correct for integer cycle ambiguities of the carrier phase.

Note that the pseudo range difference in Equation 9.25 is scaled by the f_2 carrier frequency squared but more importantly also scaled by the inverse of the difference of the two carrier frequencies squared. Thus an error in the pseudo range measurements will be amplified by this scale factor. In the case of GPS, the overall scale factor is 2.54, which is not large. However, this amplification of the error would be

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 19 Nov 2026 at 17:36:53, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009606226.010

larger if the two frequencies were closer together. Next, we will discuss the same method for correcting the phase advance of an interferogram due to the ionosphere.

To date there are no dual-frequency InSAR satellites, so the main approach to ionospheric correction is to artificially split the SAR range spectrum into two subbands (e.g., Meyer et al. (2006); Gomba et al. (2015); Liang et al. (2019)). Recall, the slant-range resolution of a SAR is equal to the inverse of the radar bandwidth as described in Chapter 4. A typical bandwidth for an L-band satellite such as ALOS-2 is 28 MHz, the typical bandwidth for a C-band satellite such as Sentinel-1 TOPS-mode is 43 MHz, and the bandwidth for an X-band satellite such as TerraSAR-X is sometimes very high 150/300 MHz.

The original SLC or raw signal data is band-pass filtered to create two SLC images having subbands with a small separation as shown in Figure 9.8. The center frequency of the lower subband is f_l and the upper subband is f_u. Then one creates two interferograms from two SAR images. The phase of the low interferogram $\Delta\phi_l$ is generated from the two low SLC images and the phase of the upper interferogram $\Delta\phi_u$ is generated from the two high SLC images using the same approach of

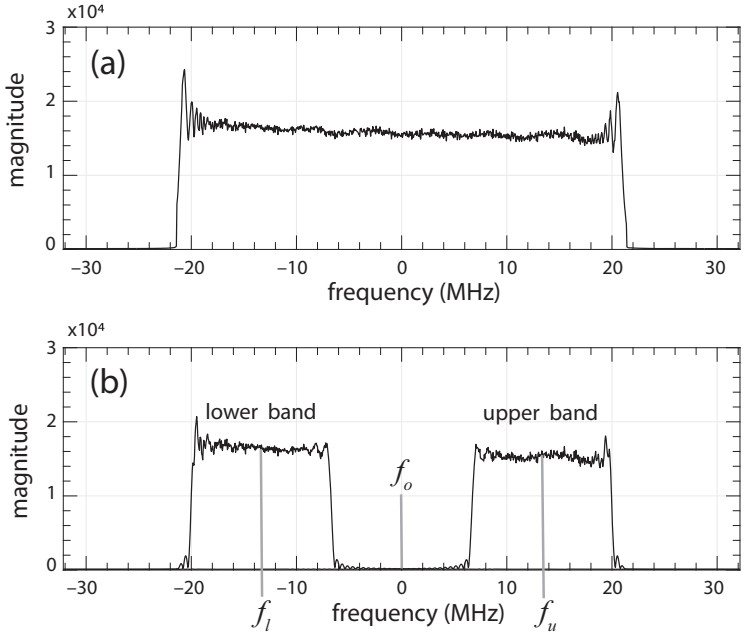


Figure 9.8 (a) Amplitude spectrum (flattened) of the radar signal of the Sentinel-1 TOPS mode has a bandwidth of 43 MHz and a center frequency of 5.405 GHz. Bandpass filtering is used to split the range signal into two subbands at frequencies of +/–14 MHz with respect to the carrier frequency. The accuracy of the split spectrum approach is best when each subband is 1/3 of the total bandwidth (Brcic et al., 2011).

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 19 Nov 2026 at 17:36:53, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009606226.010

complex multiplication as the standard interferogram (Equations 5.2 and 5.3). To calculate the phase advance, each interferogram must be unwrapped and filtered and the 2π phase ambiguity must be accurately estimated (Figure 9.9). The formula for the ionospheric phase correction to be applied to the original interferogram is:

$$\Delta\phi_{ion} = \frac{f_l f_u}{f_o\left(f_u^2 - f_l^2\right)}\Delta\phi_l f_u - \Delta\phi_u f_l \tag{9.26}$$

where f_o is the carrier frequency.

Similar to the calculation of the noise amplification factor for the GPS case, we find that an error in the noise of the difference interferogram in Equation 9.26 will be amplified by 96.5 times in the case of the Sentinel-1 data having the 43 MHz bandwidth (Figure 9.8). This means that any small amount of noise will dominate the correction, so the interferograms must be heavily filtered (Liang et al. (2019); Xu and Sandwell (2019)). Indeed, because the noise is amplified by nearly 100 times, the InSAR correlation must be nearly perfect, or the correction will add more noise than the ionospheric advance signal. NISAR will overcome this noise amplification issue by using a radar having a true split spectrum of 72.5 MHz. For the NISAR case at L-band, with a carrier frequency of 1 300 MHz, the noise amplification factor will be only 4.77, which is 20 times better than the split spectrum approach applied to Sentinel-1 data and only 1.9 times worse than the GPS dual-frequency approach.

The ionosphere has two additional effects on radar interferometry that are more fully discussed in other publications (e.g., Gray et al. (2000)). First, any along-track gradients in the TEC will cause an azimuth shift of the pixels between the reference and repeat images. The amount of azimuth shift can be estimated by first performing a geometric co-registration of the images using the precise orbits. Then one can perform a patch-wise cross-correlation between the images similar to the pixel tracking approach discussed in Chapter 10. Note that the satellite is orbiting within and just above the F-layer of the ionosphere so if the TEC gradients have a shorter length scale than the length of the synthetic aperture (Table 4.3), the resulting image will be defocused in the azimuth direction. The third effect is related to polarized SAR images where the polarization vector is rotated with respect to the original polarization by Faraday rotation (Gomba et al., 2015).

9.5 Tides and Aliasing

Corrections for the surface deformation due to the solid earth tide are an essential step for precise GNSS analysis because tidal amplitudes are large (~150 mm) and the tide models are extremely accurate (~1 mm) (McCarthy et al., 2004). Similarly, one should correct interferograms for the LOS displacement change between

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 19 Nov 2026 at 17:36:53, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009606226.010

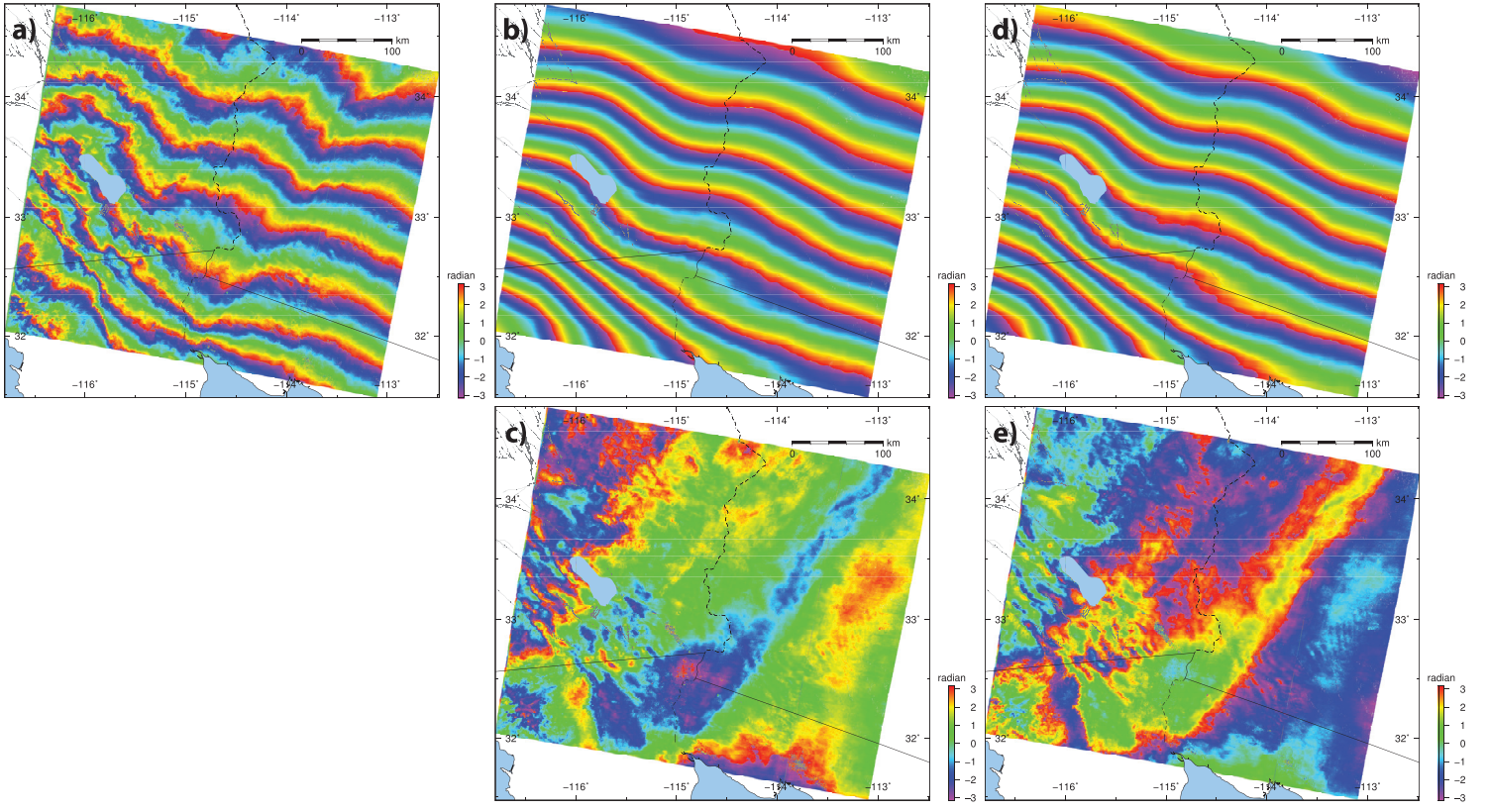


Figure 9.9 Split-spectrum ionospheric correction applied to an ALOS-2 ScanSAR interferogram (from Xu and Sandwell (2019)). (a) Original interferogram. (b) Ionospheric phase estimated using the split-spectrum method. (c) Ionospheric phase-corrected interferogram. (d) Estimated ionosphere correction if unwrapping ambiguity exists in both band-passed interferograms. (e) Applying the wrong correction from (d) to (a) results in a phase shift. Reprinted with permission from *IEEE Transactions on Geoscience and Remote Sensing*.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 19 Nov 2026 at 17:36:53, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009606226.010

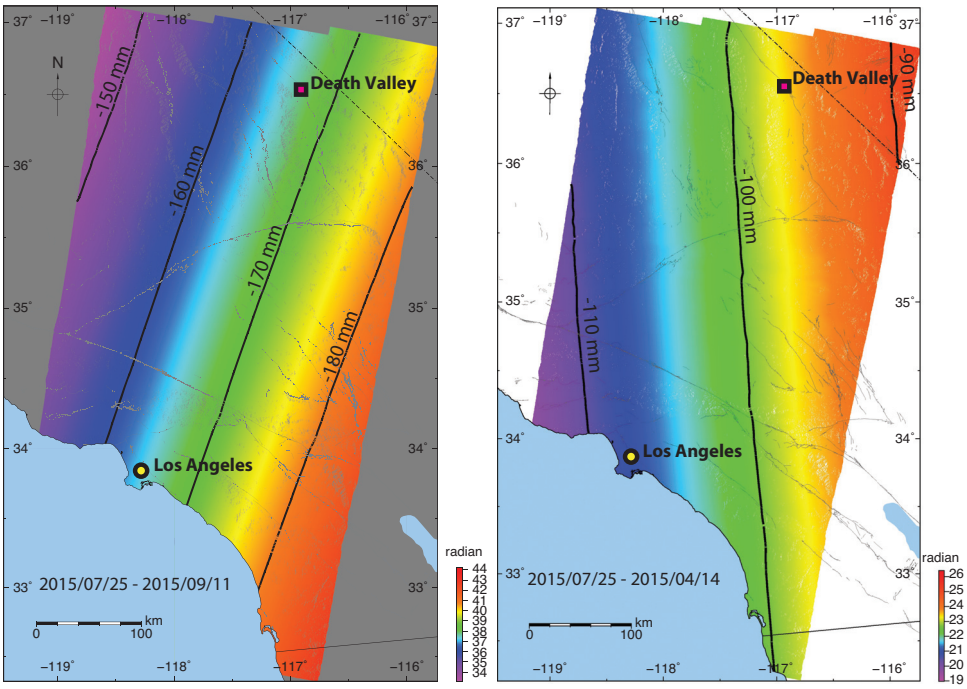


Figure 9.10 Solid earth tide phase/displacement for two Sentinel-1 TOPS interferograms (July 25, 2015–September 11, 2015 and July 25, 2015–April 14, 2015). Reprinted with permission from *IEEE Transactions on Geoscience and Remote Sensing*.

the reference and repeat images due to the solid earth tide (*solid_tide* in GMTSAR). Most InSAR studies ignore the tidal deformation signal because it is largely a constant over distances of 100 km. However, there are two important enhancements provided by the new generation of SAR satellites that make the tide correction important. First, the new satellites have greater swath widths and lengths of 250 km or greater, so the tide signal is not uniform in the swath. Second, the newer satellites have a regular acquisition cadence, so accurate time series relies on improved corrections.

Figure 9.10 illustrates how the mostly large spatial scale of the solid earth tide produces large-scale tilt signals in interferograms in both range and azimuth. The trend in the range is mostly due to the change in incidence angle across the swath. Consider Sentinel-1 TOPS-mode interferograms. For a maximum tide amplitude of 180 mm, the LOS component will be 156 mm in the near range (29°) and 124 mm in the far range (46°) resulting in a 32-mm trend across the interferogram in range. The tilt in azimuth depends on the change in latitude within a scene. As an example, the trend between Long Beach, CA, and Death Valley, CA (~300 km N-S

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 19 Nov 2026 at 17:36:53, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009606226.010

Table 9.2 *Aliasing of solid earth tide by InSAR satellites*

	repeat (days)	M2	S2	K1	O1	P1
amplitude (mm)		384.83	179.05	191.78	158.11	70.88
ERS/Envisat	35	94.50	–	365.23	75.06	365.23
RADARSAT	24	64.07	–	365.23	77.70	365.23
ALOS-1	46	398.63	–	365.23	190.60	365.23
ALOS-2	14	270.08	–	365.23	1036.73	365.23
Sentinel-1A/B	12	64.07	–	365.23	77.04	365.23
NISAR	12	64.07	–	365.23	77.04	365.23
Sentinel-1A&B	6	14.77	–	365.23	14.19	365.23

distance), has a peak-to-trough amplitude of ~20 mm. These trends in tide in both the range and latitude directions will go directly into the unwrapped interferogram, so they should be removed. Note the effects are identical at X, C, and L-band in terms of LOS displacement.

The more important issue is tidal aliasing. The sun-synchronous orbits of the current generation of SAR satellites cause the diurnal and semidiurnal tides to be aliased into much longer periods. For example, the semidiurnal solar tide S2 has a period of one day and an amplitude of 180 mm (Table 9.2). This tide component is aliased into zero frequency when sampled at an exact integer number of days. In general, the alias frequency f_a due to sampling a tide of frequency f_t at a frequency f_s is given by the following formula:

$$f_a = f_t - f_s \text{round}\left(\frac{f_t}{f_s}\right) \tag{9.27}$$

For example, the M2 tide, with a period of 0.517542 days, when sampled at an interval of 12 days, has an alias period of 64.07 days (Table 9.2). The next largest tidal component K1 has an alias period of 365.23 days when sampled at an integer number of days. The overall result is that the 12-day sampled tide has alias periods of 64.07, 365.23, and 77.04 days (Table 9.2). The aliases at 64 and 77 days could be filtered out with smoothness in an small baseline subset methods SBAS analysis. However, the aliases having a 1-year period could be confused with real deformation signals associated with seasonal variations in water loading. Since the solid earth tide can be computed accurately, it should be removed. Indeed, GMT now has an earthtide tool to compute deformation from the solid earth tide.

Ocean tides, which deform the crust significantly within about 100 km of a coastline (Agnew, 2010), will have similar aliasing properties. However, these tides are less predictable, and the loading models depend on the local elastic properties of the

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 19 Nov 2026 at 17:36:53, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009606226.010

crust and mantle, so accurate corrections are more challenging – although probably important for very accurate InSAR time series analysis in coastal areas.

9.6 Problems

1. Integrate the first term in Equation 9.7 using the hydrostatic approximation and the first term in the ideal gas law (Equation 9.8) to obtain the first term in Equation 9.9 that relates the ZHD to the surface pressure.
2. Use the definition of the group velocity of a wave and the dispersion formula for the phase velocity of a wave propagating in a plasma (Equation 9.17) to derive the formula for the group velocity of a wave propagating in a plasma (Equation 9.18). Show that the product of the phase velocity and the group velocity is the speed of light squared.
3. Use the formula for the one-way path delay of a radar pulse traveling through the ionosphere (Equation 9.24) to derive the range correction for a dual-frequency GPS measurement given the pseudorange at two frequencies (Equation 9.25).

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 19 Nov 2026 at 17:36:53, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009606226.010