G

GAMMA AND X-RADIATION

Enrico Costa and Fabio Muleri
Istituto di Astrofisica e Planetologia Spaziali, INAF, Rome, Italy

Synonyms

Gamma radiation; Gamma rays; High-energy photons; High-energy radiation; X-radiation; X-rays; γ-rays

Definition

X-radiation. Electromagnetic radiation with wavelength λ shorter than ultraviolet light but longer than γ-rays, $0.01\ \text{nm} \leq \lambda \leq 10\ \text{nm}$.
Gamma radiation. Electromagnetic radiation with energy higher (or, equivalently, shorter wavelength/higher frequency) than X-radiation.

Introduction

X-rays and gamma rays are the shorter wavelength or, equivalently, the higher-energy side of the electromagnetic spectrum. The separation of these two bands is quite conventional: X-rays are considered to have an energy from that of ultraviolet light to the binding energy of inner electronic shells of atoms, whereas above this value there are gamma rays which, broadly speaking, are produced by nuclear activity. For the sake of simplicity, we name as X-radiation those electromagnetic waves with wavelength λ between 10 and 0.01 nm, while gamma rays are assumed to have a value of λ below 0.01 nm. A common way of describing gamma and X-radiation is by the energy expressed in units of electron volt (eV), where 1 eV is the energy acquired by an electron which passes through a difference of potential of 1 V. With this definition, X-rays have energies between about 100 eV and 100 keV (1 keV = 1,000 eV). The relation between the wavelength in nm and the energy of photons expressed in eV is given by:

$$E = \frac{1239.84}{\lambda_{nm}}\ \text{eV}. \tag{1}$$

X-rays were discovered by Wilhelm Roentgen in 1895 as a new kind of emission obtained by the interaction of accelerated cathode rays, that is, electrons, with a target (Shamos, 1987). Their characteristics were a relatively long range and the propagation in straight lines, even in the presence of a magnetic field, which suggested to rule out the hypothesis that X-rays were charged particles. However, it was not earlier than 1912 that the real nature of X-rays as short wavelength electromagnetic radiation came out. Max von Laue showed that X-rays are diffracted by crystal lattices thus proving that X-rays are actually electromagnetic radiation with wavelength comparable to lattice spacing. Gamma rays were discovered by Paul Villard in 1900, who recognized the presence, beyond the α-rays (helium nuclei) and β-rays (electrons), of a component unaffected by magnetic field and scarcely absorbed in the emission of radioactive materials (Gerward, 1999). Also in this case the definite proof that gamma rays are photons with short wavelength came from their diffraction on crystal lattice.

Interaction of X-rays and gamma rays with matter

The short wavelength makes the behavior of gamma and X-radiation much similar to that of discrete particles. Notwithstanding, there is a fundamental difference in the interaction with matter, that is, photons are destroyed when they interact whereas charged particles lose energy continuously.

There are three main processes through which gamma and X photons interact with matter and the probability of occurring of each depends on the energy of the radiation.

E.G. Njoku (ed.), *Encyclopedia of Remote Sensing*, DOI 10.1007/978-0-387-36699-9,

In the low-energy part of the X-ray band (soft X-rays), the prevailing interaction is with electrons in inner shells of atoms via the photoelectric effect (or photoabsorption). An electron (named photoelectron) is extracted from the atom with a kinetic energy equal to the difference between the energy of the photon and the binding energy of the electron. Only interactions with electrons bound with an energy lower than that of the photon are possible, and therefore, inner electronic shells are being involved only for increasing photon energy. The contribution of the most internal shell is always predominant with respect to the outer ones. The probability of interaction with a certain shell, that is, the cross section of the process σ, is a steep decreasing function of energy, $\sigma \sim E^{-3.5}$. This dependency is characterized by an abrupt increase of σ as soon as the energy is sufficient to extract photoelectrons from a more inner shell (*absorption edge*). The probability of photoelectric absorption has a strong dependence on the atomic number Z ($\sigma \sim Z^5$) and thus, heavy elements absorb photons with a much more efficiency. The ejection of the photoelectron leaves the ionized atom in an excited status with a vacancy in an inner shell. It can decay via two major processes: the ejection of a fluorescence photon or the ejection of a self-ionization electron (named Auger electron). In both cases, the total energy released is just below the binding energy of the photoelectron. The detection of fluorescence emission is a powerful tool for determining the composition of materials because each element produces radiation with a characteristic pattern, that is, only at those energies which correspond to transitions between its own electronic shells. Therefore, by measuring the energy and the intensity of fluorescence photons, it is possible to identify the atomic elements and their abundance without destroying the sample under test.

At energies higher than the binding energy of most inner electrons (belonging to the K-shell), the probability of photoabsorption decreases quickly. When the energy is tens of times the K-shell energy, namely, between a few tens to a few hundreds of kilo electron volts depending on the atomic number of the material, the prevailing interaction of photons becomes the Compton effect. In this case, the photon is scattered off an electron and the former transfers a part of its energy to the latter. Even if electrons in materials are bound in atoms, for what concern the process of interaction they can be regarded as to be free, because the binding energy is in any case negligible with respect to the energy of the photon. By the conservation of the energy and of the momentum, the energy of the scattered photon E' can be calculated as a function of the scattering angle θ from the energy of the incident photon E:

$$E' = \frac{E}{1 + \frac{E}{mc^2}(1 - \cos\theta)}. \tag{2}$$

The energy acquired by the electron is the difference between E and E'. The probability of scattering and the angular distribution of photons are described by the Klein-Nishina formula:

$$\frac{d\sigma}{d\Omega} = \frac{1}{4}\left(\frac{e}{mc^2}\right)^2 \frac{E'^2}{E^2}\left(\frac{E}{E'} + \frac{E'}{E} - \sin^2\theta\right). \tag{3}$$

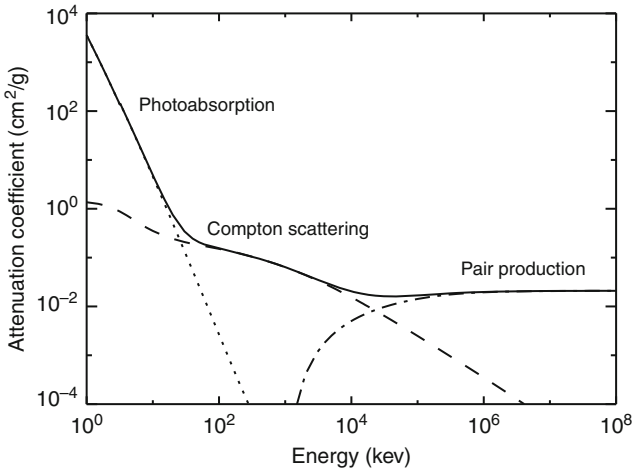
Another process of interaction between photons and matter is possible, that is, the pair production. If the photon has an energy higher than two times the rest mass of electrons (>1.022 MeV, 1 MeV = 1,000 keV), the photon can split into an electron-positron pair. A third body, usually the field of a nucleus but also that of an electron, must be present to allow for the conservation of both energy and momentum, and it can absorb a part of the initial photon energy. Above some tens of MeV, pair production is the most probable interaction and the cross section becomes basically independent from energy.

Interaction of X-rays and gamma rays with Earth's atmosphere

Earth's atmosphere strongly absorbs high-energy radiation, especially in the soft X-ray range. The fraction of photons which pass through it without interacting, that is, the transparency T, as a function of energy is calculated by using the attenuation coefficient μ, whose value can be found in the literature (see, e.g., the "XCOM: Photon Cross Sections Database," Berger et al.). If the depth of the medium is x and ρ is its density, we have:

$$T = \exp(-\mu\rho x). \tag{4}$$

In Figure 1, we report the attenuation coefficient for the Earth atmosphere as a function of energy. The absorption



Gamma and X-Radiation, Figure 1 Attenuation coefficient as a function of energy for Earth's atmosphere. The contributions of photoabsorption (*dotted line*), of Compton scattering (*dashed line*), and of pair production (*dash-dot line*) are distinguished. It is assumed a composition of N_2 (78.1 %), O_2 (20.9 %), and Ar (0.01) (NASA's Earth Fact Sheet). The value of the attenuation coefficient is retrieved from XCOM database (Berger et al.).

is dominated by photoelectric effect, Compton scattering, or pair production at different energies, but its value remains significant all through the spectrum. For example, the attenuation length, defined as the distance at which the transmission is reduced to $1/e$, is 0.23 cm at 1 keV, and the transparency of 1 cm is about 1 % (the density of the atmosphere is assumed to be 1.2×10^{-3} g/cm^3). At higher energies, the absorption is less effective but even at the minimum of the attenuation coefficient around 40 MeV, the attenuation length is only a few hundreds of meters. Moreover, in the range of maximum transparency, the prevailing interaction is the Compton scattering. The radiation interacting with atmosphere is scattered forward with typical angles around 30° and with an energy significantly lower than that of the original photon. Therefore, even if a significant number of photons can survive a thickness of a few kilometers, any spectral or imaging information is lost. As a consequence remote sensing is not an effective tool in these energy ranges for the study of the surface of our planet. This is true for the Earth but not necessarily for other planets: Any celestial body, satellite, or minor body, without an atmosphere, is a potential target for remote sensing in X-/gamma rays – actually a few such studies of remote sensing have been performed and are planned for future space missions.

Detectors for X-/gamma rays

Modern detectors allow to measure X- and gamma radiation with a fine sensitivity (Leo, 1994; Lutz, 2007; Knoll, 2010). Spectroscopy, at the level of 1 eV, and imaging, with a spatial resolution up to 1 μm, are the more developed fields, particularly in the soft X-ray band, but also polarimetry is progressing. X- and gamma rays can be detected only after they have interacted with matter via one of the processes described above, which destroys the initial photon and results in the transfer of its energy to one or more electrons. These electrons propagate in the matter while being scattered by interactions with atomic nuclei and slowed down by interactions with atomic electrons. The energy lost along the path ionizes atoms, producing other electrons, and excites electrons of external shells, which shortly afterward decay and emit UV photons. Electrons extracted by ionization propagate at short distance because they have a very low energy, and hence, the result of the interaction of the photon is a small cloud of ion-electron pairs close to the interaction point and a small flash of UV photons.

We can classify detectors in the following three groups:

Detectors that collect the charges produced by ionizing electrons

One way to detect high-energy photons is to measure the charge deposited in the sensitive volume of the instrument by the electrons produced in consequence of the initial interaction. These charges are produced near the point where the photon interacted and are proportional, in number, to the energy of the photon. Moreover, they are produced along the path of initial electrons, which contains information on the polarization of the photon. For example, the average energy necessary to the electrons to produce other electron/ion pairs in a gas by ionization is 20–25 eV, and therefore, the energy of a photon of 10 keV is converted in around 400 electrons. An electric field is applied in the sensitive volume of the instrument to drift the electron/ion pairs toward an anode/cathode which collects them. To overcome the noise of read-out electronics, often the pairs are multiplied before collection (of a factor spanning from 10^3 to 10^6 and above) in a region with an intense electric field, which accelerates the electrons so that they can ionize other atoms.

The *proportional counter* was the first device developed for detecting high-energy radiation, capable to measure the energy of individual photons, and it can illustrate very well the functioning of detectors that collect the charges produced by ionizing electrons. The sensitive volume is a chamber filled with a gas, where photons are absorbed and electrons are drifted by an electric field to an anode wire (15–50 μm thick) at high voltage. The field close to the wire is very high and accelerates the electrons that collide with atoms producing more electrons. The original cloud of electron is amplified of a factor 10^3–10^5 and is collected by the anode wire, while the ions are collected by other wires or by the box of the detector acting as a cathode. This is the basic function of proportional counters. With the use of many wires, they can act as imaging devices with a resolution down to 0.1 mm. The energy resolution is of the order of 20 % at 6 keV and scales with the square root of the energy.

Semiconductor detectors are another type of instrument which measures the charge produced as the result of the photon interaction. Instead of gas, the sensitive volume is a junction of a crystalline semiconductor (mostly silicon and germanium) inversely polarized. When an ionization cloud is produced inside the depletion layer of the junction, the electrons and holes are immediately collected at the electrodes. Since the energy to produce an electron/hole pair in silicon is 3.4 eV, photons of the same energy produce more charges in a semiconductor than in a gas, and this results in an enhancement of the energy resolution. Another advantage is that solid-state instruments are more compact with respect to gas detectors of comparable efficiency.

Semiconductor detectors are usually used without charge multiplication and hence need very low-noise electronics. This requires a low leakage current, a low capacity of the detector, and an operation at low temperature. Since the depleted layer is the actual sensitive part, it is enhanced by building junctions on two sides of an intrinsic, high-resistivity crystal. The capacity is reduced by reducing the surface of the read-out anode, which means detectors with a smaller surface or instruments which exploit a drift field to collect the charges produced in a large volume on a smaller anode (*silicon drift detector*; see below). Imaging detectors can be built by dividing a detector into regions each one with a dedicated electronics to read out the signal. This can be

done in one dimension with the so-called microstrip detectors or in two dimensions with the so-called pixel detectors. By increasing the number of pixels, the encumbrance, the power consumption, and the difficulty to bond each pixel with its electronic chain with 100 % efficiency has oriented the technology toward an X-ray adaptation of the charge-coupled device (CCD). In this case, the charge produced by each absorbed photon is collected and temporarily stored in the pixel closer to the interaction point, but the read-out of the pixels occurs sequentially with a single low-noise electronics. The potentials of the pixels are modified periodically to repeatedly shift the charge from the pixel to its neighbor and, eventually, to the output where the charge is preamplified and converted from analog to digital. A CCD devoted to X-rays is substantially the same of a CCD for optical application with some minor changes: (1) A window in the entrance prevents the optical photons to reach the sensitive volume, since it would be saturated by the low-energy electrons generated from the absorption of visible/UV photons. (2) The depletion layer is made thicker to increase the efficiency. (3) The dead layers are minimized in order to minimize the absorption of X-rays. While in an optical CCD each photon creates 1 (or 0) electron and the charge collected in the pixel is proportional to the number of impinging photons, a CCD for X-rays is usually discharged very frequently so that the probability to have more than one photon detected in the same pixel is low and the charge measured from the pixel is proportional to the energy of the photon. Therefore, the CCD is an excellent imager (down to resolutions of the order of 10 μm) and an excellent spectroscopic device (down to resolutions of the order of 150 eV). The major limitations are in the surface that cannot exceed a few square centimeters and in the time resolution that cannot be better than a fraction of milliseconds because of the relatively long time required to transfer the charge of the whole matrix of pixels to the output without losses. A recent evolution of CCDs, still in the development stage, is the *advanced pixel sensor* that reads out the charge without transferring it, with a more parallel read-out and a better energy resolution. More in general the CCD suffers the limitations due to the use of silicon as absorbing material, which becomes quite inefficient above some tens of keV, the actual value depending on the thickness. To detect photons of higher energies, detectors have been studied based on materials of higher atomic number such as GaAs and CdTe, but the technologies have not arrived to the level of development of CCDs yet. Other alternative technology is the *silicon drift detectors* where the charge produced is promptly drifted to the output and the coordinate in the drift direction is measured from the drift time, the drift velocity being known. These have a good energy resolution at room temperature but are compatible only with one-dimensional imaging. A further improvement is the development of *avalanche silicon detectors* that, by exploiting the small difference in the mobility of electrons and holes, can perform a proportional multiplication of the charge.

Detectors that measure the light

X-ray photons can also be detected by measuring the light emitted as a consequence of ionization/excitation processes in suitable crystals. These photons are typically in the UV band and are absorbed in few tens of microns, making impossible their direct detection. However, if crystals are doped with elements that create suitable metastable levels, the absorption of the UV photons creates new photons shifted to longer wavelength. The latter are capable to travel across the crystal to an optical window without being absorbed. A photomultiplier with its electronics detects the light, giving an electrical pulse whose height is proportional to the number of photons collected from the cathode and thence to energy released in the interaction.

Scintillators to detect X- and gamma rays are based on inorganic crystals. The most common have been for years NaI (thallium activated), the CsI (thallium or sodium activated), because of good efficiency and high light yield. The $Bi_4Ge_3O_{12}$ (that does not require activation) is most common to detect higher-energy gammas but has a poor light yield. The more recently discovered $LaBr_3$ (cerium activated) is nowadays the best performing material since it combines a high quantum efficiency with a high light yield that results in a higher-energy resolution. Scintillators are usually employed to detect hard X-rays ($>$10 keV) and soft gamma rays. They are suitable to achieve large volumes, and smaller crystals can be made position sensitive when coupled with multi-pixel photomultipliers. Larger crystals in the form of bars or disks can be made position sensitive by recording how light is shared by different photomultipliers in different parts of the crystal. Gamma cameras for CAT are a classical application of this method.

Detectors that measure the heath

After the very first interaction, all the energy absorbed from the original photon is eventually converted into heath. If a crystal is small and is kept at very low temperature, where the specific heath is extremely small, also the small amount of energy carried by an X-ray photon is sufficient to produce a measurable increase of the temperature. The microcalorimeters (or X-ray bolometers) are based on this effect. The detector is conceptually composed by an absorber that converts the X-rays, a thermometer in good thermal contact to it and a weak link to a thermal bath to restore the initial condition after the photon absortpion. Although the more mature tecnology is to use semiconductor thermistors to measure the small temperature variation of the absorber, the more promising devices exploit a superconductor kept just below the critical temperature. The heath is sufficient to induce the transition from superconducting to normal state and the change of the current flowing in it is detected by a SQUID (Superconducting Quantum Interference Device). The total energy absorbed is derived from the sequence of the

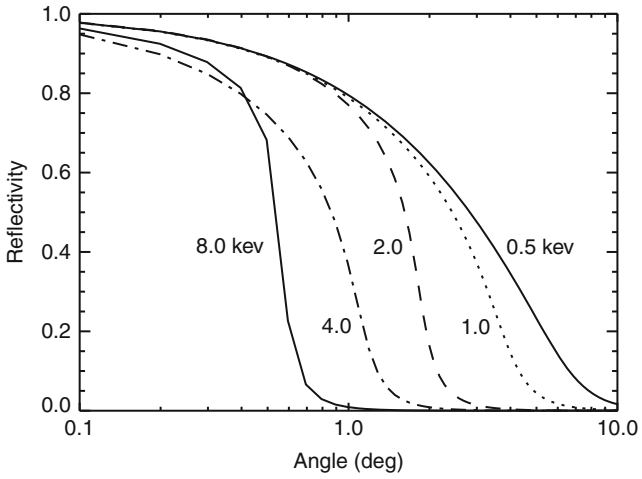
transition as mapped by the SQUID. This is the Transition Edge Sensor, nowadays, the best performing detector in terms of energy resolution (the current record is 1.8 eV at 6 keV). The development of large matrices of TES and of the read-out of individual pixels is the new frontier and the step needed to give practical application to this technology.

Optics and other imaging techniques

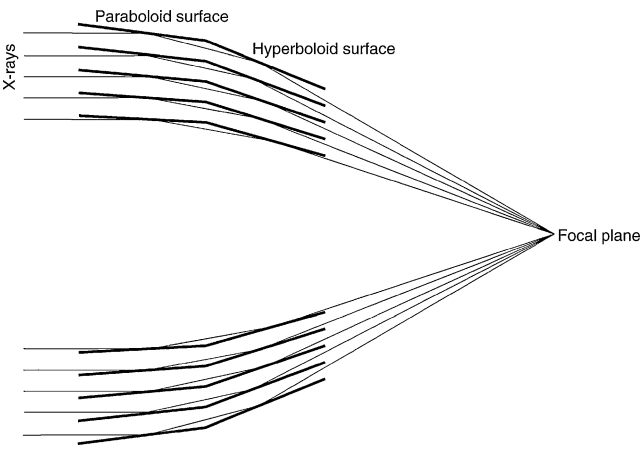
X-ray and soft gamma rays can be focused by means of suitable reflective mirrors like visible light, even if with quite severe limitations. The main difference is that high-energy radiation is reflected only if the grazing angle is below a certain threshold of the order of a few degrees, and this value decreases while increasing energy. This is because the index of refraction at these energies is very close to one. The critical angle above which reflection substantially does not occur is larger for materials of higher atomic number, and indeed it is roughly proportional to the square root of the density of the reflecting material. For this reason a heavy and malleable material as gold is used. As an example, in Figure 2 the reflectivity of a mirror made of gold is reported as a function of the grazing angle and for different energies between 0.5 and 8 keV. Unfortunately, the use of high Z materials causes also a large drop in the reflectivity for energies close to the absorption edges because of the high probability of photoabsorption. This can be filled with a suitable coating on the reflecting material, that is, a material of different atomic number which can reflect X-rays in that particular energy band.

The possibility of reflecting X-rays only for angles below a few degrees implies that the telescope must have a long focal length. This makes the telescopes rather cumbersome. The cost of such a device is also very high, because each reflective surface is to be manufactured with very low roughness, comparable to the wavelength of the radiation to be reflected, precisely aligned and supported by a mechanical structure which provides an adequate stiffness.

X-ray telescopes find their most important use in astronomy since they enable to resolve extended sources and crowded regions, reduce the background, and provide an adequate collecting area. The optics comprises many nested and concentric shells which are formed by two surfaces, the first is a paraboloid and the second is a hyperboloid. The radiation is focused in the focal plane after two reflections on the two surfaces. This configuration, named *Wolter type I*, allows to produce sharp images even off-axis and to reduce the focal length and so the volume (see Figure 3). Inner shells contribute less to the total area because of the smaller radii, but external ones can provide a significant contribution only at low energy since they must reflect photons at larger angles. This actually causes a fast decrease of the effective area with energy, and as a matter of fact, these telescopes can work only below 10–15 keV.



Gamma and X-Radiation, Figure 2 Reflectivity for gold as a function of the grazing angle and for different photon energies. *Solid*, *dotted*, *dashed*, *dot-dash*, and *dash-dot-dot* refer to radiation at 0.5, 1.0, 2.0, 4.0, and 8.0 keV, respectively (Data from Henke et al. (1993)).



Gamma and X-Radiation, Figure 3 Sketch of a telescope for X-ray astronomy. Tens of concentric shells are nested and each contributes to the total area, even if external shells can provide a significant contribution only at low energy because of the larger reflection angles required. X-rays are reflected two times on a paraboloid and on a hyperboloid surfaces to provide good imaging capability even off-axis and to reduce the focal length.

The most advanced X-ray telescope is that onboard the Chandra X-ray Observatory, orbiting the Earth since 1999 (see the Chandra X-ray Observatory website). It was designed to achieve the outstanding angular resolution of 0.5 arcsec, and this has required a focal length of 10 m. The telescope comprises of four nested paraboloid/hyperboloid shells which provide a collecting area of 400 cm^2 at 1 keV. The working energy range is between 0.1 and 10 keV, and only the telescope with the support required to keep aligned the mirrors weighs about 1 t.

A new technology, that is, the multilayer coating, allows to focus X-rays with grazing incidence optics up to hundreds of keV. Instead of a single material with a coating, the shells are made of stacked alternating layers of low and high atomic number materials, deposited via vacuum processes such as evaporation or sputtering. Each layer has a thickness of the order of 1–10 nm which, acting as a crystal with such lattice spacing, diffracts X-rays. Diffraction is a phenomenon which occurs when an electromagnetic wave encounters a spatially repeated structure, such as a grating, with spacing comparable to its wavelength. For X-rays, this occurs with crystal lattices which typically have spacing of the order of a few angstroms (1 Å = 0.1 nm). The planes of the crystal act as reflecting surfaces and the radiation sum, in phase, in those directions which make the difference of optical path between reflections on different planes an integer value of wavelength. Bragg's law relates the energy and the angle at which the diffraction can occur:

$$E = \frac{nhc}{2d \sin \vartheta} \tag{5}$$

where h and c are Plank's constant and the speed of light in vacuum, respectively, d is the crystal lattice spacing, ϑ is diffraction angle, and n an integer which indicates the order of diffraction. Multilayer optics can reflect photons at higher energy, or at larger glancing angles, than "classical" telescopes. By substituting $\vartheta = 1^\circ$ and $d = 1$ nm in Equation 5, it results that the energy diffracted is 35 keV (for $n = 1$). This is a great improvement with respect to single-layer optics, by which photons are reflected at angles larger than 1° only below a few kiloelectron volts. The diffraction from the single layer is nearly monochromatic since, for a fixed value of the spacing of the layer d, the energy is also fixed except for the diffraction order. To diffract continuum spectra, the thickness of the layers is gradually decreased with depth so that the total response is the convolution of nearly monochromatic lines at different energies. Multilayer optics in X-ray astronomy was tested for the first time between 5 and 80 keV by the NASA mission NuSTAR, which was launched in 2012 (Harrison et al., 2013).

Above the hard X energy range, the imaging of radiation cannot be performed with multilayer optics, because it is difficult to make layer with thickness below 1 nm and low roughness. The most promising technique to concentrate soft gamma rays is the use of Laue lenses: photons pass through crystals and are diffracted at grazing angles on the lattice, whose spacing is of the order of a few angstroms (Laue geometry). Laue lenses actually do not focus the image on a focal plane but they just concentrate the radiation. Although the image of a point source on-axis is really a punct, the quality of the image basically depending on the alignment of crystals which compose the lens, it appears like a ring off-axis. Since diffraction on crystal is nearly monochromatic, the efficiency for continuum spectra is increased, at the expense of the image quality, with the use of mosaic crystals, which are composed of small domains slightly and regularly misaligned acting as independent crystals. Laue lenses are at the moment in development stage, and their use can be expected in the near future.

Imaging of hard X- and gamma rays can be performed also with *coded masks*. These are an evolution of the simple pinhole cameras: The direction of photons is reconstructed by the shadow projected on a sensitive plane. A coded mask instrument is composed of a mask, which is divided in elements, and of a position-sensitive detector. The elements are arranged in a predetermined pattern; some of them are transparent to the radiation to be detected, while the others are opaque. The shift of the mask shadow with respect to center and the distance between the detection plane and the mask enable to reconstruct the direction of impinging photons. The former is measured with a detector with a spatial resolution better than the size of the mask elements. If there is more than one source, the shadows projected by each are summed but the arrangement of the mask elements is chosen so that in any case the position of the sources can be reconstructed without ambiguity by means of a deconvolution process. The intensity of each source is proportional to the "strength" of the corresponding shadow, while the angular resolution is dictated by geometrical parameters, such as the spatial resolution of the detector, the size of the mask elements, and the distance with the detection plane. As a matter of fact, angular resolutions are limited to a few arc minutes because better values would require long focal length and hence quite large volumes.

Coded masks provide a low cost and compact solution to image high-energy radiation up to soft gamma rays, even if the sensitivity of a focusing instrument is always better for an equal collecting area. A coded mask detector measures as a whole the shadow, and hence the noise, of all sources. As a result, the noise which affects the single source is the sum of all contributions, while focusing instruments resolve them spatially. Moreover, the actual geometrical area of the detector in a coded mask device is always larger than its collecting area because about a half of flux must be absorbed by the mask. To reach a large collecting area, an even larger sensitive area is required and this implies a higher background. On the contrary, the flux of sources detected by instruments inserted in focal planes are affected only by the background in the small region where photons are concentrated, that is, the point spread function of the telescope. An application where coded masks are largely used is the imaging of a large fraction of the sky. Grazing incidence telescopes have a field of view smaller than 1 square degree, while a single-coded mask instrument can cover even more than one tenth of the sky, and hence, the entire sky coverage can be performed with a small cluster of detectors.

Another interesting possibility to build large field of view and sensitive instruments is the use of *lobster eye* optics, which is based on the same principle as the

eye of crustaceans (Angel, 1979). The optics is composed of an array of square channels, which are arranged so that their axes are orthogonal to a spherical surface with a radius R. The internal surfaces of channels are made reflective and a true image is formed by those rays which experience an odd number of reflections on both the couples of parallel walls of the channel. Lobster eye lenses are optically equivalent to a concave spherical mirror of radius R and indeed the focal plane is also spherical, with a radius equal to $R/2$. Photons which are reflected an even number of times, also if in just one direction, produce a linear image, parallel to the sides of the channels and passing through the focused image, and, together with photons which are not reflected, are to be regarded as an additional noise. The diagonal of the channel is also the diameter of a point source image on the focal plane, because rays focused by any channel emerge parallel from it with a beam size equal to the diagonal. In principle, lobster eye optics can have a field of view as large as desired up to 4π steradian, that is a complete angular coverage, because there is not any preferred direction, but actually there are technical limitations to build spherical optics and detectors. Also the angular resolution can be very good, and it can be of the order of arc minutes or below. It is basically determined by the ratio between the size of the channel and the focal length, that is, $R/2$.

Lobster eye telescopes are built by exploiting slumped lead glass microchannel plates (MCP, Fraser et al., 1993). Coatings such as iridium or nickel are used to enhance the reflectivity, and the reflection up to hard X-rays has been reported (Price et al., 2002a). An improvement to the basic design is also the packing of a pair of MCPs to form an approximate Wolter type I optics, with each MCP composed of square channels arranged radially (Willingale et al., 1998; Price et al., 2002b). This design improves the performance from many points of view, such as better imaging capabilities and higher effective area, but the field of view is narrower because a preferred direction is introduced by the system.

Recently, lobster eye optics has been also developed with metal ribs coated with gold (Gertsenshteyn et al., 2005). The possibility to polish the walls of the channels allows for the diffraction up to hard X-rays (up to 40 keV) and has opened the way for a number of applications. One of the most interesting is the possibility to inspect objects in a fast and nondestructive way with a portable instrument. This comprises of an X-ray generator, which illuminates the material, and a detector in the focus of lobster eye optics. Photons which are scattered back toward the detector allow for imaging different materials in the object, since the probability of scattering depends on the number of electrons and then on the atomic number. Such an application requires hard X-rays, which are quite penetrating and interact with matter predominantly with Compton scattering. Moreover, the use of a lobster eye telescope allows to collect enough photons in a reasonable time without the use of a cumbersome grazing incident multilayer optics.

At even larger energies, the interactions produce high-energy electrons that are highly penetrating. In experiments active above 50 MeV photons are converted on a thin layer of a high Z material (W or Pb) into an electron-positron pair. These two particles propagate through a tower of tracking silicon detectors. From the analysis of tracks, the original vertex of interaction is reconstructed and the direction of the original photon is derived. Eventually the pair is absorbed on a thick scintillator to measure the energy.

The Earth from outside: X- and gamma rays from atmosphere

The atmosphere of the Earth seen from space is an intense source of X-rays and gamma rays. Also an intense flux of X-rays and γ-rays falls on the atmosphere from the space: Most of them come from the cosmic diffuse background plus many cosmic X-ray sources, and for the half of the Earth exposed to the Sun, they also arrive from solar flares and steady coronal emission. Some of these photons are absorbed by photoelectric effect, while at higher energies the dominant interaction is the Compton effect. A fraction of photons, of the order of 10 % around 80 keV and decreasing at lower and higher energies, is backscattered to space with a significant shift toward longer wavelengths. But the largest component of photons emitted by the atmosphere derives from the interactions with cosmic rays, which are energetic particles coming from a number of classes of astrophysical sources. Therefore, even though atmospheric emission is usually named "albedo" also in X- and gamma rays, the meaning of this term is quite different from the usual optical definition because actually photons emitted are, for the largest part, not reflected.

Cosmic rays have a large flux with a spectrum extending up to very high energies but quickly decreasing at increasing energies. They are composed mainly of protons (i.e., nuclei of hydrogen for $\sim$89 %), α-particles (i.e., nuclei of helium for $\sim$10 %) plus nuclei of higher atomic number elements (for the lasting $\sim$1 %). In paths of the order of few tens of grams per square centimeter, which correspond by dividing by the density of the upper layers of the atmosphere to a height of a few tens of kilometers, these particles collide with the protons or the neutrons in the nuclei of atmosphere constituents, losing a part of their energy and producing short lifetime particles named pions. The neutral pions π° decay into two photons, while charged pions π^{+} and π^{-} decay into μ^{+} and μ^{-} mesons, that further decay into electrons and positrons, plus a bunch of elusive neutrinos, that have a very low probability to further interact. Due to conservation of momentum, all these products (photons and electrons) have still directions close to the original direction of the cosmic ray particle. The particles and the photons produced after the first interactions have sufficient energy to interact again and generate further particles, giving rise to a so-called shower. We have, therefore, a bunch of photons

and particles moving downward and continuously interacting, increasing in number and decreasing in energy. After crossing a certain amount of atmosphere, the components of the shower lose energy by photoelectric absorption and by ionization, and the number of particles/photons starts to decrease. Eventually only a few μ mesons arrive to the ground. Also, as long as the shower develops, the angular distribution increases and some photons are backscattered. While most of the energy of the shower is lost in the atmosphere, a certain fraction of photons, nuclei, and electrons move upward and succeed to escape. This is the main constituent of the so-called albedo. For satellites in low Earth orbit, between 400 and 2,000 km, this is a significant source of instrumental background. In the past, many semiempirical formulas have been given to describe the albedo. More recently, given the presence of excellent codes to simulate all these interactions and the good knowledge of the Cosmic ray spectra, detailed computations have been performed of the X-/γ-ray fluxes at different depths. Sazonov et al. (2007) derive from its simulations that the spectrum of X-/γ-ray albedo from the atmosphere in the range 25–300 keV is well fitted by the function:

$$\frac{dN_\gamma}{dE} = \frac{C}{\left(\frac{E}{44\,keV}\right)^{-5} + \left(\frac{E}{44\,keV}\right)^{1.4}}\ photons/keV \qquad (6)$$

The spectrum peaks around 50–60 keV. At lower energies we have a fast drop, due to the fact that at lower energies photons are absorbed, and the emission involves thinner and more external layers of the atmosphere. When compared with the sky, the Earth as seen from the space is very bright in γ-rays, while in soft X-rays is dark. The constant C in Equation 6 depends on the magnetic latitude, on the phase of the solar cycle and on the zenith angle. Formulas to compute this normalization can be found in Sazonov et al. (2007). At the peak around ~60 keV, the flux integrated on all directions is of the order of 4×10^{-3} photons/(cm^2 s keV) at the magnetic equator and around six times higher at a magnetic latitude of 65°. Also some nuclear gamma ray lines in the MeV range are present, but they are a minor fraction of the total flux.

The remote sensing of the Earth from space in the hard X/soft gamma band is mainly probing the interaction of Cosmic Rays with the upper layers of the atmosphere. This can identify a method to study the atmosphere of other planets and has been actually used with some interplanetary probe.

A particular case is that of terrestrial gamma-ray flashes (TGFs). They were discovered by the Burst and Transient Source Experiment aboard the Compton Gamma Ray Observatory (Fishman et al., 1994) and are very short bursts of γ-rays, lasting from 0.5 to 4 ms, arriving from the direction of Earth and characterized by a much harder spectrum than typical celestial gamma-ray burst. The very short duration of these events puts a requirement on time resolution and on triggering capability of the instrument, which is usually satisfied by modern satellites. More events were reported by the Reuven Ramaty High Energy Solar Spectroscopic Imager (Smith et al., 2005) and AGILE arrived to detect a TGF event with a maximum energy of 43 MeV (Marisaldi et al., 2010). The geographic distribution of TGF is peculiar: There is a concentration at very low latitudes, with a higher frequency above the dry land than above the oceans.

While the fact that TGFs are originated within the atmosphere seems out of question, a large debate is in progress about their nature. There is a certain similarity with X-/γ-radiation observed on ground from lightning. Dwyer (2008) proposes that the energetic seed particle production most likely involves either relativistic feedback or runaway electron production in the strong electric fields associated with lightning leaders or streamers. The phenomenon would originate at relatively low levels in the atmosphere (10–20 km). The Atmosphere Space Interaction Monitor, aboard the International Space Station, will study TGFs, trying to connect them with storms and with the elusive Transient Luminous Events. ASIM includes a coded mask and an array of Hard X-ray detectors to better determine the location of the phenomenon (Neubert et al., 2006).

Remote sensing of solar system bodies: planets, satellites, and minor bodies

If the remote sensing (RS) of the Earth is limited to few topics that do not include the solid or liquid surfaces, the situation is different for planets or other bodies of the Solar System without an atmosphere or with one thin enough to be transparent to X- and gamma rays. Actually experiments of RS have been embarked aboard missions flying by and orbiting around some of these bodies.

In the domain of X-rays, the spectrum of photons from the surface can determine, through the fluorescence lines and the absorption edges, the chemical composition and the abundances of the surface. Depending on the energy, this analysis can probe layers spanning from a few tens of microns for Si to, for instance, 1 mm for nickel. One major problem is that planetary surfaces do not emit X-rays but they can only reprocess those coming from an external source. Usually solar flares are providing such a source but, for an interpretation of fluorescence spectra in terms of quantitative chemical composition, a good knowledge of the impinging spectrum is mandatory. However, this is not very easy to achieve because every solar flare has its own spectrum. In the gamma-ray band, a celestial body can be itself a source of radiation deriving from the intrinsic radioactivity of the rocks or from that produced by the impact of cosmic rays, both primary and solar. Since the cosmic rays are ubiquitous, their characteristic affects less the results than the impinging spectrum in X-rays spectroscopy. The gamma-ray lines provide unique signatures for the presence of elements in a layer of the order of 10 cm.

A gamma-ray spectrometer (GRS) was mounted aboard Apollo 15 and Apollo 16. It was based on a NaI(Tl)

scintillator read with a photomultiplier sensitive in the 0.065–27.5 MeV range. Orbiting around the Moon, these instruments covered about 20 % of the surface. They had no directionality and a row mapping was achieved simply on the basis of proximity to various regions at various orbital phases. By comparison of data accumulated in orbit with samples of the Moon collected from various missions in different sites and returned to the Earth, these data allowed to assess that anorthositic fragments are the main constituents of the Moon's surface.

The Near Earth Asteroid Rendezvous Shoemaker spacecraft orbited around asteroid 433 EROS from February 14, 2000, and after about a year it landed on the surface. It included an X-ray/gamma-ray spectrometer. The X-ray detector was a set of three proportional counters capable to measure spectra in the 1–10 keV range. The gamma-ray detector was made of an internal scintillator of NaI(Tl) with a shield made of BGO, each read by its own photomultiplier. The X-ray spectrometer detected the X-rays by solar flares backscattered with superimposed absorption edges and fluoresce lines. From a modeling of solar flares spectrum, the composition of surface layers was derived. The gamma-ray spectrometer detected nuclear lines from a few elements. This can be considered the first case of remote sensing of a reasonable completeness.

After this first experience, X-ray and gamma-ray spectrometers were included in planetary probes. SMART-1, an ESA mission that orbited around the Moon from 2003 to 2006, included the Demonstration of a Compact Imaging X-ray Spectrometer (D-CIXS), an X-ray spectrometer based on silicon detectors with a narrow field collimator (Grande et al., 2007). An important improvement is the presence of a detector of solar flares (Huovelin et al., 2010), providing the input spectrum needed to convert the measured spectra into abundance of elements. A similar instrument, Chandrayaan-1 X-ray Spectrometer (C1XS) was launched on October 22, 2008, on India's Chandrayaan-1 mission to the Moon (Swinyard et al., 2009).

Another X-ray spectrometer (XRS) and gamma-ray spectrometer (GRS) were aboard the Japanese satellite Kaguya/SELENE operative in 2007–2008 (Hasebe et al., 2008; Okada et al., 2009). The spectral quality was much better than previous instruments because it used CCDs for XRS (combined with a solar flare monitor) and a high-resolution germanium detector for the GRS.

MESSENGER is a mission inserted in orbit around Mercury in 2011 (Schlemm et al., 2007) and, since then, mapping elemental abundances in the surface of the planet (Nittler et al., 2011). The X-ray spectrometer onboard is based on silicon detectors set with a collimator to detect the X-rays from the Sun and another set to detect those reflected by Mercury. A germanium detector is dedicated to gamma rays and neutrons (Goldsten et al., 2007).

DAWN mission to Vesta and Ceres includes a gamma-ray and neutron detector to detect gamma-ray lines (Prettyman et al., 2004). A first map of elemental abundances of Vesta minerals, acquired with low altitude orbits was published (Prettyman et al., 2012). So far we only described experiments that measure X-ray and gamma-ray spectra, in some case with a field limiter, but never imaging. The first imager devoted to remote sensing will be hosted aboard the Mercury Planetary Orbiter of the BepiColombo mission (Fraser et al., 2009). The mercury imaging X-ray spectrometer uses two focusing instruments composed of two identical silicon active pixel sensors and two X-ray optics based on microchannel plates. One telescope is a lobster eye lens, while the other is in a Wolter type I configuration. With this instrument X-ray remote sensing would significantly decrease the gap with respect to remote sensing in other wavelengths.

Summary

X-rays and gamma rays cover the short wavelength side of the electromagnetic spectrum. Their more frequent interactions with matter are the photoelectric absorption, the Compton scattering, and the pair production, whose relative probability depends on the energy of the photon and on the atomic number of the material. X-rays and gamma rays are detected with proportional counters, semiconductor detectors, scintillators or bolometers. X-rays can be reflected at grazing angles by very smooth surfaces. This allows for the manufacture of telescopes that can arrive to good angular resolutions on very small fields of view. To perform imaging on large field of view particular telescopes, named lobster eye, can be used. At higher energies and for wide field imaging systems combining a mask designed with a particular code and position-sensitive detectors are used.

Our Earth observed from the space is a bright source of X-rays and gamma rays, mainly originated by the interaction of cosmic rays with the top layers of the atmosphere. Therefore, remote sensing in these bands is mainly probing this region. On the contrary remote sensing in X-ray and gamma-ray band is performed on bodies of the Solar System without atmosphere (the Moon, Mercury, asteroids) or with a thin one (Mars), with instruments aboard flying-by, orbiting, or landing probes. This technique is expected to contribute significantly to the study of chemical composition of planetary surfaces.

Bibliography

Angel, J. R. P., 1979. Lobster eyes as X-ray telescopes. *The Astrophysical Journal*, **233**, 364.

Berger, M. J., et al., XCOM: photon cross sections database. Available from World Wide Web: http://www.nist.gov/physlab/data/xcom/index.cfm.

Chandra X-ray Observatory website. Available from World Wide Web: http://chandra.harvard.edu/press/chandra_specs.html.

Dwyer, J. R., 2008. Source mechanisms of terrestrial gamma-ray flashes. *Journal of Geophysical Research*, **113**, D10103, doi:10.1029/2007JD009248.

Fishman, G. J., et al., 1994. Discovery of intense gamma-ray flashes of atmospheric origin. *Science*, **264**, 1313.

Fraser, G. W., et al., 1993. X-ray focusing using square-pore microchannel plates: first observation of cruxiform image structure. *Nuclear Instruments and Methods in Physics Research*

Section A: Accelerators, Spectrometers, Detectors and Associated Equipment, **324**, 404.
Fraser, G., et al., 2009. The mercury imaging X-ray spectrometer (MIXS) on bepicolombo. *Planetary and Space Science*, **58**, 79.
Gertsenshteyn, M., Jannson, T., and Savant, G., 2005. Staring/focusing lobster-eye hard X-Ray imaging for non-astronomical objects. *Proceedings of SPIE*, **5922**, 59220N–59221N.
Gerward, L., 1999. Paul Villard and his discovery of gamma rays. *Journal Physics in Perspective (PIP)*, **1**, 367.
Goldsten, J. O., et al., 2007. The MESSENGER gamma-ray and neutron spectrometer. *Space Science Reviews*, **131**, 339.
Grande, M., et al., 2007. The D-CIXS X-ray spectrometer on the SMART-1 mission to the moon – first results. *Planetary and Space Science*, **55**, 494.
Harrison, F. A., et al., 2013. The nuclear spectroscopic telescope array (NuSTAR) high-energy X-ray mission. *The Astrophysical Journal*, **770**, 103.
Hasebe, N., et al., 2008. Gamma-ray spectrometer (GRS) for lunar polar orbiter SELENE. *Earth, Planets and Space*, **60**, 299.
Henke, B. L., Gullikson, E. M., and Davis, J. C., 1993. X-ray interactions: photoabsorption, scattering, transmission, and reflection at E = 50-30000 eV, Z = 1-92. *Atomic Data and Nuclear Data Tables*, **54**(2), 181–342. Available on-line at http://henke.lbl.gov/optical_constants/.
Huovelin, J., et al., 2010. Solar intensity X-ray and particle spectrometer (SIXS). *Planetary and Space Science*, **58**, 96.
Knoll, G. F., 2010. *Radiation Detection and Measurement*. Hoboken, NJ: Wiley.
Leo, W. R., 1994. *Techniques for Nuclear and Particle Physics Experiments*. Berlin: Springer.
Lutz, G., 2007. *Semiconductor Radiation Detectors: Device Physics*. Berlin: Springer.
Marisaldi, M., et al., 2010. Detection of terrestrial gamma ray flashes up to 40 MeV by the AGILE satellite. *Journal of Geophysical Research*, **115**, A00E13, doi:10.1029/2009JA014502.
NASA's Earth Fact Sheet. http://nssdc.gsfc.nasa.gov/planetary/factsheet/earthfact.html.
Nittler, L. R., et al., 2011. The major-element composition of mercury surface from MESSENGER X-ray spectrometry. *Science*, **333**, 1847.
Neubert, T., et al., 2006. The atmosphere-space interactions monitor (ASIM) for the International space station. In Gopalswamy, N., and Bhattacharyya, A. (eds.), *Proceedings of the ILWS Workshop, Goa, India,* p. 448.
Okada, T., et al., 2009. X-Ray Fluorescence Spectrometer (XRS) on Kaguya: current status and results. In *Proceeding of "40th Lunar and Planetary Science Conference* (2009)*"*, The Woodlands, TX, ID: 1897.
Prettyman, T. H., et al., 2004. Mapping the elemental composition of Ceres and Vesta: Dawn's gamma ray and neutron detector. *Proceedings of SPIE*, **5660**, 107.
Prettyman, T. H., et al., 2012. Elemental mapping by Dawn reveals exogenic H in Vesta's Regolith. *Science*, **338**, 242.
Price, G. J., et al., 2002a. Hard X-ray imaging with microchannel plate optics. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, **490**, 276.
Price, G. J., et al., 2002b. X-ray focusing with Wolter microchannel plate optics. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, **490**, 290.
Sazonov, S., Churazov, E., Sunyaev, R., and Revnivtsev, M., 2007. Hard X-ray emission of the Earth's atmosphere: Monte Carlo simulations. *Monthly Notices of the Royal Astronomical Society*, **377**(4), 1726–1736.
Schlemm, C. E., et al., 2007. The X-Ray spectrometer on the MESSENGER spacecraft. *Space Science Reviews*, **131**, 393.
Shamos, M. H., 1987. *Great Experiments in Physics: Firsthand Accounts from Galileo to Einstein*. New York: Dover.
Smith, D. M., Lopez, L. I., Lin, R. P., and Barrington-Leigh, C., 2005. Terrestrial gamma-ray flashes observed up to 20 MeV. *Science*, **307**, 1085.
Swinyard, B. M., et al., 2009. X-ray fluorescence observations of the moon by SMART-1/D-CIXS and the first detection of Ti Kα from the lunar surface. *Planetary and Space Science*, **57**, 744.
Willingale, R., et al., 1998. Hard X-ray imaging with microchannel plate optics. *Experimental Astronomy*, **8**, 281.

Cross-references

Electromagnetic Theory and Wave Propagation
Radiation (Natural) Within the Earth's Environment
Radiation, Electromagnetic
Radiation, Galactic, and Cosmic Background
Severe Storms

GEODESY*

Calvin Klatt
Geodetic Survey Division, Natural Resources Canada, Ottawa, ON, Canada

Definition

Geodesy. The branch of Earth Sciences concerned with the size, shape, orientation, and gravity field of the Earth.
Global Navigation Satellite Systems (*GNSS*). Satellite navigation systems that provide geo-spatial positioning with global coverage.
Very Long Baseline Interferometry (*VLBI*). A radio-astronomical technique utilizing globally distributed antennas having independent atomic clocks. The antennas collectively form an interferometer or synthetic aperture.
Gravimetry. That branch of the geosciences concerned with the measurement of the force or acceleration of gravity.
International Terrestrial Reference Frame (*ITRF*). The internationally agreed terrestrial reference frame, defining longitude and latitude at a number of reference points, thus forming the basis for much global positioning.

Introduction

Geodesy is comprised of three "pillars": the shape of the Earth, its orientation, and its gravity field.

Geodesists first consider the global scale and then proceed to the local. Regional mapping, much geophysics, and positioning with the Global Positioning System (GPS) require first that global aspects be well understood. This has driven geodesy to become perhaps the most international of the sciences.

International services and projects, organized under the auspices of the International Association of Geodesy (IAG) (http://www.iag-aig.org/), are needed to coordinate work which cannot be conducted on national or

*Her Majesty the Queen in Right of Canada

continental scales. Technique-specific services today include the International GNSS Service (IGS) (http://www.igs.org/), the International VLBI Service for Geodesy and Astrometry (IVS) (http://www.ivscc.gsfc.nasa.gov/), the International DORIS Service (IDS) (http://ids.cls.fr/), and the International Laser Ranging Service (ILRS) (http://www.ilrs.gsfc.nasa.gov/). The International Earth Rotation and Reference Systems Service (IERS) (http://www.iers.org/) is responsible for producing products based on information from these services. Products include reference frames and Earth's orientation information. The IAG International Gravity Field Service (IGFS) (http://www.igfs.net/) coordinates activities related to the Earth's changing gravity field.

The Global Geodetic Observing System (GGOS) (http://www.ggos.org/) is an IAG effort to further integrate international geodetic activities. GGOS contributes to the Global Earth Observing System of Systems (GEOSS) (http://www.earthobservations.org/) and is consistent with the Integrated Global Observing Strategy (IGOS) (http://www.igospartners.org/) (see *Global Earth Observation System of Systems (GEOSS)*).

The shape of the earth and reference frames

The Earth's shape is approximately that of an oblate spheroid with an equatorial radius of about 6,378 km. This radius is approximately 0.3 % greater than that at the poles, or 21 km. The topography varies from about +9 km (Mount Everest) to −11 km (Mariana Trench). The relative size of these variations accounts for the appearance of the Earth from space as a nearly perfect sphere.

For geodesists, the shape of the Earth is described by the location and velocity of reference points in a geocentric terrestrial Earth-fixed reference frame, called the International Terrestrial Reference Frame (ITRF). Locations may be given in geocentric X, Y, and Z coordinates, or in terms of latitude, longitude, and height above an ellipsoid that approximates the Earth's shape.

The ITRF and the International Terrestrial Reference System (ITRS) are products of the IERS. The ITRS defines the frame origin, scale, orientation, and time evolution. The ITRF is a practical realization of the ITRS: three-dimensional coordinates and velocities of reference points on the surface of the Earth. The ITRF is updated periodically as more observational data is obtained or with improvements in theory or analysis.

The IERS publishes conventions to be followed by contributors, providing agreed numerical standards, reference frame information, geophysical models, and models for propagation of observables (McCarthy and Petit, 2003).

The ITRF approximates a no-net-rotation surface, resulting from an averaging of tectonic and other motions on the Earth's crust. As a result, the velocity of a point in the ITRF will be the sum of that due to the motion of the local tectonic plate, intraplate motion/deformation, and other local effects relative to a global mean of motions.

The ITRF is produced from the combination of data from a number of different techniques. Of interest to users are the low cost and easily deployed Global Navigation Satellite System (GNSS) receivers, dominated today (2008) by the Global Positioning System (GPS). There is more than one GNSS constellation. Apart from the operational US GPS system, Russia's GLONASS is near full operational capability, the upcoming European GALILEO system holds much promise, and China may extend the regional Beidou/Compass system to provide global coverage. Regional Navigation Satellite Systems have been proposed by Japan and India.

GNSS augmentation is the use of additional information to improve that broadcast by the satellites. Augmentation can occur in real time or in postprocessing. A tradeoff has always existed between latency and accuracy, but improvements in both are occurring at a very rapid pace. User-obtained accuracy is dependent on many factors, including mode of operation (static or dynamic), single or dual frequency observations, length of observation, and equipment used. Postprocessing of data sets (from anywhere on Earth) can now be done on the internet to cm-level accuracy with a delay of only 90 min after data collection (Mireault et al., 2008). The Wide Area Augmentation System (WAAS) is a real-time augmentation system operated by the US Federal Aviation Administration that provides accuracy at nearly the 1 m level. WAAS is available in the Continental United States as well as most of Canada and Alaska. Similar systems exist in much of the world.

The primary benefit of global reference frames is in navigation, positioning, and Earth science studies. For example, the driving force behind the development of the GPS constellation was the military need for precise positioning anywhere on Earth. Use of global reference frame information benefits users who integrate georeferenced data collected by different people for many different purposes at different times. Related fields include mapping, surveying, and hydrography (see *Geological Mapping Using Earth's Magnetic Field*).

In remote sensing, georeferencing may be performed through occupation with GNSS receivers of identifiable points on imagery. With continued improvement of Earth Observation resolution, accurate orthorectification is increasingly important. Additionally, remote sensing satellites or aircraft may be dynamically positioned via GNSS. Augmentation to decimeter-level precision of kinematic GPS data collected during airborne remote sensing can be easily performed (Mireault et al., 2008). Precise global reference frames are required for many scientific efforts, including global sea-level rise studies and satellite laser altimetry missions.

As precision in geodesy has advanced, the ability to observe contemporary change has arisen (Chao, 2003). Networks to monitor motion are established near tectonic plate boundaries, in regions experiencing glacial isostatic adjustment (postglacial rebound) or other regions of interest. Regional, national, continental, and global networks

with varying data latency, precision, and spatial density exist to respond to geoscientific needs (see *Subsidence*).

Geodetic reference frame analyses require accurate modeling of many geophysical effects which affect the observables or instruments, including tidal effects, atmospheric loading, and antenna offset/deformation (McCarthy and Petit, 2003).

Effects on geodetic signals (electromagnetic radiation), experienced while traversing the Earth's ionosphere and troposphere, are very significant. Ionospheric effects can be precisely quantified by observation of multiple frequencies since the refractive index is a strong function of frequency. Taking advantage of this, GNSS systems are being deployed as ionospheric sensors (see *Ionospheric Effects on the Propagation of Electromagnetic Waves*).

Hydrostatic delays and those due to water vapor in the troposphere corrupt observables in a manner that is difficult to correct for due to the chaotic nature of the weather. Extended data collection and sensitive analysis are necessary to separate these effects from the desired geodetic information (see *Water Vapor*).

GNSS systems are deployed in meteorology in two ways. First, the analysis of fixed continuously operating GNSS systems yields information on atmospheric pressure (hydrostatic delays), including water vapor content above the antennae. Second, limb sounding of GNSS satellite signals from other satellites in lower orbit provides information on atmospheric density profiles (see *Limb Sounding, Atmospheric*).

Earth's orientation

The orientation of the Earth is given by the relationship between the terrestrial reference frame and another representing inertial space. This relationship involves the rotation of the Earth, polar motion (geographic position of the rotation axis of the Earth), and the coordinates of the rotation axis in the celestial sphere ("celestial pole"). Each of these Earth orientation components exhibits some degree of motion that is currently unpredictable.

Polar motion includes a drift of about 12 m per century superimposed on the Chandler Wobble (having a diameter of about 15 m and a period of approximately 433 days) and a smaller seasonal signal of 8 m diameter. The slow drift of the pole is due to large-scale mass redistribution such as postglacial rebound.

Motion of the celestial pole is the sum of precession (one cycle each 25,800 years) and nutation (a series of many shorter period terms).

Following the invention of accurate clocks based on crystal oscillators during the Second World War, it soon became clear that the Earth's rotation was irregular. Instrumentation was in place to measure the changing Earth rotation prior to the introduction of conventional Universal Time in 1960. The rotation rate may be expressed as its inverse, the length of day.

The Earth system ("solid Earth," oceans, atmosphere) approximates a closed system which conserves angular momentum. Thus, the counterpoint to variations in Earth's rotation/orientation ("solid Earth") can be seen in atmospheric and oceanic changes. The Earth's orientation thus provides a global sensor of mass movements both within (the "solid" Earth is not solid) and outside the Earth.

The IERS publishes Earth Orientation Parameters which relate the ITRF with the International Celestial Reference Frame (ICRF). The ICRF approximates an inertial reference frame and is primarily expressed by a set of coordinates of quasars, extragalactic objects at great distance from the Earth. This great distance ensures that quasars show no apparent movement (unlike visible stars) and are thus nearly ideal points of reference. The astronomical Very Long Baseline Interferometry (VLBI) technique is critical to establish the ICRF and its link to the ITRF (i.e., the Earth's orientation). Satellite systems are crucial in establishing the origin of the ITRF.

The historical link between time and longitude continues today in the use of GNSS systems as time references (telecommunications) and for precise time transfer, most notably among timing laboratories used in Universal Time definition.

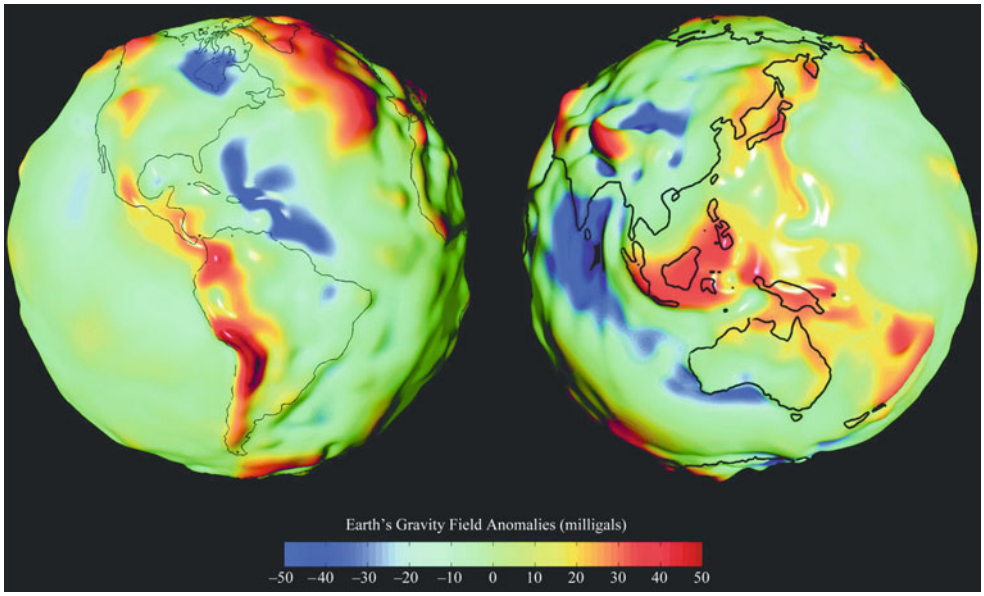
Physical geodesy

The study of the gravity field of the Earth is known as physical geodesy. The force of gravity varies at the surface of the Earth due to distance from the geocenter, density variations in the crust, and other effects.

The Earth is neither homogeneous nor smooth, so the gravitational potential at its surface varies spatially. To describe these variations, we define an equipotential surface (the "geoid") which corresponds to the mean ocean surface of the Earth ("mean sea level"). The geoid is typically described by a smooth ellipsoid plus local "geoid undulations" which can be as large as 107 m (NIMA World Geodetic System Technical Report 2000) (Figure 1).

Heights on mapping products are given as heights above sea level; thus, they are representative of the inhomogeneous Earth and are useful for management of water. Heights have traditionally been obtained by sequential spirit level measurements from a reference point (e.g., tide gauge), but today it is possible in many regions to use GNSS and a geoid model. Geoid models are used with measurements of sea-surface height to identify sea-surface topography associated with currents (see *Sea Surface Wind/Stress Vector*).

Gravity field information contributes to petroleum and mineral exploration, as mineral deposits will have characteristic gravitational signatures. Similarly a characteristic gravitational signature occurs where continental crust meets oceanic crust; thus, gravimetry is used to support efforts to define limits to continental shelves, crucial for delimiting national borders under the United Nations Law of the Sea (see *Solid Earth Mass Transport*).



Geodesy, Figure 1 The Earth's gravity field derived from GRACE satellite data (Image courtesy of NASA's Gravity Recovery and Climate Experiment, The University of Texas at Austin's Center for Space Research).

Gravimetry (at the surface, airborne, or from space) can be considered a remote sensing tool that sees through solid objects. Gravity combined with other information (magnetic, seismic, surface geology, etc.) is used to produce three-dimensional maps of potential oil- or mineral-bearing regions. New technologies fusing gravity data with other remotely sensed information, such as hyperspectral surface geochemistry or RADAR interferometry, have the potential to further contribute.

Models of the Earth's gravity field are also used in satellite orbit determination and to predict missile trajectories.

The Earth's gravity field has a temporal structure as well. An instrument on the Earth's surface will record large diurnal changes resulting from solid Earth tides (on the order of 30 cm), since the instrument will be sensitive to geocentric distance. Gravity measurements on the Earth's surface in regions experiencing glacial isostatic adjustment (postglacial rebound) detect both the rising surface of the Earth above the geocenter and mass redistribution in the mantle.

Other temporal changes in Earth's gravity are associated with hydrology, as these are the largest mass redistributions on a time scale of months to years. Satellite missions such as GRACE (http://www.csr.utexas.edu/grace/) are capable of remotely sensing the changing water mass on and in the Earth (Velicogna and Wahr, 2006). Based on observations of the motion of and relative motion between the two GRACE satellites, a gravity field model is determined at regular (monthly) intervals and expressed in spherical harmonic functions. Such satellites are entirely unlike other remote sensing satellites which use electromagnetic radiation, frequently incorporate CCD sensors, and produce images composed of pixels.

Conclusion

Geodesy is one of the oldest and most international of the sciences. Rapid technological improvements have transformed our capabilities in recent years, enabling exciting new discoveries about the Earth system. It has been argued that there is a space-geodetic "Moore's law" in which we see a tenfold improvement every decade in measurement precision as well as, in some cases, spatial and temporal resolution (Chao, 2003). These advances are opening up new opportunities such as real-time GNSS-based tsunami warning systems and hydrology from satellite gravimetry.

User access to the ITRF is expected to rapidly improve in the near future as the capacity to perform satellite orbit/clock modeling in near-real time becomes a reality and as multiple satellite constellations are integrated. With this will come many challenges as well as opportunities to learn more about the Earth system.

The VLBI community is focused on a new operational model entitled VLBI2010 which promises significant advances in precision with lower cost operations. Similarly the satellite laser ranging (ILRS) community has a modernization effort underway entitled SLR2000. Refer to the respective IGS service websites for additional information.

The GRACE satellites represent a revolutionary new technique for which applications are still being discovered. Next-generation satellites of this type (in the planning stage) will provide significant improvements and will likely lead to many new benefits.

Acknowledgment

This research was carried out at the Jet Propulsion Laboratory, California Institute of Technology, under a contract with the NASA.

Bibliography

Chao, B., 2003. Geodesy is not just for static measurements any more. *Eos*, **84**(16), 145–150.
http://earth-info.nga.mil/GandG/publications/tr8350.2/wgs84fin.pdf.
McCarthy D. D., Petit G., (eds.), 2003. IERS Conventions. ftp://tai.bipm.org/iers/conv2003/tn32.pdf.
Mireault, Y., Tetreault, P., Lahaye, F., Heroux, P., and Kouba, J., 2008. Online precise point positioning: a new, timely service from natural resources Canada. *GPS World*, **19**(9), 59–64.
NIMA World Geodetic System Technical Report, Amendment 1, 2000. NIMA TR8350.2.
Velicogna, I., and Wahr, J., 2006. Acceleration of Greenland ice mass loss in spring 2004. *Nature*, **443**, 329–331.

Cross-references

Geological Mapping Using Earth's Magnetic Field
Global Earth Observation System of Systems (GEOSS)
Ionospheric Effects on the Propagation of Electromagnetic Waves
Limb Sounding, Atmospheric
Sea Surface Wind/Stress Vector
Solid Earth Mass Transport
Subsidence
Water Vapor

GEOLOGICAL MAPPING USING EARTH'S MAGNETIC FIELD

Vernon H. Singhroy[1] and Mark Pilkington[2]
[1]Applications Development Section, Natural Resources Canada, Canada Centre for Remote Sensing, Ottawa, ON, Canada
[2]Geological Survey of Canada, Ottawa, ON, Canada

Definition and introduction

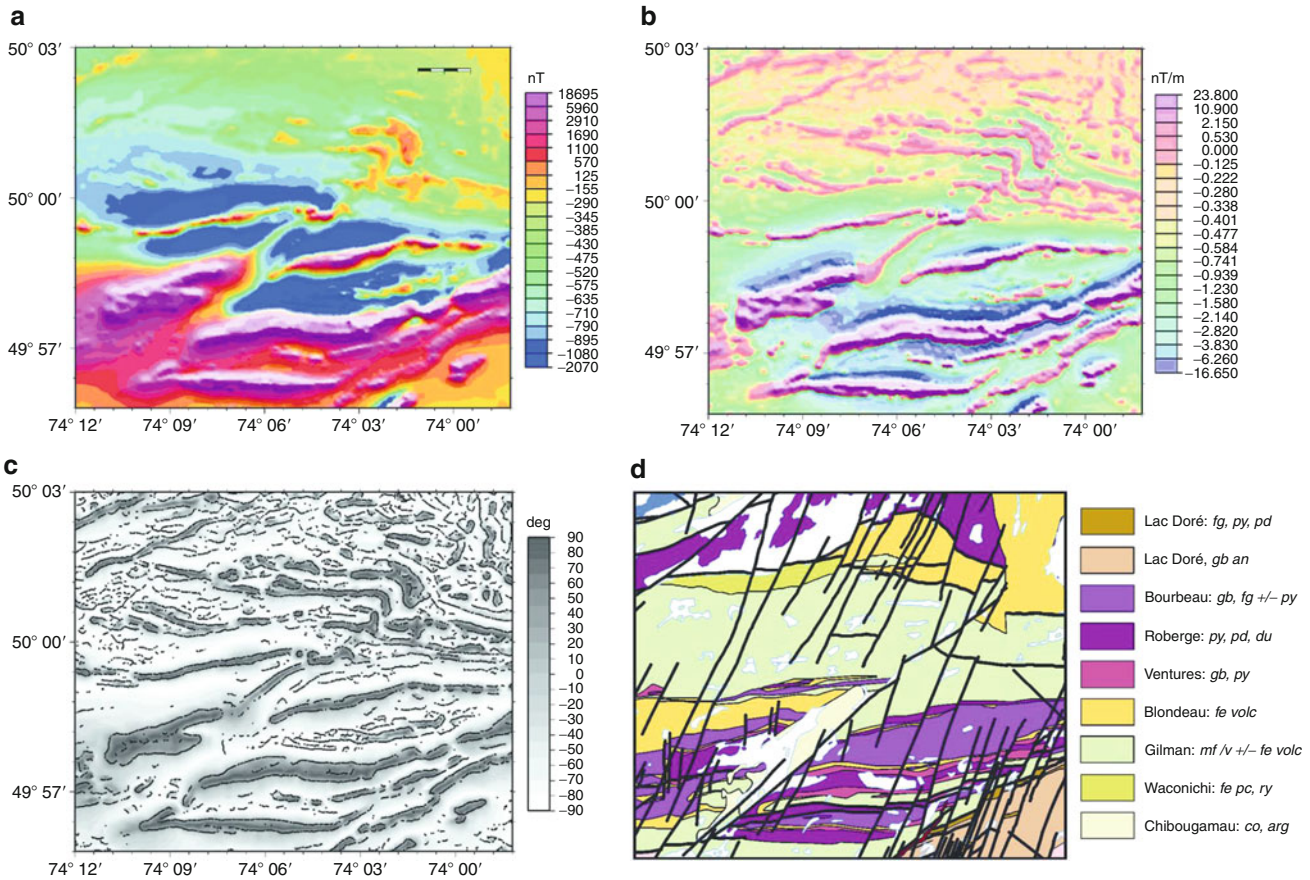
The Earth's magnetic field comprises two main components: the core field, caused by motions of the electrically conductive, liquid iron mixture in the outer core, and the crustal field, caused by magnetic minerals within rocks in the Earth's crust. Geological applications focus on the crustal field, which has spatial variations with wavelengths <1,000 km and amplitudes generally <1,000 nT (nanotesla). The separation of core and crustal field is effected through removal of a mathematical model of the core field known as the IGRF (International Geomagnetic Reference Field). The IGRF is based on mainly magnetic observatory, ship- and satellite-based magnetometer measurements and is updated every 5 years. Once isolated, the (crustal) magnetic field can be used to infer the magnetic properties of the rocks within the Earth's crust and hence provide indications of geological structure and composition. Measurements of the Earth's magnetic field for mapping purposes are carried out on a variety of moving platforms: satellites, ships, airplanes, helicopters, and vehicles. Ground-based, handheld magnetometers may also be used for small-scale surveys. The type of measurement platform will be determined by the size of survey area, its topography, location, and the mapping goals. For geological mapping purposes, airborne (fixed-wing aircraft) surveys are the most cost-effective and informative approach for data gathering.

Geological mapping applications of magnetic field data

The fundamental tenet of geological mapping with magnetic fields is *trends of features in the magnetic field mimic those of the outcropping geology*. This is especially the case where the geology consists of magnetic metamorphic and igneous lithologies (Figure 1). Since surficial cover such as vegetation, soil, desert sands, glacial till, water, and many sedimentary rocks is effectively nonmagnetic, the observed magnetic field can be directly related to the underlying rock types. Although structural features and changes in lithology are reflected in the magnetic field maps, a one-to-one connection between some geological feature, e.g., a pluton, and the resulting magnetic field is not always guaranteed. There must be a contrast in the physical property (i.e., magnetization) between the pluton and its surroundings that is large enough to produce a magnetic anomaly which can be detected by a magnetic survey.

The identification of lithology from magnetic field maps can also be ambiguous as the magnetic properties of a given rock type are controlled by the amount and composition of the accessory minerals present, which only constitute about 1.5 % of crustal minerals (Clark and Emerson, 1991). Accessory minerals are generally ignored in petrological classifications (based on silicate mineralogy), and so specific lithologies may exhibit a wide range of magnetic properties. Nevertheless, individual anomalies may signify the presence of a specific lithology or formation. Patterns of anomalies and the shapes of groups of anomalies may reflect certain lithologies, formations, or geological domains. Since there is ambiguity present in what lithology produces a given anomaly, it is desirable to have some knowledge of the magnetizations of the rock types likely encountered in a given area. This knowledge can come from known (published) compilations of physical rock properties, or preferably, from measurements on samples taken from the area of interest or from adjacent areas.

The magnetization of a rock consists of two components: induced and remanent. Induced magnetization is quantified by the rock's susceptibility and is only produced in the presence of the Earth's magnetic field. Remanent magnetization is a permanent magnetization acquired at some point in the rock's history. Both types of magnetization may be present in a rock, but magnetic anomalies are predominantly caused by induced magnetization. The main contributor to the induced magnetization is magnetite (Fe_3O_4), a member of the magnetite-ulvospinel solid-solution series known as



Geological Mapping Using Earth's Magnetic Field, Figure 1 (**a**) Magnetic field over portion of Abitibi Greenstone Belt, Quebec, Canada. (**b**) Magnetic vertical gradient. (**c**) Tilt of magnetic field with superimposed calculated contact locations. (**d**) Geological map: *fg* ferrogabbro, *py* pyroxenite, *pd* peridotite, *gb an* gabbroic anorthosite, *gb* gabbro, *du* dunite, *fe volc* felsic volcaniclastics, *mf lv* mafic lavas, felsic pyroclastics, *co* conglomerate, arg argillite, *ry* rhyolite. *Solid black lines* faults.

titanomagnetites (Reynolds et al., 1990). The susceptibility of titanomagnetites is effectively zero for ulvospinel contents greater than about 70 % but is significant and fairly uniform for the rest of the series. Other magnetic minerals of lesser importance are hematite (αFe_2O_3), common in oxidized igneous rocks and sediments formed in oxidizing conditions, and pyrrhotite (Fe_7S_8), found in basic igneous rocks, low- to medium-grade metamorphic rocks, and sedimentary rocks.

Interpretation methods

How magnetic maps are interpreted depends on the goal of the study, but the aim is always to characterize and classify the magnetic sources that produce the observed magnetic field. Interpretation methods can be divided broadly into two categories: quantitative and qualitative. The difference between the two is that the former aims to estimate the values of some or all of the physical quantities (e.g., depth, shape, magnetization) of the magnetic sources present, while the latter attempts to classify the magnetic field (or processed versions of it) into regions that can be interpreted in terms of lithology and structure. For geological mapping applications, qualitative interpretation methods are the most appropriate since the aim is to take the magnetic field data and convert them into a pseudo-geological map. Although automated methods for classifying images are available, most magnetic map interpretations of this type are done manually. Quantitative approaches are usually brought into play afterward, to investigate certain anomalies of interest by, e.g., estimating the dip or depth of a given unit.

For geological mapping applications, the magnetic field map (or image) is not the only piece of information available for interpretation. Numerous processing methods can be applied to the magnetic field data to produce transformed data (or images) where certain features that will aid in the mapping process are preferentially enhanced. Several commonly used transformations are described in the following section.

Magnetic maps

Magnetic field maps

An example of a magnetic field map, covering a portion of the Abitibi Greenstone Belt, Quebec, is presented in Figure 1a. The patterns of colors depicted on the map show the strength of the measured magnetic field and hence reflect how strongly magnetized the rocks are; reds denote magnetic highs and blues represent magnetic lows. Trends in the anomalies reflect the distribution of magnetic minerals (predominantly magnetite) in the surface geology in the area.

The character of individual anomalies is dependent on the shape of the magnetized region or source, its depth, and its magnetization. As the distance from the measurement point to the magnetized source increases, the resulting anomaly decreases in amplitude and increases in width or wavelength. Therefore, the field observed in Figure 1a is mostly caused by the rocks close to or at the Earth's surface, i.e., the outcropping geology. There will be (generally weaker) contributions from deeper magnetized sources, i.e., buried geological units, but these effects can be considered unimportant when using the data for geological mapping purposes. Figure 1a shows clearly the predominant east–west striking geological trends in this region (see Figure 1d for the geology), with the more magnetic units (stronger magnetic field) occurring in the south. Some broad correlations can be made: The highest-amplitude magnetic highs on the map are caused by peridotitic rocks of the Roberge formation. Smaller-amplitude highs are associated with ferrogabbros occurring within anorthosite of the Lac Doré formation in the southeast. Magnetic lows or subdued zones are generally associated with the felsic Blondeau and Waconichi formations. On the other hand, some mafic lithologies do not show a significant magnetic effect, e.g., the Bourbeau and Ventures rocks. Often, more mafic lithologies have more Ti-rich magnetites which results in lower susceptibilities. Within the Gilman formation (volcanics), many magnetic trends can be seen suggesting possible subdivision into mafic and felsic flows. The correlation between geology and magnetic anomalies is made clearer through transformations of the magnetic field as outlined below and shown in Figure 1b and c.

Magnetic vertical gradient maps

The vertical gradient (or first vertical derivative) of magnetic fields has been used since the 1960s to improve the resolution of features on a magnetic map. The vertical gradient can be measured directly when surveying but is more commonly calculated from the magnetic field. Calculating the vertical gradient is similar to high-pass filtering the data, so long-wavelength anomalies (usually associated with deeper features) are suppressed and gradient anomalies produced by near-surface geological features are emphasized. Closely spaced geological units are better resolved because the vertical gradient of a given anomaly is narrower than its magnetic field counterpart. Consequently, the vertical gradient map is a standard product used in mapping. As with many enhancement procedures, the problem of noise amplification is present when gradients are calculated; thus, some filtering (usually low-pass) may be required.

Figure 1b shows the magnetic vertical gradient over the study area. Note that the high-amplitude E–W trending anomalies in the southern part of the map are now better resolved; the anomaly widths have decreased and in some cases a single magnetic field anomaly in Figure 1a is now resolved into two separate gradient anomalies (e.g., at 74° 09′W 49° 58′N).

Tilt maps

The tilt is defined as the arctangent of the ratio of the vertical gradient to the magnitude of horizontal gradient of the magnetic field (Miller and Singh, 1994). The horizontal gradient achieves a maximum near or over source edges and tends to zero elsewhere. The vertical gradient is positive over the source, zero over or near source edges, and negative elsewhere. As a result, the tilt of the magnetic field tends to be positive over sources bodies, zero over source edges, and negative elsewhere. One advantage of the use of the tilt is that, being a ratio, anomalies due to both weak and strongly magnetized sources are given equal weight. Hence, subtle (and often short-wavelength) anomalies are enhanced. Linear and quasi-linear geological features such as faults appear as minima, whereas dykes often appear as maxima in the tilt map.

Figure 1c shows the tilt of the magnetic field over the study area as a grey shade map (the superimposed dots are discussed below). Note that the broad change from high to low magnetic field values going northward (Figure 1a) is not apparent on the tilt map. Also, the prominence of high-amplitude (in the south) versus low-amplitude (in the north) anomalies is removed, and all anomalies have similar expressions (amplitudes) in the tilt map. Both of these effects are due to the balancing effect of calculating a ratio of quantities based on the magnetic field.

Contact maps

Mapping the locations of lateral magnetization contrasts, i.e., the edges of magnetic sources, is one of the most basic and useful applications of magnetic survey data in geological mapping. Delineating the extent of units with similar magnetic properties follows the geological mapping approach of dividing the surface into rocks having similar properties. Whether or not the magnetization changes exactly mimic lithological changes, mapping lateral contrasts provides invaluable information on structural regimes, and deformation styles and trends. There is more than one method of contact mapping, each with its own pros and cons. All methods aim to define contact locations based on maximizing a given function over the source body edge. The locations of these maxima can then be plotted over the magnetic field, vertical derivative, etc., to enhance the interpretation. Details on how the different methods are implemented are given in Pilkington and

Keating (2004). No one method has all of the desirable properties for a reliable contact mapper; hence, using more than one approach is always recommended since colocated solutions from different methods provide increased confidence in the reliability of a given contact location and lessen the adverse effects of source magnetization direction, geological dip, and depth extent.

Figure 1c shows contact locations as black dots superimposed over the tilt of the magnetic field. These contacts were computed from the absolute value of the horizontal gradients of the magnetic field in the N–S and E–W directions (Pilkington and Keating, 2004).

Analytic signal maps

The analytic signal is defined as the absolute value of the square root of the sum of the two horizontal and the vertical derivatives of the total magnetic field (Roest et al., 1992). The main property of the analytic signal is it reaches a maximum value directly over lateral magnetization contrasts (contacts) and that this maximum's location is independent of the orientation of the Earth's magnetic field and of the magnetization direction of the bodies. The benefit of these properties is that no assumptions about the nature of the magnetization of the sources are required. However, this is only true for 2D structures. For 3D structures (e.g., intrusives) the shape of the analytic signal is almost but not completely independent of the Earth's field and magnetization directions. The analytic signal gives maxima directly over contacts and sheets (thin dykes). For a thick dyke, two series of maxima slightly offset outward from their true location are observed. As a result, dykes can appear a bit thicker than they really are. The use of derivatives to define the analytic signal accentuates near-surface effects and noise in the data. As a result, the analytic signal maxima often form discontinuous trends, even along seemingly continuous total field anomalies. Nevertheless, the location of analytic signal maxima can be taken to indicate the position of contrasts between differently magnetized bodies and, as such, can be used to locate geological contacts (cf., Pilkington and Keating, 2004).

Shaded relief magnetic maps

Sometimes the intensity and spatial distribution of magnetic variations within a region are such that a magnetic field image may appear somewhat muted, containing little information. Such images can be enhanced using a technique known as shading or shaded relief. The process simulates the effects of light directed on the three-dimensional surface defined by the grid of magnetic anomaly values. The intensity of light reflected back from a slope on the surface will vary according to the steepness of the slope. The proportion of illuminating light reflected back is known as the reflectance and attains a maximum for slopes that are perpendicular to the direction of the light. The overall effect is one of areas of varying brightness and areas of shadow, an effect very similar to that created by the sun shining over a range of hills. The light direction and its inclination from the horizontal can be varied, thereby providing different images of the same data set. Depending on the direction and inclination, certain features may be enhanced or suppressed. Generally, magnetic features are enhanced when the light direction is oriented at right angles to them.

Magnetic and radar fused image

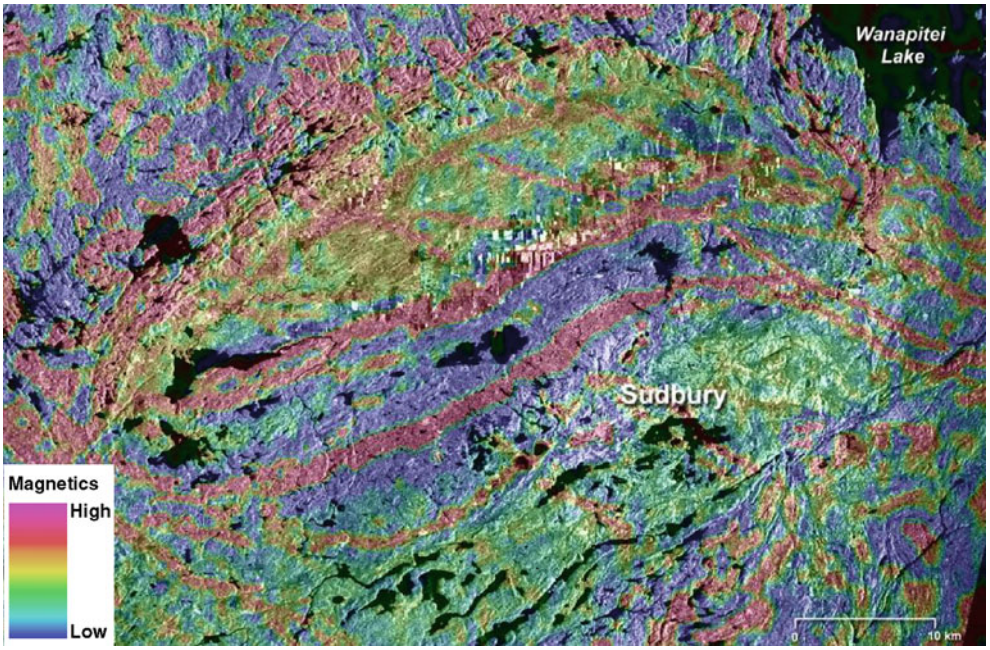
The fusion of magnetic and radar images has been used to assist in geological mapping and mineral exploration (Singhroy and Molch, 2004). An example of the uses of image fusion techniques is presented.

The Sudbury Basin is one of Canada's richest mining areas, with world-class mineral deposits and the world's oldest (2 billion years), largest, and best-exposed meteorite impact structure. Figure 2 shows the 300 km diameter elliptical impact structure known as the Sudbury. Over the past 100 years, about $135 billion of nickel and copper ores have been mined from more than 90 mines distributed around the rim of the Sudbury Basin. Current production is about $2 billion a year.

Much remains to be learned about the structural evolution of the Sudbury Basin and the mineral occurrences around it. The RADARSAT-1 image provides an excellent synoptic view of the topography and structural features of the Sudbury area. Structure is expressed by differential erosion controlled by faults, dykes, and different rock types. The fused RADARSAT and magnetic images shown in Figure 2 are produced from special image processing techniques using the Intensity-Hue-Saturation (IHS) integration procedure (Singhroy and Molch, 2004). The RADARSAT-1 image provides terrain information and the surface expressions of structures seen on the image, which appear as a network of linear features. The low-resolution vertical magnetic gradient image (100 m), obtained from the Geological Survey of Canada, shows geological units as expressed from their magnetic signatures. The variation in the concentration of magnetic minerals can be used to map different rock types. Therefore, the combined RADARSAT-1 and the magnetic image is useful in the geological mapping of structures and rock types. The circular area in red (Figure 2) represents the high magnetic signature of the Sudbury Igneous complex. Magma-filled dykes representing NE–SW trending lineaments also have a high magnetic signature. The other colors shown on the fused image correspond to different rock types with different magnetic signatures. These type of fused image maps are used by geologists and mining companies to facilitate geological mapping and mineral exploration projects.

Interpreting magnetic field maps

Geological mapping from magnetic data is essentially an exercise in pattern recognition. The basic approach is to divide the region into areas that have similar character, based on the amplitude, shapes, and textures seen on the magnetic maps. Areas can be initially divided up based on the



Geological Mapping Using Earth's Magnetic Field, Figure 2 Fusion of RADARSAT and vertical gradient magnetic image of the Sudbury Basin impact structure, Ontario, Canada. This type of image map is used to map geological structures and rock types for mineral exploration (Singhroy and Molch, 2004).

Geological Mapping Using Earth's Magnetic Field, Table 1 Magnetic field products and their utility for geological mapping applications

Quantity	Use	Mapping application
Magnetic field	General	Lithology + structure
Shaded relief	Emphasize specific directional trends + fine detail	Mainly structure
Vertical magnetic gradient	Emphasize fine detail	Lithology + structure
Tilt	Emphasize low-amplitude areas	Mainly structure
Analytic signal	Highlight contacts, remove magnetic field inclination effects	Mainly contacts
Contact	Map contacts	Structure + contacts

amplitude of the anomalies, i.e., separating the area into regions where amplitudes are high and nonmagnetic regions where amplitudes are low. Negative anomalies on maps do not necessarily imply negative or reversely magnetized rocks; they generally just show areas of weakly magnetized lithologies. Once amplitude information is taken into account, then the patterns/textures of trends/features within these areas can be evaluated. Different "units" can be distinguished based on the strike of anomaly trends by grouping together regions having similar strike (assuming a constant direction is apparent). Similarly, regions can be defined on the basis of similar texture or pattern of anomalies. The resulting areas do not yet have a geological interpretation; they are merely areas defined by similarities in amplitude, trends, and patterns of anomalies. Broad geological inferences can, however, be made. Outlines of the subdivisions may indicate lithology or structure, e.g., ovoid, magnetic regions suggest intrusions, particularly if they interrupt the surrounding dominant strike direction(s). Sinuous trends suggest ductile deformation, while truncation of trends or region outlines would suggest brittle deformation and faulting. Inferring rock type in a region defined by the magnetic data requires additional independent geological knowledge, usually from geological mapping within or adjacent to the area of interest. In this process, magnetic anomaly characteristics can be compared and calibrated with the known geology and correlations between the two can be established. Importantly, this comparison may dictate which lithologies can be distinguished and which cannot. It is more likely that more than one lithology will produce similar strength anomalies, since magnetic properties may not vary significantly between different rock types, so this ambiguity will remain in the interpreted map.

No one type of magnetic map will provide all the information to make a complete interpretation. Table 1 lists the most useful map products and what their main

applications are. Clearly, there is overlap between the different products, but since they all provide somewhat independent representations of the original data, it is recommended that all that are available be used in an interpretation.

Bibliography

Clark, D. A., and Emerson, D. W., 1991. Notes on rock magnetization characteristics in applied geophysical studies. *Exploration Geophysics*, **22**, 547–555.

Miller, H. G., and Singh, V., 1994. Potential field tilt – a new concept for location of potential field sources. *Journal of Applied Geophysics*, **32**, 213–217.

Pilkington, M., and Keating, P., 2004. Contact mapping from gridded magnetic data – a comparison of techniques. *Exploration Geophysics*, **35**, 306–311.

Reynolds, R. L., Rosenbaum, J. G., Hudson, M. R., and Fishman, N. S., 1990. Rock magnetism, the distribution of magnetic minerals in the Earth's crust, and aeromagnetic anomalies. In Hanna, W. F. (ed.), *Geologic Applications of Modern Aeromagnetic Surveys*. Denver, CO: U.S. Geological Survey. U.S. Geological Survey Bulletin 1924, pp. 24–45.

Roest, W. R., Verhoef, J., and Pilkington, M., 1992. Magnetic interpretation using the 3-D analytic signal. *Geophysics*, **57**, 116–125.

Singhroy, V., and Molch, K., 2004. Geological applications of RADARSAT-2. *Canadian Journal of Remote Sensing*, **30**(6), 893–902.

GEOMORPHOLOGY

David Pieri
Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA, USA

Synonyms

Landscape geology; Physical geography

Definitions

Uniformitarianism. The assumption that processes operating at present are the same processes that have operated in the past – essentially that "the present is the key to the past." The utility of the uniformitarianist paradigm for geomorphology depends on the assumption of a roughly constant range of surface environments, which is not true over on timescales comparable to the age of the planet.

Baselevel. The lowest level to which a stream, river, or groundwater can flow. For large rivers it is usually at the river mouth, typically sea level. For enclosed drainage basins, baselevel can be elevated or "perched." For most landscapes baselevel limits the depth of erosion.

Pediment. Gently sloping surface of bedrock, often disjoint from steeply sloping mountain foothills, which forms a steady-state equilibrium surface of transport for laminar sheetwash flows. They are often veneered by alluvial gravels or soils.

Scarp. Sometimes also called an escarpment, it is a steep difference in slope, forming a cliff that is often the boundary between different physiographic units within a landscape due to differential susceptibility to erosion. A scarp may also form because of differential uplift along a fault.

Dendrochronology. Also called "tree-ring dating," is a method of scientific dating of timber to specific years by counting the annual growth patterns of tree rings. As a technique, it has proved effective in the calibration of ages determined by radiocarbon dating over approximately the last 8,000–10,000 years.

Bioturbation. The mixing of sediment grains by the action of plants and animals in the soil.

Introduction

Geomorphology is the study of the configuration and history of the surface of the earth and the processes that shape it. As a scientific discipline, its goals are to understand how and why the landscape has come to look the way it does, through an understanding of dynamic physical and chemical processes through time, and to predict how they will shape its form in the future. Its practitioners, *geomorphologists*, traditionally have approached this challenge with empirical field observations, laboratory and field experiments, and by using numerical and analytical modeling. Modern geomorphic techniques include computer graphics visualization and airborne and orbital remote sensing. While in much of the world, especially the United States, geomorphology is taught and practiced as part of the science of geology, elsewhere, especially in the United Kingdom and former commonwealth countries, its practice and teaching has resided within the field of physical geography. Environmental and civil engineering fields also draw heavily from and contribute to the field. Geomorphology is also often central to the analysis of geological hazards. Most recently, *planetary geomorphologists* have turned their attention to studies of the surfaces of other planets and their satellites as the related field of *comparative planetology* has developed, enabled by spacecraft remote-sensing observations over the last 40 years or more.

While *geomorphological* analyses can be traced back as far as the ancient Chinese literature, the first modern Western geomorphic model was developed by the English pioneer in the field, William Morris Davis (1850–1934), who modeled the cycle of landscape erosion, following the earlier *uniformitarianist* thesis of James Hutton (1726–1797). The classic Davisian model involves riverine downcutting (baselevel lowering), followed by general erosion of the surrounding adjacent landscape to a lower elevation (the peneplain), and followed by uplift to restart the cycle. While conceptually valuable, its gross oversimplification limits its utility as a practical tool. Alternative models by Albrecht Penck (Austria, 1858–1945) and son Walther (Austria, 1888–1923; simultaneous uplift and erosion) and Lester King (South Africa, 1907–1989;

pedimentation via scarp retreat – e.g., King, 1967), competed with the classic Davisian cycle approach, but none were universally applicable. Current geomorphic models embrace the geological paradigm of plate tectonics as one of the underlying forces driving the geomorphic engine as it sculpts the earth. For the other planets, geomorphic processes have played out under a variety of exotic (compared to the earth) boundary conditions, including differing gravitational fields, surface temperature regimes (e.g., cryogenic or hyper-thermal), atmospheric conditions, impact rates, and unique tectonic regimes. G. K. Gilbert (1943–1918), the important early American geomorphologist (dynamic equilibrium model), also was one of the first to apply geomorphology to another planetary surface in his analyses of lunar craters from telescopic observations.

During the so-called quantitative revolution in geomorphology in the middle twentieth century, analyses of drainage networks as a fundamental landscape element were prominent. The fundamental modern contribution to the analyses of drainage networks was made by US Geological Survey hydrologist Robert Horton (1875–1945) in his analyses of the hierarchy of branching stream networks, whereby the smallest unbranched tributaries were assigned order unity, and order added arithmetically downstream throughout the network. Subsequently, Horton's model was modified by the American geologist Arthur Strahler (1918–2002) to include the concept of stream power (the higher the order, the higher the power). Subsequent work by Adrian Scheidegger and Ronald Shreve addressed mathematical inconsistencies in the Horton and Strahler schemes, and Shreve demonstrated the topological randomness of drainage basins, from which much of the systematic character of the stream ordering systems devolved. Time-independent steady-state equilibrium for humid temperate drainage basins, proposed by John Hack (1913–1991), stood somewhat in contrast to the erosion cycle models.

Modern geomorphic work has included an emphasis on quantitative techniques in the field and laboratory and on the physical connection between process and form, and particularly the testing of ideas about the *time-dependent* evolution of landscapes (Ritter et al., 2002). The advent of modern radiometric, dendrochronologic, thermal luminescent, and other absolute dating techniques has made this possible. Tectonic geomorphology has also taken on more importance, particularly over the last 40 years or so as plate tectonics has become the dominant paradigm for earth crustal evolution and dynamics.

Geomorphology of the earth

When seen from space, 75 % of the Earth's surface is dominated by its oceans of liquid water, while the remaining 25 % of nonmarine subaerial land, lying mainly in the Northern Hemisphere, where most of the world's population lives, has been the subject of nearly all historical geological and *geomorphological* study. The Southern Hemisphere is dominated by oceans, some subaerial continental and archipelago land masses (mainly parts of Africa, South America, Southeast Asia, and Australia), and the large, mainly subglacial, island continent of Antarctica.

Despite 200 years of geological and geographical scientific practice, it has only been during the last 40 years or so that detailed mapping and geophysical explorations of the submarine land surface revealed evidence of "plate tectonics," namely, the presence of intensely volcanic mid-oceanic ridges and their associated parallel-paired geomagnetic domains, along with submarine trenches along continental or island-arc margins (Heezen and Tharp, 1977). Mid-oceanic ridges are where new volcanically generated material accretes to oceanic plates, where trenches harbor deep subduction, and where oceanic crust is consumed beneath other overriding crustal plates. Less dense continental plates form over time as more silica-rich island-arc material generated volcanically at oceanic plate margins accretes, eventually becoming persistent continental cratons, resistant to subduction because of their lower density, thus having the propensity to "float" over more magnesium and iron-rich oceanic plates. Oceanic plates tend to be mobile and are subducted, thus having typical lifetimes of 100–200 Myr, while continental landmasses may persist up to 2–3 Byr, preserving evidence of much of the history of the earth, including some ancient landscapes. Tectonic plates are the Earth's most fundamental geomorphic units.

Geomorphically, submarine oceanic basins areally dominate terrestrial landforms. The main features of oceanic basins are the oceanic ridge and rise systems. Sometimes locally very rugged, they can rise to several km above the average oceanic depth. In the Atlantic Ocean, rise systems exhibit a central rift valley that is at the center of the rise, which is not always true in the Pacific. Older crust within oceanic basins can have gently rolling abyssal hills. Predominantly flat abyssal plains that stretch for thousands of kilometers are usually also covered with accumulated marine sediment. In places they are punctuated by seamounts – conical topographic rises often topped by coral lagoons or residing just beneath the oceans' surface. These are undersea volcanoes formed as island arcs or midplate hot spots, such as the mid-Pacific Emperor Seamount chain, which terminates in the Hawaiian Islands at its southeastern end. Though areally smaller, oceanic margins are another important submarine landform province. "Atlantic style" continental margins tend to exhibit substantial ancient sediment accumulations and a shelf-slope-rise overall morphology, which probably represents submerged subaerial landscapes remnant from the last Ice Age. Continental shelves are usually less than about 100 km in width and have very shallow ($\sim 0.1^{\circ}$) topographic slopes. Submarine canyons (also probably remnant from the last Ice Age) can deeply cut the continental shelf and slope and terminate in broad submarine sediment fan deposits at the seaward canyon outlet. "Pacific style" oceanic margins, present along the margins

of the Pacific Rim, are even narrower, consisting of a short shelf and slope that can terminate into deep submarine trenches, including subduction zones (e.g., South America, Kamchatka, Aleutian, and Kurile Islands), up to 10 km deep. Shallower "back-arc" basins occur on the overriding plate behind island arcs (e.g., Sea of Okhotsk).

The subjects of classic geomorphological investigations are the Earth's "subaerial" landscapes. While mainly continental, some important subaerial landscapes (particularly volcanic ones, e.g., Hawaii, Galapagos Islands) exist on oceanic islands. Terrestrial subaerial landform suites are the classic landscapes studied in geomorphology (Snead, 1980). Drainage basins exist at nearly all spatial scales and range from very active drainage basins in humid and semiarid climatic zones to only occasionally active or relict drainages in arid zones. Drainage basin morphology is strongly influenced by a combination of mountain building related to plate tectonics and prevailing climatic regimes (Schumm, 2005).

Land surface volcanic processes produce characteristic landscapes in all climates. They occur mainly at plate boundaries, with isolated oceanic (e.g., Hawaii) and continental (e.g., the San Francisco volcanic field in Northern Arizona, the Columbia and Snake River volcanic plains in the US Pacific Northwest, the Deccan Traps in India) examples occurring away from plate boundaries. Central vent volcanic landforms range from strato-cone volcanic structures to large collapse and resurgent calderas (e.g., Krakatau, Indonesia). More areally extensive and lower subaerial shield volcanoes, formed by more fluid lavas, exist in places like the Hawaiian Islands or the Galapagos Islands, for example (Francis and Oppenheimer, 2004).

The expenditure of energy in the landscape erodes the earth's surface and transports sediment. Destructional processes, such as rainfall-driven runoff and streamflow, are essentially exogenic processes (i.e., the energy that drives the evaporation and precipitation, and winds that transport water vapor, comes from an exterior source – the Sun). Constructive landscape processes are mostly endogenic – derived from the internal dynamo processes that drive plate tectonics and subsequent mountain building (Bloom, 1998).

In familiar ways, such destructional geomorphic processes work to reduce the "topographic disequilibria" that constructive landscapes represent. For instance, the relatively low and ancient Appalachian Mountains were pushed up during one of the collisions between the North American and European continental landmasses and probably exhibited relief comparable to the Himalayan mountain range today. The former topographic relief represented strong disequilibria and thus potential energy to be converted to kinetic energy by erosive runoff. Once continental collision processes slowed to a stop and tectonic uplift ceased, erosion and surface transport processes (such as rainfall, associated runoff, snowfall, and glaciation) over only a few tens of millions of years reduced the proto-Appalachian Mountains to the gently sloping and relatively low-relief state we see today. Volcanogenic landscapes are good examples of how competition between destructive and constructive processes sculpts the earth. For example, Mt. Fuji in Japan is an active volcano that erupts approximately every 100–150 years. Fuji's classic symmetrical conical shape is the result of volcanic eruptions that deposit material, on the average, faster than it can be eroded. When Fuji enters a dormant phase, it will become deeply incised by stream erosion, and it will rapidly lose its symmetric conical profile over a geologically short interval.

Weathering is another key geomorphic process through which consolidated material is broken down into constituent grains by chemical and physical means. Chemical weathering occurs when natural acids act on carbonates, such as limestone or sandstones, releasing the residual silicate grains. Mechanical weathering occurs when the hydrostatic pressure of ice in freeze-thaw cycles overcomes the brittle strength thresholds in rock at microscopic and macroscopic scales. Likewise, salt crystals can exert mechanical energy to break up rocks and can chemically weather rocks, mostly in arid areas. Mineral oxidation, particularly iron-containing minerals, is yet another form of chemical weathering. Biological weathering can take the form of chemical weathering by biogenic acids, particularly in tropical areas. It can also occur mechanically by bioturbation of soils and sediments, as well as by the physical pressure of root and stem turgor in cracks and fissures within solid rock. It is significant that all three main forms of weathering are enhanced or enabled by the presence of water.

Finally, mass wasting is perhaps the most dramatic form of nonvolcanic landscape alteration. The term "mass wasting" is applied to processes such as landslides; creep, snow, and debris avalanches; submarine slides and slumps; volcano-tectonic sector collapses; and scour related to the action of glaciers. While such processes tend to affect only a small part of the Earth's surface, when they occur near populated areas, their effects can be devastating.

New technology and global geomorphology

Stimulated by the desire to understand the Earth's history in the nineteenth century, geomorphology evolved from a highly descriptive natural history to a modern physical science by the progressive adoption of analytical techniques and technologies as they became applicable and available for landscape-related science. New airborne survey techniques developed as a result of World War II provided comprehensive information on the topography and near-subsurface structure and composition of landscapes. Improved aerial photography and photogrammetry greatly impacted quantitative approaches to geomorphology. The Newtonian approaches of Strahler (1950, 1952), Bagnold (1941), Horton (1945), and Leopold and Maddock (1953), emphasizing direct observables, and the advent of absolute dating methods, especially radiocarbon methods (Libby, 1955) for time periods more recent than 30Ka BP, had a profound effect on the field, finally allowing accurate estimations of surface process rates. The

expansion of modern technology within geomorphology also facilitated quantitative statistical approaches to landscape analyses (e.g., Schumm, 1956; Melton, 1958; Lubowe, 1964; Shreve, 1975; and Pieri, 1984) and allowed more complex analyses, such as attempts to illuminate the self-organization of landscapes (e.g., Phillips, 1995, 2003, 2009).

Most recently given the proliferation of large (10–100 Tb) databases produced by a new generation of late twentieth and early twenty-first century orbital observational instruments (e.g., Pieri and Abrams, 2004), high spatial and spectral resolution orbital imaging at wavelengths from visible through thermal infrared to microwave detailed stereophotogrammetric, radar, and LIDAR modeling of land surface topography, and relatively frequent temporal sampling are now available. Modern earth orbital missions routinely allow detailed physical and chemical characterization of the Earth's surface at 10–100 m scales, or better (e.g., ASTER, SRTM, EO-1, Radarsat, ICESAT, and others). The emergence and ready availability of remote-sensing technology and its influence on the science of geomorphology have fostered the modern view geomorphology as a "system science" (Church, 2010) with an integrated global perspective, especially among younger *geomorphologists* who are technologically adept.

These new technologies are now being used to monitor and study the dynamics and history of earth surface processes, land use, and societal vulnerabilities at a variety of scales and to synergistically complement in situ field observations, across the planet. Spatial continuity of data is now available down to the centimeter scale over many large areas (Church, 2010). Global topographic databases (e.g., ASTER Global Digital Elevation Model) with resolutions of tens of meters, and for local areas LIDAR imagery, can provide data at almost granular spatial scales. The availability of such thoroughly quantitative measures of the global landscape has opened up possibilities of analytical analysis that were undreamed of just a few decades ago (Lane et al., 1993). One of the most recent remote-sensing-related technologiesthat have been of critical importance has been the Global Positioning System (GPS). It has provided unprecedented ability to determine location on the Earth's surface, typically to within meters. Especially when combined with radar interferometry, these techniques allow the dynamic nature of rapidly changing topography (e.g., pre-eruption inflation of volcanoes) to come within view. In the aggregate, this postmillennial remote-sensing and measurement technology bonanza has found important application in volcanology, glacial and periglacial processes, hydrology and fluvial geomorphology, mass wasting analyses, coastal studies, eolian studies, and geomorphology as related to climate change. Perhaps most profoundly, the precision and accuracy of such techniques now allow us to document the impact of man as the principal agent of geomorphic change on Earth (Church, 2010; Hooke, 1994, 2000).

Summary

The Earth had liquid water oceans for most of its history, has a highly mobile crust, and a dynamically convecting interior. Its surface has been constantly driven by the movement of the interior causing the periodic conglomeration and separation of continental landmasses, the opening and closing of oceans, and the construction and destruction of mountain ranges. Such globally dynamic geomorphology profoundly impacted the global climate and thus biological evolution in general, and specifically human habitation patterns and the development of civilization. Given that as a species we have been shaped and affected by (and now, ourselves, shape) the patterns and intensity of surface geological processes – geomorphic processes – it is hard to overemphasize the importance of geomorphology as a discipline that offers environmental insights significant to our continued well-being as Earth inhabitants. New emerging techniques and approaches to landscape analyses using remote sensing on a planetary scale offer additional significant advantages for unification of regional geomorphological insights into an emerging global science (Baker, 1986; Short and Blair, 1986; Church, 2010). Such knowledge will most likely be even more important as we try to cope with and mitigate the effects of global climate change on society.

Acknowledgment

This research was carried out at the Jet Propulsion Laboratory, California Institute of Technology, under a contract with the NASA.

Bibliography

Bagnold, R. A., 1941. *The physics of blown dunes*. London: Methuen, p. 265.

Baker, V. R., 1986. Introduction: regional landforms analysis. In Short, N. M., and Blair, R. W., Jr. (eds.), *Geomorphology from Space: A Global Overview of Regional Landforms*. Washington, DC: NASA Scientific and Technical Information Branch, pp. 1–26.

Bloom, A. L., 1998. *Geomorphology: A Systematic Analysis of Late Cenozoic Landforms*, 3rd edn. Upper Saddle River: Prentice Hall.

Church, M., 2010. The trajectory of geomorphology. *Progress in Physical Geography*, **34**, 265–286.

Francis, P. W., and Oppenheimer, C., 2004. *Volcanoes*. Oxford: Oxford University Press.

Heezen, B., and Tharp, M., 1977. *Panoramic Maps of the Ocean Floor*. Washington, DC: United States Navy.

Hooke, R. L. B., 1994. On the efficacy of humans as geomorphic agents. *GSA Today*, **4**, 224–225.

Hooke, R. L. B., 2000. On the history of humans as geomorphic agents. *Geology*, **28**, 843–846.

Horton, R. E., 1945. Erosional development of streams and their drainage basins: hydrophysical approach to quantitative morphology. *Geological Society of America Bulletin*, **56**, 275–370.

King, L. C., 1967. *Morphology of the Earth*, 2nd edn. Edinburgh: Oliver and Boyd.

Lane, S. N., Richards, K. S., and Chandler, J. H., 1993. Developments in photogrammetry: the geomorphic potential. *Progress in Physical Geography*, **17**, 306–328.

Leopold, L. B., and Maddock, T. 1953. The hydraulic geometry of stream channels and some physiographic implications. Reston, VA: United States Geological Survey. Professional Paper 252, 57 pp.
Libby, W. F., 1955. *Radiocarbon Dating*, 2nd edn. Chicago: University of Chicago Press, p. 175.
Lubowe, J. K., 1964. Stream junction angles in dendritic drainage pattern. *American Journal of Science*, **262**, 325.
Melton, M. A., 1958. Geometric properties of mature drainage systems and their representation in an E4 phase space. *Journal of Geology*, **66**, 35–56.
Phillips, J. D., 1995. Self-organization and landscape development. *Progress in Physical Geography*, **19**, 309–321.
Phillips, J. D., 2003. Sources of nonlinearity and complexity in geomorphic systems. *Progress in Physical Geography*, **27**, 1–23.
Phillips, J. D., 2009. Landscape evolution space and the relative importance of geomorphic processes and controls. *Geomorphology*, **109**, 79–85.
Pieri, D. C., 1984. Junction angles in drainage networks. *Journal of Geophysical Research*, **89**, 6878–6884 (NB8).
Pieri, D., and Abrams, M., 2004. ASTER watches the world's volcanoes: a new paradigm for volcanological observations from orbit. *Journal of Volcanology and Geothermal Research*, **135** (1–2), 13–28.
Ritter, D. F., Kochel, R. C., and Miller, J. R., 2002. *Process Geomorphology*, 4th edn. New York: McGraw-Hill.
Schumm, S. A., 1956. Evolution of drainage systems and slopes in badlands at Perth Amboy. *New Jersey, Geological Society of America Bulletin*, **67**, 597–646.
Schumm, S. A., 2005. *River Variability and Complexity*. New York: Cambridge University Press, p. 234.
Short, N. M., and Blair, R. W., Jr. (eds.), 1986. *Geomorphology from Space: A Global Overview of Regional Landforms*. Washington, DC: NASA Scientific and Technical Information Branch. NASA SP-486.
Shreve, R. L., 1975. The probabilistic-topologic approach to drainage-basin geomorphology. *Geology*, **3**, 527–529.
Snead, R. E., 1980. *World Atlas of Geomorphic Features*. Huntington/New York: Robert E. Krieger/Van Nostrand Reinhold.
Strahler, A. N., 1950. Equilibrium theory of erosional slopes approached by frequency distribution analysis. *American Journal of Science*, **248**, 673–696.
Strahler, A. N., 1952. Dynamic basis of geomorphology. *Geological Society of America Bulletin*, **63**, 923–937.

GEOPHYSICAL RETRIEVAL, FORWARD MODELS IN REMOTE SENSING

Eugene Ustinov
Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA, USA

Definition

Forward model. Quantitative tool for simulation of *observables* for a given set of *model parameters*.
Forward problem. System of differential equations with initial and/or boundary conditions, solution of which, the *forward solution*, is used for simulation of *observables*.
Forward solution. Solution of the *forward problem*.
Instrument model. Quantitative procedure for evaluation of observables from the *forward solution*, which models the characteristics of the instrument and the procedure for measurements.
Model parameters. Parameters of the forward problem contained in its differential equations and initial and/or boundary conditions.
Observables. Output parameters of the *forward model*, which simulate the quantities measured in the geophysical retrieval.
Sensitivities. Partial or variational derivatives of *observables* with respect to *model parameters*.

Introduction

A general qualitative description of forward models used in remote sensing, their components, and functions is provided in the entry *Geophysical Retrieval, Overview*. Two major components of the forward model are as follows:

1. *Forward problem* – provides a quantitative description of physical processes in the object of remote sensing, which result into observable quantities used in the geophysical retrieval. In a general case, the forward problem has a form of a system of equations – differential equations with initial and/or boundary conditions. Its solution, the *forward solution*, serves as an input for the procedure of computation of observables.
2. *Instrument model* – an expression converting the forward solution into observables. In a general case, it has a form of a functional, which maps the forward solution, a function, into the observable, a discrete quantity. As such, the instrument model is not a mathematical problem, which has to be solved. It is an expression, which has to be calculated.

Two major functions of the forward model are:

1. *Computation of observables* – involves application of the instrument model to the forward solution to obtain the observables.
2. *Computation of sensitivities*– involves application of the linearized forward problem and linearized instrument model to obtain sensitivities of observables to retrieval parameters.

Implementation of the second function of the forward model can be accomplished in three different ways:

1. Finite-difference approach
2. Linearization approach
3. Adjoint approach

An upper-level quantitative description of major components and functions of the forward model is provided in this entry along with illustrations based on simple analytic demo forward problem and instrument model.

Forward problem, instrument model, and computation of observables

In the general case, the forward problem represents a set of nonlinear differential equations with initial and/or boundary conditions. In high-level notation, the forward

problem can be written in the form of a generally nonlinear operator equation

$$N[X] = S \tag{1}$$

where the operator N combines all operations on the forward solution X, whereas the right-hand term S combines all inhomogeneous terms. All model parameters are contained either in the operator N or in the right-hand term S, which are some functions of these parameters.

As an example, consider a demo forward problem:

$$\begin{cases} \dfrac{\mathrm{d}X}{\mathrm{d}t} + a(t)X^2(t) = 0 \\ X(t_0) = X_0 \end{cases} \tag{2}$$

It has two model parameters: $a(t)$ and X_0. This problem can be written in the form of Eq. 1, where

$$N[X] = \frac{\mathrm{d}X}{\mathrm{d}t} + a(t)X^2(t) + \delta(t - t_0)X(t)$$

and

$$S(t) = \delta(t - t_0)X_0$$

and the Dirac δ-function enforces action of the initial condition only at $t = t_0$. This forward problem has an analytic solution

$$\begin{aligned} X(t) &= \left(\int_{t_0}^{t} a(t')\mathrm{d}t' + (X_0)^{-1} \right)^{-1} \\ &= \left(A(t) + (X_0)^{-1} \right)^{-1}, \end{aligned} \tag{3}$$

where

$$A(t) = \int_{t_0}^{t} a(t')\mathrm{d}t' \tag{4}$$

is a primitive function of $a(t)$.

Using similar high-level notations, computation of observables R from the forward solution X can be represented in the form

$$R = M[X] \tag{5}$$

where M is, in general, a nonlinear functional describing the procedure of modeling the observables. In many practical applications, this functional is linear, and the above procedure can be represented in the form of the inner product

$$R = (W, X) \tag{6}$$

where the function W describing this procedure has the same domain of arguments as the forward solution X. It has a meaning of a weighting function, "carving" the observables from the forward solution and is referred to here as *observable weighting function*.

As an example, consider an observable obtained from the solution of the demo forward problem Eq. 2 in the form

$$R = X(t_1) = \int_{t_0}^{t_1} \delta(t - t_1)X(t)\,\mathrm{d}t, \text{ and } W(t) = \delta(t - t_1) \tag{7}$$

Since the forward solution is available in analytic form Eq. 3, dependence of this observable on model parameters $a(t)$ and X_0 is also available analytically:

$$R = \left(A(t_1) + (X_0)^{-1} \right)^{-1} \tag{8}$$

This analytic dependence will be used to illustrate various applications below.

As a more realistic example, consider a forward model of radiances in the thermal spectral region at the top of a plane-parallel non-scattering planetary atmosphere. The forward problem has the form of a radiative transfer (RT) problem:

$$\begin{aligned} u\frac{\mathrm{d}I_\nu}{\mathrm{d}z} + \kappa_\nu(z)I_\nu(z,u) &= \kappa_\nu(z)B_\nu(z), \\ I_\nu(0,u) &= 0 \text{ for } u > 0, \\ I_\nu(z_0,\, u) &= \varepsilon_\nu B_{s,\nu} \text{ for } u < 0 \end{aligned} \tag{9}$$

Here, the forward solution $I_\nu(z,u)$ is monochromatic intensity of radiation at frequency ν depending on vertical and angular coordinates: depth into the atmosphere z and cosine of the nadir angle of propagation of radiation u. The forward problem Eq. 9 has two continuous and two discrete model parameters: atmospheric absorption coefficient $\kappa_\nu(z)$, Planck function $B_\nu(z)$, surface emissivity ε_ν, and Planck function $B_{s,\nu}$. The geophysical parameters of interest, the atmospheric and surface temperature, as well as the composition of the atmosphere and surface, are encapsulated in the radiative parameters of the RT problem, and dependence of the radiative parameters on retrieval geophysical parameters is assumed to be known.

The instrument model here is specified by instrument spectral and angular responses. Assuming for simplicity that these responses are linear and decoupled into $\psi(\nu - \nu_0)$ and $\Psi(u - u_0)$, respectively, the radiances measured at the top of the atmosphere ($z = 0$) are represented as

$$R = \int_{\Delta\upsilon} \mathrm{d}\nu\psi(\nu - \nu_0) \int_{-1}^{1} \mathrm{d}u\Psi(u - u_0) \int_{0}^{z_0} \mathrm{d}z\delta(z)I_\nu(z,u)$$

Thus, observables in this forward model can be represented in the form of a linear functional

$$R = \int_{\Delta\nu} \mathrm{d}\nu \int_{-1}^{1} \mathrm{d}u \int_{0}^{z_0} \mathrm{d}z W_\nu(z,u)\, I_\nu(z,u) = (W, I)$$

with the observable weighting function

$$W_v(z,u) = \psi(v - v_0)\Psi(u - u_0)\delta(z)$$

Sensitivities to discrete and continuous model parameters: an overview

The dependence of observables on discrete parameters is essentially that of a function of multiple variables. Thus, the sensitivity of the observable R with respect to a discrete parameter p_k is simply a partial derivative $\partial R/\partial p_k$. If a discrete parameter p_k experiences a variation δp_k, then corresponding variation of the observable has the form

$$\delta_{p_k} R = \frac{\partial R}{\partial p_k}\delta p_k \tag{10}$$

The dependence of observables on continuous parameters is essentially that of a functional dependence on a function. Correspondingly, the sensitivities of observables with respect to continuous parameters are variational (aka functional) derivatives, which can be introduced as follows. Let $F[f]$ be a functional defined in the domain of functions $f(x)$, which, in their turn, are defined in the domain Dx of arguments x. By definition, the variational derivative $\delta F/\delta f(x)$ of the functional $F[f]$ with respect to $f(x)$ is a kernel $K(x)$ in the linear integral relation

$$\delta_f F = \int_{Dx} K(x)\delta f(x)\,\mathrm{d}x \tag{11}$$

and correspondingly

$$\frac{\delta F}{\delta f(x)} \equiv K(x) \tag{12}$$

If the function $f(x)$ is approximated by its values $\{f_j\}$ on a finite-dimensional grid of arguments $\{x_j\}$, then the functional $F[f]$ is approximated by a function of multiple variables $\tilde{F}(f_j)$. Correspondingly, variation $\delta_f \tilde{F}$ can be approximated as

$$\delta_{f_j}\tilde{F} \approx \sum_j \frac{\partial \tilde{F}}{\partial f_j}\delta f_j \tag{13}$$

On the other hand, the above integral expression for the variation $\delta_f F$ Eq. 11 can be approximated as

$$\delta_f F \approx \sum_j K_j \delta f_j \Delta_j x \tag{14}$$

where

$$K_j = \left.\frac{\delta F}{\delta f(x)}\right|_{x=x_j}$$

This yields an interrelation between the variational derivative $\delta F/\delta f(x)$ and its finite-dimensional approximation $\partial\tilde{F}/\partial f_j$:

$$\left.\frac{\delta F}{\delta f(x)}\right|_{x=x_j} \approx \frac{1}{\Delta_j x}\cdot\frac{\partial\tilde{F}}{\partial x_j}, \text{ and } \frac{\partial\tilde{F}}{\partial x_j} \approx \left.\frac{\delta F}{\delta f(x)}\right|_{x=x_j}\cdot\Delta_j x \tag{15}$$

Note that the value of the function $f(x)$ at some fixed value of the argument $x = x'$ can also be considered as a functional defined on this function. To find the variational derivative $\delta f(x')/\delta f(x)$, we observe that

$$f(x') = \int_{Dx} \delta(x' - x) f(x)\,\mathrm{d}x$$

Comparison with Eqs. 11 and 12 yields

$$\frac{\delta f(x')}{\delta f(x)} = \delta(x' - x) \tag{16}$$

This definition Eqs. 11 and 12 represents a practical recipe used to derive expressions for computation of variational derivatives and, thus, of sensitivities of observables to continuous parameters. If the parameter $p(x)$ experiences a variation $\delta p(x)$ and corresponding variation of the observable R can be represented in the form of Eq. 11

$$\delta_p R = \int_{Dx} K_p(x)\delta p(x)\,\mathrm{d}x \tag{17}$$

then sensitivity $\delta R/\delta p(x)$ is the kernel of this linear integral expression

$$\frac{\delta R}{\delta p(x)} \equiv K_p(x) \tag{18}$$

Using known solution Eq. 3, we can directly obtain sensitivities of observable in the above demo forward model Eq. 2 with respect to model parameters $a(t)$ and X_0. Sensitivity to the discrete parameter X_0 is obtained as a partial derivative:

$$\frac{\partial R_1}{\partial X_0} = -\left(A(t_1) + (X_0)^{-1}\right)^{-2}\cdot(X_0)^{-2} = \left(\frac{R}{X_0}\right)^2 \tag{19}$$

Sensitivity to the continuous parameter $a(t)$ is obtained using the definition of the variational derivative Eqs. 11 and 12. We have

$$\delta_a R = -\left(A(t) + (X_0)^{-1}\right)^{-2}\int_{t_0}^{t_1}\delta a(t)\mathrm{d}t = -R^2\int_{t_0}^{t_1}\delta a(t)\mathrm{d}t$$

Then

$$\frac{\delta R}{\delta a(t)} = -R^2 \tag{20}$$

Sensitivity analysis: computation of sensitivities of observables with respect to model parameters

Finite-difference approach

This approach is based on using differences between values of observables $R = M[X]$ computed from the solution X of the *baseline* (non-perturbed) forward problem $N[X] = S$ and values of observables $\tilde{R}_p = M[\tilde{X}_p]$ computed from the solution $\tilde{X}_p$ of the forward problem $\tilde{N}_p[X] = \tilde{S}_p$ with the operator and right-hand term perturbed due to perturbations $\tilde{p} = p + \Delta p$ of the retrieval parameter. Both components of the forward model, $\tilde{N}_p[X] = \tilde{S}_p$, and $\tilde{R}_p = M[\tilde{X}_p]$ are run separately for each discrete retrieval parameter and for each gridpoint of each continuous parameter. The matrix of sensitivities $K = \partial R/\partial p$ is approximated as

$$K = \frac{\partial R}{\partial p} \approx \frac{\tilde{R}_p - R}{\tilde{p} - p}$$

No analytic work is required here, but computer time, as compared to a single run of the forward model, increases in a direct proportion to the number of (gridpoints of) model parameters of interest.

Linearization approach

In this approach, the forward model Eq. 1 has to be linearized assuming that all model parameters of interest experience infinitesimally small perturbations. The resulting linearized forward problem has the form

$$L\delta X = \delta S - \delta N[X] \tag{21}$$

where the linear operator L represents a linearization of the operator N around the baseline solution of the forward problem X.

Note that only the right-hand term of the linearized forward problem Eq. 21 depends on variations of model parameters; the operator of the linearized forward problem L remains the same. This provides substantial savings of computer time.

After the linearized forward problem is formulated, it is solved for sensitivities of the forward solution $X(x)$ to model parameters: discrete p_j and continuous $p(x')$:

$$L\frac{\partial X}{\partial p_j} = \frac{\partial S}{\partial p_j} - \frac{\partial N[X]}{\partial p_j}, \text{ and } L\frac{\delta X}{\delta p(x')} = \frac{\delta S}{\delta p(x')} - \frac{\delta N[X]}{\delta p(x')}$$

Recalling that both N and S depend on continuous parameters as functions, and using the property Eq. 16, the linearized forward problem for sensitivities of the forward solution to continuous model parameters can be rewritten as

$$L\frac{\delta X}{\delta p(x')} = \delta(x' - x)\left(\frac{\partial S}{\partial p(x)} - \frac{\partial N[X]}{\partial p(x)}\right) \tag{22}$$

Further on, the instrument model Eq. 5 is linearized with respect to the forward solution X, resulting into a linear functional, which is written in the form of an inner product

$$\delta R = (W, \delta X) \tag{23}$$

This yields resulting expressions for sensitivities of observables to discrete and continuous parameters:

$$\frac{\partial R}{\partial p_j} = \left(W, \frac{\partial X}{\partial p_j}\right), \quad \frac{\delta R}{\delta p(x')} = \left(W, \frac{\delta X}{\delta p(x')}\right) \tag{24}$$

As an example, consider the linearization of the demo forward problem Eq. 2:

$$\begin{cases} \left(\dfrac{\mathrm{d}}{\mathrm{d}t} + 2a(t)X(t)\right)\delta X(t) = -X^2(t)\delta a(t) \\ \delta X(t_0) = \delta X_0 \end{cases}$$

Here we have

$$\begin{aligned} &L\delta X = \left(\frac{\mathrm{d}}{\mathrm{d}t} + 2a(t)X(t) + \delta(t - t_0)\right)\delta X(t), \\ &\delta S(t) = \delta(t - t_0)\delta X_0, \quad \delta N[X] = X^2(t)\delta a(t) \end{aligned} \tag{25}$$

We obtain sensitivities to parameters X_0 and $a(t)$ using known analytic solutions of this problem.

For X_0, corresponding problem has the form

$$\begin{cases} \left(\dfrac{\mathrm{d}}{\mathrm{d}t} + 2a(t)X(t)\right)_{\delta X_0} = X(t) = 0 \\ \delta_{X_0} X(t_0) = \delta X_0 \end{cases}$$

Comparing its solution

$$\delta_{X_0} X(t) = \delta X_0 \exp\left(-2\int_{t_0}^{t} a(t')X(t')\mathrm{d}t'\right)$$

with definition Eq. 10, we have

$$\frac{\partial X(t)}{\partial X_0} = \exp\left(-2\int_{t_0}^{t} a(t')X(t')\mathrm{d}t'\right)$$

Using $A(t)$, a primitive function of $a(t)$ defined by Eq. 4, and transforming

$$\begin{aligned} \int_{t_0}^{t} a(t')X(t')\mathrm{d}t' &= \int_{t_0}^{t} X(t')\mathrm{d}A(t') \\ &= \ln\left[A(t') + (X_0)^{-1}\right]\Big|_{t_0}^{t} = \ln\left(\frac{X_0 A(t) + 1}{X_0 A(t_0) + 1}\right) \\ &= \ln\left(\frac{X_0}{X(t)}\right) \end{aligned}$$

yield

$$\frac{\partial X(t_1)}{\partial X_0} = \left(\frac{X(t_1)}{X_0}\right)^2$$

Thus

$$\frac{\partial R}{\partial X_0} = \left(\frac{R}{X_0}\right)^2 \tag{26}$$

in accordance with Eq. 19 obtained above using direct linearization.

For $a(t)$, the solution of the corresponding problem

$$\left(\frac{\mathrm{d}}{\mathrm{d}t} + 2a(t)X(t)\right)\delta_a X(t) = -X^2(t)\delta a(t), \delta_a X(t_0) = 0$$

has the form

$$\delta_a X(t) = -\int_{t_0}^{t} \mathrm{d}t'\delta a(t')X^2(t') \exp\left(-2\int_{t_0}^{t} a(t'')X(t'')dt''\right)$$

Using $A(t)$, Eq. 4 and transforming

$$\int_{t'}^{t} a(t'')X(t'')\,\mathrm{d}t = \int_{t'}^{t} X(t'')\,\mathrm{d}A(t'')$$

$$= \ln\left[A(t'') + (X_0{}^{-1})\right]\Big|_{t'}^{t} = \ln\left(\frac{X_0A(t)+1}{X_0A(t')+1}\right) = \ln\left(\frac{X(t')}{X(t)}\right)$$

yields the solution $\delta_a X(t)$ in the form suitable for direct computation of the sensitivity $\delta X(t)/\delta a(t')$:

$$\delta_a X(t) = -X^2(t)\int_{t_0}^{t} \mathrm{d}t'\delta a(t'), \text{ and } \frac{\delta X(t)}{\delta a(t')} = -X^2(t)$$

Then

$$\frac{\delta X(t_1)}{\delta a(t)} = -X^2(t_1)$$

and

$$\frac{\delta R}{\delta a(t)} = -R^2 \tag{27}$$

in accordance with Eq. 20 obtained above using direct linearization.

Adjoint approach

In this approach, the operator L and right-hand term $\delta S - \delta N[X]$ of the linearized forward problem Eq. 21, along with the observables weighting function W of the linearized instrument model Eq. 23, are used in an alternative way involving the operator L^*, which is adjoint to L. This approach is more efficient when the number of parameters to be retrieved (counted by separate gridpoint values for continuous parameters) exceeds the number of observables.

Rewriting for brevity the linearized forward model Eq. 21 as

$$LX' = S' \tag{28}$$

where $S' = \delta S - \delta N[X]$ and rewriting the linearized instrument model Eq. 23 as

$$R' = (W, X') \tag{29}$$

this approach can be outlined as follows. The adjoint operator L^* is derived from the requirement that the identity

$$(g, Lf) = (L^*g, f) \tag{30}$$

be satisfied for an arbitrary pair of functions f and g in the domain of L. Then, the solution X^* of the *adjoint problem*

$$L^*X^* = W \tag{31}$$

provides an alternative way to compute the linearized observables:

$$R' = (X^*, S') \tag{32}$$

This can be demonstrated by multiplying the adjoint problem by X' and the linearized forward problem by X^* and comparing the left and right sides of the resulting equalities:

$$(L^*X^*, X') = (W, X'),\ (X^*, LX') = (X^*, S')$$

Since the left-hand terms (L^*X^*, X') and (X^*, LX') are equal by definition of the adjoint operator, the right-hand terms (W, X') and (X^*, S') are also equal. Since $(W, X') = R'$, then $R' = (X^*, S')$.

Now, returning back from abbreviated Eq. 28 to Eq. 21 and replacing in Eq. 32 $R' \to \delta R$ and $S' \to \delta S - \delta N[X]$, we obtain a direct expression of the variation of the observables through variations of the operator and right-hand term of the forward problem:

$$\delta R = (X^*, \delta S - \delta N[X]) \tag{33}$$

This yields expressions for sensitivities of observables to discrete and continuous parameters in the form

$$\frac{\partial R}{\partial p_j} = \left(X^*, \frac{\partial S}{\partial p_j} - \frac{\partial N}{\partial p_j}X\right), \text{ and } \frac{\delta R}{\delta p(x')} = \left(X^*, \frac{\delta S}{\delta p(x')} - \frac{\delta N}{\delta p(x')}X\right) \tag{34}$$

As an example, consider a matrix forward problem

$$\begin{cases} \dfrac{\mathrm{d}\mathbf{X}}{\mathrm{d}t} + \mathbf{a}(t)\mathbf{X}(t) = \mathbf{b}(t) \\ \mathbf{X}(t_0) = \mathbf{X}_0 \end{cases} \tag{35}$$

where $\mathbf{a}(t)$ is a $n \times n$-matrix and $\mathbf{X}(t)$, $\mathbf{b}(t)$, and $\mathbf{X}_0$ are n-vectors. The forward operator has the form

$$\mathbf{LX} = \left(\frac{\mathrm{d}}{\mathrm{d}t} + \mathbf{a}(\mathrm{t})\right)\mathbf{X}(t) + \delta(t - t_0)\mathbf{X}(t) \tag{36}$$

To derive the operator adjoint to $\mathbf{L}$, we write down the left side of the identity Eq. 30 and transform it to the right side of this identity. For the left side of Eq. 30, we have

$$(\mathbf{X}^*, \mathbf{LX}) = \int_{t_0}^{t_1} \mathbf{X}^{*T}(t)\left(\frac{\mathrm{d}\mathbf{X}}{\mathrm{d}t} + \mathbf{a}(t)\mathbf{X}(t) + \delta(t - t_0)\mathbf{X}(t)\right)\,\mathrm{d}t$$

Performing integration by parts and canceling terms with opposite signs

$$\int_{t_0}^{t_1} \mathbf{X}^{*T}(t)\frac{\mathrm{d}\mathbf{X}}{\mathrm{d}t}\mathrm{d}t + \int_{t_0}^{t_1} \delta(t - t_0)\mathbf{X}(t)\mathbf{X}^{*T}(t)\mathrm{d}t =$$

$$-\int_{t_0}^{t_1} \frac{\mathrm{d}\mathbf{X}^{*T}}{\mathrm{d}t}\mathbf{X}(t)\mathrm{d}t + \int_{t_0}^{t_1} \delta(t_1 - t)\mathbf{X}^{*T}(t)\mathbf{X}(t)\mathrm{d}t$$

we rewrite $(\mathbf{X}^*, \mathbf{LX})$ in the form of the right side of Eq. 30:

$$\int_{t_0}^{t_1} \left(-\frac{\mathrm{d}\mathbf{X}^*}{\mathrm{d}t} + \mathbf{a}^T(t)\mathbf{X}^*(t) + \delta(t_1 - t)\mathbf{X}^*(t)\right)^T \mathbf{X}(t)\,\mathrm{d}t$$
$$= (\mathbf{L}^*\mathbf{X}^*, \mathbf{X})$$

This yields the expression for the operator $\mathbf{L}^*$ adjoint to operator $\mathbf{L}^*$, Eq. 36:

$$\mathbf{L}^*\mathbf{X}^* = \left(-\frac{\mathrm{d}}{\mathrm{d}t} + \mathbf{a}^T(t)\right)\mathbf{X}^*(t) + \delta(t_1 - t)\mathbf{X}^*(t) \tag{37}$$

To formulate the adjoint problem in the form of Eq. 31, we assume observables as an m-vector $\mathbf{R}$. This yields the linearized instrument model in the form

$$\mathbf{R} = (\mathbf{W}, \mathbf{X}) = \int_{t_0}^{t_1} \mathbf{W}^T(t)\mathbf{X}(t)\mathrm{d}t \tag{38}$$

with observables weighting function $\mathbf{W}(t)$ being an $m \times n$-matrix. Assuming the vector $\mathbf{R} = \mathbf{W}_0\mathbf{X}(t_1)$ is to be computed using a constant matrix $\mathbf{W}_0$ applied to the forward solution $X(t)$ at $t = t_1$, we have $\mathbf{W}(t) = \delta(t_1 - t)\mathbf{W}_0$. Corresponding adjoint problem has the form

$$\begin{cases} -\dfrac{\mathrm{d}\mathbf{X}^*}{\mathrm{d}t} + \mathbf{a}^r(t)\mathbf{X}^*(t) = 0 \\ \mathbf{X}^*(t_1) = \mathbf{W}_0^T \end{cases} \tag{39}$$

As an illustration, consider the nonlinear demo forward model Eqs. 2 and 7. Repeating above derivations for the operator of corresponding linearized forward problem, Eq. 25, we obtain

$$L^*X^* = \left(-\frac{\mathrm{d}}{\mathrm{d}t} + 2a(t)X^*(t) + \delta(t - t_1)\right)X^*(t) \tag{40}$$

Thus, corresponding adjoint problem has the form

$$\begin{cases} -\dfrac{\mathrm{d}X^*}{\mathrm{d}t} + 2a(t)X(t)X^*(t) = 0 \\ X^*(t_1) = 1 \end{cases} \tag{41}$$

Its solution (note that integration proceeds backward):

$$X^*(t) = \exp\left(-2\int_t^{t_1} a(t')X(t')\,\mathrm{d}t'\right)$$

To obtain sensitivities to parameters, we will use Eq. 33. For the sensitivity to the discrete parameter X_0, we have

$$\delta_{X_0}R = (X^*, \delta_{X_0}(S)) = \int_{t_0}^{t_1} X^*(t)\,\delta\,(t - t_0)\,\delta X_0\mathrm{d}t$$
$$= X^*(t_0)\delta X_0 = \exp\left(-2\int_{t_0}^{t_1} a(t')X(t')\mathrm{d}t'\right)\delta X_0$$
$$= \left(\frac{R}{X_0}\right)^2 \delta X_0$$

Thus

$$\frac{\partial R}{\partial X_0} = \left(\frac{R}{X_0}\right)^2 \tag{42}$$

in accordance with Eqs. 19 and 26 obtained above using direct linearization and the linearization approach.

For the sensitivity to the continuous parameter $a(t)$, we have

$$\delta_a R = (X^*, -(\delta_a N)[X])$$

Rewriting $X^*(t)$

$$X^*(t) = \exp\left(-2\int_t^{t_1} a(t')X(t')\,\mathrm{d}t'\right)$$
$$= \exp\left[-2\left(\int_{t_0}^{t_1} a(t')X(t')\,\mathrm{d}t' - \int_{t_0}^{t} a(t')X(t')\,\mathrm{d}t'\right)\right]$$
$$= \frac{\exp\left(-2\int_{t_0}^{t_1} a(t')X(t')\,\mathrm{d}t'\right)}{\exp\left(-2\int_{t_0}^{t} a(t')X(t')\,\mathrm{d}t'\right)} = \frac{(X(t_1)/X_0)^2}{(X(t)/X_0)^2} = \frac{R^2}{X^2(t)}$$

we have

$$\delta_a R = (X^*, -(\delta_a N)[X]) = -\int_{t_0}^{t_1} X^*(t)X^2(t)\delta_P a(t)\mathrm{d}t$$

$$= -R^2 \int_{t_0}^{t_1} \delta a(t)\mathrm{d}t$$

Thus

$$\frac{\delta R}{\delta a(t)} = -R^2 \tag{43}$$

in accordance with Eqs. 20 and 27 obtained above using direct linearization and the linearization approach.

Summary

Forward modeling is a necessary component of geophysical retrieval algorithms. It includes the forward problem (the quantitative description of the subject of the geophysical retrieval itself) and the instrument model (the quantitative description of the procedure of measurements). In the most general case, the forward problem is represented by a system of differential equations with initial and/or boundary conditions, which, in general, can be solved only numerically. On the other hand, the instrument model is represented by an analytic expression, which converts the solution of the forward problem into observables, which simulate measured quantities used in geophysical retrievals.

Together, the forward model serves two functions. Besides simulated observables, it provides sensitivities of these observables with respect to the model parameters intended to be retrieved. This can be done in three different ways depending on the preferences of the researcher and on the relation between numbers of observables and model parameters to be retrieved. The finite-difference approach is most straightforward, but also least computer efficient at the same time. The linearization approach is more sophisticated, and it is preferable when the number of parameters to be retrieved is less than the number of observables. The adjoint approach is the most sophisticated and is preferable when parameters to be retrieved outnumber the observables, like in the retrievals of atmospheric profiles in atmospheric remote sensing.

A substantial amount of literature exists on forward modeling and sensitivity analysis (although under various names) in various areas of remote sensing, and the reader is encouraged, once he/she gets a big picture, to do independent search for this literature and to surf the Internet for this purpose. The references below will help in implementation of practical algorithms.

Acknowledgment

This research was carried out at the Jet Propulsion Laboratory, California Institute of Technology, under a contract with the NASA.

Bibliography

Korn, G. A., and Korn, T. M., 2000. *Mathematical Handbook for Scientists and Engineers: Definitions, Theorems, and Formulas for Reference and Review*. New York: Dover Civil and Mechanical Engineering.

Press, W. H., Teukolsky, S. A., Vetterling, W. T., and Flannery, B. P., 2007. *Numerical Recipes*, 3rd edn. New York/Melbourne/Madrid/Cape Town/Singapore/Saõ Paulo: Cambridge University Press. The Art of Scientific Computing.

GEOPHYSICAL RETRIEVAL, INVERSE PROBLEMS IN REMOTE SENSING

Eugene Ustinov
Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA, USA

Definition

Forward model. Quantitative tool for simulation of *observables* for a given set of *model parameters*.
Forward problem. System of differential equations with initial and/or boundary conditions, the solution of which, the *forward solution*, is used for simulation of *observables*.
Forward solution. Solution of the *forward problem*.
Instrument model. Quantitative procedure for evaluation of observables from the *forward solution*, which models the characteristics of the instrument and the procedure for measurements.
Inverse problem. Mathematical problem, usually in the form of a matrix equation, the solution of which, the *inverse solution*, represents an estimate of the *state vector*.
Model parameters. Parameters of the forward problem contained in its differential equations and initial and/or boundary conditions.
Observables. Output parameters of the *forward model*, which simulate the quantities measured in the geophysical retrieval.
Sensitivities. Partial or variational derivatives of *observables* with respect to *model parameters*.
State vector. Subset of the set of model parameters, which is intended to be retrieved.

Introduction

In contrast to *forward modeling*, which proceeds from assumed values of model parameters of the subject of study to simulated observables, the input in inverse modeling consists of measured observables, and the output consists of retrieved model parameters. Also, in contrast to forward modeling, where modeling of observables includes two

major components – *forward problem* (conversion of model parameters into the *forward solution*) and *instrument model* (conversion of the forward solution into simulated observables) – inverse modeling consists of only one major component, *inverse problem*, which converts the measured observables into estimates of the *state vector*, a subset of model parameters subject to the retrieval.

The value of any physical measurement – direct, like in situ measurements, or indirect, like in geophysical retrievals – has to be associated with its uncertainty, measurement error. Thus, it is highly desirable to retrieve the values of model parameters along with their *retrieval errors* (aka error bars). The existing methods of solution of the inverse problem deliver the solution along with its retrieval errors derived from errors of measurements driven by the instrument design.

Formulation and information analysis of the inverse problem

As stated in the accompanying article "*Forward Modeling*," the quantitative relation between model parameters **p**, and simulated observables **R** is, in general, nonlinear. The inverse problem is formulated, essentially, as inversion of this relation. Using high-level notations, we have:

$$\mathbf{F}(\mathbf{p}) = \mathbf{R} \tag{1}$$

It is assumed that all continuous model parameters are represented by their values on suitable grids of arguments and, correspondingly, **F** is a vector function of a vector argument.

The necessary premise for the formulation of the inverse problem is the availability of a computational procedure, which, for a given set of n model parameters described by a vector **p**, provides a capability to compute a set of m observables described by a vector **R** along with their partial derivatives $\partial\mathbf{R}/\partial\mathbf{p}$ treated as a single $m \times n$ matrix of sensitivities **K**. Another premise, which is necessary for the error analysis of the resulting inverse problem, as well as for its solution afterward, is knowledge of measurement errors – uncertainties of measured observables described by an $m-$ vector $\boldsymbol{\varepsilon}$. Computation of the corresponding covariance matrix is a necessary attribute of the algorithm of solution of the inverse problems in geophysical retrievals.

Differences between measured and simulated observables are commonly referred to as *residuals*. The nonlinear inverse problem is solved iteratively, and at each iteration, the solution of the corresponding linearized inverse problem yields corrections to the current approximation of the set of model parameters. These corrections are commonly referred to as *increments*.

The nonlinear inverse problem is solved iteratively. The corresponding linearized inverse problem in high-level notations has the form:

$$\mathbf{K}\mathbf{x} = \mathbf{y} \tag{2}$$

where $\mathbf{y} = \mathbf{R}_{obs} - \mathbf{R}$ is the $m-$ vector of residuals and $\mathbf{x} = \mathbf{p}_{next} - \mathbf{p}$ is the $n-$ vector of increments.

Based on the sensitivity matrix **K**, a preliminary information analysis of the inverse problem can be conducted. In general, not all elements of the measurement set are independent from each other, and, as a result, the number of linearly independent rows of **K** may be less than the total number of the rows. The number $r \le m$ of linearly independent rows is referred to as the *rank* of the matrix **K**.

Rank of matrix **K** can be determined using a technique called singular value decomposition. An eigenvalue problem is considered in the form:

$$\begin{pmatrix} \mathbf{0} & \mathbf{K} \\ \mathbf{K}^T & \mathbf{0} \end{pmatrix} \begin{pmatrix} \mathbf{u} \\ \mathbf{v} \end{pmatrix} = \lambda \begin{pmatrix} \mathbf{u} \\ \mathbf{v} \end{pmatrix} \text{ or, in explicit form :}$$
$$\mathbf{K}\mathbf{v} = \lambda\mathbf{u}$$
$$\mathbf{K}^T\mathbf{u} = \lambda\mathbf{v} \tag{3}$$

Here **u** is an $m-$ vector in the space of measurement vectors and **v** is an $n-$ vector in the space of state vectors. This problem can be rewritten in the form of two eigenvalue problems for matrices $\mathbf{K}^T\mathbf{K}(n \times n)$ and $\mathbf{K}\mathbf{K}^T(m \times m)$:

$$\mathbf{K}^T\mathbf{K}\mathbf{v} = \lambda^2\mathbf{v} \tag{4}$$

$$\mathbf{K}\mathbf{K}^T\mathbf{u} = \lambda^2\mathbf{u} \tag{5}$$

which are solved by standard methods. The numbers of resulting nonzero eigenvalues of the matrices $\mathbf{K}^T\mathbf{K}$ and $\mathbf{K}\mathbf{K}^T$ coincide and yield the rank p of the matrix **K**.

Number of degrees of freedom (*DOF*) *for signal* d_s is another quantitative measure of quality of the measurements selected for the retrieval. It indicates how many independent quantities can be measured. In general, it is not an integer number:

$$d_{\mathrm{s}} = \mathrm{tr}\left(\boldsymbol{\Lambda}^2\left(\boldsymbol{\Lambda}^2 + \mathbf{I}_m\right)^{-1}\right) = \sum_{i=1}^{m} \frac{\lambda_i^2}{\lambda_i^2 + 1} \tag{6}$$

An indicator complementary to d_s is the number of degrees of freedom for noise d_n:

$$d_{\mathrm{n}} = \mathrm{tr}\left(\left(\boldsymbol{\Lambda}^2 + \mathbf{I}_m\right)^{-1}\right) = \sum_{i=1}^{m} \frac{1}{\lambda_i^2 + 1} \tag{7}$$

Shannon information content H of measurements used in retrievals with a given matrix **K** can be estimated assuming Gaussian noise:

$$H = \frac{1}{2}\sum_{i=1}^{m} \ln\left(1 + \lambda_i^2\right) \tag{8}$$

Solution of Inverse Problems

Unlike solution of forward problems, which, in principle, can be accomplished with any numerical accuracy, solution of inverse problems is associated with inherent uncertainty, for two general reasons: presence of measurement errors, and the indirect, inference-like nature of the process of retrieval, which is implemented by solution of the inverse problem.

The presence of measurement errors forces to consider the residuals vector $\mathbf{y}$ as a random quantity. It is customary to replace $\mathbf{y} \to \mathbf{y} + \boldsymbol{\varepsilon}$, where the vector of measurement errors $\boldsymbol{\varepsilon}$ is assumed to obey the Gaussian distribution with an average (mathematical expectation) $\bar{\boldsymbol{\varepsilon}} = 0$ and a covariance matrix $\mathbf{C}_\mathbf{y}$. The corresponding *probability distribution function* (PDF) has the form:

$$P(\boldsymbol{\varepsilon}) \propto \exp\left(-\frac{1}{2}\,\boldsymbol{\varepsilon}^T \mathbf{C}_\mathbf{y}^{-1}\,\boldsymbol{\varepsilon}\right) \tag{9}$$

where the symbol " $\propto$ " means proportionality. In the case of non-correlating measurement errors $s_i (\mathrm{i} = 1, \ldots m)$ Eq. 9 can be rewritten as

$$P(\boldsymbol{\varepsilon}) \propto \exp\left(-\frac{1}{2}\sum_{i=1}^{m}\frac{\varepsilon_i^2}{s_i^2}\right) \tag{10}$$

With measurement errors included explicitly, the inverse problem Eq. 2 takes the form

$$\mathbf{K}\mathbf{x} = \mathbf{y} + \boldsymbol{\varepsilon} \tag{11}$$

where the vector $\mathbf{x}$ is treated as a random quantity, which also obeys the Gaussian distribution with the PDF:

$$P(\mathbf{x}) \propto \exp\left(-\frac{1}{2}\,(\mathbf{x} - \bar{\mathbf{x}})^T \mathbf{C}_\mathbf{x}^{-1}(\mathbf{x} - \bar{\mathbf{x}})\right) \tag{12}$$

The solution of the inverse problem Eq. 2 is sought as the average $\bar{\mathbf{x}}$ of this distribution, whereas the covariance matrix $\mathbf{C}_\mathbf{x}$ describes uncertainty (retrieval errors) of this solution.

By definition of the covariance matrix, $\mathbf{C}_\mathbf{x}$ is symmetric, i.e., $\mathbf{C}_\mathbf{x} = \mathbf{C}_\mathbf{x}^T$, and $\mathbf{C}_\mathbf{x}^{-1} = \left(\mathbf{C}_\mathbf{x}^{-1}\right)^T$. Then, the PDF Eq. 12 can be transformed as

$$P(\mathbf{x}) \propto \exp\left(-\frac{1}{2}\,\mathbf{x}^T\mathbf{B}\mathbf{x} + \mathbf{b}^T\mathbf{x}\right) \tag{13}$$

where

$$\mathbf{B} = \mathbf{C}_\mathbf{x}^{-1} \tag{14}$$

and

$$\mathbf{b} = \mathbf{B}\bar{\mathbf{x}} \tag{15}$$

On the other hand, the PDF $P(\mathrm{x})$ can be derived from the PDF $P(\boldsymbol{\varepsilon})$ Eq. 9 by the substitution $\boldsymbol{\varepsilon} = \mathbf{K}\mathbf{x} - \mathbf{y}$. Essentially, this is an *a posteriori* PDF for the random quantity $\mathbf{x}$ with quantity $\mathbf{y}$ known:

$$P(\mathbf{x}|\mathbf{y}) \propto \exp\left(-\frac{1}{2}\,(\mathbf{K}\mathbf{x} - \mathbf{y})^T\,\mathbf{C}_y^{-1}(\mathbf{K}\mathbf{x} - \mathbf{y})\right) \tag{16}$$

We have:

$$\begin{aligned}&(\mathbf{K}\mathbf{x} - \mathbf{y})^T\mathbf{C}_\mathbf{y}^{-1}\,(\mathbf{K}\mathbf{x} - \mathbf{y})\\ &= (\mathbf{K}\mathbf{x})^T\mathbf{C}_\mathbf{y}^{-1}\mathbf{K}\mathbf{x} - (\mathbf{K}\mathbf{x})^T\mathbf{C}_\mathbf{y}^{-1}\,\mathbf{y} - \mathbf{y}^T\mathbf{C}_\mathbf{y}^{-1}\mathbf{K}\mathbf{x} + \mathbf{y}^T\mathbf{y}\end{aligned} \tag{17}$$

Note that all terms in Eq. 17 are scalars, and thus are equal to themselves transposed. In particular

$$\mathbf{y}^T\mathbf{C}_\mathbf{y}^{-1}\mathbf{K}\mathbf{x} = (\mathbf{K}\mathbf{x})^T\mathbf{C}_\mathbf{y}^{-1}\mathbf{y}$$

Observing also that the term $\mathbf{y}^T\mathbf{y}$ in Eq. 17 does not depend on $\mathbf{x}$ we can rewrite Eq. 16 as:

$$P(\mathbf{x}|\mathbf{y}) \propto \exp\left(-\frac{1}{2}\,\mathbf{x}^T\mathbf{K}^T\mathbf{C}_\mathbf{y}^{-1}\mathbf{K}\mathbf{x} + \mathbf{K}^T\mathbf{C}_\mathbf{y}^{-1}\mathbf{y}\right) \tag{18}$$

Comparing Eq. 18 with the general form of PDF Eq. 13, we obtain:

$$\mathbf{B} = \mathbf{K}^T\mathbf{C}_\mathbf{y}^{-1}\mathbf{K} \tag{19}$$

and

$$\mathbf{b} = \mathbf{K}^T\mathbf{C}_\mathbf{y}^{-1}\mathbf{y} \tag{20}$$

Comparison with the equality Eq. 15 yields a matrix equation for the solution $\bar{\mathbf{x}}$ of the inverse problem Eq. 2:

$$\mathbf{K}^T\mathbf{C}_\mathbf{y}^{-1}\mathbf{K}\,\bar{\mathbf{x}} = \mathbf{K}^T\mathbf{C}_\mathbf{y}^{-1}\mathbf{y} \tag{21}$$

along with the expression for the covariance matrix of this solution:

$$\mathbf{C}_\mathbf{x} = \left(\mathbf{K}^T\mathbf{C}_\mathbf{y}^{-1}\,\mathbf{K}\right)^{-1} \tag{22}$$

In the particular case of non-correlated equal measurement errors, when $s_i \equiv s$, the matrix equation Eq. 21 and expression for the covariance matrix of its solution Eq. 22 reduce to:

$$\mathbf{K}^T\mathbf{K}\bar{\mathbf{x}} = \mathbf{K}^T\mathbf{y} \tag{23}$$

and

$$\mathbf{C}_\mathbf{x} = s^2(\mathbf{K}^T\mathbf{K})^{-1} \tag{24}$$

If the PDF of $\mathbf{x}$ can be sufficiently constrained based on some additional, *a priori* information, then the number of

measurements m can be less than the dimension n of $\mathbf{x}$. This is the case, e.g., in atmospheric remote sensing, when information about correlation between values of the atmospheric parameter to be retrieved can be invoked in the form of a covariance matrix

$$\mathbf{C}_{\mathbf{x},a} = \overline{\mathbf{x}\mathbf{x}^T} \tag{25}$$

Then the a priori PDF of $\mathbf{x}$ can be represented in the form of the corresponding Gaussian distribution

$$P_a(\mathbf{x}) \propto \exp\left(-\frac{1}{2}\,\mathbf{x}^T\mathbf{C}_{\mathbf{x},a}^{-1}\,\mathbf{x}\right) \tag{26}$$

and the resulting constrained PDF is a product of PDFs Eqs. 18 and 26:

$$P_c(\mathbf{x}|\mathbf{y}) = P(\mathbf{x}|\mathbf{y}) \cdot P_a(\mathbf{x})$$

$$\propto \exp\left(-\frac{1}{2}\,\mathbf{x}^T\left(\mathbf{K}^T\mathbf{C}_{\mathbf{y}}^{-1}\mathbf{K} + \mathbf{C}_{\mathbf{x},a}^{-1}\right)\mathbf{x} + \mathbf{K}^T\mathbf{C}_{\mathbf{y}}^{-1}\,\mathbf{y}\right) \tag{27}$$

Correspondingly, the matrix equation for the solution $\bar{\mathbf{x}}$ of the *regularized* inverse problem Eq. 2 and expression for the covariance matrix $\mathbf{C}_{\mathbf{x}}$ of this solution, respectively, take the form:

$$\left(\mathbf{K}^T\mathbf{C}_{\mathbf{y}}^{-1}\,\mathbf{K} + \mathbf{C}_{\mathbf{x},a}^{-1}\right)\bar{\mathbf{x}} = \mathbf{K}^T\mathbf{C}_{\mathbf{y}}^{-1}\mathbf{y} \tag{28}$$

$$\mathbf{C}_{\mathbf{x}} = \left(\mathbf{K}^T\mathbf{C}_{\mathbf{y}}^{-1}\mathbf{K} + \mathbf{C}_{\mathbf{x},a}^{-1}\right)^{-1} \tag{29}$$

In practice, the number of measurements m in atmospheric remote sensing can be larger, and even much larger than n, e.g., in retrievals from data of high-resolution spectral measurements with the Tropospheric Emissions Spectrometer (TES) flown on the Aura spacecraft. But the rank r of the corresponding sensitivity matrix $\mathbf{K}(m \times n)$ is of the order of the number of atmospheric scale heights over the altitude range covered by retrievals. Of course, the number of necessary grid points n is substantially larger, and, thus, invoking of a priori information is necessary.

Recall that in general the inverse problem $\mathbf{F}(\mathbf{p}) = \mathbf{R}$ is nonlinear and it has to be solved using the linearized inverse problem $\mathbf{K}\mathbf{x} = \mathbf{y}$. If the first guess of the state vector $\mathbf{p}_0$ is too far from the solution, then a straightforward application of the above approach at each iteration may not converge to the solution. A more robust approach named Levenberg–Marquardt method, which has a long history of development (see Rodgers (2000) for a review), makes it possible to cope with this difficulty. At each iteration, the size of the step $\bar{\mathbf{x}}$ is regulated by introducing an additional matrix term proportional to a scalar γ, which is chosen based on some semiempirical rules. If no regularization is necessary, this matrix term is $\gamma\mathbf{D}$, where $\mathbf{D}$ is a diagonal scaling matrix with elements corresponding to magnitudes and dimensions of the elements of the state vector $\mathbf{p}$. If the vector $\mathbf{p}$ consists of elements of same magnitudes and dimensions, then $\mathbf{D}$ reduces to the identity matrix $\mathbf{I}$. If regularization is necessary, Rodgers (2000) suggests this term in the form $\gamma\mathbf{C}_{\mathbf{x},a}^{-1}$. Thus, the step $\bar{\mathbf{x}}$ is sought as a solution of the above matrix equations modified accordingly:

$$\left(\mathbf{K}^T\mathbf{C}_{\mathbf{y}}^{-1}\,\mathbf{K} + \gamma\mathbf{D}\right)\bar{\mathbf{x}} = \mathbf{K}^T\mathbf{C}_{\mathbf{y}}^{-1}\mathbf{y} \tag{30}$$

or

$$\left(\mathbf{K}^T\mathbf{C}_{\mathbf{y}}^{-1}\mathbf{K} + (1+\gamma)\mathbf{C}_{\mathbf{x},a}^{-1}\right)\bar{\mathbf{x}} = \mathbf{K}^T\mathbf{C}_{\mathbf{y}}^{-1}\,\mathbf{y} \tag{31}$$

Error analysis of inverse problems

The covariance matrix $\mathbf{C}_{\mathbf{x}}$ provides a straightforward way for estimations of the retrieval errors (error bars) of retrieved profiles defined as variances derived from the covariance matrix

$$\sigma_j = \sqrt{\overline{(x_j - \bar{x}_j)^2}} = \sqrt{(\mathbf{C}_{\mathbf{x}})_{jj}},\ (j = 1, \ldots n) \tag{32}$$

When additional information about the solution in the form of the a priori covariance matrix $\mathbf{C}_{\mathbf{x},a}$ is used, then additional, *smoothing* errors emerge. Whereas error bars represent uncertainties of retrieved values, the smoothing errors represent uncertainties of benchmarking of retrieved values, e.g., retrieved atmospheric profiles, to specific reference arguments, e.g., altitude in the atmosphere. This smoothing error is represented by the $n \times n$ matrix $\mathbf{A} = \partial\bar{\mathbf{x}}/\partial\mathbf{x}$ called the *averaging kernel*. Using the substitution $\mathbf{y} = \mathbf{K}\mathbf{x}$ in the right-hand term of the regularized inverse problem in Eq. 28, and solving it for $\bar{\mathbf{x}}$, we have:

$$\bar{\mathbf{x}} = \left(\mathbf{K}^T\mathbf{C}_{\mathbf{y}}^{-1}\,\mathbf{K} + \mathbf{C}_{\mathbf{x},a}^{-1}\right)^{-1}\mathbf{K}^T\mathbf{C}_{\mathbf{y}}^{-1}\mathbf{K}\mathbf{x}$$

Thus the averaging kernel $\mathbf{A}$ has the form:

$$\mathbf{A} = \left(\mathbf{K}^T\mathbf{C}_{\mathbf{y}}^{-1}\mathbf{K} + \mathbf{C}_{\mathbf{x},a}^{-1}\right)^{-1}\mathbf{K}^T\mathbf{C}_{\mathbf{y}}^{-1}\mathbf{K} \tag{33}$$

It should be emphasized that the accuracy of estimation of the smoothing error represented by the matrix $\mathbf{A}$ depends on the accuracy of the knowledge of a priori covariance matrix $\mathbf{C}_{\mathbf{x},a}$. On the other hand, if the solution does not need to be constrained, then a priori PDF $P(\mathbf{x}) \propto 1$ and correspondingly $\mathbf{C}_{\mathbf{x},a}^{-1} = 0$, and the averaging kernel $\mathbf{A}$ reduces to the $n \times n$ identity matrix $\mathbf{I}_n$.

Finally, retrieval errors may be associated with uncertainties in parameters of the forward model, which are not being retrieved and have to be assumed based on some independent information. General analysis of associated uncertainties is given in the monograph of Clive Rodgers (2000).

Summary

Formulation of the inverse problem, analysis of its information content, choice of the method of its solution, and analysis of resulting retrieval errors represent main phases of development of the inversion algorithm in

geophysical retrievals. The key premise is availability of the adequate forward model, which, for a given set of model parameters, provides accurate values of simulated observables along with the matrix of sensitivities of observables with respect to elements of the state vector. This makes it possible to formulate the linearized inverse problem. If this problem is ill posed, then the algorithm of its solution involves regularization using some a priori information about this solution. The solution is sought as an average over a probability distribution, which is driven by the probability distribution of measurement errors around the measured observables and by the matrix of sensitivities provided by the forward model. The variance of the probability distribution of the solution provides the estimate of retrieval errors. If necessary, this estimate needs to be complemented by estimates of errors due to uncertainties of the model parameters outside the state vector and due to the approximate nature of the forward model.

There is a vast amount of literature describing formulation and solution of inverse problems in geophysical retrievals. The reader is encouraged, once he/she gets a big picture, to do an independent search for this literature and to surf the Internet for this purpose. The monograph by Clive Rodgers (2000) is a rich source of information on formulation and solution of inverse problems in atmospheric remote sensing, which, as a rule, are ill-posed and need regularization. Two other monographs provide valuable information on various aspects of practical implementation of algorithms of solution of inverse problems.

Acknowledgment

This research was carried out at the Jet Propulsion Laboratory, California Institute of Technology, under a contract with the NASA.

Bibliography

Korn, G. A., and Korn, T. M., 2000. *Mathematical Handbook for Scientists and Engineers: Definitions, Theorems, and Formulas for Reference and Review*. New York: Dover Civil and Mechanical Engineering.

Press, W. H., Teukolsky, S. A., Vetterling, W. T., and Flannery, B. P., 2007. *Numerical Recipes*, 3rd edn. New York/Melbourne/Madrid/Cape Town/Singapore/Saõ Paulo: Cambridge University Press/The Art of Scientific Computing.

Rodgers, C. D., 2000. *Inverse Methods for Atmospheric Sounding: Theory and Practice*. Singapore: World Scientific.

GEOPHYSICAL RETRIEVAL, OVERVIEW

Eugene Ustinov
Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA, USA

Definition

Forward model. Quantitative tool for simulation of *observables* for a given set of *model parameters*.

Forward problem. System of differential equations with initial and/or boundary conditions, solution of which, the *forward solution*, is used for simulation of *observables*.

Forward solution. Solution of the *forward problem*.

Instrument model. Quantitative procedure for evaluation of observables from the *forward solution*, which models the characteristics of the instrument and the procedure for measurements.

Inverse problem. Mathematical problem, usually in the form of a matrix equation, solution of which, the *inverse solution*, represents an estimate of the *state vector*.

Model parameters. Parameters of the forward problem contained in its differential equations and initial and/or boundary conditions.

Observables. Output parameters of the forward model, which simulate the quantities measured in the geophysical retrieval.

Sensitivities. Partial or variational derivatives of *observables* with respect to *model parameters*.

State vector. Subset of the set of model parameters, which is intended to be retrieved.

Introduction

Geophysical retrievals are indirect measurements. In contrast to direct measurements, where there is a direct relation between the measured quantity (e.g., length of the mercury column in the thermometer and inferred quantity – temperature), in geophysical retrievals, the quantities to infer (retrieve) are related to measured quantities via an intermediary, a *forward model*, which provides a cause-to-effect relation between quantities to infer (a subset of *model parameters* constituting the *state vector*) and measured quantities (*observables*). For example, in atmospheric remote sensing, the parameters of the atmosphere/surface system are retrieved using measurements of radiances measured at the top of the atmosphere. Thus, in the retrieval process, we proceed back, from the effect, i.e., observed radiances, to the cause – atmospheric and surface parameters – specific values of which result in specific values of the measured radiances, which are used for the retrieval.

To adequately describe the dependence of measured quantities on the model parameters, the forward model has to have two mandatory components. Its first component is the *forward problem* – the description of the subject of retrieval, based on known physics of it, usually, in the form of a system of differential equations with corresponding initial and/or boundary conditions. In the case considered above, the forward problem consists of the equation of radiative transfer (RT) with boundary conditions complemented by known dependencies of radiative parameters directly entering the RT equation and boundary conditions, such as the Planck function, on geophysical parameters of interest, such as atmospheric temperature. In high-level notation, the forward problem can be written in the form of an operator equation:

$$LX = S$$

where the operator L denotes all homogeneous operations on the solution X contained in the differential equations and initial and/or boundary conditions of the forward problem and the right-hand term S combines all nonhomogeneous right-hand terms of the equations and conditions.

The second component of the forward model is the *instrument model*. In the above case, this model consists of angular and spectral distributions of the instrument's response. Convolution of the solution of the forward problem with functions describing the instrument model provides observables, which simulate measured quantities. In high-level notation, the expression for observables can be written in the form of an inner product of the forward solution X with the function W describing the response of the instrument:

$$R = (W, X)$$

Besides observables, the forward model provides another entity – *sensitivities* of observables to the parameters to be retrieved. Essentially, sensitivities are derivatives of observables with respect to the model parameters to be retrieved. For example, in atmospheric remote sensing, temperature weighting functions are sensitivities of observed radiances with respect to temperature. Sensitivities computed along with observables provide premises necessary to solve the *inverse problem*. In contrast to the forward model, the input of the inverse problem is the set of measured observables, and its output is the set of retrieved model parameters providing the best fit of computed observables to measured observables. Solution of the inverse problem implements the aforementioned path from effect to cause. This circumstance is the key reason of complications associated with the solution of the inverse problems.

Forward model: forward problem and sensitivity analysis

Solution of the forward problem and computation of sensitivities constitute two major components of the algorithm implementing the forward model. There are many software packages available, which make implementation of the forward problem in the forward model to be rather a craft than a skill. On the other hand, computation of sensitivities can be implemented in a number of ways, depending on preferences of the algorithm developer.

The simplest but most computer-intensive approach to computation of sensitivities is termed the *finite-difference approach*. The forward model is run once to obtain the baseline solution for the given set of values of model parameters. Then, the parameters to be retrieved are varied one by one, gridpoint by gridpoint. For each such variation, the forward problem is run separately. Resulting variations of observables are divided by variations of parameters yielding the sensitivity matrix – a matrix of partial derivatives of observables with respect to the model parameters to be retrieved.

Two other approaches require some mathematical skills and additional computer programming. For their application, the forward problem and instrument model of the forward model need to be linearized around the baseline solution. The resulting *linearized forward problem* and *linearized instrument model* can be used in two ways, complementing each other, depending on the relation between the number of observables and the number of retrieval parameters.

The *linearization approach* is based on the linearized forward problem formulated for derivatives of the baseline solution with respect to retrieval parameters. The linearized instrument model is applied to obtained derivatives of the baseline solution to calculate partial derivatives of observables with respect to retrieval parameters. The key point is that only right-hand terms of equations (differential equations and initial and/or boundary conditions) in the linearized forward problem vary with the specific retrieval parameter. The left-hand operations on the derivatives of the solution are the same. This provides substantial savings as compared to the finite-difference approach. An illustrative analogy here is the solution of a system of linear algebraic equations for a number of cases where only right-hand inhomogeneous terms vary, but coefficients in the left-hand homogeneous terms remain the same.

The *adjoint approach* is also based on the use of the linearized forward problem and linearized instrument model, but in a different way. It requires more mathematical skills than the linearization approach, as well as additional computer programming. All left-hand operations in the linearized forward problem are combined into a linear operator of this problem. The corresponding adjoint operator is derived based on the definition of this operator. The adjoint problem is formulated using the obtained adjoint operator with the linearized instrument models for given observables as right-hand terms. It turns out that the corresponding *adjoint solution* provides a way to obtain sensitivities with respect to all model parameters at once. Similarly to the linearization approach, all left-hand operations in the adjoint problem are the same, and only right-hand terms vary with the specific observable.

Comparison of performance of the linearization and adjoint approaches is based, essentially, on the comparison of the number of observables with the number of retrievable parameters. It is important to point out here that in the case of continuous parameters, such as atmospheric parameters, the value at each gridpoint should be counted as a separate parameter. With this in mind, if the number of observables exceeds the number of retrievable parameters, then the linearization approach is more efficient. In the opposite case, the adjoint approach is more efficient. For example, in many cases of atmospheric remote sensing, the observables are spectral radiances: a single radiance or some number of them observed at different nadir angles. The retrieval parameters are atmospheric profiles, usually specified on a few tens of gridpoints in the atmosphere. Objectively, the adjoint approach is more efficient

here. But historically, the linearization approach is still dominating, apparently because it is easier to comprehend and implement.

In a few practical situations, the forward problems can be solved analytically. For example, this is the case for radiative transfer of thermal radiation in the non-scattering atmosphere, which may be a good approximation, especially in the microwave spectral region. Then, observed radiances are also available analytically, and corresponding expressions can be linearized to yield sensitivities to desired parameters to be retrieved. The equation of radiative transfer is linear per se, so no linearization is necessary. Application of the adjoint approach in this case results in the adjoint problem, which also can be solved analytically. Of course, resulting analytic expressions for sensitivities are identical to those obtained directly.

Inverse problems and their solution

As was pointed out in the Introduction, the forward models proceed from model parameters (cause) to observables (effect) in two steps. The first step is implemented by the forward problem, which yields a forward solution providing a quantitative description of the subject of retrieval, per se, as is. This description is independent on the instrument model, which implements the second step of the forward model. Application of the instrument model to the forward solution yields simulated observables.

In the inverse process of retrieval, we proceed from observables to the subset of model parameters intended to be retrieved, usually referred to as the *state vector*. Proceeding from the effect to the cause does result in specific complications, usually nonexistent in the process of forward modeling. In general, inverse problems are nonlinear and have to be solved iteratively. At each step, a corresponding linearized inverse problem is solved, which has the form of a matrix equation

$$\mathbf{K}(\mathbf{p}_{k+1} - \mathbf{p}_k) = \mathbf{R}_{obs} - \mathbf{R}_k$$

where the vector of observables $\mathbf{R}_k$ and matrix of sensitivities $\mathbf{K}$ are calculated for the current iteration of the state vector $\mathbf{p}_k$, $\mathbf{R}_{obs}$ is the vector of measured observables, and $\mathbf{p}_{k+1}$ is the next iteration of the state vector. Depending on the physics of the forward model, there are two types of inverse problem: well-posed problems and ill-posed problems.

Well-posed problems do not require any additional information about the state vector, other than that provided by the measured observables themselves. The number of independent observables has to be no less than the number of model parameters in the state vector. Additional observables – dependent or independent – will result in decrease of retrieval errors.

Ill-posed problems cannot be solved in principle without invoking some additional information about the state vector known prior to retrieval, referred to as *a priori information*. This happens, for example, when the state vector includes continuous parameters, which are functions of relevant arguments – usually, space coordinates and/or time. Vertical profiles of atmospheric parameters provide a good example. Another example, although not from the area of geophysics, is the trajectory of a celestial object – its position in the selected system of coordinates as a function of time.

The a priori information can be of various forms. In such case as retrievals of profiles of atmospheric parameters, this information can be provided in the form of covariance matrices of these parameters – derived from statistics of previous measurements, or just guessed from some considerations. In such case as trajectory retrievals, the a priori information can be in the form of requirements on smoothness of the solution in terms of the time derivative of a suitable order.

In most practical cases, the forward models are nonlinear and so are the corresponding inverse problems. They have to be linearized and solved iteratively. At each iteration, the solution sought is the correction of the state vector providing the best fit to the residuals – differences between measured and simulated observables. The linearized inverse problem has a form of a linear algebraic matrix equation, which can be solved using existing numerical recipes.

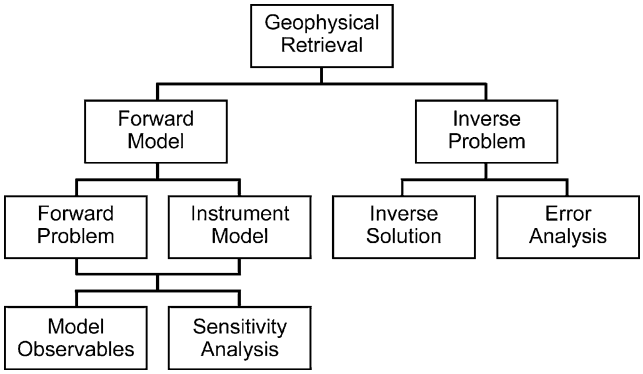
Any physical measurement is meaningful if and only if its result is obtained along with an estimate of its uncertainty. Thus, the solution of inverse problems in geophysical retrievals needs to be complemented by its error analysis. If the forward model accurately describes the subject of retrieval and the procedure of measurement and the resulting inverse problem is well-posed, then the measurement errors are the only source of retrieval errors. Resulting retrieval errors are commonly referred to as *error bars*.

If the inverse problem is ill-posed, then invoking a priori information results in an additional kind of retrieval errors. For example, vertical profiles retrieved in atmospheric remote sensing have finite vertical resolution, i.e., there are uncertainties in benchmarking of retrieved values against the vertical coordinate in the atmosphere. Resulting retrieval errors are referred to as *smoothing errors*.

Finally, any forward model represents only an approximation of the reality, and corresponding *forward model errors* contribute to retrieval errors. If formulation of the forward model itself is accurate, then the only source of forward model errors is due to uncertainties of model parameters outside of the state vector subset. But the forward model may contain uncertainties in its formulation. As Clive Rodgers states in his monograph (Rodgers, 2000), "then modeling error can be tricky to estimate."

Summary

A necessary premise for development of geophysical retrieval algorithms is an availability of an adequate forward model, which, for the given set of model parameters, provides accurate values of simulated observables along with sensitivities of the observables to the state vector.



Global Climate Observing System, Figure 1 Block diagram of principal components of geophysical retrieval.

Another necessary premise is an availability of the adequately formulated inverse problem, solution of which provides the retrieval – a best estimate of the state vector corresponding to measured values of observables – along with estimates of uncertainties of this retrieval due to measurement errors, specifics of the inverse problem, and quality of the forward model. An assembly of the forward model and inverse problem into a robust framework of the algorithm of geophysical retrieval is a challenge, and success in meeting this challenge is a necessary premise of the success of any remote sensing project.

Principal components of geophysical retrieval discussed in this entry are illustrated in Figure 1. (For further details, see *Geophysical Retrieval, Forward Models in Remote Sensing* and *Geophysical Retrieval, Inverse Problems in Remote Sensing*).

There is a vast amount of literature describing geophysical retrieval as a whole, as well as different aspects of it, both in forward modeling and in solution of corresponding inverse problems. A short list below is notably incomplete and reflects personal preferences of the author. The reader is encouraged, once he/she gets a big picture to do independent search for this literature and to surf the Internet for this purpose.

Acknowledgment

This research was carried out at the Jet Propulsion Laboratory, California Institute of Technology, under a contract with the NASA.

Bibliography

Korn, G. A., and Korn, T. M., 2000. *Mathematical Handbook for Scientists and Engineers: Definitions, Theorems, and Formulas for Reference and Review*. New York: Dover Civil and Mechanical Engineering.

Press, W. H., Teukolsky, S. A., Vetterling, W. T., and Flannery, B. P., 2007. Numerical recipes. In *The Art of Scientific Computing*, 3rd edn. New York/Melbourne/Madrid/Cape Town/Singapore/Saõ Paulo: Cambridge University Press.

Rodgers, C. D., 2000. *Inverse Methods for Atmospheric Sounding: Theory and Practice*. Singapore: World Scientific.

GLOBAL CLIMATE OBSERVING SYSTEM

Jean-Louis Fellous
Committee on Space Research (COSPAR) Secretariat, Paris, France

Definition

Albedo. The fraction of solar energy reflected from the Earth back into space.

Fraction of absorbed photosynthetically active radiation (*FAPAR*). The solar radiation reaching the surface on the 0.4–0.7 μm spectral region is known as the photosynthetically active radiation (PAR). FAPAR refers to the fraction of PAR that is absorbed by a vegetation canopy.

Leaf area index. The ratio of total upper leaf surface of vegetation divided by the surface area of the land on which the vegetation grows.

Permafrost. Soil at or below the freezing point of water (0 °C or 32 °F) for 2 or more years.

Phytoplankton. Photosynthetic or plant constituent of plankton, mainly composed of unicellular algae.

Introduction

The Global Climate Observing System (GCOS) was established in 1992. GCOS is an international mechanism aimed at coordinating observing systems and networks for meeting the needs for climate observation at national and global level. GCOS serves as the climate component of the *Global Earth Observation System of Systems (GEOSS)*. It is cosponsored by four international bodies: the World Meteorological Organization (WMO), the Intergovernmental Oceanographic Commission (IOC) of UNESCO, the United Nations Environment Programme (UNEP), and the International Council for Science (ICSU).

GCOS objectives

GCOS's main role is to ensure that the observations required to meeting national and international needs for climate data and information are well identified, gathered, and made widely available. Its ultimate goal is the establishment of a global, reliable, comprehensive, and sustained climate observing system, giving access to the physical, chemical, and biological properties as well as the atmospheric, oceanic, hydrological, terrestrial, and cryospheric processes that contribute the total Earth climate system and capable of detecting and monitoring its natural and man-induced changes.

GCOS activities provide the necessary support to the components of the World Climate Programme, including the World Climate Research Programme, to the assessment role of the Intergovernmental Panel on Climate Change (IPCC), and to the international policy development undertaken through the United Nations Framework Convention on Climate Change (UNFCCC).

GCOS builds on the existing observing systems operated in the context of the WMO World Weather Watch Global Observing System and Global Atmosphere Watch, the *Global Ocean Observing System* (*GOOS*) led by the IOC, and the *Global Terrestrial Observing System* (*GTOS*) led by the Food and Agriculture Organization (FAO) and associated Global Terrestrial Networks. It essentially relies upon observing system elements funded and operated at national level. It includes in situ, airborne, and space-based observational components.

GCOS implementation

A GCOS Steering Committee is charged with the responsibility of providing scientific and technical guidance to the sponsoring and participating organizations for the planning and implementation of the observing system. Scientific advisory panels have also been established to provide expert advice in each domain (atmosphere, ocean, and land surface). These panels are consulted on the appropriate observing strategy and the measurement requirements and contribute to assessing the status of observing networks and systems. A Secretariat located at the WMO Headquarters in Geneva, Switzerland, supports the activities of the Steering Committee, the GCOS panels and the GCOS program as a whole.

The overall plan for the global climate observing system was developed over the 1992–1995 period. GCOS, in consultation with its partners, further prepared and published in 2004 an Implementation Plan (GCOS, 2004) that addressed the requirements identified in the Second Report (GCOS, 2003) on the Adequacy of Global Observing Systems for Climate in Support of the United Nations Framework Convention on Climate Change. This plan specifically responded to the request of the Conference of the Parties to the UNFCCC. It was further complemented in 2006 with a Supplement (GCOS, 2006) specifically dedicated to space-based observations.

Essential climate variables

The GCOS Implementation Plan (GCOS-IP) is based on the needs and requirements relating to the observation of a key list of "essential climate variables" (see Table 1 below) that were identified as technically and economically feasible to observe and having critical impact with respect to the scientific requirements for systematic climate monitoring. The GCOS-IP specified 131 actions in the atmospheric, oceanic, and terrestrial domains to address the observing system needs of the UNFCCC. It outlined a comprehensive program to implement these actions by national, regional, and international entities and focused on needed improvements for observation of 44 essential climate variables (ECVs) for the atmosphere, oceans, and land.

Climate monitoring principles

Global climate change provides a specific challenge for climate monitoring, both through the need for global coverage and because the rate of change of climate variables (such as average temperature, sea level, and rainfall) tends to be small compared with the "background noise" of natural climate variability. Thus, particular attention to the quality and consistency of observations is needed for climate monitoring. In an attempt to help extend the compliance with the general standards and good practices for climate observation, GCOS has thus developed a set of ten "climate monitoring principles" for the collection, archival, and analysis of observations, which were endorsed by the UNFCCC in 1999. GCOS further proposed ten additional satellite monitoring principles that were agreed by the world's space agency members of the Committee on Earth Observation Satellites (CEOS) in 2003. These principles are reproduced in Table 2 below.

Global Climate Observing System, Table 1 Essential climate variables

Domain	Essential climate variables	
Atmospheric (over land, sea, and ice)	Surface	Air temperature, precipitation, air pressure, surface radiation budget, wind speed and direction, water vapor
	Upper air	Earth radiation budget (including solar irradiance), upper-air temperature, wind speed and direction, water vapor, cloud properties
	Composition	Carbon dioxide, methane, ozone, other long-lived greenhouse gases, aerosol properties
Oceanic	Surface	Sea-surface temperature, sea-surface salinity, sea level, sea state, sea ice, current, ocean color (for biological activity), carbon dioxide partial pressure
	Subsurface	Temperature, salinity, current, nutrients, carbon, ocean tracers, *phytoplankton*
Terrestrial	River discharge, water use, ground water, lake levels, snow cover, glaciers and ice caps, permafrost and seasonally frozen ground, albedo, land cover (including vegetation type), fraction of absorbed photosynthetically active radiation, leaf area index, biomass, fire disturbance, soil moisture	

The status of the global climate observing system

In December 2005, the GCOS Secretariat was invited to provide a comprehensive report in June 2009 on progress with actions recommended in the GCOS-IP to maintain, strengthen, or otherwise facilitate global observations of the climate system, including adherence to the GCOS Climate Monitoring Principles. In response to this invitation, a report was submitted to the UNFCCC in April 2009 and

Global Climate Observing System, Table 2 Climate monitoring principles

Effective monitoring systems for climate should adhere to the following principles:

1. The impact of new systems or changes to existing systems should be assessed prior to implementation
2. A suitable period of overlap for new and old observing systems is required
3. The details and history of local conditions, instruments, operating procedures, data processing algorithms, and other factors pertinent to interpreting data (i.e., metadata) should be documented and treated with the same care as the data themselves
4. The quality and homogeneity of data should be regularly assessed as a part of routine operations
5. Consideration of the needs for environmental and climate monitoring products and assessments, such as IPCC assessments, should be integrated into national, regional, and global observing priorities
6. Operation of historically uninterrupted stations and observing systems should be maintained
7. High priority for additional observations should be focused on data-poor regions, poorly observed parameters, regions sensitive to change, and key measurements with inadequate temporal resolution
8. Long-term requirements, including appropriate sampling frequencies, should be specified to network designers, operators, and instrument engineers at the outset of system design and implementation
9. The conversion of research observing systems to long-term operations in a carefully planned manner should be promoted
10. Data management systems that facilitate access, use, and interpretation of data and products should be included as essential elements of climate monitoring systems

    Furthermore, operators of satellite systems for monitoring climate need to:

    (a) Take steps to make radiance calibration, calibration monitoring, and satellite-to-satellite cross-calibration of the full operational constellation a part of the operational satellite system

    (b) Take steps to sample the Earth system in such a way that climate-relevant (diurnal, seasonal, and long-term interannual) changes can be resolved. Thus satellite systems for climate monitoring should adhere to the following specific principles:
11. Constant sampling within the diurnal cycle (minimizing the effects of orbital decay and orbit drift) should be maintained
12. A suitable period of overlap for new and old satellite systems should be ensured for a period adequate to determine inter-satellite biases and maintain the homogeneity and consistency of time-series observations
13. Continuity of satellite measurements (i.e., elimination of gaps in the long-term record) through appropriate launch and orbital strategies should be ensured
14. Rigorous prelaunch instrument characterization and calibration, including radiance confirmation against an international radiance scale provided by a national metrology institute, should be ensured
15. Onboard calibration adequate for climate system observations should be ensured and associated instrument characteristics monitored
16. Operational production of priority climate products should be sustained and peer-reviewed new products should be introduced as appropriate
17. Data systems needed to facilitate user access to climate products, metadata and raw data, including key data for delayed-mode analysis, should be established and maintained
18. Use of functioning baseline instruments that meet the calibration and stability requirements stated above should be maintained for as long as possible, even when these exist on decommissioned satellites
19. Complementary in situ baseline observations for satellite measurements should be maintained through appropriate activities and cooperation
20. Random errors and time-dependent biases in satellite observations and derived products should be identified

was also made available for open review by the community (GCOS, 2009).

The progress assessment for all 131 actions identified in the GCOS-IP showed that good to moderate progress had been achieved for the majority of actions. Nevertheless, in 11 % of the actions, little or no progress could be reported. The report showed evidence that developed countries had improved many of their climate observation capabilities, although commitment to sustained long-term operation remained to be secured for several important observing systems. Developing countries, however, had made only limited progress in filling gaps in their in situ observing networks, with some evidence of decline in some regions. Satellite agencies had improved both mission continuity and observational capability and were increasingly meeting the needs for data reprocessing, product generation, and access. Overall, the global climate observing system had progressed significantly, but still fell short of meeting all the climate information needs of the UNFCCC and broader user communities at regional and national levels.

As a result, a 2010 update of the "Implementation Plan for the Global Observing System for Climate in Support of the UNFCCC" (GCOS, 2010) was prepared in response to a request by Parties to the UNFCCC expressed at the 30th session of the UNFCCC Subsidiary Body for Scientific and Technological Advice (SBSTA) in June 2009 and confirmed in UNFCCC Decision 9/CP.15 (December 2009). This updated GCOS-IP includes 138 actions and takes account of the latest status of observing systems, recent progress in science and technology, the increased focus on adaptation, enhanced efforts to optimize mitigation measures, and the need for improved predictions of climate change.

The supplemental details to the satellite-based component of the 2010 updated GCOS-IP were publicly reviewed, and comments were taken into account after consultation with the broader GCOS expert community.

The so-called Satellite Supplement (GCOS, 2011) provides additional technical detail to the actions and needs identified in the 2010 updated GCOS-IP related to satellite-based observations for climate for each of the essential climate variables. In particular, it details the specific satellite data records that should be sustained in accordance with the GCOS Climate Monitoring Principles, as well as other important supplemental satellite observations that are needed on occasion or at regular intervals.

Summary

Established in 1992, GCOS is a joint undertaking of the World Meteorological Organization (WMO), the Intergovernmental Oceanographic Commission (IOC) of the United Nations Educational Scientific and Cultural Organization (UNESCO), the United Nations Environment Programme (UNEP), and the International Council for Science (ICSU). Its goal is to provide comprehensive information on the total climate system, involving a multidisciplinary range of physical, chemical, and biological properties and atmospheric, oceanic, hydrological, cryospheric, and terrestrial processes. GCOS is intended to be a long-term, user-driven operational system capable of providing the comprehensive observations required for monitoring the climate system, detecting and attributing climate change, assessing impacts of, and supporting adaptation to, climate variability and change, as well as for application to national economic development and for research to improve understanding, modeling, and prediction of the climate system.

The increasing profile of climate change has reinforced worldwide awareness of the importance of an effective global climate observing system. This system has significantly improved over the past years, but much remains to be done to meet the needs of science and society. The consequence of not meeting these requirements would be to seriously compromise the information on, and predictions of, climate variability and change.

Bibliography

GCOS, 2003. *Second Report on the Adequacy of the Global Observing System for Climate in Support of the UNFCCC (GCOS-82), WMO/TD 1143*. Geneva: World Meteorological Organisation.

GCOS, 2004. *Implementation Plan for the Global Observing System for Climate in Support of the UNFCCC (GCOS-92), WMO/TD 1219*. Geneva: World Meteorological Organisation.

GCOS, 2006. *Systematic Observation Requirements for Satellite-Based Products for Climate – Supplemental Details to the Satellite-Based Component of the Implementation Plan for the Global Observing System for Climate in Support of the UNFCCC (GCOS-107)*. Geneva: World Meteorological Organisation.

GCOS, 2009. *Progress Report on the Implementation of the Global Observing System for Climate in Support of the UNFCCC 2004–2008 (GCOS-129)*. Geneva: World Meteorological Organisation.

GCOS, 2010. *Implementation Plan for the Global Observing System for Climate in Support of the UNFCCC (2010 update) (GCOS-138), WMO/TD 1523*. Geneva: World Meteorological Organisation.

GCOS, 2011. *Systematic Observation Requirements for Satellite-Based Data Products for Climate (2011 Update), Supplemental Details to the Satellite-Based Component of the "Implementation Plan for the Global Observing System for Climate in Support of the UNFCCC (2010 Update)" (GCOS-154)*. Geneva: World Meteorological Organisation.

Cross-references

Cryosphere, Measurements and Applications
Global Earth Observation System of Systems (GEOSS)
Global Land Observing System
Ocean, Measurements and Applications

GLOBAL EARTH OBSERVATION SYSTEM OF SYSTEMS (GEOSS)

Steffen Fritz
International Institute for Applied Systems Analysis, Laxenburg, Austria

Definition

Global Earth Observation System of Systems (*GEOSS*) aims to involve all countries of the world to integrate ground-based (in situ), airborne, and space-based observation networks. Those Earth observation (EO) systems which participate in GEOSS retain their existing mandates but share primary observational data as well as information derived from those observations. The sharing of and access to data is enabled through common data standards. GEOSS is designed to address nine societal benefit areas, namely, ecosystems, biodiversity, health, disasters, energy, climate, weather, water, and agriculture. GEOSS seeks to connect the producers of environmental data and decision-support tools with the end users of these products, with the aim of enhancing the relevance of Earth observations to global issues. The end result is a global public infrastructure that generates comprehensive, near-real-time environmental data, information, and analyses for a wide range of users. The Group on Earth Observations (GEO www.earthobservations.org) is coordinating efforts to build GEOSS on the basis of a 10 Year Implementation Plan (GEO, 2005) running from 2005 to 2015. GEO is a voluntary partnership of governments and international organizations. GEO meets in plenary at the senior official level and periodically at the ministerial level. Decisions are taken via consensus of GEO members.

Introduction

We are currently faced with major challenges due to the powerful processes which drive global change. These processes operate at a global scale and can only be observed, understood, and predicted by a system that operates at a supranational level. These processes can have important consequences for human well-being, and the monitoring of these processes is critical in order to understand the

complex system of the Earth's terrestrial and maritime biospheres.

It is envisaged that through GEOSS, the information available to decision makers at all levels will be improved, specifically relating to human health and safety, protection of the global environment, the reduction of losses from natural disasters, and achieving sustainable development. GEOSS is founded on the principle that better international cooperation in the collection, interpretation, and sharing of EO information is an important and cost-effective mechanism for achieving this aim (Fritz et al., 2008). GEOSS will yield a broad range of societal benefits, notably:

1. Reducing loss of life and property from natural and human-induced disasters
2. Understanding environmental factors affecting human health and well-being
3. Improving the management of energy resources
4. Understanding, predicting, mitigating, and adapting to climate variability and change
5. Improving water resource management via better understanding of the water cycle
6. Improving weather information, forecasting, and warning
7. Improving management and protection of terrestrial, coastal, and marine ecosystems
8. Supporting sustainable agriculture and combating desertification
9. Understanding, monitoring, and conserving biodiversity

These correspond to the nine societal benefit areas (SBAs) referred to as disaster, health, energy, climate, water, weather, ecosystems, agriculture, and biodiversity.

Origin of GEOSS

As a result of the Earth Summit in Brazil in 1992, Agenda 21 identified the bridging of the gap between data collection and information required by decision makers as a key priority. As a result of the Summit, three different observing systems were formed, namely, the Global Climate Observing System (GCOS), the Global Ocean Observing System (GOOS), and the Global Terrestrial Observing System (GTOS). The 2002 World Summit on Sustainable Development in Johannesburg highlighted the urgent need for coordinated observations relating to the state of the Earth. It was realized that only by linking and coordinating the current observing systems could complex Earth processes (in an increasingly environmentally stressed world) be understood (GEO, 2005). The First Earth Observation Summit convened in Washington, D.C., in July 2003, adopted a Declaration establishing the ad hoc intergovernmental Group on Earth Observations (ad hoc GEO) to draft a 10 Year Implementation Plan. The Second Earth Observation Summit in Tokyo, Japan, in April 2004 adopted a Framework Document defining the scope and intent of a Global Earth Observation System of Systems (GEOSS). The Third Earth Observation Summit, held in Brussels in February 2005, endorsed the GEOSS 10 Year Implementation Plan and established the intergovernmental Group on Earth Observations (GEO) to carry it out (Plag, 2006).

The Group on Earth Observation (GEO)

GEO is building GEOSS on the basis of a 10 Year Implementation Plan (GEO, 2005). The GEO member governments and participating organizations are supported by the GEO secretariat based in Geneva, Switzerland. The secretariat consists of a director appointed by the executive committee, several international civil servants, and 8–10 national technical and scientific experts who are seconded to the secretariat for 2 or 3 years. The secretariat is responsible for coordinating the tasks and other activities that are driving the 10 Year Implementation Plan for GEOSS. The secretariat also services the plenary and the committees and implements outreach and other support activities.

Earth observation systems

EO systems consist of instruments and models designed to measure, monitor, and predict the physical, chemical, and biological aspects of the Earth system. Buoys floating in the oceans monitor temperature and salinity; meteorological stations and balloons record air quality and rainwater trends; sonar and radar systems estimate fish and bird populations; seismic and Global Positioning System stations record movements in the Earth's crust and interior; some 60-plus high-tech environmental satellites scan the planet from space; powerful computerized models generate simulations and forecasts; and early warning systems issue alerts to vulnerable populations.

These various systems have typically operated in isolation from one another. In recent years, however, sophisticated new technologies for gathering vast quantities of near-real-time and high-resolution EO data have become operational. At the same time, improved forecasting models and decision-support tools are increasingly allowing decision makers and other users of EO to fully exploit this widening stream of information.

With investments in EO now reaching a critical mass, it has become possible to link diverse observing systems together to paint a full picture of the Earth's condition. Because the costs and logistics of expanding EO are daunting for any single nation, linking systems together through international cooperation also offers cost savings.

Implementing GEOSS

As a networked system, GEOSS is owned by all of the GEO members and participating organizations. Partners maintain full control of the components and activities that they contribute to the system of systems. Implementation is being pursued through a work plan (GEO, 2008) currently consisting of over 70 tasks. Each task supports

one of the nine SBAs or four transverse areas and is carried out by interested organizations.

GEO work plans are triennially revised and provide the agreed framework for implementing the GEOSS 10 Year Implementation Plan 2005–2015 (www.earthobservations.org/documents). These work plans consist of a set of practical tasks that are carried out by various GEO members and participating organizations. Connections will be realized between diverse observing, processing, data-assimilation, modeling, and information-dissemination systems. This will make it possible to obtain a considerably increased range of data sets, products, and services on the key aspects of the Earth system. The plans are also focusing on enhancing the role of users and reflect the inputs and engagement of the communities of practice, taking full account of the Integrated Global Observing Strategy (IGOS, www.igospartners.org/) transition into GEO.

Governments and participating organizations have advanced GEOSS by contributing a variety of "Early Achievements." These "First 100 Steps to GEOSS" (http://www.earthobservations.org/documents/the_first_100_steps_to_geoss.pdf) were presented to the 2007 Cape Town Ministerial Summit.

Worldwide, several parallel initiatives are contributing their data to GEOSS. Inter alia, Europe is establishing Global Monitoring for Environment and Security (GMES www.gmes.info), the USA is building the Integrated Earth Observation System (IOES http://usgeo.gov/), and China and Brazil are collaborating through the China-Brazil Earth Resources Satellite Programme (CBERS, www.cbers.inpe.br), which is launching a new Earth observation service that will provide state-of-the-art images of the planet to end users throughout Africa free of charge.

The ultimate objective of GEOSS is to develop the use of EO by a broad range of user communities – from both developed and developing countries and ranging from decision- and policy makers to scientists, industry, international governmental, and nongovernmental organizations. Engagement of these communities to identify their needs for new or improved data is essential to enhance the adequacy of provided services and products for a wide diversity of applications.

Data standards and data dissemination

Due to the fact that EO data is obtained from a multitude of sources, an enormous effort is required among different governments and user groups to achieve true data interoperability (Durbha et al., 2008). Therefore, common standards for architecture and data sharing are essential (see *GEOSS Best Practices* WIKI http://wiki.ieee-earth.org/ and the GEOSS Standards and Interoperability Registry www.earthobservations.org/gci_sr.shtml). Each contributor to GEOSS must subscribe to the GEO data-sharing principles, which aim to ensure the full and open exchange of data, metadata, and products. These issues are fundamental to the successful operation of GEOSS.

The architecture of an Earth observation system refers to the way in which its components are designed so that they function as a whole. Each GEOSS component must be included in the GEOSS registry and configured so that it can communicate with the other participating systems. The GEOSS Components and Services Registry provides a formal listing and description of all the Earth observation systems, data sets, models, and other services and tools that together constitute the Global Earth Observation System of Systems (www.earthobservations.org/gci_cr.shtml).

GEOSS will disseminate information and analyses directly to users. GEO is developing the GEOPortal (www.geoportal.org) as a single Internet gateway to the data produced by GEOSS. The purpose of GEOPortal is to make it easier to integrate diverse data sets, identify relevant data and portals of contributing systems, and access models and other decision-support tools. For users without good access to high-speed Internet, GEO has established GEONETCast (http://www.earthobservations.org/geonetcast.shtml), a system of four communications satellites that transmit data to low-cost receiving stations maintained by the users.

Activities under the SBAs

1. SBA Disasters GEOSS implementation will bring a more timely dissemination of information through better coordinated systems for monitoring, predicting, risk assessment, early warning, mitigating, and responding to hazards at local, national, regional, and global levels (GEO, 2005). Cozannet et al. (2008) describe a prototype catalogue that was developed to improve access to information about sensor networks surveying geological hazards (geohazards). Related Project: for example, GFMC (www.fire.uni-freiburg.de/)
2. SBA Human Health GEOSS will improve the flow of appropriate environmental data and health statistics to the health community, promoting a focus on prevention and contributing to continued improvements in human health worldwide (GEO, 2005). Related Project: for example, PROMOTE (http://www.gse-promote.org/)
3. SBA Energy Resources GEOSS outcomes in the energy area will support environmentally responsible and equitable energy management, better matching of energy supply and demand, reduction of risks to energy infrastructure, more accurate inventories of greenhouse gases and pollutants, and a better understanding of renewable energy potential (GEO, 2005). Related Project: for example, ENVISOLAR (http://www.envisolar.com/)
4. SBA Climate GEOSS outcomes will enhance the capacity to model, mitigate, and adapt to climate change and variability. Better understanding of the climate and its impacts on the Earth system, including its human and economic aspects, will contribute to improved climate prediction and facilitate sustainable development while avoiding dangerous perturbations

to the climate system (GEO, 2005). A global climate observation system is essential to improve our understanding of the climate system and our ability to anticipate trends (Fellous, 2008). Related Project: for example, APEC (http://www.apcc21.net/)

5. SBA Water GEOSS implementation will improve integrated water resource management by bringing together observations, prediction, and decision-support systems and by creating better linkages to climate and other data. In situ networks and the automation of data collection will be consolidated, and the capacity to collect and use hydrological observations will be built where it is lacking (GEO, 2005). Related Project: for example, GRDC (http://grdc.bafg.de/servlet/is/Entry.987.Display/)
6. SBA Weather GEOSS can help fill critical gaps in the observation of, for example, wind and humidity profiles, precipitation, and data collection over ocean areas; extend the use of dynamic sampling methods globally; improve the initialization of forecasts; and increase the capacity in developing countries to deliver essential observations and use forecast products. Access to weather data for the other SBAs will be facilitated (GEO, 2005). Related Project: for example, TIGGE (http://tigge.ecmwf.int/)
7. SBA Ecosystems GEOSS implementation will seek to ensure that methodologies and observations are available on a global basis to detect and predict changes in ecosystem condition and to define resource potentials and limits. Ecosystem observations will be better harmonized and shared, spatial and topical gaps will be filled, and in situ data will be better integrated with space-based observations. Continuity of observations for monitoring wild fisheries, the carbon and nitrogen cycles, canopy properties, ocean color, and temperature will be set in place (GEO, 2005). This SBA is strongly linked to supporting the monitoring of the state of forests and to provide essential information to the UNFCCC process for REDD activities as well as the monitoring of illegal logging. Related Project: for example, POSTEL (http://postel.mediasfrance.org)
8. SBA Agriculture GEOSS implementation will address the continuity of critical data, such as high-resolution observation data from satellites. A truly global mapping and information service, integrating spatially explicit socioeconomic data with agricultural, forest, and aquaculture data will be feasible, with applications in poverty and food monitoring, international planning, and sustainable development (GEO, 2005). Related Project: for example, GAMS (http://www.earthobservations.org/cop_ag_gams.shtml)
9. SBA Biodiversity Implementing GEOSS will unify many disparate biodiversity-observing systems and create a platform to integrate biodiversity data with other types of information. Taxonomic and spatial gaps will be filled, and the pace of information collection and dissemination will be increased (GEO, 2005). Since biodiversity data in general is not lacking, but often uneven in its spatial, temporal, and topological coverage as well as physically dispersed and unorganized, the GEOBON project tries to overcome these shortcomings by installing a system which aims to organize the information, increase the exchange between suppliers and users, and to create a mechanism whereby data of different kinds, from many sources, can be combined (Scholes et al., 2008). Related Project: for example, GEOBON www.earthobservations.org/cop_bi_geobon.shtml

Conclusions

Without a global effort to link all current observing systems to build the Global Earth Observation System of Systems (GEOSS), modern civilization will be struggling to understand the complex chemical, biological, and physical processes of the Earth system. Therefore, GEOSS is needed more than ever to acquire comprehensive, near-real-time environmental data, information, and analysis by users as well as decision makers to respond more effectively to the plethora of environmental challenges.

Since the establishment of the Group on Earth Observations (GEO), many early achievements have been realized. These are documented in www.earthobservations.org under each of the nine GEOSS themes or societal benefit areas (SBAs). However, the active mobilization of data users and providers will remain necessary to make GEOSS a success (Fellous, 2008).

Many EO resources have been created and are available to the global community in order to support scientists, decision makers, and the general populace. To realize a successful GEOSS, the key is to provide mechanisms that enable EO data and geospatial data from those resources to be processed, shared, and coordinated. To this end, the GEOPortal (www.geoportal.org) has been developed as the Internet gateway to the data produced by GEOSS.

Global Earth Observations (EO) may be instrumental to achieve sustainable development, but to date there have been no integrated assessments of their economic, social, and environmental benefits. The project Global Earth Observation – Benefit Estimation (GeoBene www.geo-bene.eu) is developing methodologies and analytical tools to assess the societal benefits of GEOSS. First results from the GeoBene project illustrate that the overall societal benefit is by far higher than the incremental costs necessary to establish GEOSS.

Bibliography

Cozannet, G. L., Hosford, S., Douglas, J., Serrano, J. J., Coraboeuf, D., and Comte, J., 2008. Connecting hazard analysts and risk managers to sensor information. *Sensors*, **8**, 3932–3937.

Durbha, S. S., King, R. L., and Younan, N. H., 2008. An information semantics approach for knowledge management and interoperability for the global earth observation systems of systems. *IEEE Systems Journal*, **2**, 358–365.

Fellous, J. L., 2008. Towards a global climate observing system. *Interdisciplinary Science Reviews*, **3**, 83–94.

Fritz, S., Scholes, R. J., Obersteiner, M., Bouma, J., and Reyers, B., 2008. A conceptual framework for assessing the benefits of a global earth observation system of systems. *IEEE Systems Journal*, **2**, 338–348.
GEO, 2005. *Global Earth Observation System of Systems: 10-Year Implementation Plan Reference Document*. ESA, www.earthobservations.org/documents, 209 pp.
GEO, 2008. *GEO* 2007–2009 *Work Plan Toward Convergence*. GEO, www.earthobservations.org/documents/wp0709_v6.pdf, 30 pp.
http://nng.esoc.esa.de/ws2006/Papers/01_Plag_geo_pospaper_11pt_new.pdf.gz.
Plag, H. P., 2006. *Geo, Geoss and Igos-P: The Framework of Global Earth Observations*. ftp://igscb.jpl.nasa.gov/pub/resource/pubs/06_darmstadt/IGS%20WS%202006%20Papers%20PDF/1_Plag_geo_pospaper_11pt_new.pdf
Scholes, R. J., Mace, J. M., Turner, W., Geller, G. N., Jürgens, N., Larigauderie, A., Muchoney, D., Walther, B. A., and Mooney, H. A., 2008. Towards a global biodiversity observing system. *Science*, **321**, 1044–1045.

Cross-references

Cryosphere and Polar Region Observing System
Global Climate Observing System
Global Land Observing System

GLOBAL LAND OBSERVING SYSTEM

Johannes A. Dolman
Department of Earth Sciences, VU University Amsterdam, Amsterdam, The Netherlands

Definition

An integrated system of in situ and remote sensing observation that provides information about the state of the global land surface.

Introduction

Land has a wide variety of natural features, slopes, vegetation, and soils that affect water budgets, carbon fluxes, and the reflective properties of the surface. Land is often covered by vegetation; importantly, almost 40 % of the Earth's land surface is now under some form of management. Land use changes the characteristics of the land surface and thus can induce important local climate effects, especially through changes in albedo, roughness, soil moisture, and evapotranspiration. Precise quantification of the rate of change is important to determine whether feedback or amplification mechanisms are operating through terrestrial processes to affect the climate system.

Humans have long history of observing state of the land surface through in situ observations. This relates not only to the earliest instrumental observations of surface air temperature by Buys Ballot, but also travelers generated extended descriptions of the state of the land surface they encountered on their travels. The advance of satellite technology has made important breakthroughs in our observing capability, as now for the first time in history we are able to observe at global level the state of the Earth's surface. In 1983, a group of around 100 scientists met to initiate the International Satellite Land Surface Climatology Project (ISLSCP). This project was set out (Seller and Hall, 1992) to (1) monitor global-scale fluctuations of the land surface caused by human interference or climatic fluctuations, (2) further develop mathematical models designed to predict or simulate climate on various time scales, and (3) permit inclusion of land surface climatological variables in diagnostic and empirical studies of climatological variations. In 1984, CEOS (Committee on Earth Observation Satellites) was established in response to a recommendation from a Panel of Experts on Remote Sensing from Space, under the aegis of the G7 Economic Summit. This group, recognizing the multidisciplinary nature of satellite Earth observation, aims to coordinate satellite missions between the various space agencies, particularly with a view to avoid the occurrence of gaps. For the land surface, there is a specific constellation (http://wgiss.ceos.org/lsip/) for Land Surface Imaging established. There are three ways to observe the state of the land surface through remote sensing, in the optical, thermal, and microwave frequencies.

Microwave sensors

Active microwave instruments (radars) transmit at frequencies of around 1–10 GHz and measure the backscattered signals to generate microwave images of Earth's surface at high spatial resolutions (between 1 and 100 m) and with a swath width of around 100 km. The images produced have a similar resolution to those from high-resolution optical imagers, but radars have the capability to "see" through clouds providing data on all weather, on day/night basis. This is particularly useful to assess tropical deforestation. SAR interferometry can record the phase shift between two SAR images recorded at slightly different times, thereby providing highly accurate information on the motion of surfaces. Examples of such systems are the ASAR on ESA's ENVISAT platform, PALSAR (JAXA), and RADARSAT. An important application is also the estimation of surface soil moisture at high resolution (1 km) from active radar systems (Wagner et al., 2008) and the behavior of ice (Rignot and Kanagaratnam, 2006). Passive microwave techniques receive the microwave signal emitted from the Earth's surface. This signal can be used to retrieve soil moisture (Owe et al., 2008), surface temperature (Holmes et al., 2009), vegetation water content (Shi et al., 2008), and snow characteristics (Chang et al., 1987) at relatively large scales. The Soil Moisture and Salinity Mission from ESA (SMOS) has been successfully launched in 2009, while NASA's SMAP (Soil Moisture Active and Passive) is planned for launch in 2015. The use of P-band Synthetic Aperture Radar (SAR) is currently investigated to assess its capability to observe forest biomass from space.

Optical and thermal infrared sensors

Mid-resolution optical satellite systems image the Earth's land surface in visible, near-infrared, shortwave infrared, and thermal infrared wavelengths with spatial resolutions between 10 and 100 m. Examples include the Landsat series, the first launched in 1972, with now Landsat 7 nearing its final operational time. In 2008, the US Geological Survey made the Landsat archives – which extend back to 1972 – freely available. These data are being used to measure rates and patterns of land cover change, first on a continental and pan-tropical scale, and eventually globally. Other satellites include the Japanese ADEOS satellite, the French SPOT, and the Brazilian-Chinese CBERS satellite. High-resolution imagery is provided at 1 m resolution by the IKONOS satellite, while QUICKBIRD can achieve resolutions better than 1 m. Optical sensors are increasingly used to study fundamental properties of the land surface such as fraction of photosynthetically active radiation and Leaf Area Index (LAI). These properties can be used to assess physiological changes in the land surface and as such go beyond the more classical land use classifications (e.g., Gobron and Belward, 2009). The thermal infrared sensors have mainly been used to assess hydrometeorological state variables and fluxes (e.g., Norman et al., 1995) and geological mapping (Kahle and Goetz, 1983).

Future and new developments

Europe's first ice mission, CryoSat-2 carries a precise radar altimeter to measure changes at the margins of the ice sheets that overlay Greenland and Antarctica. By measuring thickness change in ice, CryoSat-2 will provide information on the stability of the Earth's ice sheets.

Among promising new technologies in land surface remote sensing is lidar (Dubayah et al., 2010). Spaceborne lidar has the potential to retrieve many aspects of forest structure important for carbon and ecosystem studies, including canopy height, leaf distribution, and aboveground biomass stocks. The ICESAT (Ice, Cloud, and land Elevation Satellite) mission is currently in orbit, while the planned DESDynI (Deformation, Ecosystem Structure, and Dynamics of Ice) mission will combine a multi-beam lidar with polarimetric and interferometric SAR capability to measure forest structure, biomass, and their dynamics.

Integrated global land observations

Foundations for both in situ observation networks and space-based observing components for the land surface are in place, and continuity of missions and validation still need to be strengthened. Improvements in understanding of the terrestrial components of the climate system, the causes and response of this system to change, and consequences in terms of impact and adaptation are vital to society. Increasing significance is thus being placed on terrestrial data for estimating climate forcing and better understanding of climate change and variability, as well as for impact and mitigation assessment. Recognition of this need makes integration of remote sensing-based and in situ data a priority issue in global change research.

Summary

Land has a wide variety of natural features, slopes, vegetation, and soils that affect water budgets, carbon fluxes, and the reflective properties of the surface. To be able to observe the properties of the land surface, both in situ and remote sensing tools are needed. There are three ways to observe the state of the land surface through remote sensing, in the optical, thermal, and microwave frequencies. We briefly review the most important land applications in these domains.

We conclude that foundations exist for both the in situ observation networks and space-based observing components, but that continuity of missions and validation still need to be strengthened. Improvements in understanding of the terrestrial components of the climate system, the causes and response of this system to change, and consequences in terms of impact and adaptation are vital to society.

Bibliography

Chang, A., Foster, J., and Hall, D., 1987. Nimbus-7 SMMR derived globals now cover parameters. *Annals of Glaciology*, **9**, 39–44.

Dubayah, R. O., Sheldon, S. L., Clark, D. B., Hofton, M. A., Blair, J. B., Hurtt, G. C., and Chazdon, R. L., 2010. Estimation of tropical forest height and biomass dynamics using lidar remote sensing at La Selva, Costa Rica. *Journal of Geophysical Research*, **115**, G00E09, doi:10.1029/2009JG000933.

Gobron, N., and Belward, A., 2009. Global vegetation condition, in state of the climate in 2008. *Bulletin of the American Meteorological Society*, **90**, S44–S45.

Holmes, T. R. H., De Jeu, R. A. M., Owe, M., and Dolman, A. J., 2009. Land surface temperatures from Ka-Band (37 GHz) passive microwave observations. *Journal of Geophysical Research*, **114**, D04113, doi:10.1029/2008JD010257.

Kahle, A., and Goetz, A., 1983. Mineralogic information from a new airborne thermal infrared multispectral scanner. *Science*, **222**, 24–27.

Norman, J., Kustas, W., and Humes, K., 1995. A two source approach for estimating soil and vegetation energy fluxes from observations of directional radiometric surface temperature. *Agricultural and Forest Meteorology*, **77**, 263–293.

Owe, M., De Jeu, R. A. M., and Holmes, T. R. H., 2008. Multisensor historical climatology of satellite-derived global land surface moisture. *Journal of Geophysical Research*, **113**, F01002, doi:10.1029/2007JF000769.

Rignot, E., and Kanagaratnam, P., 2006. Changes in the velocity structure of the Greenland ice sheet. *Science*, **311**(5763), 986–990.

Seller, P. J., and Hall, F. G., 1992. FIFE in 1992: Results, scientific gains, and future research directions. *J. Geophys Res*., **97**, 19091–19109.

Shi, J., Jackson, T., Tao, J., Du, J., Blindish, R., Lu, L., and Chen, K. S., 2008. Microwave vegetation indices for short vegetation covers form satellite passive microwave sensor AMSR-E. *Remote Sensing of Environment*, **112**(2008), 4285–4300.

Wagner, W., Pathe, C., Doubkova, M., Sabel, D., Bartsch, A., Hasenauer, S., Blöschl, G., Scipal, K., Martínez-Fernández, J., and Löw, A., 2008. Temporal stability of soil moisture and radar backscatter observed by the advanced synthetic aperture radar (ASAR). *Sensors*, **8**, 1174–1197.

GLOBAL PROGRAMS, OPERATIONAL SYSTEMS

Mary Kicza
National Oceanic and Atmospheric Administration (NOAA), Washington, DC, USA

Definition

Global programs, operational systems are international coordination mechanisms that have been established to coordinate the world's operational Earth remote sensing activities.

Introduction

Understanding the global environment requires an international effort: The mission objectives of Earth observation systems overlap, remote sensing science is conducted in universities and research centers around the world, and remotely sensed data have interdisciplinary applications. A number of international coordination mechanisms have been established to manage the world's operational Earth remote sensing activities. They address issues ranging from satellite system design and deployment to data acquisition and dissemination. Some of the major mechanisms are summarized below.

Asia-Pacific satellite data exchange and utilization group (APSDEU)

The Asia-Pacific Satellite Data Exchange and Utilization Group was established to increase the amount of data exchanged among agencies in the Asia-Pacific region over existing communications links, to apprise regional agencies of changes and improvements to communications and satellite systems, and to identify means of improving existing data utilization. Participants include the Japanese Meteorological Agency, the China Meteorological Administration, the Korean Meteorological Administration, the Australia Bureau of Meteorology, the Hong Kong Observatory, the Canadian Meteorological Centre, and the United States National Oceanic and Atmospheric Administration.

Committee on Earth observation satellites (CEOS)

CEOS was established in 1984 at the request of the Economic Summit of Industrialised Nations Working Group (G7) as the international forum for Earth observing space agencies. CEOS contributes the space component of the Group on Earth Observation's Global Earth Observation System of Systems and supports key stakeholders with a wide range of Earth observation data, products, and expertise.

CEOS coordinates civil spaceborne observations of the Earth and operates through the best efforts of its 30 members (space agencies) and 22 associates via voluntary contributions. Participating agencies strive to address critical scientific questions and to plan satellite missions without unnecessary overlap. CEOS has three primary objectives in pursuing this goal: (1) to optimize benefits of spaceborne Earth observations through cooperation of its members in mission planning and in development of compatible data products, formats, services, applications, and policies; (2) to serve as a focal point for international coordination of space-related Earth observation activities; and (3) to exchange policy and technical information to encourage complementarity and compatibility of observation and data exchange systems.

CEOS is composed of four main implementation mechanisms: the CEOS Strategic Implementation Team (SIT); a permanent CEOS Secretariat; four Working Groups; and seven CEOS Virtual Constellations for GEO.

Coordination group for meteorological satellites (CGMS)

CGMS was formed in 1972 and currently has 19 member agencies and 7 observer agencies. CGMS provides a forum for the exchange of technical information on geostationary and polar-orbiting meteorological satellite systems, as well as on research and development satellite missions. CGMS works to harmonize meteorological satellite mission parameters, such as orbits, sensors, data formats, and downlink frequencies. CGMS also encourages compatibility and mutual backup for system failures through cooperative mission planning, compatible data products and services, and the coordination of space and data-related activities, thereby complementing the work of other international satellite coordinating mechanisms.

Global climate observing system (GCOS)

The Global Climate Observing System was established in 1992 as a result of the second World Climate Conference in 1990. GCOS is cosponsored by the World Meteorological Organization, the Intergovernmental Oceanographic Commission, the United Nations Environment Programme, and the International Council for Science and provides a framework for meeting worldwide requirements for a sustained and robust global system of climate observations from atmospheric, oceanic, terrestrial, and remote sensing platforms. GCOS constitutes the climate observing component of the Group on Earth Observation's Global Earth Observation System of Systems.

While GCOS neither makes observations nor generates data products itself, it stimulates, encourages, coordinates, and otherwise facilitates the taking of the needed observations by national or international organizations. GCOS provides an operational framework for integrating the observational systems of participating countries and organizations into a comprehensive climate observing system. As such, GCOS serves as the formal observations compliance mechanism for the United Nations Framework Convention on Climate Change and depends on a network of national GCOS focal points, scientists, and a multitude of operational and research agencies to implement its requirements for observing the GCOS Essential Climate Variables (ECV).

Group on Earth observations (GEO)

GEO is coordinating international efforts to build a Global Earth Observation System of Systems (GEOSS). This emerging public infrastructure is interconnecting a diverse and growing array of instruments and systems for monitoring and forecasting changes in the global environment. This "system of systems" supports policymakers, resource managers, science researchers, and many other experts and decision makers.

GEO was launched in response to calls for action by the 2002 World Summit on Sustainable Development and by the G8 (Group of Eight) leading industrialized countries. These high-level meetings recognized that international collaboration is essential for exploiting the growing potential of Earth observations to support decision making in an increasingly complex and environmentally stressed world.

GEO is a voluntary partnership of governments and international organizations. It provides a framework within which these partners can develop new projects and coordinate their strategies and investments. As of March 2012, GEO's Members included 88 governments and the European Commission. In addition, 64 intergovernmental, international, and regional organizations with a mandate in Earth observation or related issues have been recognized as Participating Organizations.

International charter "space and major disasters"

The International Charter "Space and Major Disasters" was initiated at the UNISPACE III conference in Vienna in 1999. The founding members were the European Space Agency, the French Centre National d'Etudes Spatiales, and the Canadian Space Agency.

The Charter represents a joint effort by global space agencies to put Earth observation resources at the service of rescue authorities responding to major disasters. Recognizing that no single satellite operator can address all the challenges associated with disaster management, each member agency commits resources to support the provisions of the Charter and to help mitigate the effects of disasters on human life and property.

North America-Europe data exchange group (NAEDEX)

The North America-Europe Data Exchange Group was established in 1988 to optimize, maintain, and improve the exchange of meteorological data and products between Europe and North America; to maintain and update consolidated European and North American user requirements; and to provide a forum for discussions on planned changes and problems in data dissemination, processing, and formats, as well as optimization of telecommunications and IT infrastructure. Participants include Deutscher Wetterdienst, the European Center for Medium-Range Weather Forecasts, the United Kingdom Met Office, the Meteorological Service of Canada, Météo-France, the United States National Oceanic and Atmospheric Administration, and the European Organisation for the Exploitation of Meteorological Satellites.

World meteorological organization space programme (WMOSP)

The World Meteorological Organization Space Programme coordinates environmental satellite matters and activities throughout all WMO Programmes and provides guidance on the potential of remote sensing techniques in meteorology, hydrology and related disciplines, and applications. Its work includes the optimization of international space-based Earth observing capabilities with the goal of enhancing the space-based component of WMO's World Weather Watch Global Observing System.

Summary

Understanding the global environment requires an international effort. A number of international coordination mechanisms have been established to coordinate the world's operational Earth remote sensing activities that address issues ranging from satellite system design and deployment to data acquisition and dissemination. These include the Asia-Pacific Satellite Data Exchange and Utilization Group, the Committee on Earth Observation Satellites, the Coordination Group for Meteorological Satellites, the Global Climate Observing System, the Group on Earth Observations, the International Charter "Space and Major Disasters," the North America-Europe Data Exchange Group, and the World Meteorological Organization Space Programme.

Bibliography

http://gcos.wmo.int/
http://gosic.org
http://www.ceos.org/
http://www.cgms-info.org/
http://www.disasterscharter.org/
http://www.earthobservations.org/

Cross-references

Data Access
Data Policies
Global Climate Observing System
Global Earth Observation System of Systems (GEOSS)
Global Land Observing System
International Collaboration

GPS, OCCULTATION SYSTEMS

Chi O. Ao
Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA, USA

Synonyms

GNSS (Global Navigation Satellite System) occultation; GNSS Radio Occultation (GNSS RO); GPS Radio Occultation (GPS RO or GPSRO)

Definition

GPS/GNSS Radio Occultation. An atmospheric profiling technique whereby the amplitude and phase delay of the radio signals from the global positioning satellites are observed across the Earth's limb as the satellites rise or set with respect to the receiving platform.

Introduction

A Global Positioning System (GPS) radio occultation (RO) system consists of a GPS receiver on a low-Earth-orbit spacecraft that is capable of tracking the radio signals transmitted by one of the GPS satellites as it rises or sets through the Earth's atmosphere. For each occultation event, the recorded measurements consist of the time series of the carrier phase and the amplitude for each of the two GPS frequencies. With precise knowledge of orbits and clocks for the transmitter and receiver, these measurements can be directly inverted to derive a vertical profile of refractive index n (or refractivity $N = (n - 1) \times 10^6$) from the surface of the Earth to the ionosphere. In the neutral atmosphere, the refractive index can be used to retrieve atmospheric pressure, temperature, and tropospheric humidity profiles with additional constraints. In the ionosphere, electron density profiles can be derived from the refractive index (Figure 1).

The GPS occultation technique has several unique properties compared to other remote sensing instruments. First, it is a coherent, active limb sounding technique that is capable of yielding atmospheric profiles with very high vertical resolution (from 100 m in the lower troposphere to about 1 km up to the middle stratosphere). Second, the GPS RO measurements are "self-calibrating," meaning that they do not need to be adjusted against external "standard" references that could drift over time. This makes the GPS occultation a valuable dataset for climate benchmarking. Third, the GPS signal frequencies are in the L-band and, therefore, essentially unaffected by the presence of clouds and precipitation. This ensures that the GPS occultation measurements will not be degraded or biased in cloudy regions or under severe weather conditions.

Instrumentation, missions, and data coverage

The RO technique was first developed to probe planetary atmospheres starting in the 1960s. The availability of freely accessible radio signals from the GPS constellation made this concept very appealing for Earth remote sensing in terms of cost-effectiveness and scientific merits (Gurvich and Krasil'nikova 1987; Yunck et al., 1988).

The first GPS RO mission was the proof-of-concept GPS/Meteorology (GPS/MET) that operated in 1995–1997 (Ware et al., 1996). It was equipped with the NASA/JPL TurboRogue GPS receiver modified to acquire and track occultation signals. The success of GPS/MET led to the inclusion of the more advanced "BlackJack" GPS RO receivers on the CHAMP (CHAllenging Minisatellite Payload) (Wickert et al., 2001) and SAC-C (Satélite de Aplicaciones Cientificas-C) (Hajj et al., 2004) satellites, which have produced GPS RO data nearly continuously from 2001 to 2008. The twin satellites of Gravity Recovery And Climate Experiment (GRACE) also carried the BlackJack receivers, and routine occultation measurements have been made since 2006. The most dramatic addition to the suite of GPS occultation sensors in space is the six-spacecraft COSMIC (Constellation Observing System for Meteorology, Ionosphere, and Climate, also known as FORMOSAT-3) constellation launched in April 2006 (Anthes et al., 2008), which more than double the existing number of occultation soundings observed each day. The COSMIC spacecraft are equipped with the IGOR (Integrated GPS Occultation Receiver), which is based on the BlackJack design. The BlackJack/IGOR can be configured to track the GPS signal in open loop instead of the traditional phase-locked loop. Open-loop tracking is necessary for tracking the rising occultation and is essential for accurately tracking the highly dynamic signals through the moist lower troposphere (Sokolovskiy et al., 2006; Ao et al., 2009).

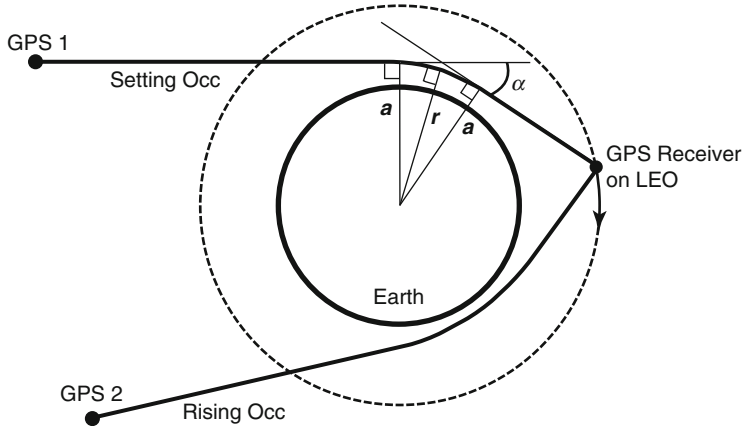
In addition, the ESA/EUMETSAT weather satellites MetOp are equipped with the GRAS GPS occultation receivers. The first of the series, MetOp-A, was launched in October 2006 and have been producing quality occultation data soon afterward (von Engeln et al., 2009). Other new missions currently flying (e.g., TerraSAR-X, TanDEM-X, MetOp-B, OCEANSAT-2) or in planning promise continued availability of GPS RO measurements.

The missions mentioned above are primarily polar orbit satellites in low Earth orbits. Some of the satellites are equipped with both fore- and aft-viewing antennas and could perform occultation measurements in both setting and rising configurations. Approximately 250 occultations, distributed quasi-randomly around the globe, can be obtained per antenna per satellite per day.

Atmospheric retrieval

There are four basic steps in retrieving atmospheric profiles from the GPS RO measurements: (1) calculation of bending angles, (2) removal of ionospheric effects, (3) calculation of refractivity profile, and (4) derivation of the temperature, pressure, and humidity profiles.

The bending angle profile represents a ray optics description of the GPS RO measurements. Traditionally, it is calculated from the Doppler shift of the signal, which is proportional to the rate of change of the phase (Fjeldbo et al., 1971). The "Doppler" method works well under most circumstances; however, it has two shortcomings. First, it is based on geometric optics and hence could not resolve vertical structures smaller than a Fresnel diameter (approximately 1 km in the upper troposphere and lower stratosphere). In addition, the Doppler method assumes that only one ray reaches the receiver at each time. In the moist lower troposphere, atmospheric multipaths occur frequently due to the presence of strong vertical gradients in the refractive index (Sokolovskiy, 2001). To untangle the atmospheric multipaths, diffraction-based



GPS, Occultation Systems, Figure 1 Illustration of GPS RO geometry. A GPS receiver on the low Earth orbit (LEO) measures the changes in the GPS signals as the GPS satellite rises or sets. Due to the vertical variation of the refractivity, the GPS signals are bent and are characterized by the bending angle α and impact parameter a.

methods must be used to transform the received signal from the time domain to impact parameter domain using the Canonical Transform Method (Gorbunov, 2002) or to Doppler frequency domain using the Full Spectrum Inversion (Jensen et al., 2003).

The bending angles are computed for each GPS frequency. Because the GPS signals traverse the ionosphere on its way to the receiver, the bending angles contain contributions from both the ionosphere and the neutral atmosphere. Thus, it is important to remove the ionospheric effects from the bending angles. Due to the dispersive nature of the ionosphere, the first-order (inverse frequency-squared) ionospheric effects can be removed through a linear combination of the dual-frequency bending angles (Vorob'ev and Krasil'nikova, 1994).

The ionosphere-corrected bending angle profile can now be used to compute the refractivity profile directly using the Abel inversion integral (Fjeldbo et al., 1971). Due to the exponentially decreasing atmospheric density with altitude, the bending angles in the upper stratosphere or above are small and susceptible to measurement noise. To reduce the noise in the retrieved refractivity, it is often necessary to replace the bending angles at high altitudes with simple approximations. This can introduce a significant source of uncertainty in stratospheric retrievals (Kursinski et al., 1997; Ao et al., 2006).

Finally, temperature, pressure, and water vapor profiles are derived from the refractivity profile through the use of the microwave refractivity equation (Smith and Weintraub, 1953) and the hydrostatic equation. When water vapor can be neglected, the refractivity is proportional to air density; the temperature and pressure can be determined unambiguously except for the initialization of the hydrostatic equation. At low altitudes, it is not possible to solve simultaneously for these variables without additional constraint. There exist two common approaches. The first is to assume that temperature is known and solve for pressure and water vapor pressure (e.g., Hajj et al., 2002). The second is to use a one-dimensional variational (1-D Var) technique that provides an optimal estimation of the all these variables assuming known error characteristics of the refractivity measurements and the a priori background fields (Healy and Eyre, 2000).

The accuracy and precision of the GPS RO retrievals have been well studied (Kursinski et al., 1997; Kuo et al., 2004). Generally speaking, the GPS RO refractivity and temperature retrievals are most accurate in the altitude range of 8–25 km (upper troposphere and lower stratosphere, or UTLS), with refractivity error of 0.2 % and temperature error less than 1 K. Above 25 km, the ray bending is small so that the retrieval is susceptible to instrument noise and other systematic errors. Below 8 km, the horizontal structure of the atmosphere becomes a significant error source, leading to a refractivity error of 1–3 %.

Comparisons between GPS RO retrievals with collocated radiosonde sounding, global weather analyses, as well as infrared and microwave satellite measurements have shown that the standard deviations in UTLS are about 1 % in refractivity and 1.5 K in temperature. Analysis of collocated CHAMP and SAC-C occultations found them to be consistent to 0.1 K in the mean and 1 K in standard deviation in UTLS (Hajj et al., 2004). Even better agreement can be found among collocated COSMIC occultations, which are shown to have a standard deviation of 0.2 % in refractivity in UTLS (Schreiner et al., 2007).

Applications

GPS occultation measurements have been used to improve weather forecasts and to improve our understanding of the atmosphere and climate.

Recognizing the unique values of GPS occultation, a number of weather centers around the world have been ingesting GPS bending angle and/or refractivity measurements into their operational numerical weather prediction (NWP) models. Past studies have shown that assimilation of GPS occultation data yielded positive impacts on the forecasts (e.g., Healy and Thépaut, 2006; Cucurull et al., 2008). A key advantage of assimilating GPS occultation is that the measurements can be assimilated into the NWP models without a bias correction and thus can serve as important anchor for assimilating other measurements that do require bias correction (Poli et al. 2010).

One of the most promising applications of GPS occultation is its potential for climate benchmarking (Goody et al., 1998). Detecting small long-term climate change signals requires precise measurements that do not vary as a result of changes in instrumentation over time. The self-calibrating nature of the GPS occultation measurements means that the measurements are not subject to biases between different satellites or time-dependent drifts due to orbit changes, problems that have plagued the efforts to infer long-term temperature trends from the Microwave Sounding Unit (MSU) (Schrøder et al., 2003; Steiner et al., 2007).

Another scientific area where GPS occultation has proven useful is in delineating the characteristics of the tropopause, which separates the convectively mixed troposphere (where temperature decreases with height) and the convectively stable stratosphere (where temperature increases with height). The tropopause plays a crucial role in tropical dynamics and the vertical transport of trace gases; moreover, the tropopause height can be a sensitive indicator of climate change. The high vertical resolution that GPS occultation temperature profiles makes them especially suitable for studying the tropopause (Nishida et al., 2000; Randel et al., 2003; Schmidt et al., 2004; Son et al., 2011).

There has also been active research in using GPS occultation data to infer the height of the planetary boundary layer (PBL) (Sokolovskiy et al., 2006; Basha and Ratnam, 2009; Guo et al., 2011; Ao et al., 2012). The PBL height is a key parameter that describes much of the diurnal, synoptic, and climatological processes associated with the PBL in a given region, including its cloud characterization and connections between the surface and free troposphere. It is, however, poorly determined due to a lack of global observations. The fine vertical resolution of the GPS occultation profiles along with its cloud-penetrating ability could potentially be used to provide insights into the boundary layer processes.

Summary

The GPS occultation technique offers unique capabilities in the remote sensing of the Earth's atmosphere. It provides high vertical resolution, all-weather profiling data from the planetary boundary layer to the stratosphere that are valuable for a wide range of weather and climate applications. In particular, the measurements possess qualities that make them particularly useful for monitoring long-term climate change. The future of GPS occultation looks exciting, with a proliferation of GNSS transmitters from the European Galileo, Russian GLONASS, and Chinese Compass navigation satellites. Increased spatial-temporal coverage would especially benefit studies of tropical storms and other mesoscale phenomena.

Acknowledgment

This research was carried out at the Jet Propulsion Laboratory, California Institute of Technology, under a contract with the NASA.

Bibliography

Anthes, R. A., et al., 2008. The COSMIC/FORMOSAT-3 mission: early results. *Bulletin of the American Meteorological Society*, **89**, 313–333.

Ao, C. O., Hajj, G. A., Iijima, B. A., Mannucci, A. J., Schrøder, T. M., de la Torre Juárez, M., and Leroy, S. S., 2006. Sensitivity of stratospheric retrievals from radio occultations on upper boundary conditions. In Foelsche, U., Kirchengast, G., and Steiner, A. (eds.), *Atmosphere and Climate: Studies by Occultation Methods*. Berlin: Springer, pp. 17–26.

Ao, C. O., Hajj, G. A., Meehan, T. K., Dong, D., Iijima, B. A., and Mannucci, A. J., 2009. Rising and setting GPS occultations by use of open-loop tracking. *Journal of Geophysical Research*, **114**, D04101, doi:10.1029/2008JD010483.

Ao, C. O., Waliser, D. E., Chan, S. K., Li, J.-L., Tian, B., Xie, F., and Mannucci, A. J., 2009. Planetary boundary layer heights from GPS radio occultation refractivity and humidity profiles. *Journal of Geophysical Research*, **117**, D16117, doi:10.1029/2012JD017598, 2012.

Basha, G., and Ratnam, M. V., 2009. Identification of atmospheric boundary layer height over a tropical station using high resolution radiosonde refractivity profiles: comparison with GPS radio occultation measurements. *Journal of Geophysical Research*, **114**, D16101, doi:10.1029/2008JD011692.

Cucurull, L., Derber, J. C., Treadon, R., and Purser, R. J., 2008. Preliminary impact studies using Global Positioning System radio occultation profiles at NCEP. *Monthly Weather Review*, **136**, 1865–1877.

Fjeldbo, G., Kliore, A. J., and Eshleman, V. R., 1971. The neutral atmosphere of Venus as studied with the Mariner V radio occultation experiments. *Astronomy Journal*, **76**, 123–140.

Goody, R., Anderson, J., and North, G., 1998. Testing climate models: an approach. *Bulletin of the American Meteorological Society*, **79**(11), 2541–2549.

Gorbunov, M. E., 2002. Canonical transform method for processing radio occultation data in the lower troposphere. *Radio Science*, **37**, 5, doi:10.1029/2000RS002592.

Guo, P., Kuo, Y.-H., Sokolovskiy, S. V., and Lenschow, D. H., 2011. Estimating atmospheric boundary layer depth using COSMIC radio occultation data. *Journal of Atmospheric Science*, **68**, 1703–1713.

Gurvich, A. S., and Krasil'nikova, T. G., 1987. Navigation satellites for radio sensing of the earth's atmosphere (in Russian). *Soviet Journal of Remote Sensing*, **6**, 89–93.

Hajj, G. A., Kursinski, E. R., Romans, L. J., Bertiger, W. I., and Leroy, S. S., 2002. A technical description of atmospheric sounding by GPS occultation. *Journal of Atmospheric and Solar-Terrestrial Physics*, **64**(4), 451–469.

Hajj, G. A., Ao, C. O., Iijima, B. A., Kuang, D., Kursinski, E. R., Mannucci, A. J., Meehan, T. K., Romans, L. J., de la Torre Juárez, M., and Yunck, T. P., 2004. CHAMP and SAC-C atmospheric occultation results and intercomparisons. *Journal of Geophysical Research*, **109**, D6, doi:10.1029/2003JD003909.

Healy, S. B., and Eyre, J. R., 2000. Retrieving temperature, water vapour and surface pressure information from refractive-index profiles derived by radio occultation: a simulation study. *Quarterly Journal of the Royal Meteorological Society*, **126**(566), 1661–1683.

Healy, S. B., and Thépaut, J. N., 2006. Forecast impact experiment with GPS radio occultation measurements. *Quarterly Journal of the Royal Meteorological Society*, **132**, 605–623.

Jensen, A. S., Lohmann, M. S., Benzon, H.-H., and Nielsen, A. S., 2003. Full spectrum inversion of radio occultation signals. *Radio Science*, **38**, 3, doi:10.1029/2002RS002763.

Kuo, Y.-H., Wee, T., Sokolovskiy, S., Rocken, C., Schreiner, W., Hunt, D., and Anthes, R., 2004. Inversion and error estimation of GPS radio occultation data. *Journal of the Meteorological Society of Japan*, **82**(1B), 507–531.

Kursinski, E. R., Hajj, G. A., Schofield, J. T., Linfield, R. P., and Hardy, K. R., 1997. Observing earth's atmosphere with radio occultation measurements using the Global Positioning System. *Journal of Geophysical Research*, **102**(D19), 23429–23465.

Nishida, M., Shimizu, A., Tsuda, T., Rocken, C., and Ware, R. H., 2000. Seasonal and longitudinal variations in the tropical tropopause observed with the GPS occultation technique (GPS/MET). *Journal of the Meteorological Society Japan*, **78**, 691–700.

Poli, P., Healy, S. B., and Dee, D. P., 2010. Assimilation of Global Positioning System radio occultation data in the ECMWF ERA–Interim reanalysis. *Quarterly Journal of the Royal Meteorological Society*, **136**, 1972–1990, doi:10.1002/qj.722.

Randel, W. J., Wu, F., and Rios, W. R., 2003. Thermal variability of the tropical tropopause region derived from GPS/MET observations. *Journal of Geophysical Research*, **108**(D1), 4024.

Schmidt, T., Wickert, J., Beyerle, G., and Reigber, C., 2004. Tropical tropopause parameters derived from GPS radio occultation measurements with CHAMP. *Journal of Geophysical Research*, **109**, D13105.

Schreiner, W., Rocken, C., Sokolovskiy, S., Syndergaard, S., and Hunt, D., 2007. Estimates of the precision of GPS radio occultations from COSMIC/FORMOSAT-3. *Geophysical Research Letters*, **34**, 4, doi:10.1029/2006GL027557.

Schrøder, T., Leroy, S., Stendel, M., and Kaas, E., 2003. Validating the microwave sounding unit stratospheric record using GPS occultation. *Geophysical Research Letters*, **30**(14), 1734, doi:10.1029/2003GL017588.

Smith, E. K., and Weintraub, S., 1953. The constants in the equation for atmospheric refractive index at radio frequencies. *Proceedings of the IRE*, **41**, 1035–1037.

Sokolovskiy, S., 2001. Modeling and inverting radio occultation signals in the moist troposphere. *Radio Science*, **36**, 3, doi:10.1029/1999RS002273.

Sokolovskiy, S., Kuo, Y.-H., Rocken, C., Schreiner, W. S., Hunt, D., and Anthes, R. A., 2006. Monitoring the atmospheric boundary layer by GPS radio occultation signals recorded in open-loop mode. *Geophysical Research Letters*, **33**, L12813, doi:10.1029/2006GL025955.

Son, S.-W., Tandon, N. F., and Polvani, N. F., 2011. The fine-scale structure of the global tropopause derived from COSMIC GPS radio occultation measurements, *Journal of Geophysical Research*, **116**, D20, doi:10.1029/2011JD016030.

Steiner, A. K., Kirchengast, G., Borsche, M., Foelsche, U., and Schoengassner, T., 2007. A multi-year comparison of lower stratospheric temperatures from CHAMP radio occultation data with MSU/AMSU records. *Journal of Geophysical Research*, **112**, D22110, doi:10.1029/2006JD008283.

von Engeln, A., Healy, S., Marquardt, C., Andres, Y., and Sancho, F., 2009. Validation of operational GRAS radio occultation data. *Geophysical Research Letters*, **36**, 17, doi:10.1029/2009GL039968.

Vorob'ev, V. V., and Krasil'nikova, T. G., 1994. Estimation of the accuracy of the atmospheric refractive index recovery from Doppler shift measurements at frequencies used in the NAVSTAR system. *Physics of the Atmosphere and Ocean*, **29**, 602–609.

Ware, R., et al., 1996. GPS sounding of the atmosphere from low earth orbit – preliminary results. *Bulletin of the American Meteorological Society*, **77**, 19–40.

Wickert, J., Reigber, C., Beyerle, G., König, R., Marquardt, C., Schmidt, T., Grunwaldt, L., Galas, R., Meehan, T. K., Melbourne, W. G., and Hocke, K., 2001. Atmospheric sounding by GPS radio occultation: first results from CHAMP. *Geophysical Research Letters*, **28**(17), 3263–3266.

Yunck, T. P., Lindal, G. F., and Liu, C. H., 1988. The role of GPS in precise earth observation. In *IEEE Position, Location and Navigation Symposium*, November 29–December 2, Orlando.