How caves form

Some chemistry

Most of the longest and deepest caves in the world are formed by chemical corrosion processes in rocks particularly water-soluble thanks to their mineral composition. These corrosion processes are known as karst processes.

Water and rock

All minerals are more or less soluble in water, but some are much more soluble than others, and require a very short time span to dissolve (in a geological sense, naturally), while others require much longer periods of time, and are therefore considered practically insoluble. Rocks made of the most soluble minerals are the ones in which karst features develop more easily, even though the karst process is a complex one in which rock composition is only one of the many factors involved in the development of the phenomenon.

When studying the solubility of the main minerals forming the most diffused rocks of the Earth's surface, it can be observed how the solubility of the different minerals varies greatly. For this reason, rocks such as rock salt (made of sodium chloride NaCl, or halite), one of the most water-soluble minerals, are practically absent in humid climates as they dissolve rapidly. In quartzite rocks, made of quartz (SiO_2), one of the minerals most resistant to weathering, karst features may develop only in particular climatic conditions and in areas in which waters had a very long time, i.e. millions of years, to dissolve the rocks (e.g. the Amazon Tepuy quartzite cave systems).

Not only water

In nature, however, things are not that simple. In fact, natural waters are never pure waters, but they are water solutions containing various ions dissolved within them that can increase the aggressiveness and corrosive action on some types of rocks, thus complicating the simple dissolution reaction. The process is well known to those who have to clean the bathroom at home. In order to remove the calcareous incrustations that ruin all the bathroom fittings (geologically speaking, these are calcium carbonate crystals, $CaCO_3$, calcite; rocks that prevalently contain calcite are called limestones), we use aqueous solutions enriched with acids that increase their corrosive streght, such as hydrochloric acid (also known to housewives as muriatic acid), or acetic acid, that are present in many house-cleaning products. These substances make the removal of incrustations easier in two ways: on one hand they increase the solubility of calcite, on the other they greatly accelerate the speed of reaction (that is very rapid and violent, in fact bubbles form because of gases freed when these products are used). Even pure water could obtain the same results, but the time taken would be decidedly beyond the scale of human observation…. and by far too long for the housewives' patience! In fact natural water behaves in the same way as the detergents, but since the acid solution is much more diluted and with much weaker acids, chemical reactions are much slower, at least on a human-observation scale. Nature, unlike housewives, has no hurry, and the results are even more spectacular!

Carbon dioxide: allied with karst

Among the substances that can increase the corrosive power of natural water, carbon dioxide (CO_2) has the most important role. It is already present in meteoric waters, as it is one of the gases of the atmosphere, however its percentages are very low (0.03 atm). Its concentration increases greatly in waters that cross through thick layers of soil covered with dense vegetation. Enrichment with CO_2 and other organic acids produced by vegetation and biological activity of water coming into contact with rock can lead to a remarkable increase in the solubility of minerals like calcite (calcium carbonate, $CaCO_3$) and dolomite (calcium carbonate and magnesium, $CaMg(CO_3)_2$), which increases from 10-12 mg/l to 200-400 mg/l. Other minerals, such as quartz or rock salt are on the contrary insensitive to the presence of CO_2 in the water they come into contact with, and their solubility therefore remains unchanged.

The dissolution reaction of calcium carbonate responsible for the formation of most caves is:

$CaCO_3 + CO_2 + H_2O = Ca_2+ + 2HCO_{3-}$

In the case of dolomite, the reaction is very similar :

$CaMg(CO_3)_2 + 2CO_2 + 2H_2O = Ca_2+ + Mg_2+ + 4HCO_{3-}$

Rocks that are most suited

Calcite and dolomite (calcium carbonate and calcium carbonate and magnesium, respectively) are minerals that are very abundant on the Earth's surface and are the main constituents of particular sedimentary rocks, known as carbonatic rocks, such as limestone and dolomite rocks. These rocks are not the most prone to karstification (gypsum and rock salt are much more soluble), however they are the most common rocks in which karstification processes take place, the most widespread and the ones able to support the existence of large volume rooms and galleries without collapsing, unlike other "weaker" rocks. For this reason the longest and deepest caves on the planet are to be found in this type of rocks. (In dolomite rocks, that are actually as soluble as calcite, the dissolving reaction is much slower and as a residue of the reaction, the rock tends to produce a sand that fills the fissures where water circulates, and this finally slows karstification down. Therefore karstification in dolomite rocks is a much less important process than in limestone rocks).

Underground water

Water commonly found underground creating caves is mainly of meteoric origin, however other kinds of water may be mixed to it in various ways. The following can be found: "Connate" water, i.e. ancient water that was trapped in a sedimentary rock during its formation, and generally very rich in salts, and therefore potentially very aggressive; deep so-called "juvenile" water, produced by magma activity, often very hot and aggressive; or meteoric water that reaches the deeper layers where it is heated and enriched with salts and acids and then comes out at the surface again through faults, generally with the characteristics of hydrothermal water. These are almost always very aggressive waters and generally their temperature is high. When these deep waters come into contact with the rock, they give rise to very rapid and intense dissolution processes, known as hyperkarst processes creating particular caves, called hypogenic caves (i.e. generated from the deep) as for example the Grotta Giusti cave near Pistoia.

Hot waters, cold waters

The possibility of water to corrode the rocks it comes into contact with depends on the chemical composition of both water and minerals building rocks. However other factors intervene to make reactions complex. In particular, temperature is a fundamental factor that acts in two ways that are apparently in contrast one with the other. In fact, the quantity of CO_2 that can dissolve in water depends on temperature – (at the same pressure) the lower the temperature, the greater the amount of CO_2 that can dissolve (also pressure is important, but as it can be affected by relatively small variations at normal environmental conditions, it is neglected for the sake of simplicity). How temperature influences the quantity of CO_2 dissolved in water can easily be seen when a bottle of carbonated mineral water is opened : the action may happen quite peacefully if the bottle has just been taken from the refrigerator, however it can have explosive results if the operation is carried out on the same bottle that has been left inside a car parked in the sun. In fact hot water cannot contain the same amount of CO_2 as cold water, and the excess amount is freed in the form of bubbles as soon as it is given the chance (i.e. when we remove the cap, decreasing pressure). This theoretically tends to favour karst processes in mountain areas at high altitudes, where water temperatures are low and therefore can contain a greater quantity of CO_2 . However two other contrary factors must be considered : first the speed of reaction is higher when temperature is higher, and second at high temperatures vegetation, and as a consequence the production of CO_2 in the soil are much greater, thus, in practice, making warm waters much more aggressive than cold ones (furthermore the content of CO_2 in the atmosphere decreases with altitude, therefore mountain waters, although they are colder, have a lower content of CO_2).
Once again boring domestic chores offer an occasion for an experiment : if we heat the products used for cleaning, the reaction, and therefore the action of removing the calcareous incrustations, is even more rapid (this operation is not really advisable because detergents also contain other substances that can release toxic and irritating vapours when heated – therefore it's best to carry out this experiment with a glass of hot vinegar, as our wise grandmothers used to do!). It is of fundamental importance, when studying karstification processes, to know the diagram that shows the

solubility of calcite at different temperatures and with different quantities of dissolved CO_2.
As a consequence of this, therefore, areas that are most favourable for karstification are those in which, besides an abundance of carbonatic rocks there is abundance of water with a large amount of CO_2 and at a high temperature. These conditions are present in the intertropical zone. Most cavities in Italy, in fact, are the relicts of caves that formed when climate was humid tropical even at our latitudes.

Complicated reactions

Another factor able increase the dissolution of calcite and dolomite in water solutions is the presence of particular ions, such as chlorine and magnesium, sulphurs or sulphates. For this reason, sea water and even more, sea water mixed with fresh water form a very aggressive mixture. Aggressive mixtures also originate when waters with different chemical properties mix. Corrosion due to water mixing, or Boegli's effect, named after the speleologist who discovered this phenomenon, is responsible for the formation of underground galleries and conduits. In fact aggressiveness would otherwise tend to wear out as water corrodes rocks thus increasing its calcium carbonate content, but can be "renewed" by these waters. A proof of this effect is witnessed by those privileged cave-divers when they penetrate karst conduits that have been widened by sea water and observe the gradual widening of the conduits near the area in which sea water and fresh water mix. Or when, in the presence of fresh water, a sudden widening can be observed, forming particular forms of corrosion, such as dissolution pockets, cupolas, and bell holes, in galleries joining and anastomosing into each other, where waters with a different chemical content flow and mix. In order to activate this mechanism, large quantities of water are not necessary. Forms of water mixing corrosion can be seen on the walls or on the ceilings of galleries, close to very small conduits, at times simple fractures (cupolas and bell holes formed by corrosion due to mixing waters are recognizable from similar forms, created by mechanical erosion, as they become narrower in the deeper parts and they always form near conduits or fractures, even of very small diameters).

Aggressive waters, saturated waters

Water that comes into contact with rocks initially is undersaturated, i.e. it can dissolve the minerals of the rock and progressively be enriched with the ions freed in the dissolving reaction, till it reaches a point of saturation, i.e. the water solution contains the maximum quantity possible of a particular ion in certain temperature conditions, atmospheric pressure or content of other acids. When this condition of saturation is reached, water no longer has a chemical effect on rock, and therefore it can only affect rock with mechanical erosion processes (as in the case of surface free flowing water).
If variations in temperature, in CO_2 content or in concentration of the solution (e.g. in the case of evaporation) occur, saturated water becomes oversaturated, i.e. it contains an excessive quantity of dissolved calcium carbonate that should therefore be deposited in the form of calcite crystals, forming speleothemes, among which stalactites and stalagmites are the better known forms; however speleothemes may exhibit a vast range of forms and colours, at times bizarre and curious (when drops of water that carelessly drip on the sink evaporate, these form the detested calcareous deposits, which are a form of concretion).

A bit of geology

The fundamental ingredients to produce karst processes are abundant water rich in CO_2 and organic acids, and a favourable type of rock, but in order to have long and deep cave systems, these conditions are not sufficient.

Pores and fissures

Carbonatic rocks, which are most favourable for karst process, are generally very compact rocks. The granules forming them are very densely packed, the mechanisms by which they were formed led to a primary porosity (i.e. a percentage of voids) and permeability (i.e. a percentage of interconnecting voids that allow the passage of water) of these rocks that are very low. They are practically almost impermeable (the primary porosity of a limestone generally varies from 1 to 20%; only reef limestones have a higher primary porosity). In these conditions, karst processes can act only on the surface of the rocks, creating surface karst forms, such as *karren* (furrows and small depressions created by the dissolution of the rock): water has no possibility of filtering underground which is an indispensable condition for the

formation of caves deep underground.
In order to be affected by karst processes deep underground, a rock must be characterized by discontinuities, through which water can filter and begin to percolate underground. Discontinuities that are very useful for this purpose are bedding planes, which often characterize limestones (dolomites instead are often massive and do not have any stratifications). Bedding planes form during sediment deposition, however most discontinuities in carbonatic rocks are secondary, of a tectonic origin. In fact carbonatic rocks are very fragile and get fractured easily if they are subjected to mechanical stress. Bedding planes are initially horizontal, but subsequent tectonic deformations can incline and fold them in various ways.

Chemical deposits

All caves are filled with more or less large amounts of chemical deposits, minerals and physical deposits, sediments of various types which are mostly transported by water into caves. These, as a whole, are known as speleothemes, and are a very precious database regarding the geological, environmental and especially climatic evolution of the past. Chemical sediments form when water saturated with calcium carbonate is subjected to variations in temperature or CO_2 content, or gets concentrated due to evaporation, thus becoming oversaturated. Therefore excess carbonate deposits in the form of concretions, known as speleothemes, which can have various shapes and morphology depending on where they form, the way the minerals precipitate, etc. Most speleothemes are made of calcite, which is surely the most widespread cave-mineral. Almost all speleothemes form in a sub-aerial environment. However, in particular conditions, in small closed basins with over-saturated water, even underwater speleothemes can form. Most speleothemes that can be observed in flooded caves formed in the sub-aerial zone, and were subsequently brought to the phreatic zone thanks to a flooding of the conduits with fresh or salt water, or as a consequence of a subsequent rise in the base-level. Speleothemes form more rapidly and more abundantly in warm climates. Growth generally takes place in concentric bands, and their chemical composition (particularly with regard to the oxygen isotopes) is controlled by that of both water and atmosphere in which they formed. Therefore they provide important data regarding the climate of the past.

Physical deposits

Physical deposits include a large variety of materials that accumulate in caves thanks to gravity (deposits, blocks abd boulders due to break- down) or transported by water. Sediments may be autochthonous, produced in the cave (as in the case of blocks and boulders due to breaking down, or the clay formed by the insoluble minerals in the calcareous rock), or allochthonous, transported into the caves by different agents, generally by water. The material transported by water can be distinguished because its rounded characteristics, the more rounded the grains, the longer is the time they have been being transported and the softer the type of rock.
Water capacity to transport material or competence, depends on its energy, in particular its speed and its density and, naturally, the density and weight of the material to be transported. Since most rocks have a density around 2.7 g/cm3, which is therefore the same for all types of rock, this can be considered a constant parameter. Instead of the weight, the mean diameter of the granules or pebbles of material to be transported can be used, i.e. what is called the granulometry. The higher the speed of water, the greater the size of the particles the current is able to transport. The size of the particles vary from millimetric to metric, as for example in large floods where water is dense with sediments. In a cave, the transport of large-sized materials is rare, because even if water were to have a competence able to move large blocks, transportation would soon be stopped by the size of the conduits. The most commonly transported materials are generally sand and pebbles.
When the speed of the current decreases, also the competence decreases and water abandons the coarser material, thus carrying out a granulometric classification of the material, i.e. a separation depending on the size of the particles. If coarse material is found in a flooded conduit this means that the current may have high speeds. Since the water is constantly moving, and with very variable speeds, it is not rare to observe a continuous reorganization and changing in the shapes and granulometry of the deposits on the bottom of a gallery. A visit to a well known gallery immediately after a flooding event can hide some surprises that are not always welcome, as for example the occlusion of the narrow and

smaller passages, which must be opened again with difficult digging operations, or the presence, mainly near sinkholes, of material carried by the flow into the cave, such as large tree trunks or vegetable rubbish. These changes in the deposits can also cause barrages in the outflow of the vadose zone, thus giving origin to the formation of lakes or siphons. The presence of fine-grained silts and clays on the ceiling and on the walls of a sub-aerial cavity, specially where the material seems humid and fresh, not dusty or dry, can indicate that the gallery may undergo a total flooding, and therefore great care must be paid while exploring during particularly rainy periods.
Physical deposits, specially those coming from outside, can provide precious information with regard to the evolution in the region. In fact these may contain remains of rock formations that are now completely worn out by erosion, or testify the passage of glaciers, or document the alternation of hot and cold periods. A detailed study of the chemical and physical deposits in a cave is fundamentally important for the reconstruction of the most recent geological and climatic history.