2
Fundamental Concepts

This morning, the authors of this book begins their day in different parts of the province of Pisa, Italy, each making their way to their offices in the CNR campus. At lunchtime, they join colleagues at the institute's canteen, a daily ritual embedded in the rhythms of scientific life at CNR. In the afternoon, they both move to a meeting at the University of Pisa, in the city centre. Then, their evening activities diverged entirely: one attended a classical music concert, the other visited a gym on the city's outskirts before heading to a pub in the historic center.

These movements can be formally described as **mobility trajectories**, sequences of spatial positions traced over time (see **??**). Each spatial position along the way corresponds to a geographic coordinate: a pair of latitude and longitude values that precisely locates a point on Earth's surface. When such coordinates are chronologically ordered, they capture the structure of an individual's movements. Thousands of individuals move daily through the city of Pisa, each tracing their unique mobility trajectory.

To systematically study mobility trajectories – their origins, destinations, lengths – we typically divide geographic space into spatial units using **spatial tessellations**. These can take the form of administrative divisions, such as neighbourhoods or municipalities, or follow geometric schemes like squares or hexagons. Within this partitioned space, we can define **mobility flows**, which quantify the number of individuals moving from one tile to another within a given time frame (see **??**b).

This chapter introduces the essential concepts and tools for analysing human mobility. We begin by describing the geographic coordinate system, then move to the definition of individual trajectories. We then show how spatial tessellations allow us to aggregate these trajectories into mobility flows, providing a structured way to interpret the patterns that shape how people move through space.

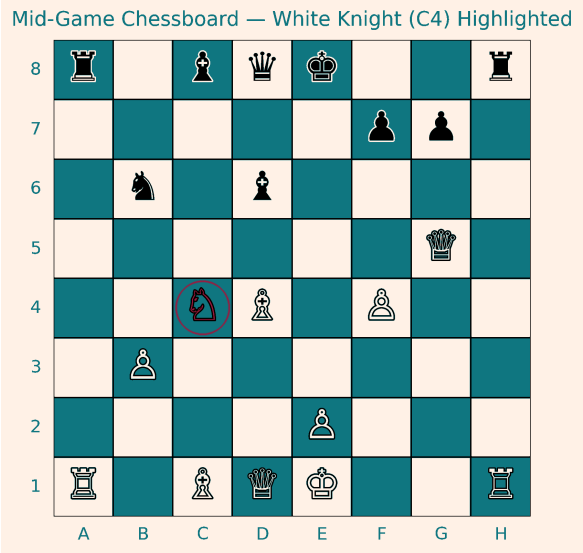

Figure 2.1 Example of a mid-game chessboard configuration. The board shows both white and black pieces positioned dynamically, with the white knight highlighted near the center at C4. The diagram illustrates how spatial positions can be represented using a coordinate system (A–H, 1–8), a useful analogy for describing discrete spatial locations in mobility analysis.

2.1 Geographic coordinates

Where are you now? Although this question sounds simple, a precise answer depends on the coordinate system at hand.

In games like Battleship or chess, location is unambiguous: letters label one axis, numbers the other, and the ordered pair (x, y) identifies a single square on a flat board. That grid is a complete, fixed reference frame (Figure 2.1).

Real spaces are less obliging. Even within a room there is no canonical coordinate system. One must choose a local frame – for example, select a corner as the origin, define two perpendicular axes along the walls, and express position as "x centimetres from one wall and y centimetres from the other." Many such frames are possible, each adequate once its origin, axes, units, and orientation are specified.

The Earth is neither a board nor a flat room. It is (approximately) a spheroid, and we occupy a curved surface embedded in three dimensions. A grid of equal squares that works on a plane will distort – stretch, skew, or tear – when imposed on a globe. Any rigorous statement of position on Earth therefore requires a

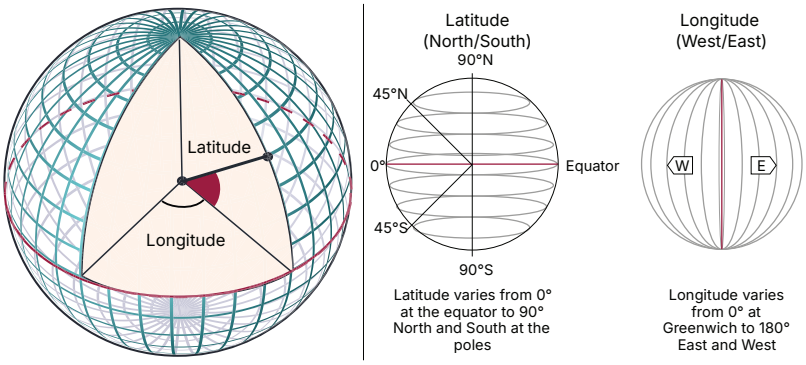

Figure 2.2 Latitude and longitude in the geographic coordinate system. Left: Angular measurements of latitude and longitude on a spherical model of Earth. Latitude measures the angular distance north or south of the equator, while longitude measures the angular distance east or west of the prime meridian. Right: Diagram of meridians and parallels. Parallels are horizontal lines parallel to the equator and represent lines of equal latitude; meridians are vertical lines connecting the poles and represent lines of equal longitude. Sources. Left: Peter Mercator, Own work, derived from Sphere wireframe 10deg 10r.svg, Public Domain. Available at: https://commons.wikimedia.org/w/index.php?curid=12226167 Right: Unknown author, vector version of a U.S. government image, Public Domain. Available at: https://commons.wikimedia.org/w/index.php?curid=92353626

properly defined **geographic coordinate system** (GCS), not merely a planar grid. A GCS is a reference framework used to specify the location of objects on the Earth's surface. It relies on angular measurements – latitude and longitude – which represent angles measured from the Earth's center to locations on its surface (see Figure 2.2).

These coordinates enable the unique identification of locations on the Earth's curved, three-dimensional surface. Latitude refers to the angular distance of a location north or south of the equator, which is defined as 0° latitude. Lines of equal latitude, called parallels, run horizontally around the Earth and are always parallel to the equator. Latitude values range from 0° to 90°, and must be specified as either north (N) or south (S) of the equator. Longitude is the angular measurement that defines a location's position east or west of the prime meridian, an imaginary line set at 0° longitude that passes through Greenwich, UK. Lines of equal longitude, known as meridians, extend from the North Pole to the South Pole and converge at both poles. Longitude values range from 0° to 180°, and each must be designated as either east (E) or west (W) of the prime meridian.

Since latitude and longitude are angles, they are described in degrees-minutes-

seconds (DMS) format. However, the decimal degrees (DD) format is more common in practice. Coordinates in DMS format can be converted in DD format using the formula:

$$\text{decimal degrees} = deg + \frac{min}{60} + \frac{s}{3600} \tag{2.1}$$

where *deg* indicate the degrees, *min* the minutes and *s* the seconds of the angle. If the coordinate lies in the southern or western hemisphere, the resulting value should be made negative. For example, 45° 30′ 00″ S corresponds to $-(45 + \frac{30}{60} + \frac{0}{3600}) = -(45 + 0.5 + 0) = -45.5$ decimal degrees. The minus sign reflects the fact that the location lies south of the equator and therefore represents a negative latitude. Coordinates in DD format are typically presented in the order (latitude, longitude). For example, the coordinates of Paris are commonly written as (48.8566, 2.3522), where 48.8566 represents the latitude and 2.3522 the longitude.

In some cases, coordinates are expressed in radians rather than decimal degrees. Since one degree corresponds to $\frac{\pi}{180}$ radians, the conversion from DMS to radians is given by

$$\text{radians} = \left(deg + \frac{min}{60} + \frac{s}{3600}\right) \cdot \frac{\pi}{180} \tag{2.2}$$

Again, the result should be made negative if the direction is South or West. For instance, 45° 30′ 00″ S corresponds to $-(45 + \frac{30}{60} + \frac{0}{3600}) \cdot \frac{\pi}{180} \approx -0.7941$ radians.

CURIOSITY 2.1

Discovering the Longitude

For centuries, determining longitude at sea was one of the greatest scientific and navigational challenges (Sobel, 1996). While latitude could be estimated with reasonable ease by measuring the angle of the Sun or stars above the horizon, longitude required knowing the time difference between a reference location and the local time. Since the Earth ro-

tates 15° per hour, every hour of time
difference corresponds to 15° of longitude.
But in the 17th and 18th centuries, keeping accurate time on a moving ship was nearly impossible. Pendulum clocks were unreliable at sea due to constant motion and variations in temperature and humidity. This uncertainty made longitudinal navigation so difficult that "searching for the longitude" became a proverbial expression for attempting the impossible — so much so that Swift's *Gulliver's Travels* lists the "discovery of the longitude" among fanciful, world-perfecting inventions we would achieve if immortal (Part III, Ch. X) (Swift, 2025).
The breakthrough came in the 18th century, when English clockmaker John Harrison developed the first highly accurate marine chronometers, culminating in the celebrated H4 timekeeper. Harrison's work, funded in part by the British government's Longitude Act of 1714, eventually allowed sailors to compare local solar time with Greenwich Mean Time (GMT) and thus compute their longitude with precision.
The problem of finding the longitude not only revolutionized navigation and global exploration, but also spurred innovations in astronomy, timekeeping, and geodesy. Today, we take for granted the instantaneous precision of GPS, but it was the centuries-long "search for the longitude" that paved the way.

2.1.1 World Geodetic System 1984

Defining the origin lines for latitude and longitude – the equator and the prime meridian – is a necessary step in establishing a GCS, but it is not sufficient. To fully describe the position of a point on the Earth's surface, a geodetic datum must be specified.

A geodetic datum is a mathematical model that defines how the Earth's surface is represented in the GCS. It consists of three components:

(i) The latitude and longitude of an initial reference point (the origin of the system);
(ii) A reference ellipsoid that approximates the shape of the Earth;
(iii) The vertical separation between the ellipsoid and the geoid (i.e., the Earth's actual physical surface) at the origin.

The second component is necessary because the Earth is not a perfect sphere. Due to its rotation, the Earth is slightly flattened at the poles and bulges at the equator, a shape more accurately modelled as an ellipsoid defined by its semi-major axis a (the radius at the equator) and semi-minor axis b (the radius from the center to the poles). The flattening parameter, defined as $f = (a - b)/a$, quantifies how much the ellipsoid deviates from a perfect sphere and is crucial for accurately modelling global positions.

The third component of the datum concerns the alignment of the ellipsoid relative to the Earth's surface. Because the Earth's shape is irregular (characterised by mountains, valleys, depressions and gravitational anomalies), this alignment determines how closely the ellipsoid matches the real-world terrain. Traditional local datums (such as the European Datum 1950 or the Indian Datum) were designed to provide high positional accuracy within a specific geographic region. To achieve this, the ellipsoid was often shifted and tilted to fit the local geoid better, resulting in excellent local precision but poor global consistency.

Modern global applications, such as GPS, satellite imagery, and international mapping, require a consistent, Earth-centred reference system. This need led to the adoption of the World Geodetic System 1984 (WGS84), a globally standardised datum developed by the U.S. Department of Defense for the Global Positioning System.[1] WGS84 defines an Earth-centred ellipsoid with the following parameters:

- Semi-major axis: $a = 6,378,137$ meters;
- Semi-minor axis: $b = 6,356,752.3142$ meters;
- Flattening parameter: $f \approx 0.003352810664747$.

The prime meridian in WGS84 corresponds to the IERS Reference Meridian, which closely aligns with the historical Greenwich Meridian and defines longitude 0°.

2.1.2 Geodesic Distance

Since the Earth is a three-dimensional, curved surface rather than a flat plane, traditional Euclidean distance formulas are not suitable for measuring distances between points on its surface. Instead, we must use the **geodesic distance**, which accounts for the Earth's curvature and provides a more accurate representation of spatial separation. The most straightforward method for computing the geodesic distance between two points uses the haversine formula, assuming Earth is

[1] Department of Defense World Geodetic System 1984: Its definition and relationships with local geodetic systems (TR8350.2). Bethesda, MD: NIMA; 2000.

spherical (Inman, 2012; Shylaja, 2015). Mathematically, given two spatial points $l_1 = (x_1, y_1)$ and $l_2 = (x_2, y_2)$, the geodesic distance is computed as:

$$d(l_1, l_2) = 2R \arcsin \sqrt{\sin^2\left(\frac{\Delta\phi}{2}\right) + \cos\phi_1 \cos\phi_2 \sin^2\left(\frac{\Delta\lambda}{2}\right)} \tag{2.3}$$

where R is the Earth's radius (mean radius $R \approx 6371$ km); ϕ_1 and ϕ_2 are the latitudes of l_1 and l_2 in radians; $\Delta\phi = \phi_2 - \phi_1$ is the difference in latitudes; λ_1 and λ_2 are the longitudes of l_1 and l_2 in radians; and $\Delta\lambda = \lambda_2 - \lambda_1$ is the difference in longitudes. See Advanced Topics 2.A for details on how equation (2.3) is derived from the haversine formula.

Note that, since Earth is not a sphere (but an ellipsoid), using the haversine formula introduces a relative error that can reach at most approximately 0.5% of the actual distance (Earle, 2006). More accurate geodesic methods, such as Vincenty's algorithm, Karney's algorithm or solutions based on the WGS84 ellipsoid (Karney, 2013), account for this shape and provide improved precision. While spherical approximations are sufficient for many applications, they may yield significant errors over long distances or when high spatial accuracy is required (Earle, 2006). As an example, consider the computation of the distance between Paris (latitude 48.8566, longitude 2.3522) and New York City (latitude 40.7128, longitude −74.0060). If we apply the haversine formula, the resulting distance is approximately 5, 837, 241 meters. Using Karney's algorithm, the computed distance is higher, approximately 5, 852, 935 meters. This yields an absolute difference of 15, 694 meters (roughly 15.7 kilometres) and a relative error of about 0.27%. This discrepancy underscores the importance of selecting an appropriate distance computation method, especially in applications where precision is critical.

2.1.3 Visualizing geographic coordinates on your screen

When we visualise geographic locations, we almost always do so on a flat medium – paper maps, computer monitors, smartphone screens. Converting positions expressed as latitude and longitude on an ellipsoidal Earth into planar (x, y) coordinates is the role of a map projection.

A map projection systematically transforms parallels and meridians to a plane, enabling two practical advantages: *(i)* we can use two-dimensional maps and *(ii)* perform calculations in a planar coordinate system rather than angular units. The trade-off is unavoidable distortion: no projection can simultaneously preserve area, shape, distance, and direction everywhere. Consequently, hundreds of projections exist, each optimised for certain properties and regions; the Mercator

projection, for example, preserves local shape but greatly inflates area at high latitudes.

Because the choice of projection depends on purpose and place, a detailed treatment lies beyond the scope of this book; readers seeking a fuller discussion should consult standard GIS texts, e.g., Chapter 2 in Chang (2018).

CURIOSITY 2.2

What if... Earth were flat?

If Earth were flat, the Euclidean distance would be the correct measure of the distance between any two points. Given two points $l_1 = (x_1, y_1)$ and $l_2 = (x_2, y_2)$, their Euclidean distance is provided by:

$$d(l_1, l_2) = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2}.$$

However, the Earth is not flat and, contrary to a common misconception, ancient and medieval scholars knew this. Since the time of the Ancient Greeks, the Earth's sphericity was widely accepted; most famously, Eratosthenes measured its circumference in the 3rd century BCE.

Neither did educated individuals in the Middle Ages subscribe to the notion of a flat Earth. This misconception was largely invented in the 19th century to portray the medieval period as a time of ignorance and superstition. The spherical nature of the Earth—though we now know the planet is more accurately described as an ellipsoid—was well understood by medieval scholars, who drew on classical knowledge preserved

through figures such as Isidore of Seville, Bede the Venerable, and Thomas Aquinas.
A compelling visual testament to this understanding appears in countless medieval paintings and sculptures across Europe, where monarchs, emperors, and religious figures are depicted holding the *globus cruciger* (a globe surmounted by a cross) symbolising both the Earth's spherical shape and Christ's dominion over the world. Although medieval cosmology placed Earth at the centre of the universe (a geocentric model), it still conceived of the Earth as a sphere.
The flat-Earth myth gained traction in the 1800s through writers like Washington Irving and John William Draper, but it has no basis in historical fact.

Behind the Curve (Netflix, 2018), a documentary exploring contemporary Flat Earthers and their belief in a conspiracy to hide the "truth" that the Earth is flat – behindthecurvefilm.com.

In the previous page, above: Rowbotham's Map of the Earth, from Earth Not a Globe. Below: Frederick II surrounded by his subjects. Detail from a miniature in the "Exultet", first half of the 13th century. Picture taken from the Museo Diocesano in Salerno, Italy.

2.2 Trajectories, Tessellations and Flows

The study of human mobility relies on a small set of foundational objects that serve as building blocks for modelling, analysing, and interpreting movement patterns. Three concepts are particularly central:

- Mobility trajectories describe individuals' movements through space and time.
- Spatial tessellations partition the geographic space into discrete regions.
- Mobility flows represent the aggregated movements between such regions.

Most of the literature in human mobility is structured around these three elements, which are used either independently or in combination to characterise human mobility at different scales and resolutions.

2.2.1 Mobility trajectories

Definition 2.1 (Mobility Trajectory) A trajectory T is a temporally ordered sequence of spatio-temporal points $T = (p_1, p_2, \ldots, p_n)$, where each point p_i is defined as a tuple $p_i = (x_i, y_i, t_i)$, with (x_i, y_i) representing the spatial coordinates, and t_i denoting the corresponding timestamp. The sequence is strictly ordered in time: $t_1 < t_2 < \cdots < t_n$, which ensures that the points in the trajectory are chronologically arranged. Equivalently, the trajectory can be written as: $T = ((x_1, y_1, t_1), (x_2, y_2, t_2), \ldots, (x_n, y_n, t_n))$.

A trajectory captures the movement of an entity, such as a human, animal, vehicle, object, or particle, through space over time. In the context of human mobility, spatial coordinates are typically limited to two dimensions: latitude (x) and longitude (y). A third dimension (z), corresponding to altitude, is often omitted due to its limited availability or relevance in most mobility datasets. As a result, trajectories in mobility studies are frequently represented as sequences of 3-tuples (x_i, y_i, t_i) or even 2-tuples if the temporal dimension is separately encoded.

Mobility trajectories are often visualised in two dimensions – longitude and latitude – while ignoring altitude and time. Figure 2.3 (left) shows a trajectory $T = (p_1, p_2, \ldots, p_6)$ representing a sequence of places visited in Pisa, Italy, plotted in two dimensions. Each point is recorded at a constant temporal interval of five minutes. Figure 2.3 (right) displays the same trajectory in three dimensions by adding the temporal axis. Although the two-dimensional representation is more commonly used in the scientific literature, the three-dimensional visualisation offers a more complete and informative view of the trajectory, capturing both spatial and temporal aspects.

As we will discuss in Chapter 3, the structure and informativeness of a mobility trajectory are influenced by the characteristics of the underlying data, particularly the sampling rate and spatial-temporal granularity. High-resolution data sources, such as GPS, provide frequent and precise location updates, enabling detailed analysis of mobility trajectories, including velocity and acceleration. In contrast, sparse and irregular data like mobile phone records or social media check-ins offer limited temporal resolution and often require data preprocessing techniques (discussed in Chapter 4) to reconstruct meaningful trajectories. The choice of sensing modality entails trade-offs between accuracy, coverage, and privacy, ultimately shaping the insights that can be drawn from the resulting mobility traces. Note that a mobility trajectory can represent the movement of an object with varying degrees of detail. For instance, it might depict an individual's complete daily trajectory, which likely includes periods of immobility, such as time spent at an office. This semantic information is not directly ascertainable

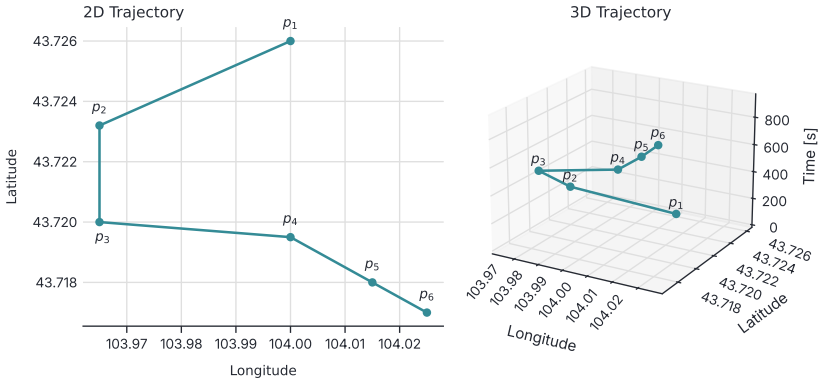

Figure 2.3 Visualising a mobility trajectory in space and time. Left: The trajectory T = (p1, p2, , p6) is plotted in two dimensions (longitude and latitude), as commonly done in mobility studies. Each point is sampled at a constant interval of 5 minutes. Right: The same trajectory is shown as a spatiotemporal curve in three dimensions, with time (in seconds since the start) added as the vertical axis. This 3D representation reveals the temporal structure of the movement, offering a more complete view of the trajectory's evolution over time.

from the raw mobility trajectory. Data preprocessing steps, as discussed in Chapter 4, are required to extract such semantic insights from the trajectory data.

A mobility trajectory reveals indirect aspects of mobility behaviour. For example, from a trajectory we can extract the following features:

- Velocity: the average velocity can be computed by measuring the spatial distance between consecutive points and dividing it by the time interval between them. For a pair of points $p_i = (x_i, y_i, t_i)$ and $p_{i+1} = (x_{i+1}, y_{i+1}, t_{i+1})$, with $t_{i+1} > t_i$, the average velocity is given by

$$v_i = \frac{d(p_i, p_{i+1})}{t_{i+1} - t_i}$$

  where $d(p_i, p_{i+1})$ denotes the geodesic distance between the two points.
- Directionality: The heading or direction of movement can be derived from the orientation of consecutive trajectory segments p_i and p_{i+1}. It is typically computed as the angle of the vector between p_i and p_{i+1} relative to a reference axis:

$$\theta_i = \arctan 2(y_{i+1} - y_i, x_{i+1} - x_i)$$

  where (x_i, y_i) and (x_{i+1}, y_{i+1}) are the coordinates of p_i and p_{i+1}, respectively.

Figure 2.4 Cave Canem (Beware of dog), a Roman mosaic from Pompeii, now at the Archaeological Museum of Naples, Italy. Roman mosaics were made throughout the Roman Republic and later Empire, and were used in a variety of private and public buildings, on both floors and walls [9]. They are constructed from geometrical blocks called tesserae, placed together to create the shapes of figures, motifs and patterns. Figure from this link under the Creative Commons Attribution-Share Alike 3.0 Unported license.

The resulting angle θ_i expresses the instantaneous direction of movement, providing insight into the preferred travel orientations of the individual.

- Duration: The duration of the movement captured by the trajectory is simply the time elapsed between the first and last points in the sequence, i.e., $t_n - t_1$, or between two consecutive points, i.e., $t_i - t_{i-1}$. Duration helps characterise the temporal extent of the mobility trajectory.

- Total distance: The sum of distances between consecutive points along a trajectory. The total distance of the trajectory, computed as the distance between consecutive points of the trajectory

$$\sum_{i=1}^{n-1} d(p_i, p_{i+1})$$

where $d(p_i, p_{i+1})$ is the geodesic distance.

- Personal preferences: By analysing the recurrence patterns in a mobility trajectory, such as frequently visited locations, time spent at specific places, or preferred travel times, one can infer individual preferences and habits. For example, the most frequently visited location may correspond to home, and temporal regularities may suggest commuting routines. This will be covered in Chapter **??**.

2.2.2 Spatial Tessellations

In many mobility tasks, the geographic space is discretised by mapping the coordinates to a spatial tessellation, i.e., a covering of the two-dimensional space using a countable number of geometric shapes called tiles, with no overlaps and no gaps (Grünbaum and Shephard, 1986). The term tessellation derives from the Latin word *tessella*, which referred to a small square tile used in mosaic work. Tessella is a diminutive of *tessera*, meaning "tile" or "cube".

Definition 2.2 (Spatial tessellation) Given a geographic area $\mathcal{A}$, a set of geographic polygons $\mathcal{S} = s_i : i = 1, \ldots, n$ is a spatial tessellation of $\mathcal{A}$ if:

(i) Each s_i is a polygon called a **tile**.
(ii) The tiles are non-overlapping, i.e., $s_i \cap s_j = \emptyset, \forall i \neq j$.
(iii) The tiles completely cover the area: $\cup_{i=1}^{n} s_i = \mathcal{A}$.

Spatial tessellations can be broadly classified into regular and irregular types, based on the geometry of their tiles and the rules by which they are constructed.

Regular Tessellations

A regular tile is one with equal side length and equal internal angles. Examples are the equilateral triangle, with three equal sides and 60° internal angles; the square, with four equal sides and 90° angles; and the regular hexagon, with six equal sides and 120° angles.

Although there are infinitely many spatial tessellations made from a single type of regular tile, only few of them meet the stricter condition that every point in the tessellation is of the same type, i.e., the arrangement of tiles around each point is identical. These are known as uniform tessellations (Grünbaum and Shephard, 1986). Among them, the tessellations composed entirely of equilateral triangles, squares, or regular hexagons are called **regular tessellations**, as they use a single regular polygon that fits together without gaps or overlaps.

The geometric simplicity, ease of implementation, and computational efficiency of regular tessellations have made them a common choice in human mobility studies, particularly when dealing with large-scale data. However, their regularity comes at a cost, as they impose a rigid and arbitrary spatial structure whose boundaries rarely align with natural features, urban infrastructure, or meaningful patterns of human mobility. Among regular tessellations, the most widely used in mobility analysis are square tessellations and hexagonal tessellations.

Square tessellations. Also named square grids, they divide the spatial domain into a lattice of equally sized square cells (see **??**a). They are straightforward to construct, easy to index, and computationally efficient, all properties that make

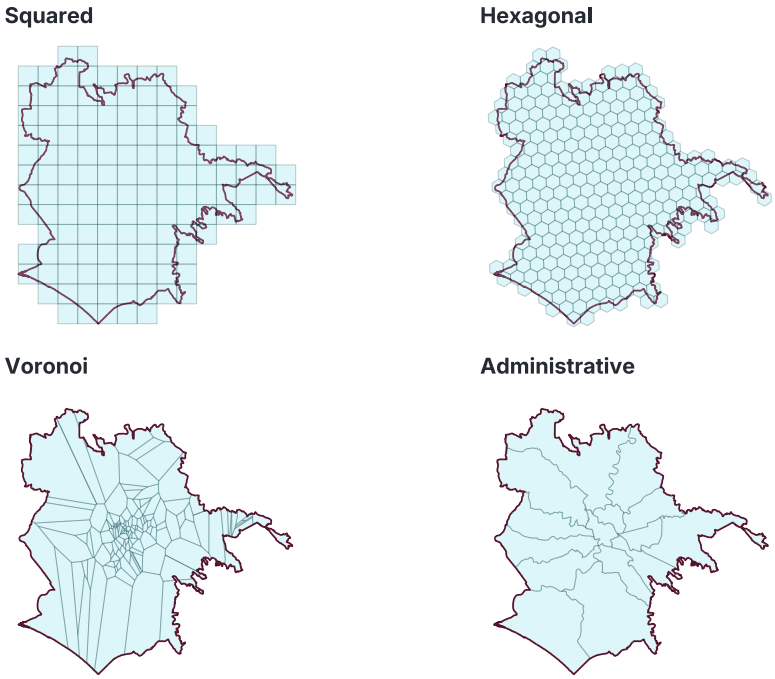

Figure 2.5 Different types of tessellation over the same geographic area (City of Rome)

them attractive for data storage and spatial queries. However, square tessellations do have a main limitation: the unequal distance between adjacent and diagonal neighbours (see Figure 2.6) can affect spatial smoothing, interpolation, and mobility modelling, leading to results that may reflect the shape of the grid more than the underlying phenomena.

Hexagonal tessellations. Also known as hexagonal grids, they use regular hexagons to tile the space (see **??**b). They are often preferred in spatial analysis due to their geometric advantages, as they possess the following two properties:

- *Uniform adjacency*, whereby distances between the centroid of a given tile and those of neighbouring tiles are the same. Each hexagonal tile has six neighbours at equal distance from its centre (see Figure 2.6). This ensures isotropic connectivity, i.e., there is no preferred direction, and spatial relationships are more balanced.

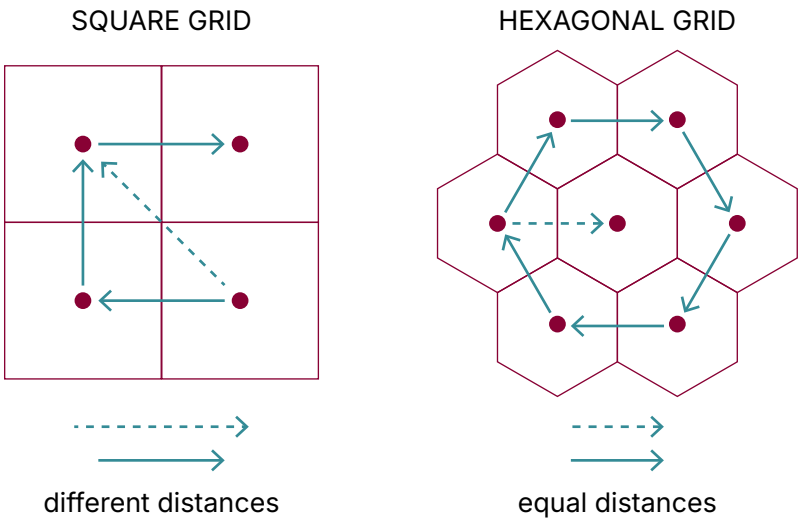

Figure 2.6 Adjacency distances in square (a) vs hexagonal (b) tessellations.

- *Uniform orientation*, which means that all tiles in the tessellation are aligned in the same direction, none are rotated or flipped relative to the others.

Square tessellations possess the uniform orientation property but not uniform adjacency. Hexagons also more closely approximate circular areas of influence, which can be a more realistic representation of how individuals interact with space. Their compact and symmetrical shape helps reduce angular distortion. Despite these benefits, like square grids, they are context-blind: their cell boundaries are externally imposed and independent of the actual distribution of people, activities, or places.

An advantage of all regular tessellations is their flexibility in spatial resolution. We can easily adjust the size of tiles to match the scale of interest. Yet this flexibility also introduces a trade-off: larger cells may obscure fine-grained spatial and mobility dynamics, while smaller ones can lead to sparse data, increased noise, and unstable estimates (Openshaw, 1984; Gehlke and Biehl, 1934). Thus, while regular grids offer practical advantages, they require careful calibration to avoid misrepresenting the very spatial patterns they are meant to capture.

Irregular Tessellations

Irregular tessellations divide space into units of varying shapes and sizes, often shaped by pre-existing geographical, social, or administrative boundaries. Unlike regular tessellations, which impose a uniform and artificial structure on space, irregular tessellations aim to capture more nuanced or context-sensitive spatial divisions. In human mobility analysis, two types of irregular tessellations are common: administrative unit tessellations and Voronoi tessellations.

Administrative unit tessellations. These types of tessellations arise from political, institutional, or historical processes and include spatial divisions such as countries, municipalities, districts, and census tracts (see **??**d). Their widespread use is largely due to their familiarity and the availability of socio-economic data aggregated at these levels. These tessellations also present several limitations. Administrative boundaries often encompass areas with contrasting land uses, population densities, and socio-economic characteristics, resulting in heterogeneous zones that can distort aggregate measures. Analytical outcomes may therefore vary significantly depending on the size, shape, and configuration of the units used. Furthermore, administrative units are rigid and predefined, offering limited adaptability to the spatial processes being studied. Their scale and geometry can differ considerably even within the same urban area, which complicates cross-regional comparisons. As a result, researchers must interpret findings with caution, as observed patterns may stem from zoning artefacts rather than genuine mobility behaviours.

Voronoi tessellations. Voronoi diagrams partition space according to a set of seed points (or generators), with each resulting tile containing all locations that are nearer to its seed point than to any other (see **??**c). Mathematically, let $P = p_1, \ldots, p_n$ be a finite set of n distinct seed points located in geographic area $\mathcal{A}$, and let $d(p, p_i)$ be the geodesic distance between a location p and seed point p_i. We define a region given by

$$V(p_i) = \{p | p \in \mathcal{A}, d(p, p_i) < d(p, p_j), j \neq i, j = 1, \ldots, n\}$$

as the Voronoi tile (or Voronoi polygon) associated with point p_i and the set given by $v(P) = \{V(p_1), \ldots, V(p_n)\}$ as the Voronoi diagram of P. The tiles of $V(p_i)$ contain all locations in the geographic area $\mathcal{A}$ which are closer to p_i than any other seed point, while the lines and the vertices of $V(p_i)$ represent those locations which are equidistant from two or more seed points. Voronoi tessellations are inherently adaptive: their geometry reflects the spatial distribution of the seed points. When seed points correspond to meaningful features (such as mobile phone towers or urban amenities), the resulting cells can approximate zones of influence or service areas, making Voronoi tessellations a powerful tool for modelling spatial accessibility or flow. Compared to regular

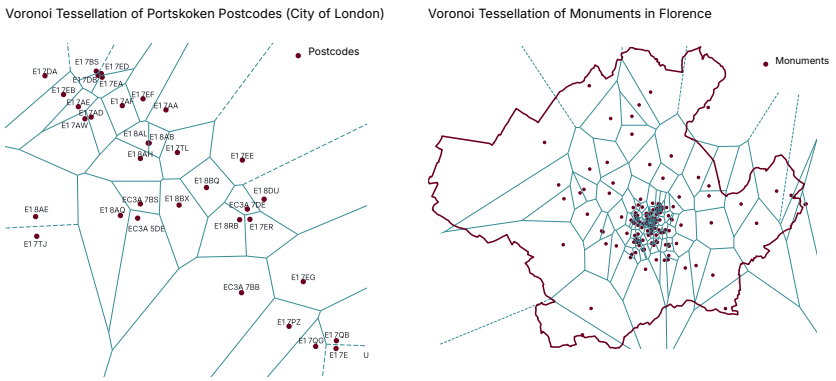

Figure 2.7 (Left) Voronoi tessellation based on postcode centroids in the Portsoken ward (City of London). (Right) Voronoi tessellation over the municipality of Florence, generated using main tourist attractions as seed points.

grids, they offer a more flexible and heterogeneous representation of space, one that can better align with underlying phenomena.

A common application of Voronoi tessellations is to reconstruct spatial divisions when explicit boundaries are missing or incomplete. For instance, in the UK, the postcode system is widely used to georeference socio-economic data. However, individual unit postcodes do not have officially defined boundaries, which creates challenges when integrating such data with census-based geographic units. A practical solution is to generate Voronoi polygons around each address point within a postcode, thereby approximating the spatial extent of each unit (see Figure 2.7a). Another example, illustrated in Figure 2.7b, involves creating zones of influence around tourist attractions. Each Voronoi tile represents the area that is geographically closest to a specific attraction, effectively partitioning space into catchment areas based on proximity. This method is particularly useful for analysing spatial accessibility or estimating the spatial reach of points of interest.

Note that the adaptability of Voronoi tessellations comes with caveats. The quality and interpretability of the tessellation depend entirely on the seed points. Sparse distributions lead to oversized and coarse partitions, while poorly placed or arbitrarily generated seeds can produce cells that do not reflect any meaningful spatial structure. Additionally, Voronoi tessellations assume space is continuous and isotropic – i.e., that movement or influence spreads equally in all directions – a condition that rarely holds in real-world urban environments characterised by physical barriers, infrastructure, or zoning constraints. While Voronoi-based partitions can reduce the artificiality of regular grids and mitigate some of their

biases, they remain sensitive to design assumptions and data limitations. The selection of an appropriate spatial tessellation is a critical step in any human mobility study, as it directly influences the insights gained and the conclusions drawn from the analysis. Each type offers unique advantages and disadvantages, making the choice dependent on the specific context and objectives of the research. Table 2.1 summarises the pros and cons of each tessellation type.

Tessellation type	**Reg./Irreg.**	**Pros**	**Cons**	**Best used when**
Square	Regular	Simple to implement and index; flexible in spatial resolution.	Unequal neighbor distance; context-blind.	You need simple spatial partitioning.
Hexagonal	Regular	Uniform adjacency and orientation; better approximates circular areas.	More complex to implement; context-blind.	You want isotropic tiles for uniform spatial coverage.
Administrative	Irregular	Familiar and interpretable; available socio-economic data.	Rigid and predefined; can be misaligned with mobility patterns.	You are working with official statistics.
Voronoi	Irregular	Adapts to spatial data; reflect influence zones; data driven.	Dependent on seeds; ignore real barriers; complex computation.	You need to model zones of influence.

Table 2.1 *Pros and cons of tessellation types.*

CURIOSITY 2.3

Voronoi tessellations are everywhere

From the irregular spots on a giraffe's coat to the delicate vein structure in a dragonfly's wing, many natural patterns resemble Voronoi tessellations. Even something as familiar as a honeycomb hints at this structure when the hive cells are viewed as regular Voronoi partitions.

Researchers have leveraged this idea to predict the surprising geometry of epithelial cells, i.e., those lining our skin, organs, and blood vessels. Their models anticipated a peculiar, previously unknown shape now dubbed the "scutoid", which was subsequently found to be ubiquitous in human biology, thanks in part to its Voronoi-like organizing principles.

We also use Voronoi tessellations everyday, even without realizing it. When we seek the nearest café, or when urban planners determine service area for hospitals, and when regional planners outline school districts, we all consider Voronoi tessellations. Each café, school, or hospital is the seed point from which a Voronoi tessellation is generated.

Voronoi tessellations and scutoids are everywhere, (Scientific American blog post).

One algorithm can describe Giraffe skin, Leaf structure, and Cracked earth? Voronoi! (video on YouTube).

In the previous page: bee on his alvear - (Scientific American). Image licensed under Creative Commons.

2.2.3 Mobility Flows

Mobility flows describe the movement of groups of people between pairs of locations on a geographic area. A typical example is migratory flows among countries or within statistical units within countries. The description of mobility flows requires the geographic area to be divided into meaningful spatial units using a spatial tessellation.

DEFINITION 2.1

Mobility Flow

Let $\mathcal{S} = s_1, \ldots, s_n$ be a spatial tessellation of a geographic area $\mathcal{A}$, where each $s_i \in \mathcal{A}$ is a tile. A mobility flow from tile s_i to tile s_j is the count of movements observed from individuals travelling from s_i to s_j within a given time window Δ.

A mobility flow represents the movement of individuals from one tile in the spatial tessellation to another. For example, if we apply an irregular spatial tessellation to New York City, such as one based on neighbourhoods, a mobility flow could capture how many people travel from Manhattan to Queens over the course of a day.

To systematically capture all mobility flows within a geographic area, we use a mobility flow matrix.

DEFINITION 2.2

Mobility flow matrix

A mobility flow matrix $M^{(\Delta)} \in \mathbb{R}^{n \times n}$, where n is the number of tiles in the spatial tessellation and each element $m_{i,j} \in M^{(\Delta)}$ denotes the mobility flow from tile s_i to s_j within a given time window Δ. The matrix M_Δ is generally asymmetric, i.e., $m_{i,j} \neq m_{j,i}$.

The mobility flow matrix $M^{(\Delta)}$, also known as the origin-destination (OD) matrix, is a square matrix where each row and column corresponds to a tile in the

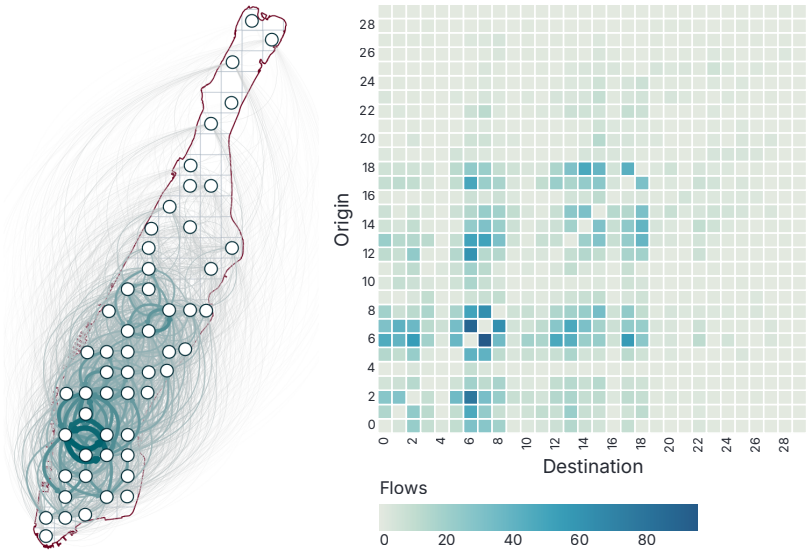

Figure 2.8 IMAGE OF FLOW NETWORK and the corresponding ADJACENCY MATRIX.

spatial tessellation. The entry $m_{i,j} \in M^{(\Delta)}$ indicates the number of individuals moving from tile s_i to tile s_j during a given time window Δ. Since the temporal window is typically understood from context, it is common to omit it from the notation and refer to the matrix simply as M.

The mobility flow matrix can be represented in two equivalent ways: as a directed, weighted network, where nodes correspond to spatial units and edges indicate flows between them; and as an adjacency matrix, where each entry $m_{i,j}$ records the number of movements from tile s_i to tile s_j. These representations enable the application of network theory and matrix-based analysis to the study of human mobility.

From trajectories to flows

Intuitively, to compute a mobility flow matrix from individual mobility trajectories, we need to translate raw movement paths into aggregated flows between spatial units. By associating each point in a trajectory with a tile from a given spatial tessellation, we can identify movements from one tile to another. Counting these transitions across all trajectories yields the entries of the flow matrix.

Mathematically, the mobility flow matrix can be computed from a collection of individual mobility trajectories as follows. Let $S = \{s_1, \ldots, s_n\}$ be a spatial tessellation of a geographic area, where each s_i is a tile in the tessellation. Let

$\mathcal{T} = \{T^{(1)}, \ldots, T^{(m)}\}$ be a collection of individual mobility trajectories that occur within a time window. Here, we assume that each trajectory $T^{(i)}$ describes a single semantic trip of an individual, i.e., the user starts at the first point of the trajectory and ends its trip at the last point of it. We assume that each trajectory $T^{(k)}$ represents a single semantic trip of an individual, wherein the user commences at the initial point of the trajectory and concludes the journey at its final point.

To compute the flow matrix $M \in \mathbb{R}^{n \times n}$, we perform the following steps:

(i) Take the starting and ending point of each trajectory. For each $T^{(k)} \in \mathcal{T}$, take the first point $(x_0^{(k)}, y_0^{(k)}, t_0^{(k)})$ and the last point $(x_n^{(k)}, y_n^{(k)}, t_n^{(k)})$. These are the origin and destination point of the trajectory, respectively.

(ii) Map origin and destination points to the corresponding tile: For each trajectory $T^{(k)}$, take the origin point $(x_0^{(k)}, y_0^{(k)}, t_0^{(k)})$ and determine which tile $s_{\text{origin}} \in \mathcal{S}$ contains the spatial location $(x_0^{(k)}, y_0^{(k)})$. Analogously, take the destination point $(x_n^{(k)}, y_n^{(k)}, t_n^{(k)})$ and determine which tile $s_{\text{destination}} \in \mathcal{S}$ contains the spatial location $(x_n^{(k)}, y_n^{(k)})$. s_{origin} and $s_{\text{destination}}$ are the origin and destination tiles of the trajectory $T^{(k)}$.

(iii) Count transitions between origin and destination tiles: Create a flow matrix $M \in \mathbb{R}^{n \times n}$ where all elements have value 0. For each trajectory $T^{(k)}$, increment the value of the element $m_{i,j}$ of the matrix if i and j are the origin and destination tiles of $T^{(k)}$, i.e., $m_{s_{\text{origin}}, s_{\text{destination}}} \mathrel{+}= 1$.

Note that while a mobility flow matrix can be derived from a collection of individual mobility trajectories, the reverse is not possible. This is because the flow matrix provides an aggregate view of movements between spatial units without retaining any information about the temporal ordering, intermediate locations, or individual-level trajectories. As a result, it lacks the granularity needed to reconstruct complete trajectories, making it a one-way transformation from detailed to aggregated data.

The mobility flow matrix derived from a given set of individual mobility trajectories can vary significantly depending on the spatial tessellation employed. In the case of regular tessellations, the size of the tiles further influences the resulting flow patterns (see Figure 2.9). As such, the choice of tessellation (and, where applicable, its resolution) is a critical methodological decision. It directly affects the structure of the resulting flow matrix and, consequently, the interpretation of any analytical results derived from it.

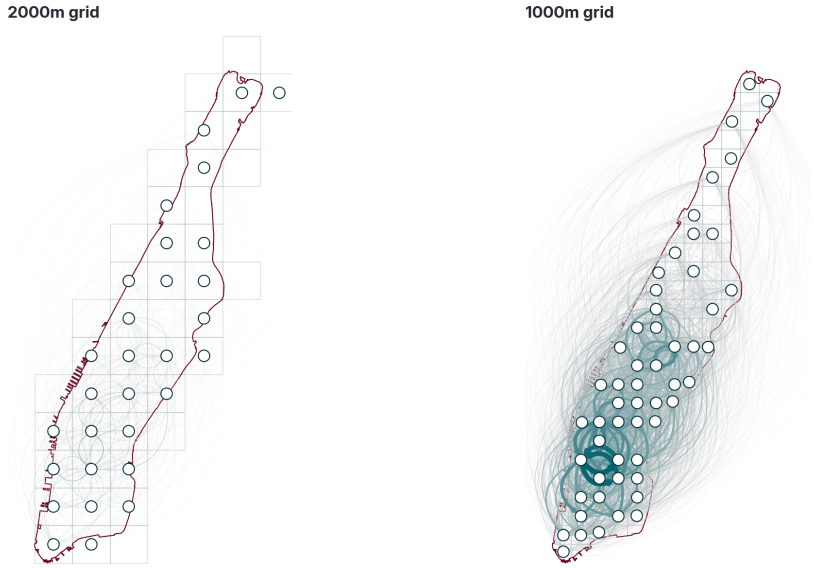

Figure 2.9 IMAGE OF A SET OF INDIVIDUAL TRAJECTORIES (on the left) AND THE CORRESPONDING MOBILITY FLOWS varying the type of tessellation.

2.3 Homeworks and exercises

EXERCISE 2.1

Why does longitude range from 0° to 180°, while latitude only ranges from 0° to 90°? Consider this question by reflecting on the geometric meaning of longitude and latitude on the Earth's surface, what these angular measurements represent, and how they are defined. Once you have formed your explanation, compare it with the answers provided by large language models such as ChatGPT and Google Gemini.

EXERCISE 2.2

Compute the distance from your home to the five largest capitals in the European Union (EU), using both the geodesic distance and the Euclidean distance. The website https://www.itilog.com/ provides the latitude and longitude of any address. Create a bar chart comparing geodesic and Euclidean distances for each capital. Are the geodesic and Euclidean distances coherent?

EXERCISE 2.3

What is the most "central" capital in the EU, i.e., the one with the lowest average geodesic distance to the other EU capitals? Create a bar chart with the average distance for each EU capital, sorted in increasing order Repeat the exercise for at least another continent

EXERCISE 2.4

Ask ChatGPT to find at least three examples of regular or irregular tessellations in natural or artificial ecosystems (other than those presented in this chapter). Why do the tiles have precisely that shape?

EXERCISE 2.5

Visualize your own trajectory. Track your movements for an entire week, tracking all points of interest you visited (e.g., home, friends' home, university, supermarket, gym, bars). Use https://www.itilog.com/ to detect the latitude and longitude of places based on their address. Visualise your trajectory using any package in Python.

EXERCISE 2.6

What is the theoretical upper bound of the mapping error of a trajectory of n points into a square tessellation where each side of the tile is s meters? And what is this upper bound for a hexagonal or triangular tessellation with a side of s meters? Use a real trajectory dataset and map the trajectory into a tessellation. Compute the error. Is it lower than the theoretical upper bound?

EXERCISE 2.7

Compute the probability of a randomly generated latitude and longitude pair falling on the land. Generate n=10k points randomly Select only the points that fall on the land and compute the probability that a random point falls into the land. Repeat this experiment 100 times and make a box plot showing the distribution of probabilities. What is the mean and the standard deviation of these probabilities? Does the average probability decrease if you execute 1000 experiments?

EXERCISE 2.8

To plan a reduction of the environmental impact of the Champions League, the Union of European Football Associations (UEFA) aims to compute the total distance travelled by all clubs during the matches in the first round of the competition. Take the position of the city of each club qualified for the Champions League this year Compute the sum of the distances travelled by each club to play its matches Make a bar chart to show the total distance travelled by each club What's the club that travels the most? And the club travelling the least? Compute and visualise the mobility flows generated by the first round of the competition

2.4 Advances Topics

2.4.1 Derivation of the geodesic distance from the haversine formula

In 1835, Sir James Inman introduced the haversine function in his book "Navigation and Nautical Astronomy for the Use of British Seamen" (Inman, 2012; Shylaja, 2015). This function was specifically designed for calculating distances on the surface of a sphere (assuming the Earth was such).

The distance d between two any points p_i and p_j along a great circle of a sphere is given by $d = r\theta$, where θ is the central angle between p_i and p_j on the sphere, and r is the radius of the sphere (the mean radius of Earth is $r \approx 6371$ km). The *haversine formula* allows the haversine of θ to be computed directly from the latitude (represented by ϕ) and longitude (represented by λ) of p_i and p_j:

$$hav\left(\frac{d}{r}\right) = hav(\Delta\phi) + \cos\phi_1 \cos\phi_2 hav(\Delta\lambda) \tag{2.4}$$

where:

- ϕ_1 and ϕ_2 are the latitudes of p_1 and p_2 in radians;
- λ_1 and λ_2 are the longitudes of p_1 and p_2 in radians;
- $\Delta\phi = \phi_2 - \phi_1$ is the difference in latitudes;
- $\Delta\lambda = \lambda_2 - \lambda_1$ is the difference in longitudes;
- $hav(\Theta) = \sin^2\left(\frac{\Theta}{2}\right)$ is the haversine function.

Substituting the haversine function into Equation (2.4) gives:

$$\sin^2\left(\frac{d}{2r}\right) = sin^2\left(\frac{\Delta\phi}{2}\right) + \cos\phi_1 \cos\phi_2 sin^2\left(\frac{\Delta\lambda}{2}\right)$$

To isolate $\sin^2\left(\frac{d}{2r}\right)$, take the square root of both sides:

$$\sin\left(\frac{d}{2r}\right) = \sqrt{sin^2\left(\frac{\Delta\phi}{2}\right) + \cos\phi_1 \cos\phi_2 sin^2\left(\frac{\Delta\lambda}{2}\right)}$$

To eliminate sin from the left-hand side, apply the function arcsin (the inverse of sin) to both sides:

$$\frac{d}{2r} = \arcsin\sqrt{sin^2\left(\frac{\Delta\phi}{2}\right) + \cos\phi_1 \cos\phi_2 sin^2\left(\frac{\Delta\lambda}{2}\right)}$$

Finally, multiply both sides of the equation by $2r$ to solve for the distance d:

$$d = 2r \arcsin \sqrt{sin^2\left(\frac{\Delta\phi}{2}\right) + \cos\phi_1 \cos\phi_2 sin^2\left(\frac{\Delta\lambda}{2}\right)} \tag{2.5}$$

Bibliography

Ahas, Rein, Silm, Siiri, Jarv, Olle, Saluveer, Erki, and Tiru, Margus. 2010. Using Mobile Positioning Data to Model Locations Meaningful to Users of Mobile Phones. *Journal of Urban Technology*, **17**(1), 3–27.

Amigo, Daniel, Pedroche, David Sánchez, García, Jesús, and Molina, José Manuel. 2021. Review and classification of trajectory summarisation algorithms: From compression to segmentation. *International Journal of Distributed Sensor Networks*, **17**(10), 15501477211050729.

Bonavita, Agnese, Guidotti, Riccardo, and Nanni, Mirco. 2021. Individual and collective stop-based adaptive trajectory segmentation. *GeoInformatica*, **26**, 451 – 477.

Chang, Kang-Tsung. 2018. *Introduction to Geographic Information Systems*. 9 edn. Columbus, OH: McGraw-Hill Education.

Deville, Pierre, Linard, Catherine, Martin, Samuel, Gilbert, Marius, Stevens, Forrest R., Gaughan, Andrea E., Blondel, Vincent D., and Tatem, Andrew J. 2014. Dynamic population mapping using mobile phone data. *Proceedings of the National Academy of Sciences*, **111**(45), 15888–15893.

Dijkstra, E. W. 2022. *A Note on Two Problems in Connexion with Graphs*. 1 edn. New York, NY, USA: Association for Computing Machinery. Page 287–290.

Douglas, David H, and Peucker, Thomas K. 1973. Algorithms for the reduction of the number of points required to represent a digitized line or its caricature. *Cartographica: the international journal for geographic information and geovisualization*, **10**(2), 112–122.

Driemel, Anne, Har-Peled, Sariel, and Wenk, Carola. 2010. Approximating the Fréchet distance for realistic curves in near linear time. Page 365–374 of: *Proceedings of the Twenty-Sixth Annual Symposium on Computational Geometry*. SoCG '10. New York, NY, USA: Association for Computing Machinery.

Earle, Michael A. 2006. Sphere to Spheroid Comparisons. *Journal of Navigation*, **59**, 491–496.

Furletti, Barbara, Cintia, Paolo, Renso, Chiara, and Spinsanti, Laura. 2013. Inferring human activities from GPS tracks. In: *Proceedings of the 2nd ACM SIGKDD International Workshop on Urban Computing*. UrbComp '13. New York, NY, USA: Association for Computing Machinery.

Garcez Duarte, Mariana M, and Sakr, Mahmoud. 2024. An experimental study of

existing tools for outlier detection and cleaning in trajectories. *Geoinformatica*, **29**(1), 31–51.

Gehlke, C. E., and Biehl, K. 1934. Certain Effects of Grouping upon the Size of the Correlation Coefficient in Census Tract Material. *Journal of the American Statistical Association*, **29**(185A), 169–170.

Grünbaum, Branko, and Shephard, G. C. 1986. *Tilings and Patterns*. New York: W. H. Freeman.

Inman, James. 2012. *Navigation and Nautical Astronomy, for the Use of British Seamen*. Rarebooksclub.com.

Iovan, Corina, Olteanu-Raimond, Ana-Maria, Couronné, Thomas, and Smoreda, Zbigniew. 2013. *Moving and Calling: Mobile Phone Data Quality Measurements and Spatiotemporal Uncertainty in Human Mobility Studies*. Cham: Springer International Publishing. Pages 247–265.

Karney, Charles F. F. 2013. Algorithms for Geodesics. *Journal of Geodesy*, **87**, 43–55.

Laube, Patrick, and Purves, Ross S. 2011. How fast is a cow? Cross-Scale Analysis of Movement Data. *Transactions in GIS*, **15**(3), 401–418.

Mannila, Heikki. 2002. Local and Global Methods in Data Mining: Basic Techniques and Open Problems. Pages 57–68 of: Widmayer, Peter, Eidenbenz, Stephan, Triguero, Francisco, Morales, Rafael, Conejo, Ricardo, and Hennessy, Matthew (eds), *Automata, Languages and Programming*. Berlin, Heidelberg: Springer Berlin Heidelberg.

Milner, Greg. 2016. *Pinpoint: How GPS Is Changing Technology, Culture, and Our Minds*. New York: W. W. Norton & Company.

Newson, Paul, and Krumm, John. 2009. Hidden Markov map matching through noise and sparseness. Page 336–343 of: *Proceedings of the 17th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*. GIS '09. New York, NY, USA: Association for Computing Machinery.

Openshaw, S. 1984. *The Modifiable Areal Unit Problem*. Norwich, UK: Geo Books.

Perrault, Charles. 1697. *Histoires ou contes du temps passé. Avec des moralités*. Paris: Claude Barbin. Includes the tale *Le Petit Poucet* (Tom Thumb).

Rinzivillo, Salvatore, Gabrielli, Lorenzo, Nanni, Mirco, Pappalardo, Luca, Pedreschi, Dino, and Giannotti, Fosca. 2014. The purpose of motion: Learning activities from Individual Mobility Networks. Pages 312–318 of: *2014 International Conference on Data Science and Advanced Analytics (DSAA)*.

Shylaja, B. S. 2015. From Navigation to Star Hopping: Forgotten Formulae. *Resonance*, **20**, 352–359.

Sila-Nowicka, Katarzyna, Vandrol, Jan, Oshan, Taylor, Long, Jed A., Demšar, Urška, and Fotheringham, A. Stewart. 2016. Analysis of human mobility patterns from GPS trajectories and contextual information. *International Journal of Geographical Information Science*, **30**(5), 881–906.

Sobel, Dava. 1996. *Longitude: The True Story of a Lone Genius Who Solved the Greatest Scientific Problem of His Time*. New York, NY: Penguin Putnam.

Swift, Jonathan. 2025. *Gulliver's Travels*. Cby Press. Modern edition reprint of the original 1726 publication.

Tan, P.N., Steinbach, M., and Kumar, V. 2016. *Introduction to Data Mining*. Pearson India.

Zhu, Lei, Holden, Jacob R., and Gonder, Jeffrey D. 2017. Trajectory Segmentation Map-Matching Approach for Large-Scale, High-Resolution GPS Data. *Transportation Research Record*, **2645**(1), 67–75.