Cell-Capacity Methodology

Taking advice from the international literature review, we are able to recommend a series of criteria for the minimum density of weather stations required in order to build a global network of weather stations that will be capable of representing weather conditions across the globe in high accuracy.

WMO recommends a minimum density of ~1 station per 600km^2 in flat areas. Such an area is better represented by the cell size 5 hexagon (where a hexagon covers an area of 253km^2) rather than the other H3 sizes provided. Considering that 2,016,830 size 5 hexagons cover the entire planet and that 71% of the Earth is covered by oceans, **the minimum number of required weather stations should be ~584,881 if all areas across the Earth were flat**. However, this number is expected to be much higher (described in the following lines), as a cell may need more than one weather station depending on its topography and land use. This practically means that in a significant number of cells the density may be larger than 1 station per 10km.

After conducting deep scientific literature research, we summarise and present our methodology for the optimal positioning of weather stations in the following section. Instead of providing only the exact location of stations, we provide flexibility by defining zones which could host a station so that cases where inaccessible areas can be avoided. Our methodology is devised in two parts, namely the topographic analysis and the land use analysis.

It is also noted that as a general principle and in order to reduce the number of stations within a small distance between adjacent cells, we define an internal buffer within each cell of 0.25km where stations cannot be placed.

a. Topographic analysis

Topographic analysis refers to the introduction of weather stations responsible for capturing variation due to landscape and orography. The process begins with evaluating whether the maximum elevation difference within the cell exceeds the 100m threshold (**Figure 5**). Identifying cases where this threshold is exceeded is useful for capturing harmful frosts which may occur at the lower ground areas of a valley or for describing significant changes of various meteorological parameters (i.e., air flow) due to a complex topography. If this 100m-threshold is not exceeded then the cell is considered flat and only one weather station is required. In the opposite case, we proceed with the creation of two zones of points, one with all points above (*high-zone*) and another with all points below (*low-zone*) the critical altitude (z_{crit}). z_{crit} is defined as the maximum altitude in the cell (z_{max}) minus 100m ($z_{crit} = z_{max} - 100$). We locate one station in high-zone, ideally at the location with altitude z_{max}. For points in the low-zone, we proceed with the identification of continuous sub zones with the same aspect which we refer to as *aspect-zones* (**Figure 6**). If multiple aspect-zones with the same aspect (e.g. two aspect-zones with North aspect) are identified, then we keep the largest one. Finally, we place one station in each aspect-zone, ideally at the lowest point of each aspect zone.

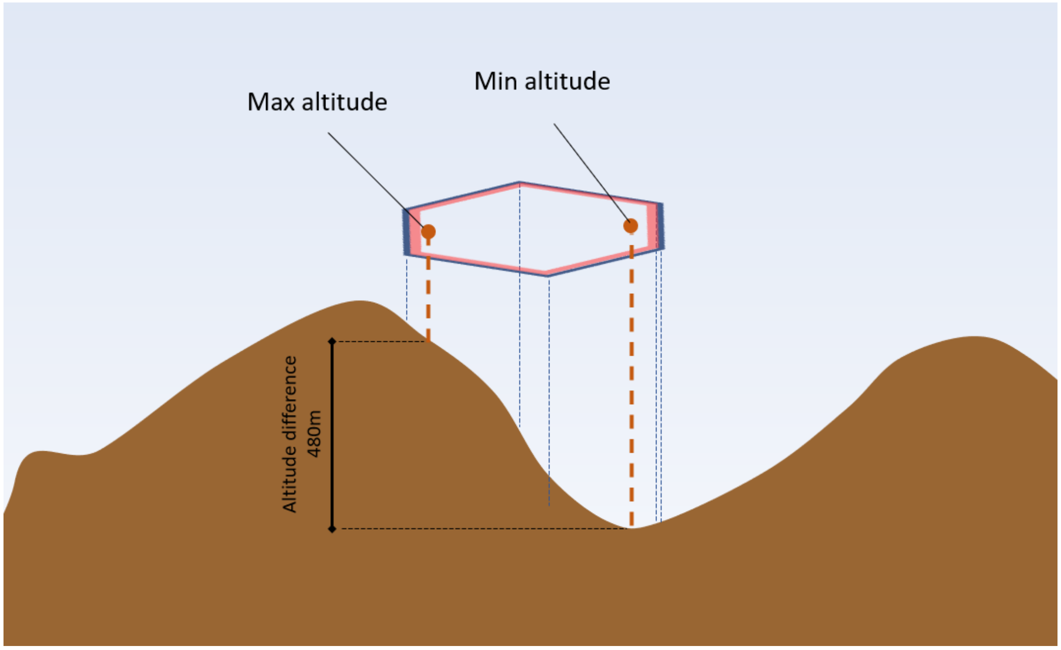

Figure 5. Positions of weather stations (orange bullets) over a complex topography. Red zones indicate the areas that should be avoided (only for determining the ideal locations of stations). For simplicity reasons points instead of zones are used in the scheme.

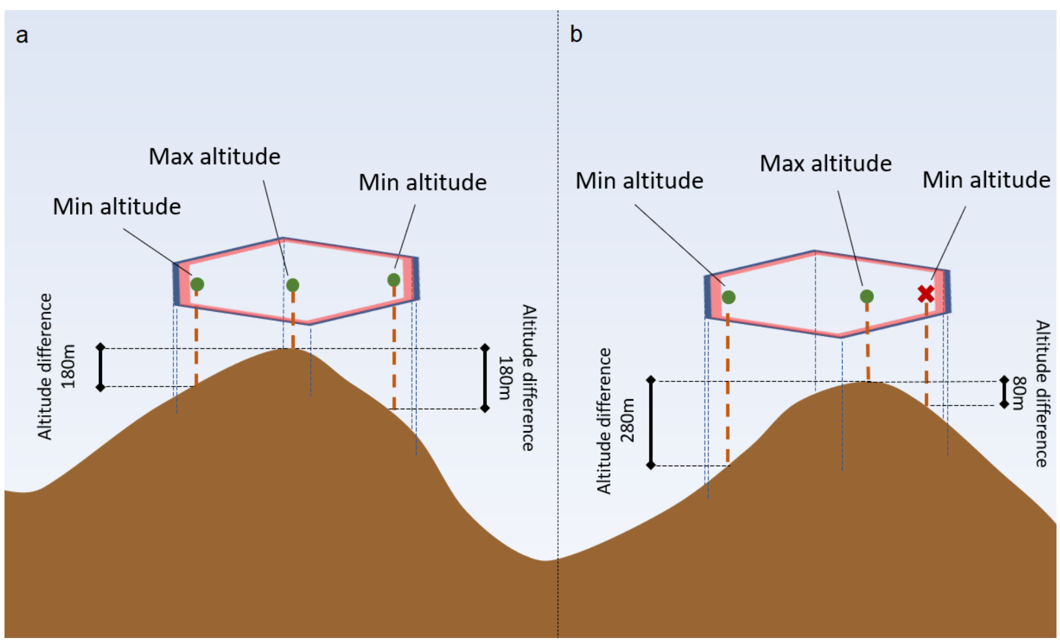

Figure 6. Positions of weather stations (green bullets) over a complex topography with differently oriented slopes. Red zones indicate the areas that should be avoided (only for determining the ideal locations of stations). The “x” symbol indicates an example of a non-recommended location. For simplicity reasons points instead of zones are used in the scheme.

As an additional example, in an extreme case that a mountain/hill with a “pyramidal” shape is included in a cell (assuming that its base points are at least 100m lower than the top), then 5 weather stations should be placed, the 4 of them at the lowest possible location of each differently oriented surface and one at the highest point (**Figure 7**). Similarly, an area with four differently aspected surfaces and a lowest point in the middle should also be represented by 5 weather stations.

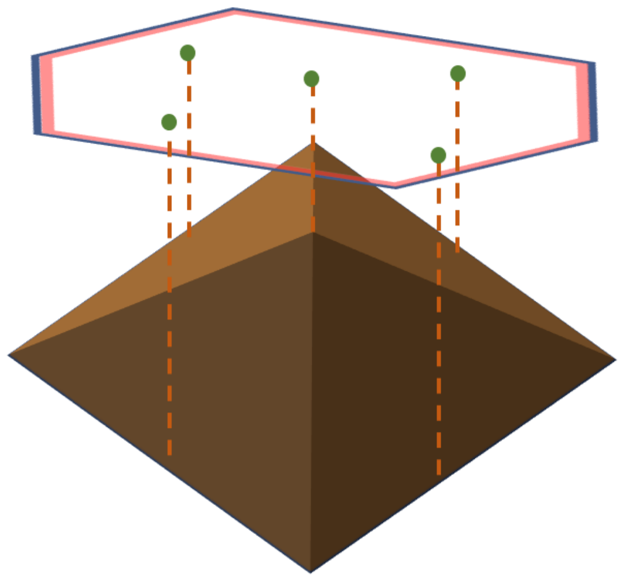

Figure 7. An extreme case of a “pyramidal” shaped mountain and the recommended locations of the stations. For simplicity reasons points instead of zones are used in the scheme.

b. Urban Land use analysis

Urban land use analysis is responsible for identifying the location of stations which capture variations of weather conditions due to the effect of urban areas. To simplify the analysis we consider three main land use categories as the most influential ones, namely urban (buildings, houses, offices, etc.), green (e.g. parks) and coastline. The procedure begins with the identification of urban and green areas, which are buffered by 100m, and then dissolved in order to combine nearby elements with the same land use type. Then, the process filters out urban zones with area less than $0.4km^2$ and green areas smaller than $0.015km^2$. In order to further reduce the number of required stations, for each cell we keep only the largest urban and green zone (within the largest urban zone), assuming that the effect of the land use is adequately captured by a single sensor per land use type. In addition, when the cell intersects with the coastline we place an additional station anywhere on a 50m buffered area along the coastline. For cases where multiple islands are present within the cell, the station is placed on the longer coastline. The previous section is visualised in **Figure 8**.

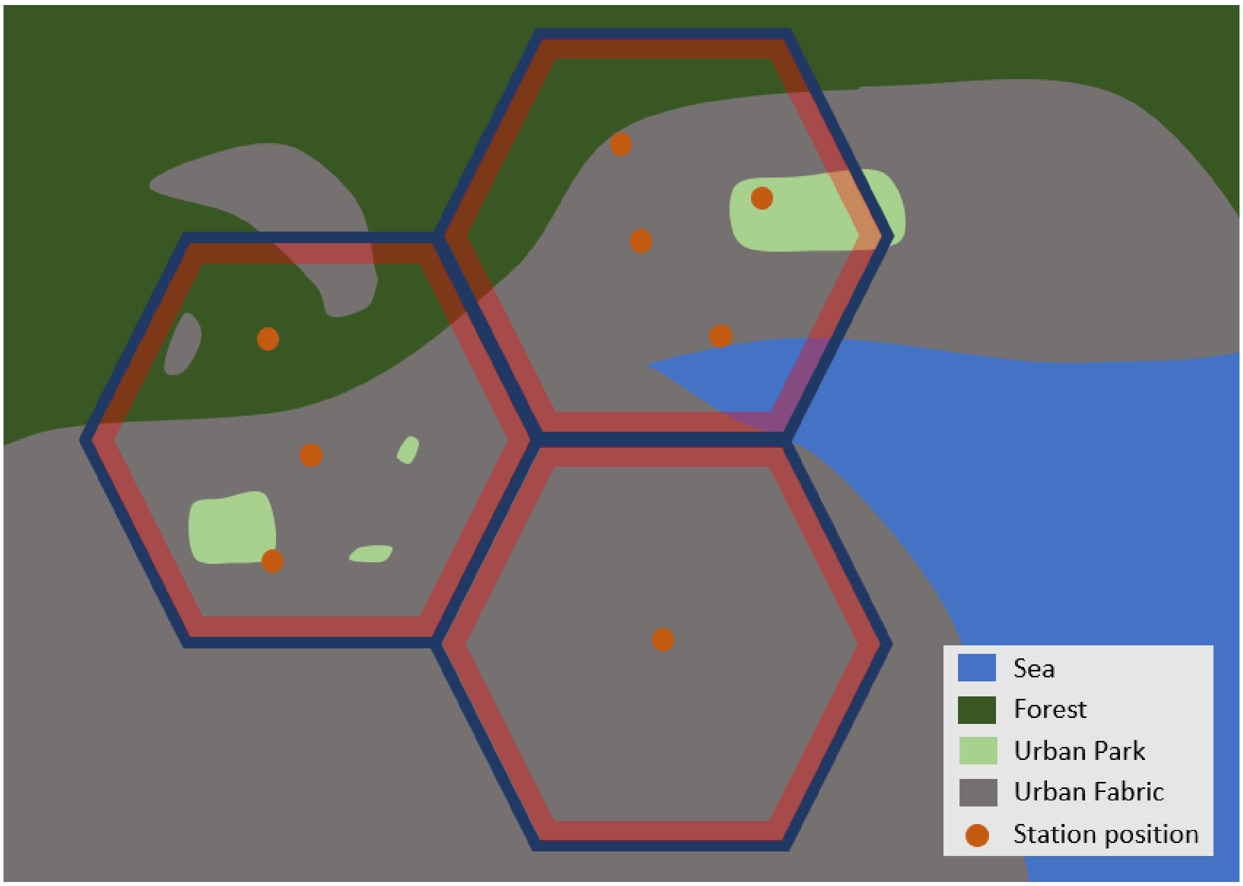

Figure 8. Minimum required stations (orange bullets) over an urban environment. Red zones indicate the areas that should be avoided (only for determining the ideal locations of stations). For simplicity reasons points instead of zones are used in the scheme.

Summary

Applying the above-described methodology, a cell may eventually require up to 8 weather stations (4 for differently oriented aspects, 1 for the highest area, 1 for the largest urban area, 1 for the largest urban green and 1 for the coastal area of a cell). Considering that convective phenomena can present an extent of **2-20km** (according to literature review) and WMO suggests a density of **one station per 1-10km for measuring convective activity**, the present methodology may achieve a density of 1 station per 5.6km, which agrees with the already recommended densities. Dividing the world into hexagon shaped cells (according to the UBER H3 model), we try to define the weather representativeness of weather stations within a cell and also to facilitate the WeatherXM rewarding mechanism. Finally, it is important to highlight that the present study defines the locations of well deployed weather stations. However, a number of additional weather stations around a well-deployed (reference) one should also exist, being able to confirm that the reference station works properly and preserve its condition.