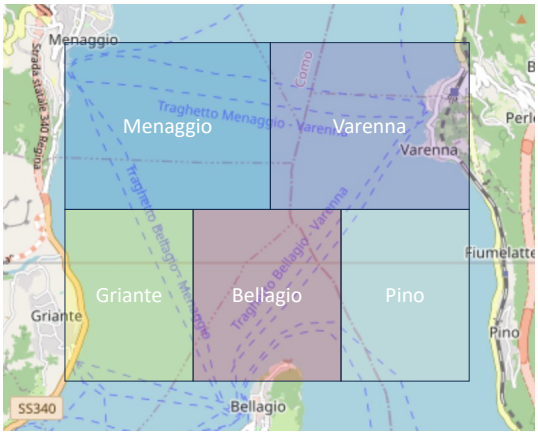
Politecnico di Milano
Scuola di Ingegneria Industriale e dell'Informazione
Applied Statistics
June 13th, 2023

Problem n.2

The dataframe `Chlorophyll.txt` contains data related to the concentration of chlorophyll (`chlorophyll` $\in \mathbb{R}$) in the centre of Como Lake. The sample consists of 210 locations (`x` and `y`) in UTM coordinates. Moreover, information relative to the depth of the Lake in each location is given (`depth` $\in \mathbb{R}$), together with the zone to which the locations belong (`zone` $\in \{$*Menaggio, Varenna, Griante, Bellagio, Pino*$\}$), i.e. the closest village to the location, as shown in Figure.



a) Consider for `chlorophyll` $z(s_k)$ the following model

$$z(s_k) = \beta_0 + \delta(s_k) \quad \text{for} \quad k = 1, ..., 210$$

where $\delta(s_k)$ a stationary residual with *Exponential* model *with* Nugget for spatial dependence structure.

- Estimate the sill and the range.
- Estimate the parameters β_0 and $\delta(s_k)$ of the model fitted via *Generalized Least Squares*.

b) Let's consider now the variable `zone` as a random intercept; take into account all the covariates except the locations `x` and `y` and fit a suitable model for accounting the hierarchy.

Report the parameterization of the model and compute the model unknowns.

c) Compute and report the PVRE index and interpret the obtained result.

d) Report the dot plot of the estimated random intercepts. Net to the effect of fixed effect covariates, which is the zone associated to the lowest concentration of `chlorophyll`?

Upload your solution here