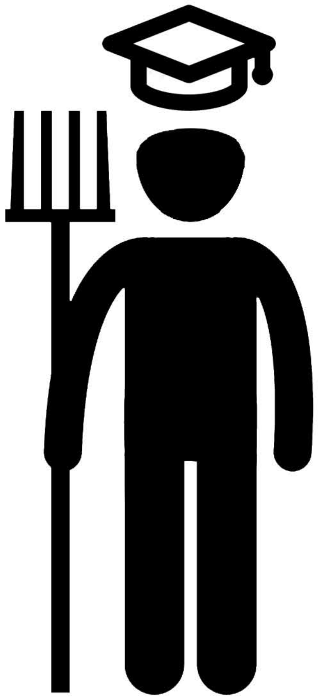
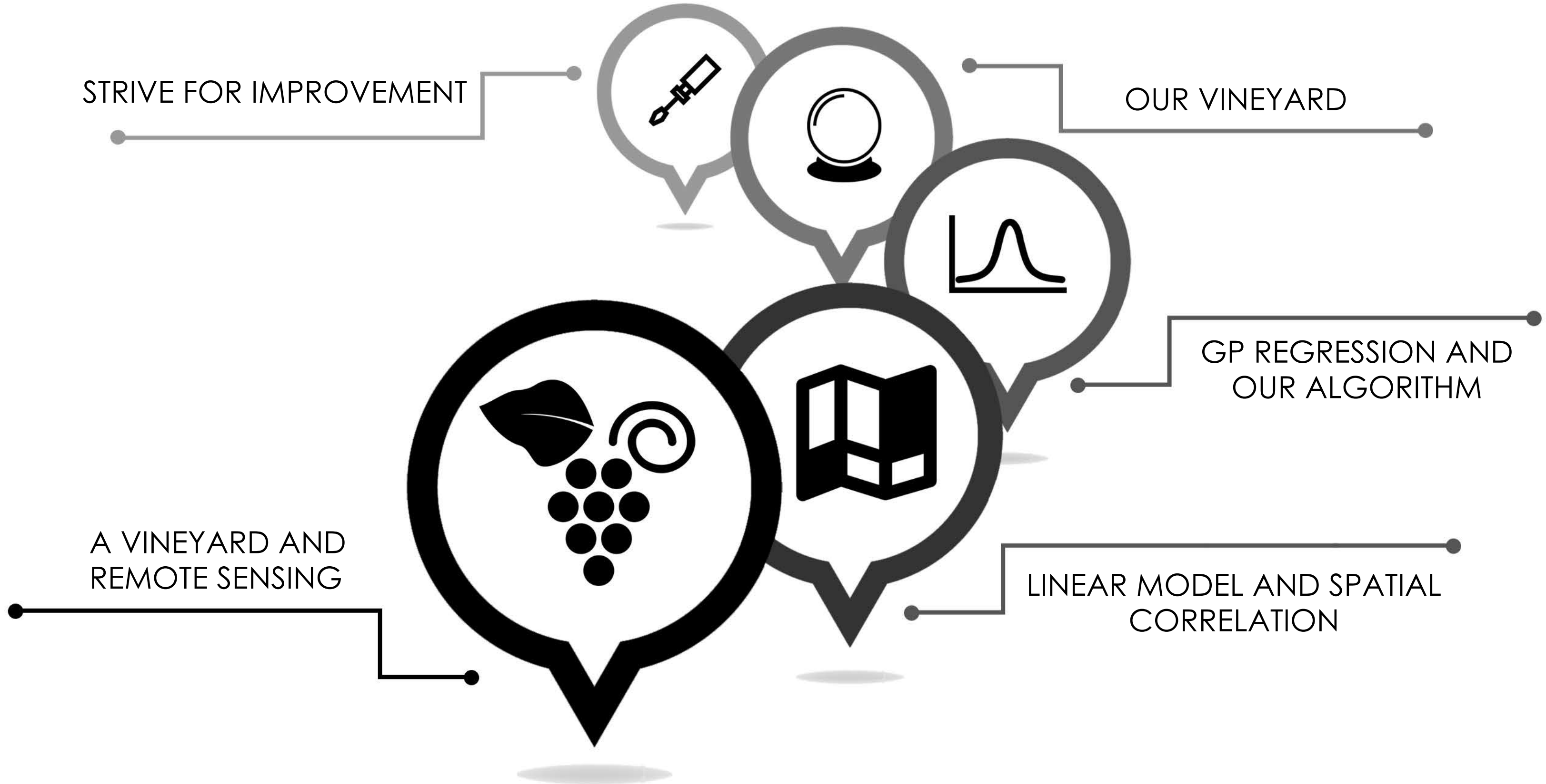
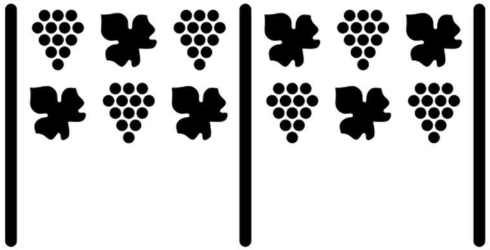
“ A spatial analytics approach for the ripeness estimation inside a vineyard ”



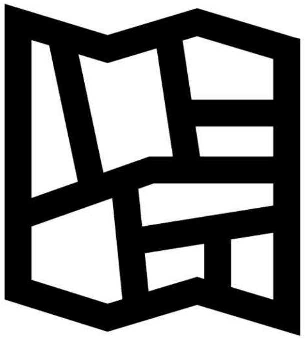
Student:
Karvouniaris Alexander
Academic Supervisor:
Michalis Titsias
Company Supervisor:
Damianos Oikonomidis

TIMELINE


STRIVE FOR IMPROVEMENT
OUR VINEYARD
GP REGRESSION AND OUR ALGORITHM
LINEAR MODEL AND SPATIAL CORRELATION
A VINEYARD AND REMOTE SENSING



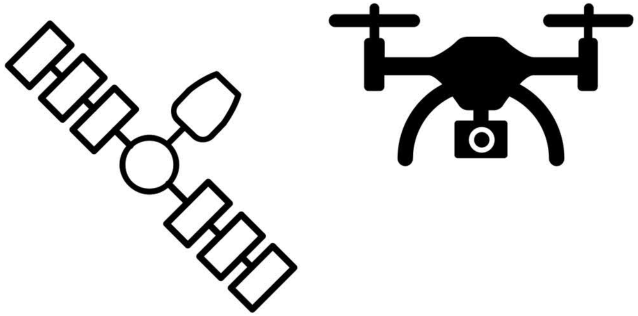
Variability



Selective harvesting

probable alcoholic degree/ total acidity

Map the ripeness proxy

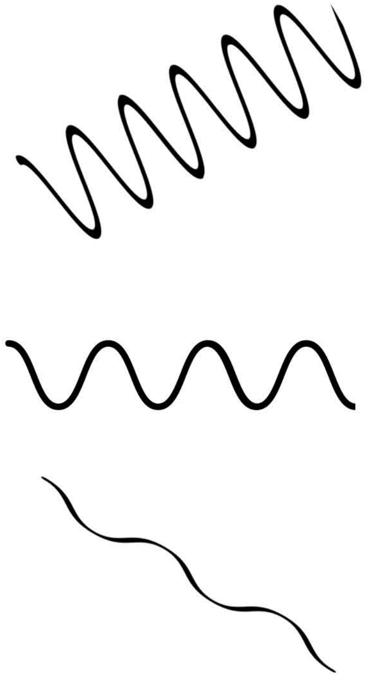


Remote Sensing

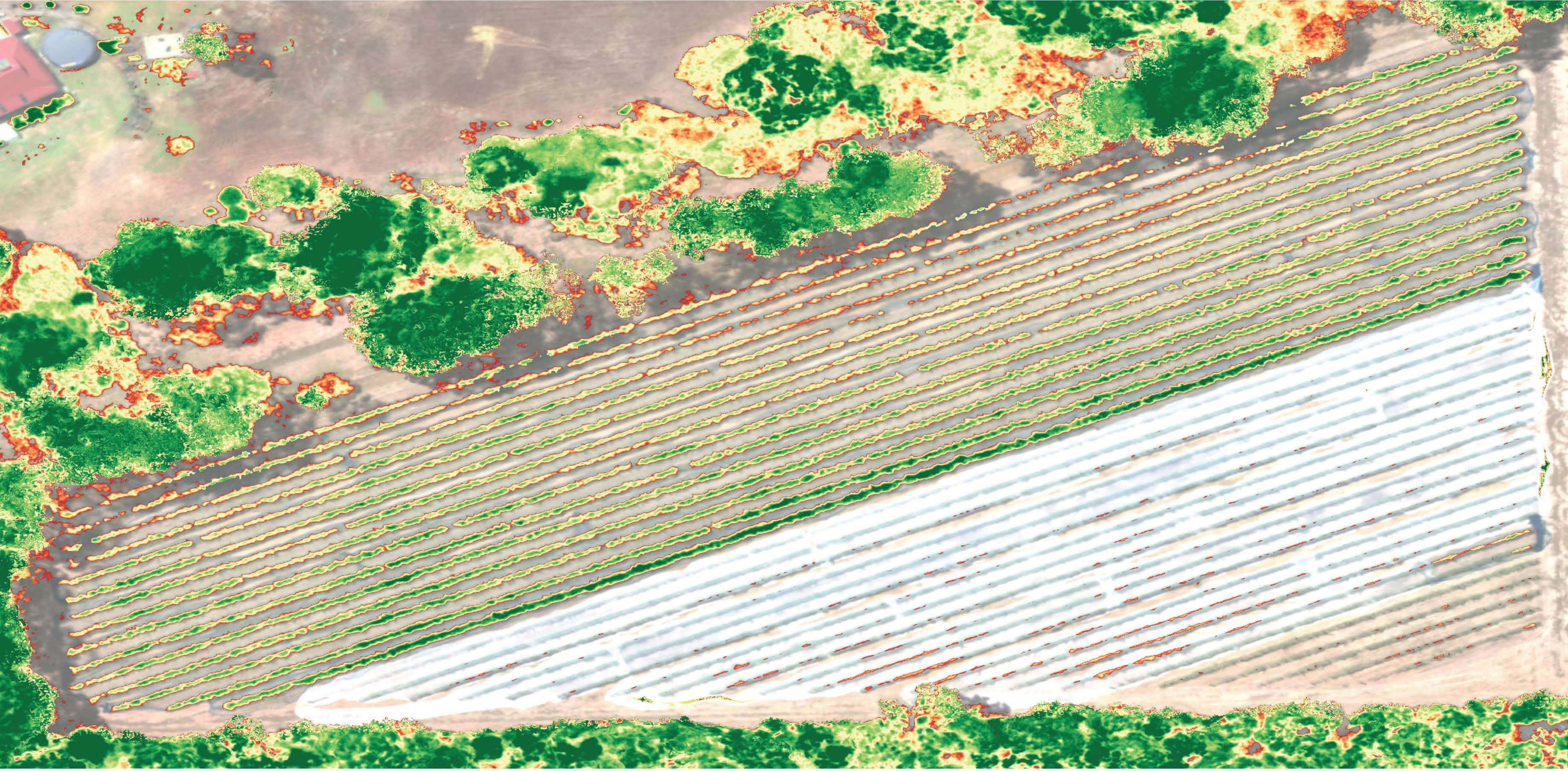
Na

Cl

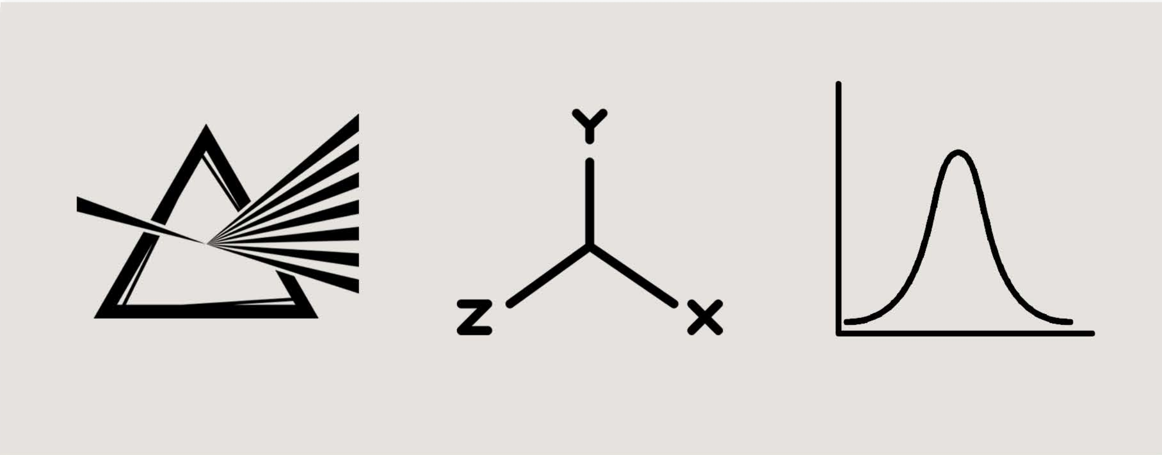
Zinc



Spectroscopy

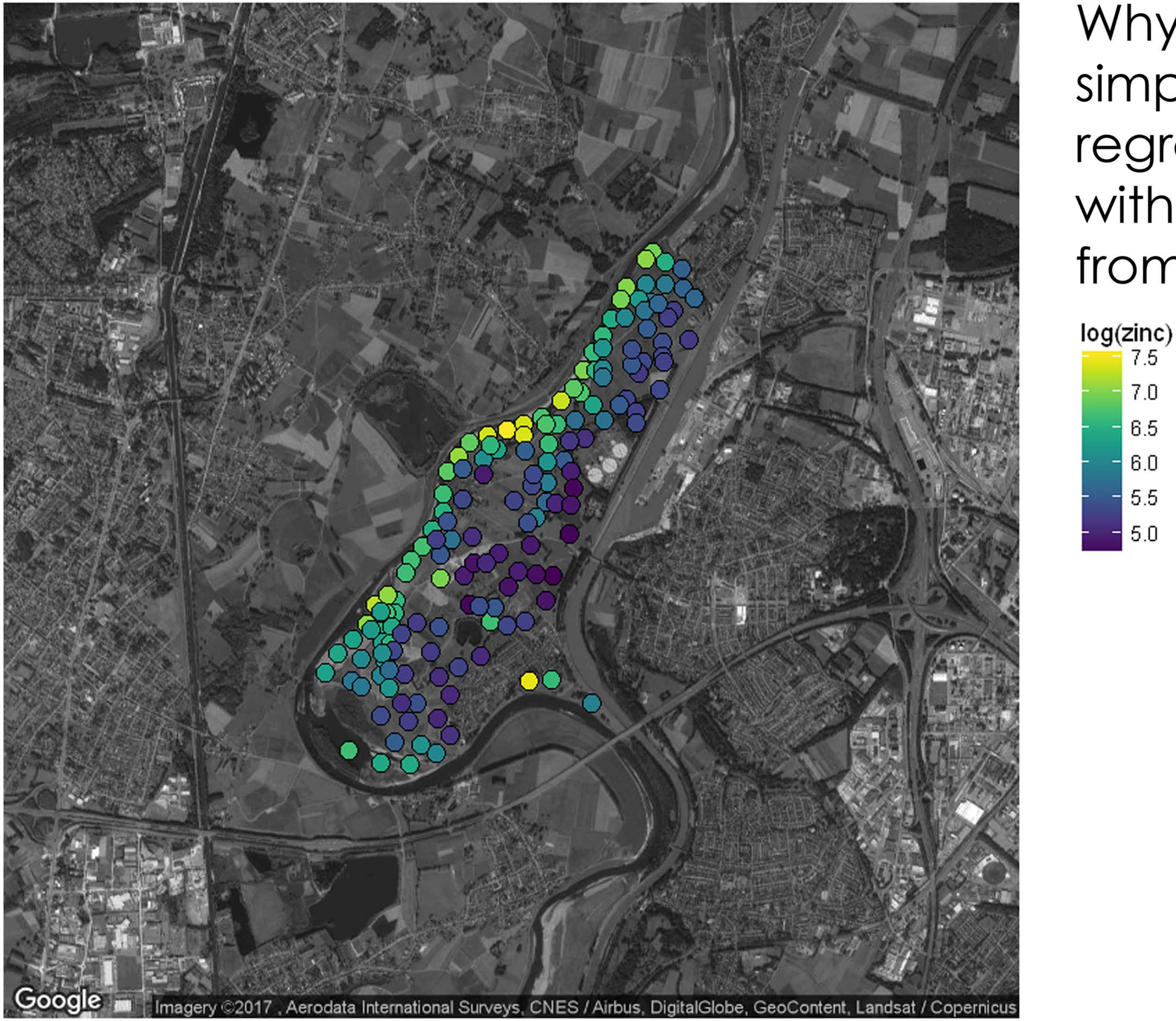


$$NDVI=(Nir - Red)/(Nir + Red)$$ Chlorophyl estimator

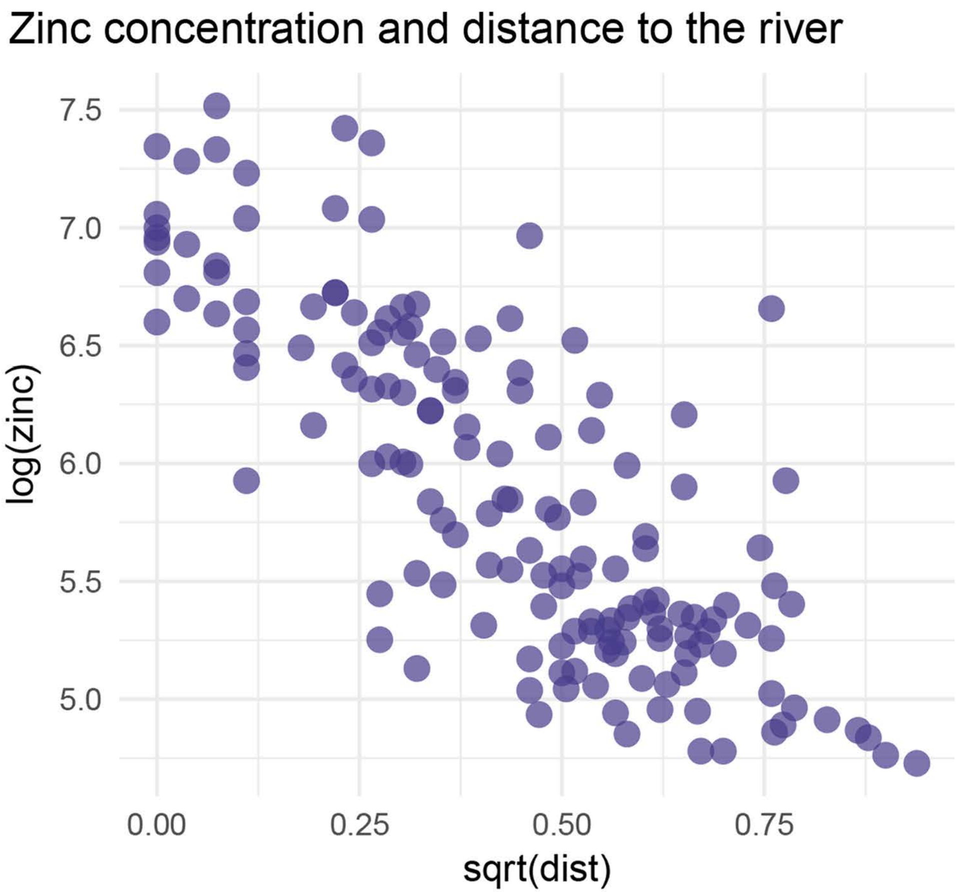


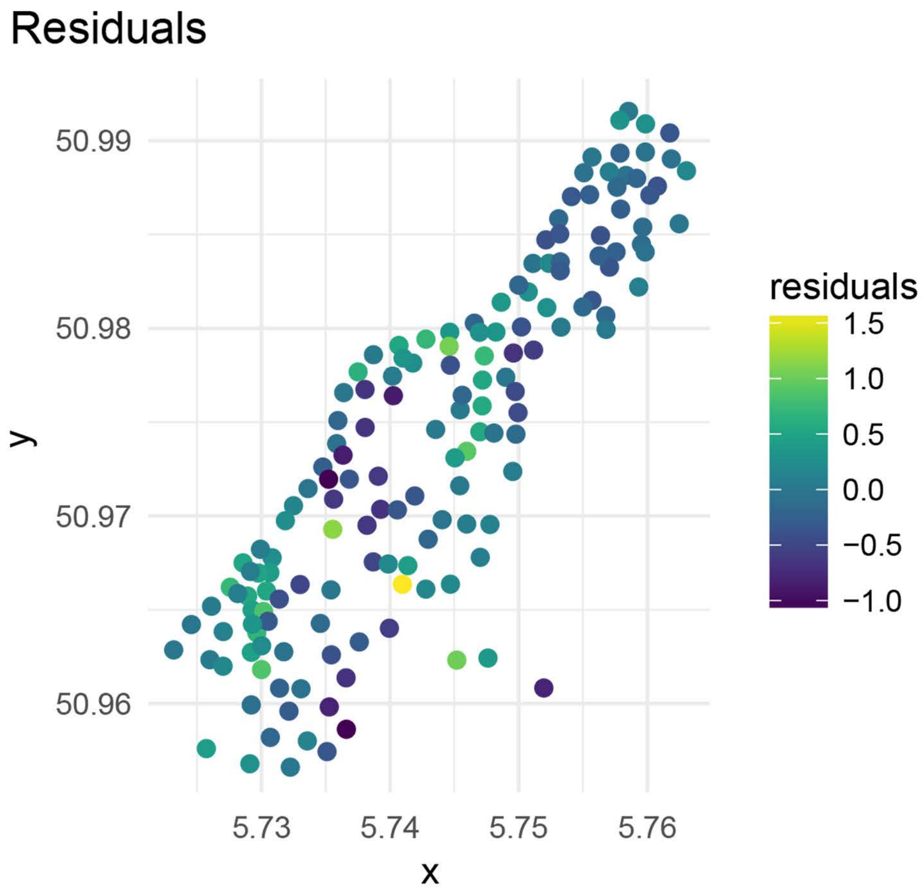
Spatial machine learning

Logged Zinc Concentration



Why not use simple linear regression with distance from river?



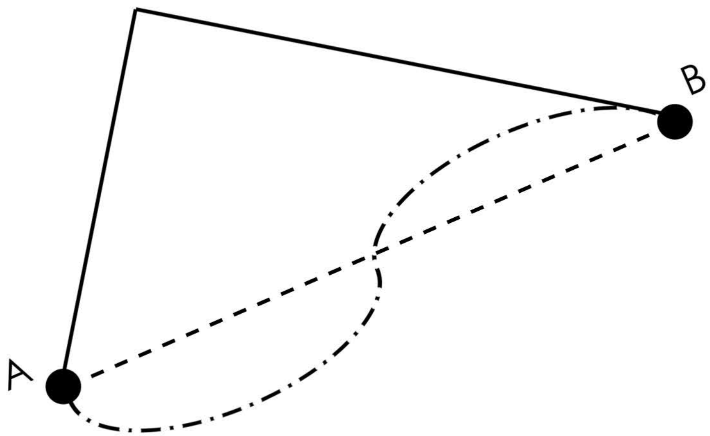


Linearity ✓

Homoscedasticity ✗

$$y_t = \mu + \phi_1 y_{t-1} + \varepsilon_t.$$

Temporal distance by 1



Spatial distance though can have many interpretations

A Gaussian process is a collection of random variables, any finite number of which have a joint Gaussian distribution

Definition

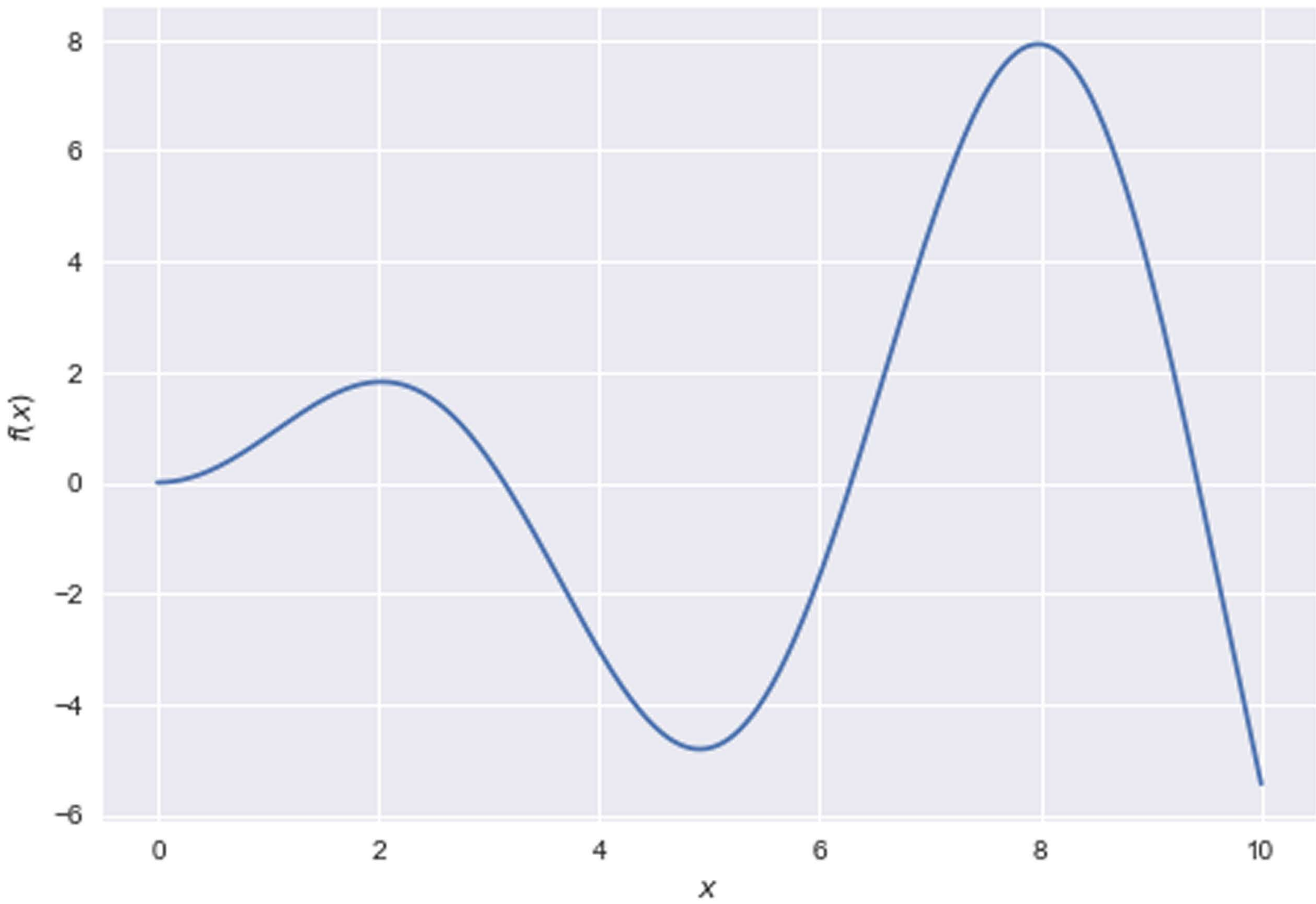
$$f(\boldsymbol{x}) \sim \mathcal{GP}(m(\boldsymbol{x}), k(\boldsymbol{x}, \boldsymbol{x}'))$$

Symbolism

$$\begin{aligned} m(\boldsymbol{x}) &= \mathbb{E}[f(\boldsymbol{x})], \\ k(\boldsymbol{x}, \boldsymbol{x}') &= \mathbb{E}[(f(\boldsymbol{x}) - m(\boldsymbol{x}))(f(\boldsymbol{x}') - m(\boldsymbol{x}'))] \end{aligned}$$

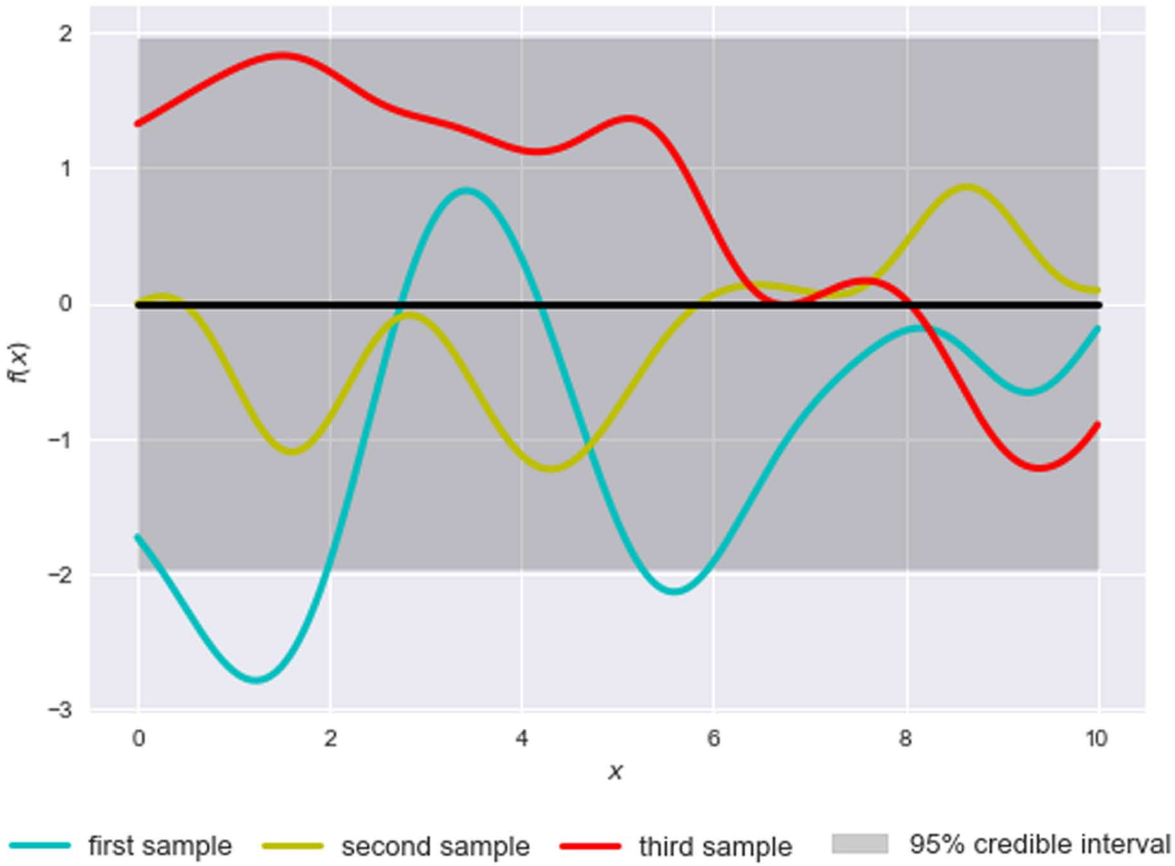
What we need to compute

Example on f(x)=xsin(x)



$$f_* \sim \mathcal{N}(\mathbf{0}, K(X_*, X_*))$$

prior predictive distribution



prior samples

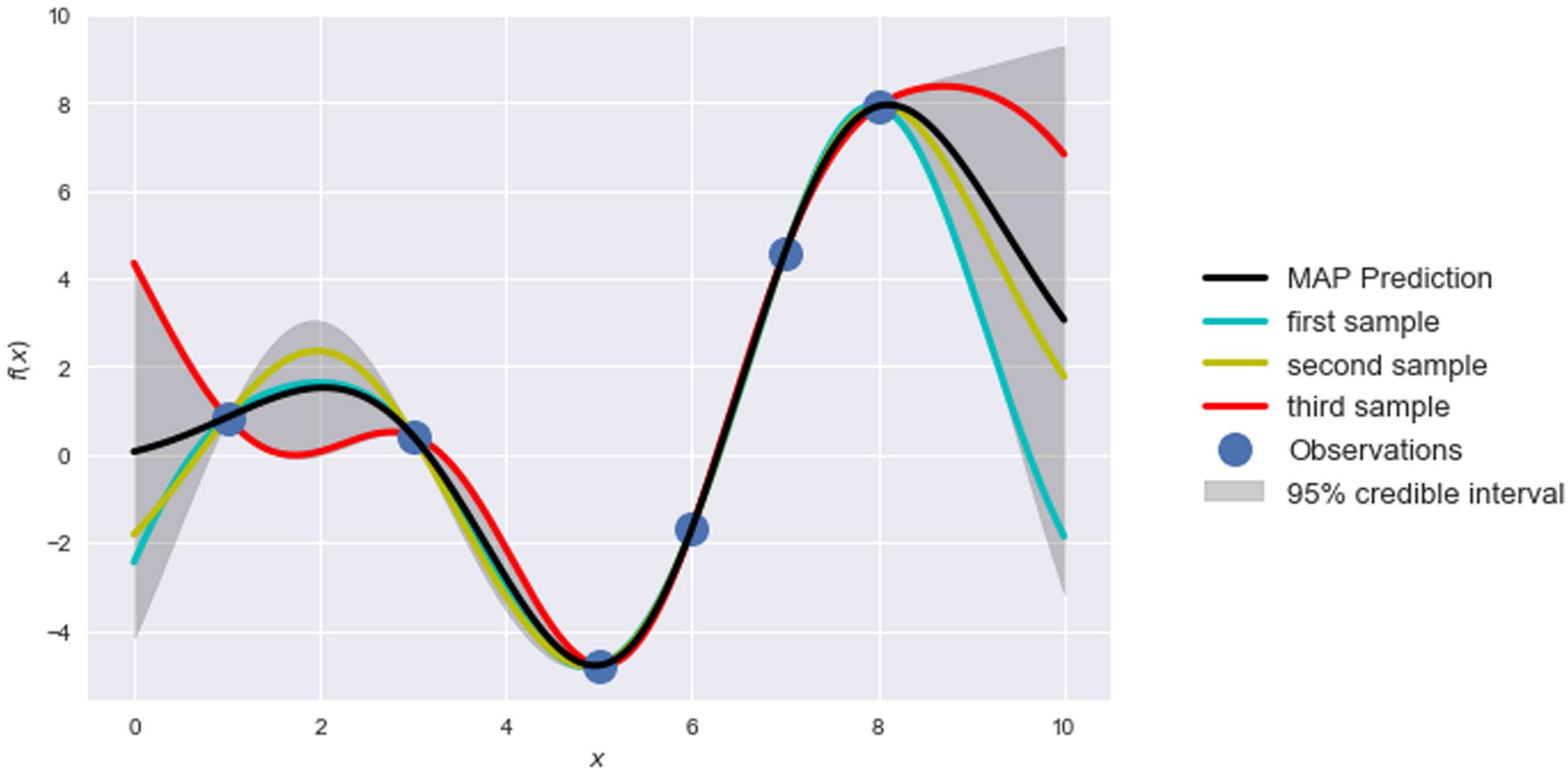
$$\begin{bmatrix} \boldsymbol{f} \\ \boldsymbol{f}_* \end{bmatrix} \sim \mathcal{N}\left(\mathbf{0}, \begin{bmatrix} K(X,X) & K(X,X_*) \\ K(X_*,X) & K(X_*,X_*) \end{bmatrix}\right)$$

Joint Gaussian

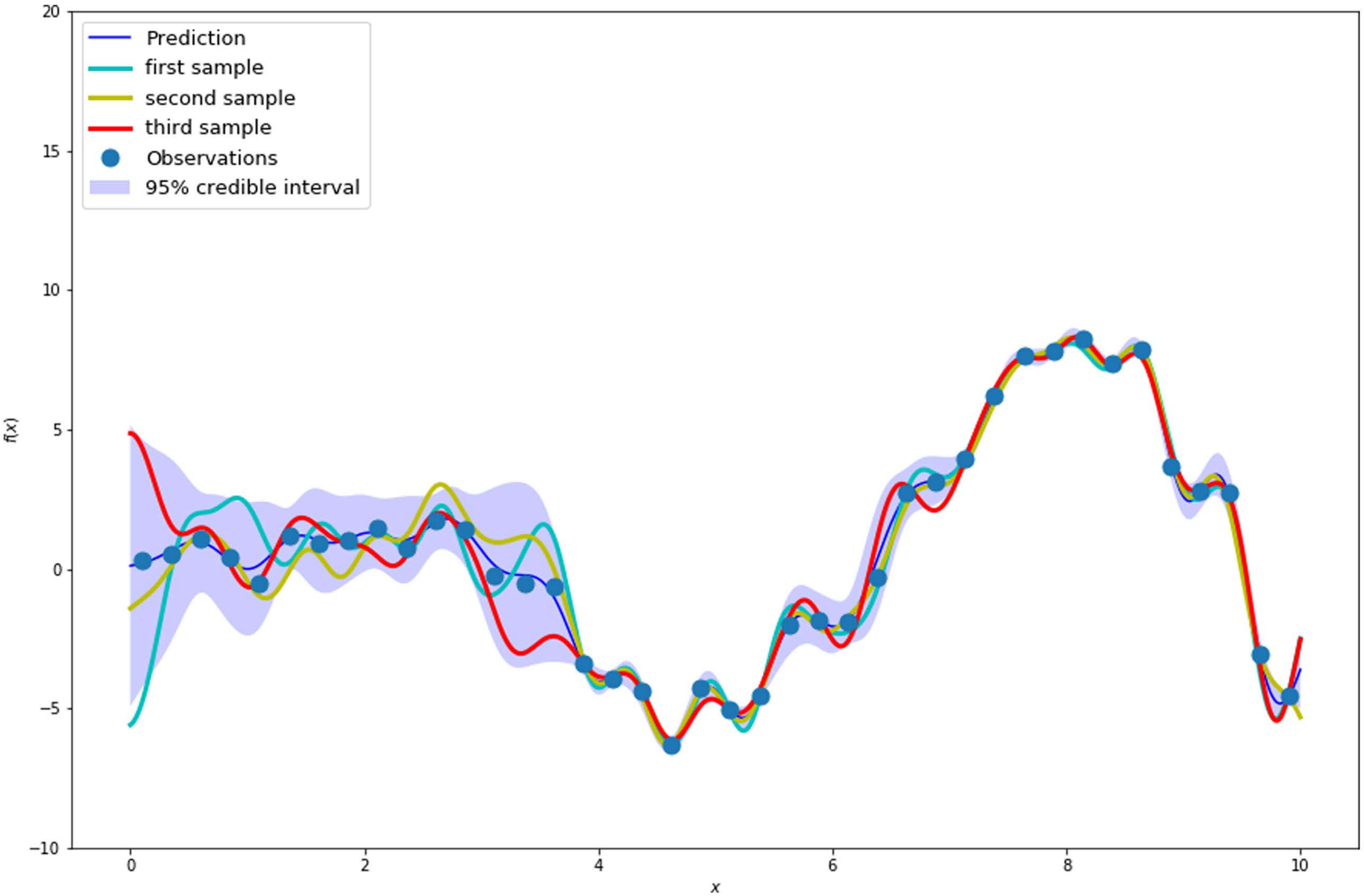
$$\boldsymbol{f}_* | X_*, X, \boldsymbol{f} \sim \mathcal{N}(K(X_*,X)K(X,X)^{-1}\boldsymbol{f}, \\ K(X_*,X_*) - K(X_*,X)K(X,X)^{-1}K(X,X_*))$$

After marginalization

Posterior predictive samples and mean



Posterior predictive samples and mean on noisy data



$$k_{SE}(r) = \exp\left(-\frac{r^2}{2l^2}\right)$$

Squared exponential

$$k_{Matern}(r) = \frac{2^{1-\nu}}{\Gamma(\nu)}\left(\frac{\sqrt{2\nu}r}{l}\right)^{\nu} K_{\nu}\left(\frac{\sqrt{2\nu}r}{l}\right)$$

Matern

$$k_{RQ}(r) = \left(1 + \frac{r^2}{2al^2}\right)^{-a}$$

Rational Quadratic

$$p(\mathcal{H}_i|\boldsymbol{y}, X) = \frac{p(\boldsymbol{y}|X, \mathcal{H}_i)p(\mathcal{H}_i)}{p(\boldsymbol{y}|X)}$$

$$p(\boldsymbol{\theta}|\boldsymbol{y}, X, \mathcal{H}_i) = \frac{p(\boldsymbol{y}|X, \boldsymbol{\theta}, \mathcal{H}_i)p(\boldsymbol{\theta}|\mathcal{H}_i)}{p(\boldsymbol{y}|X, \mathcal{H}_i)}$$

Optimization methods on hierarchical modelling

$$p(\boldsymbol{w}|\boldsymbol{y}, X, \boldsymbol{\theta}, \mathcal{H}_i) = \frac{p(\boldsymbol{y}|X, \boldsymbol{w}, \mathcal{H}_i)p(\boldsymbol{w}|\boldsymbol{\theta}, \mathcal{H}_i)}{p(\boldsymbol{y}|X, \boldsymbol{\theta}, \mathcal{H}_i)}$$

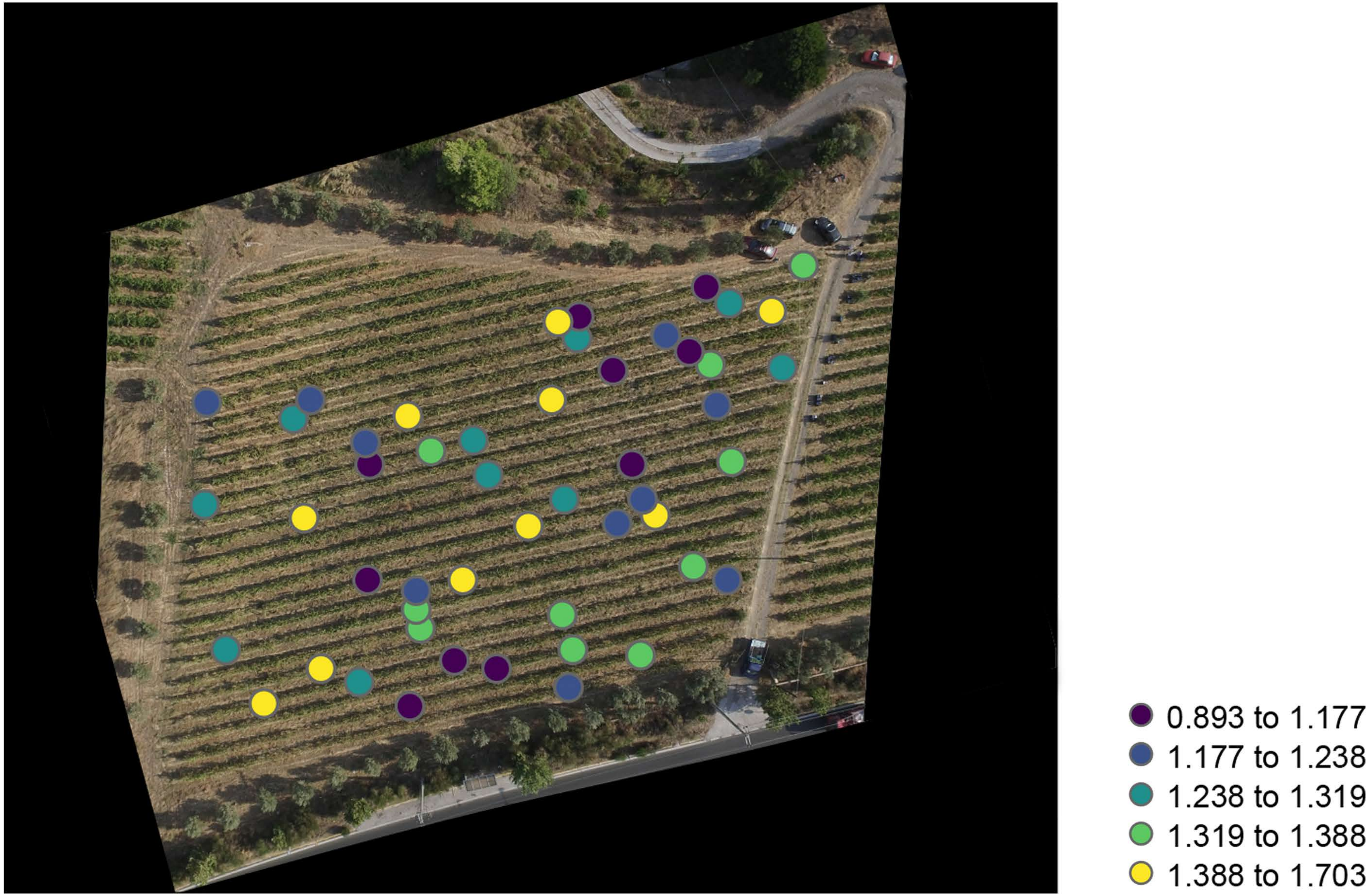
Our approach

K1, K2 valid kernels: K1+K2, K1*K2 valid kernels

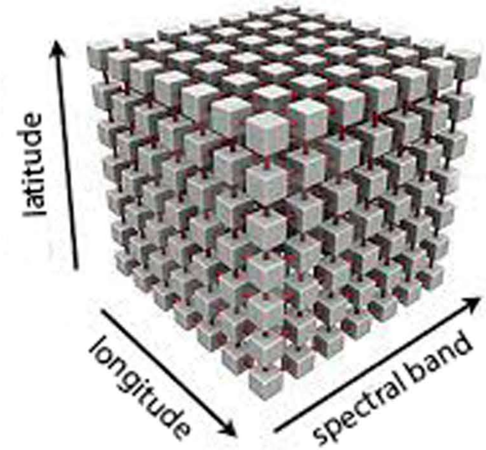
For each kernel {
    Cross-validation{
        “learn” optimal theta
        Count SMSE, NLPD}
    Count Mean SMSE, Mean NLPD

Lytras Winery, Pikermi
5000 km^2
1197 vines
50 observations

Observations

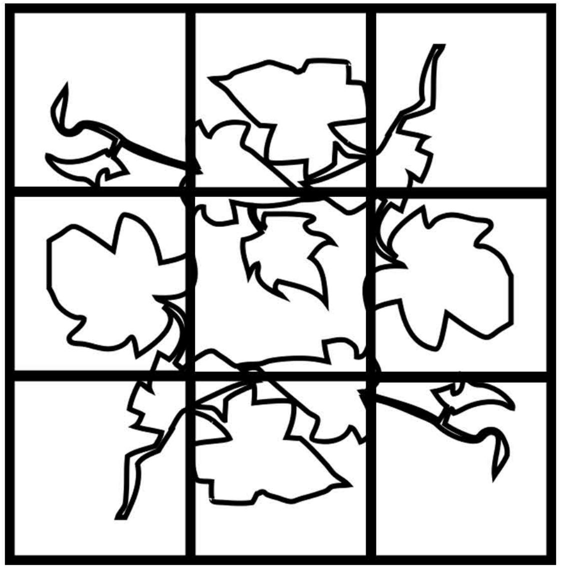


Visual inputs

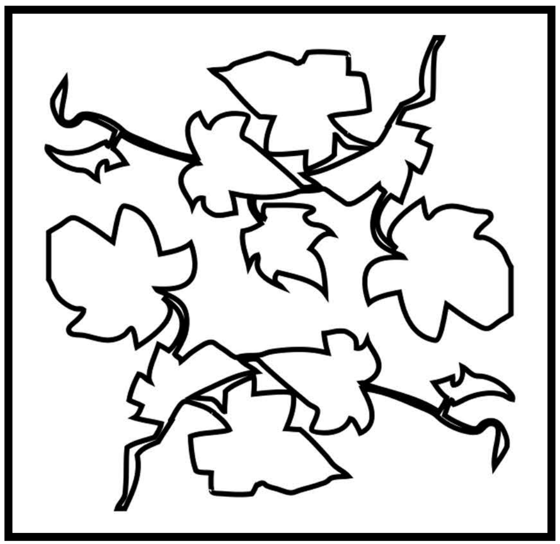


Data cube

Spatial Resolution



Too detailed



Use mean reflectance

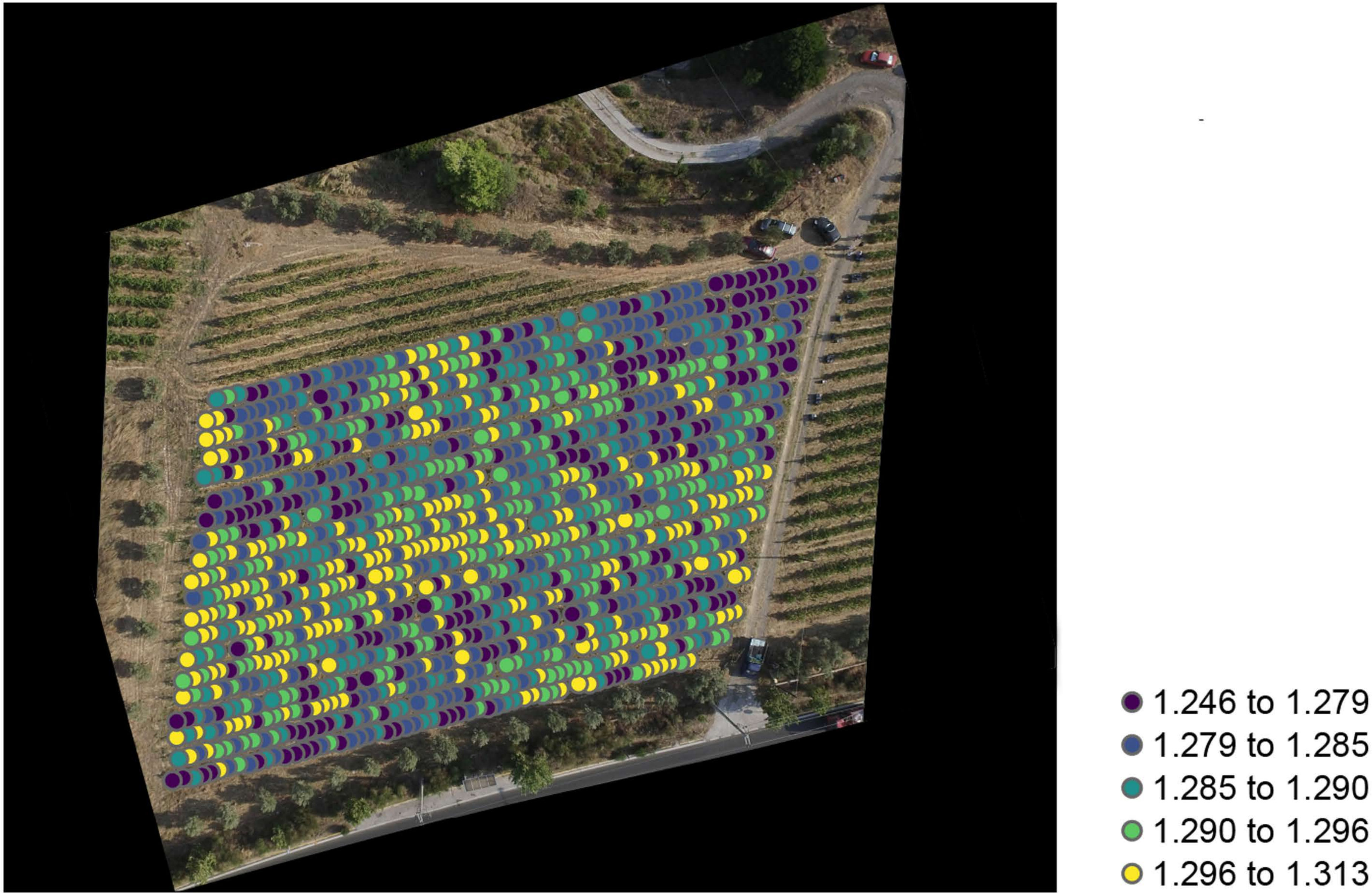
Our final dataset

red	green	blue	ir	ndvi	x	y	TA.OO
47.454519	48.289348	36.637008	106.88	0.385081	23.951851	38.005501	1.227723
75.508617	68.965478	44.009110	111.43	0.196959	23.951954	38.005538	1.343750
99.689756	93.139695	63.837746	104.54	0.033965	23.951624	38.005480	0.927419
66.554706	66.616330	46.729278	109.90	0.262506	23.951857	38.005544	1.347368
91.807513	87.032076	59.069802	113.33	0.114995	23.951748	38.005523	1.096154

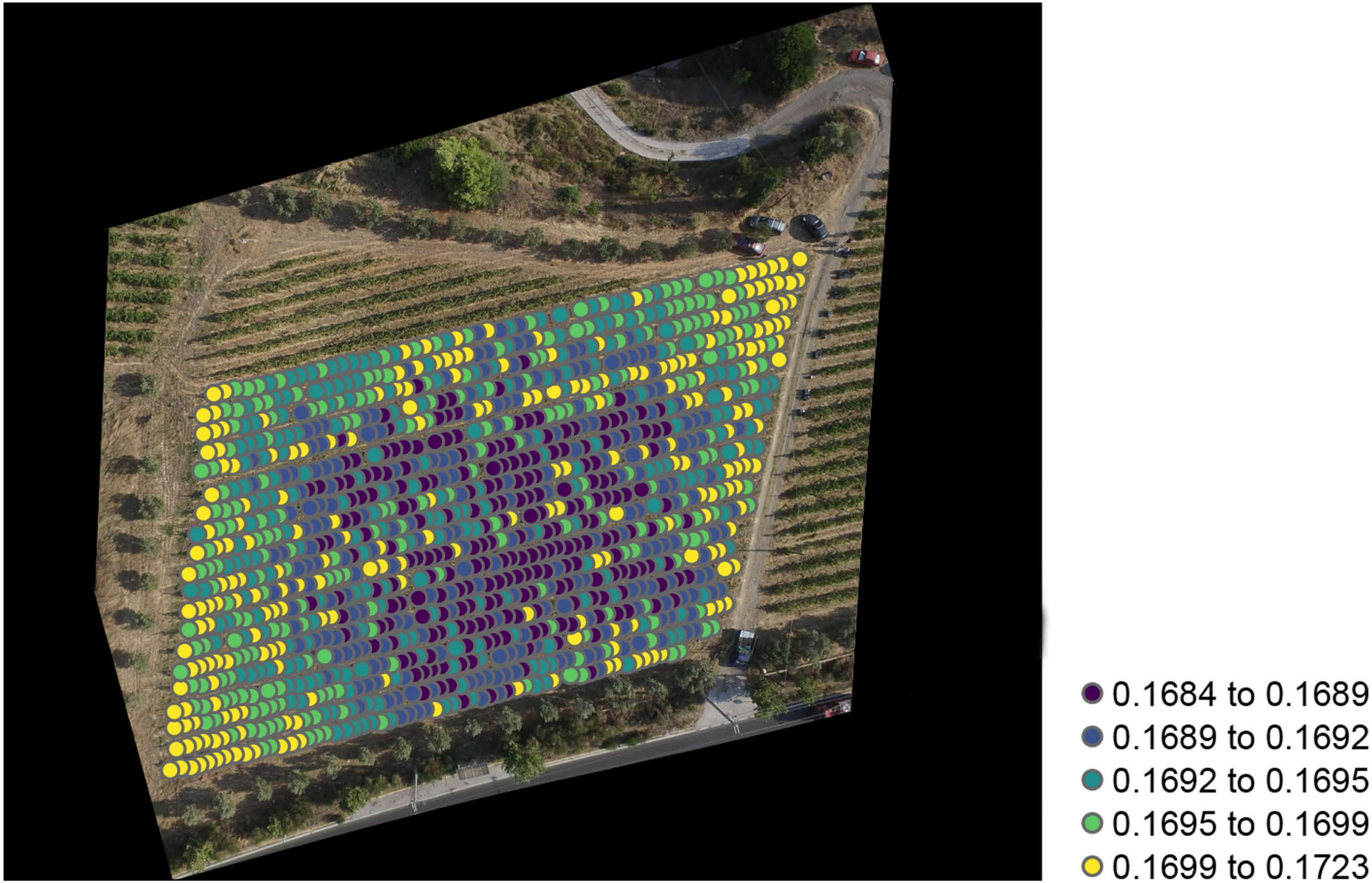
Benchmarking

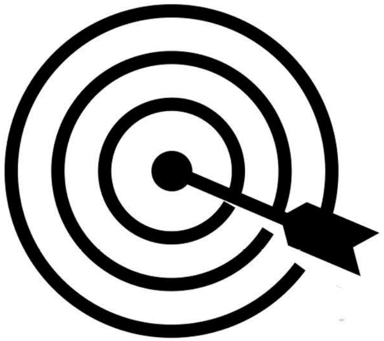
posterior covariance	SMSE	NLPD
1.29**2 * RBF(length_scale=31) + WhiteKernel(noise_level=0.0278)	1.083689	0.117819
1.29**2 * Matern(length_scale=52.8, nu=1.5) + WhiteKernel(noise_level=0.0278)	1.083825	0.117908
1.28**2 * RationalQuadratic(alpha=0.00146, length_scale=25.7) + WhiteKernel(noise_level=0.0277)	1.111897	0.120626
1.29**2 * ExpSineSquared(length_scale=7.79, periodicity=24.9) + WhiteKernel(noise_level=0.0278)	1.072662	0.167999
1.29**2 * RBF(length_scale=31) * RBF(length_scale=1.07e+03) + WhiteKernel(noise_level=0.0278)	1.083686	0.117819
1.29**2 * RBF(length_scale=1.03e+04) * Matern(length_scale=52.8, nu=1.5) + WhiteKernel(noise_level=0.0278)	1.084541	0.117923
1.29**2 * Matern(length_scale=52.8, nu=1.5) * Matern(length_scale=7.78e+04, nu=1.5) + WhiteKernel(noise_level=0.0278)	1.084027	0.117912
1.29**2 * RBF(length_scale=31) + 0.00473**2 * RBF(length_scale=118) + WhiteKernel(noise_level=0.0278)	1.125954	0.122336
1.15**2 * ExpSineSquared(length_scale=0.228, periodicity=7.18e+03) + 0.574**2 * RBF(length_scale=13.9) + WhiteKernel(noise_level=0.0278)	1.060862	0.117571
0.684**2 * ExpSineSquared(length_scale=21.1, periodicity=0.000491) + 3.1**2 * RationalQuadratic(alpha=1.43e-05, length_scale=0.0424) + WhiteKernel(noise_level=0.0248)	1.088955	0.117613
0.0416**2 * ExpSineSquared(length_scale=0.0729, periodicity=17.6) + 1.29**2 * Matern(length_scale=61.9, nu=1.5) + WhiteKernel(noise_level=0.0263)	1.125467	0.120851

Predictions



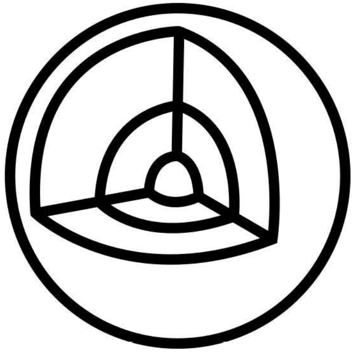
Uncertainties





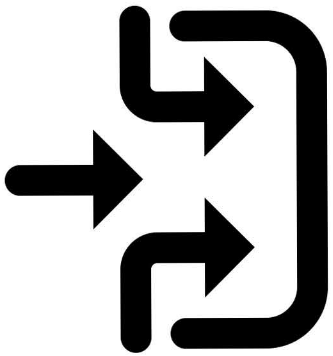
We hit our primary target

So much room for improvement

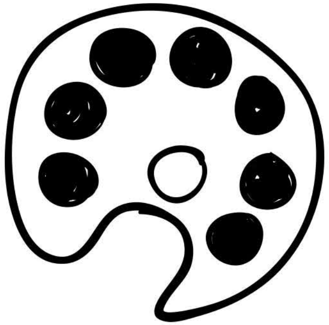


Extra base kernels
(e.g. heteroscedastic)

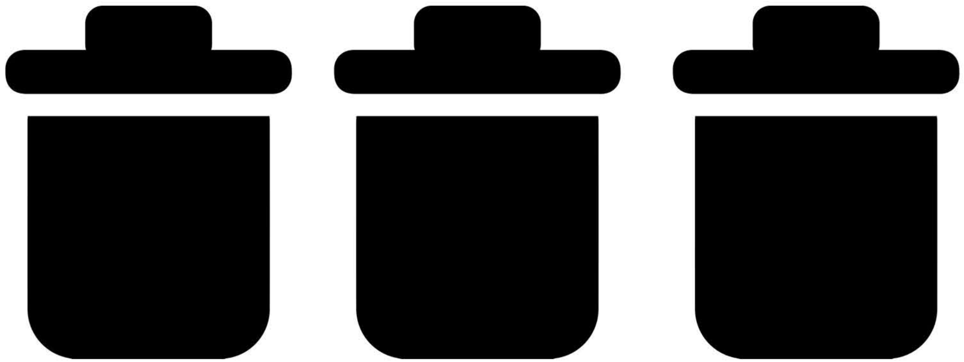
Spatio-temporal



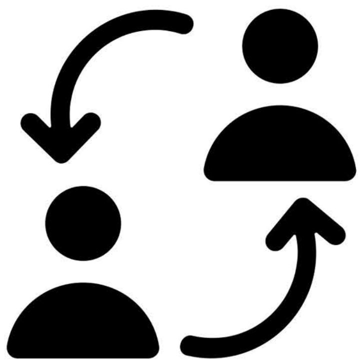
Additional inputs
(bands, humidity,
sunlight exposure,
soil composition)



How many colors will we chose?



Binning: (quantile, equal, fixed etc.)



Consult with the expert for the values of interest

Thank you for your attention!