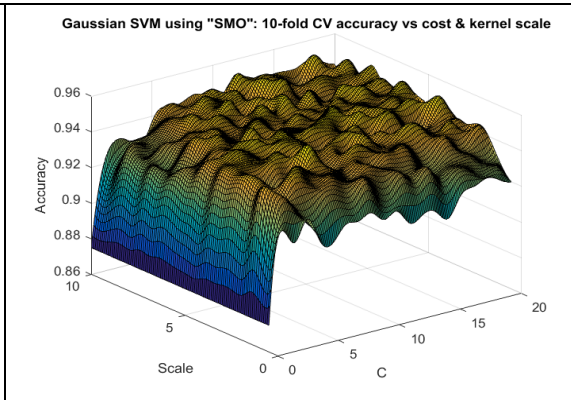
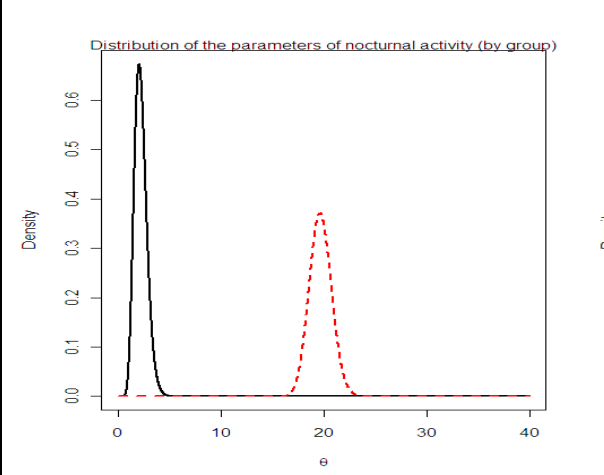


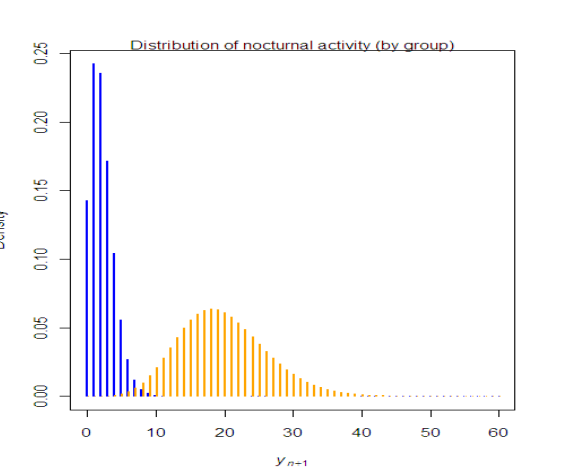
Robust regression: spline-smoothed predicted response surface ( $\approx$  beta-binomial model, not shown)



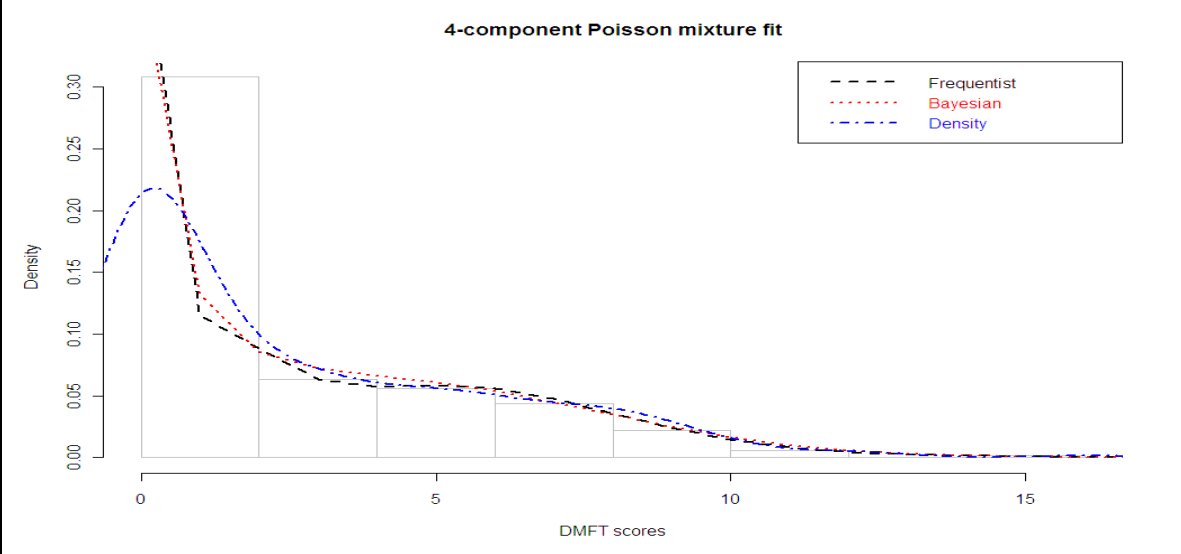
Support Vector Machine “SMO” using Gaussian kernel optimisation (*Matlab*)



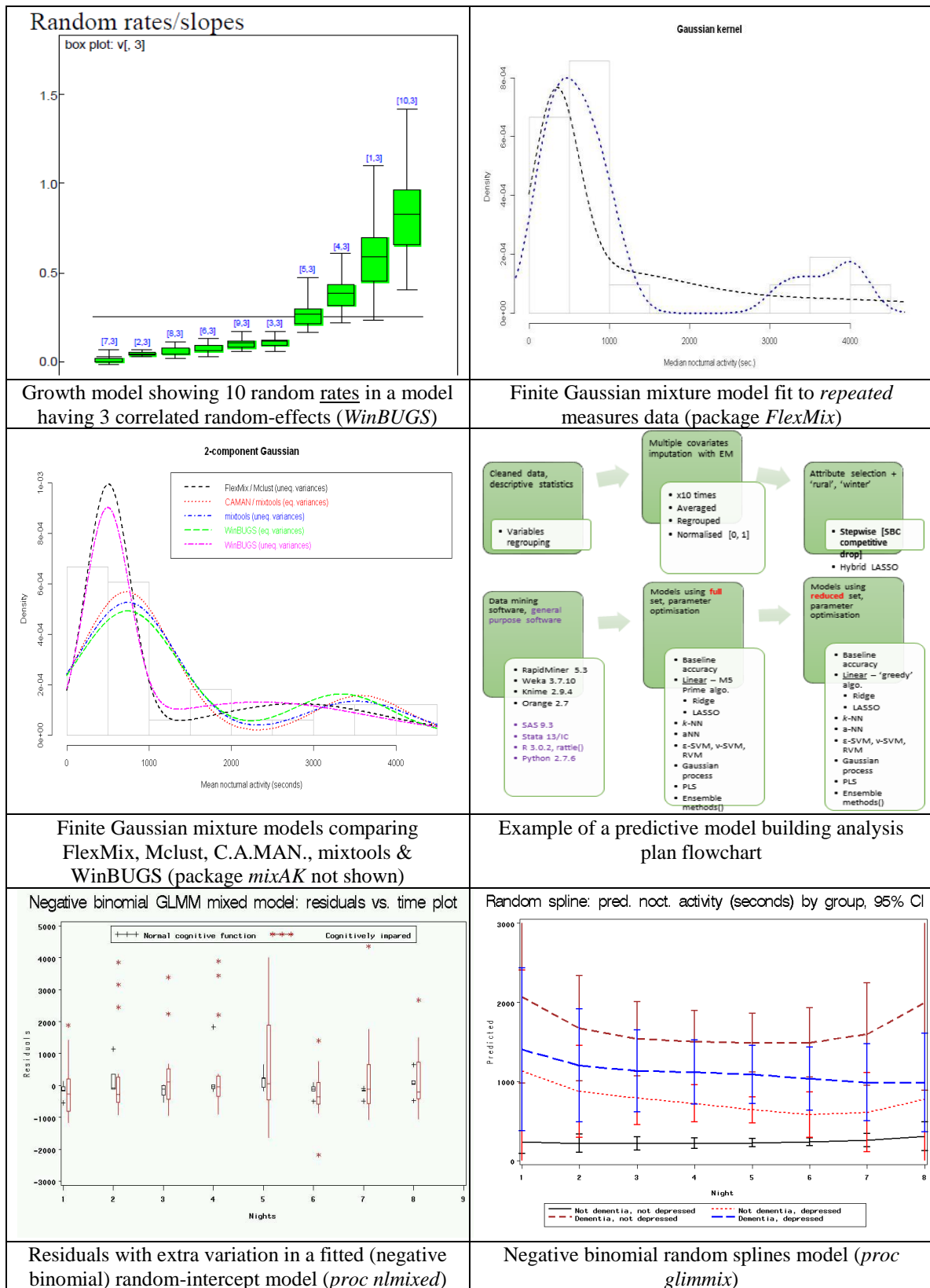
Gamma posterior distribution of  $\theta|Y$  (by group)



Negative binomial posterior predictive distribution of  $\hat{Y}|Y$  (by group)



Finite Poisson mixture model: comparison of C.A.MAN. vs. WinBUGS fits



### Finite Gaussian mixture models comparing *FlexMix*, *Mclust*, *C.A.M.A.N.*, *mixtools* & *WinBUGS* (package *mixAK* not shown)

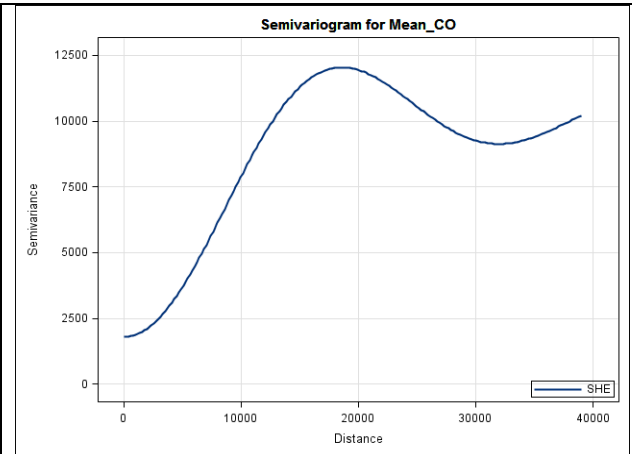
### Example of a predictive model building analysis plan flowchart

### Negative binomial GLMM mixed model: residuals vs. time plot

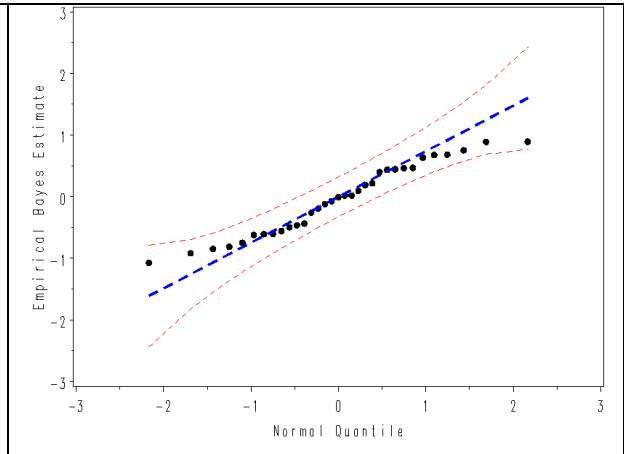
### Random spline: pred. noct. activity (seconds) by group, 95% CI

### Residuals with extra variation in a fitted (negative binomial) random-intercept model (*proc nlmixed*)

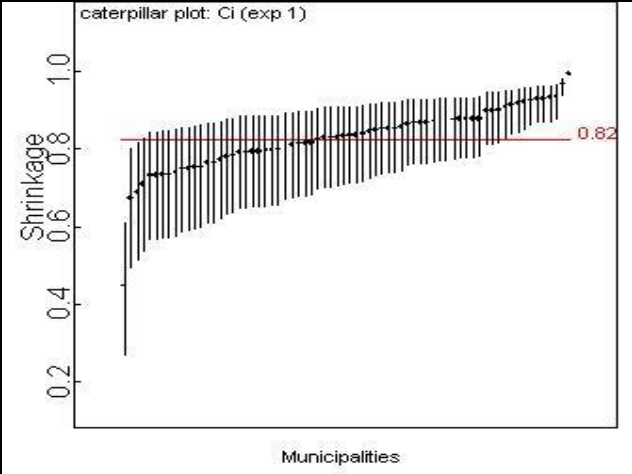
### Negative binomial random splines model (*proc glimmix*)



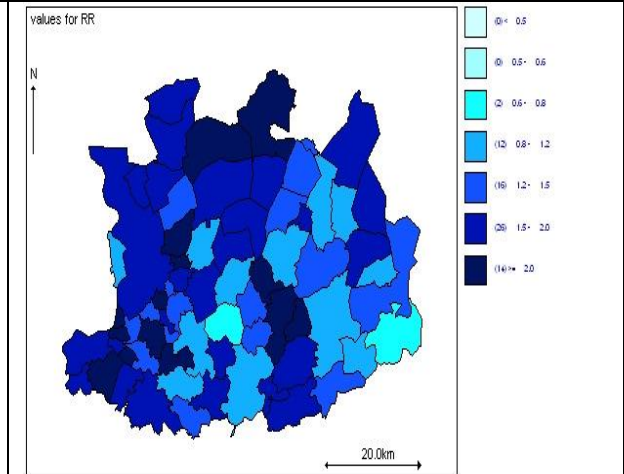
Semivariogram with ‘sine hole effect’ of air pollution data (showing nugget effect) (*proc krige2d, proc variogram*)



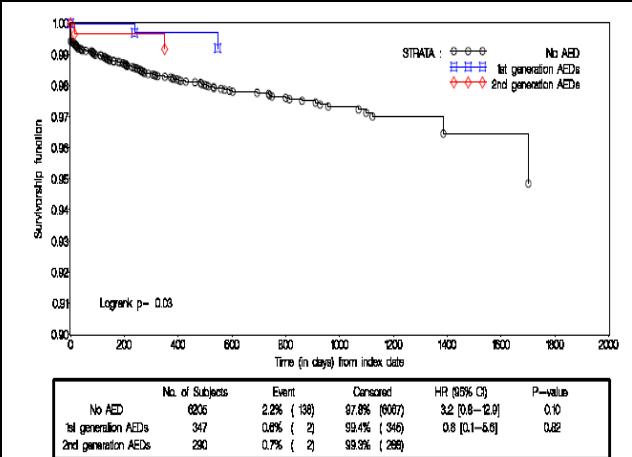
Empirical Bayes *qqplot* ( $\pm$  95% CI) of random intercepts



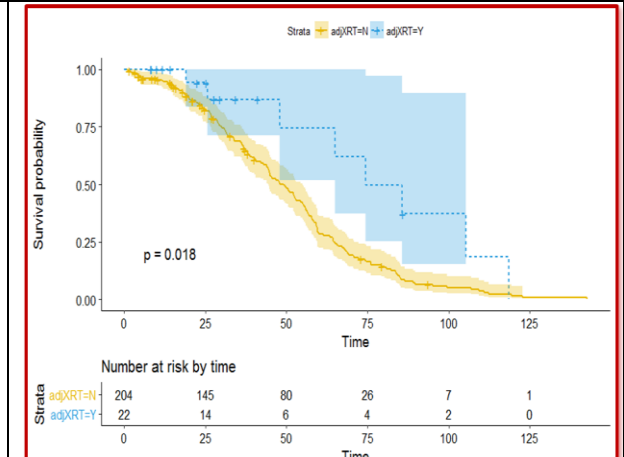
Disease mapping: ‘Bayesian shrinkage’ – ranked caterpillar plot in a Poisson-gamma model (*WinBUGS*)



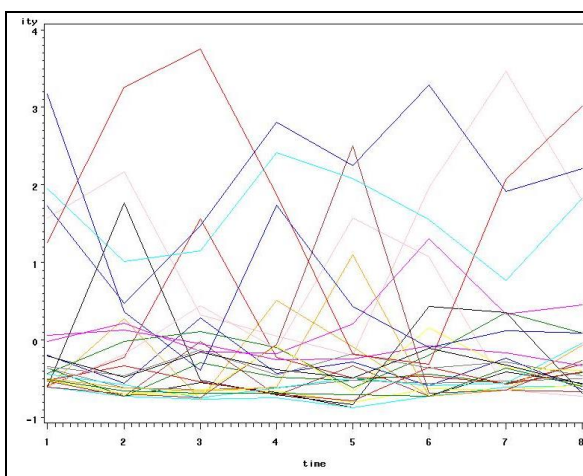
Disease mapping: CAR model of cervix cancer incidence in the Antwerp district (*WinBUGS*)



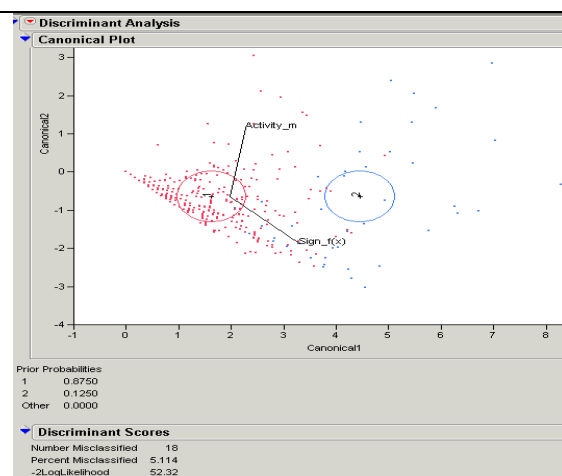
Suicide attempts ↓ with AED (any generation) treatment contrary to FDA alert (unmatched, unadjusted analysis)



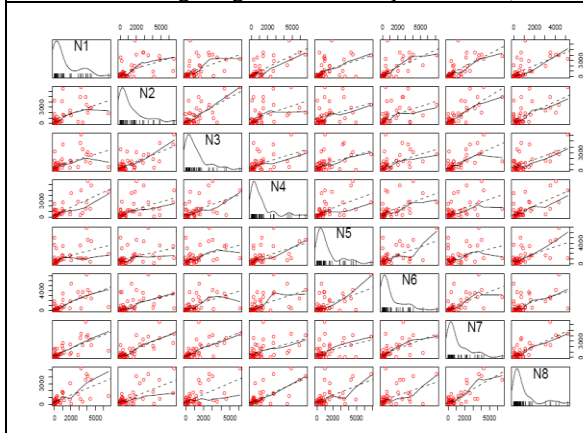
KM plot comparing RTX + CTX treated vs. untreated CRC patients (package *survminer*)



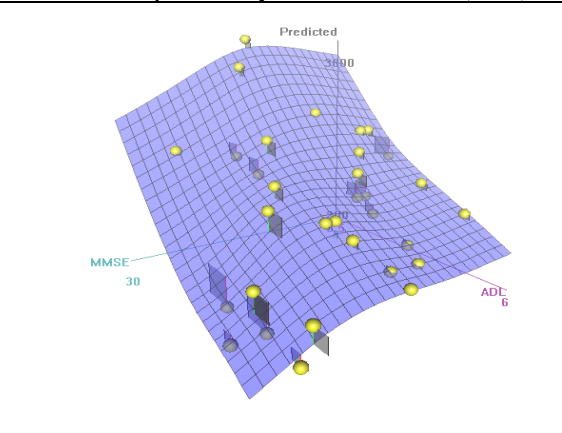
Tracking longitudinal data (*proc stdize*)



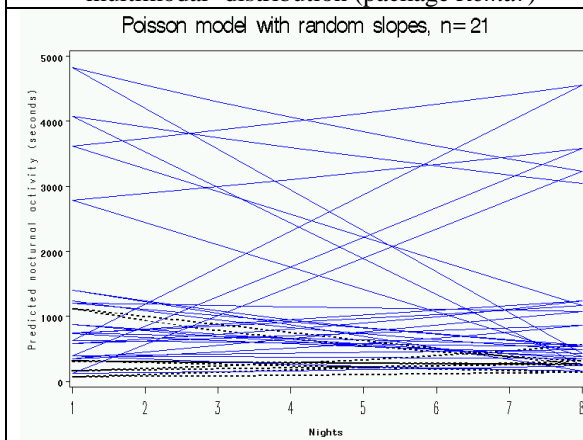
Linear empirical Bayes discrimination (*JMP*)



Correlation scatterplot between 8 repeated measurements; with diagonals showing right skewed 'multimodal' distribution (package *Rcmdr*)

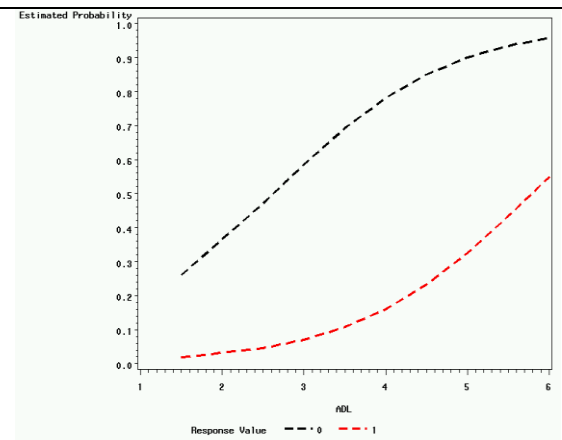


Spline-smoothed predicted response surface with squared residuals (screenshot during 360° rotation)

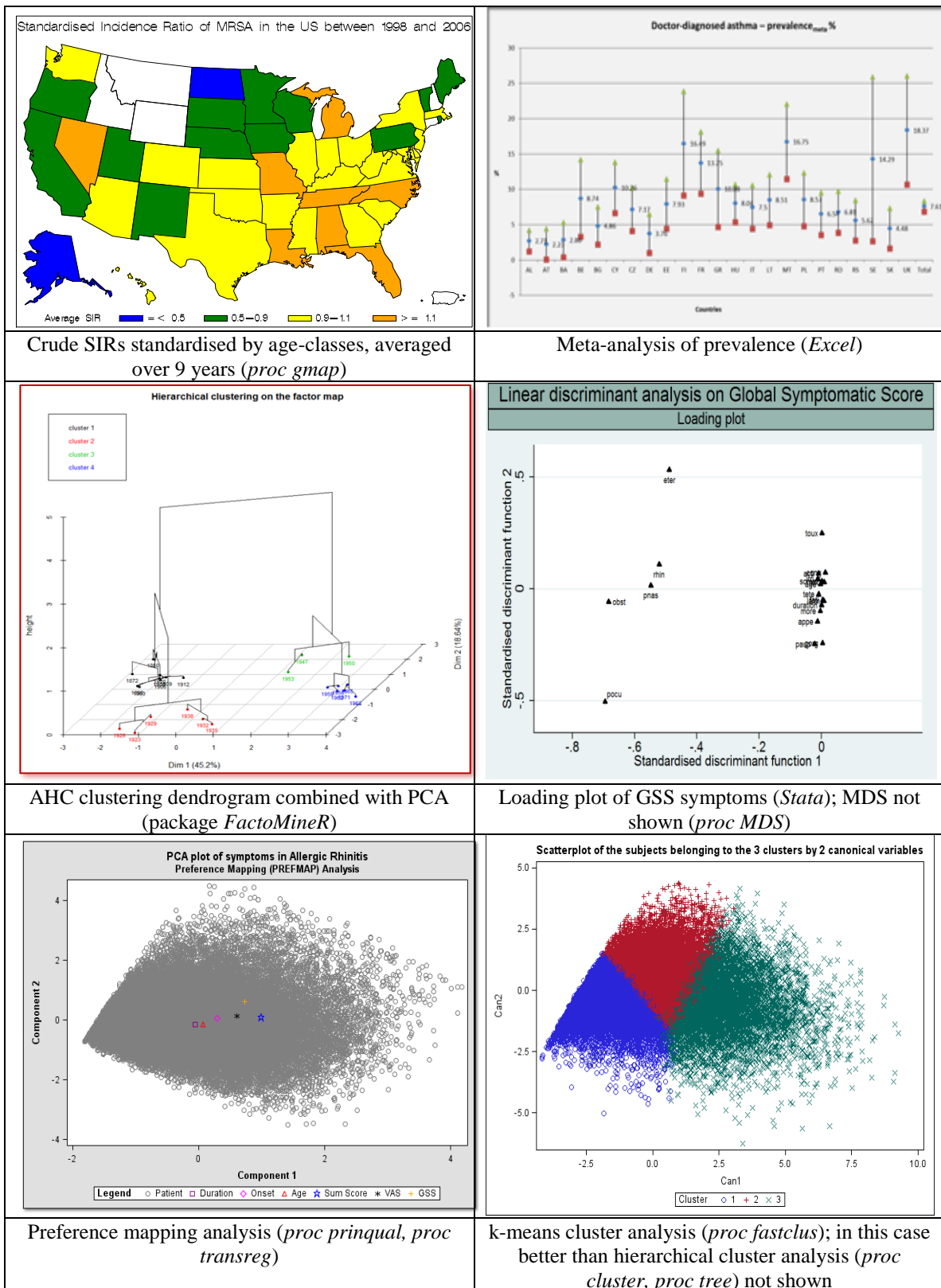


Poisson model with random slopes,  $n=21$

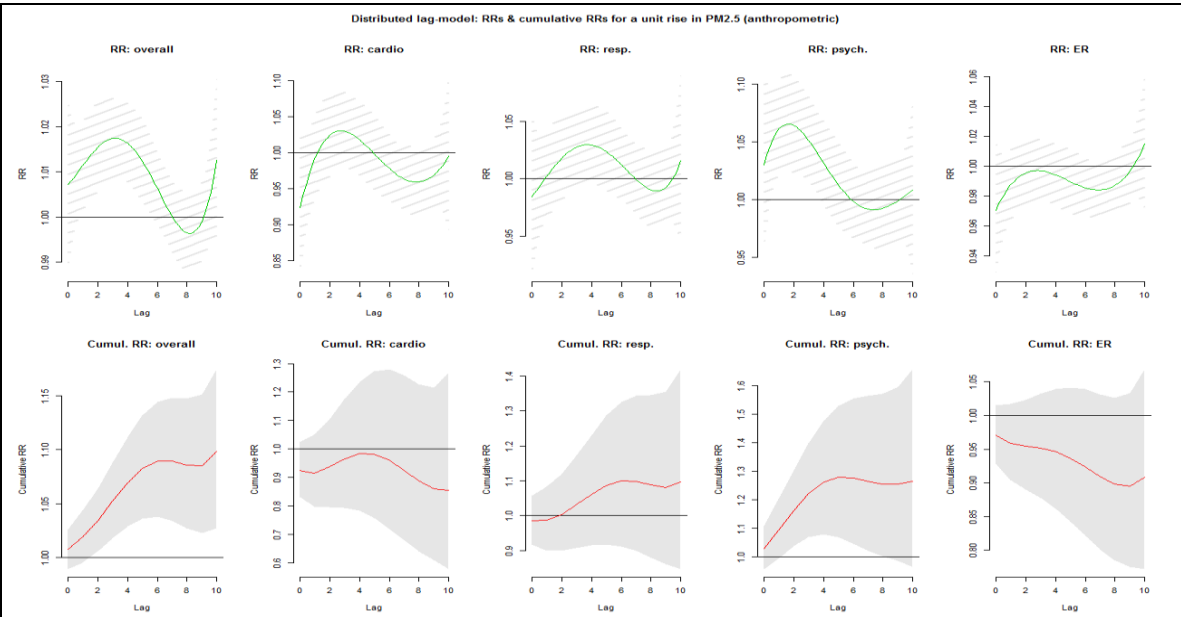
Random slopes by group: Poisson model (*proc nlmixed*)



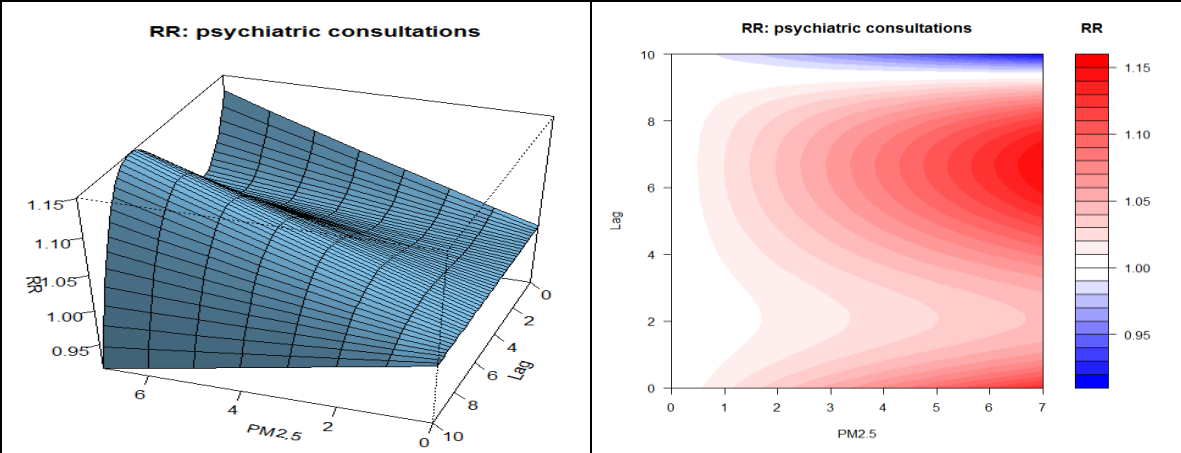
P-O model comparing 3 categories: 'worsening' (red), 'stable' (black) & 'improving' (reference) → subjects have a constant impact (OR) of ADL (exposure) across any level



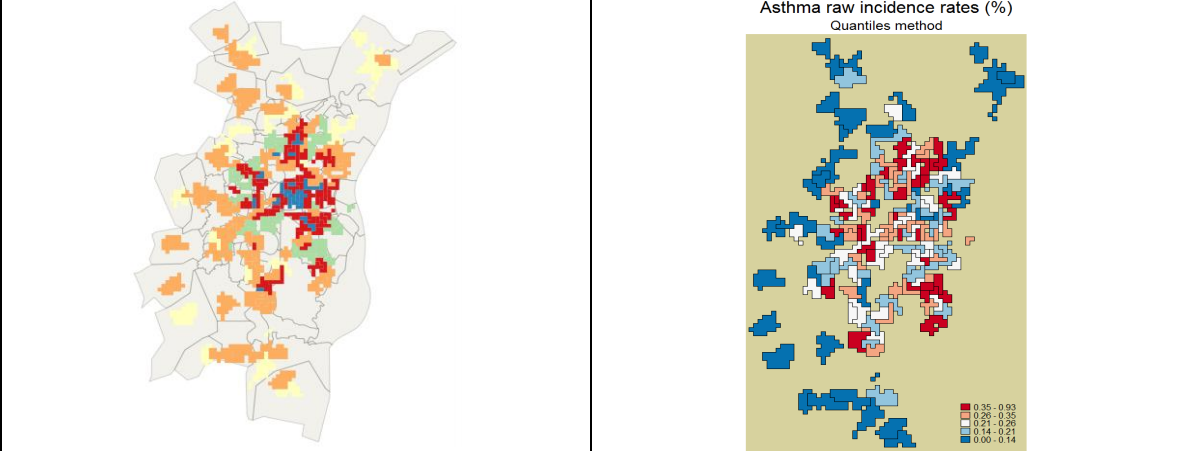




Time-Series Analysis (TSA) with GAMs: Distributed Lag Models (DLM) (package *dlm*)



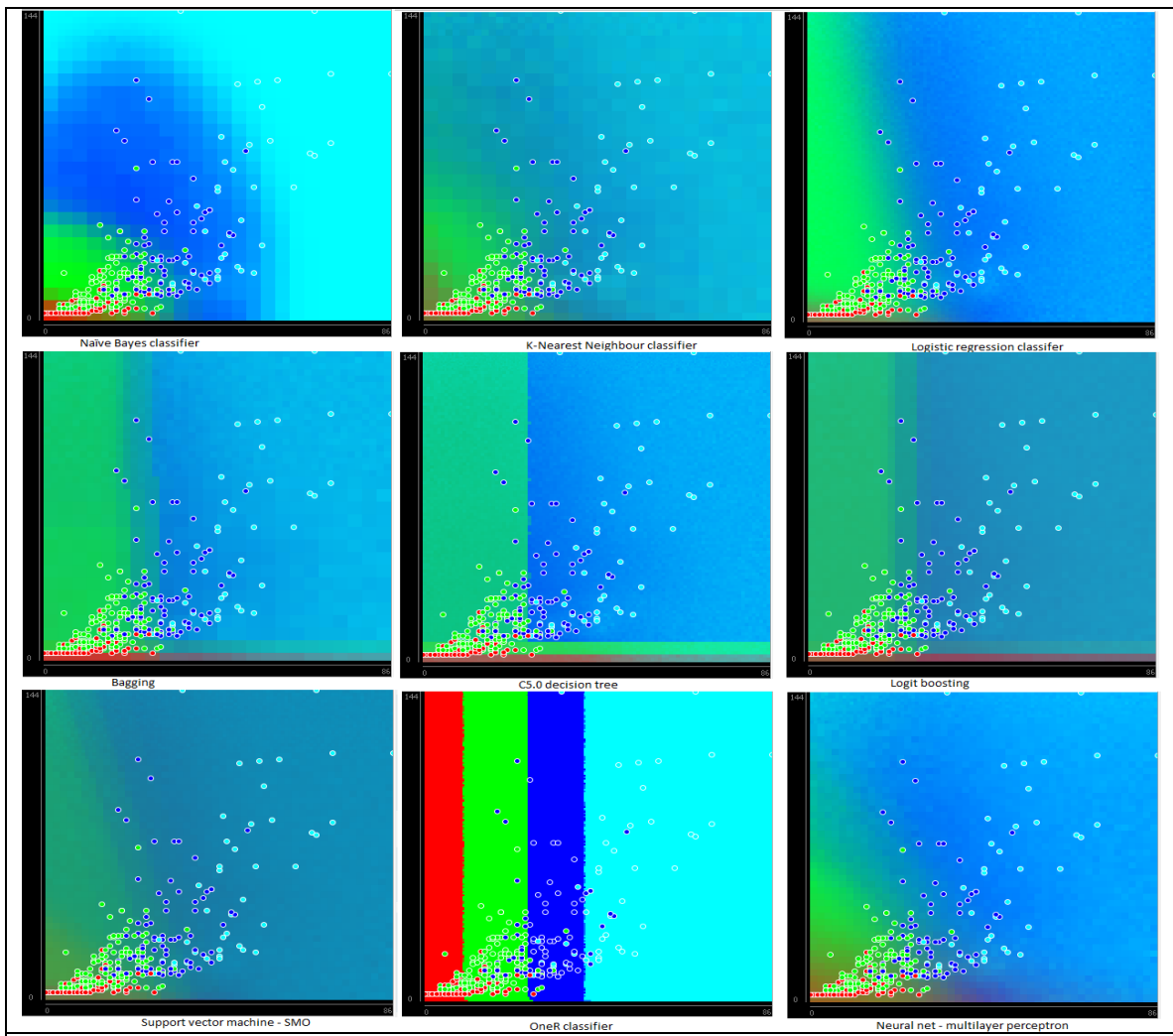
TSA with GAMs: DLM (package *dlm*)



IRIS map overlaid by new zoning map showing neighbourhood deprivation index (*QGIS*)

Raw incidence rates in the new zonings (*Stata*)





Decision boundaries of different classifiers (Weka)

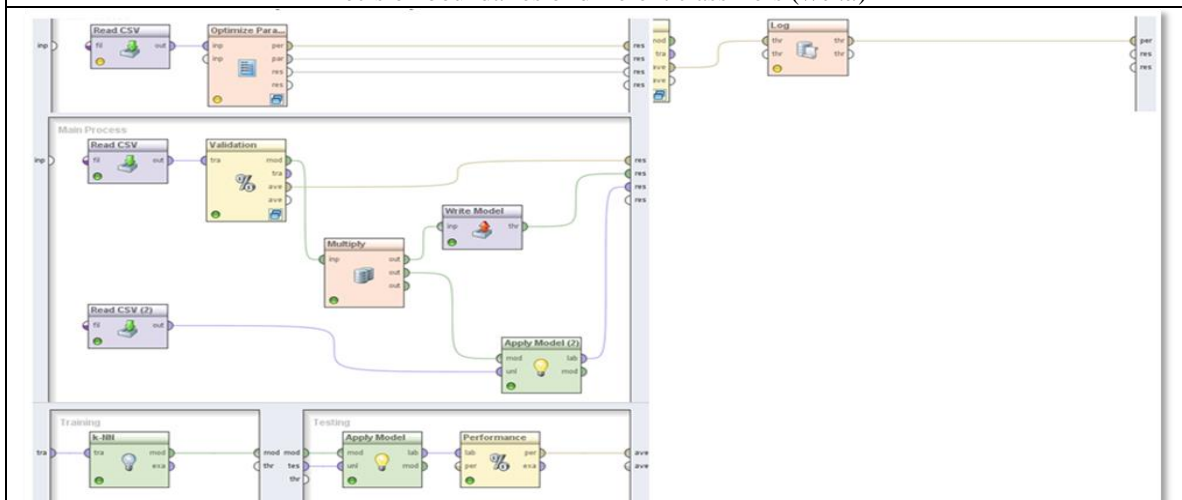
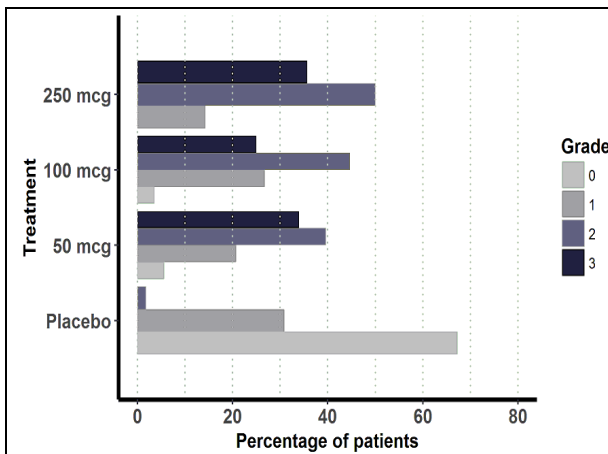


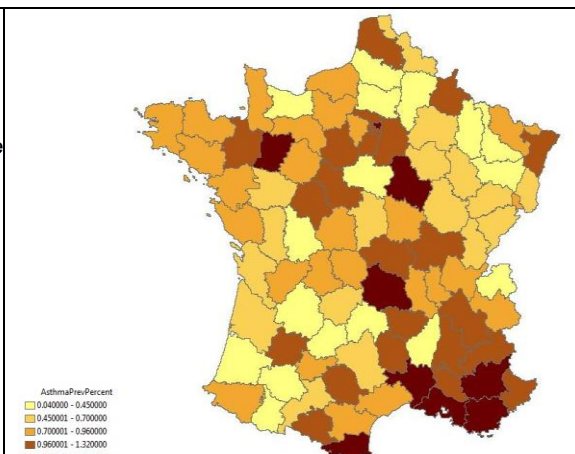
Figure 1. A flowchart on grid search parameter optimisation (top 2) followed by  $k$ -fold crossvalidation (bottom 2) (here,  $k$ -NN) using RapidMiner.

Parameter optimisation & predictive modelling flowchart (RapidMiner)

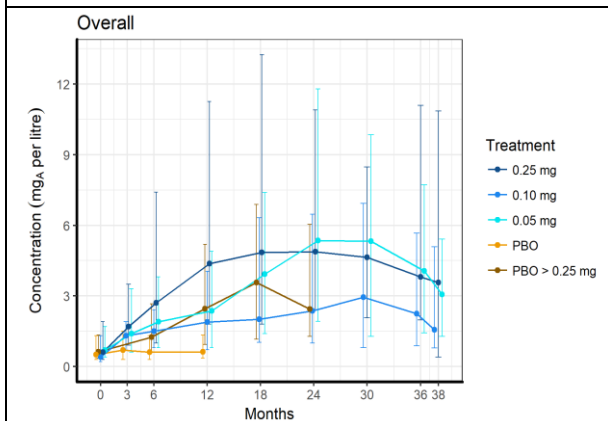




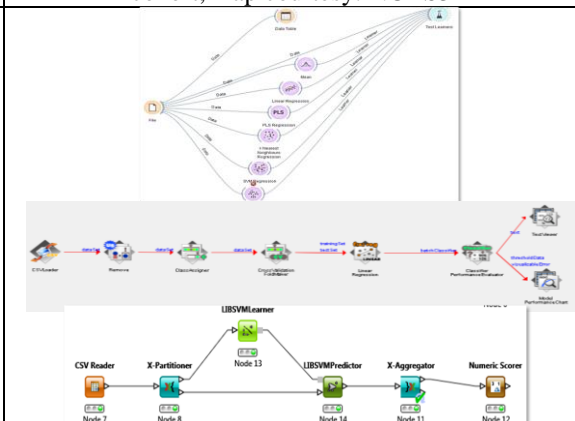
Barchart of % by treatment, grade (package *ggplot2*)



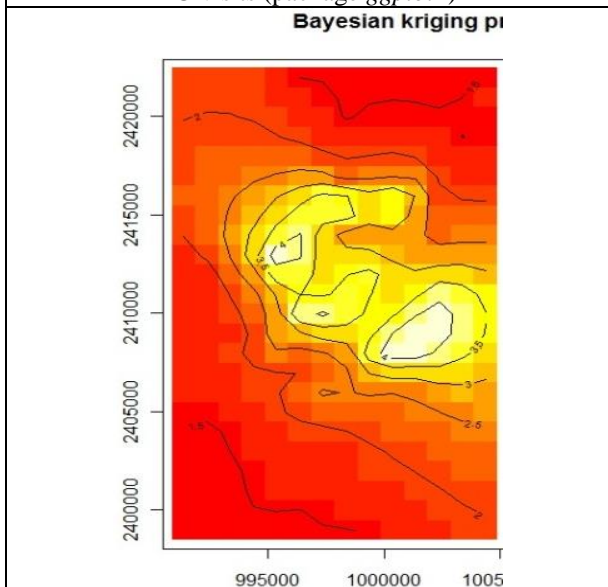
Prevalence % in France (*ArcGIS*) from the EGB cohort, map courtesy: NUTS3



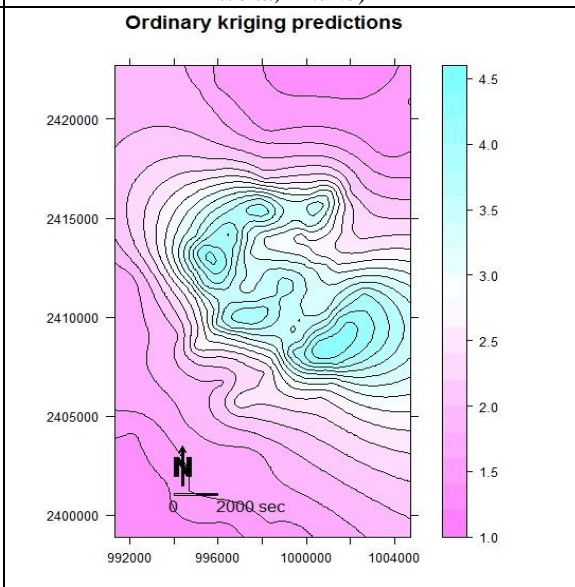
Median (IQR) by treatment w.r.t. study duration @ FU visits (package *ggplot2*)



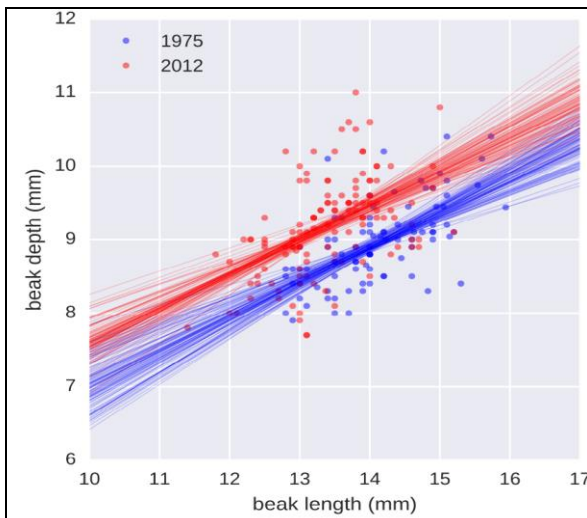
Predictive modelling process flow (*Orange miner*, *Weka*, *Knime*)



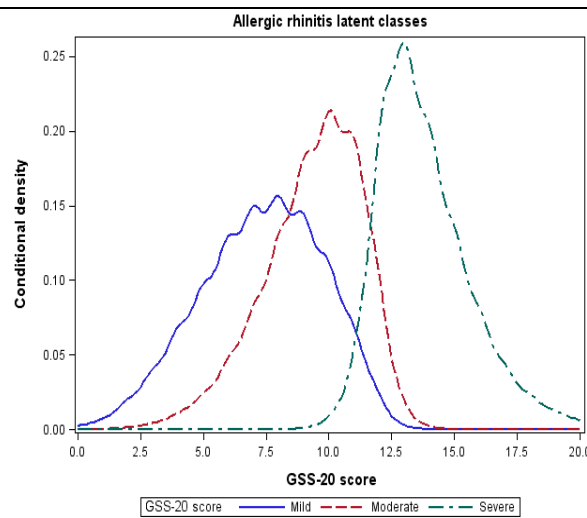
Bayesian Kriging of SES data (package *sgeostat*)



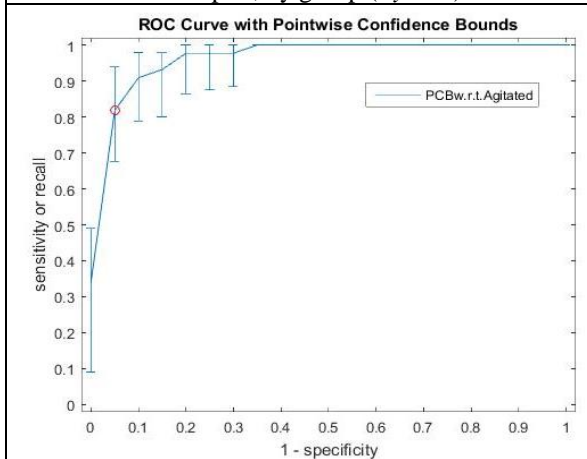
Kriging of SES data (package *gstat*)



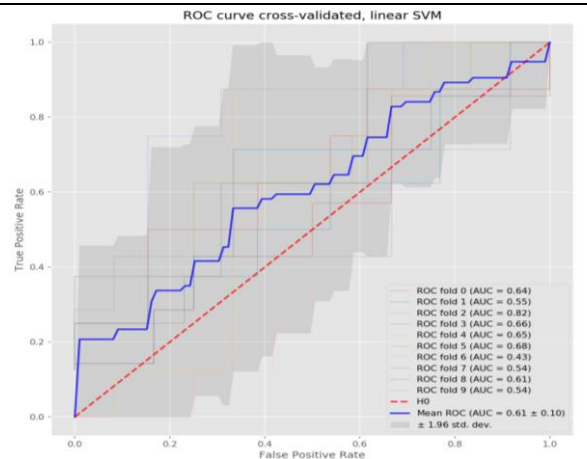
Bootstrapped linear regression overlaying the scatter plot, by group (*Python*)



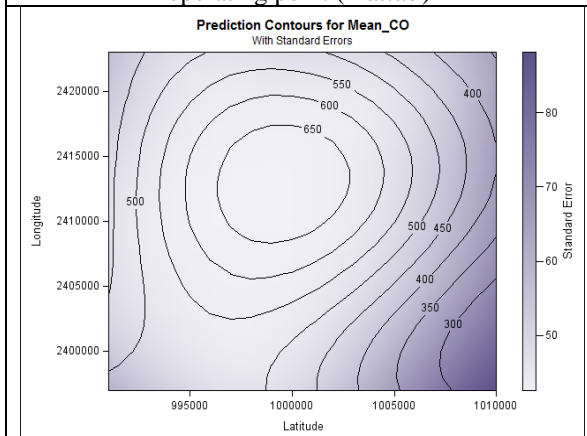
Latent class analysis (*proc lca*)



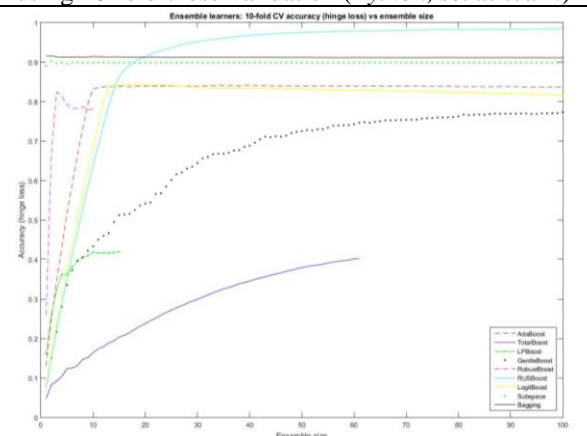
ROC curve with bootstrapped 95% CI & optimal operating point (*Matlab*)



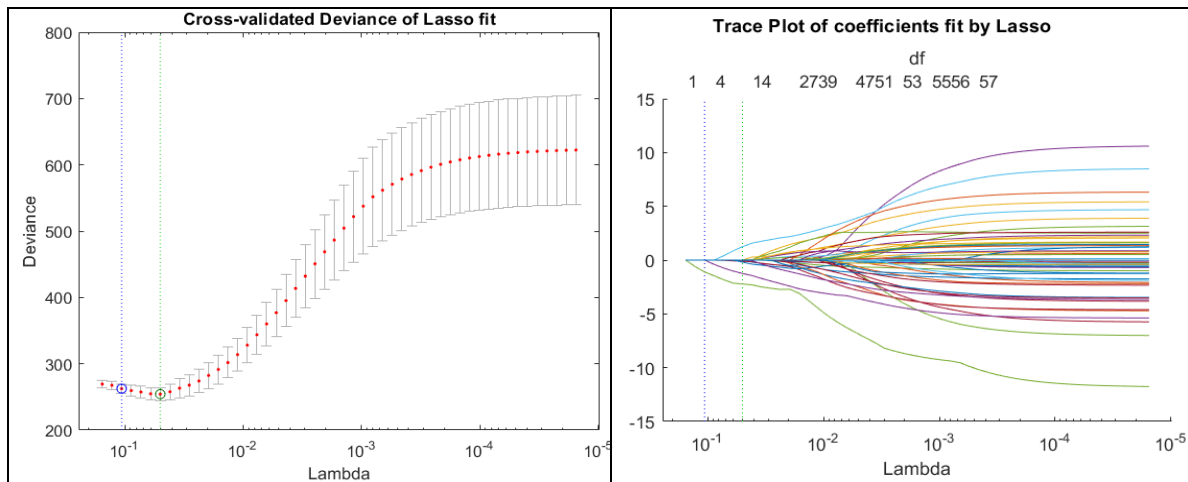
ROC curve with 95% CI in a linear SVM model using 10-fold cross-validation (*Python, scikit-learn*)



Kriging of air pollution (CO) data (*proc krige2d, proc variogram*)

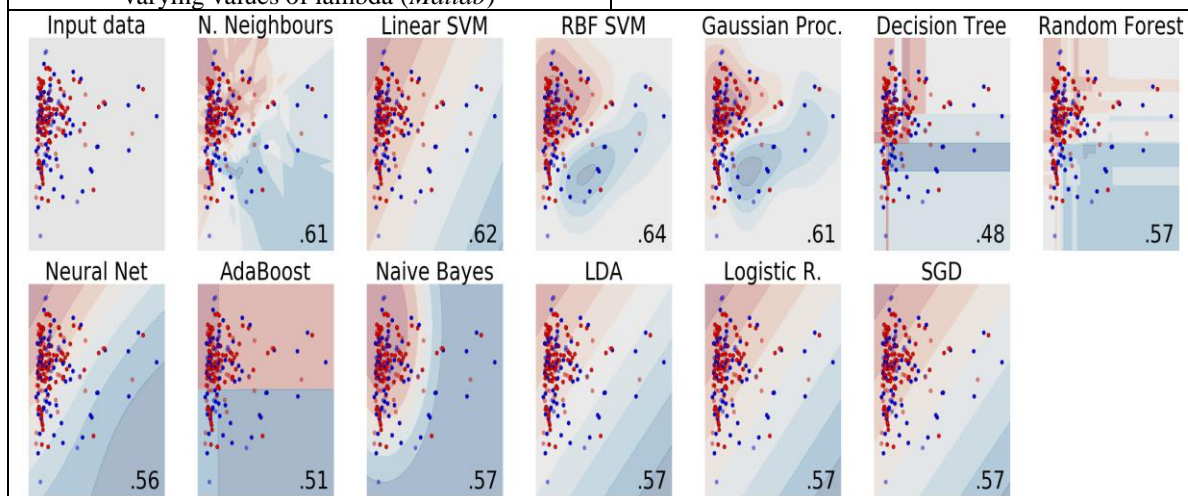


Ensemble supervised machine learning methods (boosting, bagging, random subspace) (*Matlab*)

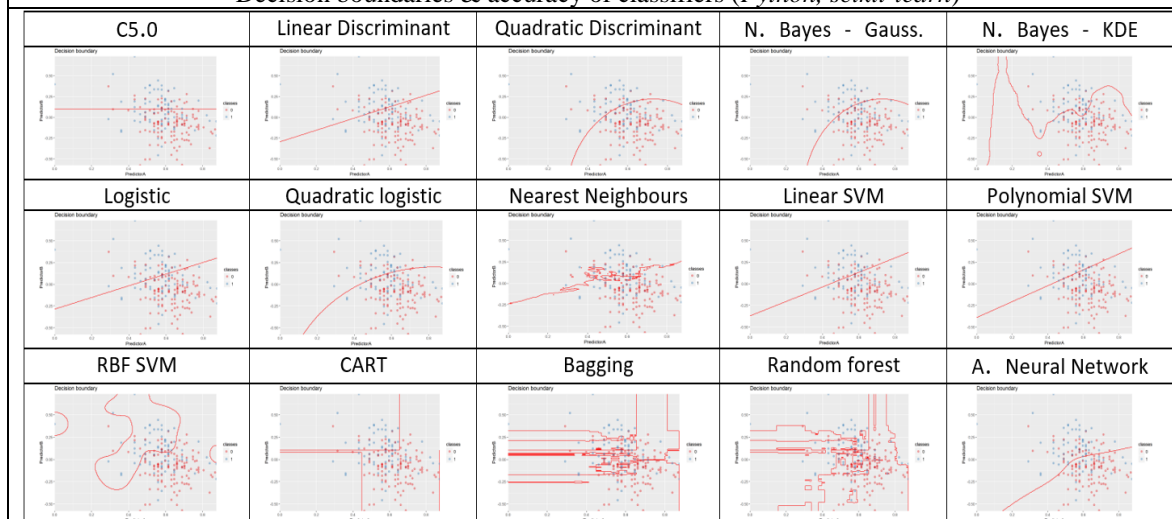


Lasso cross-validated deviance (95% CI) plot for 50 varying values of lambda (*Matlab*)

Lasso trace plot of coefficients (*Matlab*)

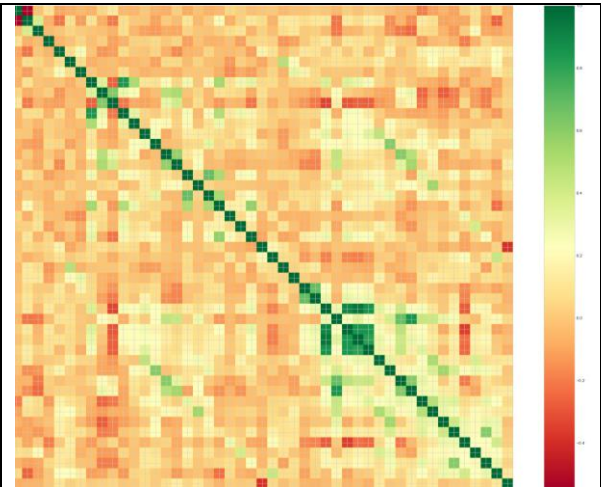


Decision boundaries & accuracy of classifiers (*Python, scikit-learn*)

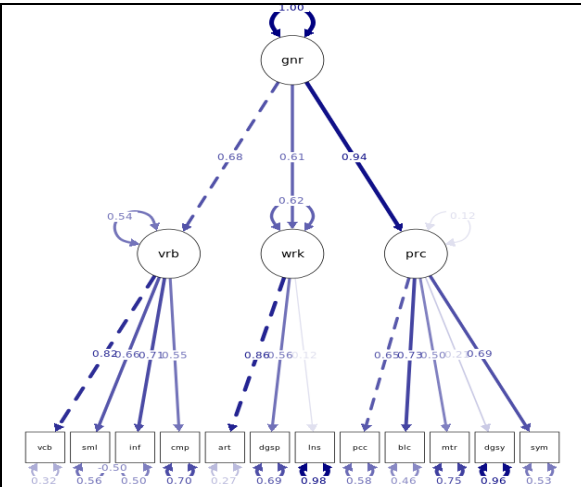


Decision boundaries of classifiers (package *caret*)

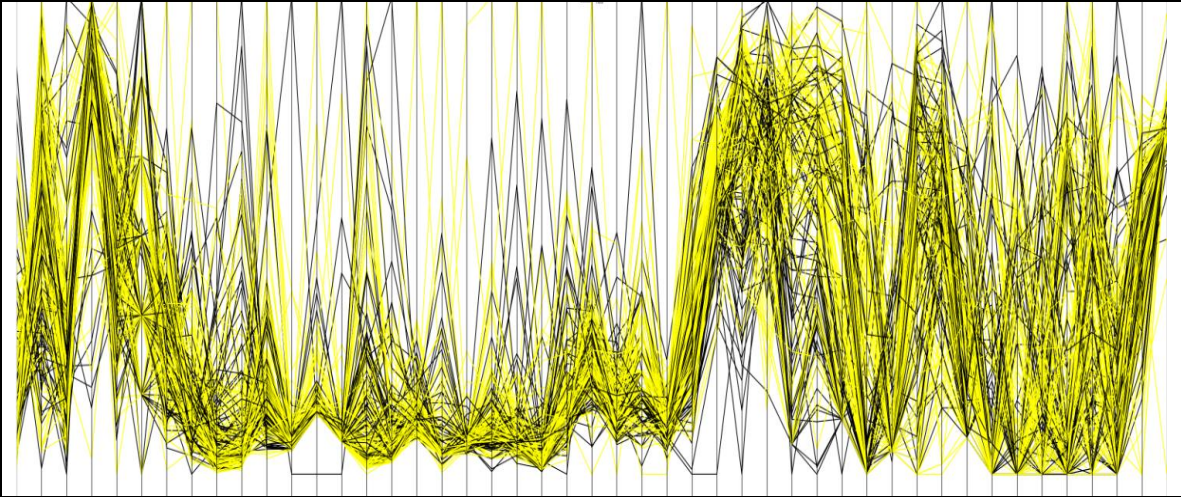




Correlation heat map after normalisation (*Python*)



Hierarchical S.E.M. plot (package *lavaan*)



Parallel coordinates plot by group (*Python, seaborn*)
