

A bibliographic review of non-parametric methods and their application

Emmanuelle Comets^{1*} Julie Antic^{2,3} and Rada Savic^{1,4}

¹ INSERM U738, Paris, France; Université Paris Diderot (Paris 7), UFR de Médecine, Paris, France ² UMR 181, Ecole Nationale Vétérinaire, Toulouse, France ³ Institut de recherches internationales Servier, Courbevoie, France ⁴ Uppsala University, Uppsala, Sweden

Objective: To survey the use of nonparametric methods within the non-linear mixed effects framework through a bibliographic study

Methods

Article selection

- Database search
 - PubMed : biomedical and life sciences articles
 - MathsScinet : maths journals
 - Keywords used in both searches: NPEM OR NPAG OR NPML OR ((non-parametric or nonparametric) and mixed and effects and model))
- Full text obtained through online libraries, mail service or direct contact. Articles divided in 3 groups:
- each group read by two readers
 - cross-checking to resolve any disagreement
 - papers not in English excluded from database

Data abstraction form

Relevant information extracted using a data abstraction form (DAF):

- general characteristics: address, affiliation, journal, pathology; drugs tested, population, study objectives
 - study design: number of subjects, number of observations per subject
 - NP methods: estimation methods, software, other methods used
 - model development: models tested, number of fixed and random effects, error model, model building
 - results reported
 - model evaluation: uncertainty, graphs, sensitivity analysis
- The DAF was built during two consensus meetings between the 3 authors, and modified after reading a first set of 3 papers.

Statistical analysis

- descriptive statistics
- data management and statistical analyses using R

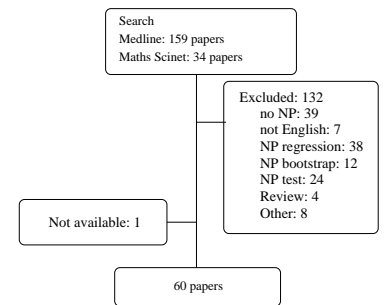


Figure 1: Flow-chart of the selection process (the searches were performed in March 2009).

Results

General results

Journal type	Pharmacology	39
	Medicine	12
	Statistics	9
Affiliation	Academia	49
	Both (including 4 grants)	10
	Industry	1
Paper focus	Application	43
	Methodology	13
	Both	4
Data type	PK	47
	PD	2
	PK/PD	3
	other	8
Indication	infection	19
	none	12
	other	9
	cancer	8
	analgesic/anesthesia	4
	epilepsy	4
	transplantation	3
	antibiotics	1
Data used	real	50
	both	6
	simulated	3
	none	1
Sampling design	sparse	18
	rich	17
	medium	8
	mix	7
	both	4
	not applicable	6

Table 1: Study description and data.

- Mostly used by academia
- Study characteristics
 - many applications in infectious diseases (antibiotics)
- Study design
 - very variable number of subjects: median 50, range 8-696 (interquartile range 25-100)
 - number of observations per subject also very variable: median 4, range 1-24 (missing or unclear in 12 papers)

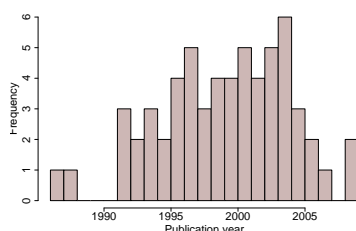


Figure 2: Evolution with time of the number of nonparametric analyses published.

Models

Covariates in dataset	no	3
	yes	57
Several models tested	no	45
	yes	15
Model comparison	Statistical criteria	7
	Predictive performance	5
	Other	2
	NR	3
Estimation method	NPEM/NPAG	43
	NPML	10
	Bayes	2
	Other	5
	Other estimation method used	no
	yes	36

Table 2: Models tested, covariate models and model building.

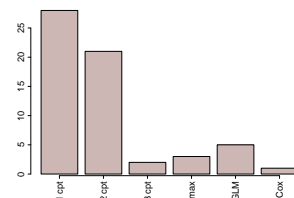


Figure 3: Models used in the analyses.

- Models
 - structural models: mostly simple compartment models
 - residual error models: polynomial models of degree 0 (homoscedastic) to 3
 - frequently only one model tested
- Parameters
 - fixed effects: range 0-7
 - * parameters of the residual error model most often fixed
 - * few estimation methods
 - random effects: median 3, range 1-10 (including covariates in NPML)
- Covariates present in 57 papers
 - included in the model in 44 papers
 - usually according to a predetermined relationship (N=33)
 - in NPML, covariates considered as additional locations (N=4)
 - statistical criterion used: AIC or log-likelihood (N=2), entropy reduction (N=1), predictive performance (N=1)
 - not reported in N=3
 - in general, methods for covariate inclusion/selection lacking
- Parameter estimation
 - estimation for fixed effects part of the estimation algorithm in N=4 methodological papers
 - software: USC*PACK (N=35), NPML (N=9), own code (N=6)
- Other estimation methods used in over half of the papers
 - IT2B (N=12): mostly to provide preliminary estimates and parameter ranges in NPEM analyses
 - FO/FOCE (N=10), two-stage methods (N=10), Bayesian methods (N=5): comparison of results
- Diffusion of non-parametric estimation methods
 - 28 analyses published outside the group developing the method
 - including 13 with acknowledgements for help

Results reported and model evaluation

Random effects reported	Mean/median and SD/CV	48
	Distribution	7
	NR	6
Uncertainty reported	no	52
	yes	8
Sensitivity analysis	yes	4
	no	55
	NR	1
GOF plots	Observations versus predictions	19
	Residuals	2
	Other	1
	None	37
External evaluation	no	52
	yes	8

Table 3: Results reported in the non-parametric analyses.

- Results reporting
 - random parameters usually reported as mean/median and SD: insufficient to use the results
 - distribution often plotted to detect multinormality
- Uncertainty
 - few methods available to obtain an estimation of the uncertainty on the nonparametric distribution
 - jackknife and bootstrap procedures proposed
- Model evaluation
 - individual predictions: no consensus but often MAP or EAP
 - model diagnostics: gof plots using observation-based metrics such as predictions, residuals, ...

Discussion

Main findings:

- Type of data: continuous or survival data
 - never applied to categorical, count or time to event data
- Fixed effects
 - seldom estimated, usually fixed to predetermined value
 - * residual variability often fixed to assay error
 - * NONMEM VI: use preliminary parametric analysis
 - some original methods:
 - * NPML: iterative procedure but convergence not proven
 - * EM algorithms including estimation step for fixed effects
 - * Bayesian methods: integration
- No large scale evaluation of nonparametric estimation methods
- Essentially no model building
 - frequently only one model tested, sometimes relying on a prior parametric estimation
 - no complex models tested: runtime limitations?
 - evaluation of model building procedures lacking
- Evaluation of uncertainty still an issue for non-Bayesian methods
 - recent developments in bootstrap methods
- Limited model diagnostics
- Decreased use in recent years

Conclusion:

In general, estimation methods have good potential but there is a room for improvements (model building procedures, evaluation of uncertainty, fixed effect estimation) to ensure their wider use.