KU KDM 2018
Konferencja Użytkowników Komputerów Dużej Mocy
Zakopane, 7-9 marca 2018 r.

ON PROPOSED ANALYSIS OF SCARCE AND LOW PRECISION MEASURED DATA

Janusz Orkisz
Institute for Computational Civil Engineering
Cracow University of Technology

CONTENTS

- Introduction
- Measured data characteristics
- New heuristic rule based uncertainty solution approach to analysis of scarce poor quality data.

  General concept – proposed weighted error functionals
- Exemplary numerical analysis – a comparitive study including the standard statistical analysis
- On error analysis
- Final remarks

INTRODUCTION

Problem characteristics

Muscle strength measurements (ATLAS – Medical Center 4M+) and data analysis

Relation: Experimental data and reality (truth)

DATA	NUMEROUS	SCARCE
PRECISION	VERY GOOD	GOOD
POOR QUALITY	GOOD	VERY POOR

Main idea of this reseach

- use heuristic rule to improve poor data

Research objective

Find various adequate ways (specific procedures) of scarce and poor quality experimental data analysis, based on different heuristic rule concepts, but providing roughly similar type results.

Find next a well based summarizing approach unifying results of all these procedures mentioned above.

INTRODUCTION – cont.

Basic assumptions

(i) Available data is scarce and of poor quality

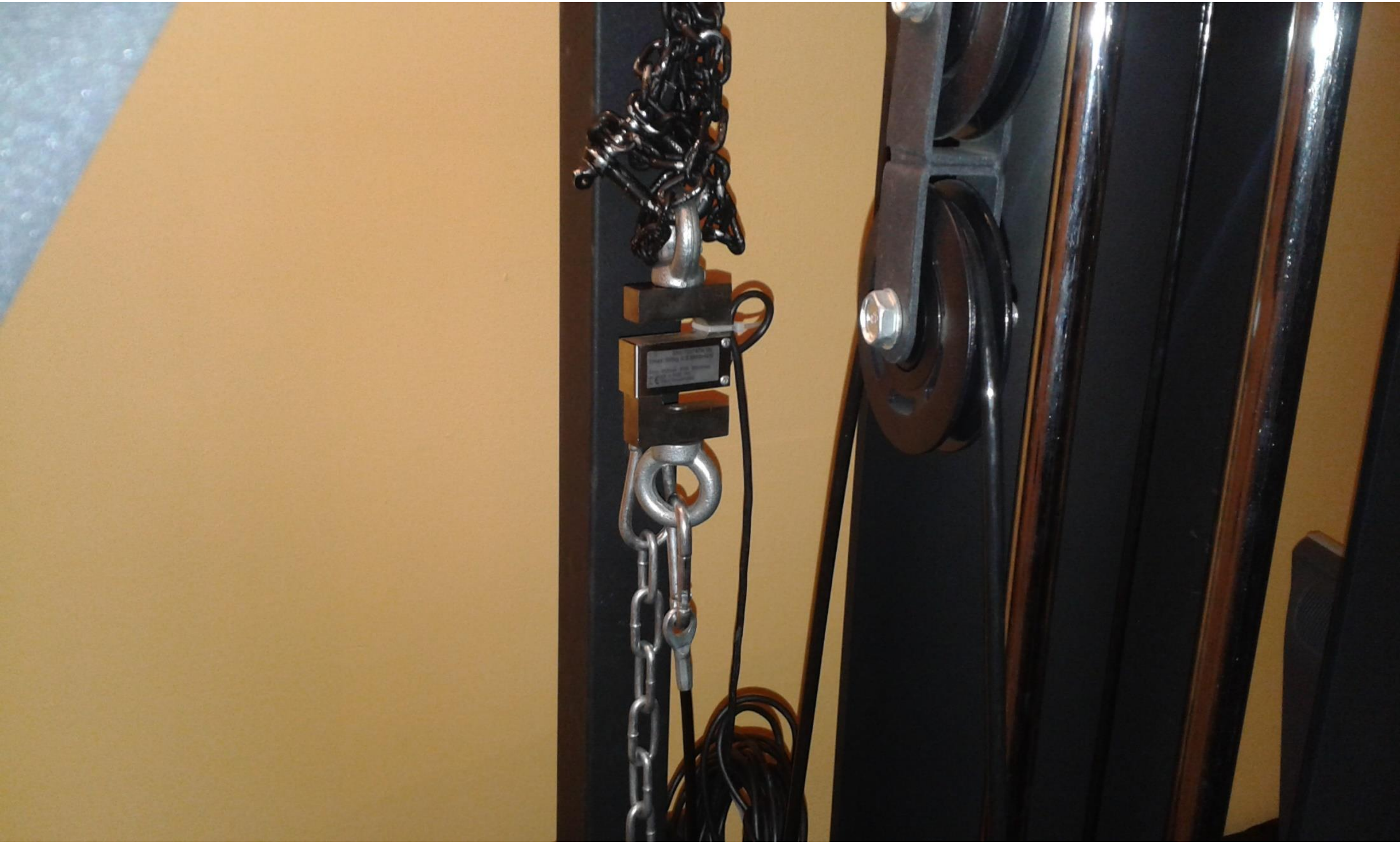
(ii) Information about the reality (the truth) provided by this data is not sufficient. However, it may be completed on various heuristic rule bases applied to generation of error functionals by means of appropriately chosen weighting factors

(iii) The final solution is obtained by means of the proposed special summation procedures unifying results found for the standard and all various, particular heuristic rule principles applied

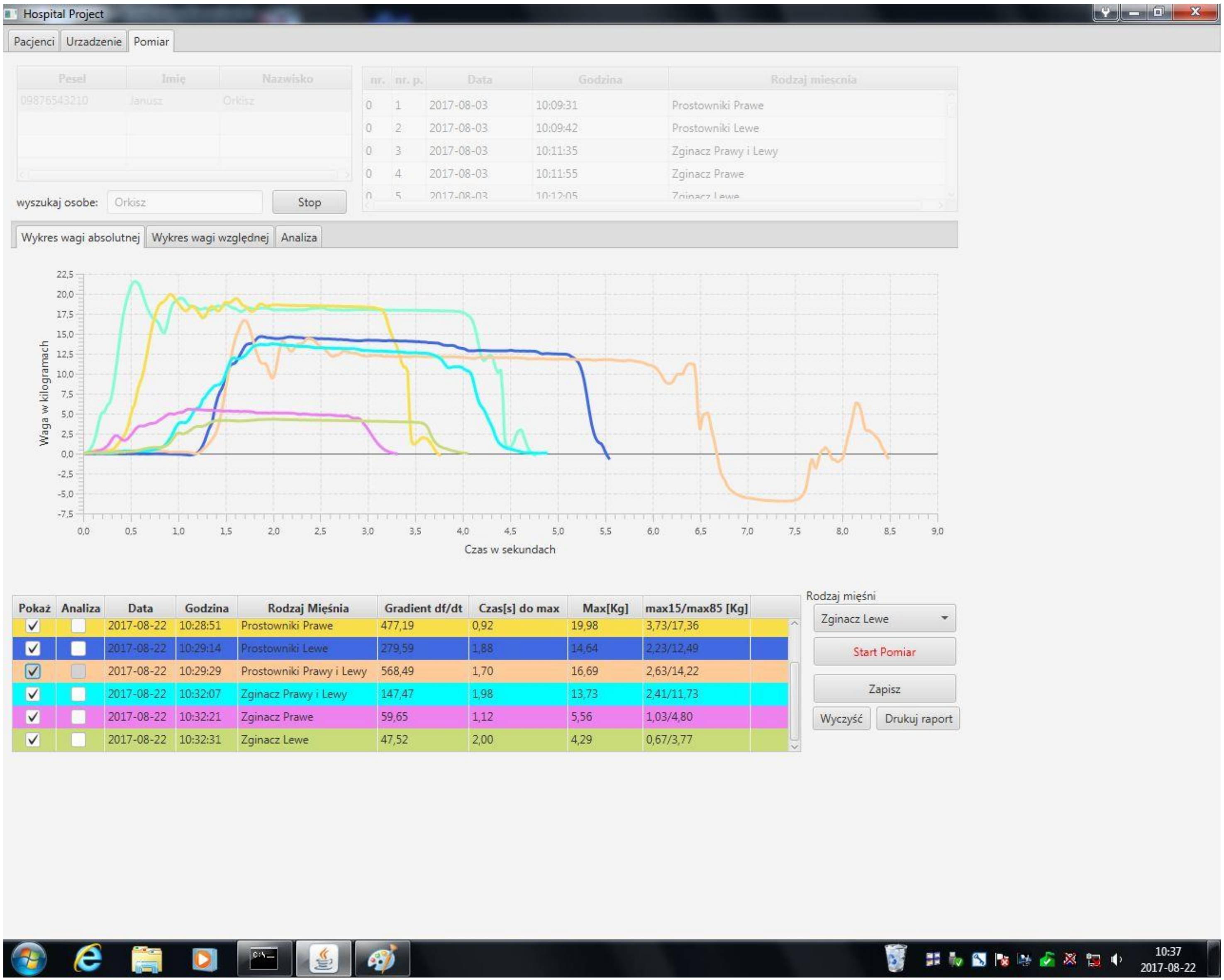
(iv) Comparison of the initial patient condition, based on measurement results, with the one after treatment, training or desease

MEASUREMENTS

MEASUREMENTS



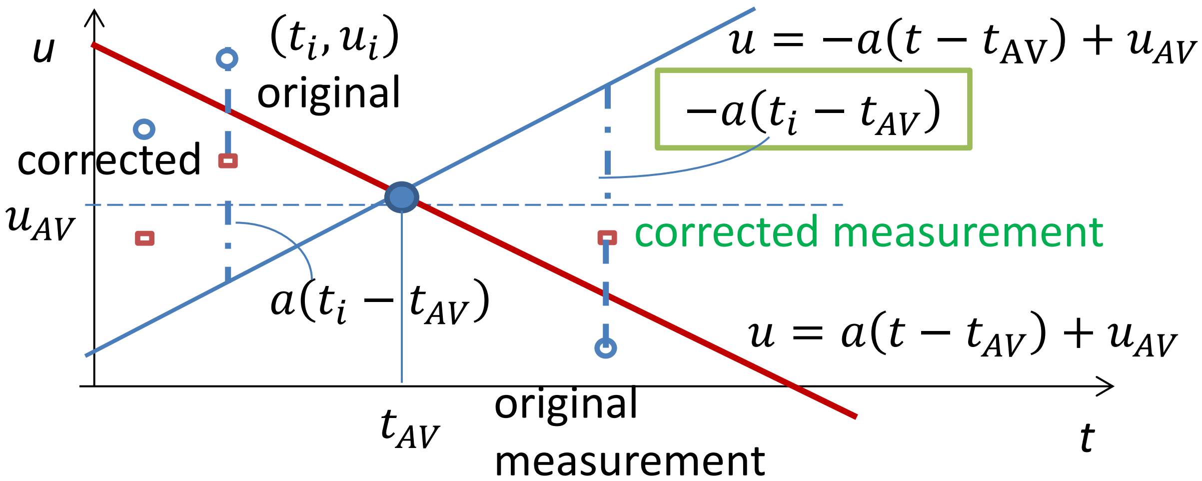
TYPICAL DATA REGISTRATION



GENERAL CONCEPT AND PROCEDURE OF ANALYSIS

- Start with 4 initial measurements of chosen muscle strength
- Preliminary data correction for muscle weariness using the linear regresion approach
- Generation of newly introduced error functionals developed on heuristic rule bases and normalized
- Parallel application of 8 (4 standard + 4 new) specific solution procedures, each one providing the expected muscle strength mean value and corresponding standard deviation
- Afterwords application of 2 final procedures unifying results obtained from all 8 particular procedures mentioned above
- Testing
- Preliminary error analysis
- Critical review of the final results obtained, and decision made whether these results are precise enough to reliably determine patients condition.

INITIAL DATA CORRECTION FOR WEARINESS
LINEAR REGRESION FIT



where

$$t_{AV} = \frac{1}{m}\sum_{i=1}^{m} t_i\,, \qquad u_{AV} = \frac{1}{m}\sum_{i=1}^{m} u_i$$

$$a = \frac{\sum_{i=1}^{m} u_i(t_i - t_{AV})}{\sum_{i=1}^{m}(t_i - t_{AV})^2}\,, \qquad b = u_{AV}$$

correction rule

$$u_i^{corrected} = u_i^{original} - a(t_i - t_{AV})$$

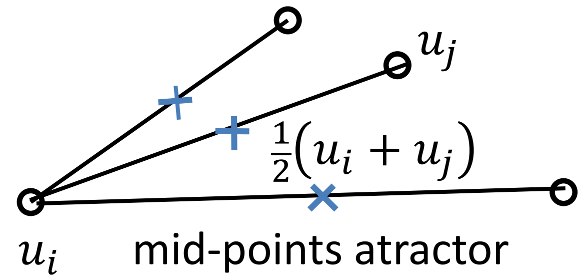
LIST OF FUNCTIONALS DEVELOPED

No	item	functional
1.	Simplest standard statistic analysis	$J(u(1),1)$
2.	Standard statistic analysis – linear regression	$J(u(2),1)$
3.	Standard statistic analysis with one outlier data removed	$J_1(u(3),1)$
4.	Standard statistic analysis with one outlier data removed after linear regresion	$J_2(u(3),1)$
5.	Functional with mid-point principle atraction complete version	$I(u,w1)$
6.	Functional with mid-point principle atraction simplyfied version	$I(u,w2)$
7.	Minimum data density principle atraction - version 1	$J(u,w1)$
8.	Minimum data density principle atraction - version 2	$J(u,w2)$
9.	Inequqlities unifying procedure	$J_1(u,\sigma)$
10.	All above unifying procedure	$J_2(u,\sigma)$

NEW WEIGHTED ERROR FUNCTIONALS AND THEIR NORMALIZATION – cont.

(i) MID-POINT CONCEPT COMPLETE FUNCTIONAL

$$I(u, w_i) = \sum_{i=j+1}^{m} \sum_{j=1}^{m-1} w_{ij} \left(u - \frac{u_i + u_j}{2} \right)^2$$



The functional normalization

by weighting factors

$$w_{ij} = \frac{(u_i - u_j)^{-2k}}{\sum_{n=l+1}^{m} \sum_{l=1}^{m-1} (u_n - u_l)^{-2k}} \quad \text{where} \quad \sum_{i=j+1}^{m} \sum_{j=1}^{m-1} w_{ij} = 1\,, k = 1$$

Expected mean value and the standard deviation

$$\bar{u} = \frac{1}{2} \sum_{i=j+1}^{m} \sum_{j=1}^{m-1} w_{ij} \, (u_i + u_j)$$

$$\sigma_I = \sqrt{I(\bar{u}, w)}$$

TYPICAL FORMULA STRUCTURE

Example

Corrected data $u_1 = 12.156, \; u_2 = 8.214, \; u_3 = 7.902, \; u_4 = 6.588$

Expected mean value

$$\bar{u} = \frac{\frac{6.588 + 7.902}{1.314^2} + \frac{7.902 + 8.214}{0.312^2} + \frac{8.214 + 12.156}{3.942^2} + \cdots}{2\left[\frac{1}{1.314^2} + \frac{1}{0.312^2} + \frac{1}{3.942^2} + \cdots\right]} = 8.027$$

$$I(\bar{u}, w) = \frac{\frac{(7.245 - 8.027)^2}{1.314^2} + \frac{(8.058 - 8.027)^2}{0.312^2} + \frac{(10.185 - 8.027)^2}{3.942^2} + \cdots}{\frac{1}{1.314^2} + \frac{1}{0.312^2} + \frac{1}{3.942^2} + \cdots} = 0.0608$$

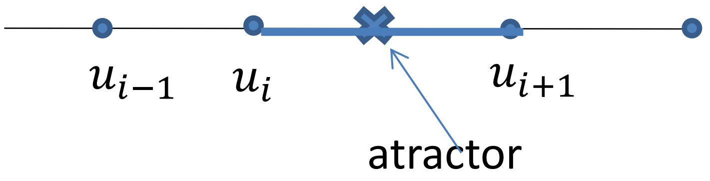
$$\sigma_I(\bar{u}) = \sqrt{I(\bar{u})} = 0.2466$$

MID-POINT CONCEPT - SIMPLIFIED FUNCTIONAL

When measured data is given in the growing order, in the simplified functional concept we replace complete formula by a simplified one

$$\frac{1}{2}(u_i + u_j) \Longrightarrow \frac{1}{2}(u_i + u_{i+1})$$

$$u_i - u_j \Longrightarrow u_i - u_{i+1}$$



Corresponding functional is denoted as $I(u, w_2)$

(ii) HIGH DENSITY CONCEPT WEIGHTED FUNCTIONAL (wariance)

$$J(u,w) = \sum_{i=1}^{m} w_i (u - u_i)^2$$ normalized by weighting factors

$$w_i = \frac{\varepsilon_i^{-2k}}{\sum_{j=1}^{m} \varepsilon_j^{-2k}}$$, where $\sum_{i=1}^{m} w_i = 1$

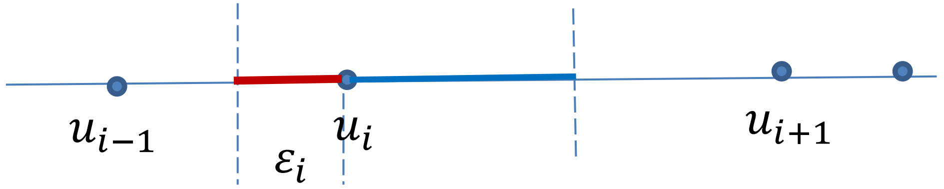
EXPECTED MEAN VALUE AND STANDARD DEVIATIONS

$$\bar{u} = \frac{\sum_{i=1}^{m} u_i \varepsilon_i^{-2k}}{\sum_{i=1}^{m} \varepsilon_i^{-2k}} \quad , \qquad \sigma = \sqrt{J(\bar{u}, w)}$$

The atractor is here measurements density. It may be defined in several ways by means of parameter ε_i

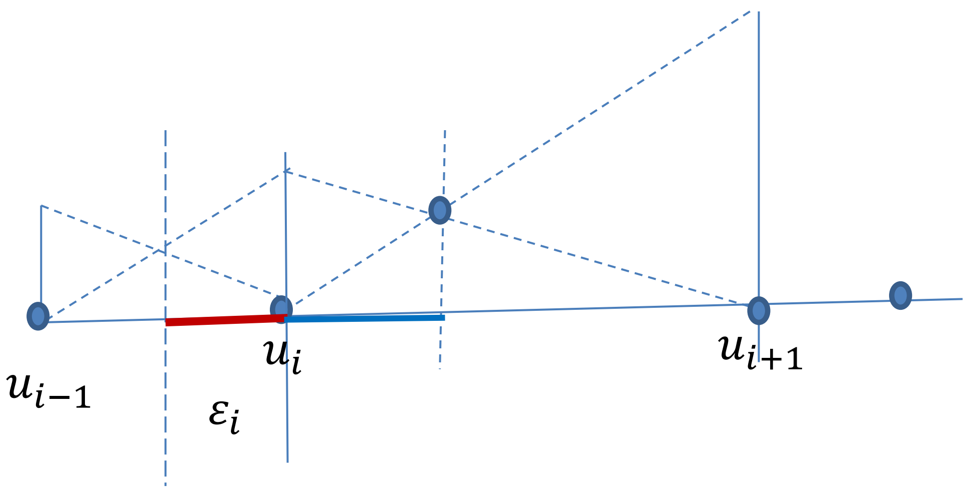
a) $\varepsilon_i = \frac{1}{2}\min\left[u_{i+1} - u_i, u_i - u_{i-1}\right]$ where $u_{j+1} \geq u_j, j = 1,2,.m$

ε_i presents here a halve of the shortest distance from u_i to the neigboring measurents



b) distance ε_i is based on the proportionality rule

$$\varepsilon_{ij} = \min\left[\frac{u_i - u_{i-1}}{u_i + u_{i-1}} u_i \;\; ; \quad \frac{u_{i+1} - u_i}{u_{i+1} + u_i} u_i\right]$$

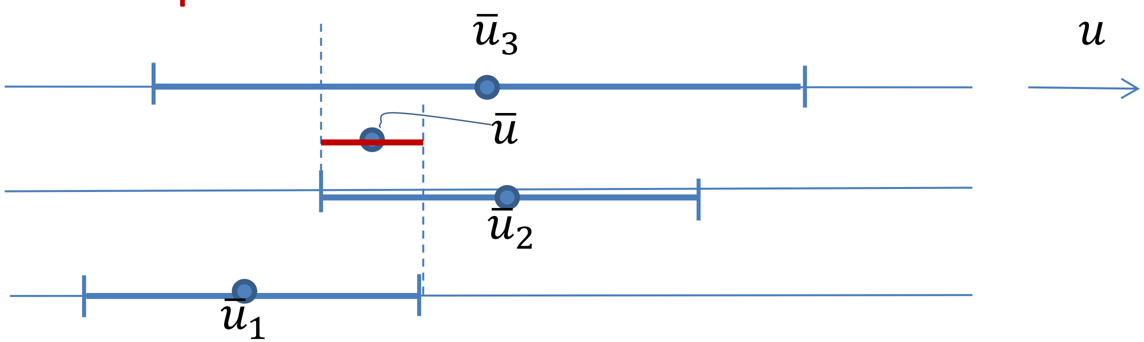


c) $\varepsilon_i = \sigma_i$

when the standard deviation σ is known the corresponding error functional is denoted as $J(u, \sigma)$ while $J(u, w_1)$, $J(u, w_2)$ are related to the cases a) and b) defined above

INEQUALITIES BASED UNIFYING PROCEDURE

Given are inequalities : confidence interval



$$\bar{u}_j - \sigma_j \leq \; u \leq \; \bar{u}_j + \sigma_j \quad , \quad j = 1,2,\ldots$$ confidence interval

Maximum lower bound — Minimum upper bound

$$\max_j(\bar{u}_j - \sigma_j) \quad \lesseqgtr ? \quad \min_j(\bar{u}_j + \sigma_j)$$

We introduce a correction factor α and require

$$\max_j \; (\bar{u}_{jmax} - \propto \sigma_{jmax}) \; = \; \min_j(\bar{u}_{jmin} + \propto \sigma_{jmin}) \; \Longrightarrow \; \alpha = \frac{\bar{u}_{jmax} - \bar{u}_{jmin}}{\sigma_{jmax} + \sigma_{jmin}}$$

hence

$$\propto = 0 \qquad \text{when} \qquad \bar{u}_{jmax} = \bar{u}_{jmin} \equiv \bar{u} \ ,$$

and

$$\sigma_{jmax} = \sigma_{jmin} \equiv \sigma$$

Otherwise the resultant expected value and standard deviation are

$$\bar{u} = \bar{u}_{jmax} - \propto \sigma_{jmax} = \bar{u}_{jmin} + \propto \sigma_{jmin}$$

$$\sigma = \sqrt{J(\bar{u}, w)}$$

where the normalized error functional

$$J(\bar{u}, w) = \left[\left(\frac{\bar{u}_{jmax} - \bar{u}}{\sigma_{jmax}} \right)^2 + \left(\frac{\bar{u}_{jmin} - \bar{u}}{\sigma_{jmin}} \right)^2 \right] \frac{1}{\frac{1}{\sigma_{jmax}^2} + \frac{1}{\sigma_{jmin}^2}}$$

GENERAL SUMMARIZING PROCEDURE UNIFYING RESULTS OF ALL PREVIOUS ONES

Given results of all other procedures used here as the pseudo experimental initial data

$(\bar{u}_j, \sigma_j)$ expected mean value, standard deviation

$j = 1,2, \ldots n, \quad n$ – number of procedures applied so far

Weighted functional, normalized by means weighting factors

$$J(u,\sigma) = \sum_{j=1}^{m} w_j \left(u - \bar{u}_j\right)^2, \text{ where } w_j = \frac{\sigma_j^{-2k}}{\sum_{i=1}^{m} \sigma_i^{-2k}}, \qquad \sum_{j=1}^{m} w_j = 1, \qquad k = 1$$

Resultant expected mean value and standard deviation

$$\bar{u} = \sum_{j=1}^{m} w_j \bar{u}_j \quad , \qquad \sigma = \sqrt{J(\bar{u}, \sigma_j)}$$

EXEMPLARY NUMERICAL ANALYSIS – A COMPARATIVE STUDY OF VARIOUS PROCEDURES

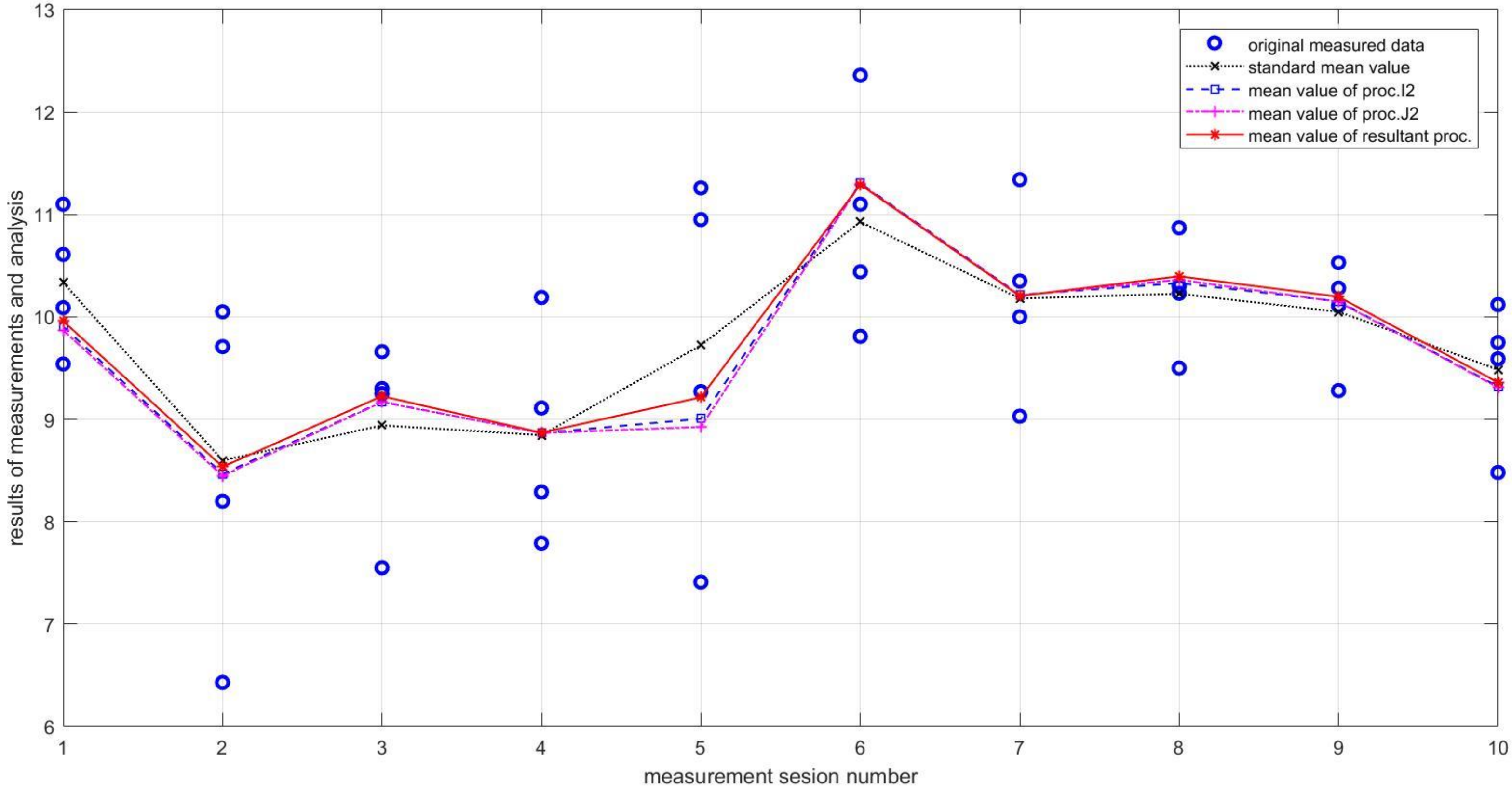
Characteristics of measured experimental data

LEG/MUSCLES	LEFT	RIGHT	BOTH
EXTENSOR	10x4	10x4	10x4
FLEXOR	10x4	10x4	10x4

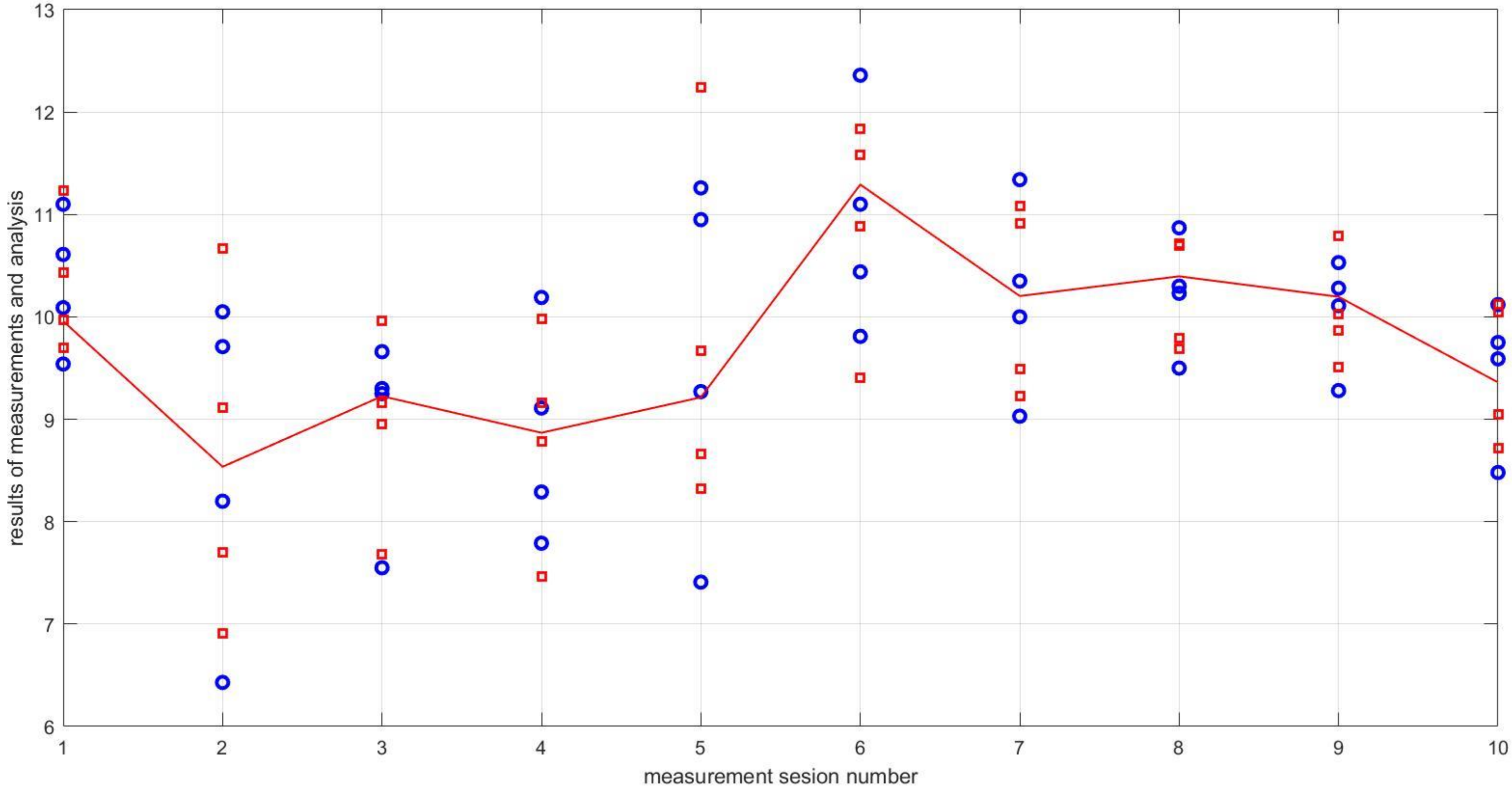
Problems tested

- Results of experiment repetition – best procedures?
- Procedures comparizon
- Noise level indicative results
- Influence of preliminary data correction
- Results precision – various error norms investigated

COMPARISON of EXPECTED MEAN VALUES FOR CHOSEN PROCEDURES LEFT FLEXOR

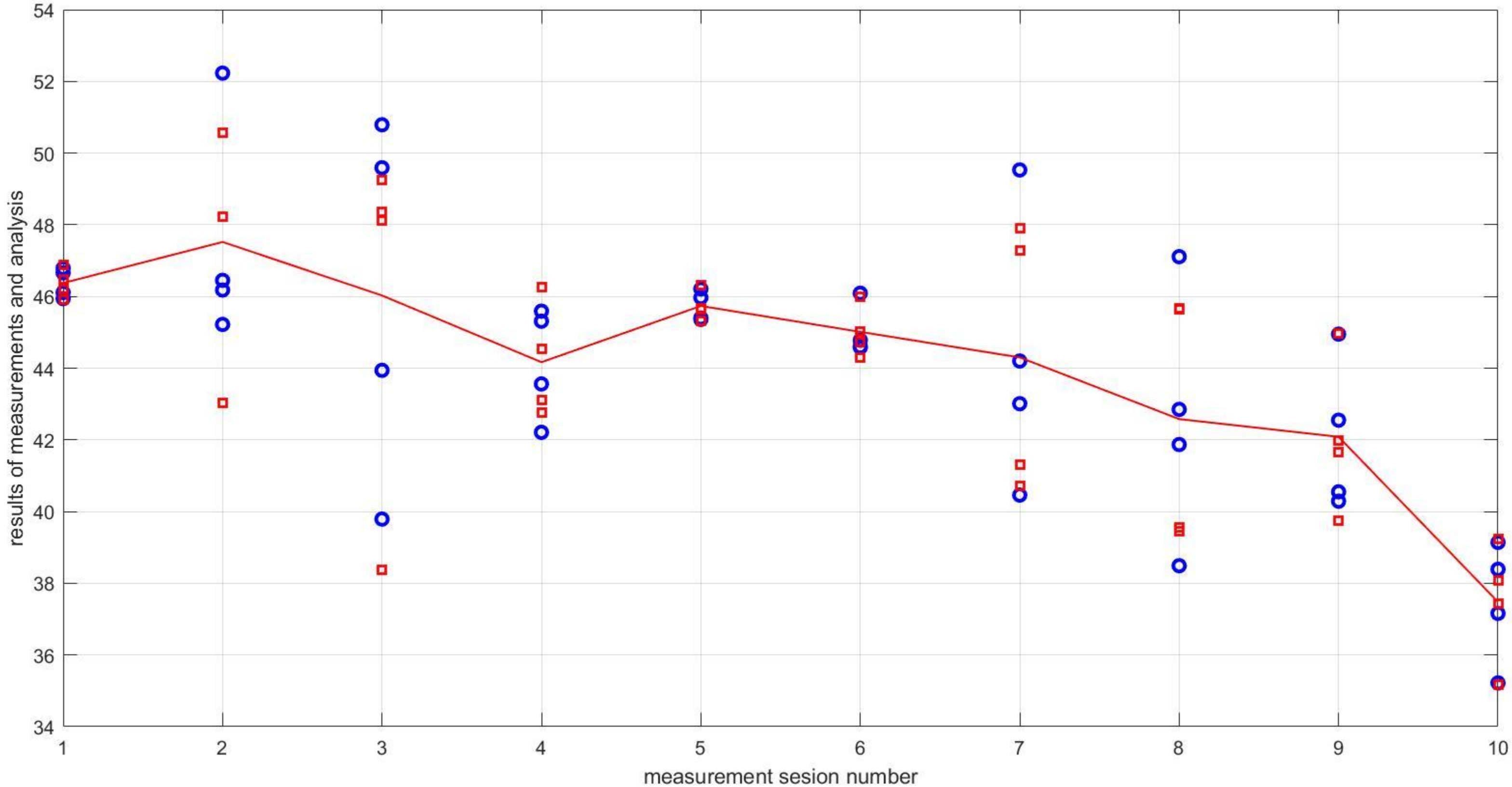


COMPARISON of ORIGINAL and CORRECTED DATA
LEFT FLEXOR

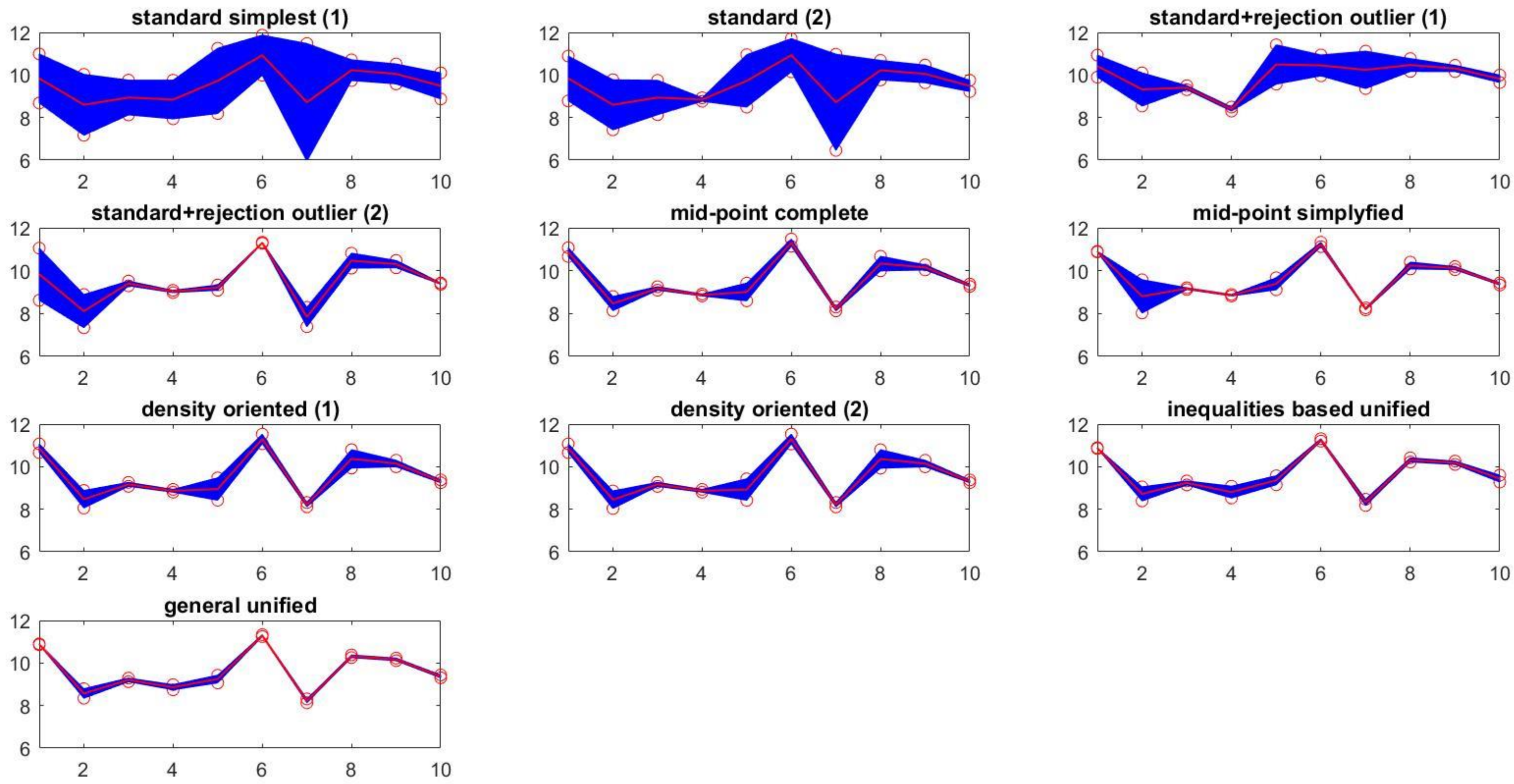


Original measured data   corrected data   mean value for resultant proc.

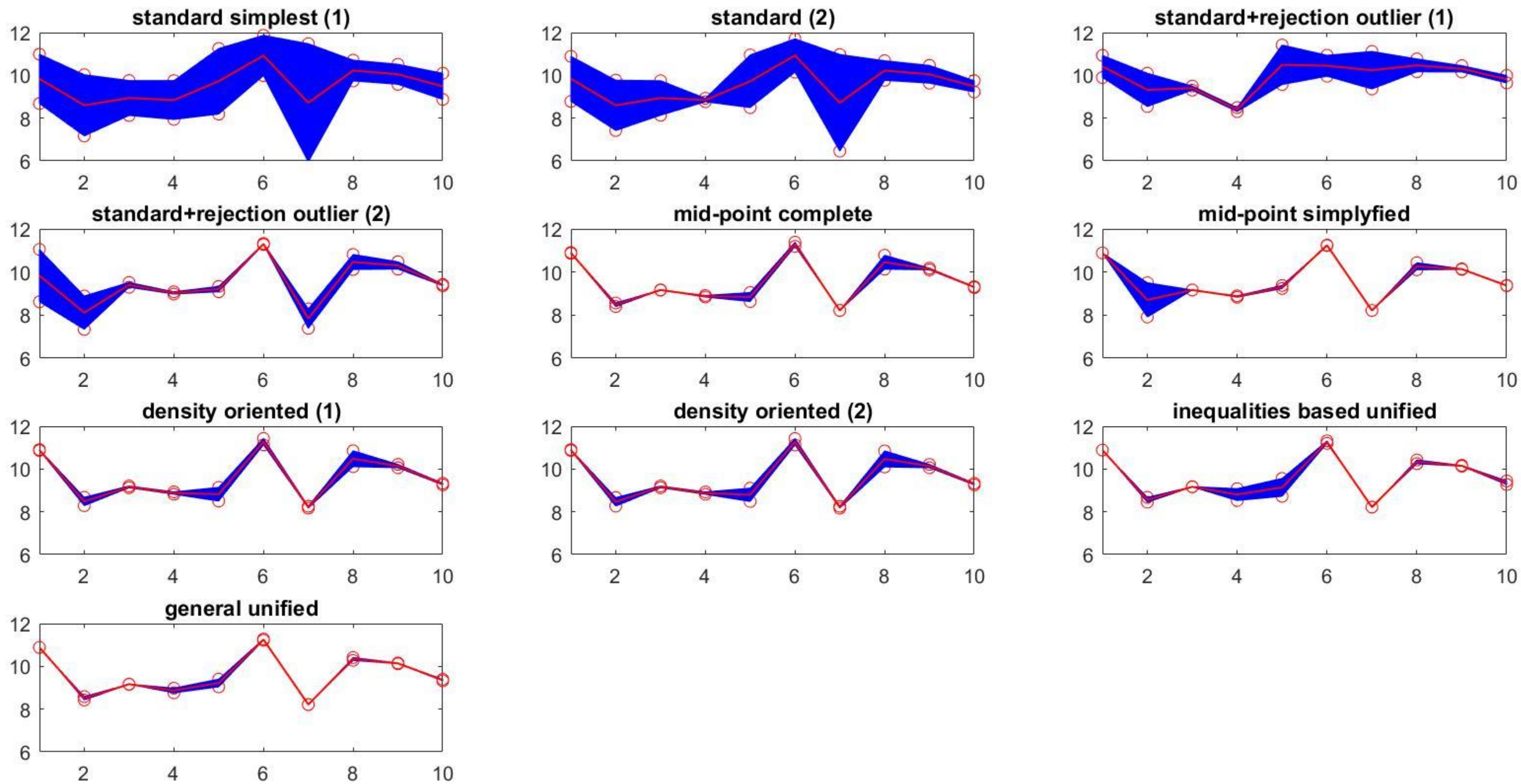
COMPARISON of ORIGINAL and CORRECTED DATA
LEFT and RIGHT EXTENSOR



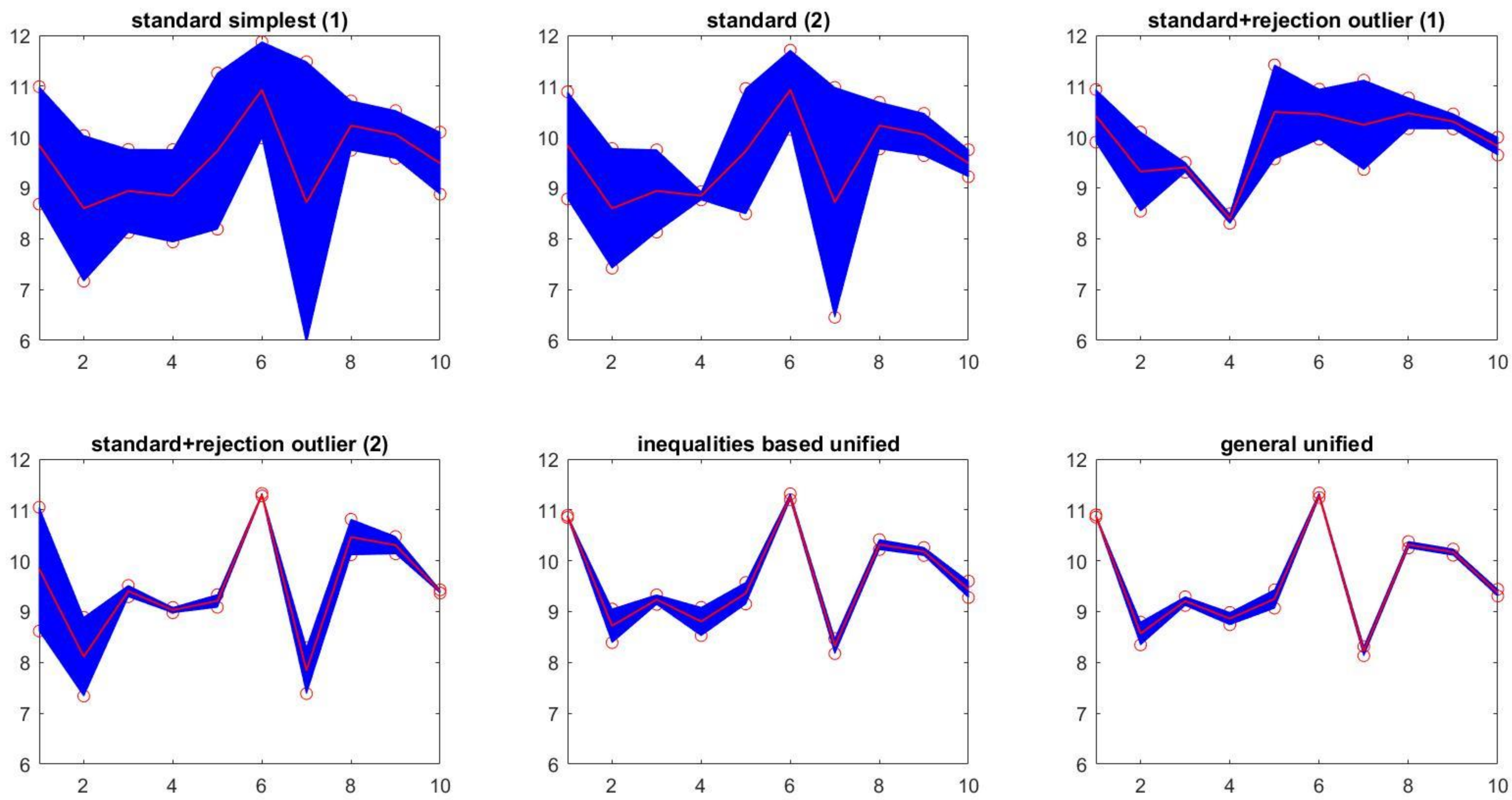
CONFIDENCE INTERVAL AND MEAN VALUE DISTRIBUTIONS
LEFT FLEXOR, k=1



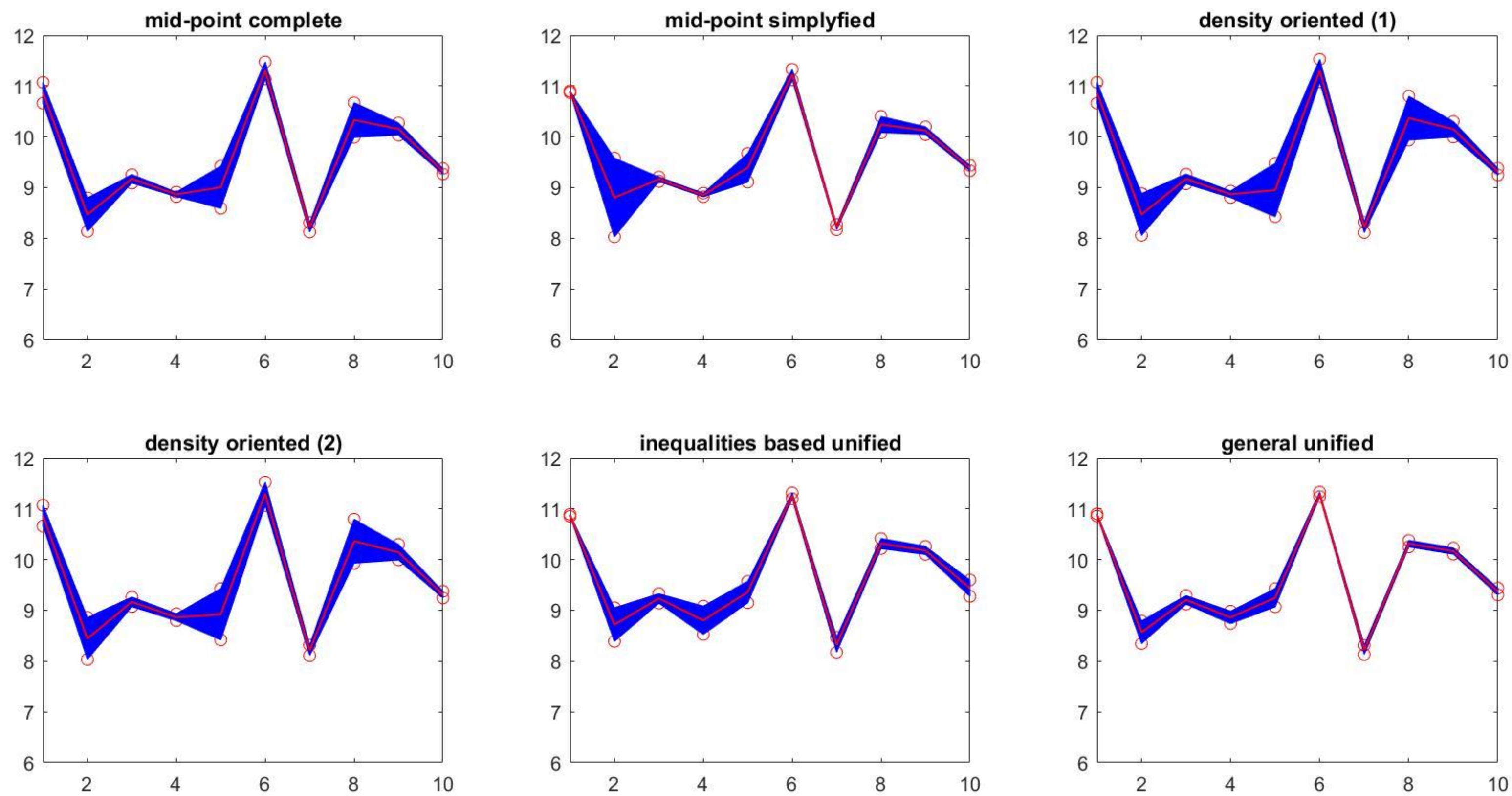
CONFIDENCE INTERVAL AND MEAN VALUE DISTRIBUTIONS
LEFT FLEXOR, k=2



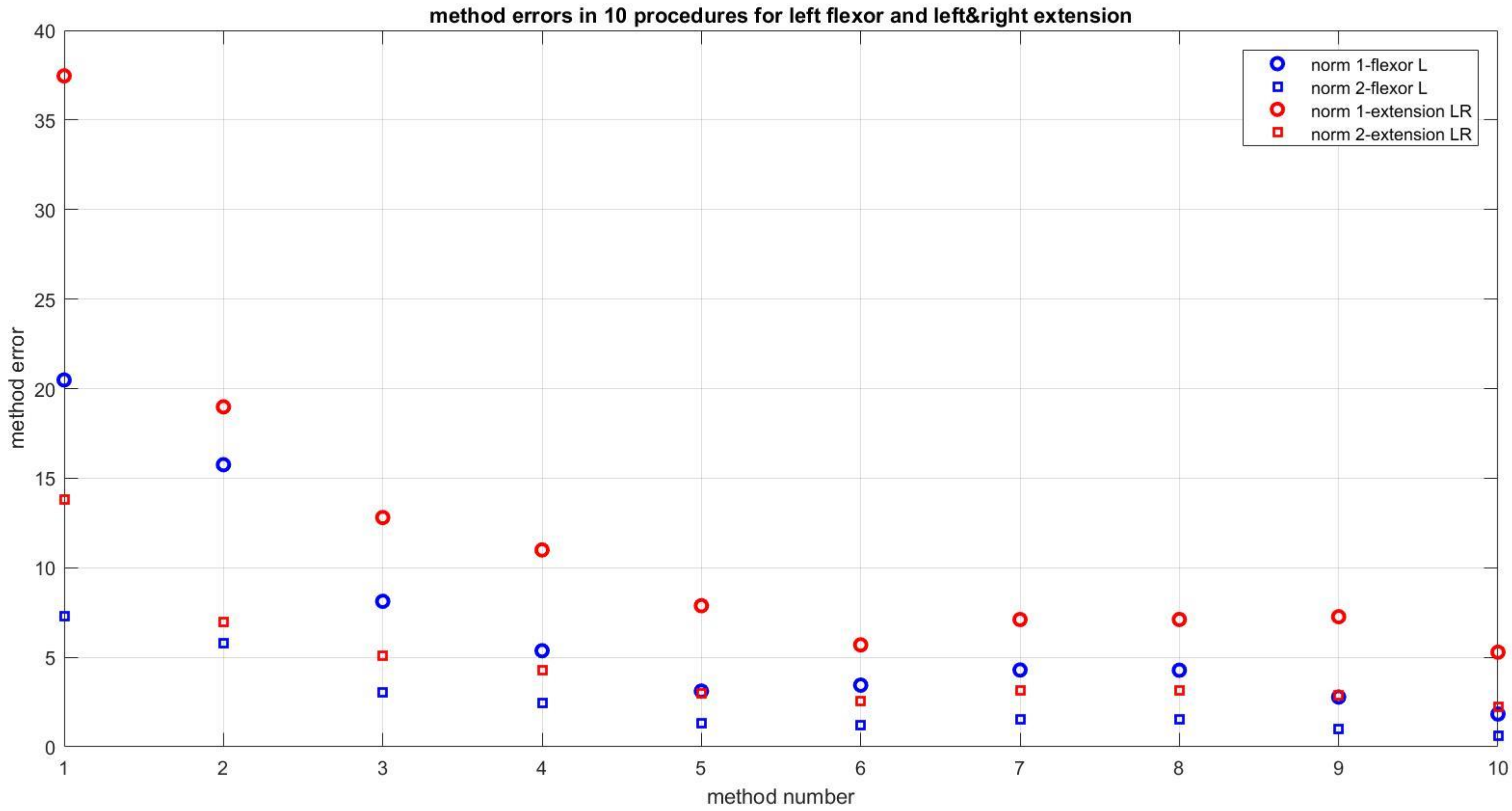
CONFIDENCE INTERVAL AND MEAN VALUE DISTRIBUTIONS
LEFT FLEXOR, k=1



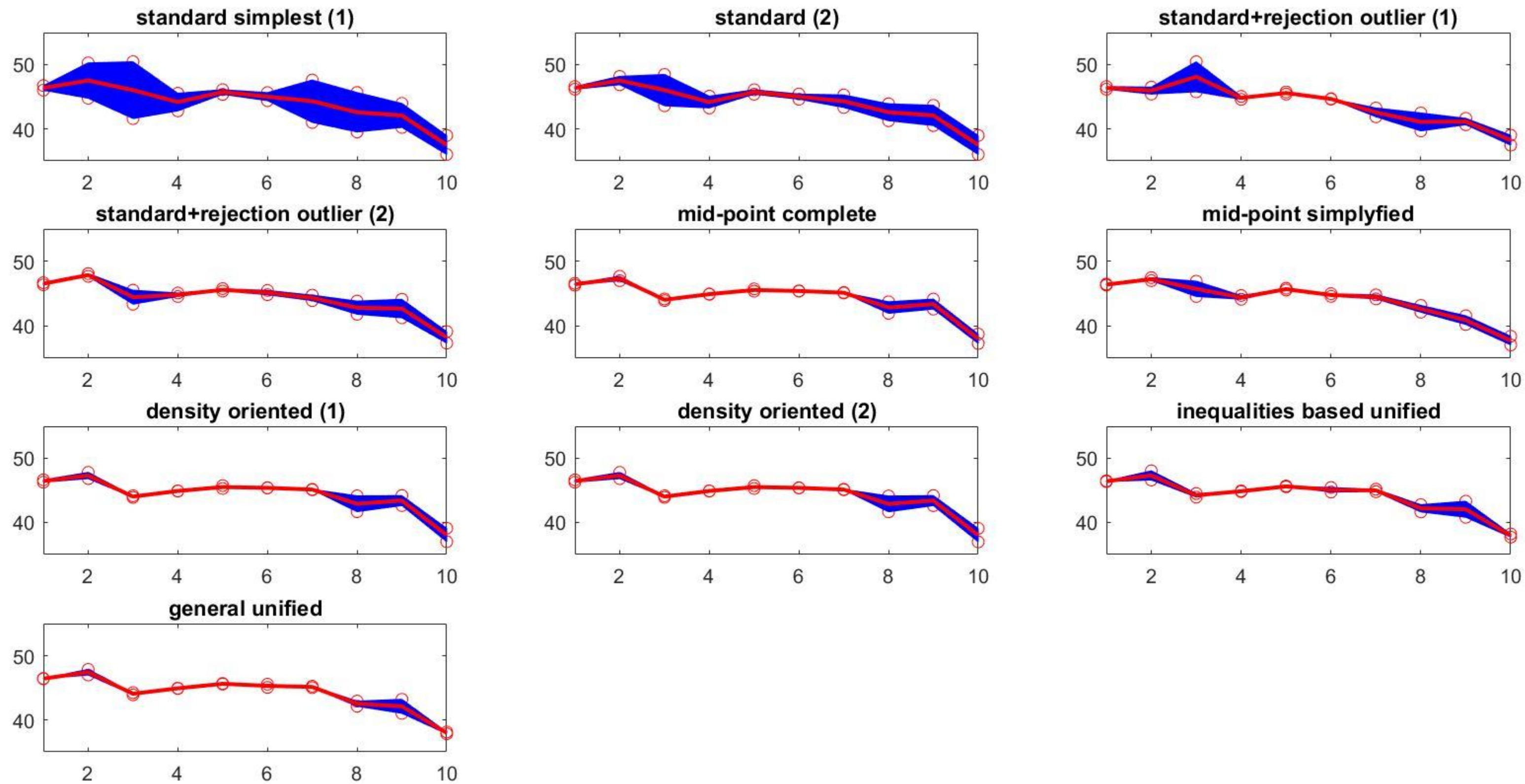
CONFIDENCE INTERVAL AND MEAN VALUE DISTRIBUTIONS
LEFT FLEXOR, k=1



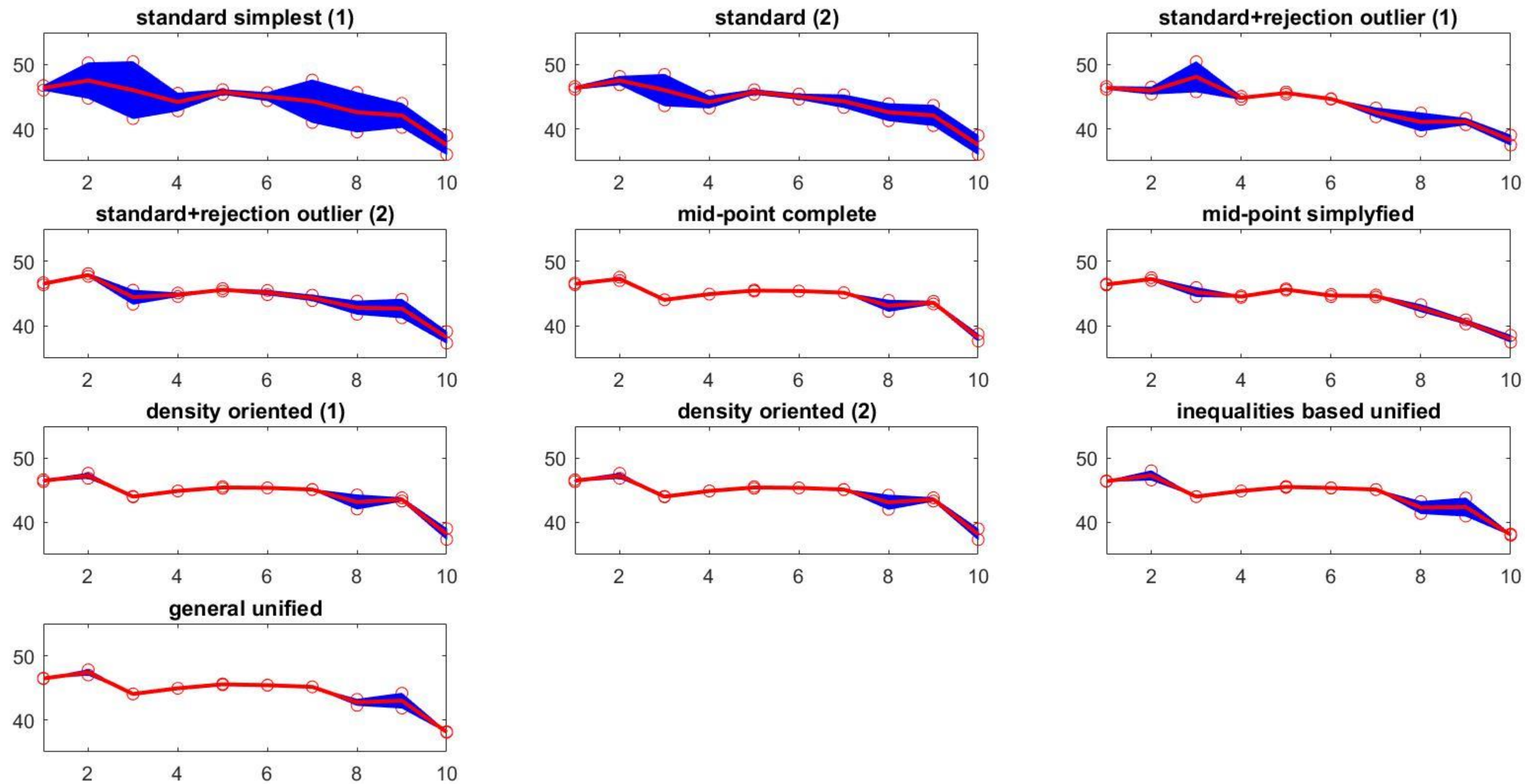
METHOD ERRORS in 10 PROCEDURES for LEFT FLEXOR and LEFT&RIGHT EXTENSOR



CONFIDENCE INTERVAL AND MEAN VALUE DISTRIBUTIONS
LEFT and RIGHT EXTENSOR, k=1



CONFIDENCE INTERVAL AND MEAN VALUE DISTRIBUTIONS
LEFT and RIGHT EXTENSOR, k=2



ON ERROR ANALYSIS: MEASUREMENTS AND NUMERICAL ANALYSIS

1		2		3		4
measured muscles strength	$\Rightarrow$	evaluated discrete u_i values data correction	$\Rightarrow$	specific solution procedures	$\Rightarrow$	resultant unifying procedures

$u_i(t)$, $i = 1,2, \ldots m \Longrightarrow$

- u_{imax}
- $u_{iAV} = \frac{1}{b-a}\int_a^b u_i(t)dt$

$\Big\} = u_i$

- $\to \bar{u}_j$ expected mean value
- $\to \sigma_j$ standard deviation

$\Big\} \sum_j \Longrightarrow$ final $\bar{u}$, final σ $\Big\} \Longrightarrow$ analysis

$\Longrightarrow$ analysis

- comparative study of solution procedures
- solution jumps
- error noise level

Errors investigation on subsequent stages of analysis

1 { measurements errors patient and device

2 { measurements and calculation errors

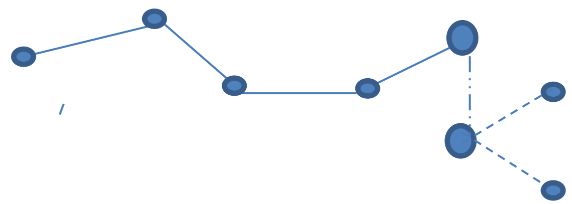
3 { solution procedure errors

4 { final unifying procedures errors

5 { post processing analysis

APPLICATION OF STATISTICAL ANALYSIS RESULTS

1. Verification of results repeatability
2. Investigation of single patient
3. Interpretation of analysis results

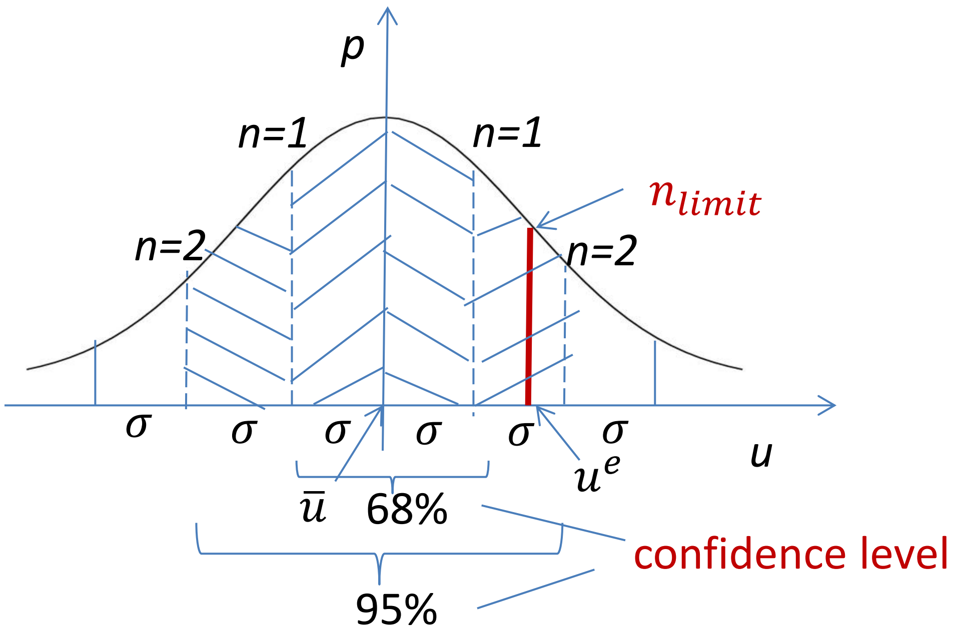


Gaussian probability density

$$p = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(u-\bar{u})^2}{2\sigma^2}\right)$$

confidence interval

$$\bar{u} - n\sigma \;\leq\; u \;\leq\; \bar{u} + n\sigma$$



assume n (mostly n = 2)

answer quations

- does measured data $u^e \in [\bar{u} - n\sigma, \bar{u} + n\sigma]$?
- which is confidence level limit for u^e ?

$$\left.\begin{array}{l} u^e > 0 \;\rightarrow\; \bar{u} + n_{limit}\sigma \;= u^e \\ u^e < 0 \;\rightarrow\; \bar{u} - n_{limit}\sigma \;= u^e \end{array}\right\} \Rightarrow n_{limit}$$

FINAL REMARKS

Done

- Considered was a heuristic rule based uncertainty approach to analysis of scarce and low precision experimental data. This approach was applied to evaluation of muscle strength using both results of appropriate measurements carried out on an adopted „ATLAS” device, and proposed innovative solution procedure
- Proposed and applied were
  - weighted normalized error functionals based on various heuristic rule principles
  - resulting specific solution procedures as well as summarizing ones
  - corresponding algorithms

- Generated was a relevant preliminary computer program
- A variety of tests were carried out including
  - Specific solution procedures verification and comparison
  - Preliminary error analysis

These tests showed good agreement of the solution procedures. Moreover clear precision advantage of the new procedures over the standard ones was noticed.

However, the general accuracy of the final results represented by the standard deviation magnitiude is not fully satisfactory as yet. Further standarization of data acquirement is needed including both the „ATLAS” measurement device, and patients position.

Further research

Both the method development, and application oriented research will be continued including:

- Investigation of the statistic noise limits (relation statistic noise – true changes?)
- Use in analysis rather average (in time) muscle strength u_{AV} than its maximum value u_{max}
- Search for the probability distribution possibly close to the true statistics of the muscle strength measurements (histogram)
- Error analysis
- A variety of tests the Student one including
- Interpretation of results obtained from this statistic analysis in the form of understandable communique needed by patients and their physiotherapists
- Measurements and analysis done as the regular service for patients
- Use results of this preliminary research for analysis of other similar type problems

THANK YOU FOR ATTENTION