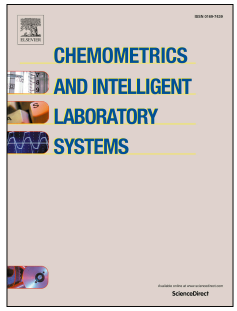
Journal Pre-proof



Concurrent multiresponse multifactorial screening of an electrodialysis process of polluted wastewater using robust non-linear Taguchi profiling

George J. Besseris

PII: S0169-7439(19)30362-4

DOI: https://doi.org/10.1016/j.chemolab.2020.103997

Reference: CHEMOM 103997

To appear in: *Chemometrics and Intelligent Laboratory Systems*

Received Date: 2 June 2019

Revised Date: 9 March 2020

Accepted Date: 11 March 2020

Please cite this article as: G.J. Besseris, Concurrent multiresponse multifactorial screening of an electrodialysis process of polluted wastewater using robust non-linear Taguchi profiling, *Chemometrics and Intelligent Laboratory Systems* (2020), doi: https://doi.org/10.1016/j.chemolab.2020.103997.

CReiT author statement

George Besseris: Conceptualization, Methodology, Data Analysis, Writing-Editing

Concurrent multiresponse multifactorial screening of an electrodialysis process of polluted wastewater using robust non-linear Taguchi profiling.

George J Besseris*

Mechanical Engineering Department, The University of West Attica, Egaleo, Attica, Greece

And

Advanced Industrial and Manufacturing Systems, Kingston University, London, UK.

*Email address: besseris@uniwa.gr

Abstract

Electrodialysis is an important chemical process that separates pollutants from wastewater pools to produce clean water for consumption and irrigation. Initial wastewater concentration of chemical elements always differs. Chemical components are strongly dependent on the efflux origin and treatment. To optimize an electrodialysis process is congruent to improved key water quality characteristics. To predict optimal electrodialysis performance there will always be a need to conduct a small number of structured experiments. This is because wastewater conditions are usually different in each situation thus requiring reliable evidence-based design decisions to be delivered timely and low-cost. We study a real example from crucial dessert wastewater operations that aim to supply clean water for irrigation. Several issues are scrutinized that are often overlooked when carrying out multi-response multi-factorial statistical optimization in environmental screening. Programming fast-cycle trials with Taguchi-type factorial recipes reaps quick information for new development and improvement projects. But it also introduces phenomena such as saturation, unreplication and non-linearity that could undermine the optimization effort. The showcased paradigm uses popular Taguchi methods to organize a rapid and short round of trials in order to investigate the behavior of four electrodialysis controlling factors: 1) the dilute flow, 2) the cathode flow, 3) the anode flow and 4) the voltage. The three monitored water quality indices are: 1) the percentage of removed sodium cations, 2) the sodium adsorption ratio and 3) the sodium ratio. We discuss the intricacies that emerge from the synthetic type of the electrodialysis data: non-normality, non-linearity and messiness. We propose a robust and agile method to conduct the multi-response multi-factorial optimization for electrodialysis of polluted wastewater. It is based on super-ranking and distribution-free profiling. Comparison with other profiling methods is provided and main advantages are commented from a chemical engineering perspective.

Keywords: Wastewater, electrodialysis, multi-response optimization, robust multi-factorial process profiling, non-linear non-normal data, data messiness.

1. Introduction

Water is the ultimate commodity on this planet since life without water cannot exist. It is oxymoron that while water covers 70% of the earth's surface, a scarce portion of less than 1% is only available for human consumption. This minimal amount is forecasted not to be sustainable to quench the household and farming requirements for a rapidly growing human population in the near future [1, 2]. Population migration trends away from arid areas would exacerbate the problem as the planet eco-system warms-up. However, sprouting water-treatment technologies seem to promise opportunities for broader accessibility to potable water [3, 4]. Brackish and saline sea water sources are considered for immediate exploitation but wastewater supplies are also not to be overlooked. Desalination and water recycling are at the forefront of treatment options for both purposes: 1) to store drinkable water and 2) to irrigate farms [5, 6]. Roughly three quarters of the distributed water is directed to farming. It is foreseeable then that large-scale irrigation operations should draw more engineering attention. There are quite a few engineering options that might be attuned to supply enough water to agricultural land [7]. Highest priority projects to water accessibility are those that align toward reaching 'Goal 6' of the United Nations Sustainable Development [48] congruent to disadvantaged human ecosystems. Automatically, in an upward chain-reaction fashion, accomplishing 'Goal 6' directly aids in attaining 'Goal 2' (zero hunger) [49] by also offering opportunities to cultivate arid land and consequently inching closer to 'Goal 1' (Poverty termination) [50]. Engineering solutions based on membrane separation technology – forward or reverse osmosis – seem to be more frequent but their high cost of ownership has not established them as a universal cure-all [8]. In particular, when the feed source is wastewater, issues of biofouling and chemical element adjustment need to be addressed, making reverse osmosis rather an expensive alternative and suitable only for high added-value cultivations. Recently, electrodialysis has been studied as a potentially useful option to treat drainage wastewater and other polluted water bodies for large-scale planting [7]. Developing farming conditions in semi-arid or arid areas around the planet is indispensable. Electrodialysis could aid in this direction by toning down sodium content while balancing soil minerals - calcium, potassium and magnesium - to favorable concentrations for plant growth. For arable crops, readily available soil potassium correlates positively with yield [9]. Moreover, electrodialysis (ED) of wastewater could control outflow water potassium content such that to facilitate the compensation of leached sandy soils, especially when such soils are comprised of little clay and organic matter. A recent study by Abou-Shady [7] brought up the idea of upscaling the ED-tuning of wastewater reserves to right-balancing irrigation water constituents. It was demonstrated how to manipulate four specific ED-process factors in order to promote optimal salinity in complex futuristic large-scale irrigation projects. At the core of that research stands out a key recommendation for the effective use of water qualimetrics ('aquametrics') [10] that utilize Taguchi-type screening techniques [11].

Taguchi-type design of experiments (DOE) methods is useful for quick-and-economical, environmentally-friendly, evidence-based screening as well as optimization studies [12-14]. In its backbone, it is the 'lean-and-agile' philosophy that has been applied successfully in designing and improving intricate manufacturing processes. It is 'lean' because minimizes wasted materials, energy, equipment-availability and man-hours that are required for large-scale industrial trials. It is 'agile' because it adapts quickly to the operational demands where Taguchi-methods need to be deployed, thus exploiting any opportunity for rapid discovery. Hidden 'lean-and-agile' benefits are also to be reaped indirectly by halving the total experimental effort and duration of the two typical and sequential trial phases; *factor screening* and *parameter design* are to be conducted in a single concurrent step [11]. Screening experiments are characterization experiments that require two distinct sequential steps: 1) factor profiling and 2) identification of the strong factors. In the profiling step, the screening dataset is processed in order to quantify - in statistical terms - all factorial influences against one or more characteristics. Once the treatment effects have been quantified, then, the strong influences are selected based on a statistical rule. A statistical processor is used to determine those effects that are greater than a critical value; the one-sided cut-off value corresponds to a preset significance level, α. The identification process is an optimization step because it involves a uni-directional search to locate and select out the strong effect(s), i.e. those effects that perform below a minimum statistical significance constraint. Identification follows the general optimization process that given a set **A** of k effects a^i $(1 \leq i \leq k)$ $\forall$ $a^i \in \Re$, and a function f: $\mathbf{A} \rightarrow \mathbf{p}$ with statistical significance $p_i \in \mathbf{p}\ \forall\ p_i \in \Re$, we seek a subset $\mathbf{x_o} \subseteq \mathbf{A}$ such that $f(\mathbf{x}_0) \leq f(\mathbf{x})$ for all $\mathbf{x} \in A$ subject to the constraint $p_i < \alpha$. Thus, the screening phase leads to a reduction of the initial group of factors. Strong effects are considered for the next phase, which is the parameter optimization. Screening may reduce significantly the amount of experimental work that is to be forwarded to the parameter design phase. But chemical screening is a cost driver that intelligent discovery systems seek to minimize by emphasizing rapid cycle times [15]. Parameter design refines the strong factors that precipitated from the screening phase such that to optimally predict one or more product or process response(s) [16]. Obviously, this tactic of 'two-in-one' in Taguchi's strategy shortens the overall optimization study cycle while lowering materials and energy consumption [12]. There are two economic gains then, one from curtailing trial-related costs and another from making an optimal product that generates less waste while completed in reduced cycle times.

Deeper environmental awareness is to be envisaged in chemical processes. Taguchi methods have been implemented to optimize wastewater treatment with reverse osmosis and to recover heavy metals for quite some time [17, 18]. They have been employed to investigate even difficult wastewater treatment cases where there was a need for improving the conditions of a coagulation-flocculation process [19]. Ramping-up processing efficiencies with characterized flocs may also be achieved with Taguchi DOE techniques targeting harsh agro-industrial wastewater treatments [20]. Desalination filtering operations are amenable to Taguchi-type screening and optimization when using modern carbon nanotube membranes [21]. When complex datasets are collected to optimize a forward osmosis process, a combination of Taguchi-type tools and neural networks have proved to be effective [22]. The Taguchi toolbox has been applied successfully in upscaled Fenton-SBR industrial operations that produced wastewater from bamboo treatment [23]. In chemometrics, the classical Taguchi method has been entrusted in optimizing measurement accuracy of UPLC isocyanate [24], optimal mixture settings for enhancing concrete properties [25], Diazinon cloud point extraction [26], and optimized multianalyte determination with biosensors [27].

Technically, Taguchi's DOE methodology in aquametrics is achieved on two ends. At the frontend, Taguchi methods demand small but structured trials. For this to happen, the DOE framework needs to obey a few predetermined factorial recipes. The experimental recipes rely on the combinatorial rules of fractional factorial designs (FFDs) [28]. The particular Taguchi-type FFD plans belong to the family of orthogonal arrays (OAs) [11]. At the backend, Taguchi methods institute two utilities: 1) the use of the signal-to-noise ratio (SNR) concept in order to compress the collected dataset streaks and 2) the standard deployment of the analysis of variance (ANOVA) to relay statistical significance to the strength of the examined effects. Maximum utilization of the frontend capabilities occurs when a selected OA trial-plan is saturated with tested controlling factors [28]. Saturation locks the requirement for minimum number of experimental runs with respect to the number of the investigated effects. Saturation maximizes the number of effects that are allowed to deliver information given a data-collection OA-plan. To illustrate the importance and ramifications of these aquametrics concepts in screening and optimizing wastewater treatment, in this work, we will take up the interesting four-factor three-response ED-process optimization paradigm of Abou-Shady [7]. We contemplate that it is a unique case as we will explain along because of the nature and the relationships among the selected water characteristics. We will not work out one 'response-at-a-time' as it is common in most wastewater treatment studies that employed Taguchi optimization. Instead, it might be useful to generalize the feasibility of the study to a more pragmatic rationale by attempting a concurrent multi-response optimization. The suggested frontend design ($L_9(3^4)$ Taguchi-type OA) in Abou-Shady's experiments was saturated [7]. It was selected such that to simultaneously screen, optimize and track down the potential influence of non-linearity for each of the tested effects. At saturation point, the constraint for the minimum number of required experiments is $n=(2 \cdot m)+1$; n is the number of trials and m is the number of the examined effects. Furthermore, the experimental design by Abou-Shady [7] featured still another property conducive to rapid, economical and lean-and-green data-generation; experimental recipes were not replicated. By undertaking an unreplicated [29] and saturated OA-scheme, the collected data was ensured to be delivered in low cost, fast turnaround time and minimum material/energy losses [30]. Unfortunately, when designing processes or products by exploiting synchronously the profitable conditions of saturation and unreplication, frontend and backend synchronicity is bound to break down in Taguchi methods. This is because the simultaneous presence of the two conditions eliminates the chance to obtain an estimate for the residual error in ANOVA, since no degrees of freedom for the error are left over [31]. Hence, no statistical inference is possible with ordinary means and no objective sizing of the effects is feasible in such an occurrence. Generally speaking, the "unreplication" condition is inherent to Taguchi methods. The prescribed SNR transformation step will always convert even replicated data to an "unreplicated response" vector form [11, 32, 33]. Undisputedly, it was recognized that the analysis of the unreplicated factorial experiments was instrumental in discovering in short time those effects that were to play a role behind an intricate landscape in industrial operations [51]. The accompanying comparative study of as many as twenty-four methods attested to such need while concluding to no single 'all-purpose' front-runner approach [51]. It definitely encouraged the development of new techniques. Recently, an important study tested leading unreplicated factorial solvers - part of modules of several mainstream software packages [52]; it indicated that benchmarked predictions varied significantly among packages. This justifies the impetus for proposing new unreplicated factorial solvers with robust capabilities. It is a main motivation point for our study. It is the "saturation" condition that may be construed as optional but encouraged from an engineering perspective due to optimal data utilization. As perplexing as it sounds, statistical profiling of an unreplicated-saturated OA-dataset may still be accomplished with specialized handling and data manipulation. Irrespective of the setbacks that may be lurking in interpreting regular Taguchi-type optimization studies [34, 35], successful wastewater research has been published as discussed previously, and clearly attesting to that this subject is in demand. One of the purposes of this work is to explore complications on the way to achieving optimal ED-process performance through multi-response multi-factorial non-linear screening/optimization aquametrics [36-38]. Hopefully, some aspects that will be discussed may lay ground for robust and agile ED-process predictions [39, 40]. We show how to upgrade the Taguchi analysis for unreplicated-saturated OA ED-trials such that to transcend from the subjective limitations of descriptive statistics to meaningful inferences.

The motivation for the selection of the exploratory desert development project [7] to be re-examined in this work becomes more transparent now. Its principal outlook aligns in accord toward to the general 'Goal 6' of the United Nations Sustainable Development [48]. The study by Abou-Shady [7] is unique because it seeks to optimize three different characteristics that all pertain to the behavior of suspended sodium in the feed wastewater. The concurrent screening/optimization of sodium content in three different chemometrical landscapes has not been undertaken before. One characteristic is the percentage of removed sodium cations (ReNa). It characterizes the *electrodialysis process* itself in a given time interval. It is dependent on the initial sodium cation concentration. It is a process quality index that tracks ED cell performance. The larger the value of the percentage of removed sodium the higher the effectiveness of the ED unit. Being a percentage-based response requires non-conventional handling. Data types in percentage form are distinct for their inherent poor additivity properties in practical situations [41]. This is because intermediate arithmetical operations with percentages are not permitted to exceed the two realistic bounds (0% and 100%). In this specific situation, the SNR transformation [42] is not an appropriate data compressor to be used as in the small and dense dataset of Abou-Shady [7]. Instead, the omega (Ω) conversion method is usually recommended which replaces the quadratic loss in the SNR with the odds ratio in Ω [41]. The formula for Ω then becomes:

$$\Omega(\mathrm{db}) = 10 \log\left(p/(1-p)\right) \text{ with } \quad 0 < p < 1 \tag{1}$$

Nevertheless, it is known that the omega function is conditionally applicable. This is because the Ω value tends to infinity if measurements approach either of the two bounds. Another noteworthy issue is that for classical SNR transformations (Taguchi-type) to be meaningful, the original (raw) dataset must: 1) be in replicated form and 2) obey normality. For each executed experimental OA-recipe, at least two replicates are necessary to recover a signal (average estimation) and a noise (variability estimation). These two critical conditions are absent in the experimental design of Abou-Shady [7]. It is a main motivation of this work to show how one might circumvent this quandary by proposing an alternative approach which relies on distribution-free statistics. The proposed approach offers simplicity, transparency, robustness and agility in the optimization cycle. Thus, the aim is to aid in deciphering complex, small and dense DOE datasets in ED-optimization studies. In turn, analysis results are pivotal to reliable decision-making for large-scale chemical operations.

The second characteristic is the sodium adsorption ratio (SAR). SAR is a water quality trait that quantifies the water suitability which is intended for crop irrigation. Even though it is a single index, SAR delivers complex and crucial information. SAR monitors the soil flocculation status by measuring the balancing act of the soil conditioners. Both, flocculation inhibitors (sodium cations) and promoters (calcium and magnesium ions) tweak soil permeability and hence the water infiltration rate. Furthermore, SAR tracks the aqueous colloid suspension stability status. It is also a standard reliability measure since it diagnoses the sodicity hazard for a farmland. Irrigation water quality is optimal when SAR is minimized and there is a critical value for flocculation.

$$SAR = \frac{Na^{+}}{\sqrt{\frac{1}{2}\left(Ca^{2+} + Mg^{2+}\right)}} \qquad (2)$$

SAR is a ratio quantity. Thus, the discussion regarding the appropriateness of Ω over SNR in ReNa response data above is also pertinent here. It is remarked that SAR is a product (outflow water) characteristic as opposed to ReNa. Also, SAR and ReNa follow opposite response directions in an optimization exercise; the former is minimized the latter maximized.

The salinity status of the electrodialyzed outflow water may also be expanded to account for all four key (monovalent and divalent) cations. The competing potassium content is thus added in determining the sodium ratio (Na^+ ratio):

$$Na^{+}\ \text{ratio} = \frac{Na^{+}}{\left(Na^{+} + K^{+} + Ca^{2+} + Mg^{2+}\right)} \qquad (3)$$

The Na^+ ratio (NaRa) is also a percentage-based quantity that is sought to be minimized just as the SAR. The previous arguments about recommending the Ω-conversion over the SNR-transformation are maintained for this index, too. Similar to the SAR applicability, NaRa reflects water product quality. In the original optimization scheme [7], the 'smaller-is-better' expression for the Taguchi-defined SNR data conversion was used for all three examined characteristics (ReNa, SAR, NaRa):

$$\mathrm{SNR} = -10log_{10}\left[\left(\sum_{i=1}^{n}\frac{1}{y_i^2}\right)/n\right] \tag{4}$$

However, ReNa is a 'larger-is-better' characteristic and thus the proper Taguchi-defined expression for the SNR data conversion would have been instead:

$$\mathrm{SNR} = -10log_{10}\left[\left(\sum_{i=1}^{n} y_i^2\right)/n\right] \tag{5}$$

This article is structured as follows. A methodology is proposed to formulate the concurrent multi-response screening/optimization aquametrics for a wastewater ED-process. The methodology also extends the potential for "two-in-one" combo-solution which is inspired by Taguchi's main DOE planning strategy but which is restricted to a single response case. This means that it is pursued besides a concurrent multi-response multi-factorial screening-and-optimization solution, the possibility of the influence of non-linearity in the examined effects. On this endeavor, we directly deal with conditions of data "messiness" [43] that Taguchi methods are not tuned to handle. Abou-Shady's [7] experimental work demonstrated the natural emergence of messiness in the Taguchi-DOE ED-trials. We show the implications of messiness in the Results section where we use modern data fusion techniques for unreplicated-saturated Taguchi-type OA datasets. Messiness brings out the realistic demands in predictions where the phenomena may not really be canned in some parametric modelling [44]. Therefore, we demonstrate the robust, lean and agile screening prediction of ED performance based on ReNA, SAR and NaRa indicators against the four relevant controlling factors [45]. The justification for this new proposal, when compared to the other available published techniques on the subject of the unreplicated factorial analysis [51], relies on accumulating several tangible traits that are not found in the previous approaches. Key features are:

the new technique converts constant-free (non-subjectively) and distribution-free (robustly) unreplicated dataset predictions even at the limiting condition of saturation. The former feature is an advantage over the Lenth test [53] and the latter over the half-normal test [47]; they are the two premier tests with great representation in most commercial statistical software packages. But the primary advantage of the new method is that it also delivers distribution-free statistical significance for multi-response unreplicated-saturated OA-datasets. This is in contrast to the leading alternative method, the desirability analysis [54], which instead provides a score estimation in lieu of a significance measure, which would be based on a statistical reference law. Furthermore, in comparison to the desirability analysis, the new method eliminates the (manual) trial-and-error search – a subjective step – which the desirability optimizer depends upon to generate a solution. This would mean discovering an appropriate set of weights to parametrize each of the partaking response functions before succeeding to compute their composite desirability score.

The new method does not involve regression coefficients, hence it is computationally simpler. Therefore, it produces no residuals. Moreover, residuals are extremely sensitive to outliers. If appearing in small datasets, outliers become particularly risky. Still, residual analysis requires inspecting for independence of errors (autocorrelation effect) which is a non-relative condition in the proposed method. Since regression methods implicate mean estimators during the data fitting process, the characteristic 0% breakdown point of the method is ominously present. On the other hand, our method uses rank-sums (median estimator), which protects the data reduction process to a breakdown point of 50% [55]; it provides the maximum possible protection as a clear technique advantage. Summarizing, our technique is simpler and more robust from other alternative profilers/optimizers; this might be a desirable amelioration according to the Occam's Razor principle.

2. Methodology

2.1 A brief description of the wastewater electrodialysis experiments

Abou-Shady's design selection is a saturated (three-level) non-linear Taguchi-type ($L_9(3^4)$) orthogonal array [7]. The unreplicated-saturated $L_9(3^4)$ OA has been featured as a preferred trial planner in non-linear screening/optimization in diverse areas of studies that involve complex chemometrics [56-70]. Parenthetically, the methodology is construed to be extended for non-linear effects also tested in four or higher settings [71-74]. Selecting a four-setting or higher OA design usually attempts to ensure that curvature tendencies will be probed more intensely by inquiring information from one or more additional observations properly located between the two operating end-points.

Due to realistic constraints on time and resources, the design was reasonably decided to be carried out once for each setup recipe. In its saturated and unreplicated form, the design prescribed the maximization of resource utilization, the minimization of trial costs and thus overall accelerated the experimental process. The experimental plan engaged four controlling factors: 1) dilute flow (DF), 2) cathode flow (CF), 3) anode flow (AF) and 4) Voltage (V). The completed $L_9(3^4)$ OA rubric with the associated factor-setting loadings are tabulated in Tables 3 and 4 in ref. [7]. They are assorted with the three-way synchronous response data for ReNa, SAR and NaRa. The three response vectors contain information about the non-linearity and normality of the four effects. At this stage, the analysis by Abou-Shady [7] proceeded by considering the inner workings for one characteristic at a time. The fact that the design was saturated could not permit a formal application of ANOVA. Therefore, the results were unavoidably discussed in an exploratory manner. However, to instill vigor in the analysis, the three characteristics will need to be processed concurrently in one-pass simultaneously to gage factor strength, non-linearity and optimal effect adjustment.

2.2 Preliminary data analysis

A preliminary data analysis is required to test potential correlations among the three responses. If there is a correlation in some of the responses then it is plausible that they could be eliminated from further consideration. Since OAs generate small data, the best tactic is to pick and correlate two responses at a time using linear regression analysis. It is important to provide 95% confidence intervals of the line fittings such that a spread uniformity of the data-points could be inspected. We test this using the linear regression module of MINITAB 17.1.

Equally important is to have a view of how the location and dispersion interplay simultaneously fare for each group of data per factor setting. To obtain a 'location-and-dispersion' screening for the various effects, the best way is to pin up all individual factor-setting boxplots on a "clothesline". Box plots with median confidence intervals should also be drawn for each response separately such that to gauge the overall behavior of their central tendency and spread around the median. In both situations, the plots are easily constructed using the graph module for boxplots in MINITAB 17.1.

2.3 Setting up the saturated-unreplicated OA for distribution-free super-ranking

We consider the minimum-size non-linear (three-level) Taguchi-type $L_n(3^m)$ OA [11] with the imposed condition for unreplication and saturation such that $n=2m+1$; n is the number of experimental recipes and m is the number of the examined effects which are labeled as: $X_1, X_2,\ldots,X_m$. Then, their respective predetermined settings on an (i,j) OA arrangement may be written as $x_{1i}, x_{2i}, \ldots, x_{mi}$ ($i=1, 2\ldots n$). The output from executing the n recipes is a group of a total of r (unreplicated) responses: $R_1, R_2,\ldots,R_r$. The vector elements for each response may be symbolized as: $r_{1j}, r_{2j}\ldots r_{rj}$ ($j=1,2,\ldots,n$). A comprehensive depiction of the relevant input/output OA arrangement where the factor settings (input) and the response vector group (output) are positioned on the left- (input) and right-hand side of the design, respectively, follows as:

$$\begin{pmatrix} x_{11} & x_{21} & \cdots & x_{m1} \\ x_{12} & x_{22} & \cdots & x_{m2} \\ \vdots & \vdots & \cdots & \vdots \\ x_{1n} & x_{2n} & \cdots & x_{mn} \end{pmatrix} \begin{pmatrix} r_{11} \\ r_{12} \\ \vdots \\ r_{1n} \end{pmatrix} \begin{pmatrix} r_{21} \\ r_{22} \\ \vdots \\ r_{2n} \end{pmatrix} \cdots \begin{pmatrix} r_{r1} \\ r_{r2} \\ \vdots \\ r_{rn} \end{pmatrix} \tag{6}$$

The new approach does not require to omega-transform datasets that pertain to characteristics which are collected in terms of percentages. The elements for each characteristic are rank-ordered according to the optimal direction that has been prescribed for each response independently. The most desirable value on a response column gets a rank of '1' and the counting continues until the least desirable entry gets a rank of '*n*'. Ties are permitted in this formulation. Generally speaking, a measured characteristic is optimized in one of the three possible directions: 1) "smaller-is-better" (minimization), 2) "larger-is-better" (maximization) or 3) "nominal-is-best" (minimization towards a target value). In the ED-case that we will analyze in the next section, it is the first and second kind that will become pertinent.

In its generic form, a rank-ordering converts the response vector elements to ordered response vectors $O_1, O_2, \ldots, O_r$ and hence it becomes:

$$\cdots\begin{pmatrix} r_{i1} \\ r_{i2} \\ \vdots \\ r_{in} \end{pmatrix}\cdots \rightarrow \cdots\begin{pmatrix} o_{i1} \\ o_{i2} \\ \vdots \\ o_{in} \end{pmatrix}\cdots \text{ with } o_{1j},\ o_{2j} \ldots o_{rj}\ (j=1,2,\ldots,n) \tag{7}$$

Using the simple super-ranking process [14, 32], we compound the homogenized behavior of all of the responses in a single vector, the sum of the squared ranks, **SSR:** $\{SSR_i \mid \forall\ 1 \le i \le n\}$:

$$\begin{pmatrix} o_{11} \\ o_{12} \\ \vdots \\ o_{1n} \end{pmatrix}\begin{pmatrix} o_{21} \\ o_{22} \\ \vdots \\ o_{2n} \end{pmatrix}\cdots\begin{pmatrix} o_{r1} \\ o_{r2} \\ \vdots \\ o_{rn} \end{pmatrix} \rightarrow \begin{pmatrix} o^2{}_{11} + o^2{}_{21} + \cdots + o^2{}_{r1} \\ o^2{}_{12} + o^2{}_{22} + \cdots + o^2{}_{r2} \\ \vdots \\ o^2{}_{1n} + o^2{}_{2n} + \cdots + o^2{}_{rn} \end{pmatrix} \rightarrow \begin{pmatrix} SSR_1 \\ SSR_2 \\ \vdots \\ SSR_n \end{pmatrix} \tag{8}$$

The final input-output arrangement is depicted as:

$$\begin{pmatrix} x_{11} & x_{21} & \cdots & x_{m1} \\ x_{12} & x_{22} & \cdots & x_{m2} \\ \vdots & \vdots & \cdots & \vdots \\ x_{1n} & x_{2n} & \cdots & x_{mn} \end{pmatrix}\begin{pmatrix} SSR_1 \\ SSR_2 \\ \vdots \\ SSR_n \end{pmatrix} \tag{9}$$

Based on the generic structure of the relationship between the OA and the generated "super-rank response" SSR in equation 9, we tabulate next the corresponding arrangement specifically for the $L_9(3^4)$ OA that we will manipulate on the next section:

$$\begin{array}{c} Run\# \\ 1\\2\\3\\4\\5\\6\\7\\8\\9 \end{array} \begin{pmatrix} X_1 & X_2 & X_3 & X_4 \\ 1&1&1&1\\1&2&2&2\\1&3&3&3\\2&1&2&3\\2&2&3&1\\2&3&1&2\\3&1&3&2\\3&2&1&3\\3&3&2&1 \end{pmatrix} \begin{array}{c} \text{SSR} \\ \begin{bmatrix} \text{SSR}_1\\ \text{SSR}_2\\ \text{SSR}_3\\ \text{SSR}_4\\ \text{SSR}_5\\ \text{SSR}_6\\ \text{SSR}_7\\ \text{SSR}_8\\ \text{SSR}_9 \end{bmatrix} \end{array} \tag{10}$$

2.4 The distribution-free analysis of a saturated-unreplicated OA for messy datasets

Since saturated-unreplicated datasets do not allow any degrees of freedom to peruse uncertainty, they become messy because the unexplainable error remains inestimable. Messiness also ensues because response distributions may vary even within each of the m factor-setting combinations [43, 44]. Messiness is a complication that requires a more sophisticated treatment. We consider the general distribution-free analysis of a super-ranked quantity that fuses information from r water-quality responses. To detect potential non-linearity, two endpoint settings are needed to frame the experimental boundary - the operating range. A third setting is placed in between the two endpoints to snoop on non-linearity. A single execution of a minimal non-linear OA requires gathering observations from n (=2m+1) predefined recipes. The resulting super-ranked quantity is symbolized as $\{SSR_{i_1,i_2,\ldots,i_m}\}$ where each i_j (j=1, 2,…,m) identifies the setting status of the j^{th} influence. Thus, planning with a three-setting OA, there are only three admissible states appointed to each i_j. Each i_j may be coded by assigning to it the generic ordinal values: '1', '2' and '3'. By default, we let '1' and '3' to represent the two operating endpoints. We propose a non-linear effects model as [31]:

$$SSR_{i_1,i_2,\cdots,i_m} = M + \sum_{j=1}^{m} D_j + \varepsilon_{i_1,i_2,\cdots,i_m} \tag{11}$$

The error term, $\varepsilon_{i_1,i_2,\cdots,i_m}$, is not bound to any particular distribution. Simply, it should be checked for statistical symmetry across the three settings, for each examined factor individually, before attempting to explain the results of the effect contrasting. The overall (grand) median, M in equation 11, for all n ***SSR*** entries is defined as:

$$M = \mathrm{Med}\left(\left\{ SSR_{i_1,i_2,\cdots,i_m} \right\}\right) \tag{12}$$

The median values of the ***SSR*** response at their three respective factor settings are: M_j^1, M_j^2 and M_j^3 with $1 \le j \le m$. The setting measure, M_j, represents a median estimation of a group of observations that share the same factor setting i_j ($1 \le j \le m$):

$$M_j = \begin{Bmatrix} M_j^1 = \mathrm{Med}\left(\left\{ SSR_{\ldots,i_j,\ldots} \right\}\right) \text{if } i_j \to 1 \\ M_j^2 = \mathrm{Med}\left(\left\{ SSR_{\ldots,i_j,\ldots} \right\}\right) \text{if } i_j \to 2 \\ M_j^3 = \mathrm{Med}\left(\left\{ SSR_{\ldots,i_j,\ldots} \right\}\right) \text{if } i_j \to 3 \end{Bmatrix} \text{for all } i_j \tag{13}$$

From equation 11, the indexed quantity D_j is the difference between M_j and M which quantifies the i_j^{th} partial (relative) effect due to the j^{th} factor with respect to the grand median:

$$D_j = \begin{Bmatrix} D_j^1 = M_j^1 - M \quad \text{if } i_j \to 1 \\ D_j^2 = M_j^2 - M \quad \text{if } i_j \to 2 \\ D_j^3 = M_j^3 - M \quad \text{if } i_j \to 3 \end{Bmatrix} \tag{14}$$

After fitting equation 11, we unstack the partial effect terms to create a new simpler response that packs information for only a specific effect and is denoted as $RSS'_{i_1,i_2,\cdots,i_m}$. Thus, the reconstructed response is sub-divided as: 1) the grand median, M, 2) the partial effect, D_j, and 3) the corresponding error contribution, $\varepsilon_{\ldots,i_j,\ldots}$ for all i_j or:

$$RSS'_{\ldots,i_j,\ldots} = M + D_j + \varepsilon_{\ldots,i_j,\ldots} \quad \text{for all}\, i_j \text{ and } 1 \le j \le m \tag{15}$$

For each effect separately, we rank-order $RRS'_{i_1,i_2,\cdots,i_m}$ to transform it to the rank response, $r_{i_1,i_2,\cdots,i_m}$:

$$RRS'_{\ldots,i_j,\ldots} \rightarrow r_{\ldots,i_j,\ldots} \quad \text{for all}\, i_j \text{ and } 1 \le j \le m \tag{16}$$

We next form the mean rank sums for all three settings of the j^{th} effect, $\overline{R}_j^1$, $\overline{R}_j^2$ and $\overline{R}_j^3$:

$$\left.\begin{aligned} \overline{R}_j^1 &= \frac{\sum_{i_j} r_{\ldots,i_j,\ldots}}{(n/3)} \text{ if } i_j \rightarrow 1 \\ \overline{R}_j^2 &= \frac{\sum_{i_j} r_{\ldots,i_j,\ldots}}{(n/3)} \text{ if } i_j \rightarrow 2 \\ \overline{R}_j^3 &= \frac{\sum_{i_j} r_{\ldots,i_j,\ldots}}{(n/3)} \text{ if } i_j \rightarrow 3 \end{aligned}\right\} \text{for all } i_1, i_2, \cdots, i_m \tag{17}$$

The Kruskal-Wallis test statistic [46], H_j $(1 \le j \le m)$, is appropriate for testing the one-way fluctuation of ranks across the three settings for each effect:

$$H_j = \left[\frac{12}{n(n+1)} \sum_{k=1}^{3} (n/3)\left(\overline{R}_j^k\right)^2\right] - 3(n+1) \tag{18}$$

Prior to delivering a screening prediction is imperative to ensure the uniformity and stability of the de-fractionated residual error in the preceding ordering operations. For this purpose, an effect-free vector is generated to carry the discrepancies. The uncertainty vector is $SSR''_{i_1,i_2,\cdots,i_m}$ such that:

$$SSR''_{\ldots,i_j,\ldots} = M + \varepsilon_{\ldots,i_j,\ldots} \quad \text{for all } i_j \text{ and } 1 \le j \le m \tag{19}$$

Proceeding to rank-order the $RSS''_{\ldots,i_j,\ldots}$ will yield the transformed response, $r'_{i_1,i_2,\cdots,i_m}$:

$$RSS''_{\ldots,i_j,\ldots} \rightarrow r'_{\ldots,i_j,\ldots} \quad \text{for all}\, i_j \text{ and } 1 \le j \le m \tag{20}$$

Forming the mean rank sums of the $r'_{i_1,i_2,\cdots,i_m}$ for all three settings of the j^{th} effect, $\overline{\mathrm{Re}}_j^k$, with k = 1, 2 or 3, we obtain:

$$\left.\begin{aligned}\overline{\mathrm{Re}}_j^1 &= \frac{\sum_{i_j} r'_{\dots,i_j,\dots}}{(n/3)} \text{ if } i_j \to 1\\ \overline{\mathrm{Re}}_j^2 &= \frac{\sum_{i_j} r'_{\dots,i_j,\dots}}{(n/3)} \text{ if } i_j \to 2\\ \overline{\mathrm{Re}}_j^3 &= \frac{\sum_{i_j} r'_{\dots,i_j,\dots}}{(n/3)} \text{ if } i_j \to 3\end{aligned}\right\} \text{for all } i_1, i_2, \cdots, i_m \tag{21}$$

The Kruskal-Wallis test statistic for the uncertainty is similarly defined as: He_j, $(1 \le j \le n)$:

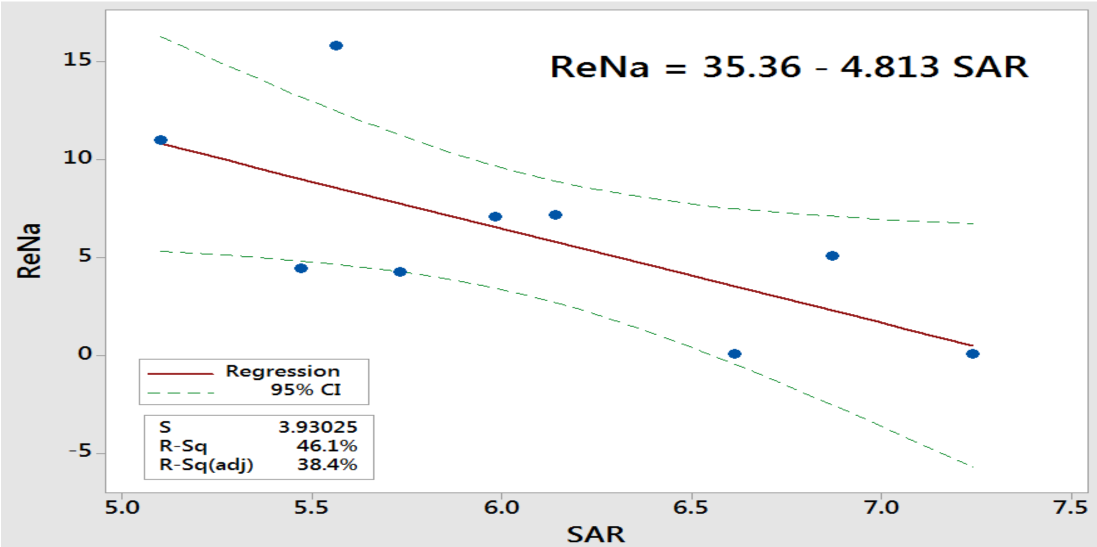
$$He_j = \left[\frac{12}{n(n+1)} \sum_{k=1}^{3} (n/3)\left(\overline{\mathrm{Re}}_j^k\right)^2\right] - 3(n+1) \tag{22}$$

The quantity He_j tracks underlying intrusions in the dataset that could destabilize the validity for each observation. Intense sporadic fluctuations of the uncertainty could blemish the significance of the screening results (equation 18). If the n calculated contrasts (equation 22) show that there is no statistical significant relationship between the m controlling factors and the experimental uncertainty, then, we may proceed to proposing any strong effects from the statistical profiling (equation 18). The exact Kruskal-Wallis test significances are computed with the statistical software package STATISTICA 9 (StatSoft).

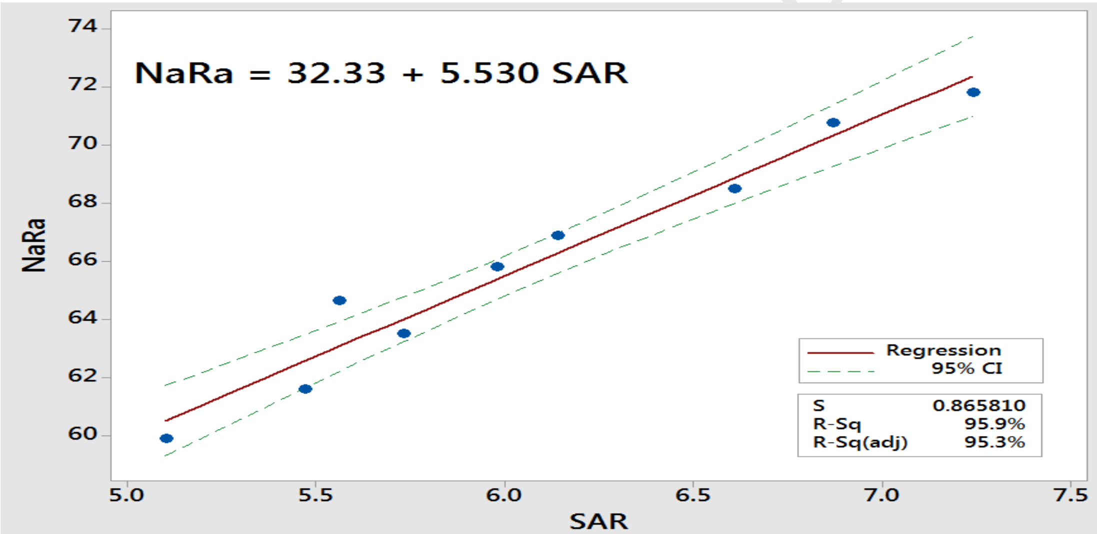
3. Results

3.1 Preliminary data analysis

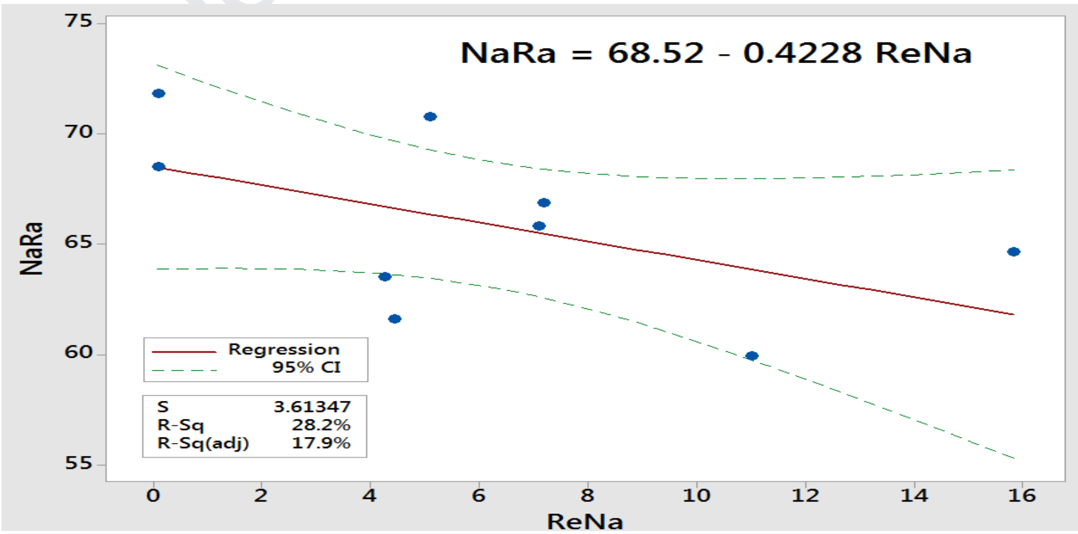
The three water characteristics generate parallel and probably overlapping information in tracking the wastewater treatment performance. Therefore, the ED-process efficiency (ReNa) and the two water quality indices (SAR and NaRa) should be tested for possible correlations between them. If they found to significantly correlate with each other, then, some of them should be eliminated from further modelling consideration. They would merely provide redundant information. The three possible correlation comparisons among the three responses are fitted and shown in Figure 1. The linear regression fittings do not reveal any relationship that could be established from any of the two-way response contrasting. This is owing to the fact that in all three cases there are in total 9 plotted data points and no data point is expected to be situated outside the 95%-CI. On the contrary, we observe at least one point to always hang out of the CI bands and several other points to populate near to their respective CI boundaries. This is true even in the case of regression analysis of NaRa versus SAR where it appears that the coefficient of determination (R^2) covers adequately the goodness-of-fit criterion; R^2 is calculated to be 95.3%. With regards to inspecting for a NaRa versus SAR correlation, we also notice that there are two points out of the total nine out of the 95%-confidence-interval band, i.e. 22% of the count, when less than 5% was to be expected. Summarizing this overall behavior, we simply conclude that the three pairs of correlations cannot be assessed at all leaving us clueless. We witness that there are no tendencies between them, irrespective of the evaluations of the slope and the goodness of fit. This exercise aids to realize one source of inherent messiness in the dataset. No conclusive judgement maybe drawn with ordinary analysis means. Ostensibly, all three responses should be maintained in the analysis. Their concurrent processing is the proper step for proceeding with sizing the effects. A preliminary "clothesline" boxplot screening for each participating factor setting, with respect to each individual chemometrical response, reveals a great range of skewness variation to include various forms of symmetrical and unsymmetrical tendencies (Part C - Supplementary Material). Additionally, their boxplot spreads exhibit a broad variability.



A)



B)



C)

Figure 1: Fitted line plots for: A) ReNa vs SAR, B) NaRa vs SAR, C) NaRa vs ReNa.

It would necessitate a rather robust and agile solver to delve into the obscure statistical relationships between factors and responses. Thus, the motive for this work is now justified.

3.2 Distribution-free analysis of the saturated-unreplicated wastewater dataset

In Table 1, we listed the rank-ordered response elements for vectors ReNa, SAR, and NaRa. It is also tabulated the corresponding sum of squared ranks (SSR). A "clothesline" box-plot screening of the SSR against the four controlling factors is shown in Figure 2, for each factor setting separately. The anisotropy in data location and dispersion persists in the fused SSR vector elements. It exacerbates the tendency for extreme skewness. Only in two out of the twelve (17%) box-plots exhibit symmetric behavior (DF2 and V80). Similarly, only two out of the twelve (17%) box-plots post a decent (contained) variation (DF2 and DF5). Even so, the 95% confidence interval of the medians coincide with the box length exposing a large variability in the compounded SSR quantity. The grand (concurrent) median of the SSR vector is computed to be 74 (Table 2). We observe that the DF-factor causes the greatest disturbance to the SSR. Its endpoints traverse from a low 44 to a high 170.3. In Table 3, we provide the detailed analysis for the reconstructed error and factor vectors which are described by the models in equations 15 and 19, SSR″ and SSR′, respectively. Utilizing the equations 18 and 22, we calculate the Kruskal-Wallis estimators (H and He) for the factor and reconstructed error vectors along with their statistical significance which is expressed in terms of p-values in Table 4. At a level of significance of 0.05, we observe that the errors across all factor settings contribute symmetrically to the effects. At a level of significance of 0.05, it is the DF-factor that barely misses to make the cut. Therefore, at this stage it is not conclusive that any factor could optimally adjust all three characteristics to their best performance. This essentially means that any factor-setting may be picked for operating the ED cell within the tested ranges. Based on that data alone, the final decision merely rests on economic and practical constraints that have not been included in the published model [7].

Table 1: The three rank ordered responses and their sum of squared ranks (SSR).

Run #	RReNa	RSAR	RNaRa	SSR
1	6	2	2	44
2	7	4	3	74
3	2	1	1	6
4	3	6	6	81
5	1	3	4	26
6	4	5	5	66
7	8.5	9	9	234.25
8	5	8	8	153
9	8.5	7	7	170.25

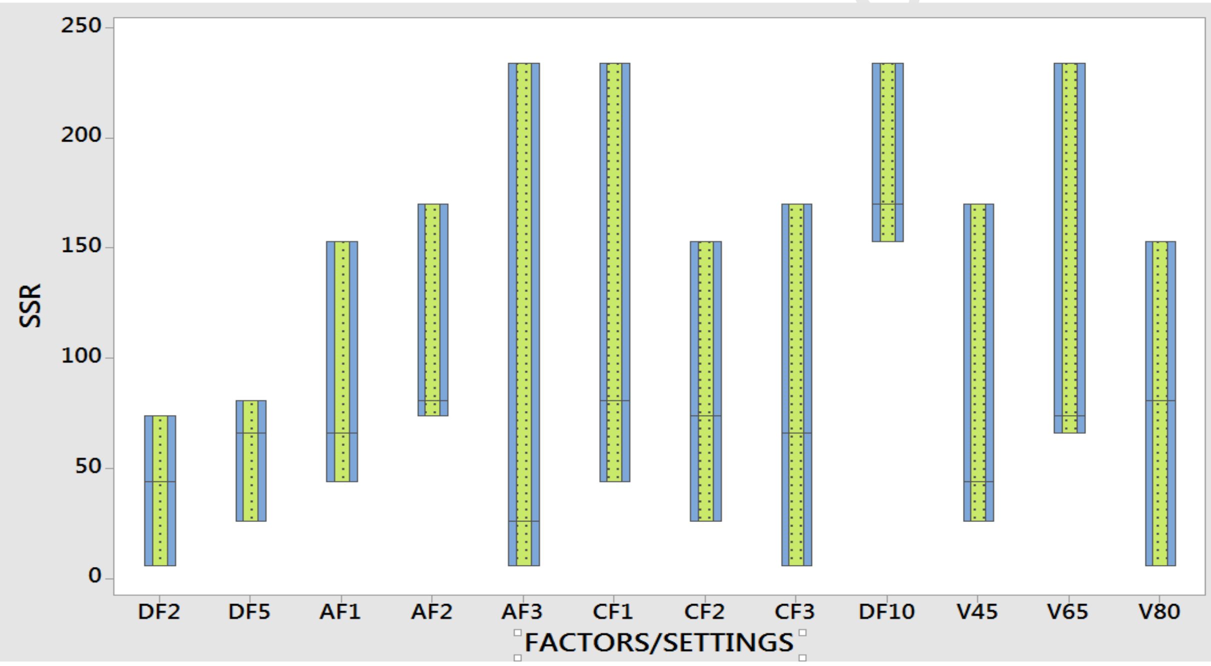


Figure 2: Box-plots for all-effect screening of SSR.

Table 2: Median (concurrent) SSR estimations and relative effects for all factor settings.

	Level	Median	Relative Effect
DF	2	44	-30
	5	66	-8
	10	170.3	96.3
CF	1	81	7
	2	74	0
	3	66	-8
AF	1	66	-8
	2	81	7
	3	26	-48
V	45	44	-30
	65	74	0
	80	81	7
GRAND MEDIAN		**74**	

Table 3: Reconstructed error (SSR″) and factor-specific (SSR′) vectors.

Run #	SSR' (DF)	SSR' (CF)	SSR' (AF)	SSR' (V)	SSR''
1	75.0	112.0	97.0	75.0	105.0
2	67.0	97.0	104.0	97.0	97.0
3	55.0	77.0	37.0	92.0	85.0
4	60.0	75.0	75.0	75.0	68.0
5	104.0	112.0	64.0	82.0	112.0
6	82.0	82.0	82.0	90.0	90.0
7	275.3	186.0	131.0	179.0	179.0
8	154.0	57.7	49.7	64.7	57.7
9	201.3	97.0	112.0	75.0	105.0

Table 4: Effect symmetry and strength significance for SSR using the Kruskal-Wallis test.

	Error Symmetry		Effect Strength	
Factor	He-estimator	p-value	H-estimator	p-value
DF	0.09	0.957	5.96	0.051
CF	0.62	0.733	0.96	0.618
AF	1.69	0.43	1.07	0.587
V	5.6	0.061	4.32	0.115

4. Discussion

The concurrent optimization of the wastewater ED-process may be assessed by reviewing the individual behaviors against the respective (ordinary) main effects plots (Part D - Supplementary Material). Briefly, the DF-effect plays the predominant role in all three screenings. Since NaRa and SAR are ought to be both minimized, we see that this could be conveniently achieved because their behavior appears to be linear. The suggested optimal dilute flow is located in the lower endpoint, at the value of 2 L/h. However, the ReNa response should be maximized and the experimental evidence shows that the suggested optimal dilute flow should move to an adjustment of 5 L/h. ReNa produces non-linear profiles for all four factors. This justifies the selected non-linear framework to deal with the experimental design of the wastewater trials. Finding the maximum ReNa performance is less clear if it is to consider all four effects. This implies that a realistic search for an optimum recipe would be derived only from a concurrent profiling. However, the statistical significance of their magnitudes cannot be obtained with ordinary means. The classical approach of using ANOVA treatment does not lead to an inference (Part A - Supplementary Material). This is because F-test comparisons cannot be executed for the saturated designs. In all three ANOVA screenings, the DF-factor appears to precede the other three effects. This observation is in agreement for both regular approaches, i.e. either using: 1) the relative magnitudes of the general linear model (GLM) coefficients or alternatively 2) the adjusted mean squares in ANOVA. Of course, this is a subjective opinion because as we saw in the previous section the three responses tend to strongly depart from normality although both comparison tests take normality as a key assumption. Thus, we do not know actually how reliable these ANOVA or GLM estimations are because no significance can be extracted from them. Similarly, the disturbances caused by effects CF, AF and V on the SAR and NaRa responses individually are clearly weak when sized against the influence of DF (Part D - Supplementary Material). It is hard to discriminate how really strong is the presence of DF since it only causes a decrease of 22% and 13% on SAR and NaRa values, respectively, in spite of stretching the ReNa range by 456%. At this point, the initial decision to doubt a potential correlation between the two responses SAR and NaRa becomes more evident. For SAR and NaRa, the optimum occurs at DF-settings 2 L/h (minimum) and 10

L/h (maximum). For SAR, the sample mean (m) and its standard error (se) are: 1) at setting 2 L/h, m= 5.43 and se = 0.18, and 2) at setting 10 L/h, m= 6.91 and se = 0.18. For $t_{n-1,\alpha} = t_{2,0.025} = 4.3$, we observe that the 95% confidence intervals of the two limiting settings overlap potential signaling that there is no detected effect. Similarly, for NaRa, the sample mean (m) and its standard error (se) are: 1) at setting 2 L/h, m= 61.68% and se = 1.04%, and 2) at setting 10 L/h, m= 70.34% and se = 0.98%. We observe then that their 95% confidence intervals are barely overlapping, and hence it is hinted a possibly 'no-effect' status. Therefore, the two responses SAR and NaRa indeed might not correlate also from this standpoint.

Table 5: Median ReNa response and relative effect for all factor settings.

	Level	Median	Relative Effect
DF	2	4.42	-0.66
	5	7.17	2.09
	10	0.08	-5
CF	1	4.42	-0.66
	2	5.08	0
	3	7.08	2
AF	1	5.08	0
	2	4.25	-0.83
	3	11	5.92
V	45	4.42	-0.66
	65	4.25	-0.83
	80	7.17	2.09
GRAND MEDIAN		**5.08**	

Table 6: Reconstructed error (ReNa″) and factor-specific (ReNa′) vectors.

Run #	ReNa'(DF)	ReNa'(CF)	ReNa'(AF)	ReNA'(V)	ReNa''
1	5.7	5.7	6.4	5.7	6.4
2	5.9	6.6	5.7	5.7	6.6
3	1.0	3.7	7.6	3.7	1.7
4	6.6	3.8	3.7	6.6	4.5
5	10.6	8.5	14.4	7.8	8.5
6	5.9	5.8	3.8	3.0	3.8
7	-4.4	0.0	6.6	-0.2	0.7
8	3.0	8.0	8.0	10.1	8.0
9	-0.4	6.6	3.7	3.9	4.6

Table 7: Effect symmetry and strength significance for ReNa response using Kruskal-Wallis test.

	Error Symmetry		Effect Strength	
Factor	He-estimator	p-value	H-estimator	p-value
DF	0.09	0.957	6.12	0.047
CF	5.42	0.067	5.54	0.063
AF	0.62	0.733	5.07	0.079
V	1.87	0.393	2.89	0.236

It is tenable to repeat the same distribution-free screening that was conducted for SSR this time only on the ReNa response. Working with a single response maintains the intricacy of dealing with saturated-unreplicated designs. Subsequently, in Table 5, we tabulate the median response and the relative effect for all four factors on ReNa. The grand median is 5.08%. DF and AF appear to contribute the most in tweaking the ReNa reaction. To run significance diagnostics, first the reconstructed error (ReNa″) and the factor-specific (ReNa′) vectors are prepared (Table 6). Subsequently, the Kruskal-Wallis estimator is evaluated and the statistical significance for error symmetry and effect strength is obtained (Table 7). We observe that the error symmetry is well-balanced across all factors at a level of significance of 0.05. We notice that for a single ReNa response screening, the DF-effect makes the only strong influence at a level of significance of 0.05. The response graph for main effects of SSR (Figure 3) portrays in a descriptive fashion a situation where the DF-effect stands out as a component that cannot be comparable to any of the other factors.

However, the data noise is as severe as we saw in the previous section that the protracted, nearly linear, inclination of the SSR vs DF fitting cannot be taken advantage of to satisfy ED-designing objectives. To verify this result, we also use the classical half-normal plot [47] to test from a different angle our findings. In Figure 4, we graph linear and quadratic effects of SSR for all four factors using the half-normal plot. This means that there must be eight points on the graph to make the distinction between the two curvature types. We discern that the linear part of the DF-effect tends to deviate from the string of points (rest of the effects) to the right. However, the slope of the line formed by the string of the rest of the effect-points also bends away from zero. Thus, it is hard to document the strength of the DF effect. The fact that the linear DF contribution scores below the 95% limit also may uphold this perspective.

Finally, to be consistent with the discussion on the tendencies of the individual screenings, three normal plots are prepared for each response (Fig. 5). There is no evidence of substantial divergence of any factor from the rest of the group in the ReNa half-normal plot. This comes in contrast to our result in Table 7, which asserts that the DF is a vital effect and capable of influencing the ReNa response. On the other hand, the linear part of DF seems to strongly depart from the behavior of the rest of the effects in the SAR and NaRa half-normal plots. This is in agreement with the behavior on the corresponding main effects plots.

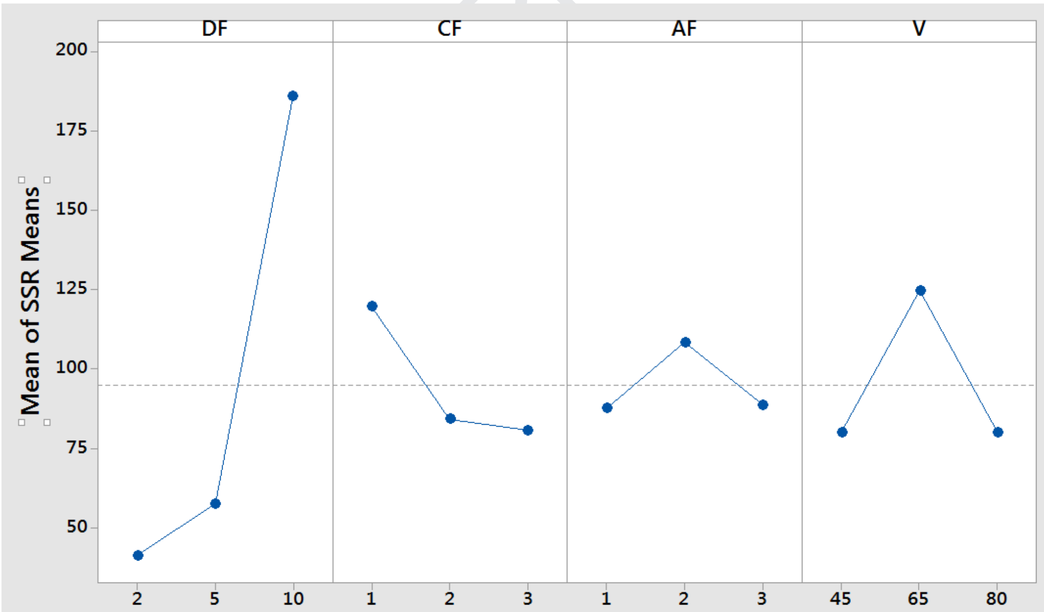


Figure 3: Main effects plots for SSR means.

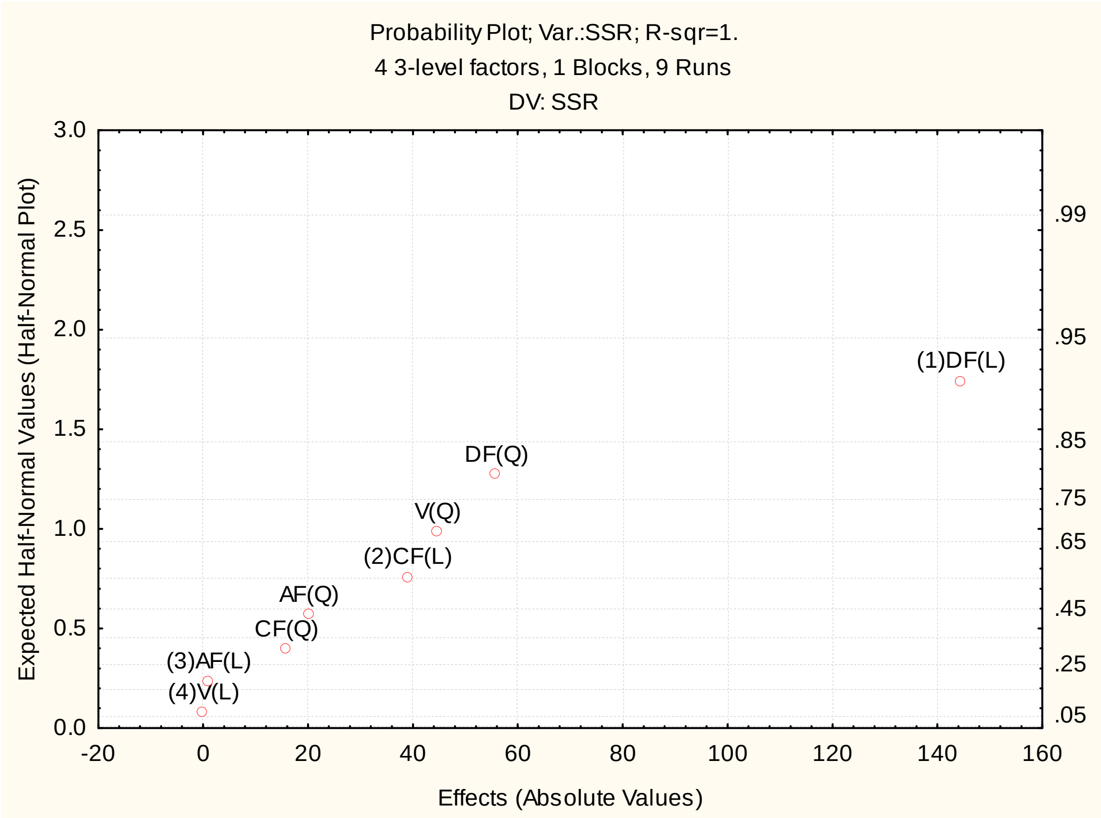
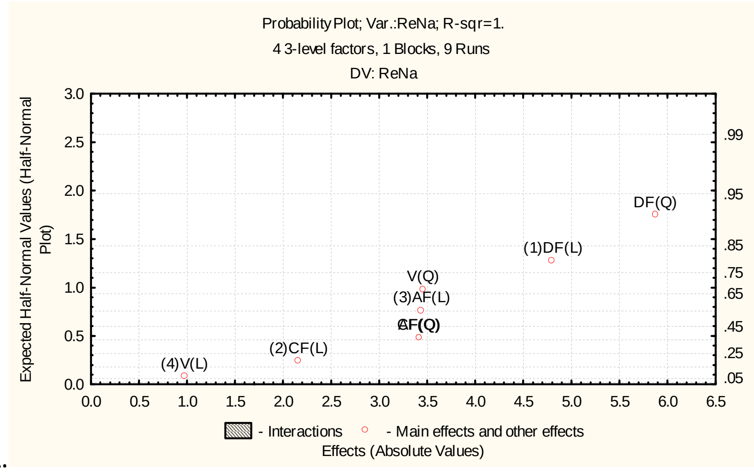


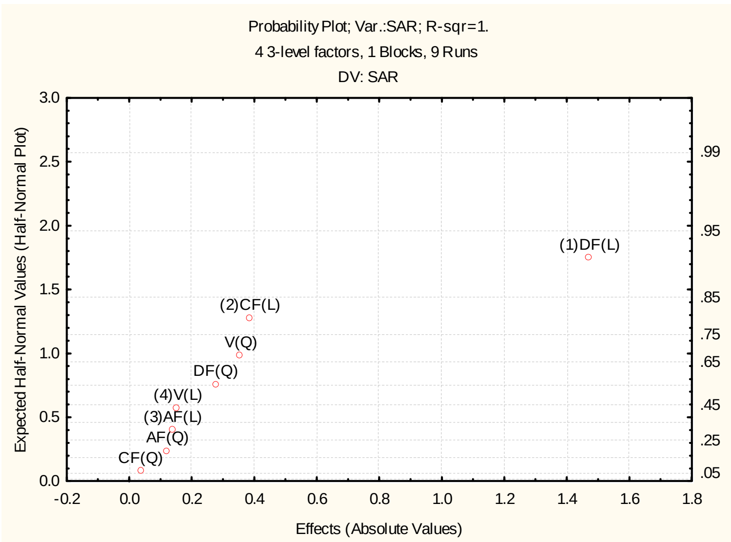
Figure 4: Half-normal plot for SSR response.

But again this result is not significant at a level of 0.05. It becomes clear from this screening/optimization effort that the settings that generate low SSR values are favored. This means that for the specific ranges that have been worked out in this paradigm the adjustments should be: DF=2 L/h, CF=3 L/h, AF=3 L/h and V=45V (Table 2). This predicts ReNa, SAR and NaRa values of 11.68%, 4.63, and 59%, respectively.

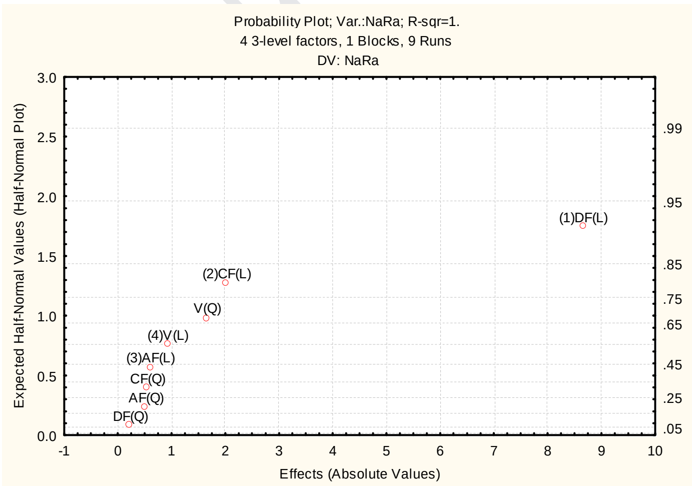
Furthermore, using the Taguchi-defined Ω-transformation for percentages, we depict the ReNa and NaRA characteristics in terms of their main effects plot for means in Part B (Supplementary Material). Checking the effect tendencies with the Ω-transformation for ratios, it becomes imperative for the response observations of ReNa due to the fact that 44% (4 out of 9) of the total observations have been measured under 5% [26], while runs # 7 and 9 (22%) had produced ReNa magnitudes close to 0 (0.08%). By employing the Ω function, now, it becomes more obvious now that besides the influence of dilute flow on ReNa variable, there is also a trend on NaRa variable. It is demonstrated that the dilute-flow low limit favors the maximization of the ReNa variable and the minimization of the NaRa variable.



A.



B.



C.

Figure 5: Half-normal plots for ReNa (A), SAR(B) and NaRa(C).

The main effects of Ω in Part B (Supplementary Material) aid in understanding that there is not significant difference between dilute flows of 2 and 5 L/hr. Visual tendencies match to the effect strength of the dilute flow on the concurrent multi-response adjustment (Table 4).

Finally, the initial optimization procedure by Abou-Shaby [7], that was also re-examined in this article, led to further modifications on the original design that as anticipated provided even greater efficiency for the ED cell after a second round of optimization. The promising concept of using an ED process as it was described in Abou-Shaby [7] should be further tested to accommodate even larger scale demands for irrigating even wider areas of crops. As the ED tank size should substantially increase in size from its current specifications, new optimization effort should be attempted. In that case, it would be interesting to take in account the possibility of introducing in the study effects additional opportunities for ED-performance enhancement such as: various exchange membrane types, optimal electrode dimensions, cell stack configuration, compartment configuration, ion-exchange resin-bead transport bridging, the influence of the origin of the feedwater sources and its associated mixture optimality on the overall ED efficiency and so forth.

5. Conclusions

Managing to extract water for household and irrigation needs from polluted wastewater pools is a major modern environmental challenge. Special engineering methods are needed to be adapted each time to the particular kind of local water demands in order to ensure adequate water supply. One of the most promising chemical processes to assist such plans is electrodialysis. For optimum feed recovery, sophisticated optimization methodologies are necessitated to deal with the complexity of the electrodialysis process at hand. Due to the intricate nature of environmental phenomena, robust aquametrics optimization techniques should be employed to assure that water quality indices are optimized. Design of experiments in a distribution-free framework may aid to surpass several cell design and process design sticking issues. Since experiments are needed to describe each time the type of effluent to work with along with the cell conditions, it is only practical to make measurements in electrodialysis operations only in small samples as in Abou-Shady's investigation. We showed in this work how to extract information on difficult multi-response multifactorial datasets from a real published wastewater study. The study was intriguing because it exposed various complications that will be confronted in chemometrics when trying to improve electrodialysis performance. Therefore, in this effort, we made an attempt to distinctly demonstrate the underlying complications that could undermine a robust decision in improving polluted wastewater operations.

The aspects that the unique multi-response aquametrics affected the interpretation of Abou-Shady's electrodialysis trials and were elucidated in this work were:

1) data smallness,

2) data non-linearity,

3) data non-normality,

4) data messiness,

5) effect saturation,

6) trial unreplication,

7) statistical multifactorial optimization,

8) concurrent multiresponse optimization.

Robust and agile tools were shown to be necessitated for such pragmatic situations in order to lead to fast and reliable inference. We proposed a method that encompasses:

1) the superanking approach to fuse and handle multiple water-quality indices,
2) the robust screening from a non-linear multifactorial statistical profiler to overcome the many data prerequisites discussed above and were not covered by the basic assumptions in the classical Taguchi approach.

It was found that:

1) A concurrent solution does not promote specific settings
2) Two out of the three responses were not affected substantially in the investigated ranges

3) Ω-transformation gives better resolution than the classical SNR transformation
4) The problem could be reduced to a single response, not because of correlated response pairs
5) The removed sodium was shown to be controlled by setting the diluted flow at 5 L/h which may be dropped to 2 L/h (potentially even lower) for the multiresponse screening case.
6) The rest of the factors are not statistically significant.
7) Inactive factors could be adjusted based on the multireponse screening solution or from practical and economic considerations.

To recapitulate the findings, it is recommended that the lower operating end of dilute flow could be tweaked for magnitudes lower than 1 L/h. Conducting again the trials would provide even better resolution for the underlying separation phenomena. Future research could also include in the design different types of efflux, mixtures of effluxes from different sources, number of cell stacks, energy requirements, types of membranes (polymer structure) and so forth.

Acknowledgements: We thank the Editor in Chief and the three reviewers for their critical comments that led to the improvement of this work.

References

[1] Y. Jake, C. Robert, S. Jesse, A. Caroline, A thirsty world, Science 313 (2006) 1088.

[2] M. Shannon, P. Bohn, M. Elimelech, J. Georgiadis, B. Marinas, A. Mayes, Science and technology for water purification in coming decades. Review, Nature 452 (2008) 301–310.

[3] U. Yermiyahu, A. Tal, A. Ben-Gal, A. Bar-Tal, J. Tarchitzky, O. Lahav, Rethinking desalinated water quality and agriculture, Science 318 (2009) 920–921.

[4] C.A. Quist-Jensena, F. Macedonioa, E. Drioli, Membrane technology for water production in agriculture: desalination and wastewater reuse, Desalination 364 (2015) 17–32.

[5] R. Zito, Electrochemical water processing, Hoboken: Wiley-Scrivener, 1st ed; 2011.

[6] Y. Tanaka, Ion exchange membrane electrodialysis: Fundamentals, desalination, separation, Nova Science, 2013.

[7] A. Abou-Shady, Recycling of polluted wastewater for agriculture purpose using electrodialysis: Perspective for large scale application, Chemical Engineering Journal 323 (2017) 1-18.

[8] H.B. Park, J. Kamsev, L.M. Robeson, M. Elimelech, B.D. Freeman, Maximizing the right stuff: The trade-off between membrane permeability and selectivity, Science 356 (2017) 6343.

[9] A.E Johnston, Understanding potassium and its use in agriculture. Brussels: European Fertilizer Manufacturers Association Publication; 2015.

[10] D.L. Massart, B.G.M. Vandeginste, L.M.C. Buydens, S. de Jong, P.J. Lewi, J. Smeyers-Verbeke, Handbook of Chemometrics and Qualimetrics, Part A, Elsevier, Amsterdam, 1997.

[11] G. Taguchi, S. Chowdhury, Y. Wu, Quality Engineering Handbook. Hoboken:Wiley-Interscience; 2004.

[12] G. Taguchi, S. Chowdhury, S. Taguchi, Robust Engineering: Learn How to Boost Quality While Reducing Costs and Time to Market. New York: McGraw-Hill; 2000.

[13] R.A. Stone, A. Veevers. The Taguchi influence on designed experiments. Journal of Chemometrics 8 (1994) 103-110.

[14] G.J. Besseris, Eco-design in total environmental quality management: Design for environment in milk-products industry, TQM Journal 24 (2012) 47-58.

[15] S.Y. Dai, H. Sang, K.-M. Lee, T.J. Herman. Cost of speed: A practical approach to evaluate a screening method from a Bayesian perspective. Chemometrics and Intelligent Laboratory Systems 156 (2016) 273-279.

[16] S.S. Lepeniotis, M.J. Vigezzi. Lowering manufacturing cost of material by formulating it through statistical modeling and design. Chemometrics and Intelligent Laboratory Systems 29 (1995) 133-139.

[17] S.S. Madeni, S. Koocheki, Application of Taguchi method in the optimization of wastewater treatment using spiral-wound reverse osmosis element, Chemical Engineering Journal 119 (2006) 37-44.

[18] N.M.S. Kaminari, D.R. Schultz, M.J.J.S. Ponte, H.A. Ponte, C.E.B. Marino, A.C. Neto, Heavy metals recovery from industrial wastewater using Taguchi method, Chemical Engineering Journal 126 (2007) 139-146.

[19] S. Aber, D. Salari, M.R. Parsa, Employing the Taguchi method to obtain the optimum conditions of coagulation-flocculation process in tannery wastewater treatment, Chemical Engineering Journal 162(2010) 127-134.

[20] K.P.Y Shak, T.Y Wu, Coagulation-flocculation treatment of high-strength agro-industrial wastewater using natural Cassia obtusifolia seed gum: Treatment efficiencies and flocs characterization, Chemical Engineering Journal 256 (2014) 293-305.

[21] M. A. Tofighy, Y. Shirazi, T. Mohammadi, A. Pal. Salty water desalination using carbon nanotubes membrane. Chemical Engineering Journal 168 (2011) 1064-1072.

[22] P.M. Pardeshi, A.A. Mungray, A.K. Mungray, Determination of optimum condition in forward osmosis using a combined Taguchi-neural approach, Chemical Engineering Research Design 109 (2016) 215-225.

[23] D.-L. Wu, W. Wang, Q.-W. Guo, Y.-H. Shen, Combined Fenton-SBR process for bamboo industry wastewater treatment, Chemical Engineering Journal 214(2013) 278-284.

[24] C. Andre, F. Jorge, I. Castanheira, A. Matos. Optimizing UPLC isocyanate determination through a Taguchi experimental design approach. Journal of Chemometrics 27 (2013) 91-98.

[25] B. Simsek, Y.Tansel ic, E.H. Simsek. A TOPSIS-based Taguchi optimization to determine optimal mixture proportions of the high strength self-compacting concrete. Chemometrics and Intelligent Laboratory Systems 125 (2013) 18-32.

[26] M. R. Sohrabi, S. Jamshidi, A. Esmaeilifar. Cloud point extraction for determination of Diazinon: Optimization of the effective parameters using Taguchi method. Chemometrics and Intelligent Laboratory Systems 110 (2012) 49-54.

[27] R. Baronas, J. Kulys, A. Zilinskas, A. Lancinskas, D. Baronas. Optimization of the multianalyte determination with biased biosensor response. Chemometrics and Intelligent Laboratory Systems 126 (2013) 108–116.

[28] G.E.P. Box, W.G. Hunter, J.S. Hunter, Statistics for experimenters – design, innovation, and discovery, 2nd ed. New York: Wiley; 2005.

[29] R. Sundberg. Interpretation of unreplicated two-level factorial experiments, by examples, Chemometrics and Intelligent Laboratory Systems 24 (1994) 1–17.

[30] G.J. Besseris, A distribution-free multi-factorial profiler for harvesting information from high-density screenings, PLoS One 8 (2013) e73275.

[31] G.J. Besseris, A fast-and-robust profiler for improving polymerase chain reaction diagnostics, PLoS One 9 (2014) e108973.

[32] G.J. Besseris. Multi-response multi-factorial master ranking in non-linear replicated-saturated DOE for qualimetrics. Chemometrics and Intelligent Laboratory Systems 116 (2012) 47-56.

[33] G.E.P. Box, Signal-to-noise ratios, performance criteria and transformation. Technometrics 30 (1988) 1-17.

[34] S. Maghsoodloo, G. Ozdemir, V. Jordan, C.-H. Huang, Strengths and limitations of Taguchi's contributions to quality, manufacturing, and process engineering, Journal of Manufacturing Systems 23 (2004) 73-126.

[35] J.J. Pignatiello, J.S. Ramberg, Top ten triumphs and tragedies of Genichi Taguchi, Quality Engineering 4 (1992) 211-225.

[36] R. Carlson, A. Nordahl, T. Barth, R. Myklebust, An approach to evaluating screening experiments when several responses are measured, Chemometrics and IntelligentLaboratory Systems 12 (1992) 237–255.

[37] Y.-Z. Liang, K.-T. Fang, Q.-S. Xu, Uniform design and its applications in chemistry and chemical engineering, Chemometrics and Intelligent Laboratory Systems 58 (2001) 43–57.

[38] J. Goupy. Unconventional experimental designs theory and application. Chemometrics and Intelligent Laboratory Systems 33 (1996) 3–16.

[39] S.N. Deming, J.A. Palasota, J.M. Palasota, Experimental design in chemometrics, Journal of Chemometrics 5 (1991) 181–192.

[40] S.F. Moller, J. Von Frese, R. Bro, Robust methods for multivariate data analysis, Journal of Chemometrics 19 (2005) 549–563.

[41] P.J. Ross, Taguchi Techniques for Quality Engineering. New York: McGraw-Hill Professional; 1995.

[42] N. Silver, The signal and the noise: Why so many predictions fail-but some don't, 1st ed. New York: Penguin; 2015.

[43] G.A. Milliken, D.E. Johnson, Analysis of Messy Data, Volume I: Designed Experiments. Boca Raton: Chapman and Hall/CRC; 2004.

[44] G.A. Milliken, D.E. Johnson, Analysis of Messy Data, Volume II: Nonreplicated Experiments. Boca Raton: Chapman and Hall/CRC; 1989.

[45] M. Daszykowski, K. Kaczmarek, Y. Vander Heyden, B. Walczak, Robust statistics in data analysis—a review. Basic concepts, Chemometrics and Intelligent Laboratory Systems 85 (2007) 203–219.

[46] W.H. Kruskal, W.A. Wallis, Use of ranks in one-criterion variance analysis, Journal of the American Statistical Association 47 (1952) 583-621.

[47] C. Daniel, Use of the half-normal plots in interpreting factorial two-level experiments, Technometrics 1 (1959) 311-341.

[48] Goal 6: Ensure access to water and sanitation for all, United Nations Sustainable Development, https://www.un.org/sustainabledevelopment/water-and-sanitation/ (Accessed 5/28/2019).

[49] Goal 2: Zero hunger, United Nations Sustainable Development, https://www.un.org/sustainabledevelopment/hunger/ (Accessed 5/28/2019).

[50] Goal 1: End poverty in all its forms everywhere, United Nations Sustainable Development, https://www.un.org/sustainabledevelopment/poverty/ (Accessed 5/28/2019).

[51] M. Hamada, N. Balakrishnan, Analyzing unreplicated factorial experiments: A review with some new proposals. Statistica Sinica, 8 (1998) 1-41.

[52] S. Fontdecaba, P. Grima, X. Tort-Martorell. Analyzing DOE with Statistical Software Packages: Controversies and proposals, The American Statistician 68 (2014) 205-211.

[53] R. V. Lenth. Quick and easy analysis of unreplicated factorials. Technometrics 31 (1989) 469–473.

[54] G. Derringer, R. Suich. Simultaneous optimization of several response variables, Journal of Quality Technology 12 (1980) 214-219.

[55] D.C. Hoaglin, F. Mosteller, J.W. Tukey, Understanding Robust and Exploratory Data Analysis, Wiley-Interscience, Hoboken, NJ; 2000.

[56] B. D. Cobb, J. M. Clarkson, A simple procedure for optimizing the polymerase chain reaction using modified Taguchi methods, Nucleic Acids Research 22 (1994) 3801-3805.

[57] C. M. Watrefall, B. D. Cobb, Single tube genotyping of sickle cell anaemia using PCR-based SNP analysis, Nucleic Acids Research 29 (2001) e119.

[58] G. A. Khoudoli, I. M. Porter, J. J. Blow, J. R. Swedlow, Optimisation of the two-dimensional gel electrophoresis protocol using the Taguchi approach, Proteome Science 2 (2004) 1-12.

[59] P. Thanakiatkrai, L. Welch, Using the Taguchi method for rapid quantitative PCR optimization with SYBR Green I, International Journal of Legal Medicine 126 (2012) 161-165.

[60] W. Dabrowski, U. Czekajlo-Kolodziej, D. Medrala, S. Giedrys-Kalemba, Optimization of AP-PCR fingerprinting discriminatory power for clinical isolates of Pseudomonas aeruginosa. FEMS Microbiology Letters 218 (2003) 51-57.

[61] J. D.Castaño, C. Cruz , E. Torres, Optimization of the production, purification and characterization of a laccase from the native fungus Xylaria sp., Biocatalysis and Agricultural Biotechnology 4 (2015) 710–716.

[62] S. Chakraborty, D. Mohanty, S. Ghosh, D. Das, Improvement of lipid content of Chlorella minutissima MCC 5 for biodiesel Production, Journal of Bioscience and Bioengineering 122 (2016) 294-300.

[63] J. Samuelsson, M. Leśko, M. Enmark, J. Högblom, A. Karlsson, K. Kaczmarski, Optimizing Column Length and Particle Size in Preparative Batch Chromatography Using Enantiomeric Separations of Omeprazole and Etiracetam as Models: Feasibility of Taguchi Empirical Optimization, Chromatographia 81 (2018) 851–860.

[64] V. Mandal, Y. Mohan, S. Hemalatha, Microwave assisted extraction of curcumin by sample–solvent dual heating mechanism using Taguchi L9 orthogonal design, Journal of Pharmaceutical and Biomedical Analysis 46 (2008) 322–327.

[65] R.B. Shah, Y. Yang, M.A. Khan, P.J. Faustino, Molecular weight determination for colloidal iron by Taguchi optimized validated gel permeation chromatography, International Journal of Pharmaceutics 353 (2008) 21–27.

[66] M. A. Tofighy, Y. Shirazi, T. Mohammadi, A. Pak, Salty water desalination using carbon nanotubes membrane, Chemical Engineering Journal 168 (2011) 1064–1072.

[67] S.-C. Wang, H.-J. Liao,W.-C. Lee, C.-M. Huang, T.-H. Tsai, Using orthogonal array to obtain gradient liquid chromatography conditions of enhanced peak intensity to determine geniposide and genipin with electrospray tandem mass spectrometry, Journal of Chromatography A, 1212 (2008) 68–75

[68] G. Zhu, Y. Li, H. Zhou, J. Liu, W. Yang, Microwave synthesis of high performance FAU-type zeolite membranes: Optimization, characterization and pervaporation dehydration of alcohols, Journal of Membrane Science 337 (2009) 47–54.

[69] S. M. Kim, K. S. Park, K. D. Kim, S. D. Park, H. T. Kim, Optimization of parameters for the synthesis of bimodal Ag nanoparticles by Taguchi method, Journal of Industrial and Engineering Chemistry 15 (2009) 894–897.

[70] Y. Chang, Y. Cao, J. Zhang, Y. Wen, Q. Ren, Purification of d-a-tocopheryl polyethylene glycol 1000 succinate (TPGS) by a temperature-modulated silica gel column chromatography: Use of Taguchi method to optimize purification conditions, Journal of Pharmaceutical and Biomedical Analysis 56 (2011) 804–808.

[71] Y. Liu, C. Liu, W. Liu, Y. Ma, S. Tang, C. Liang, Q. Cai, C. Zhang, Optimization of parameters in laser powder deposition AlSi10Mg alloy using Taguchi method, Optics and Laser Technology 111 (2019) 470–480.

[72] M. Savari, S. H. Z. Esfahani, M. Edalati, D. Biria, Optimizing conditions for production of high levels of soluble recombinant human growth hormone using Taguchi method, Protein Expression and Purification 114 (2015) 128–135.

[73] A. Nematollahzadeh, A. Dabaleh, N. Ahadi-Jomairan, S. Torabi. Iron-oxide nano-particles effect on the blood hemodynamics in atherosclerotic coronary arteries. Chemical Engineering Science 177 (2018) 293–300.

[74] A. E. Mofrad, A. Moheb, M. Masigol, M. Sadeghi, Farzaneh Radmanesh , An investigation into electrochemical properties of poly(ether sulfone)/poly(vinyl pyrrolidone) heterogeneous cation-exchange membranes by using design of experiment method, Journal of Colloid and Interface Science 532 (2018) 546–556.

HIGHLIGHTS

- Electrodialysis (ED) is an important chemo-process for polluted wastewater treatment
- Taguchi methods are essential for quick planning of ED chemometric trials
- Robust and agile chemometric methods are important for multiresponse multifactorial ED screening/optimization.
- We discuss several complications for robust decision-making on a real paradigm.

Declaration of interests

☒ The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

☐The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: