

# The Mathematical Gnostics (*Advanced Data Analysis*)

Pavel Kovanic<sup>(✉)</sup>

Institute of Information Theory and Automation of  
Czech Academy of Sciences, Prague, Czech Republic  
kovanic@tiscali.cz

**Abstract.** A brief survey of mathematical gnostics is presented. Mathematical gnostics is a tool of advanced data analysis, consisting of

1. theory of individual uncertain data and small samples,
2. algorithms to implement the theory,
3. applications of the algorithms.

The axioms and definitions of the theory are inspired by the Laws of Nature dealt with by physics and the investigation of data uncertainty follows the methods of analysis of physical processes. The first axiom is a reformulation of the measurement theory which mathematically formalizes the empirical cognitive activity of physics. This axiom enables the curvature of the data space to be revealed and quantified. The natural affinity between uncertain data and relativistic mechanics is also shown. Probability, informational entropy and information of individual uncertain data item are inferred from non-statistical Clausius' thermodynamical entropy. The quantitative cognitive activity is modeled as a closed cycle of quantification and estimation, which is proved to be irreversible and maximizes the result's information. A proper estimation of the space's curvature ensure a reliable robustness of the algorithms successfully proven in many applications. Gnostic formulae of data weights and errors, probability and information, which has been proved as valid for small samples of strongly uncertain data converge to statistical ones when uncertainty becomes weak. From this point of view, the mathematical gnostics can be considered as an extension of statistics useful under heavy-duty conditions.

## 1 A Gnostic System

The notion of the *gnostic system* has been applied in [1] to a general model of recognition characterized as the pairing of a real object and of a subject, its observer. The observation activity  $object \rightarrow subject$  is followed by the feedback  $subject \rightarrow object$  the purpose of which is using the evaluated information in

---

Motto: This approach is a deterministic theory of indeterminism.

Prof. J.A. VÍšek, Charles University, Prague.

P. Kovanic—Retired from the Institute of Information Theory and Automation of Czech Academy of Sciences.

manipulating, exploiting or control the object. In the special case of *quantitative recognition*, the observation represents the mapping of a real quantity onto numbers called *quantification*, the feed-back being the *estimation* of the true quantity's value. The necessity of quantification originated with the development of the market and the measuring became the task for physics. Mathematical modeling of counting and measuring – the measurement theory [2] – considers the quantification as a consistent mapping of structures of empiric quantities (sets endowed with some relations and operations) onto numeric structures. This theory deals with precise quantification only, leaving the treatment of imprecise quantification to mathematical statistics. Such a quantification process can be named *ideal quantification*.

## 2 Axiom of Real Quantification

As known from measurement theory, to ensure consistency of the ideal quantification, the relations between quantities and the operations on them must be subject to several logical conditions. This requirement was substituted in [3] by the idea of ideal quantification as the commutative (Abelian) group<sup>1</sup> If the real quantitative observation process would actually be the Abelian group, the estimation would be simply the inverse of this group. Unfortunately, real observations are disturbed by uncertain impacts. But these impacts are as real as the observed quantity. Moreover, their nature is the same: electrical measurements are subject to electrical disturbance. The uncertain impacts can thus be considered as countable or measurable sets and endowed with the same operation as the true observed quantities. Real quantification can be therefore modeled as a pair of two Abelian groups, one of the true and one of disturbing quantities. Considering one single quantitative observation, one actually obtains one single real number of the form of

$$A = A_0 + S\Phi \quad (1)$$

with the *true* real value  $A_0$ , real *uncertain* value  $\Phi$  and a positive dimensionless *scale parameter*  $S$ . Both  $A_0$  and  $\Phi$  are numerical images of elements of empirical structures forming Abelian groups. The multiplicative form of the additive relation 1 is obtained by exponentiation as

$$Z = Z_0 \exp(S\Phi) \quad (2)$$

Quantities  $A$  are real numbers and they can have both finite and infinite values when considered as theoretical objects. However, as numeric images of actual quantities, they have values within some finite bounds. This is why the theory involves regular transformations of the actual finite data domains onto the infinite domain to introduce and analyze the corresponding functions of data.

---

<sup>1</sup> Abel group is a set endowed with a binary operation satisfying the conditions of closedness, associativity, commutativity and existence of an identity element and of the inverse element to each element.

### 3 Geometry of Real Quantification, Quantifying Error and Weight

The observed data value  $Z$  (2) is represented by a point in the bi-dimensional plane  $(Z_0, S\Phi)$ . Observation is a discrete event, however let us consider the virtual path of a continuous variable  $z$  from the true value  $Z_0$  to the observed value  $Z$  under the impact of the uncertainty  $\varphi$  changing from the zero starting value to an unknown value  $\Phi$ . The length of this path is the observation error. A non-trivial question arises, which of many existing geometries is to be applied to quantify the error? Using the identity  $\exp(\alpha) = \cosh(\alpha) + \sinh(\alpha)$  and introducing hyperbolic Cartesian coordinates

$$x_Q = Z_0 \cosh(2S\Phi) \quad y_Q = Z_0 \sinh(2S\Phi) \quad (3)$$

one comes to the relation

$$Z_0 = \sqrt{(x_Q^2 - y_Q^2)} \quad (4)$$

The plane of observed data is thus endowed with the Minkowskian metric and the path of virtual movement is the Minkowskian circle. The number  $Z_0$  is a circle's radius and invariant of the movement. Multiplier 2 of Minkowskian angle  $S\Phi$  results from accepting the angular distance between  $\Phi$  and  $-\Phi$  (the mirrored point's angle) as the angular error. Relative coordinates

$$w_Q = \cosh(2S\Phi) \quad h_Q = \sinh(2S\Phi) \quad (5)$$

called *quantifying weight* and *quantifying irrelevance* have important interpretation in quantification of the uncertainty as data error weight and data error value. These names are motivated by the relation

$$\sinh(2S\Phi) = \int_0^{2S\Phi} \cosh(x) d(x) \quad (6)$$

where  $w_Q$  determines the weight of the differential data error  $d(2S\Phi)$  thus playing the role of metrical tensor in the sense of Riemannian geometry.

### 4 Geometry of Estimation, Estimating Error and Weight

An observer aims to use the best available way of measuring for quantification, but he must accept the observed value "as it is" without the chance of choosing the virtual quantification path determined by Nature. However, he knows from geometry, that the length of the quantifying path measured by Minkowskian geometry is an *extremal*: its length between two points exceeds the lengths of each of the other path between the same points. This means that the uncertainty  $S\Phi$  makes the observed value as bad as possible by maximizing its distance from the true value. The observer has a chance for his best "countermove" in his game with Nature by choosing the best virtual path of estimation from the known observed value  $Z$  back to the unknown true value  $Z_0$  thus minimizing the

resulting error. As shown in gnostic theory [4], such a path exists and its points have coordinates

$$x_E = Z_0 \cos(2S\varphi) \quad y_E = Z_0 \sin(2S\varphi) \quad (7)$$

where the relative coordinates

$$w_E = \cos(2S\varphi) \quad h_E = \sin(2S\varphi) \quad (8)$$

are *estimating weight* and *estimating irrelevance*, for which an analogue of the 6 exists. Thus this path has the form of Euclidean circle. It means that the observation plane is endowed by two metrics, quantifying (Minkowskian) and estimating (Euclidean) ones. The Euclidean angle  $\varphi$  is related with the Minkowskian ones  $\Phi$  by

$$\tan(S\varphi) = \tanh(S\Phi) \quad (9)$$

Thus each point of the observation plane has double interpretation, a quantifying and an estimating one.

## 5 Uncertainty and Curvature

The additive formulae 1 represents the quantity  $S\Phi$  as a cause of observed values' uncertainty. It is frequently used as an evaluation of the uncertainty's "size" and its square as an element of the data variance or data "weight". The latter notion has a classic statistical background. As proved in [10], the best asymptotically unbiased and asymptotically normally distributed estimate of the mean of differently dispersed data is a weighted mean where the weights are proportional to the reciprocal value of the data variance. It means, that measurement in different points of the observation space is to be done differently. In terms of Riemannian geometry: the metric tensor is a function of the coordinates of the space and that space is curved. "Locally dependent" metrics have been introduced into statistics as well by using the influence functions to improve robustness of the regression analysis [11]. There are many approaches to this task supported by statistical assumptions and tailored to different data classes. The influence functions derived from gnostic axioms were presented in [6].

Locally dependent metrics are also introduced by quantifying and estimating weights (5) and (8). Their non-linearity with respect to data is obviously exhibiting two types of forms, convex and concave. The scale parameter  $S$  is a function of the curvature's radius. It is closely connected with the robustness of the estimation of uncertainty.

## 6 Entropy of a Datum and Entropy Fields

C.E. Shannon's information entropy is the negative value of L. Boltzmann's statistical entropy. A complete system of probabilities of events is necessary for

the evaluation of this entropy. The pre-statistical concept of Clausius' thermodynamic entropy makes use only of the heat amount and the absolute temperature. A Gedanken-experiment helped in [5] to represent the entropy of a single uncertain datum in the Clausius' manner by introduction of the proportional mapping of the squared data value onto the absolute temperature and onto the heat flow. Substitution into Clausius' formula shows that the changes of the thermodynamic entropy of an uncertain datum within quantification and estimation is proportional to the changes of the corresponding data weights,

$$\delta\mathcal{E}_Q = w_Q - 1 \quad \delta\mathcal{E}_E = w_E - 1 \tag{10}$$

if the coefficients of proportionality of the mapping are suitably chosen. The plane of observation is formed by possible data values, each of which has its quantifying and estimating weight attached. Formulae 10 therefore define two scalar fields of entropy. Gradients of these fields can be shown to be proportional to the corresponding irrelevances  $h_Q$  and  $h_E$ .

## 7 Information and Probability of an Individual Datum

The source of a scalar field  $\mathcal{E}$  is known to result from the operation  $\text{div grad } \mathcal{E}$ , i.e. by application of the Laplace's operator  $\Delta$ . Looking for the source of the entropy field of  $\mathcal{E}_Q$  in the point  $(x, y)$  one comes to relation

$$(x^2 + y^2)\Delta\mathcal{E}_Q = \frac{1}{p * (1 - p)} \tag{11}$$

where

$$p = (1 - h_E)/2 \tag{12}$$

Introducing the quantity

$$\mathcal{I}(p) = -p * \ln(p) - (1 - p) * \ln(1 - p) \tag{13}$$

one has the relation

$$\frac{1}{p * (1 - p)} = \frac{d^2(\mathcal{I}(1/2) - \mathcal{I}(p))}{dp^2} \tag{14}$$

saying that the right hand side of Eq. 11 is a source of the field of  $\mathcal{I}$ . The quantity  $\mathcal{I}(p)$  would be formally identical with the Shannon's information of an event, the probability of which would be  $p$ . Moreover, there is a large set of conditions in [7] under which a quantity is to be accepted as information. As shown in [4], all such conditions are satisfied by  $\mathcal{I}$  which thus deserves to be accepted as *information of an individual uncertain datum* and its argument  $p$  as the datum's *probability*. Equation 11 can be thus formulated as a general statement: *The source of entropy of an individual uncertain datum is proportional to the source of its information. This equation describes the conversion of entropy to*

information and vice versa. It thus can be considered to be a mathematical model of the Maxwell's demon<sup>2</sup>

## 8 Ideal Gnostic Cycle and its Features

The observed point, interpreted by quantifying coordinates  $(x_Q, y_Q)$  or by estimating ones  $(x_E, y_E)$  has its mirrored image  $(x_Q, -y_Q)$  and  $(x_E, -y_E)$ . They are two arcs of virtual paths connecting the observed points with their mirrored images, the “hyperbolic” arc of a Minkowskian circle and an “ordinary” (Euclidean) ones. This closed path is called the *Ideal Gnostic Cycle (IGC)*. Changes of entropy and information of a datum 10 and 14 enable the important features of the *IGC* to be proved:

- [A] Data transformations following the closed path of *IGC* provide the best estimate of the true value in the sense of maximization of results' information and minimization of its entropy.
- [B] The closed *IGC* is irreversible: none estimation can completely eliminate the error of an uncertain observation.

Thus the *IGC* according to [A] provides a theoretical model for programs of estimation, but establishes by [B] unsurpassable limits for data analysis like the second law of thermodynamics does for heat transformation.

## 9 What Should Data Say for Themselves

The ideal of data treatment frequently formulated as “Let data speak for themselves!” resulted from the requirement of maximum objectivity. The more a priori assumptions on data, the more subjectivity is increasing the danger of discrepancy between assumed models and actual features of data. The goal of data treatment is information being brought by data, but reaching it is critically limited by the knowledge of data features. This knowledge requires answering a series of questions:

- What kind of geometry should be applied (Euclidean, Minkowskian, Riemannian)?
- What curvature of the space of uncertain data characterizes the given data?
- Is the data structure additive or multiplicative?
- Are the data homo- or heteroscedastic?
- Is there a data trend?
- Are the data cross- or autocorrelated?
- Are the data homogeneous?
- What is form of the probability and density distribution?

Some of these questions are not asked in statistics, others are answered by assumptions. Mathematical gnostics derives all the answers from data. The crucial point is the robust kernel estimation of probability distributions.

<sup>2</sup> As described in [8], Maxwell introduced in his Gedanken-experiment a virtual creature capable to use the information on movement of molecules to convert it into decrease of entropy.

## 10 The Unique Kernel for Robust Kernel Estimation

The kernel estimation of a probability density function was introduced in [9] along with five conditions necessary for asymptotic convergence to true density. A lot of kernels can be found in literature satisfying these conditions and giving estimates of different quality dependent on the kernel's form. Kernels are ordinarily defined over the domain of the independent variable using their natural additive or multiplicative scale. Unlike this, the individual data item's probability 12 is defined over the infinite (positive) domain obtained by transformation of the actual data domain. Its density was shown to satisfy all Parzen's conditions. Its application to kernel estimation is not only justified, but advantageous: its form is universally applicable and as a result of the theory, it is unique and optimal. The location of the kernel is determined by the (known) observed value and its "width" by the scale parameter  $S$  which is to be estimated by data.

## 11 Aggregation of Kernels

The Parzen's kernel estimating method creates the density estimate by additive aggregation of kernels without the consideration of any alternatives. It may seem natural, because the historical mathematical forerunners of kernel estimation like Green and Duhamel<sup>3</sup> did essentially the same because of linearity. However, the aggregation of gnostic kernels deserves a special consideration. The space of observed data within the quantification process has been shown as a Minkowskian plane with coordinates proportional to the hyperbolic cosine ( $w_Q$ ) and the hyperbolic sine ( $h_Q$ ). But a two-dimensional plane depicting the moment and energy of a relativistic charge-free particle would be endowed by the same geometry. This means that there exists (at least mathematically) a consistent linear mapping of the pair  $(w_Q, h_Q)$  onto the pair of *(energy, momentum)* of a relativistic particle moving with velocity corresponding to the argument of said hyperbolic functions. Moreover, this mapping is Lorentz-invariant, i.e. it is valid for all data uncertainties and corresponding particle's velocities. This mapping *uncertain data*  $\Leftrightarrow$  *relativistic particle* can be applied to several data. The aggregation law of relativistic particles is known, it is the Momentum-Energy Conservation Law, which is additive with respect to pairs *(energy, momentum)*. To preserve the mapping for a data set, one must aggregate the pairs  $(w_Q, h_Q)$  additively as well, although they are nonlinear functions of data. The second axiom of the gnostic theory extends this way of aggregating from quantifying weights and irrelevances to estimating ones to preserve the mapping of quantifying variables to estimating ones and vice versa.

However, a sum of cosine is not a cosine and a sum of sines is not a sine. Therefore, sum of weights (and irrelevances) of a data set will represent the weight (irrelevance) of the whole set but not a pair *(weight, irrelevance)* of a possible single data item. This is why a proper normalization of additively

---

<sup>3</sup> See Green's function and Duhamel's integral.

aggregated weights and irrelevances should be applied instead of their simple addition.

The form of both quantification and estimation kernels can be shown to be similar, differing only by scale parameters. However, the results of aggregation of kernels depend on metric.

## 12 Applications of the Gnostic Kernels

The kernel presented above was the derivative of probability. The linearity of this operation allows us to obtain and use kernels of both density and probability. Library of gnostic algorithms includes the following applications of kernel estimation:

### 12.1 Local Probability and Density Distribution

Local distributions are obtained as means of kernels. They possess a full flexibility controlled by the choice of scale parameter. This feature make them an ideal instrument for revealing the detailed structure of a data set and to perform the marginal analysis showing the data clusters and outliers in a non-homogeneous data set. A special kind of inner robustness of these distributions allows a deep insight into a homogeneous data set to be obtained allowing robust bounds of its important subintervals to be estimated.

### 12.2 Global Probability and Density Distribution

Each pair (*weight, irrelevance*) has its module determined as the Minkowskian or Euclidean length of the observed point's radius vector. The global probability distribution function is obtained as the mean of integral kernels divided by the module of sums of cosines and sines by using the proper metric. The global density distribution is the first derivative of the global probability distribution. There are two types of the global distributions differing by robustness, the estimating one is robust with respect to outlying data and peripheral clusters while the quantifying distribution is robust with respect to inner disturbances and noises of the treated data sample. Unlike the high flexibility of the local distribution functions, the global ones are more rigid. This feature makes them applicable to robust probability and density estimation, to reliable tests of data homogeneity and to estimation of the observed data's true values and of bounds of data support. Global distribution enable three types of censored data to be estimated: both left- and right censored ones and interval data.

The advantage of gnostic distribution functions lies in their independence on the a priori assumptions, objectivity due to the reliance only on data alone, suitability for small data samples and a much broader application field than standard statistical distributions including small data samples.



### 12.3 Robust Curve Fitting

Frequently used curve fitting by means of polynomials or by sets of other functions including the orthogonal ones can suffer from un-robustness in the case of application to uncertain data. A careful preliminary gnostic analysis providing reliable estimates of individual data weights, proper geometry and scale parameters of gnostic kernels used for the fit enable the maximum of resulting information to be reached.

### 12.4 Analysis of Dependencies

The correlation coefficient can say that there exists an interdependence between two vectors, but the interpretation of the interaction is easy only in the case close to the linear relationship. The application of kernel estimates is suitable especially for presentation of non-monotonous dependencies.

## 13 Robust Regression

The approach to the task of robust multi-dimensional regression modeling based on mathematical gnostics has been demonstrated in [6]. The gist was the choice of a criterion function for the evaluation of model's residuals. Instead of some formal "purely mathematical" functions, natural features of uncertainty were used such as the common source of fields of entropy and information. Results were shown to be applicable as special kinds of influence functions used in robust statistics for the Iterated Weighted Least Squares Method completed by a feed-back filter. Extensive comparisons with statistical models of this type demonstrated the priority of the gnostic approach resulting in better estimates of curvature of the space of uncertain data and in information maximization of the estimation process.

The standard case of a regression model representing the dependent vector as a linear combination of explanatory vectors can be called *explicit*. The *implicit* regression model is obtained from the explicit one by division of all equations of the system by the values of the dependent variable (which must be non-zero). There are some advantages of the implicit regression, e.g. uniqueness of the model independent on the exchange of roles of explanatory/dependent variables, comparability and evaluation of relative impacts of variable.

## 14 Robust Correlation

The availability of reliable robust regression techniques enabled a new approach to robust correlation coefficients to be introduced. The proportionality between two centered vectors  $x$  and  $y$  is considered twice, as  $x = c \cdot y$  and  $y = k \cdot x$  with scalars  $c$  and  $k$ , which are estimated by the robust regression. The square root of products of estimates can be used as robust estimate of the correlation coefficient.

## 15 Testing of Hypotheses

The crucial problem of statistical testing of hypotheses is the decision making on the probability distribution of the underlying data. Some statistical tests are based on the Gaussian assumption, but experienced analyst know that the “normal” distribution is not always normal. Relying on a priori assumptions as well as data violation by “normalizing” transformations can lead to incorrect decision making. The availability of robust probability distributions described above allows not only tests on a required significance level to be performed, but actual significance of the required decision to be evaluated.

## 16 Homogeneity Problem

Problems with data homogeneity can be demonstrated on the task of statistical investigation of political preferences. The careful selection of people used as a data source cannot warrant the homogeneity (similarity, affinity, comparability, closeness) of meaning of all individuals or groups. There is an extensive amount of factors influencing the measurable parameters. All these factors cannot be under the control of the survey’s organizers. Increasing the survey’s size can be even counter-productive: the more cases, the broader the spectrum of factors. Moreover, it is not always safe to assume that the demanding conditions of Central Limit Theorem are satisfied.

A non-homogeneity of a one-dimensional data set is sensitively and reliably detected by the appearance of a second maximum in gnostic global density distribution. This enables a reliable homogenization to be performed.

## 17 Robust Cluster Analysis

The local distribution functions enable the homogenization of a one-dimensional non-homogeneous data sample to be implemented by the identification of outliers, inliers and sub-clusters causing the non-homogeneity. Thus a non-homogeneous data set consisting of several homogeneous clusters can be subjected to robust marginal analysis. This approach is efficiently generalized to robust multi-dimensional cluster analysis by a marginal analysis of residuals of an implicit multi-dimensional model. A multi-dimensional non-homogeneous data set is then replaced by several homogeneous clusters.

The robustness of this approach enables the multi-dimensional objects (represented by rows of the model) to be ordered in a rational and reliable way.

## 18 Implementation

Methods of mathematical gnostics have been implemented as computer programs during the last several decades and the implementation efforts are continuing today, as well. Their application in many fields, including technology,

economy, medical, environmental investigation and others, were used not only for tests of their efficacy, but also as motivation and initiation of further development. The long-term experience confirms the usefulness of this approach to uncertainty. Many applications (especially to economic problems like financial statement analysis and financial control, marketing and financial markets) are described in [12]. The gnostic methodology of analysis of environmental parameters was investigated within the framework of two research projects of the European Union [13, 14]. Programs based on mathematical gnostics became the main data analytical tool in the Institute of Chemical Process Fundamentals of the Czech Academy of Sciences as documented by series of publications (e.g. [15–18]). Recent results enable a complete automation of the exploratory phase of data analysis providing robust information on actual data model, which offers the rising of the quality assessment control to the level unreachable by other methods ([19]).

## 19 Conclusions

Mathematical gnostics, which is based on the axiomatic theory of individual uncertain data and small samples and supported by laws of physics develops advanced methods for the treatment of strongly uncertain data. These methods maximize the resulting information and are naturally robust. Their applications also extend the range of tasks solvable by statistical methods.

## References

1. Rastrigin, P.A., Markov, V.A.: *Cybernetic Models of Recognition*. Zinatne, Riga (1976). (in Russian)
2. von Helmholtz, H.: *Zahlen und Messen erkenntniss-theoretisch betrachtet*. In: *Philosophische Aufsätze Eduard Zeller Gewidmet*, Leipzig, pp. 17–52(1887). (in German)
3. Kovanic, P.: Gnostical theory of individual data. *Prob. Control Inform. Theory* **13**, 259–271 (1984)
4. Kovanic P.: *Gnostická teorie neurčitých dat*, (Gnostic Theory of Uncertain Data), doctor (DrSc.) dissertation, The Institute of Information Theory and Automation, Czechoslovak Academy of Sciences, Prague, 161 pp. (1990). (in Czech)
5. Kovanic, P.: On relations between information and physics. *Prob. Control Inf. Theory* **13**, 383–399 (1984)
6. Kovanic, P.: A new theoretical and algorithmical basis for estimation. *Ident. Control Autom.* **V22**(6), 657–674 (1986)
7. Perez, A.: *Mathematical theory of information*. *Appl. Math.* **3**(1), 81–99 (1958). (in Czech)
8. von Bayerer, H.C.: *Maxwell's Demon*. Random House Inc., New York (1998)
9. Parzen, E.: On estimation of a probability density function and mode. *Ann. Math. Statist.* **33**, 1065–1076 (1962)
10. Linnik, Y.V.: *The Least Square Method and Basics of Observation Treatment*. GIM-FL, Moscow (1962). (in Russian)

11. Coleman, D., Holland, P., Kaden, N., Klema, V., Peters, S.C.: A system of sub-routines for iteratively re-weighted least-squares computation. *ACM Trans. Math. Softw.* **6**, 327–336 (1980)
12. Kovanic, P., Humber, M.B.: *The Economics of Information*, 717 p. (2015). [www.math-gnostics.com](http://www.math-gnostics.com)
13. Focus on Key Sources of Environmental Risk. [www.projectfoks.eu](http://www.projectfoks.eu)
14. European Project 2-FUN: Improving Risk Assessment. [www.2-fun.org](http://www.2-fun.org)
15. Jacquemin, J., Bendová, M., Sedláková, Z., Holbrey, J.D., Mullan, C.L., Youngs, T.G.A., Pison, L., Wagner, Z., Aim, K., Costa Gomes, M.F., Hardacre, C.: Phase behaviour, interactions, and structural studies of (Amines+Ionic Liquids) binary mixtures. (Eng) *Chem. Phys. Chem.* **13**(7), 1825–1835 (2012)
16. Borsós, T., Řimnáčová, D., Ždímal, V., Smolík, J., Wagner, Z., Weidinger, T., Burkart, J., Steiner, G., Reischl, G., Hitzemberger, R., Schwarz, J., Salma, I.: Comparison of particulate number concentrations in three central european capital cities. (Eng) *Sci. Total Environ.* **433**, 418–426 (2012)
17. Setničková, K., Wagner, Z., Noble, R., Uchytíl, P.: Semi-empirical model of toluene transport in polyethylene membranes based on the data using a new type of apparatus for determining gas permeability, diffusivity and solubility. (Eng) *J. Membr. Sci.* **66**(22), 5566–5574 (2011)
18. Andresová, A., Storch, J., Traikia, M., Wagner, Z., Bendová, M., Husson, P.: Branched and cyclic alkyl groups in imidazolium-based ionic liquids: molecular organization and physicochemical properties. (Eng) *Fluid Phase Equilib.* **371**, 41–49 (2014)
19. Wagner, Z., Kovanic, P.: *Advanced Data Analysis for Industrial Applications, Modelling Smart Grids 2015*, Prague, September 10–11 (2015). <http://www.smartgrids2015.eu/>