BIOTECHNO 2018

The Tenth International Conference on Bioinformatics, Biocomputational Systems and Biotechnologies

ISBN: 978-1-61208-639-2

BIONATURE 2018

The Ninth International Conference on Bioenvironment, Biodiversity and Renewable Energies

May 20 - 24, 2018

Nice, France

BIOTECHNO 2018 Editors

Birgit Gersbeck-Schierholz, Leibniz Universität Hannover, Germany
Valentinas Klevas, Lithuanian Energy Institute, Lithuania

BIOTECHNO 2018

Foreword

The Tenth International Conference on Bioinformatics, Biocomputational Systems and Biotechnologies (BIOTECHNO 2018), held between May 20 - 24, 2018 - Nice, France, covered these three main areas: bioinformatics, biomedical technologies, and biocomputing.

Bioinformatics deals with the system-level study of complex interactions in biosystems providing a quantitative systemic approach to understand them and appropriate tool support and concepts to model them. Understanding and modeling biosystems requires simulation of biological behaviors and functions. Bioinformatics itself constitutes a vast area of research and specialization, as many classical domains such as databases, modeling, and regular expressions are used to represent, store, retrieve and process a huge volume of knowledge. There are challenging aspects concerning biocomputation technologies, bioinformatics mechanisms dealing with chemoinformatics, bioimaging, and neuroinformatics.

Biotechnology is defined as the industrial use of living organisms or biological techniques developed through basic research. Bio-oriented technologies became very popular in various research topics and industrial market segments. Current human mechanisms seem to offer significant ways for improving theories, algorithms, technologies, products and systems. The focus is driven by fundamentals in approaching and applying biotechnologies in terms of engineering methods, special electronics, and special materials and systems. Borrowing simplicity and performance from the real life, biodevices cover a large spectrum of areas, from sensors, chips, and biometry to computing. One of the chief domains is represented by the biomedical biotechnologies, from instrumentation to monitoring, from simple sensors to integrated systems, including image processing and visualization systems. As the state-of-the-art in all the domains enumerated in the conference topics evolve with high velocity, new biotechnologes and biosystems become available. Their rapid integration in the real life becomes a challenge.

Brain-computing, biocomputing, and computation biology and microbiology represent advanced methodologies and mechanisms in approaching and understanding the challenging behavior of life mechanisms. Using bio-ontologies, biosemantics and special processing concepts, progress was achieved in dealing with genomics, biopharmaceutical and molecular intelligence, in the biology and microbiology domains. The area brings a rich spectrum of informatics paradigms, such as epidemic models, pattern classification, graph theory, or stochastic models, to support special biocomputing applications in biomedical, genetics, molecular and cellular biology and microbiology. While progress is achieved with a high speed, challenges must be overcome for large-scale bio-subsystems, special genomics cases, bio-nanotechnologies, drugs, or microbial propagation and immunity.

Special aspects closely related to the natural aspects and environment, like biodiversity and invasion, renewable and sustainable energies etc. are tackled within BIONATURE 2018.

BIOTECHNO 2018 also featured the following:

- BIONATURE 2018: The Ninth International Conference on Bioenvironment, Biodiversity and Renewable Energies

We take here the opportunity to warmly thank all the members of the BIOTECHNO 2018 Technical Program Committee, as well as the numerous reviewers. The creation of such a high quality conference program would not have been possible without their involvement. We also kindly thank all the authors who dedicated much of their time and efforts to contribute to BIOTECHNO 2018.

Also, this event could not have been a reality without the support of many individuals, organizations, and sponsors. We are grateful to the members of the BIOTECHNO 2018 organizing committee for their help in handling the logistics and for their work to make this professional meeting a success.

We hope that BIOTECHNO 2018 was a successful international forum for the exchange of ideas and results between academia and industry and for the promotion of progress in the fields of bioinformatics, biocomputational systems and biotechnologies.

We also hope that Nice provided a pleasant environment during the conference and everyone saved some time for exploring this beautiful city.

BIOTECHNO 2018 Chairs:

BIOTECHNO Steering Committee

Gilles Bernot, University Nice Sophia Antipolis, France
Birgit Gersbeck-Schierholz, Leibniz Universität Hannover, Germany
Hesham H. Ali, University of Nebraska at Omaha, USA
Erliang Zeng, University of South Dakota, USA
Y-h. Taguchi, Chuo University, Japan
Hunter Moseley, University of Kentucky, USA
Magnus Bordewich, Durham University, UK

BIOTECHNO Industry/Research Advisory Committee

Steffen Heber, North Carolina State University, USA
Alexandru Floares, SAIA Institute, Romania

BIONATURE Advisory Committee

Suhkneung Pyo, Sungkyunkwan University | School of Pharmacy, Republic of Korea
Vladimir Strezov, Macquarie University - Sydney, Australia
Silvia Paola Assini, University

BIOTECHNO 2018

Committee

BIOTECHNO Steering Committee

Gilles Bernot, University Nice Sophia Antipolis, France
Birgit Gersbeck-Schierholz, Leibniz Universität Hannover, Germany
Hesham H. Ali, University of Nebraska at Omaha, USA
Erliang Zeng, University of South Dakota, USA
Y-h. Taguchi, Chuo University, Japan
Hunter Moseley, University of Kentucky, USA
Magnus Bordewich, Durham University, UK

BIOTECHNO Industry/Research Advisory Committee

Steffen Heber, North Carolina State University, USA
Alexandru Floares, SAIA Institute, Romania

BIOTECHNO 2018 Technical Program Committee

Antonino Abbruzzo, Università degli Studi di Palermo, Italy
Hesham H. Ali, University of Nebraska at Omaha, USA
Joel P. Arrais, University of Coimbra, Portugal
Erich J. Baker, Baylor University, USA
Andrés Becerra Sandoval, Independent researcher, Colombia
Vincenzo Belcastro, Philip Morris International, Neuchatel, Switzerland
Kais Belwafi, King Saud University, Kingdom of Saudi Arabia
Gilles Bernot, University Nice Sophia Antipolis, France
Razvan Bocu, Transilvania University of Brasov, Romania
Magnus Bordewich, Durham University, UK
Klaus Brinker, Hamm-Lippstadt University of Applied Sciences, Germany
Rita Casadio, University of Bologna, Italy
Matthias Chung, Virginia Tech, USA
Peter Clote, Boston College, USA
Jean-Paul Comet, University Nice Sophia Antipolis, France
Qin Ding, East Carolina University, USA
Maria Evelina Fantacci, University of Pisa, Italy
Alexandru Floares, SAIA Institute, Romania
Sebastian Fudickar, University of Oldenburg, Germany
Said Gaci, Sonatrach, Algeria
Xin Gao, King Abdullah University of Science and Technology (KAUST), Saudi Arabia
Birgit Gersbeck-Schierholz, Leibniz Universität Hannover, Germany
Radu Grosu, Vienna University of Technology, Austria
Steffen Heber, North Carolina State University, USA
Asier Ibeas, Universitat Autònoma de Barcelona, Spain
Sohei Ito, National Fisheries University, Japan
Filip Jagodzinski, Western Washington University, USA

BIONATURE Advisory Committee

Vladimir Strezov, Macquarie University - Sydney, Australia
Silvia Paola Assini, University

BIONATURE 2018 Technical Program Committee

Silvia Paola Assini, University of Pavia, Italy
Dhinesh Balasubramanian, Mepco Schlenk Engineering College, Tamil Nadu, India
Yoseph Bar-Cohen, Jet Propulsion Lab (JPL), NASA
Benjamin Carbonnier, University Paris-East, France
Longjian Chen, China Agricultural University, China
Pedro Luiz Corrêa Pizzigatti is Escola Politécnica da USP (EPUSP), Brazil
Hany A. El-Shemy, Cairo University, Egypt
Mahdi Esfahanian, Florida Atlantic Univerisity, USA
Mohammad Ahsanul Haque, Aalborg University, Denmark
Valentinas Klevas, Lithuanian Energy Institute, Lithuania
Man-Kee Lam, Universiti Teknologi PETRONAS, Malaysia
Ana Jesús López Menéndez, Universidad de Oviedo, Spain
José Augusto Medrano Hernandez, Centro de Estudio de Tecnologías Energéticas Renovables (CETER) - CUJAE, Cuba
Son V. Nghiem, Jet Propulsion Laboratory, California Institute of Technology - Pasadena, USA
Constantin Paleologu, University Politehnica of Bucharest, Romania
Suhkneung Pyo**,** Sungkyunkwan University | School of Pharmacy, Republic of Korea
Francisca Segura, University of Huelva, Spain
Luciano Silva, Universidade Federal do Paraná, Brazil
Vladimir Strezov, Macquarie University - Sydney, Australia
Gökçen Firdevs Yücel, İstanbul Aydın University, Turkey

Copyright Information

Table of Contents

Predictive Analytics to Determine the Potential Occurrence of Genetic Disease and their Correlation: Osteoporosis and Cardiovascular Disease

Kae Sawada
Computer Science Department
California State University, Los Angeles
Los Angeles, USA
ksawada@calstatela.edu

Michael W. Clark
Biology Department
Pasadena City College
Pasadena, USA
mclark7@pasadena.edu

Zilong Ye
Computer Science Department
California State University, Los Angeles
Los Angeles, USA
zye5@calstatela.ed

Nabil Alshurafa
Medicine and of Computer Science Department
Northwestern University
Evanston, USA
nabil@northwestern.edu

Mohammad Pourhomayoun
Computer Science Department
California State University, Los Angeles
Los Angeles, USA
mpourho@calstatela.edu

***Abstract*— In this paper, a Predictive Analytics Model is designed, developed, and validated to determine the risk of manifesting osteoporosis in later life using big data processing. The proposed model leverages the novel genetic pleiotropic information in the 1,000 Genome Project of over 2,500 individuals world-wide. Also, the mutations associated with osteoporosis and cardiovascular disease are specifically analyzed. The study proposes the automatic histogram clustering as an effective and intuitive visualization method for high dimensional dataset. The results demonstrate a significant correlation between a person's regional background and the frequency of occurrence of the 35 Single Nucleotide Polymorphisms (SNPs) associated with osteoporosis and/or cardiovascular disease (CVD). Machine learning algorithms, such as Logistic Regression, Adaboost, and KNN are then applied to predict the occurrence of 7 osteoporosis-related-SNPs based on the existing CVD-related-SNPs input. Finally, the developed model is evaluated using a separate dataset obtained through Affymetrix microarray mRNA expression signal values for the specific SNP(s) in individuals with and without osteoporosis.**

Keywords- osteoporosis; Predictive Model; Genome Wide Association Study (GWAS); Clustering.

I. Introduction

This research is a continuation of an investigation previously presented in CSCI 2017 [34] by the same authors. The previous paper examined the dataset obtained from the 1000 Genome Project for its mutation variants with known relations to osteoporosis and cardiovascular disease. This paper proposes an novel approach to automatic data-driven clustering of histogram presented data for verification and validation of disease related expression in different human populations. As explained in later sections, high dimensional datasets can be effectively and intuitively visualized by the algorithm generating the histogram clustering. This automated process could aid in understanding existing correlations among various types of large datasets. Here, this automated clustering reproduced the manual clustering of the disease related SNPs geographic profiles. This study also confirmed the feasibility of the previously proposed predictor through a set of microarray analysis result dataset.

A. Osteoporosis and Space Exploration

Fractures, as a result of osteoporosis, have a significant negative impact on an individual's health, quality of life, and work performance. Some modern occupations inevitably expose workers to a significantly increased risk of developing osteoporosis. An obvious example is space flight. Within a few days of Zero Gravity (zero G), astronauts begin to lose both muscle and bone mass. There are also a zero G suppression effect on the immune system, as well as an increase in the aging process of particular cells. These effects occurred for all individuals who exposed themselves to zero G environment [25]. It is reasonable to assume that a predisposition to osteoporosis might increase the rate of occurrence of the condition. Having a reasonably reliable predictive model to reveal the predisposition for osteoporosis could be of great benefit for the future commercialization of space.

Over 7 TB of genomic data has been sequenced by the 1000 Genome Project [1]. These datasets consist of over 2,500 individuals from all around the world [2]. Such large genomic datasets allow for the study of complex disease states, such as osteoporosis and cardiovascular disease. These pathologies involve not only multiple gene mutation interactions, but also nutritional and age-related components. [9][20].

There are over 100 genes that are proven to be related to osteoporosis, according to Genome-Wide Association Studies (GWAS) [7]. However, many of these gene mutations still maintain some level of phenotypic ambiguity [22]. Understanding and treating such complex conditions could benefit through modern Machine Learning analysis. An effective method for early forecast of a patient-specific frequency of occurrence for a particular genetic disorder in later life would allow for appropriate preventative measures to be followed.

B. Mutations: SNPs

This paper analyzes 35 SNPs, that are commonly observed as genetic disease related mutations. It is a mutation of one base for another, which occurs in more than one percent of the general population [14]. 7 of these SNPs have direct indications in the expression of osteoporosis [9], while the other 28 SNPs have implications in both CVD and osteoporosis [4].

C. Preceding Related Work

i. Osteoporosis-related-SNP selection

The 7 osteoporosis-related-SNPs were chosen based on the study led by Hsu et al [9], published in 2010. The phenotypic association of these SNPs to osteoporosis was demonstrated by the GWAS study [26].

ii. Genetic Pleiotropy

Using False Discovery Rate (FDR) statistical methods, Reppe et al [4] revealed a potential genetic link between Cardio-Vascular Disease (CVD) and Osteoporotic conditions. In this paper, the potential mutant gene interactions between the osteoporosis-related SNPs and CVD SNPs are analyzed with: big data processing and analytics, predictive analytics based on machine learning algorithms, data visualization, and clustering.

This paper is organized as follows: in Section II, the methods and materials employed during the analysis and experimental developments are introduced. Included topics are the datasets and methods used, feature selections and dataset label, and evaluation methods. In Section III, the results and observations are discussed.

II. Methods and Materials

This section discusses the datasets employed, the design of the analysis and research, algorithms and techniques utilized in the study so far.

A. Datasets

Two big datasets have been used in this paper:

The first one is the genotypes from the 1000 Genome Project for the 35 SNPs. This dataset includes 2504 human subjects, containing both male and female from 26 regions worldwide [2].

The second dataset is from Reppe et al's study mentioned in Section I-C-ii [28]. The samples of this dataset are collected from the 84 post-menopausal females between the ages of 50 to 86 years old in Lovisenberg Diakonale Hospital located in Sweden. There are two components to this dataset:

(1) The result of Affymetrix Microarray analysis of the patients, the Affymetrix microarray signal values per sample, one ".CELL" file per patient.

(2) A set of biopsy results, sample ID (anonymous), age, gender, and the biopsy results of bone density scores, both T- and Z-scores, consisting of average neck, total hip, and average spine of each subject.

In this paper, all ".CELL" files from (1) were processed and interfaced with the library, pd.hg.u133.plus.2 [27], to obtain the gene symbol per an array cell. Then, it was interfaced with SNP identifier (e.g., #rsxxx..x, with x composing an integer) to acquire the existing SNP(s) data per each sample. In acquisition of SNP existence, the signal threshold value was determined the average score of all samples per column (per array cell). Such determination was made based on the preceding study which defined signal thresholds in DNA microarray analysis [29].

The label per sample was acquired based on the data from (2). Samples with the T-scores of -2.5 or less in one or more of neck, hip, and spine are marked as osteoporotic. The threshold was determined, following the World Health Organization (WHO) international reference standard for osteoporosis diagnosis [30]. T-scores were used instead of Z-scores, based on the sample's age and the guidelines provided by WHO.

B. Method Design

i. Problem Definition

The previous study investigated the occurrence of 7 selected osteoporosis-related SNPs [3] and its correlation to 28 CVD with 6 subdivisions, related SNPs [4], shown in Table 1 and 2. The CVD related SNPs were divided into six subcategories: High Density Lipids proteins cholesterol (HDL), Low Density Lips proteins cholesterol (LDL), Systolic Blood Pressure (SBP), Diatonic Blood Pressure (DBP), Type 1 Diabetes (T1D), and Triglycerides (TG). The table of osteoporosis-related SNPs describes the SNP identification number, normal (ancestral) base, high-risk (mutated) base, and the homozygous base pairs that are associated with a high risk of Bone Mineral Density (BMD) loss, and consequently the development of osteoporosis. (See Section I-C-i for osteoporosis-related SNPs). In this paper, we also validate the predictor design by utilizing the new dataset introduced in Section I-A.

ii. Feature Selection and Label

As mentioned in the Section II-C-ii, 28 CVD-related SNPs, and age of each subject were fed to the predictive model. The output of this predictor is a boolean value, whether the individual has developed osteoporosis or not.

iii. Predictive Analytics Algorithms

To develop a software method for predicting a tendency for disease occurrence, various machine learning algorithms were tried for building the predictive model: KNN (the best value of K determined by trial and error was K=9), Logistic Regression, Decision Tree, Naive Bayes, Adaboost, Random Forest, and Support Vector Machine (the best parameters of SVM determined by Grid Search include C=1, kernel='rbf', gamma=0.0007). A systematic aggregation of these classifier results is future work, for example using an ensemble learning algorithm.

TABLE 1: SNP ASSOCIATED WITH OSTEOPOROSIS

SNP (rs ID)	Ancestral allele	Mutated allel	Possible pair	High Risk Genotype	Phenotype (associated condition)
rs2278729	G	A	AA, GG, AG, GA	AA	Osteoporosis
rs12808199	A	G	AA, GG, AG, GA	GG	Osteoporosis
rs7227401	G	T	GG, TT, GT, TG	TT	Osteoporosis
rs494453	T	C	TT, CC, TC, CT	CC	Osteoporosis
rs12151790	G	A	AA, GG, AG, GA	AA	Osteoporosis
rs2062375	C	G	CC, GG, CG, GC	GG	Osteoporosis
rs17184557	T	A	TT, AA, TA, AT	AA	Osteoporosis

TABLE 2: SNP ASSOCIATED WITH CVD

SNP (rs ID)	Ancestral allele	Mutated allel	Possible pair	High Risk Genotype	Phenotype (associated condition)
rs4957742	A	G	AA, GG, AG, GA	GG	DBP
rs665556	C	T	TT, CC, TC, CT	TT	DBP
rs10779702	A	G	AA, GG, AG, GA	GG	HDL
rs12137389	T	C	TT, CC, TC, CT	CC	HDL
rs9309664	G	A	AA, GG, AG, GA	AA	HDL
rs7594560	T	C	TT, CC, TC, CT	CC	HDL
rs10953178	C	T	TT, CC, TC, CT	TT	HDL
rs980299	T	C	TT, CC, TC, CT	CC	HDL
rs10746070	T	C	TT, CC, TC, CT	CC	HDL
rs7175531	C	T	TT, CC, TC, CT	TT	HDL
rs3198697	C	T	TT, CC, TC, CT	TT	HDL
rs756632	C	T	TT, CC, TC, CT	TT	HDL
rs4820539	G	A	AA, GG, AG, GA	AA	HDL
rs6583337	G	A	AA, GG, AG, GA	AA	LDL
rs11809524	C	T	TT, CC, TC, CT	TT	SBP
rs11675051	G	A	AA, GG, AG, GA	AA	SBP
rs13005335	A	G	AA, GG, AG, GA	GG	SBP
rs12995369	A	G	AA, GG, AG, GA	GG	SBP
rs10464592	G	A	AA, GG, AG, GA	AA	SBP
rs1670346	A	G	AA, GG, AG, GA	GG	SBP
rs13272568	A	C	AA, CC, AC, CA	CC	SBP
rs600231	G	A	AA, GG, AG, GA	AA	SBP
rs258415	C	A	AA, CC, AC, CA	AA	SBP
rs11614913	C	T	TT, CC, TC, CT	TT	SBP
rs199529	C	A	AA, CC, AC, CA	AA	SBP
rs8090312	G	A	AA, GG, AG, GA	AA	T1D
rs2282930	G	A	AA, GG, AG, GA	AA	TG
rs10851498	T	C	TT, CC, TC, CT	CC	TG

Affymetrix microarray consists of a grid of oligonucleotide probes produced to have a known DNA sequence. The grid Microarray thus holds a specific SNP mutation at a specific locus on the grid. Preparations of labelled mRNA (cDNA/cRNA) taken from the individual patients can then be exposed to the entire grid containing the variety of SNP mutations. Identification of a specific SNP in the patient is determined by the measured level of hybridization with the corresponding target grid position and the labelled cDNA/cRNA. The corresponding SNP IDs were mapped through the affy ID and a gene symbol that are assigned to each cell, as well as the manual mappings of the target SNPs through a capability of the genome browser provided by University of California, Santa Cruz [31].

iv. Evaluation Method

The accuracy of the predictive model was evaluated using 30-fold cross-validation [32]. The performance was measured by the Area Under Curve (AUC) of the Receiver Operator Curve (ROC) . ROC AUC method accounts for the number of true positive predictions against the number of false positive predictions. Statistically, only about 30% of all postmenopausal women develop osteoporosis during their life time [35] thus, the testing dataset is unbalanced. The ROC-AUC can be an effective metric to measure and present the performance of the prediction model.

v. Phenotype Expression Measurement

The predictor execution on this particular dataset was expected to evaluate the geno-pheno-transfer rate. This will also validate the predictor developed in the previous study.

vi. Visualization

This study visualizes the dataset of over 30 essential features by generating various histograms and applying a K-mean clustering algorithm to automatically cluster the resulting plots into groups in a data-driven manner. As demonstrated in the result section, existing correlations among the dataset is clearly displayed through this visualization method. The proposed visualization method can aid the observers in developing an intuitive understanding of the dataset, and effectively mirroring the datasets' characteristics and patterns in them.

III. RESULTS AND DISCUSSION

This section discusses the result obtained from the predictor model and the histogram clustering automation.

A. Predictor results – Dataset from 1000 Genome Project

In the previous study, the predictive models were built and tested on 32 sets of inputs, where each set is an element of the powerset of 6 CVD related conditions. The results obtained from various combinations of SNP inputs showed a strong correlation between the 7 osteoporosis-related SNPs and the HDL2 SNPs. Similarly, another strong correlation was found between the 7 osteoporosis-related SNPs and SBP2 SNPs. The best classification result was achieved using Logistic Regression classifier with the accuracy of 0.7769. Furthermore, there is a strong implication that the likelihood of developing osteoporosis could be determined/predicted by the existing CVD-related factors such as Cholesterol levels, Blood Pressure, and Triglycerides, levels [34].

B. Predictor results – Affy dataset

The performances of the predictive model for 3 different scenarios are listed in the following. The performance measurement was obtained through ROC AUC as mentioned in Section II-B-v. Unlike the previous experiment, this experiment setting did not distinguish high-risk heterozygous pair vs. high-risk homozygous pair. Here are the results for 3 different scenarios:

Scenario 1: Osteoporosis SNPs only – AUC = 0.7285
Scenario 2: CVD SNPs only – AUC = 0.7569
Scenario 3: Both CVD and Osteo SNPs - AUC = 0.8571

The predictor strategy developed here has been confirmed with the available disease related database of Reppe et al. The predictor was shown to be correct in the majority of the patients. These results also confirm the correlation of osteoporosis SNPs with CVD SNPs reported in the above Section. In most cases, the chance of a mutant genotype expressing its aberrant phenotype is much less than 100%. As a well-studied example, the two mutations related to Breast Cancer, BRCA – 1 and BRAC – 2, show such facts. The chances these two mutations expressing their phenotypes by the time a woman is 70 years old are as follows, according to a report by the Susan G. Komen Foundation [33]:

- BRCA 1 - 55%
- BRCA 2 - 45%

The accuracy of the predictor developed with this paper of close to 70% demonstrated an effective prediction. Such a score would thus be reasonably sufficient to caution individuals about having a higher risk of BMD loss. With such knowledge, individuals can take the necessary preventative measures to prevent the development of undesirable conditions and disease, starting at an early stage of their lives.

C. Data Visualization

Various histograms were drawn and clustered effectively to demonstrate the correlations among these high dimensional datasets. The results demonstrated distinct region-specific SNP frequency profiles. They confirmed the genetic links between osteoporosis, and displayed the intricacy of the interrelationship among the SNPs. Grouping the results demonstrated the divides in sample's SNP characteristics. These results are consistent with the idea that an individual's genetic and gene variant profile is related to the region in which his/her ancestors came from. As it was introduced in the previous study [34]. Figure 1 demonstrates the claimed regional divides clearly. CVD (Systolic Blood Pressure)-related SNPs are colored in green, and osteoporosis-related SNPs are in orange. The trends that are obvious in all 28 regions are similar here. Again, the regional divides are well maintained. This type of histogram comparison table was originally introduced in the previous study [34].

The result was also consistent with the fact that European women show higher incidents of osteoporosis and hip fractures when compared with populations from other regions. Europeans have higher frequency of osteoporotic mutations than African [8].

D. Clustering Automation

An automated clustering process was developed for grouping the histograms to better visualize the results. The results of clustering clearly confirms the regional divide of the histogram profiles. Figure 2 displays a sample result of this automated clustering. Sample groups obtained in automated histogram clustering result. The results verify the regional divide of osteoporosis- and CVD-related SNP profiles. The left most column of Figure 2 shows European region, Iberian region in Spain, Utah region (CEPH) with Northern and Western European Ancestry, British region in England and Scotland, Finnish region in Finland, Toscani region in Italia. The middle column shows Bengali region of Bangladesh, Sri Lankan region, Tamil from the UK, Indian region, Telugu region in the UK, Gujarati Indian ancestry in Houston, Texas, Punjabi region from Lahore, Pakistan. The right most column shows Peruvians from Lima, Peru, Puerto Ricans from Puerto Rico, Mexican Ancestry from Los Angeles USA, Colombians from Medellin, Colombia.

To demonstrate the value of this clustering automation, the final number of clusters to be formed by the algorithm was varied. As shown in Figure 3, when the number of clusters is set to 12, the algorithm groups South Asian regional SNP profiles and American regional SNP profiles into two separate groups, perfectly distinguishing the two separate regions. However, when the number of cluster is set to 11,

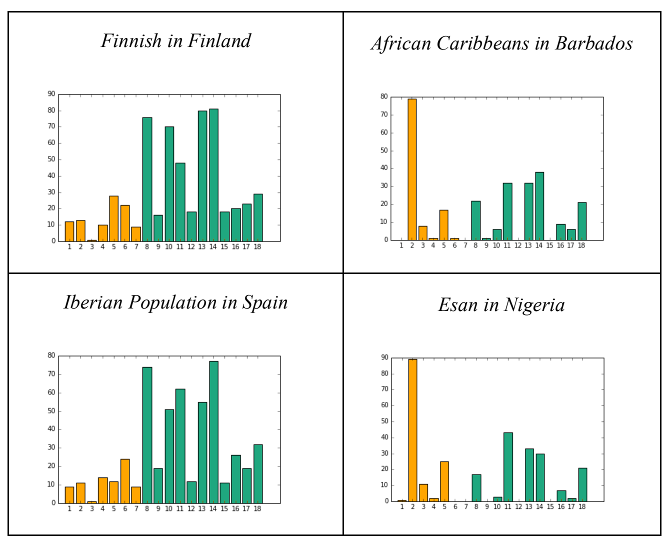


Figure 1: SBP SNP Profile, European Region vs. African Region

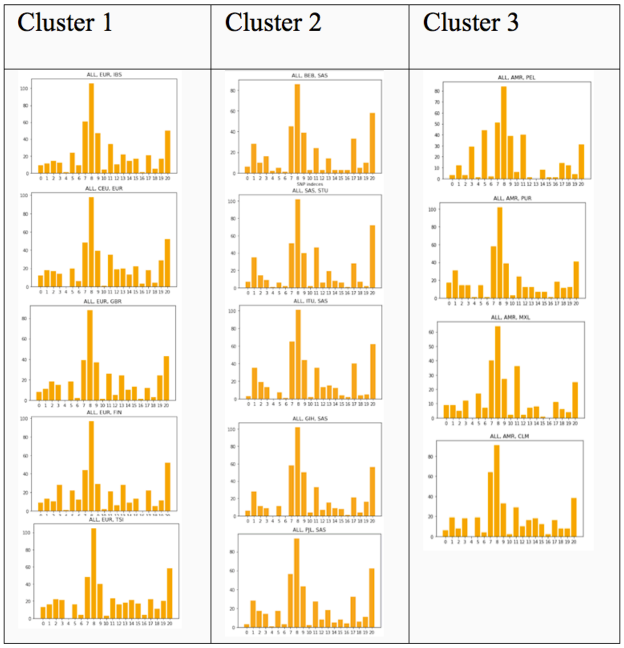


Figure 2: A Sample Histograms Demonstrating Regional Devides

the South Asian regional profiles and American regional profiles are bundled up into a single group. Such a result shows the possibility of further dissection of the regionalism seen here by the automated process. Automation of the histogram profiles also allowed increased number of SNPs to be analyzed all at once.

E. Summary of the Results

i. The high correlation between a person's regional background and the occurrence of the selected 35 SNPs associated with osteoporosis and/or CVD shown in the previous investigation were confirmed by this new methodology. It was confirmed that the highest fracture rates and osteoporosis related SNPs are found in white women of European descent; African Americans tend to have higher Bone Mineral Density (BMD) with few osteoporosis related SNPs [8], as introduced in the previous study [34].

ii. The set of CVD-/osteosis-related SNPs from Reppe et al's dataset can be used to predict a person's likelihood of developing osteoporosis with the accuracy of approximately 64% on their own.

iii. The CVD-osteoporosis related SNPs from Reppe et al's dataset outperformed osteoporosis-related SNPs (see Section I-C-i) in predicting osteoporosis.

IV. Conclusions and Discussions

This paper extended on the previous publication [34] with the following contributions: 1) Validated the predictor model introduced in the previous publication based on the second dataset, using "Area Under Curve" of the "Receiver Operator Curve" evaluation method. The model predicts an individual's likelihood of BMD loss over time. 2) Confirmed that an individual's predisposition to osteoporosis can be linked a predisposition to CVD related conditions. 3) Confirmed a correlation between individual's regional background and the SNP occurrence profiles.

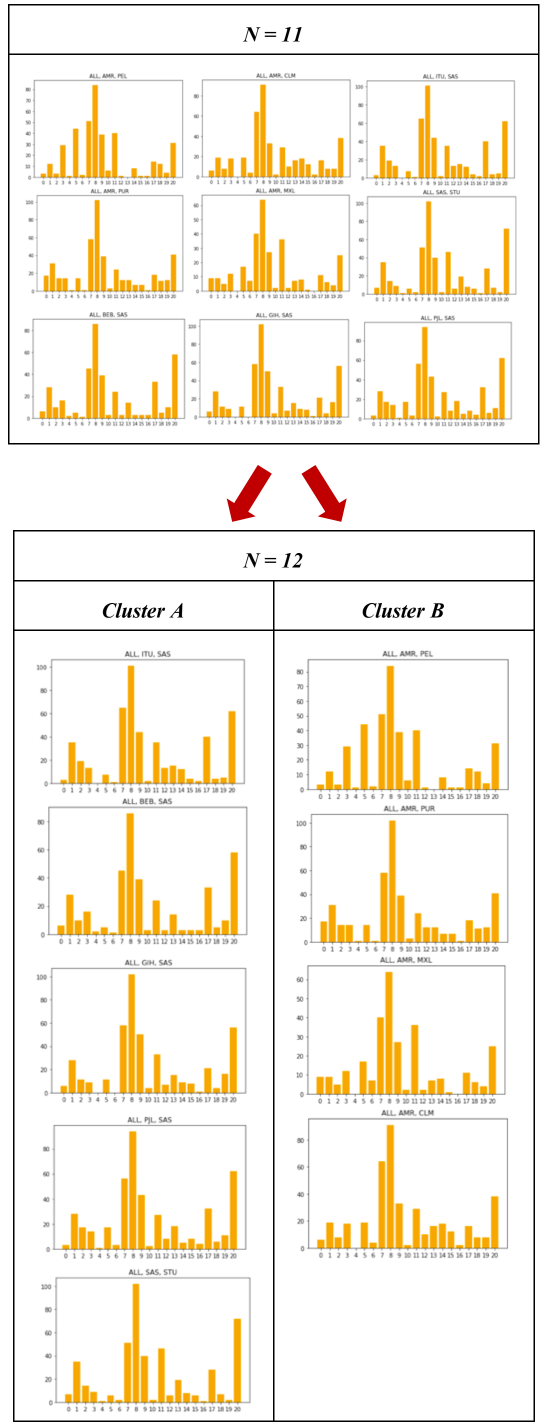


Figure 3: A Sample Auto-Clustering Result

4) Automated clustering of the histograms of SNP occurrence profiles, using the K-mean clustering algorithm. This study also proposed a histogram auto-clustering

approach to effectively visualize the correlations that exists among a high dimensional dataset.

The United Kingdom is now half way through its 100,000 genome project. The goal of obtaining this massive human genome dataset was to determine genetic predisposition to genetic related diseases, such as cancer, CVD and osteoporosis. The automated systems developed here will be of great assistance in achieving these goals.

Revisiting our future in space, as introduced in Section I-A, the exact effects of the outer space specific environments on human biological systems are unknown. Understanding such effects on skeletal system, nervous system, reproductive system, and our genome is crucial when attempting to adapt to an unknown environment. NASA and other space agencies are actively investigating these issues. Being able to use an astronaut's genetic predisposition to predict how that individual's body might respond to zero or low Gravity could help individuals take proper precautions, improve the outcome of space flight, and support the expansion of humanity into space. It will also provide an equal and safer opportunity to everyone who wishes to go to outer space. As we explore into various unknown environments of outer space, capabilities to quickly collect and analyze large amounts of information will be essential. Placing such data into the learning algorithms to update the predictor will allow the explorer appropriately respond to the new environments. With such a system on board, astronauts' well-being and success of the mission can be supported.

References

[1] IGSR and the 1000 Genomes Project. [Online] Available at: http://n.internationalgenome.org/ [retrieved: Nov. 2017].

[2] Which populations are part of your study? [Online] Available at: http://www.internationalgenome.org/faq/which-populations-are-part-your-study [retrieved: Nov. 2017].

[3] D. Karasik, H. Yi‐Hsiang, Y. Zhou, L. A. Cupples, D. P. Kiel, and S. Demissie. "Genome‐wide pleiotropy of osteoporosis‐related phenotypes: the Framingham study." Journal of Bone and Mineral Research 25, no. 7, pp. 1555-1563, 2010.

[4] S. Reppe, Y. Wang, W.K. Thompson, L. K. McEvoy, A. J. Schork, V. Zuber et al. GEFOS Consortium. "Genetic sharing with cardiovascular disease risk factors and diabetes reveals novel bone mineral density loci." PloS one 10, no. 12, e0144531, 2015.

[5] O. Hitomi, S. Sasaki, H. Horiguchi, E. Oguma, K. Miyamoto, Y. Hosoi, M. K. Kim, and F. Kayama. "Dietary patterns associated with bone mineral density in premenopausal Japanese farmwomen–." The American journal of clinical nutrition 83, no. 5, pp. 1185-1192, 2006.

[6] J. B. Richards, H. F. Zheng, and T. D. Spector, "Genetics of osteoporosis from genome-wide association studies: advances and challenges," Nature Reviews Genetics, 2012.

[7] GWAS Catalog. [Online] Available at: https://www.ebi.ac.uk/gwas/ [retrieved: Nov. 2017].

[8] N. Hae-Sung, S. S. Kweon, J. S. Choi, J. M. Zmuda, P. C. Leung, L. Y. Lui, et al. "Racial/ethnic differences in bone mineral density among older women." Journal of bone and mineral metabolism 31, 2013.

[9] H. Yi-Hsiang, et al. "An integration of genome-wide association study and gene expression profiling to prioritize the discovery of novel susceptibility Loci for osteoporosis-related traits." PLoS genetics 6, no. 6, e1000977, 2010.

[10] L. Langsetmo, et al. "Using the same bone density reference database for men and women provides a simpler estimation of fracture risk." Journal of Bone and Mineral Research, no.10, pp. 2108-2114, 2010.

[12] J. Greenbaum, K. Wu, L. Zhang, H. Shen, J. Zhang, and H. W. Deng. "Increased detection of genetic loci associated with risk predictors of osteoporotic fracture using a pleiotropic cFDR method." Bone 99, pp. 62-68, 2017.

[13] D. Karasik, and M. Cohen-Zinder. "Osteoporosis genetics: year 2011 in review." BoneKEy reports 1, no. 8, 2012.

[14] U.S. National Library of Medicine | What are Single Nucleotide Polymorphisms (SNPs)? [Online] Available at: https://ghr.nlm.nih.gov/primer/genomicresearch/snp [retrieved: 12 Nov. 2017].

[15] Y. H. Hsu, et al. "An integration of genome-wide association study and gene expression profiling to prioritize the discovery of novel susceptibility Loci for osteoporosis-related traits." PLoS genetics 6, no. 6, e1000977, 2010.

[17] I. M. Meyer, and R. Durbin. "Gene structure conservation aids similarity based gene prediction." Nucleic acids research 32, no. 2, pp. 776-783, 2004.

[18] R. J. Carter I. Dubchak, and S. R. Holbrook. "A computational approach to identify genes for functional RNAs in genomic sequences." Nucleic acids research 29, no. 19, pp. 3928-3938, 2001.

[19] T. Chen, M. Y. Kao, M. Tepel, J. Rush, and G. M. Church. "A dynamic programming approach to de novo peptide sequencing via tandem mass spectrometry." Journal of Computational Biology 8, no. 3, pp. 325-337, 2001.

[20] US Department of Health and Human Services. "Bone health and osteoporosis: a report of the Surgeon General." Rockville, MD: US Department of Health and Human Services, Office of the Surgeon General 87, 2004.

[21] P. Salari, and M. Abdollahi. "The influence of pregnancy and lactation on maternal bone health: a systematic review." Journal of family & reproductive health 8, no. 4, p. 135, 2014.

[22] L. Qin, Y. Liu, et al "Computational characterization of osteoporosis associated SNPs and genes identified by genome-wide association studies." PloS one 11, no. 3, e0150070, 2016.

[23] E. R. Spector, S. M. Smith, and J. D. Sibonga. "Skeletal effects of long-duration head-down bed rest." Aviation, space, and environmental medicine 80, no. 5, A23-A28, 2009.

[24] The Matplotlib API [Online] Available at: https://matplotlib.org/2.0.2/api/index.html [retrieved: Nov. 2017].

[25] E. Roberge. "The Gravity of It All: From osteoporosis to immunosuppression, exploring disease in a microgravity environment holds promise for better treatments on Earth." IEEE pulse 5, 2014.

[26] Ensemble Genome Browser [Online] Available at: https://uswest.ensembl.org/index.html [retrieved: Jan. 2018].

[27] Bioconductor 3.6 Annotation package, pd.hg.u133.plus.2 [Online] Available at: http://bioconductor.org/packages/release/data/annotation/html/pd.hg.u133.plus.2.html [retrieved: Jan. 2018].

[28] E-MEXP-1618 - Transcription profiling of bone biopsies from postmenopausal females identifies 8 genes highly associated with bone mineral density [Online] Available at: https://www.ebi.ac.uk/arrayexpress/experiments/E-MEXP-1618/files/ [retrieved: Jan. 2018].

[29] M. Bilban, L. K. Buehler, S. Head, G. Desoye, and V. Quaranta. "Defining signal thresholds in DNA microarrays: exemplary application for invasive cancer." BMC genomics 3, no. 1, 19, 2002.

[30] 2015 ISCD Official Positions – Adult [Online] Available at: https://www.iscd.org/official-positions/2015-iscd-official-positions-adult/ [retrieved: Jan. 2018].

[31] UCSC Genome Browser on Human Dec. 2013 (GRCh38/hg38) Assembly [Online] Available at: https://genome.ucsc.edu/ [retrieved: Jan. 2018].

[32] The documentation for *sklearn.model_selection.cross_val_score* [Online] Available at: http://scikit-learn.org/stable/modules/generated/sklearn.model_selection.cross_val_score.html [retrieved: Jan. 2018].

[33] Susan G. Komen Available at: ww5.komen.org [retrieved: Jan. 2018].

 ISBN: 978-1-61208-639-2

[34] K. Sawada, M. W. Clark, N. Alshurafa, and M. Pourhomayoun, "Analyzing the Mutation Frequencies and Correlation of Genetic Diseases in Worldwide Populations Using Big Data Processing, Clustering, and Predictive Analytics," International Conference on Computational Science and Computational Intelligence, pp. 1459 – 1464, 2017.

[35] Epidemiology by International Osteoporosis Foundation [Online] Available at: https://www.iofbonehealth.org/epidemiology.

Automated Quantification of the Resistance of Epithelial Cell Layers from an Impedance Spectrum

Thomas Schmid
Department of Computer Engineering
Universität Leipzig
Leipzig, Germany
schmid@informatik.uni-leipzig.de

Dorothee Günzel
Institute of Clinical Physiology
Charité - Universitätsmedizin Berlin
Berlin, Germany
dorothee.guenzel@charite.de

Martin Bogdan
Department of Computer Engineering
Universität Leipzig
Leipzig, Germany
bogdan@informatik.uni-leipzig.de

***Abstract*—Quantifying the permeability of intestinal epithelia is a central task in assessing tissue dysfunctions. This can be achieved conveniently by determining the electric resistance of the epithelial tissue by impedance spectroscopy. While in clinical practice this parameter is often estimated by manual extrapolation of discrete two-dimensional plots of impedance spectra, this approach is known to be particularly unprecise for spectra that deviate from a semicircular shape. Previous computational approaches achieved less then ten percent deviation from the known target value on average, but outliers exhibited significantly larger maximum deviations. Here, we show that systematic feature extraction and selection allow estimation of the epithelial resistance with less than one percent deviation from the known target value on average and less than ten percent at maximum. As a result of detailed modeling of cell culture lines and functional states, epithelial resistance for the cell lines HT-29/B6, IPEC-J2 and MDCK I can be quantified reliably under control conditions, as well as under influence of EGTA and nystatin.**

Keywords–Physiology; Epithelia; Impedance Spectroscopy; Feature Selection; Artificial Neural Networks; Random Forests.

I. Introduction

Intestinal epithelial cells form the primary barrier of the gut between the body's interior and the exterior environment (i.e., the gut contents). In general, all epithelial and endothelial cells are connected by arrays of transmembrane proteins, called tight junctions that seal the space between two neighbouring cells. Tight junction properties are determined by their major constituents, the members of the claudin protein family. While certain claudins strengthen the barrier function (e.g., claudin-1 in skin, claudin-5 in the blood-brain barrier [1]), others convey specific charge and/or size selectivity [1][2]. Regulation and dysregulation of these channel-forming claudins appear to be paramount in the pathophysiology of numerous diseases [3].

As a whole, the epithelium regulates transport of molecules and establishes a tight barrier against toxins and pathogens. The permeability of epithelial tissue to ions is reflected by its electric resistance, called transepithelial resistence (R^T, often also abbreviated as TER). R^T is the sum of the subepithelial resistance R^{sub} and the epithelial resistance R^{epi}. These partial resistances are individual and independent parameters of the tissue, and exact knowledge of their values is crucial in the analysis of epithelial dysfunctions. For patients with inflammatory intestinal diseases, e.g, it was observed that due to the inflammation R^{sub} increased whereas R^{epi} decreased.

The most common way to assess tissue permeability is a direct measurement of fluxes, e.g., by using radioactive isotopes or labelled substances. Alternatively, permeability of the two major extracellular ion species, Na^+ and Cl^-, is determined by measuring tissue conductance or its reciprocal, resistance, e.g., by using "chopstick electrodes" or Ussing chambers [4]. A more convenient way to assess tissue permeability is impedance spectroscopy. Typically, up to 50 complex-valued impedances Z are obtained by measuring current-voltage relationships under alternate current (AC) with varying frequencies [5]. A common representation of these measurements are Nyquist diagrams, where the real part $\Re(Z)$ of an impedance is plotted against the imaginary part $\Im(Z)$ (Figure 1). To analyze samples, an equivalent electric circuit of appropriate complexity is modeled [6]. The simplest circuit that incorporates R^{epi} is a resistor-capacitor (RC) circuit (Figure 2a). To represent physiological polarity of epithelial cells, a circuit with two RC subcircuits in series and a resistor in parallel may be used (Figure 2b). In both cases, the subepithelium can be represented by a further resistor in series.

In contrast to DC resistance measurements, impedance spectroscopy allows to distinguish between R^{sub} and R^{epi}. Under in vivo conditions the subepithelium does not contribute

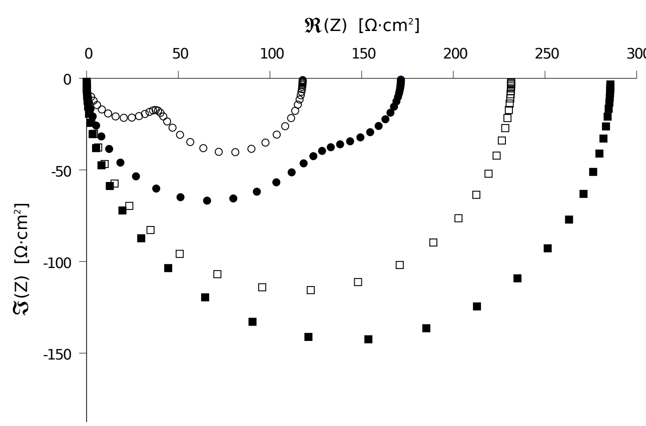


Figure 1. Overlay of two semi- and two non-semicircular impedance spectra with 42 frequencies where real ($\Re$) and imaginary ($\Im$) part of each complex-valued impedance Z are plotted against each other.

to the barrier function as subepithelial capillaries are in close contact with the basolateral membrane of the epithelial cells. Further, impedances reflect not only conductive but also capacitive properties and allow to derive the epithelial capacitance which directly depends on the epithelial surface area [7].

In previous work, we have demonstrated that traditional ways to estimate R^{epi} from an impedance spectrum, e.g. visual extrapolation, can lead to serious errors in analysis of epithelial characteristics. We have also demonstrated that estimations can be improved by applying machine learning techniques on complex-valued impedances of error-prone [8] or on extracted features of ideal impedance spectra, respectively [9]. We have introduced detailed and realistic models of the electric behavior of three epithelial cell lines [10]: the human colon carcinoma cell line HT-29/B6, the porcine jejunum cell line IPEC-J2 and the canine kidney cell line MDCK I under control conditions, as well as under influence of EGTA, nystatin or both. Rationale behind these models is that for a given electric circuit, the theoretical impedance at a given frequency can be calculated if the values of all circuit components are given.

Here, we combine systematic modeling of impedance spectra with systematic feature extraction in order to improve quantification of R^{epi}. We show that employing a mixture of both measured and extracted features reduces relative deviation from the target value when applying supervised learning techniques like random forests. Compared with previous approaches, the estimation error is significantly reduced.

II. Methods

A. Modeling Impedance Spectra

For all three modeled cell lines, an equivalent circuit consisting of two RC subcircuits a (R_a, C_a) and b (R_b, C_b) located in series and a resistor in parallel (R_p) is assumed (Figure 2). In accordance with Kirchhoff's laws, the corresponding impedance Z at an angular frequency ω can be derived from the impedances of the circuit components:

$$Z(\omega) = \frac{R_p(R_a + R_b) + i\omega[R_p(R_a\tau_b + R_b\tau_a)]}{R_a + R_b + R_p(1 - \omega^2\tau_a\tau_b) + i\omega[R_p(\tau_a + \tau_b) + R_a\tau_b + R_b\tau_a]} \quad (1)$$

where $i = \sqrt{-1}$, and $\tau_a = R_aC_a$ and $\tau_b = R_bC_b$.

In the measurements to be modeled, ten frequencies per decade are used. Based on a lowest frequency f_1 of 1.3 Hz,

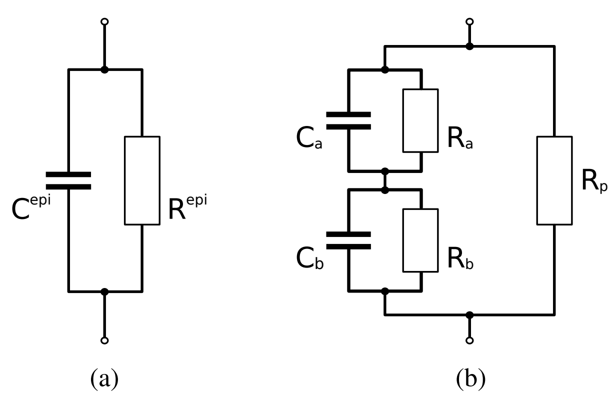


Figure 2. Equivalent circuits. (a) A simple resistor-capacitor (RC) circuit yields semicircular impedance spectra. (b) A 2-RC circuit yields semicircular or non-semicircular impedance spectra.

higher frequencies f_i with $1 < i \leq n = 42$ are multiples by a factor of $10^{0.1}$ (~1.26). Note that the value of f_1 is a chosen to a avoid obtaining multiples of 50 Hz (mains frequency) and that for application with Eq. 1, $\omega_i = 2\pi/f_i$ is calculated.

Further, we model and apply synthetic data scatter reflecting deviations from the theoretical impedance value caused by the electrophysiological measurement set-up. This scatter is here modeled based on relative deviations of real part $\Re$ and imaginary part $\Im$ of measured impedances from theoretical values. For a given impedance Z at frequency $f = 2\pi\omega$, relative deviation σ of $\Re(Z)$ is approximated as second-order Fourier series (n=2) and relative deviations of $\Im(Z)$ is approximated as fourth-order polynomial function (n=4):

$$\sigma_{\Re}(f) = a_0 + \sum_{i=1}^{n} a_i \cdot cos(nwf) + b_i \cdot sin(nwf) \quad (2)$$

$$\sigma_{\Im}(f) = a_0 + \sum_{i=1}^{n} a_i \cdot f^i \quad (3)$$

where coefficients w, a_0, a_1, b_1, a_2, b_2 or a_0, a_1, a_2, a_3, a_4 respectively were determined by function fitting and a_0 is modeled as dependend on R^T. For further details and comparison with measured data, see [8] or [10].

For all synthetic samples used in the following, first complex impedances are calculated according to model parameters. Then data scatter is added, and finally, polar impedances are calculated from the scattered complex impedances.

B. Sampling Cell Lines and Functional States

While IPEC-J2 and MDCK I cells typically show relatively high R^T values under physiological conditions, for HT-29/B6 cells relatively low values are measured. Based on these findings, as well as on further published measurement results, parameter ranges of the components of the equivalent circuit were defined and confirmed in previous work [10]. By extending this modeling approach for individual or combined application of EGTA and nystatin, impedance spectra for a total of 11 measurement scenarios are synthesized; note that for MDCK I cells, only measurements for Control, nystatin and EGTA were available for modeling [10]. Randomized samples of each scenario are combined to a data set of 275,000 samples (Tab. I), which is then split into a training set of 200,000 and a test set of 75,000 samples.

C. Reference Methods to Quantify Epithelial Resistance

Reliability of the hereby introduced approach was compared to a conventional approach to estimate R^{epi}, as well as to two approaches that we have proposed in previous work.

TABLE I. SAMPLE SIZES FOR MODELED EPITHELIAL CELL LINES AND THEIR FUNCTIONAL CONDITIONS.

Condition	HT-29/B6	IPEC-J2	MDCK I
Control	25.000	25.000	25.000
Nystatin	25.000	25.000	25.000
EGTA	25.000	25.000	25.000
EGTA+nystatin	25.000	25.000	—

1) Frequency-blind Circle Fit (Method M1): Each impedance spectrum was regarded as Nyquist diagram and a Cole-Cole fit [11] was applied, i.e., a circle was fitted as described by Kasa [12]. For this task, the *R* package *conicfit* and its function *CircleFitByKasa* was used. The intercept with the x-axis at the low frequency end was taken as R^{epi}.

2) Neural Network Prediction (Method M2): Complex-valued impedances Z_1,...,Z_{10} of the first decade of frequencies (f_1,...,f_{10}) were used to train an artificial neural network. This implies a 20-dimensional input feature space I:

$$I_{M2} = \{\Re(Z_1), ..., \Re(Z_{10}), \Im(Z_1), ..., \Im(Z_{10})\} \quad (4)$$

While this approach follows a method previously published by us [8], we improved the precision of predictions by using the Rprop algorithm [13] instead of conventional backprogation. Employed Rprop parameters were η^+=1.2, η^-=1.2, Δ_{max}=50.0, Δ_{min}=0.000001 and Δ_0=0.07. We used the *Fast Artificial Neural Network* library [14] to perform automated evaluation of 30 feed-forward networks with one hidden layer and $n \in \{1, ..., 30\}$ hidden neurons (20-n-1 architecture). Training these networks in parallel with the same data, the architecture 20-9-1 was identified as best performer and used for predictions.

3) Magnitude at 1.3 Hz (Method M3): In previous evaluations of features for prediction of R^{epi} from an impedance spectrum, we found that the magnitude r observed at 1.3 Hz is a reliable approximation for semicircular spectra [9]. While reliability decreases with non-semicircular spectra, this one-feature approach is ultimately easy to implement and a feasible method for comparison. For a given spectrum, we simply used the value of $r(Z_1) \in S_r$ as prediction for R^{epi}.

D. Feature Extraction-based Approach

To improve prediction of R^{epi} from impedance spectra, we suggest a supervised learning approach that uses systematic feature extraction and selection. After extracting implicit features for each spectrum, we ranked explicit and implicit features. Finally, subsets of these features were assessed by applying supervised learning.

From calculations, n=42 tupels of real and imaginary parts $((\Re(\omega_0), \Im(\omega_0)), \ldots, (\Re(\omega_{n-1}), \Im(\omega_{n-1})))$, are obtained. Real and imaginary parts of a spectrum can be regarded as separate feature sets $S_\Re$ and $S_\Im$:

$$S_\Re = \{\Re(\omega_0), \ldots, \Re(\omega_{n-1})\} \quad (5)$$

$$S_\Im = \{\Im(\omega_0), \ldots, \Im(\omega_{n-1})\} \quad (6)$$

As an alternative representation, these complex values were transformed into polar coordinates, i.e., into phase angle ϕ and magnitude r. This resulted in two alternative sets S_ϕ and S_r:

$$S_\phi = \{\phi(\omega_0), ..., \phi(\omega_{n-1})\} \quad (7)$$

$$S_r = \{r(\omega_0), ..., r(\omega_{n-1})\} \quad (8)$$

1) Feature Extraction: Based on the features that can be explicitly measured or synthesized, respectively, a set I_e of explicit input features was defined:

$$I_e = S_\Re \times S_\Im \times S_\phi \times S_r \quad (9)$$

where $S_\Re$, $S_\Im$, S_ϕ and S_r contained n=42 features each.

Additionally, the development of magnitude and phase over time was represented in sets of differential features:

$$S_{\Delta\phi} = \{\Delta\phi | \Delta\phi_i = \phi(\omega_{i+1}) - \phi(\omega_i), 0 \le i < n-1\} \quad (10)$$
$$S_{\Delta r} = \{\Delta r | \Delta r_i = r(\omega_{i+1}) - r(\omega_i), 0 \le i < n-1\} \quad (11)$$

which are combined to a set I_d of differential input features:

$$I_d = S_{\Delta\phi} \times S_{\Delta r} \quad (12)$$

where $S_{\Delta\phi}$ and $S_{\Delta r}$ contained n=41 features each.

Further, a total of 16 statistical properties like minimum, maximum and range were determined for each feature set (for details see appendix).

$$G_\Re = \{x \mid x \mathrel{\hat=} \text{statistical property of } S_\Re\} \quad (13)$$
$$G_\Im = \{x \mid x \mathrel{\hat=} \text{statistical property of } S_\Im\} \quad (14)$$
$$G_\phi = \{x \mid x \mathrel{\hat=} \text{statistical property of } S_\phi\} \quad (15)$$
$$G_r = \{x \mid x \mathrel{\hat=} \text{statistical property of } S_r\} \quad (16)$$
$$G_{\Delta\phi} = \{x \mid x \mathrel{\hat=} \text{statistical property of } S_{\Delta\phi}\} \quad (17)$$
$$G_{\Delta r} = \{x \mid x \mathrel{\hat=} \text{statistical property of } S_{\Delta r}\} \quad (18)$$

which were combined to a set I_i of implicit input features:

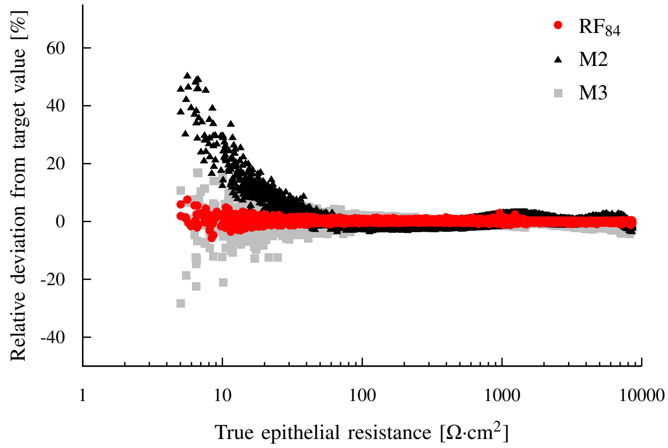
$$I_i = G_\Re \times G_\Im \times G_\phi \times G_r \times G_{\Delta\phi} \times G_{\Delta r} \quad (19)$$

By combining explicit, differential and implicit features, an 346-dimensional input feature space I was obtained that represents 42 measured impedances:

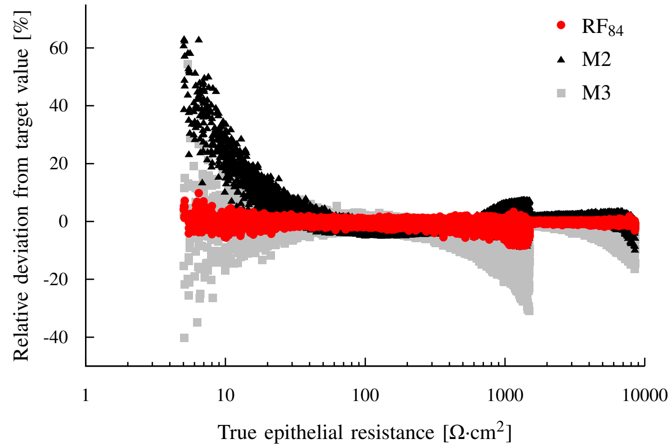
$$I = I_e \times I_d \times I_i \quad (20)$$

2) Feature Selection: Three alternative feature selection approaches were employed and assessed:

- Filter-based approach. We applied Correlation-based Feature Selection (CFS) on the input space I of the 200,000 training samples. For this task, the *R* package *FSelector* and its function cfs was used. In contrast to other multivariate filters, CFS evaluates feature subsets instead of single features [15]. In particular, CFS returns not a ranking of features but a fixed subset of features. Here, CFS evaluated a one-dimensional subset best that consists only of the range of S_r.
- Expert knowledge-based approach (EXP). As a matter of theory, low-frequency impedances are closely related to R^{epi} [6]. Also, the magnitude of an impedance is more meaningful than other representations. Therefore, we manually selected the feature sets G_r and $G_{\Delta r}$, as well as magnitudes of the low-frequency impedances Z_1,...,Z_5 and related differences ($r(Z_{i+1})$-$r(Z_i)$). This yielded a 42-dimensional representation.
- Wrapper-based approach (RF_{84}). Variable importance was assessed for the input space I of the 200,000 training samples. Using the *R* package *randomForest*, a random forest (RF) consisting of 500 trees was trained to predict R^{epi} and used in a wrapper-like way. After training, variable importance values were used to rank the input features, and a subset of 84 top-ranked features was chosen as representation of an impedance spetrum. This number of selected features is simply motivated by the initial number of 84 features per spectrum. Automated wrapper techniques like sequential feature selection or recursive feature elimination were omitted here due to the high computational costs.



(a) Semicircular test data (n=33,069)



(b) Non-semicircular test data (n=41,906)

Figure 3. Relative deviations of estimated values from true R^{epi}. Deviations achieved by the feature extraction-based approach RF_{84}, as well as for the two best performing reference methods (M2, M3) are plotted against the respective true target value. Values shown refer either to (a) semicircular or (b) nonsemicircular spectra, which are discriminated by the relation between apical and basolateral time constants.

3) Evaluation: All three feature selection approaches CFS, EXP and RF_{84} were used to estimate R^{epi} for the 75,000 samples of the test data. Deviations of the obtained estimations from the target were compared to reference methods M1, M2 and M3.

III. Results

The reliability of estimations for R^{epi} was assessed. For each approach, the relative deviation of an predicted value t_p from the target value t, or error Δt respectively, was used:

$$\Delta t = \frac{t_p - t}{t} \quad (21)$$

Based on Δt, both absolute-valued deviations, as well as signed deviations were analyzed.

A. Absolute-valued Deviations

Assessing absolute-valued relative deviation from the target $|\Delta t|$, for reference methods, as well as feature extraction-based approaches a mean deviation of less than ±2.0 % was observed. Maximum deviations, however, laid between ±54.4 % and ±92.1 % for the reference methods and between ±9.8 % and ±85.9 % for feature extraction-based approaches. Maximum deviations for the 84- and 20-dimensional cartesian representations M1 and M2 were comparable to maximum deviations of the two one-dimensional polar representations M3 and CFS. The mixed 42- and mixed 84-dimensional representations learned by random forests both showed significantly lower maximum deviations. Lowest mean and maximum deviations of all methods assessed were obtained by the random forest trained with the 84 top-ranked features. Figure 4 shows boxplots of $|\Delta t|$ for all methods evaluated.

B. Estimations for Semi- and Non-semicircular Spectra

Deviations of the random forest trained with the 84 top-ranked features (RF_{84}) were further assessed by separating semicircular and non-semicircular spectra of the 75,000 test samples. In continuity with our previous work [7], spectra from equivalent circuits with $\tau_a \leq 5 \cdot \tau_b$ or $\tau_b \leq 5 \cdot \tau_a$ were considered as semicircular spectra; all spectra with greater difference between τ_a and τ_b were considered as non-semicircular.

As Figure 3 illustrates, estimations by reference methods M2 and M3 tended to become unprecise with low values of R^{epi}; this held true for both semi- and non-semicircular spectra. Using the feature extraction-based approach RF_{84}, deviations for high and low target values laid below 10 % and thereby in a significantly smaller range than for M2 and M3.

For semicircular spectra with larger target values (Figure 3a), deviations of estimations by the feature extraction-based approach RF_{84} laid in a range comparable to M2 and M3. For non-semicircular spectra with larger target values (Figure 3b), RF_{84} showed in most cases small improvements over M2; using method M3 appeared to induce systematic error that causes greater deviations than for M2 and RF_{84}.

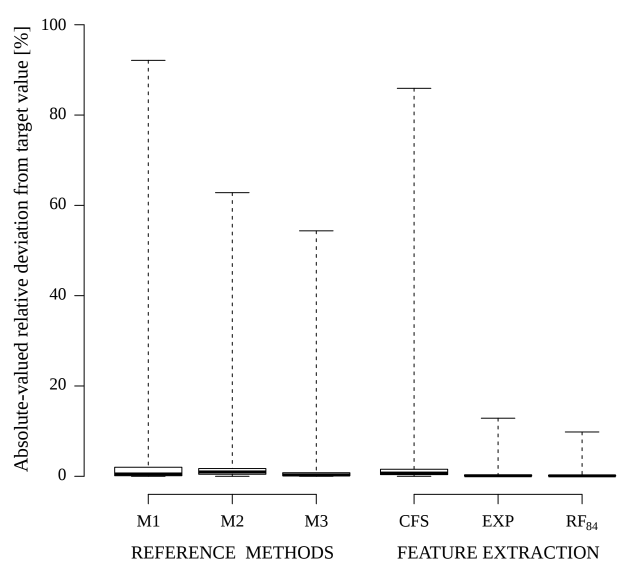


Figure 4. Boxplots of absolute-valued relative deviations from true values of R^{epi} [±%]. Each box reflects median, 1. and 3. quartile.

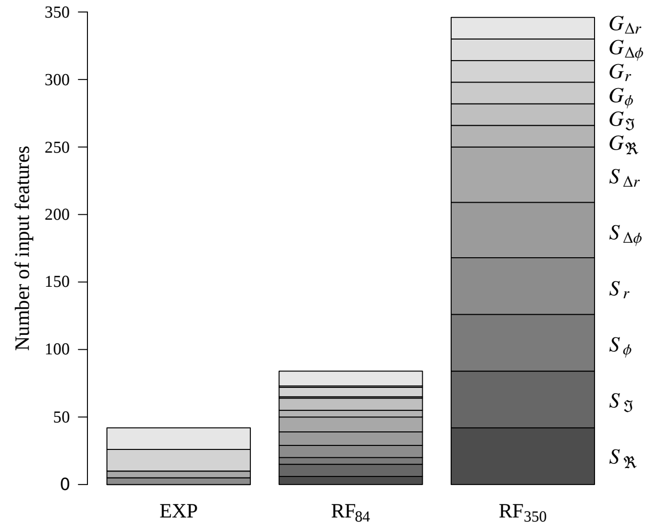


Figure 5. Usage of feature subsets. The approaches RF_{84} and RF_{350} use features from all subsets, EXP only from the subsets S_r, $S_{\Delta r}$, G_r and $G_{\Delta r}$.

IV. Discussion

A. Comparison of Problem Representations

Evaluation of the test samples showed that the filter-based (CFS), as well as the expert-based (EXP) and wrapper-based subset (RF_{84}) yielded estimations with little average deviation from the target. While using the one-dimensional CFS representation exhibits similiar maximum error as reference methods, EXP and RF_{84} allowed for significantly lower maximum error than the reference methods. Yet, the features used in both approaches exhibited different characteristics (Figure 5). EXP employed only features of direct physical relevance and only half as many dimensions as RF_{84}. RF_{84} on the other hand employed more features, but reduced the maximum error to a level comparable to a much more high-dimensional representation of 350 features (Tab. II).

B. Use of Random Forests

Exploiting the variable importance determined by random forests allowed to efficiently combine supervised learning with a wrapper-based feature selection. As in previous studies [16], the ranking of all 350 features by a random forests could be used to identify features appropriate for the given task. In particular, top-ranked features yielded estimations with lesser deviations than a feature set determined by filter-based CFS.

TABLE II. RELATIVE DEVIATIONS FROM R^{EPI} [±%].

	EXP	RF_{84}	RF_{350}
Maximum	12.9	9.81	9.24
3. Quartile	0.30	0.23	0.25
Mean	0.29	0.22	0.23
Median	0.12	0.10	0.11
1. Quartile	0.05	0.04	0.05
Minimum	0.00	0.00	0.00

Compared with the previously developed ANN approach (M2), the random forest approaches EXP and RF_{84} performed with similar precision for samples with large target values and with much better precision for small target values (R^{epi} < 25 Ωcm^2). This is in accordance with findings in previous comparisons of random forests with ANNs [16]. As a result, both EXP and RF_{84} allowed for the first time reliable estimations for impedance spectra reflecting epithelia with low R^{epi}.

C. Limitations

The feature extraction-based approach suggested here to estimate R^{epi} work equally well for cell types and conditions where time constants of the apical and basolateral cell membranes are similar (as usually found in cell culture) as for conditions where these time constants differ considerably (as found in many tissues). Moreover, Figure 3 illustrates that reliable quantification of R^{epi} from an impedance spectrum is feasible for resistance values spanning several orders of magnitude. As the presented method involves modeled spectra, however, applicability is naturally depending on modeling and sampling of the data. The previously modeled data that is used here is in good accordance with measured data [10]. While a variety of cell lines and functional states is considered, results and estimations obtained here are still limited to the scenarios modeled. Also, characteristics of the training and test data influence characteristics of estimation methods. For example, precision of supervised learning methods tends to decrease if the number of samples decreases. To this end, e.g., the density of target values needs to be considered (cf. Figure 6).

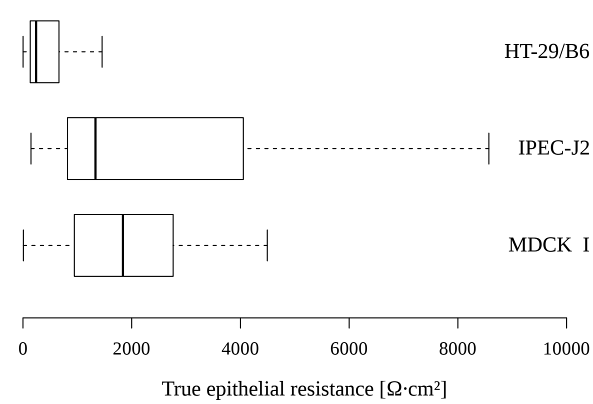


Figure 6. R^{epi} target values [$\Omega \cdot cm^2$] of the test data (n=75,000) grouped by cell lines. Each box reflects median, 1. and 3 quartile.

D. Relevance for Biomedical Applications

Impedance spectroscopic techniques are increasingly gaining importance in biomedical applications like monitoring growth of cultured epithelial and endothelial cells (e.g., retinal pigment epithelium, gastrointestinal tract cells, pulmonary cells, blood-brain-barrier models [17][18]), or alterations of barrier function during pharmacological studies [19][20][17]. Furthermore, impedance spectroscopy is the only technique that allows functional distinction between epithelial and subepithelial properties of ex vivo tissue, such as intestinal biopsies of patients with suspected barrier impairment.

 ISBN: 978-1-61208-639-2

If the technique is to be used on a routine basis, however, reliable automatization for the evaluation of impedance spectra is indispensible. On one hand, manual evaluation of impedance spectra to extract the physiologically relevant parameter, R^{epi}, requires extensive user training and is time consuming, as individual spectra need to be fitted by complex equations [9]. On the other hand, currently available systems usually only record R^T (i.e., the sum of the subepithelial and epithelial resistance) or even only relative alterations in R^T over the time-course of an experiment, as estimation of absolute R^T values is too error-prone [21][17].

V. Conclusions

While impedance spectroscopy is a convenient measurement technique, determining the resistance of an epithelial tissue with traditional approaches is error-prone under certain circumstances. Here, we have shown that this clinically important parameter can be quantified with high precision by training random forests with features extracted from modeled impedance spectra. In particular, we have shown that this approach outperforms both traditional estimation techniques and a state-of-the-art artificial neural network approach. Due to detailed and realistic modeling, we suggest this approach is valid for the epithelial cell lines HT-29/B6, IPEC-J2 and MDCK I under control conditions, as well as under the influence of nystatin and EGTA.

Appendix

For the explicit feature sets $S_{\Re}$, $S_{\Im}$, S_{ϕ} and S_r, as well as for the differential feature sets $S_{\Delta\phi}$ and $S_{\Delta r}$ a total of 16 statistical features (Tab. III) was calculated. By combining these statistical features of set S, a novel set G of related global features was extracted.

TABLE III. STATISTICAL FEATURES FOR FEATURE SET $S_{\Re}$.

Feature	Definition with $x_i \in S_{\Re}$ and $n = \#(S_{\Re})$	Description
$min(S_{\Re})$	$\{x : x \leq x_i \;\; \forall \; x_i \in S_{\Re}\}$	Minimum
$max(S_{\Re})$	$\{x : x_i \leq x \;\; \forall \; x_i \in S_{\Re}\}$	Maximum
$R(S_{\Re})$	$max(S_{\Re}) - min(S_{\Re})$	Range
$P_{0,1}(S^*_{\Re})$	$x_{\lfloor 0,1 \cdot n+1 \rfloor}$	1. Percentile
$Q_{0,25}(S^*_{\Re})$	$x_{\lfloor 0,25 \cdot n+1 \rfloor}$	1. Quartile
$Q_{0,75}(S^*_{\Re})$	$x_{\lfloor 0,75 \cdot n+1 \rfloor}$	3. Quartile
$P_{0,9}(S^*_{\Re})$	$x_{\lfloor 0,9 \cdot n+1 \rfloor}$	9. Percentile
$R_{IQ}(S_{\Re})$	$Q_{0,75}(S^*_{\Re}) - Q_{0,25}(S^*_{\Re})$	Interquartile distance
$R_{IP}(S_{\Re})$	$Q_{0,9}(S^*_{\Re}) - Q_{0,1}(S^*_{\Re})$	Interpercentile distance
$\bar{x}_{med}(S^*_{\Re})$	$\begin{cases} x_{(\frac{n+1}{2})}, & \text{if } n \text{ odd} \\ \frac{1}{2}(x_{(\frac{n}{2})} + x_{(\frac{n}{2}+1)}), & \text{if } n \text{ even} \end{cases}$	Median
$\bar{x}_{arithm}(S_{\Re})$	$\frac{1}{n} \sum_{i=1}^{n} x_i$	Arithmetic mean
$\bar{x}_{geom}(S_{\Re})$	$\sqrt[n]{\prod_{i=1}^{n} x_i}$	Geometric mean
$\bar{x}_{harm}(S_{\Re})$	$\frac{n}{\sum_{i=1}^{n} \frac{1}{x_i}}$	Harmonic mean
$s^2(S_{\Re})$	$\frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x}_{arithm}(S_{\Re}))^2$	Variance
$s(S_{\Re})$	$\sqrt{s^2(S^*_{\Re})}$	Standard deviation
$R_{MM}(S_{\Re})$	$\bar{x}_{med}(S_{\Re}) - \bar{x}_{arithm}(S_{\Re})$	Distance between median and arithmetic mean

References

[1] D. Günzel and A. S. L. Yu, “Claudins and the modulation of tight junction permeability,” Physiological Reviews, vol. 93, no. 2, 2013, pp. 525–569.

[2] S. M. Krug et al., “Charge-selective claudin channels,” Annals of the New York Academy of Sciences, vol. 1257, 2012, pp. 20–28.

[3] S. M. Krug, J.-D. Schulzke, and M. Fromm, “Tight junction, selective permeability, and related diseases,” Seminars in Cell & Developmental Biology, vol. 36, 2014, pp. 166–176.

[4] B. Jovov, N. Wills, and S. Lewis, “A spectroscopic method for assessing confluence of epithelial cell cultures,” American Journal of Physiology, vol. 261, no. 6 Pt 1, 1991, pp. C1196–203.

[5] J. R. Macdonald and W. B. Johnson, Fundamentals of Impedance Spectroscopy. John Wiley & Sons, Inc., 2005, pp. 1–26.

[6] D. Günzel et al., “From TER to trans-and paracellular resistance: lessons from impedance spectroscopy,” Annals of the New York Academy of Sciences, vol. 1257, no. 1, 2012, pp. 142–151.

[7] T. Schmid, D. Günzel, and M. Bogdan, “Automated quantification of the capacitance of epithelial cell layers from an impedance spectrum,” in Proceedings of the 7th International Conference on Bioinformatics, Biocomputational Systems and Biotechnologies, 2015, pp. 27–32.

[8] T. Schmid, M. Bogdan, and D. Günzel, “Discerning apical and basolateral properties of HT-29/B6 and IPEC-J2 cell layers by impedance spectroscopy, mathematical modeling and machine learning,” PLOS ONE, vol. 8, no. 7, 2013, p. e62913.

[9] T. Schmid, D. Günzel, and M. Bogdan, “Efficient prediction of x-axis intercepts of discrete impedance spectra,” in Proceedings of the 21st European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning (ESANN), 2013, pp. 185–190.

[10] T. Schmid, “Automatisierte Analyse von Impedanzspektren mittels konstruktivistischen maschinellen Lernens,” Ph.D. dissertation, Universität Leipzig, Germany, 2018.

[11] K. S. Cole and R. H. Cole, “Dispersion and absorption in dielectrics I. alternating current characteristics,” The Journal of Chemical Physics, vol. 9, no. 4, 1941, pp. 341–351.

[12] I. Kasa, “A circle fitting procedure and its error analysis,” IEEE Transactions on Instrumentation and Measurement, vol. 1001, no. 1, 1976, pp. 8–14.

[13] M. Riedmiller and H. Braun, “A direct adaptive method for faster backpropagation learning: the RPROP algorithm,” IEEE International Conference on Neural Networks, 1993, pp. 586–591.

[14] S. Nissen, “Implementation of a fast artificial neural network library,” University of Copenhagen, Denmark, Tech. Rep., 2003.

[15] M. A. Hall, “Correlation-based feature selection for machine learning,” Ph.D. dissertation, University of Waikato, New Zealand, 1999.

[16] T. Schmid, D. Günzel, and M. Bogdan, “Automated quantification of the relation between resistor-capacitor subcircuits from an impedance spectrum,” in Proceedings of the International Conference on Bio-Inspired Systems and Signal Processing, 2014, pp. 141–148.

[17] B. Srinivasan et al., “TEER measurement techniques for in vitro barrier model systems,” Journal of Laboratory Automation, vol. 20, no. 2, 2015, pp. 107–126.

[18] N. Onnela et al., “Electric impedance of human embryonic stem cell-derived retinal pigment epithelium,” Medical & Biological Engineering & Computing, vol. 50, no. 2, 2012, pp. 107–116.

[19] F. A. Atienzar, H. Gerets, K. Tilmant, G. Toussaint, and S. Dhalluin, “Evaluation of impedance-based label-free technology as a tool for pharmacology and toxicology investigations,” Biosensors, vol. 3, no. 1, 2013, pp. 132–156.

[20] S. Ramasamy, D. Bennet, and S. Kim, “Drug and bioactive molecule screening based on a bioelectrical impedance cell culture platform,” International Journal of Nanomedicine, vol. 9, no. 1, 2014, pp. 5789–5809.

[21] Z. Maherally et al., “Real-time acquisition of transendothelial electrical resistance in an all-human, in vitro, 3-dimensional, bloodbrain barrier model exemplifies tight-junction integrity,” The FASEB Journal, vol. 32, no. 1, 2018, pp. 168–182.

 ISBN: 978-1-61208-639-2

The Sustainable Growth Use of Renewable Energy Based on Spatial Energy Planning

Valentinas Klevas and Audrone Kleviene
Laboratory for Renewable Energy and Energy Efficiency
Lithuanian Energy Institute
Kaunas, Lithuania
Valentinas.Klevas@lei.lt

***Abstract*— The aim of the research is to apply the approach enabling assessment of external benefit for the use of renewable energy sources, which is usually not assessed for individual projects. Regional programs may be appropriate tools for reducing renewable energy adaptation costs, increasing competitiveness in the market and promoting the development. Support measures must shape uniform technological programs for long-term periods. The sustainable growth of renewable energy is based on the formulated methodology, which enables to achieve maximal benefit with minimal support for beneficiaries as well as for the state. Modified levelized cost of energy method enables analysis of economic benefit for the selected project investor, as well as relates the achieved results to macroeconomic indicators using external parameters. After the evaluation of external benefit for solar, wind, and geothermal energy projects, the results could be interpolated and used for assessment of selected technology or support scheme for wider use in the context of spatial energy planning in urban areas.**

Keywords- renewable energy, sustainable development, evaluation principles, external benefits.

I. INTRODUCTION

Economic research of Renewable Energy Sources (further in the text – RES) has intensified in the recent years through implementation of European Union directives, with particular emphasis on the role of a territorial aspect (cities, districts). Incentives for wider use of RES are declared both in the European Union and in Lithuanian laws. Adoption of the Directive 2009/28/EC "On the promotion of the use of energy from renewable sources" [1] required the development of a National Renewable Energy Action Plan for 2010-2020 and many other documents at the local level. This indicates that the development of RES takes an increasingly more significant role in the energy policy of all member states, including the Republic of Lithuania.

The main obstacle to the widespread use of RES is the failure to address the key issue – to measure the benefits to the society (reduction of pollution, energy supply for the future generations, huge and never-ending potential of energy resources such as solar and wind energy), when introducing RES-technologies and, on the basis of these benefits, to encourage investors. Therefore, when formally calculating, RES-technologies are not sufficiently competitive compared to fossil-fuel technologies. This work deals specifically with the problem of the impact of RES-technologies. The impact analysis of RES-technologies on social welfare (on job creation) and on health (on the basis of environmental impact) has allowed the formation of a targeted, reliable methodology compatible with the status of existing statistical information. On the basis of it, algorithms calculating the impact for individual types of RES are formed.

Given that in the newly developed Energy Strategy the high hopes to use RES-technologies are linked to the extremely low used energy sources, such as solar, wind and geothermal energy, significant efforts are needed to organize the accounting and statistics of decentralized RES-producers. Apart from this important condition, it is difficult to expect the proper process management of mastering the RES. The analysis of positive and negative impact of RES usage in the research of Lithuania and other countries has shown that the major part of the positive impact of the RES development is related to Green House Gas (further in the text – GHG) and other pollutant emissions, job creation and rural development. The positive effects of biofuels, biogas and biofuel manufacturing sectors are mainly reflected on the increase of employment level in the regions, the reduction of heat prices and lower emissions of some fossil fuels, which in the long run could affect the decline in the incidence of certain diseases. The development of biogas production and use also plays an important role in solving the problem of organic waste management. The main positive effects of solar and wind energy technologies are the reduction of GHG emissions compared to the use of fossil fuels and biofuels as well as the creation of temporary jobs in Lithuania by installing new solar power plants or wind farms. Developing the manufacturing technology industry in Lithuania also has a positive effect on the creation of additional permanent jobs.

Section II describes the concept of external utility of implementation of RES technologies. Evaluation of such utility may show the advantages, which are underestimated in in the investment decisions. Section III presents a solution of the identified scientific problem; it is based on a system of territorial urban planning. In Section IV is presented a description of how RES programs can consolidate related urban development programs. This may serve to formation financing assumptions and sources for their implementation. The methodology, which is based on

 ISBN: 978-1-61208-639-2

spatial planning, identifies certain aspects that usually are not covered by routine investment valuation techniques, namely, the existing problems in cities: high atmospheric pollution, unemployment, etc., may determine the validity conditions that allowing interpolating results to assess technology for spatial energy planning for solving problems mentioned above. Long distances to centralized electricity and heat networks are defind by geolocation factors and may determine financial acceptability of small decentralized renewable energy technologies due to their technical benefits.

II. CONCEPT OF EXTERNAL BENEFIT OF RES TECHNOLOGIES

In recent years, consumption of fossil fuel and mitigation of climate change have become major challenges for governments all over the world. To engage these challenges, many countries are pursuing research, development, and demonstration of RES [2]. In the past few years, usage of RES rapidly increased all over the world. RES have become important alternative energy sources to realize energy diversification. During the last few years, political support for renewable energies has been growing continuously both at the national and international level [3][4].

Looking at renewable energy use since 2006, it was a natural gas price boom, increase of social and political pressure for fast development in clean energy, and financial crisis, which required adequate government measures to stimulate the economy. The industry of renewables could be important to generate employment and stimulate growth [5]. Investment in RES may bring considerable profits, so more and more enterprises will be involved in this field. The increased use of RES in the heat market can significantly alleviate the negative effects of high energy costs on the national economy. Successful commercialization of indigenous, non-fossil energy resources is expected to promote regional economic development and employment, enable to increase national energy security and to reduce a substantial portion of the increasing trade deficit necessity to import fossil fuels [6].

The use of renewable energy in rural as well as urban areas became the significant development thus adding to mitigation of climate change [7][8], reducing differences between rural and urban dwelling options [9], bringing new RES options for diversifying energy supply [10][11][12]. The role of local governance for energy and urban development appears highly important here [13].

The stimulation of energy producers and consumers to use RES is one of the major goals of energy policy in Lithuania. Policies and measures that aim to enhance the use of RES are mainly driven by EU policy. Unstable state energy policy and changes of incentive measures destabilize the investment initiatives. The lack of economic evaluation on both demand and supply sides is the main challenge to achieve the expected goal. Unreasonably huge role is assigned to biomass in usage projections of RES. Biomass is an energy resource that is the result of economic activity and depends on continuity of economic activity. Meanwhile, the use of an inexhaustible solar and geothermal energy potential is insufficient.

The scientific problem is evaluation of the social utility of RES that can show the advantages, which is underestimated in the investment decisions [14]. The most important benefits of RES are inexhaustibility and possibility to ensure sufficiency of energy resources for future generations. Also, utilization of particular RES technologies, such as solar energy, solves environmental issues. Therefore, RES may be additionally financed from other sources. The main issue is the diversity of RES utilization opportunities and incentives. The energy phenomenon is inexhaustibility of RES that could change exhaustible types of fossil fuel, such as oil and natural gas, and could be replaced with few different types of energy. This could be solar, wind, and geothermal energy. On the other hand, social utility of RES differs from the social utility of oil and natural gas, because RES guarantees the supply of energy resources for future generations.

Renewable energy policy is a complex system, where a balance among three aspects of sustainability – environment, economy, and social life – is needed. A good performance and well-balanced RES policy requires efforts from different stakeholders. Although each country has different starting conditions, namely, developed technical infrastructure and own energy resources, development of energy from RES should be shaped on the basis of long-term macroeconomic policy [15]. This means that a good and scientifically-based theoretical background is needed as well as indicators reflecting economic development and a method for assessing the impact of certain type of energy from RES on these indicators are required. In other words, this is a reliable energy supply at an affordable price, which causes as positive environmental impact as possible.

Support for RES is required to promote a wider use of renewable energy. What is more, energy market failure is a serious obstacle to promote the development of RES. External positive benefits of using RES, such as public interest, are not evaluated here. This benefit is related to introduction of new technologies, their development, and positive impact on environment [16][17]. Subsidies for RES must be based on a positive external utility. Therefore, one of the most effective methods could be integration of RES technologies into regional energy development [18]. Regional programs may be appropriate tools for subordinating market mechanisms (reducing renewable energy adaptation costs, increasing competitiveness in the market and promoting the development), which are required to overcome market barriers [19][20].

As a rule, the impact of RES-technologies on economic and social indicators is supported by statistical correlation analysis based on historical trends in RES development and their impact on macroeconomic indicators. However, the anonymity of this method and the probability of a certain error cannot provide a clear understanding of the actual impact of specific parameters when analysing specific projects under individual RES-technologies [21]. Also, this method does not provide tools or ways to optimize the development of the RES-Sector or to put in place support measures to maximize macroeconomic benefits with minimal resources. Therefore, valuation is inseparable from the microeconomic analysis of the project chosen by the individual investor (taking into account the economic benefits to the developer), using external parameters for macroeconomic indicators [22]. Results obtained in one project may be interpolated to assess the potential impact of the support scheme at the urban level.

Levelized Cost Of Energy (LCOE) is one of the most popular approaches for comparison of different energy generation options on equal basis - present value of total life-cycle cost. The elaborated LCOE approach, which was applied in a paper of V. Bobinaite and D. Tarvydas [23] gives an opportunity to quantitatively assess the influence of certain RES support measures on the cost of energy production. Also, it provides the investors a tool which could be used to compare different RES investment projects.

III. Spatial Aspect for the Assessment of Benefit from Using RES Technologies in Urban Areas

The use of renewable energy in rural as well as urban areas became the significant development thus adding to mitigation of climate change, reducing differences between rural and urban dwelling options [24], bringing new RES options for diversifying energy supply [25]. The role of local governance for energy and urban development appears highly important here [26].

The stimulation of energy producers and consumers to use RES is one of the major goals of energy policy in Lithuania. Policies and measures that aim at enhancing the use of RES are mainly driven by EU policy. The scientific problem is evaluation of the social utility of RES that can show the advantages, which is underestimated in the investment decisions [27][28]. Moreover, it is transfer of their economic interpretations on uniform rules and economic laws into specified dimension, in this case trying to reduce different opinions as much as possible [29].

The main indicators of external benefit are considered to be the use of domestic capital; taxes paid to the state budget and newly created job positions. Combination of different types of RES may create a large external effect and make a significant impact not only on individual consumers of RES, but also on the society. Therefore, the benefit from using RES, especially in urban areas, should be evaluated during the decision making process.

Each RES has different value in terms of external benefit, thus respective promotion schemes can and must differ significantly. This depends on the environmental situation in certain area. External conditions and current infrastructure for RES technologies differ, thus spatial planning, such as urban energy development plans, is the most appropriate tool for the benefit assessment. All countries pay little attention to the support of RES technologies on the demand side. Support for the consumers is the most appropriate via support programs (and not by supporting scattered consumers), as it enables to achieve actual effect. The external benefit analysis for small wind power plants, geothermal heat pumps, and solar collectors was performed to find their niche in urban and rural areas. The external effect indicators for different scenario cases are shown in Table 1. Data in Table 1 show the comparison of external benefit using different types and scale of RES technologies. The main indicators of external benefits are considered to be the use of domestic capital; taxes paid to the state budget and newly created job places.

TABLE 1. External benefit of using solar, wind and geothermal energy technologies (case of Lithuania)

Indicators	**Wind power plants**		**Geothermal heat pump and solar collectors in multifamily building**	**Solar collectors in district heating system**
	4 kW	***10kW***	***126kW***	***7000kW***
Production, MWh	6,8	17,5	185	4500
Domestic capital, 1000 EUR	8.02	21.08	25.01	1380.88
Social insurance, 1000 EUR	0.81	1.51	9.27	119.23
Personal income taxes, 1000 EUR	0.29	0.55	3.46	44.46
New jobs (man months)	2.3	4.2	31.9	409.4

Combination of different types of RES may create a large external effect and make significant impact not only on individual consumers of RES, but on the society as well. Therefore, the benefit from using RES, especially in urban areas, should be evaluated during the decision making process. If direct support is applied for RES or environmental technologies, the impact will depend on the volume of support and the source of funding. For example, if the direct assistance is financed from the state budget, the promotion measures would increase the budget deficit in addition to the positive aspects mentioned above. Therefore, it is essential that the promotion would be cost-effective: the benefits should exceed the negative effects. Inexhaustible types of RES (wind, solar, and geothermal energy) enable greater benefit than the subsidies it might require, while penetration into the market is growing.

The idea of eco-villages has been analysed since the last decades of XX century all over the world. Modelling methods were used for resources management and sustainability assessment [30]. The most recent research concentrates on eco-innovations, eco-efficiency and eco-effectiveness, cultivations of eco-sustainability on various social-economic-environmental levels. The investigations also involve the role of eco-cultural diversity, renewable energy in eco-communities and understanding the eco-complexity and ecosystem approach [31].

The principle of ecological movement can be expressed by the quote “think globally, act locally”. Progress in solving environmental problems can be achieved more quickly if the principles of our activities and the functions of natural laws system are harmonized. This was the main principle of the development of green settlements. The energy sector of green settlements duplicates the processes in the nature which is a perfect example of waste-free technology. The volume of waste generated in energy sector can be reduced by increasing the usage efficiency of primary energy sources and promoting the use of RES. The negative impact to the environment can be mitigated through more efficient use of RES in energy sector developed under the circular economy model and bioregional development principles. The use of RES in the regions can create unique environment for resource utilization. The exploitation of energy resources must not exceed the nature's resilience limits otherwise RES can be treated as non-renewable resources. The development of regional energy sector, according to the principles of circular economy, is encouraging the use of a wide range of local and renewable resources.

IV. BALANCING OF RES SUPPORT FORMS BY USING SPATIAL PLANNING METHOD

Different types of RES have a very broad and distinctive scale of use, therefore, each renewable requires its own specific support. This topic has been focussed on the potential analysis of wind, solar, and geothermal energy for small scale applications. Every renewable has a wide developing scale of technologies, which is becoming cheaper due to the growth of supply. What is more, every renewable has different significance level in terms of external benefit, thus support schemes can differ significantly. This depends on the environmental situation in various areas. External conditions and current infrastructure, where RES technologies could be applied, differ as well; therefore, spatial planning via urban energy development plans is the most appropriate tool for benefit assessment.

Most countries support RES technologies mainly on the supply side, while existing opportunities to use solar energy at the consumers' side, such as small systems of solar collectors in multi-apartment buildings, are significant as well. Support for the consumers is the most appropriate via support programs (and not by supporting scattered consumers), as it enables achieving actual effect. Renovation of buildings by districts enables planned actions, which could solve technical and financial problems for district heating companies due to reduced consumption.

Application of the schemes mentioned above for small producers increases administration costs and is often not acceptable due to high complexity level. Therefore, territorial administration is needed. The main problem is that the programs for implementation of legally delegated functions have no consolidating unity. This might lead to serious problems: renovation of multi-apartment residential houses would reduce the heat consumption significantly. As a result, the district heating tariffs would increase due to growing fixed heat generation and supply costs.

Although the EU directives and national laws provide for the functions of the municipalities of the Republic of Lithuania in developing software documents for the energy sector in general and for RES-technologies are rather precise, in practice, plans are just beginning to be formed. Having analyzed the current laws of the Republic of Lithuania, it is clear that they regulate in a sufficiently detailed manner the responsibilities of municipalities and the development of the use of RES. But the main problem is that separate programs are being developed for the implementation of individual EU directives, which do not have a consolidated unity. Meanwhile, functions such as the Special Needs Planning Development Plan, as well as future Environmental Protection Plans, which are already beginning to plan for the implementation of commitments made under the Paris Agreement, are directly linked to the development of RES, are formed separately. In this way, the financial resources and modest forces of specialists working in municipalities are dispersed. A stable and balanced program with carefully selected subsidizing, enabling the interaction between consumers and heat suppliers, is necessary for solving these problems, as proposed in Fig. 1.

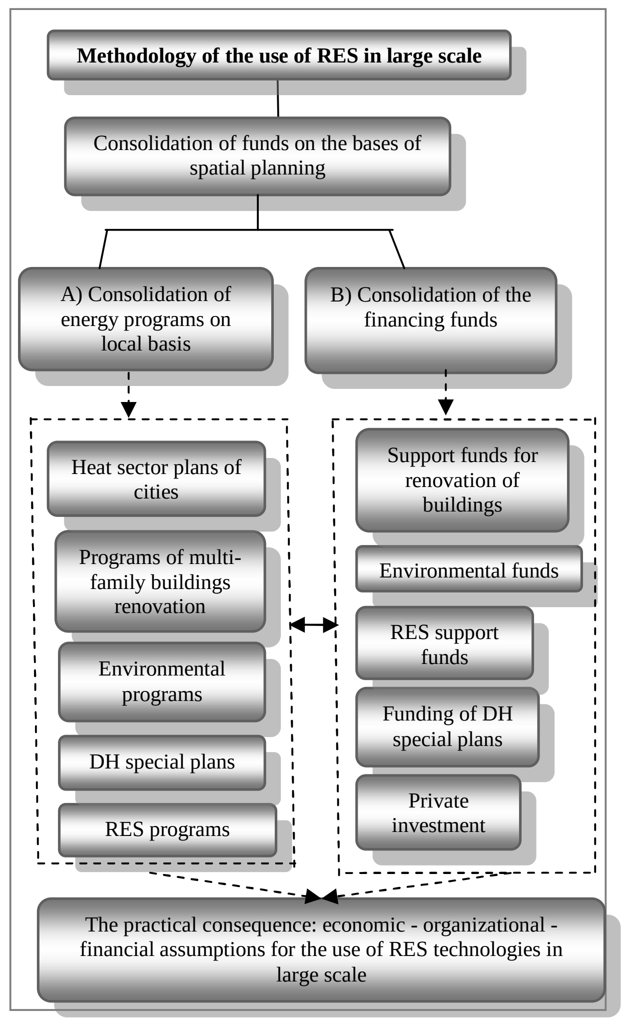


Figure 1. Methodology of the use of RES in large scale.

Periodization of economic age and assessment of various RES support forms, with regard to the specifics, has

 ISBN: 978-1-61208-639-2

different goals and tasks from the macroeconomic perspective.

There are two blocks of local programs and financing sources, which are to be described: A) and B).

A. Consolidation of cities energy programs on the uniform basis of knowledge economy

Lithuanian legal system of self-government has granted fairly broad opportunities to participate in increasing energy efficiency and development of renewable energy. Law on Local Self-Government in Lithuania distinguishes independent functions of municipalities, such as preparation and implementation of municipal strategic planning documents and planning documents implementing them. Law on Energy from Renewable Sources initiates preparing local RES development plans. Municipalities also administer the implementation of measures of Rural Development Programme such as carrying out the delegated function of the State. Law on Local Self-Government establishes that municipalities organize the heat supply within their territory. Law on Heat Sector regulates the special planning of heat, which is one of the implementation mechanisms of those obligations. In addition to broad responsibilities for the preparation of planning documents, the role of municipalities has been reinforced for renovation (modernization) of multi-family buildings in recent years.

The need for single energy policy formation with an economic support system for RES (as one of the compounds) has been notified for several years already.

Such program consolidation might be organizational and the integration of technologies could be financed using Structural Funds.

B. The second block shows consolidation of the financing funds program for sustainable development.

It is necessary not only to foresee the results that are expected in the use of RES but also to form the best prerequisites and financial resources that it would be possible to realize positive external effects. However, considering the possibility of incorporating efficient use of resources and RES into planning documents of municipalities, there is an issue of financing. Independent functions of municipalities, including implementation of the program documents, are funded from local budgets. These funds are limited; therefore, it is difficult to expect a decision of allocating funds to these areas on a larger scale. Therefore, the main role of municipalities is organizational, which would allow mobilizing larger resources. A special issue is lack of energy specialists who could prepare complex programs.

Scattered generation character and the fact that most of these installations can be implemented on demand side are left for private initiative. In order to involve hardly competing energy sources, the institution for investment management is required. Municipalities could act as institutions seeking for the development of wide scale RES technologies. Actual legislation in Lithuania shows that municipal powers and responsibilities regarding RES are regulated in detail. There are no barriers, except financing issues for the actions of municipalities. Some of them have signed the Covenant of Mayors; some have adopted RES development action plans, however, only a few of them are actually acting.

The promotion of RES on wider scale was implemented in EU by creating green settlements or so-called eco villages. [32]. For example, the state funded the project Climate Menu in Netherlands, which enabled towns to select sectors which needed exceptional attention in order to meet national obligations. Environmental policy is implemented by towns, where the best possibilities are concentrated and problems are the most visible.

V. CONCLUSION

A particular obstacle to the widespread use of RES is the lack of evaluation of the benefits to society (reduction of pollution, energy supply of the future for generations, huge and never-ending potential of energy resources such as solar, wind energy) when introducing RES-technologies. Therefore, when formally calculating, RES-technologies are not sufficiently competitive compared to fossil-fuel technologies. This work deals specifically with the problem of the impact of RES-technologies.

The analysis of the positive and negative impact of the use of RES on research in Lithuania and other countries has shown that the major part of the positive impact of the development of RES is related to GHG and other pollutant emissions, job creation and rural development, increase in employment in the regions, the reduction of heat prices and the lower emissions of some fossil fuels. Investigation has allowed the formation of a targeted, reliable methodology compatible with the status of existing statistical information.

Each RES has a different value in terms of external benefit, thus respective promotion schemes can and must differ significantly. This depends on the environmental situation in certain area. External conditions and current infrastructure for RES technologies differ, thus spatial planning, such as urban energy development plans, is the most appropriate tool for the benefit assessment.

Combination of different types of RES may create large external effect and make a significant impact not only on individual consumer of RES, but on the whole society as well. Therefore, the benefit from using RES, especially in urban areas, should be evaluated during the decision making process.

Urban programs should be the main form for the consolidation of investment funds and promotion on the basis of possible rational use of investment as a complex macroeconomic effect can be measured and achieved on the territorial basis.

ACKNOWLEDGEMENT

This research was supported by the Research Council of Lithuania (09.3.3-LMT-K-712-06-0236).

REFERENCES

[1] European Council. Directive 2009/28/EC of the European Parliament and of the Council on the promotion of the use of energy from renewable sources and amending and

subsequently repealing Directives 2001/77/EC and 2003/30/EC. Off. J. Eur. Communities, 5.6.2009, L 140/16, 2009.

[2] Y. C. Shen, G.T. Lin, K. P. Li, B. J. and C Yuan, "An assessment of exploiting renewable energy sources with concerns of policy and technology," Energy Policy, vol. 38, pp. 4604–4616, 2010. doi:10.1016/j.enpol.2010.04.016.

[3] European Commission. State aid: Commission adopts new rules on public support for environmental protection and energy [Online]. Available from: http://europa.eu/rapid/press-release_IP-14-400_en.htm. 2014.

[4] European Commission. European Commission guidance for the design of renewable energy support schemes, Commission staff working document, 2013.

[5] A. C. Marques, J. A. Fuinhasand and J. R. Pires Manso, "Motivations driving renewable energy in European countries: a panel data approach," Energy Policy, vol. 38, pp. 6877–6885, 2010 doi:10.1016/j .enpol. 2010.07.003.

[6] R. Camagni, "On the concept of territorial competitiveness: sound or misleading?," Urban Stud., vol. 39, pp. 2395–2411, 2002.

[7] B. J. de Vries, D. P. van Vuuren and M. M. Hoogwijk, "Renewable energy sources: their global potential for the first-half of the 21st century at a global level," Energy Policy, vol. 35, pp. 2590–2610, 2007. doi:10.1016/j.enpol.2006.09.002.

[8] N. L. Panwar, S. C. Kaushik and S. Kothari "Role of renewable energy sources in environmental protection: a review." Renew Sustain Energy Rev., vol. 15, pp. 1513–152, 2011. doi:10.1016/j.rser.2010.11.037.

[9] A. Bergmann, S. Colombo and N. Hanley, "Rural versus urban preferences for renewable energy developments" Ecol Econ. , vol. 65,:616–625, 2008. doi:10.1016/j.ecolecon. 2007.08.011.

[10] J. Hofierka and J. Kaňuk," Assessment of photovoltaic potential in urban areas using open-source solar radiation tools," Renew Energy, vol. 34, pp. 2206–2214, 2009. doi:10.1016/j.renene.2009.02.021.

[11] W. T. Chong, M. S. Naghavi, S. C. Poh, T. M. I. Mahlia and K. C. Pan, "Techno-economic analysis of a wind–solar hybrid renewable energy system with rainwater collection feature for urban high-rise application," Appl Energy, vol. 88, pp. 4067–4077, 2011 doi:10.1016/j.apenergy.2011.04.042.

[12] P. Droege, Urban Energy Transition. Elsevier, 2008. doi:10.1016/B978-0-08-045341-5.00029-3.

[13] V. Castán Broto and H. Bulkeley,"A survey of urban climate change experiments in 100 cities," Glob Environ Change, vol. 23, pp. 92–102, 2013. doi:10.1016/j.gloenvcha.2012.07.005.

[14] E. Strantzali and K. Aravossis, "Decision making in renewable energy investments: A review," Renew Sustain Energy Rev., vol. 55, pp. 885–898, 2016. doi:10.1016/j.rser.2015.11.021

[15] V. Klevas, Justification of long-term economic policy of renewable energy sources. ISBN 978-1-63483-203-8 (print). ISBN 978-1-63484-020-0 (e-book). New York: Nova Science Publishers, 2015.

[16] G. Tanguay, J. Rajaonson, J. Lefevre and P. Lanoie, "Measuring the sustainability of cities: An analysis of the use of local indicators," Ecol Indic., vol. 10, pp. 407–418, 2010.

[17] Y. Xing, R. Horner, M. El-Haram and J. Bebbington, "A framework model for assessing sustainability impacts of urban development," Account Forum, vol. 33, pp. 209–224, 2009.

[18] V. Klevas, Regional approach for policies and measures aiming to sustainable energy development. In: Paths to sustainable energy, Rijeka: InTech, , pp. 117-132, 2010 ISBN 978-953-307-401-6.

[19] V. Klevas, Analysis of support assumptions and measures for promotion of renewable energy sources demand in regional aspect. In: Advances in energy research, New York: Nova Science Publishers, pp. 99–122, 2012. ISBN 978-1-61470-485.

[20] V. Klevas, K. Bieksa and L. Murauskaite, "Innovative method of RES integration into the regional energy development scenarios," Energy Policy, vol. 64, pp. 324–336, 2014. doi:10.1016/j.enpol.2013.08.088.

[21] R. Wang, D. Hu, F. Li and B. Larry, "Understanding eco-complexity: Social-Economic-Natural Complex Ecosystem approach," Ecol Complex., vol. 8, pp. 15–29, 2011. doi:10.1016/j.ecocom.2010.11.001.

[22] L. Dusonchet, E. Telaretti, "Economic analysis of different supporting policies for the production of electrical energy by solar photovoltaics in eastern European Union countries," Energy Policy, vol. 38, pp. 4011–4020, 2010. doi:10.1016/j.enpol.2010.03.025.

[23] V. Bobinaite and D. Tarvydas, "Financing instruments and channels for the increasing production and consumption of renewable energy: Lithuanian case," Renew Sustain Energy Rev., vol. 38, pp. 259–276, 2014. doi:10.1016/j.rser.2014.05.039.

[24] I. Harmaajärvi, "EcoBalance model for assessing sustainability in residential areas and relevant case studies in Finland," Environ Impact Assess Rev., vol. 20, pp. 373–380, 2000. doi:10.1016/S0195-9255(00)00048-2.

[25] A.P. Palojarvi, Experiences with ecological technologies and practices. Vilnius: BMK Press, 2013.

[26] K. Tzoulas et al., "Promoting ecosystem and human health in urban areas using Green Infrastructure: a literature review," Landsc. Urban Plan., vol. 81, pp. 167–178, 2007. doi:10.1016/j.landurbplan.2007.02.001.

[27] H. Bulkeley and M. Betsill, "Rethinking Sustainable Cities," Multilevel Governance and the "Urban" Politics of Climate Change, vol.14, pp. 42–63, 2005. doi:10.1080/0964401042000310178.

[28] C. Kalenda, Ecological ethics [in Lithuanian]. Vilnius: Rosma, 2007.

[29] B. Bayulken and D. Huisingh, "Are lessons from eco-towns helping planners make more effective progress in transforming cities into sustainable urban systems: a literature review," Clean Prod., 2015. doi:10.1016/j.jclepro.2014.12.099.

[30] D. Hu and R. Wang, "Exploring eco-construction for local sustainability: An eco-village case study in China," Ecol. Eng., Vol. 11, pp. 167–176, 1998. doi:10.1016/S0925-8574(98)00032-9.

[31] Y.Y. Zheng and S. Gao, "Exploration on the Eco-Village Planning from the Perspective of Low-Carbon," Appl. Mech. Mater., pp. 357–360,2013. doi:10.4028/www.scientific.net/AMM.357-360.1945.

[32] H. Barton, Sustainable Communities: The Potential for Eco-neighbourhoods. London: Earthscan Publications, 2000.