Structure-property based model estimation of alkanes boiling points

SORANA-DANIELA BOLBOACĂ[1], LORENTZ JÄNTSCHI[2]

[1] "Iuliu Haţieganu" University of Medicine and Pharmacy
13 Emil Isac Street, 400023 Cluj-Napoca, Romania
http://sorana.cademicdirect.ro
sorana@j.academicdirect.ro

[2] Technical University of Cluj-Napoca,
15 Constantin Daicoviciu Street, 400020 Cluj-Napoca, Romania
http://lori.academicdirect.org
lori@j.academicdirect.org

Key Words and Phrases: Molecular Descriptor Family on Structure-Property Relationships (MDF SPR), Models assessment, Boiling point, Alkanes

AMS Subject Classification: 03H05 (Nonstandard models in mathematics), 62P35 (Applications to physics), 93E24 (Least squares and related methods), 93E35 (Stochastic learning and adaptive control)

This study discusses a family of molecular descriptors on structure-property relationships (MDF SPR) to model the boiling points of alkanes based on their chemical structure.

The proposed approach uses the complex information obtained from the all alkanes from C3 to C9 structures in order to generate and calculate the molecular descriptors family. The structure-property relationship models were built based on the generated descriptors.

The obtained models (model with one and two descriptors, respectively) were validated through the assessment of the cross-validation leave-one-out score. The comparison between the uni-varied model and the model with two descriptors was performed using Steiger's Z test. The best performing MDF SPR model was validated, and its correlation coefficient was compared with a previously reported model.

The analysis of the statistical characteristics of the obtained models demonstrated that the model with two descriptors has greater abilities in estimation and prediction compared with the model with one descriptor. This observation was also sustained by the results of training versus test analysis.

The results of this study revealed that the MDF SPR approach is a useful method to model the boiling points of alkanes providing stable models.

Structure-Property Based Model Estimation of Alkanes Boiling Points

Sorana-Daniela BOLBOACA & Lorentz JANTSCHI

“Iuliu Haţieganu” University of Medicine and Pharmacy & Technical University - Cluj-Napoca Romania

Outline

- **Introduction – modeling of compounds properties**
- **Research aim**
- **Material – alkanes**
- **Method – Molecular Descriptors Family on Structure-Property Relationships (MDF SPR)**
- **Results**
- **Conclusions**

Introduction

- **Quantitative Structure-Property Relationship – QSPR**

- **1868: physiological action of the ammonium salts [Crum-Brown and Fraser, 1868]**
- **1893: chemical structure and oil-water partition coefficient [Richet, 1893]**

Crum-Brown, A. and Fraser, T.R. Trans. R. Soc. Edinbours 25 (1868) 151–203.
Richet, M. C. Compt. Rend. Soc. Biol. 45 (1893) 775–776.

Introduction

- **surface tension [***Delgado & Diaz, 2006***]**
- **electrochemical degradation of substituted phenols [***Yuan et al., 2006***]**
- **infinite dilution activity coefficients [***Tamm & Burk, 2006***]**
- **n-octanol water partition coefficients [***Zhou et al., 2005***]**
- **Henry's law constant [***Modarresi et al., 2005***]**

Delgado, E. J., Diaz, G. A. SAR QSAR Environ. Res. 17 (2006) 483–496.
Modarresi, H., Modarress, H., Dearden, J. C. SAR QSAR Environ. Res. 16 (2005) 461–482.
Tamm, K., Burk, P. J. Mol. Modeling 12 (2006) 417–421.
Zhou, W., Zhai, Z., Wang, Z., Wang, L. J. Mol. Struct. THEOCHEM 755 (2005) 137–145.
Yuan, S., Xiao, M., Zheng, G., Tian, M., Lu, X. SAR QSAR Environ. Res. 17 (2006) 473–481.

Introduction

- **Boiling point of alkanes [*Toropov et al., 1998*] - 3D molecular descriptors**

$$Bp(°C) = 727.26 \cdot {}^3D0\chi - 19.46 \cdot 3DSRW2 + 7.99 \cdot M2 - 779.42$$

Eq1

n = 73; r = 0.998; s = 2.17; F = 8340

  - **3D0χ and 3DSRW2 are MIS (Method of Ideal Symmetry) indices**
  - **M2 is a 3D modification of the Zagreb index**

Toropov, A., Toropov, A., Ismailov, T., Bonchev, D. J. Mol. Struct. THEOCHEM 424 (1998) 237–247.

Introduction

- **Molecular Descriptor Family on Structure-Property Relationships (MDF SPR) [***Jäntschi, 2005***]**
- **Estimation and prediction abilities [***Jäntschi and Bolboaca, 2007***]:**
  - **retention chromatography index**
  - **octanol/water partition coefficients**
  - **water activated carbon adsorption**
  - **molar refraction**

Jäntschi, L. Leonardo Electronic Journal of Practices and Technologies 6 (2005) 76-98.

Jäntschi, L. and Bolboaca, S.D., International Journal of Molecular Sciences 8 (2007) 189-203

Research Aim

- **Is there any linear relationship between alkanes structure and boiling point?**
- **Is a strong relationship?**
- **The model has prediction abilities?**

Alkanes set & Boiling points

- **Alkanes: 73**
  - **1: C3 alkane**
  - **2: C4 compounds**
  - **3: C5 compounds**
  - **5: C6 compounds**
  - **9: C7 compounds**
  - **18: C8 compounds**
  - **35: C9 compounds**

Alkanes set & Boiling points

- **experimental boiling points [*Basak et al., 1991*]**

- **6 boiling points were corrected by Herndon [*Herndon, unpublished*]**

Basak, S. C., Niemi, G. J., Veith, G. D.: Predicting properties of molecules using graph invariants, J. Math. Chem. 7 (1991) 243–272.
Herndon, W. C. Unpublished communication.

MDF on SPR

- **Step I:** drawn and optimized HyperChem 7.0
- **Step II:** file with experimental boiling points
- **Step III:** **g**enerate the MDF members
  - **73 compounds**
  - **Seven-characters name**

MDF on SPR

- **Step IV:** Finding the MDF SPR models
- **Step V:** Models validation
  - sqaured correlation coefficient (r^2)
  - regression parameters
  - the cross-validation leave-one-out analysis ($r^2_{cv\text{-}loo}$)

http://vl.academicdirect.org/molecular_topology/mdf_findings/loo

MDF on SPR

- **Step VI: Models validation**
  - **Training versus test analysis**
  - **24 investigations:**
    - Training: 40 to 63
    - Test: 33 to 10
  - **Fisher's Z-test: significance level of 5% [*Steiger, 1980*]**

http://vl.academicdirect.org/molecular_topology/qsar_qspr_s/

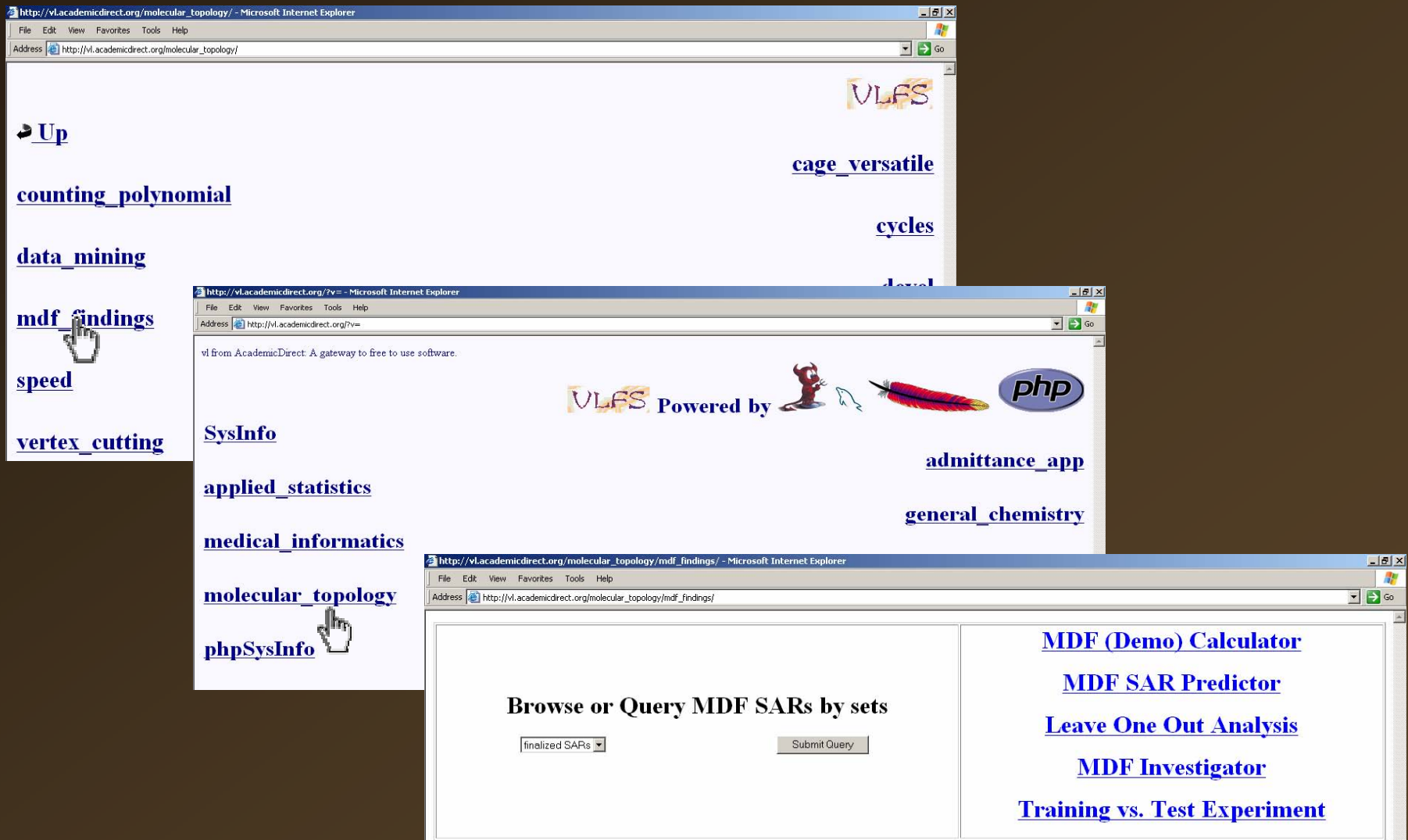
Steiger, J.H. Psychol. Bull. 87 (1980) 245–251.

MDF on SPR

- Step VII: Comparisons
  - **Steiger's Z test [*Steiger, 1980*]**
  - **significance level: 5%**

Steiger, J.H. Psychol. Bull. 87 (1980) 245–251.

Results: online



Results: MDF on SPR models

- **MDF on SPR model with one descriptor:**

$$\hat{Y}_{1D} = -507.95 + 188.40 \cdot lbMdsHg$$

Eq2

- **MDF-SAR model with two descriptors:**

$$\hat{Y}_{2D} = -129.20 - 67.45 \cdot lGDrtGt + 4.89 \cdot IbDrfHt$$

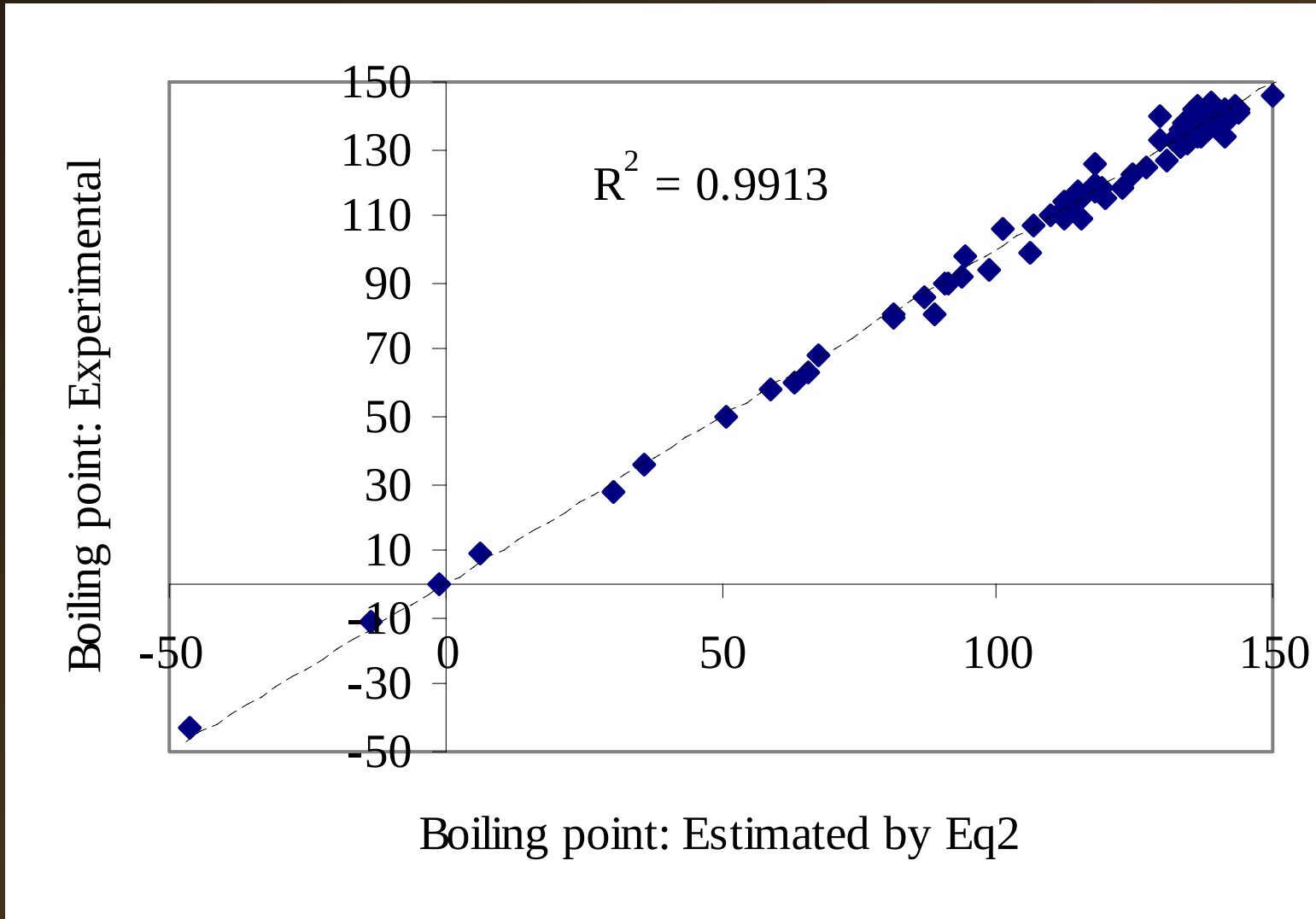
Eq3

Results: statistics

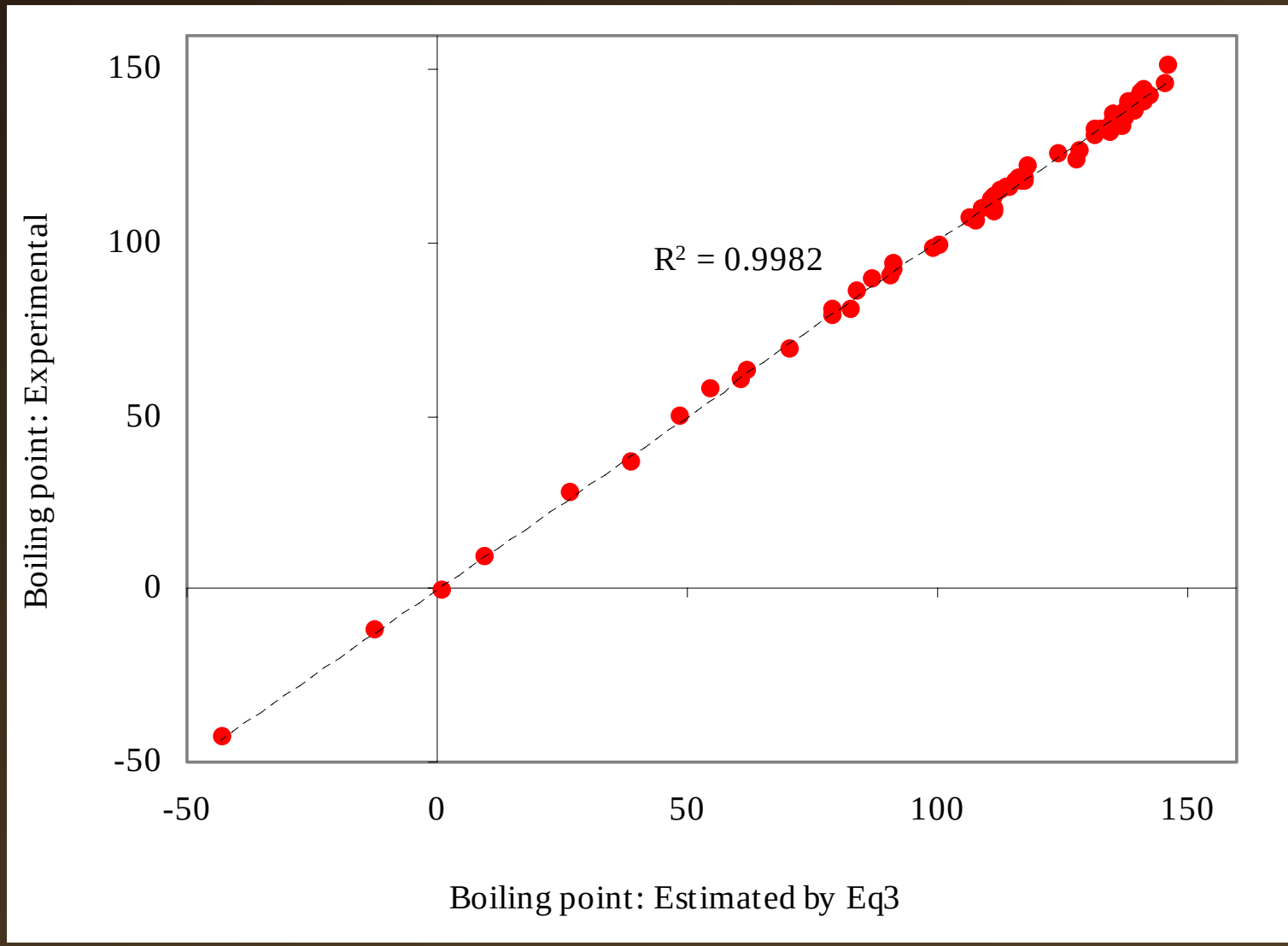
Eq.	$95\%CI_{inter}$	$95\%CI_{C\text{-}Desc}$	r^2_{MDF}	F_{MDF}	s_{MDF}	r^2_{loo}	F_{loo}	s_{loo}
2	[-521.70,-494.22]	[184.22, 192.60]	0.991	8048‡	3.81	0.991	7654‡	3.91
3	[-132.23,-126.16]	[-68.30, -66.60] [4.81, 4.97]	0.998	19361‡	1.74	0.998	17837‡	1.82

$95\%CI_{inter}$, $95\%CI_{C\text{-}Desc}$ = the 95% confidence interval for the intercept, and for the coefficient of descriptor; r^2_{MDF}, r^2_{loo} = the squared correlation of the MDF SPR model, and for the leave-one-out analysis; F_{MDF}, F_{loo} = the Fisher parameter of the MDF SPR, and leave-one-out regression models; s_{MDF}, s_{loo} = standard error for the MDF SPR model, and leave-one-out model, respectively; ‡ $p < 0.0001$

Results: Graphical Representation



Results: Graphical Representation



Results: Eq2 vs Eq3

- **Eq2:**
  - **r = 0.9956; 95%CI = [0.9929-0.9972]**
- **Eq3:**
  - **r = 0.9991; 95%CI = [0.9985-0.9994]**

- **Steiger's Z-parameter = 7.016**
- **p < 0.0001**

Results: Training vs Test Analysis

Training set			Test set			
No_{tr}	r_{tr}	$95\%CIr_{tr}$	No_{ts}	r_{ts}	$95\%CIr_{ts}$	Zr_{tr}-vs r_{ts}
40	0.9992	[0.9987 - 0.9994]	33	0.9990	[0.9984 - 0.9993]	0.454
41	0.9993	[0.9988 - 0.9995]	32	0.9986	[0.9977 - 0.9991]	1.406
42	0.9989	[0.9982 - 0.9993]	31	0.9993	[0.9988 - 0.9995]	0.913
43	0.9988	[0.9980 - 0.9992]	30	0.9994	[0.9990 - 0.9996]	1.392
44	0.9987	[0.9979 - 0.9991]	29	0.9994	[0.9990 - 0.9996]	1.543
45	0.9993	[0.9988 - 0.9995]	28	0.9986	[0.9977 - 0.9991]	1.373
46	0.9991	[0.9985 - 0.9994]	27	0.9993	[0.9988 - 0.9995]	0.493
47	0.9990	[0.9984 - 0.9993]	26	0.9992	[0.9987 - 0.9994]	0.434
48	0.9993	[0.9988 - 0.9995]	25	0.9979	[0.9966 - 0.9986]	2.113†
49	0.9994	[0.9990 - 0.9996]	24	0.9985	[0.9976 - 0.9990]	1.74†
50	0.9984	[0.9974 - 0.9989]	23	0.9995	[0.9992 - 0.9996]	2.179†
51	0.9992	[0.9987 - 0.9994]	22	0.9987	[0.9979 - 0.9991]	0.896

Results: Training vs Test Analysis

Training set			Test set			
No_{tr}	r_{tr}	$95\%CIr_{tr}$	No_{ts}	r_{ts}	$95\%CIr_{ts}$	Zr_{tr}-vs r_{ts}
51	0.9992	[0.9987 - 0.9994]	22	0.9987	[0.9979 - 0.9991]	0.896
52	0.9991	[0.9985 - 0.9994]	21	0.9992	[0.9987 - 0.9994]	0.214
53	0.9991	[0.9985 - 0.9994]	20	0.9992	[0.9987 - 0.9994]	0.21
54	0.9991	[0.9985 - 0.9994]	19	0.9993	[0.9988 - 0.9995]	0.439
55	0.9990	[0.9984 - 0.9993]	18	0.9994	[0.9990 - 0.9996]	0.872
56	0.9992	[0.9987 - 0.9994]	17	0.9985	[0.9976 - 0.9990]	1.047
57	0.9991	[0.9985 - 0.9994]	16	0.9992	[0.9987 - 0.9994]	0.191
58	0.9991	[0.9985 - 0.9994]	15	0.9992	[0.9987 - 0.9994]	0.185
59	0.9993	[0.9988 - 0.9995]	14	0.9965	[0.9944 - 0.9978]	2.442†
60	0.9990	[0.9984 - 0.9993]	13	0.9995	[0.9992 - 0.9996]	1.011
61	0.9992	[0.9987 - 0.9994]	12	0.9962	[0.9939 - 0.9976]	2.177†
62	0.9992	[0.9987 - 0.9994]	11	0.9920	[0.9872 - 0.9949]	3.061†
63	0.9992	[0.9987 - 0.9994]	10	0.9971	[0.9953 - 0.9981]	1.614

† $p < 0.05$

Results: Eq1 vs Eq3

- **Eq1: r = 0.998**
- **Eq3: r = 0.999; 95%CI = [0.998 - 0.999]**

- **Steiger's Z parameter = 2.8**
- **$p = 2.6 \cdot 10^{-3}$**

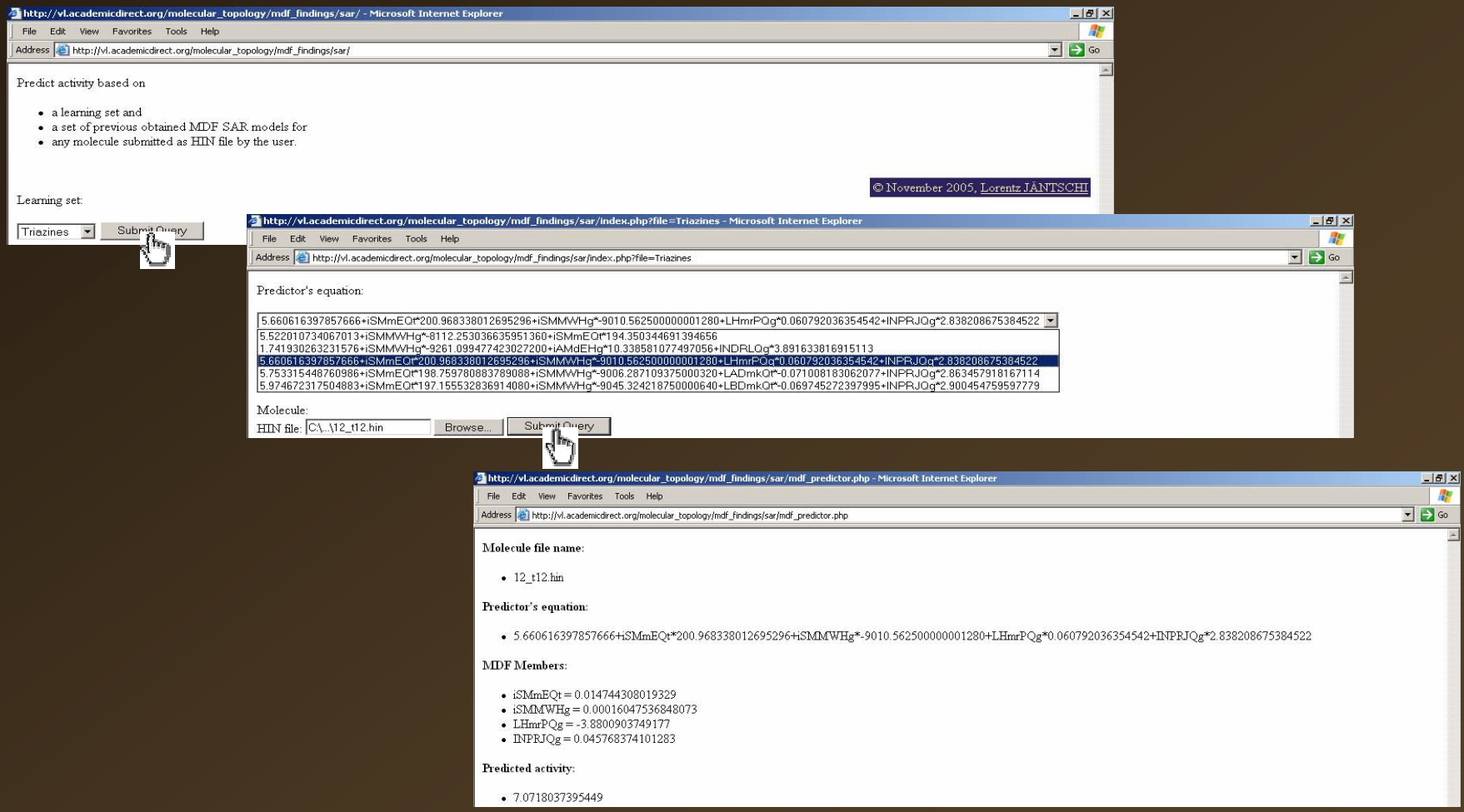
Conclusions: MDF on SPR

- **Two MDF SPR models with two descriptors proved to be able to estimate and predict the boiling points of the alkanes with variable number of atoms (3-9).**
- **The analysis of the correlation coefficients of the MDF SPR models revealed that the model with two descriptors is better than the model with one descriptor.**

Conclusions: MDF on SPR

- **The descriptors involved in the MDF SPR models were calculated solely from the chemical structure and shown that the boiling points of the studied alkanes depend on the topology of the compounds and correlate with the group electronegativity and with the number of directly bonded hydrogens.**
- **The internal validation of the MDF SPR model with two descriptors demonstrates the stability and reliability of the model.**

Conclusions: new alkanes virtual investigation



Conclusions

- **The MDF on SPR methodology opens a new pathway in:**
  - **understanding the relationships between alkanes structure and boiling point**
  - **characterization**
  - **investigation**

of other alkanes in an virtual experimental lab, free of experimental errors

Acknowledgments

- **UEFISCSU Romania – ET36/2005 & ET108/2006**

- **Fourth International Conference of Applied Mathematics and Computing Organizers – Professor DRUMI BRAINOV**

Thank you for attention!

- **Sorana D. BOLBOACĂ**

  http://sorana.academicdirect.org

  &

- **Lorentz JÄNTSCHI**

  http://lori.academicdirect.org