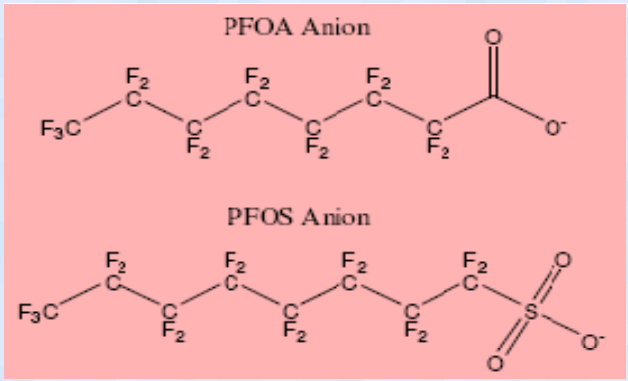
QSPR Studies for Predicting Physico-chemical Properties of Perfluorinated Compounds

Barun Bhhatarai, Paola Gramatica
QSAR and Environmental Chemistry Research Unit
Department of Structural and Functional Biology (DBSF)
University of Insubria, Varese, Italy

Conferentia Chemometrica
2009 Siòfok, Hungary
28 September, 2009

Introduction

Perfluorinated chemicals (PFCs) are compounds containing a long fluorinated carbon tail attached to different functional groups esp. polar heads



- PFCs as perfluoro-octanesulfonate (PFOS), perfluoro-octanoate (PFOA) and perfluoro-octane sulfonylamide (PFOSA) are stable chemicals with a wide range of industrial and consumer applications [*Inoue 2004*]
- Degradable products of commercial PFCs are found in environment and biota but physico-chemical and toxicological data are scarce
- PFCs are considered emerging pollutants and are believed to have potential toxic effects in humans and wildlife [*Lau 2007*]

- chemically and thermally inert
- resistant to stress
- hydrophobic and hydrophilic

- non-adhesives, waterproof fabrics, pesticides …

Environmental & Health concerns

Water → **A large portion of PFCs ends up in untreated wastewater, the discharge of water from sewage treatment ends up in streams and rivers.**

Wildlife → **down … nment … vildlife**

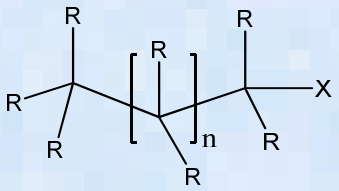
- longer fluorocarbon aliphatic chain containing compounds are incorporated into bilayers of biological membranes
- diPAPs (a group of PFCs used on food wrappers) was recently reported in human blood [*Renner 2009*]

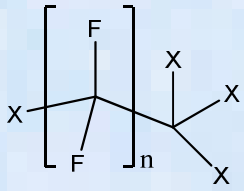
Developmer & Reproduc toxicity → **…4) and … been studied and reported for PFCs**

General toxicity → **PFCs suppresses antibody production, causes thymic and splenic atrophy, induce tumor, alters T-cell populations and are endocrine disruptors (Lau 2007, Kovarova 2008)**

Available Data & Predictions

- Data belonging to Toxicity and Physico-chemical properties of Perfluoro chemicals are limited
- Data expanded by adding multifluoro chemicals
- Prediction only for Perfluoro chemicals (PFCs)





R={H,F} X={-H, -OH, $-SO_2$, -COOH,...} X= {-F, -H, -OH, -alkyl, -aryal, -halo, -nitro,...}

A: Perfluoro chemicals B: Multifluoro chemicals

n= 1,2,3...

Available data points for QSPR modeling of physico-chemical properties

111 Melting Point,
180 Boiling Point and
36 Vapor Pressure

Predictive QSPR approach on physico-chemical properties of PFCs

248 Melting Point,
290 Boiling Point and
243 Vapor Pressure

EU-FP7 project: CADASTER*

4 classes of emerging pollutants studied:

Flame retardants, Fragrances, PerFluoro Chemicals **and (benzo)Triazoles**

WP3: QSAR model development and validation

- Source: SRC-PhysProp and PERFORCE# data
- Descriptor: DRAGON descriptors, selected by Genetic Algorithm
- Method: MLR models with OLS
- External Validation by *a priori* splitting of data (Random by activity and Self Organizing Map)
- Applicability Domain by leverage approach
- Prediction of properties for unavailable PFCs data

*CADASTER: *CAse studies on the Development and Application of in Silico Techniques for Environmental hazard and Risk asessment, 2009-2012.*

#*PERFORCE: PERFluorinated ORganic Chemicals in the European environment, EU - FP6 project, 2004-2006.*

Melting point: Data splitting

Perfluorinated chemicals (PERFORCE)

17 compounds
Prediction II

Melting point (94+17)

Variables	Train	Set	Q^2loo	R^2	Q^2boot	RMSE train	RMSE ext	Q^2ext-F1	Q^2ext-F3*	R^2ext	R^2Yscr
AAC F02[C-F] C-013	53	Prediction I **SOM** 41 test	73.35	77.11	71.89	40.86	46.65	78.84	70.16	75.30	5.18
		Prediction II 17 test			71.90		25.04	77.05	91.40	65.21	5.16
	48	Prediction I **Response** 46 test	79.30	82.85	77.48	38.07	48.52	70.70	72.16	71.77	5.84
		Prediction II 17 test			77.36		24.60	88.31	92.84	65.95	6.59
	Total 111		76.82	78.45	76.60	40.36	41.86 (cv)	-	-	-	2.82

*Consonni, V., et al. *J. Chem. Inf. Model., 49, 1669-1678.*

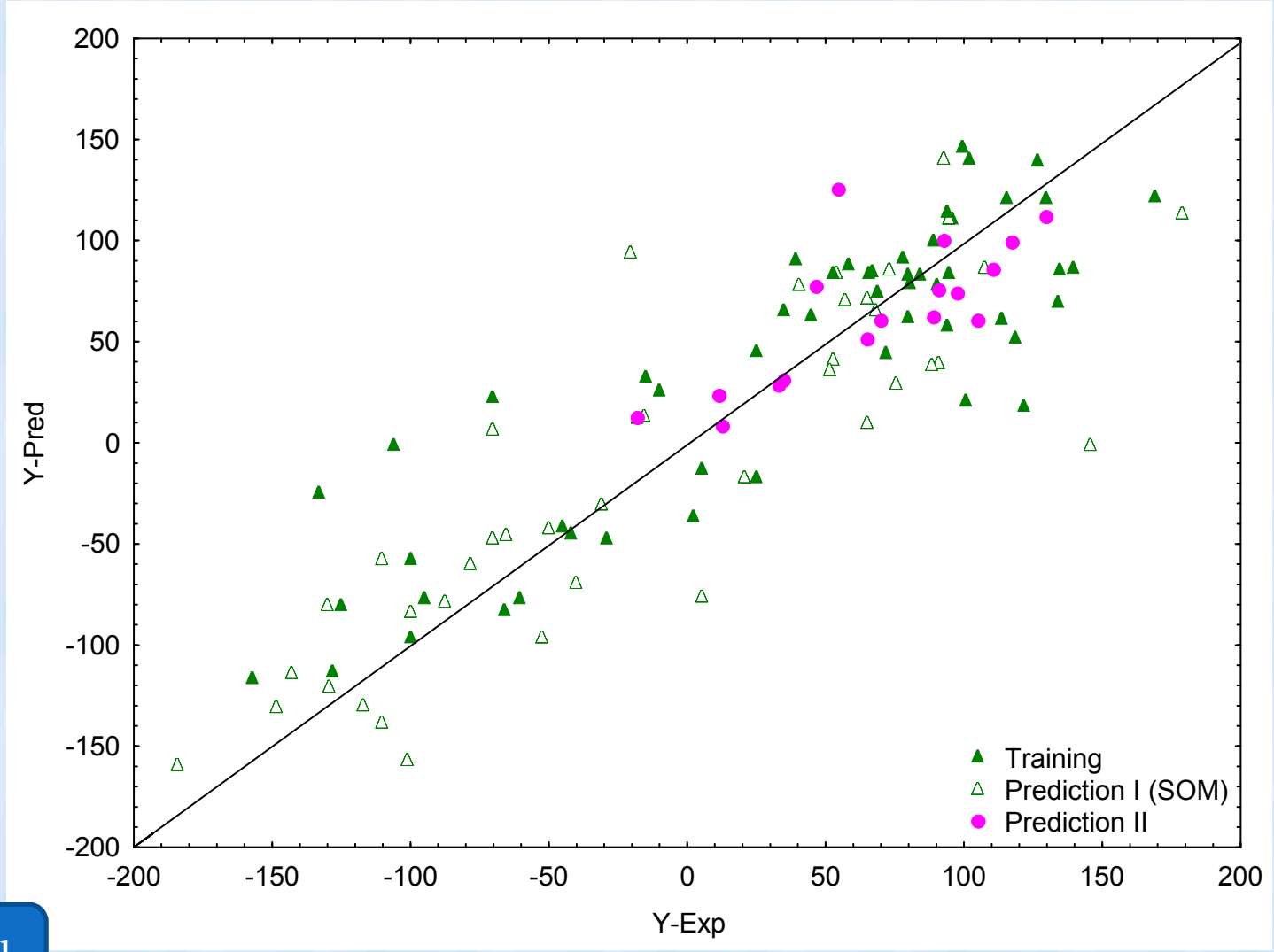
AAC = mean information index on atomic correlations, information indices

F02[C-F] = frequency of C-F at topological distance 02, 2D frequency fingerprint

C-013 = corresponds to CRX3 (X =electronegative atom), atom-centered fragments

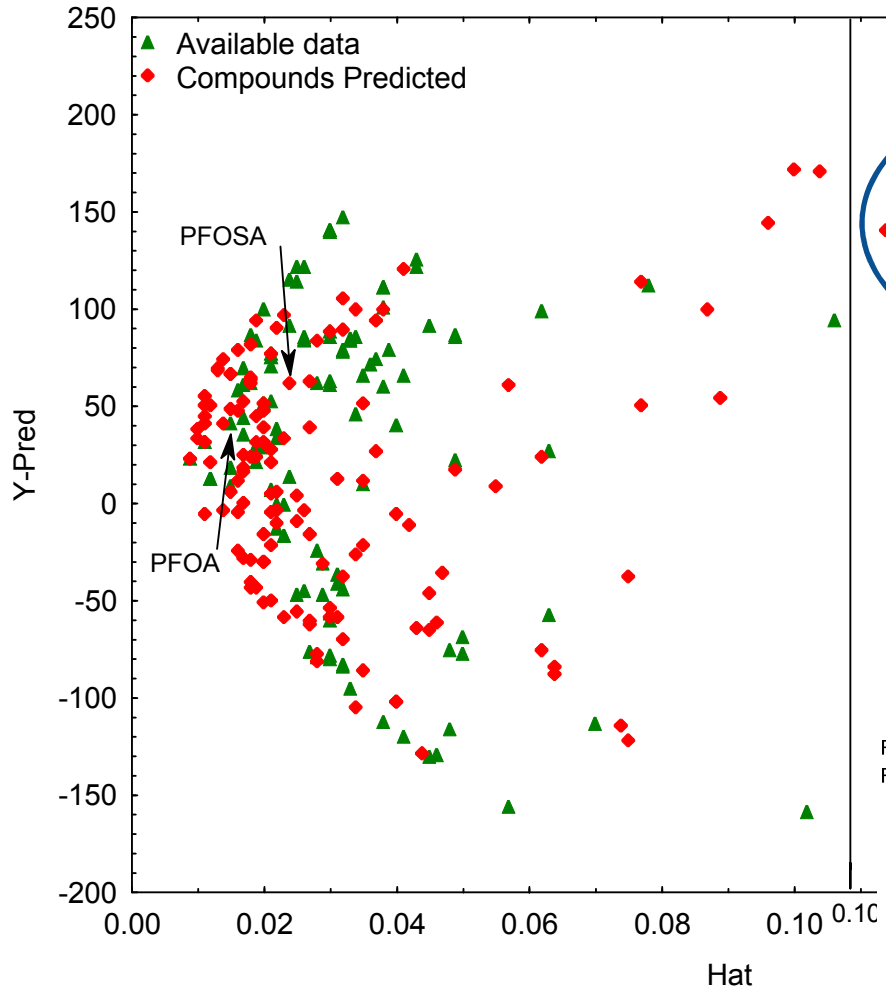
Regression plots (MP)

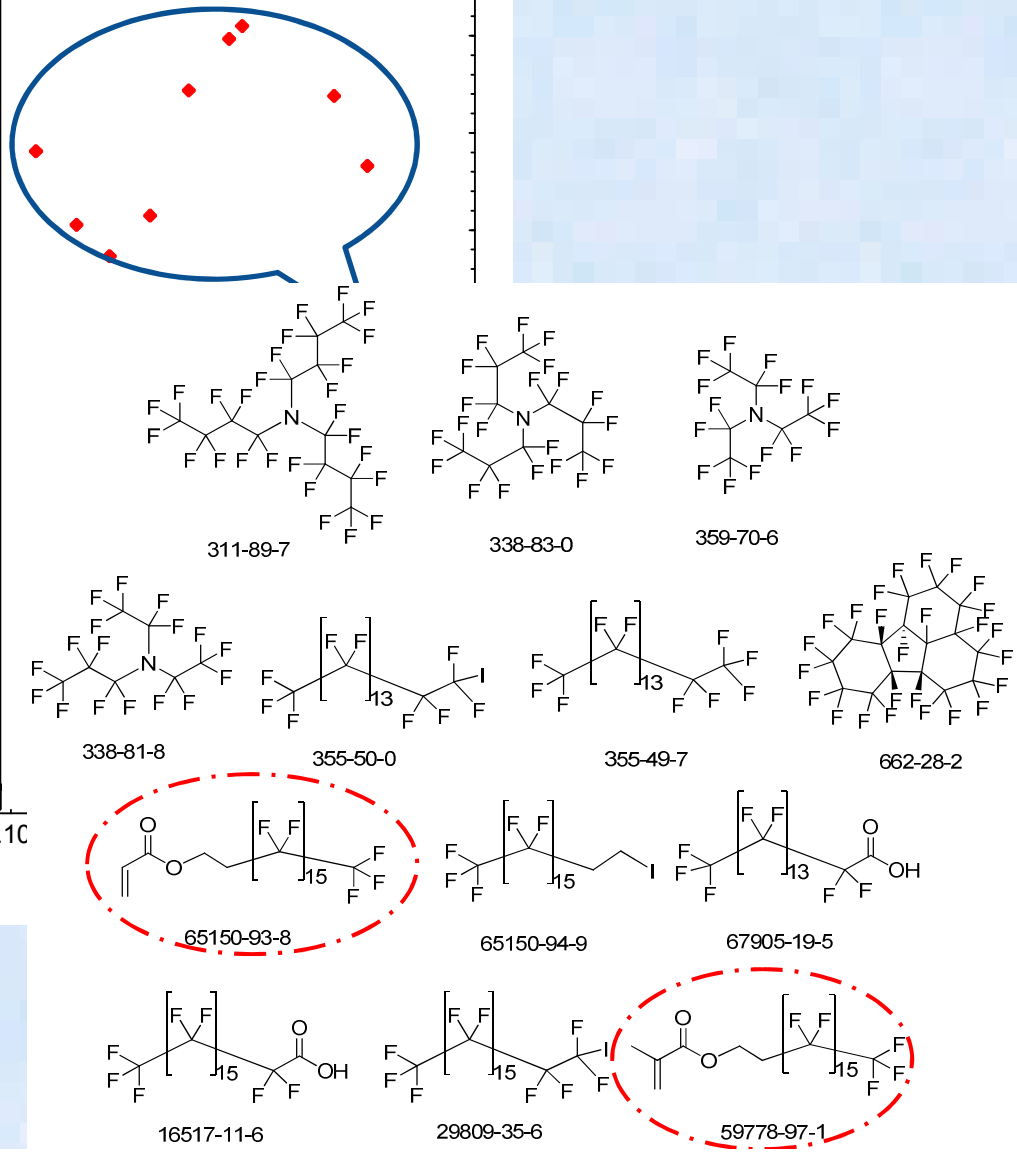
MP = 148.81 (±18.43) AAC + 4.03 (±0.66) F02[C-F] - 14.47 (±6.88) C-013 - 269.25
n=111



Full model

Applicability domain (MP)





Boiling point: Data splitting

Perfluorinated chemicals (PERFORCE)

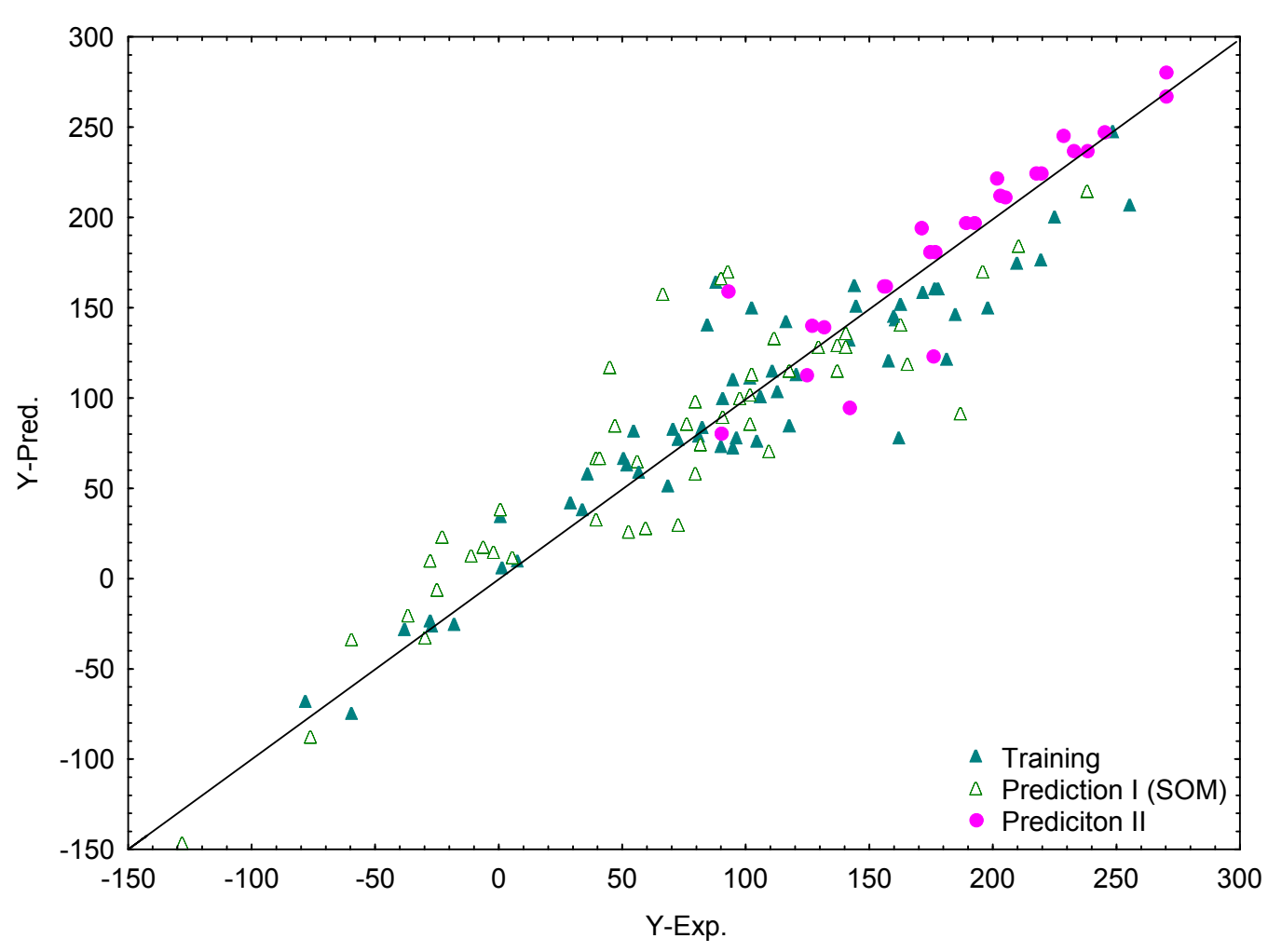
25 compounds Prediction II

Boiling point (105+25)

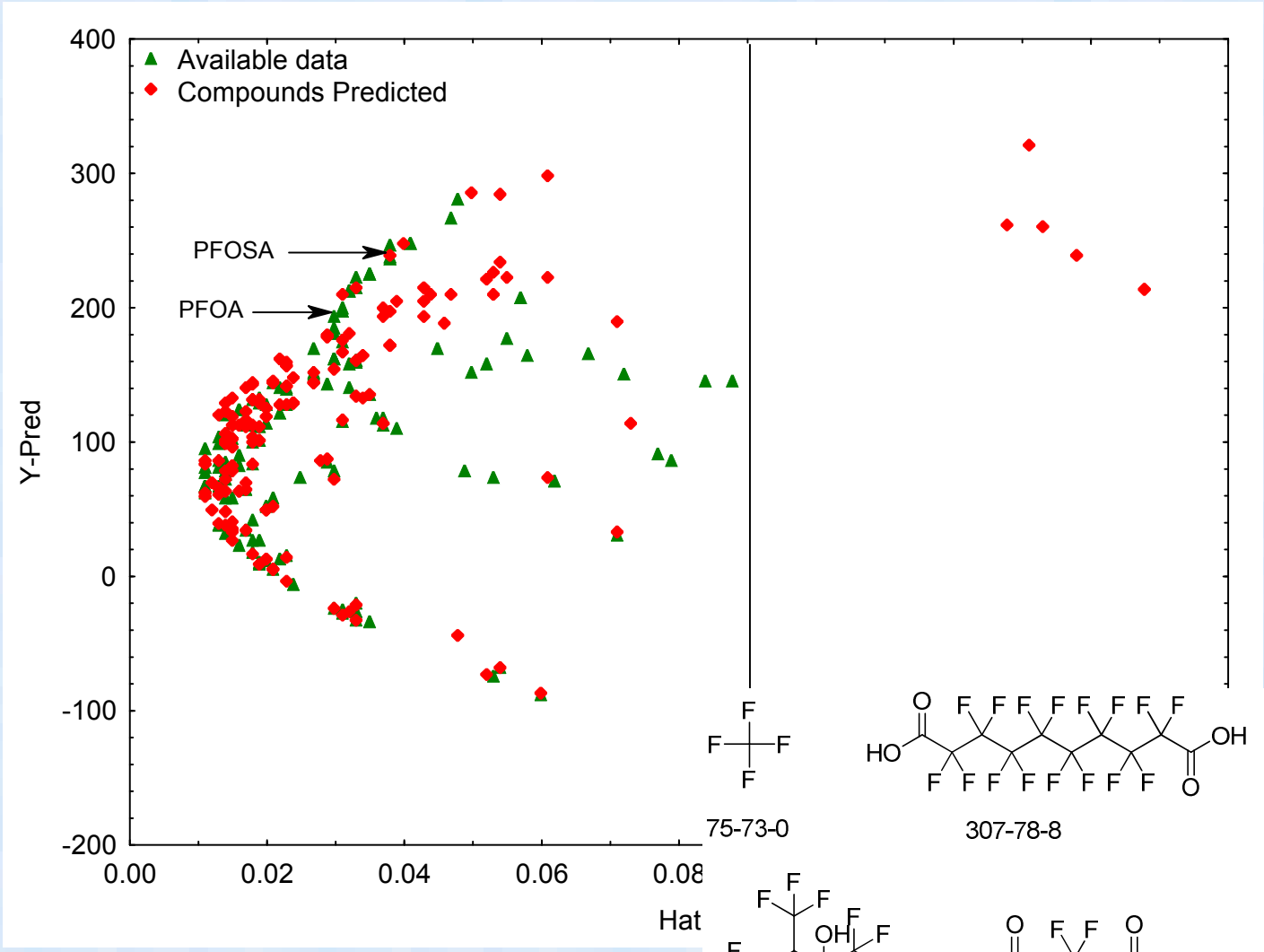
Variables	Train	Set	Q^2loo	R^2	Q^2boot	RMSE train	RMSE ext	Q^2ext -F1	Q^2ext -F3	R^2ext	R^2Yscr
Ms ATS1m nROH	55	Prediction I **SOM** 50 test	85.25	87.50	83.16	24.78	34.54	84.92	75.71	84.38	5.73
		Prediction II 25 test			86.26		29.14	91.36	85.17	83.03	5.55
	53	Prediction I **Response** 52 test	83.55	86.40	81.38	30.23	28.98	84.74	87.50	84.91	6.12
		Prediction II 25 test			80.78		26.20	94.34	89.53	83.49	5.35
	Total 130		87.54	88.54	87.37	28.21	29.42 (cv)	-	-	-	2.41

Ms = mean electro-topological state, constitutional descriptor
ATS1m = Autocorrelation of a topological structure, 2D autocorrelations
nROH = number of OH groups, functional group counts

Regression plots (BP) 130

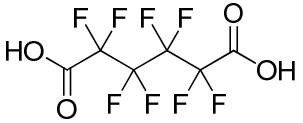


Applicability domain (BP)

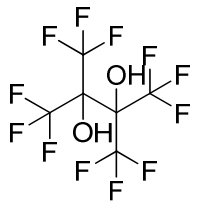


75-73-0

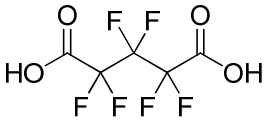
307-78-8



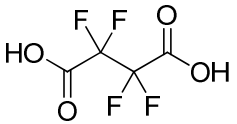
336-08-3



918-21-8



376-73-8



377-38-8

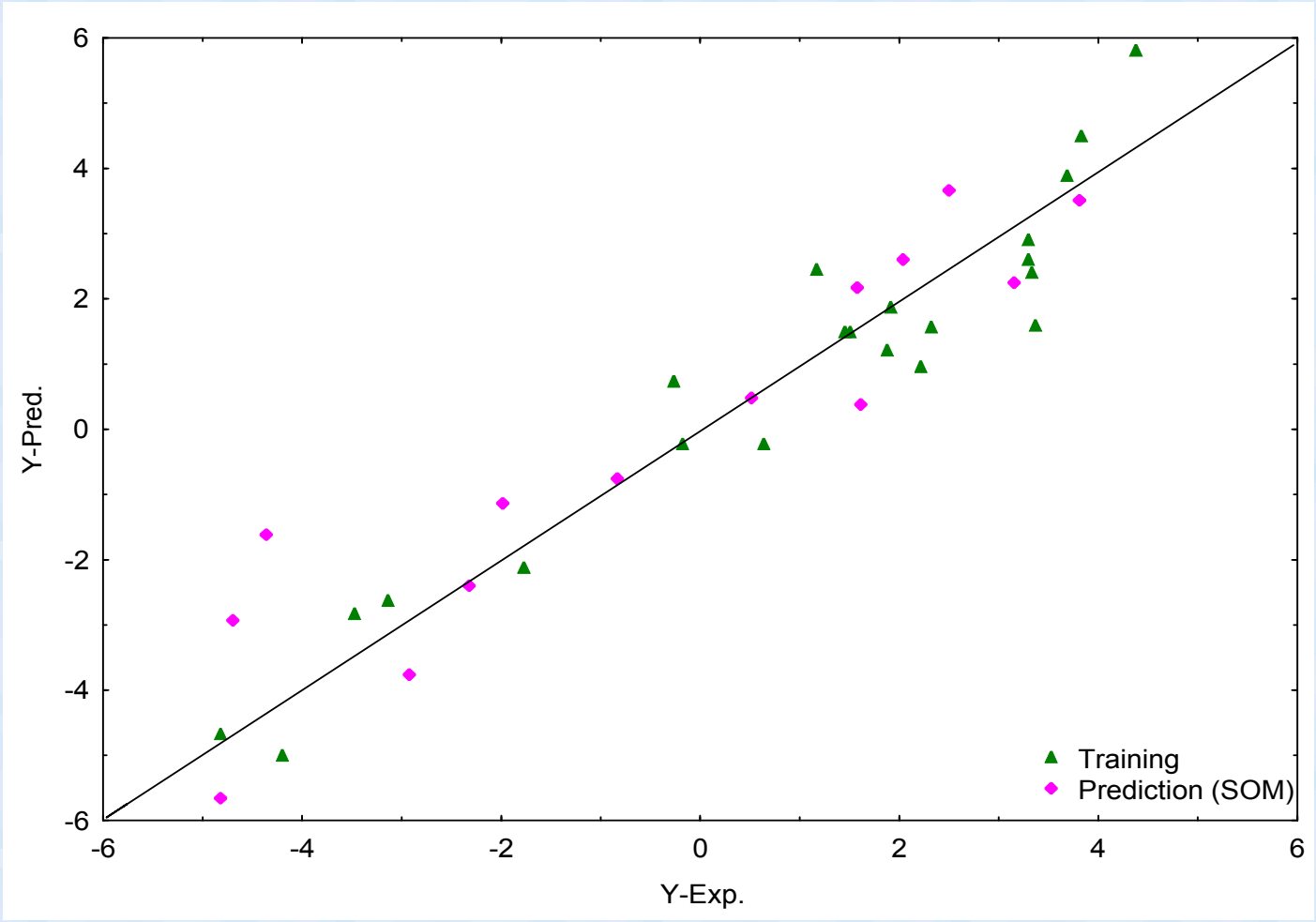
Vapor Pressure: Data splitting

Vapor Pressure (36)

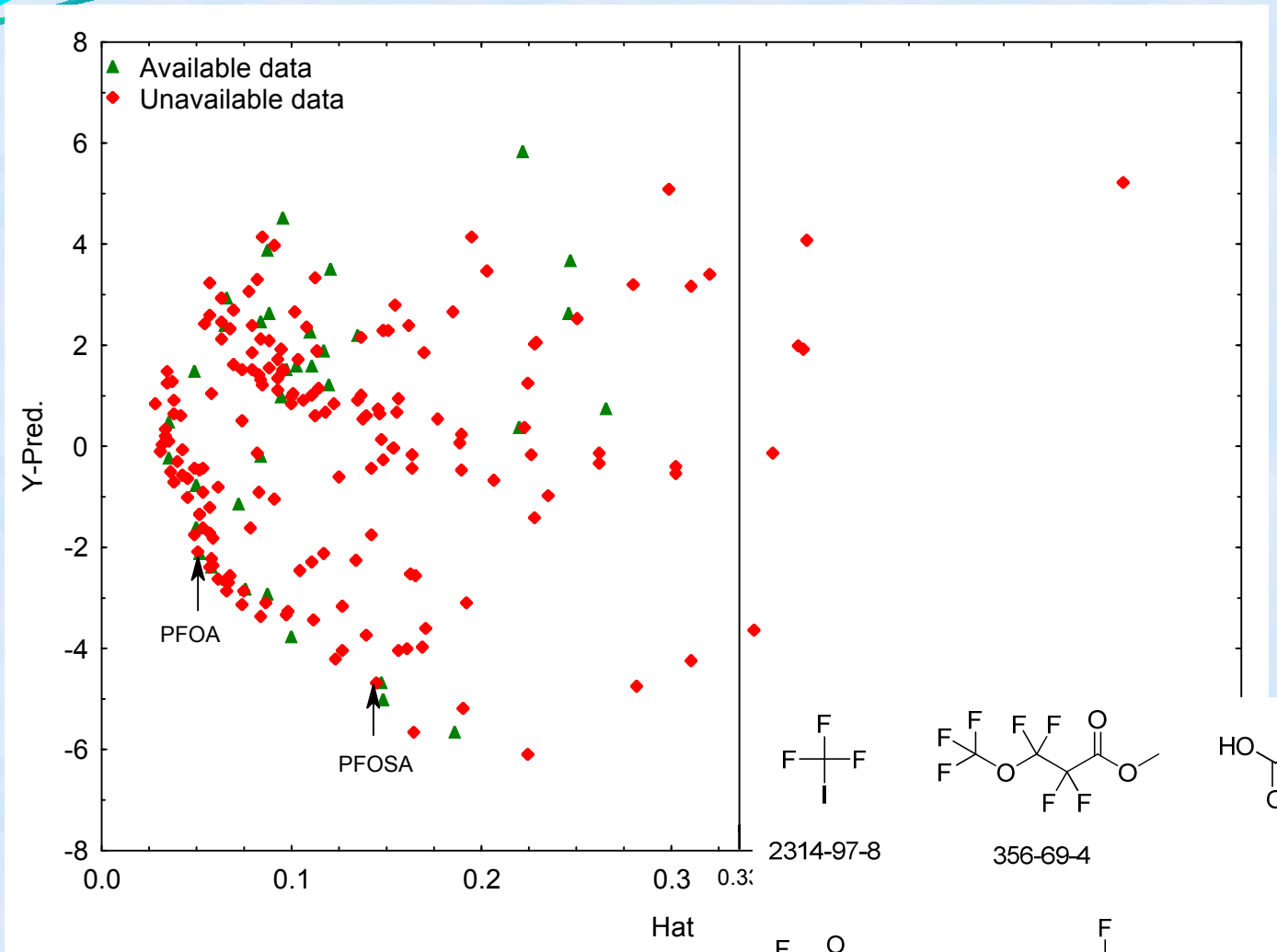
Variables	Set	Q^2loo	R^2	Q^2boot	RMSE train	RMSE ext	Q^2ext-F1	Q^2ext-F3	R^2ext	R^2Yscr
CIC0 MATS1v TPSA(Tot)	Prediction I **SOM** 14 test	91.57	94.61	80.30	0.63	1.25	85.27	78.94	83.24	14.73
	Prediction I **Response** 12 test	87.72	91.08	82.65	0.87	0.80	91.91	92.39	93.24	13.29
	Total 36	89.36	91.61	87.88	0.84	0.94 (cv)	-	-	-	8.58

Regression plots (VP)

log VP = – 0.11 (±0.01) TPSA(Tot) – 1.18 (±0.32) CIC0 – 6.35 (±3.41) MATS1v + 4.93
N=36

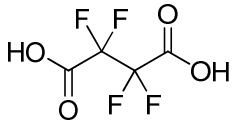


Applicability domain (VP)

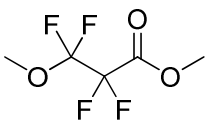


2314-97-8

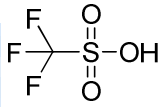
356-69-4



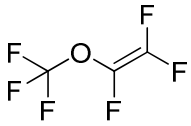
377-38-8



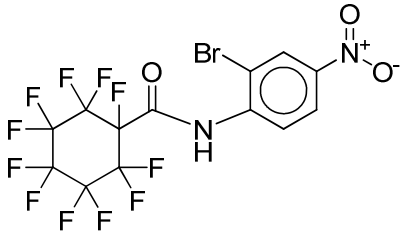
755-73-7



1493-13-6



1187-93-5



107349-85-9

Summay of QSPR on PFCs:

End point	Variables	Total	Q^2loo	R^2	Q^2boot	RMSE train	RMSE cv	(n) AD%	RMSE EPI*
Melting Point	AAC F02[C-F] C-013	111	76.82	78.45	76.10	40.36	41.86	(248) 94.7	46.678
Boiling Point	Ms ATS1m nROH	130	87.55	88.54	87.27	27.56	29.12	(290) 97.9	43.046
Vapor Pressure	CIC0 MATS1v TPSA(Tot)	36	89.36	91.61	87.88	0.84	0.94	(243) 97.1	1.03

* http://www.epa.gov/oppt/exposure/pubs/episuite.htm

Presentation of CADASTER results on PFC:

B. Bhhatarai, P. Gramatica, "**QSAR studies on Mouse Inhalation LC50 data of perfluoro- compounds**". (Poster)
ICCE 2009 - 12th EuCheMS International Conference on Chemistry & the Environment, Stockholm, Sweden, 14-17 June 2009.

B. Bhhatarai, P. Gramatica, M. Luini, E. Papa, "**QSAR modelling of toxicity endpoints of emerging pollutants: Fragrances and Perfluorinated compounds**". (Oral)
CMTPI-2009 - Fifth International Symposium on Computational Methods in Toxicology and Pharmacology Integrating Internet Resources, Istanbul, Turkey, 4-8 July 2009.

B. Bhhatarai, P. Gramatica, "**QSPR Studies for Predicting Physico-chemical Properties of Perfluorinated Compounds**". (Oral)
Conferentia Chemometrica, Siófok, Hungary, September 27-30, 2009.

B. Bhhatarai, P. Gramatica, "**Perfluoro Toxicity (LC_{50} inhalation) in *Rattus* and *Mus* species using QSAR**". (Poster)
SETAC North America 30th Annual Meeting, New Orleans, Louisiana, USA, 19-23 November 2009.

Future Work

- Validity of models on new Experimental data produced
- Experimental Design of suggested compounds
- Collaboration with CADASTER partners and other EU projects (OPENTOX)
- QSAR modeling of Acute toxicity data and other relevant SIDS end points

References:

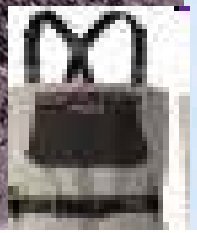
1. www.cadaster.eu
2. Todeschini, R. et al. DRAGON v.5, www.talete.it.
3. Gramatica, P. 2007, QSAR Comb. Sci. 26, 694-701.
4. Inoue K. et al., 2004, Environ. Health Perspect. 112, 1204–1207.
5. Lau, C. et al., 2007, Toxicol. Sci. 99, 366–394.
6. Renner, R., 2003, Environ. Sci. Technol. 37, 201A
7. Lau, C. et al., 2004, Toxicol. Appl. Pharmacol. 198, 231–241.
8. Lau, C. et al., 2003, Toxicol. Sci. 74, 382–392.
9. Kovarova, J. et al., 2008, Neuro Endocrinol Lett. 29, 599-608.
10. Renner, R., 2009, Environ. Sci. Technol., 43, 4219.

This study has been financially supported by the FP7th-EU Project CADASTER (n. 212668)

Acknowledgements

Dr. Ester Papa, Dr. Jiazhong Li, Simona Kovarich, Mara Luini

Thanks for your attention !!



Gore-Tex

breathable fabric