

České vysoké učení technické v Praze

Fakulta elektrotechnická

Katedra teorie obvodů



Diplomová práce

Identifikace plynů a těkavých látek pomocí chemorezistivních
senzorů

Identification of Gases and Volatile Substances Using
Chemoresistive Sensors

Autor: Bc. Daria Pecheritca

Vedoucí práce: Ing. Alexandr Laposka, Ph.D.

Studijní program: Lékařská elektronika a bioinformatika

Specializace: Lékařská technika

Praha 2024

I. OSOBNÍ A STUDIJNÍ ÚDAJE

Příjmení: **Pecheritca** Jméno: **Daria** Osobní číslo: **483883**
Fakulta/ústav: **Fakulta elektrotechnická**
Zadávající katedra/ústav: **Katedra teorie obvodů**
Studijní program: **Lékařská elektronika a bioinformatika**
Specializace: **Lékařská technika**

II. ÚDAJE K DIPLOMOVÉ PRÁCI

Název diplomové práce:

Identifikace plynů a těkavých látek pomocí chemorezistivních senzorů

Název diplomové práce anglicky:

Identification of Gases and Volatile Substances Using Chemoresistive Sensors

Pokyny pro vypracování:

1. Zpracujte rešerši o využití chemorezistivních senzorů pro identifikaci plynů a těkavých látek. Zaměřte se na různé typy senzorů, jejich principy fungování a specifika použití v medicíně.
2. Proveďte rozbor metod zpracování a vyhodnocení naměřených dat k přesnější identifikaci detekované látky.
3. S využitím vhodné vybrané matice chemorezistivních senzorů sestavte testovací pracoviště a stanovte vhodný postup měření a proveďte měření pro různé koncentrace vybraných plynů a těkavých látek.
4. Naměřená data vyhodnoťte a proveďte návrh a implementaci algoritmu pro určení detekované látky. Zahrňte například metody redukce počtu proměnných, klasifikace dat a nástroje pro zpracování vícerozměrných dat.
5. Zhodnoťte dosažené výsledky a navrhnete další postup výzkumu.

Seznam doporučené literatury:

- [1] Z. Chen, Z. Chen, Z. Song, W. Ye, and Z. Fan, "Smart gas sensor arrays powered by artificial intelligence," J. Semicond., vol. 40, no. 11, p. 111601, 2019, doi: 10.1088/1674-4926/40/11/111601.
- [2] S. Huang et al., "Machine Learning Enabled Smart Gas Sensing Platform for Identification of Industrial Gases," Adv. Intell. Syst., vol. 4, no. 4, 2022, doi: 10.1002/aisy.202200016.
- [3] S. Feng et al., "Review on smart gas sensing technology," Sensors, vol. 19, no. 17, p. 3760, 2019.

Jméno a pracoviště vedoucí(ho) diplomové práce:

Ing. Alexandr Laposa, Ph.D. katedra mikroelektroniky FEL

Jméno a pracoviště druhé(ho) vedoucí(ho) nebo konzultanta(ky) diplomové práce:

Datum zadání diplomové práce: **04.02.2024** Termín odevzdání diplomové práce: **24.05.2024**

Platnost zadání diplomové práce: **21.09.2025**

Ing. Alexandr Laposa, Ph.D.
podpis vedoucí(ho) práce

doc. Ing. Radoslav Bortel, Ph.D.
podpis vedoucí(ho) ústavu/katedry

prof. Mgr. Petr Páta, Ph.D.
podpis děkana(ky)

III. PŘEVZETÍ ZADÁNÍ

Diplomantka bere na vědomí, že je povinna vypracovat diplomovou práci samostatně, bez cizí pomoci, s výjimkou poskytnutých konzultací. Seznam použité literatury, jiných pramenů a jmen konzultantů je třeba uvést v diplomové práci.

Datum převzetí zadání

Podpis studentky

Prohlášení

Prohlašuji, že jsem předloženou práci vypracovala samostatně a že jsem uvedla veškeré použité informační zdroje v souladu s Metodickým pokynem o dodržování etických principů při přípravě vysokoškolských závěrečných prací.

V Praze dne 20.5.2024

.....

Poděkování

Ráda bych poděkovala vedoucímu diplomové práce Ing. Alexandru Laposovi, Ph.D. za odborné vedení mé diplomové práce, jeho vstřícný přístup a mnoho užitečných rad.

Také bych ráda poděkovala Ing. Jiřímu Kroutilovi, Ph.D. za pomoc při sestavování testovacího pracoviště.

V neposlední řadě bych ráda poděkovala své rodině a přátelům za podporu, kterou mi poskytli během studia.

Práce byla realizována za podpory projektu základního výzkumu Grantové agentury České republiky 22-04533S "Tištěná pole vysoce citlivých a selektivních heterogenních senzorů plynu"

Abstrakt

Daná diplomová práce se zabývá problematikou chemorezistivních senzorů plynu, jejich křížovou citlivostí a metodikou testování těkavých látek. Bylo navrženo a realizováno testovací pracoviště, s jehož pomocí byly naměřeny zvolené těkavé látky. Vybrané látky jsou potenciálními biomarkery nemocí a byly naměřeny ve třech koncentracích pomocí devíti tlustovrstvých chemorezistivních senzorů plynu. Následně s využitím metod strojového učení byla provedena vizualizace a klasifikace vybraných těkavých látek.

Klíčová slova: chemorezistivní senzory plynu, elektronický nos, VOC, klasifikace, křížová validace.

Abstract

This diploma thesis deals with the issue of chemoresistive gas sensors, their cross-sensitivity, and the methodology of testing volatile substances. A test bed was designed and implemented, using which the selected substances were measured. Chosen volatile substances, which are potential biomarkers of disease, were measured at three concentrations using nine thick film chemoresistive gas sensors. Subsequently, visualization and classification of the selected volatiles was performed using machine learning methods.

Keywords: chemoresistive gas sensors, electronic nose, VOC, classification, cross-validation.

Seznam použitých zkratek

VOC	Těkavá organická látka	Volatile Organic Compound
CHOPN	Chronická obstrukční plicní nemoc	Chronic Obstructive Pulmonary Disease
TBC	Tuberkulóza	Tuberculosis
CKD	Chronické onemocnění ledvin	Chronic Kidney Disease
MOX	Oxid kovu	Metal Oxide
PCA	Analýza hlavních komponent	Principal Component Analysis
GMM		Gaussian Mixture Model
SVM		Support Vector Machine
k-NN	k-nejbližších sousedů	k-Nearest Neighbors
ANOVA	Analýza rozptylu	Analysis of Variance
BSS		Between Sum of Squares
BMS		Between Mean Square
WSS		Within Sum of Squares
WNS		Within Mean Square
t-SNE		t-distributed Stochastic Neighbor Embedding
RF		Random Forest
LDA	Lineární diskriminační analýza	Linear Discriminant Analysis
TP	Skutečně pozitivní	True Positive
TN	Skutečně negativní	True Negative
FP	Falešně pozitivní	False Positive
FN	Falešně negativní	False Negative
CV	Křížová validace	Cross-Validation
LOOCV		Leave-One-Out Cross-Validation
LPG	Zkapalněný ropný plyn	Liquefied Petroleum Gas
IPA	Isopropylalkohol	
POC		Point of Care
UV	Ultrafialové	Ultraviolet
NPLC	Počet cyklů napětí sítě	Number of Power Line Cycle
FFT	Rychlá Fourierova transformace	Fast Fourier Transform
DWT	Diskrétní vlnková transformace	Discrete Wavelet Transform
ANN	Umělá neuronová síť	Artificial Neural Network

Seznam obrázků

Obrázek 1: Snímání signálů a zpracování dat ze senzoru plynu. Převzato z [1].	4
Obrázek 2: Blokové schéma systému elektronického nosu (viz 1.5). Převzato z [1].	4
Obrázek 3: Příklad zapojení tlustovrstvého MOX senzoru. Převzato z [11].	7
Obrázek 4: Doba odezvy senzoru v závislosti na koncentraci analytu, kde R_0 [Ω] je odpor senzoru na začátku měření, R_S [Ω] je odpor v ustáleném stavu, R_f [Ω] je 10% odezvy senzoru na konci doby snímání, τ_r [s] je doba odezvy, τ_f [s] je doba zotavení (v tomto případě se uvažuje 10 % až 90 % R_S). Upraveno z [14].	10
Obrázek 5: a) profil průběhu odezvy senzoru plynu pro cyklické měření, b) schéma profilu průběhu odezvy $S(t)$ pro měření se skládajícího ze stabilizačního procesu (t_0-t_1) pod proudem čistého N_2 k nalezení základní úrovně, fáze expozice (t_1-t_2) a fáze proplachování (t_2-t_3). Převzato z [9].	13
Obrázek 6: Technologie elektronického nosu (Smart Gas Sensing). Převzato z [14]	14
Obrázek 7: Blokové schéma udávající postup pro analýzu dat ze senzoru plynu pomocí algoritmů strojového učení. Převzato z [1].	15
Obrázek 8: Rozdělení metod strojového učení pro analýzu dat ze senzoru plynu. Převzato z [4].	15
Obrázek 9: Hlavní kroky pro počítání PCA. Převzato z [21].	21
Obrázek 10: Princip klasifikace algoritmu k-NN. Převzato z [1].	23
Obrázek 11: Princip SVM algoritmu: a) různé možné hranice pro rozdělení tříd, b) optimální hranice s maximálním tzv. marginem. Převzato z [1].	24
Obrázek 12: Základní koncept klasifikátoru RF. Převzato z [30].	25
Obrázek 13: Matice záměn (příklad pro dvě třídy).	27
Obrázek 14: Na obrázku vlevo je ukázka nedostatečného přizpůsobení modelu (tzv. underfitting), optimálního přizpůsobení modelu (uprostřed) a vpravo je ukázka nadměrného přizpůsobení modelu (tzv. overfitting), což může mít za následek špatnou predikci na budoucích, dosud neznámých datech. Převzato z [33].	29
Obrázek 15: Princip k-násobné křížové validace. Převzato z [17].	30
Obrázek 16: Princip nested křížové validace. Převzato z [33]	31
Obrázek 17: Fotografie celé sady použitých senzorů (a) pohled shora, (b) pohled zdola.	32
Obrázek 18: Popis součástí senzoru plynu MQ3. Červenou barvou jsou značené součástky integrovaného obvodu MQ senzoru, modrou barvou jsou značené piny integrovaného obvodu LM393 IC (v této práci se nevyužívají), žlutou barvou jsou značené elektrody MQ senzoru.	33
Obrázek 19: Senzor plynu MQ3, (a) celkový pohled, (b) detail vlastního snímacího prvku s měřicími elektrodami (odpor měřen mezi piny 1, 3 a piny 4,6) a výhřívací spirálou (piny 2 a 5), material citlivé vrstvy SnO_2	34
Obrázek 20: Schéma zapojení chemorezistivního senzoru plynu MQ3.	34
Obrázek 21: Graf závislosti relativní změny hodnot odporů citlivé vrstvy pro různé teploty okolí a dvě urovně relativní vlhkosti (33 % a 85 %) pro senzor MQ3. Ro je odpor citlivé vrstvy při teplotě 20 °C a detekované látky ethanol s koncentrací	

400 ppb (0,4 mg/l). Rs je odpor při teplotě a vlhkosti dle grafu, detekovaná látka je ethanol s koncentrací 400 ppb. Upraveno z [34].	35
Obrázek 22: Zapojení matice senzorů plynu s použitím nepájivého pole.	36
Obrázek 23: Zapojení matice senzorů plynu s použitím nepájivého pole a testovací komory o objemu 1,65 l.	36
Obrázek 24: První varianta měřicího pracoviště s použitím testovací komory o objemu 1,65 l.	37
Obrázek 25: Průběhy odezvy jednotlivých senzorů plynu při měření ethanolu s použitím první varianty zapojení senzorů a testovací komory objemem 1,65 l.	39
Obrázek 26: Relativní změny hodnot napětí senzorů plynu při měření ethanolu s použitím první varianty zapojení senzorů s označením jednotlivých fází měření.	40
Obrázek 27: Zapojení matice senzorů plynu s použitím plošného spoje.	41
Obrázek 28: Zapojení matice senzorů plynu s použitím plošného spoje a testovací komory o objemu 4,06 l.	42
Obrázek 29: Druhá varianta měřicího pracoviště s použitím testovací komory o objemu 4,06 l.	43
Obrázek 30: Nastavení Keithley Kickstart verze 2.11.1 pro snímané odezvy MQx senzorů.	46
Obrázek 31: Nastavení vzorkování SW Keithley Kickstart verze 2.11.1 pro snímané odezvy MQx senzorů.	46
Obrázek 32: Průběhy odezvy jednotlivých senzorů plynu při měření ethanolu s koncentrací 1 ppm s použitím druhé varianty zapojení senzorů a testovací komory objemem 4,06 l.	47
Obrázek 33: Průběhy odezvy jednotlivých senzorů plynu při měření ethanolu s koncentrací 5 ppm s použitím druhé varianty zapojení senzorů a testovací komory objemem 4,06 l.	47
Obrázek 34: Průběhy odezvy jednotlivých senzorů plynu při měření ethanolu s koncentrací 10 ppm s použitím druhé varianty zapojení senzorů a testovací komory objemem 4,06 l.	48
Obrázek 35: Relativní změny hodnot napětí senzorů plynu při měření ethanolu s koncentrací 1 ppm s použitím druhé varianty zapojení senzorů s označením jednotlivých fází měření.	49
Obrázek 36: Relativní změny hodnot napětí senzorů plynu při měření ethanolu s koncentrací 5 ppm s použitím druhé varianty zapojení senzorů s označením jednotlivých fází měření.	49
Obrázek 37: Relativní změny hodnot napětí senzorů plynu při měření ethanolu s koncentrací 10 ppm s použitím druhé varianty zapojení senzorů s označením jednotlivých fází měření.	50
Obrázek 38: Průměrný časový průběh odezvy senzoru MQ2 se standardní směrodatnou odchylkou během čtyř měření ethanolu s koncentrací 1 ppm.	51
Obrázek 39: Průměrný časový průběh odezvy senzoru MQ2 se standardní směrodatnou odchylkou během čtyř měření ethanolu s koncentrací 5 ppm.	51
Obrázek 40: Průměrný časový průběh odezvy senzoru MQ2 se standardní směrodatnou odchylkou během čtyř měření ethanolu s koncentrací 10 ppm.	52

Obrázek 41: Postup zpracování dat naměřených pomocí matice MQ senzorů plynu pro identifikace VOC.	53
Obrázek 42: Korelační matice pro extrahované parametry z časové oblasti s využitím Spearmanova korelačního koeficientu.	59
Obrázek 43: Vizualizace acetonu, ethanolu a toluenu pomocí PCA.	63
Obrázek 44: Vizualizace acetonu, ethanolu a toluenu pomocí t-SNE, hodnota hyperparametru „perplexita“ je 35.	64
Obrázek 45: Krabicové grafy pro menší soubor dat, kde A1: parametr citlivost pro aceton, A2: parametr citlivost pro ethanol, A3: parametr citlivost pro toluen, B1: parametr energie signálu pro aceton, B2: parametr energie signálu pro ethanol, B3: parametr energie signálu pro toluen, C1: parametr nejrychlejší doba náběhu napětí pro aceton, C2: parametr nejrychlejší doba náběhu napětí pro ethanol, C1: parametr nejrychlejší doba náběhu napětí pro toluen.	66
Obrázek 46: Krabicové grafy pro větší soubor dat, kde A1: parametr citlivost pro aceton, A2: parametr citlivost pro ethanol, A3: parametr citlivost pro toluen, B1: parametr energie signálu pro aceton, B2: parametr energie signálu pro ethanol, B3: parametr energie signálu pro toluen, C1: parametr nejrychlejší doba náběhu napětí pro aceton, C2: parametr nejrychlejší doba náběhu napětí pro ethanol, C1: parametr nejrychlejší doba náběhu napětí pro toluen.	67
Obrázek 47: Matice záměn pro výsledky klasifikace pomocí klasifikátoru SVM s RBF kernelem na menším souboru dat, kde třída 1 je aceton, třída 2 je ethanol, třída 3 je toluen.	69
Obrázek 48: Matice záměn pro výsledky klasifikace pomocí klasifikátoru k-NN na menším souboru dat, kde třída 1 je aceton, třída 2 je ethanol, třída 3 je toluen.	70
Obrázek 49: Matice záměn pro výsledky klasifikace pomocí klasifikátoru RF na menším souboru dat, kde třída 1 je aceton, třída 2 je ethanol, třída 3 je toluen.	70
Obrázek 50: Matice záměn pro výsledky klasifikace pomocí klasifikátoru k-NN na větším souboru dat, kde třída 1 je aceton, třída 2 je ethanol, třída 3 je toluen.	71
Obrázek 51: Matice záměn pro výsledky klasifikace pomocí klasifikátoru RF na větším souboru dat, kde třída 1 je aceton, třída 2 je ethanol, třída 3 je toluen.	71
Obrázek 52: Matice záměn pro výsledky klasifikace pomocí klasifikátoru SVM s RBF kernelem na větším souboru dat, kde třída 1 je aceton, třída 2 je ethanol, třída 3 je toluen.	72

Seznam tabulek

Tabulka 1: Příklady endogenních látek a možných onemocnění	3
Tabulka 2: Typické parametry časové odezvy senzoru.	11
Tabulka 3: Značení výstupních pinů modulu MQ senzoru.	33
Tabulka 4: Vlastností MQ senzorů plynu dle dostupných katalogových listů.	35
Tabulka 5: Zvolené VOC pro první variantu experimentu.	38
Tabulka 6: Požadovaný objem VOC pro cílovou koncentraci 5 ppm v 1,65 l vzduchu.	38
Tabulka 7: Jednotlivé fáze měření VOC.	39
Tabulka 8: Požadovaný objem zvolených VOC a plynu pro cílovou koncentraci 1 ppm v 4,06 l vzduchu.	44
Tabulka 9: Požadovaný objem zvolených VOC a plynu pro cílovou koncentraci 5 ppm v 4,06 l vzduchu.	44
Tabulka 10: Požadovaný objem zvolených VOC a plynu pro cílovou koncentraci 10 ppm v 4,06 litrech vzduchu.	44
Tabulka 11: Tenze par vybraných VOC.	45
Tabulka 12: Jednotlivé fáze měření VOC (druhá varianta experimentu).	45
Tabulka 13: Výsledky Kolmogorov-Smirnov testu normality.	59
Tabulka 14: Silně korelované páry parametrů.	60
Tabulka 15: Vybrané parametry pomocí jednosměrné ANOVA.	62
Tabulka 16: Hyperparametry použitých klasifikačních algoritmů.	68
Tabulka 17: Odhad výkonu klasifikačních algoritmů použitých pro klasifikace VOC na menším datovém souboru.	69
Tabulka 18: Odhad výkonu klasifikačních algoritmů použitých pro klasifikaci VOC na větším datovém souboru.	69

Obsah

1.	Úvod.....	1
2.	Chemorezistivní senzory plynu	2
2.1.	Využití senzorů plynu.....	2
2.2.	Princip činnosti	3
2.3.	Typy chemorezistivních senzorů	6
2.4.	Charakteristické parametry	8
2.5.	Průběh odezvy senzoru.....	11
2.6.	Problém selektivity	14
3.	Algoritmy strojového učení.....	15
3.1.	Předzpracování dat.....	16
3.2.	Selekce parametrů	17
3.2.1.	Test korelace.....	17
3.2.2.	ANOVA F-test	18
3.3.	Vizualizace multidimenzionálního prostoru	20
3.3.1.	PCA	20
3.3.2.	t-SNE.....	22
3.4.	Klasifikace	23
3.4.1.	k-NN.....	23
3.4.2.	SVM	24
3.4.3.	Random Forest	25
3.4.4.	LDA.....	26
3.5.	Odhad výkonu modelu	27
3.5.1.	Matice záměn	27
3.5.2.	Křížová validace.....	28
4.	Testování VOC	32
4.1.	MQ senzory plynu.....	32
4.2.	Metodika experimentu	36
4.2.1.	První varianta experimentu	36
4.2.2.	Hodnocení první varianty experimentu.....	40
4.2.3.	Druhá varianta experimentu	41

4.2.4.	Hodnocení druhé varianty experimentu	52
5.	Zpracování dat.....	53
5.1.	Extrakce parametrů	55
5.2.	Selekce parametrů	58
5.2.1.	Test korelace.....	58
5.2.2.	Jednosměrná ANOVA.....	60
5.3.	Vizualizace.....	63
5.3.1.	PCA	63
5.3.2.	t-SNE.....	64
5.4.	Klasifikace	65
5.4.1.	Volba CV.....	65
5.4.2.	Porovnání modelů.....	66
5.5.	Hodnocení klasifikace.....	73
5.6.	Diskuse	74
6.	Závěr.....	76
	Použitá literatura.....	78
	Přílohy	I
	Příloha A - Průběhy časových odezev senzorů s použitím 1. varianty zapojení	
	senzorů plynu:.....	II
	Příloha B - Průběhy časových odezev senzorů s použitím 2. varianty zapojení	
	senzorů plynu	IX
	Příloha C - Výsledky selekce parametrů a vizualizace	XLVI
	Příloha D - Testovací komora, varianta 2	LI
	Seznam a popis příložených souborů	LIV

1. Úvod

Tato diplomová práce se zabývá identifikací těkavých látek pomocí chemorezistivních senzorů plynu. V současné době jsou chemorezistivní senzory plynu běžně používané zejména pro kontrolu kvality ovzduší, potravin a detekci nebezpečných látek. Příkladem je chemorezistivní senzor oxidu uhelnatého (CO), který detekuje nebezpečnou hladinu plynu v domácnostech a na pracovištích. Zároveň se v poslední době objevuje více vědeckých studií na téma možnosti využití chemorezistivních senzorů plynu v medicínských aplikacích. Především pro detekci biomarkerů v dechu člověka, které jsou indikátory onemocnění. Detekce biomarkerů z dechu může být přínosná pro včasnou diagnostiku nemocí, což zvýší šance uzdravení pacientů a zároveň se sníží zátěž zdravotnického systému. Hlavní výhodou chemorezistivních senzorů plynu jsou nízké pořizovací náklady. Hlavní nevýhodou je pak křížová citlivost senzorů plynu (popsáno v kapitole č. 2).

K současným částečným řešením křížové citlivosti patří modifikace povrchu snímací vrstvy chemorezistivních senzorů plynu a použití metod strojového učení pro následné zpracování naměřených dat a identifikaci cílových těkavých látek a plynů.

Prvním cílem této diplomové práce bylo provést rešerši tématu využití chemorezistivních senzorů pro identifikaci plynů a těkavých látek. Výsledky této rešerše jsou zohledněny v kapitole č. 2.

Druhým úkolem této diplomové práce bylo provést rozbor metod zpracování a vyhodnocení naměřených dat k přesnější identifikaci detekované látky. Tento rozbor je zpracován v kapitole č. 3.

Dalším úkolem bylo vybrat vhodnou matici chemorezistivních senzorů, kde hlavním požadavkem byla nízká cena a široká dostupnost senzorů, aby byla umožněná opakovatelnost měření na jiných pracovištích, jinými výzkumnými skupinami. Dalším bodem bylo navrhnout a realizovat testovací pracoviště, s jehož pomocí bylo následně provedeno měření vybraných těkavých látek ve třech různých koncentracích (1 ppm, 5 ppm, 10 ppm). Výsledky realizace testovacího pracoviště a testování látek jsou uvedené v kapitole č. 4.

Čtvrtým cílem této práce bylo zpracovat naměřená data s využitím vybraných metod strojového učení k dosažení přesnější identifikace naměřených látek. Výsledek zpracování naměřených dat lze najít v kapitole č. 5.

Poslední kapitola č. 6 diplomové práce je věnována vyhodnocení výsledků a návržení dalšího postupu výzkumu.

2. Chemorezistivní senzory plynu

2.1. Využití senzorů plynu

Senzory plynu se používají pro monitorování životního prostředí. Pomocí tzv. systému elektronického nosu lze detekovat nebezpečné těkavé organické látky (VOC) a plyny (například metan) za účelem snížení negativního dopadu na životní prostředí. V důsledku toho se monitorování životního prostředí rozšířilo na kontrolu kvality ovzduší, kontrolu kvality vody a potravin. Dále může být detekce toxických plynů přínosná pro minimalizaci nehod v dolech. Produkce metanu v podzemních dolech může totiž vést k malým požárům, při nichž vznikají nebezpečné plyny, jako jsou NO_2 a CO . Údaje získané ze senzorů plynu mohou být zpracovány v reálném čase nebo ukládány na cloudový server pro další použití. Získané údaje se porovnávají s příslušnými prahovými hodnotami plynu, aby bylo možné v případě potřeby přijmout případná opatření ke zmírnění následků. Příkladem dalšího využití chemorezistivních senzorů plynu pro kvalitu potravin je hodnocení čerstvosti masa a detekce toxinů v potravinách. [1]

Existuje také řada metod detekce, prognózy a diagnostiky nemocí z dechu s využitím senzorů plynu (pomocí technologie elektronického nosu). Tímto způsobem rozhodně nelze nahradit laboratorní vyšetření, které vykazuje mnohem vyšší přesnost oproti měření pomocí senzorů plynu. Hlavní výhodou je však to, že diagnostika nemocí z dechu může pomoci včasné prevenci onemocnění nebo detekci onemocnění v raném stadiu, což sníží negativní následky na zdraví. Kombinací zapojení chemorezistivních senzorů do matice s blokem pro analýzu dat lze vytvořit systém elektronického nosu. Výhodou takového systému i přes nižší přesnost jsou nižší náklady oproti laboratornímu vyšetření a neinvazivnost měření. Rozhodujícím krokem je detekce specifických cílových biomarkerů, které ukazují na možný zdravotní problém. V Tabulka 1 jsou uvedené příklady endogenních látek a možných onemocnění [1], [2], [3], [4], [5], [6], [7], [8]:

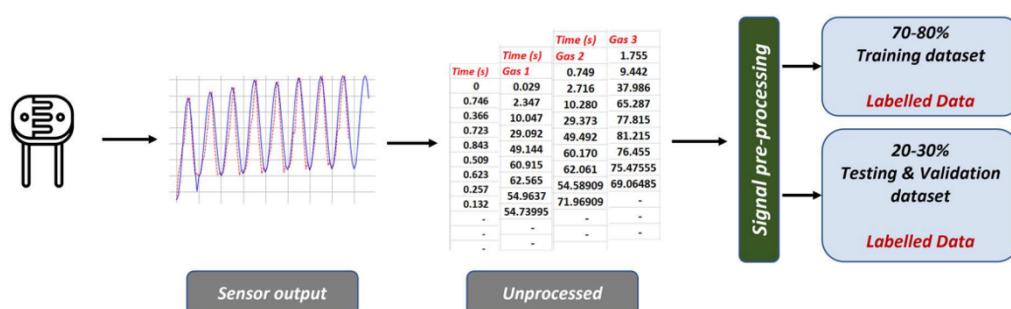
Tabulka 1: Příklady endogenních látek a možných onemocnění

Onemocnění	Endogenní látky (biomarkery)	Typ endogenních látek
Onemocnění plic / zánět / ARDS	Pentan, isopren	Těkavá organická látka (VOC)
Zánět COVID-19 neurologický	CO, NO	Anorganický plyn
Infekce horních dýchacích cest	Acetaldehyd	VOC
Poruchy gastrointestinální motility	Vodík (Hydrogen)	VOC
Gastrointestinální projevy COVID-19	Methanol	VOC
CHOPN, zánět plic, plicní tuberkulóza (TBC)	Heptan, ethan	VOC
	2 - propanol, 2- butyloktanol	Alkohol
COVID-19	Aceton, izopren, heptanal, propanol, propanal, butanon, oktanal, metylpent-2-enal, 1-chloroheptan a nonanal	VOC
Cukrovka	Aceton	VOC
Rakovina plic	Formaldehyd, hexanal, toluen	VOC
Kardiovaskulární nemoci (infarkt myokardu)	Ethanol, toluen, isobutan	VOC
Parkinsonova choroba	Perillový aldehyd	VOC
	eikosan (ikosan)	Alkán
Chronická onemocnění ledvin (CKD) a jaterní onemocnění	Amoniak	VOC

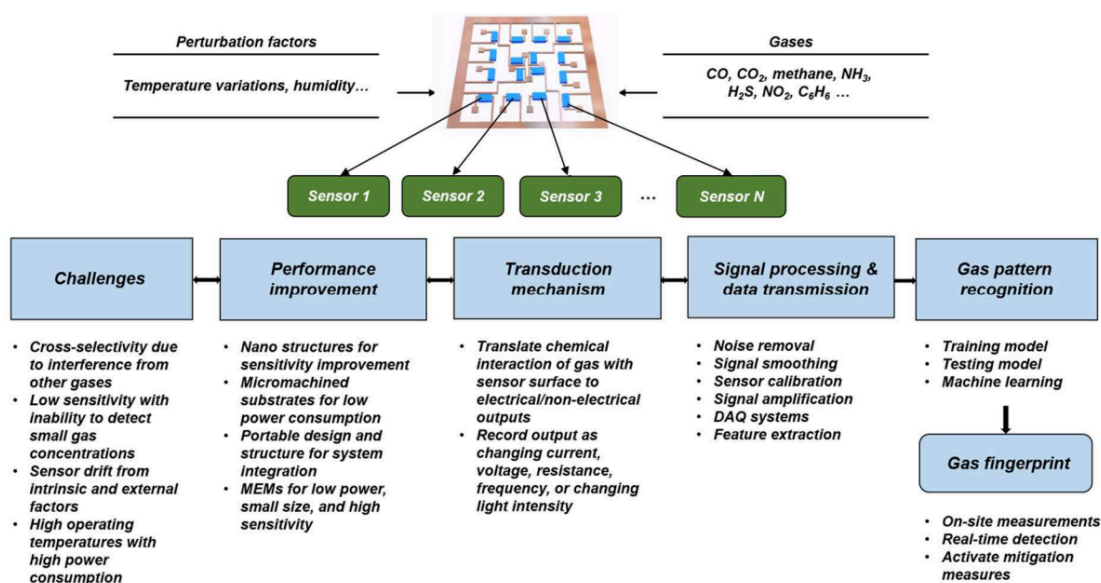
2.2. Princip činnosti

Obvykle je chemorezistivní senzor plynu určen pro jeden úkol detekce za předem definovaných podmínek. Zařízení, schopné detekovat a rozlišovat různé plyny současně, funguje díky integraci několika částečně selektivních senzorů do jednoho čipu. Strojové učení (například umělé neuronové sítě) pomáhá zvýšit výkon zařízení, protože běžně senzory plynu jsou jenom částečně selektivní. [9]

Princip činnosti chemorezistivního senzoru plynu je založen na reakci mezi analytem (ve formě plynu) a snímacím materiálem, jinými slovy adsorpce a desorpce plynu na povrchu materiálů vede ke změně elektrického odporu snímacího materiálů. Po vystavení analytu se odpor senzoru změní v důsledku výměny nosičů náboje mezi plynem a snímacím materiálem. Následně jsou parametry použité k reprezentaci analytu extrahovány z charakteristického profilu časové odezvy senzoru. Pomocí algoritmů strojového učení se zpracovávají data a vyhodnocuje se identifikace každého analytu. Na Obrázek 1 je uveden zjednodušený postup zpracování dat ze senzoru plynu. Na Obrázek 2 je znázorněno zapojení chemorezistivních senzorů do matice včetně bloku pro analýzu dat, které tak tvoří systém elektronického nosu (podrobněji viz kapitola 2.6). [4], [9]



Obrázek 1: Snímání signálů a zpracování dat ze senzoru plynu. Převzato z [1].



Obrázek 2: Blokové schéma systému elektronického nosu (viz 1.5). Převzato z [1].

K výhodám chemorezistivních senzorů patří [4]:

- nízké náklady,
- dlouhá životnost,
- snadná miniaturizace,
- široký rozsah detekovatelných plynů a VOC.

K nevýhodám chemorezistivních senzorů plynu patří [1], [4]:

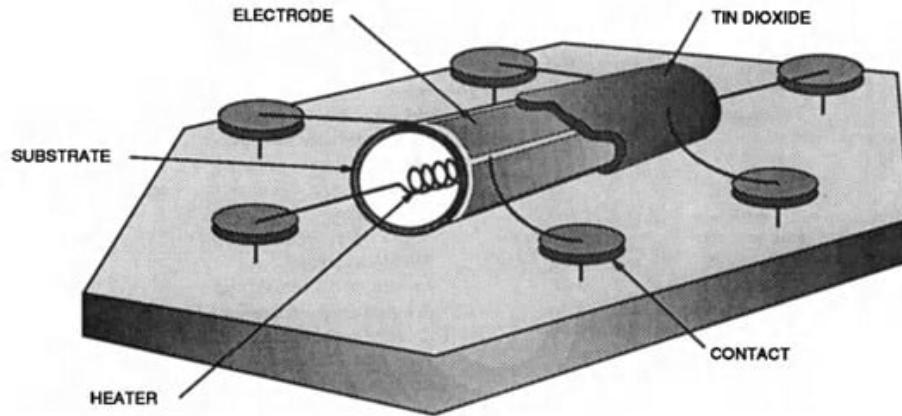
- Křížová citlivost: při detekci určitého analytu může být senzor plynu vystaven působení jiných plynů (a VOC), které mohou mít podobné vlastnosti jako cílová látka, což může způsobit horší detekci analytu.
- Závislost na vlhkosti a teplotě okolního prostředí: změny teploty a vlhkosti by mohly ovlivnit schopnost senzoru plynu detekovat analyzovaný plyn a snížit citlivost senzoru.
- Drift v čase: je možné rozdělit na drift prvního řádu a drift druhého řádu. K driftu prvního řádu dochází v důsledku interakce mezi plynem a povrchem senzoru plynu, což má za následek znečištění senzoru plynu. Drift druhého řádu nastává v důsledku vlivu vnějších provozních podmínek, například hluku, kolísání teploty a vlhkosti.
- Sensory plynu běžně pracují při vysoké teplotě (v důsledku toho chemorezistivní senzory mají velkou spotřebu energie).

2.3. Typy chemorezistivních senzorů

Jak bylo zmíněno výše, princip činnosti chemorezistivních senzorů plynu je založen na změně elektrického odporu snímacího materiálu při chemické interakci s analytem. Pro snímací vrstvu se nejčastěji používají polovodičové materiály na bázi oxidů kovů (MOX). MOX senzory vykazují vysokou citlivost a reverzibilitu pro široké spektrum aplikací, avšak díky vysoké pracovní teplotě mají vyšší spotřebu energie a nutnost použití teplotně odolnějších materiálů. Ve většině případů se pracovní teplota chemorezistivních MOX senzorů pohybuje v rozsahu 200 - 400 °C z toho důvodu, že pracovní teplota ovlivňuje tvorbu iontů kyslíku a se zvyšující teplotou se zvyšuje vodivost snímací vrstvy. MOX senzory bývají dvou typů: polovodič typu N, například na bázi ZnO , SnO_2 , TiO_2 , In_2O_3 , a polovodič typu P, například na bázi CuO . Nejčastěji se snímací vrstva vyrábí z materiálů, které patří k polovodičům typu N. MOX senzory typu N vykazují vyšší citlivost oproti MOX senzorům typu P. Oproti tomu MOX senzory typu P jsou méně ovlivňované parazitními vlastnostmi vlhkosti. Dalšími materiály, které se používají pro výrobu snímací vrstvy, jsou například vodivé polymery a uhlíkové nanotrubičky. [10]

U MOX senzorů typu N dochází při interakci molekul redukčních plynů s ionty kyslíku adsorbovanými na povrchu snímacího materiálu k uvolňování elektronů zpět do vodivostního pásu snímacího materiálu. Tento proces vede k redukci tloušťky elektronové vrstvy a k celkovému snížení elektrického odporu snímacího materiálu. V případě expozice oxidačním plynem dochází k opačné reakci, během které plyn pohlcuje elektrony ze snímacího materiálu, což vede ke zvýšení elektrického odporu snímacího materiálu typu N. U MOX senzorů typu P snímací materiál akceptuje elektrony, čímž se vytváří akumulací oblast. Při interakci s redukčním plynem se zmenšuje počet děr (pozitivních nosičů náboje) díky jejich rekombinaci s elektrony z plynů. Následkem je zvýšení celkového elektrického odporu senzoru. Při expozici oxidačními plyny se u MOX senzorů typu P odpor snižuje. [10]

Metal-oxidové senzory (MOX) lze rozdělit na dva základní typy podle technologie výroby: tlustovrstvé a tenkovrstvé senzory. Tloušťka snímacího materiálu tlustovrstvého senzoru je v rozsahu několika jednotek až několika set μm , u tenkovrstvého senzoru je to v rozsahu několika jednotek až několika set nm . Na Obrázek 3 je uveden princip zapojení tlustovrstvých MOX senzorů, které se skládají z elektrod a ohřívací cívky. Výhodou takového zapojení jsou nižší výrobní náklady a jednoduchost zapojení, nevýhodou je nízká selektivita a velká spotřeba energie. [11]



Obrázek 3: Příklad zapojení tlustovrstvého MOX senzoru. Převzato z [11].

Výhodou tenkovrstvých senzorů jsou malé rozměry, vyšší selektivita a nízká spotřeba energie. Nevýhodou jsou vyšší výrobní náklady.

2.4. Charakteristické parametry

Existuje řada faktorů ovlivňujících časovou odezvu senzoru plynu uvedených níže.

- **Teplota**

Většina používaných komerčních senzorů plynu na bázi MOX obvykle pracuje při zvýšené teplotě v rozmezí 200 - 400 °C, protože tepelná energie je nezbytná pro aktivaci adsorpce ionizovaných forem kyslíku a překonání energetické bariéry zkoumaných reakcí. [9]

- **Vlhkost**

Vlhkost okolního prostředí může mít velký vliv na přesnost detekce chemorezistivních senzorů plynu. Obecně je vlhkost považována za parazitní parametr. Zvyšující se vlhkost okolního prostředí má obvykle za následek snížení odezvy MOX senzorů, jelikož zvyšuje elektrickou vodivost snímací vrstvy. [12]

- **Koncentrace**

Obecně platí, že za identických okolních podmínek se se zvyšující koncentrací analytu zvyšuje odezva senzoru plynu (až po dosažení tzv. saturačního limitu, kdy senzor už nereaguje na zvýšení koncentrace analytu). Pro vyjádření koncentrace plynu se nejčastěji používají jednotky ppm a ppb, viz (1).

$$1 \text{ ppm} = 1000 \text{ ppb} = 1 \frac{\text{mg}}{\text{kg}} = 100 \frac{\text{mg}}{\text{l}} 0,0001 \% \quad (1)$$

Dále jsou uvedené charakteristické parametry chemorezistivních senzorů plynu.

- **Citlivost**

Citlivost (angl. sensitivity, někdy také response) je jedním z klíčových parametrů senzorů plynu, udává poměr změny výstupní snímané veličiny senzoru (odpor nebo napětí senzoru při působení analytu o dané koncentraci) ke vstupní snímané veličině senzoru (odpor nebo napětí senzoru v přítomnosti vzduchu bez zkoumaného analytu). Lze ji zlepšit například zmenšením velikosti částic materiálu snímací vrstvy senzoru, provozem při vysoké teplotě nebo v režimu aktivace světlem (například pomocí ultrafialového záření). Citlivost S [-] (resp. [%]) lze vyjádřit několika způsoby (2) - (4), kde R_a [Ω] je odpor senzoru v přítomnosti čistého vzduchu bez zkoumaného analytu, R_g [Ω] je odpor senzoru v přítomnosti analytu a ΔR je rozdíl hodnot odporu senzoru R_g a R_a [4], [13]:

$$S = \frac{R_g}{R_a} \quad (2)$$

$$S = \frac{R_g - R_a}{R_a} = \frac{\Delta R}{R_a} \quad (3)$$

$$S = \frac{R_g - R_a}{R_a} \times 100\% = \frac{\Delta R}{R_a} \times 100\% \quad (4)$$

- **Selektivita**

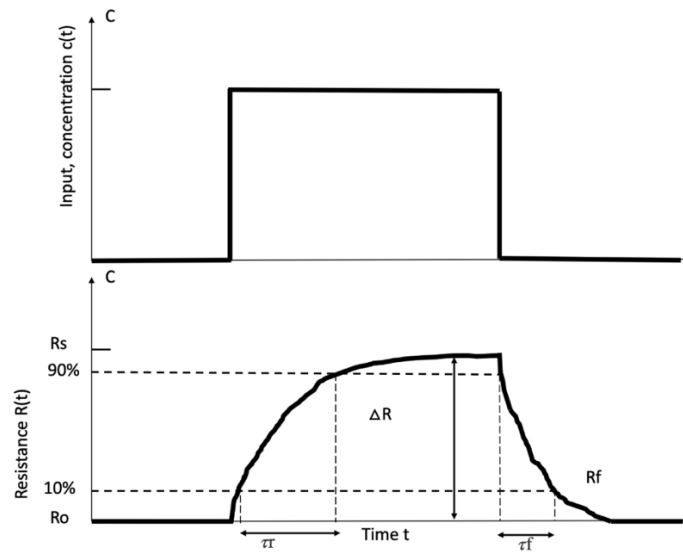
Selektivita materiálu senzoru plynu je preferenční chemorezistivní snímání konkrétního plynu C_g [-] v přítomnosti ostatních plynů C_i [-] za identických provozních podmínek. Selektivitu [-] lze vyjádřit pomocí vztahu (5) [13]:

$$\text{Selektivita} = \frac{\text{Citlivost senzoru vůči plynu na rozhraní } (C_i)}{\text{Citlivost senzoru vůči analyzovanému plynu } (C_g)} \quad (5)$$

- **Doba odezvy**

Doba odezvy je časový interval od zahájení expozice senzoru analyzovaným plynem po okamžik, kdy odezva senzoru dosáhne stabilní hodnoty na určité procentuální úrovni (nejčastěji 90 % maximální hodnoty, značí se T_{90} [s]). Čím kratší je doba odezvy, tím lepší jsou snímací vlastnosti senzoru. Také obvykle čím vyšší je koncentrace analyzovaných plynů, tím je kratší doba odezvy senzoru. Dalšími parametry, které mohou ovlivnit dobu odezvy, jsou průtok plynu, teplota a tlak analyzovaného plynu. [13]

Na Obrázek 4 je zobrazená doba odezvy v závislosti na koncentraci analytu (10 % až 90 % maximální hodnoty).



Obrázek 4: Doba odezvy senzoru v závislosti na koncentraci analytu, kde R_0 [Ω] je odpor senzoru na začátku měření, R_S [Ω] je odpor v ustáleném stavu, R_f [Ω] je 10% odezvy senzoru na konci doby snímání, τ_r [s] je doba odezvy, τ_f [s] je doba zotavení (v tomto případě se uvažuje 10 % až 90 % R_S). Upraveno z [14].

- **Doba zotavení**

Doba zotavení je čas, který senzor potřebuje k tomu, aby se po odstranění analyzovaného plynu a následném vyčištění senzoru čistým vzduchem průběh odezvy senzoru vrátil na hodnotu základní úrovně signálu. Doba zotavení senzoru by měla být krátká, aby bylo možné senzor používat opakovaně. [13]

- **Stabilita a opakovatelnost**

Stabilita snímacího materiálu znamená jeho schopnost opakovaně a dlouhodobě zachovávat své snímací vlastnosti. Tyto parametry závisí na velikosti částic a morfologii snímacího materiálu, typu interakcí a reakcí probíhajících mezi plynem a snímacím materiálem, provozních podmínkách senzoru atd. [13]

2.5. Průběh odezvy senzoru

Obvykle je analýza surových dat ze senzoru plynu obtížná, proto se pro zlepšení detekce analytu pro analýzu používají extrahované parametry ze surových dat (podrobněji je motivace extrakce parametrů popsána v kapitole č. 3.1). V Tabulka 2 jsou uvedené typické parametry signálu senzoru plynu [15]:

Tabulka 2: Typické parametry časové odezvy senzoru.

Doména	Parametr	Charakteristika
Časová	Polynomiální koeficienty A_i , $i = 0, 1, 2, 3 \dots$	$S(t) = \sum_{i=0}^N A_i t^i$, kde $S(t)$ je snímaný průběh odezvy, N je stupeň fitované polynomiální funkce
	$S'(t)_{max}$, $S'(t)_{min}$	První derivace $S'(t)$, $S(t) = \sum_{i=0}^N A_i t^i$, kde $S(t)$ je snímaný průběh odezvy
	$S''(t)_{max}$, $S''(t)_{min}$	Druhá derivace $S''(t)$, $S(t) = \sum_{i=0}^N A_i t^i$, kde $S(t)$ je snímaný průběh odezvy
	Plocha	Integrál, plocha = $\int_0^t S(t)dt$, kde $S(t)$ je snímaný průběh odezvy
	Exponenciální koeficienty, a_1 , b_1, c_1, a_2, b_2, c_2 ,	$S(t) = a_1 \left(1 - e^{\frac{-t}{b_1}}\right) + c_1$, exponenciální aproximace pro expoziční křivku (průběh odezvy senzoru během fáze expozice) $S(t) = a_2 e^{\frac{-t}{b_2}} + c_2$, exponenciální aproximace pro křivku zotavení, kde $S(t)$ je snímaný průběh odezvy a t je čas
Frekvenční	Frekvenční koeficienty f	Rychlá Fourierova transformace snímaného signálu $S(t)$, frekvenční koeficienty f
	Vlnkové koeficienty c_A, c_D	Diskrétní vlnková transformace snímaného signálu $S(t)$, aproximační koeficienty c_A a detailní koeficienty c_D

Profily průběhu odezvy pro každý analyt se skládají ze série reprodukováných měření. Každé jednotlivé měření obsahuje tři fáze: fáze nalezení základní úrovně signálu (angl. baseline), fáze expozice analytem a fáze proplachování [9]:

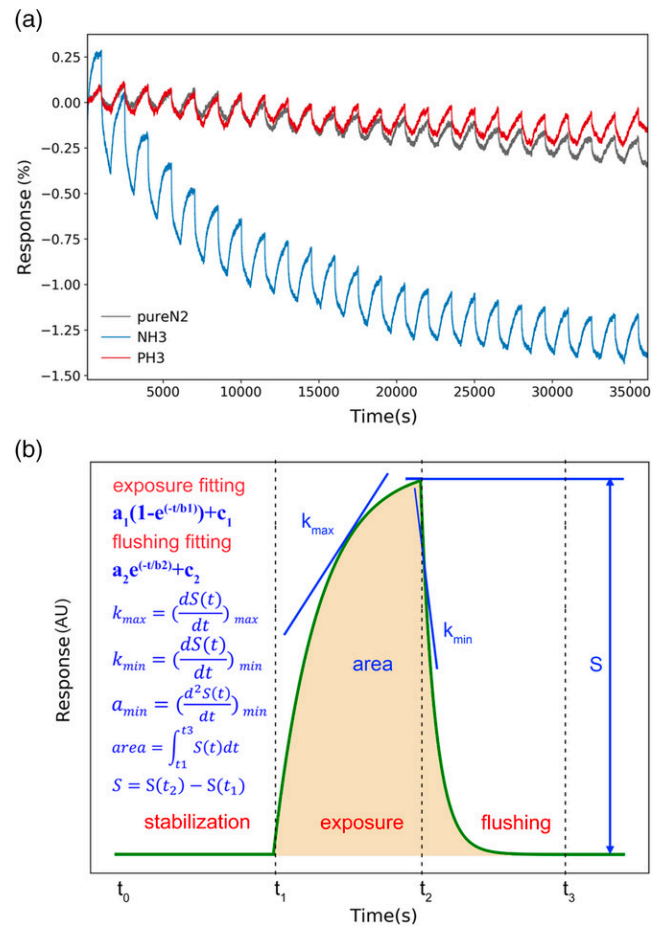
- 0. fáze - nalezení základní úrovně signálu: čištění suchým N_2 (případně syntetickým vzduchem) nebo čistým vzduchem bez zkoumaného analytu pro odstranění nečistot na povrchu senzoru a stabilizaci signálu senzoru,
- 1. fáze - expozice: expozice senzoru plynu analytem,
- 2. fáze - proplachování: senzor plynu se obnovuje suchým N_2 (případně syntetickým vzduchem) nebo čistým vzduchem bez zkoumaného analytu.

1. a 2. krok se cyklicky opakuje.

Na Obrázek 5 je uveden příklad průběhu odezvy senzoru plynu. Profily průběhu odezvy během fáze expozice analytem (t_1-t_2) a fáze proplachování (t_2-t_3) jsou proloženy exponenciálními funkcemi.

Parametry průběhu odezvy (snímaného signálu senzoru plynu) uvedené na Obrázek 5 jsou následující:

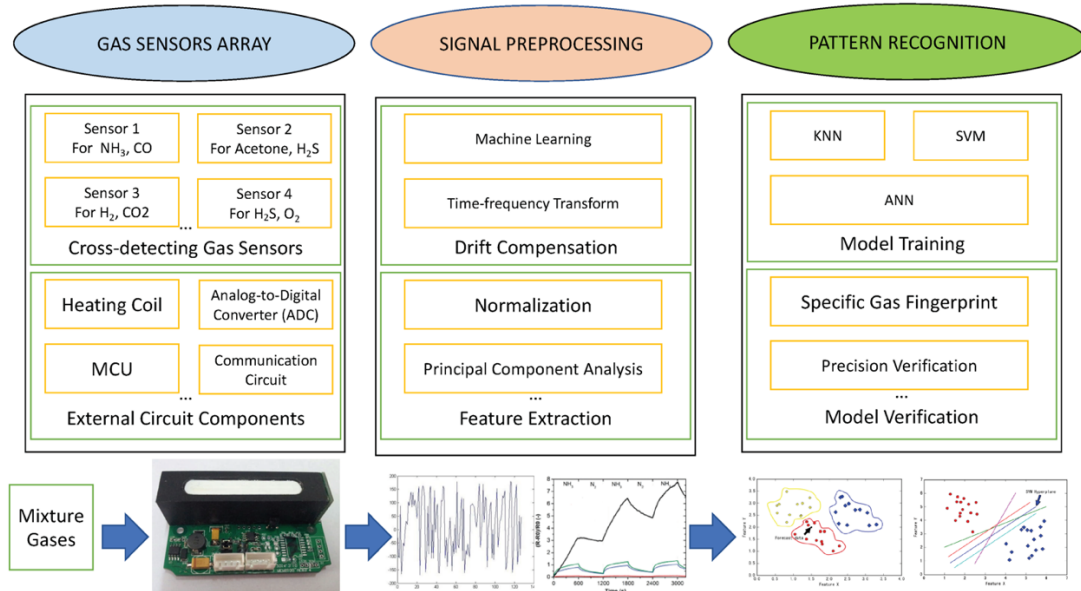
- (t_1-t_2): koeficienty a_1 , b_1 a c_1 ,
- (t_2-t_3): koeficienty a_2 , b_2 a c_2 ,
- maximální hodnota k_{max} a minimální hodnota k_{min} 1. derivace průběhu odezvy podle času,
- minimální hodnota a_{min} 2. derivace průběhu odezvy podle času,
- rozdílová odezva pro fáze expozice $S = S(t_2) - S(t_1)$,
- plocha, kterou ohraničuje časový průběh odezvy v časech t_1 až t_3 .



Obrázek 5: a) profil průběhu odezvy senzoru plynu pro cyklické měření, b) schéma profilu průběhu odezvy $S(t)$ pro měření se skládajícího ze stabilizačního procesu (t_0-t_1) pod proudem čistého N_2 k nalezení základní úrovně, fáze expozice (t_1-t_2) a fáze proplachování (t_2-t_3). Převzato z [9].

2.6. Problém selektivity

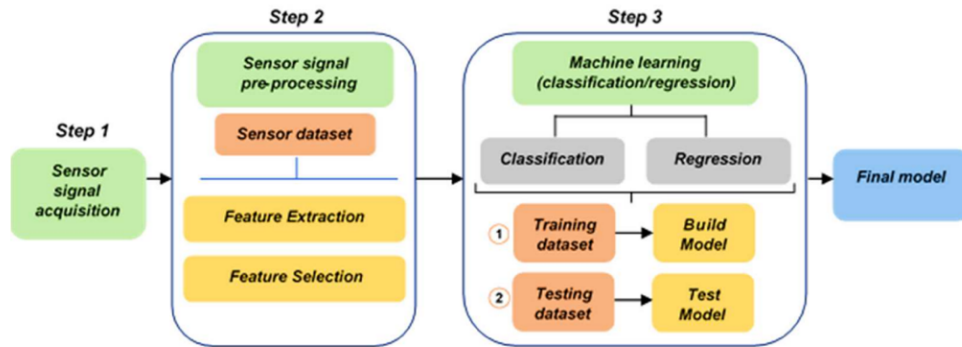
Chemorezistivní senzory mají nevýhodu, která spočívá v nižší selektivitě. Jedním z možných řešení je modifikace povrchu snímacího materiálu a zlepšení selektivity pomocí katalyzátoru. Také se používají filtry, které propouští k senzoru pouze molekuly požadovaného plynu. Praktické využití takových způsobů zvýšení selektivity senzorů pro identifikaci plynů je omezené, protože křížová citlivost senzorů plynů na různé druhy plynů a VOC není zcela eliminována. Dalším možným způsobem zvýšení selektivity je použití matice senzorů (elektronického nosu, angl. e-nose), kde každý senzor vykazuje jinou reaktivitu vůči specifickým plynům. Pro zpracování dat ze senzorů plynu se používají algoritmy strojového učení, například analýza hlavních komponent (PCA), Gaussian Mixture Model (GMM), klasifikátory support vector machine (SVM) a k-nearest neighbors (k-NN, možným překladem by mohlo být k-nejbližších sousedů) a algoritmy strojového učení založené na neuronových sítích. Nevýhoda použití matice senzorů spočívá v nákladech, spotřebě energie a objemu celého systému, které úměrně rostou se zvyšujícím se počtem použitých senzorů plynu. Na Obrázek 6 je uveden postup snímání a zpracování dat pro technologie elektronického nosu (neboli Smart Gas Sensing). [16]



Obrázek 6: Technologie elektronického nosu (Smart Gas Sensing). Převzato z [14]

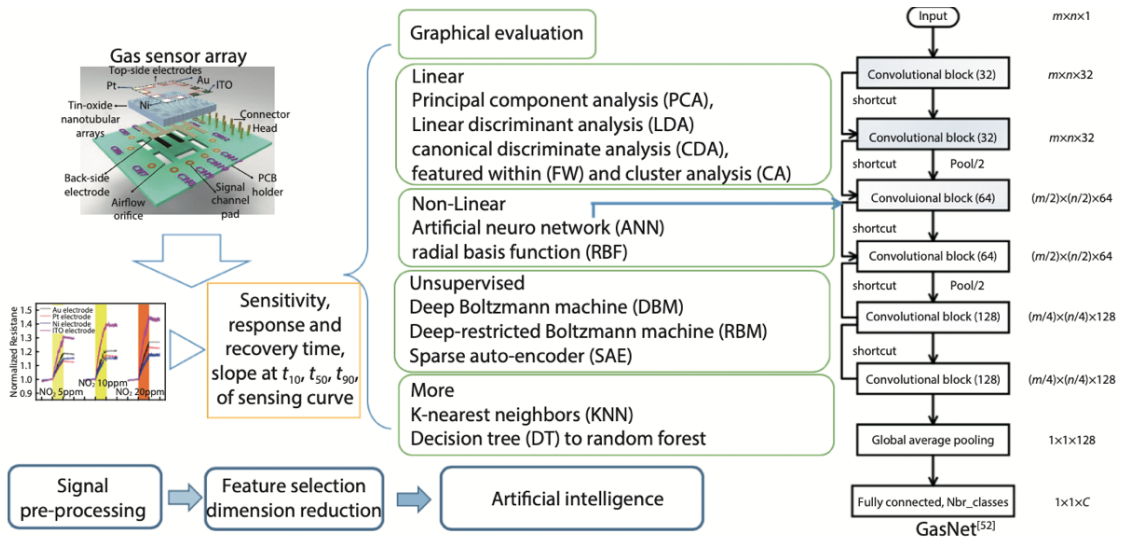
3. Algoritmy strojového učení

Algoritmy strojového učení pomáhají řešit problémy křížové citlivosti elektrochemických senzorů plynů, drift základní úrovně signálu v čase a vysoké dimenzionality dat. Na Obrázek 7 je uvedené blokové schéma udávající postup pro analýzu dat ze senzoru plynu zahrnující: získání surových dat, předzpracování dat včetně extrakce a selekce parametrů a následné využití algoritmů strojového učení pro řešení klasifikačních nebo regresních úloh.



Obrázek 7: Blokové schéma udávající postup pro analýzu dat ze senzoru plynu pomocí algoritmů strojového učení. Převzato z [1].

Na Obrázek 8 je uvedené rozdělení metod strojového učení pro analýzu dat ze senzoru plynu.



Obrázek 8: Rozdělení metod strojového učení pro analýzu dat ze senzoru plynu. Převzato z [4].

3.1. Předzpracování dat

V některých případech je před zpracováním dat nutné provést předzpracování dat, které může zahrnovat například manipulaci se základní úrovní signálu pro kalibraci driftu v čase, redukci šumu (například pomocí průměrování) a následnou normalizaci (případně standardizaci) dat pro kompenzaci rozdílů ve stupnicích jednotlivých senzorů. [1]

Následně se pro snížení rizika přeučení modelu strojového učení a redukci dimenze provádí extrakce příznaků (parametrů). Velký soubor dat průběhů odezev senzorů plynu na určitý analyt bude mít pravděpodobně vysokou redundanci v důsledku koexistence datových bodů. Transformace dat ze senzoru na charakteristické parametry (angl. features), které obsahují podstatnou část informací, pomáhá minimalizovat složitost původních dat a tím snížit výpočetní náročnost modelu strojového učení, vyhnout se nadměrnému přizpůsobení modelu strojového učení (tzv. overfitting, viz kapitola 3.5.2) a poskytuje jednodušší interpretaci dat. Existuje několik druhů metod extrakce příznaků, které se používají pro e-nosy. Jednou z nich je kategorie metod zaměřujících se na extrakci dílčích příznaků (parametrů) signálu v časové oblasti. Takové příznaky lze extrahovat pro různé fáze měření, například citlivost, sklon odezvy senzoru, dobu náběhu odezvy a průměrnou hodnotu signálu. Některé techniky se naopak zaměřují na extrakci ustálených a přechodových příznaků, nízkofrekvenčních a vysokofrekvenčních příznaků získaných například pomocí rychlé Fourierovy transformace (FFT) nebo diskrétní vlnkové transformace (DWT). Podrobněji se možnými parametry odezvy zabývá kapitola 2.5 (viz tabulka 2). [1], [17]

Pro korektní volbu algoritmu strojového učení pro následné zpracování dat je nutné provést testování normality dat a ověření předpokladů pro konkrétní algoritmus. Podrobněji se tím zabývá praktická část této práce (viz kapitola 5).

3.2. Selekce parametrů

Pro snížení redundance dat a zlepšení výsledků dalšího zpracování (např. vizualizace, klasifikace, shlukování) předzpracování dat obvykle zahrnuje výběr (selekci) důležitých parametrů. Získané parametry se použijí jako vstup například do PCA (analýza hlavních komponent) pro vytvoření unikátního markeru pro konkrétní analyt. [17]

3.2.1. Test korelace

Pro zjištění, jestli mezi dvěma kvantitativními proměnnými existuje vztah, případně jak moc silný tento vztah je, se po extrakci parametrů provádí test korelace, který měří sílu a směr vztahu mezi proměnnými. V případě, kdy mezi proměnnými existuje silná korelace, se pro další zpracování používá pouze jedna proměnná z tohoto páru. V důsledku toho se odstraní redundantní proměnné a multikolinearita.

Pro kvantifikaci lineárního vztahu mezi dvěma kvantitativními proměnnými, které mají normální rozdělení, reprezentované náhodnými veličinami X a Y se používá výběrový Pearsonův korelační koeficient vypočítaný na základě výběrového souboru. Výběrový Pearsonův korelační koeficient nabývá hodnot v intervalu $(-1,1)$, při jeho výpočtu se vychází z realizace dvojice pozorovaných hodnot těchto náhodných veličin [18]:

$$\begin{pmatrix} x_1 \\ y_1 \end{pmatrix}, \begin{pmatrix} x_2 \\ y_2 \end{pmatrix}, \dots, \begin{pmatrix} x_n \\ y_n \end{pmatrix} \quad (6)$$

n ... celkový počet datových bodů (pozorování).

Výběrový Pearsonův korelační koeficient je dán vztahem (7):

$$r = \frac{\sum_{i=1}^n x_i y_i - n\bar{x}\bar{y}}{(n-1)s_x s_y} \quad (7)$$

r ... Pearsonův korelační koeficient,

\bar{x} a \bar{y} ... výběrové průměry,

s_x a s_y ... výběrové směrodatné odchylky.

Neparametrický Spearmanův korelační koeficient se používá pro identifikaci monotónní závislosti mezi proměnnými X a Y , což může zahrnovat i nelineární vztahy. Spearmanův korelační koeficient je robustní vůči odlehlým hodnotám a nemá předpoklad normálního rozdělení dat, jehož hodnoty jsou také v intervalu $(-1,1)$. Spearmanův korelační koeficient lze vypočítat pomocí vztahu (8) [18]:

$$r_s = \frac{\sum_{i=1}^n x_{ri} y_{ri} - n \bar{x}_r \bar{y}_r}{(n-1) s_{x_r} s_{y_r}} \quad (8)$$

r_s ... Spearmanův korelační koeficient,

x_{ri} ... pořadí hodnoty x_i v rámci vzestupně uspořádaných hodnot x_1, \dots, x_n ,

y_{ri} ... pořadí hodnoty y_i v rámci vzestupně uspořádaných hodnot y_1, \dots, y_n ,

\bar{x}_r a \bar{y}_r ... průměry hodnot x_{ri} , respektive y_{ri} ,

s_{x_r} a s_{y_r} ... směrodatné odchylky hodnot x_{ri} , respektive y_{ri} .

3.2.2. ANOVA F-test

Jednou z metod používaných pro selekci příznaků pro následující klasifikaci je jednosměrná ANOVA (analýza rozptylu, angl. (one-way Analysis of Variance)).

Statistická metoda ANOVA (někdy se také používá pojem ANOVA F-test) se používá k porovnání hodnot „vícenásobných průměrů“ (angl. „multiple means“) souboru dat a k vizualizaci toho, zda existuje nějaký významný rozdíl mezi průměrnými hodnotami více skupin (tříd). Výstupem ANOVA je F-scóre (případně F-statistika), které lze vypočítat pomocí následujících kroků [19]:

1. Variance (rozptyl) mezi skupinami BMS (angl. between mean squares) je dán vztahem (10):

$$\text{Between sum of squares (BSS)} = n_1(\bar{X}_1 - \bar{X})^2 + n_2(\bar{X}_2 - \bar{X})^2 + \dots \quad (9)$$

k ... počet skupin,

n_k ... počet vzorků ve skupině k ,

\bar{X}_k ... průměr skupiny k ,

\bar{X} ... celkový průměr všech vzorků.

$$\text{Between mean squares (BMS)} = \frac{BSS}{df} \quad (10)$$

df ... tzv. „between-group“ počet stupňů volnosti (počet skupin $k-1$).

2. Variance uvnitř skupiny WMS (angl. within mean squares) se počítá podle vztahu (12):

$$\text{Within sum of squares (WSS)} = (n_1 - 1)\sigma_1^2 + (n_2 - 1)\sigma_2^2 + \dots \quad (11)$$

σ_k ... standardní směrodatná odchylka skupiny k .

$$\text{Within mean squares (WMS)} = \frac{\text{WSS}}{df_w} \quad (12)$$

df_w ... tzv. „within-group“ počet stupňů volnosti ($N - k$, kde N je počet vzorků).

3. Následně F-scóre (F-statistika) je dáno vztahem (13):

$$F = \frac{BMS}{WMS} \quad (13)$$

Metoda selekce příznaků pomocí jednosměrné ANOVA vychází z předpokladu, že parametry s nízkou variancí obsahují méně užitečných informací. Pro každý parametr (prediktor) se provádí jednosměrná ANOVA. Parametry, které mají menší F-scóre, než předem nastavený práh se pro další analýzu nepoužijí. [20]

3.3. Vizualizace multidimenzionálního prostoru

Pro vizualizaci multidimenzionálního prostoru se používá řada metod, které lze rozdělit na lineární (například PCA) a nelineární (například t-SNE, ISOMAP, kernel PCA). Existuje velké množství metod, avšak v této kapitole budou podrobně rozebrány metody, které se následně používaly v praktické části této práce.

3.3.1. PCA

PCA (analýza hlavních komponent) je algoritmus učení bez učitele. Je to neparametrická statistická lineární metoda využívaná především pro redukcí dimenzionality nebo kompresi vysokorozměrného souboru dat na nízkorozměrný podprostor příznaků s cílem zachovat většinu relevantních informací. Přestože PCA snižuje dimenzi, důležité informace ze vstupních dat jsou efektivně zachovány promítnutím původních dat do nového prostoru, jehož osy co nejlépe odpovídají směrům s největší variancí. [9], [17]

PCA algoritmus je následující [17], [21]:

0. Normalizace skóre x_{ti} :

$$x_{ti} = \frac{(r_{ti} - \bar{r}_t)}{\sigma_t}, \quad (14)$$

kde x_{ti} je normalizovaná hodnota skóre pro i -tou časovou odezvu senzoru s indexem t , i je index časové odezvy senzoru, r_{ti} je původní i -tá časová odezva senzoru s indexem t , \bar{r}_t je průměrná hodnota časové odezvy s indexem t a σ_t je standardní odchylka časové odezvy s indexem t .

1. Korelační matice A je dána vztahem (15):

$$A = \begin{pmatrix} C_{11} & C_{12} & \dots & C_{1m} \\ C_{21} & C_{22} & \dots & C_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ C_{m1} & C_{m2} & \dots & C_{mm} \end{pmatrix}, \quad (15)$$

kde C_{ab} je korelační koeficient mezi senzory a a b ($C_{11} = C_{22} = C_{mm} = 1$) a m je maximální počet senzorů.

Zároveň platí:

$$|A - \lambda E| = 0, \quad (16)$$

kde λ je vlastní číslo ($\lambda_1 > \lambda_2 > \dots > \lambda_m$) a E je jednotková matice. Vlastní číslo lze získat pro každý senzor:

$$AV_j = \lambda_j V_j, \quad (17)$$

kde V_j je matice vlastních vektorů.

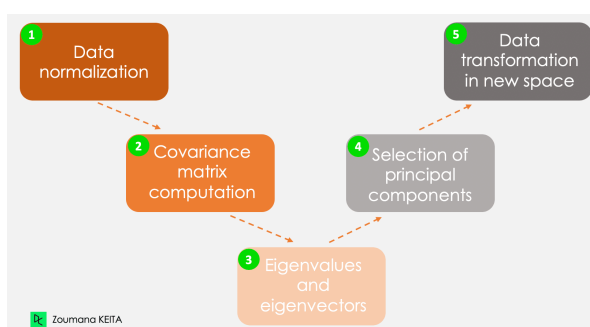
Vlastní vektor s největším vlastním číslem odpovídá první hlavní komponentě. Druhé hlavní komponentě odpovídá vlastní vektor s druhým největším vlastním číslem atd.

2. Nakonec je hlavní komponenta (Z_{ji}) vypočítána jako součet součinu normalizovaného skóre x a vlastních vektorů v :

$$Z_{ji} = v_{1j}x_{1i} + v_{2j}x_{2i} + v_{mj}x_{mi}. \quad (18)$$

3. Transformace dat do nového dimenzionálního prostoru zahrnuje mapování původních dat do nového podprostoru definovaného hlavními komponentami. Mapování se provádí vynásobením původních dat vlastními vektory. Tato transformace nemění samotná původní data, ale poskytuje novou perspektivu, která data lépe reprezentuje.

Na Obrázek 9 je uveden postup pro počítání PCA.



Obrázek 9: Hlavní kroky pro počítání PCA. Převzato z [21].

3.3.2. t-SNE

t-SNE (t-distributed Stochastic Neighbor Embedding) je nelineární metoda redukce dimenzí, která se používá pro vizualizaci vysokodimenzionálních dat v nízkodimenzionálním prostoru. Algoritmus zjišťuje podobnost bodů ve vysokodimenzionálním prostoru a snaží se ji zachovat při reprezentaci bodů v nízkodimenzionálním prostoru. Metoda t-SNE efektivně zachovává jak lokální strukturu dat, tak zohledňuje i globální sktrukturu dat. Hyperparametrem „perplexita“ je možné nastavit rovnováhu mezi lokální a globální strukturou dat, hodnota hyperparametru je obvykle v rozmezí 5 až 50. „Perplexita“ určuje, jak široké bude okolí každého bodu při výpočtu párových pravděpodobností podobnosti bodů. Tento hyperparametr ovlivňuje efektivní počet sousedů, které každý datový bod zohledňuje v průběhu procesu redukce dimenze. Vystup t-SNE není statický a podléhá změnám v závislosti na procesu optimalizace. Algoritmus je obvykle iterován až do dosažení stabilní konfigurace (např. se provádí 5000 iterací). [22], [23]

Metoda t-SNE modeluje bod, který je vybrán jako soused jiného bodu ve vysoko a nízko dimenzionálním prostoru. Princip je založen na výpočtu párové podobnosti mezi všemi datovými body ve vysokodimenzionálním prostoru pomocí Gaussovského jádra. Body, které mají menší vzdálenost od sebe, mají vyšší pravděpodobnost, že budou vybrány jako sousední ve srovnání s body, které jsou více vzdálené od sebe (u nichž je tato pravděpodobnost nižší). Můžou být použité různé metriky vzdálenosti, v praktické části této diplomové práce byla použita metoda t-SNE se standardní Euklidovskou vzdáleností (výchozí nastavení funkce t-SNE v MATLAB). Algoritmus se snaží mapovat datové body vysokodimezionálního prostoru do prostoru nižší dimenze při zachování párových podobností minimalizací Kullback-Leiblerovou divergencí mezi pravděpodobnostním rozdělením původního vysokodimenzionálního prostoru a rozdělením v prostoru nižší dimenze (minimalizací rozdílu mezi pravděpodobnostními distribucemi v obou prostorových reprezentacích). Algoritmus používá k minimalizaci divergence gradientní sestup (při kterém se minimalizuje ztrátová funkce, což je míra rozdílu podobností datových bodů v původním vysokodimenzionálním prostoru a jejich odpovídající podobností v nízkodimenzionálním prostoru). V prostoru nižší dimenze jsou párové pravděpodobnosti podobnosti rekonstruovány pomocí Studentova t-rozdělení. Povaha této práce neumožňuje detailnější popis (vysvětlení) metody t-Distributed Stochastic Neighbor Embedding (t-SNE). Ten však lze nalézt v publikaci [24] od autorů tohoto algoritmu. [22], [23], [25]

3.4. Klasifikace

3.4.1. k-NN

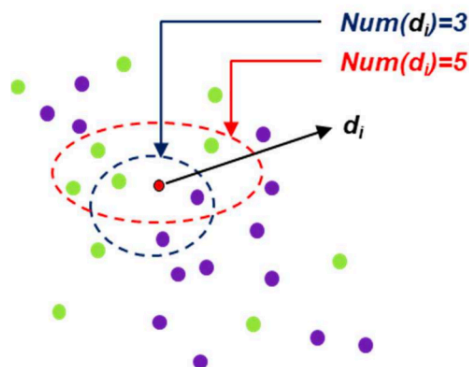
Algoritmus k-nearest neighbors (k-NN, mohlo by být přeloženo jako k-nejbližších sousedů) je nelineární klasifikační algoritmus učení s učitelem, který ke klasifikaci používá metriku vzdálenosti. Počet nejbližších sousedů k a metrika vzdálenosti jsou dva hlavní hyperparametry tohoto algoritmu. Na Obrázek 10 je znázorněn princip, na kterém je založen algoritmus k-NN. Pomocí metriky vzdálenosti (například Euklidovské vzdálenosti) algoritmus najde k nejbližších sousedů (vzdálenost se počítá pro všechny vzorky). k nejbližších sousedů je reprezentováno pomocí $Num(d_i)$ v rámci lokalizované oblasti a je určena podobnost mezi všemi vzorky dat v této oblasti (algoritmus předpokládá, že podobné vzorky jsou vždy vedle sebe). Výsledkem je, že každý nový datový bod d_i , který spadá do tohoto lokálního regionu, je na základě podobnosti přiřazen ke skupině, která má v daném okolí největší zastoupení. Pro výpočet vzdálenosti mezi vzorky dat lze použít například Minkowského vzdálenost, Euklidovskou a Manhattanickou vzdálenost. [1], [15], [26]

Minkowského rovnice vzdálenosti je zobecněná a je možné z ní odvodit Manhattanickou a Euklidovskou vzdálenost, kde $d(x, y)$ je Minkowského vzdálenost mezi datovými body x a y , p je parametr, který určuje typ vzdálenosti a n je počet dimenzí [26]:

$$d(x, y) = \left(\sum_{i=1}^n |x_i - y_i|^p \right)^{\frac{1}{p}} \quad (19)$$

$p = 1$... Manhattanická vzdálenost,

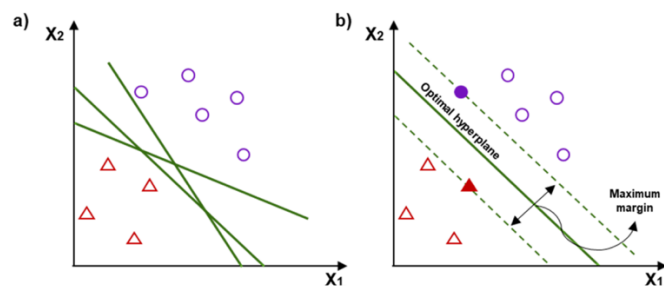
$p = 2$... Euklidovská vzdálenost.



Obrázek 10: Princip klasifikace algoritmu k-NN. Převzato z [1].

3.4.2. SVM

Algoritmus Support Vector Machine (SMV) je klasifikační algoritmus učení s učitelem (existuje však rozšíření SVM, které patří k tzv. semi-supervised learning). Hlavním cílem algoritmu SVM je nalézt hyperplochu pro rozdělení datových bodů, která maximalizuje tzv. margin, který je definovaný jako vzdálenost mezi hyperplochou a nejbližšími body (tzv. support vektory). Na Obrázek 11 je uveden příklad pro dvoudimenzionální prostor. V případě, že data nejsou lineárně separovatelná (data nelze rozdělit přímkou či rovinou), pomocí jádrové funkce (angl. kernel function) algoritmus transformuje (mapuje) původní datové body do prostoru příznaků vyšší dimenze, kde data jsou lineárně separovatelná pomocí hyperplochy. K nejpoužívanějším jádrovým funkcím patří lineární kernel, polynomiální kernel a radiální bázová funkce (RBF). [1], [27]



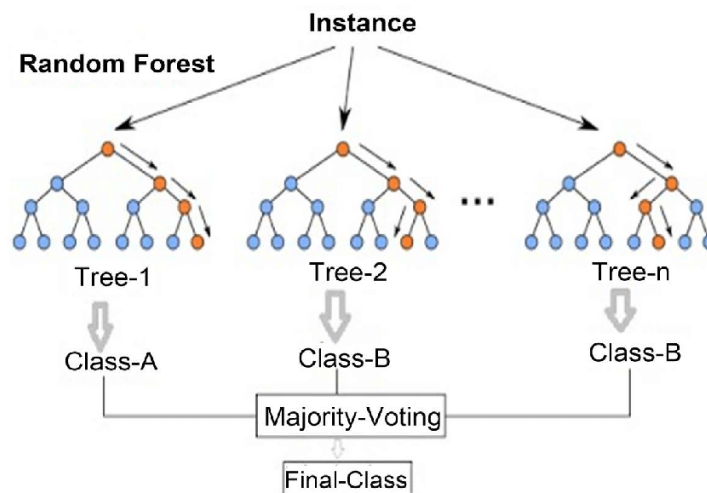
Obrázek 11: Princip SVM algoritmu: a) různé možné hranice pro rozdělení tříd, b) optimální hranice s maximálním tzv. marginem. Převzato z [1].

V praktické části této práce pro klasifikaci se používá SVM s RFB kernelem. Tento algoritmus má několik hyperparametrů, které lze optimalizovat pro dosažení lepšího výkonu modelu. Popis těchto hyperparametrů a vliv jejich hodnot na nedostatečné přizpůsobení modelu (angl. underfitting, podrobně je rozebíráno v kapitole 3.5.2) a nadměrné přizpůsobení modelu (angl. overfitting, podrobně je rozebíráno v kapitole 3.5.2) lze nalézt v článku [28].

3.4.3. Random Forest

Decision tree (DT, mohlo by být přeloženo jako Rozhodovací stromy) je algoritmus používaný pro klasifikační a regresní úlohy. Algoritmus využívá hierarchickou strukturu podobnou stromu. Tzv. root node (mohlo by být přeloženo jako kořenový uzel) představuje kompletní datový soubor. Od tzv. root node (kořenového uzlu) se strom rozděluje na větve, jejichž výsledkem je n -tice uzlů, každý vnitřní uzel může být dále rozdělen. Algoritmus je ukončen, když je dosaženo tzv. leaf node (mohlo by být přeloženo jako listový uzel). [1]

Random Forest (RF, mohlo by být přeloženo jako Náhodný les) je algoritmus založený na metodě baggingu, která vytváří více paralelních (slabých neboli méně přesných) modelů současně, v tomto případě více DT. Původní datový soubor je náhodně rozdělen na n podmnožin, každá z nich je použita pro trénování jednotlivých DT. Tento přístup umožňuje, aby byl každý DT trénován na nekorelovaných datech, čímž se zvyšuje generalizace klasifikace a snižuje se riziko přeučení modelu. Výsledek klasifikace pomocí metody RF je dán většinovým hlasováním (agregací predikcí z jednotlivých DT). Na Obrázek 12 je uveden základní koncept klasifikátoru RF. [1], [29]



Obrázek 12: Základní koncept klasifikátoru RF. Převzato z [30].

RF má několik hyperparametrů, které lze optimalizovat pro dosažení lepšího výkonu modelu, tyto hyperparametry a možnosti optimalizace jejich hodnot jsou uvedené v článku [31].

3.4.4. LDA

Algoritmus LDA (lineární diskriminační analýza, angl. linear discriminant analysis) je klasifikátor snažící se najít podprostor příznaků, který optimalizuje separabilitu tříd. Je to algoritmus učení s učitelem. Hlavním účelem LDA je minimalizovat rozptyl mezi body v rámci jedné třídy a zároveň maximalizovat rozdíl středních hodnot jednotlivých tříd. Výhodou LDA je zřetelná separabilita tříd a snadná interpretace výsledků. V případě, že třídy nejsou lineárně separovatelné, LDA není schopná korektně separovat třídy. [1], [9]

Algoritmus LDA lze popsat následovně [26]:

V datovém souboru je C tříd, μ_i je vektor středních hodnot každé třídy i , $i = 1, 2, 3, \dots, C$. M_i je počet vzorků v rámci třídy i , $i = 1, 2, 3, \dots, C$. $M = \sum_{i=1}^C M_i$ je celkový počet vzorků.

Matice rozptylu uvnitř třídy je dána vztahem (20):

$$S_w = \sum_{i=1}^C \sum_{j=1}^{M_i} (X_j - \mu_i)(X_j - \mu_i)^T \quad (20)$$

Matice rozptylu mezi třídami je dána vztahem (21), kde $\mu = \frac{1}{C} \sum_{i=1}^C \mu_i$ je střední hodnota celého datového souboru:

$$S_b = \sum_{i=1}^C (\mu_i - \mu)(\mu_i - \mu)^T \quad (21)$$

Projekce dat je $y = U^T x$, kde U^T je projekční matice.

LDA vypočítá transformaci, která maximalizuje rozptyl mezi třídami a minimalizuje rozptyl uvnitř třídy, kde \widetilde{S}_b a \widetilde{S}_w jsou matice rozptylu v prostoru projekcí:

$$\max \frac{|U^T S_b U|}{|U^T S_w U|} = \max \frac{|\widetilde{S}_b|}{|\widetilde{S}_w|} \quad (22)$$

3.5. Odhad výkonu modelu

3.5.1. Matice záměn

Matice záměn (angl. confusion matrix) je matice, která se používá k vyhodnocení výkonu klasifikačních modelů pro danou testovací sadu. Lze ji určit pouze v případě, že jsou známy skutečné hodnoty testovacích dat. Matice záměn, která je uvedena na Obrázek 13, udává výkon modelu ve formě matice. Existuje několik metrik výkonu modelu, nejčastěji se používá přesnost (angl. accuracy) modelu a F1-scóre (někdy se také používá pojem F-measure). Přesnost je dána vztahem (23) [15]:

$$A = \frac{TP + TN}{TP + FP + TN + FN} \quad (23)$$

A ... přesnost,

TP ... skutečně pozitivní hodnoty (true positive),

FN ... falešně negativní hodnoty (false negative),

FP ... falešně pozitivní hodnoty (false positive),

TN ... skutečně negativní hodnoty (true negative).

		True Class	
		Positive	Negative
Predicted Class	Positive	TP	FP
	Negative	FN	TN

Obrázek 13: Matice záměn (příklad pro dvě třídy).

F1-scóre udává rovnováhu mezi tzv. precision a tzv. recall, kde metrika precision je dána vztahem (24), metrika recall je dána vztahem (25), F1-scóre je dáno vztahem (26). Existuje více typů F1-scóre, například tzv. mikro F1-scóre a makro F1-scóre, v článku [32] lze nalézt více informací včetně vzorců. [32]

$$Precision = \frac{TP}{TP + FP} \quad (24)$$

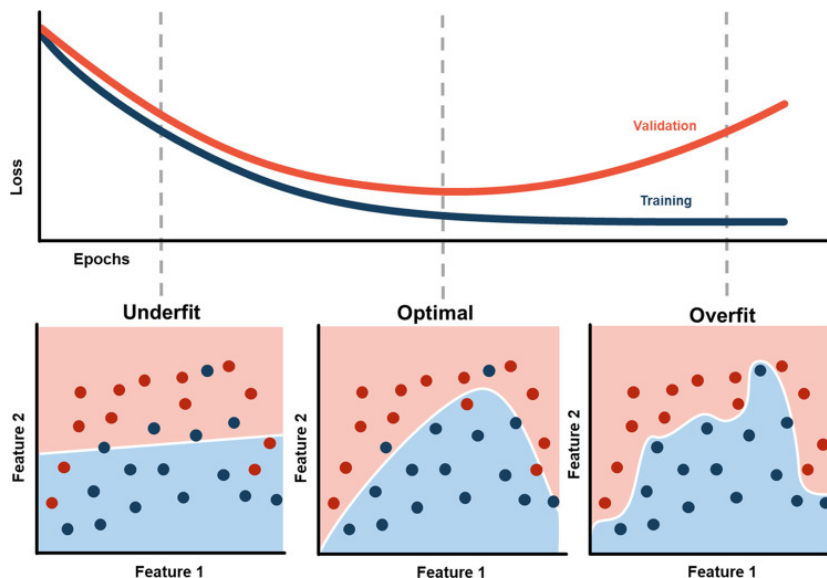
$$Recall = \frac{TP}{TP + FN} \quad (25)$$

$$F1 - score = 2 \cdot \frac{Precision \cdot Recall}{Precision + Recall} \quad (26)$$

3.5.2. Křížová validace

Křížová validace (CV) je statistická metoda používaná k aproximaci přesnosti modelu strojového učení a odhadu predikční chyby modelu. CV pomáhá předejít nadměrnému přizpůsobení (přeučení) modelu (angl. overfitting) a získat maximální a stabilní přesnost predikce. Metoda se také používá k optimalizaci (ladění) hyperparametrů a výběru (selekcii) algoritmu. [15], [17], [33]

Potřeba použití křížové validace (CV) vyplývá z toho, že algoritmy umělé inteligence jsou náchylné k nadměrnému přizpůsobení, viz Obrázek 14. Nadměrné přizpůsobení modelu nastává, když se algoritmus naučí dělat předpovědi na základě přítomnosti specifických rysů v trénovacím datovém souboru, které nejsou obecně generalizované na nová data. V důsledku toho přesnost předpovědí modelu na trénovacím souboru dat není spolehlivým ukazatelem budoucího výkonu modelu. Řešením je pro odhad očekávaného výkonu modelu použít testovací sadu dat, která je nezávislá na trénovací sadě. [33]



Obrázek 14: Na obrázku vlevo je ukázka nedostatečného přizpůsobení modelu (tzv. underfitting), optimálního přizpůsobení modelu (uprostřed) a vpravo je ukázka nadměrného přizpůsobení modelu (tzv. overfitting), což může mít za následek špatnou predikci na budoucích, dosud neznámých datech. Převzato z [33].

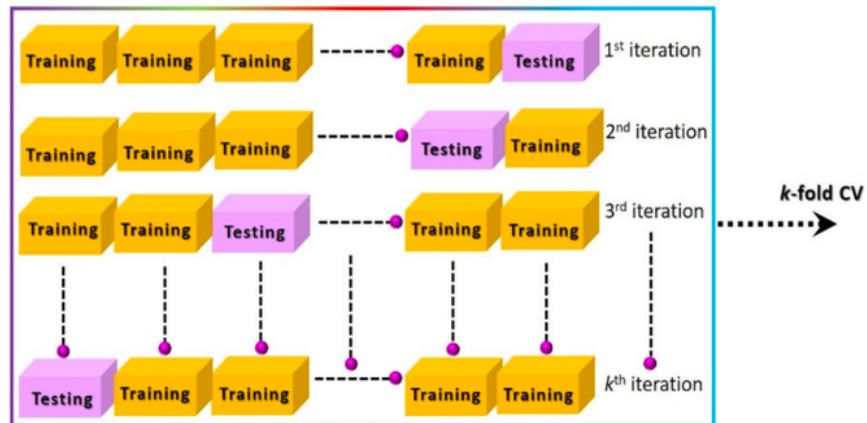
Dále jsou v této práci rozebrány pouze některé z řady metod CV. Například v článku [33] lze nalézt porovnání různých metod CV včetně složitosti různých modelů a scénářů, za kterých by se jednotlivé přístupy měly používat.

- **Metoda k-násobné CV**

Jednou z nejčastěji používaných metod je metoda k-násobné křížové validace (K-Fold CV). Při k-násobné křížové validaci je vstupní datový soubor k -krát rozdělen na k stejně velkých částí vzorků (tzv. foldů, podmnožin). Predikční funkce používá $k - 1$ částí vzorků pro trénování a zbytek je použit pro testování. Algoritmus k-násobné CV je uveden na Obrázek 15, metodu lze popsat pomocí následujících kroků [15], [17]:

1. Vstupní datový soubor je rozdělen na k stejně velkých částí.
2. Každá část je jedenkrát použita jako testovací soubor dat a zbývající části tvoří trénovací soubor pro danou iteraci (celkově je provedeno k iterací).

Celkový odhad přesnosti (případně jiné metriky výkonu) modelu se vypočítá v každé iteraci a po k -té iteraci se zprůměruje, aby se dosáhlo konečné míry přesnosti modelu. Volba optimální hodnoty k je závislá na řadě proměnných (např. velikostí souboru dat), ale obvykle se používá $k = 5$ nebo $k = 10$. [17], [33]



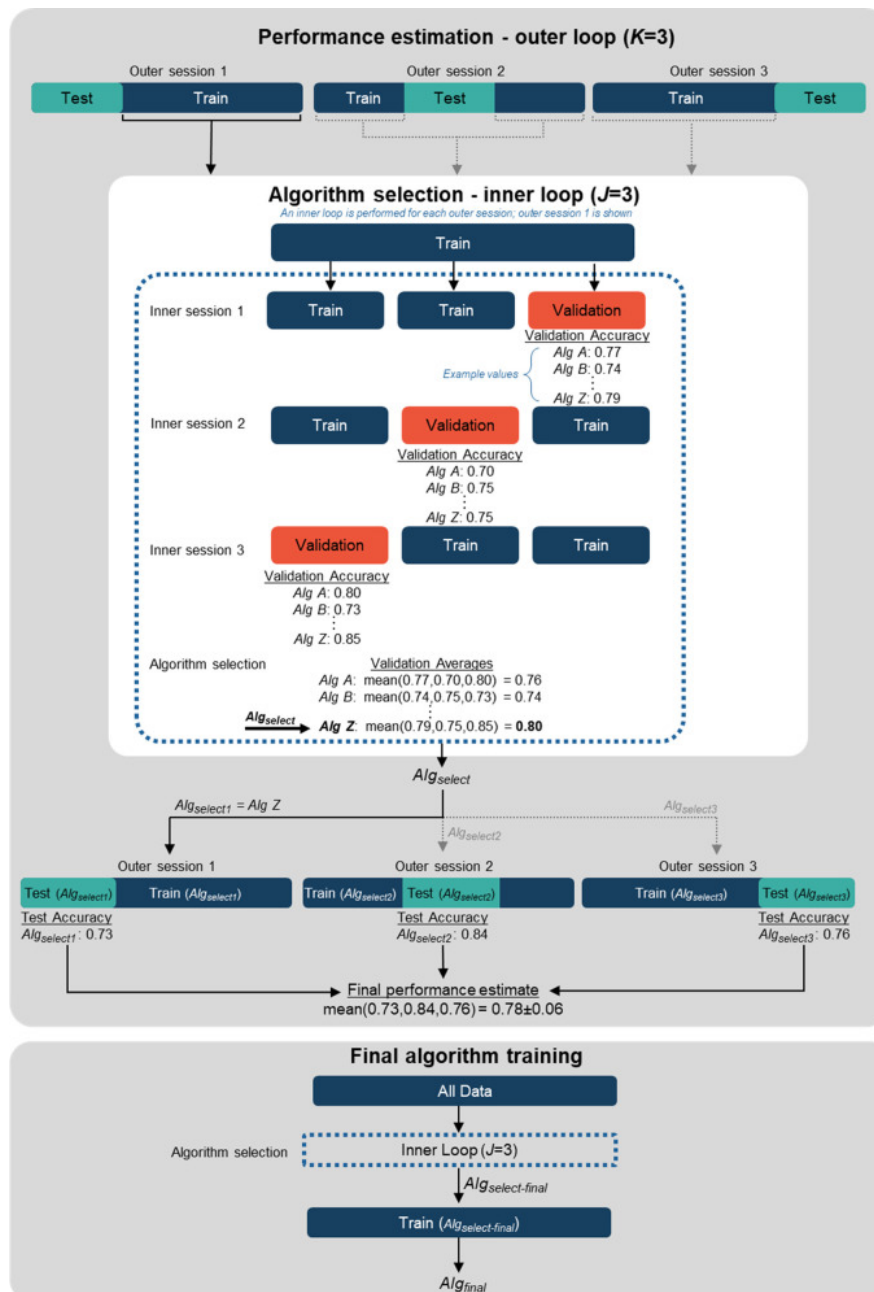
Obrázek 15: Princip k-násobné křížové validace. Převzato z [17].

- **Metoda LOOCV**

Metoda Leave-one-out křížová validace (LOOCV) je speciální případ k-násobné křížové validace, kde počet iterací k se rovná počtu vzorků n . Každý vzorek datové sady bude použit jednou pro testování. Výhoda této metody spočívá v maximalizaci dat použitých pro trénování, nevýhodou je velká výpočetní náročnost, proto je tato metoda vhodnější pro menší datové sady. Celkem je natrénováno N modelů, celkový odhad výkonu (přesnosti) je průměrem z N měření výkonu. [33]

- **Metoda Nested CV**

Metoda nested CV (možným překladem by mohlo být vnořená křížová validace) umožňuje jak odhad výkonu modelu, tak optimalizaci (ladění) hyperparametrů modelu a je užitečná pro malé soubory dat. Při nested CV vytváří dvě smyčky CV: vnější smyčka s k stejně velkých podmnožin (tzv. foldů) a vnitřní smyčka s j stejně velkých podmnožin (tzv. foldů). Vnitřní smyčka se používá k optimalizaci hyperparametrů (nalezení optimálních hodnot hyperparametrů) a vnější smyčka se používá k odhadu výkonu modelu (algoritmu) vybraného ve vnitřní smyčce (model, který měl ve vnitřní smyčce nejvyšší přesnost) s optimálními hyperparametry. Celkový odhad přesnosti (nebo jiné metriky výkonu modelu) se vypočítá jako průměr odhadů přesnosti různých k modelů (ve vnější smyčce nested CV). [33]



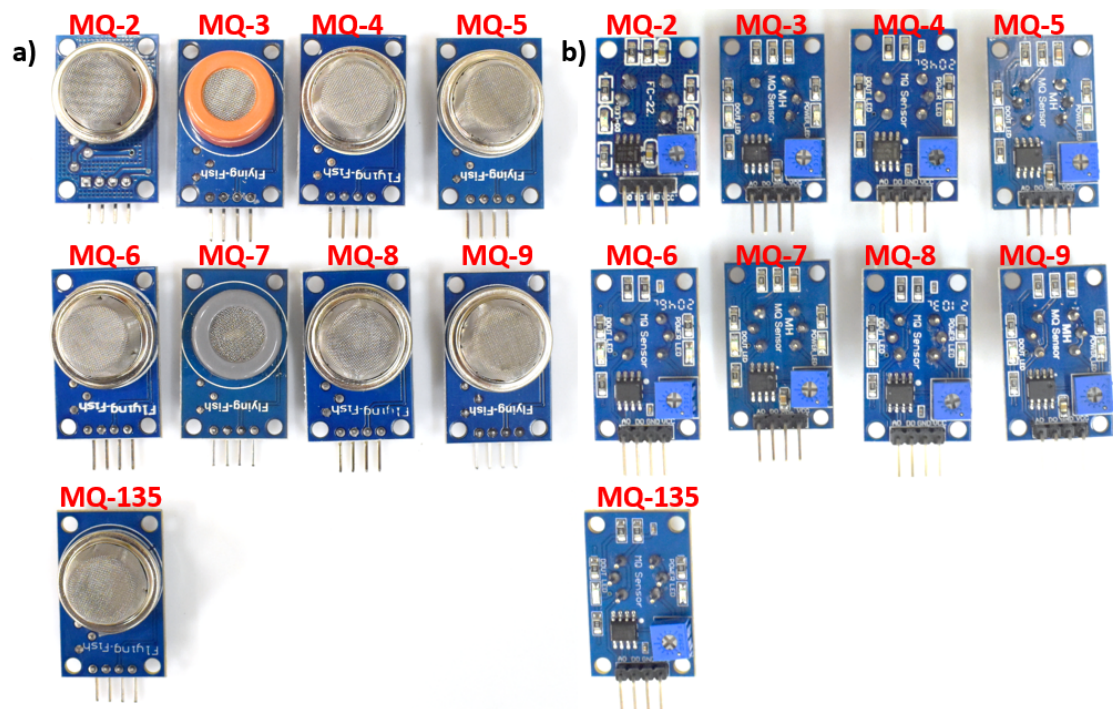
Obrázek 16: Princip nested křížové validace. Převzato z [33]

Na Obrázek 16 je uveden příklad nested CV, kde $k=3$, $j=3$. V horní části obrázku je uveden princip odhadu výkonu, kde pro každou vnější podmnožinu (tzv. fold) se provádí vnitřní smyčka, která vybírá nejlepší algoritmus. Nejlepší algoritmus Alg_{select} , který byl zvolen ve vnitřní smyčce, se následně používá pro trénování a testování vnější podmnožiny (tzv. foldu). Konečný odhad výkonu je průměr výsledků testů pro každou vnější podmnožinu (tzv. fold). V dolní části obrázku je zobrazen princip trénování finálního modelu, který také zahrnuje vnitřní smyčku pro výběr algoritmu (nalezení optimálních hodnot hyperparametrů). Finální model $Alg_{select-final}$ je natrénován na všech datech. [33]

4. Testování VOC

4.1. MQ senzory plynu

Pro měření VOC byla zvolena sada tlustovrstvých chemorezistivních senzorů plynu viz Obrázek 17 (senzor MQ9 byl použit pouze v druhé variantě zapojení matice senzoru, viz kapitola 4.2.3):



Obrázek 17: Fotografie celé sady použitých senzorů (a) pohled shora, (b) pohled zdola.

- MQ2 je senzor pro detekci hořlavých plynů propanu, metanu, butanu, vodíku;
- MQ3 je senzor pro detekci alkoholu, ethanolu, benzenu;
- MQ4 je senzor pro detekci hořlavých a zemních plynů;
- MQ5 je senzor pro detekci LPG, zemních a topných plynů;
- MQ6 je senzor pro detekci LPG, isobutanu a propanu;
- MQ7 je senzor pro detekci oxidu uhelnatého;
- MQ8 je senzor pro detekci vodíku;
- MQ9 je senzor pro detekci oxidu uhelnatého, hořlavých plynů;
- MQ135 je senzor kvality vzduchu.

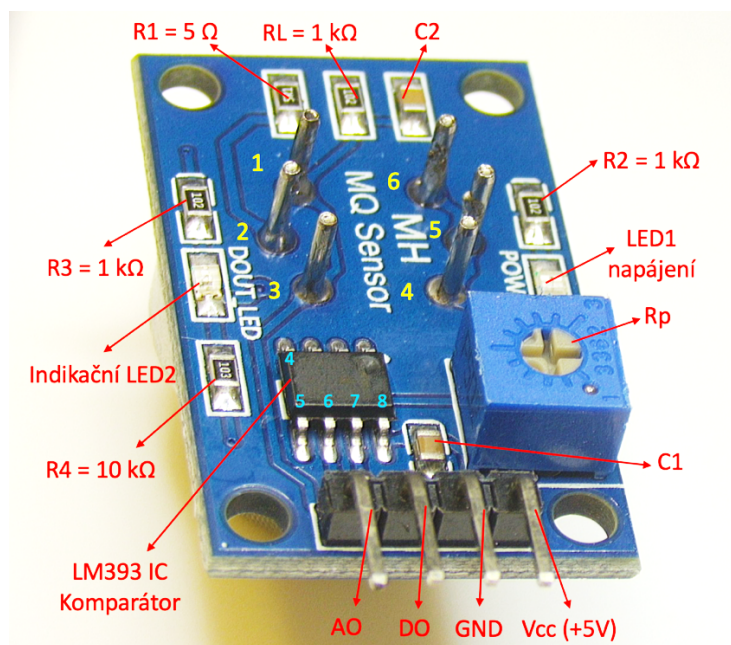
Výhoda MQ senzorů plynu spočívá v nízkých pořizovacích nákladech (cena jednoho MQ senzoru je kolem 60 Kč). Nevýhoda zvolených MQ senzorů plynu spočívá ve velké spotřebě energie, která je přibližně 100 mW, křížové citlivosti a větších rozměrech (například rozměr MQ4 senzoru je 32 x 20 x 22 mm).

Modul senzoru plynu MQ3 je zobrazen na Obrázek 18 až 20. Ovládacím prvkem R_p (potenciometrem) se nastavuje citlivost. LED1 je indikátor napájení. LED2 je indikátor digitálního výstupu (D0). Mezi elektrodami 1 a 6, 3 a 4 je odpor citlivé vrstvy R_{sens} , mezi elektrodami 2 a 5 je odpor R_h (heater). Komparátor LM393 IC je určen pro získání digitálního výstupu (výstup ve stavu on nebo off), v této práci se tato funkce nevyužívá.

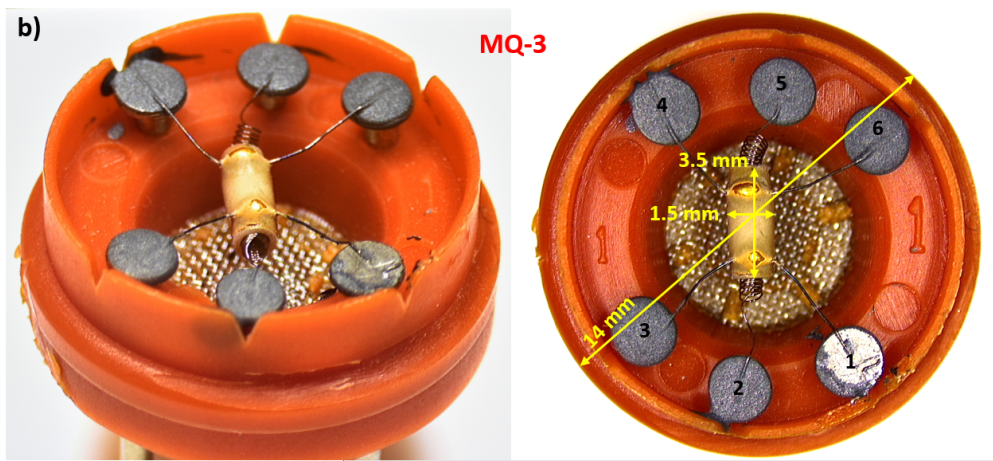
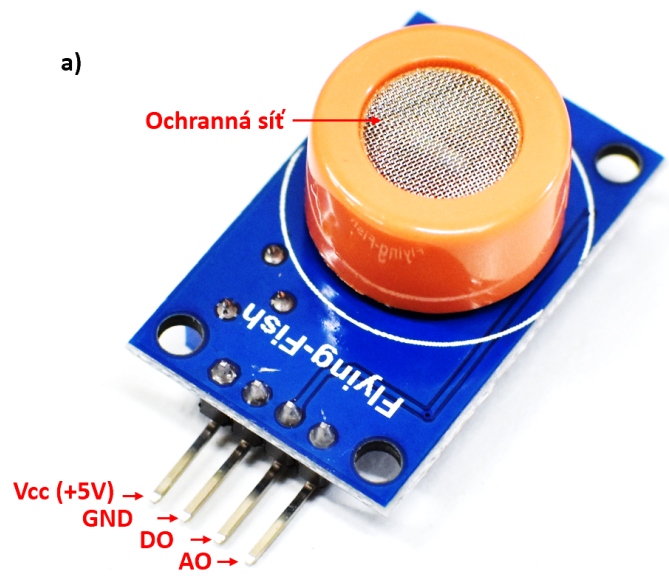
Popis výstupních pinů MQ3 senzoru jsou uvedené v Tabulka 3:

Tabulka 3: Značení výstupních pinů modulu MQ senzoru.

Č. pinu	Název pinu
1	Vcc (napájení +5 V)
2	GND
3	D0 (digitální výstup)
4	A0 (analogový výstup)

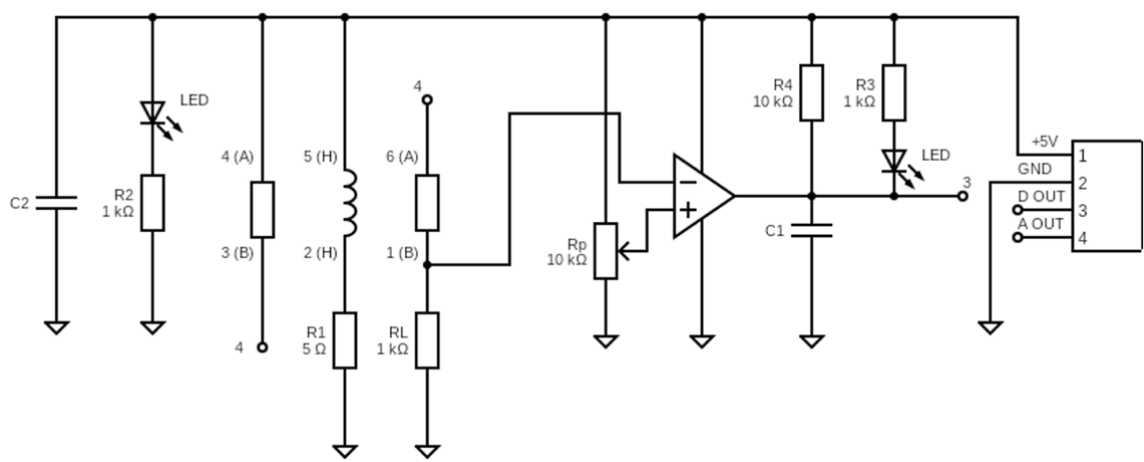


Obrázek 18: Popis součástek senzoru plynu MQ3. Červenou barvou jsou značené součástky integrovaného obvodu MQ senzoru, modrou barvou jsou značené piny integrovaného obvodu LM393 IC (v této práci se nevyužívají), žlutou barvou jsou značené elektrody MQ senzoru.



Obrázek 19: Senzor plynu MQ3, (a) celkový pohled, (b) detail vlastního snímacího prvku s měřicími elektrodami (odpor měřen mezi piny 1, 3 a piny 4,6) a výhřívací spirálou (piny 2 a 5), material citlivé vrstvy SnO_2

Schéma zapojení senzoru plynu MQ3 je uvedené na obrázku 19.



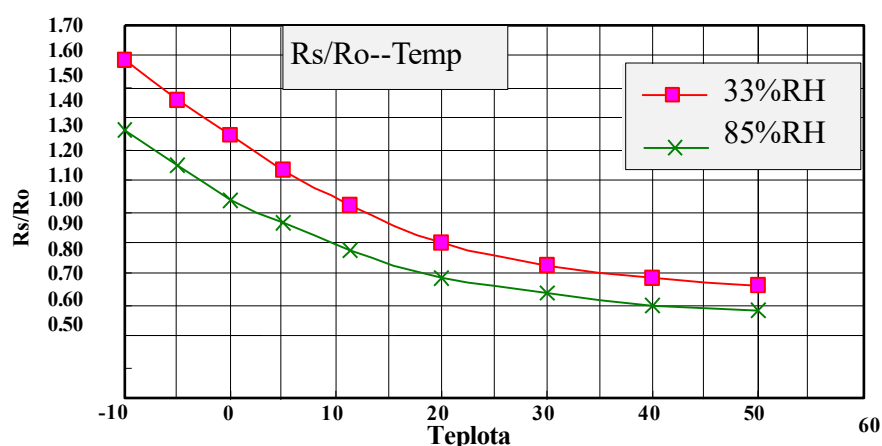
Obrázek 20: Schéma zapojení chemorezistivního senzoru plynu MQ3.

Vlastnosti použitých MQ senzorů plynu jsou shrnuty v Tabulka 4:

Tabulka 4: Vlastností MQ senzorů plynu dle dostupných katalogových listů.

	MQ -2	MQ -3	MQ -4	MQ -5	MQ -6	MQ -7	MQ -8	MQ -9	MQ -135
Napájecí napětí (Vc), (Vh)	5V ± 0.1V DC	5V ± 0.1V DC	5V ± 0.1V DC	5V ± 0.1V DC	5V ± 0.1V DC	5V ± 0.1V DC	5V ± 0.1V DC	5V ± 0.1V DC	5V ± 0.1V DC
Zatěžovací odpor (RL)	20 kΩ	200 kΩ	20 kΩ	20 kΩ	20 kΩ	20 kΩ	10 kΩ	20 kΩ	20 kΩ
Odpor snímací vrstvy (Rs)	3 kΩ – 30 kΩ	1 MΩ – 8 MΩ	10 kΩ – 60 kΩ	10 kΩ – 60 kΩ	10 kΩ – 60 kΩ	2 kΩ – 20 kΩ	10 kΩ – 60 kΩ	2 kΩ – 20 kΩ	30 kΩ – 200 kΩ
Odpor výhřevu (Rh)	33 Ω ± 5%	33 Ω ± 5%	33 Ω ± 5%	33 Ω ± 5%	33 Ω ± 5%	33 Ω ± 5%	33 Ω ± 5%	33 Ω ± 5%	33 Ω ± 5%
Tepelný výkon	< 800 mW	< 750 mW	< 750 mW	< 800 mW	< 750 mW	350 mW	< 800 mW	< 340 mW	< 800 mW
Teplota okolí	-20°C až 50°C	-10°C až 50°C	-10°C až 50°C	-10°C až 50°C	-10°C až 50°C	-20°C až 50°C	-10°C až 50°C	-20°C až 50°C	-10°C až 45°C
Relativní vlhkost	< 95% RH	< 95% RH	< 95% RH	< 95% RH	< 95% RH	< 95% RH	< 95% RH	< 95% RH	< 95% RH

Vliv vlhkosti na změnu hodnoty odporu snímací vrstvy senzoru MQ3 je ukázán na Obrázek 21.

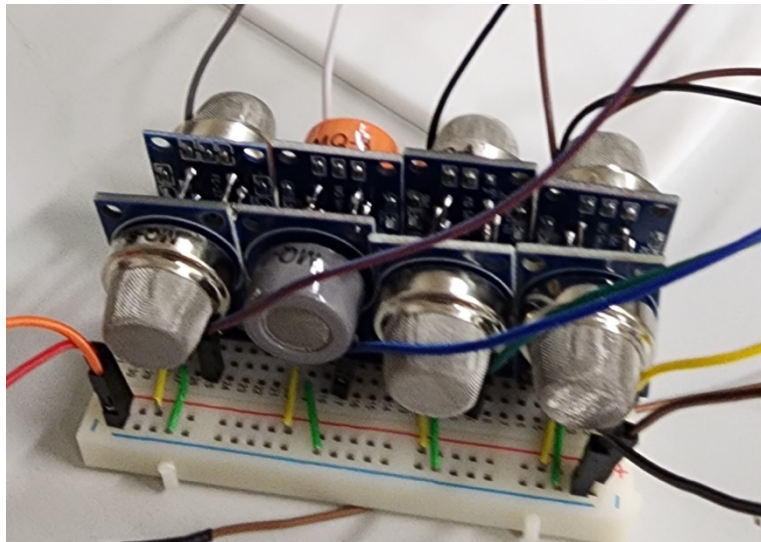


Obrázek 21: Graf závislosti relativní změny hodnot odporů citlivé vrstvy pro různé teploty okolí a dvě úrovně relativní vlhkosti (33 % a 85 %) pro senzor MQ3. Ro je odpor citlivé vrstvy při teplotě 20 °C a detekované látky ethanol s koncentrací 400 ppb (0,4 mg/l). Rs je odpor při teplotě a vlhkosti dle grafu, detekovaná látka je ethanol s koncentrací 400 ppb. Upraveno z [34].

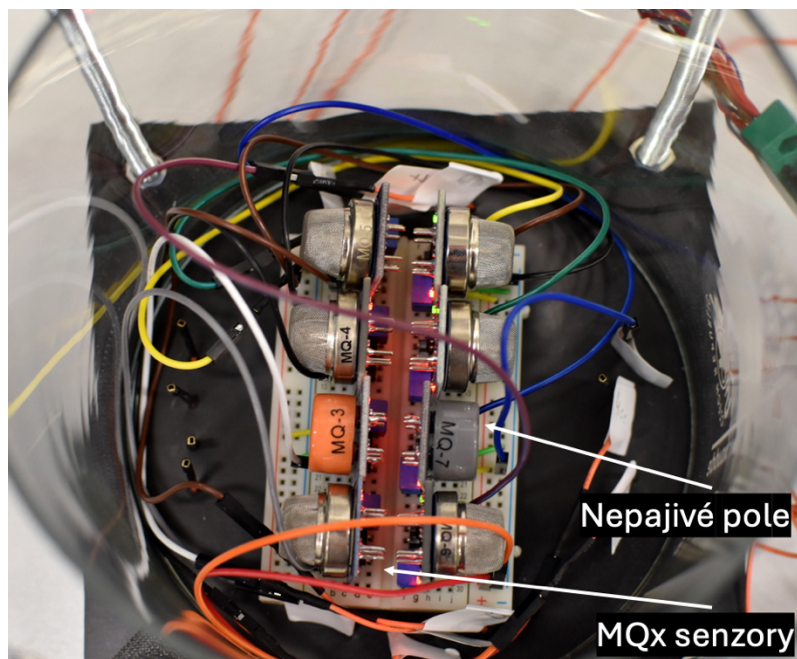
4.2. Metodika experimentu

4.2.1. První varianta experimentu

První varianta zapojení matice senzorů plynu je uvedena na Obrázek 22. Pro zapojení matice senzorů bylo použité nepájivé pole. První varianta testovací komory je znázorněná na Obrázek 23, objem komory je 1,65 l.



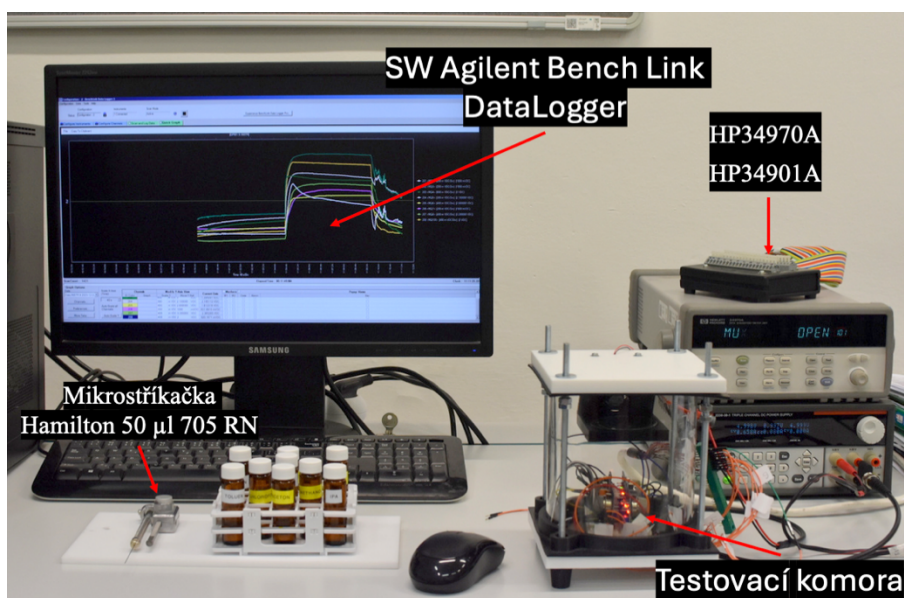
Obrázek 22: Zapojení matice senzorů plynu s použitím nepájivého pole.



Obrázek 23: Zapojení matice senzorů plynu s použitím nepájivého pole a testovací komory o objemu 1,65 l.

Součástí měřicího pracoviště, které je uvedené na Obrázek 24, byly:

- měřicí ústředna HP34970A a multiplexerová karta HP34901A,
- SW Agilent Bench Link DataLogger verze 3.2,
- chemorezistivní senzory MQx,
- mikrostříkačka Hamilton 50 μ l 705 RN pro dávkování VOC,
- testovací komora objemu 1,65 l.



Obrázek 24: První varianta měřicího pracoviště s použitím testovací komory o objemu 1,65 l.

Pro testovací měření bylo zvoleno osm následujících VOC, které byly vyhodnoceny jako dostupné a relativně bezpečné:

- toluen,
- ethanol,
- benzen,
- IPA,
- chloroform,
- aceton,
- xylen,
- methanol.

Tabulka 5 obsahuje seznam změřených těkavých látek včetně jejich chemických vzorců a relativní molekulové hmotnosti M_r , které byly pořízené od společnosti Penta Chemicals Unlimited.

Tabulka 5: Zvolené VOC pro první variantu experimentu.

VOC	Chemický vzorec	$M_r[-]$	Čistota
Toluen	C_7H_8	92,14	p.a.
Ethanol	C_2H_6O	46,07	99.8 % pro UV spektroskopii
Benzen	C_6H_6	78,12	Čistý
IPA	C_3H_8O	60,10	p.a.
Chloroform	$CHCl_3$	119,38	p.a.
Aceton	C_3H_6O	58,08	p.a.
Xylen	C_8H_{10}	106,17	p.a.
Methanol	CH_4O	32,04	p.a.

Pro každou VOC s koncentrací 5 ppm s využitím matice chemorezistivních senzorů bylo provedeno jedno měření při reálných okolních podmínkách. Objem VOC pro cílovou koncentrací 5 ppm byl spočítán podle vztahu (27). V Tabulka 6 je uveden požadovaný objem VOC (V_1) pro cílovou koncentrací 5 ppm (C_2) v 1,65 litrech vzduchu (V_2):

$$C_1 \cdot V_1 = C_2 \cdot V_2 \quad (27)$$

C_1 [ppm] ... počáteční koncentrace látky, která je dána hustotou látky ($g/m^3 = 1 \text{ mg/L} = 1 \text{ ppm}$),

C_2 [ppm] ... požadovaná (cílová) koncentrace látky,

V_1 [l] ... objem látky, který je nutné přidat pro koncentraci C_2 ,

V_2 [l] ... celkový objem.

Tabulka 6: Požadovaný objem VOC pro cílovou koncentrací 5 ppm v 1,65 l vzduchu.

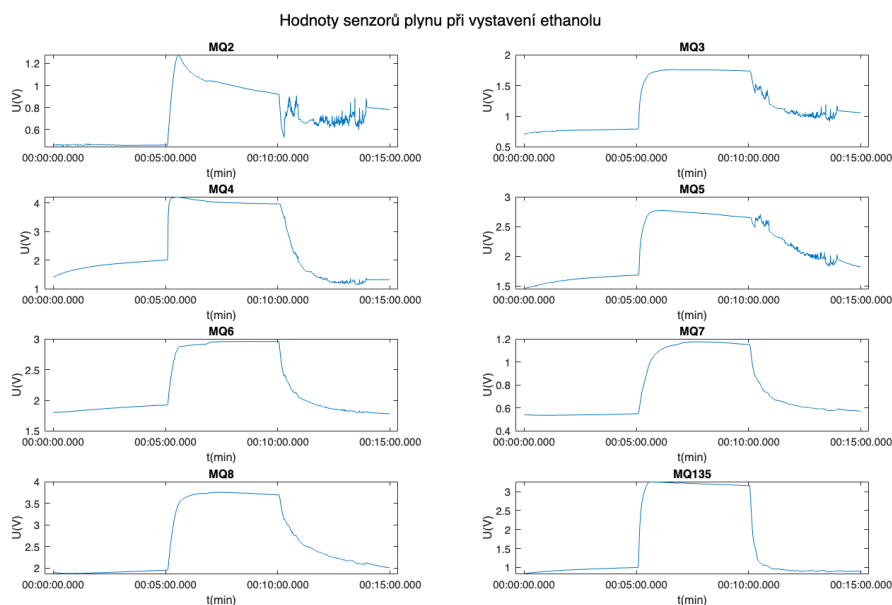
VOC	C_1 (hustota) [$g/m^3 = ppm$]	V_1 [μl]
Toluen	867000	10
Ethanol	789000	10
Benzen	876000	9
IPA	786000	10
Chloroform	1490000	6
Aceton	784000	11
Xylen	860000	10
Methanol	792000	10

Měření probíhalo ve třech fázích, uvedených v Tabulka 7. Před začátkem měření VOC látek se senzory ohřívaly po dobu 15 min. Hodnota napájecího napětí byla nastavena na 5 V. Mezi jednotlivými měření VOC byl časový rozestup 30 min nutný pro obnovu povrchu snímací vrstvy senzorů.

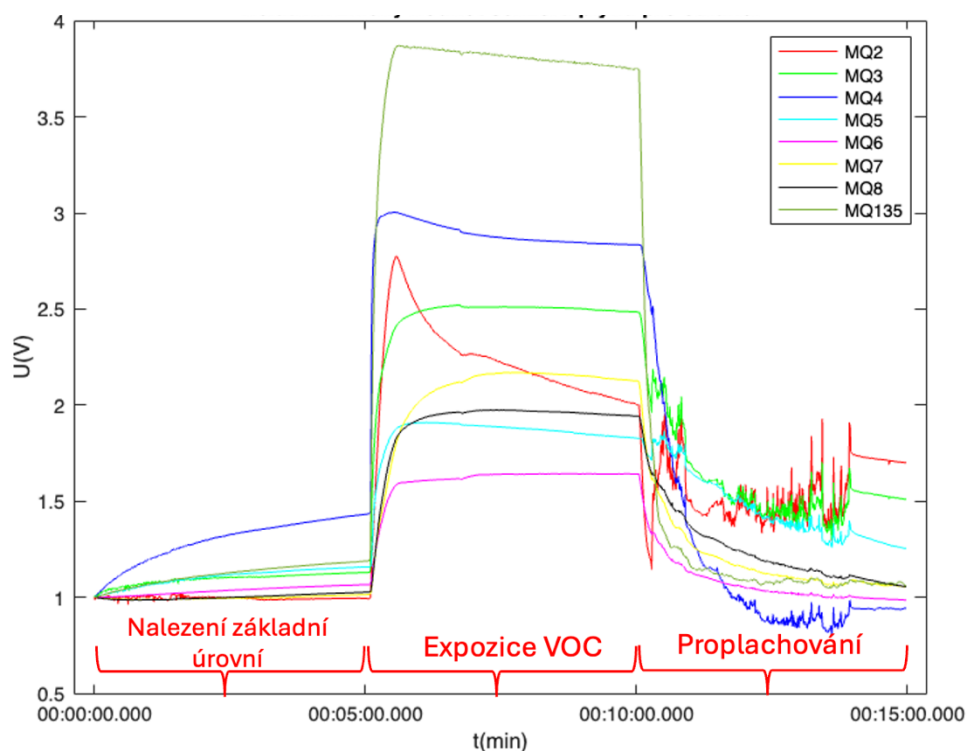
Tabulka 7: Jednotlivé fáze měření VOC.

Fáze	Čas (doba trvání)	Prostředí
0. fáze: nalezení základní úrovně signálu	5 min	Vzduch
1. fáze: expozice	5 min	VOC s koncentrací 5 ppm
2. fáze: proplachování	5 min	Vzduch

Na Obrázek 25 jsou uvedeny průběhy odezvy jednotlivých senzorů plynu při měření ethanolu. Na Obrázek 26 jsou zobrazeny relativní změny hodnot napětí senzorů plynu při expozici ethanolu, kde jsou také označené jednotlivé fáze měření.



Obrázek 25: Průběhy odezvy jednotlivých senzorů plynu při měření ethanolu s použitím první varianty zapojení senzorů a testovací komory objemem 1,65 l.



Obrázek 26: Relativní změny hodnot napětí senzorů plynu při měření ethanolu s použitím první varianty zapojení senzorů s označením jednotlivých fází měření.

V příloze A lze nalézt průběhy odezev senzorů plynu a grafy relativních změn hodnot snímaného napětí senzorů při měření ostatních VOC.

4.2.2. Hodnocení první varianty experimentu

Jednoduchost řešení představuje klíčovou výhodou prvního návrhu experimentu. Realizace zapojení matice senzorů plynu pomocí nepájivého pole je časově méně náročná než realizace zapojení s použitím plošného spoje a neklade požadavky na specifické znalosti v oblasti výroby plošných spojů. Nevýhoda takového zapojení na nepájivém poli spočívá v nízké odolnosti proti rušení a působení vnějších vlivů (z výstupu měření na Obrázek 25 až 26 lze vidět, že průběh je zašuměný), navíc vzhledem k rozměru testovací komory byly senzory blízko sebe a docházelo k vzájemnému ohřevu a k driftu signálu, proto nebylo možné dostat stabilní základní úroveň signálu. Kromě výše uvedených problémů, nebyly dostatečně stabilní kontakty a při manipulaci s komorou docházelo k jejich uvolňování, což může značně ovlivnit reprodukovatelnost měření.

Také bylo zjištěno, že pro fázi proplachování není pět minut dostatečné k tomu, aby se časová odezva senzoru vrátila na hodnotu základní úrovně.

Další nevýhodou první varianty experimentu je překmit časové odezvy některých senzorů, například překmit časové odezvy senzoru MQ2, který nastal při měření toluenu (viz. příloha A). Překmit časové odezvy senzorů při měření určitých VOC by mohl být způsobený dynamikou

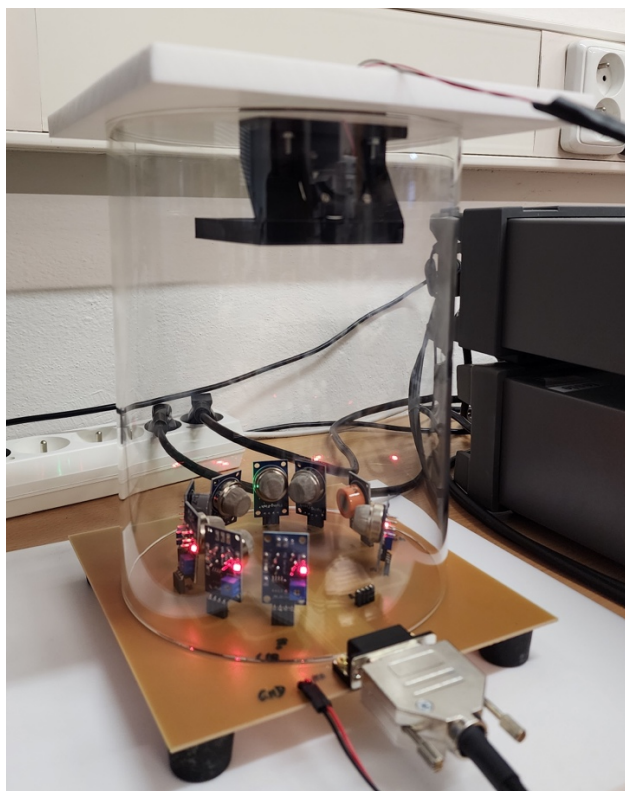
výparu VOC (při přechodu z kapalného do plynného stavu sloučeniny) a interakcí VOC se snímací vrstvou senzoru.

4.2.3. Druhá varianta experimentu

Předchozí zapojení s využitím nepájivého pole mělo určité nevýhody (viz. kapitola č. 4.2.2), zapojení matice senzorů plynu s využitím plošného spoje (viz příloha D) a kabelu se stíněním by mělo odstranit šum a kolísání základní úrovně signálů senzorů a tím zlepšit kvalitu snímaných průběhů odezvy senzorů. Použití komory, která má větší objem, a ventilátoru by mělo vyřešit problém překmitu průběhu odezvy. Na Obrázek 27 je uvedena druhá varianta zapojení matice senzorů plynu, na Obrázek 28 lze vidět zapojení matice senzorů s využitím testovací komory o objemu 4,06 l.



Obrázek 27: Zapojení matice senzorů plynu s použitím plošného spoje.



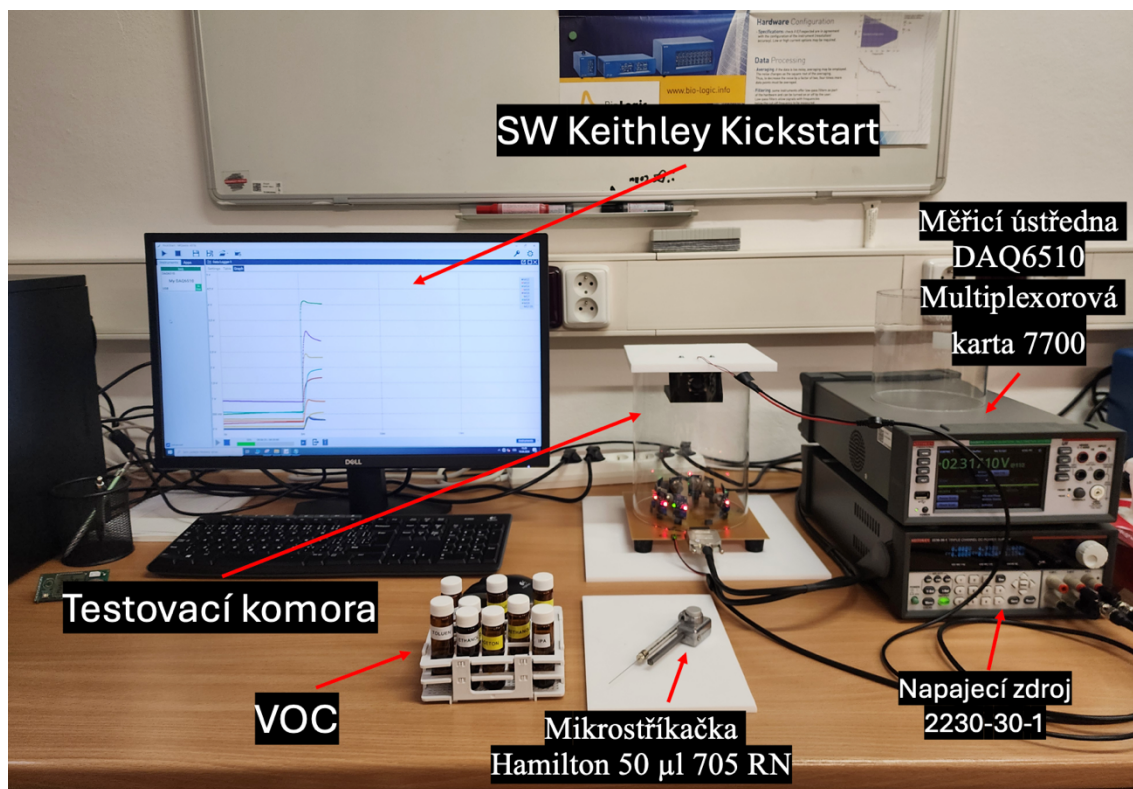
Obrázek 28: Zapojení matice senzorů plynu s použitím plošného spoje a testovací komory o objemu 4,06 l.

S využitím druhé varianty zapojení bylo provedeno výsledné měření zvolených VOC. Naměřená data byla následně zpracována (viz. kapitola č. 5).

Součástí měřicího pracoviště, které je uvedené na Obrázek 29, byly:

- SW Keithley Kickstart verze 2.11.1;
- chemorezistivní senzory MQ1, MQ2, MQ3, MQ4, MQ5, MQ6, MQ7, MQ8, MQ9, MQ135;
- mikrostříkačka Hamilton 50 μ l 705 RN pro dávkování VOC;
- Měřicí ústředna Keithley DAQ 6510 a multiplexorová karta Keithley 7700;
- Testovací komora o objemu 4,06 l;
- Napájecí zdroj Keithley 2230-30-1.

Na Obrázek 29 je uvedená konfigurace měřicího pracoviště.



Obrázek 29: Druhá varianta měřicího pracoviště s použitím testovací komory o objemu 4,06 l.

Pro měření a následné zpracování byly zvoleny následující VOC, které jsou potenciálními biomarkery nemocí:

- Ethanol (C_2H_5OH) je VOC, podle literatury [1], [8] může být potenciálním biomarkerem pro identifikaci kardiovaskulárních nemocí a cukrovky;
- Toluén (C_7H_8) je VOC, podle literatury [1], [5] je potenciální biomarker rakoviny plic a kardiovaskulárních nemocí (infarktu myokardu);
- Aceton (C_3H_6O) je VOC, podle literatury [4], [5] je potenciální biomarker cukrovky.

V rámci této práce bylo rozhodnuto změřit pouze VOC pomocí chemorezistivních senzorů plynu (místo měření VOC a plynů zároveň). Důvodem je to, že měření plynů (např. CO) by vyžadovalo jiný experimentální postup než postup měření VOC, který by zahrnoval kontrolu průtoku plynů, regulaci vlhkosti plynů a použití syntetického vzduchu namísto okolního (přirozeného) vzduchu během fáze nalezení základní úrovně signálu a fáze proplachování. V tomto případě by tím objektivita výsledků klasifikace mohla být zkreslená.

Pro měření a následující zpracování byly zvoleny tři koncentrace VOC: 1 ppm, 5 ppm a 10 ppm. Motivace pro volbu těchto specifických koncentrací vychází ze skutečnosti, že koncentrace biomarkerů v dechu zdravého jedince a nemocného jedince, například

s cukrovkou, se výrazně liší. Například podle literatury [35] se koncentrace acetonu v dechu zdravých lidí pohybuje v rozsahu 300 až 900 ppb, v dechu jedinců, kteří mají cukrovku, koncentrace acetonu může dosahovat 1800 ppb (1,8 ppm) a více.

Pro každou koncentraci těkavých organických látek (VOC) bylo měření provedeno čtyřikrát a to při reálných okolních podmínkách. Tento přístup umožňuje ověřit reprodukovatelnost a přesnost detekce ve variabilním prostředí, což je klíčové pro aplikace v reálném světě (například v rámci POC analyzátoru dechu). Celkem bylo provedeno 36 měření.

V Tabulka 8 až 10 je uveden vypočítaný objem zvolených VOC (V_1) pro cílovou koncentraci 1, 5 a 10 ppm (C_2) v 4,06 l vzduchu (V_2) podle vztahu (27).

Tabulka 8: Požadovaný objem zvolených VOC a plynu pro cílovou koncentraci 1 ppm v 4,06 l vzduchu.

VOC	C_1 (hustota) [g/m ³ = ppm]	V_1 [μl]
Toluen	867000	5
Ethanol	789000	5
Aceton	784000	5

Tabulka 9: Požadovaný objem zvolených VOC a plynu pro cílovou koncentraci 5 ppm v 4,06 l vzduchu.

VOC	C_1 (hustota) [g/m ³ = ppm]	V_1 [μl]
Toluen	867000	23
Ethanol	789000	26
Aceton	784000	26

Tabulka 10: Požadovaný objem zvolených VOC a plynu pro cílovou koncentraci 10 ppm v 4,06 litrech vzduchu.

VOC látka	C_1 (hustota) [g/m ³ = ppm]	V_1 [μl]
Toluen	867000	47
Ethanol	789000	51
Aceton	784000	52

V Tabulka 11 je uvedena tenze par příslušných VOC. Čím větší je tenze par, tím rychleji se VOC odpařuje:

Tabulka 11: Tenze par vybraných VOC.

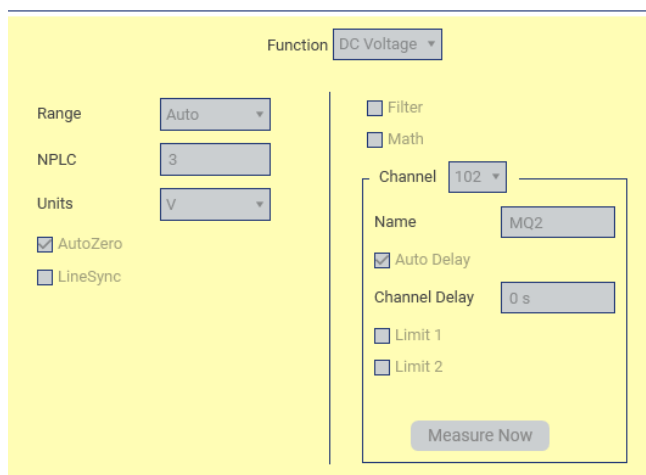
VOC	Tenze par při 20 °C
Ethanol	5,95 kPa
Toluen	2,9 kPa
Aceton	12,36 kPa

Měření probíhalo ve třech fázích, viz Tabulka 12. Stejně jako v první variantě experimentu se před začátkem měření senzory ohřívaly po dobu 15 min, hodnota napájecího napětí byla nastavena na 5 V. Na rozdíl od první varianty experimentu byla doba fáze proplachování stanovena na 10 min, aby se snímaný signál senzorů vrátil na hodnotu základní úrovně. Mezi jednotlivými měřeními stejné VOC byl časový rozestup 15 min nutný pro obnovu povrchu snímací vrstvy senzorů, mezi měřeními různých VOC byl minimální časový rozestup stanoven na 25 min. Pro zajištění konzistentního prostředí a udržení stálého toku vzduchu kolem senzoru během fáze expozice VOC senzorů plynu bylo využito ventilátoru, hodnota proudu byla nastavená na 65 mA. Překmit časové odezvy senzorů, který byl pozorován během první varianty experimentu (viz kapitola 4.2.2) by tím měl být minimalizovaný nebo by se neměl vůbec vyskytovat.

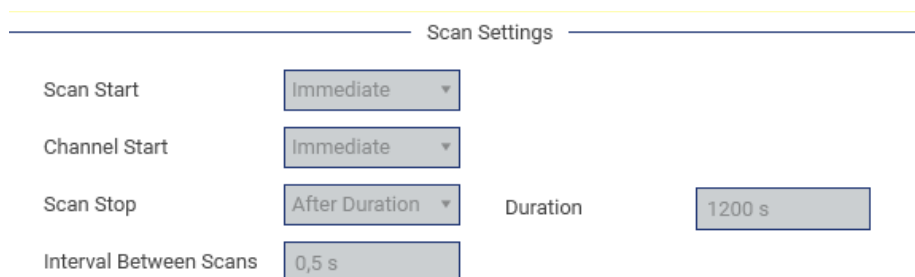
Tabulka 12: Jednotlivé fáze měření VOC (druhá varianta experimentu).

Fáze	Čas (doba trvání)	Prostředí
0. fáze: nalezení základní úrovně signálu	5 min	Vzduch
1. fáze: expozice	5 min	VOC s cílovou koncentrací
2. fáze: proplachování	10 min	Vzduch

Nastavení SW je uvedeno na Obrázek 30 a 31, kde doba integrace NPLC byla nastavená na 3 vzorky, vzorkovací frekvence je 2 Hz.



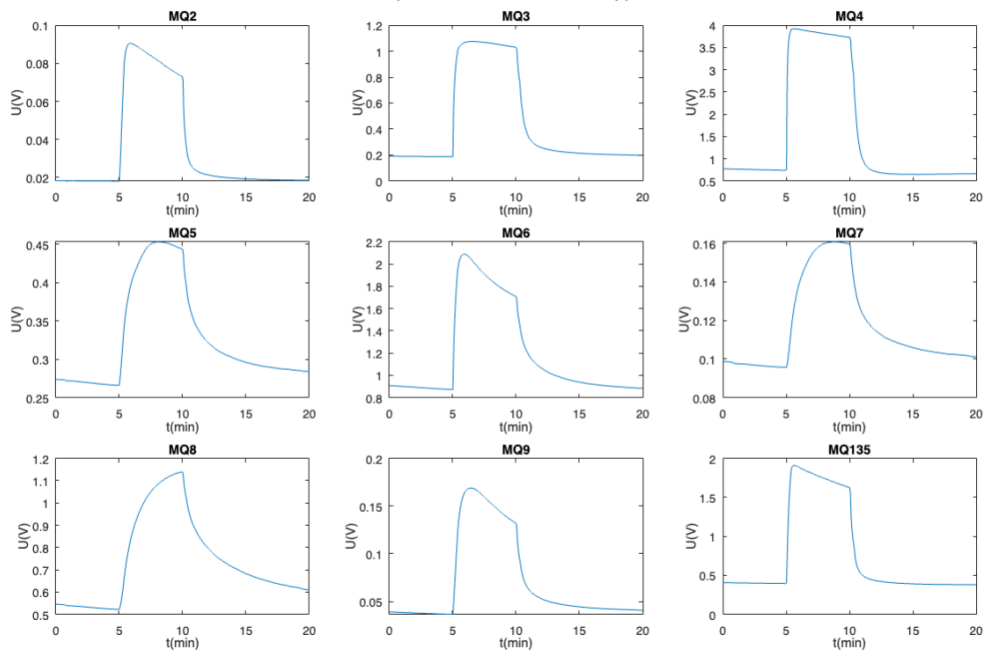
Obrázek 30: Nastavení Keithley Kickstart verze 2.11.1 pro snímané odezvy MQx senzorů.



Obrázek 31: Nastavení vzorkování SW Keithley Kickstart verze 2.11.1 pro snímané odezvy MQx senzorů.

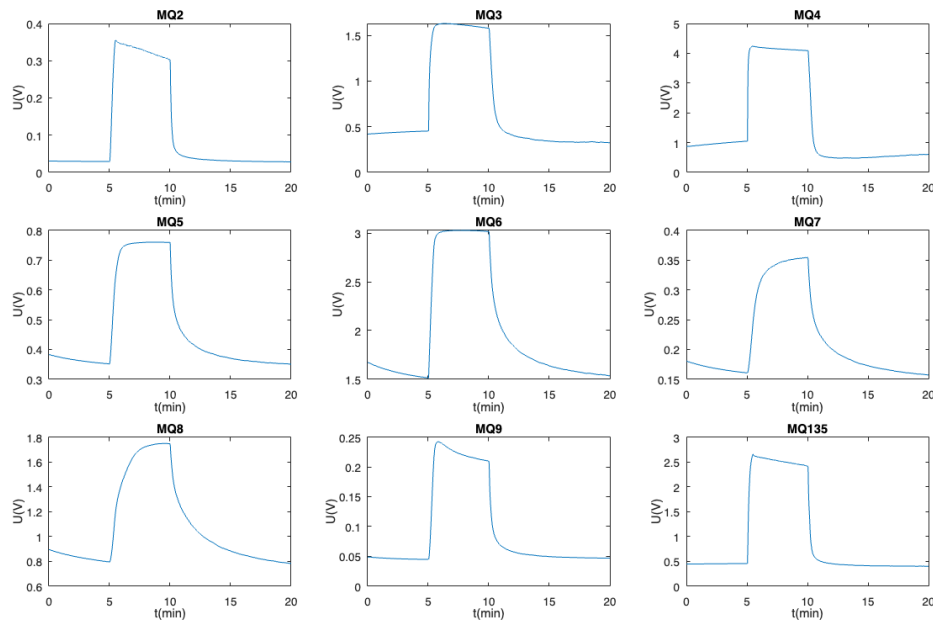
Na Obrázek 32 až 34 jsou uvedeny průběhy odezvy jednotlivých senzorů plynu při 1. měření ethanolu s koncentrací 1, 5 a 10 ppm.

Odezva senzorů pro ethanol s koncentrací 1 ppm 1. měření

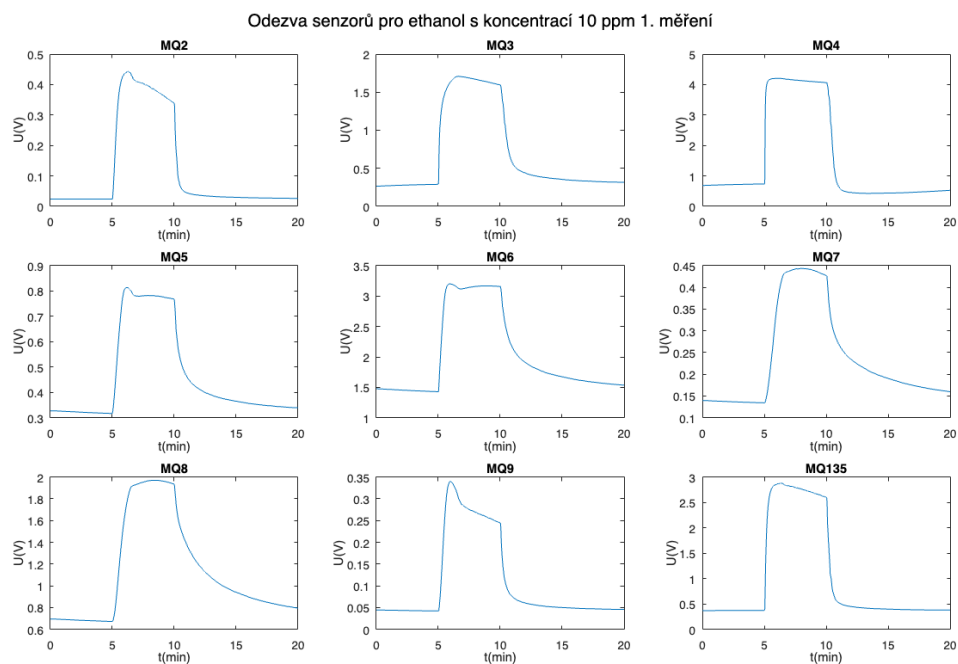


Obrázek 32: Průběhy odezvy jednotlivých senzorů plynu při měření ethanolu s koncentrací 1 ppm s použitím druhé varianty zapojení senzorů a testovací komory objemem 4,06 l.

Odezva senzorů pro ethanol s koncentrací 5 ppm 1. měření

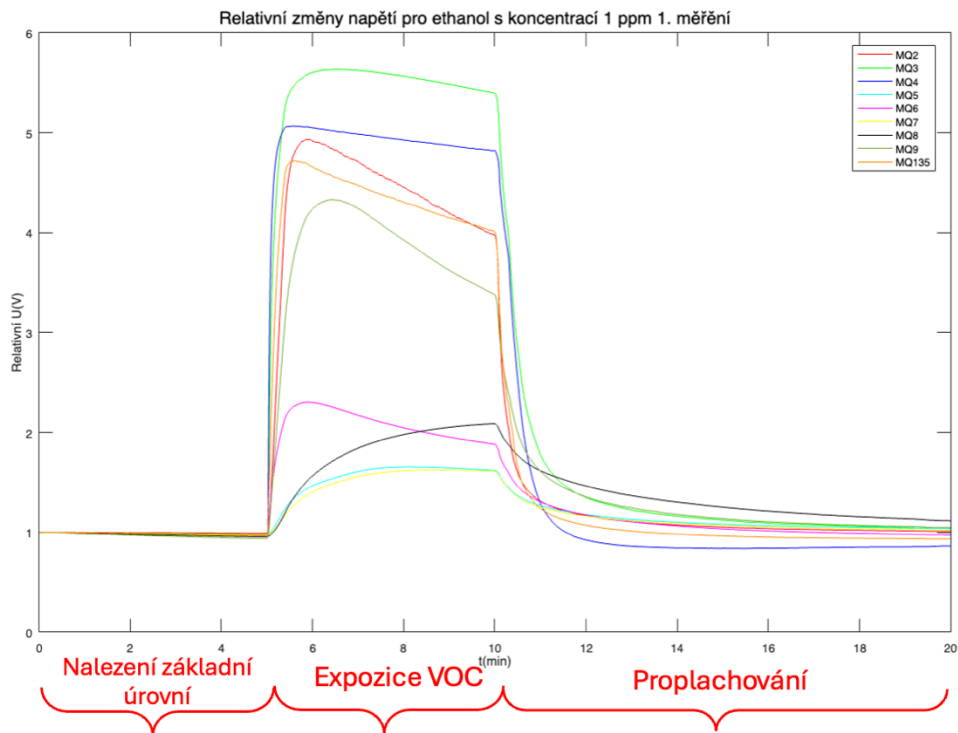


Obrázek 33: Průběhy odezvy jednotlivých senzorů plynu při měření ethanolu s koncentrací 5 ppm s použitím druhé varianty zapojení senzorů a testovací komory objemem 4,06 l.

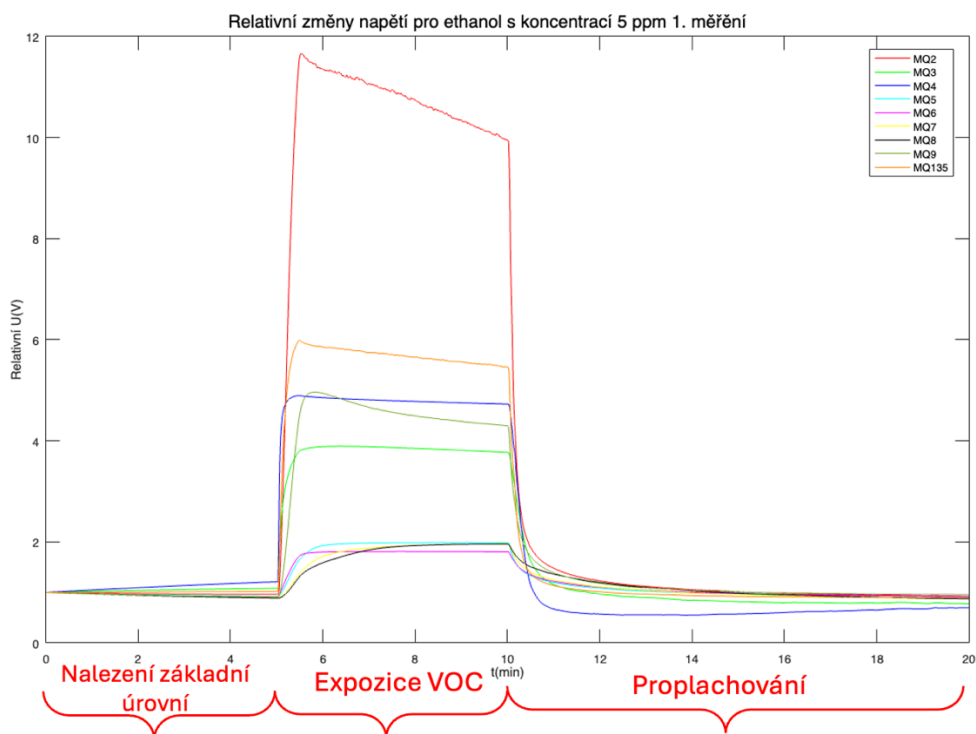


Obrázek 34: Průběhy odezvy jednotlivých senzorů plynu při měření ethanolu s koncentrací 10 ppm s použitím druhé varianty zapojení senzorů a testovací komory objemem 4,06 l.

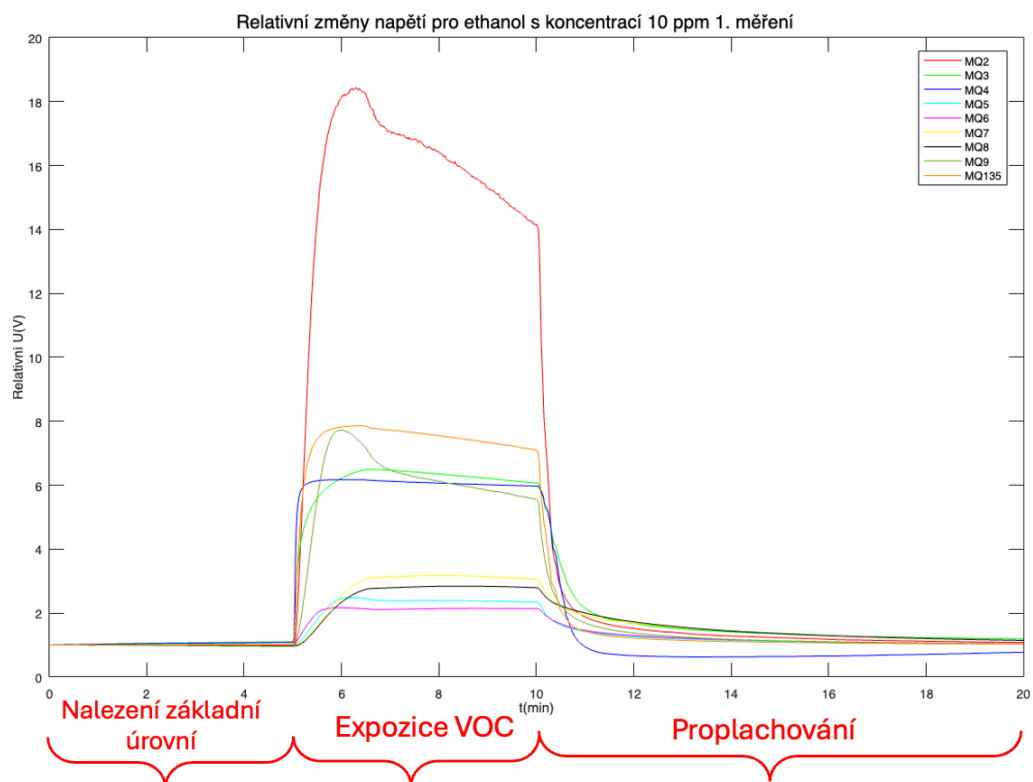
Na Obrázek 35 až 37 jsou zobrazeny relativní změny hodnot napětí senzorů plynu při expozici ethanolu s koncentrací 1, 5 a 10 ppm (1. měření každé koncentrace), kde jsou také označeny jednotlivé fáze měření.



Obrázek 35: Relativní změny hodnot napětí senzorů plynu při měření ethanolu s koncentrací 1 ppm s použitím druhé varianty zapojení senzorů s označením jednotlivých fází měření.

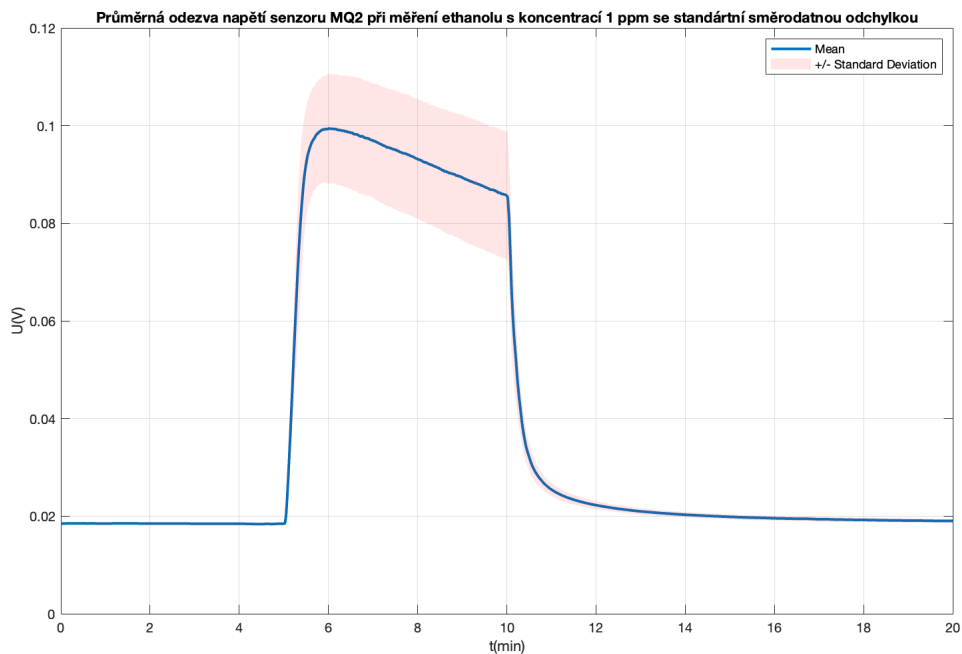


Obrázek 36: Relativní změny hodnot napětí senzorů plynu při měření ethanolu s koncentrací 5 ppm s použitím druhé varianty zapojení senzorů s označením jednotlivých fází měření.

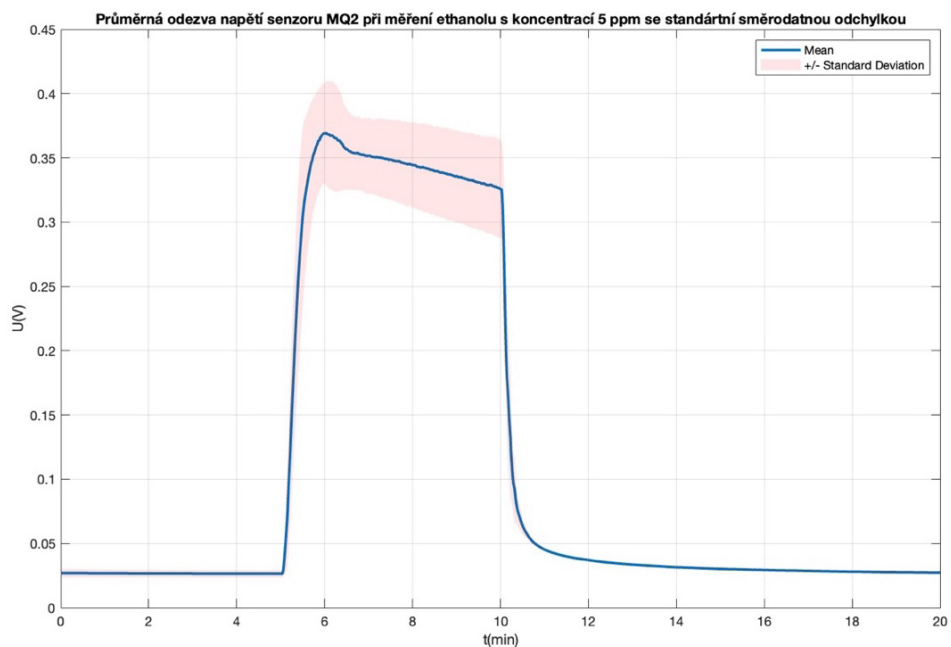


Obrázek 37: Relativní změny hodnot napětí senzorů plynu při měření ethanolu s koncentrací 10 ppm s použitím druhé varianty zapojení senzorů s označením jednotlivých fází měření.

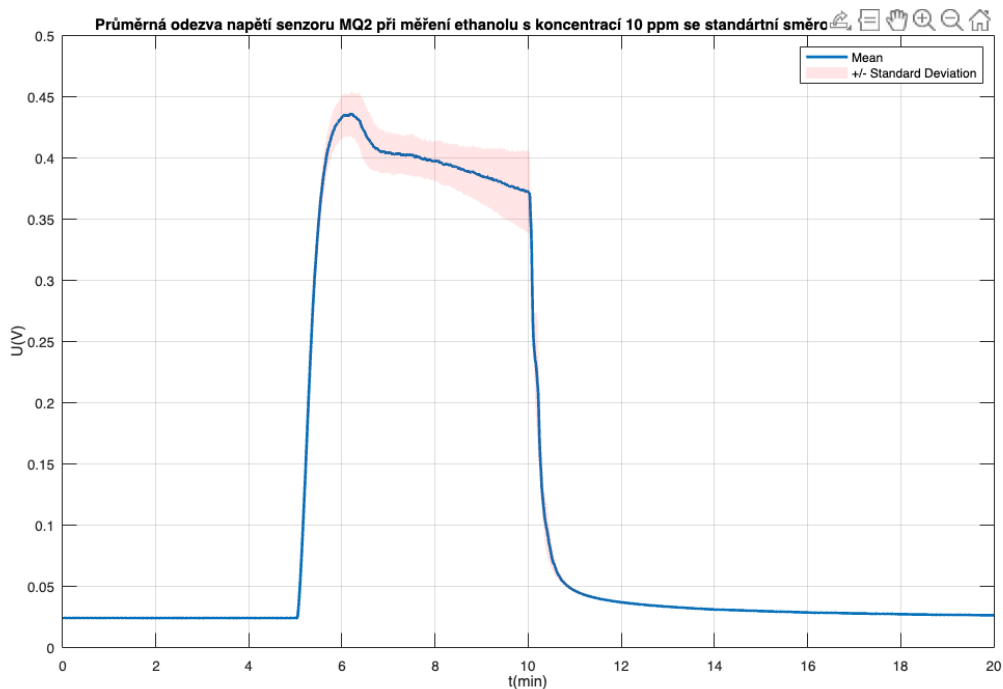
Na Obrázek 38 až 40 je zobrazen průměrný časový průběh odezvy senzoru MQ2 se standardní směrodatnou odchylkou během čtyř opakování měření ethanolu s koncentrací 1, 5 a 10 ppm. Průběhy odezvy senzorů plynu, relativní změny časové odezvy snímané během dalších opakování měření ethanolu a během měření ostatních VOC jsou uvedené v příloze B. Průměrné odezvy senzoru MQ2 se standardní směrodatnou odchylkou při měření dalších VOC jsou také uvedené v příloze B.



Obrázek 38: Průměrný časový průběh odezvy senzoru MQ2 se standardní směrodatnou odchylkou během čtyř měření ethanolu s koncentrací 1 ppm.



Obrázek 39: Průměrný časový průběh odezvy senzoru MQ2 se standardní směrodatnou odchylkou během čtyř měření ethanolu s koncentrací 5 ppm.



Obrázek 40: Průměrný časový průběh odezvy senzoru MQ2 se standardní směrodatnou odchylkou během čtyř měření ethanolu s koncentrací 10 ppm.

4.2.4. Hodnocení druhé varianty experimentu

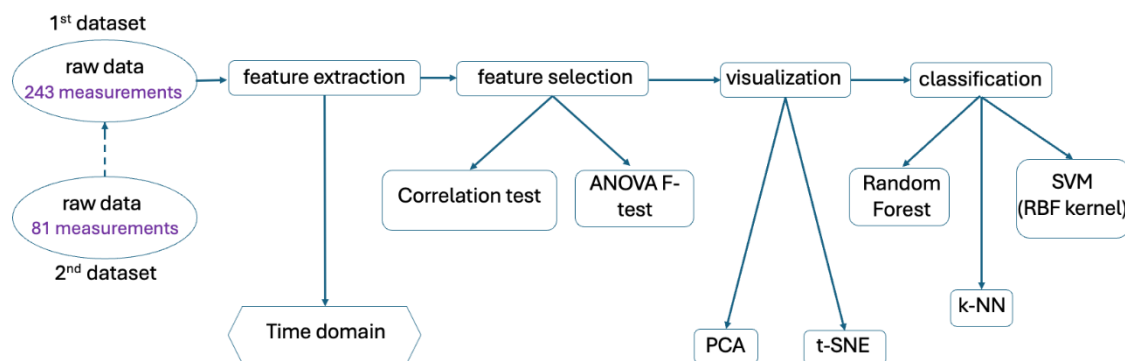
Podle Obrázek 36 a 37 lze vidět, že se nepodařilo zcela zbavit překmitu časové odezvy senzoru, jak bylo zmíněno v kapitole č. 4.2.2, překmit by mohl být způsobený dynamikou výparu VOC. Také podle Obrázek 36 a 37 vyplývá, že se překmit zvyšuje se zvyšující se koncentrací VOC, což souvisí s metodou dávkování VOC pomocí mikrostřičkačky (vyšší koncentrace vyžaduje větší kapku, což vede k lokálnímu zvýšení koncentrace a větší tvorbě tzv. oblaku VOC). Zvětšení objemu testovací komory a použití ventilátoru na to nemělo zásadní vliv. Ačkoliv v tomto případě se hledalo z cenového hlediska nejdostupnější komerční řešení. Použití jiné konfigurace komory (výměna pořadí senzorů, změna umístění, otáček či rozměru ventilátoru) nebo použití membrány by mohlo vyřešit nežádoucí překmit časové odezvy. Úplná eliminace nežádoucích překmitů by mohla být dosažena použitím oddělovací membrány anebo použitím jiné varianty konstrukce senzorů, např. miniaturizované matice senzorů.

Použití plošného spoje a stínění kabelu vedlo k větší stabilitě měření, odolnosti proti rušení, eliminaci driftu signálu a lepší reprodukovatelnosti, viz Obrázek 38 až 40.

Také lze udělat závěr, že prodloužení fáze proplachování po dobu 10 min bylo korektní, časová odezva většiny senzorů se vrátila na hodnotu základní úrovně.

5. Zpracování dat

Na Obrázek 41 je uveden postup zpracování dat získaných pomocí matice senzorů plynu pro identifikaci VOC.



Obrázek 41: Postup zpracování dat naměřených pomocí matice MQ senzorů plynu pro identifikaci VOC.

Pro zpracování naměřených dat bylo zvoleno programovací prostředí MATLAB verze R2023b. Byly použité následující toolboxy (knihovny):

- Statistics and Machine Learning Toolbox,
- Signal Processing Toolbox.

Za účelem zkoumání parazitních vlivů teploty a vlhkosti na snímané signály senzorů a tím negativní dopad na následnou klasifikaci VOC bylo rozhodnuto provést klasifikaci pro menší a větší soubor dat, viz Obrázek 41. Menší datová sada obsahuje 27 měření VOC pomocí devíti chemorezistivních senzorů plynu (celkem 243 měření), kde pro každou ze třech VOC s koncentrací 1, 5 a 10 ppm byla provedena tři měření. Měření probíhalo za okolních podmínek, kde se teplota pohybovala v rozmezí 21 až 23 °C, vlhkost okolního vzduchu se pohybovala v rozmezí 30 až 43 % RH. Větší datový soubor se skládá z 36 měření VOC (kde 27 měření jsou z menšího souboru dat) pomocí devíti chemorezistivních senzorů plynu (celkem 324 měření), kde pro každou ze třech VOC s koncentrací 1, 5 a 10 ppm byla provedena čtyři měření. Měření probíhalo za reálných okolních podmínek, kde se teplota pohybovala v rozmezí 21 až 25 °C, vlhkost okolního vzduchu se pohybovala v rozmezí 30 až 55 % RH. Podle grafů relativních změn napětí senzorů, které lze nalézt v příloze B, je vidět, že snímané časové odezvy senzorů během čtvrtého měření každé VOC pro všechny koncentrace jsou poměrně odlišné v porovnání s časovými odezvami senzorů z předchozích měření. Jedním z možných důvodů tohoto rozdílu je větší vlhkost a teplota okolního prostředí, za kterých byla provedena čtvrtého měření každé VOC.

Výsledky selekce parametrů a vizualizace většího souboru dat lze nalézt v příloze C. Výsledky klasifikace menšího souboru dat a většího souboru dat včetně porovnání odhadů výkonu různých klasifikačních modelů jsou uvedeny v kapitole 5.5.

V následujících částech této práce jsou diskutované pojmy hladina statistické významnosti α , chyby I. a II. typu, také tzv. problém násobného testování (angl. multiple significance tests). Definice těchto pojmů lze nalézt v literatuře [36].

5.1. Extrakce parametrů

Parametry, které lze extrahovat ze surových dat získaných pomocí senzorů plynu, byly podrobněji rozebírány v kapitole č. 2.4. V praktické části této práce byly pro vizualizaci a klasifikaci použité parametry z časové oblasti. Důvodem byla snadnost interpretace těchto parametrů. Následně se ukázalo, že pro řešení dané klasifikační úlohy to bylo postačující.

Z časové oblasti byly extrahované následující parametry:

- **Citlivost**

Citlivost senzoru (viz kapitola č. 2.4) byla vypočítaná podle vztahu (28):

$$S = \frac{0,9 \cdot U_g - U_a}{U_a} \quad (28)$$

S [-] ... citlivost chemorezistivního senzoru,

U_g [V] ... maximální hodnota napětí senzoru v přítomnosti analytu,

U_a [V] ... maximální hodnota napětí senzoru v přítomnosti čistého vzduchu.

- **Doba náběhu (diference)**

Doba difference je definována jako časový interval mezi zahájením expozice a okamžikem, kdy napětí dosahuje svého (prvního) maxima.

$$t_{dif} = t(U_g) - t_e \quad (29)$$

t_{dif} [s] ... doba difference,

$t(U_g)$ [s] ... čas, kdy napětí senzoru dosahuje svého (prvního) maxima,

t_e [s] ... doba zahájení fáze expozice.

- **Plocha pod křivkou**

Tento parametr je určen jako celková plocha, kterou ohraničuje časový průběh odezvy během fáze expozice a fáze proplachování.

$$A = \int_{t_e}^{t_f} U(t) dt \quad (30)$$

A [V·s] ... celková plocha ohraničená časovým průběhem odezvy,

t_e [s] ... doba zahájení fáze expozice,

t_f [s] ... doba ukončení fáze proplachování,

U [V] ... snímané napětí senzoru.

- **Energie**

Energie signálu je definována jako integrál kvadrátu časového průběhu odezvy během fáze expozice a fáze proplachování.

$$E = \int_{t_e}^{t_f} U(t)^2 dt \quad (31)$$

E [V²·s] ... energie signálu chemorezistivního senzoru,

t_e [s] ... doba zahájení fáze expozice,

t_f [s] ... doba ukončení fáze proplachování,

U [V] ... snímané napětí senzoru.

- **Nejrychlejší doba náběhu napětí**

Nejrychlejší doba náběhu snímaného napětí senzoru je vypočítaná jako časový interval, ve kterém dochází k maximální změně napětí za jednotku času.

$$\Delta t_{max} = \arg \max_{t \in [t_e, t_f]} \left(\frac{dU}{dt} \right) \quad (32)$$

Δt_{max} [s] ... nejrychlejší doba náběhu snímaného napětí,

U [V] ... snímané napětí senzoru,

t_e [s] ... doba zahájení fáze expozice,

t_f [s] ... doba ukončení fáze proplachování.

- **Variabilita signálu**

Variabilita signálu je dána standardní směrodatnou odchylkou snímaného napětí senzoru (časové odezvy) během fáze expozice a fáze proplachování.

$$\sigma_U = \sqrt{\frac{1}{N-1} \sum_{i=i_e}^{i_e+N-1} (U_i - \bar{U})^2} \quad (33)$$

σ_U [V] ... standardní směrodatná odchylka snímaného signálu napětí,

N [-] ... celkový počet vzorků snímaných během fáze expozice a fáze proplachování,

i_e [-] ... index prvního vzorku napětí v souboru dat po zahájení fáze expozice,

U_i [V] ... i -tý vzorek napětí v souboru dat,

\bar{U} [V] ... průměrná hodnota napětí od začátku fáze expozice.

Podle článku [37] byl také vyzkoušen postup extrakce parametrů z frekvenční oblasti. Pomocí rychlé Fourierove transformace (FFT) bylo získáno prvních pět frekvenčních složek výkonové spektrální hustoty. Následně se ukázalo, že na naměřených datech získané (extrahované) parametry z frekvenční oblasti vykazují silnou multikolinearitu (viz příloha C). Tyto parametry pro další analýzu nebyly použité, proto detailní postup této metody nebude v této práci uveden.

5.2. Selekce parametrů

Celkem bylo z časové oblasti extrahováno šest parametrů. Pro snížení redundance a zvýšení výkonu klasifikačních modelů před klasifikací byla provedena selekce parametrů. V této části budou uvedené výsledky pro menší soubor dat, jak bylo zmíněno výše, výsledky selekce parametrů pro větší soubor dat lze nalézt v příloze C.

5.2.1. Test korelace

Ověření normality

Pro správnou volbu testu korelace je nutné ověřit, jestli parametry mají normální rozdělení. Pro tento účel byl proveden jednovýběrový Kolmogorov–Smirnov test normality pomocí funkce *kstest()* (více ohledně tohoto testu normality a funkce *kstest()* lze nalézt v literatuře [38]). Testovala se alternativní hypotéza H_A na hladině statistické významnosti $\alpha = 5\%$.

Nulová hypotéza je $H_0 = \text{„Parametr má normální rozdělení“}$.

Alternativní hypotéza je pak $H_A = \text{„Parametr nemá normální rozdělení“}$.

Tím, že byl test normality proveden pro každý parametr, mohl vzniknout tzv. problém násobného testování. Aby se pravděpodobnost chyby I. typu nezvýšila, byla provedena korekce hladiny statistické významnosti α pomocí Bonferroniho korekce podle (34):

$$\alpha_{korig} = \frac{\alpha}{n} = \frac{0,05}{6} = 0,008 \quad (34)$$

α ... původní hladina statistické významnosti,

α_{korig} ... korigovaná hladina statistické významnosti,

n ... počet provedených testů.

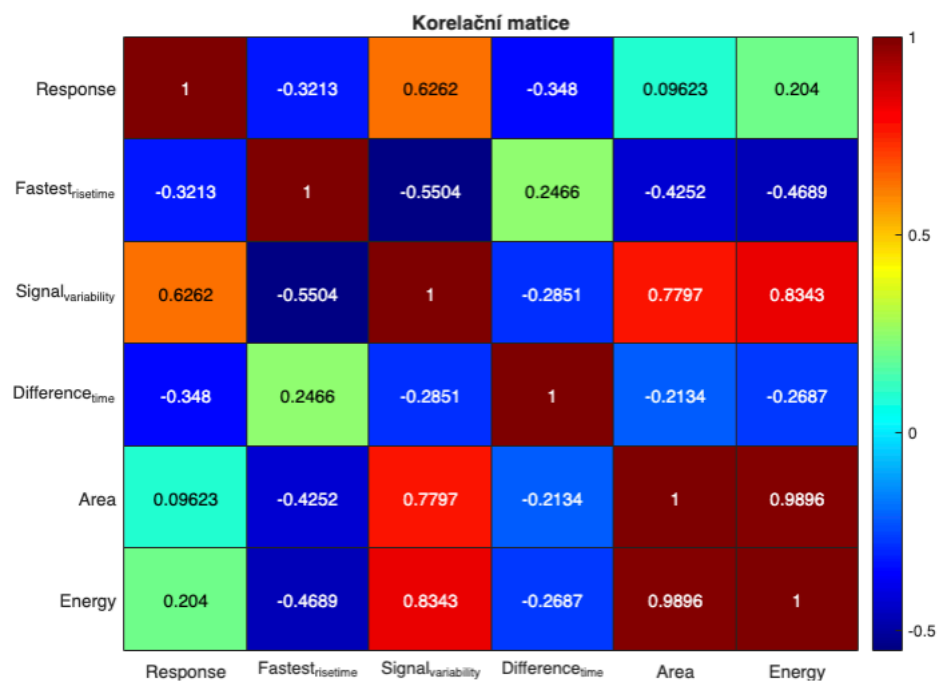
Pro všechny parametry extrahované z časové oblasti platí alternativní hypotéza a z toho vyplývá, že parametry nemají normální rozdělení. Výsledky testu normality jsou uvedené v Tabulka 13.

Tabulka 13: Výsledky Kolmogorov–Smirnov testu normality.

Parametr	p-hodnota	Platná hypotéza
Citlivost	<0,001	H_A
Nejrychlejší doba náběhu napětí	<0,001	H_A
Variabilita signálu	<0,001	H_A
Doba difference	<0,001	H_A
Plocha pod křivkou	<0,001	H_A
Energie signálu	<0,001	H_A

Test korelace

Aby se předešlo možnému zkreslení výsledků další analýzy, před provedením selekce parametrů pomocí jednosměrné ANOVA (viz kapitola 5.2.2) byl proveden test korelace. Vzhledem k tomu, že parametry nemají normální rozdělení, byl použit neparametrický Spearmanův korelační koeficient r_s . Výsledky testování jsou uvedené na Obrázek 42.



Obrázek 42: Korelační matice pro extrahované parametry z časové oblasti s využitím Spearmanova korelačního koeficientu.

Z každého páru parametrů, pro který hodnota korelačního koeficientu $|r_s|$ přesáhla 0,5, byl odstraněn jeden z parametrů, aby se eliminovala multikolinearita a zvýšila se tak spolehlivost další analýzy a klasifikace. Odstranění redundantních proměnných je jedním z klíčových kroků pro dosažení dobré generalizace modelu na neznámých (nových) datech.

V Tabulka 14 jsou uvedené páry parametrů, které mají Spearmanův korelační koeficient $|r_s|$ větší než 0,5.

Tabulka 14: Silně korelované páry parametrů.

Pár parametrů	Spearmanův korelační koeficient r_s
Citlivost a variabilita signálu	0,63
Nejrychlejší doba náběhu napětí a variabilita signálu	-0,55
Variabilita signálu a plocha pod křivkou	0,78
Variabilita signálu a energie signálu	0,83
Plocha pod křivkou a energie signálu	0,99

Parametry variabilita signálu a plocha pod křivkou vykazují silnou multikolinearitu, následně, jak bylo zmíněno výše, byly tyto parametry z analýzy odstraněny.

5.2.2. Jednosměrná ANOVA

Ověření platnosti předpokladů:

1. Nezávislost vzorků

Za předpokladu, že experiment byl správně navržen a proveden, předpoklad nezávislosti vzorků platí.

2. Homogenita rozptylu (homoskedasticita)

Podle literatury [39] jestli poměr skupin je menší než 1,5, porušení předpokladu homogenity rozptylu se neprojeví na výsledcích ANOVA F-testu. Experiment byl navržen a proveden tak, že všechny skupiny (VOC) jsou zastoupené stejným počtem vzorků (a tím je poměr skupin 1), tudíž předpoklad homogenity rozptylu lze v tomto případě zanedbat.

3. Normalita

Pro ověření platnosti předpokladů normality byl proveden Kolmogorov–Smirnov test normality pro všechny parametry napříč všemi skupinami (VOC). Aby se pravděpodobnost chyby I. typu nezvýšila, byla také provedena korekce hladiny statistické významnosti α pomocí

Bonferroniho korekce. Téměř pro všechny parametry napříč všemi skupinami (VOC) tento předpoklad neplatí (viz příloha C), většina parametrů nepochází z normálního rozdělení.

Existuje neparametrická varianta Kruskal-Wallis test, avšak ten má menší statistickou sílu v porovnání s ANOVA (větší pravděpodobnost chyby II. typu). Přestože jeden předpoklad jednosměrné ANOVA neplatí (předpoklad normality dat) a tím se zvyšuje pravděpodobnost chyby I. typu, pro selekci příznaků bylo zvoleno provést jednosměrnou ANOVA, protože má větší statistickou sílu v porovnání s neparametrickým Kruskal-Wallis testem a je obecně považovaná za poměrně odolný test vůči porušení předpokladů.

Pro analýzu vlivu jednotlivých prediktorů byla aplikována jednosměrná ANOVA. V Tabulka 15 jsou uvedené nalezené statisticky významné parametry. Selekcce parametrů se prováděla na základě předpokladu, který byl popsán v kapitole č. 3.2.2, že parametry s větší F-hodnotou jsou víc užitečné pro klasifikaci.

Testovala se nulová hypotéza H_0 :

$H_0 =$ „*Variance mezi skupinami není statisticky významně ovlivněná změnami daného prediktoru*“.

V případě, že je p-hodnota menší než hladina statistické významnosti, nulová hypotéza je zamítnutá. V tomto případě platí, že daný parametr (prediktor) má vliv na závislou proměnnou (skupinu neboli VOC).

Byly zvolené parametry, jejichž p-hodnota je menší než hladina statistické významnosti α a F-statistika ($F(df_{columns}, df_{error})$), kde $df_{columns}$ je počet stupňů volnosti mezi skupinami, je dána vztahem $df_{columns} = \text{počet skupin} - 1$, df_{error} je počet stupňů volnosti uvnitř skupin, je dáno vztahem $df_{error} = \text{počet vzorků} - \text{počet skupin}$ větší než pět (empirický nastavený práh).

Tím, že jednosměrná ANOVA byla provedena zvlášť pro každý parametr, mohl vzniknou tzv. problém násobného testování. Aby se pravděpodobnost chyby I. typu nezvýšila, byla také provedena korekce hladiny statistické významnosti α pomocí Bonferroniho korekce podle (35):

$$\alpha_{korig} = \frac{\alpha}{n} = \frac{0,05}{4} = 0,0125 = 1,25\% \quad (35)$$

α ... původní hladina statistické významnosti,

α_{korig} ... korigovaná hladina statistické významnosti,

n ... počet provedených testů.

Tabulka 15: Vybrané parametry pomocí jednosměrné ANOVA.

Parametr	p-hodnota	F($df_{columns}, df_{error}$)
Citlivost	<0,001	37(2, 240)
Nejrychlejší doba náběhu signálu	0,001	7(2, 240)
Energie signálu	0,003	6(2, 240)

5.3. Vizualizace

V této kapitole budou uvedené výsledky vizualizace pro menší soubor dat. Vizualizace parametrů pro větší soubor dat lze nalézt v příloze C.

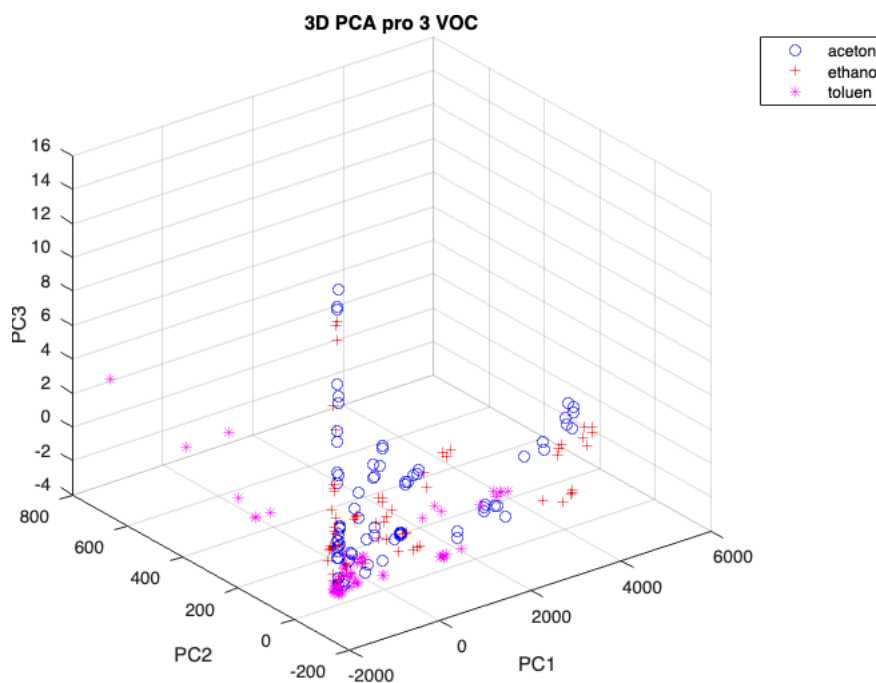
Pro vizualizaci byly použity následující parametry, které byly zvoleny po testu korelace a selekce parametrů pomocí jednosměrné ANOVA:

- citlivost,
- nejrychlejší doba náběhu napětí,
- energie signálu.

Vizualizace byla provedena pomocí lineární metody redukce dimenze PCA a nelineární metody vizualizace multidimenzionálního prostoru t-SNE.

5.3.1. PCA

Na Obrázek 43 je uvedena vizualizace změřených VOC (aceton, ethanol a toluen) s koncentrací 1 ppm, 5 ppm a 10 ppm pomocí metody PCA ve třech dimenzích. Lze vidět, že se data neshlukují do třech shluků. Je možné udělat závěr, že v tomto případě metoda PCA není vhodná pro vizualizaci.

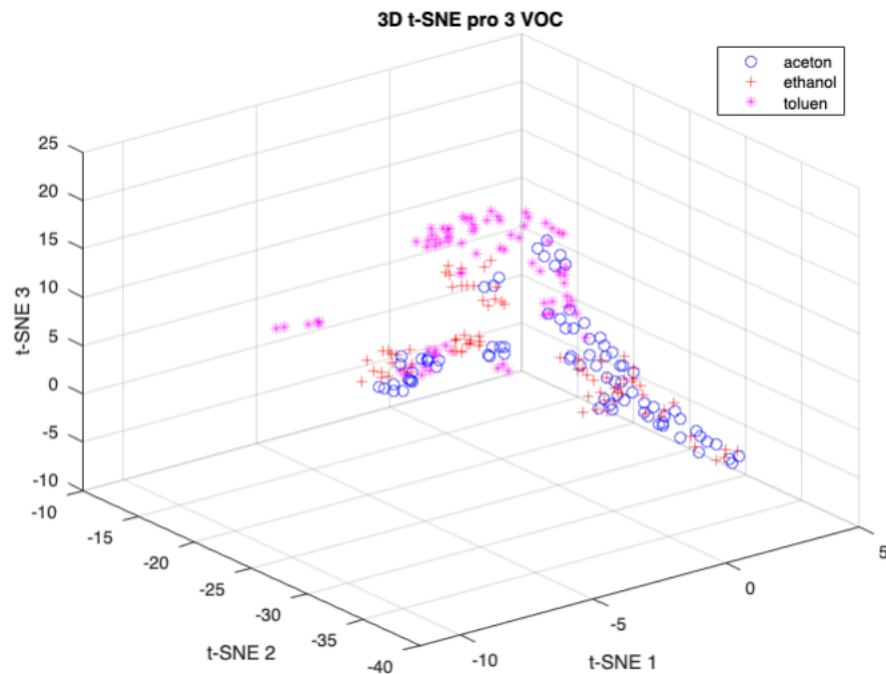


Obrázek 43: Vizualizace acetonu, ethanolu a toluenu pomocí PCA.

5.3.2. t-SNE

Pro vizualizaci naměřených dat byla také použita metoda t-SNE. Jak bylo uvedeno v kapitole 3.3.2. metoda t-SNE má hyperparametr „perplexita“. Hodnota hyperparametru „perplexita“ byla empiricky zvolena a nastavena na 35 (byly vyzkoušeny hodnoty od 5 až 40 s krokem 5). Pro dosažení stabilního výsledku byl nastaven maximální počet iterací na 5000, avšak výsledek není stabilní. Parametry byly standardizované pomocí z-scóre.

Na Obrázek 44 je uveden výsledek vizualizace pomocí metody t-SNE. Na Obrázek 44 lze vidět, že data tvoří tři shluky, ale nejsou separátní. Přesto je výstup přehlednější a použití metody t-SNE pro vizualizaci multidimenzionálního prostoru je v tomto případě vhodnější oproti metodě PCA.



Obrázek 44: Vizualizace acetonu, ethanolu a toluenu pomocí t-SNE, hodnota hyperparametru „perplexita“ je 35.

5.4. Klasifikace

5.4.1. Volba CV

Správná volba metody odhadu výkonu modelu je zásadní krok, který ovlivní celou analýzu. Pro větší datové sady je korektní rozdělit datový soubor na trénovací (např. 80 % dat) a testovací sadu (např. 20 % dat). Trénování modelu (validace a případná optimalizace hyperparametrů) se provádí na trénovací sadě za použití metody CV (nejčastěji k -násobné CV), poté se výsledný odhad výkonu modelu počítá na testovací sadě. Tento postup poskytuje nezkreslený odhad výkonu modelu strojového učení.

Datové soubory, které byly použité pro analýzu a klasifikaci, se skládají celkem z 243 a 324 měření, ze kterých byly následně extrahované parametry použité pro klasifikaci. Vzhledem k tomu, že velikost souborů dat je relativně malá, bylo považováno za nevhodné rozdělení dat na trénovací a testovací sadu. Tento přístup by zmenšil původní již tak malé sady dat použité pro trénování modelu a následkem by mohlo být nedostatečné přizpůsobení modelu strojového učení (angl. underfitting).

Pro ladění hyperparametrů modelů a odhad výkonu modelů byla zvolena tzv. nested CV (možným překladem by mohlo být vnořená CV, viz kapitola č. 3.5), která je vhodná pro malé datové soubory, protože ve vnitřní smyčce CV optimalizuje hodnoty hyperparametrů, ve vnější smyčce počítá nezkreslený odhad výkonu modelu. Byla zvolená hodnota počtu podmnožin (tzv. foldů) pro vnější smyčku CV $k = 10$, hodnota počtu podmnožin (tzv. foldů) pro vnitřní smyčku CV $j = 10$.

Vzhledem k tomu, že v datových souborech je rovnoměrné zastoupení skupin (devět měření pomocí devíti senzorů pro každou VOC pro menší datovou sadu, případně 12 měření pomocí devíti senzorů pro každou VOC pro větší datovou sadu), bylo vhodné zvolit jako metriku výkonu přesnost modelu (viz kapitola č. 3.5.1), která v tomto případě udává nezkreslený odhad výkonu modelu.

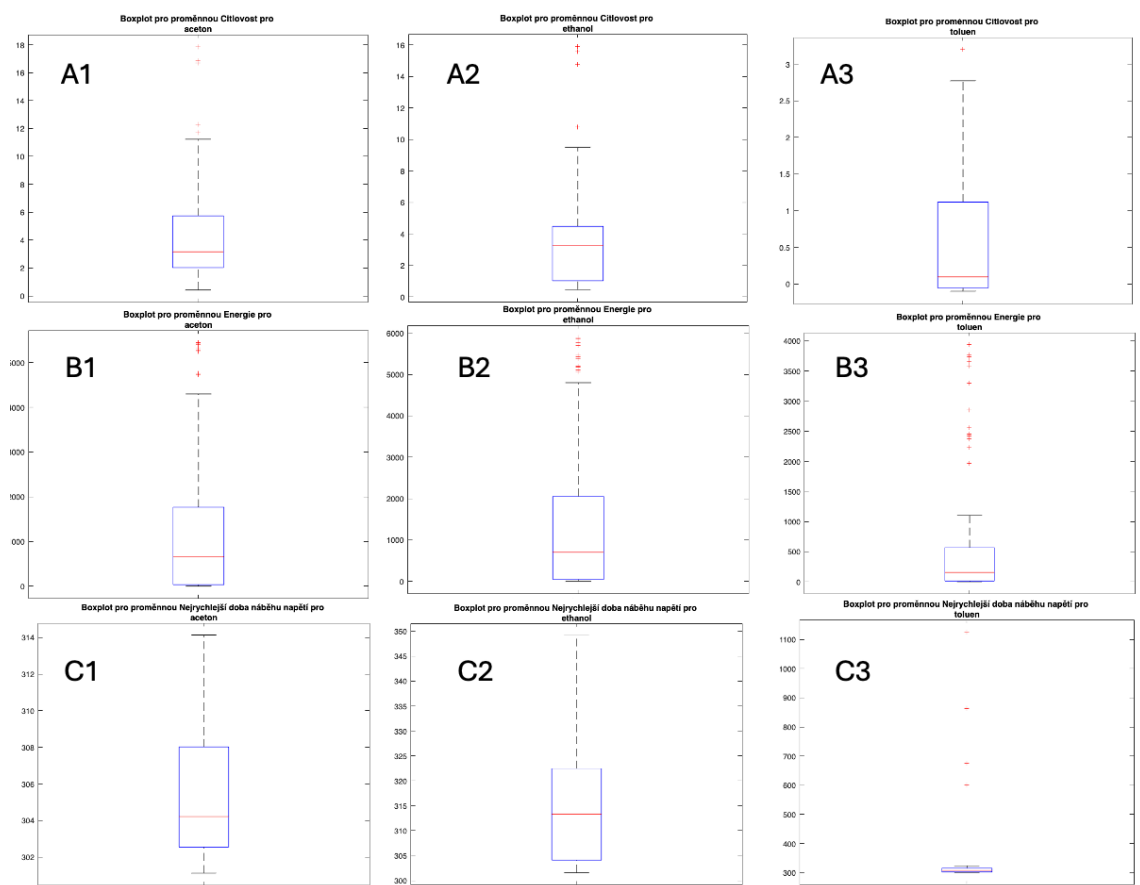
Pro větší přehlednost byla ve vnější smyčce CV vypočítána matice záměn (angl. confusion matrix) a průměrné makro F1-score (tato metrika výkonu modelu je vhodná i v případě nevyváženého souboru dat).

5.4.2. Porovnání modelů

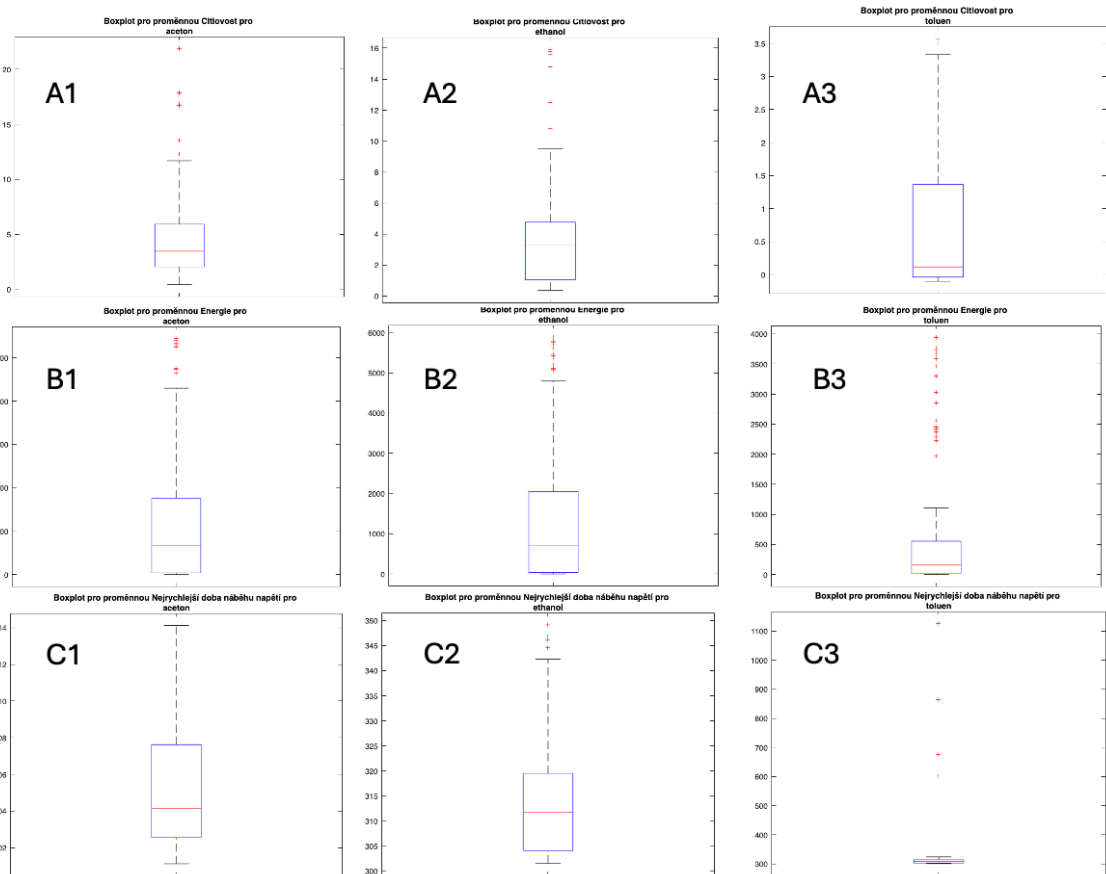
Po provedení selekce parametrů pomocí testu korelace a jednosměrné ANOVA byly pro klasifikaci zvoleny následující parametry:

- citlivost,
- nejrychlejší doba náběhu signálu,
- energie signálu.

Pro správnou volbu klasifikačních algoritmů byla provedena explorační analýza dat. Na Obrázek 45 a 46 jsou uvedené krabicové grafy (někdy se také používá pojem boxplot) každé proměnné pro každou VOC.



Obrázek 45: Krabicové grafy pro menší soubor dat, kde A1: parametr citlivost pro aceton, A2: parametr citlivost pro ethanol, A3: parametr citlivost pro toluen, B1: parametr energie signálu pro aceton, B2: parametr energie signálu pro ethanol, B3: parametr energie signálu pro toluen, C1: parametr nejrychlejší doba náběhu napětí pro aceton, C2: parametr nejrychlejší doba náběhu napětí pro ethanol, C3: parametr nejrychlejší doba náběhu napětí pro toluen.



Obrázek 46: Krabicové grafy pro větší soubor dat, kde A1: parametr citlivost pro aceton, A2: parametr citlivost pro ethanol, A3: parametr citlivost pro toluen, B1: parametr energie signálu pro aceton, B2: parametr energie signálu pro ethanol, B3: parametr energie signálu pro toluen, C1: parametr nejrychlejší doba náběhu napětí pro aceton, C2: parametr nejrychlejší doba náběhu napětí pro ethanol, C3: parametr nejrychlejší doba náběhu napětí pro toluen.

Podle krabicových grafů lze vidět, že parametr energie signálu má poměrně velký počet odlehlých hodnot. Také z Obrázek 45 a 46 vyplývá, že toluen má poměrně odlišné hodnoty parametrů v porovnání s acetonem a ethanolom. Po vizuálním posouzení krabicových grafů lze udělat předpoklad, že toluen bude nejlépe predikovaná skupina, oproti tomu predikce acetonu a ethanolu může být obtížná.

Pro klasifikaci byly zvoleny následující klasifikační algoritmy, které nevyžadují velkou datovou sadu (např. oproti umělým neuronovým sítím), nemají předpoklad normálního rozdělení dat (např. oproti LDA) a mají hyperparametry:

- SVM s RBF kernelem (jádrovou funkcí),
- Random Forest,
- k-NN.

Tyto algoritmy mají větší flexibilitu oproti algoritmům, které nemají hyperparametry. Flexibilita spočívá v možnosti optimalizace hyperparametrů. Jednou z běžných metod optimalizace hyperparametrů je tzv. Grid Search. Tato metoda vyhodnocuje model pro každou kombinaci předem definovaných hodnot hyperparametrů. Výstupem je taková kombinace hodnot hyperparametrů, která maximalizuje odhad výkonu modelu.

Hyperparametry optimalizované během analýzy v rámci této práce jsou uvedené v Tabulka 16. Klasifikátor RF má mnohem více hyperparametrů (viz literatura [31]), avšak pro tuto práci byly pro jednoduchost řešení zvoleny pouze dva hyperparametry. Hodnoty hyperparametrů byly optimalizované ve vnitřní smyčce desetinasobné CV pomocí metody Grid Search. Pro klasifikátor k-NN byla použita Euklidovská metrika vzdálenosti.

Tabulka 16: Hyperparametry použitých klasifikačních algoritmů.

Klasifikační algoritmus	Hyperparametr
SVM s RBF kernelem	Box Constraint, Kernel Scale
Random Forest	Počet stromů v lese, minimální počet vzorků potřebných v listovém uzlu
k-NN	Počet nejbližších sousedů (k)

Klasifikační algoritmus k-NN byl proveden pomocí funkce *fitcknn()*, pro klasifikátor RF byla použita funkce *TreeBagger()*, pro SVM s RFB kernelem byly použité funkce *templateSVM()* a *fitcecoc()*. Funkce *fitcecoc()* umožňuje trénování SVM modelu pro klasifikaci s více třídami (byla zvolena metoda „one-vs-one“). Pro každý použitý klasifikátor byla provedena standardizace parametrů (pomocí odpovídajícího parametru „Standardize“).

V Tabulka 17 až 18 jsou uvedené výsledky klasifikace po provedení 10×10 nested CV pro menší a větší soubor dat. Metrikami výkonu klasifikačních modelů byly zvolené přesnost a makro F1-score vypočítané pomocí vnější smyčky desetinasobné CV.

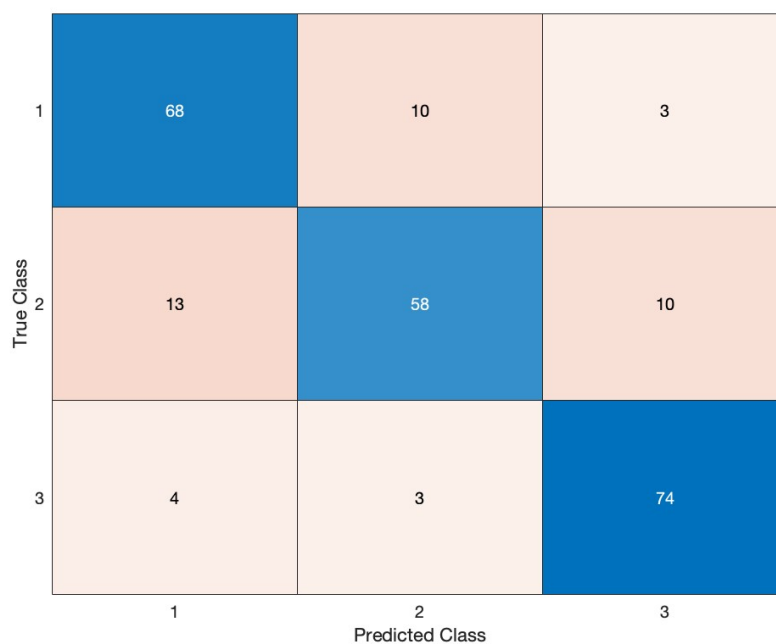
Tabulka 17: Odhad výkonu klasifikačních algoritmů použitých pro klasifikaci VOC na menším datovém souboru.

Klasifikační algoritmus	Průměrná přesnost	Průměrné F1-scóre
k-NN	90 %	0,89
Random Forest	89 %	0,89
SVM s RBF kernelem	82 %	0,81

Tabulka 18: Odhad výkonu klasifikačních algoritmů použitých pro klasifikaci VOC na větším datovém souboru.

Klasifikační algoritmus	Průměrná přesnost	Průměrné F1-scóre
k-NN	86 %	0,86
Random Forest	83 %	0,83
SVM s RBF kernelem	81 %	0,81

Na Obrázek 47 až 49 jsou uvedené matice záměn vypočítané pomocí vnější smyčky desetinasobné CV pro každý klasifikační algoritmus použitý pro klasifikaci VOC na menším datovém souboru.



Obrázek 47: Matice záměn pro výsledky klasifikace pomocí klasifikátoru SVM s RBF kernelem na menším souboru dat, kde třída 1 je aceton, třída 2 je ethanol, třída 3 je toluen.

1	72	8	1
2	11	69	1
3	1	3	77
	1	2	3

Predicted Class

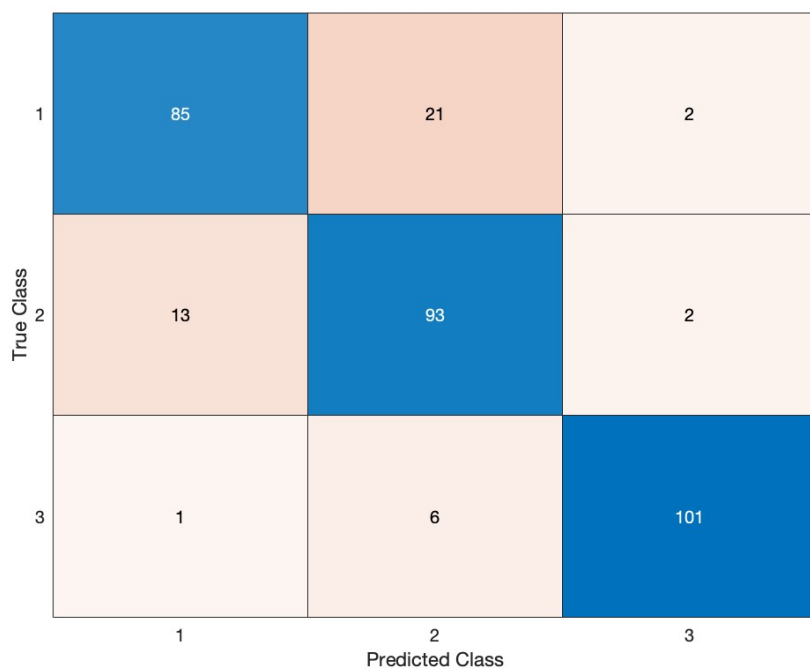
Obrázek 48: Matice záměn pro výsledky klasifikace pomocí klasifikátoru k-NN na menším souboru dat, kde třída 1 je aceton, třída 2 je ethanol, třída 3 je toluen.

1	72	8	1
2	11	69	1
3	2	4	75
	1	2	3

Predicted Class

Obrázek 49: Matice záměn pro výsledky klasifikace pomocí klasifikátoru RF na menším souboru dat, kde třída 1 je aceton, třída 2 je ethanol, třída 3 je toluen.

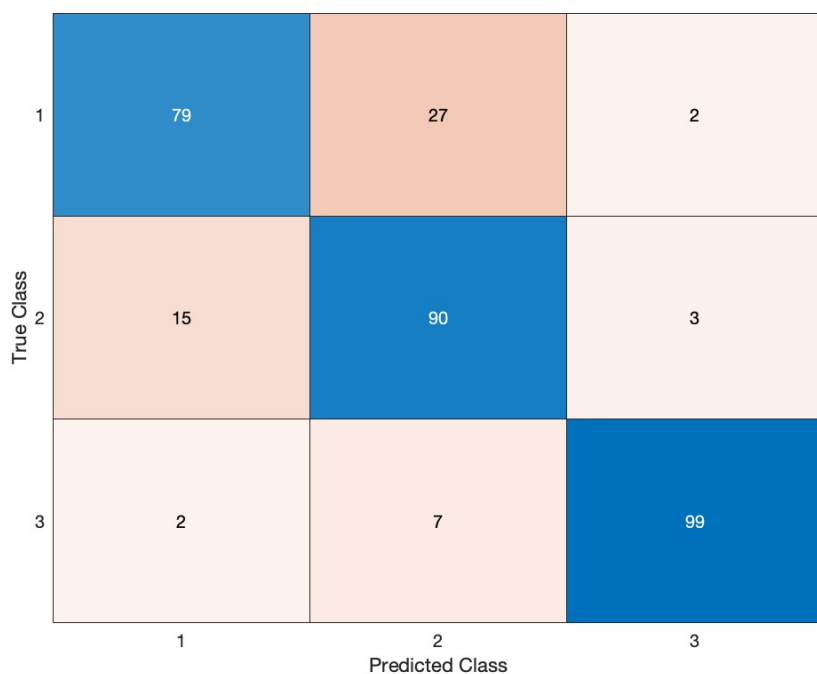
Na Obrázek 50 až 52 jsou uvedené matice záměn vypočítané pomocí vnější smyčky desetinasobné CV pro každý klasifikační algoritmus použitý pro klasifikace VOC na větším datovém souboru.



A confusion matrix for a k-NN classifier. The y-axis is labeled 'True Class' and has categories 1, 2, and 3. The x-axis is labeled 'Predicted Class' and has categories 1, 2, and 3. The matrix cells contain the following values: (1,1)=85, (1,2)=21, (1,3)=2, (2,1)=13, (2,2)=93, (2,3)=2, (3,1)=1, (3,2)=6, (3,3)=101. The diagonal elements are highlighted in blue, while off-diagonal elements are in shades of orange.

True Class \ Predicted Class	1	2	3
1	85	21	2
2	13	93	2
3	1	6	101

Obrázek 50: Matice záměn pro výsledky klasifikace pomocí klasifikátoru k-NN na větším souboru dat, kde třída 1 je aceton, třída 2 je ethanol, třída 3 je toluen.



A confusion matrix for an RF classifier. The y-axis is labeled 'True Class' and has categories 1, 2, and 3. The x-axis is labeled 'Predicted Class' and has categories 1, 2, and 3. The matrix cells contain the following values: (1,1)=79, (1,2)=27, (1,3)=2, (2,1)=15, (2,2)=90, (2,3)=3, (3,1)=2, (3,2)=7, (3,3)=99. The diagonal elements are highlighted in blue, while off-diagonal elements are in shades of orange.

True Class \ Predicted Class	1	2	3
1	79	27	2
2	15	90	3
3	2	7	99

Obrázek 51: Matice záměn pro výsledky klasifikace pomocí klasifikátoru RF na větším souboru dat, kde třída 1 je aceton, třída 2 je ethanol, třída 3 je toluen.

True Class	Predicted Class		
	1	2	3
1	84	21	3
2	14	82	12
3	7	4	97

Obrázek 52: Matice záměn pro výsledky klasifikace pomocí klasifikátoru SVM s RBF kernelem na větším souboru dat, kde třída 1 je aceton, třída 2 je ethanol, třída 3 je toluen.

5.5. Hodnocení klasifikace

Z Tabulka 17 až 18 lze vidět, že klasifikátor SVM s RBF kernelem má nejnižší odhad přesnosti a F1-scóre. Nižší přesnost klasifikace by mohla být způsobená tím, že SVM vyžaduje větší velikost datové sady (s tím, že jednotlivé měření byly provedeny za podobných okolních podmínek, větší velikost datové sady použité pro klasifikace může být prospěšná v dosažení vyšší přesnosti klasifikace s použitím algoritmu SVM s RBF kernelem). Zároveň odhad přesnosti klasifikace na větším souboru, který obsahuje měření provedené za více odlišných okolních podmínek než menší soubor dat, klesl pouze o 1 %. Z toho vyplývá, že SVM s RBF kernelem je nejvíc robustní klasifikační algoritmus z použitých algoritmů pro účely klasifikace v této práci.

Nejvyšší odhad přesnosti klasifikace dosáhl klasifikační algoritmus k-NN, který je 90 % (odhad F1-scóre je 0,89) pro menší soubor dat a 86 % (odhad F1-scóre je 0,86) pro větší soubor dat. Tento klasifikátor je jednodušší na implementaci a interpretaci kvůli tomu, že má pouze jeden hyperparametr tzv. počet nejbližších sousedů (někdy se také optimalizuje metrika vzdálenosti, tato práce se tím nezabývá). V tomto případě je klasifikační algoritmus k-NN nejvhodnější pro danou klasifikační úlohu.

Odhad přesnosti klasifikace pomocí algoritmu RF dosáhl na menším souboru dat 89 % (odhad F1-scóre je 0,89). Zároveň když byl tento klasifikátor použit na větší datové sadě, nastal největší pokles odhadu přesnosti, a to o 6 %.

Podle matic záměn (viz Obrázek 47 až 52) byla predikce toluenu nejpřesnější (největší počet TP). A to pro každý klasifikační algoritmus i větší a menší datovou sadu.

Predikce ethanolu a acetonu byla více problematická (poměrně velký počet FN a FP pro tyhle třídy). Tím se potvrdil předpoklad, který byl udělán podle vizuálního posouzení krabicových grafů parametrů použitých pro klasifikaci.

5.6. Diskuse

Teplota a vlhkost okolního prostředí jsou parazitními vlivy a negativně ovlivňují identifikaci (klasifikaci) VOC. To se ukázalo i v této diplomové práci vzhledem k tomu, že klasifikace byla provedená pro dva typy souborů dat. První, menší soubor dat, který se skládal z měření provedených s nižším rozsahem okolních podmínek, kdy se teplota pohybovala v rozmezí 21 až 23 °C, vlhkost okolního vzduchu se pohybovala v rozmezí 30 až 43 % RH. A druhý, větší soubor dat, který se skládal z měření provedených za více odlišných okolních podmínek, kdy se teplota pohybovala v rozmezí 21 až 25 °C, vlhkost okolního vzduchu se pohybovala v rozmezí 30 až 55 % RH.

Vlivem většího rozptylu hodnot parazitních parametrů byl průměrný odhad přesnosti klasifikace pomocí klasifikačního algoritmu k-NN snížen o 4 % i přesto, že byl proveden na větším souboru dat, který obsahoval navíc 81 dalších měření získaných pomocí chemorezistivních senzorů plynu za větší vlhkosti a teploty okolního prostředí.

Vzhledem k velikosti datového souboru, byla pro odhad výkonu modelu zvolena tzv. nested CV, která efektivně používá data pro trénování a optimalizaci hyperparametrů a testování. Avšak tato metoda má velkou výpočetní náročnost ($O(najk)$, kde n je počet vzorků, a je počet hyperparametrů, j je počet podmnožin ve vnitřní smyčce CV, k je počet podmnožin ve vnější smyčce CV podle [33]) a pro větší datové sady by nebyla vhodná. Jak bylo uvedeno v kapitole č. 5.5.1 v případě analýzy větší datové sady by bylo vhodnější rozdělit soubor dat na testovací sadu (např. 80 % dat) a trénovací sadu (např. 20 % dat). Testovací sadu lze použít pro validaci a hledání optimálních hodnot hyperparametrů klasifikačních modelů pomocí metody k-násobné CV (k-folds CV), která je méně výpočetní náročná ($O(nak)$, kde n je počet vzorků, a je počet hyperparametrů, k je počet podmnožin (tzv. foldů) CV podle [33]) oproti metodě nested CV. Odhad výkonu modelu lze pak vypočítat na testovací sadě, která nikdy nebyla použita pro trénování. Tento postup tak poskytne skutečný odhad, jak model bude schopný generalizovat na neviditelná (nová) data.

Extrakce, selekce parametrů a následná optimalizace hyperparametrů klasifikačních algoritmů se ukázaly klíčovými kroky ovlivňujícími odhad výkonu modelu. Ačkoliv cílem analýzy bylo dosáhnout optimální rovnováhy mezi zaujatostí a rozptylem (tzv. bias-variance tradeoff), nejvyšší průměrný odhad přesnosti klasifikace dosáhl 90 %. Nedostatečné přizpůsobení modelů mohlo být způsobeno malou velikostí datového souboru a malým počtem parametrů extrahovaných ze surových dat. Možným řešením by mohlo být provedení více měření za stálých podmínek a použití metody augmentace datového souboru, čímž se zvětší velikost trénovací množiny.

V dané práci byly pro analýzu použity parametry z časové oblasti, které lze snadno interpretovat. Ačkoliv byl pro řešení dané klasifikační úlohy takový přístup extrakce parametrů postačující (maximální přesnost klasifikace dosáhla 90 %), existuje velký potenciál extrakce parametrů z frekvenční oblasti, například pomocí diskrétní vlnkové transformace (angl. discrete wavelet transform, DWT), pomocí které lze nalézt skryté rysy, které budou přínosné pro klasifikaci.

Dále v této práci byla pro selekci parametrů použita jednosměrná ANOVA. Tento přístup je jednoduchý na interpretaci a obecný (selekce parametrů je závislá na varianci parametrů, nikoliv na konkrétním klasifikačním algoritmu, který bude následně aplikován na vybrané parametry). Jednou z alternativních možností selekce parametrů, která se obvykle používá je tzv. sekvenční selekce parametrů, kterou lze provést například pomocí funkce *sequentialfs()* v MATLAB. Metoda sekvenční selekce parametrů má výhodu oproti předchozí metodě selekce parametrů v tom, že nemá předpokladů, penalizační kritérium lze nastavit uživatelem. Nevýhoda takové metody spočívá v tom, že při provedení sekvenční selekce parametrů pomocí funkce *sequentialfs()* není výstup obecný, protože funkce *sequentialfs()* zahrnuje uživatelskou funkci *fun()*, která se nastavuje pro konkrétní klasifikační algoritmus.

Další možností dosažení vyššího výkonu klasifikace (a lepší generalizace modelů) je použití jiné metriky vzdálenosti pro klasifikátor k-NN a optimalizaci dalších hyperparametrů klasifikátoru RF a úprava mřížky hyperparametrů pro metodu Grid Search (podrobnější popis hyperparametrů lze nalézt v článku [31]).

6. Závěr

Předložená diplomová práce se zabývá problematikou chemorezistivních senzorů plynu, testováním těkavých látek a následným zpracováním naměřených dat s využitím metod strojového učení. Hlavním cílem bylo navrhnout a realizovat vhodné testovací pracoviště pro měření VOC a zpracovat naměřená data pro přesnější identifikaci detekovaných látek.

Byly realizované a vyhodnocené dvě varianty testovacího pracoviště. První varianta testovacího pracoviště byla provedena s využitím nepájivého pole, druhá varianta pracoviště byla realizována na plošném spoji. Ukázalo se, že varianta testovacího pracoviště s využitím plošného spoje je vhodnější, jelikož poskytuje stabilnější výsledky měření a je odolnější vůči rušení. Pomocí druhé varianty testovacího pracoviště s využitím devíti tlustovrstvých chemorezistivních senzorů byly naměřené tři vybrané VOC ve třech koncentracích. Celkem bylo provedeno 36 měření (každá VOC pro každou koncentraci byla změřena čtyřikrát). Pro ilustraci časové náročnosti každé měření včetně přípravy trvalo minimálně 60 minut, celkově tedy 36 hodin, ale díky použití matice devíti chemorezistivních senzorů byl vytvořen soubor dat velikosti celkem 324 surových měření. Vybrané VOC jsou potenciálními biomarkery nemocí. Měření bylo provedeno za variabilních okolních podmínek, jelikož je velice pravděpodobné, že při aplikaci chemorezistivních senzorů plynu v reálném světě teplota a vlhkost okolního prostředí nebudou stacionární, což může značně ovlivnit výsledky měření a přesnost identifikace cílové látky. Použité chemorezistivní senzory byly vybrány na základě nízké ceny a široké dostupnosti tak, aby byla umožněna snadná reprodukce měření na jiných pracovištích. Společně s postupem měření a zpracováním dat předložená diplomová práce poslouží jako návod pro práce dalších studentů.

V rámci zpracování dat byla pro znázornění parazitních vlivů teploty a vlhkosti naměřená data rozdělena do dvou datových souborů. První soubor obsahuje 243 měření provedených za stabilních okolních podmínek (rozsah teplot je 21 až 23 °C, rozsah vlhkosti je 30 až 43 % RH). Druhý soubor zahrnuje 324 měření (k prvnímu souboru je přidáno 81 dalších měření) včetně těch, která byla provedena za teploty a vlhkosti v rozmezí 23 až 25 °C a 45 až 55 % RH.

Pro analýzu byla provedena extrakce parametrů snímaných průběhů odezev senzorů z časové oblasti a selekce parametrů využitím testu korelace a série testů jednosměrné ANOVA.

Následně byly pro vizualizaci a klasifikaci použity pouze tři parametry. Vizualizace vysokodimenzionálního prostoru byla provedena za použití lineární metody PCA a nelineární

metody t-SNE. Ačkoliv výsledek není stabilní, metoda t-SNE se ukázala vhodnější pro vizualizaci naměřených dat.

Pro klasifikaci byly použity celkem tři klasifikační algoritmy: k-NN, RF a SVM s RBF kernelem. Vzhledem k velikosti datových souborů pro optimalizaci hyperparametrů a hodnocení výkonu jednotlivých klasifikátorů byla zvolena metoda nested CV. Dle této metody má nejvyšší odhad přesnosti a F1-scóre klasifikační algoritmus k-NN. Nejvyšší odhad přesnosti klasifikace VOC na menším datovém souboru dosáhl 90 % (odhad F1-scóre je 0,89) a na větším souboru dat dosáhl 86 % (odhad F1-scóre je 0,86). Tím se ukázalo, že teplota a vlhkost okolního prostředí mohou značně ovlivnit výsledky klasifikace, zejména v případě, že klasifikační algoritmus není dostatečně robustní.

Možnost rozvinutí této práce vidím v jejím použití pro další výzkum chemorezistivních senzorů pro nalezení optimální metody extrakce parametrů (například z frekvenční oblasti pomocí DWT nebo FFT) a nalezení robustnějšího klasifikačního algoritmu. Navíc provedení dalších měření VOC umožní použití tzv. umělé neuronové sítě (angl. Artificial Neural Network, ANN), pomocí níž lze potenciálně dosáhnout mnohem lepších výsledků klasifikace než pomocí metod strojového učení použitých v této diplomové práci.

Další možností rozvinutí této práce je optimalizace testovacího pracoviště, kde testovací komora bude obsahovat separační membránu (oddělovací komoru) umožňující interakci VOC a snímacího materiálu chemorezistivních senzorů až po úplném vypaření VOC, čímž by mohl být eliminován překmit průběhů odezev chemorezistivních senzorů.

Použitá literatura

- [1] L. Mahmood, M. Ghommem, and Z. Bahroun, “Smart Gas Sensors: Materials, Technologies, Practical Applications, and Use of Machine Learning – A Review,” *Journal of Applied and Computational Mechanics*, vol. 9, no. 3, pp. 775–803, 2023, doi: 10.22055/jacm.2023.41985.3851.
- [2] F. South, T. Usf, D. Graduate, G. Usf, D. Theses, and T. C. Miller, “Metal Oxide Sensor Array Test Bed Prototype for Diagnostic Breath Analysis,” 2021.
- [3] “Insights into VOC biomarkers of diabetes.” Accessed: Oct. 24, 2023. [Online]. Available: <https://www.owlstonemedical.com/about/blog/2022/feb/22/voc-biomarkers-diabetes/>
- [4] Z. Chen, Z. Chen, Z. Song, W. Ye, and Z. Fan, “Smart gas sensor arrays powered by artificial intelligence,” *Journal of Semiconductors*, vol. 40, no. 11. Institute of Physics Publishing, 2019. doi: 10.1088/1674-4926/40/11/111601.
- [5] H. Wang, X. Shi, F. Liu, T. Duan, and B. Sun, “Non-Invasive Rapid Detection of Lung Cancer Biomarker Toluene with a Cataluminescence Sensor Based on the Two-Dimensional Nanocomposite Pt/Ti3C2Tx-CNT,” *Chemosensors*, vol. 10, no. 8, Aug. 2022, doi: 10.3390/chemosensors10080333.
- [6] M. Bhogadia, M. Edgar, K. Hunwin, G. Page, and M. Grootveld, “Detection and Quantification of Ammonia as the Ammonium Cation in Human Saliva by 1H NMR: A Promising Probe for Health Status Monitoring, with Special Reference to Cancer,” *Metabolites*, vol. 13, no. 7, p. 792, Jun. 2023, doi: 10.3390/metabo13070792.
- [7] A. Shanmugasundaram, S. V Manorama, D. S. Kim, Y. J. Jeong, and D. Weon Lee, “Toward Point-of-Care chronic disease Management: Biomarker detection in exhaled breath using an E-Nose sensor based on rGO/SnO2 superstructures,” *Chemical Engineering Journal*, vol. 448, Nov. 2022, doi: 10.1016/j.cej.2022.137736.
- [8] A. K. Pathak *et al.*, “Recent Advances in Sensing Materials Targeting Clinical Volatile Organic Compound (VOC) Biomarkers: A Review,” *Biosensors*, vol. 13, no. 1. MDPI, Jan. 01, 2023. doi: 10.3390/bios13010114.
- [9] S. Huang *et al.*, “Machine Learning-Enabled Smart Gas Sensing Platform for Identification of Industrial Gases,” *Advanced Intelligent Systems*, vol. 4, no. 4, Apr. 2022, doi: 10.1002/aisy.202200016.

- [10] B. Yang, N. V. Myung, and T. T. Tran, “1D Metal Oxide Semiconductor Materials for Chemiresistive Gas Sensors: A Review,” *Advanced Electronic Materials*, vol. 7, no. 9. John Wiley and Sons Inc, Sep. 01, 2021. doi: 10.1002/aelm.202100271.
- [11] P. Chesler and C. Hornoïu, “MOX-Based Resistive Gas Sensors with Different Types of Sensitive Materials (Powders, Pellets, Films), Used in Environmental Chemistry,” *Chemosensors*, vol. 11, no. 2. MDPI, Feb. 01, 2023. doi: 10.3390/chemosensors11020095.
- [12] A. N. Abdullah *et al.*, “Correction Model for Metal Oxide Sensor Drift Caused by Ambient Temperature and Humidity,” *Sensors*, vol. 22, no. 9, May 2022, doi: 10.3390/s22093301.
- [13] R. Kumar, O. Al-Dossary, G. Kumar, and A. Umar, “Zinc oxide nanostructures for no2 gas–sensor applications: A review,” *Nano-Micro Letters*, vol. 7, no. 2. SpringerOpen, pp. 97–120, Apr. 01, 2015. doi: 10.1007/s40820-014-0023-3.
- [14] S. Feng *et al.*, “Review on smart gas sensing technology,” *Sensors (Switzerland)*, vol. 19, no. 17. MDPI AG, Sep. 01, 2019. doi: 10.3390/s19173760.
- [15] S. R. Rajanalli, S. Endoli, M. Magi, and N. Ramavenkateswaran, “Machine learning algorithms for classification of gas sensor array dataset.”
- [16] I. Cho *et al.*, “Deep-learning-based gas identification by time-variant illumination of a single micro-LED-embedded gas sensor,” *Light Sci Appl*, vol. 12, no. 1, Dec. 2023, doi: 10.1038/s41377-023-01120-7.
- [17] U. Yaqoob and M. I. Younis, “Chemical gas sensors: Recent developments, challenges, and the potential of machine learning—a review,” *Sensors*, vol. 21, no. 8. MDPI AG, Apr. 02, 2021. doi: 10.3390/s21082877.
- [18] T. Pavlík and L. Dušek, *Biostatistika*. Akademické nakladatelství CERM, 2012.
- [19] M. Kumar, N. K. Rath, A. Swain, and S. K. Rath, “Feature Selection and Classification of Microarray Data using MapReduce based ANOVA and K-Nearest Neighbor,” in *Procedia Computer Science*, Elsevier, 2015, pp. 301–310. doi: 10.1016/j.procs.2015.06.035.
- [20] R. Dhanya, I. R. Paul, S. S. Akula, M. Sivakumar, and J. J. Nair, “F-test feature selection in Stacking ensemble model for breast cancer prediction,” in *Procedia Computer Science*, Elsevier B.V., 2020, pp. 1561–1570. doi: 10.1016/j.procs.2020.04.167.
- [21] “Principal Component Analysis in R Tutorial.” Accessed: Oct. 18, 2023. [Online]. Available: <https://www.datacamp.com/tutorial/pca-analysis-r>

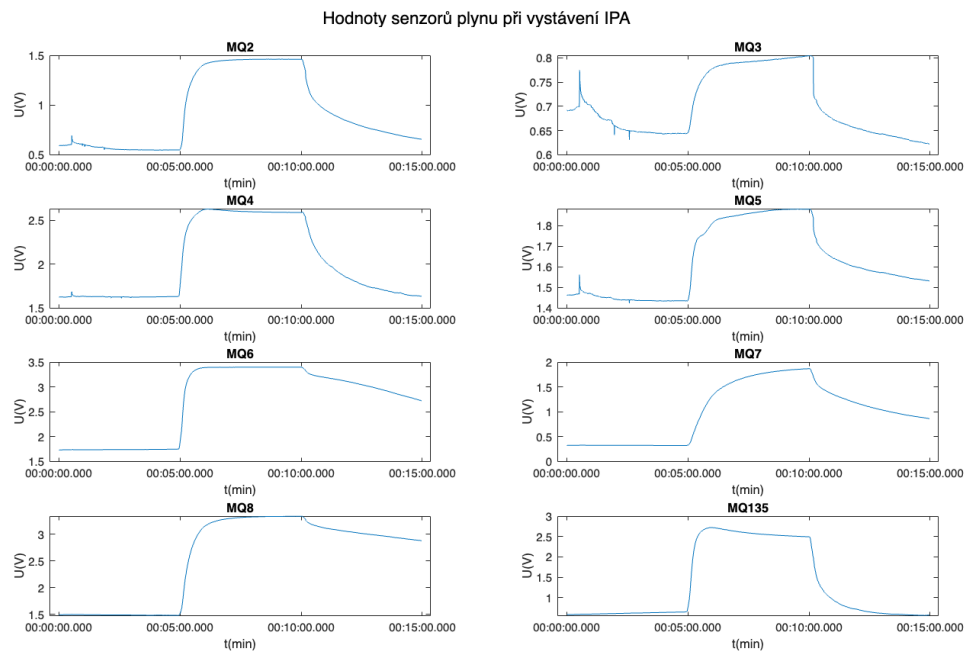
- [22] “Introduction to t-SNE.” Accessed: Apr. 04, 2024. [Online]. Available: <https://www.datacamp.com/tutorial/introduction-t-sne>
- [23] “How to Use t-SNE Effectively.” Accessed: Apr. 04, 2024. [Online]. Available: <https://distill.pub/2016/misread-tsne/>
- [24] L. Van Der Maaten and G. Hinton, “Visualizing Data using t-SNE,” 2008.
- [25] “t-SNE.” Accessed: Apr. 19, 2024. [Online]. Available: <https://www.mathworks.com/help/stats/t-sne.html>
- [26] V. H. Kamble and M. P. Dale, “Machine learning approach for longitudinal face recognition of children,” in *Machine Learning for Biometrics: Concepts, Algorithms and Applications*, Elsevier, 2022, pp. 1–27. doi: 10.1016/B978-0-323-85209-8.00011-0.
- [27] J. Tan and J. Xu, “Applications of electronic nose (e-nose) and electronic tongue (e-tongue) in food quality-related properties determination: A review,” *Artificial Intelligence in Agriculture*, vol. 4, pp. 104–115, 2020, doi: 10.1016/j.aiaa.2020.06.003.
- [28] F. Budiman, “SVM-RBF parameters testing optimization using cross validation and grid search to improve multiclass classification,” *Scientific Visualization*, vol. 11, no. 1, pp. 80–90, 2019, doi: 10.26583/sv.11.1.07.
- [29] W. Ni *et al.*, “Classification and Concentration Predictions of Volatile Organic Compounds Using an Electronic Nose Based on XGBoost-Random Forest Algorithms,” *IEEE Sens J*, vol. 24, no. 1, pp. 671–678, Jan. 2024, doi: 10.1109/JSEN.2023.3304355.
- [30] B. Alghamdi and F. Alharby, “An Intelligent Model for Online Recruitment Fraud Detection,” *Journal of Information Security*, vol. 10, no. 03, pp. 155–176, 2019, doi: 10.4236/jis.2019.103009.
- [31] P. Probst, M. N. Wright, and A. L. Boulesteix, “Hyperparameters and tuning strategies for random forest,” *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, vol. 9, no. 3. Wiley-Blackwell, May 01, 2019. doi: 10.1002/widm.1301.
- [32] M. Heydarian, T. E. Doyle, and R. Samavi, “MLCM: Multi-Label Confusion Matrix,” *IEEE Access*, vol. 10, pp. 19083–19095, 2022, doi: 10.1109/ACCESS.2022.3151048.
- [33] T. J. Bradshaw, Z. Huemann, J. Hu, and A. Rahmim, “A Guide to Cross-Validation for Artificial Intelligence in Medical Imaging,” *Radiol Artif Intell*, vol. 5, no. 4, Jul. 2023, doi: 10.1148/ryai.220232.

- [34] “MQ3 Gas Sensor Technical Data”, Accessed: May 20, 2024. [Online]. Available: https://www.laskakit.cz/user/related_files/mq-3_datasheet.pdf
- [35] V. Saasa, T. Malwela, M. Beukes, M. Mokgotho, C.-P. Liu, and B. Mwakikunga, “Sensing Technologies for Detection of Acetone in Human Breath for Diabetes Diagnosis and Monitoring,” *Diagnostics*, vol. 8, no. 1, p. 12, Jan. 2018, doi: 10.3390/diagnostics8010012.
- [36] M. Bland, “An Introduction to Medical Statistics,” 2015.
- [37] S. Acharyya, S. Nag, and P. K. Guha, “Ultra-selective tin oxide-based chemiresistive gas sensor employing signal transform and machine learning techniques,” *Anal Chim Acta*, vol. 1217, Jul. 2022, doi: 10.1016/j.aca.2022.339996.
- [38] “kstest.” Accessed: May 05, 2024. [Online]. Available: <https://www.mathworks.com/help/stats/kstest.html>
- [39] M. J. Blanca, R. Alarcón, J. Arnau, R. Bono, and R. Bendayan, “Effect of variance ratio on ANOVA robustness: Might 1.5 be the limit?,” *Behav Res Methods*, vol. 50, no. 3, pp. 937–962, Jun. 2018, doi: 10.3758/s13428-017-0918-2.

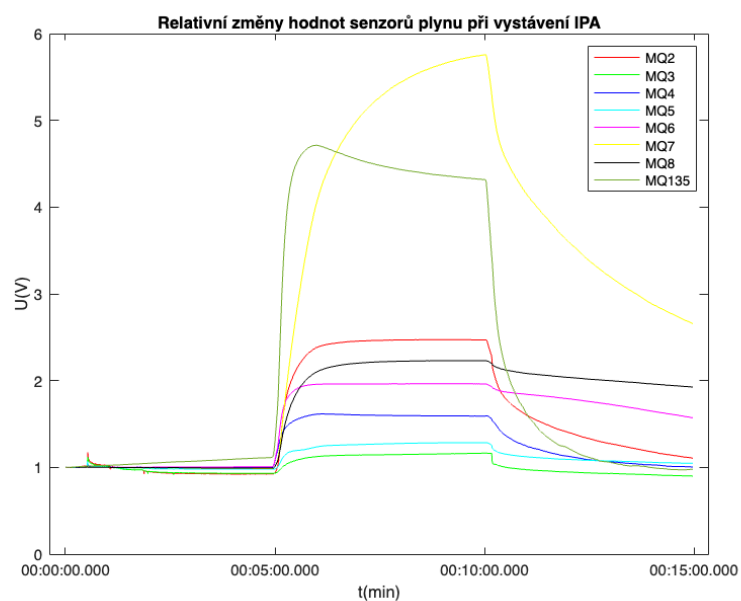
Přílohy

Příloha A - Průběhy časových odezev senzorů s použitím 1. varianty zapojení senzorů plynu:

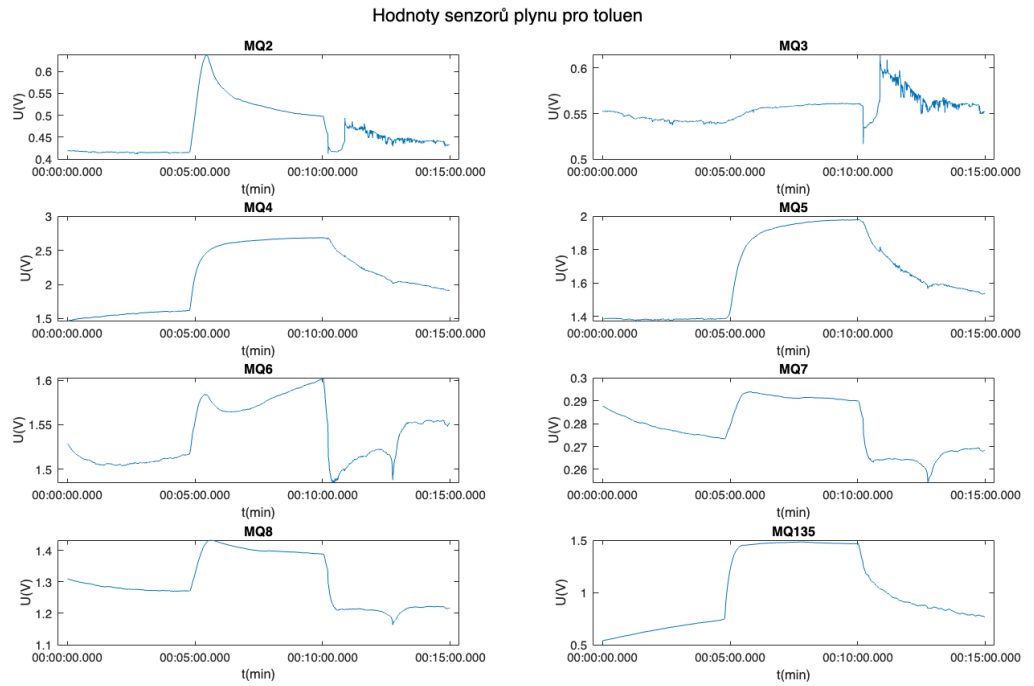
- Průběhy odezev chemorezistivních senzorů při měření IPA s koncentrací 5 ppm.



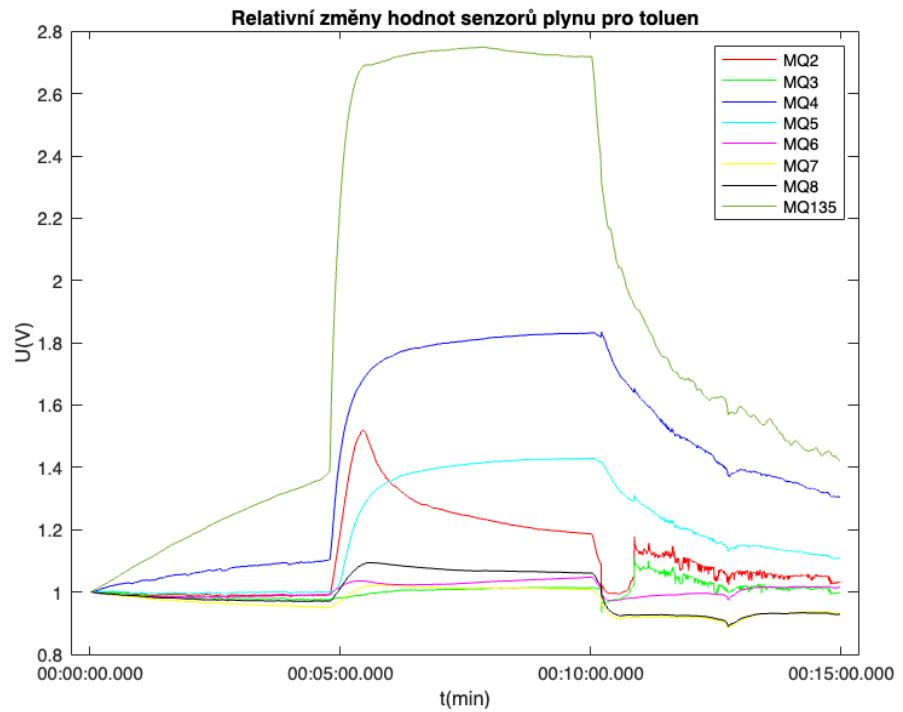
- Relativní změny časových odezev chemorezistivních senzorů při měření IPA s koncentrací 5 ppm.



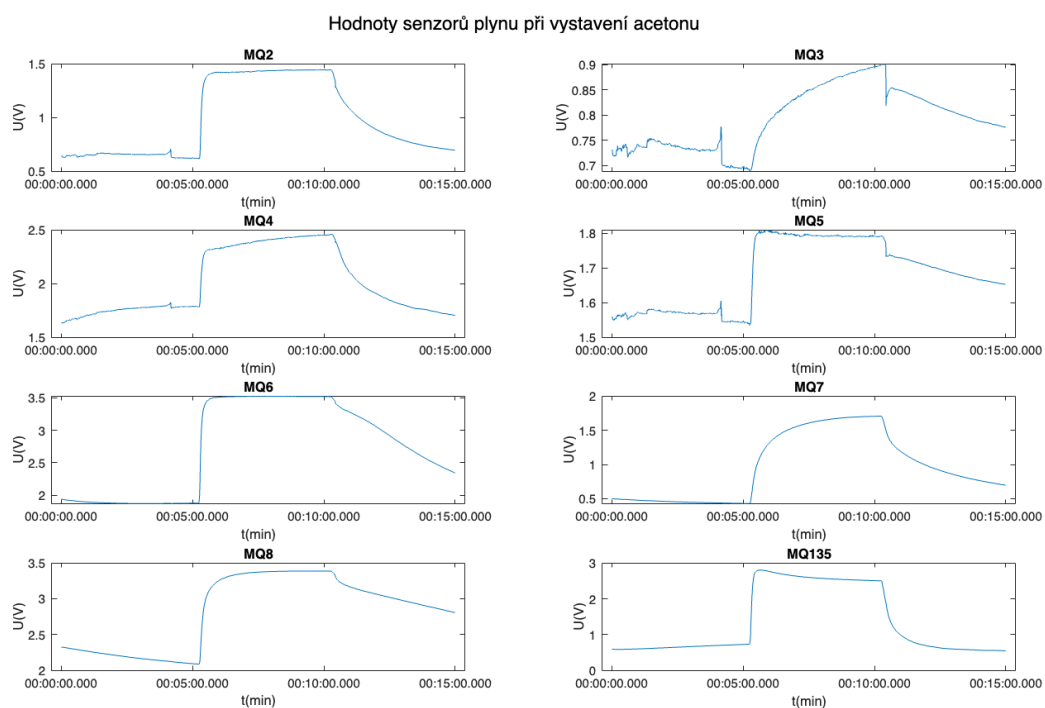
- Průběhy odezev chemorezistivních senzorů při měření toluenu s koncentrací 5 ppm.



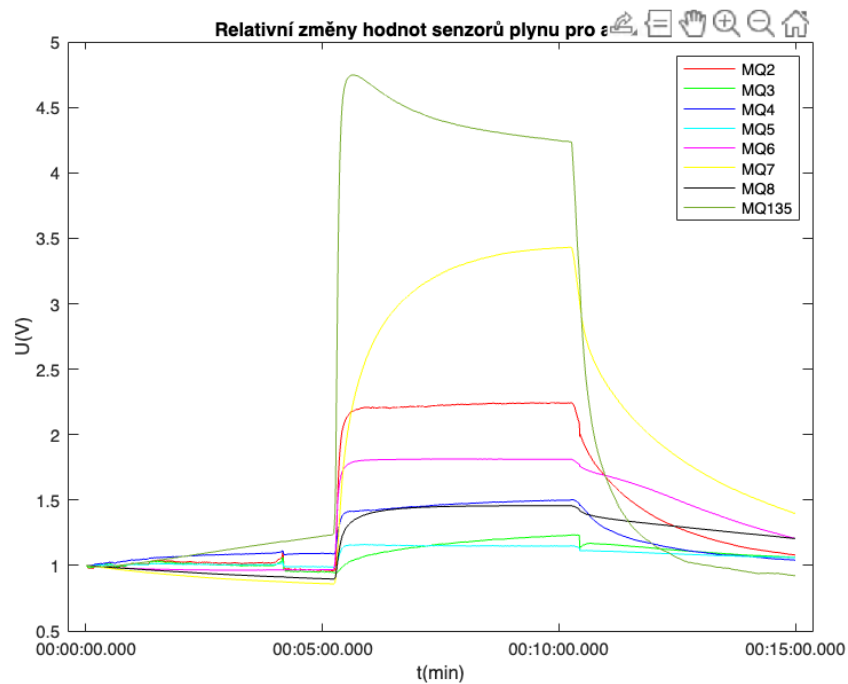
- Relativní změny časových odezev chemorezistivních senzorů při měření toluenu s koncentrací 5 ppm.



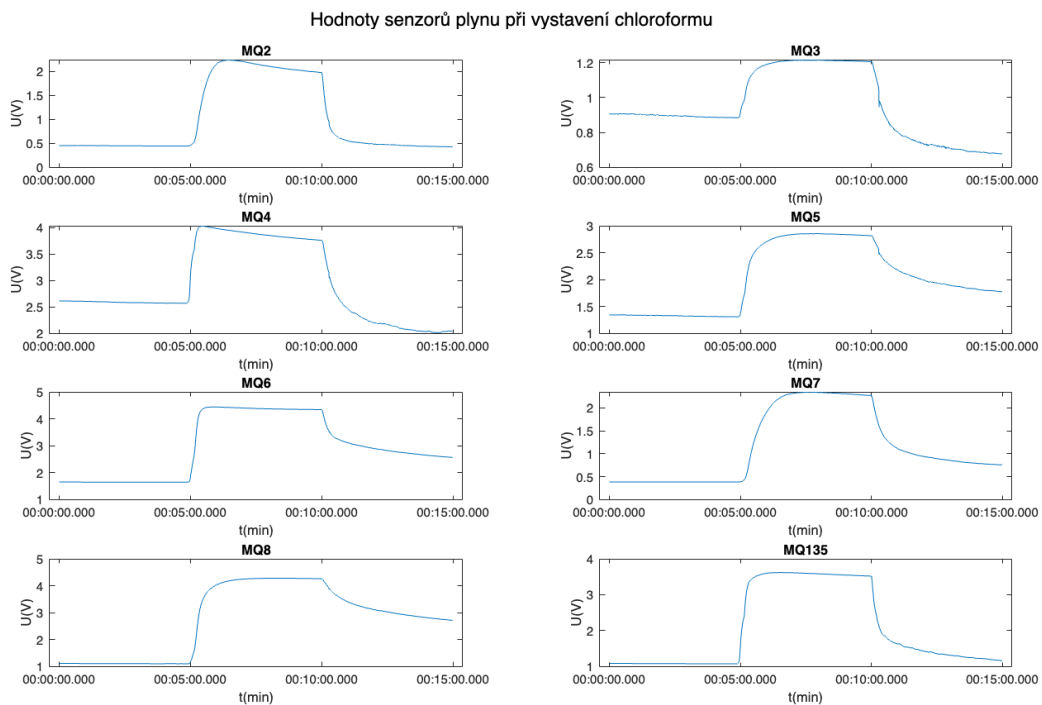
- Průběhy odezev chemorezistivních senzorů při měření acetonu s koncentrací 5 ppm.



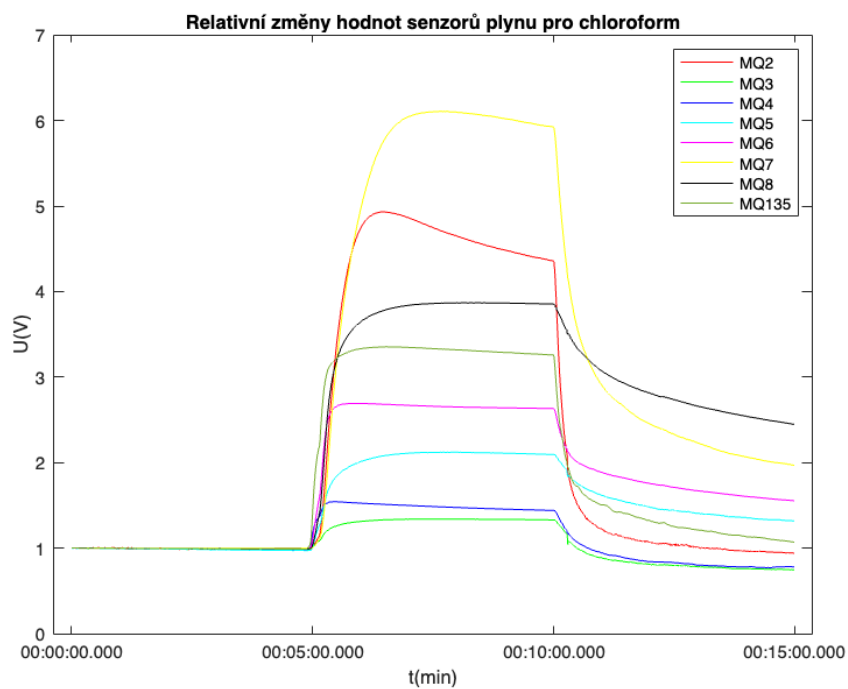
- Relativní změny časových odezev chemorezistivních senzorů při měření acetonu s koncentrací 5 ppm.



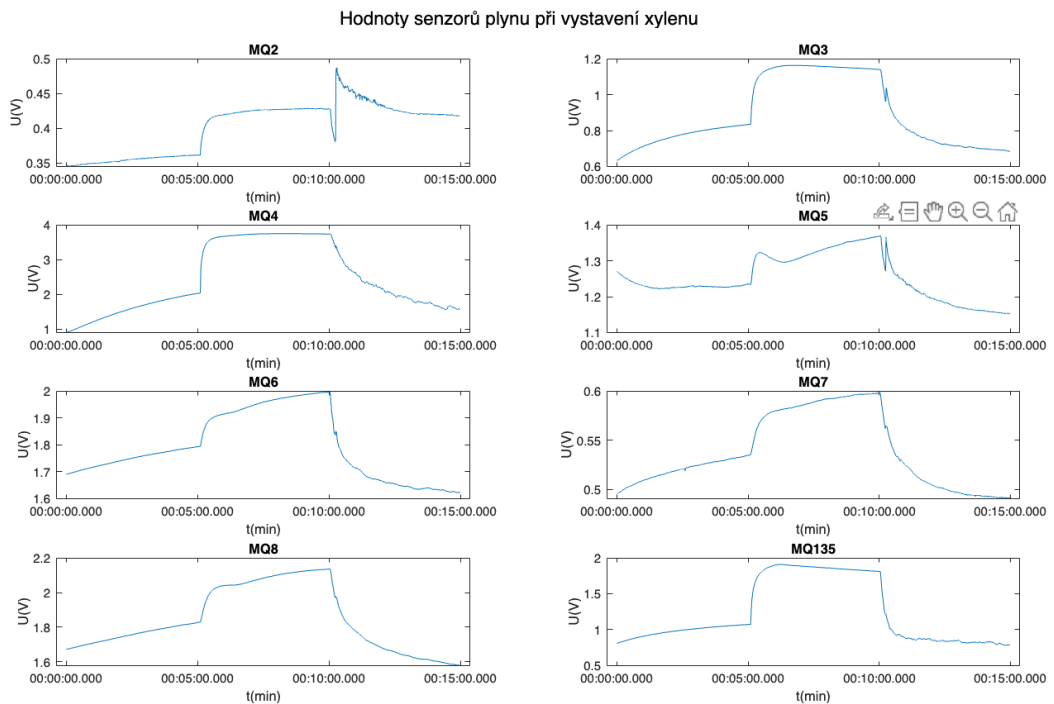
- Průběhy odezev chemorezistivních senzorů při měření chloroformu s koncentrací 5 ppm.



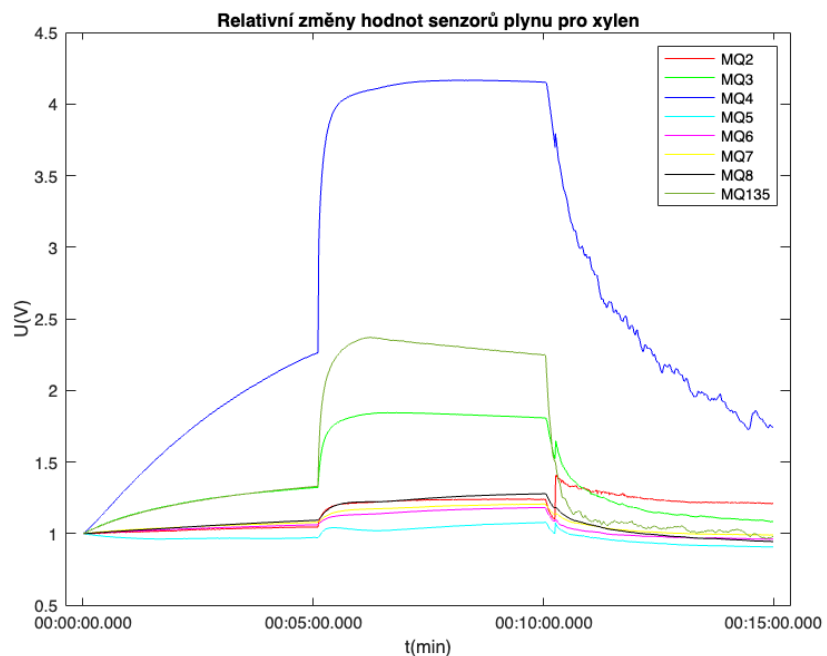
- Relativní změny časových odezev chemorezistivních senzorů při měření chloroformu s koncentrací 5 ppm.



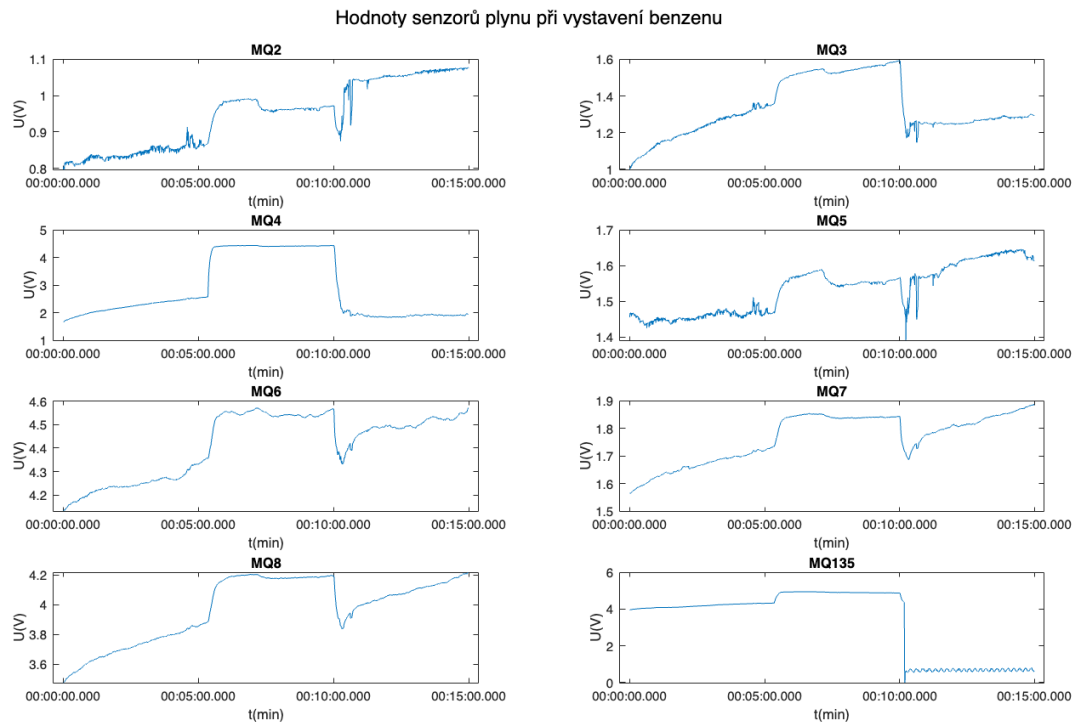
- Průběhy odezev chemorezistivních senzorů při měření xylenu s koncentrací 5 ppm.



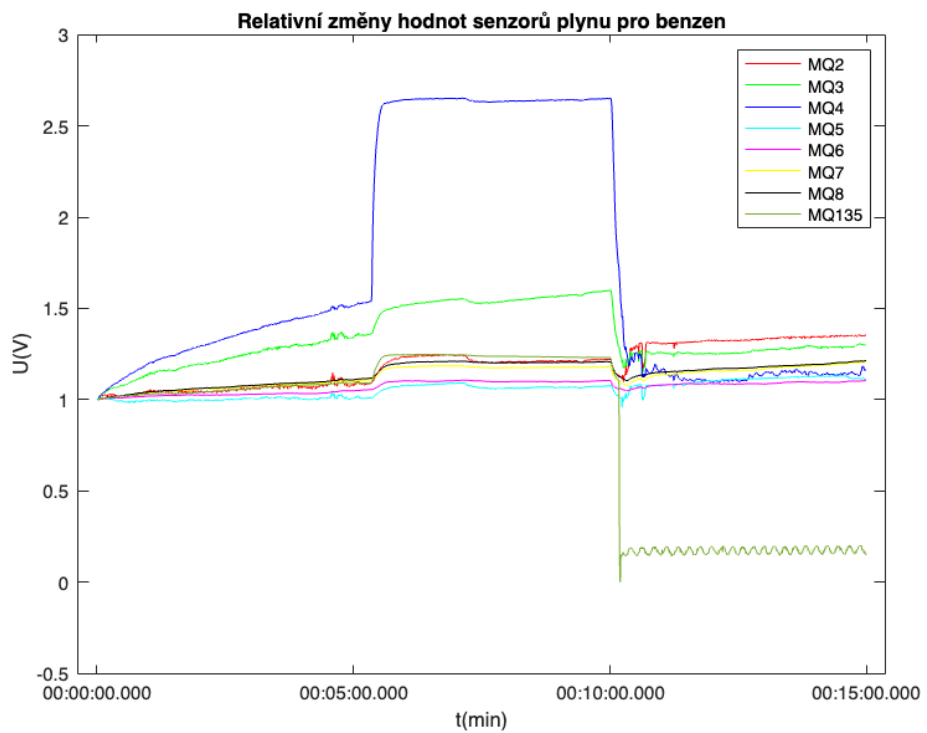
- Relativní změny časových odezev chemorezistivních senzorů při měření xylenu s koncentrací 5 ppm.



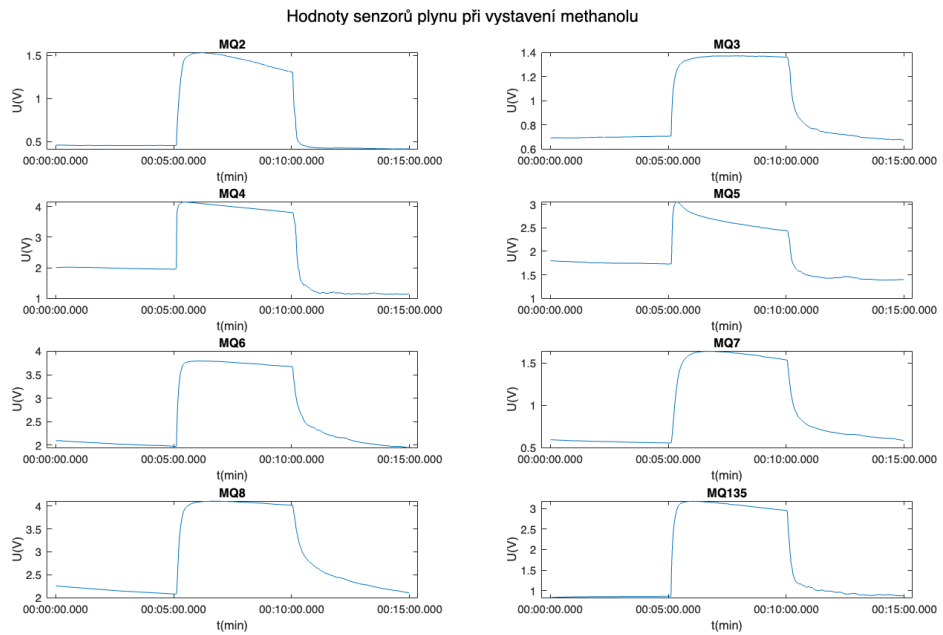
- Průběhy odezev chemorezistivních senzorů při měření benzenu s koncentrací 5 ppm.



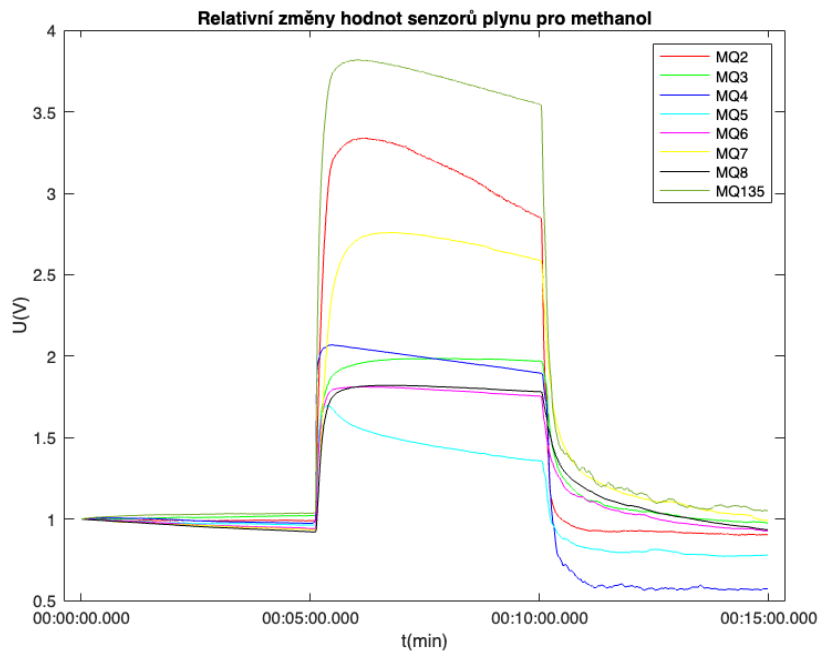
- Relativní změny časových odezev chemorezistivních senzorů při měření benzenu s koncentrací 5 ppm.



- Průběhy odezev chemorezistivních senzorů při měření methanolu s koncentrací 5 ppm.



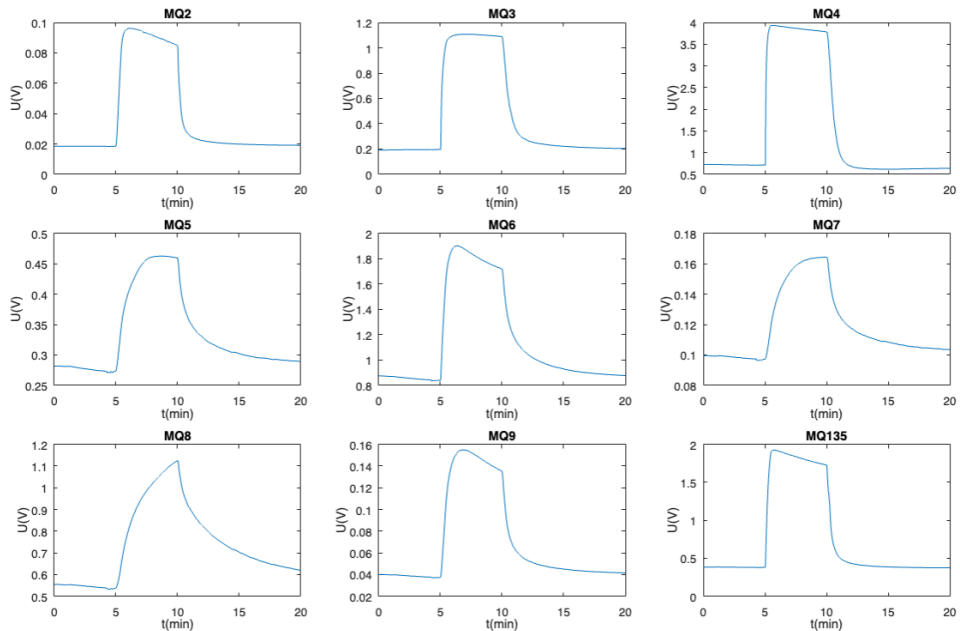
- Relativní změny časových odezev chemorezistivních senzorů při měření methanolu s koncentrací 5 ppm.



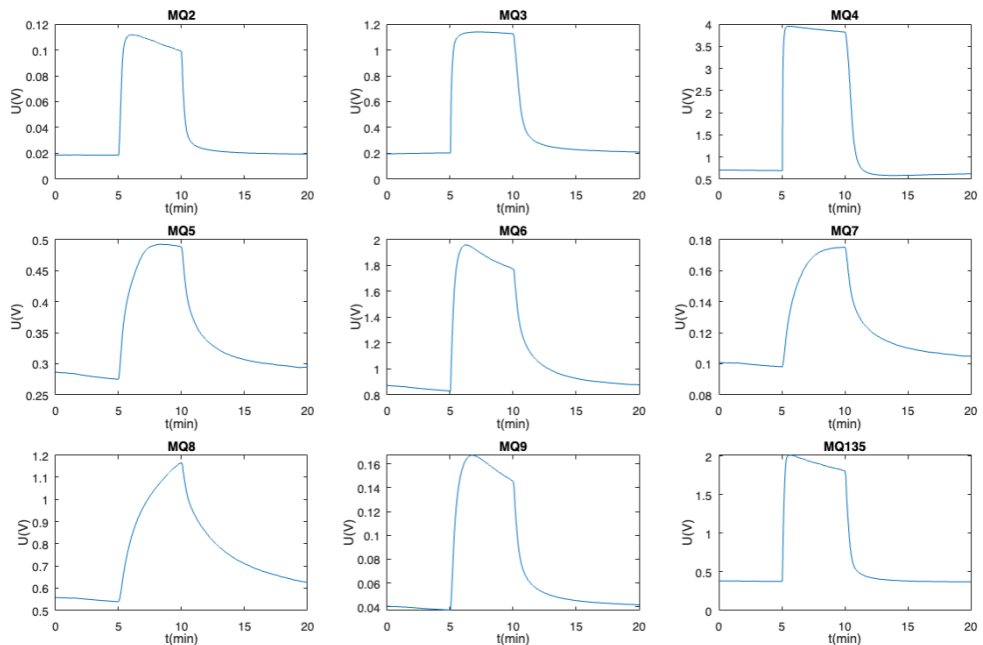
Příloha B - Průběhy časových odezev senzorů s použitím 2. varianty zapojení senzorů plynu

- Průběhy odezev chemorezistivních senzorů při měření ethanolu s koncentrací 1 ppm.

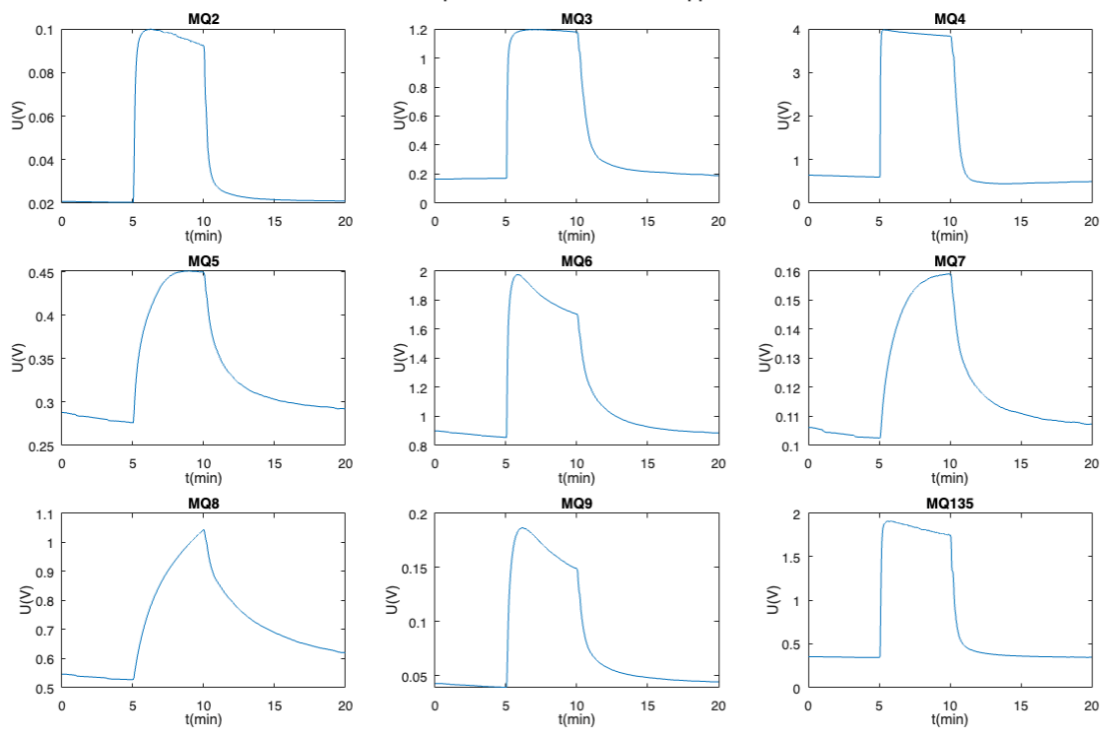
Odezva senzorů pro ethanol s koncentrací 1 ppm 2. měření



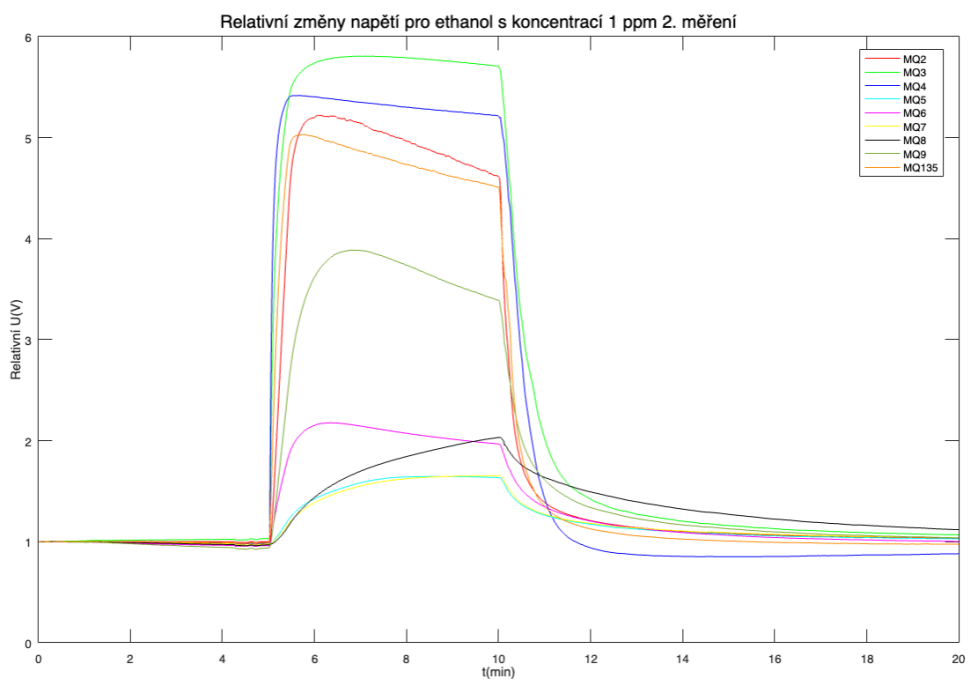
Odezva senzorů pro ethanol s koncentrací 1 ppm 3. měření

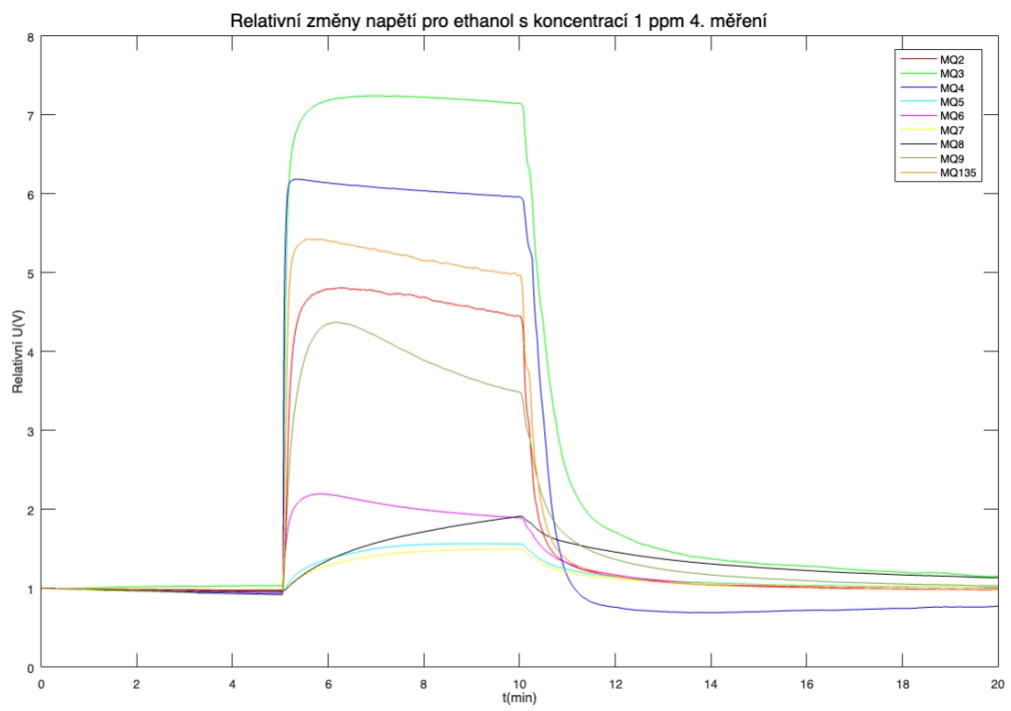
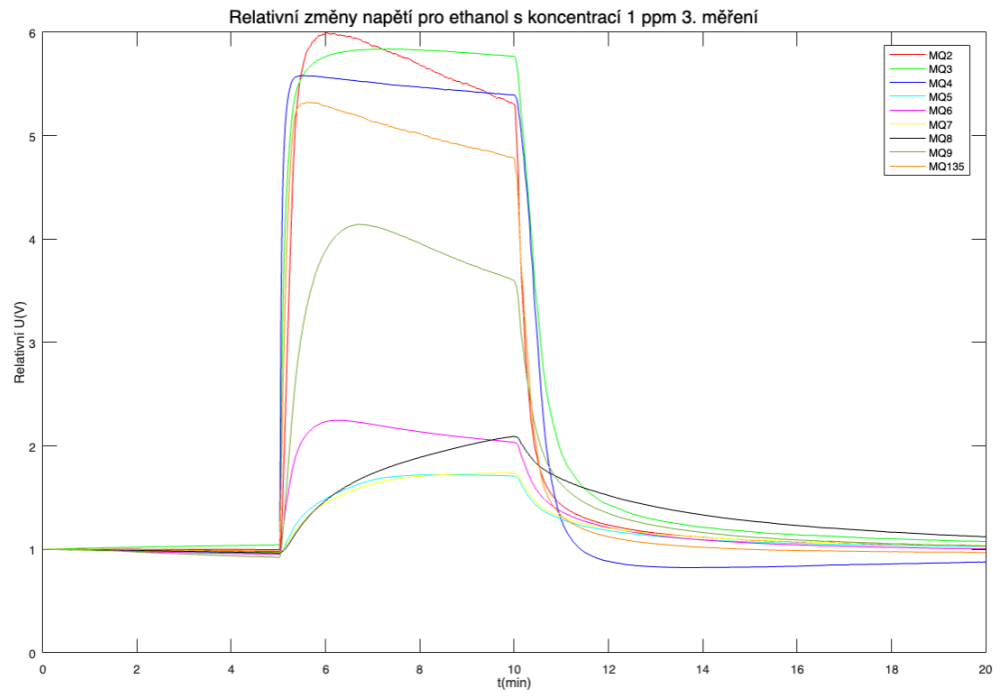


Odezva senzorů pro ethanol s koncentrací 1 ppm 4. měření



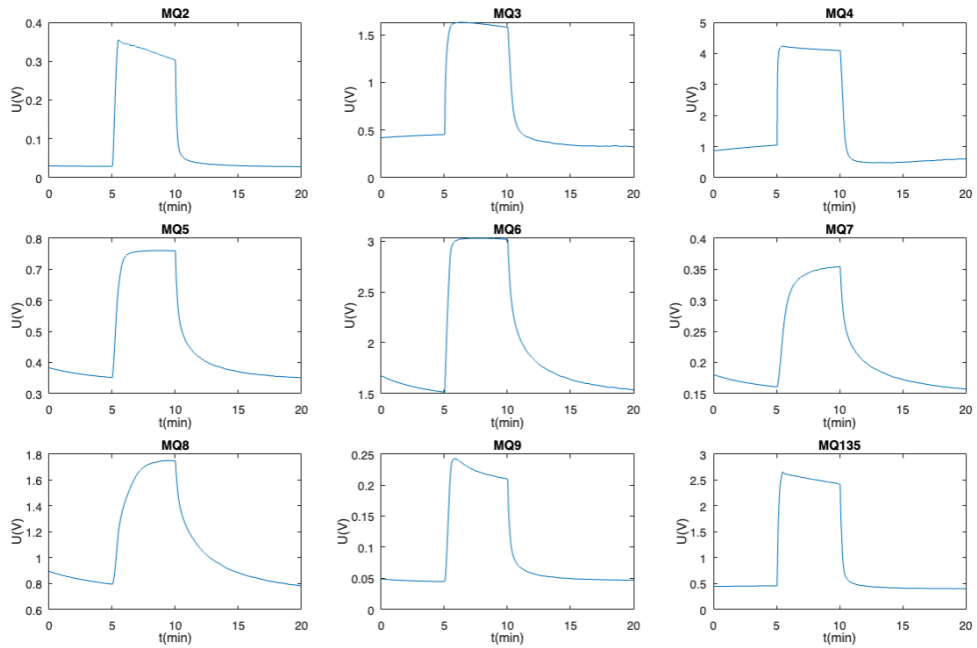
- Relativní změny časových odezev chemorezistivních senzorů při měření ethanolu s koncentrací 1 ppm.



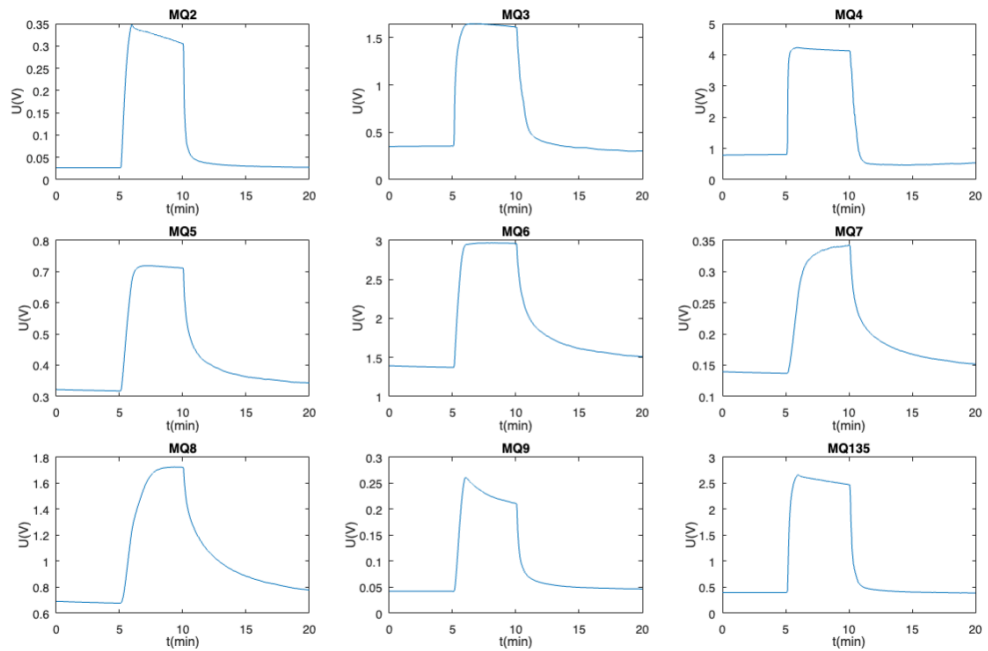


- Průběhy odezvy chemorezistivních senzorů při měření ethanolu s koncentrací 5 ppm.

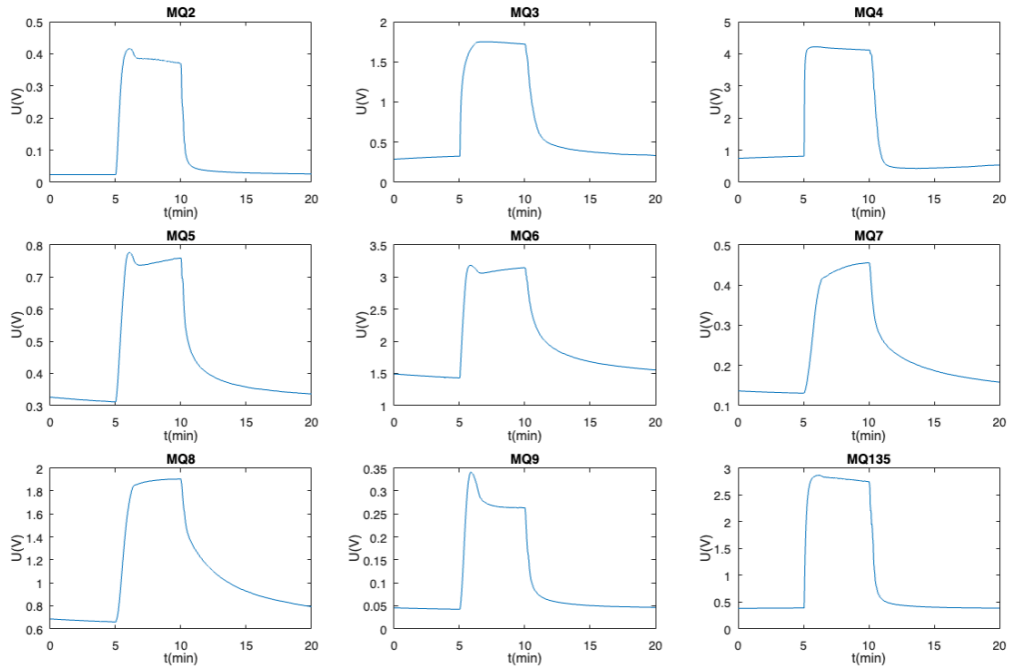
Odezva senzorů pro ethanol s koncentrací 5 ppm 1. měření



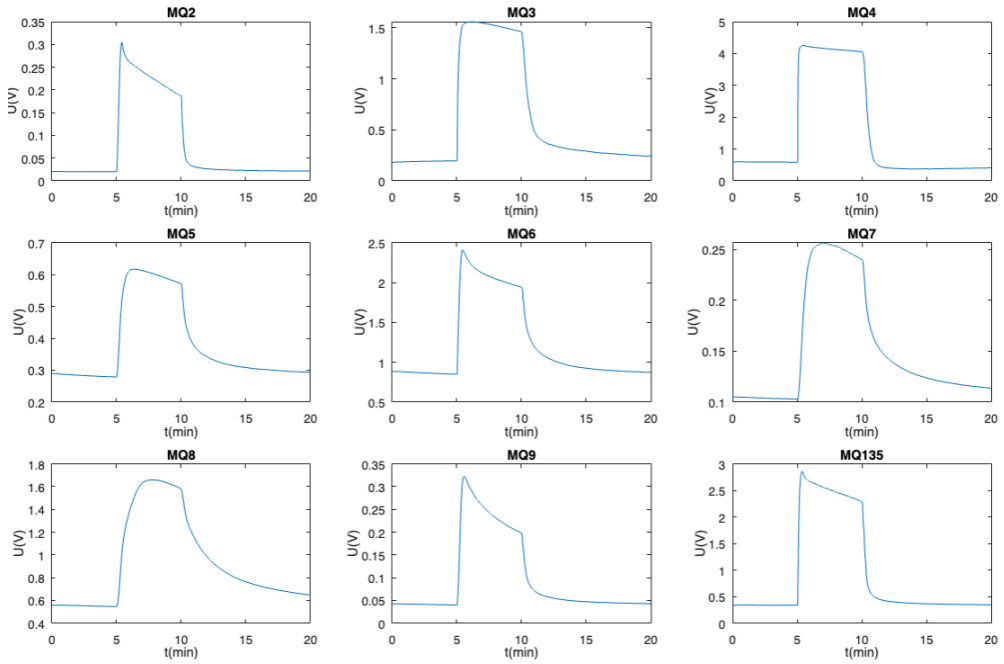
Odezva senzorů pro ethanol s koncentrací 5 ppm 2. měření



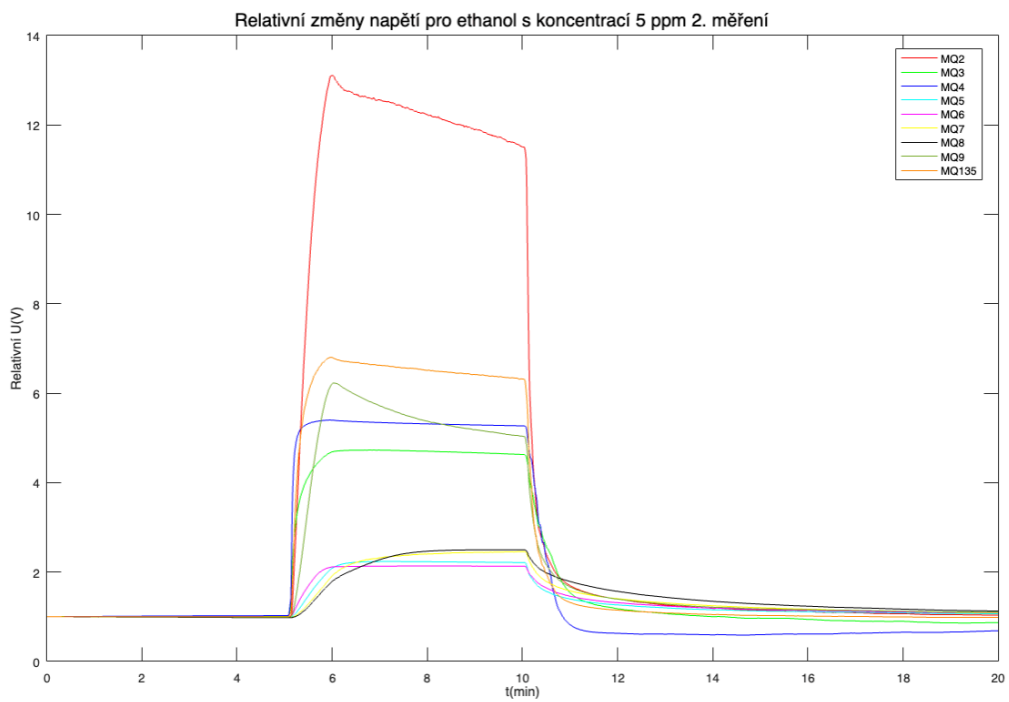
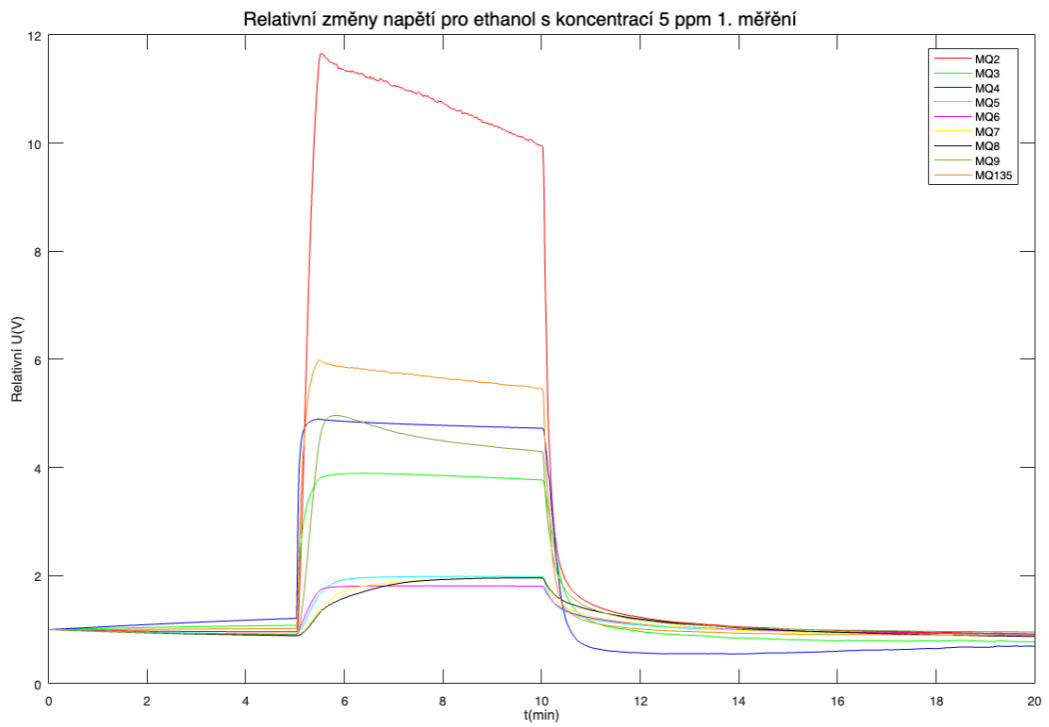
Odezva senzorů pro ethanol s koncentrací 5 ppm 3. měření

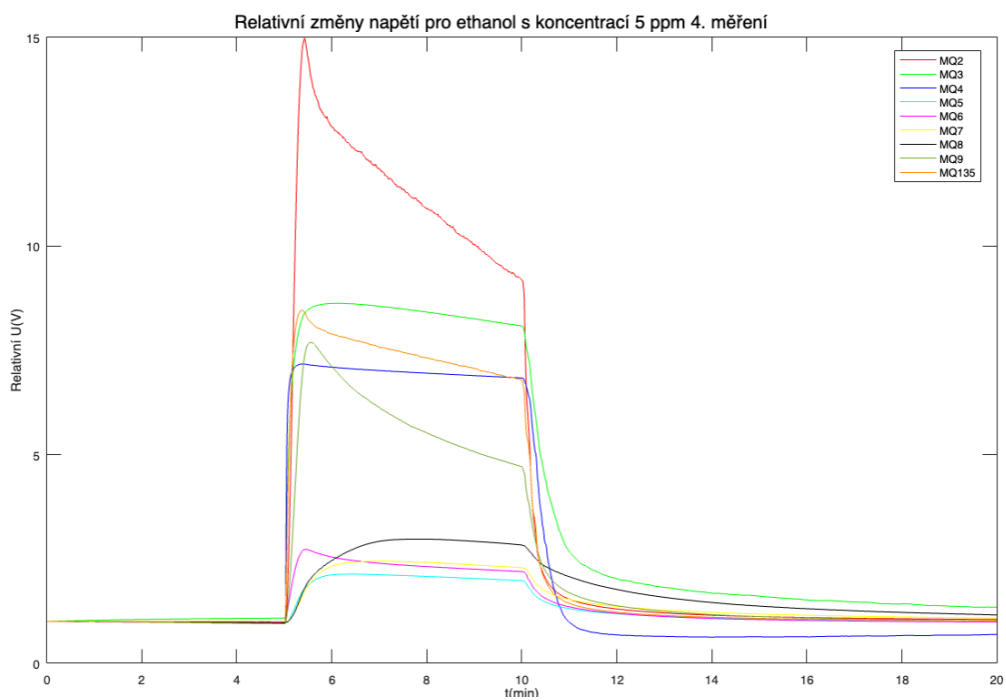
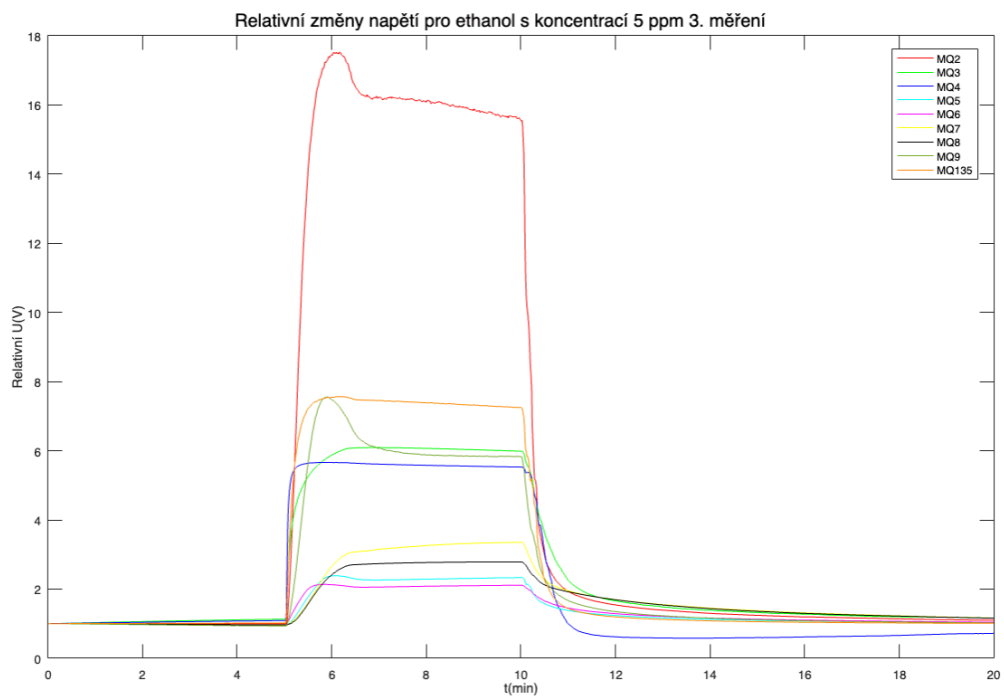


Odezva senzorů pro ethanol s koncentrací 5 ppm 4. měření



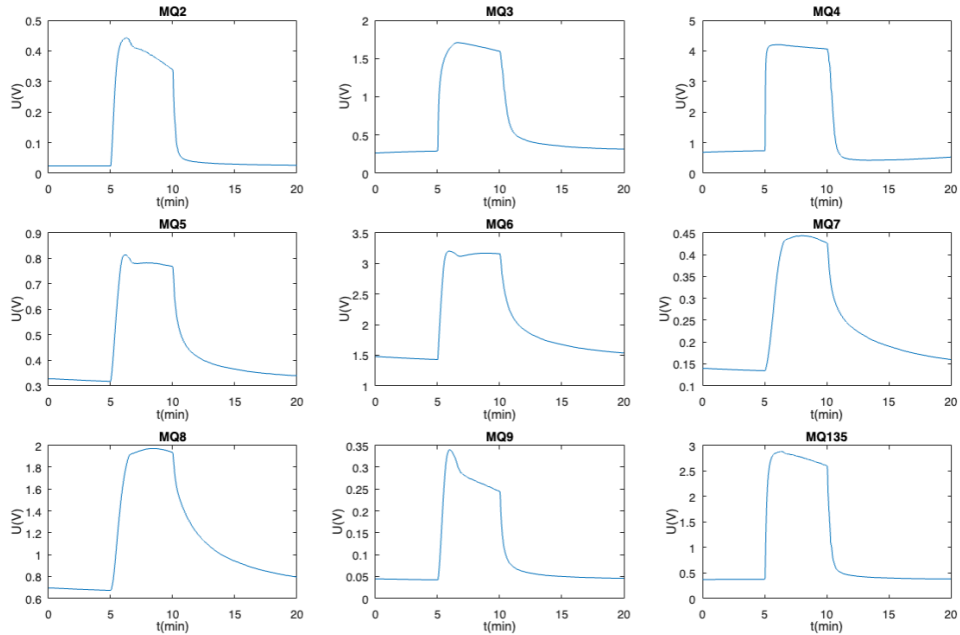
- Relativní změny časových odezev chemorezistivních senzorů při měření ethanolu s koncentrací 5 ppm.



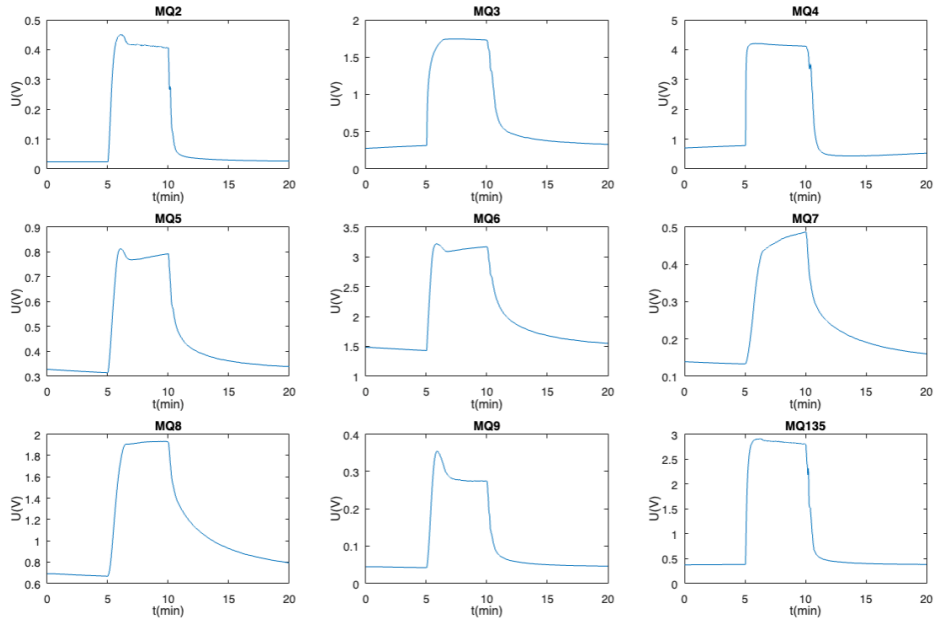


- Průběhy odezvy chemorezistivních senzorů při měření ethanolu s koncentrací 10 ppm.

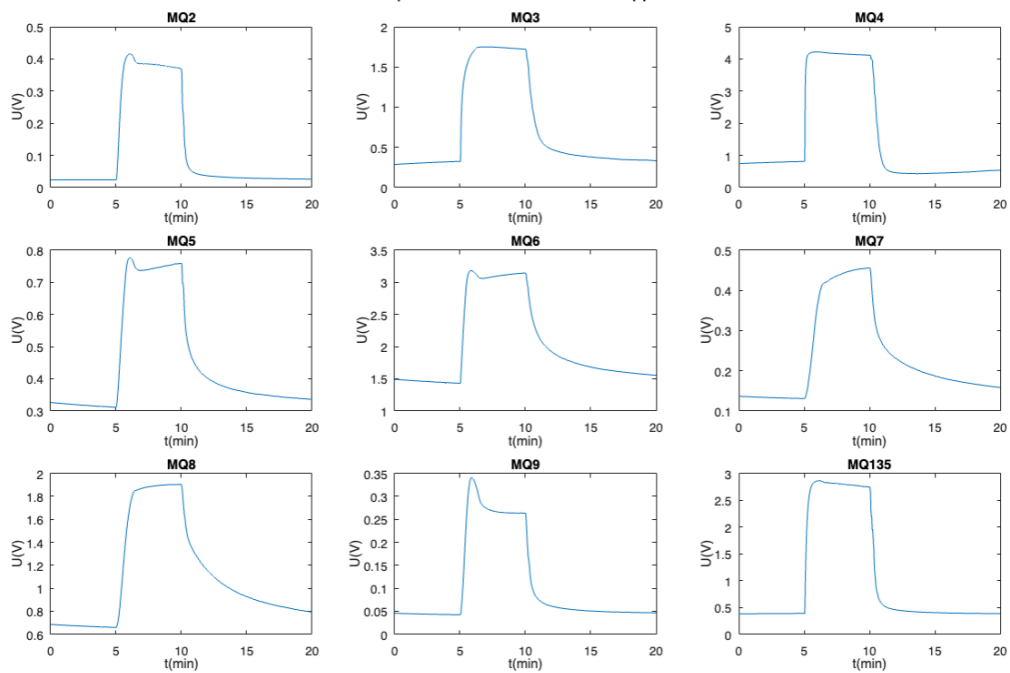
Odezva senzorů pro ethanol s koncentrací 10 ppm 1. měření



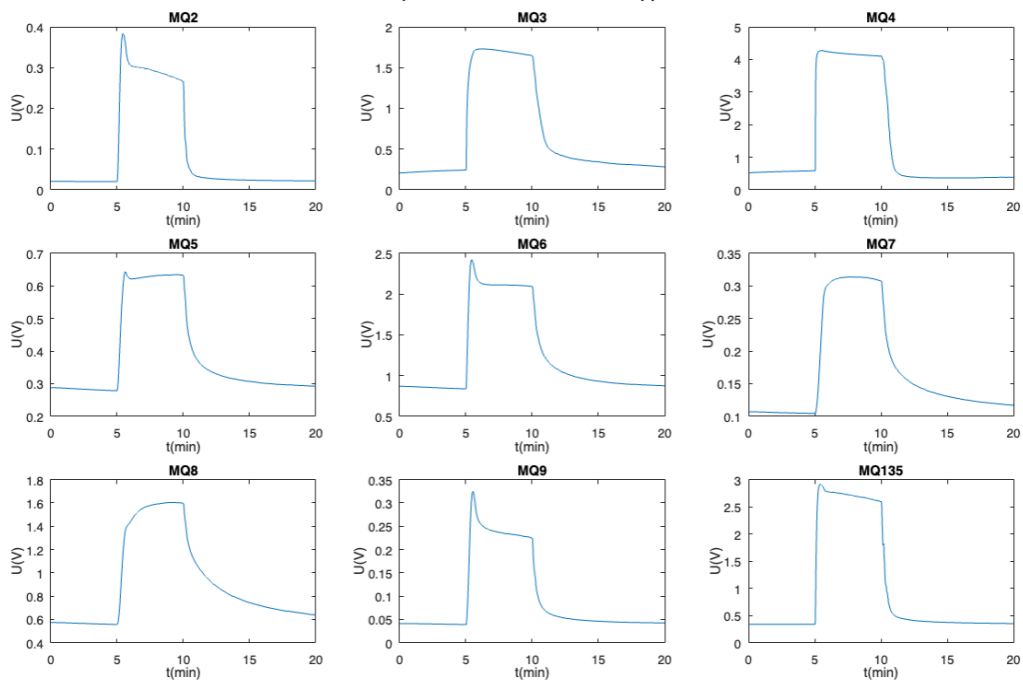
Odezva senzorů pro ethanol s koncentrací 10 ppm 2. měření



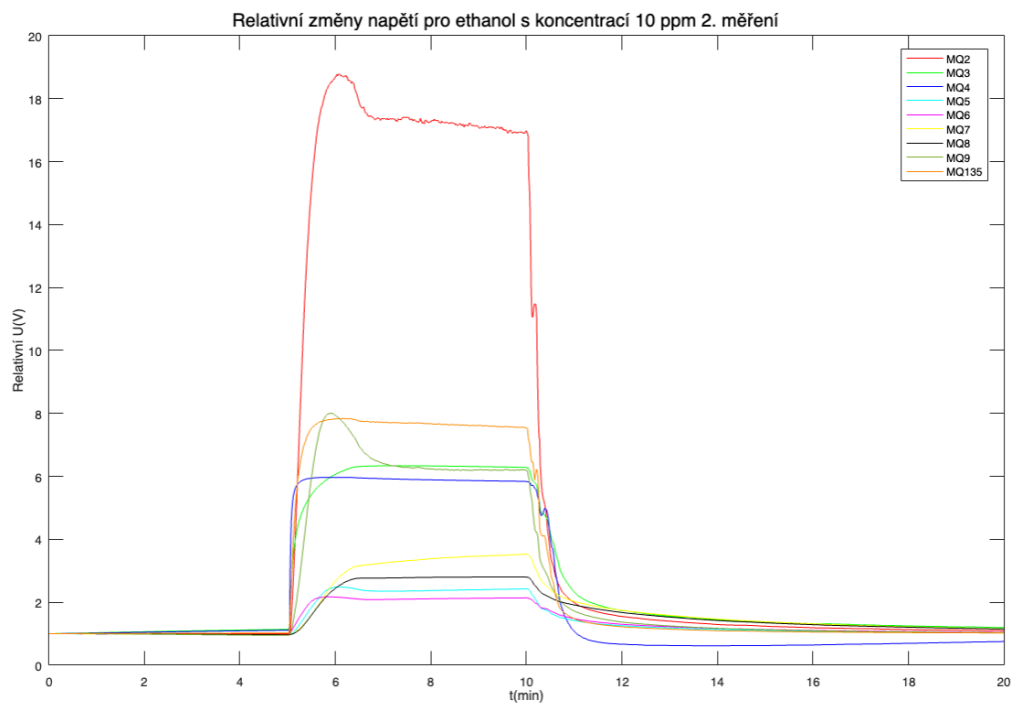
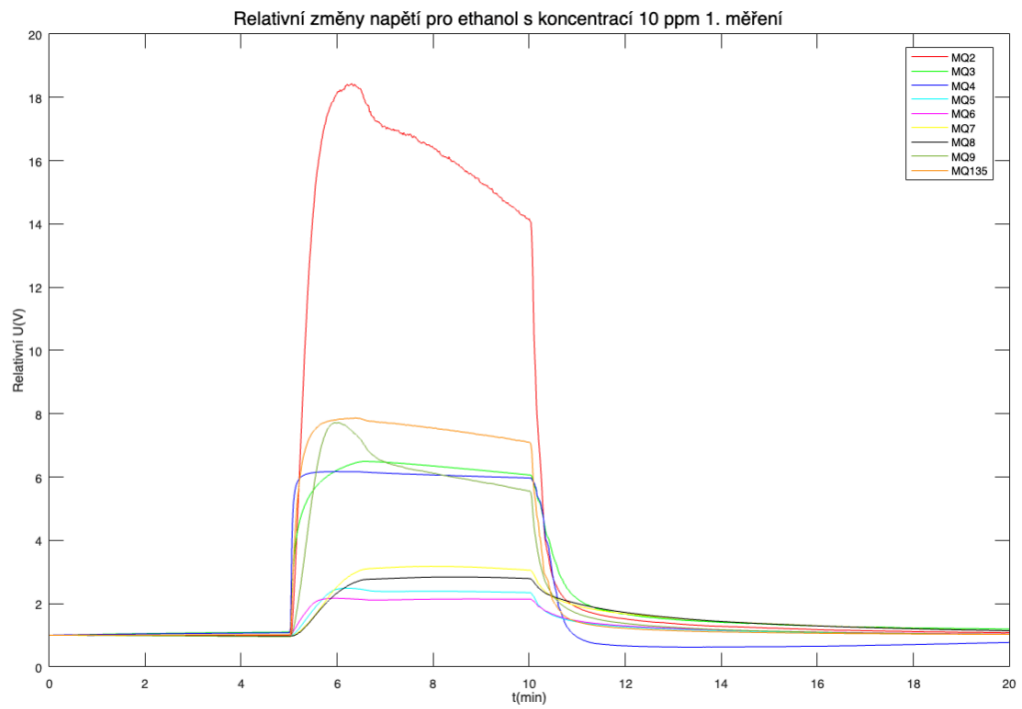
Odezva senzorů pro ethanol s koncentrací 10 ppm 3. měření

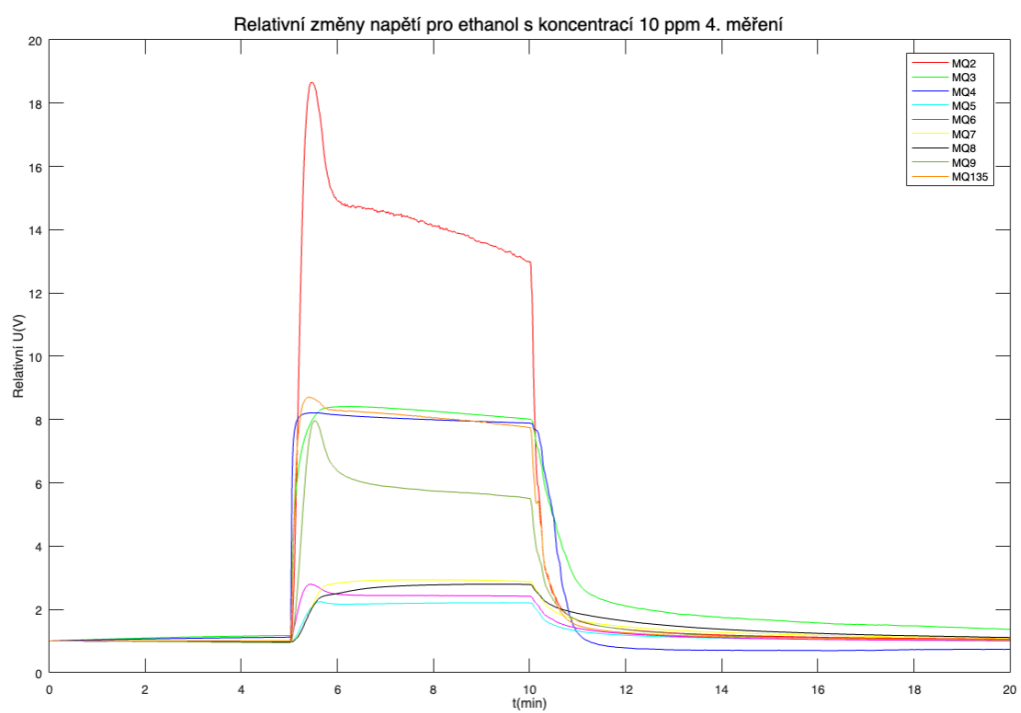
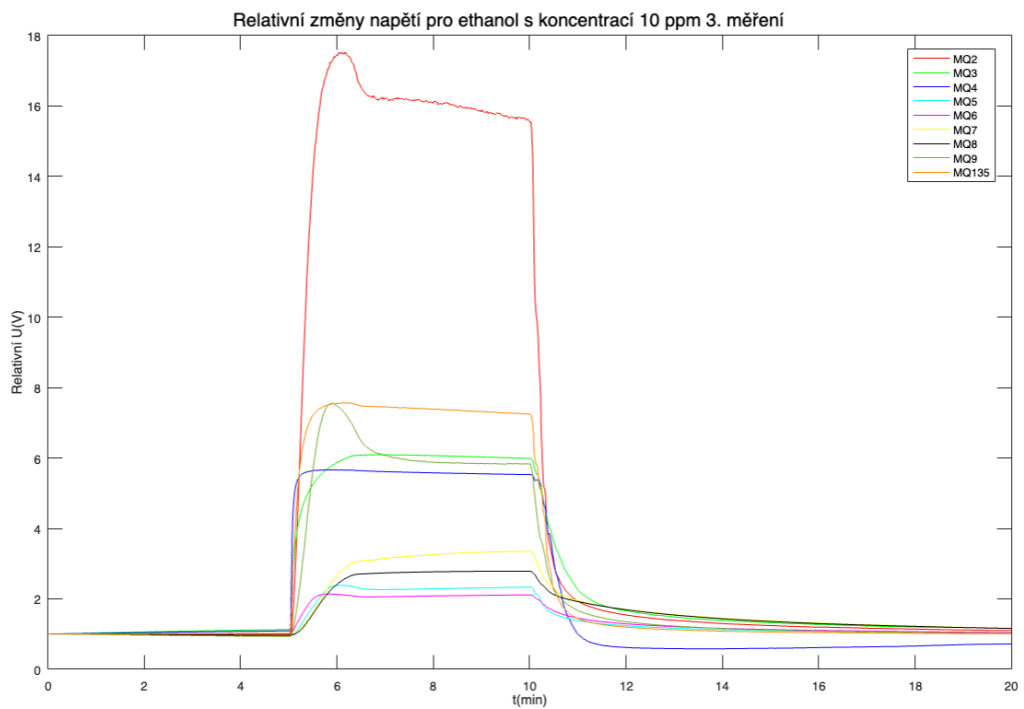


Odezva senzorů pro ethanol s koncentrací 10 ppm 4. měření



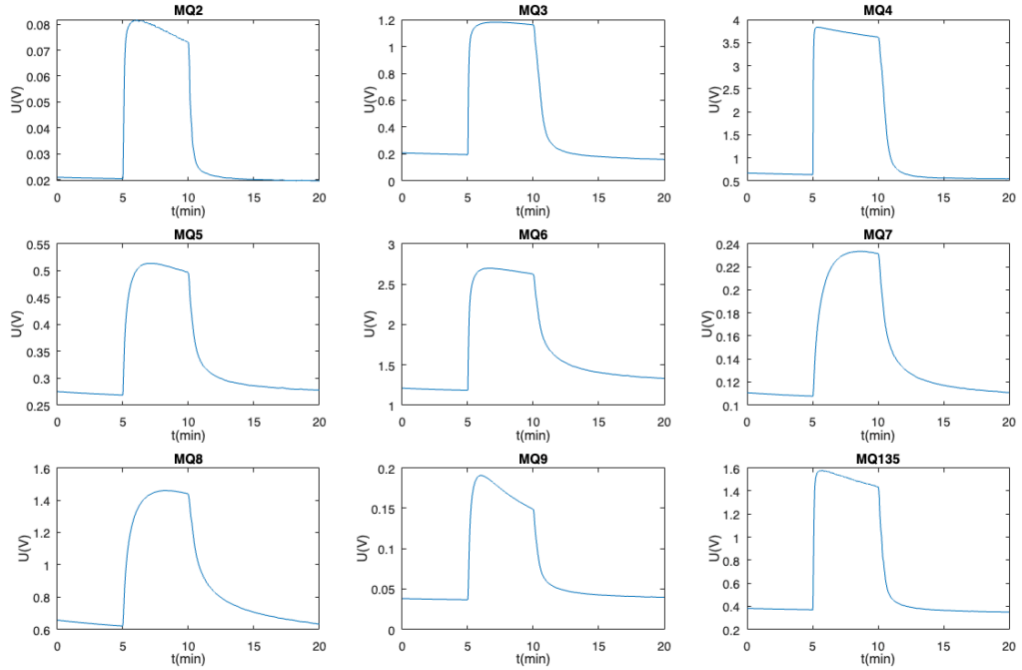
- Relativní změny časových odezev chemorezistivních senzorů při měření ethanolu s koncentrací 10 ppm.



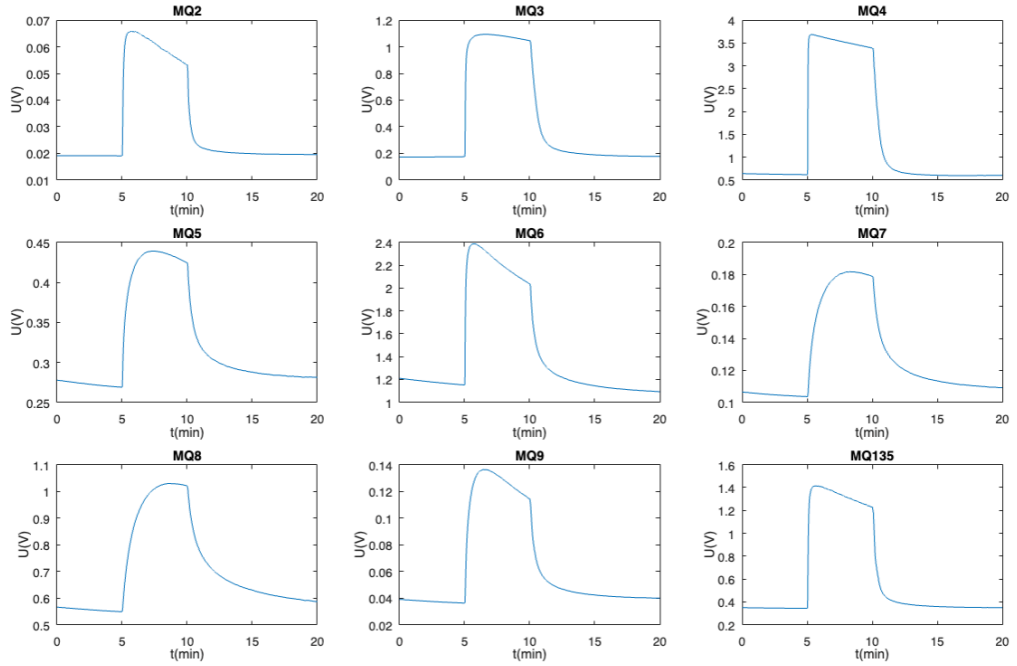


- Průběhy odezvy chemorezistivních senzorů při měření acetonu s koncentrací 1 ppm.

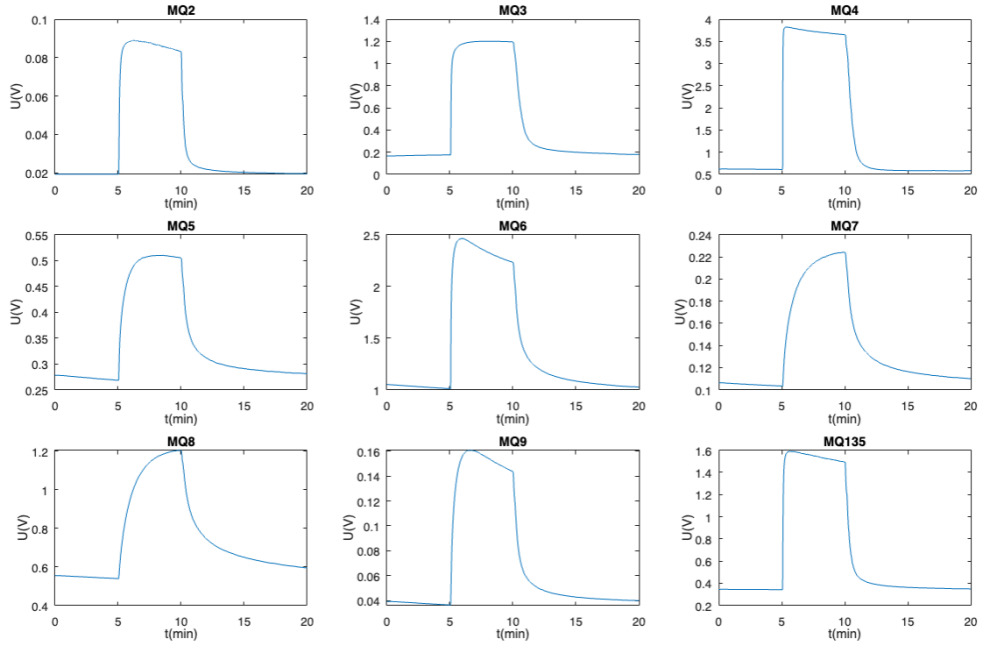
Odezva senzorů pro aceton s koncentrací 1 ppm 1. měření



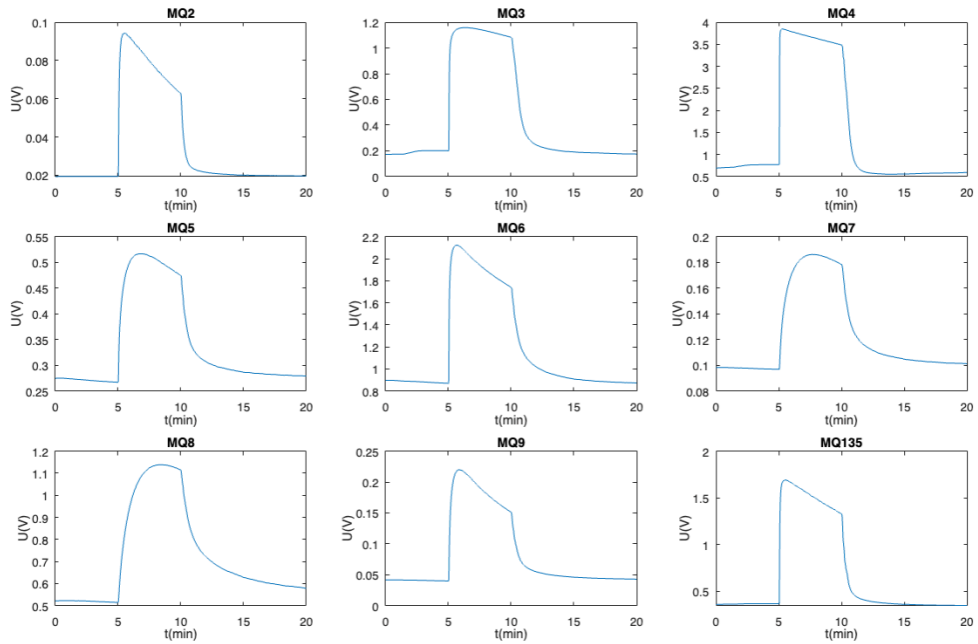
Odezva senzorů pro aceton s koncentrací 1 ppm 2. měření



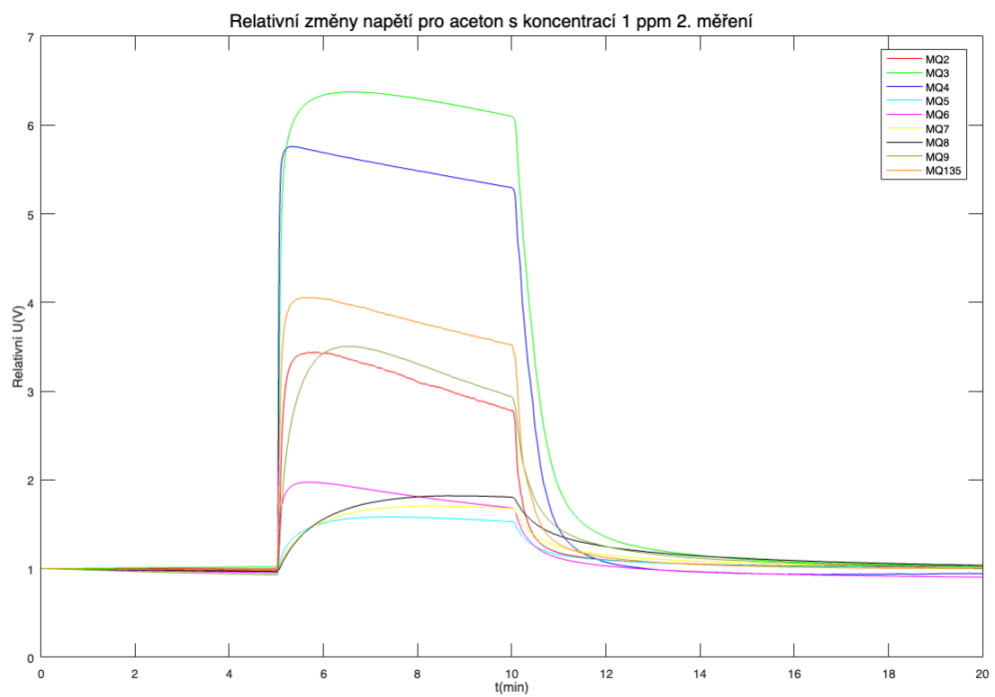
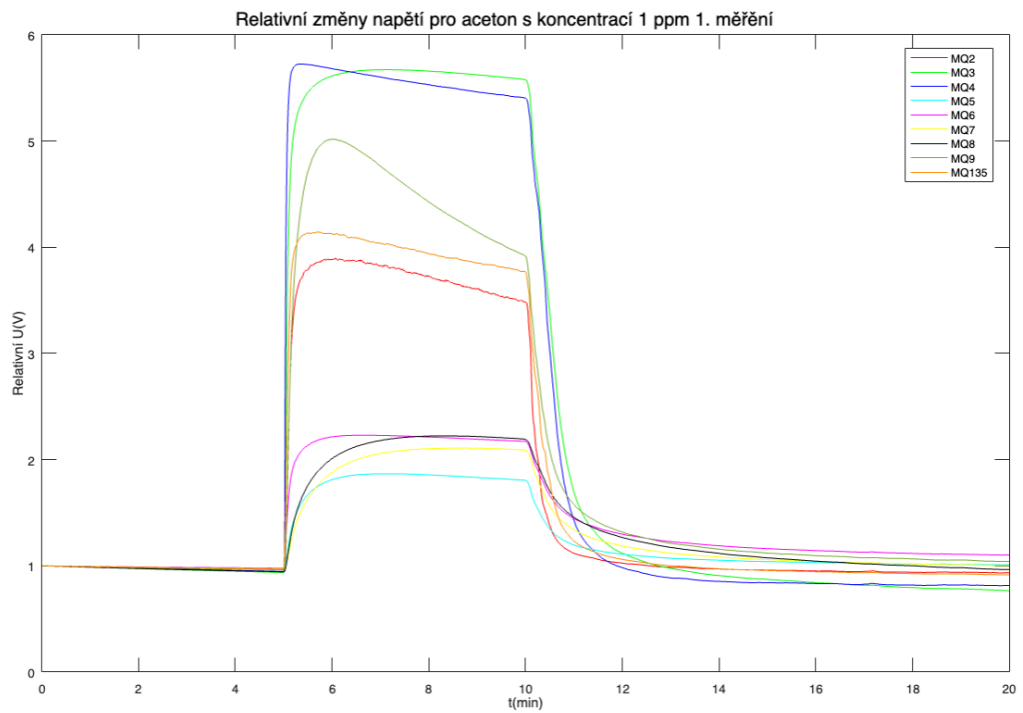
Odezva senzorů pro aceton s koncentrací 1 ppm 3. měření

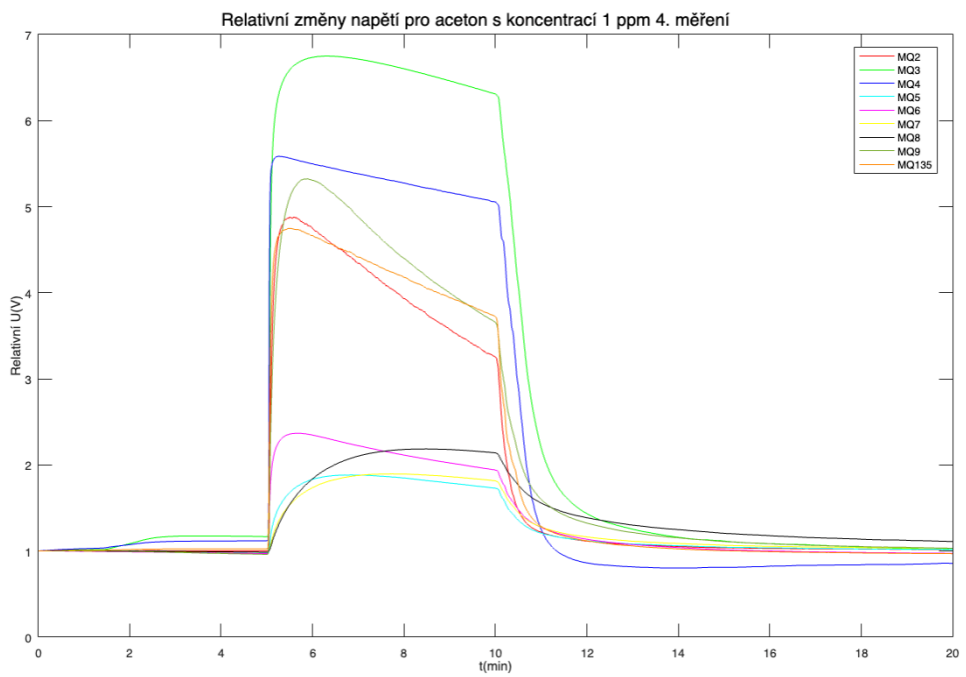
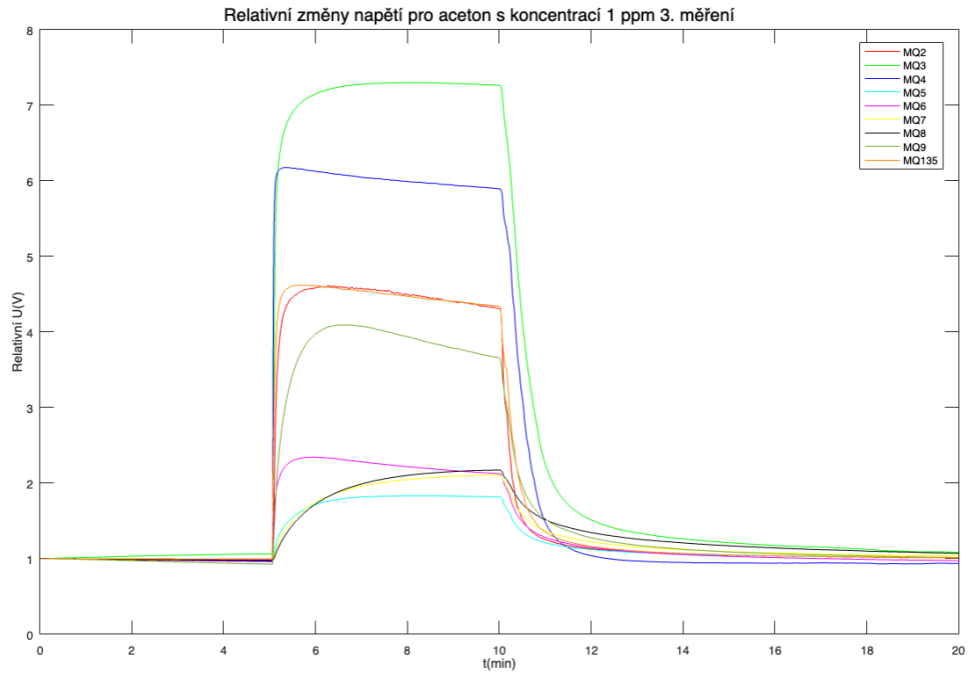


Odezva senzorů pro aceton s koncentrací 1 ppm 4. měření



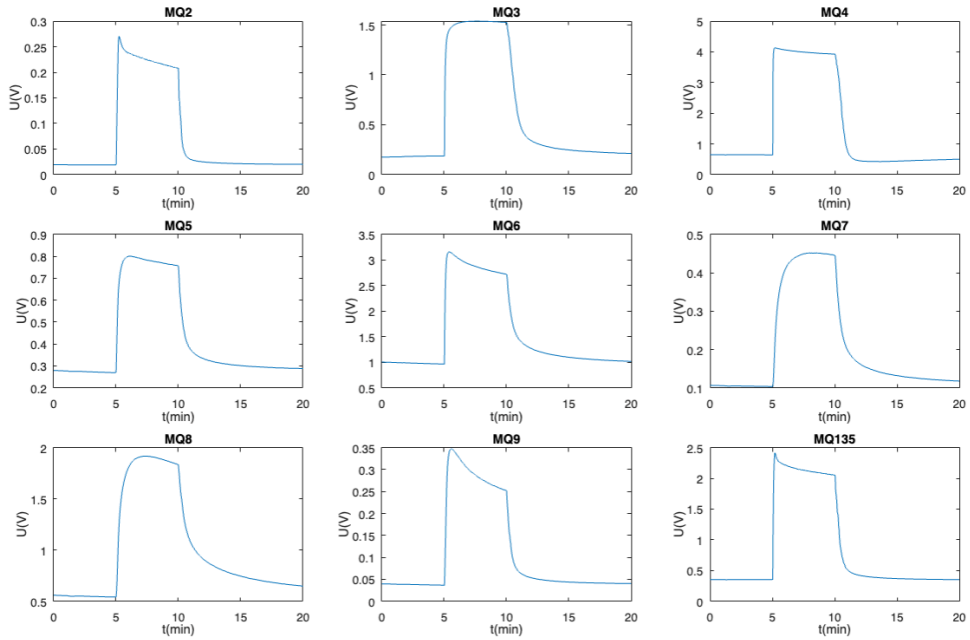
- Relativní změny časových odezev chemorezistivních senzorů při měření acetonu s koncentrací 1 ppm.



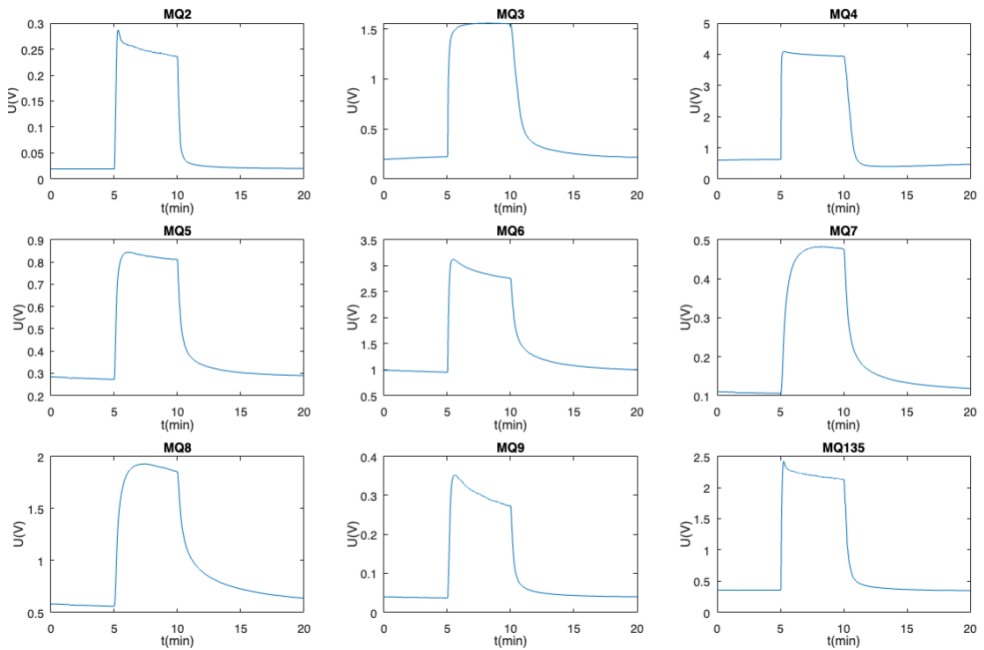


- Průběhy odezvy chemorezistivních senzorů při měření acetonu s koncentrací 5 ppm.

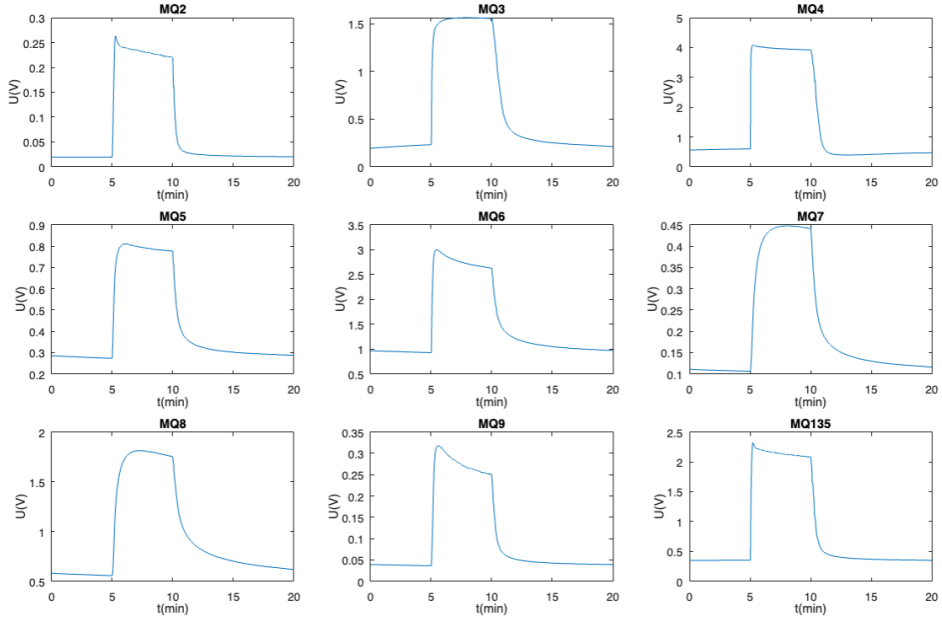
Odezva senzorů pro aceton s koncentrací 5 ppm 1. měření



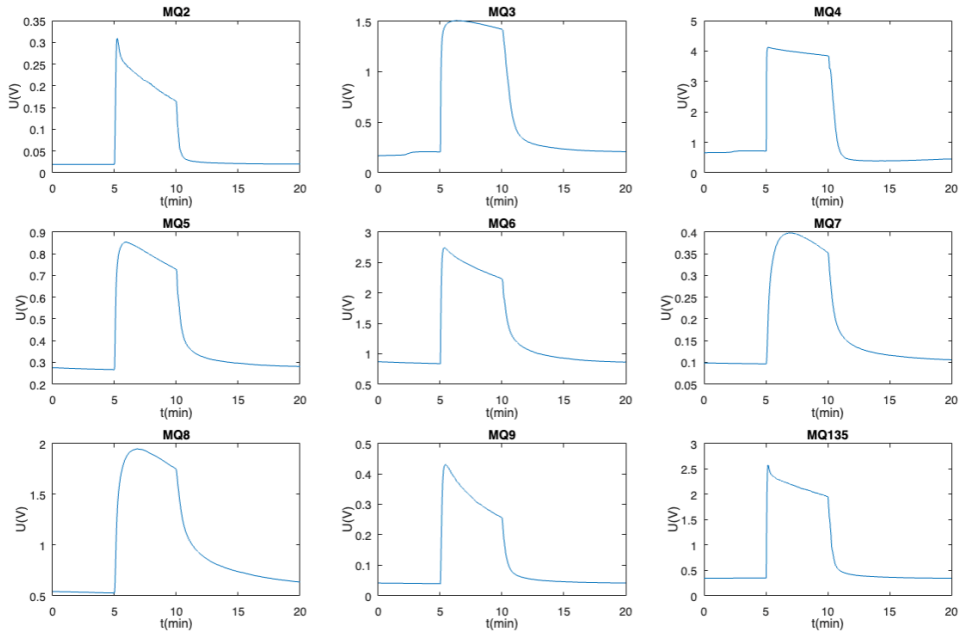
Odezva senzorů pro aceton s koncentrací 5 ppm 2. měření



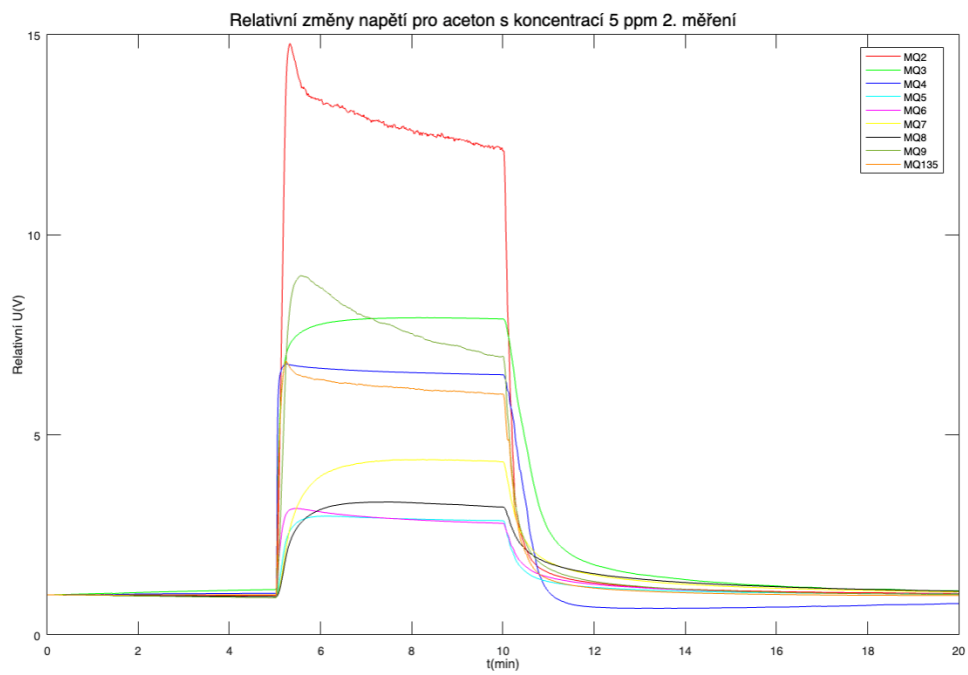
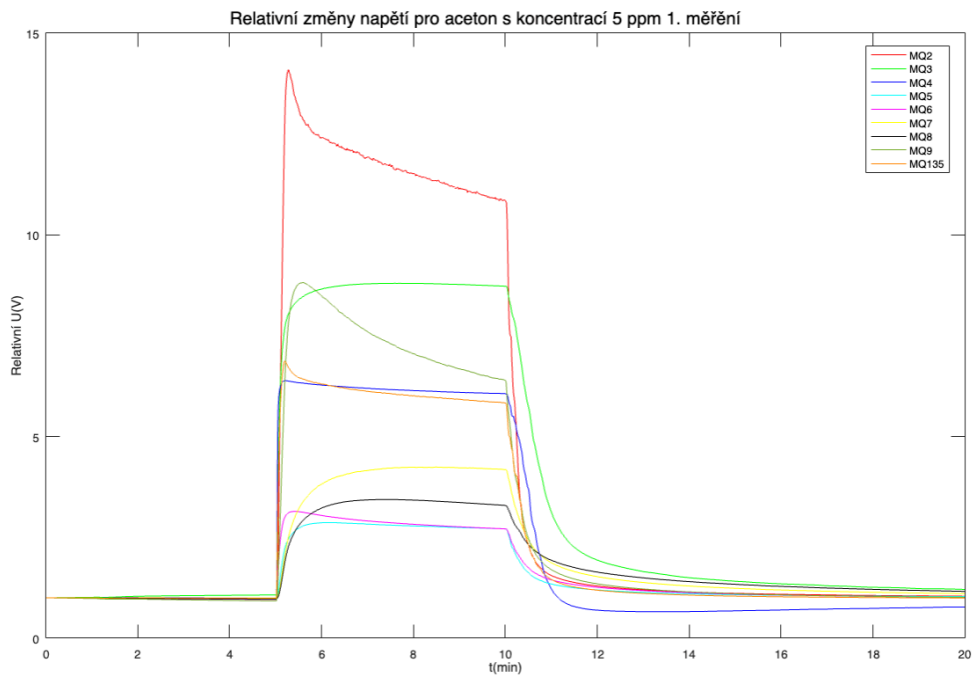
Odezva senzorů pro aceton s koncentrací 5 ppm 3. měření

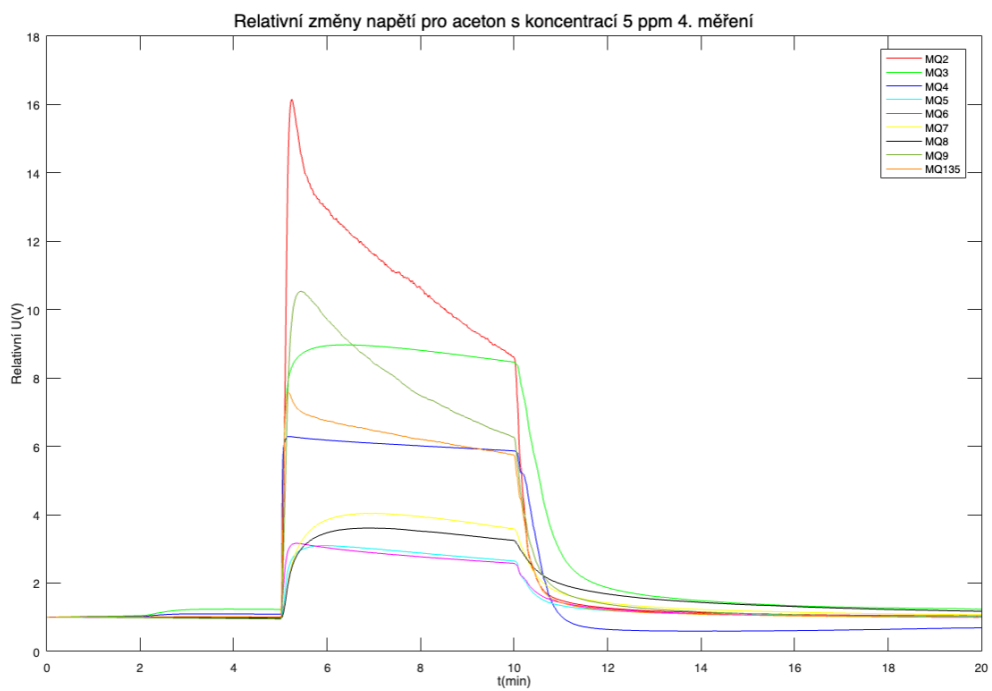
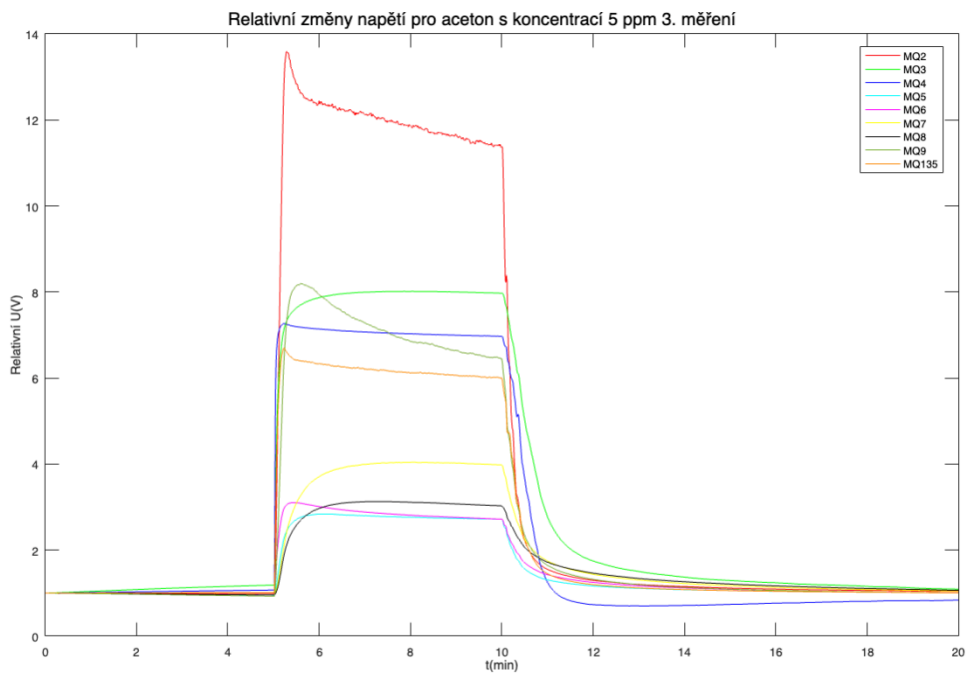


Odezva senzorů pro aceton s koncentrací 5 ppm 4. měření



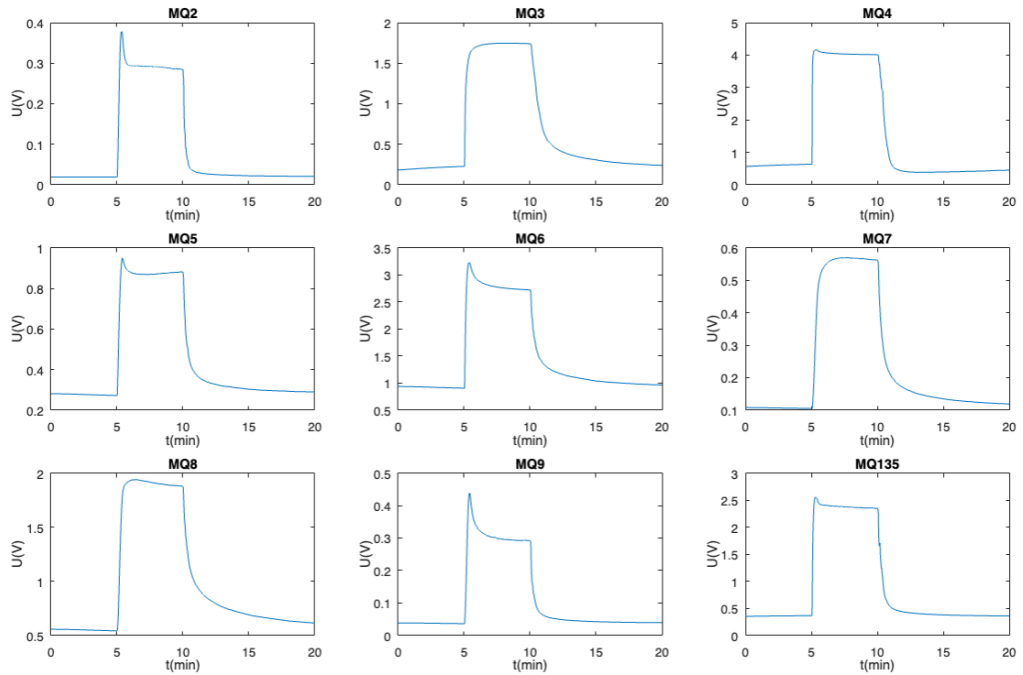
- Relativní změny časových odezev chemorezistivních senzorů při měření acetonu s koncentrací 5 ppm.



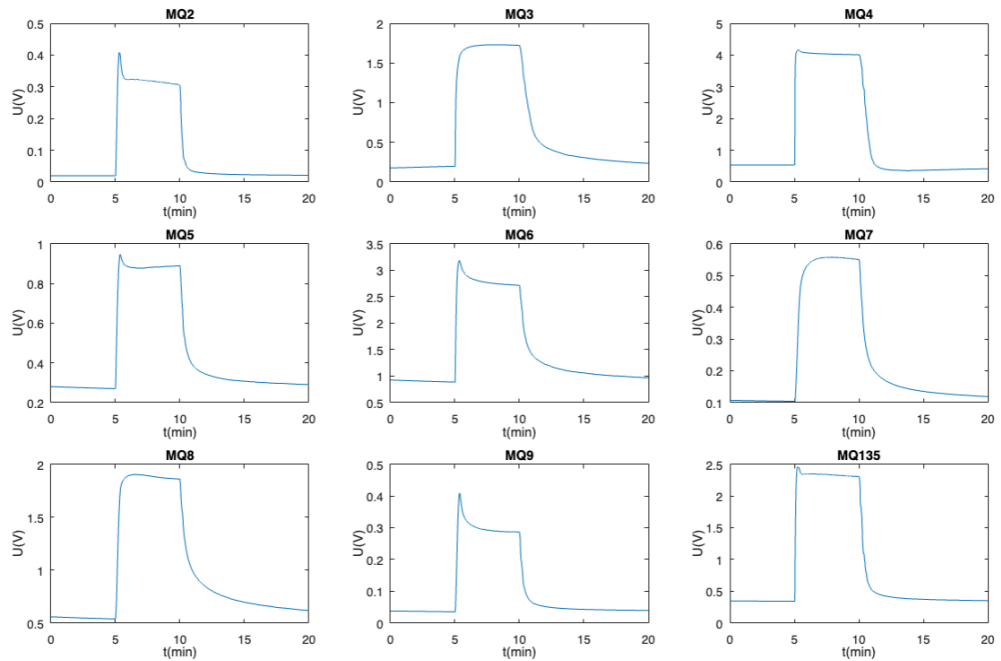


- Průběhy odezvy chemorezistivních senzorů při měření acetonu s koncentrací 10 ppm.

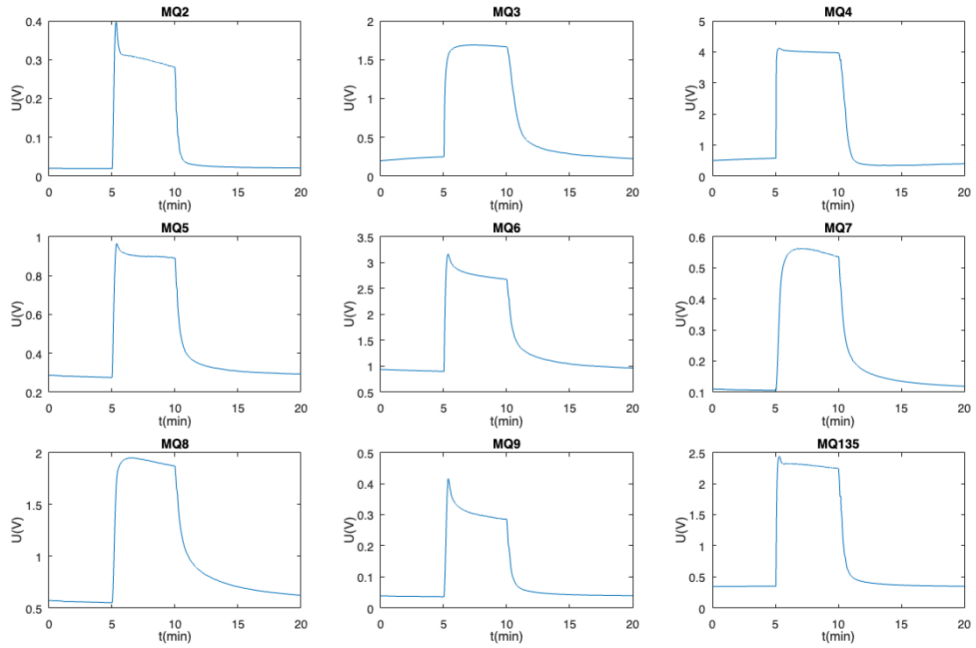
Odezva senzorů pro aceton s koncentrací 10 ppm 1. měření



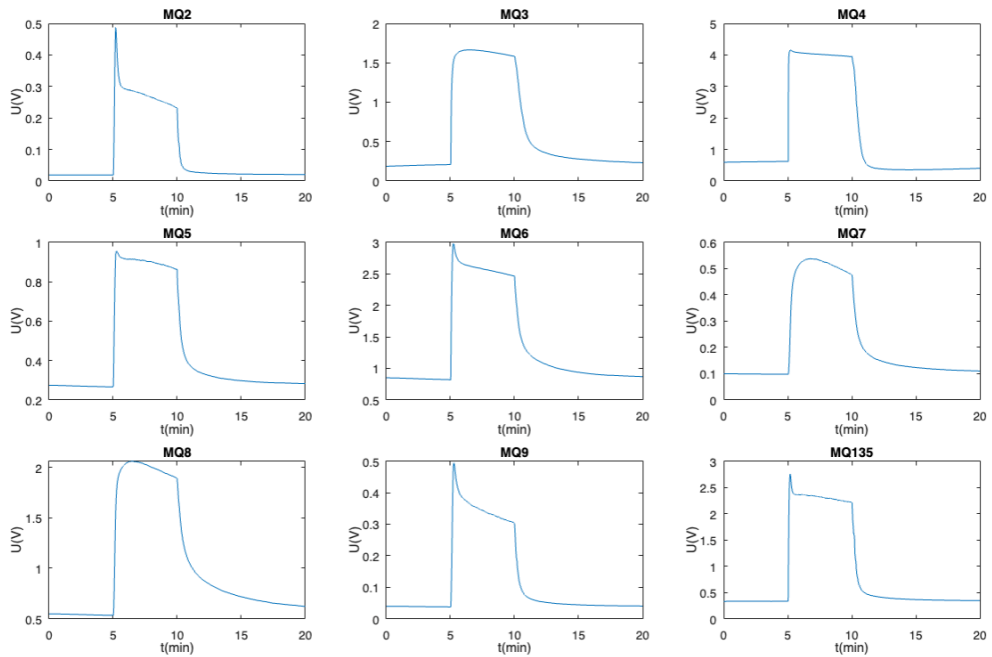
Odezva senzorů pro aceton s koncentrací 10 ppm 2. měření



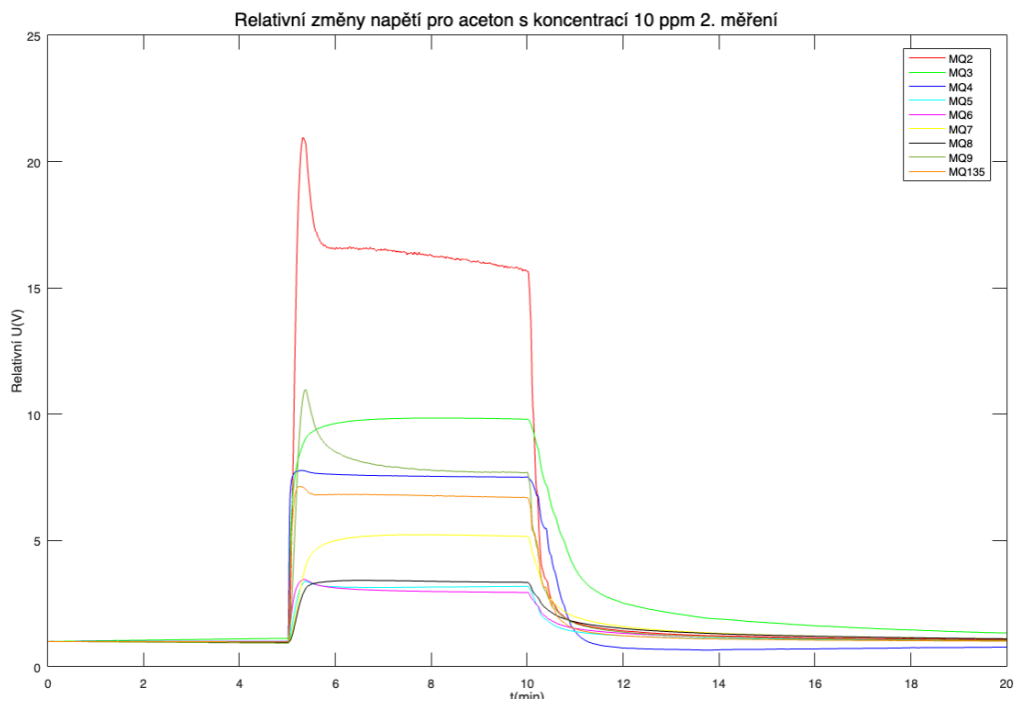
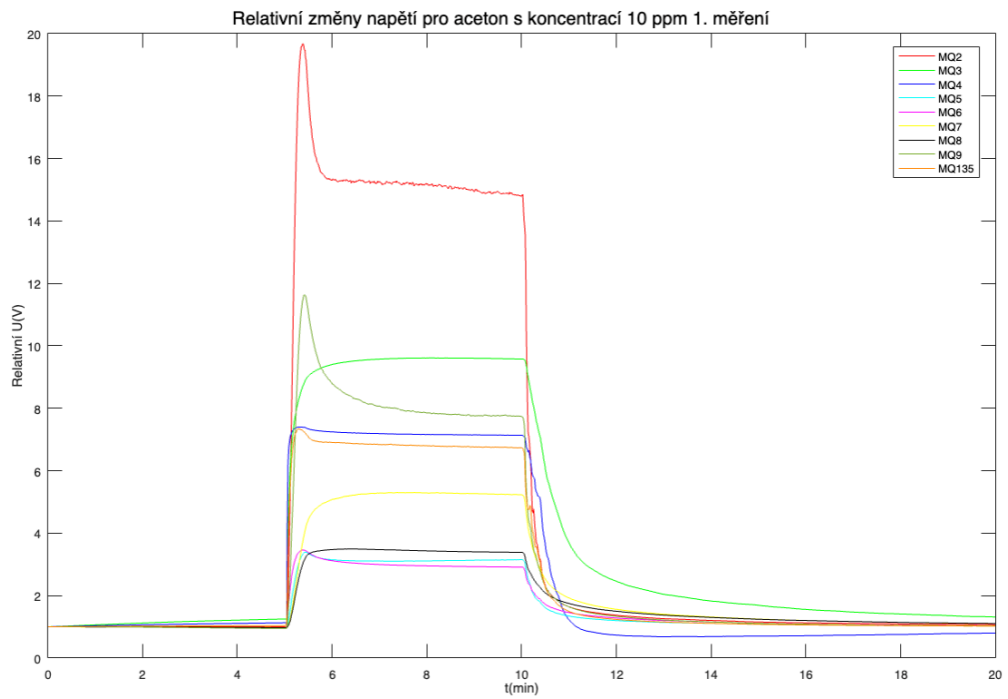
Odezva senzorů pro aceton s koncentrací 10 ppm 3. měření

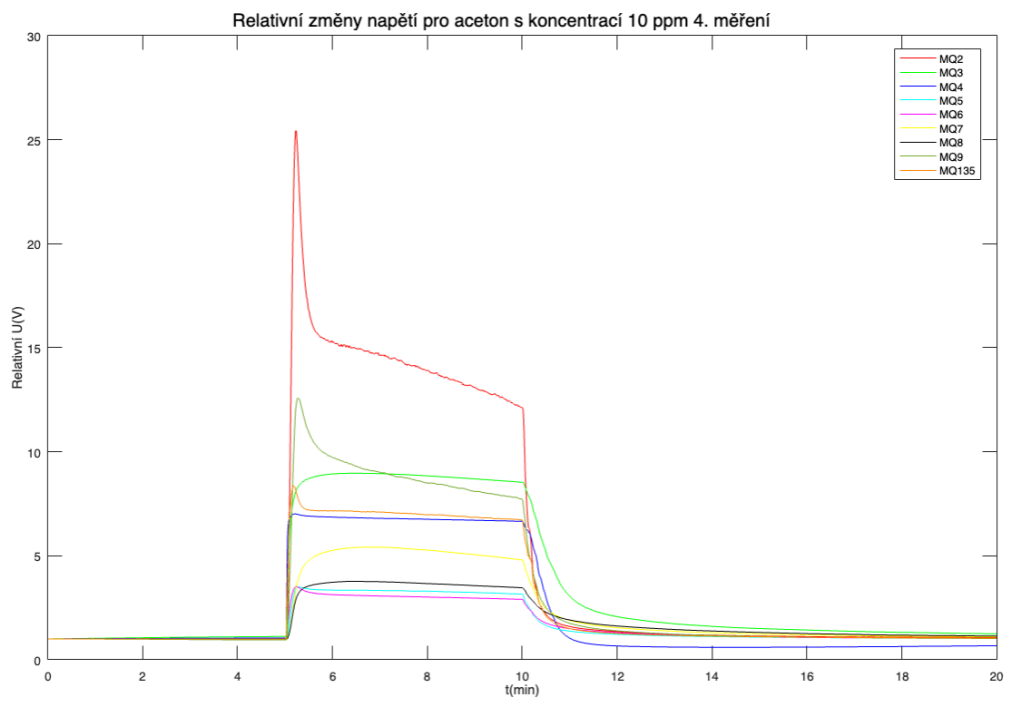
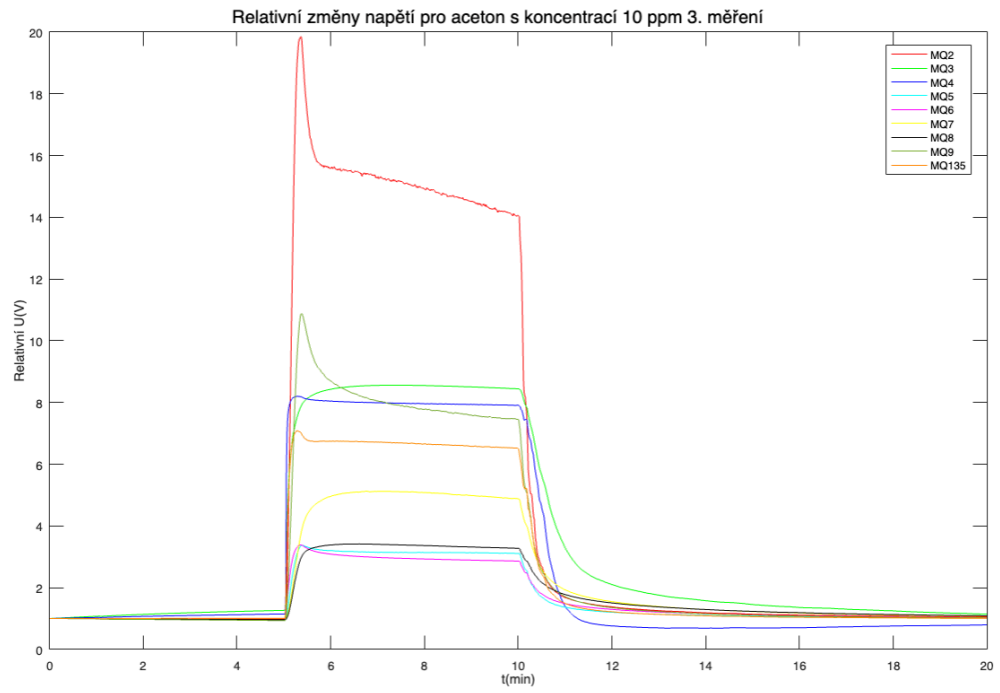


Odezva senzorů pro aceton s koncentrací 10 ppm 4. měření



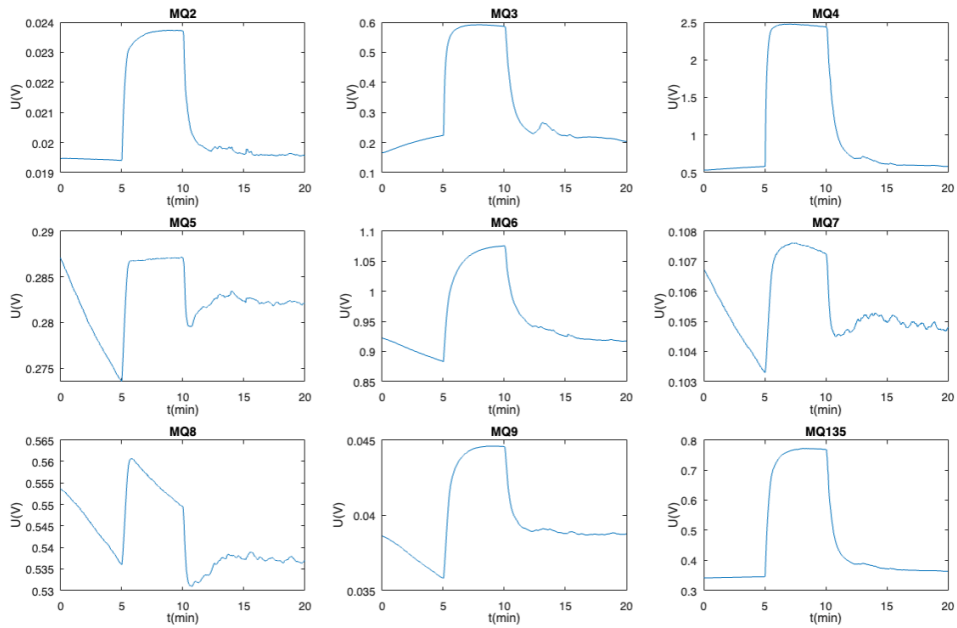
- Relativní změny časových odezev chemorezistivních senzorů při měření acetonu s koncentrací 10 ppm.



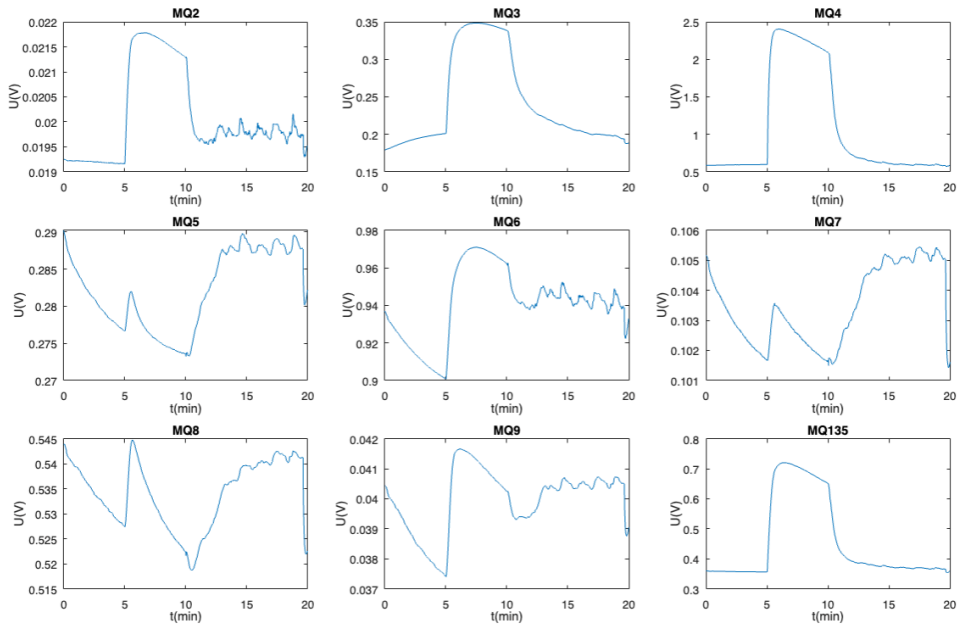


- Průběhy odezvy chemorezistivních senzorů při měření toluenu s koncentrací 1 ppm.

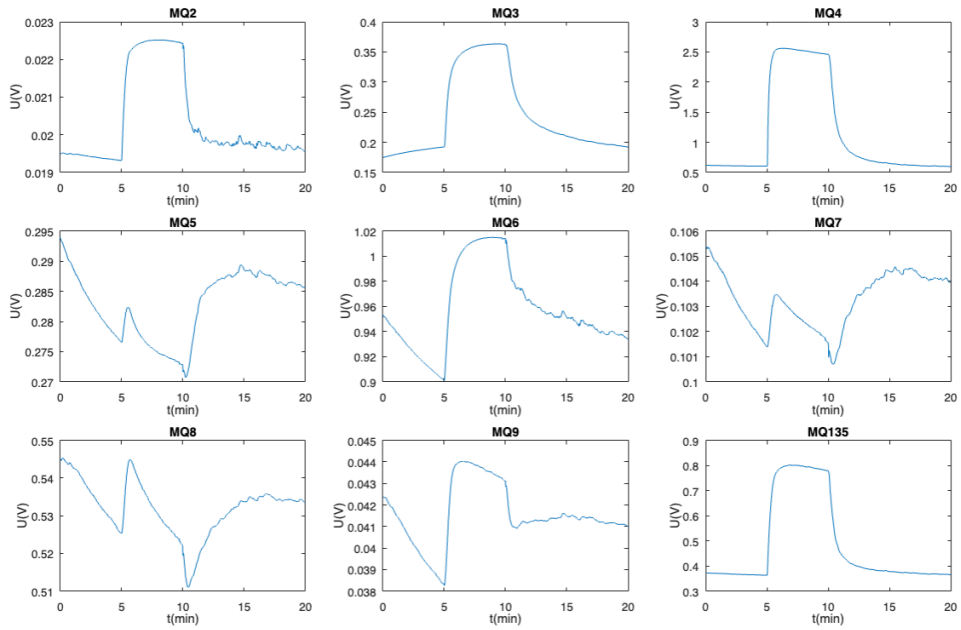
Odezva senzorů pro toluen s koncentrací 1 ppm 1. měření



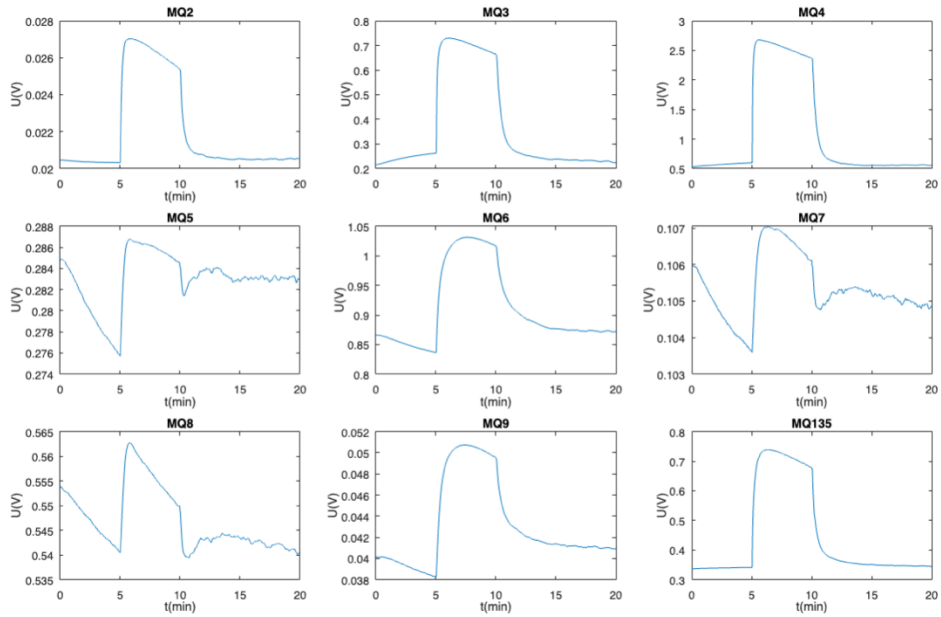
Odezva senzorů pro toluen s koncentrací 1 ppm 2. měření



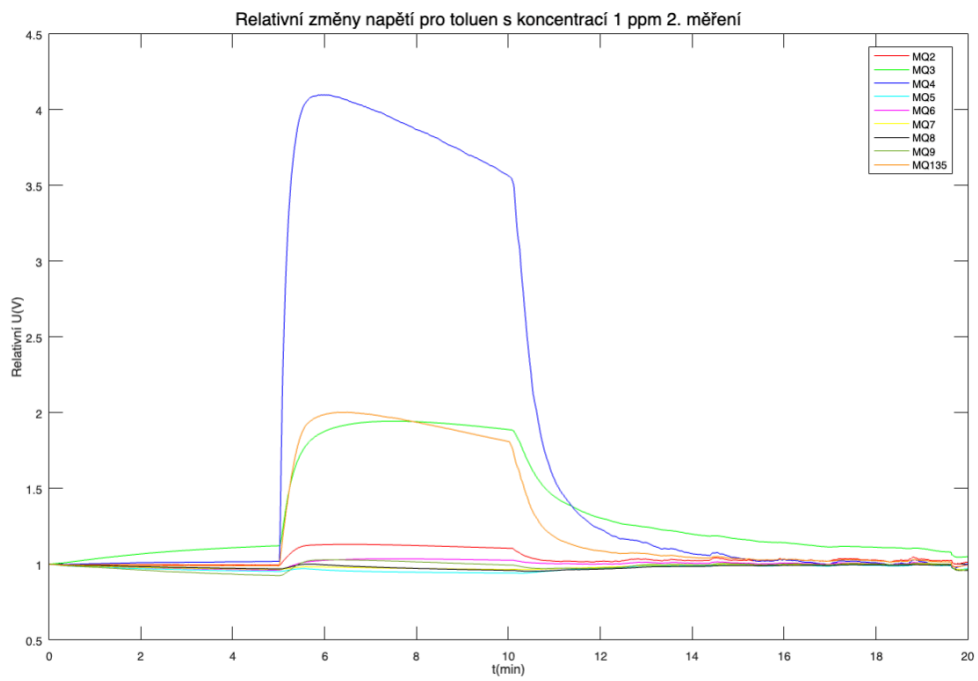
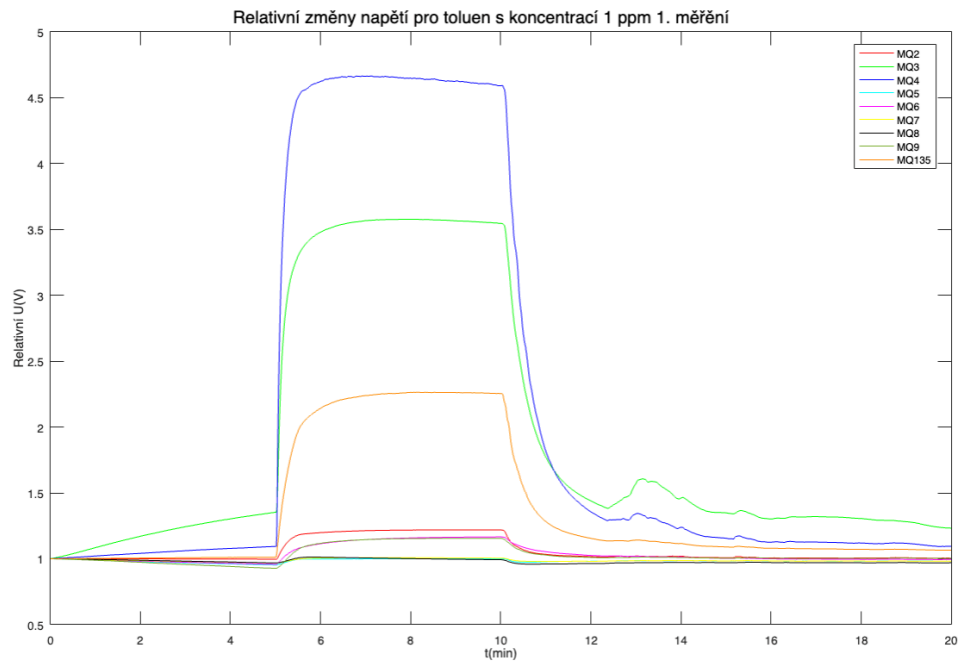
Odezva senzorů pro toluen s koncentrací 1 ppm 3. měření

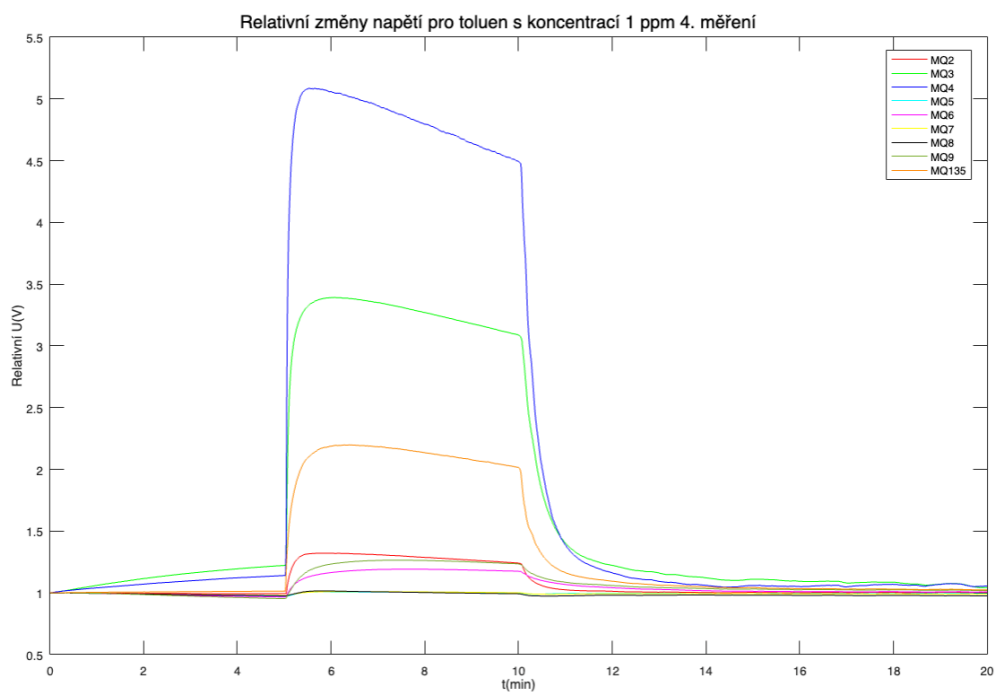
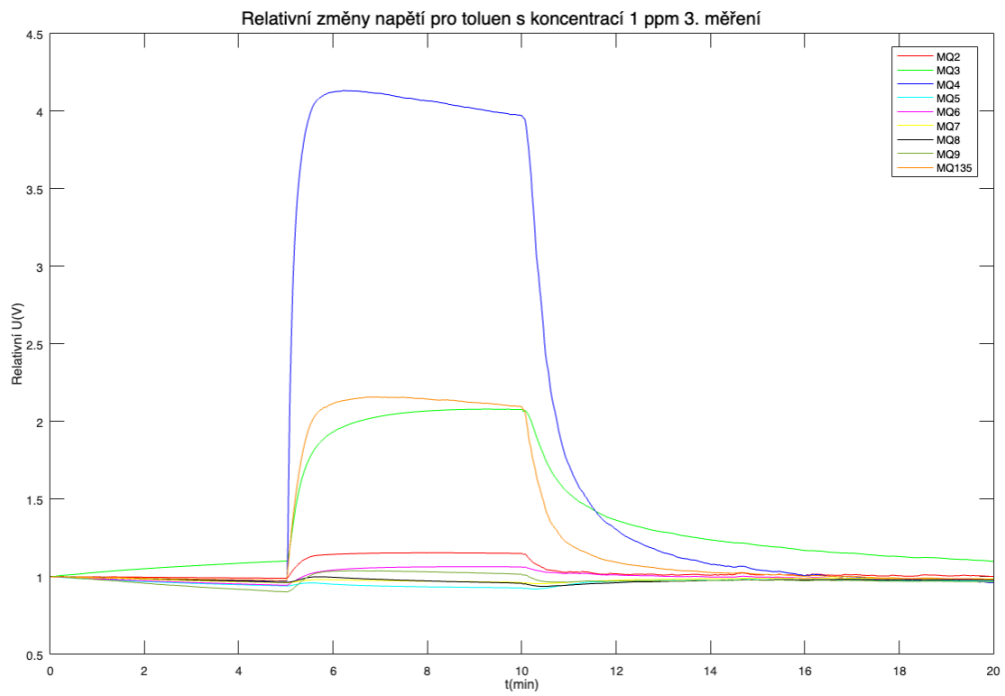


Odezva senzoru pro toluen s koncentrací 1 ppm 4. měření



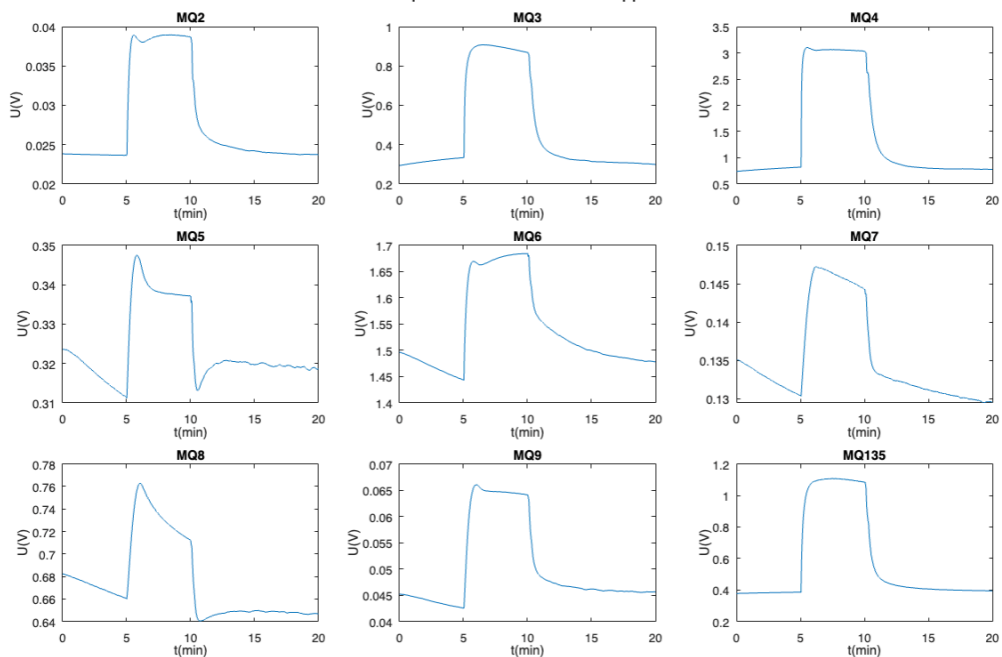
- Relativní změny časových odezev chemorezistivních senzorů při měření toluenu s koncentrací 1 ppm.



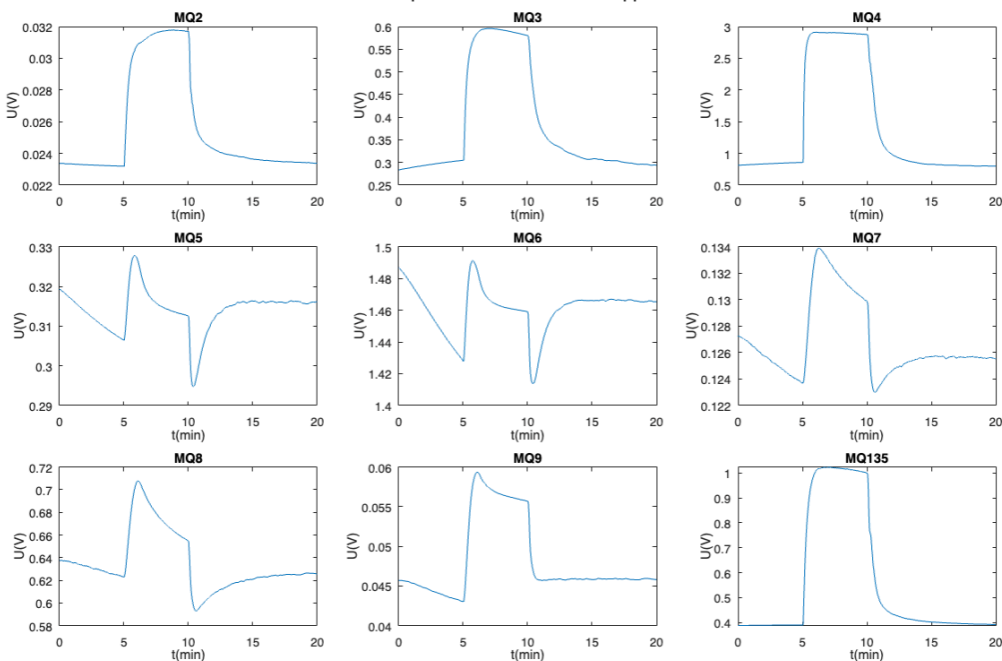


- Průběhy odezvy chemorezistivních senzorů při měření toluenu s koncentrací 5 ppm.

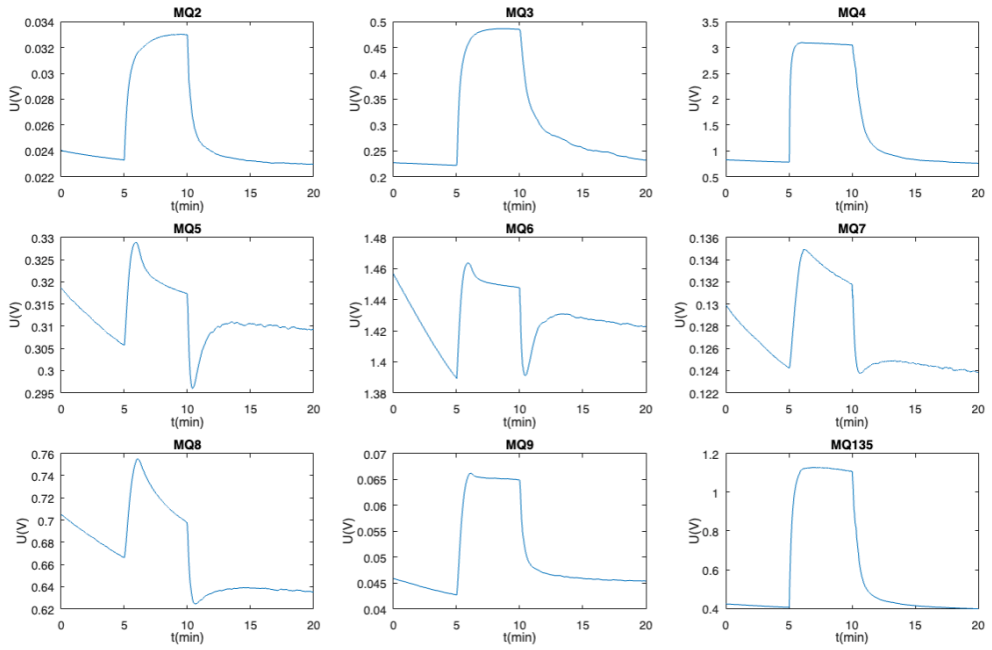
Odezva senzorů pro toluen s koncentrací 5 ppm 1. měření



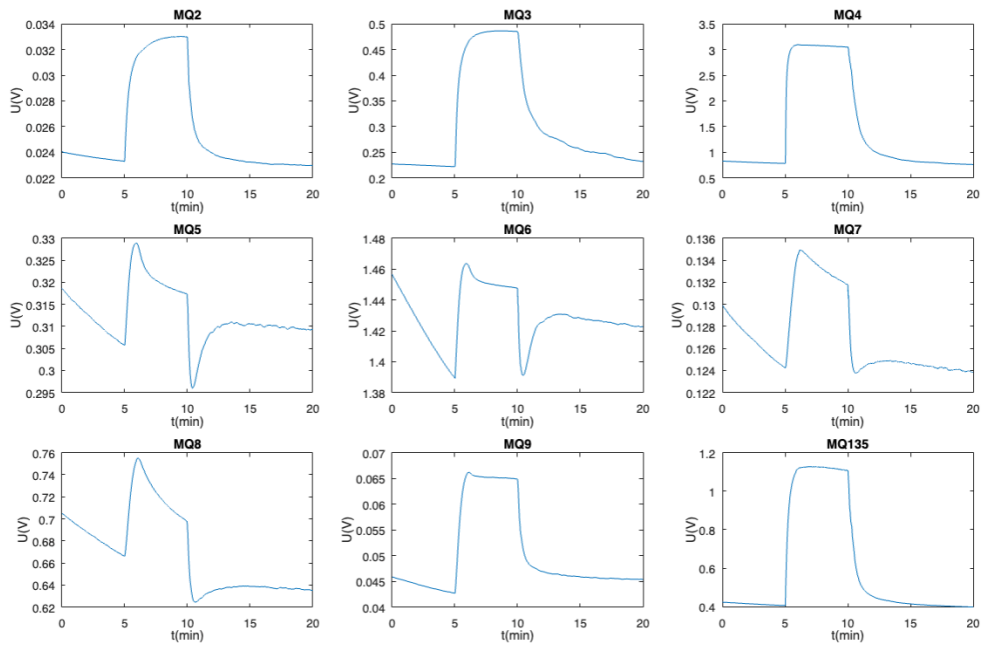
Odezva senzorů pro toluen s koncentrací 5 ppm 2. měření



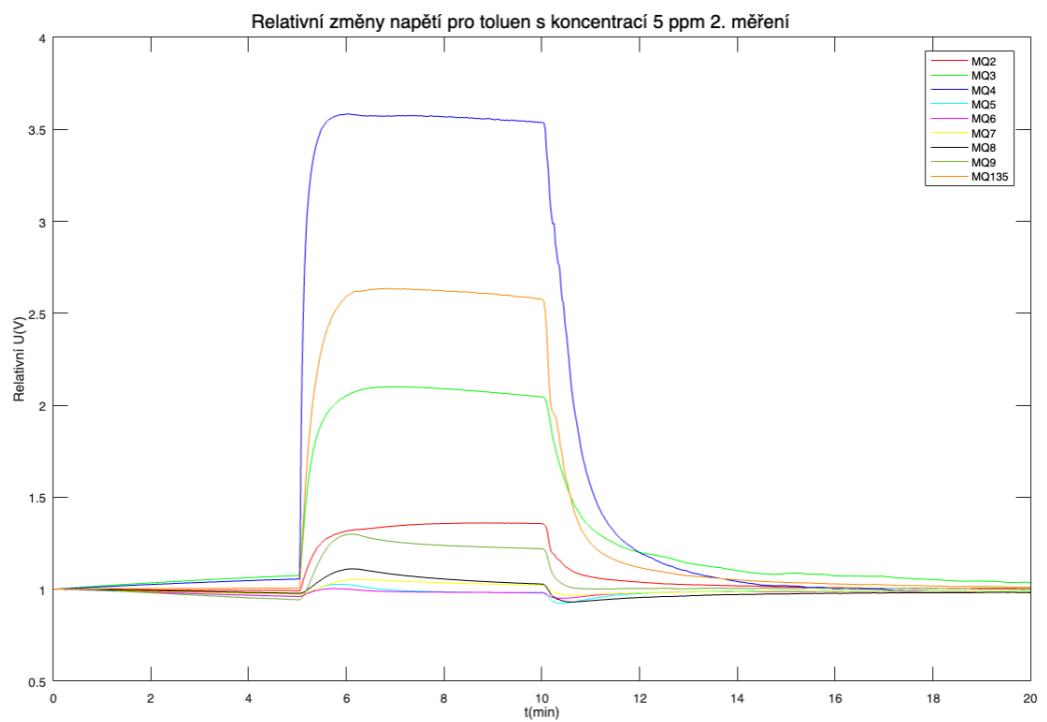
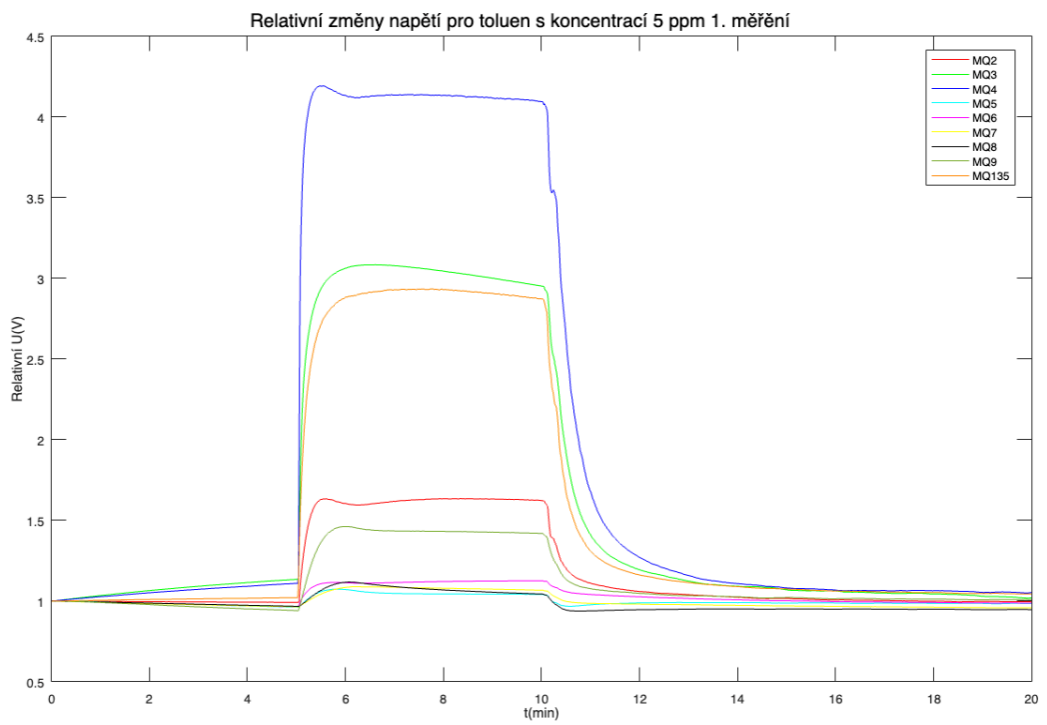
Odezva senzorů pro toluen s koncentrací 5 ppm 3. měření

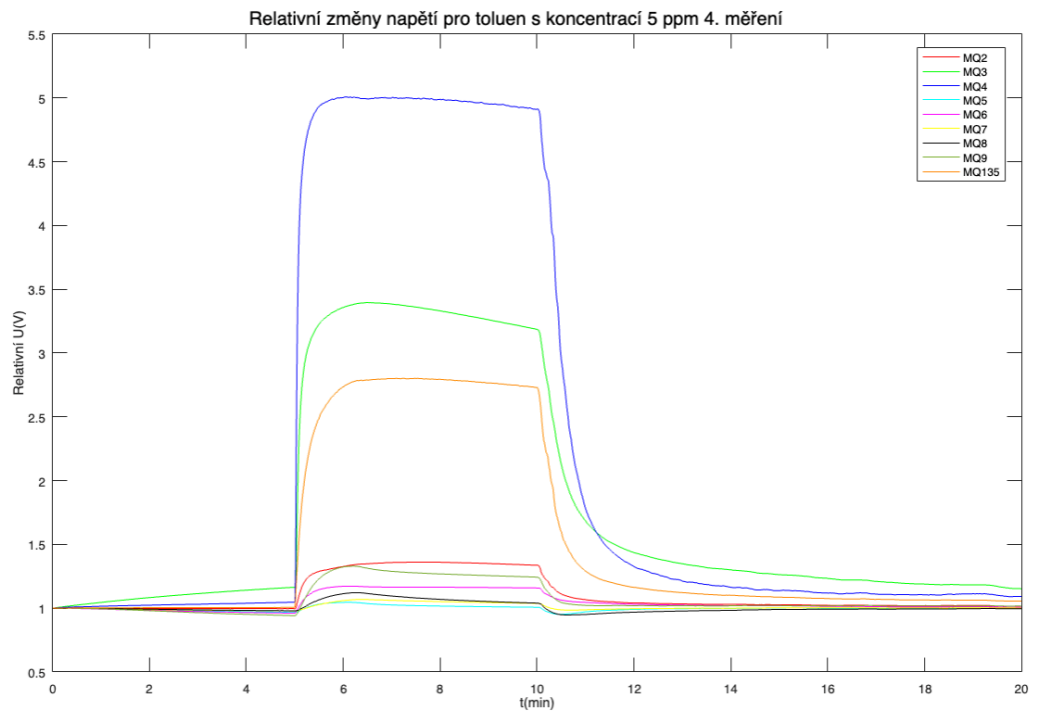
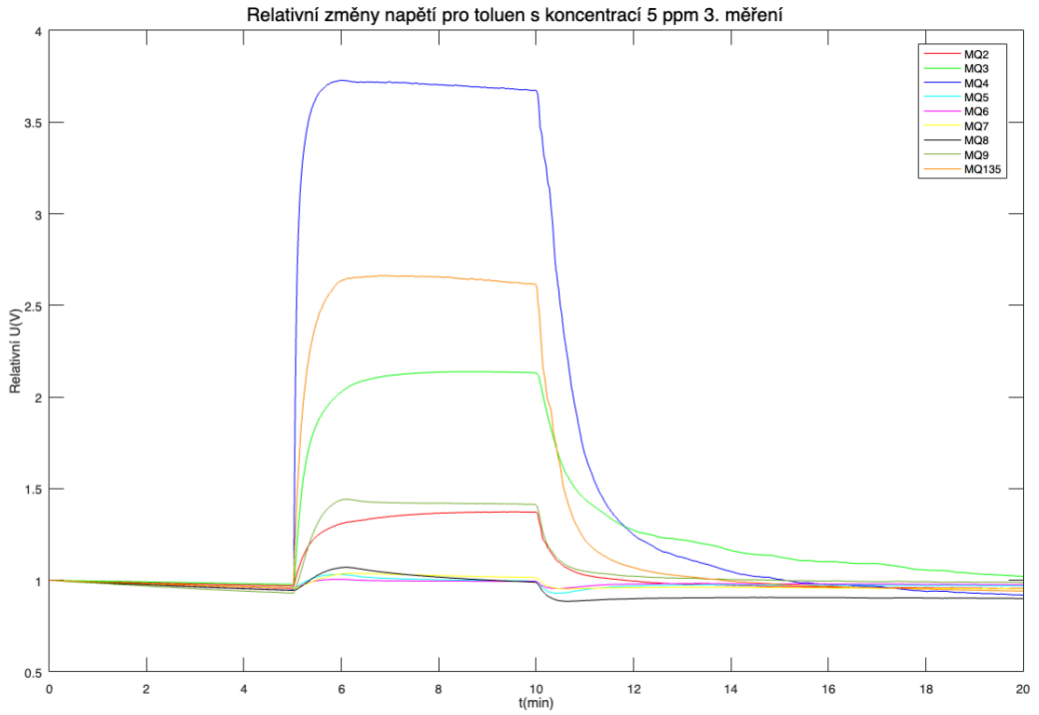


Odezva senzorů pro toluen s koncentrací 5 ppm 3. měření



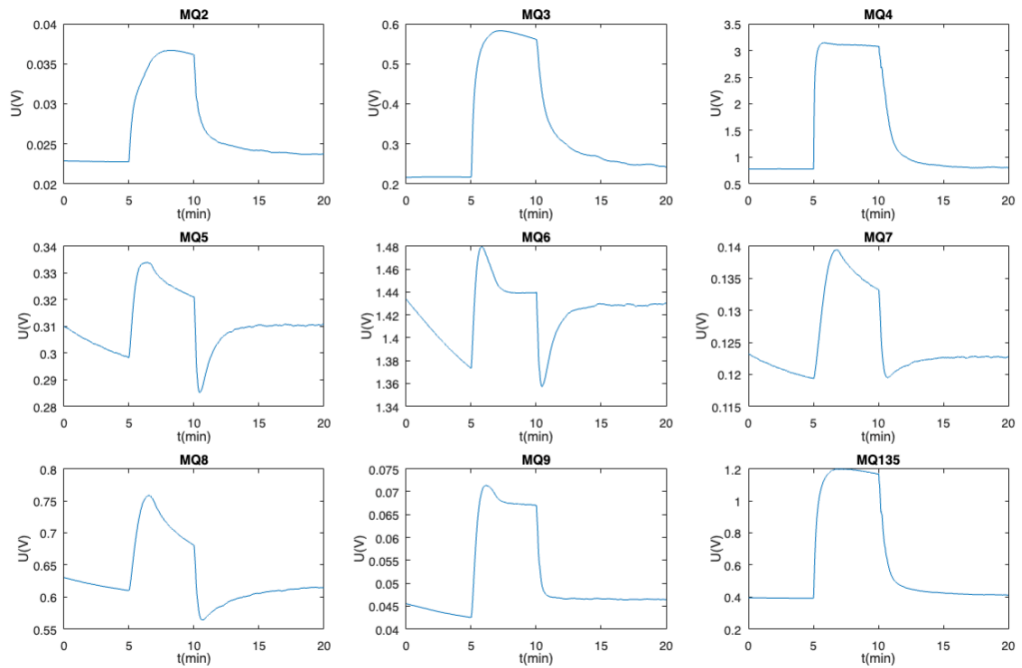
- Relativní změny časových odezev chemorezistivních senzorů při měření toluenu s koncentrací 5 ppm.



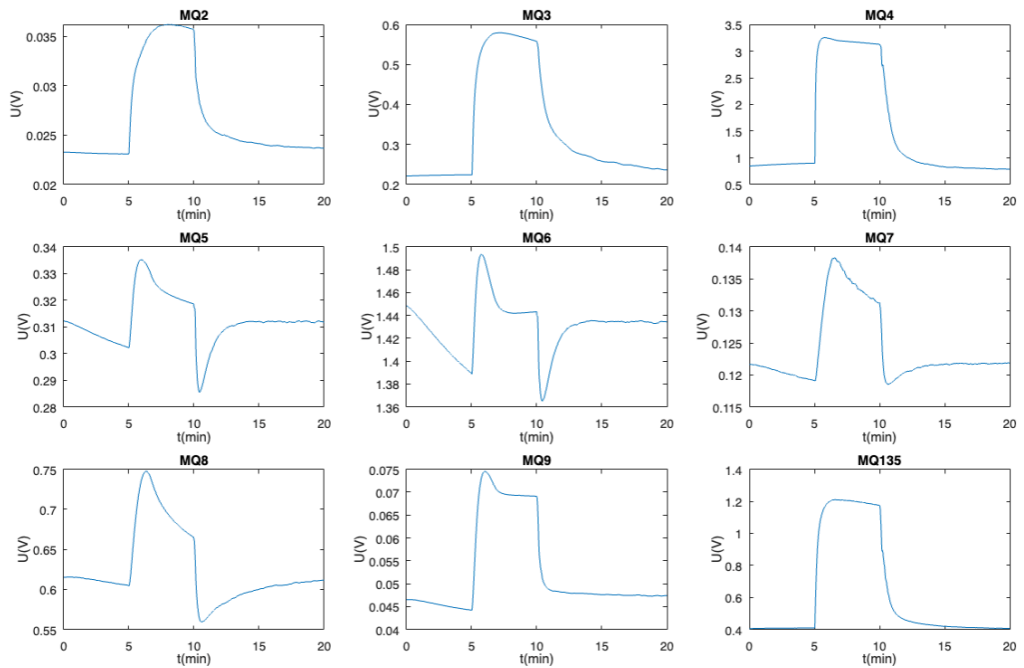


- Průběhy odezvy chemorezistivních senzorů při měření toluenu s koncentrací 10 ppm.

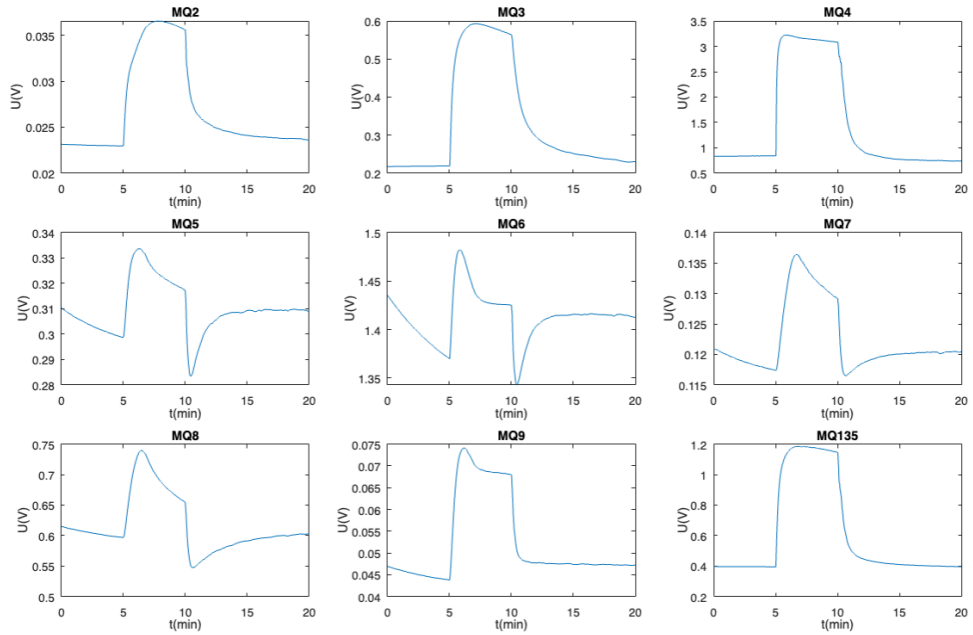
Odezva senzorů pro toluen s koncentrací 10 ppm 1. měření



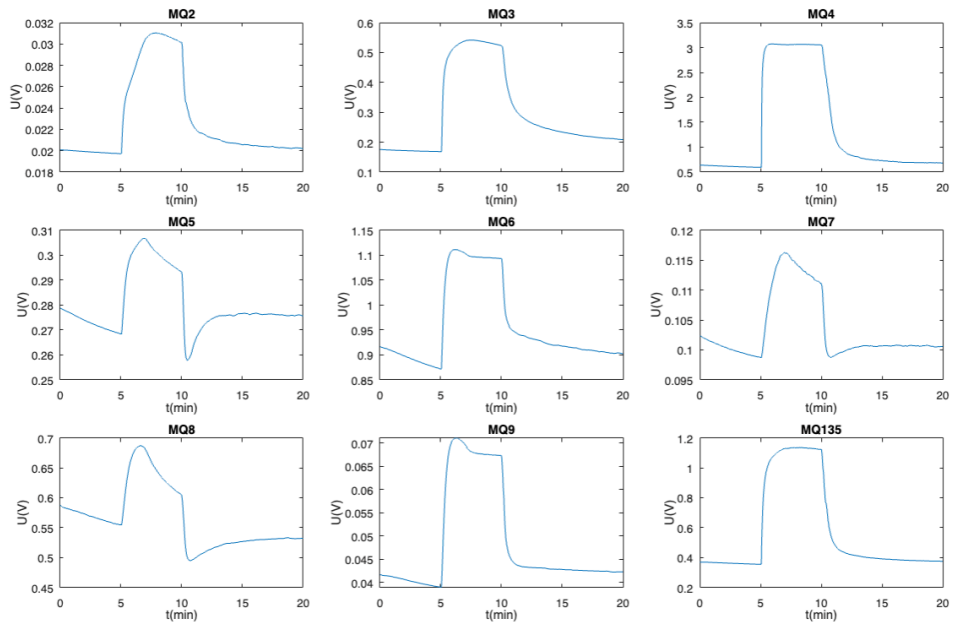
Odezva senzorů pro toluen s koncentrací 10 ppm 2. měření



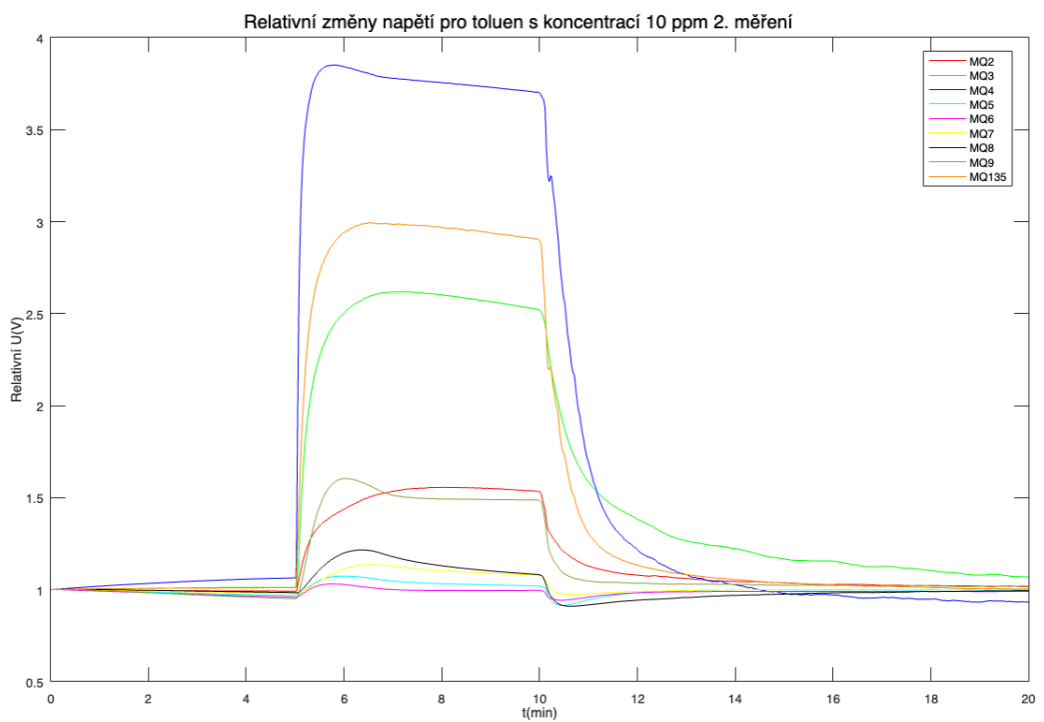
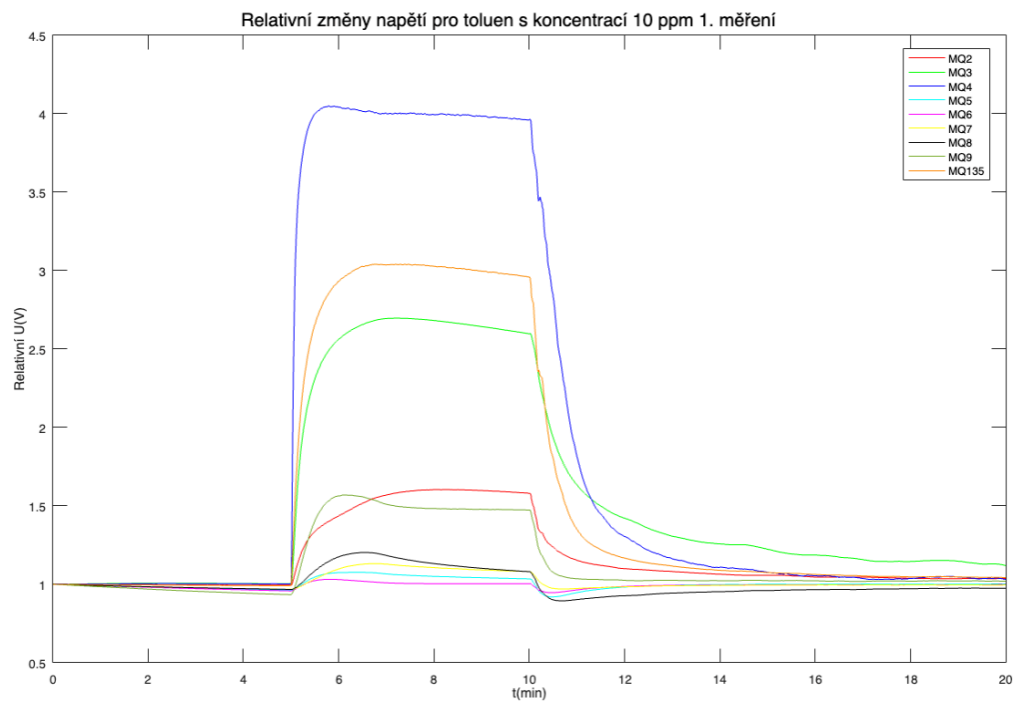
Odezva senzorů pro toluen s koncentrací 10 ppm 3. měření

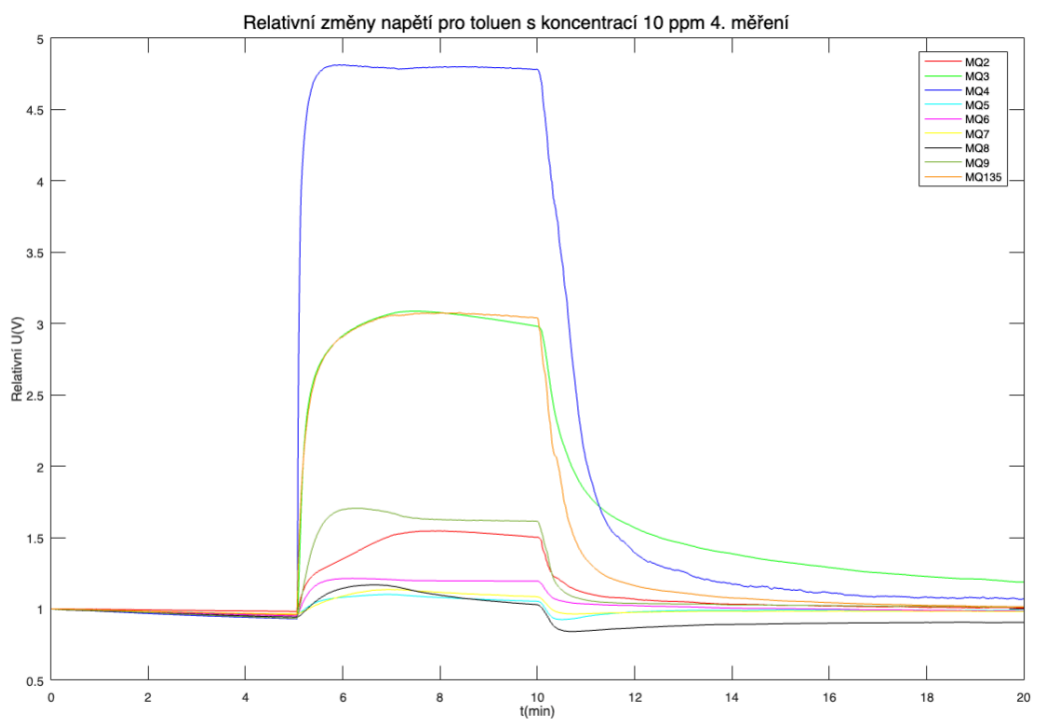
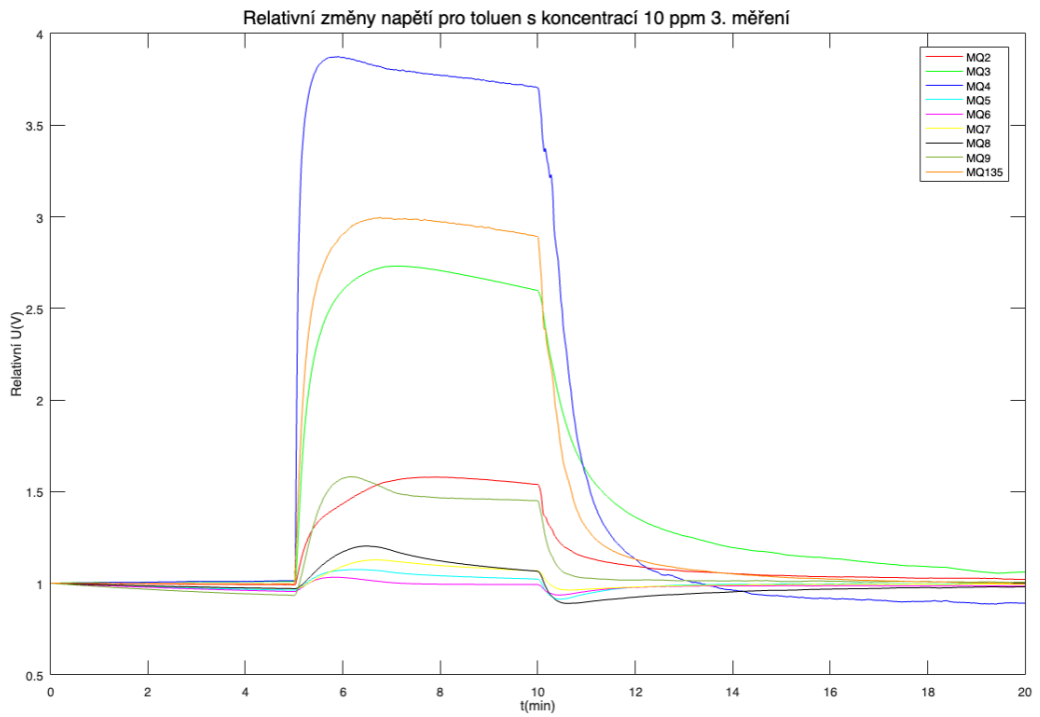


Odezva senzorů pro toluen s koncentrací 10 ppm 4. měření

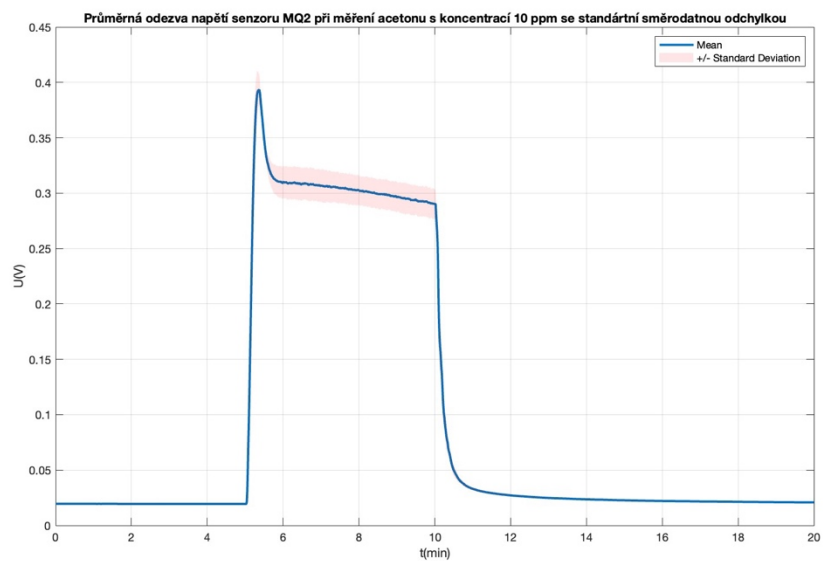
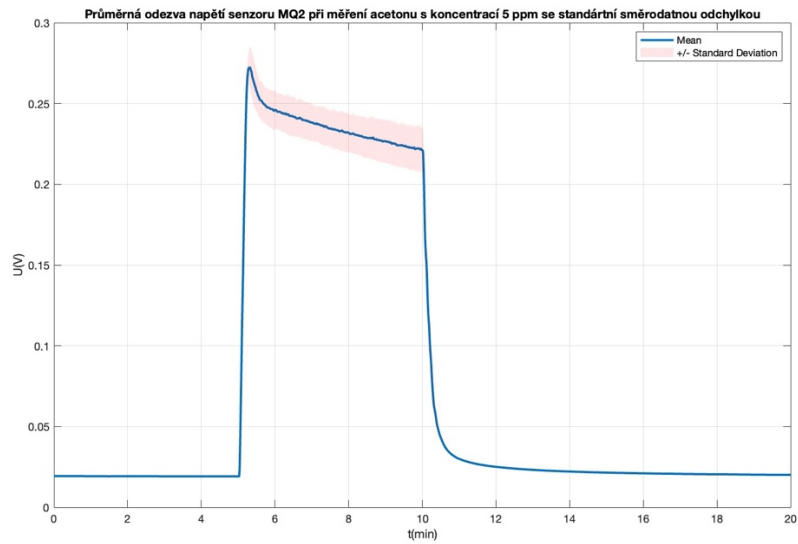
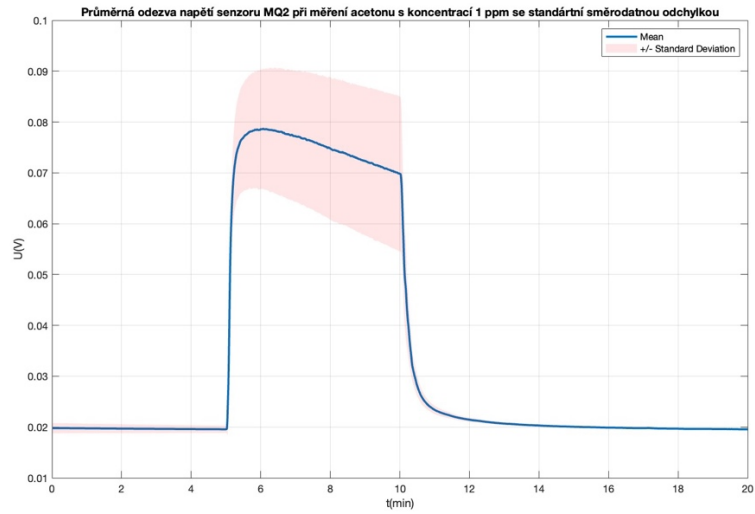


- Relativní změny časových odezev chemorezistivních senzorů při měření toluenu s koncentrací 10 ppm.

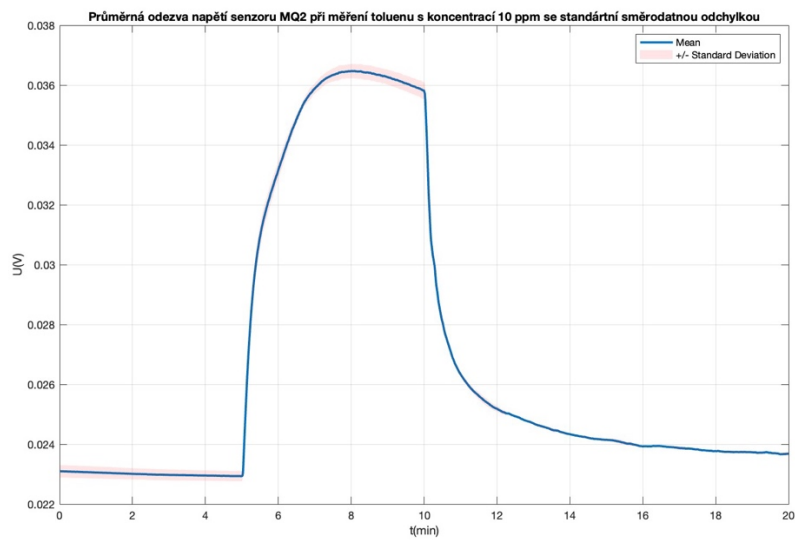
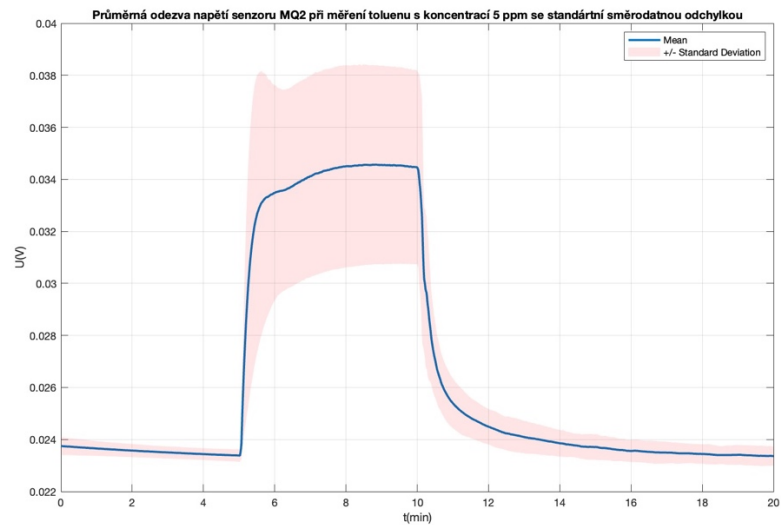
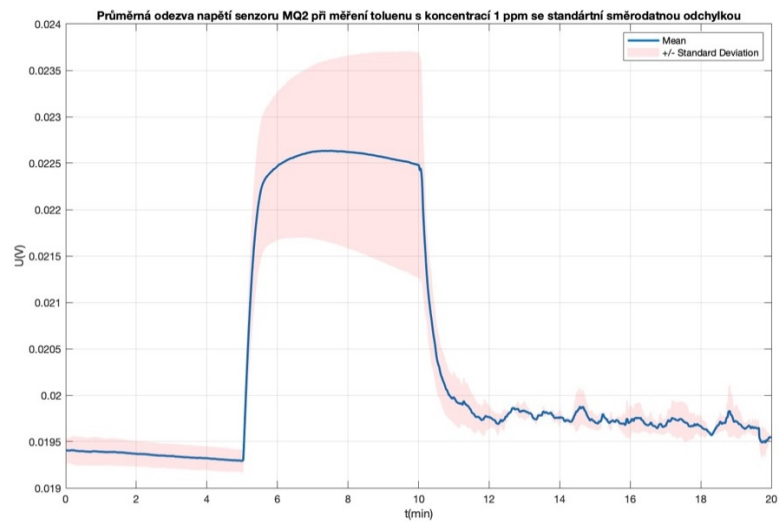




- Průměrné časové odezvy se standardní směrodatnou odchylkou chemorezistivního senzoru MQ2 při měření acetonu s koncentrací 1, 5 a 10 ppm.



- Průměrné časové odezvy se standardní směrodatnou odchylkou chemorezistivního senzoru MQ2 při měření toluenu s koncentrací 1, 5 a 10 ppm.



Příloha C - Výsledky selekce parametrů a vizualizace

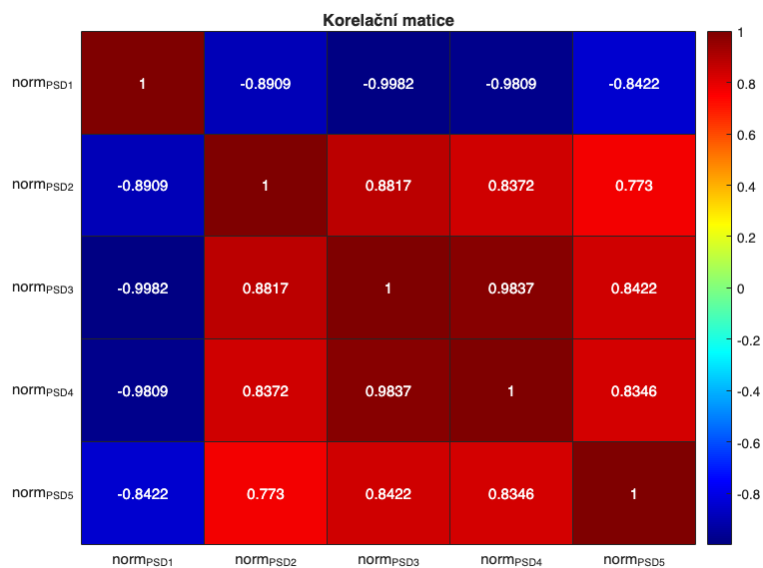
1. Menší soubor dat (měření VOC za víc podobných okolních podmínek)

- Ověření předpokladu normality dat pro jednosměrnou ANOVA pomocí Kolmogorov –Smirnov testu normality pro všechny VOC (test normality napříč skupinami) pro parametry extrahované z časové oblasti:

Tabulka C.1: Výsledky Kolmogoro–vSmirnov testu normality napříč skupinami.

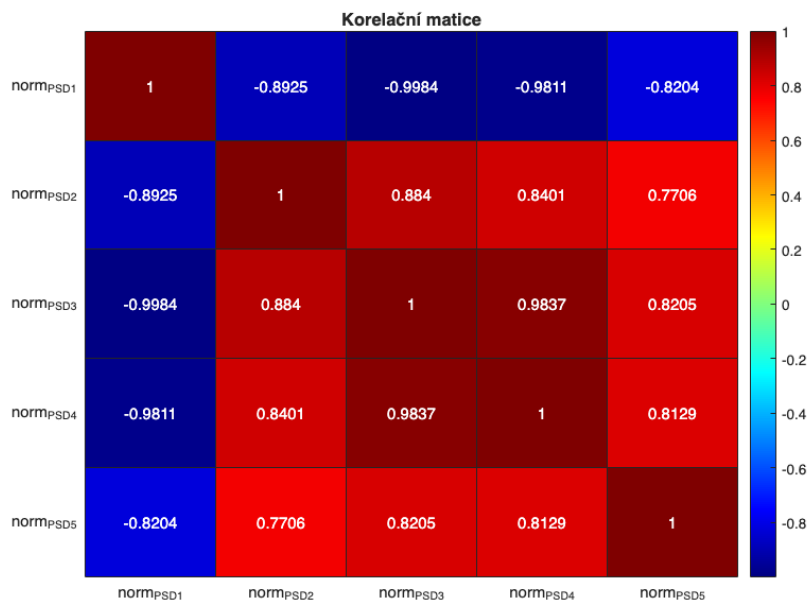
VOC	Parametr	p-hodnota	Platná hypotéza
Aceton	Citlivost	0.074	H_0
Aceton	Nejrychlejší doba náběhu napětí	0.005	H_A
Aceton	Doba diference	0.001	H_A
Aceton	Energie signálu	< 0.001	H_A
Ethanol	Citlivost	< 0.001	H_A
Ethanol	Nejrychlejší doba náběhu napětí	< 0.001	H_A
Ethanol	Doba diference	< 0.001	H_A
Ethanol	Energie signálu	< 0.001	H_A
Toluen	Citlivost	< 0.001	H_A
Toluen	Nejrychlejší doba náběhu napětí	< 0.001	H_A
Toluen	Doba diference	< 0.001	H_A
Toluen	Energie signálu	< 0.001	H_A

- Výsledky testu korelace s využitím Spearmanova korelačního koeficientu pro parametry extrahované z frekvenční oblasti:



2. Větší soubor dat (měření VOC za víc variabilních okolních podmínek)

- Výsledky testu korelace s využitím Spearmanova korelačního koeficientu pro parametry extrahované z frekvenční oblasti:

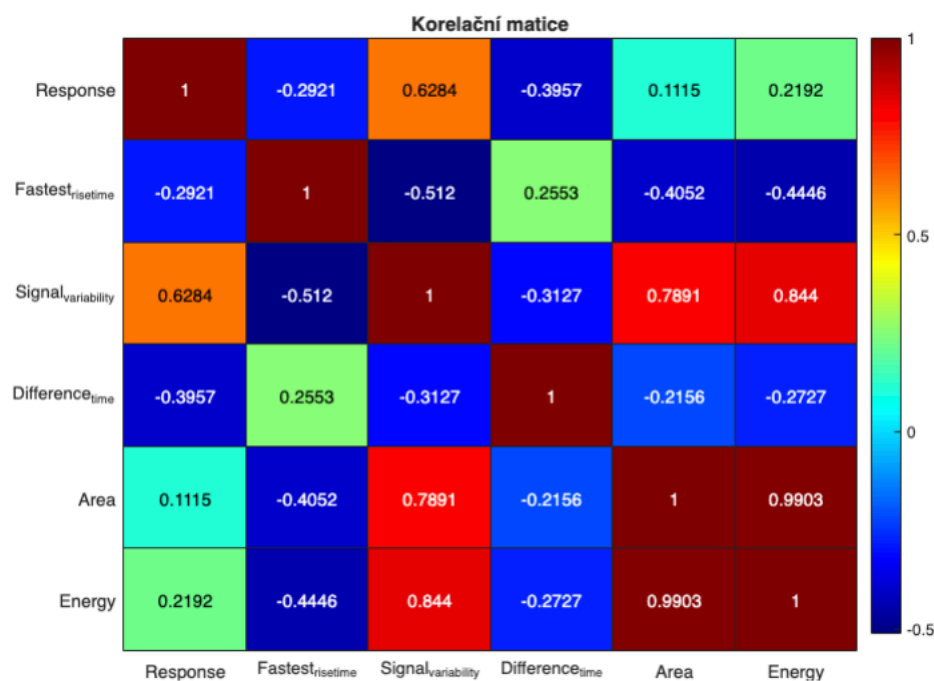


- Výsledky Kolmogorov–Smirnov testu normality pro parametry extrahované z časové oblasti (pro větší soubor dat):

Tabulka C.2: Výsledky Kolmogorov–Smirnov testu normality.

Citlivost	< 0.001	H_A
Nejrychlejší doba náběhu napětí	< 0.001	H_A
Variabilita signálu	< 0.001	H_A
Doba difference	< 0.001	H_A
Plocha pod křivkou	< 0.001	H_A
Energie signálu	< 0.001	H_A

- Výsledky testu korelace s použitím Spearmanůva korelačního koeficientu pro parametry extrahované z časové oblasti (pro větší soubor dat):



Tabulka C.3: Silně korelované páry parametrů.

Pár parametrů	Spearmanův korelační koeficient r_s
Citlivost a variabilita signálu	0.63
Nejrychlejší doba náběhu napětí a variabilita signálu	-0.51
Variabilita signálu a plocha pod křivkou	0.79
Variabilita signálu a energie	0.84
Plocha pod křivkou a energie	0.99

- Ověření předpokladu normality dat pro jednosměrnou ANOVA pomocí Kolmogorov Smirnov testu normality pro všechny VOC (test normality napříč skupinami):

Tabulka C.4: Výsledky Kolmogorov–Smirnov testu normalitu napříč skupinami.

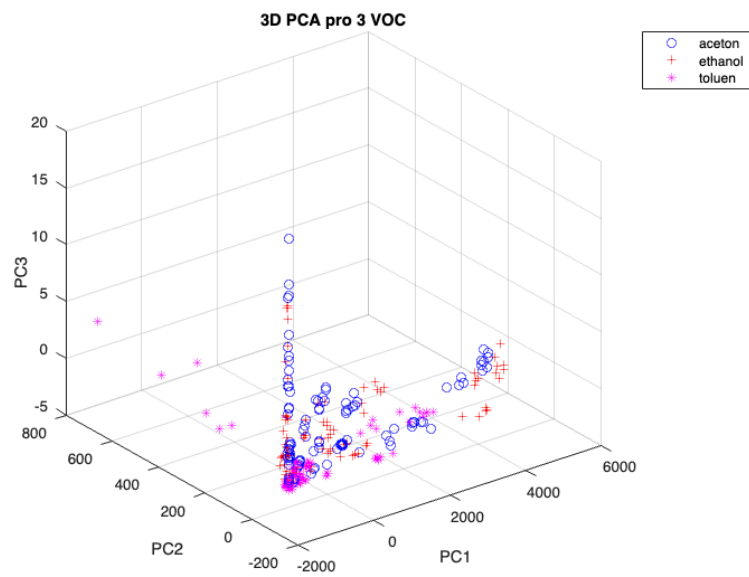
VOC	Parametr	p-hodnota	Platná hypotéza
Aceton	Citlivost	0.017	H_0
Aceton	Nejrychlejší doba náběhu napětí	< 0.001	H_A
Aceton	Doba diference	< 0.001	H_A
Aceton	Energie signálu	< 0.001	H_A
Ethanol	Citlivost	0.004	H_A
Ethanol	Nejrychlejší doba náběhu napětí	0.002	H_A
Ethanol	Doba diference	< 0.001	H_A
Ethanol	Energie signálu	< 0.001	H_A
Toluen	Citlivost	< 0.001	H_A
Toluen	Nejrychlejší doba náběhu napětí	< 0.001	H_A
Toluen	Doba diference	< 0.001	H_A
Toluen	Energie signálu	< 0.001	H_A

- Výsledky selekce parametrů pomocí jednosměrné ANOVA (pro větší soubor dat):

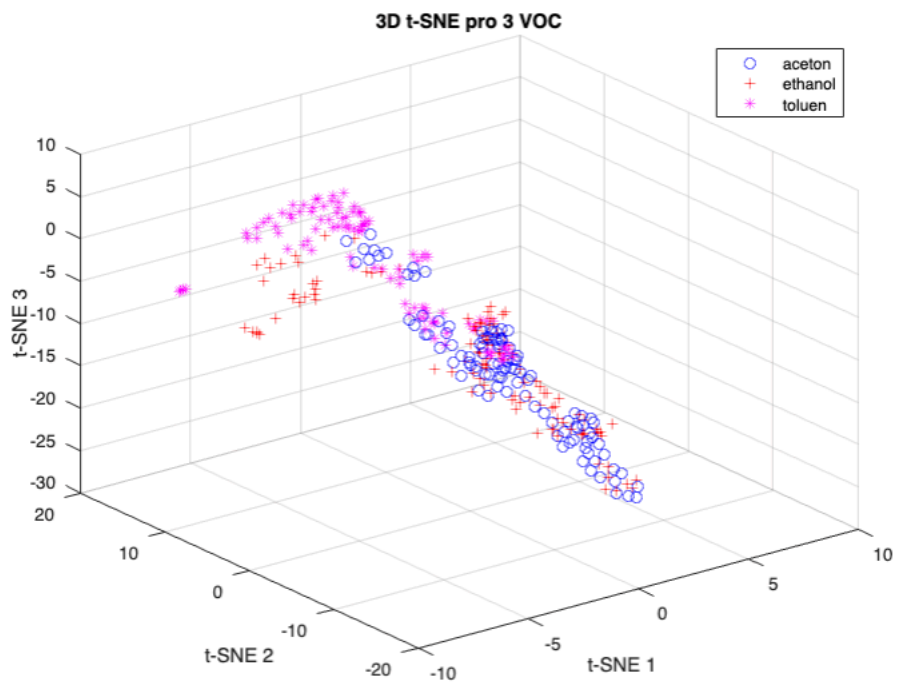
Tabulka C.5: Vybrané parametry pomocí jednosměrné ANOVA.

Parametr	p-hodnota	$F(df_{columns}, df_{error})$
Citlivost	< 0.001	47(2, 240)
Nejrychlejší doba náběhu signálu	< 0.001	8(2, 240)
Energie signálu	0.001	7(2, 240)

- Výsledky vizualizace VOC pomocí PCA na větším souboru dat:

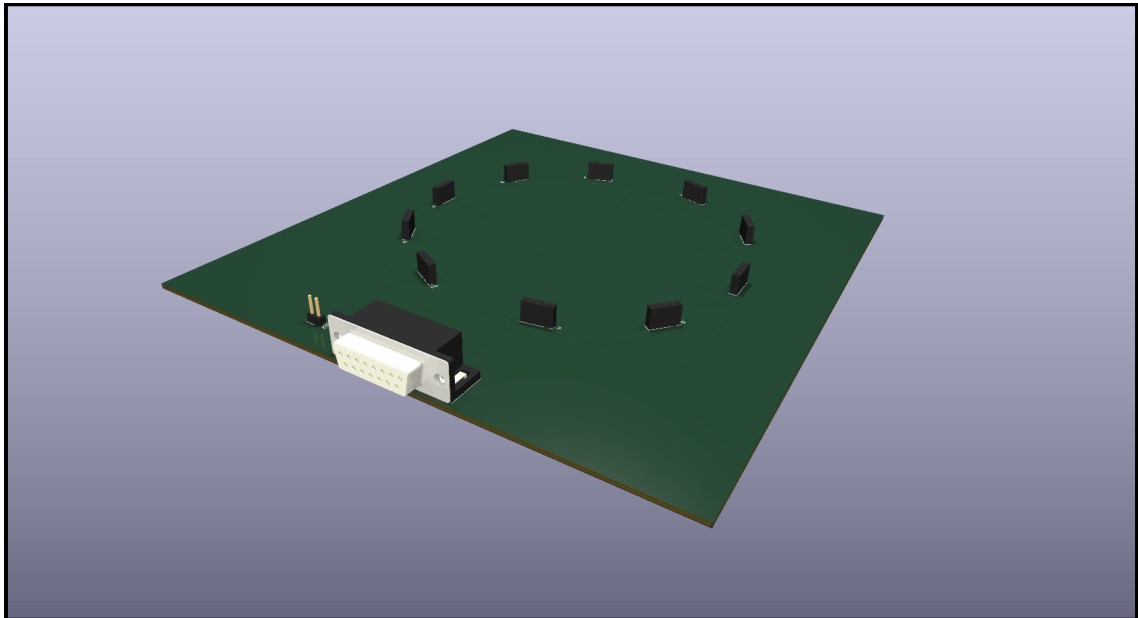


- Výsledky vizualizace VOC pomocí t-SNE na větším souboru dat, hodnota hyperparametru „perplexita“ je 35:

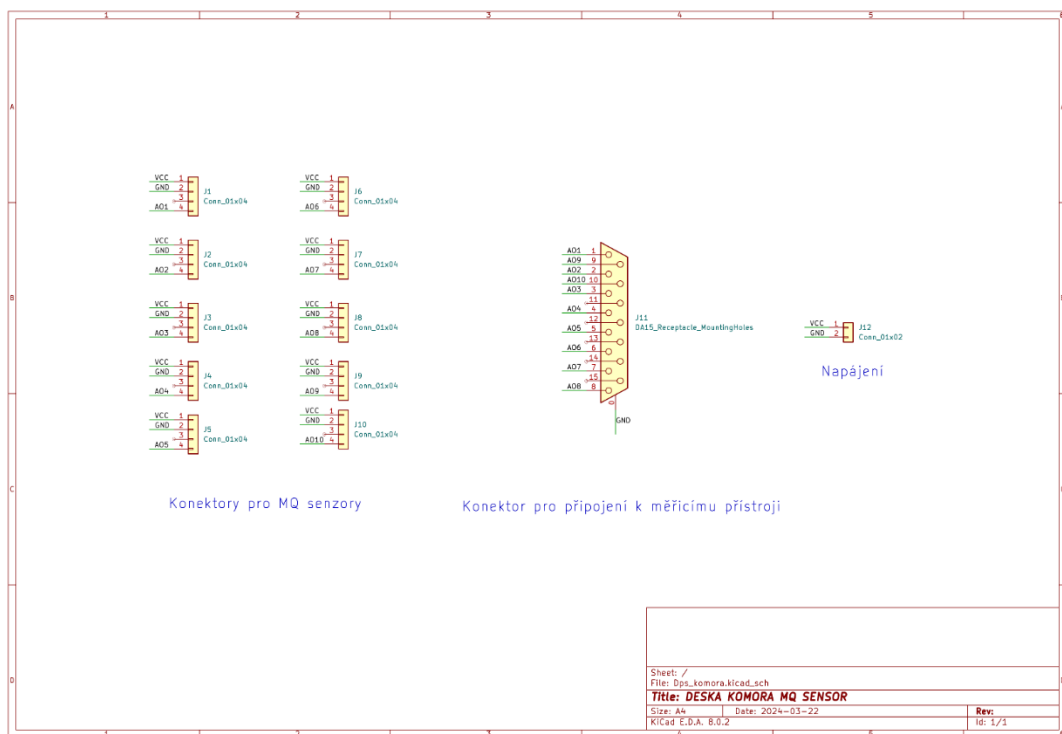


Příloha D - Testovací komora, varianta 2

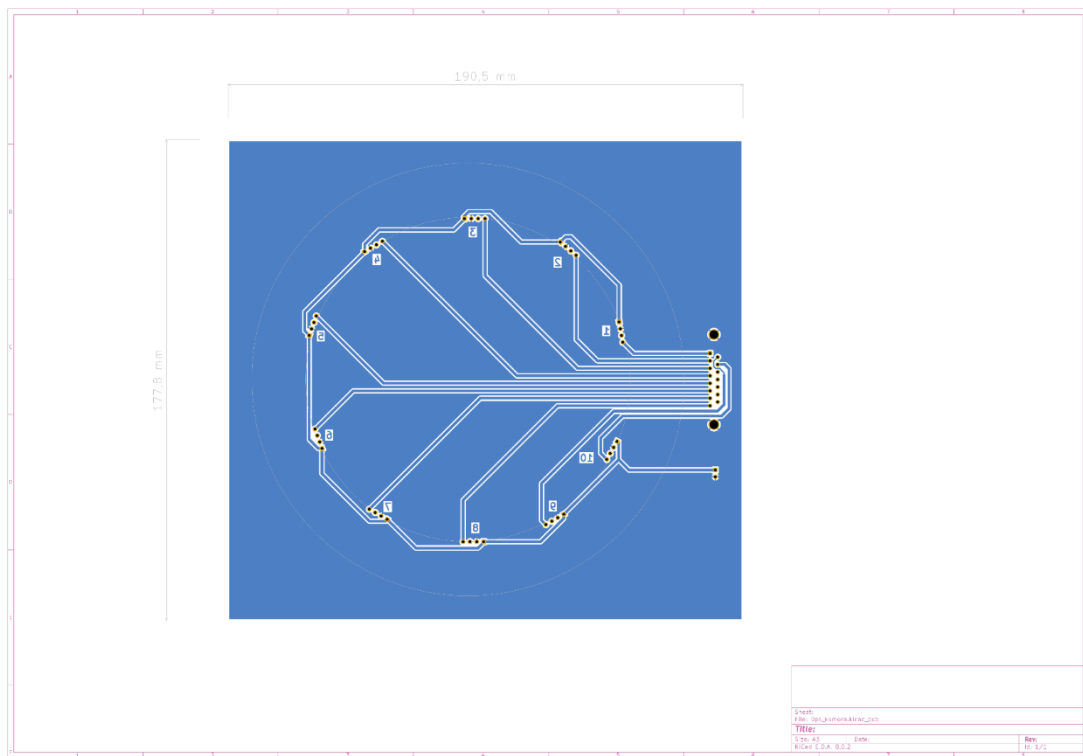
- 3D pohled na desku s plošnými spoji (DPS)



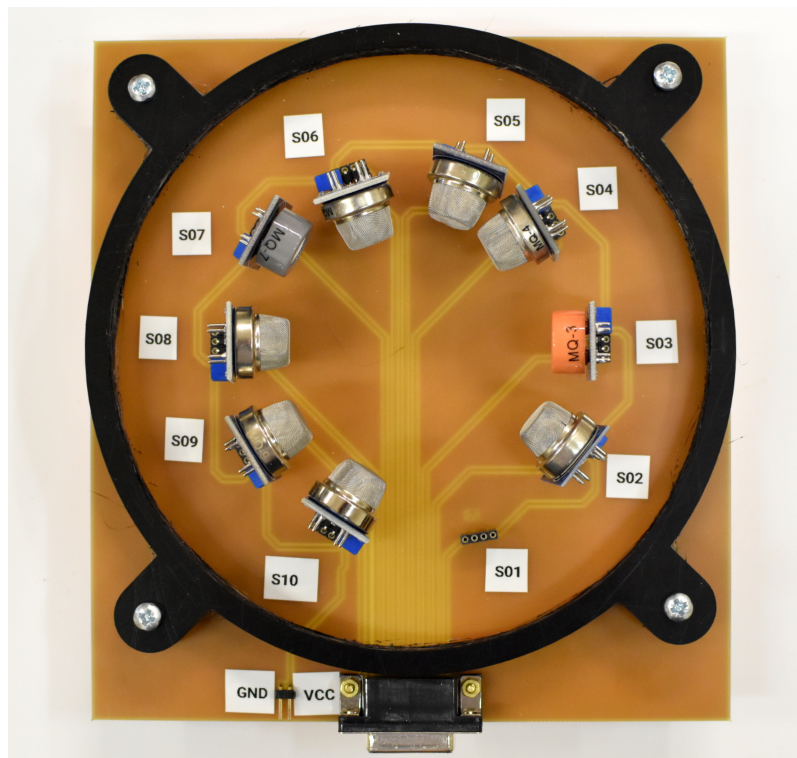
- Schéma pro DPS



- Motiv mědi na spodní straně DPS



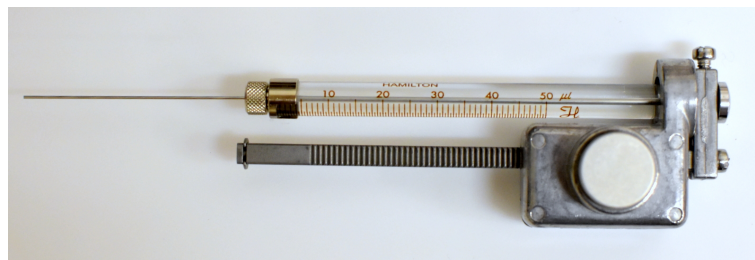
- Fotografie hotové DPS, včetně MQ senzorů a vymešovaciho rámu



- Fotografie celé komory, v horní části je držák s hodinovým sklíčkem a ventilátorem pro dávkování VOC, vnější průměr komory 16 cm



- Fotografie mikrostříkačky Hamilton s dávkovacím nastavcem pro přesné dávkování VOC, rozsah dávkování 1 μ l až 50 μ l



Seznam a popis příložených souborů

- Popis:
 - README.txt
- Naměřená data
- Zdrojový kód:
 - Výpočet koncentrací VOC:
 - kontrace.mlx
 - Grafy průběhů odezev:
 - prubehy_odezev.mlx
 - mean_std_prubehy.mlx
 - Extrakce parametrů:
 - parametry_cas.m
 - parametry_frekvence.mlx
 - Selektce a vizualizace parametrů:
 - selekce.mlx
 - vizualizace.mlx
 - features_boxplot.mlx
 - Klasifikace:
 - nested_CV.m