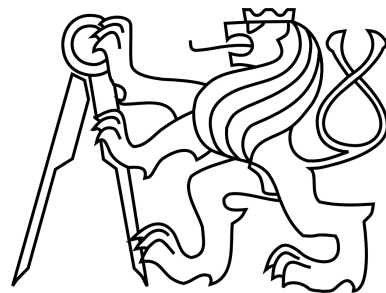


ČESKÉ VYSOKÉ UČENÍ TECHNICKÉ V PRAZE
FAKULTA ELEKTROTECHNICKÁ
KATEDRA KYBERNETIKY



DIPLOMOVÁ PRÁCE

Detekce poruch senzorů

Praha, 2014

Autor: Petr Marek

Autor: Petr Marek
Studijní program kybernetika a robotika (magisterský),
Obor robotika

Vedoucí: Doc. Ing. Radislav Šmíd, Ph.D.
Katedra měření, Fakulta elektrotechnická,
České vysoké učení technické v Praze

Oponent: Ing. Eva Jerhotová, Ph.D.
Honeywell s.r.o.,
V Parku 2326/18, Praha 4

Rok: 2014

ZADÁNÍ DIPLOMOVÉ PRÁCE

Student: Bc. Petr Marek
Studijní program: Kybernetika a robotika (magisterský)
Obor: Robotika
Název tématu: Detekce poruch senzorů

Pokyny pro vypracování:

1. Na základě bibliografické rešerše navrhnete nové metody pro detekci poruch senzorů v aplikacích v budovách, zejména vzduchotechnice (například senzor teploty venkovního vzduchu OAT). Zaměřte se na metody užívající externí informaci (např. z blízkého objektu stejného typu, veřejného datového zdroje nebo redundantních senzorů) a ohodnocení kvality informace. Detekujte zejména trvalé nesprávné chování senzoru a změnu chování senzoru vůči počáteční situaci.
2. Vyhodnoťte a porovnejte úspěšnost metod.
3. Metody implementujte v Matlabu ve formě toolboxu.

Seznam odborné literatury:

- [1] Chatfield, Ch.: The Analysis of Time Series: An Introduction, CRC Pr I Llc, 2003.
[2] Kreibich, O.; Neužil, J.; Šmíd, R.: Quality-based Multiple Sensor Fusion in an Industrial Wireless Sensor Network for MCM, to appear in IEEE Transactions on Industrial Electronics, 2013.

Vedoucí diplomové práce: doc. Ing. Radislav Šmíd, Ph.D.

Platnost zadání: do konce letního semestru 2014/2015

L.S.

doc. Dr. Ing. Jan Kybic
vedoucí katedry

prof. Ing. Pavel Ripka, CSc.
děkan

V Praze dne 10. 12. 2013

Prohlášení

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

V Praze dne _____

Petr Marek

Poděkování

Děkuji především vedoucímu diplomové práce Doc. Ing. Radislavu Šmídovi, Ph.D. za vedení této diplomové práce a také kolegovi Ing. Vladimíru Horynovi, za ochotu a cenné rady.

Dále bych rád poděkoval své rodině a blízkým přátelům za jejich podporu během celé doby mého studia.

Abstrakt

Cílem této diplomové práce je na základě bibliografické rešerše navrhnout nové metody pro detekci poruch senzorů v aplikacích v budovách, zejména ve vzduchotechnice a spotřebě elektrické energie. Základem zde představených metod je využití externí informace z blízkého objektu stejného typu, veřejného datového zdroje nebo redundantních senzorů. Hlavním úkolem je detekovat zejména trvale nesprávné chování senzoru a změnu chování senzoru vůči počáteční situaci. V práci jsou porovnány metody pracující na základě ARX (AutoRegressive model with eXternal input), GMM (Gaussian Mixture Model) a NARXnet (Nonlinear AutoRegressive neural network with eXternal input). Výsledkem diplomové práce jsou funkční metody pro detekci poruch senzorů implementované v programu Matlab, které jsou v závěru vyhodnoceny a porovnány mezi sebou.

Klíčová slova

senzory, diagnostika, detekce poruch

Abstract

The aim of this thesis is to develop new sensor fault detection methods. These methods are to be applied within buildings, especially for air conditioning monitoring and fault detection using evaluation of energy consumption. All the methods use data from a nearby object of the same type, a public data source or redundant sensors. The main objective is to detect persistent faults and changes in comparison to initial situation. For the purpose of this thesis three methods are compared: ARX (AutoRegressive model with eXternal input), GMM (Gaussian Mixture Model) and NARXnet (Nonlinear AutoRegressive neural network with eXternal input). As a result the thesis presents functional fault detection methods implemented in Matlab which are evaluated in the conclusion.

Key words

sensors, diagnostics, fault detection

Obsah

Seznam obrázků	ix
Seznam tabulek	xvii
1 Úvod	1
1.1 Motivace	1
1.1.1 Plýtvání energií v budovách	2
1.1.2 Dopad poruchy senzoru OAT na funkci HVAC	3
1.2 Data ke zpracování a zdroje externích informací	5
1.2.1 Data o spotřebě elektrické energie	5
1.2.2 Naměřené teploty ze čtyř meteorologických stanic	9
2 Cíl diplomové práce	13
3 Současný stav	15
3.1 Bibliografická rešerše	15
3.2 Detekce poruch	16
3.3 Poruchy	18
4 Teoretická část	21
4.1 Základní princip použitých metod detekce poruch	21
4.2 Analýza časové řady a její predikce	23
4.2.1 Cíle analýzy časových řad	23
4.2.2 Předpovídání	24
4.3 ARX (AutoRegressive model with eXternal input)	25
4.3.1 SISO ARX (Single Input–Single Output)	26
4.3.2 Predikce modelu	27
4.3.3 Učení modelu	27

4.3.4	MISO ARX (Multiple Input–Single Output)	28
4.4	GMM (Gaussian Mixture Model)	29
4.4.1	Gaussian mixture model	30
4.4.2	Výběr počtu komponent	31
4.4.3	Predikce modelu	31
4.4.4	Interval spolehlivosti predikce	32
4.5	NARXnet	33
4.5.1	Úvod do neuronových sítí	33
4.5.2	NARXnet (Nonlinear AutoRegressive neural network with eXternal input)	36
5	Implementace	39
5.1	Hlavní struktura systému	39
5.2	Implementace funkcí ARX	43
5.2.1	Učení modelu	43
5.2.2	Stanovení parametrů pro nastavení mezí	45
5.2.3	Predikce	45
5.3	Implementace funkcí GMM	47
5.3.1	Učení modelu	47
5.3.2	Stanovení parametrů pro nastavení mezí	50
5.3.3	Predikce	50
5.4	Implementace funkcí NARXnet	51
5.4.1	Učení modelu	51
5.4.2	Stanovení parametrů pro nastavení mezí	51
5.4.3	Predikce	52
5.4.4	Stav	53
6	Experimenty a zhodnocení výsledků	55
6.1	Problematika učení modelů, hledání parametrů, testování	55
6.2	Výsledky	58
6.2.1	ARX (AutoRegressive model with eXternal input) uspořádání číslo 1	58
6.2.2	ARX (AutoRegressive model with eXternal input) uspořádání číslo 2	70
6.2.3	GMM (Gaussian Mixture Model)	81
6.2.4	NARXnet (Nonlinear AutoRegressive neural network with eXternal input) uspořádání číslo 1	93

6.2.5	NARXnet (Nonlinear AutoRegressive neural network with eXternal input) uspořádání číslo 2	104
6.3	Zhodnocení	115
6.3.1	Shrnutí	117
7	Závěr	119
7.1	Návrh na další práci	119
A	Grafické přílohy	I
B	Tabulky	XV
C	Obsah přiloženého CD	XXI

Seznam obrázků

1.1	Schéma vzduchotechnické jednotky sloužící k úpravě a cirkulaci vzduchu jako součást HVAC.	4
1.2	Spotřeba elektrické energie dvou domácností za 1 rok (pro ilustraci průběhů).	6
1.3	Spotřeba elektrické energie dvou domácností za 1 týden.	7
1.4	Průměrná spotřeba elektrické energie dvou domácností za den.	7
1.5	Histogramy spotřeby energie dvou domácností.	8
1.6	Spotřeba elektrické energie dvou domácností za 1 rok (pro ilustraci průběhů).	8
1.7	Histogramy spotřeby energie dvou domácností.	9
1.8	Průběh teplot ze čtyř meteorologických stanic za 1 rok (pro ilustraci průběhů).	10
1.9	Průběh teplot ze čtyř meteorologických stanic za 1 měsíc.	10
1.10	Histogramy teplot ze čtyř meteorologických stanic.	11
2.1	Principiální schéma pro detekci poruch senzorů.	14
3.1	Schéma modelu detekce poruch, [18].	16
3.2	Schéma rozhodování o poruše.	17
3.3	Časově závislé aditivní chyby: a) náhlá porucha, b) pozvolná porucha (drift), c) občasná porucha, [18].	18
3.4	Ukázka vlivu různých poruch na statický senzor (přečteno $Y(t)$, naměřeno $Y_0(t)$): a) posunutí počátku, b) změna zesílení (gain), c) změna odezvy, d) změna hystereze, [18].	19
4.1	Zjednodušené schéma učení matematického modelu.	22
4.2	Zjednodušené schéma detekce poruch pomocí GMM.	22
4.3	Zjednodušené schéma detekce poruch pomocí ARX (NAXRnet).	22
4.4	Blokové schéma SISO ARX modelu.	26
4.5	Blokové schéma MISO ARX modelu.	29
4.6	Struktura dvouvrstvé dopředné neuronové sítě.	33

4.7	Model umělého neuronu.	34
4.8	Závislost velikosti chyb na délce trénovací množiny vlevo a závislost velikosti chyb na počtu neuronů ve skryté vrstvě vpravo, převzato z [15]. . .	35
4.9	Ukázka struktury NARXnet.	36
4.10	Obrázek paralelní a sério-paralelní architektury pro učení, převzato a upraveno z [12].	36
4.11	Výsledný model NARXnet, převzato a upraveno z [12].	37
5.1	Blokové schéma vyjadřující strukturu toolboxu v Matlabu.	39
5.2	Principiální schéma vyhodnocení poruch.	42
5.3	Schéma učení ARX modelu.	43
5.4	Ukázka hledání optimálního nastavení parametrů ARX modelu pro teplotní data.	44
5.5	Ukázka hledání optimálního nastavení parametrů ARX modelu pro data spotřeby energie „Coastal Family“.	44
5.6	Schéma predikce a stanovení mezí ARX.	46
5.7	Schéma učení GMM modelu.	47
5.8	Ukázka hledání optimálního počtu komponent GMM modelu pro teplotní data.	48
5.9	Ukázka rozdělení teplotních dat GMM modelu.	48
5.10	Ukázka hledání optimálního počtu komponent GMM modelu pro data spotřeby energie „Coastal Family“.	49
5.11	Ukázka rozdělení dat spotřeby energie „Coastal Family“ GMM modelu. .	49
5.12	Schéma predikce GMM.	50
5.13	Schéma učení NARXnet modelu.	51
5.14	Schéma predikce NARXnet.	52
6.1	Ukázka teplotních dat - <i>Budihostice, Červený Újezd,...</i>	56
6.2	Ukázka dat určených na učení modelů, hledání parametrů a testování - spotřeba energie <i>Coastal Multi a Single Family</i>	57
6.3	Ukázka dat určených na učení modelů, hledání parametrů a testování - spotřeba energie <i>Desert Multi a Single Family</i>	57
6.5	Detail výsledků ARX uspořádání číslo 1 - teploty <i>Budihostice, Červený Újezd, Chloumek, ČZU (peak)</i>	62
6.6	Detail výsledků ARX uspořádání číslo 1 - teploty <i>Budihostice, Červený Újezd, Chloumek, ČZU (skok)</i>	62

6.7	Detail výsledků ARX uspořádání číslo 1 - teploty <i>Budihostice, Červený Újezd, Chloumek, ČZU</i> (drift).	63
6.8	Detail výsledků ARX uspořádání číslo 1 - teploty <i>Budihostice, Červený Újezd, Chloumek, ČZU</i> (gain).	63
6.10	Detail výsledků ARX uspořádání číslo 1 - spotřeba energie <i>Coastal Multi a Single Family</i> (peak).	65
6.11	Detail výsledků ARX uspořádání číslo 1 - spotřeba energie <i>Coastal Multi a Single Family</i> (skok).	65
6.12	Detail výsledků ARX uspořádání číslo 1 - spotřeba energie <i>Coastal Multi a Single Family</i> (drift).	66
6.13	Detail výsledků ARX uspořádání číslo 1 - spotřeba energie <i>Coastal Multi a Single Family</i> (gain).	66
6.15	Detail výsledků ARX uspořádání číslo 1 - spotřeba energie <i>Desert Multi a Single Family</i> (peak).	68
6.16	Detail výsledků ARX uspořádání číslo 1 - spotřeba energie <i>Desert Multi a Single Family</i> (skok).	68
6.17	Detail výsledků ARX uspořádání číslo 1 - spotřeba energie <i>Desert Multi a Single Family</i> (drift).	69
6.18	Detail výsledků ARX uspořádání číslo 1 - spotřeba energie <i>Desert Multi a Single Family</i> (gain).	69
6.20	Detail výsledků ARX uspořádání číslo 2 - teploty <i>Budihostice, Červený Újezd, Chloumek, ČZU</i> (peak).	73
6.21	Detail výsledků ARX uspořádání číslo 2 - teploty <i>Budihostice, Červený Újezd, Chloumek, ČZU</i> (skok).	73
6.22	Detail výsledků ARX uspořádání číslo 2 - teploty <i>Budihostice, Červený Újezd, Chloumek, ČZU</i> (drift).	74
6.23	Detail výsledků ARX uspořádání číslo 2 - teploty <i>Budihostice, Červený Újezd, Chloumek, ČZU</i> (gain).	74
6.25	Detail výsledků ARX uspořádání číslo 2 - spotřeba energie <i>Coastal Multi a Single Family</i> (peak).	76
6.26	Detail výsledků ARX uspořádání číslo 2 - spotřeba energie <i>Coastal Multi a Single Family</i> (skok).	76
6.27	Detail výsledků ARX uspořádání číslo 2 - spotřeba energie <i>Coastal Multi a Single Family</i> (drift).	77

6.28	Detail výsledků ARX uspořádání číslo 2 - spotřeba energie <i>Coastal Multi a Single Family</i> (gain).	77
6.30	Detail výsledků ARX uspořádání číslo 2 - spotřeba energie <i>Desert Multi a Single Family</i> (peak).	79
6.31	Detail výsledků ARX uspořádání číslo 2 - spotřeba energie <i>Desert Multi a Single Family</i> (skok).	79
6.32	Detail výsledků ARX uspořádání číslo 2 - spotřeba energie <i>Desert Multi a Single Family</i> (drift).	80
6.33	Detail výsledků ARX uspořádání číslo 2 - spotřeba energie <i>Desert Multi a Single Family</i> (gain).	80
6.35	Detail výsledků GMM - teploty <i>Budihostice, Červený Újezd, Chloumek, ČZU</i> (peak).	85
6.36	Detail výsledků GMM - teploty <i>Budihostice, Červený Újezd, Chloumek, ČZU</i> (skok).	85
6.37	Detail výsledků GMM - teploty <i>Budihostice, Červený Újezd, Chloumek, ČZU</i> (drift).	86
6.38	Detail výsledků GMM - teploty <i>Budihostice, Červený Újezd, Chloumek, ČZU</i> (gain).	86
6.40	Detail výsledků GMM - spotřeba energie <i>Coastal Multi a Single Family</i> (peak).	88
6.41	Detail výsledků GMM - spotřeba energie <i>Coastal Multi a Single Family</i> (skok).	88
6.42	Detail výsledků GMM - spotřeba energie <i>Coastal Multi a Single Family</i> (drift).	89
6.43	Detail výsledků GMM - spotřeba energie <i>Coastal Multi a Single Family</i> (gain).	89
6.45	Detail výsledků GMM - spotřeba energie <i>Desert Multi a Single Family</i> (peak).	91
6.46	Detail výsledků GMM - spotřeba energie <i>Desert Multi a Single Family</i> (skok).	91
6.47	Detail výsledků GMM - spotřeba energie <i>Desert Multi a Single Family</i> (drift).	92
6.48	Detail výsledků GMM - spotřeba energie <i>Desert Multi a Single Family</i> (gain).	92

6.50	Detail výsledků NARXnet uspořádání číslo 1 - teploty <i>Budihostice, Červený Újezd, Chloumek, ČZU</i> (peak).	96
6.51	Detail výsledků NARXnet uspořádání číslo 1 - teploty <i>Budihostice, Červený Újezd, Chloumek, ČZU</i> (skok).	96
6.52	Detail výsledků NARXnet uspořádání číslo 1 - teploty <i>Budihostice, Červený Újezd, Chloumek, ČZU</i> (drift).	97
6.53	Detail výsledků NARXnet uspořádání číslo 1 - teploty <i>Budihostice, Červený Újezd, Chloumek, ČZU</i> (gain).	97
6.55	Detail výsledků NARXnet uspořádání číslo 1 - spotřeba energie <i>Coastal Multi a Single Family</i> (peak).	99
6.56	Detail výsledků NARXnet uspořádání číslo 1 - spotřeba energie <i>Coastal Multi a Single Family</i> (skok).	99
6.57	Detail výsledků NARXnet uspořádání číslo 1 - spotřeba energie <i>Coastal Multi a Single Family</i> (drift).	100
6.58	Detail výsledků NARXnet uspořádání číslo 1 - spotřeba energie <i>Coastal Multi a Single Family</i> (gain).	100
6.60	Detail výsledků NARXnet uspořádání číslo 1 - spotřeba energie <i>Desert Multi a Single Family</i> (peak).	102
6.61	Detail výsledků NARXnet uspořádání číslo 1 - spotřeba energie <i>Desert Multi a Single Family</i> (skok).	102
6.62	Detail výsledků NARXNET uspořádání číslo 1 - spotřeba energie <i>Desert Multi a Single Family</i> (drift).	103
6.63	Detail výsledků NARXnet uspořádání číslo 1 - spotřeba energie <i>Desert Multi a Single Family</i> (gain).	103
6.65	Detail výsledků NARXnet uspořádání číslo 2 - teploty <i>Budihostice, Červený Újezd, Chloumek, ČZU</i> (peak).	107
6.66	Detail výsledků NARXnet uspořádání číslo 2 - teploty <i>Budihostice, Červený Újezd, Chloumek, ČZU</i> (skok).	107
6.67	Detail výsledků NARXnet uspořádání číslo 2 - teploty <i>Budihostice, Červený Újezd, Chloumek, ČZU</i> (drift).	108
6.68	Detail výsledků NARXnet uspořádání číslo 2 - teploty <i>Budihostice, Červený Újezd, Chloumek, ČZU</i> (gain).	108
6.70	Detail výsledků NARXnet uspořádání číslo 2 - spotřeba energie <i>Coastal Multi a Single Family</i> (peak).	110

6.71	Detail výsledků NARXnet uspořádání číslo 2 - spotřeba energie <i>Coastal Multi a Single Family</i> (skok).	110
6.72	Detail výsledků NARXnet uspořádání číslo 2 - spotřeba energie <i>Coastal Multi a Single Family</i> (drift).	111
6.73	Detail výsledků NARXnet uspořádání číslo 2 - spotřeba energie <i>Coastal Multi a Single Family</i> (gain).	111
6.75	Detail výsledků NARXnet uspořádání číslo 2 - spotřeba energie <i>Desert Multi a Single Family</i> (peak).	113
6.76	Detail výsledků NARXnet uspořádání číslo 2 - spotřeba energie <i>Desert Multi a Single Family</i> (skok).	113
6.77	Detail výsledků NARXnet uspořádání číslo 2 - spotřeba energie <i>Desert Multi a Single Family</i> (drift).	114
6.78	Detail výsledků NARXnet uspořádání číslo 2 - spotřeba energie <i>Desert Multi a Single Family</i> (gain).	114
A.6	Hledání optimálního nastavení parametrů ARX modelu pro data spotřeby energie „Desert Family“.	VII
A.7	Hledání optimálního nastavení parametrů ARX modelu pro data spotřeby energie „Inland Family“.	VII
A.8	Hledání optimálního nastavení parametrů ARX modelu pro data spotřeby energie „Mountain Family“.	VIII
A.9	Hledání optimálního počtu komponent GMM modelu pro data spotřeby energie „Desert Family“.	VIII
A.10	Rozdělení dat spotřeby energie „Desert Family“ GMM modelu.	IX
A.11	Hledání optimálního počtu komponent GMM modelu pro data spotřeby energie „Inland Family“.	IX
A.12	Rozdělení dat spotřeby energie „Inland Family“ GMM modelu.	X
A.13	Hledání optimálního počtu komponent GMM modelu pro data spotřeby energie „Mountain Family“.	X
A.14	Rozdělení dat spotřeby energie „Mountain Family“ GMM modelu.	XI
A.15	Hledání optimálního počtu komponent GMM modelu pro teplotní data.	XI
A.16	1) Rozdělení teplotních dat GMM modelu.	XII
A.17	2) Rozdělení teplotních dat GMM modelu.	XII
A.18	3) Rozdělení teplotních dat GMM modelu.	XIII
A.19	4) Rozdělení teplotních dat GMM modelu.	XIII

A.20 5) Rozdělení teplotních dat GMM modelu.	XIV
A.21 6) Rozdělení teplotních dat GMM modelu.	XIV

Seznam tabulek

4.1	Proměnné a parametry SISO ARX modelu.	26
4.2	Proměnné a parametry MISO ARX modelu.	28
6.1	Výsledky ARX uspořádání číslo 1 - teploty <i>Budihostice, Červený Újezd, Chloumek, ČZU</i>	59
6.2	Výsledky ARX uspořádání číslo 1 - spotřeba energie <i>Coastal Multi a Single Family</i>	60
6.3	Výsledky ARX uspořádání číslo 1 - spotřeba energie <i>Desert Multi a Single Family</i>	60
6.4	Výsledky ARX uspořádání číslo 2 - teploty <i>Budihostice, Červený Újezd, Chloumek, ČZU</i>	71
6.5	Výsledky ARX uspořádání číslo 2 - spotřeba energie <i>Coastal Multi a Single Family</i>	71
6.6	Výsledky ARX uspořádání číslo 2 - spotřeba energie <i>Desert Multi a Single Family</i>	71
6.7	Výsledky GMM - teploty <i>Budihostice, Červený Újezd, Chloumek, ČZU</i>	82
6.8	Výsledky GMM - spotřeba energie <i>Coastal Multi a Single Family</i>	82
6.9	Výsledky GMM - spotřeba energie <i>Desert Multi a Single Family</i>	83
6.10	Výsledky NARXnet uspořádání číslo 1 - teploty <i>Budihostice, Červený Újezd, Chloumek, ČZU</i>	93
6.11	Výsledky NARXnet uspořádání číslo 1 - spotřeba energie <i>Coastal Multi a Single Family</i>	94
6.12	Výsledky NARXnet uspořádání číslo 1 - spotřeba energie <i>Desert Multi a Single Family</i>	94
6.13	Výsledky NARXnet uspořádání číslo 2 - teploty <i>Budihostice, Červený Újezd, Chloumek, ČZU</i>	104

6.14	Výsledky NARXnet uspořádání číslo 2 - spotřeba energie <i>Coastal Multi a Single Family</i>	104
6.15	Výsledky NARXnet uspořádání číslo 2 - spotřeba energie <i>Desert Multi a Single Family</i>	105
6.16	Vybrané výsledky - teploty <i>Budihostice, Červený Újezd, Chloumek, ČZU</i> .	116
6.17	Vybrané výsledky - spotřeba energie <i>Coastal Families</i>	116
B.1	Výsledky ARX uspořádání číslo 1 - spotřeba energie <i>Inland Multi a Single Family</i>	XV
B.2	Výsledky ARX uspořádání číslo 1 - spotřeba energie <i>Mountain Multi a Single Family</i>	XVI
B.3	Výsledky ARX uspořádání číslo 2 - spotřeba energie <i>Inland Multi a Single Family</i>	XVI
B.4	Výsledky ARX uspořádání číslo 2 - spotřeba energie <i>Mountain Multi a Single Family</i>	XVI
B.5	Výsledky GMM - spotřeba energie <i>Inland Multi a Single Family</i>	XVII
B.6	Výsledky GMM - spotřeba energie <i>Mountain Multi a Single Family</i>	XVII
B.7	Výsledky NARXnet uspořádání číslo 1 - spotřeba energie <i>Inland Multi a Single Family</i>	XVIII
B.8	Výsledky NARXnet uspořádání číslo 1 - spotřeba energie <i>Mountain Multi a Single Family</i>	XVIII
B.9	Výsledky NARXnet uspořádání číslo 2 - spotřeba energie <i>Inland Multi a Single Family</i>	XIX
B.10	Výsledky NARXnet uspořádání číslo 2 - spotřeba energie <i>Mountain Multi a Single Family</i>	XIX

Kapitola 1

Úvod

1.1 Motivace

Mým cílem pro vypracování této diplomové práce je především snaha o změnu, konkrétně o změnu týkající se snížení plýtvání energií, ale i o snahu odhalit špatné fungování některých senzorů, které by mohly způsobit neefektivitu či až selhání řídicích systémů v různých odvětvích průmyslu nebo obyčejných domácností.

Detekce poruch senzorů je velice podstatné téma, jelikož na jejich správném fungování je postaveno velké množství řídicích systémů, aplikací a podobně. Například v bezpečnostních aplikacích by chybné údaje ze senzorů mohly způsobit nejen ztrátu s ohledem na ekonomická hlediska, ale především na lidské životy, u kterých je jejich cena nevyčíslitelná. Uvedený příklad se řadí ke krajním případům, ale na něm nejvíce si člověk uvědomí, jak jsou správně fungující senzory či včasná detekce poruch nepostradatelné. Částečným zvýšením bezpečnosti fungování systémů je použití redundantního počtu senzorů.

V práci se zaměřuji na metody detekce poruch senzorů, jejichž základem je využití externí informace z blízkého objektu stejného typu, veřejného datového zdroje nebo redundantních senzorů. Konkrétně jsem se zabýval problematikou plýtvání energie v systému HVAC (Heating, Ventilation and Air Conditioning), který je blíže popsán v kapitole 1.1.2. Detekce poruch byla prováděna u senzoru venkovní teploty vzduchu (OAT), jehož chybná informace může degradovat funkci řídicího systému HVAC, což vede ke zvýšené spotřebě energie. Jako zdroj externí informace jsem použil veřejně dostupná data z několika meteorologických stanic. Dalším případem byla detekce zvýšené spotřeby energie v rámci celého domu, kde jako externí informace byla použita data z blízké budovy stejného typu.

1.1.1 Plýtvání energií v budovách

Budovy představují jednu z nejrychleji rostoucích oblastí s vysokou spotřebou energie na Zemi. Podle zprávy „US Department of Energy 2009 Building Energy Data Book“ na budovy v USA připadá národní spotřeba elektřiny 72 %, zemního plynu 54 % a 39 % celkové spotřeby energie národa, viz [17].

Příčiny neefektivního využívání energie:

- Příkladem může být absence energetické koncepce způsobená architektonickým návrhem, který nebere v úvahu fyzikální vlastnosti budovy. Následně již projektanti stavební části instalují taková technická zařízení pro vytápění, větrání a chlazení, aby splnila zadané požadavky. Při použití koncepce by bylo možné uvažovat o zařízení s rekuperací tepla či dalších možných alternativách, převzato z [16].
- Energetická náročnost bývá často opomíjena, jelikož velké množství budov je objednáváno u dodavatele na klíč, což vede ke snaze minimalizovat investiční náklady, tedy k použití nejlevnějšího řešení na úkor budoucí energetické náročnosti.
- Další příčinou neefektivního využívání energie může být BMS (Building Management System), který se stará o vytápění a podobně. Tento systém může být různé kvality, jak v možnosti ovládání jednotlivých prvků, tak i z hlediska technického vybavení. Škoda může být způsobena i špatně poučenou obsluhou tohoto systému.
- Poruchy senzoru neboli odlehle hodnoty, skokové změny, drift či změna zesílení u naměřených hodnot vedou k tomu, že dochází ke zbytečnému topení či chlazení, protože řídicí systém má ze senzoru informaci, která neodpovídá skutečnosti.

Konkrétně HVAC systémy jsou jedním z hlavních spotřebitelů energie v budovách (uvádí se až 40 %, například v USA až 57 %). Aby bylo dosaženo vysoké energetické účinnosti, je nutné neustále sledovat a analyzovat výkonnost HVAC systémů po celou dobu jejich životnosti.

Detekce a diagnostika poruch HVAC systémů

Uvádí se, že 25 % až 40 % spotřebované energie v HVAC systémech je způsobeno poruchami, uvedeno v [13]. Aby však byly závady odhaleny, je potřeba získat a analyzovat nejrozumnější informace a data ze softwarových a hardwarových komponent HVAC.

Nicméně velmi málo reálných zařízení je detekcí a diagnostikou poruch vybaveno. Jedním z důvodů je, že by provozovatel systému musel ručně shromažďovat informace, které tyto algoritmy vyžadují, a interpretovat jejich výstupy pro rekonfiguraci HVAC systému. Vysoký potenciál v úspoře energie bude využit až v případě nasazení algoritmů detekce a diagnostiky poruch, které budou schopny automaticky řídit HVAC systém.

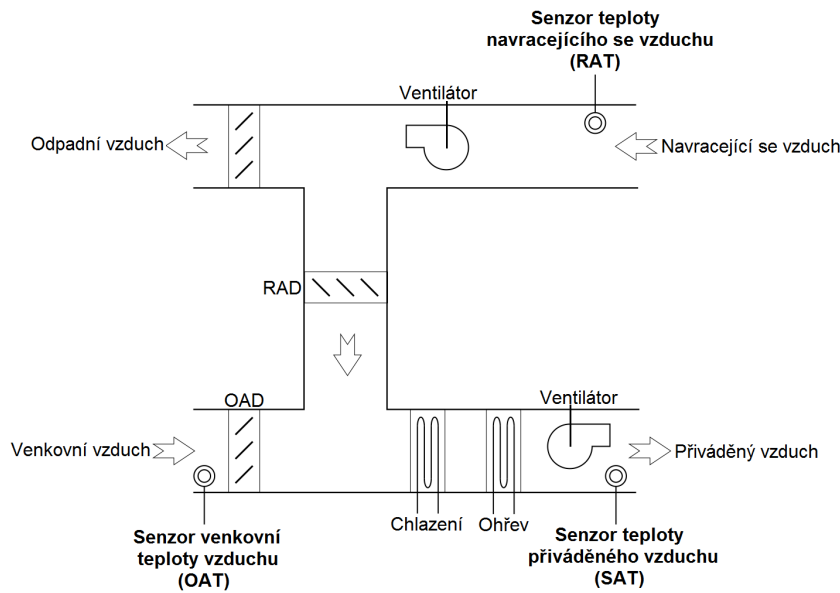
Porovnání účinnosti různých HVAC systémů

Účinnost HVAC systému je dána komponenty, konfigurací, strategií, údržbou, parametry budovy, klimatickými podmínkami a podobně. Samotné hodnocení účinnosti může probíhat na základě:

- Srovnání minulých a aktuálních hodnot, například spotřeby energie, bez ohledu na dynamiku provozu a změny, které mohly nastat, například jiné klimatické podmínky.
- Srovnání výkonnostních hodnot, například spotřeby energie, se základními hodnotami systému bez ohledu na složitost. Příkladem může být systém obsluhující budovu s významnými tepelnými úniky, kde jeho vyšší spotřeba energie nemusí nutně ukazovat na neefektivitu systému. Tedy i systém s vyšší spotřebou energie na plochu nemusí být méně účinný než ten s nižší spotřebou.

1.1.2 Dopad poruchy senzoru OAT na funkci HVAC

HVAC systém kombinuje tři základní funkce vytápění, větrání a klimatizace, které jsou vzájemně propojeny. Tento vynález je produktem průmyslové revoluce a následného zdokonalování v čase (zvyšování jeho účinnosti). HVAC systém má zajistit tzv. tepelný komfort a přijatelnou kvalitu vnitřního ovzduší s možným ohledem na cenu energie. Samotné zařízení nasává obvykle 15 % čerstvého vzduchu, který je přefiltrován a smíchán s částí dosavadního vzduchu. Dalším krokem je jeho ochlazení či ohřev a rozvedení po budově pomocí ventilátoru. Na závěr je část navraceného vzduchu odvedena pryč a část smíchána s čerstvým vzduchem, viz obrázek 1.1.



Obrázek 1.1: Schéma vzduchotechnické jednotky sloužící k úpravě a cirkulaci vzduchu jako součást HVAC.

Funkce celého systému je závislá na informacích, které dostává od senzorů teploty vzduchu. Porucha na některém ze senzorů by vedla k degradaci řídicí funkce systému, což by mělo za následek plýtvání s energií. Z tohoto důvodu bylo navrženo několik metod pro detekci poruch senzorů, které by plýtvání měly předejít. Základem metod, jak jsem již v kapitole 1.1 uvedl, je využití externí informace z blízkého objektu stejného typu, veřejného datového zdroje nebo redundantních senzorů, podrobně v kapitole 2. Rozhodnutí, zda má být vzduch v budově ohřát nebo ochlazen, je závislé nejen na vnějších klimatických podmínkách, konstrukci budovy, ale především i na správnosti měřených veličin detekovaných senzory.

Představte si situaci podle obrázku 1.1, kdy je skutečná teplota venkovního vzduchu například 19°C a teplota uvnitř budovy 22°C . Dále uvažujeme případ, že senzor venkovní teploty vzduchu (OAT) je vystaven prudkému slunečnímu záření nebo okolo senzoru proudí teplý vzduch od vyústění kuchyňské digestoře a podobně. V tomto okamžiku by daný senzor OAT začal ukazovat teplotu špatně, například o 5°C vyšší, tedy 24°C místo 19°C . Následně by vlivem chladnějšího počasí venku klesla teplota v budově z nastavených 22°C na 21°C . Systém HVAC by na tuto situaci musel zareagovat. Při správné funkci senzoru OAT by systém věděl, že venkovní teplota je nižší než v budově, a spustil by cirkulaci vzduchu uvnitř s mírným dohříváním vzduchu o 1°C . Ovšem při nesprávné

signalizaci senzoru OAT by systém HVAC situaci vyhodnotil tak, že bude navracející se vzduch o teplotě 21 °C vypouštět ven a venkovní vzduch podle signalizované teploty 24 °C nasávat dovnitř, což by mělo vést k ohřátí vzduchu. Ale teplota venkovního vzduchu je ve skutečnosti pouhých 19 °C, což způsobí dokonce ochlazení cirkulovaného vzduchu. Následně je nutný mnohem větší ohřev vzduchu, který vede k plýtvání s energií oproti případu se správnou funkcí senzoru venkovní teploty (OAT). Zde jsme si ukázali, že detekce podobných poruch je velice důležitá.

Diplomová práce se snaží řešit problematiku neefektivního využívání energie způsobem brzkého detekování poruchy (teplotního) senzoru nebo způsobem rozpoznání zcela odlišné spotřeby elektrické energie u blízkých objektů stejného typu na základě externích informací. Detaily jsou zmíněny v kapitole 2.

1.2 Data ke zpracování a zdroje externích informací

V této části se zaměřím na data, která byla poskytnuta pro testování vyvinutého algoritmu, a představím jejich charakteristické vlastnosti.

1.2.1 Data o spotřebě elektrické energie

Data o spotřebě elektrické energie byla stažena z veřejně dostupné internetové stránky <http://www.greenbuttondata.org/>, která podporuje myšlenku, že by odběratelé elektrické energie měli mít možnost si zjistit a stáhnout podrobné informace o spotřebě elektrické energie. Spotřebitelé si na základě těchto informací mohou nastavit vytápění a klimatizaci s ohledem na pohodlí versus šetření, vypočítat případné investice s cílem zvýšení energetických i finančních úspor nebo stanovit počet potřebných solárních panelů pro výrobu elektrické energie na střechu a podobně.

Jelikož základním kamenem všech metod, které jsou popsány v této diplomové práci, je využití externí informace z blízkého objektu stejného typu, tak jsem se rozhodl stáhnout tato data¹ o spotřebě elektrické energie:

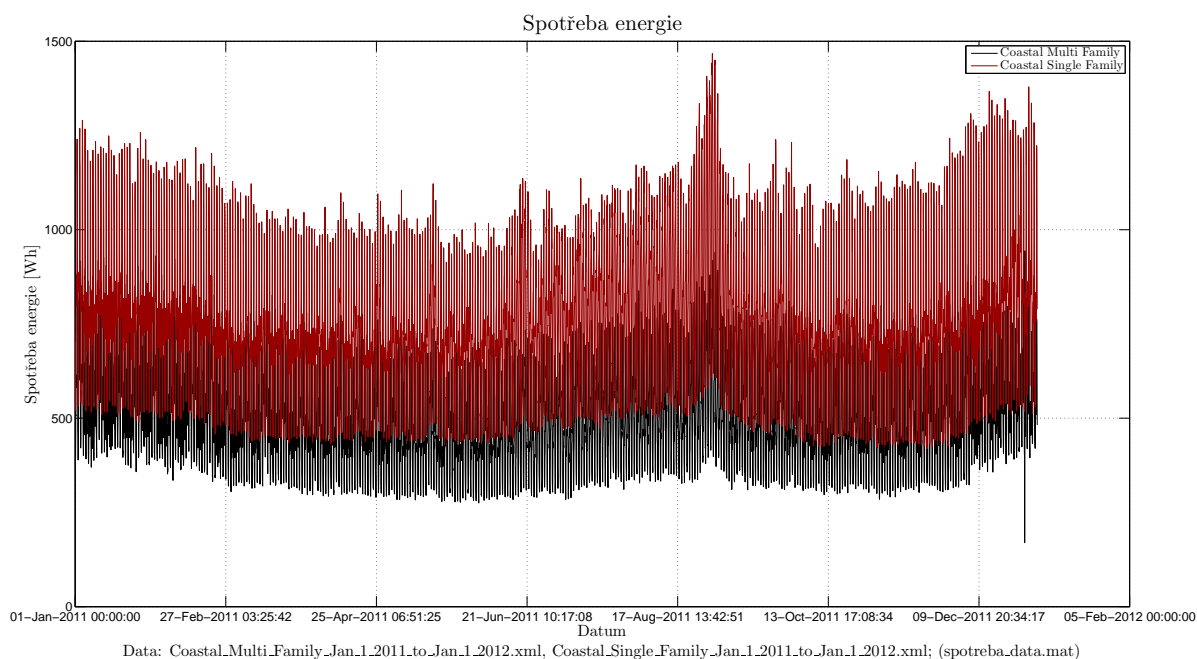
- Coastal_Multi_Family_Jan_1_2011_to_Jan_1_2012.xml,
Coastal_Single_Family_Jan_1_2011_to_Jan_1_2012.xml

¹Data o spotřebě elektrické energie byla následně převedena do souboru *spotreba_data.mat*.

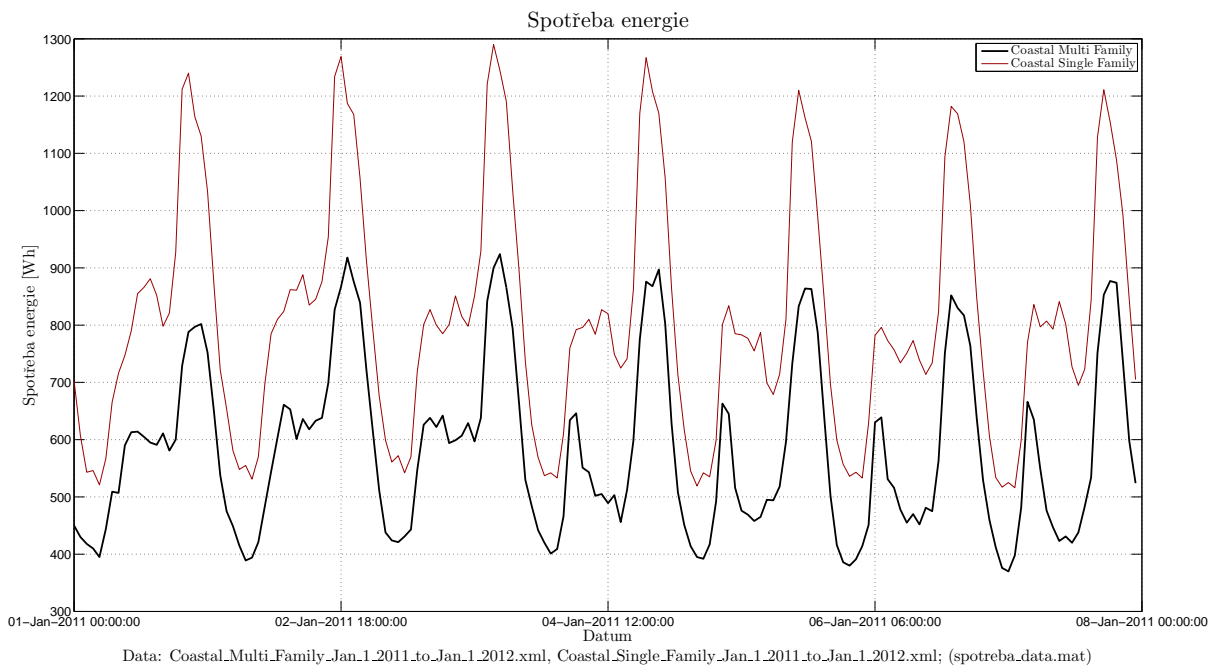
- Desert_Multi_Family_Jan_1_2011_to_Jan_1_2012.xml,
Desert_Single_Family_Jan_1_2011_to_Jan_1_2012.xml
- Inland_Multi_Family_Jan_1_2011_to_Jan_1_2012.xml,
Inland_Single_Family_Jan_1_2011_to_Jan_1_2012.xml
- Mountain_Multi_family_Jan_1_2011_to_Jan_1_2012.xml,
Mountain_Single_family_Jan_1_2011_to_Jan_1_2012.xml

Vždy se jedná o roční záznam spotřeby elektrické energie v období od 1.1.2011 00:00h do 1.1.2012 00:00h vzorkovaný po jedné hodině. Zvolená data jsou po dvojicích z různých oblastí (pobřeží, poušť, vnitrozemí, hory). Dvojice představuje dvě domácnosti s podobnou spotřebou elektrické energie. Při řešení jsou považována data z jedné domácnosti jako lokální a z druhé domácnosti jako externí.

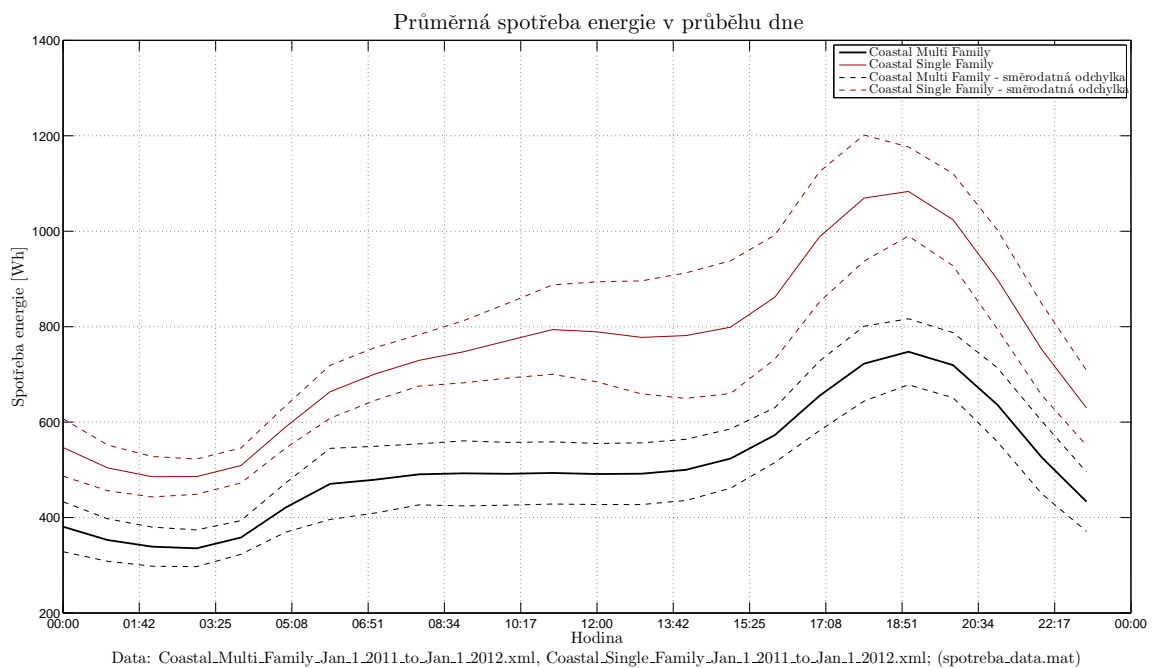
Obrázek 1.2 představuje ukázkou průběhu spotřeby elektrické energie dvou domácností z pobřežní oblasti v průběhu jednoho roku a obrázek 1.3 zobrazuje detail jednoho týdne, ze kterého je patrné, že tvar průběhů spotřeby energie jednotlivých dní je velice podobný. Tato závislost je dokázána a vykreslena na obrázku 1.4, kde vidíme průměrnou spotřebu elektrické energie za den včetně směrodatné odchylky. Data mají přibližně tvar jednoho nebo kombinace dvou normálních rozdělení, viz obrázek 1.5.



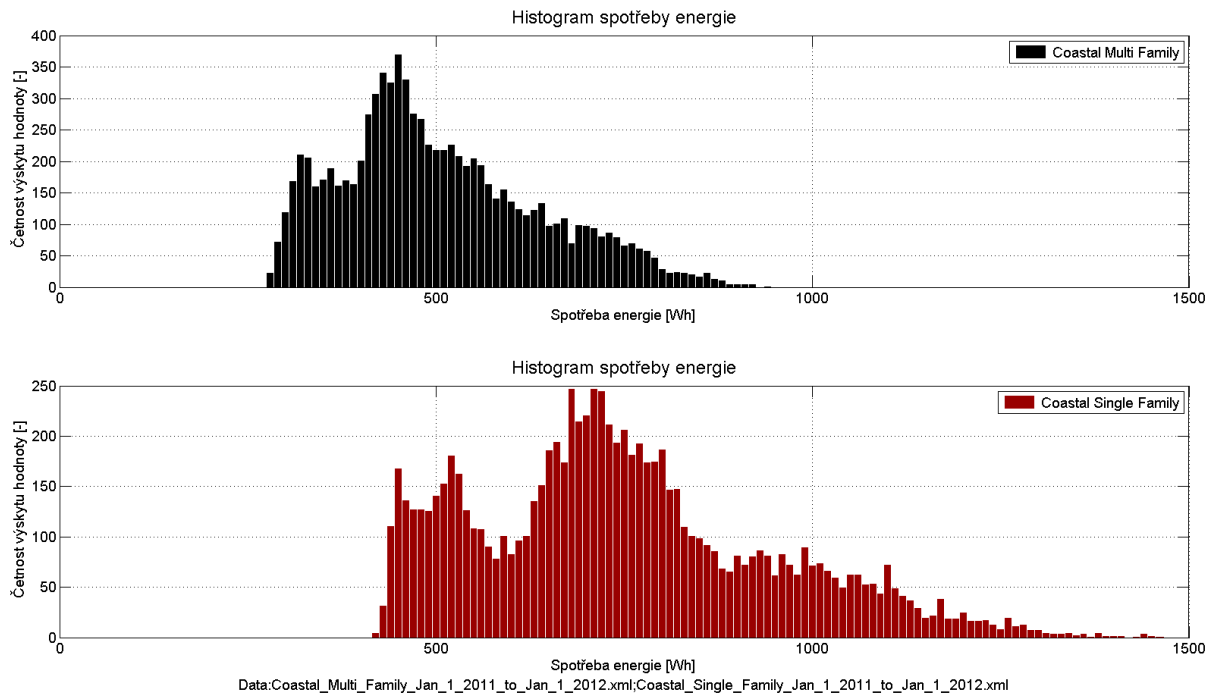
Obrázek 1.2: Spotřeba elektrické energie dvou domácností za 1 rok (pro ilustraci průběhů).



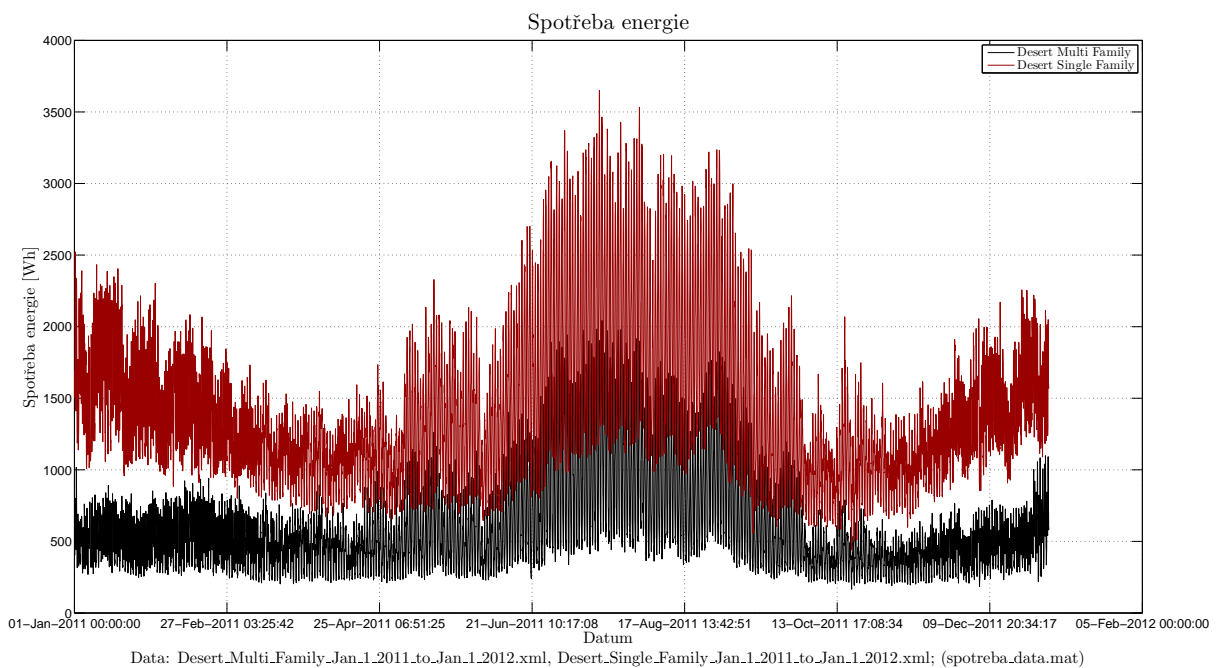
Obrázek 1.3: Spotřeba elektrické energie dvou domácností za 1 týden.



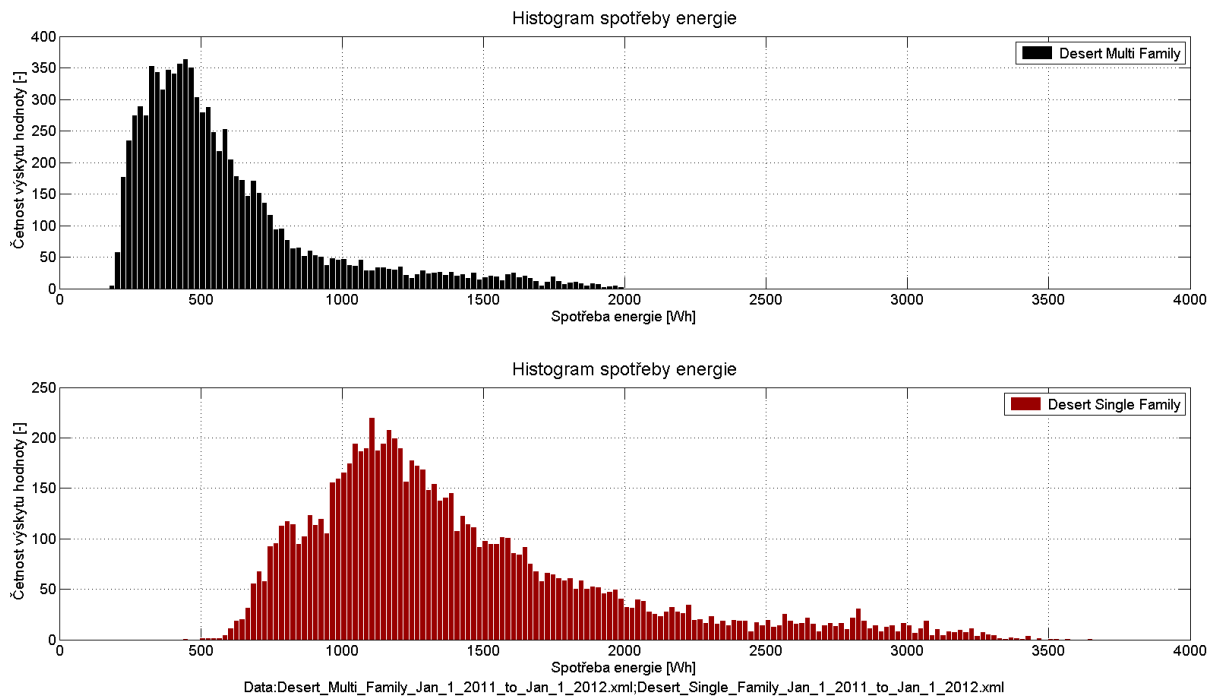
Obrázek 1.4: Průměrná spotřeba elektrické energie dvou domácností za den.



Obrázek 1.5: Histogramy spotřeby energie dvou domácností.



Obrázek 1.6: Spotřeba elektrické energie dvou domácností za 1 rok (pro ilustraci průběhů).

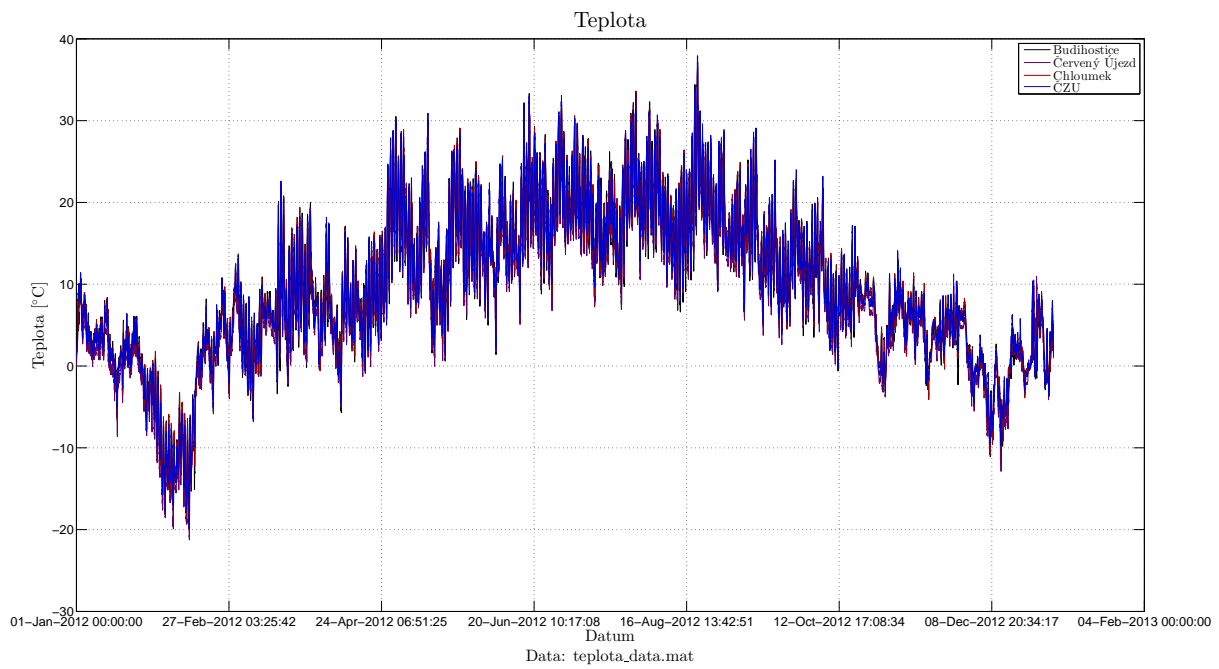


Obrázek 1.7: Histogramy spotřeby energie dvou domácností.

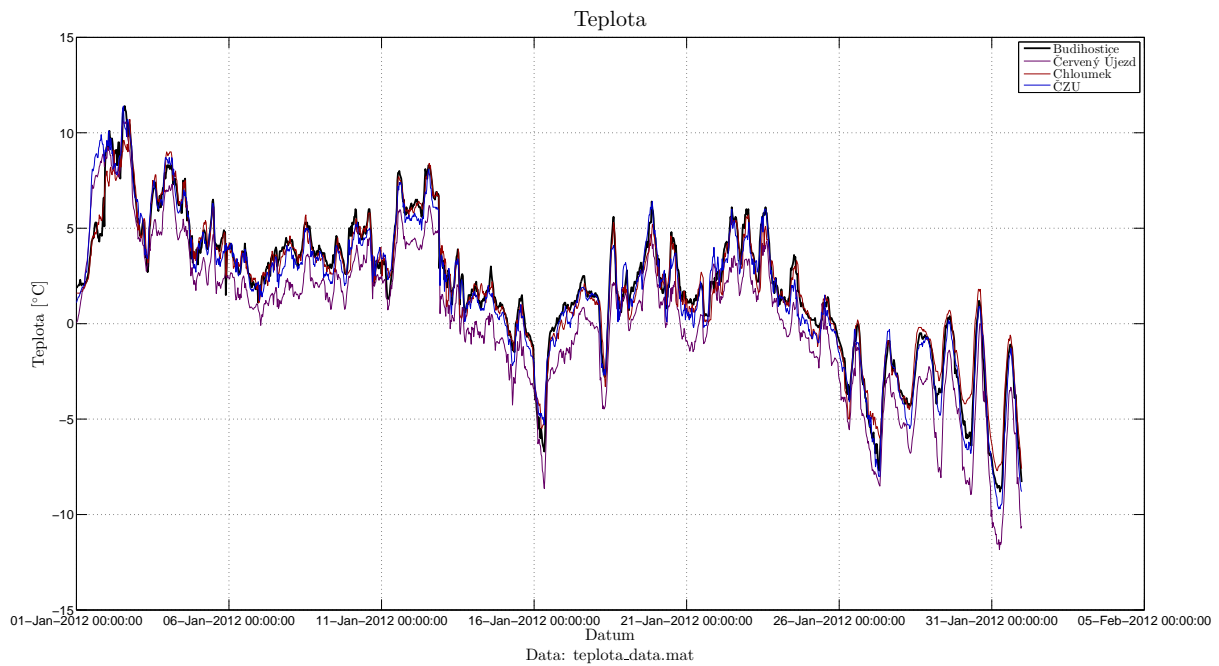
V rámci diplomové práce jsou tato data použita pro otestování algoritmu detekce poruch. Myšlenkou je včasné upozornění odběratele energie, že došlo k nepředpokládanému zvýšení odběru energie vůči obvyklému stavu.

1.2.2 Naměřené teploty ze čtyř meteorologických stanic

Zde uvažujeme naměřená data ze čtyř meteorologických stanic z jedné oblasti, která mi byla poskytnuta vedoucím diplomové práce. Konkrétně se jedná o naměřené hodnoty venkovní teploty v období od 1.1.2012 00:00h do 1.1.2013 00:00h vzorkované po třiceti minutách. Při řešení uvažujeme jednu sadu teplot jako lokální a zbývající trojici teplot jako externí informaci, jedná se tedy o situaci s redundantním počtem senzorů. Poskytnutá data jsou uložena v příloženém souboru *teplotni_data.mat*, který obsahuje naměřené teploty z meteorologických stanic Budihostice, Červený Újezd, Chloumek, ČZU (Česká zemědělská univerzita v Praze).

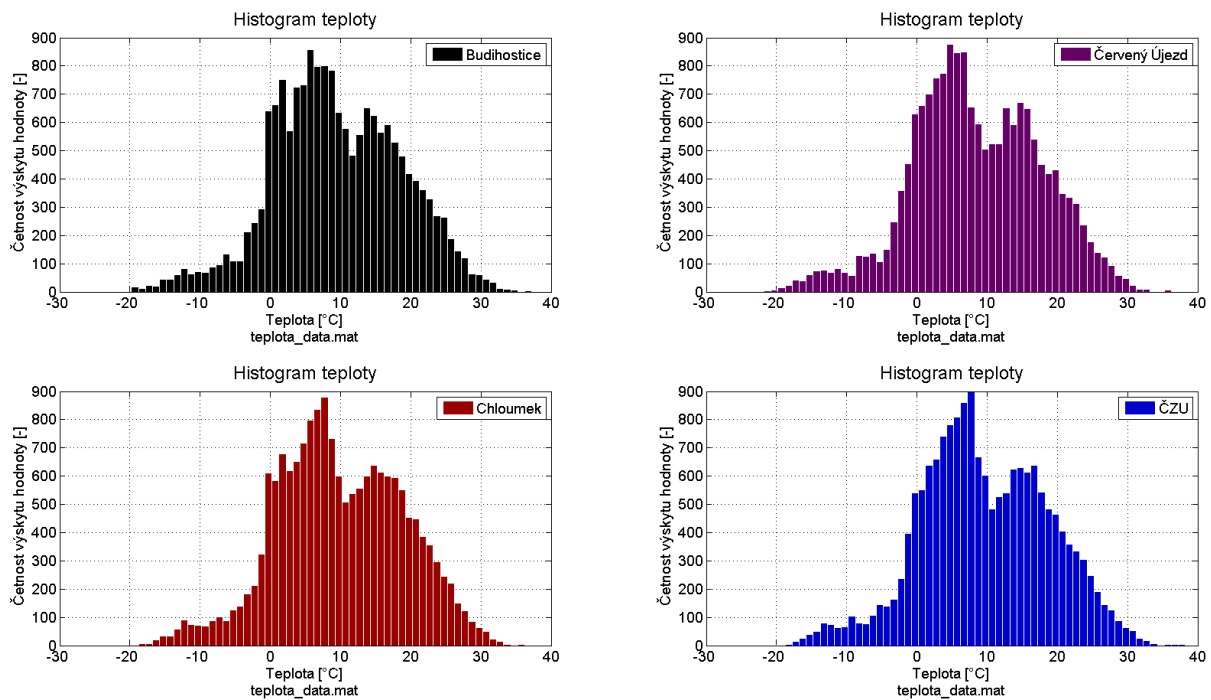


Obrázek 1.8: Průběh teplot ze čtyř meteorologických stanic za 1 rok (pro ilustraci průběhů).



Obrázek 1.9: Průběh teplot ze čtyř meteorologických stanic za 1 měsíc.

Na obrázku 1.8 vidíme ukázkou průběhu naměřených teplot ze čtyř meteorologických stanic v průběhu jednoho roku a na obrázku 1.9 je detail jednoho měsíce, ze kterého je patrné, že jednotlivé teploty si jsou velice podobné, až na hodnoty z meteorologické stanice Červený Újezd, které jsou mírně posunuté. Tato odchylka může být způsobena prostředím nebo polohou této meteorologické stanice. Jednotlivá data mají přibližně tvar jednoho nebo kombinace více normálních rozdělení, viz obrázek 1.10.



Obrázek 1.10: Histogramy teplot ze čtyř meteorologických stanic.

Na teplotních sadách dat je testována schopnost algoritmu detekovat nesprávné chování senzoru vůči počáteční situaci, jako je detekce odlehlé hodnoty, skokové změny a podobně. Cílem všech metod popsaných v kapitole 4 je informovat o špatné funkci lokálního senzoru, který by například mohl způsobit zbytečné zapnutí klimatizační jednotky, což by vedlo ke zvýšené spotřebě elektrické energie. Pro odhalení poruchy senzoru by bylo nejvhodnější použít strategii několika redundantních senzorů v daném místě, což z hlediska nákladů nemusí být nejlevnější možnost. Náš přístup může zmiňovanou redundanci vyřešit pomocí externích zdrojů informace, více viz kapitola 2.

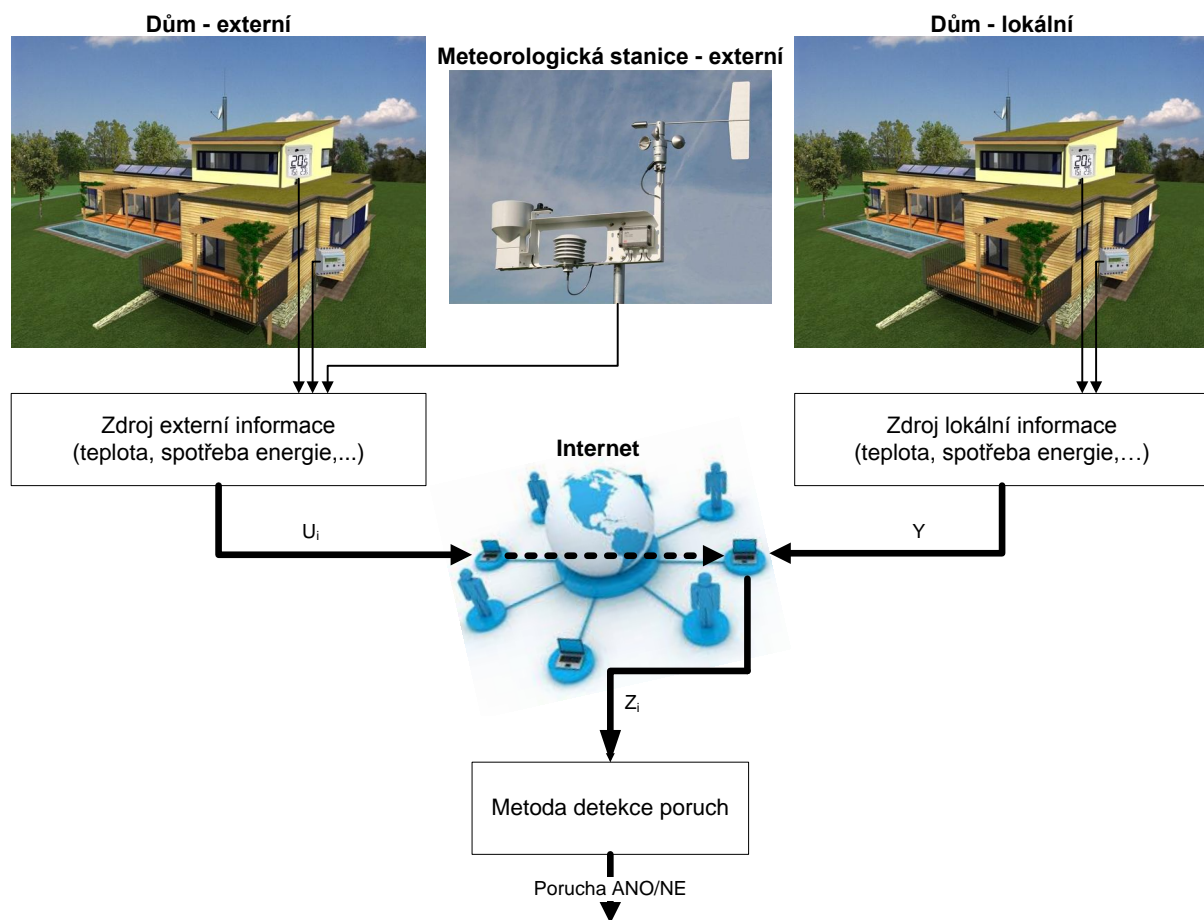
Kapitola 2

Cíl diplomové práce

Cílem této diplomové práce je na základě bibliografické rešerše navrhnout nové metody pro detekci poruch senzorů v aplikacích v budovách, zejména ve vzduchotechnice a spotřebě elektrické energie.

- Základním kamenem všech metod je využití externí informace z blízkého objektu stejného typu, veřejného datového zdroje nebo redundantních senzorů.
- Na základě bibliografické rešerše a po konzultaci s vedoucím diplomové práce jsou zvoleny tyto metody detekce poruch:
 - ARX (AutoRegressive model with eXternal input)
 - GMM (Gaussian Mixture Model)
 - NARXnet (Nonlinear AutoRegressive neural network with eXternal input)
- Úkolem je detekovat především trvale nesprávné chování senzoru a změnu jeho chování vůči počáteční situaci. Příkladem může být detekce odlehle hodnoty, skokové změny, driftu nebo změny zesílení.
- Dalším krokem je implementace příslušných metod v prostředí Matlab¹.
- Celkové zhodnocení a porovnání jednotlivých přístupů mezi sebou a následná diskuze nad dosaženými výsledky diplomové práce.

¹K implementaci byl konkrétně použit Matlab R2013b.



Obrázek 2.1: Principiální schéma pro detekci poruch senzorů.

Kapitola 3

Současný stav

Kapitola popisuje současný stav s ohledem na požadavky specifikované v kapitole 2 a v následující části se zaměřuje na problematiku detekce poruch a jejich druhy.

3.1 Bibliografická rešerše

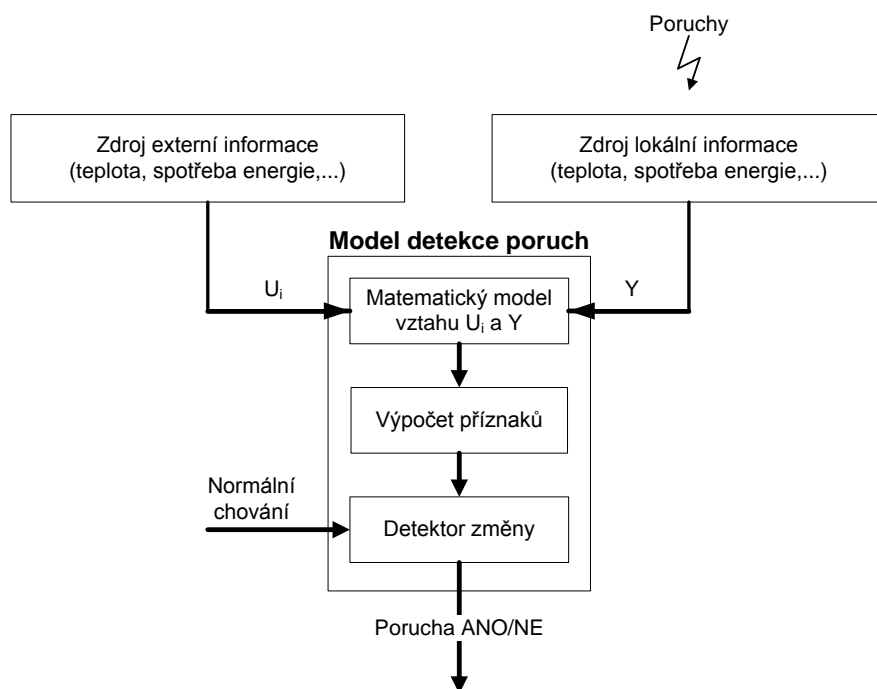
Jak bylo již zmíněno v kapitole 2, prvním krokem bylo provedení bibliografické rešerše neboli pátrání v oblasti veřejně dostupných publikací, zda se již někteří autoři danou problematikou nezabývali a jak případně postupovali. Mohu konstatovat, že existuje mnoho různých, ať obecných či speciálních přístupů detekce poruch, příkladem v tomto ohledu může být několik článků od pana *R. Isermanna*. Autoři uvažovali například průběžné nastavování lokálních mezí místo globálních, pozorovatele systému a podobně, ale nikde se mi nepodařilo nalézt případ, který by pracoval na základě externí informace dle schématu na obrázku 3.1. Vždy jsem narazil pouze na příklad, kdy autoři rozhodují o poruše na základě vstupu do systému a jeho výstupu.

Metody používající ARX (AutoRegressive model with eXternal input) se využívají v oblasti energetické efektivity budov, HVAC (vytápění, větrání a klimatizace), vibrodiagnostice a podobně. V článku [6] používají ARX k výpočtu příznaků, které slouží k detekci poruch AHU (Air Handling Unit), konkrétně predikují teplotu přiváděného vzduchu, tlak a průtok vzduchu a sledují zda nedošlo k překročení stanovených hranic a zda tedy nedošlo k poruše. Obdobný postup je popsán i v článku [19]. Metody pracující s GMM (Gaussian Mixture Model) modelem se používají v HVAC, chemických procesech atd. V článku [9] používají GMM v procesu čištění odpadních vod. Naopak článek [7]

se snažil ukázat, že statistické metody založené na podmíněné hustotě pravděpodobnosti jsou vhodné pro některé klasifikační úlohy. Článek [2] se zaměřuje na predikci spotřeby energie budov na základě vnější teploty, vlhkosti, slunečního záření a na určení nejistoty této predikce pomocí GMM. Metoda využívající NARXnet (Nonlinear AutoRegressive neural network with eXternal input) pro detekci poruch je například zmíněna v článku [5], který se zaměřuje na diagnostiku poruch v chemických procesech, nebo v článku [11], který se zabývá spolehlivostí softwaru. Nicméně žádný z uvedených článků nevyužívá znalosti externí informace k detekci poruch sensorů jako v našem případě, viz obrázek 2.1 a 3.1.

3.2 Detekce poruch

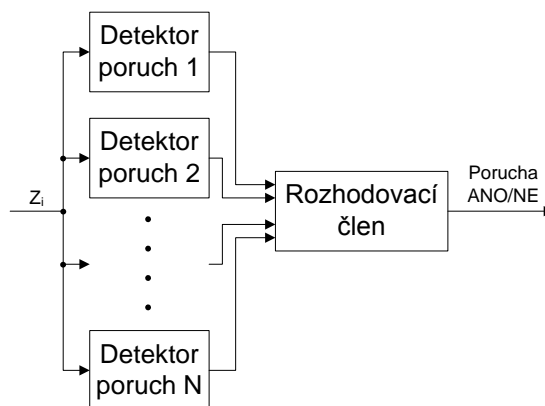
Základní metody detekce poruch využívají různé vztahy mezi měřenými veličinami s cílem získat informace o jejich možných změnách způsobených chybami. Vztahy bývají obvykle v analytickém tvaru v podobě matematických rovnic, ale mohou být vyjádřeny i formou nějakého pravidla.



Obrázek 3.1: Schéma modelu detekce poruch, [18].

Na obrázku 3.1 vidíme naše zjednodušené schéma modelu detekce poruch. Vztahy mezi měřenými vstupními hodnotami U a Y jsou právě zdefinovalé pomocí matematického modelu. Jeho výstupem jsou příznaky jako různé parametry, stavové proměnné a podobně. Následně jejich porovnáním v detektoru změn s nominální hodnotou je rozhodnuto, zda výstupem detekční metody poruch je standardní chování nebo zda byla detekována porucha, [18].

Naše řešení však rozhodně nekončí úvahou nad jednou metodou či modelem detekce poruch, který by rozhodoval, zda k poruše došlo či nedošlo. Myšlenka je dále rozvinuta do podoby uvedené na obrázku 3.2, kde uvažujeme, že celkové rozhodnutí o případné chybě je stanoveno pomocí většinového výstupu detektorů poruch nebo jejich vážením. Proměnná Z_i obsahuje veškeré vstupní hodnoty z obrázku 3.1.



Obrázek 3.2: Schéma rozhodování o poruše.

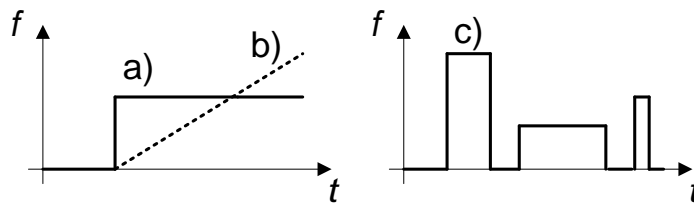
Nyní se vrátíme zpět k metodám detekujícím poruchy, u kterých je nutné pro správnou funkci zajistit adekvátní model těchto chyb, viz [18]. Realistický přístup předpokládá porozumění mezi skutečnými fyzickými poruchami a jejich vlivem na matematický či jiný model, což není vždy úplně možné. Důvodů pro vznik chyb je velká řada, například špatný design nebo provoz, špatná montáž či stárnutí. Tyto závady mohou být přítomny od začátku nebo se mohou objevit náhle s malou nebo velkou velikostí, v krocích a nebo postupně jako drift, ale výsledně se dají považovat za deterministické poruchy. Větší problém nastává u poruch náhodných.

3.3 Poruchy

Porucha může být definována jako nedovolená odchylka alespoň jednoho charakteristického příznaku od obvyklého stavu, viz 3.2. Příkladem chyby je například hodnota, která neleží v předpokládaném intervalu, který je odvozen ze standardního chování.

Poruchy, kterými se budu nadále především zabývat, jsou tyto:

- a) aditivní náhlá porucha v podobě skoku,
- b) aditivní pozvolná porucha neboli drift,
- c) aditivní občasná chyba,
- d) multiplikativní porucha (změna zesílení).



Obrázek 3.3: Časově závislé aditivní chyby: a) náhlá porucha, b) pozvolná porucha (drift), c) občasná porucha, [18].

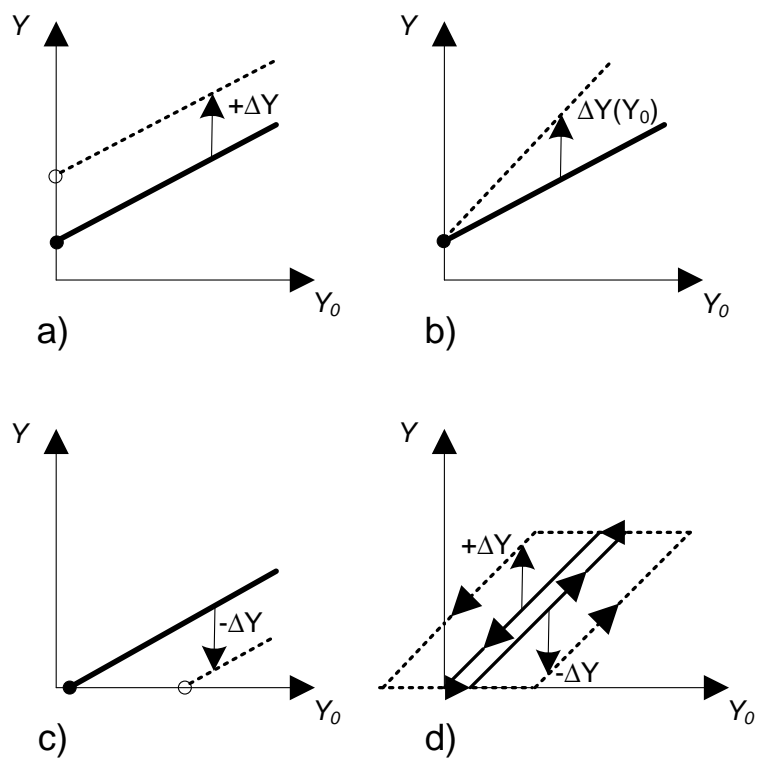
Změna signálu představuje aditivní chybu, protože k proměnné $Y_u(t)$ je přičteno $f(t)$, viz rovnice 3.1,[18]:

$$Y(t) = Y_u(t) + f(t) \quad (3.1)$$

Naopak multiplikativní chybu představuje změna parametru, protože proměnná $U(t)$ je vynásobena $f(t)$, rovnice 3.2,[18]:

$$Y(t) = (a + \Delta a(t))U(t) = aU(t) + \Delta a(t)U(t) = Y_u(t) + f(t)U(t) \quad (3.2)$$

V případě aditivní chyby je změna $\Delta Y(t)$ proměnné nezávislá na jakémkoli jiném signálu, $\Delta Y(t) = f(t)$. Nicméně v případě multiplikativní chyby je změna $\Delta Y(t)$ proměnné závislá na $U(t)$, konkrétně $\Delta Y(t) = f(t)U(t)$. Z toho vyplývá, že aditivní chyba je detekovatelná vždy, ale multiplikativní pouze pokud $U(t) \neq 0$.



Obrázek 3.4: Ukázka vlivu různých poruch na statický senzor (přečteno $Y(t)$, naměřeno $Y_0(t)$): a) posunutí počátku, b) změna zesílení (gain), c) změna odezvy, d) změna hystereze, [18].

Kapitola 4

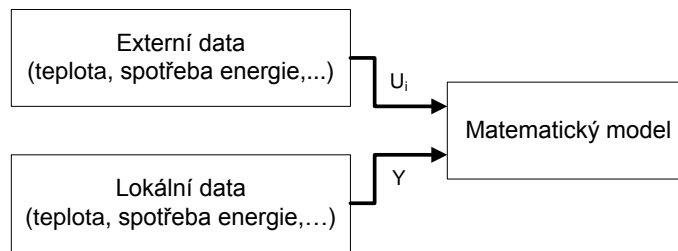
Teoretická část

V první části této kapitoly je naznačen princip jednotlivých metod detekce poruch a s jakými daty pracují. Druhá část obsahuje úvod do analýzy časových řad a na něj následně plynule navazuje teorie použitých metod detekce poruch.

4.1 Základní princip použitých metod detekce poruch

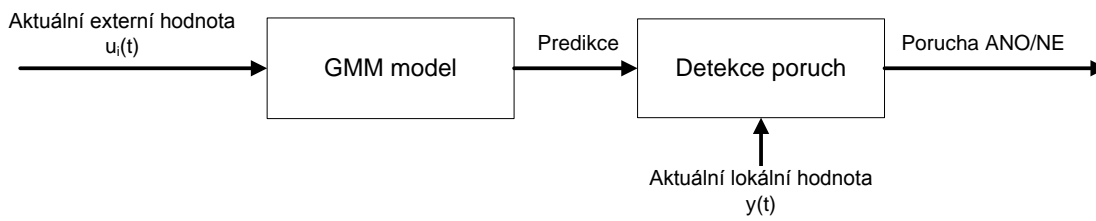
Veškeré uvedené metody detekce poruch, viz kapitola 2, které jsou použity v této diplomové práci, předpokládají, že existuje vztah mezi lokálními a externími daty z blízkého objektu stejného typu, veřejného datového zdroje nebo redundantních senzorů, jinak by tyto metody nebyly použitelné. Uvažuje se tedy uspořádání, které je zobrazeno na obrázku 2.1 a 3.1. Dále, jak bylo uvedeno v kapitole 1.2, k testování vyvinutých metod jsem dostal k dispozici teplotní data, u kterých je patrná silná vazba mezi lokálními a externími daty, a data spotřeby energie, kde je vazba slabší.

Prvním krokem je vytvoření matematického modelu, který popisuje vztah mezi lokálními a externími daty, viz obrázek 4.1. V případě ARX (AutoRegressive model with eXternal input) se jedná o sadu parametrů popisující model, u GMM (Gaussian Mixture Model) jde o vyjádření podmíněné pravděpodobnosti a u NARXnetu (Nonlinear Auto-Regressive neural network with eXternal input) to spočívá v trénování neuronové sítě. Podrobnější informace o vytvoření modelů u jednotlivých metod jsou popsány níže.



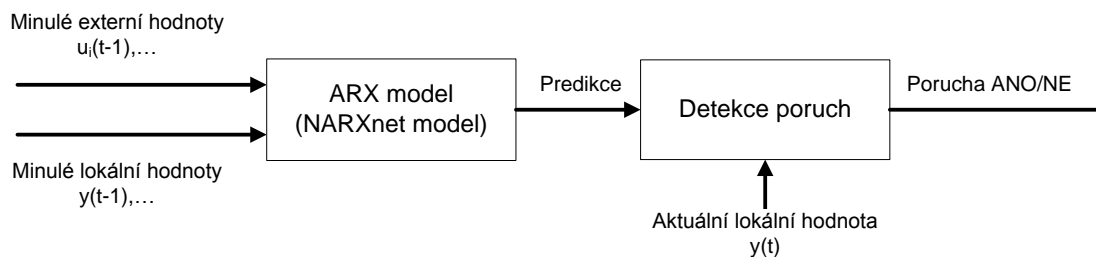
Obrázek 4.1: Zjednodušené schéma učení matematického modelu.

U GMM lze poté na základě stanoveného vztahu pro podmíněnou pravděpodobnost a aktuální externí hodnoty predikovat lokální hodnotu. Porovnáním této predikce s aktuální lokální hodnotou je možné vyhodnotit, zda nedošlo k poruše, viz obrázek 4.2.



Obrázek 4.2: Zjednodušené schéma detekce poruch pomocí GMM.

V případě ARX (AutoRegressive model with eXternal input) a NARXnet (Nonlinear AutoRegressive neural network with eXternal input) se vychází z minulých externích a lokálních hodnot, viz obrázek 4.3, na jejichž základě je predikována hodnota, která dále slouží k vyhodnocení, zda aktuální lokální hodnota je chybná. Tato problematika poukazuje na potřebu analýzy časové řady, která bude nyní vysvětlena.



Obrázek 4.3: Zjednodušené schéma detekce poruch pomocí ARX (NARXnet).

4.2 Analýza časové řady a její predikce

Časové řady jsou posloupnosti pozorování přijatých postupně v čase, viz [3]. Vyskytují se v takových oblastech, jako je ekonomika, přírodní nebo sociální věda, technika a podobně. Jejich vlastností obvykle bývá to, že sousední pozorování jsou na sobě závislá, což pro nás má značný praktický význam. Technika zabývající se touto závislostí se nazývá analýza časových řad.

Za předpokladu po sobě jdoucích pozorování, která jsou na sobě závislá, mohou být budoucí hodnoty predikovány pomocí minulých pozorování, viz [4]. V případě přesné predikce se jedná o deterministickou časovou řadu. Nicméně, většina časových řad je stochastická, jelikož jejich budoucí pozorování jsou určena jen částečně s určitým rozdělením pravděpodobnosti na základě minulých hodnot. Přesnost odhadů může být vyjádřena pomocí dolní a horní pravděpodobnostní hranice či meze, což představuje pravděpodobnost, se kterou reálná hodnota bude ležet v tomto intervalu mezi dolní a horní hranicí.

4.2.1 Cíle analýzy časových řad

Existuje několik možných cílů analýzy časových řad, může to být snaha o popis, vysvětlení, predikci nebo řízení, viz literatura [4].

Popis

Prvním krokem při analýze časové řady by mělo být její vykreslení, ze kterého je možné získat jednoduchý přehled o situaci a vlastnostech řady. Z obrázku 1.8 může být patrný například sezónní efekt nebo z dalšího obrázku 1.4 je vidět přibližně stejný opakující se průběh pozorování každý den. Obzvláště sofistikovanější metody vyžadují adekvátní analýzy s ohledem na vytvoření komplexnějšího modelu. Problémem mohou být například odlehlé hodnoty (outliers), které lze obvykle při vykreslení spatřit. Některé metody mohou být v závislosti na odlehlých hodnotách nefunkční, ale existují i robustní metody, kterým toto nevádí. Dále je vhodné hledat náhlé či pozvolné změny ve vlastnostech časové řady jako skoková změna, zlom vývoje trendu ze vzestupného na klesající a naopak nebo podobné nespojitosti, [4].

Vysvětlení

Při pozorování dvou nebo více proměnných je možné změny v jedné časové řadě použít k vysvětlení změn v některé z dalších časových řad, což může napomoci k hlubšímu pochopení mechanismu, který danou řadu vytvořil. Ačkoli více regresní modely jsou v některých případech užitečné, ne vždy jsou vhodné pro práci s časovými řadami, [4].

Predikce

Předpovídání budoucí hodnoty časové řady na základě minulých pozorování se řadí mezi nejobtížnější úlohy. Tato predikce má široké uplatnění například při předpovídání prodejů, při analýze hospodářských či průmyslových výsledků a podobně, [4].

Řízení

Cílem analýzy časových řad je zlepšit řízení či kontrolu nad nějakým ať už fyzickým či ekonomickým systémem. Problémy s řízením bývají úzce spojeny s předpověďmi v mnoha situacích. Příkladem může být přijetí vhodných nápravných opatření v případě, že predikovaná hodnota bude ležet mimo interval spolehlivosti. Řídicí postupy se velice liší, je možné provádět akce na základě jednoduchých diagramů, ale též sofistikovaněji použitím různých modelů, u kterých je možné volit optimální strategie pro úlohy, [4].

4.2.2 Předpovídání

Existuje mnoho postupů pro předpovídání budoucích hodnot, ale je nutné si uvědomit, že žádná jednotlivá metoda není univerzální, viz [4]. Je nutné, aby analytik zvolil takovou variantu, která nejlépe splňuje zadanou sadu podmínek, ale musí brát v úvahu i to, že předpovídání je forma extrapolace se všemi svými riziky. Dále by mělo být možné analýzu na základě vnější informace změnit.

Metody pro předpovídání lze rozdělit například do těchto skupin, [4]:

- **Subjektivní:** Tato metoda využívá úsudek, intuici, zpětnou vazbu od jiných analytiků a další relevantní znalosti, ale my požadujeme alespoň částečně objektivní odhad.
- **Jednorozměrné:** Jednorozměrná prognóza řady je založena pouze na minulých pozorováních, případně může být rozšířena a jednoduchou funkcí času. Nebere v úvahu další faktory.

- **Vícerozměrné:** Vícerozměrná prognóza závisí alespoň částečně na hodnotách jedné či více dalších řad nazývaných prediktor, a tímto se především budeme zabývat.

Další možné dělení je na:

- **Automatické,** které nevyžadují lidský zásah.
- **Neautomatické,** které ho naopak vyžadují.

Podobných rozdělení je možné vytvořit více například podle použití, počtu minulých pozorování atd.

Vícerozměrná regrese

Tento model předpokládá, že proměnná y je lineárně úměrná p proměnným x_1, x_2, \dots, x_p . Pak můžeme vícenásobný regresní model napsat jako:

$$y = \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p + e, \quad (4.1)$$

kde β_i jsou konstanty a e je šum. Rovnice 4.1 je lineární, ale parametry β_i mohou zahrnovat nelineární funkce pozorovaných proměnných. Dále v rovnici není výslovně zahrnut čas, ale lze jej zavést jako jednu z vysvětlujících proměnných. Obecně každá proměnná potřebuje index, který určuje, kdy byla změřena. Charakter modelu může být pozměněn, jestliže známe minulé hodnoty proměnné y , viz [4].

4.3 ARX (AutoRegressive model with eXternal input)

ARX (AutoRegressive model with eXternal input) model se řadí mezi nejpoužívanější lineární dynamické modely a jeho predikce patří mezi nejúčinnější z polynomiálních metod, protože výsledkem je řešení lineárních regresních rovnic v analytickém tvaru, navíc toto řešení je unikátní. Jinými slovy řešení vždy splňuje globální minimum ztrátové funkce. Nevýhodou ARX modelu je šum, který je součástí dynamiky systému.

Jestliže se nejedná o bílý šum, tak vazba mezi deterministickou a stochastickou dynamikou může zkreslit predikci ARX modelu. Pro minimalizaci chyby výpočtu lze nastavit řád modelu vyšší, než je skutečný, obzvláště když je poměr signál šum malý. Nicméně pokud se nastaví vyšší řád, tak to může ovlivnit některé dynamické vlastnosti modelu, například stabilitu.

4.3.1 SISO ARX (Single Input–Single Output)

SISO ARX model a jeho následná struktura je definována rovnicí 4.2.

$$\begin{aligned} y(t) + a_1 y(t-1) + a_2 y(t-2) + \dots + a_{n_a} y(t-n_a) = \\ = b_1 u(t-n_k) + b_2 u(t-n_k-1) + \dots + b_{n_b} u(t-n_k-n_b+1) + \epsilon(t), \end{aligned} \quad (4.2)$$

kde jednotlivé proměnné a parametry ARX modelu jsou uvedeny v tabulce 4.1.

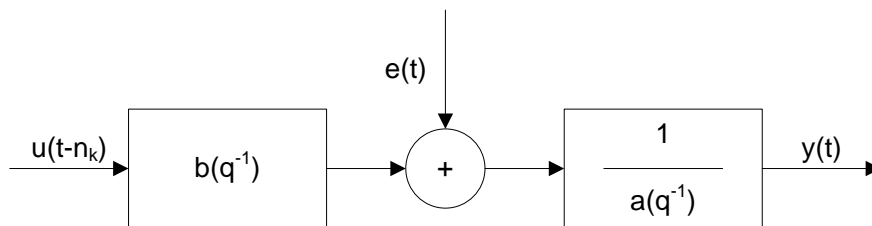
Tabulka 4.1: Proměnné a parametry SISO ARX modelu.

$y(t)$	hlavní výstup v čase t
a_1, \dots, a_n a b_1, \dots, b_n	hledané parametry ARX modelu
n_a	počet pólů systému
$n_b - 1$	počet nul systému
n_k	dopravní zpoždění
$y(t-1), \dots, y(t-n_a)$	předchozí výstupy, na kterých je závislý hlavní výstup
$u(t-n_k), \dots, u(t-n_k-n_b+1)$	předchozí vstupy, na kterých je závislý hlavní výstup
$\epsilon(t)$	šum

Model definovaný rovnicí 4.2 přepíšeme do jednodušší formy, viz blokové schéma na obrázku 4.4.

$$y(t) = \frac{b(q)}{a(q)} u(t-n_k) + \frac{1}{a(q)} \epsilon(t), \quad (4.3)$$

kde q je zpoždění (například $q^{-n}u(k) = u(k-n)$), $a(q) = 1 + a_1q^{-1} + \dots + a_{n_a}q^{-n_a}$ je polynom jmenovatele a $b(q) = b_1 + b_2q^{-1} + \dots + b_{n_b}q^{-n_b+1}$ je polynom čitatele.



Obrázek 4.4: Blokové schéma SISO ARX modelu.

4.3.2 Predikce modelu

Podle literatury [8] a rovnice 4.3 je patrné, že přenos deterministické části systému lze zvolit libovolně, zatímco tvarovací filtr náhodné složky je implicitně definován jako jmenovatel přenosu deterministické části modelu, viz blokové schéma 4.4. V důsledku tohoto významného omezení je predikovaná střední hodnota výstupu $\hat{y}(t)$ lineární funkcí měřitelných dat. K predikci modelu proto můžeme použít lineární regresi.

Provedeme lineární regresi, viz [8] a [14].

$$\hat{y}(t) = \sum_{i=1}^{n_b} b_i u(t - n_k - i + 1) - \sum_{i=1}^{n_a} a_i y(t - i) \quad (4.4)$$

neboli

$$\hat{y}(t) = b(q)u(t - n_k) + (1 - a(q))y(t) \quad (4.5)$$

Vektor hledaných parametrů modelu označíme jako $\theta = [a_1, a_2, \dots, a_{n_a}, b_1, b_2, \dots, b_{n_b}]^T$ a $z(t) = [-y(t-1), -y(t-2), \dots, -y(t-n_a), u(t-n_k), u(t-n_k-1), \dots, u(t-n_k-n_b+1)]^T$ jako vektor dat neboli regresor. Zjednodušený zápis závislosti predikované střední hodnoty výstupu na parametrech:

$$\hat{y}(t) = z^T(t)\theta = \theta^T z(t) \quad (4.6)$$

Zatímco reálná hodnota výstupu je:

$$y(t) = \hat{y}(t) + \epsilon(t) = z^T(t)\theta + \epsilon(t) = \theta^T z(t) + \epsilon(t) \quad (4.7)$$

4.3.3 Učení modelu

Dodefinováním vektoru měření výstupu

$$\mathbf{y} = [y_1(t), \dots, y_N(t)]^T \quad (4.8)$$

a matice dat

$$\mathbf{Z} = [z_1^T(t), \dots, z_N^T(t)]^T, \quad (4.9)$$

kde N je délka dat určených na učení modelu, samotné hledání parametrů ARX modelu spočívá ve vyřešení následující rovnice:

$$(\mathbf{Z}^T \mathbf{Z}) \theta = \mathbf{Z}^T \mathbf{y}, \quad (4.10)$$

$$\theta = (\mathbf{Z}^T \mathbf{Z})^{-1} \mathbf{Z}^T \mathbf{y}, \quad (4.11)$$

Je vidět, že rovnice 4.11 je řešena na množině dat určených k učení ARX modelu. Výsledkem jsou pak parametry modelu θ , které nejlépe splňují danou rovnici. Podrobný postup a vysvětlení identifikace parametrů lineárního regresního modelu je uvedeno v literatuře [8]. Implementace byla provedena v programu Matlab¹.

4.3.4 MISO ARX (Multiple Input–Single Output)

V této části bych se rád zmínil ještě o MISO ARX (Multiple Input–Single Output) modelu, který je spolu se SISO ARX modelem použit v praktické části.

MISO ARX model a jeho následná struktura je definována rovnicí 4.12.

$$\begin{aligned} y(t) + A_1 y(t-1) + A_2 y(t-2) + \dots + A_{n_a} y(t-n_a) = \\ = B_1 u(t-n_k) + B_2 u(t-n_k-1) + \dots + B_{n_b} u(t-n_k-n_b+1) + \epsilon(t), \end{aligned} \quad (4.12)$$

kde jednotlivé proměnné a parametry ARX modelu jsou uvedeny v tabulce 4.2.

Tabulka 4.2: Proměnné a parametry MISO ARX modelu.

$y(t)$	hlavní výstup v čase t
n_y	počet výstupů
n_u	počet vstupů
a_1, \dots, a_{n_a} a $b_{11}, \dots, b_{n_b n_u}$	hledané parametry ARX modelu
n_a	počet pólů systému
$n_b - 1$	počet nul systému
n_k	dopravní zpoždění
$y(t-1), \dots, y(t-n_a)$	předchozí výstupy, na nichž je závislý hlavní výstup
$u(t-n_k), \dots, u(t-n_k-n_b+1)$	vektor předchozích vstupů, na kterých je závislý hlavní výstup
$\epsilon(t)$	šum

Model definovaný rovnicí 4.12 přepíšeme do jednodušší formy, viz blokové schéma 4.5.

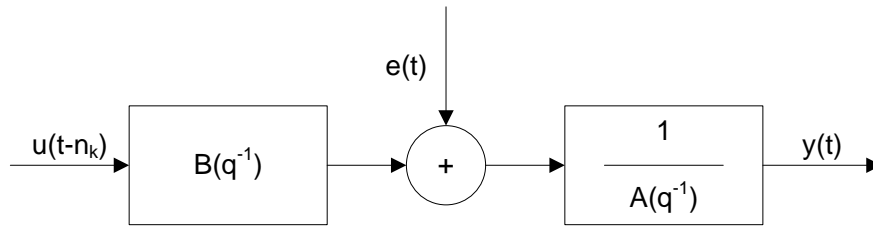
$$y(t) = \frac{B(q)}{A(q)} u(t-n_k) + \frac{1}{A(q)} \epsilon(t), \quad (4.13)$$

kde q je zpoždění, $A(q)$ je pro MISO ARX model pouze polynom a $B(q)$ je matice polynomů pro jednotlivé vstupy.

¹Pro nalezení parametrů ARX modelu byla použita funkce `arx`, více v nápovědě.

$$A(q) = 1 + a_1q^{-1} + \dots + a_{n_a}q^{-n_a}, \quad (4.14)$$

$$B(q) = \begin{bmatrix} b_{11} & b_{12} & \dots & \cdot \\ b_{21}q^{-1} & b_{22}q^{-1} & \dots & \cdot \\ \cdot & \cdot & \dots & \cdot \\ b_{n_b1}q^{-n_b+1} & b_{n_b2}q^{-n_b+1} & \dots & b_{n_bn_u}q^{-n_b+1} \end{bmatrix} \quad (4.15)$$



Obrázek 4.5: Blokové schéma MISO ARX modelu.

Samotné učení a predikce MISO ARX modelu probíhá podle stejných pravidel jako u SISO ARX modelu, viz 4.3.3 a 4.3.2.

4.4 GMM (Gaussian Mixture Model)

GMM (Gaussian Mixture Model) patří k dalším metodám pro predikci hodnot, kterou zde popíši. Zjednodušeně řečeno se jedná o proložení naměřených hodnot určených k naučení modelu pomocí směsi Gaussovských rozdělání. Podkapitola vychází z článku [2]. Velikou předností tohoto modelu je, že dokáže provádět dobrou predikci i u závislostí, které nejsou lineární. Konfidenční intervaly vychází z kumulativní distribuční funkce odezvy, což má za následek snížení počtu falešných poplachů. Častým nedostatkem jednotlivých metod je jejich omezená schopnost kvantifikovat nejistotu predikce, obvykle řešeno globálním konfidenčním intervalem. Tento přístup však vede nejen ke konzervativnímu odhadu nejistoty, ale také k zanesení nejistoty do celého rozsahu naměřených hodnot. A proto bývají odhady nejistot na globální úrovni vyšší než na lokální.

Výhody GMM (Gaussian Mixture Model)

- *Integrovaná predikce a kvantifikace nejistoty:* Predikce hodnot a jejich nejistota je určena pomocí podmíněné hustoty pravděpodobnosti predikce chyby. Výpočet parametrů modelu je proveden pomocí maximální věrohodnosti. Není nutné provádět lokální nebo globální odhad spolehlivosti.
- *Malý vliv korelovaných regresorů:* Korelované regresory mohou být vysoce citlivé na šum, což může způsobit špatné určení parametrů modelu a proto je velice důležité brát v úvahu závislosti mezi jednotlivými proměnnými, viz [2].
- *Menší citlivost na řídkost hodnot:* Metoda GMM (Gaussian Mixture Model) je méně citlivá na nízkou hustotu dat, viz [2].
- *Výběr struktury modelu:* Optimální počet komponent (Gaussovských rozdělení) GMM modelu je určen pomocí BIC (Bayesian Information Criteria).

4.4.1 Gaussian mixture model

Uvažujeme sadu vstupů (regresorů) U a výstupů (odezva) Y , které jsou dohromady reprezentovány proměnnou $Z = (Y, U)$. Dále tedy vektor $z = [y, u_1, \dots, u_d]^T$ je délky $(d+1)$, kde d je počet regresorů predikujících skalární hodnotu y , [2]. Hustotu pravděpodobnosti směsi Gaussovských rozdělení $\psi(z; \theta)$, kde K je počet komponent GMM modelu, lze psat pomocí rozdělení pravděpodobnosti proměnné Z s parametry θ takto:

$$\psi(z; \theta) = \sum_{k=1}^K \lambda^k \phi(z; \mu^k, \Sigma^k), \quad (4.16)$$

kde součet kladných skalárních směšovacích poměrů λ^k je roven jedné a kde hustota vícerozměrného normálního rozdělení ϕ pro každou komponentu je parametrizována vektorem středních hodnot μ^k délky $(d+1)$ a symetrickou pozitivně definitní kovarianční maticí Σ^k o velikosti $(d+1) \times (d+1)$, viz literatura [10] a rovnice 4.17.

$$\phi(z; \mu^k, \Sigma^k) = \frac{1}{(2\pi)^{d/2} |\Sigma^k|^{1/2}} \exp\left(-\frac{1}{2} (z - \mu^k)^T [\Sigma^k]^{-1} (z - \mu^k)\right) \quad (4.17)$$

Parametry modelu θ jsou vypočítány pomocí metody maximální věrohodnosti² na N datech určených pro naučení modelu, viz literatura [10] a rovnice 4.18.

$$\ell(\theta | z^{(i)}_1^N) = \sum_{i=1}^N \ln(\psi(z^{(i)}; \theta)) \quad (4.18)$$

²Implementováno pomocí statistického toolboxu pro GMM v programu Matlab.

4.4.2 Výběr počtu komponent

Pro určení optimálního počtu komponent GMM modelu jsem zvolil kritérium BIC (Bayesian Information Criteria). Nejprve vypočítáme celkový počet parametrů M :

$$M = \underbrace{K(d+1)}_{\text{vektor středních hodnot}} + \underbrace{\frac{1}{2}K(d+1)(d+2)}_{\text{kovariační matice}} + \underbrace{(K-1)}_{\text{směšovací poměry}} \quad (4.19)$$

Optimální počet komponent GMM modelu je vybrán jako hodnota K minimalizující BIC v rovnici 4.20:

$$BIC = -2\ell(\theta|z^{(i)N}_1) + M \ln(N) \quad (4.20)$$

4.4.3 Predikce modelu

Tato část popisuje postup predikce hodnot a s tím související chybu odhadu vzhledem k pozorovaným hodnotám regresorů. Přístup je velice efektivní, jelikož se jedná o vyřešení několika algebraických rovnic. Cílem je určit podmíněnou střední hodnotu $E(Y|U; \theta)$. Podmíněná hustota odezvy je dána rovnicí 4.21, viz [2].

$$\psi(y; \theta(u)) = \sum_{k=1}^K \lambda^k(u) \times \frac{1}{\sqrt{2\pi}\sigma_{Y|U}^k(u)} \exp\left(-\frac{1}{2} \left(\frac{y - \mu_{Y|U}^k(u)}{\sigma_{Y|U}^k(u)}\right)^2\right) \quad (4.21)$$

Z rovnice 4.21 je patrné, že se jedná o jednorozměrné GMM parametrizované parametry $\theta(u) = \lambda^k(u), \mu_{Y|U}^k(u), \sigma_{Y|U}^k(u)$, kde tyto parametry závisí na hodnotě u z regresorů U . Nyní už můžeme vyjádřit podmíněnou střední hodnotu a rozptyl náhodné veličiny Y , viz [2] a [10], s ohledem na rovnici 4.21. Proměnná $\hat{y}(u)$ představuje predikovanou hodnotu závislou na regresorech.

$$\hat{y}(u) = E_{\psi}(Y|U = u) = \sum_{k=1}^K \lambda^k(u) \mu_{Y|U}^k(u), \quad (4.22)$$

$$Var_{\psi}(Y) = \sum_{k=1}^K \lambda^k(u) (\sigma_{Y|U}^k(u)^2 + \mu_{Y|U}^k(u)^2) - \left(\sum_{k=1}^K \lambda^k(u) \mu_{Y|U}^k(u)\right)^2 \quad (4.23)$$

Prvním krokem je určení parametrů $\theta(u)$ GMM modelu pomocí hodnot u z regresorů U . Následně pomocí rovnice 4.22 a 4.23 vypočítáme odhad a jeho rozptyl.

Přehled

$$\mu^k = \begin{bmatrix} \mu_Y^k \\ \mu_U^k \end{bmatrix} \text{ s rozměry } \begin{bmatrix} 1 \times 1 \\ d \times 1 \end{bmatrix}, \quad (4.24)$$

$$\Sigma^k = \begin{bmatrix} \Sigma_{YY}^k & \Sigma_{YU}^k \\ \Sigma_{UY}^k & \Sigma_{UU}^k \end{bmatrix} \text{ s rozměry } \begin{bmatrix} 1 \times 1 & 1 \times d \\ d \times 1 & d \times d \end{bmatrix} \quad (4.25)$$

Pomocí této dekompozice můžeme definovat marginální hustotu normálního rozdělení k – té komponenty pro odezvu a regresory jako $\psi_Y^k(y; \mu_Y^k, \Sigma_{YY}^k)$ a $\psi_U^k(u; \mu_U^k, \Sigma_{UU}^k)$. Pak tedy:

$$\lambda^k(u) = \frac{\lambda^k \times \phi_U(u; \mu_U^k, \Sigma_{UU}^k)}{\sum_{i=1}^K \lambda^i \times \phi_U(u; \mu_U^i, \Sigma_{UU}^i)}, \quad (4.26)$$

$$\mu_{Y|U}^k(u) = \mu_Y^k + \Sigma_{YU}^k (\Sigma_{UU}^k)^{-1} (u - \mu_U^k), \quad (4.27)$$

$$(\sigma_{Y|U}^k(u))^2 = \Sigma_{YY}^k - \Sigma_{YU}^k (\Sigma_{UU}^k)^{-1} \Sigma_{UY}^k \quad (4.28)$$

4.4.4 Interval spolehlivosti predikce

V této části se zaměříme na výpočet intervalů spolehlivosti predikce GMM, viz [2]. Integrací podmíněné hustoty uvedené v rovnici 4.21 od $-\infty$ do y vypočítáme kumulativní distribuční funkci $\Psi(y; \theta(u))$.

$$\Psi(y; \theta(u)) = \sum_{k=1}^K \frac{\lambda^k(u)}{2} \left(1 + \operatorname{erf} \left(\frac{y - \mu_{Y|U}^k(u)}{\sigma_{Y|U}^k(u) \sqrt{2}} \right) \right) \quad (4.29)$$

Samotné konfidenční intervaly vypočítáme pro určitou hladinu významnosti α . Dolní a horní mez určíme vyřešením rovnic 4.30 a 4.31 s ohledem na y . Vzhledem k tomu, že chybová funkce erf má spojitou derivaci, tak lze tyto rovnice řešit iterativním způsobem.

$$\frac{\alpha}{2} = \sum_{k=1}^K \frac{\lambda^k(u)}{2} \left(1 + \operatorname{erf} \left(\frac{y - \mu_{Y|U}^k(u)}{\sigma_{Y|U}^k(u) \sqrt{2}} \right) \right) \quad (4.30)$$

$$1 - \frac{\alpha}{2} = \sum_{k=1}^K \frac{\lambda^k(u)}{2} \left(1 + \operatorname{erf} \left(\frac{y - \mu_{Y|U}^k(u)}{\sigma_{Y|U}^k(u) \sqrt{2}} \right) \right) \quad (4.31)$$

Dolní a horní mez predikované hodnoty se tedy určí vyřešením následujících rovnic 4.32 a 4.33.

$$\hat{y}_{DM} = \arg_y \left[\frac{\alpha}{2} - \sum_{k=1}^K \frac{\lambda^k(u)}{2} \left(1 + \operatorname{erf} \left(\frac{y - \mu_{Y|U}^k(u)}{\sigma_{Y|U}^k(u)\sqrt{2}} \right) \right) = 0 \right] \quad (4.32)$$

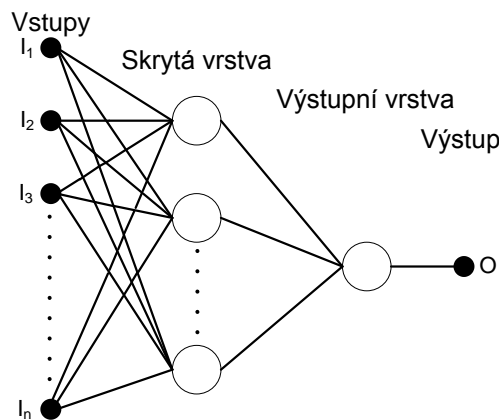
$$\hat{y}_{HM} = \arg_y \left[1 - \frac{\alpha}{2} - \sum_{k=1}^K \frac{\lambda^k(u)}{2} \left(1 + \operatorname{erf} \left(\frac{y - \mu_{Y|U}^k(u)}{\sigma_{Y|U}^k(u)\sqrt{2}} \right) \right) = 0 \right] \quad (4.33)$$

4.5 NARXnet

Na úvod se zmíním obecně o neuronových sítích a postupně se zaměřím na zvolenou metodu NARXnet (Nonlinear AutoRegressive neural network with eXternal input).

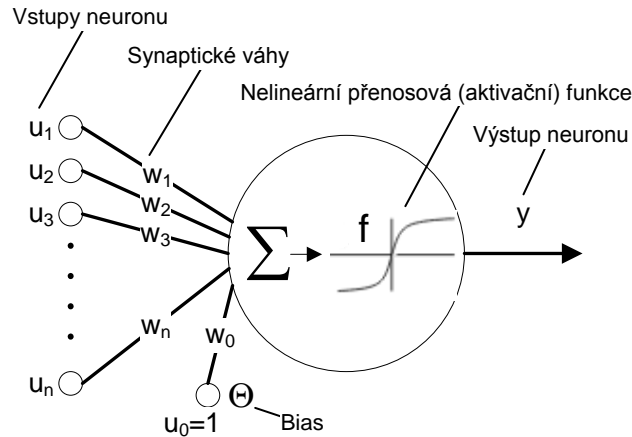
4.5.1 Úvod do neuronových sítí

Jedna z definic neuronové sítě praví, že se jedná o distribuovaný výpočetní systém sestavený z dílčích podsystémů (neuronů), který je inspirován neurofyziologickými poznatky o struktuře a činnosti neuronů a nervových systémů živých organismů a který je ve větší či menší míře realizuje, viz [1]. Neuronové sítě jsou definovány jako nedeklarativní systémy umělé inteligence. Nemusíme dopředu definovat pravidla, kterými se neuronová síť řídí. Zaměřím se na vícevrstvé neuronové sítě, které jsou schopny řešit širší okruh problémů než jednovrstvé.



Obrázek 4.6: Struktura dvouvrstvé dopředné neuronové sítě.

Základní jednotkou vícevrstevných neuronových sítí je neuron (perceptron). Tento neuron je samostatně pracující element s mnoha vstupy a většinou jedním výstupem. Jeho výstupní signál je závislý pouze na vstupních úrovních.



Obrázek 4.7: Model umělého neuronu.

Model umělého neuronu se skládá ze vstupů u_i , vah w_i , těla, biasu Θ a jednoho výstupu y . Tento model transformuje vektor vstupů $[u_1, \dots, u_n]^T$ na výstup y dle vzorce 4.34.

$$y(u_i) = f\left(\Theta + \sum_{i=1}^n w_i u_i\right) = f\left(\sum_{i=0}^n w_i u_i\right) = f(\zeta), \quad (4.34)$$

kde bias Θ lze nahradit vahou w_0 na vstupu $u_0 = 1$ a kde f označuje aktivační funkci popsanou níže.

Váhy neuronu nastavují jednotlivým vstupům určitou důležitost a jejich nastavení či adaptace probíhá v procesu učení. V neuronu probíhají dva procesy: výpočet vnitřního (post-synaptického) potenciálu a výpočet hodnoty výstupu pomocí aktivační funkce. Znalost vah a biasu postačuje k popisu prvního:

$$\zeta = \Theta + \sum_{i=1}^n w_i u_i, \quad (4.35)$$

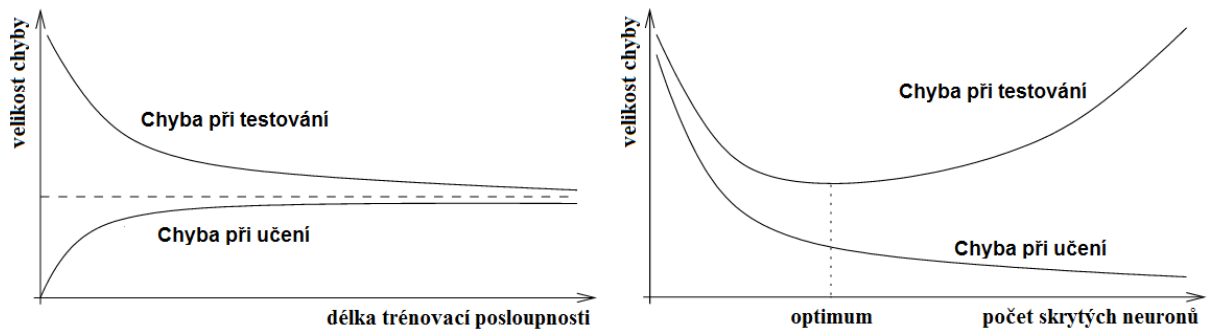
kde ζ se nazývá vnitřní potenciál.

Tělo neuronu tvoří přenosová (aktivační) funkce, jejichž typů je hodně, příkladem může být signum, která ale není diferencovatelná, a proto se u vícevrstevných sítí používají jiné funkce, které se jí svým průběhem blíží, ale diferencovatelné jsou, viz [15]. Nejčastější funkcí bývá sigmoida:

$$f(\zeta) = \frac{1}{1 + e^{-\zeta}} \quad (4.36)$$

Obecně hlavní výhodou neuronových sítí je schopnost naučení se na daný problém pouze na základě předkládaných vzorů.

Data určená pro naučení neuronové sítě bývají obvykle rozdělena do tří skupin na trénovací, validační a testovací. Trénovací množina obsahuje vzory sloužící k nastavení vah a k hledání nejlepší konfigurace. Proces učení vícevrstvé neuronové sítě se snaží nalézt takové váhy w , které minimalizují odchylku mezi skutečným a požadovaným výstupem na trénovací množině. Díky diferencovatelnosti aktivační funkce lze pro učení použít gradientní metodu optimalizace. Ta je založena na výpočtu gradientu ztrátové funkce a následném upravení všech vah ve směru gradientu, viz [15]. Úpravy se provádí pro každý vzorek z trénovací množiny tak dlouho, dokud nenalezneme přijatelné řešení. Validací data jsou určena k testování chyby při učení a stopnutí trénování, aby nedošlo k přeučení. Poslední, testovací množina slouží k testování naučené sítě. Na obrázku 4.8 bych rád ještě ukázal, jaký vliv má délka trénovací množiny nebo počet neuronů ve skryté vrstvě.



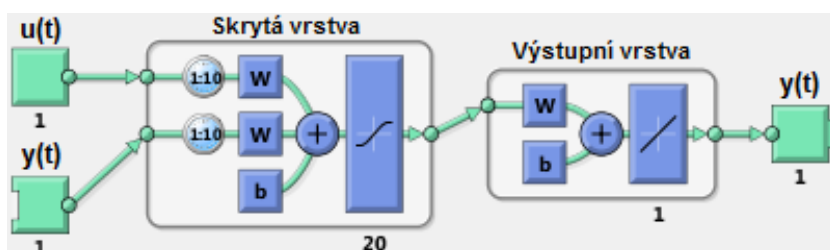
Obrázek 4.8: Závislost velikosti chyb na délce trénovací množiny vlevo a závislost velikosti chyb na počtu neuronů ve skryté vrstvě vpravo, převzato z [15].

Přesnost vícevrstvých neuronových sítí nebo chybu na testovacích datech může ovlivnit, [15]:

- Použitý algoritmus učení a počet iterací - určuje, jak dobře je síť naučená
- Délka trénovací množiny - určuje, jak přesně reprezentují vzorky z trénovací množiny hledanou funkci
- Počet neuronů ve skryté vrstvě - udává „vyjadřovací schopnost“ sítě, u hladkých funkcí jich stačí méně než u složitějších funkcí

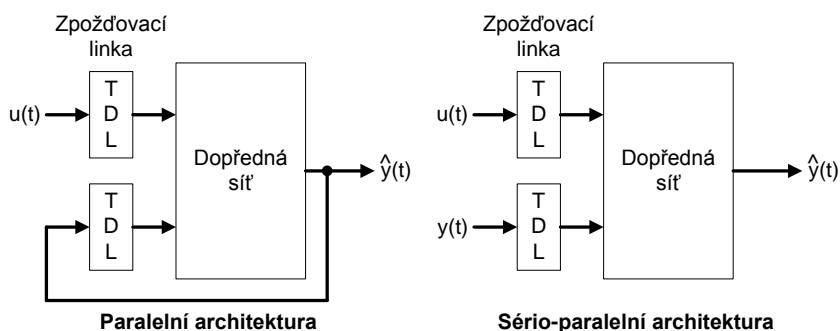
4.5.2 NARXnet (Nonlinear AutoRegressive neural network with eXternal input)

Model NARXnet je založen na lineárním modelu ARX, viz kapitola 4.3, který se běžně používá v modelování časových řad.



Obrázek 4.9: Ukázka struktury NARXnet.

Obrázek 4.9 znázorňuje standardní NARXnet síť. Jedná se o dvouvrstvou dopřednou síť s aktivační funkcí sigmoida ve skryté vrstvě a lineární aktivační funkcí ve výstupní vrstvě. Dále tato síť používá zpožďovací linky pro ukládání minulých vstupů $u(t)$ a výstupů $y(t)$. Konkrétně na obrázku 4.9 vidíme, že zde máme 1 vstup a 1 výstup, zpožďovací linku na 10 hodnot (trénování začne až s 11. prvkem), skrytou vrstvu s 20 neurony a výstupní vrstvu s 1 neuronem.



Obrázek 4.10: Obrázek paralelní a sério-paralelní architektury pro učení, převzato a upraveno z [12].

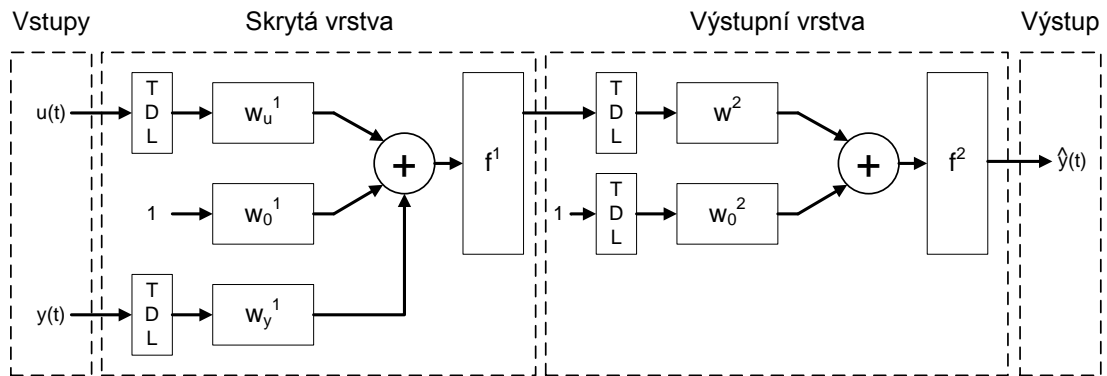
Na obrázku 4.10 vidíme dvě principiální schémata učení modelu NARXnet. V našem případě uvažujeme sério-paralelní architekturu, která má dvě výhody. První z nich je, že vstupy do dopředné sítě jsou přesnější, a druhou výhodou je, že lze použít „back

propagation“ neboli algoritmus zpětného šíření pro trénování neuronové sítě, kde se chyba šíří zpětně přes všechny vrstvy k první vrstvě, podrobněji v [15]. Musí ale být známa vstupní a výstupní dvojice hodnot, které v našem případě známe.

Model NARXnet je definován rovnicí:

$$y(t) = g(y(t-1), \dots, y(t-n_y), u(t-1), \dots, u(t-n_u)), \quad (4.37)$$

kde výstupní hodnota $y(t)$ je určena pomocí minulých vstupních, výstupních hodnot a kde n_u je počet zpožděných vstupů a n_y je počet zpožděných výstupů. Naše implementace tohoto modelu spočívá v aproximaci funkce g pomocí dopředné neuronové sítě, viz obrázek 4.11, aby byla minimalizována odchylka mezi skutečným a požadovaným výstupem. Díky tomu dostaneme nejlepší hledanou konfiguraci s nastavením jednotlivých vah, kterou lze použít pro předpovídání hodnot. Toto provedení umožňuje použití více vstupů a výstupů.



Obrázek 4.11: Výsledný model NARXnet, převzato a upraveno z [12].

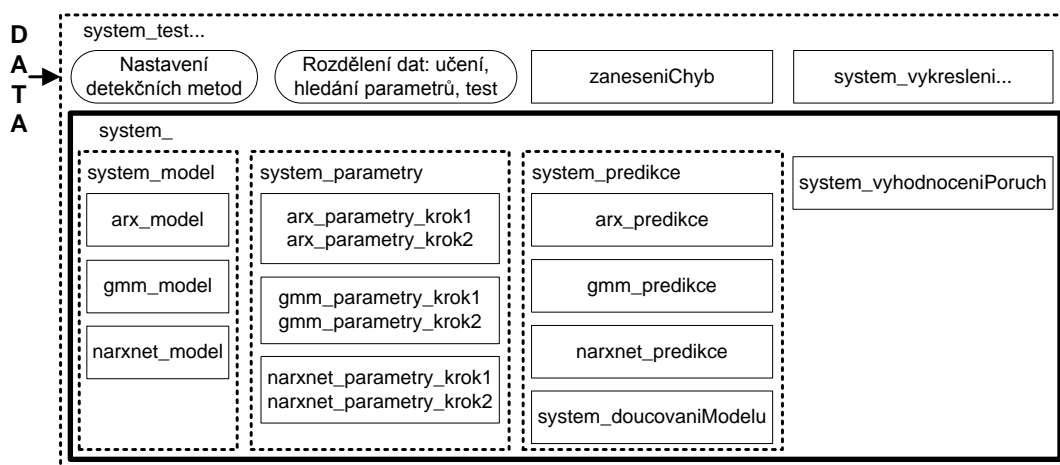
Kapitola 5

Implementace

Kapitola popisuje skladbu implementovaného toolboxu v programu Matlab R2013b a seznamuje s jednotlivými funkcemi. První část je zaměřena na hlavní funkce celého systému, zatímco druhá část obsahuje konkrétnější podobu implementace jednotlivých metod pro detekci poruch senzorů.

5.1 Hlavní struktura systému

Program je rozdělen do několika částí, které obstarávají jednotlivé úkoly. Postupně dochází k naučení modelů jednotlivých metod pro detekci poruch, hledání mezí, predikci a na závěr k vyhodnocení poruch.



Obrázek 5.1: Blokové schéma vyjadřující strukturu toolboxu v Matlabu.

1) **system_test.m** - Tato část je vyčleněna pro práci uživatele, který zde provede načtení dat určených k diagnostice na přítomnost poruch u senzorů. Tato data musí být ve formátu matice tak, že první sloupec bude obsahovat lokální data určená k otestování a ostatní sloupce budou obsahovat externí data z blízkého objektu stejného typu, veřejného datového zdroje nebo redundantních senzorů. V mém případě se jednalo o naměřené hodnoty spotřeby energie a teplot uložené v souborech *spotreba_data.mat* a *teplota_data.mat*. Dále je nutné data rozdělit do tří skupin určených na učení modelu, na hledání parametrů ke stanovení intervalu spolehlivosti predikce neboli dolní a horní meze a na samotné testování přítomnosti poruch u senzorů. Dostal jsem za úkol, aby data na samotné učení a stanovení mezí byla co nejkratší. Důvodem bylo, aby byly metody detekce poruch použitelné v praxi. Data jsem se proto rozhodl rozdělit přibližně takto: měsíční záznam dat na učení, měsíční na hledání potřebných parametrů a zbylých 10 měsíců dat na testování algoritmů. Samotné rozdělení dat do takto velkých skupin může velice ovlivnit kvalitu systému detekce poruch senzorů, a to především v případě, kdy testovací data mají velice odlišný rozsah a průběh než data vyčleněná na učení modelů. Posledním a nejdůležitějším krokem uživatele je zvolení příslušné testovací metody detekce poruch a zadání jejího vstupního nastavení.

- **nastaveni.ARX** = struct('na', -, 'nb', -, 'nk', -, 'on_off_hledani', -, 'on_off_vykresleni', -, 'procentoChyb', -, 'maxPorucha', -, 'indexChyb', -);
- **nastaveni.GMM** = struct('k', -, 'on_off_hledani', -, 'on_off_vykresleni', -, 'alfa', -, 'procentoChyb', -, 'maxPorucha', -, 'indexChyb', -);
- **nastaveni.NARXNET** = struct('nu', -, 'ny', -, 'neurony', -, 'procentoChyb', -, 'maxPorucha', 'indexChyb', -);
 - Parametry příslušející jednotlivým metodám n_a , n_b , n_k , k , $alfa$, n_u , n_y , $neurony$ jsou popsány v kapitole 4.
 - V případě nastavení parametru *on_off_hledani* na 0/1 dojde ke spuštění algoritmu na hledání optimálního nastavení již zmíněných hodnot n_a , n_b nebo k . V případě metody založené na ARX se hodnotí kvalita modelu při změnách počtu pólů a nul n_a , n_b od jedné do jejich nastavené hodnoty. U metody GMM je vybírán optimální počet komponent k modelu od jedné do nastavené hodnoty, který nejlépe splňuje kritérium BIC popsané v kapitole 4.
 - Nastavení *on_off_vykresleni* na 0/1 spouští vykreslení závislostí při hledání optimálního nastavení v minulém bodě a navíc u metody GMM se zobrazí

2D a 3D model Gaussovských rozdělení.

- Parametr *procentoChyb* slouží k nastavení mezí okolo predikované hodnoty, ale také k tomu, aby bylo možné mezi sebou porovnat použité metody detekce poruch. Představuje tzv. procento dovolených chyb neboli procentuální počet překročení mezí testovanou lokální hodnotou. Nízká hodnota způsobí nastavení širokého konfidenčního intervalu a naopak vysoká způsobí označení většiny hodnot za poruchu.
- Hodnota *maxPorucha* slouží k výslednému vyhodnocení, zda se skutečně jedná o chybu nebo ne. Rozhodnutí, zda došlo k poruše, je řešeno pomocí součtu velikostí překročení dolní nebo horní meze skutečnou lokální hodnotou za sebou. Jestliže je jejich součet větší než nastavená hodnota, jsou tyto hodnoty označeny za poruchy, viz obrázek 5.2.
- Proměnná *indexChyb* obsahuje informaci, kde na testovacích datech byla uměle vytvořena porucha, aby bylo možné na závěr vyhodnotit úspěšnost jednotlivých metod.

Po nastavení veškerých parametrů se zavolá funkce *system_(...)*, která ve formě struktury vrátí výsledky, které jsou uvedeny a graficky vyobrazeny v kapitole 6. Pro jejich vykreslení je možné použít připravenou funkci *system_vykresleni(...)*.

- 2) **system_.m** - Zde jsou na základě vstupních dat určených na učení modelu, hledání parametrů ke stanovení mezí a na testování zavolány v daném pořadí příslušné funkce, které zajišťují celý chod algoritmu detekce poruch sensorů, viz blokové schéma 5.1. Naučení zvolených modelů zajišťuje funkce *system_model(...)*, stanovení parametrů souvisejících s nastavením dolních a horních mezí obstarává *system_parametry(...)*, predikce je zajištěna pomocí funkce *system_predikce(...)* a na závěr vyhodnocení poruch pomocí *system_vyhodnoceniPoruch(...)*. Podrobnější informace ohledně implementace jednotlivých metod detekce poruch jsou uvedeny v kapitolách 5.2, 5.3 a 5.4. Výstupem této části ve formě struktury jsou informace o jednotlivých modelech, dále vektory predikovaných hodnot a jejich chyba oproti skutečnosti, vektory mezí a konečně i vyhodnocení výsledků poruch na testovacích datech.
- 3) **system_model.m** - Podle nastavení uživatelem v úvodu jsou zde volány příslušné funkce jednotlivých metod detekce poruch sensorů, viz blokové schéma 5.1, které zajišťují vytvoření jednotlivých modelů z dat k tomuto účelu určených. Samotným výstupem je struktura obsahující vytvořené modely (ARX, GMM, NARXnet).

- 4) **system_parametry.m** - V tomto procesu se na základě zadaného vstupního argumentu *procentoChyb* uvedeného výše hledají takové vnitřní parametry, aby byly nastaveny dolní a horní mez okolo predikované hodnoty podle požadavku.
- 5) **system_predikce.m** - Zde jsou prováděny predikce hodnot na základě vytvořených modelů, nastavení uživatele a získaných parametrů, detaily v kapitolách 5.2, 5.3 a 5.4. K predikovaným hodnotám je vypočítána chyba vzhledem ke skutečným hodnotám a také jsou zde dopočítány meze a o kolik byly případně překročeny. Součástí je i funkce *system_doucovaniModelu(...)*, která v případě potřeby provádí opětovné doučování na počátku vytvořených modelů.
- 6) **system_doucovaniModelu.m** - Samotná funkce doučování modelů definovaných na začátku pracuje na principu zavolání opětovného učení modelů, pokud data určená pro testování vybočí z původního rozsahu a zároveň u lokálních dat není detekována porucha. Například jestliže 20 testovaných hodnot překročilo maximum původního rozsahu a zároveň u těchto hodnot nebyla detekována porucha, budou přidány k datům určeným pro učení a bude provedeno doučení modelů. Výstupem je rozhodnutí, zda má dojít k doučení a jaká nová data k tomuto účelu mají být použita.
- 7) **system_vyhodnoceniPoruch.m** - Zde se provádí vyhodnocení, zda u lokálních dat došlo k poruše. Jestliže součet chyb překročení mezí za sebou přesáhne uživatelem zvolenou hodnotu *maxPorucha*, jsou tyto lokální hodnoty označeny za poruchu, viz obrázek 5.2. Díky informaci o uměle zavedených chybách do testovacích dat *indexChyb* a chybách zjištěných metodami detekce poruch je vypočítána jejich úspěšnost. Výsledky mají následující podobu: „true positive, true negative, false positive, false negative, specificita a senzitivita“, viz kapitola 6.



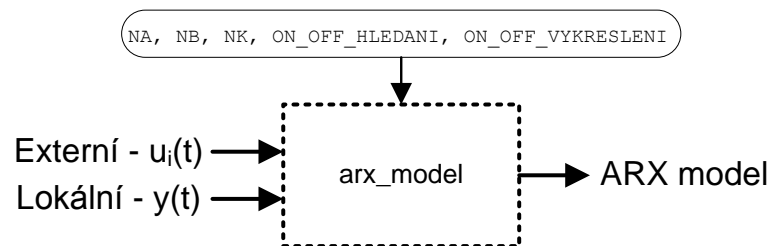
Obrázek 5.2: Principiální schéma vyhodnocení poruch.

Podstatná je skutečnost, že detekované poruchy, které jsou nahodilé či způsobené šumem, se vyznačují tím, že mají malou úroveň překročení a že se neopakují.

5.2 Implementace funkcí ARX

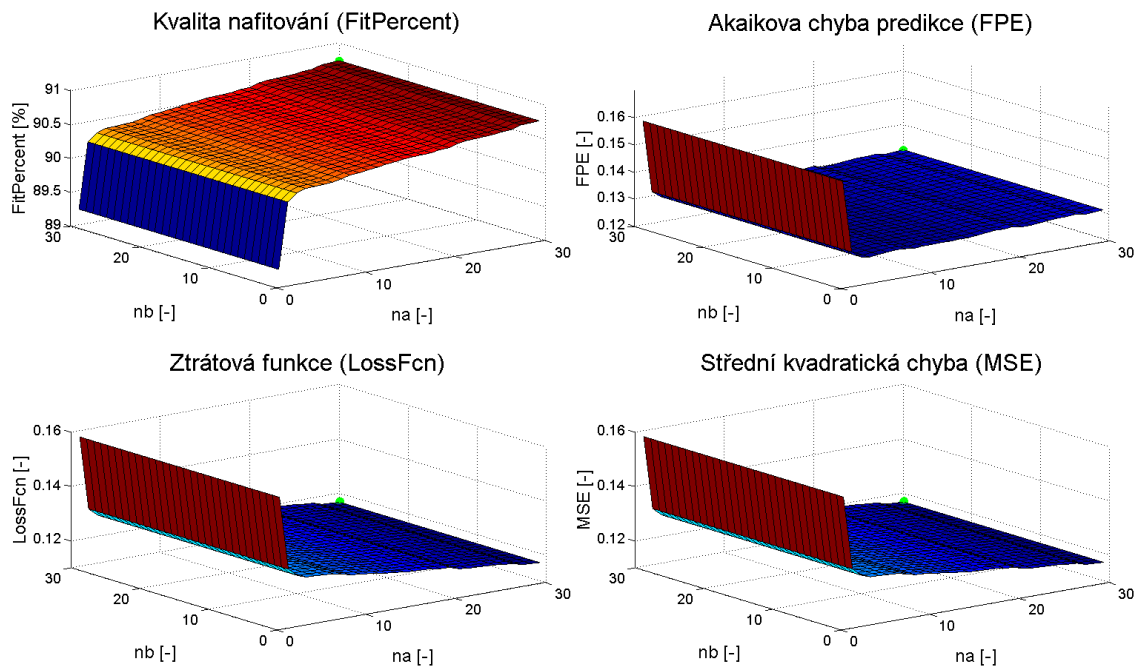
5.2.1 Učení modelu

Učení ARX modelu probíhá přibližně podle schématu uvedeného na obrázku 5.3. Hlavní funkcí je zde `arx_model(...)`, která na základě vstupních externích, lokálních dat a nastavení, které zvolil sám uživatel, vytvoří příslušný ARX model, viz kapitola 4.3.

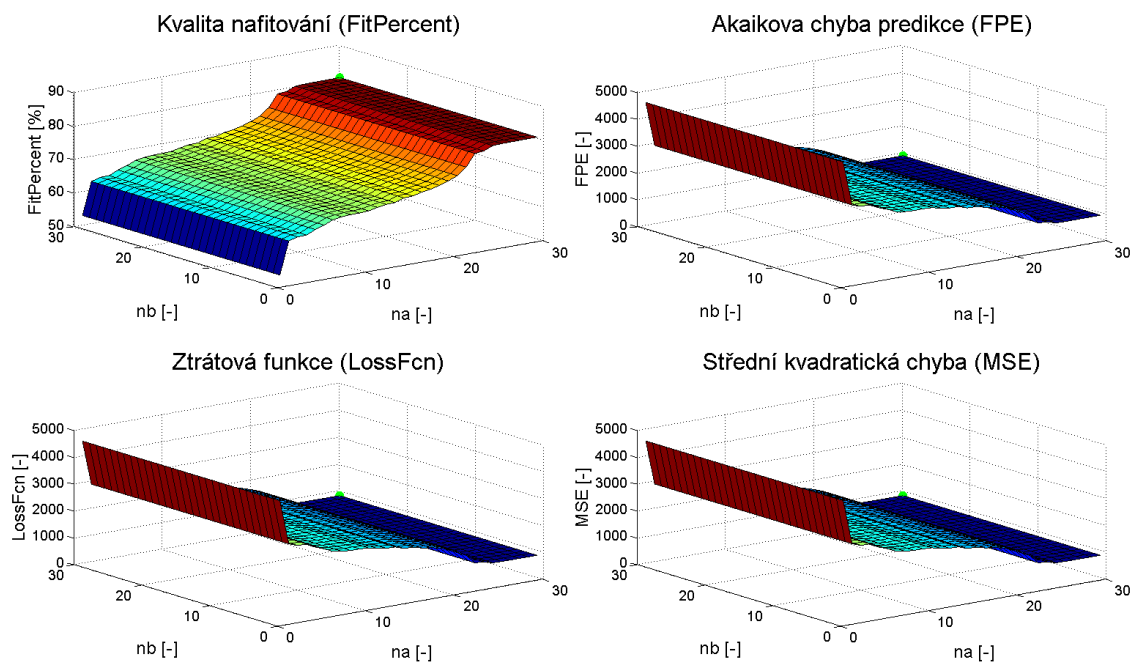


Obrázek 5.3: Schéma učení ARX modelu.

Vytvoření příslušného modelu je možné dvěma způsoby. První možností je nastavení parametru $ON_OFF_HLEDANI = 0$, což vytvoří ARX model na základě zadaných hodnot NA , NB a NK . Druhá volba $ON_OFF_HLEDANI = 1$ slouží k hledání nejlepšího modelu neboli nejmenší chyby modelu na učicích datech pomocí vyzkoušení kombinací hodnot parametrů NA a NB od jedné do jejich původně zvolené hodnoty. Výsledkem tohoto hledání je model s nejmenší chybou. Pomocí $ON_OFF_VYKRESLENI = 1$ je případně možné vykreslit závislosti kvality nařizování modelu nebo jeho chyby na parametrech NA a NB . Ukázky dvou příkladů jsou na obrázcích 5.4 a 5.5, více v grafické příloze A. Z obrázků je patrné, že záleží především na délce historie lokálních dat.



Obrázek 5.4: Ukázka hledání optimálního nastavení parametrů ARX modelu pro teplotní data.



Obrázek 5.5: Ukázka hledání optimálního nastavení parametrů ARX modelu pro data spotřeby energie „Coastal Family“.

5.2.2 Stanovení parametrů pro nastavení mezí

V případě ARX jsou prováděny dvě různé predikce, které jsou popsány v kapitole 5.2.3, a proto jsou zde vypočítány vnitřní parametry pro obě.

Základní nastavení mezí spočívá ve dvou krocích. První část je zajištěna pomocí funkce **arx_parametry_krok1(...)**, která na základě vstupních dat provádí predikci hodnot pro dvě varianty uspořádání, viz kapitola 5.2.3 a obrázek 5.6. Následně vypočítává maximální chybu rozdílu skutečné lokální hodnoty a *Predikce 1*, která v další části poslouží jako krajní hodnota při hledání mezí u první varianty, aby byl splněn parametr *procentoChyb* popsany v 5.1. Dále je zde vypočítán parametr *DIFERENCE*, který obsahuje minimální a maximální rozdíl lokální hodnoty a *Predikce 2*. Tato hodnota je určující pro stanovení mezí u druhé varianty.

Druhá část **arx_parametry_krok2(...)**, která slouží pouze první variantě uspořádání, spočívá v nalezení parametru *MEZ* tak, aby byl splněn požadavek na množství povolených chyb ve vstupních datech *procentoChyb*. *MEZ* slouží ke stanovení dolní a horní meze *Predikce 1*.

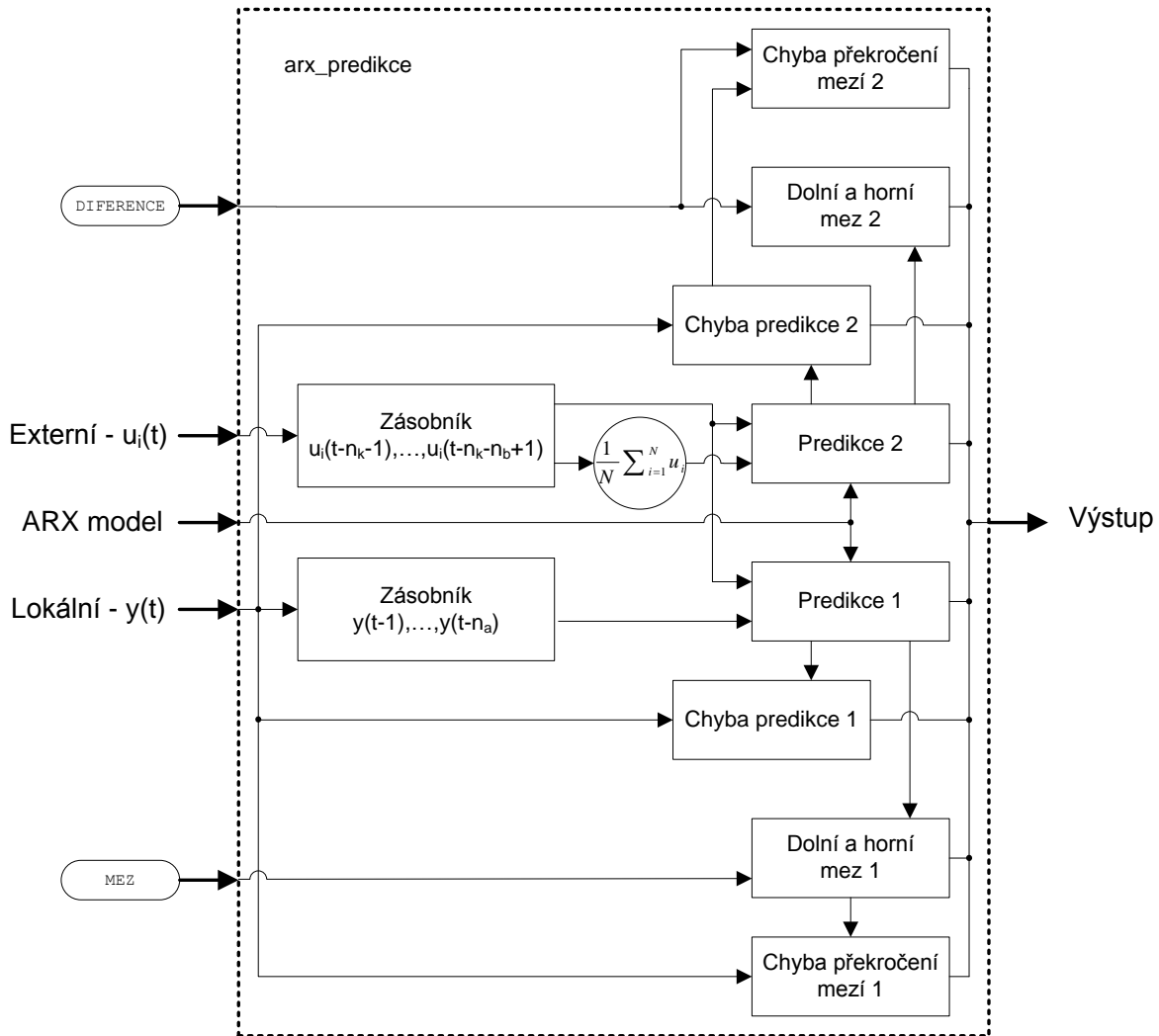
5.2.3 Predikce

Celá funkce **arx_predikce(...)** předpovídání hodnot, určení dolní a horní meze nebo chyby predikce či překročení mezí je zobrazena na blokovém schématu 5.6.

Jak jsem se již zmínil v kapitole 5.2.2 a je patrné z obrázku 5.6, jsou zde prováděny dvě různé predikce hodnot (*Predikce 1* a *Predikce 2*).

První z nich, *Predikce 1*, je standardní a jako vstupy používá minulá externí a lokální data přesně podle naučeného ARX modelu. Dále je vypočítána chyba této predikce vůči skutečné lokální hodnotě. Následně pomocí parametru *MEZ*, který byl vypočítán v kapitole 5.2.2, jsou stanoveny dolní a horní mez související s *Predikcí 1*. Na závěr je ještě stanoveno, zda lokální hodnota vybočuje z intervalu těchto mezí a případně o kolik.

Predikce 2 stojí na myšlence, že naučený ARX model obsahuje vztah mezi externími a lokálními hodnotami. Provedená predikce pouze na základě externích dat se jim blíží, ale není stejná. Samotná druhá varianta je postavena na sledování difference mezi *Predikcí 2* a lokální hodnotou. Následně je vypočítán rozdíl lokálních dat a této predikce (*Chyba predikce 2*), která společně s parametrem *DIFERENCE* z kapitoly 5.2.2 poslouží k vyhodnocení, zda došlo k překročení mezí.

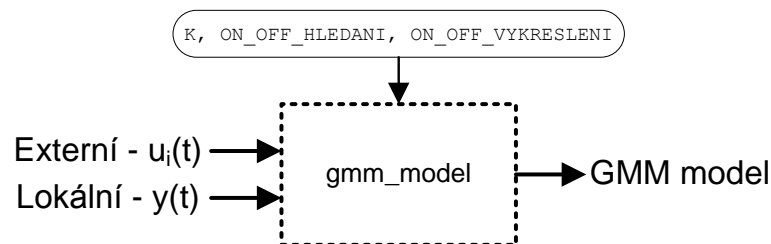


Obrázek 5.6: Schéma predikce a stanovení mezí ARX.

5.3 Implementace funkcí GMM

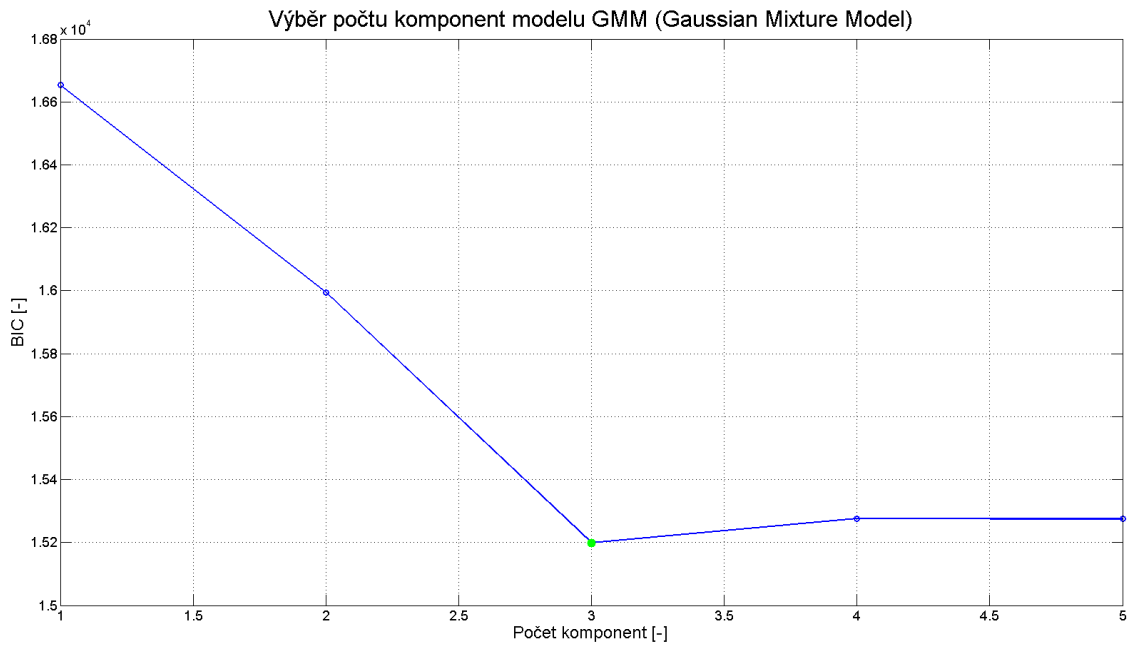
5.3.1 Učení modelu

Na obrázku 5.3 vidíme schéma, podle kterého dochází k naučení GMM modelu. Vše je postaveno na funkci `gmm_model(...)`, která na základě vstupních externích, lokálních dat a nastavení, které zvolil sám uživatel, vytvoří příslušný GMM model, viz kapitola 4.4. Stejně jako u implementace modelu ARX zde máme možnost nastavení parametru `ON_OFF_HLEDANI`. Jestliže je roven nule, dojde k vytvoření modelu na základě zadaného počtu komponent K , ale pokud je roven jedné, je spuštěno hledání optimálního počtu komponent podle BIC kritéria, viz 4.4.2, který je následně použit.

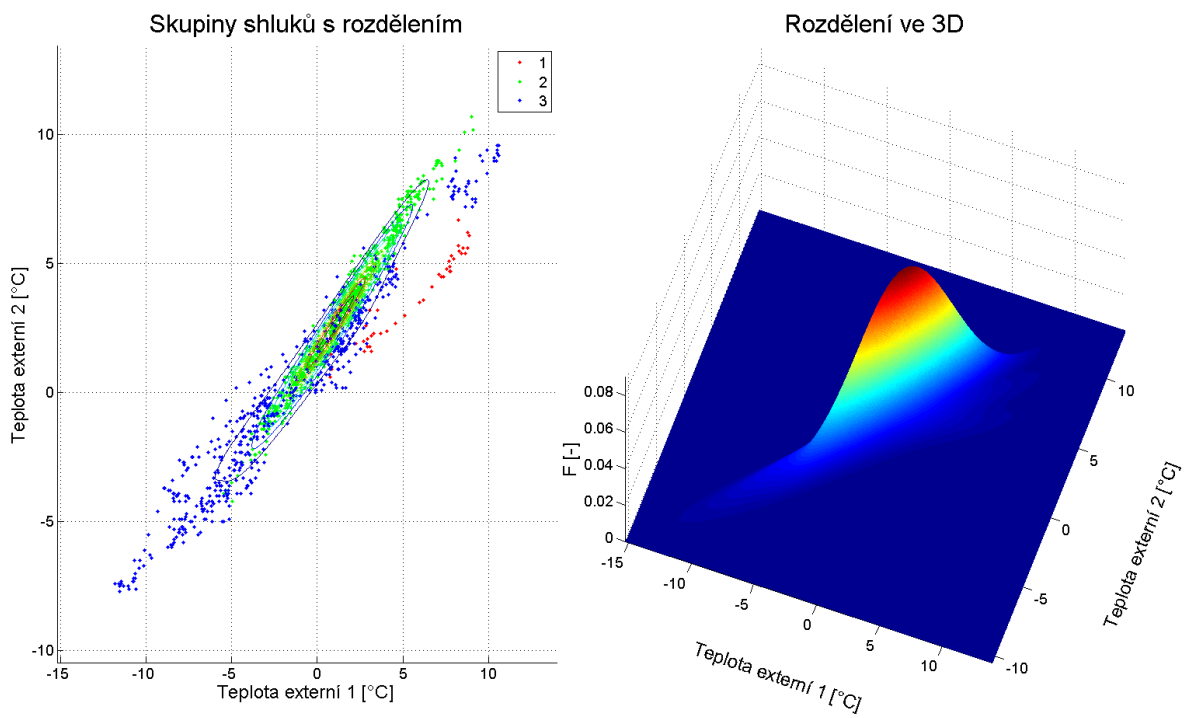


Obrázek 5.7: Schéma učení GMM modelu.

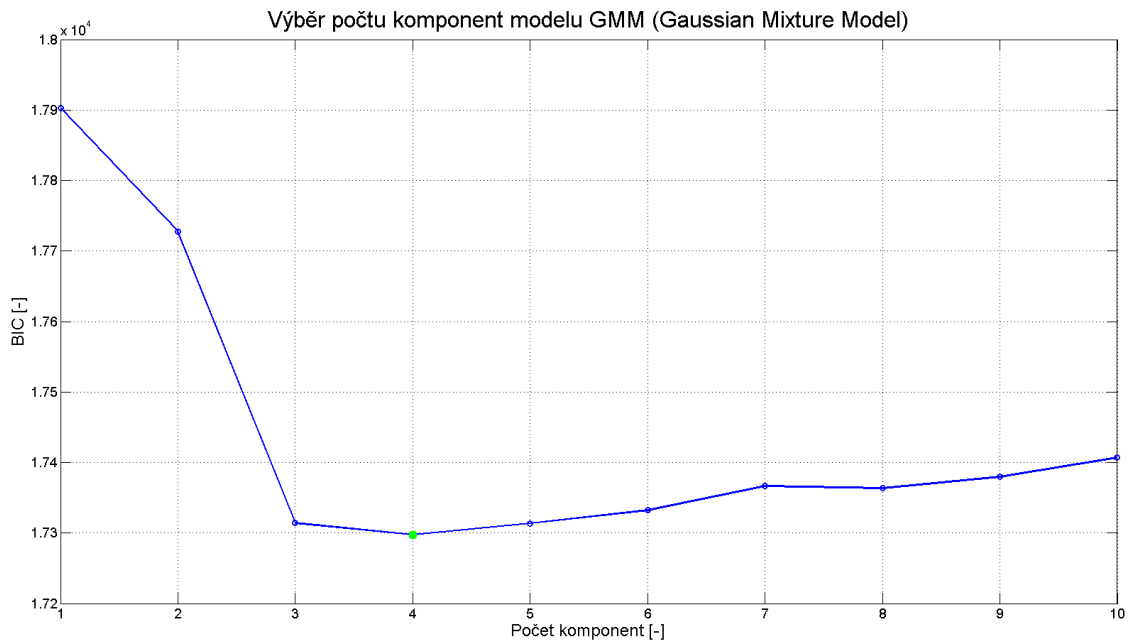
Zobrazení závislost BIC kritéria na počtu komponent společně s vykreslením rozdělení je možné provést nastavením parametru `ON_OFF_VYKRESLENI = 1`, viz ukázka na obrázcích 5.8, 5.9 pro teplotní data a také na obrázcích 5.10, 5.11 pro jedna data o spotřebě energie, ostatní jsou zobrazeny v příloze A. Z obrázků je patrné, že v případě teplotních dat se jedná o přibližně lineární závislost s přiměřeným rozptylem dat, ale jsou zde patrné i odlehlejší hodnoty, které mírně ovlivní stanovení konfidenčního intervalu. Naopak co se týká dat spotřeby energie, je vidět, že zde už nelze tvrdit, že se jedná o lineární závislost. Dále je vidět, že rozdělení má velký rozptyl, což se velice projeví na šířce intervalu spolehlivosti.



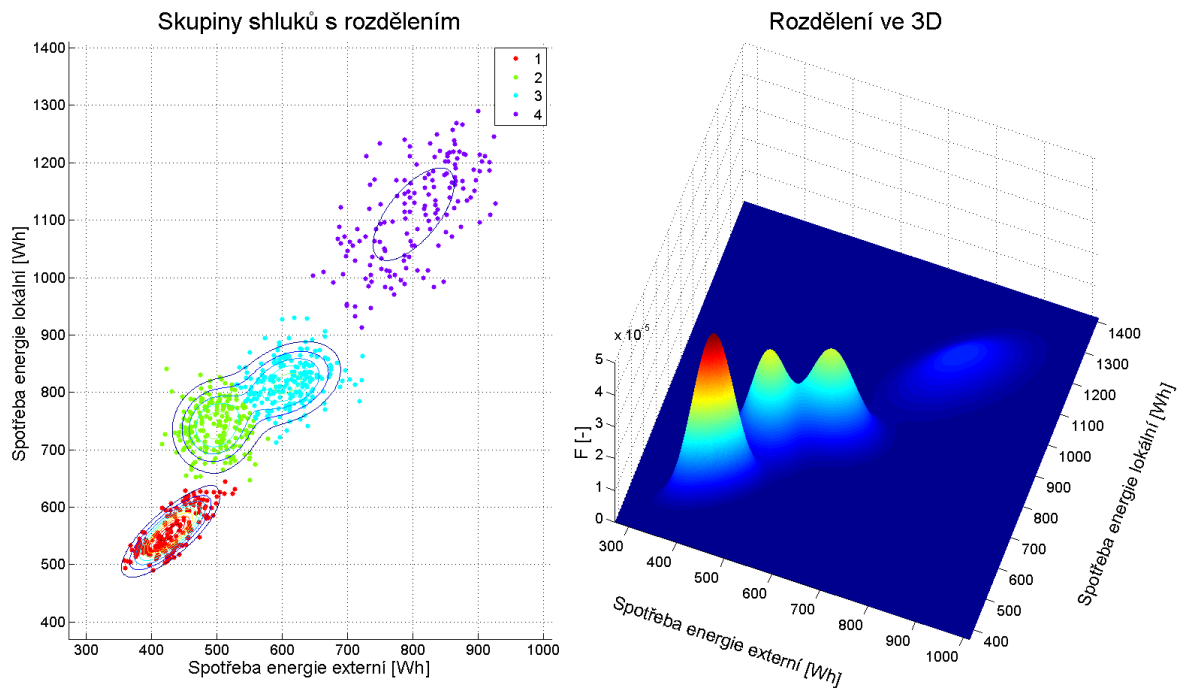
Obrázek 5.8: Ukázka hledání optimálního počtu komponent GMM modelu pro teplotní data.



Obrázek 5.9: Ukázka rozdělení teplotních dat GMM modelu.



Obrázek 5.10: Ukázka hledání optimálního počtu komponent GMM modelu pro data spotřeby energie „Coastal Family“.



Obrázek 5.11: Ukázka rozdělení dat spotřeby energie „Coastal Family“ GMM modelu.

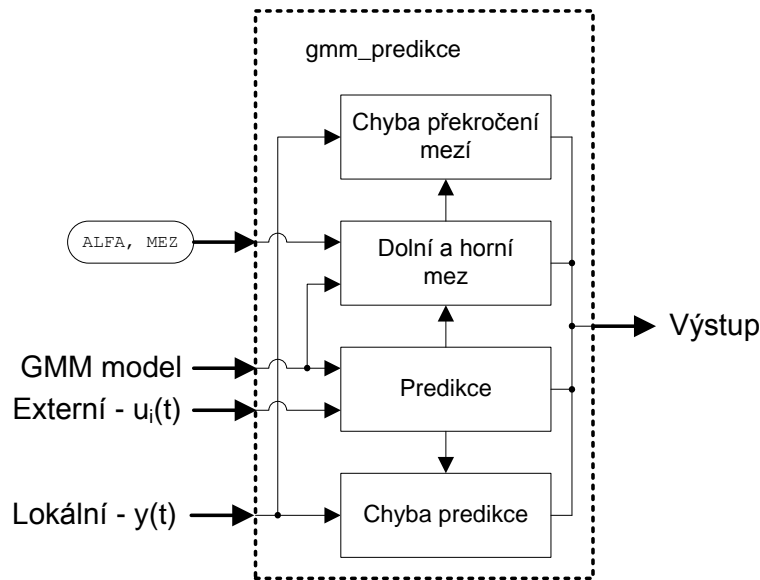
5.3.2 Stanovení parametrů pro nastavení mezí

Výpočet parametrů probíhá ve dvou krocích. V prvním `gmm_parametry_krok1(...)` jsou z externích dat predikovány lokální hodnoty, viz kapitola 5.3.3, obrázek 5.12 a teorie 4.4.3. Následně je vypočítána maximální chyba rozdílu skutečné lokální hodnoty a *Predikce*, která ve druhém kroku poslouží jako krajní hodnota při stanovování mezí.

Druhý krok `gmm_parametry_krok2(...)` spočívá v nalezení parametru *MEZ* tak, aby byl splněn požadavek na množství povolených chyb ve vstupních datech *procentoChyb*. Rozdílem oproti stanovení dolní a horní meze u ARX, kde je výpočet založen pouze na parametru *MEZ*, je v tom, že zde navíc používáme parametr *ALFA* (hladinu významnosti), který je popsán v kapitole 4.4.4.

5.3.3 Predikce

Předpovídání hodnot, stanovení dolní a horní meze, výpočet chyby predikce a chyby překročení mezí zajišťuje funkce `gmm_predikce(...)` a vychází ze zjednodušeného blokového schématu 5.12.



Obrázek 5.12: Schéma predikce GMM.

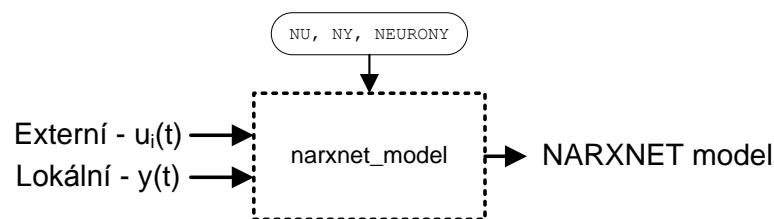
Základem samotné predikce lokálních hodnot je podmíněná střední hodnota, viz 4.4, díky které je možné provádět predikci pouze na základě znalosti modelu GMM a externích dat. Následně je vypočítána chyba predikce jako rozdíl skutečné lokální a od-

hadnuté hodnoty. Dolní a horní mez jsou stanoveny pomocí zvoleného *ALFA*, viz 4.4.4, parametru *MEZ* z kapitoly 5.3.2 a vypočítané predikce. Na závěr je ještě vyhodnoceno, zda a případně o kolik došlo k překročení mezí skutečnou lokální hodnotou.

5.4 Implementace funkcí NARXnet

5.4.1 Učení modelu

Učení NARXnet modelu probíhá přibližně podle schématu uvedeného na obrázku 5.13. Hlavní funkcí je zde `narxnet_model(...)`, která na základě vstupních externích, lokálních dat a nastavení (*NU*, *NY*, *NEURONY*), které zvolil sám uživatel, vytvoří pomocí trénování příslušný NARXnet model, viz kapitola 4.5.



Obrázek 5.13: Schéma učení NARXnet modelu.

5.4.2 Stanovení parametrů pro nastavení mezí

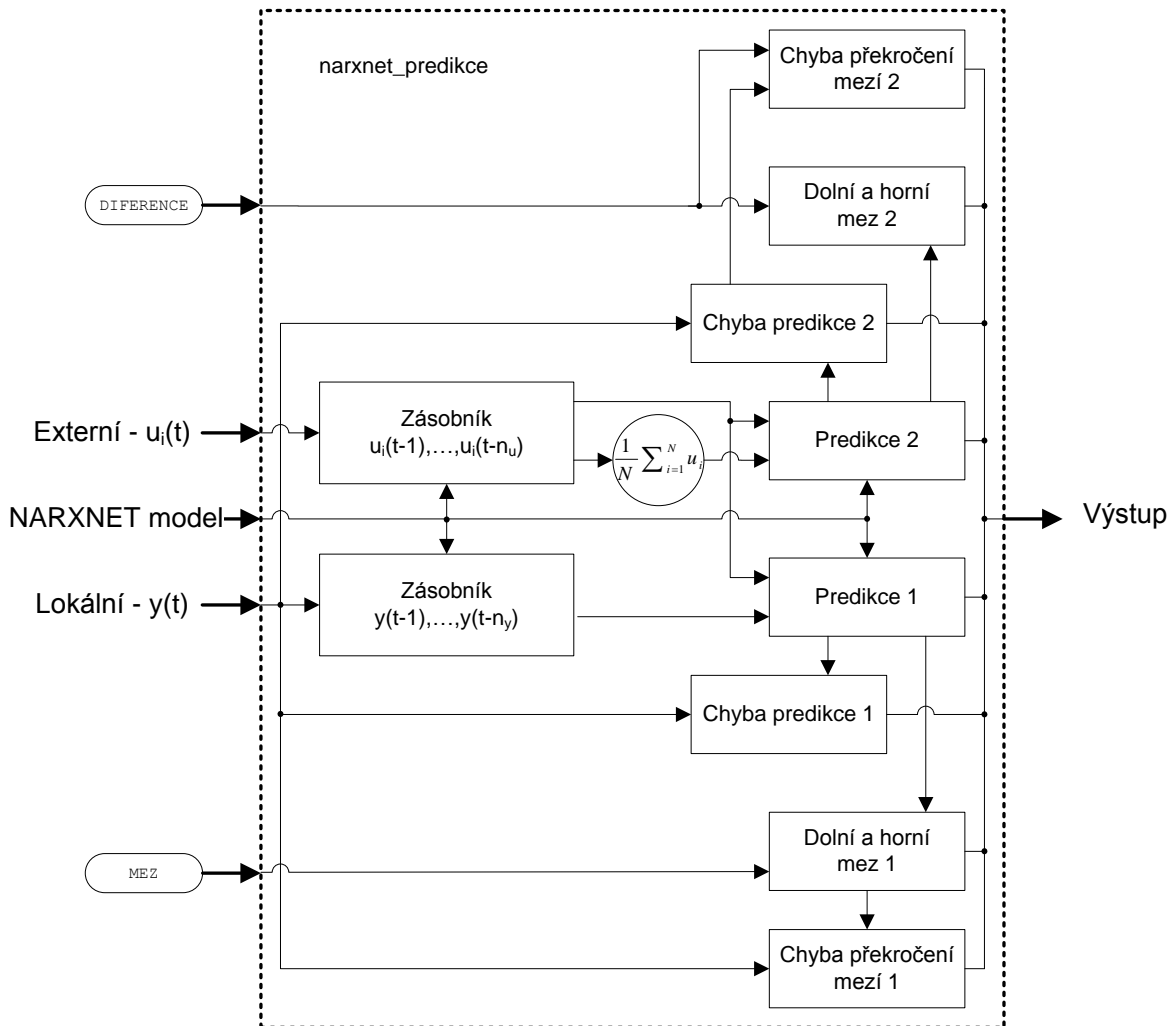
V případě NARXnet jsou opět prováděny dvě různé predikce, které jsou popsány v kapitole 5.4.3, a proto jsou zde vypočítány vnitřní parametry pro obě jako u ARX.

Stanovení parametrů pro nastavení mezí spočívá ve dvou krocích. První část je zajištěna pomocí funkce `narxnet_parametry_krok1(...)`, která na základě vstupních dat provádí predikci hodnot pro dvě varianty uspořádání, viz kapitola 5.4.3 a obrázek 5.14. Dále vypočítá maximální chybu rozdílu skutečné lokální hodnoty a *Predikce 1*, která ve druhém kroku poslouží jako krajní hodnota při hledání mezí u první varianty uspořádání. Také je zde vypočítán parametr *DIFERENCE*, který obsahuje minimální a maximální rozdíl lokální hodnoty a *Predikce 2*. Tato hodnota je určující pro stanovení mezí u druhé varianty uspořádání.

Druhá část `narxnet_parametry_krok2(...)`, která slouží pouze první variantě uspořádání, spočívá v nalezení parametru *MEZ* tak, aby byl splněn požadavek na množství povolených chyb ve vstupních datech *procentoChyb*. *MEZ* slouží ke stanovení dolní a horní meze *Predikce 1*.

5.4.3 Predikce

Předpovídání hodnot, stanovení dolní a horní meze, výpočet chyby predikce nebo překročení mezí má na starosti funkce `narxnet_predikce(...)`, jejíž blokové schéma je na obrázku 5.14. Podrobnější informace o předpovídání hodnot jsou uvedeny v kapitole 4.5.2.



Obrázek 5.14: Schéma predikce NARXnet.

Z obrázku 5.6 je patrné, že jsou zde prováděny dvě různé predikce hodnot (*Predikce 1* a *Predikce 2*).

První z nich, *Predikce 1*, je standardní a jako vstupy používá externí a lokální data přesně podle naučeného NARXnet modelu. Dále je vypočítána chyba této predikce vůči skutečné lokální hodnotě. Následně pomocí parametru *MEZ*, který byl vypočítán v kapitole 5.4.2, jsou stanoveny dolní a horní mez související s *Predikcí 1*. Na závěr je ještě stanoveno, zda lokální hodnota vybočuje z intervalu těchto mezí a případně o kolik.

Predikci 2 lze použít hlavně pro předpovídání velice podobných dat, například hodnot z téměř identicky měřících redundantních senzorů, v mém případě teplot. Tato varianta je postavena na sledování difference mezi *Predikcí 2* a lokální hodnotou. Následně je vypočítán rozdíl lokálních dat a této predikce (*Chyba predikce 2*), která společně s parametrem *DIFERENCE* z kapitoly 5.4.2 poslouží k vyhodnocení, zda došlo k překročení mezí.

5.4.4 Stav

Metodu detekce poruch pomocí neuronové sítě NARXnet (Nonlinear AutoRegressive neural network with eXternal input) se podařilo implementovat, ale v rámci pokračování či budoucí práce by bylo vhodné provést odladění trénování neuronové sítě či naprogramovat optimální hledání parametrů pro naučení modelu.

Kapitola 6

Experimenty a zhodnocení výsledků

Tato kapitola popisuje výsledky provedených experimentů v oblasti detekce poruch senzorů pomocí několika metod, které byly popsány v kapitole 4. Výsledky jsou prezentovány pomocí tabulek a obrázků pro použitá testovací data. V této kapitole jsou blíže popsána pouze teplotní data a dvojice data spotřeby energie, protože zbývající data spotřeby jim jsou velice podobná. Tabulky a část grafů pro popisovaná data jsou uvedeny zde, zbývající grafy jsou uloženy v elektronické podobě na přiloženém CD. Výsledky pro zbývající dvojici dat jsou uvedeny v příloze B a příslušné grafy na CD.

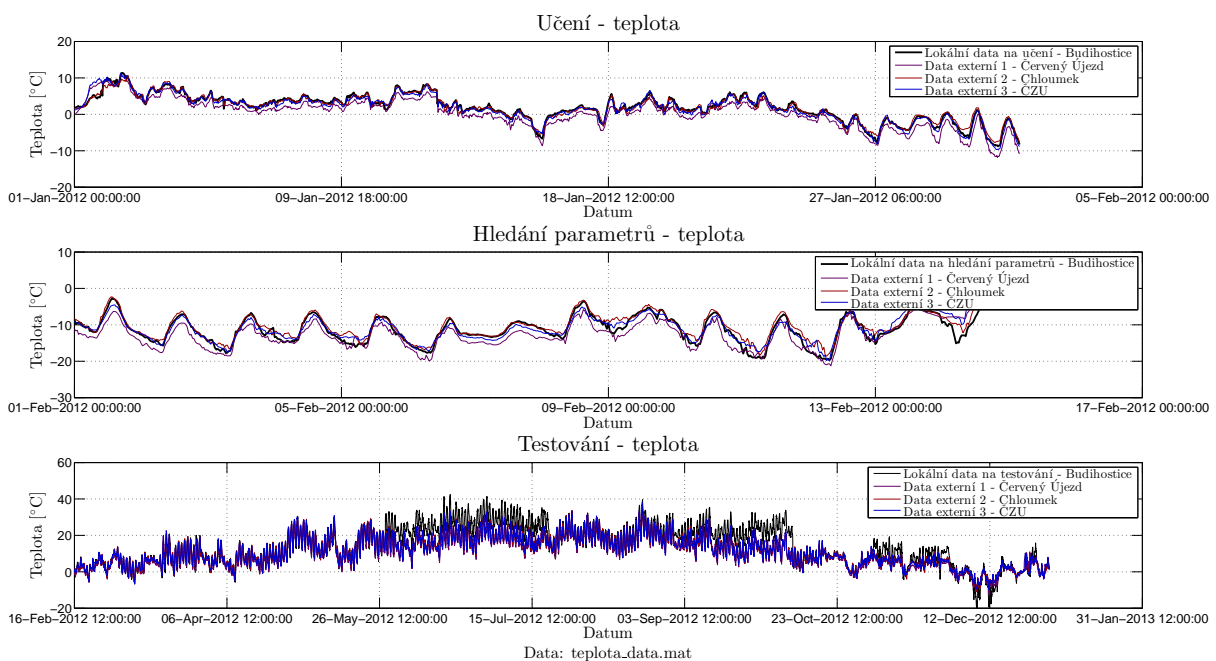
V první části se zaměřím na popis problematiky dat určených na učení modelů, hledání parametrů a testování. Dále představím výsledky jednotlivých metod a na závěr zhodnotím celkové dosažené výsledky.

6.1 Problematika učení modelů, hledání parametrů, testování

Pro samotný popis problematiky učení modelů, hledání parametrů a testování jsem zvolil teplotní data z meteorologických stanic *Budihostice*, *Červený Újezd*, *Chloumek*, *ČZU* a dále dvojice data spotřeby energie *Coastal Multi Family*, *Coastal Single Family* a *Desert Multi Family*, *Desert Single Family*. Ukázky průběhů bez zavedené chyby a s ní jsou uvedeny v grafické příloze A.

Teplotní data *Budihostice, Červený Újezd, Chloumek, ČZU*

Z obrázku 6.1 je patrné, že první problém při učení modelů by mohla způsobit teplota z meteorologické stanice *Červený Újezd*, jelikož tato data jsou mírně posunutá od ostatních. Ovšem hlavním problémem je skutečnost, že jednotlivé modely jsou učený na datech v rozsahu přibližně -10 až $+10$ °C, zatímco testování probíhá v rozsahu -20 až $+40$ °C, což může výrazně ovlivnit výsledky, přestože je implementováno doučování modelů v průběhu, jelikož k doučování dojde pouze pokud nebyla detekována porucha.



Obrázek 6.1: Ukázka teplotních dat - *Budihostice, Červený Újezd,...*

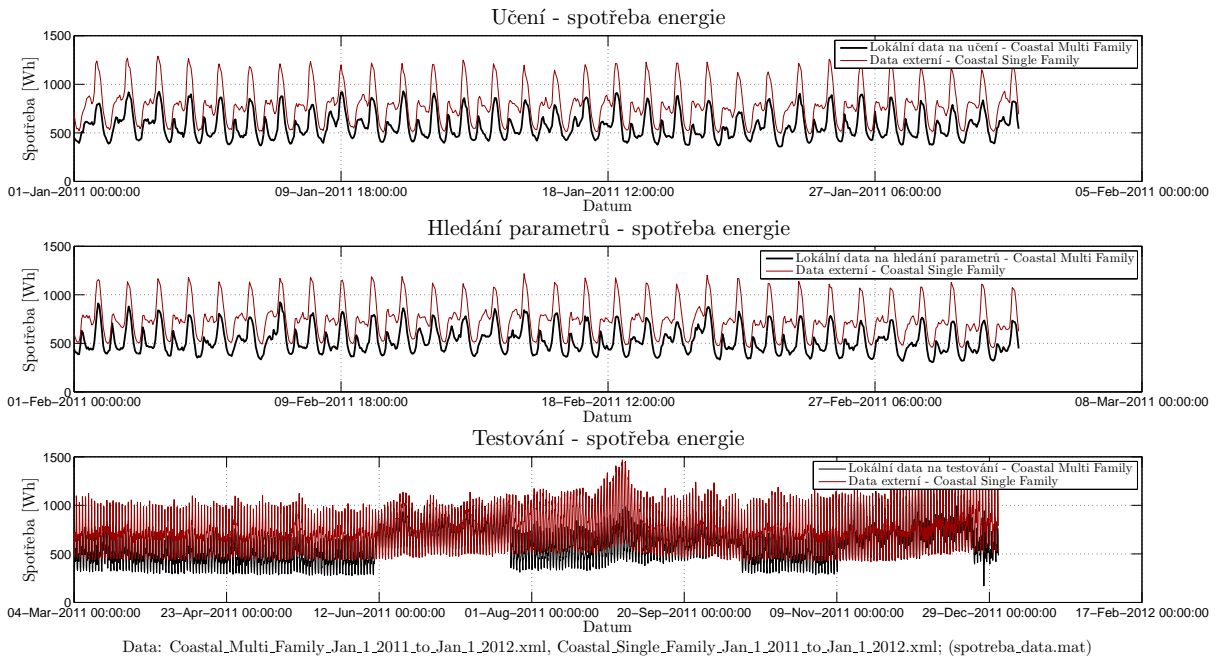
Spotřeba energie *Coastal Multi Family, Coastal Single Family*

Podle průběhů na obrázku 6.2 lze předpokládat, že na těchto datech by neměl být problém detekovat poruchy senzorů, jelikož se zdá, že se data chovají stále přibližně stejně.

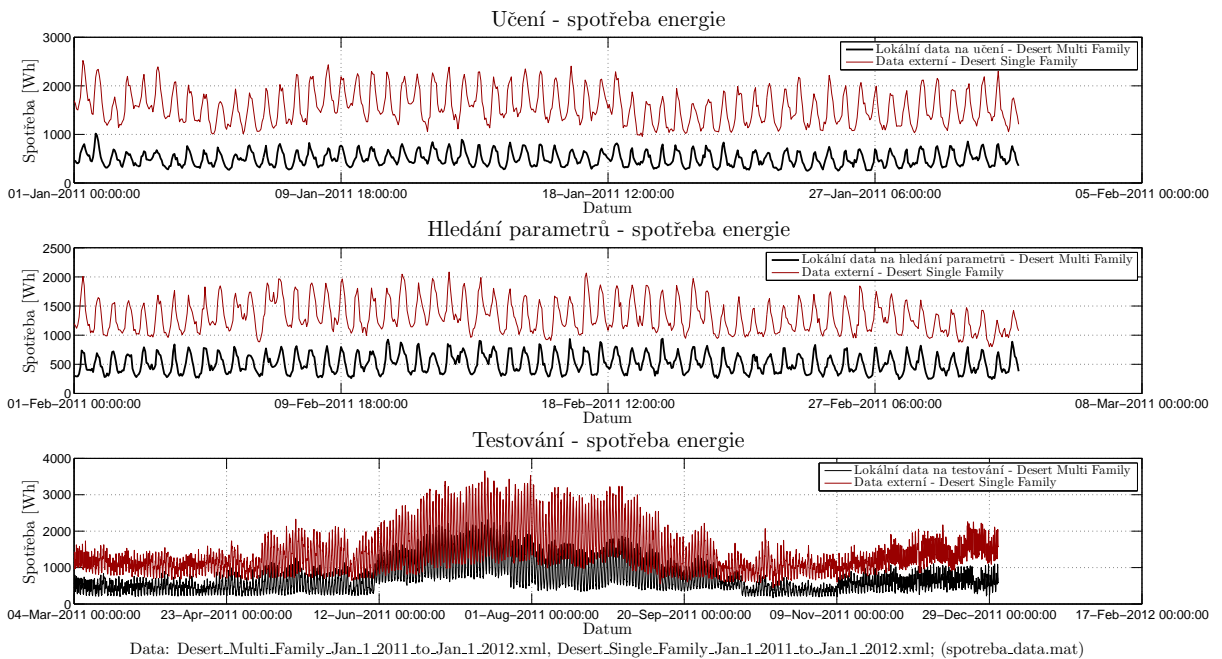
Spotřeba energie *Desert Multi Family, Desert Single Family*

Tato data, obrázek 6.3, jsou velice problematická, jelikož hodnoty určené na učení modelů velice málo odpovídají hodnotám testovacím. Velký problém se ukrývá v měnící se diferencii, rozsahu a rozptylu lokálních a externích dat. Toto jsou jedny z důvodů, proč implementované metody detekce poruch senzorů jsou na těchto datech neúčinné.

6.1. PROBLEMATIKA UČENÍ MODELŮ, HLEDÁNÍ PARAMETRŮ, TESTOVÁNÍ⁵⁷



Obrázek 6.2: Ukázka dat určených na učení modelů, hledání parametrů a testování - spotřeba energie *Coastal Multi* a *Single Family*.



Obrázek 6.3: Ukázka dat určených na učení modelů, hledání parametrů a testování - spotřeba energie *Desert Multi* a *Single Family*.

6.2 Výsledky

Výsledky jsou uvedeny pomocí tabulek u jednotlivých metod detekce poruch senzorů. Na základě dat určených pro testování, vstupních parametrů *procentoChyb* a *maxPorucha*, viz kapitola 5, byla provedena detekce poruch pomocí jednotlivých metod na přítomnost chyb v lokálních datech. Jejich výstupem bylo vyhodnocení *true positive (TP)*, *true negative (TN)*, *false positive (FP)*, *false negative (FN)*, *specificita* (true negative rate = $TN/(FP+TN)$) a *senzitivita* (true positive rate = $TP/(TP+FN)$), které popisují jejich úspěšnost detekce chyb.

Výsledky se v některých případech mohou zdát špatné, ale to může být způsobeno tím, že do diagnostikovaných dat byla zavedena malá chyba nebo se tato chyba projeví postupem času, až vzroste. Příkladem může být drift, u kterého trvá déle, než se chyba projeví, ale parametr *indexChyb*, že byla chyba zavedena, je již od počátku nastaven na jedna (nula znamená lokální hodnotu bez zavedené chyby).

Vysvětlení a odůvodnění jednotlivých výsledků je uvedeno vždy konkrétně u každé metody, včetně popisu co je vidět na obrázcích či proč zvolená metoda nefungovala správně.

6.2.1 ARX (AutoRegressive model with eXternal input)

uspořádání číslo 1

Tato metoda detekce poruch se podle výsledků uvedených v tabulkách 6.1, 6.2, 6.3, B.1 a B.2 řadí mezi méně úspěšné. Její velkou nevýhodou je fakt, že se dokáže velice dobře přizpůsobit lokálním hodnotám, které obsahují poruchy, což způsobí, že nejsou detekovány. Samotným důvodem je naučený model, který při následné predikci lokální hodnoty dává velkou váhu minulé lokální hodnotě a menší váhu ostatním externím či lokálním hodnotám.

Na obrázcích 6.4, 6.9 a 6.14 vidíme náhled na testovaná data, stanovené meze a vyhodnocení, zda byla zavedená chyba detekována.

Obrázky 6.5, 6.10 a 6.15 ukazují jakým způsobem je detekována větší odlehlá hodnota (občasná porucha). Za povšimnutí stojí způsob, jakým se daná chyba projeví v průběhu překročení mezí. Je zde patrné vychýlení jedním směrem a následně směrem druhým, což je právě způsobeno tím, že tento model dává při predikci velkou váhu na minulou lokální hodnotu. Dále je zde vidět i několik falešných poplachů.

Na obrázcích 6.6 a 6.11 je zobrazen projev skokové poruchy. Je vidět, že detekce této chyby je velmi malá, protože se predikce přizpůsobila lokálním datům, které tuto poruchu

obsahují. Dále je zde také patrné, jakou chybu překročení mezí vyvolá právě skok, na začátku je indikován větším překročením mezí nahoru a konec skoku překročením dolů. Na obrázku 6.16 je detekce ovlivněna samotným průběhem dat.

Projev driftu (pozdvolné poruchy) je zobrazen na obrázcích 6.7, 6.12 a 6.17, opět zde vidíme, jak se predikce přizpůsobí, a proto nedochází k detekci chyby.

Detekce změny zesílení (gain) senzoru je uvedena na obrázcích 6.8, 6.13 a 6.18. Znovu tady vidíme stejný problém s přizpůsobováním. Změna zesílení je detekována pouze v případě velké změny trendu lokálních dat, a to ze stoupajícího na klesající a obráceně.

Tabulka 6.1: Výsledky ARX uspořádání číslo 1 - teploty *Budihostice, Červený Újezd, Chloumek, ČZU*.

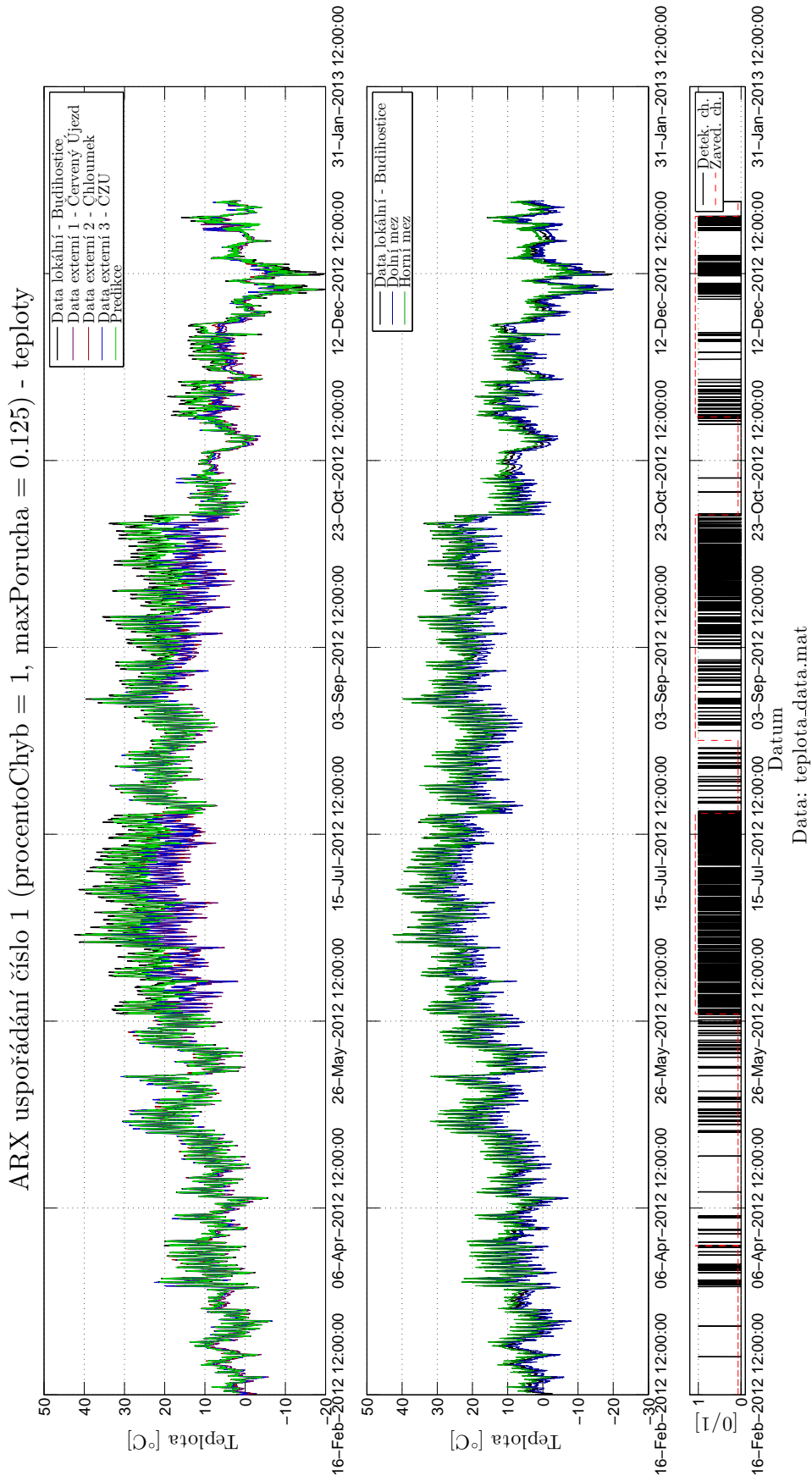
Procento chyb [%]	Max porucha [°C]	TP [-]	FP [-]	FN [-]	TN [-]	Specificita [%]	Senzitivita [%]
0,5	0,125	1135	59	6914	7228	99,19	14,10
	0,500	870	33	7179	7254	99,55	10,81
	1,500	503	11	7546	7276	99,85	6,25
1,0	0,125	1803	131	6246	7156	98,20	22,40
	0,500	1518	68	6531	7219	99,07	18,86
	1,500	1070	19	6979	7268	99,74	13,29
1,5	0,125	1937	218	6112	7069	97,01	24,07
	0,500	1647	109	6402	7178	98,50	20,46
	1,500	1114	28	6935	7259	99,62	13,84

Tabulka 6.2: Výsledky ARX uspořádání číslo 1 - spotřeba energie *Coastal*
Multi a Single Family.

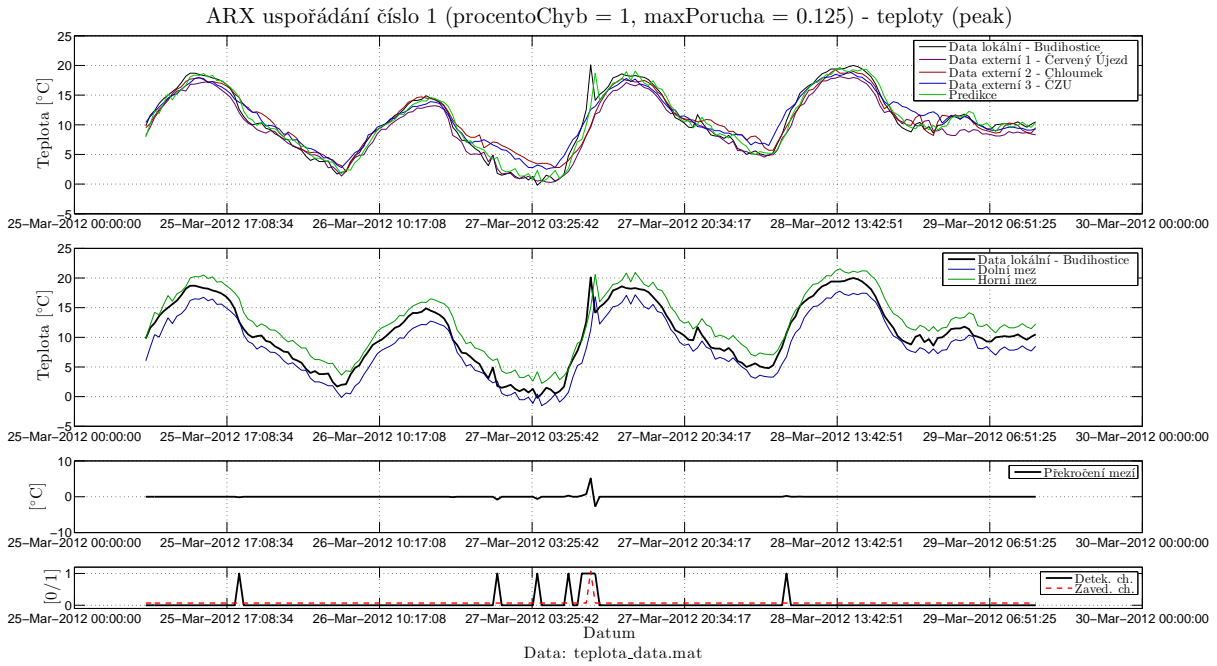
Procento chyb [%]	Max porucha [Wh]	TP [-]	FP [-]	FN [-]	TN [-]	Specifická [%]	Senzitivita [%]
0,5	12,5	120	27	3225	3900	99,31	3,59
	50,0	53	15	3292	3912	99,62	1,58
	150,0	9	8	3336	3919	99,80	0,27
1,0	12,5	159	33	3186	3894	99,16	4,75
	50,0	73	16	3272	3911	99,59	2,18
	150,0	21	8	3324	3919	99,80	0,63
1,5	12,5	169	37	3176	3890	99,06	5,05
	50,0	79	16	3266	3911	99,59	2,36
	150,0	19	9	3326	3918	99,77	0,57

Tabulka 6.3: Výsledky ARX uspořádání číslo 1 - spotřeba energie *Desert*
Multi a Single Family.

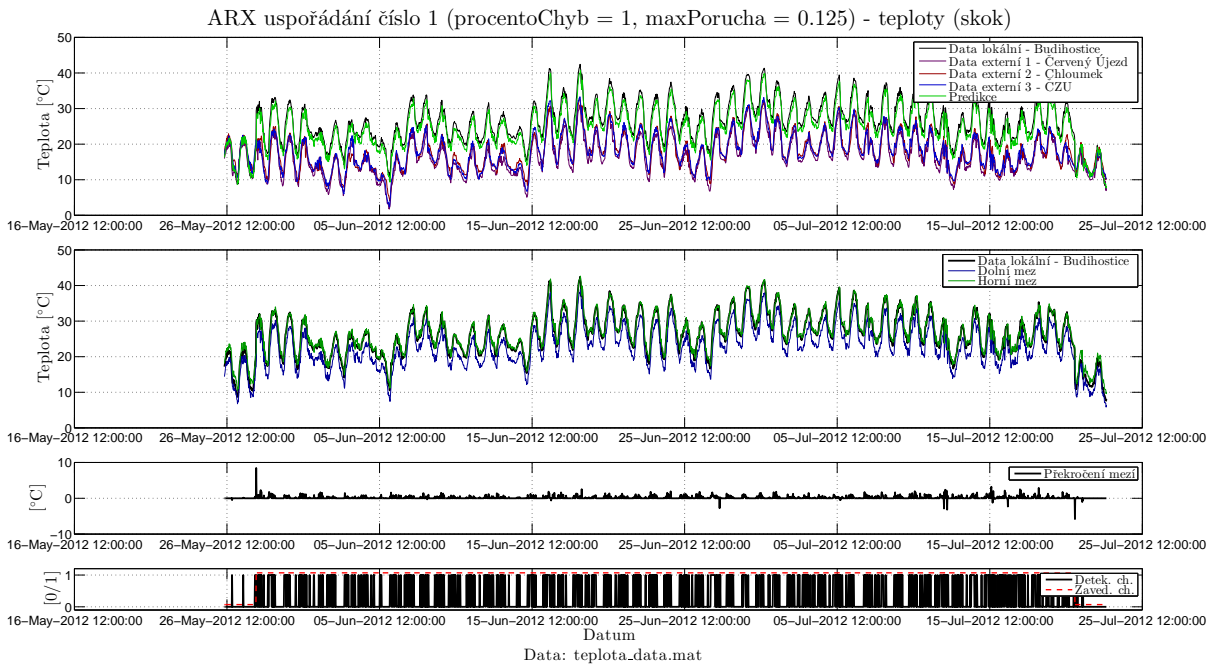
Procento chyb [%]	Max porucha [Wh]	TP [-]	FP [-]	FN [-]	TN [-]	Specifická [%]	Senzitivita [%]
0,5	12,5	425	177	2920	3750	95,49	12,71
	50,0	335	114	3010	3813	97,10	10,01
	150,0	204	59	3141	3868	98,50	6,10
1,0	12,5	377	127	2968	3800	96,77	11,27
	50,0	293	83	3052	3844	97,89	8,76
	150,0	192	38	3153	3889	99,03	5,74
1,5	12,5	381	141	2964	3786	96,41	11,39
	50,0	303	91	3042	3836	97,68	9,06
	150,0	193	48	3152	3879	98,78	5,77



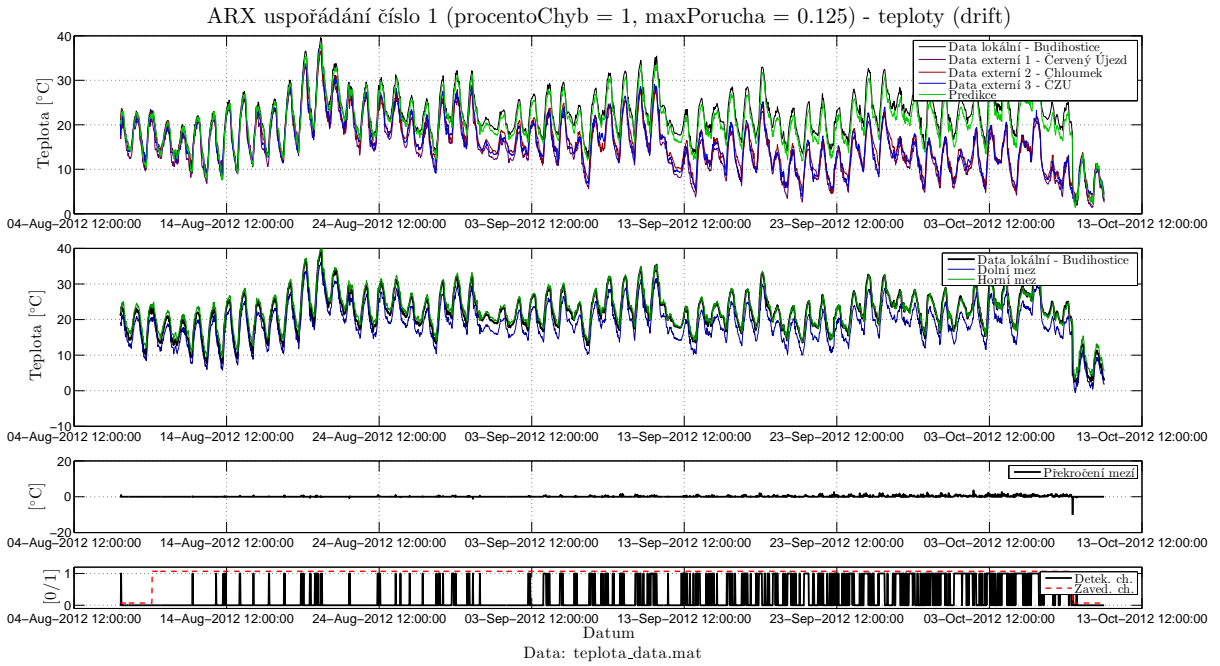
Obrázek 6.4: Výsledky ARX uspořádání číslo 1 - teploty Budihostice, Červený Újezd, Chloumek, ČZU.



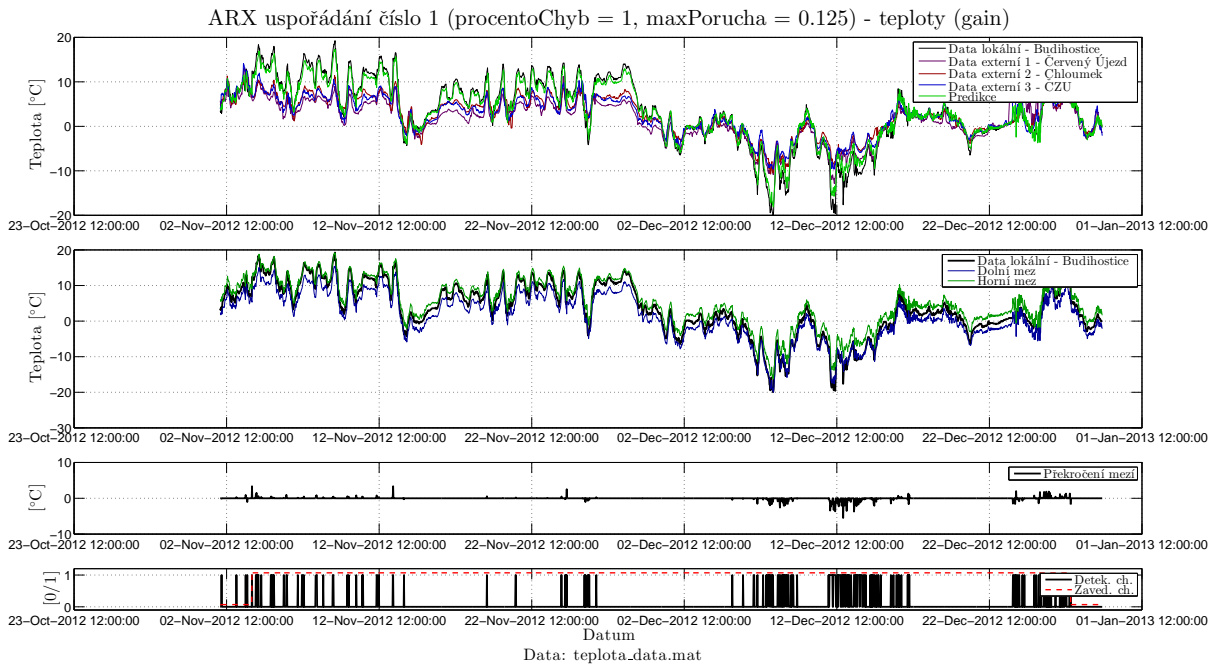
Obrázek 6.5: Detail výsledků ARX uspořádání číslo 1 - teploty *Budihostice*, *Červený Újezd*, *Chloumek*, *ČZU* (peak).



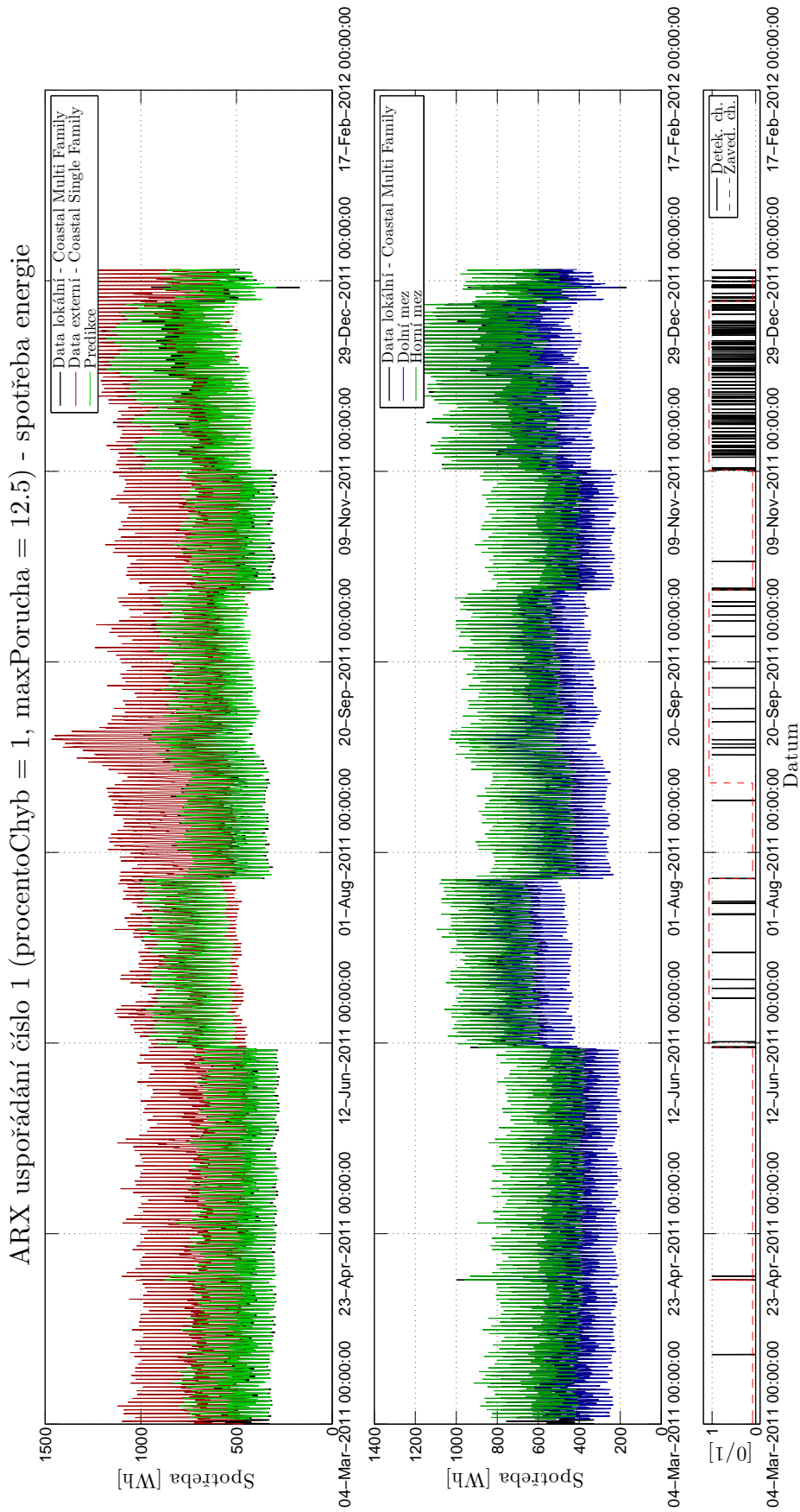
Obrázek 6.6: Detail výsledků ARX uspořádání číslo 1 - teploty *Budihostice*, *Červený Újezd*, *Chloumek*, *ČZU* (skok).

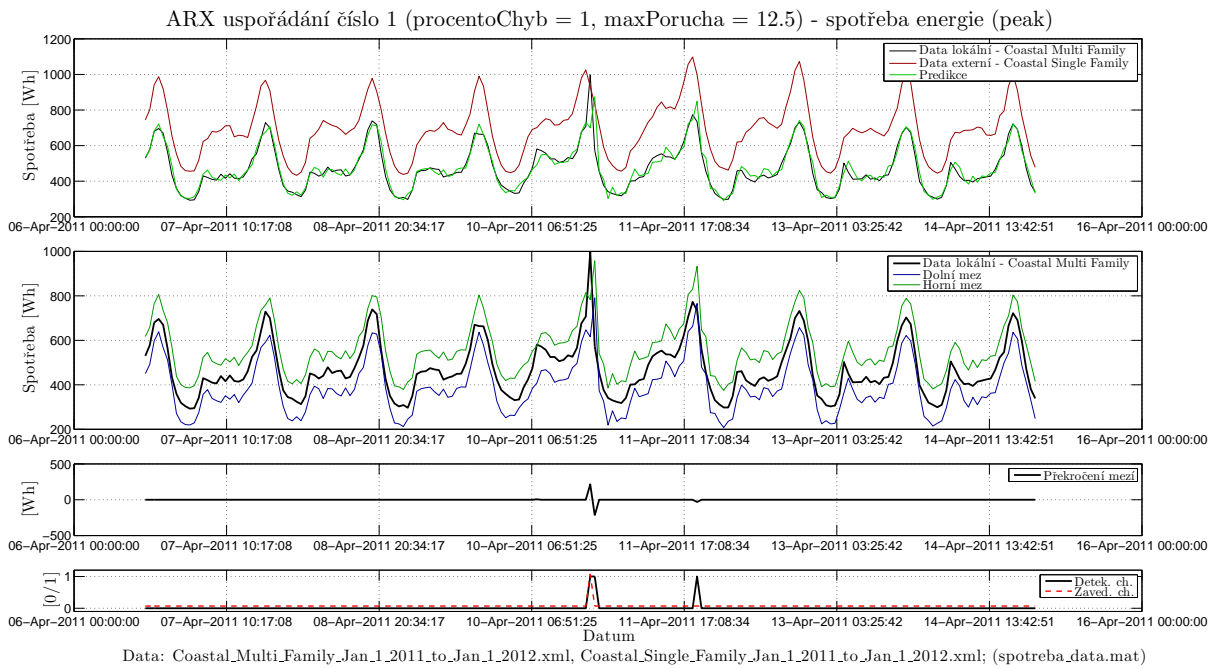


Obrázek 6.7: Detail výsledků ARX uspořádání číslo 1 - teploty *Budihostice*, *Červený Újezd*, *Chloumek*, *ČZU* (drift).

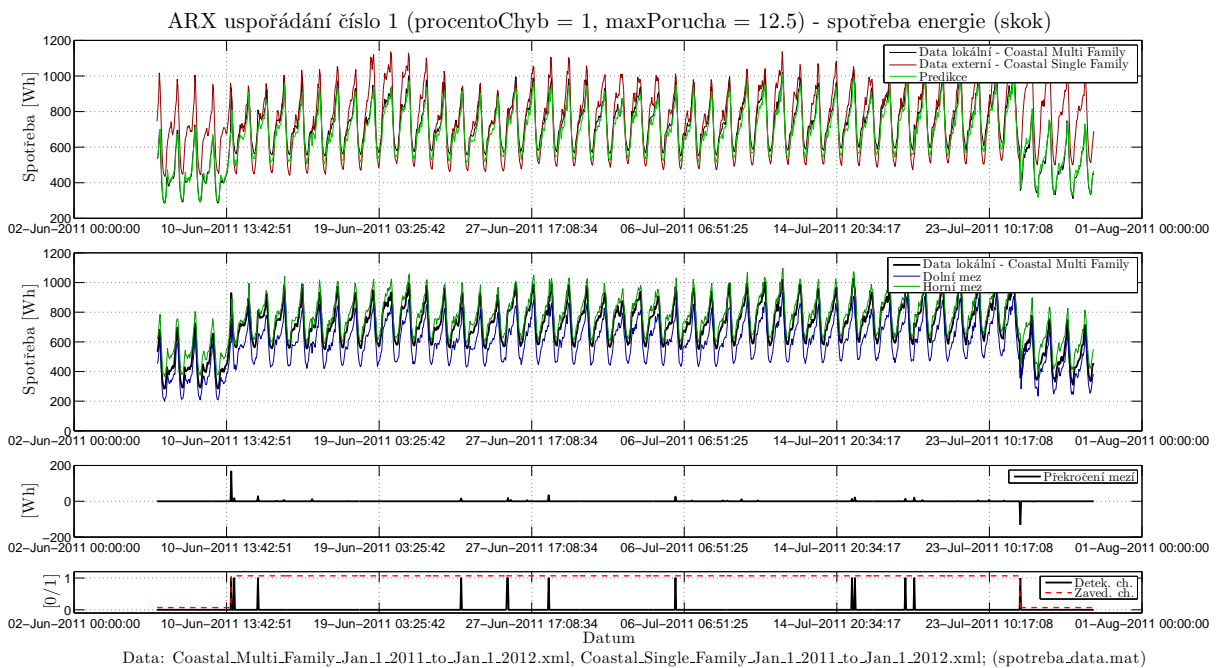


Obrázek 6.8: Detail výsledků ARX uspořádání číslo 1 - teploty *Budihostice*, *Červený Újezd*, *Chloumek*, *ČZU* (gain).

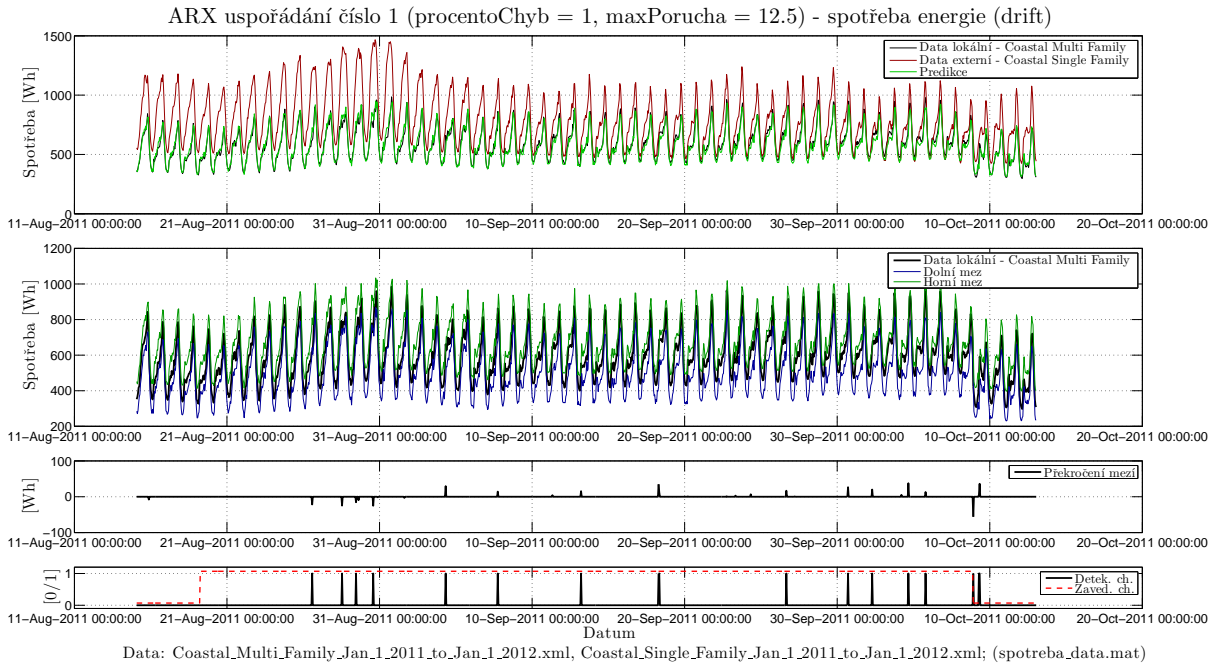
Obrázek 6.9: Výsledky ARX uspořádání číslo 1 - spotřeba energie *Coastal**Multi a Single Family*



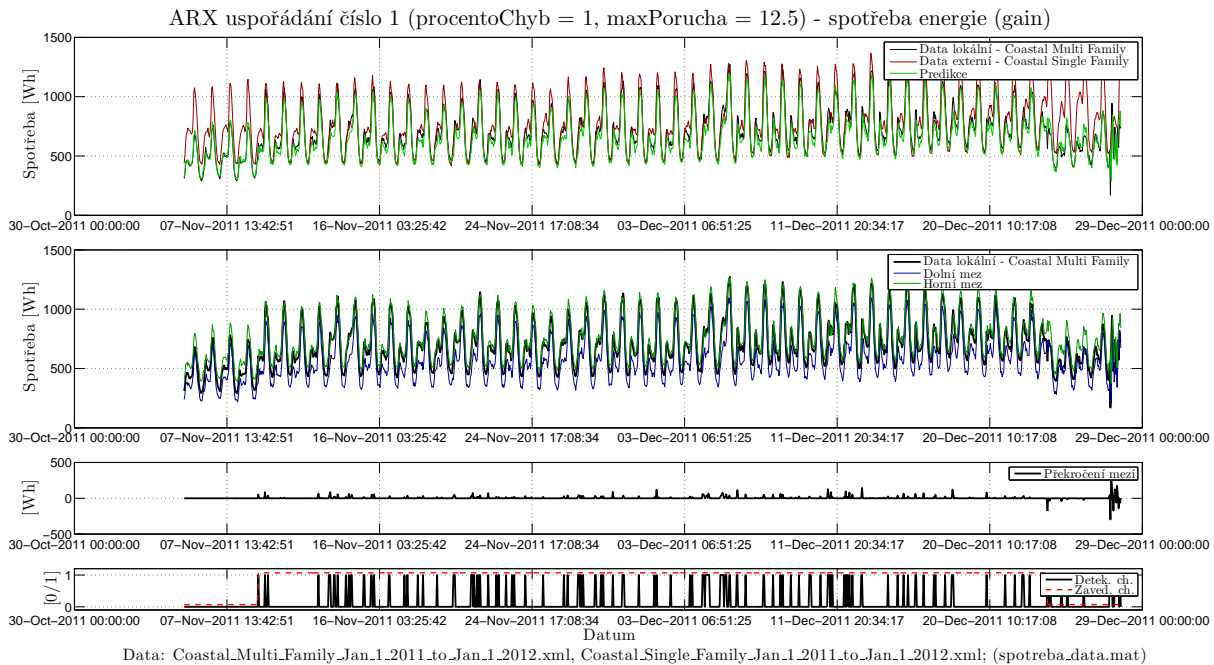
Obrázek 6.10: Detail výsledků ARX uspořádání číslo 1 - spotřeba energie *Coastal Multi a Single Family* (peak).



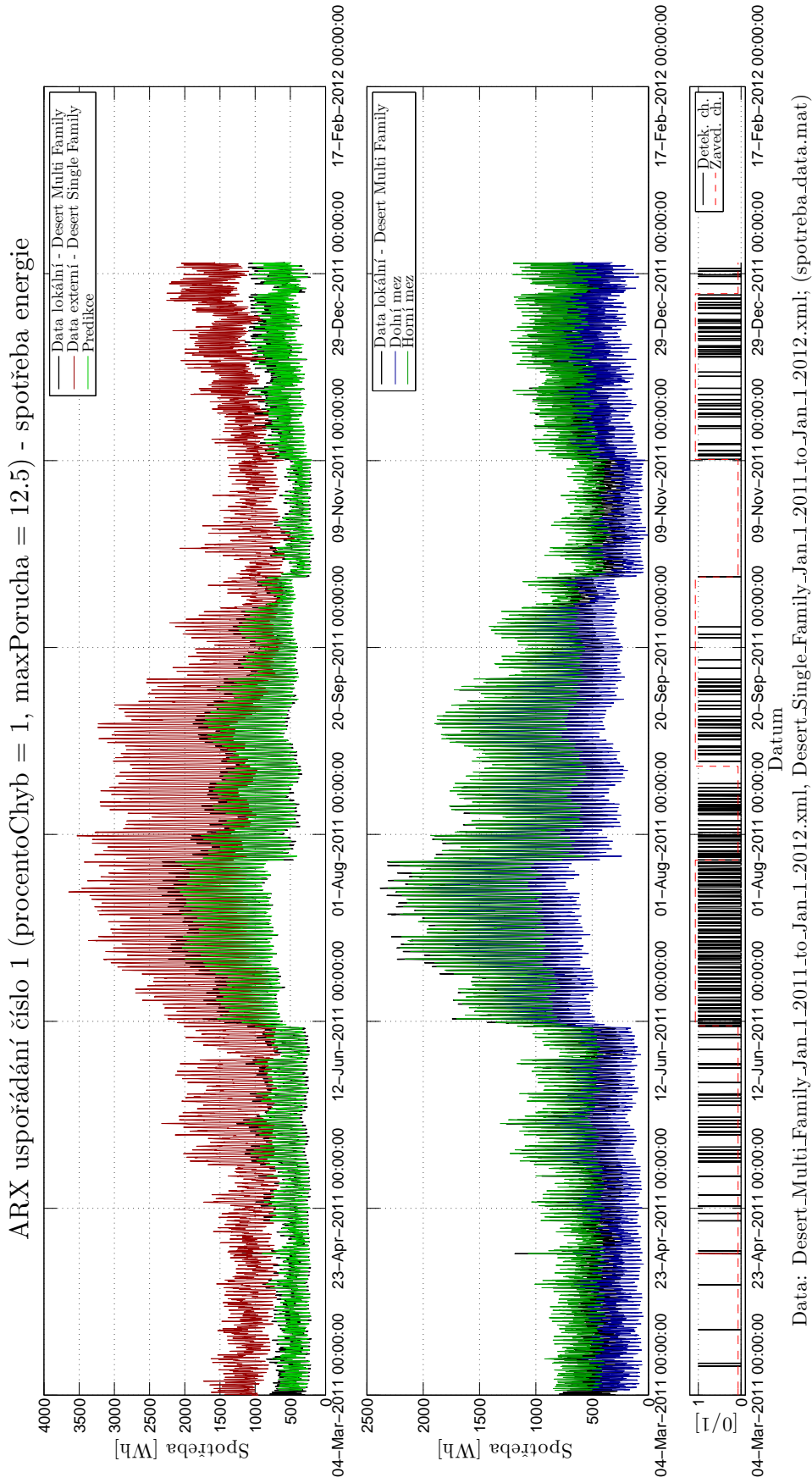
Obrázek 6.11: Detail výsledků ARX uspořádání číslo 1 - spotřeba energie *Coastal Multi a Single Family* (skok).

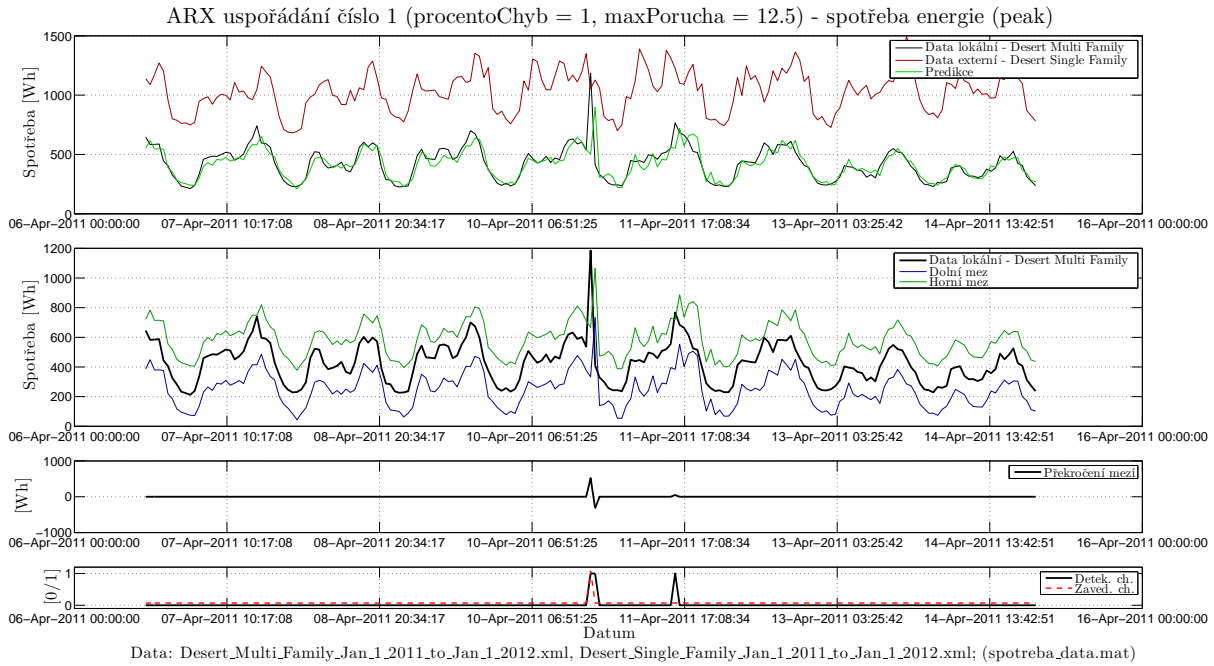


Obrázek 6.12: Detail výsledků ARX uspořádání číslo 1 - spotřeba energie *Coastal Multi a Single Family* (drift).

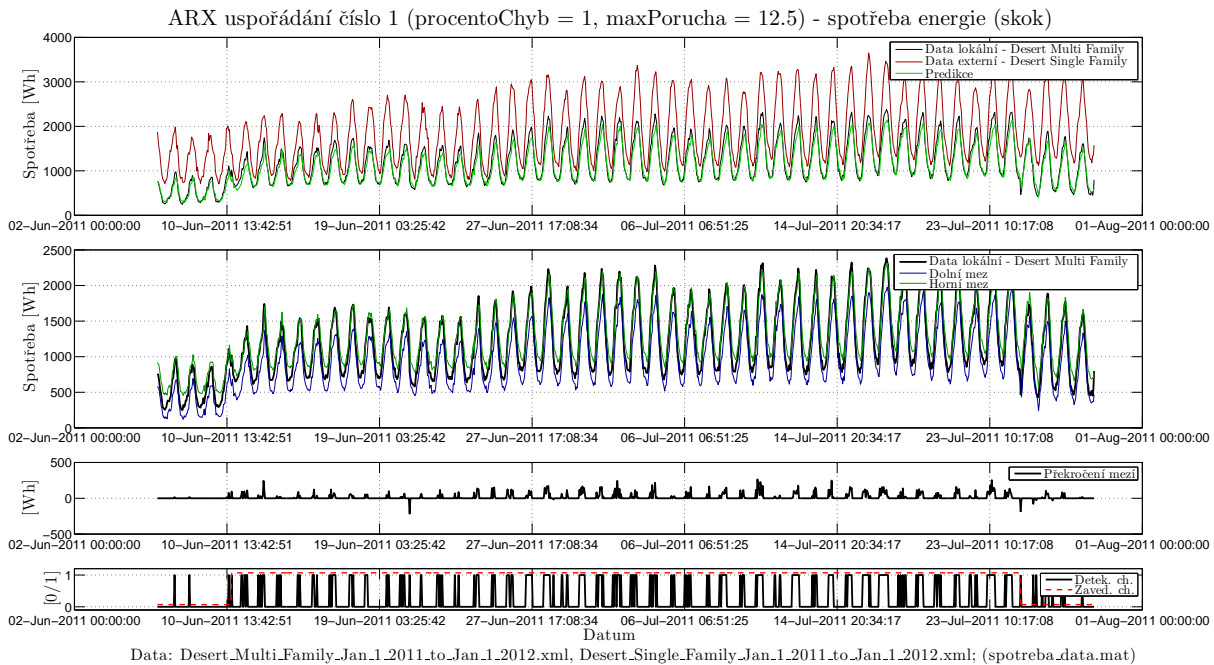


Obrázek 6.13: Detail výsledků ARX uspořádání číslo 1 - spotřeba energie *Coastal Multi a Single Family* (gain).

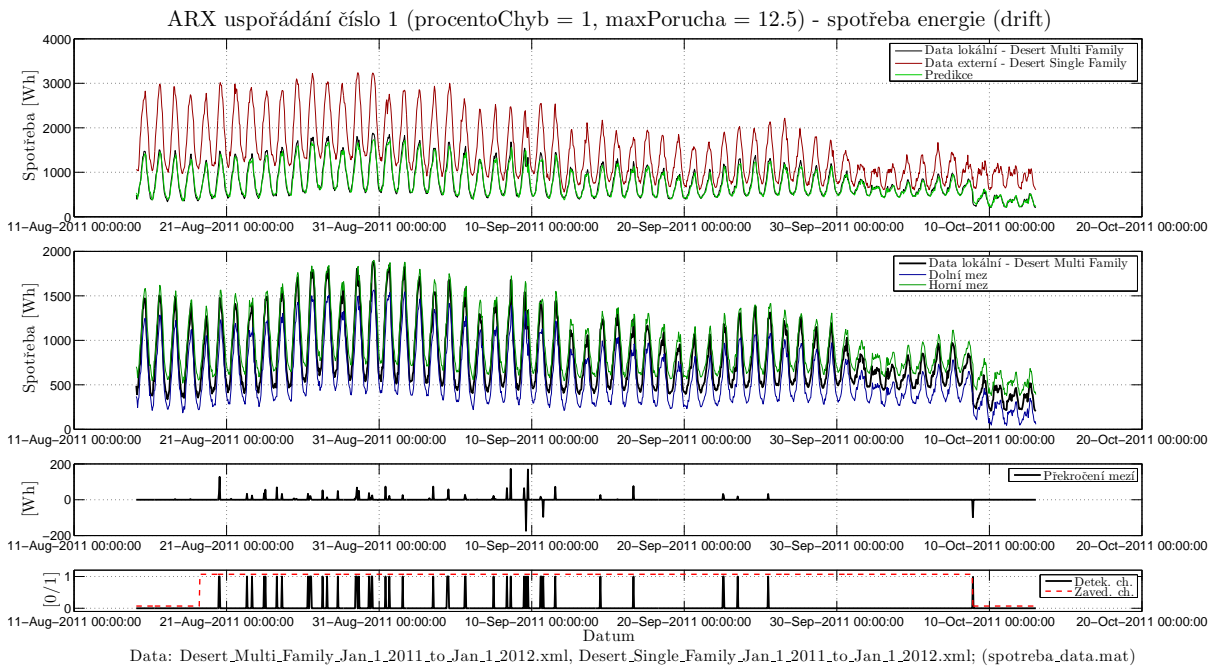
Obrázek 6.14: Výsledky ARX uspořádání číslo 1 - spotřeba energie *Desert**Multi a Single Family*



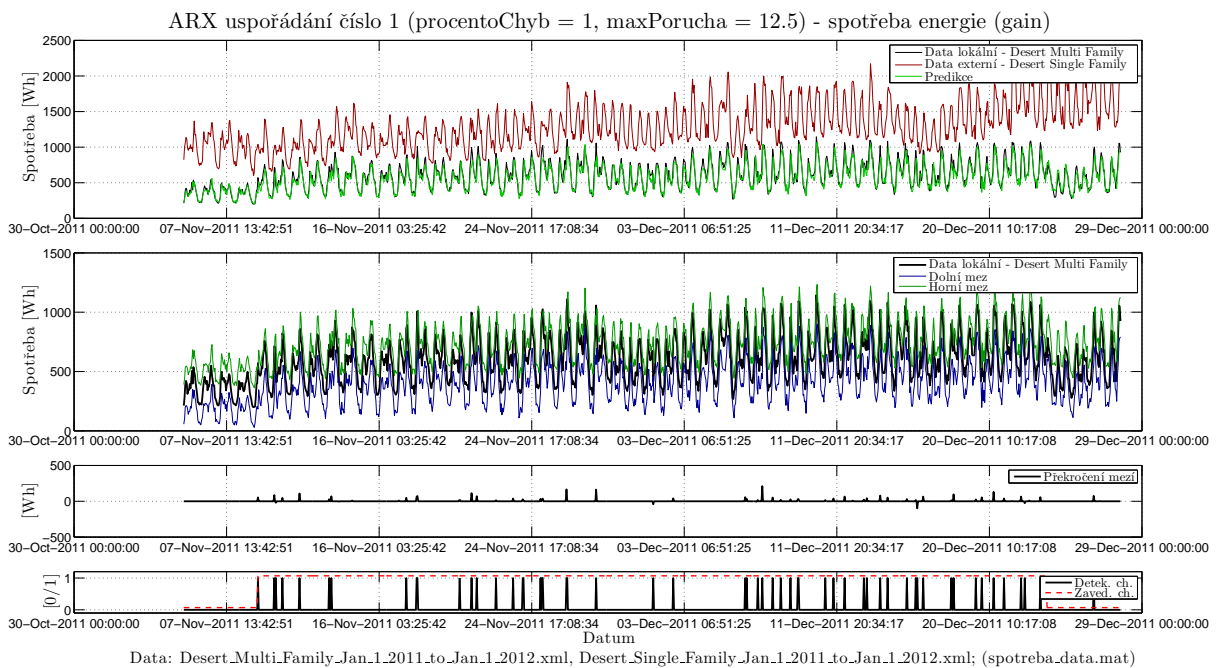
Obrázek 6.15: Detail výsledků ARX uspořádání číslo 1 - spotřeba energie
Desert Multi a Single Family (peak).



Obrázek 6.16: Detail výsledků ARX uspořádání číslo 1 - spotřeba energie
Desert Multi a Single Family (skok).



Obrázek 6.17: Detail výsledků ARX uspořádání číslo 1 - spotřeba energie *Desert Multi a Single Family* (drift).



Obrázek 6.18: Detail výsledků ARX uspořádání číslo 1 - spotřeba energie *Desert Multi a Single Family* (gain).

6.2.2 ARX (AutoRegressive model with eXternal input) uspořádání číslo 2

Z výsledků v tabulkách 6.4, 6.5, 6.6, B.3 a B.4 je patrné, že metoda detekce poruch pomocí ARX (uspořádání číslo 2) je celkem účinná. Metoda dosahuje výborných výsledků u teplotních dat *Budihostice*, *Červený Újezd*, *Chloumek*, *ČZU*, viz tabulka 6.1, a přiměřených u dat spotřeby energie. Výjimku tvoří velice komplikovaná data *Desert Multi Family*, *Desert Single Family*, kde je úspěšnost nízká. Důvodem je, že je tato metoda založena na sledování difference mezi predikcí externích hodnot a lokálními daty. Neboli při učení modelu dojde k výpočtu rozmezí, ve kterém by se měla difference pohybovat. V případě dat spotřeby energie *Desert Multi Family*, *Desert Single Family* se při testování tato difference mezi predikcí externích hodnot a lokálními daty velice často a výrazně mění, a proto je použití této metody na tato data nepříliš vhodné.

Na obrázcích 6.19, 6.24 a 6.29 je zobrazen celkový pohled na testovaná data, stanovené meze a vyhodnocení, zda byla zavedená chyba detekována.

Detekce odlehle hodnoty pro jednotlivá data je zobrazena na obrázcích 6.20, 6.25 a 6.30. Vidíme, že tato hodnota je správně detekována a projeví se předpokládaným překročením mezí v daném směru chyby.

Na obrázku 6.21 s teplotními daty vidíme, že skoková změna hodnoty senzoru je výborně detekována. U dat spotřeby energie 6.26 a 6.30 je detekce této chyby nižší. Důvodem je široký konfidenční interval v kontrastu se zavedenou velikostí chyby a tedy i menší počet překročení stanovených mezí.

Obrázek 6.22 ukazuje téměř příkladnou detekci driftu senzoru u teplotních dat. Z překročení mezí je patrná stoupající tendence driftu. Na obrázku 6.27 spotřeby energie vidíme, že množství detekovaných poruch způsobených driftem je nižší, ale je zde též patrné, že postupem času by toto množství stouplo a byl by zde totožný trend překročení mezí jako u teplotních dat.

Detekovaná změna zesílení (gain) senzoru je na obrázcích 6.23, 6.28 a 6.33, ze kterých je vidět, že pokud je změna dostatečná a meze nejsou příliš široké, je chyba detekována.

Tabulka 6.4: Výsledky ARX uspořádání číslo 2 - teploty *Budihostice, Červený Újezd, Chloumek, ČZU*.

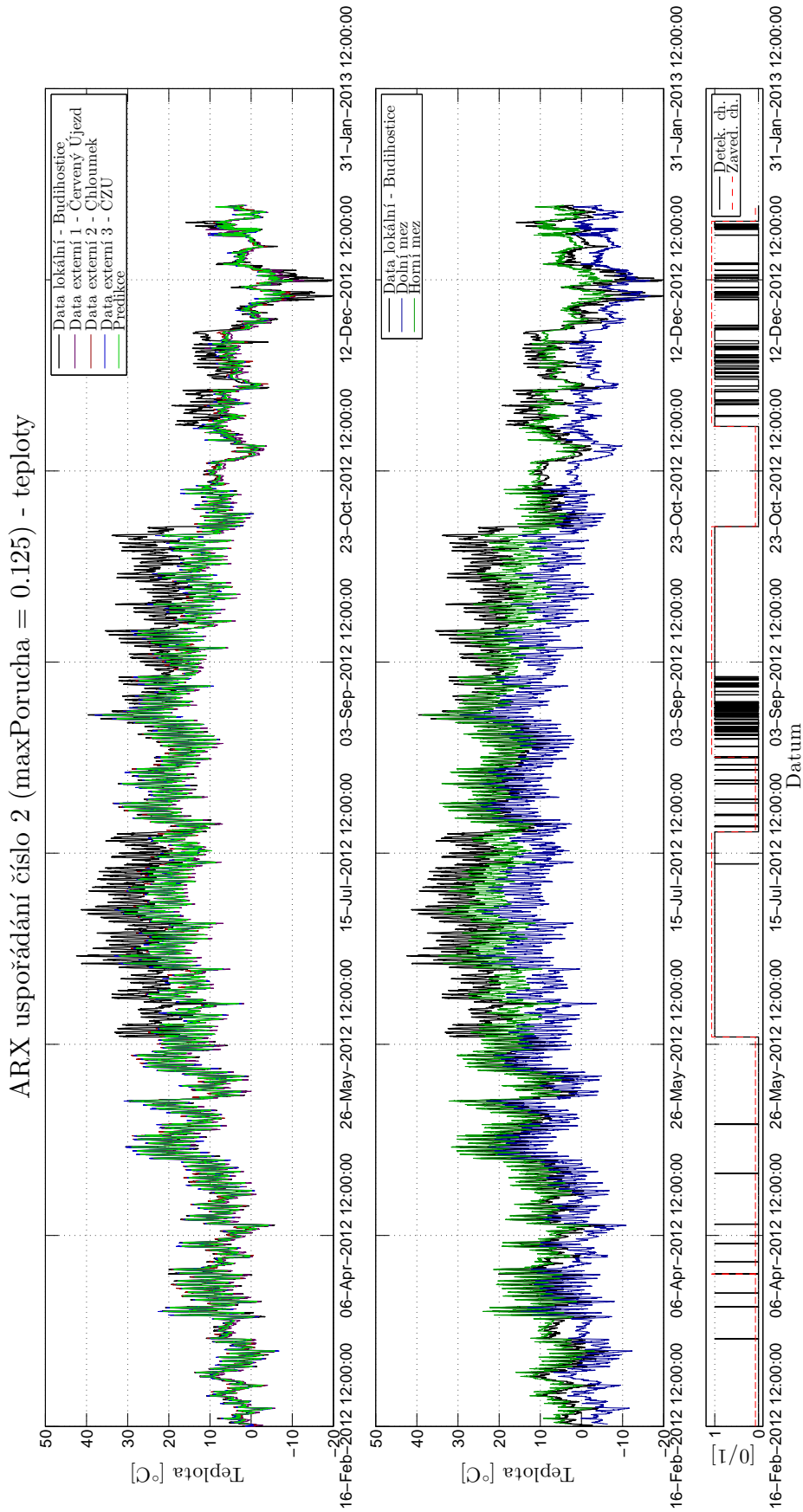
Max porucha [°C]	TP [-]	FP [-]	FN [-]	TN [-]	Specificita [%]	Senzitivita [%]
0,125	6010	42	2039	7245	99,42	74,67
0,500	5970	26	2079	7261	99,64	74,17
1,500	5920	20	2129	7267	99,73	73,55

Tabulka 6.5: Výsledky ARX uspořádání číslo 2 - spotřeba energie *Coastal Multi a Single Family*.

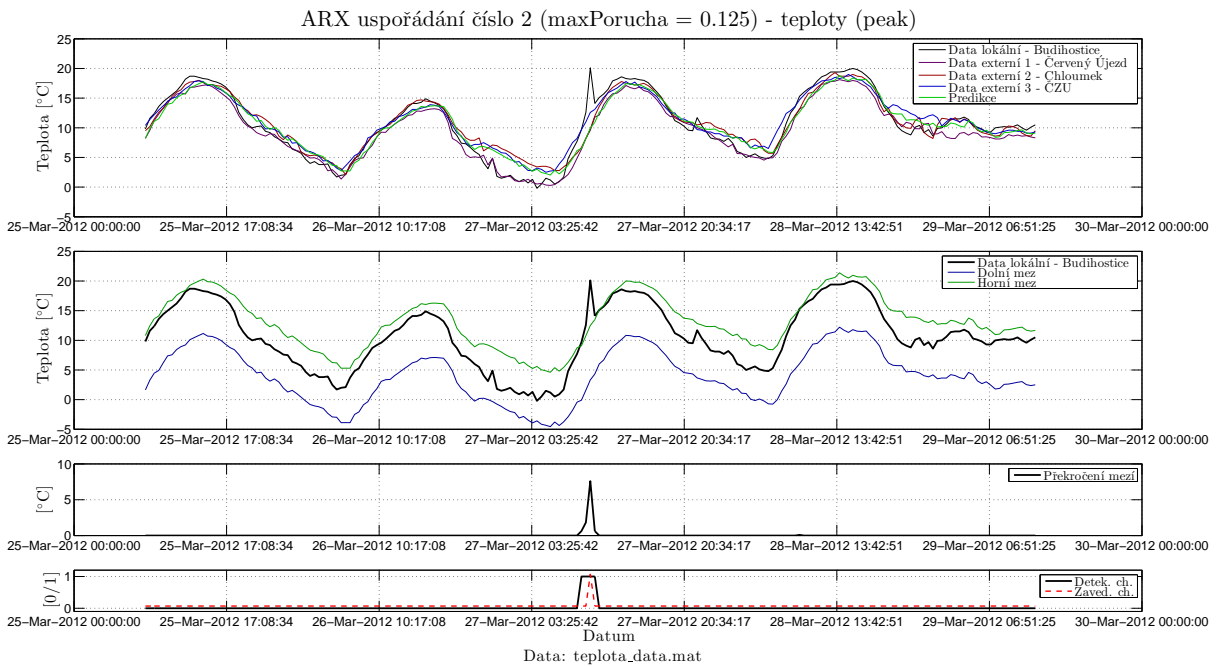
Max porucha [Wh]	TP [-]	FP [-]	FN [-]	TN [-]	Specificita [%]	Senzitivita [%]
12,5	1783	39	1562	3888	99,01	53,30
50,0	1712	31	1633	3896	99,21	51,18
150,0	1612	22	1733	3905	99,44	48,19

Tabulka 6.6: Výsledky ARX uspořádání číslo 2 - spotřeba energie *Desert Multi a Single Family*.

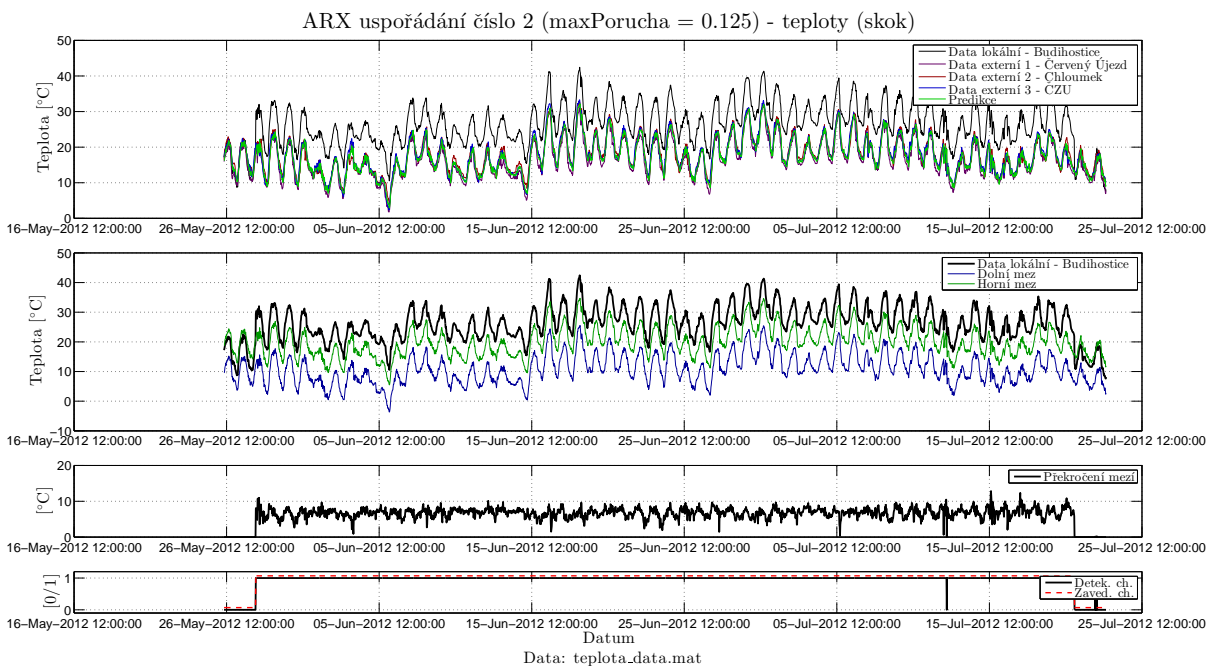
Max porucha [Wh]	TP [-]	FP [-]	FN [-]	TN [-]	Specificita [%]	Senzitivita [%]
12,5	548	118	2797	3809	97,00	16,38
50,0	516	104	2829	3823	97,35	15,43
150,0	407	84	2938	3843	97,86	12,17



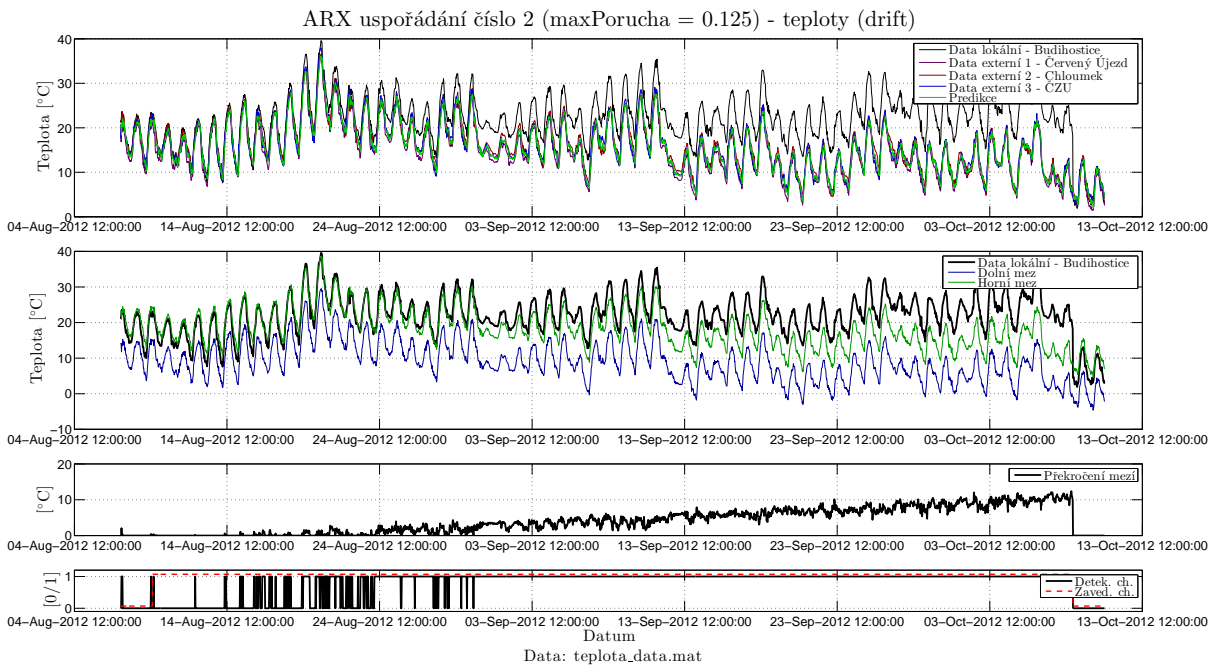
Obrázek 6.19: Výsledky ARX uspořádání číslo 2 - teploty *Budihostice*,
Červený Újezd, *Chloumek*, *ČZU*.



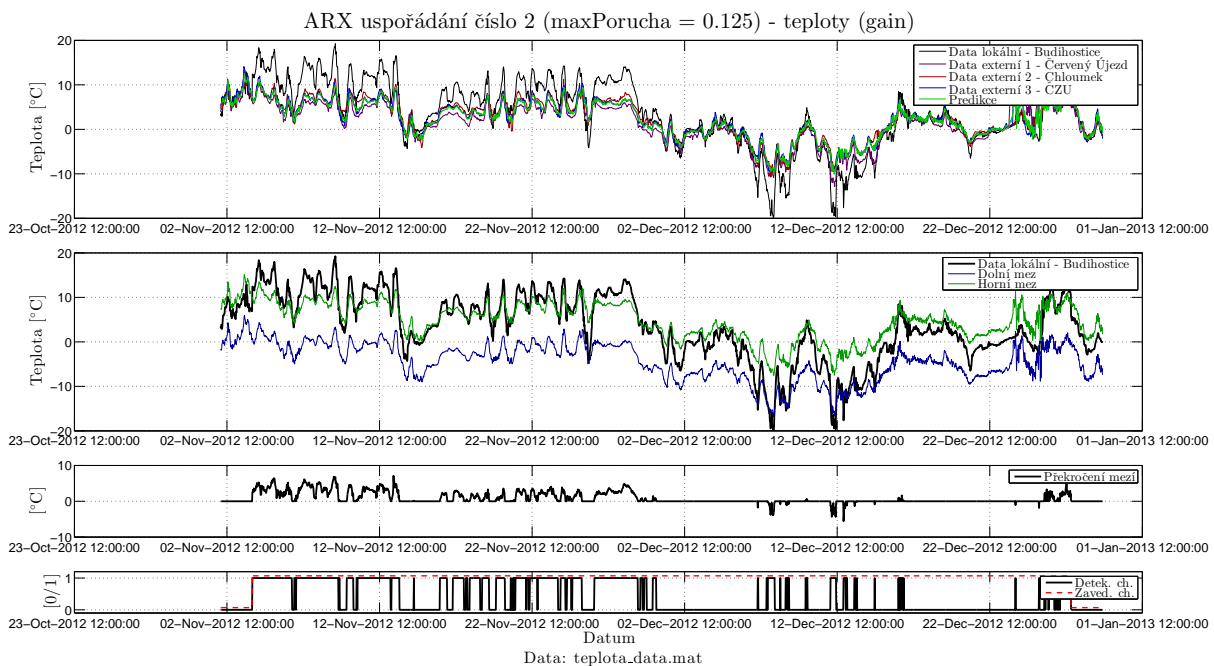
Obrázek 6.20: Detail výsledků ARX uspořádání číslo 2 - teploty *Budihostice*, *Červený Újezd*, *Chloumek*, *ČZU* (peak).



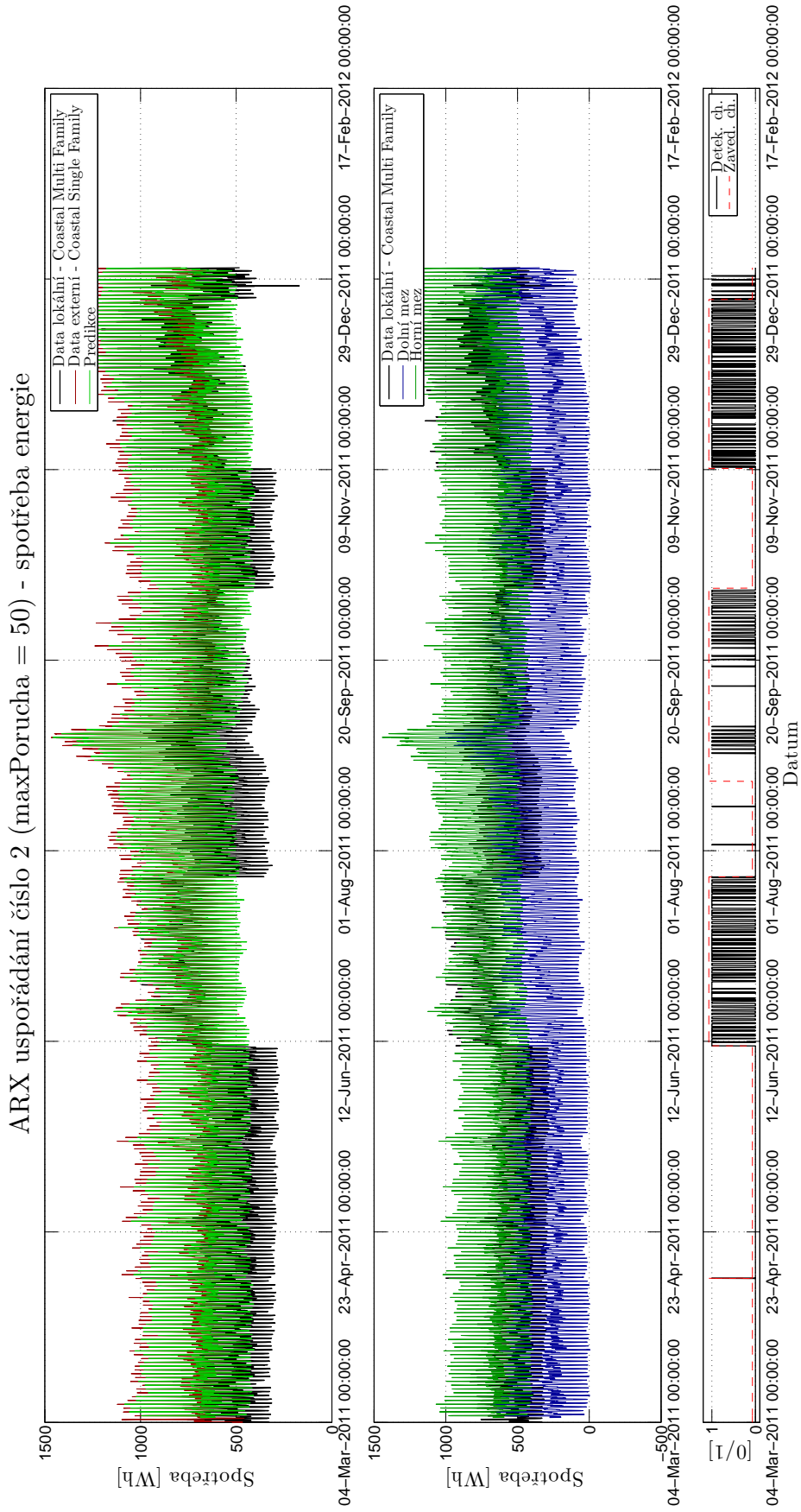
Obrázek 6.21: Detail výsledků ARX uspořádání číslo 2 - teploty *Budihostice*, *Červený Újezd*, *Chloumek*, *ČZU* (skok).

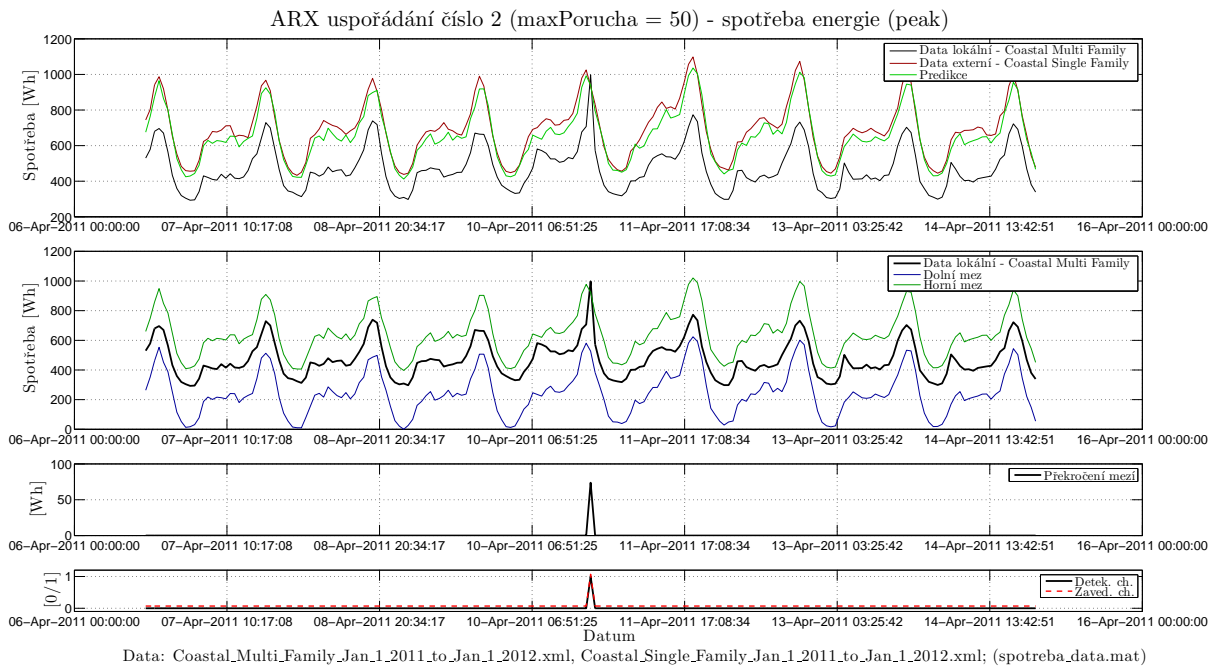


Obrázek 6.22: Detail výsledků ARX uspořádání číslo 2 - teploty *Budihostice*, *Červený Újezd*, *Chloumek*, *ČZU* (drift).

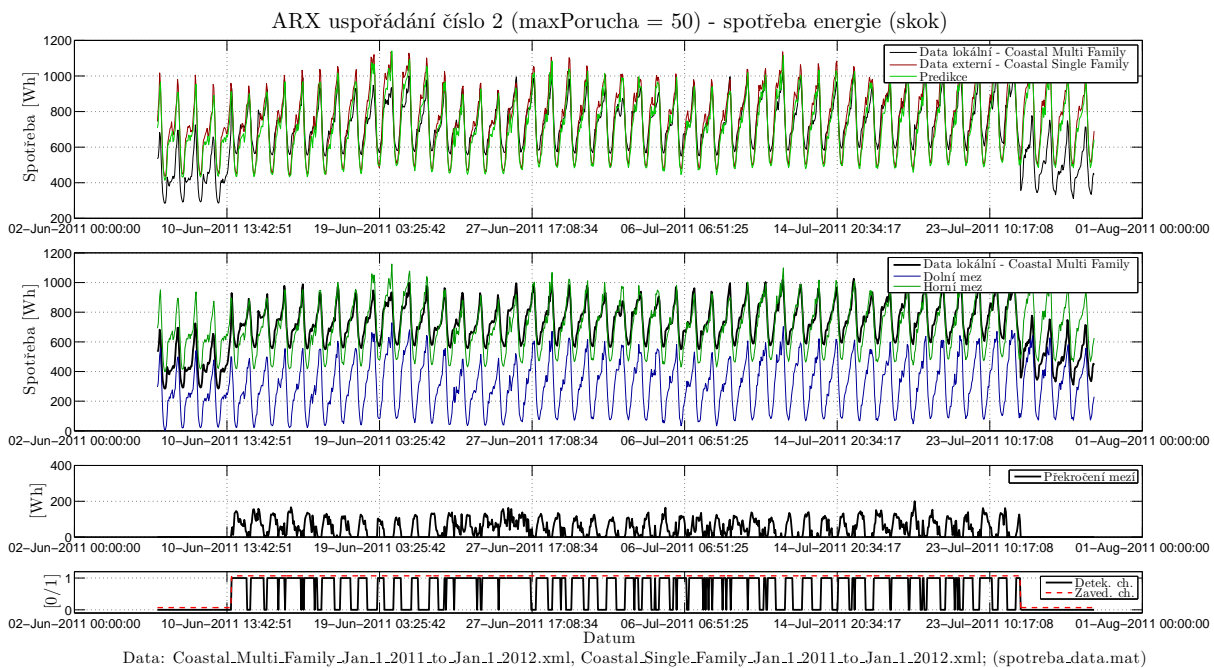


Obrázek 6.23: Detail výsledků ARX uspořádání číslo 2 - teploty *Budihostice*, *Červený Újezd*, *Chloumek*, *ČZU* (gain).

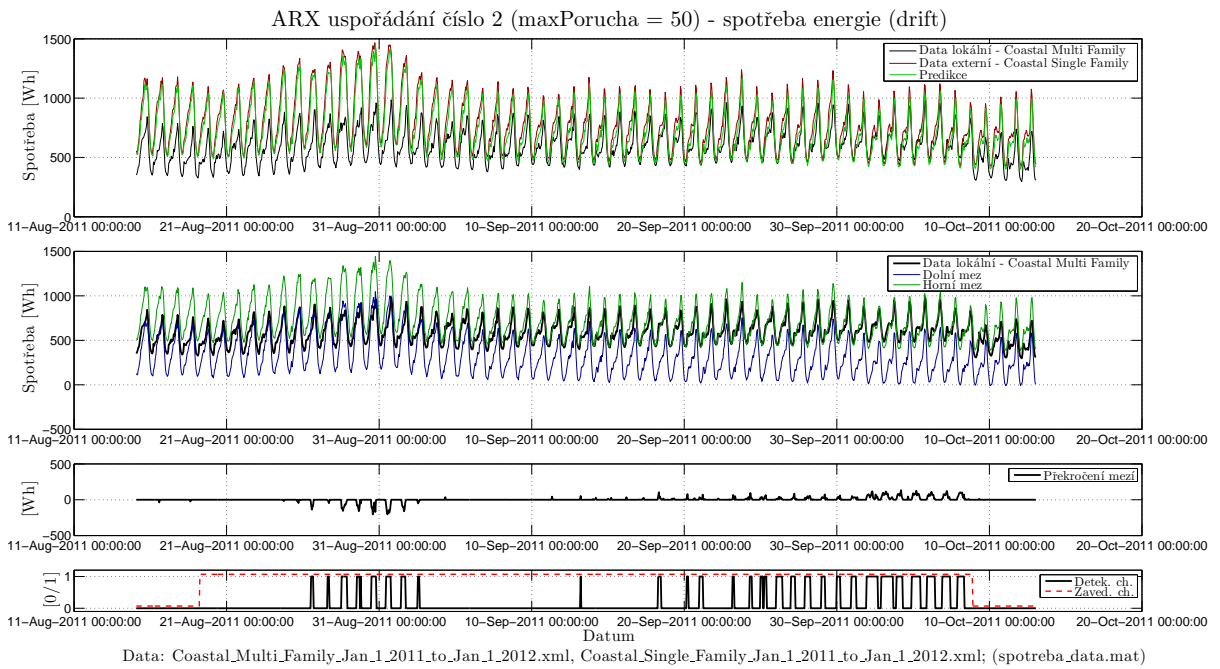
Obrázek 6.24: Výsledky ARX uspořádání číslo 2 - spotřeba energie *Coastal**Multi a Single Family*



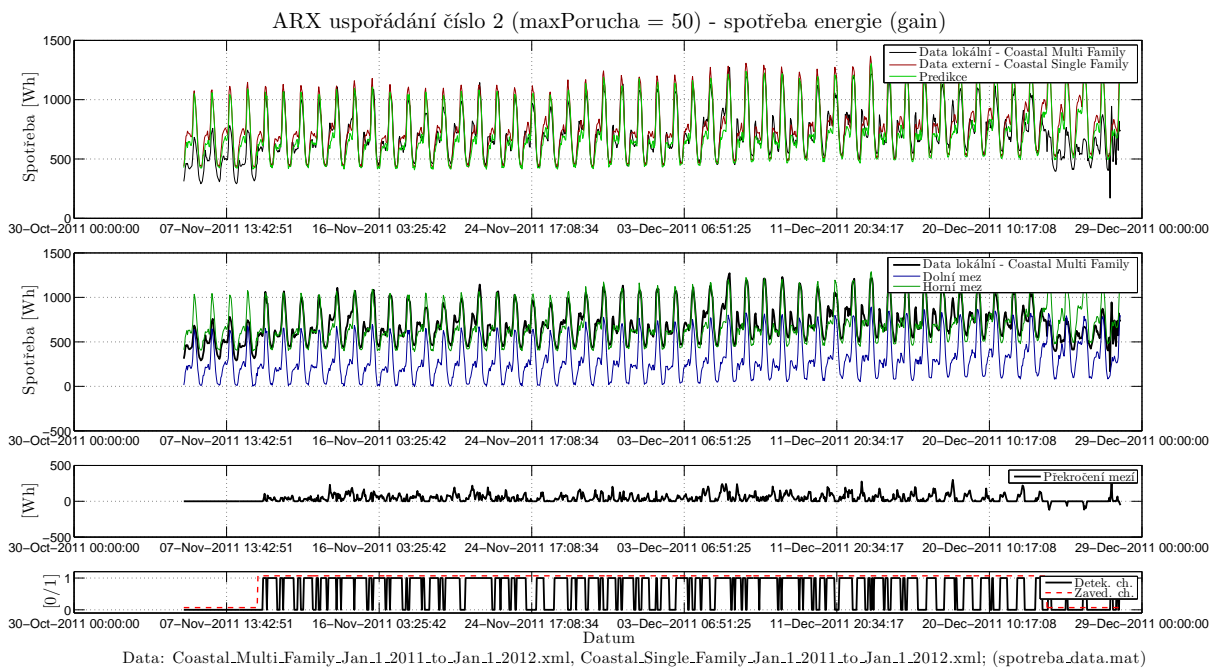
Obrázek 6.25: Detail výsledků ARX uspořádání číslo 2 - spotřeba energie *Coastal Multi a Single Family* (peak).



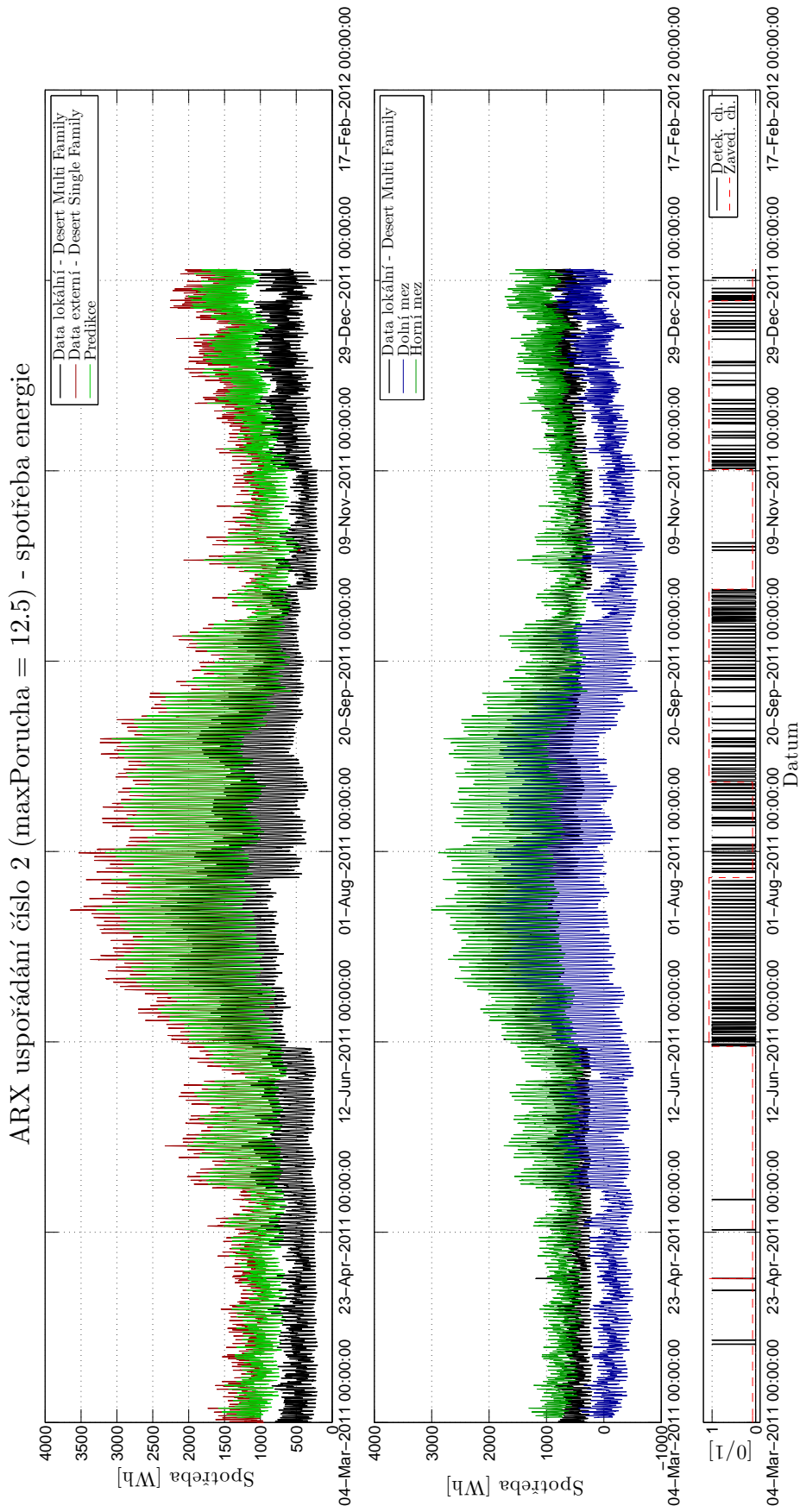
Obrázek 6.26: Detail výsledků ARX uspořádání číslo 2 - spotřeba energie *Coastal Multi a Single Family* (skok).

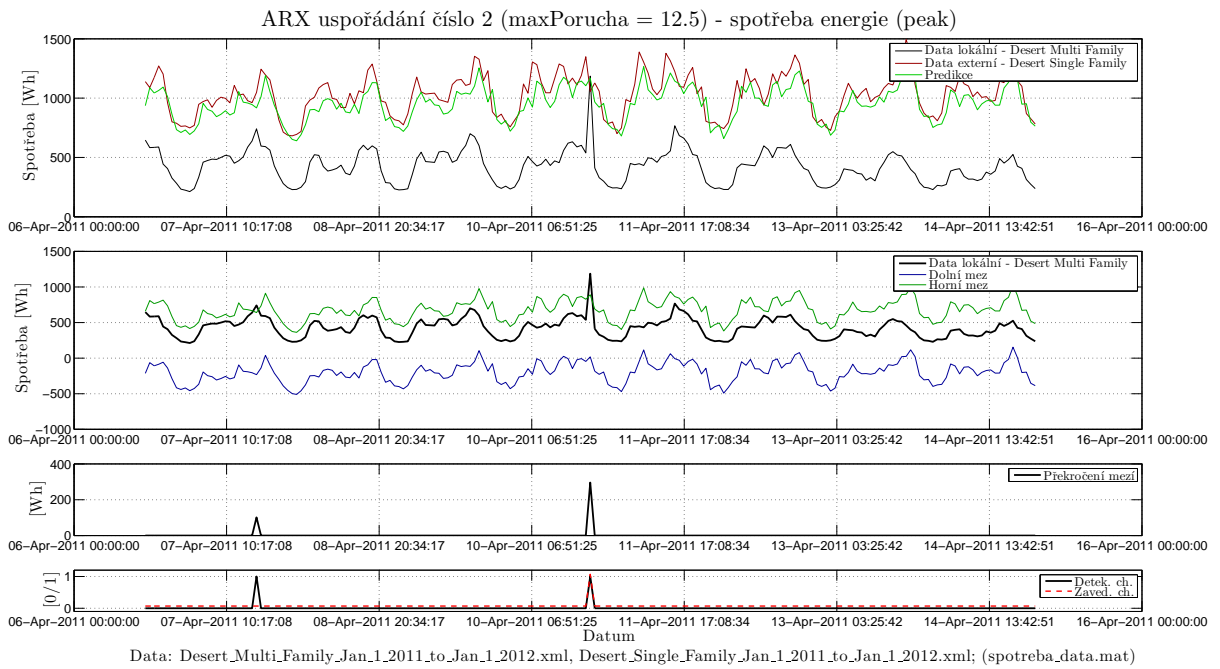


Obrázek 6.27: Detail výsledků ARX uspořádání číslo 2 - spotřeba energie
Coastal Multi a Single Family (drift).

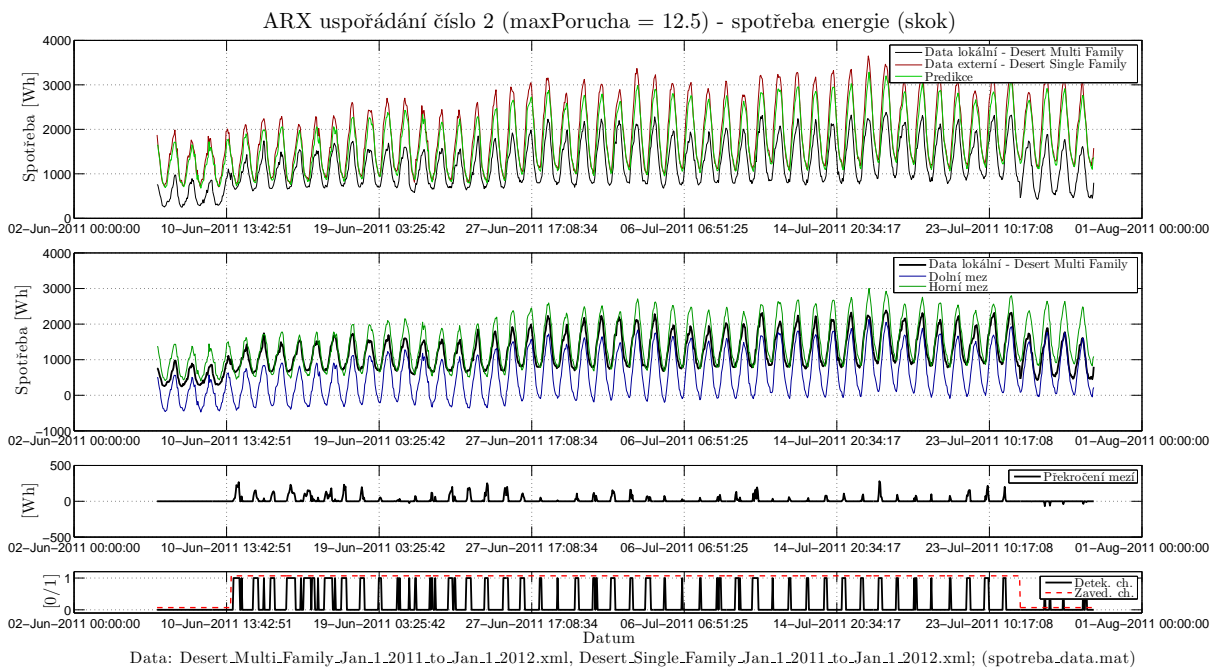


Obrázek 6.28: Detail výsledků ARX uspořádání číslo 2 - spotřeba energie
Coastal Multi a Single Family (gain).

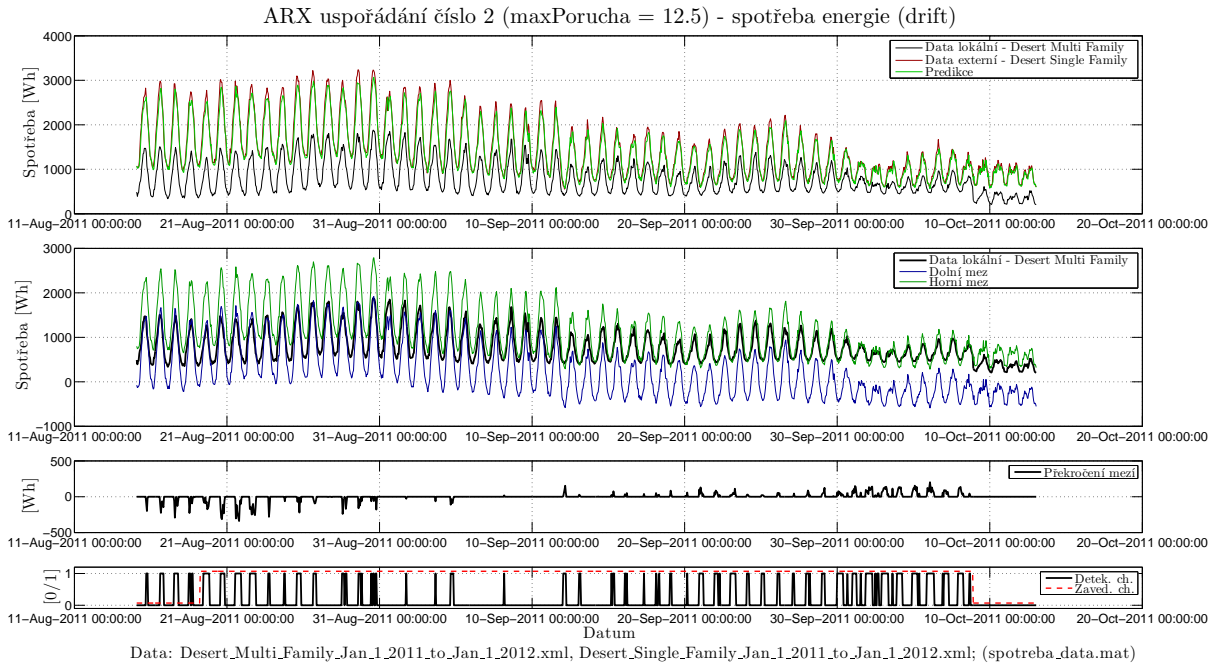
Obrázek 6.29: Výsledky ARX uspořádání číslo 2 - spotřeba energie *Desert**Multi a Single Family*



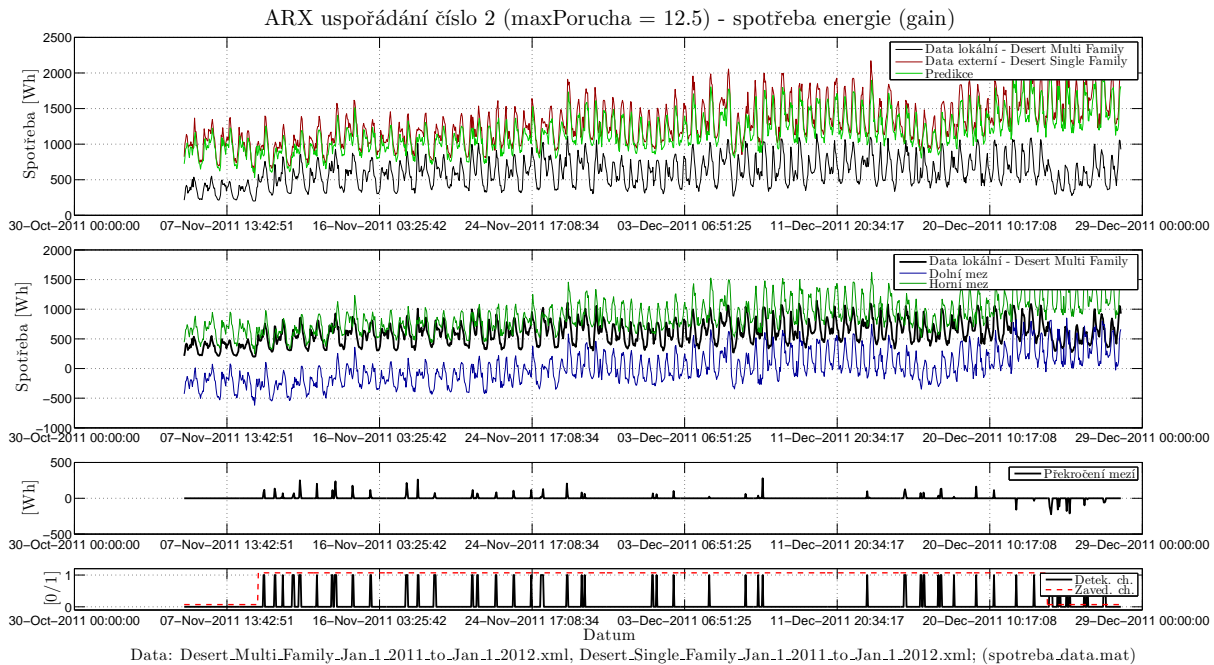
Obrázek 6.30: Detail výsledků ARX uspořádání číslo 2 - spotřeba energie *Desert Multi a Single Family* (peak).



Obrázek 6.31: Detail výsledků ARX uspořádání číslo 2 - spotřeba energie *Desert Multi a Single Family* (skok).



Obrázek 6.32: Detail výsledků ARX uspořádání číslo 2 - spotřeba energie
Desert Multi a Single Family (drift).



Obrázek 6.33: Detail výsledků ARX uspořádání číslo 2 - spotřeba energie
Desert Multi a Single Family (gain).

6.2.3 GMM (Gaussian Mixture Model)

Metoda detekce poruch senzorů pomocí GMM (Gaussian Mixture Model) podle výsledků uvedených v tabulkách 6.7, 6.8, 6.9, B.5 a B.6 patří mezi velice účinné metody. Podobně jako ARX (uspořádání číslo 2) tato metoda dosahuje nejlepších výsledků na teplotních datech, což je způsobeno přibližně lineární závislostí mezi lokálními a externími daty, viz kapitola 5.3. U dat spotřeby energie je detekce poruch nižší, jelikož zde již závislost není tolik lineární, ale především je zde patrný větší rozptyl, což vede k širšímu konfidenčnímu intervalu a nižší detekci poruch. Výjimku ovšem opět tvoří velice komplikovaná data *Desert Multi Family*, *Desert Single Family*, kde je podle tabulky 6.9 úspěšnost vysoká, ale z obrázku 6.44 jasně vyplývá, že jsou tyto výsledky zkresleny. Problémem je špatně naučený model, který byl naučen na velice odlišná data, než na kterých byl nyní testován.

Obrázky 6.34, 6.39 a 6.44 opět zachycují náhled na testovaná data, stanovené meze a vyhodnocení, zda byla zavedená chyba detekována.

Na dalších obrázcích 6.35, 6.40 a 6.45 je zobrazen detail při detekci odlehle hodnoty, která se projeví v překročení mezí. Za povšimnutí na těchto obrázcích stojí také detail tvaru a šířky konfidenčního intervalu, který je závislý na parametrech *ALFA*, *MEZ*, viz kapitola 5.3. Z obrázků 6.35, 6.40 je vidět, že dolní a horní mez kopírují tvar testovaných lokálních dat a samotná šířka konfidenčního intervalu je přiměřená.

Skoková změna hodnoty senzoru, drift i změna zesílení jsou touto metodou u teplotních dat a dat spotřeby energie *Coastal Multi*, *Single Family* úspěšně detekovány, viz obrázky 6.36, 6.41, 6.37, 6.42, 6.38 a 6.43. V případě detekce driftu je vidět postupný zvyšující se trend detekce této poruchy.

Tabulka 6.7: Výsledky GMM - teploty *Budihostice, Červený Újezd, Chloumek, ČZU*.

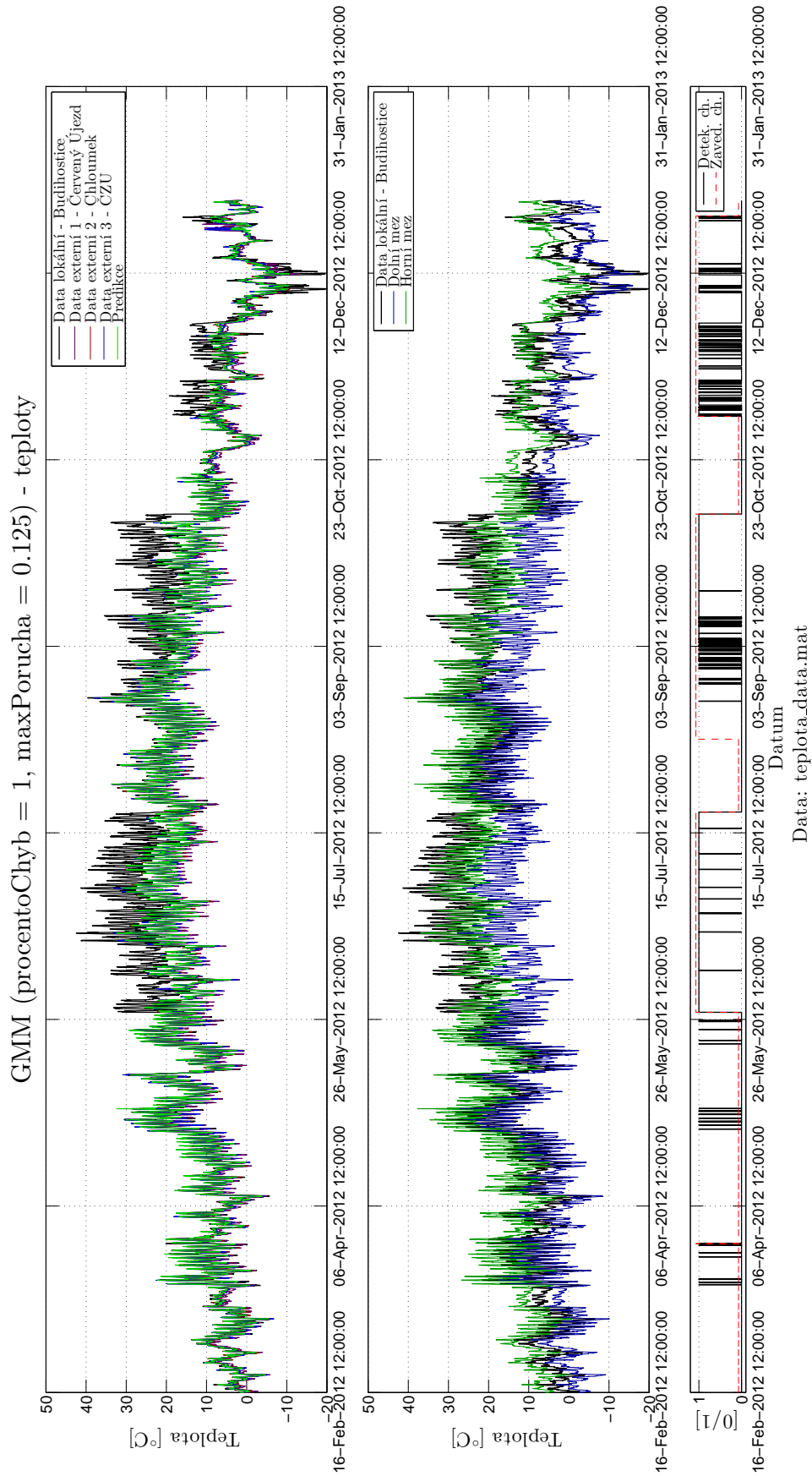
Procento chyb [%]	Max porucha [°C]	TP [-]	FP [-]	FN [-]	TN [-]	Specificita [%]	Senzitivita [%]
0,5	0,125	4683	25	3366	7262	99,66	58,18
	0,500	4795	64	3254	7223	99,12	59,57
	1,500	4726	103	3323	7184	98,59	58,72
1,0	0,125	4919	56	3130	7231	99,23	61,11
	0,500	4879	69	3170	7218	99,05	60,62
	1,500	4805	59	3244	7228	99,19	59,70
1,5	0,125	5129	232	2920	7055	96,82	63,72
	0,500	5290	207	2759	7080	97,16	65,72
	1,500	5243	113	2806	7174	98,45	65,14

Tabulka 6.8: Výsledky GMM - spotřeba energie *Coastal Multi a Single Family*.

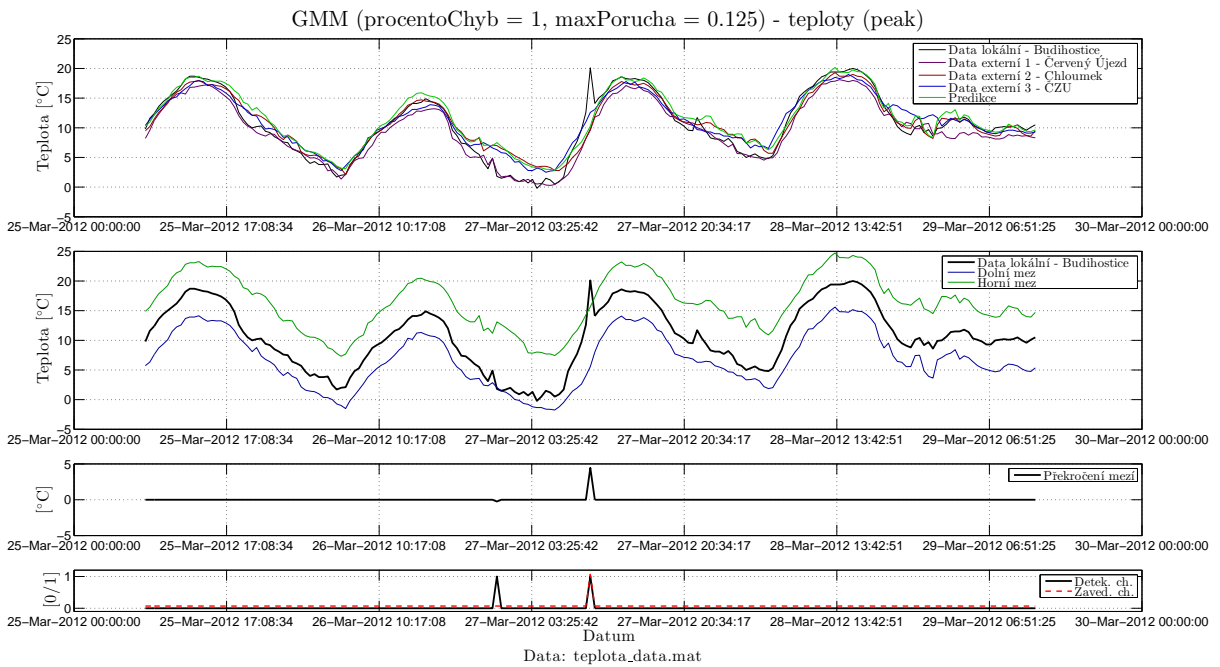
Procento chyb [%]	Max porucha [Wh]	TP [-]	FP [-]	FN [-]	TN [-]	Specificita [%]	Senzitivita [%]
0,5	12,5	2021	122	1324	3805	96,89	60,42
	50,0	2011	94	1334	3833	97,61	60,12
	150,0	1947	53	1398	3874	98,65	58,21
1,0	12,5	2094	214	1251	3713	94,55	62,60
	50,0	2061	123	1284	3804	96,87	61,61
	150,0	1981	81	1364	3846	97,94	59,22
1,5	12,5	2217	277	1128	3650	92,95	66,28
	50,0	2044	120	1301	3807	96,94	61,11
	150,0	2090	97	1255	3830	97,53	62,48

Tabulka 6.9: Výsledky GMM - spotřeba energie *Desert Multi a Single Family*.

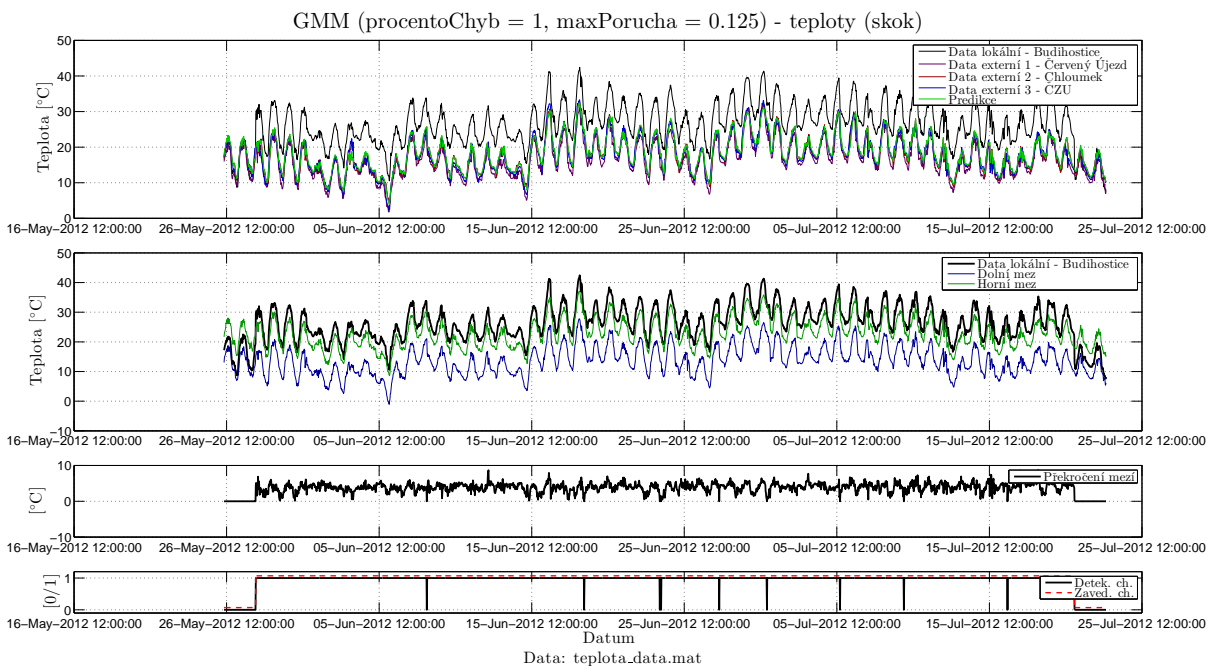
Procento chyb [%]	Max porucha [Wh]	TP [-]	FP [-]	FN [-]	TN [-]	Specificita [%]	Senzitivita [%]
0,5	12,5	1719	336	1626	3591	91,44	51,39
	50,0	1661	316	1684	3611	91,95	49,66
	150,0	1586	286	1759	3641	92,72	47,41
1,0	12,5	1889	404	1456	3523	89,71	56,47
	50,0	1842	369	1503	3558	90,60	55,07
	150,0	1754	335	1591	3592	91,47	52,44
1,5	12,5	2035	448	1310	3479	88,59	60,84
	50,0	2002	435	1343	3492	88,92	59,85
	150,0	1913	378	1432	3549	90,37	57,19



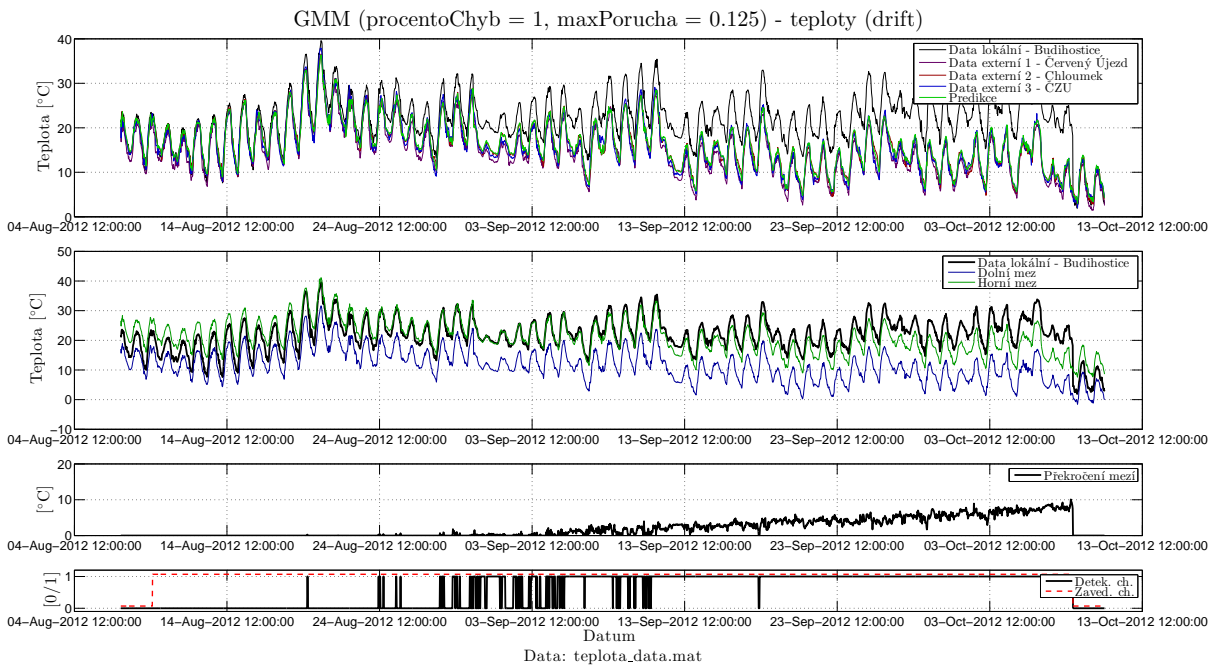
Obrázek 6.34: Výsledky GMM - teploty Budihostice, Červený Újezd,
Chloumek, ČZU.



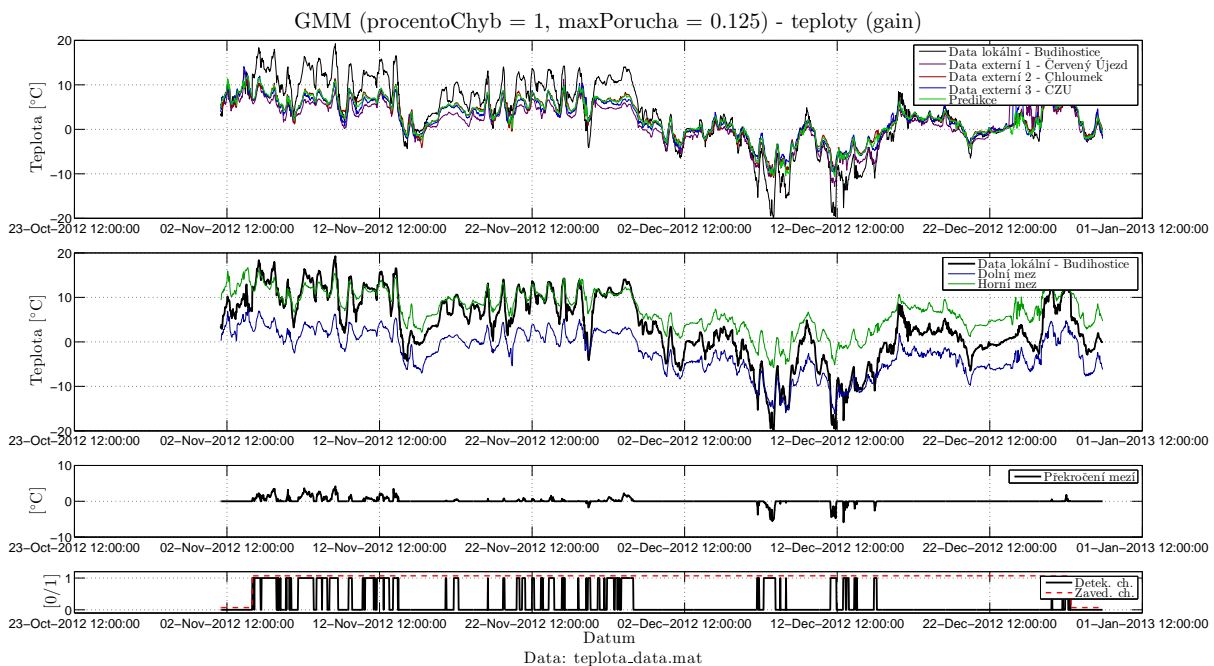
Obrázek 6.35: Detail výsledků GMM - teploty *Budihostice, Červený Újezd, Chloumek, ČZU* (peak).



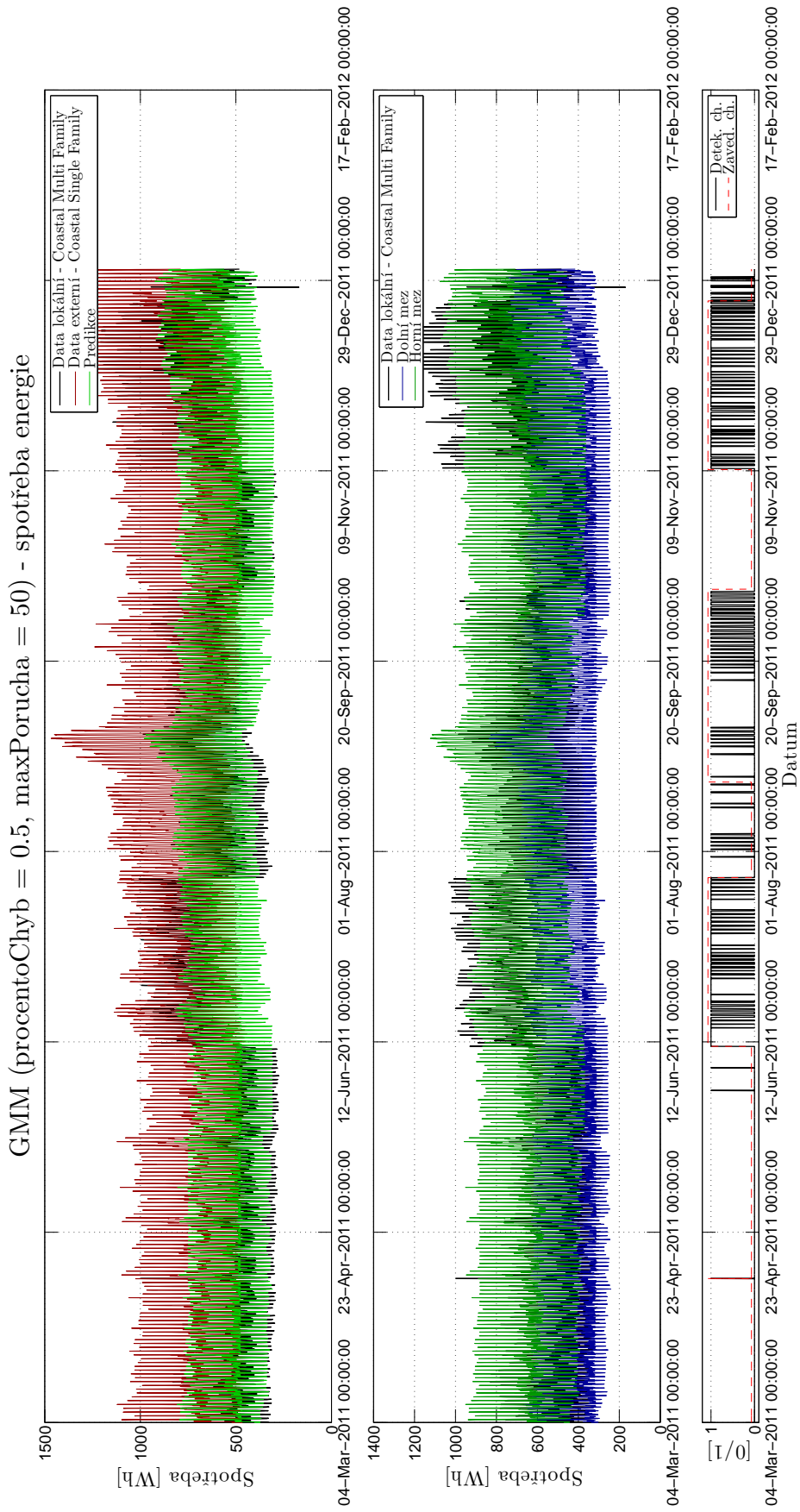
Obrázek 6.36: Detail výsledků GMM - teploty *Budihostice, Červený Újezd, Chloumek, ČZU* (skok).



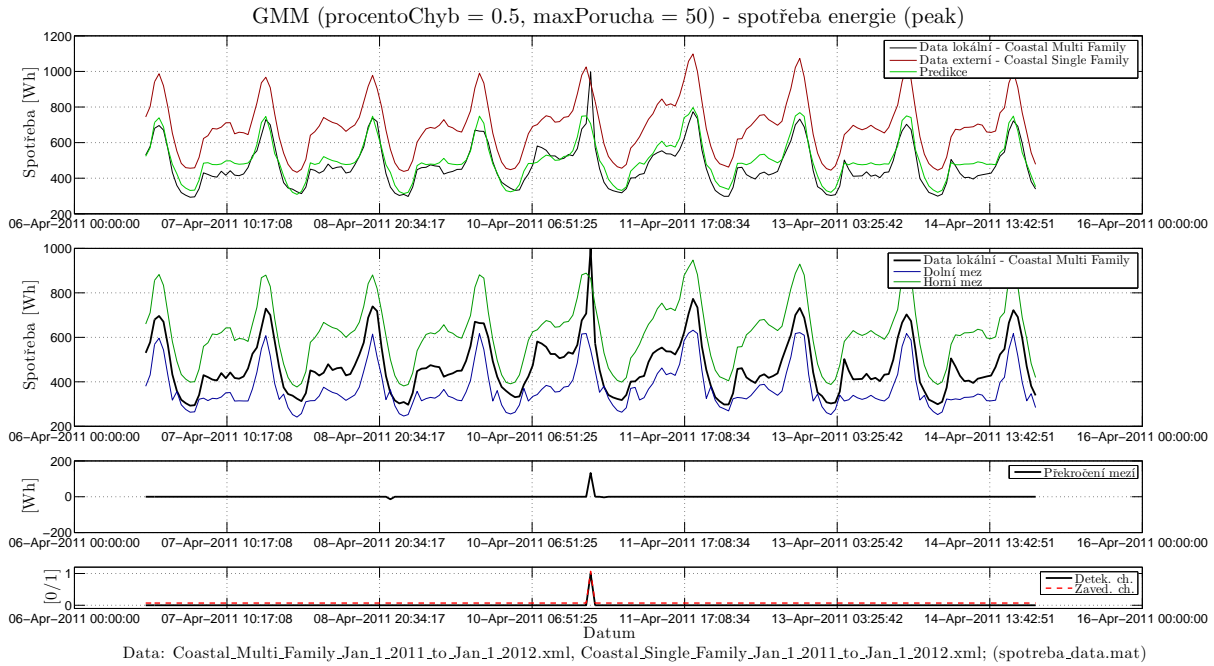
Obrázek 6.37: Detail výsledků GMM - teploty *Budihostice, Červený Újezd, Chloumek, ČZU* (drift).



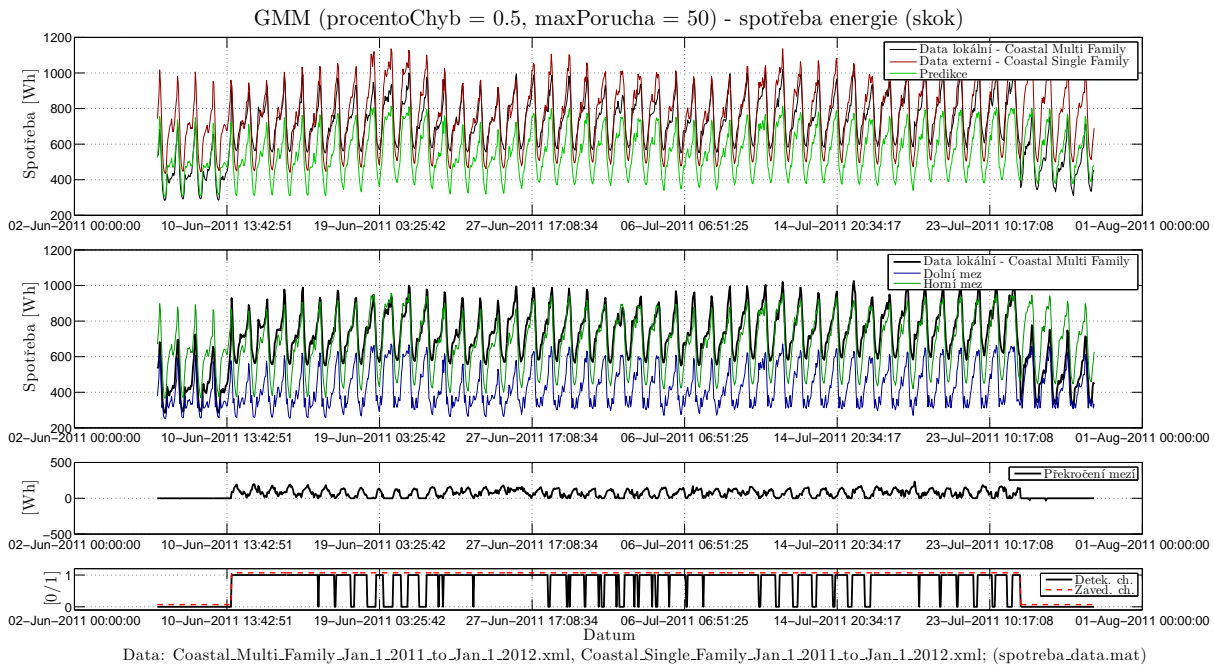
Obrázek 6.38: Detail výsledků GMM - teploty *Budihostice, Červený Újezd, Chloumek, ČZU* (gain).



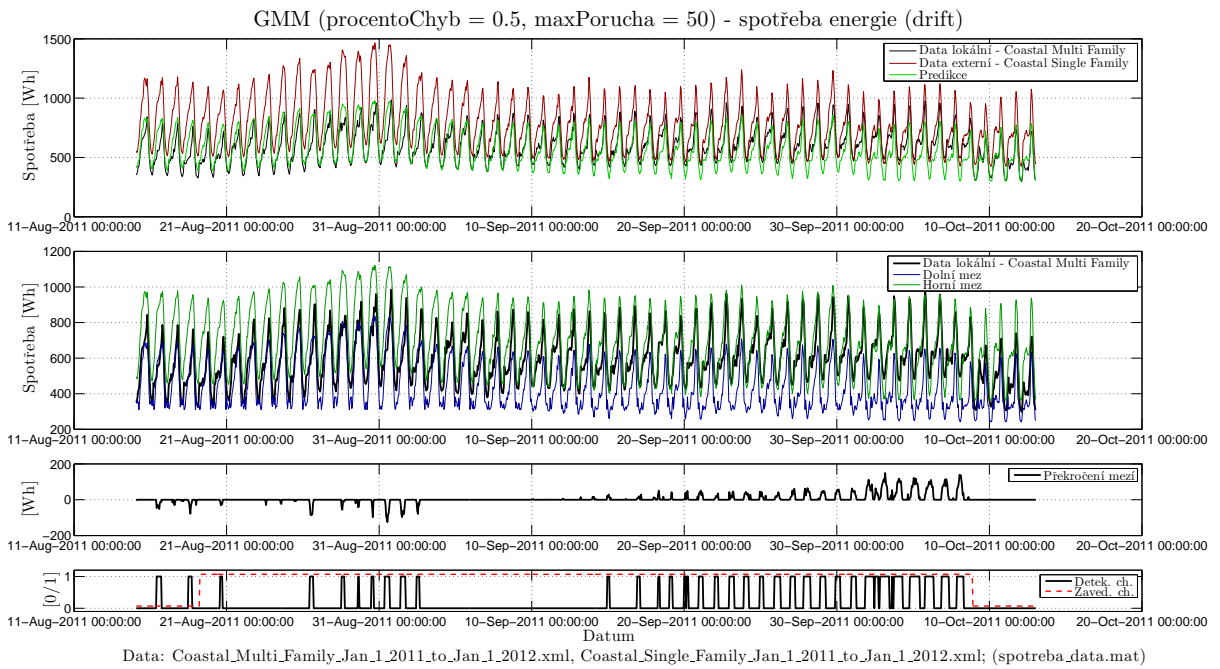
Obrázek 6.39: Výsledky GMM - spotřeba energie *Coastal Multi a Single Family*



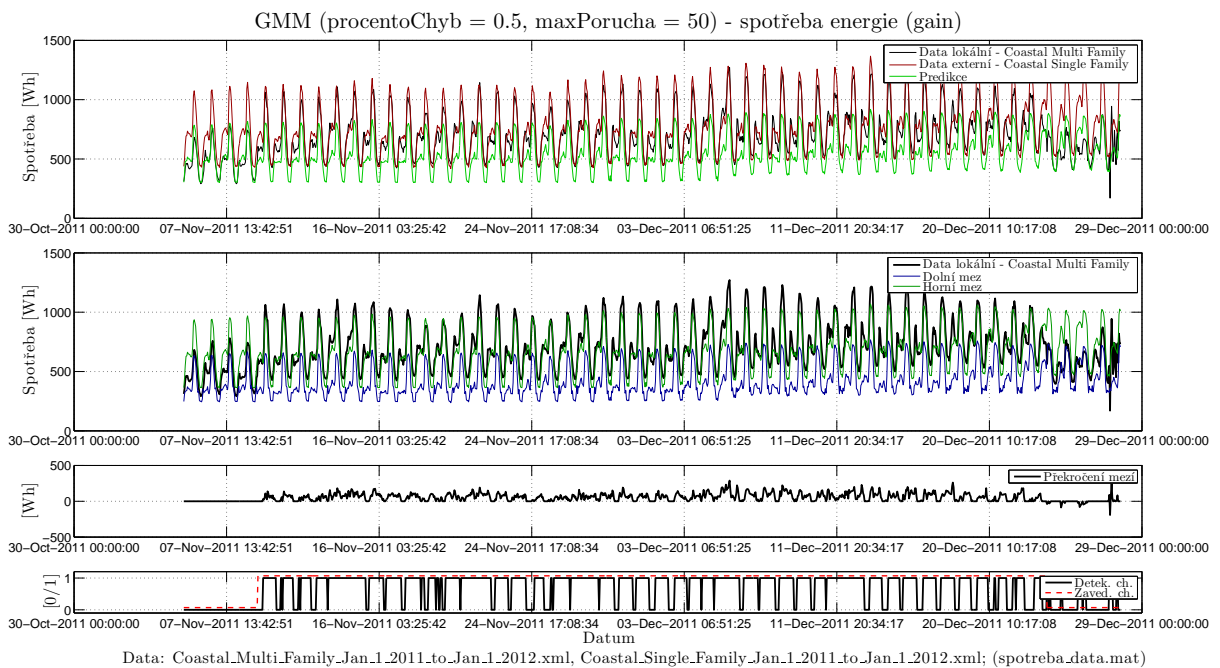
Obrázek 6.40: Detail výsledků GMM - spotřeba energie *Coastal Multi a Single Family* (peak).



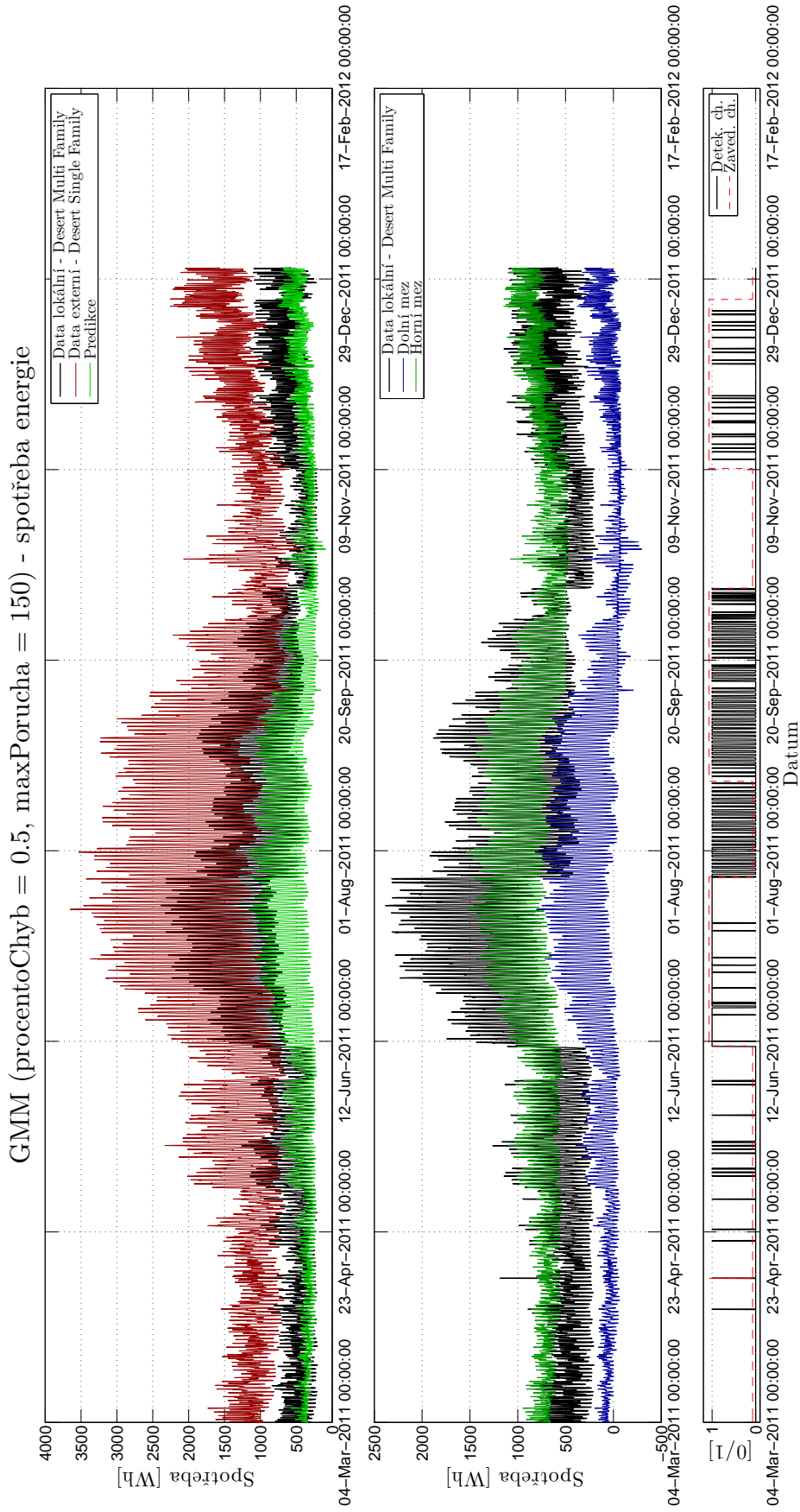
Obrázek 6.41: Detail výsledků GMM - spotřeba energie *Coastal Multi a Single Family* (skok).



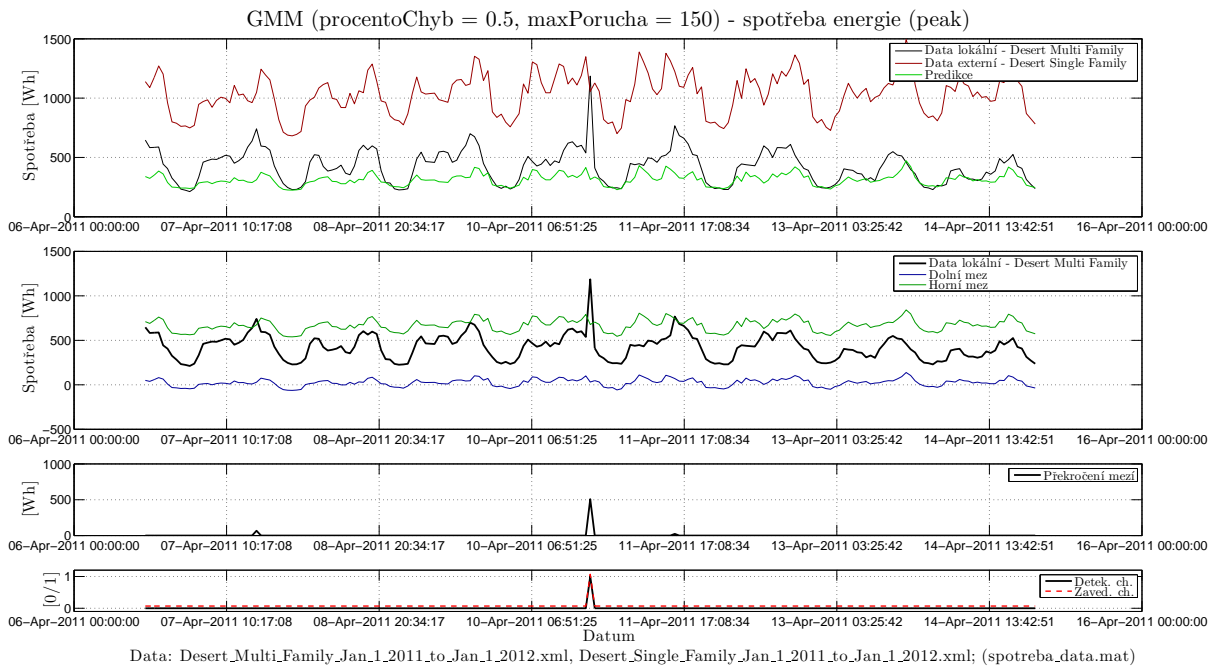
Obrázek 6.42: Detail výsledků GMM - spotřeba energie *Coastal Multi a Single Family* (drift).



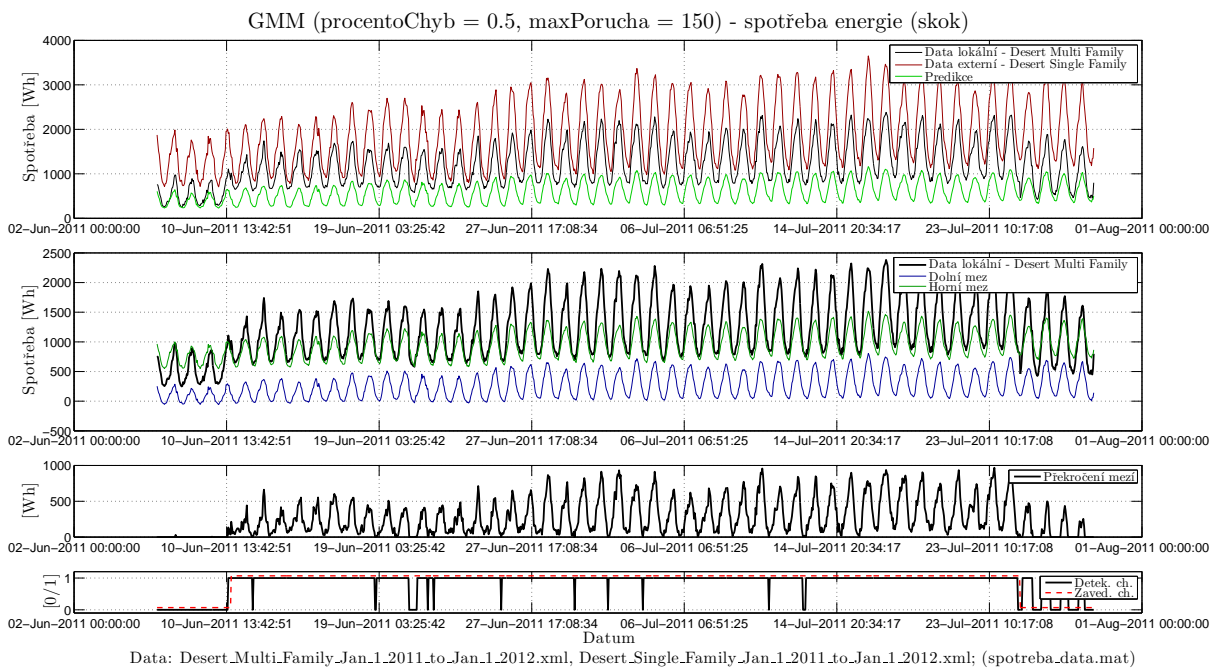
Obrázek 6.43: Detail výsledků GMM - spotřeba energie *Coastal Multi a Single Family* (gain).

Obrázek 6.44: Výsledky GMM - spotřeba energie *Desert Multi a Single*

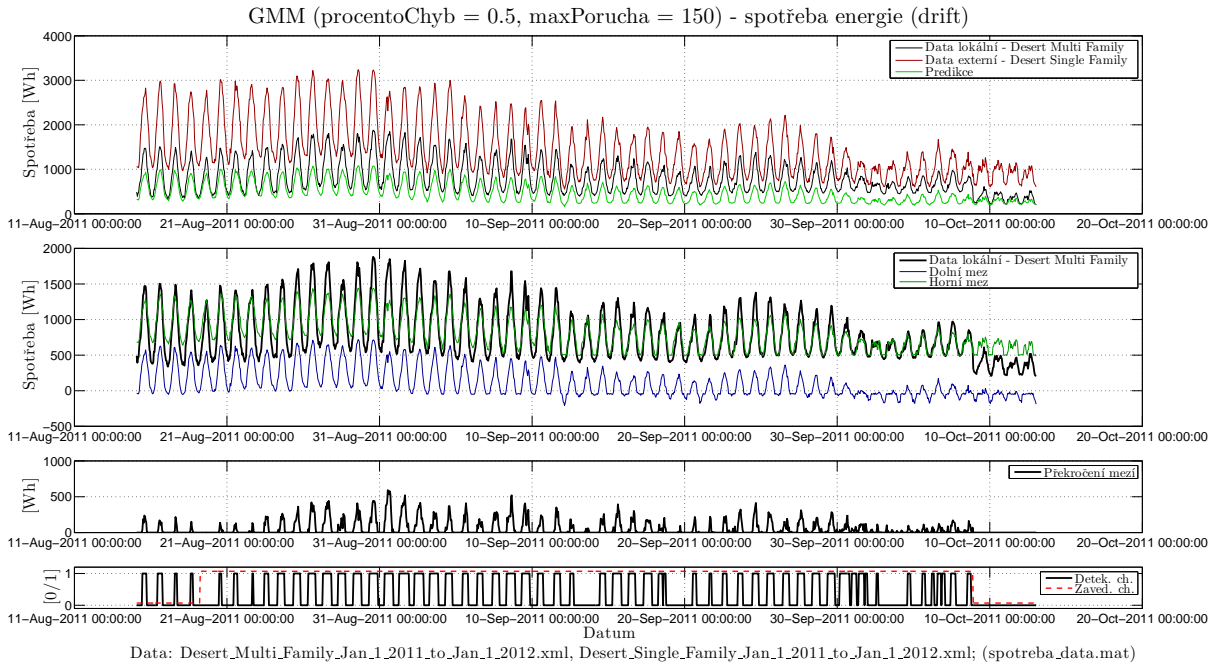
Familie



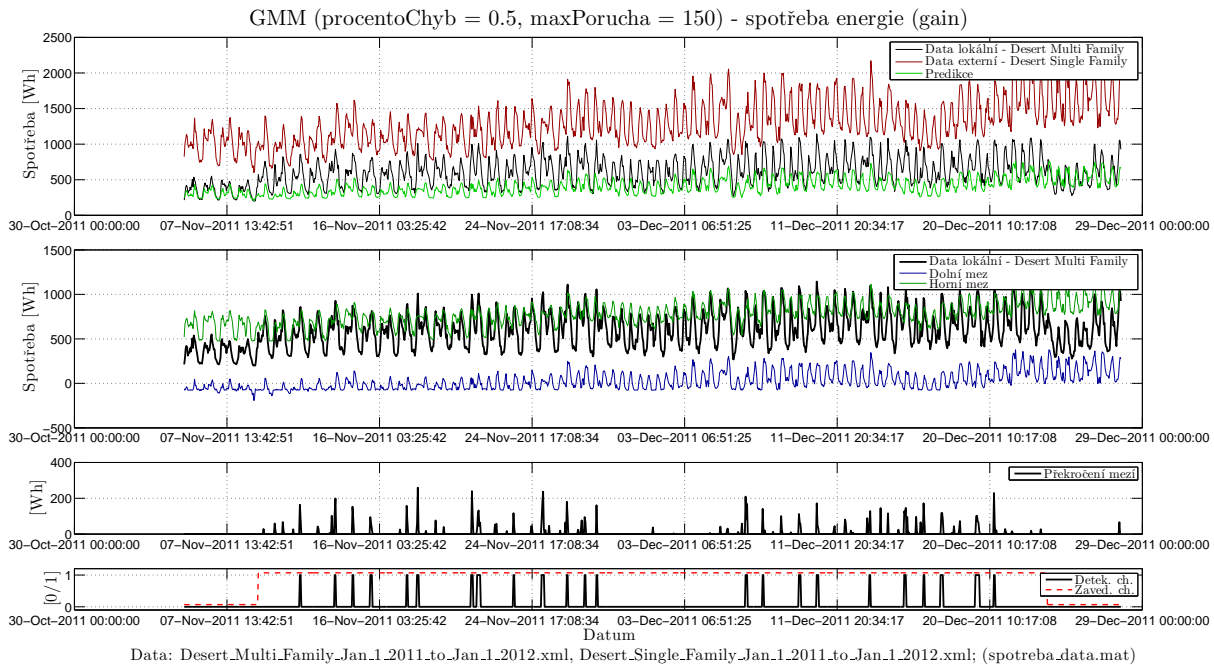
Obrázek 6.45: Detail výsledků GMM - spotřeba energie *Desert Multi a Single Family* (peak).



Obrázek 6.46: Detail výsledků GMM - spotřeba energie *Desert Multi a Single Family* (skok).



Obrázek 6.47: Detail výsledků GMM - spotřeba energie *Desert Multi a Single Family* (drift).



Obrázek 6.48: Detail výsledků GMM - spotřeba energie *Desert Multi a Single Family* (gain).

6.2.4 NARXnet (Nonlinear AutoRegressive neural network with eXternal input) uspořádání číslo 1

Výsledky této metody uvedené v tabulkách 6.10, 6.11, 6.12, B.7 a B.8 jsou v některých případech jiné, než bychom předpokládali. Tuto skutečnost připisují rozdílné kvalitě naučení modelů. Důvodem může být, že hodnoty pro trénování, validaci a testování neuronové sítě byly vybírány náhodně. Zlepšení by mohlo přinést případné implementování funkce, která by hledala vhodné nastavení pro učení modelu. Samotným problémem může také být velký rozdíl mezi daty určenými pro učení a testování modelu. Důsledek tohoto rozdílu je vidět na obrázku 6.59 dat spotřeby energie *Desert Multi a Single Family*, kde výstupem je špatná predikce, a proto zkusíme dále rozebírat.

Celkový pohled na testovaná data, stanovené meze a vyhodnocení je vidět na obrázcích 6.49, 6.54 a 6.59.

Z výsledků v tabulkách 6.11, 6.12, B.7, B.8 a na obrázku 6.54 vyplývá, že u detekce poruch u dat spotřeby energie je detekováno množství falešných poplachů. Detekce odlehlých hodnot je zachycena na obrázcích 6.50 a 6.55. Dále na obrázcích 6.51 a 6.56 jsou zobrazeny výsledky skokové změny hodnoty senzoru, která je velice často detekována správně. Detekce driftu je u této metody pomalá, ale na obrázku 6.52 je vidět, že v závěru je již drift detekován. Na obrázcích 6.53 a 6.58 vidíme detekci změny zesílení senzoru.

Tabulka 6.10: Výsledky NARXnet uspořádání číslo 1 - teploty *Budihostice, Červený Újezd, Chloumek, ČZU*.

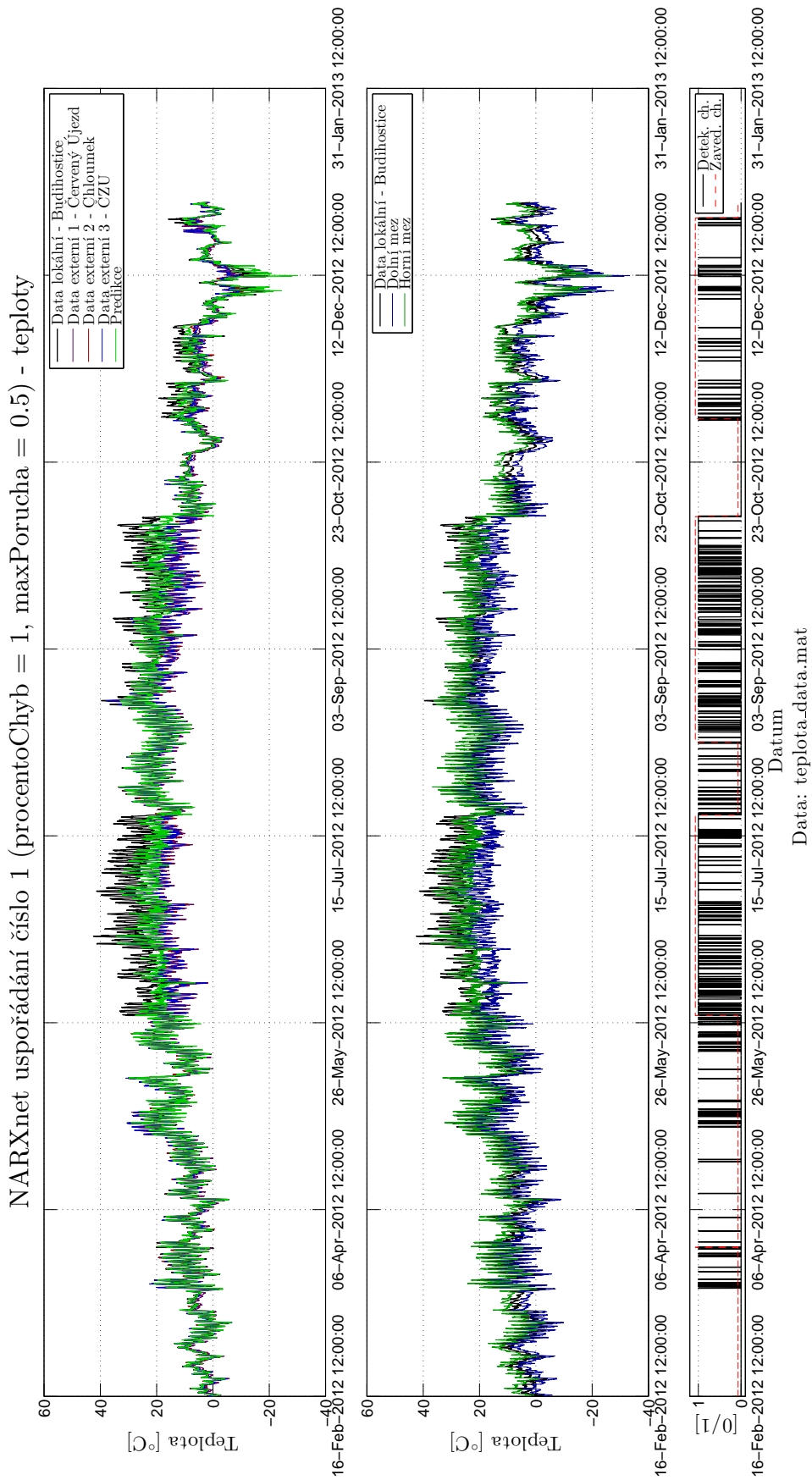
Procento chyb [%]	Max porucha [°C]	TP [-]	FP [-]	FN [-]	TN [-]	Specificita [%]	Senzitivita [%]
0,5	0,125	1518	104	6531	7183	98,57	18,86
	0,500	1290	119	6759	7168	98,37	16,03
	1,500	715	226	7334	7061	96,90	8,88
1,0	0,125	3978	442	4071	6845	93,93	49,42
	0,500	3524	195	4525	7092	97,32	43,78
	1,500	805	110	7244	7177	98,49	10,00
1,5	0,125	3478	174	4571	7113	97,61	43,21
	0,500	1754	29	6295	7258	99,60	21,79
	1,500	0	0	8049	7287	100,00	00,00

Tabulka 6.11: Výsledky NARXnet uspořádání číslo 1 - spotřeba energie
Coastal Multi a Single Family.

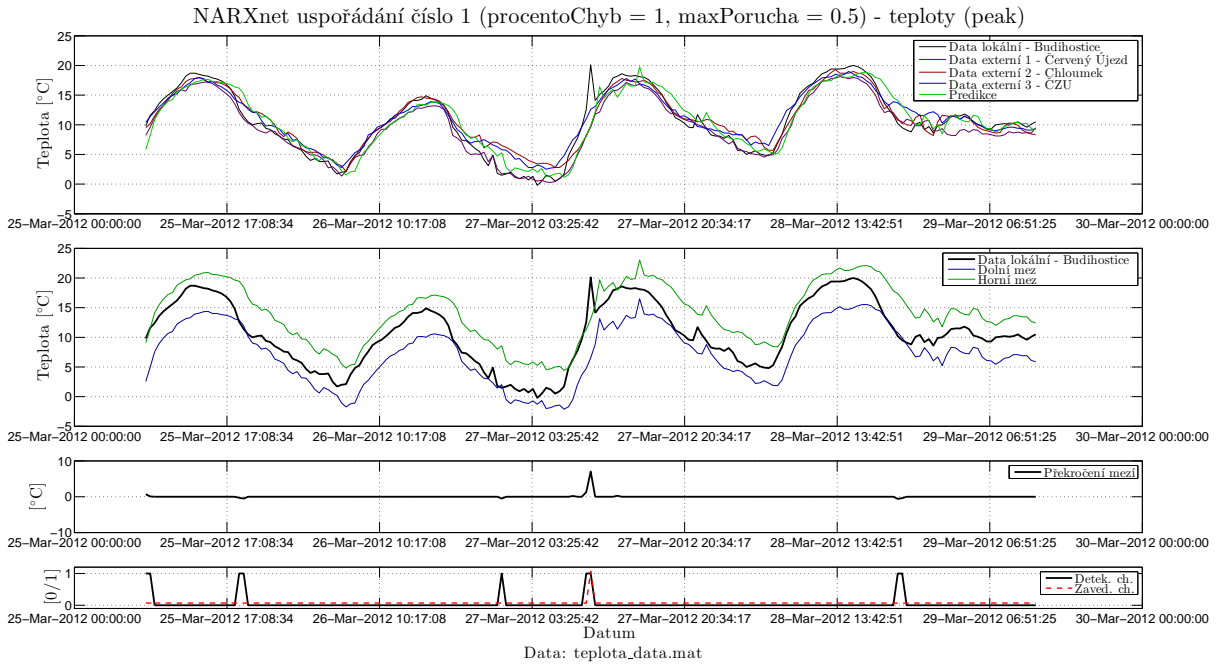
Procento chyb [%]	Max porucha [Wh]	TP [-]	FP [-]	FN [-]	TN [-]	Specifická [%]	Senzitivita [%]
0,5	12,5	960	975	2385	2952	75,17	28,70
	50,0	745	513	2600	3414	86,94	22,27
	150,0	305	275	3040	3652	93,00	9,12
1,0	12,5	552	571	2793	3356	85,46	16,50
	50,0	1050	525	2295	3402	86,63	31,39
	150,0	1587	391	1758	3536	90,04	47,44
1,5	12,5	1355	877	1990	3050	77,67	40,51
	50,0	1531	736	1814	3191	81,26	45,77
	150,0	645	629	2700	3298	83,98	19,28

Tabulka 6.12: Výsledky NARXnet uspořádání číslo 1 - spotřeba energie
Desert Multi a Single Family.

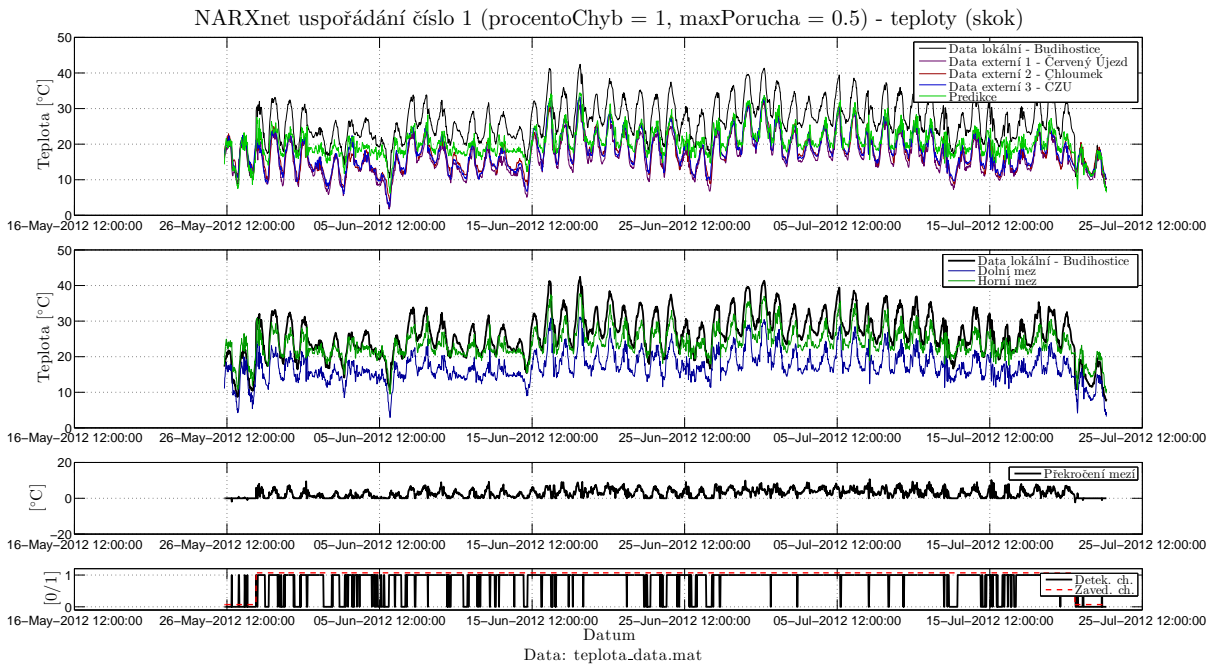
Procento chyb [%]	Max porucha [Wh]	TP [-]	FP [-]	FN [-]	TN [-]	Specifická [%]	Senzitivita [%]
0,5	12,5	1709	649	1636	3278	83,47	51,09
	50,0	1571	767	1774	3160	80,47	46,97
	150,0	1148	454	2197	3473	88,44	34,32
1,0	12,5	1200	738	2145	3189	81,21	35,87
	50,0	1677	770	1668	3157	80,39	50,13
	150,0	1382	562	1963	3365	85,69	41,32
1,5	12,5	2105	896	1240	3031	77,18	62,93
	50,0	1991	688	1354	3239	82,48	59,52
	150,0	1489	703	1856	3224	82,10	44,51



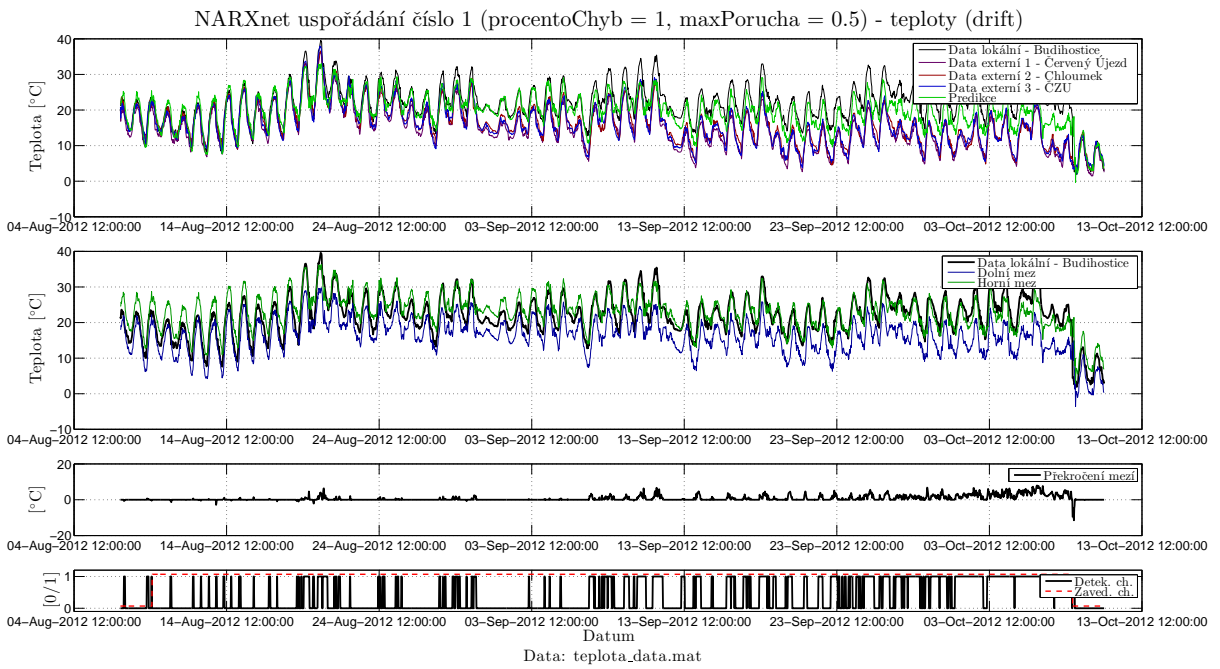
Obrázek 6.49: Výsledky NARXnet uspořádání číslo 1 - teploty *Budihostice*, *Červený Újezd*, *Chloumek*, *ČZU*.



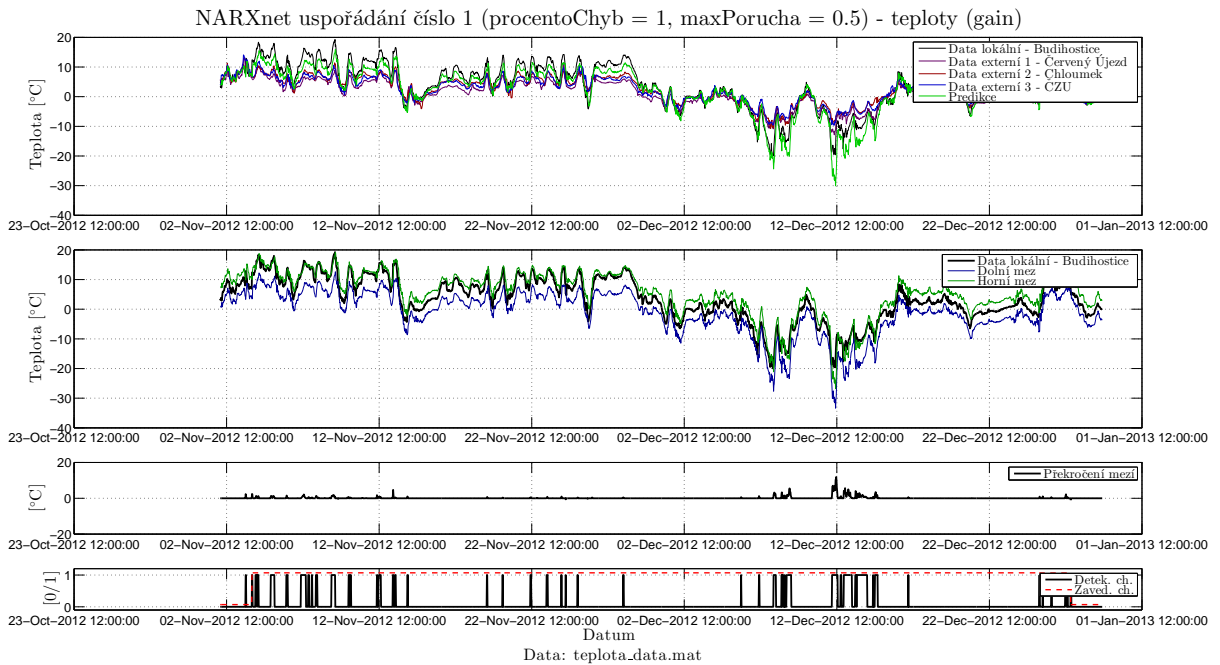
Obrázek 6.50: Detail výsledků NARXnet uspořádání číslo 1 - teploty *Budihostice*, *Červený Újezd*, *Chloumek*, *ČZU* (peak).



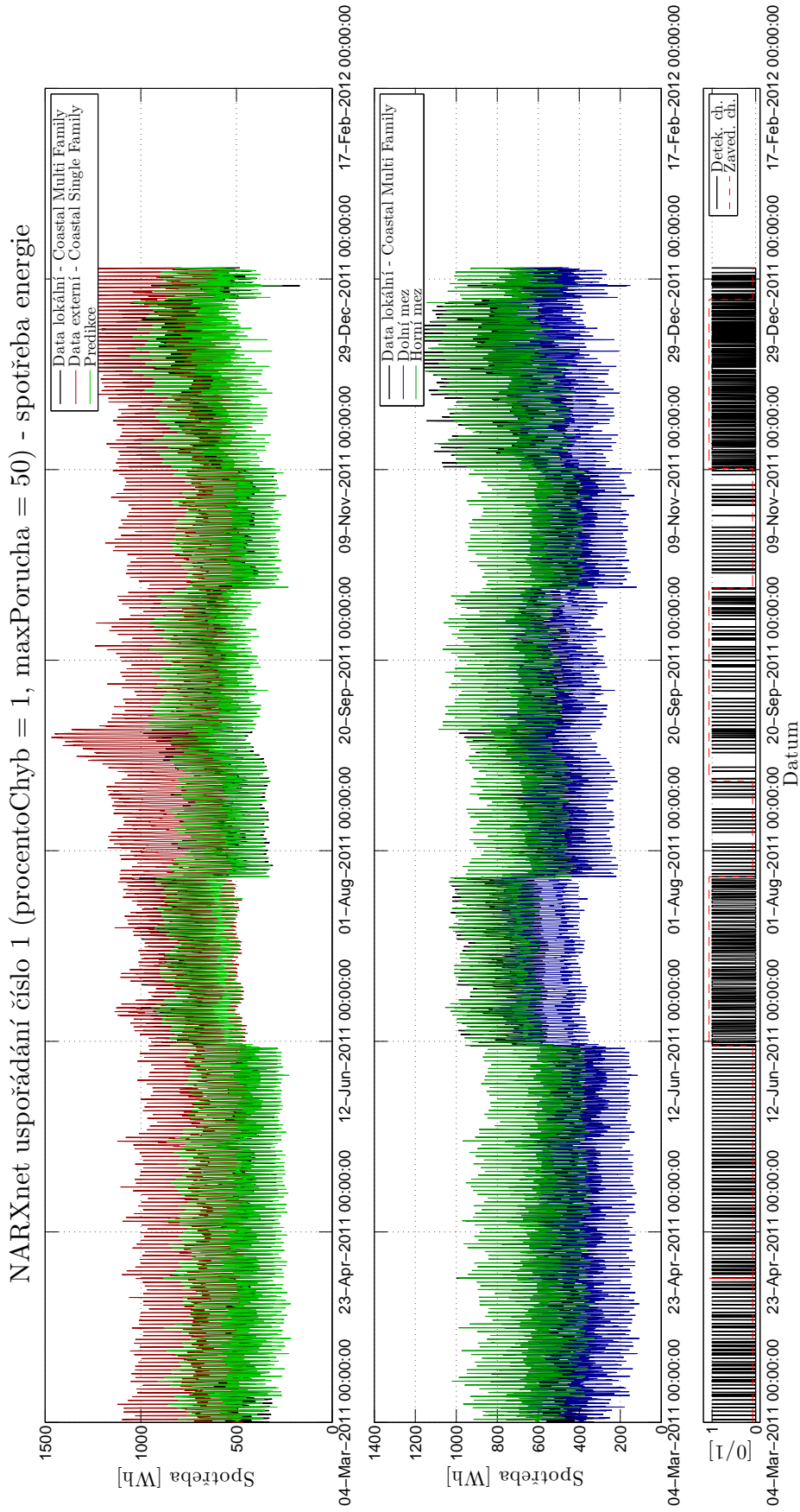
Obrázek 6.51: Detail výsledků NARXnet uspořádání číslo 1 - teploty *Budihostice*, *Červený Újezd*, *Chloumek*, *ČZU* (skok).



Obrázek 6.52: Detail výsledků NARXnet uspořádání číslo 1 - teploty *Budihovice, Červený Újezd, Chloumek, ČZU* (drift).

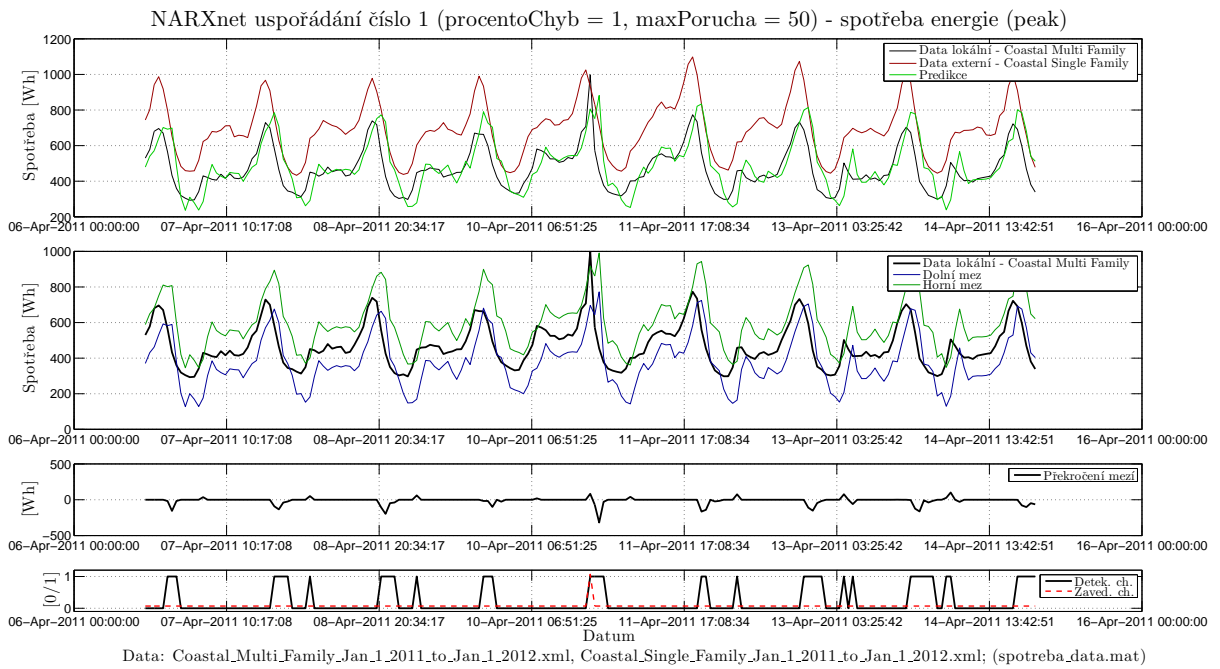


Obrázek 6.53: Detail výsledků NARXnet uspořádání číslo 1 - teploty *Budihovice, Červený Újezd, Chloumek, ČZU* (gain).

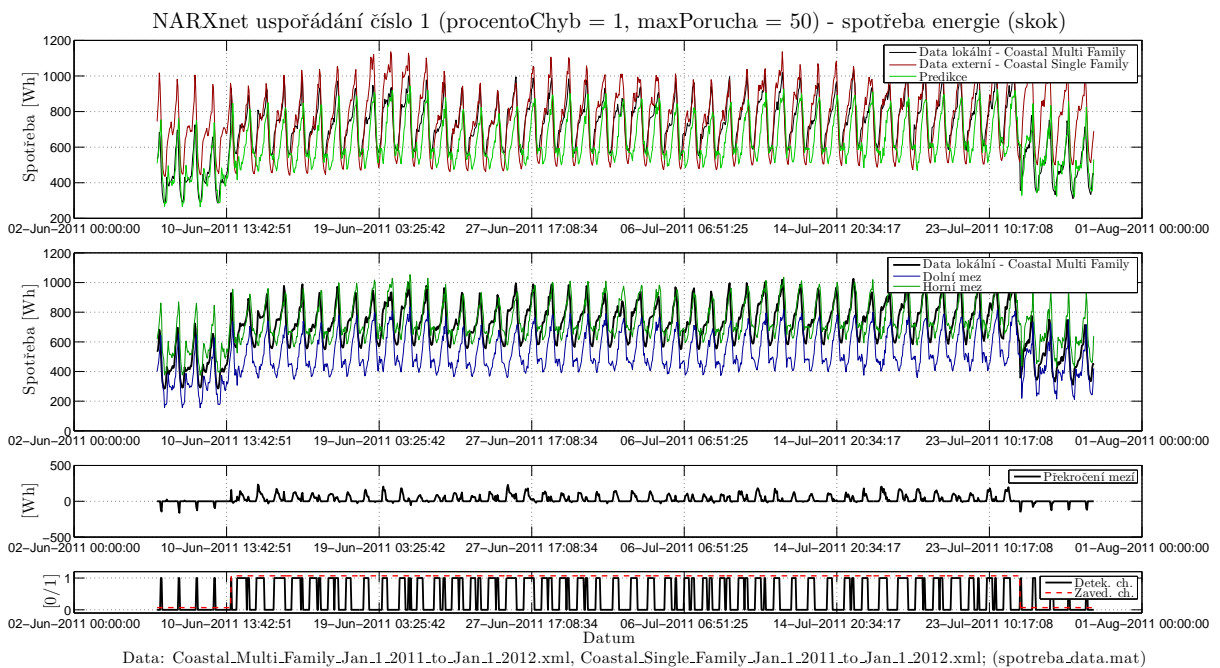


Obrázek 6.54: Výsledky NARXnet uspořádání číslo 1 - spotřeba energie

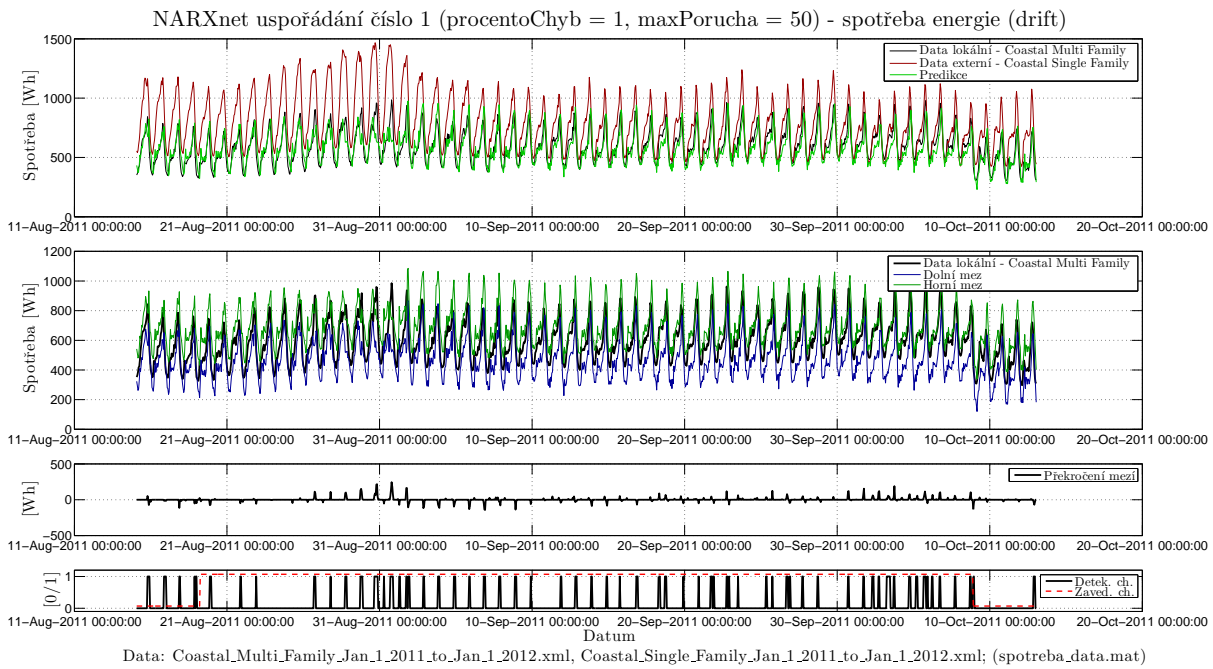
Coastal Multi a Single Family



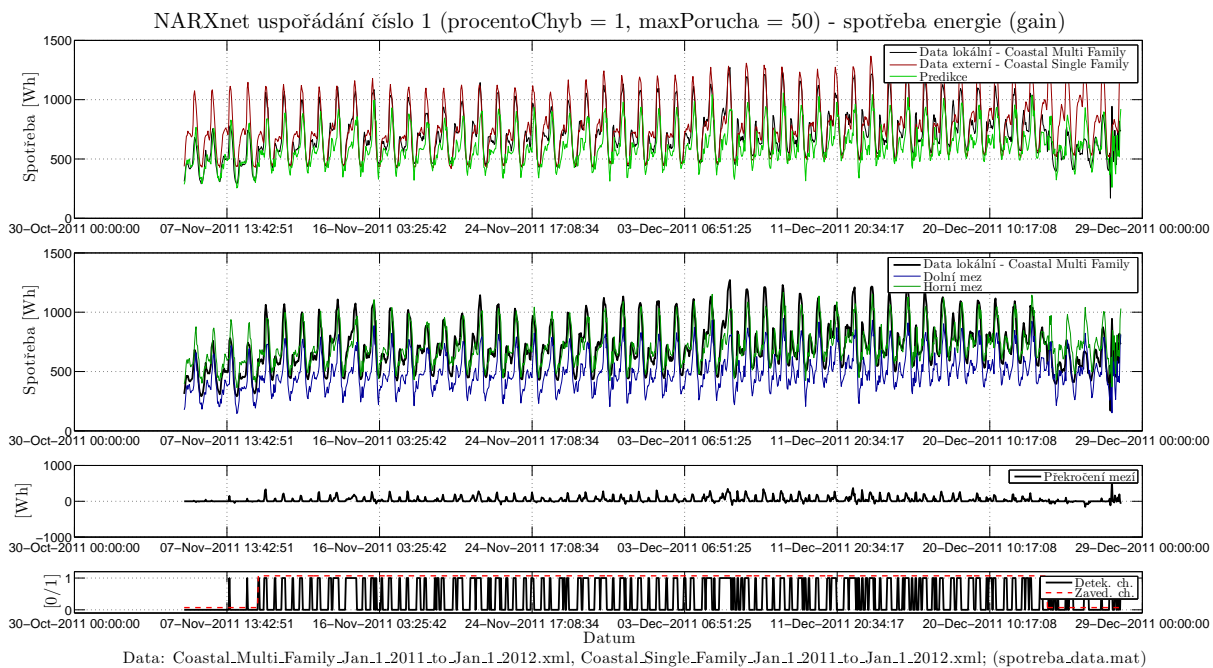
Obrázek 6.55: Detail výsledků NARXnet uspořádání číslo 1 - spotřeba energie *Coastal Multi a Single Family* (peak).



Obrázek 6.56: Detail výsledků NARXnet uspořádání číslo 1 - spotřeba energie *Coastal Multi a Single Family* (skok).

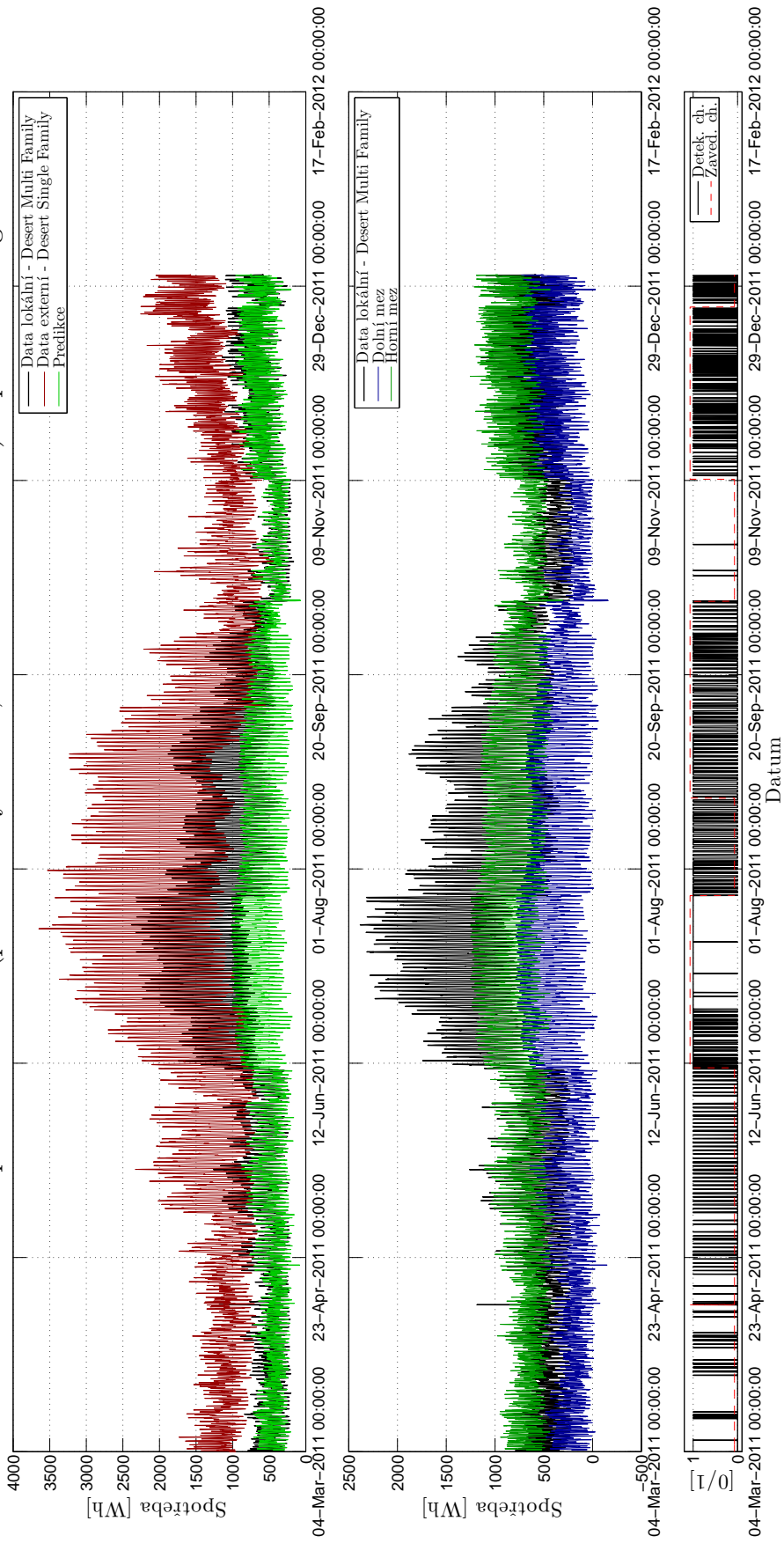


Obrázek 6.57: Detail výsledků NARXnet uspořádání číslo 1 - spotřeba energie *Coastal Multi a Single Family* (drift).



Obrázek 6.58: Detail výsledků NARXnet uspořádání číslo 1 - spotřeba energie *Coastal Multi a Single Family* (gain).

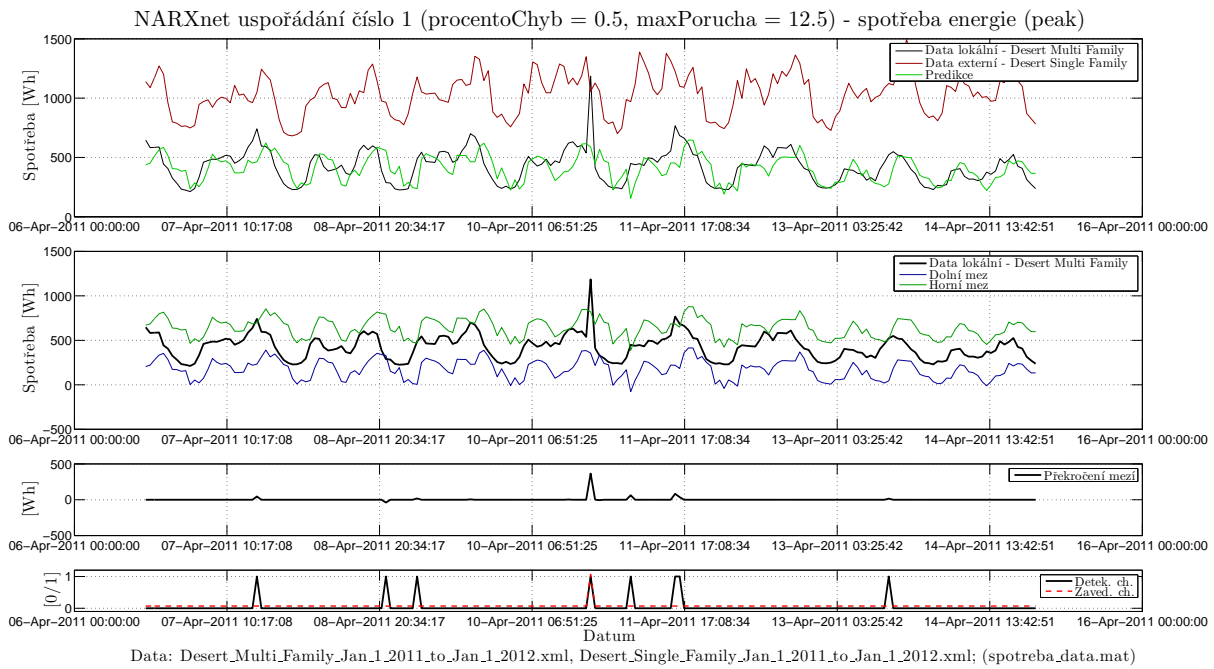
NARXnet uspořádání číslo 1 (procentoChyb = 0.5, maxPorucha = 12.5) - spotřeba energie



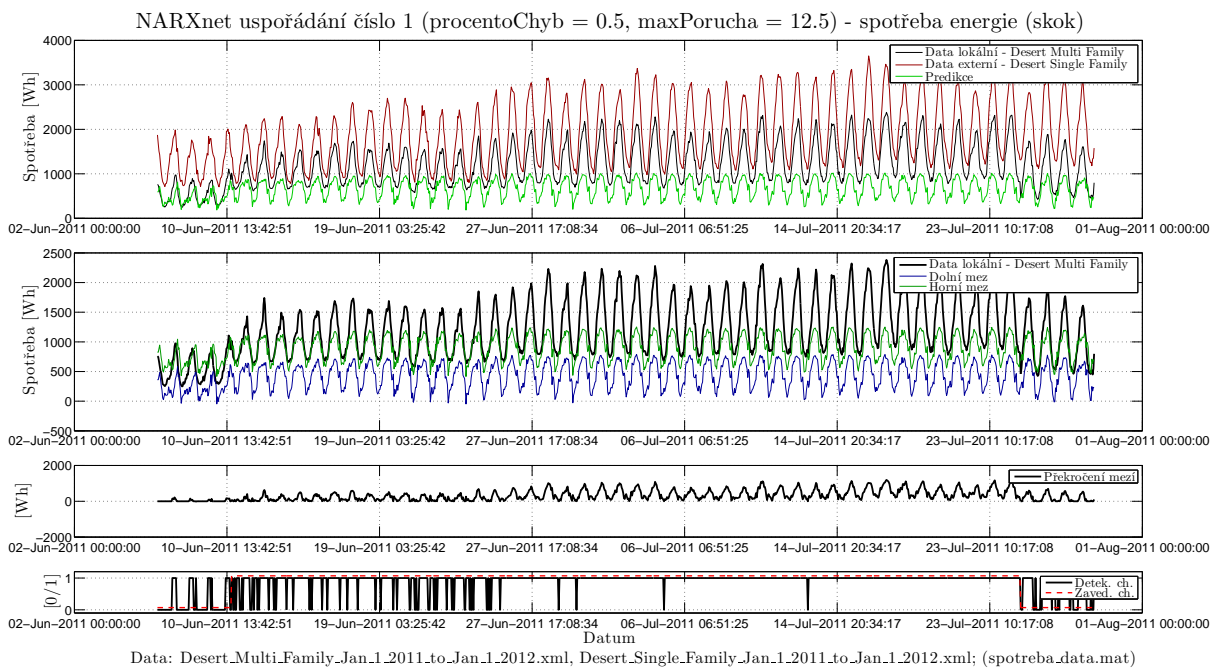
Data: Desert_Multi-Family-Jan.1.2011.to-Jan.1.2012.xml, Desert_Single-Family-Jan.1.2011.to-Jan.1.2012.xml; (spotřeba-data.mat)

Obrázek 6.59: Výsledky NARXnet uspořádání číslo 1 - spotřeba energie

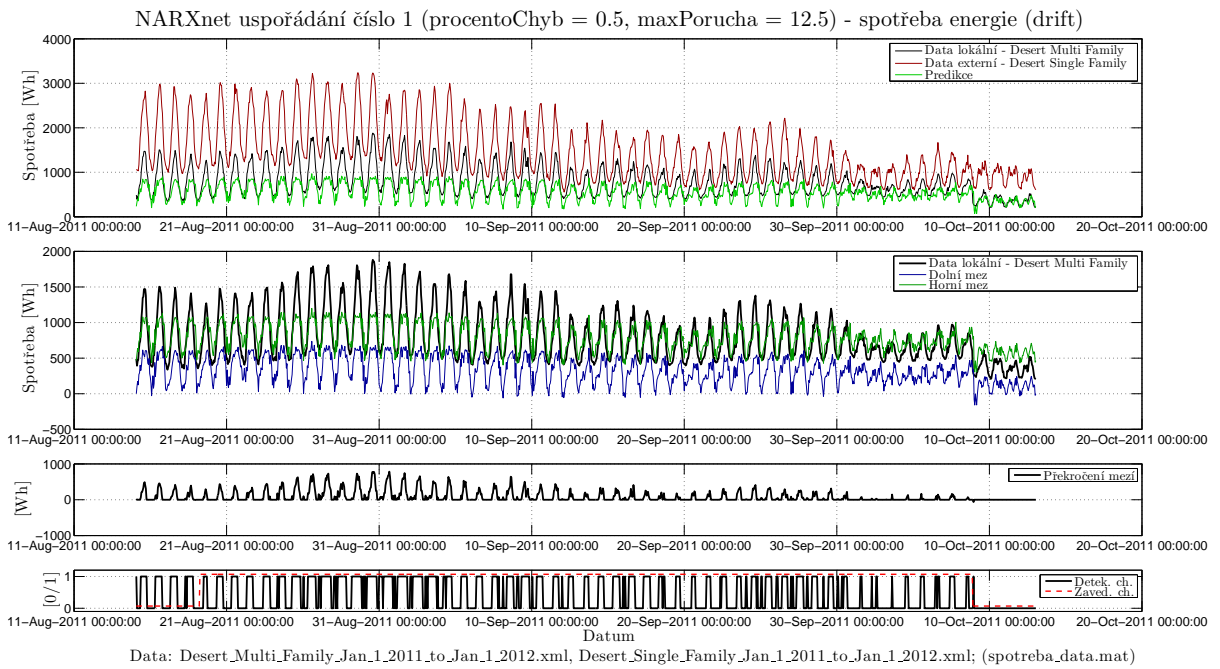
Desert Multi a Single Family



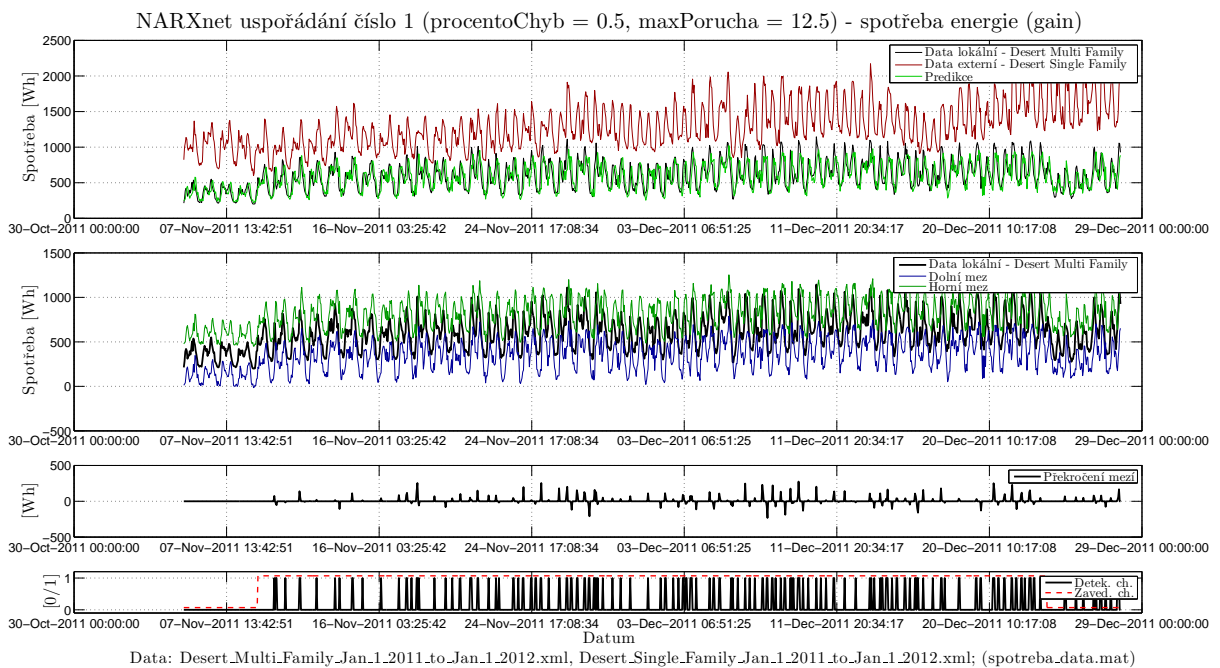
Obrázek 6.60: Detail výsledků NARXnet uspořádání číslo 1 - spotřeba energie *Desert Multi a Single Family* (peak).



Obrázek 6.61: Detail výsledků NARXnet uspořádání číslo 1 - spotřeba energie *Desert Multi a Single Family* (skok).



Obrázek 6.62: Detail výsledků NARXNET uspořádání číslo 1 - spotřeba energie *Desert Multi a Single Family* (drift).



Obrázek 6.63: Detail výsledků NARXnet uspořádání číslo 1 - spotřeba energie *Desert Multi a Single Family* (gain).

6.2.5 NARXnet (Nonlinear AutoRegressive neural network with eXternal input) uspořádání číslo 2

Z výsledků v tabulkách 6.13, 6.14, 6.15, B.9 a B.10 je patrné, že se výsledky občas liší v závislosti na kvalitě naučení modelu. Tato problematika byla již podrobně popsána v kapitole 6.2.4, a proto ji zde již nebudu rozebírat.

Na obrázcích 6.64, 6.69 a 6.74 je opět vidět náhled na testovaná data, stanovené meze a vyhodnocení, zda byla zavedená chyba detekována. U dat spotřeby energie je detekováno množství falešných poplachů.

Detekce odlehle hodnoty pro jednotlivá data je zobrazena na obrázcích 6.65, 6.70, ze kterých je vidět, že tato hodnota je správně detekována. Na obrázku 6.66 s teplotními daty a obrázku 6.71 dat spotřeby energie je patrné, že skoková změna hodnoty senzoru je úspěšně detekována. Obrázek 6.67 a 6.72 zobrazuje téměř příkladnou detekci driftu senzoru. Detekovaná změna zesílení (gain) senzoru je na obrázcích 6.68 a 6.73, ze kterých je vidět, že chyba je z velké části správně odhalena.

Tabulka 6.13: Výsledky NARXnet uspořádání číslo 2 - teploty *Budihostice, Červený Újezd, Chloumek, ČZU*.

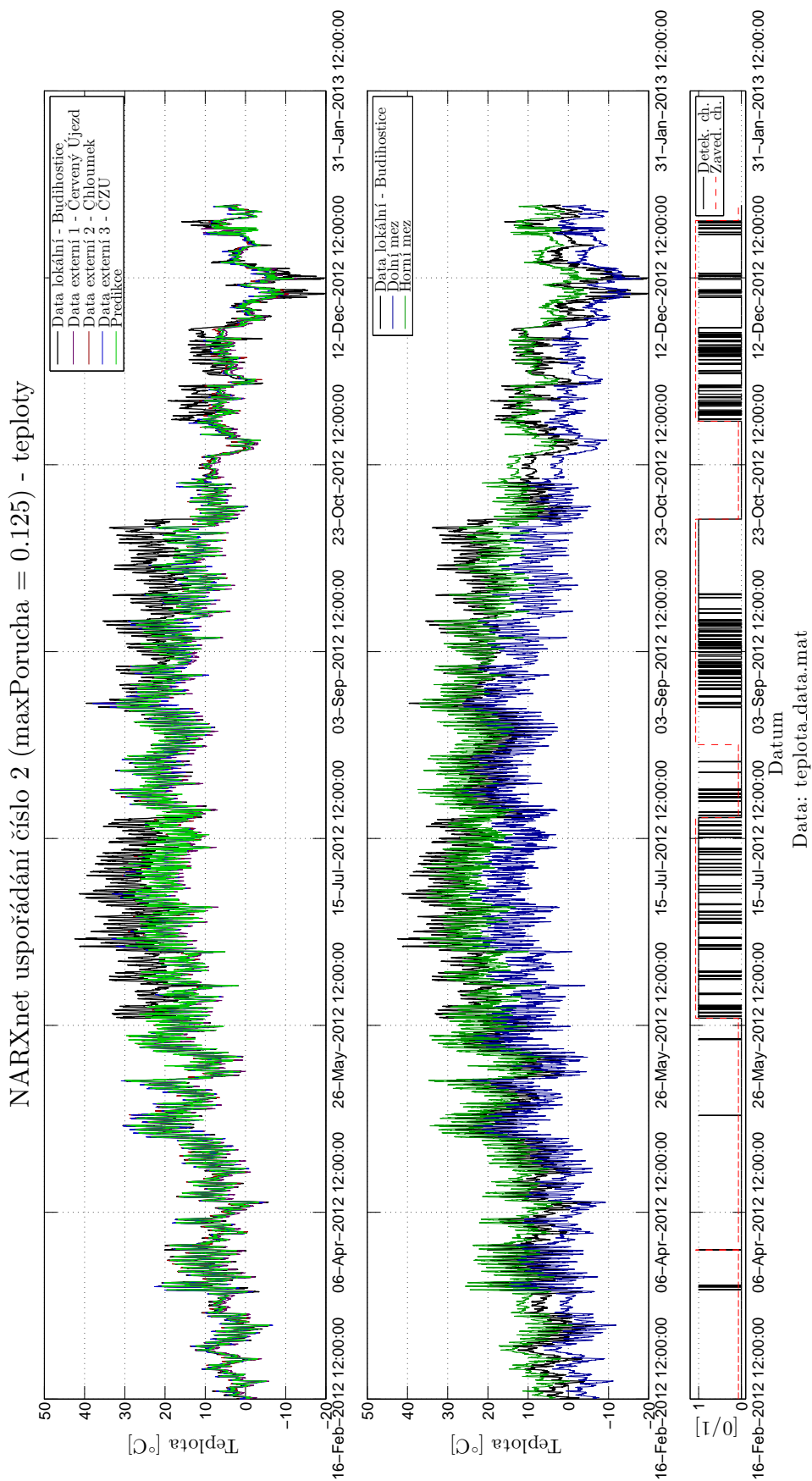
Max porucha [°C]	TP [-]	FP [-]	FN [-]	TN [-]	Specificita [%]	Senzitivita [%]
0,125	4791	43	3258	7244	99,41	59,52
0,500	5995	240	2054	7047	96,71	74,48
1,500	5365	166	2684	7121	97,72	66,65

Tabulka 6.14: Výsledky NARXnet uspořádání číslo 2 - spotřeba energie *Coastal Multi a Single Family*.

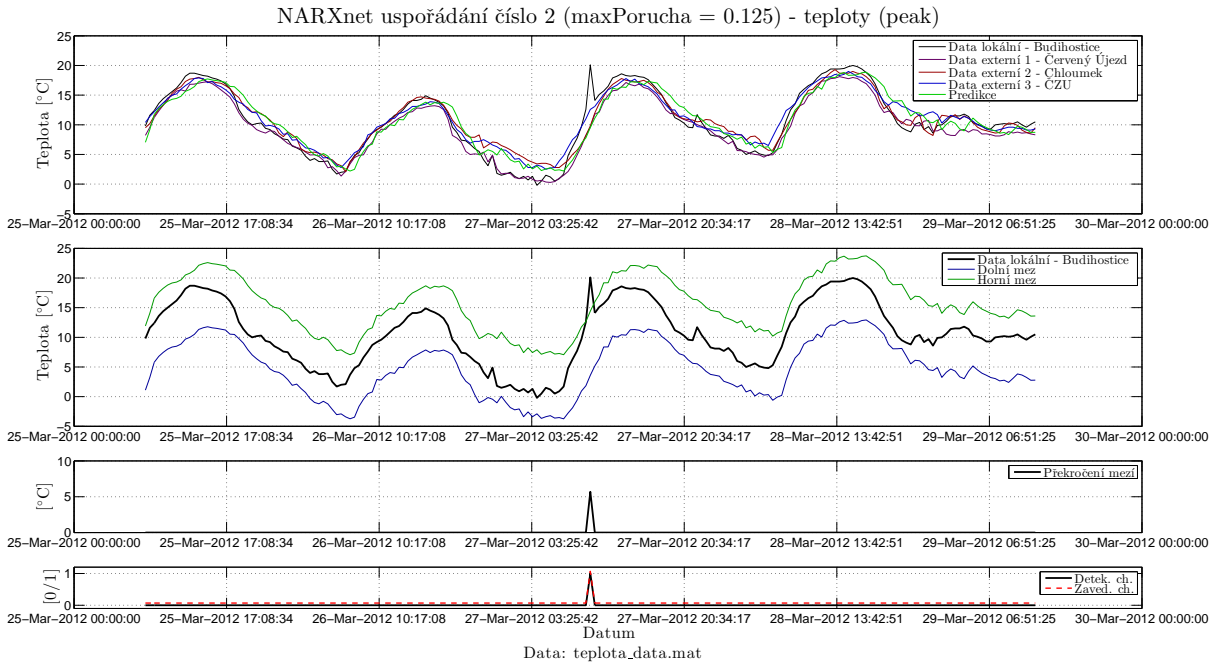
Max porucha [Wh]	TP [-]	FP [-]	FN [-]	TN [-]	Specificita [%]	Senzitivita [%]
12,5	1903	225	1442	3702	94,27	56,89
50,0	1381	94	1964	3833	97,61	41,29
150,0	861	508	2484	3419	87,06	25,74

Tabulka 6.15: Výsledky NARXnet uspořádání číslo 2 - spotřeba energie
Desert Multi a Single Family.

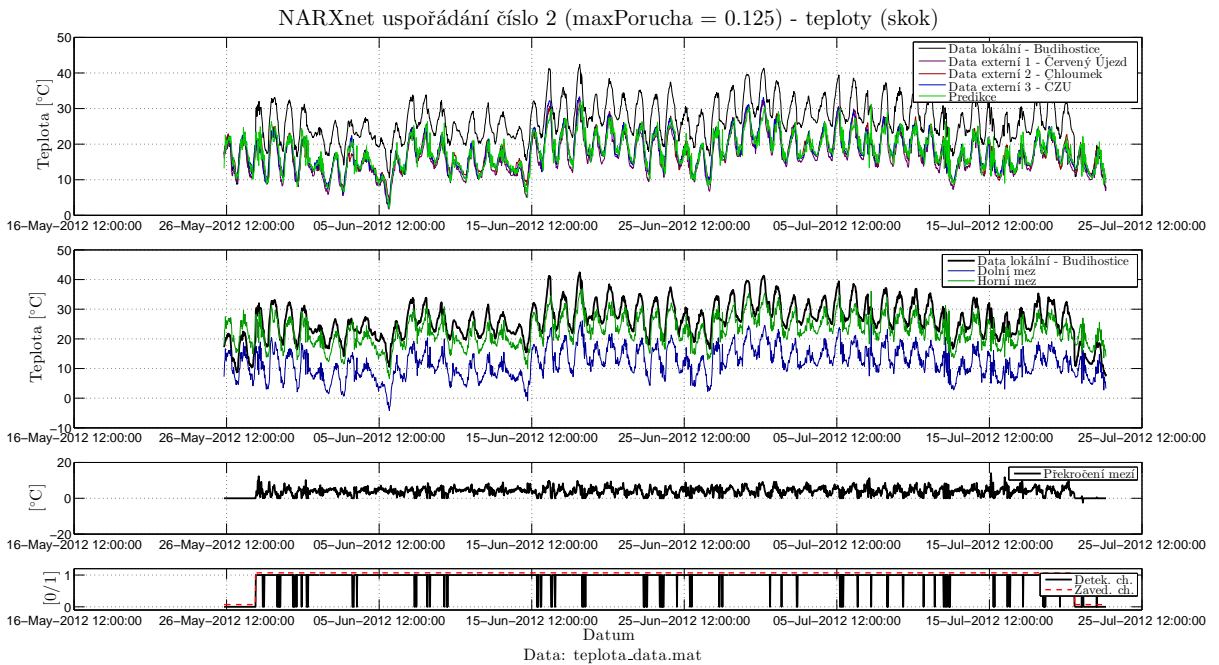
Max porucha [Wh]	TP [-]	FP [-]	FN [-]	TN [-]	Specificita [%]	Senzitivita [%]
12,5	1482	749	1863	3178	80,93	44,30
50,0	1644	1153	1701	2774	70,64	49,15
150,0	1898	1199	1447	2728	69,47	56,74



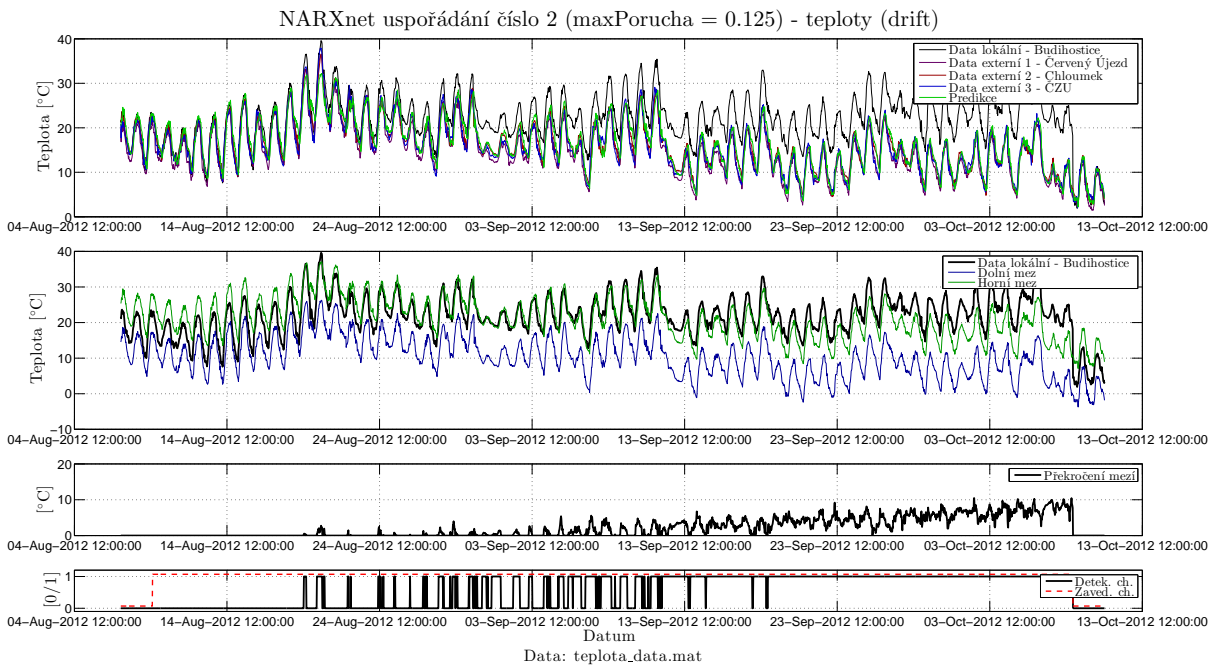
Obrázek 6.64: Výsledky NARXnet uspořádání číslo 2 - teploty *Budíhošovice*, *Červený Újezd*, *Chloumek*, *ČZU*.



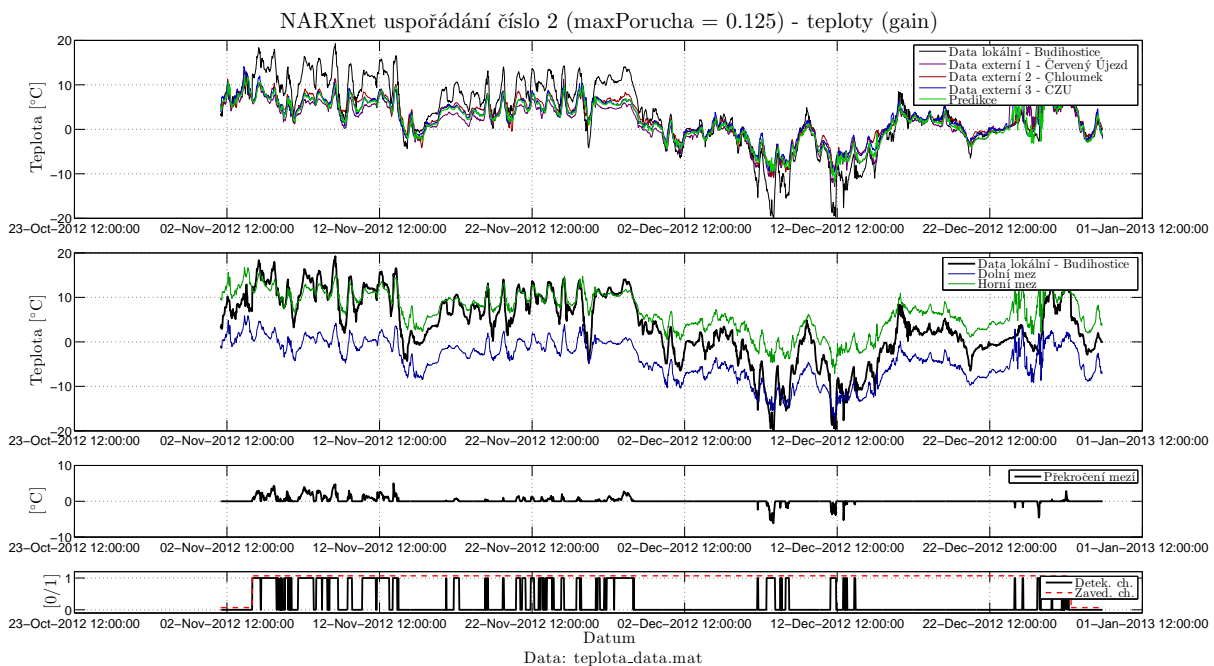
Obrázek 6.65: Detail výsledků NARXnet uspořádání číslo 2 - teploty *Budihostice*, *Červený Újezd*, *Chloumek*, *ČZU* (peak).



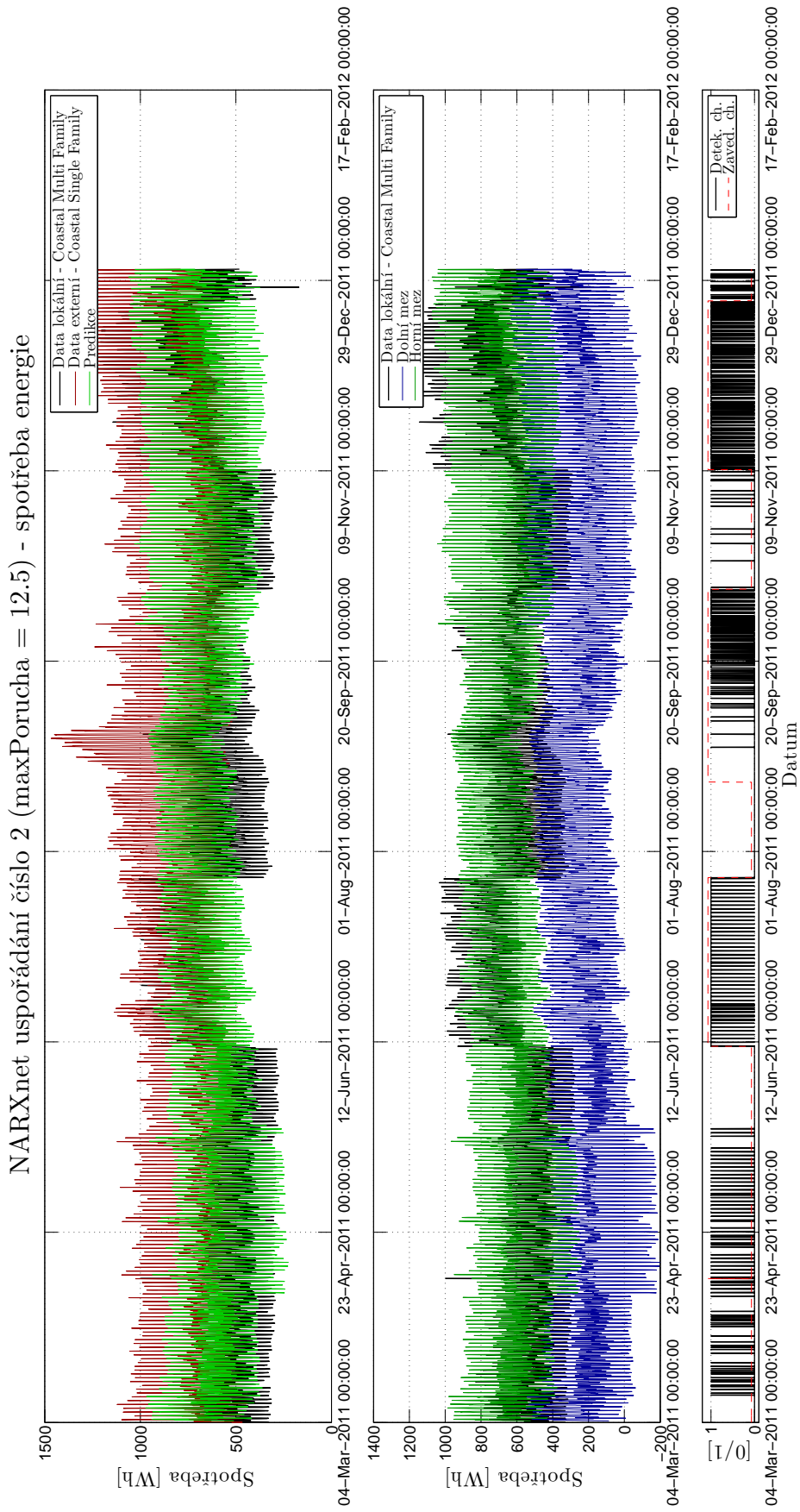
Obrázek 6.66: Detail výsledků NARXnet uspořádání číslo 2 - teploty *Budihostice*, *Červený Újezd*, *Chloumek*, *ČZU* (skok).



Obrázek 6.67: Detail výsledků NARXnet uspořádání číslo 2 - teploty *Budihostice*, *Červený Újezd*, *Chloumek*, *ČZU* (drift).

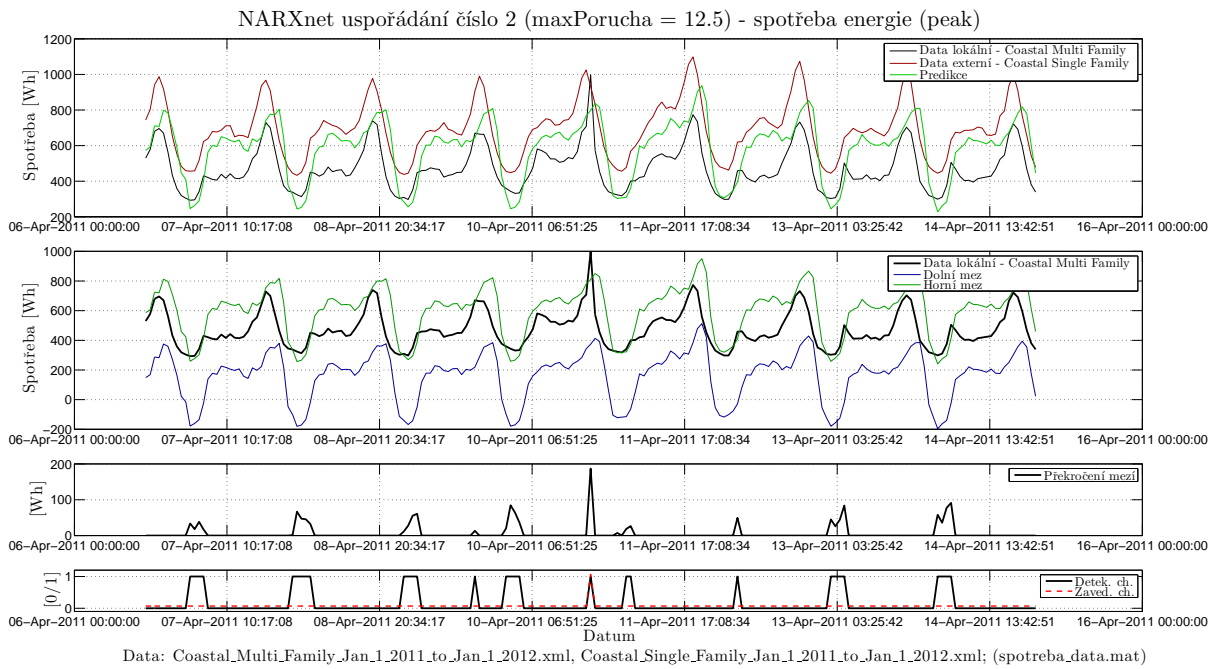


Obrázek 6.68: Detail výsledků NARXnet uspořádání číslo 2 - teploty *Budihostice*, *Červený Újezd*, *Chloumek*, *ČZU* (gain).

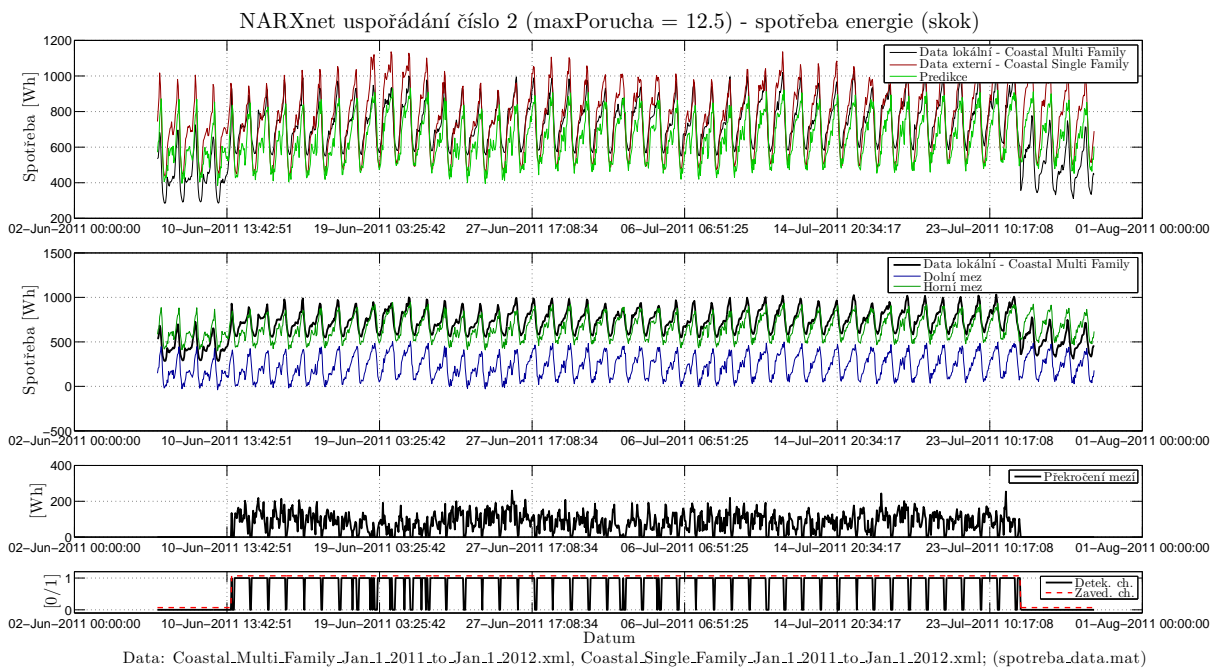


Obrázek 6.69: Výsledky NARXnet uspořádání číslo 2 - spotřeba energie

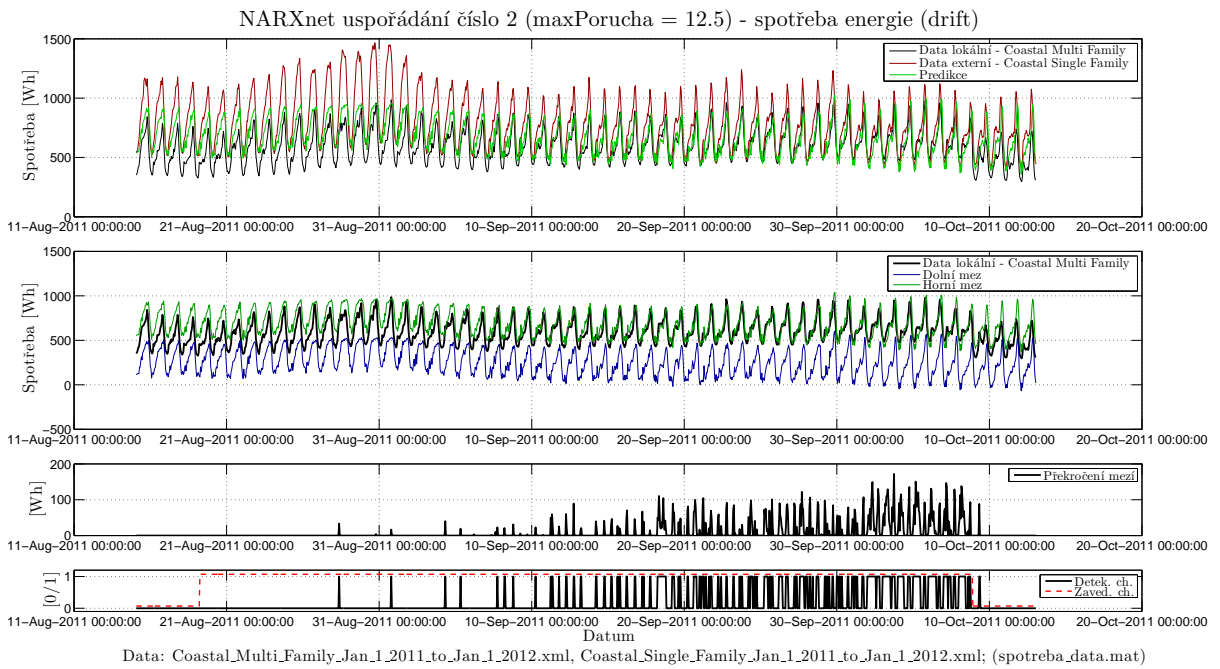
Coastal Multi a Single Family



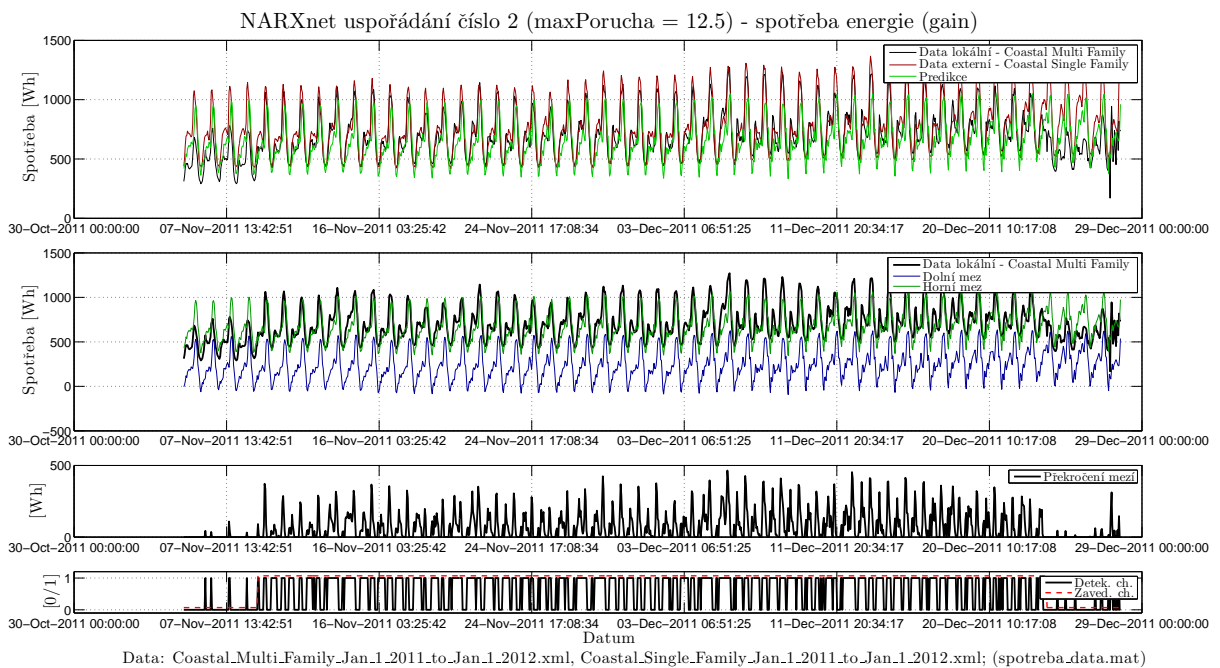
Obrázek 6.70: Detail výsledků NARXnet uspořádání číslo 2 - spotřeba energie *Coastal Multi a Single Family* (peak).



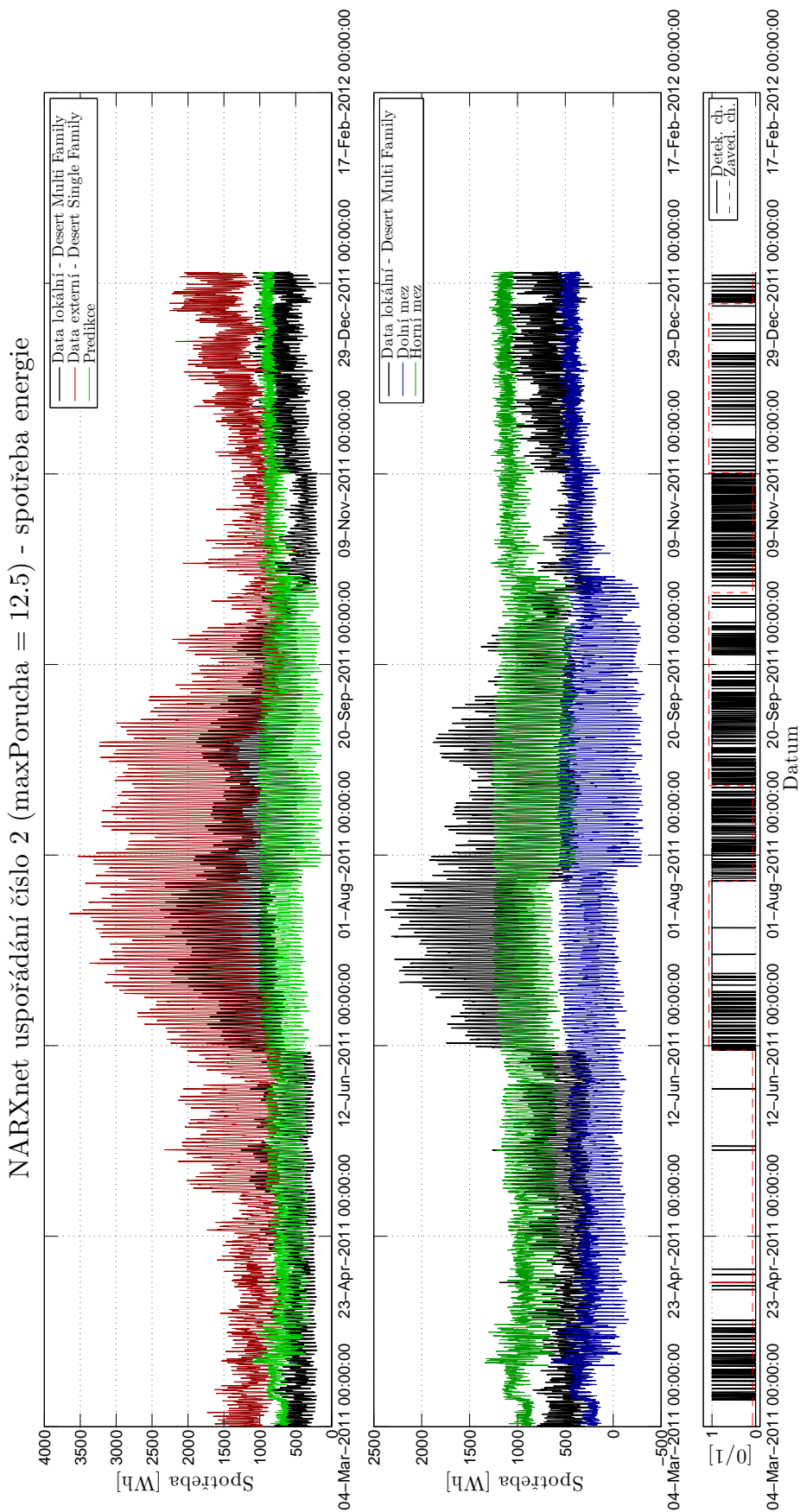
Obrázek 6.71: Detail výsledků NARXnet uspořádání číslo 2 - spotřeba energie *Coastal Multi a Single Family* (skok).



Obrázek 6.72: Detail výsledků NARXnet uspořádání číslo 2 - spotřeba energie *Coastal Multi a Single Family* (drift).



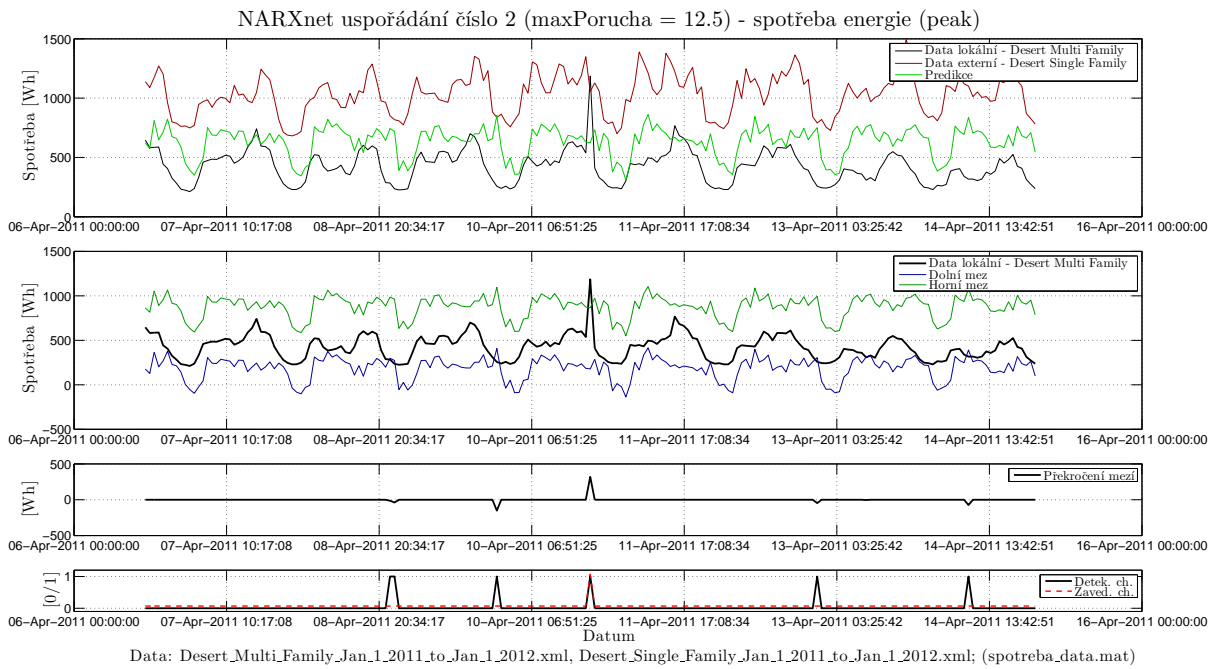
Obrázek 6.73: Detail výsledků NARXnet uspořádání číslo 2 - spotřeba energie *Coastal Multi a Single Family* (gain).



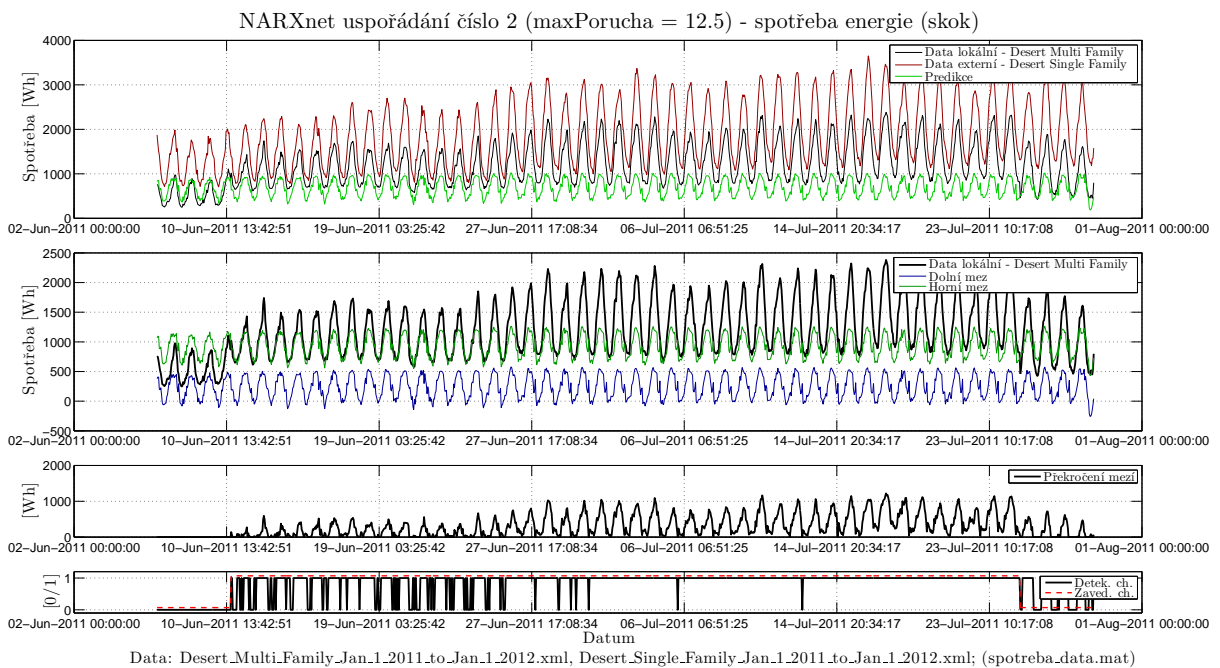
Data: Desert_Multi_Family_Jan_1_2011_to_Jan_1_2012.xml, Desert_Single_Family_Jan_1_2011_to_Jan_1_2012.xml; (spotřeba_data.mat)

Obrázek 6.74: Výsledky NARXnet uspořádání číslo 2 - spotřeba energie

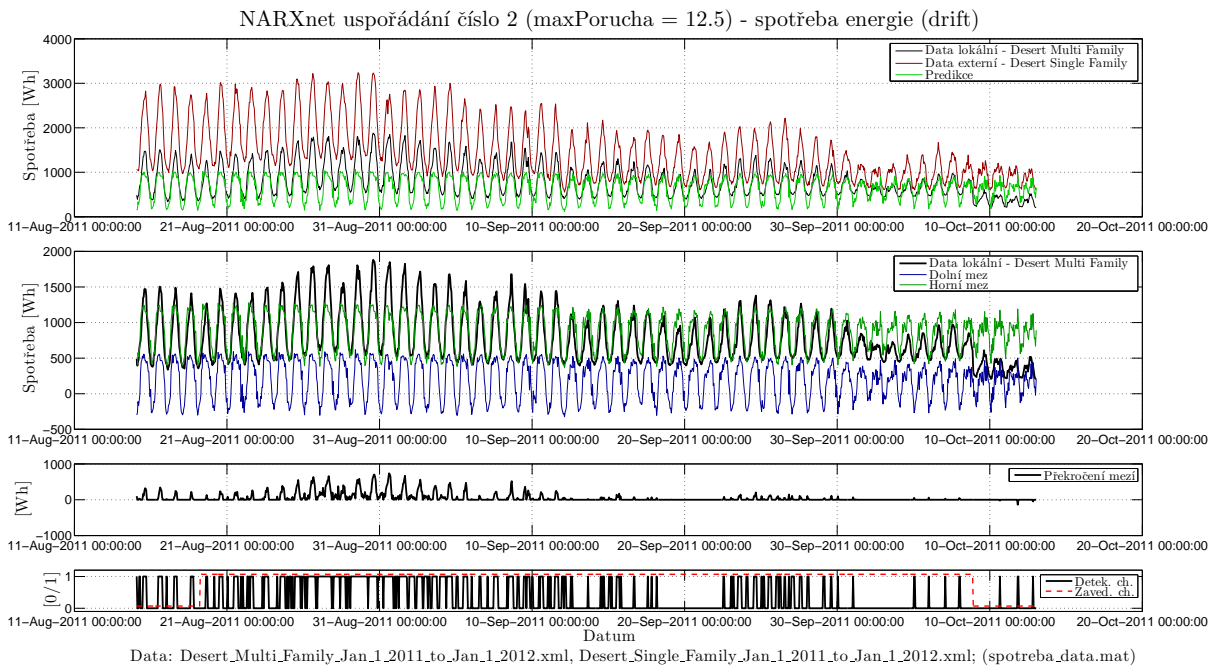
Desert Multi a Single Family



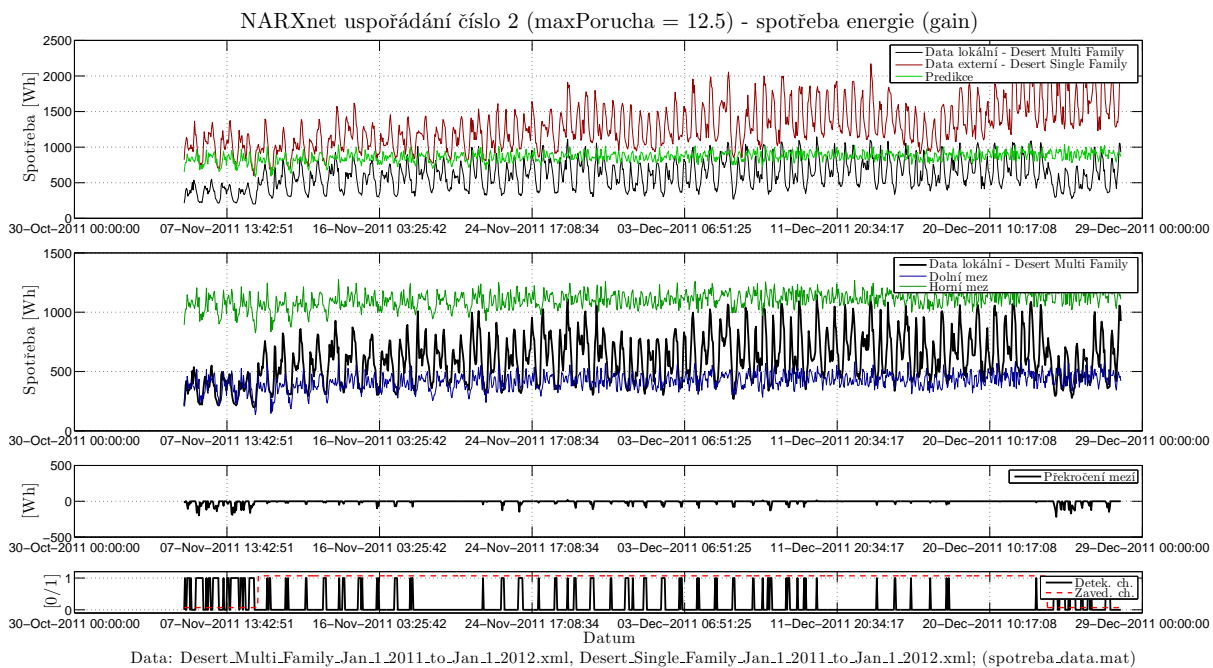
Obrázek 6.75: Detail výsledků NARXnet uspořádání číslo 2 - spotřeba energie *Desert Multi a Single Family* (peak).



Obrázek 6.76: Detail výsledků NARXnet uspořádání číslo 2 - spotřeba energie *Desert Multi a Single Family* (skok).



Obrázek 6.77: Detail výsledků NARXnet uspořádání číslo 2 - spotřeba energie *Desert Multi a Single Family* (drift).



Obrázek 6.78: Detail výsledků NARXnet uspořádání číslo 2 - spotřeba energie *Desert Multi a Single Family* (gain).

6.3 Zhodnocení

- Z výsledků uvedených výše jasně vyplývá, že velikou roli v detekci poruch hraje způsob a kvalita naučení jednotlivých modelů. Aby byla zajištěna lepší funkce metod detekce poruch, musely by být modely naučeny na větším množství rozdílných hodnot, které pokrývají celý průběh testovaných dat, což zde použito nebylo z důvodu použitelnosti v praxi. Při zavádění systému obvykle nemáme například rok staré hodnoty použitelné pro učení, ale požadujeme, aby detekce poruch fungovala co nejdříve.
- Dále stojí za povšimnutí, že nejlepších výsledků dosáhly metody, které při detekci poruch senzorů vycházely pouze z externích dat a nebyly tedy ovlivněny zavedenou chybou v testovaných lokálních datech.
- Cílem bylo vyvinout několik metod detekujících poruchy senzorů s využitím externí informace z blízkého objektu stejného typu, veřejného datového zdroje nebo redundantních senzorů. V případě teplotních dat bylo dosaženo některými metodami vynikajících výsledků a dobrých výsledků bylo dosaženo i u dat spotřeby, kde se průběh dat určených na učení příliš nelišil od průběhu na testování. Výjimku tvoří problematická data spotřeby energie jako například *Desert Multi Family*, *Desert Single Family*, kde se žádné metodě detekce poruch nepodařilo dosáhnout uspokojivých výsledků. Na obrázcích 6.14, 6.29, 6.44, 6.59 a 6.74 jsou vidět výsledky jednotlivých metod pro tato data.
- Dalším zajímavým poznatkem je, že metoda detekce poruch založená na GMM (Gaussian Mixture Model), která provádí detekci pouze na základě aktuální externí hodnoty, předčila výsledky metod, které pracují s historií lokálních dat (ARX 1, NARXnet 1).
- Nevýhodou ARX uspořádání číslo 1 je to, že se dokáže velice dobře přizpůsobit lokálním hodnotám, které obsahují poruchy, což způsobí, že nejsou detekovány. Samotným důvodem je naučený model, který při následné predikci lokální hodnoty dává velkou váhu minulé lokální hodnotě a menší váhu ostatním externím či lokálním hodnotám.
- Úspěšnost ARX uspořádání číslo 2, s výjimkou zmíněných dat spotřeby energie *Desert Multi Family*, *Desert Single Family*, přisuzují faktu, že vychází pouze z ex-

terních dat a že je založena na principu sledování difference mezi predikcí externích hodnot a lokálními daty.

- Domnívám se, že nedostatkem obou uspořádání NARXnet je, že hodnoty pro trénování, validaci a testování neuronové sítě byly vybírány náhodně, a proto se kvalita výsledných modelů a následně výsledků často lišila. S kvalitou naučeného modelu následně souvisí počet falešných poplachů.

Tabulka 6.16: Vybrané výsledky - teploty *Budihostice, Červený Újezd, Chloumek, ČZU*.

Metoda	Procento chyb [%]	Max porucha [°C]	TP [-]	FP [-]	FN [-]	TN [-]	Specificita [%]	Senzitivita [%]
ARX 1	1		1803	131	6246	7156	98,20	22,40
ARX 2	-		6010	42	2039	7245	99,42	74,67
GMM	1	0,125	4919	56	3130	7231	99,23	61,11
NARXnet 1	1		3978	442	4071	6845	93,93	49,42
NARXnet 2	-		4791	43	3258	7244	99,41	59,52

Tabulka 6.17: Vybrané výsledky - spotřeba energie *Coastal Families*.

Metoda	Procento chyb [%]	Max porucha [Wh]	TP [-]	FP [-]	FN [-]	TN [-]	Specificita [%]	Senzitivita [%]
ARX 1	0,5		53	15	3292	3912	99,62	1,58
ARX 2	-		1712	31	1633	3896	99,21	51,18
GMM	0,5	50	2011	94	1334	3833	97,61	60,12
NARXnet 1	0,5		745	513	2600	3414	86,94	22,27
NARXnet 2	-		1381	94	1964	3833	97,61	41,29

6.3.1 Shrnutí

Nejllepších výsledků na testovaných datech bylo dosaženo pomocí metody GMM (Gaussian Mixture Model). Dobrých výsledků dosáhla také metoda ARX (AutoRegressive model with eXternal input) uspořádání číslo 2. Jako třetí metodu bych uvedl NARXnet (Nonlinear AutoRegressive neural network with eXternal input) uspořádání číslo 2.

Kapitola 7

Závěr

Cílem této diplomové práce bylo vyvinout několik metod detekujících poruchy senzorů. Základní vlastností těchto metod je využití externí informace z blízkého objektu stejného typu, veřejného datového zdroje nebo redundantních senzorů. Konkrétně byly v programu Matlab R2013b implementovány metody využívající ARX (AutoRegressive model with eXternal input), GMM (Gaussian Mixture Model) a NARXnet (Nonlinear AutoRegressive neural network with eXternal input). Jejich úkolem bylo detekovat především trvale nesprávné chování senzoru a změnu jeho chování vůči počáteční situaci. Příkladem může být občasná chyba (odlehlá hodnota), náhlá porucha (skok), pozvolná porucha (drift) nebo změna zesílení (gain).

Tento cíl byl splněn s výjimkou problematických dat spotřeby energie jako například *Desert Multi Family*, *Desert Single Family*. Nejlepších výsledků bylo dosaženo pomocí metody GMM (Gaussian Mixture Model) a dále pak pomocí metody ARX (AutoRegressive model with eXternal input) uspořádání číslo 2.

Výsledkem této práce jsou zdrojové kódy k navrženým metodám a jsou uloženy na příloženém CD.

7.1 Návrh na další práci

Pokračování v této diplomové práci by spočívalo v odladění trénování modelu NARXnet. Dále by bylo vhodné upravit funkci, která provádí doučování jednotlivých modelů tak, aby bylo doučování častější a aby k němu bylo použito mnohem více hodnot než nyní. V poslední řadě bych doporučil provést spojení nejlepších metod detekce poruch.

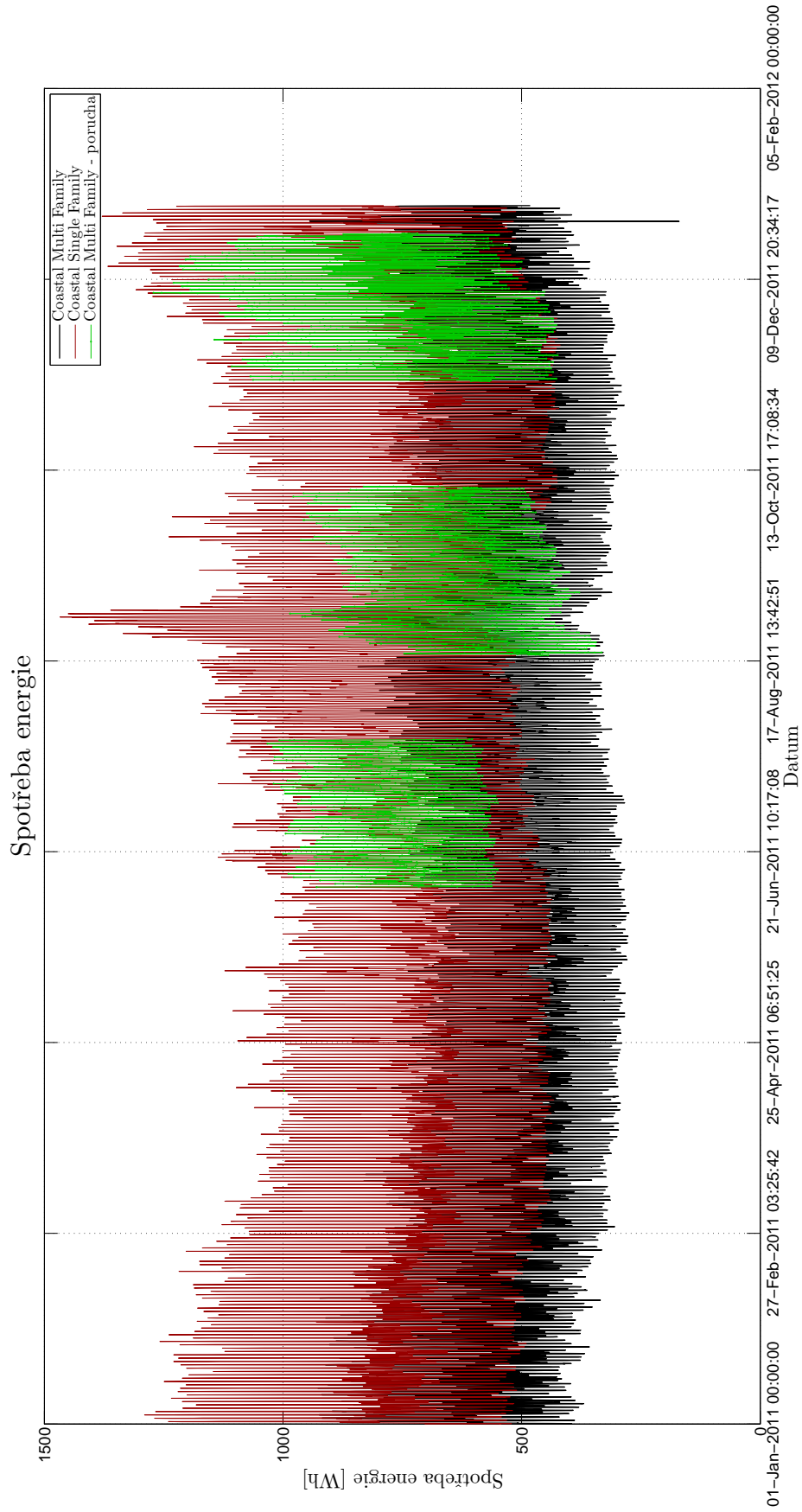
Literatura

- [1] MAŘÍK V.; ŠTĚPÁNKOVÁ O.; LAŽANSKÝ J. a kolektiv. *Umělá inteligence (4)*. 1. vydání. Academia, Praha, 2003, ISBN 80-200-1044-0.
- [2] SRIVASTAV A.; TEWARI A.; DONG B. *Baseline building energy modeling and localized uncertainty quantification using Gaussian mixture models*. Elsevier Energy and Buildings, Vol. 65, pp. 438-447, 2013.
- [3] BOX G. E. P.; JENKINS G. M.; REINSEL G. C. *Time Series Anylysis: Forecasting and Control, Wiley Series in Probability and Statistics*. 4th ed. John Wiley and Sons, USA, 2008, ISBN 978-0-470-27284-8.
- [4] CHATFIELD CH. *The Analysis of Time Series: An Introduction*. 6th ed. Chapman and Hall/CRC, UK, 2003, ISBN 1-58488-317-0.
- [5] YU D. L.; GOMM J. B.; WILLIAMS D. *Sensor fault diagnosis in a chemical process via RBF neural networks*. Elsevier Control Engineering Practice, Vol. 7, pp. 49-55, 1999.
- [6] LEE W.Y.; PARK C.; KELLY G.E. *Fault detection in an air-handling unit using residual and recursive parameter identification methods*. ASHRAE, Vol. 102, 1996.
- [7] PAALANEN P.; KAMARAINEN J. K.; ILONEN J.; KÄLVIÄINEN H. *Feature representation and discrimination based on Gaussian mixture model probability densities—Practices and algorithms*. Elsevier Pattern Recognition, Vol. 39, pp. 1346-1358, 2006.
- [8] HAVLENA V.; ŠTECHA J. *Moderní teorie řízení*. Ediční středisko ČVUT, Praha 6, Zikova 4, 1999.

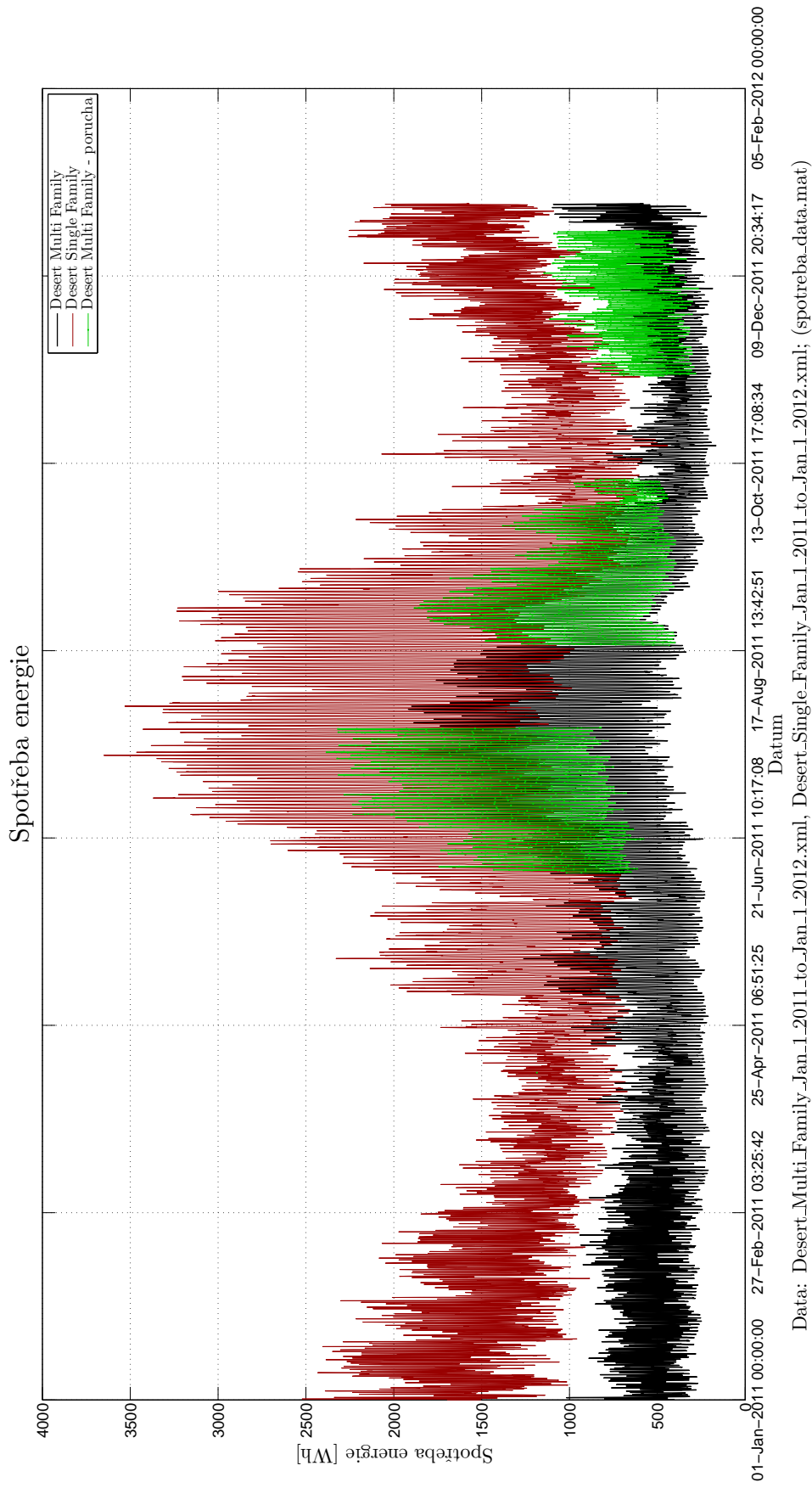
- [9] YU J. *A nonlinear kernel Gaussian mixture model based inferential monitoring approach for fault detection and diagnosis of chemical processes*. Elsevier Chemical Engineering Science, Vol. 68, pp. 506-519, 2012.
- [10] NAVARA M. *Pravděpodobnost a matematická statistika*. 1. vydání. ČVUT, Praha, 2007, ISBN 978-80-01-03795-9.
- [11] CHATTERJEE S.; NIGAM S. SINGH J. B.; UPADHYAYA L. N. *Software fault prediction using Nonlinear Autoregressive with eXogenous Inputs (NARX) network*. Springer Science + Business Media, 2011.
- [12] [online]. Design time series narx feedback neural networks. <http://www.mathworks.com/help/nnet/ug/design-time-series-narx-feedback-neural-networks.html>.
- [13] AKINCI B.; BERGÉS M.; EYBPOOSH M.; LIU X. [online]. A framework for supporting performance evaluation of hvac systems. <http://inferlab.org/projects/self-managing-hvac/>.
- [14] HAVLENA V. [online]. Odhadování, filtrace a detekce, 2010. https://moodle.dce.fel.cvut.cz/pluginfile.php/6198/mod_page/content/6/0FD_slides_2up.pdf.
- [15] KRÁKORA M.; SOJKA M. [online]. Vícevrstvý perceptron a rbf sítě, 2001. <http://gerstner.felk.cvut.cz/biolab/X33BMI/slides/NN.pdf>.
- [16] MACHOLDA F. [online]. Plýtvání energií ve výrobních a obchodních budovách, 2011. <http://bit.ly/1heMuP5>.
- [17] WU S.; SUN J.Q. [online]. Multilevel fault detection and diagnosis on office building hvac systems. https://www.academia.edu/581430/Multilevel_Fault_Detection_and_Diagnosis_on_Building_HVAC_Systems.
- [18] ISERMANN R. *Fault-Diagnosis Systems: An Introduction from Fault Detection to Fault Tolerance*. Springer, Berlín, 2006, ISBN 3-540-24112-4.
- [19] YOSHIDA H.; KUMAR S. *ARX and AFMM model-based on-line real-time data base diagnosis of sudden fault in AHU of VAV system*. Elsevier Energy Conversion and Management, Vol. 40, pp. 1191-1206, 1999.

Příloha A

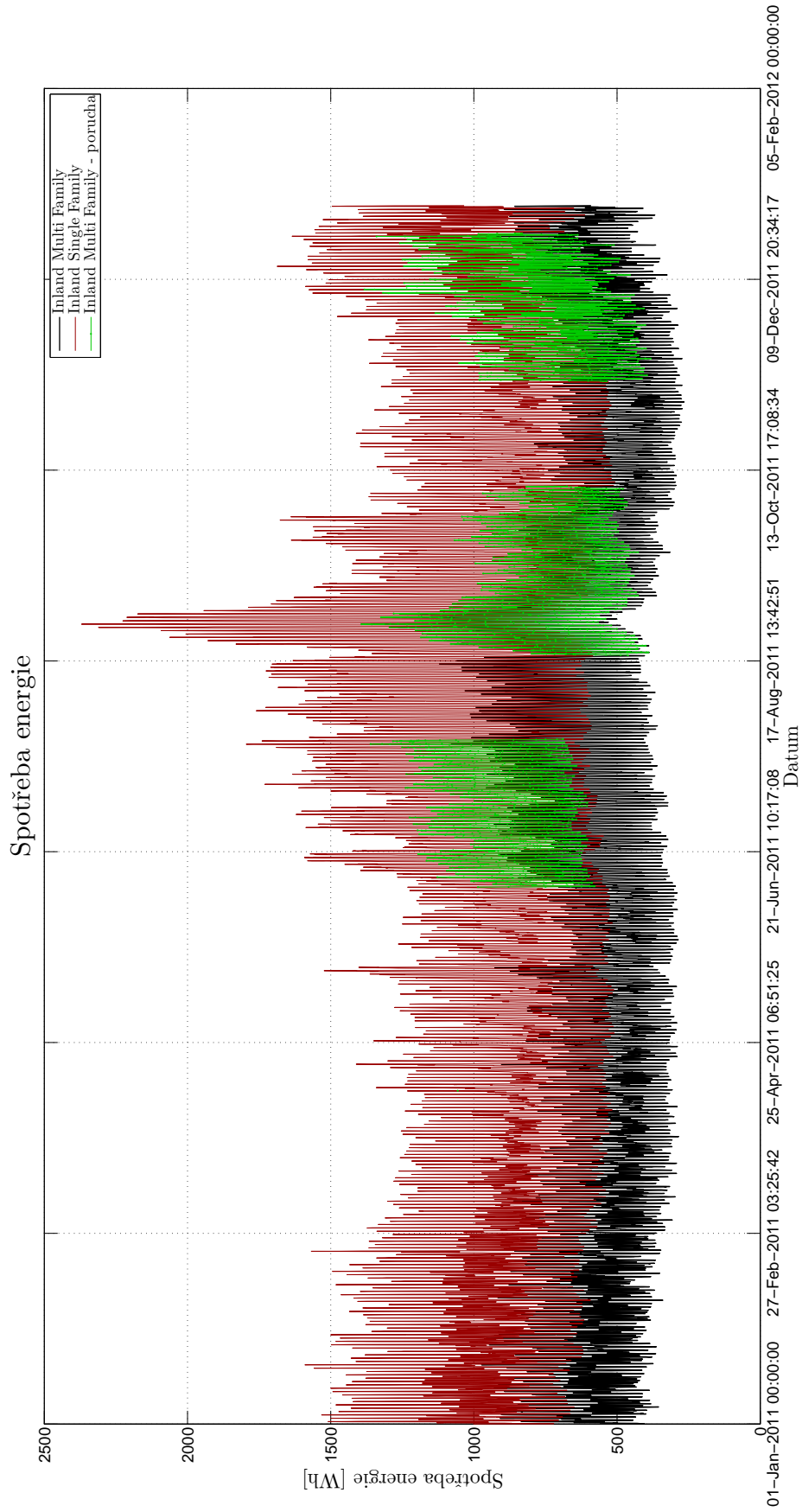
Grafické přílohy



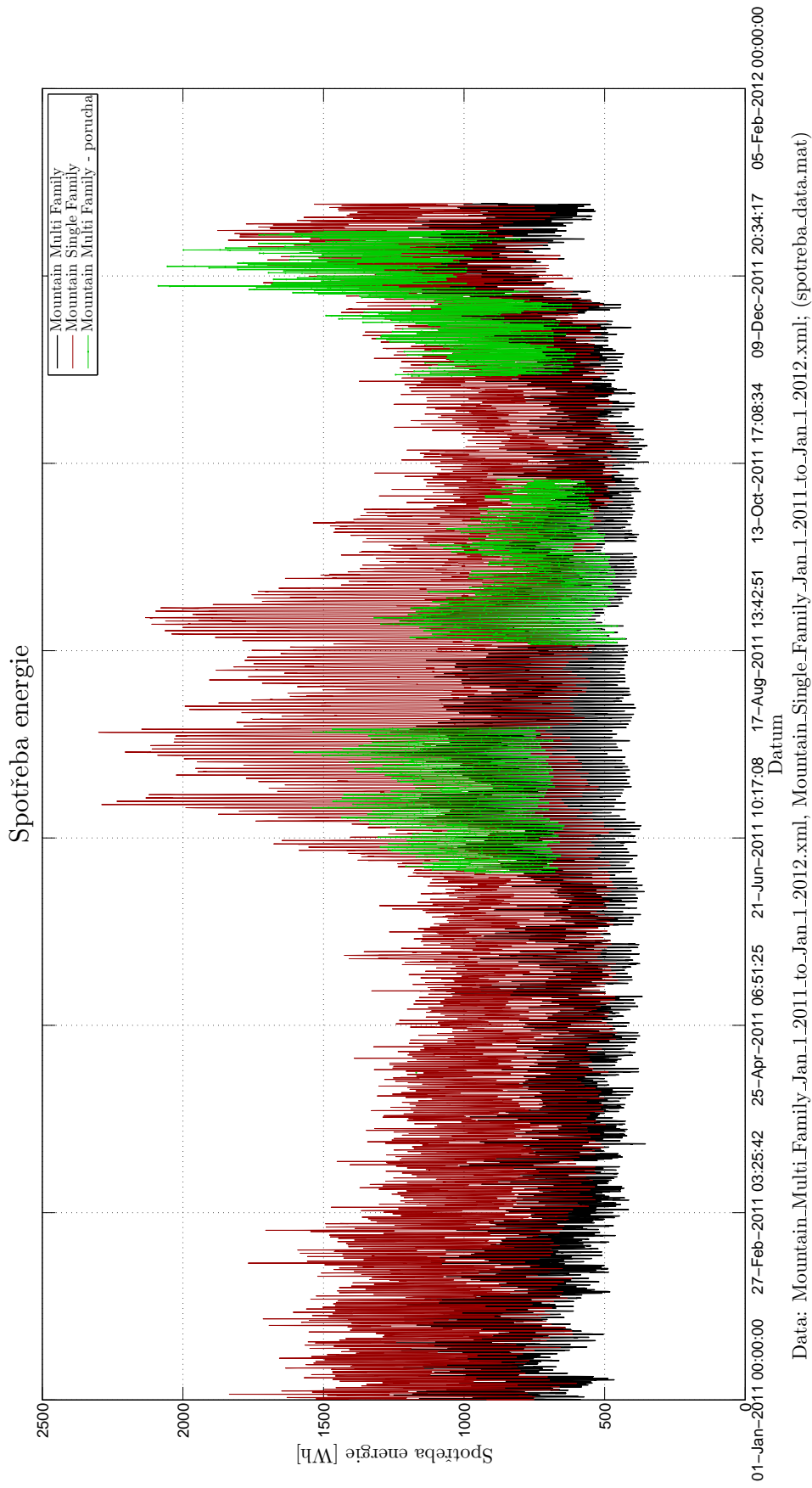
Obrázek A.1: Spotřeba elektrické energie „Coastal Family“ s průběhem poruchy.



Obrázek A.2: Spotřeba elektrické energie „Desert Family“ s průběhem poruch.

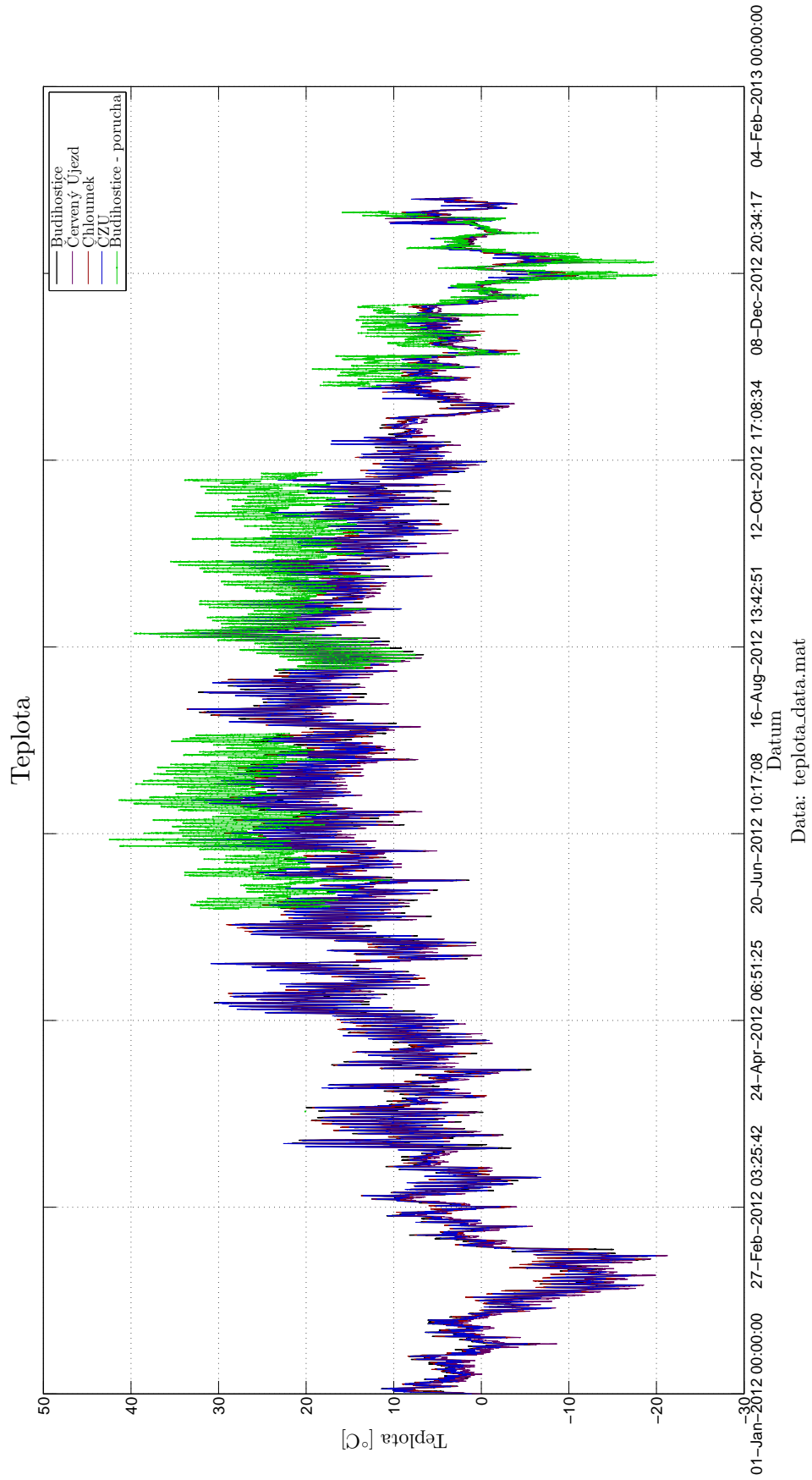


Obrázek A.3: Spotřeba elektrické energie „Inland Family“ s průběhem poruch.

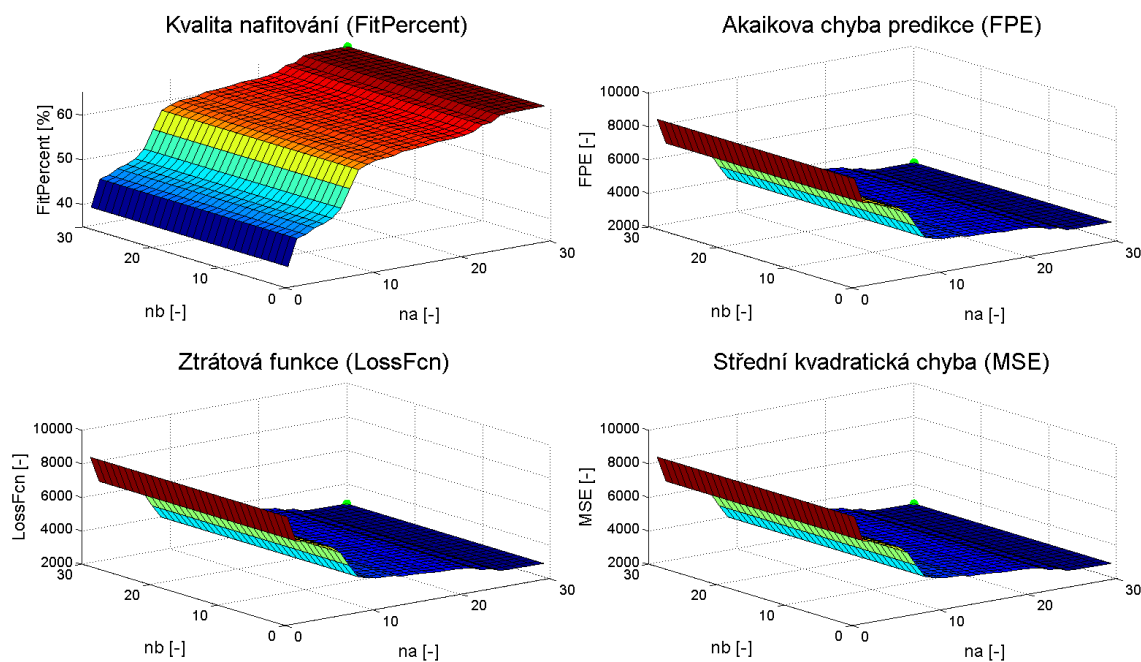


Data: Mountain_Multi_Family_Jan.1.2011_to_Jan.1.2012.xml, Mountain_Single_Family_Jan.1.2011_to_Jan.1.2012.xml; (spotreba_data.mat)

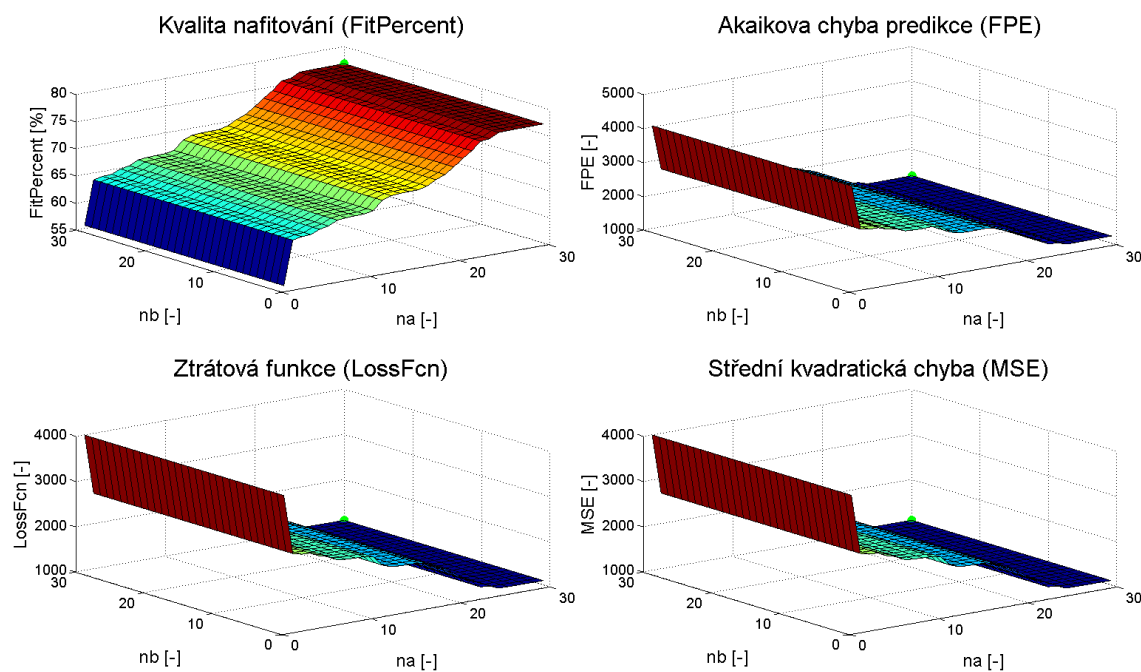
Obrázek A.4: Spotřeba elektrické energie „Mountain Family“ s průběhem poruchy.



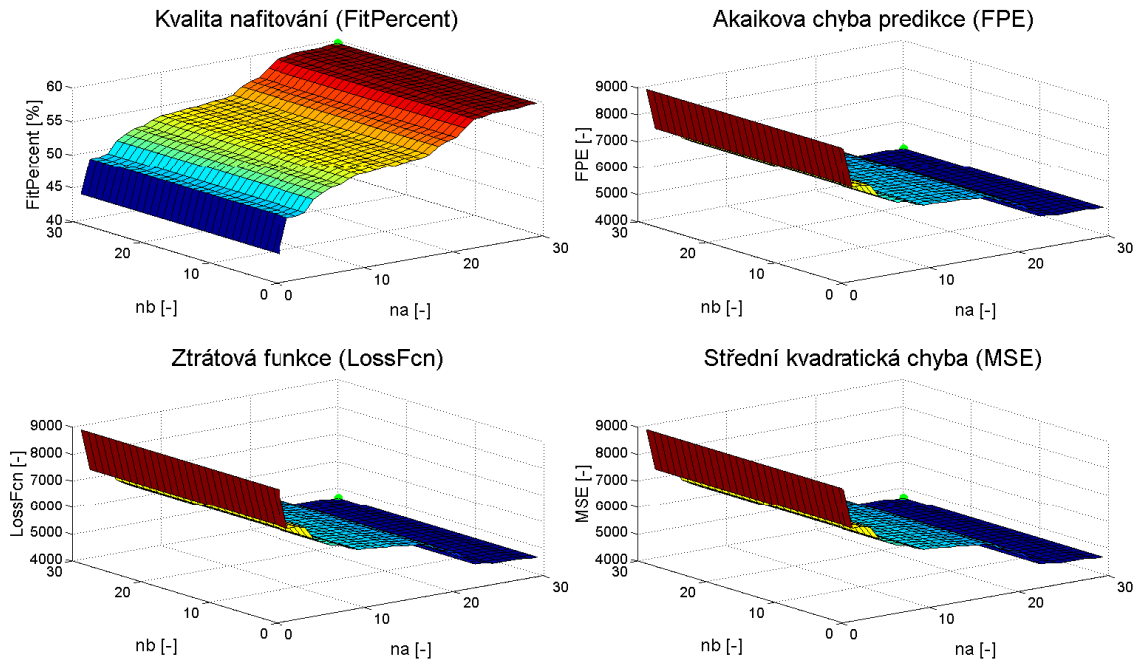
Obrázek A.5: Průběh teplot ze čtyř meteorologických stanic s průběhem poruchy.



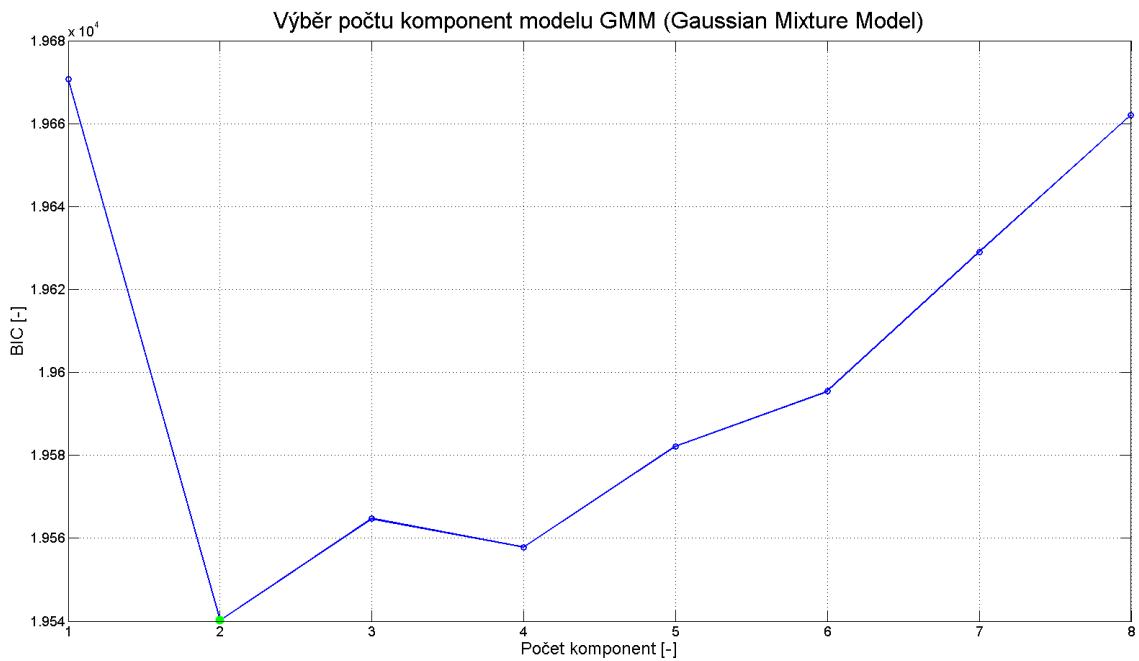
Obrázek A.6: Hledání optimálního nastavení parametrů ARX modelu pro data spotřeby energie „Desert Family“.



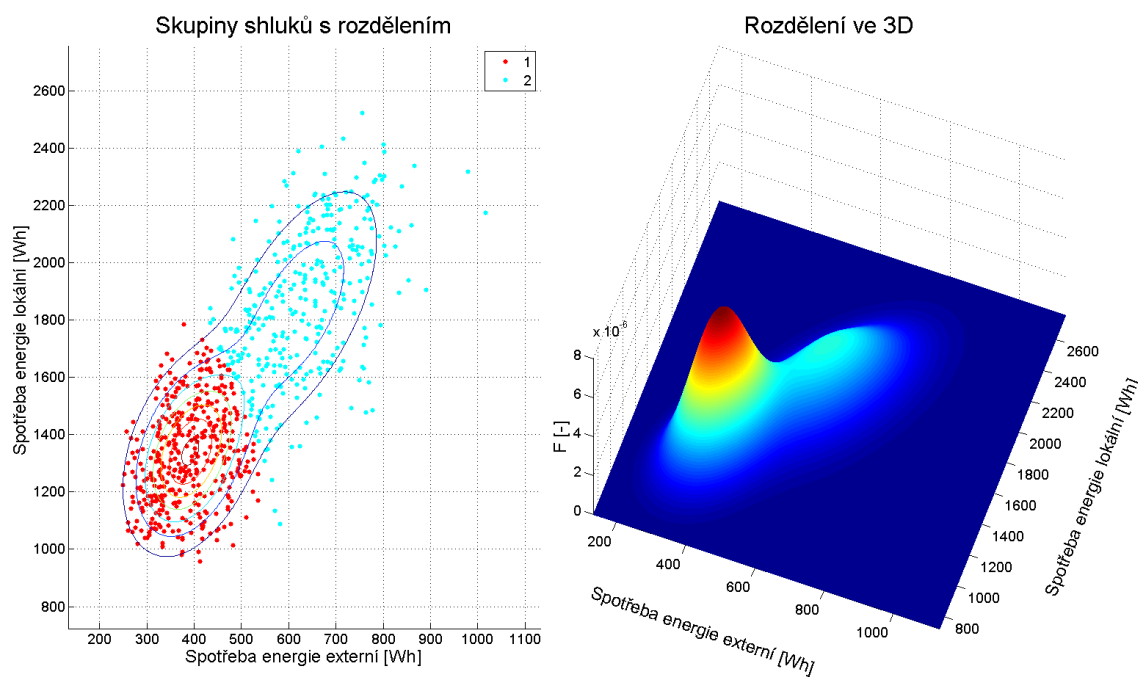
Obrázek A.7: Hledání optimálního nastavení parametrů ARX modelu pro data spotřeby energie „Inland Family“.



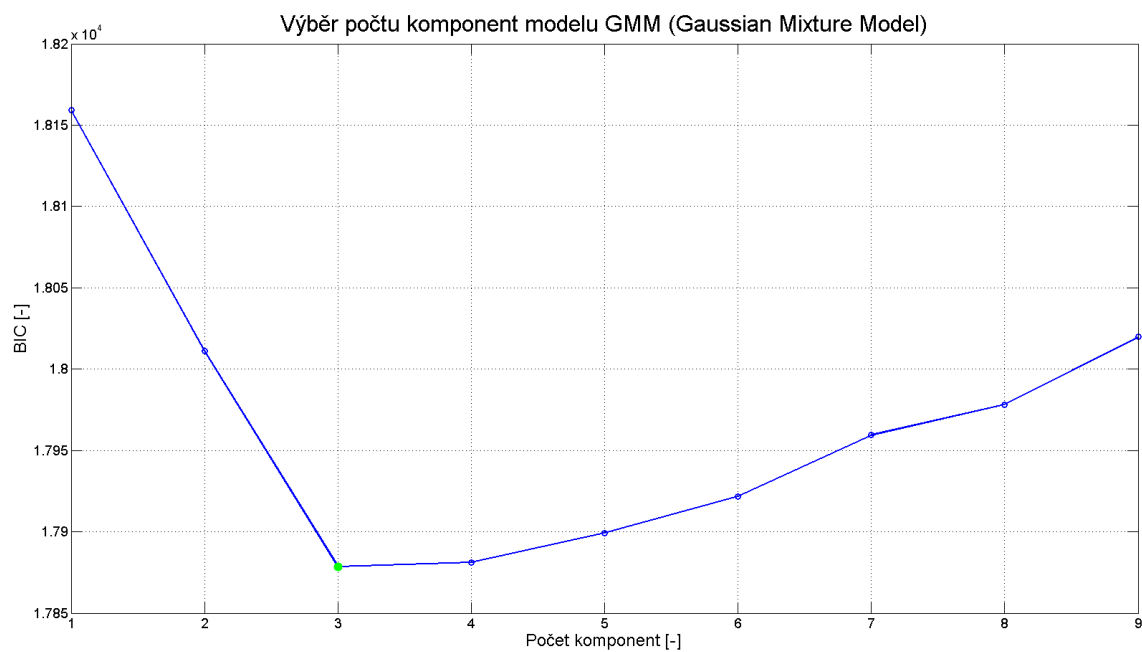
Obrázek A.8: Hledání optimálního nastavení parametrů ARX modelu pro data spotřeby energie „Mountain Family“.



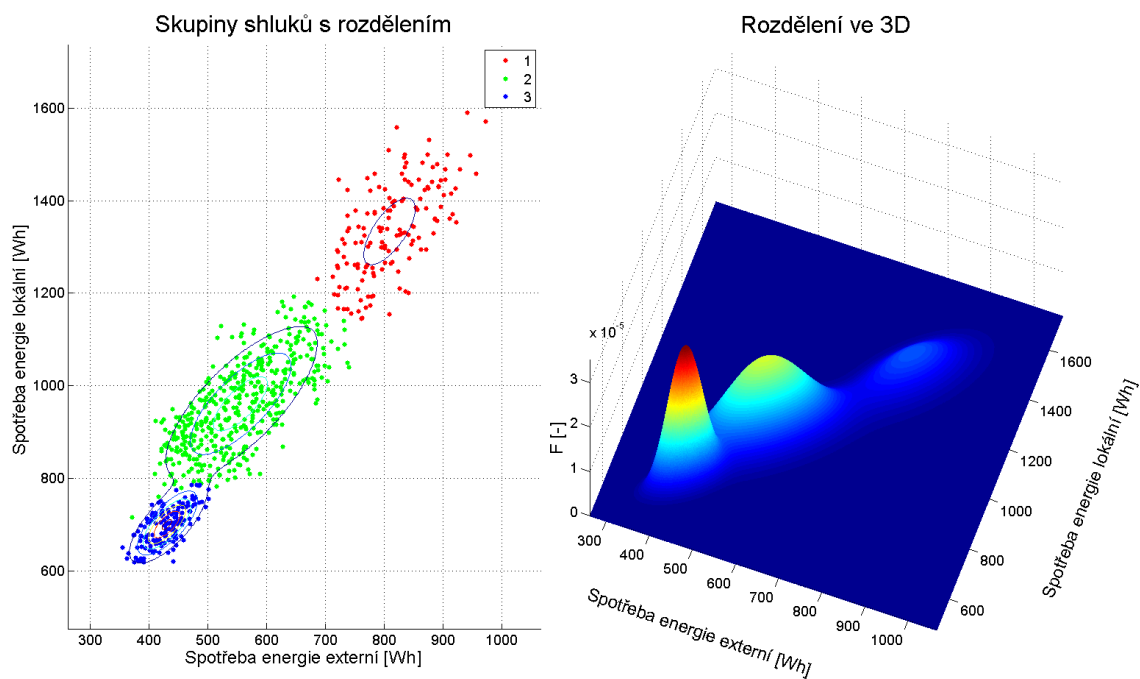
Obrázek A.9: Hledání optimálního počtu komponent GMM modelu pro data spotřeby energie „Desert Family“.



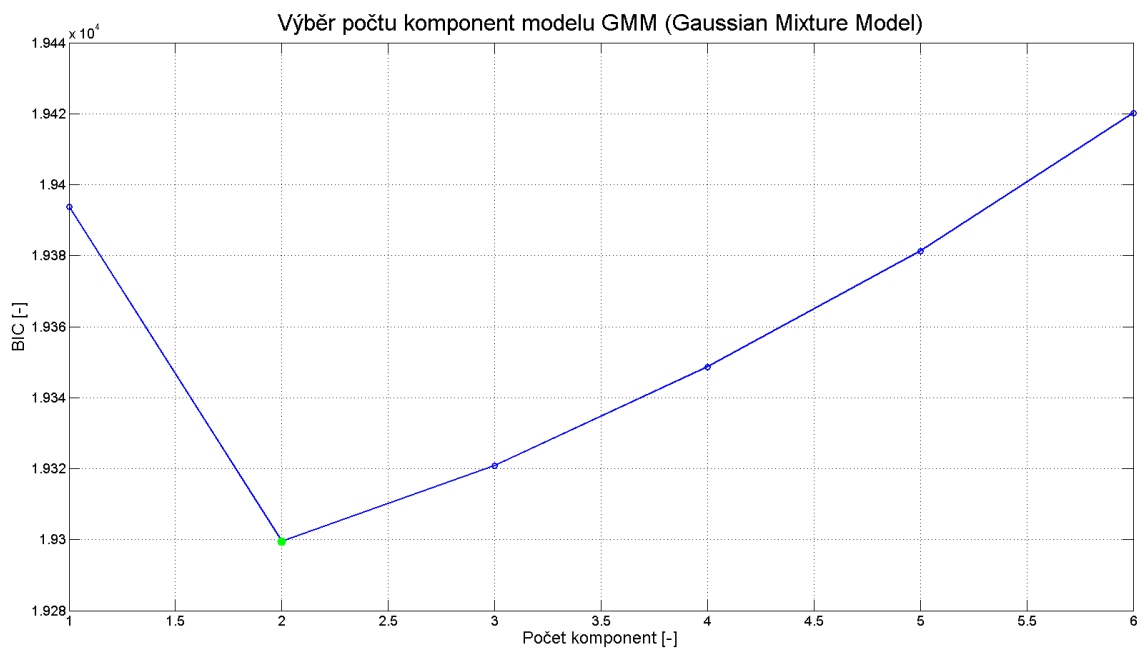
Obrázek A.10: Rozdělení dat spotřeby energie „Desert Family“ GMM modelem.



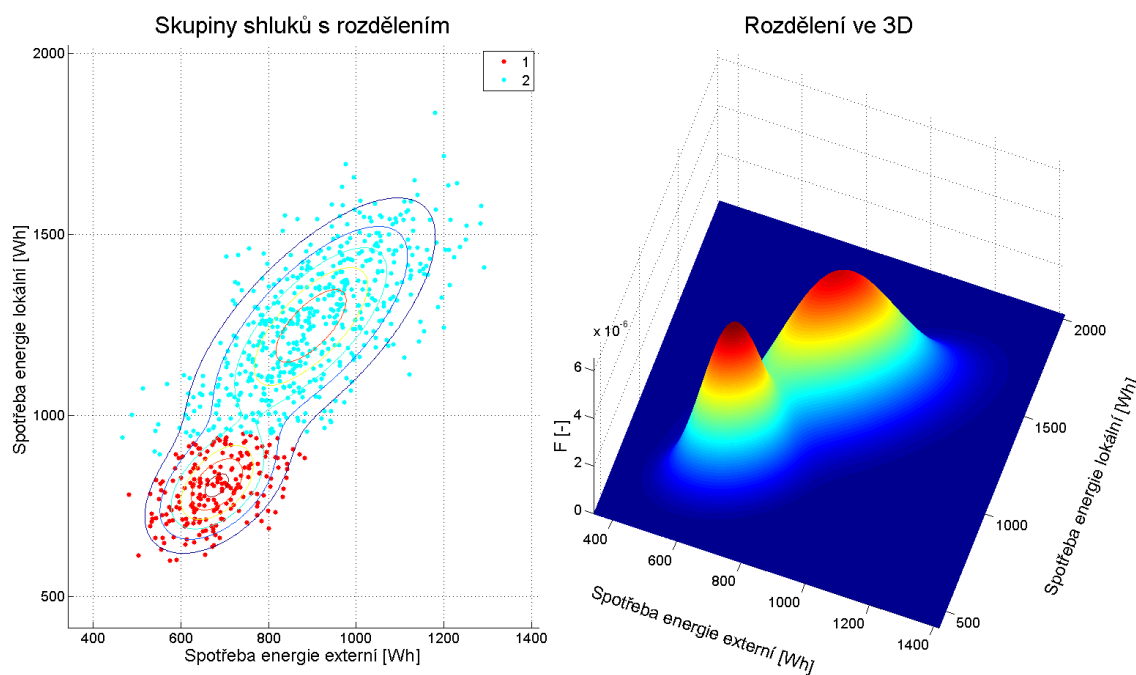
Obrázek A.11: Hledání optimálního počtu komponent GMM modelu pro data spotřeby energie „Inland Family“.



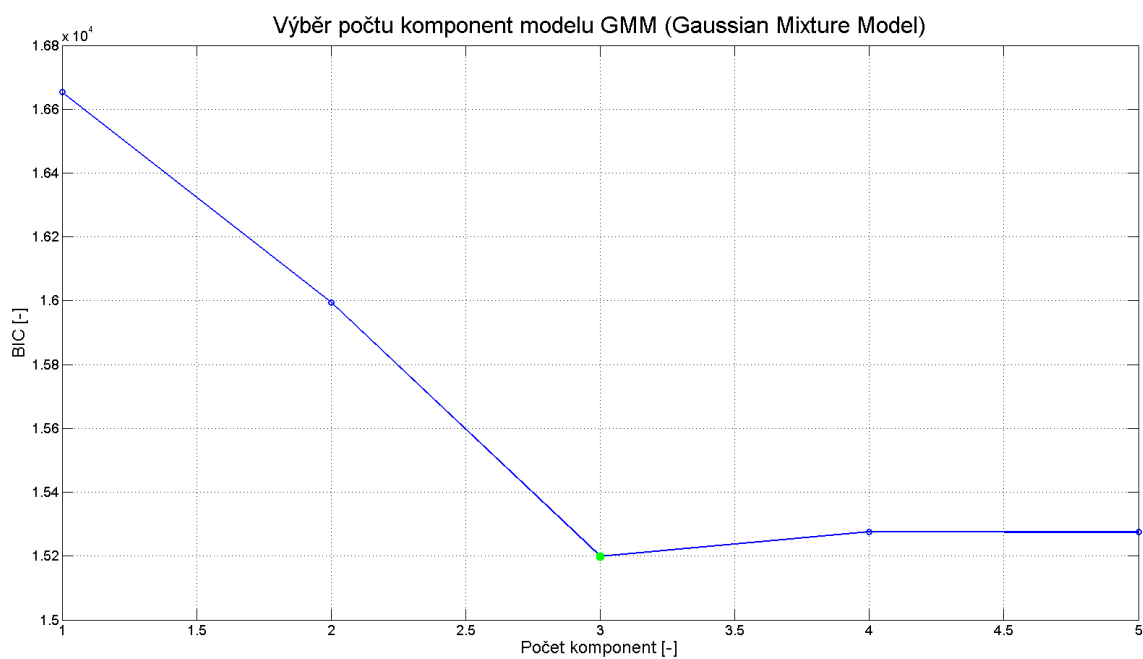
Obrázek A.12: Rozdělení dat spotřeby energie „Inland Family“ GMM modelem.



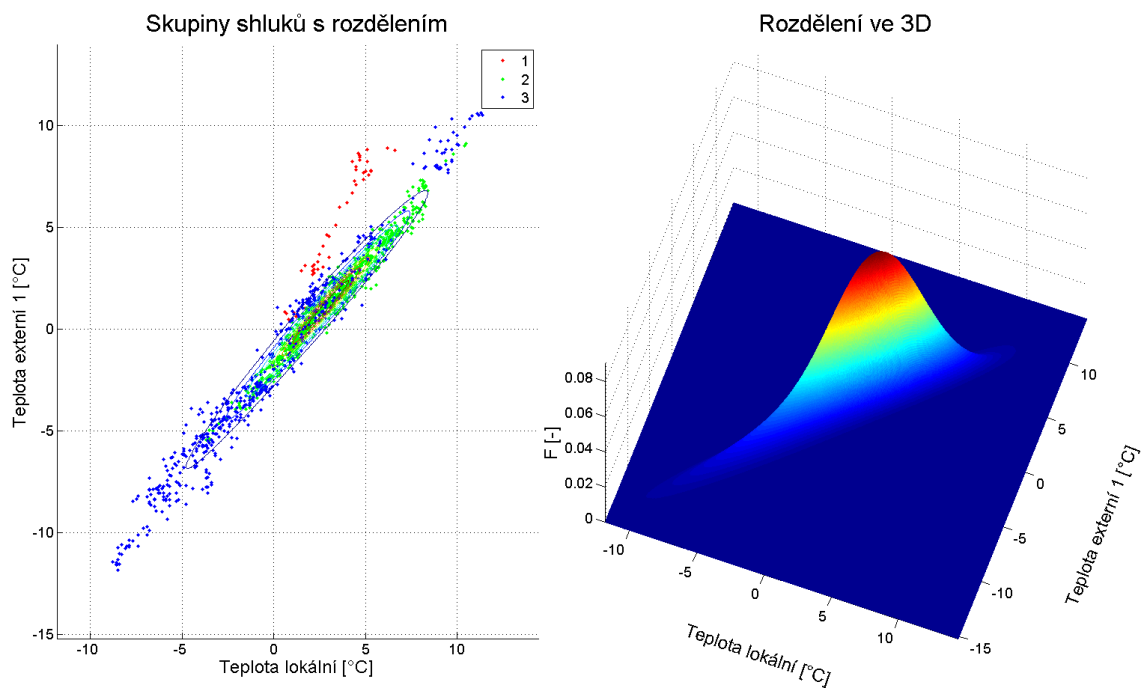
Obrázek A.13: Hledání optimálního počtu komponent GMM modelu pro data spotřeby energie „Mountain Family“.



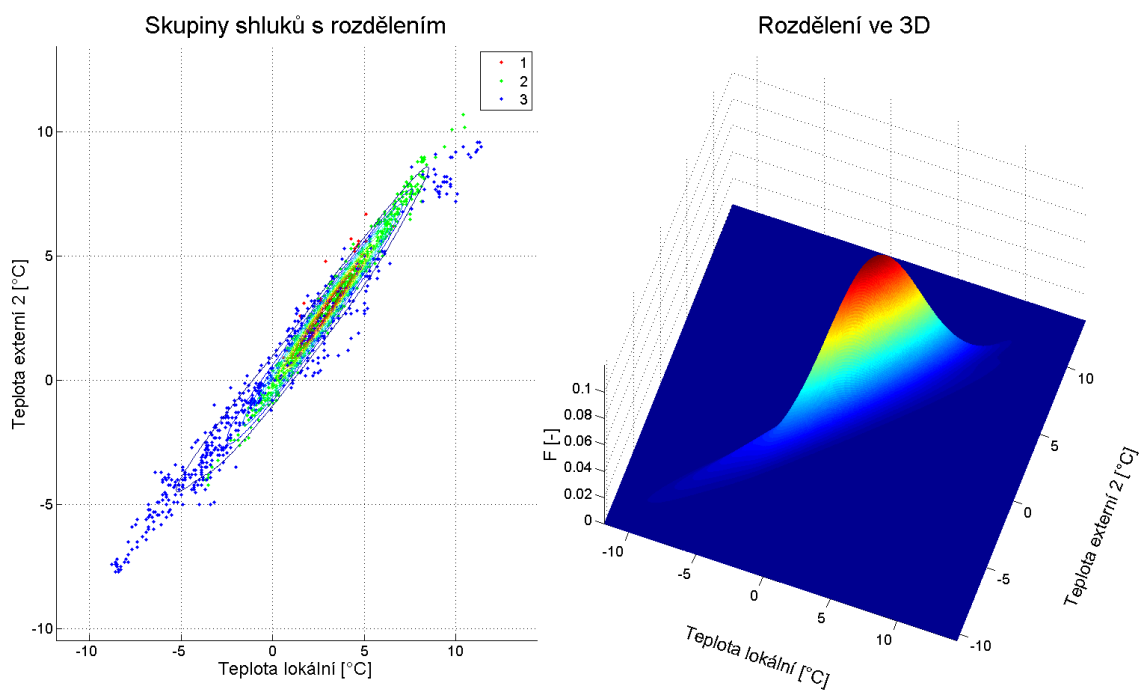
Obrázek A.14: Rozdělení dat spotřeby energie „Mountain Family“ GMM modelu.



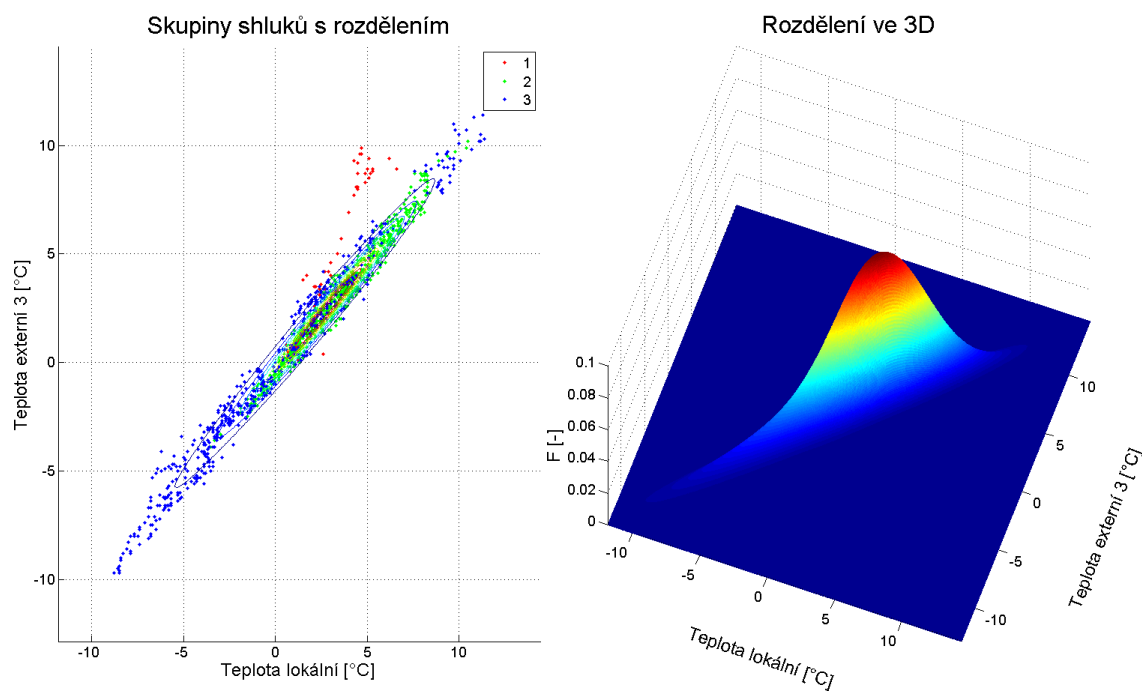
Obrázek A.15: Hledání optimálního počtu komponent GMM modelu pro teplotní data.



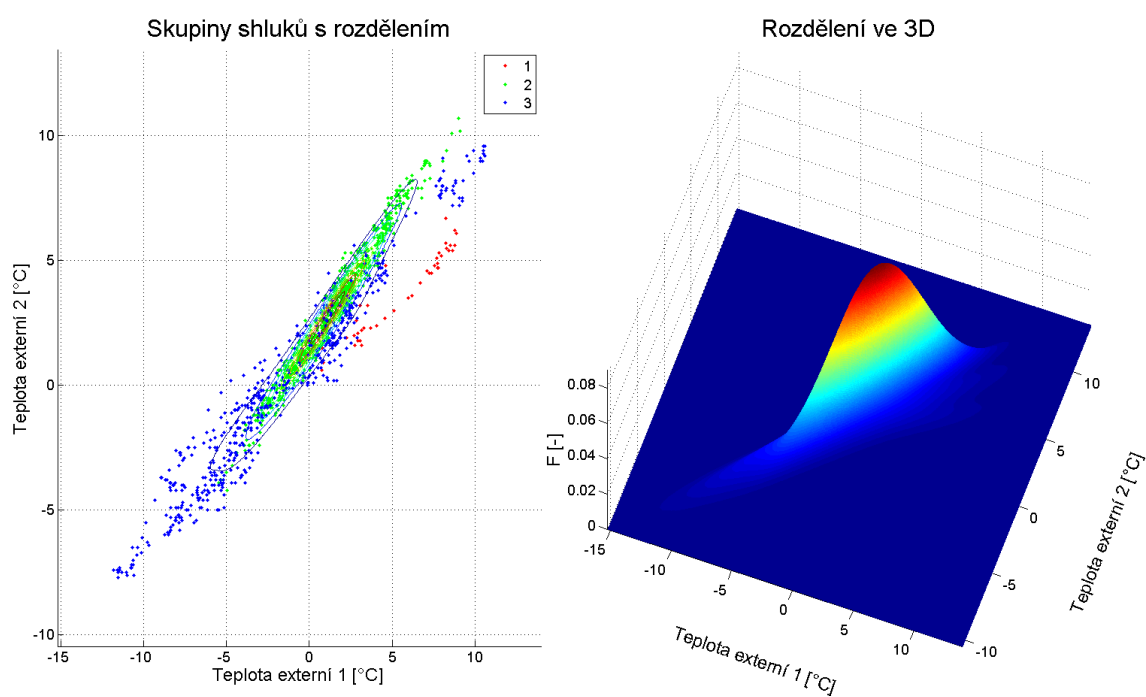
Obrázek A.16: 1) Rozdělení teplotních dat GMM modelu.



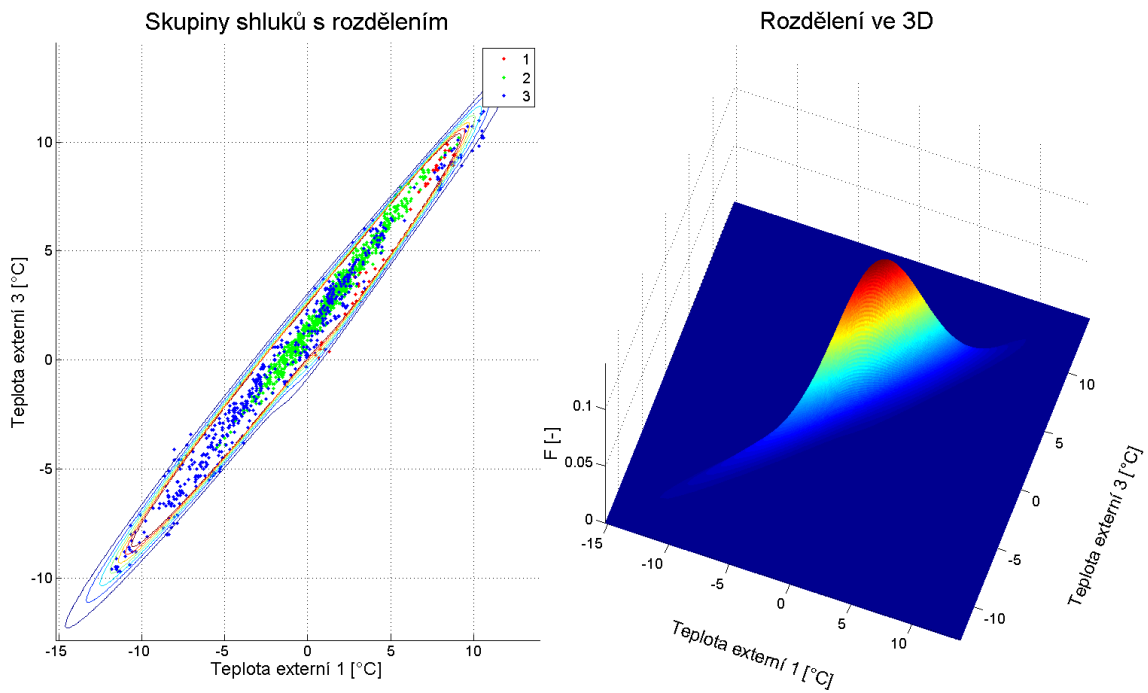
Obrázek A.17: 2) Rozdělení teplotních dat GMM modelu.



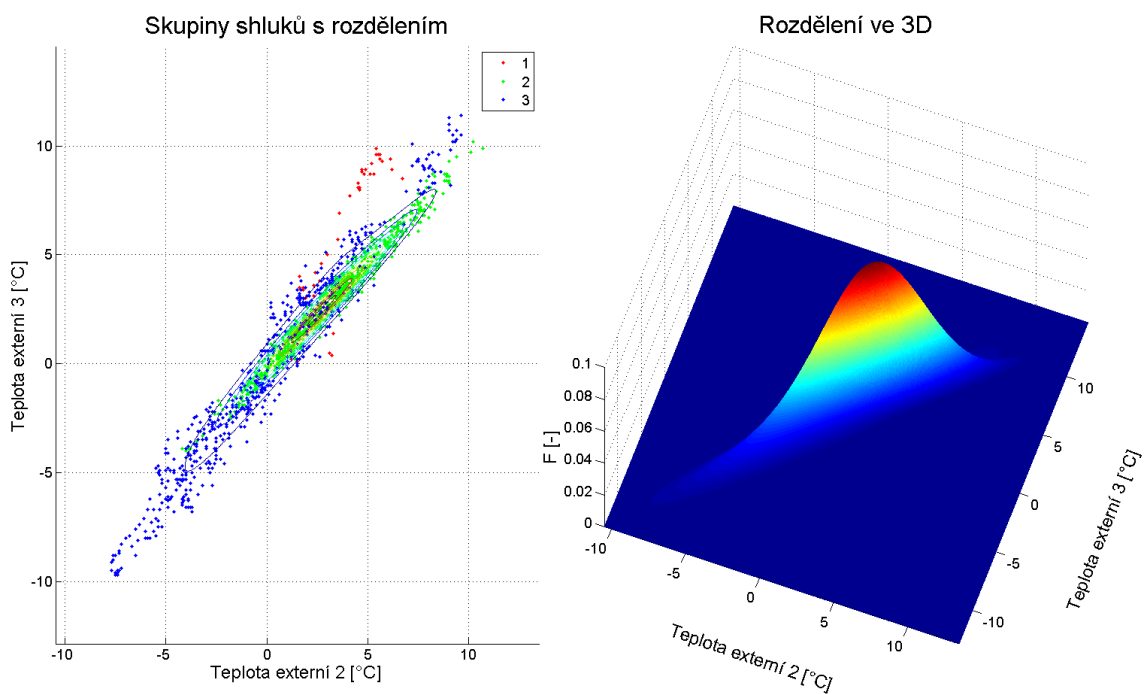
Obrázek A.18: 3) Rozdělení teplotních dat GMM modelu.



Obrázek A.19: 4) Rozdělení teplotních dat GMM modelu.



Obrázek A.20: 5) Rozdělení teplotních dat GMM modelu.



Obrázek A.21: 6) Rozdělení teplotních dat GMM modelu.

Příloha B

Tabulky

Tabulka B.1: Výsledky ARX uspořádání číslo 1 - spotřeba energie *Inland Multi a Single Family*.

Procento chyb [%]	Max porucha [Wh]	TP [-]	FP [-]	FN [-]	TN [-]	Specificita [%]	Senzitivita [%]
0,5	12,5	1308	21	2037	3906	99,47	39,10
	50,0	1130	7	2215	3920	99,82	33,78
	150,0	929	1	2416	3926	99,97	27,77
1,0	12,5	1619	43	1726	3884	98,91	48,40
	50,0	1115	7	2230	3920	99,82	33,33
	150,0	846	1	2499	3926	99,97	25,29
1,5	12,5	1361	45	1984	3882	98,85	40,69
	50,0	1146	8	2199	3919	99,80	34,26
	150,0	894	1	2451	3926	99,97	26,73

Tabulka B.2: Výsledky ARX uspořádání číslo 1 - spotřeba energie *Mountain Multi a Single Family*.

Procento chyb [%]	Max porucha [Wh]	TP [-]	FP [-]	FN [-]	TN [-]	Specificita [%]	Senzitivita [%]
0,5	12,5	142	48	3203	3879	98,78	4,25
	50,0	81	17	3264	3910	99,57	2,42
	150,0	19	3	3326	3924	99,92	0,57
1,0	12,5	169	67	3176	3860	98,29	5,05
	50,0	97	21	3248	3906	99,47	2,90
	150,0	22	3	3323	3924	99,92	0,66
1,5	12,5	206	83	3139	3844	97,89	6,16
	50,0	109	27	3236	3900	99,31	3,26
	150,0	27	3	3318	3924	99,92	0,81

Tabulka B.3: Výsledky ARX uspořádání číslo 2 - spotřeba energie *Inland Multi a Single Family*.

Max porucha [Wh]	TP [-]	FP [-]	FN [-]	TN [-]	Specificita [%]	Senzitivita [%]
12,5	2220	67	1125	3860	98,29	66,37
50,0	1809	25	1536	3902	99,36	54,08
150,0	1711	3	1634	3924	99,92	51,15

Tabulka B.4: Výsledky ARX uspořádání číslo 2 - spotřeba energie *Mountain Multi a Single Family*.

Max porucha [Wh]	TP [-]	FP [-]	FN [-]	TN [-]	Specificita [%]	Senzitivita [%]
12,5	1459	220	1886	3707	94,40	43,62
50,0	1402	183	1943	3744	95,34	41,91
150,0	1317	101	2028	3826	97,43	39,37

Tabulka B.5: Výsledky GMM - spotřeba energie *Inland Multi a Single Family*.

Procento chyb [%]	Max porucha [Wh]	TP [-]	FP [-]	FN [-]	TN [-]	Specificita [%]	Senzitivita [%]
0,5	12,5	2349	92	996	3835	97,66	70,22
	50,0	2295	34	1050	3893	99,13	68,61
	150,0	2178	4	1167	3923	99,90	65,11
1,0	12,5	2433	120	912	3807	96,94	72,74
	50,0	2336	59	1009	3868	98,50	69,84
	150,0	2261	4	1084	3923	99,90	67,59
1,5	12,5	2434	169	911	3758	95,70	72,77
	50,0	2359	78	986	3849	98,01	70,52
	150,0	2249	7	1096	3920	99,82	67,23

Tabulka B.6: Výsledky GMM - spotřeba energie *Mountain Multi a Single Family*.

Procento chyb [%]	Max porucha [Wh]	TP [-]	FP [-]	FN [-]	TN [-]	Specificita [%]	Senzitivita [%]
0,5	12,5	654	35	2691	3892	99,11	19,55
	50,0	607	24	2738	3903	99,39	18,15
	150,0	530	13	2815	3914	99,67	15,84
1,0	12,5	709	57	2636	3870	98,55	21,20
	50,0	660	33	2685	3894	99,16	19,73
	150,0	578	14	2767	3913	99,64	17,28
1,5	12,5	779	106	2566	3821	97,30	23,29
	50,0	733	52	2612	3875	98,68	21,91
	150,0	648	26	2697	3901	99,34	19,37

Tabulka B.7: Výsledky NARXnet uspořádání číslo 1 - spotřeba energie
Inland Multi a Single Family.

Procento chyb [%]	Max porucha [Wh]	TP [-]	FP [-]	FN [-]	TN [-]	Specifická [%]	Senzitivita [%]
0,5	12,5	1524	680	1821	3247	82,68	45,56
	50,0	1087	326	2258	3601	91,70	32,50
	150,0	1045	114	2300	3813	97,10	31,24
1,0	12,5	1872	652	1473	3275	83,40	55,96
	50,0	1250	397	2095	3530	89,89	37,37
	150,0	1215	237	2130	3690	93,96	36,32
1,5	12,5	1334	992	2011	2935	74,74	39,88
	50,0	1523	595	1822	3332	84,85	45,53
	150,0	1009	213	2336	3714	94,58	30,16

Tabulka B.8: Výsledky NARXnet uspořádání číslo 1 - spotřeba energie
Mountain Multi a Single Family.

Procento chyb [%]	Max porucha [Wh]	TP [-]	FP [-]	FN [-]	TN [-]	Specifická [%]	Senzitivita [%]
0,5	12,5	614	199	2731	3728	94,93	18,36
	50,0	363	137	2982	3790	96,51	10,85
	150,0	445	113	2900	3814	97,12	13,30
1,0	12,5	383	178	2962	3749	95,47	11,45
	50,0	491	284	2854	3643	92,77	14,68
	150,0	420	96	2925	3831	97,56	12,56
1,5	12,5	366	272	2979	3655	93,07	10,94
	50,0	650	341	2695	3586	91,32	19,43
	150,0	295	143	3050	3784	96,36	8,82

Tabulka B.9: Výsledky NARXnet uspořádání číslo 2 - spotřeba energie
Inland Multi a Single Family.

Max porucha [Wh]	TP [-]	FP [-]	FN [-]	TN [-]	Specificita [%]	Senzitivita [%]
12,5	584	175	2761	3752	95,54	17,46
50,0	492	14	2853	3913	99,64	14,71
150,0	1303	719	2042	3208	81,69	38,95

Tabulka B.10: Výsledky NARXnet uspořádání číslo 2 - spotřeba energie
Mountain Multi a Single Family.

Max porucha [Wh]	TP [-]	FP [-]	FN [-]	TN [-]	Specificita [%]	Senzitivita [%]
12,5	923	153	2422	3774	96,10	27,59
50,0	1299	91	2046	3836	97,68	38,83
150,0	598	71	2747	3856	98,19	17,88

Příloha C

Obsah přiloženého CD

- Tento text ve formátu PDF.
- Grafické přílohy, které se do tohoto textu nevešly.
- Zdrojové kódy k implementovaným metodám detekce poruch senzorů.