

---

**TECHNICKÁ UNIVERZITA V LIBERCI**

Studentská 2, 461 17 Liberec 1

**Fakulta mechatroniky, informatiky a mezioborových studií**

Ústav mechatroniky a technické informatiky

---

**Autoreferát disertační práce**

**Optimalizace archivu měření  
kvality elektrické energie**

Liberec 2011

Autor: Ing. Jan Kraus  
Vedoucí práce: Prof. Ing. Aleš Richter, CSc.

---



---

**Autoreferát disertační práce**

**Optimalizace databází měření  
kvality elektrické energie**

Studijní program: P 2612 Elektrotechnika a informatika

Studijní obor: 2612V045 Technická kybernetika

Pracoviště: Ústav mechatroniky a technické informatiky  
Fakulta mechatroniky, informatiky a mezioborových studií  
Technická univerzita v Liberci  
Studentská 2, 461 17, Liberec

Autor: Ing. Jan Kraus

Školitel: Prof. Ing. Aleš Richter, CSc.

---

**Abstract:** Techniques for evaluation of power quality are analysed and described in depth. In relation to this general specification all required standards and other requirements for a power quality monitor and monitoring system is analysed. The main topic for the presented work is an effective storage of the measured data in device and in the final power quality database (in respect to the consumed storage space and fitness for further analysis). For different types of archives and encodings various lossless compression algorithms are evaluated. The LZMA algorithm is identified as a good out-of-shelf choice. Further improvements of the compression (of aggregated main archive data) in the presented experiments was achieved by application of time series modelling techniques. The polynomial and spline interpolation, physical relationship modelling, statistical characteristics on intervals and fast Fourier transform was evaluated. To achieve the overall best compression ratio the best prediction technique (model) turns out to be a minimum value on the interval. The span of interval is best defined by corresponding metric. From many possible and published choices none was efficient enough so the new LogHist metric is proposed and experimental results are also presented. Unlike the more common metrics the proposed LogHist metric is also useful for multi-variable input files and is showing its effectivity in determination of the node/interval distribution on the original data. Compression of series modeled with intervals of balanced LogHist have shown the best compression performance. The proposed was implemented in a compact and limited hardware platform based on ARM microcontroller using also unique non-power of two implementation of FFT algorithm and new fast algorithm for calculation of reactive power.

**Keywords:** power quality, data compression efficiency, time series of measurement, data modelling and prediction

**Abstrakt:** práce popisuje techniky pro efektivní vyhodnocování kvality elektrické energie a zaměřuje se na optimalizaci archivů měření. V této souvislosti byly analyzovány požadavky norem a různých typických uživatelů na komplexní hodnotící systém. Hlavním náplní této práce je návrh optimalizace archivů měření. Pro různé typy archivů, archivovaných dat a metod kódování je analyzována efektivita bezztrátových kompresních algoritmů. LZMA algoritmus je vyhodnocen jako nejvhodnější. Dalšího zlepšení komprese dat archivu je dosaženo s využitím matematických modelů hodnot časové řady veličin. Použité metody zahrnují aproximace (polynomem, spline křivkami), odhad hodnot veličin dle přibližně platných vztahů, využití charakteristické hodnoty řady na dílčích intervalech a modelování pomocí FFT. S ohledem na efektivitu komprese je jako nejvhodnější technika predikce hodnot vyhodnoceno modelování minimem na vhodně zvolených intervalech s následnými lokálními optimalizacemi. Pro potřebu určení správného počtu uzlů/intervalů modelu byla navržena unikátní metrika LogHist, která kvantifikuje charakter datového toku kombinace model-residuum s ohledem na použitý typ komprese a její minimalizaci při konstrukci modelu je dosaženo nejlepších kompresních poměrů. Byly také navrženy a implementovány konkrétní postupy hodnocení kvality elektrické energie. Při jeho implementaci jsou využity netradiční algoritmy pro výpočet FFT a součástí implementace je i zcela nová metoda pro rychlý výpočet jalového výkonu.

**Klíčová slova:** kvalita elektrické energie, efektivita komprese dat, časové řady hodnot měření, modelování a predikce dat

# Obsah

<b>1</b>	<b>Cíle disertační práce</b>	<b>2</b>
<b>2</b>	<b>Stav řešené problematiky</b>	<b>4</b>
2.1	Datová úložiště	4
2.2	Algoritmy bezztrátové komprese	5
2.3	Typy archivovaných dat	6
2.4	Komprese hlavního archivu analyzátoru	7
<b>3</b>	<b>Definice a vztahy</b>	<b>8</b>
3.1	Základní veličiny	8
3.2	Hodnocení kvality elektrické energie	11
<b>4</b>	<b>Optimalizace archivů</b>	<b>13</b>
4.1	Archivace dat v analyzátoru	13
4.2	Archivace měření v PC	13
4.2.1	Binární soubor jako zdroj dat	14
4.2.2	SQL Server jako zdroj dat	14
4.3	Komprese hlavního archivu	15
4.3.1	Parametry modelovaných veličin	15
<b>5</b>	<b>Modelování časových řad</b>	<b>19</b>
5.1	Běžné metriky	19
5.2	Navržená LogHist metrika	20
5.3	Typy modelů	20
5.4	Konstrukce optimálních modelů	22
5.5	Popis knihovny pro modelování	25
5.6	Shrnutí výsledků měření	27

# Seznam pojmů

- AC Aritmetické kódování
- AIC Akaikeho informační kritérium (Akaike Information Criterion)
- BIC Bayesovské informační kritérium (Bayesian Information Criterion)
- CIREN International Conference on Electricity Distribution
- DFT diskrétní Fourierova transformace (Discrete Fourier Transform)
- EMC Elektromagnetická kompatibilita
- EPRI Electric Power Research Institute
- FFT rychlá Fourierova transformace (Fast Fourier Transform)
- HUF Huffmanovo kódování
- IEC International Electrotechnical Commission
- IEEE The Institute of Electrical and Electronics Engineers
- LZ Lempel-Ziv, rodina slovníkových kompresních algoritmů
- LZMA Lempel-Ziv-Markov Chain, varianta LZ algoritmu
- RC Range Coder
- RVC rychlé změny napětí (Rapid Voltage Changes)
- SF Shannon-Fano, entropický kódér

Symbol	Popis	Jednotka
$f$	frekvence	$Hz$
$f_{NOM}$	nominální frekvence systému	$Hz$
$f_i$	frekvence $i$ -té harmonické složky signálu	$Hz$
$U_{ef}, U_i$	efektivní hodnota napětí, $i$ -té fáze	$V$
$u(t)$	okamžitá hodnota napětí	$V$
$u_i(t)$	okamžitá hodnota napětí $i$ -té harm. složky	$V$
$u_H(t)$	harmonické napětí	$V$
$I_{ef}, I_i$	efektivní hodnota proudu, $i$ -té fáze	$A$
$i(t)$	okamžitá hodnota proudu	$A$
$i_i(t)$	okamžitá hodnota proudu $i$ -té harm. složky	$A$
$i_H(t)$	harmonický proud	$A$
$E_p$	činná energie	$Wh$
$p(t)$	okamžitý výkon	$W$
$S$	zdánlivý výkon	$VA$
$P$	činný výkon	$W$
$P_i$	činný výkon $i$ -té harm. složky	$W$
$P_H$	harmonický činný výkon	$W$
$Q_{\sim}$	neaktivní výkon	$var$
$Q$	jalový výkon	$var$
$Q_i$	jalový výkon $i$ -té harm. složky	$var$
$Q_H$	harmonický jalový výkon	$var$
$D$	deformační výkon	$var$
$\lambda$	účinnost	—
$\cos(\varphi)$	cos fázového posunu 1. harmonické napětí a proudu	—
$g_i$	činitel deformace proudu	—
$k_i$	činitel zkreslení	—
$k_1$	faktor zpětné složky napětí/proudu	—
$k_0$	faktor nulové složky napětí/proudu	—
$\epsilon$	residuum (rozdíl hodnot originálu a modelu)	—

# Úvod

V dizertační práci se zabývám problematikou měření kvality elektrické energie, přesněji návrhem a realizací optimálnějších technik pro její dlouhodobé monitorování. Potřeba měřit parametry elektrické energie má nejméně dva klíčové důvody. Nekvalitní elektřina jednak způsobuje škody a nefunkčnost prvků distribuční sítě a jednak z pohledu liberalizace trhu s elektrickou energií vzniká potřeba ověřovat kvalitu dodaného/zakoupeného zboží - elektrické energie - mezi dodavateli a zákazníky. Kontinuálně sbíraná data jsou jedním ze základních vstupů k plánování optimalizace celé infrastruktury.

Problematika měření kvality elektrické energie je dlouhodobě zpracovávána zejména mezinárodními standardizačními organizacemi v úzké spolupráci s výrobci a spotřebiteli elektřiny. Cílem tohoto procesu je vytvoření specifikací pro provádění všech nezbytných měření tak, aby metodika byla co nejméně závislá na konkrétním řešení konkrétního dodavatele. Jednotlivá měření by měla být pokud možno nezávislé na použitém přístroji konkrétního výrobce. S tím souvisí i snaha umožnit provázání konkrétního měření fyzikálních veličin smluvními vztahy a garantovat důvěryhodnost prováděných měření.

Základní normou pro měření kvality je ČSN EN 50160 [13], definující klíčové ukazatele a principy měření. Norma odkazuje na další normy, předepisující vlastnosti měřicích přístrojů. Zejména se jedná o IEC 61000-4-7 [26] specifikující metody měření harmonických a mezipharmonických, IEC 61000-4-15 [24] popisující funkci a naznačující implementaci měřiče blikání a IEC 61000-4-30 [25] s definicemi a metodami určenými pro měření kvality napájecího napětí a odvozených veličin a ukazatelů. Pro měření energie jsou směrodatné normy IEC 62053-21 [22] pro měřidla činné energie a IEC 62053-23 [23] pro měření jalové energie. Pro celkovou definici funkcí měřicího přístroje pak slouží IEC 61557-12 [27]. Problematikou definice a měření různých výkonů se zabývá také IEEE 1459-2000 [2].

Konkrétním a poměrně závažným problémem systému monitoringu kvality dle výše zmíněných požadavků je efektivita archivace s ohledem na rychlou dostupnost dat a jejich rozumné využití pro další optimalizace a analýzy. V textu disertační práce jsou prezentovány některé techniky, které právě tuto problematiku řeší.



# Kapitola 1

## Cíle disertační práce

Dizertační práce se především zaměřuje na otázky související s problematikou archivace dat měření kvality elektrické energie. Hlavním cílem této práce je návrh korektního a optimálního rámce (systému) pro vyhodnocování kvality elektrické energie s důrazem kladeným na efektivní ukládání archivovaných dat i s ohledem na jejich další zpracování při pozdějším vyhodnocování měření.

Obsah a cíle práce lze shrnout do následujících bodů:

1. analýza požadavků na způsoby získávání, archivaci a vyhodnocení ukazatelů pro potřeby monitoringu PQ
2. implementace navrhovaného experimentálního měřicího systému a analytického software
3. ověření aplikovatelnosti a efektivity bezeztrátových algoritmů kompresních na vytvořených souborech měření PQ v PC
4. výběr efektivního bezeztrátového kompresního algoritmu pro použité embedded zařízení
5. návrh technik modelování hodnot v archivu měření za účelem zvýšení kompresního poměru
6. hodnocení navrhovaných technik modelování s ohledem na (a) dosažený poměr komprese a (b) využitelnost při následném zpracování archivovaných dat.

Dalším cílem doktorského studia bylo zanalyzovat požadavky na analyzátor kvality elektrické energie z hlediska platných norem i očekávání zákazníků. Navržený firmware analyzátoru bylo nutné implementovat na existující platformě s mikrokontrolérem ARM. Práce z této implementace analyzátoru kvality vychází, popisuje některé použité optimalizace a využívá data získaná měření uskutečněných s těmito přístroji pro další zpracování. Spolu s přístrojem vzniklo i programové vybavení pro analýzu záznamů PQ z analyzátoru, implementující popisované techniky v reálném prostředí se skutečnými daty.

Zabývá se vlivy použití různých technik předzpracování archivních dat na dosažený kompresní poměr při archivaci reálných měření. Kombinací kompresních algoritmů a vhodných technik předzpracování množiny hodnot veličin (modelování/predikce hodnot časové řady měření) je možné zlepšit celkový kompresní poměr dat měření. Zároveň správně zvolený model časové řady měření lze beze ztráty obecnosti využít pro některé následné analýzy kvality elektrické

energie a výrazně tak uspořít výpočetní prostředky. Modelování hodnot měřených veličin pak v neposlední řadě umožňuje navrhnout nové metody pro analýzu a prezentaci dat obsluze, čímž by mohl v budoucnu být narušen dnes běžný trend data sbírat, ukládat a dále je jen víceméně ignorovat. Navržené matematické modely usnadňují návrh automatických metod pro hodnocení dat a data mining rozsáhlých množin měření.

### **Zdůvodnění významu očekávaných výsledků:**

Aktuální rychlé rozšiřování chytrých prvků na všech úrovních energetické infrastruktury (chytré elektroměry, analyzátoři kvality, autonomní ochrany, centrální sběr dat) vznikají enormně velké databáze měření s poměrně málo využívaným potenciálem pro další analýzy. Běžně taková databáze obsahuje výrazné množství informací naprosto zbytných pro vyhodnocování vlivu měřených jevů na bezproblémový provoz sítě. Současná běžná praxe přitom definuje víceméně jen agregace v pevně normou daných intervalech jako jedinou možnou formu decimace získaných PQ dat. Jedním z výstupů práce je srovnání různých technik pro (bezeztrátovou) kompresi těchto agregovaných dat.

V první části práce jsou popisovány realizované postupy pro efektivní a korektní vyhodnocování PQ a další zpracování charakteristických veličin při vyhodnocování kvality elektrické energie [29, 43, 45, 30]. Uvedené optimalizace algoritmů v embedded systému umožňují realizaci celého měřicího přístroje třídy S na hardware s celkově nižšími nároky (což je v přímém kontrastu s většinou výrobců takových systémů) nebo případně rozšíření funkcí přístroje o nová vyhodnocení. Za předpokladu budoucího masového nasazení monitorů kvality v tzv. smart grids je žádoucí vyvíjet levná řešení monitoringu s minimální vlastní spotřebou, aby případný přínos uplatnění těchto monitorů v praxi nebyl snadno zpochybnitelný argumentem jejich vlastní spotřeby.

Pro ukládání měřených dat jsou primárně využity binární soubory v proprietárním formátu CEA a nebo SQL databáze aplikace ENVIS. Data z obou těchto zdrojů lze přímo vyhodnocovat a vizualizovat v aplikaci ENVIS. Do databáze ENVIS lze prostřednictvím CSV souborů importovat i data jiných analyzátorů třetích stran pro případné ověřování jejich vlastostí.

Praktická implementace popisovaných technik přináší výraznou úsporu potřebných diskových kapacit, což v případě rozsáhlých systémů například distribuční společnosti výrazně sníží pořizovací i provozní náklady celého systému. I pro popisovaný embedded systém s omezeným výpočetním výkonem je možné využít vhodně optimalizovanou kompresi. Kompresi implementovaná již na úrovni embedded systému šetří kromě permanentní paměti přístroje i čas potřebný pro přenos záznamů z přístrojů do PC po komunikačním kanále.

Na popisovaných postupech je pomalá obvykle tvorba optimálního modelu a následná komprese modelovaných dat. Naopak dekomprese a evaluace modelu není z pohledu běžných PC nijak výpočetně náročná a je dostatečně rychlá. Vhodně implementovaný model a komprese dat proto zrychluje častější operaci načítání dat z databáze.

Techniky konstrukce predikčních modelů pro zlepšení komprese v některých případech respektují charakter dat a požadavky na korektní vyhodnocení PQ ukazatelů a tudíž je lze bez úhony na obecnosti využít pro rychlejší a méně paměťově náročnou analýzu archivovaných dat.

## Kapitola 2

# Stav řešené problematiky

Kvalita elektrické energie je pojem, popisující stavy a události v elektrické rozvodné síti. Určování kvality elektrické energie je ve své podstatě problém elektromagnetické kompatibility (EMC) a klade si za cíl určit zda parametry energie dodané ze sítě jsou slučitelné s připojenými zařízeními. Zabývá se zejména sledováním a vyhodnocováním jevů, jejichž důsledky mohou poškodit připojené spotřebiče nebo narušit správnost jejich funkce, případně snižují účinnost využití dodané energie. Ideální stav v systémech se střídavým napětím lze definovat poměrně jednoznačně – představuje ho kontinuální dodávka harmonického napětí nominální frekvence s nulovou impedancí pro všechny frekvence.

Hlavní snahou mezinárodních organizací jako např. IEEE, IEC, EPRI, CIRED a jiných je dosáhnout maximální možné míry interoperability používaných přístrojů a sjednocování způsobů deklarace a specifikace použitých funkcí, algoritmů a metodik pro vyhodnocování měření. Předními světovými výrobci monitorů kvality el. energie jsou firmy ELSPEC, Dranetz-BMI, Schneider, Siemens, ABB, Janitza a další. V České republice se touto problematikou hlouběji zabývá například katedra elektrických měření VŠB TU Ostrava (ve spolupráci s firmou ELCOM) a katedra elektroenergetiky na ČVUT FEL v Praze, firma Mega, EnerGoConsult, EGU atd. V tomto oboru vyvíjí svá řešení i firma KMB systems, se kterou je tato dizertační práce v rámci úzké spolupráce řešena.

Mezinárodní standardy pro EMC a vyhodnocování kvality definují mnoho parametrů a způsob jejich určování včetně rozsahů a tolerancí a způsobu vyhodnocení. Specifikace PQ dále definují výstupy měření a způsob odezvy měřicího systému v situacích, kdy měření některé ze základních veličin nelze korektně uskutečnit.

### 2.1 Datová úložiště

Data v systémech vyhodnocování kvality elektrické energie se nacházejí obecně v (a) paměti měřicího přístroje, (b) v komunikačním kanále, (c) v databázi nebo (d) souboru nadřazeného systému. Hlavní důraz této práce je kladen na optimalizaci variant c a d. Varianta a a b jsou zmíněny pouze okrajově, nicméně jedná se o podobnou problematiku, kde jediným přidaným kritériem je nižší výkon embedded systému.

Ukládání dat v souborech na PC se v praxi realizuje obvykle v proprietárních souborech, jejichž struktura není veřejná. Výjimkou je například formát PQDIF (v aktuální verzi 1.5, [1, 15]),

který je otevřený, zdokumentovaný a podporovaný několika různými přístroji a programy různých firem. Kompresi formát přímo nepodporuje, ale autor v roce 2009 krátce polemizoval s využitím RAR nebo ZIP komprese[39]. Podporu aktuální verze formátu PQDIF v třetích aplikacích umožňuje autorova vlastní zveřejněná a zdokumentovaná knihovna. S využitím této knihovny je základní podpora formátu PQDIF doplněna jako modul i v aplikaci ENVIS v1.1.

Další otevřený formát je tzv. IEEE Standard Common Format for Transient Data Exchange for Power Systems, uzpůsobený zejména pro výměnu informací o událostech a přechodových jevech. V českém prostředí standardizaci datových toků nabízí produkt PAQNET, bohužel bez bližší dohledatelné a dostupné dokumentace či otevřené implementace je pro potřeby mé práce nepoužitelný. Nicméně například vstup dat do poměrně rozšířené databáze systému pro archivaci a hodnocení měření kvality elektřiny DAM probíhá i dnes na bázi nespolehlivého a neefektivního importu textových nebo csv souborů. Zajímavé a nestandardní řešení představuje ztrátový patentovaný formát PQZIP, který dle výrobce dosahuje kompresního poměru až 10000:1 při rozumném zachování kvality informace.

## 2.2 Algoritmy bezztrátové komprese

V této práci se zejména zabývám aplikací vhodných bezztrátových kompresních algoritmů na data měření kvality el. energie. Teoretické základy kompresních algoritmů položil Shannon ve své práci [42], v níž mimo jiné definuje pojem informační entropie a kompresní poměr.

Mnou popisované metody využívají různé slovníkové a entropické algoritmy bezztrátové komprese. Podrobný popis nejčastěji používaných algoritmů je uveden v [40] v kapitolách 5 a 6. V experimentální části této práce byly použity pouze implementace s otevřeným zdrojovým kódem, u kterých je dostatečně ověřena spolehlivá funkce - jsou využívány v reálných aplikacích. Dalším kritériem byla i dostupnost konkrétního zdrojového kódu algoritmu s možností dalších případných modifikací<sup>1</sup>. Pro nasazení v měřicím přístroji v prostředí embedded systému s velmi omezenými prostředky jsou důležité i praktické nároky konkrétního algoritmu na systémové prostředky.

Komprimované soubory obsahují zejména časové posloupnosti veličin kódovaných jako 8, 16, 32 nebo 64 bitové celočíselné proměnné (integer) a 32 či méně časté 64 bitové proměnné v plovoucí řádové čárce (Single, Double). U celočíselných typů je i při kompresi a modelování řad zachován princip úplné bezztrátovosti. S ohledem na vlastnosti čísel s plovoucí řádovou čárkou toto ani při základních aritmetických operacích nelze zaručit[19]. Z tohoto důvodu operace s veličinami v plovoucí řádové čárce jsou realizovány tak, aby výsledná chyba byla řádově menší, než požadovaná přesnost konkrétní veličiny.

Pro reprezentaci desetinných čísel (např. výkony) v archivu PQ je také možné využít kódování v tzv. pevné řádové čárce (fixed-point arithmetic). To v podstatě představuje celočíselnou aritmetiku se všemi jejími výhodami. Tato reprezentace je především rychlá a deterministická. Při použití pevné čárky pro kódování je nutné kontrolovat rozsahy čísel a ošetřovat přetečení.

<sup>1</sup>Zdrojový kód použité knihovny je uvolněn pod některou z tzv. open source licencí.

**Entropické kodéry** vytvářejí kódy znaků, které reflektují četnost výskytu kódovaných znaků v původním textu a vytvářejí pro častěji se vyskytující data kratší binární sekvence. Pokročilé entropické algoritmy jsou tzv. Shannon-Fano (SF) a Huffmanovo (HUF, [21]) kódování. Obě metody mají proměnnou délku znaku a častější symboly kódují kratším počtem bitů. Oba algoritmy mají stejnou asymptotickou složitost, nicméně HUF obvykle dosahuje lepší komprese, protože vždy vytváří kódování, které je optimální.

Mezi entropické kodéry patří i aritmetické kódování (AC, [3]) a rychlejší Range kódování (RC, [12]). Tyto algoritmy efektivněji komprimují soubory, kde pravděpodobnosti výskytu jednotlivých znaků nejsou zlomkem mocniny dvou (HUF každému znaku přiřazuje binární kód délky  $N$ ). Varianty HUF a nebo RC bývají obvykle integrovány jako vstupní nebo výstupní modul většiny používaných algoritmů komprese.

**Slovníkové metody** obecně nahrazují posloupnosti symbolů kratšími posloupnostmi (vytvářejí slovník). Některé implementace namísto fixního slovníku zpětně referují na výskyty stejných sekvencí v předchozím toku dat. Jejich výhodou je obvykle snadná a rychlá dekomprese. Nevýhodou pak jsou obecně větší paměťové nároky na konstrukci slovníku a s tím související pomalejší komprese. Základním algoritmem této skupiny je Lempel-Ziv (LZ, [47]). LZ metody jsou nejčastěji aplikovány a vykazují nejlepší dosažené výsledky pro kompresi textových souborů. Optimalizované varianty tohoto algoritmu se obvykle liší způsobem tvoření slovníku, jeho velikostí a několika dalšími drobnými vylepšeními. Existuje mnoho variant LZ algoritmů a mnoho z nich je nebo bylo chráněno patenty.

**Moderní algoritmy** kombinují obvykle některou LZ variantu s modelováním pro predikci následných symbolů [11, 16, 17]. Praktické implementace tohoto algoritmu jsou kodéry BZip2 a LZMA (7-Zip). Součástí LZMA je vestavěný Range kodér, který zaručuje optimální kódování symbolů s nízkou entropií. LZMA je také optimalizované na rychlou implementaci dekodéru. V různých odvětvích bylo LZMA úspěšně vyhodnoceno jako optimální bezztrátový kompresní algoritmus, viz např. [6, 44].

Nejnovější série algoritmů z rodiny PAQ [34] využívá predikci s využitím tzv. kontext mixing, které vhodným způsobem kombinuje výstupy různých predikčních modelů (ne nezbytně optimálních) posloupnosti následujících symbolů, kterým je zpravidla jediný bit. Některé algoritmy z této rodiny dosáhly rekordní komprese textových i jiných souborů, nicméně jejich velkou nevýhodou je extrémní náročnost na paměť a výpočetní výkon, která tyto algoritmy diskvalifikuje pro využití v praktických aplikacích.

## 2.3 Typy archivovaných dat

Problematika měření kvality elektrické energie se dělí na samostatné kategorie ověřovaných parametrů; kvality. Díky rozmanitosti těchto úloh je nutné z hlediska volby optimálního způsobu komprese jednotlivých kategorií k těmto přistupovat samostatně.

**Periodické záznamy (agregace)** představují obvykle hlavní výstup měření. Většinou jsou tyto archivy nejrozsáhlejší a tak každá jejich optimalizace přináší nejvýraznější zlepšení využití

dostupných prostředků. Vzhledem k nenáhodnému charakteru většiny v něm obsažených veličin je vhodná pro aplikaci některých statistických kompresních algoritmů. Tato práce se zabývá zejména kompresí archivu periodických odečtů.

**Odečty elektroměru** jsou charakteristické tím, že hodnoty ukládaných registrů v typickém případě neklesají. Z hlediska objemu dat se při monitorování kvality obvykle nejedná o kritický archiv. Diametrálně odlišná situace však nastává v jiné aplikaci - tzv. *smart meteringu* a *smart grids* obecně, kde jsou data sbírána z velkého množství měřidel.

**Vyhodnocení měření dle EN 50 160** představuje cyklické záznamy stavu čítačů a booleanových proměnných, které určují na základě přednastavených hodnot splnění či nesplnění jednotlivých kritérií PQ dle EN 50 160.

**Statistické vyhodnocení událostí** vytváří archiv záznamů, které obsahují dvojici časů a jednu nebo více agregovaných hodnot charakterizujících tzv. událost.

**Krátkodobé průběhy efektivních hodnot** zaznamenávají posloupnosti efektivních hodnot sledovaných veličin v krátkých intervalech v průběhu detekované události.

**Oscilogramy** obsahují hodnoty jednotlivých vzorků sledované veličiny V literatuře [41, 38] jsou popsány různé metody pro ztrátovou kompresi těchto archivů, využívající obvykle 2D Fourierovu či wavelet transformaci a následnou normalizaci spektra.

**Provozní deník (LOG)** ukládá do odděleného paměťového prostoru stavové, analytické, ladicí a jiné pomocné informace.

Prakticky lze rozlišit problematiku archivace měření na ukládání dat při měření ve vlastním měřicím přístroji (kap. 4.1) a archivaci či vyhodnocování (tzv. postprocessing) měření v počítači (kap. 4.2). Tyto dvě platformy se liší jak obecnými požadavky na podsystém archivace tak i dostupnými prostředky a výpočetním výkonem platformy.

## 2.4 Komprese hlavního archivu analyzátoru

Obecný přehled dat popisujících kvalitu energie a možností použití kompresních algoritmů je stručně uveden v [7]. Paradoxní je, že tento i většina dále zmiňovaných článků se soustředí zejména na možnosti komprese záznamů přechodových jevů, událostí a dalších krátkodobých záznamů (popisované v kapitole 2.3). Z pohledu běžných analyzátorů se přitom obvykle jedná o nadstandardní funkce archivu a naopak klíčovou a běžnou vlastností je ukládání archivů agregovaných hodnot měřených veličin.

# Kapitola 3

## Měření veličin a použité vztahy

### 3.1 Základní veličiny

**Efektivní hodnota napětí**

$$U_{eff} = \sqrt{\frac{1}{T} \int_0^T u^2(t) dt} \quad (3.1)$$

Pro případ periodického signálu lze dále zavést následující vztahy pro okamžitou hodnotu napětí  $u(t)$ , napětí první harmonické  $u_1(t)$  a tzv. harmonické napětí  $u_H(t)$ :

$$u(t) = U_0 + \sum_{n=1}^{\infty} U_n \sin(n\omega t + \varphi_n) \quad (3.2)$$

$$u_1(t) = U_1 \sin(\omega t + \varphi_1) \quad (3.3)$$

$$u_H(t) = \sum_{n=2}^{\infty} U_n \sin(n\omega t + \varphi_n) \quad (3.4)$$

kde  $U_0$  je stejnosměrná složka a druhý člen představuje součet harmonických složek, kde  $U_n$  je amplituda  $n$ -té harmonické složky signálu a  $\varphi_n$  je fázový posun příslušné složky. Pro diskrétní signál lze dále zavést vztahy pro fázové ( $U_{LN}$ ) a analogicky i sdružené ( $U_{LL}$ ) napětí.  $U_{Li}$  představuje  $i$ . vzorek fázového napětí příslušné fáze,  $U_{Ni}$  představuje  $i$ . vzorek napětí ochranného vodiče a  $N$  je počet vzorků signálu. Pro  $U_{LN}$  platí:

$$U_{LN} = \sqrt{\frac{\sum_{i=0}^{N-1} (U_{Li} - U_{Ni})^2}{N}} \quad (3.5)$$

**Proudy** jsou vyhodnocovány ve stejném základním cyklu jako napětí a spolu s ním jsou využívány k výpočtu dalších veličin. Pro efektivní hodnotu proudu platí:

$$I_{eff} = \sqrt{\frac{1}{T} \int_0^T i^2(t) dt} = \sqrt{\frac{\sum_{i=0}^{N-1} I_i^2}{N}} \quad (3.6)$$

Pro rozklad na jednotlivé harmonické složky opět analogicky platí následující vztah, vyjadřující hodnotu proudu v čase  $t$  pomocí součtu jednotlivých složek:

$$i(t) = I_0 + \sum_{n=0}^{\infty} I_n \sin(n\omega t + \varphi_n) \quad (3.7)$$

**Elektrický výkon** je v podstatě svázán s přenosem náboje, který je nositelem energie, ze zdroje do soustavy. Náboj, který projde soustavou za dobu  $dt$  je  $i(t)dt$ , způsobí pokles potenciálu o hodnotu  $u(t)$ . Tím dojde k poklesu potenciální energie

$$dE_p = dq(t)du(t) = i(t)u(t)dt \quad (3.8)$$

$$E_p = \int_{t_1}^{t_2} i(t)u(t)dt \quad (3.9)$$

kde  $i(t)u(t) = p(t)$  je okamžitý výkon.

Zavedení různých systémů definice elektrických výkonů vychází z požadavků nalézt přesný a jednoduchý popis vztahů mezi napětím a proudem v časové rovině. Hlavním cílem zavedení libovolné takové ucelené teorie tedy bývá snaha o:

1. vysvětlení fyzikálních zákonitostí spojených s přenosem energie
2. definice veličin, které popisují tok energie, efektivitu jejího využívání a možnosti škálování různých síťových zařízení
3. vytvoření aparátu pro vývoj zařízení, která obecně vylepšují míru využití energie, případně pro dimenzování kapacity infrastruktury sítě
4. položení základních vztahů mezi odběrateli a dodavateli el. energie

Popisem vhodného systému vztahů (teorie) pro měření výkonů v analyzátoru kvality se zabývá hned několik standardů a publikací [2, 22, 23, 27], ačkoliv přímo součástí problematiky vyhodnocování kvality (napětí) zatím není, stejně jako např. problematika monitoringu proudů. V následujícím textu jsou zavedeny jednotlivé veličiny, které je možné měřit a vyhodnocovat pomocí platformy analyzátoru SMP a vycházející z požadavků výše zmíněných standardů.

Pro konkrétní výpočet je vždy nutné z obecné definice odvodit její diskrétní podobu vhodnou pro implementaci v mikroprocesoru. Mikroprocesor analyzátoru vzorkuje téměř synchronně vůči proměnlivé frekvenci napětí a proud se zanedbatelnou odchylkou, v našem případě pravidelných intervalech. Výsledné průběhy ukládá do pole  $N$  vzorků, které je dále numericky zpracováno. Frekvence je měřena průběžně v cca vteřinových intervalech a rychlost vzorkování je přizpůsobována změnám frekvence. V textu práce jsou zavedeny diskrétní vztahy pro výpočet konkrétních výkonů (okamžitý, zdánlivý, činný, jalový, neaktivní, deformační a další). Práce také popisuje problematiku definice a měření výkonů ve vícefázových soustavách.

**Frekvence sítě** se měří v intervalu 10 s, její změna v tomto intervalu ovlivňuje všechny ostatní právě měřené veličiny, které je pak třeba označovat příznakem. Frekvence je měřena pomocí čítače průchodů nulou. Při velmi zarušeném signálu je metoda ošetřena proti čítání nadbytečných pulzů pomocí vhodně dimenzované filtrace signálů rušivých frekvencí v hardware i firmware přístroje. Fázový závěs pro převodníky automaticky zajišťuje synchronní vzorkování U a I.



**Účinit**  $\lambda$  definuje fiktivní fázový posuv harmonického proudu a napětí, který by měly stejné efektivní hodnoty a činný výkon. Protože používá celkové  $S$  a celkové  $P$ , zahrnuje v sobě i vliv vyšších harmonických v případech, kdy je signál obsahuje (platí obecně i pro neharmonické průběhu napětí a proudu).

$$\lambda = \frac{|P|}{S} \quad (3.10)$$

**Cos  $\varphi$**  využívá definice činného a jalového výkonu první harmonické ( $P_1$  resp.  $Q_1$ ) a vyjadřuje jejich vzájemný fázový posun.

$$\begin{aligned} |\cos\varphi| &= \frac{|P_1|}{S_1} \\ \cos(\varphi) &= \cos\left(\arctan\left(\frac{Q_1}{P_1}\right)\right) \end{aligned} \quad (3.11)$$

**Celkové harmonické zkreslení THD** opisuje souhrnně obsah všech složek napětí nebo proudu.

$$THD_U = \sqrt{\frac{\sum_{n=2}^{50} U_n^2}{U_1^2}} = \sqrt{\frac{U_{ef}^2}{U_1^2} - 1} = \sqrt{\sum_{n=2}^{50} u_n^2}, \quad u_n = \frac{U_n}{U_1} \quad (3.12)$$

$$THD_I = \sqrt{\frac{\sum I_{ef}^2}{I_1^2} - 1}, \dots \quad (3.13)$$

**Napěťová nesymetrie** představuje jev ve více-fázových sítích, který může významně ovlivnit míru využití dodané elektrické energie [20]. Trvalá nebo dlouhodobá nesymetrie (ustálený stav) vzniká zejména vlivem rozdílné zátěže jednotlivých fází při spotřebě, krátkodobá nesymetrie nastává při vzniku zemních spojení a jiných přechodových jevů. Podobně jako definujeme symetrii fázových napětí, můžeme definovat a vyhodnocovat i symetrii proudů [37].

V Evropských normách je zakotven způsob, který vychází z tzv. teorie symetrických komponent [18, 5, 3]. Tato metoda matematicky rozkládá nesymetrický systém na tři symetrické složky - souslednou ( $U_1, V^+$ ), zpětnou ( $U_2, V^-$ ) a nulovou složku ( $U_0, V^0$ ). Pomocí tohoto rozkladu je zaveden faktor zpětné složky (napětí, proudu)  $k_1 = \frac{U_2}{U_1}$  a faktor nulové složky (napětí)  $k_0 = \frac{U_0}{U_1}$ . Takto zavedená nesymetrie se v literatuře často nazývá též VUF nebo IEC definice.

$$\bar{U}_1 = \frac{1}{\sqrt{3}}(\bar{U}_{12} + a\bar{U}_{23} + a^2\bar{U}_{31}) \quad (3.14)$$

$$\bar{U}_2 = \frac{1}{\sqrt{3}}(\bar{U}_{12} + a^2\bar{U}_{23} + a\bar{U}_{31}) \quad (3.15)$$

$$\bar{U}_0 = \frac{1}{\sqrt{3}}(\bar{U}_{12} + \bar{U}_{23} + \bar{U}_{31}) \quad (3.16)$$

$$a = e^{\frac{j2\pi}{3}} = -\frac{1}{2} + i\frac{\sqrt{3}}{2} \quad (3.17)$$

Nulová složka  $V_0$  v tří-vodičových sítích nemůže působit a proto má faktor  $k_0$  pro analýzu PQ v 3-f sítích velmi malý praktický význam. V praxi se obvykle vyhodnocuje pouze faktor  $k_1$ , který lze zjednodušeně určit také z následujících vztahů (odvození viz. např. [46]) pouze se znalostí amplitud sdružených napětí (1. harmonické).

$$k_1 = VUF = \frac{U_2}{U_1} = \sqrt{\frac{1 - \sqrt{3 - 6\beta}}{1 + \sqrt{3 - 6\beta}}} \times 100 [\%] \quad (3.18)$$

$$\beta = \frac{(U_{12}^4 + U_{23}^4 + U_{31}^4)}{(U_{12}^2 + U_{23}^2 + U_{31}^2)^2} \quad (3.19)$$

## 3.2 Hodnocení kvality elektrické energie

**Frekvence** je základní veličina, v analyzátorech dle IEC norem musí být měřena kontinuálně a vypočítávána každých 10 s.

$$f_{din} = \frac{\sum PulsPřiPrůchoduNulou}{2t} [Hz] \quad (3.20)$$

Přístroj pro 50 Hz síť musí správně fungovat a měřit při frekvenci 42,5 Hz až 57,5 Hz. Pro nominální frekvenci 50 Hz je většina dalších veličin ( $U, I, P, Q, S, D, \dots$ ) vyhodnocována z 10-ti periodového intervalu vzorkování a výsledné hodnoty jsou vstupem pro další dlouhodobé průměrování. Princip kontinuálního měření vyžaduje, aby za žádné běžné situace nemohlo dojít k výpadku vzorkování některé z těchto hodnot. Statistické vyhodnocení ukazatelů kvality se pak provádí na definovaných intervalech (většinou týden, rok) a pro zadané percentily.

**Efektivní hodnoty napětí** se vypočítávají v 10-ti periodových intervalech ukládají se do hlavního archivu měření. Pro agregaci hodnot je předepsáno použití kvadratického průměrování. U desetiminutového intervalu začíná vzorkování na časové značce 10 min přesně. Při další časové značce okamžitě začne měřit nový interval (třída A) a zároveň dokončí výpočty aktuálního měření a ukončí jej. Třída S a B může použít i jednodušší metody synchronizace. Normou nespecifikované intervaly lze přizpůsobit dle výše uvedených principů.

**Harmonické rušení** představuje jakoukoliv odchylku napětí a proudu od harmonického ustáleného (ideálního) stavu. Pro potřeby PQ monitoringu bývá nejčastěji vyhodnocováno harmonickou analýzou v základním pásmu do 2 KHz, což představuje 50. harmonickou. Výpočet používá DFT nebo speciální FFT[14] na navzorkovaných deseti periodách<sup>1</sup> signálu pro  $f_{nom} = 50 Hz$ . Pro archivaci a vyhodnocování se používá agregace v čase (podobně jako u běžných veličin) a tzv. grupování.

Trend v analýze harmonického rušení dle standardů IEC směřuje i k měření a archivaci harmonických složek (napětí) do 9 kHz [8] stejným způsobem jako je dnes prováděno do 2 kHz

<sup>1</sup>Základní harmonická v matematickém smyslu tohoto pojmu je při faktickém měření deseti period signálu 5Hz a vyšší harmonické jsou ve výsledném spektru rozmístěny po 5Hz. V reálném systému se složky s  $f_h < f_{NOM}$  se nazývají subharmonické,  $f_{NOM} = 50 Hz$  resp. 60 Hz, složky s  $f_h = N \times f_{NOM}$  jsou tzv. harmonické, složky s  $f$  se nazývají mezi-harmonické.

a zkoumány jsou i vhodné inovativní metody agregace pro kontinuální analýzu spektra řádově až do stovek  $kHz$  viz např. [32].

Data o harmonickém rušení představují největší objem archivovaných dat a ve všech třech uvedených případech je nutné používat agregaci a případně selekci pro ukládání pouze relevantních dat. Současný stav norem jen minimálně odráží potřeby dalšího praktického využití informací o harmonickém rušení a existuje velký prostor pro další studium a optimalizace tohoto procesu.

**Nesymetrie třífázové soustavy** se agreguje v intervalu dle hlavního archivu. Z hlediska elektromagnetické kompatibility se sleduje a vyhodnocuje VUF pro napětí a PPDS vyžaduje také měření zpětné složky proudu  $i_2$  a její úhel. Platforma SMP měří a archivuje VUF, zpětnou složku napětí i proudu a úhel zpětné složky proudu. Vliv těchto veličin na velikost celkového archivu respektive jeho kompresní poměr je zanedbatelný a dále se jimi práce specificky nezabývá.

**Blikání** (flicker) obecně představuje rušivý vjem, způsobený nestálostí osvětlení prostředí. Je způsoben cyklickými změnami intenzity osvětlení, způsobené změnami napájecího napětí svítidla. Vliv těchto fluktuací závisí na jejich amplitudě, periodě opakování a jejich dalších charakteristikách. Měření vlivu flickeru je komplexní proces, modelující reakci lidského oka a mozku na změny osvětlení.

**Rychlé změny napětí** (RVC, Rapid Voltage Changes) představují relativně novou charakteristiku kvality napětí, jejíž definice stále není pevně ustálená a sjednocená konkrétní mezinárodní normou. Parametry výpočtu a implementace je podrobně popsána v [9].

**Události napětí**, např. překročení, pokles nebo úplná ztráta napětí je považováno za napěťovou událost. Vyhodnocuje se z měřených efektivních hodnot za půl periody. Hlubší náhled na ovlivňování různých typů spotřebičů různými jevy je uveden v [35]. Technicky se rozlišuje krátký pokles napětí (dip,sag), přerušování napájení a přepětí. Pro potřeby analýzy těchto jevů se zaznamenává čas začátku a konce události a extrém dosažené hodnoty. K podrobnější analýze problému lze využít kontinuální záznam události formou odečtů efektivních hodnot napětí (a proudu) v půlperiodě – archiv PQ Event Trend. Nejpodrobnější záznam události pak představuje tzv. oscilogram.

**Třífázové události** jsou definovány na základě záznamu jednofázových událostí. Vyhodnocují se ze všech měřených fází najednou, kdy událost začíná výskytem na kterékoliv fázi a končí v okamžiku, kdy na všech fázích je opět normální stav.

## Kapitola 4

# Optimalizace archivů

Hlavním cílem popisovaných experimentů s generovanými i skutečnými daty bylo zvolit optimální kompresi s ohledem na dosahovanou účinnost komprese (CR), stabilitu dosažené účinnosti, spolehlivost a také vzhledem k aplikovatelnosti té které metody v prostředí embedded systému s omezenými výpočetními prostředky a minimem dostupné operační paměti.

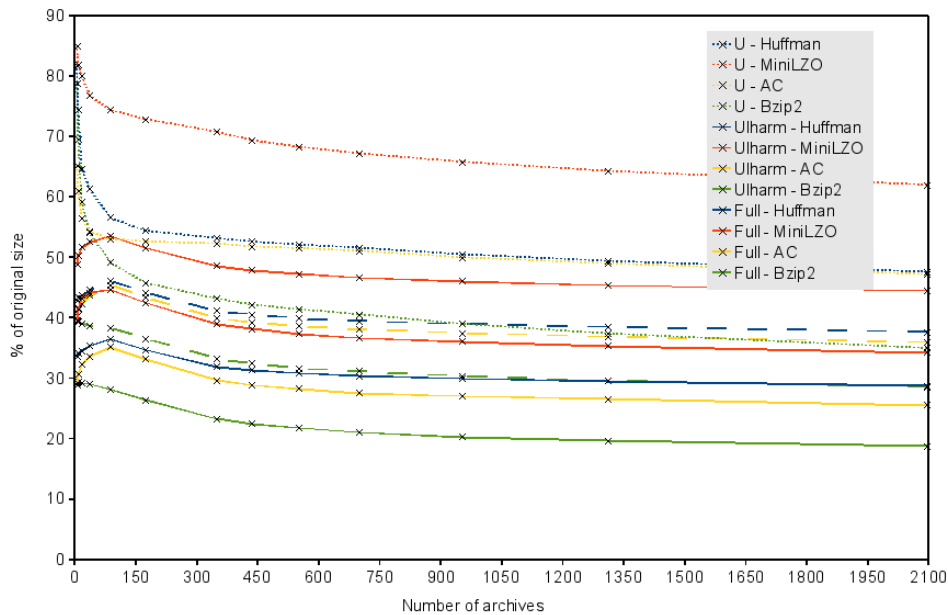
Pro účely testování různých algoritmů byl vytvořen modulární generátor PQ archivů [10, 28, 31], který lze snadno modifikovat pro potřeby experimentů. Tento generátor byl použit jako výchozí nástroj pro benchmark RLE, různých variant LZ kompresí, Huffmanova kódování a bzip2. Výsledky provedených experimentů jsou porovnány s ohledem na spotřebu paměti v průběhu komprese, na dobu komprese, složitost implementace a dosažené účinnosti v [28]. Další experimenty byly prováděny na datech získaných z měření reálnými analyzátory SMPQ a SIMON.

### 4.1 Archivace dat v analyzátoru

Zefektivnění formátu ukládání dat je realizováno na úrovni kódování jednotlivých veličin s ohledem na jejich rozsah a požadovanou přesnost. Další úspory je možné dosáhnout i s využitím vrstvy pro kompresi kódovaných záznamů. Výkonnost těchto algoritmů byla experimentálně ověřena na testovacích i skutečných datech. Výsledky experimentů jsou publikovány v [28, 31]. Jedním z podstatných závěrů je vzhledem k výraznému omezení operační paměti i výpočetního výkonu doporučení pro využití méně efektivních ale nenáročných kompresních algoritmů - například aritmetického kódování nebo kompresní algoritmus MiniLZO.

### 4.2 Archivace měření v PC

V publikacích [28, 31, 10] byl jako jeden z hlavních sledovaných parametrů testován vliv množství vstupních dat na účinnost komprese. Bylo ukázáno že pro běžné algoritmy lze pouze odhadnout optimální velikost vstupního souboru. Agregace dat (typicky 2 hodiny, 10 minut, 1 minuta, 150 period, 10 period) způsobuje vyhlazení krátkodobého kolísání. Tyto delší agregace se obvykle archivují spolu s původními daty a slouží tak i jako jeden možný (průměrovací) model pro eliminaci entropie originálních dat.



Obrázek 4.1: Účinnost různých kompresních algoritmů při aplikaci na soubor reálných dat [31].

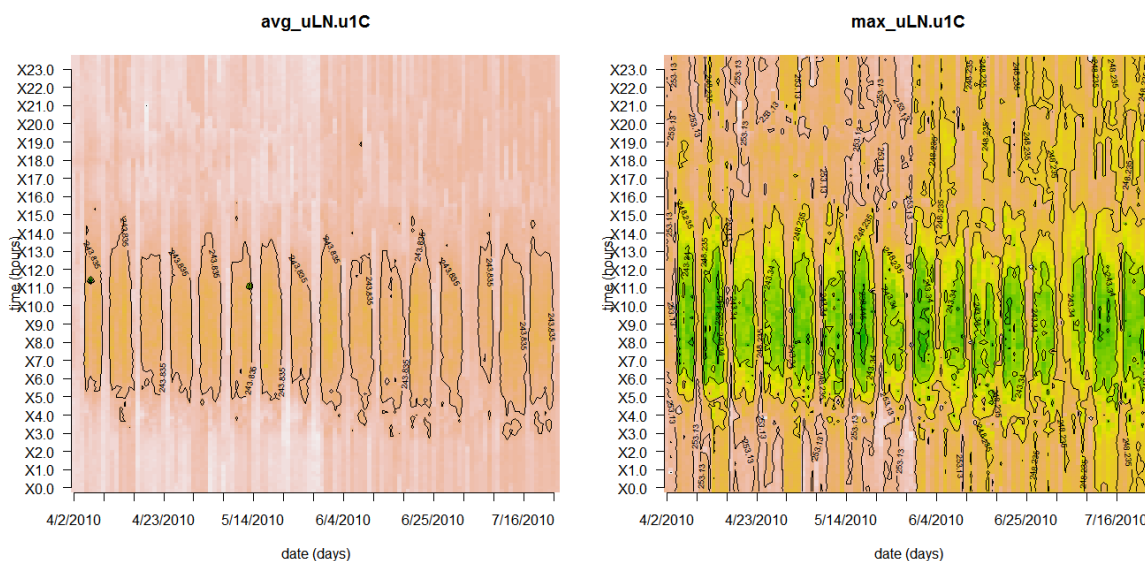
#### 4.2.1 Binární soubor jako zdroj dat

Původní soubory pro archiv měření (formát CETIS) představují sekvenci binárních záznamů konfiguračních struktur v hlavičce souboru a bloky záznamů měření. Data nejsou komprimována. Formát RETIS vychází z výše popsaného formátu, doplňuje pouze strukturovanou vnitřní hierarchii v rámci souboru. S pomocí datové vrstvy organizuje jednotlivé záznamy po skupinách ve virtuálních podadresářích. Datová vrstva analytického software ENVIS představuje úložiště, které umožňuje rychlé vyhledávání provázaných informací s ohledem na implementované rozhraní. Optimalizace velikosti archivu je realizována s využitím volitelného algoritmu komprese souboru.

#### 4.2.2 SQL Server jako zdroj dat

SQL databáze jsou nejvyužívanější technologií pro ukládání rozličných dat a popis vzájemných vztahů v datovém souboru. ENVIS podporuje SQL Server 2008 firmy Microsoft. Využívá řádkovou a stránkovou kompresi a podporu FILESTREAM. Cílem provedených experimentů bylo vyhodnotit vliv těchto technik na objem ukládaných dat a rychlost načítání. Podrobněji se touto problematikou zabývá práce [36].

Kompresi na úrovni SQL serveru může být použita na tabulkách (bez indexů, s klastrovanými a neklastrovanými indexy), indexovaných pohledech i samotných indexech. ROW (řádková) komprese ukládá hodnoty typů s fixní délkou jako typy s proměnnou délkou. PAGE (stránková) komprese minimalizuje místo obsazené opakujícími se hodnotami v jednotlivých sloupcích na stránce. Funkce FILESTREAM umožňuje ukládat větší bloky binárních dat jako binární soubory přímo do filesystému a na úrovni SQL databáze s nimi pracovat s pomocí odkazů.



Obrázek 4.2: Příklad rozložení průměrných a maximálních hodnot v delším intervalu.

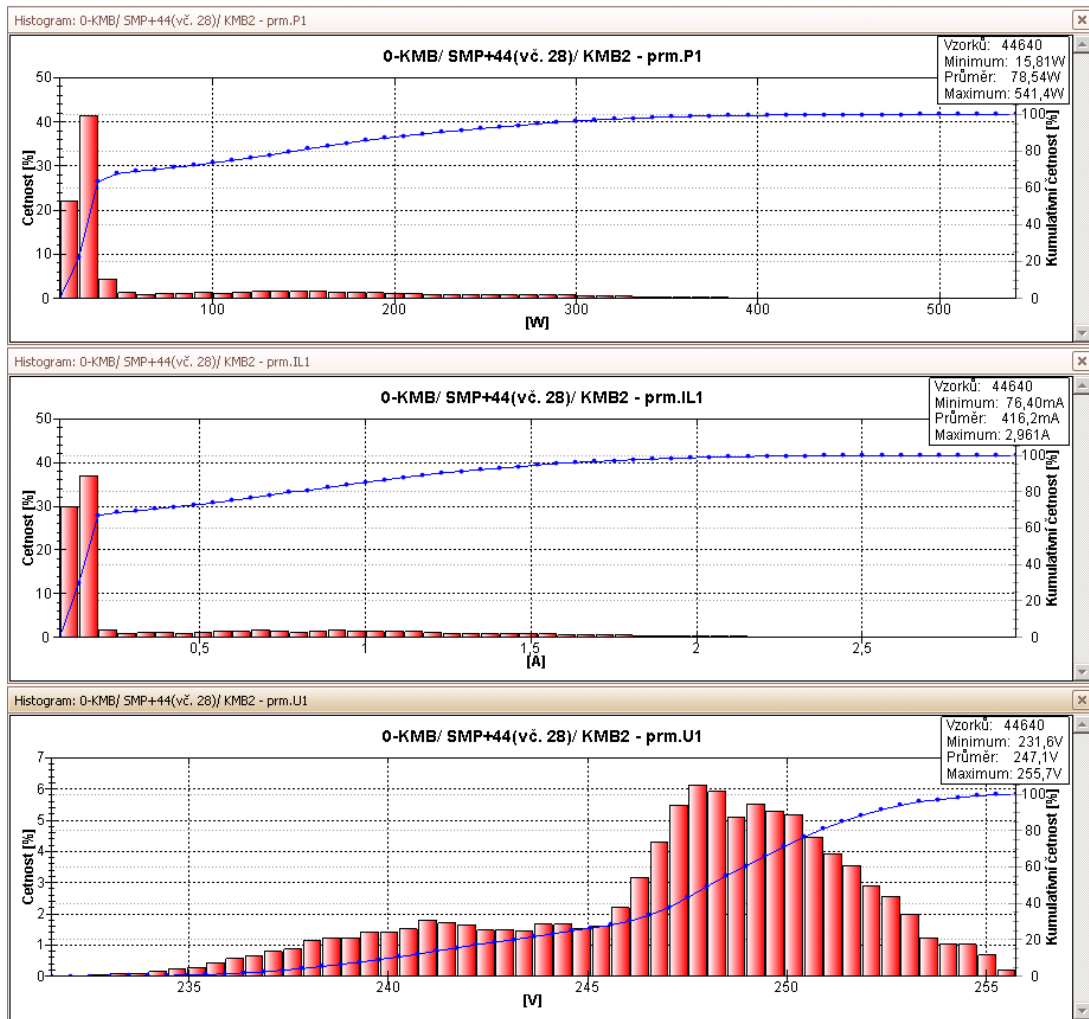
### 4.3 Komprese hlavního archivu

Agregovaná data různých veličin v hlavním archivu mají různá statistická rozložení respektive vykazují rozdílné charakteristiky. Pro optimalizaci komprese těchto záznamů je nutné použít některé heuristicky orientované metody a případně vybírat konkrétní metodu podle aktuálního výsledku. Hodnoty frekvence, napětí, proudů, blikání, THD a některé další jsou kódovány jako celá čísla s rozsahem 1, 2 nebo 4 bajty. Výkony jsou kódovány jako 32-bitové veličiny s plovoucí řádovou čárkou (float). Obr.: 4.4 srovnává CR vybraných veličin hlavního archivu. Běžné nároky na provoz elektrických zařízení eliminují příliš velké kolísání jednotlivých veličin. Prudké změny hodnot některých veličin či dlouhodobější setrvání hodnoty mimo vyhrazený rozsah je omezeno. Obrázek 4.2 zobrazuje konkrétní hodnoty měřeného proudu a napětí v průběhu dne (osa Y) po dobu několika měsíců (osa X). Z grafů je názorně vidět typické rozložení hodnot v čase pro tuto konkrétní aplikaci a velikost ploch stejné barvy naznačuje míru redundance dat i časové rozložení.

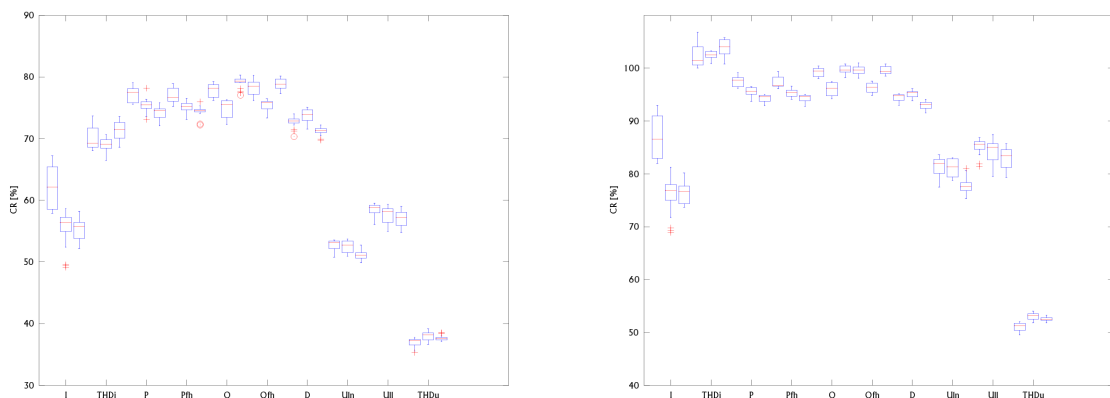
Efekt aplikace komprese nelze pro obecně různá vstupní data spolehlivě predikovat. Přitom aby taková aplikace měla smysl, je nutný předpoklad, že komprese povede k úspoře místa v archivu. V grafech 4.5 a 4.4 je pro různé soubory měření znázorněn charakter dosaženého CR. Je patrný vliv velikosti vstupního souboru a kódování veličiny na dosaženém CR. V některých případech je rozptyl CR poměrně velký - obvykle je tento jev způsoben tím, že zkoumaná data obsahují úseky měření s výrazně odlišným charakterem.

#### 4.3.1 Parametry modelovaných veličin

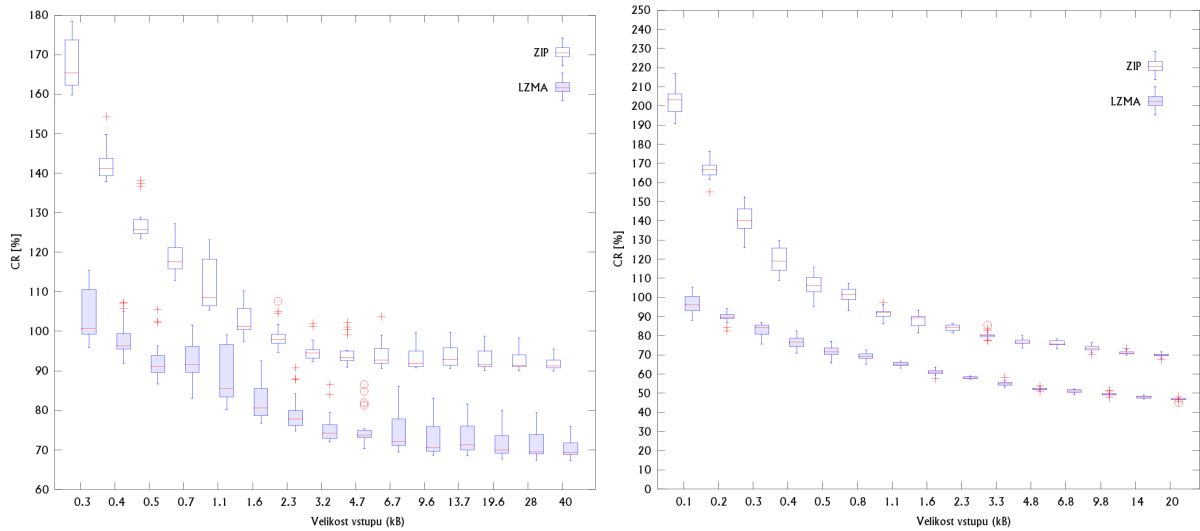
**Frekvence** vykazuje obvykle minimální kolísání v okolí nominální hodnoty (50 nebo 60 Hz). Rozsah hodnot lze omezit a snížit tak i počet bitů nezbytných pro zakódování všech platných hodnot.



Obrázek 4.3: Histogram (distribuce) hodnot napětí, proudu a výkonu v souboru KMB2.



Obrázek 4.4: Rozptyl kompresních poměrů algoritmu LZMA a ZIP pro 10000 hodnot měření vybraných veličin (20 náhodných výběrů z měsíčního měření, soubor KMB2).



Obrázek 4.5: Rozptyl kompresních poměrů algoritmů ZIP a LZMA v závislosti na velikosti vstupních dat pro 20 náhodných výběrů. Vlevo je činný výkon, vpravo napětí.

**Napětí** představuje 7 datových toků (16-bit int). Přibližným modelem závislostí mezi těmito veličinami je fázorový diagram a poměry z něj plynoucí pro tyto veličiny. V archivu je ukládána také 1. harmonická složka napětí  $U_1$ ,  $THD_U$ , hodnoty jednotlivých harmonických a meziharmonických a hodnota napěťové nesymetrie  $UNB_U$ . Z grafu na obr. 4.3 je zřejmé typické rozložení hodnot měřeného napětí v souboru.

**Proudy** jsou ukládány jako 4 datové toky (16 bit int), v praxi stejně jako u napětí lze předikovat hodnotu všech fází pomocí změny hodnoty v jedné fázi. Proudy lze více využívat při kompresi výkonů viz. následující odstavec. Jako zvláštní veličina je ukládána také první harmonická složka  $I_1$ ,  $THD_I$ , hodnoty jednotlivých harmonických a meziharmonických a hodnota proudové nesymetrie  $UNB_I$ .

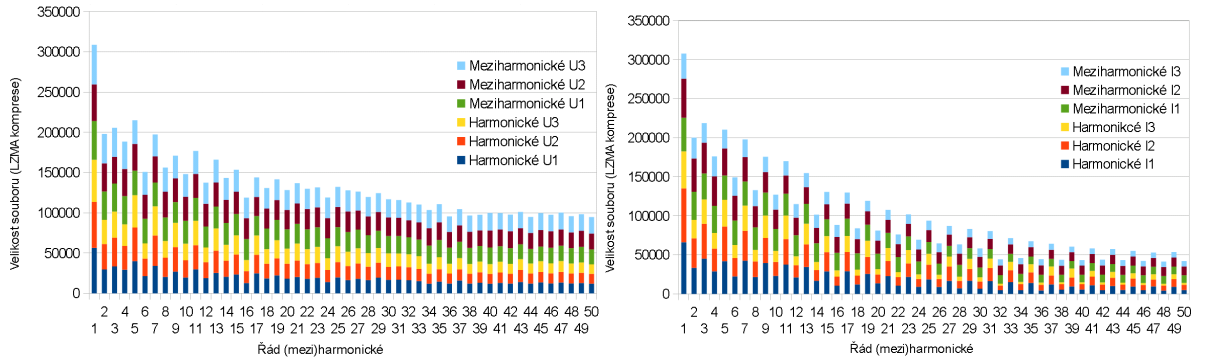
**Výkony** představují významný datový objem hlavního archivu. V původní verzi jsou ukládány jako float hodnoty. Ukládají se hodnoty pro  $P$ ,  $Q$ ,  $P_1$ ,  $Q_1$  a  $D$ . Pro modelování platí, že zdánlivý výkon je určen jako součin agregované efektivní hodnoty napětí a proudu.  $P_{fh} < P$  a pro jejich rozdíl  $P_H$  obvykle platí, že  $P_H < P_{fh}$ . Obdobný vztah a předpoklad platí i pro  $Q$ ,  $Q_1$  a  $Q_H$ . Je tedy výhodnější modelovat a ukládat dvojice veličin  $P_1$  a  $P_H$ , resp.  $Q_1$  a  $Q_H$ . Stejně tak hodnoty  $D$  a  $U_{LN}$  mají obvykle menší rozptyl a dobře se komprimují. Nejprve tedy pro  $P$  platí:

$$\begin{aligned} P_H &= \epsilon_P = P - P_1 \\ P &= P_1 + P_H \end{aligned} \quad (4.1)$$

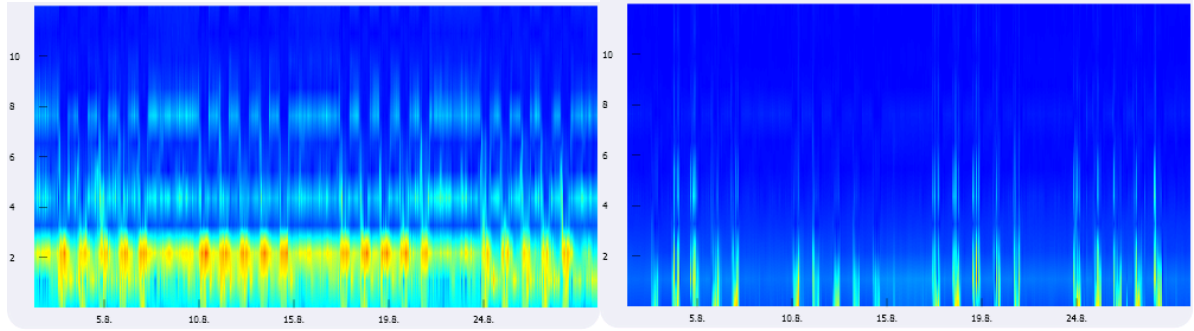
, kde dvojice  $P_H$  a  $P_1$  úplně a zároveň s menší entropií definují hodnoty činného výkonu ( $Q$ ,  $Q_h$  a  $Q_1$  analogicky pro jalový). Vztahy pro  $U$ ,  $U_1$ ,  $I$  a  $I_1$  a výkony jsou následující<sup>1</sup>.

<sup>1</sup>Z hlediska rozložení lze předpokládat, že platí  $P > P_1 > P_H$ ,  $Q > Q_1 > Q_H$  a  $Q > D$ .





(a) Kompresi harmonických složek napětí a proudu.



(b) Diagram harmonických složek napětí a proudu.

Obrázek 4.6: Kompresi archivu zaznamenaných harmonických a mezipharmonických (1. až 50. harmonická, agregace 1s). Soubor KMB2 obsahuje celkem cca 85 tis. odečtů.

$$\epsilon_{I_{ef}} = UI - Q^2 - P^2 - D^2 \quad (4.2)$$

$$I_{ef} = \frac{\epsilon_I + Q^2 + P^2 + D^2}{U} \quad (4.3)$$

$$\epsilon_{I_1} = U_1 I_1 - Q_1^2 - P_1^2 \quad (4.4)$$

$$I_1 = \frac{\epsilon_{I_1} + Q_1^2 + P_1^2}{U_{fh}} \quad (4.5)$$

**Harmonické napětí a proudu** V závislosti na konkrétním nastavení archivace mohou záznamy jednotlivých harmonických a mezipharmonických složek a fázových posunů mezi nimi představovat zdaleka největší objem ukládaných dat. Efektivní hodnota dané veličiny je úměrná součtu efektivních hodnot jejích jednotlivých harmonických a mezipharmonických složek, což je však pro využití pro kompresi málo užitečné. V reálných systémech však lze také využít předpokladu, že vyšší řád harmonických je měřen s nižší absolutní přesností a zároveň je pravděpodobné, že reálné měřené hodnoty vyšších harmonických budou absolutně nižší než hodnoty nižších harmonických.

# Kapitola 5

## Modelování časových řad

Modely časových řad měření jednotlivých veličin jsou zvoleny jako prostředek vhodný pro optimalizaci vstupních dat před kompresní fází. Jak bylo experimentálně ověřeno, tento krok může vylepšit výsledný kompresní poměr v řádu jednotek až desítek procent. V textu práce jsou popsány použité modely, možnosti jejich automatické konstrukce pro datové řady s obecně neznámou charakteristikou i praktické výsledky pro aplikaci modelů při kompresi.

### 5.1 Běžné metriky

Model časové řady je mechanismus, s jehož pomocí lze odhadovat hodnotu  $y_i$  pro libovolné  $i$  z definovaného intervalu. Residuum  $e_i$  představuje rozdíl skutečné hodnoty vzorku  $y_i$  a modelované hodnoty  $f(x_i)$  pro nezávislou proměnnou  $x_i$ , tj. odchylku modelu v bodě  $i$  od modelované reality. Pro nalezení nejúčinnějšího modelu pro kompresi lze vyhodnocovat přímo účinnost komprese šumu (residuí) a modelu vůči kompresi originálních dat. Nevýhodou této metriky je ale její vysoká časová i paměťová náročnost a proto lze využít i jiné standardní metriky, například viz vztahy 5.1, 5.3 či 5.4.

$$RSS = \sum_{i=1}^n e_i^2 \quad (5.1)$$

$$MSE = \frac{RSS}{n} = \frac{1}{n} \sum_{i=1}^n (f(x_i) - y_i)^2 \quad (5.2)$$

$$RMSE = \sqrt{MSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (f(x_i) - y_i)^2} \quad (5.3)$$

$$MAE = \frac{1}{n} \sum_{i=1}^n |f(x_i) - y_i| \quad (5.4)$$

## 5.2 Navržená LogHist metrika

Efektivita současných kompresních algoritmů souvisí úzce s rozložením jednotlivých symbolů v komprimovaném datovém toku. Optimální rozložení představuje soubor se znaky položenými v blízkosti nuly. Vzdálenost kódu znaku od nuly lze definovat různými způsoby, triviálně například ve smyslu kartézské vzdálenosti.

Mnou navržená metrika vychází z této úvahy a respektuje některé kroky kompresního algoritmu a penalizuje prvky souboru, které přispívají k horšímu výslednému CR. Tato její vlastnost a použití metriky jako kritérium pro optimalizaci modelů je účinné (např. ve srovnání s optimalizací, která jako metriku používá RMSE) pro zlepšení dosažitelného CR. Metrika v první řadě konstruuje histogram výskytu symbolů v residuu každého intervalu. Pro 8mi-bitové symboly tak například pro každý modelovaný interval vzniká 256 hodnot čítačů, jejichž konkrétní hodnota popisuje, jak úspěšná bude prováděná komprese.

Pro účinnost komprese je relevantní počet bitů, potřebných k zakódování jednotlivých znaků v souboru respektive v jeho části a proto jsem v metrice jednotlivé skupiny symbolů rozdělil podle indexu nejvýznamnějšího nenulového bitu v kódu každého symbolu. Četnosti jednotlivých znaků jsem sečetl vynásobené vahou příslušné skupiny, čímž vznikla celková agregace (metrika) pro hodnocení intervalu modelu z hlediska potencionální účinnosti komprese viz. vztah 5.5. Pro posuzování jednotlivých intervalů lze tento vztah dále doplnit o normalizaci počtem hodnot v intervalu, získáme tak pro každý interval průměrné *LogHist* viz vztah 5.6.

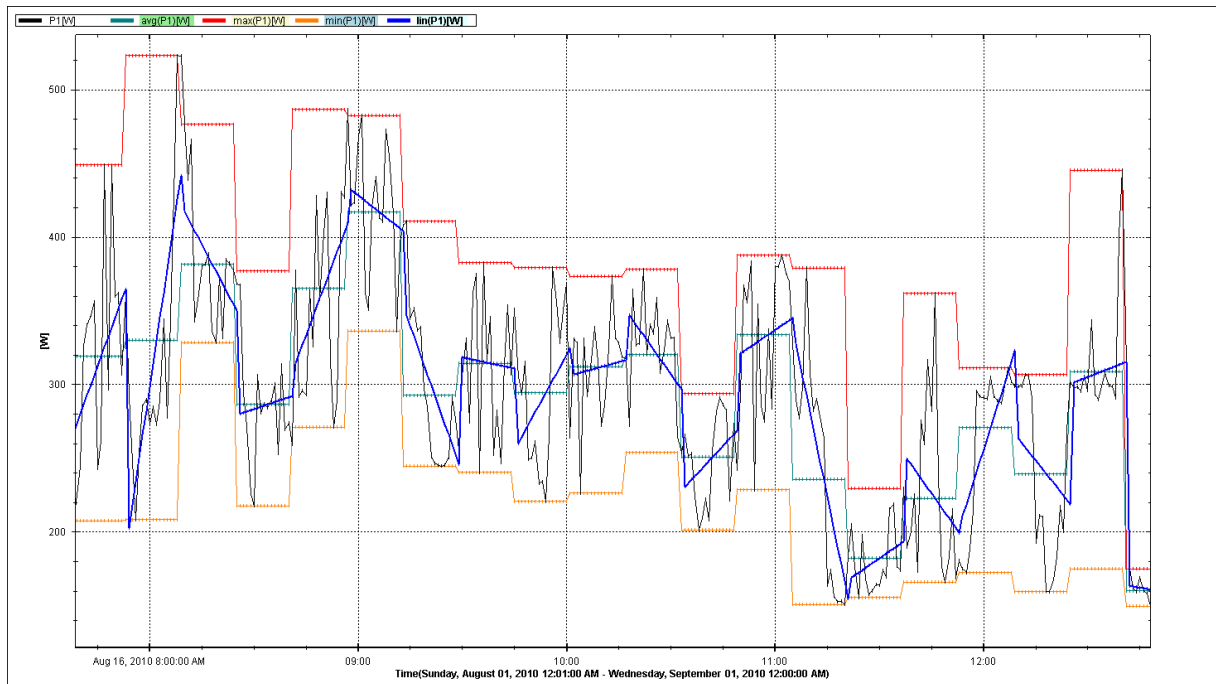
$$\text{LogHist} = \sum_{i=0}^{Max} \text{Hist}_i \times (\text{int})(\log_N(i) + 1) \quad (5.5)$$

$$\text{LogHist}_{norm} = \frac{\text{LogHist}}{\sum_{i=0}^{Max} \text{Hist}_i} \quad (5.6)$$

kde *Max* je maximální možná hodnota symbolu ( $Max = 2^N$ , kde *N* je délka kódového slova), *Hist<sub>i</sub>* je počet výskytů *i*. symbolu v intervalu a  $(\text{int})(\log_N(i) + 1)$  je (lineární) váha skupiny symbolů zvoleného *N*-árního kódování. Běžně *N*=2, kromě proporcionalního váhování lze použít i jiné závislosti pro výpočet vah jednotlivých skupin symbolů. Takto definovaná optimalizační metrika přináší obvykle lepší výsledky než například optimalizace vůči RMSE nebo MAE. Její obrovskou výhodou oproti výše zmiňovaným metrikám je její přirozená multikriterialita tzn. schopnost optimalizovat model pro kompresní fázi pro více časových řad najednou, její účinnost není počtem vstupů (řad) ovlivněna.

## 5.3 Typy modelů

**Diferenciální kódování** veličin je metoda předzpracování dat, která v časové posloupnosti měření veličin snižuje entropii predikcí následující hodnoty hodnotou předchozí. Hodnoty reziduí diferenciálního kódování pak představují vlastní model. Pro získání originálních hodnot ve zvoleném intervalu je nutné dekódovat kompletní úsek diferenciálně uložené posloupnosti.



Obrázek 5.1: Originální data souboru KMB2 modelovaná pomocí charakteristických hodnot intervalu (min, avg, max) a lineární aproximací.

**Model MIN/AVG/MAX aktuálních hodnot** využívá platnosti nerovnosti  $X_{min} \leq X_{avg} \leq X_{max}$ . Tohoto pravidla lze snadno využívat pro jednoduchou a rychlou redukci entropie dat.

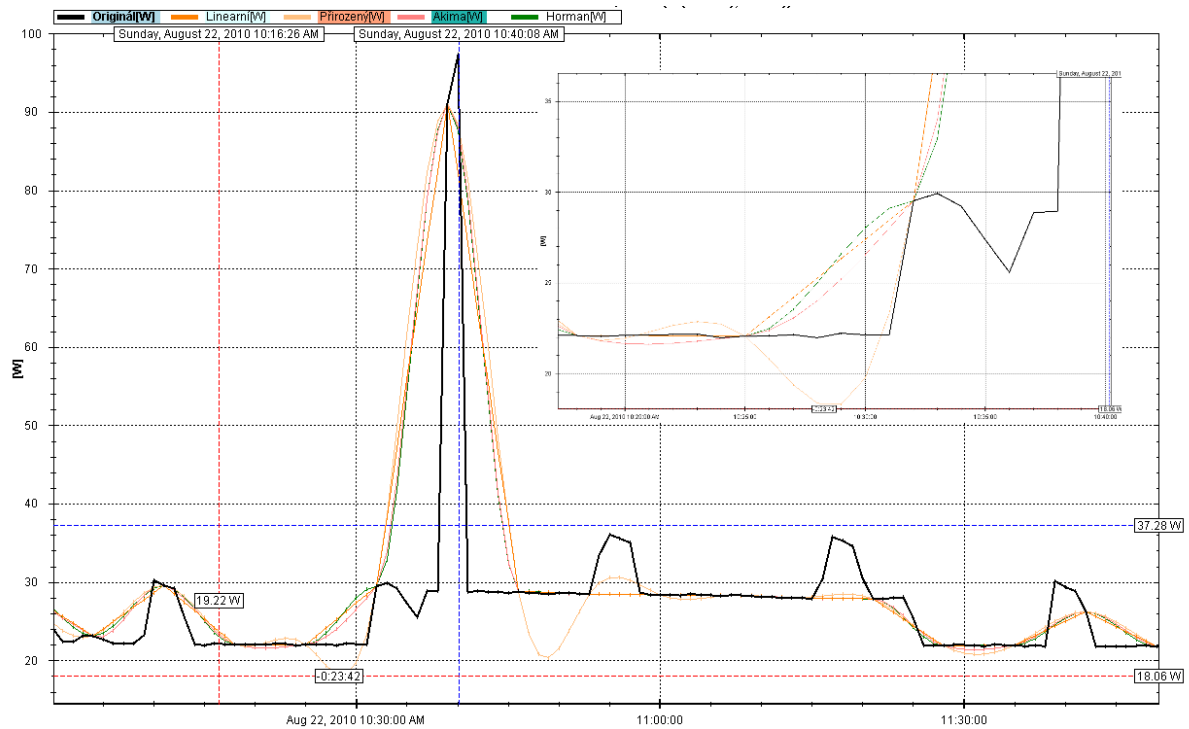
$P = P_1 + P_H$  a  $Q = Q_1 + Q_H$  popisuje vztahy mezi v databázi uloženým činným resp. jalovým výkonem a jeho harmonickými složkami.

$P = (U_{ef}I_{ef})^2 - Q^2$ ,  $P = (U_{ef}I_{ef})^2 - Q^2 - D^2$  je model podobný předchozímu, ovšem vychází z jiných vztahů mezi činným, jalovým, zdánlivým a deformačním výkonem. Výhodou oproti předchozímu modelu je použití v praxi častěji zkoumaných veličin. U a D představují nejlépe komprimovatelné veličiny, mezi P a I obvykle existuje silnější korelace a obvykle také platí, že  $P \gg Q$ .

$THD_U, THD_I$  Celkové harmonické zkreslení ovlivňuje relace mezi agregovanými hodnotami v archivu. Určit lze buď ze znalosti všech harmonických složek signálu nebo z efektivní hodnoty měřené veličiny a hodnoty 1. harmonické.

**Vztah mezi  $U_{LL}$  a  $U_{LN}$**  za předpokladu, že jednotlivé fázory napětí jsou stejnoměrně posunutá o  $\frac{2}{3}\pi$ [RAD] lze využít s využitím kosinové věty k odhadu průměrných fázových napětí ze sdružených.

**Charakteristická hodnota v intervalu** modeluje časovou řadu ve zvolených intervalech vhodnou (typickou) hodnotou. Efektivnější pro aplikaci komprese je minimální resp. maximální model (modeluje interval hodnotou extrémů tohoto intervalu) a tyto dva jsou srovnatelně účinné.



Obrázek 5.2: Modelů využívající různé spline interpolace. Vpravo detailní část výřezu.

**Polynomiální model** využívá k popisu funkční závislosti měřených hodnot na čase polynom určitého stupně  $N_p$  s vhodně zvolenými koeficienty

**Spline modely** využívají k modelování tzv. spline křivky. Jednotlivé typy spline se liší použitou funkcí (polynom  $N$ . stupně, racionální lomená funkce, ...) a okrajovými podmínkami, které jsou použity pro definování koeficientů těchto interpolačních funkcí. Spline využívají modely výrazně nižšího stupně než je velikost vstupních dat. V implementaci byly použity kubické spline interpolace, Hermitova či Akima interpolace a další.

**Fourierova transformace** popisuje periodické jevy v signálu. Vhodným výběrem či změnou složek spektra a jeho zpětnou transformací jsou získány hodnoty modelu odpovídajících vlastností. Vliv postupného odstranění nejméně významných složek spektra signálu na model je názorně zobrazen na obrázku 5.6. Tento způsob modelování je pro zlepšení bezztrátové komprese málo účinný - součet zkomprimovaných dat modelu a reziduí výrazně převyšuje velikost zkomprimovaného původního modelu. FFT modely jsou vhodné pro predikci periodických dat a lze je využít pro velmi efektivní ztrátovou kompresi. Celková chyba efektivní hodnoty modelu je přímo úměrná součtu efektivních hodnot složek, které odstraníme ze spektra modelu.

## 5.4 Konstrukce optimálních modelů

Pokud vyhodnocujeme kvalitu různých modelů vůči stejné množině dat, je možné použít přímo některý z ukazatelů, uvedených v kapitole 5.3. Pro eliminaci vzniků přeúčených modelů lze použít kritéria, která berou v úvahu kromě absolutní chyby také složitost modelu a vhodným způsobem

ji penalizují. Mezi taková kritéria patří například Akaikeho informační kritérium (AIC)[4] a Bayesovské informační kritérium (BIC)[33].

$$BIC = \log \left[ \frac{1}{n} \sum_{i=1}^n (f(x_i) - y_i)^2 \right] + \frac{2p}{n} \quad (5.7)$$

$$AIC = \log \left[ \frac{1}{n} \sum_{i=1}^n (f(x_i) - y_i)^2 \right] + \frac{(\log n)p}{n} \quad (5.8)$$

V obou vztazích  $p$  představuje počet uzlů modelu, které lze vybrat z  $n$  vzorků. Obě kritéria druhým členem penalizují složitost modelu.

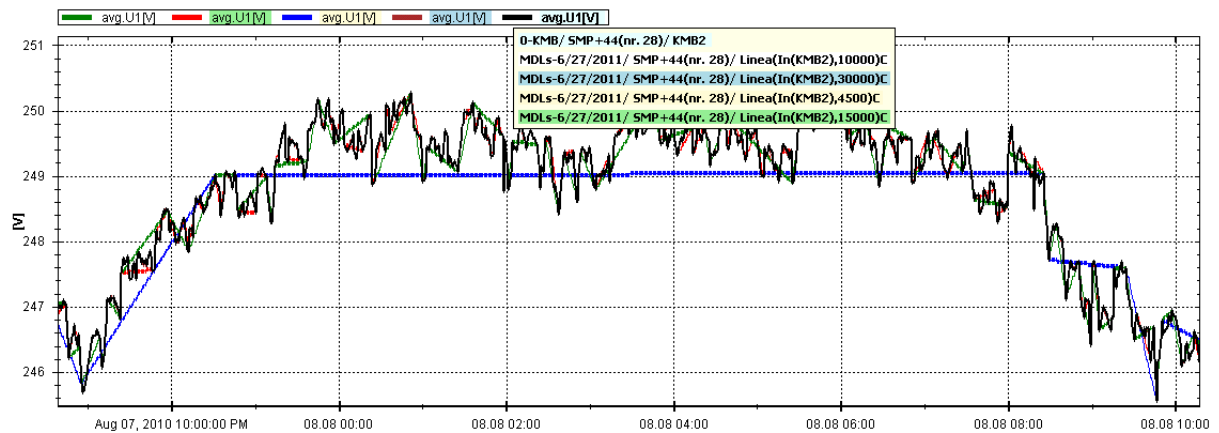
Všechny modely s výjimkou fyzikálních modelují časové řady horizontálně - tzn. aproximují různými způsoby hodnoty časových řad v libovolném okamžiku. Při konstrukci konkrétního modelu je tak klíčovým krokem výběr optimální variace časů (uzlů) z originálního měření. Pro vyhodnocování optimality jednotlivých výběrů pro model lze využít některé z výše uvedených kritérií. V následujících podkapitolách jsou proto podrobněji popsány některé možné optimalizace a techniky konstrukce vhodných modelů.

**Rovnoměrně rozložené intervaly** představují základní metodu konstrukce modelu časové řady, která zaznamenané hodnoty dělí na úseky stejné délky. Jedná se o jednoduchou techniku konstrukce modelu, která však pro praktické soubory dat vrací obvykle dostatečně vhodné aproximace. Optimální počet uzlů modelů s intervaly a jejich konkrétní rozmístění (dělení modelu na intervaly) je možné určovat kontinuálně dle aktuální hodnoty zvolené metriky modelu. Vztah 5.3 se jeví jako vhodný pro tento účel, ale konkrétní aplikace neprokázala výrazné zlepšení dosahovaného CR. Metrika LogHist daná vztahem 5.5 naopak vytváří tímto postupem instance modelů výrazně vhodnější pro LZ kompresní algoritmy. Na obrázku 5.5 jsou znázorněny četnosti jednotlivých symbolů (v 2D histogramu) pro ekvidistantní Min model a pro stejný typ modelu, konstruovaný s využitím metriky LogHist. V grafu 5.4 tato technika v kombinaci s diferenciálním kódováním hodnot modelu dosahuje vůbec nejlepšího CR pro archiv hodnot proudu v souboru KMB2.

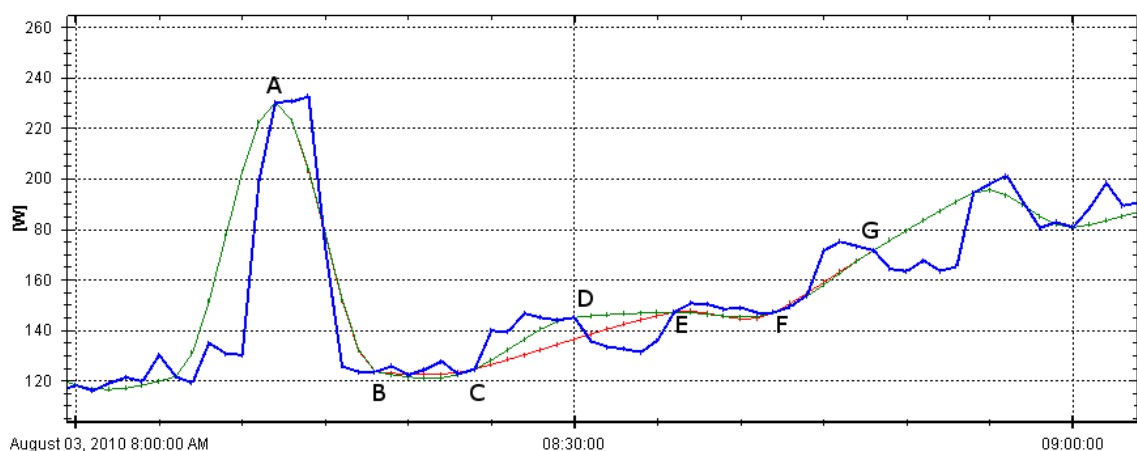
**Logaritmické škálování** je systematický postup, který metodou shora dolů v  $O(\log_2 N)$  krocích postupně zjemňuje model s rovnoměrně rozloženými intervaly dokud následující krok algoritmu přináší úsporu. Tato technika je použitelná pro odhad optimálního poměru uzlů pro konkrétní data neznámých či nepředvídatelných vlastností většího rozsahu.

**Racionální škálování** představuje podobný postup jako předchozí algoritmus. Liší se volbou velikosti kroku při snižování počtu uzlů a je vhodný pro systematické ověření parametrů modelu vůči kvalitě komprese pro menší množiny dat a nebo pro precizní doladění dosažené komprese v okolí jiným způsobem nalezeného optima.

**Gradientní hledání lokálních optim** vychází z modelu s rovnoměrně rozloženými intervaly - tzn. že v prvním kroku vytvoří takový rovnoměrný model. V následujících několika krocích



(a) Výřez srovnání originálu a několika vybraných modelů při optimalizaci postupným odebráním nejméně významných uzlů - model interpolující data přímkou.



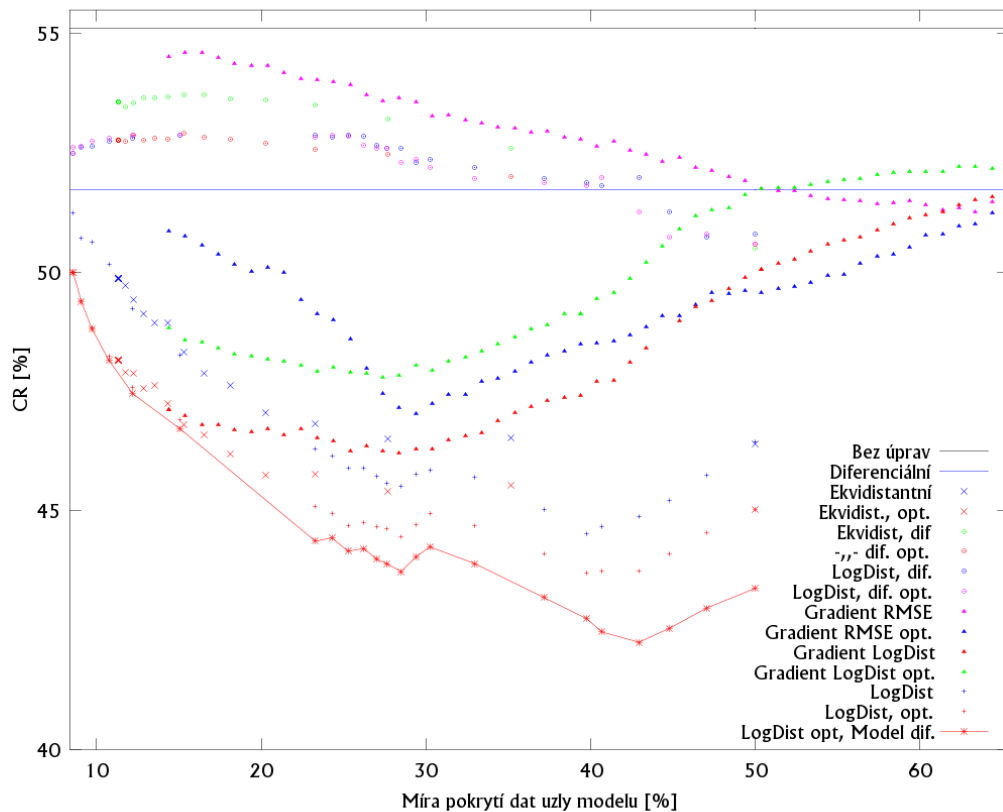
(b) Lokální změna modelu odebráním jednoho uzlu (bod D) z Akima interpolačního modelu. Odchylka hodnot oproti původnímu modelu se projeví pouze v intervalu mezi body A a G. Modrá křivka představuje původní data, zelená je původní model, červená je model s odebráním bodem D.

Obrázek 5.3: Ukázky vybraných modelovacích metod na reálných datech.

postupně mění výběr jednotlivých intervalů (rozložení uzlů) tak, aby v každém vybraném čase dosáhla minimální odchylky. Výběr každého uzlu lze postupně posouvat ve směru gradientu funkce celkové chyby.

**Optimalizace záměnou uzlů** představuje jednoduchý a stabilní hladový algoritmus. Vstupem tohoto algoritmu je libovolný model, daný volbou  $k$  uzlů. Z optimalizovaného modelu lze postupně systematicky odebírat  $N$  uzlů, jejichž odebráním způsobíme nejmenší nárůst celkové chyby modelu. Odebraný bod umístíme do intervalu, kde je největší lokální chyba buď na střed či jinak empiricky vypočítanou pozici v intervalu či na místo, které nejvíce sníží celkovou chybu modelu.

**Dělení intervalů s největší chybou** pokrývá data uzly modelu v intervalu, jehož rozdělením vznikne model s nejmenší celkovou chybou. Algoritmus končí v okamžiku, kdy vyčerpá počet přidělených uzlů.



Obrázek 5.4: Srovnání efektivity různých metod optimalizace při aplikaci pro kompresi archivu měření PQ (44639 hodnot  $I_1$  ze souboru KMB2, RawDataMinModel).

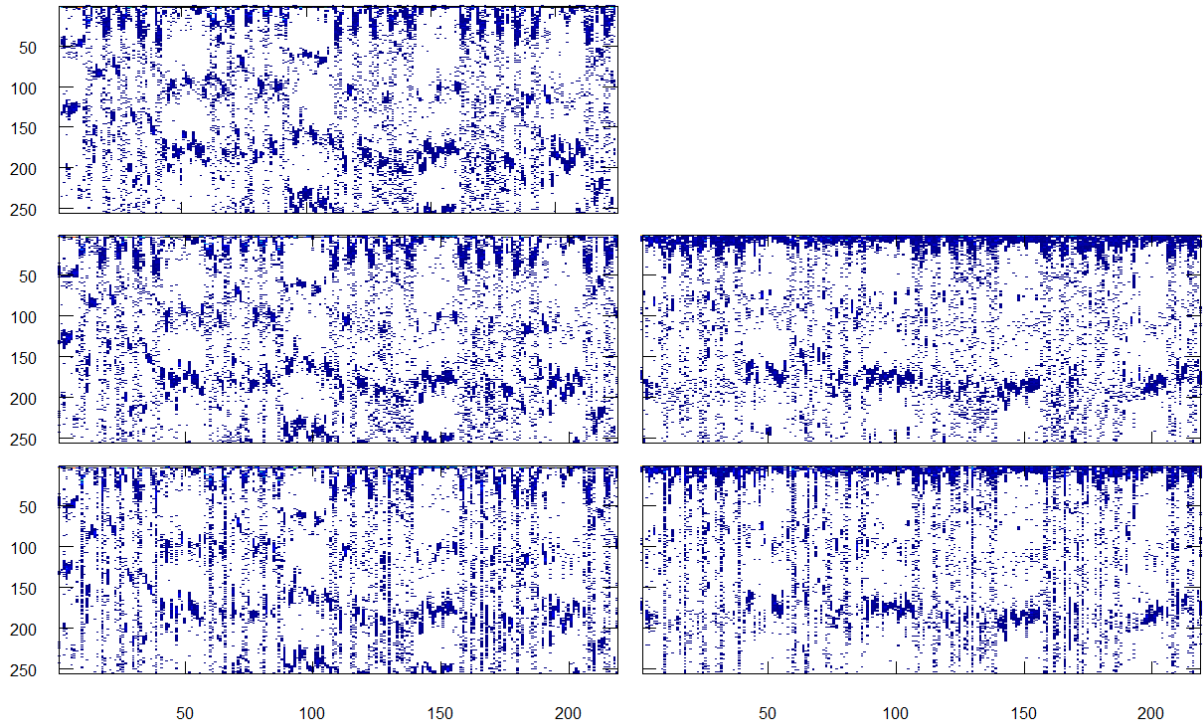
**Postupné odebírání uzlů** představuje podobně jako předchozí jednoduchou metodu pro postupnou konstrukci přiměřeně dobrého modelu pomocí série lokálně optimálních kroků. V tomto případě je výchozím stavem model s uzly rozmístěnými ve všech bodech (nebo v rovnoměrně rozmístěných intervalech) vstupních dat. Algoritmus postupně odebírá ty uzly, jejichž odebráním se celková chyba modelu nejméně zhorší. Algoritmus skončí v okamžiku, kdy dosáhne stavu, kdy model obsahuje požadovaný počet uzlů.

**Optimální Fourierův model** je představován filtrovaným spektrem signálu, které je tvořeno magnitudami a fázovým posunem jednotlivých harmonických složek. Pro snižování entropie residuálních hodnot je nutné sjednotit (normalizovat) podobné složky spektra.

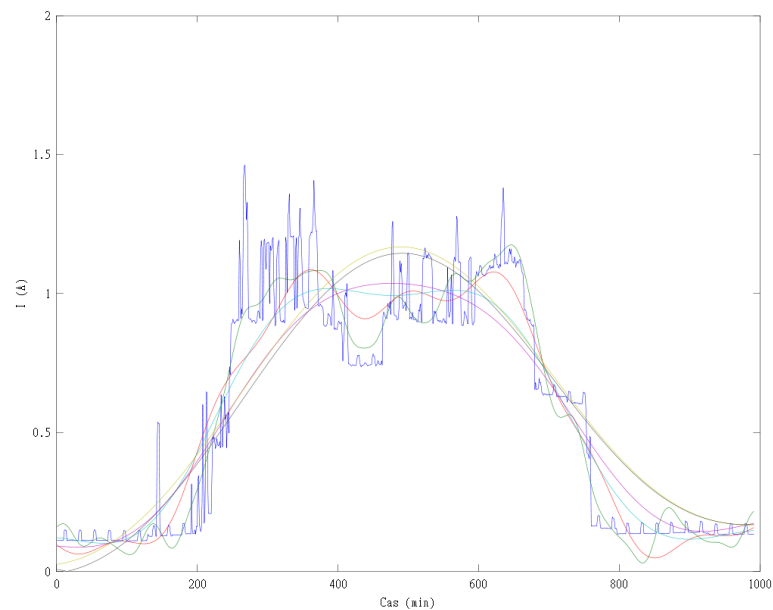
## 5.5 Popis knihovny pro modelování

Pro měření efektivity a správnosti výše zmiňovaných modelů jsem vytvořil knihovnu tříd KMB Analytics v jazyce C#. Knihovna obsahuje rozhraní definující společné vlastnosti a funkce všech modelů. Toto rozhraní umožňuje jednotlivé modely využívat transparentně ve všech popisovaných experimentech. Data (časové řady měření různých veličin) jsou ve zmiňované knihovně reprezentovány vhodnou univerzální datovou strukturou `IList<T>[]`, kde T představuje požado-





Obrázek 5.5: Rozložení symbolů při modelování minimální hodnotou s ekvidistantními intervaly vlevo a při vytváření intervalů s konstantní chybou dle vztahu 5.5 vpravo. Shora dolů jednotlivé grafy představují 2d histogram výskytu symbolů v originálních datech, v residuu a v residuu modelu, který je doladěn expanzí intervalů. Vpravo lze pozorovat výrazný úbytek znaků s delším kódem.



Obrázek 5.6: Modelování časové řady FFT modelu s různými úrovněmi zachování původní informace - výřez denního průběhu měření činného výkonu.

vaný datový typ, nejčastěji double, int nebo short. Nultá kolekce v tomto poli obvykle reprezentuje nezávislou veličinu (čas). Všechny ostatní kolekce obsahují hodnoty příslušných (závislých) veličin v konkrétním čase.

K implementaci interpolačních modelů byla použita knihovna Math.Net Numerics, obsahující funkce pro generování potřebných polynomiálních, racionálních a spline interpolací. Součástí knihovny je i třída implementující FFT/IFFT algoritmus. Pro interpolaci a odhad koeficientů metodou nejmenších čtverců jsem využil knihovnu Meta Numerics. Nedílnou součástí knihovny jsou třídy s implementací funkcí vhodných pro analýzu časových řad. Knihovna na kolekcích časových řad implementuje v práci popisované metriky. Zavádí také další běžné statistické výpočty, transformace a operace pro snadnější manipulaci s množinami měření a vyhodnocování jejich vlastností. Interoperabilita a rozšiřitelnost je zaručena využíváním díky využívání standardních kontejnerových typů pro manipulaci s daty a využitím obecných rozhraní pro definování funkcionality.

Pro experimenty s kompresí byly použity algoritmy a konkrétní knihovny pro prostředí .NET, jejichž licence umožňuje volné použití zdrojových kódů těchto algoritmů pro nekomerční aplikace. Knihovna KMB.Analytics definuje obecné rozhraní kompresního kodéru, jehož výstupem je přímo zakódované pole bajtů. Toto rozhraní implementují jednotlivé pomocné třídy, které sjednocují používání různých kompresních/dekompresních funkcí, vyvinutých přímo v KMB.Analytics i z knihoven třetích stran (LZMA 9.20 projektu 7ZIP ([www.7-zip.org](http://www.7-zip.org)) a ZIP a BZIP2 komprese knihovny SharpZipLib v. 0.86). Pro potřeby vizualizace, ověřování a dalšího zpracování je knihovna doplněna o vstupně výstupní vrstvu s možností exportu do formátu souborů pro prostředí MATLAB (.m) a EXCEL/SCALC (.csv, .xls). Modelovaná data, residua a další výstupní posloupnosti lze také ukládat do SQL databáze aplikace ENVIS.

## 5.6 Shrnutí výsledků měření

V kapitole 5 jsou diskutovány možnosti modelovat množiny časových řad různými metodami. Hlavní důraz při hodnocení je kladen na možnost zefektivnění CR při následné aplikaci komprese na data modelu a residua oproti CR komprese, aplikované na data originálu. Bylo též ukázáno, že existuje souvislost mezi kvalitou modelu, která může být vyjádřena například pomocí RMSE, AIC, BIC či jiných metrik, a mírou zlepšení CR.

Nejvhodnější zkoumaná metoda pro kompresi popisovaných dat představuje kombinaci LZMA komprese, modelování dat na intervalech jejich charakteristickou (minimální) hodnotou a s daty modelu následně kódovanými diferenciálně. Délku a počet jednotlivých intervalů lze určit fixně, ale lepších výsledků dosahuje optimalizace intervalů na konstantní RMSE a v experimentech vůbec nejlepší metoda na určení rozložení intervalů je navržená LogHist metrika.

# Zhodnocení disertační práce

Modelování dat měření představuje nástroj pro automatickou analýzu a optimalizaci velmi rozsáhlých databází měření. Modely lze využít k vyhledávání charakteristických vlastností měřených uzlů sítě či naopak k odhalování poruchových stavů. Modelováním spotřeby lze charakterizovat různé profily uživatelů a vytvářet respektive analyzovat typické okamžiky. S touto přidanou informací lze lépe plánovat optimalizaci provozu měřené sítě a připojených zdrojů či spotřebičů. Pro další rozvoj datových skladů pro PQ je zajímavá a dostatečně nevyužitá možnost pravidelně agregovat starší měření, které neobsahuje data zajímavá buď z pohledu vyhodnocení kvality nebo nebo představují „typické“ měření. Pro plnohodnotné nasazení testovaných algoritmů je nezbytně nutné provést optimalizaci jejich implementace - ta je sice s ohledem na výsledek správná, ale nekladla si za cíl dosáhnout i časové či paměťové efektivity. Většinu v práci použitých algoritmů lze například jednoduchým způsobem převést na paralelní. Významnou úsporu výkonu by představovala implementace vybraných metod již na úrovni databázového serveru například s využitím vložených procedur.

Díličí cíle, jichž bylo dosaženo, lze shrnout do následujících bodů:

- Byl zanalyzován soubor požadavků na přístroje, vyhodnocující kvalitu elektrické energie.
- Na základě této analýzy byl realizován experimentální analyzátor kvality elektrické energie dle EN 50160 ed. 3 třída S a analytický software ENVIS.
- Pro potřeby ověření vlastností bezztrátových algoritmů pro kompresi dat měření byl vytvořen virtuální měřicí přístroj.
- Na datech vytvořeného analyzátoru byly testovány různé kompresní algoritmy a vliv volby kódování/uspořádání hodnot veličin na kompresní poměr.
- Jako optimální varianta byl vyhodnocen LZMA algoritmus v kombinaci s modelem intervalů měření minimální hodnotou.
- Rozložení intervalů modelu je nejlépe určeno s využitím v práci navržené metriky LogHist a lokálních gradientních optimalizací.

Přínosy disertační práce lze spatřovat zejména v tom, že:

- Vznikl datový formát pro záznam měření, který implementuje LZ kompresi.
- Vznikl návrh pro efektivnější využití kapacity paměti v měřicím přístroji.
- Následným modelováním průběhu měření bylo docíleno zlepšení celkového kompresního poměru v databázi záznamu měření.

# Závěr

Dizertační práce vychází z analýzy a specifikace metod a algoritmů pro kontinuální měření a vyhodnocování parametrů kvality elektrické energie popsané v kapitolách 3 a 3.2. S využitím vlastních řešení je implementován firmware měřícího přístroje pro mikrokontrolér ARM. Experimentální firmware analyzátoru používá pro použitý procesor speciálně optimalizovaný algoritmus FFT a korigovaný rychlý výpočet jalového výkonu. Obě tyto optimalizace umožnily realizaci analyzátoru kvality dle požadavků a s nadstandardními funkcemi na minimalistickém hardware. Také byl navržen a realizován vizualizační software pro PC. Tento software ukládá data v binárních souborech či do SQL databáze.

V průběhu vývoje výše uvedených se jedním z hlavních problémů obecných systémů pro monitoring kvality ukázal být efektivní, tedy hlavně rychlý a spolehlivý, přístup k datům na všech úrovních popisovaného systému. Protože se zároveň jedná o nedostatečně publikovanou a jak bylo ukázáno i nevyužívanou problematiku, bylo hlavní náplní této práce navrhnout a realizovat nové postupy, které povedou k zefektivnění využití úložného prostoru. V experimentech s fiktivními i reálnými archivními daty byl jako nejvhodnější algoritmus bezztrátové komprese pro další experimenty vyhodnocen bezztrátový kompresní algoritmus LZMA.

Charakter časových řad hodnot měřených veličin a jejich různé vzájemné vztahy a závislosti umožnily další optimalizaci kompresního poměru s využitím vybraných technik matematického modelování a predikce s cílem snížit celkovou entropii datového souboru. Práce rozebírá jednotlivé navržené metody a jejich vliv na případné zlepšení komprese původního souboru. Je ukázáno, že vhodnou volbou typu a parametrů modelu lze dále zlepšit účinnost komprese.

Nejvhodnější zkoumaná metoda pro kompresi popisovaných dat představuje kombinaci LZMA komprese, modelování dat na intervalech jejich charakteristickou (minimální) hodnotou a s daty modelu následně kódovanými diferencially. Délku a počet jednotlivých intervalů lze určit fixně, ale nejlepších výsledků dosahuje autorem navržená LogHist metrika. V práci jsou také diskutovány další výhody i nevýhody navrhovaných metod ukládání archivních dat.

# Literatura

- [1] IEEE Recommended Practices and Requirements for Harmonic Control in Electrical Power Systems. *IEEE Std 519-1992*, 1993, doi:10.1109/IEEESTD.1993.114370.
- [2] IEEE Trial-Use Standard Definitions for the Measurement of Electric Power Quantities Under Sinusoidal, Non-sinusoidal, Balanced, Or Unbalanced Conditions. *IEEE Std 1459-2000*, 2000: s. i–44, doi:10.1109/IEEESTD.2000.93398.
- [3] Abramson, N.: *Information theory and coding*. McGraw-Hill electronic sciences series, New York, NY: McGraw-Hill, 1963.
- [4] Akaike, H.: A New Look at the Statistical Model Identification. *IEEE Transactions on Automatic Control*, ročník 19, č. 6, 1974: s. 716–723.
- [5] Antonio, M. (editor): *Power Quality, Mitigation Technologies in a Distributed Environment*. Springer London, 2007, ISBN 978-1-84628-772-5, doi:10.1007/978-1-84628-772-5.
- [6] Balaji, A.; Sharma, G.; Shaw, M. Q.; aj.: Preprocessing Methods for Improved Lossless Compression of Color Look-up Tables. *Journal of Imaging Science and Technology*, ročník 52, 2008, doi:10.2352/J.ImagingSci.Technol.(2008)52:4(040901).
- [7] Bingham, R. P.; Kreiss, D.; Santoso, S.: Advances in Data Reduction Techniques for Power Quality Instrumentation.
- [8] Bollen, M. H. J.; Ribeiro, P. F.; Larsson, E. O. A.; aj.: Limits for voltage distortion in the frequency range 2 to 9 kHz. *IEEE TRANSACTIONS ON POWER DELIVERY*, ročník 23, č. 3, JUL 2008: s. 1481–1487, ISSN 0885-8977, doi:{10.1109/TPWRD.2008.919180}.
- [9] Brekke, K.; Seljeseth, H.; Mogstad, O.: Rapid voltage changes - Definition and minimum requirements. In *Electricity Distribution - Part 1, 2009. CIRED 2009. 20th International Conference and Exhibition on*, june 2009, ISSN 0537-9989, s. 1 –4.
- [10] Bubla, V.: *Virtuální měřicí a monitorovací přístroj v prostředí GNU/Linux*. Technická univerzita v Liberci, 2009.
- [11] Burrows, M.; Wheeler, D. J.: A block-sorting lossless data compression algorithm. *Technická Zpráva* 124, 1994.
- [12] Campos, A.: Range coder. 1999.  
URL [http://www.arturocampos.com/ac\\_range.html](http://www.arturocampos.com/ac_range.html)

- [13] CENELEC (editor): *ČSN EN 50160 ed. 3, Charakteristiky napětí elektrické energie dodávané z veřejné distribuční sítě*. Praha: Český Normalizační Institut, 2010.
- [14] Cooley, J. W.; Tukey, J. W.: An Algorithm for the Machine Calculation of Complex Fourier Series. *Mathematics of Computation*, ročník 19, č. 90, 1965: s. 297–301, ISSN 00255718, doi:10.2307/2003354.  
URL <http://dx.doi.org/10.2307/2003354>
- [15] Dabbs, W.; Sabin, D.: Employing Extensible Markup Language (XML) in IEEE Std 1159.3-2002 PQDIF. In *Transmission and Distribution Conference and Exhibition, 2005/2006 IEEE PES*, may 2006, s. 370–375, doi:10.1109/TDC.2006.1668521.
- [16] Deorowicz, S.: Context exhumation after the Burrows-Wheeler transform. *Information Processing Letters*, ročník 95, č. 1, 2005: s. 313 – 320, ISSN 0020-0190, doi: DOI:10.1016/j.ipl.2005.03.008.  
URL <http://www.sciencedirect.com/science/article/B6V0F-4G1R404-1/2/7b37c129d0d3fb836fa2e9e5651e10ba>
- [17] Fenwick, P.: Burrows-Wheeler compression: Principles and reflections. *Theor. Comput. Sci.*, ročník 387, November 2007: s. 200–219, ISSN 0304-3975, doi:http://dx.doi.org/10.1016/j.tcs.2007.07.012.  
URL <http://dx.doi.org/10.1016/j.tcs.2007.07.012>
- [18] Fortescue, C.: Method of Symmetrical Coordinates Applied to the Solution of Polyphase Networks. *Trans. AIEE*, ročník 37, 1918: s. 1027–1140.
- [19] Goldberg, D.: What every computer scientist should know about floating-point arithmetic. *ACM Comput. Surv.*, ročník 23, March 1991: s. 5–48, ISSN 0360-0300, doi:http://doi.acm.org/10.1145/103162.103163.  
URL <http://doi.acm.org/10.1145/103162.103163>
- [20] Gosbell, V.: Voltage Unbalance. *Electric Power Systems Research*, 2002, doi:10.1016/j.epsr.2003.07.010.  
URL <http://www.elec.uow.edu.au/iepqrc/files/technote6.pdf>
- [21] Huffman, D.: A Method for the Construction of Minimum-Redundancy Codes. *Proceedings of the IRE*, ročník 40, č. 9, Zář 1952: s. 1098–1101, ISSN 0096-8390, doi:10.1109/JRPROC.1952.273898.  
URL <http://dx.doi.org/10.1109/JRPROC.1952.273898>
- [22] IEC-TC13 (editor): *IEC 62053-21 Electricity metering equipment (a.c.) Particular requirements Part 21: Static meters for active energy (classes 1 and 2)*. Geneve: International Electrotechnical Commission, 2003.
- [23] IEC-TC13 (editor): *IEC 62053-23 Electricity metering equipment (a.c.) Particular requirements Part 23: Static meters for reactive energy (classes 2 and 3)*. Geneve: International Electrotechnical Commission, 2003.

- [24] IEC-TC77A (editor): *IEC 61000-4-15, Electromagnetic compatibility (EMC) - Part 4: Testing and measurement techniques - Section 15: Flickermeter - Functional and design specifications*. Geneve: International Electrotechnical Commission, 2003.
- [25] IEC-TC77A (editor): *IEC 61000-4-30, Electromagnetic compatibility (EMC) - Part 4-30: Testing and measurement techniques - Power quality measurement methods*. Geneve: International Electrotechnical Commission, 2003.
- [26] IEC-TC77A (editor): *Electromagnetic compatibility (EMC) - Part 4-7: Testing and measurement techniques - General guide on harmonics and interharmonics measurements and instrumentation, for power supply systems and equipment connected thereto*. Geneve: International Electrotechnical Commission, 2009.
- [27] IEC-TC85 (editor): *IEC 61557-12: Electrical safety in low voltage distribution systems up to 1000 V a.c. and 1500 V d.c. - Equipment for testing, measuring or monitoring of protective measures - Part 12: Performance measuring and monitoring devices (PMD)*. Geneve: International Electrotechnical Commission, 2007.
- [28] Kraus, J.; Bubla, V.: Optimal Methods for Data Storage in Performance Measuring and Monitoring Devices. In *EPE: PROCEEDINGS OF THE 9TH INTERNATIONAL SCIENTIFIC CONFERENCE ELECTRIC POWER ENGINEERING 2008*, 2008, ISBN 978-80-214-3650-3, s. 131–133, 9th International Scientific Conference on Electric Power Engineering, Brno, Czech Republic, MAY 13-15, 2008.
- [29] Kraus, J.; Tobiška, T.: SMP - Platform for Compact Power Quality Meter. In *Proceedings of ELEN 2008*, Prague, Czech Republic: CTU, Faculty of Electrical Engineering, Department of Electroenergetics, 2008, ISBN ISBN 978-80-254-2293-9.
- [30] Kraus, J.; Tobiška, T.: Metoda pro rychlé a korektní kontinuální měření jalového výkonu pro výpočetně omezené mikroprocesory. In *Electronic Power Engineering 2011*, Brno, Czech Republic: Brno University of Technology, 2010, ISBN ISBN 978-80-214-4094-4.
- [31] Kraus, J.; Tobiška, T.; Bubla, V.: Lossless encodings and compression algorithms applied on power quality datasets. In *Electricity Distribution - Part 1, 2009. CIRED 2009. 20th International Conference and Exhibition on*, 2009, ISSN 0537-9989.
- [32] Larsson, E. O. A.; Bollen, M. H. J.; Wahlberg, M. G.; aj.: Measurements of High-Frequency (2-150 kHz) Distortion in Low-Voltage Networks. *IEEE TRANSACTIONS ON POWER DELIVERY*, ročník 25, č. 3, JUL 2010: s. 1749–1757, ISSN 0885-8977, doi: {10.1109/TPWRD.2010.2041371}.
- [33] Liddle, A. R.: Information criteria for astrophysical model selection. Únor 2007, [astro-ph/0701113](https://arxiv.org/abs/astro-ph/0701113).  
URL <http://arxiv.org/abs/astro-ph/0701113>
- [34] Mahoney, M.: The ZPAQ Open Standard Format for Highly Compressed Data - Level 1. 2009.  
URL <http://www.mattmahoney.net/dc/zpaq1.pdf>

- [35] Math, B.: *Understanding power quality problems: voltage sags and interruptions*. New York: IEEE Press, 2009, ISBN 9780780347137.
- [36] Polívka, P.: *Effective Data Access in Large Database of Electrical Quantities Measurements*. Technická univerzita v Liberci, 2011.
- [37] Provozovatelé distribučních soustav (editor): *Pravidla provozování distribučních soustav. Příloha 3: Kvalita elektřiny v distribuční soustavě, způsoby jejího zjišťování a hodnocení*. Praha: Energetický regulační úřad, 2009.  
URL <http://www.eon-distribuce.cz/cs/distribuce-elektriny/>
- [38] Ribeiro, M. V.; Park, S. H.; Romano, J. M. T.; aj.: A novel MDL-based compression method for power quality applications. *IEEE TRANSACTIONS ON POWER DELIVERY*, ročník 22, č. 1, JAN 2007: s. 27–36, ISSN 0885-8977, doi:{10.1109/TPWRD.2006.887091}.
- [39] Sabin, D.: IEEE P1159.3 PQDIF Task Force Meeting. 2009.  
URL <http://grouper.ieee.org/groups/1159/3/docs/2009-07-chair-notes.pdf>
- [40] Salomon, D.; Motta, G.: *Handbook of Data Compression*. Springer London, 2010, ISBN 978-1-84882-903-9.
- [41] Santoso, S.; Powers, E.; Grady, W.: Power quality disturbance data compression using wavelet transform methods. *IEEE TRANSACTIONS ON POWER DELIVERY*, ročník 12, č. 3, JUL 1997: s. 1250–1257, ISSN 0885-8977, doi:{10.1109/61.637001}, 1996 IEEE/PES Summer Meeting, DENVER, CO, JUL 28-AUG 01, 1996.
- [42] Shannon, C. E.: A mathematical theory of Communication. *The Bell system technical journal*, ročník 27, Červenec 1948: s. 379–423.
- [43] Tobiška, T.: *Firmware přístroje pro měření kvality elektrické energie*. Diplomová práce, Technická univerzita v Liberci, 2008.
- [44] Tu, Z.; Zhang, S.: A Novel Implementation of JPEG 2000 Lossless Coding Based on LZMA. In *Computer and Information Technology, 2006. CIT '06. The Sixth IEEE International Conference on*, sept. 2006, str. 140, doi:10.1109/CIT.2006.19.
- [45] Štěpán, P.: *Návrh a realizace systému pro měření blikání - flickermeter*. Diplomová práce, Technická univerzita v Liberci, 2009.
- [46] Zbigniew, O.; Andrzej, W.: Methods of Voltage Unbalance Estimation in Electric Power Networks. *Electrical Power Quality and Utilisation*, ročník XII, č. 2, 2006: s. 72–76.
- [47] Ziv, J.; Lempel, A.: A universal algorithm for sequential data compression. *IEEE TRANSACTIONS ON INFORMATION THEORY*, ročník 23, č. 3, 1977: s. 337–343.



# Vybrané publikace autora<sup>1</sup>

- [1] Derbek, V.; Kraus, J.; Tichy, T.; aj.: Kansas State University's Autonomous Robot Borivoj: An Entry in the AAAI2002's Robot Host Competition. In *AAAI Mobile Robot Competition'02*, 2002, s. 15–17.
- [2] Kraus, J.: *Texado - Application Framework for a Distributed Agents Architecture*. Diplomová práce, Czech Technical University, 2004.
- [3] Kraus, J.: New Technologies in Automatic Power Factor Control - Comprehensive Solution for Power Quality. *ENERTIA*, 2009.
- [4] Kraus, J.: Černá skříňka pro vaši technologii - inovované měřicí přístroje SMV, SMP a SMPQ. *AZElektro*, 2009.
- [5] Kraus, J.: Černá skříňka pro vaši technologii - inovované měřicí přístroje SMV, SMP a SMPQ. *Elektrotechnika v praxi*, 2009.
- [6] Kraus, J.; Bleha, M.: *ENVIS Software User Guide for Supported Measuring Instruments*. KMB systems, s.r.o., Liberec, Czech Republic, 2010.
- [7] Kraus, J.; Bubla, V.: Optimal Methods for Data Storage in Performance Measuring and Monitoring Devices. In *EPE: PROCEEDINGS OF THE 9TH INTERNATIONAL SCIENTIFIC CONFERENCE ELECTRIC POWER ENGINEERING 2008*, 2008, ISBN 978-80-214-3650-3, s. 131–133, 9th International Scientific Conference on Electric Power Engineering, Brno, CZECH REPUBLIC, MAY 13-15, 2008.
- [8] Kraus, J.; Kukacka, L.; Bubla, V.: Data Modeling for Reduction of Volume in Large Archives of Power Quality Data. In *Electricity Distribution - Part 1, 2011. CIRED 2011. 21th International Conference and Exhibition on*, 2011, ISSN 0537-9989.
- [9] Kraus, J.; Tobiska, T.: Testing SMP Platforms for measuring and evaluating the quality of electricity. In *PROCEEDINGS OF THE 10TH INTERNATIONAL SCIENTIFIC CONFERENCE ELECTRIC POWER ENGINEERING 2009*, editace Rusek, S and Gono, R, 2009, ISBN 978-80-248-1947-1, s. 143–146, 10th International Scientific Conference Electric Power Engineering, Kouty nad Desnou, CZECH REPUBLIC, MAY 12-14, 2009.
- [10] Kraus, J.; Tobiska, T.: Method for Fast and Precise Calculation of Reactive Power for Computationally Limited Microprocessors. In *11TH INTERNATIONAL SCIENTIFIC CONFERENCE ELECTRIC POWER ENGINEERING 2010, PROCEEDINGS*, editace Drapela, J

---

<sup>1</sup>Číslování zde uvedené je pouze orientační a nesouvisí s číslováním citací v předchozím seznamu i textu.

- and Machacek, J, 2010, ISBN 978-80-214-4094-4, s. 563–566, 11th International Scientific Conference on Electric Power Engineering 2010, Brno, CZECH REPUBLIC, MAY 04-06, 2010.
- [11] Kraus, J.; Tobiška, T.: SMP - Platform for Compact Power Quality Meter. In *Proceedings of ELEN 2008*, Prague, Czech Republic: CTU, Faculty of Electrical Engineering, Department of Electroenergetics, 2008, ISBN ISBN 978-80-254-2293-9.
- [12] Kraus, J.; Tobiška, T.: Metoda pro rychlé a korektní kontinuální měření jalového výkonu pro výpočetně omezené mikroprocesory. In *Electronic Power Engineering 2011*, Brno, Czech Republic: Brno University of Technology, 2010, ISBN ISBN 978-80-214-4094-4.
- [13] Kraus, J.; Tobiška, T.; Bubla, V.: Looseless encodings and compression algorithms applied on power quality datasets. In *Electricity Distribution - Part 1, 2009. CIREN 2009. 20th International Conference and Exhibition on*, 2009, ISSN 0537-9989.
- [14] Kraus, J.; Václavík, J.: Monitoring of Power Consumption and Harmonic Distortion on City-Lighting System. In *Light Sources and Light Systems - Technology for the Future*, Heraklion, Crete: COST Strategic Workshop on Lighting, 2006.
- [15] Ruscassié, R.; Kraus, J.; Zissis, G.: Comparaison de Deux Types de Systemes Gradables pour Éclairage Urbain. In *EF 2007*, Toulouse, France: ENSEEIHT, 2007.