

УНИВЕРЗИТЕТ У БЕОГРАДУ
ЕЛЕКТРОТЕХНИЧКИ ФАКУЛТЕТ

мр Совјетка Крстонијевић

**ОКВИР ЗА ПРОГНОЗУ ПОТРОШЊЕ
ЕЛЕКТРИЧНЕ ЕНЕРИЈЕ У
ИНТЕЛИГЕНТНИМ
ЕЛЕКТРОЕНЕРГЕТСКИМ МРЕЖАМА
ЗАСНОВАН НА ГЕНЕРАЛИЗОВАНОМ
АДИТИВНОМ МОДЕЛУ**

докторска дисертација

Београд, 2024

UNIVERSITY OF BELGRADE
SCHOOL OF ELECTRICAL ENGINEERING

MSc Совјетка Крстонијевић

**FRAMEWORK FOR ELECTRICITY LOAD
FORECASTING IN SMART GRID
BASED ON A
GENERALIZED ADDITIVE MODEL**

Doctoral Dissertation

Belgrade, 2024

Подаци о ментору и члановима Комисије за усмену одбрану дисертације

Ментор:

др Предраг Пејовић, редовни професор
Универзитет у Београду — Електротехнички факултет

Чланови Комисије:

др Горан Добрић, ванредни професор
Универзитет у Београду — Електротехнички факултет

др Јелена Поповић Божовић, доцент
Универзитет у Београду — Електротехнички факултет

др Жељко Деспотовић, научни саветник
Универзитет у Београду — Институт Михајло Пупин

Датум усмене одбране: _____

Захвалница

Пре свега, захваљујем се ментору, проф. др. Предрагу Пејовићу, на указаном поверењу, подршци, вођењу кроз процес реализације научно-истраживачког рада и смерницама за израду и писање тезе.

Члановима Комисије се захваљујем на издвојеном времену, сугестијама и кориговању текста, што је омогућило финализацију тезе.

Преко две деценије рада у Институту Михајло Пупин били су од пресудног значаја за истраживачки рад. Будући да је теза проистекла из конкретног радног задатка, свакодневне консултације са колегама, размена мишљења, па и критике, помогли су да се сагледају практични аспекти истраживања и правци даљег развоја. Стога, хвала др Нинелу Чукалевском, Горану Јакуповићу и осталим сарадницима, на свему наученом, али и на инспиративној радној атмосфери.

Породици дугујем велику захвалност за подршку, разумевање и стрпљење. Посебно се захваљујем родитељима за све животне лекције из посвећености и истрајности, за безусловну љубав.

А мојој Каји - велико хвала на свему важном у животу.

Наслов докторске дисертације:

Оквир за прогнозу потрошње електричне енергије у интелигентним електроенергетским мрежама заснован на генерализованом адитивном моделу

Сажетак:

Дисертација се бави унапређењем функционалности краткорочне прогнозе потрошње електричне енергије (*short term load forecasting, STLF*) у правцу примене у будућим интелигентним електроенергетским мрежама. Нови услови децентрализованог и локалног управљања електроенергетским системима, у присуству производње из обновљивих извора енергије и утицаја локалног тржишта електричне енергије, чине постојеће прогностичке алате неадекватним. Алати за генерисање прогноза који су у употреби, махом, су намењени централизованом управљању на нивоу система или веће групе потрошача, развијани и прилагођени подацима за потрошњу са, типично, једног места у систему. Нова *STLF* функционалност треба да обезбеди задовољавајућу тачности прогнозе за више различитих прогностичких области (различитог типа и броја потрошача и осталих фактора који утичу на ниво утрошене енергије), за променљиве услове прогнозирања. Полазећи од овакве спецификације, у тези је предложена методологија која омогућава краткорочно прогнозирање потрошње за карактеристична места управљања у интелигентним електроенергетским мрежама, од нивоа веће области до индивидуалног потрошача. Основни кораци у развоју методологије су били избор приступа моделовања, развој функционалности која задовољава горње захтеве и евалуација на релевантним подацима. Прегледом литературе је констатован генерални недостатак истраживања приступа који би подржали функционалност задатих карактеристика, са изузетком тек неколико публикација новијег датума, што показује да је ова област истраживања на свом почетку. Анализом предности и недостатака постојећих метода (или истраживања), као оптималан приступ је одабран шири оквир генерализованог адитивног модела, ГАМ. Карактеристике ГАМ-а које га кандидују за развој нових *STLF* апликација су задовољавајућа тачност прогнозирања добијена за потрошње на различитим места у систему, једноставан поступак спецификације модела, брзи естимациони алгоритми, аутоматска процедура избора модела, интерпретабилност и могућност даљег проширења. ГАМ потрошње је збир *data-driven* подмодела појединачних компоненти потрошње (ГАМ чланова), стандардизованих зависно од нивоа агрегације дате потрошње, расположивости података (метеоролошке и остале спољашње варијабле) и задатог временског хоризонта (унутар дана, дан унапред). Развијена методологија се заснива на 1) каталогу ГАМ чланова потрошњи за дато место у електроенергетском систему и услове прогнозирања, и 2) процедури динамичког моделовања којим се редовно, са сваким новим податком, модел поново обучава и прилагођава подацима и у времену. За генерализовани каталог са ГАМ члановима за различита места у систему, 1) и 2) чине *оквир за прогнозу потрошње електричне енергије*. Методологија је евалуирана стандардним критеријумима за предиктивне моделе, за репрезентативан и свеобухватан сет потрошњи који обухвата

агрегиране потрошње на нивоу области са већим бројем потрошача (веће градске зоне, трафо рејоне), за комерцијалне и резиденцијалне зграде и индивидуалну потрошњу на нивоу домаћинства. За практичну примену, развијени оквир омогућава генерисање конкретног прогностичког *STLF* алата за произвољно место у систему, али и развој симулација рада на бази прогнозираних потрошњи, издвајање и прогнозирање појединачних компоненти потрошње, итд. Поред моделоване и прогнозиране вредности потрошње, за корисника су од значаја и опсези предиктивних интервала, што није разматрано у тези и може се сматрати њеним основним недостатком.

Главни доприноси дисертације су 1) систематизација литературе из предметне области и формализација захтева нових *STLF*, 2) систематизација карактеристика временске серије потрошње за различита места и хијерархијске нивое потрошње у електроенергетском систему и 3) прикупљање базе података која обухвата потрошње са статистички различитим карактеристикама (агрегирана, дисагрегирана потрошња и потрошња на нивоу зграда). Научни доприноси дисертације су 1) развој методологије која задовољава формулисане захтеве нове *STLF* функционалности и њена имплементација у *R* језику, 2) први предлог примене проширене верзије ГАМ-а за прогнозирање потрошње индивидуалног потрошача и 3) резултати краткорочне прогнозе, бољи до компарабилни у односу на референтне резултате добијене у литератури.

Кључне речи: краткорочна прогноза потрошње електричне енергије, интелигентне електроенергетске мреже, адаптивна прогноза потрошње електричне енергије, прогноза потрошње у зградама, прогноза потрошње индивидуалног потрошача, генерализовани адитивни модел, хијерархијски генерализовани адитивни модел.

Научна област: техничке науке, електротехника

Научна подобласт: електроника

Dissertation title:

A framework for electricity load forecasting in Smart Grid based on a generalized additive model

Abstract:

The dissertation considers the improvement of short-term electricity load forecasting (STLF) towards its application in Smart Grid. The new conditions of decentralized and local power system management, in the presence of renewable energy sources and the local electricity market, make existing prognostic tools inadequate. Most of the electricity load forecasts' generators, currently in use, are intended for centralized management, at the level of a system or a larger group of consumers, typically, developed and adapted to particular load in the system. The new STLF functionality should provide satisfactory accuracy for different forecasting areas, of different types and number of consumers and other load factors, for variable forecasting conditions. Starting from this specification, the thesis proposes a methodology for developing the STLF functionality that can be applied to all characteristic places of Smart Grid control. The main steps in its development were the choice of an appropriate modeling approach, the deployment of functionality that meets the above requirements and its evaluation on relevant data. A review of the literature revealed a general lack of research on holistic approaches that would support the functionality of the given characteristics, with the exception of a few recent publications, which shows that this area of research is in its infancy. By analyzing the advantages and disadvantages of existing methods (or research), a broader framework of the generalized additive model, GAM, was chosen as the optimal approach. The properties of GAM that candidates it for new STLF applications are competitive accuracy obtained for loads in different places in the system, simple model specification procedure, fast estimation algorithms, automatic model selection procedure, interpretability and the possibility of further expansion. GAM of electricity load is a sum of data-driven sub-models of individual load components (GAM terms), standardized for particular level of consumers' aggregation, availability of data (meteorological and other externals) and given time horizon (intraday, day ahead). The proposed methodology is based on 1) a dictionary of such GAM load terms for a particular load in system and for given forecasting conditions and 2) a procedure for dynamic GAM modeling, which regularly retrains the model and adapts it to the data and in time. For a generalized dictionary that includes GAM terms for different locations in the system, respecting different operating conditions, 1) and 2) form *a framework for short-term electricity load forecasting*. Standard criteria for predictive models were used to evaluate the proposed methodology for representative data sets, including aggregated load (larger urban areas, substations), for commercial and residential buildings and individual consumer. For practical application, the developed framework, primarily, enables the generation of a specific prognostic STLF tool for an arbitrary place in the system, but also, various simulations based on forecasted load values, the extraction and forecasting of individual load components, etc. In addition to modeled and forecasted electricity load values, the estimates of predictive intervals' are also practically important, which, however, is not considered in the thesis.

The main contributions of the dissertation are 1) the systematization of the literature in the subject area and the formalization of the requirements of the new STLF, 2) the

systematization of the main factors, features and drivers of the load time series for different places and hierarchical levels in the power system and 3) searching open databases and collecting data that includes loads with statistically different characteristics, (aggregated, disaggregated load and load from residential and commercial buildings). The main scientific contributions are 1) the development of a methodology that meets the formulated requirements of the new STLF functionality and its implementation in the *R* language, 2) the first propose of extended GAM version to individual load forecasting and 3) the results of short-term forecasts with the accuracy, better or comparable to those reported in literature.

Key Words: short-term load forecasting, Smart Grid, adaptive load forecasting, building load forecasting, household load forecasting, generalized additive model, hierarchical generalized additive model.

Scientific field: technical sciences, electrical engineering

Scientific subfield: electronics

Садржај

1. Увод.....	1
1.1 Предмет, циљ и значај истраживања.....	4
1.2 Анализа литературе и полазне хипотезе.....	5
1.3 Методе истраживања	11
1.4 Структура докторског рада	11
1.5 Објављени научни радови	12
2. Слободан софтвер и отворени подаци	13
2.1 Слободан софтвер коришћени у раду	13
2.2 Отворени подаци коришћени у раду.....	13
2.2.1 NYISO база података (NY).....	13
2.2.2 Ausgrid база података (AU)	14
2.2.3 EnerNOC база података (EN)	15
2.2.4 ASHRAE база података (AS).....	16
2.2.5 London Carbon Project база података (LCL)	16
3. Генерализовани адитивни модел, ГАМ	18
3.1 Избор типа базисних функција.....	20
3.2 Хијерархијски генерализовани адитивни модел, ХГАМ	26
3.3 Естимација ГАМ модела	27
3.4 Избор ГАМ модела	28
3.5 (Х)ГАМ примена на временске серије и методологија динамичког прогнозирања.....	30
3.5.1 Изведена (Х)ГАМ методологија за прогнозирање скупа временских серија	31
4. Карактеристике временских серија потрошње	33
4.1 Календарске променљиве.....	33
4.2 Метеоролошки фактори потрошње	36
4.3 Утицај претходних вредности потрошње.....	40
4.4 Карактеристике потрошње у зависности од нивоа агрегације потрошача.....	41
4.5 Одступање од регуларних профила потрошње и ефекат специјалних дана	45
4.6 Daylight saving	46
4.7 Компонента тренда потрошње.....	47
4.8 Карактеристике расподеле временске серије потрошње	47
5. Резултати.....	49
5.1 Верификација модела и предиктивне перформансе	49
5.2 Пример потрошње већих градских зона – NY база података	51
5.2.1 Резултати анализе резидуала за NYISO потрошње	56

5.2.2	Неколико коментара о резултатима селекције ГАМ модела	60
5.2.3	Тестирање предиктивних перформанси модела	61
5.2.4	Резултати прогнозе за 24 сата унапред	63
5.2.5	Кратак резиме и коментари.....	66
5.3	Пример потрошње произвољне области у дистрибутивном систему – АУ база података	67
5.3.1	Кратак резиме.....	72
5.4	Пример потрошње у зградама – ЕН и АС база података.....	73
5.5	Пример потрошње индивидуалног и произвољне агрегације потрошача – LCL база података	79
5.5	Кратак резиме.....	85
6.	Закључак	87
6.1	Преглед научних доприноса	88
6.2	Практична примена.....	89
6.3	Ограничење истраживања	90
6.4	Правци даљег истраживања	90
	Референце	91

Списак слика

Слика 1.1: Број публикованих радова за петогошње интервале за категорије прогностичких алата зависно од временског хоризонта	3
Слика 1.2: Број публикованих радова који се баве <i>STLF</i> прогнозирањем локалних потрошњи... 7	7
Слика 2.1: Сатна потрошња за NORTH зону, за период 2009-2013 године	14
Слика 3.1: Пример базисних функција кардиналног кубног сплајна	22
Слика 3.2: Пример базисних функција цикличног кубног сплајна.....	23
Слика 3.3: Пример базисних функција <i>TPRS</i> сплајна.....	23
Слика 3.4: Тензорски производ базисних функција цикличног кубног сплајна.....	24
Слика 3.5: Моделовање годишње сезоналности стандардним кубним сплајном (NYC зона)	25
Слика 3.6: Моделовање годишње сезоналности цикличним кубним сплајном (NYC зона)	26
Слика 3.7: Временска шема унакрсне валидације за временске серије.....	28
Слика 3.8: Временска шема процедуре за прогнозу са клизајућим прозором.....	31
Слика 3.9: Шема ГАМ базиране прогнозе више независних временских серија	32
Слика 3.10: Шема ХГАМ базиране прогнозе више корелираних временских серија.	32
Слика 4.1: Сатне вредности потрошње током године и летње/зимске седмице за NYC зона	34
Слика 4.2: Сатне вредности потрошње током године и летње/зимске седмице за AU рејон.....	34
Слика 4.3: Сатне вредности потрошње током године и летње/зимске седмице за EN зграду	34
Слика 4.4: Сатне вредности потрошње током године и летње/зимске седмице за 2 индивидуална LCL потрошача	34
Слика 4.5: Зависност потрошње од температуре за GENESE зону, за дане у седмици (горе), за месеце у години (доле)	37
Слика 4.6: Зависност потрошње од температуре за EN1 зграду, за дане у седмици (горе), за месеце у години (доле)	38
Слика 4.7: Промене дневног профила потрошње електричне енергије за различите нивое пенетрације солара.....	40
Слика 4.8: Зависност потрошње од основних фактора за 1, 50 и 600 индивидуалних потрошача	41
Слика 4.9: Зависност MAPE грешке од нивоа потрошње за SARMA(1,0)(1,0) ₂₄ модел	42
Слика 4.10: Пример хетерогене (горе) и хомогене (доле) групе од 10 LCL индивидуалних потрошача.....	44
Слика 4.11: Ефекат специјалних дана на примеру <i>Дана сећања</i> за NYC зону	45
Слика 4.12: <i>DST</i> ефекат за 2013. годину за NYC зону.....	46
Слика 4.13: Приказ расподеле потрошње за а), б) и ц)	47
Слика 4.14: Приказ расподеле трансформисане потрошње за а), б) и ц)	48
Слика 5.2.1: Резултати моделовања моделом M1	52
Слика 5.2.2: Резултати моделовања моделом M2	52
Слика 5.2.3: Резултати моделовања моделом M3	53
Слика 5.2.4: Аутокорелација/парцијална корелација резидуала M3 модела за 8.сат.....	54
Слика 5.2.5: Резултати моделовања моделом M4	54
Слика 5.2.6: Аутокорелација/парцијална корелација резидуала M4 модела за 8.сат.....	54

Слика 5.2.7: Аутокорелација/парцијална корелација резидуала М5 модела за 8.сат.....	55
Слика 5.2.8: Резултати моделовања моделом М6.....	55
Слика 5.2.9: Аутокорелација/парцијална корелација резидуала М6 укупне потрошње.....	56
Слика 5.2.10: Аутокорелација/парцијална корелација резидуала М7 укупне потрошње.....	56
Слика 5.2.11: Резултати анализе резидуала за 2. сат, NYC зона.....	58
Слика 5.2.12: Резултати анализе резидуала за 15. сат, NYC зона.....	58
Слика 5.2.13: Реалне и моделова сатне вредности за период 2009-2012 год. за NORTH зону ...	59
Слика 5.2.14: Временски график и аутокореалција резидуала за 3. сат, NYC зона.....	60
Слика 5.2.15: Расподела MAPE/MAE грешке по сатима за NYISO зоне.....	62
Слика 5.2.16: Ефекат моделовања специјалних дана на примеру Дана независности, NYC.....	65
Слика 5.3.1: Резултати прогнозе за <i>RNS Hospital</i> трафо рејон.....	70
Слика 5.3.2: Резултати прогнозе за <i>Leichhardt</i> трафо рејон.....	71
Слика 5.3.3: Расподела APE грешке по сатима за <i>Cronulla</i> трафо рејон за 11. август 2022.год.	71
Слика 5.3.4: Резултати анализе резидуала модела потрошње 15. сата за <i>Cronulla</i> трафо рејон.	72
Слика 5.3.5: Ефекат агрегације потрошње за AU трафо рејоне на грешку прогнозе.....	73
Слика 5.4.1: Сатне потрошње за 40 дана за зграде EN1 и EN2.....	74
Слика 5.4.2: Сатне потрошње за 40 дана за зграде AS1 и AS3 зграду.....	74
Слика 5.4.3: Резултати анализе резидуала модела потрошње 15. сата за EN1 зграду.....	76
Слика 5.4.4: Резултати прогнозе за EN1 зграду, за летњи период.....	77
Слика 5.4.5: Резултати прогнозе за EN1 зграду, за зимски период.....	77
Слика 5.4.6: Резултати прогнозе за EN2 зграду, за зимски период.....	77
Слика 5.4.7: Резултати прогнозе за EN2 зграду, за летњи период.....	78
Слика 5.4.8: Резултати прогнозе за AS1 зграду.....	78
Слика 5.4.9: Резултати прогнозе за AS2 зграду.....	79
Слика 5.4.10: Резултати прогнозе за AS3 зграду.....	79
Слика 5.5.1: Резултати анализе резидуала ХГАМ модела потрошње 15. сата, за одабраног LCL потрошача.....	81
Слика 5.5.2: Резултати ГАМ прогнозе за одабраног потрошача.....	81
Слика 5.5.3: Резултати ГАМ прогнозе репрезента потрошње за произвољну групу потрошача..	82
Слика 5.5.4: Резултати ГАМ прогнозе репрезента потрошње за хомогену групу потрошача.....	83
Слика 5.5.5: Резултати анализе резидуала ХГАМ модела потрошње 15. сата, за одабраног LCL потрошача из произвољне групе потрошача.....	84
Слика 5.5.6: Резултати анализе резидуала ХГАМ модела потрошње 15. сата, за одабраног LCL потрошача из хомогене групе потрошача.....	84
Слика 5.5.7: Резултати ХГАМ прогнозе индивидуалног потрошача из произвољне и хомогене групе потрошача.....	85

Списак табела

Табела 2.1: Листа и карактеристике <i>NYISO</i> зона	14
Табела 2.2: Листа и просечна потрошње <i>Ausgrid</i> зона	15
Табела 2.3: <i>EnerNOC</i> зграде, њихова просечна потрошња, летњи и зимски максимум	16
Табела 2.4: <i>ASHRAE</i> резиденцијалне зграде, њихова просечна потрошња и опис	16
Табела 3. 1: Типови ГАМ чланова који се користе у раду	19
Табела 5.2.1: Каталог ГАМ чланова потрошње за <i>NYISO</i> зоне	57
Табела 5.2.2: MAPE/MAE резултати за један сат унапред, за <i>NYISO</i> зоне.....	61
Табела 5.2.3: Упоредни MAPE/MAE резултати прогнозирања модела добијеног поступком унакрсне валидације и предложеном методологијом.....	62
Табела 5.2.4: MAPE/MAE резултати за 24 сата унапред, за <i>NYISO</i> зоне.....	63
Табела 5.2.5: MAPE резултати без/са урачунатим специјалним данима, за <i>NYISO</i> зоне.....	64
Табела 5.2.6: MAPE резултати без/са урачунатим <i>DST</i> ефектом, за <i>NYISO</i> зоне	65
Табела 5.3.1: Генерализовани каталог ГАМ чланова потрошње.....	68
Табела 5.3.2: MAPE/MAE резултати за један и 24 сата унапред, за AU прогностичке области .	69
Табела 5.4.1: Каталог ГАМ чланова потрошње за зграде	75
Табела 5.4.2: MAPE, MAE, CVRMSE и MBE вредности за EN зграде, за летњи и зимски период	76
Табела 5.4.3: MAPE, MAE, CVRMSE и MBE вредности за AS зграде	78
Табела 5.5.1: Каталог ГАМ чланова потрошње за индивидуалног потрошача	80
Табела 5.5.2: Каталог ХГАМ чланова потрошње за индивидуалног потрошача.....	83

1. Увод

Актуелне промене у енергетском сектору које се односе на нове начине потрошње и производње електричне енергије представљају изазов за постојеће конвенционалне електроенергетске системе (ЕЕС). Поред константног раста захтева за електричном енергијом, присуство све већег броја обновљивих извора енергије (ОИЕ) и нових потрошача, као што су електрична возила, значајно повећава ниво неодређености у систему. Због своје интермитентне и стохастичке природе која прети да угрози ниво квалитета и поузданост напајања у ЕЕС-у, управљање овим ресурсима је сложеније и превазилази могућности постојећих система. С друге стране, тржиште електричне енергије, које прати динамику дешавања у ЕЕС-у, поставља стриктне захтеве који треба да обезбеде захтевану ефикасности рада, техничку и економску. С обзиром да се већина промене дешава на дистрибутивном нивоу и локално, са аспекта конвенционалног тзв. “*top-down*” концепта рада ЕЕС-а су “неприступачне” за проактивно и ефикасно деловање и управљања системом се помера ближе њиховом извору, чиме се оно децентрализује и локализује. Значајан корак ка напреднијем ЕЕС-у је интегрисање комуникационо-информационих технологија са постојећом инфраструктуром у оквиру концепта интелигентних мрежа. Тиме је, пре свега, омогућена доступност мерењима релевантних величине на различитим местима и хијерархији ЕЕС-а. Порастом броја имплементираних паметних бројила, омогућава се ефикаснија имплементација програма управљања потрошњом (*demand side management, DSM¹*), одзива потрошње (*demand response, DR²*) и могућност активног учешћа потрошача у управљању тржиштем електричне енергије. Закључак је да се управљање савременим и будућим ЕЕ системима заснива на захтевном задатку оптимизовања ресурса и чинилаца унутар одговарајућих временских периода које осигурава квалитетно и одрживо снабдевање електричном енергијом, обезбеђујући економске бенефите како за испоручиоце тако и за потрошаче.

У раду *SPEN – DSO VISION* [1.1], из 2017. године, који се бави аспектима транзиције електроенергетских система и њиховом даљом перспективом, као главни механизми ефикаснијег и ‘интелигентнијег’ управљања будућим ЕЕ системима се наводе интезивирање праћења рада система, његова аутоматизација, заштита података и, значајно, унапређење модела и предиктивних алата. Будући електроенергетски системи се, суштински, заснивају на *real-time* подацима за праћење и аквизицију стања у мрежи и на тржишту, и софтверским алатима који те податке користе за благовремено одлучивање и локално деловање. С обзиром на критичне временске оквире и велики проток података, од посебног је значаја да сва решења буду аутоматизована и да осигуравају задати ниво флексибилности и прилагодљивости, аутономно, без интервенције човека. Правовремене одлуке и активности у ЕЕС-у су засноване на проценама стања система и релевантних величина за краће и дуже будуће периоде, где је њихова тачност економски и технички критична варијабла. Из тог разлога су прогностички алати кључни за ефикасно управљање ЕЕС-ом, од којих су они

¹ **Управљање потрошњом** подразумева скуп активности испоручиоца електричне енергије са циљем да изазову жељене промене у облику дијаграма потрошње, у времену и нивоу, [1.17].

² **Одзив потрошње** је механизам управљања потрошњом зависно од услова снабдевања, [1.17].

намењени прогнози потрошње електричне енергије од највећег значаја, судећи по резултатима анкете *SAS*-а, спроведене међу 136 енергетских компанија [1.2].

Прогнозирање потрошње електричне енергије која обухвата и примену у интелигентним мрежама је значајно комплекснија и потреба за њом постоји на нивоу веће области или његове подобласти [1.3], трафо рејона [1.4], зграде [1.5] или индивидуалног потрошача [1.6]. Прогностички алати у ЕЕС-у се користе као подршка оптималном оперативном планирању и управљању [1.7], за генерисање подлога за имплементацију програма управљања потрошњом [1.8], за правовремену детекцију аномалија у мрежи [1.9], координацију пуњења електричних возила [1.10], развој стратегија енергетске ефикасности у зградама [1.11], различите апликације на бази мерења са паметних бројила [1.12], итд.

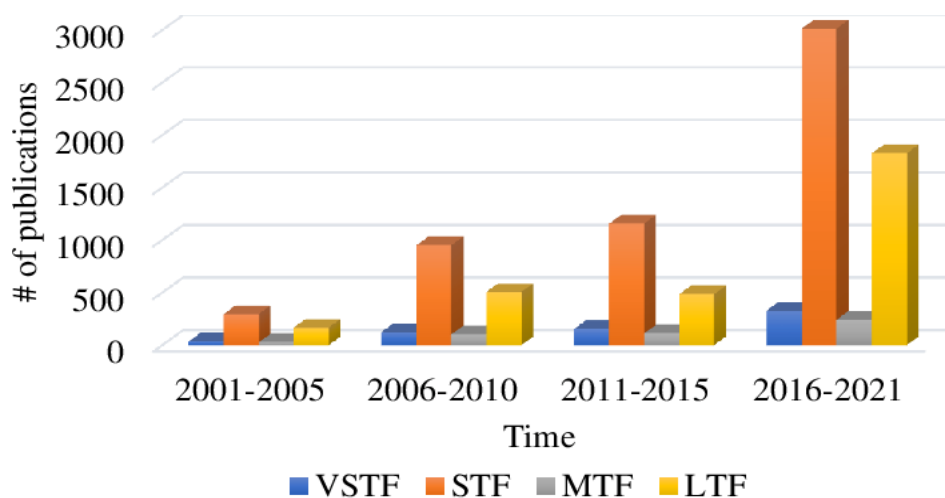
Временски хоризонти прогнозирања потрошње који су од интереса обухватају периоде од наредног минута до неколико година унапред, категоризовано у врло краткорочну (*very short term load forecasting, VSTLF*), краткорочну (*short term load forecasting, STLF*), средњерочну (*middle term load forecasting, MTLF*) и дугорочну (*long term load forecasting, LTLF*) функционалност. У литератури постоји неконзистентност у дефинисању дужине ових интервала, као и њихове примене, и интерпретација се може разликовати од публикације до публикације, што се види из Табеле 2 у раду [1.13]. Међутим, најчешћа категоризација конвергира оној датој у [1.14] по којој:

- *VSTLF* функционалност генерише прогнозе за период од једног минута до неколико сати унапред (0-3h),
- *STLF* прогнозе обухватају период од наредног сата (или полусата) до неколико дана унапред,
- *MTLF* функционалности покрива период од неколико дана до неколико месеци унапред и
- *LTLF* је намењен генерисању прогноза од неколико месеци до неколико година.

За већину дневних оперативних задатака у систему од највећег значаја је временски хоризонт од наредног сата (полусата) до наредног дана, (*intra-day* и *day-ahead* функционалност), што спада у домен краткорочне прогнозе потрошње (*STLF*), што је и предмет истраживања у тези. Последице, већина публикација која је посвећена развоју предиктивних алата је намењена, управо, овој функционалности. Слика 1.1, преузета из поменутог рада [1.14] показује расподелу публикованих радова за петогодишње интервале, у периоду од 2001. до 2021. године на основу временског хоризонта прогнозе, где се уочава доминантно већи број публикација посвећених *STLF* у односу на друге категорије прогноза.

Такође, значајан и нагли пораст публикованих радова који се баве краткорочном прогнозом потрошње последњих година, указује на потребу за даљим унапређењем. Основни разлог је што су постојећи *STLF* алати који су тренутно у употреби и који су верификовани у пракси, махом, намењени централизованом управљању, на нивоу система или веће групе потрошача, значајно, развијани и прилагођени подацима прикупљеним са конкретног места и прилагођени конкретној потрошњи. Међутим, код дистрибутивног управљања, *STLF* функционалност је намењена прогнозирању више локалних потрошњи које обухватају различите прогностичке области и подобласти (укључујући и индивидуалног потрошача на нивоу домаћинства), који се, генерално, разликују по величини, релевантним факторима потрошње (утицај метеоролошких и економских параметара, зависност од календарских варијабли, итд.), комплексности и броју и саставу обухваћених ОИЕ јединица. Изазов у развоју генералног приступа њиховог прогнозирања представља велика хетерогеност

профила потрошње, која захтева већи ниво адаптивности, генерално, несвојствен конвенционалним прогностичким алатима.



Слика 1.1: Број публикованих радова за петогошње интервале за категорије прогностичких алата зависно од временског хоризонта

Од посебног значаја је и чињеница да доступност мерених потрошњи на различитим хијерархијским нивоима и за различите агрегације потрошача омогућава развој прогностичког алата који је у стању да обавља прогнозу за *различите напонске нивое у систему*. С обзиром да је оваква функционалност пројектована за будуће интелигентне мреже, развој тзв. хијерархијског прогнозирања потрошње³ представља императив даљег развоја *STLF*-а. Овим се, међутим, у проблематику уводи додатни ниво хетерогености профила, што захтева већи ниво адаптивности прогнозирања потрошње, вертикално у ЕЕС-у.

Други проблем је што, на локалном нивоу, потрошња мањег броја потрошача има већу варијабилност и комплексније временске профиле. Како се прогноза потрошње заснива на моделу временске серије, ово се, пре свега, односи на одабир одговарајућег модела и/или методологије и идентификацију одговарајућих фактора потрошње. За већу групу потрошача, са географски веће области, главни фактори потрошње су утицај метеорологије (за коју, у већини случајева, постоје мерења) и, релативно, предвидљиви годишњи, седмични и дневни профили укупне потрошње за целу област. Са друге стране, потрошња индивидуалног потрошача је доминантно одређена стохастичким навикама употребе електричне енергије у домаћинству (нетривијално за идентификовање и мерење), са слабије предвидљивим обрасцима понашања потрошње. За развој хијерархијског приступа, проблем је у избору модела и методологије који урачунава разлике у комплексности профила и статистичким карактеристикама временских серија потрошње за различит број потрошача.

Такође, од посебног значаја је могућност динамичког прилагођавања модела на услове прогнозирања. Модел иницијално обучен за дате вредности потрошње временом више није актуелан и не даје добре резултате, те је потребно његово поновно обучавање на новим вредностима. Разлог су, пре свега, промене основних фактора потрошње у времену, поготово утицај метеоролошких параметара. Потреба за динамичким прилагођавањем је израженија у

³ Термин “хијерархијско прогнозирање потрошње” је увео *Tao Hong* у раду [1.18]

новим условима у систему, у присуству ОИЕ и имплементираних *DSM/DR* ефеката, који узрокују промене потрошње на краћим временским интервалима. У раду [1.15], где је анализирана применљивост стандардно коришћених модела на локалну потрошњу, као примарни недостаци који онемогућавају њихову примену у новим условима су наведени дуго време обучавања модела и нестабилни оптимизациони алгоритми, као и велика осетљивост на хиперпараметре (параметри структуре модела) који се, за већину модела у употреби, дефинишу дуготрајном процедуром. Оба кључна разлога из поменуте референце представљају директна ограничења у реализацији динамичког прилагођавања *STLF*-а.

Коначно, карактеристика оптималног прогностичког алата, је и транспарентност. Савремени *STLF* алати не представљају само тзв. прогностичке машине (*forecasting engine*), већ имају и ширу предиктивну сврху. С обзиром да се ослањају на велики број података, како техничких из мреже, тако и из спољашњих извора (метеоролошки, географски информациони системи, економски параметри), користе се и за низ других апликација, на пример, процену понашања потрошача, анализу потрошње, итд. Због тога се, као додатна ставка при избору оптималног *STLF* модела, разматрају интерпретабилне технике моделовања. Неки аспекти ове карактеристике *STLF* модела дати су у [1.16].

Већина постојећих прогностичких алата, а и истраживања уважава специфичности места у ЕЕС-у или примене на конкретан задатак. На пример, истраживања која се баве *STLF* алатима за дистрибутивне системе се, махом, фокусирају на решавање проблема адаптације за различите прогностичке области, док, на пример, она за ниво индивидуалног потрошача, на комплексност профила потрошње. Међутим, развој приступа који би задовољио опште услове и критеријуме рада у интелигентним мрежама, који подразумева одговарајућу тачност прогнозе, прилагодљивост у простору и ЕЕС хијерархији и у времену, применљивост за конкретан прогностички задатак, репродуцибилност и транспарентност, захтева даље истраживање.

1.1 Предмет, циљ и значај истраживања

С обзиром на захтеве наведене у претходном поглављу, **предмет истраживања** ове тезе је развој генералног приступа краћокорочне прогнозе потрошње електричне енергије, као подлоге за генерисање *STLF* алата за ширу примену на различитим местима у новом ЕЕС-у.

Уведене претпоставке и полазишта истраживања су:

Контекст примене у интелигентним мрежама у тези, а који стоји и у њеном наслову, се, преваходно, односи на претпоставку о доступности података о потрошњи на различитим местима у систему и уважавање њихових разлика, као предуслов за реализацију *STLF* функционалности и њено тестирање.

За развој конкретне *STLF* функционалности, на основу карактеристика наведених у претходном поглављу, могуће је формулисати 4 основне ставке спецификације оптималног прогностичког алата, које служе као смерница при избору одговарајуће методе:

1. С обзиром на разлике у комплексности профила за различите агрегације потрошача, која се повећава за локалне потрошње и за мањи број потрошача, главни критеријум при избору методологије је могућност постизања задовољавајуће тачности прогнозе.
2. Прогностички алати у интелигентним мрежама намењени су прогнозирању великог броја хетерогених временских серија потрошње, прикупљених са више места у систему и за различите агрегације потрошача. Могућност просторне, и генерално, хијерархијске

адаптације прогнозирања потрошње и очувања квалитета прогнозе је додатни критеријум за избор модела.

3. Услови прогнозирања се мењају током времена и како би се квалитет прогнозирања аутоматски прилагодио, потребна је методологија динамичког прилагођавања условима прогнозирања.
4. При избору модела и методологије је преферабилно усвајање интерпретабилних модела.

Значај развоја *STLF* функционалности горе наведених карактеристика је у томе што би омогућио њену ширу примену у ЕЕС-у и имплементацију важних апликација које се заснивају на прогнозама потрошње. Сходно томе, постављени **циљеви истраживања** су:

- претрага, анализа и избор одговарајућег модела и/или методологије за развој генералног приступа прогнозирања потрошње електричне енергије који задовољава горњу спецификацију,
- реализација конкретне *STLF* функционалности за одређено место у ЕЕС-у. Резултати добијени за потрошње за карактеристична места управљања у интелигентним мрежама, служе за евалуацију предложеног приступа.

1.2 Анализа литературе и полазне хипотезе

Први корак истраживања је преглед литературе из предметне области са циљем констатовања нивоа решености проблема у свету и претраге оптималног приступа моделовања. Прогнозирање потрошње електричне енергије је због свог значаја дуго изучавано и као основно полазиште у разматрању опсежне литературе кориштени су прегледни радови који нуде систематичан приступ проблематици и анализу актуелног стања. Претраживани су резултати објављени, већином, за период последњих 20 година, филтрирани по релевантности, након чега је одабрано 34 публикације, датих референцама [1.19-1.52]. Осим резултата који се нуде у одабраним прегледима, детаљно је прегледано неколико публикација са конкретним моделом, приступом и резултатима, наведено у наставку.

Прелиминарно, истраживања се крећу у два главна правца. С једне стране је фокус на развоју конкретне *STLF* функционалности која урачунава специфичности потрошњи и система за који се развија. Ови приступи најчешће користе, унапређују и прилагођавају већ постојеће методе моделовања и тиме дају допринос њиховој верификацији и смернице за даљи развој. С друге стране су истраживања која се баве развојем нових метода, које, често, користе пример временске серије потрошње за евалуацију. Оваква истраживања махом прате развој савремених техника машинског учења и вештачке интелигенције, изучавају њихове предности и алгоритамска ограничења и препоручују нове методологије. Као резултат оваквих тенденција, историјска подела метода које се користе за прогнозу потрошње је на конвенционалне, засноване, типично, на статистичким методама и оне савремене, којима је обухваћене широка област техника машинског учења и вештачке интелигенције. (У литератури се често може наћи и терминологија и подела на статистичке и интелигентне [1.19]). Због значајног броја публикованих радова, резултати у којима се примењују технике дубоког учења се могу издвојити као засебна категорија. У пракси, с друге стране, значајно место заузимају методологије које комбинују више модела, као што су хибридни модели и ансамбли.

С обзиром на велики број публикација који се бави *STLF* алатом и несагледив број примењиваних модела и методологија, спроведена је анализа оних приступа који се најчешће користе и који су највише заступљени у публикацијама, са новијим датумом. Претпоставка је да модели који су ‘опстали’ током дугогодишњег истраживања и који су верификовани у пракси имају перспективу за даљи развој. На примеру 113 различитих апликација краткорочне прогнозе из 41 публикације у прегледу из референце [1.52], 240 радова публикованих за период од 2000. до 2019. године из [1.25] и 77 рада, публикованих у периоду од 2010. до 2020. године у прегледу [1.20], као најчешћи приступи који се користе за решавање проблема краткорочне прогнозе потрошње су:

- Модели вишеструке регресије (*multiple linear regression, MLR*)
- Модели временских серија, (*ARMA, ARIMA* и варијанте са сезоналностима)
- Модели стабла одлучивања (*decision tree*), од којих је посебно коришћен модел случајне шуме (*random forest, RF*),
- Модел потпорних вектора (*support vector regression, SVR*),
- Неуралне мреже (*neural network, NN*), од којих је најчешће у употреби *multi-layer perceptron, MLP* и *radial basis functions, RBF*,
- Дубоке неуралне мреже (*deep neural network, DNN*), поготово су интересантне топологије за моделовање временских серија као што су, *recurrent neural network (RNN)*, *convolutional neural network (CNN)* и, посебно, *long-short term memory (LSTM)* мреже.

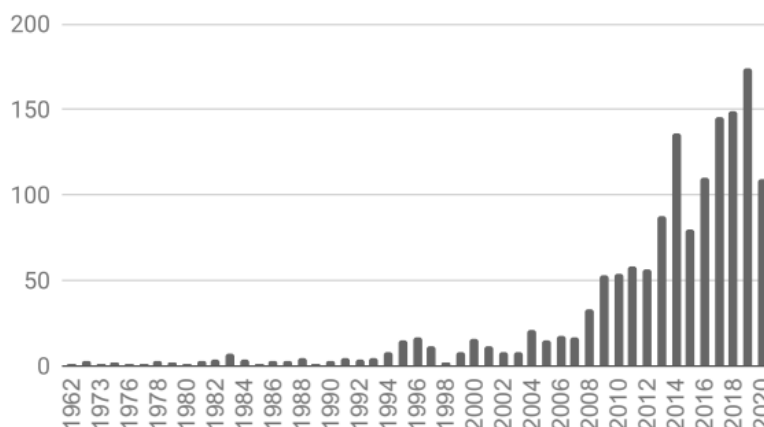
Поред појединачних типова модела, у пракси значајно место заузимају методологије које комбинују два или више модела, као што су хибридни модели и ансамбли.

На основу прегледа више од 300 публикација из наведених референци, закључено је да је за реализовање *STLF* функционалности најчешће у употреби неки од типова неуралних мрежа. До истог резултата се дошло и у [1.53] где су се на основу 160 радова и 60 пројеката показале као преферабилан избор за генерисање подлога за одзиве потрошње, *DR*. Упоредивањем различитих топологија и типова неуралних мрежа, а на основу проблема који се решава, коришћених алгоритама, улазних варијабли, карактеристика података и добијених резултата, из 51 рада, обрађеног у [1.54], се показало да највећу тачност прогнозе дају *RNN* и *LSTM* модели. Међутим, аутори указују на чињеницу да се оваква статистика, изведена на основу сингуларних резултата за специфичне податке, не може генерализовати и да је потребна даља верификација ових модела над већим скупом података. И иначе, као кључни недостаци неуралних мрежа у [1.51] наводи се потенцијално слаба генерализација за податке који су ван опсега оних за које су обучавани и стандардан проблем преобучавања, осетљивост на избор хиперпараметара модела и немогућност динамичке адаптације на нове услове прогнозирања и, коначно, недостатак интерпретабилности.

С друге стране, поред тачности прогнозирања, у практичним условима избор модела и методологије зависи од примене и конкретних захтева који се имају за предиктивним алатом. На пример, у раду [1.55] из 2022 су испитана 264 различита модела и неколико десетина методологија из 47 публикације, како би се унапредиле перформансе постојећег прогностичког алата за Пољски ЕЕС. Као најекономичније решење су се показали модели временских серија. Детаљнија студија која се бави компарацијом 9 најпопуларнијих модела који се користе за прогнозирање је [1.56], где су аутори упоредили временске серије и регресионе моделе са дубоким мрежама, за потрошње неколико кинеских провинција, је показала да се највећа тачност прогнозе добија за *MLR* и *SVR* модели. Из сличних примера је могуће закључити да теоријски напредније технике моделовања нису гарант унапређења у

практично значајним условима прогнозирања. У неколико радова (као, на пример, у [1.57]) се истиче да је од већег значаја припрема и препроцесирање података који се користе за обучавање модела и прецизан избор одговарајућих параметара модела, од самог типа модела. Предности и недостаци појединачних модела који се користе за развој *STLF* функционалности, намећу потребу њиховог комбиновања, са намером да се побољша тачност прогнозирања. Хибридни модели се пројектују тако да искористе добре особине различитих приступа за унапређење. На пример, у [1.58] је временска серија потрошње декомпонована на нискофреквентну компоненту, која је моделована линеарним моделом, и високофреквентну за коју је коришћен *LSTM* модел. У литератури се, типично, налазе хибриди засновани на *SVR* и *NN* моделима, и један од детаљнијих прегледа је дат референцом [1.59]. Међу дубоким неуралним мрежама, најбољу тачност прогнозирања даје комбинација *CNN-LSTM*, како је закључено у [1.60] на основу 1000 прегледаних публикација. Међутим, иако се комбиновањем индивидуалних модела, свакако, може очекивати побољшање тачности моделовања, главни недостатак комбинованих приступа може бити велико време извршавања које представља ограничење за неке *STLF* примене.

Други критеријум при прегледу литературе који је био од значаја је избор оптималне методе прогнозирања зависно од места у систему, поготово, хијерархијски, за различит број потрошача. Карактеристике временске серије потрошње се разликују зависно од типа и броја потрошача којим је обухваћена прогностичка област, превасходно, због разлика у факторима потрошње (на пр., за индустријске и комерцијалне потрошаче је утицај метеорологије мање значајан у односу на резиденцијалне потрошаче). То има за последицу разлике у оптималним моделима за различите агрегације потрошача. На пример, у [1.61] су аутори дошли до закључка да је за индивидуалног потрошача најбоља примена *MLP*, док је док је *SVR* бољи за агрегације потрошача веће од 32 домаћинства. Овакве недоумице су интензивирале истраживања која се баве развојем *STLF*-а на локалном нивоу, што показује и тренд броја радова на слици 1.2, преузетој из прегледа [1.32].



Слика 1.2: Број публикованих радова који се баве *STLF* прогнозирањем локалних потрошњи

Оно што је констатовано је да је већина резултата из поменутог прегледа посвећена прогнози на нивоу индивидуалног потрошача, док је број радова који се бави потрошњама на осталим нивоима нисконапонске мреже знатно мањи. Као разлог се наводи недостатак података и отворених база које би се користила за развој и тестирање.

На основу статистичких карактеристика временске серије потрошње, профила и фактора потрошње које су значајне за реализацију *STLF* функционалности, могуће је идентификовати 3 нивоа потрошње у систему: 1) ниво индивидуалне потрошње (неагрегирана потрошња), на бази мерења са паметних бројила индивидуалног потрошача, 2) ниво агрегиране потрошње (област, подобласт) која се добија или мерењем на одређеном месту у систему, или агрегацијом индивидуалних потрошњи и 3) ниво потрошње објекта одређене намене, односно, комерцијалне, индустријске и резиденцијалне зграде, [1.62].

Већина литературе је посвећена агрегираној потрошњи. Специфичност истраживања које се бави прогнозом на нивоу индивидуалног потрошача је - доминантна примена дубоких неуралних мрежа, поготово *LSTM*. Будући да се ради о моделима великог капацитета, основна полазна претпоставка је да, управо, ови модели имају најбоље шансе за постизање задовољавајуће тачности за комплексне профиле индивидуалне потрошње. Компаративне студије на контролисаном скупу података су показале супериорност дубоких неуралних мрежа у односу на већину често коришћених метода, што је и показано у референцама [1.63], [1.64], [1.65], [1.66]. Међутим, проблем дуготрајног, итеративног одређивања великог броја хиперпараметара модела и рачунска захтеваност, као и проблем преприлагођавања, за сада, могу представљати ограничење за њихову ширу практичну примену. Такође, унапређења у тачности моделовања имају ограничења, док дуго време потребно за обучавање модела и отежано одржавање квалитета прогнозирања у пракси [1.67], неконзистентно време тренирања за различите потрошаче, величину података и различите факторе потрошње [1.68], значајно успорава њихову примену. Интересантан закључак за, генерално, неуралне мреже се наводи у опсежној студији (70) из 2020. године, где се наводи: “*ANNs were never applied in context of wide-scale day-ahead predictions on a statistically relevant sample of local loads where historic data is limited, manual adjustment is not possible and the loads can be highly volatile.*“

На основу прегледних радова и неколико конкретних примера, направљен је кратак преглед примењивости популарних техника моделовања за три карактеристична нивоа потрошње.

Конвенционалне методе, иако типично верификоване на потрошњама на нивоу система [1.70], [1.71], своју примену налазе и за прогнозирање потрошње индивидуалног потрошача [1.72] и зграда [1.73]. Њихова главна предност су једноставни и брзи алгоритми који могу да подрже динамичке режиме прогнозирања потрошње, тако да су преферабилан избор за, на пример, динамичке режиме одзива потрошње (*dynamic demand response*), као у раду [1.74]. Унапређења у моделовању нелинеарности временске серије потрошње и тачности прогнозе, нуде технике машинског учења, као на пример, *SVR*, примењен на нивоу система [1.75], у дистрибуцији [1.76], за зграде [1.77]. *RF* се показала као алгоритамски једноставан техника која се користи такође на нивоу целог система [1.78], али и за различите типова кластера зграда [1.79]. У пракси су најчешће неуралне мреже, које се користе на нивоу целог система [1.80], за тафо рејон [1.81], зграде [1.82], индивидуалног потрошача [1.83]. Дубоке неуралне мреже су посебно интересантне за примену на потрошњу индивидуалног или групе индивидуалних потрошача, као у већ поменутих радовима, али налазе примену и на нивоу целог система и зграде. У пракси су, међутим, највероватније методе где се комбинују два или више модела, било њиховом хибридизацијом, најчешће неуралних мрежа и временских серија [1.84], или ансамблима где се резултати прогноза више модела комбинују за постизање одговарајуће тачности прогнозе, [1.85].

Резултати из побројаних референци показује да је на располагању више метода које се могу применити за моделовање појединачних потрошњи на различитим местима у систему, односно, оних које задовољавају захтев 1, спецификације из 1.1. Њихова генерализација која би омогућила рад за више потрошњи (захтев 2, спецификације из 1.1) зависи од врсте модела и постиже се, типично, или кроз избор улазних променљивих (*feature selection, variable selection*) и/или одговарајућим избором параметара и хиперпараметара модела. На пример у већ наведеном раду [1.76], аутори су развили аутоматску процедуру за моделовање *SVR* модела, којом се подешавају параметри модела и улазне променљиве и, на тај начин, врши аутоматско прилагођавање конкретној потрошњи. У [1.79] је вршен избор улазних променљивих *RF* модела, тестирано на већем скупу на нивоу зграде. За неуралну мрежу из [1.83], је усвојена јединствена топологија (један сет хиперпараметара) за прогнозирање 37 временских серија потрошње, добијен итеартивним претраживањем. У [1.82] су, претходно испитиване статистичке особине временске серије потрошње како би се подесили хиперпараметри *CNN* модела. За зграде, у [1.86], за *CNN* модел је усвојен заједнички сет хиперпараметара (број слојева и чворова мреже), док су коефицијенти мреже калкулисани индивидуално за сваког потрошача. Како би се избегла претходна опсежна анализа у дефинисању структуре дубоких мрежа, и аутоматизовао процес избора оптималне структуре модела у [1.87], аутори су предложили *GA-LSTM* приступ где се користи генетски алгоритам (*GA*) за избор улазних променљивих и хиперпараметара *LSTM* модела, за прилагођавање.

Начини динамичког прилагођавања модела променама потрошње у времену (захтев 3) зависи од типа модела, односно, његовог капацитета и од тога који део модела је могуће прилагођавати у задатим временским интервалима тако да прати значајне промене. На пример, у раду [1.88] је кориштена метода базирана на једноставном *ARIMA* моделу са клизајућим прозором која омогућава поновно обучавање комплетног модела (реда *ARIMA* модела и коефицијената) у реалном времену, захваљујући брзим и једноставним алгоритмима. Међутим, код, на пример, неуралних мрежа, методе итеративног подешавања хиперпараметара било ручно или коришћењем спорих оптимизационим алгоритма (као што је *GA*) представљају проблем за поновно обучавање. Како би се превазишао овај проблем, у раду [1.89], динамички се врши избор улазних променљивих за *LSTM* модел, док је у [1.90] је предложена методологија којом је омогућено динамичко подешавање групе хиперпараметара дубоке неуралне мреже.

Што се тиче интерпретабилности (захтев 4), ту је избор на страни конвенционалнијих методе које су, по својој структури, отворене у односу на већину напредних техника, које су *black-box* модели.

Важна констатација је да је веома мали број радова посвећен истраживању и развоју *STLF* алата који је систематично и експлицитно евалуиран за више нивоа агрегације потрошње у мрежи. У раду [1.91] из 2023 је предложен оквир (базиран на дубоким неуралним мрежама) за различита сценарија прогнозирања, тестиран на више индивидуалних потрошача, њихових агрегација и трафо-рејона. У закључку аутори наглашавају потребу развоја методологије динамичког прилагођавања за развој *real-time* верзије њиховог модела или, алтернативно, препоручују периодично ремоделовање. У раду [1.92] из исте године је предложен пробабилистички приступ реализован фузијом модела (базиран на *GAMLSS*⁴ моделу) за процену времена и нивоа дневног максимума за наредни дан, тестиран за више нивоа

⁴ *GAMLSS-Generalized Additive Model for Location, Space and Shape*

агрегације у систему. У овом раду се, такође, наглашава важност адаптације модела на промене услова прогнозирања и, значајно, потребу да се истраже карактеристике временске серије потрошње на различитим нивоима агрегације у систему. У раду [1.93], такође из 2023., предлаже се јединствена софтверска платформа, *eForecaster*, базирана на неколицини техника машинског учења намењених прогнозирању потрошње електричне енергије, за различите прогностичке задатке у ЕЕ систему. У [1.94], из исте године, аутори су развили *STLF* методологију која је тестирана на нивоу целог система и за трафо рејоне. Дакле, тек неколико резултата који се баве генералним приступом *STLF* функционалности у ЕЕС-у, који су скоријег датума и са закључцима који указују да је потребно даље унапређење, представљају јасан показатељ да постоји недостатак овог типа истраживања.

Закључно, прегледана литература је показала да избор модела *само* на основу тачности моделовања није оптималан, већ су за практично одрживу *STLF* функционалност важне и имплементационе карактеристике, брзина алгоритама и извршавања, тип података, комплексност профила, конкретна примена. Са аспекта потребне флексибилности којом се обезбеђује тачност прогнозе и њено одржавање у простору, ЕЕС хијерархији и у времену, препознате су две тенденције. Са једне стране су у употреби мање комплексни модели, али са методологијама припреме и поделе података и брзим алгоритмима. На пример, у раду [1.95] је развијена аутоматска процедура која динамички бира између 7 стандардних модела и врши прогнозу потрошње за различитих места у систему, број потрошача, расположивост података и различите временске периоде. Иако, теоријски, овакав приступ обезбеђује висок степен флексибилности, оперативно, може постојати проблем са различитим брзинама естимационих процедура и њиховом неусаглашеношћу. С друге стране су комплексне структуре, типично, дубоке неуралне мреже, унапред дефинисане дуготрајним поступком, које се користе за развој глобалних модела. На пример, у раду [1.96] је описан пример глобалног модела са 2.5 милиона параметара, обучен за 1000 потрошњи на нивоу индивидуалног потрошача, нисконапонског извода и трафо рејона. Код оваквих модела проблем може представљати немогућност динамичког прилагођавања.

Као потенцијално оптималан избор, у тези се предлаже интерпретабилна техника машинског учења која припада оквиру генерализованог адитивног модела, ГАМ, чији су основни концепти и примена на задатак прогнозе потрошње електричне енергије дати у референцама [1.97]-[1.117]. На основу литературе, карактеристике ГАМ-а које га кандидују за развој нових *STLF* апликација је:

- *верификована висока тачност прогнозирања* добијена за конкретне потрошње, на нивоу целог система, региона, подстаннице, зграда, индустријског потрошача, и мање агрегације индивидуалних потрошача, чиме се потврђује његова адекватност за моделовање различитих профила потрошње унутар ЕЕС-а и потенцијал за развој генералне методологије. Такође, веома је индикативно, да популарност ГАМ модела за примену на *STLF* значајно расте са развојем генералних приступа. На пример, методологије предложене у наведеним публикацијама, [1.92], [1.93] и [1.94], користе ГАМ, у основној или проширеној верзији.
- *једноставан поступак спецификације модела*. Потрошња моделована ГАМ-ом представља суму индивидуалних подмодела фактора потрошње (сезоналности, календар, метеоролошки утицаји и остали спољашњи фактори, итд.), тј. чланова ГАМ-а. Хиперпараметри ових подмодела се једноставно дефинишу, неспецифични су и некритични (предиктивне перформансе нису претерано осетљиве на њихов избор).

- *брзи и једноставни естимациони алгоритми*, засновани на класичним регресионим техникама. Алгоритамска унапређења, дата у [1.103] и [1.107], омогућила су рад са великим подацима и значајно убрзали процес ГАМ естимације.
- *аутоматска процедура избора ГАМ чланова*. Постоји неколико метода за избор ГАМ чланова, уграђених у брзу естимациону процедуру [1.117], чиме се пружа могућност развоја динамичког прилагођавања конкретној потрошњи, које за потребе прогнозе које се имају, могу бити са сваким новим податком.
- *интерпретабилност* као кључна особина ГАМ модела која омогућава издвајање појединачних ефеката потрошње и њихову анализу, (захтев 4).
- *могућност проширења* и даље генерализације за различите расподеле променљиве која се моделује ГАМЛСС, на хијерархијску верзију, ХГАМ, дубоку ГАМ верзију, итд. Константан развој ГАМ оквира је предуслов даљег унапређења и ГАМ-базиране *STLF* функционалности.
- *open source окружење* за развој ГАМ модела које омогућава бржи развој и имплементацију ГАМ методологија.

На основу побројаних карактеристика, **претпоставка** је да је ГАМ математички и алгоритамски подржан да може да задовољи спецификацију 1-4 из 1.2 и да је погодан за развој нових *STLF* алата.

1.3 Методе истраживања

Методе истраживања које су се користиле при изради тезе су:

- Направљен је опсежан преглед литературе из области ради анализе постојећих модела и методологија, сагледавања њихових предности и недостатака, на основу чега је извршен избор одговарајућег модела.
- Методе прикупљања, препроцесирање и припреме података за тестирање и формирање локалне базе података. Користе се отворени извори података за потрошњу и остале варијабле, за различитих места у систему, водећи рачуна да подаци садрже репрезентативну базу за евалуацију *STLF* функционалности.
- Методе анализа временске серије потрошње, као и досад установљена знања о потрошњи како би се идентификовали основни фактори потрошње и изучила њихова релевантност за различите прогностичке области и припремиле за моделовање.
- Развој и софтверска имплементација усвојеног приступа (*R* језик), њено тестирање коришћењем података из локалне базе и евалуација стандардним критеријумима за прогностичке моделе.

1.4 Структура докторског рада

Теза садржи шест основних поглавља, списак литературе, податке о ментору и члановима Комисије, сажетак, биографију и библиографију аутора. У првом поглављу је описана проблематика која се обрађује у тези, дат преглед литературе и образложење. У другом поглављу је дат списак отворених база података и начин формирања локалних база, које се користе за тестирање модела. Треће поглавље описује основе генерализованог адитивног модела, са аспекта који се користи у тези. Четврта област даје преглед карактеристика временских серија потрошње унутар ЕЕС-а, и начин имплементирања ГАМ моделом. У петој области су приказани добијени резултати за све податке из локалне базе. У шестом поглављу је дат закључак, преглед научних доприноса, практична примена, главни недостаци добијених резултата из тезе и правци даљег истраживања.

1.5 Објављени научни радови

Објављени научни радови из области којом се бави теза су:

- I. **S. Krstonijević**, “Adaptive Load Forecasting Methodology Based on Generalized Additive Model with Automatic Variable Selection,” in *Sensors*, vol. 22, no. 19: 7247, doi: 10.3390/s22197247. ISSN 1424-8220. IF = 3.847. (M21)
Број прегледа: **1789** (приступано 20.11.2024.)
Број цитата: **3**
- II. **S. Krstonijević**, “Generalized Additive Model for Electricity Load Prediction in R,” in *Primena slobodnog softvera i otvorenog hardvera (Application of Free Software and Open Hardware, PSSOH)*, Belgrade, Serbia, 24 October 2020 (Session: Free Software and Open Hardware in Electrical Engineering and Computer Science, Part 2), doi: 10.5281/zenodo.4748377. (M61)
Број прегледа: **3244**
Број цитата: **1**
- III. **S. Krstonijević**, “Predlog prediktivnog modela potrošnje u pametnim elektroenergetskim mrežama”, 15. Savjetovanje BH CIGRE 2021, Zbornik radova R.D2.06, Neum, BH, 17-20 Oct., 2021. (M33)

2. Слободан софтвер и отворени подаци

Предложени оквир за прогнозу потрошње електричне енергије је развијан коришћењем слободног софтвера и евалуиран на отвореним подацима.

Отворени ресурси доприносе репродукцибилности, лакшој евалуацији резултата, омогућавају сваки вид компарације, коначно, будући доступни широј истраживачкој заједници, бржем развоју области.

2.1 Слободан софтвер коришћени у раду

Окружење на коме је развијана методологија је *RStudio*, [2.1] и коришћени су специјализовани *R* пакети: *mgcv*, *forecast*, *R.utils*, *ggplot2*, *simstudy*, *reshape*, *feather*, *arrow*, *TTR*.

2.2 Отворени подаци коришћени у раду

У наставку су побројане и описане отворене базе података које су коришћене у раду. Сваком сету података је додељен идентификациона ознака његове локалне базе која се користи, ради лакшег кретања кроз рад. Такође, описан је начин филтрирања података и њиховог иницијалног препроцесирања и генерисања локалних база података, припремљених за директну примену.

2.2.1 *NYISO* база података (NY)

New York Independent System Operator (NYISO) је независни оператор који контролише територију шире области *New York*, подељену на 11 зона. *NYISO* зоне се разликују по величини површине коју обухватају и броју потрошача, од веома насељене која обухвата ширу област *New York* а, географски велике, али слабо насељене *Mohawk Valley*, руралне области са веома малим бројем становника, *North*, до, површином мале, зоне *Millwood*. Подаци за потрошњу и метеоролошке вредности за период од 01.01.2009 до 31.12.2013 су преузети са интернет локације [2.2]. *NYISO* зоне, њихов званични назив, просечна потрошња за посматрани период године и број становника дати су табелом Табела 2.1.

Подаци садрже вредности за потрошњу и максималну и минималну дневну вредност температуре са сваку зону. Предност ове базе података је што поред реалних вредности температуре садржи и историјске вредности прогноза температуре, што омогућава тестирање за реалан сценарио прогнозирања. Такође, одабрани временски период садржи десетогодишњи рекордни максимум потрошње од ~ 34 GW (19.07.2013 у 17 сати).

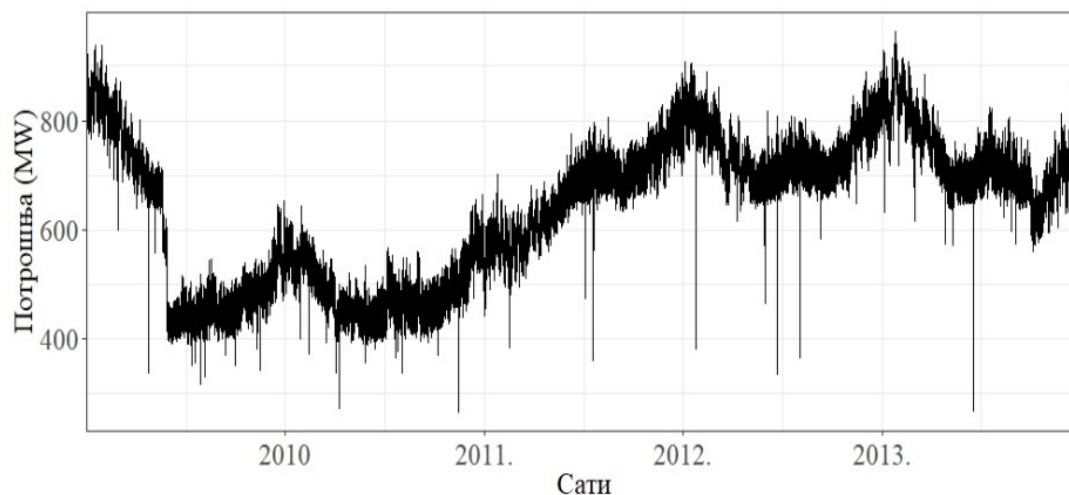
С обзиром да су *NYISO* подаци реална мерења, у бази је постојао извештај број недостајућих података и оутлејера⁵. Недостајућа мерења која су, махом, на нивоу једног податка су прорачуната усредњавањем суседних вредности. Осим визуелно уочљивих оутлејера код неких зона, који су третирано као недостајућа мерења, даља препроцесирања нису спроведена. Уочено је да NORTH зона има неуобичајене профиле потрошње у односу на остале зоне (приказано сликом 2.1), и, како се наводи у [2.3], ради се о индустријској зони.

⁵ Оутлејер – екстремне, нестандартне и ван опсега вредности

Слободан софтвер и отворени подаци

Табела 2.1: Листа и карактеристике NYISO зона

Зона	ID	Просечна потрошња (MW)	Број становника (x 10 ³)
<i>West</i>	WEST	1790	1532
<i>Genese</i>	GENESE	1140	1003
<i>Central</i>	CENTRL	1850	1384
<i>Capital</i>	CAPITL	1330	1215
<i>Millwood</i>	MILLWD	330	190
<i>Dunwoodie</i>	DUNWD	670	760
<i>New York City</i>	NYC	6120	8186
<i>Long Island</i>	LONGIL	2540	2835
<i>Mohawk Valley</i>	MHKVL	910	891
<i>Hudson Valley</i>	HUDVL	1150	1372
<i>North</i>	NORTH	540	82



Слика 2.1: Сатна потрошња за NORTH зону, за период 2009-2013 године

За сваку зону су припремљене сатне вредности потрошње, дневна максимална и минимална реална вредност температура, као и њихове прогнозе. Са интернета су прикупљени датуми значајнијих празника, Нова Година (31.12. и 1.1.), Ускрс (Велики петак и Ускршња недеља), као и за Меморијални дан (*Memorial Day*), Дан независности (*Independence Day*) и Колумбов дан (*Columbus Day*), за 2009-2013. годину. Такође, подаци за *day-light saving*, односно датуми преласка са зимског на летње рачунање времена (и назад) су прикупљени са интернета.

2.2.2 Ausgrid база података (AU)

Ausgrid је највећи дистрибутер електричне енергије на источној обали Аустралије, који електричном енергијом снабдева око 1.8 милиона потрошача са територије од преко 20 хиљада километара квадратних, обухватајући ширу област Сиднеја. Он управља мрежом која

Слободан софтвер и отворени подаци

укључује преко 180 TS 33/11kV, 66/11kV, 132/11kV трансформаторских станица. Састав потрошача које обухватају трафо-рејони укључује резиденцијалне и руралне потрошаче, мала предузећа и велике мултинационалне корпорације, тако да база података обухвата различите профиле потрошње.

Једном годишње *Ausgrid* на свом сајту [2.4] објављује историјске податке за потрошњу, за протеклих 12 месеци. Историјски подаци садрже 15-минутна мерења потрошње прикупљена са *Ausgrid* SCADA система, за већину трансформаторских станица које се надгледају. Вредности представљају ‘сирова’ мерења, која нису третирана нити анализирана од стране оператера.

Прегледом *Ausgrid* базе података констатовано је да постоји велики број недостајућих мерења, краткотрајних и дуготрајних искакања изван опсега мерене потрошње. С обзиром да је на располагању велики број података, а како би се избегло њихово опсежно препроцесирање и предобрада, за потребе тестирања у раду изабрано је 12 трансформаторских станица, код којих су подаци били коректни, за 2021 и 2022. годину. Листа одабраних трафо објеката и просечна потрошња електричне енергије трафо рејона које покривају, за одабрани период, дати су у Табели 2.2.

Табела 2.2: Листа и просечна потрошње *Ausgrid* зона

Локални ID	Назив трафо станице	Просечна потрошња (MW)
AU1	<i>RNS Hospital</i>	7
AU2	<i>Auburn</i>	11
AU3	<i>Olympic Park</i>	15
AU4	<i>Chatswood</i>	19
AU5	<i>Leichhardt</i>	20
AU6	<i>Matrville</i>	21
AU7	<i>Cronulla</i>	21
AU8	<i>Clovelly</i>	21
AU9	<i>Maroubra</i>	25
AU10	<i>North Sydney</i>	27
AU11	<i>Burwood</i>	33
AU12	<i>City North</i>	45

Такође, са интернет локације [2.5] преузете су средње сатне вредности неколико метеоролошких варијабли за Сиднеј, за исти временски период.

2.2.3 *EnerNOC* база података (EN)

EnerNOC база података садржи вредности потрошње за око 100 комерцијалних, индустријских зграда и школа, за 2012. годину, доступна на сајту [2.6]. За тестирање су одабране 2 зграде, школа и комерцијални објекат, у Њујорку, чије су просечне потрошње за целу годину, као и летњи и зимски максимуми дати су у Табели 2.3.

Са интернет локације [2.5] прикупљене су вредности средње сатне температуре, смештене у локалну базу података, заједно са сатним вредностима за потрошњу.

Табела 2.3: *EnerNOC* зграде, њихова просечна потрошња, летњи и зимски максимум

ID у бази података	Локални ID	Просечна потрошња, зимски/летњи максимум	Опис
137	EN1	110kW 180kW (зимски макс.) 380kW (летњи макс.)	Основна/средња школа површине око 7000 m ²
341	EN2	400kW 450kW (зимски макс.) 650kW (летњи макс.)	Продајни објекат површине око 5000 m ²

2.2.4 *ASHRAE*⁶ база података (AS)

ASHRAE база података садржи сатне вредности за потрошњу електричне енергије за око 1000 зграда. Подаци су преузети са *Kaggle* платформе [2.7], који су постављени за потребе *Great Energy Predictor III (GEP III)* такмичења из 2020. Поред вредности за потрошњу, у бази постоје и сатне вредности за температуру и неколико секундарних метео варијабли, за 2016. годину. Из базе која садржи податке за различите типова зграда, издвојени су подаци за 3 резиденцијане зграде, чије су карактеристике дате у Табели 2.4.

Табела 2.4: *ASHRAE* резиденцијалне зграде, њихова просечна потрошња и опис

ID у бази података	Локални ID	Просечна потрошња, зимски/летњи максимум	Опис
167	AS1	80 kW	8000 m ² из 1956. године
202	AS2	200 kW	7000 m ² из 2006. године
133	AS3	56 kW	6000 m ² из 1960. године

2.2.5 *London Carbon Project* база података (LCL)

LCL база садржи мерења прикупљена са бројила за домаћинстава у оквиру *Low Carbon London* пројекта, спроведеног од стране *UK Power Networks* [2.8]. Подаци за потрошњу електричне енергије за око 5000 индивидуалних потрошача са ширег подручја Лондона,

⁶ *American Society of Heating, Refrigerating and Air-Conditioning Engineers*

Слободан софтвер и отворени подаци

подељених у 6 гео-демографских категорија, покривају период од новембра 2011 до фебруара 2014. База садржи велики број недостајућих мерења и за потребе тестирања су коришћени подаци за потрошњу 114 потрошача из једне категорије (просечна годишња потрошња до 8.44 kWh), за 2013. годину, за које су подаци били коректни.

3. Генерализовани адитивни модел, ГАМ

Генерализовани адитивни модел⁷, ГАМ, је проширење линеарног модела који, као чланове, укључује глатке нелинеарне функције улазних променљивих. Предложен је у раду [3.1] и даља теоријска и алгоритамска унапређења су покривена опсежном литературом, од чега је неколико значајнијих публикација које су послужиле за објашњење наведено у референцама [3.2-3.15]. За развој ГАМ базираних апликација и имплементацију у R програмском језику, као и остале аспекте примене ГАМ модела, од посебног је значаја књига *Simon Wood*-а [3.15], као кључна референца која се користи у тези. Алгоритамска подршка за развој ГАМ модела и пропратне процедуре су имплементирани у *mgcv* R пакету (истог аутора), који је главно развојно окружење предложене методологије [3.16].

ГАМ обухвата различите структуре модела, дефинисане збиром линеарних и нелинеарних чланова, једне или више променљивих, као и њихових међусобних интеракција. Избор улазних променљивих, као и претпоставке о линеарности, нелинеарности или заједничком ефекту које имају на излазну променљиву, а које се уносе при спецификацији модела, омогућавају конструкцију различитих типова чланова ГАМ-а. Генерална структура ГАМ-а која се користи у докторској дисертацији дефинисана је скупом чланова као у једначини 3.1 и сви примери дати на даље су њене инстанце и служи као основа за развој методологије на коме се заснива предложени оквир.

По формули 3.1, очекивана вредност излазне променљиве $\mu = E(y)$, која припада експоненцијалној фамилији расподела, моделована је збиром неколико типова ГАМ чланова, стандардизованих по свом функционалном облику, до остатка $e \sim \mathcal{N}(0, \sigma)$. g је функција везе, односно, *link* функција. Типови чланова, као и функције и променљиве које им припадају су у једначини индексирани са i, j, k и l и њихов опис дат је у Табели 3.1.

$$g(\mu) = \sum_i \beta_{0i} x_i + \sum_j f_j(x_j) + \sum_k x_{1,k} f(x_{2,k}) + \sum_l f_l(x_{1,l}, x_{2,l}) + e \quad 3.1$$

Основне претпоставке ГАМ модела су:

- Функција везе g је специфицирана тако да обезбеђује линеарну зависност ефекта улазне променљиве и трансформације средње вредности излазне променљиве. Најчешће су употреби, идентичка (μ), логаритамска ($\ln(\mu)$), квадратни корен ($\sqrt{\mu}$), *логистичка* функција ($\ln \frac{\mu}{1-\mu}$).
- Адитивност модела која подразумева да је променљива која се моделује резултат сумирања идентификованих, појединачних ефеката улазних променљивих која омогућава њихово независно моделовање. Претпоставка адитивности значајно поједностављује структуру модела и олакшава интерпретабилност модела.
- Глаткост нелинеарних ГАМ функција којом се претпоставља да су нелинеарне зависности излаза модела од улаза глатке, непрекидне нелинеарне функције.
- Претпоставка да су улазни подаци модела међусобно независни, наслеђена из линеарних модела.
- Расподела излазне променљиве је из фамилије експоненцијалних функција као што су нормална расподела, *Gamma*, биномна, мултиномна, *Poisson*-ова, итд.

⁷ Користи се и превод Уопштени адитивни модел

Генерализовани адитивни модел

Табела 3. 1: Типови ГАМ чланова који се користе у раду

	Тип ГАМ члана	Ефекат
1	x_i	Параметарски члан карактеристичан за линеарне моделе којим се моделује линеарна ефекат улазне на излазну променљиву. По природи може бити: <ul style="list-style-type: none"> • континуална променљива или њена трансформација • категоричка варијабла за моделовање више група (категорија) • средња вредност излазне променљиве за нулти остатак модела (<i>intercept</i>)
2	$f_j(x_j)$	Глатка функција којом се моделује произвољни нелинеарни ефекат улазне променљиве x_k на излазну променљиву.
3	$f_k(x_{1,k})x_{2,k}$	Заједнички ефекат или интеракција параметарске променљиве и нелинеарности <ul style="list-style-type: none"> • за случај кад је $x_{1,k}$ категоричка варијабла, моделује се посебна нелинеарност за сваку од категорија. • за случају кад је $x_{1,k}$ бинарно, могуће је моделовати присуство/одсуство нелинеарности
4	$f_l(x_{1,l}, x_{2,l})$	Заједнички, здружени ефекат или глатка интеракција између две променљиве.

Нелинеарности у ГАМ-у представљају непараметарске чланове, којима се при спецификацији модела не уносе претпоставке о облику функције коју моделују и чији се облик прилагођава подацима у поступку естимације модела. То омогућава квалитативне формулације при дефинисању ових ГАМ чланова (линеарност, нелинеарност, интеракција) и флексибилност која је погодна за примене код моделовања вођеног подацима. Естимација ГАМ-а је базирана на регресионим методама, и у ту сврху су ове нелинеарности апроксимирани линеарним моделима коришћењем базисних функција сплајна.

Нелинеарна функција једне променљиве (ГАМ члан типа 2. Табела 3.1), је представљена линеарном комбинацијом функција одређеног типа сплајна, b_k , и одговарајућих коефицијената, β_k по:

$$f(x) = \sum_{k=1}^K b_k(x)\beta_k \quad 3.2$$

Базисне функције дефинисане за вредности x_n , $n = 1, \dots, N$, које се користе за естимацију ГАМ-а, формирају $N \times K$ матрицу модела, \mathbf{B}_x ,

$$\mathbf{V}_x = \begin{bmatrix} 1 & b_1(x_1) & \cdots & b_K(x_1) \\ \vdots & \vdots & \ddots & \vdots \\ 1 & b_1(x_N) & \cdots & b_K(x_N) \end{bmatrix} \quad 3.3$$

тако да $f(x)$ трансформација из 3.2 представљена производом $\mathbf{V}_x \boldsymbol{\beta}_x$, где је $\boldsymbol{\beta}_x$ вектор коефицијената које је потребно естимирати.

ГАМ члан типа 3 из Табеле 3.1 који показује интеракцију између нелинеарности и параметарске променљиве, $f(x_1)x_2$ је тзв, модел са променљивим коефицијентима. За вредности $x_{1,n}, x_{2,n}, n = 1, \dots, N$, које се користе за естимацију модела, у генералном случају, матрица модела овог ГАМ члана може бити представљена као

$$\text{diag}(\mathbf{x}_2) \mathbf{V}_{x_1} \quad 3.4$$

где је \mathbf{V}_{x_1} матрица модела функције $f(x_1)$ и $\text{diag}(\mathbf{x}_2)$ је дијагонална матрица са $x_{2,n}$ вредностима на n -том месту на главној дијагонали.

Карактеристичан пример који је значајан у овом раду, је онај када је x_2 категоричка променљива. Типична примена је кад се у модел желе укључити промене $f(x_1)$ нелинеарности за различите категорије података. Категоричка променљива је дефинисана коришћењем тзв. *dummy* променљивих, за сваку категорију, посебно. На примеру где су подаци подељени у 2 категорије, дефинишу се две променљиве, x_2^1 и x_2^2 , којима је кодира припадност $x_{1,n}$ податка одређеној категорији, првој, за $x_{2,n}^1 = 1$ и $x_{2,n}^2 = 0$, односно, другој за $x_{2,n}^1 = 0$ и $x_{2,n}^2 = 1$. На тај начин се, формално, $f(x_1)x_2$ члан се, може разложити на две подкомпоненте по формули 3.5:

$$f(x_1)x_2^1 + f(x_1)x_2^2 \quad 3.5$$

На основу 3.3 и 3.4, матрица модела \mathbf{V}_{x_1} се процењују одговарајући коефицијенти за сваку категорију, посебно. У специјалном случају када је x_2 бинарна променљива, могуће је контролисати присуство нелинеарности $f(x_1)$ у моделу за податке за које је $x_{2,n} = 1$, односно њено одсуство, када је $x_2 = 0$.

3.1 Избор типа базисних функција

Избор типа сплајна, спецификација базисних функција и њихов број K који се користе за апроксимацију ГАМ нелинеарности, бирају се зависно од конкретне примене. За развој генералних приступа и раду са већим бројем података које је потребно моделовати, за модел са потенцијално већим бројем чланова, од интереса је да се спецификацијом уноси што мање улазних претпоставки. Како би се оправдао избор базисних функција који се користе у моделима у тези, у наставку је дато неколико основних објашњења.

Сплајн f је секвенца полинома M -тог реда над интервалима x променљиве, дефинисане тзв., чворовима сплајна (*knots*), $x_1 < \dots < x_{k_n}$ који задовољава услов континуалности за цео домен улазне променљиве. Најчешће је у употреби кубни сплајн ($M = 3$), за који важе следећи услови:

- покрива интервале између чворова $(-\infty, x_1], [x_1, x_2], \dots, [x_{k_n}, \infty)$ полиномима трећег степена,
- први и други извод функције f_x и f_{xx} су континуални на сваком чвору x_1, \dots, x_{k_n} ,

За дате локације чворова, односно вредности x_1, \dots, x_{k_n} , функција може, на еквивалентан начин, бити представљена базисним функцијама различитих типова. Параметризација базисних функција за различите типове сплајна је различита, тако да избор зависи од процене флексибилности апроксимације која је потребна за конкретан проблем и броја параметара које је потребно одредити.

Најједноставнији је пример полиномског кубног сплајна за који су базисне функције дефинисане са 3.6.

$$b_1(x) = 1, \quad b_2(x) = x, \quad b_3(x) = x^2, \quad b_4(x) = x^3$$

$$b_{i+4}(x) = (x - x_i)^3 \quad i = 1, \dots, k_n \quad 3.6$$

Као нумерички стабилнији, са стриктно локалним базисима, често се користи и В-сплајн, дефинисан са 3.7.

$$b_i^m(x) = \frac{x - x_i}{x_{i+m+1} - x_i} b_i^{m-1}(x) + \frac{x_{i+m+2} - x}{x_{i+m+2} - x_{i+1}} b_{i+1}^{m-1}(x), \quad i = 1, \dots, K$$

$$b_i^{-1}(x) = \begin{cases} 1, & x_i \leq x < x_{i+1} \\ 0, & \text{иначе} \end{cases} \quad 3.7$$

Даља унапређења којим се решава проблем веће варијабилности на рубовима интервала сплајна, је увођењем природног (*natural*) базисних функција. Од интереса у тези је и циклични кубни сплајн, којим се моделује глатка периодична нелинеарност, за коју је услов континуалности задовољен на спојевима између периода (што ће бити приказано касније). Детаљније објашњења карактеристика и начина конструкције различитих типова сплајнова наведен је у [3.15].

За одабрани тип сплајна, избор локације и броја чворова зависи од поступка естимације модела. За једноставан случај моделовања у променљиве, гаусовске расподеле, глатком функцијом једне променљиве $f(x)$, коришћењем стандардне регресије решава се проблем 3.8, где је \mathbf{y} вектор y_n вредности за $n = 1, \dots, N$, $\mathbf{B} \in \mathbb{R}^{N \times K}$ матрица модела добијена по 3.2 и 3.3 и $\boldsymbol{\beta} = (\beta_1, \dots, \beta_K)$ вектор коефицијената који се процењује.

$$\min_{\boldsymbol{\beta}} \sum_{n=1}^N \left(y_n - \sum_{k=1}^K \beta_k b_k(x_n) \right)^2 \Leftrightarrow \min_{\boldsymbol{\beta}} \|\mathbf{y} - \mathbf{B}\boldsymbol{\beta}\|_2^2 \quad 3.8$$

У овом случају, за одабрани тип базиса, апроксимација је потпуно одређена избором чворова, односно њиховом локацијом и њиховим бројем. С обзиром да оптимално решење веома зависи од облика нелинеарности која се моделује, то значи да би за сваку нелинеарност и сваки сет података било потребно њихово дефинисање, што је неприхватљиво решење за задатак који се решава у тези.

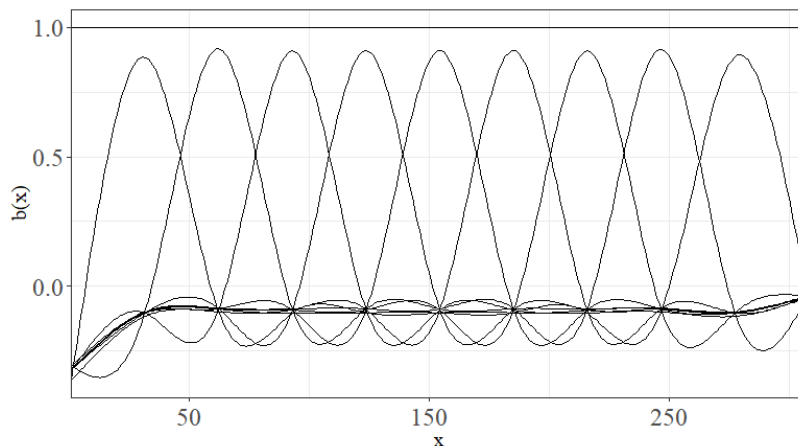
Горњи проблем је мање критичан код пенализованих сплајнова. У том случају се при процени коефицијената контролише брзина промене функције, мерене њеним другим изводом, f_{xx} , тако да се добије глатка апроксимација. Тада се за претходни пример, коефицијенти добијају решавањем проблема 3.9.

$$\min_f \sum_{n=1}^N \left(y_n - \sum_{k=1}^K \beta_k b_k(x_n) \right)^2 + \lambda \int f_{xx}^2 dx \Leftrightarrow \min \| \mathbf{y} - \mathbf{B}\boldsymbol{\beta} \|^2 + \lambda \boldsymbol{\beta}^T \mathbf{S}\boldsymbol{\beta} \quad 3.9$$

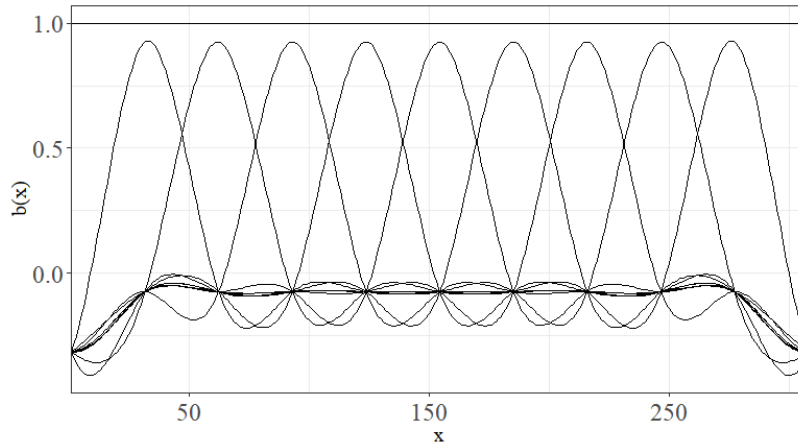
У горњој једначини, \mathbf{S} је фиксна матрица пенала која комбинује коефицијенте модела тако да важи $\int f_{xx}^2 dx = \boldsymbol{\beta}^T \mathbf{S}\boldsymbol{\beta}$, карактеристична за одређени тип сплајна. Величина пенализације у 3.9 је контролисана параметром изглађивања (*smoothing parameter*) λ , са вредностима од $\lambda = 0$ за непенализовани сплајн, до максималне пенализације $\lambda \rightarrow \infty$, када је резултат процене функције права линија, односно, добија се линеаран ГАМ члан. Моделована нелинеарност је контролисана тако да за функције које се брзо мењају на кратком интервалу простора улазне променљиве, вредности параметара изглађивања ће бити веће. Као таква, веома зависи од природе функције и нивоа њене варијабилности. У *mgcv* пакету, параметар изглађивања се добија аутоматски на основу података који се моделују. За пенализоване сплајнове, чворови се, типично, равномерно распоређују по опсегу вредности аргумента x (*knots-based* приступ). Алтернативно, може се користити *eigen-based* приступ, код кога се чворови дефинишу за сваку вредност аргумента (*full spline*), одговарајући пенал $\boldsymbol{\beta}^T \mathbf{S}\boldsymbol{\beta}$ се репараметризује тако да се, из простора описаних базисима, омогући избор оних K базисних функција са најмањом брзином промена, [3.8]. Дакле, за постојеће стратегије параметризације пенализованог сплајна, могућа је реализација ГАМ члана која се, за дати тип сплајна, заснива на одабиру *само* броја базисних функција, K .

Имплементација кубног сплајна у *mgcv* пакету користи базисне функције тзв. природног, кардиналног сплајна [3.15]. На сликама 3.1 и 3.2 је приказано 10 базисних функција кардиналног сплајна, генерисаних у *mgcv* пакету, за кубни сплајн и цикличну верзију, респективно.

Унапређење у правцу реализације *data-driven* ГАМ члана модела је коришћење *thin plate regression spline*, *TPRS*. Он је скраћена верзију *thin plate spline* *TPS*, који представља оптимално решење проблема налажења глатке функције више променљивих пенализованим сплајном. По основној дефиницији, *TPS* је применљив за функције произвољног броја променљивих и за произвољан ред диференцирања у пенализацији. На примеру функције $f(x)$ једне променљиве, пенализоване другим изводом (који је интересантан у раду), базисима *TPS* сплајна добијеним минимизацијом проблема 3.10. се даје одговор на питање оптималне глатке апроксимације функције [3.15, страна 150-152].



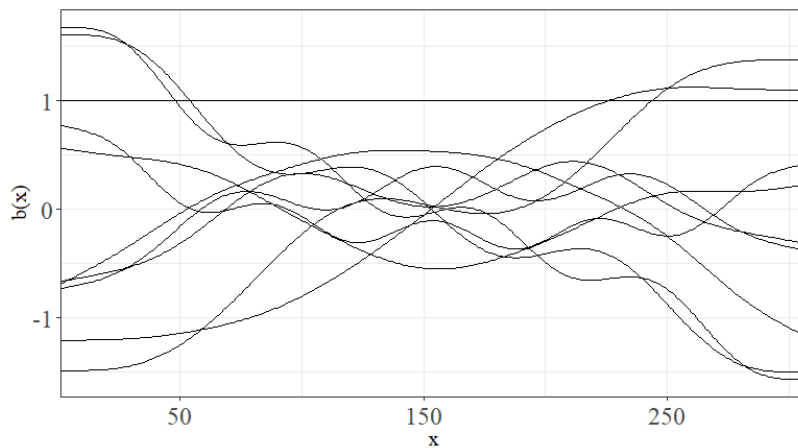
Слика 3.1: Пример базисних функција кардиналног кубног сплајна



Слика 3.2: Пример базисних функција цикличног кубног сплајна

$$\|y - f\|^2 + \lambda \int f_{xx}^2 dx \quad 3.10$$

TPS сплајн је дефинисан као *full spline* и његова директна практична примена је рачунски лимитирана. Зато се користи скраћена верзија, коришћењем (*eigen-based* приступ) и начини конструкције *TPRS* сплајна се могу наћи у [3.15, страна 153]. Оно што је важно закључити из приказаног поступка је да за спецификацију *TPS* сплајна не захтева унапред дефинисање нити одређивање локације чворова, нити се унапред врши избор типа базисних функција, јер оба проистичу као решење постављеног математичког проблема минимизације 3.10, за дате податке. Сликаом 3.3 приказано је 10 базисних функција *TPRS* сплајна.

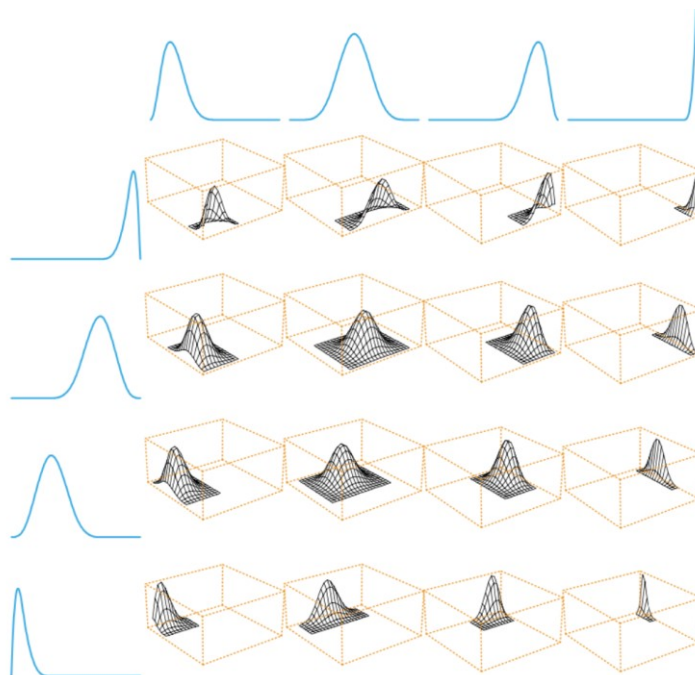


Слика 3.3: Пример базисних функција *TPRS* сплајна

Моделовање функције више променљивих *TPRS* сплајном има недостатак што се брзине промена за сваку од променљивих контролишу истим параметром изглађивања λ (својство изотропности), што је прихватљиво за променљиве које су на истој скали, као, на пример, моделовање у простору, где су улазне променљиве географска ширина/дужина. За функцију две променљиве које су на различитим опсезима вредности и, генерално, различите природе, користи се модел заснован на тензорима, који омогућава различиту пенализацију. Модел функције две променљиве, x_1 и x_2 , (ГАМ члан типа 4, Табела 3.1), које су, индивидуално, представљене базисним функцијама, b_k и b_j , респективно, дат је једначином:

$$f(x_1, x_2) = \sum_{k=1}^K \sum_{j=1}^J \beta_{k,j} b_k(x_1) b_j(x_2) \quad 3.11$$

Горња једначина је изведена из тензорског производа матрица модела појединачних функција $f_{x_1}(x_1)$ и $f_{x_2}(x_2)$, \mathbf{V}_{x_1} и \mathbf{V}_{x_2} и детаљна објашњења овог поступка се налазе у референци [3.15, одељак 4.1.8]. (На слици 3.4, преузетој из референце [3.17] приказан је тензорски производ кубних сплајнова две функције). Оно што је значајно је да је тензорски производ могућ за различите типове сплајнова и различит број базиса.



Слика 3.4: Тензорски производ базисних функција цикличног кубног сплајна

Сумом из једначине 3.6 моделована је укупна промена две променљиве, која садржи индивидуалне ефекте $f_{x_1}(x_1)$ и $f_{x_2}(x_2)$ и компоненту промена *само* услед њихове интеракције, $f_{x_1, x_2}(x_1, x_2)$. То даје могућност да се функција $f(x_1, x_2)$ може представити трима члановима ГАМ-а којима се засебно моделују појединачни и заједничка компонента по формули:

$$f(x_1, x_2) = f_{x_1}(x_1) + f_{x_2}(x_2) + f_{x_1, x_2}(x_1, x_2) \quad 3.12$$

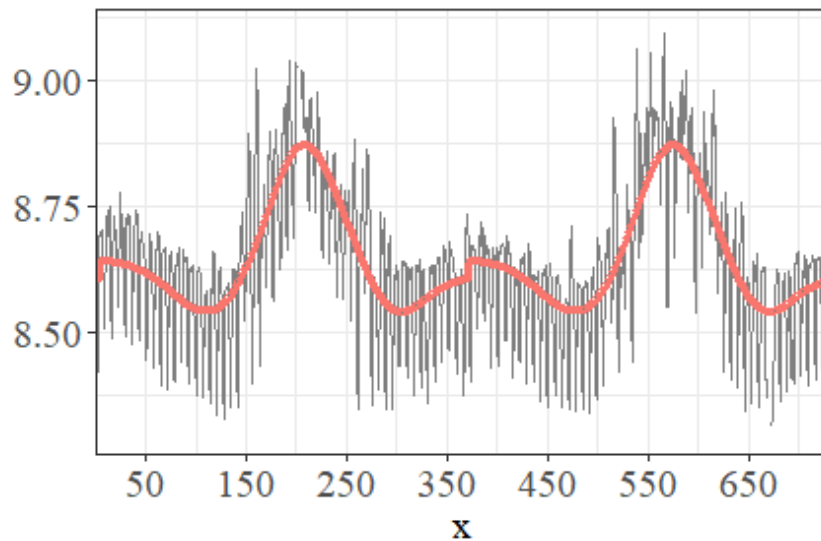
Неколико је разлога због којих је декомпонована репрезентација ГАМ функције две променљиве преферабилан избор за практичну примену. Пре свега, на тај начин се подкомпоненте из 3.12 могу посматрати и анализирати одвојено, као, на пример, приликом бирања модела, где се независно разматра утицај сваке од њих на излазну променљиву. Такође, у моделима где постоје две, или више, функција две променљиве које имају исту улазну променљиву, засебним моделовањем индивидуалних компоненти и интеркација се избегава преклапање ефеката. Поступком декомпоновања се из тензорског производа издваја скуп базиса матрице \mathbf{V}_{x_1, x_2} , којим је обухваћена компонента $f_{x_1, x_2}(x_1, x_2)$, а индивидуалне нелинеарности одвојено укључују у модел. Поступак је описан у [3.15, страна 203], и имплементиран у *mgcv* пакету.

За функцију две променљиве, $f(x_1, x_2)$, реализовану тензорским производом, брзине промена услед x_1 и x_2 су пенализоване различитим параметрима изглађивања, λ_1 и λ_2 , [3.15, страна 161]. За изведену декомпоновану верзију 3.12, поред засебних параметара изглађивања за индивидуалне функције, постоје још два пенала интеракције, $f_{x_1, x_2}(x_1, x_2)$, дат са 3.13.

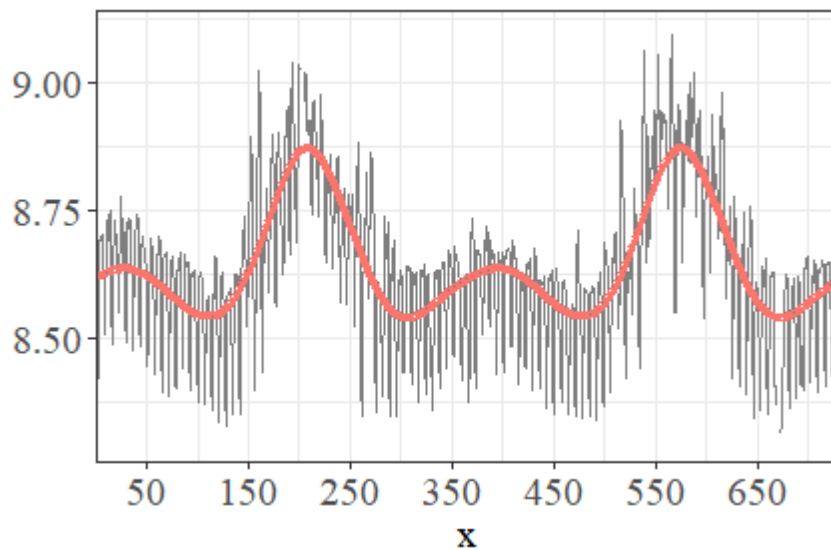
$$\lambda_3 \beta_3^T \tilde{\mathbf{S}}_3 \beta_3 + \lambda_4 \beta_4^T \tilde{\mathbf{S}}_4 \beta_4 \quad 3.13$$

Математички детаљи, услови декомпоновања, идентификабилност и остала потребна објашњења везана за декомпоновану верзију, могу се се наћи у референцама [3.15] и [3.18]. Са аспекта имплементације, од користи може бити чињеница да већи број пенала са различитим параметрима изглађивања, омогућава већи степен флексибилности при моделовању функције две променљиве.

У тези су коришћена ГАМ модели са два типа сплајн базисних функција. Први је циклични кардинални сплајн којим се моделују периодичности код временских серија. Већина реалних временских серија којима се описују људске активности (типично, у економији, енергетици, итд.), садрже годишњу, седмичну, дневну периодичну компоненту (сезоналност). Стандардно, годишња периодична нелинеарност се моделује за дане у години, $x=1:365$ (366), са 12 равномерно распоређених базисних функција, док су седмичне периодичности, дефинисане за сваки дан у седмици, $x=1:7$, моделоване са 7 базисних функција. На слици 3.5 је приказан пример временске серије потрошње електричне енергије у GW (подаци за NYC зону, NY подаци). Приказане су вредности за две узастопне године (сиво) са резултатима моделовања годишње сезоналности (црвено) стандардним кубним сплајном, са 12 равномерно распоређених чворова. Оно што се уочава да на прелазу из једне у другу годину (датици 31.12. и 01.01.) постоји дисконтинуитет, који је неповољан ефекат за многе примене модела. Разлог је што стандардни кубни сплајн сезоналност моделује као независну нелинеарност која се понавља, не водећи рачуна о континуитету рубних вредности. Циклична верзија кубног сплајна је конструисана тако да обезбеђује да вредности моделоване функције на почетку и на крају периоде задовољавају услов глаткости (иста вредност и 1 извод) и обезбеђује бољу апроксимацију рубова криве. На слици 3.6 је приказан исти пример моделован цикличним кубном сплајном, са видљивим побољшањем.



Слика 3.5: Моделовање годишње сезоналности стандардним кубним сплајном (NYC зона)



Слика 3.6: Моделовање годишње сезоналности цикличним кубним сплајном (NYC зона)

За све нелинеарности, за које не постоје чврсте претпоставке о облику или понашању криве, користи се *TPRS* сплајн. Избор броја *TPRS* базисних функција је, махом, арбитран и зависи од конкретних података и примене модела. У поступку конструкције *TPRS*-а, простор који је обухваћен *TPS* базисима се скраћује на простор мање брзих промена, одређено бројем базиса K . С друге стране, у поступку естимације модел функције се адаптира подацима. На пример, уколико је реална функција линеарна, без обзира колико велики број базисних функција је специфициран, њена апроксимација ће бити блиска линеарном моделу. Дакле, при одабиру броја *TPRS* базисних функција, потребно је водити рачуна само да њихов број не буде рестриктивно мали за дату функцију и онемогући моделовање важних варијабилности. У *mgcv* пакету је подразумевана вредност $K=10$, која се показала довољном за моделовање већине реалних нелинеарности, што је и овде уважено за све ГАМ чланове реализоване *TPRS*-ом.

3.2 Хијерархијски генерализовани адитивни модел, ХГАМ

У тези се разматра и проширење стандардног ГАМ модела које урачунава хетерогеност у подацима увођењем случајних ефеката у модел. Пример примене од интереса је развој модела на бази корелисаних група података, на пример, прикупљених из више извора (са више места, локација). Сваки идентификовани извор података се препознаје као засебна категорија, G . Међутим, за разлику од члана 3 из Табеле 3.1 где се подаци који припадају различитим категоријама независно моделују, овде се претпоставља да постоји сличност функција зависности излазне променљиве од улазних вредности које припадају различитим категоријама, (мерено средњим квадратним одступањем).

Случајни ефекти су карактеристични за групу мешовитих модела, односно, генерализованих адитивних мешовитих модела, ГАММ [3.15, поглавље 6], који се базирају на засебним процедурама и методама, распложивим у одговарајућим софтвером у R -у [3.18]. Међутим, у тези је од интереса могућност увођења случајног ефекта коришћењем стандардних метода и алгоритама, поготово оних који подржавају методологију која се овде предлаже, што је дато у [3.19]. У поменутом раду се предлаже хијерархијски ГАМ-ом (ХГАМ), где се уводе ХГАМ чланови случајних ефеката, реализовани као инстанце

стандардних ГАМ чланова. Специфичност ХГАМ-а је што омогућава моделовање случајног ефекта глатком, нелинеарном функцијом, $f_q(x_q, G)$.

Комбиновањем стандардних ГАМ и ХГАМ чланова у реализацији модела је могуће покрити различита сценарија, као у неколико примера у поменутом раду [3.19]. Овде се разматра опција где је генерална (заједничка) нелинеарност моделована стандардним ГАМ чланом $f_k(x_k)$, којој се додаје члан индивидуалних случајних промена, $f_k(x_k, G)$ као што је дато једначином:

$$f_k(x_k) + f_k(x_k, G) \tag{3.14}$$

На основу тога, развој ХГАМ модела у тези се заснива, пре свега, на 1) идентификовању стандардног ГАМ модела којим се имплементирају основни фактори потрошње, 2) идентификовању група у подацима и оних ГАМ чланова за које постоји претпоставка да су извор хетерогености међу категоријама и 3) допуњавањем модела њиховим случајним еквивалентом.

Матрица пенала за случајне глатке чланове типа $f_q(x_q, G)$ је инверзна матрица корелације коефицијената базисних функција, [3.19], [3.15], док је параметар изглађивања исти за све категорије података.

3.3 Естимација ГАМ модела

Као што је већ наведено у претходним поглављима, естимација ГАМ-а се заснива на стандардним методама регресије са регуларизацијом брзине промена нелинеарних функција у моделу [3.17]. Развојем свих нелинеарних функција ГАМ-а (3.1) у сплајнове, добија се генерализовани линеарни модел $\mathbf{y} \approx \mathbf{V}\boldsymbol{\beta}$, где је излазна променљива, дефинисана за N узорака, $\mathbf{y} = (y_1, y_2, \dots, y_N)^T$ представљена матрицом модела $\mathbf{V} = (\mathbf{X} \dots \mathbf{V}_x \dots \mathbf{V}_{x_1, x_2} \dots)$, са векторима колона свих параметарских чланова, \mathbf{X} , базисних функција за чланове једне променљиве, \mathbf{V}_x и интеракције две променљиве, \mathbf{V}_{x_1, x_2} . $\boldsymbol{\beta} = (\beta_0 \dots \beta_x \dots \beta_{x_1, x_2} \dots)^T$ је вектор одговарајућих коефицијената који се, у генерализованом случају, одређује естиматором максималне веродостојности (*maximal likelihood estimation*, MLE).

Код адитивних модела се сваком нелинеарном члану модела додељује један (или више) пенала облика $\boldsymbol{\beta}^T \mathbf{S} \boldsymbol{\beta}$, (описано у претходном подпоглављу, за сваки тип ГАМ члана). Процењени коефицијенти модела $\hat{\boldsymbol{\beta}}$ се добијају решавањем оптимизационог проблема 3.15, где је функција веродостојности, $l(\boldsymbol{\beta})$, проширена сумом пенала за све функције модела (3.17 страна 164).

$$\hat{\boldsymbol{\beta}} = \underset{\boldsymbol{\beta}}{\operatorname{argmax}} l(\boldsymbol{\beta}) - \frac{1}{2} \sum_j \lambda_j \boldsymbol{\beta}_j^T \mathbf{S}_j \boldsymbol{\beta}_j \tag{3.15}$$

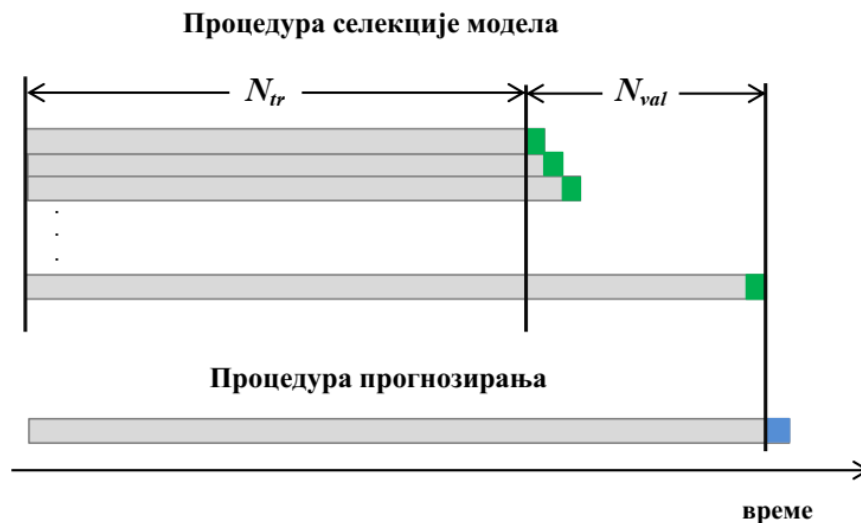
Параметри изглађивања, односно, јачина регуларизације модела се аутоматски одређује у естимацији једном од метода које су доступне у ГАМ развојном окружењу, типично, генерализована унакрсна валидација (*generalized cross validation*, GCV), *Un-Biased Risk Estimator*, UBRE, AIC, или, опција која се користи у тези, *Restricted Maximal Likelihood* (REML), [3.15, страна 128]. Веома важна алгоритамска унапређења, дата у [3.20], [3.31], омогућила су рад са великим подацима и значајно убрзала процес ГАМ естимације. Добијени

резултати и евалуација на проблему великих димензија је основно полазиште за разматрање динамичког моделовања ГАМ-а.

3.4 Избор ГАМ модела

Селекција ГАМ-а подразумева избор скупа ГАМ чланова који, за дате податке и дату примену, даје жељене перформансе моделовања. (Поступак селекције може обухватати и избор типа сплајна и број базисних функција). Скуп ГАМ чланова који је задат на основу почетних претпоставки и доменског знања, може бити или недовољан или, вероватније, делом нерелевантан и редундантан, што доводи до преобучавања и лоших резултата предикције. Поступак селекције релевантног ГАМ модела подразумева претрагу различитих комбинација ГАМ чланова и евалуацију одговарајућег модела одабраним критеријумом, и коначан избор оне комбинације која је дала најбоље резултате. За моделе мањих димензија, односно, за мањи број задатих ГАМ чланова, поступак претраживања се може обављати ручно [3.21] или се, иначе, може аутоматизовати, користећи различите шеме претраге, [3.22]. Критеријуми који се користе за избор ГАМ-а зависе од његове примене, али се најчешће користи неки од информационих критеријума као *Akaike Information Criterion* (AIC), *Bayesian Information Criterion* (BIC), или се модел бира, на пример, на основу анализе остатка од моделовања [3.23]. Специфично за прогностичке моделе, стандардни критеријум за избор модела је на бази грешке прогнозе, типично, *root-mean square error* (RMSE), *mean absolute error* (MAE), *mean absolute percentage error* (MAPE), [3.21].

Као и код осталих техника машинског учења, и у ГАМ-у, ‘златни стандард’ у генерализацији процене перформанси модела и његовој селекцији је унакрсна валидација (*cross validation*). Суштина овог поступка је да се модел не евалуира на подацима за које се обучавао, чиме се тачније процењују његове предиктивне перформансе. Ради се о *resampling* методи, где се у више итерација врши подела расположивих података на оне намењене обучавању модела и оне за тестирање, и перформансе модела се добијају усредњавањем резултата за различите поделе. Овако добијена процена перформанси модела се користи као критеријум за његову селекцију. За временске серије, поступак унакрсне валидације је специфичан и неке смернице се могу наћи у [3.25] и [3.24]. Шема избора модела временске серије унакрсном валидацијом, дата је сликом 3.7.



Слика 3.7: Временска шема унакрсне валидације за временске серије

Процедура се спроводи тако што се од података који су на располагању у тренутку моделовања, N_{val} најскоријих вредности издвоје за валидацију модела, док се претходних N_{tr} користе за његово обучавање у пропорцији, која се, кроз литературу, креће од 50:50% до 95:5%. По горњој шеми, за дати скуп ГАМ чланова, модел се итеративно, естимира користећи податке одвојене за обучавање (сиво), врши прогноза наредне вредности (зелено) и израчуна грешка прогнозе, све до краја података на располагању. Описани поступак се понавља за различите комбинације ГАМ чланова, и бира се она за коју се добија најмања средња грешка прогнозирања. У процедури прогнозирања се изабрани модел, поново естимира за укупне расположиве податке, $(N_{val}+N_{tr})$ и, на даље, врши прогноза наредне вредности (плаво). Шеме претраживања ГАМ чланова се могу заснивати на различитим поступним процедурама, на пример, сукцесивном додавању значајних чланова (*forward selection*) или искључивању оних нерелевантних (*backward selection*), до задовољења задатког критеријума. Најчешће се користи *stepwise* процедуре, која комбинује поступак сукцесивног додавања и елиминације ГАМ чланова и омогућава флексибилну претрагу. У пракси, се, за велики број ГАМ чланова који се разматра, а сходно проблему и примени модела, претраживање може спровести и парцијално за одређени скуп чланова или независно за различите подгрупе чланова [3.24]. Коначно, за претражене *све* комбинације ГАМ чланова (тзв. *best subset*), описани поступак унакрсне валидације је дефинитиван и референтан приступ за избор предиктивног модела и евалуацију алтернативних метода.

Основни проблем наведених стандардних методологија избора модела је да се ради о итеративним поступцима. У многим практично значајним случајевима, поготово за моделе намењене раду са подацима из више, могуће великог броја, извора, од интереса је избећи итеративане и интерактивне процедуре. Њихова аутоматизација, такође, није решење, јер тај приступ има своја предвидива ограничења, за већи број ГАМ чланова уколико се разматра могућност динамичког моделовања.

Као алтернатива у тези су размотрене методе додатне регуларизације естимационе процедуре, које су се показале компјутерски ефикасније и које дају резултате компарабилне са итеративним методама [3.26]. Идеја ових метода је да се поступка естимације модела додатно пенализује, тако да омогућава нулте вредности коефицијента⁸, последично, искључење одговарајућих чланова из модела. Пример примене селекције адитивних модела регуларизацијом је дат у радовима [3.27] где се користи групни LASSO (*Least Absolute Shrinkage and Selection Operator*) и AIC, BIC и *Generalized Cross Validation (GCV)* критеријуми за селекцију ГАМ чланова. Поред стандардних метода регуларизације регресионих модела, развијено је и неколико метода, специјално за ГАМ, [3.26, 3.28-3.30] У тези је разматрана метода дуплих пенала (*double penalty*), предложена у референци [3.26]. Избор ове методе је вођен, превасходно једноставношћу и лаком интеграцијом у естимациону процедуру. Такође, аутори су навели да су резултати компарабилни са оним добијеним унакрсном валидацијом методом. Практично значајна карактеристика ове методе је у томе што се избор ГАМ чланова обавља паралелно са естимацијом, у једном пролазу, производећи само једна модел.

Метода дуплих пенала се заснива на декомпозицији матрице пенала \mathbf{S}_j на компоненту пенализовану у естимационој процедури (*range space*) и непенализовани остатак (*null space*). Идентификација нултог простора сплајна се добија декомпозицијом у простор сопствених вектора по формули:

⁸ Ове методе се још и називају *shrinkage* методе

$$\mathbf{S}_j = \mathbf{U}_j \mathbf{\Lambda}_j \mathbf{U}_j^T \quad 3.16$$

где је \mathbf{U}_j је матрица сопствених вектора, а $\mathbf{\Lambda}_j$ је одговарајућа дијагонална матрица сопствених вредности. Нулти простор матрице пенала је онај за које $\mathbf{\Lambda}_j$ има нуле и одговарајуће сопствене векторе \mathbf{U}_j^* . Тада се формира додатна матрица пенала дефинисана са 3.17, која се додаје постојећим пеналима модела у 3.15, тако да се процена коефицијената модела, са уграђеном процедуром избора ГАМ чланова, добија решавањем оптимизационог проблема 3.18.

$$\mathbf{S}_j^* = \mathbf{U}_j^* \mathbf{U}_j^{*T} \quad 3.17$$

$$\hat{\boldsymbol{\beta}} = \operatorname{argmax} l(\boldsymbol{\beta}) - \frac{1}{2} \sum_j (\lambda_j \boldsymbol{\beta}_j^T \mathbf{S}_j \boldsymbol{\beta}_j + \lambda_j^* \boldsymbol{\beta}_j^T \mathbf{S}_j^* \boldsymbol{\beta}_j) \quad 3.18$$

За разлику од итеративних поступака који представљају дискретну процедуру избора ГАМ модела, методама регуларизације се континуално претражује простор дефинисан унапред специфицираним члановима, типом сплајна и бројем базисних функција за сваки. За конкретну примену је важно нагласити да метода дуплих пенала има своја ограничења.

1. Ограничење методе дуплих пенала је што има ефекат само на непараметарске чланове модела који су реализовани сплајновима. Овом методом, међутим, није могуће обавити селекцију фиксних, параметарских чланова модела (тип 1, Табела 3.1). Један од начина да се, ипак, генерализује поступак избора модела методом дуплих пенала, а који је кориштен у раду, је да се све континуалне улазне променљиве, за које је то оправдано, моделују као непараметарски чланови. За економично решење је могуће, на пример, користити мањи број базисних функција за њихово моделовање. Изузетак су категоричке варијабле које остају у моделу непенализоване.
2. Друга ‘мера опреза’ је да се приликом димензионисања модела води рачуна да метода дуплих пенала има ефекат уколико је број узорака који се користе за обучавање, минимално, једнак броју коефицијената модела. То значи, да уколико је на располагању мали број података за естимацију модела, потребно је водити рачуна о броју ГАМ чланова и базисних функција којима се моделују.

3.5 (X)ГАМ примена на временске серије и методологија динамичког прогнозирања

Домен примене ГАМ (ХГАМ) модела који је од интереса у тези су временске серије односно, временски зависне променљиве, чије се промене хронолошки прате. Прогнозирање временске серије је задатак предикције њених будућих вредности, а на основу тренутних и оних из претходног периода, урачунавајући, генерално, и утицаје спољашњих фактора. Формално, (X)ГАМ модел временске серије у се описује једначином:

$$y_t = (h)gam(y_{t-1}, y_{t-2}, \dots, \mathbf{x}_t, \mathbf{x}_{t-1}, \dots) + e_t \quad 3.19$$

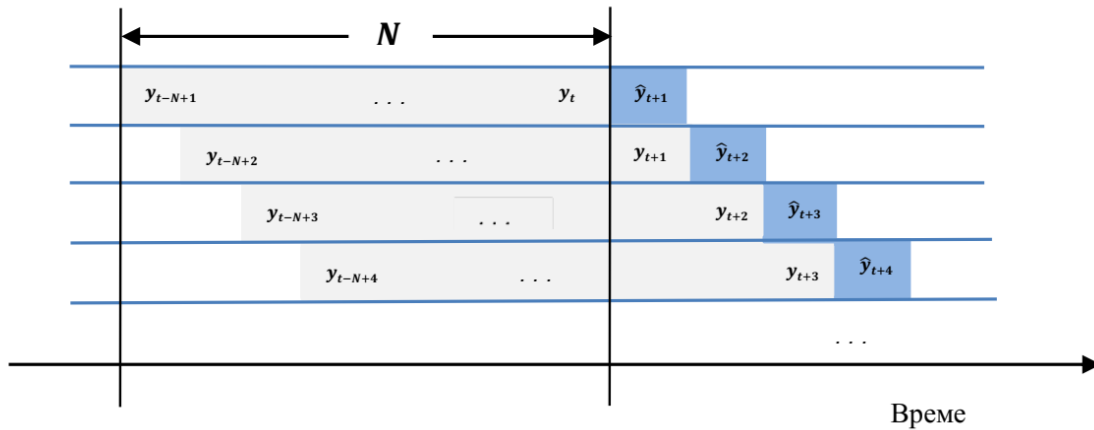
где y у времену t зависи од својих претходних вредности у временима $t - 1, t - 2, \dots$ и од тренутних и претходних вредности спољашњих променљивих из вектора \mathbf{x} . $(h)gam$ у једначини означава збир иницијално дефинисаног скупа (X)ГАМ чланова, одабраног за моделовање ових зависности.

Користећи се горњом једначином, развијена је процедура за прогнозирање са клизајућим прозором по временској шеми са слике 3.7, где се у сваком тренутку посматрања временске серије обављају два задатка:

- врши естимација и селекција $(h)gam$ модела, у складу са 3.19. За естимацију модела се користе актуелне вредности, y_t и \mathbf{x}_t и временски прозор од N претходних вредности y и свих улаза модела. Као резултат овог поступка добија се модел са актуелним члановима и вредностима коефицијената, $(h)gam_t$.
- естимирани $(h)gam_t$ врши прогнозу наредне вредности временске серије, \hat{y}_{t+1} , користећи претходне и актуелне вредности улаза модела и прогнозе (или реалне вредности) наредних вредности спољашњих променљивих, $\hat{\mathbf{x}}_{t+1}$, по једначини:

$$\hat{y}_{t+1} = (h)gam_t(y_t, y_{t-1}, \dots, \hat{\mathbf{x}}_{t+1}, \mathbf{x}_t, \dots) \quad 3.20$$

Поступак и прозор са подацима се помера за један корак напред и процедура се понавља.



Слика 3.8: Временска шема процедуре за прогнозу са клизајућим прозором

Горњи поступак намењен прогнози наредне вредности временске серије се може искористити за генерисање прогноза које покривају већи временски хоризонт, за времена, $t + 1, t + 2, \dots$. Типично, вредности \hat{y}_{t+2}, \dots се добијају рекурзијом, на основу прогноза $\hat{\mathbf{x}}_{t+2}$ и естимираних вредности \hat{y}_{t+1} .

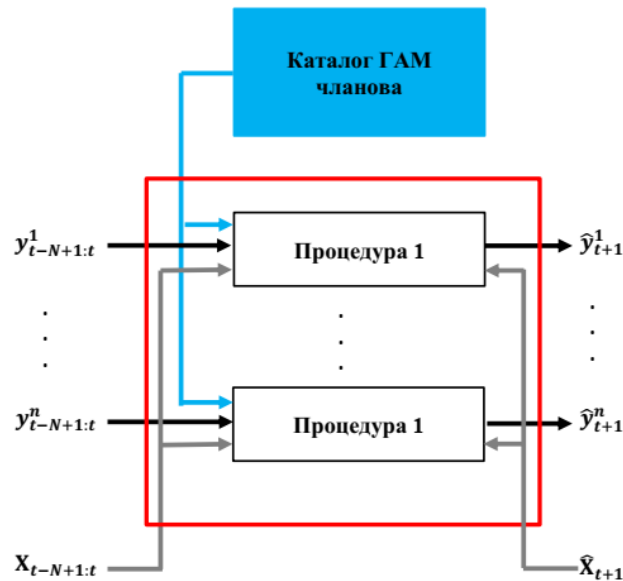
Специфичност предложене процедуре је што се поред регуларног и континуираног прорачуна коефицијената модела, са сваким новим податком, врши и поновни избор селекције (X)ГАМ чланова чиме се пројектује боље динамичко адаптирање модела на промене услова прогнозирања, као и бољи резултати прогнозе. Описана процедура (у наставку Процедура 1 за стандардан ГАМ модел, односно, Процедура 2, за XГАМ модел) подржава прогностичку машину (*forecasting engine*), за аутоматско генерисање прогноза сваке временске серије за коју је иницијални сет (X)ГАМ чланова из *gam* и величина временског прозора, N , валидна претпоставка. Ова чињеница је искоришћена за развој методологија прогнозирања више временских серија, базираних на ГАМ и XГАМ моделима, како је описано у наставку.

3.5.1 Изведена (X)ГАМ методологија за прогнозирање скупа временских серија

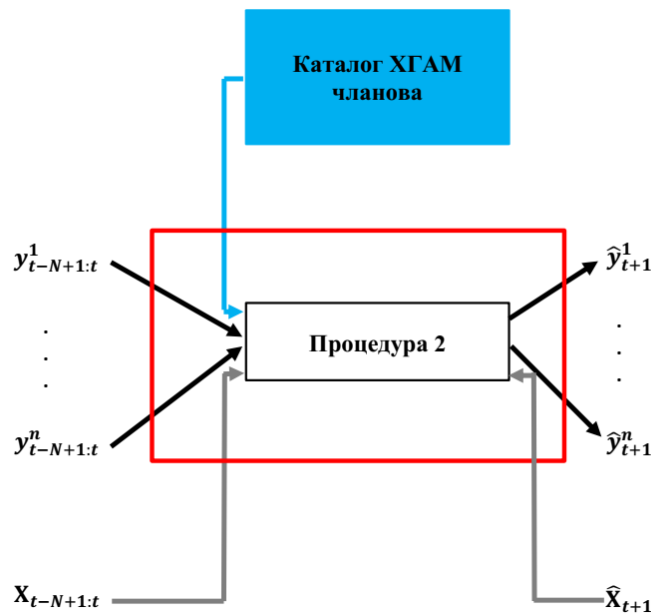
На слици 3.9 је дата функционална шема ГАМ базиране методологије прогнозирања n независних временских серија. Методологија се заснива на каталогу иницијално

дефинисаних ГАМ чланова којима се описују све посматране временске серије и Процедури 1 која се спроводи за сваку од њих. Претпоставке практичне примене приказане методологије су да је за дати проблем могуће дефинисати јединствен каталог ГАМ чланова и да је Процедура 1 довољно брза да подржи захтевани прогностички задатак.

На слици 3.10 је дата функционална шема ХГАМ базиране методологије прогнозе n корелисаних временских серија. Она се заснива на каталогу иницијално дефинисаних ХГАМ чланова и Процедури 2, која се симултано спроводи за све временске серије. Као проширење ГАМ модела за скуп корелисаних временских серија, додатна претпоставка практичне примене ХГАМ модела је корелисаност временских серија за које се врши прогноза.



Слика 3.9: Шема ГАМ базиране прогнозе више независних временских серија.



Слика 3.10: Шема ХГАМ базиране прогнозе више корелисаних временских серија.

4. Карактеристике временских серија потрошње

Како би се применила методологија описана у претходном одељку на временску серију потрошње електричне енергије, који ће се користити за развој функционалности њеног прогнозирања, у наставку су размотрени начини дефинисања каталога ГАМ чланова потрошње. Чланови ГАМ-а имплементирају факторе потрошње, који представљају њене појединачне компоненте које зависе од календара, метеоролошких услова, економских параметара, итд. Услов који је потребно задовољити при њиховом избору је да, за расположиве податке, дефинисани скуп буде довољан, односно, да у потпуности описују потрошњу која се моделује. За функционалност која је намењена прогнозирању више потрошњи, дефинише се јединствен скуп ГАМ чланова где претходни услов треба да буде задовољен за све потрошње из посматраног скупа. Коначно, крајњи циљ је генерализација приступа избора чланова ГАМ-а и развој методологије која би задовољила претходне услове за различита места у електроенергетском систему.

На основну досадашњих знања о карактеристикама временске серије потрошње, као и на бази резултата истраживање, у дефинисању номинованог сета улазних променљивих и пропратне методологије разматрани су:

- Основни фактори потрошње за различита места у систему и различите агрегације потрошача, како би се нашао одговор на питање које улазне променљиве одабрати за специфично место у систему.
- Релације и међусобне зависности између појединачних фактора потрошње, за дефинисање чланова интеракције у ГАМ моделу.
- Разлике у релевантности фактора потрошње за различите временске серије потрошње и у времену и оправданост динамичке селекције ГАМ чланова.

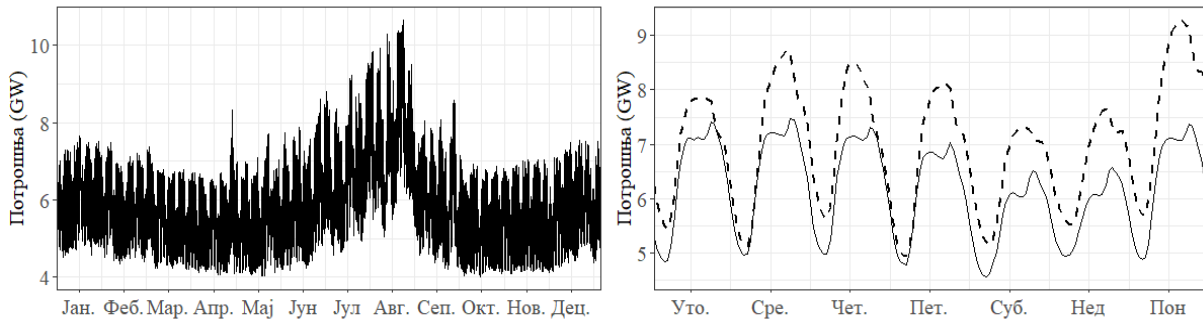
У наставку су размотрене основне компоненте потрошње и начини конструкције одговарајућих ГАМ чланова који су коришћени у тези.

Потрошња електричне енергије представља временску серију чије су основне карактеристике мултисезоналност и утицај спољашњих фактора. Основни фактори сваке потрошње електричне енергије, који су у мањој или већој мери изражени су подељени у три категорије: зависност од календарских променљивих, од метеоролошких услова и њених вредности из претходног периода. Поред основних, постоје и специфичности одређене потрошње, као на пример, утицај специјалних дана, глобални тренд, нови фактори потрошње тј. утицаји из ОИЕ и DSM/DR, неочекивани догађаји (COVID), итд.

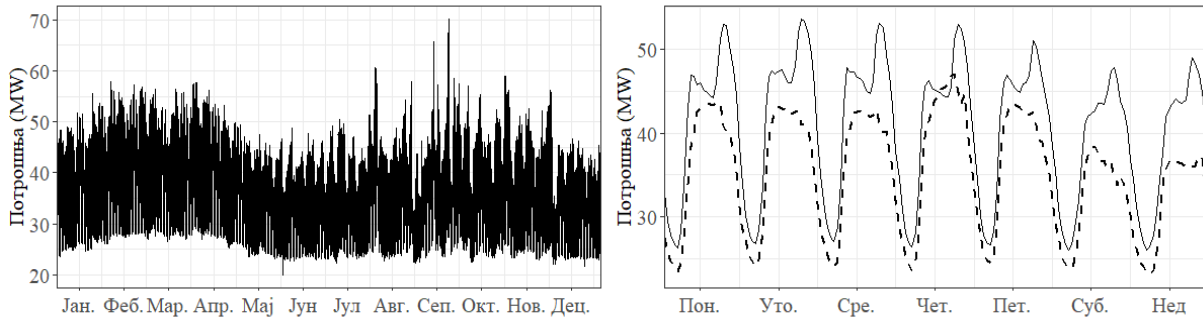
4.1 Календарске променљиве

Мултисезоналност временске серије потрошње се односи на постојање годишње, седмичне и дневне сезоналности, које су резултат календаром одређених периодичних активности и навика потрошача, карактеристичних за одређени дан у години, дан у седмици и доба дана, дефинисаних тзв. календарским променљивим. Обично се идентификује као годишњи, седмични и дневни профил потрошње, чији облик зависи од географског подручја, климатских услова, типа потрошача (индустријски, комерцијални, резиденцијални), састава становништва (урбано, рурално), начина коришћења, коначно, броја потрошача којом је обухваћена прогностичка област. За илустрацију, на сликама 4.1-4.4 приказане су сатне вредности потрошње у току једне године (лева колона) и током седмице (десна колона), за четири различита система и нивоа агрегације.

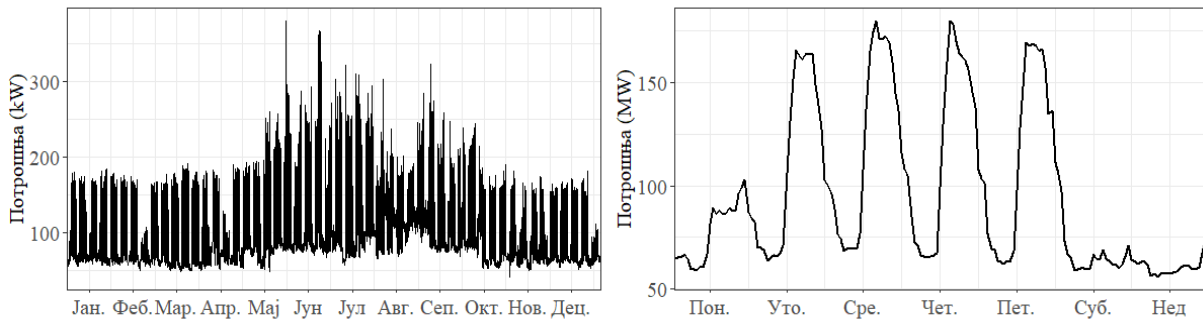
Карактеристике временске серије потрошње



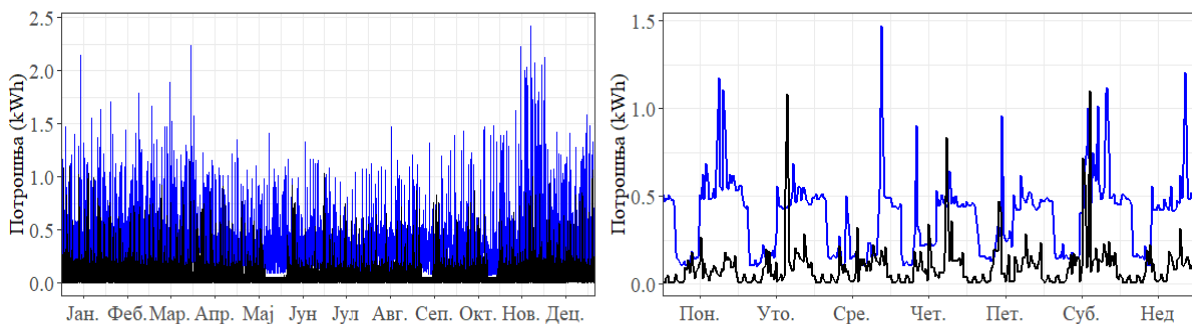
Слика 4.1: Сатне вредности потрошње током године и летње/зимске седмице за NYC зона



Слика 4.2: Сатне вредности потрошње током године и летње/зимске седмице за AU рејон



Слика 4.3: Сатне вредности потрошње током године и летње/зимске седмице за EN зграда



Слика 4.4: Сатне вредности потрошње током године и летње/зимске седмице за 2 индивидуална LCL потрошача

На слици 4.1, лево, приказан је годишњи профил за NYC зону (NY подаци) са израженим летњим максимумом. Повећање потрошње за период јул/август у односу на годишњи просек је последица знатно веће употребе хлађења лети, што је посебно наглашено за потрошње већег броја потрошача. Десно су приказани примери седмичне потрошња лети (испрекидана линије) и зими (пуна линија), са типичним падом потрошње за дане викенда.

Карактеристике временске серије потрошње

Карактеристичне су разлике у летњим и зимским дневним профилима и распореду максимума потрошње у току дана у зависности од употребе хлађења. На слици 4.2, су приказани исти профили само за потрошњу трафо рејона AU1 у околини Сиднеја, где се, у односу на претходни пример, виде разлике у профилима потрошње за различито географско подручје и различит број потрошача. Слика 4.3 је приказан годишњи и седмични профили потрошње нерезиденцијалне зграде EN1 (EN подаци). EN зграде користе гас за грејање, тако да је прираштај потрошње електричне енергије у односу на просечну, махом, у летњим месецима. Седмични профил је карактеристичан за нерезиденцијалне потрошаче са изражено ниском потрошњом за дане викенда. На слици 4.4 су приказани годишњи и седмични профили два резиденцијална индивидуална потрошача са истог географског подручја (LCL). Уочавају се разлике у годишњим профилима које карактерише повећање потрошње зими (плаво) или периоди одсуства укућана (црно), слабо изражен седмични профил, док дневним профилом доминира стохастичка компонента дневне потрошње уређаја у домаћинству, итд. Дати преглед профила и пропратна објашњења за различите системе и број потрошача у ЕЕС-у служи увиду у различитост нелинеарности које је потребно моделовати у ГАМ-у.

Календарске променљиве дефинисане су временским индексима, на пример, за сат у дану, дан у седмици и дан у години. Неколико је стандардних начина њиховог дефинисања који се могу користити и у ГАМ-у, најчешће *dummy* варијабалама, тригонометријским функцијама или засебним моделовањем сезона. Профил потрошње моделован *dummy* варијабалама x_s , за S сезона (на пример, за дане у седмици, $S=7$) је дат формулом 4.1.

$$\sum_{s=1}^S a_s x_s \quad 4.1$$

За нерезиденцијалну зграду са слике 4.1ц, где су од интереса само промене потрошње током радних дана, могуће је усвојити 6 категорија, понедељак до петак и викенд променљива, док код индустријских или већих комерцијалних потрошача, за радни/нерадни дан. За дан у години је могуће увести дводимензиони вектор, за месец и дан у седмици [4.1], или временски индекси могу бити угњеждени, за дан у седмици, седмица у месецу, месец у сезони [4.2], итд.

Коришћењем синусних и косинусних функција, сезоналност се моделује по 4.2. и пример је дат у [4.3].

$$\sum_{k=1}^{S/2} \left\{ \alpha_k \cos\left(\frac{2\pi k}{S}\right) + \beta_k \sin\left(\frac{2\pi k}{S}\right) \right\} \quad 4.2$$

Коначно, могуће је увести засебан модел за сваку сезону (категорију), на пример, за сваки сат у дану [4.4] или дан у седмици [4.5], итд. У тези су дневни профили потрошње реализовани засебним моделовањем потрошњи које одговарају сваком сату у дану⁹.

Поред стандардних метода, специфично за ГАМ моделе, могућа је конструкција глатког, нелинеарног профила, реализованог цикличним кубним сплајном са задатим бројем базисних функција по 4.3.

⁹ Подразумева се генерализација на друге интервале унутар дана, на пример, полусатне, 15-минутне, зависно од расположивих мерења и апликације модела.

$$f_s(x_s) = \sum_{s=1}^s a_s b_s(x_s) \quad 4.3$$

У пракси се сплајном моделован седмични профил, $f_s(x_s)$ типично реализује са 7, а годишњи профил, $f_g(x_g)$, са 12, еквидистантно распоређених цикличних кубних сплајнова.

Ефекат промена седмичних профила током године (разлика између летњег и зимског профила на сликама 4.1а-б десно), се често изводи категоричким варијаблама, за лето/зима (или месечно) коришћењем члана 2 и 3 из Табеле 3.1. У ГАМ-у је могуће ову интеракцију реализовати увођењем глатког члана интеракције $f_{gs}(g, s)$ (члан 4., Табела 3.1). Предност овакве репрезентације интеракције променљивих у односу на коришћење категоричких варијабли је што представља континуалну генерализацију заједничког ефекта две променљиве, бржа се естимира за већи број категорија и преферабилан је избор у ГАМ-у.

4.2 Метеоролошки фактори потрошње

Утицај метеоролошких услова је најзначајнији спољашњи фактор потрошње електричне енергије, поготово важан за делове система са претежно резиденцијалним потрошачима, где је утрошак електричне енергије услед употребе грејања/хлађења значајнији. Узрочник промена потрошње електричне енергије услед метеоролошких услова је субјективни доживљај температуре амбијента и за моделовање ове компоненте потрошње, поред вредности температуре ваздуха, користе се и друге метеоролошке варијабле. На пример, зими, за дату температуру ваздуха, већа брзина ветра доводи до повећања потрошње која се користи за грејање, мање значајно у летњим месецима. С друге стране, лети, ниво влажности ваздуха утиче на субјективни доживљај топлоте и повећану употребу клима уређаја. Поред наведених, прецизнијем моделовању метеоро фактора потрошње могу допринети и урачунавање, на пр. количина падавина, облачност, ваздушни притисак, сунчево зрачење. Важно је истаћи да се „нетемпературне“, тзв. секундарне метеоро варијабле ретко користе независно и махом се комбинују са температуром ваздуха.

У прелиминарном прегледу из референце [4.6], наведени су типично коришћени начини уврштавања метеоролошких варијабли у модел. Најједноставнији приступ је директно коришћење мерења као улаза модела. За температуре се, типично, користе сатне или дневне вредности (средња дневна вредност, дневни максимум и минимум). Такође, честа је и употреба *heating degree-days* (HDD) and *cooling degree-days* (CDD) величина, које су дефинисане као одступање средње дневне вредности температуре од унапред задатих температурних прагова коришћења грејања/хлађења. Брзина ветра се, због велике варијабилности, ретко директно уводи у модел и најчешће се користи изведена метеоро величине која комбинује брзину ветра и температуру, као на пример, *wind chill index* (WCI), [4.7] или *cooling power of the wind* (CP). Слично, ниво влажности ваздуха се, најчешће, представља варијаблама прорачунатим у односу на температуру *relative humidity*, *wet-bulb, dew point temperature*) или комбиновано са њом (*heat index* (HI), *apparent temperature* и *temperature-humidity index* (THI)), [4.8]. Посебно *real sense of temperature* је изведена метеоро варијабла која комбинује температуру, ниво влажности и брзину ветра.

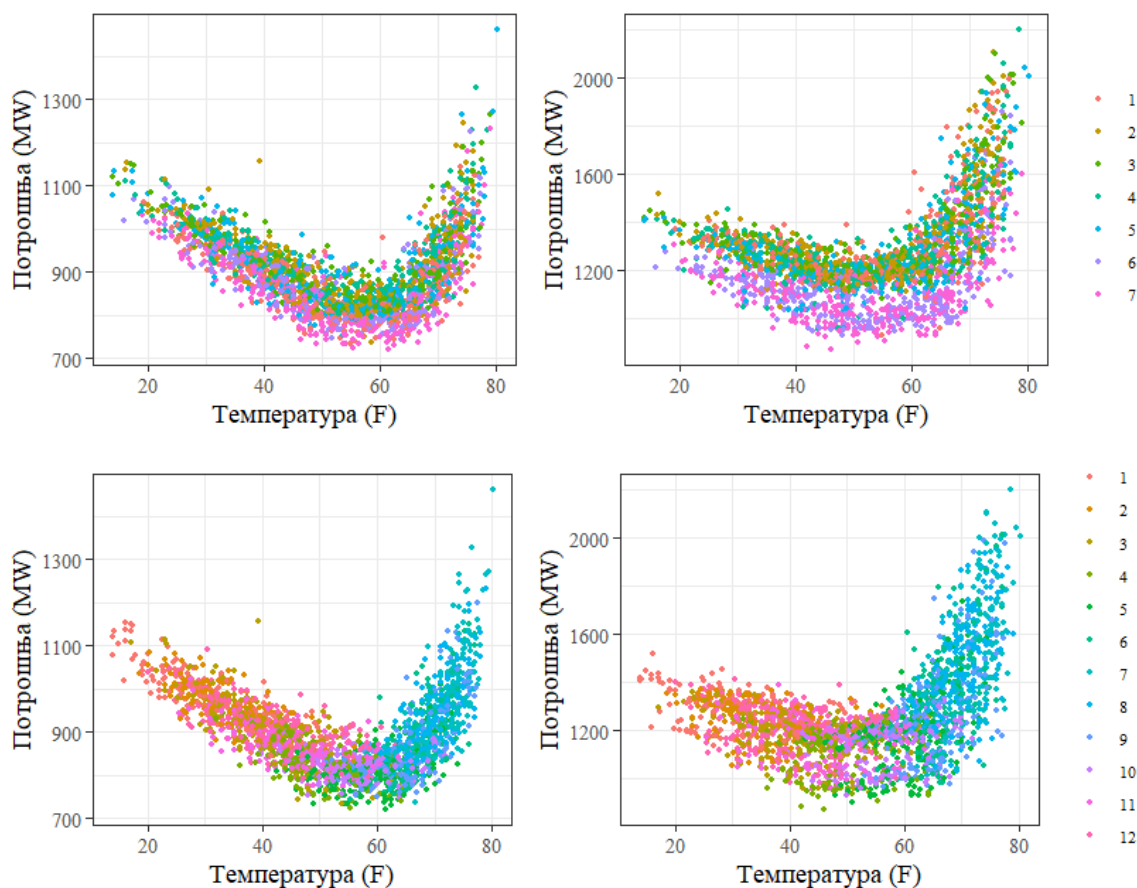
Један од значајних метеоролошких ефеката на потрошњу је и температурна инерција, односно, утицај вредности температура из претходног периода. Начини урачунавања овог ефекта зависе од усвојеног приступа моделовања и најједноставнији је онај где се вредности температуре из претходног периода (претходних дана или сати) директно уводе у модел.

Карактеристике временске серије потрошње

Алтернативно, могу се користити и њихове средње вредности или тзв. ефективна температура која се добија њиховим експоненцијалним изглађивањем. Број претходних вредности за температуру и секундарне варијабле се дефинишу унапред или на основу претходне анализе или поступком избора модела.

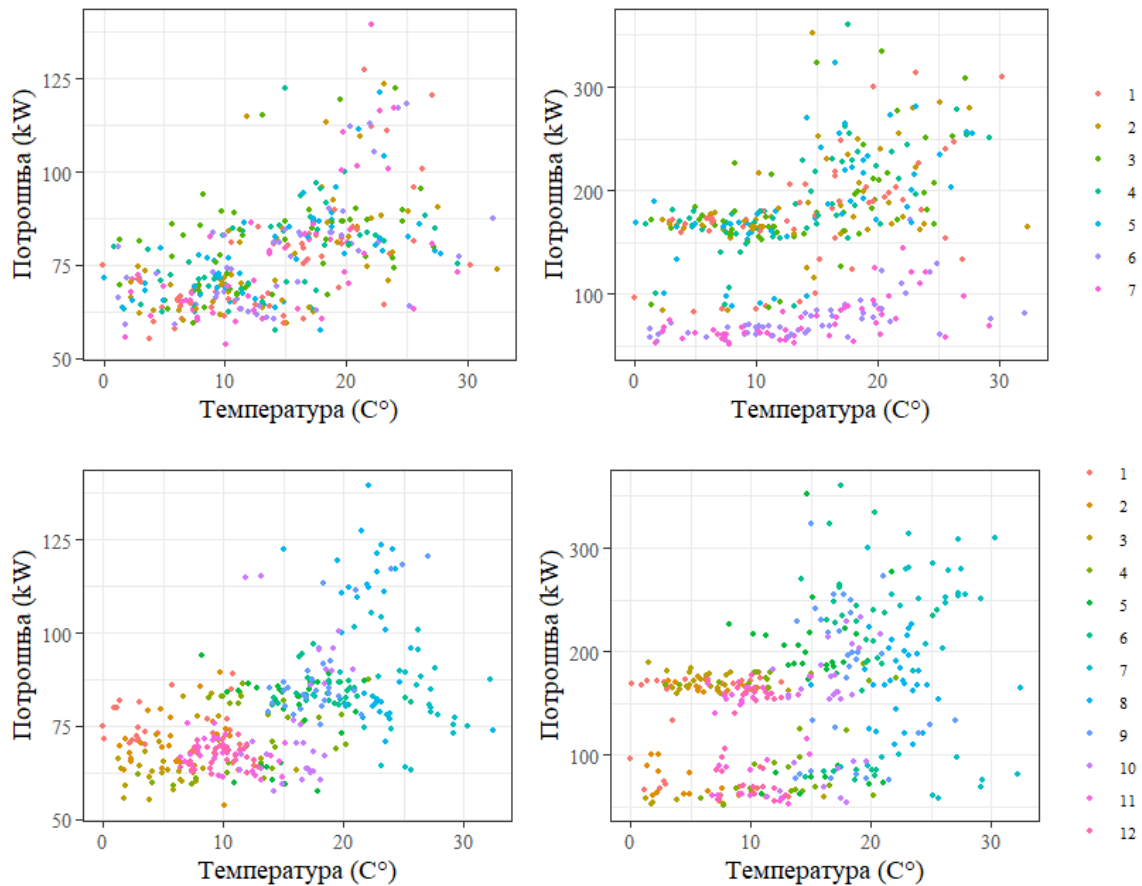
Већина *STLF* алата у прегледаној литератури, а у којој се не разматра ОИЕ утицај, засновано је само на температурама ваздуха, било због недоступности мерења секундарних варијабли, или је констатовано да унапређење тачности прогнозе није загарантовано њиховим увођењем [4.9], или се показало да су за одређена места у систему нерелевантне [4.10], итд. У највероватнијем случају у интелигентним мрежама, кад не постоји проблем расположивости мерења и кад постоји методологија избора релевантних метеоролошких величина за дату прогностичку област, као критичан, остаје проблем квалитетних прогноза за поједине метео варијабле (на пр. брзину ветра).

На сликама 4.5 и 4.6 су приказане зависности потрошње од средње дневне температуре, за GENESE зону (NYC) и EN1 зграду, лево за 3. сат, десно за 12. сат. Истакнуте су разлике за различите дане у седмици (горе) и за различите месеце унутар године (доле).



Слика 4.5: Зависност потрошње од температуре за GENESE зону, за дане у седмици (горе), за месеце у години (доле)

Карактеристике временске серије потрошње



Слика 4.6: Зависност потрошње од температуре за EN1 зграду, за дане у седмици (горе), за месеце у години (доле)

Генерално, зависност потрошње од температуре је U-крива, где доњи крак криве приказује повећање потрошње услед повећаног грејања на ниским температурама и горњи услед повећаног хлађења при високим температурама. Ова нелинеарност се, за географски различите прогностичке области и зависно од састава потрошача, може значајно разликовати. На пример, Европске земље су подељене на три типа земаља на основу зависности потрошње од температуре: хладне/топле, за које горња зависност има само доњи/горњи крак која може бити блиска линеарној, и остале земље чија зависност потрошње од температуре, мање-више следи облик криве. Зграда са слике 4.6, на пример, користи гас за грејање и уочљиво је повећање потрошње само на вишим температурама услед повећане употребе клима уређаја, док је за ниже вредности температуре њен утицај на потрошњу мање релевантан. У генералном случају, за моделовање приказане зависности сваке потрошње, уводи се нелинеарни ГАМ члан $f_T(T)$, где је T вредност температуре, реализована $TPRS$ сплајном, који ће се у поступку естимације модела прилагодити подацима за одређену прогностичку област. На даље, промене ове зависности током седмице и током године, односно, разлике у реакцији потрошача на температуру за различите дане у седмици и години, моделоване су члановима интеракције $f_{T,g}(T, g)$ и $f_{T,s}(T, s)$, за g и s дефинисане као у 4.1. Са ова три члана се моделују промене потрошње услед температуре ваздуха унутар дана. За урачунавање ефекта инерције температуре, односно, утицаја вредности температуре из претходног периода (у овом случају, из претходних дана) на тренутну вредност потрошње, могуће је дефинисати исти сет чланова са температурама из претходног периода (претходних дана).

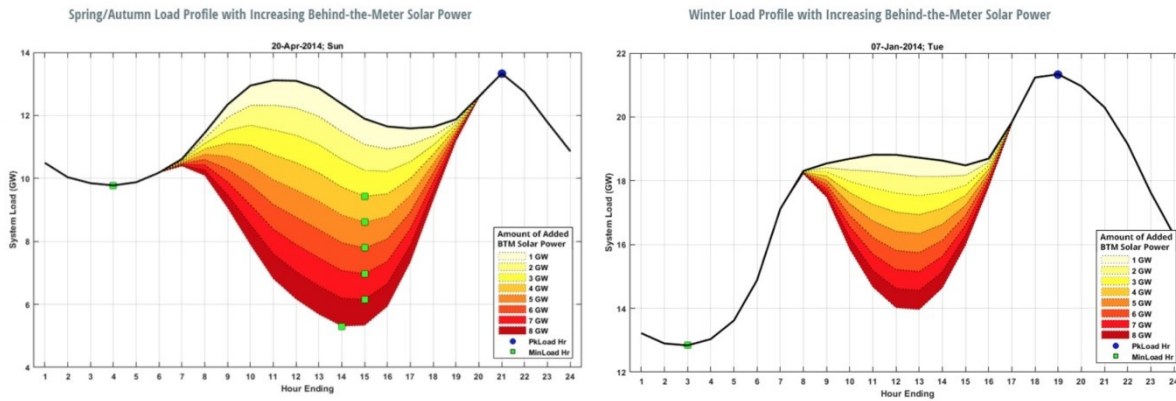
Зависност потрошње од брзине ветра, влажности ваздуха, или неке друге секундарне варијабле, SV , као и њихове промене у зависности од дана у години и седмици се, као и за температуру, могу моделовати као нелинеарна зависност $f_{SV}(SV)$ и члановима интеракције $f_{SV,g}(SV, g)$ и $f_{SV,s}(SV, s)$, респективно. Међутим, значајна је њихова интеракција са температуром, чији се утицај може генерализовати уврштавањем додатног члана, типа $f_{T,SV}(T, SV)$, за сваку од урачунатих секундарних варијабли, SV . За разлику од појединих, горе наведених, изведених метео величина, прорачунатих за унапред задате параметре и вредности температура, овако дефинисан модел комбиновања температуре са осталим секундарним варијаблама је у потпуности вођен подацима.

Компонента потрошње електричне енергије која је, такође, базирана на метеорологији, представљају промене услед утицаја ОИЕ производње [4.11], [4.12], [4.13]. Део истраживања који се бави управљањем ЕЕС-ом, заснован је на независним прогнозама за ОИЕ производњу и потрошњу електричне енергије, са или без урачунате њихове међусобне корелисаности [4.11]. У том случају, сваки од модела који се користи за прогнозу има засебно моделовану метео компоненту и засебан скуп метео варијабли. Овакав приступ постаје проблематичан за области са значајним бројем имплементираних „*behind the meter*“ дистрибуираних извора енергије, због слабије расположивости мерења произведене и потрошене електричне енергије са конкретне локације. За такву прогностичку област, прогнозира се потрошња измерена у мрежи, тзв. *net load*, која бележи разлику тоталне потрошње електричне енергије (без утицаја ОИЕ производње) и укупне ОИЕ производње. Тада су промене услед утицаја ОИЕ производње уграђене у профиле потрошње из мреже, што захтева проширење модела додатним спољашњим варијаблама.

На слици 4.7¹⁰ су приказане промене летњег и зимског дневног профила потрошње електричне енергије из мреже за различите нивое пенетрације солара (тзв. *duck curve*). У односу на криву тоталне потрошње (црно), промене су карактеристичне за одређена времена у току дана и за различите периоде године, које коинцидирају са производњом, сходно нивоу пенетрације. Ове промене се, индиректно, моделују додавањем релевантних метео варијабли, које, поред температуре, укључују брзину и правац ветра, влажност, притисак, облачност, количину падавина, видљивост, екстратерестријалну ирадијацију, глобалну ирадијацију, итд. У раду [4.12] су упоређивани резултати прогнозе потрошње из мреже добијене њеним директним моделовањем са оним добијеним као разлика прогноза тоталне потрошње и производње из солара, за област трафо-рејона, за 0, 17.8% and 35.6%. ниво пенетрације. За директно моделовање конкретне потрошње у мрежи кориштено је 17 метеоролошких варијабли, за тоталну потрошњу 8 и производњу из солара 16, тестирано за неколико модела. Компаративна тачност прогнозирања која је добијена за ова два приступа, показује да је, правилним избором метео варијабли, могуће постићи тачност и у недостатку „*behind the meter*“ мерења. Генерално, даља унапређења ове компоненте потрошње иду у правцу развоја техника генерисања метео параметара, као што су, на пример *sky-imaging* технике, препоручене за прогностичке области са високим степеном пенетрације солара, [4.13], итд.

¹⁰ Слика је преузета са сајта <https://www.greentechmedia.com/articles/read/massachusetts-is-staring-down-a-duck-curve-of-its-own-storage-could-help>

Карактеристике временске серије потрошње



Слика 4.7: Промене дневног профила потрошње електричне енергије за различите нивое пенетрације солара

У ГАМ-у се ове промене могу моделовати проширењем каталога ГАМ чланова додатним секундарним метео варијаблама и интеракцијама, на начин како је описано у 4.2. Међутим, додавањем великог броја улазних променљивих, значајно се усложњава модел и време његовог обучавања. С друге стране, постоје значајне разлике у броју и саставу обухваћених ОИЕ јединица, раположивости мерења и квалитетних прогноза, за различите области, дакле, релевантности појединих метео варијабли. Један од могућих приступа је да се, за конкретну област, број ГАМ чланова из иницијално задатог каталога сузи додатним поступком избора ГАМ чланова.

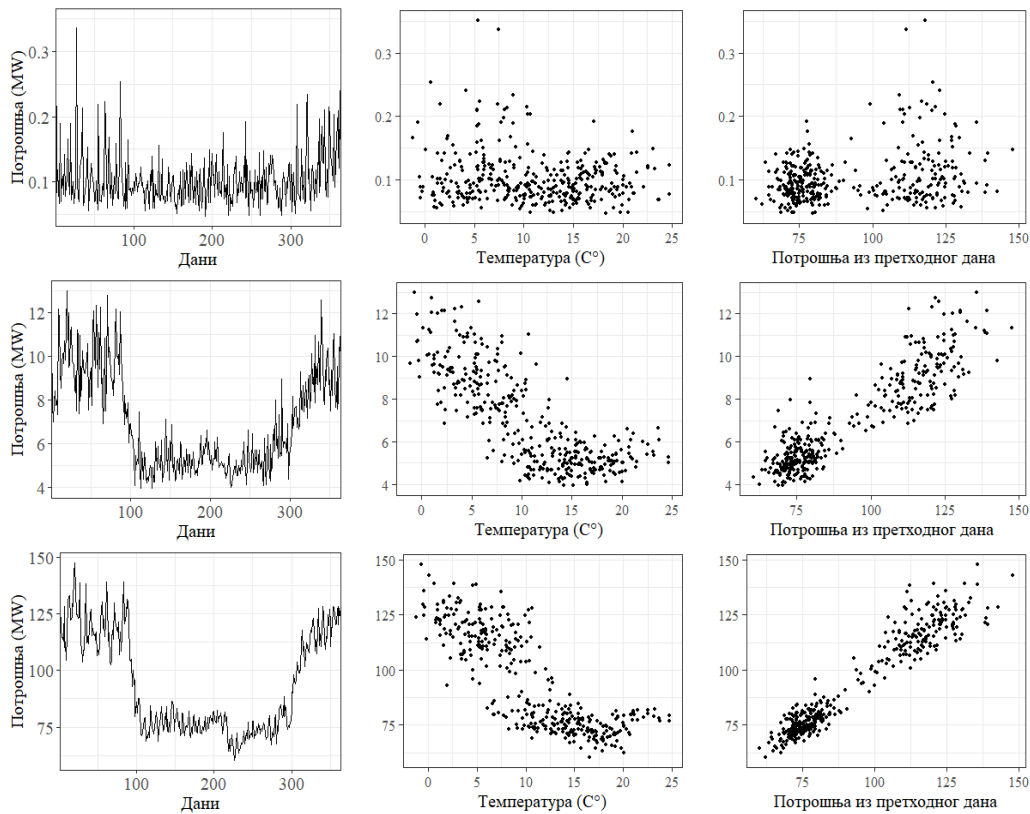
4.3 Утицај претходних вредности потрошње

Карактеристика потрошње је серијска корелација и, генерално, зависност актуелне вредности од реализација потрошње из претходног периода. Корелација која није обухваћена основним профилима и њене промене током времена се, за већину популарних прогностичких модела, урачунава додавањем вредности из претходних сати, дана и седмица. Структура временске серије потрошње, односно, број и времена претходних потрошњи који улазе у модел, је карактеристична за сваку појединачну потрошњу и захтева претходну анализу. Међутим, код генералног приступа, а и у регуларном поступку моделовања, овај избор се махом ослања на *a priori* процену колико вредности потрошње из претходног периода укључити у модел да допринесе побољшању тачности прогнозирања, а не повећа, значајно, његову комплексност. Минимална структура за сатне подмоделе која се користи код *STLF*-а, типично, урачунава вредност потрошње из претходног дана и претходне седмице, чиме се у обзир узимају промене дневне и седмичне сезоналности. Ова основна структура се, на даље, проширује додавањем потрошњи из претходних 2 или више дана и/или седмица. За сатни модел потрошње, типичан приступ урачунава утицај потрошњи из претходних 7 дана, односно, за потрошње у сату h , и дану d , P_h^d , њена вредност у текућем дану, $d = 0$, зависи од реализација из претходних, $d = -1: -7$, дана, за исти сат, $P_h^{-1}, \dots, P_h^{-7}$. За урачунавање унутар-дневне корелације, у модел се укључују и утицај потрошњи из претходног сата (односно, од претходних мерених или израчунатих вредности потрошње, за периоде од пола сата, 15 мин, 5 мин, итд.). На основу препорука из [4.10], оптималан избор за моделовање корелација између сатних потрошњи је урачунавање ефекта утицаја потрошњи из протекла два сата, $h = -1: -2$, P_{-1} и P_{-2} . ГАМ реализација ове компоненте потрошње је увођењем засебног TPRS члана модела за сваку урачунату потрошњу из

претходног периода, тако да је препоручена група ГАМ чланова којом се моделује $f^{-1}(P^{-1}), \dots, f^{-7}(P^{-7}), f_{-1}(P_{-1}), f_{-2}(P_{-2})$.

4.4 Карактеристике потрошње у зависности од нивоа агрегације потрошача

Један од увида са слика 4.1-4.4 су значајне разлике у профилима за различит број потрошача. Полазећи од агрегираних потрошњи као, на пример, оних на сликама 4.1 и 4.2, са јасно дефинисаним профилима, како се силази ниже ка индивидуалном потрошачу са слике 4.4, профили постају све комплекснији, мање структурирани, са израженом стохастичком компонентом. Смањење комплексности временске серије потрошње са растом броја потрошача је познато као ефекат агрегације и представља индиректну меру предвидљивости њених профила. На слици 4.8 су приказани примери годишњег профила (1. колона), зависности потрошње од средње дневне температуре (2. колона) и од потрошње из претходног дана, за 1, 50 и 600 произвољно одабраних потрошача (*LCL* подаци), 1-3 ред, респективно. Слабија корелисаност потрошње мањег броја потрошача са основним факторима потрошње који се користе за развој прогностичких модела утиче на могућност постизања задовољавајуће тачности моделовања. Последишно, значајне су разлике у тачностима краткорочне прогнозе добијене актуелним методама за различите агрегације потрошача и, на основу прегледане литературе, вредност *MAPE* грешке¹¹ се креће у распону од < 1% на нивоу целог система или већег региона, до преко 40% за индивидуалног потрошача.



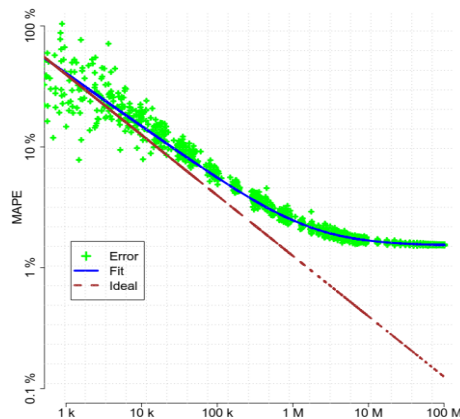
Слика 4.8: Зависност потрошње од основних фактора за 1, 50 и 600 индивидуалних потрошача

11

$MAPE = 100 \sum_k |P_k - \hat{P}_k| / P_k$, где је P_k реална вредност, \hat{P}_k њена прогноза

Карактеристике временске серије потрошње

Анализа ефекта агрегације за различите типове модела у [4.14] је показала да средња вредност MAPE грешке монотono опада са растом нивоа потрошње, до одређеног нивоа, тзв. критичне потрошње када достиже zasiћење. Вредност критичне потрошње је карактеристична за одређени тип модела и, као пример, на слици 4.9, преузетој из поменутог рада, приказана је ова зависност је приказана за SARMA(1,0)(1,0)₂₄ модел, за *Pacific Gas & Electric (PG&E)* податке (референца је дата у поменутом раду), са процењеном вредности критичне потрошње од 2250 kWh. Модел криве којом је квантификован ефекат агрегације, за конкретан прогностички модел, служи *a priori* процени његовог потенцијала у постизању одређене тачности прогнозе.



Слика 4 9: Зависност MAPE грешке од нивоа потрошње за SARMA(1,0)(1,0)₂₄ модел

Резултати из литературе која се бави поређењем различитих модела на основу зависности грешке прогнозе од нивоа агрегације потрошње, вертикално у систему, су показала да не постоји фаворит међу *моделима* и унапређења се, пре, обезбеђују одговарајућим *методологијама*. На пример, у [4.15] је неколико стандардних линеарних и нелинеарних модела, тестирани за ниво *VN/SN* трансформаторске станице (100 < 1000 потрошача), на нивоу *SN/NN* трансформаторске станице (од неколико стотина потрошача), агрегације од неколико до десетине потрошача и ниво индивидуалног потрошача, нису довели до унапређења тачности прогнозирања за ниже агрегације потрошње у односу на референтну, тзв. наивну прогнозу (код које је $\hat{P}_{t+1} = P_t$). Слично је констатовано у раду [4.16] где је размотрено више техника машинског учења која укључује и дубоке неуралне мреже, валидираним за различите нивое агрегације резиденцијалних и комерцијалних потрошача је показала да решења проблема нису у софистициранијим моделима и њиховом димензионисању, него, пре, у повећању предвидљивости самих података.

Уважавајући ефекат агрегације и њене последице на тачност прогнозе, као и број потрошњи за који се развија неколико је стратегија и методологија које се користе у пракси при реализацији практично корисних решења прогнозирања, а која обухвата и ниже нивое агрегације

Један од начина је увођење пробабилистичког приступа [4.17] које, поред прогнозе очекиване вредности потрошње, даје и процену опсега њених могућих вредности за период прогнозирања (у виду квантила, интервала или густине расподеле). Са аспекта претходне анализе, код овог приступа се у моделовању ниво неодређености - процењује и доставља кориснику алата као информација која му користи у интерпретацији резултата прогнозе. Типичан пример примене код кога је пробабилистички приступ од посебног значаја је код

прогнозирања дневног максимума потрошње [4.18]. Прогнозе дневног максимума потрошње је критична у дистрибуцији и кориснику је значајна за многе апликације, као на пример, за задатке координисања електричних возила, DSM/DR програме, итд.

Друга стратегија је повећање тачности прогнозе проширењем модела додатним варијаблама. Овај приступ користи предности све већег броја мерења у интелигентним мрежама и могућност прецизнијег моделовања. Проширење се односи на додавање параметара спољашњих факторе потрошње, као на пример додавањем секундарних метео варијабли за софистицираније моделовање метео компоненте потрошње, [4.19] или додавање мерења са већих потрошача у домаћинству [4.20], итд. Међутим, иако се додавањем нових варијабли повећава вероватноћа побољшања тачности прогнозе, код развоја генералних приступа, проблем може представљати различит статус расположивости одговарајућих мерења за прогностичке области за које се користи алат.

Унапређења тачности прогнозирања које се може постићи постојећим моделима је препроцесирање и припрема података који се користе за моделовање тако да се повећа њихова предвидљивост, како је и предложено у [4.16]. Тамо је *a priori* анализа предвидљивости потрошње за различите агрегације потрошача показала да чак и једноставно диференцирање временске серије потрошње даје побољшање. Овај приступ се, међутим, типично, заснива на сегментацији или декомпозицији временске серије потрошње на подкомпоненте које су једноставније за моделовање и које се засебно моделују стандардним моделима. Најједноставнији пример, детаљно описан у раду [4.21] је декомпоновање потрошње на основу календарских променљивих. На пример, засебно се моделују потрошње које припадају различитим данима у седмици (7 модела), сатима у дану (24 модела), или за сате у дану у седмици (168 модела). У ову групу спада и подела на сатне подмоделе која је усвојена у тези, посебно значајна за индивидуалне и ниже агрегације потрошње. Поделом на сатне подмоделе се избегава моделовање комплексног дневног профила индивидуалне потрошње (слика 4.4). С друге стране, овим приступом је омогућена већа флексибилност урачунавања промена у факторима потрошње унутар дана, за различите сате. На даље, потрошња се може декомпоновати на бази фреквенцијског садржаја [4.22], коришћењем *wavelet* трансформације [4.23] или *empirical mode decomposition* код [4.24]. Поред ових метода код којих је подела временске серије потрошње унапред дефинисана, сегментација потрошње се може обавити и сходно конкретним подацима, као на пример, по сличности дневних профила (тзв. профилисање потрошње). Изведени профили се, типично, добијају аутоматски, најчешће, поступком кластеризације¹² и користе се за низ апликација у интелигентним мрежама, као и за прогнозирање потрошње на нивоу индивидуалног потрошача, [4.25], зграде [4.26], итд.

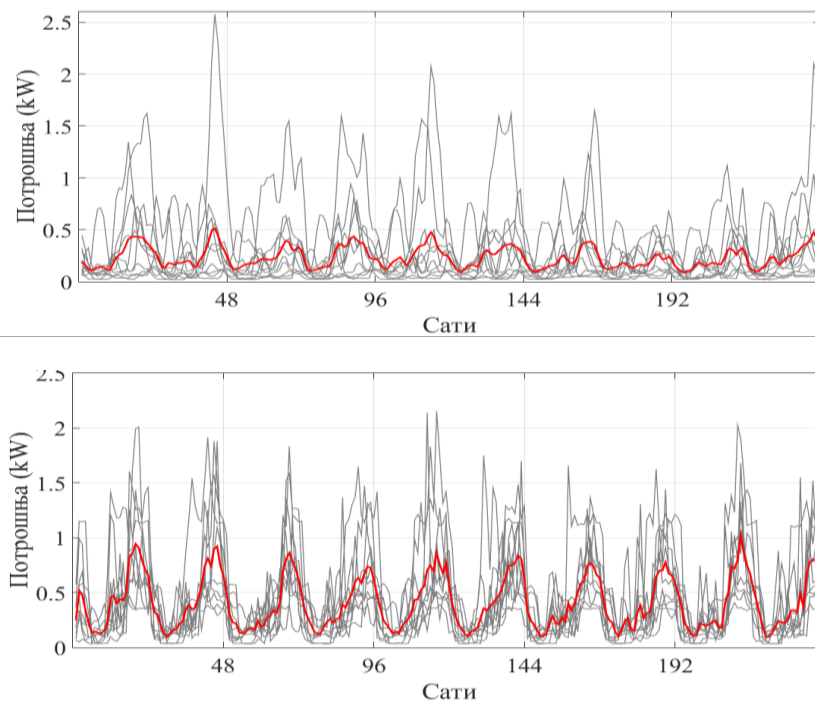
Такође, могуће је обавити профилисања потрошача, односно њихово груписање, потом, издвајање репрезента потрошње за сваку групу и моделовање на нивоу репрезента. Овај приступ директно користи ефекте агрегације за повећање предвидљивост профила где се репрезент најчешће моделује средњом потрошњом за сваку групу. Практично је интересантан, пре свега, јер уважава реалне околности где је број временских серија потрошње за које се развија *STLF* функционалност произвољно велики и препроцесирање се не обавља засебно за сваку временску серију. Друга предност је у томе што су издвојени репрезенти предвидљивији и могу се моделовати једноставнијим моделима. Коначно, велики број потрошњи за који се развија модел се своди на мањи број репрезентата, мања је

¹² Кластеризација је несупервизирана техника машинског учења за груписање података на основу унапред дефинисаног критеријума сличности

Карактеристике временске серије потрошње

хетерогеност података и, опет, једноставније моделовање. Иако карактеристичан за примену на индивидуалне потрошње, овај поступак се може користити и за потрошње на нивоу зграда, области, итд. Груписање потрошача се може обавити по области, региону, или на други начин, прописан од стране компаније. Међутим, на овај начин се, вероватније, добијају хетерогене групе потрошње, чији се профили могу значајно разликовати од репрезента, што проузрокује већу грешку прогнозирања у односу на индивидуалног потрошача унутар групе. За формирање хомогених група, код којих користи неки од аутоматских поступака груписања потрошача по сличности профила потрошње, најчешће, кластеризација [4.27], [4.28]. (За илустрацију, на слици 4.10, горе, су приказане полусатне вредности за потрошњу за 10 дана, за група од 10 произвољно одабраних потрошача и изведени репрезент (црвено), доле, за хомогену групу од 10 потрошача, добијену кластеризацијом (LCL подаци)). Одређивање броја кластера, такође, може бити условљено конкретним захтевима у пракси. На пример, у [4.27] је број кастомизираних уговара за ангажовање 6000 потрошача у DSM/DR програме унапред одређен да буде 30, тада се кластеризацијом потрошачи деле у 30 кластера, за које се издваја репрезент потрошње, за које се креирају услуге. Даља унапређење је да се број кластера оптимизује по критеријуму квалитета груписања потрошача, као што је у [4.28], итд. Описани приступи развоја функционалности прогнозе потрошње великог броја индивидуалних потрошача, иако једноставан за примену и користан за низ апликација од значаја у интелигентним мрежама, генерише прогнозе, ипак, само *на нивоу репрезента*.

Алтернатива која омогућава побољшање тачности прогнозе на нивоу индивидуалне потрошње, је новијег датума, тзв. *pooling* метода, где се истражује могућност развоја глобалног модела, који се обучава над сетом потрошњи, покупљених са различитих места. Обучавањем модела за више потрошњи уједно се инхерентно урачунава додатни фактор потрошње - просторна корелација између потрошњи из различитих извора и за неке технике



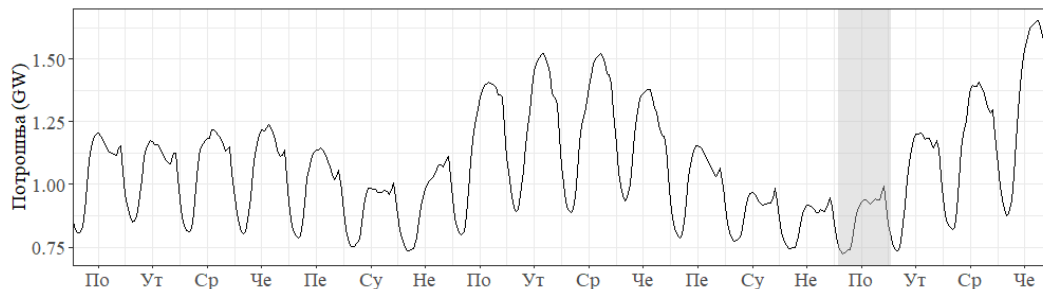
Слика 4.10: Пример хетерогене (горе) и хомогене (доле) групе од 10 LCL индивидуалних потрошача

моделовања, примарно, дубоке неуралне мреже, превенира преобучавање. Прва публикација у којој се предлаже *pooling* метода је референца [4.29], из 2017. године, где се дубока неурална мрежа обучава за потрошње 25 произвољно одабраних потрошача, унапређено у [4.30], применом за хомогену групу потрошача, добијену кластеризацијом. Варијанте примене *pooling* методе за различите моделе су дате у публикацијама [4.31] и [4.32].

4.5 Одступање од регуларних профила потрошње и ефекат специјалних дана

Претходно описани основни фактори дефинишу модел регуларних профила потрошње. Међутим, у току године постоје дани када долази до очекиваних или неочекиваних одступања од регуларног понашања потрошње, за тзв. специјални дани. Типичан пример очекиваних специјалних дана су профили потрошње током празника. Одступања од регуларне потрошње током ових дана, се генерално рефлектују као промене нивоа потрошње за одређене (или све) сате током дана, што уноси промене у карактеристичан дневни профил потрошње за тај дан и седмични профил за ту седмицу.

Као илустрација, слика 4.11 приказује сатну потрошњу за период од 13.05 до 30.05.2013. за NYC зону (NY). На слици је засенчен 27. мај када је Дан сећања (*Memorial Day*) у Њујорку, који је био понедељак. Уместо датума, приказани су дани у седмици да би се лакше уочио ефекат специјалних дана на седмични и дневни профил. У поређењу са претходна два понедељка, профил потрошње показује знатно нижи ниво потрошње, са измењеном профилом, који има већу сличност са претходна два дана викенда. Модел који не урачунава ефекат празника ће, очекивано, направити значајну грешку у прогнозама за овај дан.



Слика 4.11: Ефекат специјалних дана на примеру *Дана сећања* за NYC зону

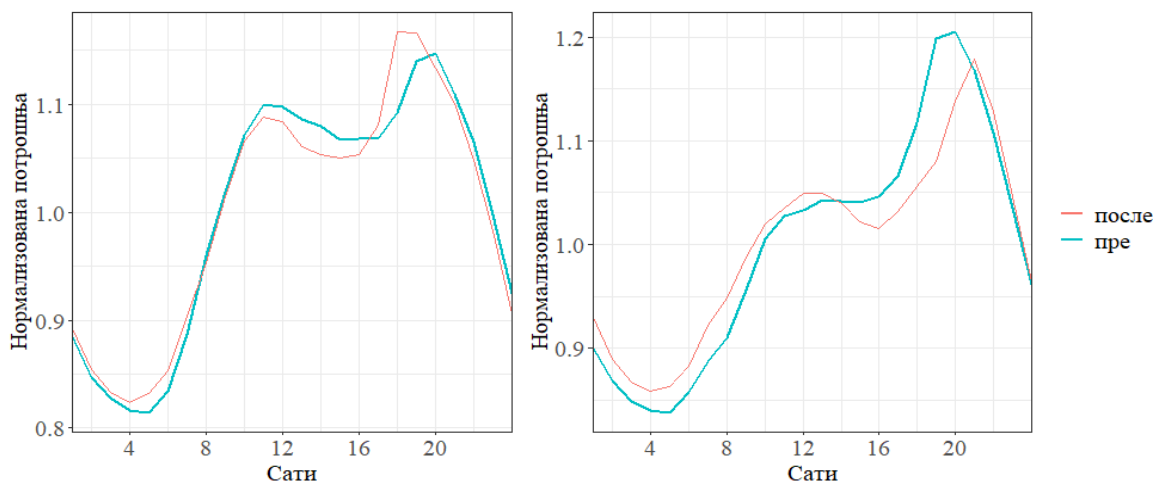
За већину система, за празнике је смањена потрошња што се, у неким случајевима, успешно моделује кад дан викенда, док се за одређене прогностичке области (на пр., туристичке) бележи, управо, њено повећање [4.33]. С друге стране, за неке системе, број специјалних дана у току године је значајан, са комплекснијим профилима потрошње, као на пример у земљама Средњег истока за време верских празника [4.34]. С обзиром на различитости и специфичности профила потрошње за празнике и за различите прогностичке области, разликују се и начини њиховог моделовања и пример детаљне студије која обухвата 64 различите методологије, дат је у [4.33]. Најједноставнији приступ, који не урачунава специфичности празника је увођење категорије регуларан дан/празник. На даље је могуће повећати број категорија поделом празника државни/верски празници, или засебаног моделовања сваког празника. Међутим, већи број категорија за тип дана у години, повећава и потребну базу података за њихово моделовање и чешће се користе једноставнија решења која имају за циљ само да *умање* нерегуларности потрошње за специјалне дане. У тези је ефекат специјалних дана моделован категоричком варијаблом, SD , за 4 категорије: регуларан

Карактеристике временске серије потрошње

дан након регуларног дана, специјални дан након регуларног дана, итд. Увођењем променљиве SD моделују се просечне промене нивоа потрошње за ове дане. Додавањем и члана интеракције са потрошњом из претходног дана, $SD f^{-1}(P_h^{-1})$ уводи се корекција утицаја претходних вредности потрошње за различите категорије дана, за сваку сатну потрошњу.

4.6 Daylight saving

Daylight saving time (DST), односно, прелазак на летње/зимско рачунање времена има за циљ уштеду електричне енергије која се користи за осветљење ефикасним искоришћењем дневне светлости - померањем једног сата унапред у пролеће и враћањем сата уназад на јесен. У Сједињеним Америчким Државама, за област за коју се користе тест подаци, DST траје 238 дана, које почиње прескакањем 2. сата, друге недеље у марту и враћа се на зимско рачунање времена прве недеље новембра. У пракси је уобичајено да се у базама података у којима се прикупљају подаци за потрошњу, интерполира вредност за недостајући 2. сат у марту са суседним вредностима из 1. и 3. сата, и усредњи вредност дуплог 2. сата у новембру, како би се обезбедио континуитет и интегритет временске серије потрошње који је битан за несметано обучавање и прогнозирање. Међутим, осим очувања 24-сатног дневног профила у оперативном раду, саме промене потрошње које уноси DST ефекат су ретко разматране. Скорије истраживање у референци [4.35] је показало да се детаљнијом анализом ових промена и укључивањем DST ефекта у прогностички модел добија побољшање тачности прогнозе. На слици 4.12 су приказани нормализовани профили потрошње за недељу пре и после промене времена у марту (лево) и у новембру (десно) за NYC зону, где се уочавају измене дневног дијаграма потрошње, поготово, померање дневног максимума. Ове промене су резултат релативног помераја компоненте потрошње која зависи од сати у дану (радно време, дневне обавезе) у односу на обданицу и за њихово моделовање је могуће користити додатне варијабле, као на пример, времена изласка и заласка сунца као што је предложено у [4.35]. За дату прогностичку област, DST ефекат зависи од доба године и дана у седмици и уноси измене у годишње и седмичне профиле потрошње, како је констатовано у поменутом раду.



Слика 4.12: DST ефекат за 2013. годину за NYC зону

Карактеристике временске серије потрошње

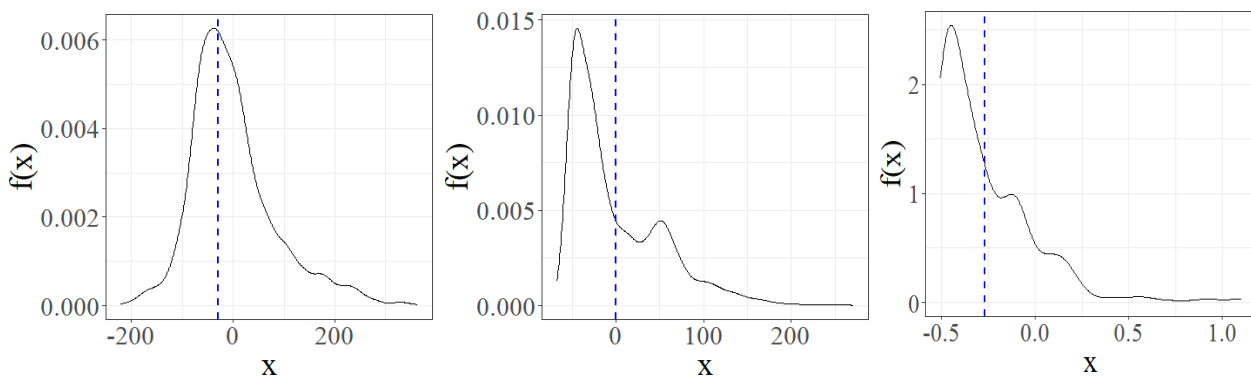
Један од могућих начина моделовања DST ефекта у ГАМ-у, без коришћења спољашњих променљивих, а који је усвојен у тези, је моделовање DST ефекта као разлике између летњих и зимских профила. У ту сврху се уводи бинарна променљива DS , чија је вредност 1 за период летњег (или зимског) рачунања времена, иначе је 0, и моделу додају чланови $DSf_g(x_g)$, and $DSf_s(x_s)$, (члан 3, Табела 3.1). На тај начин се, на основу података, DST ефекат моделује као нелинерна модификација годишњег и седмичног профила.

4.7 Компонента тренда потрошње

Дугорочни тренд је споропроменљива компонента потрошње, чији су узроци глобалне промене у навикама и броју потрошача или услед утицаја спољашњих фактора током дужег временског периода. Начин на који се тренд моделује зависи од података који се користе за обучавање модела, односно, дужине клизајућег прозора, N у процедури Процедуре 1 и расположивих мерења. На пример, за моделе који се обучавају на подацима који обухватају период од више година, тренд представља вишегодишњу промену нивоа потрошње које су, типично, демографске или економске природе и, уколико су параметри којима се ове промене мере расположиви, могу се искористити за дефинисање додатних ГАМ члан(ова) модела. С друге стране, код модела који се обучавају на краћим временским интервалима, унутар године (на пример, неколико месеци), тренд је локална промена нивоа потрошње, која је, најчешће, део годишњег профила. У генералном случају кад узроци тренда нису идентификовани или не постоје одговарајућа мерења, ова компонента се може моделовати као функција од времена. Типична реализација у ГАМ-у је коришћењем полинома или нелинеарном глатком функцијом $f(trend)$, где је $trend = 1:N$, [4.36].

4.8 Карактеристике расподеле временске серије потрошње

На слици 4.13 су приказане функције густине расподеле временских серија за а) агрегирану потрошњу (пример NYC), б) EN1 зграду и ц) индивидуалног потрошача (произвољан LCL потрошач). За агрегиране потрошње расподела је блиска нормалној, док је за локалне потрошње несиметрична и померена улево (*right-skewed*). Генерално, потрошња индивидуалног потрошача је, како се види на слици 4.4, већину дана на нижем нивоу, са повременим скоковима услед повременог коришћења појединих уређаја у домаћинству, што резултира значајном несиметријом њене расподеле. По Централној граничној теорему, са растом броја потрошача, расподела потрошње конвергира ка нормалној.



Слика 4.13: Приказ расподеле потрошње за а), б) и ц)

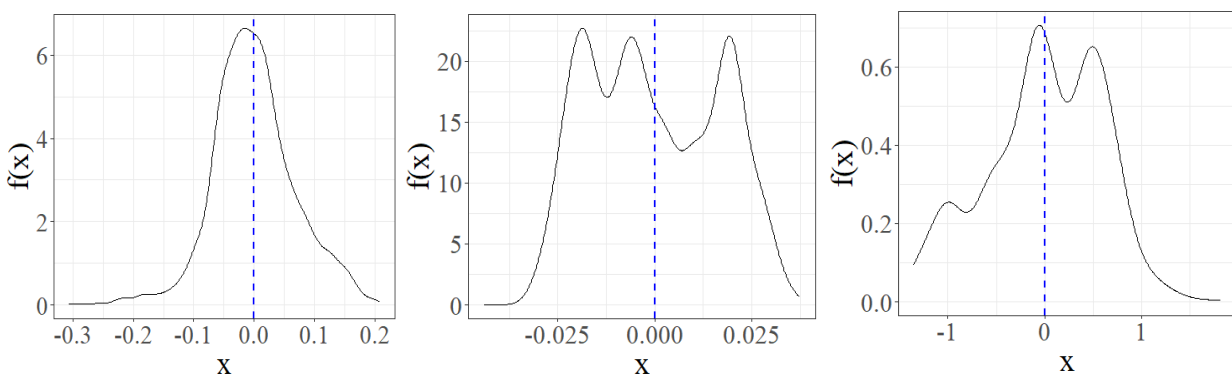
Карактеристике временске серије потрошње

ГАМ, по дефиницији, омогућава моделовање променљивих са расподелом из шире фамилије експоненцијалних функција (поглавље 3.1). Међутим, како су алгоритми за обучавање и његова унапређења прилагођенија нормалној расподели (обучавање је знатно брже, боља је покривеност потребним алгоритмима у R -у), у тези се адитивна структура из 3.1 моделује за претпостављену нормалну расподелу излазне променљиве. У том случају, решење је да се, у фази препроцесирања обави трансформација потрошње како би се постигла расподела која је ближа Гаусовској. За агрегиране потрошње, најчешће, се користи логаритамска трансформација, али су у употреби и квадратни, кубни корен, којом се жели постићи боља симетричност расподеле. Међутим, као генерализација овог поступка, применљива за различите потрошње и агрегације, користи се *Box-Cox* трансформација [4.37] дефинисана формулом 4.4:

$$y_T = \begin{cases} \log(y) & \text{за } \lambda = 0 \\ \frac{y^\lambda - 1}{\lambda} & \text{за } \lambda \neq 0 \end{cases} \quad 4.4$$

У 4.4 λ је контролни параметар за чије се карактеристичне вредности добијају стандардне трансформације, на пример, за вредност 0 је $\ln(y)$, за 1 уноси се померај средње вредности података без значајније трансформације, $\frac{1}{2}$ квадратни корен и линеарна трансформација, итд. Проблем одређивања контролног параметра је у томе што се оптимална вредност разликује за различите потрошње и за сваку би било потребно посебно је одредити. Зато се за практичну примену и рад са више података, користи аутоматизовани поступак одређивања оптималне вредности контролног параметра трансформације, који се једноставно имплементира у процедуру моделовања за потрошње, имплементиран у R пакету `forecast::lambda.BoxCox`.

Сликом 4.14 су приказане густина расподеле потрошњи са претходне слике након трансформације, где је евидентно побољшање у симетричности расподеле. Добијене вредности за -0.2, -0.7, 0.1 респективно.



Слика 4.9:

Слика 4.14: Приказ расподеле трансформисане потрошње за а), б) и ц)

5. Резултати

За тестирање предложене методологије искоришчени су подаци из прикупљених база података, описаних у поглављу 2, којим су обухваћене потрошње на различитим нивоима агрегације и за различита места у електроенергетском систему и различите прогностичке задатке. Сваки сет података служи развоју засебне функционалности прогнозе, (за градске зоне, за зграде, на нивоу трафо-рејона и индивидуалног потрошача) и за сваки од примера дефинише се каталог ГАМ чланова. У дефинисању каталога су уважени предлози из поглавља 4 и поступак прогнозирања се заснива на процедурама, описаним у поглављу 3.5.

5.1 Верификација модела и предиктивне перформансе

Верификације модела има за циљ проверу адекватности одабраног каталога ГАМ чланова у моделовању дате потрошње. Модели се, стандардно, верификују анализом резидуала модела, односно остатка од моделовања временске серије за период естимације. Са слике 3.7 из подпоглавља 3.5 резидуал модела естимираног у тренутку t , на бази N претходних вредности потрошње, $y_{t-N+1:t}$, које је проценио вредностима $\hat{y}_{t-N+1:t}$, дефинисан је секвенцом $e_{t-N+1:t}$, по 5.1.

$$e_{t-N+1:t} = y_{t-N+1:t} - \hat{y}_{t-N+1:t} \quad 5.1$$

Код коректно моделоване временске серије, узорци секвенце $e_{t-N+1:t}$ задовољавају карактеристике случајне грешке:

1. равномерно су распоређени око нулте вредности, без тренда и значајних промена нивоа
2. међусобно су независни, без остатака корелације
3. имају константну варијансу
4. подлежу Гаусовској раподели

Прогностички метод који задовољава услове 1. и 2. урачунава довољно информација потребних за моделовање тако да се не уноси отклон у прогнозу и сваки предиктивни модел чији резидуал не задовољава ове особине може бити унапређен [5.1]. Услови 3. и 4. утичу на очекивани предиктивни интервал грешке прогнозе, и, често, модел који не задовољава ове услове не може бити даље унапређен. У реалним условима где може постојати проблем недоступности података или немогућности идентификовања компоненти потрошње или њених промена, анализом ових особина се показује да ли се са расположивим подацима и усвојеним приступом може добити модел, довољно добар за предиктивни задатак. Провера резидуала се обавља, превасходно, графички, на временској оси и анализом корелограма.

Код генерализованих модела се, такође, као мера прилагођености подацима (*goodness-of-fit*) користи тзв. описана девијанса¹³, односно, *deviance explained*, DE , која, у процентима, показује колико добро актуелни модел описује податке у односу на тзв. засићени модел (*saturated model*) [5.2]. Засићени модел је модел интерполације, са бројем параметара (кофицијената модела) колико има и узорака у подацима. Тада се, за дате податке, за максималну вредност функције веродостојности естимираног посматраног модела, l и оног засићеног, l_s , девијанса дефинише формулом 5.2.

¹³ Термин је слободан превод аутора.

$$D \equiv 2(l - l_s) \quad 5.2$$

Као референтна вредност за израчунавање описане девијансе се користи модел са нултом девијансом, D_0 , односно, модел који моделује само средњу вредност података без осталих предиктора. Тада се DE дефинише по формули 5.3.

$$DE = 1 - \frac{D}{D_0} \% \quad 5.3$$

По 5.3, DE вредност модела који садржи само константу је 0%, док је за засићени модел који у потпуности описује податке, 100%. DE се користи за поређење модела и погодан је за праћење унапређења ГАМ-а додавањем нових чланова и процену њиховог доприноса у описивању моделованих података.

У оквиру процедуре валидације ГАМ модела, анализирају се и парцијални резидуали сваког ГАМ члана, као и добијени ефективни степени слободе (*effective degrees of freedom*, EDF). Циљ ове анализе је да се констатује да ли је број базисних функција довољан за моделовање појединачног члана модела. Основни поступак подразумева визуелну инспекцију резидуала око процењене вредности ГАМ члана, а потом и проверу EDF вредности. EDF је мера комплексности ГАМ нелинеарности, која у ГАМ-у зависи од броја базисних функција и од параметра изглађивања, λ : највећу вредност има за непенализовани сплајн и опада са растом пенализације λ . Ефективни степени слободе ГАМ члана је, генерално, нецелобројна вредност и при анализи резидуала, процењује се колико је блиска задатом броју базисних функција. Једноставно правило за процену правилне димензионисаности ГАМ члана је да уколико се EDF вредност значајно приближи задатом броју базисних функција, могуће је да треба повећати број базиса. (Провера EDF -а се добија функцијом `gam.summary`, у `mgcv` пакету).

Предиктивне перформансе модела су тестиране директном применом Процедуре 1, из поглавља 3.5, за унапред познате вредности потрошње и температуре за период тестирања $t + 1, t + 2, \dots$. Тада се са почетка интервала одвоји N вредности потрошње и свих улаза модела који се искористе за естимацију модела и прогнозу наредне вредности. Процедура 1 се у времену помера за један корак унапред и понавља се претходни поступак, и тако на даље, до краја интервала предвиђеног за тестирање. За периода тестирања $t + k: t + N$, квалитет прогнозе се мери грешком између реалне вредности, y_{t+k} , и њене прогнозе, \hat{y}_{t+k} , изражене стандардним критеријумима, *mean absolute error* (MAE) и *mean absolute percentage error* (MAPE), дефинисаних формулама 5.4. и 5.5.

$$MAE = \frac{1}{N_t} \sum_{k=1}^{N_t} |y_{t+k} - \hat{y}_{t+k}| \quad 5.4$$

$$MAPE = 100 \frac{1}{N_t} \sum_{k=1}^{N_t} \frac{|y_{t+k} - \hat{y}_{t+k}|}{y_{t+k}} \quad 5.5$$

5.2 Пример потрошње већих градских зона – NY база података

Потрошња електричне енергије већих градских зона из *NYISO* базе података, са минималним бројем потрошача од око 80000 и средњом потрошњом од 330 MW, има карактеристике агрегиране потрошње, са релативно, предвидљивим профилима. Међутим, разлике у саставу потрошача, начину коришћења електричне енергије и метеорологији које постоје између 11 зона, чине ове профиле довољно хетерогеним за евалуацију предложене методологије за реалан сценарио прогнозирања.

Почетне поставке које се на даље користе су:

- За естимацију *NYISO* модела кориштени су подаци за потрошњу и температуре од 4 године, што значи да је ширина прозора за Процедуру 1, за сваки сатни подмодел $N = 365 * 4 + 1$ узорака. Дакле, од расположивих података који обухватају период од 2009 до 2013 године, прве четири године су иницијално издвојене за естимацију сатних модела и генерисање прогноза за 01.01.2013 и прорачун грешака 5.4 и 5.5. На даље се клизајући прозор од 4 године, сукцесивно, помера за један сат унапред и врши прогноза, све до краја 2013, како је описано у 3.5. Резултати тестирања су приказани средњим вредностима добијене MAPE/MAE грешке.
- Засебно се моделују потрошње које одговарају различитим сатима у дану и моделује се њихова логаритамска трансформација, као прихватљиво решење за агрегиране потрошње.

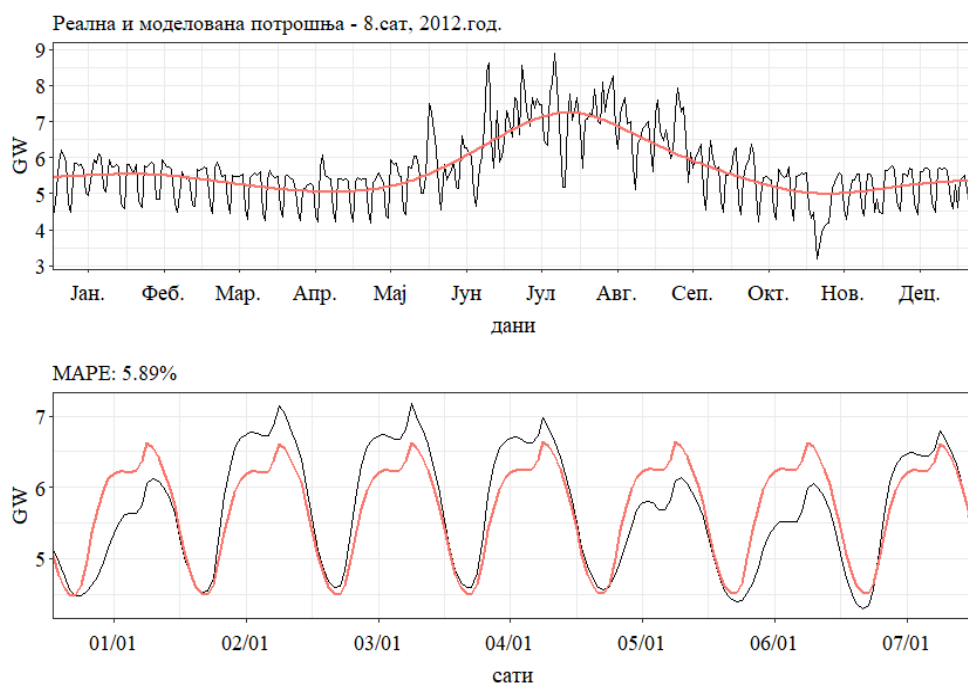
Први корак у примени Процедуре 1 за прогнозирање *NYISO* потрошњи је дефинисање иницијалног скупа ГАМ чланова (подпоглавља 3.5). У наставку је приказан поступак у коме се прате унапређења модела додавањем појединачних ГАМ чланова који су предложени у поглављу 4 и дефинисања коначног скупа, довољног да опише све *NYISO* потрошње. Резултати моделовања се посматрају на примеру потрошње која припада 8. сату у дану, а резултати прогнозе за тест период од 01.01.- 07.01.2013.

Под претпоставком да вредност потрошње у сваком сату, P , зависи само од дана у години, тада се трансформисана временска серија сатне потрошње, $\ln P$, моделује само годишњим профилем, коришћењем цикличног кубног сплајна као што је предложено у поглављу 4, по моделу **M1**:

$$\mathbf{M1:} \quad \ln P = f_g(g)$$

На слици 5.2.1 је црвеном бојом приказан моделовани годишњи профил временске серије потрошње која припада 8. сату, за који је $DE = 43.5\%$. На доњем графику су дати резултати прогнозе **M1** модела и MAPE резултат за цео тест период. Оно што се уочава је да се засебним моделовањем само годишњих промена сатних потрошњи добијају базични дневни профили NYC потрошње.

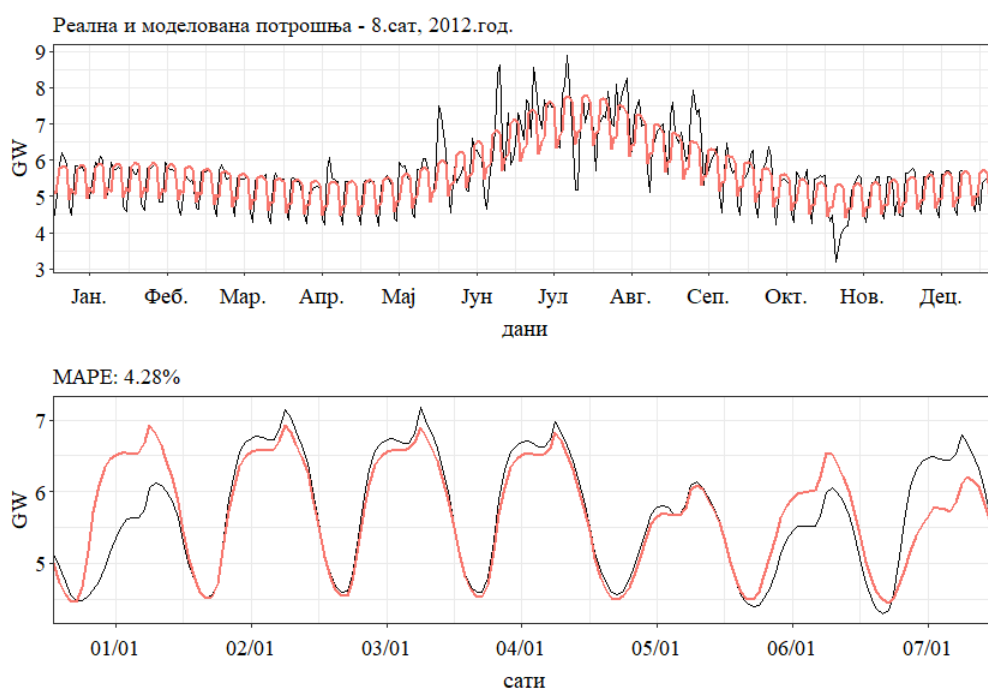
Резултати



Слика 5.2.1: Резултати моделовања моделом М1

За урачунавање компоненте потрошње која зависи и од дана у седмици, моделу **М1** се додаје члан који моделује седмични профил потрошње $f_s(s)$. Промене седмичног профила током године се моделују чланом интеракције између годишњег и седмичног профила, $f_{g,s}(g, s)$. Овим проширењем модела *DE* се повећало на 76.4%. Резултати моделовања потрошње 8. сата и прогнозе за прву седмицу 2013. године су приказани сликом 5.2.2.

$$\mathbf{M2}: \quad \ln P = \mathbf{M1} + f_s(s) + f_{g,s}(g, s)$$



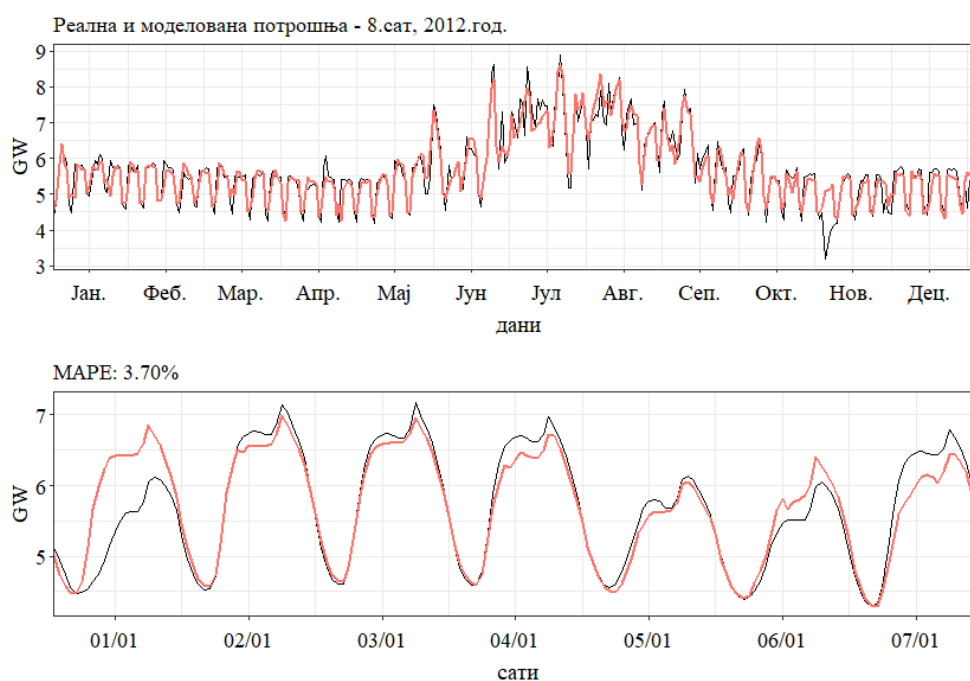
Слика 5.2.2: Резултати моделовања моделом М2

Резултати

За моделовање утицаја температуре ваздуха на потрошњу, модел M2 је проширен групом чланова, за моделовање индивидуалног ефекта температуре $f_{T_x}(T_x)$, где је T_x расположива вредност температуре (средња, максимална или минимална дневна, средња сатна) и интеракција са другим променљивим y .

$$\mathbf{M3}: \quad \ln P = \mathbf{M2} + \sum f_{T_x}(T_x) + \sum f_{T_x,y}(T_x, y)$$

За расположиве вредности максималне и минималне дневне температуре T_{mx} и T_{mn} за NYISO податке, ефекат на потрошњу је у потпуности урачунат члановима који описују засебни утицај, $f_{T_{mx}}(T_{mx})$ и $f_{T_{mn}}(T_{mn})$, реализовани *TPRS* сплајновима и чланом њихове интеракције $f_{T_{mx},T_{mn}}(T_{mx}, T_{mn})$. За урачунавање промена утицаја температура током седмице и године додају се чланови $f_{T_x,g}(T_x, g)$ и $f_{T_x,s}(T_x, s)$, где је T_x максимална/минимална температура. Утицај температура из претходних дана се моделује истим скупом ГАМ чланова са температурама из претходна два дана. Резултати моделовања на слици 5.2.3. показују доминантан утицај температуре у односу на седмичну сезоналност у летњим месецима. Описана девијанса **M3** модела за 8. сат је $DE = 88.3\%$.

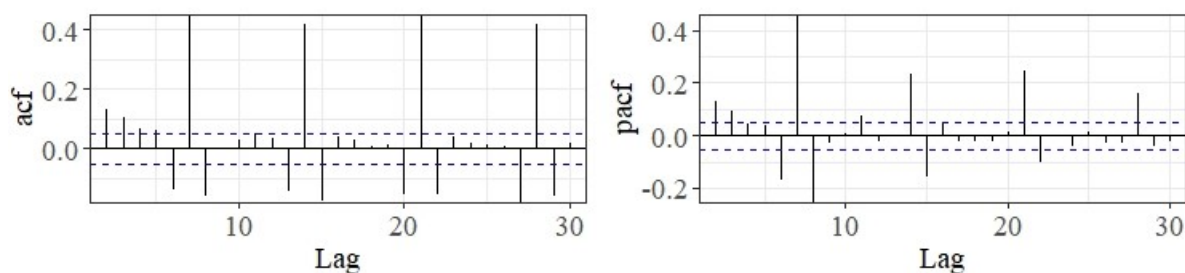


Слика 5.2.3: Резултати моделовања моделом M3

На слици 5.2.4. је приказан корелограм резидуала **M3** модела за 8. сат, са интервалима поверења (95%), приказаним испрекиданим линијама. Вредности изван интервала поверења указују на постојање немоделоване аутокорелације у резидуалу. Како би се моделом обухватила и ова компонента потрошње, додати су чланови са потрошњама из претходних 7 дана (поглавље 4.3).

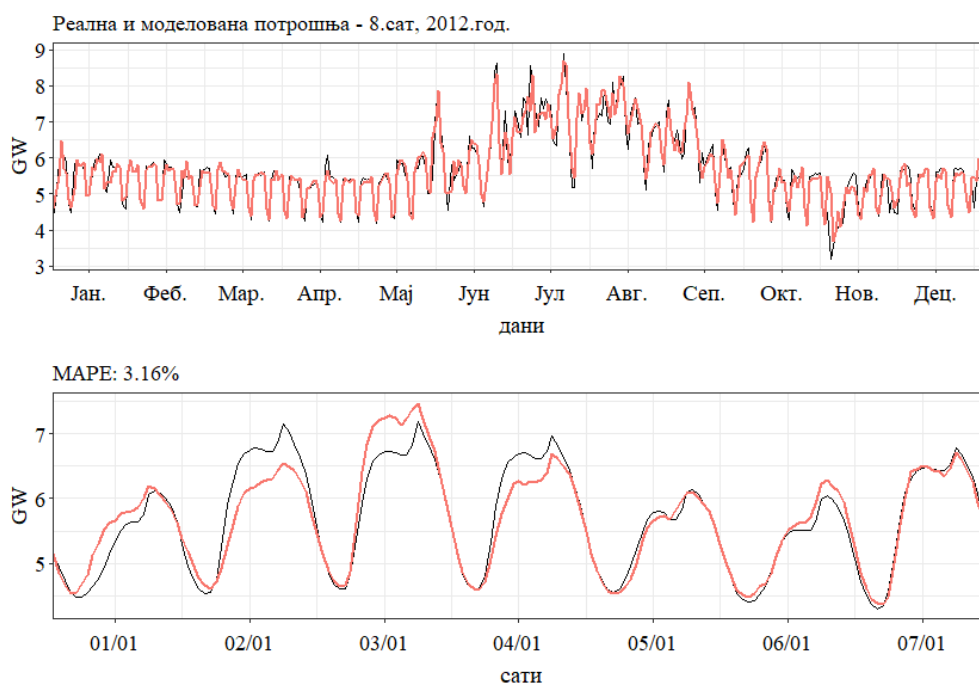
$$\mathbf{M4}: \quad \ln P = \mathbf{M3} + \sum_{d=-7}^{-1} f^d(P^d)$$

Резултати

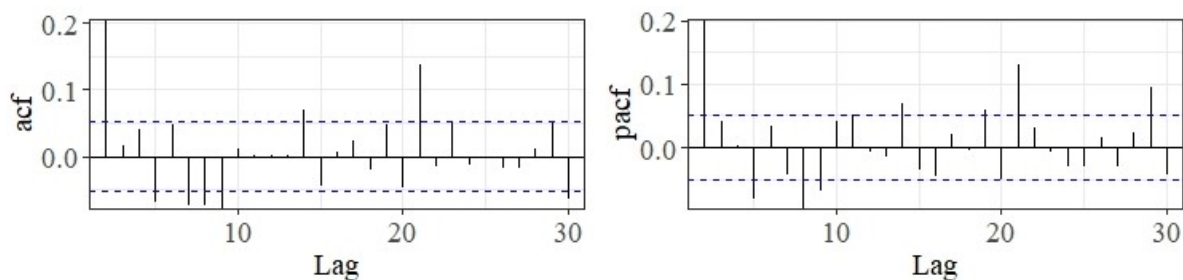


Слика 5.2.4: Аутокорељација/парцијална корељација резидуала М3 модела за 8.сат

Додавањем нових ГАМ чланова, вредност DE се повећала на 93.9%, укупна грешка прогнозе смањила за око 17% и резултати су приказани сликом 5.2.5. Корелограм резидуала новог **M4** модела за 8. сат, дат сликом 5.2.6, показује значајно побољшање. Провером корелограма резидуала модела за све сатне потрошње констатовано је да остаци немоделоване аутокорељације постоје за 6, 8, 18 и 19 сат.



Слика 5.2.5: Резултати моделовања моделом М4



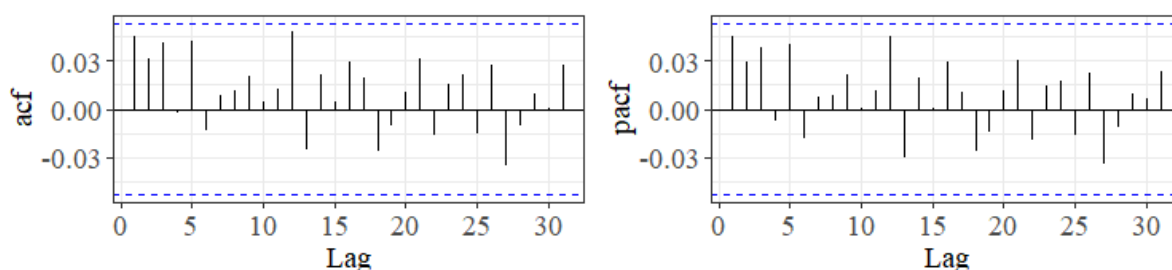
Слика 5.2.6: Аутокорељација/парцијална корељација резидуала М4 модела за 8.сат

Резултати

Урачунавање *DST* ефекта је спроведено као што је описано у поглављу 4.6, тј., додавањем чланова промена седмичног и годишњег профила, контролисаних категоријом варијаблом *DS*.

$$\mathbf{M5}: \quad \ln P = \mathbf{M4} + DS f_g(g) + DS f_s(s)$$

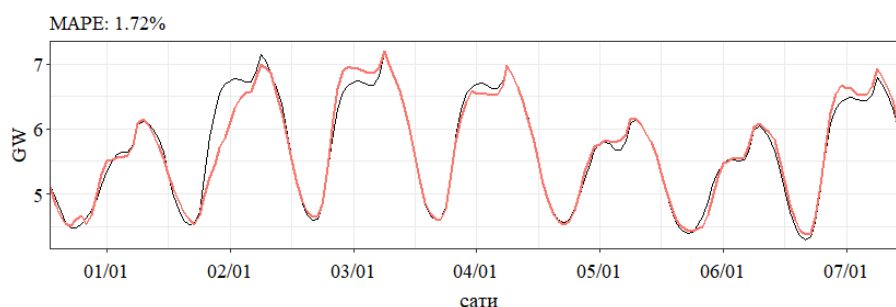
Промене урачунаог *DST* ефекта су најзначајније у седмицама након њихове имплементације и *MAPE* грешка прогнозе за седмице након преласка са зимског на летње рачунање времена и назад, за 2013. годину је смањена за 11%. Интересантан резултат је да су додавањем *DST* члана у потпуности уклоњени остаци аутокорељације у резидуалу у односу на **M4** модел, како се види на слици 5.2.7. Померање компоненте потрошње која иде на осветљење услед *DST* ефекта уноси додатну корелацију у временску серију потрошње, поготово за потрошње у сатима изласка и заласка сунца (за NYISO области, 6-9. сат ујутру, 18.-20. сат увече), која је остала немоделована у **M4** моделу. Уз смањење грешке прогнозе, добијени резултат је прихваћен као потврда добро моделованог *DST* ефекта.



Слика 5.2.7: Аутокорељација/парцијална корелација резидуала **M5** модела за 8. сат

За одабране датуме празника, **M5** моделу су додати чланови за моделовање специјалних дана, предложени у подпоглављу 4.5.

$$\mathbf{M6}: \quad \ln P = \mathbf{M5} + SD + SD f^{-1}(P^{-1})$$



Слика 5.2.8: Резултати моделовања моделом **M6**

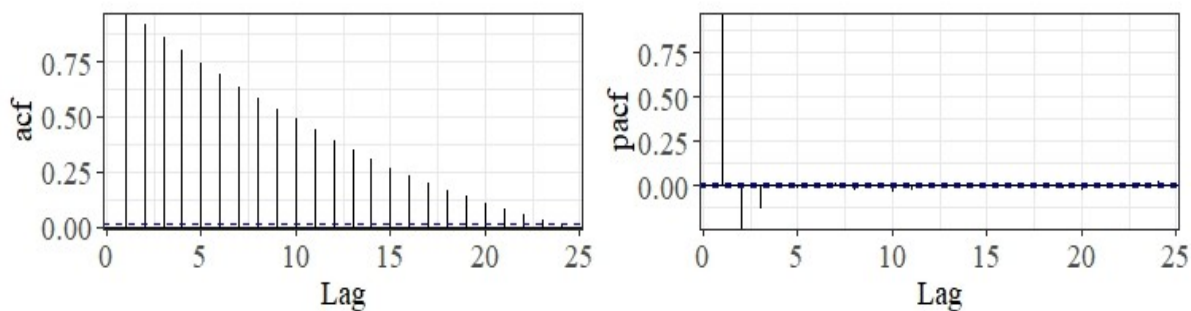
Увођењем ефекта специјалних дана у модел *MAPE* грешка новог **M6** модела се смањила за 13%, за посматрани тест период. За 1. јануар, који је специјалан дан, *MAPE* вредност у моделу **M5**, је била 3.8%, док је у моделу са урачунатим ефектом специјалних дана, **M6**, 1.3%, што је значајно унапређење. Резултати су приказани сликом 5.2.8.

M6 је модел сатне потрошње који не урачунава утицај потрошњи из претходних сати на актуелну вредност. На слици 5.2.9 приказан је корелограм укупног резидуала за NYC зону, добијен комбиновањем резидуала **M6** модела за сваки сат, који показује значајну

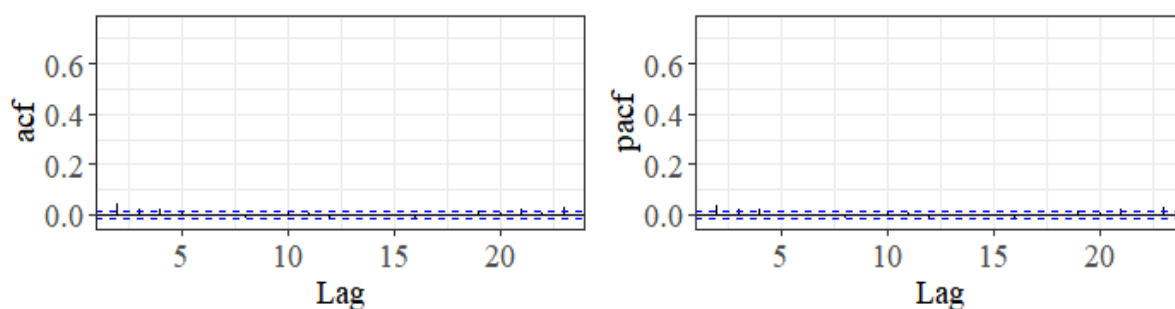
Резултати

корелисаност између сатних потрошњи. На основу предлога из 4.4, **M6** је проширен члановима са потрошњама из претходна два сата (модел **M7**), што је довело до побољшања, приказано сликом 5.2.10. Описана девијанса **M7** модела потрошње за 8. сат је $DE = 98.7\%$, $MARE$ грешка за тест период је била 0.64% .

$$\mathbf{M7}: \quad \ln P = \mathbf{M6} + \sum_{h=-2}^{-1} f_h(P_h)$$



Слика 5.2.9: Аутокорелација/парцијална корелација резидуала M6 укупне потрошње



Слика 5.2.10: Аутокорелација/парцијална корелација резидуала M7 укупне потрошње

На основу резултата, датих примером 8.сата, за NYC зону (типични за све сатне потрошње из NYISO базе података), дефинисан је каталог ГАМ чланова који је усвојен за моделовање сатних потрошњи из NYISO податке, дат у Табели 5.2.1.

5.2.1 Резултати анализе резидуала за NYISO потрошње

Модел дефинисан каталогом ГАМ чланова из Табеле 5.2.1 је естимиран за сваку од $11 \times 24 = 264$ сатних временских серија потрошње из NYISO сета. Анализирани су остаци од моделовања за период од 2009 до 2013 године. Резултати су приказани сликама 5.2.11-12, на примерима 2. и 15. сата, за NYC зону, као репрезентативни и за остале временске серије. Временски графици на сликама показују да су вредности резидуала махом равномерно распоређене у односу на нулту вредност (средње вредности $< e^{-14}$), без уочљивог тренда или промена у нивоу, али са спорадичним и израженим аутлејерима¹⁴.

Слике доле, лево показују да нема остатака аутокорелација, која се за све временске серије налазила у опсегу интервала поверења. За већину временских серија (~95%) временски график указује на релативно константну варијансу резидуала (пример 2. сата,

¹⁴ Аутлејер - нестандартна обсервација,

Резултати

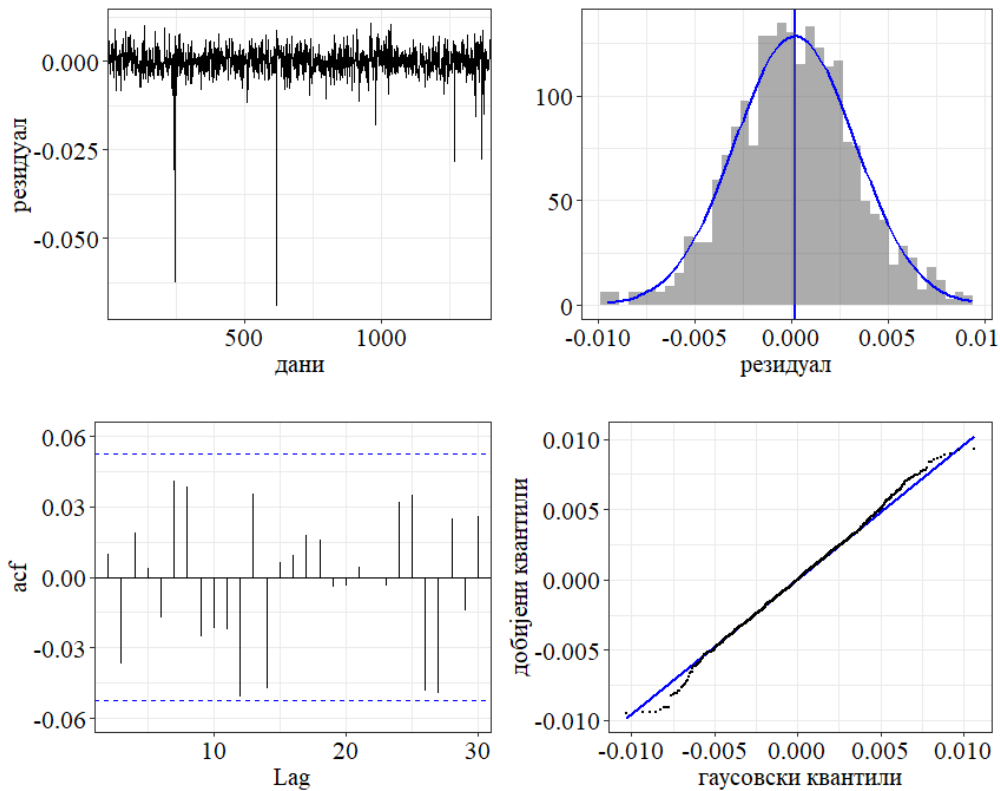
лево), док је за 14.-16. сат, за урбане зоне (NYC, LONGIL, WEST и GENESE) она повећана током летњих месеци (пример слике 15. сата).

Табела 5.2.1: Каталог ГАМ чланова потрошње за NYISO зоне

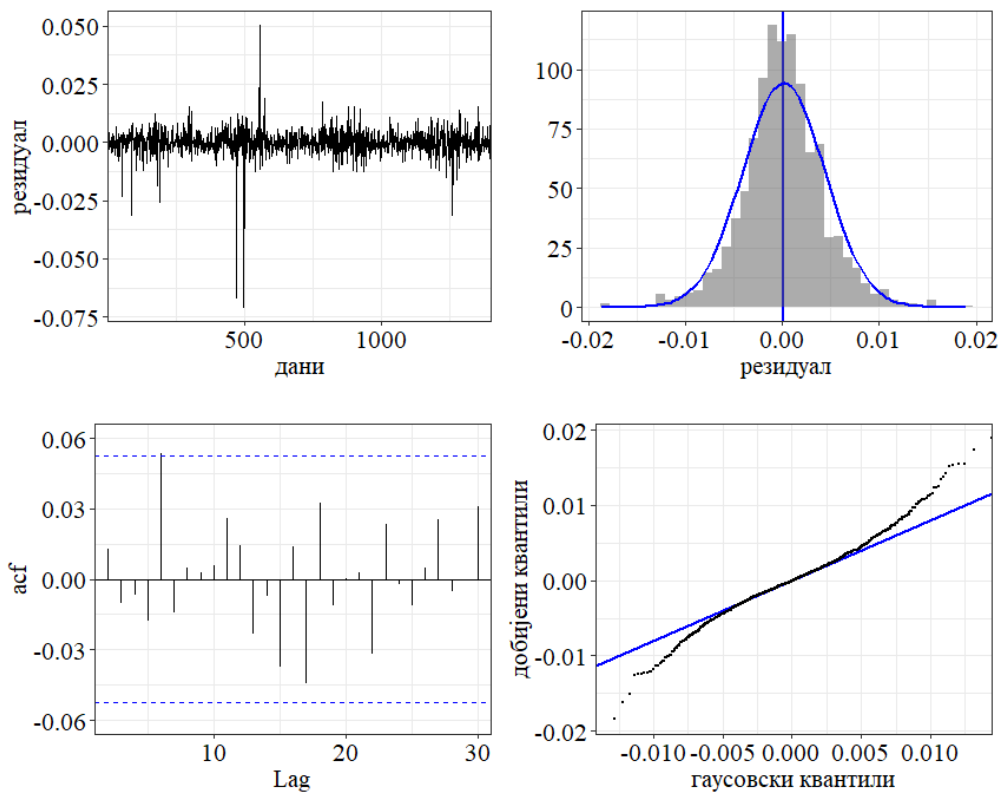
ГАМ члан	Ефекат потрошње
$f_t(trend)$	Компонента тренда
$f_g(g)$	Годишњи профил
$f_s(s)$	Седмични профил
$f_{s,g}(s, g)$	Промене седмичног профила током године
$DS f_g(g)$	Промене годишњег профила услед <i>DST</i>
$DS f_s(s)$	Промене седмичног профила услед <i>DST</i>
$f_{T_x}(T_x^d)$	Зависност од температуре x (минимална/максимална) у дану $d = 0, -1, -2$
$f_{mx, mn}(T_{max}^d, T_{min}^d)$	Интеракција максималне/минималне температуре за дане $d = 0, -1, -2$
$f_{T_x, s}(T_x^d, s)$	Промене температуре x (минимална/максимална) током седмице за дане $d = 0, -1, -2$
$f_{T_x, g}(T_x^d, g)$	Промене температуре x (минимална/максимална) током године за дане $d = 0, -1, -2$
SD	Промена нивоа потрошње за тип дана који урачунава ефекат специјалних дана
$SD f^{-1}(P^{-1})$	Утицај вредности потрошње из претходног дана, зависно од типа дана
$f^d(P^d)$	Утицај вредности потрошње за дане $d = -2, \dots, -7$
$f_h(P_h)$	Утицај вредности потрошње за сате, $h = -1, -2$

Како би се ближе испитале карактеристике расподеле резидуала, визуелно уочљиве екстремне вредности су ручно отклоњене. За резидуал који одговара моделу 2. сата је отклоњено 9 узорака (245, 511, 616, 919, 980, 1266, 1344, 1367, 1375), док је за 15. сат уклоњено њих 11 (53, 86, 180, 191, 470, 498, 555, 572, 786, 1258, 1263). Хистограми на сликама горе, десно, дефинисани су за 60 подинтервала, са маркираном референтном, теоријском (Гаусовском) густином расподеле и средњом вредности резидуала. QQ крива (доле, десно) приказује резултате поређења добијене расподеле са очекиваном, у овом случају, гаусовском. Генерално, за све временске серије, хистограми су симетрично распоређени око нулте вредности. Укупни резултати показују две тенденције: за већину временских серија расподеле резидуала је блиска гаусовској (као на примеру 2. сата са слике 5.2.11), а са друге стране, за сате летњег дневног максимума, за прогностичке области са великим бројем потрошача, расподела значајно одступа од нормалне (пример 15. сата са слике 5.2.12).

Резултати



Слика 5.2.11: Резултати анализе резидуала за 2. сат, NYC зона

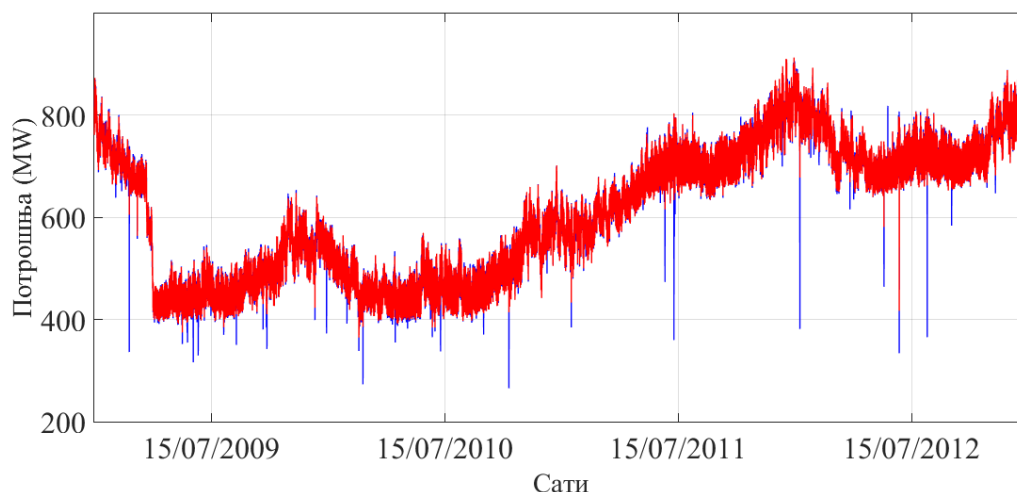


Слика 5.2.12: Резултати анализе резидуала за 15. сат, NYC зона

Резултати

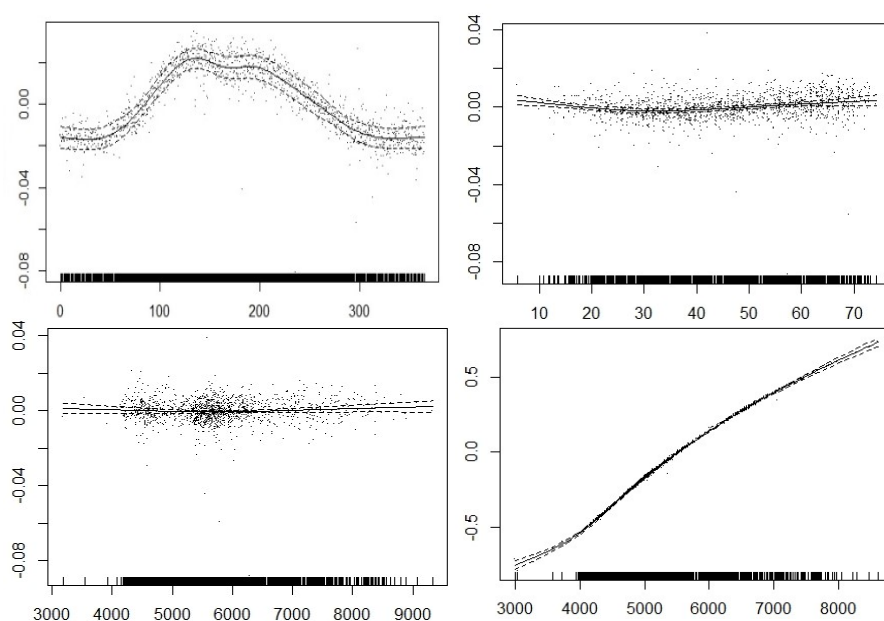
Већа варијабилност резидуала модела у летњој сезони у односу на остатак године, за неке сате и зоне, је карактеристична појава која се односи на утицаје метеоролошких фактора на потрошњу, који се може разликовати током сезоне грејања, зими, и хлађења, лети. Профил годишње потрошње за Њујорк са слике 4.1, указује да је могуће да се зими користи гас и значајнији је пораст потрошње лети услед повећаног броја укључења клима уређаја великог броја потрошача. По резултатима наведеним у референци [5.3], се показало да за већину области у Сједињеним Америчким Државама, прогнозе засноване само на температурама имају тенденцију негативне грешке током летњих месеци и да је потребно даље проширење додатним варијаблама. За податке који се користе у тези, у коме се користе дневни максимум и минимум температуре горњи резултати анализе резидуала показују да се може очекивати већа грешка прогнозе у летњим месецима, за веће температуре у односу на остатак године.

Што се тиче аутлејера, публикације у којима се користе NYISO подаци, а баве се анализом ефеката интеграције обновљивих извора [5.4], утицајем тржишта и цена [5.5], статистичким карактеристикама грешке прогнозе [5.6, 5.7] за NYISO потрошњу, показују да је проблем спорадичних екстремних вредности карактеристичан за ове податке. С друге стране, ГАМ модел се показао као, релативно, робустан на већи број аутлејера, што се и види на слици 5.2.13 где су приказани резултати моделовања за NORTH зону, (реалне вредности су приказане-плавом бојом, резултати моделовања - црвеном).



Слика 5.2.13: Реалне и моделова сатне вредности за период 2009-2012 год. за NORTH зону

На слици 5.2.14 су приказани добијени модели појединачних ГАМ чланова на примеру годишњег профила (горе, лево), зависност од максималне дневне температуре (горе, десно), зависност од потрошње из претходног дана (доле, лево) и претходног сата (доле, десно), за 14 сат, NYC зоне. (Слике су аутоматски генерисане коришћењем функције *gam.plot* из *mgcv* пакета). Графици приказују релативне промене модела појединачних ГАМ чланова трансформисане потрошње (на ординати), за вредност одговарајућих улазних променљивих на апсциси. Приказана је моделована вредност ГАМ компоненте са 95% интервалом поверења и узорцима резидуала. Узорци резидуала, добро распоређени око моделоване вредности, без систематичног одступања и коректне EDF вредности, које су за приказане ГАМ чланове били 8.01, 1.48, 6.64 и 7.98 респективно, указују на добро димензионисане чланове модела.



Слика 5.2.14: Временски график и аутокореалција резидуала за 3. сат, NYC зона

На основу укупних резултата анализе остатка од моделовања свих временских серија из NYISO сета су верификовали модел заснован на предложеном каталогу ГАМ чланова.

5.2.2 Неколико коментара о резултатима селекције ГАМ модела

Резултати селекције ГАМ чланова је могуће проверити увидом у p -вредности (< 0.05) за сваки ГАМ члан (функцијом *gam.summary*, *mgcv* пакет). С обзиром на велики број временских серија, дати су коментари укупне анализе за све сатне потрошње и све зоне. Седмичне промене потрошње као и утицај из претходна два сата су веома значајни за све временске серије. Идентификована су три типа временских серија у зависности од доминантних фактора потрошње, оних који зависе од календарских варијабли код којих су примарни годишња и седмична сезоналност, температуром, или претходним вредностима потрошње. Ноћни сати су, на пример, претежно одређени седмичним и годишњим профилем и њиховом променама, а за неке потрошње значајним потрошњом и температуром из претходног дана. За јутарње (7 – 9) и послеподневне (14–17) сате, утицај вредности потрошње из претходних дана конвергира ка својој минималној структури из претходног дана и претходне седмице. Што се тиче температура, генерално је, актуелна вредност за текући дан од највећег значаја (поготово њен максимум), док су вредности из претходних дана и њихове промене током седмице и године су спорадично значајне за неке зоне и неке сате. Сати у средини дана (12-15) су махом температурно оријентисани. Специјални дани су више значајни за дневне сате, почев од 7 сати ујутру. Интересантно, за мале зоне, добри резултати прогнозе за специјалне дане се добијају усвајањем само две категорије, тј. регуларни/специјални дани, док су за високо урбане зоне, све 4 уведене категорије дана од значаја. Промене услед *DST* ефекта највише утичу на 6.-10. сат ујутру и 17.-18. сат увече, дакле, око времена изласка и заласка сунца.

5.2.3 Тестирање предиктивних перформанси модела

Предиктивне перформансе модела се тестирају коришћењем Процедуре 1 за реалне вредности улазних променљивих (температуре, претходне вредности потрошње), како би се искључио утицај неодређености у улазним подацима. За естимацију сатног подмодела (са око 400 коефицијената) и генерисање прогнозе било је потребно < 15 секунди на AMD Ryzen 5 3500U, 2.10 GHz, 8.00 GB RAM рачунару. Вредности MAPE/MAE грешке су усредњаване на нивоу зона за 2013. годину, дато у Табели 5.2.2, док је њихова расподела по сатима и по зонама дата на слици 5.2.15.

Практично, добијене MAPE/MAE вредности представљају резултате прогнозе за *реалну STLF функционалност за један сат унапред, за реалне вредности температура*.

Иако поређење са резултатима других приступа моделовања који се налазе у литератури за NYISO податке није доследно због разлика у начину препроцесирања података и спроведеним процедурама, прелиминарно, добијене MAPE вредности показују већу тачност прогнозирања од оних из, на пример, радова [5.8], [5.9].

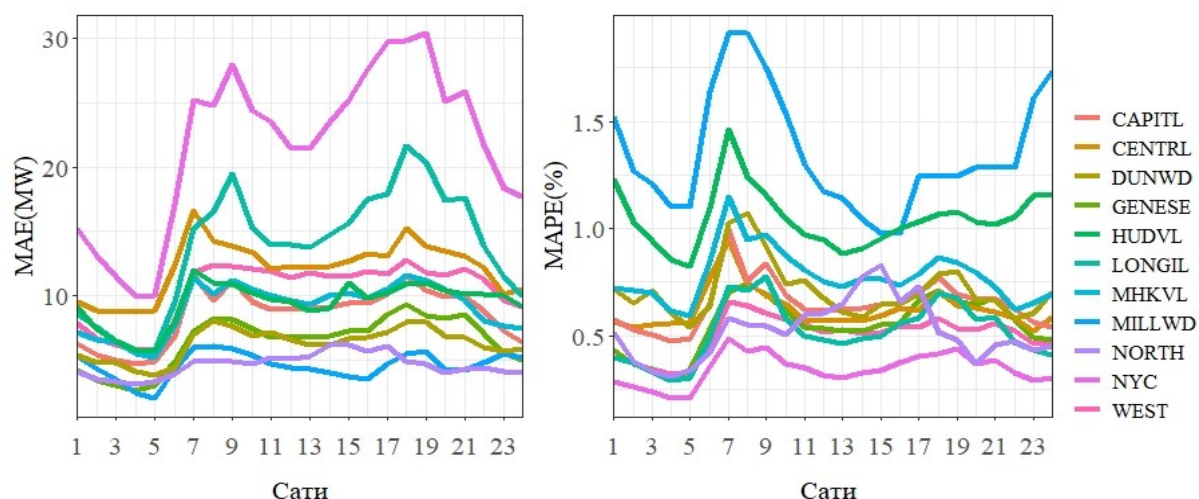
Табела 5.2.2: MAPE/MAE резултати за један сат унапред, за NYISO зоне

Зона	MAPE (%)	MAE (MW)
WEST	0.50	10.3
GENESE	0.55	6.43
CENTRL	0.62	12.1
CAPITL	0.65	8.5
MILLWD	1.31	4.5
DUNWD	0.71	6.2
NYC	0.31	21.7
LONGIL	0.52	13.7
NORTH	0.78	9.0
HUDVL	0.98	9.4
MHKVL	0.52	4.6

Једна од претпоставки тезе је да се поновном селекцијом ГАМ-а са сваким новим податком може обезбедити боља адаптивност модела у времену и боља тачност прогнозе. Како би се потврдила ова претпоставка, упоређени су резултати прогнозе са и без динамичке селекције. Резултати прогнозе су посматрани за јул 2013., који обухвата период екстремних температура, значајно јер представља сценарио у коме се модел тестира на ‘неочекиване’ вредности температура. Посматрана су 2 случаја:

1. Користи се стандардни поступак унакрсне валидација за избор модела, описан у подпоглављу 3.4. Подаци за период од 2009 до 2012 одвојени за почетно обучавање модела (N_{tr} на слици), они за период од 01.01 до 30.06. 2013 за његову валидацију, (N_{val} на слици). Претражују се модели за различите скупове ГАМ чланова, почев од седмичног и годишњег профила, затим се, један по један, додају ГАМ чланови за температуре, прво актуелне, а потом, и за претходне дане, вредности потрошње из

Резултати



Слика 5.2.15: Расподела MAPE/MAE грешке по сатима за NYISO зоне

претходних дана, као и чланови интеракција, до комплетног скупа чланова из Табеле 5.1. Поступак се спроводи за сваку временску серију потрошње и за сваку се бира модел са најмањом грешком прогнозе. Добијене MAPE вредности за период тестирања на нивоу зона, дате су у колони MAPE_{cv}, Табеле 5.2.3.

- Користи се поступак предложен у тези, где се модел бира током естимације, по Процедури 1, подобласт 3.5. За почетно обучавање модела користе се подаци за период $N = N_{tr} + N_{val}$ за сваку сатну временску серију потрошње. Коначни резултати прогнозе за период тестирања на нивоу зона, су смештени у колони MAPE_{dp}, Табеле 5.2.3.

Табела 5.2.3: Упоредни MAPE/MAE резултати прогнозирања модела добијеног поступком унакрсне валидације и предложеном методологијом

Зона	MAPE _{cv}	MAPE _{dp}	Унапређење (%)
WEST	0.85	0.67	27
GENESE	0.86	0.77	17
CENTRL	0.90	0.74	22
CAPITL	0.81	0.71	14
MILLWD	1.01	0.61	65
DUNWD	1.06	0.98	8
NYC	0.436	0.41	6
LONGIL	0.84	0.71	18
NORTH	1.24	0.84	48
HUDVL	1.29	1.19	8
MHKVL	0.77	0.66	16

Унапређења тачности прогнозе, мерене процентуалним смањењем MAPE грешке (трећа колона Табеле 5.2.3), су евидентна. Иако су апсолутне разлике мале, ова побољшања добијају на значају у реалним условима, за прогнозиране вредности температурâ и за временске хоризонте веће од једног сата. Најмања разлика у тачности прогнозе се добила за NYC зону, што указује на релативно стабилан модел потрошње за ову зону, за период тестирања. Значај предложене методологије се, међутим, највише огледа код мањих зона, са варијабилним профилима потрошње, као што је MILLWD или оних са нестандартним профилем, као што је NORTH.

Резултати

5.2.4 Резултати прогнозе за 24 сата унапред

Као подлога за основну STLF функционалност прогнозирања потрошње за наредни дан, посматрани су резултати прогнозе за наредних 24 сата, за сваку од NYISO зона. Прогноза се иницира у 00 сати, кад сваки од сатних подмодела аутоматски генерише прогнозу за наредни дан, по Процедури 1. Разлика у односу на претходни режим прогнозирања (за један сат унапред), је у томе што улазни подаци уносе неодређеност која утиче на резултате прогнозе. Пре свега, с обзиром да модел садржи чланове из протекла два сата, прогнозе за 2,...24 сат се ослањају на прогнозе из 1,... 23, што доводи до мање или веће акумулације грешке у прогнозама. Такође, у реалним условима, за наредних 24 сата, на располагању су само прогнозиране вредности температуре, које уносе додатну грешку прогнозирања. Како би се одвојено посматрала ова два случаја, прогнозирање је спроведено за реалне и прогнозиране вредности температуре.

Временски периоди од две седмице за сваку сезону, заобилазећи празнике, одабрани су за тестирање и приказ резултата: 7-20 јануар (зима), 6-19 април (пролеће), 7-20 јул (лето са летњим максимумом температуре) и 2-15 новембар (јесен), за 2013. годину. Резултати MAPE/MAE грешака за одабране тест периоде, за прогнозиране (**п**) и реалне (**р**) вредности температура, дате су у Табели 5.2.4. Средње вредности грешака које се очекују на основу резултата за одабране периоде, дате су у последњој колони.

Табела 5.2.4: MAPE/MAE резултати за 24 сата унапред, за NYISO зоне

Зона	п/р	Зима	Пролеће	Лето	Јесен	Средња вредност
WEST	п	1.22 (22.87)	1.50 (27.29)	2.31 (49.87)	1.34 (25.42)	1.59 (31.36)
	р	1.18 (22.28)	1.34 (23.41)	2.02 (42.55)	1.31 (24.90)	1.46 (28.29)
GENESE	п	1.40 (20.05)	1.46 (16.35)	2.06 (30.87)	1.15 (12.53)	1.52 (19.95)
	р	1.32 (19.01)	1.34 (14.71)	1.81 (26.43)	1.07 (12.14)	1.39 (18.07)
CENTRL	п	1.42 (28.32)	1.28 (22.28)	1.93 (43.91)	1.46 (28.61)	1.52 (30.78)
	р	1.32 (25.71)	1.20 (21.21)	1.85 (42.81)	1.42 (27.63)	1.44 (29.34)
CAPITL	п	1.53 (20.84)	1.69 (23.82)	2.01 (35.79)	1.62 (21.92)	1.71 (25.59)
	р	1.43 (19.38)	1.63 (20.62)	1.94 (34.76)	1.55 (20.81)	1.64 (23.89)
MILLWD	п	2.75 (9.36)	2.92 (9.99)	2.97 (15.58)	2.94 (9.33)	2.89 (11.06)
	р	2.35 (7.91)	2.73 (8.29)	2.88 (14.89)	2.73 (9.29)	2.67 (10.01)
DUNWD	п	2.01 (14.51)	2.00 (13.58)	3.12 (36.87)	1.51 (10.32)	2.16 (18.82)
	р	1.80 (12.26)	1.76 (12.37)	2.81 (36.87)	1.36 (9.14)	1.93 (17.66)
NYC	п	1.10 (64.60)	1.31 (74.62)	2.42 (217.54)	0.83 (54.71)	1.42 (102.87)
	р	0.90 (52.82)	1.03 (57.73)	2.06 (193.29)	0.79 (49.52)	1.20 (88.34)
LONGIL	п	1.77 (45.94)	1.72 (40.22)	2.72 (112.38)	1.79 (41.62)	2.00 (60.04)
	р	1.73 (45.73)	1.61 (36.01)	2.67 (108.23)	1.55 (35.92)	1.89 (56.47)
NORTH	п	1.17 (9.53)	1.01 (7.81)	1.15 (8.83)	1.51 (11.83)	1.21 (9.50)
	р	0.98 (8.00)	0.92 (7.06)	1.09 (8.33)	1.38 (10.64)	1.09 (8.51)
HUDVL	п	1.65 (18.77)	1.68 (17.91)	2.82 (47.90)	1.64 (18.3)	1.88 (25.72)
	р	1.60 (18.10)	1.53 (15.82)	2.69 (42.05)	1.42 (15.7)	1.81 (22.92)
MHKVL	п	1.82 (20.21)	2.28 (25.41)	2.68 (30.67)	2.06 (19.54)	2.21 (23.96)
	р	1.76 (19.03)	2.02 (20.12)	2.63 (29.60)	2.00 (18.85)	2.01 (21.90)

Резултати

Минимална тачност прогнозе је добијена за MILLWD зону, што је очекиван резултат с обзиром на најмањи број потрошача. Евидентно је смањење тачности прогнозе у летњим месецима (поготово за летњи максимум температуре) које је мање изражено за мање насељене зоне NORTH и MILLWD, док је значајно за високо урбане зоне, са значајно већим бројем потрошача.

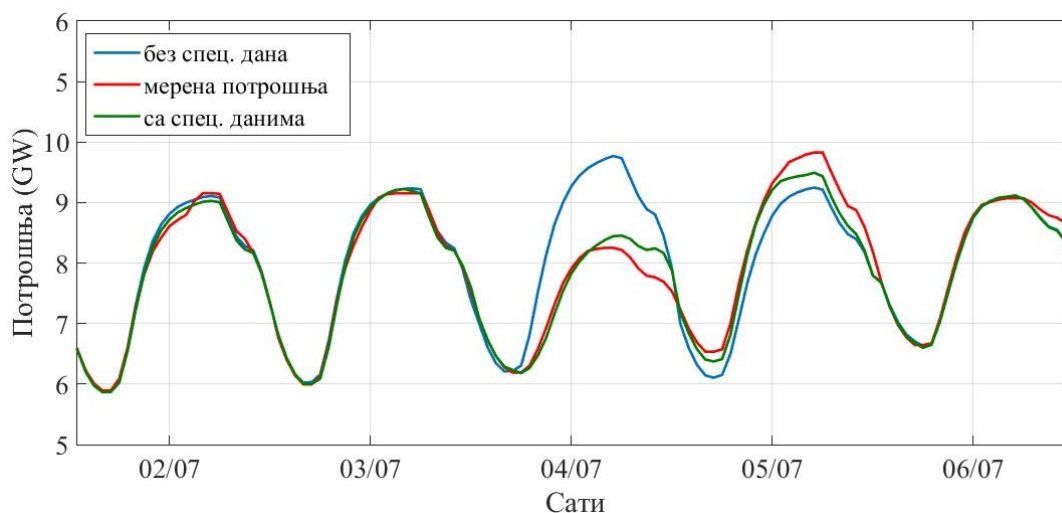
За тестирање ефекта специјалних дана, упоређивани су резултати прогнозе за датуме одабраних празника између модела код којих је њихов ефекат урачунат и оног где су ови датуми третирани као регуларни дани и MAPE резултати су дати у Табели 5.2.5, где **без/са** представља неурачунат/урачунат специјални дан.

Табела 5.2.5: MAPE резултати без/са урачунатим специјалним данима, за NYISO зоне

Зона	са/без	1.јануар	Ускрсе	Дан Сећања	Дан Независности	Колумбов дан
WEST	са	2.91	0.94	2.23	2.72	2.03
	без	7.94	5.72	10.50	8.30	2.15
GENESE	са	3.52	1.51	2.15	3.82	1.92
	без	11.59	8.72	12.50	10.10	2.14
CENTRL	са	3.62	1.58	1.74	3.34	1.24
	без	5.30	6.64	8.82	7.23	1.38
CAPITL	са	4.92	2.24	3.25	3.13	1.52
	без	10.12	3.82	11.81	7.94	1.77
MILLWD	са	2.54	3.23	3.92	2.53	2.22
	без	10.81	3.53	11.26	3.72	2.27
DUNWD	са	3.72	2.53	2.10	2.31	2.43
	без	10.12	2.95	9.74	9.34	2.55
NYC	са	3.40	1.33	1.32	2.01	3.21
	без	11.23	5.97	9.90	10.80	3.39
LONGIL	са	2.84	3.17	3.41	3.32	3.50
	без	9.16	3.95	10.92	5.27	3.67
NORTH	са	1.72	0.88	1.24	0.92	0.91
	без	1.83	1.84	2.63	2.32	1.01
HUDVL	са	3.72	1.13	4.62	1.53	1.63
	без	8.73	5.94	7.81	3.35	1.73
MHKVL	са	3.21	2.10	3.26	1.92	2.21
	без	9.97	4.72	14.62	4.56	2.34

На основу добијених резултата, за оне празнике који су и нерадан дан за већину популације, школе, пословне и комерцијалне објекте, као што је Дан сећања или Дан независности, MAPE грешка се смањила са више од 50%. Као илустрација, на слици 5.2.16 је дат пример Дана независности за NYC зону, где је без урачунатих дана грешка била 10.8% (грешка између плаве и црвене линије), док је са укљученим ефектом пала на 2% (грешка између плаве и зелене линије).

Резултати



Слика 5.2.16: Ефекат моделовања специјалних дана на примеру Дана независности, NYC

Међутим за неке празнике побољшања су слабије изражена. За Колумбов дан, на пример, MAPE грешка без урачунатих специјалних дана је у опсегу средње вредности потрошње, последично унапређење није значајно. За неке зоне, дани Ускрса, третирају се као регуларни дани, такође имају мањи пад грешке прогнозе. Могуће објашњење је да, с обзиром да је Ускрс дан викенда, пад потрошње обухваћен моделом седмичног профила и допринос који уноси модел специјалних дана је мање изражен. Закључак је да је увођење ГАМ чланова за моделовање специјалних дана допринело повећању тачности прогнозе, да су потребна даља унапређења за групу празника са комплекснијим профилима потрошње (Нова година, Ускрс) и тестирање на већем скупу података.

Промене у потрошњи услед *DST* ефекта су најочљивије у периоду непосредно после имплементације и за тестирање су посматрани резултати прогнозе за прву седмицу након пролећне промене сата (која обухвата период од 11. до 17. марта). Као и код специјалних дана, посматрани су упоредни резултати са и без урачунатих *DST* чланова модела. Средње вредности MAPE грешке, дате за сваку зону у Табели 5.2.6., показују побољшање тачности прогнозе. Моделовање *DST* ефекта потрошње је могуће унапредити коришћењем додатних варијабли (време изласка/заласка сунца) [5.10], чиме се пројектује постизање боље тачности прогнозе.

Табела 5.2.6: MAPE резултати без/са урачунатим *DST* ефектом, за NYISO зоне

Зоне	MAPE без <i>DST</i> ефекта	MAPE са <i>DST</i> ефектом
WEST	1.58	1.28
GENESE	1.84	1.65
CENTRL	2.32	2.21
CAPTIL	1.94	1.66
MILLWD	4.50	4.23
DUNWD	2.05	1.78
NYC	1.97	0.85
LONGIL	1.92	1.58
NORTH	2.16	1.38
HUDVL	2.68	2.64
MHKVL	1.88	1.85

5.2.5 Кратак резиме и коментари

У претходним подпоглављима је описан поступак формирања каталога ГАМ чланова и његова валидација за потрошње из NYISO базе података и дати резултати краткорочне прогнозе потрошње за сат и дан унапред. Приказани укупни резултати су публиковани у раду из 1.5 под 1. У наставку је дато неколико коментара који се тичу практичне примене предложене методологије.

- ГАМ каталог из Табеле 5.2.1. имплементира основне факторе сваке потрошње у ЕЕС-у и генерално је применљив за моделовање потрошње на различитим местима у систему. Измене каталога, елиминисањем или додавањем чланова, зависе од расположивих мерења и података за спољашње променљиве, расположивих података за обучавање и временског хоризонта. У приказаном примеру за NYISO податке, од спољашњих променљивих су на располагању били дневни максимум и минимум температура. Укључивање осталих метеоролошких или других спољашњих променљивих, различите резолуције мерења (сатне, дневне) се може извести на сличан начин, додавањем одговарајућих ГАМ чланова. Такође, избор ГАМ чланова из каталога који улазе у модел зависи од величине сета података који се користи за обучавање модела. На пример, модел који се обучава подацима од неколико месеци не може ‘ухватити’ годишњи профил, специјалне дане и *DST* ефекат и одговарајући ГАМ чланови се елиминишу из каталога. Чланови модела од којих директно зависи квалитет прогнозе за веће временске хоризонте прогнозирања су вредности потрошње из суседних сати, због акумулације грешке коју уносе у прогнозу (видети 5.2.4). За временски хоризонт од 24 сата унапред, овакав приступ даје добре резултате. Развој методологије прогнозирања за веће временске интервале прогнозирања је засебна област истраживања и неколико варијанти је могуће наћи у референци [5.11].
- Директна примена Процедуре 1, којом се модел обучава на регуларној бази, у пракси, може имати ограничење што јој је за обучавање модела потребан, релативно, велики број података. Међутим, предложена методологија се може експлоатисати зависно од услова, доступних података, апликације и конкретних захтева. На пример, могуће је увести периодично обучавање уместо редовног, где је од значаја истражити који је то временски период којим се може очувати квалитет прогнозе у времену, за дате податке и ГАМ чланове. Алтернативно, обучавање модела може бити повремено, као на пример, у раду [5.12] где се једноставним алгоритмом прати квалитет прогнозе и сигнализира покретање процедуре за обучавање када дође до значајних промена или погоршања тачности прогнозе.
- Адитивна структура модела пружа могућност засебног моделовања и парцијалног обучавања. На пример, дугорочне и/или “стабилне” компоненте модела (на пр. годишњи профил) се могу обучавати само повремено, док компонента којом се моделује утицај претходних потрошњи, редовно и са мањим бројем података. Овакав приступ захтева даља истраживања.

5.3 Пример потрошње произвољне области у дистрибутивном систему – АУ база података

АУ подаци су искоришћени за тестирање методологије на примеру реалног сценарија по коме је кориснику омогућен избор произвољне прогностичке области. Селекцијом зонских трансформаторских станица (или више њих) из АУ базе података дефинише се област обухваћена одговарајућим трафо-рејонима за чију се укупну потрошњу врши прогноза. АУ прогностичке области су већински резиденцијалне, али обухватају и мале и велике комерцијалне потрошаче као и објекте од јавног значаја. С обзиром на различит састав и број потрошача, постоје значајне разлике у профилима и нивоу потрошње за различите области. Важан податак је да је Аустралија, генерално, једна од водећих земаља у свету по броју *behind the meter* фотонапонских система, са преко 83% од укупног 6GW капацитета инсталираних на крововима резиденцијалних објеката, [податак из 2016. године наведен у раду 5.13]. То значи да се може очекивати да, за неке области, потрошња мерена на *Ausgrid* SCADA-и бележи укупно утрошену електричну енергију, умањену за компоненту утицаја производње из солара (објашњење у поглављу 4.2). Разлике у броју инсталираних јединица и нивоу пенетрације за различите АУ области чија се производња уграђује у профиле мерене потрошње, додатно повећавају њихову хетерогеност и неодређеност при моделовању. Проблем при дефинисању каталога ГАМ чланова који би био довољан да опише све потрошње из АУ базе је 1) што укључује, потенцијално, велики број додатних улаза, поготово метеоролошких, и 2) што постоје велике разлике у њиховој релевантности за различите АУ области.

У табели 5.3.1. је дато проширење каталога ГАМ чланова из претходног примера за моделовање сатне потрошње, који укључује стандардизоване чланове за температуру, средње сатне (актуелне и вредности из претходних M_{Th} сати), средње дневне из претходних M_{Td} дана и сатне вредности секундарне метео варијабле, SV (актуелне и из претходних M_{SV} дана). SV могу укључивати, брзину и правац ветра, влажност, падавине, ирадијацију, облачност, итд., и за сваку од њих се конструише засебан ГАМ члан. У каталог су укључени и чланови којим се моделује утицај потрошње за произвољан број претходних D дана и H сати, чланови интеракције између температуре и осталих метеоролошких променљивих, седмичног профила и температуре и/или других метео варијабли, итд. Циљ оваквог проширења је да се каталогом обухвате сви унапред идентификовани фактори моделованих потрошњи, за различита мерења која могу бити на располагању и омогући дефинисање произвољног броја, превасходно, метеоролошких променљивих.

Додавање произвољног, потенцијално, великог броја ГАМ чланова, међутим, значајно повећава димензије модела, последично, успорава поступак његове естимације и, даље, прогнозирања у Процедури 1. Зато се, пре уласка у Процедуру 1, уводи додатни корак, где се поступком унакрсне валидације, из генерализованог каталога врши ужи избор релевантних ГАМ чланова за конкретну потрошњу. Поступак унакрсне валидације који се користи описан је у поглављу 3.4., са шемом претраге ГАМ чланова слично оној датој у публикацији [3.24]. Почев од комплетног скупа ГАМ чланова, дефинисаног за *све* метеоролошке варијабле за које постоје мерења и за довољно велике вредности, D , H , M_{Td} , M_{Th} и M_{SMV} , из Табеле 5.3.1., на даље се, итеративно, тестирају различите комбинације ГАМ чланова, на начин описан у 3.4.

Резултати

Табела 5.3.1: Генерализовани каталог ГАМ чланова потрошње

ГАМ члан	Ефекат потрошње
$f_t(trend)$	Компонента тренда
$f_g(g)$	Годишњи профил
$f_s(s)$	Седмични профил
$DS f_g(g)$	Промене годишњег профила услед <i>DST</i> ефекта
$DS f_s(s)$	Промене седмичног профила услед <i>DST</i> ефекта
$f_T(T_{x,h}^d)$	Стандардни члан за моделовање зависности од температуре, за: <ul style="list-style-type: none"> • x тип мерења (средња дневна, дневни максимум/минимум, средња сатна) • за актуелни и претходне дане, $d = 0: -M_{Td}$ • за актуелни и претходне сате, $h = 0: -M_{Th}$
$f_{sv}(SV_h)$	Зависност потрошње од секундарне метео варијабле за актуелни и претходне сате, $h = 0: -M_{SV}$.
$f^d(P^d)$	Утицај вредности потрошње из претходних дана d , где је $d = -2: -D$
$f_h(P_h)$	Утицај вредности потрошње из претходних сати h , где је $h = -1: -H$
SD	Промена нивоа потрошње за тип дана који урачунава ефекат специјалних дана
$SD f^{-1}(P^{-1})$	Утицај вредности потрошње из претходног дана, зависно од типа дана
$f_{xy}(x,y)$	Интеракција или заједнички ефекат x и y променљиве, које могу бити: <ul style="list-style-type: none"> $x = s, y = g$, промене седмичног профила током године $x = T, y = SV$, интеракција температуре и секундарне мет. вар. $x = week, y = T, SV$, интеракција мет. вар. и седмичног профила $x = P^{-1}, y = T, SV$, интеракција мет. вар. и потрошње из претходног дана $x = P_h, y = T, SV$, интеракција м. вар. и потрошње из претходних сати

Засебно и сукцесивно се претражују 4 категорије ГАМ чланова: календарске (седмични профил, тренд, специјални дани, *day-light savings*), метеоролошке (температура и секундарне метео варијабле), утицај потрошњи из претходног периода и чланови интеракција. Претрага ГАМ чланова унутар категорије је као у одељку III.B поменуте референце [3.24]. За конкретну примену, претрага ГАМ чланова може бити делимична, где се чланови модела за, на пример, одређену категорију унапред задају и претражује остатак из каталога. Овде је од интереса, пре свега, претражити ГАМ чланове са имплементираним метеоролошким факторима, како би се елиминисали они који нису од значаја за потрошњу конкретне области. Описани поступак се може покретати периодично и по потреби како би се пратиле дугорочне промене у саставу потрошача, проширења ОИЕ капацитета, итд., које, потенцијално, утичу на промене фактора потрошње.

Резултати

У примеру примене на АУ податке, 15-минутне вредности за потрошњу су усредњаване на нивоу сата. Од метеоролошких варијабли, на располагању су биле температура, влажност ваздуха, брзина ветра, облачност и инсолација. Подаци за потрошњу и спољашње променљиве, за период од 01.01.2021 до 31.07.2022. је кориштен за дефинисање иницијалног сета ГАМ чланова потрошње, од тога, од 01.05.2021 до 01.04.2022. се користе за естимацију модела, остатак до краја сета за валидацију. Резултати прогнозе посматрани су за период од 01.08.2022 до 30.10.2022. Поступак селекције релевантних ГАМ чланова из Табеле 5.3.1, за сваку потрошњу из АУ базе се састоји из следећих корака:

1. годишњи профил, утицај специјалних дана и *day-light savings* ефекат су елиминисани из скупа, јер је за њихово моделовање потребна база од више година,
2. усвојено је $D=7$ и $H=2$, као што је предложено у одељку 4.3,
3. генерисани су чланови који укључује утицај температуре, влажности ваздуха, брзине ветра, облачности и ирадијације и усвојено је $M_{SMV} = M_{Tx} = M_T = 2$,
4. генерише се комплетан скуп чланова интеракције из Табеле 5.3.1 који укључују све метеоролошке променљиве,
5. за фиксиране горње вредности, унакрсном валидацијом се претражују чланови модела за метео варијабле и чланове интеракције из Табеле 5.3.1.

Горњим поступком се дефинише иницијални каталог ГАМ чланова за сваку АУ потрошњу. Извршена је *Box-Cox* трансформација потрошњи за сваку прогностичку област која улази у Процедуру 1. Грешке прогнозе за један и 24 сата унапред $MAPE_1$ (MAE_1) и $MAPE_{24}$ (MAE_{24}), респективно, дати су у Табели 5.3.3.

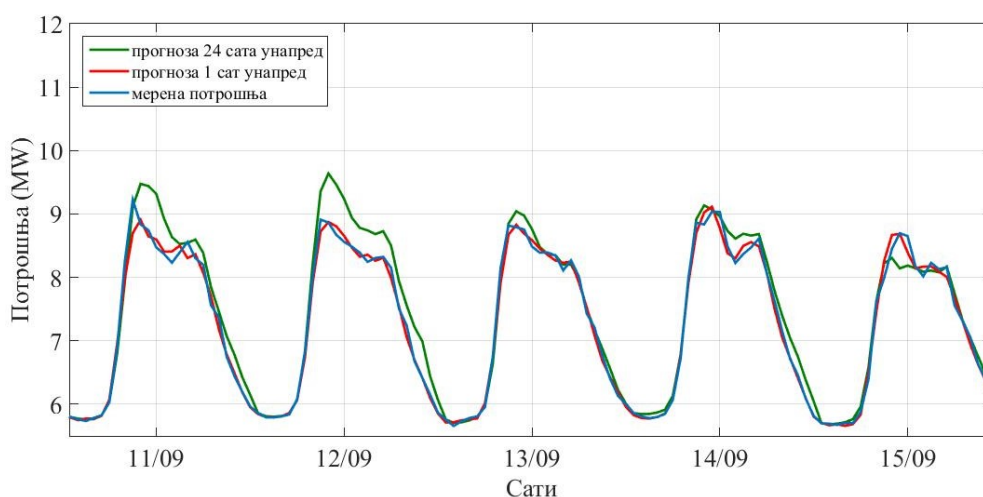
Табела 5.3.2: $MAPE/MAE$ резултати за један и 24 сата унапред, за АУ прогностичке области

Редни број	Назив области	$MAPE_1$ (MAE_1)	$MAPE_{24}$ (MAE_{24})
1	<i>RNS Hospital</i>	0.83 (0.04)	2.45 (0.16)
2	<i>Auburn</i>	1.48 (0.13)	3.88 (0.34)
3	<i>Olympic Park</i>	0.76 (0.11)	2.89 (0.43)
4	<i>Chatswood</i>	1.07 (0.22)	3.45 (0.79)
5	<i>Leichhardt</i>	1.48 (0.28)	3.91 (1.08)
6	<i>Matraville</i>	1.51 (0.29)	4.72 (1.33)
7	<i>Cronulla</i>	1.45 (0.32)	5.01 (1.04)
8	<i>Clovelly</i>	1.57 (0.33)	3.66 (0.92)
9	<i>Maroubra</i>	0.94 (0.21)	2.94 (0.62)
10	<i>North Sydney</i>	0.95 (0.32)	3.16 (1.02)
11	<i>Burwood</i>	1.16 (0.39)	3.16 (1.63)
12	<i>City North</i>	0.87 (0.20)	2.22 (0.41)

Пример STLF модела, развијаног за *Ausgrid* податке, у рефернцама [5.14], где се користи ансамбл метода за *bottom-up* приступ прогнозирања потрошње, за 155 АУ трафо рејона, добијена $MAPE$ грешка за 24 сата унапред је $>5\%$. Такође, у раду [5.15] је приказан примена LSTM модела на 15-минутне вредности за потрошњу, за коју је вредности $MAPE$ грешке наредне вредности $\sim 2\%$.

Резултати

Резултати избора релевантних ГАМ чланова поступком унакрсне валидације, за већину трафо области, је била некарактеристична, тј. добијени модели су просечних димензија са мањим бројем метео варијабли и интерактивних чланова у односу на комплетан сет из каталога. Статистички, најзначајнији чланови модела су они за температуру и њене интеракције са календарским променљивим, док је интеракција између температуре и секундарних метео варијабли, од мањег значаја. На пример, *RNS Hospital* је прогностичка област која обухвата локацију око *Royal North Shore* болнице. Показало се да сатне временске серије потрошње за ову област, претежно, зависе од календарских променљивих и дневне температуре из претходног дана, док остале метео варијабле нису биле од значаја. На слици 5.3.1 су приказани профили и резултати прогнозирања за ову област.



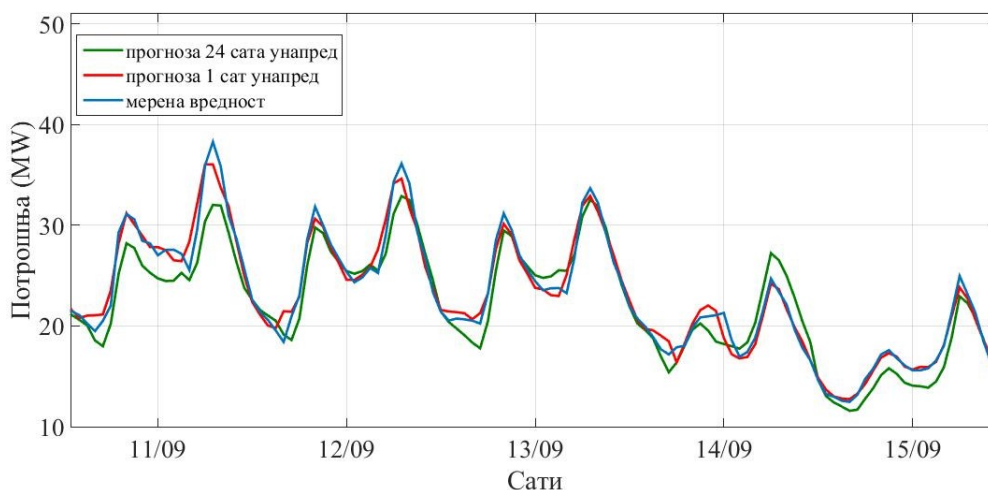
Слика 5.3.1: Резултати прогнозе за *RNS Hospital* трафо рејон

Што се тиче метеорологије, сличан резултат се добија за *Leichhardt* област, за коју је једино релевантна сатна температура. Унапређења тачности прогнозирања су, међутим, постигнута додавањем чланова интеракције тренда и вредности потрошње из претходног дана и сата са седмичним профилем. То указује на доминантну компоненту навика потрошача са ове области која зависи од дана у седмици на утрошену електричну енергију. На слици 5.3.2 су приказани профили и резултати прогнозирања за ову област.

Трафо рејон *Cronulla* је интересантан јер је изразито осетљив на метеорологију. Резултати унакрсне валидације су показали да на потрошњу ове области утичу све метео варијабле које су коришћене, температуре из претходног дана и скоро свака интеракција која укључује температуру. На слици 5.3.3 је приказана расподела *absolute percentage error*, APE ¹⁵ грешке прогнозе по сатима, за 11 август 2022. Резултати су приказани за 1) базични модел који обухвата ГАМ чланове само са календарским променљивим (плава), 2) базични модел са додатим чланом за сатну температуру (зелено), 3) базични модел са члановима за температуру и остале метео варијабле (љубичасто) и 4) базични модел са свим метео варијаблама и члановима интеракције (црвено). Укупна $MAPE$ грешка за

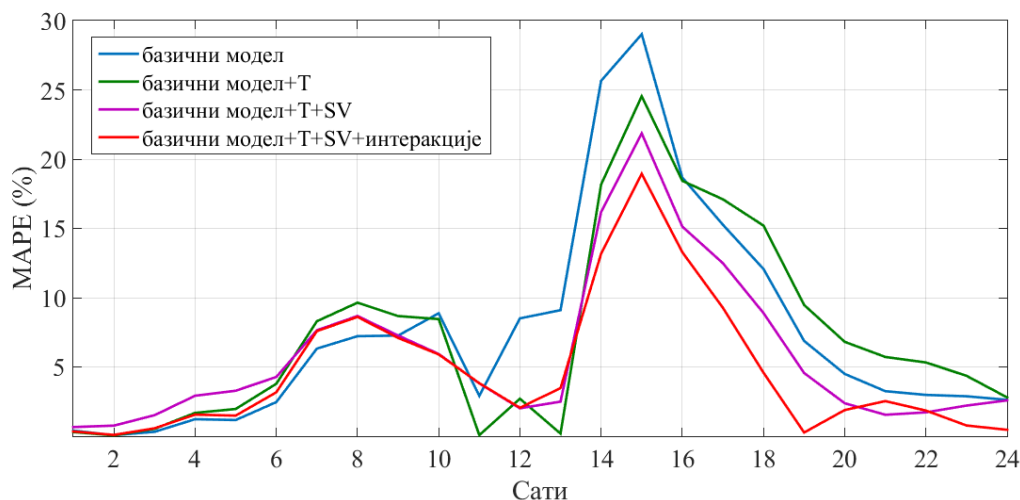
¹⁵ $APE = 100 \times |P - \hat{P}|/P$, где је P реална вредност, \hat{P} прогнозирана вредност

Резултати



Слика 5.3.2: Резултати прогнозе за *Leichhardt* трафо рејон

ове моделе, је 7.48%, 7.26%, 5.87% и 4.70%, респективно. Упоредни график резултата показује унапређење тачности прогнозирања које се добија додавањем температура (~4%) секундарних варијабли (за ~ 20%) и њихових интеракција (за ~ 40%), у односу модел који не урачунава утицај метеоролошких фактора. Такође, приказани резултати показују да постоји простор за даља унапређења, поготово за побољшање тачности прогнозе у време дневног маскимума.

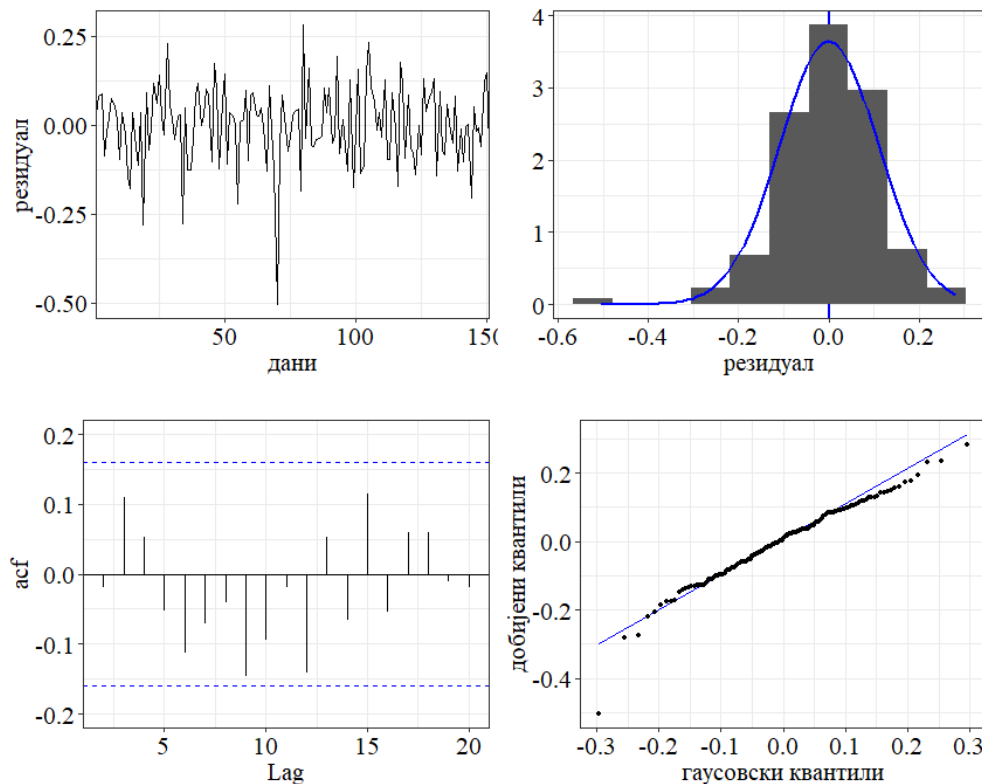


Слика 5.3.3: Расподела APE грешке по сатима за *Cronulla* трафо рејон за 11. август 2022.год.

Анализа резидуала моделовања поступком описаним у претходном одељку, приказана примером 15. сата за *Cronulla* трафо рејон, показала је коректно моделовање за све AU временске серије потрошње: без тренда, промене нивоа и остатака корелације који би се пренеле у прогнозу. Парцијални резидуали и одговарајуће *EDF* вредности су показали коректно димензионисане ГАМ чланове.

Као практично интересантан пример, анализирани су резултати прогнозе за различите агрегације трафо-области. На слици 5.3.1 су приказани резултати MAPE грешке за прогностичке области добијене сабирањем потрошње за више трафо рејона.

Резултати



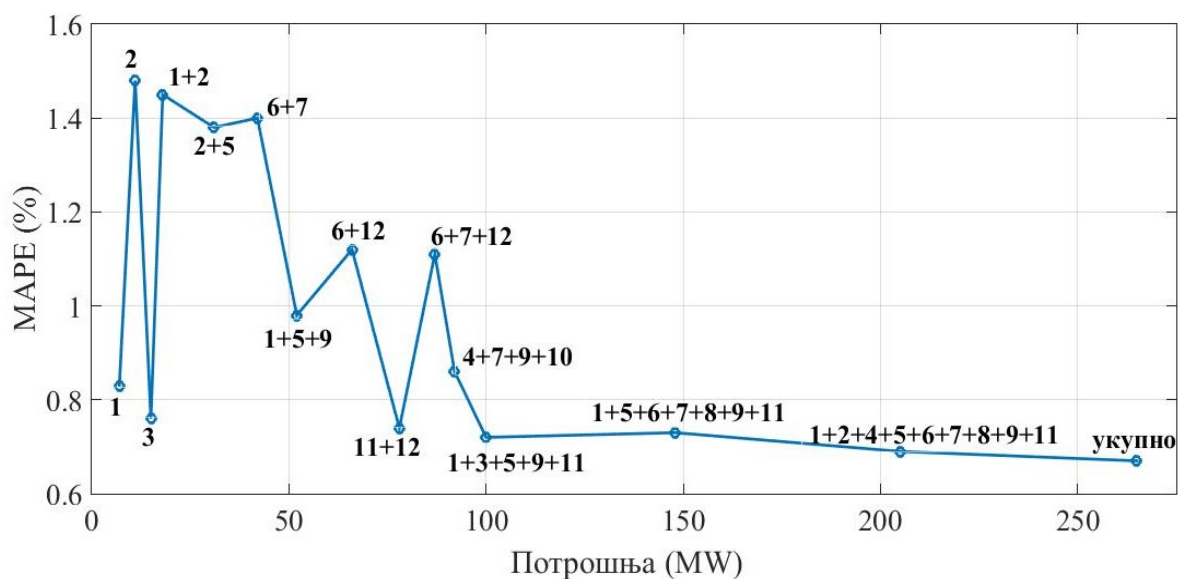
Слика 5.3.4: Резултати анализе резидуала модела потрошње 15. сата за *Cronulla* трафо рејон

Бројеви поред вредности на графику представљају редни број трафо станице у локалној бази (Табела 2.1), која је обухваћена дефинисаном облашћу. Резултати показују очекивану тенденцију смањења MAPE грешке са порастом нивоа потрошње. За ниже нивое потрошње су уочљиве разлике за различите типове потрошача: за област око RNS болнице, потрошње од 7 MW се добија већа тачност прогнозирања (0.83%) у односу на прогностичку област коју формира са резиденцијалним потрошачима на 50 MW (0.98%). Такође, за област око Олимпијског парка, који обухвата спортске и комерцијалне објекте, чија је укупна потрошња 15 MW, MAPE вредност је 0.76%. Овај ефекат постаје све мање изражен са већим агрегирањем потрошње и за област која обухвата све AU трафо рејоне, добијена MAPE грешка је 0.68%.

5.3.1 Кратак резиме

На примеру података за *Ausgrid* трафо рејоне, приказано је унапређење методологије предложене у тези за примену на потрошње са међусобно различитим факторима, чије укупно моделовање захтева већи број спољашњих варијабли. Унапређење се односи на 1) дефинисање генерализованог каталога ГАМ чланова којим су моделовани сви унапред идентификовани фактори потрошње и 2) њиховом селекцијом применом унакрсне валидације којим се каталог прилагођава конкретној потрошњи.

Резултати



Слика 5.3.5: Ефекат агрегације потрошње за АУ трафо рејоне на грешку прогнозе

Унакрсна валидација је, овде, уведена са циљем да сузи избор великог броја спољашњих променљивих које учествују у моделу, за шта је типичан пример група метеоролошких варијабли при моделовању ОИЕ утицаја на потрошњу у мрежи, за области за које је непознато да ли постоји и, ако постоји, колики је ниво пенетрације. За тестирање је коришћено 5 метео варијабли. Међутим, поступак добија на значају у пракси где је тај број знатно већи као у раду [4.12], где их је било 17. Коректни резултати анализе резидуала, за веома различите моделе добијене за различите прогностичке области, као и добра тачност прогнозе за овај ниво потрошње у систему, верификују предложени приступ.

Приказани пример развоја прогностичког алата за непознате услове прогнозирања које се односе на састав потрошача, ОИЕ присуство/одсуство, итд., представља генерализацију STLF функционалности и предложене методологије за примену на потрошњу у дистрибутивним системима.

5.4 Пример потрошње у зградама – EN и AS база података

Више од 30% укупно утрошене глобалне енергије и око 50% глобалне електричне енергије чини потрошњу у зградама, [5.16], [5.17]. С обзиром на ову статистику, процењено је да сектор зградарства има највећи потенцијал за уштеду енергије у оквиру програма енергетске ефикасности. Могућност смањења утрошка енергије за исте или боље услове живота у зградама се процењује на основу њиховог енергетског статуса, укључујући велики број мерења и информација о њиховим физичким карактеристикама. Кључну улогу у овим проценама имају енергетски модели зграда (*Building Energy Models, BEM*) и методологије прогнозирања енергије, који се користе за низ апликација, као, на пример, упоредне процене, прорачуне уштеде, генерисање прогноза као подлога за програме управљања потрошњом, итд.

С обзиром на важност тачних процена актуелног и будућег енергетског статуса у зградама, ASHRAE¹⁶ је у [5.18] увела стандардизацију мера и процедура за прорачун могуће уштеде потрошње енергије и воде у зградама. Између осталог, у овом документу се препоручују и начини евалуације предиктивних модела који се користе за прогнозирање

¹⁶ American Society of Heating, Refrigerating and Air-Conditioning Engineers

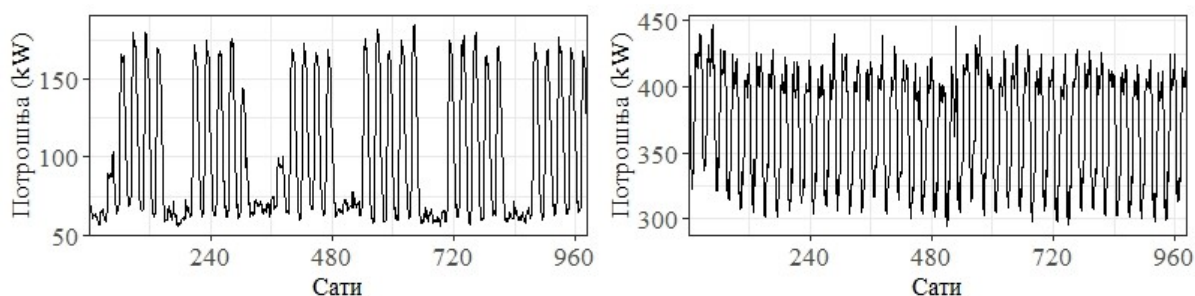
Резултати

енергије. Од важности за резултате у тези је што се овим стандардом, поред типичних начина процене предиктивних модела, за њихову евалуацију, уведе још два критеријума: *mean bias error* (МБЕ) и коефицијент варијације средњег квадратног одступања (*coefficient of variation of the root-mean-square error*, CVRMSE). За мерену вредност y_k и њену прогнозу, \hat{y}_k , ови критеријуми су дефинисани једначинама 5.6 и 5.7. По ASHRAE препорукама, сатни предиктивни модели прихватљивог квалитета испуњавају услове: МБЕ $\pm 5\%$, CVRMSE $< 15\%$ [5.19]

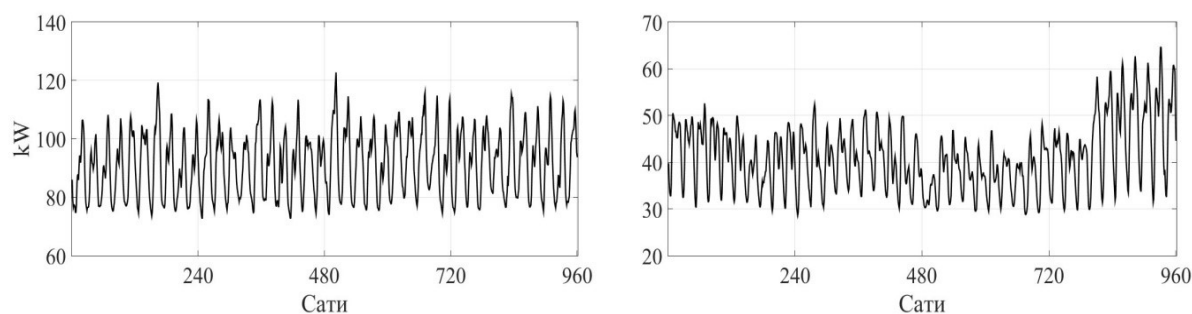
$$\text{MBE} = 100 \frac{\sum_{k=1}^N (y_k - \hat{y}_k)}{\sum_{k=1}^N y_k} \quad 5.6$$

$$\text{CVRMSE} = 100 \frac{\sqrt{\frac{1}{N-1} \sum_{k=1}^N \frac{(y_k - \hat{y}_k)^2}{y_k}}}{\frac{1}{N} \sum_{k=1}^N y_k} \quad 5.7$$

Полазећи од ASHRAE референтних вредности, у овом подпоглављу је тестирана предложена методологија на потрошњу електричне енергије на нивоу зграде, како би се испитала њена применљивост и за овај прогностички задатак. Тестирање је спроведено на малом узорку од 5 зграде различите намене, за које су подаци прикупљени из две базе, описане у подпоглављу 2.2.4. Сатне вредности за EN1, EN2, AS1 и AS3 зграде, за 40 узастопних дана приказане на сликама 5.4.1.-2.



Слика 5.4.1: Сатне потрошње за 40 дана за зграде EN1 и EN2



Слика 5.4.2: Сатне потрошње за 40 дана за зграде AS1 и AS3 зграду

С обзиром да су за све зграде из локалне базе на располагању једногодишњи подаци за потрошњу и остале метеоролошке варијабле, дефинисан је генерализовани каталог ГАМ

Резултати

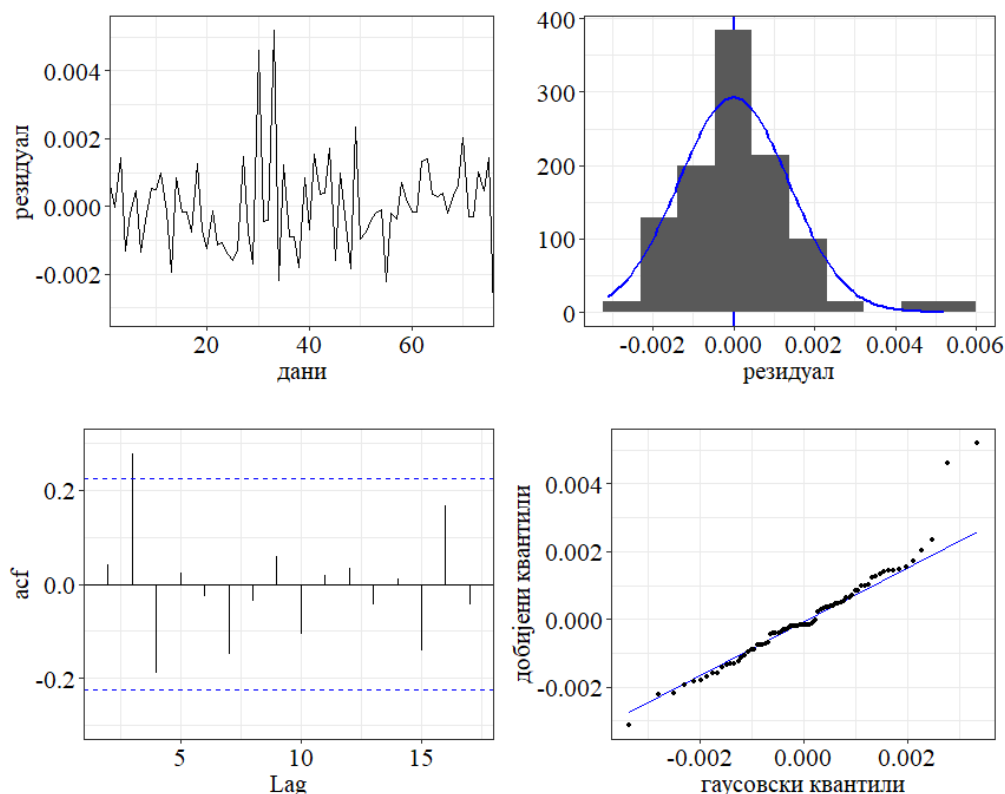
чланова којима се моделују унутар-годишњи фактори потрошње, као што је дато у Табели 5.4.1.

Табела 5.4.1: Каталог ГАМ чланова потрошње за зграде

ГАМ члан	Ефекат потрошње
$f_t(trend)$	Компонента тренда
$f(s)$	Седмични профил
$f_T(T_{x,h})$	Стандардни члан за моделовање зависности потрошње од температуре, где је x средња сатна температура, за актуелни и претходне сате $h = 0: -M_{Th}$.
$f_{SV}(SV_h)$	Зависност потрошње од секундарне метео варијабле за актуелни и претходне сате, $h = 0: -M_{SV}$.
$f^d(P^d)$	Утицај вредности потрошње из претходних дана, $d = -1: -D$
$f_h(P_h)$	Утицај вредности потрошње из претходних сати, $h = -1: -H$
$f_{xy}(x, y)$	Интеракција или заједнички ефекат x и y променљиве, које могу бити: $x = s, y = g$, промене седмичног профила током године $x = T, y = SV$, интеракција температуре и секундарне м. вар. $x = T, y = SV$

За EN1 и EN2 зграде су, ради компарације, временски периоди за које се тестирају резултати прогнозе усаглашени са онима из публикације [5.20], где се користи верзија дубоких мрежа за прогнозирање потрошње електричне енергије за исти сет података. Посматрани су резултати за зимски период (новембар до април) и летњи период (мај до октобар), конкретно, за EN1 зграду, 26. – 31. март и 11.-16. јун, за EN2, 2. – 7. април и 2.-7. јул. Димензије модела, односно, број ГАМ чланова су у овом примеру ограничени расположивим подацима који се користе за обучавање модела (видети коментар у подпоглављу 3.4). На пример, за EN1 зграду за први тест пример, за обучавање модела се могу користити само подаци за временски период од 1. јануара до 25. марта (84 дана). Зато је за моделовање из каталога одабран редуцирани скуп ГАМ чланова, који садржи само компоненту тренда (f_t), седмични профил (f_s), зависност од температуре за текући сат ($f_{T,0}$) и потрошње из претходног дана (f^{-1}) од пре 7 дана (f^{-7}) и претходног сата (f_{-1}). Овакав модел има ~50 коефицијената. Број података који се користио за обучавање модела је 75. Пре уласка у Процедуру 1, обављена је *Box-Cox* трансформација временских серија потрошње. Резултати анализе резидуала редуцираног модела (приказани на примеру потрошње за 15. сат, за EN1, на слици 5.4.3. су били без остатака тренда, са малим остацима корелације 14. 15. и 18. сат. Парцијални резидуали и одговарајуће EDF вредности су показали коректно димензионисане ГАМ чланове. Резултати MAPE и MAE грешке, за 1 и 24 сата унапред, CVRMSE и MBE, за летњи и зимски период, дати су Табелом 5.4.2.

Резултати



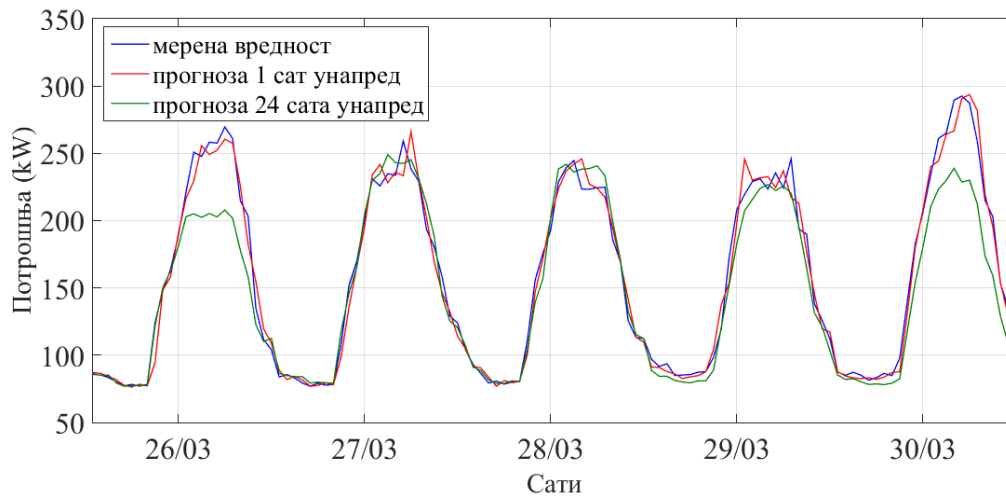
Слика 5.4.3: Резултати анализе резидуала модела потрошње 15. сата за EN1 зграду

Табела 5.4.2: MAPE, MAE, CVRMSE и MBE вредности за EN зграде, за летњи и зимски период

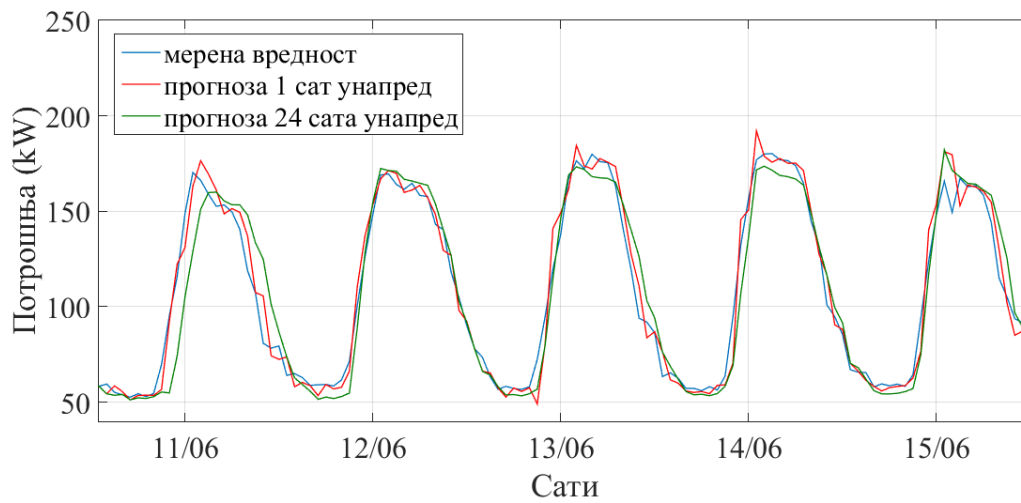
ID	MAPE _{ZI/L1}	MAE _{ZI/L1}	CVRMSE _{ZI/L1}	MBE _{ZI/L1}	MAPE _{Z24/L24}	MAE _{Z24/L24}
EN1	5.87/4.20	15,62/6,87	9.70/8.13	-2.7÷0.9/-0.8÷2.2	8.67/7.16	8.79/13.39
EN2	2.75/1.66	9.66/7.04	5.16/2.56	-1.2÷2.5/-0.9÷0.5	4.09/2.42	14.22/10.76

У поменутом референтном раду за развој функционалности прогнозе потрошње за зграде EN1 и EN2 коришћене су унапређење верзије дубоких неуралних мрежа, *Gated RNN (GRNN)* и *Gated CNN (GCNN)* модели. Од метеоролошких варијабли су кориштене вредности за температуру, влажност ваздуха, ваздушни притисак и брзина ветра. Резултати прогнозирања за један сат унапред показују да методологија предложена у тези даје већу тачност прогнозирања потрошње у односу на *GRNN* модел за летњи/зимски период за EN1 зграду (~30%) и летњи период за EN2 зграду, док је у односу на *GCNN* модел побољшање тачности прогнозирања добијено за летње месеце. На сликама 5.4.3-5.4.6 су приказани резултати прогнозе за обе EN зграде, за летњи и зимски тест период.

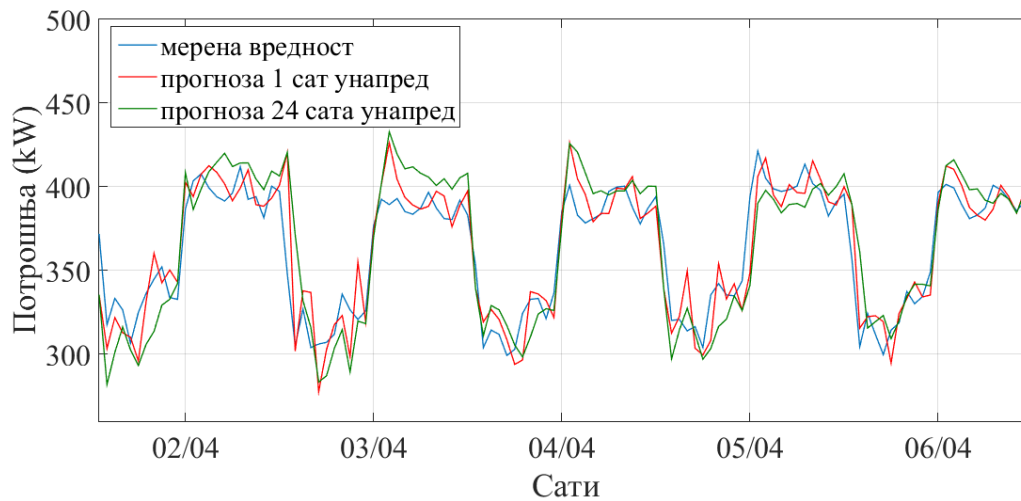
Резултати



Слика 5.4.4: Резултати прогнозе за EN1 зграду, за летњи период

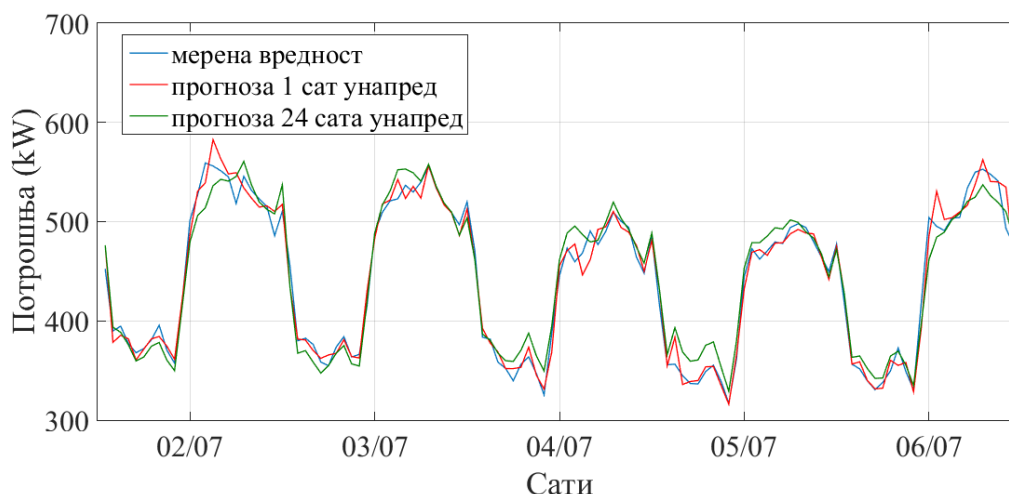


Слика 5.4.5: Резултати прогнозе за EN1 зграду, за зимски период



Слика 5.4.6: Резултати прогнозе за EN2 зграду, за зимски период

Резултати

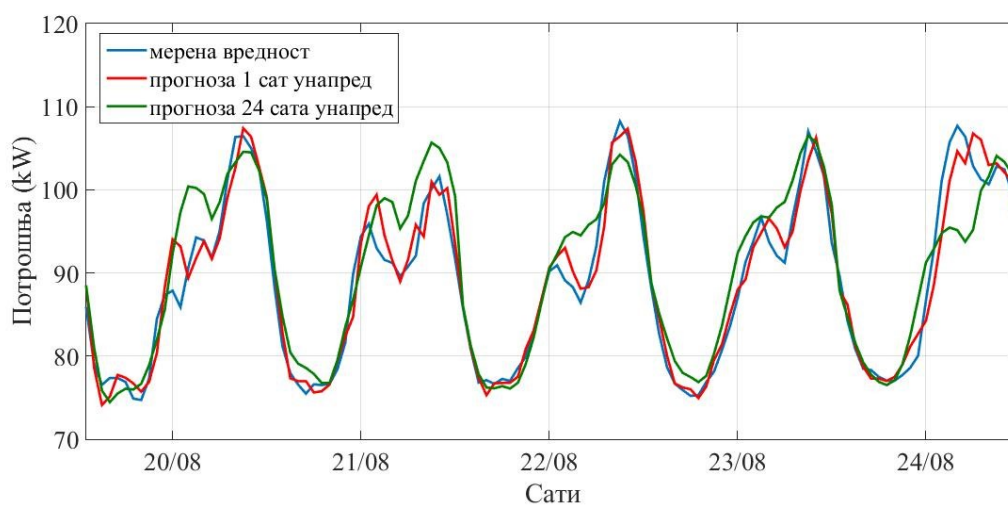


Слика 5.4.7: Резултати прогнозе за EN2 зграду, за летњи период

За моделовање потрошње за резиденцијалне *AS* зграде користиле су се сатне вредности за температуру, брзину ветра и влажност ваздуха. Поред чланова модела са имплементираним метеоролошким варијаблама, модел је укључивао и остале чланове из Табеле 5.4.1. Број података који се користио за обучавање модела је 200. Резултати прогнозирања су посматрани за период од 20.08 до 20.10. Резултати анализе резидуала моделовања, парцијалних резидуала и одговарајуће *EDF* вредности за сваку временску серију потрошње су били коректни.

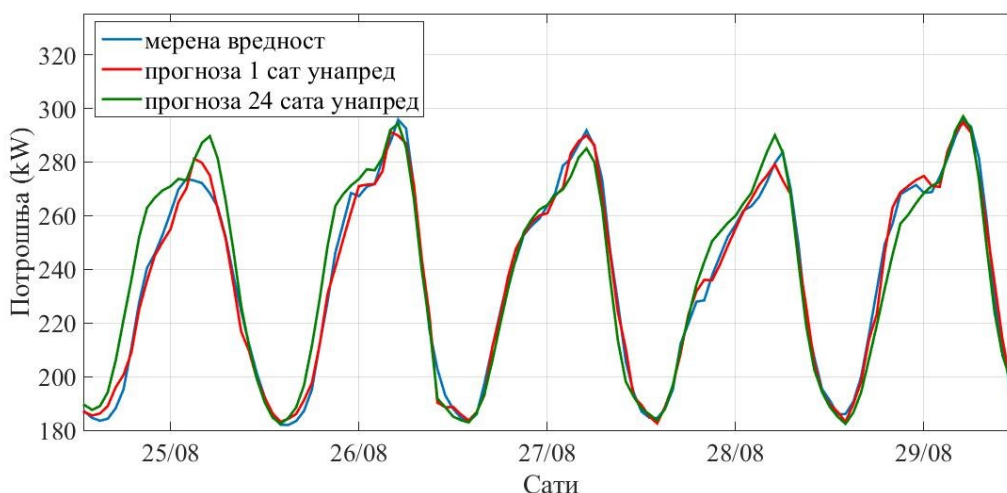
Табела 5.4.3: MAPE, MAE, CVRMSE и MBE вредности за *AS* зграде

building_ID	MAPE ₁	MAE ₁	CVRMSE	MBE	MAPE ₂₄	MAE ₂₄
AS1	2.03	1.86	2.86	-1.18÷1.36	4.02	3.70
AS2	1.49	3.46	2.24	-1.22÷1.26	3.81	8.85
AS3	2.47	1.11	3.28	-1.48÷2.29	6.33	2.82

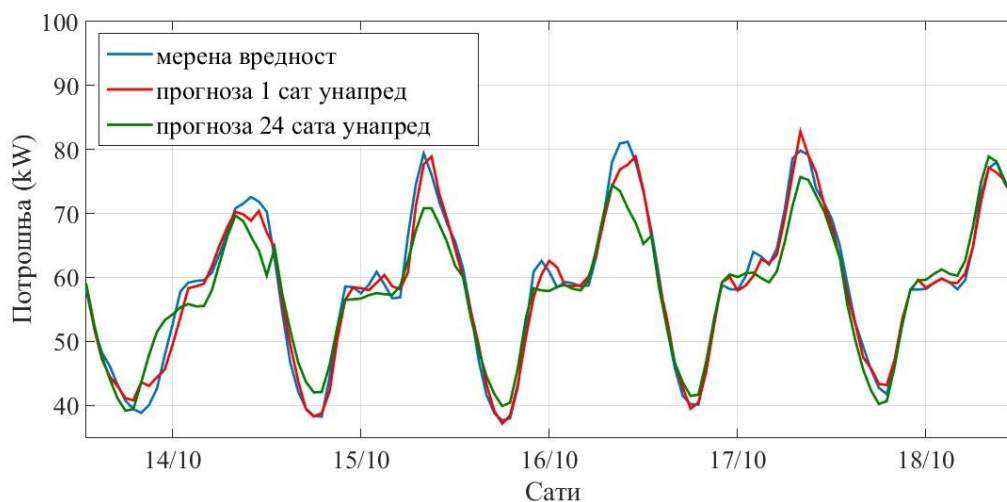


Слика 5.4.8: Резултати прогнозе за AS1 зграду

Резултати



Слика 5.4.9: Резултати прогнозе за AS2 зграду



Слика

5.4.10: Резултати прогнозе за AS3 зграду

Бољи до компарабилни резултати у односу на популарне технике моделовања, МВЕ, CVRMSE вредности које су у препорученим ASHRAE интервалима квалитета предиктивног модела, чак за ГАМ модел малих димензија у примеру са потрошњама за EN зграда, где се од спољашњих варијабли користила само средња сатна вредност температуре, верификује предложу методологију за примену на потрошњу електричне енергије у зградама.

5.5 Пример потрошње индивидуалног и произвољне агрегације потрошача – LCL база података

Мерења потрошње са паметних бројила индивидуалног потрошача су од значаја, не само што омогућавају прогностички задатак и низ апликација на нивоу индивидуалне потрошње, него се и њиховим агрегирањем, омогућава формирање прогностичких области за произвољан број потрошача. У зависности од конкретног задатка прогнозирања, начини моделовања мерених потрошњи су различити и неке од њих су наведене у подпоглављу (Поглавље 4.3). Две су основне примене *STLF* прогностичких модела индивидуалне потрошње. Једна је стандардно *STLF* прогнозирање код кога се максимизира тачност

Резултати

прогнозе за потрошњу која се мери и за коју се и развија модел, било индивидуална или она добијена агрегирањем индивидуалних. Друга примена је тзв. *bottom-up* приступ, где се прогнозе локалних потрошњи, индивидуалних или на мањим нивоима агрегације, сабирају са циљем побољшања тачности прогнозе на вишим нивоима агрегације потрошњи. Квалитет прогнозирања за обе примене суштински зависи од тачности прогнозе на нивоу индивидуалне потрошње.

У наставку је приказана примена (X)ГАМ модела за карактеристичне приступе развоја *STLF* прогноза базираних на мерењима са паметних бројила. За тестирање предложене методологије користили су се подаци из LCL базе. Како је у овом примеру превасходно интересантно направити компарацију различитих примена (X)ГАМ моделовања на индивидуалну потрошњу, посматрају се резултати прогнозе само за један сат унапред, евалуирани MAPE грешком. Такође, многи аспекти анализе модела остају исти као и у претходним примерима.

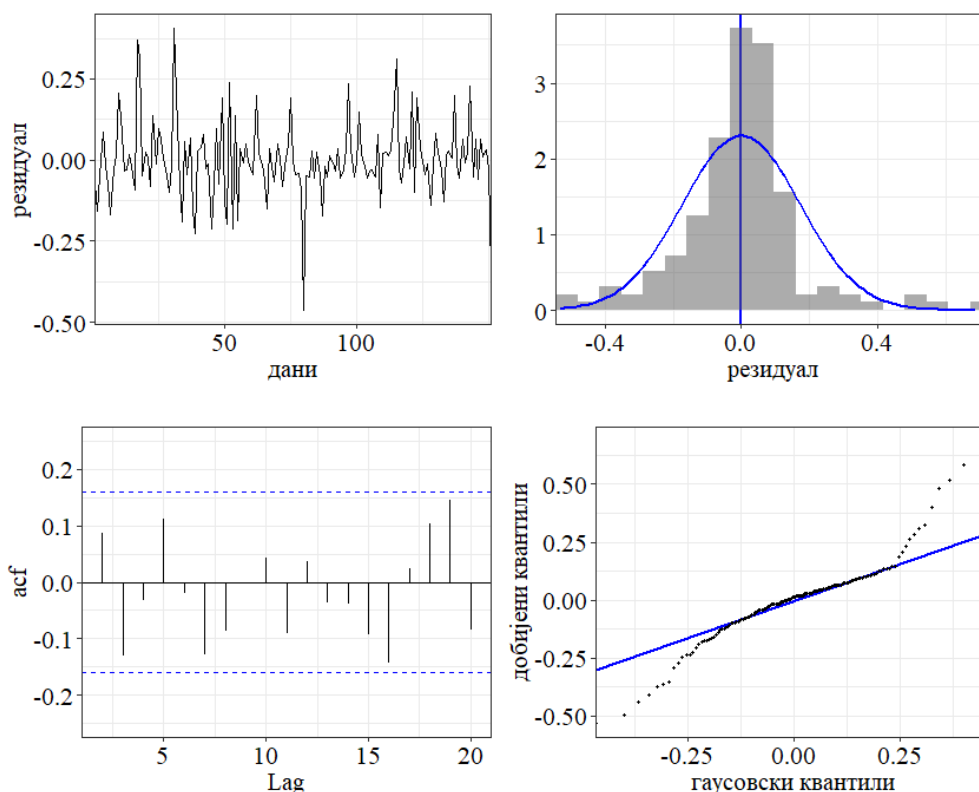
ГАМ модел се, у овом случају, користи за примену на независну временску серију потрошње на нивоу индивидуалног потрошача или репрезента (средње вредности или збира) више индивидуалних потрошњи. За LCL сатне вредности за потрошњу и температуру за 2013. годину, дефинисан је скуп стандардних ГАМ чланова за моделовање фактора потрошње унутар године, дат Табелом 5.5.1. Модели су развијени за трансформисане потрошње (*Box-Cox* трансформација), број података који се користио за обучавање модела је 120, док је период тестирања обухватао 5 месеци (мај – септембар).

Табела 5.5.1: Каталог ГАМ чланова потрошње за индивидуалног потрошача

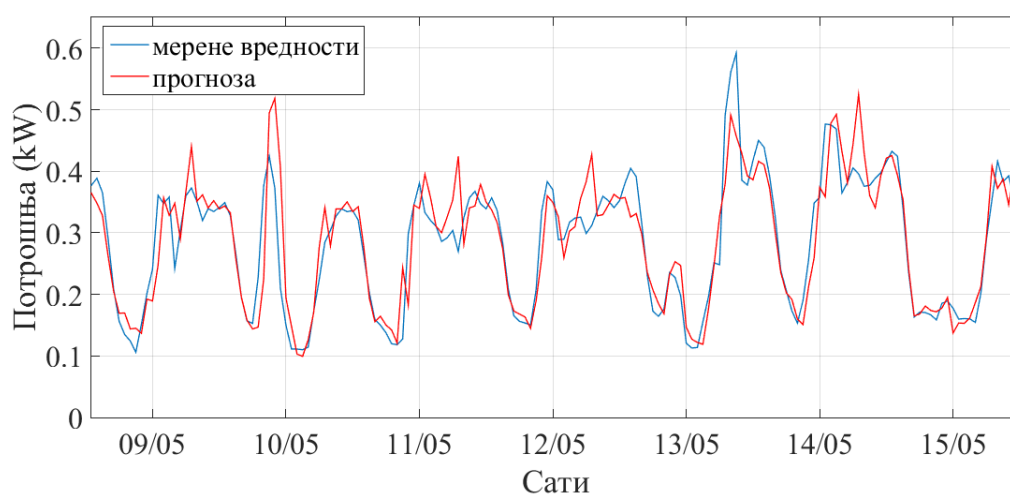
ГАМ члан	Ефекат потрошње
$f_t(trend)$	Компонента тренда
$f_s(s)$	Седмични профил
$f_T(T)$	Средња дневна вредност температуре
$f^d(P^d)$	Вредности потрошње у тренутку t , за дане $d - 1$ и $d - 7$
$f_h(P_h)$	Вредности потрошње из претходна $t - 1$

Пример 1: Пре свега, посматран је случај директне примене ГАМ методологије за индивидуалну потрошњу, за произвољно одабраног LCL потрошача. На слици 5.5.1 су приказани резултати анализе резидуала за 15. сат, а на слици 5.5.2 потрошња и резултати њене прогнозе за период од 09.05 до 15.05. Добијена MAPE грешка прогнозе је била **13.23%**, што упада у интервале резултата који се наводе у литератури (5%-43%). Међутим, директна примена предложене методологије на сваку индивидуалну потрошњу за коју се развија *STLF* функционалност није практично интересантна (поглавље 4.3) и добијени резултат служи као референца за следеће тест примере.

Резултати



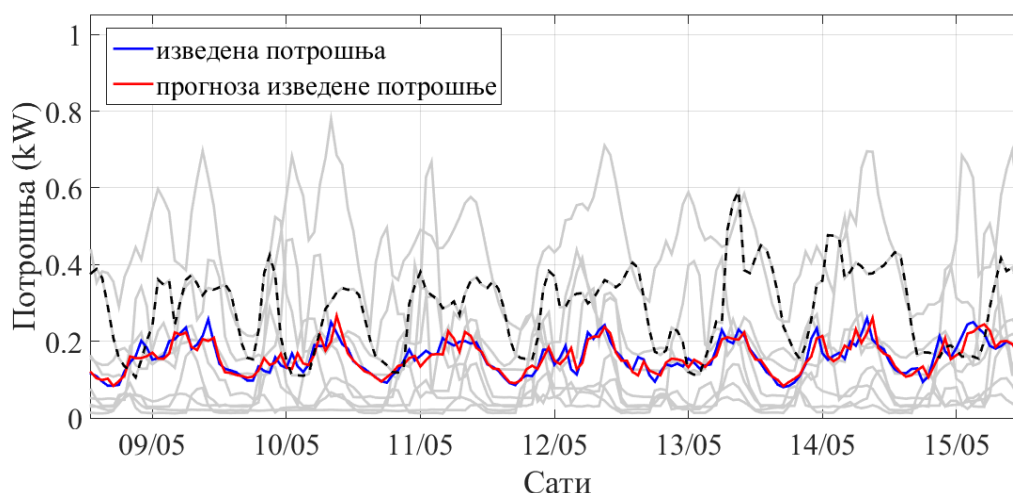
Слика 5.5.1: Резултати анализе резидуала ХГАМ модела потрошње 15. сата, за одабраног LCL потрошача



Слика 5.5.2: Резултати ГАМ прогнозе за одабраног потрошача

Пример 2: Други тест је заснован на груписању потрошача, издвајању репрезентативне временске серије потрошње и рад са изведеном потрошњом (поглавље 4.3). Овај пример је разматран за хетерогену и хомогену групу потрошача.

2.а. У првом примеру се ГАМ модел развија за средњу потрошњу групе од 10 произвољних потрошача, која укључује и потрошача из претходног примера. Исти сет ГАМ чланова и из Табеле 5.5.1 је кориштен за моделовање средње вредности трансформисане потрошње, за исти период обучавања.



Слика 5.5.3: Резултати ГАМ прогнозе репрезента потрошње за произвољну групу потрошача

На слици 5.5.3 су приказани потрошња репрезента (плаво), њена прогноза (црвено) и потрошња одабраног потрошача (црно испрекидано), за период од 09.05 до 15.05. МАРЕ грешка прогнозирања изведене потрошње је **10.04%**, док је она која је направљена у односу на одабраног потрошача **41.11%**. Средња МАРЕ грешка за све потрошње из групе у односу на прогнозу репрезента је била **38.32%**.

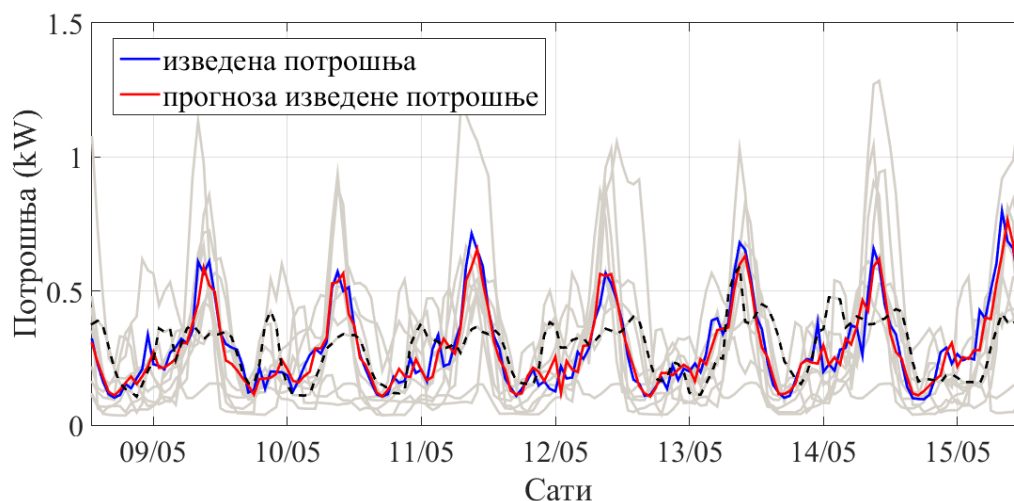
2.6. Претходни тест је поновљен за хомогену групу потрошача, формирану на бази сличности профила потрошњи. Као мера сличности се користио коефицијент корелације, дефинисан формулом 5.8, где су $L_{1,i}$ и $L_{2,i}$ i -ти узорци временске серије потрошње, средњих вредности \bar{L}_1 и \bar{L}_2 , за посматраног и произвољног LCL потрошача, респективно, [5.21]. Хомогена група обухвата посматраног и 9 LCL потрошача са највећим r коефицијентом корелације.

$$r = \frac{\sum(L_{1,i} - \bar{L}_1)(L_{2,i} - \bar{L}_2)}{\sqrt{\sum(L_{1,i} - \bar{L}_1)^2 \sum(L_{2,i} - \bar{L}_2)^2}} \quad 5.8$$

Сликом 5.5.3. је приказана потрошња репрезента (плаво), њена прогноза (црвено) и потрошња одабраног потрошача (црно испрекидано), за период од 09.05 до 15.05. МАРЕ грешка прогнозирања изведене потрошње је, у овом примеру, била **9.83%**, док је она у односу на посматраног потрошача била **32.87%**. Средња МАРЕ грешка за све потрошње из групе у односу на прогнозу репрезента је била **29.83%**. Смањење грешака прогнозе у односу на претходни пример је очекивано с обзиром на сличност профила потрошње унутар хомогене групе и на предвидљивије профиле њене средње потрошње.

Иако прогнозе за изведене репрезенте дају добре резултате за ниво укупне (или средње) потрошње за 10 индивидуалних потрошача, и применљиви су за конкретне прогностичке задатке, грешка прогнозирања која се прави у односу на потрошњу индивидуалног потрошача захтева унапређење. У наставку је тестирана примена ХГАМ модела на хетерогену и хомогену групу потрошача из претходна два теста, са циљем да се испита могућност побољшања добијених резултата.

Резултати



Слика

5.5.4: Резултати ГАМ прогнозе репрезента потрошње за хомогену групу потрошача

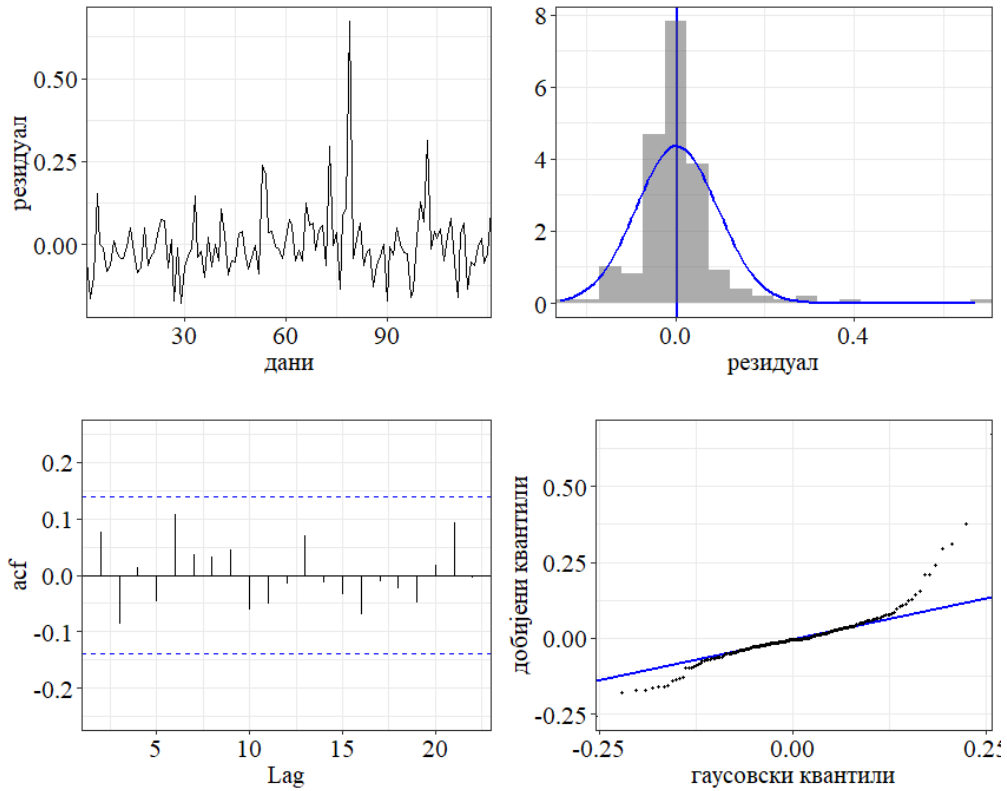
Пример 3: За разлику од ГАМ модела, ХГАМ се обучава за укупан сет потрошњи из групе и уводи претпоставке о њиховој међусобној корелисаности. Свака временска серија потрошње из групе је идентификована као засебна категорија података, тако да је за урачунавање ХГАМ чланова модела, по једначини 3.5, $G = 10$. На даље је претпостављено да сви фактори потрошње, моделовани ГАМ члановима из Табеле 5.5.1, могу варирати за различите потрошње, тако да се сваком од њих додаје еквивалентан ХГАМ члан, као што је дато у Табели 5.5.2.

Табела 5.5.2: Каталог ХГАМ чланова потрошње за индивидуалног потрошача

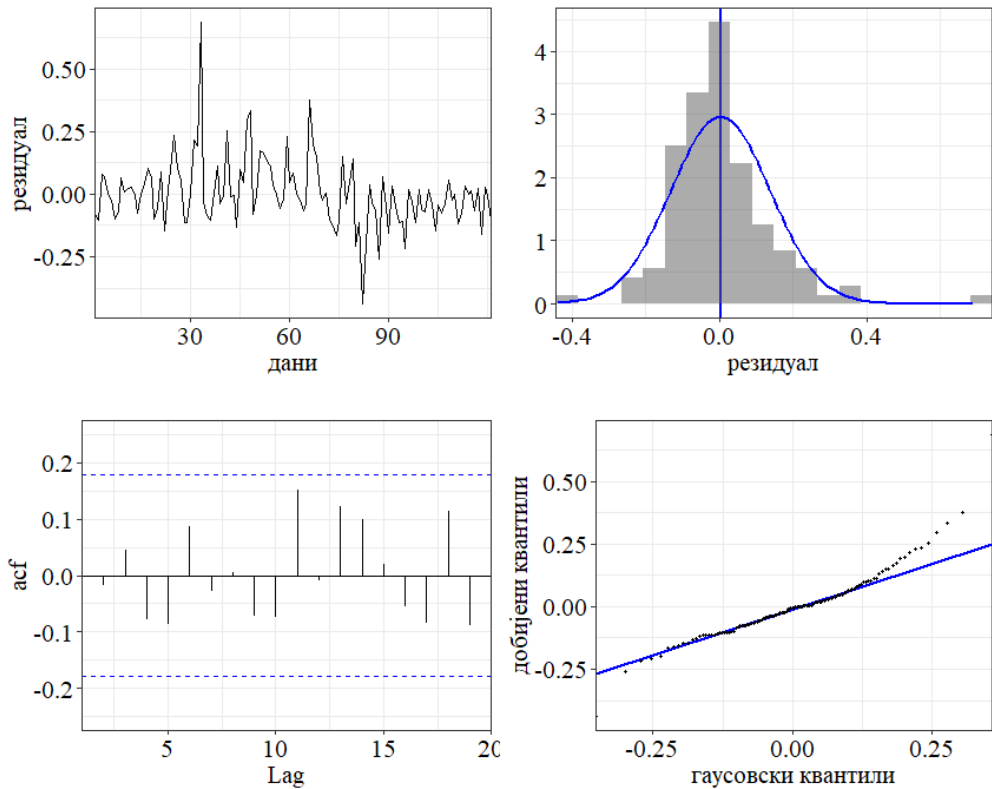
ХГАМ члан	Ефекат потрошње
$f_t(trend, G)$	ХГАМ члан компоненте тренда
$f_s(s, G)$	Седмични профил
$f_T(T, G)$	Средња дневна вредност температуре
$f_2(P_t^{d-1}, G), f_7(P_t^{d-7}, G)$	Вредности потрошње у тренутку t , за дане $d - 1$ и $d - 7$
$f_7(P_{t-1}, G)$	Вредности потрошње из претходна $t - 1$

Естимација ХГАМ модела и генерисање прогноза за све потрошње из групе се обавља симултано. За одабраног потрошача, резултати анализе резидуала, за случај произвољне и хомогене групе су дати сликама 5.5.5 и 5.5.6, резултати прогнозе за период од 09.05 до 15.05. на слици 5.5.7., са МАРЕ грешком од **10.21%** и **9.46%**, респективно. Поређењем резултата прогноза на излазу из стандардног ГАМ модела (Пример 1) у односу на резултате добијене ХГАМ-ом, за индивидуалног потрошача, МАРЕ грешка се смањила за око 30%, односно скоро 40%, за хетерогену и хомогену групу потрошача, респективно.

Резултати

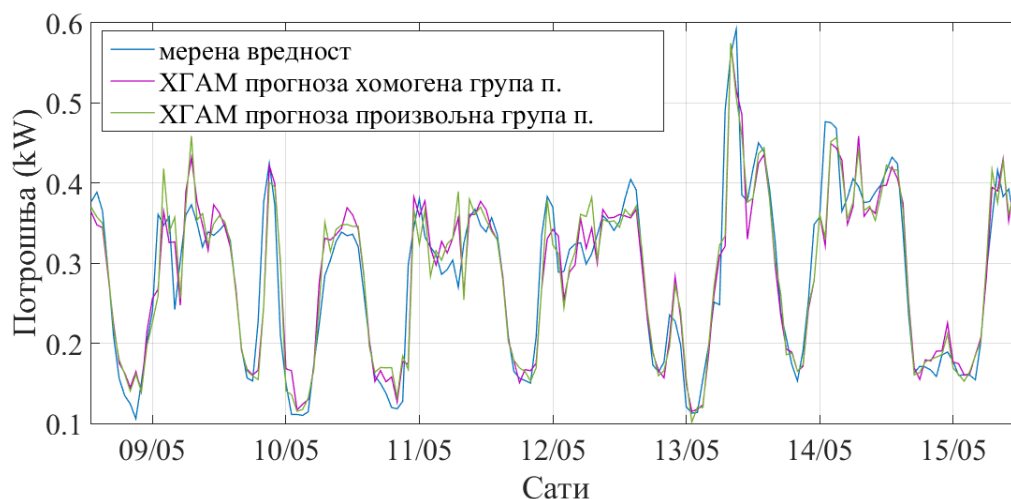


Слика 5.5.5: Резултати анализе резидуала ХГАМ модела потрошње 15. сата, за одабраног LCL потрошача из произвољне групе потрошача



Слика 5.5.6: Резултати анализе резидуала ХГАМ модела потрошње 15. сата, за одабраног LCL потрошача из хомогене групе потрошача

Резултати



Слика 5.5.7: Резултати ХГАМ прогнозе индивидуалног потрошача из произвољне и хомогене групе потрошача

Коначно, израчуната МАРЕ грешка између средње вредности (збира) добијених прогноза свих потрошњи, за произвољну и хомогену групу, у односу на реалне средње вредности је била **7.38%** и **5.76%**, што је преко 50% побољшања у односу на приступ директног моделовања средње потрошње из Примера 2.

Приказани резултат, пре свега, потврђују да се урачунавањем тзв. просторне корелације између индивидуалних потрошњи (подобласт 4.4), добијају побољшања тачности њиховог прогнозирања и то побољшање је веће за више корелисане потрошње. Побољшања се односе како за индивидуалну потрошњу, тако и за средње (укупне) вредности групе потрошњи, од значаја у пракси, за прогнозирање репрезентативне потрошње групе потрошача или за *bottom-up* сценарио прогнозирања потрошње њихове агрегације. За резултате у тези је важно је да је то побољшање могуће добити коришћењем ХГАМ модела. Практична примена добијених резултата подразумева да се потрошачи превасходно групишу у хомогене групе, типично, неким од поступака кластеризације, и сваки од кластера се моделује ХГАМ моделом дефинисаним члановима из Табела 5.5.1 и 5.5.2. Описани поступак, иако типичан за индивидуалне потрошње, концептуално је у потпуности применљив и за примену на потрошње већег броја зграда, трафо рејона, итд.

5.5 Кратак резиме

У претходним подпоглављима су дати резултати тестирања предложене методологије на примерима агрегираних потрошњи, нерезиденцијалних и резиденцијалних зграда, индивидуалне и произвољне агрегације индивидуалне потрошње. Основни фактори потрошње су моделовани ГАМ члановима из Табела 5.2.1, 5.3.1, 5.4.1, 5.5.1 и 5.5.2, за све приказане примере. Уз Процедуру 1(2), генерализовани каталог којим су обухваћени сви ГАМ чланови из поменутих табела представља основу оквира за краткорочну прогнозу потрошње у интелигентним електр енергетским мрежама. Прилагођавање каталога конкретном прогностичком задатку, зависи од:

- *расположивих мерења спољашњих променљивих*, што подразумева различите типове варијабли и различите интервале мерења. Прилагођавање се обавља једноставним уврштавањем оних променљивих у модел за које постоји мерење. Додавање секундарних метеоролошких и осталих неметеоролошких варијабли, као и њихових интеракција, постиже се проширивањем генерализованог каталога додатним ГАМ члановима. За

Резултати

велики број спољашњих променљивих, могуће је користити поступак унакрсне валидације.

- *расположивих историјских података* који утичу на величину скупа података који се користи за обучавање модела. За моделе који се обучавају подацима од више година, из предложених каталога се могу преузети сви ГАМ чланови, док је за оне који се обучавају подацима који обухватају интервале унутар године, потребно искључити чланове за годишњи профил, специјалне дане и *day-light savings*. И, иначе, моделом се обухватају они ГАМ чланови

Додатно, предложени STLF оквир укључује:

- поступак препроцесирање података за потрошње и спољашње променљиве (отклањање недостајућих података и филтрирање) и *Box-Cox* трансформацију.
- методологија груписања потрошача у хомогене групе за примену на скуп корелисаних потрошњи. Ово подразумева развој поступка кластеризације, која уважава перформансе процедуре прогнозирања, предложене у тези.

6. Закључак

Предмет истраживања докторске тезе је развој функционалности краткорочне прогнозе потрошње електричне енергије нове генерације, пројектоване за примену у интелигентним електроенергетским мрежама. Алати за генерисање прогноза који су у употреби, махом су намењени централизованом управљању на нивоу система или веће групе потрошача, развијани и прилагођени подацима за потрошњу прикупљеним са, типично, једног места у систему. Нови услови децентрализованог и локалног управљања у интелигентним енергетским мрежама, захтевају нову *STLF* функционалност која обезбеђује задовољавајућу тачности прогнозе за више различитих прогностичких области (различитог типа и броја потрошача и осталих фактора који утичу на ниво утрошене енергије), за променљиве услове прогнозирања. Са аспекта реализације конкретне функционалности, то значи да је потребно обезбедити већу прилагодљивост постојећих модела различитим потрошњама, (поготово за различите агрегације потрошњи) и у времену, како би се очувао квалитет прогнозе.

Досадашња истраживања која се баве *STLF*-ом на нивоу система, карактеристична за традиционално управљање системима, довела су до задовољавајућих резултата. На располагању је велики број приступа, од којих су многи верификовани у пракси и у употреби. Изазови који постоје код развоја *STLF* у дистрибуцији и на нивоу нисконапоснке мреже се односе, пре свега, на повећану комплексност профила потрошње мањег броја потрошача и већу динамику њихових промена услед утицаја ОИЕ производње, нових потрошача, итд., која је значајније изражена на локалном нивоу. На даље, у пракси, са аспекта управљања већим бројем локалних потрошњи, за, понекад, променљиве прогностичке области, развој засебне функционалности је неодржив. То значи да приступ мора да има тај ниво генералности да буде применљив за више потрошњи у систему, мерено потребном тачности прогнозирања. Досадашња истраживања *STLF* функционалности локалних потрошњи, нуде махом појединачна решења, евалуирана за конкретне, често, ограничене сетове података, код којих је генералност упитна, што се и истиче у закључцима публикованих радова. Пре свега, постоје реалне препреке у недоступности релевантних података које би омогућили другачији закључак, а потом и у ограниченој адаптивности постојећих модела на различите потрошње. Повећани интерес за напредним техникама дубоког учења, које преовлађују у скоријој литератури, поготово за моделовање потрошње индивидуалног потрошача, је у њиховом обећавајуће великом капацитету и могућности моделовања великог броја функција. Оно што се може закључити на основу укупних резултата је да за дати сет података, добро димензионисана дубока неурална мрежа има највећи потенцијал да да бољу тачност моделовања у односу на алтернативан приступ. Димензионисање дубоке неуралне мреже подразумева, генерално, дуготрајно, *a priori* подешавање великог броја хиперпараметара, за унапред одабрани сет репрезентативних података. Кључно питање је у којој мери унапред дефинисан модел великог капацитета може да генерализује будуће промене потрошње и очува квалитет прогнозе у времену. Последично, један део истраживања је посвећен унапређивању дубоких неуралних мрежа у правцу њихове примене за рад у реалном времену, док се други део истраживања бави сврсисходном припремом података и применом мање комплексних модела, са брзим алгоритмима за обучавање и прогнозирање. У поређењу ова два приступа, у прегледаној литератури, не постоји јасан консензус и закључци варирају од публикације до публикације. Оно што јесте генерални закључак је да се проблем прогнозирања локалне потрошње преваходно постиже одговарајућом методологијом, без обзира на избор модела. Услов да

методологија буде верификована је њена евалуација за различите скупове података, а за генерални приступ који би послужио у интелигентним мрежама, потребно је да буде применљив на различитим местима у систему и за различите агрегације потрошача. Свеобухватна истраживања генералне *STLF* функционалности ових карактеристика су скоријег датума и неки од њих су наведени у уводу. На основу закључака из поменутих радова, потребна су даља унапређења, поготово у правцу развоја динамичког прилагођавања модела променљивим условима прогнозирања. Полазећи од укупних резултата из литературе, и унапред задате спецификације оптималног прогностичког алата, користећи се подацима из отворених база, у тези се предлаже методологија за развој функционалности краткорочне прогнозе потрошње електричне енергије која би задовољила услове и критеријуме рада у интелигентним мрежама.

6.1 Преглед научних доприноса

Доприноси ове докторске дисертације су:

- обављен је детаљан преглед литературе из предметне области, са циљем да се одабере модел и/или методологија за развој нове *STLF* функционалности. На основу резултата репрезентативних истраживања извршена је категоризација различитих приступа моделовања и идентификација оних који имају најбољу перспективу за даљи развој. Полазећи од њихових недостатака или недостатака одговарајућих истраживања, као оптимално решење, одабран је алгоритамски унапређен генерализовани адитивни модел, ГАМ, и његово проширење на хијерархијску верзију, ХГАМ. Мотиви оваквог избора су, пре свега, задовољавајућа тачност краткорочне прогнозе добијене ГАМ-ом за појединачне потрошње у ЕЕС-у и флексибилност модела и пропратних алгоритама који омогућавају развој нове *STLF* функционалности.
- претражене су отворене базе података и прикупљени подаци за потрошње са, пре свега, географски различитих области, за различит број потрошача и за статистички различите категорије потрошње (агрегирана, неагрегирана и потрошња за зграде). Формирана је локална база података која садржи податке за потрошњу за 11 већих градских зона, 12 трафо рејона, 5 зграда (резиденцијалних, комерцијалних и јавних објеката) и 10 индивидуалних потрошача. Поред тога, приказани резултати су добијени за различит скуп метеоролошких променљивих, који обухвата дневне вредности температуре (средње, максималне/минимална), средње сатне вредности за температуру, брзину ветра, влажност, облачност, итд.

Кључни научни допринос тезе је развој ГАМ базираног генеричког оквира за моделовање и краткорочну прогнозу потрошње електричне енергије у интелигентним електроенергетским мрежама. Он се може формулисати кроз 4 ставке:

- дефинисање генерализованог каталога (Х)ГАМ чланова модела који имплементирају основне факторе потрошње. Спроведена је анализа временских серија потрошње за различита места унутар ЕЕС и коришћена установљена знања о потрошњи за њихово идентификовање и дефинисање одговарајућих ГАМ чланова. Специфичност ове анализе је увид у разлике за различите агрегације потрошача, као и уважавање просторне корелације између потрошњи као фактора чијим се урачунавањем побољшава тачност прогнозирања.

Закључак

- развој процедуре динамичког моделовања, засноване на динамичком избору ГАМ чланова који је интегрисан у брзу естимациону процедуру. Реализоване су Процедура 1 (2), које, на основу расположивих историјских података, за унапред задати слуп (X)ГАМ чланова, аутоматски, прилагођава модел подацима и у времену и врши прогнозу наредне вредности.
- имплементација *STLF* оквира у *R* језику, коришћењем пакета намењених развоју (X)ГАМ модела, које подразумева:
 - имплементацију нових алгоритама за брзу селекцију и естимацију ГАМ-а,
 - имплементацију каталога (X)ГАМ чланова потрошње.
- резултати прогнозирања за сат и дан-унапред, бољи или компарабилни у односу на добијене у прегледаној литератури. Методологија је евалуирана за произвољне агрегиране потрошње, нерезиденцијалне и резиденцијалне зграде, индивидуалног и групе индивидуалних потрошача, чиме су обухваћене све идентификоване, статистички различите, категорије потрошњи у ЕЕС-у.

Основна предност предложене методологије је што комплексан задатак развоја функционалности краткорочног прогнозирања за више потрошњи са различитих места у систему и различите нивое агрегације, своди на мање захтеван задатак дефинисања ГАМ каталога. Свака даља примена методологије на произвољну потрошњу у систему подразумева селекцију или проширивање каталога зависно од расположивих мерења спољашњих променљивих и расположивих података за обучавање модела. За идентификоване факторе потрошње, ГАМ чланови се лако имплементирају, без рестриктивних улазних претпоставки или критичних хиперапараметара.

6.2 Практична примена

- Основна примена развијеног оквира је генерисање практично применљивог *STLF* алата за конкретно место у ЕЕС-у. Посебан допринос приказаних резултата је што уважава разлике у хијерхији потрошње за различит број потрошача за потрошње са статистички различитим карактеристикама, 1) на нивоу произвољне агрегиране потрошње, 2) на нивоу нерезиденцијалних и резиденцијалних зграда и 3) на нивоу индивидуалног и произвољне групе индивидуалних потрошача. С обзиром да реални услови прогнозирања могу бити условљени низом оперативних захтева, корисницима се препоручује коришћење више *STLF* алата или неки вид хибридизације више приступа. У том смислу, методологија која је развијена у тези представља солидну основу која омогућава несметано прогнозирање и сваки вид симулације, као и додатну анализу потрошње у управљању системом.
- Убрзаном имплементацијом концепата интелигентних мрежа, постоји потреба за развојем платформе за симулацију рада ЕЕС-а и валидацију прогностичких алата за различита *what-if* сценарија. Будући конфигурабилан, једноставан за имплементацију, заснован на ширем статистичком оквиру који омогућава додатне анализе резултата прогнозе и моделовања и једноставан за даље надограђивање, развијени оквир представља препоручено окружење за развој симулатора.

- Шире гледајући, ван проблематике којом се бави теза, ГАМ модели се, примарно, користе за издвајање појединачних подкомпоненти, односно процену модела на нивоу једног или више ГАМ чланова. Ова могућност модела се може искористити за задатке декомпоновања потрошње на подкомпоненте, као на пример, издвајање дела потрошње само услед метеоролошких услова или карактеристичних профила потрошње и даљу анализу ‘понашања’ потрошње.

6.3 Ограничење истраживања

Резултатима приказаним у тези, до комплетне *STLF* функционалности за примену у интелигентним мрежама недостаје методологија одређивања предиктивних интервала. Кориснику алата је од великог значаја не само информација о очекиваној потрошњи, него и опсег вероватних вредности који помажу процени ризика, поготово у време дневног максимума потрошње.

6.4 Правци даљег истраживања

Правци даљег развоја се деле на неминовне и препоручљиве. Неминовно је да даље истраживање треба да реши недостатке приступа предложеног у тези. То се, пре свега, односи на:

- развој методе динамичког прорачуна предиктивних интервала, који би омогућили да се са сваком испоруком прогнозираних вредности, дају и резултати вероватних опсега. Поред стандардних методологија које се, иначе, користе и имплементирају за различите моделе, за ГАМ постоје и алгоритамска проширења одговарајућим пакетима, као што су *GAMLSS*, *quantGAM*, итд., који омогућавају прорачун предиктивних интервала.
- софистицираније моделовање метеоролошке компоненте потрошње, којом би се поред проблематике великог броја варијабли размотрили и остали аспекти њиховог прикупљања и моделовања као, на пример, за оне прогностичке области за које се метео параметри преузимају са више мерних места.

Поред превазилажења недостатака предложене методологије, даље унапређење ХГАМ модела које би омогућило комплетну функционалност прогнозирања индивидуалних потрошача, је сигуран наредни корак. То подразумева развој метода поделе потрошача у хомогене групе, типично, неким од поступака кластеризације и примена развијене ХГАМ методологије на прогнозу потрошње за сваку од група.

Референце

- 1.1 G. Boyd, SPEN – DSO Vision, 24th International Conference & Exhibition on Electricity Distribution (CIRED), 12-15 June 2017.
- 1.2 Utility analytics in 2017: Aligning data and analytics with business strategy, SAS.
- 1.3 W. Ma, S. Fang, G. Liu, R. Zhou, "Modeling of district load forecasting for distributed energy system," in *Applied Energy*, vol. 204, pp. 181-205, 2017.
- 1.4 V. Veeramsetty, R. Deshmukh, "Electric power load forecasting on a 33/11 kV substation using artificial neural networks," in *SN Appl. Sci.*, vol. 2, p. 855, 2020.
- 1.5 Y. Liu, W. Wang, N. Ghadimi, "Electricity load forecasting by an improved forecast engine for building level consumers", in *Energy*, vol. 139, pp. 18-30, 2017.
- 1.6 B. Hayes, J. Gruber and M. Prodanovic, "Short-Term Load Forecasting at the local level using smart meter data," in *IEEE Eindhoven PowerTech*, Eindhoven, Netherlands, pp. 1-6, 2015.
- 1.7 J. C. López, M. J. Rider, Q. Wu, "Parsimonious Short-Term Load Forecasting for Optimal Operation Planning of Electrical Distribution Systems," in *IEEE Transactions on Power Systems*, vol. 34, no.2, pp. 1427 - 1437, 2018.
- 1.8 K. Ullah *et al.*, "Load Forecasting Schemes and Demand Response Programs within Smart Grid," in *2020 International Conference on Electrical, Communication, and Computer Engineering (ICECCE)*, pp. 1-6, 2020.
- 1.9 G. Fenza, M. Gallo, V. Loia, "Drift-Aware Methodology for Anomaly Detection in Smart Grid," in *IEEE Access*, vol. 7, pp. 9645-9657, 2019.
- 1.10 Y. Wanjun, et al., "Study on orderly charging strategy of EV with load forecasting," in *Energy*, vol. 278, p. 127818, 2023.
- 1.11 Z. Dong, et al., "Hourly energy consumption prediction of an office building based on ensemble learning and energy consumption pattern classification," *Energy and Buildings*, vol. 241, p. 110929, 2021.
- 1.12 Y. Wang, et al., "Review of smart meter data analytics: Applications, methodologies, and challenges," in *IEEE Transactions on Smart Grid*, vol. 10, no. 3, pp. 3125-3148, 2018.
- 1.13 C. Ramón, et al., "Exogenous Data for Load Forecasting: A Review," in *IJCCI*, pp. 489-500, 2020.
- 1.14 N. Ahmad, Y. Ghadi, M. Adnan, M. Ali, "Load Forecasting Techniques for Power System: Research Challenges and Survey," in *IEEE Access*, vol. 10, pp. 71054-71090, 2022.
- 1.15 T. H. Dang-Ha, et al., "Local short term electricity load forecasting: Automatic approaches," in *2017 international joint conference on neural networks (ijcnn)*, pp. 4267-4274, 2017.
- 1.16 L. Baur, et al., "Explainability and interpretability in electric load forecasting using machine learning techniques—A review," *Energy and AI* (2024): 100358.
- 1.17 V. Šiljkut, "Upravljanje potrošnjom u inteligentnim energetske mrežama sa varijabilnom proizvodnjom.", Doktorska disertacija, *Универзитет у Београду* (2015).
- 1.18 T. Hong, S. Fan, "Probabilistic electric load forecasting: A tutorial review, *International Journal of Forecasting*," vol. 32, no. 3, 2016, pp. 914-938, 2016.
- 1.19 S. Akhtar, et al., "Short-Term Load Forecasting Models: A Review of Challenges, Progress, and the Road Ahead," in *Energies*, vol. 16, p. 4060, 2023.
- 1.20 I. K. Nti, et al., "Electricity load forecasting: a systematic review," in *Journal of Electrical Systems and Information Technology*, vol. 7, pp. 1-19, 2020.

- 1.21 N. Ahmad, et al., "Load forecasting techniques for power system: Research challenges and survey," in *IEEE Access*, vol. 10, pp. 71054-71090, 2022.
- 1.22 A. Baliyan, et al., "A review of short term load forecasting using artificial neural network models," in *Procedia Computer Science*, vol. 48, pp. 121-125, 2015.
- 1.23 K. Zor, et al., "A state-of-the-art review of artificial intelligence techniques for short-term electric load forecasting," in *6th international youth conference on energy (IYCE)*, 2017.
- 1.24 V. Y. Kondaiah, et al. "A review on short-term load forecasting models for micro-grid application," in *The Journal of Engineering*, vol. 7, pp. 665-689, 2022.
- 1.25 A. B. Nassif, et al. "Artificial intelligence and statistical techniques in short-term load forecasting: a review," in *arXiv preprint arXiv:2201.00437*, 2021.
- 1.26 S. S. Subbiah, J. Chinnappan, "A review of short term load forecasting using deep learning," in *International Journal on Emerging Technologies*, vol. 11, no. 2, pp. 378-384, 2020.
- 1.27 M. A. Hammad, et al. "Methods and models for electric load forecasting: a comprehensive review," in *Logist. Sustain. Transp*, vol. 11, no. 1, pp. 51-76, 2020.
- 1.28 Y. Eren, Yavuz, I. Küçükdemiral, "A comprehensive review on deep learning approaches for short-term load forecasting," in *Renewable and Sustainable Energy Reviews*, vol. 189, p. 114031, 2024.
- 1.29 A. Al Mamun, et al. "A comprehensive review of the load forecasting techniques using single and hybrid predictive models," in *IEEE access*, vol. 8, pp. 134911-134939, 2020.
- 1.30 K. N. Singh, K. R. Singh, "A review on deep learning models for short-term load forecasting," in *Applications of Artificial Intelligence and Machine Learning: Select Proceedings of ICAAAIML 2020*, pp. 705-721, 2021.
- 1.31 S. Salehimehr, B. Taheri, M. Sedighizadeh, "Short-term load forecasting in smart grids using artificial intelligence methods: A survey," in *The Journal of Engineering* vol.12, pp. 1133-1142, 2022.
- 1.32 S. Haben, et al. "Review of low voltage load forecasting: Methods, applications, and recommendations," in *Applied Energy*, vol. 304, p. 117798, 2021.
- 1.33 H. Habbak, et al. "Load forecasting techniques and their applications in smart grids," in *Energies*, vol. 16, no. 3, p. 1480, 2023.
- 1.34 N. B. Vanting, Z. Ma, B. N. Jørgensen, "A scoping review of deep neural networks for electric load forecasting," in *Energy Informatics*, vol. 4(2), 2022.
- 1.35 N. Ahmad, et al. "Load forecasting techniques for power system: Research challenges and survey," in *IEEE Access*, vol. 10, pp. 71054-71090, 2022.
- 1.36 F. Dewangan, A. Y. Abdelaziz, M. Biswal. "Load forecasting models in smart grid using smart meter information: A review," *Energies*, vol. 16, no. 3, 2023.
- 1.37 N. S. M. Salleh, A. Suliman, B. N. Jørgensen, "A Systematic Literature Review of Electricity Load Forecasting using Long Short-Term Memory," in *Proceedings of the 8th International Conference on Computational Science and Technology*, Singapore, 2022.
- 1.38 R. Christen, et al. "Exogenous Data for Load Forecasting: A Review," in *IJCCI*, pp. 489-500, 2020.
- 1.39 S. A. Akinola, et al. "Deep learning approach to load forecasting: A survey," in *Communication, Networks and Computing: Second International Conference, CNC 2020*, 2020.
- 1.40 D. Upadhaya, R. Thakur, N. K. Singh, "A systematic review on the methods of short term load forecasting," in *2019 2nd International Conference on Power Energy, Environment and Intelligent Control (PEEIC)*, 2019.

- 1.41 S. N. Fallah, et al. "Computational intelligence on short-term load forecasting: A methodological overview," in *Energies* vol. 12, no. 3, p.393, 2019.
- 1.42 R. Patel, M. R. Patel, R. V. Patel, "A review: Introduction and understanding of load forecasting," in *Journal of Applied Science and Computations (JASC)*, vol. 4, no. 4, pp. 1449-1457, 2019.
- 1.43 Z. Wang, et al. "A review of load forecasting of the distributed energy system," in *IOP Conference Series: Earth and Environmental Science*, vol. 237, no. 4, 2019.
- 1.44 S. Adewuyi, et al. "An overview of deep learning techniques for short-term electricity load forecasting," in *Applied Computer Science*, vol.15, no. 4, 2019.
- 1.45 I. Barić, R. Grbić, E. K. Nyarko, "Short-term forecasting of electricity consumption using artificial neural networks-an overview," in *42nd International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO)*, 2019.
- 1.46 M.M. Asiri, G. Aldehim, F. Alotaibi, M.M. Alnfai, M. Assiri, A. Mahmud, "Short-Term Load Forecasting in Smart Grids using Hybrid Deep Learning," in *IEEE Access*, 2024.
- 1.47 S. Bansal, R. S. Lodhi, P. Nema, "State of Art on Short term Load Forecasting Using Artificial Neural Network," in *IOSR Journal of Electrical and Engineering E (IOSR-JEEE)*, vol.13, no. 3, pp. 80-85, 2018.
- 1.48 M. Bourdeau, et al. "Modeling and forecasting building energy consumption: A review of data-driven techniques," in *Sustainable Cities and Society* vol. 48, p. 101533, 2018.
- 1.49 A. Baliyan, K. Gaurav, S. K. Mishra, "A review of short term load forecasting using artificial neural network models," in *Procedia Computer Science*, vol. 48, pp. 121-125, 2015.
- 1.50 A. Almalaq, G. Edwards, "A review of deep learning methods applied on load forecasting," in *16th IEEE international conference on machine learning and applications (ICMLA)*, pp. 511-516, 2017.
- 1.51 J. Runge, R. Zmeureanu, "Forecasting Energy Use in Buildings Using Artificial Neural Networks: A Review," in *Energies*, vol. 12, no. 17, p. 3254, 2019.
- 1.52 C. Kuster, Y. Rezgui, M. Mourshed, "Electrical load forecasting models: A critical systematic review," in *Sustainable Cities and Society*, vol. 35, pp. 257-270, 2017.
- 1.53 I. Antonopoulos, et al. "Artificial intelligence and machine learning approaches to energy demand-side response: A systematic review," in *Renewable and Sustainable Energy Reviews*, vol. 130, p. 109899, 2020.
- 1.54 A. Román-Portabales, M. López-Nores, J.J. Pazos-Arias, "Systematic review of electricity demand forecast using ANN-based machine learning algorithms." in *Sensors*, vol.21. no. 13, p. 4544, 2021.
- 1.55 R. Czapaj, J. Kamiński, M. Sołtysik, "A Review of Auto-Regressive Methods Applications to Short-Term Demand Forecasting in Power Systems," in *Energies*, vol. 15, p. 6729, 2022.
- 1.56 X. Liu, et al., "A comparative study of the data-driven day-ahead hourly provincial load forecasting methods: From classical data mining to deep learning," in *Renewable and Sustainable Energy Reviews*, vol. 119, p.109632, 2020.
- 1.57 S. Makridakis, E. Spiliotis, V. Assimakopoulos, "Statistical and Machine Learning forecasting methods: Concerns and ways forward," in *PloS one*, vol.13, no. 3, 2018.
- 1.58 J. Li, et.al., "A Novel Hybrid Short-Term Load Forecasting Method of Smart Grid Using MLR and LSTM Neural Network," in *IEEE Trans. Ind. Inform.*, vol. 17, no. 4, pp. 2443–2452, 2021.
- 1.59 A. A. Mamun, et al., "A comprehensive review of the load forecasting techniques using single and hybrid predictive models," in *IEEE access* 8, pp. 134911-134939, 2020.

- 1.60 N. B. Vanting, M. Zheng Ma, B. N. Jørgensen, "A scoping review of deep neural networks for electric load forecasting," in *Energy Informatics* 4.Suppl 2 (2021): 49.
- 1.61 S. Humeau, T. K. Wijaya, M. Vasirani, K. Aberer, "Electricity load forecasting for residential customers: Exploiting aggregation and correlation between households," in *2013 Sustainable Internet and ICT for Sustainability (SustainIT)*, pp. 1–6, 2013.
- 1.62 E. Iskandarnia, A. A. Hesham, A. M. Wael, "Load forecasting in different scale and horizon-a review," in *3rd Smart Cities Symposium (SCS 2020)*, vol. 2020, pp. 464-470. IET, 2020.
- 1.63 S. Ryi, H. Noh, H. Kim, "Deep Neural Network Based Demand Side Short Term Load Forecasting," in *Energies*, vol. 10, no. 3, 2021.
- 1.64 S. Muzaffar, A. Afshari, "Short-Term Load Forecasts Using LSTM Networks," in *Energy Procedia*, 158, pp. 2922–2927, 2019.
- 1.65 W. Kong, Z. Y. Dong, Y. Jia, D. J. Hill, Y. Xu, Y. Zhang, "Short-Term Residential Load Forecasting Based on LSTM Recurrent Neural Network," in *IEEE Transactions on Smart Grid*, vol. 10, no. 1, pp. 841-851, 2009.
- 1.66 A. Muneer, et al., "Short Term Residential Load Forecasting using LSTM Recurrent Neural Network," in *International Journal of Electrical & Computer Engineering (2088-8708)*, vol. 12, no. 5, 2022.
- 1.67 T.L. Yuan, et al., "Recurrent Neural Network Based Short-Term Load Forecast with Spline Bases and Real-Time Adaptation," in *Applied Sciences*, 11(13), 2021.
- 1.68 F. Ünal, A. Almalaq, S. Ekici, "A Novel Load Forecasting Approach Based on Smart Meter Data Using Advance Preprocessing and Hybrid Deep Learning," in *Appl. Sci.* 11, 2021.
- 1.69 O. Valgaev, F. Kupzog, H. Schmeck, "Adequacy of neural networks for wide-scale day-ahead load forecasts on buildings and distribution systems using smart meter data," in *Energy Inform*, vol. 3, no. 28, 2020.
- 1.70 J. R. Cancelo, A. Espasa, R. Grafe, "Forecasting the electricity load from one day to one week ahead for the Spanish system operator," in *International Journal of Forecasting*, vol. 24, pp. 588–602, 2008.
- 1.71 A. Tarsitano, I. L. Amerise, "Short-term load forecasting using a two-stage sarimax model," in *Energy*, vol. 133, 2017.
- 1.72 Ö.F. Ertugrul, H. Tekin, R. Tekin, "A novel regression method in forecasting short-term grid electricity load in buildings that were connected to the smart grid," in *Electr. Eng.*, vol. 103, pp. 717–728, 2017.
- 1.73 B. Nepal, M. Yamaha, A. Yokoe, A. Yamaji, "Electricity load forecasting using clustering and ARIMA model for energy management in buildings," in *Jpn. Archit. Rev.*, vol. 3, pp. 62–76, 2019.
- 1.74 S. Amana, et al., "Prediction Models for Dynamic Demand Response: Requirements, Challenges, and Insights," in *Proceedings of the IEEE Conference on Smart Grid Communications*, Miami, FL, USA, 2–5 November 2015.
- 1.75 M. Božić, et al., "A new two-stage approach to short term electrical load forecasting," in *Energies*, vol. 6, no. 4, pp. 2130-2148, 2013.
- 1.76 E. Ceperic, V. Ceperic, A. Baric, "A Strategy for Short-Term Load Forecasting by Support Vector Regression Machines," in *IEEE Trans. Power Syst.*, vol. 28, pp. 4356–4364, 2015.
- 1.77 Y. Che, et al., "Short-term electrical load forecasting using the Support Vector Regression (SVR) model to calculate the demand response baseline for office buildings," in *Appl. Energy*, 195, 2017.

- 1.78 A. Lahouar, J.B.H. Slama, "Random Forests Model for One Day Ahead Load Forecasting," in *Proceedings of the IREC 2015 The Sixth International Renewable Energy Congress*, Sousse, Tunisia, 24–26 March 2015
- 1.79 M. Son, J. Moon, S. Jung, E. Hwang, "A Short-Term Load Forecasting Scheme Based on Auto-Encoder and Random Forest," in *Proceedings of the 3rd International Conference on Applied Physics, System Science and Computers (APSAC 2018)*, Dubrovnik, Croatia, 26–28 September 2018.
- 1.80 H. S. Hippert, C. E. Pedreira, R. C. Souza, "Neural networks for short-term load forecasting: A review and evaluation", in *IEEE Trans. Power Syst.*, vol. 16, pp. 44–55, 2001.
- 1.81 N. Ding, C. Benoit, G. Foggia, Y. Bésanger, F. Wurtz, "Neural Network-Based Model Design for Short-Term Load Forecast in Distribution System," in *IEEE Trans. Power Syst.*, vol. 31, pp. 72–81, 2016.
- 1.82 G. Jetcheva, M. Majidpour, W. P. Chen, "Neural Network Model Ensembles for Building Level Electricity Load Forecasts," in *Energy Build.*, vol. 84, pp. 214–233, 2014.
- 1.83 Y.A. Khan, D. Jayaweera, "Smart Meter Data Based Load Forecasting and Demand Side Management in Distribution Networks with Embedded PV Systems," in *IEEE Access*, vol. 8, pp. 2631–2644, 2019.
- 1.84 M. López, S. Valero, A. Rodriguez, I. Veiras, C. Senabre, "New online load forecasting system for the Spanish Transport System Operator," in *Electr. Power Syst. Res.*, vol. 154, pp. 401–412, 2018.
- 1.85 P. Mirowski, S. Chen, T. Kam Ho, C. N. Yu, "Demand Forecasting in Smart Grids," in *Bell Labs Tech. J.*, vol. 4, pp. 135–158, 2014.
- 1.86 M. Cai, M. Pipattanasomporn, S. Rahman, "Day-ahead building-level load forecasts using deep learning vs. traditional time series techniques," in *Applied Energy*, 236, 2019.
- 1.87 A. Almalaq, J. J. Zhang, "Evolutionary Deep Learning-Based Energy Consumption Prediction for Buildings," in *IEEE Access*, vol. 7, pp. 1520–1531, 2019.
- 1.88 D. Alberg, M. Last, "Short-term load forecasting in smart meters with sliding window-based ARIMA algorithms," in *Vietnam Journal of Computer Science*, vol. 5, pp. 241-249, 2018.
- 1.89 L. Li, et al. "Short-term apartment-level load forecasting using a modified neural network with selected auto-regressive features," in *Applied Energy* 287, 2021.
- 1.90 M. N. Fekri, M. Navid, et al. "Deep learning for load forecasting with smart meter data: Online Adaptive Recurrent Neural Network," in *Applied Energy* 282, 2021.
- 1.91 A. A. Mohammad, A. Al-Jaafreh, G. Mokryani, B. Amjad, "An enhanced CNN-LSTM based multi-stage framework for PV and load short-term forecasting: DSO scenarios," in *Energy Reports*, vol. 10, pp. 1387-1408, 2023.
- 1.92 C. Gilbert, J. Browell, B. Stephen, "Probabilistic load forecasting for the low voltage network: Forecast fusion and daily peaks," in *Sustainable Energy, Grids and Networks*, vol. 34, 2023.
- 1.93 Z. Zhu, et al., "eForecaster: unifying electricity forecasting with robust, flexible, and explainable machine learning algorithms," in *Proceedings of the AAAI Conference on Artificial Intelligence*. vol. 37, no. 13, 2023.
- 1.94 M. G. Pinheiro, et al., "Short-term electricity load forecasting—A systematic approach from system level to secondary substations." *Applied Energy* 332, 2023.
- 1.95 R. Vazquez, et al., "Assessment of an adaptive load forecasting methodology in a smart grid demonstration project," in *Energies*, vol.10, no. 2, 2017.
- 1.96 M. Grabner, et al. "A Global Modeling Approach for Load Forecasting in Distribution Networks," *arXiv preprint arXiv:2204.00493* (2022).

- 1.97 S. N. Wood, *Generalized additive models: an introduction with R*. CRC press, 2017.
- 1.98 A. Pierrot, Y. Goude, "Short-term electricity load forecasting with generalized additive models," in *Proceedings of the 16th International Conference on Intelligent System Applications to Power Systems ISAP*, Hersonisos, Crete, Greece, 25–28 September, 2011.
- 1.99 Y. Goude, F. Dan, "Electric Energy Forecast for the Industrial Consumers," in *Journal of Sustainable Energy* 2, no. 1, 2011.
- 1.100 S. Fan, R. J. Hyndman, "Short-term load forecasting based on a semi-parametric additive model," in *IEEE Trans. Power Syst.*, vol. 27, no. 1, pp. 134–141, 2012.
- 1.101 Y. Goude, N. Nedellec, N. Kong, "Local short and middle term electricity load forecasting with semi-parametric additive models," in *IEEE Trans. Smart Grids*, vol. 5, no. 1, pp. 440–446, 2012.
- 1.102 J. Ploennigs, B. Chen, P. Palmes, R. Lloyd, "e2-Diagnoser: A System for Monitoring, Forecasting and Diagnosing Energy Usage," in *Proceedings of the 2014 IEEE International Conference on Data Mining Workshop*, Shenzhen, China, 14 December **2014**.
- 1.103 S. N. Wood, Y. Goude, S. Shaw. "Generalized additive models for large data sets," in *Journal of the Royal Statistical Society Series C: Applied Statistics*, vol. 64, no.1, pp. 139-155, 2015.
- 1.104 P. Pompey, et al. "Massive-scale simulation of electrical load in smart grids using generalized additive models," in *Modeling and Stochastic Learning for Forecasting in High Dimensions*; Springer: Cham, Switzerland; New York, NY, USA, pp.193-212, 2015.
- 1.105 J.B. Fiot, F. Dinuzzo, "Electricity demand forecasting by multi-task learning," in *IEEE Transactions on Smart Grid*, vol. 9, no. 2, pp. 544-551, 2016.
- 1.106 E. Sjöborg, "Modelling and Forecasting Electricity Load in Secondary Substations," (2017).
- 1.107 S. N. Wood, et al. "Generalized additive models for gigadata: modeling the UK black smoke network daily data," in *Journal of the American Statistical Association* 112.519 (2017), pp. 1199-1210.
- 1.108 D. Obst, J. de Vilmarest, Y. Goude, "Adaptive Methods for Short-Term Electricity Load Forecasting during COVID-19 Lockdown in France," *arXiv* 2020, arXiv:2009.06527v1.
- 1.109 T.R. Khamma, et al. "Generalized additive models: An efficient method for short-term energy prediction in office buildings," in *Energy* 213 (2020): 118834.
- 1.110 L. Yang, R. Ren, X. Gu, L. Sun, "Interactive Generalized Additive Model and Its Applications in Electric Load Forecasting," in *Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, 2023.
- 1.111 Z. Guo, J. R. O'Hanley, S..Gibson, "Predicting residential electricity consumption patterns based on smart meter and household data: A case study from the Republic of Ireland," in *Utilities Policy* 79 (2022): 101446.
- 1.112 T. R. Khamma, Z. Zhang, S. Guerrier, M. Boubekri, "Generalized additive models: An efficient method for short-term energy prediction in office buildings," in *Energy* 213 (2020): 118834.
- 1.113 B. Grillone, et al., "Use of generalized additive models to assess energy efficiency savings in buildings using smart metering data," in *PROCEEDINGS book 27*, 2019.
- 1.114 A. Sundararajan, B. Ollis, "Regression and generalized additive model to enhance the performance of photovoltaic power ensemble predictors," in *IEEE Access* 9, pp. 111899-111914, 2021.
- 1.115 V. Thouvenot, "*Estimation et sélection pour les modèles additifs et application à la prévision de la con sommation électrique*," Ph.D. dissertation, Université Paris-Saclay, Français, 2015.

- 1.116 V. Thouvenot, A. Pichavant, Y. Goude, A. Antoniadis, J. M. Poggi, "Electricity Forecasting Using Multi-Stage Estimators of Nonlinear Additive Models," in *IEEE Trans. Power Syst.*, vol. 36, no. 5, pp. 3665–3673, 2015.
- 1.117 G. Marra, S. N. Wood, "Practical variable selection for generalized additive models," *Computational Statistics & Data Analysis*, 2011, pp. 2372-2387.
- 2.1 <https://cran.r-project.org/web/package>
- 2.2 www.nyiso.com
- 2.3 C. E. Borges, A. Peña, Y. K. Peña, " On the influence of surrounding load demand to improve primary substation STLF," *IECON 2013 - 39th Annual Conference of the IEEE Industrial Electronics Society*, Vienna, Austria, 2013, pp. 8166-8171, doi: 10.1109/IECON.2013.6700499.
- 2.4 <https://www.ausgrid.com.au/Industry/Our-Research/Data-to-share/Distribution-zone-substation-data>
- 2.5 www.wunderground.com
- 2.6 <https://openenergoc-data.s3.amazonaws.com/anon/index.html>
- 2.7 <https://www.kaggle.com/c/ashrae-energy-prediction/overview/description>
- 2.8 <https://data.london.gov.uk/dataset/smartmeter-energy-use-data-in-london-households>
- 3.1 T. Hastie, R. Tibshirani. "Generalized Additive Models." *Statistical Science*, vol. 1, no. 3, 1986, pp. 297–310. *JSTOR*, <http://www.jstor.org/stable/2245459>
- 3.2 H. Liu, "Generalized additive model" *Department of Mathematics and Statistics University of Minnesota Duluth: Duluth, MN, USA 55812* (2008).
- 3.3 T. Hastie, R. Tibshirani, "Generalized additive models: some applications", *Journal of the American Statistical Association* 82.398 (1987): 371-386.
- 3.4 S. N. Wood, "Fast stable direct fitting and smoothness selection for generalized additive models", *Journal of the Royal Statistical Society, Series B (Statistical Methodology)*, 70(3), 495-518. <https://doi.org/10.1111/j.1467-9868.2007.00646.x>
- 3.5 S. N. Wood, "Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models", *Journal of the Royal Statistical Society (B)* 73(1):3-36 dets. *J.R.Statist. Soc. B* 70(3):495-518
- 3.6 J. M. Grego, "Generalized additive models." *Encyclopedia of Environmetrics* (2006)
- 3.7 P. Burman, "Estimation of generalized additive models", *Journal of multivariate analysis* 32.2 (1990): 230-255.
- 3.8 S. N. Wood, "Inference and computation with generalized additive models and their extensions", *Test* 29.2 (2020): 307-339.
- 3.9 R. Wong et al., "Robust estimation for generalized additive models." *Journal of Computational and Graphical Statistics* 23.1 (2014): 270-289.
- 3.10 K. Berhane, Kiros, R. J. Tibshirani. "Generalized additive models for longitudinal data." *Canadian Journal of Statistics* 26.4 (1998): 517-535.
- 3.11 M. Aerts, G. Claeskens, M. P. Wand, "Some theory for penalized spline generalized additive models", *Journal of statistical planning and inference* 103.1-2 (2002): 455-470.
- 3.12 G. Kauermann, J. D. Opsomer. "Local likelihood estimation in generalized additive models", *Scandinavian Journal of Statistics* 30.2 (2003): 317-337.
- 3.13 M. G. Schimek, A. T. Berwin, "Additive and generalized additive models", *Humboldt-Universität zu Berlin, Wirtschaftswissenschaftliche Fakultät*, 2006.

- 3.14 B. Ganguli, P. W. Matt, "Feature significance in generalized additive models", *Statistics and Computing* 17 (2007): 179-192.
- 3.15 S. N. Wood, "Generalized additive models: an introduction with R", CRC press, 2017.
- 3.16 <https://cran.r-project.org/web/packages/mgcv/index.html>
- 3.17 H. Trevor, R. Tibshirani, J. Friedman, "The Elements of Statistical Learning: Data Mining, Inference and Prediction", Springer, 2009. Second edition.
- 3.18 <https://www.rdocumentation.org/packages/mgcv/versions/1.9-1/topics/gamm>
- 3.19 E. J. Pedersen, et al. "Hierarchical generalized additive models in ecology: an introduction with *mgcv*." *PeerJ* 7 (2019): e6876.
- 3.20 S. N. Wood, et al. "Generalized additive models for gigadata: modeling the UK black smoke network daily data." *Journal of the American Statistical Association*, 112.519 (2017): 1199-1210.
- 3.21 A. Pierrot, Y. Goude, "Short-term electricity load forecasting with generalized additive models", *Proceedings of the 16th International Conference on Intelligent System Applications to Power Systems ISAP*, Hersonisos, Crete, Greece, 25–28 September 2011
- 3.22 W. Garner, G. Sciences, "Automated Forward and Best Subset Selection for Generalized Additive Models using PROC GAM", SAS technical paper.
- 3.23 A. Pierrot, Y. Goude, "Short-term electricity load forecasting with generalized additive models", *Proceedings of the 16th International Conference on Intelligent System Applications to Power Systems ISAP*, Hersonisos, Crete, Greece, 25–28 September 2011.
- 3.24 S. Fan, R. J. Hyndman, "Short-term load forecasting based on a semi-parametric additive model", *IEEE Trans. Power Syst.* 2012, 27, 134–141.
- 3.25 S. Arlot, A. Celisse, "A survey of cross-validation procedures for model selection", *Statistical Survey*, vol. 4, pp. 40–79, 2010
- 3.26 G. Marra, S. N. Wood, "Practical variable selection for generalized additive models", *Computational Statistics & Data Analysis* 55.7 (2011): 2372-2387.
- 3.27 A. Thouvenot et. al, "Electricity Forecasting Using Multi-Stage Estimators of Nonlinear Additive Models", *IEEE Trans. Power Syst.* 2016, 31, 3665–3673.
- 3.28 Y. Lin, H. H. Zhang, "Component Selection and Smoothing in Multivariate Nonparametric Regression", *Ann. Stat.* 2006, 34, 2272–2297.
- 3.29 P. Ravikumar, J. Lafferty, H. Liu, L. Wasserman, "Sparse Additive Model", *J. R. Stat. Soc. Ser. B (Stat. Methodol.)* 2009, 71, 1009–1030.
- 3.30 A. Chouldechova, T. Hastie, "Generalized Additive Model Selection", *arXiv* 2015, arXiv:1506.03850.
- 3.31 S. N. Wood, Y. Goude, S. Shaw. "Generalized additive models for large data sets." *Journal of the Royal Statistical Society Series C: Applied Statistics* 64.1 (2015): 139-155.
- 4.1 X. Liu, Z. Zhang, Z. Song, "A comparative study of the data-driven day-ahead hourly provincial load forecasting methods: From classical data mining to deep learning," in *Renewable and Sustainable Energy Reviews* 119 (2020), 109632.
- 4.2 S. Ruzic, A. Vuckovic and N. Nikolic, "Weather sensitive method for short term load forecasting in Electric Power Utility of Serbia," in *IEEE Transactions on Power Systems*, vol. 18, no. 4, pp. 1581-1586, Nov. 2003.
- 4.3 S. N. Wood, Y. Goude, S. Shaw, "Generalized additive models for large data sets," in *Journal of the Royal Statistical Society Series C: Applied Statistics*, vol. 64, no. 1, 2015, pp. 139-155.

- 4.4 G. J. Tsekouras, F. D. Kanellos, N. Mastorakis, "Short term load forecasting in electric power systems with artificial neural networks," in *Computational problems in science and engineering*, 2015, pp. 19-58.
- 4.5 J. Hinman, E. Hickey, "Modeling and forecasting short-term electricity load using regression analysis," in *Journal of Institute for Regulatory Policy Studies*, 2009.
- 4.6 M. Janicki, "Methods of weather variables introduction into short-term electric load forecasting models-a review," in *Przegląd Elektrotechniczny*, vol. 93, no. 4, 2017.
- 4.7 X. Jingrui, T. Hong, "Wind speed for load forecasting models," in *Sustainability*, vol. 9, no. 5, 2017.
- 4.8 X. Jingrui, et al., "Relative humidity for load forecasting models," in *IEEE Transactions on Smart Grid* vol. 9, no. 1, 2016, pp. 191-198.
- 4.9 M. López, S. Valero, A. Rodriguez, I. Veiras, C. Senabre, "New online load forecasting system for the Spanish Transport System Operator," in *Electr. Power Syst. Res.*, 2018, pp. 401–412.
- 4.10 X. Liu, Z. Zijun, S. Zhe, "A comparative study of the data-driven day-ahead hourly provincial load forecasting methods: From classical data mining to deep learning," in *Renewable and Sustainable Energy Reviews* 119 (2020): 109632.
- 4.11 H. Wang, et al., "A comprehensive review for wind, solar, and electrical load forecasting methods," in *Global Energy Interconnection*, vol. 5, no. 1, 2022, pp. 9-30.
- 4.12 A. Falce, et al., "Short-term net load forecast in distribution networks with PV penetration behind the meter," in *Energy Reports*, vol. 9, Supplement 3, 2023, pp. 115-122.
- 4.13 Y. Chu, et al., "Intra-Hour Net load forecasts for solar-integrated operational grid feeders", in *Solar Energy* vol. 158, 2017, pp. 236-246.
- 4.14 R. Sevlian, R. Rajagopal, "A scaling law for short term load forecasting on varying levels of aggregation", *International Journal of Electrical Power & Energy Systems*, vol. 98, 2018, pp. 350-361.
- 4.15 B. Hayes, J. G. Barry, M. Prodanovic, "Short-term load forecasting at the local level using smart meter data," in *2015 IEEE Eindhoven PowerTech*, 2015.
- 4.16 Y. Peng, et al., "Short-term load forecasting at different aggregation levels with predictability analysis," in *2019 IEEE Innovative Smart Grid Technologies-Asia (ISGT Asia)*, 2019.
- 4.17 T. Hong, S. Fan, "Probabilistic electric load forecasting: A tutorial review," in *International Journal of Forecasting*, vol. 32, no. 3, 2016, pp. 914-938.
- 4.18 G. Ciaran, J. Browell, S. Bruce, "Probabilistic load forecasting for the low voltage network: Forecast fusion and daily peaks, " in *Sustainable Energy, Grids and Networks*, vol. 34, 2023, pp. 2352-4677.
- 4.19 Y. Wang, N. Zhang, X. Chen, "A short-term residential load forecasting model based on LSTM recurrent neural network considering weather features," in *Energies* 14.10 (2021): 2737.
- 4.20 W. Kong, et al., "Short-term residential load forecasting based on resident behaviour learning," in *IEEE Transactions on power systems*, vol. 33, no. 1, 2017, pp. 1087-1088.
- 4.21 G. Dudek, "Multilayer perceptron for short-term load forecasting: from global to local approach," in *Neural Comput & Applic*, **32**, 2020, pp. 3695–3707.
- 4.22 Q. Zhang, et al., "Short-term load forecasting based on frequency domain decomposition and deep learning," in *Mathematical Problems in Engineering*, 2020, pp. 1-9.
- 4.23 N. Amjady, F. Keynia, "Short-term load forecasting of power systems by combination of wavelet transform and neuro-evolutionary algorithm," in *Energy*, vol. 34, no. 1, pp. 46–57.

- 4.24 J. Li, et al., "A Novel Hybrid Short-Term Load Forecasting Method of Smart Grid Using MLR and LSTM Neural Network," in *IEEE Trans. Ind. Inform.*, vol. 17, 2021, pp.2443–2452.
- 4.25 T. Teeraratkul, D. O’Neill, S. Lall, "Shape-based approach to household electric load curve clustering and prediction," in *IEEE Transactions on Smart Grid*, vol. 9, no. 5, 2017, pp. 5196-5206.
- 4.26 B. Nepal, et al. "Electricity load forecasting using clustering and ARIMA model for energy management in buildings," in *Japan Architectural Review*, vol. 3, no. 1, 2020, pp. 62-76.
- 4.27 R. Silipo, P. Winters, "Big data, smart energy, and predictive analytics," in *Time Series Prediction of Smart Energy Data 1* (2013): 37.
- 4.28 Z. A. Khan, J. Dilan, "Smart meter data based load forecasting and demand side management in distribution networks with embedded PV systems," in *IEEE Access* 8, 2019, pp. 2631-2644.
- 4.29 H. Shi, M. Xu, R. Li, "Deep learning for household load forecasting—A novel pooling deep RNN," in *IEEE Transactions on Smart Grid*, vol. 9, no. 5, 2017, pp. 5271-5280.
- 4.30 F. Han, et al., "Short-term forecasting of individual residential load based on deep learning and K-means clustering," in *CSEE Journal of Power and Energy Systems* 7.2 (2020), pp. 261-269.
- 4.31 J. Xu, et al. "Spatial-temporal load forecasting using AMI data," in *2016 IEEE International Conference on Smart Grid Communications (SmartGridComm)*, 2016.
- 4.32 S. Arastehfar, et al., "Short-term residential load forecasting using graph convolutional recurrent neural networks," in *Engineering Applications of Artificial Intelligence* 116 (2022): 105358.
- 4.33 F. Ziel, "Modeling public holidays in load forecasting: a German case study," in *Journal of Modern Power Systems and Clean Energy*, 6.2 (2018): 191-207.
- 4.34 M. M. Elgazzar, H. E. Elsayed, "Electrical load forecasting using Hijri causal events," in *2016 Eighteenth International Middle East Power Systems Conference (MEPCON)*, IEEE, 2016.
- 4.35 M. Lopez, S. Valero, C. Sans, C. Senabre, "Use of Available Daylight to Improve Short-Term Load Forecasting Accuracy," in *Energies*, 2021, 14, 95.
- 4.36 S. N. Wood, Y. Goude, S. Shaw, "Generalized additive models for large data sets," in *Journal of the Royal Statistical Society Series C: Applied Statistics* 64.1 (2015): 139-155.
- 4.37 G. EP. Box, D. R. Cox. "An analysis of transformations," in *Journal of the Royal Statistical Society Series B: Statistical Methodology* 26.2 (1964): 211-243.
- 5.1 R. J. Hyndman, G. Athanasopoulos, "Forecasting: principles and practice," *OTexts*, 2018.
- 5.2 S. N. Wood, *Generalized additive models: an introduction with R*. CRC press, 2017.
- 5.3 D. Maia-Silva, R. Kumar, R. Nateghi, "The critical role of humidity in modeling summer electricity demand across the United States," in *Nature Communication*, 2020, 11, 1686.
- 5.4 G. Notton, C. Voyant, "Forecasting of intermittent solar energy resource," in *Advances in renewable energies and power technologies*, Elsevier, 2018. pp. 77-114.
- 5.5 Market, NYISO Electricity. "Price Forecast Valuation for the NYISO electricity market".
- 5.6 B. M. Hodge, D. Lew, M. Milligan, "Short-term load forecast error distributions and implications for renewable integration studies," in *Proceedings of the 2013 IEEE Green Technologies Conference (GreenTech)*, 2013.
- 5.7 K. F. Forbes, "Demand for grid-supplied electricity in the presence of distributed solar energy resources: Evidence from New York City," in *Utilities Policy*, Elsevier, vol. 80(C), 2023.

- 5.8 S. S. Ahmed, et al., "A two-fold machine learning approach for efficient day-ahead load prediction at hourly granularity for NYC," in *Advances in Information and Communication: Proceedings of the 2019 Future of Information and Communication Conference (FICC), Volume 2*, pp. 84-97, Springer International Publishing.
- 5.9 S. Henselmeyer, M. Grzegorzek. "Short-term load forecasting using an attended sequential encoder-stacked decoder model with online training," *Applied Sciences* 11.11 (2021): 4927.
- 5.10 M. Lopez, S. Valero, C. Sans, C. Senabre, "Use of Available Daylight to Improve Short-Term Load Forecasting Accuracy," in *Energies*, 2021, 14, 95.
- 5.11 S. B. Taieb, R. J. Hyndman, "*Recursive and direct multi-step forecasting: the best of both worlds*," Department of Econometrics and Business Statistics, Vol. 19, Monash
- 5.12 M. N. Fekri, et al., "Deep learning for load forecasting with smart meter data: Online Adaptive Recurrent Neural Network," in *Applied Energy* 282 (2021), 116177.
- 5.13 N. Haghdad, et al., "Impact of distributed photovoltaic systems on zone substation peak demand," in *IEEE Transactions on Sustainable Energy*, 2017, pp. 621-629.
- 5.14 Y. Wang, et al., "An ensemble forecasting method for the aggregated load with subprofiles," in *IEEE Transactions on Smart Grid*, 2018, pp. 3906-3908.
- 5.15 I. A. Ibrahim, M. J. Hossain, "LSTM neural network model for ultra-short-term distribution zone substation peak demand prediction," in *IEEE Power & Energy Society General Meeting (PESGM)*, 2020, pp.1-5.
- 5.16 <https://www.iea.org/reports/the-critical-role-of-buildings>
- 5.17 L. Cao, et al., "Electrical load prediction of healthcare buildings through single and ensemble learning," in *Energy Reports*, 2020, pp. 2751-2767.
- 5.18 Guideline, A. S. H. R. A. E. "Measurement of energy, demand, and water savings." *ASHRAE guidel 4* (2014): 1-150.
- 5.19 J. Piotrowska-Woroniak, T. Szul, "Application of a model based on rough set theory (RST) to estimate the energy efficiency of public buildings," in *Energies* 15.23, 2022.
- 5.20 M. Cai, M. Pipattanasomporn, S. Rahman, "Day-ahead building-level load forecasts using deep learning vs. traditional time series techniques," in *Applied Energy*, 2019, pp. 1078–1088.
- 5.21 F. E. Nordbjerg, N. Iftikhar, "On measuring similarity in consumption patterns for smart meter time series," in UC Viden - Professionshøjskolernes Videndatabase , 2018

БИОГРАФИЈА

Совјетка Крстонијевић рођена је 14.12.1966. године у Мостару, Босна и Херцеговина. Основну и средњу школу завршила је у Титограду (сада Подгорица). Електротехнички факултет у Београду, одсек Електроника, уписала је 1985. године. Дипломирала је фебруара 1992. године, одбраном рада под називом "Видео меморије", где јој је ментор био проф. др. Миодраг Поповића. На истом факултету је уписала постдипломске студије, а магистарски рад под називом "Препознавање замора у мишићу на основу електромиографског сигнала користећи вештачке неуронске мреже", одбранила је 2001. године, код ментора проф. др. Дејана Поповића. Школске 2105/2016 уписује Докторске академске студије на Електротехничком факултету Универзитета у Београду, смер електроника.

По дипломирању, запошљава се у Телеоптик.д.о.о., прво, на програму развоја медицинске опреме, а потом бива ангажована на пројектима одељења за Аутоматику. Након стицања дипломе Магистра електротехничких наука 2001., запошљава се у Институту Михајло Пупин, где и данас ради. За цео период рада у Институту, била је учесник на пројектима Министарства просвете и науке Републике Србије, (Програм: Технолошки развој, Област: Енергетика, рударство и енергетска ефикасност). Учествовала је у изради већег броја техничких решења, од којих је на 5 први аутор. Област њеног рада и истраживања је развој различитих апликација и функционалности намењених управљању електроенергетским системом, те њихова примена и кастомизација сходно захтевима корисника, као на пример, софтверски пакети за предвиђање могућности оптерећења трансформатора и далековода (*PMOT*, *PMODV*), за краткорочну прогнозу потрошње за произвољну област електроенергетског система, дефинисану од стране корисника (*STLF RT*, *STLF Model*), као и њено даље унапређење у правцу шире примене на дистрибутивне системе и будуће интелигентне мреже.

Совјетка Крстонијевић је аутор/коаутор више радова презентованих на домаћим и међународним конференцијама, коаутор студије и радова у домаћем часопису, као и аутор рада објављеног у часопису са SCI листе у категорији M21. Током 2014/2015 била је учесник семинара за развој предузетништва у науци, финансираног од стране Европске уније, а у оквиру пројекта *PACINNO—Platform for trans-Academic Cooperation in Innovation-Managerial-Business Skills Development for Researchers*, *COBIK Slovenia*.

Изјава о ауторству

Име и презиме аутора: **мр Совјетка Крстонијевић**

Број индекса: **5049/15**

Изјављујем

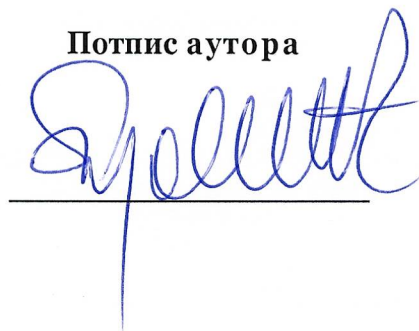
да је докторска дисертација под насловом

„Оквир за прогнозу потрошње електричне енергије у интелигентним електроенергетским мрежама заснован на генерализованом адитивном моделу“

- резултат сопственог истраживачког рада;
- да дисертација у целини ни у деловима није била предложена за стицање друге дипломе према студијским програмима других високошколских установа;
- да су резултати коректно наведени и
- да нисам кршио/ла ауторска права и користио/ла интелектуалну својину других лица.

У Београду, 25.11.2024

Потпис аутора



Изјава о истоветности штампане и електронске верзије докторског рада

Име и презиме аутора: **мр Совјетка Крстонијевић**

Број индекса: **5049/15**

Студијски програм: **електроника**

Наслов рада: „ **Оквир за прогнозу потрошње електричне енергије у интелигентним електроенергетским мрежама заснован на генерализованом адитивном моделу**“

Ментор: **др Предраг Пејовић, редовни професор**

Изјављујем да је штампана верзија мог докторског рада истоветна електронској верзији коју сам предао/ла ради похрањивања у **Дигиталном репозиторијуму Универзитета у Београду**.

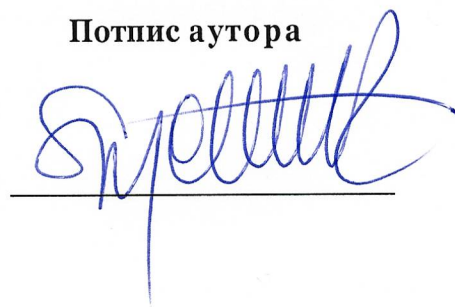
Дозвољавам да се објаве моји лични подаци везани за добијање академског назива доктора наука, као што су име и презиме, година и место рођења и датум одбране рада.

Ови лични подаци могу се објавити на мрежним страницама дигиталне библиотеке, у електронском каталогу и у публикацијама Универзитета у Београду.

У Београду,

25.11.2024

Потпис аутора



Изјава о коришћењу

Овлашћујем Универзитетску библиотеку „Светозар Марковић“ да у Дигитални репозиторијум Универзитета у Београду унесе моју докторску дисертацију под насловом:

„ Оквиp за прогнозу потрошње електричне енергије у интелигентним електроенергетским мрежама заснован на генерализованом адитивном моделу“

која је моје ауторско дело.

Дисертацију са свим прилозима предао/ла сам у електронском формату погодном за трајно архивирање.

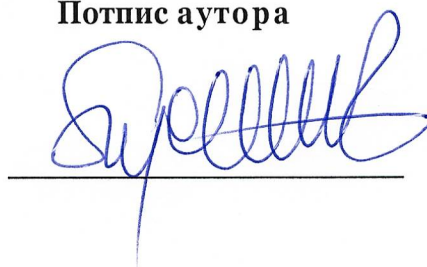
Моју докторску дисертацију похрањену у Дигиталном репозиторијуму Универзитета у Београду и доступну у отвореном приступу могу да користе сви који поштују одредбе садржане у одабраном типу лиценце Креативне заједнице (Creative Commons) за коју сам се одлучио/ла.

1. Ауторство (CC BY)
2. Ауторство – некомерцијално (CC BY-NC)
3. Ауторство – некомерцијално – без прерада (CC BY-NC-ND)
4. Ауторство – некомерцијално – делити под истим условима (CC BY-NC-SA)
5. Ауторство – без прерада (CC BY-ND)
6. Ауторство – делити под истим условима (CC BY-SA)

(Молимо да заокружите само једну од шест понуђених лиценци.
Кратак опис лиценци је саставни део ове изјаве).

У Београду, 25.11.2024

Потпис аутора



1. **Ауторство.** Дозвољаваате умножавање, дистрибуцију и јавно саопштавање дела, и прераде, ако се наведе име аутора на начин одређен од стране аутора или даваоца лиценце, чак и у комерцијалне сврхе. Ово је најслободнија од свих лиценци.
2. **Ауторство – некомерцијално.** Дозвољаваате умножавање, дистрибуцију и јавно саопштавање дела, и прераде, ако се наведе име аутора на начин одређен од стране аутора или даваоца лиценце. Ова лиценца не дозвољава комерцијалну употребу дела.
3. **Ауторство – некомерцијално – без прерада.** Дозвољаваате умножавање, дистрибуцију и јавно саопштавање дела, без промена, преобликовања или употребе дела у свом делу, ако се наведе име аутора на начин одређен од стране аутора или даваоца лиценце. Ова лиценца не дозвољава комерцијалну употребу дела. У односу на све остале лиценце, овом лиценцом се ограничава највећи обим права коришћења дела.
4. **Ауторство – некомерцијално – делити под истим условима.** Дозвољаваате умножавање, дистрибуцију и јавно саопштавање дела, и прераде, ако се наведе име аутора на начин одређен од стране аутора или даваоца лиценце и ако се прерада дистрибуира под истом или сличном лиценцом. Ова лиценца не дозвољава комерцијалну употребу дела и прерада.
5. **Ауторство – без прерада.** Дозвољаваате умножавање, дистрибуцију и јавно саопштавање дела, без промена, преобликовања или употребе дела у свом делу, ако се наведе име аутора на начин одређен од стране аутора или даваоца лиценце. Ова лиценца дозвољава комерцијалну употребу дела.
6. **Ауторство – делити под истим условима.** Дозвољаваате умножавање, дистрибуцију и јавно саопштавање дела, и прераде, ако се наведе име аутора на начин одређен од стране аутора или даваоца лиценце и ако се прерада дистрибуира под истом или сличном лиценцом. Ова лиценца дозвољава комерцијалну употребу дела и прерада. Слична је софтверским лиценцама, односно лиценцама отвореног кода.