



AKADEMIA GÓRNICZO-HUTNICZA IM. STANISŁAWA STASZICA W KRAKOWIE

WYDZIAŁ ZARZĄDZANIA

KATEDRA INFORMATYKI BIZNESOWEJ I INŻYNIERII ZARZĄDZANIA

Praca dyplomowa magisterska

*„Wykorzystanie podejścia hierarchicznego do prognozowania
zużycia energii elektrycznej w USA”*

*“Using a hierarchical approach to forecast electricity
consumption in the USA”*

Autor:	<i>Bartosz Nguyen Van</i>
Kierunek studiów:	<i>Informatyka i Ekonometria</i>
Opiekun pracy:	<i>dr inż. Bartłomiej Gaweł</i>

Kraków, 2022

Spis treści

Wstęp.....	1
1. Podejście hierarchiczne w prognozowaniu gospodarczym	4
1.1. Prognozowanie hierarchiczne.....	4
1.1.1. Prognozowanie energii, a metody hierarchiczne.....	5
1.1.2. Szereg czasowy i oznaczenia.....	6
1.1.3. Uzgadnianie prognoz.....	10
1.1.4. Opis wybranych metod uzgadniania	12
1.1.5. Prognozowanie hierarchiczne, a automatyczne.....	18
2. Prognozowanie automatyczne i inne metody	19
2.1. Istota prognozowania automatycznego	19
2.2. Metody stosowane w prognozach automatycznych	19
2.2.1. ETS	19
2.2.2. ARIMA.....	21
2.2.3. STLM	23
2.2.4. TBATS	23
2.3. Automatyczny dobór parametrów	23
2.4. Inne metody zastosowane w badaniu	24
2.4.1. SNAIVE	24
2.4.2. Prophet.....	24
2.5. Ocena błędu prognozy.....	29
2.5.1. MAPE	29
2.5.2. RMSE	29
2.5.3. MASE.....	30
2.5.4. CRPS	31
2.5.5. Skill score	31
3. Wykorzystanie prognozowania hierarchicznego do budowy prognozy zużycia energii elektrycznej w USA	33
3.1. Dane.....	33
3.1.1. Źródło danych.....	33
3.1.2. Charakterystyka zbioru danych	34
3.1.3. Struktura hierarchiczna	35
3.1.4. Preprocessing danych	37
3.1.5. Wstępna analiza danych	39
3.1.6. Podział zbioru danych	41
3.2. Stos technologiczny.....	42
3.2.1. Język R	42
3.2.2. Pakiety R	42
3.2.3. RStudio	44
3.3. Eksperymenty	44
3.3.1. Eksperyment 1 – ocena trafności prognozowania metod bazowych na najwyższym poziomie hierarchii.....	48
3.3.2. Eksperyment 2 – ocena trafności prognozowania metod bazowych na najniższym poziomie hierarchii.....	49
3.3.3. Eksperyment 3 – zastosowanie metod uzgadniania	50
3.3.4. Eksperyment 4 – porównanie metod pod względem prognozy punktowej.....	57
3.3.5. Eksperyment 5 - porównanie metod pod względem prognozy przedziałowej.....	62
3.3.6. Eksperyment 6 - porównanie metod pod względem czasu wykonywania	65
3.4. Wnioski	68
Zakończenie.....	71

<i>Bibliografia</i>	73
<i>Spis tabel</i>	76
<i>Spis rysunków</i>	77

Wstęp

We współczesnym świecie energia elektryczna jest czynnikiem niezbędnym dla prawidłowego funkcjonowania społeczeństwa i gospodarki. Przerwanie dostaw elektryczności praktycznie paraliżuje cały oparty na niej system, przez co wiele funkcji nieodzownych dla dzisiejszego życia ulega zatrzymaniu. Praktycznie niemożliwe jest również oszacowanie rzeczywistej wielkości roli, jaką odegrała w budowaniu dzisiejszej cywilizacji. Dostępność ogromnych ilości energii elektrycznej w dzisiejszych czasach spowodowała skrócenie dnia pracy, zwiększenie produkcji rolnej i przemysłowej, zdrowszą i bardziej zrównoważoną dietę, lepsze możliwości transportowe oraz wiele innych aspektów, które ułatwiły życie człowieka.

Mimo, że energia elektryczna jest wykorzystywana przez ludzkość już od XIX wieku to sektor energetyczny wciąż się zmienia. W ostatnich latach zmiany zachodzą, między innymi, jeśli chodzi o strukturę źródeł skąd jest pozyskiwana elektryczność, gdyż coraz w większym stopniu wykorzystywane są surowce odnawialne, przez co dąży się do zmniejszenia uzależnienia od paliw kopalnych i osiągnięcia lepszej, zrównoważonej przyszłości. Ponadto w wielu krajach, zwłaszcza tych najbardziej rozwiniętych gospodarczo, istnieje trend, by systemy elektroenergetyczne były coraz bardziej „inteligentne”, czyli w pewnym zakresie autonomiczne w działaniach mających na celu optymalizację wydajności energetycznej.¹

Biorąc pod uwagę zachodzące przemiany oraz rosnącą ilość danych pochodzących z coraz bardziej zaawansowanych systemów energetycznych, wciąż istnieje nieodzowna i rosnąca potrzeba, aby przedsiębiorstwa użyteczności publicznej określały ilościowo niepewność przyszłej produkcji i zapotrzebowania energii. Dokładne prognozy zapotrzebowania na energię elektryczną są krytyczne dla zakładów energetycznych w wielu zadaniach operacyjnych i planistycznych. Posiadanie takich predykcji umożliwia efektywne i bezpieczne zarządzanie sieciami energetycznymi. Dodatkowo, trafne prognozy są również bardzo ważne ze względu na specyfikę energii elektrycznej tj. z powodu konsekwencji wynikających z możliwej nadmiarowości jej generowania, mianowicie, energii elektrycznej nie da się łatwo przechowywać na dużą skalę, a w przypadku wytwarzania jej z paliw kopalnych istnieją istotne problemy środowiskowe. Dostawy muszą być wystarczające, aby zaspokoić przewidywane zapotrzebowanie, ponieważ nieodpowiednie lub zawodne dostawy mogą być kosztowne dla osób, organizacji oraz środowiska naturalnego.

¹ Kranz J., Kolbe L.M., Koo C., *Smart energy: where do we stand and where should we go?*, Electron Markets, 2015

Mając świadomość jak ważne problemy pomaga rozwiązać trafne zaprognozowanie zapotrzebowania energii elektrycznej, badacze już od dziesięcioleci wykorzystują różnorodne techniki i modele prognostyczne, których celem jest osiągnięcie jak najlepszego rezultatu w tym aspekcie. W ostatnich latach popularne stało się zwłaszcza podejście nazywane prognozowaniem hierarchicznym, której wykorzystanie pozwala na uzyskiwanie prognoz opierając się na powiązaniach w całej strukturze wybranej sieci elektroenergetycznej.

Celem pracy było zbadanie czy wykorzystanie podejścia hierarchicznego w prognozowaniu zużycia energii elektrycznej pozwala na zwiększenie trafności. Poza tym czynnikiem wzięty pod uwagę został także drugi aspekt użycia wspomnianych metod, a mianowicie, czas trwania obliczeń. W badaniu zostało rozważonych kilka najpopularniejszych rodzin metod prognostycznych oraz różne podejścia prognozowania hierarchicznego, co miało pozwolić na wybór optymalnego podejścia, jeśli chodzi o ich kombinacje. Co ważne, badanie miało też umożliwić ocenę czy przy budowie prognoz w szeregach czasowych charakteryzujących się strukturą hierarchiczną warto wykorzystywać występujące w tej kompozycji powiązania lub traktować je niezależnie. Ostatecznie, poprzez wybór najlepszego podejścia prognozowania hierarchicznego celem było także określenie, czy wystarczy prognozować jedynie wybrany poziom struktury hierarchicznej, by osiągnąć zadowalające wyniki (np. tylko najniższy poziom, a pozostałe wyższe poziomy uzyskiwać w oparciu o niego).

Osiągnięcie wszystkich powyższych celów pozwala na uzyskanie wiedzy oraz perspektywy, które mogą w pewnym stopniu pomóc ocenić czy implementacja badanych metod prognozowania jest opłacalna dla przedsiębiorstw, urzędów i innych jednostek. Często nawet minimalna poprawa trafności i szybkości uzyskiwania prognoz pozwala na znaczne oszczędności pod względem finansowym.

Niniejsze opracowanie składa się z trzech głównych rozdziałów. W rozdziale pierwszym skupiono się na problematyce dotyczącej prognozowania hierarchicznego. Dokładnie omówiono na czym polega, z czym jest powiązane oraz opisano jego aspekt teoretyczny. W kolejnym rozdziale również zamieszczono teorię, lecz w tym wypadku powiązaną z prognozowaniem automatycznym, innymi ważnymi metodami użytymi w badaniu oraz sposób w jaki oceniano błędy prognozy. Dużą uwagę poświęcono zwłaszcza głównie użytym w badaniu rodzinom modeli, czyli Prophet, ETS i ARIMA. Następnie zaczyna się część praktyczna, czyli rozdział trzeci. W tym miejscu wykorzystano prognozowanie

hierarchiczne do budowy prognozy zużycia energii elektrycznej w USA. Najpierw zostały przedstawione wykorzystane w niniejszej pracy dane, omówiono ich pochodzenie, zastosowane transformacje i przeprowadzono wstępną analizę. Kolejną część tego rozdziału stanowi opis technologii, narzędzi i bibliotek użytych do wykonania badania. Po tym następuje część empiryczna składająca się z sześciu różnych eksperymentów badających i porównujących użyte metody względem trafności oraz czasu wykonywania obliczeń. Na samym końcu znajdują się wnioski wyciągnięte na podstawie wyników uzyskanych w przeprowadzonych doświadczeniach.

1. Podejście hierarchiczne w prognozowaniu gospodarczym

1.1. Prognozowanie hierarchiczne

Istnieje wiele metod wykorzystywanych do prognozowania począwszy od trywialnych metod naiwnych do zaawansowanych np. opartych na sieciach neuronowych. Podobnie jak w każdej dziedzinie są podejścia bardziej i mniej popularne, każda z nich ma swoje cechy charakterystyczne, zbudowana jest w oparciu o inne założenia i może służyć prognozowaniu konkretnych zjawisk. W zależności od specyfiki problemu jedne będą dawać lepsze wyniki, a inne gorsze, a odpowiedni dobór metody pozwoli na uzyskanie jak najlepszych rezultatów. Najpopularniejszym podejściem w praktyce jest wyodrębnienie jednej, najbardziej trafnej prognozy do wybranego szeregu czasowego. Niestety bardzo często taki wybór nie jest możliwy ze względu na brak odpowiedniej ilości danych, które pozwoliłyby trafnie zaprognozować dane zjawisko. By poradzić sobie z tym problemem badacze opracowali pewne metody, które pozwalają wykorzystać wiedzę zawartą w innych szeregach czasowych np. poprzez występujące powiązania strukturalne, a jednym z takich podejść jest prognozowanie hierarchiczne.

Prognozowanie hierarchiczne w porównaniu do innych metod jest dosyć unikalne, można je traktować w pewien sposób jako ich rozwinięcie, gdyż metoda ta bazuje na prognozach otrzymanych przy pomocy właśnie tych innych metod, a po ich uzyskaniu, są one „uzgadniane” w taki sposób by wartości zgadzały się w całej strukturze hierarchicznej szeregów czasowych w danym problemie, co w konsekwencji prowadzi do nowych wyników prognoz.² W przypadku braku takiej struktury hierarchicznej zastosowanie tej metody nie miałoby sensu i prognoza równałaby się tej bazowej otrzymanej przy pomocy np. modelu ARIMA.

Prognozy hierarchiczne są szeroko wykorzystywane w problemach i zagadnieniach, gdzie występują powiązania strukturalne. Przykładowo, w dziedzinie makroekonomii PKB powstaje jako agregat konsumpcji, wydatków rządowych, inwestycji i eksportu netto, przy czym każdy z tych składników powstaje również jako agregat innych zmiennych ekonomicznych. Wykorzystując nieodłączne struktury agregacji takich zmiennych, stosowane są metody uzgadniania prognoz w celu wygenerowania prognoz, które są spójne z

² George Athanasopoulos, Puwasala Gamakumara, Anastasios Panagiotelis, Rob J Hyndman, Mohamed Affan, Hierarchical forecasting. *Macroeconomic forecasting in the era of big data*, ed. P. Fuleky, Springer, 2020, Chapter 21, pp.689-719.

ograniczeniami agregacji. Pokazali to George Athanasopoulos, Puwasala Gamakumara, Anastasios Panagiotelis, Rob J. Hyndman w swojej pracy „Hierarchical forecasting. *Macroeconomic forecasting in the era of big data*” (2020) na przykładzie australijskiego PKB, że uzgadnianie prognoz nie tylko zwraca spójne prognozy, ale również poprawia ogólną dokładność prognoz zarówno w przypadku prognoz punktowych, jak i probabilistycznych.³

1.1.1. Prognozowanie energii, a metody hierarchiczne

W niniejszym opracowaniu prognozy hierarchiczne będą budowane w oparciu o dane pochodzące z jednej z najważniejszych we współczesnym świecie dziedzin gospodarki stworzonej przez człowieka - sektora energetycznego. Prognozowanie hierarchiczne jest bardzo szeroko wykorzystywane w tym segmencie systemu gospodarczego, gdyż cechą charakterystyczną obciążenia elektrycznego jest częste reprezentowanie jej jako zagregowanego obciążenia wielu jednostek (np. na poziomie elektrowni, miasta, czy regionu).⁴ Powoduje to, że prognozy dotyczące energii elektrycznej powinny być spójne w tym sensie, że prognoza szeregu zagregowanego powinna być równa sumie prognoz odpowiadających mu w strukturze szeregów zdezagregowanych (zdekomponowanych). Spójność jest niezbędna, gdyż przydział energii elektrycznej na jednym poziomie hierarchii zależy od dostarczenia odpowiedniej ilości z poprzedniego poziomu. Uzyskanie spójnych i dokładnych hierarchicznych prognoz przyczynia się do usprawnienia procesu podejmowania decyzji dla sieci energetycznych, a zwłaszcza takich, gdzie wprowadzono inteligentne systemy monitorowania i zarządzania.

Hierarchiczne prognozy energii elektrycznej są dokonywane zarówno punktowo, jak i przedziałowo. W ostatnich latach zyskały na popularności szczególnie te drugie, ponieważ biorą pod uwagę również niepewność w prognozie, dzięki czemu szersza staje się ogólna perspektywa danego problemu dla osoby podejmującej wybór co w konsekwencji przekłada

³ George Athanasopoulos, Puwasala Gamakumara, Anastasios Panagiotelis, Rob J Hyndman, Mohamed Affan, Hierarchical forecasting. *Macroeconomic forecasting in the era of big data*, ed. P. Fuleky, Springer, 2020, Chapter 21, pp.689-719.

⁴ Souhaib Ben Taieb, Raphael Huser, Rob J Hyndman, Marc G Genton, *Forecasting uncertainty in electricity smart meter data by boosting additive quantile regression.*, *IEEE Transactions on Smart Grid* 7(5), 2016, 2448-2455.

się na optymalność decyzji.⁵ Probabilistyczne prognozy znajdują zastosowanie w bardzo wielu aspektach, są między innymi wykorzystywane w stochastycznych modelach zaangażowania jednostek, planowaniu zaopatrzenia w energię, wyznaczaniu rezerwy operacyjnej, prognozowaniu cen oraz handlu na rynku energii elektrycznej.⁶ Używane są także do zapewnienia równowagi pomiędzy wytwarzaniem i zużyciem na poziomie regionalnym⁷ i dystrybucyjnym⁸. Natomiast na poziomie gospodarstw domowych dostawcy mogą wykorzystywać prognozy probabilistyczne do aktualizacji cen i systemów motywacyjnych dla poszczególnych odbiorców.⁹ Zaletą generowania prognoz probabilistycznych dla całej hierarchii jest również to, że pozwala na to, aby prognozy dla każdego szeregu w hierarchii mogły korzystać z informacji w innych miejscach tej struktury (innymi słowy niepewność na poziomach najniższych jest uwzględniana również na pozostałych poziomach w strukturze). Z kolei prognozy punktowo można uznać za bardziej tradycyjne podejście, chociaż w ostatnich latach również pojawiały się prace wykorzystujące ten sposób, np. w prognozowaniu popytu energii w inteligentnych sieciach.¹⁰ W niniejszej pracy będzie oceniana zarówno prognoza punktowa jak i przedziałowa.

1.1.2. Szereg czasowy i oznaczenia

Szeregiem czasowym nazywany jest taki ciąg danych liczbowych, w którym każda obserwacja y związana jest z konkretnym momentem w czasie t , obserwacja taka jest oznaczana jako y_t dla $t = 1, 2, \dots, T$.¹¹ Na potrzeby przedstawionego w niniejszej pracy

⁵ Berrocal, V. J., Raftery, A. E., Gneiting, T. & Steed, R. C., *Probabilistic weather forecasting for winter road maintenance*, *Journal of the American Statistical Association*, 2010, 105(490), 522–537.

⁶ Hong, T., Pinson, P., Fan, S., Zareipour, H., Troccoli, A. & Hyndman, R. J., *Probabilistic energy forecasting: Global energy forecasting competition 2014 and beyond*, *International Journal of Forecasting* 32(3), 2016, 896–913.

⁷ Cabrera, B. L. & Schulz, F., *Forecasting generalized quantiles of electricity demand: A functional data approach*, *Journal of the American Statistical Association* 112(517), 2017, 127–136.

⁸ Sun, X., Luh, P. B., Cheung, K. W., Guan, W., Michel, L. D., Venkata, S. S. & Miller, M. T., 'An efficient approach to Short-Term load forecasting at the distribution level', *IEEE Transactions on Power Systems* 31(4), 2016, 2526–2537.

⁹ Arora, S. & Taylor, J. W., 'Forecasting electricity smart meter data using conditional kernel density estimation', *Omega* 59, 2016, Part A, 47–59.

¹⁰ F. L. Quilumba, W. Lee, H. Huang, D. Y. Wang, and R. L. Szabados, *Using smart meter data to improve the accuracy of intraday load forecasting considering customer behavior similarities*, *IEEE Transactions on Smart Grid*, 2015, vol. 6, no. 2, pp. 911–918.

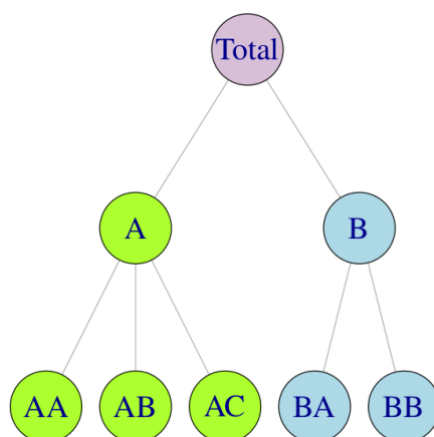
¹¹ Cieślak M., *Prognozowanie Gospodarcze Metody i Zastosowania*, Wydawnictwo Naukowe PWN, Warszawa, 2005

badania wprowadzona zostaje również zmienna \hat{y}_h oznaczająca wartość prognozowanej wartości na podstawie szeregu y dla horyzontu h (zapis skrócony dla $\hat{y}_{T+h|T}$). W opracowaniu występują także wektory, takie obiekty oznaczone są poprzez **pogrubienie** czcionki.

Zagadnienia teoretyczne i notacja zastosowane w badaniu zostały w głównej mierze zaczerpnięte z książki „*Forecasting: principles and practice, 3rd edition*” autorstwa R. J. Hyndman i G. Athanasopoulos.

Hierarchiczny szereg czasowy

W celu zrozumienia w jaki sposób działa prognozowanie hierarchiczne podstawowym elementem będzie zaznajomienie się z pojęciem hierarchii. Mianowicie, jest to układ elementów jakiejś struktury uporządkowany od najwyższych do najniższych według określonego kryterium.¹² Na świecie istnieje niezliczona ilość takich struktur i naturalnie szeregi czasowe również mogą podlegać dekompozycji biorąc pod uwagę jakiś wybrany wyznacznik. Dla przykładu, całkowita liczba produktów sprzedanych w sklepie komputerowym może być podzielona wedle typu produktu na mniejsze kategorie takie jak laptopy, smartfony, komputery stacjonarne itp., a każdy z tych typów na jeszcze mniejsze podkategorie i tak dalej, są one zagnieżdżone w większych grupach, co świadczy o tym, że taka kolekcja szeregów czasowych ma strukturę hierarchiczną. Innym sztandarowym przykładem hierarchii może być również podział administracyjny i biorąc pod uwagę Polskę takimi kategoriami są województwa, powiaty i gminy. Podsumowując, są to szeregi czasowe, pomiędzy którymi występują powiązania ze względu na hierarchię, co można wykorzystać (i nawet należy) w ich prognozowaniu.



Rysunek 1 Diagram dwupoziomowego drzewa hierarchicznego

Źródło: Hyndman, R.J., & Athanasopoulos, G., *Forecasting: principles and practice*, 3rd edition, 2021

¹² Słownik języka polskiego PWN, <https://sjp.pwn.pl/slowniki/hierarchia.html> [dostęp: 01.07.2022]

Na rysunku 1 zamieszczono przykład prostej struktury hierarchicznej. Na samym jej szczycie znajduje się „Total” czyli najwyższy poziom agregacji w tym wypadku. Każda z obserwacji dla szeregu „Total” niech będzie oznaczona jako y_t dla $t = 1, \dots, T$. Następnie ten poziom jest rozdzielony na dwa pomniejsze szeregi czasowe A i B, które z kolei są rozdzielone kolejno na 3 i 2 szeregi czasowe na najniższym poziomie hierarchii. Poniżej najwyższego poziomu używa się oznaczenia $y_{j,t}$, które wskazuje na obserwację t szeregu na gałęzi j . Przykładowo, $y_{A,t}$ oznacza obserwację t na gałęzi A, $y_{AB,t}$ oznacza obserwację t na gałęzi AB itd.

Dla każdego czasu t , obserwacje z niższego poziomu hierarchii sumują się do obserwacji szeregów z wyższego poziomu, np.:

$$y_t = y_{AA,t} + y_{AB,t} + y_{AC,t} + y_{BA,t} + y_{BB,t}, \quad (1.1)$$

$$y_{A,t} = y_{AA,t} + y_{AB,t} + y_{AC,t} \text{ oraz } y_{B,t} = y_{BA,t} + y_{BB,t} \quad (1.2)$$

Podstawiając (1.2) do (1.1) otrzymuje się $y_t = y_{A,t} + y_{B,t}$.

Taką strukturę hierarchiczną można bardziej efektywnie przedstawić zapisując je macierzowo, co pozwoli na lepsze zobrazowanie działania poszczególnych podejść w prognozowaniu hierarchicznym w dalszej części. Na przykładzie struktury z rysunku 1 taka macierz może wyglądać następująco:

$$\begin{bmatrix} y_t \\ y_{A,t} \\ y_{B,t} \\ y_{AA,t} \\ y_{AB,t} \\ y_{AC,t} \\ y_{BA,t} \\ y_{BB,t} \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} y_{AA,t} \\ y_{AB,t} \\ y_{AC,t} \\ y_{BA,t} \\ y_{BB,t} \end{bmatrix} \quad (1.3)$$

Co można zapisać jako¹³:

$$\mathbf{y}_t = \mathbf{S}\mathbf{b}_t \quad (1.4)$$

Gdzie:

- \mathbf{y}_t – n -wymiarowy wektor wszystkich obserwacji w hierarchii w czasie t ,

¹³ Hyndman, R.J., & Athanasopoulos, G., *Forecasting: principles and practice*, 3rd edition, 2021, OTexts: Melbourne, Australia. OTexts.com/fpp3.

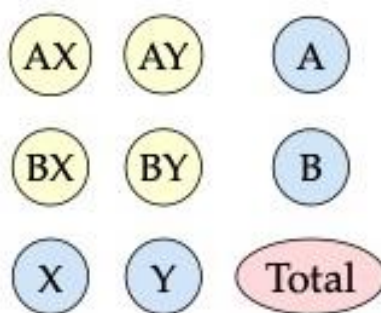
- S – macierz sumaryczna, która wskazuje w jaki sposób szeregi są agregowane, tam, gdzie szeregi z niższego poziomu są elementem danego szeregu z wyższego poziomu (z wektora y_t) występują 1,
- b_t – m -wymiarowy wektor wszystkich obserwacji na najniższym poziomie hierarchii w czasie t .

Ogólnie, $y_t \in \mathbb{R}^n$ oraz $b_t \in \mathbb{R}^m$ gdzie n oznacza liczbę wszystkich szeregów w strukturze, m liczbę szeregów na najniższym poziomie. Zawsze $n > m$. W tym przykładzie, $n = 8, m = 5$.¹⁴ Na podstawie wprowadzonego w tym podrozdziale zapisu macierzowego będą opisywane wybrane metody uzgadniania prognoz w dalszej części opracowania.

Zgrupowany szereg czasowy

Inną ważną alternatywną strukturą agregacji jest zgrupowany szereg czasowy. W tym wypadku szereg na najwyższym poziomie jest dekomponowany według wybranych atrybutów, które są krzyżowane, a nie zagnieżdżone tak jak to ma miejsce w przypadku hierarchicznych szeregów czasowych.

Na przykład, produkcja przemysłowa może być zdekomponowana wzdłuż linii geograficznych lub sektorowych lub obu. Określane jest to jako struktura zgrupowana.



Rysunek 2 Prosta dwupoziomowa zgrupowana struktura.

Źródło: George Athanasopoulos, Puwasala Gamakumara, Anastasios Panagiotelis, Rob J Hyndman and Mohamed Affan (2020), Hierarchical forecasting. *Macroeconomic forecasting in the era of big data*.

Na rysunku 2 przedstawiono prostą dwupoziomową strukturę zgrupowaną. Można zauważyć, że szereg „Total” dekomponuje się na $y_{A,t}$ i $y_{B,t}$, ale także na $y_{X,t}$ i $y_{Y,t}$, a następnie na najniższym poziomie w $b_t = (y_{AX}, y_{AY}, y_{BX}, y_{BY})'$. Zatem, w przeciwieństwie do struktur

¹⁴ Hyndman, R.J., & Athanasopoulos, G., *Forecasting: principles and practice*, 3rd edition, 2021, OTexts: Melbourne, Australia. OTexts.com/fpp3.

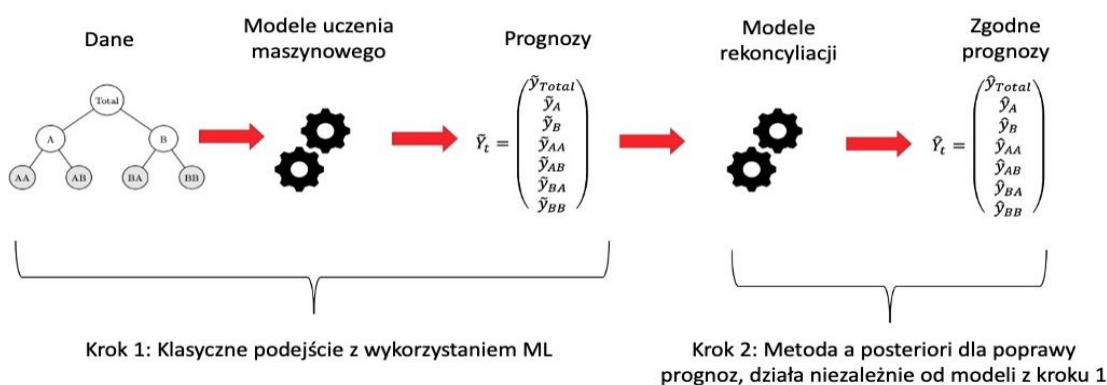
hierarchicznych, struktury zgrupowane nie ulegają naturalnej dekompozycji w unikalny sposób.¹⁵

W niniejszym opracowaniu badane są wyłącznie struktury hierarchiczne.

1.1.3. Uzgadnianie prognoz

Prognozy są najczęściej wymagane dla wszystkich szeregów czasowych na każdym poziomie hierarchii i naturalnym jest, żeby otrzymane wartości poszczególnych przewidywań sumowały się tak, aby były zgodne z tym co otrzymano dla wyższego poziomu (np. suma zaprognozowanego zapotrzebowania na prąd w powiatach powinna się równać prognozie dla całego województwa itp.). Niestety najczęściej jest tak, że nie są one zgodne (ang. incoherent), a wynika to z tego, że dla każdego szeregu czasowego tworzony jest osobny model (często za pomocą innej metody), wybrany w taki sposób by dany szereg dawał jak najlepszych wyniki, a zatem wszystkie prognozy są generowane niezależnie, co powoduje problem niezgodności.

By poradzić sobie z tym problemem z pomocą przychodzi uzgadnianie (ang. reconciliation), która ma celu dostosowanie wartości prognoz otrzymanych przy pomocy poszczególnych modeli prognostycznych. Jej sposób działania wygląda tak, że po procesie prognozowania, wybrana metoda uzgadniania tak modyfikuje wyniki, aby były one zgodne tj. agregowały się poprawnie na wszystkich poziomach hierarchii. Jak pokazują badania, taki proces uzgadniania również poprawia dokładność uzyskanych prognoz.¹⁶



Rysunek 3 Schemat działania uzgadniania prognoz

Źródło: https://medium.com/@adeforceville_96412/hierarchical-timeseries-reconciliation-58addce2aeb7 [dostęp: 01.07.2022]

¹⁵ George Athanasopoulos, Puwasala Gamakumara, Anastasios Panagiotelis, Rob J Hyndman, Mohamed Affan, Hierarchical forecasting. *Macroeconomic forecasting in the era of big data*, ed. P. Fuleky, Springer, 2020, Chapter 21, pp.689-719.

¹⁶ Tamże.

Najważniejsze podejścia do uzgadniania prognoz:

1. **Bottom-up** - jest to metoda uzgadniania, gdzie tworzone są prognozy wyłącznie dla szeregów czasowych na najniższym poziomie, które następnie są sumowane w celu otrzymania prognoz na pozostałych (wyższych) poziomach w danej hierarchii. Tak uzyskane wartości są zgodne w całej strukturze.
2. **Top-down** - jest to metoda można powiedzieć odwrotna to opisywanej poprzednio bottom-up, w tym wypadku najpierw generowana jest prognoza dla szeregów czasowych znajdujących się na najwyższym poziomie hierarchii, a następnie następuje dekompozycja na niższe poziomy. W praktyce dokonywane jest to za pomocą zestawu proporcji dezagregacji (uzyskanych w oparciu o dane), które określają, w jaki sposób mają być rozłożone prognozy szeregu z wyższego poziomu na niższy.
3. **Middle-out** - trzecie opisywane podejście łączy zarówno elementy bottom-up, jak i top-down. Najpierw wybierany jest „środkowy” poziom w hierarchii, a następnie dla wszystkich szeregów na tym szczeblu tworzone są prognozy. Po tym kroku dla szeregów na wyższych poziomach względem wybranego generowane są zgodne prognozy przy użyciu bottom-up, natomiast dla poziomów poniżej wybranego, środkowego poziomu generowane są zgodne prognozy przy użyciu podejścia top-down.
4. **Metody MinT** - metody, które powstały w latach 2018-2020. W tym podejściu należy obliczyć taką macierz \mathbf{G} , gdzie minimalizowana jest całkowita wariancja łączna wszystkich otrzymanych prognoz zgodnych – to podejście nazywane jest optymalnym uzgadnieniem MinT (Minimum Trace).¹⁷ W zależności od zastosowanego w tym podejściu założenia wyróżnia się 4 różne metody MinT: OLS, WLS Var, WLS Struct i Shrink.

W niniejszym opracowaniu zostaną użyte wszystkie wymienione główne metody z wyjątkiem middle-out. Spośród metod MinT zostaną zastosowane OLS i Shrink.

¹⁷ Wickramasuriya, S. L., Athanasopoulos, G., & Hyndman, R. J., *Optimal forecast reconciliation for hierarchical and grouped time series through trace minimization*, *Journal of the American Statistical Association*, 2019, 114(526), 804–819.

1.1.4. Opis wybranych metod uzgadniania

Metody uzgadniania bottom-up oraz top-down są omówione na przykładzie struktury hierarchicznej z rysunku 1.

Bottom-up

W podejściu bottom-up najpierw generowane są prognozy bazowe o horyzoncie h dla każdego szeregu czasowego z najniższego poziomu, czyli:

$$\hat{y}_{A,h}, \hat{y}_{AB,h}, \hat{y}_{AC,h}, \hat{y}_{BA,h}, \text{ oraz } \hat{y}_{BB,h}. \quad (1.5)$$

Sumując je, można otrzymać zgodne prognozy o horyzoncie h dla pozostałych szeregów (zapis \tilde{y}_h oznacza prognozę zgodną):

$$\begin{aligned} \tilde{y}_h &= \hat{y}_{AA,h} + \hat{y}_{AB,h} + \hat{y}_{AC,h} + \hat{y}_{BA,h} + \hat{y}_{BB,h} \\ \tilde{y}_{A,h} &= \hat{y}_{AA,h} + \hat{y}_{AB,h} + \hat{y}_{AC,h} \\ \tilde{y}_{B,h} &= \hat{y}_{BA,h} + \hat{y}_{BB,h} \end{aligned} \quad (1.6)$$

Stosując zapis macierzowy dla prognoz bazowych z równania (1.5) można je oznaczyć jako $\hat{\mathbf{y}}_{i,h}$, gdzie i to poszczególne gałęzie struktury hierarchicznej. W takim wypadku zestaw zgodnych prognoz dla całej hierarchii jest zapisywany jako¹⁸:

$$\widetilde{\mathbf{y}}_h = \mathbf{S}\mathbf{G}\hat{\mathbf{y}}_h \quad (1.7)$$

Gdzie:

- \mathbf{G} - macierz, która mapuje wszystkie bazowe prognozy na najniższym poziomie hierarchii,
- \mathbf{S} - macierz sumująca prognozy bazowe na podstawie struktury tej agregacji,
- $\widetilde{\mathbf{y}}_t$ - otrzymane prognozy zgodne.

¹⁸ Hyndman, R.J., & Athanasopoulos, G., *Forecasting: principles and practice*, 3rd edition, 2021, OTexts: Melbourne, Australia. OTexts.com/fpp3.

Macierz \mathbf{G} jest definiowana na podstawie użytego podejścia, w tym wypadku wygląda w ten sposób:

$$\mathbf{G} = \begin{bmatrix} 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \quad (1.8)$$

Pierwsze trzy kolumny zawierają wyłącznie 0, ponieważ nie odnoszą się do najniższego poziomu, 1 występują tylko tam, gdzie są prognozy bazowe na najniższym poziomie.

Zaletami tej metody jest brak utraty informacji, gdyż prognozowanie jest dokonywane na najniższym poziomie hierarchii, oraz to, że lepiej można uchwycić dynamikę tych szeregów. Z kolei wadami są: duża liczba szeregów czasowych, których prognozy należy dokonać, bardziej skomplikowane tworzenie modeli ze względu na zaszumione dane, co charakteryzuje szeregi na tym poziomie.¹⁹

Top-down

Niech p_1, \dots, p_m oznaczają zestaw proporcji określający sposób w jaki sposób szereg czasowy „Total” ma być rozdzielony na niższych poziomach hierarchii.

Niech proporcje mają oznaczenie p_1, \dots, p_5 , wtedy:

$$\tilde{y}_{AA,t} = p_1 \hat{y}_t, \quad \tilde{y}_{AB,t} = p_2 \hat{y}_t, \quad \tilde{y}_{AC,t} = p_3 \hat{y}_t, \quad \tilde{y}_{BA,t} = p_4 \hat{y}_t \text{ oraz } \tilde{y}_{BB,t} = p_5 \hat{y}_t. \quad (1.9)$$

Jak tylko prognoza o horyzoncie h dla najniższego poziomu zostanie w powyższy sposób wygenerowana, to wtedy następuje agregacja tych prognoz do wyższych poziomów w celu uzyskania prognoz zgodnych.

Występuje kilkanaście różnych odmian metody top-down, główną różnicą wśród nich jest sposób w jaki tworzone są proporcje na podstawie których przeprowadzana jest dekompozycja. Najbardziej popularne stosują podział opierając się na kryterium historii danych lub prognozowanych danych (tj. jak kształtują się proporcje w stosunku do wyższego poziomu biorąc pod uwagę dane szeregów). W tym badaniu zostanie zastosowana metoda

¹⁹ Hyndman, R.J., & Athanasopoulos, G., *Forecasting: principles and practice*, 3rd edition, 2021, OTexts: Melbourne, Australia. OTexts.com/fpp3.

oparta na prognozowanych danych (ang. forecast proportions), ponieważ tylko ona uwzględnia zmiany proporcji w prognozowanym czasie, a inne proporcje tj. historyczne stosowane do dekompozycji nie uwzględniają tego.²⁰

Proporcje na bazie historycznych danych

Pierwsza z tych metod nosi nazwę „Całkowity procent całości”. Omawiając to podejście na przykładzie prognozowania sprzedaży dla okręgów agregujących się do jednego regionu, to proporcje byłyby obliczane jako stosunek całkowitej sprzedaży historycznej w danym, wybranym okręgu, do całkowitej sprzedaży historycznej na wyższym poziomie agregacji, czyli w tym wypadku regionu. I tak kolejno dla poszczególnych okręgów i ich regionów.

Druga z metod opartych na historycznych danych to „Procent okresu całości”. Dana formuła jest podobna do wcześniej opisanej, aczkolwiek, jest ona powtarzana dla każdego okresu, a nie dla sumy we wszystkich okresach występujących w wybranym szeregu czasowym. Wyjaśniając to na tym samym przykładzie co w metodzie „Całkowity procent całości” – przy obliczaniu proporcji będą również brane pod uwagę poszczególne okresy w szeregu, tzn., jeśli zagregowano historyczną sprzedaż na poziomie regionu, proporcja dla okręgu w danym okresie, czyli np. stycznia, wynosi wielkość sprzedaży w tym miesiącu podzieloną przez sprzedaż styczniową dla całego regionu. I tak kolejno dla okresów, okręgów i regionów. Metoda ta jest wrażliwa na wahania sezonowe.²¹

Proporcja dla całego szeregu jest obliczana jako średnia tych poszczególnych proporcji uzyskanych za pomocą jednej z tych metod.

Proporcje na bazie otrzymanych prognoz

Rozważana w tym przypadku hierarchia jest jednopoziomowa. W takim przypadku pierwszym krokiem jest wygenerowanie prognoz o horyzoncie h dla wszystkich szeregów. Są to prognozy „wstępne” i niezgodne. Następnie obliczana jest proporcja każdej takiej wstępnej prognozy o horyzoncie h na najniższym poziomie hierarchii do zagregowanej sumy wszystkich tych prognoz na tym poziomie. Tak otrzymane wartości są nazywane proporcjami

²⁰ Hyndman, R.J., & Athanasopoulos, G., *Forecasting: principles and practice*, 3rd edition, 2021, OTexts: Melbourne, Australia. OTexts.com/fpp3.

²¹ Blessington, Mark. *Sales Forecasting: A Practical Guide*, CreateSpace Independent Publishing Platform, 2015

prognostycznymi i są następnie użyte do dekompozycji prognozy o horyzoncie h na najwyższym poziomie. Dzięki temu są otrzymywane prognozy zgodne dla całej hierarchii.

Dla hierarchii z K -poziomami, proces ten jest powtarzany dla każdej gałęzi, idąc od góry do najniższego poziomu. Stosując to, podejście prowadzi do głównej zasady otrzymywania omawianych proporcji²²:

$$p_j = \prod_{l=0}^{K-1} \frac{\hat{y}_{j,h}^{(l)}}{\hat{S}_{j,h}^{(l+1)}} \quad (1.10)$$

Gdzie $j = 1, 2, \dots, m$, natomiast $\hat{y}_{j,h}^{(l)}$ jest prognozą wstępną o horyzoncie h odpowiadającą gałęzi, która jest l poziomów wyżej nad j , z kolei $\hat{S}_{j,h}^{(l+1)}$ to suma wszystkich prognoz wstępnych o horyzoncie h poniżej gałęzi, która jest l poziomów wyżej nad gałęzią j oraz bezpośrednio połączonych z tą gałęzią. Te proporcje pozwalają na dekompozycję prognoz wstępnych o horyzoncie h z głównego szeregu do zgodnych prognoz o horyzoncie h dla niższych poziomów.

MinT (minimalny ślad)

Tak jak wspomniano wcześniej w podejściu MinT stosuje się taką macierz \mathbf{G} dzięki której dokonuje się minimalizacji całkowitej wariancji łącznej wszystkich otrzymanych prognoz zgodnych.

Aby to podejście zadziałało pierwszym krokiem jest nałożenie ograniczenia na macierz \mathbf{G} , aby prognozy były obiektywne (ang. unbiased). Z tego wynika, że aby prognozy bazowe $\hat{\mathbf{y}}_h$ były obiektywne wtedy prognozy zgodne $\tilde{\mathbf{y}}_h$ takie będą, jeśli $\mathbf{SGS} = \mathbf{S}$.

Następnie należy otrzymać błędy prognoz. Wickramasuriya²³ pokazał, że macierz wariancji-kowariancji zgodnych prognoz horyzoncie h wygląda w sposób następujący²⁴:

$$\mathbf{V}_h = \text{Var}[\mathbf{y}_{T+h} - \tilde{\mathbf{y}}_h] = \mathbf{SGW}_h \mathbf{G}' \mathbf{S}' \quad (1.11)$$

²² Hyndman, R.J., & Athanasopoulos, G., *Forecasting: principles and practice*, 3rd edition, 2021, OTexts: Melbourne, Australia. OTexts.com/fpp3.

²³ Wickramasuriya, S. L., Athanasopoulos, G., & Hyndman, R. J., *Optimal forecast reconciliation for hierarchical and grouped time series through trace minimization*, *Journal of the American Statistical Association*, 114(526), 2019, 804–819.

²⁴ Hyndman, R.J., & Athanasopoulos, G., *Forecasting: principles and practice*, 3rd edition, 2021, OTexts: Melbourne, Australia. OTexts.com/fpp3.

Gdzie $\mathbf{W}_h = \text{Var}[(\mathbf{y}_{T+h} - \tilde{\mathbf{y}}_h)]$ to macierz wariancji-kowariancji korespondujących błędów prognoz bazowych dla horyzontu h .

Następnie celem jest znalezienie takiej macierzy \mathbf{G} , która minimalizuje błędy wariancji prognoz zgodnych. Błędy tych wariancji znajdują się na przekątnej macierzy \mathbf{V}_h , a więc suma wszystkich tych błędów jest otrzymywana jako ślad macierzy \mathbf{V}_h . Ostateczna macierz \mathbf{G} , która minimalizuje ślad \mathbf{V}_h pod warunkiem $\mathbf{SGS} = \mathbf{S}$, wygląda tak²⁵:

$$\mathbf{G} = (\mathbf{S}'\mathbf{W}_h^{-1}\mathbf{S})^{-1}\mathbf{S}'\mathbf{W}_h^{-1} \quad (1.12)$$

Zatem, optymalne prognozy uzyskane dzięki uzgadnianiu otrzymuje się za pomocą równania:

$$\tilde{\mathbf{y}}_h = \mathbf{S}(\mathbf{S}'\mathbf{W}_h^{-1}\mathbf{S})^{-1}\mathbf{S}'\mathbf{W}_h^{-1}\hat{\mathbf{y}}_h \quad (1.13)$$

W praktyce otrzymanie macierzy \mathbf{W}_h może być skomplikowane i dla ułatwienia tego procesu mogą być stosowane cztery przybliżenia upraszczające otrzymując w ten sposób cztery różne metody MinT²⁶:

1. Niech $\mathbf{W}_h = k_h\mathbf{I}$ dla wszystkich h , gdzie $k_h > 0$.

Najbardziej upraszczające założenie ze wszystkich czterech, oznacza, że macierz \mathbf{G} jest niezależna od danych, dzięki czemu całe równanie znacznie upraszcza się obliczeniowo. Wadą tego założenia jest brak uwzględniania różnic w skali między poziomami struktury oraz zależności pomiędzy szeregami.

Stosowanie założenia $\mathbf{W}_h = k_h\mathbf{I}$ w równaniu 1.13 daje estymator MNK. Jest to metoda OLS.

2. Niech $\mathbf{W}_k = k_h \text{diag}(\widehat{\mathbf{W}}_1)$ dla wszystkich h , gdzie $k_h > 0$,

²⁵ Hyndman, R.J., & Athanasopoulos, G., *Forecasting: principles and practice*, 3rd edition, 2021, OTexts: Melbourne, Australia. OTexts.com/fpp3.

²⁶ Tamże.

$$\widehat{\mathbf{W}}_1 = \frac{1}{T} \sum_{t=1}^T \mathbf{e}_t \mathbf{e}_t' \quad (1.14)$$

\mathbf{e}_t to n -wymiarowy wektor reszt modeli, które wygenerowały prognozy bazowe (kolejność taka sama jak w danych). Stosowanie tego założenia pozwala skalować bazowe prognozy używając wariancji reszt, przez to określana jest jako MinT WLS Var (ważona metoda MNK z użyciem skalowania wariancji).

3. Niech $\mathbf{W}_h = k_h \mathbf{\Lambda}$ dla wszystkich h , gdzie $k_h > 0$, $\mathbf{\Lambda} = \text{diag}(\mathbf{S}\mathbf{1})$, $\mathbf{1}$ jest wektorem jednostkowym o wymiarze m (liczba szeregów najniższego poziomu hierarchii).

To przybliżenie zakłada, że każdy błąd prognoz bazowych najniższego poziomu ma wariancję k_h i jest nieskorelowany z innymi. Zatem każdy element przekątnej macierzy $\mathbf{\Lambda}$ zawiera liczb wariancji błędu prognozy przypadającej na każdy węzeł struktury. Estymator ten zależy jedynie od struktury agregacji, a nie od rzeczywistych danych. Z tego względu jest on określany mianem skalowania strukturalnego. Zastosowanie specyfikacji skalowania strukturalnego jest szczególnie użyteczne w przypadkach, gdy nie są dostępne reszty, a więc nie można zastosować skalowania wariancji. Zatem jest to ważona MNK, lecz z użyciem skalowania strukturalnego – WLS Struct.

4. Niech $\mathbf{W}_h = k_h \mathbf{W}_1$ dla wszystkich h , gdzie $k_h > 0$. W tym wypadku występuje założenie, że macierze kowariancji błędów są proporcjonalne do siebie, i bezpośrednio estymowana jest pełna jednostopniowa macierz kowariancji \mathbf{W}_1 . W tym celu używany jest specjalny estymator, który pozwala zapisać otrzymane wartości kowariancji w macierzy diagonalnej. Zastosowanie tego uproszczenia pozwala otrzymać metodę nazywaną MinT Shrink.

Podsumowując, w odróżnieniu od innych istniejących podejść, optymalne prognozy uzgadniania są generowane z wykorzystaniem wszystkich informacji dostępnych w ramach struktury hierarchicznej lub zgrupowanej. Jest to ważne, ponieważ poszczególne poziomy agregacji lub grupowania mogą ujawnić cechy danych, które są interesujące dla użytkownika i ważne do modelowania. Cechy te mogą być całkowicie ukryte lub trudne do zidentyfikowania na innych poziomach.

Na przykład, rozważając dane dotyczące turystyki, gdzie struktura hierarchiczna odpowiada geograficznemu podziałowi kraju na stany i regiony. Niektóre obszary będą w dużej mierze destynacjami letnimi, podczas gdy inne mogą być destynacjami zimowymi. Różnice te zostaną wygładzone na poziomie kraju ze względu na agregację danych.

1.1.5. Prognozowanie hierarchiczne, a automatyczne

Prognozowanie hierarchiczne niesie za sobą również wiele wyzwań. Po pierwsze, szeregi na różnych poziomach hierarchii, odpowiadające różnym stopniom agregacji, mogą wykazywać wyraźnie różne cechy. Każdy z szeregów ma zazwyczaj charakterystyczne dla siebie właściwości np. stosunkowo dużą zmienność, trend czy występowanie skośności. Dodatkowo, drugim ważnym aspektem jest liczność szeregów w takiej strukturze, przykładowo, nawet przy prognozowaniu hierarchicznym dla polskiego podziału administracyjnego, na najniższym poziomie tej hierarchii występowałoby aż 2477 gmin, czyli 2477 szeregów czasowych, dla których należałoby wybrać odpowiednie parametry w bardziej skomplikowanych modelach, by otrzymać prognozy. Można z pewnością stwierdzić, że manualne dokonanie tych czynności byłoby bardzo praco- oraz czasochłonne, na szczęście, z pomocą przychodzi tutaj prognozowanie automatyczne.

2. Prognozowanie automatyczne i inne metody

2.1. Istota prognozowania automatycznego

Termin prognozowanie automatyczne opisuje system prognozowania, który, poza pewnymi wstępnymi specyfikacjami, wymaga jedynie wejścia obserwowanego szeregu czasowego w celu wygenerowania zestawu prognoz. Oznacza to, że wybór schematu prognozowania, z pewnego wstępnie określonego zestawu możliwości, odbywa się bez interwencji użytkownika.²⁷ W odróżnieniu od prognozowania manualnego użytkownik nie musi sam określać parametrów, obliczać prognoz, czy wybierać odpowiedniego modelu prognostycznego – wszystko jest dokonywane przez algorytm. Automatyczne algorytmy muszą być odporne na nietypowe wzorce szeregów czasowych i mieć zastosowanie do dużej liczby szeregów bez interwencji użytkownika.²⁸

W problemach biznesowych są najczęściej wykorzystywane do prognozy dużej liczby jednowymiarowych szeregów czasowych np. ponad tysiąca linii produktów, które wymagają prognozowania co najmniej raz w miesiącu. A nawet jeśli wymagana jest mniejsza liczba prognoz, to często w przedsiębiorstwach może nie być nikogo o odpowiednich kwalifikacjach do stworzenia manualnej prognozy. W takiej sytuacji, algorytm automatycznego prognozowania jest niezbędnym narzędziem.

2.2. Metody stosowane w prognozach automatycznych

Najpopularniejsze algorytmy automatycznego prognozowania oparte są na modelach wygładzania wykładniczego (ETS) lub ARIMA. Dodatkowo dużym uznaniem cieszą się między innymi takie metody jak STLM, czy TBATS.²⁹

2.2.1. ETS

Rodzina ETS (metody wygładzania wykładniczego) to jedna z najpopularniejszych podejść do modelowania szeregów czasowych, a główny nacisk jest w niej położony na opis zachowania trendu i sezonowości danych. Głównym założeniem, na którym oparte jest

²⁷ Ord, Keith & Lowe, Sam., *Automatic Forecasting*, American Statistician - AMER STATIST. 50. 88-94., 1996, 10.1080/00031305.1996.10473549.

²⁸ Rob J Hyndman, Yeasmin Khandakar, *Automatic time series forecasting: the forecast package for R*, *Journal of Statistical Software* 27 (3), 2008

²⁹ Rob J Hyndman, *Automatic forecasting algorithms*, 2017, <https://robjhyndman.com/files/2-AutomaticForecasting.pdf> [dostęp: 15.07.2021]

tworzenie prognoz w rodzinie ETS to waga obserwacji. Dokładniej rzecz ujmując, prognozy są średnimi ważonymi poprzednich obserwacji, przy czym wagi maleją wykładniczo w miarę starzenia się obserwacji. Innymi słowy, im nowsza obserwacja, tym wyższa związana z nią waga. Taki sposób pozwala na wiarygodne i szybkie prognozy dla szerokiego zakresu szeregów czasowych, co jest wielką zaletą i ma duże znaczenie dla wielu zastosowań.³⁰

Istnieje cała rodzina modeli ETS, które można rozróżnić na podstawie rodzaju stosowanego błędu, trendu i sezonowości. Ogólnie rzecz biorąc, trend może być nieistniejący, addytywny, multiplikatywny, addytywny wygaszony lub multiplikatywny wygaszony. Sezonowość może być nieobecna, addytywna lub multiplikatywna. Błąd może być addytywny lub multiplikatywny, aczkolwiek rozróżnienie tych dwóch opcji ma konsekwencje tylko dla przedziałów predykcji, a nie dla prognoz punktowych. W sumie istnieje 30 modeli z różnymi kombinacjami błędu, trendu i sezonowości.³¹ Dla rozróżnienia modeli stosowana jest notacja ETS (błąd, trend, sezonowość). Możliwości dla każdego z nich są następujące: błąd = {A, M}, trend = {N, A, A_d} oraz sezonowość = {N, A, M}.

Oznaczenia:

- N – brak komponentu w modelu,
- A – komponent addytywny,
- A_d – komponent addytywny wygaszony (tylko trend),
- M – komponent multiplikatywny.

Jeśli chodzi o budowę modeli ETS to występuje w nich główne równanie prognostyczne (ang. forecast equation), oraz w zależności od obecności komponentów, dodatkowe równania wygładzające (ang. smoothing equations) opisujące komponenty obecne w równaniu prognostycznym. Podstawowe wygładzanie wykładnicze może być rozszerzone o modelowanie wspomnianych wcześniej składników szeregu, czyli trendu, sezonowości i reszt, gdzie trend opisuje długoterminowy kierunek szeregu, a część sezonowa obejmuje powtarzające się składniki o znanej okresowości, natomiast komponent resztowy odnosi się do niemożliwych do zaprognozowania składników. Przykładowo, bazowe równania modelu ETS (A, N, N) w formie komponentowej można zapisać w tym wypadku jako:

³⁰ Hyndman, R.J., & Athanasopoulos, G., *Forecasting: principles and practice*, 3rd edition, 2021, OTexts: Melbourne, Australia. [OTexts.com/fpp3](https://otexts.com/fpp3).

³¹ Christoph Bergmeir, Rob J Hyndman, José M Benítez, *Bagging exponential smoothing methods using STL decomposition and Box-Cox transformation*, *International Journal of Forecasting* **32**(2), 2016, 303-312.

$$\hat{y}_{t+1|t} = l_t \quad (1.15)$$

$$l_t = \alpha y_t + (1 - \alpha)l_{t-1}, \quad (1.16)$$

gdzie (1.15) to główne równanie prognostyczne, a (1.16) to równanie wygładzające. l_t to poziom szeregu (wygładzona wartość) w czasie t , natomiast α to parametr wygładzający (małe wartości tego parametru to większa waga dla obserwacji z dalekiej przeszłości, duża wartość to duża waga dla najnowszych). W przypadku uwzględnienia również trendu b_t czy sezonowości s_t w równaniu (1.15) pojawiłyby się te czynniki w postaci dodatkowych równań wygładzających. Na rysunku 4 przedstawiono równania poszczególnych modeli ETS w zależności od obecności trendu i sezonowości.

Trend	Seasonal		
	N	A	M
N	$\hat{y}_{t+h t} = \ell_t$	$\hat{y}_{t+h t} = \ell_t + s_{t+h-m(k+1)}$	$\hat{y}_{t+h t} = \ell_t s_{t+h-m(k+1)}$
	$\ell_t = \alpha y_t + (1 - \alpha)\ell_{t-1}$	$\ell_t = \alpha(y_t - s_{t-m}) + (1 - \alpha)\ell_{t-1}$	$\ell_t = \alpha(y_t/s_{t-m}) + (1 - \alpha)\ell_{t-1}$
		$s_t = \gamma(y_t - \ell_{t-1}) + (1 - \gamma)s_{t-m}$	$s_t = \gamma(y_t/\ell_{t-1}) + (1 - \gamma)s_{t-m}$
A	$\hat{y}_{t+h t} = \ell_t + hb_t$	$\hat{y}_{t+h t} = \ell_t + hb_t + s_{t+h-m(k+1)}$	$\hat{y}_{t+h t} = (\ell_t + hb_t)s_{t+h-m(k+1)}$
	$\ell_t = \alpha y_t + (1 - \alpha)(\ell_{t-1} + b_{t-1})$	$\ell_t = \alpha(y_t - s_{t-m}) + (1 - \alpha)(\ell_{t-1} + b_{t-1})$	$\ell_t = \alpha(y_t/s_{t-m}) + (1 - \alpha)(\ell_{t-1} + b_{t-1})$
	$b_t = \beta^*(\ell_t - \ell_{t-1}) + (1 - \beta^*)b_{t-1}$	$b_t = \beta^*(\ell_t - \ell_{t-1}) + (1 - \beta^*)b_{t-1}$	$b_t = \beta^*(\ell_t - \ell_{t-1}) + (1 - \beta^*)b_{t-1}$
Ad	$\hat{y}_{t+h t} = \ell_t + \phi_h b_t$	$\hat{y}_{t+h t} = \ell_t + \phi_h b_t + s_{t+h-m(k+1)}$	$\hat{y}_{t+h t} = (\ell_t + \phi_h b_t)s_{t+h-m(k+1)}$
	$\ell_t = \alpha y_t + (1 - \alpha)(\ell_{t-1} + \phi b_{t-1})$	$\ell_t = \alpha(y_t - s_{t-m}) + (1 - \alpha)(\ell_{t-1} + \phi b_{t-1})$	$\ell_t = \alpha(y_t/s_{t-m}) + (1 - \alpha)(\ell_{t-1} + \phi b_{t-1})$
	$b_t = \beta^*(\ell_t - \ell_{t-1}) + (1 - \beta^*)\phi b_{t-1}$	$b_t = \beta^*(\ell_t - \ell_{t-1}) + (1 - \beta^*)\phi b_{t-1}$	$b_t = \beta^*(\ell_t - \ell_{t-1}) + (1 - \beta^*)\phi b_{t-1}$
		$s_t = \gamma(y_t - \ell_{t-1} - \phi b_{t-1}) + (1 - \gamma)s_{t-m}$	$s_t = \gamma(y_t/(\ell_{t-1} + \phi b_{t-1})) + (1 - \gamma)s_{t-m}$

Rysunek 4 Równania ETS

Źródło: Hyndman, R.J., & Athanasopoulos, G., *Forecasting: principles and practice*, 3rd edition, 2021, OTexts: Melbourne, Australia. OTexts.com/fpp3.

2.2.2. ARIMA

W odróżnieniu od modeli ETS, bazujących na zachowaniu trendu i sezonowości, w modelach ARIMA główny nacisk jest położony na opisanu autokorelacji pomiędzy badanymi danymi.

Modele ARIMA to połączenie modelu autoregresyjnego (AR), modelu średniej ruchomej (MA) oraz uwzględnienia różnicowania danych (I). Taki model można zapisać następująco:

$$y'_t = c + \phi_1 y'_{t-1} + \dots + \phi_{1p} y'_{t-p} + \theta_1 \varepsilon_{t-1} + \dots + \theta_q \varepsilon_{t-q} + \varepsilon_t \quad (1.17)$$

W równaniu 1.17 y' oznacza zróżnicowane dane (mogą one być zróżnicowane więcej niż raz). Zmienne objaśniające po prawej stronie równania to opóźnione wartości y_t i opóźnione błędy ε_t . Taki model nazywany jest ARIMA (p, d, q), gdzie:

- p – wielkość opóźnienia części AR,
- d – stopień różnicowania danych I,
- q – wielkość opóźnienia części MA³².

Część AR

Ta część modelu opisuje, w jaki sposób każda obserwacja jest funkcją poprzednich p obserwacji. Na przykład, jeśli $p = 1$, to każda obserwacja jest funkcją tylko jednej poprzedniej obserwacji.

Część I

Część modelu określająca czy obserwowane wartości są modelowane bezpośrednio, czy też zamiast tego modelowane są różnice między kolejnymi obserwacjami. Jeśli $d = 0$, obserwacje są modelowane bezpośrednio. Jeśli $d = 1$, modelowane są różnice między kolejnymi obserwacjami. Jeżeli $d = 2$, modelowane są różnice różnic. W praktyce d rzadko jest większe niż 2.

Część MA

Część modelu opisująca, w jaki sposób każda obserwacja jest funkcją poprzednich q błędów. Na przykład, jeśli $q = 1$, to każda obserwacja jest funkcją tylko jednego poprzedniego błędu.

Sezonowość

Modele ARIMA są w stanie również modelować dane sezonowe (model SARIMA), w takim wypadku model ten składa się z dodatkowej części opisującej sezonowość. Zapisywane to jest w ten sposób:

$$ARIMA(p, d, q)(P, D, Q)_m$$

Gdzie parametr m określa okres sezonowości (np. liczba obserwacji na rok). Wielkimi literami zapisywana jest część sezonowa, natomiast małymi część niesezonowa.

³² Rob J Hyndman, *ARIMA processes*, Informed Student Guide to Management Science, 2002

2.2.3. STLM

W tej metodzie stosowana jest dekompozycja STL w celu rozdzielenia poszczególnych komponentów szeregu czasowego tj. składnika sezonowego, składnika trendu oraz składnika resztowego. Następnie składnik sezonowy oraz dane wyrównane sezonowo (tj. dane po usunięciu sezonowości, czyli trend i składnik resztowy) są prognozowane oddzielnie, a następnie takie prognozy są łączone, by uzyskać wartości prognoz dla oryginalnych danych.

2.2.4. TBATS

Model TBATS został stworzony w celu prognozowania szeregów czasowych ze złożonymi wzorcami sezonowymi przy użyciu wygładzania wykładniczego³³. W celu zbudowania tej metody zastosowano sezonowość trygonometryczną, transformację Boxa-Coxa, model ARMA, komponenty trendu i sezonowości. Konsekwencją takiej budowy jest większa przestrzeń parametrów co może prowadzić do lepszych prognoz.

2.3. Automatyczny dobór parametrów

Algorytmy prognozowania automatycznego pozwalające na odpowiedni dobór parametrów wyglądają najczęściej różnie w zależności od wybranej metody. Przykładowo przy modelu wygładzania wykładniczego w algorytmie używa się kryterium informacyjnego AIC, podczas gdy w ARIMIE, poza tym wskaźnikiem stosowane są również testy pierwiastka jednostkowego. Użyty algorytm pozwalający na dobór parametrów zależy od budowy danej metody, jednakże schemat działa zazwyczaj w nich na podobnej zasadzie. Schemat dla metody ETS wygląda następująco³⁴:

1. Dla każdego szeregu zastosowanie wszystkich modeli, które są odpowiednie, optymalizując parametry (zarówno parametry wygładzania, jak i początkową zmienną stanu) modelu w każdym przypadku.
2. Wybranie najlepszego z modeli zgodnie z kryterium informacyjnym AIC (inne kryteria np. BIC też mogą być użyte).
3. Opracowanie prognoz punktowych przy użyciu najlepszego modelu (ze zoptymalizowanymi parametrami) dla tylu kroków naprzód, ile jest wymagane.

³³ Alysha M De Livera, Rob J Hyndman, Ralph D Snyder, *Forecasting time series with complex seasonal patterns using exponential smoothing*, *Journal of the American Statistical Association* **106**(496), 2011, 1513-1527.

³⁴ Hyndman, R. J., & Khandakar, Y., *Automatic Time Series Forecasting: The forecast Package for R*, *Journal of Statistical Software*, 27(3), 2008, 1–22.

4. Uzyskanie przedziałów predykcji dla najlepszego modelu przy użyciu jednej z metod analitycznych.

2.4. Inne metody zastosowane w badaniu

W badaniu do celów porównawczych wykorzystano również metodę SNAIVE (jako benchmark) oraz model Prophet.

2.4.1. SNAIVE

SNAIVE to jedna z podstawowych metod prognozowania, która posłuży jako „benchmark”, czyli innymi słowy jako wzorzec, punkt odniesienia, według którego dalsze, bardziej skomplikowane modele będą oceniane i porównywane. Pozwoli to na identyfikację różnic między ocenami prognoz wykorzystanych metod, a w konsekwencji na wybór tej najbardziej skutecznej dla badanego problemu.

Metoda SNAIVE to zmodyfikowana metoda naiwna, jest ona zmieniona w taki sposób by również uwzględniać sezonowość danych. Tak jak w przypadku metody naiwnej prognoza jest równa ostatniej zaobserwowanej wartości, lecz dla odpowiedniego czasu, np. dla danych miesięcznych prognoza na lipiec bieżącego roku, to będzie ta sama wartość co w lipcu poprzedniego roku. Formalnie prognozę dla czasu $T + h$ można zapisać jako:

$$\hat{y}_{T+h|T} = y_{T+h-m(k+1)} \quad (1.18)$$

Gdzie:

- m – okres sezonowy,
- k – liczba całkowita części $((h - 1)/m)$ - tzn. liczba pełnych lat w prognozowanym okresie do czasu $T + h$.

2.4.2. Prophet

Prophet to model prognostyczny zaproponowany przez firmę Facebook (autorzy - S. J. Taylor & B. Lethan) w 2017 roku. Model został stworzony w oparciu o naturę danych pochodzących z portalu społecznościowego Facebook (trendy, sezonowość wielokrotna, święta zmienne) oraz pod kątem wykorzystania do problemów związanych z zagadnieniem nazwanym przez autorów „*forecasting at scale*”, co wiąże się z takim dostosowaniem tego modelu by był on łatwo skalowalny dla różnych potrzeb biznesowych. Dzięki tej elastyczności

może on być np. wykorzystany i konfigurowalny przez osoby nie będące ekspertami w dziedzinie prognozowania, ale z dużą wiedzą dziedzinową, czy również do dużej różnorodności problemów prognostycznych.³⁵

Podstawą jego budowy jest dekompozycyjny model szeregu czasowego z trzema głównymi komponentami: trendem, sezonowością i dniami wolnymi. Jest on wyrażony następującym równaniem:

$$y_t = g_t + s_t + h_t + \epsilon_t \quad (1.19)$$

Gdzie:

- g_t – funkcja trendu, która modeluje nieokresowe zmiany wartości szeregu czasowego,
- s_t – reprezentuje zmiany okresowe (np. sezonowość roczną i tygodniową),
- h_t – reprezentuje efekty dni wolnych, świąt itd., występujących w nieregularnych harmonogramach przez jeden lub więcej dni,
- ϵ_t – reprezentuje wszelkie zmiany, które nie są uwzględnione przez model (błędy, które są białym szumem), ϵ_t ma rozkład normalny.

Twórcy tego rozwiązania stworzyli dwa główne modele trendu g_t : model wzrostu nasyconego (ang. saturating growth model), oraz model liniowy z podziałem na części (ang. piecewise linear model). Są one używane w zależności od tego czy w danym problemie prognostycznych występuje trend nasycony lub stały.

Model wzrostu nasyconego

Tak jak sugeruje nazwa, w przypadku, gdy trend zwiększa się wraz z upływem czasu jako model trendu jest stosowany nasycony model wzrostu, który w jego najprostszej formie wygląda w ten sposób:

$$g_t = \frac{C}{1 + \exp(-k(t - m))} \quad (1.20)$$

³⁵ Taylor SJ, Letham B., *Forecasting at scale*, *PeerJ Preprints*, 2017

Gdzie:

- C – pojemność nośna,
- k – tempo wzrostu,
- m – parametr przesunięcia.

Jest to wyłącznie model wyjściowy, ponieważ stwierdzono, że istnieją dwa ważne aspekty wzrostu w Facebooku, które nie zostały uchwycone w równaniu 1.20. Po pierwsze, pojemność nośna nie jest stała - wraz ze wzrostem liczby ludzi na świecie, którzy mają dostęp do Internetu, rośnie też pułap wzrostu. Zatem zasadnym jest, by stałą pojemność C zastąpić zmienną w czasie pojemnością C_t .

Po drugie, tempo wzrostu nie jest stałe. Nowe produkty mogą znacząco zmienić tempo wzrostu w danym regionie, więc model musi być w stanie uwzględnić zmienne tempo, aby dopasować się do danych historycznych. Zmiany te są uwzględniane za pomocą tak zwanych punktów zmiany (ang. changepoints), są one oznaczane jako S w punktach czasu s_j dla $j = 1, \dots, S$. Dodatkowo należy również zdefiniować wektor poszczególnych zmian jako $\boldsymbol{\delta} \in \mathbb{R}^S$, gdzie δ_j to zmiana współczynnika w punkcie czasowym s_j . Zatem wartość tego współczynnika w czasie t to współczynnik bazowy k plus wszystkie zmiany, które zaszły do tego punktu czasowego:

$$k + \sum_{j:t > s_j} \delta_j \quad (1.21)$$

Można to również zapisać w bardziej przejrzysty sposób definiując wektor $\mathbf{a}(t) \in \{0, 1\}^S$ taki, że:

$$\alpha_{j,t} = \begin{cases} 1, & \text{jeśli } t \geq s_j, \\ 0, & \text{dla pozostałych.} \end{cases}$$

W takim wypadku współczynnik w punkcie czasowym t można zapisać $k + \mathbf{a}_t^T \boldsymbol{\delta}$. Jeśli współczynnik k jest zmieniony to wtedy również musi być odpowiednio dostosowany parametr m w celu połączenia punktów końcowych poszczególnych segmentów. Wtedy prawidłowe dopasowanie w punkcie zmiany j jest obliczane następująco:

$$\gamma_j = (s_j - m - \sum_{l < j} \gamma_l) \left(1 - \frac{k + \sum_{l < j} \delta_l}{k + \sum_{l \leq j} \delta_l}\right) \quad (1.22)$$

Zatem wprowadzając dwie zmiany z pojemnością nośną oraz tempem wzrostu model wyjściowy można przekształcić do następującego modelu:

$$g_t = \frac{C_t}{1 + \exp(-(k + \mathbf{a}_t^T \boldsymbol{\delta})(t - (m + \mathbf{a}_t^T \boldsymbol{\gamma})))} \quad (1.23)$$

Model liniowy z podziałem na części

W przypadku problemów prognostycznych, które nie wykazują nasyczonego wzrostu, stosuje się odcinkowo-stały wzrost, co zapewnia optymalny model o równaniu:

$$g(t) = (k + \mathbf{a}(t)^T \boldsymbol{\delta})t + (m + \mathbf{a}(t)^T \boldsymbol{\gamma}) \quad (1.24)$$

Gdzie:

- k – tempo wzrostu,
- $\boldsymbol{\delta}$ – wektor zmian tempa wzrostu,
- m – parametr przesunięcia,
- $\boldsymbol{\gamma}$ – wektor dopasowania zmian w punktach zmian j , poszczególne γ_j są obliczane jako $-s_j \delta_j$ by zapewnić ciągłość funkcji

W obu opisanych funkcjach trendu punkty zmiany mogą być ustawiane ręcznie przez analityków lub automatycznie przy pomocy rozkładu Laplace’a – częstość występowania punktów zmiany jest określana a priori na podstawie danych historycznych, a następnie przy pomocy podejścia bayesowskiego uzyskuje się rozkład a posteriori. Dzięki takiemu podejściu w prognozach brana jest pod uwagę niepewność trendu.

Sezonowość

Modele sezonowości wykorzystane w modelu Prophet są okresowymi funkcjami t i opierają się na szeregach Fouriera. Niech P będzie okresem regularnym, którego oczekuje się od szeregu czasowego (np. $P = 365.25$ dla danych rocznych, $P = 7$ dla danych tygodniowych). Wtedy można aproksymować arbitralnie gładkie efekty sezonowe za pomocą:

$$s_t = \sum_{n=1}^N \left(a_n \cos\left(\frac{2\pi nt}{P}\right) + b_n \sin\left(\frac{2\pi nt}{P}\right) \right) \quad (1.25)$$

Następnie dopasowanie sezonowości wymaga estymacji $2N$ parametrów $\beta = [a_1, b_1, \dots, a_N, b_N]^T$. Jest to dokonywane poprzez konstrukcję macierz z wektorami sezonowości dla każdej wartości t w historycznych i przyszłych danych. Przykładowo, dla rocznej sezonowości oraz $N = 10$:

$$\mathbf{X}_t = \left[\cos\left(\frac{2\pi(1)t}{365.25}\right), \dots, \sin\left(\frac{2\pi(10)t}{365.25}\right) \right] \quad (1.26)$$

Wtedy komponent sezonowy modelu wygląda następująco:

$$\mathbf{s}_t = \mathbf{X}_t \boldsymbol{\beta} \quad (1.27)$$

W modelu generatywnym przyjmuje się $\boldsymbol{\beta} \sim N(0, \sigma^2)$.

Święta, dni wolne i inne wydarzenia

Ostatnim głównym komponentem modelu Prophet są święta i wydarzenia, które dostarczają duże, lecz najczęściej przewidywalne wstrząsy dla wielu szeregów czasowych przedsiębiorstw. Uwzględnienie tych efektów jest możliwe za pomocą sztucznych zmiennych oraz założeniu, efekty świąt są niezależne.

Dla każdego święta i , niech D_i będzie zbiorem przeszłych i oraz przyszłych dat tego święta. Dalej dodawana jest funkcja wskaźnika reprezentującą, czy czas t jest w czasie święta i , następnie przypisywany jest każdemu świętu parametr κ_i , który jest odpowiednią zmianą w prognozie. Odbywa się to w podobny sposób jak w przypadku sezonowości poprzez wygenerowanie macierzy regresorów:

$$\mathbf{Z}_t = [\mathbf{1}(t \in D_1), \dots, \mathbf{1}(t \in D_L)] \quad (1.28)$$

Wtedy dany komponent:

$$h_t = Z_t \kappa \quad (1.29)$$

Również jak w przypadku sezonowości, występuje założenie a priori, że $\kappa \sim N(0, \nu^2)$.

Często ważne jest, aby uwzględnić efekty dla okna dni wokół konkretnego święta, np. cały weekend Wielkanocny. Aby to osiągnąć, dołączane są dodatkowe parametry dla dni otaczających święto, traktując każdy z dni w oknie wokół święta jako osobne święto.

2.5. Ocena błędu prognozy

Do weryfikacji prognoz otrzymanych przez wcześniej wymienione metody i oceny jej jakości posłużą cztery miary błędów prognozy ex post: MAPE, RMSE, MASE oraz CRPS. Pierwsze trzy oceniają prognozę punktową, natomiast ostatni przedziałową. W dalszej części y_t oznacza wartość zaobserwowaną dla czasu t , natomiast \hat{y}_t wartość prognozowaną dla czasu t .

2.5.1. MAPE

W celu umożliwienia porównania trafności prognoz otrzymywanych z różnych modeli stosuje się średni absolutny procentowy błąd prognozy (*ang. mean absolute percentage error*). Informuje on o średniej wielkości błędów prognoz dla okresu $t = 1, 2, \dots, T$, wyrażonych w procentach rzeczywistych wartości prognozowanej zmiennej.³⁶

$$MAPE = \frac{1}{T} \sum_{t=1}^T \left| \frac{y_t - \hat{y}_t}{y_t} \right| \cdot 100 \quad (1.30)$$

$$MAPE = \frac{MAE}{|\hat{y}_t|} \cdot 100 \quad (1.31)$$

2.5.2. RMSE

Jest to pierwiastek średniego kwadratowego błędu prognozy (*ang. root mean square error*) wyrażany wzorem przedstawionym w równaniu 1.32.

³⁶ Skoczypiec S., *Metody prognozowania: Jakość prognoz*, 2012, http://m6.pk.edu.pl/materialy/mp/MP_02_bledy_prognozy.pdf [dostęp: 01.07.2022]

$$RMSE = \sqrt{\frac{1}{T} \sum_{t=1}^T (y_t - \hat{y}_t)^2} \quad (1.32)$$

2.5.3. MASE

Jest to średni absolutny skalowany błąd (ang. mean absolute scaled error) zaproponowany przez R. J. Hyndman i A. B. Koehler w 2006 roku jako alternatywa dla stosowania błędów procentowych przy porównywaniu trafności prognoz dla szeregów o różnych jednostkach.³⁷

Miara tego błędu jest obliczana w oparciu o błąd z prostej metody prognozowania, zazwyczaj metody naiwnej. Niech e_j oznacza błędy modelu, dla którego obliczana jest miara *MASE*, wtedy dla szeregów bez sezonowości równanie wygląda tak:

$$q_j = \frac{e_j}{\frac{1}{T-1} \sum_{t=2}^T |y_t - y_{t-1}|} \quad (1.33)$$

Natomiast dla szeregów z sezonowością jest używana metoda naiwna sezonowa, przez co błąd obliczany jest w inny sposób:

$$q_j = \frac{e_j}{\frac{1}{T-m} \sum_{t=m+1}^T |y_t - y_{t-m}|} \quad (1.34)$$

m oznacza okres sezonowy.

Następnie błąd *MASE* to:

$$MASE = \frac{1}{T} \sum_{j=1}^T |q_j| \quad (1.35)$$

³⁷ Rob J Hyndman, Anne B Koehler, *Another look at measures of forecast accuracy*, *International Journal of Forecasting* **22**(4), 2006, 679-688.

Gdy $MASE < 1$, badana metoda daje średnio mniejsze błędy niż błędy o jednym kroku t z metody naiwnej, natomiast gdy $MASE > 1$ to wtedy większe błędy.

2.5.4. CRPS

Jest to miara błędu dla prognozy przedziałowej i została zaproponowana przez T. Gneiting i M. Katzfuss w 2014 roku.³⁸ Do obliczenia tej miary wykorzystywany jest tzw. ocena kwantylowa (ang. quantile score), czyli miara poprawności estymacji kwantyla rzędu p . Niech prognoza tego kwantyla będzie oznaczona jako $f_{p,t}$, natomiast y_t obserwacją w czasie t , wtedy ocena kwantylowa jest obliczana jako:

$$Q_{p,t} = \begin{cases} 2(1-p)(f_{p,t} - y_t), & \text{jeśli } y_t < f_{p,t} \\ 2p(y_t - f_{p,t}), & \text{jeśli } y_t \geq f_{p,t} \end{cases} \quad (1.36)$$

Niższe wartości $Q_{p,t}$ oznaczają lepszą estymację kwantyla. Jeśli $p > 0,5$, to negatywny wpływ na wartość $Q_{p,t}$ jest większy w sytuacji, kiedy obserwacja ma większą wartość niż estymowany kwantyl, a mniejszy, gdy ta obserwacja jest niższa. Jeśli $p < 0,5$ to jest odwrotnie.

Miara CRPS służy do oceny całego rozkładu takiej prognozy i polega na uśrednieniu wszystkich wyników ocen kwantyli dla wszystkich wartości p , jest trochę jak ważony błąd bezwzględny obliczony z całego rozkładu prognozy, gdzie waga uwzględnia prawdopodobieństwa.

2.5.5. Skill score

Innym podejściem do oceny prognozy jest wykorzystanie tzw. skill score, co może być wykorzystane zarówno dla oceny trafności prognozy punktowej jak i prognozy przedziałowej.³⁹ W tej metodzie miara błędu prognozy jest obliczana w stosunku do jakiejś metody odniesienia, najczęściej jednej z prostych metod prognostycznych. Na przykład jako punkt odniesienia dla prognoz otrzymanych za pomocą metody dryfu może zostać użyta

³⁸ Gneiting, T., & Katzfuss, M., *Probabilistic forecasting*, *Annual Review of Statistics and Its Application*, 1(1), 2014, 125–151.

³⁹ Hyndman, R.J., & Athanasopoulos, G., *Forecasting: principles and practice*, 3rd edition, 2021, OTexts: Melbourne, Australia. OTexts.com/fpp3.

metoda naiwna, i w takim wypadku skill score np. miary CRPS dla metody dryfu jest w stosunku do metody naiwnej obliczana jako:

$$\frac{CRPS_{naiwna} - CRPS_{dryf}}{CRPS_{naiwna}} \quad (1.37)$$

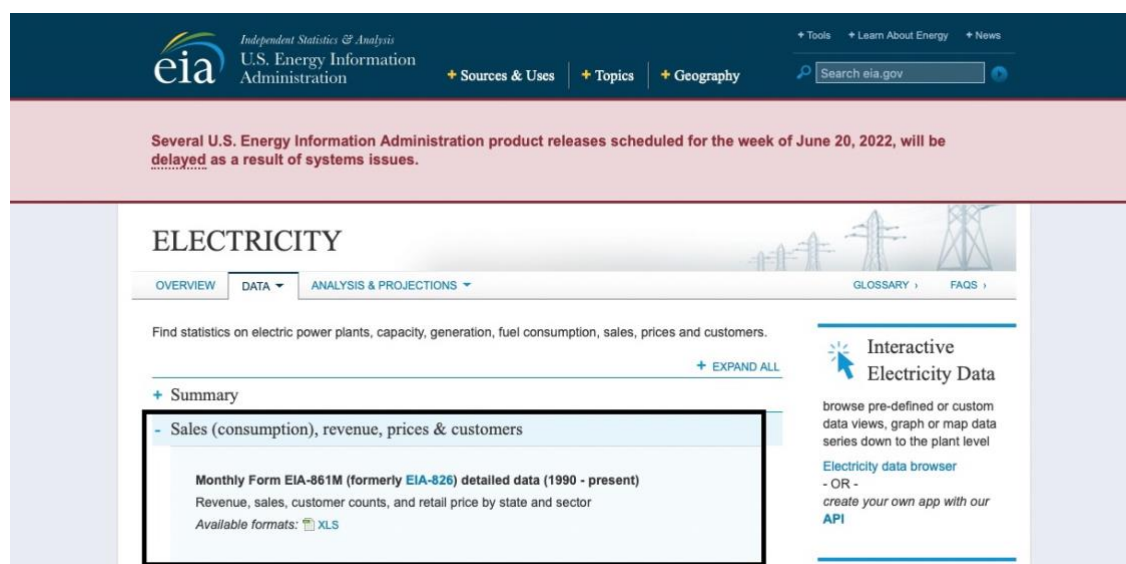
3. Wykorzystanie prognozowania hierarchicznego do budowy prognozy zużycia energii elektrycznej w USA

Aby zrealizować główne cele postawione na początku niniejszego badania zostanie zbudowana prognoza zużycia energii elektrycznej w oparciu o dane pochodzące z USA. Dzięki temu będzie możliwe przeprowadzanie sześciu różnych eksperymentów, których zadaniem jest porównanie poszczególnych rodzin modeli i metod prognozowania hierarchicznego. Ich przebieg i ocena pozwolą na sformułowanie odpowiednich wniosków.

3.1. Dane

3.1.1. Źródło danych

Wykorzystane w badaniu dane pochodzą w całości z bazy danych *U.S. Energy Information Administration (EIA)*.⁴⁰ Jest to to agencja statystyczna oraz analityczna działająca w ramach Departamentu Energii USA. Podmiot ten gromadzi, analizuje i rozpowszechnia bezstronne informacje na temat energii w celu promowania rozsądnego kształtowania polityki, efektywnych rynków oraz publicznego zrozumienia energii i jej interakcji z gospodarką oraz środowiskiem⁴¹. Programy EIA obejmują dane dotyczące węgla, ropy naftowej, gazu ziemnego, energii elektrycznej, odnawialnej i jądrowej.



Rysunek 5 Strona internetowa EIA - miejsce skąd pobrano dane

Źródło: <https://www.eia.gov/electricity/data.php> [dostęp: 01.07.2022]

⁴⁰ U.S. Energy Information Administration, <https://www.eia.gov/electricity/data.php> [dostęp: 01.07.2022]

⁴¹ U.S. Energy Information Administration, https://www.eia.gov/about/mission_overview.php [dostęp: 01.07.2022]

Dane zostały pobrane z miejsca na stronie agencji EIA wskazanego na rysunku 5. W pobranym pliku .xls znajdowały się dane dotyczące sprzedaży (konsumpcji), przychodów, cen i liczby odbiorców energii elektrycznej w USA z podziałem na poszczególne stany (rysunek 6).

	A	B	C	D	Y	Z	AA	AB	AC
1					TOTAL				
2					Revenue	Sales	Customers	Price	
3	Year	Month	State	Data Status	Thousand Dollars	Megawatthours	Count	Cents/kWh	
784	2020	12	KS	Final	322 965	3 257 379	1 547 344	9,91	
785	2020	12	KY	Final	546 064	6 395 482	2 350 519	8,54	
786	2020	12	LA	Final	557 166	7 147 143	2 433 411	7,80	
787	2020	12	MA	Final	763 691	4 253 069	3 229 358	17,96	
788	2020	12	MD	Final	597 744	5 328 475	2 655 460	11,22	
789	2020	12	ME	Final	131 599	986 471	822 245	13,34	
790	2020	12	MI	Final	1 037 427	8 295 290	5 000 920	12,51	
791	2020	12	MN	Final	574 519	5 683 857	2 795 424	10,11	
792	2020	12	MO	Final	561 748	6 655 101	3 258 628	8,44	
793	2020	12	MS	Final	348 970	3 764 667	1 565 101	9,27	
794	2020	12	MT	Final	120 824	1 314 220	648 434	9,19	
795	2020	12	NC	Final	1 040 009	10 958 963	5 460 841	9,49	
796	2020	12	ND	Final	160 283	2 043 445	473 990	7,84	
797	2020	12	NE	Final	222 594	2 691 285	1 092 178	8,27	

Rysunek 6 Nieprzetworzone dane w pliku .xls

Źródło: Opracowanie własne

3.1.2. Charakterystyka zbioru danych

Nieprzetworzone dane przedstawione na rysunku 6 składają się z 19686 wierszy oraz 8 kolumn. Uzyskane obserwacje są miesięczne i pochodzą z lat 1990-2022, według informacji na stronie EIA aktualizowane są one na bieżąco i w momencie ich pobrania (maj 2022), jeśli chodzi o rok 2022, dostępne były tylko dane ze stycznia oraz lutego. Jako że podjęto decyzję, aby zużycie energii elektrycznej badać na poziomie kwartałów zadecydowano o pominięciu roku 2022. Zatem ostatecznie rozważane szeregi czasowe są z okresu 1990-2021. O roku i miesiącu z jakich pochodzą wartości mówią kolejno dwie pierwsze kolumny.

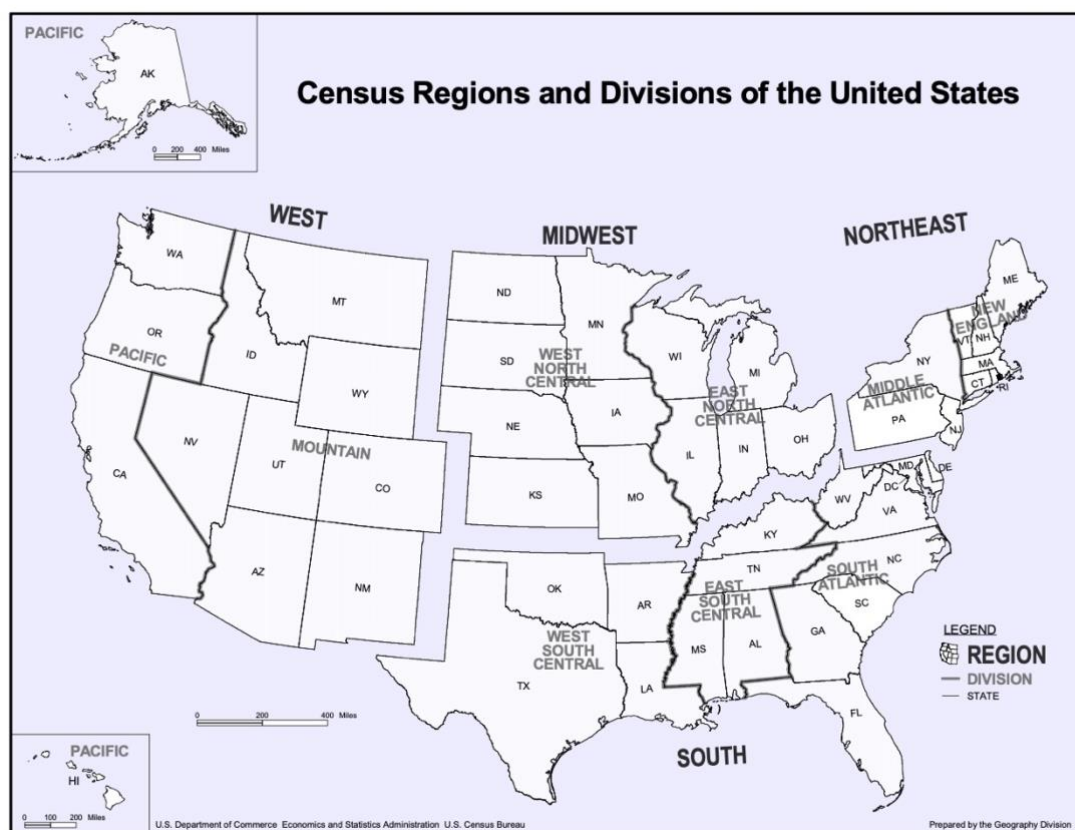
Kolejna kolumna jest oznaczona jako „State” i wskazuje ona na stan USA, którego dotyczy dany rekord. Łącznie występuje w niej 51 unikalnych wartości – oprócz 50 stanów USA wyodrębniony został również Dystrykt Kolumbii (stolica USA – Waszyngton, D.C.).

Następna kolumna tj. „Data Status” mówi o statusie w jakim znajdują się wyciągnięte dane, jeśli chodzi o możliwe korekty. W tym wypadku wszystkie rekordy w arkuszu do roku 2020 mają status „Final”, co oznacza wartości ostateczne. Dla roku 2021 „Data Status” jest określany jako „Preliminary” co oznacza wartości wstępne i możliwe lekkie korekty. Mimo wszystko zdecydowano o pozostawieniu danych o tym statusie w badaniu, gdyż zmiany te z reguły nie są znaczne i ostatecznie dane końcowe są bardzo zbliżone do danych początkowych.

Ostatnią część arkusza z nieprzetworzonymi danymi stanowią cztery kolumny z wartościami i odnoszą się one kolejno do: przychodów ze sprzedaży energii elektrycznej, wielkości sprzedaży (konsumpcja), liczba odbiorców oraz średniej ceny za energię. W niniejszym opracowaniu badane jest zużycie energii elektrycznej, a zatem czynnikiem brany

pod uwagę jest wyłącznie drugi z wymienionych kolumn. Jest on wyrażony w megawatogodzinach (MWh), czyli ilości energii elektrycznej równej pracy wykonywanej przez prąd o mocy 1 megawata w ciągu jednej godziny.

3.1.3. Struktura hierarchiczna



Rysunek 7 Podział USA na jednostki statystyczne

Źródło: https://www2.census.gov/geo/pdfs/maps-data/maps/reference/us_regdiv.pdf [dostęp: 01.07.2022]

Jak wspomniano w rozdziale opisującym prognozowanie hierarchiczne, bardzo ważnym aspektem w niniejszym badaniu jest struktura hierarchiczna jaką charakteryzują się poddawane analizie dane. W niniejszym opracowaniu struktura została oparta na najczęściej używanym systemie klasyfikacji do gromadzenia i analizy danych w USA jakim jest podział opracowany przez Biuro Spisu Ludności USA (ang. *the United States Census Bureau*) stworzony w 1950 roku. W tym systemie na najniższym poziomie są poszczególne stany, następnie grupowane one są w 9 dywizji, one z kolei zawarte są w 4 głównych regionach, co

już ostatecznie składa się na całe państwo.⁴² W niniejszym opracowaniu pominięto poziom regionów, występują wyłącznie trzy kategorie: stany, dywizje i całe państwo. Strukturę przedstawiono na rysunkach 7 i 8. Portoryko i wszystkie inne terytoria USA niewymienione na nich nie są częścią żadnego regionu spisowego ani oddziału spisowego.

U.S. Census Bureau

Census Bureau Regions and Divisions with State FIPS Codes

Region 1: Northeast

**Division 1:
New England**

Connecticut (09)
Maine (23)
Massachusetts (25)
New Hampshire (33)
Rhode Island (44)
Vermont (50)

**Division 2:
Middle Atlantic**

New Jersey (34)
New York (36)
Pennsylvania (42)

Region 2: Midwest*

**Division 3:
East North Central**

Indiana (18)
Illinois (17)
Michigan (26)
Ohio (39)
Wisconsin (55)

**Division 4:
West North Central**

Iowa (19)
Kansas (20)
Minnesota (27)
Missouri (29)
Nebraska (31)
North Dakota (38)
South Dakota (46)

Region 3: South

**Division 5:
South Atlantic**

Delaware (10)
District of Columbia (11)
Florida (12)
Georgia (13)
Maryland (24)
North Carolina (37)
South Carolina (45)
Virginia (51)
West Virginia (54)

**Division 6:
East South Central**

Alabama (01)
Kentucky (21)
Mississippi (28)
Tennessee (47)

**Division 7:
West South Central**

Arkansas (05)
Louisiana (22)
Oklahoma (40)
Texas (48)

Region 4: West

**Division 8:
Mountain**

Arizona (04)
Colorado (08)
Idaho (16)
New Mexico (35)
Montana (30)
Utah (49)
Nevada (32)
Wyoming (56)

**Division 9:
Pacific**

Alaska (02)
California (06)
Hawaii (15)
Oregon (41)
Washington (53)

*Prior to June 1984, the Midwest Region was designated as the North Central Region.

Rysunek 8 Podział stosowany przez Biuro Spisu Ludności USA

Źródło: https://www2.census.gov/geo/pdfs/maps-data/maps/reference/us_regdiv.pdf [dostęp: 01.07.2022]

⁴²United States Census Bureau, https://www.census.gov/programs-surveys/economic-census/guidance-geographies/levels.html#par_textimage_34 [dostęp: 15.07.2022]

3.1.4. Preprocessing danych

Tak jak wspomniano wcześniej, badanie zostanie przeprowadzone na poziomie kwartałów, a zatem pierwszym krokiem w przetwarzaniu danych było dodanie kolumny „Quarter” wskazującej na kwartał w roku.

Kolejny, niemniej ważny krok, to dodanie do zbioru danych struktury hierarchicznej USA, co pozwoli w dalszej części na zastosowanie prognozowania hierarchicznego. Najpierw dodano całe nazwy wszystkich stanów, dzięki czemu łatwiej można było rozpoznawać te jednostki. Następnie opierając się na nich dodano odpowiednie dywizje w kolumnie „Division Name”. Stworzono również kolumnę „Country” z nazwą kraju, co ułatwiło agregację.

Po usunięciu niepotrzebnych kolumn z wartościami i pozostawieniu wyłącznie wielkości sprzedaży wyrażonej w MWh (kolumna „Total”) otrzymano dane wstępnie przetworzone dane przedstawione na rysunku 9.

	A	B	C	D	E	F	G	H
1	Year	Quarter	Month	State	State Name	Division Name	Country	Total
2	2022	1	2	AK	Alaska	Pacific	United States	510 261
3	2022	1	2	AL	Alabama	East South Central	United States	6 826 948
4	2022	1	2	AR	Arkansas	West South Central	United States	4 005 699
5	2022	1	2	AZ	Arizona	Mountain	United States	5 229 574
6	2022	1	2	CA	California	Pacific	United States	16 541 601
7	2022	1	2	CO	Colorado	Mountain	United States	4 343 202
8	2022	1	2	CT	Connecticut	New England	United States	2 214 557
9	2022	1	2	DC	District of Columbia	South Atlantic	United States	805 119
10	2022	1	2	DE	Delaware	South Atlantic	United States	993 852
11	2022	1	2	FL	Florida	South Atlantic	United States	18 069 369
12	2022	1	2	GA	Georgia	South Atlantic	United States	10 792 840
13	2022	1	2	HI	Hawaii	Pacific	United States	653 066

Rysunek 9 Wstępnie przetworzone dane

Źródło: Opracowanie własne

Dotychczasowe kroki zostały przeprowadzone w arkuszu kalkulacyjnym Excel, natomiast poczynając od tego etapu dalszy preprocessing dokonany był w języku R, gdzie załadowano dane do ramki danych (ang. dataframe).

W tym miejscu w pierwszej kolejności stwierdzono, że dane są kompletne, tzn. brak było pustych lub zerowych wartości w badanym zbiorze. Dalej usunięto występujące tzw. „whitespaces”, czyli puste znaki, dzięki czemu możliwa była również zamiana poszczególnych kolumn na odpowiedni typ danych.

Po uzyskaniu tak przetworzonych danych kolejnym aspektem było usunięcie wszystkich obserwacji z roku 2022, a następnie agregacja danych z miesięcy do poziomu kwartałów. W ten sposób liczba wierszy zmniejszyła się z 19686 do 6528. W międzyczasie usunięto również niepotrzebne kolumny. Przekształcone dane z tego etapu przedstawiono na rysunku 10.

	Quarter	StateName	DivisionName	MWh
1	1990 Q1	Alabama	East South Central	13884045
2	1990 Q1	Alaska	Pacific	1207145
3	1990 Q1	Arizona	Mountain	9637034
4	1990 Q1	Arkansas	West South Central	6258447
5	1990 Q1	California	Pacific	50919235
6	1990 Q1	Colorado	Mountain	7975731
7	1990 Q1	Connecticut	New England	7216696
8	1990 Q1	Delaware	South Atlantic	2152873
9	1990 Q1	District of Columbia	South Atlantic	2423308
10	1990 Q1	Florida	South Atlantic	32123975

Rysunek 10 Dane wstępnie przetworzone, kwartalne

Źródło: Opracowanie własne

Na tak przygotowanych danych postanowiono sprawdzić czy w szeregach czasowych dla poszczególnych stanów występują obserwacje odstające. W tym celu użyto metody opisanej przez R. J. Hyndmana w artykule internetowym „Detecting time series outliers”.⁴³ Stosuje on metodę wykrywania outlierów w oparciu o dekompozycję, estymator regresji nieparametrycznej "Friedman's Super Smoother"⁴⁴ i przedział międzykwantylowy IQR. Cały algorytm jest zaimplementowany w funkcji *tsoutliers()* w pakiecie *forecast*. Sprawdzono wszystkie 51 szeregów i stwierdzono, że nigdzie nie występują obserwacje odstające.

Ostatnimi dwoma krokami w preprocessingu danych była zamiana obiektu *dataframe* na *tsibble* (funkcja *as_tsibble()* z pakietu *tsibble*), co w następnym kroku pozwoliło na stworzenie na bazie tego obiektu struktury hierarchicznej (użycie funkcji *aggregate_key()* z pakietu *fable*). W ten sposób dokonana się również agregacja poszczególnych stanów do dywizji, a następnie do całego kraju, co spowodowało wzrost rekordów do 7808 (9 nowych szeregów dla dywizji, 1 dla całego kraju). Finalne dane po preprocessingu przedstawiono na rysunku 11.

	Quarter	DivisionName	StateName	MWh
1	1990 Q1	<aggregated>	<aggregated>	668304422
2	1990 Q2	<aggregated>	<aggregated>	640265910
3	1990 Q3	<aggregated>	<aggregated>	748884728
4	1990 Q4	<aggregated>	<aggregated>	655099602
5	1991 Q1	<aggregated>	<aggregated>	674567539
6	1991 Q2	<aggregated>	<aggregated>	661616318
7	1991 Q3	<aggregated>	<aggregated>	759650457
8	1991 Q4	<aggregated>	<aggregated>	666168736
9	1992 Q1	<aggregated>	<aggregated>	684789788
10	1992 Q2	<aggregated>	<aggregated>	652926693

Rysunek 11 Finalnie przetworzone dane

Źródło: Opracowanie własne

⁴³ Hyndman R, *Detecting time series outliers*, 2021, <https://robjhyndman.com/hyndsight/tsoutliers/> [dostęp: 27.07.2022]

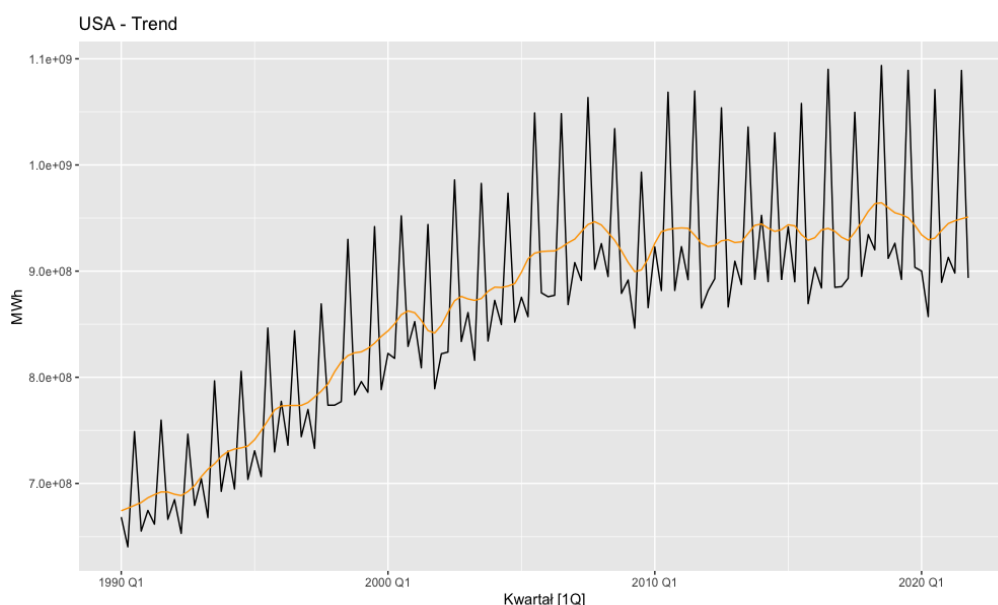
⁴⁴ Luedicke Joerg, *Friedman's Super Smoother*, 2015.

3.1.5. Wstępna analiza danych

Wstępna analiza danych zostanie przeprowadzona na podstawie szeregu czasowego na najwyższym poziomie hierarchii, czyli całego USA. Pozwoli to na zapoznanie się z ogólną charakterystyką danych, czyli np. dostrzeżeniem w jaki sposób zachowuje się w nich trend czy sezonowość.

Trend

Na rysunku 12 przedstawiono wykres zagregowanego zużycia energii elektrycznej dla całego USA. Pomarańczowa linia przedstawia trend w danych uzyskany za pomocą dekompozycji STL. Jak można zauważyć, trend w badanym zjawisku jest wzrostowy, choć w ostatnich lat wzrost jest coraz niższy i stabilizuje się. Wspomniana stabilizacja zaczęła się od około 2009 roku, od tego czasu siła trendu jest mniejsza, można to także zauważyć na rysunku 13, lata ubiegłe bardziej nakładają się na siebie niż te odleglejsze. Trend nie jest również idealnie liniowy, występują pewne fluktuacje i były krótkie okresy, gdzie trend był malejący.



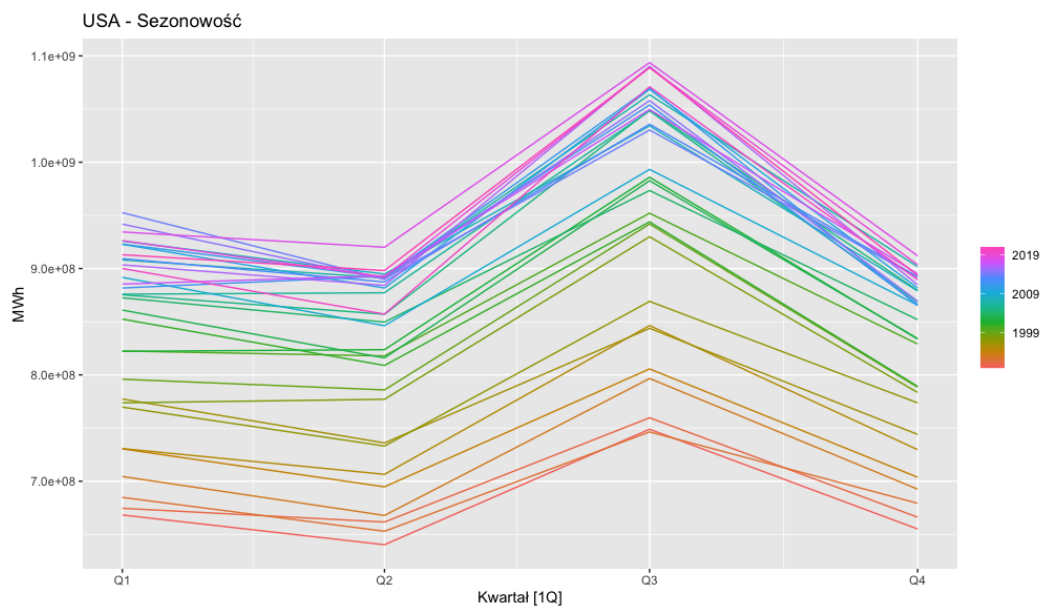
Rysunek 12 USA - trend

Źródło: Opracowanie własne

Sezonowość

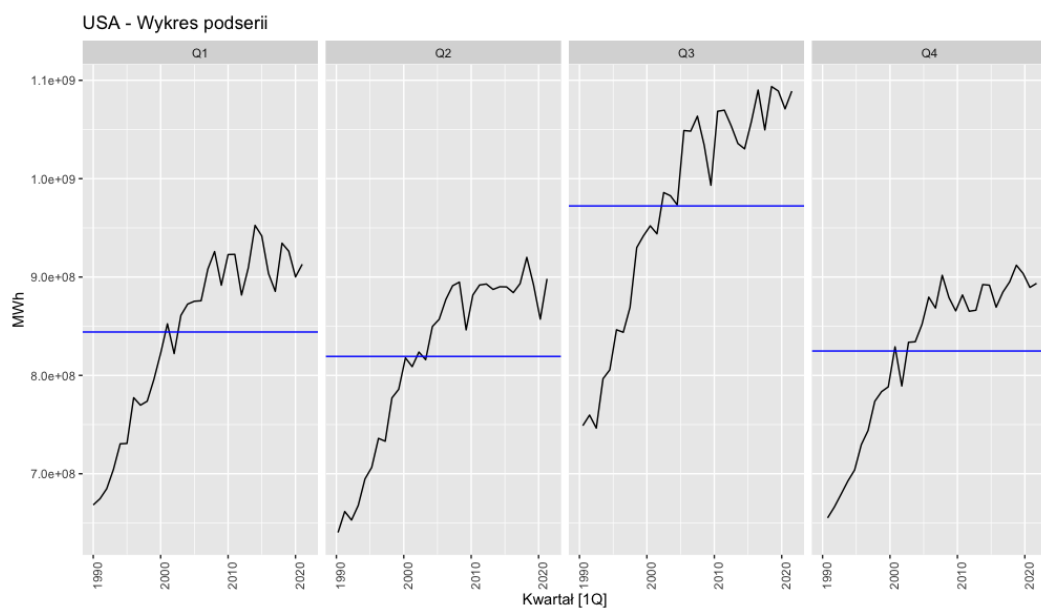
Sezonowość w badanych danych jest bardzo wyraźna i widać ją na wszystkich przedstawionych w tym rozdziale rysunkach (12-15). Zwłaszcza na wykresach przedstawionych na rysunkach 13 i 14 można zauważyć, że największe zużycie energii występuje w kwartale Q3, natomiast kwartały Q1, Q2 i Q4 charakteryzują się podobnym poziomem. Nie jest to zaskoczeniem, gdyż w USA znajduje się w większości taki klimat, gdzie

występują pory roku. Kwartał Q3 to część roku o najwyższych temperaturach powietrza, a co za tym idzie klimatyzacja w amerykańskich domach czy biurach jest w użyciu niemalże codziennie. Taki stan rzeczy powoduje największe zużycie w miesiącach letnich.⁴⁵



Rysunek 13 USA - sezonowość

Źródło: Opracowanie własne



Rysunek 14 USA - wykres podserii

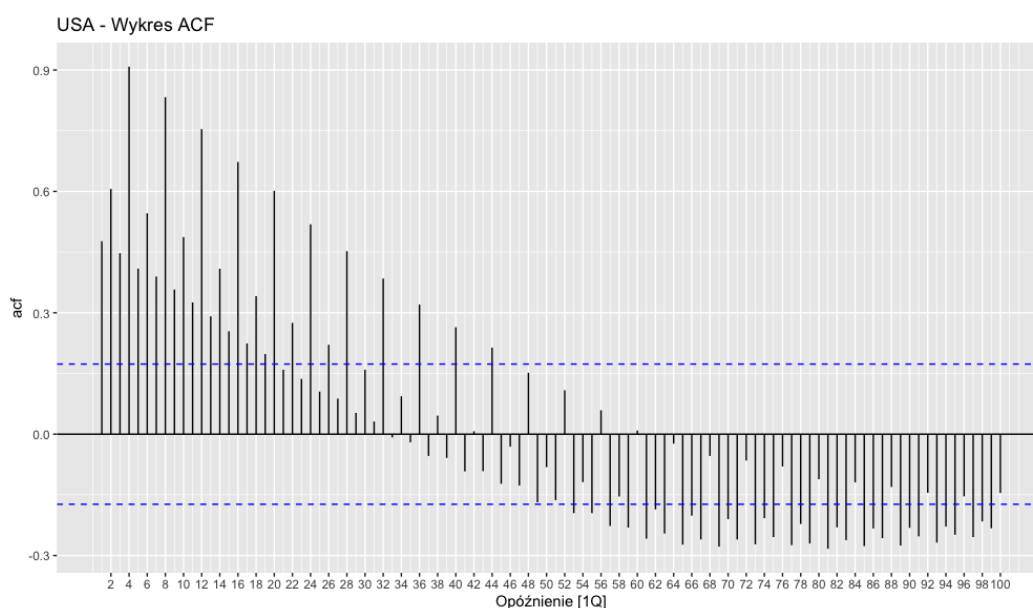
Źródło: Opracowanie własne

⁴⁵ U.S. Energy Information Administration, <https://www.eia.gov/todayinenergy/detail.php?id=42915> [dostęp: 29.07.2022]

Autokorelacja

Na wykresie ACF (rysunek 15) dla 100 opóźnień bardzo dobrze widać dużą wartość funkcji autokorelacji dla wielokrotności liczby 4, co również wskazuje na potwierdzenie sezonowości kwartalnej badanych danych.

Można zauważyć również tendencję, że z wartości dodatnich funkcji autokorelacji, osiągają one powoli coraz mniejsze wartości wraz ze wzrostem opóźnień, co jest charakterystyczne dla ACF szeregu czasowego z trendem.



Rysunek 15 USA - ACF

Źródło: Opracowanie własne

3.1.6. Podział zbioru danych

Rozmiar zbioru testowego wynosi zazwyczaj około 20% całej próbki, chociaż wartość ta zależy od tego, jak długa jest próba i jak daleko w przyszłość będzie dokonywana prognoza. Idealnie zbiór testowy powinien być co najmniej tak duży, jak wymagany maksymalny horyzont prognozy.⁴⁶

W niniejszym badaniu zdecydowano się na dosyć małą próbkę testową wynoszącą 12,5% całego zestawu danych. Zdecydowano się na taki krok ze względu na dosyć krótki szereg czasowy oraz na relatywnie mały horyzont prognozy. Zbiór testowy zawiera 16

⁴⁶Hyndman, R.J., & Athanasopoulos, G., *Forecasting: principles and practice*, 3rd edition, 2021, OTexts: Melbourne, Australia. [OTexts.com/fpp3](https://otexts.com/fpp3).

obserwacji (lata od 2018), podczas gdy zbiór treningowy ma ich 112 (lata do 2017 włącznie). Horyzont prognozy w badaniu wynosi 16 kwartałów, co oznacza, że zbiór testowy jest tak samo liczny.

3.2. Stos technologiczny

3.2.1. Język R

R to język programowania dla obliczeń statystycznych i grafiki wspierany przez R Core Team i R Foundation for Statistical Computing. Język R jest dialektem języka S, który został zaprojektowany w latach 80. i od tego czasu jest szeroko stosowany w środowisku statystycznym. Składnia języka wykazuje powierzchowne podobieństwo do C, ale semantyka jest odmiany FPL (functional programming language) z silniejszymi pokrewieństwami z Lispem i APL. W szczególności pozwala on na "computing on the language", czyli umożliwianie pisanie funkcji przyjmujących wyrażenia jako dane wejściowe, co jest często przydatne w modelowaniu statystycznym i grafice. R i jego biblioteki implementują różne techniki statystyczne, w tym modelowanie liniowe, uogólnione liniowe i nieliniowe, klasyczne testy statystyczne, analizę przestrzenną i szeregów czasowych, klasyfikację, grupowanie i inne. Inną mocną stroną R jest statyczna grafika, może on tworzyć wykresy o jakości publikacji, zawierające symbole matematyczne.⁴⁷

Jest on głównym narzędziem za pomocą, którego przeprowadzono badanie w niniejszym opracowaniu. Dokonywano w nim zarówno wszystkich obliczeń w celu budowy modeli oraz prognoz, a także stworzono większość wizualizacji. Bardzo pomocne w tych aspektach okazały się również stworzone przez społeczność pakiety dostępne w R.

3.2.2. Pakiety R

Pakiety R to zbiory funkcji R, skompilowanego kodu i przykładowych danych, które są przechowywane w katalogu o nazwie *library* w środowisku R. W badaniu użyto kilkanaście z nich. Ważny jest fakt, że pakiety te zazwyczaj są powiązane z innymi pakietami, stąd wypisanie ich wszystkich byłoby dosyć skomplikowanym zadaniem, a zatem zostaną wymienione tylko te najważniejsze:

- *fable* - zawiera zbiór powszechnie stosowanych jedno- i wielowymiarowych modeli prognozowania szeregów czasowych, w tym automatycznie wybierane modele

⁴⁷ The R Project for Statistical Computing, <https://www.r-project.org/about.html> [dostęp: 29.07.2022]

wygładzania wykładniczego (ETS) i autoregresyjnej średniej ruchomej (ARIMA). Modele te działają w ramach *fable* dostarczanych przez pakiet *fabletools*, który dostarcza narzędzi do oceny, wizualizacji i łączenia modeli w sposób spójny z *tidyverse*.⁴⁸

- *fabletools* - dostarcza narzędzi, pomocników i struktur danych do tworzenia modeli i funkcji szeregów czasowych dla *fable* i pakietów rozszerzających. Narzędzia te wspierają spójny i uporządkowany interfejs do modelowania i analizy szeregów czasowych.
- *fable.prophet* - pozwala na używanie modeli Prophet z pakietu *prophet* zgodnie z uporządkowanym interfejsem modelowania *fabletools*. Rozszerza pakiet *prophet*, aby zapewnić ulepszoną specyfikację i zarządzanie modelami, metody oceny wydajności i narzędzia do łączenia modeli.⁴⁹
- *forecast* - pakiet ten implementuje automatyczne prognozowanie przy użyciu wygładzania wykładniczego (ETS), modeli ARIMA, metody Theta, oraz również innych popularnych metod prognozowania.⁵⁰ Zawiera również wiele funkcji przydatnych w analizie szeregów np. *tsouliers()*.
- *feasts* - udostępnia zbiór cech, metod dekompozycji, podsumowań statystycznych i funkcji graficznych do analizy uporządkowanych danych z szeregów czasowych. Nazwa pakietu *feasts* jest akronimem zawierającym jego kluczowe cechy: Feature Extraction And Statistics for Time Series.
- *tsibble* - zapewnia klasę 'tbl_ts' (*tsibble*) dla danych czasowych w formacie zorientowanym na dane i model. *tsibble* dostarcza narzędzi do łatwego manipulowania i analizowania danych czasowych, takich jak wypełnianie luk czasowych i agregowanie w okresach kalendarzowych.
- *tictoc* - udostępnia funkcje pomiaru czasu *tic()* i *toc()*, które mogą być zagnieżdżone. Pozwala na nagrywanie wszystkich czasów podczas działania złożonego skryptu, a następnie na zbadanie tych wartości.

⁴⁸ The R Project for Statistical Computing, <https://cran.r-project.org/web/packages/fable/index.html> [dostęp: 29.07.2022]

⁴⁹ The R Project for Statistical Computing, <https://cran.r-project.org/web/packages/fable.prophet/index.html> [dostęp: 29.07.2022]

⁵⁰ Rob J Hyndman, Yeasmin Khandakar, *Automatic time series forecasting: the forecast package for R*, Journal of Statistical Software 27(3), 2008

- *tidyverse* - jest zbiorem pakietów R zaprojektowanych dla Data Science. Wszystkie pakiety mają wspólną filozofię projektowania, gramatykę i struktury danych.⁵¹ Najważniejsze pakiety zawarte w tej kolekcji i używane w badaniu to *ggplot2* (wizualizacja danych), *dplyr* (manipulacja danych), *tidyr* (czyszczenie danych).
- *lubridate* – zawiera funkcje do pracy z czasem i przedziałami czasowymi, zapewnia szybkie i przyjazne dla użytkownika parsowanie danych czasowych, ekstrakcje i aktualizacje składników czasu (lata, miesiące, dni, godziny, minuty i sekundy), algebraiczne manipulacje na obiektach czasu i przedziałach czasowych.

3.2.3. RStudio

RStudio to zintegrowane środowisko programistyczne (IDE) dla języka programowania R. Zawiera konsolę, edytor z podświetlaniem składni, który obsługuje bezpośrednie wykonywanie kodu, a także narzędzia do tworzenia wykresów, historii, debugowania i zarządzania przestrzenią roboczą. Cała część badania, gdzie wykorzystywano język R została opracowana z wykorzystaniem tego narzędzia.

3.3. Eksperymenty

Tak jak wspomniano w poprzednim rozdziale rozważanymi rodzinami modeli są ETS, ARIMA i Prophet – w dalszej części będą one również określane jako metody/modele bazowe. Wykorzystując metody każdej z tych rodzin zostały zbudowane prognozy kwartalne na lata 2018-2021 (parametry dla poszczególnych szeregów zostały dobrane dzięki pomocy algorytmów prognozowania automatycznego), a następnie tak otrzymane wartości zostały poddane uzgadnianiu za pomocą podejść: Bottom-up, Top-down, MinT OLS i MinT Shrink. Biorąc pod uwagę tę kolejność skrótowe oznaczenia dla tych metod używane w tym badaniu wyglądają następująco: BU, TD, OLS, MinT. Prognozy niepoddane uzgadnianiu są nazywane prognozami bazowymi (BASE), a te otrzymane w takim procesie są określane jako prognozy zgodne. Występuje także rozróżnienie prognoz bazowych na „BASE Train” i „BASE Test”, pierwsze oznaczenie oznacza trafność dla prognozy bazowej w oparciu o zbiór treningowy, natomiast drugie równa się wartości uzyskanej na podstawie zbioru testowego. Obecne w badaniu trzy poziomy hierarchii, idąc do najbardziej zagregowanego stopnia, są nazywane kolejno: poziom State (poziom stanu), poziom Division (poziom dywizji) i poziom USA (całego kraju). Dla oceny trafności prognoz punktowych wykorzystano miary MAPE, MASE

⁵¹ Tidyverse, <https://www.tidyverse.org> [dostęp: 29.07.2022]

i RMSE, natomiast dla przedziałowej CRPS. Jeśli na danym poziomie był więcej niż jeden szereg czasowy to oceną trafności prognozy jest średnia uzyskanych miar błędów dla tych szeregów.

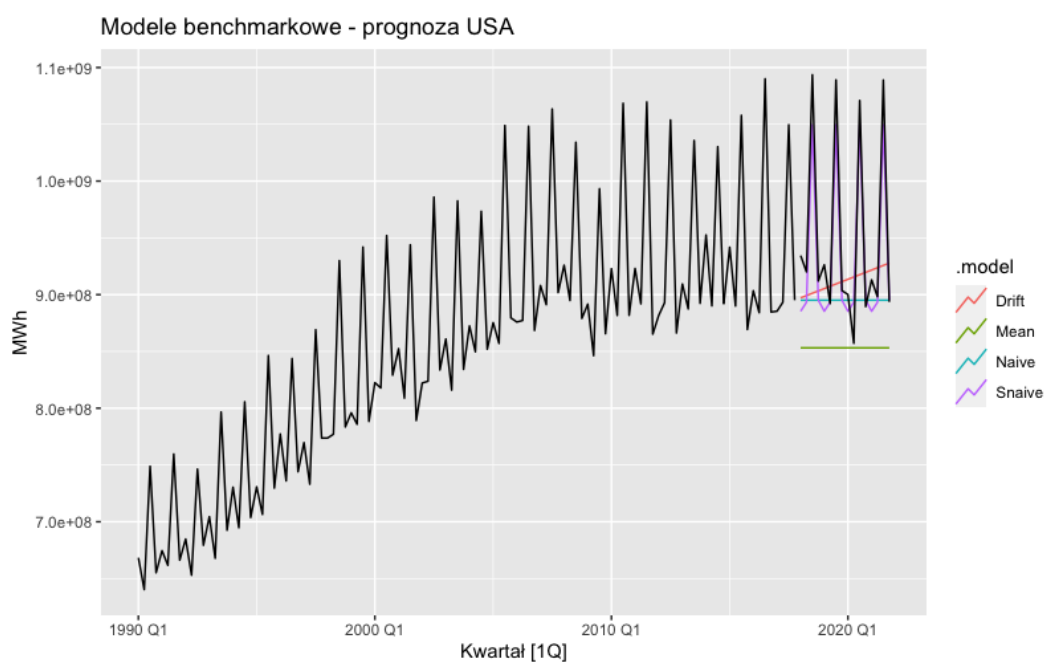
Opierając się na tak zbudowanej prognozie zostało przeprowadzonych sześć eksperymentów:

- Eksperyment 1 – pierwszy z eksperymentów miał na celu sprawdzenie na ile poszczególne podejścia ETS, ARIMA, Prophet są lepsze od metody benchmarkingowej. W ten sposób zostało określone czy dane metody ze wspomnianych rodzin modeli radzą sobie ze zbiorem danych. Przeprowadzone to zostało na najwyższym poziomie hierarchii.
- Eksperyment 2 – drugi z eksperymentów miał również na celu sprawdzenie na ile poszczególne metody radzą sobie ze zbiorem danych (tak jak pierwszy z eksperymentów), ale biorąc pod uwagę najniższy poziom hierarchii. Porównano średni błąd prognozy uzyskany dla wszystkich szeregów czasowych obecnych na tym stopniu struktury.
- Eksperyment 3 – trzeci z eksperymentów miał na celu sprawdzenie czy wybrane metody uzgadniania korzystnie wpływają na otrzymaną prognozę bazową uzyskaną przez poszczególne metody z rozważanych rodzin modeli bazowych.
- Eksperyment 4 – celem czwartego z eksperymentów było porównanie wykorzystanych metod prognozowania wraz z uzgadnianiem, jeśli chodzi o uzyskane wyniki prognozy punktowej.
- Eksperyment 5 – w piątym eksperymencie również dokonywano porównania wykorzystanych metod prognozowania wraz z uzgadnianiem, aczkolwiek w tym wypadku brano pod uwagę wyniki prognozy przedziałowej.
- Eksperyment 6 – w ostatnim eksperymencie został porównany czas budowy prognozy pomiędzy użytymi metodami prognozowania oraz sprawdzono czy różnica pomiędzy otrzymaniem prognoz bazowej, a uzgodnionej jest istotna.

Metoda benchmarkingowa

W eksperymentach jako punkt odniesienia zastosowano wybraną wcześniej metodę benchmarkingową. Została ona wytypowana spośród 4 podstawowych i prostych metod prognozowania: średniej, naiwnej, sezonowej naiwnej i dryfu (są one w dalszej części oznaczane kolejno jako: Mean, Naive, Snaive i Drift). W tym celu stworzono prognozy w

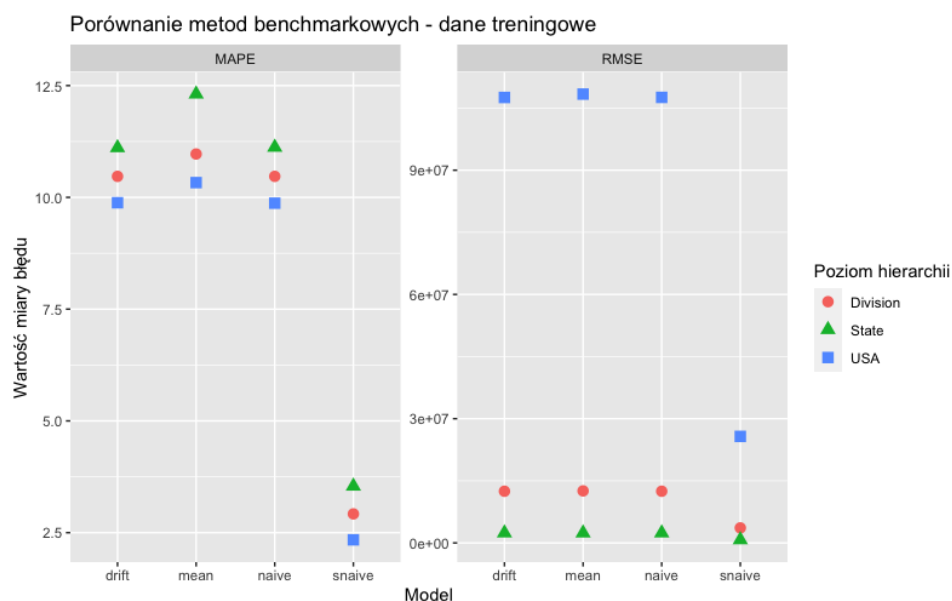
oparciu o każdą z nich, co zwizualizowano na rysunku 16. Opierając się wyłącznie na tym wykresie można od razu stwierdzić, że najbardziej odpowiednia jest metoda sezonowa naiwna ze względu na obecność sezonowości w badanych danych.



Rysunek 16 Metody benchmarkingowe - prognoza

Źródło: Opracowanie własne

Dla potwierdzenia tego wniosku zaprezentowano uzyskane miary błędów MAPE i RMSE dla danych treningowych (rysunek 17 i tabela 1) oraz dla danych testowych (rysunek 18 i tabela 2).



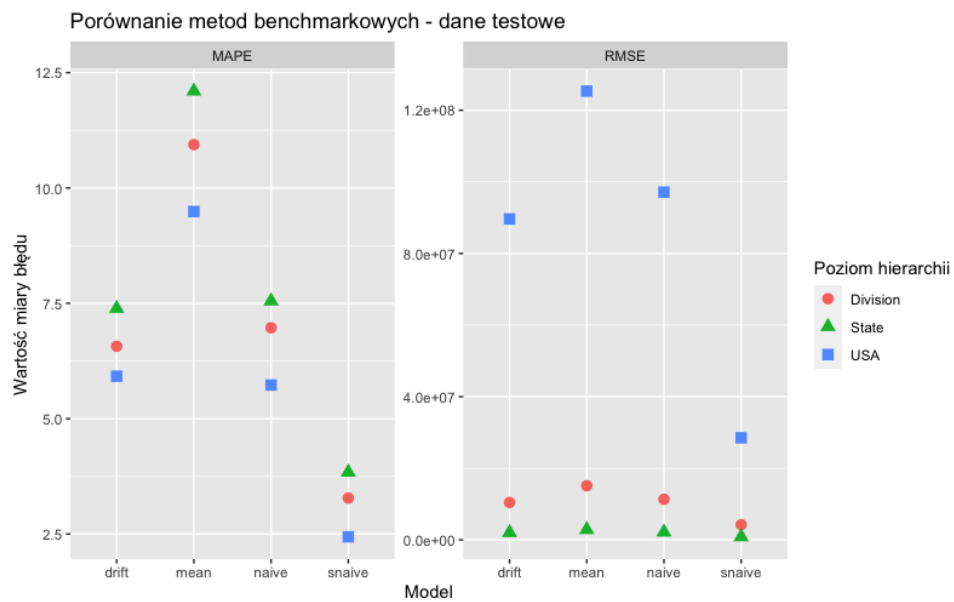
Rysunek 17 Metody benchmarkingowe - porównanie dokładności na bazie danych treningowych

Źródło: Opracowanie własne

Tabela 1 Metody benchmarkingowe – MAPE i RMSE dla danych treningowych

	MAPE			RMSE		
	USA	Division	State	USA	Division	State
Drift	9,88	10,47	11,11	107562612,40	12482468,16	2400528,30
Mean	10,33	10,97	12,31	108377377,10	12566218,66	2389259,43
Naive	9,87	10,47	11,12	107582016,80	12484857,49	2400991,39
Snaive	2,34	2,92	3,54	25724826,84	3628692,03	750025,63

Źródło: Opracowanie własne



Rysunek 18 Metody benchmarkingowe - porównanie dokładności na bazie danych testowych

Źródło: Opracowanie własne

Tabela 2 Metody benchmarkingowe – MAPE i RMSE dla danych testowych

	MAPE			RMSE		
	USA	Division	State	USA	Division	State
Drift	5,92	6,57	7,39	89632994,22	10445071,74	2026659,98
Mean	9,49	10,94	12,10	125335765,80	15127812,19	2884824,01
Naive	5,73	6,97	7,55	97127237,93	11330588,36	2169730,43
Snaive	2,44	3,28	3,84	28506621,69	4276157,15	860306,96

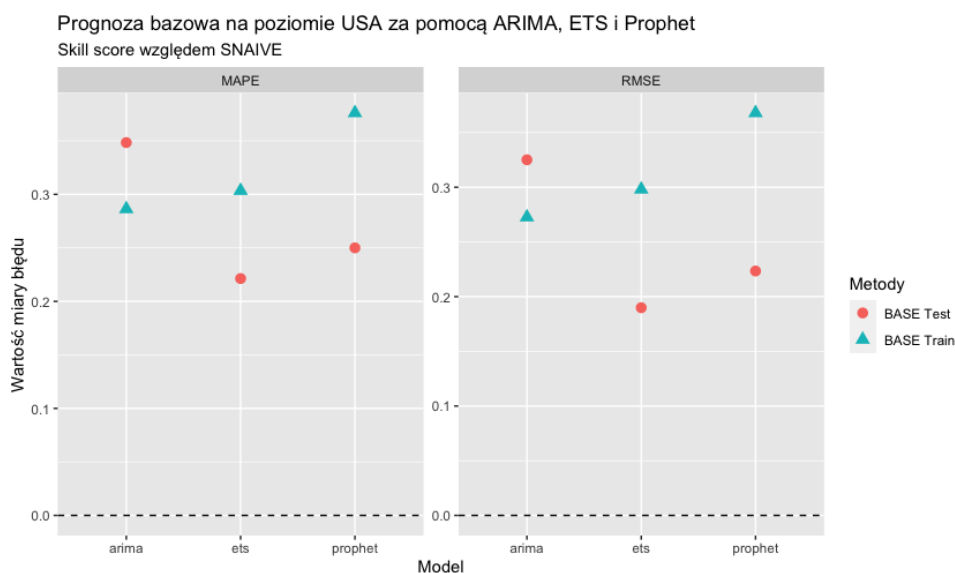
Źródło: Opracowanie własne

Dla każdego z rozważanych poziomów struktury hierarchicznej najlepsze wyniki trafności zarówno dla MAPE jak i RMSE uzyskała metoda Snaive, dlatego to właśnie ona będzie stosowana jako benchmark w dalszych etapach pracy. Z tego względu również skill

score dla poszczególnych metod w dalszej części opracowania będzie obliczany w stosunku do wybranego w tym podrozdziale podejścia.

3.3.1. Eksperyment 1 – ocena trafności prognozowania metod bazowych na najwyższym poziomie hierarchii

Pierwszy eksperyment obecny w badaniu został wykonany w celu określenia w jakim stopniu wybrane rodziny modeli radzą sobie lepiej w porównaniu do metody benchmarkingowej. By zrealizować to zadanie obliczono miary błędów MAPE i RMSE dla uzyskanych prognoz bazowych otrzymanych dzięki metodom ETS, ARIMA i Prophet. Rezultaty przedstawione na rysunku 19 to skill score poszczególnych miar względem SNAIVE, natomiast w tabeli 3 zamieszczono oryginalne wartości.



Rysunek 19 Porównanie trafności prognozy modeli bazowych na poziomie USA

Źródło: Opracowanie własne

Jak widać na rysunku 19 prognozy wszystkich modeli bazowych mają większą trafność niż wartości uzyskane przy pomocy metody benchmarkingowej. Skill score wskazuje, że wspomniane wartości są lepsze średnio o około 30% w porównaniu do obranego punktu odniesienia. Ciekawy jest fakt, że ARIMA w próbie testowej osiągnęła lepsze wyniki niż w uczącej.

Tabela 3 Modele bazowe i SNAIVE – poziom USA

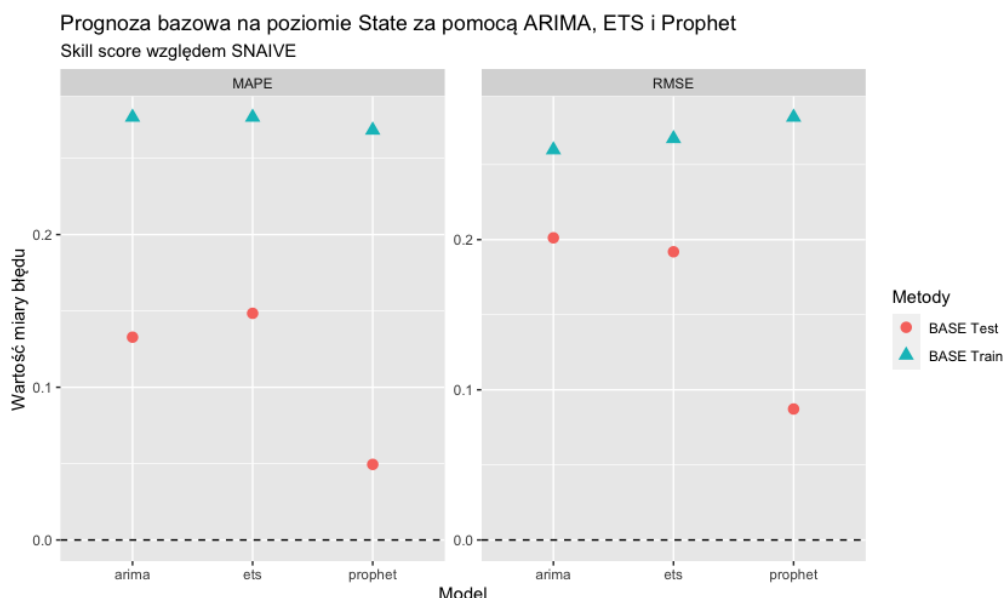
	MAPE		RMSE	
	BASE Train	BASE Test	BASE Train	BASE Test
Snaive	2,34	2,44	25724826,84	28506621,69
ETS	1,63	1,90	18056559,08	23093877,63
ARIMA	1,67	1,59	18710699,40	19238836,13
Prophet	1,46	1,83	16259447,26	22137613,74

Źródło: Opracowanie własne

Bazując na otrzymanych wynikach w tym eksperymencie można stwierdzić, że poszczególne metody bazowe dobrze radzą sobie z badanym zbiorem danych, gdyż miary MAPE dla każdej z nich osiągnęły wartości mniejsze, niż 2% (wedle literatury jest to bardzo dokładna prognoza)⁵², a także wyniki w ich przypadku wskazują na lepszą trafność, niż w metodzie benchmarkingowej.

3.3.2. Eksperyment 2 – ocena trafności prognozowania metod bazowych na najniższym poziomie hierarchii

W kolejnym eksperymencie dokonano porównania metod bazowych pod względem trafności na poziomie pojedynczych szeregów czasowych. Zastosowane miary oceny trafności i prezentacja wyników są identyczne jak poprzednio, różnicą jest, że w tym przypadku użyto średnich wartości miar błędu prognozy, gdyż na tym poziomie jest ich więcej, niż jeden.



Rysunek 20 Porównanie trafności modeli bazowych na poziomie State

Źródło: Opracowanie własne

⁵² Skoczypiec S., *Metody prognozowania: Jakość prognoz*, 2012, http://m6.pk.edu.pl/materialy/mp/MP_02_bledy_prognozy.pdf [dostęp: 16.08.2022].

Analogicznie jak w eksperymencie 1 w tym wypadku także wszystkie modele bazowe osiągnęły lepsze wyniki, niż metoda benchmarkingowa, aczkolwiek warto zauważyć, że procentowo nie była to tak samo duża różnica w jakości. Tym razem było to około 10-20%, spadek widać zwłaszcza w przypadku trafności w próbie testowej dla Prophet.

Tabela 4 Modele bazowe i SNAIVE – poziom State

	MAPE		RMSE	
	BASE Train	BASE Test	BASE Train	BASE Test
Snaive	3,54	3,84	750025,63	860306,96
ETS	2,56	3,27	549577,78	695203,02
ARIMA	2,56	3,33	555190,97	687205,79
Prophet	2,59	3,65	538895,53	785268,62

Źródło: Opracowanie własne

W porównaniu do otrzymanych wyników trafności na poziomie USA wyniki osiągnięte na najniższym poziomie hierarchii dały gorsze rezultaty. Mimo wszystko, wciąż można uznać, że wszystkie rodziny modeli także na tym stopniu poradziły sobie z prognozowaniem wartości badanych danych. Biorąc pod uwagę miarę błędu MAPE (tabela 4) to wedle literatury otrzymane prognozy można uznać za dokładne.⁵³

3.3.3. Eksperyment 3 – zastosowanie metod uzgadniania

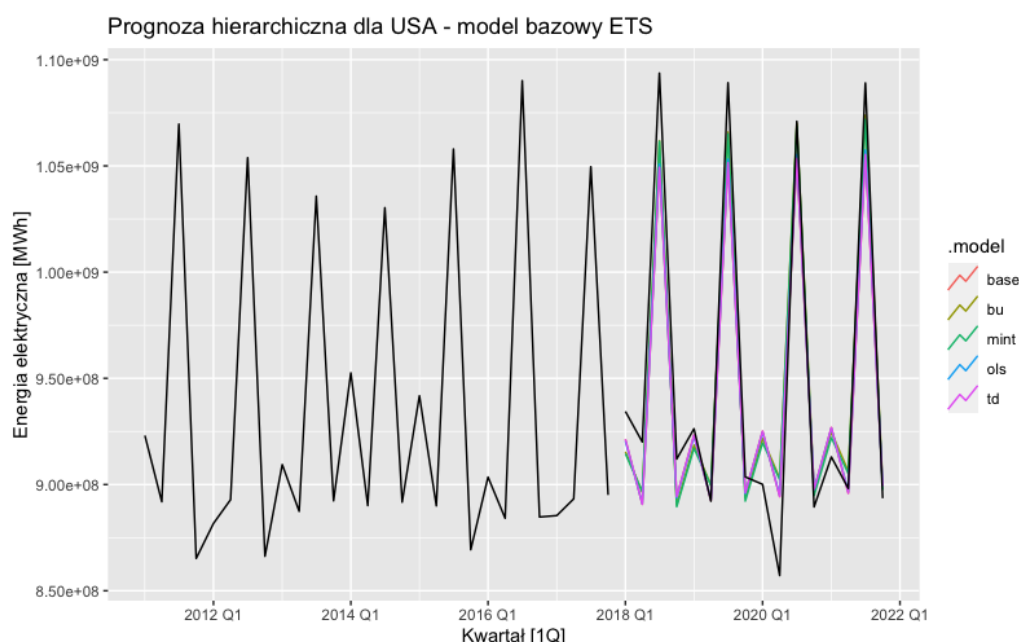
Wiedząc na podstawie poprzednich eksperymentów, że modele bazowe dobrze radzą sobie z prognozowaniem wykorzystywanego zbioru danych, celem kolejnego eksperymentu było ustalenie czy zastosowanie podejścia hierarchicznego, a tym samym wybranych metod uzgadniania pozwoli zwiększyć trafność prognozy bazowej dla poszczególnych rodzin modeli. Począwszy od tego eksperymenty skupiono się już na całej hierarchii.

ETS

Na rysunku 21 przedstawiono wykres obrazujący otrzymane wyniki prognoz bazowych oraz zgodnych zbudowanych w oparciu o ETS dla najwyższego poziomu hierarchii, czyli całego USA. Dla lepszej czytelności pokazano szereg czasowy od roku 2011. Pierwsza rzecz jaką można zauważyć to bardzo zbliżone wartości dla wszystkich metod, które w dużym stopniu pokrywają się lub są bliskie z danymi rzeczywistymi oznaczonymi czarnym kolorem.

⁵³ Skoczypiec S., *Metody prognozowania: Jakość prognoz*, 2012, http://m6.pk.edu.pl/materialy/mp/MP_02_bledy_prognozy.pdf [dostęp: 16.08.2022].

Najwyraźniejsza różnica między prognozowanymi liczbami, a rzeczywistymi występuje w drugim kwartale 2020. Wystąpił wtedy znaczny spadek zużycia co ma najprawdopodobniej związek z początkiem pandemii COVID-19, której skutkiem były ograniczenia dla społeczeństwa w pewnych dziedzinach (tzw. „lockdown”) w USA.⁵⁴



Rysunek 21 Wizualizacja prognozy dla modelu bazowego ETS

Źródło: Opracowanie własne

Dokładne wyniki z podziałem na wszystkie poziomy struktury i metody przedstawiono w tabeli 5. Skupiono się na ocenie tylko prognozy punktowej, do czego wykorzystano miary trafności MAPE, MASE i RMSE.

Tabela 5 Model bazowy ETS - MAPE, MASE, RMSE

	MAPE			MASE			RMSE		
	USA	Division	State	USA	Division	State	USA	Division	State
BASE Train	1,63	2,08	2,56	0,7	0,72	0,72	18056559,08	2642956,66	549577,78
BASE Test	1,90	2,64	3,27	0,92	0,96	1,04	23093877,63	3191532,28	695203,02
BU	1,73	2,32	3,27	0,81	0,87	1,04	19757221,92	2920672,31	695203,02
MinT	1,71	2,45	3,25	0,80	0,92	1,03	19723009,33	3068657,01	692600,53
OLS	1,87	2,78	4,11	0,90	1,04	1,34	22547010,43	3345437,50	757529,19
TD	1,90	2,73	3,53	0,92	1,01	1,12	23093877,63	3434375,44	759099,07

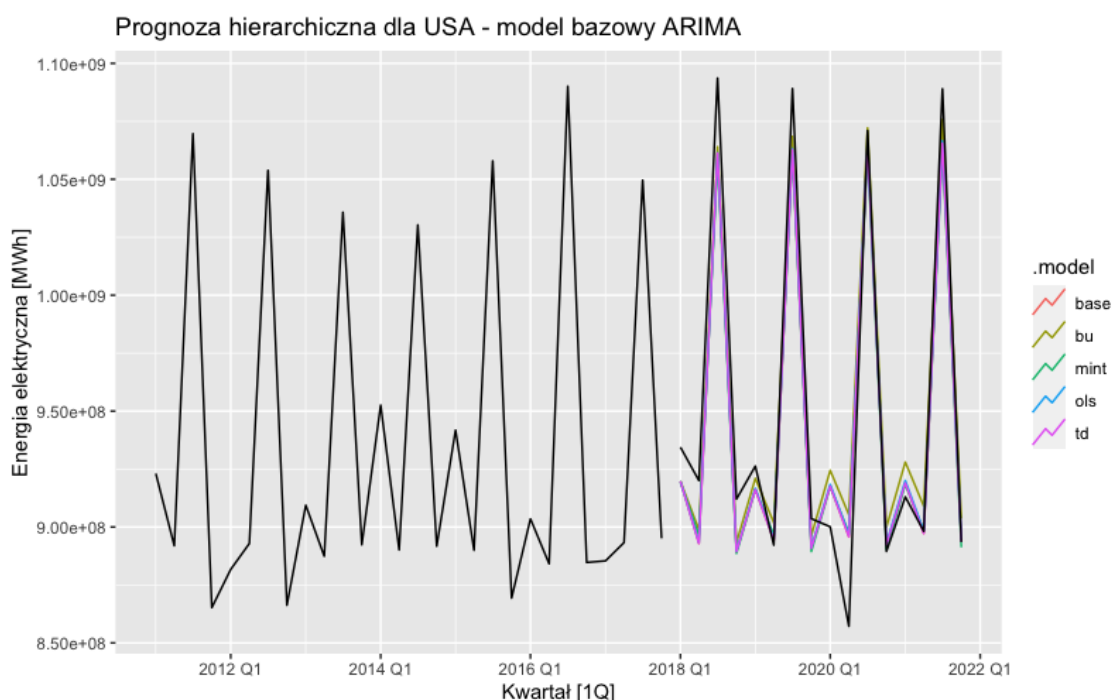
Źródło: Opracowanie własne

⁵⁴ Li Z, Ye H, Liao N, Wang R, Qiu Y, Wang Y, *Impact of COVID-19 on electricity energy consumption: A quantitative analysis on electricity*, International Journal of Electrical Power & Energy Systems, 2022

Najwyższe wyniki trafności uzyskała prognoza bazowa dla danych treningowych, aczkolwiek w porównaniu do niej prognoza bazowa dla danych testowych otrzymała relatywnie nieznacznie gorsze wyniki. Przechodząc do metod uzgadniania widać, że ich zastosowanie miało tutaj sens na poziomie USA, gdyż w większości przypadków poprawiło wyniki prognozy bazowej, z kolei dla Division poprawę było widać tylko dla BU i MinT. Inaczej sytuacja wyglądała w przypadku najniższego poziomu, tj. State, różnica pomiędzy trafnością w próbce treningowej, a testowej i uzgodnionymi prognozami była bardziej wyraźna. W tym wypadku uzgadnianie prognoz poprawiło minimalnie wyniki tylko przy metodzie MinT.

ARIMA

Tak jak widać na rysunku 22 prognozy otrzymane przy pomocy modelu ARIMA są bardzo podobne do tych uzyskanych z pomocą rodziny ETS. Tutaj również na prawie całym horyzoncie prognozy otrzymane wartości pokrywają się lub są zbliżone do danych rzeczywistych, największa różnica także występuje przy spadku w drugim kwartale 2020.



Rysunek 22 Wizualizacja prognozy dla modelu bazowego ARIMA

Źródło: Opracowanie własne

Analizując wyniki miar trafności dla modelu ARIMA pokazane w tabeli 6 od razu można zauważyć interesujący fakt, że na poziomie USA prognozy bazowe w próbie testowej dla MAPE uzyskały lepsze wyniki, niż wartości dopasowane. Jednak niestety żadna z metod uzgadniania nie poprawiła tych wartości pod kątem trafności (TD ma taką samą wartość, ponieważ w tym podejściu otrzymane wartości dla poziomu najwyższego są zawsze równe prognozie bazowej). Na kolejnym poziomie, czyli Division, sytuacja jest odwrotna i próbka testowa ma mniejszą trafność oraz uzgadnianie prognoz poprawia jej wynik. Wyniki poziomu State zachowują się analogicznie jak przy ETS.

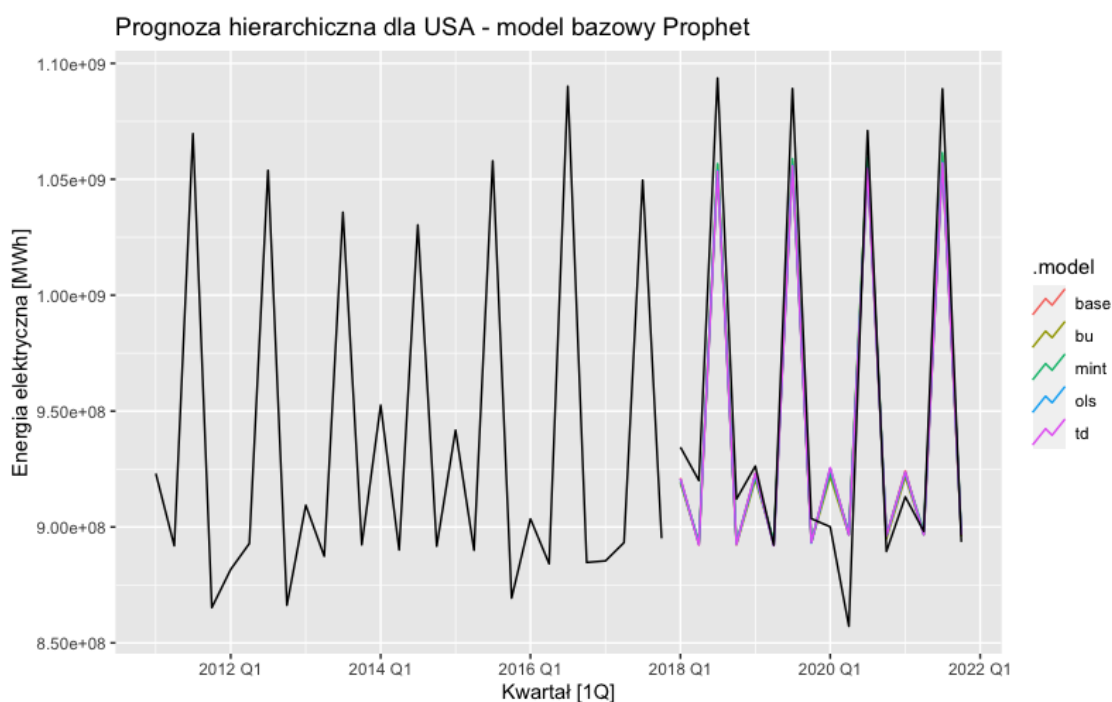
Tabela 6 Model bazowy ARIMA - MAPE, MASE, RMSE

	MAPE			MASE			RMSE		
	USA	Division	State	USA	Division	State	USA	Division	State
BASE Train	1,67	2,11	2,56	0,73	0,73	0,72	18710699,40	2665800,12	555190,97
BASE Test	1,59	2,90	3,33	0,76	1,11	1,06	19238836,13	3432738,94	687205,79
BU	1,74	2,26	3,33	0,81	0,84	1,06	19575605,81	2825530,93	687205,79
MinT	1,67	2,50	3,47	0,80	0,95	1,10	20142848,91	3163539,74	734370,60
OLS	1,60	2,82	5,23	0,76	1,07	1,85	19185215,48	3300917,26	741196,94
TD	1,59	2,87	3,61	0,76	1,10	1,15	19238836,13	3422399,44	756839,49

Źródło: Opracowanie własne

Prophet

W ostatnim modelu bazowym uzyskana prognoza wizualnie osiąga podobne wartości jak w dwóch poprzednich przypadkach. Jej wyniki również są zbliżone do wartości rzeczywistych, a największa różnica między nimi także występuje w drugim kwartale 2020. Natomiast można zauważyć (rysunek 23), że porównaniu do pozostałych metod wyniki uzyskane za pomocą uzgadniania mniej się różnią od prognozy bazowej (poszczególne linie na wykresie niemal się pokrywają).



Rysunek 23 Wizualizacja prognozy dla modelu bazowego Prophet

Źródło: Opracowanie własne

Poprzednie przypuszczenie potwierdzają także wyniki zaprezentowane w tabeli 7. Różnice pomiędzy uzyskanymi miarami błędów są mniejsze niż poprzednio. W tym przypadku występuje typowa sytuacja i największa trafność jest w obrębie próbki uczącej. Następnie następuje spadek trafności dla próbki testowej. Zastosowanie uzgadniania daje tutaj pozytywne rezultaty przynajmniej w jednym z podejść dla każdego poziomu, dla USA następuje to w każdej z metod uzgadniania, dla Division i State jest to osiągnięte wyłącznie dzięki MinT.

Tabela 7 Model bazowy Prophet - MAPE, MASE, RMSE

	MAPE			MASE			RMSE		
	USA	Division	State	USA	Division	State	USA	Division	State
BASE Train	1,46	2,10	2,59	0,62	0,74	0,73	16259447,26	2555169,93	538895,53
BASE Test	1,83	2,93	3,65	0,88	1,13	1,17	22137613,74	3554554,53	785268,62
BU	1,73	2,96	3,65	0,83	1,13	1,17	20900541,53	3491340,92	785060,66
MinT	1,73	2,86	3,56	0,83	1,09	1,14	20789909,44	3415433,54	769817,64
OLS	1,81	3,04	3,97	0,87	1,17	1,29	21985996,85	3631917,05	808355,59
TD	1,82	3,00	3,76	0,88	1,16	1,21	22130762,28	3636864,10	815788,49

Źródło: Opracowanie własne

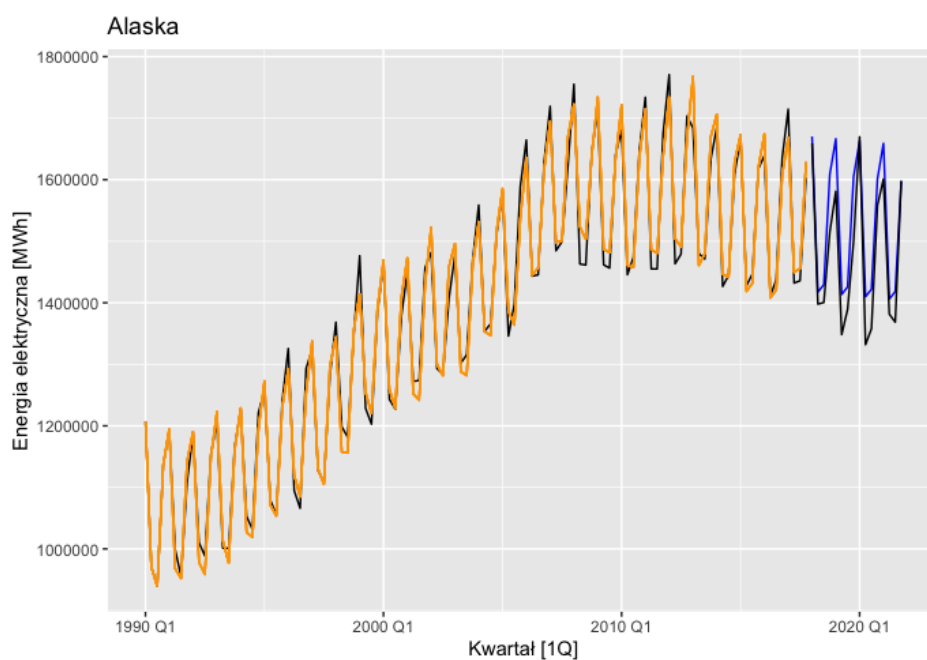
Podsumowanie eksperymentu 3

Ogólnie można stwierdzić, że kombinacje złożone z metod pochodzących z rodzin modeli ETS, ARIMA i Prophet wraz z wybranymi podejściami uzgadniania zostały odpowiednio dobrane do badanego zjawiska w niniejszej pracy. Podobnie jak w poprzednich eksperymentach, stosując zaproponowane w literaturze kryteria dopuszczalności prognoz oparte o miarę MAPE stwierdzono, że uzyskana prognoza przez te kombinacje jest bardzo dokładna dla poziomów USA i Division (błąd jest mniejszy, niż 3%) oraz dokładna dla State (błąd mniejszy niż 5%).⁵⁵

W większości przypadkach największą trafność uzyskiwano w zbiorze uczącym, następnie miara ta pogarszała się w próbie testowej, by następnie relatywnie nieznacznie osiągnąć w niej lepsze wyniki dzięki metodom uzgadniania. W ostatnim wymienionym aspekcie najlepsze poprawienie wyników uzyskiwano dla poziomów USA i Division. Dla poziomu State zdarzyło się to tylko dwa razy – dla ETS i Prophet dzięki pomocy MinT. Zatem w większości przypadków zastosowanie uzgadniania przyniosło pozytywny efekt.

W eksperymencie zauważono także inny interesujący fakt, który dotyczył kształtowania się trafności prognoz na poszczególnych poziomach hierarchii. Okazało się, że im niższy stopień hierarchii tym uzyskiwano wyższe średnie błędy dla poszczególnych prognoz. Na taki stan rzeczy ma najprawdopodobniej wpływ występowanie na niższych poziomach większej zmienności i ewentualnych anomalii, których zaprognozowanie jest bardzo trudne lub wręcz niemożliwe (np. niespodziewany spadek zużycia w czasie pandemii). Dodatkowo, ze względu na skalę, ich obecność w pojedynczych stanach w większym stopniu rzutuje na trafność prognozy, aniżeli gdyby brać pod uwagę cały kraj, gdzie błędy z pojedynczych jednostek są zagłuszane przez ogół struktury. Na rysunkach 24-26 przedstawiono szeregi o największych błędach MAPE dla próbki testowej w każdym z bazowych modeli, co w pewnym stopniu obrazuje skąd biorą się wspomniane wyższe błędy na tym poziomie struktury, a także może wyjaśniać różnicę w trafności między zbiorem uczącym, a testowym. Dane rzeczywiste są oznaczone kolorem czarnym, dopasowane pomarańczowym, a prognozowane niebieskim.

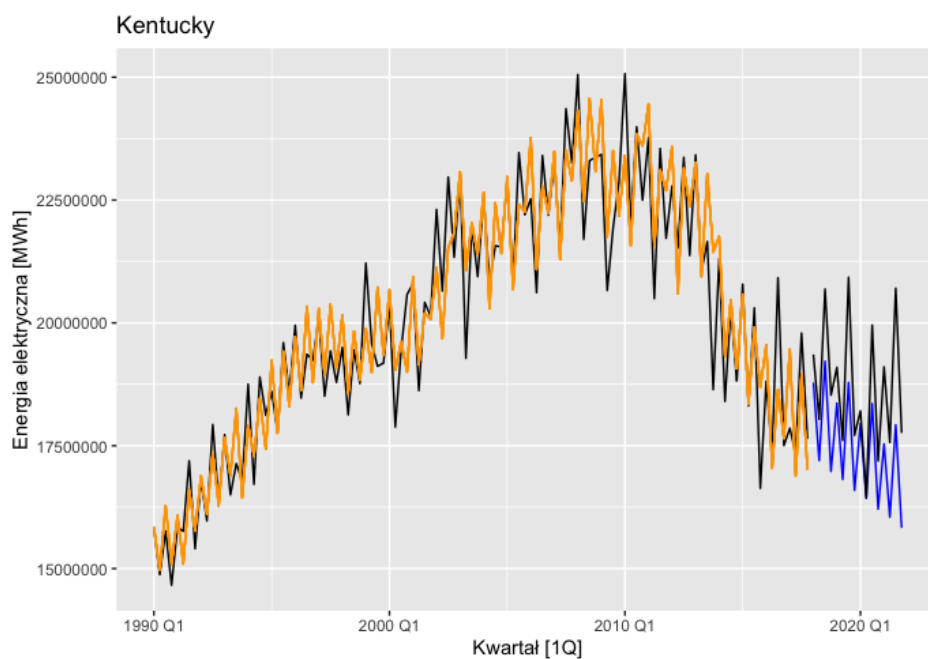
⁵⁵ Skoczypiec S., *Metody prognozowania: Jakość prognoz*, 2012, http://m6.pk.edu.pl/materialy/mp/MP_02_bledy_prognozy.pdf [dostęp: 16.08.2021].



Rysunek 24 ARIMA – szereg o największym błędzie MAPE dla próbki testowej

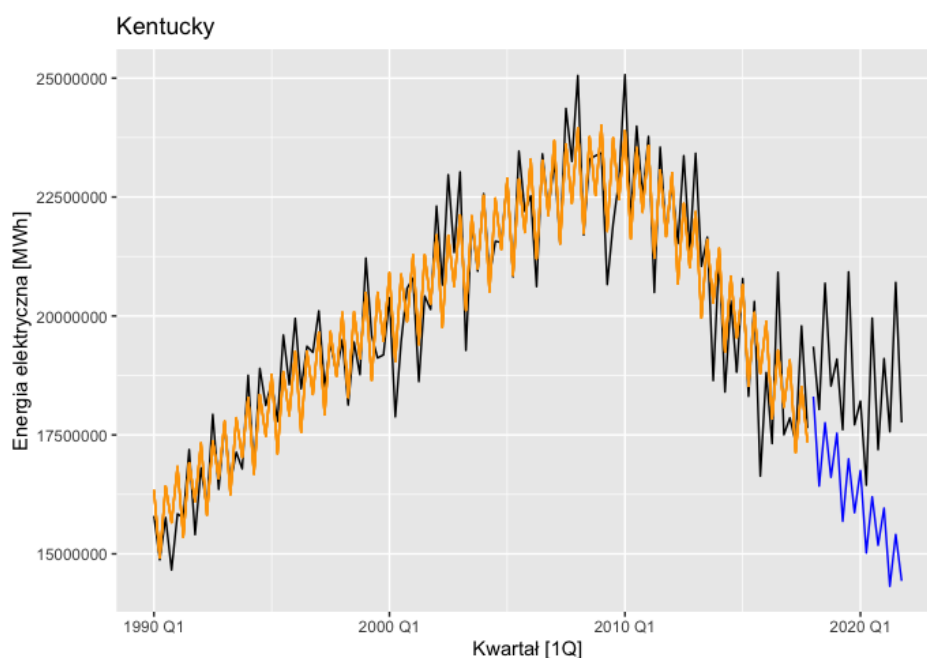
Źródło: Opracowanie własne

Największy błąd MAPE dla próbki testowej model bazowy ARIMA osiągnął dla stanu Alaska, wyniósł on 3,31, podczas gdy dla próbki uczącej miał wartość 1,48. Jak widać na rysunku 24 wynikało to ze zmniejszenia zużycia energii w ostatnich latach w tym miejscu.



Rysunek 25 ETS - szereg o największym błędzie MAPE dla próbki testowej

Źródło: Opracowanie własne



Rysunek 26 Prophet - szereg o największym błędzie MAPE dla próbki testowej

Źródło: Opracowanie własne

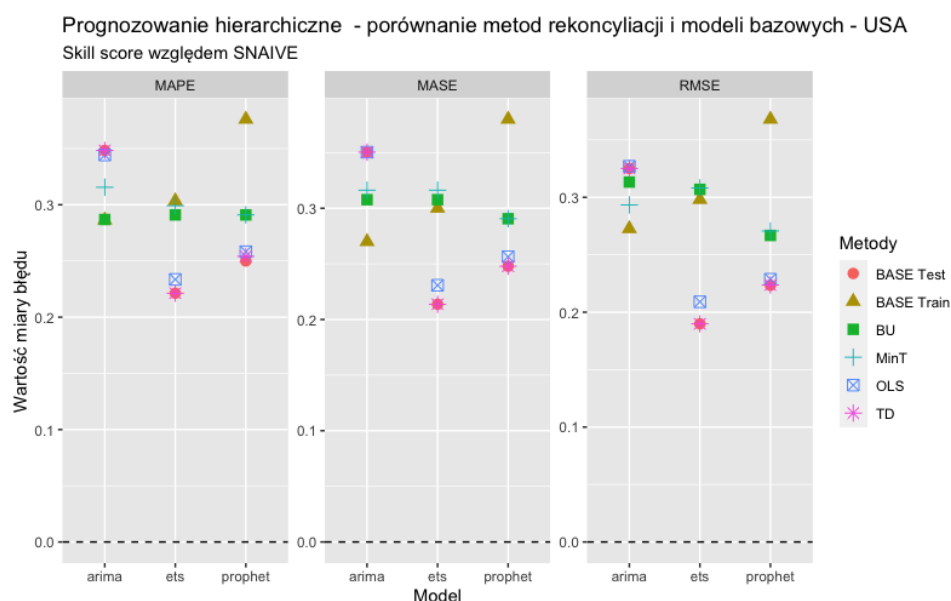
Największe błędy MAPE dla próbki testowej modeli opartych na ETS i Prophet uzyskały dla tego samego szeregu – stanu Kentucky. W pierwszym z nich wyniósł on 6,52 (w próbce uczącej 3,06), a w drugim aż 13,4 (w uczącej 2,99). Na rysunkach 25 i 26 można zauważyć, że nastąpiła w tym szeregu nagle zmiana trendu, jest negatywny i ustabilizował się w prognozowanym okresie, co wyjaśnia w pewnym stopniu otrzymany błąd.

3.3.4. Eksperyment 4 – porównanie metod pod względem prognozy punktowej

Biorąc pod uwagę wyniki i obserwacje z eksperymentu 3, gdzie stwierdzono, że dobrane metody uzgadniania zwiększają trafność prognozy bazowej i pozwalają odpowiednio prognozować badane zjawisko, kolejnym celem będzie wybór kombinacji metody bazowej i uzgadniania, która daje najlepsze wyniki spośród wszystkich rozważanych. Dla bardziej klarownego porównania zastosowano tu również metodę benchmarkingową jako punkt odniesienia, co pozwoliło łatwiej oceniać trafność doboru zastosowanych podejść.

Porównanie wyników uzyskanych dzięki pomocy poszczególnych metod zostało podzielone na trzy poziomy struktury hierarchicznej charakteryzujących badane dane (USA, Division, State) oraz na wszystkie szeregi składające się na tę strukturę wspólnie. Dla każdego z nich stworzono wizualizację, gdzie wykorzystano tzw. skill score dla MAPE, MASE i RMSE w odniesieniu do wybranej poprzednio metody benchmarkingowej. Dodatkowo zamieszczono również tabele z oryginalnymi wartościami tych trzech wspomnianych miar.

Poziom USA



Rysunek 27 Poziom USA - porównanie prognozy punktowej względem SNAIVE

Źródło: Opracowanie własne

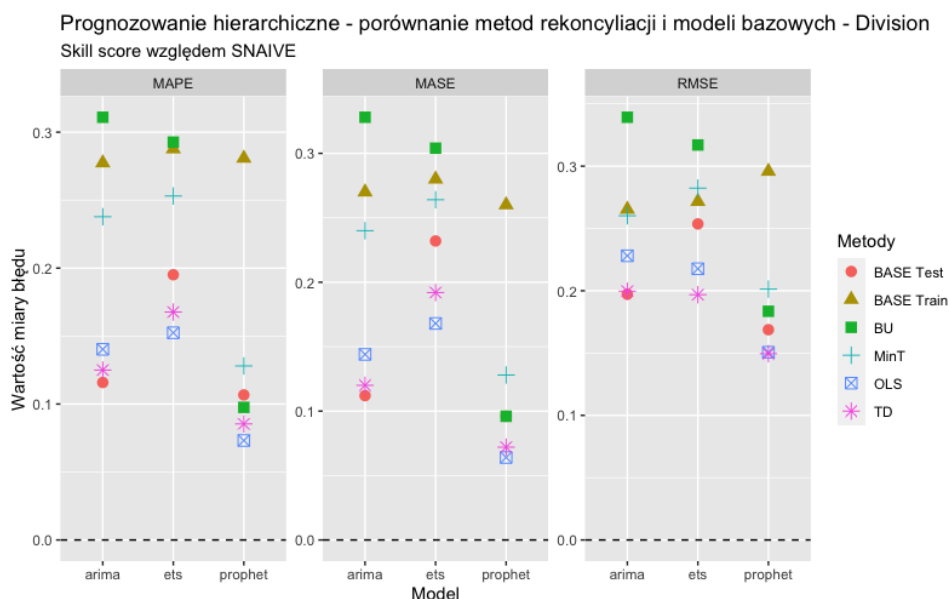
Odnosząc się do rysunku 27 i tabeli 8, dla poziomu USA spośród trzech bazowych rodzin modeli w tym badaniu średnio najlepsze wyniki prognozy punktowej uzyskała ARIMA, która też jako jedyna miała lepsze wyniki w próbce testowej, niż w uczącej. Jednak w jej wypadku zastosowanie uzgadniania nie polepszyło trafności prognozy. Na drugim miejscu plasuje się ETS za sprawą wyników osiągniętych dzięki BU i MinT, ostatni Prophet także uzyskał w tych podejściach najlepsze, lecz minimalnie gorsze od poprzednika prognozy. Co warto wspomnieć, to że wszystkie rodziny modeli bazowych uzyskały o wiele lepsze wyniki od swojej metody referencyjnej (niektóre nawet o 35% lepsze).

Tabela 8 Poziom USA - MAPE, MASE, RMSE

	MAPE			MASE			RMSE		
	ARIMA	ETS	Prophet	ARIMA	ETS	Prophet	ARIMA	ETS	Prophet
BASE Train	1,67	1,63	1,46	0,73	0,70	0,62	18056559,08	18710699,40	16259447,26
BASE Test	1,59	1,90	1,83	0,76	0,92	0,88	23093877,63	19238836,13	22137613,74
BU	1,74	1,73	1,73	0,81	0,81	0,83	19757221,92	19575605,81	20900541,53
MinT	1,67	1,71	1,73	0,80	0,80	0,83	19723009,33	20142848,91	20789909,44
OLS	1,60	1,87	1,81	0,76	0,90	0,87	22547010,43	19185215,48	21985996,85
TD	1,59	1,90	1,82	0,76	0,92	0,88	23093877,63	19238836,13	22130762,28

Źródło: Opracowanie własne

Poziom Division



Rysunek 28 Poziom Division - porównanie prognozy punktowej względem SNAIVE

Źródło: Opracowanie własne

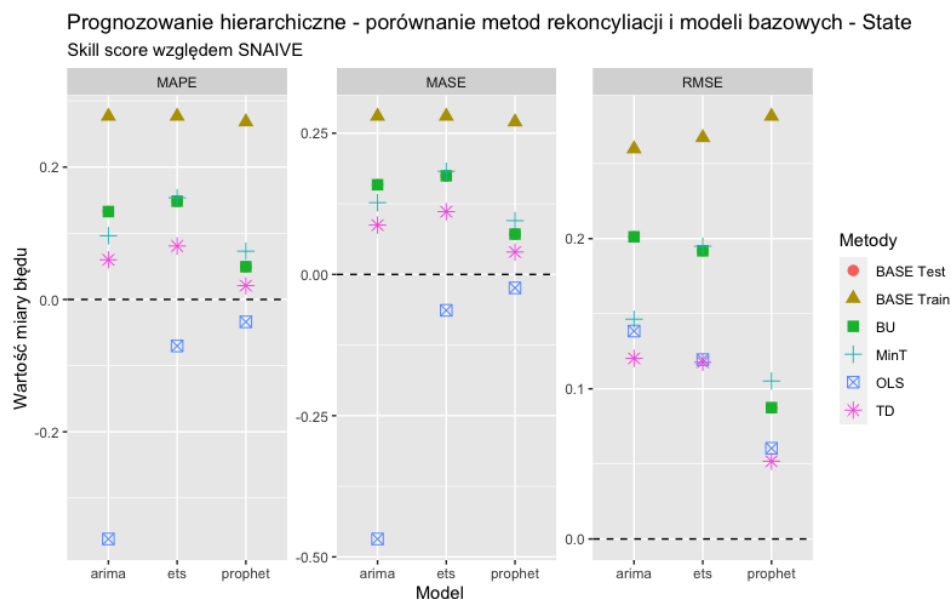
Na środkowym poziomie badanej struktury hierarchicznej również wszystkie zastosowane metody uzyskały trafniejsze wyniki od metody benchmarkingowej. Jak widać na rysunku 28 dla ARIMA i ETS bardzo dobre prognozy otrzymano w pierwszej kolejności dzięki wykorzystaniu metody uzgadniania BU (około 31% lepsze od referencyjnej dla każdej z miar), podejście MinT także dało bardzo dobre wyniki (około 27 % lepsze). Interesujący wydaje się natomiast fakt, że na tym stopniu hierarchii Prophet radzi sobie zauważalnie gorzej niż swoi konkurenci. Warto także wspomnieć, że w przypadku rodzin ARIMA i ETS zastosowanie uzgadniania daje relatywnie lepsze wyniki w porównaniu do prognozy bazowej, niż ma to miejsce dla modelu Prophet.

Tabela 9 Poziom Division - MAPE, MASE, RMSE

	MAPE			MASE			RMSE		
	ARIMA	ETS	Prophet	ARIMA	ETS	Prophet	ARIMA	ETS	Prophet
BASE Train	2,11	2,08	2,10	0,73	0,72	0,74	2665800,12	2642956,66	2555169,93
BASE Test	2,90	2,64	2,93	1,11	0,96	1,13	3432738,94	3191532,28	3554554,53
BU	2,26	2,32	2,96	0,84	0,87	1,13	2825530,93	2920672,31	3491340,92
MinT	2,50	2,45	2,86	0,95	0,92	1,09	3163539,74	3068657,01	3415433,54
OLS	2,82	2,78	3,04	1,07	1,04	1,17	3300917,26	3345437,50	3631917,05
TD	2,87	2,73	3,00	1,10	1,01	1,16	3422399,44	3434375,44	3636864,10

Źródło: Opracowanie własne

Poziom State



Rysunek 29 Poziom State - porównanie prognozy punktowej względem SNAIVE

Źródło: Opracowanie własne

Na poziomie State pierwszy raz występuje sytuacja, że jakakolwiek z metod uzyskała gorsze wyniki od metody referencyjnej. Co ciekawe, jest to metoda uzgadniania, a mianowicie OLS. W tym wypadku zastosowanie uzgadniania tym podejściem przyniosło odwrotny skutek od oczekiwanego, np. dla modelu bazowego ARIMA jej wartości są nawet prawie o 50% gorsze niż metoda referencyjna (błąd MASE, rysunek 29).

Mimo wszystko pozostałe prognozy bazowe i zgodne uzyskały lepszą trafność od punktu odniesienia Snaive. Najlepiej okazała się działać metoda uzgadniania MinT, dzięki której otrzymano korzystniejsze wartości prognoz dla zarówno dla modeli ETS jak i Prophet (należy pamiętać, że na tym poziomie BU równa się prognozie bazowej). W przypadku ARIMY uzyskano za pomocą tego podejścia nieznacznie gorsze wartości od prognozy bazowej. Tak jak poprzednio, w przypadku tego poziomu Prophet również uzyskał najgorsze wyniki dla próbki testowej.

Tabela 10 Poziom State - MAPE, MASE, RMSE

	MAPE			MASE			RMSE		
	ARIMA	ETS	Prophet	ARIMA	ETS	Prophet	ARIMA	ETS	Prophet
BASE Train	2,56	2,56	2,59	0,72	0,72	0,73	549577,78	555190,97	538895,53
BASE Test	3,27	3,33	3,65	1,04	1,06	1,17	695203,02	687205,79	785268,62
BU	3,27	3,33	3,65	1,04	1,06	1,17	695203,02	687205,79	785060,66
MinT	3,25	3,47	3,56	1,03	1,10	1,14	692600,53	734370,60	769817,64
OLS	4,11	5,23	3,97	1,34	1,85	1,29	757529,19	741196,94	808355,59
TD	3,53	3,61	3,76	1,12	1,15	1,21	759099,07	756839,49	815788,49

Źródło: Opracowanie własne

Wszystkie szeregi

Dla szerszej perspektywy stworzono również wizualizację (rysunek 30) i tabelę (tabela 11) obrazujące wyniki miar błędów biorąc pod uwagę wszystkie szeregi czasowe badanej struktury hierarchicznej razem.



Rysunek 30 Wszystkie szeregi - porównanie prognozy punktowej względem SNAIVE

Źródło: Opracowanie własne

Z racji budowy badanej struktury hierarchicznej, tj. szeregi poziomu najniższego State są od razu agregowane jedynie do kilku jednostek Division na następnym poziomie, ogólne wyniki trafności i zestawienie poszczególnych podejść względem siebie jest podobna do najniższego poziomu hierarchii. Można łatwo to zauważyć porównując rysunki 29 i 30, gdzie najgorsze miary posiadają OLS i TD, a najlepsze charakteryzują BU i MinT.

Porównując wyniki poszczególnych podejść prognozowania do metody benchmarkingowej, widać, że zdecydowanie najgorzej radziła sobie metoda OLS. Tylko w jej przypadku osiągnięto gorszy wyniki od wybranego punktu odniesienia. Pozostałe podejścia uzyskały wyniki powyżej tych wartości.

Tabela 11 Wszystkie szeregi - MAPE, MASE, RMSE

	MAPE			MASE			RMSE		
	ETS	ARIMA	Prophet	ETS	ARIMA	Prophet	ETS	ARIMA	Prophet
BASE Train	2,48	2,48	2,50	0,72	0,73	0,73	1145436,65	1164223,60	1094092,60
BASE Test	3,15	3,24	3,52	1,02	1,06	1,16	1430705,28	1396409,53	1545445,86
BU	3,10	3,15	3,52	1,01	1,02	1,16	1336043,06	1312342,28	1514147,47
MinT	3,11	3,29	3,42	1,01	1,07	1,13	1355140,16	1410944,38	1486497,22
OLS	3,88	4,82	3,80	1,29	1,72	1,26	1496556,34	1421221,55	1570062,51
TD	3,39	3,47	3,62	1,10	1,14	1,20	1519955,89	1453102,38	1580634,65

Źródło: Opracowanie własne

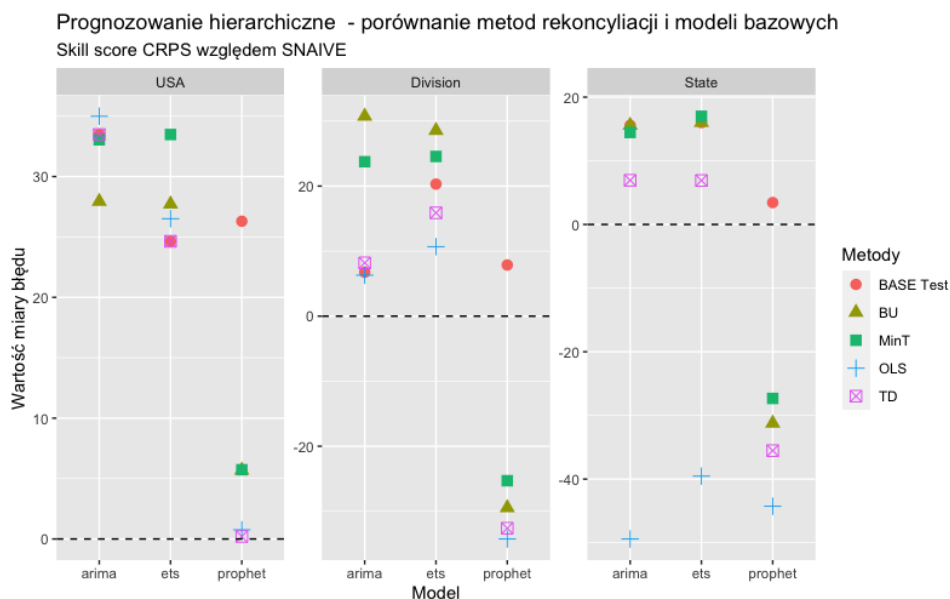
Podsumowanie eksperymentu 4

Celem drugiego eksperymentu było wybranie najlepiej spisującej się kombinacji modelu bazowego i metody uzgadniania w kontekście trafności prognozy punktowej. W zależności od poziomu struktury hierarchicznej uzyskiwano różnorodne wyniki, jednakże na podstawie otrzymanych rezultatów uznano, że najlepsze prognozy otrzymywano przy pomocy modelu bazowego ETS wraz z zastosowaniem uzgadniania BU lub MinT. Biorąc pod uwagę wszystkie szeregi struktury trafność dla tych kombinacji była największa. Bardzo dobre wyniki uzyskano, także dla ARIMY w sytuacji, gdzie uzgodniono jej wyniki za pomocą BU. Natomiast zauważalnie najgorzej w porównaniu do pozostałych poradził sobie z prognozowaniem Prophet, mimo, że w każdym z przypadków zastosowanie MinT poprawiło trafność jego prognozy bazowej.

3.3.5. Eksperyment 5 - porównanie metod pod względem prognozy przedziałowej

Cel eksperymentu 5 był taki sam jak w eksperymencie 4 z różnicą, że w tym wypadku badany był wynik prognozy przedziałowej zamiast punktowej. By porównać poszczególne podejścia w tej kwestii zastosowano skill score zbudowany na mierze błędu przedziałowego CRPS w odniesieniu do sezonowej metody naiwnej. Wyniki tej oceny prognozy zaprezentowano na rysunku 31 i w tabeli 12.

Poszczególne poziomy struktury



Rysunek 31 Porównanie prognozy przedziałowej względem SNAIVE – poszczególne poziomy struktury

Źródło: Opracowanie własne

Analizując rysunek 31 począwszy od poziomu USA, można zauważyć, że na tym stopniu struktury wszystkie metody uzyskały wyniki nie gorsze, niż metoda benchmarkingowa. Najlepsze prognozy otrzymano przy zastosowaniu modeli bazowych ARIMA i ETS, natomiast Prophet osiągnął relatywnie gorsze rezultaty. W przypadku tych dwóch pierwszych metody uzgadniania poprawiły wyniki prognoz bazowych.

Na poziomie Division sytuacja była podobna, aczkolwiek tutaj już nie wszystkie podejścia okazały się być trafniejsze niż benchmark. Dotyczy to prognoz zgodnych, gdzie modelem bazowym jest Prophet, gdyż niektóre są one nawet o 30% gorsze w porównaniu do punktu odniesienia. Z metod uzgadniania najlepiej sobie radzą na tym poziomie BU oraz MinT.

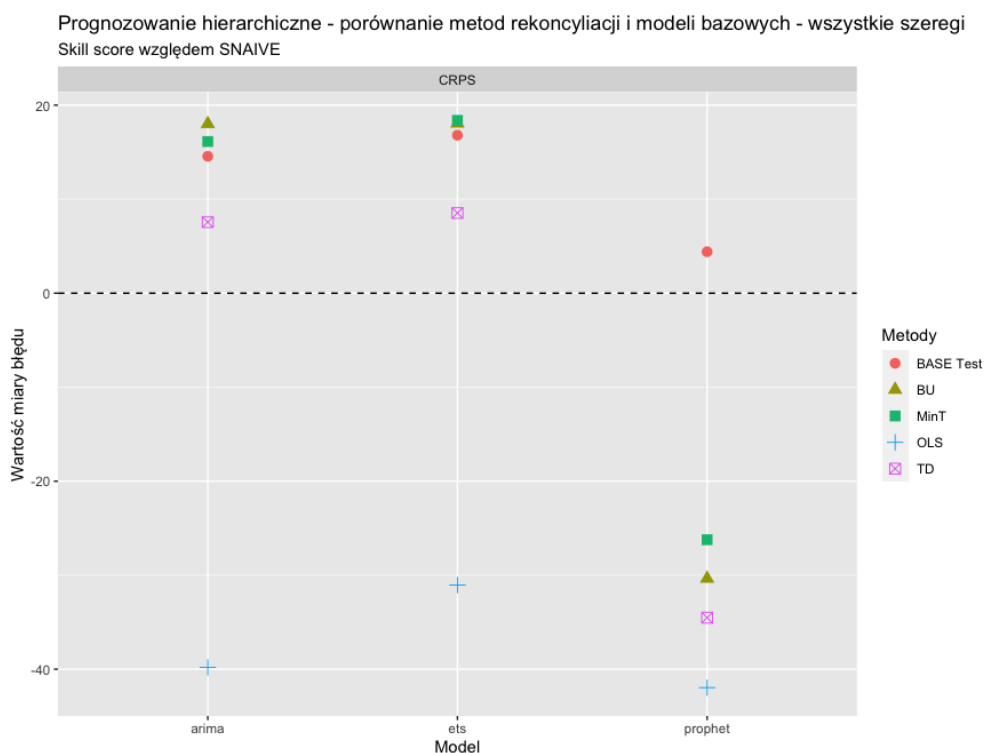
Na najniższym stopniu hierarchii do podejść z poprzedniego poziomu, które uzyskały gorsze wyniki od metody benchmarkingowej dołącza również metoda OLS, zarówno w modelu bazowym ARIMA, jak i ETS. Bazując na widocznych wynikach, można stwierdzić, że metoda uzgadniania OLS radzi sobie o wiele gorzej od innych podejść. Z kolei tak jak poprzednio najlepsze prognozy uzyskano dzięki BU (tutaj równa prognozie bazowej) i MinT.

Tabela 12 CRPS dla poszczególnych poziomów struktury

	USA			Division			State		
	ARIMA	ETS	Prophet	ARIMA	ETS	Prophet	ARIMA	ETS	Prophet
BASE Test	33,48	24,64	26,30	6,76	20,31	7,87	15,57	16,02	3,44
BU	27,94	27,72	5,69	30,74	28,55	-29,46	15,57	16,02	-31,23
MinT	33,04	33,48	5,73	23,75	24,56	-25,30	14,45	17,00	-27,33
OLS	34,99	26,51	0,76	6,30	10,69	-34,26	-49,41	-39,54	-44,27
TD	33,48	24,64	0,21	8,20	15,89	-32,61	6,94	6,91	-35,53

Źródło: Opracowanie własne

Wszystkie szeregi



Rysunek 32 Porównanie prognozy przedziałowej względem SNAIVE – wszystkie szeregi

Źródło: Opracowanie własne

Tak jak opisano w poprzednim eksperymencie ze względu na budowę hierarchii wykres przedstawiony na rysunku 32 jest podobny do poziomu State znajdującego się na rysunku 31. Jednakże dzięki wzięciu pod uwagę wyników z wyższych poziomów można zauważyć, że miary trafności prognoz poszczególnych metod mniej nakładają się na siebie, co sugeruje różne wartości błędów na poszczególnych stopniach struktury. Ostatecznie jednak, na tym rysunku także widać, że najlepiej poradziły sobie modele bazowe ARIMA i ETS wraz z metodami uzgadniania BU i MinT. Z kolei wartości modelu Prophet, a zwłaszcza prognozy zgodne otrzymane dzięki OLS uzyskiwały najgorsze rezultaty (OLS uzyskało najmniej trafne prognozy wśród wszystkich metod uzgadniania także dla pozostałych modeli bazowych).

Tabela 13 CRPS dla wszystkich szeregów struktury

	CRPS (skill score)		
	ETS	ARIMA	Prophet
BASE Test	16,80	14,56	4,41
BU	18,06	18,01	-30,36
MinT	18,38	16,13	-26,23
OLS	-31,05	-39,81	-41,96
TD	8,53	7,56	-34,52

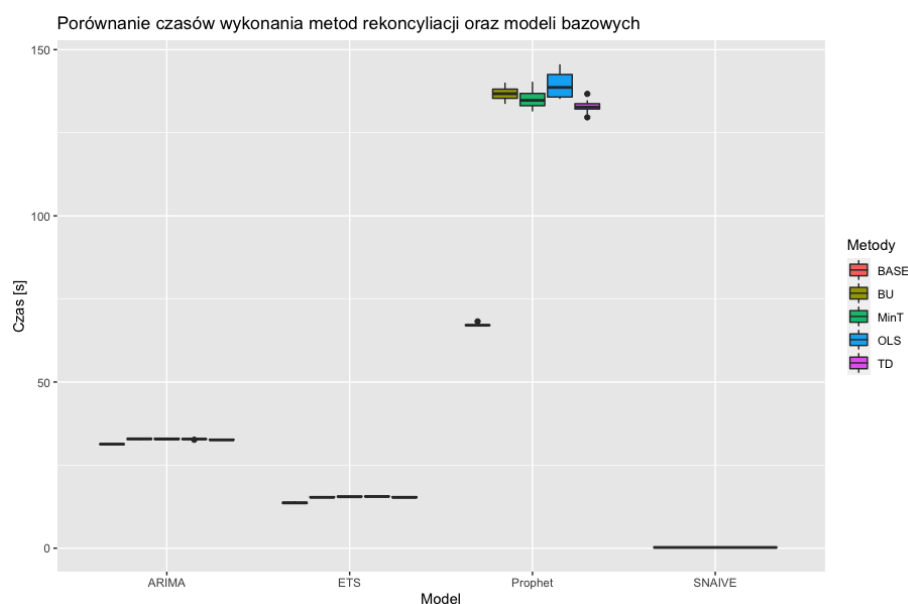
Źródło: Opracowanie własne

Podsumowanie eksperymentu 5

Podsumowując, wyniki eksperymentu 5 okazały się być zgodne z wynikami eksperymentu 4, tak jak w przypadku oceny prognozy punktowej, tutaj także kombinacjami, które najlepiej sobie poradziły z prognozowaniem były ETS i BU oraz ETS i MinT. Również tak jak wcześniej bardzo dobre wyniki uzyskano w przypadku ARIMY i tych samych metod uzgadniania co dla ETS, aczkolwiek trafność w tym wypadku była nieznacznie gorsza. Trzeci z modeli bazowych tj. Prophet znów osiągnął najgorsze wyniki w porównaniu do pozostałych (wszystkie prognozy zgodne oparte na tym modelu okazały się być gorsze niż benchmark). Ta sama sytuacja dotyczy metody uzgadniania OLS, zastosowanie jej nie przynosiło poprawy prognoz zgodnych, a nawet zmniejszały ich trafność.

3.3.6. Eksperyment 6 - porównanie metod pod względem czasu wykonywania

Eksperyment 6 miał na celu dwie główne intencje. Po pierwsze, również tak jak eksperymenty 4 i 5, jego zadaniem było porównanie poszczególnych podejść, ale biorąc pod uwagę inny czynnik – czas wykonywania obliczeń (czas uzyskania prognoz). Po drugie, miał pomóc dostrzec czy pomiędzy czasem wykonywania obliczeń dla modelu bazowego, a model wraz z uzgadnianiem występuje istotna różnica. Aby tego dokonać mierzono czas działania skryptu dla poszczególnych modeli bazowych i uzgadniania, na co pozwoliło użycie pakietu `tictoc()` z biblioteki R. W taki sposób dla każdego z podejść uzyskano 10-elementową próbkę czasu wykonywania (w sekundach), której średnią i odchylenie standardowe przedstawiono w tabeli 14. Wyniki te zwizualizowano także za pomocą wykresów pudełkowych na rysunku 33. Model SNAIVE został zamieszczony jedynie dla porównania.



Rysunek 33 Wykresy pudełkowe dla czasów wykonywania poszczególnych metod

Źródło: Opracowanie własne

Tabela 14 Średnie i odchylenia standardowe czasów wykonywania poszczególnych metod

Model bazowy	Metoda uzgadniania	Średnia [s]	Odchylenie Std. [s]
ARIMA	BASE	31,33	0,11
	BU	32,87	0,20
	MinT	32,92	0,23
	OLS	32,83	0,10
	TD	32,64	0,12
ETS	BASE	13,69	0,25
	BU	15,37	0,14
	MinT	15,56	0,13
	OLS	15,58	0,11
	TD	15,37	0,15
Prophet	BASE	67,21	0,37
	BU	136,77	2,29
	MinT	135,06	2,74
	OLS	139,26	3,80
	TD	132,88	2,01
SNAIVE	BASE	0,23	0,02

Źródło: Opracowanie własne

Pierwsze co można zauważyć analizując rysunek 33 to jednorodność próbek czasu wykonywania dla ARIMY i ETS. Poszczególne obserwacje wewnątrz ich zbiorów oscylowały wokół tych samych wartości, potwierdzają to również otrzymane dla nich odchylenia standardowe (tabela 14). Jedynym miejscem, gdzie uzyskiwane wartości były relatywnie bardziej różnorodne (w porównaniu do pozostałych metod) to prognozy oparte na modelu

Prophet. Podczas gdy dla poprzednich podejść odchylenia standardowe wynosiły zazwyczaj około 0,1 – 0,2 sekundy, to w tym wypadku były one o wiele wyższe, mieściły się w przedziale 2 – 4 sekund.

Natomiast skupiając się na średnim czasie wykonywania głównych modeli bazowych omawianych w tym badaniu wyraźnie było widać, że najlepiej poradziła sobie rodzina ETS, następnie ARIMA, a na samym końcu (z wyraźnym dystansem) umiejscowił się Prophet. Taka sama kolejność wśród modeli występuje również po zastosowaniu uzgadniania.

W celu zbadania czy dla rozważanych modeli bazowych, a ich prognozami otrzymanymi dzięki uzgadnianiu występuje istotna różnica wykorzystano statystyczny test Wilcoxona (znany również jako test Manna-Whitneya). Jest to nieparametryczna wersja testu t-Studenta, który można wykorzystać do porównania dwóch niezależnych próbek.⁵⁶ Hipotezy testu Wilcoxona są następujące:

H_0 : Średnie dla próbek są takie same.

H_1 : Średnie dla próbek nie są takie same.

Tabela 15 Wyniki testu Wilcoxona dla każdego modelu.

Model	p-value	Wynik
ARIMA	1,31E-06	odrzuć hipotezę zerową
ETS	1,31E-06	odrzuć hipotezę zerową
Prophet	1,31E-06	odrzuć hipotezę zerową

Źródło: Opracowanie własne

Jak widać w tabeli 15 we wszystkich przypadkach odrzucono hipotezę zerową, a zatem różnice w czasie pomiędzy otrzymaniem prognoz bazowych, a zgodnych dla każdego z omawianych modeli są istotne.

Zatem biorąc pod uwagę drugi z rozważanych aspektów w tym eksperymencie otrzymano wynik, że czas otrzymywania prognoz bazowych i zgodnych jest istotnie różny. Ciekawy jest natomiast fakt, że zastosowanie uzgadniania w Prophet znacznie wydłuża średni czas wykonywania skryptu (o około 70 sekund więcej niż dla prognoz bez uzgadniania),

⁵⁶ STHDA Statistical tools for high-throughput data analysis,
<http://www.sthda.com/english/wiki/unpaired-two-samples-wilcoxon-test-in-r> [dostęp: 26.08.2022]

podczas gdy dla rodzin ARIMA i ETS wartość ta była o wiele mniejsza (około 1,5 sekundy więcej).

Podsumowanie eksperymentu 6

Biorąc pod uwagę czas wykonywania najlepiej spisującą się metodą okazał się być model oparty na rodzinie ETS. Drugim w kolejności pod względem badanego czynnika była ARIMA, z kolei wyraźnie najgorsze wyniki osiągnął Prophet. Takie wyniki uzyskano zarówno dla czasu obliczeń dla prognoz bazowych jak i zgodnych tego modelu. Jeśli chodzi o różnice pomiędzy poszczególnymi podejściami uzgadniania wśród danego modelu bazowego to można uznać, że nie były one wyraźne. Znaczniejszy kontrast istniał pomiędzy czasem wykonywania samych prognoz bazowych, a otrzymaniem prognoz zgodnych. Potwierdziły to także przeprowadzone testy Wilcoxona dla poszczególnych rodzin modeli.

3.4. Wnioski

Głównym celem pracy było zbadanie czy wykorzystanie podejścia hierarchicznego w prognozowaniu zużycia energii elektrycznej pozwala na zwiększenie trafności budowanych prognoz. By ten zamysł osiągnąć zastosowano oraz porównano różne podejścia prognostyczne z wykorzystaniem wielu narzędzi i technik (głównie język R i powiązanie z nimi biblioteki) w oparciu o dane zużycia energii elektrycznej w USA. Biorąc pod uwagę wszystkie dokonane w niniejszej pracy eksperymenty i otrzymane wyniki można uznać, że powyższa teoria jest zasadna i zastosowanie badanej metody ma pozytywny wpływ na jakość prognozy, jednakże występują pewnie wyjątki w tym aspekcie dla wybranych modeli i metod.

Podstawowy argument przemawiający za faktem poprawy trafności prognoz można dostrzec w wynikach otrzymanych w eksperymentach 3-5, gdzie stwierdzono, że zawsze przynajmniej jeden z czterech rozważanych sposobów uzgadniania poprawiał wartości prognoz bazowych na każdym z poziomów hierarchii. Najlepszą procentową poprawę zaobserwowano na najwyższym poziomie, a idąc niżej w hierarchii była ona niższa, aczkolwiek wciąż otrzymane prognozy zgodne w większości były lepszej jakości niż bazowe. Warto także wspomnieć, że biorąc pod uwagę kryteria dopuszczalności prognoz oparte o miarę MAPE, uzyskano „bardzo dokładne” lub „dokładne” wyniki. Dodatkowo należy pamiętać, że oprócz samej poprawy tego aspektu, wykorzystanie prognozowania hierarchicznego niesie ze sobą zgodność wartości na poszczególnych poziomach hierarchii, co także jest argumentem przemawiającym za tym podejściem.

Wiedząc, że stosowanie prognozowania hierarchicznego jest generalnie korzystne w celu otrzymania lepszej trafności prognoz, pojawia się również pytanie jaka kombinacja modelu bazowego i metody uzgadniania zadziałała najlepiej w powyższym badaniu. W niniejszym opracowaniu okazało się, że najlepsze wyniki wśród modeli bazowych uzyskuje ETS, a w procesie uzgadniania MinT (co jest też zgodne z literaturą związaną z tym tematem⁵⁷). Taki sam wniosek można nawet przyjąć, gdyby brać te dwa podejścia oddzielnie (nie jako kombinacja). Poprzednie dwa stwierdzenia potwierdzają wyniki otrzymane w eksperymentach 4 i 5, zarówno w prognozie punktowej i przedziałowej spisywały się one najlepiej. ETS oraz MinT uzyskiwały najlepsze wyniki miar błędów w tych dwóch kategoriach w większości badanych poziomach struktury hierarchicznej, a biorąc pod uwagę wszystkie szeregi osiągnęły pierwszy rezultat. Również, jeśli wziąć pod uwagę czas wykonywania obliczeń to rodzina ETS spisała się najlepiej z grupy porównywanych modeli bazowych, a jeśli chodzi kombinacje wraz z uzgadnianiem to był to jeden z najlepszych wyników.

Interesującym faktem, który pojawił się w badaniu jest kwestia, że poza MinT bardzo dobre wyniki osiągnięto dla metody uzgadniania BU. Podejście to dobrze współpracowało zwłaszcza z modelami opartymi na rodzinie ETS i ARIMA, gdzie jej zastosowanie w prawie wszystkich przypadkach pozwalało otrzymać jedno z najlepszych prognoz zgodnych pod względem trafności. Jest to zaskakujące, gdyż BU to metoda, która w literaturze raczej jest niezalecana ze względu na to, że pomija dużo informacji obecnej w strukturze⁵⁸, to w tym wypadku jej kombinacja z ETS osiągnęła drugi najlepszy wynik spośród wszystkich. Opierając się na wspomnianych wynikach być może warto rozważyć niekiedy jej użycie, plusem w wielu zastosowaniach może być też jej prostota. W związku z tym można by było nawet pokusić się o stwierdzenie, że w wybranych przypadkach wystarczy prognozować jedynie najniższy poziom struktury hierarchicznej, a zadowalające wyniki dla wyższych poziomów również zostaną osiągnięte.

Tak jak wspomniano, na ogólnym poziomie prognozowanie hierarchiczne poprawiało trafność prognozy bazowej, jednak istniały przypadki, gdy zastosowanie jej przynosiło odwrotny skutek. Taka sytuacja występowała najczęściej dla podejść OLS i TD. Zarówno w

⁵⁷ George Athanasopoulos, Puwasala Gamakumara, Anastasios Panagiotelis, Rob J Hyndman, Mohamed Affan, Hierarchical forecasting. *Macroeconomic forecasting in the era of big data*, ed. P. Fuleky, Springer, 2020, Chapter 21, pp.689-719.

⁵⁸ Wickramasuriya, S. L., Athanasopoulos, G., & Hyndman, R. J., *Optimal forecast reconciliation for hierarchical and grouped time series through trace minimization*, *Journal of the American Statistical Association*, 2019, 114(526), 804–819.

prognozie punktowej jak i przedziałowej ich zastosowanie przyniosło gorszą trafność w prognozach zgodnych w porównaniu do bazowej, a niekiedy nawet metody benchmarkingowej. Bardzo źle wypadła zwłaszcza pierwsza z nich. Z kolei, jeśli chodzi o modele bazowe to najgorzej w niniejszym badaniu radził sobie Prophet, W jego przypadku uzyskiwano zazwyczaj najgorsze wyniki spośród trzech badanych rodzin, a także połączenie tego modelu z uzgadnianiem w wielu miejscach dało gorszą trafność niż brany pod uwagę punkt odniesienia (SNAIVE). Taka sytuacja miała miejsce zwłaszcza w eksperymencie 5, gdzie wszystkie kombinacje tej metody z uzgadnianiem uzyskały ujemne wartości w skill score względem metody benchmarkingowej.

Jeśli chodzi natomiast o drugi czynnik wzięty pod uwagę w ocenie czy warto stosować prognozowanie hierarchiczne, czyli czas obliczeń, to tutaj nie było niespodzianki i jak można było się spodziewać zastosowanie uzgadniania prognoz powoduje wydłużenie jego wykonywania. Potwierdziły to testy statystyczne przeprowadzone w eksperymencie 6. Należy jednak zauważyć, że różnice pomiędzy otrzymywaniem wartości prognozowanych, a zgodnych był inne dla każdego z modeli i przykładowo dla rodziny ETS oraz ARIMA nie były one znaczne (w przeciwieństwie do modelu Prophet, gdzie uzgadnianie wydłużyło ten czas ponad dwukrotnie, co daje drugi argument w tym badaniu przeciwko jego stosowaniu w danym problemie). Także bazując na tych wynikach wydaje się, że w zależności od metody zastosowanie uzgadniania może przynieść więcej korzyści (trafność), niż negatywnych aspektów (czas otrzymania prognozy).

Zakończenie

Dzięki wykorzystaniu występujących powiązań w strukturze hierarchicznej badanych danych uzyskano dla nich spójne i trafne prognozy. Biorąc jednak pod uwagę całe dokonane opracowanie należy pamiętać, że zastosowano w nim jedynie jeden zbiór danych dotyczący wybranej tematyki - zużycia energii elektrycznej w USA. Należy mieć na uwadze, że otrzymane wyniki można uogólnić w największym stopniu odnosząc się raczej jedynie do tej dziedziny (sektora energetycznego). Mimo to, wciąż istnieje duże prawdopodobieństwo, że wykorzystując obecne w niniejszym opracowaniu rodziny modeli i metody do np. danych pochodzących z innego państwa można by otrzymać zupełnie inne wnioski. Idąc tym tokiem warto byłoby opierając się na wnioskach z tego badania przeprowadzić podobny eksperyment opierając się na innej perspektywie i założeniach.

Bazując na tej idei oraz na wynikach z niniejszego opracowania istnieje bardzo wiele dróg w jaki można by było rozwinąć badania na tenże temat. Po pierwsze warto byłoby się przyjrzeć jak będzie spisywało się prognozowanie hierarchiczne w zależności od wielkości i częstotliwości zestawu danych (tutaj warto byłoby rozważyć zestawy danych o większej częstotliwości, np. dzienne lub nawet śróddzienne). Po drugie, bardzo ciekawą analizę można by było przeprowadzić badając o wiele bardziej skomplikowaną strukturę, przykładowo, o znacznie większej liczbie szeregów czasowych lub sytuacji, gdzie oprócz samych hierarchicznych szeregów byłyby obecne również zgrupowane. Natomiast skupiając się na poprawie samej trafności prognoz można by także użyć bardziej skomplikowanych modeli bazowych, czy np. zastosować ich kombinacje. W dodatku istnieje też kilka innych metod uzgadniania, które można by było wykorzystać i porównać. Oczywiście należy również pamiętać, że badanie zużycia energii elektrycznej to nie jedyna problematyka, gdzie występuje struktura hierarchiczna, tak jak wspomniano w początkowych rozdziałach, na świecie jest obecnych wiele innych dziedzin, które można zbadać w kontekście słuszności wykorzystywania prognozowania hierarchicznego.

Reasumując, a zarazem odnosząc się do postawionego we wstępie tego opracowania stwierdzenia dotyczącego wiedzy i perspektywy niezbędnej dla interesariuszy w sprawie decyzji czy korzystać z prognozowania hierarchicznego, należy podsumować, że jeśli chodzi o zużycie energii elektrycznej to badanie wskazało ważne argumenty, które przemawiają za tym faktem. Stosowanie i rozwijanie badanego w niniejszej pracy podejścia w miejscach związanych z sektorem energetycznym, gdzie wymagana jest prognoza szeregów czasowych charakteryzujących się strukturą hierarchiczną, mogłoby przynieść wiele korzyści, zakładając

oczywiście, że do rozwiązania danego problemu zostaną wybrane odpowiednie kombinacje modelu bazowego i metody uzgadniania.

Bibliografia

Kranz J., Kolbe L.M., Koo C., *Smart energy: where do we stand and where should we go?*, Electron Markets, 2015

George Athanasopoulos, Puwasala Gamakumara, Anastasios Panagiotelis, Rob J Hyndman, Mohamed Affan, Hierarchical forecasting. *Macroeconomic forecasting in the era of big data*, ed. P. Fuleky, Springer, 2020, Chapter 21, pp.689-719.

Souhaib Ben Taieb, Raphael Huser, Rob J Hyndman, Marc G Genton, *Forecasting uncertainty in electricity smart meter data by boosting additive quantile regression.*, *IEEE Transactions on Smart Grid* 7(5), 2016, 2448-2455.

Berrocal, V. J., Raftery, A. E., Gneiting, T. & Steed, R. C., *Probabilistic weather forecasting for winter road maintenance*, *Journal of the American Statistical Association*, 2010, 105(490), 522–537.

Hong, T., Pinson, P., Fan, S., Zareipour, H., Troccoli, A. & Hyndman, R. J., *Probabilistic energy forecasting: Global energy forecasting competition 2014 and beyond*, *International Journal of Forecasting* 32(3), 2016, 896–913.

Cabrera, B. L. & Schulz, F., *Forecasting generalized quantiles of electricity demand: A functional data approach*, *Journal of the American Statistical Association* 112(517), 2017, 127–136.

Sun, X., Luh, P. B., Cheung, K. W., Guan, W., Michel, L. D., Venkata, S. S. & Miller, M. T., *'An efficient approach to Short-Term load forecasting at the distribution level'*, *IEEE Transactions on Power Systems* 31(4), 2016, 2526–2537.

Arora, S. & Taylor, J. W., *Forecasting electricity smart meter data using conditional kernel density estimation*, *Omega* 59, 2016, Part A, 47–59.

F. L. Quilumba, W. Lee, H. Huang, D. Y. Wang, and R. L. Szabados, *Using smart meter data to improve the accuracy of intraday load forecasting considering customer behavior similarities*, *IEEE Transactions on Smart Grid*, 2015, vol. 6, no. 2, pp. 911–918.

Cieślak M., *Prognozowanie Gospodarcze Metody i Zastosowania*, Wydawnictwo Naukowe PWN, Warszawa 2005

Słownik języka polskiego PWN, <https://sjp.pwn.pl/slowniki/hierarchia.html> [dostęp: 01.07.2022]

Hyndman, R.J., & Athanasopoulos, G., *Forecasting: principles and practice*, 3rd edition, 2021, OTexts: Melbourne, Australia. OTexts.com/fpp3.

Wickramasuriya, S. L., Athanasopoulos, G., & Hyndman, R. J., *Optimal forecast reconciliation for hierarchical and grouped time series through trace minimization.*, *Journal of the American Statistical Association*, 2019, 114(526), 804–819.

Blessington Mark, *Sales Forecasting: A Practical Guide*, CreateSpace Independent Publishing Platform, 2015

Ord Keith & Lowe Sam, *Automatic Forecasting*, American Statistician - AMER STATIST. 50. 88-94., 1996

Rob J Hyndman, Yeasmin Khandakar, *Automatic time series forecasting: the forecast package for R*, *Journal of Statistical Software* **27** (3), 2008

Rob J Hyndman, *Automatic forecasting algorithms*, 2017, <https://robjhyndman.com/files/2-AutomaticForecasting.pdf> [dostęp: 15.07.2021]

Christoph Bergmeir, Rob J Hyndman, José M Benítez, *Bagging exponential smoothing methods using STL decomposition and Box-Cox transformation*, *International Journal of Forecasting* **32**(2), 2016, 303-312.

Rob J Hyndman, *ARIMA processes*, In: *Informed Student Guide to Management Science*, 2002, ed., Hans Daellenbach and Robert Flood, Thomson: London.

Alysha M De Livera, Rob J Hyndman, Ralph D Snyder, *Forecasting time series with complex seasonal patterns using exponential smoothing*, *Journal of the American Statistical Association* **106**(496), 2011, 1513-1527.

Hyndman, R. J., & Khandakar, Y., *Automatic Time Series Forecasting: The forecast Package for R*, *Journal of Statistical Software*, 27(3), 2008, 1–22.,

Taylor SJ, Letham B., *Forecasting at scale*, *PeerJ Preprints*, 2017

Skoczypiec S., *Metody prognozowania: Jakość prognoz*, 2012, http://m6.pk.edu.pl/materialy/mp/MP_02_bledy_prognozy.pdf [dostęp: 01.07.2022]

Rob J Hyndman, Anne B Koehler, *Another look at measures of forecast accuracy*, *International Journal of Forecasting* **22**(4), 2006, 679-688.

Gneiting, T., & Katzfuss, M., *Probabilistic forecasting*, *Annual Review of Statistics and Its Application*, 1(1), 2014, 125–151.

U.S. Energy Information Administration, <https://www.eia.gov/electricity/data.php> [dostęp: 01.07.2022]

U.S. Energy Information Administration, https://www.eia.gov/about/mission_overview.php [dostęp: 01.07.2022]

United States Census Bureau, https://www.census.gov/programs-surveys/economic-census/guidance-geographies/levels.html#par_textimage_34 [dostęp: 15.07.2022]

Hyndman R, *Detecting time series outliers*, 2021, <https://robjhyndman.com/hyndsight/tsoutliers/> [dostęp: 27.07.2022]

Luedicke Joerg, *Friedman's Super Smoother*, 2015.

U.S. Energy Information Administration,

<https://www.eia.gov/todayinenergy/detail.php?id=42915> [dostęp: 29.07.2022]

The R Project for Statistical Computing, <https://www.r-project.org/about.html> [dostęp: 29.07.2022]

The R Project for Statistical Computing, <https://cran.r-project.org/web/packages/fable/index.html> [dostęp: 29.07.2022]

The R Project for Statistical Computing, <https://cran.r-project.org/web/packages/fable.prophet/index.html> [dostęp: 29.07.2022]

Rob J Hyndman, Yeasmin Khandakar, Automatic time series forecasting: the forecast package for R, *Journal of Statistical Software* 27(3), 2008

Tidyverse, <https://www.tidyverse.org> [dostęp: 29.07.2022]

Li Z, Ye H, Liao N, Wang R, Qiu Y, Wang Y, *Impact of COVID-19 on electricity energy consumption: A quantitative analysis on electricity*, *International Journal of Electrical Power & Energy Systems*, 2022

STHDA Statistical tools for high-throughput data analysis,
<http://www.sthda.com/english/wiki/unpaired-two-samples-wilcoxon-test-in-r> [dostęp: 26.08.2022]

Spis tabel

Tabela 1 Metody benchmarkingowe – MAPE i RMSE dla danych treningowych	47
Tabela 2 Metody benchmarkingowe – MAPE i RMSE dla danych testowych	47
Tabela 3 Modele bazowe i SNAIVE – poziom USA	49
Tabela 4 Modele bazowe i SNAIVE – poziom State	50
Tabela 5 Model bazowy ETS - MAPE, MASE, RMSE	51
Tabela 6 Model bazowy ARIMA - MAPE, MASE, RMSE	53
Tabela 7 Model bazowy Prophet - MAPE, MASE, RMSE	54
Tabela 8 Poziom USA - MAPE, MASE, RMSE	58
Tabela 9 Poziom Division - MAPE, MASE, RMSE	59
Tabela 10 Poziom State - MAPE, MASE, RMSE	61
Tabela 11 Wszystkie szeregi - MAPE, MASE, RMSE	62
Tabela 12 CRPS dla poszczególnych poziomów struktury	64
Tabela 13 CRPS dla wszystkich szeregów struktury	65
Tabela 14 Średnie i odchylenia standardowe czasów wykonywania poszczególnych metod	66
Tabela 15 Wyniki testu Wilcozona dla każdego modelu.	67

Spis rysunków

Rysunek 1 Diagram dwupoziomowego drzewa hierarchicznego	7
Rysunek 2 Prosta dwupoziomowa zgrupowana struktura.	9
Rysunek 3 Schemat działania uzgadniania prognoz	10
Rysunek 4 Równania ETS	21
Rysunek 5 Strona internetowa EIA - miejsce skąd pobrano dane	33
Rysunek 6 Nieprzetworzone dane w pliku .xls.....	34
Rysunek 7 Podział USA na jednostki statystyczne.....	35
Rysunek 8 Podział stosowany przez Biuro Spisu Ludności USA	36
Rysunek 9 Wstępnie przetworzone dane	37
Rysunek 10 Dane wstępnie przetworzone, kwartalne	38
Rysunek 11 Finalnie przetworzone dane	38
Rysunek 12 USA - trend	39
Rysunek 13 USA - sezonowość	40
Rysunek 14 USA - wykres podserii.....	40
Rysunek 15 USA - ACF	41
Rysunek 16 Metody benchmarkingowe - prognoza	46
Rysunek 17 Metody benchmarkingowe - porównanie dokładności na bazie danych treningowych.....	46
Rysunek 18 Metody benchmarkingowe - porównanie dokładności na bazie danych testowych	47
Rysunek 19 Porównanie trafności prognozy modeli bazowych na poziomie USA	48
Rysunek 20 Porównanie trafności modeli bazowych na poziomie State	49
Rysunek 21 Wizualizacja prognozy dla modelu bazowego ETS	51
Rysunek 22 Wizualizacja prognozy dla modelu bazowego ARIMA	52
Rysunek 23 Wizualizacja prognozy dla modelu bazowego Prophet	54
Rysunek 24 ARIMA – szereg o największym błędzie MAPE dla próbki testowej.....	56
Rysunek 25 ETS - szereg o największym błędzie MAPE dla próbki testowej	56
Rysunek 26 Prophet - szereg o największym błędzie MAPE dla próbki testowej	57
Rysunek 27 Poziom USA - porównanie prognozy punktowej względem SNAIVE	58
Rysunek 28 Poziom Division - porównanie prognozy punktowej względem SNAIVE	59
Rysunek 29 Poziom State - porównanie prognozy punktowej względem SNAIVE	60
Rysunek 30 Wszystkie szeregi - porównanie prognozy punktowej względem SNAIVE	61
Rysunek 31 Porównanie prognozy przedziałowej względem SNAIVE – poszczególne poziomy struktury	63
Rysunek 32 Porównanie prognozy przedziałowej względem SNAIVE – wszystkie szeregi..	64
Rysunek 33 Wykresy pudełkowe dla czasów wykonywania poszczególnych metod	66