

Data oddania: _____

Ocena: _____

Mateusz Walczak 216911

Konrad Kajszczak 216790

Zadanie 2: Lingwistyczne podsumowania baz danych*

1. Cel

Praca w toku

2. Wprowadzenie

Praca w toku

3. Opis implementacji

Praca w toku

4. Materiały i metody

Wybrana przez nas baza danych zawiera historyczne pomiary pogodowe z Holandii [1]. Dane zostały zgromadzone przez KNMI (*Dutch weather institute* - Holenderski instytut pogody) na przestrzeni lat 1901-2018 i pochodziły z 50 różnych stacji pogowych znajdujących się na terenie całego kraju.

* SVN: <https://github.com/Walducha1908/KSR2>

Ze względu na fakt, iż oryginalna baza danych składa się z 804099 krotek, postanowiliśmy wybrać tylko niewielką część z dostępnych danych. Zdecydowaliśmy się na najnowsze dane pomiarowe - z lat 2016-2018. W ten sposób ograniczyliśmy liczbę wykorzystywanych krotek do 17000.

4.1. Wybór kolumn

W celu analizy bazy danych i tworzenia jej lingwistycznych podsumowań wybraliśmy następujące 10 kolumn z danymi liczbowymi:

- FG - średnia prędkość wiatru przez cały dzień [$0.1 \frac{m}{s}$].
- FHX - najwyższa średnia prędkość wiatru w ciągu jednej godziny [$0.1 \frac{m}{s}$].
- FHN - najniższa średnia prędkość wiatru w ciągu jednej godziny [$0.1 \frac{m}{s}$].
- FXX - najszybszy podmuch wiatru w ciągu całego dnia [$0.1 \frac{m}{s}$].
- TG - średnia dzienna temperatura [$0.1^{\circ}C$].
- TN - minimalna dzienna temperatura [$0.1^{\circ}C$].
- TX - maksymalna dzienna temperatura [$0.1^{\circ}C$].
- T10N - minimalna dzienna temperatura na wysokości 10 cm od poziomu gruntu [$0.1^{\circ}C$].
- Q - nasłonecznienie, energia słoneczna przypadająca na powierzchnię [$\frac{J}{cm^2}$].
- RH - suma opadów atmosferycznych w ciągu całego dnia [$0.1mm$].

Oprócz wyżej opisanych danych liczbowych, w naszej bazie znajdują się także dwie dodatkowe kolumny, służące do identyfikacji pomiaru:

- STN - numer stacji badawczej wykonującej pomiar.
- YYYYMMDD - data pomiaru w formacie opisanym przez nazwę kolumny.

	A	B	C	D	E	F	G	H	I	J	K	L
1	STN	YYYYMMDD	FG	FHX	FHN	FXX	TG	TN	TX	T10N	Q	RH
2	380	20181231	23	30	10	60	83	74	91	70	96	19
3	370	20181231	27	40	20	70	89	73	99	72	115	5
4	350	20181231	28	40	20	80	89	75	98	72	132	5
5	375	20181231	28	50	20	90	90	73	98	71	105	1
6	290	20181231	36	60	20	90	88	74	99	73	126	2
7	275	20181231	30	50	20	90	86	72	98	70	132	1
8	279	20181231	40	60	20	100	86	78	97	77	150	-1
9	260	20181231	27	40	20	100	89	77	100	76	137	1
10	269	20181231	36	50	20	80	87	78	100	75	155	-1
11	280	20181231	40	60	20	110	88	80	99	77	174	5
12	240	20181231	45	70	30	110	90	83	109	76	205	-1
13	344	20181231	35	50	20	90	91	79	104	77	207	-1
14	215	20181231	42	60	20	90	89	82	97	80	212	0
15	235	20181231	51	80	40	110	90	83	98	77	229	-1
16	270	20181231	50	80	30	120	87	82	97	78	236	-1
17	310	20181231	48	60	30	90	88	73	102	71	335	3
18	375	20181230	32	40	20	90	79	53	92	42	137	31
19	350	20181230	29	50	20	80	82	56	93	49	139	14
20	260	20181230	26	50	20	100	85	62	95	52	112	10
21	370	20181230	33	50	20	90	80	51	91	45	156	11
22	269	20181230	33	60	20	110	82	63	92	54	119	8
23	344	20181230	36	60	20	120	86	50	97	36	141	4
24	215	20181230	43	60	20	100	87	57	99	48	127	7
25	275	20181230	35	70	20	120	77	55	90	48	138	24
26	279	20181230	39	80	20	160	76	67	84	62	148	10

Rysunek 1. Fragment widoku bazy w formacie *xlsx*

4.2. Wykorzystywane funkcje przynależności

W celu zdefiniowania zmiennych lingwistycznych i kwantyfikatorów, posłużyliśmy się dwoma rodzajami funkcji przynależności:

- funkcją trójkątną opisaną wzorem:

$$f_{troj}(x) = \begin{cases} \frac{x-a}{b-a} & \text{jeśli } a \leq x < b \\ 1 & \text{jeśli } x = b \\ \frac{c-x}{c-b} & \text{jeśli } b < x \leq c \\ 0 & \text{w przeciwnym wypadku.} \end{cases} \quad (1)$$

- oraz funkcją trapezoidalną opisaną wzorem:

$$f_{trap}(x) = \begin{cases} \frac{x-a}{b-a} & \text{jeśli } a \leq x < b \\ 1 & \text{jeśli } b \leq x \leq c \\ \frac{d-x}{d-c} & \text{jeśli } c < x \leq d \\ 0 & \text{w przeciwnym wypadku.} \end{cases} \quad (2)$$

gdzie a, b, c oraz d są parametrami funkcji przynależności - wierzchołkami trójkąta lub trapezu na wykresie.

4.3. Zmienne lingwistyczne

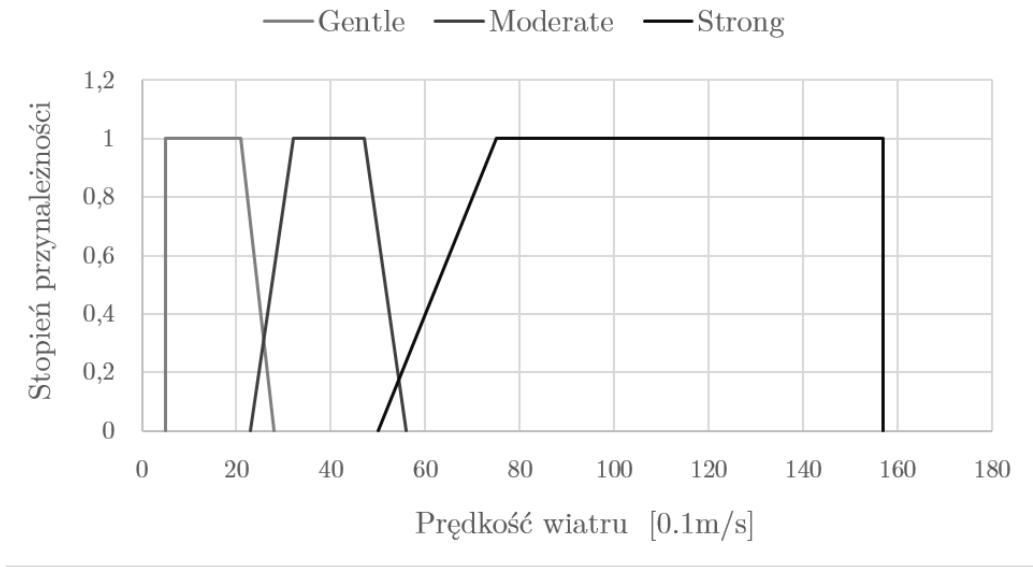
W tym rozdziale przedstawimy wzory i wykresy opisujące zaproponowane przez nas zmienne lingwistyczne. We wszystkich przypadkach, wykorzystywanymi przez nas funkcjami przynależności są funkcje trapezoidalne.¹

Aby nie duplikować treści wzorów, niepotrzebnie zwiększając w ten sposób objętość sprawozdania, zdecydowano się na zamieszczenie tabel z parametrami etykiet zmiennych lingwistycznych, odnoszącymi się do wzorów z poprzedniego podrozdziału.

¹ Wartości prezentowane w tabelach są tylko propozycjami. Autorzy sprawozdania zastrzegają sobie możliwość do ich późniejszej modyfikacji

4.3.1. Kolumna FG

Wykres opisujący zmienną lingwistyczną dla kolumny zawierającej wartości średniej prędkości wiatru przez cały dzień (FG), zamieszczono poniżej.



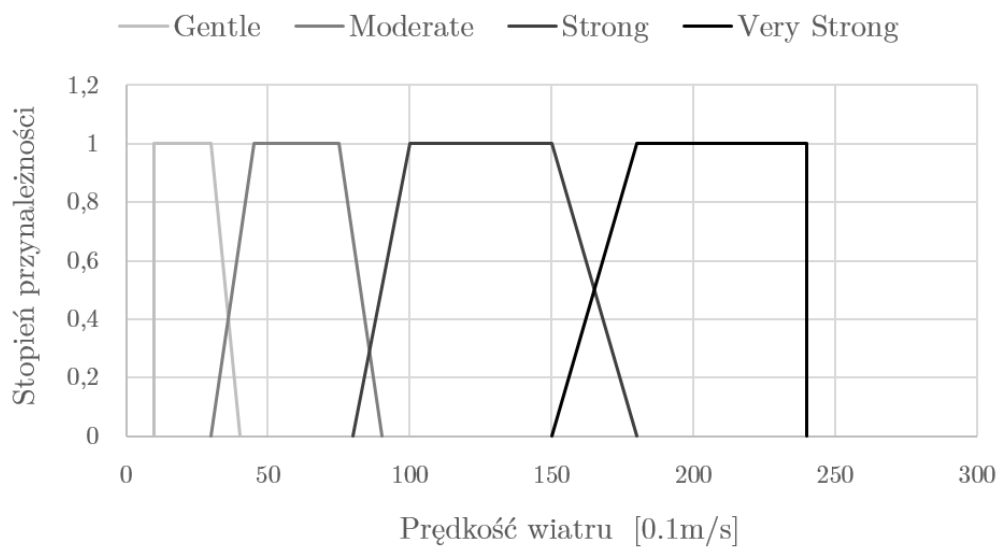
Rysunek 2. Wykres opisujący zmienną lingwistyczną dla kolumny FG.

Etykieta	a	b	c	d
Gentle	5	5	21	28
Moderate	23	32	47	56
Strong	50	75	157	157

Tabela 1. Przyporządkowane parametry funkcji trapezoidalnej dla kolumny FG.

4.3.2. Kolumna FHX

Wykres opisujący zmienną lingwistyczną dla kolumny zawierającej najwyższą średnią prędkości wiatru w przeciągu jednej godziny (FHX), zamieszczono poniżej.



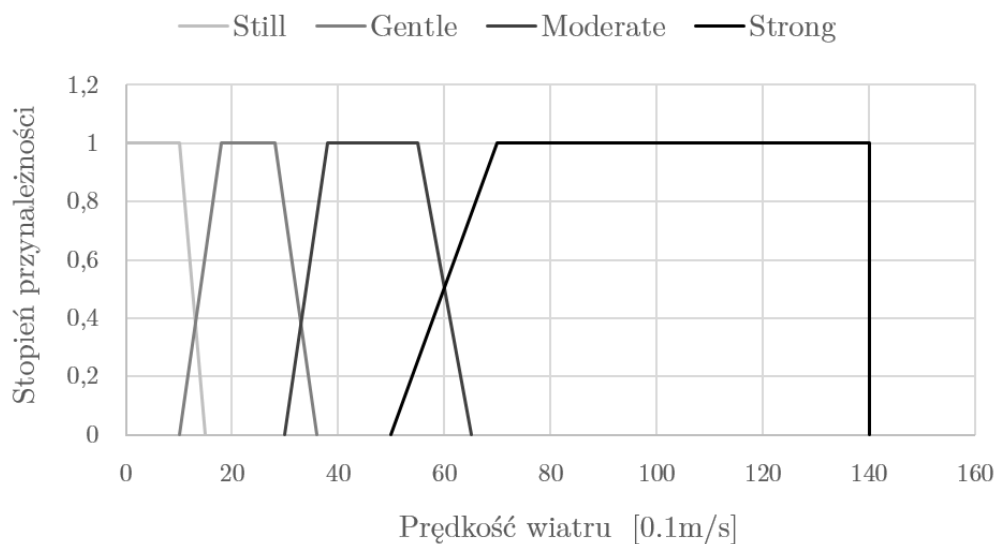
Rysunek 3. Wykres opisujący zmienną lingwistyczną dla kolumny FHX.

Etykieta	a	b	c	d
Gentle	10	10	30	40
Moderate	30	45	75	90
Strong	80	100	150	180
Very strong	150	180	240	240

Tabela 2. Przyporządkowane parametry funkcji trapezoidalnej dla kolumny FHX.

4.3.3. Kolumna FHN

Wykres opisujący zmienną lingwistyczną dla kolumny zawierającej najniższą średnią prędkości wiatru w przeciągu jednej godziny (FHN), zamieszczono poniżej.



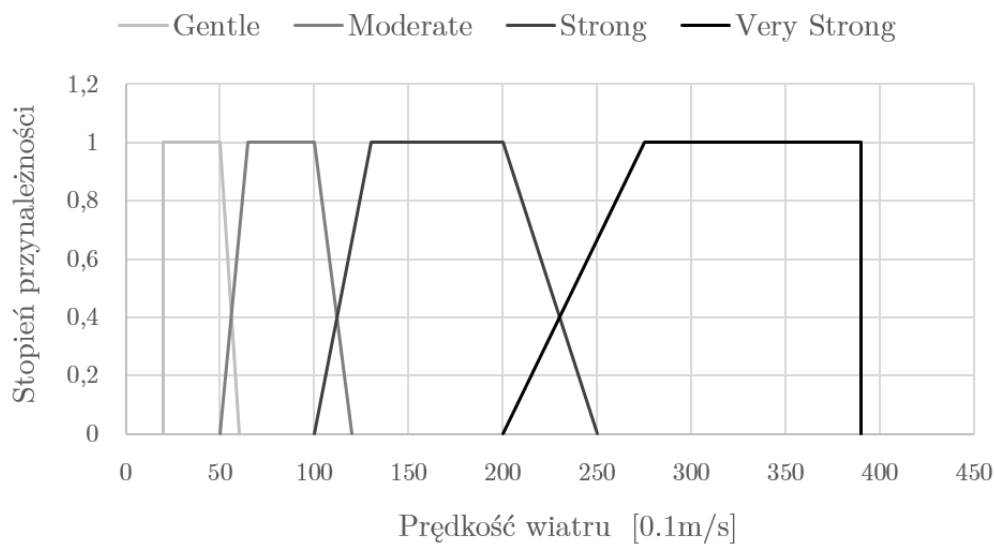
Rysunek 4. Wykres opisujący zmienną lingwistyczną dla kolumny FHN.

Etykieta	a	b	c	d
Gentle	0	0	10	15
Moderate	10	18	28	36
Strong	30	38	55	65
Very strong	50	70	140	140

Tabela 3. Przyporządkowane parametry funkcji trapezoidalnej dla kolumny FHN.

4.3.4. Kolumna FXX

Wykres opisujący zmienną lingwistyczną dla kolumny zawierającej najsilniejszy powiew wiatru (FXX), zamieszczono poniżej.



Rysunek 5. Wykres opisujący zmienną lingwistyczną dla kolumny FXX.

Etykieta	a	b	c	d
Gentle	20	20	50	60
Moderate	50	65	100	120
Strong	100	130	200	250
Very strong	200	275	390	390

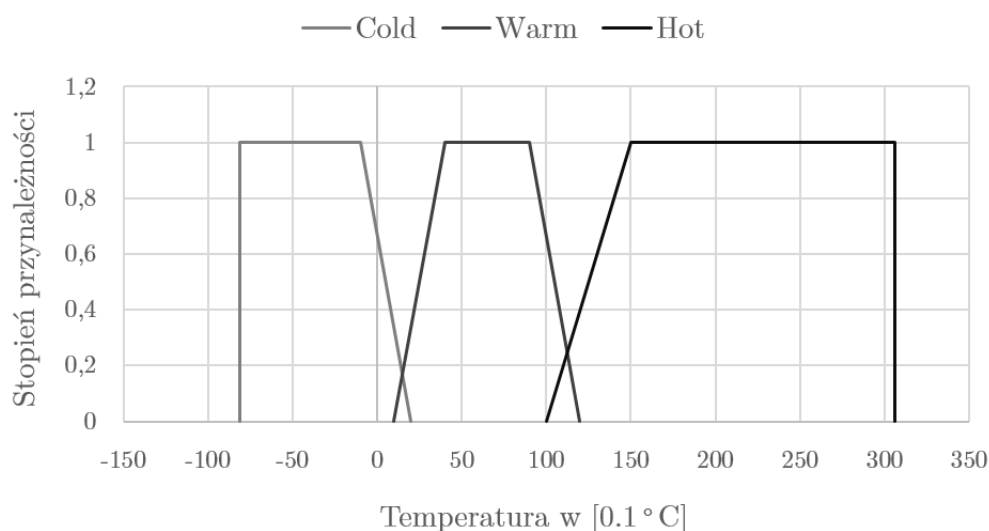
Tabela 4. Przyporządkowane parametry funkcji trapezoidalnej dla kolumny FXX.

4.3.5. Kolumna TG

W przypadku średniej dziennej temperatury (TG) oraz innych kolumn związanych z temperaturą (TN, TX oraz T10N), zdecydowaliśmy się podzielić nasze rozważania ze względu na pory roku. Dlatego też przyjęliśmy trzy różne warianty zmiennej lingwistycznej dla kolumny TG:

- TGW - dla pomiarów uzyskanych podczas astronomicznej zimy (litera *W* od *Winter*),
- TGSA - dla pomiarów uzyskanych podczas astronomicznej wiosny lub jesieni (*S* od *Spring*, *A* od *Autumn*),
- TGS - dla pomiarów uzyskanych podczas astronomicznego lata (litera *S* od *Summer*).

Rozpocznijmy od zmiennej lingwistycznej TGW.

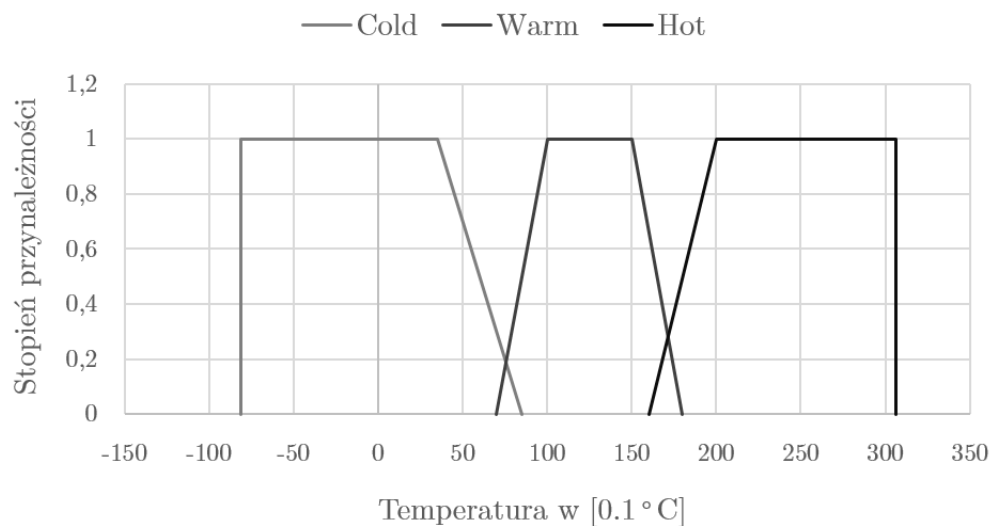


Rysunek 6. Wykres opisujący zmienną lingwistyczną dla kolumny TG dla pomiarów wykonanych astronomiczną zimą.

Etykieta	a	b	c	d
Cold	-81	-81	-10	20
Warm	10	40	90	120
Hot	100	150	306	306

Tabela 5. Przyporządkowane parametry funkcji trapezoidalnej dla zmiennej TGW.

Następną prezentowaną zmienną, będzie zmienna lingwistyczna TGSA.

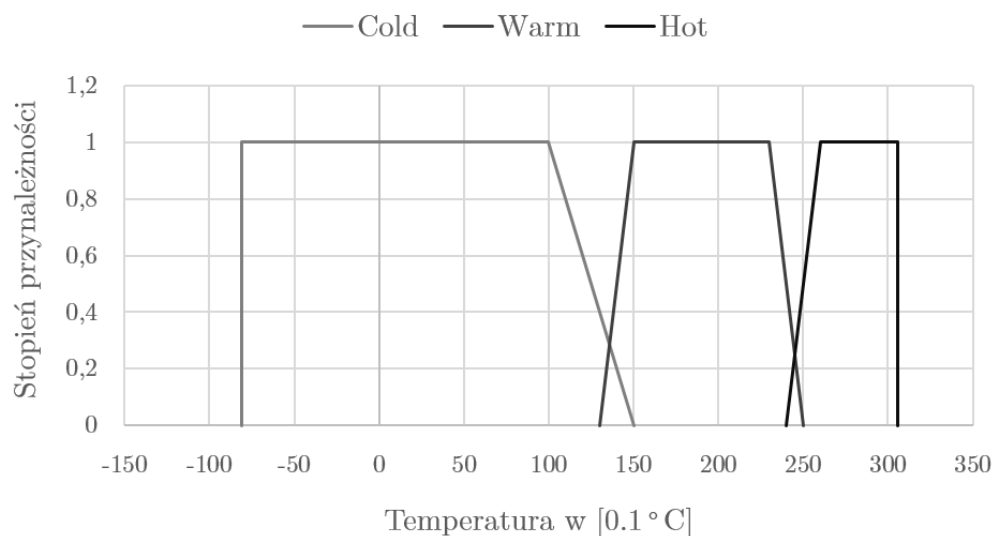


Rysunek 7. Wykres opisujący zmienną lingwistyczną dla kolumny TG dla pomiarów wykonanych astronomiczną wiosną i jesienią.

Etykieta	a	b	c	d
Cold	-81	-81	35	85
Warm	70	100	150	180
Hot	160	200	306	306

Tabela 6. Przyporządkowane parametry funkcji trapezoidalnej dla zmiennej TGSA.

Ostatnią zmienną dla kolumny TG będzie zmienna dotycząca pomiarów letnich - TGS.



Rysunek 8. Wykres opisujący zmienną lingwistyczną dla kolumny TG dla pomiarów wykonanych astronomicznym latem.

Etykieta	a	b	c	d
Cold	-81	-81	100	150
Warm	130	150	230	250
Hot	240	260	306	306

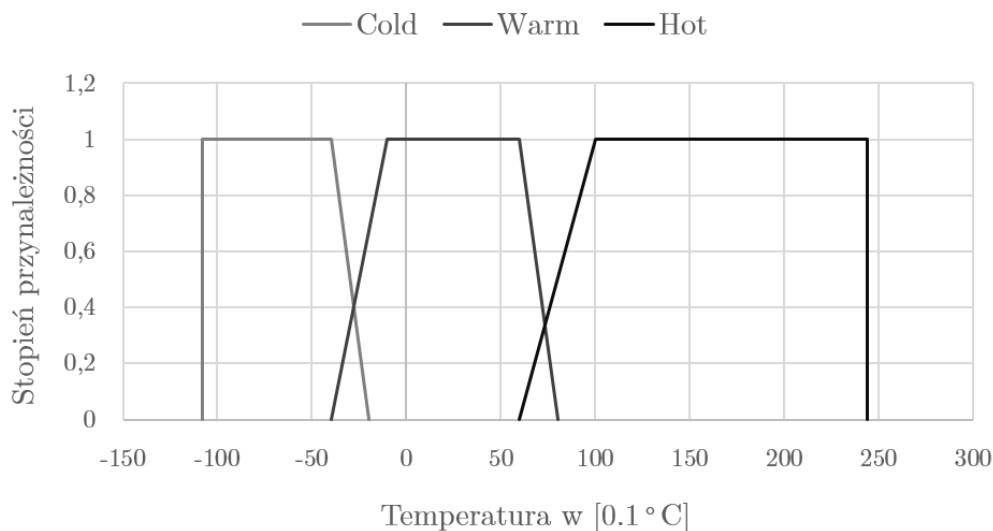
Tabela 7. Przyporządkowane parametry funkcji trapezoidalnej dla zmiennej TGS.

4.3.6. Kolumna TN

Kolumna TN zawiera najniższą temperaturę powietrza w ciągu dnia. Wszystkie trzy warianty zmiennej lingwistycznej dla kolumny TN zaprezentowano poniżej:

- TNW - dla pomiarów uzyskanych podczas astronomicznej zimy,
- TNSA - dla pomiarów uzyskanych podczas astronomicznej wiosny lub jesieni,
- TNS - dla pomiarów uzyskanych podczas astronomicznego lata.

Pomiary zimowe - zmienna TNW.

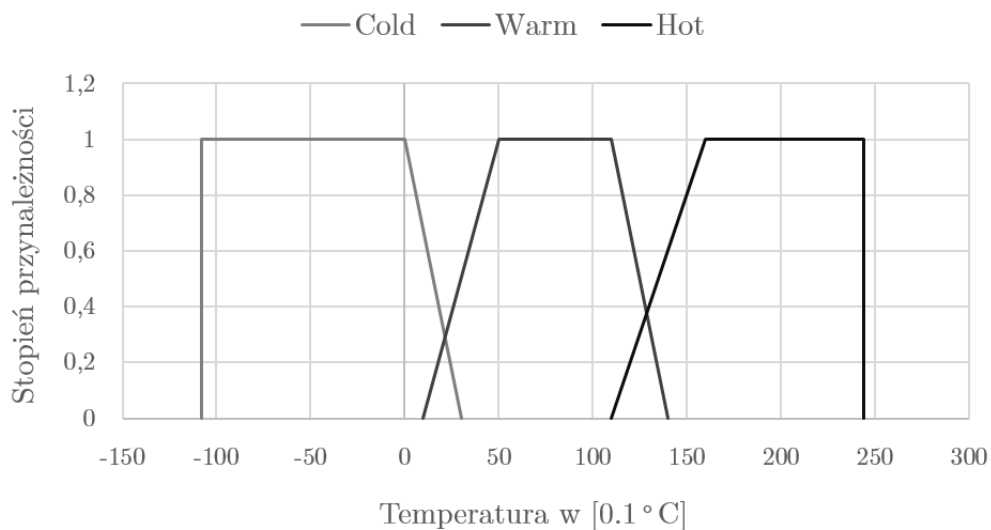


Rysunek 9. Wykres opisujący zmienną lingwistyczną dla kolumny TN dla pomiarów wykonanych astronomiczną zimą.

Etykieta	a	b	c	d
Cold	-108	-108	-40	-20
Warm	-40	-10	60	80
Hot	60	100	244	244

Tabela 8. Przyporządkowane parametry funkcji trapezoidalnej dla zmiennej TNW.

Pomiary wiosenne i jesienne - zmienna TNSA.

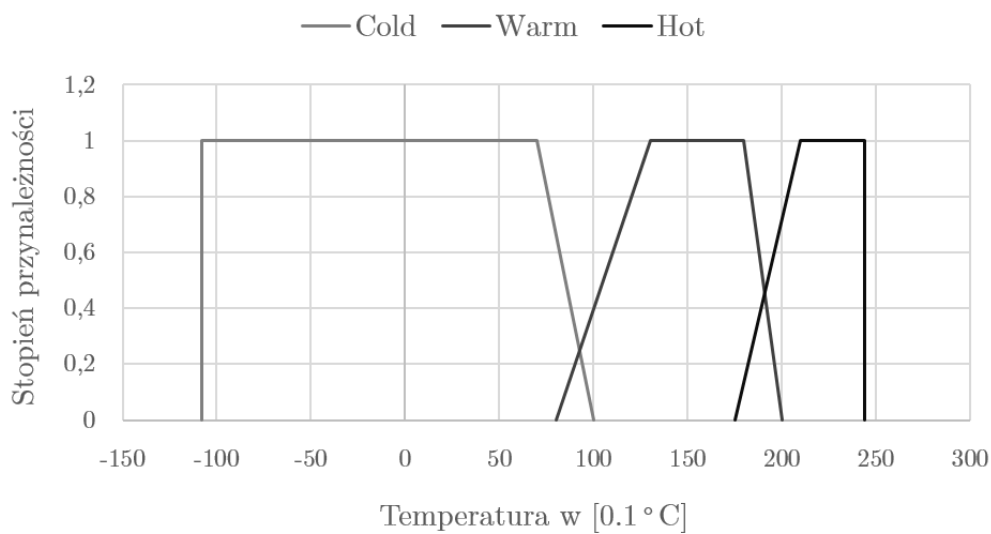


Rysunek 10. Wykres opisujący zmienną lingwistyczną dla kolumny TN dla pomiarów wykonanych astronomiczną wiosną i jesienią.

Etykieta	a	b	c	d
Cold	-108	-108	0	30
Warm	10	50	110	140
Hot	110	160	244	244

Tabela 9. Przyporządkowane parametry funkcji trapezoidalnej dla zmiennej TNSA.

Pomiary letnie - zmienna TNS.



Rysunek 11. Wykres opisujący zmienną lingwistyczną dla kolumny TN dla pomiarów wykonanych astronomicznym latem.

Etykieta	a	b	c	d
Cold	-108	-108	70	100
Warm	80	130	180	200
Hot	175	210	244	244

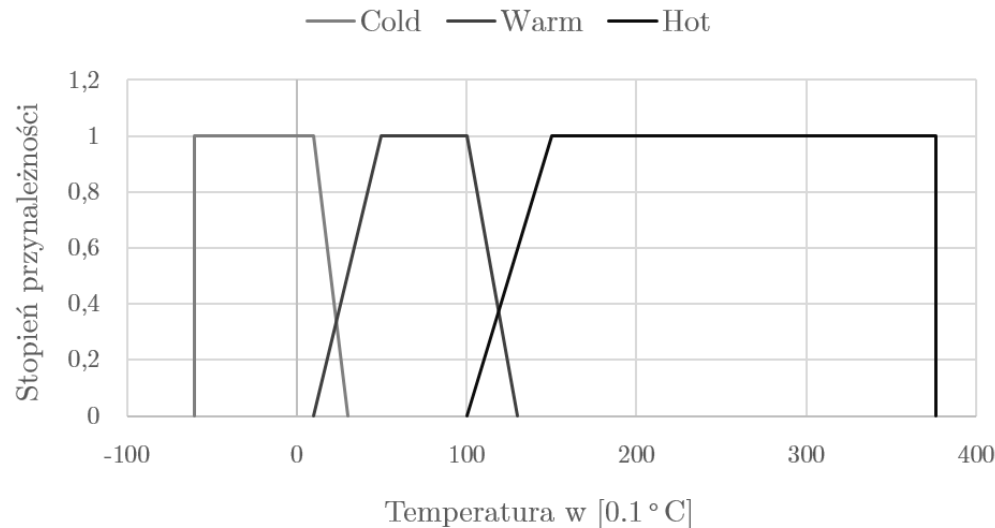
Tabela 10. Przyporządkowane parametry funkcji trapezoidalnej dla zmiennej TNS.

4.3.7. Kolumna TX

Kolumna TX zawiera najwyższą temperaturę powietrza w ciągu dnia. Wszystkie trzy warianty zmiennej lingwistycznej dla kolumny TX zaprezentowano poniżej:

- TXW - dla pomiarów uzyskanych podczas astronomicznej zimy,
- TXSA - dla pomiarów uzyskanych podczas astronomicznej wiosny lub jesieni,
- TXS - dla pomiarów uzyskanych podczas astronomicznego lata.

Pomiary zimowe - zmienna TXW.

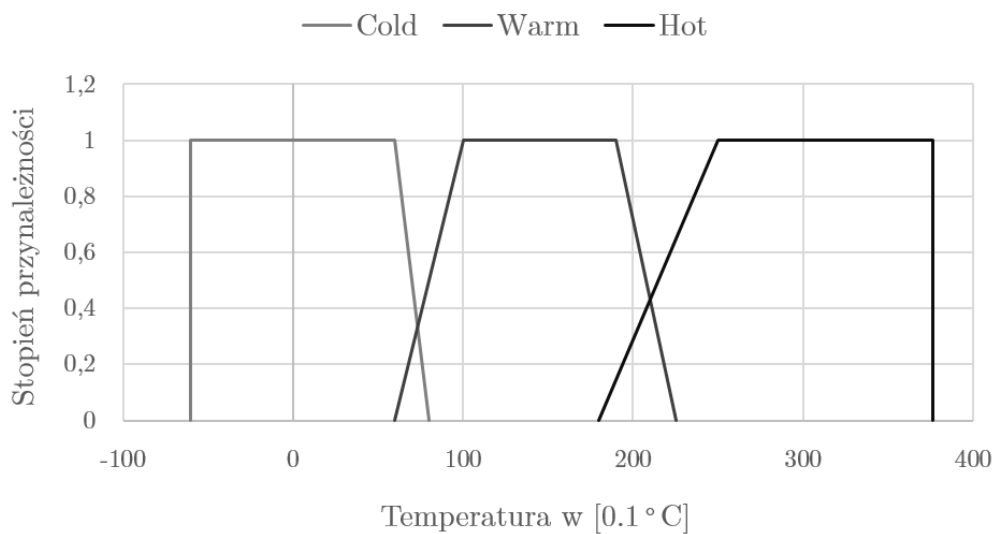


Rysunek 12. Wykres opisujący zmienną lingwistyczną dla kolumny TX dla pomiarów wykonanych astronomiczną zimą.

Etykieta	a	b	c	d
Cold	-60	-60	10	30
Warm	10	50	100	130
Hot	100	150	376	376

Tabela 11. Przyporządkowane parametry funkcji trapezoidalnej dla zmiennej TXW.

Pomiary wiosenne i jesienne - zmienna TXSA.

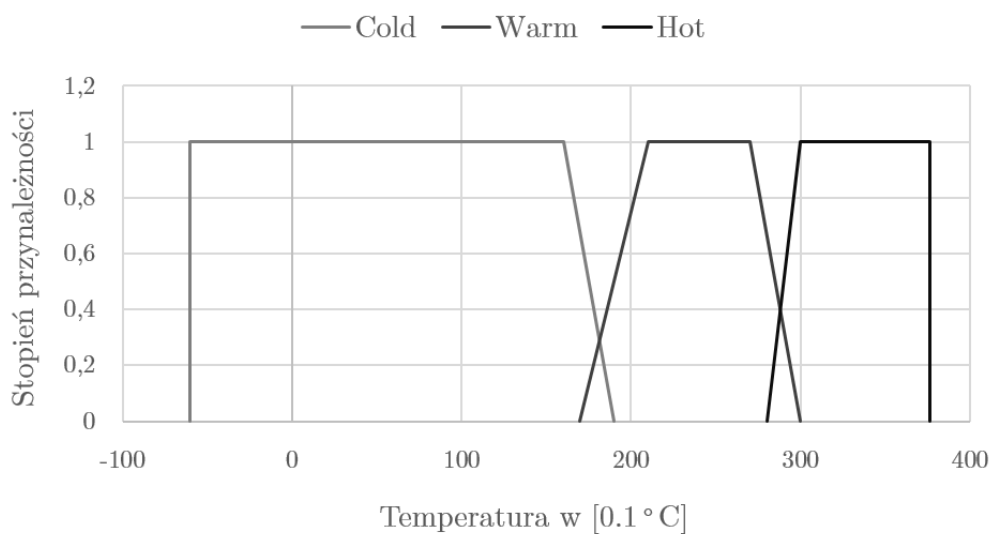


Rysunek 13. Wykres opisujący zmienną lingwistyczną dla kolumny TX dla pomiarów wykonanych astronomiczną wiosną i jesienią.

Etykieta	a	b	c	d
Cold	-60	-60	60	80
Warm	60	100	190	225
Hot	180	250	376	376

Tabela 12. Przyporządkowane parametry funkcji trapezoidalnej dla zmiennej TXSA.

Pomiary letnie - zmienna TXS.



Rysunek 14. Wykres opisujący zmienną lingwistyczną dla kolumny TX dla pomiarów wykonanych astronomicznym latem.

Etykieta	a	b	c	d
Cold	-60	-60	160	190
Warm	170	210	270	300
Hot	280	300	376	376

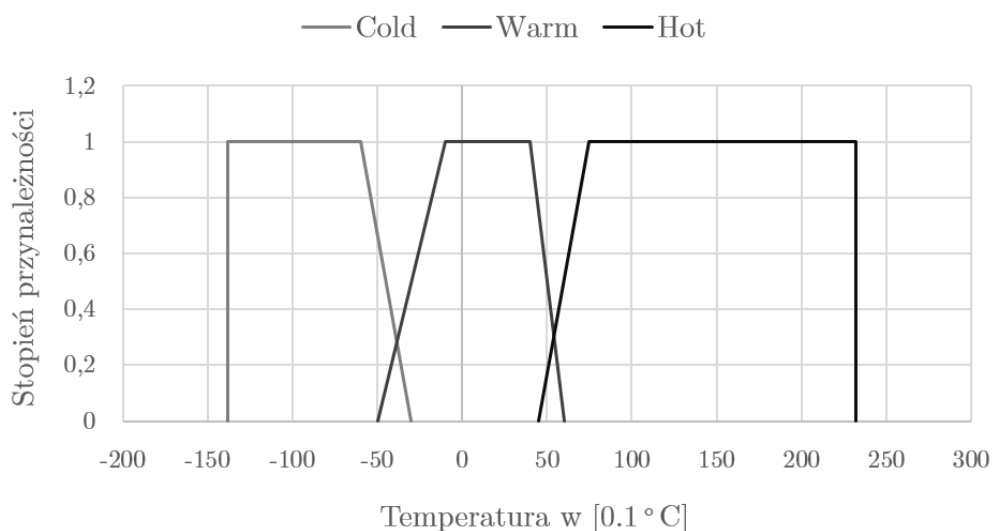
Tabela 13. Przyporządkowane parametry funkcji trapezoidalnej dla zmiennej TXS.

4.3.8. Kolumna T10N

Kolumna T10N zawiera najmniejszą temperaturę w ciągu dnia zmierzoną na wysokości 10cm od poziomu gruntu. Wszystkie trzy warianty zmiennej lingwistycznej dla kolumny T10N zaprezentowano poniżej:

- T10NW - dla pomiarów uzyskanych podczas astronomicznej zimy,
- T10NSA - dla pomiarów uzyskanych podczas astronomicznej wiosny lub jesieni,
- T10NS - dla pomiarów uzyskanych podczas astronomicznego lata.

Pomiary zimowe - zmienna T10NW.

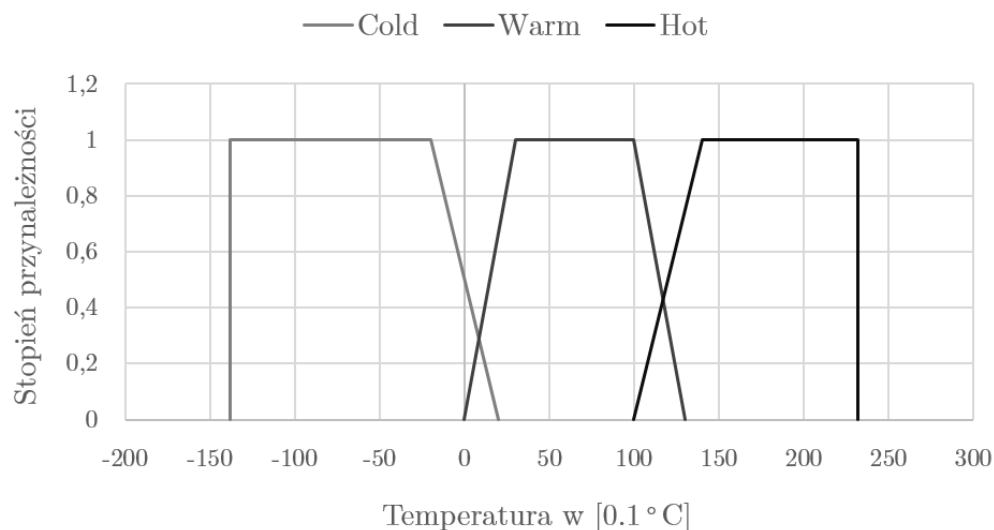


Rysunek 15. Wykres opisujący zmienną lingwistyczną dla kolumny T10N dla pomiarów wykonanych astronomiczną zimą.

Etykieta	a	b	c	d
Cold	-138	-138	-60	-30
Warm	-50	-10	40	60
Hot	45	75	232	232

Tabela 14. Przyporządkowane parametry funkcji trapezoidalnej dla zmiennej T10NW.

Pomiary wiosenne i jesienne - zmienna T10NSA.

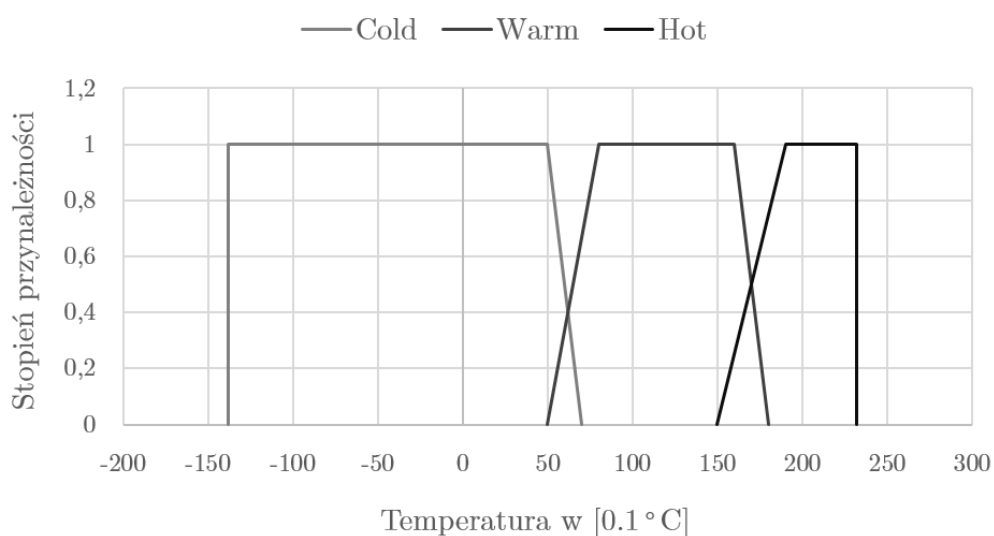


Rysunek 16. Wykres opisujący zmienną lingwistyczną dla kolumny T10N dla pomiarów wykonanych astronomiczną wiosną i jesienią.

Etykieta	a	b	c	d
Cold	-138	-138	-20	20
Warm	0	30	100	130
Hot	100	140	232	232

Tabela 15. Przyporządkowane parametry funkcji trapezoidalnej dla zmiennej T10NSA.

Pomiary letnie - zmienna T10NS.



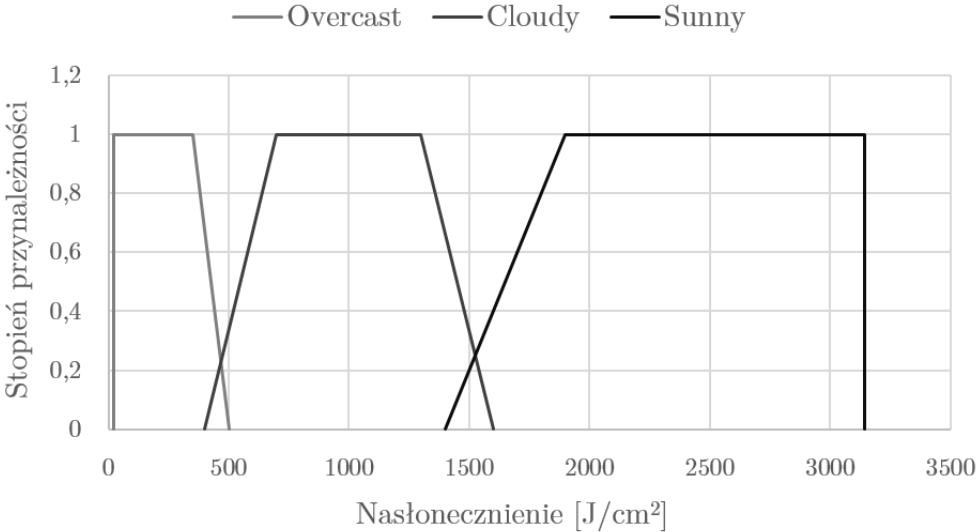
Rysunek 17. Wykres opisujący zmienną lingwistyczną dla kolumny T10N dla pomiarów wykonanych astronomicznym latem.

Etykieta	a	b	c	d
Cold	-138	-138	50	70
Warm	50	80	160	180
Hot	150	190	232	232

Tabela 16. Przyporządkowane parametry funkcji trapezoidalnej dla zmiennej T10NS.

4.3.9. Kolumna Q

Wykres opisujący zmienną lingwistyczną dla kolumny zawierającej wartości nasłonecznienia (Q), zamieszczono poniżej.



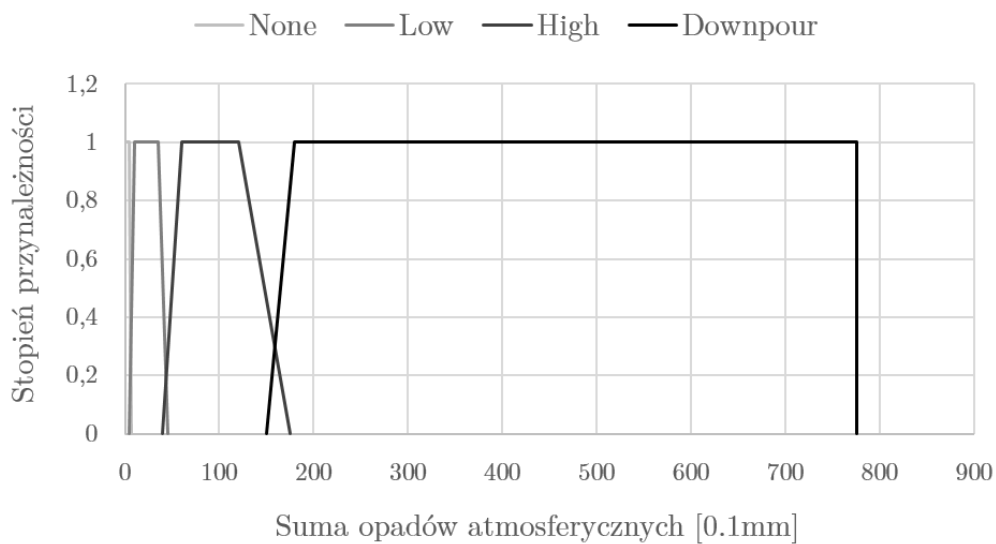
Rysunek 18. Wykres opisujący zmienną lingwistyczną dla kolumny Q

Etykieta	a	b	c	d
Overcast	24	24	350	500
Cloudy	400	700	1300	1600
Sunny	1400	1900	3145	3145

Tabela 17. Przyporządkowane parametry funkcji trapezoidalnej dla kolumny Q.

4.3.10. Kolumna RH

Wykres opisujący zmienną lingwistyczną dla kolumny zawierającej sumę opadów atmosferycznych w ciągu całego dnia (RH), zamieszczono poniżej.



Rysunek 19. Wykres opisujący zmienną lingwistyczną dla kolumny RH

Etykieta	a	b	c	d
None	-1	-1	5	7
Low	5	10	35	45
High	40	60	120	175
Downpour	150	180	776	776

Tabela 18. Przyporządkowane parametry funkcji trapezoidalnej dla kolumny RH.

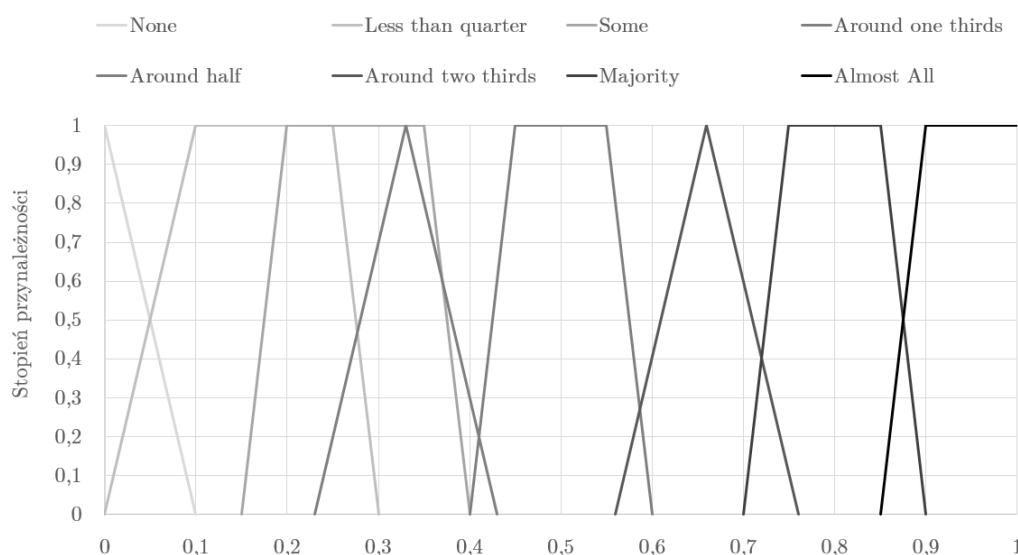
4.4. Kwantyfikatory

W tym rozdziale skoncentrujemy się na zaproponowanych przez nas kwantyfikatorach. Prezentację rozpoczniemy od kwantyfikatorów względnych, aby następnie omówić kwantyfikatory bezwzględne.

W przypadku kwantyfikatorów, wykorzystywanymi przez nas funkcjami przynależności są zarówno funkcje trapezoidalne jak i funkcje trójkątne. Za każdym razem w tabelach podano, z jakiej funkcji skorzystano przy definicji danej etykiety, co ma swoje odzwierciedlenie na prezentowanych wykresach.

4.4.1. Kwantyfikatory względne

Wykres ilustrujący wszystkie kwantyfikatory względne, zamieszczono poniżej.



Rysunek 20. Kwantyfikatory względne

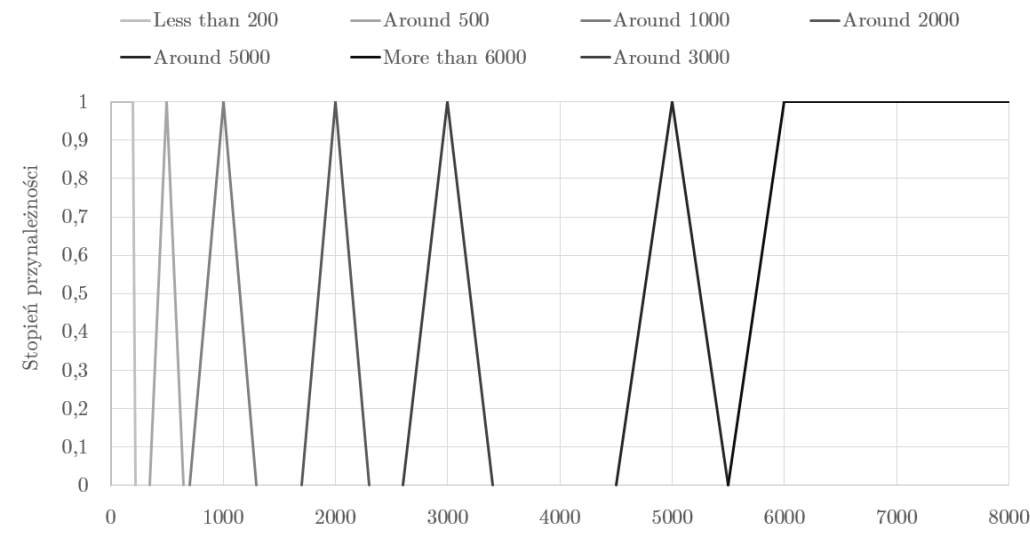
Tabela zawierająca parametry i nazwy funkcji przynależności, dla poszczególnych kwantyfikatorów, prezentuje się następująco.

Kwantyfikator	Funkcja przynależności	a	b	c	d
None	Trójkątna	0	0	0.1	-
Less than quarter	Trapezoidalna	0	0.1	0.25	0.3
Some	Trapezoidalna	0.15	0.2	0.35	0.4
Around one thirds	Trójkątna	0.23	0.33	0.43	-
Around half	Trapezoidalna	0.4	0.45	0.55	0.6
Around two thirds	Trójkątna	0.56	0.66	0.67	-
Majority	Trapezoidalna	0.7	0.75	0.85	0.9
Almost all	Trapezoidalna	0.85	0.9	1	1

Tabela 19. Funkcje przynależności kwantyfikatorów względnych - nazwy wraz z parametrami.

4.4.2. Kwantyfikatory bezwzględne

Wykres ilustrujący wszystkie kwantyfikatory bezwzględne, zamieszczono poniżej.



Rysunek 21. Kwantyfikatory bezwzględne

Tabela zawierająca parametry i nazwy funkcji przynależności, dla poszczególnych kwantyfikatorów, prezentuje się następująco.

Kwantyfikator	Funkcja przynależności	a	b	c	d
Less than 200	Trapezoidalna	0	0	200	220
Around 500	Trójkątna	350	500	650	-
Around 1000	Trójkątna	700	1000	1300	-
Around 2000	Trójkątna	1700	2000	2300	-
Around 3000	Trójkątna	2600	3000	3400	-
Around 5000	Trójkątna	4500	5000	5500	-
More than 6000	Trapezoidalna	5500	6000	8000	17000

Tabela 20. Funkcje przynależności kwantyfikatorów bezwzględnych - nazwy wraz z parametrami.

5. Wyniki

Praca w toku

6. Dyskusja

Praca w toku

7. Wnioski

Praca w toku

Literatura

- [1] Baza danych - "*Historical weather in the Netherlands 1901-2018*"