Politechnika Śląska

Wydział Mechaniczny Technologiczny

KATEDRA MECHANIKI I INŻYNIERII OBLICZENIOWEJ

ROZPRAWA DOKTORSKA

mgr inż. Adam Rurański

METAMODELOWANIE I OPTYMALIZACJA W PROJEKTOWANIU I TESTACH NUMERYCZNYCH WYBRANYCH ELEMENTÓW BATERII LITOWO-JONOWYCH

Promotor:

dr hab. inż. Wacław Kuś, prof. PŚ

Gliwice 2026

Tę pracę dedykuję mojemu synowi, Ireneuszowi Rurańskiemu.

W szczególności chciałbym podziękować:

dr inż. Przemysławowi Sebastjanowi za inspirację i motywację do podjęcia doktoratu.

dr inż. Antoniemu Skrobolowi za wsparcie merytoryczne i mentorskie,

dr inż. Barbarze Dołżańskiej za motywację, wsparcie oraz serdeczność,

dr hab. inż. Wacławowi Kusiowi, prof. PŚ, za pomoc merytoryczną oraz zaangażowanie,

mojej żonie, Katarzynie Rurańskiej, za nieustającą wiarę we mnie.

Ponadto dziękuję wszystkim, którzy okazali mi wsparcie i wiarę we mnie, a których nie sposób tutaj wszystkich wymienić.

Spis treści

Wykaz Symboli 6

Wykaz akronimów 13

Wykaz terminów 13

1. Wstęp 15

1.1. Struktura i zakres pracy 19

1.2. Zagrożenia i wyzwania związane z budową pakietów baterii litowo-jonowych. 20

1.3. Przegląd rozwiązań konstrukcyjnych obudów baterii litowo-jonowych 25

1.4. Motywacja badawcza oraz teza pracy 28

2. Wybrane metody symulacji i optymalizacji konstrukcji 30

2.1. Metoda elementów skończonych 30

2.1.1. Rozwiązywanie równań dynamiki MES 35

2.1.2. Modele plastyczności w zadaniach nieliniowych 39

2.1.3. Modele inicjacji i ewolucji zniszczenia 43

2.2. Wybrane metody metamodelowania 49

2.2.1. Przestrzeń projektowa oraz metody losowania próbek 52

2.2.2. Wielomianowe aproksymatory 57

2.2.3. Metoda Kriginga, Kriging 59

2.2.4. Sztuczne Sieci Neuronowe 62

2.2.5. Dobór parametrów metamodeli 67

2.3. Ewolucyjny Algorytm Inspirowany Zjawiskami Kwantowymi 70

3. Metoda optymalizacji obudów baterii 74

3.1. Sformułowanie problemu optymalizacji obudowy baterii 74

3.2. Metoda obliczania wartości funkcji celu 76

3.3. Bazy danych wyników symulacji numerycznych 79

3.4. Połączenie algorytmu optymalizacji QEA oraz MES 81

3.5. Połączenie algorytmu optymalizacji z uproszczonymi bazami danych 83

3.6. Metoda hybrydowa optymalizacji wykorzystująca algorytm optymalizacji oraz metamodelowanie 85

4. Przykład numeryczny optymalizacji obudowy baterii 91

4.1. Model numeryczny 91

4.2. Bazy danych zawierające wyniki analiz MES użyte do budowy metamodeli 97

4.3. Strojenie hiperparametrów oraz testy metamodeli 101

4.3.1. Hiperparametry Sztucznej Sieci Neuronowej 101

4.3.2. Dobór parametrów Kriging 112

4.3.3. Dobór parametrów regresji wielomianowej 116

4.4. Optymalizacja obudowy baterii z użyciem QEA 120

4.4.1. Optymalizacja obudowy z użyciem QEA bez metamodelowania 121

4.4.2. Optymalizacja z użyciem metody HYBRYDA1 124

4.4.3. Optymalizacja z użyciem metody HYBRYDA2 128

5. Wnioski 132

Bibliografia 136

Streszczenie 146

Abstract 147

Wykaz Symboli

$\mathbf{A}$ - macierz mapowania stopni swobody elementu na globalne stopnie swobody

a, b - współczynniki dobrane w celu równomiernego rozkładu wartości funkcji celu FF2

$\mathbf{A}_e$ - macierz mapowania dla elementu skończonego

$\mathbf{B}$ - macierz pochodnych funkcji kształtu

$\mathbf{b}$ - wektor sił masowych

b_0 - bias neuronu

$\mathbf{C}$ - macierz tłumienia

$\mathcal{C}$ - zbiór dopuszczalnych parametrów projektowych

c_{coh} - kohezja

c_e - prędkość rozchodzenia się fali mechanicznej w elemencie

D - skalarowa miara postępu inicjacji zniszczenia

$\ddot{\mathbf{d}}$ - wektor przyspieszeń uogólnionych

$\dot{\mathbf{d}}$ - wektor prędkości uogólnionych

$\mathbf{d}$ - wektor globalnych stopni swobody

$\mathbf{D}$ - macierz konstytutywna materiału

$\mathbf{d}_0$ - początkowy wektor przemieszczeń

d_{kl} - odległość między punktami w przestrzeni projektowej

E - moduł sprężystości materiału nominalnego

$\bar{E}$ - zredukowany moduł sprężystości materiału uszkodzonego

E^2 - błąd kwadratowy

F - funkcja odwzorowania modelu wysokiej wierności

$\mathbf{F}$ - globalny wektor sił zewnętrznych

$\mathcal{F}$ - odwzorowanie modelu wysokiej wierności

$f(\mathbf{P})$ - trend/deterministyczna część modelu Kriginga

$\mathbf{f}_{\text{int}}$ - wektor sił wewnętrznych

$\bar{\mathbf{g}}$ - zadany wektor trakcji na brzegu

$\mathbf{g}$ - wektor trakcji

g - przyspieszenie ziemskie

G_c - energia pękania

h - wysokość spadku

$i_{\max}$ - zadana liczba iteracji algorytmu optymalizacyjnego

i_{opt} - licznik aktualnej iteracji algorytmu optymalizacyjnego

I_x, I_y, I_z - momenty bezwładności mas zastępczych

J_0 - funkcja celu

J_1, J_2, J_3 - niezmienniki tensora naprężeń

$\mathbf{K}$ - macierz sztywności

$\mathbf{K}^{(e)}$ - macierz sztywności elementu skończonego

k_{DP} - parametr modelu Druckera-Pragera

$\mathbf{K}_T$ - styczna macierz sztywności

$\mathbf{L}$ - tensor naprężeń dewiatorowych

ℓ_e - charakterystyczna długość elementu skończonego

l_e - charakterystyczna długość elementu skończonego

M - macierz mas

m - funkcja metamodelu

$\mathbf{M}^{(e)}$ - macierz masowa elementu skończonego

m_{DB} - długość osobnika w algorytmie QEA

MSE_s - błąd średniokwadratowy dla zbioru s

N - macierz funkcji kształtu

n - wektor normalny do powierzchni

N_b - funkcja kształtu

N_{bazy} - liczebność bazy danych

n_{DOE} - liczba punktów planu eksperymentu

n_e - liczba elementów skończonych

N_{faza2} - minimalna liczebność bazy wymagana do przejścia do fazy hybrydowej

N_{new} - liczba nowych osobników od ostatniego uczenia metamodelu

N_s - liczebność zbioru s

N_{train} - liczba nowych osobników wymagana do aktualizacji metamodelu

P_{max} - maksymalna wartość parametru projektowego

P_{min} - minimalna wartość parametru projektowego

P - wektor parametrów projektowych

$\mathcal{P}$ - przestrzeń parametrów projektowych

$\bar{q}$ - średnia wartość odpowiedzi w analizowanym zbiorze

$\mathbf{q}$ - wektor wartości odpowiedzi referencyjnej

$\hat{q}$ - wartość predykcji metamodelu

q - wartość funkcji celu modelu wysokiej wierności

$\mathbf{q}_{\text{bit}}$ - wektor reprezentujący kubity osobnika w algorytmie QEA

$q_{\text{bit},j}$ - reprezentacja j-tego kubitu w algorytmie QEA

r - promień w płaszczyźnie dewiatorowej

R^2 - współczynnik determinacji

RMSE_s - pierwiastek błędu średniokwadratowego dla zbioru s

R^2_{PNR} - autorska miara jakości metamodelu regresji wielomianowej

R^2_{test} - współczynnik determinacji na zbiorze testowym

R^2_{train} - współczynnik determinacji na zbiorze treningowym

R^2_{val} - współczynnik determinacji na zbiorze walidacyjnym

$\mathbf{S}$ - operator odkształceń

$\mathbf{s}$ - tensor naprężeń dewiatorowych

SSE - suma kwadratów reszt

SST - całkowita suma kwadratów odchyleń

t - czas

t^0 - czas początkowy

t^n - czas w chwili n

$t^{n+\frac{1}{2}}$ - czas w kroku połówkowym

$\bar{\mathbf{u}}$ - zadane przemieszczenie na brzegu

$\ddot{\mathbf{u}}$ - wektor przyspieszeń

$\dot{\mathbf{u}}$ - wektor prędkości

$\mathbf{u}$ - wektor przemieszczeń

$\tilde{\mathbf{u}}_b$ - wektor przemieszczeń węzłowych

$\mathbf{v}_0$ - początkowy wektor prędkości

$\bar{\mathbf{v}}_0$ - wektor prędkości początkowych podczas spadku

V_{del} - objętość elementów usuniętych w wyniku spełnienia kryterium zniszczenia

$\mathbf{W}$ - macierz regresji

w_1 - graniczne rozwarcie szczeliny w całce energii pękania

w_{slot} - rozwarcie szczeliny

$\mathbf{x}$ - wektor położenia lub wektor wejściowy, zależnie od kontekstu równania

$Z(\mathbf{P})$ - funkcja kowariancji

α' - parametr modelu Druckera-Pragera

α_Q, β_Q - amplitudy superpozycji stanów w algorytmie QEA

α_{crit} - współczynnik bezpieczeństwa dla kroku czasowego

$\alpha_{\text{damp}}, \beta_{\text{damp}}$ - współczynniki tłumienia Rayleigha

$\boldsymbol{\beta}$ - wektor współczynników regresji wielomianowej

$\widehat{\boldsymbol{\beta}}$ - wektor oszacowanych współczynników regresji

n_β - liczba współczynników regresji wielomianowej

Γ - brzeg ograniczający obszar ciała

Γ_t - część brzegu z warunkiem Neumanna

Γ_u - część brzegu z warunkiem Dirichleta

γ - parametr lepkości

$\gamma(h)$ - semiwariogram w modelu Kriginga

$U(\Delta\theta_{Q,i})$ - macierz bramki obrotu w algorytmie QEA

ΔP - rozdzielczość zmiennej projektowej w kodowaniu QEA

Δt - krok czasowy

Δt^n - krok czasowy dla indeksu n

Δt_{crit} - krytyczny krok czasowy

$\Delta\theta_{Q,i}$ - przyrost kąta bramki obrotu w algorytmie QEA

$\delta\boldsymbol{\varepsilon}$ - wirtualne odkształcenie

$\delta\mathbf{u}$ - wirtualne przemieszczenie

$\boldsymbol{\varepsilon}$ - wektor odkształceń

$\boldsymbol{\varepsilon}^e$ - część sprężysta tensora odkształceń

$\boldsymbol{\varepsilon}^p$ - część plastyczna tensora odkształceń

ε_f - lokus zniszczenia

$\varepsilon_{p,\max}$ - maksymalne zredukowane odkształcenie plastyczne znalezione w analizie

$\varepsilon_{p,\max}^{\text{mat}}$ - maksymalne dopuszczalne odkształcenie plastyczne materiału

$\hat{\varepsilon}_f$ - funkcja odkształcenia granicznego, lokus zniszczenia

$\tilde{\varepsilon}$ - odkształcenie efektywne w opisie degradacji materiału

$\bar{\varepsilon}^p$ - zredukowane odkształcenie plastyczne

$\bar{\varepsilon}_f$ - zredukowane odkształcenie plastyczne przy zniszczeniu

η - współczynnik trójosiowości naprężenia

$\bar{\theta}$ - znormalizowany parametr kąta Lodego

θ - kąt Lodego

$Y(\kappa_{\text{izo}})$ - aktualna granica uplastycznienia zależna od wzmocnienia izotropowego

κ_{izo} - parametr wzmocnienia izotropowego

κ_{kin} - parametr wzmocnienia kinematycznego

$\dot{\lambda}$ - mnożnik plastyczny lub szybkość mnożnika plastycznego

λ_k - wartość własna dla trybu k

$\boldsymbol{\Xi}$ - zbiór parametrów modelu SSN

ξ - znormalizowana wartość trzeciego niezmiennika dewiatora naprężeń

Π - funkcja plastyczności lub funkcja nadaprężenia

ρ - gęstość

ϱ - funkcja aktywacji neuronu

$\bar{\sigma}$ - naprężenie zredukowane

$\boldsymbol{\sigma}$ - tensor naprężeń Cauchy'ego

σ_{dev} - miara części dewiatorowej naprężenia

σ_{slot} - funkcja rozwarcia szczeliny

σ_m - naprężenie hydrostatyczne

$\tilde{\sigma}$ - naprężenie efektywne w opisie degradacji materiału

$\boldsymbol{\Phi}_k$ - wektor własny dla trybu k

ϕ - funkcja lepkoplastyczna

ϕ_{fric} - kąt tarcia wewnętrznego

χ - miara ewolucji zniszczenia i degradacji materiału

$|\psi\rangle$ - stan superpozycji

Ω - obszar ciała

$\Omega^{(e)}$ - obszar elementu skończonego

ω_I^e - częstość własna elementu e

ω_{max} - maksymalna częstość układu liniowego

ω_k - częstość własna trybu k

Wykaz akronimów

MES - Metoda Elementów Skończonych

QEA - (ang. Quantum Evolutionary Algorithm) - Ewolucyjny Algorytm Inspirowany Zjawiskami Kwantowymi.

SSN - Sztuczne Sieci Neuronowe (ang. Artificial Neural Network)

PNR - (ang. Polynomial Regression) - Aproksymator Wielomianowy

KRG - metamodel Kriginga

TPE - (ang. Tree-structured Parzen Estimator) - przeszukiwanie bayesowskie

FF1 - Pierwszy wariant funkcji celu

FF2 - Drugi wariant funkcji celu

Wykaz terminów

Baza Full - baza danych zawierająca osobniki, w których doszło do zniszczenia, oraz w których nie doszło do zniszczenia

Baza NoDamage - baza danych powstała z *bazy Full* po usunięciu osobników, w których doszło do zniszczenia

Metamodel Full - metamodel stworzony na podstawie pełnej bazy danych z osobnikami ze zniszczeniem oraz bez zniszczenia

Metamodel NoDamage - metamodel stworzony na podstawie bazy danych z osobnikami tylko bez zniszczenia

LHS5000 - baza danych powstała z obliczeń MES z osobnikami wylosowanymi poprzez losowanie metodą LHS spośród 5000 próbek

LHSX - baza danych powstała w wyniku losowania LHS X próbek, po czym na podstawie tych próbek znaleziono najbliższe im odpowiadające odpowiedniki z Bazy LHS5000

QEA1687 - baza danych powstała z obliczeń MES z osobnikami zebranymi podczas kilku zadań QEA. Została poddana losowemu sortowaniu

QEAX - baza danych powstała w wyniku mieszania oraz obcinania zakresu bazy QEA1687 do X liczby próbek

1. Wstęp

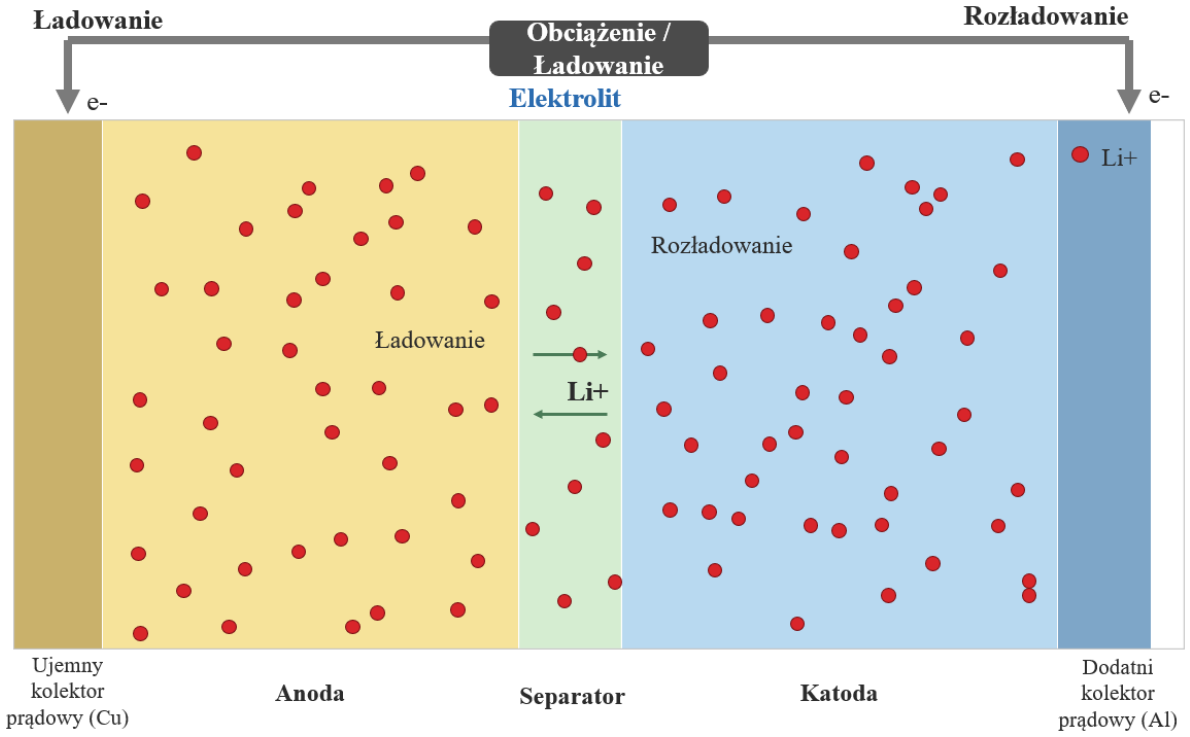
Dynamiczny rozwój technologii ogniw elektrochemicznych w ostatnich latach, w szczególności w obszarze wzrostu gęstości energii, umożliwił upowszechnienie rozwiązań, które wcześniej pozostawały niszowe lub były nieopłacalne z uwagi na ograniczenia masowo-gabarytowe [43]. Historycznie, wysoka masa i objętość baterii stanowiły barierę użytkową, natomiast niska gęstość energii skutkowała niewystarczającym czasem pracy urządzeń lub zasięgiem pojazdów. Wraz ze wzrostem gęstości energii parametry baterii przestały być dominującym ograniczeniem projektowym. Dzięki temu możliwe stało się wdrażanie produktów i funkcji, które wcześniej nie mogły się upowszechnić, ponieważ zbyt duża masa lub zbyt krótki czas pracy nie zapewniały wystarczającej wartości użytkowej i nie generowały popytu na aplikacje końcowe.

Geneza ogniw bateryjnych sięga przełomu XVIII i XIX wieku, gdy stos Volty zapewnił pierwsze praktyczne źródło prądu stałego. W kolejnych dekadach rozwijano bardziej stabilne ogniwa, w tym ogniwo Daniella z 1836 roku [4], które dostarczało stały prąd o większej stabilności niż ogniwo Volty i znalazło zastosowania techniczne. Kluczowym krokiem w kierunku nowoczesnego magazynowania energii było opracowanie pierwszego praktycznego akumulatora kwasowo-ołowiowego przez Plantégo w latach 1859-1860, co otworzyło drogę do powtarzalnego cyklu ładowanie-rozładowanie w zastosowaniach użytkowych [4]. Kamienie milowe w rozwoju baterii dobrze pokazują, że rozwój baterii był w dużej mierze odpowiedzią na potrzebę zwiększania pojemności, stabilności pracy oraz możliwości wielokrotnego wykorzystania źródła energii.

Przed upowszechnieniem ogniw litowo-jonowych (Li-ion) w zastosowaniach przenośnych dominowały akumulatory o elektrolitach wodnych, przede wszystkim kwasowo-ołowiowe oraz niklowe-kadmowe (NiCd), później niklowo-metalowo-wodorkowe (NiMH). Elektrolit wodny jest z natury niepalny, ale ma istotne ograniczenie fizykochemiczne: stabilne napięcie roztworów wodnych wynosi 1,2 V, co ogranicza napięcie pojedynczego ogniwa i utrudnia uzyskanie wysokiej gęstości energii bez dużej masy i objętości. Równolegle rozwijano akumulatory litowe z anodą z litu metalicznego, które obiecywały wysoką energię, lecz w praktyce ich problemem były dendryty litu, czyli nieregularne, igłowe lub włókniste struktury metalicznego litu, które mogą narastać na powierzchni elektrody podczas nierównomiernego osadzania litu [16], [72]. Dendryty stanowią zagrożenie, ponieważ mogą lokalnie uszkodzić separator, doprowadzić do zwarcia wewnętrznego, wzrostu temperatury, a w skrajnych przypadkach do pożaru [4], [75], [88].

Przełom w konstrukcjach ogniw litowo-jonowych polegał na tym, że podczas pracy ogniwa lit przemieszcza się między anodą i katodą w postaci jonów, natomiast elektrony płyną wyłącznie przez obwód zewnętrzny (rysunek 1.1). Dzięki temu nie ma potrzeby osadzania litu metalicznego na elektrodach, co ogranicza kluczowy mechanizm ryzyka typowy dla akumulatorów z anodą litową, czyli tworzenie dendrytów [72]. W elektrolicie nośnikiem ładunku są jony litu Li^+, natomiast elektrony e^- nie przechodzą przez elektrolit i przemieszczają się tylko w przewodniku zewnętrznym. W fazie ładowania źródło zewnętrzne wymusza przepływ elektronów e^- z elektrody dodatniej do ujemnej, a odpowiadające temu jony Li^+migrują przez separator z katody do anody. Po stronie anody jony Li^+ulegają wbudowaniu w materiał elektrody ujemnej (interkalacji), co prowadzi do magazynowania energii w postaci różnicy potencjałów chemicznych. W fazie rozładowania proces zachodzi w kierunku przeciwnym, tj. jony Li^+przemieszczają się z anody do katody przez elektrolit i separator. Jednocześnie elektrony e^- płyną przez odbiornik w obwodzie zewnętrznym, wykonując pracę elektryczną. Separator zapewnia przewodnictwo jonowe, a zarazem zapobiega zwarciu elektronowemu między elektrodami [50], [97]. Ta koncepcja, połączona z materiałami elektrod o wysokim napięciu pracy, umożliwiła uzyskanie znacznie większej gęstości energii niż w typowych akumulatorach wodnych, co bezpośrednio przełożyło się na miniaturyzację i mobilność urządzeń [4], [12]. Jednocześnie technologia litowo-jonowa bywa postrzegana jako bardziej wymagająca pod kątem bezpieczeństwa, ponieważ wysoka gęstość energii oraz obecność palnych elektrolitów organicznych powodują, że w rzadkich scenariuszach awarii termicznej możliwy jest gwałtowny przebieg reakcji spalania z emisją palnych gazów i ryzykiem pożaru. Zwykle nie obserwuje się takich zjawisk w tej samej formie w chemiach wodnych, choć te również bywają niebezpieczne, ze względu na możliwość wycieku żrącego elektrolitu, czy wydzielanie toksycznych gazów.

W praktyce rynkowym czasowym punktem odniesienia jest rozpoczęcie masowej sprzedaży pierwszych ładowalnych akumulatorów litowo-jonowych na początku lat 90-tych i kolejnych latach. Później technologia ta stała się podstawą zasilania elektroniki przenośnej (m.in. elektronarzędzia) i zastosowań w kolejnych segmentach przemysłu.



Rysunek 1.1 Zasada działania ogniwa litowo-jonowego.

Wpływ wzrostu gęstości energii na rozwój produktów i ukształtowanie rynku jest szczególnie widoczny w transporcie i elektronarzędziach. Koncepcja pojazdów elektrycznych pojawiła się już w XIX wieku, natomiast praktyczne wdrożenie zależało od ówczesnych parametrów baterii i nie spotkało się z sukcesem. Dopiero wdrożenie technologii baterii o odpowiedniej gęstości energii doprowadziło do spopularyzowania samochodów elektrycznych w ciągu ostatnich trzech dekad [5]. Współcześnie akumulatory litowo-jonowe są technologią wiodącą w elektronice użytkowej [12], pojazdach elektrycznych, rowerach elektrycznych oraz w szybko rosnącym segmencie magazynów energii. Równocześnie odegrały one kluczową rolę w rozwoju elektronarzędzi akumulatorowych. Użycie ogniw litowo-jonowych umożliwiło znaczące obniżenie masy, poprawę czasu pracy oraz lepszą charakterystykę użytkową w porównaniu z wcześniejszymi generacjami akumulatorów. W praktyce oznaczało to zmianę popularności z narzędzi przewodowych lub ciężkich rozwiązań akumulatorowych na narzędzia bezprzewodowe o wysokiej użyteczności w środowisku przemysłowym, warsztatowym i domowym. Wraz ze wzrostem liczby zastosowań wzrosła również liczba niebezpiecznych incydentów, co zwiększyło zainteresowanie organów normalizacyjnych oraz przełożyło się na bardziej restrykcyjne wymagania certyfikacyjne oraz testowe. Obecnie w większym stopniu akcentuje się bezpieczeństwo mechaniczne pakietów bateryjnych, w tym integralność konstrukcji oraz odporność na udary i np. powtarzalne upuszczenia, które są typowym scenariuszem

użytkowania w elektronarzędziach. Powyższe scenariusze testowe są wyzwaniem podczas procesu projektowania z uwagi na wysoki koszt testów laboratoryjnych oraz mniejszy, ale również znaczący koszt obliczeniowy symulacji, co stwarza potrzebę poszukiwania nowych metod obliczeniowych przyspieszających poszukiwanie rozwiązań konstrukcyjnych.

Równolegle do postępu w chemii i konstrukcji ogniw rozwijały się również metody obliczeniowe stosowane m.in. w analizie konstrukcji bateryjnych. Szczególne znaczenie w tym obszarze ma metoda elementów skończonych, która umożliwia ocenę odpowiedzi mechanicznej konstrukcji w warunkach obciążeń dynamicznych. W przypadku obudów baterii MES pozwala analizować rozkłady naprężeń, odkształceń oraz obszary potencjalnej inicjacji uszkodzeń. Modele inspirowane sieciami neuronowymi były formułowane już w połowie XX wieku, w tym perceptron Rosenblatta [54] jednak ich praktyczna użyteczność w zadaniach inżynierskich przez długi czas była ograniczona dostępnością mocy obliczeniowej oraz narzędzi numerycznych. Dopiero rozwój możliwości obliczeniowych oraz metod uczenia maszynowego umożliwił szerokie wykorzystanie modeli zastępczych do predykcji wyników kosztownych obliczeń numerycznych. W projektowaniu inżynierskim metamodelowanie jest stosowane jako sposób przybliżania odpowiedzi modeli wysokiej wierności, w tym modeli MES, co pozwala ograniczyć liczbę pełnych analiz wymaganych podczas eksploracji przestrzeni projektowej, analizy wrażliwości i optymalizacji [45], [71], [84], [93]. Obecnie badania nad kierunkiem rozwoju w tym obszarze są bardzo intensywne [19], [38], [48], [51], [58], [59], [60], [61], [62], [63], [65], [69], [82], [83], [92], pojawiają się próby wdrożenia metod opartych na sztucznej inteligencji w komercyjnych oprogramowaniach MES [107], [108], [109]. Mimo to w dalszym ciągu pojawia się potrzeba tworzenia dedykowanych narzędzi do celów specjalistycznych.

W konsekwencji, połączenie metod symulacyjnych z metamodelowaniem tworzy spójny kierunek rozwoju narzędzi projektowych dla systemów bateryjnych, gdzie koszty obliczeniowe i czas analiz stają się dużym wyzwaniem w procesie projektowym. Równolegle rosnące wymagania projektowe stawiane bateriom oraz rozrost mocy obliczeniowej komputerów, a co za tym idzie możliwość stosowania metod sztucznej inteligencji przyczyniło się do przedsięwzięcia prac nad tematem niniejszej dysertacji.

1.1. Struktura i zakres pracy

Rozdział pierwszy pełni funkcję wprowadzającą do tematyki pracy. W podrozdziale 1.2 przedstawiono budowę ogniw litowo-jonowych oraz wynikające z niej zagrożenia i wyzwania, które determinują wymagania projektowe wobec obudów. Następnie w 1.3 zaprezentowano przegląd rozwiązań konstrukcyjnych obudów, z naciskiem na podział na typy konstrukcji oraz konsekwencje tych wyborów dla odporności na obciążenia dynamiczne. W podrozdziale 1.4 sformułowano motywację badawczą oraz tezę pracy, wskazując lukę metodyczną związaną z kosztami obliczeń MES w zadaniach optymalizacji odporności na wielokrotne upadki. Podrozdział 1.4 domyka rozdział pierwszy, definiując zakres pracy oraz przepływ pomiędzy kolejnymi rozdziałami.

Rozdział drugi zawiera opis metod obliczeniowych i optymalizacyjnych wykorzystanych w pracy. W części 2.1 przedstawiono fundamenty metody elementów skończonych w ujęciu dynamiki, co obejmuje sformułowanie równania ruchu oraz jawny schemat całkowania w 2.1.1, kluczowy dla tego typu analiz stosowanych w symulacjach zderzeniowych i upadków. W 2.1.2 omówiono modele plastyczności, a w 2.1.3 modele inicjacji i ewolucji zniszczenia, ponieważ to właśnie nieliniowość materiałowa i degradacja nośności decydują o mechanizmach uszkodzeń obudowy podczas uderzeń. W części 2.2 wprowadzono metody metamodelowania, zaczynając od definicji przestrzeni projektowej oraz sposobów doboru punktów próbkowania w 2.2.1, a następnie opisując reprezentatywne klasy modeli zastępczych, czyli aproksymatory wielomianowe w 2.2.2, Kriging w 2.2.3 oraz sieci neuronowe w 2.2.4. Rozdział 2.3 domyka temat optymalizacyjny, przedstawiając ewolucyjny algorytm inspirowany zjawiskami kwantowymi jako narzędzie przeszukiwania przestrzeni projektowej w warunkach ograniczonej liczby kosztownych obliczeń funkcji celu.

Rozdział trzeci prezentuje autorską metodę optymalizacji obudów baterii, zorientowaną na praktyczne uruchomienie procesu obliczeniowego oraz kontrolę kosztu obliczeń. W 3.1 sformułowano problem optymalizacji, w tym postać zmiennych projektowych, ograniczeń oraz oczekiwaną rolę funkcji celu w ocenie odporności na wielokrotne upadki. W 3.2 przedstawiono sposób obliczania funkcji celu na podstawie wyników analiz numerycznych, tak aby wynik był jednoznaczny, porównywalny pomiędzy wariantami oraz możliwy do wykorzystania przez algorytm optymalizacyjny. W 3.3 opisano strukturę bazy danych wyników, która spina warstwę symulacji z warstwą uczenia modeli zastępczych oraz zaprezentowano podejście do strojenia parametrów, a następnie w 3.4 omówiono konfigurację oraz sposób działania algorytmu optymalizacyjnego w odniesieniu do zdefiniowanego zadania. Kolejne podrozdziały mają charakter wariantów realizacji optymalizacji, gdzie 3.5 dotyczy podejścia metamodel plus QEA opartego wyłącznie na predykcji modeli zastępczych, natomiast 3.6 opisuje metodę hybrydową,

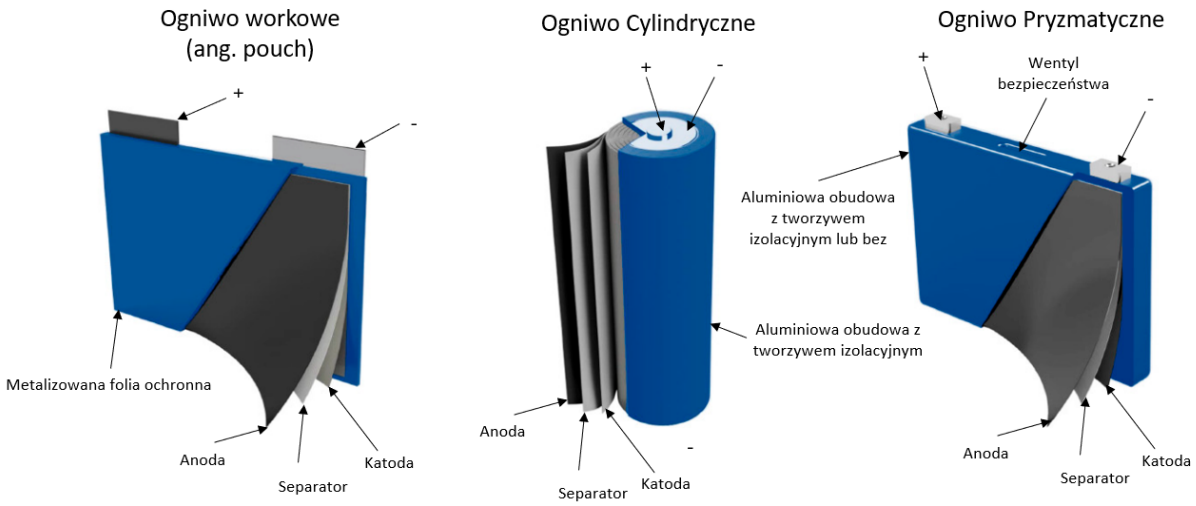
w której część osobników jest oceniana przez MES, a część przez metamodel, co ma na celu redukcję liczby analiz wysokiej wierności przy zachowaniu kontroli jakości rozwiązania.

Rozdział czwarty stanowi studium numeryczne, w którym cała zaproponowana metodyka jest weryfikowana na konkretnym przykładzie obudowy baterii. W 4.1 zdefiniowano model numeryczny wraz z kluczowymi założeniami modelowania, warunkami brzegowymi oraz sposobem odwzorowania scenariuszy obciążeń dynamicznych. W 4.2 scharakteryzowano zbiory danych wykorzystywane w procesie uczenia i walidacji metamodeli oraz w pętli optymalizacji. W 4.3 przedstawiono strojenie hiperparametrów i testy metamodeli, aby ilościowo ocenić ich przydatność w przewidywaniu funkcji celu w przestrzeni projektowej. Część 4.4 zawiera wyniki i wnioski z porównania strategii obliczeniowych, gdzie osobno zaprezentowano przebieg oraz efekty podejścia referencyjnego QEA z pełną ewaluacją funkcji celu metodą MES w 4.4.1 oraz wariantów hybrydowych w 4.4.2 i 4.4.3, a całość zamknięto wnioskami w 5 w odniesieniu do skuteczności, kosztu obliczeniowego i ograniczeń zaproponowanej metody.

Pracę zamyka rozdział 5, w którym zawarto zestawienie porównawcze metod oraz omawiane są wnioski.

1.2. Zagrożenia i wyzwania związane z budową pakietów baterii litowo-jonowych.

Istnieje kilka podstawowych typów konstrukcji ogniw litowo-jonowych (rysunek 1.2). W zastosowaniach związanych z elektromobilnością powszechnie spotyka się ogniwa pryzmatyczne. Równolegle szeroko stosowane są ogniwa cylindryczne, występujące w wielu rozmiarach, z których do najpopularniejszych należą formaty 18650 (około 18 mm średnicy i 65 mm długości) oraz 21700 (około 21 mm średnicy i 70 mm długości). Ponadto stosuje się również ogniwa w postaci tzw. worków (ang. pouch). Ze względu na ograniczoną odporność mechaniczną samego opakowania, rozwiązania pouch są zwykle integrowane z dodatkowymi elementami usztywniającymi lub zabudowywane w obudowach, które przejmują funkcję ochronną, nierzadko prowadząc w efekcie do konfiguracji zbliżonej funkcjonalnie do konstrukcji pryzmatycznej.



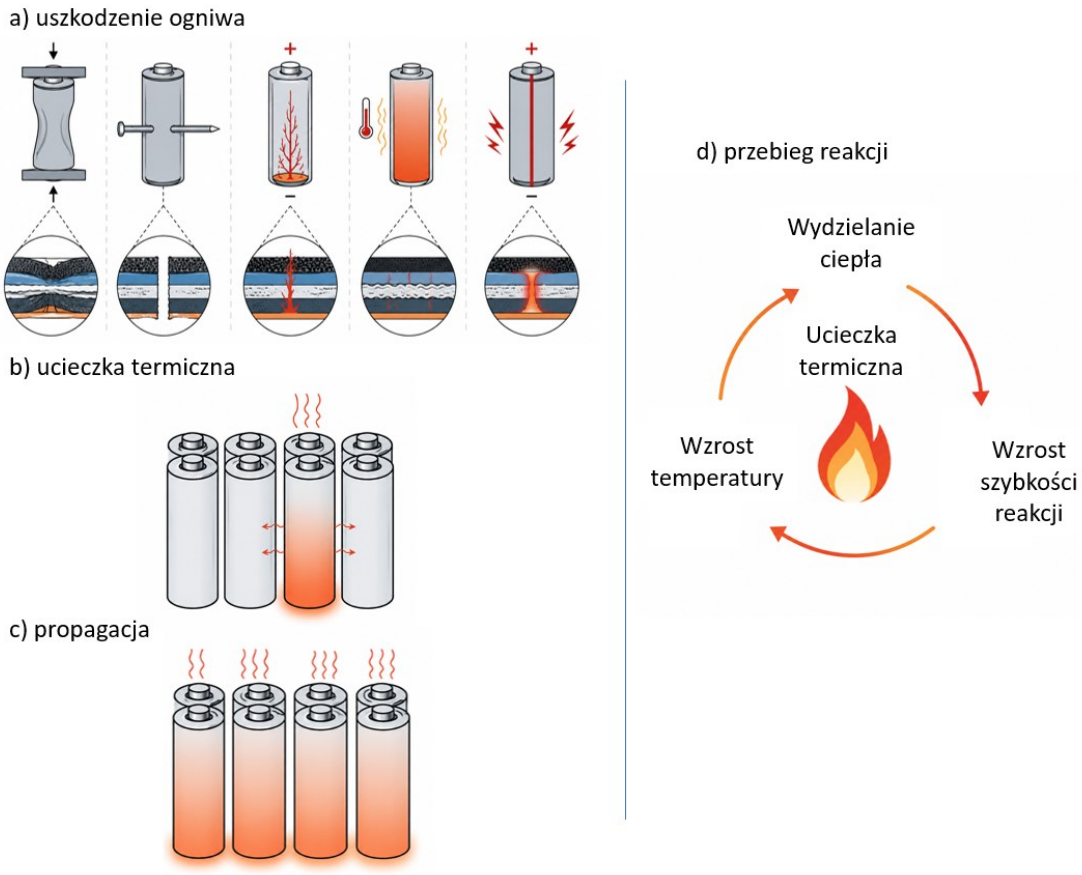
Rysunek 1.2 Rodzaje konstrukcji ogniw litowo-jonowych [52].

Niezależnie od typu konstrukcyjnego, ogniwa litowo-jonowe wykazują wspólne cechy wynikające z podobnej architektury wewnętrznej i doboru materiałów, co przekłada się na zbliżone mechanizmy awarii. Nie można pominąć konsekwencji wynikających z ich budowy, w szczególności w obszarze bezpieczeństwa. Podczas zdarzenia awaryjnego dochodzi do uwolnienia znacznej ilości palnych gazów, dlatego w ogniwach cylindrycznych oraz pryzmatycznych stosowane są wentyle bezpieczeństwa mające nie dopuścić do gwałtownego uwolnienia gazów (czyli wybuchu).

Istotnym aspektem bezpieczeństwa baterii litowo-jonowych jest zjawisko ucieczki termicznej [23], [24], [81], [91], [94]. Zjawisko to przedstawia rysunek 1.3. Inicjacja tego zjawiska może mieć kilka źródeł: elektryczne, takie jak przeładowanie, mechaniczne, w tym uszkodzenia prowadzące do zwarć wewnętrznych oraz termiczne, związane z pracą poza dopuszczalnym zakresem temperatur [22], [98], [101]. W literaturze i praktyce przemysłowej wskazuje się także czynniki wewnętrzne, wynikające z wad materiałowych [72] lub defektów produkcyjnych ogniwa. Ucieczka termiczna w ogniwie litowo-jonowym to proces wzrostu temperatury, w którym ciepło generowane wewnątrz ogniwa przekracza zdolność jego odprowadzania. Wraz ze wzrostem temperatury przyspieszają reakcje egzotermiczne i pogarszają się własności elementów wewnętrznych, co dodatkowo zwiększa wydzielanie ciepła. Skutkiem może być gwałtowne odgazowanie, wzrost ciśnienia, rozszczelnienie, a w skrajnych przypadkach zapłon lub eksplozja. Zjawisko może też propagować na sąsiednie ogniwa znajdujące się w baterii, jeśli ciepło i produkty reakcji przeniosą się do kolejnych ogniw.

Ograniczanie ryzyka ucieczki termicznej realizuje się równolegle na poziomie konstrukcji ogniwa, systemów zabezpieczeń elektrycznych (ang. Battery Management System, BMS) [18]

oraz rozwiązań konstrukcyjnych pakietu, w tym takich, które redukują prawdopodobieństwo krytycznych uszkodzeń mechanicznych.

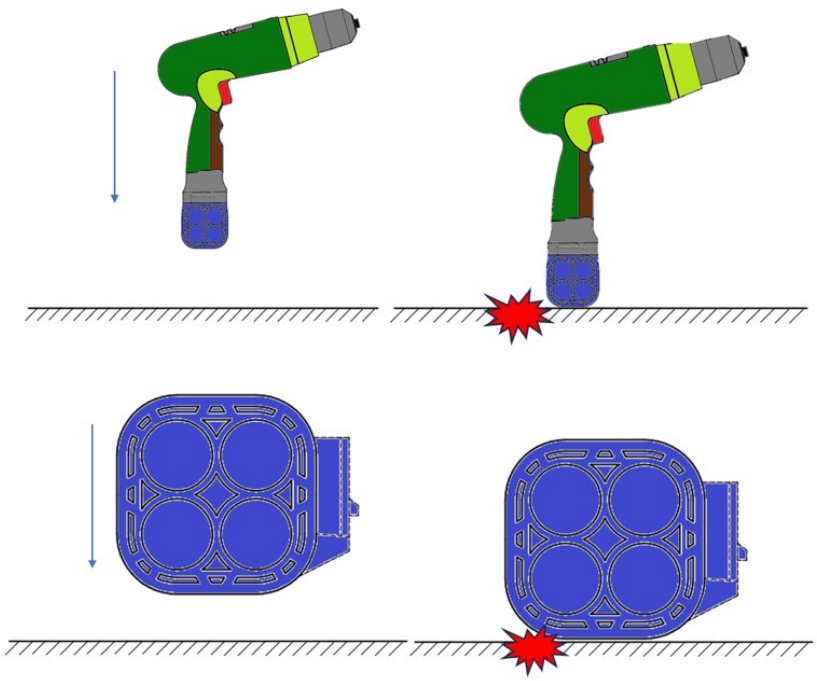


Rysunek 1.3 Zjawisko ucieczki termicznej (ang. thermal runaway) [85].

Z uwagi na specyfikę użytkowania elektronarzędzi, pakiety bateryjne są regularnie narażone na upadki oraz udary, czyli przypadkowe obciążenia dynamiczne. Dlatego najbardziej popularnymi ogniwami są ogniwa cylindryczne zapewniające dobry kompromis pomiędzy bezpieczeństwem a pojemnością. Pomimo tego ciągłość konstrukcyjna pakietu, rozumiana jako brak pęknięć, rozszczelnień oraz utraty zamocowania ogniw w obudowie, staje się krytycznym aspektem projektowym, ponieważ uszkodzenia mechaniczne mogą prowadzić nie tylko do degradacji funkcjonalnej, ale również do ucieczki termicznej. Na poziomie pojedynczego ogniwa mechaniczne odkształcenie może prowadzić do wewnętrznego zwarcia, co stanowi jeden z bezpośrednich mechanizmów inicjacji zdarzeń awaryjnych [79].

Przedmiotem niniejszej pracy są akumulatory litowo-jonowe stosowane w elektronarzędziach, ze szczególnym uwzględnieniem odporności konstrukcji oraz obudów na obciążenia udarowe i scenariusze testów zrzutowych.

W reakcji na incydenty pożarowe szeroko nagłaśniane w przestrzeni publicznej [14], [89], [106], [110] zaostrzono regulacje oraz standardy branżowe dotyczące zarówno transportu, jak i użytkowania baterii w aplikacjach końcowych [49]. W praktyce łańcucha dostaw podstawowym wymaganiem w obszarze transportu jest podsekcja normy transportowej UN38.3 [104]. Obejmuje ona m.in. próby mechaniczne symulujące warunki przewozu, w tym wibracje oraz wstrząs mechaniczny. Kryteria akceptacji testu koncentrują się na braku stanów niebezpiecznych (np. pożar, rozszczelnienie, rozerwanie) oraz na utrzymaniu parametrów elektrycznych w granicach wymagań po teście. W dalszej kolejności normalizacja obejmuje środowisko pracy i aplikacje końcowe. Dla obciążeń mechanicznych w warunkach pojazdowych stosuje się normę ISO 16750-3 [31], [41], która opisuje potencjalne obciążenia środowiskowe i zalecane badania dla urządzeń elektrycznych i elektronicznych w zależności od miejsca ich montażu. Obejmuje ona m.in. obciążenia o charakterze drgań i udarów. Równolegle funkcjonują normy bezpieczeństwa dla przenośnych ogniw i baterii, takie jak norma IEC 62133-2 [17]. Norma ta definiuje wymagania i zestaw prób dla bezpiecznej pracy w warunkach użytkowania i przewidywalnego niewłaściwego użycia. W obszarze mechanicznym obejmuje m.in. swobodny upadek, wibracje oraz wstrząsy mechaniczne, a ocenę opiera się na kryterium braku zdarzeń niebezpiecznych. Dla samego poziomu ogniwa, w kontekście kwalifikacji bezpieczeństwa używa się również normy bezpieczeństwa UL 1642 [17], gdzie w obszarze mechanicznym definiuje się m.in. próby zgniatania, udaru, wstrząsu i wibracji.



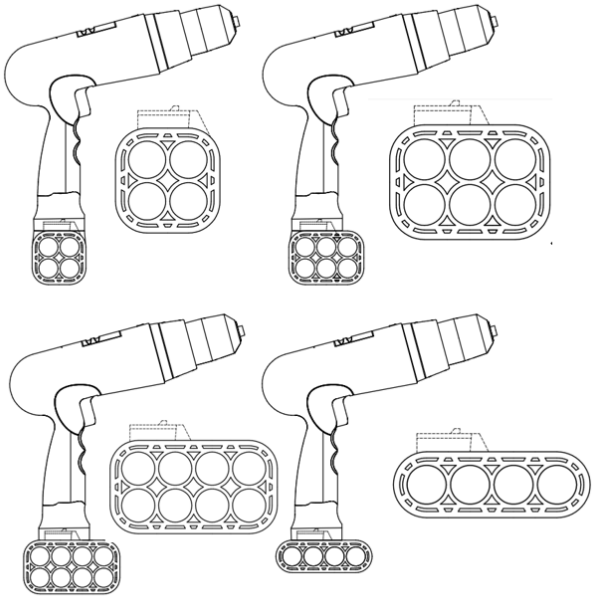
Rysunek 1.4 Test upuszczenia elektronarzędzia oraz samej obudowy ogniw.

Niezależnie od norm, w praktyce wielu producentów urządzeń (w tym elektronarzędzi) stosuje także wewnętrzne procedury kwalifikacyjne, często ukierunkowane na specyficzne scenariusze użytkowe, takie jak wielokrotne upuszczenia w zdefiniowanych orientacjach. Jeden z testów dla klienta przemysłowego obejmuje serię aż 90 zrzutów (rysunek 1.4.) Test polega na trzykrotnym zrzucie baterii z każdej z 10 zdefiniowanych wysokości (od 30 cm do 210 cm - co 20 cm) na każdą z trzech ściśle zdefiniowanych stron (prawa, przednia, lewa - patrząc od przodu na wrzeciona elektrycznej wiertarki). Dla poszczególnych przedziałów wysokości są przyjęte różne kryteria pozytywnej weryfikacji. Kryterium A nie dopuszcza żadnych uszkodzeń oraz utraty funkcjonalności, kryterium B dopuszcza delikatne uszkodzenia wewnętrzne, w tym słyszalne dźwięki podczas poruszania obudowy. Kryterium C dopuszcza utratę funkcjonalności urządzenia, jednak nie dopuszcza zniszczeń potencjalnie niebezpiecznych typu utrata integralności obudowy - brak szczelności. Po każdym zrzucie w tabeli zaznaczane jest spełnienie kryteriów. Przeprowadzenie tak rozbudowanego testu fizycznego jest kosztowne, ponieważ wymaga przygotowania prototypów, stanowiska badawczego oraz czasochłonnej procedury obejmującej wiele powtarzalnych zrzutów. Z perspektywy projektowej stanowi to wyjątkowo trudne zadanie, ponieważ przewidzenie skutków wielokrotnych upadków już na etapie projektu, przed wykonaniem fizycznych prototypów, jest obarczone wysoką niepewnością. Symulacja wielokrotnych upadków metodami numerycznymi wiąże się z bardzo wysokimi kosztami obliczeniowymi, a proces iteracyjnej optymalizacji, wymagany do uzyskania najlepszych rozwiązań konstrukcyjnych, staje się w praktyce trudny i czasochłonny [28], [39]. Z tego powodu istotnym problemem badawczym jest opracowanie podejścia, które pozwoli efektywnie oceniać i optymalizować konstrukcję obudowy przy zachowaniu adekwatnego poziomu wiarygodności predykcji w warunkach wielokrotnych udarów.

Należy podkreślić, że obudowa pakietu bateryjnego pełni równocześnie szereg funkcji wykraczających poza samą ochronę przed skutkami upadków. W praktyce konstrukcja musi ograniczać możliwość przemieszczania się ogniw i elementów wewnętrznych podczas wibracji oraz zapewniać warunki sprzyjające odprowadzaniu ciepła, na przykład przez odpowiednio zaprojektowany przepływ powietrza lub inne rozwiązania wspierające chłodzenie. W konsekwencji projekt obudowy jest zadaniem wielowątkowym, w którym odporność na obciążenia mechaniczne jest jednym z wymagań, ale musi być rozpatrywana łącznie z ograniczeniami funkcjonalnymi i bezpieczeństwa. Podczas procesu projektowania należy uwzględnić również ograniczenia technologiczne wynikające z procesu wytwarzania elementów obudowy. Istotna jest również masa, a każda obudowa musi być szczelna z uwagi na ryzyko zwarcia.

1.3. Przegląd rozwiązań konstrukcyjnych obudów baterii litowo-jonowych

W niniejszej pracy skoncentrowano się na elektronarzędziach. W tego rodzaju urządzeniach szczególnie rozpowszechnione są rozwiązania modułowe (rysunek 1.5) o wspólnym interfejsie mechaniczno-elektrycznym, w których użytkownik dobiera pakiet z tej samej rodziny kompatybilnych rozwiązań, ale o różnej pojemności, zależnie od zapotrzebowania na czas pracy i moc.

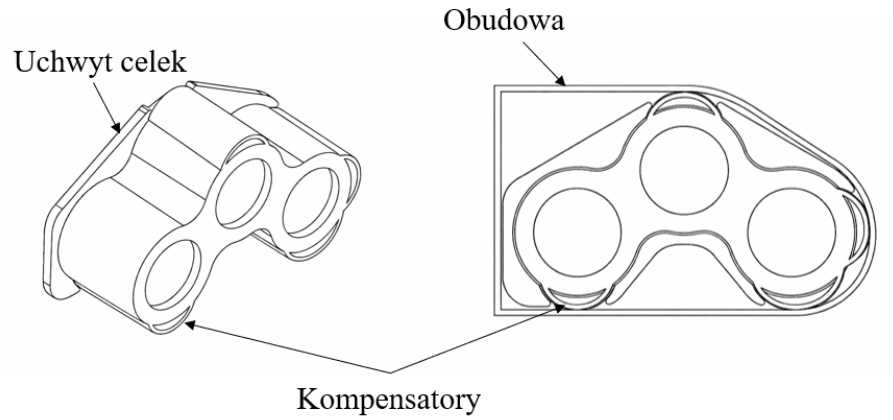


Rysunek 1.5 Przykład modułowej konstrukcji baterii do elektronarzędzi.

W rodzinie baterii do elektronarzędzi, w oparciu o rozwiązania projektowe stosowane w przedsiębiorstwie, można wyróżnić dwie główne gałęzie konstrukcyjne obudów: wieloelementową (rysunek 1.7) oraz integralną (rysunek 1.8). W praktyce najczęściej spotykanym rozwiązaniem jest obudowa wieloelementowa, w której kluczowe funkcje są rozdzielone pomiędzy uchwyt ogniw i obudowę zewnętrzną.

Pierwszym zasadniczym elementem jest uchwyt ogniw (ang. cellholder) (rysunek 1.6), którego zadaniem jest ustalenie położenia ogniw względem siebie oraz zapewnienie stabilnego mocowania elementów elektronicznych związanych z pakietem, w szczególności płytki sterującej i połączeń elektrycznych. Drugą istotną funkcją uchwytu jest zapewnienie kontrolowanego dopasowania do obudowy zewnętrznej poprzez rozwiązania kompensujące tolerancje wymiarowe i montażowe (rysunek 1.6). Realizowane jest to za pomocą elementów

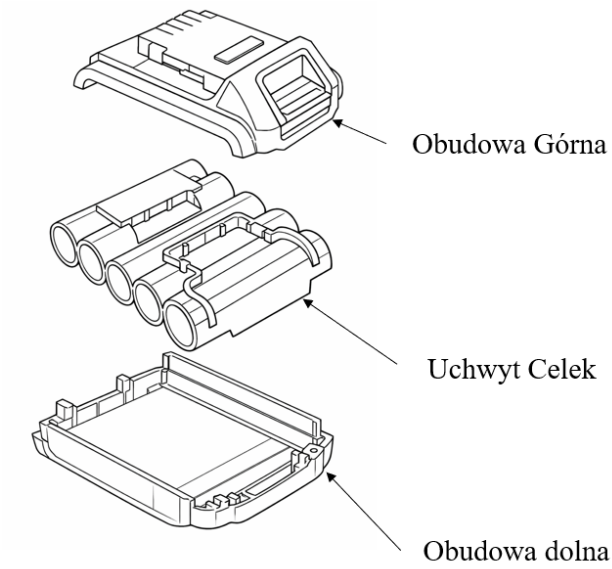
kompensujących zaprojektowanych albo na uchwycie ogniw, albo na powierzchniach wewnętrznych obudowy zewnętrznej. Taka kompensacja ogranicza luzy montażowe, poprawia dopasowanie elementów oraz zmniejsza ryzyko przemieszczania się ogniw i wkładu wewnętrznego względem obudowy podczas obciążeń dynamicznych, takich jak drgania i udary.



Rysunek 1.6 Przykład konstrukcji uchwytu ogniw (ang. cellholder). Rysunek z patentu autora pracy [53].

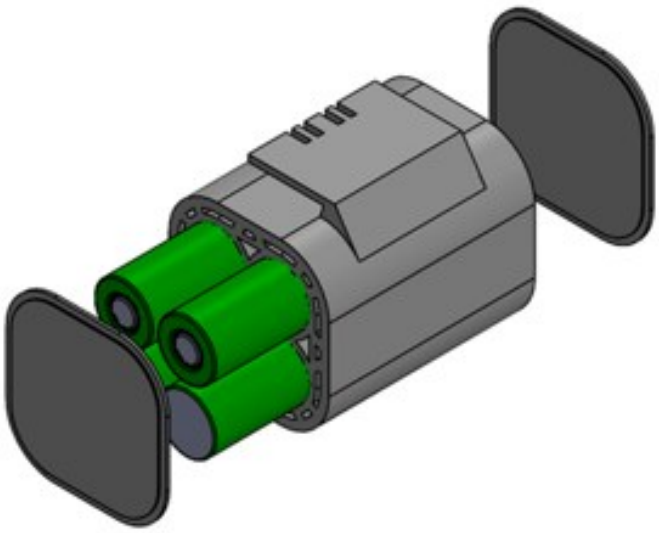
Drugim elementem jest obudowa zewnętrzna, która pełni funkcję osłonową. Od strony mechanicznej zapewnia ochronę przed czynnikami środowiskowymi i oddziaływaniami zewnętrznymi, w tym uszkodzeniami mechanicznymi, a także współuczestniczy w realizacji wymagań dotyczących szczelności pakietu. Najczęściej obudowa zewnętrzna ma postać konstrukcji dwuczęściowej (część górna i dolna), jednak spotyka się również warianty o podziale bocznym, zależnie od wymagań montażowych i uwarunkowań geometrycznych konkretnej platformy akumulatorowej.

Pierwszą gałąź konstrukcji obudów stanowią obudowy wieloelementowe (rysunek 1.7). Ten rodzaj obudowy jest wyjątkowo korzystny od strony produkcyjnej z uwagi na możliwość podzielenia operacji montażu na kilka osobnych etapów, jednak od strony odporności mechanicznej posiada pewne ograniczenia.



Rysunek 1.7 Przykład konstrukcji wieloelementowej z podziałem poziomym.

Drugą gałęzią konstrukcyjną są obudowy integralne, w których funkcje uchwytu ogniw oraz obudowy zewnętrznej zostały połączone w jeden element konstrukcyjny. Oznacza to, że geometria obudowy nie tylko pełni rolę osłony przed oddziaływaniami zewnętrznymi, lecz także bezpośrednio pozycjonuje i unieruchamia ogniwa oraz komponenty elektroniczne. Taka architektura zwykle stawia wyższe wymagania montażowe, ponieważ dokładność wykonania detali ustalających oraz sekwencja montażu muszą zapewnić poprawne osadzenie ogniw bez możliwości korekty poprzez osobny uchwyt celek.



Rysunek 1.8 Przykład konstrukcji integralnej.

Z punktu widzenia obciążeń dynamicznych obudowa integralna (rysunek 1.8) może oferować korzystniejsze warunki przenoszenia obciążeń, ponieważ ogranicza liczbę interfejsów i płaszczyzn styku pomiędzy elementami, a tym samym zmniejsza ryzyko powstawania luzów montażowych oraz przemieszczeń względnych podczas drgań, udarów i upadków. W wielu rozwiązaniach eliminuje się potrzebę stosowania odrębnej kompensacji dopasowania pomiędzy uchwytem ogniw a obudową zewnętrzną, co upraszcza problem kontroli przesuwania się uchwytu wewnętrznego ogniw względem osłony. Dodatkowo w tego typu konstrukcjach często projektuje się elementy ochronne w postaci odbojników lub stref kontrolowanego zgniotu, których zadaniem jest ograniczenie nagłych skoków przyspieszeń i poprawa pochłaniania energii w trakcie uderzeń, co sprzyja zachowaniu integralności mechanicznej pakietu.

W obudowach integralnych elementami dodatkowymi są zwykle tylko pokrywy lub klapy boczne, które zamykają pakiet, wspierają wymagania szczelności oraz stabilizują ogniwa w założonym położeniu. W praktyce dobór tej architektury jest kompromisem pomiędzy łatwością montażu i serwisowania, oczekiwaną sztywnością, a odpornością na wymuszenia dynamiczne.

W obu opisanych gałęziach konstrukcyjnych (wieloelementowej i integralnej) dominują tworzywa termoplastyczne przetwarzane metodą wtrysku, ponieważ pozwalają one integrować w jednym elemencie funkcje mechaniczne (żebra usztywniające, gniazda śrub, prowadzenia, zatrzaski, gniazda uszczelek, strefy kontrolowanego ugięcia) przy relatywnie niskiej masie i kosztach wytwarzania wielkoseryjnego. Dla obudów i elementów ustalających najczęściej używa się poliamidów wzmacnianych włóknem szklanym (np. PA6-GF30, PA66-GF), polimery ABS oraz mieszaniny PC+ABS, również PBT-GF lub PP, zależnie od wymagań środowiskowych i norm palności.

1.4. Motywacja badawcza oraz teza pracy

Motywacją do podjęcia pracy jest praktyczny i jednocześnie metodyczny problem projektowania obudów i integracji baterii litowo-jonowych z uwzględnieniem odporności na oddziaływania mechaniczne wynikające z typowych scenariuszy użytkowania i obsługi. W wielu zastosowaniach, w tym w elektronarzędziach, urządzeniach przenośnych oraz w systemach o podwyższonych wymaganiach bezpieczeństwa, zdarzenia upadku nie są incydentami jednostkowymi, lecz mogą powtarzać się wielokrotnie w cyklu życia wyrobu. W konsekwencji o rzeczywistej niezawodności nie decyduje wyłącznie odpowiedź konstrukcji na pojedynczy upadek, ale także kumulacja uszkodzeń i degradacja nośności w sekwencji kolejnych uderzeń. Tego typu zjawiska są silnie nieliniowe, zależne od kontaktu,

lokalnych odkształceń plastycznych, geometrii oraz warunków brzegowych, przez co ich wiarygodna ocena wymaga złożonych symulacji numerycznych, oraz odpowiednio zdefiniowanych miar uszkodzenia. Jednocześnie klasyczne podejście optymalizacyjne oparte wyłącznie na obliczeniach MES dla każdego wariantu projektowego jest mało efektywne obliczeniowo. Optymalizacja globalna wymaga dużej liczby ewaluacji funkcji celu, a w przypadku testów wielokrotnych upadków każda pojedyncza ewaluacja jest kosztowna, bo obejmuje sekwencję zdarzeń dynamicznych, możliwe zmiany kontaktu i propagację stref uplastycznienia lub uszkodzeń. W praktyce prowadzi to do kompromisów: upraszczania modelu, ograniczania liczby wariantów lub redukowania problemu do pojedynczego scenariusza obciążenia, co osłabia zarówno kompletność analizy, jak i jakość decyzji projektowych. Stąd potrzeba opracowania metody obliczeniowej, która pozwoli prowadzić optymalizację projektową dla sekwencji upadków w sposób wykonalny czasowo, przy zachowaniu spójności z ograniczeniami konstrukcyjnymi i wymogami formalnymi stawianymi wyrobom.

W niniejszej pracy przyjęto podejście ukierunkowane na możliwość praktycznego zastosowania wyników w procesie projektowym. W środowisku przemysłowym wymagane jest iteracyjne porównywanie wielu wariantów geometrii i parametrów konstrukcyjnych, jednak w przypadku wielokrotnych upadków koszt obliczeniowy pełnej oceny MES pojedynczego wariantu istotnie ogranicza liczbę możliwych iteracji. W efekcie projektowanie jest często upraszczane do pojedynczego zdarzenia lub zawężane do niewielkiego zbioru rozwiązań, co nie odpowiada rzeczywistym warunkom użytkowania baterii litowo-jonowych, w tym w elektronarzędziach. Zasadniczym celem metodycznym jest więc opracowanie takiej procedury obliczeniowej, która umożliwi efektywne przeszukiwanie przestrzeni projektowej dla sekwencji upadków, redukując liczbę koniecznych pełnych symulacji przy jednoczesnym utrzymaniu jakości oceny i zgodności z ograniczeniami projektowymi.

Teza badawcza pracy została sformułowana następująco:

Zastosowanie hybrydowej metody optymalizacji, łączącej kwantowo inspirowany algorytm optymalizacji, modelowanie MES oraz metamodelowanie, umożliwia ograniczenie nakładu obliczeniowego optymalizacji wybranych elementów konstrukcyjnych baterii litowo-jonowych poddawanych obciążeniom dynamicznym, przy zachowaniu poprawności wnioskowania dotyczącego ich odpowiedzi mechanicznej i rozwoju uszkodzeń.

2. Wybrane metody symulacji i optymalizacji konstrukcji

W pracy rozważany jest proces projektowania, w którym jakość konstrukcji ocenia się na podstawie symulacji numerycznych. W takim ujęciu kluczową rolę pełni metoda elementów skończonych (MES) [8], [9], [35], [102], [103], ponieważ umożliwia modelowanie zjawisk nieliniowych, w tym nieliniowości materiałowej, geometrycznej wynikającej z kontaktu, a wyniki można interpretować w kategoriach wielkości mechanicznych, takich jak przemieszczenia, odkształcenia i naprężenia.

Koszt obliczeniowy analiz o wysokiej wierności rośnie szybko wraz ze wzrostem liczby stopni swobody, liczby wariantów geometrycznych oraz złożoności przyjętego modelu materiałowego. W praktyce projektowej prowadzi to do ograniczenia liczby możliwych do wykonania analiz. Z tego powodu można zastosować metamodelowanie, czyli budowę modeli zastępczych aproksymujących zależność wejście-wyjście na podstawie próbek z analiz wysokiej wierności (analiz MES). Metamodel może przejąć rolę funkcji celu, redukując liczbę kosztownych uruchomień pełnego modelu MES podczas optymalizacji.

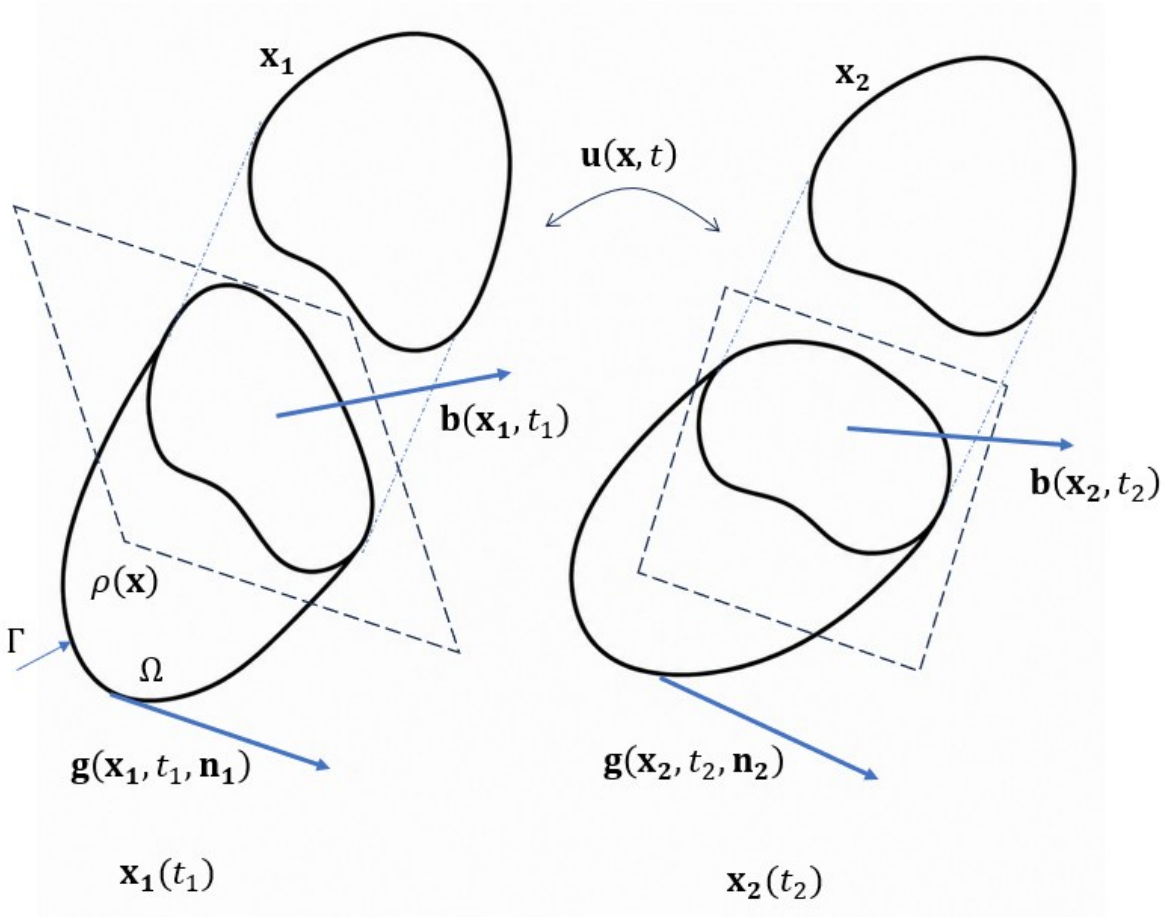
Ostatecznie, system hybrydowy obejmuje trzy współpracujące komponenty: model MES generujący dane referencyjne, metamodel przewidujący odpowiedź konstrukcji dla nowych punktów w przestrzeni projektowej oraz algorytm optymalizacyjny sterujący poszukiwaniem rozwiązania. W niniejszym rozdziale przedstawiono podstawy teoretyczne każdego z tych elementów.

2.1. Metoda elementów skończonych

Metoda elementów skończonych jest podejściem numerycznym przeznaczonym do rozwiązywania zadań mechaniki ośrodka ciągłego w przypadkach, w których ze względu na złożoną geometrię, warunki brzegowe, zmienność materiału lub nieliniowość nie da się uzyskać rozwiązania analitycznego. Jej podstawy rozwijały się od lat 40. XX wieku, natomiast jako metoda inżynierskiej analizy konstrukcji MES została ukształtowana głównie w latach 50. i 60. XX wieku[8], [9], [33], [35], [102], [103]. Metoda powstała i była rozwijana najpierw jako narzędzie inżynierskie do analizy konstrukcji, a następnie została uogólniona na inne klasy problemów. Współcześnie MES jest podstawowym narzędziem do analizy statycznej i dynamicznej, a typowy proces modelowania i obliczeń obejmuje kolejne kroki od idealizacji problemu fizycznego, przez model matematyczny, model obliczeniowy, po interpretację wyników.

W niniejszej pracy rozpatruje się przede wszystkim przypadki obciążeń dynamicznych, czyli takich, w których ciało może poruszać się i odkształcać w czasie, a istotną rolę odgrywa bezwładność. W praktyce oznacza to, że równowaga nie jest rozumiana jako brak ruchu, lecz jako zgodność pomiędzy działającymi siłami i uzyskiwanym przyspieszeniem. W dalszych rozważaniach przyjmuje się hipotezę małych przemieszczeń i małych odkształceń, co pozwala na zastosowanie liniowej miary odkształcenia oraz klasycznej relacji kinematycznej pomiędzy przemieszczeniami a odkształceniami.

Rozważane ciało zajmuje obszar $\Omega \subset \mathbb{R}^3$ o brzegu ograniczającym bryłę Γ (rysunek 2.1). Wektor położenia oznaczono przez $\mathbf{x}$, a czas przez t. Podstawową niewiadomą jest pole przemieszczeń $\mathbf{u}(\mathbf{x}, t)$. Pole przemieszczeń można interpretować jako wektor $\mathbf{u}(\mathbf{x}, t)$, o który przesunął się punkt w chwili czasowej t. Gęstość ciała oznaczono przez $\rho(\mathbf{x})$. Obciążenia objętościowe (siły masowe) opisuje wektor $\mathbf{b}(\mathbf{x}, t)$. Na brzegu mogą działać obciążenia powierzchniowe $\mathbf{g}(\mathbf{x}, t, \mathbf{n})$.



Rysunek 2.1 Siły wewnętrzne i zewnętrzne w ciele Ω[33].

Tensor naprężeń Cauchy'ego $\boldsymbol{\sigma}(\mathbf{x}, t)$, opisuje stan sił wewnętrznych w ciele. Z tensora $\boldsymbol{\sigma}$ wyznacza się trakcję $\mathbf{g}(\mathbf{x}, t, \mathbf{n})$, czyli wektor siły powierzchniowej na jednostkę pola, działający na powierzchnię o normalnej $\mathbf{n}$. Holzapfel podaje twierdzenie Cauchy'ego o naprężeniach w postaci relacji liniowej pomiędzy trakcją i wektorem normalnym do powierzchni [33]:

$$\mathbf{g}(\mathbf{x}, t, \mathbf{n}) = \boldsymbol{\sigma}(\mathbf{x}, t)\, \mathbf{n} \tag{2.1}$$

Tensor $\boldsymbol{\sigma}$ pozwala wyznaczyć siłę powierzchniową działającą na dowolnie zorientowaną płaszczyznę przechodzącą przez punkt ciała.

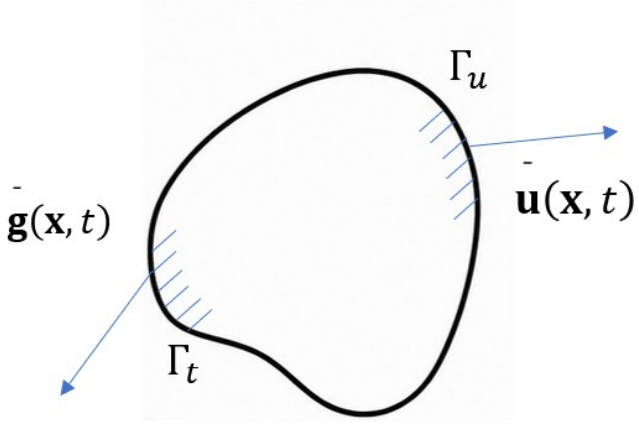
Równanie równowagi w objętości Ω jest podstawowym równaniem ruchu w mechanice ośrodka ciągłego [102]:

$$\sigma_{ji,j} + b_i = \rho\, \ddot{u}_i, i, j = 1,2,3 \tag{2.2}$$

gdzie σ_{ji} to składowe tensora naprężeń Cauchy'ego, b_i to składowe sił masowych, ρ to gęstość, a $\ddot{u}_i$ to składowe przyspieszenia. Tensor naprężeń jest symetryczny:

$$\sigma_{ij} = \sigma_{ji} \tag{2.3}$$

Warunki brzegowe formułuje się przez podział brzegu Γ na część Γ_u, na której narzuca się przemieszczenia (warunek Dirichleta), oraz część Γ_t, na której narzuca się trakcje (warunek Neumanna) (rysunek 2.2).



Rysunek 2.2. Podział brzegu na część z zadanymi przemieszczeniami oraz część z zadanymi trakcjami [102].

Wektor trakcji $\mathbf{g}$ wynika z działania tensora naprężenia Cauchy'ego σ na powierzchnię o normalnej zewnętrznej n. W zapisie indeksowym zależność tę można zapisać jako:

$$g_i = \sigma_{ji}\, n_j \tag{2.4}$$

gdzie g_i oznacza składowe wektora trakcji, σ_{ji} oznacza składowe tensora naprężenia Cauchy'ego, a $n_j$oznacza składowe jednostkowego wektora normalnego zewnętrznego do brzegu. Powtarzający się indeks j oznacza sumowanie po składowych przestrzennych.

Warunek Neumanna określa zadane obciążenia powierzchniowe na części brzegu Γ_t:

$$\boldsymbol{\sigma}(\mathbf{x}, t)\, \mathbf{n} = \bar{\mathbf{g}}(\mathbf{x}, t) \quad \text{na } \Gamma_t \tag{2.5}$$

gdzie $\bar{\mathbf{g}}(\mathbf{x}, t)$ oznacza zadany wektor obciążenia powierzchniowego. Warunek Dirichleta określa zadane przemieszczenia na części brzegu Γ_u:

$$\mathbf{u}(\mathbf{x}, t) = \bar{\mathbf{u}}(\mathbf{x}, t) \quad \text{na } \Gamma_u \tag{2.6}$$

Równanie ruchu w obszarze Ω (2.2), warunki brzegowe (2.5 oraz 2.6) tworzą opis problemu brzegowo-początkowego dla rozważanego zagadnienia mechaniki ośrodka ciągłego.

Wprowadzamy wirtualne przemieszczenia δu_i oraz odpowiadające im wirtualne odkształcenia $\delta \varepsilon_{ij}$:

$$\delta \varepsilon_{ij} = \frac{1}{2}\left(\frac{\partial\, \delta u_i}{\partial x_j} + \frac{\partial\, \delta u_j}{\partial x_i}\right) \tag{2.7}$$

gdzie δu_i oraz δu_j oznaczają składowe pola przemieszczeń wirtualnych, $\delta \varepsilon_{ij}$ jest składową liniowego tensora odkształceń wirtualnych, x_i oraz x_j oznaczają składowe wektora położenia $\mathbf{x}$, natomiast indeksy $i, j = 1,2,3$ odnoszą się do kierunków przyjętego układu współrzędnych.

Końcowe sformułowanie słabe dla dynamiki to warunek całkowy, w którym występują kolejno cztery całki: bezwładnościowa, pracy wewnętrznej, sił zewnętrznych oraz trakcji:

$$\int_\Omega \delta\, u_i\, \rho\, \ddot{u}_i \, \mathrm{d}\Omega + \int_\Omega \delta\, \varepsilon_{ij}\, \sigma_{ij} \, \mathrm{d}\Omega - \int_\Omega \delta\, u_i\, b_i \, \mathrm{d}\Omega - \int_{\Gamma_t} \delta\, u_i\, g_i \, \mathrm{d}\Gamma = 0 \tag{2.8}$$

gdzie dΩ oznacza element całkowania po obszarze Ω, dΓ oznacza element całkowania po brzegu, a b_i jest składową siły objętościowej przypadającej na jednostkę objętości. Równanie wyraża warunek równowagi dynamicznej w postaci całkowej.

Dyskretyzacja przestrzenna polega na podziale obszaru Ω na elementy i aproksymacji pola przemieszczeń przez funkcje kształtu oraz przemieszczenia węzłowe. Przemieszczenia aproksymuje się funkcją:

$$\mathbf{u}(\mathbf{x},t) \approx \hat{\mathbf{u}} = \sum_{b} N_b\,(\mathbf{x})\,\tilde{\mathbf{u}}_b(t) = \mathbf{N}(\mathbf{x})\,\mathbf{d}(t) \tag{2.9}$$

gdzie N_b są funkcjami kształtu elementu skończonego, a $\mathbf{u}_b(t)$ są przemieszczeniami węzłowymi zależnymi od czasu. W wyniku dyskretyzacji przestrzennej ciągłe pole przemieszczeń $\mathbf{u}(\mathbf{x},t)$ zostaje aproksymowane jako kombinacja funkcji kształtu i wartości węzłowych. Przemieszczenia węzłowe tworzą wektor stopni swobody $\mathbf{d}(t)$, który stanowi główną niewiadomą w sformułowaniu przemieszczeniowym MES.

W celu powiązania odkształceń i przemieszczeń wygodnie jest zastosować zapis operatorowy:

$$\boldsymbol{\varepsilon} = \mathbf{S}\mathbf{d} = \mathbf{S}\mathbf{N}\mathbf{d} = \mathbf{B}\mathbf{d} \tag{2.10}$$

gdzie $\boldsymbol{\varepsilon}$ jest wektorem odkształceń, $\mathbf{d}$ jest wektorem przemieszczeń, a $\mathbf{S}$ jest operatorem odkształceń. Macierz $\mathbf{B}$, jest wynikiem działania operatora różniczkowego $\mathbf{S}$ na funkcje kształtu $\mathbf{N}$.

Po podstawieniu aproksymacji MES do równania słabego otrzymuje się zapis zależny od skończonej liczby stopni swobody. Nieznane pole ciągłe zostaje zastąpione wartościami węzłowymi, zebranymi w wektorze $\mathbf{d}(t)$, oraz jego pochodnymi czasowymi. Współczynniki stojące przy tych wielkościach wynikają z całkowania funkcji kształtu i ich pochodnych w obrębie elementów skończonych. Człon bezwładnościowy prowadzi do macierzy masy. Człon pracy wewnętrznej prowadzi do wektora sił wewnętrznych, a w przypadku liniowej sprężystości do macierzy sztywności. Człony obciążeń objętościowych i brzegowych tworzą wektor sił zewnętrznych. Wielkości te są najpierw wyznaczane na poziomie elementu, a następnie składane do układu globalnego przez mapowanie lokalnych stopni swobody elementu na globalny wektor stopni swobody.

Dla liniowej sprężystości (w sensie liniowego prawa konstytutywnego) zapis wynikający z dyskretyzacji przestrzennej globalnej macierzy sztywności $\mathbf{K}$ ma postać:

$$\mathbf{K} = \sum_{e=1}^{n_e} \mathbf{A}_e^{\mathrm{T}}\ \mathbf{K}^{(e)}\ \mathbf{A}_e, \mathbf{K}^{(e)} = \int_{\Omega^{(e)}} \mathbf{B}^{(e)\mathrm{T}}\,\mathbf{D}\mathbf{B}^{(e)}\,\mathrm{d}\Omega \tag{2.11}$$

Dla równania ruchu (człon bezwładności) macierz masy $\mathbf{M}$ zapisuje się jako:

$$\mathbf{M} = \sum_{e=1}^{n_e} \mathbf{A}_e^{\mathrm{T}}\ \mathbf{M}^{(e)}\ \mathbf{A}_e, \mathbf{M}^{(e)} = \int_{\Omega^{(e)}} \rho\ \mathbf{N}^{(e)\mathrm{T}} \mathbf{N}^{(e)}\ \mathrm{d}\Omega \tag{2.12}$$

gdzie $\mathbf{K}^{(e)}$ oznacza macierz sztywności elementu skończonego o numerze e, stanowiącą wkład elementowy wynikający z całkowania po objętości tego elementu. Symbol $\mathbf{M}^{(e)}$ oznacza macierz masy elementu skończonego o numerze e, natomiast e jest numerem elementu skończonego w sumowaniu po elementach siatki. Liczbę wszystkich elementów skończonych w dyskretyzacji obszaru oznaczono przez n_e. Macierz $\mathbf{A}_e$ jest macierzą składania przypisaną do elementu skończonego o numerze e i służy do odwzorowania lokalnych stopni swobody elementu na odpowiadające im stopnie swobody globalne. Macierz $\mathbf{B}^{(e)}$ opisuje zależność odkształceń od przemieszczeń w elemencie skończonym o numerze e, a $\mathbf{D}$ jest macierzą konstytutywną materiału. Symbol $d\Omega$ oznacza element miary całkowania po objętości elementu, natomiast $\Omega^{(e)}$ oznacza objętość elementu skończonego o numerze e. Macierz $\mathbf{N}^{(e)}$ jest macierzą funkcji kształtu dla pola przemieszczeń w elemencie skończonym o numerze e, a ρ oznacza gęstość materiału.

Macierz tłumienia $\mathbf{C}$ w wielu zadaniach przyjmujemy jako przybliżenie:

$$\mathbf{C} = \alpha_{damp}\ \mathbf{M} + \beta_{damp}\ \mathbf{K} \tag{2.13}$$

gdzie α_{damp} i β_{damp} dobiera się tak, aby uzyskać zakładane charakterystyki tłumienia.

2.1.1. Rozwiązywanie równań dynamiki MES

W analizach zależnych od czasu niewiadomą jest wektor przemieszczeń węzłowych $\mathbf{d}(t)$. W najprostszym przypadku, przy braku bezwładności i tłumienia, równanie statyki liniowej ma postać:

$$\mathbf{Kd} = \mathbf{F} \tag{2.14}$$

W tym wariancie zakłada się liniową zależność sił wewnętrznych, a czas nie występuje.

W przypadku nieliniowości, geometrycznych i/lub materiałowych, relacja pomiędzy przemieszczeniami a siłami wewnętrznymi nie jest liniowa:

$$\mathbf{K}(\mathbf{d})\ \mathbf{d} = \mathbf{F} \tag{2.15}$$

gdzie $\mathbf{K}(\mathbf{d})$ należy rozumieć jako operator zależny od stanu, który spełnia zależność $\mathbf{K}(\mathbf{d})\,\mathbf{d} = \mathbf{f}_{\mathrm{int}}(\mathbf{d})$. W ujęciu ogólnym wielkością podstawową jest zatem wektor sił wewnętrznych, a standardowa postać statyki nieliniowej zapisywana jest jako:

$$\mathbf{f}_{\mathrm{int}}(\mathbf{d}) = \mathbf{F} \tag{2.16}$$

W rozwiązaniu przyrostowo-iteracyjnym (np. metodą Newtona) występuje dodatkowo styczna macierz sztywności $\mathbf{K}_{\mathrm{T}}$, zdefiniowana jako pochodna sił wewnętrznych względem przemieszczeń,

$$\mathbf{K}_{\mathrm{T}}(\mathbf{d}) = \frac{\partial \mathbf{f}_{\mathrm{int}}(\mathbf{d})}{\partial \mathbf{d}} \tag{2.17}$$

W przypadku analiz związanych z określeniem częstości i postaci drgań własnych korzystamy z równania:

$$\mathbf{M}\ddot{\mathbf{d}} + \mathbf{K}\mathbf{d} = \mathbf{0} \tag{2.18}$$

Zadanie własne, pozwala wyznaczyć częstości własne i odpowiadające im postacie drgań [35]:

$$(\mathbf{K} - \lambda_k \mathbf{M})\boldsymbol{\Phi}_k = \mathbf{0} \tag{2.19}$$

gdzie $\boldsymbol{\Phi}_k$ jest wektorem własnym (postacią drgań), a λ_k jest wartością własną związaną z częstością własną ω_k relacją $\lambda_k = \omega_k^2$.

Równanie uwzględniające tłumienie oraz siły masowe opisane jest jako:

$$\mathbf{M}\ddot{\mathbf{d}} + \mathbf{C}\dot{\mathbf{d}} + \mathbf{K}\mathbf{d} = \mathbf{F} \tag{2.20}$$

Równanie jest układem równań różniczkowych drugiego rzędu. Jego rozwiązanie wymaga zatem posiadania warunków początkowych dla przemieszczeń i prędkości, np.:

$$\mathbf{d}(0) = \mathbf{d}_0 \tag{2.21}$$

$$\dot{\mathbf{d}}(0) = \mathbf{v}_0 \tag{2.22}$$

gdzie $\mathbf{d}_0$ oznacza wektor początkowych przemieszczeń w chwili $t = 0$, natomiast $\mathbf{v}_0$ oznacza wektor początkowych prędkości w tej samej chwili. Warunki te określają stan początkowy

układu dynamicznego i są niezbędne do jednoznacznego wyznaczenia odpowiedzi czasowej układu opisanego równaniem ruchu drugiego rzędu.

W zadaniach, w których występują nieliniowości materiałowe, geometryczne lub zjawiska kontaktu, siły wewnętrzne nie są liniową funkcją przemieszczeń. Wtedy postać równania ruchu ujmuje się jako równowagę pomiędzy bezwładnością, tłumieniem, siłami wewnętrznymi i zewnętrznymi:

$$\mathbf{M}\ddot{\mathbf{d}} + \mathbf{C}\dot{\mathbf{d}} + \mathbf{f}_{\text{int}}(\mathbf{d}) = \mathbf{F} \tag{2.23}$$

Rozwiązanie równań 2.21 i 2.24 wymaga zastosowania jednego z dwóch podejść. Jednym z nich jest schemat jawny. Polega na wyznaczaniu stanu układu w kolejnym kroku czasowym bezpośrednio na podstawie wielkości znanych z kroku bieżącego lub poprzedniego, bez iteracyjnego rozwiązywania globalnego układu równań równowagi w każdym kroku. Z tego powodu jest szczególnie przydatny w analizach silnie nieliniowych, obejmujących kontakt, lokalizację zniszczenia, utratę nośności lub usuwanie elementów, pod warunkiem zastosowania odpowiednio małego kroku czasowego wynikającego z warunku stabilności schematu.

Po dyskretyzacji przestrzennej otrzymuje się układ równań ruchu w dziedzinie czasu. W podejściu jawnym (ang. explicit) wielkości w kroku $n+1$ wyznacza się wyłącznie na podstawie informacji znanych w kroku n (oraz ewentualnie $n-1$), bez rozwiązywania globalnego układu równań [36].

Rozważamy ogólny układ równań różniczkowych zwyczajnych pierwszego rzędu:

$$\dot{y} = f(t, y), t \geq t^0, y(t^0) = y^0 \tag{2.24}$$

Po wprowadzeniu dyskretyzacji czasu $t^{n+1} = t^n + \Delta t^n$ jawna metoda Eulera (ang. Euler forward) ma postać:

$$y^{n+1} = y^n + \Delta t^n \, f\,(t^n, y^n), \;\; n = 0,1, \ldots, \tag{2.25}$$

Jest to schemat jawny i jednokrokowy, ponieważ prawa strona wykorzystuje wyłącznie wielkości znane w chwili t_n.

W MES dla dynamiki problemów drugiego rzędu w czasie stosuje się schemat różnic centralnych, który może być interpretowany jako jawna aktualizacja typu Eulera, lecz realizowana na siatce czasowej z przesunięciem o pół kroku dla prędkości.

W metodzie różnic centralnych [9] wprowadza się kroki czasu związane z chwilami całkowitymi t^n i połówkowymi $t^{n\pm1/2}$:

$$\Delta t^{\,n+\frac{1}{2}} = t^{n+1} - t^n, \Delta t^{\,n} = t^{n+\frac{1}{2}} - t^{n-\frac{1}{2}} = \frac{1}{2}\left(\Delta t^{\,n+\frac{1}{2}} + \Delta t^{\,n-\frac{1}{2}}\right) \tag{2.26}$$

Prędkość w chwili połówkowej aproksymuje się różnicą w przemieszczeniu:

$$\dot{\mathbf{d}}^{\,n+\frac{1}{2}} = \frac{\mathbf{d}^{\,n+1} - \mathbf{d}^{\,n}}{\Delta t^{\,n+\frac{1}{2}}} \tag{2.27}$$

Stąd wynika jawna aktualizacja przemieszczeń:

$$\mathbf{d}^{\,n+1} = \mathbf{d}^{\,n} + \Delta t^{\,n+\frac{1}{2}}\,\dot{\mathbf{d}}^{\,n+\frac{1}{2}} \tag{2.28}$$

Przyspieszenie $\ddot{\mathbf{d}}^{\,n}$ w chwili całkowitej t^n oraz odpowiadająca mu jawna aktualizacja prędkości mają postać:

$$\ddot{\mathbf{d}}^{\,n} = \frac{\dot{\mathbf{d}}^{\,n+\frac{1}{2}} - \dot{\mathbf{d}}^{\,n-\frac{1}{2}}}{\Delta t^{\,n}}, \dot{\mathbf{d}}^{\,n+\frac{1}{2}} = \dot{\mathbf{d}}^{\,n-\frac{1}{2}} + \Delta t^{\,n}\,\mathbf{d}^{\,n} \tag{2.29}$$

Dla stałego kroku czasu klasyczna postać różnicy centralnej dla przyspieszenia jest równa:

$$\ddot{\mathbf{d}}^{\,n} = \frac{\mathbf{d}^{\,n+1} - 2\mathbf{d}^{\,n} + \mathbf{d}^{\,n-1}}{(\Delta t)^2} \tag{2.30}$$

W kroku n przyspieszenia wyznacza się z równowagi w postaci:

$$\mathbf{M}\,\mathbf{d}^{\,n} = \mathbf{F}(\mathbf{d}^{\,n}, t^n) - \mathbf{f}_{\mathrm{int}}(\mathbf{d}^{\,n}, t^n) \tag{2.31}$$

Następnie wykonuje się aktualizację prędkości $\dot{\mathbf{d}}^{\,n+\frac{1}{2}}$ oraz przemieszczeń $\mathbf{d}^{n+1}$ zgodnie z relacjami (2.30) oraz (2.29). Przemieszczeniowe warunki brzegowe realizuje się przez wymuszenie wartości odpowiednich składowych wektora $\mathbf{d}^{n+1}$ po wykonaniu kroku aktualizacji.

Schemat jawny jest warunkowo stabilny [9], a dopuszczalny krok czasowy wynika z analizy stabilności liniowego układu dynamicznego bez tłumienia, w którym decydującą rolę odgrywa najwyższa częstość własna układu:

$$\Delta t = \alpha_{crit}\,\Delta t_{\mathrm{crit}}, \Delta t_{\mathrm{crit}} = \frac{2}{\omega_{\max}} \leq \min_{e,I} \frac{2}{\omega_I^e} = \min_{e} \frac{\ell_e}{c_e} \tag{2.32}$$

gdzie $\omega_{\max}$ oznacza maksymalną częstość kołową układu liniowego, a ω_I^e jest częstością własną elementu skończonego o numerze e. Symbol α_{crit} oznacza współczynnik bezpieczeństwa stosowany przy wyznaczaniu stabilnego kroku czasowego. Wielkość c_e określa prędkość rozchodzenia się fali mechanicznej w elemencie skończonym, natomiast ℓ_e oznacza charakterystyczną długość elementu, przyjmowaną jako odległość pomiędzy dwoma najbliższymi węzłami siatki.

Interpretacja falowa ostatniej nierówności jest jednoznaczna wymiarowo: Δt_{crit} musi być mniejsze niż czasu przejścia fali mechanicznej przez najmniejszy element. To oznacza, że wielkość kroku czasowego jest uzależniona od wielkości najmniejszego elementu w zadaniu (jeżeli występuje tylko jeden materiał).

Metodę całkowania jawnego stosuje się, ponieważ umożliwia jawne wyznaczanie kolejnych wartości prędkości i przemieszczeń na podstawie wielkości znanych z poprzednich kroków czasowych, bez konieczności iteracyjnego rozwiązywania globalnego układu równań w każdym kroku analizy. Jednocześnie schemat wykorzystuje przesunięcie prędkości o pół kroku czasowego względem przemieszczeń i przyspieszeń, co poprawia jego przydatność w analizach zjawisk dynamicznych, falowych i oscylacyjnych w porównaniu z prostym schematem jednokrokowym, w którym wszystkie wielkości są określane w tych samych chwilach czasowych.

2.1.2. Modele plastyczności w zadaniach nieliniowych

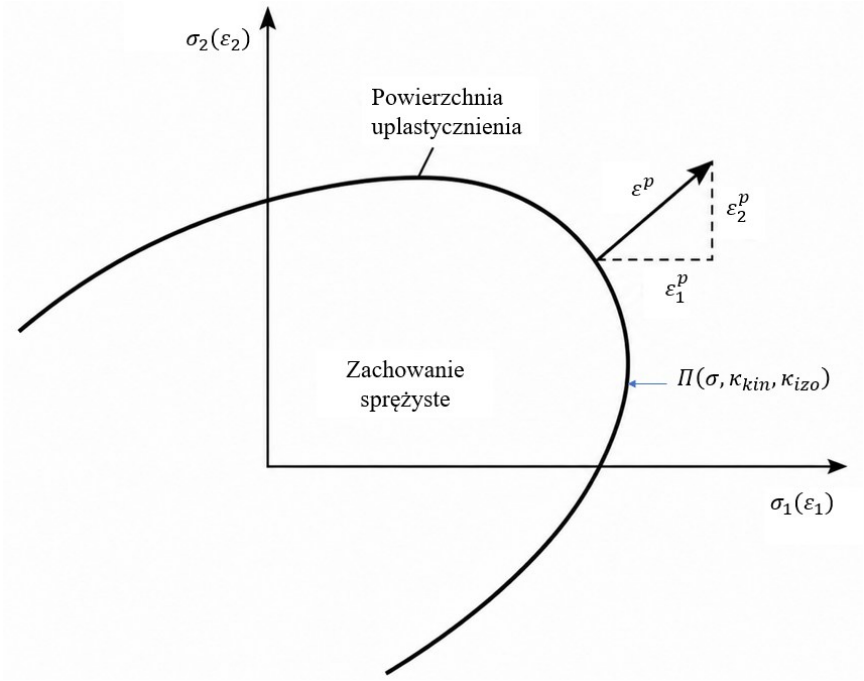
Plastyczność jest własnością materiału polegającą na występowaniu nieodwracalnych odkształceń po odciążeniu, jeśli stan naprężenia osiągnie poziom odpowiadający uplastycznieniu [102]. W klasycznej teorii sprężysto-plastyczności dla małych odkształceń przyjmuje się addytywny rozkład całkowitego tensora odkształcenia $\boldsymbol{\varepsilon}$ na część sprężystą $\boldsymbol{\varepsilon}^e$ oraz plastyczną $\boldsymbol{\varepsilon}^p$:

$$\boldsymbol{\varepsilon} = \boldsymbol{\varepsilon}^e + \boldsymbol{\varepsilon}^p \tag{2.33}$$

Przejście od zachowania sprężystego do sprężysto-plastycznego opisuje się powierzchnią plastyczności (powierzchnią uplastycznienia), definiowaną przez funkcję uplastycznienia Π zależną od tensora naprężenia $\boldsymbol{\sigma}$ oraz parametrów umocnienia:

$$\Pi(\boldsymbol{\sigma}, \kappa_{kin}, \kappa_{izo}) = 0 \tag{2.34}$$

W zapisie przyjętym w literaturze cytowanej powyżej symbol κ występuje zarówno dla parametrów umocnienia kinematycznego κ_{kin}, jak i izotropowego κ_{izo}. Zbiór stanów naprężenia spełniających $\Pi = 0$ stanowi powierzchnię plastyczności w przestrzeni naprężeń, natomiast znak Π rozróżnia stany bez płynięcia plastycznego i stany związane z uplastycznieniem. Przykładową powierzchnię uplastycznienia ilustruje rysunek 2.3.



Rysunek 2.3. Powierzchnia uplastycznienia [102].

Dla materiału izotropowego funkcję Π wygodnie opisuje się przez niezmienniki tensora naprężenia oraz dewiatora naprężenia. Wprowadza się pierwszy niezmiennik tensora naprężenia oraz drugi i trzeci niezmiennik dewiatora naprężenia:

$$J_1 = \sigma_{ii}, 2J_2 = s_{ij}s_{ji} = |\mathbf{s}|^2, 3J_3 = s_{ij}s_{jk}s_{ki} = \det \mathbf{s} \tag{2.35}$$

gdzie σ_{ij} są składowymi tensora naprężenia Cauchy'ego $\boldsymbol{\sigma}$, s_{ij} są składowymi dewiatora naprężenia $\boldsymbol{s}$, $|\mathbf{s}|$ jest normą dewiatora w sensie $|\mathbf{s}|^2 = s_{ij}s_{ji}$, natomiast J_1, J_2, J_3

są niezmiennikami opisującymi stan naprężenia w sposób niezależny od orientacji osi układu odniesienia.

W celu zapisu klasycznych powierzchni plastyczności stosuje się parametryzację niezmienników przez naprężenie średnie σ_m, miarę intensywności części dewiatorowej σ_{dev} oraz kąt Lodego θ (wyjaśniony w rozdziale 2.1.3):

$$3\sigma_m = J_1, \sigma_{dev} = \sqrt{J_2}, \quad 3\theta = sin^{-1}\left(-\frac{3\sqrt{3}\,J_3^{1/3}}{2\bar{\sigma}}\right) \; gdzie \; -\frac{\pi}{6} \leq \theta \leq \frac{\pi}{6} \tag{2.36}$$

Wielkość σ_m opisuje składową hydrostatyczną stanu naprężenia, σ_{dev} jest miarą „odległości" od osi hydrostatycznej w przestrzeni naprężeń (intensywnością części postaciowej), natomiast θ parametryzuje wpływ trzeciego niezmiennika dewiatora i w konsekwencji wpływa na kształt powierzchni uplastycznienia w przestrzeni naprężeń głównych.

Powierzchnię uplastycznienia dla kryterium Treski można zapisać jako:

$$\Pi = 2\sigma_{dev} cos\,\theta - Y(\kappa_{izo}) = 0 \tag{2.37}$$

A dla kryterium Hubera von Misesa:

$$\Pi = \sqrt{2}\,\sigma_{dev} - \sqrt{\frac{2}{3}}\,Y(\kappa_{izo}) = |\mathbf{s}| - \sqrt{\frac{2}{3}}\,Y(\kappa_{izo}) = 0 \tag{2.38}$$

W powyższych równaniach $Y(\kappa_{izo})$ jest miarą aktualnej granicy uplastycznienia w ujęciu izotropowego umocnienia, zależną od parametru κ_{izo}.

Dla materiałów zależnych od ciśnienia i tarcia wewnętrznego uwzględnia się bezpośrednio wpływ naprężenia średniego. Kryterium Moora Coulomba w zapisie $(\sigma_m, \sigma_{dev}, \theta)$ma postać:

$$\Pi = \sigma_m sin\,\phi_{fric} + \sigma_{dev}\left(cos\,\theta - \frac{1}{\sqrt{3}} sin\,\phi_{fric}\; sin\,\theta\right) - c_{coh}\cos\phi_{fric} = 0 \tag{2.39}$$

gdzie c_{coh} jest kohezją, a ϕ_{fric} jest kątem tarcia wewnętrznego, przy czym oba parametry mogą zależeć od parametru umocnienia κ_{izo}.

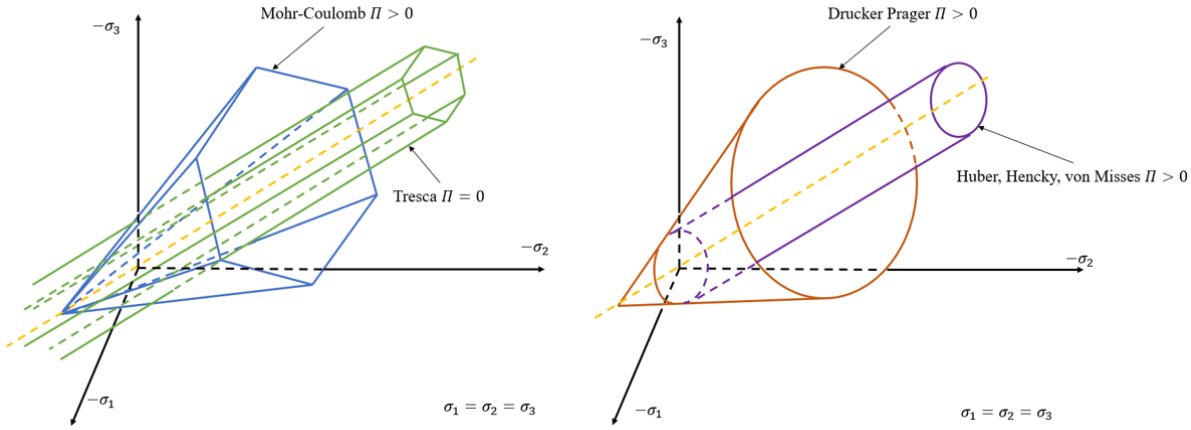
Jako gładką aproksymację zachowania tarciowego często stosuje się powierzchnię Druckera Pragéra:

$$\Pi = 3\alpha'(\kappa_{izo})\sigma_m + \sigma_{dev} - k_{DP}(\kappa_{izo}) = 0 \tag{2.40}$$

gdzie:

$$\alpha' = \frac{2sin\,\phi_{fric}}{\sqrt{3}(3 - sin\,\phi_{fric})}, \qquad k_{DP} = \frac{6cos\,\phi_{fric}}{\sqrt{3}(3 - sin\,\phi_{fric})} \tag{2.41}$$

Powierzchnie uplastycznienia ilustruje rysunek 2.4.



Rysunek 2.4. Powierzchnie uplastycznienia dla różnych modeli plastyczności [102].

W analizach udarowych (wysokie prędkości odkształcenia) istotne jest, że część materiałów wykazuje wrażliwość na prędkość odkształcenia, co może zmieniać efektywną granicę uplastycznienia i poziom dyssypacji energii. W ujęciu lepkoplastycznym, opisującym zależność od prędkości, wprowadza się modyfikację warunku uplastycznienia przez relację wiążącą tempo mnożnika plastycznego $\dot{\lambda}$ z funkcją $\phi(\Pi)$ oraz parametrem lepkości γ [102]:

$$\gamma\langle\phi(\Pi)\rangle - \dot{\lambda} = 0 \tag{2.42}$$

Nawiasy $\langle\cdot\rangle$oznaczają część dodatnią, co formalnie zapisano jako:

$$\langle\phi(\Pi)\rangle = \begin{cases} 0 & \text{gdy } \Pi \leq 0, \\ \phi(\Pi) & \text{gdy } \Pi > 0. \end{cases} \tag{2.43}$$

W konsekwencji, dla $\Pi \leq 0$ nie występuje płynięcie plastyczne ($\dot{\lambda} = 0$), natomiast dla $\Pi > 0$ dopuszcza się tempo płynięcia kontrolowane przez γ i postać $\phi(\Pi)$.

W niniejszej pracy, ze względu na charakter analizowanego problemu oraz potrzebę zachowania stabilności i powtarzalności obliczeń w licznych wariantach symulacji, jako model uplastycznienia przyjęto kryterium Hubera-von Misesa-Henckiego. Przyjęcie tej postaci oznacza, że uplastycznienie jest determinowane przede wszystkim przez składowe postaciowe (opisane przez σ_{dev} lub $|\, s \,|$), a wpływ składowej hydrostatycznej σ_m na sam warunek uplastycznienia jest pomijany. Z punktu widzenia numerycznego skutkuje to gładką, wypukłą powierzchnią uplastycznienia i stabilniejszą implementacją w obliczeniach nieliniowych w porównaniu z kryteriami o niegładkiej geometrii powierzchni.

Jednocześnie dla tworzyw sztucznych należy wskazać ryzyka modelowe tego uproszczenia. Jeżeli materiał jest wrażliwy na ciśnienie hydrostatyczne (istotny wpływ σ_m na uplastycznienie) lub wykazuje asymetrię w rozciąganiu i ściskaniu, to kryterium von Misesa może nie odtwarzać poprawnie granicy uplastycznienia w całym zakresie stanów naprężenia. Sens tych ograniczeń jest widoczny na tle kryteriów Moora Coulomba i Druckera Pragéra, w których σ_m wchodzi jawnie do warunku $\Pi = 0$, co jakościowo zmienia przewidywania w stanach o dużej trójosiowości naprężenia. W tej pracy przyjęcie kryterium Hubera-von Misesa-Henckiego traktuje się jako świadome uproszczenie modelowe, którego celem jest uchwycenie dominującego mechanizmu uplastycznienia w analizowanej konstrukcji przy zachowaniu rozwiązywalności procesu optymalizacji.

2.1.3. Modele inicjacji i ewolucji zniszczenia

W modelowaniu zniszczenia materiału rozróżnia się inicjację zniszczenia (moment spełnienia lokalnego kryterium, po którym rozpoczyna się degradacja nośności) oraz ewolucję zniszczenia (opis postępu degradacji aż do utraty nośności i ewentualnego usunięcia elementu)[10]. W praktyce obliczeń MES rozdzielenie tych etapów pozwala z jednej strony zdefiniować warunek rozpoczęcia degradacji na podstawie mierzalnych wielkości (np. zredukowanego odkształcenia plastycznego), a z drugiej strony kontrolować proces osłabiania odpowiedzi materiału w sposób bardziej stabilny numerycznie i mniej wrażliwy na zagęszczenie siatki.

W kryteriach dla materiałów ciągliwych inicjacja zniszczenia jest wiązana z osiągnięciem przez materiał krytycznej wartości odkształcenia plastycznego, przy czym wartość krytyczna nie jest stała, lecz zależy od stanu naprężenia [99]. Stan naprężenia jest parametryzowany m.in. przez trójosiowość naprężenia (ang. stress triaxiality) oraz, przez parametr związany z kątem Lodego (ang. Lode angle) [6].

Trójosiowość naprężenia η jest definiowana jako stosunek naprężenia średniego do naprężenia zredukowanego von Misesa, czyli miara udziału składowej hydrostatycznej względem składowej dewiatorowej.

$$\eta = \frac{\sigma_m}{\bar{\sigma}} \tag{2.44}$$

Naprężenie średnie σ_m oraz naprężenie zredukowane $\bar{\sigma}$ są opisane:

$$\sigma_m = \frac{1}{3}(\sigma_1 + \sigma_2 + \sigma_3) \; = \; \frac{1}{3}\, tr(\sigma) \tag{2.45}$$

$$\bar{\sigma} = \sqrt{\frac{1}{2}[(\sigma_1 - \sigma_2)^2 + (\sigma_2 - \sigma_3)^2 + (\sigma_3 - \sigma_1)^2]} \tag{2.46}$$

Sam parametr η nie rozróżnia jednak w pełni „kształtu" stanu dewiatorowego. Do tego potrzebna jest informacja o trzecim niezmienniku dewiatora, która jest parametryzowana kątem Lodego lub jego funkcją. W praktyce oznacza to, że dwa stany o tej samej trójosiowości mogą mieć różną skłonność do zniszczenia, jeżeli różni się udział ścinania względem rozciągania w sensie układu naprężeń głównych. Klasyczne ujęcia zniszczenia często pomijały wpływ ciśnienia (składowej hydrostatycznej) oraz wpływ trzeciego niezmiennika, natomiast dane doświadczalne wskazują, że obie zależności mogą być istotne i powinny być uwzględniane w opisie zniszczenia materiałów podatnych na odkształcenia plastyczne.

Drugi parametr rozróżniający typ stanu dewiatorowego jest związany z kątem Lodego θ. Bai i Wierzbicki [6] wiążą znormalizowany trzeci niezmiennik dewiatora ξ z kątem Lodego następująco:

$$\xi = \left(\frac{r}{\bar{\sigma}}\right)^3 = \cos(3\theta)\,, \bar{\theta} = 1 - \frac{6\theta}{\pi} = 1 - \frac{2}{\pi}\arccos\xi \tag{2.47}$$

gdzie $\bar{\theta}$ to znormalizowany parametr kąta Lodego do wartości [-1,1], a r jest skalarną miarą zależną od trzeciego niezmiennika tensora naprężeń dewiatorowych L(tj. $J_3 = \det(L)$), stosowaną do opisu wpływu kąta Lodego.

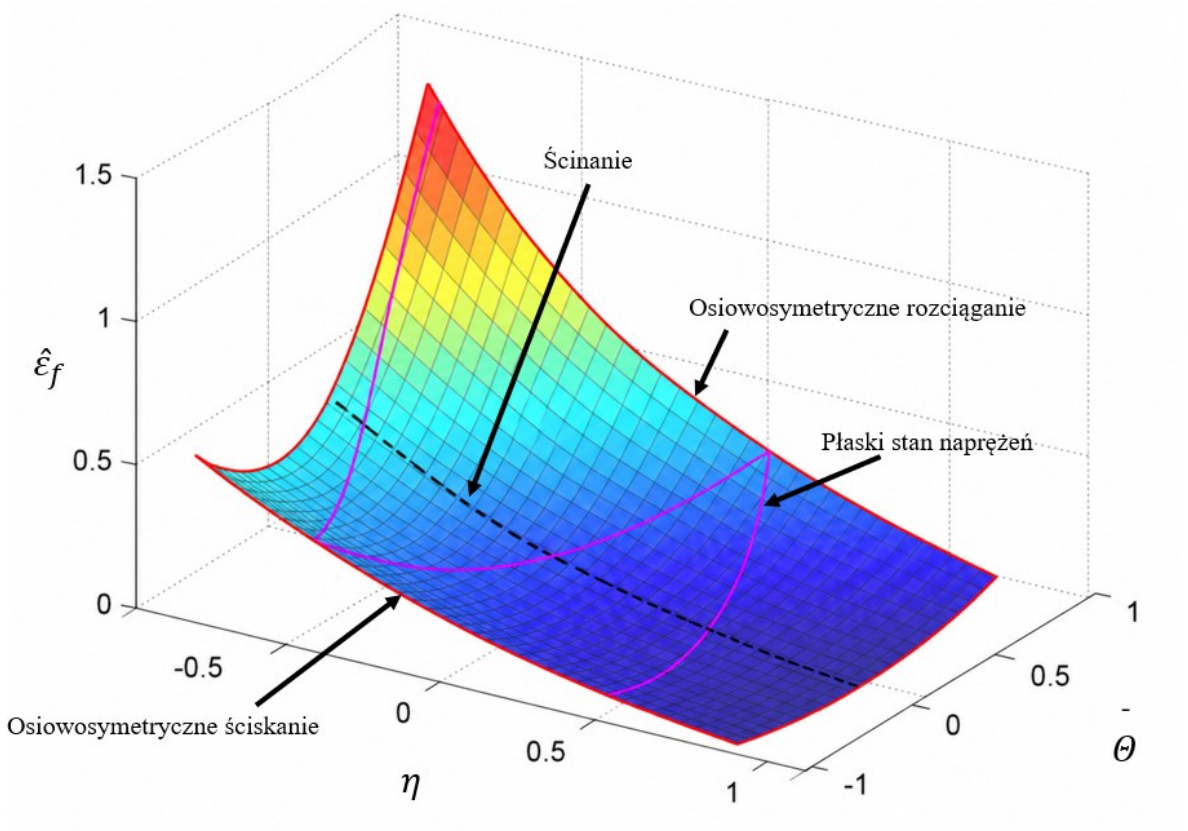
Sama trójosiowość η opisuje jaki jest udział stanu hydrostatycznego, ale nie rozróżnia w pełni, czy stan dewiatorowy jest bliższy ścinaniu czy stanom osiowo-symetrycznym w przestrzeni naprężeń głównych. Parametr $\bar{\theta}$ uzupełnia tę informację. W modelach inicjacji zniszczenia oznacza to, że dwa punkty o tej samej η mogą mieć inne odkształcenie do inicjacji,

jeżeli różni się charakter stanu dewiatorowego, co jest dokładnie powodem wprowadzania zależności od kąta Lodego.

W ujęciu Bai i Wierzbickiego odkształcenie do zniszczenia jest opisane jako funkcja stanu naprężenia, w szczególności trójosiowości i parametru związanego z kątem Lodego:

$$\bar{\varepsilon}_f = \hat{\varepsilon}_f(\eta, \bar{\theta}) \tag{2.48}$$

gdzie $\bar{\varepsilon}_f$ jest odkształceniem plastycznym zredukowanym przy zniszczeniu, a $\hat{\varepsilon}_f$ jest tzw. lokusem zniszczenia (funkcją wyznaczaną doświadczalnie i kalibrowaną), zależnym od η i $\bar{\theta}$. Zastosowanie tej funkcji pozwala zastąpić pojedynczą, stałą wartość odkształcenia granicznego wartością zależną od stanu naprężenia. Takie ujęcie jest zgodne z obserwacją, że odkształcenie prowadzące do zniszczenia jest zazwyczaj mniejsze w próbach charakteryzujących się wysoką trójosiowością naprężenia niż w próbach zdominowanych przez ścinanie. Wpływ trójosiowości i kąta Lodego ilustruje rysunek 2.5.



Rysunek 2.5. Odkształcenie przy zerwaniu w zależności od współczynnika trójosiowości i kąta Lodego [6].

Definicja miary akumulacji zniszczenia (wskaźnika) w postaci całki po przyroście zredukowanego odkształcenia plastycznego [6]:

$$D(\bar{\varepsilon}^p) = \int_0^{\bar{\varepsilon}^p} \frac{\mathrm{d}\bar{\varepsilon}^p}{f(\eta, \bar{\Theta})}, D\left(\bar{\varepsilon}_0^p\right) = 0, D\left(\bar{\varepsilon}_f^p\right) = 1 \tag{2.49}$$

W tym zapisie $D(\bar{\varepsilon}^p)$ jest skalarową miarą postępu zniszczenia (od 0 do inicjacji), $\bar{\varepsilon}^p$ jest zredukowanym odkształceniem plastycznym, natomiast $f(\eta, \bar{\Theta})$ jest funkcją lokusu zniszczenia, której interpretacja jest zgodna z $\hat{\varepsilon}_f(\eta, \bar{\Theta})$ z poprzedniego wzoru. Zaletą postaci całkowej jest możliwość uwzględnienia historii obciążenia: jeżeli w trakcie procesu odkształcania zmienia się stan naprężenia, to wkład do D jest sumowany inkrementalnie w kolejnych przyrostach $\bar{\varepsilon}^p$, a wpływ bieżącej η i $\bar{\Theta}$ działa poprzez mianownik [99].

Po inicjacji zniszczenia konieczne jest zdefiniowanie ewolucji zniszczenia, czyli sposobu degradacji sztywności i nośności materiału, aż do osiągnięcia pełnej degradacji (zniszczenia), co w praktyce MES często prowadzi do usunięcia elementu [34].

W mechanice ośrodków ciągłych uszkodzenia wygodnym i jednoznacznym interpretacyjnie wyborem jest zdefiniowanie zmiennej uszkodzenia jako udziału powierzchni (lub objętości) nieprzenoszącej obciążeń. W szczególności, przy interpretacji geometrycznej [67]:

$$\chi = \frac{f_\blacksquare}{F_\blacksquare} = A_A \tag{2.50}$$

gdzie, $f_\blacksquare$ oznacza pole powierzchni uszkodzonej, $F_\blacksquare$ całkowite pole przekroju, zaś A_A są odpowiednio ułamkiem objętościowym i ułamkiem powierzchniowym defektów. Z definicji $f_\blacksquare \leq F_\blacksquare$, a więc $0 \leq \chi \leq 1$, co jest spójne z intuicją: $\chi = 0$ odpowiada stanowi nieuszkodzonemu, natomiast wzrost χ oznacza postępującą degradację.

W modelu Lemaitre'a kluczowym założeniem jest równoważność odkształceń (ang. strain equivalence), zgodnie z którą odkształcenia makroskopowe są takie same w materiale rzeczywistym i w materiale „efektywnym" (pozbawionym defektów, ale obciążonym naprężeniem efektywnym) [67]. Prowadzi to do związku pomiędzy modułem sprężystości E materiału nominalnego i modułem zredukowanym $\bar{E}$:

$$\chi = 1 - \frac{\bar{E}}{E} \tag{2.51}$$

a także do definicji naprężenia efektywnego $\tilde{\sigma}$:

$$\tilde{\sigma} = \frac{\sigma}{1-\chi} \tag{2.52}$$

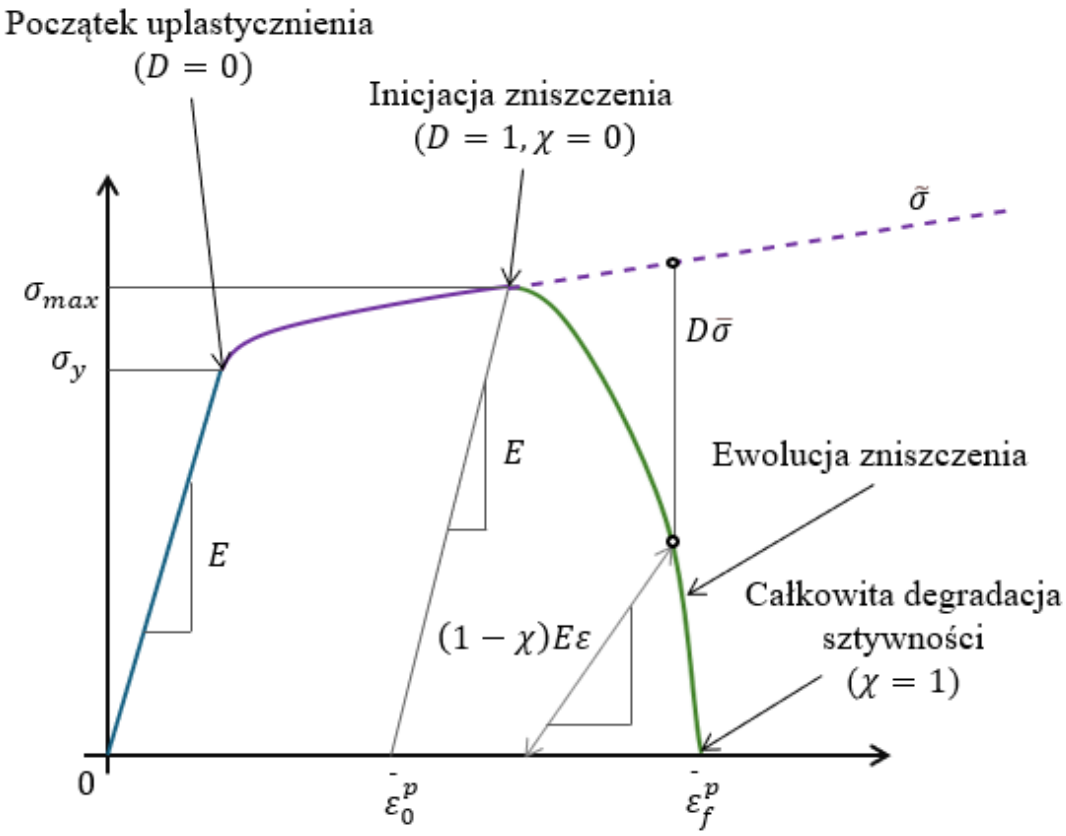
Dalej hipoteza równoważności odkształceń:

$$\tilde{\varepsilon} = \varepsilon \tag{2.53}$$

W praktyce oznacza to, że degradacja materiału może być interpretowana jako redukcja efektywnej sztywności przy zachowaniu tej samej miary odkształcenia. Dla jednoosiowego przypadku sprężystego relacja naprężenie-odkształcenie w obecności uszkodzenia przyjmuje postać:

$$\sigma = (1-\chi)E\varepsilon \tag{2.54}$$

Takie sformułowanie jest użyteczne jako punkt wyjścia do powiązania ewolucji zmiennej χ z miarami odkształceń nieodwracalnych. Zależność pomiędzy wskaźnikiem inicjacji zniszczenia D, zmienną degradacji sztywności χ oraz odpowiedzią naprężenie-odkształcenie przedstawiono schematycznie na rysunku 2.6.



Rysunek 2.6. Wykres rozciągania z modelem zniszczenia.

Powiązanie numerycznej utraty nośności strefy zniszczenia z bilansem energetycznym i charakterystycznym wymiarem dyskretyzacji wynika z faktu, że w modelach lokalnych energia rozpraszania może zależeć od rozmiaru elementu, jeśli nie narzuci się warunku energetycznego. Klasycznym sposobem zapewnienia poprawnej energii pękania jest podejście Hillerborga (model szczeliny kohezyjnej), w którym definiuje się energię pękania G_c jako pole pod krzywą zależności naprężenia od rozwarcia szczeliny $\sigma_{slot}(w_{slot})$ (gdzie w_{slot} jest szerokością rozwarcia) [32]:

$$G_c = \int_0^{w_1} \sigma_{slot}(w_{slot})\, \mathrm{d}w_{slot} \tag{2.55}$$

Warunek energetyczny w tej postaci stanowi jednoznaczne sprzęgnięcie modelu zniszczenia z energią rozpraszania na jednostkę nowej powierzchni pęknięcia, co jest krytyczne przy porównywaniu wyników dla różnych siatek, ponieważ kontroluje całkowitą energię „wydatkowaną” na wytworzenie zniszczenia.

W praktyce MES ujęcie energetyczne umożliwia definiowanie prawa degradacji w kategoriach dyssypacji energii na jednostkę powierzchni pęknięcia, co ogranicza wpływ

rozmiaru elementu na całkowitą energię zniszczenia i stabilizuje prognozę momentu utraty nośności w analizach z lokalizacją odkształceń.

Z punktu widzenia konsekwencji praktycznych dla obliczeń, nieuwzględnienie trójosiowości i parametru Lodego w kryterium inicjacji prowadzi do stosowania jednego „uśrednionego" odkształcenia granicznego, które nie rozróżnia mechanizmów inicjacji w ścinaniu i w rozciąganiu hydrostatycznym. Taki model bywa zdolny do dopasowania jednej klasy prób (np. jednoosiowego rozciągania), lecz w stanach kontaktowych i udarowych może systematycznie przeszacowywać lub niedoszacowywać odporność na zniszczenie w zależności od lokalnej trójosiowości i udziału ścinania. Uwzględnienie η i $\bar{\Theta}$ (lub ξ) umożliwia natomiast kalibrację lokusu $\hat{\varepsilon}_f(\eta, \bar{\Theta})$w sposób rozróżniający typowe stany naprężenia, co jest zgodne z przesłanką, że wpływ ciśnienia i trzeciego niezmiennika dewiatora powinien być w wielu materiałach istotny dla opisu zniszczenia.

2.2. Wybrane metody metamodelowania

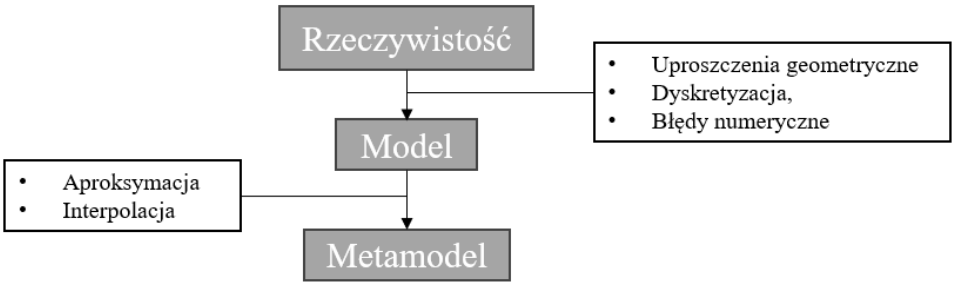
Metamodelowanie, określane również jako modelowanie zastępcze lub surogacyjne, polega na budowie uproszczonego modelu matematycznego, którego zadaniem jest przybliżenie odpowiedzi uzyskiwanej z bardziej złożonego modelu obliczeniowego. W takim ujęciu metamodel nie opisuje bezpośrednio pełnej fizyki analizowanego zjawiska, lecz odwzorowuje relację pomiędzy zmiennymi wejściowymi a odpowiedzią wyznaczoną wcześniej za pomocą modelu wysokiej wierności.

W zadaniach inżynierskich, w których wymagane jest wielokrotne wyznaczanie odpowiedzi układu dla różnych wariantów parametrów konstrukcyjnych, takie podejście pozwala ograniczyć koszt obliczeń numerycznych. W kontekście analiz MES opisanych w podrozdziale 2.1 pojedyncza symulacja może być na tyle kosztowna obliczeniowo, że jej wielokrotne wykonywanie staje się nieefektywne.

Metamodel jest uczony na zbiorze par danych wejście-wyjście, obejmujących wektory zmiennych projektowych oraz odpowiadające im wartości odpowiedzi obliczone za pomocą modelu wysokiej wierności. Takie uczenie relacji z par danych wejście-wyjście jest przykładem uczenia nadzorowanego (ang. supervised learning), przy czym typowo liczba dostępnych obserwacji jest niewielka, ponieważ każdy punkt wymaga kosztownej oceny odpowiedzi [27].

W niniejszym rozdziale przyjmuje się następującą notację: wektor zmiennych projektowych oznacza się przez $\mathbf{P} \in \mathcal{P} \subset \mathbb{R}^k$, gdzie k jest liczbą parametrów projektowych, a $\mathcal{P}$ jest rozważaną przestrzenią projektową.

Proces metamodelowania można przedstawić jako łańcuch zależności: rzeczywistość fizyczna jest przybliżana modelem numerycznym (np. z użyciem MES), natomiast metamodel jest kolejną, jeszcze bardziej uproszczoną warstwą, aproksymującą odpowiedź modelu numerycznego na podstawie ograniczonego zbioru wyników. W tym ujęciu metamodel nie zastępuje mechaniki opisanej w rozdziale 2.1, lecz pełni funkcję narzędzia obliczeniowego umożliwiającego szybkie przeszukiwanie wariantów konstrukcji, analizę wrażliwości, budowę map odpowiedzi oraz prowadzenie optymalizacji przy akceptowalnym koszcie czasowym (rysunek 2.7).



Rysunek 2.7. Zależności pomiędzy rzeczywistością, modelem, a metamodelem.

Metamodel w ujęciu ogólnym jest funkcją aproksymującą, która na podstawie wektora parametrów projektowych $\mathbf{P} \in \mathcal{P} \subset \mathbb{R}^k$ zwraca przybliżenie odpowiedzi modelu wysokiej wierności. Przyjmując, że odpowiedź referencyjna (np. z obliczeń MES) jest opisana odwzorowaniem $\mathcal{F}: \mathcal{P} \to \mathbb{R}$, to metamodel $m: \mathcal{P} \to \mathbb{R}$ stanowi przybliżenie $\mathcal{F}$ w rozważanym zakresie parametrów i generuje predykcję $\hat{q}$ według zależności:

$$q = \mathcal{F}(\mathbf{P}), \quad \hat{q} = m(\mathbf{P}) \tag{2.56}$$

Metamodel można interpretować jako funkcję, która jest pewną powierzchnią odpowiedzi. Dla przypadku o jednym parametrze powierzchnia odpowiedzi będzie krzywą, dla 2 zmiennych będzie to powierzchnia. Parametrów sterujących może być więcej, a wtedy powierzchnia odpowiedzi przyjmuje postać hiperpowierzchni w przestrzeni n-wymiarowej.

Warunkiem uzyskania użytecznego metamodelu jest właściwe przygotowanie danych modelu. Po pierwsze, parametry projektowe powinny być przeskalowane do porównywalnego zakresu, co poprawia uwarunkowanie numeryczne i upraszcza dobór parametrów metamodeli; powszechną praktyką jest normalizacja zmiennych do hipersześcianu jednostkowego $(0,1)^k$. Po drugie, próbki powinny zostać podzielone na część służącą konstrukcji metamodelu oraz część służącą weryfikacji jakości predykcji (zbiory testowe i walidacyjne), aby ocena dokładności nie była obciążona dopasowaniem do danych uczących [37].

Jakość metamodelu jest oceniana miarami, które pozwalają zarówno na interpretację względną, jak i bezwzględną błędu predykcji. Współczynnik determinacji R^2 jest miarą względną, odnoszącą błąd dopasowania do zmienności danych referencyjnych. Dla zbioru weryfikacyjnego o liczności n (tu: liczba punktów testowych) oraz dla wartości referencyjnych q_i i predykcji $\hat{q}_i$ metamodelu, R^2 definiuje się jako:

$$R^2 = 1 - \frac{\text{SSE}}{\text{SST}} \tag{2.57}$$

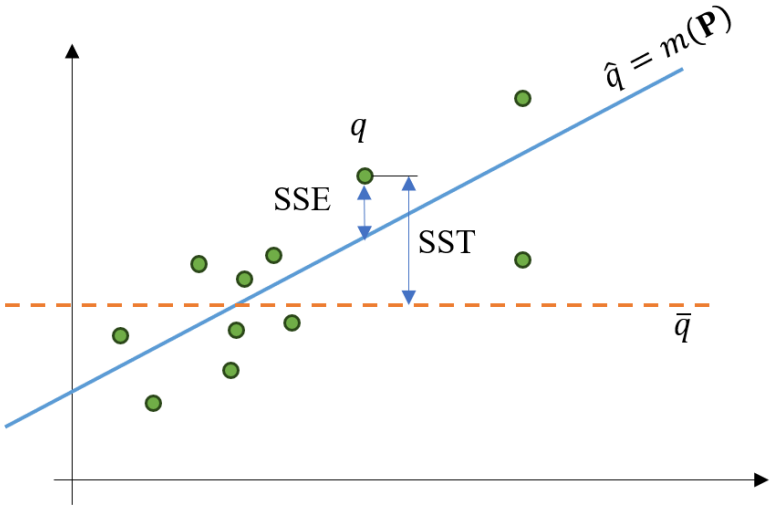
W powyższym SSE (ang. Sum of Squared Errors) jest sumą kwadratów reszt, czyli skumulowanym błędem dopasowania metamodelu:

$$\text{SSE} = \sum_{i=1}^{n} (q_i - \hat{q}_i)^2 \tag{2.58}$$

Natomiast SST (ang. Total Sum of Squares) jest całkowitą sumą kwadratów odchyleń wartości referencyjnych od ich średniej $\bar{q}$ w zbiorze testowym:

$$\text{SST} = \sum_{i=1}^{n} (q_i - \bar{q})^2 \tag{2.59}$$

Zależności tych miar pokazuje rysunek *2.8*.



Rysunek 2.8. Wizualizacja współczynników do obliczania współczynnika determinacji.

Współczynnik determinacji R^2 porównuje błąd metamodelu z „naturalnym rozrzutem" danych. SSE mierzy, jak bardzo predykcje $\hat{q}_i$ różnią się od wartości referencyjnych q_i, a SST

mierzy, jak bardzo same wartości q_i różnią się od swojej średniej $\bar{q}$; dlatego $R^2 = 1$ oznacza dopasowanie idealne, $R^2 = 0$ oznacza, że model nie jest lepszy niż przewidywanie stałej równej średniej $\bar{q}$, a $R^2 < 0$ oznacza, że model jest gorszy niż ta średnia. Jednocześnie R^2 bywa mylące w zadaniach aproksymacji, ponieważ wysoka wartość tej miary nie zawsze oznacza wysoką dokładność predykcji, a w pewnych sytuacjach może wzrastać także wskutek dodawania składników nieistotnych dla jakości predykcji. Z tego powodu w metamodelowaniu równolegle stosuje się miary bezwzględne, wyrażone w skali odpowiedzi q. Miary te są wyznaczane osobno dla zbioru uczącego, walidacyjnego oraz testowego. W dalszym zapisie indeks s oznacza rozpatrywany zbiór danych, gdzie $train$, val oraz $test$ oznaczają odpowiednio zbiór uczący, walidacyjny i testowy. Średni błąd kwadratowy MSE dla zbioru s, zawierającego N_s punktów, ma postać:

$$\mathrm{MSE}_s = \frac{1}{N_\mathrm{s}} \sum_{i=1}^{N_\mathrm{s}} (q_i - \hat{q}_i)^2 \ , s \in \{train, val, test\} \tag{2.60}$$

Pierwiastek z MSE, czyli błąd RMSE, zachowuje jednostkę odpowiedzi q i jest interpretowany jako typowy (średniokwadratowy) poziom odchyłki predykcji:

$$\mathrm{RMSE}_s = \sqrt{\mathrm{MSE}_s} = \sqrt{\frac{1}{N_\mathrm{s}} \sum_{i=1}^{N_\mathrm{s}} (q_i - \hat{q}_i)^2} \tag{2.61}$$

W szczególności wartości $\mathrm{MSE}_{\mathrm{train}}$, $\mathrm{MSE}_{\mathrm{val}}$ oraz $\mathrm{MSE}_{\mathrm{test}}$ oznaczają odpowiednio błąd średniokwadratowy na zbiorze uczącym, walidacyjnym i testowym. Nazywa się je również stratami odpowiednio treningowymi walidacyjnymi oraz testowymi. Analogicznie definiuje się $\mathrm{RMSE}_{\mathrm{train}}$, $\mathrm{RMSE}_{\mathrm{val}}$ oraz $\mathrm{RMSE}_{\mathrm{test}}$.

W dalszych podrozdziałach miary te używane są do porównania metamodeli. Należy w tym miejscu zaznaczyć, że równie dobrym parametrem sprawdzającym predykcję jest samo sprawdzenie predykcji z wynikami modelu wysokiej wierności (np. MES).

2.2.1. Przestrzeń projektowa oraz metody losowania próbek

Planowanie eksperymentów (ang. Design of Experiments, DOE) rozwinęło się jako odpowiedź na ograniczenia podejścia polegającego na zmianie jednego czynnika naraz oraz na problemy z interpretacją danych pozyskanych z badań realizowanych bez zaplanowanego przebiegu. Kluczowym krokiem było sformalizowanie zasad randomizacji, replikacji i blokowania oraz powiązanie ich z analizą wariancji i planami czynnikowymi, co umożliwiło obiektywne wnioskowanie o wpływach czynników i ich interakcjach przy ograniczonej liczbie

prób. Montgomery [56] wskazuje cztery ery rozwoju projektowania eksperymentów: erę rolniczą (Fisher) [26], następnie erę przemysłową napędzaną przez metody RSM i eksperymentowanie sekwencyjne (Box i Wilson) [13], później falę jakościową od końca lat 70. oraz wzrost znaczenia metod komputerowych i narzędzi obliczeniowych, które ułatwiły praktyczne stosowanie bardziej zaawansowanych planów i kryteriów optymalności. W erze przemysłowej metody DOE upowszechniały się szczególnie w przemyśle procesowym, w tym chemicznym, gdzie szybka obserwowalność odpowiedzi i możliwość prowadzenia eksperymentów etapami sprzyjały podejściu sekwencyjnemu [56].

W zastosowaniach obliczeniowych, gdzie eksperyment jest rozwiązaniem problemu sformułowanego jako model numeryczny, rola DOE przesuwa się z kontroli błędów pomiarowych na efektywne i możliwie równomierne pokrycie przestrzeni projektowej rozwiązań [29]. Istotne są plany pozwalające na wypełnianianie przestrzeni rozwiązań (ang. space-filling), projektowane w znormalizowanym obszarze poszukiwań, ponieważ ich celem jest zapewnienie możliwie jednolitej informacji o odpowiedzi w całym obszarze rozważań [27], [78].

Wyróżnia się kilka podstawowych metod planowania eksperymentów:

- Losowe próbkowanie (ang. Random Sampling) polega na generowaniu punktów niezależnie zgodnie z przyjętym rozkładem zmiennych projektowych, bez narzucania dodatkowej struktury geometrycznej planu. Zaletą jest minimalna złożoność, jednak konsekwencją w zadaniach metamodelowania bywa nierównomierne pokrycie obszaru poszukiwań, szczególnie przy małej liczbie punktów. W praktyce prowadzi to do współwystępowania obszarów słabo próbkowanych oraz lokalnych skupisk punktów, co ogranicza jakość danych treningowych i może pogarszać stabilność oraz jakość predykcji metamodelu w słabiej pokrytych regionach.

- Plan pełnoczynnikowy (ang. Full Factorial) tworzy regularną siatkę poziomów w każdym wymiarze, co zapewnia uporządkowaną strukturę punktów i sprzyja klasycznej analizie efektów dla poziomów dyskretnych. W zastosowaniach z ciągłym obszarem poszukiwań kluczową konsekwencją jest jednak szybki wzrost liczności planu wraz z liczbą zmiennych, co przy kosztownych uruchomieniach modelu prowadzi do braku akceptowalności obliczeniowej. Dodatkowo regularna „kratowa" geometria może być mniej efektywna z punktu widzenia własności wypełniania przestrzeni projektowej, ponieważ punkty są rozmieszczone w sposób silnie deterministyczny i nie zawsze zapewniają równomierne próbkowanie całej objętości obszaru poszukiwań.

- Plan ułamkowy (ang. Fractional Factorial) stanowi redukcję liczby uruchomień względem pełnego planu dwupoziomowego poprzez zastosowanie ułamka i odpowiedni dobór struktury

planu. Takie podejście skutecznie obniża koszt eksperymentu, lecz jego konsekwencją jest częściowe nakładanie się informacji o efektach, co wymusza założenia dotyczące pomijalności interakcji wyższych rzędów. W kontekście metamodelowania dla zmiennych ciągłych dodatkowym ograniczeniem jest to, że plany te zachowują „kratowy" charakter i nie są projektowane pierwotnie pod równomierne wypełnienie obszaru poszukiwań, które jest kluczowe dla stabilnego uczenia metamodeli w ujęciu wypełniania przestrzeni projektowej.

- Ciągi Sobola należą do sekwencji quasi-losowych, które w praktyce zapewniają bardziej równomierne pokrycie obszaru poszukiwań niż losowe próbkowanie, szczególnie dla większych liczności próbek. Istotną własnością z punktu widzenia obliczeń sekwencyjnych jest zagnieżdżanie, czyli możliwość rozszerzania planu przez dopisywanie kolejnych punktów bez utraty spójności już wykonanych obliczeń. Konsekwencją jest wysoka użyteczność w sytuacjach, gdy budżet obliczeniowy jest niepewny lub plan jest realizowany etapami, przy czym należy zauważyć, że quasi-losowość nie wymusza wprost podziału zakresu każdej zmiennej na przedziały obsadzone pojedynczymi próbkami (stratyfikacja jednowymiarowa), tak jak Metoda Hipersześcianu Łacińskiego, co może mieć znaczenie przy małej liczbie próbek [27].

- Metoda Hipersześcianu Łacińskiego (Latin Hypercube Sampling, LHS) LHS stanowi plan przeznaczony do ciągłych obszarów poszukiwań, który wymusza równomierne pokrycie projekcji jednowymiarowych przez podział zakresu każdej zmiennej na n równych przedziałów i przypisanie każdemu punktowi unikalnego przedziału w każdym wymiarze. Konsekwencją jest ograniczenie ryzyka sytuacji, w której część zakresu danej zmiennej pozostaje niepróbkowana, co bezpośrednio zwiększa informacyjność danych dla uczenia metamodelu przy ograniczonym budżecie uruchomień. Z tego powodu ta metoda jest najbardziej pożądana z punktu widzenia metamodelowania. Metoda ta została pierwotnie zaproponowana jako sposób doboru wartości zmiennych wejściowych w analizie odpowiedzi kodu komputerowego, przy zachowaniu stratyfikacji zakresów poszczególnych zmiennych [55].

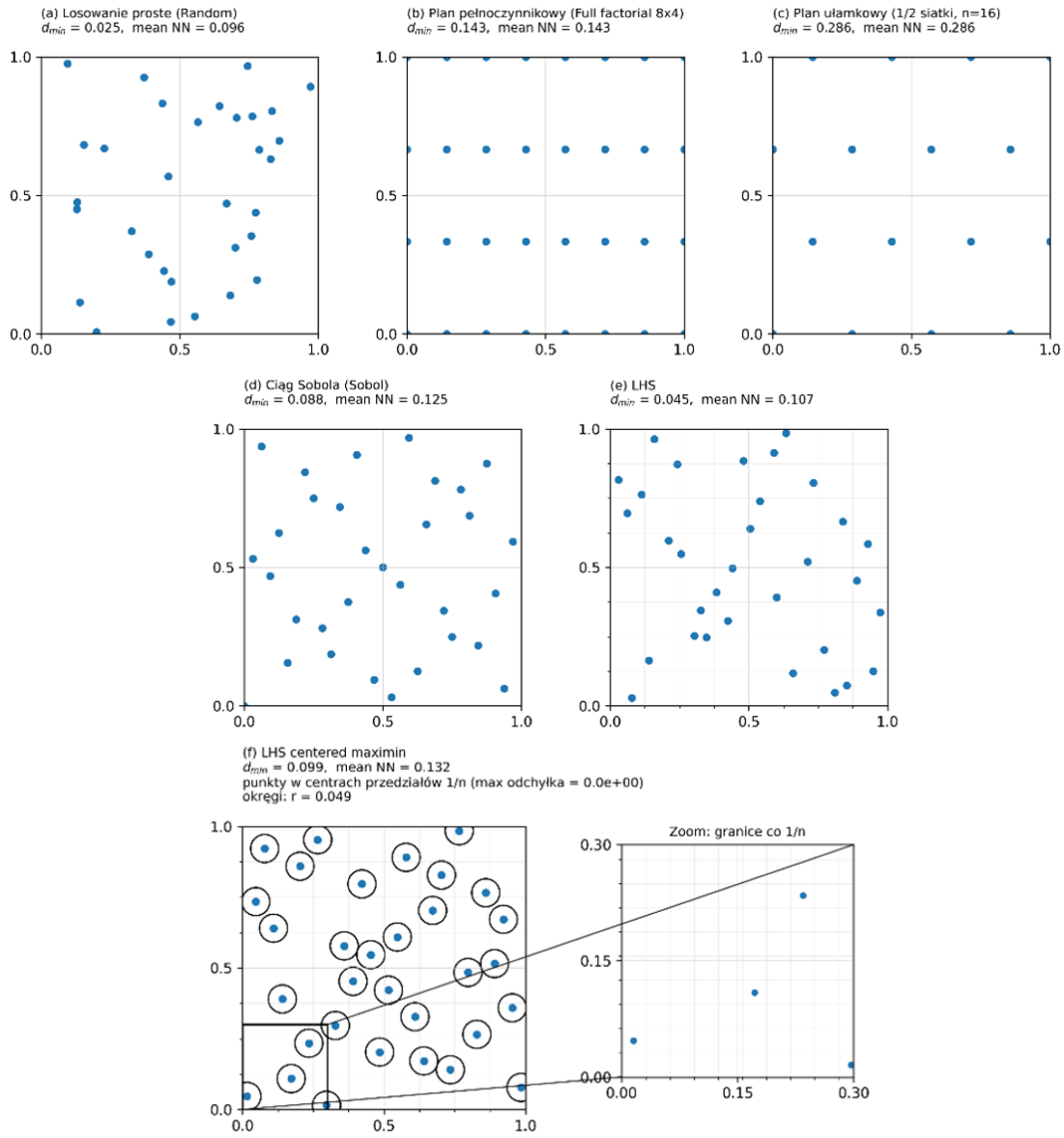
Sam podział zakresu każdej zmiennej na przedziały obsadzone pojedynczymi próbkami nie gwarantuje dobrego wypełnienia przestrzeni wielowymiarowej. Forrester [27] wskazuje, że plan spełniający kryterium stratyfikacji może nadal mieć niekorzystny rozkład, przykładowo, gdy punkty leżą blisko przekątnej przestrzeni zmiennych projektowych, co intuicyjnie nie daje równomiernego „wypełnienia" obszaru poszukiwań. Z tego powodu potrzebna jest miara równomierności umożliwiająca rozróżnianie realizacji „dobrych" i „złych" w obrębie rodziny LHS.

Jedną z metod ulepszonego LHS jest wariant centralny (centralizujący) z maksymalizacją minimalnych odległości (ang. centered maximin). Z jednej strony zachowuje stratyfikację jednowymiarową, a z drugiej poprawia własność wypełniania przestrzeni w przestrzeni wielowymiarowej. Najpierw generuje się realizację LHS przez permutacje poziomów $\{1, \dots, n\}$ w każdej kolumnie oraz skaluje do obszaru poszukiwań znormalizowanego w trybie centrowania, tj. lokując punkty w środkach przedziałów stratyfikacji. Następnie, ponieważ sama stratyfikacja nie gwarantuje dobrego wypełnienia przestrzeni, stosuje się kryterium maksymalizacji minimalnych odległości, które wybiera taki plan, aby maksymalizować najmniejszą odległość pomiędzy dowolną parą punktów, przy czym odległość definiuje się standardowo przez p-normę. Poniższy wzór definiuje odległość d_p pomiędzy dwoma punktami planu doświadczeń w przestrzeni k-wymiarowej, zgodnie z klasyczną definicją odległości opartej o p-normę:

$$d_p\left(\mathbf{P}^{(i_1)}, \mathbf{P}^{(i_2)}\right) = \left(\sum_{j=1}^{k} \mid P_j^{(i_1)} - P_j^{(i_2)} \mid^p\right)^{1/p} \tag{2.62}$$

gdzie $\mathbf{P}^{(i_1)}$oraz $\mathbf{P}^{(i_2)}$oznaczają dwa różne punkty planu, czyli dwie realizacje eksperymentu zapisane jako wektory zmiennych projektowych w przestrzeni $\mathbb{R}^k$. Składowa $P_j^{(i)}$ oznacza j-tą współrzędną i-tego punktu, a więc wartość j-tej zmiennej wejściowej w tym punkcie. Symbol k określa liczbę zmiennych wejściowych, czyli wymiar przestrzeni eksperymentu, natomiast j jest indeksem sumowania po kolejnych wymiarach, przyjmującym wartości $j = 1, \dots, k$. Parametr p określa rodzaj normy. Dla $p \geq 1$ powyższy zapis definiuje normę, a tym samym metrykę odległości. W praktyce najczęściej stosuje się $p = 2$, co odpowiada odległości euklidesowej albo $p = 1$, co odpowiada odległości Manhattan.W kontekście LHS i kryterium *maximin* ta odległość jest liczona dla wszystkich par punktów, a następnie wykorzystywana do oceny „rozsiania" planu przez maksymalizację najmniejszej z tych odległości.

Rysunek 2.9 przedstawia wizualizację metod losowania próbek opisanych wcześniej.



Rysunek 2.9. Wizualizacja wybranych metod losowania próbek.

W konsekwencji wariant centralizujący z maksymalizacją minimalnych można interpretować jako połączenie dwóch elementów koncepcyjnych: po pierwsze, konstrukcji LHS z lokowaniem punktów w środkach przedziałów warstwowania w przestrzeni znormalizowanej, co stabilizuje stratyfikację w każdym wymiarze, oraz po drugie, selekcji realizacji LHS przez maksymalizację minimalnej odległości między punktowej, co poprawia wypełnianie przestrzeni dla $k > 1$. Ta metoda jest najczęściej stosowana oraz zalecana przy użyciu metamodeli.

2.2.2. Wielomianowe aproksymatory

Wielomianowe aproksymatory (regresja wielomianowa) [21], [64] należą do klasycznych narzędzi budowy modeli empirycznych, w których nieznaną zależność wejście-wyjście przybliża się funkcją o zadanej (znanej z góry i ustalonej) strukturze algebraicznej. Intuicyjnie współczynniki wielomianu są wyznaczane w taki sposób, aby minimalizować błąd względem znanych danych, co prowadzi do uzyskania gładkiej powierzchni aproksymującej zależność wejście-wyjście. W przeciwieństwie do interpolacji, model ten nie musi przechodzić dokładnie przez punkty uczące - jego celem jest możliwie wierne odwzorowanie ogólnego trendu. Aproksymatory wielomianowe działają dobrze, gdy dane mają widoczny trend. W przypadku danych bardzo niejednoznacznych, powierzchnia odpowiedzi może ulec bardzo silnym fluktuacjom i gwałtownym zakrzywieniom. Popularność tego metamodelu wynika z prostoty implementacji, interpretowalności współczynników oraz faktu, że po rozwinięciu bazy (np. o potęgi i iloczyny zmiennych) problem pozostaje liniowy względem parametrów modelu, co umożliwia stabilne dopasowanie parametrów metodą najmniejszych kwadratów.

W praktyce metamodelowania często stosuje się modele niskiego rzędu, w szczególności rzędu drugiego, ponieważ stanowią kompromis między elastycznością a liczbą parametrów [56]. Pełny wielomian drugiego rzędu dla n_p parametrów można zapisać jako:

$$\hat{q} = \beta_0 + \sum_{i=1}^{k} \beta_i P_i + \sum_{i=1}^{k} \beta_{ii} P_i^2 + \sum_{ij(\mathrm{i<j})} \beta_{ij} P_i P_j \tag{2.63}$$

gdzie β_0 jest wyrazem wolnym, β_i współczynnikami członów liniowych, β_{ii} współczynnikami członów kwadratowych, a β_{ij} współczynnikami interakcji. Składniki mieszane $P_i P_j$ odpowiadają sytuacjom, w których wpływ jednego parametru na odpowiedź zależy od poziomu innego parametru.

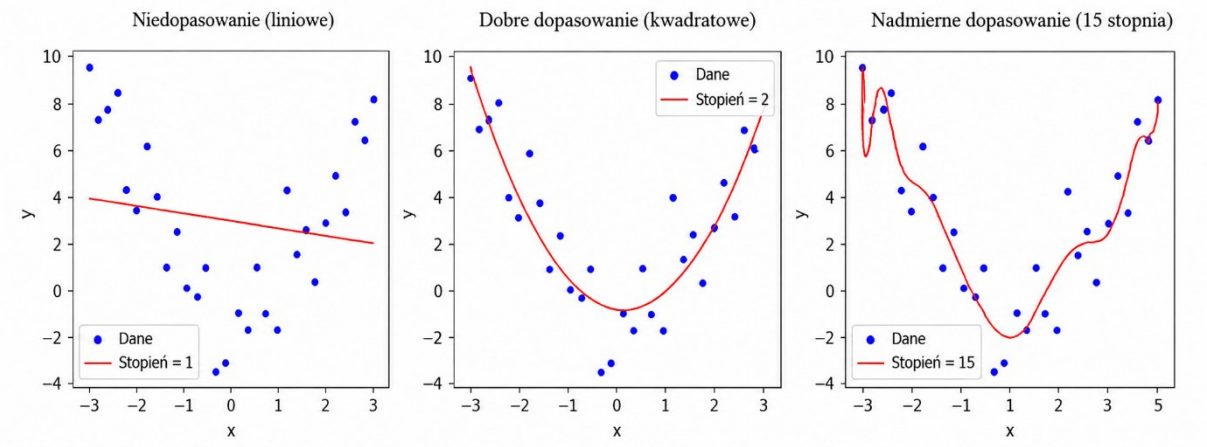
Dopasowanie regresji wielomianowej realizuje się zwykle metodą najmniejszych kwadratów. Niech $\mathbf{W} \in \mathbb{R}^{n_{\mathrm{DOE}} \times n_\beta}$ będzie macierzą regresji zbudowaną z wybranych monomów (kolumny odpowiadają składnikom bazy), a $\mathbf{q} \in \mathbb{R}^{n_{\mathrm{DOE}}}$ wektorem odpowiedzi w punktach planu. Wówczas estymator współczynników ma postać:

$$\widehat{\boldsymbol{\beta}} = (\mathbf{W}'\mathbf{W})^{-1}\mathbf{W}'\mathbf{q} \tag{2.64}$$

gdzie $\mathbf{W}$ oznacza macierz regresji o rozmiarze $n_{DOE} \times n_\beta$, w której wiersze odpowiadają punktom danych wejściowych, natomiast kolumny odpowiadają składnikom bazy, czyli monomom. Symbol $\mathbf{W}'$oznacza transpozycję macierzy $\mathbf{W}$. Iloczyn $\mathbf{W}'\mathbf{W}$ tworzy macierz

o rozmiarze $n_\beta \times n_\beta$, nazywaną macierzą normalną. Wektor $\mathbf{q}$ zawiera wartości odpowiedzi, czyli wyniki obliczone w punktach danych wejściowych i ma rozmiar $n_{DOE} \times 1$. Z kolei $\widehat{\boldsymbol{\beta}}$ oznacza wektor oszacowanych współczynników regresji o rozmiarze $n_\beta \times 1$. Z punktu widzenia stabilności numerycznej kluczowe jest uwarunkowanie $\mathbf{W}'\mathbf{W}$, które pogarsza się m.in. przy dużych różnicach skal parametrów oraz przy współliniowości składników bazy. Dlatego standardowym zabiegiem jest skalowanie lub kodowanie parametrów $\mathbf{P}$ przed budową bazy wielomianowej.

Wzrost stopnia wielomianu zwiększa zdolność dopasowania danych uczących, ale jednocześnie podnosi ryzyko przeuczenia, czyli sytuacji, w której model „uczy się" także przypadkowych fluktuacji danych zamiast stabilnej zależności fizycznej (rysunek 2.10). W modelach wielomianowych przeuczenie wzmacnia się dodatkowo przez szybki przyrost liczby wyrazów bazy i współczynników, co zwiększa wariancję estymacji oraz wrażliwość na szum i korelacje wejść. Z punktu widzenia zastosowań obliczeniowych oznacza to, że sam wysoki poziom dopasowania na zbiorze uczącym nie jest wystarczającym argumentem, aby przyjąć model o wysokim stopniu.



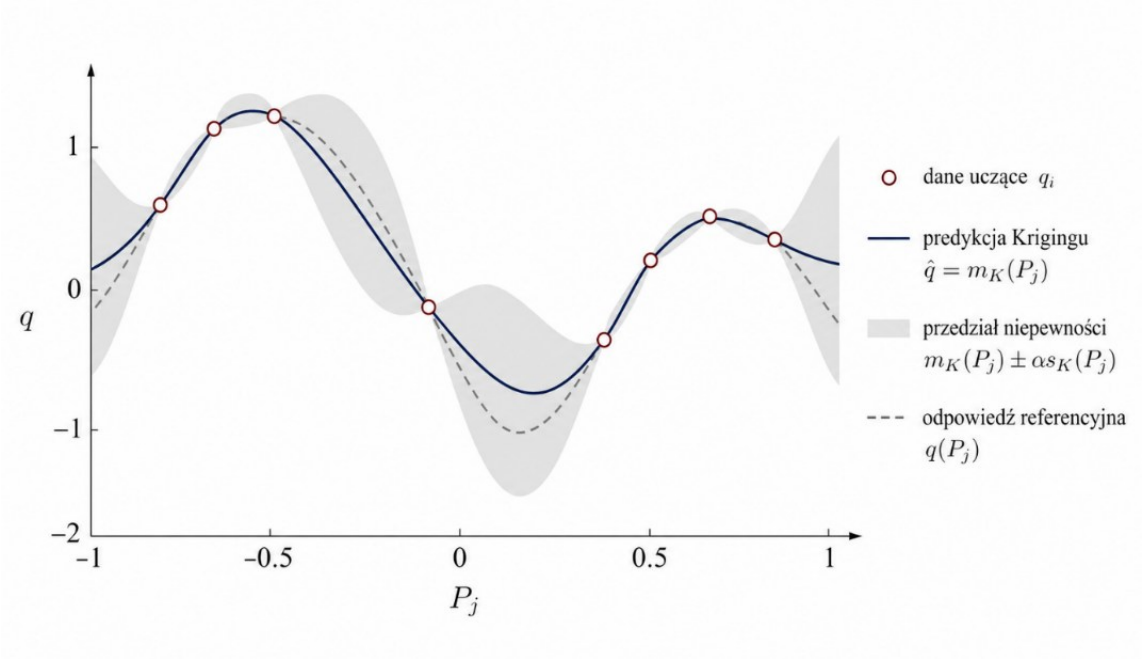
Rysunek 2.10. Wpływ stopnia wielomianu na dopasowanie powierzchni odpowiedzi dla przypadku dwóch zmiennych.

W konsekwencji, przy ograniczonej liczbie próbek oraz przy danych obarczonych szumem, preferowane są modele niskiego stopnia, w szczególności pełny lub zredukowany model drugiego rzędu, o ile diagnostyka reszt i walidacja nie wskazują systematycznych błędów modelu. Dalsze zwiększanie stopnia wielomianu powinno być traktowane jako etap wymagający rygorystycznej walidacji oraz kontroli uwarunkowania numerycznego dopasowania, ponieważ w przeciwnym razie uzyskana poprawa dopasowania może być pozorna i nieprzenośna na nowe przypadki.

W metamodelowaniu opartym na regresji wielomianowej jakość dopasowania nie może być oceniana wyłącznie przez maksymalizację pojedynczej miary na jednym podzbiorze danych, ponieważ taki wybór sprzyja zarówno niedouczeniu (zbyt niski stopień i brak uchwycenia krzywizny), jak i przeuczeniu (zbyt wysoki stopień i dopasowanie do przypadkowych fluktuacji), a ponadto bywa wrażliwy na konkretny podział próbek na zbiory treningowy i walidacyjny. Z punktu widzenia zastosowań inżynierskich, gdzie metamodel ma służyć do predykcji w nowych punktach przestrzeni projektowej, kluczowe jest nie tylko uzyskanie wysokiej wartości miary dopasowania, lecz także wykazanie stabilności tego dopasowania pomiędzy niezależnymi podzbiorami danych, rozumianej jako brak istotnych rozbieżności jakości predykcji pomiędzy treningiem, walidacją i testem. Dlatego w dalszej części pracy wprowadzono kryterium wyboru konfiguracji PNR, które łączy informację o średnim poziomie dopasowania (uśrednienie miary typu R^2 w kilku podzbiorach) z karą za niespójność wyników między tymi podzbiorami, tak aby preferowane były modele jednocześnie dokładne i uogólniające, a nie jedynie dobre dla danych uczących lub przypadkowo korzystne dla konkretnego podziału próbek.

2.2.3. Metoda Kriginga, Kriging

Kriging to interpolacyjna metoda metamodelowania [27], [73], [80]. Intuicyjnie Kriging tworzy powierzchnię odpowiedzi, która dla danych deterministycznych może przechodzić dokładnie przez wyniki znanych symulacji, a pomiędzy nimi estymuje wartości przy założeniu, że punkty bardziej podobne w przestrzeni parametrów mają bardziej zbliżone wyniki niż punkty mniej podobne. Założenie o gładkości powierzchni jest sterowane przez funkcję kowariancji, dzięki czemu model potrafi uogólniać wyniki z ograniczonej liczby prób na całą przestrzeń parametrów. W przypadku deterministycznych eksperymentów komputerowych interpolacyjny charakter Krigingu jest korzystny, ponieważ umożliwia zachowanie zgodności predykcji z punktami uczącymi, jednak zdolność uogólniania zależy od doboru funkcji kowariancji, parametrów modelu oraz rozmieszczenia punktów w przestrzeni projektowej [78]. Dzięki temu, że funkcje estymujące są znane metoda ta, jest w stanie podać jaką niepewność posiada wynik dla przewidzianej wartości (rysunek 2.11).



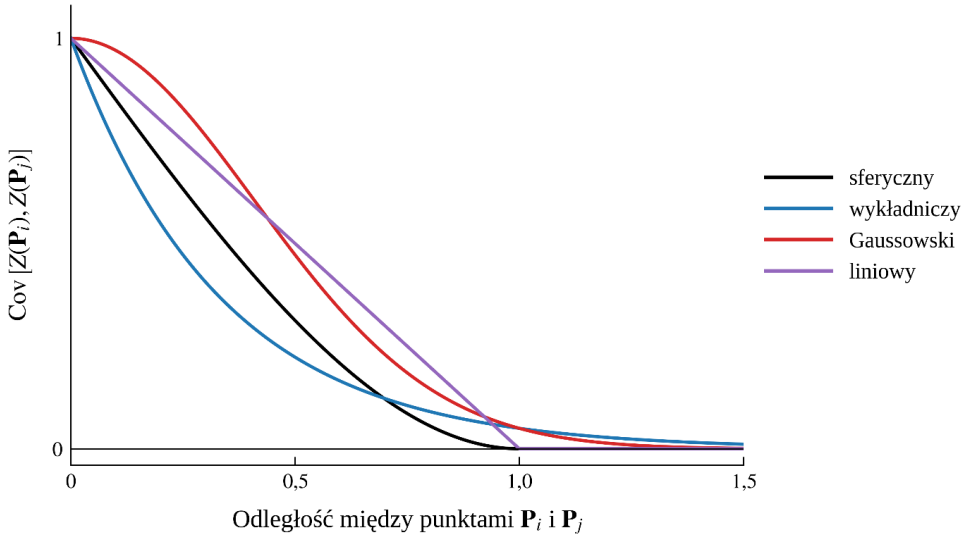
Rysunek 2.11. Przestrzeń niepewności w metamodelu Kriginga.

Metodę najczęściej interpretuje się jako regresję z procesem Gaussa, w której globalny trend $f(\mathbf{P})$ opisuje ogólny poziom odpowiedzi, a składnik losowy $Z(\mathbf{P})$ modeluje lokalne odchylenia skorelowane w przestrzeni parametrów. W tym ujęciu predykcja jest sumą trendu i realizacji procesu Gaussa, a niepewność predykcji wynika z rozkładu warunkowego tego procesu:

$$\hat{q}(\mathbf{P}) = f(P) + Z(P) \tag{2.65}$$

Zapis ten oznacza, że $f(\mathbf{P})$ jest częścią deterministyczną (np. stałą lub wielomianem niskiego stopnia), natomiast $Z(\mathbf{P})$ jest realizacją procesu Gaussa opisaną przez przyjętą funkcję kowariancji, która narzuca stopień gładkości i zasięg zależności przestrzennej między punktami. Dzięki temu, poza wartością predykcji, model dostarcza również informację o niepewności estymacji zależnej od położenia punktu względem danych uczących [73].

Dobór modelu kowariancji lub równoważnie modelu semiwariogramu jest kluczowym elementem strojenia metamodelu, bo determinuje gładkość przewidywanej funkcji, zachowanie w pobliżu zera, które odpowiada czułości na małe zmiany parametrów. Z tego powodu praktyczne zestawy modeli obejmują zarówno warianty bardzo gładkie (Gaussowski), jak i „bardziej szorstkie” (wykładnicze), a także modele o skończonym zasięgu (sferyczne) oraz modele bez progu nasycenia (potęgowy i liniowy). Przykłady zilustrowano na rysunku 2.12.

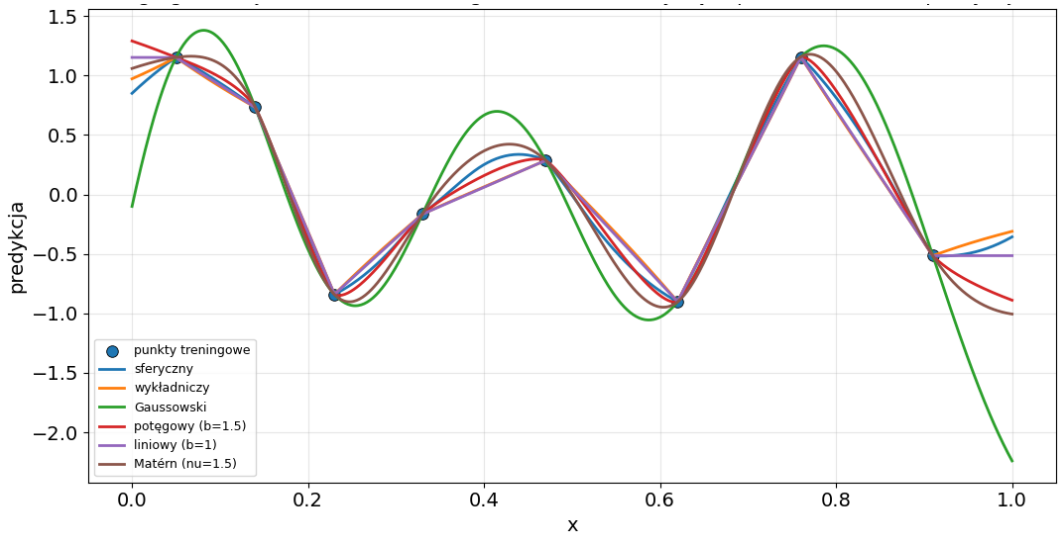


Rysunek 2.12. Przykładowe funkcje kowariancji.

W zapisie funkcji jednej zmiennej dla odległości $h = \| \mathbf{h} \|$(dla znormalizowanych parametrów projektowych) typowe modele semiwariogramu, można zebrać w postaci:

$$\gamma(h)=c_0+c\cdot\begin{cases} 1.5\,\frac{h}{a}-0.5\left(\frac{h}{a}\right)^3 & 0\le h\le a \text{ (sferyczny)},\\ 1 & h>a \text{ (sferyczny)},\\ 1-\exp\left(-\frac{h}{a}\right) & h\ge 0 \text{ (wykładniczy)},\\ 1-\exp\left(-\left(\frac{h}{a}\right)^2\right) & h\ge 0 \text{ (Gaussowski)},\\ \left(\frac{h}{a}\right)^b & h\ge 0,\ 0<b<2 \text{ (potęgowy)},\\ \frac{h}{a} & h\ge 0 \text{ (liniowy, : } b=1). \end{cases} \tag{2.66}$$

W elementarnym przykładzie jednowymiarowym, konsekwencje stosowania różnych funkcji kowariancji przedstawiono na rysunku 2.13:



Rysunek 2.13. Konsekwencje użycia różnych funkcji kowariancji dla modelu Kriging w funkcji jednej zmiennej.

W praktyce wybór modelu kowariancji jest kompromisem między gładkością powierzchni odpowiedzi a zdolnością odtwarzania lokalnych zmian, co bezpośrednio wpływa na predykcję oraz na rozkład niepewności. Dlatego parametrów i typu funkcji nie dobiera się „z definicji", tylko stroi na danych, np. przez dopasowanie do semiwariogramu empirycznego oraz walidację krzyżową jakości predykcji.

2.2.4. Sztuczne Sieci Neuronowe

Sztuczne sieci neuronowe (SSN) rozwijane były od prostych modeli perceptronu do współczesnych, wielowarstwowych architektur zdolnych do aproksymacji złożonych zależności nieliniowych. Klasyczny perceptron Rosenblatta (1957) [74] był formalizacją neuronu o ważonych wejściach i pojedynczym wyjściu, stanowiąc punkt wyjścia do dalszych uogólnień w kierunku sieci wielowarstwowych [25] .

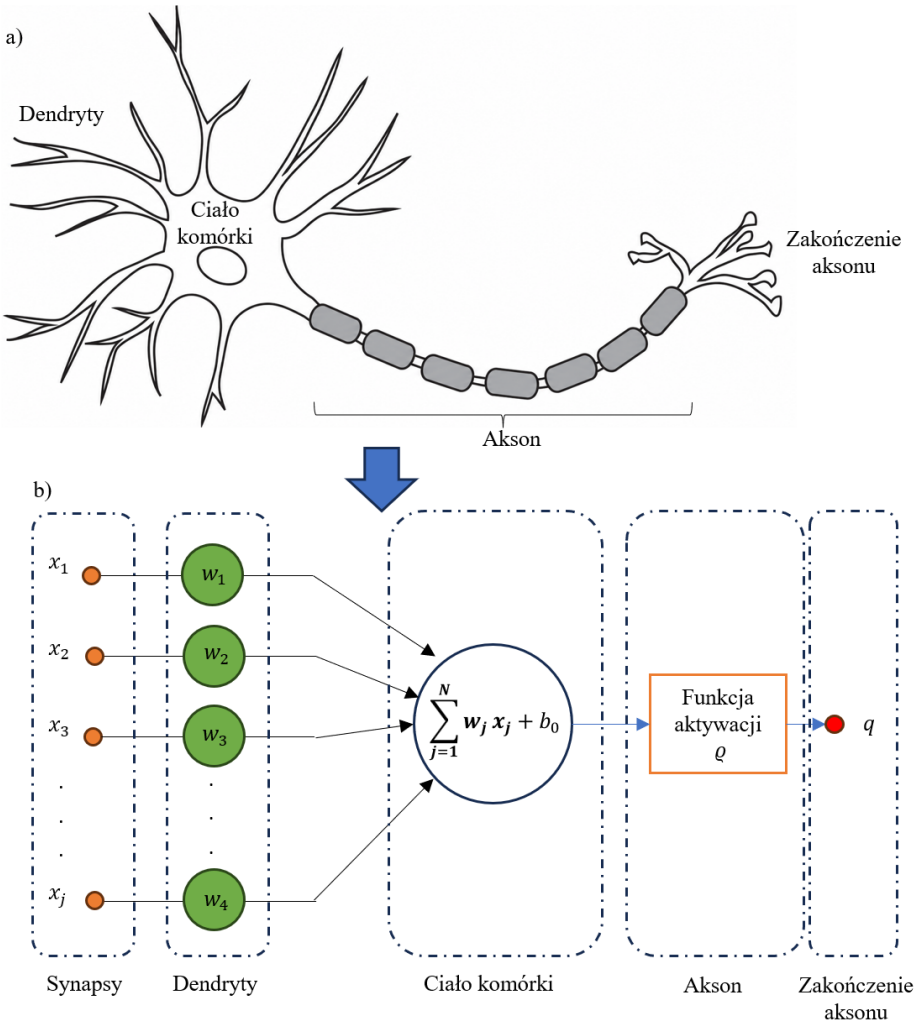
Intuicyjnie sztuczna sieć neuronowa tworzy powierzchnię odpowiedzi, która dopasowuje się do danych treningowych, ale nie musi przechodzić dokładnie przez każdy punkt. Jej celem jest uchwycenie ogólnego kształtu zależności między wejściem a wyjściem - czyli aproksymacja. W zależności od liczby warstw, neuronów i sposobu uczenia, sieć może tworzyć bardzo złożone, nieregularne powierzchnie odpowiedzi. Jeśli jednak sieć jest zbyt rozbudowana i nie ma ograniczeń (np. regularyzacji), może zacząć „uczyć się na pamięć" i zachowywać jak interpolator - przechodząc dokładnie przez punkty uczące, ale tracąc zdolność uogólniania. Dlatego sieci neuronowe są z natury aproksymatorami, ale ich zachowanie zależy od sposobu konstrukcji i treningu. W przeciwieństwie do aproksymatorów wielomianowych SSN nie mają

z góry narzuconej postaci funkcji aproksymującej, takiej jak stopień wielomianu i zestaw jego składników. Kształt zależności między zmiennymi wejściowymi i odpowiedzią jest wyznaczany w procesie uczenia, co stanowi ich istotną zaletę, ponieważ pozwala dostosować model do złożonych danych. Jednocześnie wymaga to kontroli zjawiska przeuczenia.

Podstawowym elementem SSN jest sztuczny neuron, będący silnie uproszczonym modelem neuronu biologicznego. Istota działania jest następująca: sygnały wejściowe są ważone, sumowane, a następnie przekształcane przez funkcję aktywacji; dodatkowo stosuje się składnik bias, który realizuje przesunięcie (ang. offset) argumentu funkcji aktywacji. Matematycznie, dla neuronu o wejściach x_j i wagach w_j, wyjście q definiuje się jako wynik działania funkcji aktywacji ϱ na sumę ważoną sygnałów wejściowych powiększoną o składnik bias:

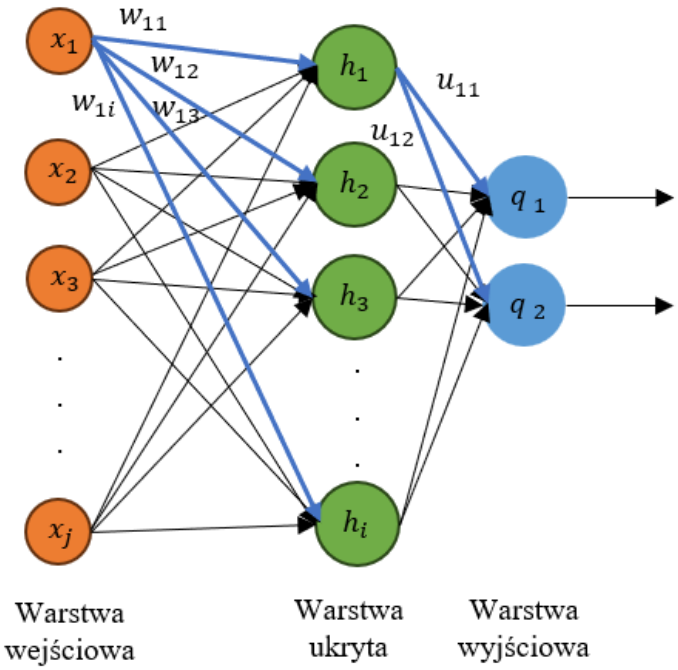
$$q = \varrho\left(\sum_{j=1}^{N} \boldsymbol{w_j}\, \boldsymbol{x_j} + b_0\right) \tag{2.67}$$

gdzie w_j jest wagą dla j-tego wejścia, x_j jest j-tym sygnałem wejściowym, q jest wyjściem neuronu, ϱ jest funkcją aktywacji, a b_0 jest składnikiem bias (przesunięciem), w praktyce realizującym przesunięcie argumentu funkcji aktywacji. Model neuronu oraz jego matematyczną interpretację zobrazowano na rysunku 2.14.



Rysunek 2.14. a) Schemat neuronu b) matematyczny model neuronu [30].

Sieć neuronowa powstaje przez połączenie wielu neuronów w strukturę warstwową, w której sygnały z wyjść neuronów jednej warstwy są wejściami neuronów warstwy następnej. W ogólnym przypadku połączenia między neuronami nie muszą mieć charakteru pełnego, czyli typu każdy z każdym, a sposób ich definiowania może stanowić jeden z parametrów architektury modelu. Typowy układ ma warstwę wejściową (nieliczoną jako warstwa neuronów obliczeniowych), co najmniej jedną warstwę ukrytą oraz warstwę wyjściową. Taki wariant, w którym sygnał przepływa w jednym kierunku, jest klasycznym przykładem sieci typu sprzężenie do przodu (ang. feedforward); najczęściej stosowaną architekturą w tej klasie jest perceptron wielowarstwowy (MLP). Taką strukturę przedstawiono na rysunku 2.15.



Rysunek 2.15. Wielowarstwowa SSN.

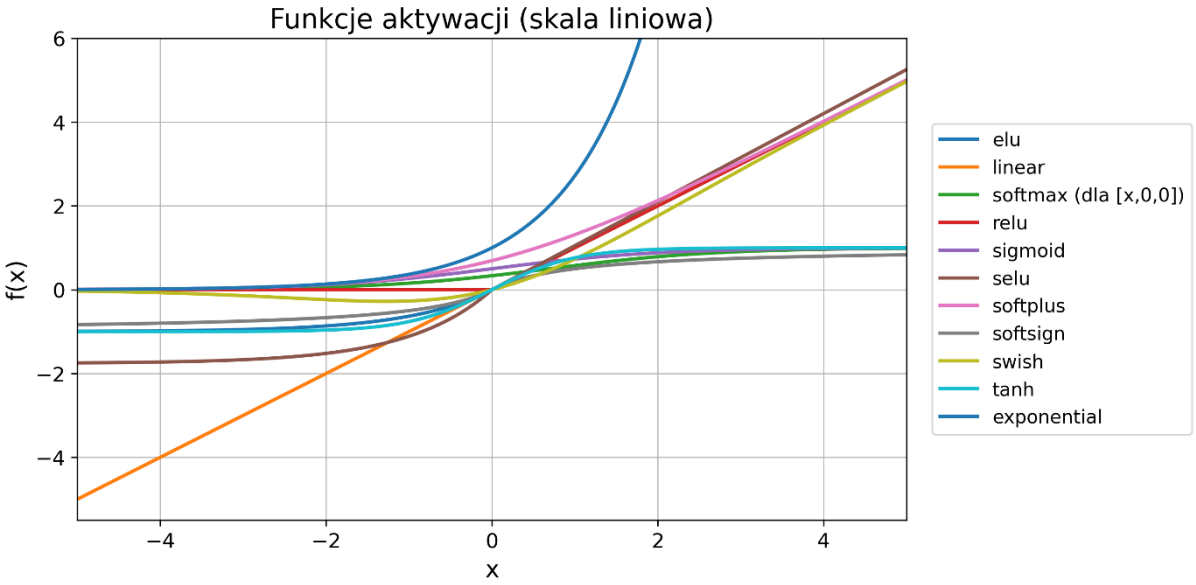
W kontekście metamodelowania architektura MLP jest używana jako parametryczna funkcja $f(\mathbf{P};\Xi)$, gdzie Ξ oznacza komplet wag i biasów wszystkich neuronów. Dla wektora wejściowego $\mathbf{x} \in \mathbb{R}^n$ sieć zwraca predykcję $\hat{q}$ odpowiedzi modelu:

$$\hat{q} = f(\mathbf{P};\Xi) \tag{2.68}$$

Powyższy zapis jest celowo ogólny, konkretna postać $f(\cdot)$ wynika z przyjętej liczby warstw, liczby neuronów w kolejnych warstwach oraz doboru funkcji aktywacji, natomiast mechanizm elementarny zawsze opiera się o neuron w postaci sumy ważonej z biasem oraz nieliniowości $\varrho(\cdot)$.

Funkcja aktywacji w neuronie przekształca sumę ważoną wejść na sygnał wyjściowy, a więc decyduje o tym, czy pojedyncza warstwa pozostaje modelem liniowym, czy cała sieć uzyskuje zdolność modelowania nieliniowych zależności wejście-wyjście. Dobór funkcji aktywacji silnie wpływa również na przebieg uczenia, ponieważ typowe metody optymalizacji oparte o gradient wymagają funkcji różniczkowalnych. Pewne funkcje, zwłaszcza sigmoidalne, mogą wchodzić w zakres nasycenia, w którym pochodna funkcji jest bliska zeru, co osłabia gradienty i pogarsza stabilność oraz szybkość zbieżności. W konsekwencji funkcja aktywacji jest hiperparametrem wpływającym na zdolność modelu do odwzorowania złożonych

zależności wejście-wyjście, stabilność procesu uczenia oraz koszt obliczeniowy. Przykładowe funkcje aktywacji przedstawia rysunek 2.16.



Rysunek 2.16. Przykładowe funkcje aktywacji.

Uczenie SSN jest formułowane jako problem minimalizacji błędu na zbiorze uczącym. Dla regresji typowym wyborem jest błąd kwadratowy (sumaryczny na zbiorze), zapisany jako [25]:

$$E^2 = \sum_{i=1}^{n} \left(q_i - \hat{q}(x_i)\right)^2 \tag{2.69}$$

gdzie q_i jest wartością referencyjną, $\hat{q}(x_i)$ predykcją modelu dla próbki x_i, a n licznością zbioru (w metamodelowaniu x_i odpowiada wektorowi parametrów $\mathbf{P}^{(i)}$).

W metamodelowaniu dla modeli kosztownych obliczeniowo kluczowe znaczenie ma kontrola uogólniania, sieć może bardzo dobrze dopasować zbiór uczący, a jednocześnie pogorszyć predykcję poza nim (nadmierne dopasowanie). Z tego powodu standardem jest podział danych na zbiory: uczący, walidacyjny i testowy, a także stosowanie regularyzacji i kryteriów zatrzymania uczenia. W praktyce inżynierskiej, gdzie liczba próbek jest ograniczona, preferuje się architektury o umiarkowanej liczbie parametrów, świadomy dobór funkcji aktywacji oraz procedury walidacyjne oparte o miary jakości dopasowania.

Odrębną klasą decyzji, istotną dla powtarzalności wyników, są hiperparametry, czyli wielkości sterujące procesem uczenia i strukturą sieci, które nie są wyznaczane bezpośrednio metodą gradientową. Do najważniejszych należą: liczba warstw i neuronów, funkcje aktywacji,

rozmiar podzbioru danych wykorzystywanego w pojedynczej aktualizacji wag, maksymalna liczba przejść przez zbiór uczący, kryteria zatrzymania, a także techniki regularyzacyjne, np. losowe wygaszanie neuronów (ang. dropout). Ich dobór realizuje się przez strategię przeszukiwania przestrzeni hiperparametrów, m.in. przeszukiwanie siatki, losowe próbkowanie oraz metody optymalizacji bayesowskiej.

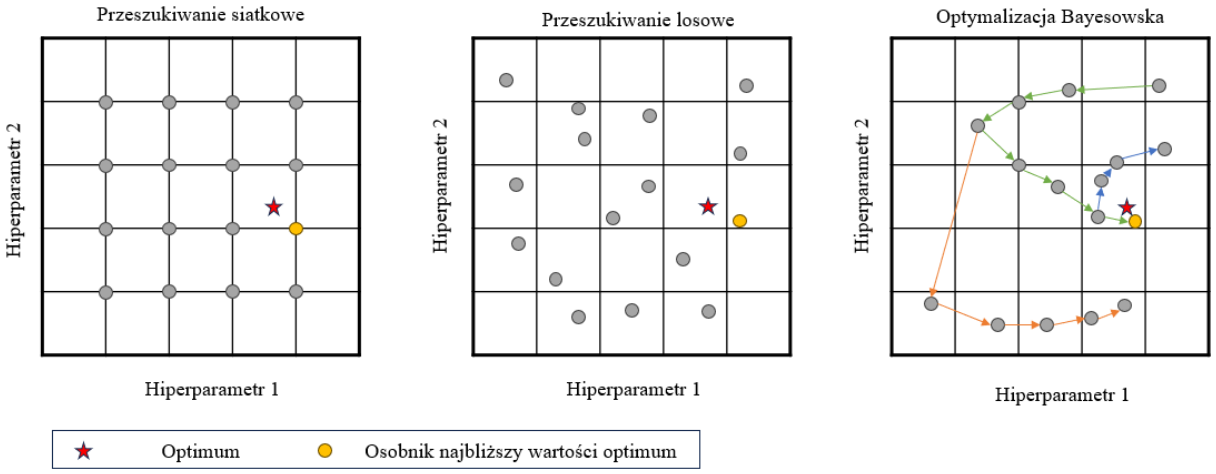
Uczenie SSN polega na takim doborze parametrów $\boldsymbol{\Xi}$, czyli wag i składników bias, aby zminimalizować błąd między wartościami referencyjnymi q_i a predykcjami sieci $\hat{q}\left(\mathbf{x}^{(i)}\right)$. W praktyce oznacza to iteracyjną aktualizację parametrów modelu na podstawie wartości funkcji błędu, której przykładową postać przedstawiono równaniem (2.69). Dla sieci wielowarstwowych wpływ poszczególnych wag na błąd wyznacza się z wykorzystaniem propagacji wstecznej, co umożliwia aktualizację parametrów w kolejnych warstwach modelu. Proces uczenia powinien być kontrolowany na zbiorze walidacyjnym, ponieważ zbyt dobre dopasowanie do zbioru uczącego może prowadzić do przeuczenia i pogorszenia jakości predykcji dla nowych punktów przestrzeni projektowej.

2.2.5. Dobór parametrów metamodeli

W niniejszej pracy zastosowano trzy metody metamodelowania: regresję wielomianową (rozdział 2.2.2), metodę Kriginga (2.2.3) oraz sztuczne sieci neuronowe opisane w rozdziale (2.2.4).

Wraz z rosnącą popularnością sztucznych sieci neuronowych w różnych dziedzinach przemysłu w ostatnich latach opracowano szereg narzędzi umożliwiających optymalizację hiperparametrów [1]. Do najczęściej stosowanych metod strojenia hiperparametrów należą metoda przeszukiwania siatką (ang. *Grid Search)* oraz losowe przeszukiwanie (ang. *Random Search)* [11]. Przeszukiwanie siatkowe pozwala na pełne przeszukanie przestrzeni konfiguracji, jednak charakteryzuje się wysoką złożonością obliczeniową. Z kolei przeszukiwanie losowe jest mniej dokładne (istnieje ryzyko pominięcia konfiguracji optymalnej), lecz znacznie ogranicza koszt obliczeń. W praktycznych zastosowaniach coraz częściej wykorzystuje się adaptacyjne metody strojenia hiperparametrów, w których kolejne konfiguracje nie są dobierane z góry (jak w przeszukiwaniu siatkowym) ani losowane niezależnie (jak w przeszukiwaniu losowym), lecz są proponowane na podstawie wyników uzyskanych w poprzednich próbach. Do tej grupy należą w szczególności metody optymalizacji bayesowskiej (np. podejścia oparte na modelowaniu zależności między hiperparametrami a wartością funkcji celu i wyborze kolejnych punktów w sposób ukierunkowany na poprawę jakości) oraz algorytmy sekwencyjnego modelowania i próbkowania (ang. TPE, Tree-structured Parzen Estimator), które efektywnie koncentrują obliczenia w obiecujących

obszarach przestrzeni poszukiwań [1]. Istotnym uzupełnieniem takich podejść jest mechanizm wczesnego odrzucania prób (ang. *early stopping/pruning*), polegający na przerywaniu konfiguracji, które na wczesnym etapie uczenia nie wykazują potencjału poprawy metryki walidacyjnej, co pozwala istotnie ograniczyć koszt obliczeniowy. Metody te są dostępne w postaci współczesnych bibliotek do optymalizacji hiperparametrów, które dodatkowo zapewniają obsługę przestrzeni warunkowych, rejestrowanie wyników i możliwość równoległego uruchamiania prób. Przedstawione trzy strategie przeszukiwania ilustruje rysunek 2.17.



Rysunek 2.17. Trzy wybrane metody strojenia hiperparametrów.

Sztuczne sieci neuronowe wymagają dostrojenia wielu hiperparametrów. W analizie należy uwzględnić różne funkcje aktywacji: *ReLU*, *tanh*, *sigmoid*, *softmax*, *softplus*, *softsign*, *SELU*, *ELU*, *exponential*, *linear* oraz *swish*. Na zbiorze danych należy zastosować techniki regularyzacji, w tym metodę wygaszania (ang. *dropout)*. Wygaszanie jest techniką regularyzacji stosowaną podczas uczenia sztucznych sieci neuronowych w celu ograniczenia przeuczenia i poprawy zdolności uogólniania modelu na danych niewidzianych [87]. Mechanizm polega na losowym wyłączaniu w trakcie treningu części neuronów lub połączeń z zadanym prawdopodobieństwem, co powoduje, że w kolejnych iteracjach uczenia wykorzystywane są różne „przerzedzone" warianty tej samej sieci. W efekcie ograniczana jest współzależność neuronów (ang. co-adaptation) i wymuszane jest uczenie bardziej rozproszonych reprezentacji, mniej zależnych od pojedynczych neuronów oraz mniej wrażliwych na przypadkowe fluktuacje w zbiorze uczącym. Interpretacyjnie działanie wygaszania można traktować jako przybliżenie współpracy wielu podsieci współdzielących wagi, gdzie w fazie wnioskowania wykorzystywana jest pełna architektura odpowiadająca uśrednionemu zachowaniu tych podsieci. W praktyce zastosowanie wygaszenia zmniejsza wariancję modelu i często redukuje różnicę pomiędzy jakością dopasowania na zbiorze

treningowym i walidacyjnym, przy czym dobór prawdopodobieństwa wygaszeń oraz miejsca w architekturze wymaga strojenia empirycznego, ponieważ zbyt silny współczynnik wygaszenia może spowalniać zbieżność lub prowadzić do niedouczenia. W procesie doboru hiperparametrów sztucznej sieci neuronowej jakość kolejnych konfiguracji ocenia się z wykorzystaniem współczynnika determinacji R^2 oraz straty walidacyjnej MSE_{val}, zgodnie z definicjami przedstawionymi w rozdziale 2.2. Zbiór walidacyjny nie uczestniczy bezpośrednio w aktualizacji wag modelu, lecz służy do kontroli zdolności uogólniania oraz do identyfikacji przeuczenia. Po wyborze konfiguracji jakość modelu sprawdza się dodatkowo na zbiorze testowym, wyznaczając stratę testową MSE_{test}. Wartość MSE_{test} może być raportowana w tabelach porównawczych jako informacja kontrolna, jednak nie stanowi kryterium strojenia hiperparametrów.

W przypadku metamodelu Kriginga (opisanego w rozdziale 2.2.3) kluczowym etapem jest selekcja odpowiedniej funkcji wariogramu (modelu wariogramowego), opisującej zależność wariancji od odległości przestrzennej pomiędzy próbkami. Należy do nich pięć standardowych modeli: liniowy, potęgowy, Gaussa, sferyczny oraz wykładniczy. W celu przeszukania wszystkich możliwych konfiguracji zmiennych można zastosować z powodzeniem metodę przeszukiwania siatkowego z uwagi na stosunkowo niewielką liczbę możliwych konfiguracji do sprawdzenia. Dla każdej konfiguracji metamodelu Kriginga jakość aproksymacji ocenia się przez porównanie wartości przewidywanych z wartościami referencyjnymi. W tym celu wykorzystuje się współczynnik determinacji R^2 oraz błąd średniokwadratowy na zbiorze walidacyjnym MSE_{val}. Po wyborze konfiguracji modelu wariogramowego jakość wybranego wariantu sprawdza się dodatkowo na zbiorze testowym, wyznaczając MSE_{test}. W przypadku Kriginga pojęcie straty oznacza wartość błędu predykcji na wydzielonym zbiorze danych, a nie iteracyjnie minimalizowaną funkcję uczenia analogiczną do stosowanej w sztucznych sieciach neuronowych.

W konstrukcji metamodelu opartego na regresji wielomianowej (ang. Polynomial Regression) zasadniczym elementem procesu strojenia hiperparametrów jest wybór odpowiedniego stopnia wielomianu. Zbyt niski stopień może prowadzić do *niedouczenia* modelu, czyli braku zdolności do odwzorowania istotnych zależności w danych. Z kolei zbyt wysoki stopień wielomianu zwiększa ryzyko przeuczenia i nadmiernej złożoności obliczeniowej. W praktyce optymalny stopień wielomianu określa się z wykorzystaniem walidacji krzyżowej lub dedykowanego zbioru walidacyjnego. Popularne techniki strojenia hiperparametrów w regresji wielomianowej obejmują metody przeszukiwania siatkowego oraz losowego. Jednakże przy dużej liczbie potencjalnych konfiguracji korzystniejsze może być zastosowanie metod bardziej zaawansowanych, takich jak optymalizacja bayesowska, umożliwiających efektywniejsze identyfikowanie obiecujących zestawów parametrów.

Proces strojenia przeprowadza się iteracyjnie, testując modele o różnych stopniach wielomianu i parametrach regularyzacji, a następnie wybierając rozwiązanie zapewniające najlepsze wyniki - tj. najwyższą wartość współczynnika determinacji R^2 przy najniższym błędzie walidacyjnym.

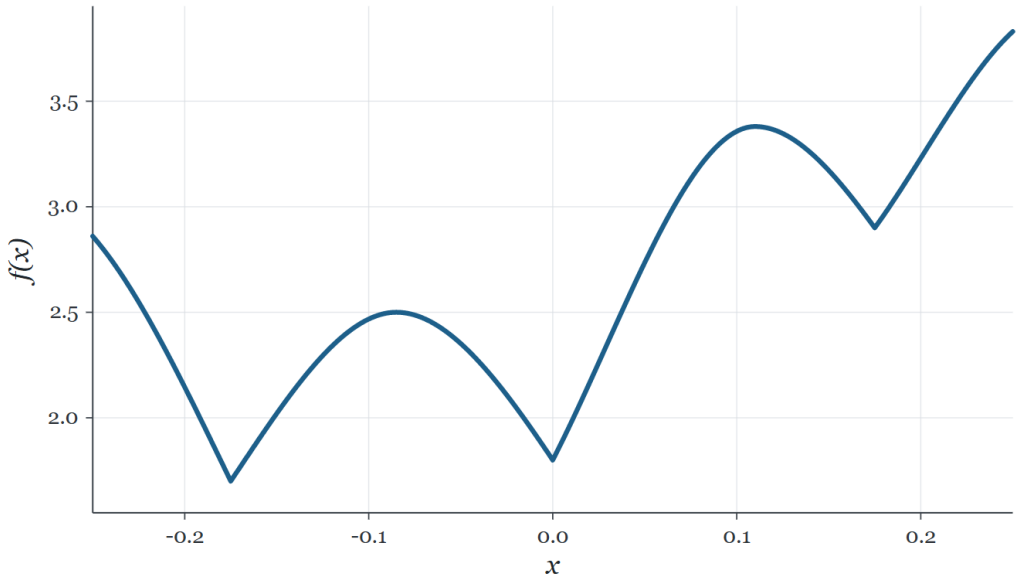
W celu wyłonienia najbardziej wiarygodnego modelu regresji wielomianowej opracowano autorskie kryterium rankingowe, zamiast prostego wyboru konfiguracji o najwyższym R^2_{test}. Dla każdego modelu (rozdział określonego przez stopień wielomianu i sposób podziału danych) obliczano trzy miary jakości: R^2_{train}, R^2_{val} oraz R^2_{test}. Ostateczny wynik stanowiła średnia arytmetyczna z tych trzech wartości, skorygowana o różnice bezwzględne pomiędzy nimi. Inaczej ujmując jest to średnia arytmetyczna karana za różnicę pomiędzy zbiorami. Matematycznie miarę tę zdefiniowano jako:

$$R^2_{\text{PNR}} = \frac{R^2_{\text{train}} + R^2_{\text{val}} + R^2_{\text{test}}}{3} - \left|R^2_{\text{train}} - R^2_{\text{val}}\right| - \left|R^2_{\text{val}} - R^2_{\text{test}}\right| \tag{2.70}$$

Takie podejście preferuje modele, które nie tylko osiągają wysoką dokładność predykcji, lecz również wykazują spójność wyników we wszystkich podzbiorach danych, co minimalizuje ryzyko przeuczenia i ograniczonej zdolności uogólniania. Konfiguracja o najwyższej wartości wskaźnika R^2_{PNR} zostanie uznana za końcowy model. Zapewnia to wybór stopnia wielomianu, który jest jednocześnie dokładny i odporny na dane nieznane w procesie uczenia.

2.3. Ewolucyjny Algorytm Inspirowany Zjawiskami Kwantowymi

W kontekście optymalizacji konstrukcji ocenianej kosztowną symulacją numeryczną, funkcja celu bywa nieliniowa, wielomodalna i niegładka, a jej wartości mogą zależeć od zjawisk kontaktowych oraz modeli materiałowych. Taką funkcję pokazuje rysunek 2.18.



Rysunek 2.18. Przykład nieliniowej, wielomodalnej i niegładkiej funkcji.

Z tego względu, obok metod klasycznych (np. gradientowych), w praktyce inżynierskiej stosuje się metody przybliżone i metaheurystyki, w tym metody populacyjne. W literaturze dotyczącej rozważanego przypadku wskazuje się przykładowo takie podejścia jak algorytmy ewolucyjne, algorytmy rojowe [15].

W ramach metod populacyjnych algorytmy kwantowo-inspirowane można traktować jako szczególną odmianę algorytmów ewolucyjnych uruchamianych na komputerze klasycznym, w których istotą jest probabilistyczna reprezentacja populacji inspirowana pojęciem kubitu. Istotne doprecyzowanie metodologiczne polega na tym, że QEA, mimo inspiracji formalizmem mechaniki kwantowej, nie jest algorytmem kwantowym w sensie uruchamiania na komputerze kwantowym, lecz algorytmem dla komputera klasycznego [46], [100].

Podstawowym obiektem tej reprezentacji jest kubit, którego stan w bazie obliczeniowej opisuje superpozycja stanów $|0\rangle$ oraz $|1\rangle$, z wagami (amplitudami) α_Q oraz β_Q:

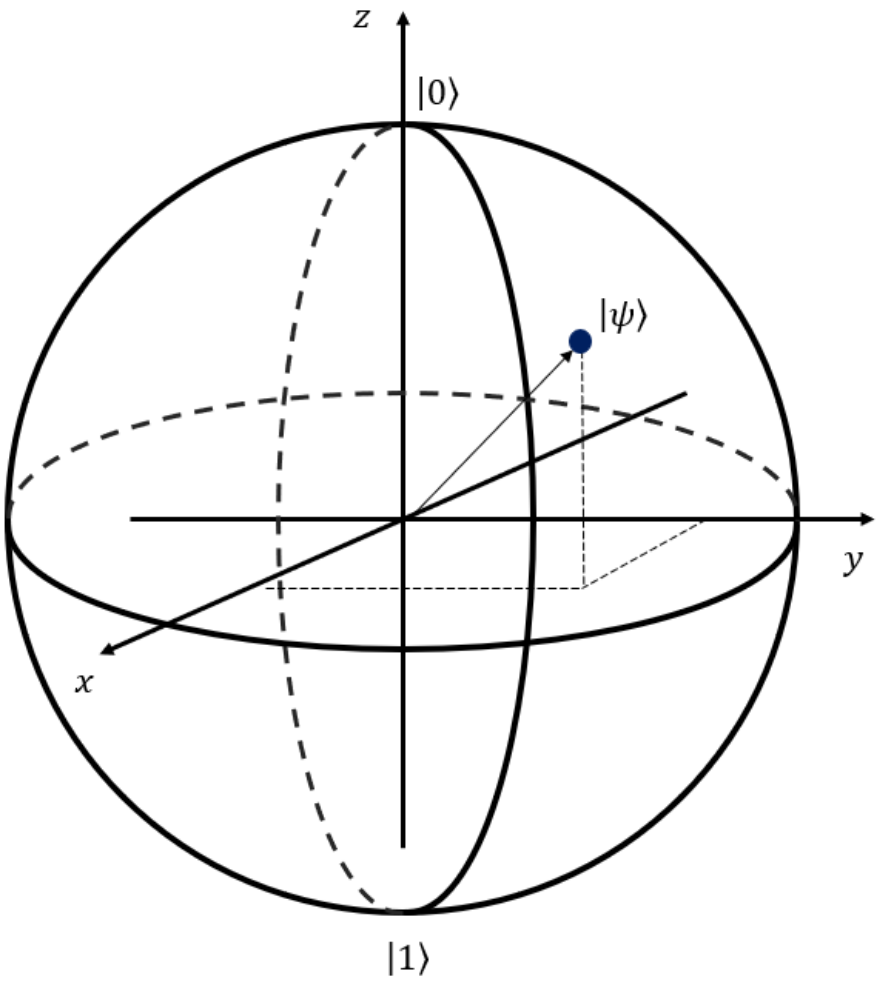
$$|\psi\rangle = \alpha_Q|0\rangle + \beta_Q|1\rangle \tag{2.71}$$

Amplitudy spełniają warunek normalizacji, co umożliwia interpretację probabilistyczną:

$$|\alpha_Q|^2 + |\beta_Q|^2 = 1 \tag{2.72}$$

Interpretacyjnie oznacza to, że w kroku odczytu uzyskuje się wartość binarną, przy czym $|\alpha_Q|^2$ odpowiada prawdopodobieństwu uzyskania stanu 0, a $|\beta_Q|^2$ prawdopodobieństwu

uzyskania stanu 1. Geometrię tego zapisu i sens „sterowania” stanem przez operacje typu bramki dobrze ilustruje reprezentacja na sferze Blocha (rysunek 2.19).



Rysunek 2.19. Sfera Blocha [47].

W algorytmie QEA osobnik o długości m_{DB} jest wektorem m_{DB} kubitów (zapisanym jako macierz $2 \times m_{DB}$), gdzie α_{Q_j} oraz β_{Q_j} są amplitudami j-tego kubitu:

$$\mathbf{q}_{\boldsymbol{bit_j}} = \begin{bmatrix} \alpha_{Q_1} & \alpha_{Q_2} & \cdots & \alpha_{Q_{m_{DB}}} \\ \beta_{Q_1} & \beta_{Q_2} & \cdots & \beta_{Q_{m_{DB}}} \end{bmatrix} \tag{2.73}$$

Taka reprezentacja jest kluczowa numerycznie, ponieważ nie przechowuje pojedynczego deterministycznego rozwiązania, lecz parametry rozkładów, z których generuje się (odczytuje) rozwiązania klasyczne w kolejnych iteracjach.

W ujęciu implementacyjnym, w zastosowaniu do problemów o zmiennych rzeczywistych, grupy kubitów przypisuje się do poszczególnych zmiennych projektowych, a po odczycie

wektor bitów mapuje się na wartości typu zmiennoprzecinkowych w dozwolonych przedziałach.

Mechanizm kierunkowej aktualizacji populacji w QEA realizuje się przez modyfikację amplitud α_Q i β_Q, a więc przez sterowanie prawdopodobieństwami przyszłych odczytów. W klasycznej formulacji Han i Kim [46] rolę operatora wariacyjnego pełni bramka obrotu (ang. rotation gate), opisana macierzą zależną od przyrostu kąta $\Delta\theta_{Q_i}$:

$$U\left(\Delta\theta_{Q_i}\right) = \begin{bmatrix} cos\left(\Delta\theta_{Q_i}\right) & -sin\left(\Delta\theta_{Q_i}\right) \\ sin\left(\Delta\theta_{Q_i}\right) & cos\left(\Delta\theta_{Q_i}\right) \end{bmatrix} \tag{2.74}$$

W praktyce algorytmu znak i wartość przyrostu kąta $\Delta\theta_{Q_i}$ dobiera się na podstawie relacji pomiędzy bieżącym rozwiązaniem a rozwiązaniem najlepszym lub referencyjnym w populacji. Celem tej operacji jest zwiększanie prawdopodobieństwa generowania konfiguracji zgodnych z najlepszymi osobnikami przy jednoczesnym zachowaniu zdolności eksploracji przestrzeni rozwiązań. W ujęciu ogólnym pętla algorytmu QEA obejmuje inicjalizację populacji, odczyt osobników, ocenę funkcji celu, selekcję, modyfikację reprezentacji kubitowej oraz sprawdzenie warunku zakończenia.

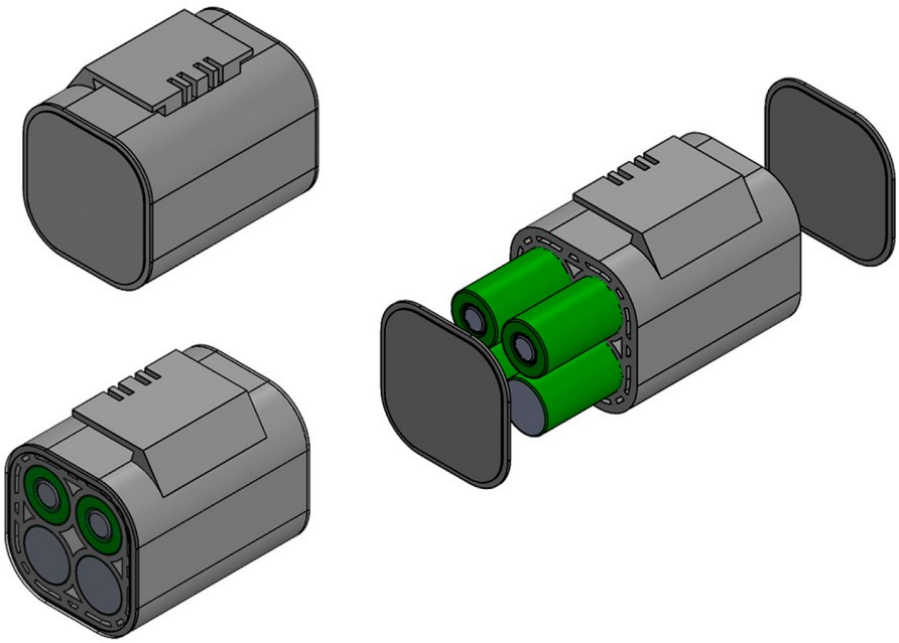
W zastosowaniach optymalizacyjnych rozpatrywanych w pracy QEA pełni rolę algorytmu nadrzędnego, który generuje kolejne osobniki, przekazuje je do oceny funkcji celu, a następnie na podstawie uzyskanych wartości aktualizuje reprezentację kubitową populacji. Populacja startowa może być zdefiniowana losowo, natomiast w opisywanym wariancie inicjalizację realizuje się przez zastosowanie bramki Hadamarda do każdego kubitu, co daje na starcie równe prawdopodobieństwa odczytu 0 i 1. Po każdym odczycie i mapowaniu do zmiennych rzeczywistych wykonywana jest ocena funkcji celu dla osobników, wyniki wracają do części optymalizacyjnej, a następnie wykonywany jest krok selekcji i modyfikacji kubitów. Iteracje są kontynuowane do spełnienia warunku stopu, który może zostać zdefiniowany przez osiągnięcie zakładanej wartości funkcji celu, brak poprawy przez określoną liczbę iteracji lub limit maksymalnej liczby iteracji.

W literaturze algorytm QEA [20], [46], [47] jest opisywany głównie w kontekście problemów kombinatorycznych i benchmarkowych, takich jak problem plecakowy oraz zadania porządkowania i permutacyjne, np. TSP. Jego wykorzystanie w typowych zadaniach inżynierskich sprzężonych z kosztowną ewaluacją funkcji celu, np. symulacjami MES, jest relatywnie mniej spotykane.

3. Metoda optymalizacji obudów baterii

W niniejszym rozdziale jest przedstawiona metoda optymalizacji obudowy baterii, oparta na technikach omówionych w rozdziale 2. Optymalizacje prowadzone są z użyciem jawnej metody całkowania MES (rozdział 2.1) wspomagane wybranymi metodami metamodelowania (rozdział 2.2), a proces optymalizacji opisano w rozdziale 2.3.

Przedstawiona metoda optymalizacji jest uniwersalna i może być zastosowana do wielu konstrukcji o podobnym przeznaczeniu, jednak w celu większej przejrzystości pracy przedstawiono przykład modelu baterii z obudową integralną (rozdział 1.3), która jest poddana analizie wielokrotnego zrzutu (rysunek 3.1).



Rysunek 3.1 Przykładowa koncepcja obudowy wraz z ogniwami.

3.1. Sformułowanie problemu optymalizacji obudowy baterii

Celem optymalizacji, w kontekście testu wielokrotnego zrzutu, jest opracowanie obudowy, która po zadanej sekwencji upadków zachowa możliwie korzystny stan mechaniczny, rozumiany jako brak aktywacji przyjętego modelu zniszczenia oraz możliwie niski poziom trwałych odkształceń plastycznych. W oparciu o metodę elementów skończonych, model plastyczności opisany w rozdziale 2.1.2 oraz model inicjacji i ewolucji zniszczenia opisany w rozdziale 2.1.3, podstawowym wskaźnikiem uszkodzenia w analizie numerycznej jest objętość elementów usuniętych po spełnieniu przyjętego kryterium zniszczenia. Usunięcie

elementów nie jest tutaj interpretowane jako bezpośrednia obserwacja rzeczywistego pęknięcia, lecz jako numeryczna reprezentacja lokalnej utraty nośności materiału zgodnie z zastosowanym modelem obliczeniowym. W przypadkach, w których nie występuje usuwanie elementów, warianty konstrukcyjne są dodatkowo porównywane na podstawie maksymalnego zredukowanego odkształcenia plastycznego. Wielkość ta nie stanowi miary uszkodzenia w sensie mechaniki zniszczenia, lecz pełni funkcję pomocniczego kryterium porządkującego warianty, dla których nie doszło do aktywacji modelu zniszczenia.

Biorąc pod uwagę powyższe stwierdzenia sformułowano następującą funkcję celu:

$$\min_{P} J_0(\mathbf{P}) \tag{3.1}$$

gdzie **P** jest wektorem zmiennych projektowych opisujących analizowany wariant geometrii obudowy, natomiast $J_0(\mathbf{P})$ oznacza funkcję celu przypisaną do danego wariantu konstrukcji. Operator $\min_{P}$ oznacza poszukiwanie takiego zestawu wartości parametrów projektowych **P**, dla którego funkcja celu przyjmuje najmniejszą wartość w rozpatrywanym obszarze zmienności zmiennych projektowych.

Wektor P jest wektorem zawierającym parametry geometryczne obudowy baterii.

$$\mathbf{P} = [P_1, P_2, \dots, P_n] \tag{3.2}$$

gdzie P_i oznacza i-ty parametr projektowy opisujący geometrię obudowy, natomiast n jest liczbą parametrów projektowych uwzględnionych w zadaniu optymalizacji. W analizowanym przypadku parametry P_i odpowiadają wymiarom geometrycznym obudowy, których wartości są zmieniane w kolejnych wariantach konstrukcyjnych.

Zadanie optymalizacyjne uzupełnione jest w rozpatrywanym problemie o ograniczenia. Parametry P_i podlegają ograniczeniom kostkowym:

$$\mathrm{C} = \{\mathbf{P}\colon \forall \mathrm{i} \in \{1, \dots, \mathrm{n}\} P_{imin} \leq P_i \leq P_{imax}\} \tag{3.3}$$

Rozważane są dwa warianty funkcji celu. Pierwszy wariant funkcji celu (FF1) można zapisać:

$$J_0(\mathbf{P}) = \mathrm{a} \cdot \mathrm{V}_{del}(\mathbf{P}) + \mathrm{b} \cdot \varepsilon_{p_max}(\mathbf{P}) \tag{3.4}$$

gdzie $V_{del}(\mathbf{P})$oznacza objętość elementów skończonych usuniętych z modelu w wyniku spełnienia kryterium zniszczenia [2], natomiast $\varepsilon_{p,\max}(\mathbf{P})$ oznacza maksymalną wartość zredukowanego odkształcenia plastycznego wyznaczoną w analizowanym wariancie

konstrukcji. Współczynniki a oraz b są współczynnikami wagowymi, które określają względny udział obu składników w funkcji celu

Zaproponowano również drugi wariant funkcji celu w postaci:

$$J_0(\mathbf{P}) = \begin{cases} \varepsilon_{p_max}(\mathbf{P}), & jeżeli\ \mathrm{V}_{del}(\mathbf{P}) = 0 \\ a \cdot \log(\mathrm{V}_{del}(\mathbf{P}) + 1) + b\,, & jeżeli\ \mathrm{V}_{del}(\mathbf{P}) > 0 \end{cases} \tag{3.5}$$

gdzie a i b są współczynnikami dobranymi w celu równomiernego rozkładu wartości funkcji celu. Funkcja ta zawiera logarytmiczne przekształcenie objętości usuniętych elementów $V_{del}(\mathbf{P})$, w celu ograniczenia nadmiernego rozrzutu wartości funkcji celu wynikającego z dużej zmienności tej wielkości. Dzięki temu przypadki z większą objętością usuniętych elementów nadal są oceniane jako mniej korzystne, ale nie dominują całkowicie nad pozostałymi wariantami konstrukcyjnymi. Pozwala to uzyskać bardziej równomierny rozkład wartości $J_0(\mathbf{P})$ w analizowanym zbiorze rozwiązań.

3.2. Metoda obliczania wartości funkcji celu

Metody optymalizacyjne opisane w rozdziale opierają się na obliczeniach funkcji celu na bazie rozwiązań zadań MES. W celu usprawnienia analiz proces ten został zautomatyzowany za pomocą dedykowanego programu sterującego pracą programu MES. Algorytm można zaimplementować z użyciem dowolnego oprogramowania MES.

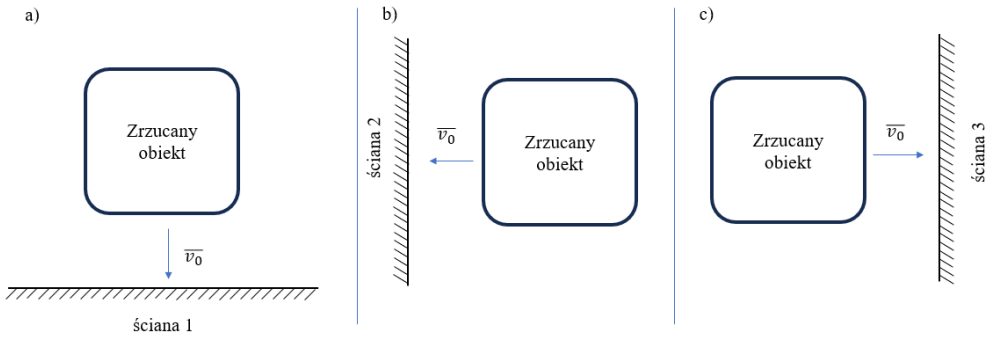
Analizowany przypadek ma charakter symulacji wielokrotnego zrzutu, w której kolejne uderzenia obudowy o sztywne podłoże są traktowane jako następujące po sobie etapy obciążenia dynamicznego. Każdy zrzut stanowi odrębne zadanie dynamiczne MES, obejmujące fazę swobodnego ruchu, kontakt z podłożem, gwałtowne wyhamowanie konstrukcji oraz ewentualny rozwój lokalnych odkształceń plastycznych i zniszczenia. Z tego względu procedura obliczeniowa nie może być ograniczona do pojedynczej analizy udarowej, lecz musi uwzględniać sekwencyjny charakter testu oraz wpływ poprzednich zrzutów na odpowiedź konstrukcji w kolejnych cyklach. W symulacji uwzględniono efekt kumulacji zniszczenia po każdym kolejnym cyklu zrzutu. Stan obudowy (deformacje, rozkład naprężeń i odkształceń) po zakończeniu jednego zrzutu, jest automatycznie wczytywany na początku symulacji kolejnego zrzutu.

Algorytm sterujący został przygotowany w taki sposób, aby można było ustawić dowolną ilość powtórzeń zrzutu z danej wysokości oraz umożliwiono wpisanie prędkości początkowych $\overline{v_0}$. Prędkość $\overline{v_0}$ określono jako:

$$\overline{v_0} = \sqrt{2gh} \tag{3.6}$$

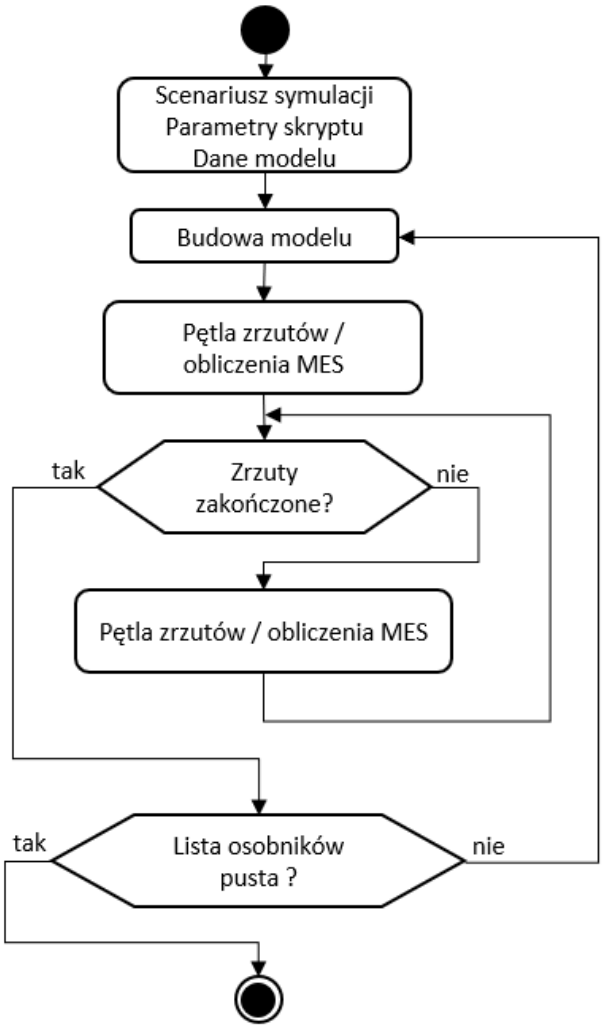
gdzie g to wartość przyspieszenia ziemskiego, a h to wysokość z jakiej jest wykonywany zrzut.

Poza wysokościami oraz ilością powtórzeń (dla konkretnej wysokości oraz kierunku) skrypt posiada również możliwość zdefiniowania strony obudowy jaka ma być zrzucana. Jest to deterministyczny wybór spośród trzech predefiniowanych kierunków zrzutów, co odpowiada rzeczywistej definicji testu opisanego w rozdziale 1.2. W modelu numerycznym jest to zrealizowane za pomocą utworzenia trzech analitycznych sztywnych ścian, które w zależności od kierunku zrzutu są aktywowane automatycznie za pomocą skryptu, a kierunek wektora prędkości początkowej $\overline{v_0}$ jest zmieniany w taki sposób, aby był skierowany w kierunku odpowiedniej ściany. Przedstawia to rysunek 3.2.



Rysunek 3.2. Schemat przedstawiający sposób realizacji zrzutów na różne kierunki.

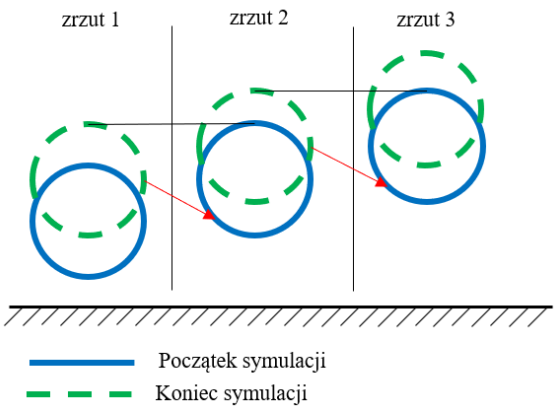
W skrypcie jest zaimplementowany sparametryzowany model, który został odpowiednio przygotowany, tak aby program mógł przejąć kontrolę nad zmianą parametrów geometrycznych czy materiałowych. Do zdefiniowania zadania i scenariusza symulacyjnego wymagane są podstawowe parametry, takie jak wartość prędkości początkowych odpowiadających zadanym wysokościom zrzutów, kierunek lub kierunki zrzutów oraz liczbę powtórzeń dla każdej konfiguracji. Po zdefiniowaniu tych danych, dalsze operacje - w tym generowanie geometrii na podstawie wartości zmiennych projektowych, symulacja serii zrzutów, przenoszenie stanu z końca do początku analizy oraz analiza wyników - wykonywane są automatycznie zgodnie z określonym schematem. Automatyzacja ta pozwala na efektywne przeprowadzenie procesu optymalizacji, minimalizując czasochłonność pracy oraz ryzyko błędów ludzkich w trakcie obliczeń (rysunek 3.3).



Rysunek 3.3. Schemat blokowy przedstawiający algorytm skryptu symulacji MES.

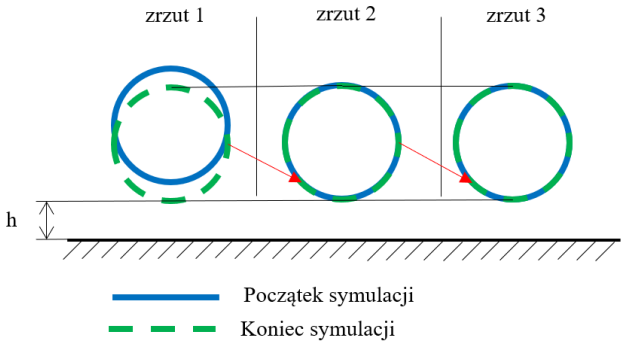
Program zapisuje wyniki obliczeń do bazy danych. Plik wyjściowy zawiera informacje o wektorze parametrów obliczonych osobników wraz z wynikami, czyli największym zredukowanym odkształceniem plastycznym ε_{p_max} oraz objętością elementów V_{del}, które zostały usunięte na skutek przekroczenia kryterium zniszczenia (rozdział 2.1.3). Objętość ta jest automatycznie obliczana na bazie indeksów zniszczenia opisanymi w rozdziale (rozdział 2.1.3).

Jednym z kluczowych problemów w analizie wielokrotnych zrzutów jest przenoszenie stanu końcowego z jednej symulacji do początkowego stanu kolejnej. Założenie, że każda analiza trwa określony, stały czas, prowadzi do sytuacji, w której przy zrzutach z większej wysokości symulacja kończy się na coraz wyższej pozycji odbijającego się obiektu (rysunek 3.4). Problem ten skutkuje przesunięciem początku kolejnej analizy, co wpływa na spójność warunków początkowych i może zakłócać poprawność wyników.



Rysunek 3.4 Wizualizacja problemu z kumulacją odległości podczas wielokrotnych zrzutów.

Problem ten rozwiązano poprzez zaimplementowanie mechanizmu opartego na punktach charakterystycznych, które monitorują przemieszczenie obiektu. Punkty te śledzą kierunek ruchu i w momencie jego zmiany (np. podczas odbicia) aktywują zatrzymanie obliczeń MES dla zadanej drogi *h* po odbiciu. Dzięki temu analiza kończy się dokładnie w momencie odbijania po przebyciu drogi *h,* co pozwala na zachowanie niezmiennych warunków początkowych dla kolejnych cykli zrzutów (rysunek 3.5). Rozwiązanie to skutecznie zapobiega efektowi stopniowego podnoszenia się obiektu, zapewniając spójność i dokładność symulacji w całym procesie optymalizacyjnym.



Rysunek 3.5 Rozwiązanie problemu kumulacji odległości za pomocą filtrów.

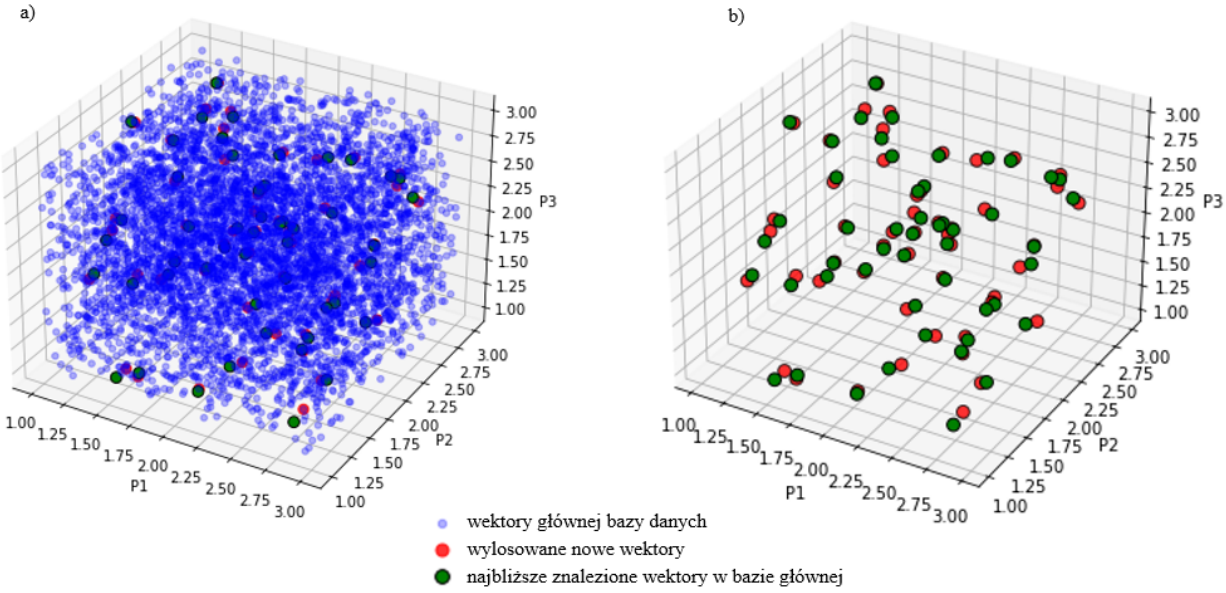
3.3. Bazy danych wyników symulacji numerycznych

Wyniki analiz MES zapisywane są w bazie danych. Każdy rekord (zapisany wynik) zawiera wartości wektora zmiennych projektowych **P** i dane obejmujące wyniki obliczeń, takie jak całkowita objętość usuniętych elementów V_{del} oraz maksymalne zredukowane odkształcenie plastyczne znalezione w analizie ε_{p_max}. Pojedynczy rekord w bazie danych przyjmuje następującą postać: $[P_1, P_2, P_3, \ldots, P_n, V_{del}, \varepsilon_{p_max}]$.

Bazy danych zawierają rekordy odpowiadające zestawom parametrów, dla których wykonano obliczenia MES. Przed uruchomieniem nowej analizy skrypt sprawdza, czy dany wektor zmiennych projektowych został już wcześniej zapisany w bazie danych. Jeżeli taki rekord istnieje, wartość funkcji celu jest pobierana z bazy bez ponownego uruchamiania analizy MES i przekazywana do bazy danych. Jeżeli rekord nie istnieje, wykonywana jest nowa analiza MES, a otrzymany wynik jest dopisywany do bazy danych. Mechanizm ten ogranicza liczbę powtarzanych obliczeń dla tych samych konfiguracji geometrycznych.

W pracy rozpatrywane będą dwie bazy danych. Pierwsza zawiera dane uzyskane z użyciem metod planowania eksperymentów próbkowania metodą Hiperześcianu Łacińskiego (LHS). Pierwszą z baz uzyskuje się metodą Hiperześcianu Łacińskiego (2.2.1). Metoda ta łączy dwie strategie ukierunkowane na poprawę równomierności rozmieszczenia próbek według kryterium wyboru środkowego przedziału (ang. „centered") oraz maksymalizacją minimalnych odległości (ang. „maxmin") (opisano w rozdziale 2.2.1). Druga baza danych jest utworzoną na podstawie danych z zadań optymalizacyjnych rozwiązanych w ramach realizacji pracy. Druga baza danych tworzona jest na podstawie wektorów zmiennych projektowych pochodzących z wcześniejszych przebiegów procesu optymalizacji prowadzonego z użyciem Ewolucyjnego Algorytmu Inspirowanego Zjawiskami Kwantowymi (QEA) (rozdział 2.3). Dane gromadzone w trakcie licznych iteracji procesu optymalizacyjnego zapewniają odpowiednią liczebność próbek do dalszych badań (rozdział 4.4.1).

W celach badawczych wykorzystuje się dwa nadrzędne zbiory danych, z których wyznacza się podzbiory o różnej liczebności próbek. Podzbiory te służą do oceny wpływu liczby dostępnych rekordów na dokładność predykcji metamodelu oraz do określenia minimalnego rozmiaru zbioru danych zapewniającego akceptowalną jakość aproksymacji. Wektory zmiennych projektowych w obu zbiorach są generowane niezależnie, dlatego punkty nie muszą pokrywać się bezpośrednio w przestrzeni projektowej. Aby ograniczyć koszt ponownego wykonywania obliczeń numerycznych, dla punktów reprezentujących podzbiory o mniejszej liczności wyszukuje się najbliższe punkty dostępne w większym zbiorze danych. Wybrane w ten sposób rekordy tworzą zredukowane podzbiory danych o zadanej liczebności, które następnie wykorzystuje się do budowy i porównania metamodeli trenowanych na zbiorach o różnym rozmiarze. Sposób wyznaczania takich podzbiorów przedstawia rysunek 3.6 .



Rysunek 3.6 Wizualizacja metody wyznaczania mniejszych baz danych na podstawie podzbiorów przybliżonych rozwiązań. a) wektory głównej bazy danych wraz z nowymi oraz najbliższymi wektorami z głównej bazy danych do nowych wektorów b) nowo wylosowane wektory oraz najbliższe im odpowiadające wektory z głównej bazy.

Odległość $d_{k\ell}$ pomiędzy punktami jest określana zgodnie z równaniem:

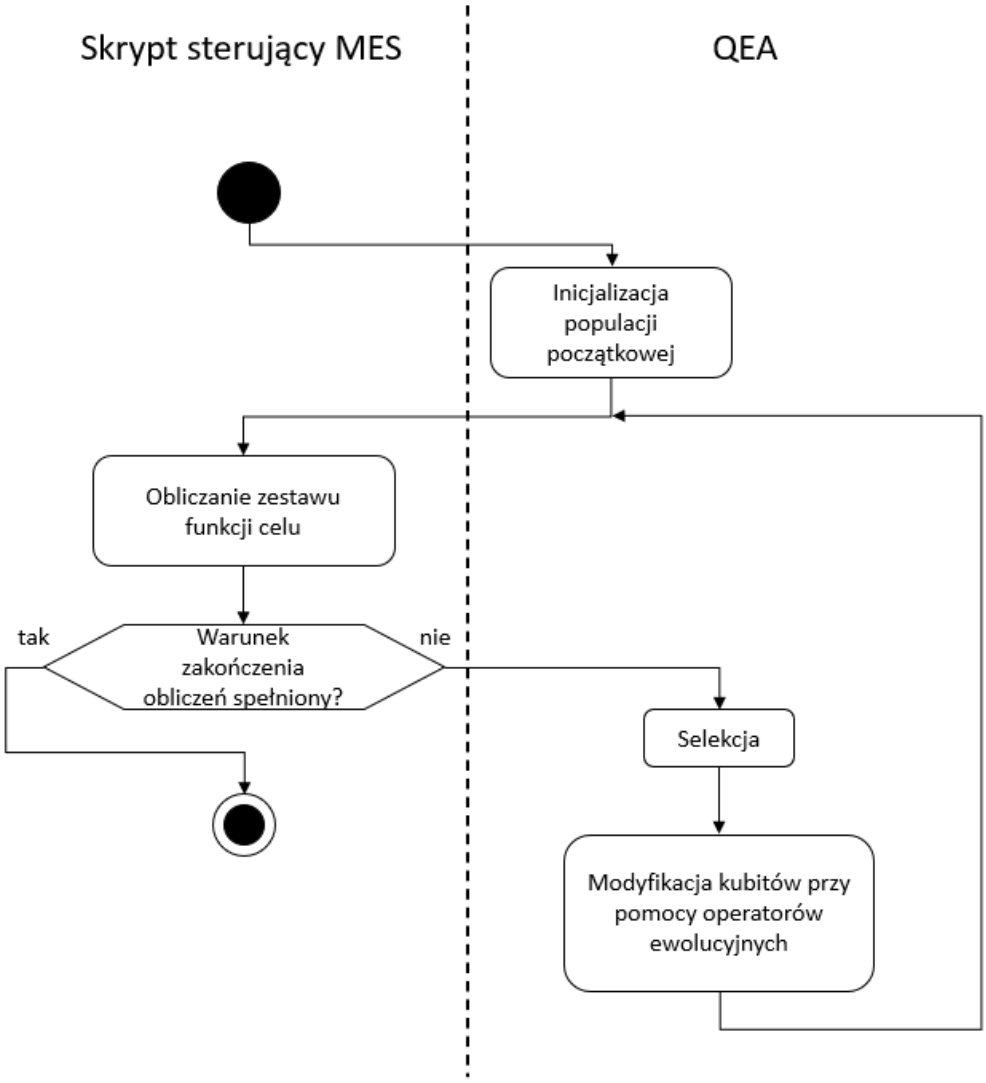
$$d_{k\ell} = \| \mathbf{x}_k - \mathbf{x}_\ell \|_2 = \sqrt{\sum_{j=1}^{n} \left(x_{k,j} - x_{\ell,j}\right)^2} \tag{3.7}$$

gdzie $\mathbf{x}_k$ oraz $\mathbf{x}_\ell$ oznaczają wartości zmiennych odpowiadających dwóm różnym punktom, natomiast j jest liczbą wymiarów analizowanej przestrzeni, a n to ilość zmiennych. Zastosowanie tego podejścia pozwoliło na istotne skrócenie czasu obliczeń przy jednoczesnym zachowaniu równomiernego rozkładu próbek w przestrzeni projektowej.

3.4. Połączenie algorytmu optymalizacji QEA oraz MES

Metoda opisana w tym rozdziale jest połączeniem Ewolucyjnego Algorytmu Inspirowanego Zjawiskami Kwantowymi z MES. Metoda łączy skrypt sterujący opisany w rozdziale 3.2 z algorytmem optymalizacyjnym. Skrypt sterujący uruchamia algorytm optymalizacyjny, który generuje pulę wektorów zmiennych projektowych. Wielkość puli jest zależna od liczby osobników, dla których konieczne jest określenie wartości funkcji celu. Na podstawie puli

wektorów zmiennych projektowych obliczane są wartości funkcji celu z użyciem wyników z analiz MES (rysunek 3.7).



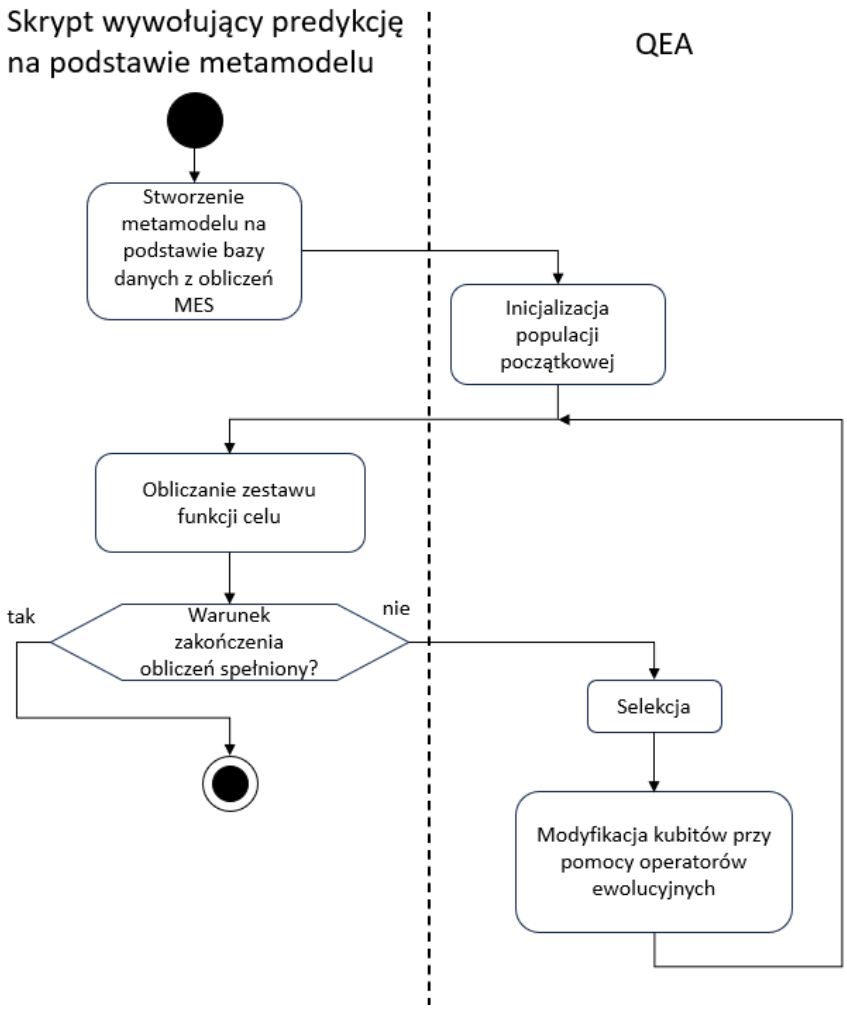
Rysunek 3.7 Sposób współdzielenia algorytmu optymalizacji oraz skryptu obliczającego wartości funkcji celu.

W trakcie obliczeń rekordy są zapisywane do bazy danych, aby umożliwić ich wykorzystanie w dalszych krokach analizy. Każdy rekord zawiera wektor zmiennych projektowych oraz odpowiadającą mu wartość funkcji celu obliczoną metodą MES. Dodatkowo kontrolowane jest, czy dany zestaw parametrów był już wcześniej obliczony. Jeżeli w kolejnych iteracjach algorytm QEA wygeneruje wektor zmiennych projektowych występujący już w zbiorze wcześniej obliczonych przypadków, skrypt sterujący odczytuje zapisaną wartość funkcji celu bez ponownego uruchamiania obliczeń MES. Pozwala to ograniczyć liczbę

kosztownych symulacji numerycznych oraz zachować komplet danych wymaganych do dalszego przebiegu optymalizacji.

3.5. Połączenie algorytmu optymalizacji z uproszczonymi bazami danych

Strojenie metamodeli oraz dobór hiperparametrów algorytmu optymalizacji jest procesem kosztownym obliczeniowo. W celu przeprowadzenia testów numerycznych zdecydowano o zastosowaniu algorytmu, który nie wykorzystuje MES w procesie obliczania funkcji celu. Zasadę działania ilustruje rysunek 3.8.

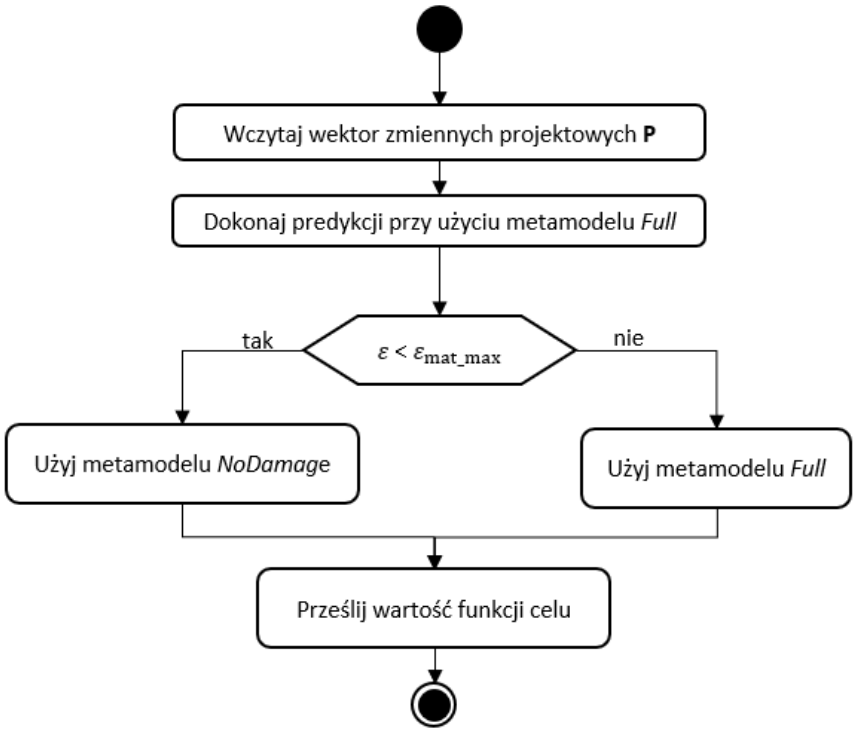


Rysunek 3.8 Zasada działania metody połączenia algorytmu optymalizacji z uproszczonymi bazami danych.

Algorytm taki nie może zostać użyty w optymalizacji rzeczywistej konstrukcji w tej postaci, ponieważ nie wykonuje nowych obliczeń MES w trakcie kolejnych iteracji. W analizowanym

wariancie obliczenia MES są zastępowane przez gotowy metamodel, zbudowany wcześniej na podstawie baz danych pochodzących z obliczeń MES. Metamodel pełni więc funkcję szybkiego aproksymatora wartości funkcji celu dla osobników generowanych przez algorytm optymalizacyjny, co pozwala skrócić czas obliczeń i sprawdzić przydatność takiego podejścia przed opracowaniem procedury hybrydowej z bieżącą aktualizacją bazy danych.

W celu uzyskania lepszej predykcji metamodelu baza danych z wynikami jest podzielona na dwie osobne bazy. Jedna baza obejmuje wszystkie wyniki i nazywała się *Full*, oraz druga baza danych, która zawierała tylko osobniki, w których nie doszło do zniszczenia o nazwie *NoDamage*. Konsekwencją tego faktu jest to, że w jednym zadaniu zostaną utworzone dwie bazy danych i do celów predykcyjnych musi zostać zaimplementowana logika, która pozwoli na właściwy wybór metamodelu, przedstawia ją rysunek 3.9.



Rysunek 3.9 Schemat decyzyjny wyboru metamodelu *Full* lub *NoDamage* na podstawie wstępnej predykcji oraz kryterium progowego związanego z $\varepsilon_{p_max_mat}$.

Dla zadanego wektora zmiennych projektowych wejściowych **P** w pierwszej kolejności wykonywana jest predykcja z użyciem metamodelu *Full*, który obejmuje cały zakres danych, tj. przypadki ze zniszczeniem oraz bez zniszczeń. Otrzymany wynik pełni rolę kryterium klasyfikującego i jest porównywany z wartością progową odpowiadającą największemu dopuszczalnemu odkształceniu plastycznemu w danym materiale $\varepsilon_{p_max_mat}$. Jeżeli

przewidywany wynik jest mniejszy od $\varepsilon_{p_max_mat}$ przypadek jest kwalifikowany do obszaru *NoDamage* i końcowa wartość funkcji celu wyznaczana jest przy użyciu metamodelu *NoDamage*, trenowanego wyłącznie na próbkach bez inicjacji uszkodzeń. W przeciwnym razie, gdy warunek nie jest spełniony, do dalszych obliczeń przyjmowana jest wykonana wcześniej predykcja metamodelu *Full*. Rozdzielenie modeli ma na celu poprawę dokładności predykcji w podprzestrzeni, w której nie występuje zniszczenie, Wynik predykcji po wyborze gałęzi jest następnie przekazywany do kolejnych etapów procedury optymalizacji. Ten sam schemat działania jest zaimplementowany w rozdziale 3.6

3.6. Metoda hybrydowa optymalizacji wykorzystująca algorytm optymalizacji oraz metamodelowanie

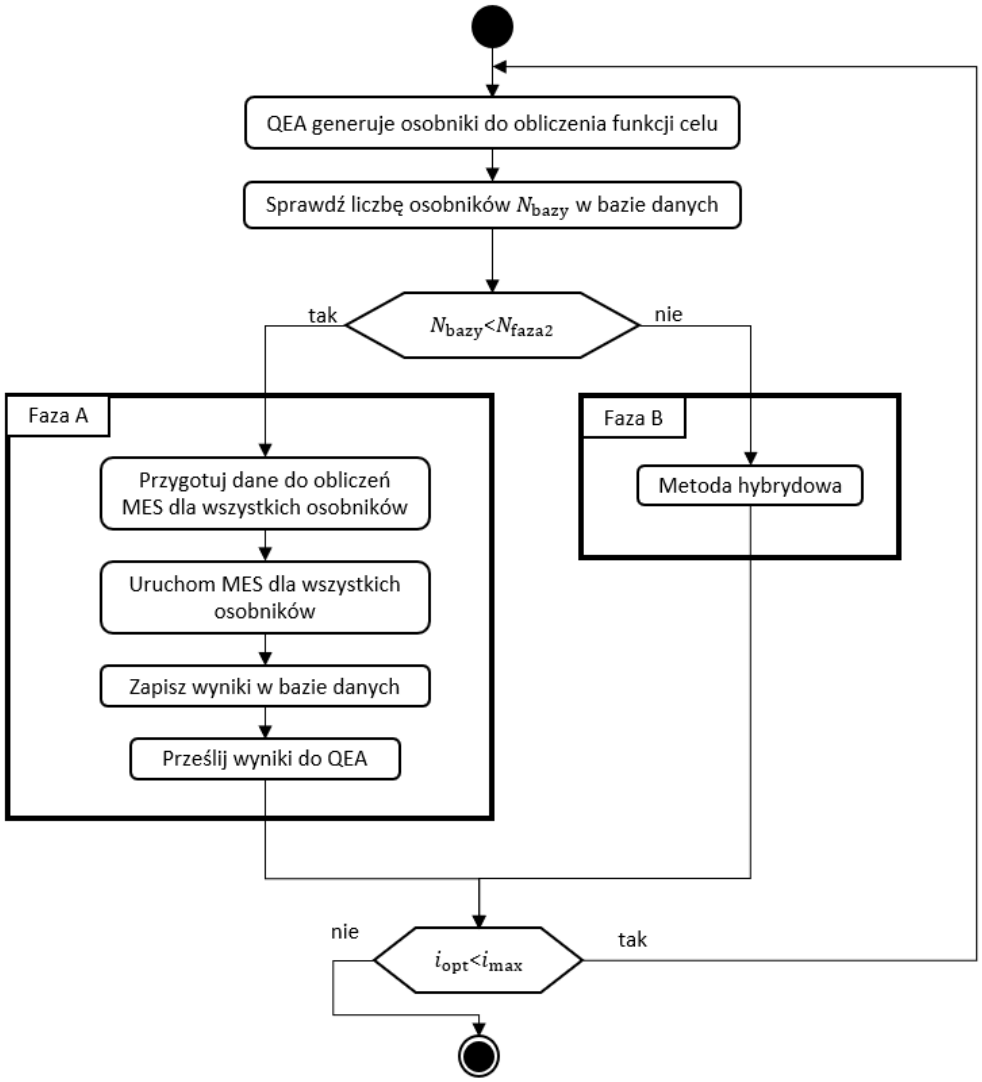
Przeprowadzanie procesu optymalizacji wyłącznie na podstawie wcześniej wygenerowanych baz danych (rozdział 3.5) oraz statycznych (bez aktualizacji) predykcji metamodeli nie stanowi realistycznego rozwiązania inżynierskiego. Rozwiązywanie zadania optymalizacyjnego przy użyciu wyłącznie algorytmu optymalizacyjnego oraz oceny funkcji celu za pomocą MES (rozdział 3.4) prowadzi do bardzo długich czasów obliczeń. Z tego względu opracowano metodę polegającą na integracji procesu optymalizacji z budową bazy danych rozwiązań w trakcie optymalizacji oraz na wykorzystaniu metamodelu do przyspieszenia ewaluacji funkcji celu.

Zaproponowana metoda zdefiniowana jest jako proces dwufazowy. W fazie A wartość funkcji celu wyznaczana jest wyłącznie na podstawie obliczeń MES. W fazie B ocena populacji jest wspierana predykcjami metamodelu, wykorzystującego dane zgromadzone w fazie A.

Faza A jest uruchamiana, jeżeli liczebność bazy N_{bazy}, rozumiana jako liczba unikalnych osobników z przypisaną wartością funkcji celu wyznaczoną metodą MES, jest mniejsza od zadanego progu minimalnego N_{faza2}. Odpowiada to procesowi, który został opisany w rozdziale 3.4. W fazie A algorytm QEA generuje pulę osobników, po czym skrypt przygotowuje komplet danych wejściowych do analiz MES dla całej puli. Następnie wykonywane są obliczenia MES dla całej puli, a uzyskane wartości funkcji celu są przekazywane do QEA. Faza A jest powtarzana iteracyjnie aż do momentu osiągnięcia minimalnego progu liczby rekordów w bazie danych N_{bazy}.

W fazie B ewaluacja funkcji celu realizowana jest z wykorzystaniem metody hybrydowej. W tej fazie wartości funkcji celu dla populacji są w pierwszej kolejności wyznaczane z użyciem metamodelu, natomiast obliczenia MES są wykonywane selektywnie zgodnie z przyjętym wariantem algorytmu hybrydowego. W pracy rozpatrzono dwa warianty fazy B: HYBRYDA1

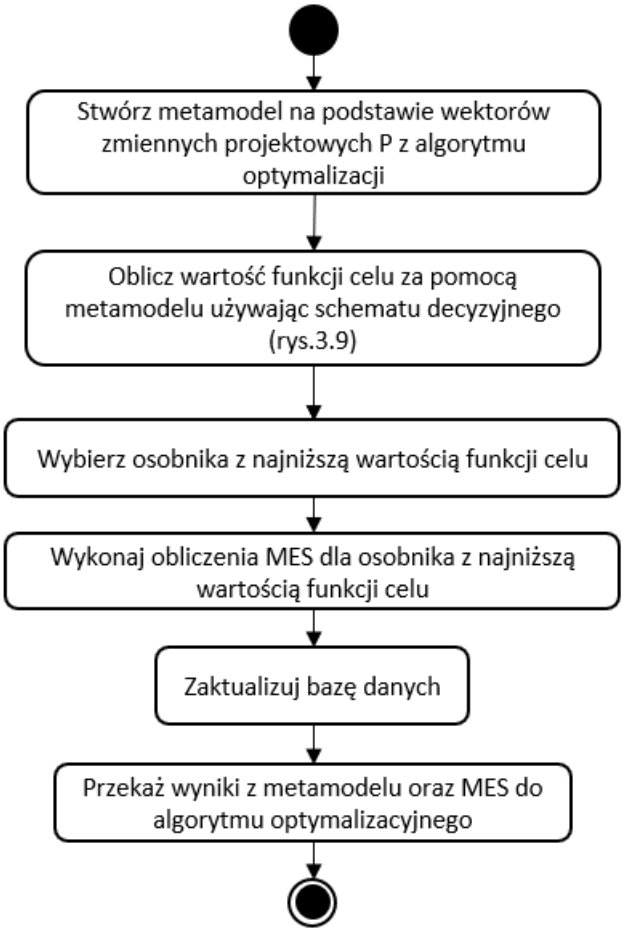
oraz HYBRYDA2. Cały proces przebiega w pętli sterowanej liczbą iteracji i_{opt}, kontynuowanej do osiągnięcia z góry zdefiniowanego limitu iteracji i_{max}. Schemat przełączania pomiędzy fazą A i fazą B oraz przepływ danych w algorytmie przedstawia rysunek 3.10.



Rysunek 3.10 Schemat działania metody hybrydowej.

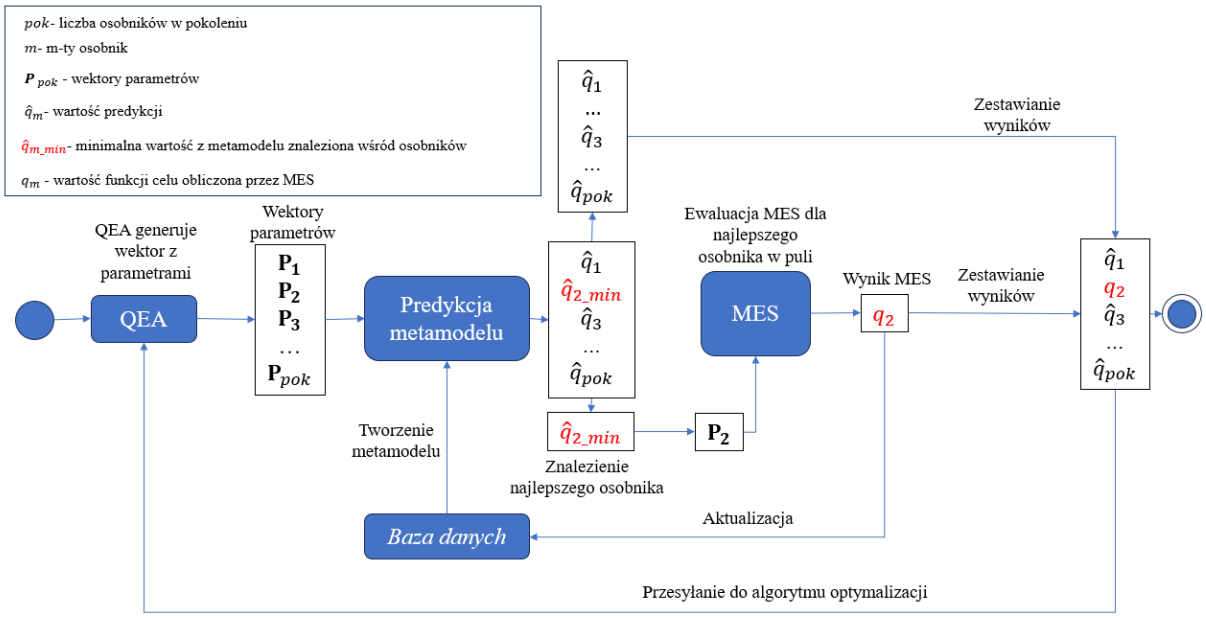
Podejście oznaczone jako HYBRYDA1 realizuje pętlę fazy B na zasadzie „metamodel dla całej populacji, MES dla najlepszego osobnika”. Na początku każdej iteracji fazy B tworzona jest określona liczba metamodeli na aktualnej bazie danych, a następnie wybierany jest najlepszy wariant metamodelu danej klasy. Kryterium wyboru zależy od zastosowanej metody metamodelowania. W przypadku sztucznych sieci neuronowych oraz metamodelu Kriginga stosuje się wartość współczynnika determinacji R^2, natomiast w przypadku regresji

wielomianowej stosuje się miarę rankingową R^2_{PNR}, zdefiniowaną wzorem 2.70. Metamodel o najlepszym wyniku służy do wyznaczenia przybliżonej wartości funkcji celu (predykcji) dla wszystkich osobników wygenerowanych przez algorytm optymalizacyjny w bieżącym pokoleniu. Na tym etapie w zależności od potrzeby jest uruchamiana procedura wyboru metamodelu (*Full* lub *NoDamage*) (rysunek 3.11).



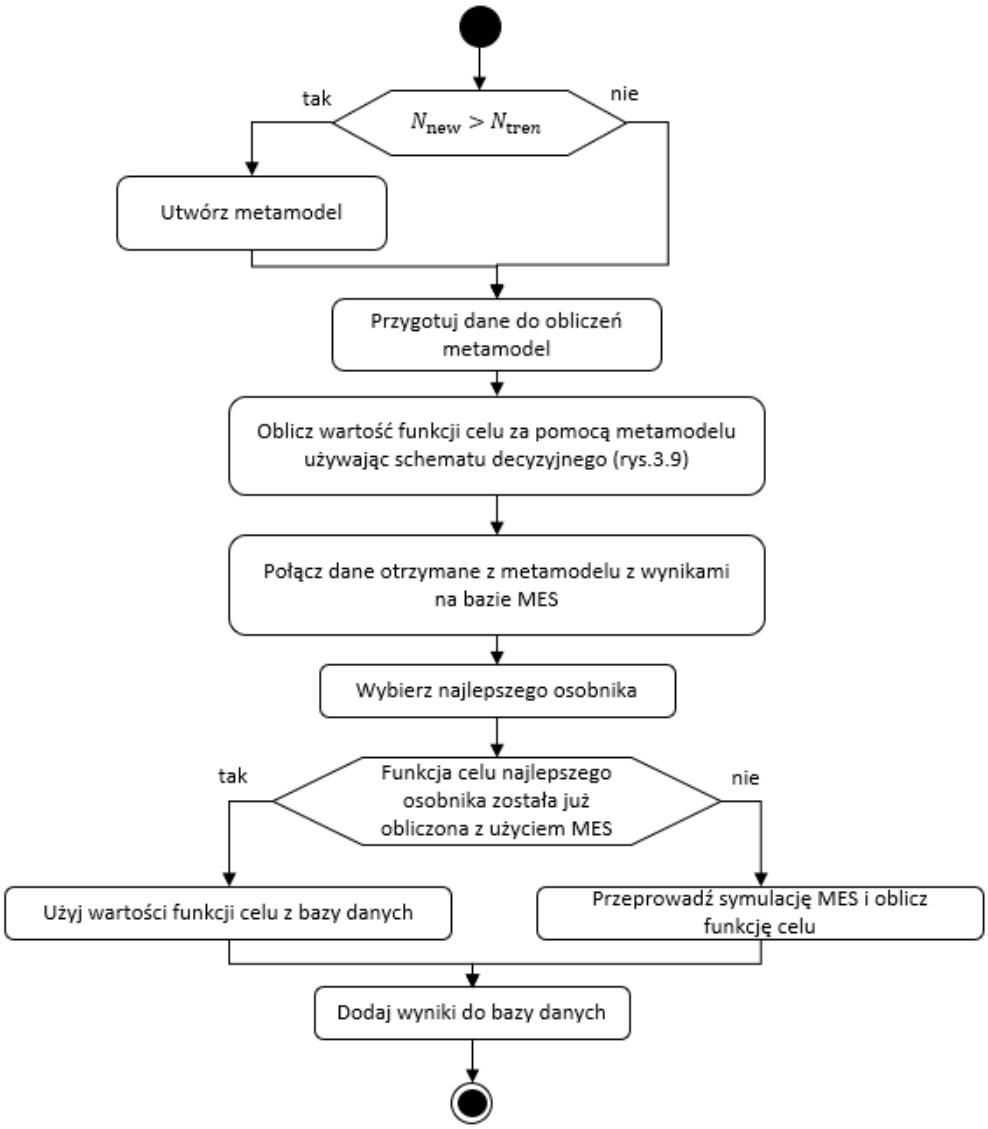
Rysunek 3.11 Schemat ilustrujący zasadę działania metody HYBRYDA1.

Na podstawie predykcji z użyciem metamodelu wybierany jest osobnik o najlepszej, tj. minimalnej, wartości funkcji celu. Dla tak wybranego osobnika wykonywana jest analiza MES, a uzyskany wynik jest dopisywany do bazy. Na zakończenie przygotowywany jest zestaw wartości funkcji celu przekazywany do algorytmu optymalizacyjnego, obejmujący wynik MES dla najlepszego osobnika oraz wartości wyznaczone przez metamodel dla pozostałych osobników (rysunek 3.12).



Rysunek 3.12 Przepływ wyników pomiędzy bazami danych w metodzie HYBRYDA1.

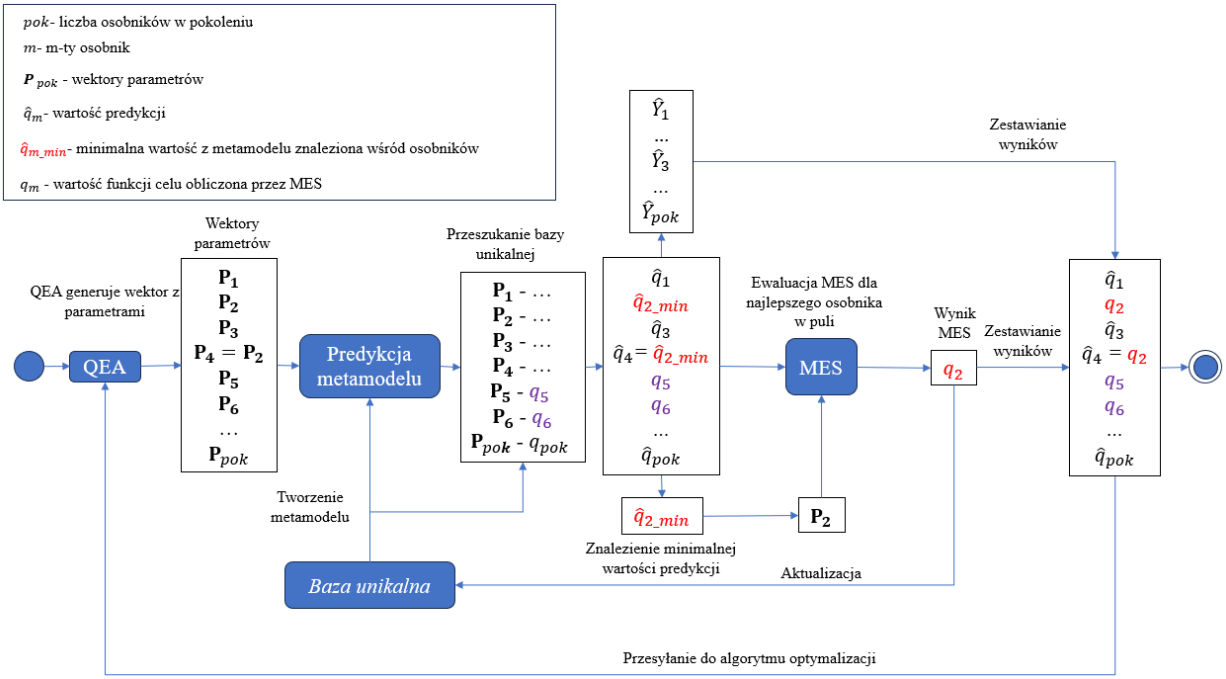
Metoda HYBRYDA2 różni się od podejścia HYBRYDA1 sposobem prowadzenia analiz MES oraz częstością ponownego uczenia metamodeli. Jeżeli jest to pierwsza iteracja fazy B lub liczba nowych punktów dodanych do bazy N_{new} od ostatniego uczenia przekroczy zdefiniowany próg N_{train}, metamodel jest tworzony na nowo lub ponownie; w przeciwnym razie etap uczenia jest pomijany i wykorzystywana jest ostatnia wersja metamodelu. Następnie dla populacji osobników wygenerowanej przez algorytm optymalizacyjny określana jest wartość funkcji celu na podstawie metamodelowania. W przypadku istnienia rozwiązania MES, ma ono wyższy priorytet niż rozwiązanie z metamodelu (rysunek 3.13.).



Rysunek 3.13 Schemat działania metody HYBRYDA 2.

Na podstawie tak przygotowanej bazy danych wybierany jest najlepszy kandydat w danym pokoleniu. Jeżeli najlepszy osobnik posiada już wynik MES zapisany w bazie, wartość funkcji celu jest odczytywana bez uruchamiania kolejnej symulacji. W przypadku braku wyniku w bazie, przygotowywany jest plik wejściowy do obliczeń MES, wykonywana jest symulacja, a uzyskana wartość funkcji celu jest przekazywana do QEA. W kolejnym kroku przekazywane są wyniki do algorytmu optymalizacji zawierające wartości funkcji celu dla całej populacji - pochodzące z symulacji MES tam, gdzie wynik jest dostępny, oraz z metamodelu w pozostałych przypadkach. W miarę nasycania bazy danych rośnie udział osobników, dla których dostępna jest wartość funkcji celu na podstawie MES, co ogranicza liczbę nowych uruchomień symulacji, natomiast okresowe ponowne tworzenie metamodelu przyspiesza

proces predykcji poprzez ograniczenie częstych zmian modelu. Schemat przepływu podejścia HYBRYDA2 przedstawia rysunek 3.14



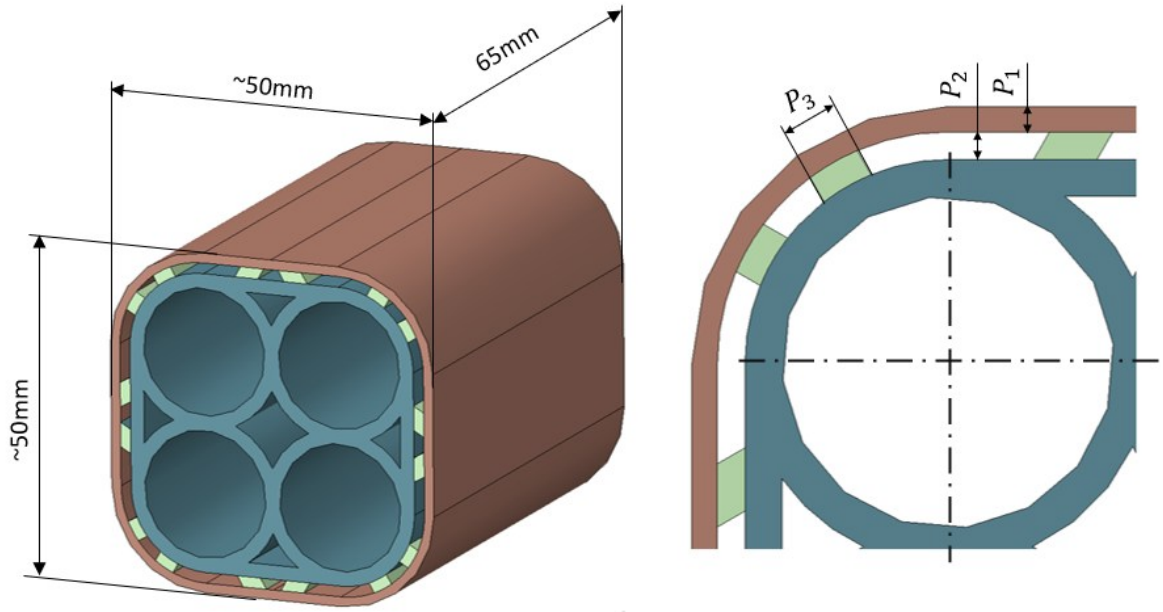
Rysunek 3.14 Przepływ wyników pomiędzy bazami danych w metodzie HYBRYDA2.

4. Przykład numeryczny optymalizacji obudowy baterii

W niniejszym rozdziale opracowano koncepcyjny model obudowy baterii w wariancie konstrukcji zintegrowanej (rozdział 1.3). Geometria obudowy została dobrana w taki sposób, aby z jednej strony zachować kluczowe cechy zintegrowanej obudowy baterii, z drugiej zaś - maksymalnie uprościć kształt w celu jego parametryzacji. Zastosowano obudowę przystosowaną do pomieszczenia czterech ogniw litowo-jonowych. Należy podkreślić, że przedstawiona w procedura może zostać bezpośrednio wykorzystana do optymalizacji dowolnego kształtu obudowy, o ile tylko możliwe jest jego opisanie za pomocą skończonego zbioru parametrów geometrycznych. Wyniki przedstawione w tym rozdziale były wcześniej częściowo publikowane przez autora dysertacji w publikacjach naukowych [76], [77]. Oprogramowaniem użytym w przykładzie numerycznym jest Abaqus 2024 [105], a skrypt sterujący programem Abaqus został napisany w języku Python (wersja 3.12) [70].

4.1. Model numeryczny

Jako bazę do modelu opisywanego w tym rozdziale posłużył model opisany w rozdziale 3 (rysunek 3.1). Model konstrukcyjny obudowy został odpowiednio uproszczony i przekształcony w model numeryczny do obliczeń metodą elementów skończonych. Z modelu usunięto niepotrzebne elementy takie jak klapy boczne, które służą jedynie do uzyskania szczelności, oraz adapter do mocowania w narzędziu końcowym. Koncepcyjna obudowa składa się z trzech zasadniczych sekcji, przedstawia to rysunek 4.1 W jej skład wchodzą: wewnętrzny element pozycjonujący ogniwa baterijne (niebieski), zewnętrzna powłoka obudowy (oznaczona kolorem czerwonym) oraz układ żeber łączących obie te części (kolor zielony). Jak wskazano w rozdziale 1.2, analizowany jest scenariusz wielokrotnych upadków obudowy na podłoże (w wariancie uproszczonym). W związku z tym, w celu zapewnienia odpowiedniej wytrzymałości konstrukcji, w dalszej części pracy dobierane są trzy parametry sterujące kształtem obudowy, oznaczone jako P_1, P_2oraz P_3.



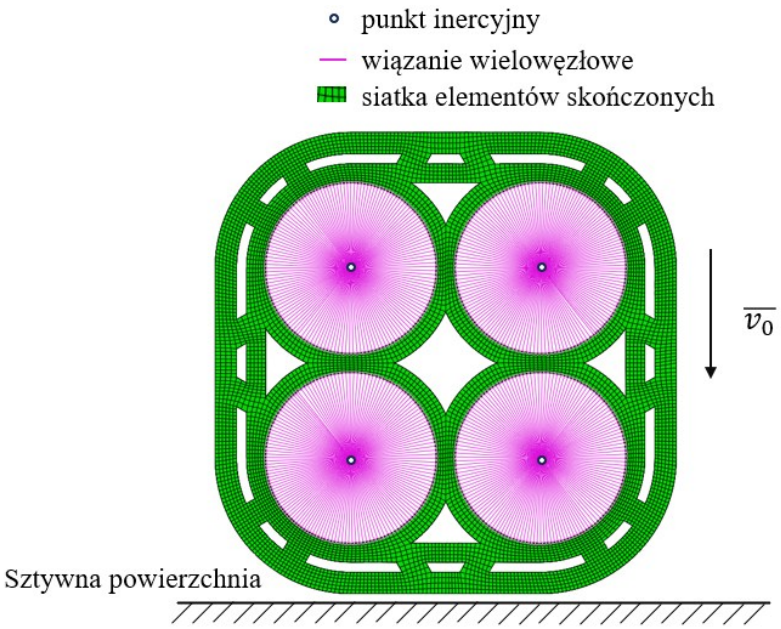
Rysunek 4.1 Koncepcyjny model obudowy.

Przestrzeń projektowa została zdefiniowana jako podzbiór trójwymiarowej przestrzeni parametrów o ograniczeniach kostkowych:

$$\mathbf{P} = \{ P_1, P_2, P_3 \mid P_1 \in (1\,\text{mm}, 3\,\text{mm}),\ P_2 \in (1\,\text{mm}, 3\,\text{mm}),\ P_3 \in (1\,\text{mm}, 3\,\text{mm}) \} \qquad (4.1)$$

Zmienność parametrów wynika przede wszystkim z ograniczeń grubości ścianek w technologii wykonania obudowy.

Objętość obudowy zdyskretyzowano przy użyciu około 64 000 (w zależności od wartości parametrów $P_1 - P_3$) ośmiowęzłowych, liniowych elementów objętościowych typu hex, przy średniej długości boku elementu ustalonej na 0,4 *mm*. Siatkę elementów skończonych dobrano w taki sposób, aby w każdym wariancie występowały minimum 4 elementy na grubości (rysunek 4.2). Analiza zbieżności siatki wykazała dobry kompromis pomiędzy jakością otrzymywanych wyników a czasem obliczeń. Ogniwa bateryjne zastąpiono punktami inercyjnymi o masie 0,07 kg i zadanych momentach bezwładności równych dla osi wzdłuż ogniwa $I_z = 2{,}835\,\text{kg} \cdot \text{mm}^2$ oraz dla osi poprzecznych do osi wzdłużnej ogniwa $I_x = I_y = 26\,\text{kg} \cdot \text{mm}^2$. Punkty te połączono z konstrukcją obudowy za pomocą wielowęzłowego wiązania, w którym odebrano im zarówno przemieszczeniowe, jak i obrotowe stopnie swobody względem ścianek obudowy, tak aby przenieść ich bezwładność na konstrukcję bez lokalnego odkształcania samych ogniw.



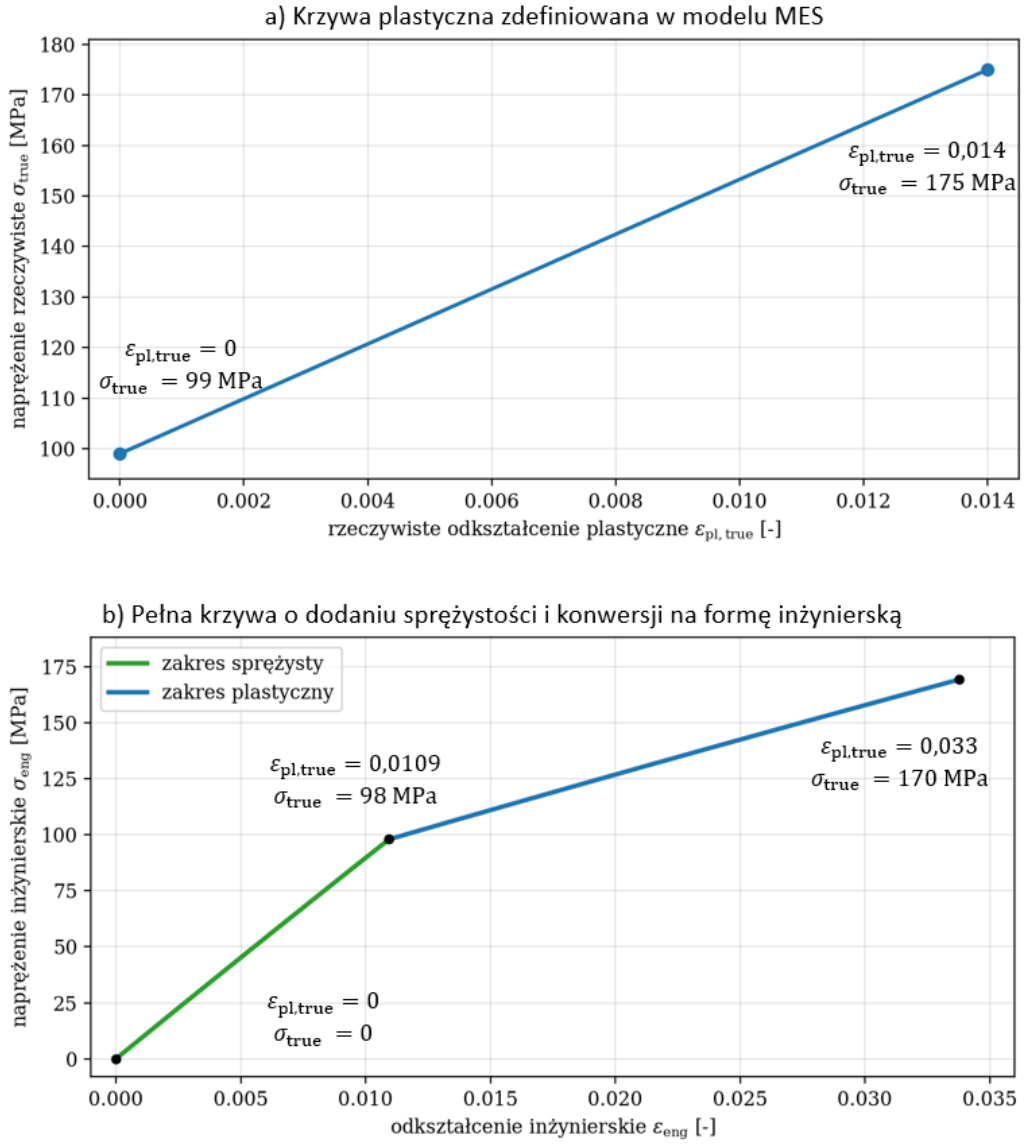
Rysunek 4.2 Model numeryczny MES.

Pod obudową zamodelowano analityczną, idealnie sztywną powierzchnię kontaktową, której wszystkie stopnie swobody zostały zablokowane. Całej obudowie nadano prędkość początkową $\overline{v_0}$ (różną w zależności od analizowanego scenariusza upadków), skierowaną prostopadle do powierzchni podłoża. W ten sposób odwzorowano warunki zrzutu z zadanej wysokości z pominięciem etapu swobodnego spadku (symulacja rozpoczyna się tuż przed uderzeniem o sztywną powierzchnię).

Jako materiał konstrukcyjny obudowy przyjęto szeroko stosowany w wielu gałęziach przemysłu, w tym również w przemyśle bateryjnym PA6-GF30 [42]. Jest to kompozytowe tworzywo sztuczne - poliamid wzmacniany krótkimi włóknami szklanymi w około 30% objętości.

W analizie numerycznej przyjęto zhomogenizowane własności sprężyste z założeniem izotropii materiału. Moduł Younga ustalono na 9.1 GPa, współczynnik Poissona na 0.34, a gęstość na $1350\ \frac{kg}{m^3}$. Model materiału ma charakter sprężysto-plastyczny z kryterium uplastycznienia Hubera-von Misesa-Henckego. Krzywą naprężenie-odkształcenie w zakresie plastycznym przyjęto w postaci biliniowej: materiał wchodzi w stan plastycznego płynięcia przy naprężeniu równym 99 MPa, natomiast drugi punkt krzywej plastycznej zdefiniowano jako 175 MPa przy zredukowanym odkształceniu plastycznym równym 0.014 $\frac{mm}{mm}$, co odpowiada stanowi zbliżonemu do zniszczenia materiału. Na rysunku 4.3 przedstawiono przyjętą charakterystykę materiałową w dwóch ujęciach. Wykres górny pokazuje wyłącznie

zakres plastyczny, czyli zależność naprężenia rzeczywistego od rzeczywistego odkształcenia plastycznego. Taki zapis odpowiada bezpośrednio punktom definiującym biliniową krzywą plastyczną w modelu materiałowym, dlatego nie obejmuje odcinka sprężystego. Wynika to z faktu, że odkształcenie plastyczne jest liczone dopiero od chwili osiągnięcia granicy plastyczności. Wykres dolny ma charakter pomocniczy i przedstawia pełną jednoosiową krzywą naprężenie-odkształcenie po doliczeniu części sprężystej oraz przeliczeniu wartości rzeczywistych na wielkości inżynierskie. Z tego względu oba wykresy nie mają tej samej interpretacji osi odkształcenia: pierwszy odnosi się do danych plastycznych wprowadzanych do modelu, natomiast drugi służy do czytelnej wizualizacji całej charakterystyki materiałowej, obejmującej zarówno zakres sprężysty, jak i plastyczny.



Rysunek 4.3 Wykresy materiałów a) wykres rzeczywistego naprężenia i odkształcenia plastycznego b) wykres naprężenia inżynierskie oraz odkształceń inżynierskich.

W celu implementacji modelu zniszczenia, opisanego szczegółowo w rozdziale 2.1.3, przyjęto kryterium inicjacji uszkodzenia zależne od stanu naprężeń, opisanego współczynnikiem trójosiowości (ang. triaxiality). Graniczne wartości odkształcenia plastycznego inicjujące uszkodzenie przypisano w postaci tabelarycznej jako funkcję trójosiowości naprężeń, zgodnie z danymi, które przedstawia Tabela 4.1. Zależność zniszczenia od stanu naprężenia jest szczególnie istotna w przypadku testu upuszczania, ponieważ przy braku takiej zależności elementy w strefie kontaktu z podłożem ulegałyby zniszczeniu w stanie bliskim jednoosiowemu ściskaniu ($\eta = -0.33$), co nie jest obserwowane w rzeczywistych konstrukcjach, dlatego w tym stanie zniszczenie następuje, przy największej wartości odkształcenia. Dla stanu czystego ścinania ($\eta = 0$) oraz jednoosiowego rozciągania ($\eta = 0.33$), wartości przyjmują wartość uzyskaną z testu jednoosiowego rozciągania.

Tabela 4.1. Parametry inicjacji uszkodzenia.

Odkształcenie przy zerwaniu $\frac{mm}{mm}$	Współczynnik trójosiowości η
1.014	-0.33
0.014	0
0.014	0.33

Zastosowany w niniejszej pracy opis konstytutywny materiału PA6-GF30 odpowiada klasycznemu, izotropowemu modelowi sprężysto-plastycznemu typu J2 z izotropowym umocnieniem (rozdział 2.1.2), rozszerzonemu o model zniszczenia typu „zniszczenie w materiałach ciągliwych" (ang. ductile damage) zależny od trójosiowości naprężeń (rozdział 2.1.3). Mimo że PA6-GF30 jest kompozytem krótkowłóknistym, o istotnie anizotropowych właściwościach lokalnych wynikających z procesu wtrysku, w skali elementu skończonego i w analizach konstrukcyjnych o charakterze globalnym materiał ten jest powszechnie modelowany jako zhomogenizowany ośrodek ciągły. W wielu pracach przyjmuje się dodatkowo założenie izotropii, co pozwala na zastosowanie standardowego kryterium plastyczności von Misesa i prostych praw umocnienia przy zachowaniu zadowalającej zgodności z eksperymentem [42]. Przykładowo, Mouti [57] analizując nisko-prędkościowe uderzenia w miskę olejową z krótkowłóknistego poliamidu przyjęli materiał jako izotropowy i niezależny od szybkości odkształcenia, zastosowali kryterium von Misesa plastyczności oraz odcinkowo-liniową krzywą σ-ε, a zniszczenie modelowali poprzez usuwanie elementów po przekroczeniu granicznego efektywnego odkształcenia plastycznego . Podobne uproszczenie zastosowali Soós i Goda [86] w analizie bossów wykonanych z Ultramidu PA6-GF30, traktując materiał jako jednorodny i izotropowy oraz wykorzystując kryterium von Misesa z izotropowym umocnieniem w połączeniu z lepkosprężystym modelem.

Jednocześnie w literaturze dostępne są bardziej złożone modele anizotropowe dla krótkowłóknistych termoplastów, uwzględniające orientację włókien, zależności szybkościowe oraz wpływ temperatury i wilgotności [44]. Prace Nouri'ego i Meraghniego [66] pokazują, że zachowanie zmęczeniowe wtryskowego PA6-GF30 może być skutecznie opisywane w ramach mechaniki uszkodzeń ciągłych, w której degradacja sztywności zależy od historii odkształceń i stanu naprężenia. Krummenacker i Hausmann [44] wykazali, że homogenizacyjne opisy kompozytów wzmacnianych krótkimi włóknami szklanymi – z wykorzystaniem izotropowych modeli konstytutywnych oraz kryteriów zniszczenia opartych na naprężeniu zredukowanym - dobrze odwzorowują inicjację uszkodzeń w krótkowłóknistych poliamidach. Z kolei nowsze prace Ammara [3] oraz Witzgalla i Wartzacka [96] integrują modele uszkodzenia wzmacniane krótkimi włóknami z symulacjami odporności na uderzenia komponentów, potwierdzając przydatność podejść ciągłych w analizach bezpieczeństwa konstrukcji z PA6-GF30 i pokrewnych kompozytów. Na tle powyższych wyników przyjęty w niniejszej pracy model materiałowy można traktować jako kompromis pomiędzy dokładnością opisu a kosztem obliczeniowym. W porównaniu z przemysłowo stosowanymi modelami izotropowymi Hubera-von Misesa-Henckiego dla elementów z krótkowłóknistych poliamidów model użyty w niniejszej pracy jest rozszerzony o zależny od trójosiowości naprężeń model zniszczenia dla materiałów ciągliwych, w którym inicjacja uszkodzenia jest opisana tablicą granicznych odkształceń plastycznych w funkcji trójosiowości. Taka formuła jest spójna z koncepcjami stosowanymi w modelach uszkodzenia SFRT dla zastosowań zderzeniowych [44], [66], a jednocześnie umożliwia efektywne generowanie wielu rozwiązań numerycznych dla potrzeb optymalizacji kształtu obudowy baterii.

Przyjęte kryterium inicjacji uszkodzenia zależne od trójosiowości naprężeń należy traktować jako fenomenologiczne uproszczenie inżynierskie. Modele lokusu zniszczenia, w których krytyczne odkształcenie plastyczne zależy od miar stanu naprężenia, były pierwotnie rozwijane głównie dla metali, między innymi w pracach Bai i Wierzbickiego [6], [7], [95]. W niniejszej pracy nie są one traktowane jako opis zniszczenia kompozytu PA6-GF30, lecz jako sposób wprowadzenia zależności inicjacji uszkodzenia od stanu naprężenia w globalnym modelu MES. Uzasadnieniem takiego ujęcia jest fakt, że w polimerach i kompozytach polimerowych odkształcenie do zniszczenia zależy od warunków obciążenia, w szczególności od udziału rozciągania, ściskania i ścinania oraz od stanów sprzyjających lokalizacji odkształceń.

Ponieważ analizowany przypadek ma charakter udarowy, dodatkowym ograniczeniem modelu jest pominięcie jawnej zależności parametrów materiałowych od prędkości odkształcenia. Dla PA6-GF30 wzrost prędkości odkształcenia może zwiększać sztywność i wytrzymałość, a jednocześnie zmniejszać odkształcenie przy zerwaniu. Z tego względu

wartości granicznych odkształceń plastycznych przyjęte w Tabeli 4.1 należy interpretować jako parametry efektywne zastosowane na potrzeby obliczeń porównawczych i optymalizacyjnych. Pełna identyfikacja modelu wymagałaby kalibracji zależności odkształcenia granicznego od trójosiowości naprężeń dla badanego materiału, technologii wytwarzania oraz zakresu prędkości odkształcenia odpowiadającego analizowanym upadkom.

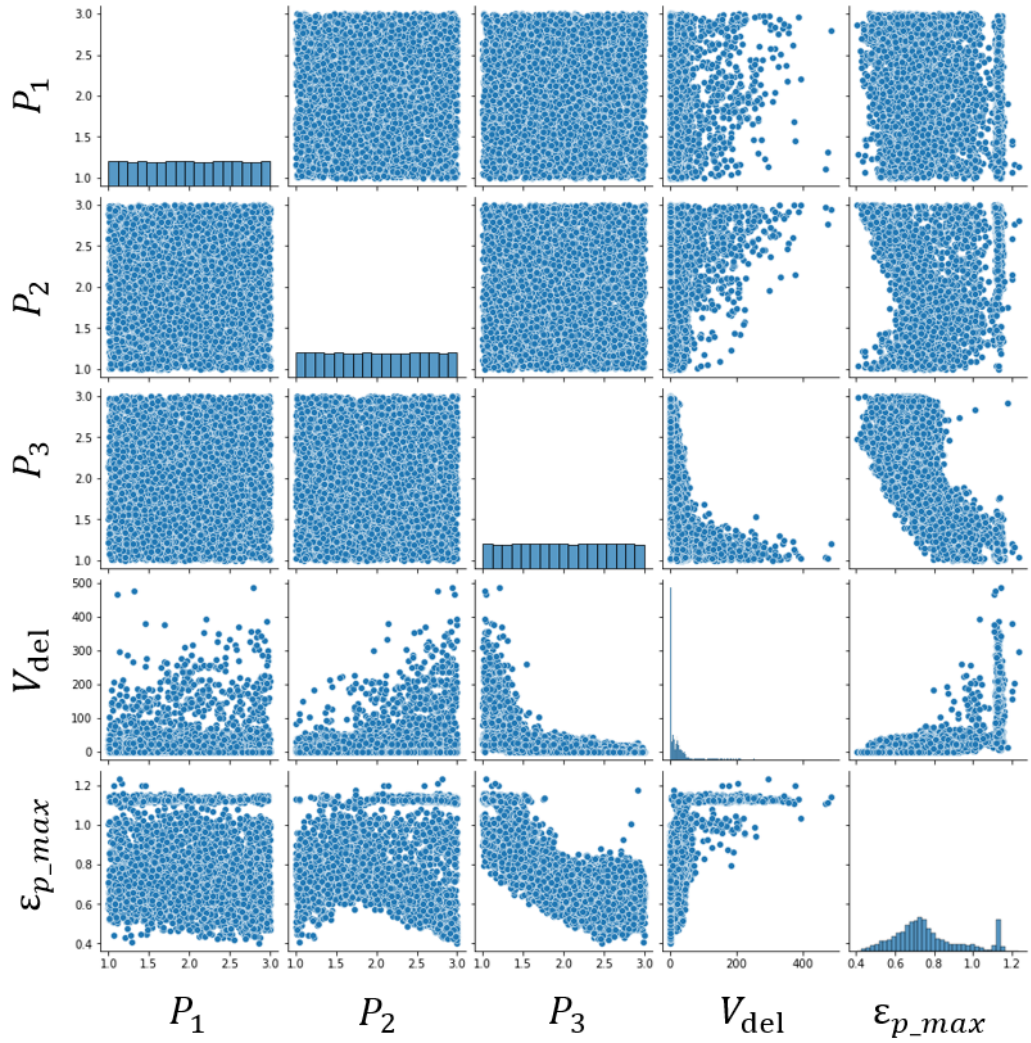
Model zaimplementowano w kodzie Abaqus jako użytkowy podprogram.

Do zastosowań badawczych, w celu zmniejszenia kosztu obliczeniowego przy równoczesnym wykorzystaniu potencjału skryptu automatyzującego obliczenia, wybrano scenariusz zrzutu składający się z dwóch kolejnych upadków w jednym kierunku z prędkościami początkowymi równymi odpowiednio 10.2 $\frac{m}{s}$ oraz 16.5 $\frac{m}{s}$. Prędkości te dobrano tak, aby otrzymywane wyniki obejmowały zarówno przypadki prowadzące do zniszczenia obudowy, jak i przypadki analiz bez zniszczeń.

4.2. Bazy danych zawierające wyniki analiz MES użyte do budowy metamodeli

W celu przeprowadzenia badań nad metamodelowaniem oraz dostrajaniu ich parametrów potrzebne są bazy danych zawierające wektory zmiennych projektowych oraz odpowiadające im wartości funkcji celu. W pierwszym etapie zbudowano bazę danych opartą na projektowaniu eksperymentu z wykorzystaniem metody Hipersześcianu Łacińskiego (LHS) z kryterium wyboru środkowego przedziału oraz maksymalizacją minimalnych odległości opisanych w rozdziale 2.2.1. Wygenerowano 5000 wektorów zmiennych projektowych $\mathbf{P} = (P_1, P_2, P_3)$, dla których przeprowadzono obliczenia numeryczne MES, tworząc bazę *LHS5000*. Każdy rekord zawiera wartości parametrów geometrycznych P_1, P_2, P_3 oraz odpowiadające im wyniki analizy: objętość usuniętych elementów V_{del} oraz maksymalne zredukowane odkształcenie plastyczne $\varepsilon_{p_{\max}}$. Przegląd rozkładu danych w przestrzeni $(P_1, P_2, P_3, V_{\mathrm{del}}, \varepsilon_{p_max})$ przedstawia rysunek 4.4. Histogramy parametrów wejściowych na przekątnej macierzy wykresów wskazują na równomierne pokrycie całego przedziału od 1mm do 3mm dla każdego z parametrów, co jest bezpośrednią konsekwencją zastosowania metody LHS. Wykresy par (P_i, P_j) nie ujawniają istotnych korelacji pomiędzy parametrami - punkty wypełniają cały obszar, zapewniając globalne odwzorowanie przestrzeni projektowej. Zależności pomiędzy parametrami a wynikami są wyraźnie nieliniowe: w szczególności w relacjach (P_2, V_{del}) oraz (P_3, V_{del}) obserwuje się silny trend malejący, co wskazuje, że wzrost P_2 i P_3 prowadzi do ograniczenia objętości zniszczonego materiału. Dla P_1 wpływ na V_{del} jest słabszy. Rozkład V_{del} jest silnie skośny, z przewagą przypadków o niewielkim ubytku objętości, natomiast relacja

$(V_{\mathrm{del}}, \varepsilon_{p_max})$ wskazuje, że większym objętościom zniszczenia towarzyszą większe maksymalne odkształcenia plastyczne.



Rysunek 4.4. Rozkład danych w bazie wygenerowanej za pomocą LHS.

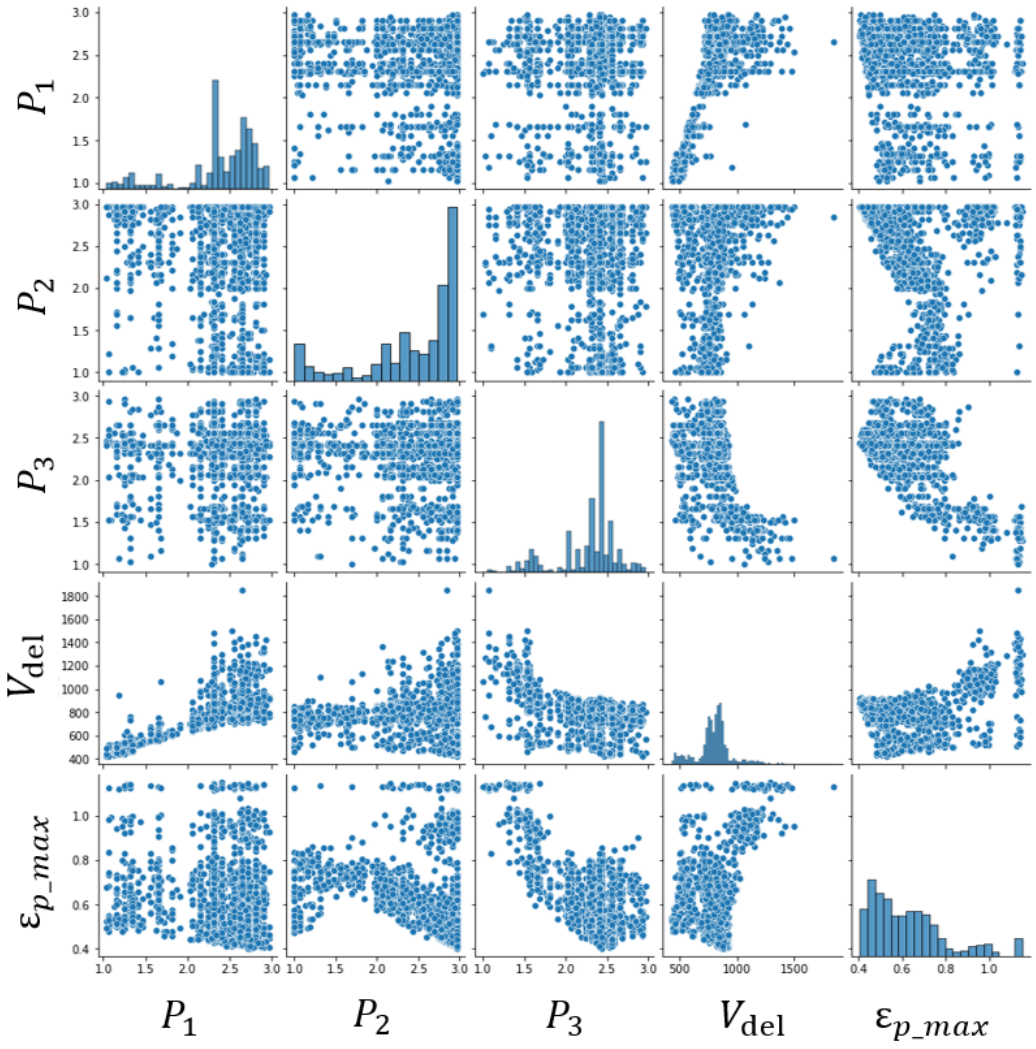
W celu zbadania wpływu liczebności zbioru danych na jakość metamodeli, z bazy *LHS5000* utworzono mniejsze, kontrolowane podzbiory: *LHS2500*, *LHS1250*, *LHS625*, *LHS312*, *LHS156*, zawierające odpowiednio 2500, 1250, 625, 312 i 156 próbek. Podzbiory o mniejszej liczebności nie były wybierane losowo bezpośrednio z bazy LHS5000. Dla każdej wymaganej liczby próbek najpierw wygenerowano niezależny plan LHS w tym samym obszarze zmiennych projektowych, a następnie dla każdego punktu tego planu wyszukano najbliższy odpowiadający mu punkt w bazie LHS5000. W ten sposób uzyskano zredukowane zbiory danych, które zachowywały globalny charakter pokrycia przestrzeni projektowej, ale zawierały stopniowo mniejszą liczbę przypadków obliczeniowych, zgodnie z metodą opisaną w rozdziale

3.3. Zbiory te zostały następnie wykorzystane w analizie wpływu rozmiaru bazy danych na proces uczenia i zdolność generalizacji metamodeli.

Drugą grupę stanowią bazy danych oparte na wynikach zadań optymalizacyjnych z zastosowaniem algorytmu optymalizacji QEA z rozdzielczością 6 kubitów dla każdej zmiennej. Rozdzielczość ΔP dla danej liczby kubitów $kubit$, przyjmując, że $P_{\max}$ oraz $P_{\min}$ to maksymalna i minimalna wartość danego parametru, można obliczyć w następujący sposób:

$$\Delta P = \frac{P_{\max} - P_{\min}}{2^{kubit}} = \frac{3mm - 1mm}{2^6} = \frac{2}{64} = 0.03125mm \tag{4.2}$$

Taka dokładność jest wystarczającym przybliżeniem z punktu widzenia procesu technologicznego. Dane z wielu przebiegów optymalizacji połączono w jeden zbiór, który następnie losowo przemieszano i podzielono na mniejsze podzbiory. Otrzymano bazy *QEA1687, QEA625, QEA312, QEA156*, zawierające kolejno 1687, 625, 312 i 156 próbek. Największy zbiór, *QEA1687*, pokazuje rysunek 4.5. W przeciwieństwie do LHS, rozkłady parametrów wejściowych w bazie QEA są wyraźnie nierównomierne: histogramy P_1, P_2 i P_3 wykazują koncentrację punktów w stosunkowo wąskich przedziałach, natomiast pozostałe obszary przestrzeni projektowej są słabiej reprezentowane. Jest to naturalny efekt działania algorytmu ewolucyjnego, który zagęszcza próby w otoczeniu rozwiązań o korzystnych wartościach funkcji celu. Wykresy par (P_i, P_j) ujawniają przy tym „pasmowe" struktury rozkładu punktów, świadczące o korelacjach pomiędzy parametrami w pobliżu potencjalnych optimum. Rozkład V_{del} w bazie QEA jest przesunięty w kierunku większych wartości, a rozkład ε_{p_max} bardziej skupiony, co oznacza, że większość próbek odpowiada konfiguracjom o istotnym poziomie zniszczenia i zbliżonym zakresie odkształceń plastycznych - zgodnie z oczekiwaniem dla danych generowanych w procesie optymalizacji.



Rysunek 4.5 Rozkład danych w bazie wygenerowanej za pomocą zbierania danych z zadań rozwiązanych z użyciem algorytmu QEA.

Z punktu widzenia budowy metamodeli obie rodziny baz danych mają ciekawe własności. Zbiory LHS zapewniają globalne, równomierne pokrycie przestrzeni parametrów (P_1, P_2, P_3) i dostarczają informacji o ogólnej strukturze funkcji celu w całym rozpatrywanym zakresie. Z kolei bazy QEA odzwierciedlają rzeczywisty przebieg procesu optymalizacji i zawierają zagęszczenie próbek w obszarach istotnych z punktu widzenia inżynierskiego - w pobliżu rozwiązań o korzystnych wartościach funkcji celu. W dalszej części pracy zbiory typu LHS oraz typu QEA są wykorzystywane rozdzielnie. Dla każdej bazy (i jej podzbiorów) budowane są niezależne metamodele, a następnie porównywana jest ich skuteczność predykcyjna. Pozwala to ocenić wpływ charakteru próbkowania przestrzeni projektowej - równomiernego, eksperymentalnego (LHS) oraz skoncentrowanego w otoczeniu dobrych rozwiązań (QEA) –

na jakość uzyskiwanych modeli zastępczych i ich przydatność w zadaniach optymalizacji konstrukcji.

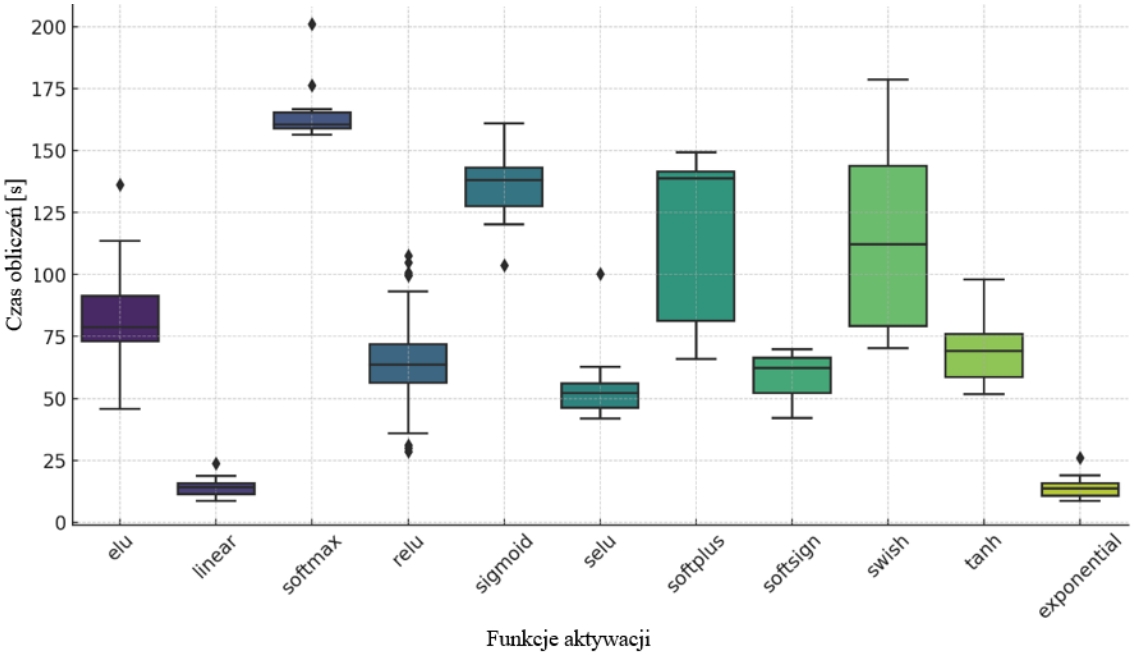
4.3. Strojenie hiperparametrów oraz testy metamodeli

W niniejszym podrozdziale przedstawiono sposób strojenia hiperparametrów metamodeli wykorzystywanych do aproksymacji funkcji celu FF1 (3.4) i FF2 (3.5) w zadaniu optymalizacji obudowy baterii litowo-jonowych.

4.3.1. Hiperparametry Sztucznej Sieci Neuronowej

Strojenie hiperparametrów dla Sztucznych Sieci Neuronowych (SSN) przeprowadzono w pierwszej kolejności dla funkcji celu FF1 (3.4). W celu wyznaczenia możliwie uniwersalnej konfiguracji, która mogłaby zostać zastosowana do wszystkich analizowanych baz, do strojenia hiperparametrów wybrano największą i najbardziej jednorodną bazę LHS5000. Zawiera ona 5000 próbek wektora zmiennych projektowych **P**, wygenerowanych metodą Latin Hypercube Sampling, co zapewnia dobre wypełnienie przestrzeni projektowej i ogranicza ryzyko nadmiernego dopasowania do lokalnych zagęszczeń danych. Na etapie przygotowania danych wejściowych zastosowano standaryzację, po czym zbiór danych podzielono na część uczącą (80%), walidacyjną (10%) oraz testową (10%). Uczenie realizowano z użyciem optymalizatora Adam [40], maksymalną liczbę epok przyjęto równą 1000, natomiast mechanizm wczesnego zatrzymania przerywał proces, jeżeli przez 50 kolejnych epok nie odnotowano poprawy wartości błędu walidacyjnego. W celu ograniczenia ryzyka przeuczenia wprowadzono regularyzację w postaci warstwy wygaszania (ang. dropout) o prawdopodobieństwie dezaktywacji neuronów równym 0,2. Optymalizację hiperparametrów przeprowadzono w 500 próbach, traktując tę liczbę jako kompromis pomiędzy dokładnością przeszukania przestrzeni parametrów a czasem obliczeń. Do realizacji przeszukiwania wykorzystano bibliotekę Optuna [68] w środowisku Python, implementującą podejście optymalizacji bayesowskiej omówione w rozdziale 2.2.5. Pierwszy etap prac poświęcono doborowi funkcji aktywacji w warstwach ukrytych. Jako konfigurację referencyjną przyjęto dwuwarstwową sieć SSN z 30 neuronami w każdej warstwie, a następnie porównano zestaw popularnych funkcji aktywacji: *ReLU, Sigmoid, Swish, Softplus, ELU, Tanh, Leaky ReLU, Softmax, Linear, Hard Sigmoid* oraz *Exponential* [90]. Wyniki zestawiono w Tabela 4.2 oraz pokazano na rysunku 4.6. Najkorzystniejsze rezultaty uzyskano dla *ReLU*, dla której średnia wartość współczynnika determinacji wyniosła $R^2 \approx 0{,}8155$, przy najniższym odchyleniu standardowym $\mathrm{std}\, R^2 \approx 0{,}0032$ oraz relatywnie krótkim czasie uczenia rzędu 64 s. Funkcje *Sigmoid, Swish, Softplus* i *ELU* osiągały wyniki zbliżone, jednak nieco gorsze pod względem R^2, a jednocześnie wymagały wyraźnie dłuższego czasu uczenia. Najsłabszą jakość aproksymacji odnotowano

dla funkcji *Softmax, Linear, Hard Sigmoid oraz Exponential*. Z perspektywy praktycznej *ReLU* zapewniała najbardziej korzystny kompromis pomiędzy dokładnością predykcji, stabilnością uczenia i kosztem obliczeniowym, dlatego przyjęto ją jako domyślną funkcję aktywacji w kolejnych analizach.



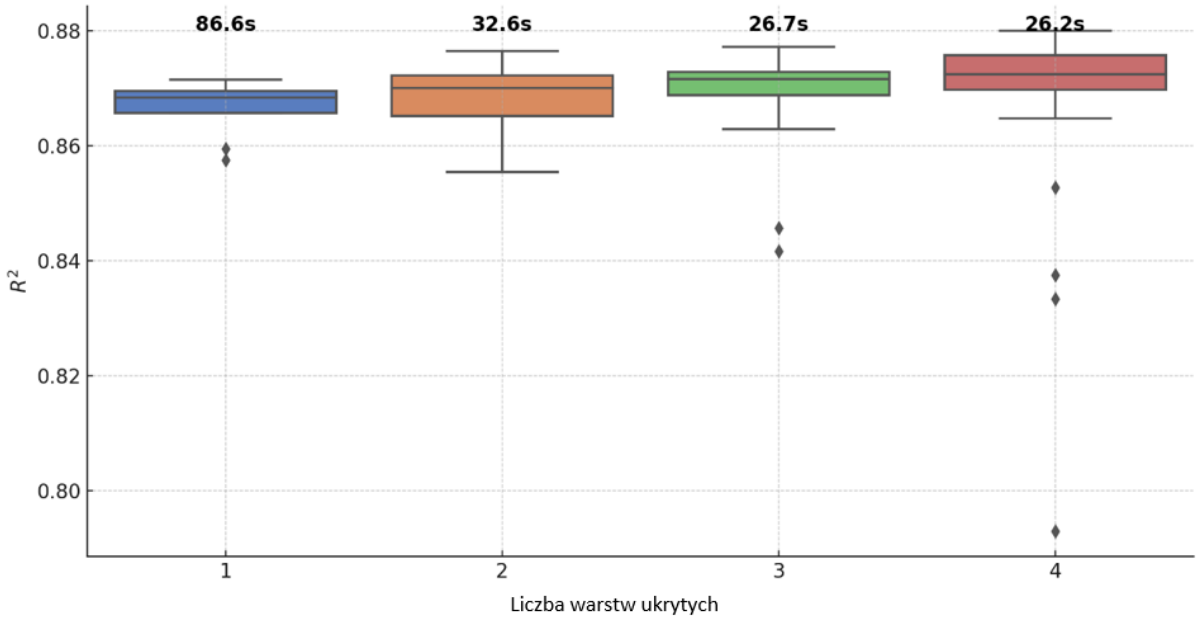
Rysunek 4.6 Rozkład czasu uczenia dla różnych funkcji aktywacji.

Tabela 4.2. Zestawienie wyników uczenia SSN dla różnych funkcji aktywacji.

Funkcja aktywacji	Liczba prób	R^2	Odch. Std. R^2	Średni czas [s]
ReLU	347	0.8155	0.0032	64.46
Sigmoid	15	0.8135	0.0022	135.32
Swish	15	0.8108	0.0053	113.73
Softplus	15	0.8014	0.0063	119.53
ELU	16	0.7924	0.0063	82.22
Tanh	16	0.7821	0.0071	109.87
Leaky ReLU	16	0.7758	0.0085	91.34
Softmax	15	0.6212	0.0124	159.34
Linear	15	0.3628	0.0093	18.69
Hard Sigmoid	15	0.3415	0.0152	22.41
Exponential	15	0.2894	0.0185	31.57

W kolejnym kroku zbadano jednoczesny wpływ liczby warstw ukrytych (od 1 do 4) oraz liczby neuronów w warstwie (od 2 do 100) na jakość predykcji i czas obliczeń.

Przeprowadzono łącznie 100 testów, których wyniki przedstawiono w postaci map i wykresów (rysunek 4.7). W tym celu wykorzystano przeszukiwanie bayesowskie (TPE) przy wykorzystaniu biblioteki Optuna. Analiza wykazała, że w rozpatrywanym problemie liczba warstw i neuronów ma stosunkowo niewielki wpływ na średnią wartość R^2, natomiast w większej mierze oddziałuje na czas obliczeń i stabilność uczenia. Sieci z jedną warstwą ukrytą wymagały zauważalnie dłuższego czasu uczenia przy porównywalnej jakości aproksymacji, natomiast sieci trój- i czterowarstwowe trenowały się szybciej i często osiągały nieco lepsze wyniki R^2, jednak wykazywały okresowo pogorszenie jakości predykcji (pojedyncze realizacje o wyraźnie niższym R^2). Z tego względu jako konfigurację bazową przyjęto sieć z dwiema warstwami ukrytymi, która łączy akceptowalny czas obliczeń z dobrą stabilnością.

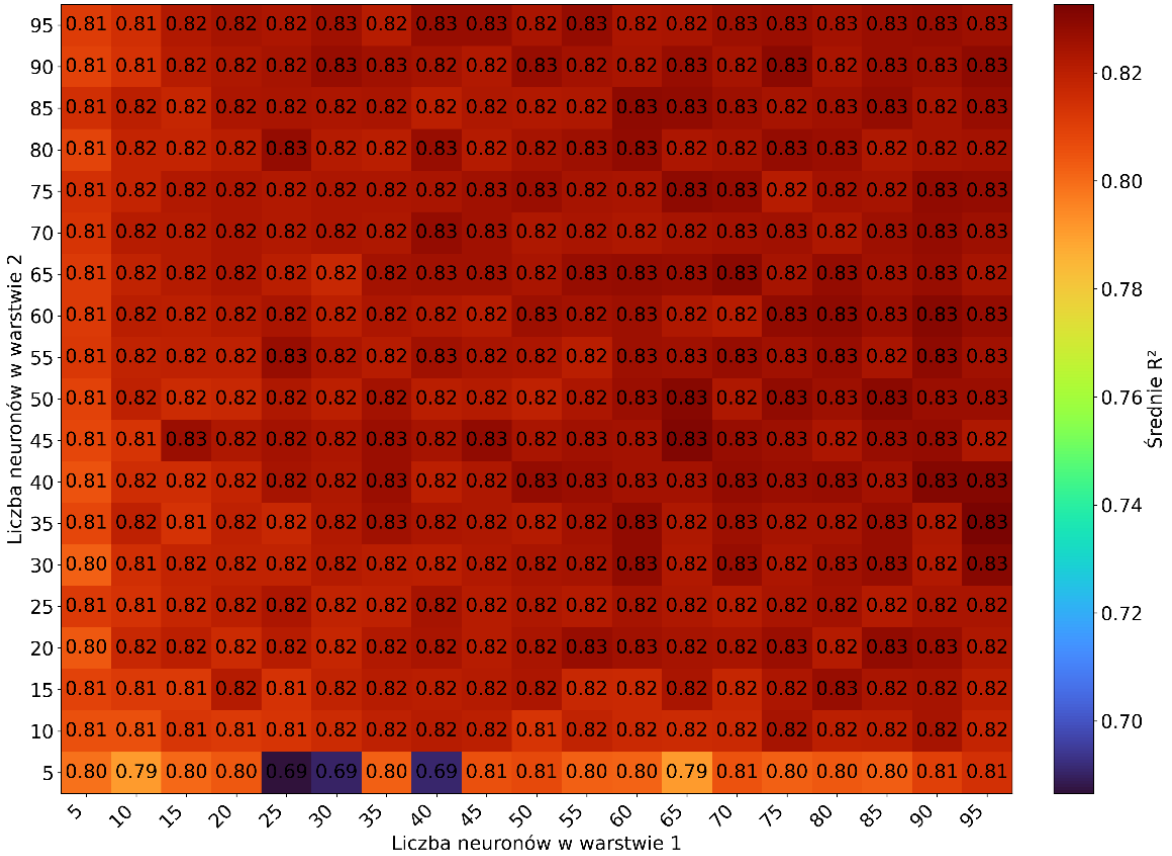


Rysunek 4.7 Rozkład współczynnika determinacji w dziedzinie liczby ukrytych warstw neuronów ze średnim czasem obliczeń.

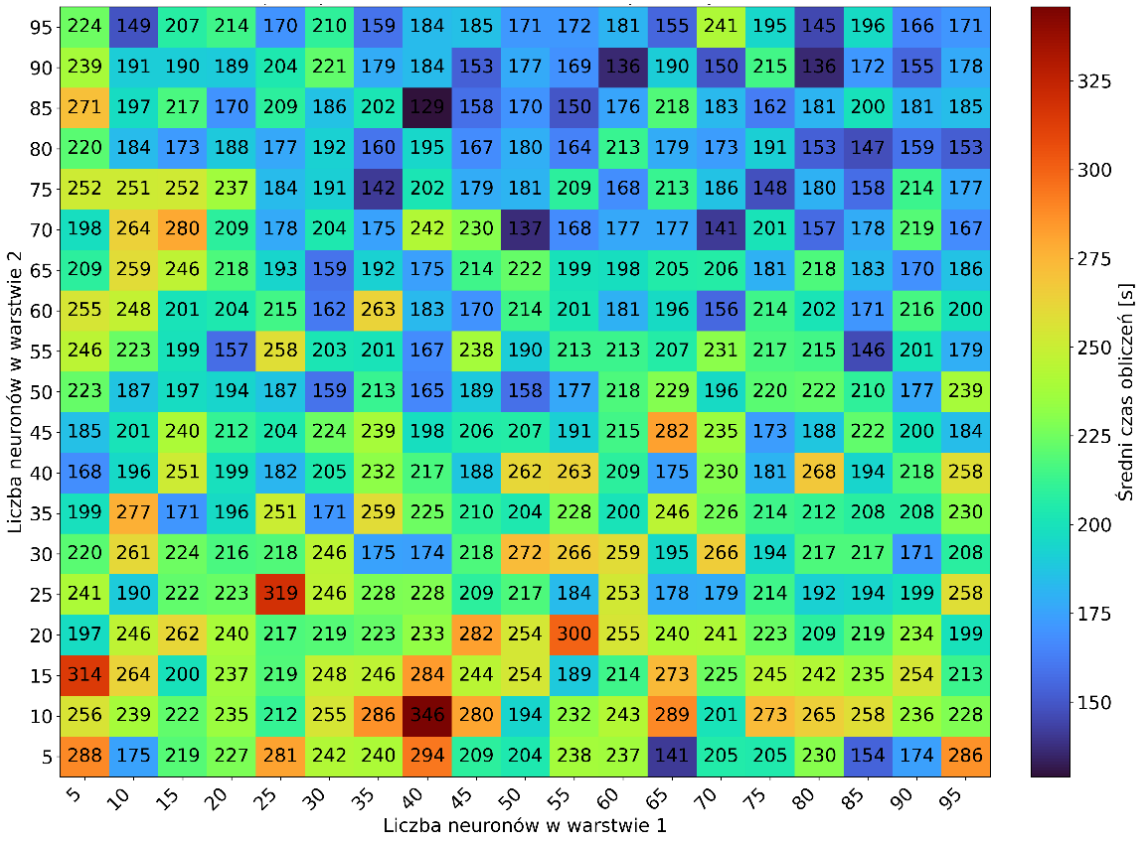
Do szczegółowego doboru liczby neuronów w warstwach zastosowano przeszukiwanie siatkowe (ang. grid search) konfiguracji, oceniając wszystkie kombinacje w zadanym zakresie i wizualizując uzyskane wartości R^2 i czasów obliczeń w postaci map ciepła (rysunek 4.8 oraz 4.9). Wykorzystano w tym celu narzędzie Optuna, ale z wymuszeniem przeszukania wszystkich konfiguracji, co odpowiada przeszukiwaniu siatkowemu. Wytrenowano łącznie 1805 sieci (po 5 na każdą konfigurację), a w celu oceny wpływu losowego podziału danych, zbiory uczący, walidacyjny i testowy były za każdym razem losowane na nowo. Analiza wykazała, że sieć zachowuje się stabilnie przy zastosowaniu funkcji aktywacji ReLU oraz dwóch warstw zawierających co najmniej 50 neuronów każda. W modelu wykorzystano

współczynnik wygaszania neuronów (ang. *dropout*) równy 0,2 oraz optymalizator Adam. Zastosowano również mechanizm wczesnego zatrzymania po 50 epokach bez poprawy wartości funkcji straty dla zbioru walidacyjnego. Wahania jakości predykcji, wynikające wyłącznie z losowego podziału danych, nie mają istotnego wpływu na dalsze analizy. Jednocześnie zmniejszanie liczby neuronów nie powoduje wyraźnego i systematycznego skrócenia czasu obliczeń.

Z tego powodu w dalszych badaniach przyjęto co najmniej 50 neuronów w każdej z dwóch warstw ukrytych.



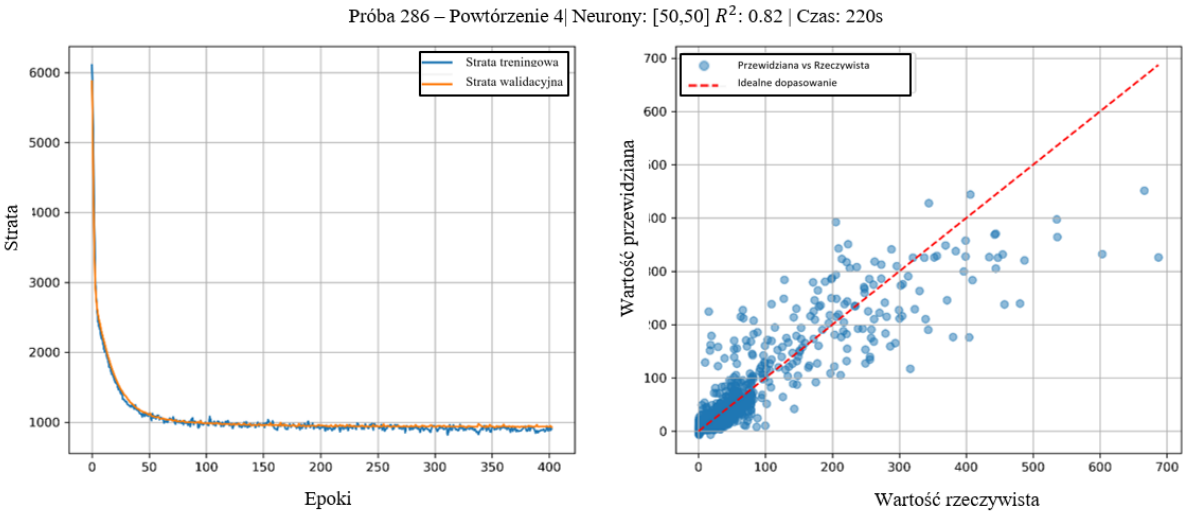
Rysunek 4.8 Mapa cieplna średnich wyników współczynnika determinacji.



Rysunek 4.9. Mapa cieplna średnich czasów obliczeń.

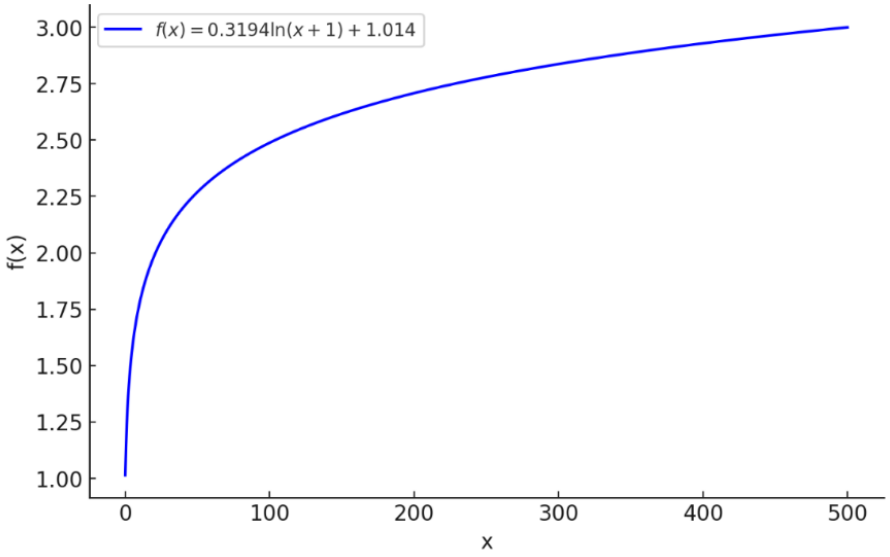
Krzywe uczenia (strata treningowa i walidacyjna) gwałtownie maleją w początkowych epokach, po czym tempo poprawy wyraźnie spada, przy czym ich przebiegi są do siebie zbliżone, co sugeruje stabilne uczenie sieci bez jednoznacznych oznak przeuczenia (rysunek 4.10). Jednocześnie wykres wartości rzeczywistych do przewidywanych pokazuje silną koncentrację próbek w obszarze małych wartości oraz systematyczne odchylenia od prostej idealnej y = x, zwłaszcza dla większych wartości rzeczywistych, gdzie predykcje są często zaniżane, a rozbieżności rzędu kilkudziesięciu jednostek są widoczne. W kontekście zadania, w którym krytyczny jest obszar wartości poniżej $\varepsilon_{p_max} = 1.014$, taki charakter błędów oznacza, że model może nie zapewniać wystarczającej rozdzielczości predykcji w pobliżu progu decyzyjnego, pomimo pozornie poprawnego przebiegu strat. Wbrew stabilnemu przebiegowi strat treningowej i walidacyjnej, analiza zależności wartości przewidywanych do rzeczywistych ujawniła błąd systematyczny oraz ryzyko niewystarczającej czułości modelu w sąsiedztwie $\varepsilon_{p\max} = 1.014$. Ponieważ w zadaniu kluczowa jest poprawna klasyfikacja względem tego progu, w kolejnym kroku podjęto prace nad modyfikacją sposobu definiowania celu optymalizacji i rozpoczęto testy funkcji FF2 (3.5). FF1(3.4) zostanie ponownie

sprawdzona w rozdziale 4.4.1 oraz 4.4.2, jednak główna część prac związana jest z optymalizacją z użyciem FF2 (3.5).

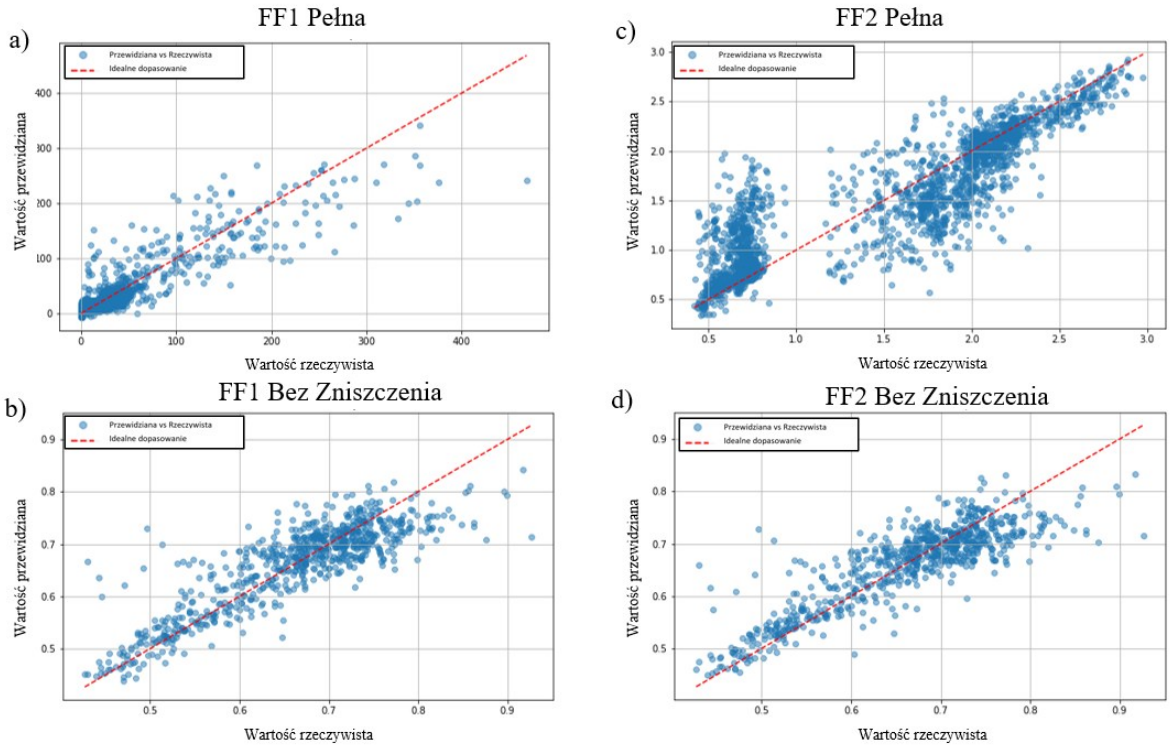


Rysunek 4.10. Przykład wykresu uczenia pochodzący z eksperymentu dla konfiguracji 50x50 neuronów dla FF1

W dalszej części rozważono funkcję celu FF2, której parametry a=0.3194 i b=1.014 dobrano w taki sposób, aby dla zerowej objętości zniszczonego materiału funkcja przyjmowała wartość 1,014 odpowiadającą maksymalnej dopuszczalnej odkształcalności plastycznej materiału ε_{p_max}. Dla wartości funkcji celu rzędu 500 wartości FF2 (3.5) rosną do ok. 3, co zapewnia bardziej zwartą i inżyniersko interpretowalną skalę wyników (rysunek 4.11). Krzywe uczenia dla FF1 (3.4) i FF2 (3.5) (rysunek 4.12) wskazują, że w obszarze granicznym pomiędzy analizami z uszkodzeniem i bez uszkodzenia sieć ma trudność w jednoznacznej aproksymacji funkcji celu, co objawia się pogorszeniem jakości predykcji w pobliżu progu inicjacji zniszczenia. Z tego powodu zdecydowano się na zastosowanie dwóch odrębnych sieci SSN: pierwszą trenowaną na pełnej bazie (*Full*), zawierającej również przypadki z uszkodzeniem elementu, oraz drugiej trenowanej na bazie ograniczonej do przypadków bez zniszczenia (*NoDamage*).



Rysunek 4.11 Funkcja skalująca wartość funkcji celu FF2 (3.5) w stosunku do FF1 (3.4).



Rysunek 4.12 Przykłady krzywych uczenia dla FF1 (3.4) oraz FF2 (3.5) oraz podziału na bazy bez zniszczenia oraz ze zniszczeniem.

Rysunek 4.12 pokazuje, że zależność „predykcja vs. wartość rzeczywista” dla modeli trenowanych na pełnej bazie (FF1 (3.4) *Full*, FF2 (3.5) *Full*) ma wyraźnie większy rozrzut względem linii idealnej niż w wariantach ograniczonych do przypadków bez zniszczenia (FF1

(3.4) *NoDamage*, FF2 (3.5) *NoDamage*). W szczególności dla FF2 (3.5) *Full* pojawiają się dwa odmienne podzbiory punktów: skupienie w dolnym zakresie wartości (typowe dla przypadków bez zniszczenia) oraz wyraźnie bardziej rozproszony zbiór punktów dla wyższych wartości (przypadki z uszkodzeniem), przy czym w obszarze przejściowym rozrzut istotnie rośnie, a błędy mają charakter nielosowy. To jest bezpośrednia przesłanka, że pojedyncza sieć uczona na mieszanej funkcji celu próbuje jednocześnie aproksymować dwa jakościowo różne zachowania wynikające z inicjacji zniszczenia, co degraduje jakość predykcji właśnie w pobliżu progu. Dla porównania, wykresy *NoDamage* pokazują znacznie bardziej zwartą chmurę punktów wzdłuż linii idealnej w ograniczonym przedziale wartości, co oznacza, że w tej podprzestrzeni odwzorowanie jest bardziej gładkie i jednorodne, a więc łatwiejsze do wiarygodnej aproksymacji. W konsekwencji rozdzielenie na dwa modele SSN jest uzasadnione empirycznie: redukuje problem mieszania próbek (bez zniszczenia i ze zniszczeniem), poprawia dopasowanie w każdej z podprzestrzeni i ogranicza błędy w obszarze granicznym.

Po wyznaczeniu konfiguracji sieci na bazie *LHS5000* przeanalizowano wpływ rozmiaru bazy danych oraz rodzaju funkcji celu (FF1 (3.4) vs FF2 (3.5)) na jakość predykcji. Tabela 4.3 zestawia wartości R^2, błędów walidacyjnych i treningowych dla szeregu baz LHS o różnej liczności (*LHS5000, LHS2500, LHS1250, LHS625, LHS312, LHS156*) oraz dla obu modeli (*Full* i *NoDamage*). Obliczone dane są uśrednionymi wartościami dla 5-ciu różnych modeli. Dla największych baz (5000-2500 punktów) uzyskano wartości R^2 na poziomie ok. 0,75-0,83, przy zbliżonych błędach treningu i walidacji, co wskazuje na poprawne uogólnianie. Wraz ze zmniejszaniem liczebności bazy jakość aproksymacji stopniowo się pogarsza, co jest szczególnie widoczne dla LHS156, gdzie dla modelu *NoDamage* obserwuje się nawet ujemne wartości R^2, świadczące o silnym przeuczeniu lub niedostatecznej informacji w danych. Porównanie dla FF1 (3.4) i FF2 (3.5) pokazuje, że dla mniejszych baz funkcja FF2 (3.5) zazwyczaj zapewnia nieco wyższe wartości R^2, co jest istotne z punktu widzenia zastosowań inżynierskich, gdzie liczba dostępnych prób może być ograniczona. Jednak na poziomie ogólnym wyniki są porównywalne i nie przemawiają na korzyść żadnej z funkcji celu. Analogiczną analizę przeprowadzono dla baz QEA (QEA1687, QEA625, QEA312, QEA156), uzyskując podobne wnioski: największe bazy QEA pozwalają na budowę metamodeli SSN o stabilnej jakości predykcyjnej, rozumianej jako uzyskiwanie dodatnich i powtarzalnych wartości R^2 w kolejnych powtórzeniach uczenia oraz brak wyraźnej rozbieżności między błędem treningowym i walidacyjnym. W przypadku najmniejszych baz jakość predykcji wyraźnie spada. Na podstawie tego badania ustalono, że bezpiecznym progiem minimalnej wielkości bazy w tym przypadku jest 300 osobników, ponieważ jest to wielkość, przy której można tworzyć pierwsze metamodele o powtarzalnej jakości predykcji. Co istotne, w wariancie FF1 (3.4), po rozdzieleniu na dwa modele (*Full* i *NoDamage*), uzyskuje się dla baz ≥ 300

dodatnie, powtarzalne wartości R^2 oraz zbliżone błędy treningu i walidacji, a różnice względem FF2 (3.5) mieszczą się w rozrzucie wynikającym z pięciu niezależnych powtórzeń uczenia, co potwierdza porównywalną jakość predykcji.

Tabela 4.3. Wyniki porównania baz danych o różnych wielkościach i sposobie utworzenia.

SSN z funkcją celu FF1 uczony na bazie LHS						
Baza LHS	Współ. determinacji R^2		MSE_{val}		MSE_{test}	
	Baza *Full*	Baza *NoDamage*	Baza *Full*	Baza *NoDamage*	Baza *Full*	Baza *NoDamage*
LHS5000	0.83	0.74	426.78	0.0022	452.45	0.0021
LHS2500	0.82	0.76	445.71	0.0021	377.84	0.0021
LHS1250	0.83	0.74	455.71	0.0022	527.78	0.0022
LHS625	0.75	0.56	813.32	0.0036	1092.68	0.0043
LHS312	0.82	0.62	583.25	0.0031	660.37	0.0057
LHS156	0.63	-0.60	731.14	0.0113	1519.44	0.0136
SSN z funkcją celu FF2 uczony na bazie LHS						
Baza LHS	Współ. determinacji R^2		MSE_{val}		MSE_{test}	
	Baza *Full*	Baza *NoDamage*	Baza *Full*	Baza *NoDamage*	Baza *Full*	Baza *NoDamage*
LHS5000	0.77	0.75	0.11	0.0021	0.1094	0.0020
LHS2500	0.75	0.76	0.12	0.0022	0.1288	0.0022
LHS1250	0.73	0.72	0.13	0.0023	0.1402	0.0023
LHS 625	0.70	0.58	0.15	0.0034	0.1702	0.0040
LHS 312	0.73	0.32	0.14	0.0053	0.1585	0.0093
LHS156	0.64	-1.02	0.17	0.0136	0.1411	0.0164
SSN z funkcją celu FF1 uczony na bazie QEA						
Baza QEA	Współ. determinacji R^2		MSE_{val}		MSE_{test}	
	Baza *Full*	Baza *NoDamage*	Baza *Full*	Baza *NoDamage*	Baza *Full*	Baza *NoDamage*
QEA1687	0.748	0.83	197.50	0.0016	255.21	0.0020
QEA625	0.51	0.84	285.90	0.0016	139.06	0.0015
QEA312	0.51	0.86	126.92	0.0014	166.39	0.0013
QEA156	0.32	0.46	70.56	0.0062	96.116	0.0060
SSN z funkcją celu FF2 uczony na bazie QEA						
	Współ. determinacji R^2		MSE_{val}		MSE_{test}	

Baza QEA	Baza *Full*	Baza *NoDamage*	Baza *Full*	Baza *NoDamage*	Baza *Full*	Baza *NoDamage*
QEA1687	0.75	0.86	0.11	0.0015	0.1171	0.0015
QEA625	0.75	0.83	0.11	0.0019	0.1335	0.0022
QEA312	0.66	0.77	0.16	0.0023	0.1815	0.0025
QEA156	0.62	0.24	0.20	0.0084	0.1312	0.0072

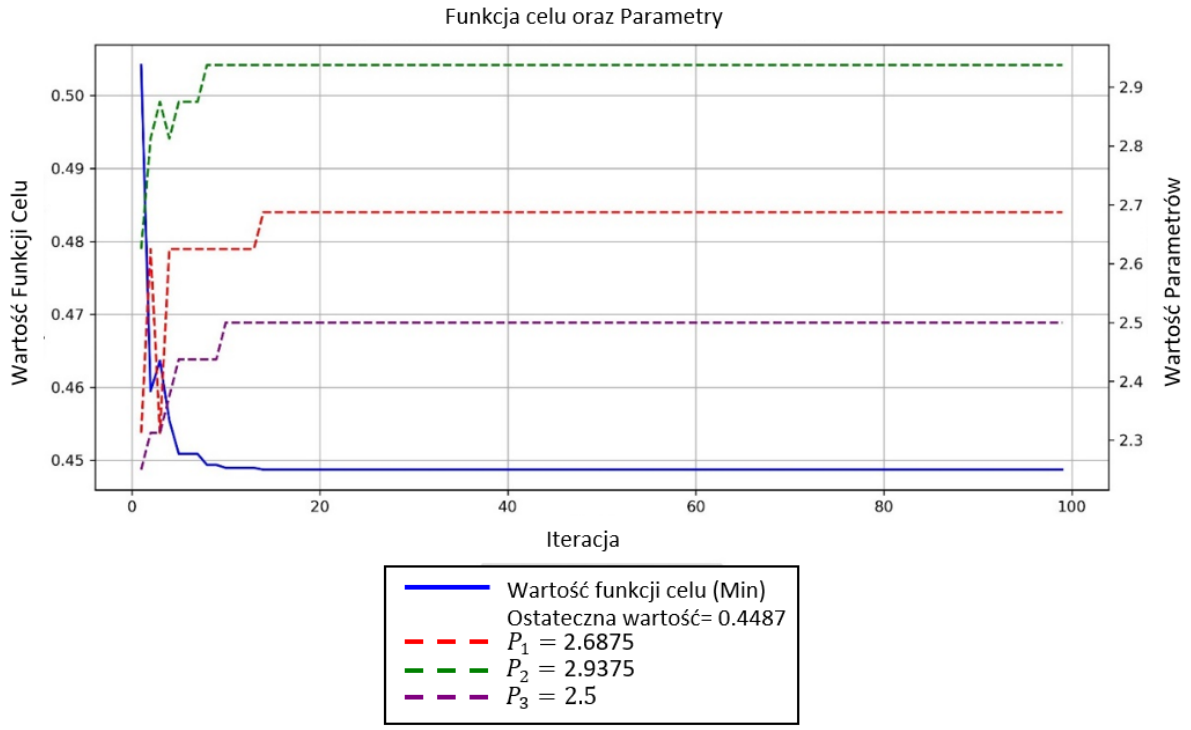
W następnym etapie sprawdzono, jak wytrenowane sieci SSN sprawdzają się w praktycznym zadaniu optymalizacji z użyciem algorytmu QEA. Dla wybranych konfiguracji metamodeli przeprowadzono serie 16 niezależnych biegów optymalizacji (po cztery na każdy wyuczony model), korzystając wyłącznie z predykcji SSN (bez dodatkowych obliczeń MES). Metoda ta została opisana w rozdziale 3.5. Badanie przeprowadzano na największych dostępnych bazach *QEA1687* i *LHS5000*. Tabela 4.4 przedstawia przykłady wyników, w których porównano wartości funkcji celu uzyskane z metamodelu z wynikami z pełnych analiz MES dla tego samego osobnika.

Tabela 4.4. Porównanie przewidywań SSN budowanych na tych samych bazach danych, na różnych treningach i różnych próbach predykcji.

Optymalizacja QEA dla FF2 - bazująca na *LHS5000*				
Model [R^2]	Próba	Wartość funkcji celu	Wektor zmiennych projektowych	Wartość funkcji celu MES
Full= 0.792 *NoDamage*=0.78	1	0.439186	2.375,2.9375,2.4375	0.44241
	2	0.435265	2.8125,2.9375,2.625	0.44878
	3			
	4	0.439186	2.375,2.9375,2.4375	0.44241
Full= 0.7847 *NoDamage* =0.771	1	0.461702	2.9375,2.9375,2.5	0.46484
	2			
	3			
	4			
Full=0.7844 *NoDamage* =0.754	1	0.45215	2.25,2.9375,2.5	0.445811
	2			
	3			
	4			
Full=0.7820	1	0.431665	2.9375,2.9375,2.5	0.46484
	2			

NoDamage =0.751	3			
	4			
Optymalizacja QEA dla FF2 - bazująca na *QEA1687*				
Full =0.7213 *NoDamage* =0.8713	1	0.437105	2.75,2.9375,2.4375	0.41423
	2			
	3			
	4			
Full =0.7140 *NoDamage* =0.8761	1	0.439528	2.6875,2.9375,2.4375	0.411295
	2			
	3	0.439374	2.5625,2.9375,2.5	0.42695
	4	0.438613	2.625,2.9375,2.4375	0.42717
Full =0.7260 *NoDamage* =0.8721	1	0.438376	2.75,2.9375,2.5	0.434021
	2			
	3			
	4			
Full =0.7396 *NoDamage* =0.9749	1	0.448706	2.6875,2.9375,2.5	0.42717
	2			
	3			
	4			

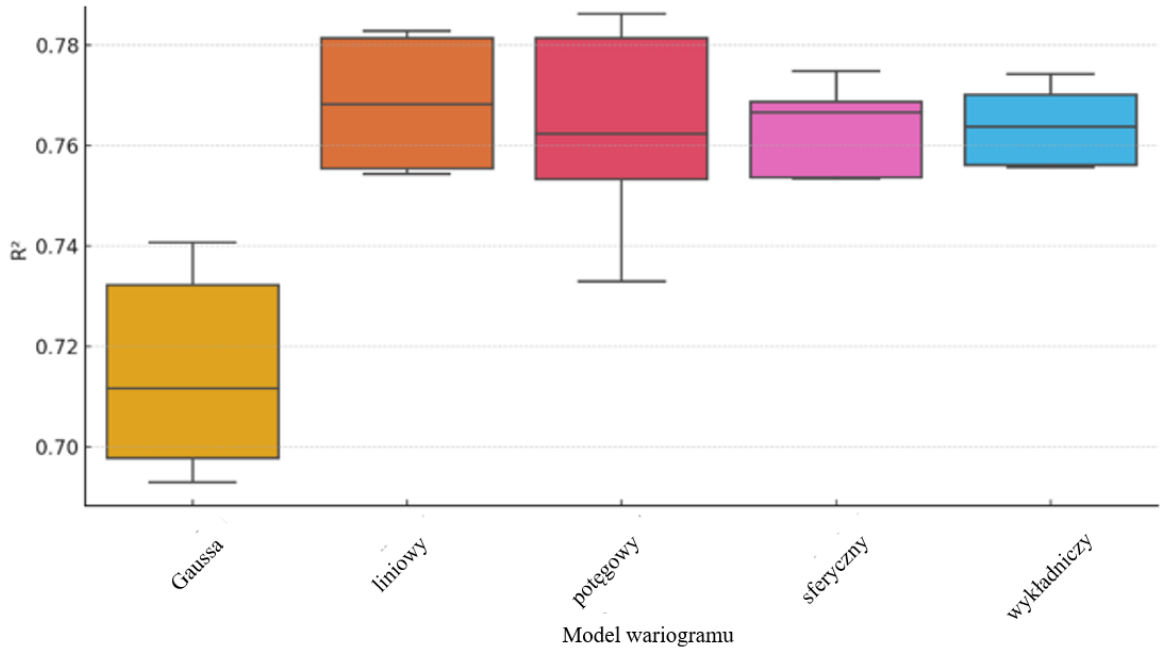
Różnice pomiędzy predykcją, a wynikiem odniesienia są niewielkie, co potwierdza przydatność sieci SSN jako przybliżenia funkcji celu w procesie optymalizacji, przy czym dokładność zależy od jakości i liczebności bazy, na której metamodel został wytrenowany (jak wykazano we wcześniejszym badaniu). Przykładowy przebieg zmiany wartości funkcji celu w funkcji numeru iteracji podczas optymalizacji przedstawia (rysunek 4.13).



Rysunek 4.13. Przykład przebiegu procesu optymalizacji, w którym wykorzystano tylko dane otrzymane z metamodelu opartego na bazie *LHS5000*.

4.3.2. Dobór parametrów Kriging

Budowa metamodelu Kriginga (2.2.3) wymaga doboru zestawu parametrów opisujących strukturę korelacji przestrzennej danych, ponieważ to one wprost determinują kształt funkcji autokowariancji i w konsekwencji - jakość predykcji. Z tego względu w dalszej części rozdziału skoncentrowano się na strojeniu hiperparametrów związanych z wariogramem, tj. na wyborze jego postaci oraz na parametrach dopasowania. Za hiperparametry kluczowe przyjęto wybór modelu wariogramu, wśród wariantów: liniowego, potęgowego, Gaussa, sferycznego oraz wykładniczego, a także parametry regulujące jego dopasowanie do danych. Dla każdej kombinacji modelu wariogramu i przyjętego schematu podziału zbioru danych przeprowadzono pięć niezależnych uruchomień uczenia, aby ograniczyć wpływ losowości i uzyskać bardziej stabilną ocenę jakości. Strojenie hiperparametrów wykonano z użyciem biblioteki Optuna w trybie GridSampler, ponieważ rozważana przestrzeń poszukiwań była dyskretna i ograniczona liczbowo, co pozwalało na pełne, deterministyczne przetestowanie wszystkich z góry zdefiniowanych konfiguracji. Strojenie hiperparametrów przeprowadzono w oparciu o Funkcję celu FF2 (3.5), wyniki badania przedstawiono na rysunek 4.14.



Rysunek 4.14. Dobór modelu wariogramu, zależność R^2od funkcji wariogramu.

Na podstawie średnich wartości R^2, wybrano wariogram liniowy zapewniający najwyższą jakość dopasowania, zarówno dla modeli *Full*, jak i *NoDamage*. Tabela 4.5 pokazuje, że wartości R^2 dla metamodelu Kriginga z funkcją celu FF2 (3.5) są zbliżone do wyników uzyskanych dla sieci SSN i mieszczą się w zakresie 0,67-0,78 dla baz LHS oraz 0,74-0,86 dla baz QEA, przy przewidywalnym spadku jakości wraz ze zmniejszeniem rozmiaru bazy.

Tabela 4.5. Wyniki sprawdzania mniejszych baz danych dla metamodelu Kriginga.

Metamodel Kriginga stworzony na funkcji celu FF2 bazujący na bazie LHS						
Baza LHS	Współ. determinacji R^2		MSE_{val}		MSE_{test}	
	Baza *Full*	Baza *NoDamage*	Baza *Full*	Baza *NoDamage*	Baza *Full*	Baza *NoDamage*
LHS5000	0.78	0.75	0.21	0.03	0.33	0.04
LHS2500	0.75	0.76	0.23	0.03	0.35	0.04
LHS1250	0.76	0.77	0.24	0.03	0.34	0.04
LHS625	0.76	0.53	0.24	0.04	0.35	0.06
LHS312	0.73	0.61	0.25	0.04	0.37	0.05
LHS156	0.66	0.22	0.30	0.05	0.40	0.07
Metamodel Kriginga stworzony na funkcji celu FF2 bazujący na bazie QEA						
	Współ. determinacji R^2		MSE_{val}		MSE_{test}	

Baza QEA	Baza *Full*	Baza *NoDamage*	Baza *Full*	Baza *NoDamage*	Baza *Full*	Baza *NoDamage*
QEA1687	0.77	0.85	0.20	0.02	0.32	0.03
QEA625	0.76	0.81	0.21	0.03	0.33	0.04
QEA312	0.73	0.77	0.24	0.03	0.35	0.04
QEA156	0.67	0.65	0.29	0.04	0.42	0.06

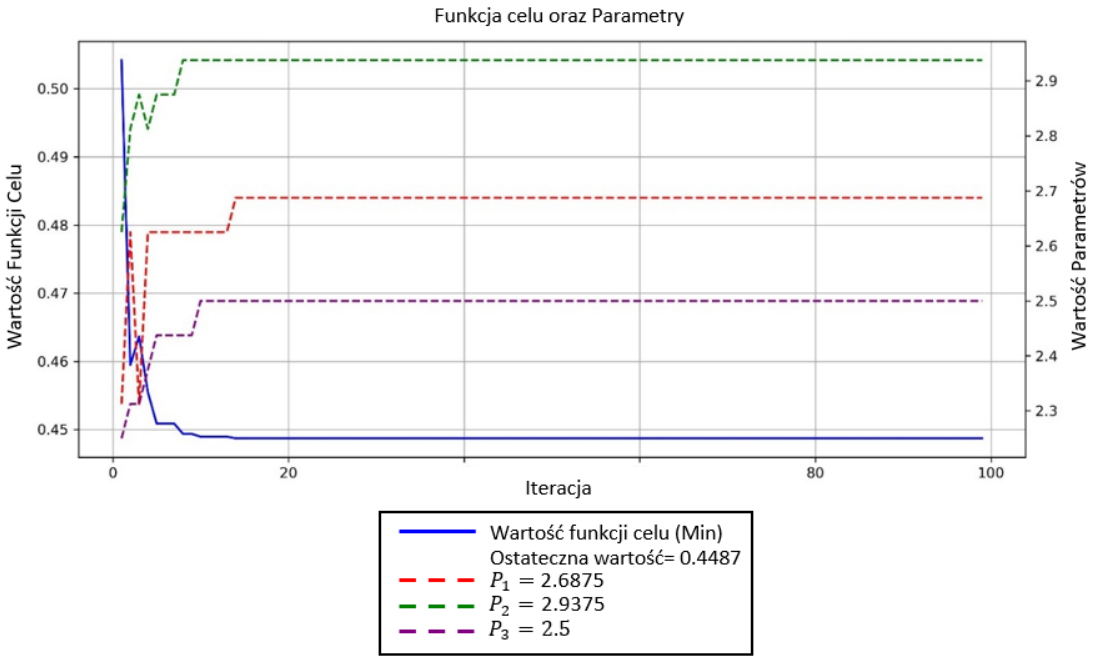
Następnie metamodel Kriginga z funkcją FF2 (3.5) wykorzystano do przeprowadzenia predykcji w zadaniach QEA jako jedyne narzędzie ewaluacji wartości funkcji celu, bez wykonywania dodatkowych obliczeń MES. W niezależnych biegach optymalizacji zastosowano cztery warianty modeli, trenowane na bazach LHS5000 oraz QEA1687, osobno dla przypadków *Full* i *NoDamage*. Zestawienie wyników przedstawia Tabela 4.6.

Tabela 4.6. Porównanie przewidywań Kriginga budowanych na tych samych bazach danych, na różnych treningach i różnych próbach predykcji.

Optymalizacja QEA dla FF2 - bazująca na *LHS5000*			
Model [R^2]	Wartość funkcji celu	Wektor zmiennych projektowych	Wartość funkcji celu MES
Full -0.7954 *NoDamage:* 0.8034	0.43023	2.6875,2.9375,2.5	0.42717
Full -0.7795 *NoDamage* -0. 7269	0.428881	2.8125,2.9375,2.4375	0.42132
Full -0.7886 *NoDamage:* 0.7942	0.423571	2.625,2.9375,2.5	0.42642
Full -0.7885 *NoDamage:* 0.78	0.43848	2.8125,2.9375,2.4375	0.42132
Optymalizacja QEA dla FF2 - bazująca na *LHS5000 QEA1687*			
*Full:*0.7993	0.427507	2.6875,2.9375,2.5	0.42717

NoDamage: 0.8657			
*Full:*0.8023 *NoDamage:* 0.870	0.425319	2.6875,2.9375,2.5	0.42717
*Full:*0.7857 *NoDamage:* 0.8735	0.422793	2.75,2.9375,2.5	0.43402
*Full:*0.7957 *NoDamage:* 0.8643	0.420768	2.75,2.9375,2.5	0.43402

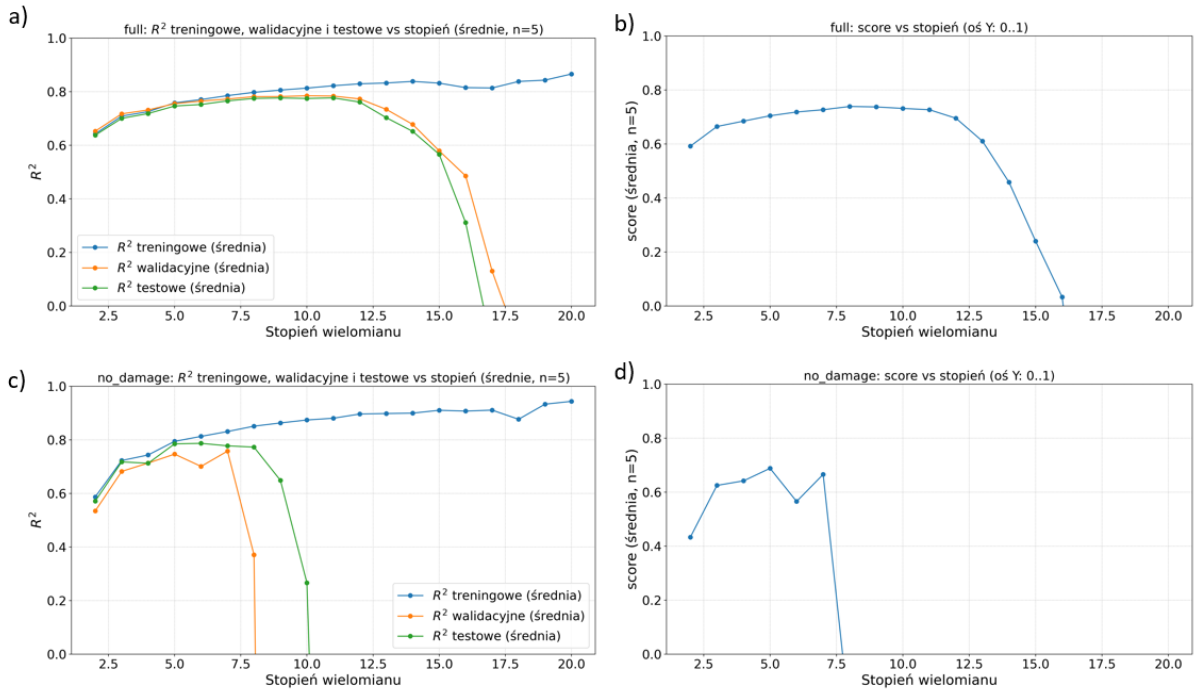
Badanie wykazało, że wartości funkcji celu wyznaczone za pomocą metamodelu Kriginga były bardzo zbliżone do wyników MES dla najlepszych osobników w danej próbie. Potwierdziło to, że metamodel Kriginga może pełnić rolę efektywnego metamodelu w procesie optymalizacji, przy czym, analogicznie jak w przypadku SSN, jakość predykcji zależy od rozmiaru i struktury bazy danych. Przykład przebiegu optymalizacji przedstawia rysunek 4.15.



Rysunek 4.15. Zmienność funkcji celu w funkcji iteracji dla przykładowego przebiegu procesu optymalizacji z użyciem metamodelu Kriging bez obliczania funkcji celu używając wyników MES.

4.3.3. Dobór parametrów regresji wielomianowej

W przypadku regresji wielomianowej najważniejszym hiperparametrem jest stopień wielomianu. Dla każdej bazy danych i dla obu modeli (*Full* oraz *NoDamage*) przeprowadzono serię 5 obliczeń dla każdego stopnia wielomianu. W tym badaniu testom poddano funkcje celu FF2 (3.5) dla bazy LHS5000. Czas tworzenia modeli regresji wielomianowej (ang. Polynomial Regression - PNR) jest stosunkowo krótki, co umożliwiło użycie przeszukiwania siatkowego - przestrzeń poszukiwań ograniczono do zakresu wystarczającego na otrzymanie pełnego przekroju wyników. Rysunek 4.16 przedstawia zależność jakości dopasowania od stopnia wielomianu: na wykresach a) i c) pokazano średnie wartości R^2 na zbiorach treningowym, walidacyjnym i testowym, wyznaczone w $n = 5$ niezależnych powtórzeniach (różne losowe podziały danych). Wykresy b) i d) prezentują przebieg autorskiej miary rankingowej R^2_{PNR}, zdefiniowanej wcześniej równaniem (rozdział 2.2.5, wzór 2.70). Miara ta agreguje trzy wartości współczynnika determinacji: R^2_{train}, R^2_{val} oraz R^2_{test}, a jednocześnie nakłada karę za rozbieżności między nimi. Dzięki temu preferowane są takie konfiguracje modelu PNR, które nie tylko uzyskują wysokie wartości R^2, lecz także zachowują zbliżoną jakość predykcji na zbiorze treningowym, walidacyjnym i testowym. W tym ujęciu stabilna generalizacja oznacza brak wyraźnego pogorszenia jakości predykcji po przejściu ze zbioru treningowego na dane niewykorzystywane bezpośrednio do dopasowania modelu.



Rysunek 4.16. Wpływ stopnia wielomianu na jakość regresji wielomianowej dla funkcji celu FF2 (3.5): a) R^2 treningowe, walidacyjne i testowe w funkcji stopnia dla wariantu *Full* (średnia

z $n = 5$); b) wartość miary R^2_{PNR} w funkcji stopnia dla wariantu *Full* (średnia z $n = 5$); c) jak w a) dla wariantu *NoDamage*; d) jak w b) dla wariantu *NoDamage*.

Zaobserwowano, że wraz ze wzrostem stopnia wielomianu wartość współczynnika determinacji R^2 na zbiorze treningowym rośnie niemal monotonicznie, natomiast na zbiorach walidacyjnym i testowym poprawa występuje jedynie do pewnego zakresu stopni, po czym następuje wyraźna degradacja jakości predykcji, co wskazuje na przeuczenie modelu o zbyt dużej złożoności. W wariancie *Full* efekt ten ujawnia się jako rozbieżność pomiędzy rosnącym R^2_{train} a spadającymi R^2_{val} i R^2_{test} dla wyższych stopni wielomianu, przy czym w skrajnych przypadkach R^2_{val} może przyjmować wartości ujemne, co oznacza predykcję gorszą niż model odniesienia oparty o średnią. Jednocześnie przebieg miary R^2_{PNR} potwierdza, że wybór stopnia odpowiada obszarowi kompromisu, w którym średni poziom dopasowania jest wysoki, a rozbieżności pomiędzy R^2_{train}, R^2_{val} i R^2_{test} pozostają ograniczone, natomiast w obszarze przeuczenia R^2_{PNR} gwałtownie spada z uwagi na narastające różnice pomiędzy zbiorami. W wariancie *NoDamage* pogorszenie jakości walidacji i teście pojawia się wcześniej niż w wariancie *Full*, co wskazuje na węższy, bardziej restrykcyjny zakres stopni zapewniających stabilną generalizację i uzasadnia dobór hiperparametru na podstawie kryterium łączącego poziom dopasowania z kontrolą zgodności wyników pomiędzy zbiorami.

Przeprowadzono również badanie wpływu wielkości zbiorów. Tabele 4.7 i 4.8 pokazują, że stopień wielomianu optymalny z punktu widzenia R^2 zależy silnie od rozmiaru bazy. Dla największych baz LHS5000 i LHS2500 uzyskano najlepsze wyniki przy stopniach odpowiednio 8-9 dla modelu *Full* oraz 5-6 dla modelu *NoDamage*. W miarę redukcji liczby próbek optymalny stopień wielomianu obniża się, osiągając wartości rzędu 2-3 dla najmniejszych baz *LHS156* czy *QEA156*. Wynika to z faktu, że wysokostopniowe wielomiany są bardzo podatne na przeuczenie przy ograniczonej liczbie danych.

Tabela 4.7. Porównanie przewidywań metamodelu Regresji Wielomianowej budowanych na tych samych bazach danych, na różnych treningach i różnych próbach predykcji.

PNR stworzony na funkcji celu FF2 bazujący na bazie LHS						
Baza LHS	R^2 trening		R^2 walidacja		R^2 test	
	Baza *Full*	Baza *NoDamage*	Baza *Full*	Baza *NoDamage*	Baza *Full*	Baza *NoDamage*
LHS5000	0.79	0.79	0.78	0.79	0.77	0.78
LHS2500	0.80	0.80	0.76	0.83	0.77	0.83
LHS1250	0.77	0.75	0.77	0.76	0.76	0.77
LHS625	0.73	0.73	0.72	0.73	0.72	0.74

LHS312	0.74	0.62	0.73	0.58	0.75	0.53
LHS156	0.76	0.79	0.77	0.58	0.79	0.49
PNR stworzony na funkcji celu FF2 bazujący na bazie QEA						
Baza QEA	R^2 trening		R^2 walidacja		R^2 test	
	Baza *Full*	Baza *NoDamage*	Baza *Full*	Baza *NoDamage*	Baza *Full*	Baza *NoDamage*
QEA1687	0.77	0.86	0.79	0.86	0.74	0.87
QEA625	0.75	0.84	0.75	0.80	0.75	0.82
QEA312	0.76	0.87	0.77	0.86	0.73	0.86
QEA156	0.66	0.90	0.55	0.88	0.61	0.93

Tabela 4.8. Stopień wielomianu dla najlepszego modelu dla danej wielkości bazy a) dla modelu stworzonego na bazie LHS b) dla modelu stworzonego na bazie QEA

a)

Baza danych LHS	Stopień Wielomianu	
	Baza *Full*	Baza *NoDamage*
LHS5000	8	5
LHS2500	9	6
LHS1250	6	3
LHS 625	3	3
LHS 312	3	2
LHS 156	3	2

b)

Baza danych QEA	Stopień Wielomianu	
	Baza *Full*	Baza *NoDamage*
QEA1687	6	4
QEA625	3	3
QEA312	4	3
QEA156	2	3

Ostatecznie przetestowano metamodel PNR w zadaniu optymalizacji QEA, analogicznie jak w przypadku SSN i metamodelu Kriginga. Tabela 4.9 przedstawia przykładowe wyniki dla modeli trenowanych na bazach LHS5000 i QEA1687 z funkcją celu FF2 (3.5). Średnio wyniki PNR są porównywalne z rezultatami innych metamodeli - wartości funkcji celu przewidywane przez PNR dla osobników najlepszych w danym biegu są zbliżone do wartości uzyskanych z MES. Problemem okazała się jednak powtarzalność: wśród pięciu przeprowadzonych prób dla konfiguracji opartych na bazie QEA1687 dwie zakończyły się wyraźną porażką metamodelu (skrajnie błędne wartości funkcji celu, w tym wartości ujemne lub nieadekwatnie duże), pomimo pozornie poprawnych wskaźników jakości na etapie walidacji. Z tego względu regresja wielomianowa została oceniona jako metoda potencjalnie użyteczna, lecz wymagająca dodatkowych mechanizmów kontroli jakości (np. powtarzanie

optymalizacji z różnymi inicjalizacjami i odrzucanie rozwiązań nieprawidłowych) lub stosowania w połączeniu z innymi metamodelami.

Tabela 4.9. Porównanie przewidywań Regresji Wielomianowej budowanej na tych samych bazach danych, na różnych treningach i różnych próbach predykcji.

Optymalizacja QEA dla FF2 - bazująca na *LHS5000*			
Model [R^2]	Wartość funkcji celu	Wektor zmiennych projektowych	Wartość funkcji celu MES
*Full:*0.77 *NoDamage:* 0.789	0.4466	2.5625,2.9375,2.625	0.4858
*Full:*0.79 *NoDamage:* 0.8	0.4377	2.6875,2.9375,2.5	0.427044
*Full:*0.79 *NoDamage:* 0.8	0.4282	2.6875,2.9375,2.5625	0.43345
*Full:*0.78 *NoDamage:* 0.81	0.4456	2.625,2.9375,2.5625	0.4306
Optymalizacja QEA dla FF2 - bazująca na *QEA1687*			
*Full:*0.77 *NoDamage:* 0.84	-0.36	1,1,2.9375	0.41940
*Full:*0.76 *NoDamage:* 0.88	0.4233	2.6875,2.9375,2.5625	0.43345
*Full:*0.78 *NoDamage:* 0.84	0.4245	2.6875,2.9375,2.5625	0.43345

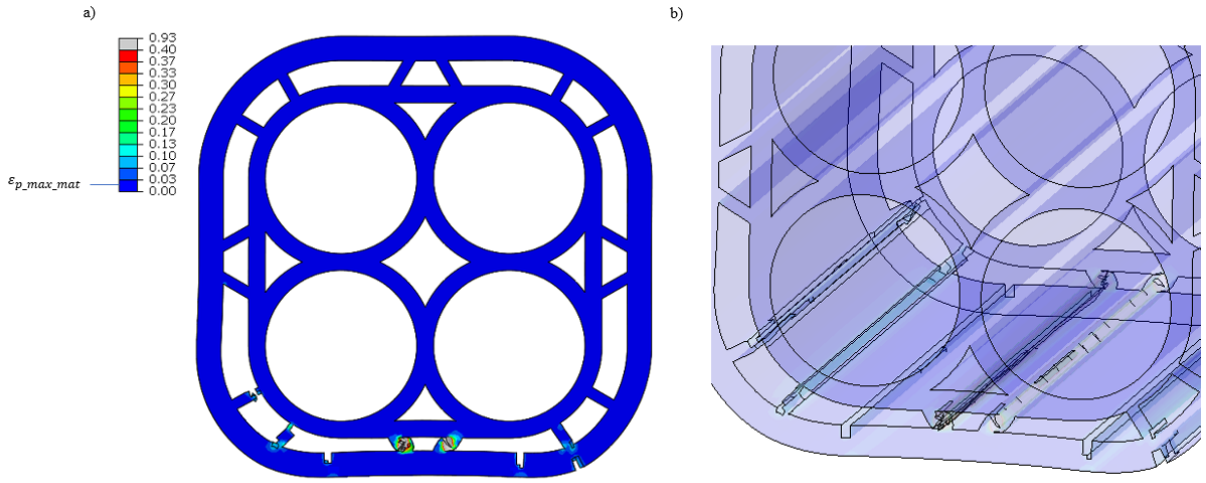
*Full:*0.73 *NoDamage:* 0.87	0.4191	2.625,2.9375,2.5625	0.4306
*Full:*0.76 *NoDamage:* 0.86	- 11.375	2.9375,2.9375,1	325.659

Na podstawie przeprowadzonych badań porównawczych należy podkreślić, że w zasadzie każdy z rozważanych metamodeli (SSN, Kriging, PNR) mógłby zostać zdefiniowany w dwóch wariantach, tj. dla bazy *Full* oraz *NoDamage*. Analizy numeryczne wykazały, że SSN zapewnia najwyższą i zarazem najbardziej powtarzalną jakość aproksymacji zarówno dla baz LHS, jak i QEA, przy niewielkiej różnicy pomiędzy błędem uczącym i walidacyjnym oraz niskim koszcie pojedynczej ewaluacji. Jednocześnie metamodel typu Kriging, osiągał jakości predykcji zbliżone do SSN i w wielu przypadkach generował wyniki porównywalne pod względem wartości współczynnika R^2, co potwierdza jego przydatność w rozpatrywanym problemie. Regresja wielomianowa (PNR) okazała się natomiast mniej stabilna w ujęciu praktycznym, ponieważ była silnie wrażliwa na wybór stopnia wielomianu i rozmiar bazy, co prowadziło do przeuczenia oraz lokalnie niestabilnych predykcji w trakcie optymalizacji, pomimo pozornie poprawnych wskaźników walidacyjnych. W konsekwencji, przy braku istotnych różnic jakościowych pomiędzy SSN a Krigingiem oraz uwzględniając słabszą niezawodność PNR, w dalszej części pracy przyjęto do analiz wyłącznie SSN, przy czym Kriging należy uznać za rozwiązanie równorzędnie użyteczne i możliwe do zastosowania w analizowanym zadaniu.

4.4. Optymalizacja obudowy baterii z użyciem QEA

Celem tego rozdziału jest przedstawienie wyników optymalizacji z użyciem QEA dla trzech wariantów, bez wspomagania metamodelem, z użyciem metamodelowania w podejściu HYBRYDA1, następnie HYBRYDA2 (rozdział 3.6).

W pierwszym kroku przeanalizowano wyniki referencyjne, tj. uzyskane bez zastosowania procedury optymalizacji, wśród których zidentyfikowano przypadki prowadzące do uszkodzeń obudowy. Przykładowy osobnik z parametrami $P_1 = 2.89553$, $P_2 = 2.99335$, $P_3 = 1.25013$ wykazuje istotne zniszczenia (rysunek 4.17).



Rysunek 4.17. Mapa odkształceń plastycznych dla modelu z konfiguracją, w której doszło do zniszczenia. a) widok z przodu b) widok izometryczny (z przeźroczystością).

Zniszczeniom uległy żebra w wyniku czego warstwa obudowy zewnętrznej uległa pęknięciu. W warunkach projektowych wyniki takiej symulacji wskazywałyby na wysokie ryzyko niepowodzenia testu eksperymentalnego.

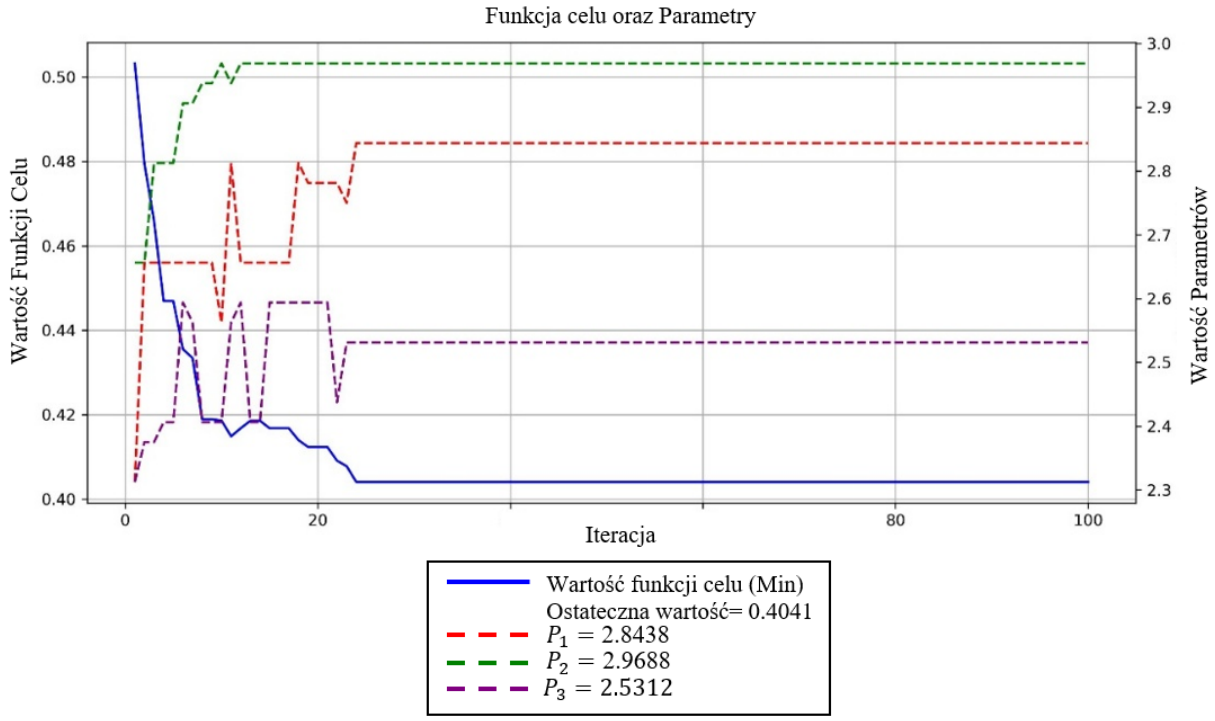
4.4.1. Optymalizacja obudowy z użyciem QEA bez metamodelowania

W tym podrozdziale zostaną przedstawione wyniki dla optymalizacji QEA z użyciem tylko MES bez wspomagania metamodelem. Dzięki temu możliwe będzie porównanie korzyści czasowej osiągniętej poprzez zastosowanie metamodelowania. Poniższa tabela reprezentuje zestawienie zadań QEA dla FF1 (3.4) oraz FF2 (3.5) uruchomionej czterokrotnie, po czym w każdym następnym uruchomieniu baza jest uaktualniana o wyniki z poprzedniego przebiegu optymalizacji. Kolorem czerwonym zaznaczono najniższą wartość funkcji celu. Ilość iteracji algorytmu optymalizacyjnego została ustawiona na 100 iteracji. Rozdzielczość przeszukiwania ustalono na 6 kubitów. W jednej iteracji algorytmu QEA znajduje się 30 osobników.

Tabela 4.10. Zestawienie wyników optymalizacji dla podejścia bez metamodelowania dla funkcji celu FF1 i FF2.

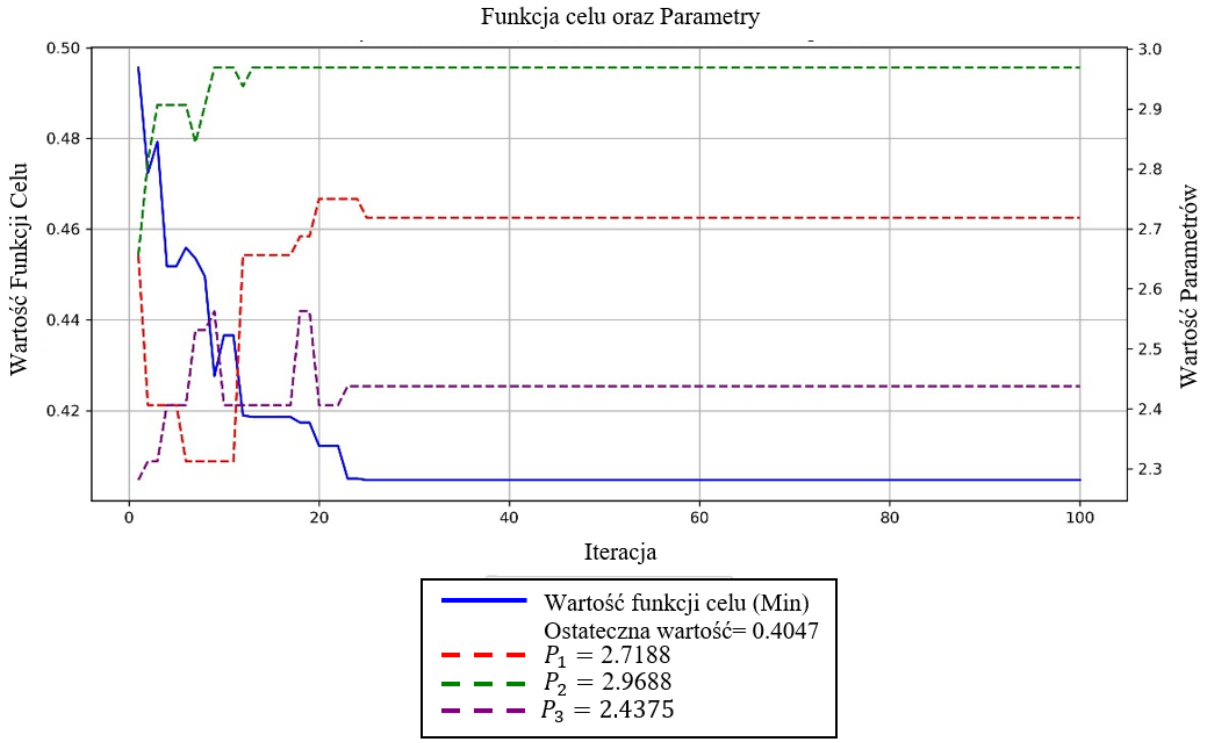
QEA FF1				
Próba	Wartość funkcji celu	Wektor zmiennych projektowych	Początkowa liczba danych z poprzednich uruchomień	Czas obliczeń
1	0.407254	2.875 2.96875 2.53125	0	38 h
2	0.404056	2.84375 2.96875 2.53125	409	41 h
3	0.407254	2.875 2.96875 2.53125	1036	31 h
4	0.404659	2.71875 2.96875 2.4375	1079	32 h
QEA FF2				
1	0.404975	2.75 2.96875 2.4375	0	40h
2	0.405483	2.90625 2.96875 2.4375	495	35 h
3	0.404659	2.71875 2.96875 2.4375	934	26 h
4	0.404659	2.71875 2.96875 2.4375	1294	32 h

Zestawienie wyników (Tabela 4.10) wskazuje, że zarówno funkcja celu FF1 (3.4), jak i FF2 (3.5) prowadzą do rozwiązań o bardzo zbliżonej jakości, ponieważ uzyskane minima mieszczą się w wąskim przedziale około 0.404 do 0.407. Dodatkowo, w obu przypadkach obserwowana jest powtarzalność najlepszych osobników w tej samej siatce dyskretnej, co sugeruje, że procedura optymalizacji konsekwentnie identyfikuje ten sam obszar przestrzeni projektowej jako korzystny w przyjętej rozdzielczości przeszukiwania. W szczególności stała wartość parametru $P2 = 2.96875$ we wszystkich próbach, niezależnie od rodzaju funkcji celu i wielkości bazy początkowej, potwierdza stabilizację rozwiązania względem tego wymiaru. W konsekwencji FF1 (3.4) i FF2 (3.5) można traktować jako równoważnie skuteczne w kontekście uzyskiwanej wartości funkcji celu i powtarzalności identyfikowanych konfiguracji.



Rysunek 4.18. Wartość funkcji celu w funkcji iteracji oraz zmiana parametrów testu QEA bez metamodelowania z funkcją celu FF1 dla najlepszego testu.

Rysunek 4.18 przedstawia przebieg minimalizowanej funkcji celu FF1 (3.4) (linia niebieska) oraz zmianę parametrów $P1$, $P2$, $P3$ (linie przerywane) w kolejnych iteracjach procesu optymalizacji. Zaobserwowano szybki spadek wartości funkcji celu w początkowej fazie obliczeń, z poziomu około 0.50 do około 0.42, a następnie wolniejszą poprawę o charakterze schodkowym aż do osiągnięcia wartości końcowej 0.4041. Po około 25 iteracji nie odnotowano dalszej poprawy, co wskazuje na stabilizację najlepszego znalezionego rozwiązania w przyjętej liczbie iteracji. Równocześnie parametry stopniowo przechodzą do wartości stałych: $P_1 = 2.8438$, $P_2 = 2.9688$, $P_3 = 2.5312$, przy czym ich stabilizacja czasowo pokrywa się z wypłaszczeniem funkcji celu, co sugeruje, że dalsze iteracje nie prowadzą do eksploracji nowych, korzystniejszych konfiguracji w rozpatrywanej przestrzeni parametrów.



Rysunek 4.19. Wartość funkcji celu w funkcji iteracji oraz zmiana parametrów testu QEA bez metamodelowania z funkcją celu FF2 dla najlepszego testu.

Podobny przebieg można zaobserwować dla funkcji celu FF2 (3.5) (rysunek 4.19).

4.4.2. Optymalizacja z użyciem metody HYBRYDA1

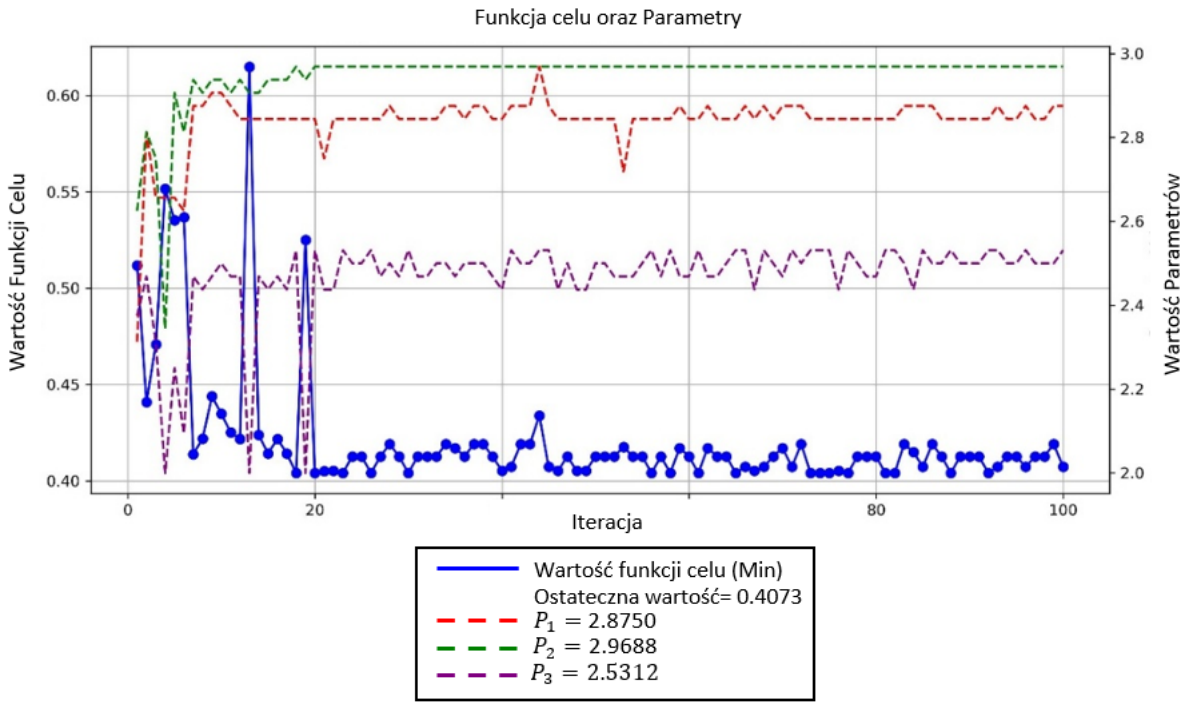
Podczas badania wpływu wielkości baz w rozdziale 4.3 ustalono, że minimalną liczebnością bazy danych, od której włączenie metamodelowania do optymalizacji przynosi zauważalny efekt, jest około 300 osobników. Poniżej zestawiono wyniki optymalizacji z użyciem podejścia HYBRYDA 1, porównując uzyskane wartości funkcji celu oraz czasy obliczeń. W dalszych testach zastosowano funkcję celu FF2 (3.5), przy czym wybór ten nie wynika z ograniczeń jakościowych FF1 (3.4). Przeciwnie, wcześniejsze przebiegi QEA wykazały, że FF1 (3.4) i FF2 (3.5) prowadzą do rozwiązań o porównywalnej jakości, a uzyskiwane minima oraz identyfikowane konfiguracje parametrów są zbliżone w granicach przyjętej rozdzielczości przeszukiwania. W związku z brakiem istotnych różnic w wynikach końcowych, dalszą analizę podejścia HYBRYDA1 przeprowadzono konsekwentnie dla FF2 (3.5), traktując FF1 (3.4) jako wariant równoważny pod względem osiąganej wartości funkcji celu.

Tabela 4.11. Zestawienie wyników optymalizacji w wykorzystaniem podejścia HYBRYDA1 dla FF2 (3.5).

HYBRYDA 1 z funkcją celu				
Próba	Wartość funkcji celu	Wektor zmiennych projektowych	Początkowa liczba danych z poprzednich uruchomień	Czas obliczeń
1	0.404056	2.84375 2.96875 2.53125	0	29h
2	0.404975	2.75 2.96875 2.4375	345	3h 30 min
3	0.404056	2.84375 2.96875 2.53125	360	3h 40 min
4	0.404056	2.84375 2.96875 2.53125	377	3h 15 min

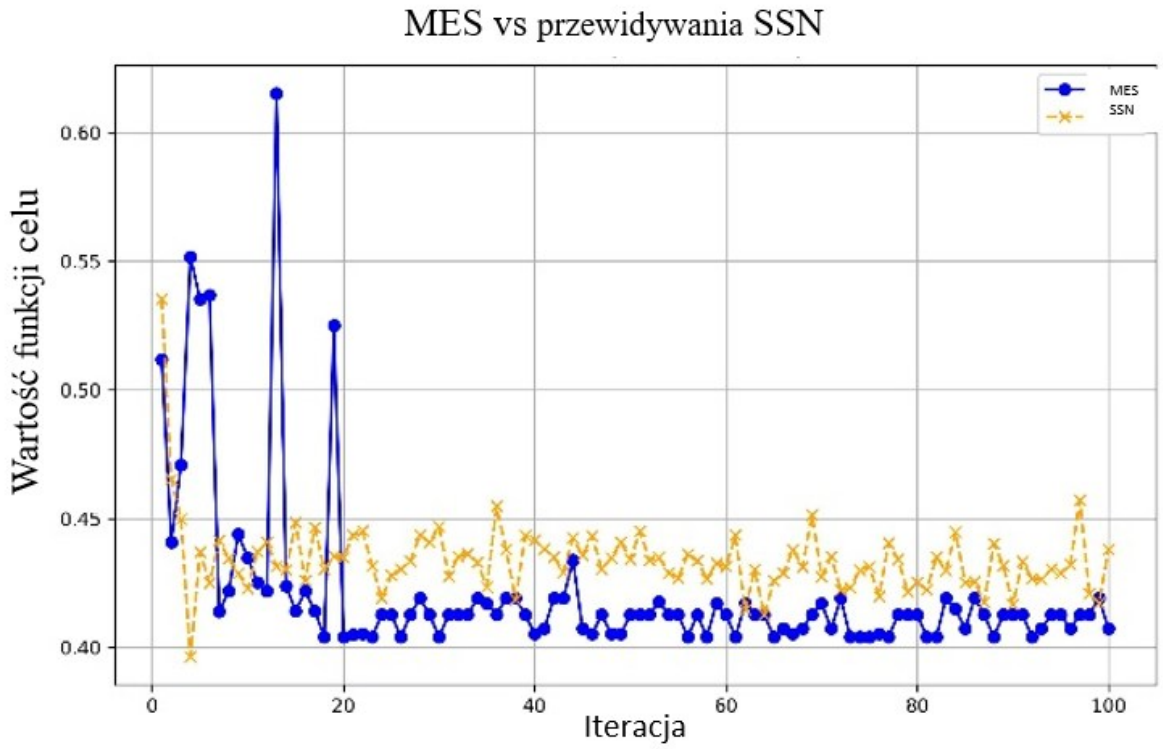
Zestawienie dla funkcji celu FF2 (3.5) wskazuje, że w podejściu HYBRYDA1 uzyskano minimum równe 0.404056, a więc poziom porównywalny z najlepszymi rezultatami wariantu uzyskanymi w poprzednim rozdziale (4.4.1), przy jednoczesnym istotnym skróceniu czasu pojedynczego uruchomienia testu. W pierwszej próbie, realizowanej bez bazy początkowej, czas wyniósł 29 h, natomiast po przekroczeniu progu liczebności bazy rzędu kilkuset próbek (345 do 377) czas kolejnych uruchomień spadł do około 3h 30 min, co potwierdza praktyczną skuteczność metamodelowania po osiągnięciu minimalnej wielkości zbioru uczącego. Jednocześnie niewielkie różnice czasów pomiędzy próbami 2 do 4, mimo wzrostu liczby danych początkowych, sugerują, że całkowity koszt procesu jest wciąż w dużej mierze determinowany przez wywołania MES, jednak ich liczba została ograniczona na tyle, że uzyskano wyraźny zysk czasowy. Z punktu widzenia zastosowań inżynierskich zysk ten jest szczególnie istotny, ponieważ proces optymalizacji musi być uruchamiany wielokrotnie z różnymi populacjami, zarówno ze względu na stochastyczny charakter algorytmu, jak i ryzyko utknięcia w minimach lokalnych, a więc redukcja czasu pojedynczego przebiegu bezpośrednio przekłada się na redukcję łącznego czasu procesu optymalizacyjnego. Powtarzalność najlepszych rozwiązań w tych samych dyskretnych punktach przestrzeni decyzyjnej, na przykład $P_1 = 2.84375,\ P_2 = 2.96875,\ P_3 = 2.53125$, pozostaje spójna z obserwowaną kwantyzacją parametrów wynikającą z przyjętej rozdzielczości przeszukiwania.

W przebiegu całej próby optymalizacji zaobserwowano kilka charakterystycznych zjawisk (rysunek 4.20): po fazie początkowej, w której wartości funkcji celu osiągają wysoki rozrzut i występują incydentalne piki, następuje szybkie zejście do poziomu około 0.41 oraz dalsza praca wykazuje wyraźne fluktuacje wartości bieżących.



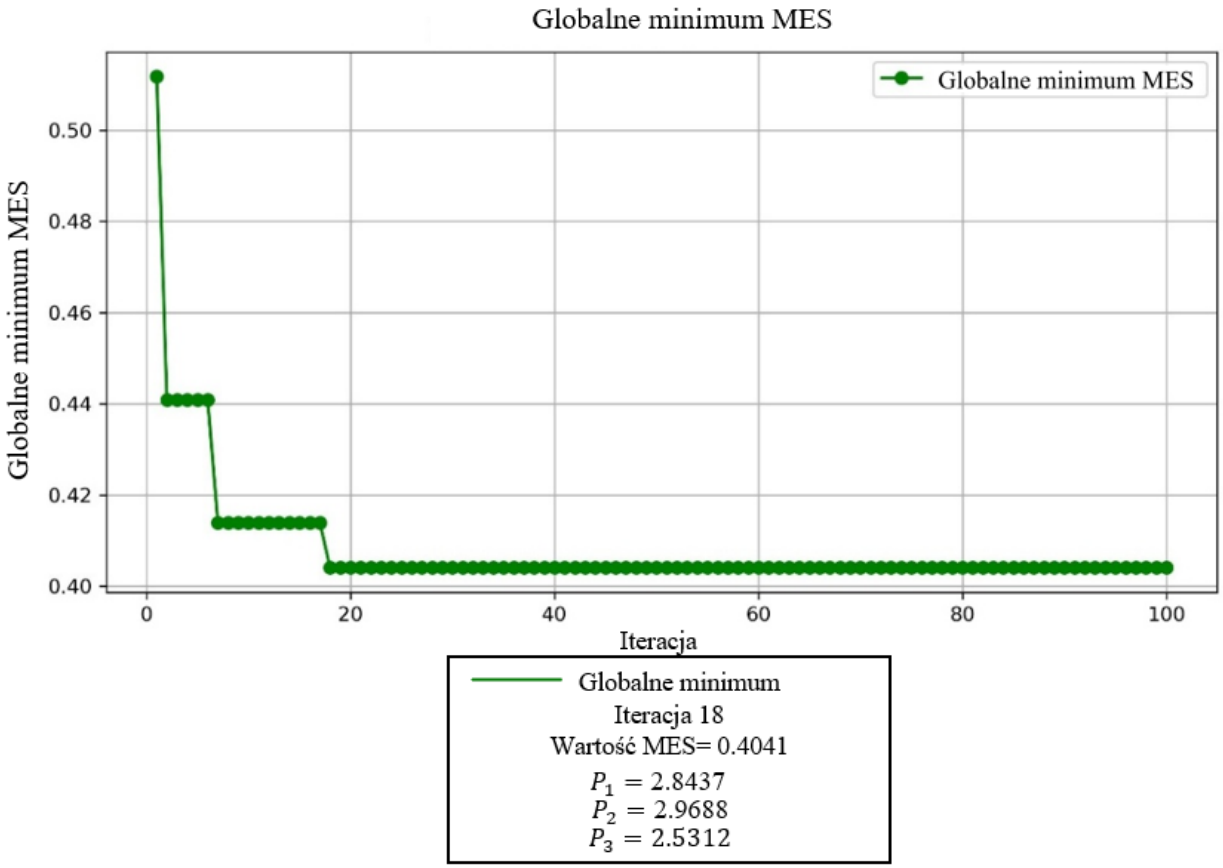
Rysunek 4.20. Wartość funkcji celu w funkcji iteracji oraz zmiana parametrów testu QEA HYBRYDA1 z funkcją celu FF2 dla najlepszego testu.

Fluktuacje te należy interpretować jako efekt trenowania SSN przy każdej iteracji od nowa, w związku z czym dla różnych zbiorów treningowych oraz walidacyjnych predykcja może dawać różne wyniki dla tych samych parametrów. Predykcja jest przeprowadzana na wszystkich 30stu osobnikach z puli i tylko jeden osobnik z najlepszą wartością funkcji celu jest poddawany ponownej ewaluacji funkcji celu na bazie wyników symulacji MES. Rysunek 4.21 przedstawia porównanie wartości funkcji celu otrzymanej z użyciem metamodelu oraz z użyciem wyników MES (dla każdej iteracji).



Rysunek 4.21. Przewidywania wartość funkcji celu SSN oraz funkcja celu na bazie MES w funkcji iteracji.

Efekt fluktuacji widoczny na wykresie wartości bieżących nie oznacza pogorszenia najlepszego znalezionego rozwiązania, lecz wynika ze zmienności predykcji metamodelu SSN oraz z oceny kolejnych populacji osobników. Z tego względu, w celu jednoznacznej oceny postępu optymalizacji, na rysunku 4.22 przedstawiono przebieg najmniejszej wartości funkcji celu znalezionej od początku procesu do danej iteracji. W każdej iteracji wykres pokazuje więc najlepszy wynik uzyskany we wszystkich wcześniejszych i bieżących obliczeniach. Dzięki temu chwilowe fluktuacje wartości bieżących nie zaburzają interpretacji procesu zbieżności.



Rysunek 4.22. Wizualizacja wykresu globalnego minimum

Pomimo fluktuacji najniższa wartość pokrywa się z wartością znalezioną w podejściu QEA z użyciem MES bez metamodelu.

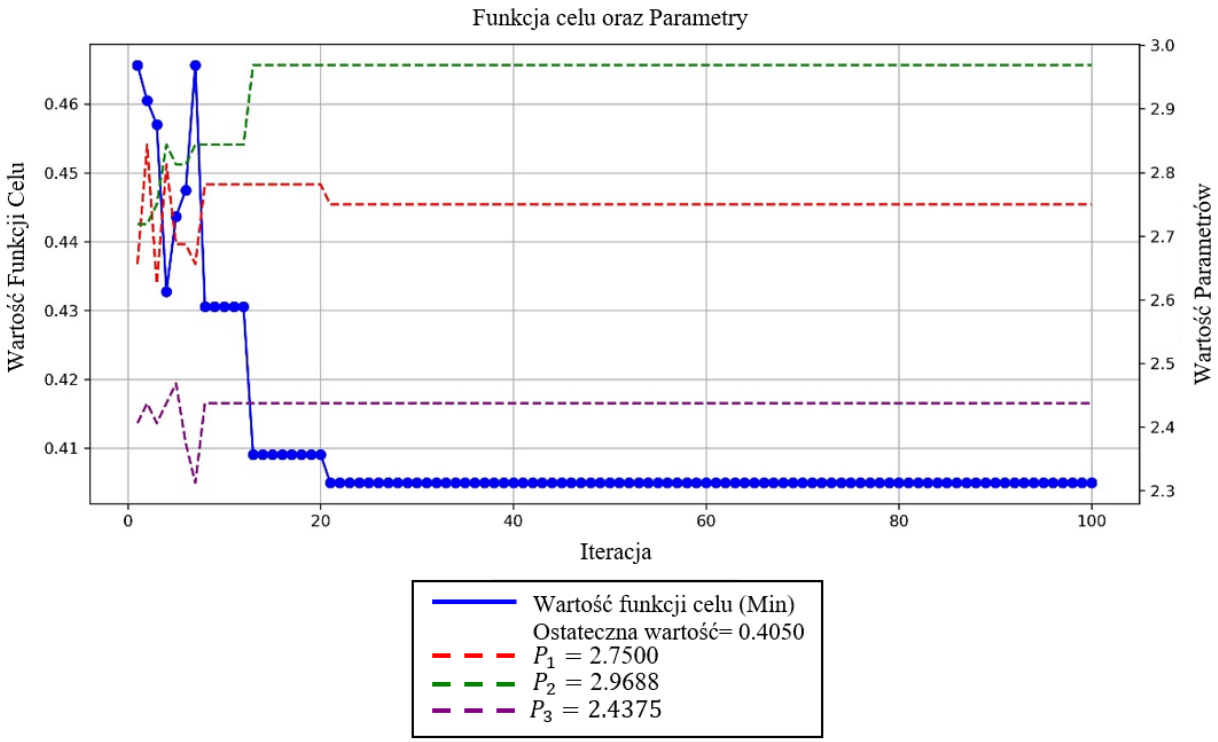
4.4.3. Optymalizacja z użyciem metody HYBRYDA2

Celem analizy podejścia HYBRYDA2 było sprawdzenie czy pominięcie części zbędnych obliczeń może skrócić czas działania algorytmu oraz poprawić spójność i stabilność uzyskiwanych wyników. Zestawienie rezultatów poszczególnych uruchomień przedstawia Tabela 4.12, obejmująca wartości funkcji celu FF2, wektory zmiennych projektowych najlepszego osobnika, liczbę dostępnych danych początkowych oraz całkowity czas obliczeń.

Tabela 4.12. Zestawienie wyników optymalizacji z wykorzystaniem podejścia HYBRYDA 2 dla FF2.

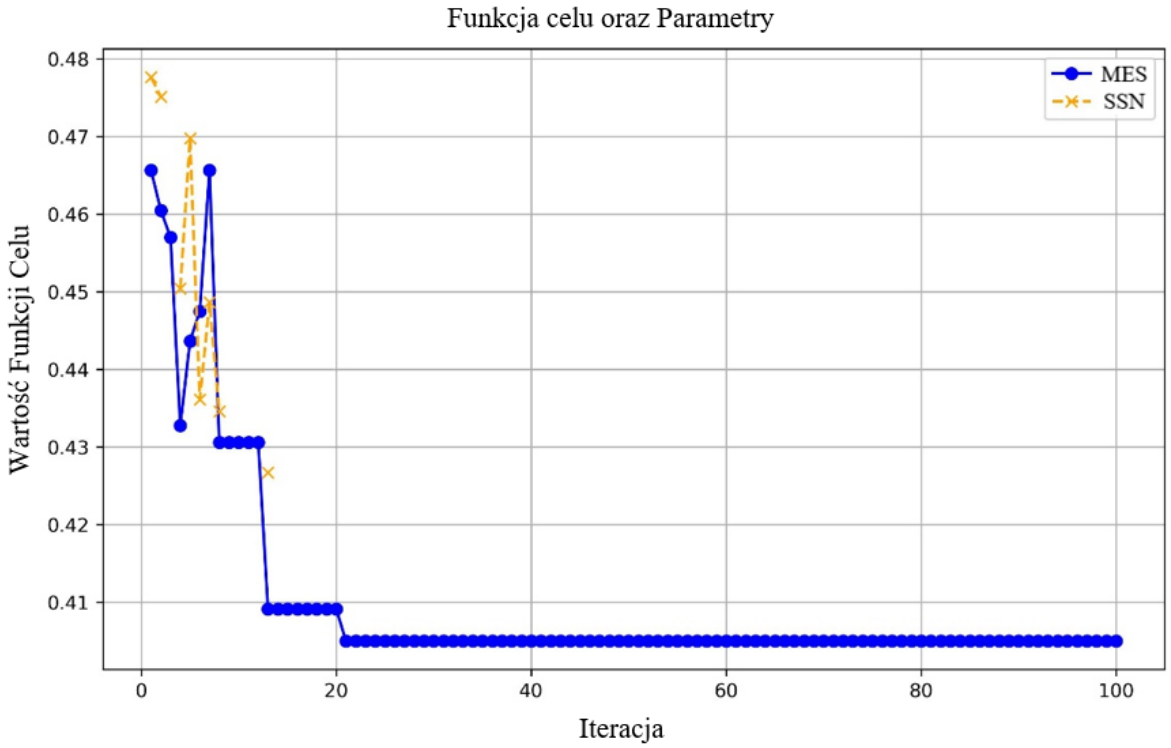
HYBRYDA2 z funkcją celu FF2				
Próba	Wartość funkcji celu	Wektor zmiennych projektowych	Początkowa liczba danych z poprzednich uruchomień	Czas obliczeń
1	0.404975	2.75 2.96875 2.4375	0	25h
2	0.442153	2.625 2.84375 2.34375	307	1h 5 min
3	0.404975	2.75 2.96875 2.4375	313	40 min
4	0.404975	2.75 2.96875 2.4375	316	1h 20 min

Zaobserwowano wyraźną poprawę powtarzalności przebiegu optymalizacji (rysunek 4.23). W trzech z czterech niezależnych uruchomień osiągnięto identyczną wartość minimum funkcji celu $FF2 = 0.404975$ oraz ten sam wektor zmiennych projektowych $P_1 = 2.7500$, $P_2 = 2.9688$, $P_3 = 2.4375$. Oznacza to, że procedura optymalizacji w większości uruchomień prowadziła do tego samego rozwiązania końcowego, pomimo losowego charakteru procesu uczenia metamodelu i kolejnych predykcji wykonywanych w trakcie optymalizacji. Rysunek 4.23 pokazuje, że po fazie wahań w początkowych iteracjach, algorytm szybko osiąga ustabilizowaną wartość funkcji celu, natomiast rysunek 4.24 potwierdza, że zarówno wartość minimalna, jak i odpowiadające jej parametry nie ulegają zmianie przez większą część przebiegu. Uzyskane minimum jest nieznacznie gorsze od najlepszego wyniku HYBRYDA1 ($FF2 = 0.404056$), ale różnica ta pozostaje pomijalna w kontekście uzyskanych oszczędności czasowych - czas obliczeń skrócono z 25 godzin (bez danych początkowych) do poniżej 1,5 godziny przy wykorzystaniu bazy uczącej zawierającej ok. 300 próbek. Odnotowano jednak pojedynczy przypadek (Próba 2), w którym algorytm zbiega do wyraźnie gorszego minimum lokalnego ($FF2 = 0.442153$), co uzasadnia konieczność wielokrotnych uruchomień przy zastosowaniu tej strategii.



Rysunek 4.23. Wartość funkcji celu w funkcji iteracji oraz zmiana parametrów testu QEA HYBRYDA2 z funkcją celu FF2 dla najlepszego testu.

Dodatkowo rysunek 4.24 przedstawia, że, w początkowej fazie przebiegu występują wyraźne rozbieżności pomiędzy wartościami przewidywanymi przez sieć neuronową (model zastępczy) a wynikami rzeczywistymi uzyskiwanymi z obliczeń metodą elementów skończonych. Różnice te są szczególnie widoczne w pierwszych kilkunastu iteracjach, co świadczy o niedostatecznym pokryciu przestrzeni projektowej w początkowym etapie. Z czasem, w miarę rozbudowy bazy uczącej, wartości predykcji oraz wartośći z MES zbliżają się do siebie, a najlepszy osobnik wyznaczany przez model zastępczy jest potwierdzany przez symulacje MES.



Rysunek 4.24. Wartość funkcji celu na bazie MES oraz wartość funkcji celu na bazie metamodelu w funkcji iteracji.

5. Wnioski

W celu końcowej oceny skuteczności opracowanych procedur zestawiono trzy warianty prowadzenia optymalizacji: klasyczne podejście, czyli MES połączony z QEA, podejście HYBRYDA1 oraz podejście HYBRYDA2. Porównanie wykonano dla funkcji celu FF2, ponieważ dla tej funkcji przeprowadzono zasadniczą część analiz metod hybrydowych. W zestawieniu uwzględniono nie tylko najlepszą uzyskaną wartość funkcji celu, lecz także całkowity czas serii prób. Takie ujęcie jest istotne, ponieważ w praktycznym zastosowaniu nowego zadania optymalizacyjnego proces nie rozpoczyna się od gotowej bazy danych, lecz od bazy pustej lub niewystarczającej do efektywnego uczenia metamodelu. Porównywanie wyłącznie czasów uzyskanych po wcześniejszym zgromadzeniu danych mogłoby prowadzić do zawyżenia rzeczywistej efektywności wariantów hybrydowych. Poniżej w tabeli 5.1 przedstawiono wyniki trzech omawianych metod optymalizacyjnych.

Tabela 5.1 Zestawienie wyników trzech metod optymalizacyjnych.

Wariant	Najlepsza uzyskana wartość funkcji celu	Wektor zmiennych projektowych dla najlepszego wyniku	Czas pierwszej próby	Łączny czas czterech prób
QEA+MES	0.404659	$P_1 = 2.71875,$ $P_2 = 2.96875,$ $P_3 = 2.4375$	40 h	133 h
HYBRYDA1	0.404056	$P_1 = 2.84375,$ $P_2 = 2.96875,$ $P_3 = 2.53125$	29 h	39 h 25 min
HYBRYDA2	0.404975	$P_1 = 2.75000,$ $P_2 = 2.96875,$ $P_3 = 2.4375$	25 h	28 h 5 min

Na podstawie zestawienia można stwierdzić, że wariant referencyjny łączący MES oraz QEA stanowi właściwy wariant odniesienia, ponieważ ocena funkcji celu opiera się bezpośrednio na obliczeniach MES. Jednocześnie jest to podejście najmniej efektywne czasowo. Łączny czas czterech prób wyniósł 133 h, co potwierdza, że pełne wykorzystanie MES w każdej istotnej ocenie funkcji celu prowadzi do wysokiego kosztu obliczeniowego. Podejście HYBRYDA1 istotnie ograniczyło łączny czas obliczeń, który wyniósł 39 h 25 min. Wariant ten zachowuje mechanizm kontroli jakości, ponieważ najlepszy osobnik wskazany przez metamodel jest następnie weryfikowany za pomocą MES. Z tego względu HYBRYDA1

można uznać za najbardziej zrównoważony wariant postępowania. Metoda ta łączy redukcję kosztu obliczeniowego z ograniczeniem ryzyka zaakceptowania rozwiązania wynikającego wyłącznie z błędu predykcji. Podejście HYBRYDA2 zapewniło największą redukcję czasu obliczeń. Łączny czas czterech prób wyniósł 28 h 5 min. W trzech z czterech prób uzyskano zbliżony wynik końcowy oraz ten sam wektor zmiennych projektowych, co wskazuje na dobrą powtarzalność w tych przebiegach. Jednocześnie jedna próba zakończyła się wyraźnie gorszym wynikiem, dlatego metoda ta wymaga dodatkowej kontroli jakości predykcji i ostrożnego stosowania przy małej lub nierównomiernie rozmieszczonej bazie danych. Ostatecznie za najbardziej bezpieczny wariant roboczy należy uznać HYBRYDA1. Podejście to zapewnia znaczące skrócenie czasu obliczeń względem wariantu referencyjnego łączącego QEA i MES, a jednocześnie utrzymuje bezpośrednią weryfikację najlepszego rozwiązania za pomocą MES. HYBRYDA2 jest wariantem najbardziej efektywnym czasowo, lecz jej stosowanie jest uzasadnione przede wszystkim wtedy, gdy dostępna baza danych ma wystarczającą liczebność i jakość, a proces predykcji metamodelu jest dodatkowo kontrolowany. Wariant referencyjny QEA z MES pozostaje natomiast punktem odniesienia do oceny metod hybrydowych, ale ze względu na całkowity czas obliczeń jest najmniej korzystny w zastosowaniach wymagających wielu powtórzeń procesu optymalizacji.

Na podstawie przeprowadzonych badań optymalizacji sformułowano następujące wnioski:

1. Dla analizowanego zadania optymalizacji wyniki końcowe uzyskiwane przy zastosowaniu FF1 (3.4) i FF2 (3.5) są jakościowo zbliżone, a wybór funkcji celu nie zmienia zasadniczo charakteru rozwiązań optymalnych w przyjętej rozdzielczości przeszukiwania.
2. Analizy metamodelowania wykazały silną zależność jakości predykcji od liczebności oraz pochodzenia danych uczących, a stabilne wsparcie optymalizacji wymaga osiągnięcia minimalnej liczby próbek rzędu około 300, poniżej której rośnie ryzyko błędnego prowadzenia poszukiwań.
3. Wprowadzenie podziału na bazy *Full* oraz *NoDamage* było uzasadnione ze względu na odmienną charakterystykę odpowiedzi w obszarze bez inicjacji uszkodzeń i pozwoliło ograniczyć błąd predykcji w obszarze krytycznym dla oceny poprawności projektu.
4. W porównaniu metamodeli SSN, Kriging i PNR uznano, że SSN oraz Kriging są rozwiązaniami użytecznymi dla rozpatrywanego problemu, przy czym SSN wybrano jako metamodel roboczy, natomiast PNR wykazywał większą wrażliwość i mniejszą stabilność, co ograniczało jego zastosowanie w pętli optymalizacyjnej.
5. Zastosowanie podejść hybrydowych przyniosło istotny zysk czasowy względem wariantu odniesienia QEA z użyciem tylko MES: HYBRYDA1 pozwalała zachować poziom optimum porównywalny z podejściem pełnym przy skróceniu czasu do kilku godzin po zgromadzeniu odpowiedniej bazy danych, natomiast HYBRYDA2 umożliwiała dalszą

redukcję czasu do około godziny i stabilniejszy przebieg, kosztem większego ryzyka pogorszenia najlepszego wyniku, co potwierdza kompromis jakości do czasu w projektowaniu procedury hybrydowej.

6. W podejściach hybrydowych kluczowe są reguły sterujące wiarygodnością oceny: predykcje metamodelu powinny być okresowo weryfikowane obliczeniami MES, a dla punktów już występujących w bazie danych jako obowiązującą należy przyjmować wartość z MES; brak takich reguł zwiększa ryzyko propagacji błędu predykcji i zbieżności do minimum lokalnego.
7. Z uwagi na stochastyczny charakter QEA oraz możliwość minimów lokalnych, ocena metody powinna opierać się na serii niezależnych uruchomień oraz na stabilności rozwiązania końcowego; w raportowaniu przebiegu zasadne jest równoległe prezentowanie minimum globalnego, ponieważ wartości pośrednie mogą ulegać fluktuacjom wynikającym z mieszanej ewaluacji (za pomocą metamodelu oraz weryfikacji MES).
8. W procedurze hybrydowej istotne znaczenie ma mechanizm eliminacji powtórzeń obliczeń (baza wyników unikalnych), ponieważ ogranicza on liczbę zbędnych wywołań MES oraz stabilizuje proces uczenia metamodelu przez zapewnienie jednoznacznych par wejście-wyjście.
9. Dla zapewnienia poprawnej pracy metamodelu w pętli optymalizacji konieczne jest utrzymanie spójności przetwarzania danych (identyczne skalowanie cech wejściowych, identyczna definicja funkcji celu, stały sposób filtracji i podziału danych), ponieważ niespójność tych etapów prowadzi do błędów systematycznych predykcji.
10. Metodycznie uzasadnione jest stosowanie strategii adaptacyjnego uzupełniania bazy (douczenie metamodelu punktami wybranymi na podstawie przebiegu optymalizacji), ponieważ zwiększa to gęstość danych w obszarach istotnych decyzyjnie i redukuje błąd aproksymacji tam, gdzie algorytm faktycznie poszukuje rozwiązań.
11. W zadaniach o odpowiedzi progowej lub o odmiennych charakterach pracy układu celowe jest stosowanie rozdziału modelowania na przypadki o różnym charakterze (np. przez osobne metamodelowanie lub regułę wyboru modelu), ponieważ pojedynczy metamodel globalny częściej generuje błąd w pobliżu granicy między charakterami pracy.
12. Przyjęta dyskretyzacja przestrzeni decyzyjnej w algorytmie QEA determinuje rozdzielczość zmian parametrów, co należy traktować jako jawne ograniczenie metody wpływające na powtarzalność wskazywanych rozwiązań oraz na interpretację „różnic” pomiędzy wariantami w kontekście projektowym.
13. Dla zapewnienia weryfikowalności i odtwarzalności wyników procedury hybrydowej konieczne jest prowadzenie pełnego logowania (wersje modeli, parametry treningu, liczebność baz, identyfikatory osobników, źródło wartości funkcji celu), ponieważ bez tego nie jest możliwe rozdzielenie wpływu losowości algorytmu od zmian jakości metamodelu.

14. Metoda elementów skończonych nie stanowi bezwzględnego odniesienia o zerowej niepewności, ponieważ wynik obliczeń jest determinowany przez przyjęte idealizacje i założenia modelowe (w szczególności opis geometrii, warunki brzegowe i kontakt, model materiału oraz parametry numeryczne), jakość dyskretyzacji, a także kompletność i poprawność danych wejściowych. Z tego względu wiarygodność wyników MES wymaga weryfikacji numerycznej, obejmującej co najmniej ocenę wrażliwości oraz analizę zbieżności, a także walidacji w odniesieniu do danych eksperymentalnych, jeżeli są dostępne.
15. Po przeprowadzeniu kalibracji parametrów modelu oraz walidacji dla reprezentatywnych przypadków obciążenia możliwe jest uzyskanie zgodności wyników MES z pomiarami w granicach przyjętych tolerancji inżynierskich. Jednocześnie należy podkreślić, że zakres poprawności predykcji jest ograniczony do domeny, dla której wykonano identyfikację parametrów i walidację, natomiast stosowanie modelu, poza tym zakresem zwiększa ryzyko błędu modelowego.

W odniesieniu do tezy pracy można stwierdzić, że przeprowadzone badania potwierdziły możliwość efektywnego zastosowania algorytmu optymalizacji, obliczeń MES oraz metamodelowania w optymalizacji wybranych elementów obudowy baterii litowo jonowej z uwzględnieniem zjawisk dynamicznych. Efektywność metody wykazano przede wszystkim przez znaczące ograniczenie czasu obliczeń w wariantach hybrydowych względem podejścia referencyjnego QEA z użyciem tylko MES, przy jednoczesnym zachowaniu wyników końcowych na porównywalnym poziomie jakościowym. Potwierdzenie tezy należy interpretować w zakresie przyjętego modelu numerycznego, zdefiniowanej przestrzeni zmiennych projektowych, zastosowanych funkcji celu oraz analizowanego scenariusza wielokrotnego zrzutu. Opracowana procedura nie zastępuje metody elementów skończonych jako źródła informacji o odpowiedzi mechanicznej konstrukcji, lecz wykorzystuje wyniki MES do budowy i aktualizacji modeli zastępczych. Dzięki temu możliwe jest ograniczenie liczby kosztownych analiz wysokiej wierności bez rezygnacji z uwzględnienia zjawisk dynamicznych, kontaktu, odkształceń plastycznych oraz inicjacji uszkodzeń. Uzyskane wyniki wskazują, że metoda może być stosowana jako narzędzie wspomagające projektowanie konstrukcji obudów baterii litowo jonowych, szczególnie na etapie przeszukiwania przestrzeni projektowej i wyboru korzystnych wariantów geometrycznych. Warunkiem jej poprawnego stosowania jest jednak utrzymanie odpowiedniej jakości bazy danych, spójności przetwarzania danych oraz okresowa kontrola predykcji metamodelu za pomocą obliczeń MES. W tym ujęciu teza pracy została potwierdzona, przy czym zakres jej obowiązywania jest ograniczony do domeny projektowej i modelowej, dla której przeprowadzono identyfikację, uczenie, testowanie oraz weryfikację procedury.

Bibliografia

[1] T. Akiba, S. Sano, T. Yanase, T. Ohta, i M. Koyama, „Optuna: A Next-generation Hyperparameter Optimization Framework", w *KDD '19: Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, Anchorage AK USA: ACM, lip. 2019, s. 2623–2631. doi: 10.1145/3292500.3330701.

[2] Z. Alomar, L. Maccioni, i F. Concli, „Development and Implementation of Element Deletion Algorithm into an Open-Source Software Based on the Fracture Locus of Materials", *Materials*, t. 16, nr 1, s. 187, grud. 2022, doi: 10.3390/ma16010187.

[3] A. Ammar, W. Leclerc, H. Haddad, A. Chetouani, i M. Guessasma, „Thermo-hygro-mechanical coupling model based on discrete element method to simulate hygrothermal-induced damage of PA6/GF30 material", *International Journal of Solids and Structures*, t. 282, s. 112450, paź. 2023, doi: 10.1016/j.ijsolstr.2023.112450.

[4] K. Antosik, „Overview of the Development of Battery Technologies Used in Electric Vehicles Over the Years", *pjti*, t. 1, nr 1, s. 9–17, cze. 2025, doi: 10.37660/PJTI.2025.25.1.2.

[5] K. Antosik i A. Zelenska, „Analiza rozwoju elektromobilności na przestrzeni lat", *pjti*, t. 1, nr 1, s. 18–29, cze. 2025, doi: 10.37660/PJTI.2025.25.1.3.

[6] Y. Bai i T. Wierzbicki, „A new model of metal plasticity and fracture with pressure and Lode dependence", *International Journal of Plasticity*, t. 24, nr 6, Art. nr 6, cze. 2008, doi: 10.1016/j.ijplas.2007.09.004.

[7] Y. Bao i T. Wierzbicki, „On fracture locus in the equivalent strain and stress triaxiality space", *International Journal of Mechanical Sciences*, t. 46, nr 1, Art. nr 1, sty. 2004, doi: 10.1016/j.ijmecsci.2004.02.006.

[8] K.-J. Bathe, *Finite element procedures*. Englewood Cliffs, N.J: Prentice Hall, 1996.

[9] T. Belytschko, W. K. Liu, B. Moran, i K. I. Elkhodary, *Nonlinear Finite Elements for Continua and Structures*, 2st ed. w New York Academy of Sciences Series. Newark: John Wiley & Sons, Incorporated, 2014.

[10] J. Besson, „Continuum Models of Ductile Fracture: A Review", *International Journal of Damage Mechanics*, t. 19, nr 1, s. 3–52, sty. 2010, doi: 10.1177/1056789509103482.

[11] B. Bischl, M. Binder, M. Lang, T. Pielok, J. Richter, S. Coors, J. Thomas, T. Ullmann, M. Becker, A. Boulesteix, D. Deng, i M. Lindauer, „Hyperparameter optimization: Foundations, algorithms, best practices, and open challenges", *WIREs Data Min & Knowl*, t. 13, nr 2, s. e1484, mar. 2023, doi: 10.1002/widm.1484.

[12] G. E. Blomgren, „The Development and Future of Lithium Ion Batteries", *J. Electrochem. Soc.*, t. 164, nr 1, Art. nr 1, 2017, doi: 10.1149/2.0251701jes.

[13] G. E. P. Box, J. S. Hunter, i W. G. Hunter, *Statistics for experimenters: design, innovation and discovery*, 2. ed. w Wiley series in probability and statistics. New York: Wiley, 2005.

[14] L. Bravo Diaz, X. He, Z. Hu, F. Restuccia, M. Marinescu, J. V. Barreras, Y. Patel, G. Offer, i G. Rein, „Review—Meta-Review of Fire Safety of Lithium-Ion Batteries: Industry Challenges and Research Contributions", *J. Electrochem. Soc.*, t. 167, nr 9, s. 090559, sie. 2020, doi: 10.1149/1945-7111/aba8b9.

[15] T. Burczyński, W. Kuś, W. Beluch, A. Długosz, A. Poteralski, i M. Szczepanik, *Intelligent Computing in Optimal Design*, t. 261. w Solid Mechanics and Its Applications, vol. 261. Cham: Springer International Publishing, 2020. doi: 10.1007/978-3-030-34161-9.

[16] J.-N. Chazalviel, „Electrochemical aspects of the generation of ramified metallic electrodeposits", *Phys. Rev. A*, t. 42, nr 12, s. 7355–7367, grud. 1990, doi: 10.1103/PhysRevA.42.7355.

[17] Y. Chen, Y. Kang, Y. Zhao, L. Wang, J. Liu, Y. Li, Z. Liang, X. He, X. Li, N. Tavajohi, i B. Li, „A review of lithium-ion battery safety concerns: The issues, strategies, and testing standards", *Journal of Energy Chemistry*, t. 59, s. 83–99, sie. 2021, doi: 10.1016/j.jechem.2020.10.017.

[18] K. W. E. Cheng, B. P. Divakar, H. Wu, K. Ding, i H. F. Ho, „Battery-Management System (BMS) and SOC Development for Electrical Vehicles", *IEEE Trans. Veh. Technol.*, t. 60, nr 1, Art. nr 1, sty. 2011, doi: 10.1109/TVT.2010.2089647.

[19] T. Chugh, K. Sindhya, J. Hakanen, i K. Miettinen, „A survey on handling computationally expensive multiobjective optimization problems with evolutionary algorithms", *Soft Comput*, t. 23, nr 9, s. 3137–3166, maja 2019, doi: 10.1007/s00500-017-2965-0.

[20] L. R. Da Silveira, R. Tanscheit, i M. M. B. R. Vellasco, „Quantum inspired evolutionary algorithm for ordering problems", *Expert Systems with Applications*, t. 67, s. 71–83, sty. 2017, doi: 10.1016/j.eswa.2016.08.067.

[21] N. R. Draper i H. Smith, *Applied Regression Analysis*, 3. wyd. w New York Academy of Sciences Series, no. v. 326. Newark: John Wiley & Sons, Incorporated, 1998.

[22] J. E, H. Xiao, S. Tian, i Y. Huang, „A comprehensive review on thermal runaway model of a lithium-ion battery: Mechanism, thermal, mechanical, propagation, gas venting and combustion", *Renewable Energy*, t. 229, s. 120762, sie. 2024, doi: 10.1016/j.renene.2024.120762.

[23] X. Feng, M. Ouyang, X. Liu, L. Lu, Y. Xia, i X. He, „Thermal runaway mechanism of lithium ion battery for electric vehicles: A review", *Energy Storage Materials*, t. 10, s. 246–267, sty. 2018, doi: 10.1016/j.ensm.2017.05.013.

[24] X. Feng, D. Ren, X. He, i M. Ouyang, „Mitigating Thermal Runaway of Lithium-Ion Batteries”, *Joule*, t. 4, nr 4, Art. nr 4, kwi. 2020, doi: 10.1016/j.joule.2020.02.010.

[25] R. Fernandes De Mello i M. Antonelli Ponti, *Machine Learning: A Practical Approach on the Statistical Learning Theory*. Cham: Springer International Publishing, 2018. doi: 10.1007/978-3-319-94989-5.

[26] R. Fisher, *Design of experiments*, 9th ed. New York: Hafner Publishing Company, 1971.

[27] A. I. J. Forrester, A. Sóbester, i A. J. Keane, *Engineering Design via Surrogate Modelling: A Practical Guide*, 1. wyd. Wiley, 2008. doi: 10.1002/9780470770801.

[28] A. Fragiadaki i K. Tserpes, „Impact Response of a Thermoplastic Battery Housing for Transport Applications”, *Batteries*, t. 11, nr 10, s. 369, paź. 2025, doi: 10.3390/batteries11100369.

[29] S. S. Garud, I. A. Karimi, i M. Kraft, „Design of computer experiments: A review”, *Computers & Chemical Engineering*, t. 106, s. 71–95, lis. 2017, doi: 10.1016/j.compchemeng.2017.05.010.

[30] V. Hedges, „Cells of the Nervous System: The Neuron”, grud. 2022, Dostęp: 14 maja 2026. [Online]. Dostępne na: https://openbooks.lib.msu.edu/introneuroscience1/chapter/the-neuron/

[31] T. Heinzen, B. Plaumann, i M. Kaatz, „Influences on Vibration Load Testing Levels for BEV Automotive Battery Packs”, *Vehicles*, t. 5, nr 2, s. 446–463, kwi. 2023, doi: 10.3390/vehicles5020025.

[32] A. Hillerborg, M. Modéer, i P.-E. Petersson, „Analysis of crack formation and crack growth in concrete by means of fracture mechanics and finite elements”, *Cement and Concrete Research*, t. 6, nr 6, Art. nr 6, lis. 1976, doi: 10.1016/0008-8846(76)90007-7.

[33] G. A. Holzapfel, *Nonlinear solid mechanics: a continuum approach for engineering*, Repr. Chichester Weinheim: Wiley, 2010.

[34] H. Hooputra, H. Gese, H. Dell, i H. Werner, „A comprehensive failure model for crashworthiness simulation of aluminium extrusions”, *International Journal of Crashworthiness*, t. 9, nr 5, s. 449–464, wrz. 2004, doi: 10.1533/ijcr.2004.0289.

[35] T. J. R. Hughes, *The finite element method: linear static and dynamic finite element analysis*, 1. wyd. Englewood Cliffs, N.J.: Prentice Hall, 1987.

[36] A. Iserles, *A first course in the numerical analysis of differential equations*, Second edition. w Cambridge texts in applied mathematics, no. 44. Cambridge: Cambridge University Press, 2012. doi: 10.1017/CBO9780511995569.

[37] P. Jiang, Q. Zhou, i X. Shao, *Surrogate Model-Based Engineering Design and Optimization*. w Springer Tracts in Mechanical Engineering. Singapore: Springer Singapore, 2020. doi: 10.1007/978-981-15-0731-1.

[38] Y. Jin, „Surrogate-assisted evolutionary computation: Recent advances and future challenges”, *Swarm and Evolutionary Computation*, t. 1, nr 2, s. 61–70, cze. 2011, doi: 10.1016/j.swevo.2011.05.001.

[39] H. S. Kim, M. Cho, C. Lee, J. Kim, i S. Kang, „Numerical and Experimental Approaches for Mechanical Durability Assessment of an EV Battery Pack Case”, *Materials*, t. 18, nr 24, s. 5683, grud. 2025, doi: 10.3390/ma18245683.

[40] D. P. Kingma i J. Ba, „Adam: A Method for Stochastic Optimization”, zaprezentowano na Proceedings of the International Conference on Learning Representations (ICLR) 2015, 2015.

[41] G. Kjell i J. F. Lang, „Comparing different vibration tests proposed for li-ion batteries with vibration measurement in an electric vehicle”, w *2013 World Electric Vehicle Symposium and Exhibition (EVS27)*, Barcelona, Spain: IEEE, lis. 2013, s. 1–11. doi: 10.1109/EVS.2013.6914869.

[42] M. Kohutiar, L. Kakošová, M. Krbata, R. Janík, J. J. Fekiač, A. Breznická, M. Eckert, P. Mikuš, i Ľ. Timárová, „Comprehensive Review: Technological Approaches, Properties, and Applications of Pure and Reinforced Polyamide 6 (PA6) and Polyamide 12 (PA12) Composite Materials”, *Polymers*, t. 17, nr 4, s. 442, lut. 2025, doi: 10.3390/polym17040442.

[43] R. Krishnan, V. Packirisamy, D. Palani, i R. Ariyamuthu, „Historical perspective in technological advances in lithium-ion battery development”, *JCIS Open*, t. 21, s. 100163, kwi. 2026, doi: 10.1016/j.jciso.2025.100163.

[44] J. Krummenacker i J. Hausmann, „Determination of Fatigue Damage Initiation in Short Fiber-Reinforced Thermoplastic through Acoustic Emission Analysis”, *J. Compos. Sci.*, t. 5, nr 8, s. 221, sie. 2021, doi: 10.3390/jcs5080221.

[45] J. Kudela i R. Matousek, „Recent advances and applications of surrogate models for finite element method computations: a review”, *Soft Comput*, t. 26, nr 24, s. 13709–13733, grud. 2022, doi: 10.1007/s00500-022-07362-8.

[46] Kuk-Hyun Han i Jong-Hwan Kim, „Quantum-inspired evolutionary algorithm for a class of combinatorial optimization”, *IEEE Trans. Evol. Computat.*, t. 6, nr 6, Art. nr 6, grud. 2002, doi: 10.1109/TEVC.2002.804320.

[47] W. Kuś i A. Mrozek, „Quantum-inspired evolutionary optimization of SLMoS2 two-phase structures”, *CMMS*, t. 22, nr 2, Art. nr 2, 2022, doi: 10.7494/cmms.2022.2.0777.

[48] W. Kuś i W. Mucha, „Memetic Inverse Problem Solution in Cyber-physical Systems”, w *Advances in Technical Diagnostics*, w Applied Condition Monitoring, vol. 10. Cham: Springer International Publishing, 2018, s. 335–341. doi: 10.1007/978-3-319-62042-8_30.

[49] X. Lai, J. Yao, C. Jin, X. Feng, H. Wang, C. Xu, i Y. Zheng, „A Review of Lithium-Ion Battery Failure Hazards: Test Standards, Accident Analysis, and Safety Suggestions”, *Batteries*, t. 8, nr 11, Art. nr 11, lis. 2022, doi: 10.3390/batteries8110248.

[50] K. Leung i C. M. Tenney, „Towards First Principles prediction of Voltage Dependences of Electrolyte/Electrolyte Interfacial Processes in Lithium Ion Batteries”, *J. Phys. Chem. C*, t. 117, nr 46, s. 24224–24235, lis. 2013, doi: 10.1021/jp408974k.

[51] J.-Y. Li, Z.-H. Zhan, i J. Zhang, „Evolutionary Computation for Expensive Optimization: A Survey”, *Mach. Intell. Res.*, t. 19, nr 1, s. 3–23, lut. 2022, doi: 10.1007/s11633-022-1317-4.

[52] H. Löbberding, S. Wessel, C. Offermanns, M. Kehrer, J. Rother, H. Heimes, i A. Kampker, „From Cell to Battery System in BEVs: Analysis of System Packing Efficiency and Cell Types”, *WEVJ*, t. 11, nr 4, s. 77, grud. 2020, doi: 10.3390/wevj11040077.

[53] M. Łozowski, B. Osmołowski, T. Orłowski, A. Rurański, i D. Reclik, „A cell holder for a battery housing of an electrical vehicle or a power tool”, WO2024022584A1, 1 lutego 2024 Dostęp: 9 czerwca 2026. [Online]. Dostępne na: https://patents.google.com/patent/WO2024022584A1/en

[54] B. Macukow, „Sieci neuronowe, historia badań i podstawowe modele.”, w *Prace Naukowe Wydziału Elektroniki i Technik Informacyjnych Politechniki Warszawskiej,* WUT Publishing House., t. 1, Prace Naukowe EiTI PW, 2020, s. 90–108.

[55] M. D. McKay, R. J. Beckman, i W. J. Conover, „Comparison of Three Methods for Selecting Values of Input Variables in the Analysis of Output from a Computer Code”, *Technometrics*, t. 21, nr 2, s. 239–245, maja 1979, doi: 10.1080/00401706.1979.10489755.

[56] D. C. Montgomery, Red., *Design and analysis of experiments*, Eighth edition. Hoboken, NJ: John Wiley & Sons, Inc, 2013.

[57] Z. Mouti, K. Westwood, D. Long, i J. Njuguna, „Finite element analysis of localised impact loading on short glass fibre- reinforced polyamide engine oil pan subjected to low velocity impact from flying projectiles”, *Steel Research International*, t. 88, nr 10, s. 963–967, 2012, doi: 10.1002/srin.201100299.

[58] W. Mucha, „A new nondeterministic method for optimal selection of master degrees of freedom for dynamic condensation based on evolutionary optimization”, *Journal of Theoretical and Applied Mechanics*, t. 58, nr 2, s. 445–458, kwi. 2020, doi: 10.15632/jtam-pl/118823.

[59] W. Mucha, „Application of dynamic condensation for model order reduction in real-time hybrid simulations”, *Meccanica*, t. 58, nr 7, s. 1409–1425, lip. 2023, doi: 10.1007/s11012-023-01675-0.

[60] W. Mucha, „Real-time Operational Load Monitoring of a Composite Aerostructure Using FPGA-based Computing System”, *Bulletin of the Polish Academy of Sciences Technical Sciences*, s. 148251–148251, lis. 2023, doi: 10.24425/bpasts.2023.148251.

[61] W. Mucha i W. Kuś, „FPGA Support in Hybrid Simulation Using Finite Element Method”, *SSP*, t. 260, s. 105–112, lip. 2017, doi: 10.4028/www.scientific.net/SSP.260.105.

[62] W. Mucha i W. Kuś, „Metamodeling as a model order reduction technique in hybrid simulation using RTFEM”, zaprezentowano na 2ND INTERNATIONAL CONFERENCE ON CHEMISTRY, CHEMICAL PROCESS AND ENGINEERING (IC3PE), Yogyakarta, Indonesia, 2018, s. 020045. doi: 10.1063/1.5066507.

[63] W. Mucha i W. Kuś, „Mountain bicycle frame testing as an example of practical implementation of hybrid simulation using RTFEM”, zaprezentowano na COMPUTER METHODS IN MECHANICS (CMM2017): Proceedings of the 22nd International Conference on Computer Methods in Mechanics, Lublin, Poland, 2018, s. 140002. doi: 10.1063/1.5019144.

[64] R. H. Myers, D. C. Montgomery, i C. M. Anderson-Cook, *Response Surface Methodology: Process and Product Optimization Using Designed Experiments*, 1st ed. w New York Academy of Sciences Series. Newark: John Wiley & Sons, Incorporated, 2016.

[65] I. Negrin, M. Kripka, i V. Yepes, „Metamodel-assisted design optimization in the field of structural engineering: A literature review”, *Structures*, t. 52, s. 609–631, cze. 2023, doi: 10.1016/j.istruc.2023.04.006.

[66] H. Nouri, C. Czarnota, F. MERAGHNI, i P. Lory, „Fatigue damage model for short glass fibre reinforced thermoplastics (PA6-GF30)”, *ICCM International Conferences on Composite Materials*, lip. 2009.

[67] A. Öchsner, *Continuum Damage and Fracture Mechanics*. Singapore: Springer Singapore, 2016. doi: 10.1007/978-981-287-865-6.

[68] Optuna Developers, „Optuna Documentation 4.9.0 dev”. Dostęp: 15 maja 2026. [Online]. Dostępne na: https://optuna.readthedocs.io/_/downloads/en/latest/pdf/

[69] H.-S. Park i X.-P. Dang, „Structural optimization based on CAD–CAE integration and metamodeling techniques”, *Computer-Aided Design*, t. 42, nr 10, Art. nr 10, paź. 2010, doi: 10.1016/j.cad.2010.06.003.

[70] Python Software Foundation, „Python 3.14 documentation”, Python documentation. Dostęp: 15 maja 2026. [Online]. Dostępne na: https://docs.python.org/3/

[71] N. V. Queipo, R. T. Haftka, W. Shyy, T. Goel, R. Vaidyanathan, i P. Kevin Tucker, „Surrogate-based analysis and optimization”, *Progress in Aerospace Sciences*, t. 41, nr 1, s. 1–28, sty. 2005, doi: 10.1016/j.paerosci.2005.02.001.

[72] T. Rahman i T. Alharbi, „Exploring Lithium-Ion Battery Degradation: A Concise Review of Critical Factors, Impacts, Data-Driven Degradation Estimation Techniques, and Sustainable Directions for Energy Storage Systems”, *Batteries*, t. 10, nr 7, s. 220, cze. 2024, doi: 10.3390/batteries10070220.

[73] C. E. Rasmussen i C. K. I. Williams, *Gaussian processes for machine learning*, 3. print. w Adaptive computation and machine learning. Cambridge, Mass.: MIT Press, 2008.

[74] F. Rosenblatt, *The Perceptron, a Perceiving and Recognizing Automaton: (Project Para)*. w Report / Cornell Aeronautical Laboratory. Cornell Aeronautical Laboratory, 1957. [Online]. Dostępne na: https://books.google.pl/books?id=P_XGPgAACAAJ

[75] M. Rosso, C. Brissot, A. Teyssot, M. Dollé, L. Sannier, J.-M. Tarascon, R. Bouchet, i S. Lascaud, „Dendrite short-circuit and fuse effect on Li/polymer/Li cells”, *Electrochimica Acta*, t. 51, nr 25, s. 5334–5340, lip. 2006, doi: 10.1016/j.electacta.2006.02.004.

[76] A. Rurański i W. Kuś, „The Quantum-Inspired Evolutionary Algorithm in the Parametric Optimization of Lithium-Ion Battery Housing in the Multiple-Drop Test”, *Batteries*, t. 10, nr 9, s. 308, sie. 2024, doi: 10.3390/batteries10090308.

[77] A. Rurański i W. Kuś, „Metamodel Selection and Hyperparameter Tuning for Hybrid Optimisation with a Quantum-Inspired Evolutionary Algorithm in Multiple Drop Test Simulation”, *Computer Assisted Methods in Engineering and Science*, 2026, doi: https://doi.org/10.24423/cames.2026.1971.

[78] J. Sacks, W. J. Welch, T. J. Mitchell, i H. P. Wynn, „Design and Analysis of Computer Experiments”, *Statist. Sci.*, t. 4, nr 4, lis. 1989, doi: 10.1214/ss/1177012413.

[79] E. Sahraei, J. Meier, i T. Wierzbicki, „Characterizing and modeling mechanical properties and onset of short circuit for three types of lithium-ion pouch cells”, *Journal of Power Sources*, t. 247, s. 503–516, lut. 2014, doi: 10.1016/j.jpowsour.2013.08.056.

[80] T. J. Santner, B. J. Williams, i W. I. Notz, *The Design and Analysis of Computer Experiments*, 2nd ed. w Springer Series in Statistics Ser. New York, NY: Springer, 2019.

[81] N. G. Sauer, B. Gaudet, i A. Barowy, „Experimental investigation of explosion hazard from lithium-ion battery thermal runaway effluent gas”, *Fuel*, t. 378, s. 132818, grud. 2024, doi: 10.1016/j.fuel.2024.132818.

[82] P. Sebastjan i W. Kuś, „Optimization of material distribution for forged automotive components using hybrid optimization techniques”, *CMMS*, t. 21, nr 2, Art. nr 2, 2021, doi: https://doi.org/10.7494/cmms.2021.2.0746.

[83] P. Sebastjan i W. Kuś, „Method for Parameter Tuning of Hybrid Optimization Algorithms for Problems with High Computational Costs of Objective Function Evaluations”, *Applied Sciences*, t. 13, nr 10, Art. nr 10, maja 2023, doi: 10.3390/app13106307.

[84] T. W. Simpson, J. D. Poplinski, P. N. Koch, i J. K. Allen, „Metamodels for Computer-based Engineering Design: Survey and recommendations”, *EWC*, t. 17, nr 2, s. 129–150, lip. 2001, doi: 10.1007/PL00007198.

[85] N. Smol, „The risk of lithium-ion batteries - Thermal Runaway in EV’s”, Fire Isolator. Dostęp: 13 stycznia 2026. [Online]. Dostępne na: https://fireisolator.com/the-risk-of-lithium-ion-batteries-thermal-runaway-in-evs/

[86] E. Soós i T. Goda, „Mechanical behavior of glass fiber-reinforced bosses: experiments and FE simulations”, *Periodica Polytechnica Mechanical Engineering.*, t. 55, nr 1, s. 3–14, 2011, doi: https://doi.org/10.3311/pp.me.2011-1.01.

[87] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, i R. Salakhutdinov, „Dropout: A Simple Way to Prevent Neural Networks from Overfitting”, *Journal of Machine Learning Research*, t. 15, nr 56, s. 1929–1958, 2014.

[88] J. Steiger, D. Kramer, i R. Mönig, „Mechanisms of dendritic growth investigated by in situ light microscopy during electrodeposition and dissolution of lithium”, *Journal of Power Sources*, t. 261, s. 112–119, wrz. 2014, doi: 10.1016/j.jpowsour.2014.03.029.

[89] R. Szabo, „«Dangerous» power tool explodes, damaging work vehicle”, Australasian Mine Safety Journal. Dostęp: 13 stycznia 2026. [Online]. Dostępne na: https://www.amsj.com.au/dangerous-power-tool-explodes-damaging-work-vehicle/

[90] T. Szandała, „Review and Comparison of Commonly Used Activation Functions for Deep Neural Networks”, w *Bio-inspired Neurocomputing*, t. 903, A. K. Bhoi, P. K. Mallick, C.-M. Liu, i V. E. Balas, Red., w Studies in Computational Intelligence, vol. 903. , Singapore: Springer Singapore, 2021, s. 203–224. doi: 10.1007/978-981-15-5495-7_11.

[91] S. Szpak, C. J. Gabriel, i J. R. Driscoll, „Catastrophic thermal runaway in lithium batteries”, *Electrochimica Acta*, t. 32, nr 2, Art. nr 2, lut. 1987, doi: 10.1016/0013-4686(87)85030-2.

[92] F. A. C. Viana, T. W. Simpson, V. Balabanov, i V. Toropov, „Special Section on Multidisciplinary Design Optimization: Metamodeling in Multidisciplinary Design Optimization: How Far Have We Really Come?”, *AIAA Journal*, t. 52, nr 4, s. 670–690, kwi. 2014, doi: 10.2514/1.J052375.

[93] G. G. Wang i S. Shan, „Review of Metamodeling Techniques in Support of Engineering Design Optimization”, *Journal of Mechanical Design*, t. 129, nr 4, s. 370–380, kwi. 2007, doi: 10.1115/1.2429697.

[94] Q. Wang, P. Ping, X. Zhao, G. Chu, J. Sun, i C. Chen, „Thermal runaway caused fire and explosion of lithium ion battery”, *Journal of Power Sources*, t. 208, s. 210–224, cze. 2012, doi: 10.1016/j.jpowsour.2012.02.038.

[95] T. Wierzbicki, Y. Bao, Y.-W. Lee, i Y. Bai, „Calibration and evaluation of seven fracture models”, *International Journal of Mechanical Sciences*, t. 47, nr 4–5, Art. nr 4–5, kwi. 2005, doi: 10.1016/j.ijmecsci.2005.03.003.

[96] C. Witzgall i S. Wartzack, „Fatigue Damage of Short Fibre-Reinforced Thermoplastics in Crashworthiness Simulation”, *Applied Mechanics*, t. 4, nr 4, s. 1188–1205, lis. 2023, doi: 10.3390/applmech4040061.

[97] M. Wu, B. Xu, i C. Ouyang, „Physics of electron and lithium-ion transport in electrode materials for Li-ion batteries”, *Chinese Phys. B*, t. 25, nr 1, s. 018206, sty. 2016, doi: 10.1088/1674-1056/25/1/018206.

[98] Y. Xiao, F. Yang, Z. Gao, M. Liu, J. Wang, Z. Kou, Y. Lin, Y. Li, L. Gao, Y. Chen, S. Ren, i X. Li, „Review of mechanical abuse related thermal runaway models of lithium-ion batteries at different scales”, *Journal of Energy Storage*, t. 64, s. 107145, sie. 2023, doi: 10.1016/j.est.2023.107145.

[99] L. Xue i T. Wierzbicki, „Ductile fracture initiation and propagation modeling using damage plasticity theory”, *Engineering Fracture Mechanics*, t. 75, nr 11, s. 3276–3293, lip. 2008, doi: 10.1016/j.engfracmech.2007.08.012.

[100] G. Zhang, „Quantum-inspired evolutionary algorithms: a survey and empirical study”, *J Heuristics*, t. 17, nr 3, Art. nr 3, cze. 2011, doi: 10.1007/s10732-010-9136-0.

[101] X. Zhang, S. Chen, J. Zhu, i Y. Gao, „A Critical Review of Thermal Runaway Prediction and Early-Warning Methods for Lithium-Ion Batteries”, *Energy Material Advances*, t. 4, s. 0008, sty. 2023, doi: 10.34133/energymatadv.0008.

[102] O. C. Zienkiewicz i R. L. Taylor, *The finite element method for solid and structural mechanics*, 6. ed., Repr. Amsterdam Heidelberg: Elsevier Butterworth-Heinemann, 2006.

[103] O. C. Zienkiewicz, R. L. Taylor, i J. Zhu, *The finite element method: its basis and fundamentals*, 6. ed., Reprint., Transferred to digital print. Amsterdam Heidelberg: Elsevier, 2010.

[104] *Recommendations on the transport of dangerous goods - manual of tests and criteria. Hauptband*, Sixth revised edition. w ST/SG/AC 10, no. 11,Rev. 6. New York: United Nations, 2015.

[105] „ABAQUS 2024 Abaqus Analysis User’s Manual, Version 2024,” *Dassault Systems Simulia*, 2024.

[106] „Battery from common household tool explodes in Melbourne home”, 7NEWS. Dostęp: 13 stycznia 2026. [Online]. Dostępne na: https://7news.com.au/news/lithium-ion-battery-from-drill-explodes-and-causes-100000-worth-of-damage-in-melbourne-home-c-14116752

[107] „AI and Machine Learning in Simulation”, Dassault Systèmes. Dostęp: 4 listopada 2025. [Online]. Dostępne na: https://www.3ds.com/products/simulia/ai-and-machine-learning-simulation

[108] „AI and Machine Learning for Engineering Design Generation and Optimization”, Altair. Dostęp: 4 listopada 2025. [Online]. Dostępne na: https://altair.com/resource/ai-and-machine-learning-for-engineering-design-generation-and-optimization

[109] „Ansys AI - AI-Augmented Simulation Technology”. Dostęp: 4 listopada 2025. [Online]. Dostępne na: https://www.ansys.com/ai

[110] „Concerns over growing number of fires linked to lithium-ion batteries in e-scooters and e-bikes - ABC News”. Dostęp: 13 stycznia 2026. [Online]. Dostępne na: https://www.abc.net.au/news/2023-01-18/e-scooter-lithium-battery-fire-risk-fears-/101863902

Streszczenie

Praca dotyczy zastosowania metamodelowania i metod optymalizacji w projektowaniu wybranych elementów konstrukcyjnych baterii litowo jonowych poddawanych obciążeniom dynamicznym. Głównym problemem badawczym jest wysoki koszt obliczeniowy optymalizacji, gdy ocena kolejnych wariantów konstrukcyjnych wymaga przeprowadzania nieliniowych analiz metodą elementów skończonych, uwzględniających kontakt, odkształcenia plastyczne oraz rozwój uszkodzeń.

Opracowano hybrydową metodę optymalizacji łączącą kwantowo inspirowany algorytm ewolucyjny, MES oraz metamodelowanie. Modele zastępcze budowano na podstawie wyników analiz MES i wykorzystywano do ograniczenia liczby kosztownych ewaluacji funkcji celu. Porównano sztuczne sieci neuronowe, Kriging oraz regresję wielomianową, analizując wpływ liczebności i pochodzenia danych uczących na jakość predykcji. Zastosowano również odrębne modele dla obszarów o różnym charakterze odpowiedzi mechanicznej, w szczególności przypadków z inicjacją uszkodzeń i bez jej wystąpienia.

Metodę zweryfikowano na parametrycznym modelu zintegrowanej obudowy baterii litowojonowej z trzema zmiennymi geometrycznymi. Model numeryczny opracowano w Abaqus, a automatyzację obliczeń zrealizowano w języku Python. Porównano podejście referencyjne, wykorzystujące MES do wyznaczania funkcji celu, z wariantami hybrydowymi wykorzystującymi predykcję metamodeli.

Uzyskane wyniki wykazały możliwość znaczącego ograniczenia czasu obliczeń przy zachowaniu zbliżonych wyników optymalizacji. Łączny czas czterech prób zmniejszono ze 133 h dla podejścia referencyjnego do 39 h 25 min oraz 28 h 5 min dla wariantów hybrydowych. Najkorzystniejszy kompromis pomiędzy czasem obliczeń a wiarygodnością rozwiązania uzyskano w wariancie, w którym najlepszy wynik wskazany przez metamodel był następnie weryfikowany za pomocą MES. Przeprowadzone badania potwierdziły możliwość efektywnego wykorzystania metamodelowania jako narzędzia wspomagającego optymalizację konstrukcji baterii poddawanych obciążeniom dynamicznym.

Abstract

The dissertation concerns the application of metamodeling and optimization methods to the design of selected structural components of lithium-ion batteries subjected to dynamic loading. The main research problem is the high computational cost of optimization when successive design variants require nonlinear finite element analyses accounting for contact, plastic deformation, and damage evolution.

A hybrid optimization method combining a quantum-inspired evolutionary algorithm, FEM, and metamodeling was developed. Surrogate models were constructed using FEM results to reduce the number of computationally expensive objective function evaluations. Artificial neural networks, Kriging, and polynomial regression were compared, with particular attention to the influence of training dataset size and origin on prediction quality. Separate surrogate models were also applied to regions characterized by different types of mechanical response, particularly cases with and without damage initiation.

The methodology was verified using a parametric model of an integrated lithium-ion battery housing with three geometrical design variables. The numerical model was developed in Abaqus, while Python was used to automate the computational procedure. The reference approach based on FEM objective function evaluations was compared with hybrid variants employing surrogate model predictions.

The results demonstrated that computational time can be significantly reduced while maintaining comparable optimization results. The total time of four optimization runs was reduced from 133 h for the reference approach to 39 h 25 min and 28 h 5 min for the hybrid variants. The most favorable compromise between computational efficiency and solution reliability was obtained when the best solution indicated by the surrogate model was subsequently verified using FEM. The study confirmed that metamodeling can effectively support the optimization of battery structures subjected to dynamic loading.