



PLANOWANIE DOŚWIADCZEŃ JAKO DROGA DO INNOWACYJNOŚCI – PRZYKŁAD OPTYMALIZACJI PROCESU PRODUKCYJNEGO

Grzegorz Harańczyk, StatSoft Polska Sp. z o.o.

W artykule pokazano, w jaki sposób, wykorzystując planowanie doświadczeń, poprawić proces produkcyjny. Optymalizowano pewien proces chemiczny ze względu na jego wydajność. Oprócz poprawy wydajności dbano o jakość procesu oraz jego realizowalność w praktyce, dlatego też ocena optymalnych parametrów analizowanego procesu była wielokryterialna.

Planowanie doświadczeń a innowacyjność w przemyśle

Podczas realizacji każdego procesu wytwarzania pojawiają się pytania, czy produkt mógłby być lepszy, czy produkcja mogłaby być tańsza, stabilniejsza lub szybsza. Rozwiązaniem może być instalacja nowej, ulepszonej technologii produkcyjnej, ale niesie to ze sobą duże koszty i częstokroć nie jest możliwe do zrealizowania. Wówczas rozwiązaniem może być optymalizacja parametrów produkcji przy zachowaniu dotychczasowej technologii wytwarzania.

Każdy proces technologiczny charakteryzuje się wieloma parametrami, których wielkość potencjalnie ma wpływ na jakość uzyskiwanego produktu. Te parametry będziemy nazywali wielkościami wejściowymi. Dokładnego zdefiniowania wymaga jakość produktu, często niedająca się wyrazić za pomocą jednego wskaźnika, a określana przez cały zespół cech, co gorsza poprawa jednego wskaźnika jakości może wiązać się z pogorszeniem innych. Typowym zagadnieniem w procesie rozwoju produktu jest znalezienie zbioru warunków i wartości wielkości wejściowych, które umożliwiają uzyskanie wyrobu najbardziej pożądanego w sensie jego charakterystyk - wielkości wyjściowych. Procedury używane do rozwiązywania tego problemu składają się z dwóch kroków:

1. Znalezienia związku pomiędzy wielkościami wejściowymi a wielkościami opisującymi jakość produktu. Zadanie to realizowane jest na ogół w dwóch etapach: najpierw znajdujemy wielkości wejściowe mające istotny wpływ na badane zjawisko, a następnie próbujemy opisać ten związek ilościowo poprzez wyznaczenie odpowiedniego modelu, najczęściej w postaci równania.
2. Znalezienia wartości wielkości wejściowych, które zapewnią uzyskanie najbardziej użytecznych (pożyczanych) aproksymowanych wartości wielkości wyjściowych.



Zależność pomiędzy parametrami wejściowymi a jakością może być prosta lub bardzo skomplikowana (jej wyznaczenie wówczas może sprawiać spore problemy). Gdy badany proces zależy tylko od jednego parametru i jego wpływ, w badanym zakresie, jest liniowy, to problem nie wymaga specjalnego podejścia. Takich procesów jednak jest mało. Jeśli proces zależy od kilku parametrów, których wpływ i wzajemne interakcje między nimi nie są znane, wówczas problem staje się złożony. Należy wówczas korzystać z wszelkiej dostępnej wiedzy i dotychczasowego doświadczenia w prowadzeniu procesu oraz przeprowadzić dodatkowe eksperymenty mające na celu poznanie procesu.

Dodatkowa trudność pojawia się, jeśli chcemy jednocześnie optymalizować kilka wielkości, a to też jest częsta sytuacja. W tym artykule właśnie na tym problemie skupimy się w dalszej części i omówimy go na przykładzie optymalizacji pewnego procesu produkcyjnego. Podobny przykład został opisany w [Montgomery 2005] oraz [Myers, Montgomery, Anderson-Cook 2009] lub [Derringer, Suich 1980].

Planowanie doświadczeń

W związku z poszukiwaniem innowacji pojawia się potrzeba uzyskania jak najpełniejszej informacji dotyczącej czynników wpływających na proces produkcji. Najlepiej oczywiście byłoby dokonać pomiarów dla każdej możliwej kombinacji wartości wielkości wejściowych. W praktyce nie jest to jednak prawie nigdy możliwe, gdyż każdy układ wielkości wejściowych wymaga często czasochłonnej i kosztownej regulacji linii produkcyjnej. Należy zatem wybrać tylko pewne punkty w przestrzeni wielkości wejściowych, w których wykonamy pomiary.

Przeprowadzając pomiary w zbyt małej liczbie punktów, możemy nie być w stanie dopasować odpowiedniego modelu. W związku z tym często postępuje się w następujący sposób: najpierw używa się mniej skomplikowanych planów, wymagających mniejszej liczby układów, aby wybrać te wielkości wejściowe, które rzeczywiście mają istotny wpływ na badane wielkości wyjściowe. A następnie dla tych wybranych krytycznych parametrów procesu wykonuje się dodatkowe eksperymenty, aby oszacować już postać związku i wielkość wpływu (tzw. efekt) na pożądane parametry procesu. Przykład takiego podejścia został opisany w pracy [Harańczyk, Woyna-Orlewicz 2009], w przykładzie 2 opisano redukcję liczb wielkości wejściowych, a następnie dokładniejszą analizę w przykładzie 3.

Wstępne analizy pozwalają nie tylko zmniejszyć liczbę potencjalnych parametrów krytycznych, ale także często skorygować ich poziomy. Często na wstępnym etapie wykorzystywane są plany frakcyjne (np. tak zwane plany 2^k) i uzyskane dzięki nim modele liniowe mogą wskazać kierunek, wzdłuż którego wraz ze zmianą wielkości wejściowych następuje najszybsza zmiana wielkości wyjściowych. Na tym etapie należy uwzględnić też posiadaną wiedzę o procesie, wynikającą z doświadczenia. Pomaga to wybrać odpowiednie wielkości krytyczne i poziomy ich wartości.

Wybierając zestaw układów wielkości wejściowych, dla których dokonamy pomiarów, należy pamiętać, że oprócz wpływu poszczególnych wielkości chcemy również ocenić ich



wzajemne interakcje. W języku eksperymentu interakcja oznacza, że wartości wielkości wyjściowej, będącej efektem łącznego działania co najmniej dwóch czynników, nie można określić na podstawie znajomości efektu każdego z tych czynników oddzielnie. Nie obsadzając pewnych układów w zaplanowanym eksperymencie, możemy nie być w stanie wykryć tego typu zjawisk.

Kolejną ważną kwestią jest decyzja, czy powtarzać pewne układy zaplanowanego eksperymentu. Wykonanie pomiarów dla każdego układu planu więcej niż jeden raz pozwala oszacować tak zwany czysty błąd (ang. *pure error*) doświadczenia. Zmienność ta wskazuje na wielkość błędu losowego w pomiarach (np. z powodu niekontrolowanych czynników, niepewności przyrządów pomiarowych itp.), gdyż powtórzone pomiary wykonywane są w identycznych warunkach (tj. ustalonych wartościach wielkości wejściowych). Jeżeli zmienność resztowa jest znacząco większa od zmienności czystego błędu, to można uznać, że w dalszym ciągu występuje pewna statystycznie istotna zmienność, którą można przypisać różnicom pomiędzy grupami, a więc występuje ogólny brak dopasowania (nieadekwatność, ang. *lack of fit*) aktualnego modelu.

W sytuacji, gdy nie jest możliwe lub celowe powtarzanie pomiarów dla wszystkich układów planu (tj. powtórzenie całego planu doświadczenia), nadal można wyznaczyć ocenę czystego błędu przez powtórzenie jedynie niektórych układów planu. Jednakże należy uważnie rozpatrzyć możliwe obciążenia estymatorów, które mogą zostać wprowadzone poprzez selektywne powtarzanie jedynie niektórych układów planu. Jeżeli powtarzane są jedynie układy, które najłatwiej powtórzyć (np. wykonanie pomiaru jest w ich przypadku najtańsze), można pechowo wybrać takie układy, dla których losowa zmienność wielkości wyjściowej jest bardzo mała (lub bardzo duża), w efekcie prawdziwa wartość czystego błędu zostanie niedoszacowana (lub przeszacowana). W takiej sytuacji należy rozważyć, najczęściej ufając własnemu doświadczeniu i wiedzy o badanym procesie, które układy planu powinny być powtórzone, to znaczy, które układy dostarczą dobrą (nieobciążoną) ocenę czystego błędu. Należy tu przypomnieć, że powtórzenie planu nie jest tożsame z wykonaniem serii pomiarów na tym samym układzie wielkości wejściowych. Za każdym razem nastawy parametrów procesu powinny być wykonane ponownie. Wykonanie serii produktów i wykonanie pomiarów dla każdego z nich zazwyczaj powoduje zaniżenie czystego błędu.

Powyżej opisano na co warto zwrócić uwagę przed przystąpieniem do badania procesu. Oczywiście można jednak wyobrazić sobie sytuację, że zwykle „próbowanie” przy różnych ustawieniach parametrów procesu da zadowalające rezultaty, ale nie ma co do tego pewności, a na pewno nie zwiększy to w dużym stopniu wiedzy o badanym procesie. Zaleca się raczej korzystanie z wypracowanych schematów podejścia systematycznego – z tzw. planów doświadczeń. Uwzględniając powyższe zalecenia, należy wybrać odpowiedni plan doświadczenia, który potem zostanie zrealizowany. Odpowiedni plan można wygenerować w module *Planowanie doświadczeń* w programie *STATISTICA*. Program oferuje również narzędzia do analizy wyników uzyskanych w trakcie realizacji zaplanowanego doświadczenia oraz poszukiwania wartości optymalnych.



Opis analizowanego procesu produkcyjnego

Naszym zadaniem będzie zoptymalizowanie pewnego procesu chemicznego. Należy tak dobrać parametry tego procesu, aby poprawić jego wydajność. Oprócz poprawy wydajności chcemy dbać o jakość procesu oraz jego realizowalność w praktyce, dlatego ocena optymalnych parametrów procesu będzie wielokryterialna. Dodatkowo zostaną uwzględnione dwie wielkości wyjściowe: lepkość oraz liczbowa średnia masa cząsteczkowa. Wymagamy od procesu, aby otrzymywany produkt miał odpowiednie parametry i aby wydajność procesu była możliwie jak największa. Zdefiniowano pożądane zakresy i tak: lepkość finalnego wyrobu powinna mieścić się w przedziale [63, 69], natomiast liczbowa średnia masa cząsteczkowa w przedziale [3300, 3500].

Po wstępnych analizach i bazując na dotychczasowej wiedzy o procesie wybrano dwie wielkości wejściowe, których wpływ wydaje się być ważny. Są to: temperatura mierzona w stopniach Celsjusza i czas mierzony w minutach. Dotychczas proces produkcyjny realizowano przez 35 minut w temperaturze 68 stopni Celsjusza. Wykonano wstępne pomiary zgodnie z planem 2^k i, korzystając z metodyki największego spadku, oceniono, że poprawa wydajności powinna nastąpić w warunkach 79 stopni Celsjusza i w czasie 85 minut. Do poszukiwaniu modelu opisującego badany proces wybrano plan centralny-kompozycyjny (z centrum w wybranym wcześniej punkcie i o ramionach długości odpowiednio: 3 stopnie Celsjusza oraz 5 minut) (por. [Konkol 2007]).

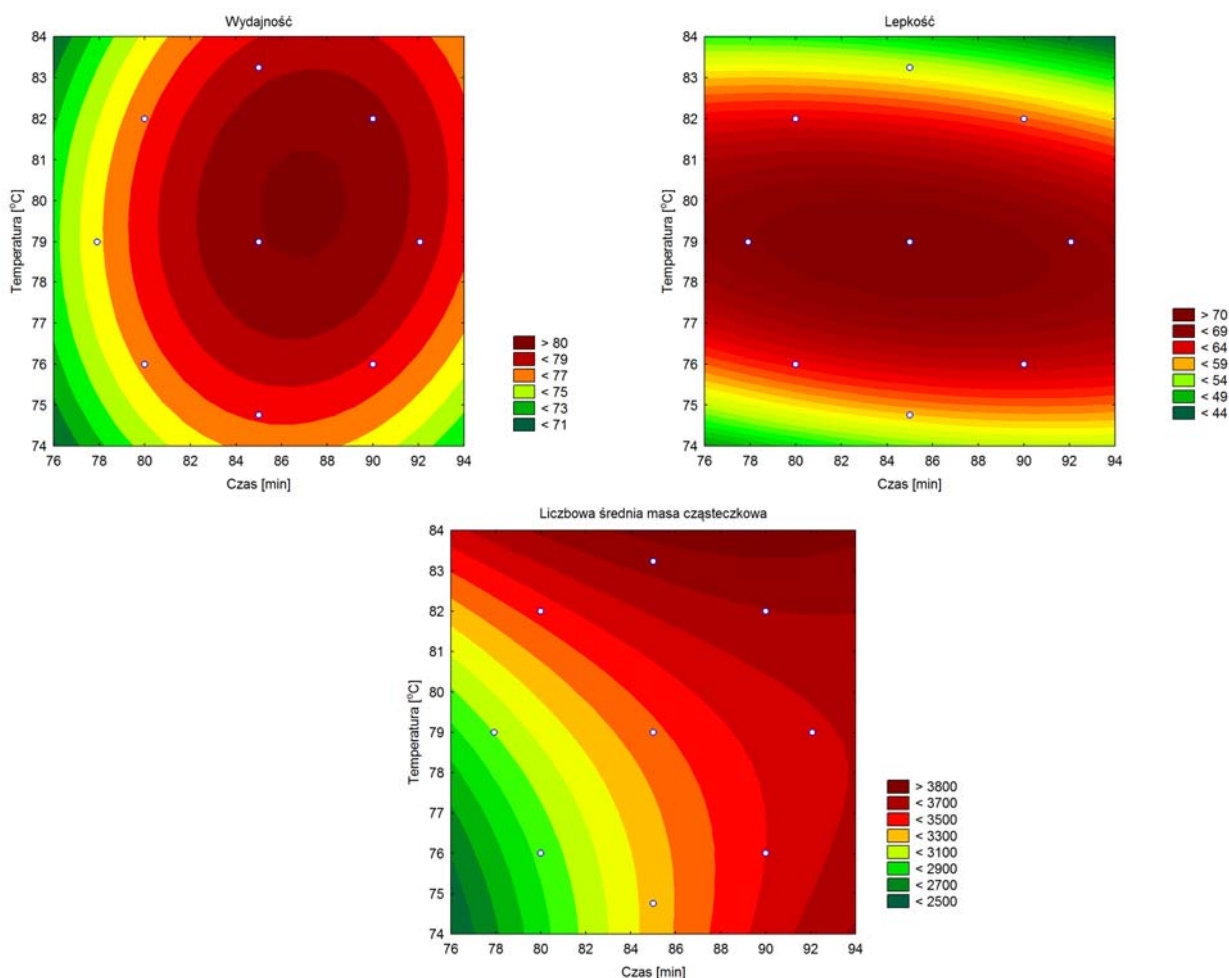
Analiza uzyskanych wyników

Przechodzimy do analizy uzyskanych wyników. Ma ona dostarczyć nam zarówno wiedzy o badanym procesie, jak i pomóc wyznaczyć takie wartości wielkości wejściowych, dla których wartości wielkości wyjściowych są najbardziej pożądane. W pierwszym kroku na podstawie wykonanych pomiarów zbadano za pomocą metod analizy wariancji, które wielkości wejściowe wpływają znacząco na wielkość wyjściową, oraz wyznaczono dla każdej ze zmiennych wyjściowych odpowiednią powierzchnię odpowiedzi (rys. 1). Dzięki odpowiednim modelom w postaci równań i miar wielkości efektu można ocenić wpływ poszczególnych zmiennych wejściowych na każdą ze zmiennych wyjściowych. Na wykresach ilustrujących te powierzchnie odpowiedzi widać także charakter wpływu oraz obszary, w których wielkości wejściowe przyjmują wartości największe. W tym przykładzie widać, że wartości pożądane dla każdej z wielkości wyjściowych są realizowane dla różnych kombinacji wielkości wejściowych. Jest to sytuacja dość typowa, trudno wyobrazić sobie, aby wszystkie cechy wyrobu zależały w ten sam sposób od warunków wytwarzania.

Określając zadowalający poziom wydajności oraz biorąc pod uwagę pozostałe ograniczenia na inne zmienne wyjściowe, można poprzestać na wyznaczeniu obszaru, dla którego wszystkie te parametry są na odpowiednim poziomie (por. [Harańczyk, Woyna-Orlewicz 2009, Przykład 1]). Można jednak również spróbować wyznaczyć wartość optymalną ze względu na wybrane parametry, jednak musimy wówczas zdefiniować nową



wielkość – jakość całkowitą oraz określić wkład poszczególnych składników jakości na jakość całkowitą.



Rys. 1. Powierzchnie odpowiedzi dla wielkości wyjściowych w zależności od temperatury i czasu.

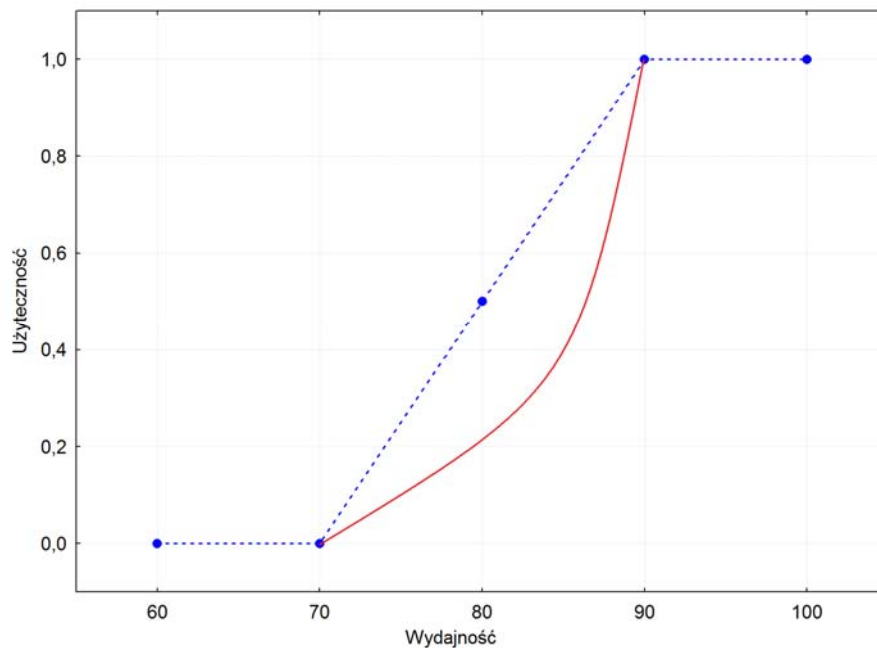
Jednoczesna optymalizacja wielu wielkości wyjściowych

Aby móc zoptymalizować proces ze względu na wiele wielkości wyjściowych, najczęściej stosuje się podejście zaproponowane w pracy [Derringer, Suich 1980]. Podejście to polega na przekształceniu aproksymowanych wartości wielu wielkości wyjściowych na pojedynczą wartość całkowitej użyteczności. Zagadnienie jednoczesnej optymalizacji wielu wielkości wyjściowych upraszcza się wtedy do znalezienia wartości wielkości wyjściowych, które maksymalizują całkowitą użyteczność odpowiedzi wielkości wyjściowych.

Zanim zdefiniujemy użyteczność całkowitą, w pierwszym kroku należy zdefiniować profil użyteczności odpowiedzi dla każdej wielkości optymalizowanej. Polega to na przypisaniu każdej z możliwych wartości badanej cechy wartości z przedziału [0,1], mówiącej o satysfakcji z uzyskania wyniku na takim poziomie. W szczególności zdefiniowanie, czy

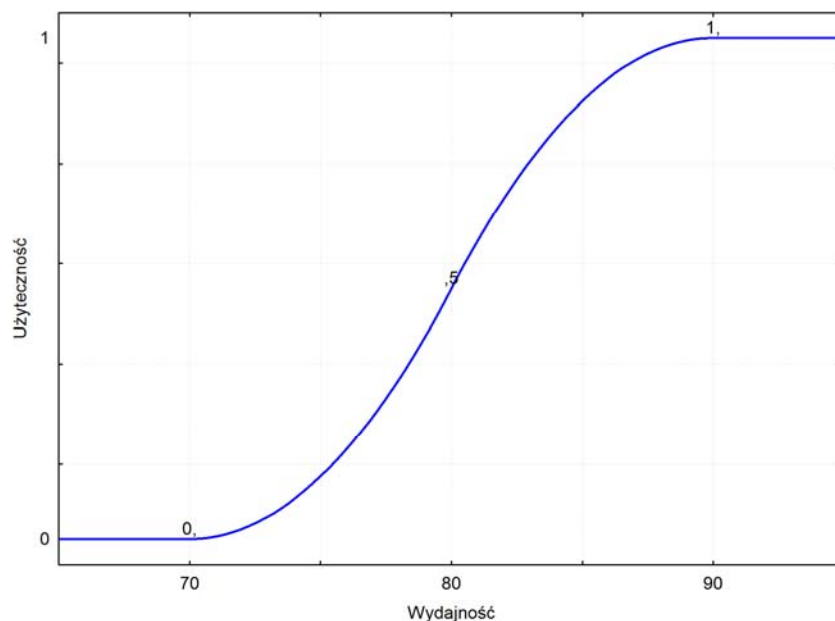


im większe natężenie optymalizowanej cechy, to tym lepiej czy tym gorzej. Pozwala to również na zdefiniowanie zakresu akceptowalnych zakresów odpowiedzi procesu.



Rys. 2. Dwa przykładowe profile użyteczności (linia przerywana – profil liniowy, linia ciągła – „bardziej wymagający” profil o zmodyfikowanej krzywiznie).

Widać, że oprócz zdefiniowania, że im większa wydajność, tym lepiej, możemy również określić tempo zmiany naszej satysfakcji wraz ze zmianą wydajności procesu. Linia ciągłą zaznaczono na rys. 2 profil „bardziej wymagający” – taką samą satysfakcję uzyskujemy przy większych wartościach wydajności niż dla profilu liniowego.

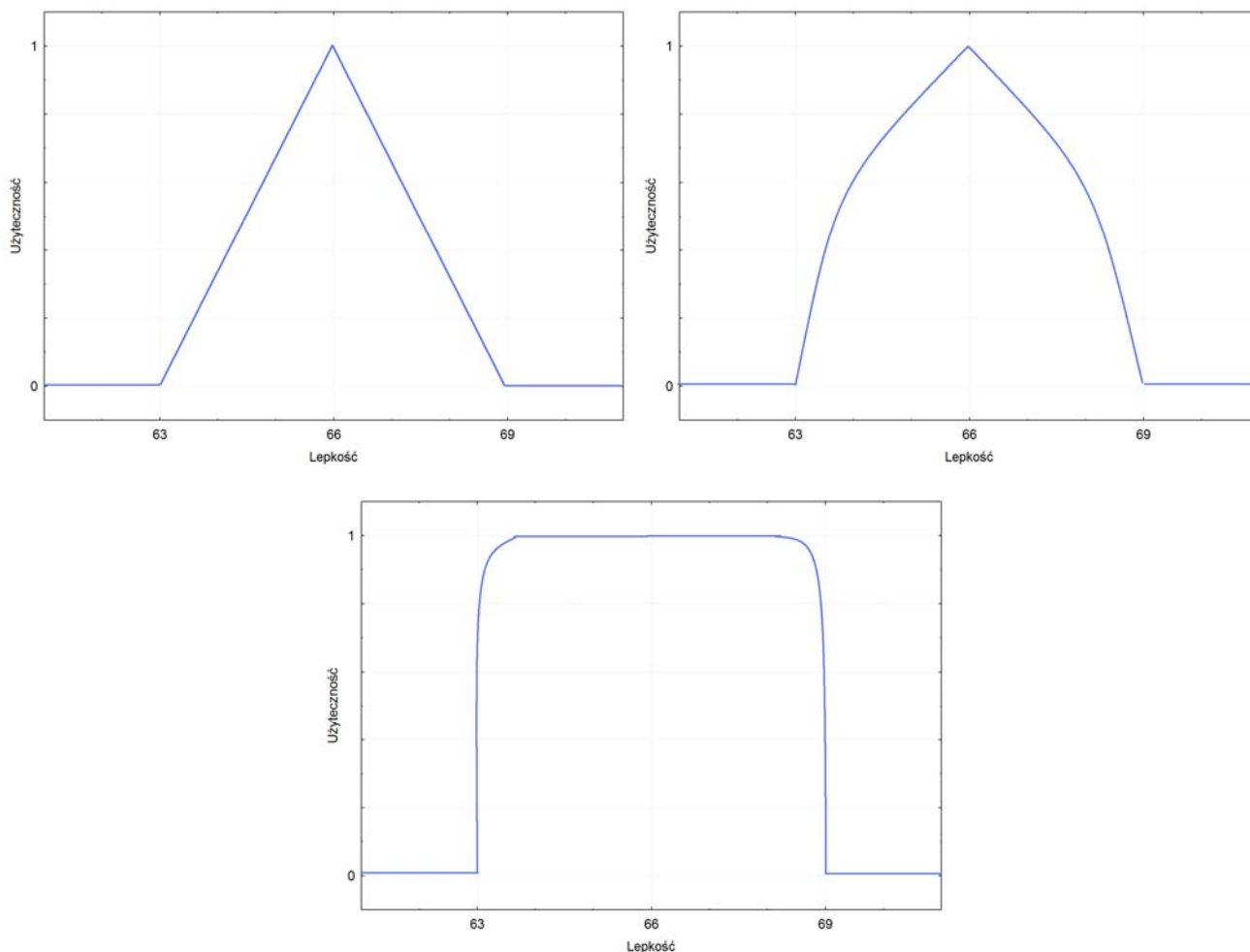


Rys. 3. Profil funkcji użyteczności dla wielkości wyjściowej Wydajność.



Dla zmiennej Wydajność definiujemy profil użyteczności odpowiedzi jak na rys. 3.

Poprzez odpowiednie zdefiniowanie profilu możemy określić także pożądany zakres wielkości optymalizowanych oraz to, jakie mamy preferencje na przykład na brzegu wybranego zakresu (por. rys. 4). Tak definiujemy profile użyteczności dla pozostałych zmiennych.

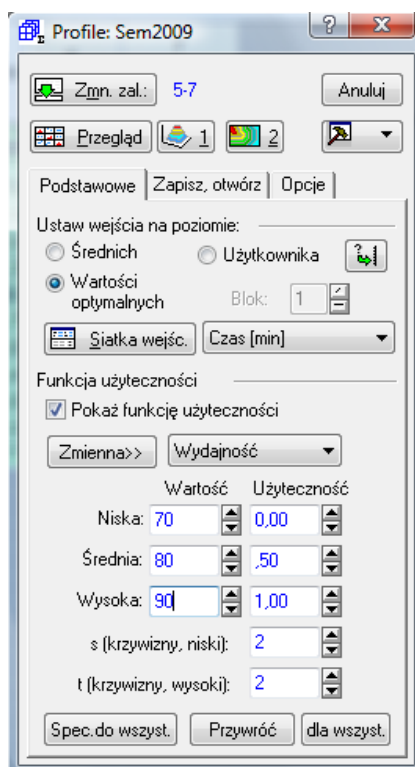


Rys. 4. Przykładowe profile użyteczności umożliwiające zdefiniowanie pożądanego zakresu wartości.

W programie *STATISTICA* profil użyteczności definiujemy za pomocą 8 parametrów: poziomów górnego, środkowego i dolnego oraz wartości funkcji użyteczności na tych odpowiednich poziomach oraz za pomocą dwóch parametrów określających krzywiznę funkcji użyteczności na odcinkach zdefiniowanych przez te punkty. Daje to dużą elastyczność i umożliwia zdefiniowanie wielu różnych profili w zależności od potrzeb.



Na przykład profil z rys. 3 zdefiniowano następująco:



Rys. 5. Definiowanie profilu użyteczności dla zmiennej Wydajność.

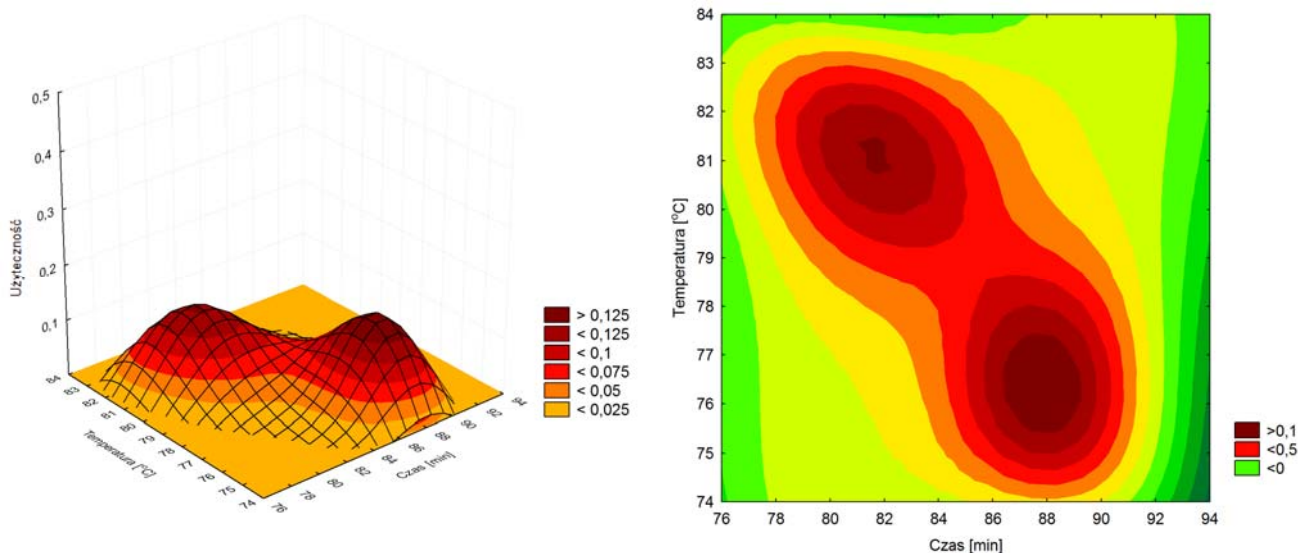
Po zdefiniowaniu funkcji użyteczności dla każdej ze zmiennych wyjściowych przechodzimy do procedury poszukującej wartości optymalnych. Dysponujemy modelami dla każdej z wielkości wyjściowych oraz profilami użyteczności.

1. Definiujemy siatkę dla każdej ze zmiennych wejściowych (zakres i liczbę wartości).
2. Kolejno w każdym punkcie siatki obliczamy wartość przewidywaną za pomocą modelu dla każdej z wielkości wyjściowych (domyślnie stosowany jest model kwadratowo-liniowy z interakcjami stopnia 2).
3. Dla każdej z wielkości wyjściowych, używając odpowiedniej funkcji użyteczności, uzyskujemy użyteczność (częstkową) z uzyskania takiej wartości zmiennej wyjściowej.
4. Następnie obliczymy wartość użyteczności całkowitej jako średnią geometryczną użyteczności częściowych z punktu 3. Wybrano średnią geometryczną (dla n liczb jest to pierwiastek n -tego stopnia z iloczynu tych n liczb), aby w sytuacji, gdy jedna z użyteczności częściowych była równa zero, również użyteczność całkowita powinna być równa zero (proces o bardzo dużej wydajności, ale niespełniający innych wymagań, jest całkowicie bezużyteczny). Zatem w każdym punkcie siatki uzyskujemy wartość użyteczności całkowitej.
5. Następnie budujemy model dla użyteczności całkowitej, wykorzystując jedną z dostępnych metod: najmniejszych kwadratów, funkcji sklepanych lub innych. Model ten ma przewidywać użyteczność całkowitą za pomocą wielkości wejściowych.



6. Poszukujemy rozwiązania optymalnego - wartości wielkości wejściowych, dla których całkowita użyteczność przyjmuje wartość największą. W tym modelu w programie *STATISTICA* zaimplementowane są dwie metody:
 - ◆ Pierwsza z to ogólna procedura optymalizacji funkcji (metoda sympleksowa) dla znalezienia optymalnych wartości wielkości wejściowych (wewnątrz podanego zakresu). Metoda ta jest zasadniczo szybsza, ale w przypadku danych, dla których funkcja całkowitej użyteczności ma wiele lokalnych minimów, albo też maksymalna użyteczność jest położona blisko brzegu obszaru badanego, zbieżność tej metody może zawieść, czyli znalezienie wartości optymalnych będzie niemożliwe.
 - ◆ Druga metoda przeszukuje wszystkie podane kombinacje wartości wielkości wejściowych w węzłach siatki zdefiniowane w punkcie 1, aby znaleźć taką, która daje w efekcie optymalną całkowitą użyteczność odpowiedzi.
7. Użyteczność całkowita w zależności od parametrów wejściowych (jeśli więcej niż dwa, to pozostałe parametry na ustalonych poziomach) jest ilustrowana na odpowiednich wykresach. Dostępne są wykresy: powierzchniowy (pary wielkości wejściowych są reprezentowane przez dwie osie, natomiast całkowita użyteczność odpowiedzi jest odkładana na osi trzeciej) i warstwiczny (wartości całkowitej użyteczności reprezentowane są przez obszary o różnym kolorze). Wykresy takie ilustrują dynamikę zmian wielkości wyjściowych ze względu na wybrane parametry oraz pozwalają ocenić, które zmienne są istotniejsze.

W naszym przykładzie zastosowano ogólną metodę optymalizacji funkcji. Otrzymaliśmy następujące optymalne parametry procesu produkcyjnego: czas równy 87,6 minut oraz temperaturę na poziomie 76,3 °C.



Rys. 6. Użyteczności dla zmiennej Wydajność.

Na rys. 6 widzimy wykres użyteczności całkowitej w zależności od temperatury i czasu. Wykres taki daje dużo informacji na temat zachowania procesu w pobliżu punktu optymalnego, dzięki czemu można ocenić stabilność procesu, czyli to, jaka będzie jakość produktu



przy zaburzeniu parametrów wejściowych. Widać, że w przestrzeni parametrów występują dwa obszary o większej użyteczności całkowitej, ale ta wybrana przez nas nie dość, że obejmuje większe wartości, to jest również większa, dzięki czemu proces powinien być stabilniejszy.

Podsumowanie

Zaprezentowane podejście daje nam nie tylko optymalne parametry produkcji, ale również zwiększa naszą wiedzę o badanym procesie produkcyjnym. Warto także zwrócić jeszcze raz uwagę na to, że możliwe jest poszukiwanie optimum ze względu na wiele cech produktu jednocześnie, co odzwierciedla typowe problemy praktyczne. Program *STATISTICA* i moduł do planowania doświadczeń i analizy ich wyników oraz poszukiwania rozwiązań optymalnych dostarczają potężnego narzędzia dla badań nad rozwojem produktów.

Prezentowany przykład dotyczył co prawda zagadnień z zakresu optymalizacji procesu produkcyjnego, ale może być również z powodzeniem wykorzystany w przypadku zagadnień pochodzących z innych dziedzin badań empirycznych.

Literatura

1. Montgomery D. C., *Design and Analysis of Experiments*, John Wiley & Sons, New York 2005.
2. Myers R. H., Montgomery D. C., Anderson-Cook C. M., *Response Surface Methodology: Process and Product Optimization Using Designed Experiments*, John Wiley & Sons, New York 2009.
3. Breyfogle III F. W., *Implementing Six Sigma*, John Wiley & Sons, New York 1999.
4. Derringer G., Suich R., Simultaneous optimization of several response variables, *Journal of Quality Technology* 12, vol. 4, 1980.
5. Harańczyk G., Woyna-Orlewicz K., Realizacja koncepcji Quality by Design, *Praktyczna Analiza Danych dla Przemysłu Farmaceutycznego*, StatSoft 2009.
6. Konkol J., Planowanie i analiza wyników badań laboratoryjnych betonów o zróżnicowanej strukturze, *Zastosowania statystyki i data mining*, StatSoft 2007.