

Analyses of Power Consumption of the Heat Pump Dryer in the Automobile Drying Process by using the Principal Component Analysis and Multiple Regression

Chang-Yong Lee* · Gensoo Song** · Jinho Kim*†

*Dept. of Industrial and Systems Engineering, Kongju National University
**R&D Center of Duksan A.C.M. Co. LTD.

주성분 분석과 다중회귀모형을 사용한 자동차 건조 공정의 히트펌프 건조기 소모 전력 분석

이창용* · 송근수** · 김진호*†

*공주대학교 산업시스템공학과
**덕산공조(주) 기업부설연구소

In this paper, we investigate how the power consumption of a heat pump dryer depends on various factors in the drying process by analyzing variables that affect the power consumption. Since there are in general many variables that affect the power consumption, for a feasible analysis, we utilize the principal component analysis to reduce the number of variables (or dimensionality) to two or three. We find that the first component is correlated positively to the entrance temperature of various devices such as compressor, expander, evaporator, and the second, negatively to condenser. We then model the power consumption as a multiple regression with two and/or three transformed variables of the selected principal components. We find that fitted value from the multiple regression explains 80~90% of the observed value of the power consumption. This results can be applied to a more elaborate control of the power consumption in the heat pump dryer.

Keywords : Heat Pump Dryer, Principal Component Analysis, Multiple Regression, Multivariate Data Analysis

1. 서론

건조 공정은 대부분의 산업 분야에서 요구되는 필수적인 공정으로 건조 공정에서는 일반적으로 열에너지를 이용하여 피 건조물을 건조시키기 때문에 많은 에너지가 소비 된다[2]. 따라서 에너지 효율을 높이기 위한 방안이 매우 중요하며, 히트펌프 건조기(heat pump dryer)는 기

존의 열풍식 건조기에 비해 약 80%까지 에너지 효율을 높일 수 있어 상당한 에너지 절약 효과가 있는 것으로 알려져 있다[8, 15]. 히트펌프 건조기가 에너지 효율이 높은 주된 이유는 히트펌프 사이클의 특성상 사용한 열을 회수할 수 있기 때문이다. 건조비용과 관련 있는 인자인 비제습률(Specific Moisture Extraction Rate, SMER)을 보면, 일반적인 열풍식 건조기는 약 0.5~1kg/kWh인데 비하여 히트펌프 건조기는 약 3kg/kWh 정도로 높은 효율을 가지고 있다[13].

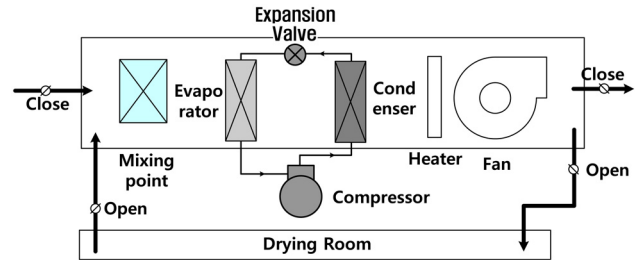
건조공정에서 히트펌프에 관한 기존 연구는 주로 히트펌프의 기계적인 성능 및 적용에 관한 연구가 주를 이

루고 있다. Bannister et al.[15]은 히트펌프의 효율적인 제습을 위하여 건조 환경, 시스템 효율성과 신뢰성, 그리고 사용 편의성 등을 통합적인 측면에서 고려할 수 있는 기술에 대한 연구를 수행하였으며, Lee et al.[13]은 히트펌프의 기계적인 건조 성능에 필요한 공기 흐름의 효과를 분석하였다. 응용측면을 보면, Wijesinghe et al.[8]은 히트펌프의 높은 에너지 효율 특성을 이용하여 농산물 건조에 적용하는 연구를 수행하였으며, Shon et al.[10, 14]은 한냉지 및 업무용 건물의 재생 가능 에너지인 지열을 이용한 히트펌프 시스템의 성능 예측을 시뮬레이션을 통해 수행하고 난방 성능 분석에 대한 연구를 수행하였다. 이러한 연구는 주로 히트펌프의 제어가 상대적으로 덜 중요한 농업 및 업무용에 적용한 연구이며, 히트펌프 건조기의 각 구성 요소의 최적 제어가 절대적으로 중요한 산업용 건조기에 적용하기 위한 연구는 아직 그 초기 단계에 있다. 즉, 열풍식 건조기에 비하여 에너지 효율이 높은 히트펌프를 생산 품질에 직접적인 영향을 미치는 산업용 건조공정에 적용하기 위해서는 히트펌프의 정밀한 제어가 필요하나 이에 대한 연구는 아직 미비한 실정이며, 본 논문은 이러한 측면에서 볼 때 기존 연구와 대비한 차별성이 있다.

일반적으로 히트펌프 건조기는 <Figure 1>과 같이 압축기, 응축기, 팽창변, 증발기 등 4개의 성분으로 구성되어 있는데, 히트펌프 건조기에서는 건조기 출구에서 나오는 일정 온도 이상의 다습한 공기는 외부로 배출되지 않고 증발기와 응축기로 순환되면서 습기를 제거하고 공기 중의 열에너지를 회수하여 다시 사용한다. 히트펌프 건조기는 높은 열효율을 가지는 장점이 있는 반면, 각 성분의 입 출구 냉매의 온도 및 유량 등에 따라 건조기의 출구 공기 온도 및 비제습률의 변화가 심하여 정밀한 온도 제어의 어려움이 있다. 이러한 이유로 히트펌프 건조기는 제어 온도 범위가 비교적 넓은 농산물 건조에 많이 이용되고 있고, 자동차 도장 건조와 같이 건조가 생산 품질에 직접적인 영향을 미치는 산업용 건조에는 아직 이용되고 있지 않은 실정이다.

히트펌프의 정밀한 온도 제어를 위해서는 히트펌프의 소모 전력에 영향을 미치는 변수들과 소모 전력 간의 관계 규명이 선행되어야 한다. 전형적인 히트펌프는 <Figure 1>과 같이 압축기, 응축기, 팽창변, 증발기 등 4가지 요소로 구성되어 있고, 4개의 구성 요소가 서로 상호 복잡하게 연관되어 있다. 히트펌프의 성능(즉, 소모 전력)에 영향을 주는 요인은 히트펌프의 작동 압력과 작동 온도이다. 작동 압력에 영향을 주는 변수는 히트펌프의 각 구성 요소를 통과하는 열 매개물질인 냉매의 유량, 응축기를 통과하는 물의 유량과 응축기를 통과하는 물의 입 출구 수온 등 4가지이고, 작동 온도에 영향을 주는 변수는

히트펌프의 4가지 구성 요소에 대한 입구 냉매 온도로 정의될 수 있다. 따라서 히트펌프의 소모 전력에 영향을 미치는 변수는 총 8개이다. 위의 변수들의 조건이 변경됨에 따라 히트펌프의 성능이 달라짐으로 본 연구에서는 소모 전력에 영향을 미치는 변수를 8가지로 정하였다. 따라서 히트펌프 건조기에 영향을 미치는 변수로는 압축기의 냉매 유량(v_1)과 입구 냉매 온도(v_2), 응축기의 유량(v_3)과 입출구 수온(v_4, v_5) 및 출구 냉매 온도(v_6), 팽창변 입구 냉매 온도(v_7), 증발기 입구 냉매 온도(v_8) 등 8개가 있다. <Figure 2>는 종속변수인 소모 전력과 8개의 독립변수 데이터의 일부를 나타낸 것이다.



<Figure 1> Schematic Diagram of a Heat Pump Dryer

건조기의 소모 전력을 분석하기 위해 가장 흔히 생각할 수 있는 방법은 다중회귀모형(multiple regression model)으로 종속변수(혹은 반응변수)인 소모 전력에 영향을 미치는 모든 독립변수(혹은 설명변수)를 모형에 포함시켜서 종속변수를 설명하는 것이다. 독립변수가 많을수록 회귀모형의 설명력인 결정계수 R^2 은 커지게 되어 회귀식의 설명력은 높아지지만, 각 독립변수의 p -값은 커져 개별 인자들이 유의하지 않을 수 있다. 또한 독립변수들 간에 높은 선형관계가 존재하는 경우에는 종속변수의 추정 값에 대한 분산이 상대적으로 커지게 되어 다중공선성(multicollinearity) 문제가 발생할 수 있다. 특히 독립변수들 간에 높은 선형관계는 회귀분석의 가정을 위배하는 것이므로 적절한 변수 선택이 중요한 문제로 부각된다.

power	v_1	v_2	v_3	v_4	v_5	v_6	v_7	v_8
53.700	1.689	16.200	2.868	39.400	43.100	25.500	12.400	5.500
66.100	1.728	17.800	2.901	39.400	44.200	25.300	13.200	1.000
74.000	1.778	18.900	2.893	39.400	44.000	25.200	14.700	-0.300
79.200	1.774	19.500	2.885	39.400	43.800	25.100	16.100	-0.800
83.100	1.762	19.900	2.894	39.400	43.700	25.000	17.300	-0.900
86.000	1.754	20.100	2.889	39.400	43.600	25.000	18.300	-0.800
88.300	1.747	20.100	2.889	39.400	43.600	24.900	19.100	-0.700
90.300	1.744	20.100	2.878	39.400	43.600	24.900	19.700	-0.500
92.000	1.732	20.100	2.902	39.400	43.700	24.900	20.100	-0.400
93.400	1.732	20.000	2.890	39.400	43.800	24.800	20.500	-0.300
94.500	1.744	19.700	2.888	39.400	43.800	24.700	20.600	-0.200
95.500	1.740	19.500	2.893	39.400	43.900	24.700	20.800	-0.300
96.200	1.744	19.300	2.886	39.400	44.000	24.700	20.800	-0.400
96.900	1.736	19.100	2.896	39.400	44.100	24.600	20.800	-0.500

<Figure 2> A Part of Data that are Analyzed in this Study

또한 많은 변수로 구성된 다변량(multi-variate) 데이터의 경우, 각 변수들이 상호 독립이면 각 변수에 대하여 단변량 분석을 할 수 있으나 변수들 간에 상관관계가 존재하는 경우에는 단변량 분석이 효율적이지 못하다. 특히 관측값에 영향을 미치는 모든 독립 변수를 고려하여 분석을 수행하는 경우에는 변수의 개수(혹은 차원, dimension)이 증가할수록 데이터의 분석 또는 모델 추정에 필요한 데이터 수가 기하급수적으로 증가하는 소위 ‘차원의 저주’(curse of dimensionality) 문제가 발생한다. 따라서 다변량 데이터 분석의 경우에는 변수들 간의 상관구조를 파악하여 차원을 축소시켜 분석을 수행하는 것이 일반적이다.

차원을 축소시키는 방법은 변수 선택(variable selection)과 변수 추출(variable extraction) 등 두 가지로 구분할 수 있다[17]. 변수 선택 방법은 전체 변수 집합에서 주요 변수로 구성된 부분집합을 선정하는 것으로, 원래 변수의 변환을 거치지 않는 장점은 있지만 주요 변수를 선택할 때 변수들 간의 상관관계를 고려하지 않기 때문에 변수들 간에 상관관계가 존재하는 경우에는 적합하지 않은 한계가 있다[4].

변수 추출 기법은 기존 변수들 간의 상관관계를 고려하여 새로운 변수들을 생성하는 방법이다[4]. 이 때 추출된 변수들은 고차원의 기존 변수들의 조합으로 정의되기 때문에 추출된 변수에 대한 해석이 어려울 수도 있지만, 기존 변수들의 상관관계를 고려할 수 있다는 장점이 있음으로 추출된 변수의 해석이 비교적 용이한 경우에 널리 사용할 수 있는 방법이다. 변수 추출은 반응변수의 존재 유무에 따라 지도 변수 추출(supervised variable extraction)과 비지도 변수 추출(unsupervised variable extraction)로 나눌 수 있다[1, 7]. 지도 변수 추출의 대표적인 기법으로 변수의 개수에 비하여 데이터의 개수가 적은 경우에 주로 적용하는 방법인 PLS 회귀(partial least square regression)[12]를 들 수 있으며, 비지도 변수 추출 기법의 대표적인 기법으로 주성분 분석(PCA, principal component analysis)[5, 6, 11]이 있다.

본 연구에서는 히트펌프 건조기의 높은 에너지 효율의 장점을 유지하면서 건조기 제어의 어려움을 극복하기 위하여, 건조기의 소모 전력에 영향을 미치는 주요한 변수들을 규명하고 소모 전력과의 관계를 분석하여 자동차 도장 공정에 히트펌프 건조기를 적용할 수 있는 기술 개발의 초석을 제공하고자 한다. 이를 위하여 소모 전력에 영향을 미치는 변수가 8개로 비교적 많음으로 각 변수들 간의 상관관계를 주성분 분석을 사용하여 차원을 축소하고, 축소된 차원을 다중회귀모형에 적용하여 변수들이 전력량에 미치는 영향을 분석하고자 한다.

본 논문은 다음과 같이 구성되어 있다. 제 2장에서는 주성분 분석을 사용한 차원 축소 및 이에 따른 변수간의

특성, 그리고 주성분 분석 사용의 정당성을 논하였고, 제 3장에서는 축소된 차원을 사용하여 다중회귀모형을 적용한 결과를 설명하였으며, 마지막 장에서는 본 논문의 결론을 맺었다.

2. 주성분 분석을 통한 차원 축소

2.1 고유치 문제(Eigenvalue Problem)와 구형 검정(Sphericity Test)

주성분 분석은 고차원(혹은 다변량) 데이터의 정보를 최대한 유지하면서 차원을 축소하여 저차원상에서 변수들의 관계를 규명하는 다변량 데이터 처리 방법으로 피어슨(Karl Pearson)과 호텔링(Herold Hotelling)에 의하여 제안되었다. 주성분 분석에서는 상관계수(혹은 공분산) 행렬로 표현되는 변수들 간의 변동을 최대한 설명하는 적은 개수(보통 2~3개)의 주성분을 원래 변수들의 선형 결합인 직교 변환(orthogonal transformation)을 통해 구한다. 즉, p 개의 변수로 구성된 데이터들의 변동을 최대한 설명하기 위하여 p 차원 공간의 축을 회전시켜 주축(principal axes)을 구하고, p 차원 데이터를 $m(m < p)$ 개의 주성분 축 방향으로 사영(projection)시킴으로써 차원을 축소한다. 따라서 주성분 분석의 핵심은 다변량 데이터에서 변동을 원래 변수들의 선형결합들로 이루어진 새로운 변수를 사용하여 표현하는 것이다. 또한 주성분 분석에서 변수 간에 값의 차이가 클 때에는 단위가 큰 변수의 값에 의하여 데이터 분석 및 해석이 좌우됨으로 일반적으로 정규화(standardization)를 사용하여 각 변수에 대한 데이터가 평균이 0이고 분산이 1이 되도록 한다. 따라서 주성분 분석에서는 공분산 행렬보다 상관계수 행렬을 주로 사용한다.

p 개의 독립변수들로 구성된 n 개의 다변량 데이터를 $n \times p$ 행렬인 $X = \{x_{ij}\} (i = 1, 2, \dots, n, j = 1, 2, \dots, p)$ 로 표현하고, 독립변수들의 선형 결합을 표현하기 위하여 단위 벡터(unit vector) $\vec{u}_i (i = 1, 2, \dots, p)$ 를 사용하면, i 번째 선형 결합 벡터 \vec{z}_i 는

$$\vec{z}_i = X\vec{u}_i, \quad i = 1, 2, \dots, p \quad (1)$$

로 표현할 수 있으며, 여기서 \vec{z}_i 는 $n \times 1$ 벡터이고, \vec{u}_i 는 $p \times 1$ 벡터이다. 위의 식 (1)을 행렬로 표현하면

$$Z = XU \quad (2)$$

가 되며, 여기서 $Z = \{\vec{z}_1, \vec{z}_2, \dots, \vec{z}_p\}$ 이며 $U = \{\vec{u}_1, \vec{u}_2, \dots, \vec{u}_p\}$ 이다.

주성분 분석은 고차원 데이터의 분산(즉, 변동)을 최대한 유지하면서 차원을 축소시키는 것임으로 Z 의 분산이 최대가 되도록 U 를 결정한다. 식 (2)에 의하여 Z 의 분산은

$$var(Z) = U^T \Sigma U, \quad \Sigma \equiv \frac{1}{n-1} X^T X \quad (3)$$

로 주어진다. 또한 상관계수 행렬 Σ 는 정규화를 고려하면

$$\Sigma_{jk} = \frac{\frac{1}{n-1} \sum_{i=1}^n (x_{ij} - \bar{x}_j)(x_{ik} - \bar{x}_k)^T}{s_j s_k}, \quad j \cdot k = 1, 2, \dots, p \quad (4)$$

로 표현되며, 여기서 $\bar{x}_j = \frac{1}{n} \sum_{i=1}^n x_{ij}$ 으로 표본 평균, $s_j = \frac{1}{n-1} \sum_{i=1}^n (x_{ij} - \bar{x}_j)^2$ 로 표본 분산에 해당한다. 주어진 Σ 에 대하여 $var(Z)$ 가 최대가 되는 U 를 결정하는 것은 고유치 문제(eigenvalue problem)으로 귀착되며

$$\Sigma U = \lambda U \text{ 혹은 } (\Sigma - \lambda I) U = 0 \quad (5)$$

으로 표현할 수 있다. 따라서 식 (5)를 풀면 p 개의 고유치(eigenvalue)와 고유벡터(eigenvector) λ_k 과 $\vec{u}_k (k=1, 2, \dots, p)$ 를 구할 수 있다.

소모 전력에 영향을 미치는 8개 변수로 구성된 데이터 <Figure 2>에 대하여 식 (4)와 식 (5)를 사용하여 고유치 문제를 푼 결과를 <Table 1>에 나타내었다. <Table 1>에서 각 고유치는 각 주성분이 설명할 수 있는 분산의 크기를 나타냄으로 $\sum_{k=1}^8 \lambda_k = 8$ 이 되고, 특히 첫 번째와 두 번째 주성분은 각각 전체 분산의 약 50%와 30%를 설명함을 알 수 있다.

주성분 분석은 변수들 간에 높은 상관관계가 존재함을 가정하고 수행하는 것임으로 만약 변수들이 서로 독립적이라면 주성분 분석은 적절하지 않다. 따라서 변수들이 충분히 높은 상관관계를 가지고 있음을 검정할 필요가 있으며, 이를 위하여 Bartlett에 의하여 제안된 구형검정(sphericity test)[18]이 일반적으로 사용된다. 이 검정은

일종의 카이제곱 검정(X^2 -test)으로 검정통계량(test statistics)는 변수들 간에 상관관계가 존재하지 않는다는 귀무가설 하에 상관계수 행렬의 행렬식(determinant)을 사용하여

$$X^2 \left[\frac{p(p-1)}{2} \right] = - \left[(n-1) - \frac{(2p+5)}{6} \right] \ln[\Sigma] \quad (6)$$

로 주어진다. 여기서 p 와 n 은 각각 변수와 데이터의 개수를 나타내며, $|\Sigma|$ 는 상관계수 행렬 Σ 의 행렬식을 나타내고, $\frac{p(p-1)}{2}$ 는 자유도이다. 구형 검정은 데이터가 추출된 모집단의 상관계수 행렬은 단위행렬(identity matrix)이라는 귀무가설 하에 수행하는 일종의 적합도(goodness-of-fit) 검정이다. 따라서 상관계수 행렬의 행렬식이 1에 가까우면 Bartlett의 검정통계량(식 (6))은 0에 가까게 되어 귀무가설을 채택하게 된다. 그와 대조적으로 변수들 간에 상관계수가 높은 경우에는 행렬식이 0에 근접하게 되고 따라서 검정통계량은 양의 큰 값을 가지게 되어 귀무가설을 기각하게 된다.

실험에 사용한 데이터 개수 $n = 3,153$ 을 사용하여 $p = 8$ 에서 Bartlett's 검정통계량을 계산하였으며, 그 결과 $-[(n-1) - \frac{(2p+5)}{6}] \ln[\Sigma] \approx 36,163$ 을 얻었다. 자유도가 $\frac{p(p-1)}{2} = 28$ 이고 95% 유의수준(significance level)에서 임계값(critical value)는 카이제곱 분포 표에서 $X^2(28) = 41.377$ 임으로 검정통계량 값이 임계값보다 크기 때문에 귀무가설을 기각할 수 있다. 따라서 이 결과는 변수들 간에 상관관계가 높아서 주성분 분석을 통한 차원 축소가 유의함을 나타낸다.

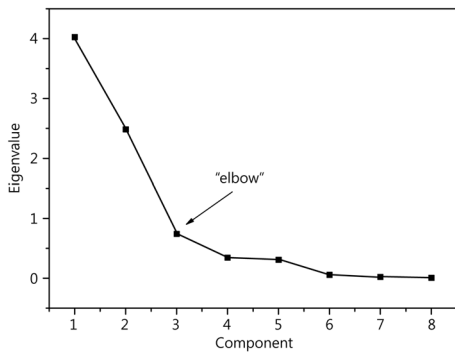
2.2 차원 축소와 주성분 적재값(Principal Component Loading)

고차원 데이터의 차원을 축소하기 위해서는 축소할 차원 수만큼의 주성분을 선택해야 한다. 주성분 선택은

<Table 1> The Eigenvalues λ_k and Corresponding Eigenvectors \vec{u}_k Obtained by Applying PCA to the Data

k	1	2	3	4	5	6	7	8
\vec{u}_k	0.14	-0.48	0.49	-0.54	0.46	-0.01	0.06	0.05
	0.46	0.14	0.21	0.30	0.24	-0.40	-0.24	-0.60
	0.36	0.24	-0.45	-0.72	-0.20	-0.17	-0.02	-0.14
	0.02	-0.59	-0.38	0.18	-0.06	-0.61	0.30	0.09
	0.25	-0.49	-0.43	0.12	0.10	0.53	-0.43	-0.15
	0.36	-0.25	0.42	0.03	-0.79	0.07	-0.04	0.03
	0.47	0.18	-0.01	0.15	0.18	-0.16	-0.30	0.76
	0.48	0.10	-0.07	0.17	0.15	0.36	0.76	-0.02
λ_k	4.027	2.484	0.743	0.346	0.309	0.058	0.024	0.009

보통 경험 법칙(rule of thumb)을 사용하여 고유치 중에서 큰 값 몇 개를 선택한다. 큰 고유치 개수를 선택하는 방법은 여러 가지가 있는데, 전체 변이에의 공헌도를 기준으로 전체 변이의 약 70~90%가 되도록 주성분수를 결정하는 방법, 고유치가 평균 이상이 되는 주성분을 설정하는 방법, 그리고 고유치 크기순서와 고유치를 축으로 하는 2차원 좌표축에 고유치를 표현하여 고유치 개수를 정하는 스크리 플롯(scree plot) 방법[16] 등이 있다. 이 중에서 가장 널리 사용되는 스크리 플롯은 도표를 사용하는 방법으로 고유치를 내림차순으로 나열하여 각 고유치를 나타낸 다음 고유치를 연결한 직선에서 “팔꿈치”(elbow) 부분에 해당하는 영역을 고려하여 팔꿈치 위부분에 속한 고유치만 주성분으로 고려하는 것을 말한다. <Figure 3>은 위의 제 2.1절에서 구한 고유치에 대하여 스크리 플롯을 나타낸 것이다. <Figure 3>에서 볼 수 있듯이 3번째로 큰 고유치에서 현저한 굴곡이 나타나고, 그 이후는 기울기가 완만해 짐을 알 수 있다. 따라서 가장 큰 2(혹은 3)개의 주성분을 선택함이 적당함을 경험적으로 알 수 있다. 이것은 원래 8개의 변수를 2(혹은 3)개의 주성분으로 차원을 축소할 수 있음을 의미하며, 2개의 주성분으로 차원을 축소하는 경우에는 \vec{z}_1 과 \vec{z}_2 가 새로운 변수가 된다.



<Figure 3> The Scree Plot by Using Eigenvalues. The Arrow Indicates the “Elbow”

고유치 문제에서 고유치와 고유벡터가 결정되면, 식 (2)를 사용하여 선형 변환된 행렬 Z 를 구할 수 있으며, 이를 이용하여 행렬 Z 와 원 데이터 행렬 X 사이의 상관 정도를 계산할 수 있는데, 이것을 주성분 적재값(principal component loading)이라 한다[9]. 주성분 적재값 F 는 주성분을 통하여 설명할 수 있는 원래 독립변수의 값들을 양적으로 표현한 것으로 고유벡터로 구성된 행렬 $U = \{\vec{u}_1, \vec{u}_2, \dots, \vec{u}_p\}$ 를 사용하여 다음과 같이 표현할 수 있다.

$$F = Corr(X, Z) = UD^{\frac{1}{2}}, D \equiv U^T \Sigma U. \quad (7)$$

<Table 2>는 위의 식 (7)과 <Table 1>의 고유벡터를 사용하여 구한 주성분 적재값을 나타낸 것으로, $\vec{x}_i (i = 1, 2, \dots, p)$ 는 i 번째 원래 변수의 데이터로 구성된 $n \times 1$ 벡터이며, $X = \{\vec{x}_1, \vec{x}_2, \dots, \vec{x}_p\}$ 이다.

<Table 2>의 주성분 적재값 결과와 변수를 통해 볼 때, 차원 축소를 위해 선택된 첫 번째 주성분 \vec{z}_1 은 주로 입구 온도(압축기 v_2 , 팽창변 v_7 , 증발기 v_8) 그리고 응축기의 유량(v_3) 및 출구 냉매 온도(v_6)와 높은 양의 상관관계를 가지고 있으며, 두 번째 주성분 \vec{z}_2 는 주로 응축기의 입출구 수온(v_4, v_5) 및 압축기의 냉매 유량(v_1)과 높은 음의 상관관계를 가지고 있음을 알 수 있다. 또한 두 개의 주성분 모두에 대하여 높은 상관관계를 가지는 변수는 없으며, 첫 번째 주성분은 양의 상관관계를 두 번째 주성분은 주로 음의 상관관계를 나타낸다.

주성분 적재값의 결과를 사용하면 차원 축소를 위해 선택한 주성분과 원래 변수 사이의 상관관계를 양적으로 표현할 수 있다. 즉, 차원 축소를 위하여 선택된 주성분이 원래 데이터를 어느 정도 설명하고 있는 가를 나타내는 주성분 적재값을 통하여 분석할 수 있다. 예를 들어 주성분 적재값 결과 <Table 2>를 사용하여 압축기의 입구 냉매 온도(v_2)인 \vec{x}_2 과 첫 번째 주성분 벡터 \vec{z}_1 사이의 상관

<Table 2> The Result of the Principal Component Loading

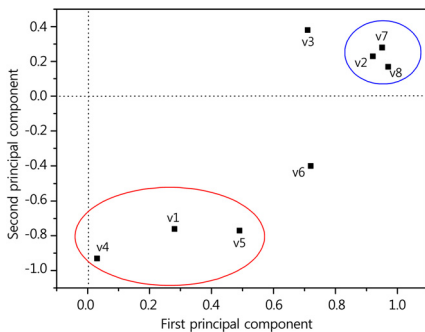
$Corr(\vec{x}_i, \vec{z}_j)$	\vec{z}_1	\vec{z}_2	\vec{z}_3	\vec{z}_4	\vec{z}_5	\vec{z}_6	\vec{z}_7	\vec{z}_8
\vec{x}_1	0.28	-0.76	0.42	-0.32	0.25	0.00	0.01	0.01
\vec{x}_2	0.92	0.23	0.18	0.18	0.13	-0.10	-0.04	-0.06
\vec{x}_3	0.71	0.38	-0.39	-0.42	-0.11	-0.04	0.00	-0.01
\vec{x}_4	0.03	-0.93	-0.32	0.11	-0.04	-0.15	0.05	0.01
\vec{x}_5	0.49	-0.77	-0.37	0.07	0.06	0.13	-0.07	-0.01
\vec{x}_6	0.72	-0.40	0.36	0.02	-0.44	0.02	-0.01	0.00
\vec{x}_7	0.95	0.28	-0.01	0.09	0.10	-0.04	-0.05	0.07
\vec{x}_8	0.97	0.17	-0.06	0.10	0.08	0.09	0.12	0.00
$Ave(R^2_j)$	0.503	0.313	0.092	0.043	0.038	0.008	0.003	0.001

관계를 보면, \vec{x}_2 의 분산 중 주성분 \vec{z}_1 을 통해 설명할 수 있는 부분은 $R_{8,1}^2 = Corr(\vec{x}_8, \vec{z}_1) = 0.92^2 = 0.85$ 가 되어 약 85%의 압축기의 입구 냉매 온도의 분산을 주성분 \vec{z}_1 을 통해 설명할 수 있음을 알 수 있다. 또한 \vec{z}_1 들은 서로 독립임으로 두 주성분 \vec{z}_1 과 \vec{z}_2 는 압축기의 입구 냉매 온도 (v_2)의 분산을 $R^2 = 0.92^2 + 0.23^2 = 0.90$, 즉 약 90%를 설명하고 있다. 또한 모든 변수에 대하여 주성분 j 의 평균 설명 정도인 $Ave(R^2)$ 는

$$Ave(R^2) = \frac{1}{p} \sum_{i=1}^p [Corr(\vec{x}_i, \vec{z}_j)]^2 \quad (8)$$

과 같이 정의하면, 첫 번째와 두 번째 주성분은 각각 약 50%와 30%를 설명하는 것을 알 수 있다. 또한 각 주성분은 서로 독립임으로 두 주성분을 합치면 모든 변수의 약 80%를 설명함을 알 수 있다.

선택된 2개의 주성분(\vec{z}_1 과 \vec{z}_2)을 축으로 하여 8개의 원래 변수(v_1-v_8)에 대하여 주성분 적재값 결과를 시각화할 수 있는데, 이를 통하여 변수들 간의 유사성을 파악할 수 있다. <Table 2>의 결과를 사용하여 2개의 주성분에 대한 주성분 적재값 결과를 <Figure 4>에 나타내었다. <Figure 4>에서 각 점은 원래 변수와 두 주성분 간의 상관관계수를 나타내며 가까운 점들끼리는 더 유사함을 의미한다. 예를 들어, 압축기 입구 냉매 온도(v_2)는 같은 압축기의 냉매 유량(v_1)에 비하여 증발기 입구 냉매 온도 (v_8)와 더 유사한 프로파일(profile)을 가지고 있음을 알 수 있다. 또한 v_2, v_7, v_8 는 각각 압축기 입구 냉매 온도, 팽창변 입구 냉매 온도, 증발기 입구 냉매 온도를 나타냄으로 ‘입구 냉매 온도’들은 유사한 프로파일을 가지고 있는 같은 그룹으로 표현할 수 있으며, v_1, v_4, v_5 는 각각 압축기 냉매 유량, 응축기 입구 수온, 응축기 출구 수온을 나타냄으로 압축기 냉매 유량은 응축기의 수온과 유사한 프로파일을 가지고 있음을 알 수 있다.



<Figure 4> Plot of the Principal Component Loading by Using the Two Principal Components

3. 다중회귀모형을 적용한 소모 전력 분석

주성분 분석을 통하여 8개의 변수를 2개의 주성분(\vec{z}_1 과 \vec{z}_2)으로 차원을 축소시킬 수 있음을 보여주므로, 이를 이용하여 소모 전력을 2개의 변환된 변수(\vec{z}_1 과 \vec{z}_2)에 대하여 다중회귀 분석을 수행할 수 있다. 즉, 식 (1)을 사용하면 변환된 변수 \vec{z}_1 과 $\vec{z}_2 = X\vec{u}_i (i=1, 2)$ 의 값을 구할 수 있으므로 이를 이용하여 소모 전력에 대하여 2개의 새로운 변수로 구성된 다중회귀모형을 적용할 수 있다. 또한 2개의 변수는 선형 종속이 아니고 서로 독립임으로 다중공선성 문제는 발생하지 않는다. 2개의 변환된 변수를 사용한 다중 회귀 모형은

$$Y_k = \beta_0 + \beta_1 Z_{1k} + \beta_2 Z_{2k} + \epsilon_k, k = 1, 2, \dots, n \quad (9)$$

로 주어지며, 여기서 $\epsilon_k \sim N(0, \sigma^2)$ 를 따른다.

위의 다중회귀모형을 소모 전력 데이터 <Figure 2>에 적용하였으며 그 결과를 <Table 3>에 나타내었다. <Table 3>에서 추정된 회귀식은

$$\hat{y} = 146.995 + 1.278z_1 + 0.553z_2 \quad (10)$$

로 표현할 수 있으며, 모형으로부터 추정된 회귀선의 표본관측점에 대한 적합도를 정량적으로 나타내는 척도인 결정계수(coefficient of determination) R^2 값은 0.8061로 추정된 회귀선이 Y 의 변동을 약 80% 정도 설명함을 알 수 있다. 새로운 변수가 추가되면 오차 제곱은 작아지기 때문에 결정계수가 커지는 문제를 보완하기 위하여 제안된 수정된 결정계수(adjusted coefficient of determination) [3]는 $R_{adj}^2 = 0.806$ 으로 결정계수 R^2 와 거의 같음을 알 수 있다. 즉, 변수가 추가되더라도 오차의 제곱은 차이가 없음을 의미하는데, 이것은 두 변수 Z_1 과 Z_2 가 서로 독립이기 때문이다.

다중회귀모형에 대한 추론은 전형적으로 단계별로 실행한다. 즉, 모든 설명변수가 집합적으로 반응변수에 영향을 미치는가에 대한 가설 검정을 먼저 수행하고, 그 다음 단계로 각 회귀계수(regression coefficient)의 유의성을 다른 변수들을 조정하면서 검정한다. 첫 번째 단계로 모든 설명변수 Z_1 과 Z_2 의 반응변수에 집합적 영향에 대한 가설 검정을 귀무가설

$$H_0: \beta_1 = \beta_2 = 0$$

을 사용하여 수행하였다. <Table 3>을 통해 볼 수 있듯이 F 통계량은 $F_0 = 6547$ 이며, 이 때 유의확률값(p -값) $< 2.2 \times 10^{-16}$ 임으로 변수 Z_1 과 Z_2 가 소모 전력에 집합적으로

<Table 3> The Result of Multiple Regression Analysis with Two Principal Components. Here, DF and (***) Respectively Stand for the Degree of Freedom and the Test Being Significant with the Significance Level of 0.001

Parameter	Estimate	Std. Error	t-value	Pr (> t)
β_0	146.995	3.726	39.45	$< 2 \times 10^{-16} (***)$
β_1	1.278	0.071	17.94	$< 2 \times 10^{-16} (***)$
β_2	0.553	0.021	26.20	$< 2 \times 10^{-16} (***)$
$R^2 = 0.8061$ $R^2_{adj} = 0.8060$ $F_0 = 6547$ $DF(2,3150)$ $p\text{-값} < 2.2 \times 10^{-16}$				

<Table 4> The Result of Multiple Regression Analysis with Three Principal Components. Here, DF and (***) Respectively Stand for the Degree of Freedom and the Test Being Significant with the Significance Level of 0.001

Parameter	Estimate	Std. Error	t-value	Pr (> t)
β_0	-112.621	4.583	-24.57	$< 2 \times 10^{-16} (***)$
β_1	-0.588	0.054	-10.93	$< 2 \times 10^{-16} (***)$
β_2	2.163	0.0278	77.90	$< 2 \times 10^{-16} (***)$
β_3	-4.888	0.073	-66.53	$< 2 \times 10^{-16} (***)$
$R^2 = 0.8061$ $R^2_{adj} = 0.8060$ $F_0 = 6547$ $DF(2,3150)$ $p\text{-값} < 2.2 \times 10^{-16}$				

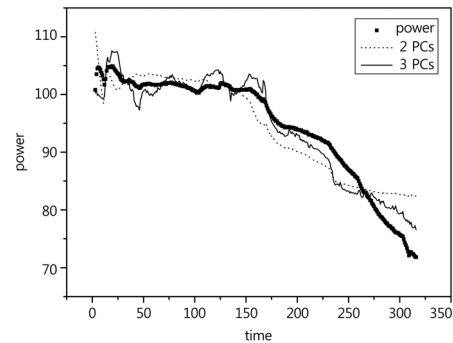
영향을 미치지 않는다는 귀무가설은 기각됨을 알 수 있다. 다음 단계로 각 설명변수에 대한 회귀계수의 유의성에 대한 검정을 보면, Z_2 의 효과를 감안하더라도 Z_1 은 $p\text{-값} < 2 \times 10^{-16}$ 로 유의하며, Z_2 의 경우도 마찬가지로임을 알 수 있다.

주성분 추가에 대한 효과를 분석하기 위하여 3개의 주성분 (z_1, z_2, z_3)을 사용하여 소모 전력에 대한 다중회귀 분석을 수행하였으며, 그 결과를 <Table 4>에 나타내었다. 이 때 추정된 회귀식은

$$\hat{y} = -112.621 - 0.588z_1 + 2.163z_2 - 4.888z_3 \quad (11)$$

으로 주어지며, 결정계수 R^2 과 R^2_{adj} 는 약 91%를 얻었다. 이것은 주성분 1개를 추가함으로써 인해 모형으로부터 추정된 회귀선의 표본관측점에 대한 적합도가 약 10% 증가됨을 의미한다. 그 외의 특성은 변수가 2개인 경우와 유사하였다.

다중회귀모형으로 적합한 회귀식에 의해 추정된 소모 전력(식 (10)과 식 (11))과 실제 소모 전력 데이터를 <Figure 5>에 비교하였다. <Figure 5>를 통해 볼 때, 회귀식들은 실제 소모 전력 데이터를 대체적으로 설명하고 있음을 알 수 있고, 3개의 주성분을 사용한 회귀식이 2개를 사용한 경우에 비하여 실제 데이터에 더욱 적합함을 알 수 있다. 이 결과를 통해 볼 때 소모 전력은 2~3개의 주성분을 구성하는 변수들에 의해 80~90% 정도의 제어가 가능함을 알 수 있다.



<Figure 5> Plots of the Power Consumption(Filled Boxes) and the Multiple Regression Results Using Two (Dotted Line) and Three(Solid Line) Principal Components

4. 결 론

본 연구에서는 자동차 건조 공정을 위한 히트펌프 건조기의 정밀 제어를 위한 기초 자료를 위해 히트펌프의 소모 전력에 영향을 미치는 변수들과 소모 전력 간의 관계를 규명하였다. 특히 소모 전력에 영향을 미치는 변수가 8개로 모든 변수를 사용하여 분석하기에 어려움이 있으므로 주성분 분석을 통하여 변수들 간의 상관관계를 분석하여 주성분을 결정하고, 이 결과를 다중회귀모형을 사용하여 주성분이 소모 전력에 미치는 영향을 분석하였다. 히트펌프의 구성 요소가 소모 전력에 미치는 영향은 그

동안 체계적인 분석이 이루어지지 않아서 주로 사용자의 경험을 통해서 제어를 수행해 왔는데, 는 영향은 그 동안 체계적인 분석이 이루어지지 않아서 주로 사용자의 경험을 통해서 제어를 수행해 왔는데, 본 연구에서는 주성분 분석과 다중회귀모형을 사용하여 히트펌프의 성능인 소모 전력에 영향을 미치는 여러 변수들 중에서 전력소모량에 가장 영향을 미치는 변수 혹은 변수들의 조합을 체계적으로 분석하였다. 이러한 분석 결과는 향 후 히트펌프 제어 장치인 PLC(Programmable Logic Control)에 적용하여 성능을 평가하는 연구에 적용할 수 있다.

주성분 분석을 통하여 8개의 변수에 대한 고유치와 고유벡터는 스크리 플롯을 사용하여 2~3개의 선형 변환된 주성분으로 차원을 축소시킬 수 있었으며, 선택된 첫 번째 주성분은 주로 입구 온도(팽창변, 증발기, 압축기)와 높은 양의 상관관계를 가지고 있으며, 두 번째 주성분은 주로 응축기와 높은 음의 상관관계를 가지고 있었다. 또한 선택된 주성분은 원래 변수들을 약 80% 설명함을 알 수 있었다. 이러한 결과를 바탕으로 다중회귀모형을 사용하여 소모 전력과 선형 변환된 변수들 간의 관계를 규명하였다. 적합된 회귀식은 실제 소모 전력 데이터를 대체적으로 잘 설명하고 있음을 알 수 있었는데, 2개의 주성분을 사용한 경우에는 90%의 소모 전력을, 3개인 경우에는 약 90%를 설명할 수 있음을 보였다. 세 번째 주성분을 포함시킴으로 소모 전력을 약 10% 추가적으로 설명할 수 있었는데, 이 주성분은 주로 응축기와 관련 있음을 알 수 있었다. 이러한 결과는 2~3개의 주성분을 구성하는 변수들에 의해 소모 전력 80~90% 정도의 제어가 가능함을 입증하는 것이라 할 수 있다.

히트펌프 건조기 소모 전력 데이터에 대한 분석 결과는 히트펌프를 사용한 건조 공정에 필요한 정밀한 제어를 통한 효율적 에너지 관리와 절감을 위한 선행 자료로 그 가치가 있다. 이러한 측면에서 볼 때 본 연구의 결과는 건조기 사용에 선행되어야 할 정밀 제어를 위한 모형 구축에 활용될 수 있다. 또한 본 연구에서 사용한 방법 외에 다양한 방법(예를 들어 ARMA 모형을 사용한 소모 전력 시계열 데이터 분석)들의 적용은 향 후 히트펌프 건조기의 소모 전력 변화를 이해하고 에너지 수요 예측 모형을 구축하는 데 적용될 수 있을 것으로 예상된다.

Acknowledgement

This work was supported by the Power Generation and Electricity Delivery Core Technology Program of the Korea Institute of Energy Technology Evaluation and Planning (KETEP) granted financial resource from the Ministry of Trade, Industry and Energy, Republic of Korea(No. 20131020400850).

References

- [1] Bannister, P., Carrington, G., and Chen, G., *Heat Pump Dehumidifier Drying Technology-Status Potential*. Proc. of 7th IEA Heat Pump Conference, 2002, Vol. 1, pp. 219-230.
- [2] Barlett, M., Test of Significance of Factor Analysis. *British Journal of Psychology*, 1950, Vol. 3, pp. 77-85.
- [3] Guyon, I. and Elisseeff, A., An introduction to variable and feature selection. *The Journal of Machine Learning Research*, 2003, Vol. 3, pp. 1157-1182.
- [4] Hotelling, H., Analysis of a complex of statistical variables into principal components. *Journal of Educational Psychology*, 1933, Vol. 24, pp. 417-441, pp. 498-520.
- [5] Hotelling, H., Relations between two sets of variates. *Biometrika*, 1936, Vol. 27, pp. 321-77.
- [6] Kim, S.B., Feature Extraction/Selection in High-Dimensional Spectral Data. In J. Wang (Ed.), *Encyclopedia of Data Warehousing and Mining*, Second Edition, Hershey, PA : Information Science Reference, 2009.
- [7] Lattin, J., Carroll, J., and Green P., *Analyzing multivariate data*, Thomson Learning, Pacific Grove, 2003.
- [8] Lee, K.H., Kim, O.J., and Lee, S.R., Analysis on the Drying Performance with the Flow Rate of Circulation Air in a Heat Pump Dryer. *Korean Journal of Air-Conditioning and Refrigeration Engineering*, 2009, Vol. 21, No. 1, pp. 1-8.
- [9] Mao, K.Z., Identifying critical variables of principal components for unsupervised feature selection, Systems, Man, and Cybernetics, Part B : Cybernetics. *IEEE Transactions*, 2005, Vol. 35, pp. 339-344.
- [10] Mujumdar, A.S., *Handbook of Industrial Drying*, Marcel Dekker. Inc., New York, 1995.
- [11] Pearson, K., On Lines and Planes of Closest Fit to Systems of Points in Space. *Philosophical Magazine*, 1901, Vol. 2, pp. 559-572.
- [12] Sohn, B.H. and Kwon, H.S., Performance Prediction on the Application of a Ground-Source Heat Pump System in an Office Building. *Korean Journal of Air-Conditioning and Refrigeration Engineering*, 2014, Vol. 26, No. 9, pp. 409-415.
- [13] Sohn, B.H., Choi, J.H., and Min, K.C., Heating Performance of Geothermal Heat Pump System Applied in Cold Climate Region. *Korean Journal of Air-Conditioning and Refrigeration Engineering*, 2015, Vol. 27, No. 1, pp. 031-038.

- [14] Tenenhaus, M., Esposito Vinzi, V., Chatelinc, Y.-M., and Lauro, C., PLS path modeling. *Computational Statistics and Data Analysis*, 2005, Vol. 48, pp. 159-205.
- [15] Theil, H., *Economic Forecasts and Policy*, Holland, Amsterdam : North, 1961.
- [16] Warne, R.T. and Larsen, R., Evaluating a proposed modification of the Guttman rule for determining the number of factors in an exploratory factor analysis. *Psychological Test and Assessment Modeling*, 2014, Vol. 56, pp. 104-123.
- [17] Widjaja, D., Varon, C., Dorado, A., Suykens, J.A., and Van Huffel, S., Application of Kernel Principal Component Analysis for Single-Lead-ECG-Derived Respiration. *Biomedical Engineering. IEEE Transactions on*, 2012, Vol. 59, pp. 1169-1176.
- [18] Wijesinghe, B., Low Temperature Drying of Food Materials Using Energy-Efficient Heat Pump Dryers. *CAD-DET Newsletter*, 1997, No. 7, pp. 4-5.

ORCID

- Chang-Yong Lee | <http://orcid.org/0000-0003-1778-6532>
Gensoo Song | <http://orcid.org/0000-0002-3185-2965>
Jinho Kim | <http://orcid.org/0000-0003-4243-7318>