

함수 주성분 분석을 이용한 일별 도시가스 수요 예측

최용옥*·박혜성**

요약 : 우리나라 도시가스 수요는 난방수요에 기인한 뚜렷한 동고하저의 계절성을 보이며, 기온에 따른 민감도는 시간에 따라 변화하는 것으로 나타났다. 본 연구에서는 시간에 따라 변화하는 계절성을 효과적으로 모형하기 위해서 시간변동 기온반응함수 개념을 도입하여 이를 해당 일의 기온분포로 적분하여 기온에 따른 수요변동을 추정한다. 또한 기상청에서 발표하는 향후 10일의 도시별 기온 예측치를 체계적으로 반영하여 도시가스 수요를 예측하는 방법론을 개발하였다. 평년기온분포를 사용한 것에 비해서 함수적 방법론을 이용하여 기상청의 기온 예측치를 기온분포예측치로 변환하여 예측했을 때 기온분포의 예측 오차율은 2배, 도시가스 수요의 예측 오차는 5배 가까이 감소하는 것을 확인하였다.

주제어 : 도시가스 수요 예측, 함수 주성분 분석, 시간에 따라 변하는 계절성, 일별 수요 예측, 기온분포 예측

JEL 분류 : C2 C5 Q3

접수일(2020년 10월 12일), 수정일(2020년 11월 29일), 게재확정일(2020년 12월 1일)

* 중앙대학교 경제학부 조교수, 제1저자(e-mail: choiyongok@cau.ac.kr)

** 성균관대학교 경제학과 석사과정, 교신저자 (e-mail: gaoyunpark@gmail.com)

Daily Gas Demand Forecast Using Functional Principal Component Analysis

Yongok Choi* and Haeseong Park**

ABSTRACT : The majority of the natural gas demand in South Korea is mainly determined by the heating demand. Accordingly, there is a distinct seasonality in which the gas demand increases in winter and decreases in summer. Moreover, the degree of sensitiveness to temperature on gas demand has changed over time. This study firstly introduces changing temperature response function (TRF) to capture effects of changing seasonality. The temperature effect (TE), estimated by integrating temperature response function with daily temperature density, represents for the amount of gas demand change due to variation of temperature distribution. Also, this study presents an innovative way in forecasting daily temperature density by employing functional principal component analysis based on daily max/min temperature forecasts for the five big cities in Korea. The forecast errors of the temperature density and gas demand are decreased by 50% and 80% respectively if we use the proposed forecasted density rather than the average daily temperature density.

Keywords : Forecasting gas demand, Functional principal component analysis, Changing seasonality, Daily gas demand, Forecasting temperature density

Received: October 12, 2020. Revised: November 29, 2020. Accepted: December 1, 2020.

* Assistant Professor, School of Economics, Chung-Ang University, First author(e-mail: choiyongok@cau.ac.kr)

** Master's Degree, Department of Economics, Sungkyunkwan University, Corresponding author(e-mail: gaoyunpark@gmail.com)

I. 서론

국내 도시가스 가정용 공급세대수가 2019년 12월을 기준으로 1,881만 6,139세대¹⁾에 달하면서 전국 도시가스 평균 보급률이 83.7%에 이르는 것(2019년 12월 도시가스 월보 기준)으로 집계되었다. 이와 같이 많은 국민이 사용하는 도시가스를 안정적으로 공급하기 위해서는 정확한 수요 예측이 필수적이지만, 도시가스 수요를 결정하는 가격 및 기온의 변동성이 매우 크기 때문에 정확한 예측을 하기가 어렵다.

특히 난방수요가 많이 발생하는 동절기의 경우, 일별 도시가스 수요는 기온에 민감하게 반응하는데, 우리나라의 경우 겨울철 기온의 변동 폭이 매우 크게 나타난다. 예를 들어서 2020년 1월 6일에서 1월 17일 사이(1월 첫째 주 및 둘째 주)의 근무일²⁾ 수요를 보면 1월 7일의 수요는 78,299톤으로 1월 15일 93,594톤 대비 16.6%나 낮게 실현되었는데, 이는 1월 7일과 1월 15일의 평균기온³⁾이 각각 8.3°C 및 -0.8°C로 크게 차이가 나기 때문인 것으로 보인다.

이는 일별 도시가스 수요 예측을 할 때, 기온을 평년 기온 값으로 가정하는 것보다 단기 예측 정보를 이용하여 예측할 필요가 있음을 시사한다. 이에 본 연구에서는 기상청의 중기 예보(10일 이내의 지역별 최저, 최고 기온 예측)를 체계적으로 이용하여 일별 수요 예측을 하는 방법을 제안한다. 보다 구체적으로는 중기 예보를 이용하여 향후 10일의 기온분포를 합수적 주성분 방식으로 예측하고 이를 이용해서 기온효과를 예측하여 도시가스 수요를 예측한다.

분포 예측과 관련해 최근 연구의 추세는 머신러닝 기법을 사용하여 기온 분포 자체를 예측하는 것이다. 대표적으로 Diebold and Shin(2019)에서는 여러 분포를 조합하여 더 나은 분포를 예측하고자 함에 있어 partially-egalitarian LASSO 방법론을 이용하였다. Partially-egalitarian LASSO 방법론은 여러 분포를 조합할 때 동일한 가중치를 적용하는 단순 평균법의 이점을 LASSO에 적용한 것으로 2단계 추정 과정을 거친다. 이 외에도 Geweke and Amisano(2011)는 인공신경망 방법(Artificial Neural Network)이자 동시에

1) LPG로 공급되는 29,005세대를 포함한 값이다.

2) 휴일의 경우에 산업용 도시가스 수요의 감소로 전반적으로 수요가 낮게 결정되므로 기온에 의한 요인을 식별하기 힘들어서 제외하였다.

3) 5대 지역(서울, 대전, 대구, 광주, 부산)의 가중 평균기온이며, 가중치는 2018년도의 권역별 판매량이다.

마코브 가우시안 혼합 모형(Marocov Normal Mixture)을 사용하여 S&P 500의 일별 주가와 1년 만기채권 수익률 데이터를 사용하여 주가 주식의 일별 분포를 예측하였다.

도시가스 수요 예측 모형은 많은 연구자에 의해서 개발됐다. 자료의 빈도에 따라서 구분을 해보면, 이승재 외(2013)에서는 1981년 이후 연별자료를 이용해서 당해 연도의 수요가 당해 연도 가격 및 총생산뿐만 아니라 전년도 수요에 의해서 결정되는 것으로 보고 장단기 소득 및 가격탄력성을 계산하였다. 월별 자료를 이용한 모형에서는 김점수 외(2011), 김인무 외(2011), 이성로(2017) 등이 있는데 이들 모형은 기본적으로 시간변동 계수를 갖는 공적분 모형으로 수요와 경제성장 사이의 시간에 따라 변화하는 균형 관계를 설정하였다. 김점수 외(2011)에서는 국내총생산과 기온변수를 설명변수로 하여 월별 수요를 예측하는 모형을 제안하였으며 김인무 외(2011)에서는 기온가격 교차반응함수를 이용하여 국내총생산 이 외에 기온 및 상대가격 등을 이용하여 월별 수요를 예측하였다. 이성로(2017)에서는 기온효과 및 상대가격을 설명변수로 이용하여 지역별로 예측할 경우에 예측성고가 개선될 수 있음을 보였다. 한편 본 연구와 같이 일별 수요를 예측하는 연구를 살펴보면, 이근철과 한정희(2015)는 회귀 모형 분석 시 선별적으로 샘플링한 일별 수요를 대상으로 하되⁴⁾, 직전일 수요, 기온, 요일 등을 이용해서 일별 총수요를 예측하는 모형을 제안하였으며, 양준모 외(2017)에서는 ARMAX 모형을 이용하여 요일효과와 월별(기온)효과를 설명변수로 하여 지역별 수요를 예측하였고 지역별 모형의 예측치를 합산하여 총수요에 대한 전망치를 구하는 방식을 제시하였다.

기존 연구의 특징을 살펴보면, 대부분의 연구에서 도시가스 수요를 결정하는 요인으로 경제성장 변수 이 외에 상대가격 및 기온을 사용하고 있음을 알 수 있다. 장기 모형일수록 상대가격 변수가 중요하고, 단기 모형일수록 요일효과 및 기온이 중요한 변수임을 확인할 수 있다.

본 연구는 기존 연구의 일별 모형과 다음과 같은 점에서 차별된다. 첫째 기존 모형에 비해서 특수일 효과를 정교하게 추정한다. 법정 공휴일뿐 아니라, 도시가스 수요에 영향을 미칠 수 있는 요인, 예를 들어서 12월 31일 효과, 명절-2, 명절+2, 명절+3일 효과, 샌드

4) 이근철과 한정희(2015)에서 소개된 선별적 샘플링의 기준은 크게 두 가지인데, 첫 번째 기준은 예측일과 유사한 과거 데이터만을 샘플링에 포함시키는 것이고, 두 번째 기준은 이상치가 포함된 과거 데이터를 샘플에서 제외시키는 것이다.

위치일 효과, 연휴 효과, 하계 집중휴가기간 효과 등을 추정하고 있다. 두 번째, 시간에 따라서 기온에 대한 민감도가 달라지는 것을 고려하여 기온효과를 측정하고 있다. 박성근(2016)에서 밝힌 바와 같이 난방용 도시가스 수요 및 난방용 전력수요는 대체 관계가 있으며, 전력을 사용하는 난방기기 보급의 증가 및 단열기술 등의 향상에 따라서 같은 일별 기온분포라고 하더라도 과거에 비해서 최근의 수요는 낮게 실현될 수 있는 것을 반영하고 있다. 세 번째, 가장 중요한 차별점은 일별 수요를 예측할 때 기상청의 정보를 체계적으로 반영해서 기온분포를 예측하고 이를 이용해서 기온효과를 예측해서 사용하는 방안을 제시한 것이다.

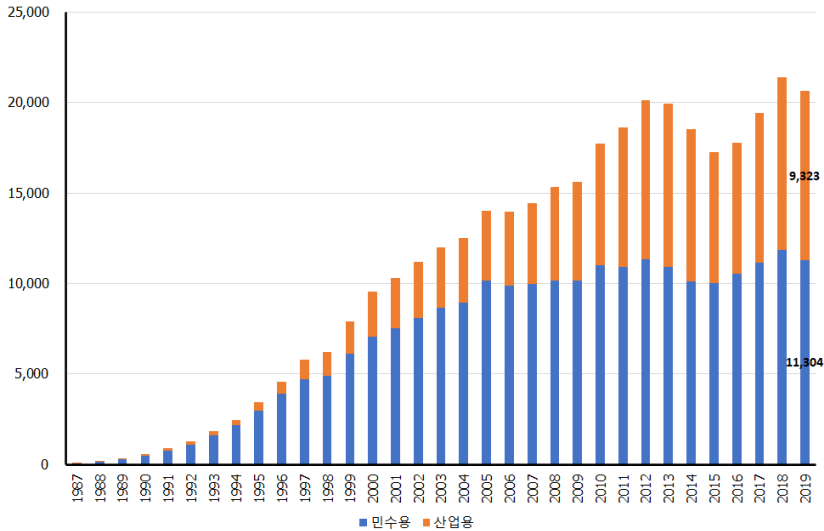
본 논문의 구성은 다음과 같다. 2장에서는 도시가스 수요의 특징을 추세 및 계절성의 측면에서 살펴본다. 3장에서는 본 연구에서 제안하는 일별 도시가스 수요 모형에 대해서 설명하고 추정결과를 정리한다. 4장에서는 함수적 주성분 방식으로 기온분포를 예측하는 과정을 설명한다. 5장에서는 본 연구에서 제안하는 방식으로 기온효과를 예측한 모형과 평년 기온분포를 이용해서 예측한 모형의 예측성과를 비교하고, 6장에서는 결론을 맺는다.

II. 도시가스 수요의 시계열적 특징

국내 도시가스⁵⁾ 수요는 1987년 보급되기 시작한 이래로 2007년까지는 급격히 증가하는 추세를 보였다. 이와 같은 급격한 증가 추세는 도시가스 공급에 필수적인 배관망이 1987년 도시가스 보급 시작 이래로 지속해서 확충되었기 때문이다. <그림 1>은 도시가스가 보급되기 시작한 1987년부터 최근 2019년까지의 도시가스 용도별 수요의 추이를 나타낸 그래프이다. 민수용과 산업용으로 구분하여 나타냈으며, 민수용은 주택용과 일반용을 합한 수요이다. 2007년까지 도시가스 수요의 증가는 주로 민수용 수요 증가에 기인하며, 2007년 이후는 산업용 수요에 의해서 도시가스의 수요의 등락폭이 결정된다. 특히 2007년 이후의 변동은 김인무 외(2011) 및 이성로(2017)에서 지적한 바와 같이 도

5) 일반적으로 도시가스는 LNG 및 LPG로 공급되는 모든 물량을 포괄하지만, 2019년 기준 LPG로 공급되는 물량은 전체의 0.1%에 불과하다(2019년 도시가스 월보기준). 본 연구에서는 LNG로 공급되는 도시가스 수요를 분석하였다.

〈그림 1〉 국내 도시가스 수요의 추세



시가스의 나프타 등 석유류의 상대가격에 따라서 산업용 수요가 변동했기 때문인 것으로 보인다.

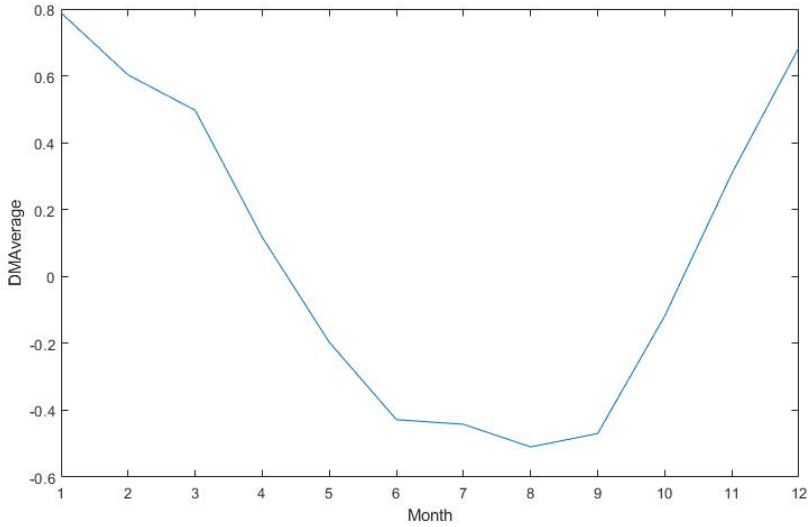
도시가스용 수요 시계열의 다른 특징은 뚜렷한 계절성을 보인다는 점이다. 도시가스용 수요는 여름철에 비해서 겨울철에 많은 특징을 보이며, 이는 주로 도시가스가 난방용 수요를 충당하는 데 사용되기 때문이다. 이와 같은 사실은 <그림 2>에서도 확인할 수 있다. <그림 2>는 도시가스용 수요의 계절성을 보기 위해서 원 시계열에서 추세를 제거한 뒤에 이를 월별로 평균하여 나타낸 것이다.

도시가스 수요는 동고하저의 계절성 이 외에도 일별 특성에 따라서도 영향을 받는다. 이는 주로 산업용 수요가 근무일 대비 공휴일에 낮게 발생하기 때문⁶⁾이다. 특히 난방수요가 작은 계절에는 민수용 수요가 산업용 수요에 비해서 작기 때문에 근무일 대비 공휴일에 수요 감소 폭이 상대적으로 크게 발생한다. 이와 같은 특징은 <그림 3>에서 확인할 수 있다. <그림 3>은 2019년도 근무일, 토요일, 일요일 혹은 공휴일일 경우를 구분하여 월별 도시가스 수요량의 평균을 로그 변환한 시계열이다. 예상한 바와 같이 일요일 및 토

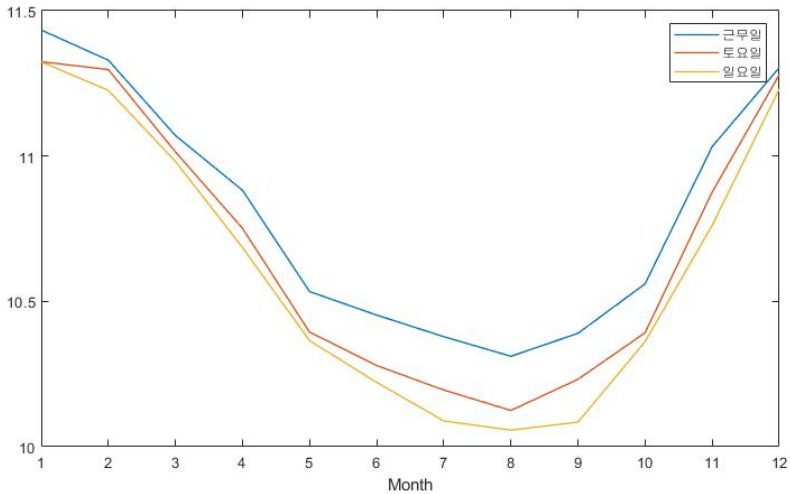
6) 이와 같은 차이로 인해 민수용과 산업용 수요를 별도로 설정 하는 것이 바람직할 수도 있으나, 일별 자료로는 민수용과 산업용이 구분되어서 관측되지 않는다.

요일 수요가 상대적으로 근무일에 비해서 낮게 나타났으며, 일요일 및 공휴일의 수요가 토요일 보다 낮게 나타났다. 특히 겨울철보다 여름철에 감소 폭이 더 크게 나타났다.

〈그림 2〉 도시가스 수요의 계절성



〈그림 3〉 요일 특성별 도시가스 수요의 계절성



III. 일별 도시가스 모형 및 추정결과

2장에서 살펴본 일별 도시가스 수요의 특징을 반영하여 본 연구에서는 도시가스 수요를 다음과 같은 모형으로 적합하고 예측한다.

$$\log(DGD_t) = \alpha + \beta_1 TE_t + \beta_2 CTE_t + \gamma \log(MADGD_{t-10}) + \sum_{k=1}^{29} \delta_k SDE_{kt} + \epsilon_t \quad (1)$$

식 (1) 일별 도시가스 수요 예측 모형은 일별 도시가스 수요량을 로그 변환한 값을 종속 변수로 사용하며, 설명 변수로는 일별 기온효과(TE_t), 누적기온효과(CTE_t), 365일 이동평균한 일별 수요량의 10일 전 변수($MADGD_{t-10}$), 특수일 효과(SDE_t) 터미변수를 사용한다.

도시가스 수요 시계열의 계절성은 대부분 난방수요에 기인한다. 이에 도시가스 수요 예측 모형에서는 기온에 따라 변화하는 도시가스 수요량을 정확하게 추정하기 위하여 기온효과(TE , *Temperature Effect*)라는 변수를 사용하고 있다. 기본적으로 김점수 외(2011), 김인무 외(2011), 이성로(2017) 등에서 사용된 변수와 동일한 변수이다. 즉, 기온효과(TE)는 주어진 기간(본 모형에서는 하루 동안)의 기온분포와 당시의 기온이 도시가스 수요에 얼마나 민감하게 영향을 미치는지를 나타내는 기온반응도에 의해서 정의된다. 구체적으로 기온효과는 다음과 같이 정의된다.

$$TE_t = \int g_t(s) f_t(s) ds \quad (2)$$

$g_t(s)$ 는 t 시점의 기온반응함수이고, $f_t(s)$ 는 t 일의 기온분포함수이다. t 일의 기온효과(TE_t)는 해당 일의 기온분포함수를 추정된 뒤 이를 기온반응함수로 리만 적분한 값이다. 기온효과는 특정 일 t 에서의 기온분포함수에 온도 별로 도시가스 수요가 민감하게 반응하는 정도를 나타내는 기온반응함수를 곱해서 합하기 때문에 도시가스 수요가 특

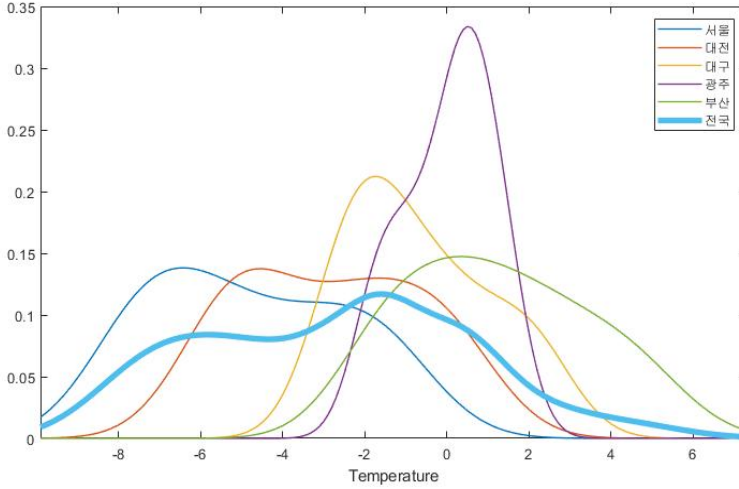
정 온도 구간에서 해당 온도에 따라 반응하는 정도를 체계적으로 반영할 수 있다. 특히 가스수요가 기온에 비선형적으로 반응하는 것을 자연스럽게 반영할 수 있다는 장점이 있다. 당연히 기온효과가 크다는 것은 해당 일의 도시가스 수요가 크게 발생할 것임을 의미한다.

기온분포함수, $f_t(s)$ 는 t 일의 한국의 기온 분포를 나타낸다. 분명히 도시가스 수요는 기온에 의해서 결정된다. 하루 전체의 도시가스 수요는 하루 전체의 기온에 따라서 결정되는 것으로 보는 것이 합리적이며, 평균기온이나, 최고·최저 기온 등의 정보뿐만 아니라, 각 기온별로 지속 시간까지 고려할 수 있다는 점에서 보다 정확히 기온효과를 측정할 수 있다. 단적으로, 평균 기온은 같은 경우에도 불구하고 일교차가 큰 날이 작은 날에 비해 상대적으로 더 큰 가스 수요가 발생하므로 평균 기온만으로는 추정하기 힘든 기온효과를 반영할 수 있다. 또한 평균 기온 이 외에 최고·최저 기온을 이용하는 경우에는, 특정한 시점의 기온이 높거나 낮을 때에는 가스 수요가 이에 비례해서 증가하지 않기 때문에 필연적으로 기온효과를 과대 측정하게 된다는 한계가 있는데, 이를 효과적으로 극복할 수 있다.

기온분포함수는 t 시점에서 기온이 특정 온도 구간에 얼마나 자주 나타났었는지의 상대빈도를 평활화한 함수로 정의될 수 있다. $f_t(s)$ 의 정의역에 해당하는 온도 구간은 우리나라 기온에서 실현되는 범위인 영하 20°C부터 영상 40°C까지 0.1°C로 설정하였다. 서울, 대전, 대구, 광주, 부산을 포함하는 5대 지역 기온 자료를 사용하여 지역별 기온분포를 비모수적으로 추정한 뒤 일별 가스 지역별 판매량을 이용하여 가중 평균 전국 기온분포를 추정하였다. 예를 들어, 이와 같은 과정을 거쳐 추정된 2019년 1월 1일의 기온분포 함수를 나타낸 것이 <그림 4>이다.

기온반응함수란 기온에 따라서 도시가스 수요가 영향을 받는 민감도의 정도를 의미하는 함수이다. 가스 수요의 경우, 특징적으로 겨울철에 낮은 온도에 비교적 더 민감하게 반응하며 기온이 올라감에 따라 여름철에는 비교적 덜 민감하게 반응한다. 이는 온도에 대한 가스 수요의 반응이 비선형적임을 뜻한다. 본 연구에서는 김점수 외(2011), 김인무 외(2011), 이성로(2017)와 달리 시간에 따라서 기온 민감도가 달라지는 것을 허용하여 기온반응함수를 추정하였다. 이와 같은 모형으로, 같은 기온 분포에도 2007년에 비해서 2020년의 가스수요가 낮게 실현되는 것을 설명할 수 있다.

〈그림 4〉 2019년 1월 1일 지역별, 전국 기온분포 추정 예시



기온반응함수는 다음과 같이 설정한다.

$$g_t(s) = b_{1t}(s - a)^2(1 - G(s)) + b_{2t}(s - a)^2G(s) \quad (3)$$

where $G(s) = \frac{1}{1 + e^{-c(s-a)}}$

여기에서 $g_t(s)$ 는 t 시점에서 기온 s 에 대한 도시가스 수요의 반응을 모수적으로 나타낸 함수이고, 모수에 따라서 다양한 형태의 비선형적인 반응도를 표현할 수 있다. 본 연구에서는 b_{1t} 와 b_{2t} 는 기온민감도가 시간에 따라서 비선형적으로 변할 수 있도록(7) 식 (4), (5)와 같이 설정하였다.

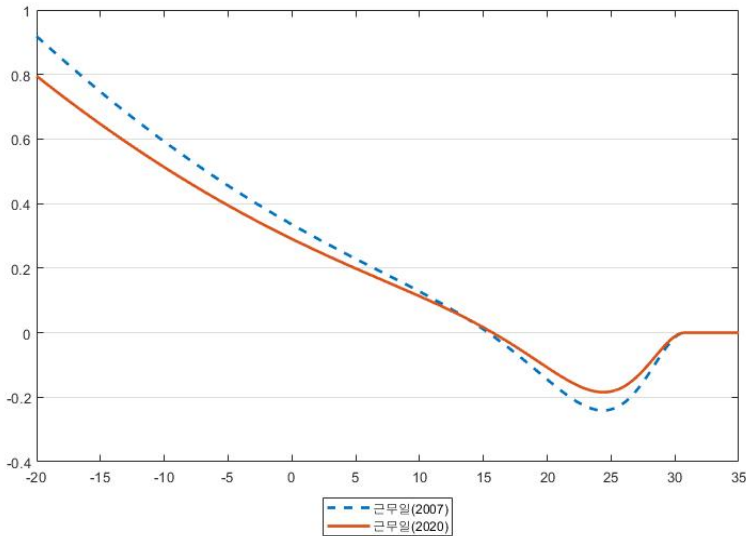
$$b_{1t} = b_1^0 + \frac{t}{T}b_1^t + b_1^c \cos\left(2\pi \frac{t}{T}\right) + b_1^s \sin\left(2\pi \frac{t}{T}\right) \quad (4)$$

7) 이와 같은 형태의 모형 설정은 한국전력거래소(2020)의 결과인 전력수요의 기온반응함수 설정과 동일한 것이다. 한국전력거래소(2020)에서는 전력수요의 기온반응함수가 시간에 따라서 비선형적으로 변하는 것을 밝혔으며, 전력수요와 도시가스수요 사이의 대체 관계를 고려할 때, 도시가스 수요의 기온반응도 역시 시간에 따라서 비선형적으로 변할 수 있도록 설정하였다.

$$b_{2t} = b_2^0 + \frac{t}{T}b_2^t + b_2^c \cos\left(2\pi \frac{t}{T}\right) + b_2^s \sin\left(2\pi \frac{t}{T}\right) \quad (5)$$

<그림 5>는 근무일의 기온반응함수를 비선형최소자승법(nonlinear least squares)으로 추정한 결과⁸⁾이다. 2007년에 비해서 2020년의 민감도가 낮아진 것을 확인할 수 있다. 이는 동일한 기온분포에 대해서 2020년의 도시가스 수요가 낮게 결정되는 것을 뜻한다.

<그림 5> 시간에 따라서 변하는 기온반응함수 추정결과: 근무일

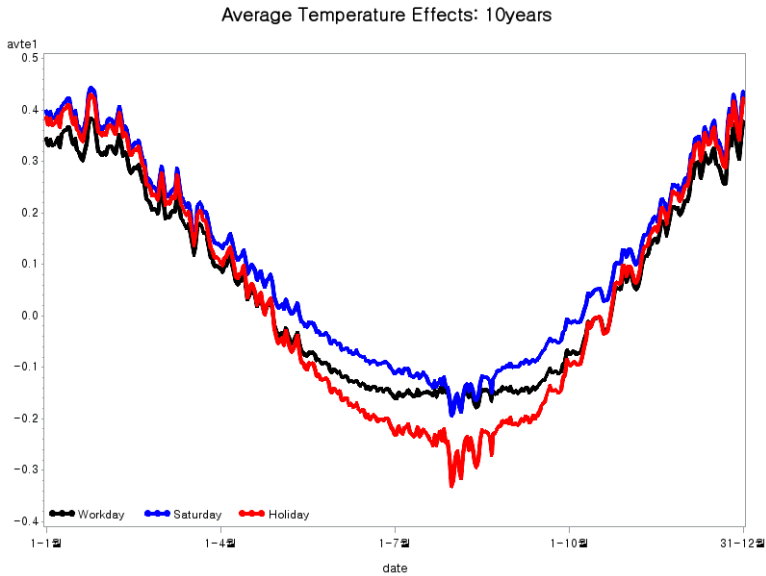


매일의 기온분포함수와 기온반응함수를 적분하면, 해당일의 일별 기온효과를 추정할 수 있다. <그림 6>은 2020년 1월 1일 시점의 기온반응함수를 최근 10년간 평균 기온분포로 적분하여 나타낸 기온효과이다. 예상과 같이 동고하저의 계절성을 보이고 있다.

일별 도시가스 수요는 당일의 기온뿐만 아니라 그 전날의 기온분포에도 영향을 받는 것으로 보인다. 예를 들어서 2020년 1월 10일과 2020년 1월 17일의 평균 기온은 각각

8) 25도 이상에서 기온반응도가 올라가게 추정된 것은 모형설정오류에 의한 것으로 보인다. 추정기간 및 설명변수의 차이에 따라서 해당 구간의 추정값이 불안정하게 바뀌는 것을 확인하였다. 그러나 이와 같은 오류가 전체 결과에 큰 영향을 미치지 않는 것으로 판단된다. 실제로 모형에서 사용되는 설명변수는 기온효과인데, <그림 6>에서 확인할 수 있는 기온효과는 기온이 높을 때 도시가스 수요가 낮아지는 것을 반영하고 있다.

〈그림 6〉 10년 평균 기온분포를 이용한 기온효과 추정결과



1.1°C와 1.3°C로 나타났는데, 도시가스 수요는 각각 85,082톤과 89,865톤으로 평균기온이 높았던 1월 17일에 5.6% 높게 실현되었다. 이는 1월 17일의 경우 1월 14일에서 1월 16일의 직전 3일의 평균기온이 -0.5°C인 반면 1월 10일의 경우 1월 7일에서 1월 9일의 평균기온이 5°C인 것에 기인한다.

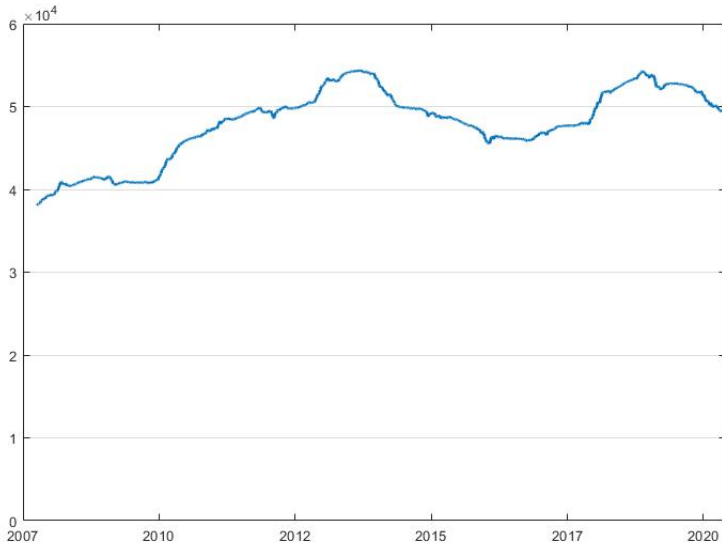
본 연구에서는 이와 같은 누적기온효과를 distributed lag model를 이용하여 반영하였다. 당일의 기온반응함수의 형태를 유지하면서, 당일의 기온 반응함수보다 하루 전의 기온 반응함수가 기온별 반응 차이가 작아지도록 설정하였는데, 1일 전의 기온반응함수의 계수에 가중치 w^1 을 곱하고, 2일 전의 기온반응함수의 계수는 당일 기온반응함수의 계수에 가중치의 제곱항, w^2 을 곱하여 사용하였고, 3일 전의 기온반응함수의 계수는 가중치의 세제곱항, w^3 을 곱하였다. 그러므로 당일에서 시간이 멀어질수록 기온반응함수의 계수가 0에 가까워지면서 평탄해져서, 3일 전의 기온변화에 따른 오늘의 기온 반응이 2일 전의 기온변화에 따른 반응보다 평탄하게 모형을 설정하였다.

9) 가중치는 RMSE가 가장 낮아지는 값(0.47)으로 선택하였다.

일별 도시가스 수요에 영향을 주는 다른 요인으로 특수일 효과를 고려하였다. 특수일 효과(Special Day Effects, SDE)는 특수일의 도시가스 수요가 근무일에 비해서 낮아지는 것을 반영한다. 본 모형에서는 법정 공휴일뿐 아니라 전통적으로 도시가스 수요가 낮아지는 특수일을 함께 고려하고 있다. 예를 들어, 일요일, 국경일¹⁰⁾, 1월 1일, 설날, 식목일, 석가탄신일, 어린이날, 현충일, 추석, 성탄일, 선거투표일, 기타 정부에서 수시로 정하는 날¹¹⁾ 등은 공휴일 요인이다. 비공휴일 요인으로는 수요가 근무일에 비해서 작게 나타나는 월요일, 토요일, 근로자의 날, 12월 31일, 샌드위치 Day, 명절 연휴 앞 날, 명절 연휴 뒤 이틀, 하계 집중 휴가기간 등이 있다.

배관망이 확충된 2007년 이후에도 산업용 도시가스 수요는 석유제품과의 상대가격에 따라서 크게 변동되므로, 단기적인 수준을 설명하기 위한 변수가 필요하다. 기존 연구인 이근철과 한정희(2015)와 양준모 외(2017)에서는 직전일의 수요를 설명변수로 포함해서 전반적인 수요 수준을 적합하는데, 본 연구에서는 365일 이동평균한 일별 수요

〈그림 7〉 일별 도시가스 판매량의 365일 이동평균 시계열(10일 전 값)



10) 삼일절, 제헌절, 광복절, 개천절.

11) 올림픽 개막식, 월드컵 휴일 등.

량의 10일 전에 상응하는 값¹²⁾을 사용한다. 이와 같이 이동평균으로 생성한 항은 계절효과 및 특수일 효과가 제거된 전반적인 수요의 수준을 설명하는 변수이다. <그림 7>은 2007년 10월 1일에서 2020년 5월 31일까지 계산한 로그를 취하기 전 이동평균 판매량이다. 산업용 수요가 급격히 증가했던 2013년에 가장 높은 수준임을 확인할 수 있다.

이 같은 변수를 이용하여 식 (1)을 공적분 모형으로 추정한 결과는 <표 1>에 정리되어 있다. 기온효과 및 누적기온효과, 이동평균 판매량의 시차변수 등이 모두 유의하게 추정되었으며, 개별 공휴일 효과에서는 주로 겨울철 공휴일의 감소 효과가 상대적으로 작게 나타났다.

<표 1> 일별 도시가스 예측 모형 추정 결과

	계수 추정치	t-value		계수 추정치	t-value
상수항	0.096	0.2158	설날-2	0.109	1.294
기온효과	1.168	12.676	설날-1	0.072	0.744
누적 기온효과	1.423	12.184	설날	-0.064	-0.663
월요일	-0.019	-1.455	설날+1	0.039	0.407
토요일	-0.207	-15.962	설날+2	0.164	1.697
휴일	-0.185	-13.904	설날+3	0.192	1.990
1월 1일	0.039	0.404	추석-2	0.134	1.589
삼일절	0.134	1.592	추석-1	-0.061	-0.635
노동절	0.046	0.553	추석	-0.309	-3.206
어린이날	0.036	0.425	추석+1	-0.219	-2.271
현충일	0.104	1.242	추석+2	-0.007	-0.072
광복절	0.245	2.897	추석+3	0.156	1.621
제헌절	0.054	0.643	선거일	0.084	0.834
성탄절	0.074	0.878	CH	-0.084	-1.719
12월 31일	0.082	0.970	샌드위치데이	-0.059	-1.355
부처님 오신 날	0.071	0.834	하계 휴가	-0.107	-2.812
			표준화된 일별 판매량	0.979	23.643

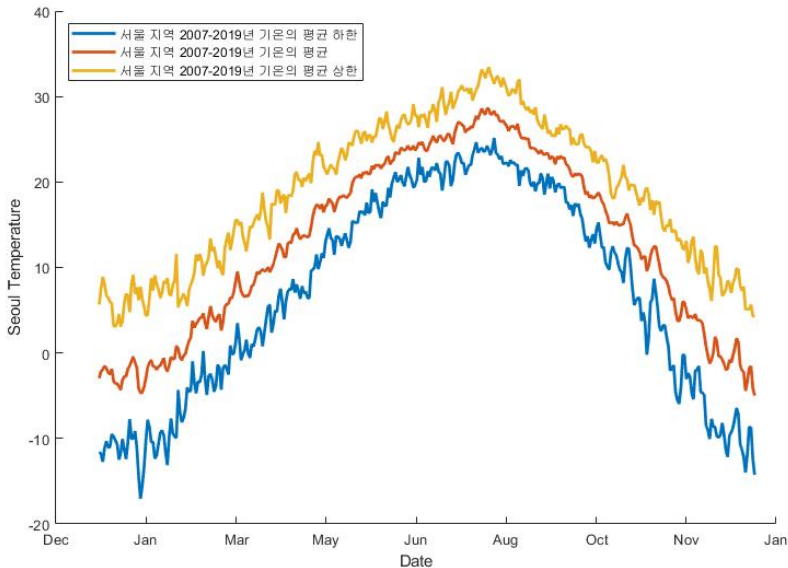
12) 식으로 표현하면 t 시점의 설명변수로는 $\log\left(\frac{1}{365} \sum_{k=1}^{365} DGD_{t-k-10}\right)$ 를 사용한다.

IV. 기온분포 예측모형

일별 도시가스 수요 예측에 가장 큰 영향을 주는 요인은 기온이다. 특히 겨울철 수요는 기온에 의해서 크게 영향을 받는다. 그런데, 겨울철 기온은 변동성이 큰 것으로 확인되었다. <그림 8>은 2007년에서 2019년 사이의 시간별 기온자료를 이용해서 서울지역의 일평균 기온을 추정하고, 일자별 평균기온과 평균기온의 표준편차를 이용해서 상하한을 나타낸 그래프이다. 여기서 상하한 폭은 표준편차의 2배로 설정하였다. 예를 들어서 과거 1월 1일의 서울지역의 평균기온은 -2.9°C 로 추정되었으며, 일평균 기온들의 표준편차는 4.3°C 로 추정되었다. 반면에 7월 31일의 서울지역의 평균기온은 27.3°C 이며 일평균 기온들의 표준편차는 1.9°C 에 불과한 것으로 나타났다.

본 연구의 모형의 기온효과를 이용해서 분석한 결과 겨울철의 경우 평균 기온이 1도 하락하면, 도시가스 수요는 평균적으로 2.4% 증가하는 것으로 분석되었다. 겨울철 기온의 변동폭이 크다는 것을 고려할 때, 일별 수요 예측의 정확도를 높이기 위해서 정확한 기온 예측치를 사용하는 것이 매우 중요한 문제가 된다.

<그림 8> 일 평균 기온의 계절에 따른 변동성: 서울지역 2007~2019년 일평균 기온



일별 도시가스 수요 예측을 수행할 때 기온 예측치를 평년 기온에 해당하는 값을 사용할 수도 있다. 구체적으로 설명하자면, 과거 10년 동안의 전국 5대 지역별 시간당 기온 자료를 이용하여 일별 평균 기온 분포를 추정하고 이를 예측 전제로 삼아 기온효과를 추정할 수 있다. 이와 같은 방식의 가장 큰 한계점은 기온의 변동이 매우 커서 일 평균 기온조차도 실제 기온에서 크게 벗어날 수 있다는 점이다. 또한 과거의 기온 분포는 일반적으로 일교차가 큰 날과 작은 날이 평균되어 전반적으로는 평균기온에 집중되는 분포를 갖게 되는데, 실제의 기온은 일교차가 크게 발생하는 경우가 잦다. 이와 같은 현상은 과거 기온분포의 평균으로는 반영할 수가 없다.

본 연구에서는 기온의 예측 정보를 활용하여 도시가스 수요에 대한 예측력을 높이는 방안을 제안한다. 대표적으로 활용할 수 있는 기온 예측 정보에는 기상청에서 제공하는 기온 예측치가 있다. 기상청은 기온 예측과 관련해 초단기, 단기, 중기예보를 제공하고 있다. 초단기예보란 예보 시점부터 6시간 이내에 대하여 행하는 예보를 의미하며, 단기 예보는 예보 시점부터 3일 이내 기간, 중기 예보는 예보 시점부터 10일 이내의 예보를 의미한다. 본 연구에서는 중기 예보의 지역별 최저 기온, 최고 기온의 예보 값을 활용하여 기온분포를 예측하는 방안을 제시한다.

본 연구에서는 함수 주성분 분석을 사용하여 일별 기온분포를 예측하였다. 이 방법은 일별 기온분포를 함수 시계열로 인식하고, 함수시계열의 변동을 설명하는 주요 요인을 주성분 분석을 통해 분석한 후 이에 상응하는 좌표과정(coordinate process)을 실제 일별 기온분포 예측의 종속 변수로 활용하는 것으로 요약할 수 있다.

$y(t, s) = y_t(s)$ 를 t 일의 기온 s 에서의 확률밀도 값을 나타내는 함수로 상정하면, $y_t(s)$ 혹은 y_t ¹³⁾를 t 에 따라서 정의된 함수 시계열로 생각할 수 있다. 만약 $y_t(s)$ 가 힐버트 공간(Hilbert Space) H 의 확률 변수(random variable)¹⁴⁾이고 힐버트 공간이 분할가능(separable)하다면

$$y_t(s) = \sum_{k=1}^{\infty} \langle v_k, y_t \rangle v_k \tag{6}$$

13) 여기서 $y_t(s)$ 와 y_t 의 차이는 argument를 특정하는가의 여부만이 다른 동일한 대상을 지칭한다.

14) $y_t(s)$ 가 힐버트 공간의 원소가 되기 위해서는 $\int y_t(s) ds = 0$ 과 같은 조건을 만족해야 하며, 본 연구에서는 해당 일의 평년 기온분포를 차감한 값으로 정의해서 사용한다.

와 같이 나타낼 수 있다. 여기서 \langle, \rangle 는 힐버트 공간에서 정의되는 내적(inner product)으로서, $\langle u, v \rangle = \int v(s)u(s)ds$, for $u, v \in H$ 를 나타내고, v_k 는 힐버트 공간의 정규직교기저(orthonormal basis)이다. 위의 식은 벡터 공간 정규직교기저 representation의 무한대 공간 버전이라고 볼 수 있다. 함수 시계열에 대한 구체적인 내용들은 Bosq(2000), Chang et al.(2019)에서 확인할 수 있다.

위와 같은 무한 차원공간을 유한차원으로 차원 축소(dimensional reduction)하는데 가장 많이 쓰이는 방법은 함수 버전의 주성분분석(PCA, Principal Component Analysis)이다. 각 t 시점의 $y_t(s)$ 는 충분히 큰 차원(m)의 벡터로 근사(approximation)가 가능하고, 벡터를 다변량분석에서의 벡터와 같이 취급하여 분산 공분산 행렬을 정의할 수 있다. 분산 공분산 행렬의 고유값(eigenvalue)값을 $\lambda_1 > \dots > \lambda_m$ 이라고 했을 때, 상대적으로 큰 고유값에 상응하는 고유벡터 v_1, \dots, v_p 를 주성분 벡터라고 하며, 식 (6)은 다음과 같이 근사될 수 있을 것이다.

$$y_t(s) = \sum_{k=1}^p \langle v_k, y_t \rangle v_k \tag{7}$$

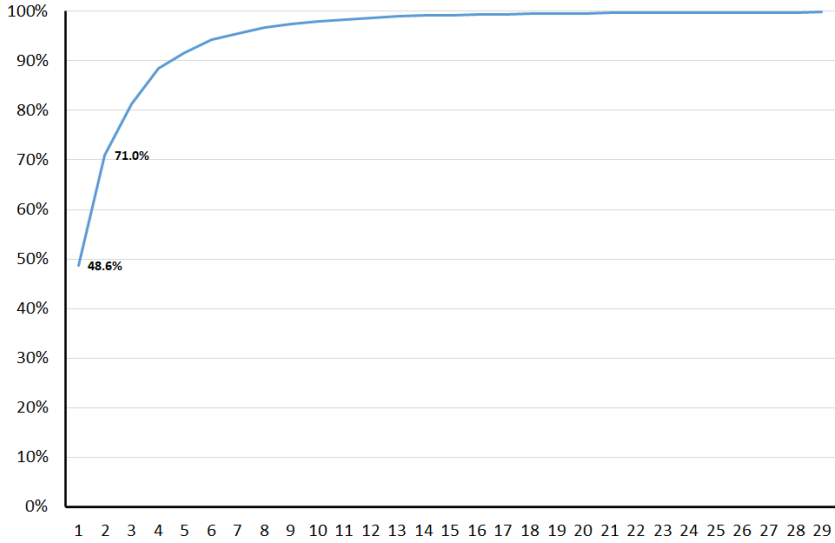
<그림 9>는 함수적 주성분 분석 결과로 도출된 cumulative scree plot이다. 실제 기온 분포함수와 평년기온분포 함수의 차이를 분석 대상이 되는 시계열로 상정하고 분석한 결과인데, 첫 번째와 두 번째까지의 주성분만으로 각각 48.6%, 71.0%의 변동을 설명 가능한 것으로 확인되었다. 이에 본 연구에서는 $p = 2$ 로 설정하고 분석을 진행하였다.

<그림 10> 과 <그림 11>은 주성분 분석에서 추정된 첫 번째 두 번째 요인, 정규직교기저이다. <그림 10>은 평균기온 대비 낮은 기온의 빈도가 높은 날에 높은 가중치를 주는 요인¹⁵⁾이고, <그림 11>은 대략적으로 평균기온 대비 높은 기온의 빈도가 높은 날에 높은 값을 주는 요인¹⁶⁾이다. 결국 두 요인을 종합적으로 고려하면, 평균 기온 대비 높거나

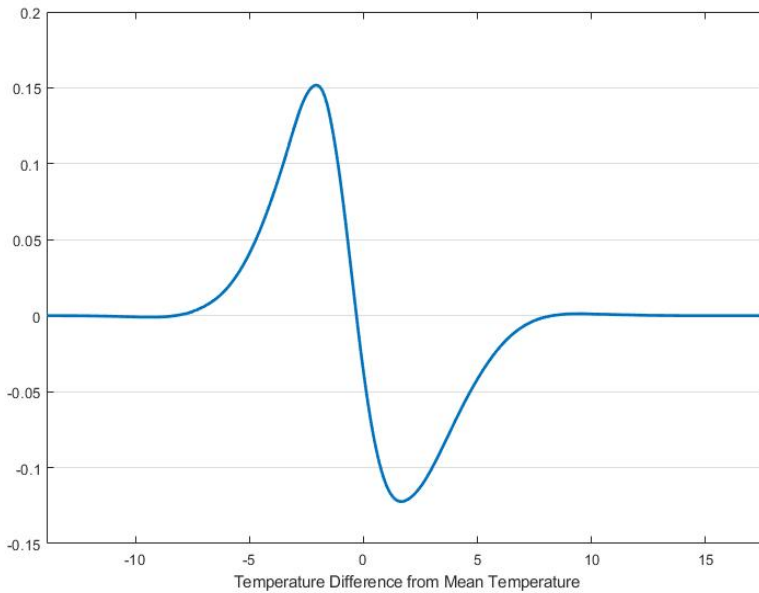
15) 여기서 분석하는 y_t 는 t 일의 기온분포에서 해당 일의 평균이 조정된 평년 기온분포를 뺀 함수이다. 따라서 주성분벡터의 정의역은 평균기온에서 벗어난 정도로 정의가 되며, 그림 10의 경우에 0보다 작은 값에서 큰 값을 갖고 0보다 작은 값에서 작은 값을 갖는 형태로 나타나는데, 평균 기온 보다 작은 기온의 분포가 클 때 $\langle v_1, y_t \rangle$ 가 큰 값을 갖게 된다.

16) 첫 번째 주성분과 달리 두 번째 주성분의 특징은 뚜렷하지 않지만, 대략적으로 평균 기온 보다 높은 기온의 분포가 클 때 $\langle v_2, y_t \rangle$ 의 값이 상대적으로 큰 것을 확인하였다.

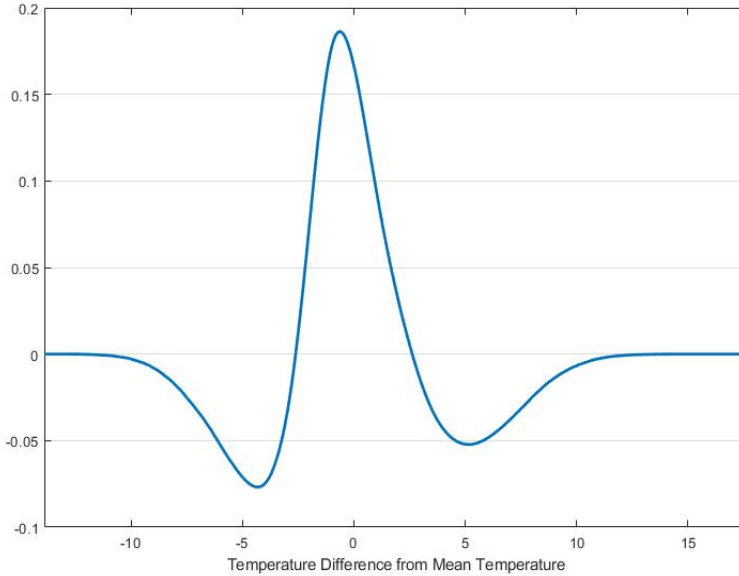
〈그림 9〉 함수적 주성분 분석의 cumulative scree plot



〈그림 10〉 주성분 분석 첫 번째 요인에 상응하는 정규직교 기저



<그림 11> 주성분 분석 두 번째 요인에 상응하는 정규직교 기저



낮은 온도의 빈도가 큰 날들, 즉 일교차가 큰 날들에서 높은 값을 갖는 요인으로 설명할 수 있다.

최종적으로 식 (7)에서 t+1기의 평년 기온분포로부터 벗어난 정도를 나타내는 함수 $y_{t+1}(s)$ 은 다음과 같이 예측된다.

$$y_{t+1}(s) = \sum_{k=1}^2 c_{k,t+1} v_k \quad (8)$$

여기서 주성분 벡터 v_1, v_2 는 시간에 따라서 고정되며, 좌표과정(coordinate process)으로 정의되는 $c_{k,t}$ 는 $c_{k,t} = \langle v_k, y_t \rangle$ 이므로 일별 기온분포 예측에서 최초(leading) 2개의 주성분의 좌표과정(coordinate process)을 예측하는 것이 가장 중요한 과정이 된다.

직교 기저의 특징에서 확인할 수 있듯이 각 요인은 일교차에 큰 영향을 받는다. 따라서 좌표과정을 예측하기 위한 모형은 다음과 같이 설정하였다.

$$\langle v_k, y_t \rangle = \alpha + \beta_1 x_{1t} + \beta_2 x_{2t} + \beta_3 x_{3t} + \epsilon_t \quad (9)$$

$$\widehat{y_{t+k}} = \langle v_1, y_{t+k} \rangle v_1 + \langle v_2, y_{t+k} \rangle v_2 \tag{10}$$

식 (9)에서 x_{1t} 는 일별 최저 기온 변수이고, x_{2t} 는 일별 평균 기온 변수이며, x_{3t} 는 일별 최고 기온에서 최저 기온을 제하여 생성한 일교차 변수이다. 설명 변수로 사용한 변수는 모두 기상청의 10일 중기예보를 활용하여 생성할 수 있다. 따라서 식 (9)로부터 $\langle v_1, y_{t+k} \rangle$ 와 $\langle v_2, y_{t+k} \rangle$ 를 예측하면 식 (10)과 같이 $t+k$ 일의 평년 기온분포로부터 벗어난 정도를 예측할 수 있고, 평년 기온분포와 더불어 $t+k$ 일의 기온분포¹⁷⁾를 예측할 수 있다.

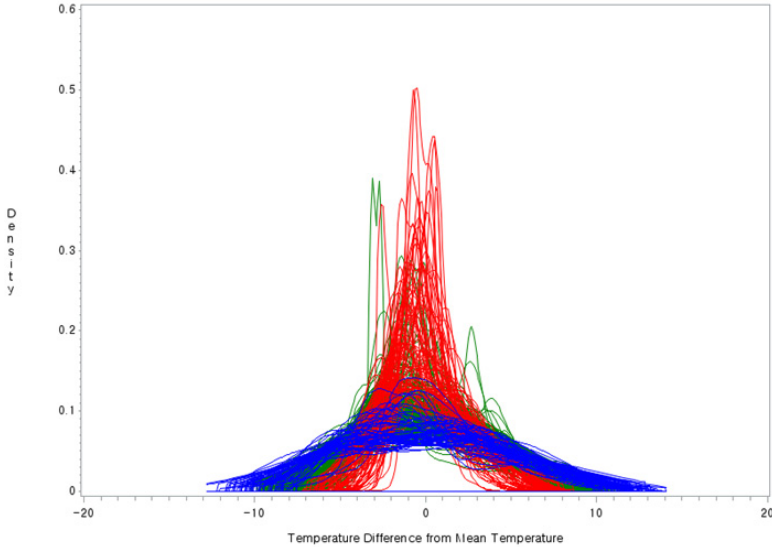
2007년 1월 1일에서 2019년 12월 31일까지의 기온자료를 활용해 식 (9)를 추정할 결과는 <표 2>에 나타내었다. 본 연구에서 제안한 설명변수들이 통계적으로 유의한 것으로 나타났다. <그림 12>는 우리나라 일별 기온분포의 특징에서 비롯되어, 본 연구에서 제안한 설명변수들이 통계적으로 유의한 것으로 나타난 배경에 관한 것이다. <그림 12>는 2019년 4월에서부터 2020년 3월까지의 일교차를 비교한다. 우리나라의 일별 기온분포는 평균의 차이를 제하더라도, 일교차의 변동이 크며, 일교차의 변동에 따라 우리나라의 일별 기온분포를 크게 일교차가 큰 날의 기온분포와 일교차가 작은 날의 분포로 구분할 수 있다. 일교차의 변동에 따라 일별 도시가스 수요에 변동이 크게 나타나므로 일별 기온분포를 예측함에 있어 일교차를 포함하는 것이 중요하다고 할 수 있다.

<표 2> 주성분 분석 회귀모형 추정 결과(괄호 안의 숫자는 t-value 의미)

	첫 번째 좌표과정 (Factor loading1)	두 번째 좌표과정 (Factor loading2)
상수항	0.537 (34.66)	0.145 (14.27)
일 최저 기온	0.221 (18.03)	-0.818 (-10.16)
일 평균 기온	-0.226 (-18.42)	0.079 (9.72)
일교차	0.060 (9.71)	-0.070 (-17.19)

17) 기온분포함수는 항상 양의 값을 가지므로, 0보다 항상 크도록 재조정하는 과정을 거쳤다.

<그림 12> 일별 기온 분포의 특징: 2019.04.~2020.03. 일교차 변동 비교



V. 예측 성과 비교

이 장에서는 기온분포예측 자체에 대한 예측력과 기온분포 예측을 활용한 도시가스 수요모형의 예측력을 검토한다.

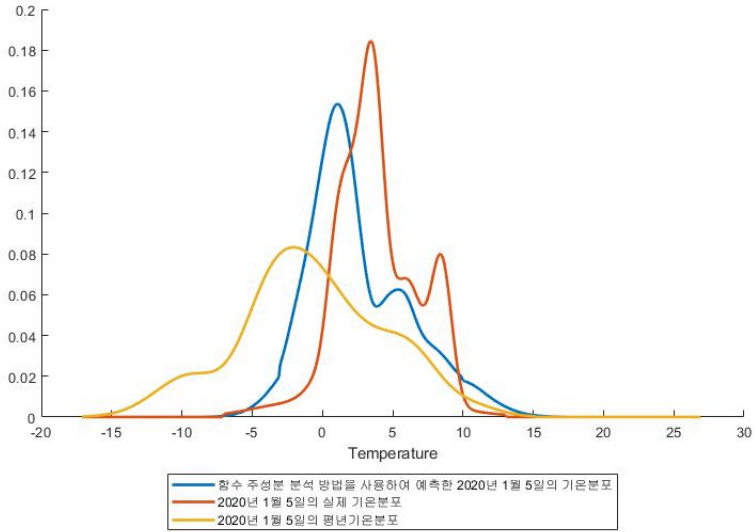
<그림 13>은 2020년 1월 5일의 실제 기온분포함수와 함수적 분석 방법을 사용해 예측한 기온분포함수, 그리고 1월 5일의 평년 기온분포함수를 비교한다. 2020년 1월 5일의 경우는 최저/최고 기온이 각각 8.6°C, -2.3°C로 나타나 일교차가 10.8°C에 이르는 등 상대적으로 큰 일교차가 발생한 날이다. 이처럼 일교차가 크게 벌어지는 날에는 실제 기온분포함수와 평년기온분포함수 간에는 비교적 큰 오차가 발생하게 되는데 이는 평년 기온분포함수가 일교차의 정도를 반영하지 못하기 때문이다.

<그림 14>는 2020년 1월과 2월의 총 60일의 기온분포를 1) 본 연구방식으로 예측¹⁸⁾ 하고 2) 평년 기온분포를 사용하여 실제 기온분포함수와의 일별 integrated squared error($\int (\hat{f}(s) - f(s))^2 ds$)를 비교한 것이다. 60일간의 MISE(mean integrated squared

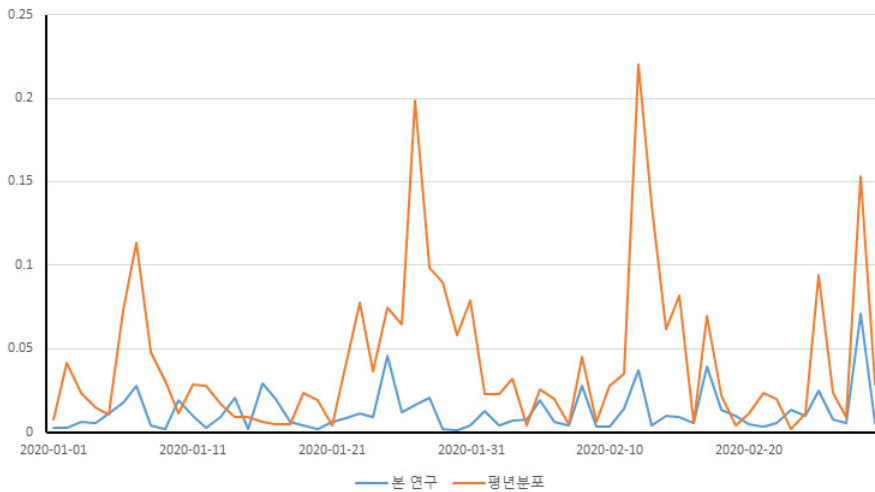
18) 5대 지역의 일 최대 최저 기온을 안다고 가정했을 경우이다.

error)는 각각 0.000197과 0.000695로 나타나서 본 연구의 방식이 기온분포를 훨씬 정확하게 예측한 것으로 볼 수 있다.

〈그림 13〉 기온분포함수 예측결과 비교



〈그림 14〉 기온분포 예측 결과: Integrated Square Error 비교



<표 3>은 앞서 생성된 기온분포를 이용해서 기온효과를 예측하고, 최종적으로 가스 수요를 예측한 결과를 정리한 표이다. 2020년 1월 1일에서부터 시작하여 2020년 2월 20일까지 향후 10일 간의 예측을 총 51회 수행하였다. 여기서 나온 결과를 바탕으로 평균 절대 백분율 편차(MAPE, Mean Average Percentage Error)를 예측 기간(horizon)별로 비교하였다. 합수 주성분 분석을 이용하여 기온분포를 예측하고 이를 일별 가스 수요 예측에 활용하였을 때, 평균적으로 오차가 5배 이상 줄어든 것으로 나타났다.

<표 3> 가스 수요 예측치 추정 결과-MAPE(%)

	합수적 주성분 분석을 이용해 예측한 기온분포 활용 시 가스 수요 예측치의 MAPE	평년평균 기온분포를 이용한 가스수요 예측치의 MAPE
1일 후 예측	7.96	16.00
2일 후 예측	5.15	21.89
3일 후 예측	2.97	24.09
4일 후 예측	3.58	25.10
5일 후 예측	4.45	24.74
6일 후 예측	4.12	24.52
7일 후 예측	2.63	23.99
8일 후 예측	3.05	23.51
9일 후 예측	5.08	23.26
10일 후 예측	6.18	23.27

VI. 결론

국내 도시가스 수요는 배관망 확충에 따라 급증하던 시기를 지나서 산업용 도시가스의 상대가격이나 기온에 의한 변동으로 결정되고 있다. 특히 일별 수요 예측의 경우는 기온에 따른 변동을 추정하고 예측하는 것이 예측력 제고에 가장 중요한 과업으로 보인다.

본 연구에서는 특수일효과, 기온효과, 누적 기온효과 등을 반영하여 일별 도시가스 수요 모형을 추정하였다. 또한 기상청의 중기예보를 이용해서 합수적 주성분 분석을 통해 기온분포 예측하고 이를 바탕으로 예측했을 경우에 평년 기온전제로 예측하는 것보다 예측오차가 1/5 수준으로 감소하는 것을 확인하였다.

도시가스용 일별 수요의 예측 정확도를 제고하기 위해서는 도시가스 수요의 기온에 대한 효과를 정확히 예측할 필요가 있다. 이를 위해서 기온에 대한 민감도를 면밀하게 분석하고 예측기온에 대한 예측을 정확하게 할 필요가 있다. 본 연구에서 제시하는 함수적 주성분 분석 방법론 이 외에도 머신러닝과 같은 첨단기법을 이용해서 기온분포를 예측 해보는 노력이 필요할 것이다.

[References]

- 김인무·김창식·박성근, “에너지 상대가격 변화에 따른 에너지 수요 예측”, 「경제학연구」, 제 59권 제1호, 한국경제학회, 2011. 3.
- 김점수·양춘승·박중구, “한국 도시가스용 천연가스의 소비함수에 대한 실증분석- 시간변동 계수(TVC) 시계열모형 활용-”, 「에너지공학」, 제20권 제4호, 한국에너지학회, 2011. 12.
- 박성근, “전력수요의 기온반응 변화 추정과 시사점”, KIET 산업경제, 2016. 5.
- 양준모·정민수·조인숙, “도시가스 수요예측 모형에 대한 소고”, 「Journal of the Korean Data Analysis Society」, 제19권 제6호, 한국자료분석학회, 2017. 12.
- 이근철·한정희, “선별적 샘플링을 이용한 국내 도시가스 일별 수요예측 절차 개발”, 「한국산 학기술학회논문지」, 제16권 제10호, 한국산학기술학회, 2015. 10.
- 이성로, “수도권과 지방권 수요예측모형을 통한 전국 도시가스수요전망의 예측력 향상”, 「자 원·환경경제연구」, 제 26권 제4호, 한국환경경제학회·한국자원경제학회, 2017. 12.
- 이승재·어승섭·유승훈, “시계열 자료를 이용한 도시가스의 수요함수 추정”, 「에너지공학」 제 22권 제4호, 한국에너지학회, 2013. 12.
- 한국전력거래소, “최근 이상기후를 반영한 냉난방부하 분석 및 예측 모형 개발”, 외부용역 연구보고서, 2020.
- Bosq, D., “Linear Processes in Function Spaces: Theory and Application,” *Springer-Verlag New York*, 2000.
- Chang, Y., Y. Choi, C. S. Kim, J. I. Miller, and J. Y. Park, “Forecasting Regional Long-Run Energy Demand: A Functional Coefficient Panel Approach,” Working Papers 1915, *Department of Economics, University of Missouri*, 2019.