

---

## Identifying Extreme Cold Events Using Phase Space Reconstruction

---

**Babatunde I. Ishola\*, Richard J. Povinelli,  
George F. Corliss, and Ronald H. Brown**

Department of Electrical and Computer Engineering,  
Marquette University,  
P.O. Box 1881, Milwaukee, WI 53201-1881, USA  
E-mail: isholabi@gmail.com  
E-mail: richard.povinelli@marquette.edu  
E-mail: george.corliss@marquette.edu  
E-mail: ronald.brown@marquette.edu  
\*Corresponding author

**Abstract:** Extreme cold events in natural gas demand are characterized by unusual dynamics that makes modeling the characteristics of the gas demand during extreme cold events a challenging task. This unusual dynamics is in the form of hysteresis, possibly due to human behavioral response to extreme weather conditions. To natural gas distribution utilities, extreme cold events represent high risk events given the associated huge demand of gas by their customers. To understand the nature of the unusual dynamics and help utilities in their decision-making process, we present a semi-supervised learning algorithm that identifies extreme cold events in natural gas time series data. Using phase space reconstruction, the input space is mapped into a phase space. In the reconstructed phase space, events with similar dynamics are closer together, while events with different dynamics are far apart. A cluster containing extreme cold events is identified by finding the nearest neighbors to an observed cold event. The learning algorithm was tested on natural gas consumption data obtained from natural gas local distribution companies. Our RPS- $k$ NN algorithm was able to identify extreme cold events in the data.

**Keywords:** Reconstructed Phase Space; Nearest Neighbor; Semi-supervised learning; Extreme Cold Events; and Energy Forecasting.

**Reference** to this paper should be made as follows: Ishola, B., Povinelli, R., Corliss, G. and Brown, R. (2016) 'Identifying Extreme Cold Events Using Phase Space Reconstruction', *Int. J. Applied Pattern Recognition*, Vol. x, No. x, pp.xx-xx.

**Biographical notes:** Babatunde Ishola is pursuing an M.S. degree in Electrical and Computer Engineering at Marquette University. He received his B.S. degree in Electronic and Electrical Engineering from Obafemi Awolowo University, Nigeria, in 2012. He is currently a supported Graduate Research Assistant on the GasDay Project at Marquette University.

Richard J. Povinelli received the B.S. degree in electrical engineering and B.A. degree in psychology from the University of Illinois, Champaign-Urbana, IL, USA, in 1987, the M.S. degree in computer and systems engineering from Rensselaer Polytechnic Institute, Troy, NY, USA, in 1989, and the Ph.D. degree

in electrical and computer engineering from Marquette University, Milwaukee, WI, USA, in 1999. From 1987 to 1990, he was a Software Engineer with General Electric (GE) Corporate Research and Development. From 1990 to 1994, he was with GE Medical Systems, where he served as a Program Manager and then as a Global Project Leader. From 1995 to 2006, he consecutively held the positions of Lecturer, Adjunct Assistant Professor, and Assistant Professor with the Department of Electrical and Computer Engineering, Marquette University, Milwaukee, WI, USA, where, since 2006, he has been an Associate Professor. His research interests include signal processing, machine learning, and chaos and dynamical systems. He has authored and coauthored over 60 publications in these areas.

George Corliss is professor of electrical and computer engineering at Marquette University and Senior Scientist of the GasDay Project. He received his B.A. in mathematics for the College of Wooster (Ohio) and his Ph.D. in mathematics from Michigan State University. He has taught and worked at the University of Nebraska, Lincoln, University of Wisconsin, Madison, Swiss Federal Institute of Technology, Argonne National Laboratory, Karlsruhe Institute of Technology (Germany), Compuware Corp., and Marquette University, as well as in several industrial and consulting positions. His research interests include scientific computation and mathematical modeling, guaranteed enclosures of the solutions of ordinary differential equations, industrial applications of mathematics and scientific computation, numerical optimization, automatic differentiation, and software engineering. He teaches courses in engineering design, computer architecture, operating systems, database design, and software engineering.

Dr. Ronald H. Brown is Associate Professor of Electrical and Computer Engineering at Marquette University and the founding Director of Marquette University's GasDay Project. Dr. Brown's research is in system modeling, identification, prediction, optimization, and control. The applications of his research has been focused on natural gas distribution and transmission since 1993, when the GasDay Project was founded as a means to connect students with the many industrial partners who support the lab's work. Over the course of the project he has worked with more than 150 undergraduate students from four colleges at Marquette directly participating in the project, and many more who have participated through classroom assignments that have "borrowed" project ideas from GasDay. He is a frequent presenter at energy industry meetings and consultant to many energy companies looking for guidance in planning for daily and peak load conditions.

---

## 1 Motivation

The most important days in natural gas demand forecasting include the days when demand is at its peak. It is important to forecast gas demand accurately during this period because it helps in infrastructure, supply, and operational planning (Lyness (1981)). Residential and commercial gas demand increases as the temperature decreases since homes and businesses use more natural gas for space heating as it gets colder (Vitullo et al. (2009)). The highest gas demand occurs during extreme cold events. An extreme cold event is a multi-day event for which the temperature is below a given threshold (specified by 1-in- $n$  years) for several consecutive days with a characteristic response in gas demand in the form of hysteresis (see Figure 1). A 1-in- $n$  temperature denotes the temperature which occurs as infrequently as once every  $n$  years. Extreme cold events are by nature rare, so they are not represented

adequately in gas demand data, leading to high forecast error during extreme cold events. Considering the financial implications as well as physical limits to the amount of gas supply that can be made available during an extreme cold event, it is important to identify extreme cold events in natural gas demand data. Identifying these events enables us to improve the gas demand forecast during such events, which may represent the most challenging days of the year for operational gas forecasters because their gas delivery systems are operating near their maximum capacities.

### 1.1 Behavioral Response

In addition to the infrequent nature of extreme events, they are also characterized by some interesting behaviors. Generally, gas demand varies linearly with temperature. For extreme cold events however, this relationship becomes non-linear. An unusual response in gas consumption in the form of hysteresis has been observed during the extreme cold events. Figure 1 shows the plot of daily natural gas consumption (flow) against wind-adjusted temperature (labeled *HDDW*), spanning a period of ten years. Figure 1b is a replica of Figure 1a with emphasis on the behavior of interest. The straight lines connect instances of natural gas consumption versus wind-adjusted temperature for five consecutive days. The days in the series identified by the lines represents the consumption for days  $t - 2$ ,  $t - 1$ ,  $t$ ,  $t + 1$ , and  $t + 2$ , with  $t$  being the coldest day in the event. The flow for the day after the coldest day ( $t + 1$ ) is much higher than the flow for day  $t$ , even though the temperature is warmer. Apparently, people tend to use more gas even when it is not as cold as the day before.

Part of this response is due to thermodynamic effects, as heat transfer is a dynamic process. There is a certain time-lag relating the reported (outside) temperature to the actual temperature (inside the building). The lag factor depends on the building's insulation system. Murat (2011) provides a good insight into the effect of thermodynamics on space heating in buildings. Attempts have been made to model the thermodynamics component by adjusting the forecast model for prior day weather effects as shown by Vitullo et al. (2009), and Brown and Matin (1995). The hypothesis here is that there is an unmodeled behavioral component, possibly due to human responses to extreme temperature and/or temperature changes (see Kalkstein et al. (1986), Brown (2014a), and Brown (2014b)), since the response to extreme cold events appears different from typical days.

In the next section, we will build a gas demand forecast model and observe the model's performance during extreme cold events.

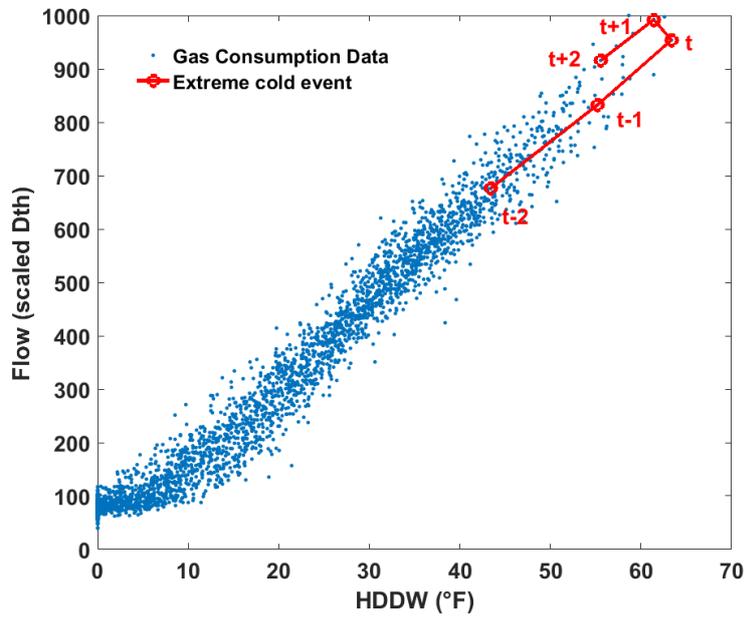
### 1.2 Gas Demand Forecast Model

This section describes a base line gas demand model that will be referenced throughout this paper. This base model is an ensemble of multiple linear regression (MLR) and artificial neural networks (ANN). The MLR model is a 13-parameter linear regression model

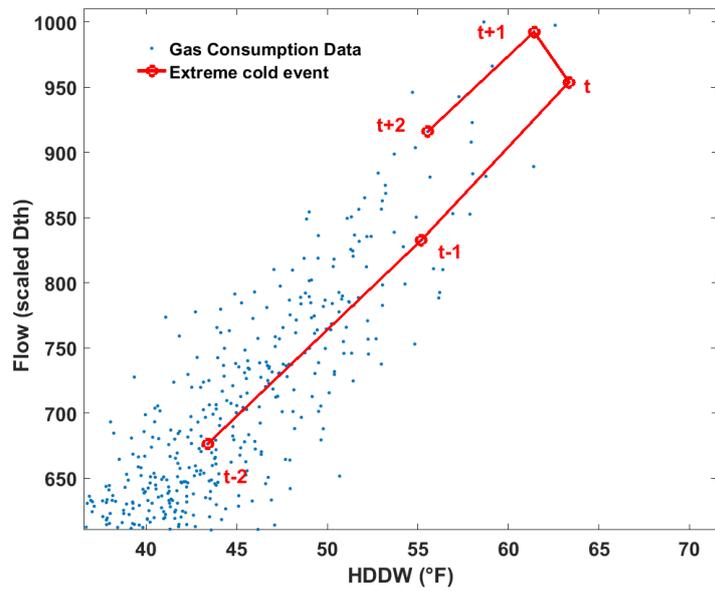
$$\hat{S}_t = \beta_0 + \sum_{i=1}^{13} \beta_i x_{i,t}, \quad (1)$$

$$S_t = \hat{S}_t + \epsilon_t, \quad (2)$$

where  $\hat{S}_t$  is the model estimate of gas demand for day  $t$ ,  $x_{i,t}$  represents input features such as temperature, prior day temperature, wind speed, day of week, and so on, with  $\beta$  being the



(a)



(b)

**Figure 1:** An extreme cold event in natural gas consumption data for a certain region in the USA. The extreme event identified can be seen to exhibit a hysteresis effect as a result of unusual (human behavioral) response to extreme temperatures. The plot in (b) is an enlarged version of (a) with focus on the extreme event.

model parameters. Let  $\epsilon_t$  be the forecast error for day  $t$ . Then the actual flow  $S_t$  is related to the estimated flow  $\hat{S}_t$  by Equation 2. The MLR component assumes a linear relationship between the dependent and independent variables, while the ANN model accounts for non-linear responses to the input features. The ANN model uses the same input features as the MLR model. The ensemble model was trained on historical data obtained from gas utilities in the USA. The learned model was used to estimate daily gas demand.

### 1.3 Base Model Performance

The base model described in Section 1.2 often over-forecasts and under-forecasts gas demand for days before and after the coldest day in an extreme cold event, respectively. Figure 2a shows an extreme cold event. If  $t$  is the index of the coldest day in the extreme cold event, on days  $t$ ,  $t - 1$ , and  $t - 2$ , the dashed line (base model estimate) is above the straight line (actual consumption), which means the demand forecast is more than the actual consumption for days before the coldest day. For days  $t + 1$  and  $t + 2$ , the dashed line is below the straight line, which means that the gas demand forecast is less than the actual consumption for days after the coldest day. This pattern of demand forecast and unusual response during extreme cold events has been observed for more than 20 operating areas (from different geographical locations), especially for those areas that have experience severe weather conditions in the past 10 years.

### 1.4 Quantifying Deviation

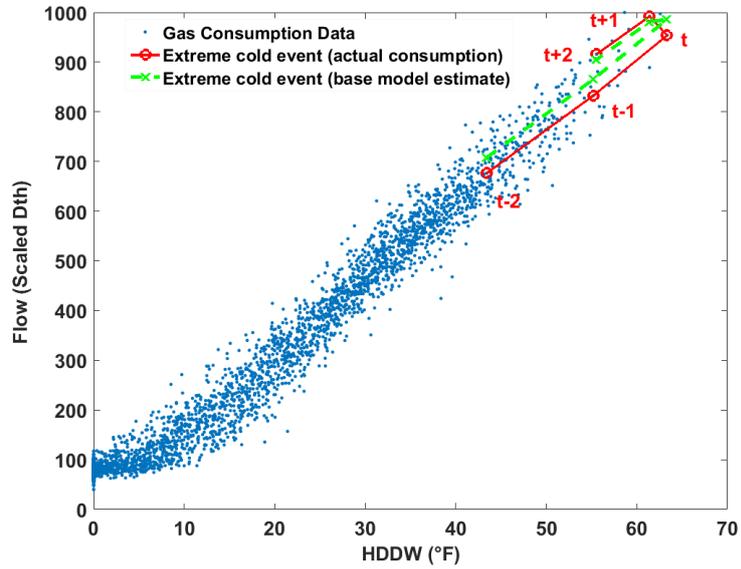
The work presented in this paper offers a strategy for adjusting the current base model estimate during extreme cold events to improve the accuracy of the gas demand forecast. Our strategy involves quantifying the deviation of the base model (which is a result of unmodeled behavioral components) from the actual demand during extreme events. A computational model is built based on the statistics of this deviation to estimate the forecast residual on extreme cold events. This is employed to estimate an adjustment to the base model.

To build a computational model that estimates forecast residual on extreme cold events, the extreme events are identified in the data. In Section 1.1, we postulated that extreme cold events have different dynamics than usual days due to the unusual behavioral response. In identifying extreme cold events, we search for events in the data with similar dynamics to a known extreme event. The events are treated as temporal patterns.

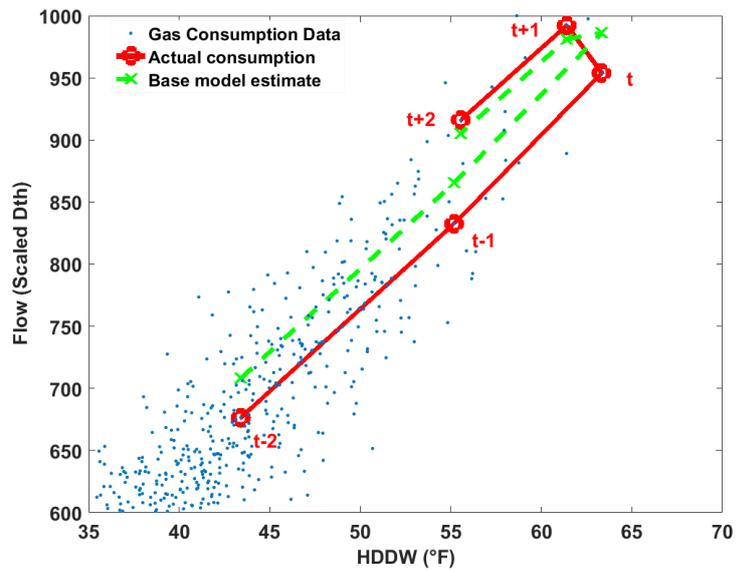
We identify temporal patterns in natural gas data that correspond to extreme cold events. This is achieved by clustering the data based on dynamics. Natural gas demand is a high dimensional system, so that events with similar dynamics may not occupy the same cluster in the input data space. For effective clustering, a low-dimensional embedding of the data is performed using phase space reconstruction (Povinelli et al. (2006)). In the reconstructed phase space, events with similar dynamics are closer to each other, while those with different dynamics are far apart. Extreme cold events are identified by finding the events that are close to a known extreme cold event in the reconstructed phase space, using a nearest neighbor algorithm.

### 1.5 Paper Overview

In the next section, we discuss important concepts on which the work presented in this paper is based, such as modeling non-stationary time series and phase space reconstruction for pattern recognition. In Section 3, we describe our approach to identifying temporal patterns



(a)



(b)

**Figure 2:** Performance of the base model on extreme cold events. For the extreme event shown, the base model (identified by the dashed line) over-forecasts gas demand for days before the coldest day  $t$  and under-forecasts for days after  $t$ . The plot in (b) is an enlarged version of (a) with focus on the extreme event.

of extreme cold events in natural gas time series data. Pseudocode also is presented. In Section 4, we discuss the performance of our approach and present results obtained when the algorithm was evaluated on six gas demand data sets from different gas utilities. As future work, we provide a brief overview on how the result of the identification is being used to estimate adjustments to gas forecast on extreme cold days.

## 2 Forecasting Non-stationary Data

In a non-stationary time series, the statistical properties of the underlying system vary over different regions of the data space. A common technique for forecasting such time series data involves building multiple models, with each model optimized for different regions of the data. The data space is partitioned into regions of similar dynamics using some clustering algorithm, and local models are learned for each identified cluster. The multiple local model approach often achieves higher forecasting accuracy than a single global model (Vilalta et al. (2010)). Global models are only well suited to stationary data, as they attempt to find an approximate representation of a system's dynamics (Pavlidis et al. (2006); Cao (2003)).

In financial forecasting, where exchange rates are highly correlated with economic, political, and psychological factors, all interacting in a highly complex manner, Pavlidis et al. (2006) employed clustering algorithms to partition the input data space into subspaces. Each subspace was learned using Feed-Forward Neural Networks (FFNN). Given test data, it was first determined to which cluster the data point belongs, and the corresponding FFNN was used to predict the exchange rate. Results reported show that the approach compares well with other established approaches. Cao (2003) employed a mixture of support vector machine (SVM) experts, with each expert optimized to forecast different regions of the input space. A self-organizing feature map was developed to cluster the input data space into several disjointed regions. With the partitioned regions having a more uniform distribution than the original input space, it becomes easier for the SVM experts to capture a stationary input-output relationship. The SVM expert that best fits a partitioned region is trained by finding the most appropriate kernel function and optimal free parameters of the SVM. Using three openly available data as test cases, Cao showed that for all the test cases, the mixture of SVM experts model achieves better performance than a single SVM model Cao (2003).

### 2.1 Clustering

The performance of the multiple model approach depends on the effectiveness of the clustering step. Clustering algorithms are used in data mining and pattern recognition tasks where items are to be separated into groups. Items in the same group are considered similar, with similarity defined only in the sense of the particular application. Metrics used in determining similarity include distance (i.e., how close the points are), density (i.e., how compact points are), and connectivity. When using a distance function as a similarity metric, it is possible for similar points to be far apart in the input data space, especially when dealing with high dimensional data. In high dimensional spaces, distances between points are relatively uniform, so the concept of closeness is meaningless (Steinbach et al. (2004)). In clustering such high-dimensional data, it is customary to perform a low-dimensional embedding, mapping the input data space into a new space where closeness is properly defined.

## 2.2 Phase Space Embedding

One common technique employed in low-dimensional embedding of high dimensional data is called phase space reconstruction. Phase space reconstruction is based on Takens' (1981) time-delay embedding theorem. Takens' theorem gives the condition under which a dynamical system can be reconstructed from a sequence of observations of the state of the system. Sauer et al. (1991) showed that for almost every time delay embedding with the appropriate selection of embedding parameters (dimension and time-lag), the reconstructed dynamics, with a probability of 1, are topologically identical to the true dynamics of the underlying system. Hence, the underlying dynamics of a system can be captured fully in a reconstructed phase space (RPS).

This technique is able to reconstruct the underlying dynamics of any complex system and map it into a new lower dimensional space. Since the RPS is equivalent to the true dynamics of the system, points with similar dynamics are guaranteed to be close in this space, while less similar points are far apart (Povinelli et al. (2004); Robinson (2005)).

## 2.3 Temporal Pattern Identification Using RPS

The RPS-based approach was demonstrated by Povinelli et al. (2006) to classify heart arrhythmia into one of four rhythms. An electrocardiogram signal was reconstructed in a phase space. The reconstructed phase was learned using a Gaussian Mixture Model (GMM) and classified using a Bayesian classifier. Povinelli et al. (2006) showed that the RPS-based approach outperformed other frequency-based methods with an accuracy of up to 95%, compared to the 44% accuracy of the frequency-based method.

While most of the existing applications of the RPS approach deal with univariate time series where the temporal pattern to be identified appears in the same feature space, the RPS approach can be extended to multivariate time series. Zhang and Feng (2012) in detecting sludge bulking, a primary cause of failure in water treatment plants, used an RPS-based approach to identify multivariate temporal patterns characteristics of sludge bulking in sludge volume index (SVI) and dissolved oxygen (DO) time series. The SVI and DO time series data are embedded in a multivariate RPS. The embedding dimension and time-lag for each signal was estimated using global false nearest-neighbors and first minimum auto-mutual information (Abarbanel (2012)). A mixture of Gaussian models is used to cluster the multivariate reconstructed phase space into three distinct classes. The result of the RPS-GMM approach was compared to other methods and was shown to perform better than both ANN and Time Series Data Mining (Povinelli et al. (2001)) approaches by at least 28%.

## 3 Identifying Extreme Cold Events

The techniques employed in identifying extreme events are similar to those described in Sections 2.1 through 2.3. This section discusses how the phase space reconstruction technique is applied to identify temporal patterns that correspond to extreme cold events in natural gas data.

Let an event be described as the dynamics between temperature and the corresponding natural gas demand over a series of five days. An event is classified as an extreme cold event if the pattern associated with the unusual behavioral response described in Section 1.1 is detected. The natural gas dataset is a multivariate time series consisting of two separate time

series; daily gas demand and daily temperature time series data. Let  $S_t$  represent natural gas consumption for day  $t$  and  $HDDW_t$  be derived from the corresponding (wind-adjusted) temperature. An extreme cold event is a multivariate temporal pattern, defined as

$$p = \{S_1, S_2, \dots, S_q; HDDW_1, HDDW_2, \dots, HDDW_q\}, \quad (3)$$

with  $p \in P \subseteq \mathcal{R}^{2q}$ ,  $q$  is the length of the temporal pattern.  $P$  represents the pattern cluster. Given a multivariate time series  $X = \{S(t); HDDW(t)\}, t = 1, 2, \dots, n$ , it is desired to identify all  $p \in P$ .

To identify all  $p \in P$ ,  $X$  is embedded in a multivariate reconstructed phase space in a way similar to Zhang and Feng (2012). Pattern cluster  $P$  is identified using a nearest neighbor algorithm in the reconstructed phase space.

### 3.1 Data Preprocessing

The datasets used in this work were obtained from natural gas utilities across the USA. This data has been anonymized to protect the identity of the utilities. Each dataset comprises ten years of actual gas consumption and weather data. The data is normalized prior to constructing a multivariate embedding. This ensures that  $S_t$  and  $HDDW_t$  are weighted equally in the reconstructed phase space such that the range of both  $S$  and  $HDDW$  is  $[0, 1]$ .

$$S_t = \frac{\max(S) - S_t}{\max(S)}, \quad (4)$$

$$HDDW_t = \frac{\max(HDDW) - HDDW_t}{\max(HDDW)}. \quad (5)$$

### 3.2 Multivariate Phase Space Embedding

The second step involves multivariate phase space embedding of the normalized time series data. According to Sauer et al. (1991), the appropriate selection of embedding parameters is necessary to ensure the reconstructed space is topologically equivalent to the original system. Takens' (1981) original work argued that choosing embedding dimension  $Q$  greater than  $2m + 1$ , where  $m$  is the dimension of the system's original state space, the time series can be completely unfolded in a phase space. Povinelli and Feng (1998), Abarbanel (2012) showed that useful information still can be extracted from the phase space by choosing a smaller  $Q$ . In most common applications (Povinelli and Feng (2003); Zimmerman et al. (2003); Povinelli et al. (2004, 2006); Zhang and Feng (2012)), time-lag  $\tau$  is estimated using the first minimum auto-mutual information, while dimension  $Q$  is estimated using the global false nearest-neighbor technique. In Povinelli and Feng (2003), embedding parameters were selected based on the of length of the temporal pattern vector to be identified.

Our selection of embedding parameters is application-specific. The dimension  $Q$  of the RPS and the time-lag  $\tau$  at which to sample the signal are selected based on our domain knowledge. The selection of  $\tau$  and  $q$  is based on the length of the temporal pattern vector to be identified. We are interested in bitter cold events about five days long, so the inter-relationship between flow  $S$  and wind-adjusted temperature  $HDDW$  for five consecutive days interests us. Multivariate embedding is done by augmenting individual univariate RPS.

Flow time series  $S(t)$  is embedded in a univariate RPS with time-lag  $\tau = 1$ , and dimension  $Q = q = 5$ .  $S$  maps into  $\mathbb{R}^q$ . The resulting phase space matrix

$$s = \begin{bmatrix} S_1 & S_2 & S_3 & S_4 & S_5 \\ S_2 & S_3 & S_4 & S_5 & S_6 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ S_i & S_{i+\tau} & \dots & S_{i+\tau(q-1)} & \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ S_{n-\tau(q-1)} & \dots & S_n & & \end{bmatrix}.$$

$HDDW(t)$  is embedded in a univariate RPS with  $\tau = 1$  and  $Q = q = 5$  in a way similar to  $S(t)$ . The resulting phase space matrix

$$hddw = \begin{bmatrix} HDDW_1 & HDDW_2 & HDDW_3 & HDDW_4 & HDDW_5 \\ HDDW_2 & HDDW_3 & HDDW_4 & HDDW_5 & HDDW_6 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ HDDW_i & HDDW_{i+\tau} & \dots & & HDDW_{i+\tau(q-1)} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ HDDW_{n-\tau(q-1)} & \dots & & & HDDW_n \end{bmatrix}.$$

The univariate phase space matrices  $s$  and  $hddw$  have equal sizes. A multivariate RPS is formed by augmenting  $s$  and  $hddw$  such that the resulting multivariate phase space matrix is

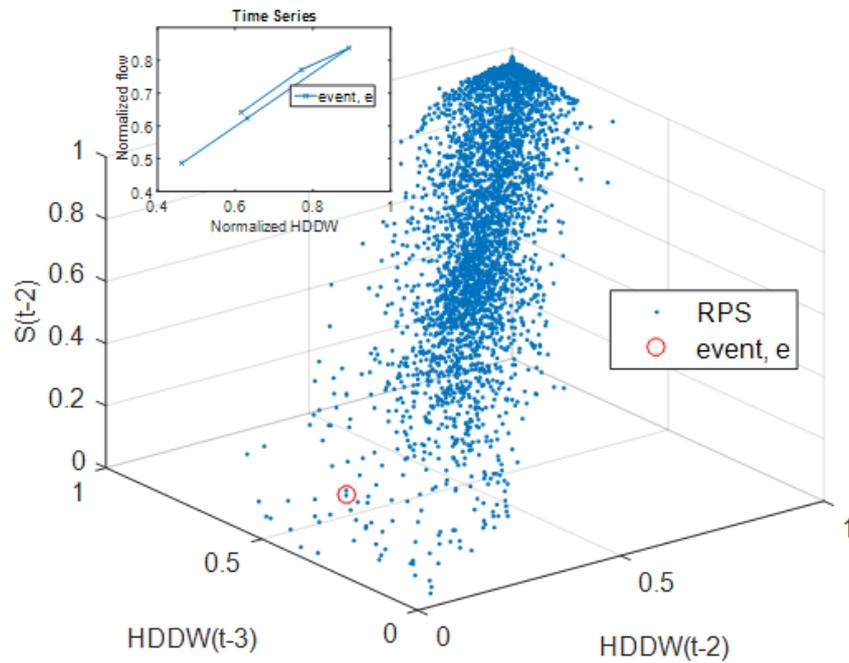
$$\begin{bmatrix} S_1 & S_2 & \dots & S_5 & & HDDW_1 & HDDW_2 & \dots & HDDW_5 \\ S_2 & S_3 & \dots & S_6 & & HDDW_2 & HDDW_3 & \dots & HDDW_6 \\ \vdots & \vdots & \vdots & \vdots & & \vdots & \vdots & \vdots & \vdots \\ S_i & & \dots & S_{i+\tau(q-1)} & & HDDW_i & & \dots & HDDW_{i+\tau(q-1)} \\ \vdots & \vdots & \vdots & \vdots & & \vdots & \vdots & \vdots & \vdots \\ S_{n-\tau(q-1)} & \dots & S_n & & & HDDW_{n-\tau(q-1)} & & \dots & HDDW_n \end{bmatrix}$$

The overall embedding dimension  $Q$  is the sum of the embedding dimensions of both variables, i.e.,  $Q = \sum_{i=1}^2 q = 10$ . Each row of the RPS matrix is a point in 10-dimensional space representing the dynamics of flow and temperature for five consecutive days.

Figure 3 shows a 3-dimensional projection of the 10-dimensional reconstructed phase space. Only three (namely  $S(t-2)$ ,  $HDDW(t-2)$ , and  $HDDW(t-3)$ ) of the 10 axes are shown for visualization purposes. Figure 3 also shows an event instance  $e$  in the time series and its corresponding mapping in the RPS. The event  $e$  shown in the time series plot has been reduced to a point in 10-dimensional space.

### 3.3 Nearest Neighbor Classifier

We desire to find the pattern cluster  $P$  that corresponds to extreme cold events. This is achieved by classifying events into one of two classes: normal and extreme cold events. Classification is done in the reconstructed phase space obtained in Section 3.2 using a



**Figure 3:** Reconstructed phase space built from natural gas consumption data. The overlaid plot (flow vs. temperature) is an event instance  $e$ . In the reconstructed phase space, the event instance  $e$  is represented by the circular marker. The reconstructed phase space is a 10-dimensional phase space with axes  $S(t), S(t-1), \dots, S(t-4)$  and  $HDDW(t), HDDW(t-1), \dots, HDDW(t-4)$ . The RPS plot shows only 3 of the 10 axes.

nearest neighbor (NN) algorithm. This is possible because closeness can be defined in this new feature space.

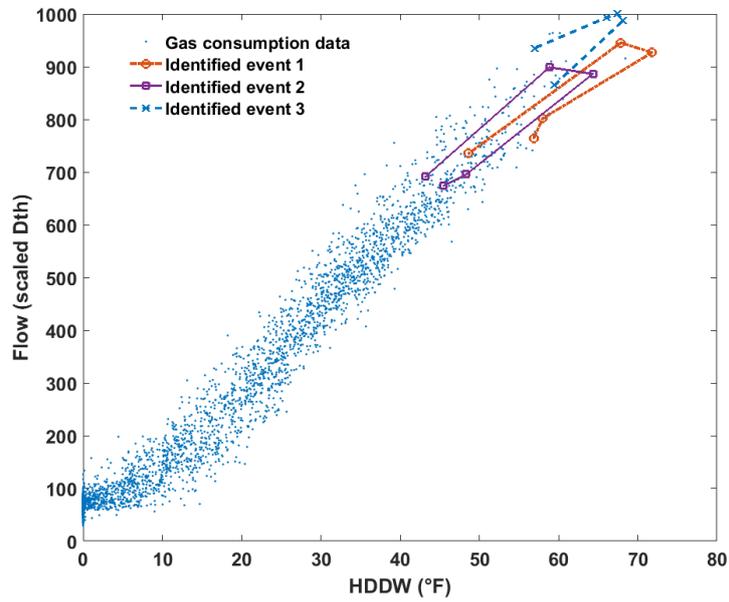
Nearest neighbor is a nonparametric classification method based on the measurement of a point's similarity to a training set containing patterns for which class labels are supplied. A nearest neighbor classifier is an instance-based learning algorithm, i.e., it does not build a model through learning, but rather aggregates the values provided by the training patterns in the vicinity of the current point. A  $k$ -Nearest Neighbor ( $k$ -NN) classifier assigns a label to a point  $x$  in the feature space based on the class assignment of its  $k$ -nearest neighbors. Decision is based on majority voting. This  $k$ -NN algorithm is supervised, requiring all training samples to have an assigned label. For an unsupervised task with unlabeled data, the  $k$ -NN algorithm no longer works. Identifying extreme events is an unsupervised task since there are no labeled datasets. To tackle the challenge of unlabeled data set, Povinelli et al. (2001) assigned class label to the training set by defining an event characterization function. In Liu et al. (2013), Liu used a semi-supervised  $k$ -NN employing instance ranking to deal with unlabeled data. To overcome the challenge of unlabeled data, we transform our unsupervised task into a semi-supervised one by assigning a class label to one of the data points. This point will be referred to as the pivot. The  $k$ -NN algorithm is modified to find the  $k$  nearest neighbors to the pivot point (inclusive). The  $k$  nearest neighbors discovered by this  $k$ -NN algorithm are assigned the same class label as the pivot. A known extreme cold event is chosen as the pivot, and the algorithm finds the  $k$  closest events to the extreme cold event. Closeness of a point (to the pivot) is determined by computing its Euclidean distance  $d(\text{pivot}, \text{event})$  from the pivot. The smaller the Euclidean distance, the higher the likelihood of the event being an extreme cold event and vice versa.

With the modified  $k$ -NN classifier described above, choosing the coldest event in the dataset as the pivot, the  $k$ -NN algorithm returns  $k$  events that have the same dynamics as the observed coldest event. The coldest event is found by manually searching the reconstructed phase space for the event with the max  $HDDW_{j+\frac{q-1}{2}}$  (i.e., lowest third day temperature for five-day events) and assigning it a class label: extreme event. Since the identification is done in the reconstructed phase space, the identified extreme events are mapped back to the original time series.

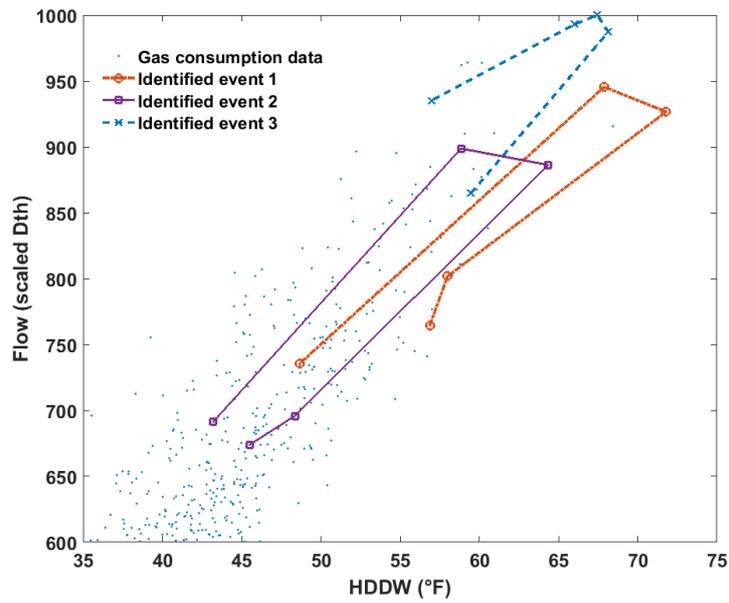
Figure 4 shows the flow and  $HDDW$  time series with extreme events identified by the  $k$ -nearest neighbor classifier. In Figure 4,  $k$  has been chosen as three for the purpose of presentation. Typical value of  $k$  might be about two events per year of available data. The event identified by the circular marker is the pivot (coldest) event. The box and 'X' markers represent the other extreme events identified by the algorithm having a similar 'unusual response' to the pivot event.

### 3.4 Algorithms

The pseudocode of the RPS- $k$ NN approach described in Sections 3.1 through 3.3 is provided in Algorithm 1. The **identifyExtremeColdEvents** function builds a multivariate RPS by merging two univariate RPS and calls the **classifyWithKNN** function to identify the extreme cold events. The **formUnivariateRPS** function builds individual RPS using the selected time lag  $\tau$  and dimension  $q$ .



(a)



(b)

**Figure 4:** Three extreme cold events that have been identified using our RPS- $k$ NN approach. The rightmost (coldest) event is chosen as the pivot. The two other events have been identified as the nearest neighbors to the coldest event in the reconstructed phase space. The plot in (b) is an enlarged version of (a) with focus on the extreme events.

---

**Algorithm 1** Reconstructed Phase Space - k Nearest Neighbor (RPS-kNN)
 

---

```

1: function identifyExtremeColdEvents(multivariateTimeseries, k)
2:   flow  $\leftarrow$  extract flow from multivariateTimeseries ▷ Preprocessing
3:   HDDW  $\leftarrow$  extract wind-adjusted temperature from multivariateTimeseries
4:   normalizedFlow  $\leftarrow$  normalize flow
5:   normalizedHDDW  $\leftarrow$  normalize HDDW

6:   choose timelag  $\tau$  and dimension  $q$  based on domain knowledge ▷ RPS
7:   rpsFlow  $\leftarrow$  formUnivariateRPS(normalizedFlow,  $\tau$ ,  $q$ )
8:   rpsHDDW  $\leftarrow$  formUnivariateRPS(normalizedHDDW,  $\tau$ ,  $q$ )
9:   rps  $\leftarrow$  merge rpsFlow and rpsHDDW to form a multivariate rps

10:  return extremeColdEvents  $\leftarrow$  classifyWithKNN(rps, k) ▷ Classification
11: end function

12: function formUnivariateRPS(data,  $\tau$ ,  $q$ )
13:   reconstructedPhaseSpace  $\leftarrow$  form a reconstructed phase space of data using the given
      $\tau$  and  $q$ 
14:   return reconstructedPhaseSpace
15: end function

16: function classifyWithKNN(rps, k)
17:    $x_i \leftarrow$  find coldest event and choose as pivot
18:   for each event  $x_j$  in rps do:
19:      $d(i, j) \leftarrow$  compute the Euclidean distance
20:   end for
21:   d  $\leftarrow$  sort(d, asc)
22:   indexes  $\leftarrow$  return the indexes of the first k elements
23:   return extremeColdEvents  $\leftarrow$  re-map indexes in the phase space to time series
24: end function

```

---

## 4 Discussion

The RPS- $k$ NN algorithm described in Algorithm 1 was tested on several datasets from different LDCs. Each dataset contains ten years of actual natural gas consumption data with the corresponding weather information. For each dataset, a multivariate reconstructed phase space is formed, and the coldest event in each dataset is chosen as the pivot. With the pivot chosen and  $k$  fixed, the  $k$ -nearest neighbor classifier returns the  $k$  most similar events to the coldest event. These events are considered to be the extreme cold events in the dataset. Figure 5 shows the events identified by the RPS- $k$ NN algorithm for six datasets. Only three of the identified events are shown, for the sake of presentation.

The performance of the classification stage will be determined by the contribution of the identified events to improving the estimate of gas forecast during extreme cold events.

### 4.1 Defining Bias

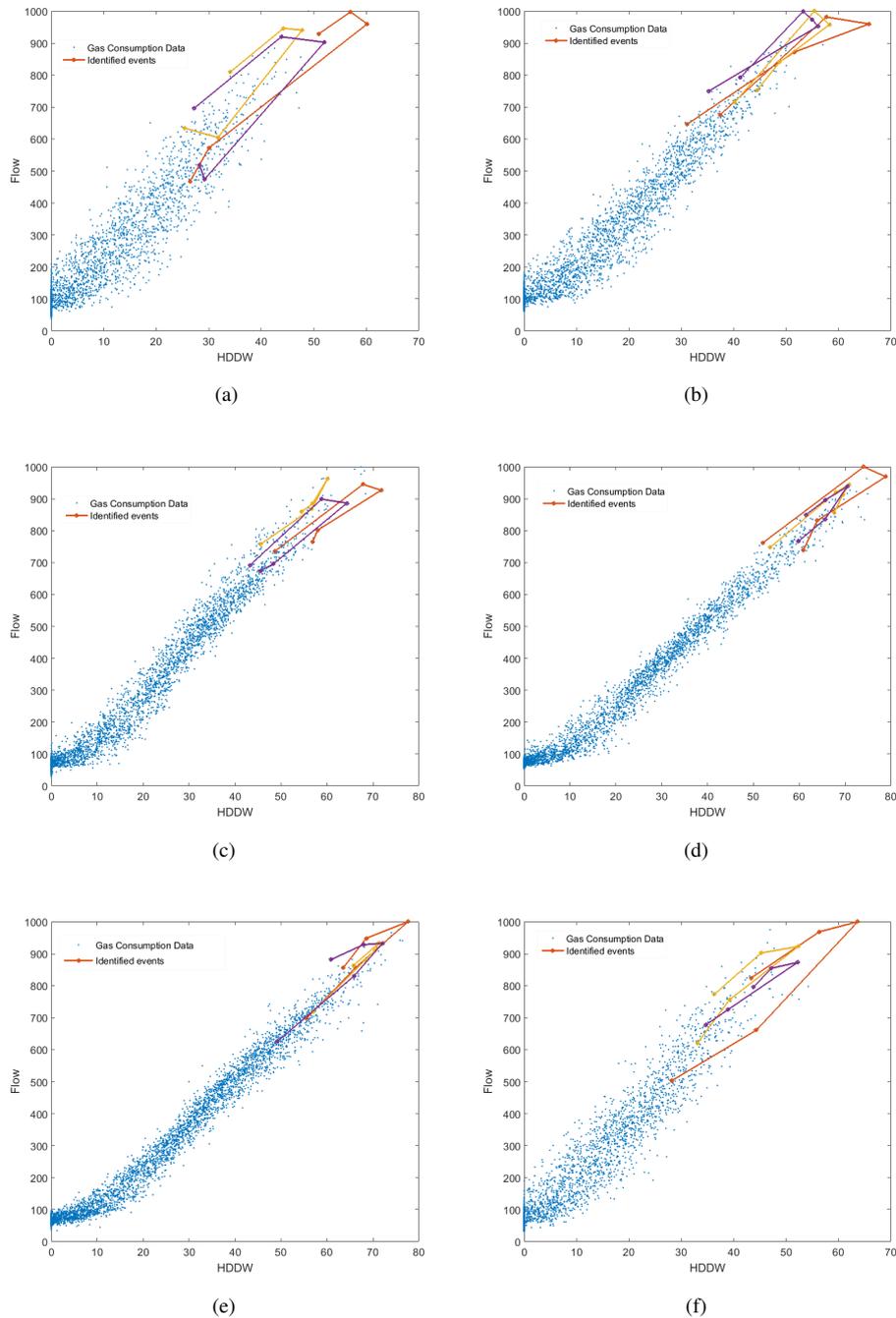
The effectiveness of the RPS- $k$ NN technique depends on the chosen value of  $k$ . As mentioned in Section 3.3, as  $k$  increases, the likelihood of event  $x_k$  having similar dynamics to the pivot (coldest extreme event) decreases. The  $k$ -nearest neighbor classifier identified  $k$  events that are closest to an observed cold event. If actual similar events are fewer than  $k$ , the classifier identifies  $k$  events regardless of how similar they actually are to the pivot event. Hence, the chosen value of  $k$  must be optimal.

In Section 1.4, we stated that part of the motivation for identifying extreme cold events is to build a computation model that estimates adjustment to gas demand forecast. The effectiveness of the computational model in estimating residual values depends on performance of our clustering. Because of this interdependence, we optimize  $k$  in the context of the residual analysis, i.e., we chose  $k$  such that identified events improve the estimate of gas demand during extreme cold events using an approach similar to Tongal (2014). Although the residual analysis is beyond the scope of this paper, we discuss briefly how this is achieved in Section 4.2.

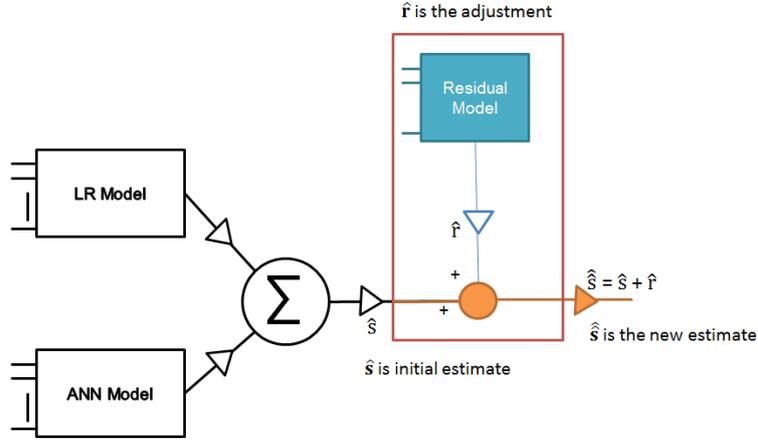
### 4.2 Optimizing $k$

Careful evaluation of the effectiveness of our identification of extreme cold events depends on the contribution of the identified events in improving the base model's estimate. In estimating adjustment to the base model (described in Section 1.2) during extreme cold events, temporal patterns in the dataset that are characteristics of extreme cold events are identified using the RPS- $k$ NN algorithm described in Algorithm 1. A residual learning model (shown in Figure 6) is built on the identified events. The residual model is a Partial Least Square (PLS) model trained to estimate base model's residuals  $\hat{r}$  for days in an extreme cold event. For the PLS model, points in the 10-dimensional reconstructed phase space were taken as predictor variables, with the base model's forecast residuals  $r = S - \hat{S}$  being the response variable. The estimated residuals  $\hat{r}$  were added as adjustment to the base models' flow estimate  $\hat{S}$  to form a new estimate  $\hat{\hat{S}}$  of gas demand i.e.,  $\hat{\hat{S}} = \hat{S} + \hat{r}$ .

The performance of the adjustment model is dependent on the value of  $k$  which is the number of identified extreme events. In determining the optimum value of  $k$  to be used in the RPS- $k$ NN procedure, the value of  $k$ , is varied from 5 to an arbitrarily large value (less than  $n$ ), and the corresponding Mean Absolute Percentage Error (MAPE) is evaluated on the validation data similar to Tongal (2014). We desire to find the  $k$  that minimizes



**Figure 5:** Events that have been identified as extreme cold events in natural gas consumption data using our RPS- $k$ NN approach. For each plot, only 3 events are shown for the sake of presentation. Plots (a) through (f) show the identification result obtained when the RPS- $k$ NN algorithm was executed on six datasets obtained from different natural gas local distribution companies in the United States. Each of the dataset used spans a period of ten years.



**Figure 6:** Adjustment model architecture for extreme cold events. Residual Model estimates the forecast residuals  $\hat{f}$  for days in an extreme cold event. A new estimate of gas demand is derived by adjusting the initial estimate  $\hat{S}$  with the residual estimate  $\hat{f}$ .

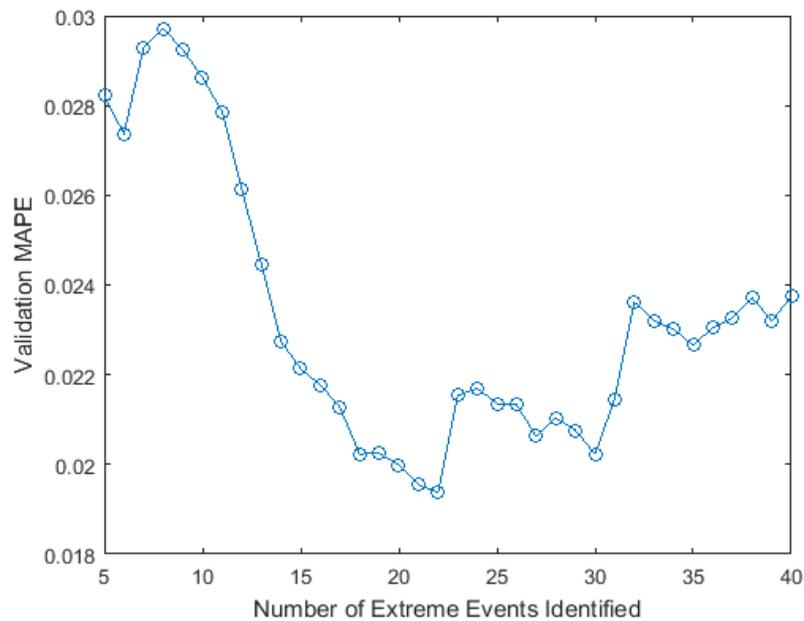
the validation MAPE between actual gas demand  $S$  and adjusted demand estimate  $\hat{\hat{S}}$ . The optimization problem is expressed as

$$\min_k \left( \frac{1}{k} \sum_{j=1}^k \frac{S_j - \hat{\hat{S}}_j}{S_j} \right). \quad (6)$$

Figure 7 shows the MAPE plotted against  $k$ , with  $k$  varying from 5 to 40. The MAPE decreases as  $k$  increases, until an optimum  $k$  is reached at  $k = 22$ , after which the MAPE starts to increase. The best value of  $k$  is set at 22 and the RPS- $k$ NN algorithm is executed again to identify the 22 useful (in the context of improving forecast during extreme cold events) extreme cold events in the data. Observe  $k = 22$  corresponds to about two extreme cold events per winter for each of the 10 years in the data set.

## 5 Conclusion

This paper introduced the problem of unusual response in natural gas demand consumption on a stretch of extremely cold days and how it affects natural gas forecasting accuracy. As a precursor to improving the forecast model, we have developed a semi-supervised learning algorithm to identify the subspace of the dataset that exhibits this unusual response. Low-dimensional embedding of the natural gas dataset was done using phase space reconstruction. A nearest neighbor algorithm was used in identifying temporal patterns relating to extreme cold events in the reconstructed phase space. The RPS- $k$ NN algorithm was tested on several datasets from different natural gas local distribution companies in United States to identify extreme cold events in each dataset. The results show that RPS provides a compact representation of the dynamics of a system, and in combination with  $k$ -nearest neighbor classifier, events of interest can be identified based on their dynamics.



**Figure 7:** Adjustment model validation error. Up till  $k = 22$ , the MAPE decreases as the number of events increases. For values of  $k$  above 22 the validation MAPE continues to increase. Optimum  $k$  is set at 22.

## References

- Lyness, F.K. (1981), 'Consistent forecasting of severe winter gas demand', *Journal of the Operational Research Society*, Vol. 32, pp. 347–359.
- Murat, O. (2011), 'Thermodynamic assessment of space heating in buildings via solar energy system', *Journal of Engineering and Technology*, Vol. 1, No. 1.
- Vitullo, S.R., Brown, R.H., Corliss, G.F. and Marx, B.M. (2009), 'Mathematical models for natural gas forecasting', *Canadian Applied Mathematics Quarterly* 17.
- Brown, R.H. and Matin, I. 1995, 'Development of artificial neural network models to predict daily gas consumption', *Industrial Electronics, Control, and Instrumentation, Proceedings of the 1995 IEEE IECON 21st International Conference on*, Vol.2, pp. 1389–1394.
- Kalkstein, L.S. and K. M. Valimont (1986), 'An evaluation of summer discomfort in the United States using a relative climatological index', *Bulletin of the American Meteorological Society*, Vol. 7, pp. 842-848.
- Brown, R.H. (2014), 'In search of the hook equation: Modeling behavioral response during bitter cold events', *2014 Gas Forecasters Forum*.
- Brown, R.H. (2014), 'Research results: The heck-with-it hook and other observations', *Southern Gas Association Conference: Gas Forecasters Forum*.

- Brown, R.H., Quinn, T.F., Corliss, G.F., Goldberg, J.R. and Nagurka, M. (2012), 'The GasDay Project at Marquette University: A Learning Laboratory in a Functioning Business', *2012 ASEE Annual Conference*.
- Povinelli, R. J., Johnson, M. T., Lindgren, A. C., Roberts, F. M. and Ye, J. (2006), 'Statistical Models of Reconstructed Phase Spaces for Signal Classification', *IEEE Transactions on Signal Processing*, Vol. 54, no. 6, June, 2178–2186.
- N. M. Public Regulation Commission informal task force investigation, 'Severe weather event of February, 2011 and its cascading impacts on NM utility service', [http://www.nmprc.state.nm.us/utilities/docs/2011-12-21\\_Final\\_Report\\_NMPRC.pdf](http://www.nmprc.state.nm.us/utilities/docs/2011-12-21_Final_Report_NMPRC.pdf), December 21, 2011.
- Hao, J., Arne, B. and Jerry, R.S. (2005), 'Modeling the Tail Distribution and Ratemaking: An Application of Extreme Value Theory', *American Agricultural Economics Association (New Name 2008: Agricultural and Applied Economics Association)*, 2005 Annual meeting, July 24-27, Providence, RI. No. 19190.
- Richard, L.S. and Ishay, W. (1985), 'Maximum Likelihood Estimation of the Lower Tail of a Probability Distribution', *Journal of the Royal Statistical Society. Series B (Methodological)*, Vol. 47, No. 2, pp. 285–298.
- D'Silva, A. (2015), 'Estimating the Extreme Low-Temperature Event Using Nonparametric Methods', *Master's thesis, Marquette University, Department of Electrical and Computer Engineering*.
- Dixon, P.M., Ellison, A.M. and Gotelli, N.J. (2005), 'Improving the precision of estimates of the frequency of rare events', *Ecology* 86:, 1114–1123.
- Brown, R.H., Vitullo, S.R., Corliss, G.F., Adya, M.P., Povinelli, R.J. and Kaefer, P.E. (2015), 'Detrending daily natural gas consumption series to improve short-term forecasts', *IEEE Power and Energy Society general meeting (PESGM)*
- Robinson, J.C. (2005), 'A topological delay embedding theorem for infinite-dimensional dynamical systems', *Nonlinearity*, Vol. 18 No. 5, pp. 2135.
- Takens, F. (1981), 'Detecting strange attractors in turbulence', *Springer Lecture Notes in Mathematics*, Vol. 898, pp 366–381.
- Sauer, T., Yorke, J. A. and Casdagli, M. (1991), 'Embedology', *Journal of Statistical Physics*, 65(3-4), pp 579–616.
- Lindgren, A. C., Johnson, M. T. and Povinelli, R. J. (2003), 'Speech Recognition using Reconstructed Phase Space Features', *International Conference on Acoustics, Speech and Signal Processing*, Vol. I, pp. 61–63.
- Pavlidis, N. G., Plagianakos, V. P., Tasoulis, D. K. and Vrahatis, M. N. (2006), 'Financial forecasting through unsupervised clustering and neural networks', *Operational Research*, 6(2), pp. 103–127.
- Povinelli, R. J. (2001), 'Identifying temporal patterns for characterization and prediction of financial time series events', *In Temporal, Spatial, and Spatio-Temporal Data Mining. Springer Berlin Heidelberg.*, pp. 46–61.

- Bernad, D. J. (1996), 'Finding patterns in time series: a dynamic programming approach', *Advances in Knowledge Discovery and Data Mining*.
- Weiss, S. M. and Indurkha, N. (1998), 'Predictive data mining: a practical guide', *Morgan Kaufmann*.
- Han, J. and Kamber, M. (2001), 'Data mining: concepts and techniques', *United States of America: Morgan Kauffmann Publishers*.
- Keogh, E. J. and Smyth, P. (1997), 'A Probabilistic Approach to Fast Pattern Matching in Time Series Databases', *In KDD*, Vol. 1997, pp. 24–30.
- Guralnik, V., Wijesekera, D. and Srivastava, J. (1998), 'Pattern Directed Mining of Sequence Data', *In KDD*, pp. 51–57.
- Rosenstein, M. T. and Cohen, P. R. (1999), 'Continuous categories for a mobile robot', *In AAAI/IAAI*, pp. 634–640.
- Povinelli, R. J. and Feng, X. (2003), 'A new temporal pattern identification method for characterization and prediction of complex time series events', *Knowledge and Data Engineering, IEEE Transactions*, Vol. 15, pp. 339–352.
- Liu, Zun-Xiong (2005), 'Short-term load forecasting method based on wavelet and reconstructed phase space', *In Machine Learning and Cybernetics. Proceedings of 2005 International Conference* Vol. 8, pp. 4813–4817.
- Povinelli, R. J., Johnson, M. T., Lindgren, A. C. and Ye, J. (2004), 'Time series classification using Gaussian mixture models of reconstructed phase spaces', *Knowledge and Data Engineering, IEEE Transactions.*, 16(6), pp. 779–783.
- Zhang, W. and Feng, X. (2012), 'Predictive temporal patterns detection in multivariate dynamic data system', *In IEEE Intelligent Control and Automation (WCICA), 2012 10th World Congress* pp. 803–808.
- Zimmerman, M. W., Povinelli, R. J., Johnson, M. T. and Ropella, K. M. (2003), 'A reconstructed phase space approach for distinguishing ischemic from non-ischemic ST changes using Holter ECG data', *In Computers in Cardiology* pp. 243–246.
- Abarbanel, H. (2012), 'Analysis of observed chaotic data', *Springer Science and Business Media*
- Povinelli, R. J. and Feng, X. (1998), 'Temporal pattern identification of time series data using pattern wavelets and genetic algorithms', *In Artificial Neural Networks in Engineering* pp. 691–696.
- Steinbach, M., Ertöz, L. and Kumar, V. (2004), 'The challenges of clustering high dimensional data', *In New Directions in Statistical Physics. Springer Berlin Heidelberg*. pp. 273–309.
- Tongal, H. (2014), 'Nonlinear forecasting of stream flows using a chaotic approach and artificial neural networks', *Earth Sciences Research Journal*. Vol. 17, No.2.

- Vilalta, R., Ocegueda-Hernandez, F. and Bagaria, C. (2010), 'A conceptual study of model selection in classification', *International Conference on Agents and Artificial Intelligence* Vol. 1.
- Liu, Z., Zhao, X., Zou, J. and Xu, H. (2013), 'A semi-supervised approach based on k-nearest neighbor', *Journal of Software* Vol. 8, No.4, pp. 768–775.
- Kim, M., Kim, Y., Kim, H., Piao, W. and Kim, C. (2015), 'Evaluation of the k-nearest neighbor method for forecasting the influent characteristics of wastewater treatment plant', *Frontiers of Environmental Science and Engineering* pp. 1–12.
- Cao, L. (2003), 'Support vector machines experts for time series forecasting', *Neurocomputing* Vol. 51, pp. 321–339.