Forecasting Supply Chain sporadic demand with Nearest Neighbor approaches

Nikolopoulos, Konstantinos; Babai, Mohammed Zied; Bozos, Konstantinos

International Journal of Production Economics

DOI:
10.1016/j.ijpe.2016.04.013

Published: 01/07/2016

Peer reviewed version

Forecasting Supply Chain sporadic demand with Nearest Neighbor approaches

Konstantinos I. Nikolopoulos[1,*], M. Zied Babai[2], Konstantinos Bozos[3]

[1] *forLAB, Bangor Business School, Prifysgol Bangor University, Bangor, Gwynedd, United Kingdom; k.nikolopoulos@bangor.ac.uk*

[2] *Kedge Business School, Bordeaux, France; mohamed-zied.babai@kedgebs.com*

[3] *Leeds University Business School, University of Leeds, Leeds, United Kingdom; k.bozos@lubs.leeds.ac.uk*

Abstract

One of the biggest challenges in Supply Chain Management (SCM) is to forecast sporadic demand. Our forecasting methods' arsenal includes Croston's method, SBA and TSB as well as some more recent non-parametric advances, but none of these can identify and extrapolate patterns existing in data; this is essential as these patterns do appear quite often, driven by infrequent but nevertheless repetitive managerial practices. One could claim such patterns could be picked up by Artificial Intelligence approaches, however these do need large training datasets, unfortunately non-existent in industrial time series. Nearest Neighbors (NN) can however operate in these latter contexts, and pick up patterns even in short series. In this research we propose applying NN for supply chain data and we investigate the conditions under which these perform adequately through an extensive simulation. Furthermore, via an empirical investigation in automotive data we provide evidence that practitioners could benefit from employing supervised NN approaches. The contribution of this research is not in the development of a new theory, but in the proposition of a new conceptual framework that brings existing theory (i.e. NN) from Computer Science and Statistics and applies it successfully in an SCM setting.

Keywords: Supply Chain; Logistics; Demand Forecasting; Exponential Smoothing; Nearest Neighbors;

[*] Corresponding author: 0044 7981 332913

Forecasting Supply Chain sporadic demand with Nearest Neighbor approaches

1. Introduction

One of the biggest challenges in Operations Management (OM) is to forecast as accurately as possible the sporadic/intermittent demand faced in most supply chain and logistics operations. Assuming we are not in a made-to-order supply context, then forecasters are facing true uncertain stochastic demand. The reason that makes the forecasting task so challenging has to be attributed to the dual uncertainty faced by decision makers due to the sporadic nature of both the actual demand volume as well as the timing of demand arrivals; the latter do not occur during each and every period of time and as such there are many periods in time with zero demand.

Especially for logistics operations where the spatial dimension comes into the equation, the appearance of sporadic demand is more often, as prima facie regular demand at the aggregate manufacturing or distribution-centre level becomes intermittent when demand is realised through the alternative geographical channels. Furthermore the alternative direct retailing channels through online and digital media create more 'channel products' for which demand needs to be forecasted and stock to be replenished.

There have been a few forecasting methods tailored for such data that have been developed over the last forty years (Petropoulos et al. 2014), mostly exponential smoothing approaches like Croston's method (1972), Syntetos and Boylan approximation–SBA (2001) and more recently a method developed from Teunter, Syntetos and Babai–TSB (2011). A handful of more advanced but at the same more complex non-parametric methods has also been proposed over the years but these do not necessarily outperform the former (Syntetos et al. 2015).

In general the assumption is that the aforementioned intermittent series present no trend or autocorrelation, no seasonality or cycles - in essence none whatsoever structural component that could be identified via a formal statistical procedure. However, this for practitioners just does not make sense as they do know that very often they apply the same tactics when they order, replenish or produce products and services; it is just that these are happening in non-periodic lags and thus becomes very difficult to spot them in the past of a time series. Attesting to that, Altay with Litteral and Rudisill have provided evidence of existence of trend (2008) and correlation (2012) in intermittent data and proved that these features significantly affect forecasting and stock control performance.

Unfortunately, none of the proposed to date methods can pick up one-off (or more frequent) patterns existing in the past of a series, patterns that may appear quite often due to irregular but repetitive managerial practices. Artificial Intelligence and other computationally intensive approaches (Haykin, 1998) that could pick up such patterns have been successfully applied in other disciplines like finance, economics, marketing and computer science. These methods however do rely on very large training datasets in order to predict satisfactorily, thus are not fit-for-purpose in our context where lengthy series are perceived as quite a luxury. Nevertheless there is one non-parametric regression smoothing method that does not require large training data sets: Nearest Neighbors (NN), which are capable of producing forecasts even when only just two or three observations are available.

In this research we propose applying NN for supply chain data and we investigate the conditions under which these perform adequately through an extensive simulation on 480 time series where we control for intermittence levels, frequency, plurality and type of patterns, and level of noise. Furthermore via an empirical investigation in automotive data over 3000 SKUs we provide evidence that practitioners could benefit from employing supervised NN approaches - supervised in the sense that you only allow the patterns to be identified if there is some evidence that these patterns do exist - this in this paper is achieved through an adjusted cumulative autocorrelation function.

The contribution of this research is not in the development of a new theory, but in the proposition of a new conceptual framework that brings existing theory (i.e. NN) from Computer Science and Statistics and applies it successfully in an SCM setting. The proposed framework comes with a series of limitations as it should be applied only when evidence of existence of patterns is apparent, and even so not left unsupervised rather than applied only in the part of the dataset that exhibits higher autocorrelations. Ceteris paribus, we still believe that we live in a world where humans drive data, through their action and behaviours eventually do introduce infrequent patterns in time series data; and as such we do believe there is merit for practitioners and academics to consider our proposition and further research it in the future.

The remaining of the paper is structured as follows: section two revisits the background literature while section three presents the conceptual framework of the NN method and gives an illustrative example of how the method could be used in a SCM context; section four compares the newly proposed method with other classical parametric approaches. Section five gives the results of an extensive empirical evaluation across a real large dataset from RAF. Section six discusses the implications for SCM theory and practice, while the last section concludes and provides the roadmaps for future research.

2. Background Literature

The first part of the literature review revisits the most important methods for forecasting intermittent demand. The second part focuses on successful applications of Nearest Neighbors in various contexts.

2.1. Forecasting Supply chain sporadic demand

For intermittent data, simple techniques such as Naïve, Moving Averages and Simple Exponential Smoothing (SES) have been quite popular among practitioners over the years due to their simplicity and accuracy but most importantly the ability to handle non-demand observations without the need of time series transformations (Petropoulos et al. 2013). The first method tailored to intermittent data came from John Croston in his seminal article Croston (1972) proposing a decomposition of the data into two subseries, one for the positive demands (excluding the zeros) and one for the arrival intervals. Syntetos and Boylan proposed (2001) and evaluated successfully (2005) SBA: a bias-correction approximation to the Croston method. More recently, Teunter et al. (2011) suggested the TSB decomposition method that builds on the separate extrapolation of the non-zero demands and the probability to have a demand; with this method being very useful in cases of obsolescence.

A handful of non-parametric methods has been proposed over the years, most notably from Willemain et al. (2004) with a bootstrapping approach that captures potential auto-correlations of the underlying demand patterns and simultaneously accounts for variability not observed in the original demand sample through the patented 'Jittering' process. AI approaches have been also proposed; most often Artificial Neural Networks (Gutierrez et al 2008) but these do need very large training datasets. In any case, there is no sufficient empirical evidence that these methods are more accurate from the simpler ones (Syntetos et al. 2015).

More recently, Nikolopoulos et al. (2011) proposed the ADIDA non-overlapping temporal aggregation forecasting framework that was successfully evaluated both in terms of forecasting accuracy as well as of stock control performance (Babai et al., 2012). The proposed framework is now perceived as a "forecasting-method improving" mechanism that through frequency transformations helps methods achieve better accuracy performance. The first theoretical developments for the framework appeared recently in the literature (Spithourakis et al., 2014; Rostami-Tabar et al., 2013, 2014). Kourentzes et al. (2014) extended this idea by means of estimating time series structural components across multiple frequencies and optimally extrapolate and combine them; with empirical results being quite promising for long-term forecasting. Petropoulos and Kourentzes (2014) also proposed forecasting method combinations on the aforementioned context with improved forecasting performance.

2.1 Nearest Neighbors

Nearest Neighbor approaches (NN/NNs, Härdle[1] 1992) are quite popular in the forecasting literature (Green 2002), largely because of their intuitively appealing simplistic nature and theoretical attributes (Yakowitz 1987). They are generally found to present distinct advantages versus their alternatives for non-linear fluctuations, since, while their parameters are linear in nature, NNs can capture the complex non-linear patterns among neighbours (Yankov et al. 2006), thus predicting composite non-linear behaviors in a fairly accurate manner. Hence, NN methods have been applied to time-series and cross-sectional data in a wide variety of domains and have often been found to outperform alternative, most of the time far more complex approaches. However to the best of our knowledge these have never been applied in an SCM sporadic demand context. We hereafter review some of the most indicative works in the most common domains.

Economics & Finance

In his attempt to beat the 'tenacious' random walk model in forecasting exchange rates, Mizrach (1992) combines k-NN estimators (using the outcomes from 'k' closest past cases=the neighbors) with a locally weighted regression procedure. He finds an improvement in the forecast accuracy in just one out of the three rates examined, a result which is however not robust to further empirical scrutiny. Along the same lines, Jaditz and Sayers (1998), compare out-of-sample forecasts between a best fitting global linear model and a local-information nearest-neighbor forecasting methodology on money supply data. They report just marginally smaller Root Mean Squared Errors for NNs than those from simple linear autoregressive (AR) models.

[1] *This is by many academics considered the most complete reference book describing the theoretical foundations as well as many practical applications of Nearest Neighbors for smoothing, classification and forecasting.*

More recent studies have offered considerable merit for the use of NN estimates in financial forecasting. In an attempt to predict short term Foreign Exchange rates, Meade (2002) compares the accuracy of a linear AR-GARCH model (Engle, 1982) to three NN methods and locally weighted regressions. He reports higher accuracy for the NN approaches, which is also found to improve, as the data frequency increases from daily to half-hourly.

In another study of US listed company share prices and dividends, Kanas (2003) compares the forecasts generated by parametric standard and Markov regime switching models with the ones produced by a simultaneous (multivariate) nearest-neighbour (SNN) approach and ANNs and finds SNN to perform similarly to the others in terms of accuracy, but not as well as the Markov switching model in terms of forecast encompassing. In a recent study of short term stock price reactions to equity offering announcements, Bozos and Nikolopoulos (2011) compare forecasts by parametric, non-parametric models and expert judges and conclude that k-NN methods ranked high in terms of economic performance of the forecasts, despite their forecast accuracy being relatively lower.

Analytics (Marketing, Logistics)

Similar attention has been paid by researchers in the broader areas of commerce, business and communications: In their competition among a large set of techniques employed to forecast TV audience ratings and respective advertising spending around special events, Nikolopoulos et al. (2007) report that a 3-NN model outperformed multiple linear regression and the more computationally-demanding neural networks ANNs. They attribute the superiority of the k-NN model to its ability to identify more complex interactions across the input variables and filter out more noise in the observations. In another example from marketing literature Mulhern and Caprara (1994) combine a multivariate NN model with regression analysis to forecast market response, using store scanner data for a consumer packaged good. Their results suggest that such an approach has obvious advantages versus the more traditional Box-Jenkins analysis as it allows time effects (traditionally filtered out in ARIMA models) to be integrated into the causal relationships.

In the field of logistics Smith et al. (2002), combine k-NN methods with heuristically improved forecast generation methods, to predict 15-minute traffic flow rates from the outer loop of the London's Orbital Highway M25. They report that their k-NN nonparametric regressions outperform naïve forecasts, but not those of the seasonal ARIMA models. Along similar lines, Sun et al. (2003), using 5-min interval traffic data from the US-290 highway in Texas, find k-NN methods to perform better than kernel smoothing, but not as well as local linear methods.

Natural Systems

In the seminal collection of forecasting competitions during the early 1990s by the Santa Fe Institute (Weigend and Gershenfeld 1994), with datasets drawn from a range of disciplines (e.g. laser lab results, physiological data, astrophysical data, etc.), k-NN methods were found to perform very well in terms of prediction error, even among more computationally-sophisticated and data-intensive methods. Specifically, Sauer's (1994) 4-NN local linear fitting algorithm ranked just second among fourteen competing approaches for a univariate dataset generated by a far-infrared-laser in a chaotic state. k-NN approaches have also been successfully applied in a variety of forecasting extreme weather-related phenomena (Brabec and Meister 2001).

3. A conceptual framework for the application of k-NN in SCM sporadic time series

NN approaches are a very simple way for forecasting the impact of future events by looking at the past for the impact of similar or analogous events (Green, 2002). In pattern recognition k-NN is a non-parametric method for classification that predicts class memberships based on the k closest training examples – the neighbors. In a time series extrapolation context, finding the neighbours requires ranking in terms of similarity historic sequences of observations, to the most recent sequence of observations; in the same fashion that the autocorrelation function ACF does. Similarity ranking of course necessitates the use of a metric of distance; various metrics such as Euclidian norms have been used in order to measure distance in the n-D space (Hardle, 1992). n-D space metrics are required since our sequences could involve several (n) observations

3.1 The framework

For a time series Y_t, t=1..N , where $Y_N \neq 0$[2], let $Now_{(N-l+1,\ N)}$ be a row-vector of length l that consists of the latest sequence on non-zero observations starting from the last observation N and moving backwards to N-l+1.

Let $Neighbor_{(N-Lag-l+1,\ N-Lag)}$ be any row-vector of the same length l taken from the past of the series when we move backwards for a specific *Lag*. This latter vector is allowed to contain zero values.

In order to measure the *distance D* of the neighboring vector (noted as *Neighbor*) from the present one (noted as *Now*), we do use the standard Euclidean norm, as follows:

$$D = ||Now_{(N-l+1,\ N)} - Neighbor_{(N-Lag-l+1,\ N-Lag)}|| = \sqrt{\left(\sum_{i=0}^{l-1}(Y_{N-i} - Y_{N-Lag-i})^2\right)} \tag{1}$$

The bigger the value of D the less similar the respective neighbour is to the present vector. D is also called the *neighboring function.* Distances are calculated for all possible neighbors from the past of the series and a *ranking* is created from the Nearest Neighbor to the less similar ones. Ties are being allowed as many neighbors can actually have exactly the same distance from the present vector.

If $Neighbor_{(N-Lag-l+1,\ N-Lag)}$ is the Nearest Neighbor NN to the present vector $Now_{(N-l+1,\ N)}$ according to the constructed ranking then:

$$F_{N+1} = Y_{N-Lag+1} \tag{2}$$

would be the one-step ahead forecast for the NN1 method that uses only one neighbour. The flexibility of the NN approaches allows us to use as many neighbors as we want and with whatever weights we want to weigh them. Typical selections include an odd number of neighbors, usually up to a maximum of five or seven neighbors, with either equal weights or with an exponentially decaying function giving respective weights as the neighboring distance D is decreasing. Kernel functions like the *Triangular* or *Epanechnicov* are used very frequently to produce the respective weights as well (Haerdle, 1992). The final forecast is constructed by

[2] *If Y_N=0 then we use the Naïve method so as to produce the one-step ahead forecasts, and we are waiting until $Y_{N+m} \neq 0$ (m>0) so as to apply the framework. The underlying principle is that we are interested to identify patterns and neighbors when there is 'activity' in the series, thus at the so-called 'issue points' where we have non-zero demand. The framework can easily be expanded to work for present sequences containing zero values as well.*

applying these weights in the actual data points right after the neighboring vectors. For example if we use an exponential function with basis α and use *k neighbors*, the one-step ahead *k-NN* forecast becomes:

$$F_{k\text{-}NN,N+1} = \left(\sum_{i=1}^{k} a^i Y_{N\text{-}Lagi+1}\right) / \sum_{i=1}^{k} a^i,\ 0<\alpha<1 \qquad (3)$$

where *Lag1* is the lag of the nearest neighbor, *Lag2* the lag of the second nearest neighbor, etc.

3.2 An illustrative example

The proposed framework with the use of only one NN - usually noted as *NN1 model* or *1-NN model*, is graphically illustrated in figure 1.

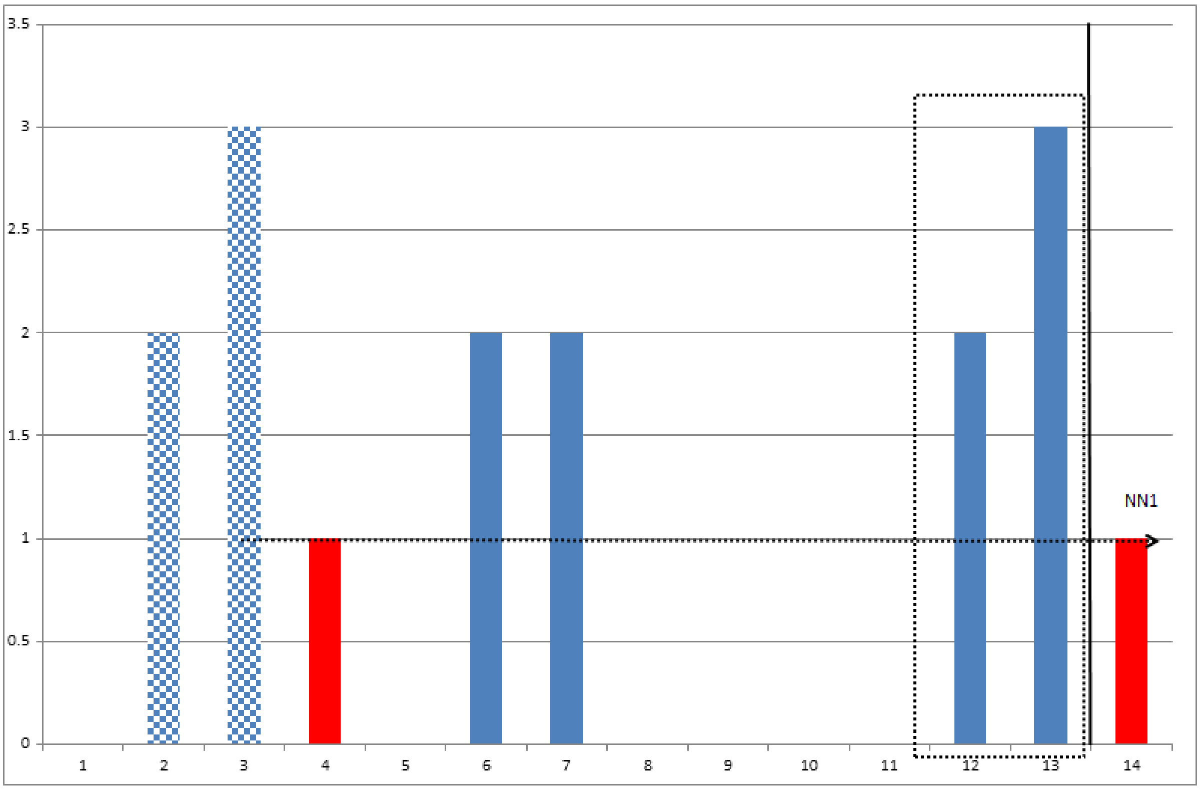

Figure 1. An example of how the proposed NN1 method would work on a sporadic SCM time simulated weekly series.

In this simulated example we do have a sporadic time series with 13 weekly observations (one quarter of a year), of which six are zero-values and seven non-zero values. The 14th value is kept as a holdout so we do use only the first thirteen observations so as to produce a one-step ahead forecast for the fourteenth observation that is equal to 1. The actual volumes are ranging from 0 to 3 units and the inter-demand intervals from 0 to 5 weeks.

Following the aforementioned conceptual framework, the present row-vector *Now* contains the latest non-zero values in observations 12 and 13 and is equal to [2,3] with length l=2 (as the 11th observation is equal to zero). We are now going to compare this row-vector with all possible row vectors of sequential observations of length l=2 from the past of the series as illustrated in Table 1:

Present row-vector: ***Now***$_{(12,13)}$ **= [2,3]**				
Observations	***Neighbor*** row-vector	Distance **D**	**Forecast** F_{14}	**Ranking**
1,2	[0,2]	$\sqrt{5}$	$Y_3=3$	3
2,3	**[2,3]**	**0**	$\mathbf{Y_4=1}$	**1**
3,4	[3,1]	$\sqrt{5}$	$Y_5=0$	3
4,5	[1,0]	$\sqrt{10}$	$Y_6=2$	8
5,6	[0,2]	$\sqrt{5}$	$Y_7=2$	3
6,7	**[2,2]**	**1**	**Y8=0**	**2**
7,8	[2,0]	3	$Y_9=0$	3
8,9	[0,0]	$\sqrt{13}$	$Y_{10}=0$	9
9,10	[0,0]	$\sqrt{13}$	$Y_{11}=0$	9
10,11	[0,0]	$\sqrt{13}$	$Y_{12}=2$	9
11,12	[0,2]	$\sqrt{5}$	$Y_{13}=3$	3

Table 1. The NN algorithm: finding the *Neighbors*, calculating the distances *D* and *Ranking*.

We do notice that in observations 2 and 3 there is the vector [2,3] ('checked' bars in figure 1, second row in table 1), that is a 'perfect match' to the present vector in observations 12 and 13; this is obviously the nearest neighbor. We do also notice that in observations 6 and 7 there is the vector [2,2] that with a distance D = sqrt($(2-2)^2 + (2-3)^2$) = 1 is the second nearest neighbor. In a similar fashion table1 is constructed and the neighbors are ranked, where many 'ties' do occur in this specific example. Alternatively we can use other non-Eucledian norms for measuring the distance like the sum of the absolute differences of the respective vector elements.

We also do notice that the proposed neighboring function gives the same score *D* for vectors [2,0] and [0,2] as it does not apply different weights depending on the order of appearance of the two elements of the vectors. We could also generalise the neighboring function for allowing the introduction of weights for the respective elements of the vectors:

$$D^* = ||Now_{(N-l+1,N)} - Neighbor_{(N-Lag-l+1,N-Lag)}||$$

$$= \sqrt{\left(\sum_{i=0}^{l-1} w_i (Y_{N-i} - Y_{N-Lag-i})^2\right) / \sum_{i=0}^{l-1} w_i} \tag{4}$$

This way we could give more weight for example to the respective differences to the very last observation, that according to the forecasting literature is the one containing most of the important information; and thus the historic consistent success of the Naïve forecasting method in forecasting evaluations.

In order to produce the forecast for the hidden 14th observation we use the actual observations right after the end of the vector of each of the respective neighbors as illustrated in table 1; thus:

$F_{NN1,14} = Y_4 = 1$, and (5)

$F_{NN2,14} = (Y_4+Y_8)/2 = ½$ (6)

We could also instead of using equal weights for the two nearest neighbors of the *NN2 model* (or *2-NN*), to use ad-hoc weights so as to weigh more the nearest neighbor, for example:

$$F_{NN2^*,\,14} = (3Y_4+Y_8)/4 = 3/4 \tag{7}$$

resulting in a 75% weight for the nearest neighbor and 25% for the second nearest neighbor. Alternatively we could use an exponential function or a kernel function for the respective purpose (see equation (3)).

If we wanted to include the third nearest neighbor, we would need to take another ad-hoc decision as to either include any (randomly) of the five equally third nearest neighbors, or (for the sake of consistency) use the average forecast of those five neighbors. As in any averaging exercise, the more elements included in the averaging, the more the smoothing that is applied and as a result the closer the final forecast would be to the mean of the data.

4. Forecasting simulated series with NN versus parametric methods

In order to see how the NN methods perform in a large dataset rather than a single time series as in the aforementioned illustrative example, and versus established benchmarks, as well as investigate the conditions under which we would expect the NN methods to perform well, we use a simulated dataset[3]. We start our simulation with four different *types of patterns* of lengths 2, 2, 3, and 4 periods respectively, and do control for three characteristics that often prevail sporadic time series:

- *frequency* and *plurality* of patterns: *low* frequency and *singularity*, where for every pattern in the time series there is one more repetition of that same pattern; *high* frequency and *singularity*, where for every pattern in the series there are four more repetitions of it; and *high* frequency and *plurality*, where for every pattern in the time series there is one more repetition of it and three more different patterns present (to the original and in-between them) - making pattern recognition very difficult!
- level of *intermittence*: *low* where 50% of the data points represent non-zero demand; and high where only 20% of the data points represent non-zero demand thus 80% of the data are zeros.
- levels of *noise*: *low* noise and *high* noise; for each level of noise there are 10 instances created for the series using different randomisation seeds

All these characteristics are superimposed under a factorial setup thus creating a pool of 4x3x2x2x10=480 series where each series is coded separately for identification and respective aggregate analysis. The length of the series varies from 8 to a 100 data points. We do use 50% of each series as a holdout over which we do a rolling evaluation of one-step ahead forecasts over the remaining periods. We do use two metrics of forecasting performance, the Mean Error (ME) that is a proxy of the *bias* of the produced forecasts from every method and the Mean Squared Error (MSE) that is a proxy of the accuracy and more specifically of the variance and the respective *uncertainty* of the provided forecasts.

[3] *The simulated dataset and the respective coding for the three control features (patterns, intermittence, noise) may be accessed at http://www.forlab.eu/forecasting-software*

We do produce forecasts with five NN approaches: a) NN1, b) NN3 with equal weights, c) NN3* with three nearest neighbors but with the nearest one getting a weight of 50% and the remaining two 30% and 20% respectively, d) NN4 with equal weights, c) NN4* with four nearest neighbors but with the nearest one getting a weight of 30% and the remaining three 40%, 30%. 20% and 10% respectively. NN1 is the most selective of the NN approaches as it uses only one neighbor so as to extrapolate.

	NN1	**NN3**	**NN3*** (0.5,0.3,0.2)	**NN4**	**NN4*** (0.4,0.3,0.2,0.1)
ME	-4.046	-1.984	-2.775	**-0.359**	-1.919
MSE	1260.569	1123.011	**1053.325**	1164.179	1059.522

Table 2a. All 480 simulated series: forecasting performance of ***NN*** methods in terms of ME and MSE

We then produce forecasts with four popular parametric approaches: SES, Croston, SBA, TSB for a wide range of their parameters with smoothing constant α ranging from 0.05 to 0.30 (step 0.05) and β for TSB set to 0.05[4]. We do yet again present the performance both for ME and for MSE in table2b.

	Methods	*Smoothing constant* 0.05	*0.10*	*0.15*	*0.2*	*0.25*	*0.30*
ME	SES	0.187	0.076	**-0.017**	-0.055	-0.047	-0.005
	SBA	5.458	4.397	3.566	2.877	2.289	1.775
	Croston	6.211	8.736	5.795	5.856	6.035	6.311
	TSB	0.501	-0.060	0.020	-0.161	-0.312	-0.438
MSE	SES	1289.911	1336.125	1379.609	1419.622	1456.529	1490.945
	SBA	1452.946	1455.140	1464.756	1476.931	1490.159	1503.851
	Croston	1471.743	1598.683	1515.813	1544.678	1575.842	1609.207
	TSB	**1287.995**	1306.861	1290.671	1293.825	1297.739	1302.198

Table 2b. All 480 simulated series: forecasting performance of ***parametric*** methods in terms of ME and MSE.

Comparing tables 2a and 2b we notice that overall and across the 480 simulated series NN3* is the most accurate method in terms of MSE while TSB is the best from the parametric methods. In fact, all five NN variants are more accurate than any parametric method. In terms of ME however parametric methods do better with the best NN one being NN4 that was better than SBA and Croston but significantly worse than SES and TSB. This is expected as these two latter parametric methods have been shown to be statistically unbiased.

What we can also add when we interpret the results of Table 2a and 2b is that when a single α value is used (as in practice), then for this α value we look at both the ME and MSE simultaneously. In that case, NN does better than even TSB because for the MSE = 1287.995, the ME is 0.501 so higher than that of NN if we compare with the two values for NN4 for ME and MSE (-0.359, 1164.179). If we opt for a low bias which is given by SES, then the MSE is much

[4] *Results under other beta values have also been generated for TSB but not reported here for the sake of economy of space, as they do not lead to different insights.*

higher than that of NN. So in all cases, it is clear using this argument that NN is the best for this dataset.

To a certain extent, we were anticipating that in a set of simulated data NN methods would illustrate superior forecasting performance; and this bring to the more interesting question: what are the conditions favouring (and not) the use of the NN framework?

4.1 Conditions for superior performance of NN approaches

Of all the potentially 12 combinations of:

(*frequency* and *plurality* of patterns) x (level of *intermittence*) x (level of *noise*)

when the plurality is set to singular - that is when there is only pattern in the series that is repeated more or less frequently, then the NN methods always perform better than the parametric ones; this is the case for all 8 such combinations tested in our simulation.

Parametric methods can outperform the NN ones only when faulty patterns do exist in the series, so in the simulated series where for every true repetition of a pattern there are three other different patterns present in the series. The result for those cases are as follows:

	NN1	NN3	NN3* (0.5,0.3,0.2)	NN4	NN4* (0.4,0.3,0.2,0.1)
ME	-6.670	-2.152	-3.773	**0.727**	-2.170
MSE	1940.323	1337.214	1359.634	1272.409	**1269.507**

Table 3a. 40 simulated series with plurality of patterns, low intermittence and low noise: forecasting performance of ***NN*** methods in terms of ME and MSE

	Methods	*Smoothing constant* 0.05	*0.10*	*0.15*	*0.2*	*0.25*	*0.30*
ME	SES	2.751	2.227	1.8403	1.584	1.431	1.353
	SBA	3.489	2.287	1.328	0.545	**-0.106**	-0.658
	Croston	4.344	3.978	3.855	3.921	4.140	4.491
		3.383	2.732	2.222	1.831	1.542	1.338
MSE	SES	1345.282	1392.534	1443.422	1495.313	1547.728	1600.654
	SBA	**1315.522**	1343.224	1372.963	1404.202	1436.613	1469.993
	Croston	1324.351	1360.207	1398.894	1440.993	1487.146	1538.094
	TSB	1340.940	1345.616	1354.375	1366.230	1380.541	1396.910

Table 3b. 40 simulated series with plurality of patterns, low intermittence and low noise: forecasting performance of ***parametric*** methods in terms of ME and MSE

Here yet again, the lowest MSE of SBA is obtained for α = 0.05. For the same α, ME of SBA is 3.489 that is higher than that of NN obtained for the lowest MSE. So, for α =0.3, NN is better than SBA overall.

For the 80 series where intermittence is low - and thus there are many non-zero demand values in the data, in the case of high noise, parametric methods are slightly better (Tables 3a/3b) while when noise is low NN methods are slightly better (Tables 4a/4b).

	NN1	NN3	NN3* (0.5,0.3,0.2)	NN4	NN4* (0.4,0.3,0.2,0.1)
ME	-6.871	-4.616	-5.255	-2.122	-4.032
MSE	1071.484	927.110	923.809	910.429	**893.287**

Table 4a. 40 simulated series with plurality of patterns, low intermittence and high noise: forecasting performance of ***NN*** methods in terms of ME and MSE

	Methods	*Smoothing constant* 0.05	0.10	0.15	0.2	0.25	0.30
ME	SES	-1.362	-0.834	-0.5179	-0.283	-0.101	0.040
	SBA	-5.306	-4.622	-4.155	-3.805	-3.523	-3.280
	Croston	-5.081	-4.123	-3.348	-2.661	-2.012	-1.370
	TSB	-1.078	-1.178	-1.263	-1.334	-1.396	-1.450
MSE	SES	889.494	920.872	949.225	974.439	997.053	1017.728
	SBA	912.595	906.205	906.378	909.986	915.671	922.762
	Croston	912.116	905.135	905.271	909.710	917.335	927.734
	TSB	**884.638**	885.375	886.316	887.411	888.630	889.959

Table 4b. 40 simulated series with plurality of patterns, low intermittence and high noise: forecasting performance of ***parametric*** methods in terms of ME and MSE

For the 80 series where intermittence is high - and thus there are very few non-zero demand values in the data, in both the case of low and high noise, parametric methods are clearly better than the NN methods (Tables 5a/5b and 6a/6b respectively).

	NN1	NN3	NN3* (0.5,0.3,0.2)	NN4	NN4* (0.4,0.3,0.2,0.1)
ME	-0.032	-2.934	-2.176	-1.065	-1.719
MSE	3474.202	**1957.513**	2068.298	1985.325	1985.345

Table 5a. 40 simulated series with plurality of patterns, high intermittence and low noise: forecasting performance of ***NN*** methods in terms of ME and MSE

	Methods	*Smoothing constant* 0.05	0.10	0.15	0.2	0.25	0.30
ME	SES	2.876	2.490	2.1319	1.853	1.660	1.542
	SBA	4.278	2.994	1.933	1.042	0.276	-0.397
	Croston	5.291	5.005	4.945	5.070	5.346	5.747
	TSB	3.523	2.830	2.274	1.833	1.487	1.220
MSE	SES	1744.364	1802.792	1864.087	1924.609	1983.687	2041.558
	SBA	**1724.997**	1752.986	1786.699	1824.098	1863.488	1903.444
	Croston	1738.474	1778.430	1824.714	1876.899	1934.522	1997.092
	TSB	1746.324	1744.605	1748.404	1756.031	1766.261	1778.197

Table 5b. 40 simulated series with plurality of patterns, high intermittence and low noise: forecasting performance of ***parametric*** methods in terms of ME and MSE

Here one also can claim that for α = 0.05, it is true that SBA gives lower MSE but it results in higher bias than NN. Only a full efficiency analysis could prove what is clearly better but nevertheless there are arguments for NN in almost all provided results.

	NN1	NN3	NN3* (0.5,0.3,0.2)	NN4	NN4* (0.4,0.3,0.2,0.1)
ME	-3.252	**0.069**	-0.453	2.285	0.766
MSE	1597.075	**1081.939**	1126.280	1105.073	1091.883

Table 6a. 40 simulated series with plurality of patterns, high intermittence and high noise: forecasting performance of ***NN*** methods in terms of ME and MSE

	Methods	*Smoothing constant*					
		0.05	*0.10*	*0.15*	*0.2*	*0.25*	*0.30*
ME	SES	-1.340	-0.907	-0.6262	-0.391	-0.189	**-0.019**
	SBA	-5.131	-4.664	-4.388	-4.206	-4.077	-3.975
	Croston	-4.865	-4.096	-3.490	-2.955	-2.449	-1.947
	TSB	-0.886	-1.053	-1.197	-1.324	-1.435	-1.534
MSE	SES	1033.686	1069.901	1102.3333	1130.971	1156.516	1179.909
	SBA	1058.587	1049.415	1047.643	1049.649	1053.855	1059.478
	Croston	1058.592	1048.845	1046.697	1048.877	1054.051	1061.664
	TSB	**1032.883**	1032.501	1032.464	1032.690	1033.121	1033.721

Table 6b. 40 simulated series with plurality of patterns, high intermittence and high noise: forecasting performance of ***parametric*** methods in terms of ME and MSE

5. Empirical evaluation via forecasting a real large-scale dataset: Automotive spare parts

Given the obvious advantage of NN methods when a pattern that is about to repeat itself exists in the past of a time series, the real question becomes: what is the potential accuracy loss if we leave the NN methods to produce forecasts unsupervised in a large dataset?

To answer this question, we employ an empirical database that consists of the individual monthly demand histories of 3000 SKUs covering 24 consecutive months from the automotive industry. Descriptive statistics (across all SKUs) are given in Table 9.

3000 SKUs	Demand Intervals		Demand Sizes		Demand per period	
	Mean	St. Deviation	Mean	St. Deviation	Mean	St. Deviation
Min	1.043	0.209	1.000	0.000	0.542	0.504
25%ile	1.095	0.301	2.050	1.137	1.458	1.319
Median	1.263	0.523	2.886	1.761	2.333	1.922
75%ile	1.412	0.733	5.000	3.357	4.167	3.502
Max	2.000	1.595	193.750	101.415	129.167	122.746

Table 7. Descriptive statistics of the empirical data

Note from Table 7 that this dataset consists of low-volume demand items with low degree of intermittence. In fact, the average inter-demand interval ranges from 1.04 to 2 months and the average demand size (positive demands excluding the zeros) is between 1 and 194 units. This results in an average demand per unit time period ranging from 0.5 to 129 units.

From the 24 data points in each series we use the last 11 as a holdout over which we do a rolling evaluation of one-step ahead forecasts (over these 11 periods). We do use the same metrics as in the simulation in the previous section and evaluate exactly the same NN approaches and respective parametric benchmarks.

	NN1	NN3	NN3* (0.5,0.3,0.2)	NN4	NN4* (0.4,0.3,0.2,0.1)
ME	0.059	0.050	0.061	-0.024	**0.011**
MSE	129.050	90.603	93.489	**84.627**	85.526

Table 8a. 3000 Automotive series: forecasting performance of ***NN*** methods in terms of ME and MSE

	Methods	*Smoothing constant* 0.05	0.10	0.15	0.2	0.25	0.30
ME	SES	-0.066	-0.068	-0.066	-0.062	-0.057	-0.052
	SBA	**-0.005**	-0.107	-0.206	-0.304	-0.403	-0.503
	Croston	0.108	0.120	0.136	0.153	0.171	0.189
	TSB	-0.056	-0.064	-0.069	-0.073	-0.075	-0.076
MSE	SES	**74.970**	75.269	76.339	77.859	79.673	81.702
	SBA	75.010	74.923	75.481	76.396	77.512	78.747
	Croston	75.173	75.199	75.917	77.077	78.542	80.239
	TSB	75.069	75.033	75.554	76.422	77.523	78.795

Table 8b. 3000 Automotive series: forecasting performance of ***parametric*** methods in terms of ME and MSE.

We note in tables 8a and 8b that the best parametric method in terms of accuracy is scoring 74.970 (MSE for SES) while the best NN method is scoring worse at 84.627 (MSE for NN4). This compiles in 12.88% worse performance over the 3000 series in terms of the specific metric employed. We will now calculate if this difference is amplified or narrows when the NN methods are applied to that part of the dataset with series that illustrate more repetitive patterns. To that end we now try to identify series within this dataset that contain more patterns than others. We do use the standard autocorrelation function (ACF) (equation 8) for that purpose, but not applied to each original series.

$$r_k = \frac{\sum_{t=k+1}^{n}(y_t-\bar{y})(y_{t-k}-\bar{y})}{\sum_{t=1}^{n}(y_t-\bar{y})^2} \qquad (8)$$

If this metric is applied to each original series, a series with lots of occurrences of zero-observations will present high ACF. We instead are interested in repeated non-zero patterns of length of two or more. As such we first strip from every series all non-zero sequences of length one (single occurrences) and then all remaining zero values. We then apply the ACF function in this modified series. We do allow for ACF to be calculated from lag=2 up to the max lag that maybe calculated in each series given the respective length. To account for the total autocorrelation in every series we introduce a cumulative metric in terms of an ACF scoring function defined as the $\sum_{k=2}^{p-1} r_k$ (sum from lag 2 to lag p-1 where p is the length of the series). We then rank the 300 series of the automotive dataset according to their ACF scoring and select the upper quartile of this non-parametric distribution - the top 25% of the series in terms of the exhibiting autocorrelations in non-zero patterns of length of two or more. These 750 series are series that have more patterns and as such the NN variants should perform better than when applied unsupervised in all 3000 series.

	NN1	**NN3**	**NN3*** (0.5,0.3,0.2)	**NN4**	**NN4*** (0.4,0.3,0.2,0.1)
ME	-0.047	0.029	**0.010**	0.055	0.039
MSE	5.372	3.539	3.674	**3.292**	3.337

Table9a. ACF top-quartile (750 Series) of the 3000 Automotive series: forecasting performance of ***NN*** methods in terms of ME and MSE.

	Methods	*Smoothing constant*					
		0.05	*0.10*	*0.15*	*0.2*	*0.25*	*0.30*
ME	SES	-0.024	-0.022	-0.019	-0.016	-0.014	**-0.012**
	SBA	<u>0.024</u>	**-0.012**	-0.044	-0.074	-0.103	-0.132
	Croston	0.059	0.058	0.061	0.067	0.074	0.081
	TSB	-0.024	-0.024	-0.023	-0.023	-0.023	-0.022
MSE	SES	<u>3.353</u>	3.283	3.279	3.314	3.374	3.450
	SBA	3.379	3.293	3.255	3.250	3.268	3.302
	Croston	3.385	3.295	3.251	**3.243**	3.261	3.298
	TSB	3.381	3.321	3.289	3.276	3.278	3.290

Table 9b. ACF top-quartile (750 Series) of the 3000 Automotive series: forecasting performance of ***parametric*** methods in terms of ME and MSE.

We note in tables 9a and 9b that the best parametric method in terms of accuracy is scoring 3.243 (MSE for Croston) while the best NN method is scoring worse at 3.292 (MSE for NN4). This compiles in just 1.51% worse performance over the 750 series in terms of the specific metric employed and that difference is actually reverted with NN4 being 1.81% better that the winner of the entire 3000 series SES when evaluated over just the 750 series (MSE at 3.352). What we can also add when we interpret the results of Table 9a and 9b (and that is clear) is that when an α value is used (as in practice) i.e. here 0.2, then for this α value we look at both the ME and MSE. In that case, NN does better than even TSB because (for the MSE=3.243) NN has a lower bias which is an argument in favour of NN.

Thus our argument is by and large supported: if NN are left unsupervised over a dataset cannot really improve accuracy, but if applied only on a specific part of a dataset exhibit higher levels of cumulative ACFS and therefore potential existence of repetitive patterns, then NN are en par with top parametric methods. Converting this finding into tangible guidance for academics, practitioners and software designer we suggest that the proposed framework should be used quite selectively:

- *Use either TSB or SBA as the default forecasting method for all the products and services in your dataset, and do flag up the ones where there is definite knowledge that a pattern do exist in the past and would at some point reappear: for these do use the NN approach*

6. Implications for SCM theory and practice

In many respects, this work can be seen as a contribution towards *bridging the gap between SCM theory and practice.* It provides the conceptual and theoretical framework for forecasting sporadic series with at least one historic pattern that is about to repeat itself: something that in practice is actually happening almost every day, and that to date could only be dealt through judgmental interventions in statistical forecasts. We do also believe that there are some direct implications for both SCM theory and practice coming out of this study.

Implications for SCM theory

- *A new framework*: the contribution of this research is not in the development of a new theory, but in the proposition of a new conceptual framework that brings existing theory from Computer Science and Statistics in the SCM setting. This to the best of our knowledge has not been done in the past and as such it is a contribution by itself.
- *Interdisciplinarity*: the proposed interdisciplinary framework brings together concepts from three disciplines: a) Computer Science and AI with the concept of automated unsupervised learning from past events/patterns, b) Statistics with the concepts of non-parametric regression smoothing techniques, and c) SCM with the sporadic nature of demand time series.
- *Simplicity*: the proposed framework is very simple to use and very easy to understand and thus drives science and theory away from 'black box' solutions and towards identifying causality in inferential tasks. Also NN can work with very small training sets something that no other AI methods can do. Furthermore it follow the KISS[5] paradigm as

[5] *Keep It Sophisticated Simple*

the proposal is as simple as possible so as to successfully identify patterns in the past of a time series.

Implications for SCM practice

- *Solving an everyday practical problem:* the proposed framework addresses an everyday practical problem that SCM and OM practitioners have to deal with. Repetitive patterns do exist in SCM data and unfortunately the existing forecasting methods cannot capture these, and in order to cope with this problem judgmental adjustments on statistical forecasts have to be part of everyday practice.
- *Automation:* a problem that needed human intervention and expert judgment can now to a large extent be sorted automatically
- *Interpretation*: practitioners do understand how this framework works as they do apply that in practice anyway. By and large NN are based on the *representativeness* heuristic that practitioners do use all the time (Tversky & Kahneman, 1974).
- *Simplicity:* the approach can work with very few data something very convenient for practitioners, as very rarely they do have enough historic data for the demand of each of their products and services. Furthermore the method can be adopted so as to look for neighbors in other time series as well and not strictly just in the past of one time series, something very useful for complementary and substituting products and services
- *Flexibility*: the framework allows the users to adjust the method to their needs by selecting the number of neighbors and by changing the neighboring function to their needs
- *Implementation*: the proposed framework should superimpose a default method for forecasting sporadic demand like SBA or TSB and triggered only if evidence of patterns exist in the series.

7. Conclusions & Further Research

In this research we have proposed a conceptual framework for applying NN for supply chain data. Furthermore we investigated the conditions under which the method can perform adequately. Our findings suggest that practitioners can benefit from employing NN approaches; however these methods should not be left unsupervised in large datasets rather than run in a selective mode where a standard parametric method is the default (SBA or TSB) and the NN is used only in the presence of reliable information attesting for the existence of patterns.

From the theoretical underpinning of the proposed framework as well as the provided illustrative examples and the extensive simulations, we believe it became evident that the NN methods can help towards improving the forecastibility of sporadic time series when applied selectively. If the proposed NN method runs unsupervised in large datasets, our empirical evidence suggests that we could be facing up to 12.88 % deterioration in terms of forecasting accuracy performance that however is disappearing if applied to the part of the dataset containing series with repetitive patterns. Thus, we reiterate the predicament that we are facing here: the NN should only allowed to run when we do know that a pattern existed in the past and it is about to recur.

The simulations also highlighted that in the presence of high plurality of patterns - where most are not repeated, NN methods' performance deteriorates significantly especially if simultaneously high levels of intermittence are present; however if the latter fall into average levels then NN methods and parametric ones perform en par.

There are clear implications for SCM and OM theory and practice from this research, most of which result from the simplicity and intuitively appealing nature of the proposed interdisciplinary framework.

As far as the future research on this topic is concerned, we do suggest that:

- The framework should be expanded to work also with present sequences containing or ending in zero values
- Fully automate the use of the framework in large datasets without supervisions through identifying which series in a dataset do have some patterns that may be starting reappearing.

Acknowledgements

We would like to thank Professor Aris Syntetos and Dr Fotios Petropoulos for their feedback and research assistance in earlier versions of this research.

References

Altay, N., Litteral, L. A. & Rudisill, F. 2012. 'Effects of correlation on intermittent demand forecasting and stock control.' *International Journal of Production Economics*, 135:1, 275-283

Altay, N., Rudisill, F. & Litteral, L. A. 2008. 'Adapting Wright's modification of Holt's method to forecasting intermittent demand.' *International Journal of Production Economics*, 111: 2, 389-408

Babai, M.Z., Ali, M., & Nikolopoulos, K. 2012. 'Impact of Temporal Aggregation on Stock Control Performance of Intermittent Demand Estimators: Empirical Analysis.' *OMEGA: The International Journal of Management Science*, 40, 713–721.

Bozos, K. & Nikolopoulos, K. 2011. 'Forecasting the value effect of seasoned equity offering announcements.' *European Journal of Operational Research,* 214:2, 418-27.

Brabec, B. & Meister, R. 2001. 'A nearest-neighbour model for regional avalanche forecasting ' *Annals of Glaciology,* 32:1, 130-34.

Engle, R. F. 1982. 'Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation,' *Econometrica* , 50:4, 987-1007.

Green, K. C. 2002. 'Forecasting decisions in conflict situations: a comparison of game theory, role-playing, and unaided judgement.' *International Journal of Forecasting,* 18:3, 321-44.

Härdle, W. 1992. *Applied Nonparametric Regression.* Cambridge University Press.

Haykin, S. 1998. *Neural networks: A Comprehensive Foundation*. International Edition. Pearson US Imports & PHIPEs.

Gutierrez, R. S., Solis, A. O. & Mukhopadhyay, S. 2008. 'Lumpy demand forecasting using neural networks,' I*nternational Journal of Production Economics*, 111: 2, 409–420.

Jaditz, T. & Sayers, C. L. 1998. 'Out-of-Sample Forecast Performance as a Test for Nonlinearity in Time Series.' *Journal of Business & Economic Statistics,* 16:1, 110-17.

Kanas, A. 2003. 'Non-linear forecasts of stock returns.' *Journal of Forecasting,,* 22:4, 299-315.

Kourentzes, N., Petropoulos, F., and Trapero, J. R. 2014. 'Improving forecasting by estimating time series structural components across multiple frequencies'. *International Journal of Forecasting*, 30, 291-302.

Meade, N. 2002. 'A comparison of the accuracy of short term foreign exchange forecasting methods. .' *International Journal of Forecasting,* 18:1, 67-83.

Mizrach, B. 1992. 'Multivariate Nearest-Neighbour Forecasts of EMS Exchange Rates.' *Journal of Applied Econometrics,* 7, S151-S63.

Mulhern, F. J. & Caprara, R. J. 1994. 'A nearest neighbor model for forecasting market response.' *International Journal of Forecasting,* 10:2, 191-207.

Nikolopoulos, K., Goodwin, P., Patelis, A. & Assimakopoulos, V. 2007. 'Forecasting with cue information: A comparison of multiple regression with alternative forecasting approaches.' *European Journal of Operational Research,* 180:1, 354-68.

Nikolopoulos, K., Syntetos, A., Boylan, J., Petropoulos, F., & Assimakopoulos, V. 2011. 'ADIDA: An aggregate/disaggregate approach for intermittent demand forecasting.' *Journal of the Operational Research Society*, 62, 544-554.

Petropoulos, F., Nikolopoulos, K., Spithourakis, G. P. & Assimakopoulos, V. 2013. 'Empirical Heuristics for Improving Supply Chain Forecasting for intermittent demand SKUs.' *Industrial Management and Data Systems*, 113:5, 683-696.

Petropoulos, F., and Kourentzes, N. 2014. 'Forecast combinations for intermittent demand'. *Journal of the Operational Research Society*. DOI 10.1057/jors.2014.62.

Petropoulos, F., Makridakis, S., Assimakopoulos, V. and Nikolopoulos, K. 2014. 'Horses for Courses' in demand forecasting', *European Journal of Operational Research* 237 (1): 152-163.

Rostami-Tabar, B., Babai, M. Z., Syntetos, A. A., & Ducq, Y. 2013. 'Demand forecasting by temporal aggregation.' *Naval Research Logistics*, 60, 479-498.

Rostami-Tabar, B., Babai, M.Z., Syntetos, A., and Ducq, Y. 2014. 'A note on the forecast performance of temporal aggregation'. *Naval Research Logistics*, 61, 489-500.

Sauer, T. 1994. 'Time series prediction by using delay coordinate embedding.' In A. S. Weigend & N. A. Gershenfeld (Eds.) *Time Series Prediction. Forecasting the Future and Understanding the Past*. Reading, MA: Addison-Wesley

Smith, B. L., Williams, B. M. & Keith Oswald, R. 2002. 'Comparison of parametric and nonparametric models for traffic flow forecasting.' *Transportation Research Part C: Emerging Technologies,* 10:4, 303-21.

Spithourakis,, G. P., Petropoulos, F., Nikolopoulos, K., & Assimakopoulos, V. 2014. 'A systemic view of the ADIDA framework.' *IMA Journal of Management Mathematics*, first published online December 2, 2012, doi:10.1093/imaman/dps031

Sun, H. Y., Liu, H. X., Xiao, H., He, R. R. & Ran, B. 2003. 'Use of local linear regression model for short-term traffic forecasting.' *Initiatives In Information Technology and Geospatial Science for Transportation. Transp. Res. Record,* 1836, 143-50.

Syntetos, A. A., & Boylan, J. E. 2001. 'On the bias of intermittent demand estimates.' International Journal of Production Economics, 71, 457-466.

Syntetos, A. A., & Boylan, J. E. 2005. 'The accuracy of intermittent demand estimates.' *International Journal of Forecasting*, 21, 303-314.

Syntetos, A. A., Boylan, J. E., & Croston, J. D. 2005. 'On the categorization of demand patterns.' *Journal of the Operational Research Society*, 56, 495-503.

Syntetos A. A., Babai M. Z. & Gardner, E. S. Jr. 2015. 'Forecasting Intermittent Inventory Demands: Parametric Methods vs. Bootstrapping.' *Journal of Business Research*, forthcoming

Teunter, R. H., Syntetos A.A., & Babai M. Z. 2011. 'Intermittent demand: Linking forecasting to inventory obsolescence.' *European Journal of Operational Research*, 214, 606-615.

Tversky, A. & Kahneman, D. 1974. 'Judgments Under Uncertainty: Heuristics and Biases.' *Science,* 185:4157, 1124–1131.

Weigend, A. S. & Gershenfeld, N. A. 1994. *Time Series Prediction: Forecasting the Future and Understanding the Past : Proceedings of the NATO Advanced Research Workshop on Comparative Time Series Analysis, Held in Santa Fe, New Mexico, May 14-17, 1992.* Addison-Wesley.

Willemain, T. R., Smart, C. N. & Schwarz, H. F. 2004. 'A new approach to forecasting intermittent demand for service parts inventories.' *International Journal of Forecasting*, 20:3, 375–387.

Yakowitz, S. 1987. 'Nearest-neighbour methods for time series analysis.' *Journal of Time Series Analysis,* 8:2, 235-47.

Yankov, D., DeCoste, D. & Keogh, E. 2006. 'Ensembles of Nearest Neighbor Forecasts.' In J. Fürnkranz, T. Scheffer & M. Spiliopoulou (Eds.) *Machine Learning: ECML 2006:* 545-56. Springer Berlin Heidelberg.