Real-time temperature prediction in a cold supply chain based on Newton's law of cooling

Iurii Konovalenko[a,*], André Ludwig[b], Henrik Leopold[c]

[a] Kühne Logistics University, Postal Address: Großer Grasbrook 17, Hamburg 20457, Hamburg, Germany, email: iurii.konovalenko@the-klu.org

[b] Kühne Logistics University, Postal Address: Großer Grasbrook 17, Hamburg 20457, Hamburg, Germany, email: andre.ludwig@the-klu.org

[c] Kühne Logistics University, Postal Address: Großer Grasbrook 17, Hamburg 20457, Hamburg, Germany, email: henrik.leopold@the-klu.org

* Corresponding Author: Tel: +49 40 32 87 07-304, Fax: +49 40 32 87 07-109

Real-time temperature prediction in a cold supply chain based on Newton's law of cooling

Abstract

Many goods, including pharmaceuticals, require close temperature monitoring. This is important not only for complying with regulations but also for guaranteeing safety of use. A particular challenge in controlling a product's temperature arises during transportation. In cold supply chains (SCs), temperature is maintained by refrigerated containers. However, many situations, e.g. cooling system failure, lead to ambient temperature changes, and this needs to be detected as early as possible to prevent product damage. Existing approaches to temperature prediction are confined to long-term forecasts with relatively stable ambient temperatures and/or rely on multiple sensors in the known fixed positions. Since interventions in a SC are required immediately, there is a need for methods that provide real-time predictions regarding regular ambient temperature instability, i.e. when the ambient temperature changes unexpectedly in the short term. We propose a novel method that extends the applicability of Newton's law of cooling (NLC) to changeable ambient temperatures based on a set of temperature stability conditions and a sensor measurement error. In the method, an optimal number of measurements that characterize stable ambient temperatures and improve prediction reliability are selected. We compare the adapted NLC with artificial neural networks and autoregressive moving average models with respect to deviation prediction, prediction error, and execution time. Our evaluation based on real-world data shows that the adapted NLC outperforms existing baseline methods. In contrast to existing solutions, our method does not require any knowledge about the positioning of products within the container, further increasing its practical value.

Keywords

Cold supply chain, temperature prediction, Newton's law, artificial neural network, ARMA, event data

1. Introduction

Temperature excursions in a cold supply chain (SC) can not only affect the quality of sensitive goods during transportation and storage but can also pose a danger to the health of end consumers. While temperature deviations affect the length of food products' shelf life, pharmaceuticals are particularly vulnerable to immediate losses in quality [1], with alarming consequences for health and life [2]. Huge economic losses are

the usual consequence of such deviations for manufacturers and other parties involved in the transportation and storage process [3]. The criticality of temperature maintenance along a cold SC is anchored in numerous legal requirements and quality assurance instructions that form the basis of a pharmaceutical business [4-7]. The challenge of continuous temperature control is exacerbated by the modern complexity of SCs, which often supply multiple geographic locations and involve various parties [8].

Current temperature monitoring solutions are enabled by the wide availability of sensor technologies for registering, storing, and transmitting measurements [9]. Tracing and tracking applications ensure either the availability of measurements for recovering the upstream path of products in an SC (tracing) or their real-time monitoring downstream of an SC (tracking) [1,10]. Rule-based and event-driven logic enables partial or complete automation of measurement evaluations in such applications [11]. Stability studies form a basis for monitoring by providing the permissible temperature ranges based on laboratory experiments [12].

Although the abovementioned types of monitoring applications made their way from research into practice [13], they can only notify about deviations after their occurrence and cannot predict them. In practice, however, there are a number of situations in which ambient temperature, and hence also the cargo's temperature, changes abruptly, thus requiring its proactive identification. These situations include but are not limited to a sensor or cooling unit malfunction, container placement on tarmac prior to loading, or physical handling such as inspections. When we look into extant research on temperature prediction methods in a cold SC, we notice that it has so far focused on the approximation of cargo's temperature with *stable, or relatively stable, ambient temperatures* over longer time periods and has *failed to address the predictions immediately adjusting to changeable, or unstable, ambient temperatures* [14-17].

However, the knowledge of cargo's future temperature trajectories under the condition of unstable ambient temperatures is required by numerous decision makers. In the short term, transport and operations managers (including temperature alarm personnel in cold SCs), dispatchers, and warehouse managers need such predictions for an immediate reaction to potential deviations. In the medium and long term, cold SC safety professionals and SC managers benefit from the ability to assess the reliability and compliance of freight forwarders and carriers (based on successfully avoided deviations due to proactive predictions) and estimate the improved reaction time of alarm personnel and its consequences for product safety. From a decision

support perspective, the following solutions can be realized thanks to real-time temperature predictions with the possibility of their integration into intelligent decision support systems:

1. Informing users of possible deviations in advance (for proactive countermeasures).
2. Forecasting temperature trajectories or interpolating measurements in the case of missing event data (e.g., no network coverage for event data transmission, delayed or out-of-sequence event data).
3. Identifying problematic segments in an SC in which deviations could have occurred if environmental conditions in the container remained unchanged (and developing possible countermeasures).
4. Ranking of identified (and forecast) deviations in terms of predicted magnitude, duration, etc.
5. Improving the service offered to customers not only in the form of real-time monitoring but also as a predictive solution, allowing for a timely reaction to or avoidance of a possible deviation.

In light of the abovementioned focus of extant research on long-term predictions with relatively stable ambient temperatures, we pose the following research question for this paper: *"How can cargo's temperature in a container be accurately predicted under the condition of unstable ambient temperatures for a proactive reaction to expected deviations in the immediate future?"*.

To answer this question, we propose the adaptation of Newton's law of cooling (NLC) for real-time temperature predictions. Normally, NLC assumes *ambient temperature stability* to be used for cargo temperature predictions in a container. In practice, however, it is hard to know if a new cargo sensor measurement signifies a stable ambient temperature or its (slight) change, especially in the presence of a measurement error. This makes the application of NLC in such cases problematic because it is not clear which (and how many) available measurements should be used for making predictions. Therefore, to realize the adaptation of NLC to unstable ambient temperatures, we propose a method for real-time selection of an optimal number of measurements that are characterized by a stable ambient temperature. Our method is based on a set of ambient temperature stability conditions and accounts for sensor measurement errors.

Additionally, in light of the abovementioned reliance of extant research on *simulated or case study data* [18-20], we perform the evaluation of adaptive NLC using longitudinal real-world data that ideally reflect the multitude of possible deviations in a cold SC. Finally, in contrast to the current research, we do not rely on the

knowledge about other factors that influence spatial temperature distribution and interpolation in a container, i.e., the number of sensors, their location, and the position of pallets in a container.

Our paper is organized as follows. First, we review the extant literature in the field of temperature control and prediction in a cold SC and identify the poorly addressed research areas (Section 2). In Section 3, we describe how we adapted NLC to unstable ambient temperatures with the proposed method. In Section 4, we first describe the business context of an air cargo cold SC, the type of SC in which the data used in our evaluations were collected. The baseline methods, a procedure for the generation of training and test data, and the experimental setups are then described. In Section 5, we discuss the evaluation results for deviation prediction, prediction error, and execution time tests. Section 6 concludes the paper by summarizing its findings and formulating the outlook for further research endeavors.

2. Related work

The extant literature on temperature control and prediction in cold SC boasts many valuable developments. Safety and quality requirements in the field of food sciences (FS) motivated much research on spatial temperature distribution (STD) and shelf-life modeling (SLM). Investigation of relevant literature allows one to single out five main categories into which prior work can be divided, namely:

a) *Demonstration and/or testing of new technologies or developed systems in a cold SC*. Most scholars whose work falls into this category explore the application of RFID-enabled temperature loggers for temperature monitoring and quality control in a cold food SC [18,21-24]. The body of research is complemented by tests and implementation of a wireless sensor network (WSN) for monitoring refrigerated perishable products [19,25-27].

b) *Development of general approaches and frameworks for enabling or practicing temperature control in a cold SC*. For example, general solutions based on complex event processing (CEP) are proposed by [28,29], who develop a framework for real-time monitoring in a cold SC and the IoT infrastructure for smart transportation applications. Similarly, [30-32] propose the design of monitoring systems to ensure food safety and reduce waste as a result of quality deterioration.

c) *Temperature mapping/profiling and ensuing quality control*. This category of the literature is more concerned with real-life applications and focuses on determining the influence of temperature on the shelf life of perishable products [18,21,33] or on estimation of temperature distribution in a container [20,34].

d) *Temperature (failure) prediction*. Research in this group goes a step further and applies physical models and artificial neural networks (ANNs) to long-term temperature trajectory approximations. For example, [14] relies on a heat transfer model in assessing the temperature stability of insulated boxes. Different types of ANN and their architectures are deployed by [15-17,35] for estimation of temperature changes during the entire transportation process.

e) *Summary or overview of approaches to temperature and quality control in a cold SC*. This category is represented by work on methods for shelf-life prediction and pallet temperature estimation as well as by overviews of field studies on time-temperature conditions at each critical stage of the cold SC [35-37].

A detailed account of extant work on temperature control and prediction, including the corresponding methodological approaches and domains, is presented in Table 1 below.

Table 1. Temperature control and prediction in a cold SC

Ref.	Contribution	Field	Methods	Data
	a) Technology demonstration			
[19]	Complex sensor network in perishable goods logistics	FS	ANN, SLM	E
[22]	RFID smart tag for real-time traceability and monitoring	FS	–	E
[25]	Test of wireless sensor motes during a real shipment in a refrigerated truck	FS	ANOVA, psychrometric charts	E
[23]	Applications survey of RFID-based temperature sensors	FS	–	E
[38]	Optimal solution of minimum vehicle traveling distance	FS	Simulated annealing	S
[26]	Application of a dynamic WSN in a cold SC	G	–	–
[24]	RFID-based shelf-life estimation in a cold SC	FS	SLM, CM	E
[27]	WSN for monitoring food products in a warehouse	FS	SLM	E
	b) General frameworks			
[28]	Framework for real-time monitoring of goods	G	CEP	S

[29]	IoT infrastructure for transportation applications	G	CEP	S
[30]	Development of an SC quality sustainability decision support system	FS	Rule learning, Dempster's rule	E
[31]	System for time-temperature monitoring in a meat SC	FS	–	E
[32]*	Framework for reducing supply chain food waste	FS	SLM	–
	c) Temperature mapping/profiling			
[18]*	Demonstration of the use of RFID temperature loggers in temperature gradient detection	FS	STD, SLM, ANOVA	E
[20]	Study of international container transport in terms of temperature distribution	FS	STD	E
[21]*	RFID-enabled product quality assessment during transport	FS	STD, SLM	E
[34]	Temperature distribution for improved food safety	FS	ANN, Kriging, HTM	E
[33]	Evaluation of the temperature performance of food SCs	FS	SLM	E
	d) Temperature (failure) prediction			
[14]	Estimation of temperature control failure	G	HTM	E
[15]	Back propagation neural network for predicting temperature shifts and trends	G	ANN, simulation, ANOVA	S, E
[16]	Approach to cold chain management based on the real-time monitoring of perishable cargo	FS	Autoregressive model, STD	E
[17]	Neural net framework based on a heat transfer model	FS	ANN, HTM	S
	e) Approaches to temperature control			
[35]**	Summary of shelf-life prediction factors	FS	ANN, Kriging, CM	E
[36]	Review of field studies on time-temperature conditions	FS	–	–
[37]*	Methods for RFID-based shelf-life prediction	FS	SLM, regression	S, E

* also belongs to category *a*; ** also belongs to category *d*; G – General; E – Experimental (e.g., case studies, laboratory experiments); S – Simulated; HTM – heat transfer model; CM – capacitor method

Although it is valuable on its merits regarding different aspects of temperature control, the extant research has so far failed to address some of the practical requirements for a cold SC. In this respect, we identify three main points that distinguish our work from previously published research.

First, we observe that the literature (especially in categories *c*, *d*, *e*, and partially *a*) assumes or relies on the knowledge about the *exact locations of sensors and pallets and the container load allocations*. Moreover, the deployment of a certain number of sensors in relevant positions is required [18-20] for temperature predictions in a STD setting. This knowledge definitely allows for building models for temperature approximations in areas without sensors. Quite common in practice, however, are cases in which there is one sensor per pallet or one sensor per shipment [17] due to cost considerations or cargo characteristics. In addition, cargo location and sensor position can remain unreported, or such placement is done voluntarily. This contradiction makes the transfer of developed solutions to real-life applications somewhat problematic. In contrast, we do not assume any knowledge of sensor or cargo position inside a container and consequently do not rely on temperature mapping inside a container with respect to other factors (e.g., cooling modes, briefly opened doors, and current container position in a controlled temperature room or on the tarmac).

Second, the literature on temperature (failure) prediction (see category *d*) focuses on *long-term temperature approximations and temperature change predictions with stable, or relatively stable, ambient temperature*. For example, [14] considers time horizons extending over months in temperature deviation predictions. Authors of [15] work on the prediction of temperature fluctuations without specification of a forecasting time frame. By predicting temperature as a function of sensor position, [16] provides long-term temperature approximations that show repeating patterns extending over longer periods of time. A long-term temperature approximation based on sensor readings in different positions is proposed in [17]. In this paper, we focus on predictions under the condition of (constantly) unstable ambient temperatures. Such instabilities, even if they are brief, directly influence cargo temperatures and affect the accuracy of cargo temperature predictions in the immediate future if they are not accounted for.

Third, in the extant research, *temperature profiles are experimentally developed* for different cooling scenarios and environmental conditions, and *measurements are either simulated or collected in the course of pilot system tests or case studies* (marked as S or E in Table 1). Although such simulated or experimentally set

up temperature profiles may reflect a multitude of typical environmental conditions in a container, some may be overrepresented, while others may skip under the radar. In our paper, we use longitudinal real-world data collected during almost six years that encompasses different cooling regimes and SC configurations. Last but not least, our focus on a cold pharmaceutical SC complements the abundant research in the field of FS.

3. Proposed adaptive method

NLC states that the rate at which an object's temperature changes is proportional to the difference between the object's temperature and the ambient temperature [39], which can be expressed in a differential equation that expresses the difference of temperatures as a function of time:

$$dT_{object}/dt \propto T_{object} - T_{ambient} \quad (1)$$

By solving this equation, exponential decay of the difference of temperatures can be shown over time. Depending on the difference of the object's and ambient temperature, the energy is flowing from or to the object [39]. To put it into plain language, if the object's temperature is much higher (slightly higher) than the ambient temperature, the former decreases at a relatively high (relatively low) rate before approaching the ambient temperature. The same logic holds true for the object's temperature much or slightly lower than the ambient temperature. At the same time, the rate at which temperature changes also depends on other physical properties of the object (e.g., as listed in Table 2 and on the nature of fluid motion, surface geometry, and an assortment of fluid thermodynamic properties [40,41]). Despite its simplicity and utility in temperature predictions, NLC assumes that *the ambient temperature remains unchanged*, the condition that is often violated in the context of a cold SC.

For the purposes of this paper, we present in this section a brief formalization of NLC and propose a method for the adaptation of NLC to the prediction of object's (cargo's) temperatures under the condition of ambient temperature instability (thus relaxing the assumption above). The notation for NLC and main variables of the proposed method are listed and explained in Table 2 below.

Table 2. Notation for NLC and the proposed adaptive method

NLC	Proposed adaptive method
$T(t)$ – cargo's temperature at time period t	e^a – actual measurement error

T_{env} – ambient temperature in container	e_t^e – error estimate at time period t
h – heat transfer coefficient	m_t^a – actual measurement at time period t
A – heat transfer surface area	m_t^e – measurement estimate at time period t
m – cargo's mass	*MEI* – measurement error interval
c_p – cargo's mass-specific heat capacity	*STM* – stable measurements for Kalman filter
t – time index (integer)	*inputVector* – stable measurements for prediction

Verbal description of NLC in the first paragraph of this section can be formalized as follows [40,41]:

$$T(t) = T_{env} + (T(0) - T_{env})\exp\left(-hAt/mc_p\right) \quad (2)$$

If we look at the right-hand part of Eq. (2), we may notice that, for a cold SC, hA/mc_p permits some simplifications. In addition to the commonly assumed temperature independence of h, A of the packaging of the transported pharmaceutical, its m, and, consequently, its c_p remain constant during transport. We can ignore possible cargo movements during physical handling that may affect A.

The independence of h from the current temperature cannot be guaranteed in classical convective heat transfer [42]. However, the changes in h that occur with changes in temperature can still be approximated, and NLC can still provide good approximations of the temperature trajectory for small temperature changes. Given the eventual variability of t h and the constant nature of the other terms, we can simplify Eq. (2) to

$$T(t) = T_{env} + (T(0) - T_{env})e^{-rt} \quad (3)$$

Eq. (3) still implies a temperature stability [43] in which T_{env} remains steady and $T(t)$ approaches T_{env} over the course of time. In practice, however, the temperature within a container unavoidably demonstrates instabilities as the ambient temperature fluctuates. This poses a challenge when the curvatures of temperature are predicted in real time. In particular, the question regarding temperature stability conditions remains open, i.e., at what number of specific measurements do we regard temperature as stable, and when does it cease to be so? Consequently, we must determine *how many previous measurements* typical of a stable temperature should be considered when making predictions and how *initial object temperature* is identified.

In our method, we consider the *temperature gradient (sign)* of the latest measurements, observe whether the temperature measurements demonstrate *exponential decay*, and allow for sensor *measurement error* to answer the above question. In essence, if the combination of these conditions suggests temperature

instability, we discard the measurements associated with the previous stable temperature and consider the last state of that stable temperature as the initial state of the new stable temperature. At the same time, $T(0)$ of the new temperature development is derived as the last state of the previous stable temperature or is approximated if the measurement error condition applies.

The *first condition* that we consider is the sign of the most recent measurements. It is quite logical to assume that the temperature is stable if the temperature gradient remains unchanged. In other words, a *minus* sign signifies a continuous decrease, and a *plus* sign, or equality of measurements over time, signifies a continuous increase. If the sign of the current measurement does not correspond either to a succession of unchanging signs or to the latest calculated sign, we observe a temperature instability.

The *second condition* for a stable temperature is the exponential decay in the magnitude of the differences between successive measurements. If we examine the right-hand side of Eq. (3), we observe that, assuming the abovementioned independence of the heat transfer coefficient and the temperature, the only changing term is the time index (T_{env} does not change for a stable temperature). Given the incrementally increasing values of t, T_{env} is appended each time with an exponentially decreasing product. This reasoning corresponds to the idea that the object's temperature changes more rapidly in the direction of the ambient temperature in the first periods of temperature stability. This condition is satisfied if $|T(t) - T(t-1)|/|T(t-1) - T(t-2)| \leq 1$.

The *third condition* assumes that incoming measurements are subject to error within the scope stipulated by the sensor's manufacturer. This means that we ignore variability in temperature readings for the first two conditions if this variability is within the range of measurement error. This condition helps avoid a situation in which temperature instability is falsely identified based on the first two conditions and allows for other physical processes in a container thermal system that we assume to have no effect on our method (e.g., other fundamental modes of heat transfer).

We use the Kalman filter for the approximation of $T(0)$ in the case of possible measurement error. Based on the estimation of a joint probability distribution, this algorithm produces a more accurate value of a certain variable over the course of time given some historical observations of that variable [44]. The general form of the Kalman filter is applicable to scenarios that involve multiple sources of data. Simplification to the scenario

in which a single sensor is used uses the recursive updating scheme illustrated in Fig. 1 [45]. Given the initial e^a (provided by the manufacturer) and the m_t^a, more accurate estimates of the true temperature can be produced in subsequent iterations. Since the original error estimate (e_0^e) and the original measurement estimate (m_0^e) do not influence the value to which the object's temperature converges, only two input values are needed for the current measurement estimation of the new stable temperature, i.e., $T(0)$.

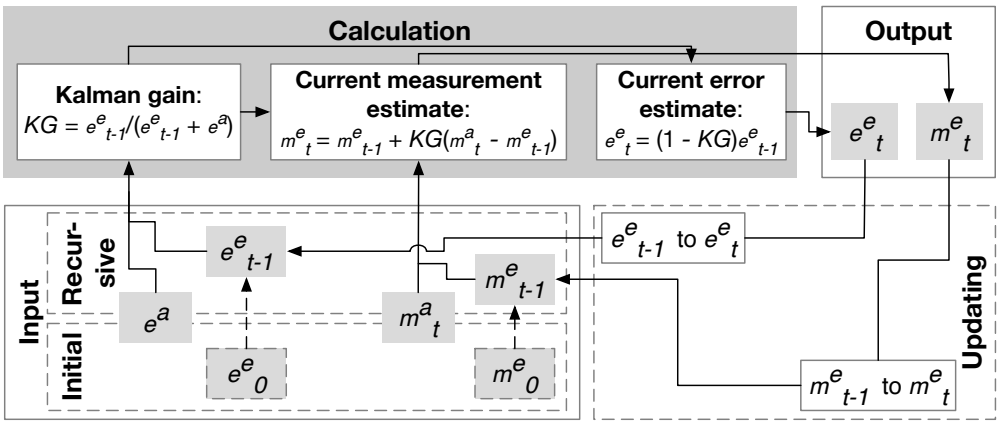

Fig. 1. Kalman filter calculation for a scenario involving a single sensor

Despite their logical clarity and simplicity, the abovementioned conditions allow for manifold combinations that either signal or dismiss temperature stability. To illustrate our method in general, we provide a pseudocode in Fig. 2 that generates with each incoming temperature reading a measurement vector that characterizes a stable temperature. As input, our procedure uses e^a and the incoming event steam containing temperature measurements. For further evaluations, we declare two variables: a) *MEI* that contains an upper and a lower bound that define the range of temperatures that fall within e^a; b) a vector with readings characterizing a stable temperature measurement (*STM*).

For each iteration in the loop, we first check whether the latest reading lies within the bounds of *MEI* (line 6 or line 26, depending on its length). If the temperature is identified as stable (lines 20–24 or 27–31), the affected variables are updated; in this case, the produced measurement vector differs from the previous vector with respect to the last appended measurement. Otherwise, (lines 8–18 or 33–54), *MEI* is reset, and if the previous measurement variation could be explained through a sensor error, $T(0)$ is estimated based on the readings in *STM* with the help of a Kalman filter. This is followed by resetting *STM* and updating other variables (lines 11–17). If the variability of the previous measurements could instead be explained by temperature instability, both *MEI* and *STM* are reset, and the sign and slope variables are updated. Based on

the updated values, one of three possible combinations of sign and slope variables is identified (lines 44–52), and the values for $T(0)$ and input vector are reassigned accordingly. Finally, the list *inputVectorList* is appended with the current input vector that characterizes a new (or existing) stable temperature and whose readings are used for training/fitting and ensuing predictions. Please note that the choice of this form of output, i.e., a list-in-list, is arbitrary and does not exclude other representations.

```
Procedure: Input vector construction
    Input: incomingMeasurements, e^a
    Initialize: MEI ← [∅], ΔT(t-1) ← 0, ΔT(t) ← 0, sign ← [∅], slope ← 0, inputVectorList ← [∅],T(0) ← 0
    Output: inputVectorList
 1  inputVector ← [m^a_0]
 2  MEI ← MEI.append(m^a_0)
 3  STM ← STM.append(m^a_0)
 4  T(0) ← m^a_0
 5  for each m^a_t in incomingMeasurements do
 6  |   if length(MEI) = 2 then
 7  |   |   if m^a_t < min(MEI) or m^a_t > max(MEI) then
 8  |   |   |   MEI ← [m^a_t]
 9  |   |   |   T(0) ← KalmanFilter(STM, e^a)
10  |   |   |   STM ← [m^a_t]
11  |   |   |   inputVector ← T(0)
12  |   |   |   inputVector.append(m^a_t)
13  |   |   |   if inputVector[–2] < inputVector[–1] then
14  |   |   |   |   sign.append(+1)
15  |   |   |   else
16  |   |   |   |   sign.append(–1)
17  |   |   |   ΔT(t-1) ← | inputVector[–2] – inputVector[–1] |
18  |   |   |   inputVectorList.append(inputVector)
19  |   |   else
20  |   |   |   STM.append(m^a_t)
21  |   |   |   inputVector.append(m^a_t)
22  |   |   |   sign.append(0)
23  |   |   |   ΔT(t-1) ← 0
24  |   |   |   inputVectorList.append(inputVector)
25  |   else
26  |   |   if |MEI[0] – m^a_t| <= e^a*2 then
27  |   |   |   MEI.append(m^a_t)
28  |   |   |   STM.append(m^a_t)
29  |   |   |   inputVector.append(m^a_t)
30  |   |   |   sign.append(0)
31  |   |   |   ΔT(t-1) ← 0
32  |   |   else
33  |   |   |   MEI ← [m^a_t]
34  |   |   |   STM ← [m^a_t]
35  |   |   |   if inputVector[–1] > m^a_t then
36  |   |   |   |   sign.append(–1)
37  |   |   |   else
38  |   |   |   |   sign.append(+1)
39  |   |   |   ΔT(t) ← | inputVector[–1] – m^a_t|
40  |   |   |   if ΔT(t-1) ≠ 0 then
41  |   |   |   |   slope ← ΔT(t)/ΔT(t-1)
42  |   |   |   else
43  |   |   |   |   slope ← 1
44  |   |   |   if slope >= 1 and sign[–2] ≠ sign[–1] then
45  |   |   |   |   T(0) ← inputVector[–1]
46  |   |   |   |   inputVector ← inputVector[–1]
47  |   |   |   elif slope >= 1 and sign[–2] = sign[–1] then
48  |   |   |   |   T(0) ← inputVector[–1]
49  |   |   |   |   inputVector ← inputVector[–1]
50  |   |   |   elif slope < 1 and sign[–2] ≠ sign[–1] then
51  |   |   |   |   T(0) ← inputVector[–1]
52  |   |   |   |   inputVector ← inputVector[–1]
53  |   |   |   inputVector.append(m^a_t)
54  |   |   |   ΔT(t-1) ← | inputVector[–2] – inputVector[–1] |
55  |   |   inputVectorList.append(inputVector)
```

Fig. 2. Input vector generation for a stable temperature

NLC, with its variations and adjustments, has been used in STD estimation in a cold SC. For example, [14] applied heat transfer modeling to produce an estimate of product temperature that depends on time and external conditions. NLC was used by [38] to simulate product temperature variability data for solving the problem of optimizing a vehicle's traveling distance considering product perishability. A capacitive heat transfer model was deployed in [34,35] for STD estimation. Finally, [17] built an ANN for temperature distribution inside a pallet inspired by the heat transfer model. We rely on this prior research as a motivation for the use and subsequent adaptation of this law in our paper.

4. Data and experiments

In this section, we first describe the business case and how temperature data were collected. This is followed by a description of the baseline methods. We then show how training and test data were generated. The section concludes with a description of the experiments and the metrics used in the methods comparison.

4.1. Business context and collected data

In this paper, we focus on the case of a cold pharmaceutical SC of a large international Europe-based logistics service provider (LSP) rendering air transportation services with temperature monitoring in over 100 countries. The LSP serves over one hundred customers on four continents; its customers require intercontinental transportation of time-temperature sensitive pharmaceuticals. At the end of 2018, this LSP had collected over 14 million temperature measurements with over 30,000 sensor devices.

The operations of the LSP reflect a typical cold air SC consisting in its simplest form of at least three legs (for an operational view of the first two legs, see Fig. 3a). The first link connects a manufacturer with the warehouse facilities of the LSP or carrier and is performed by the LSP via road transportation. At this stage, the manufacturer prepares its cargo by providing proper packaging, prestorage, and attachment of wireless sensors that are assigned to and associated with a shipment by the LSP. At the time of sensor assignment, the LSP has already been notified by the manufacturer of the results of stability studies for the transported pharmaceutical and assigns the lower and higher temperature thresholds (temperature regime).

The sensors associated with an individual shipment either temporarily save measurements with corresponding timestamps if a network is not available or build a mesh network and transmit the previously

saved and current measurement events via a gateway, enabling constant temperature control by the LSP. Event data arriving at the measurement event server are processed by the designated rule-based system based on the associated temperature regimes, and deviations are identified in real time. For a sensor network view typical of the entire cold air SC, see Fig. 3b.

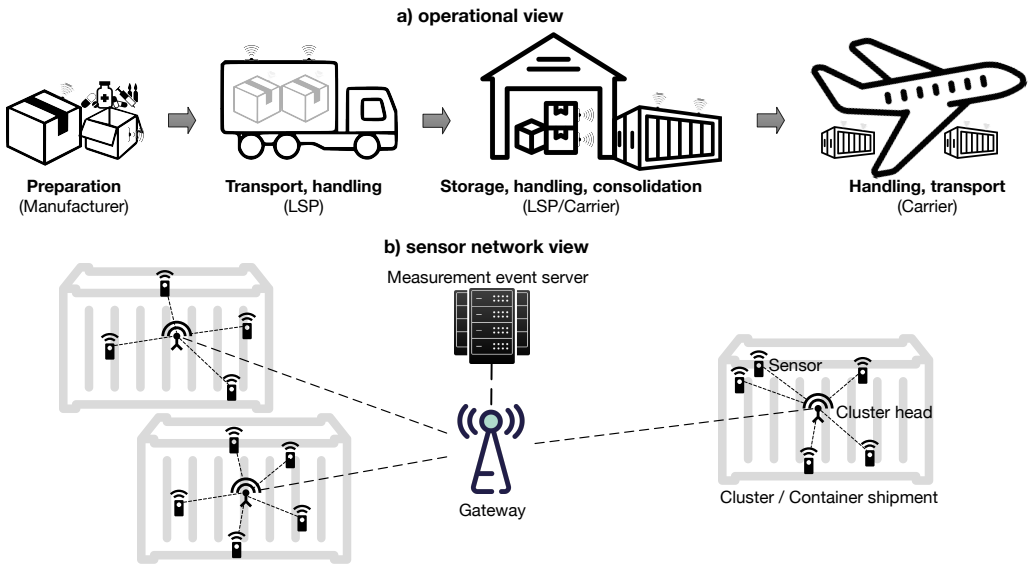

Fig. 3. Operational and sensor network views of a cold pharmaceutical air SC

As a next step, the shipment is picked up by the LSP and transported either to its warehouse facilities prior to delivery to the facilities of the carrier or directly to the carrier. This transport usually requires six hours or less and connects not very distant geographic locations. Attachment and assignment of additional sensors and shipment consolidation by the LSP are possible at this stage. After the shipment has been delivered to the carrier's facilities, temporary storage occurs before the next leg begins. The second link connects the airport of origin and the destination airport and is performed by the carrier by air. During such transport, temperature measurements are saved locally due to the unavailability of the network and are transmitted upon landing. However, in-flight temperature is controlled by the carrier's personnel. Cargo unloading is followed by temporary storage in temperature-controlled facilities before it is picked up by the LSP or its subcontractor for delivery to the consignee (the last leg). In practice, a cold SC incorporates more phases, in part depending on the origin–destination pairs, and unavoidably includes additional formalities, i.e., customs.

Taking a look at the operational view of a cold air SC, one cannot but notice that there are numerous occasions on which temperature deviations may occur. For example, each case of physical handling in which exposure to the ambient temperature is unavoidable may cause an excursion. Placement of a shipment on an

airport apron prior to loading or immediately after offloading (tarmac time) often results in deviations. The placement of cargo in temperature-controlled rooms with specific cooling profiles or erroneous storage at the wrong temperature can also lead to brief excursions or to longer ones.

Immediately after the event data have been processed by a rule-based system and a deviation has been identified, the personnel of the LSP are notified, and the case is investigated. The parties responsible for handling the affected shipment are contacted. Thereupon, corrective actions are required if it is determined that a deviation should be rectified through human involvement. Alternatively, monitoring of the situation is advised if an excursion is brief and unavoidable, e.g., cargo loading on a sunny day in a location with typically high temperatures. During the study period of 2013–2018, over 19.000 deviations were registered, and some of these called for time-consuming investigations. High time-temperature sensitivity of pharmaceutical products and the abundance of excursions during storage and transportation necessitate proactive identification of impermissibly low or high cargo temperatures.

For the development of a method for temperature prediction, we rely on the data collected by the LSP since the introduction of air transportation services for a cold chain. The data comprise measurement event data, associated temperature regimes (lower and higher thresholds), measurement timestamps, and a setpoint temperature. From the moment of sensor assignment to that of sensor deactivation, the data encompass monitoring periods of up to over a week for one shipment. Temperature measurements are collected at a predefined sampling rate (in our case, every 10 minutes).

Although the business case presented and the data used in this paper clearly characterize a temperature-controlled scenario of air cargo transportation, we do not lose generalizability for real-time temperature predictions for other types of sensitive goods if some or all of the following conditions apply:

a) The cold SC comprises more than one party responsible for maintaining permissible temperature regimes (a higher number of parties implies more physical handling points and possible resulting deviations);

b) Measurement event data arrive in real time, but some delays in transfer (may) occur;

c) Transported goods are properly packaged, implying that temperature deviations of the packaged contents are slightly delayed rather than immediate when changes in the ambient temperature occur;

d) Temperature mapping studies are not available and there is no, occasional, or incomplete knowledge of cargo position in a container or sensor location on a pallet; these conditions invalidate the application of predictive approaches based on STD.

4.2. Baseline methods

Evaluation of the performance of NLC accompanied with our adaptive method involves a comparison of its predictive accuracy with that of other methods. In this section, we briefly describe two baseline methods, ANN and autoregressive moving average (ARMA), and justify the choice of these methods for use in our paper.

The first baseline method, ANN, or multilayer perceptron, is a type of feedforward network consisting of an input layer, hidden layer(s), and an output layer; it is used in supervised learning problems (classification or regression). The idea of a perceptron dates back to fundamental research on brain cell activity [46] and to further attempts to automate the learning of optimal weight coefficients to decide whether or not a neuron will fire [47]. The literature on ANN is immense, and the method's success has been demonstrated in many applications; an interested reader is referred to the literature summarizing the recent developments in the field [48]. In Section 4.3, we discuss in detail how input vectors are generated.

ANN have been successfully applied in multiple disciplines and in recent years also for the purposes of temperature prediction in a cold SC. In particular, [15] applied a back-propagation neural network to predict temperature shifts in cold chain monitoring. Prediction of environmental parameters in container transports with the help of ANN was proposed in [19]. In [35], product temperature was estimated via ANN on the basis of sensor measurements outside the pallet. Temperature estimation of food products in different refrigeration failure scenarios with the help of ANN was demonstrated in [34]. Finally, [17] used ANN to perform temperature distribution estimation in a pallet. In light of the applicability of ANN to temperature prediction in a cold SC, we also explore its performance in our context.

The other baseline method, the autoregressive moving average (ARMA) model, was popularized by Box and Jenkins for time series analysis [49]. An ARMA model consists of two parts, i.e., autoregression (AR) and moving average (MA), that together provide a parsimonious description of a stochastic process. AR (p) involves regressing time series values on their own lagged value, and the MA (q) part models the error term as a linear combination of error terms observed at the same time and at different times in the past.

The model parameters, or the model order (p, q), should be estimated separately for each case of input time series data. A popular criterion for the choice of p and q is the Akaike information criterion. In the case of small sample sizes (actually, the situation that we face in our context), the criterion is prone to the problem of overfitting [50]. To avoid the possibility of overfitting, we use the corrected Akaike information criterion, which is calculated for sample size n as follows [51]:

$$AICc = -2\log(L) + 2(p+q+k) + \frac{2(p+q+k)(p+q+k+1)}{(n-p-q-k-1)} \quad (4)$$

where L is the maximum value of the model likelihood function, p and q are the parameters of the ARMA model, and $k = 1$ (because the variance of an error term is also estimated).

ARMA models have so far been applied to time series forecasting extending over longer periods [52-55]. Nevertheless, we choose this family of models for our analysis because manifold predictive models can be obtained through estimation of the model order that best fits previous historical temperature measurements. For instance, because our focus is on the deviations that occur within relatively short time frames, we will clearly not observe periodicity in our temperature curves. However, we can expect to obtain a series of observations that can be modeled using linear regression with lagged variables (e.g., equally paced increases or decreases). These can be approximated by an ARMA model with the corresponding order.

The choice of a baseline method that learns a target function from a set of observations, i.e. input vectors with training data, (ANN) or another baseline method that selects an optimal forecasting model based on one observation (ARMA) provides a good comparison of data-driven models with a theory-based model (NLC).

4.3. Generation of training and test data

In this paper, we conduct two main groups of experiments for NLC, ANN, and ARMA. In the *adaptive approach*, our proposed method that allows for the application of NCL to the prediction of cargo temperatures under the condition of ambient temperature instability is used for the selection of the optimal number of previous measurements (see Fig. 2 for how vectors of input data are generated). In the *non-adaptive approach*, a range of fixed numbers of historical measurements for prediction (i.e. as if we did not rely on our proposed method and wanted to experimentally find the optimal number of previous measurements). This distinction is introduced to assess the value of our method not only in the context of NLC but also in the context of ANN and ARMA. All experiments are set up for predictions of up to one hour. Given a sampling rate of one

measurement per 10 minutes, predictions for the next six observations are made[1]. We consider this period sufficient for the implementation of contingency plans in practice.

In the non-adaptive approach, we used measurement events obtained during the four hours prior to a deviation and the four hours thereafter and reserved the last six measurements for a test set[2]. The preceding 42 readings were sliced for 11 setups in which the training data contained two through 12 measurements. For example, for a data set consisting of two events for training and the customary six readings for testing, we used a sliding window of size two with a step of one measurement event. The six subsequent events were taken as test data. For each measurement vector, the procedure was repeated until the final six test observations were reached. Data sets that included the same number of training events were merged in the end. An analogous procedure was followed in the remaining ten non-adaptive setups, the only difference being the number of observations taken for training.

In the adaptive approach, our method (see Fig. 2) was deployed for the adjustment of vector length for training. To make the comparison with the non-adaptive approach possible, we focused on observations with two through 12 training points. In the end, we had 22 data sets for both approaches. For computational reasons, data set size was limited to 100.000 observations. In cases with higher numbers of rows in the generated data sets, the maximum permissible number of rows was randomly sampled.

The choice of two through 12 measurements as training data was, on the one hand, identified as appropriate in the course of experimentation. Considering that each measurement vector contained 48 events, increasing the training vector length was not feasible in the adaptive approach since most of the deviations could be observed after fewer than 12 first events. On the other hand, when temperature instability is recognized, at most two measurements are available for making a prediction. Although the use of two measurements may introduce inaccuracies in the forecasts for the next six periods, it is nevertheless relevant to investigate the reliability of forecasts made in a more challenging but still quite realistic scenario.

---

[1] We focus on short-term predictions; as a convention, among such short-term predictions, we differentiate between *shorter-term* (for 10, 20, and 30 minutes) and *longer-term* predictions (for 40, 50, and 60 minutes).

[2] We limit the whole collection of measurements to eight-hour time frames around registered deviations.

4.4. Experiments and metrics

The efficiency of the methods described in Sections 3 and 4.2 can be regarded from two main perspectives, that is, the *success of deviation prediction* and *prediction accuracy (error)*. As described in Section 4.1, the LSP personnel react to alarms triggered during event data analysis and alarm rule evaluation when a temperature measurement falls outside the permissible range. If the produced forecasts coincide with actual future measurements that lie outside the temperature range, alarm personnel can be successfully notified of possible future deviations before they occur. Although this type of deviation prediction is helpful and informative, experiments measuring its success have so far not been presented in the extant literature on cold SCs.

Therefore, in our first experiment of *deviation prediction*, we report true positives (TPs) for correctly predicted deviations, true negatives (TNs) for correctly predicted temperatures within range, false positives (FPs) for incorrectly predicted deviations, and false negatives (FNs) for incorrectly predicted temperatures within range. These metrics allow for a thorough assessment of a method's ability to reliably predict potential deviations and its proneness to omissions of critical situations.

Although the abovementioned metrics are indicative of forecasting performance, they may be not sufficient on their own to allow the selection of a better-performing method. In particular, correctly predicted deviations do not provide any information about the magnitude of the deviations. This limitation is critical in the case in which decisions should be made regarding the order in which deviations are approached (i.e., ranking of predicted deviation magnitudes). For this reason, in the second experiment of *prediction accuracy*, we calculate two additional evaluation metrics – prediction error measured as a root-mean-square error (RMSE) and its standard deviation (SD). The RMSE and its SD show how the predictions made by the compared methods differ from the actual measurements. They are also more sensitive to the exactitude of forecasts since they penalize correctly predicted deviations and measurements within the permissible range if the predicted measurements do not equal the actual ones.

Finally, we conduct the third experiment, i.e. *an execution time test*, to shed light on scalability limitations and potential trade-offs (i.e., accuracy vs. computational burden) that users may face in implementing these methods. The test is performed on a machine running Windows Server 2016 equipped with a six-core Intel Xeon Gold 6126 CPU clocked at 2.60 GHz and 64 Gb RAM. Multithread processing is not enforced.

In the case of NLC, the values of a positive constant r and T_{env} are estimated using a trust region reflective algorithm [56,57] with bounds for T_{env} based on the historically lowest and highest registered ambient temperatures and r satisfying the constraint of constant positivity. Predictions for six periods are calculated based on the estimated values of r and T_{env} by incrementally updating the time index t for six periods. Open-source Python library *SciPy* is used for estimating parameters and making predictions.

For ANN, we use grid search [58] for the optimization of hyperparameters (i.e., activation function, learning rate schedule for weights, learning rate initialization, momentum for gradient descent update, and regularization term). Architecture configurations, i.e., number of hidden layers and neurons, that help avoid overfitting are used in the grid search [59]. Optimal hyperparameters are found in a 10-fold cross-validation procedure with an L2-norm cost function. 500 epochs without early stopping are used for training ANNs before convergence is reached. Direct forecasting is deployed for ANN: ANNs are trained, depending on the evaluation setup, on two through 12 past measurements and a one-time prediction is made for $t + 1$ through $t + 6$ periods. Machine learning library *scikit-learn* is used in the implementation of this method.

For ARMA, recursive forecasting is deployed, i.e., after data fitting and model order estimation based on AICc, a prediction for period $t + 1$ is made. This prediction is appended to the training data, based on what the next five predictions are made iteratively. Search domain for p and q is equal to the number of training points in each evaluation setup. The Python module *statsmodels* is used for implementing ARMA models.

5. Results and discussion

This section presents the results of the experiments described in Section 4.4 and provides an accompanying discussion. In particular, deviation prediction analysis is performed for the adaptive and non-adaptive approaches. It is followed by prediction error analysis for the same experimental setups. Finally, the results of the execution time tests are presented to conclude the section.

5.1. Deviation prediction

Our first evaluation is based on the TP, TN, FP, FN counts, which reflects the success of predicting whether further measurements lie within or outside a permissible temperature range. Table 2 presents the average scores for predictions for $t + 1$ through $t + 6$. As mentioned in Section 4.3, data sets in evaluation setups for two through 12 training points were limited to 100.000 observations, so, instead of providing absolute counts,

we report shares of TPs, TNs, FPs, and FNs in overall predictions. Best results for each method and metric are provided in **bold** and worst – in *italics*. For page space considerations, in Table 3 and in Fig. 4 through Fig. 6 below, we report the results for two, three, seven, and 12 training points but not for all numbers of training points. The omitted results are mostly approximated by linear interpolations, except the more abrupt changes that occur between two and three training points that are always provided.

Table 3. Deviation prediction performance (adaptive approach above, non-adaptive – below)

	ARMA				**ANN**				**NLC**			
	2	**3**	**7**	**12**	**2**	**3**	**7**	**12**	**2**	**3**	**7**	**12**
TP	*35.11*	**46.15**	42.42	42.51	*34.68*	**48.27**	45.12	43.88	*30.92*	49.28	49.86	**50.21**
	41.55	44.06	43.33	40.09	40.01	44.09	44.25	39.50	42.85	42.81	41.99	40.62
TN	39.25	43.93	41.22	36.98	**44.50**	41.22	39.08	33.78	**51.12**	46.21	43.94	*39.83*
	44.06	40.95	35.71	*35.08*	39.54	42.05	30.43	*28.67*	47.83	49.12	48.69	46.91
FP	8.05	**4.10**	6.24	7.59	7.95	**3.98**	5.62	8.03	6.59	**1.99**	2.99	4.05
	6.24	7.38	8.85	*10.22*	11.13	7.21	9.33	*14.81*	5.73	4.58	5.63	*8.35*
FN	*17.59*	**5.82**	10.12	12.92	12.87	**6.53**	10.18	14.31	*11.37*	**2.52**	3.22	5.91
	8.15	7.61	12.11	14.61	9.32	6.65	15.99	*17.02*	3.59	3.49	3.69	4.12

In general, as can be seen in Table 3, we observe that all methods usually achieve best results in the adaptive approach, whereas worst results are typical of both approaches. NLC yields on average better results across all metrics and evaluation setups, whereas ANN and ARMA achieve second-best results without any systematic advantage of any method. However, ANN usually yields better results, in particular in terms of TNs, TPs, and FPs, in the adaptive approach in comparison with ARMA. Conversely, in the non-adaptive approach, ARMA shows somewhat better results than ANN.

ARMA demonstrates a good discoverability of deviations (TPs) in the *adaptive approach* beginning with at least three training points and preserves good results for up to 12 training points (performance decrease of about 8%). However, ARMA's discoverability of normal temperature (TNs) is less stable for predictions with more training points (decrease of 16% for three vs. 12 training points) and is usually lower than the share of TPs. ARMA makes least incorrect deviation predictions with three training points, whereas other numbers of

training points lead to the results of around 7% of FPs. It can also be observed that ARMA has a problem with incorrectly predicted normal temperatures (FNs). Although the method yields a comparably good result with three training points, other predictions are usually worse in comparison with ANN and NLC.

In the *non-adaptive approach*, ARMA shows no systematic advantage over its predictions of TPs and TNs in the adaptive approach. At the same time, except for predictions with two training points, shares of FPs and FNs are always higher by about 2% in the non-adaptive approach.

ANN's performance in the *adaptive approach* is usually characterized by more accurate predictions with three training points. For example, it discovers TPs well, and for seven and 12 training points their share decreases by slightly more than 1%. Share of FPs is also at its lowest for three training points, and increases slightly predictions with more previous measurements (about 2%). Conversely, share of TNs decreases abruptly with the increasing number of training points. Also, the share of FNs is constantly higher than the share of FPs and exceeds FNs achieved by ARMA and NLC. This can be disadvantageous in practice because alarm personnel will have less time to react to some overseen deviations.

In the *non-adaptive approach*, ANNs trained on at least three historical measurements yield consistently worse results than in the adaptive approach. For example, ANN usually discovers fewer TPs and TNs (by at least 1%) and suffers from misclassifications in the form of FPs (about 6% oftener) and FNs (over 5% oftener).

NLC yields generally leading performance in the *adaptive approach* when compared with other methods or its performance in the non-adaptive settings. With the exception of TPs based on predictions from two training points, NLC correctly predicts deviations up to 8% better than ANN or ARMA. NLC boasts superior predictions of normal temperatures (TNs) for any number of training points in comparison with other methods. NLC also manages to make two times fewer mistakes in predictions by achieving generally lower shares of FPs and FNs than ANN and ARMA. In the *non-adaptive approach*, NLC delivers predictions of deviations (TPs) that are generally comparable with the results of ANN and ARMA. What regards TNs, NLC manages to achieve generally better results than its performance in the adaptive approach or the performance of ANN and ARMA. This can be explained by more stable forecasts by NLC that are not immediately affected by temperature instabilities in the non-adaptive approach (the case that we usually observe with measurements within a permissible temperature range that show slight variability). At the same time, although the share of FPs of NLC is higher

in the non-adaptive approach, it is still below the levels of other methods. Also, NLC commits errors by predicting normal temperatures (FNs) two to three times less often than ANN or ARMA.

Overall, we observe that all methods begin yielding better results mostly for three training points, whereas generally better predictions are made in the adaptive approach. At least three training points are required for ARMA to discover a trend in data, for ANN to begin learning a more reliable approximation function, and for NLC to better estimate parameters r and T_{env}. Although the quality of predictions is expected to increase with the increasing number of training points in the adaptive approach, we notice that this effect is countered by the peculiarities of temperature data, i.e. the probability of temperature instability in the future increases with the increasing number of measurements characterized by stable temperature trajectory.

5.2. Prediction error

5.2.1. ANN

As mentioned in Section 4.4, findings based on TPs, TNs, FPs, and FNs oversee absolute prediction errors for both correctly and incorrectly classified deviations and normal temperature measurements. Therefore, in this section, we present the detailed results of a prediction error test. Fig. 4 shows the RMSE and SD for ANN predictions and input vectors of various lengths for both approaches.

Overall, we observe that RMSE and SD for predictions in the adaptive approach are usually lower than those for predictions in the non-adaptive approach. For input vectors with small lengths, the adaptive approach shows much better results for shorter-term predictions. Predictions made on the basis of middle-length input vectors are characterized by the lowest RMSE and SD for the adaptive approach and lower RMSE and SD for shorter-term predictions in the non-adaptive case. Increasing the input vector length negatively affects both approaches; however, the non-adaptive case shows slower increases in RMSE for longer-term predictions.

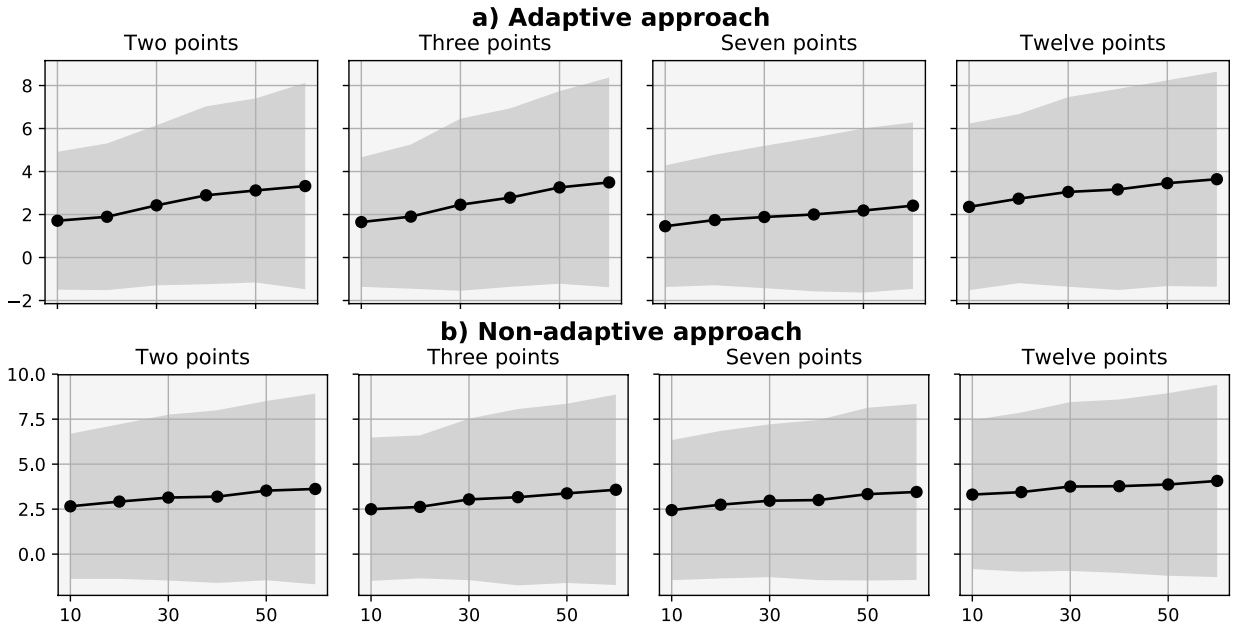

Fig. 4. RMSE (lines) and SD (light gray regions) for ANN predictions for both approaches

Despite the direct forecasting and global minimization of RMSE, ANN still struggles to develop a proper target function from the insufficient signal contained in shorter input vectors. The adaptive approach generates measurements with at least initially lower variance in the test data (better shorter-term predictions). In contrast, the non-adaptive approach unavoidably introduces more noise due to the consideration of both stable and unstable temperatures in the training data. The improved RMSE for both approaches for middle-length input vectors can be attributed to the emerging structure in the temperature data, which can obviously be better learned in the adaptive case. At the same time, the case with longer input vector is plagued, as discussed in other evaluations, with higher variances in the test data. The use of a longer input vector does not affect the non-adaptive approach. One may observe this in the example that shows RMSE increasing more slowly for longer-term predictions, in contrast to the adaptive case. On the other hand, the lower RMSE for shorter-term predictions observed in the adaptive case indicates that at least some data structure can be learned from input vectors derived from periods in which the temperature is stable.

5.2.2. ARMA

Fig. 5 illustrates the results of prediction error evaluations for ARMA. On average, the results are only insignificantly better for the adaptive approach. Additionally, the predictions for the first period are generally similar in both approaches. However, depending on the length of the input vectors, different degrees of deterioration of RMSE can be observed for the adaptive and non-adaptive approaches.

For example, predictions for input vectors of smaller length tend to yield better results for the non-adaptive approach than for the adaptive approach. Only with increasing vector length (observable for seven points in Fig. 5) do the adaptive predictions fare about as well as the non-adaptive forecasts. Lower RMSE values are typical of both shorter- and longer-term predictions in this regard. Further increases in the input vector length are associated with worsening RMSE values (noticeable for 12 points in Fig. 5). Nevertheless, the adaptive predictions tend to continuously outperform the non-adaptive forecasts.

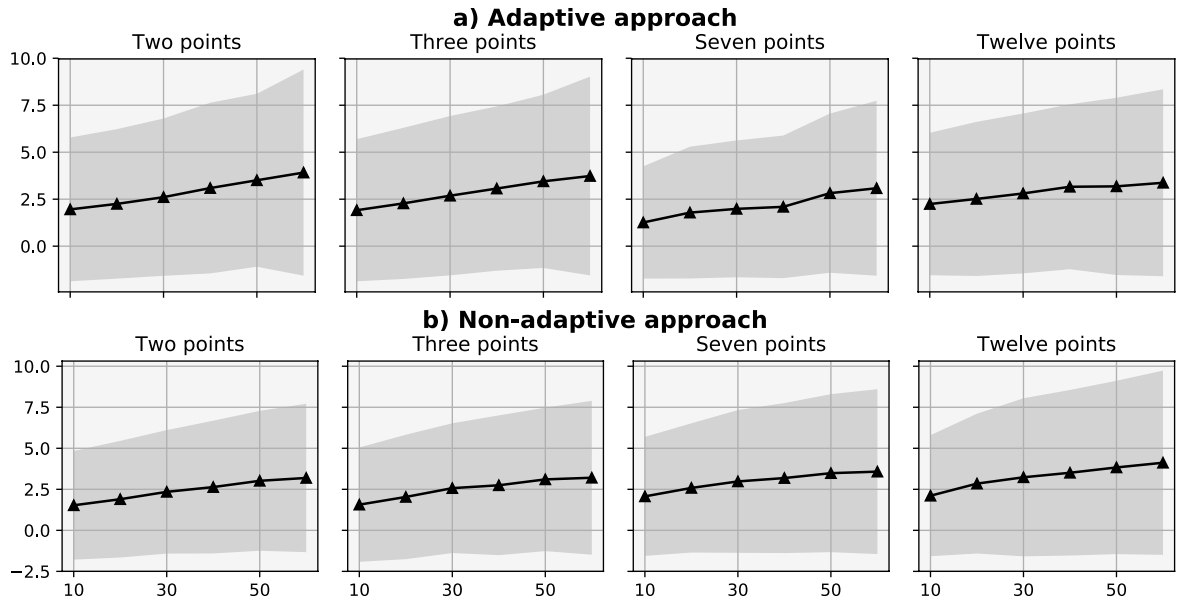

Fig. 5. RMSE (lines) and SD (light gray regions) for ARMA for both approaches

The interesting inconsistency in prediction accuracy for shorter input vectors could be attributed to the fact that a small number of observations are still insufficient as a basis for correctly estimating the trend in the data. For example, for two-dimensional input vectors, forecasts based on the mean value would be logical to assume, i.e., model order (0, 0). In this case, such input vectors of the adaptive approach contain the data for a new stable temperature for which further measurements are expected to deviate from the mean of the first two measurements. Conversely, two-dimensional vectors in the non-adaptive case can contain measurements with a stable temperature whose further measurements can be better approximated by mean predictions.

Additionally, even though the use of somewhat longer input vectors in the adaptive case may allow for the estimation of other model orders, recursive forecasting may fare worse. Due to its iterative nature, the method overshoots or undershoots the exponential behavior of changing temperature, especially for longer-term predictions. Conversely, in the non-adaptive case, such predictions may fare better for longer periods since noisier historical data would lead to estimations that are closer to the mean.

5.2.3. NLC

Fig. 6 presents the prediction error results of NLC for all experimental setups in both approaches. As usual, we observe better performance of NLC in comparison with ARMA and ANN. In the adaptive approach, for instance, the best average RMSE score achieved by the adjusted NLC is 12.61 and 11.25 times better than ARMA's or ANN's best (lowest) score, respectively. At the same time, its highest average RMSE score is 1.79 and 1.64 times lower than ARMA's and ANN's lowest scores, respectively.

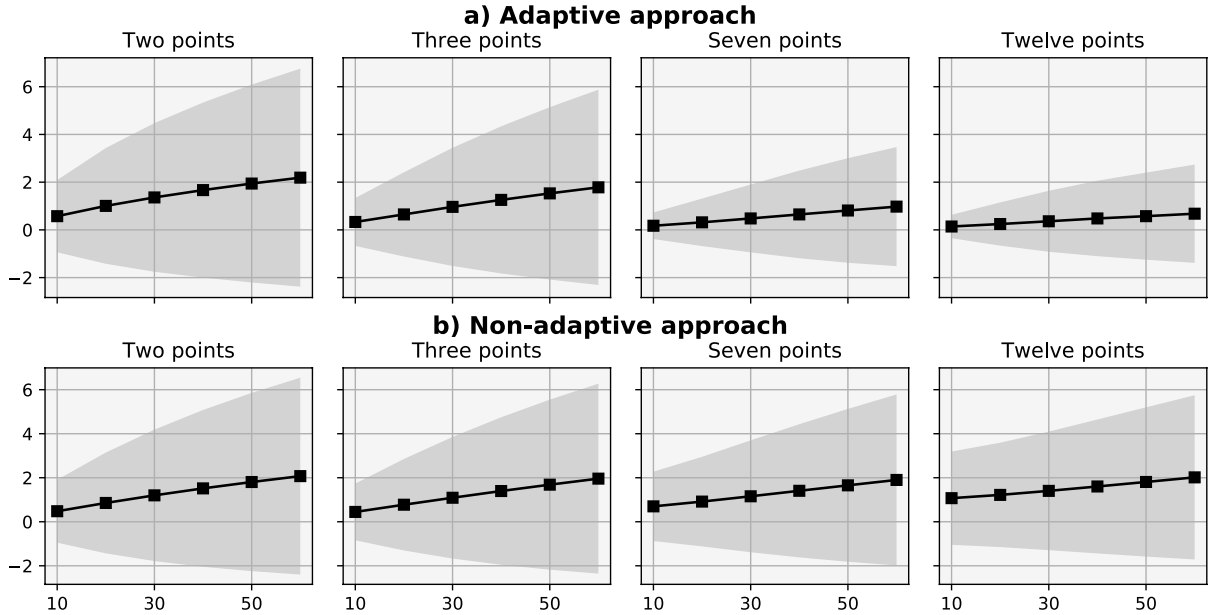

Fig. 6. RMSE (lines) and SD (light gray regions) for NLC for both approaches

A closer look at Fig. 6 further illustrates some of the observations made in Section 5.1. In particular, RMSE and SD for the adaptive forecast are always better than those for the non-adaptive approach, irrespective of the prediction period or the length of the input vector. Only two-dimensional input vectors constitute an exception. These input vectors in the adaptive approach exclusively comprise readings that characterize a new stable temperature. In light of the sampling rate interval, i.e., 10 minutes, estimation of $T(0)$ is prone to inaccuracies since the exact time of temperature instability cannot be identified. Given only two readings, estimation of r and T_{env} is usually difficult. Conversely, RMSE for the non-adaptive approach is improved due to the inclusion of input vectors typical of a stable temperature, for which parameter estimation is trivial.

On the other hand, any additional temperature measurement in an input vector improves RMSE for the adaptive approach in a way that cannot be improved upon by the non-adaptive predictions. Both RMSE and SD are significantly better for adaptive than for non-adaptive forecasts. The reason for this is that longer input

vectors of the adaptive approach are associated with higher variance in the test data. This is reflected in slightly worsening RMSE and SD, especially for longer-term predictions. Despite the higher RMSE and SD values for a two-dimensional input vector, the adaptive approach still guarantees better overall results.

5.3. Execution time

Execution time tests for all setups were run one hundred times, and the mean and SD values of the results were calculated. For ARMA, execution time included the time required for parameter estimation, i.e., p and q, and for six recursive forecasts. For ANN, hyperparameter optimization time was not included in the tests since its value depends on the size of the data set used and on the number and range of possible values of the hyperparameters; only forecasting time was considered. For the adjusted NLC, 1.000 iterations were set for parameter estimation (r and T_{env}), and the time required for prediction calculations was also considered. The results for both approaches are illustrated in Fig. 7, which shows execution time per prediction.

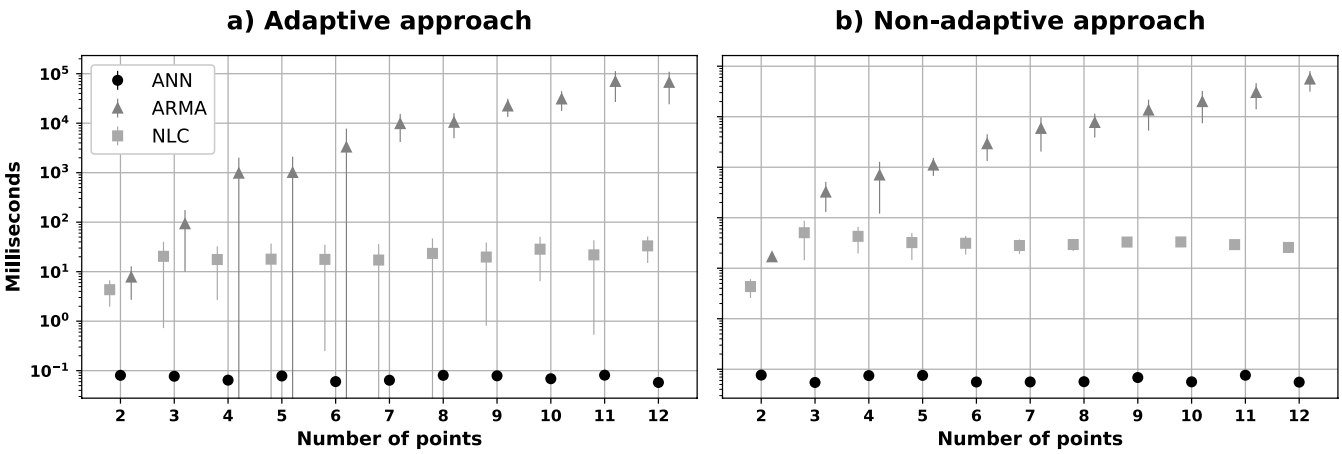

Fig. 7. Execution times for both approaches (mean and one standard deviation range)

In general, we observed no significant differences in execution times between the adaptive (a) and non-adaptive (b) approaches. The adaptive approach guarantees only slightly shorter execution times for certain numbers of input observations. The SD values show the presence of quite low execution times for the adaptive approach, in contrast to the non-adaptive one. The same holds true for SD values for the adjusted NLC. At the same time, NLC is characterized by lower execution times for all numbers of input observations in the adaptive approach. ANN demonstrates next to no variation across approaches and the number of input observations and boasts the shortest execution times.

The observation that ARMA had the longest execution times can be explained based on the computational burden for the model order search. The number of possible parameter combinations increases for input data

with more observations. Even in light of ARMA's hypothetically satisfactory forecasting, such long execution times could be a serious limiting factor in its implementation. NLC demonstrates moderate execution times for all setups. Use of the adaptive approach makes the parameter estimation task easier since approximate optimal values can be found more rapidly for the stable temperature described in Eq. (2) and Eq. (3). This is also reflected in slightly shorter execution times. ANN achieves quite short execution times due to the computational simplicity of its predictions. The forecasting task is tantamount to the multiplication of learned weights with input values, summation, and application of the activation function. If a training data set and the number of optimized hyperparameters are large and there is high retraining frequency, such computational benefits for forecasting may soon be diluted.

6. Conclusions and outlook

In this paper, we proposed a method that enables the application of NLC to cargo temperature prediction in a cold pharmaceutical SC under the condition of ambient temperature instability. NLC was initially defined for stable ambient temperatures that limited its use in dynamic temperature environments. Our method stipulates the conditions required for a real-time selection of past measurements characterized by ambient temperature stability, which overcomes the mentioned limitation of NLC. The temperature gradient, the exponential decay of the temperature slope, and the measurement error with Kalman filtering are used in the elaboration of stability criteria and in the estimation of the initial stable temperature.

According to our evaluations based on longitudinal real-world data for 2013-2018, NLC, ANN, and ARMA fared on average better when accompanied with our method (adaptive approach) in contrast to a choice of any fixed number of previous measurements for making predictions (non-adaptive approach). NLC demonstrated generally better performance than ANN and ARMA, and its superiority was especially pronounced when accompanied with our method.

In the first experiment on deviation prediction, both ANN and ARMA achieved an almost 4% reduction in FPs and FNs and an almost 5% increase in TNs and TPs when using training data with our method in contrast to the non-adaptive approach. NLC accompanied with our method yielded best results achieving on average 3–4% of FPs and FNs, over 43% of TNs and over 49% TPs. In the second experiment on prediction error, NLC accompanied with our method achieved an average RMSE varying from 0.24 for 10-minute predictions to 1.27

for 60-minute predictions. The results for ANN fluctuated on average between 1.29 and 3.01, and those for ARMA – between 1.81 and 3.46 for the respective predictions. RMSE in experimental setups without our method (non-adaptive approach) was constantly higher. The third experiment on execution time hinted at comparably higher computational costs for ARMA, the lowest burden for ANN, and very moderate execution times for NCL with no significant differences across approaches. Considering real-world implementations that include frequent retraining, ANN can soon lose their advantage to NLC.

These findings let us conclude that NLC accompanied with our method (adaptive approach) outperforms other baseline methods (ANN and ARMA) in any approach. This holds true for the reliability of deviation predictions, accuracy of predictions and, in the medium and long term, execution times. The only exception for NLC, ANN, and ARMA in the adaptive approach is the case of a two-dimensional input vector for a newly stable temperature; in that case, two measurement events are still insufficient to make a reliable prediction.

Although ANN has fared well in prior research focusing on temperature estimation, especially in STD, NLC accompanied with our method managed to outperform the former. The main reason for that could be that the previously studied business contexts, i.e. multiple sensors used for the estimation of STD, included various factors specific to one SC actor (materials used in packaging, exact position of sensors, insulation etc.), whose modelling from the physical viewpoint would be very time-consuming. ANNs fared well at learning and approximating those latent relationships between multiple factors, thus sparing a daunting modelling effort. Although our business context also included the influence of various external factors, it focused on the prediction of temperature measured by a specific sensor, which made the application and adjustment of a physical law (NLC) possible. This leads us to the managerial and research implication that a temperature prediction problem can be better addressed with our method in the context of sensor temperature predictions (see Section 4.1 for the account of contexts in which the application of our method is feasible).

Obviously, our paper leaves some issues unaddressed and opens new areas for future research. The first area concerns the investigation of cases where the assumption of independence of temperature is relaxed for h and more sophisticated methods for the estimation of r can be proposed. This is especially applicable to scenarios involving sudden temperature increases or decreases such as those that are often observed for ambient sensors. The second area is largely related to extant research on sensor fusion and STD (see Section

2) but addresses the possibility of using these techniques to predict temperatures that will occur in the foreseeable future rather than for interpolation purposes. Last, context-awareness based on sensor data that describe other physical properties, i.e., light, shock, location, could complement our method to ensure more efficient predictions. In this regard, we shift the focus not to the accuracy of future temperature trajectory approximation under the condition of temperature stability but to the estimation of the possible duration and magnitude of deviations. Currently, our method is capable of adapting the input data to temperature instabilities, but it cannot forecast possible changes in temperature stability based on historical event data.

References

[1] G.H. Haan, J.V. Hillegersberg, E. de Jong, K. Sikkel, Adoption of Wireless Sensors in Supply Chains: A Process View Analysis of a Pharmaceutical Cold Chain, J. Theor. Appl. Electron. Commer. Res. 8 (2013) 138–154. doi:10.4067/S0718-18762013000200011.

[2] C.M. Hanson, A.M. George, A. Sawadogo, B. Schreiber, Is freezing in the vaccine cold chain an ongoing issue? A literature review, Vaccine. 35 (2017) 2127–2133. doi:10.1016/j.vaccine.2016.09.070.

[3] D.M. Matthias, J. Robertson, M.M. Garrison, S. Newland, C. Nelson, Freezing temperatures in the vaccine cold chain: A systematic literature review, Vaccine. 25 (2007) 3980–3986.

[4] European Commission, Directive 2001/83/EC of the European Parliament and of the Council on the Community Code Relating to Medicinal Products for Human Use, 2001.

[5] World Health Organization, Guide to good storage practices for pharmaceuticals, World Health Organization, 2003.

[6] FanJun Li, ZhaoJiong Chen, Brief analysis of application of RFID in pharmaceutical cold-chain temperature monitoring system, in: 2011 International Conference on Transportation and Mechanical & Electrical Engineering (TMEE), IEEE, 2011: pp. 2418–2420. doi:10.1109/TMEE.2011.6199709.

[7] A.A. Abdallah, Global Pharmaceutical Supply Chain: A Quality Perspective, International Journal of Business and Management. 8 (2013) 62. doi:10.5539/ijbm.v8n17p62.

[8] S. Serdarasan, M. Tanyas, Dealing with Complexity in the Supply Chain: The Effect of Supply Chain Management Initiatives, SSRN Electronic Journal. (2012). doi:10.2139/ssrn.2056331.

[9] G. Meroni, L. Baresi, M. Montali, P. Plebani, Multi-party business process compliance monitoring through IoT-enabled artifacts, Information Systems. 73 (2018) 61–78. doi:10.1016/j.is.2017.12.009.

[10] M. Potdar, A. Sharif, V. Potdar, E. Chang, Applications of Wireless Sensor Networks in Pharmaceutical Industry, in: 2009 International Conference on Advanced Information Networking and Applications Workshops (WAINA), IEEE, 2009: pp. 642–647. doi:10.1109/WAINA.2009.193.

[11] G.L. Geerts, D.E. O'Leary, A supply chain of things: The EAGLET ontology for highly visible supply chains, Decision Support Systems. 63 (2014) 3–22. doi:10.1016/j.dss.2013.09.007.

[12] C. Ammann, Stability Studies Needed to Define the Handling and Transport Conditions of Sensitive Pharmaceutical or Biotechnological Products, AAPS PharmSciTech. 12 (2011) 1264–1275.

[13] A. Musa, A. Gunasekaran, Y. Yusuf, Supply chain product visibility: Methods, systems and impacts, Expert Systems with Applications. 41 (2014) 176–194. doi:10.1016/j.eswa.2013.07.020.

[14] A. East, N. Smale, S. Kang, A method for quantitative risk assessment of temperature control in insulated boxes, International Journal of Refrigeration. 32 (2009) 1505–1513.

[15] K.-Y. Chen, Y.-C. Shaw, Applying back propagation network to cold chain temperature monitoring, Advanced Engineering Informatics. 25 (2011) 11–22. doi:10.1016/j.aei.2010.05.003.

[16] C.C. Emenike, N.P. Van Eyk, A.J. Hoffman, Improving Cold Chain Logistics through RFID temperature sensing and Predictive Modelling, in: 2016 IEEE 19th International Conference on Intelligent Transportation Systems (ITSC), IEEE, 2016: pp. 2331–2338. doi:10.1109/ITSC.2016.7795932.

[17] S. Mercier, I. Uysal, Neural network models for predicting perishable food temperatures along the supply chain, Biosystems Engineering. 171 (2018) 91–100. doi:10.1016/j.biosystemseng.2018.04.016.

[18] R. Jedermann, L. Ruiz-Garcia, W. Lang, Spatial temperature profiling by semi-passive RFID loggers for perishable food transportation, Computers and Electronics in Agriculture. 65 (2009) 145–154.

[19] W. Lang, R. Jedermann, D. Mrugala, A. Jabbari, B. Krieg-Brückner, K. Schill, The “Intelligent Container”—A Cognitive Sensor Network for Transport Management, IEEE Sensors J. 11 (2011) 688–698.

[20] J. Rodríguez-Bermejo, P. Barreiro, J.I. Robla, L. Ruiz-Garcia, Thermal study of a transport container, Journal of Food Engineering. 80 (2007) 517–527. doi:10.1016/j.jfoodeng.2006.06.010.

[21] R. Jedermann, W. Lang, Semi-passive RFID and beyond: steps towards automated quality tracing in the food chain, Ijrfita. 1 (2007) 247. doi:10.1504/IJRFITA.2007.015849.

[22] E. Abad, F. Palacio, M. Nuin, A.G. de Zárate, A. Juarros, J.M. Gómez, et al., RFID smart tag for traceability and cold chain monitoring of foods: Demonstration in an intercontinental fresh fish logistic chain, Journal of Food Engineering. 93 (2009) 394–399. doi:10.1016/j.jfoodeng.2009.02.004.

[23] D. Delen, R. Sharda, B. Hardgrave, The promise of RFID-based sensors in the perishables supply chain, IEEE Wireless Commun. 18 (2011) 82–88. doi:10.1109/MWC.2011.5751300.

[24] I. Uysal, J.P. Emond, G. Bennett, Tag testing methodology for RFID enabled temperature tracking and shelf life estimation, in: 2011 IEEE International Conference on RFID-Technologies and Applications (RFID-TA), IEEE, 2011: pp. 8–15. doi:10.1109/RFID-TA.2011.6068608.

[25] L. Ruiz-Garcia, P. Barreiro, J.I. Robla, L. Lunadei, Testing ZigBee Motes for Monitoring Refrigerated Vegetable Transportation under Real Conditions, Sensors. 10 (2010) 4968–4982.

[26] D.J.A. Bijwaard, W.A.P. van Kleunen, P.J.M. Havinga, L. Kleiboer, M.J.J. Bijl, Industry: Using Dynamic WSNs in Smart Logistics For Fruits and Pharmacy, in: SenSys '11, ACM Press, New York, USA, 2011: p. 218.

[27] V.F. Annese, D. De Venuto, On-line shelf-life prediction in perishable goods chain through the integration of WSN technology with a 1st order kinetic model, in: 2015 IEEE 15th International Conference on Environment and Electrical Engineering (EEEIC), IEEE, 2015: pp. 605–610.

[28] S. Nechifor, B. Tarnauca, L. Sasu, D. Puiu, A. Petrescu, J. Teutsch, et al., Autonomic monitoring approach based on CEP and ML for logistic of sensitive goods, in: 2014 18th International Conference on Intelligent Engineering Systems (INES), IEEE, 2014: pp. 67–72. doi:10.1109/INES.2014.6909343.

[29] S. Nechifor, A. Petrescu, D. Damian, D. Puiu, B. Tarnauca, Predictive analytics based on CEP for logistic of sensitive goods, in: 2014 International Conference on Optimization of Electrical and Electronic Equipment (OPTIM), IEEE, 2014: pp. 817–822. doi:10.1109/OPTIM.2014.6850965.

[30] S.L. Ting, Y.K. Tse, G.T.S. Ho, S.H. Chung, G. Pang, Mining logistics data to assure the quality in a sustainable food supply chain: A case in the red wine industry, International Journal of Production Economics. 152 (2014) 200–209. doi:10.1016/j.ijpe.2013.12.010.

[31] M. Thakur, E. Forås, EPCIS based online temperature monitoring and traceability in a cold meat chain, Computers and Electronics in Agriculture. 117 (2015) 22–30. doi:10.1016/j.compag.2015.07.006.

[32] A. Pal, K. Kant, Smart Sensing, Communication, and Control in Perishable Food Supply Chain, ACM Trans. Sen. Netw. 16 (2020) 1–41. doi:10.1145/3360726.

[33] M. Göransson, F. Nilsson, Å. Jevinger, Temperature performance and food shelf-life accuracy in cold food supply chains – Insights from multiple field studies, Food Control. 86 (2018) 332–341.

[34] R. Badia-Melis, U. Mc Carthy, I. Uysal, Data estimation methods for predicting temperatures of fruit in refrigerated containers, Biosystems Engineering. 151 (2016) 261–272.

[35] M.C. do Nascimento Nunes, M. Nicometo, J.P. Emond, R.B. Melis, I. Uysal, Improvement in fresh fruit and vegetable logistics quality: berry logistics field studies, Phil. Trans. R. Soc. A. 372 (2014) 20130307.

[36] S. Mercier, S. Villeneuve, M. Mondor, I. Uysal, Time-Temperature Management Along the Food Cold Chain: A Review of Recent Developments, Comprehensive Reviews in Food Science and Food Safety. 16 (2017) 647–667. doi:10.1111/1541-4337.12269.

[37] Xuan-Tien Doan, P.T. Kidd, R. Goodacre, B.D. Grieve, Regression analysis for supply chain logged data: A simulated case study on shelf life prediction, in: 2008 9th International Conference on Signal Processing (ICSP 2008), IEEE, 2008: pp. 2717–2720. doi:10.1109/ICOSP.2008.4697709.

[38] A.G.N. Novaes, O.F. Lima Jr, C.C. de Carvalho, E.T. Bez, Thermal Performance Of Refrigerated Vehicles in the Distribution of Perishable Food, Pesqui. Oper. 35 (2015) 251–284.

[39] J.H. Lienhard IV, J.H. Liehnhard V, A Heat Transfer Textbook, Fifth Edition, Phlogiston Press, 2019.

[40] L.C. Burmeister, Convective Heat Transfer, Second, Wiley-Interscience, 1993.

[41] T.L. Bergman, A.S. Lavine, F.P. Incropera, D.P. Dewitt, Fundamentals of Heat and Mass Transfer, Seventh Edition, John Wiley & Sons, 2011.

[42] M. Favre-Marinet, S. Tardu, Convective Heat Transfer, First Edition, Wiley-ISTE, 2009.

[43] M. Gockenbach, K. Schmidtke, Newton's law of heating and the heat equation, Involve. 2 (2009) 419–437. doi:10.2140/involve.2009.2.419.

[44] R.E. Kalman, A New Approach to Linear Filtering and Prediction Problems, Journal of Basic Engineering. 82 (1960) 35–45. doi:10.1115/1.3662552.

[45] P. Zarchan, H. Musoff, Fundamentals of Kalman Filtering A Practical Approach, Third Edition, American Institute of Aeronautics and Astronautics, Inc, 2009.

[46] W.S. McCulloch, W. Pitts, A Logical Calculus of the Ideas Immanent in Nervous Activity, Bulletin of Mathematical Biophysics. 5 (1943) 115–133.

[47] F. Rosenblatt, The Perceptron: A Perceiving and Recognizing Automaton, Cornell Aeronautical Laboratory, 1957.

[48] J. Schmidhuber, Deep learning in neural networks: An overview, Neural Networks. 61 (2015) 85–117.

[49] G.E.P. Box, G.M. Jenkins, Time series analysis: Forecasting and control, Holden-Day, San Francisco, 1970.

[50] A.D.R. McQuarrie, C.-L. Tsai, Regression and Time Series Model Selection, World Scientific, 1998.

[51] C.M. HURVICH, C.-L. Tsai, Regression and time series model selection in small samples, Biometrika. 76 (1989) 297–307. doi:10.1093/biomet/76.2.297.

[52] T.C. Mills, R.N. Markellos, The Econometric Modelling of Financial Time Series, Third Edition, Cambridge University Press, 2008.

[53] D.C. Montgomery, C.L. Jennings, M. Kulahci, Introduction to Time Series Analysis and Forecasting, Second, Wiley, 2015.

[54] P.J. Brockwell, R.A. Davis, Introduction to Time Series and Forecasting, Third, Springer, 2016.

[55] A. Konar, D. Bhattacharya, Time-Series Prediction and Applications, Springer, 2017.

[56] R.H. Byrd, R.B. Schnabel, G.A. Shultz, Approximate solution of the trust region problem by minimization over two-dimensional subspaces, Mathematical Programming. 40-40 (1988) 247–263.

[57] M.A. Branch, T.F. Coleman, Y. Li, A Subspace, Interior, and Conjugate Gradient Method for Large-Scale Bound-Constrained Minimization Problems, SIAM J. Sci. Comput. 21 (1999) 1–23.

[58] J. Bergstra, Y. Bengio, Random Search for Hyper-Parameter Optimization, The Journal of Machine Learning Research. 13 (2012) 281–305.

[59] J. Heaton, Introduction to Neural Networks with Java, Second, Heaton Research, Inc, 2008.