ZigBee-based wireless sensor networks for monitoring animal presence and pasture time in a strip of new grass

Nadimi, Esmaeil; Søgaaard, Henning T.; Bak, Thomas; Oudshoorn, Frank W.

Published in:
Computers and Electronics in Agriculture

DOI (link to publication from Publisher):
10.1016/j.compag.2007.09.010

Publication date:
2008

Document Version
Early version, also known as pre-print

available at www.sciencedirect.com

ScienceDirect

journal homepage: www.elsevier.com/locate/compag

ZigBee-based wireless sensor networks for monitoring animal presence and pasture time in a strip of new grass

E.S. Nadimi[a,b,*], *H.T. Søgaard*[a], *T. Bak*[b], *F.W. Oudshoorn*[a]

[a] *Research Center Bygholm, Department of Agricultural Engineering, Aarhus University, Schuttesvej 17, DK-8700 Horsens, Denmark*
[b] *Department of Electronic Systems, Automation and Control, Aalborg University, Denmark*

ARTICLE INFO

Article history:
Received 4 May 2007
Received in revised form
13 August 2007
Accepted 11 September 2007

Keywords:
Wireless sensor networks
ZigBee
Packet delivery performance
Received signal strength
Pearson chi-square test
Animal presence monitoring
Pasture time

ABSTRACT

The problem of online monitoring of cows' presence and pasture time in an extended area covered by a strip of new grass using wireless sensor networks has been addressed. The total pasture time in the extended area was estimated by measuring the pasture time in a specific part of that area called the gateway connectivity area where sensor nodes mounted on the cows could communicate directly with a gateway. Packet loss causes a node that was present in the connectivity range of the gateway frequently to be classified as an absent node. Therefore, a moving average window with optimal window length and threshold was designed to minimize the misclassification. As the measured pasture time in the gateway connectivity area was an underestimation of the total pasture time in the extended area, an area-based correction factor, same for all individual animals was applied.

As only 23% of the animals in a herd were equipped to be monitored by sensor nodes, investigations to evaluate if the monitored number of animals could represent the whole herd were of great importance. To accomplish the investigations, the number of monitored cows by sensor nodes and the total number of cows (with and without sensor nodes) in the extended area were counted manually each minute over a period of 3 h during 3 days. Pearson chi-square test of goodness of fit showed that the number of cows in the extended area was normally distributed. Furthermore, a statistical test showed that the mean number of monitored cows in the extended area and the mean of total number of cows in the extended area corresponded with the percentage of monitored cows by sensor nodes in the herd (23%).

1. Introduction

Public perception, animal welfare, and milk quality call for a continued use of grasslands for grazing in dairy farming (Torjusen et al., 2001). To meet the public concern some milk producers offer incentives to dairy farmers if they let their dairy cows graze, but for many farmers this is impossible due to livestock management and control problems. Management and control relies on monitoring of the herd, which is significantly complicated by the inherent distribution of the animals as well as the outdoor location. Successful grazing in developed agriculture calls for automated and efficient monitoring and control of the animals. The monitoring should allow us to establish a better understanding of animal behav-

* *Corresponding author at*: Research Center Bygholm, Department of Agricultural Engineering, Aarhus University, Schuttesvej 17, DK-8700 Horsens, Denmark. Tel.: +45 89993014; fax: +45 89993100.
E-mail address: esmaeil.nadimi@agrsci.dk (E.S. Nadimi).
0168-1699/$ – see front matter
doi:10.1016/j.compag.2007.09.010

ior, detect individuals with potential health problems, and generally optimize the grazing process, all things that potentially would have a significant impact on practical farming.

The general behavior of a herd of animals is well known by farmers but not so well documented. Different aspects of animal behavior have been studied by different researchers. The position of animals in the field were tracked and monitored by White et al. (2001), Schwager et al. (2007) and Butler et al. (2004) while Oudshoorn et al. (2008) made investigation based on the positions and the velocities of the movements in the field. Different behavior phases of dairy cattle were evaluated by Munksgaard et al. (2005), Wilson et al. (2005), Nadimi et al. (2007) and Bishop-Hurley et al. (2007). None of these references, however, addressed an online monitoring system that registers the time that animals spent in specific areas of the field. Such information would be useful in strip crop grazing systems, where the animals are offered a controllable section of, e.g. new grass at regular intervals (Oudshoorn and Nadimi, 2007). The total number of animals roaming in a particular area of the field and their total pasture time in that area can be an indicator of the grass quality, and quantity and may help determine the right time to provide access to a new strip. From a strip crop grazing point of view, the question is if we can set up an automatic monitoring system that can identify animals present in the new strip, determine how long time they spend there and based on that say something about the need for a new strip of grass. In addition it is interesting to investigate if the whole herd has to be monitored or if a subset of the herd can be used as an indicator of the need for new feed. Monitoring only a subset of a herd might be more economical and practical.

The most popular system for outdoor localization is based on the Global Positioning System (GPS) (Butler et al., 2004; Oudshoorn et al., 2008) but energy consumption and cost makes it difficult to apply in practical farming. In addition satellite connection loss has been reported frequently in the research done by Oudshoorn et al. (2008). A simpler alternative is based on radio frequency identification (RFID) tags (Ng et al., 2005). Locating RFID readers strategically in the field allow animals entering a specific area to be registered. The main drawbacks of RFID technology are the relatively short communication range (1–2 m) and the fact that the devices are passive limiting future extensions such as temperature and motion monitoring. Monitoring relatively large extended areas (1800–3000 m^2) using RFID tags also demand a significant infrastructure. A more natural candidate for an online monitoring framework is based on wireless sensor network technology. By providing each animal with a sensor node, which incorporates computation, sensing, and wireless networking capabilities allows relevant health parameters and location to be collected at regular intervals on each individual. Information can flow across the group as in a modern communications network, using low-power radios with well-designed protocol stacks thereby extending the communication range of system significantly at no extra cost. This permits data to be aggregated across the network and forwarded to control and management systems. Local computational capabilities on the individual sensor node allow complex filtering and triggering functions, and application or sensor-specific data compression algorithms. The application of sensor networks for animal monitoring was addressed by Szewczyk et al. (2004), Wang et al. (2006), Bishop-Hurley et al. (2007), Nadimi et al. (2007) and Schwager et al. (2007).

The objectives of this research were to demonstrate registration of pasture time in a specific area (a strip with new grass) using a ZigBee (Szewczyk et al., 2004)-based wireless sensor network and single hop connectivity. Another objective was to prove two extensions: an area extension where knowledge about animal presence in a limited area is used to predict animal presence in a larger extended area. The other extension aims at determining the whole herd presence based on registration of a subset of tagged animals. Yet another objective was to solve a specific problem regarding packet loss using data post processing.

Each node in the network was programmed to transmit data when located within communication range of a gateway in the area with new grass as illustrated in Fig. 1.

The principle is single hop connectivity that is the gateway only registers presence when a specific node is within the communication range and actively participates in handshaking communication (Lewis, 2004). In this research, multi hop connectivity as used in modern communication networks was not utilized.

As the area defined by the communication range does not necessarily cover the same area as the new grass strip, an area-based correction factor was applied to the measured time in the gateway connectivity area to estimate the total pasture time in the new grass strip.

Most researches (e.g. White et al., 2001; Butler et al., 2004; Munksgaard et al., 2005; Nadimi et al., 2007; Oudshoorn et al., 2008) only monitored a portion of a herd of animals but the monitored behavior was generalized to the whole herd without any reliable proof. However, it is of great importance to demonstrate that the monitored subset of a herd can represent the whole herd. In the present paper, a statistical test is suggested to determine if the number of monitored animals in the new grass strip could represent the whole herd.

The remainder of this paper is organized as follows. Section 2 describes materials and methods that have been used to monitor the pasture time and animal presence in an extended area. Section 3 describes the experimental setup and results and finally, the conclusions are presented.

2. Materials and methods

2.1. *Materials*

MPR2400 Micaz sensor motes from Crossbow were used for the experiments in this paper. They have a Chipcon CC2420 radio, which uses 2.4 GHz IEEE 802.15.4/ZigBee RF transceiver with MAC support. TinyOS was running on the motes. In order to register the absolute time when the nodes were within range of communication with the gateway, the gateway was programmed to register the arrival time of the packets disseminated by the nodes as a time stamp in the received packet. When a node desired to transmit a message, handshaking protocols with the destination node were used to improve reliability. The destination and gateway transmitted alternately as follows: request to send, ready to receive, send message,

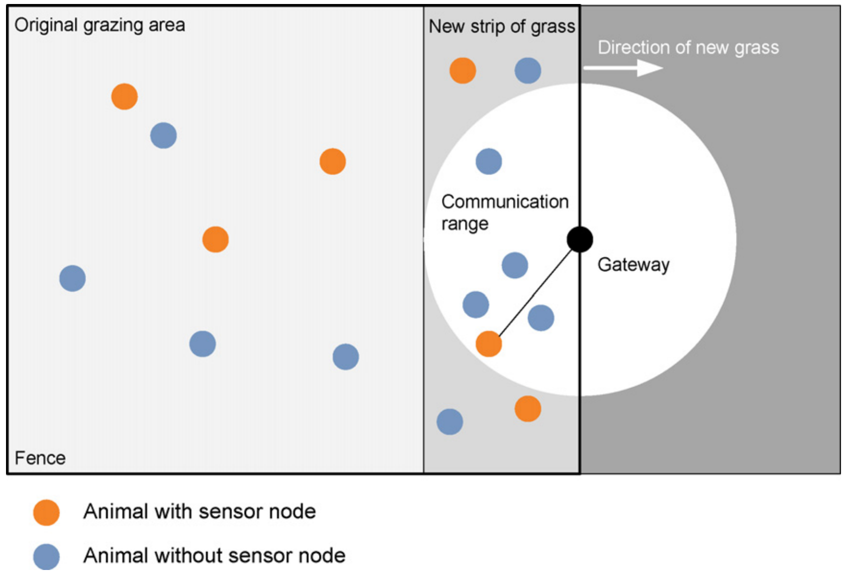

Fig. 1 – Basic strategy for strip crop grazing.

message received. The sampling rate for the packet dissemination was chosen as 1 Hz (Nadimi et al., 2006).

The CC2420 radio supports up to 255 different transmission power levels and allows for a programmable transmission frequency. In order to minimize the number of variables in the experiment, the RF transmission frequency was fixed at a single frequency band (2.48 GHz) while the transmission power (1 mW) was selected to ensure that the nodes were able to communicate with the gateway only in a certain area, i.e. a part of the new grass strip (gateway connectivity area).

2.2. *Methods*

Outdoor wireless communication channels as used in this work are inherently unreliable and the effect of packet loss cannot be neglected. Here, the basic idea is to use arrival of packets as the only indicator for classifying nodes as being within or outside the gateway communication range. Packets disseminated by each sensor node contained the identification number (node ID) of the node. The packet arrival time was registered by the gateway and indicated the presence of the node within the communication range of the gateway at that time instant. In order to minimize misclassification due to packet loss in the presence of obstacles, a moving average window was applied to packet arrival sequence. An optimization problem was set up to find the optimal window length and the optimal threshold for classification.

2.3. *Estimation of window length and threshold value*

2.3.1. *Packet delivery performance*

Once the nodes were deployed, each of them followed a sequence of instructions to gather information about its surroundings and to transmit data packets toward the gateway. Intermittent communication due to poor connectivity with the transceiver, presence of obstacles as an interferer and general unreliability in the communication channels caused loss of packets. As an example, packet delivery performance for one of the nodes in the communication range of the gateway is represented in Fig. 2 in which 1 is an indicator of packet arrival and 0 indicates packet loss. The packet loss in this example was 312 out of 1000 packets or 31.2%.

As it can be seen from Fig. 2, it would lead to a high misclassification rate if packet loss was taken as an indication of the cow being outside the communication range of the gateway. Therefore a moving average window and a threshold operation were employed to minimize misclassification, i.e. if the average of the packet delivery values in a window around a given time instant was larger than a given threshold, the cow was classified as being within the communication range of the gateway at that time instant.

To calculate the optimal window length and threshold value, another experiment was carried out in which the

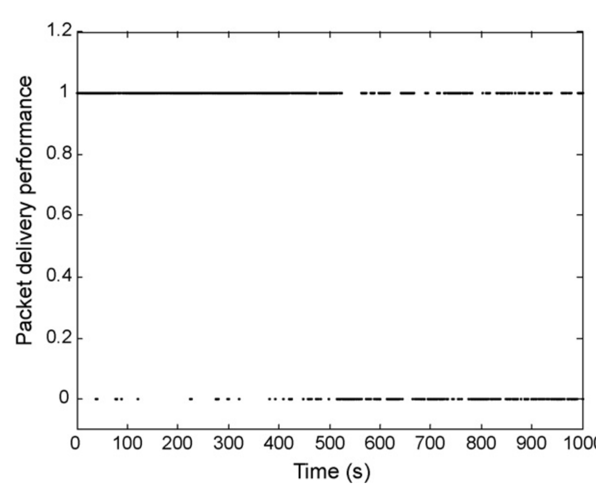

Fig. 2 – Example of packet delivery performance in the network when the sensor node is within the communication range of the gateway. 1 indicates packet arrival and 0 indicates packet loss.

gateway was programmed to measure the received signal strength (RSS) of the incoming packets to estimate the distance between sensor nodes and the gateway. Assuming that the RSS distance estimate is used as ground truth and based on packet delivery performance, an optimization problem for estimation of the optimal window length and the threshold could be formulated. The following sections present the details of the approach applied.

2.3.2. *RSS measurement analysis*

In order to convert the received signal strength to an accurate estimate of the distance between the gateway and the node, extensive preliminary field measurements and calibrations were carried out. Fig. 3 shows the graph of signal strength versus distance for one of the nodes during calibrations. The received power level can be converted to a distance estimate by using a radio wave propagation model fitted to the experimental data (Kotanen et al., 2003; Nadimi et al., 2007):

$$20 \log d = P_{\mathrm{Tx}} - P_{\mathrm{Rx}} + C \tag{1}$$

where P_{Tx} [dB m] and P_{Rx} [dB m] are the transmitted and received power levels, respectively. d [m] is the distance between transmitter and receiver. In this model, constant C represents the antenna gain and wavelength effects and was estimated by minimizing the sum of squared differences between the experimental RSS and the modeled RSS. As all the nodes have different characteristics such as different antenna gains or different radios, the graph of received signal strength versus distance (Fig. 3) is not the same for all the nodes. Therefore, the optimal constant C in Eq. (1) is different from one node to another one (the range varied between −60 dB m to −55 dB m). In the present research, the constant C calculated for one of the nodes (−56 dBm) was selected as the optimal constant representing antenna gain and wavelength effect for all the nodes. This strategy tends to diminish precision of the results of each individual node (curve fit and estimated distance between nodes and gateway) and consequently the whole system. However, that is the practical solution to implement a similar monitoring system to a large herd of animals with a large number of nodes as it will be time and energy consuming process to estimate the optimal constant C for all the nodes.

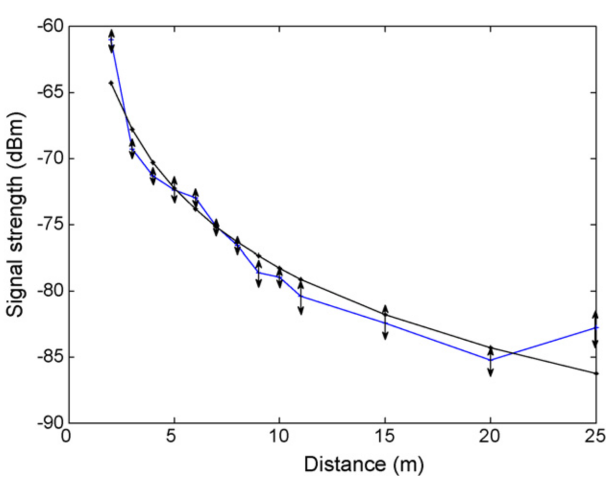

Fig. 3 – RSS vs. distance for fitted optimal propagation model and experimental data. Black curve: propagation model; blue curve: experimental data. Arrows as indicator of error bar (standard deviation) at each point. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of the article.)

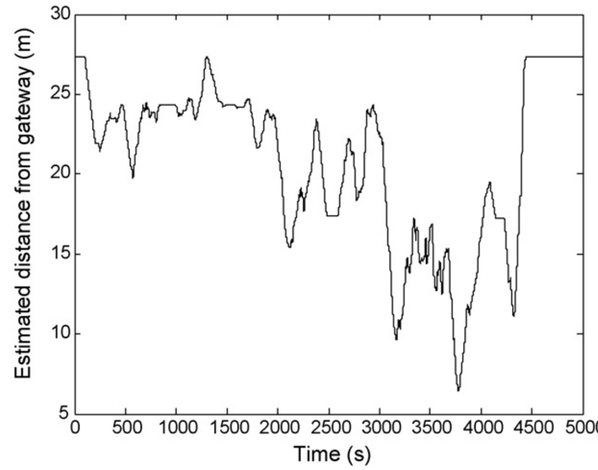

Fig. 4 – Example of estimated distance between sensor node and the gateway in field experiment with cows.

To estimate the distance between sensor nodes and the gateway in case of missing RSS data due to packet loss, a simple Kalman filter with intermittent observation was implemented to the RSS data (Sinopoli et al., 2004). Modeling the data as a discrete time Wiener process, the Kalman filter was designed to estimate the states not observed due to packet loss. Estimated distances between a sensor node and the gateway during an experiment with cows in the field are presented in Fig. 4.

2.3.3. *Optimal window length and threshold*

The packet arrival sequence $\gamma(t)$ ($t = 1, 2, \ldots$ is time (s); $\gamma(t) = 1$ if a packet arrived at time t, and $\gamma(t) = 0$ otherwise) was filtered by use of a moving average window of length WL (odd integer) to obtain a smoothed sequence $\gamma'(t)$:

$$\gamma'(t) = \frac{1}{\mathrm{WL}} \sum_{i=1}^{\mathrm{WL}} \gamma\left(t - i + \frac{\mathrm{WL}+1}{2}\right) \tag{2}$$

To classify a node as being inside or outside the gateway communication range at time t, a threshold T was introduced, i.e. the node was classified as inside if $\gamma'(t) \geq T$ and outside if $\gamma'(t) < T$. In order to find the window length, WL and threshold, T that minimized the likelihood that a node was wrongly classified (i.e. classified as being within the connectivity range r_0 when it was not and vice versa) a minimization criterion was defined:

$$J = \min_{\mathrm{WL},T} \sum e^2(t) \tag{3}$$

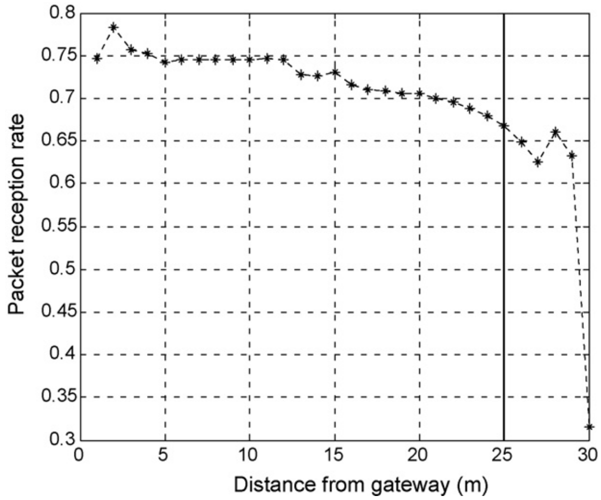

Fig. 5 – Example of packet reception rate vs. distance from gateway for a node (no obstacles between node and gateway).

where the classification error $e(t)$ was defined as:

$$e(t) = \begin{cases} \gamma'(t) - T & \text{if } \gamma'(t) \geq T \text{ and } \hat{r}(t) > r_0 \text{ or } \gamma'(t) < T \text{ and } \hat{r}(t) \leq r_0 \\ 0 & \text{otherwise} \end{cases} \quad (4)$$

In Eq. (4), $\hat{r}(t)$ is the RSS-based estimate of the distance between a node and the gateway and r_0 (gateway connectivity range) was set to 25 m based on packet reception rate (example in Fig. 5). The selected value of r_0 reflects a compromise: on one hand a large value of r_0 was desired to make the covered circle as large as possible; on the other hand r_0 should not be large to avoid unreliable classification because of low reception rate at high distances from the gateway (Fig. 5). Among all the facts which can intensify packet loss rate such as environmental condition, distance, relative height between transceiver and receiver, transmission power, data rate, packet size and the routing protocol; environmental effect and the relative height are mainly the factors that caused high rate packet loss. While the experiment and therefore the calibration process were accomplished in rainy days in the field (outdoor), low reception rate is expected due to high humidity rate where the radio waves can be more easily absorbed by the water and the wet grass. As the curve of packet reception rate versus relative height between sensor nodes and the gateway ascends until a certain relative height and then reaches the steady state, relatively short distance between nodes and the ground and the gateway (40 cm) could be a reason for high packet loss rate.

2.4. *Pasture time estimation*

In order to monitor the pasture time in the strip of new grass, the field was extended by moving a section of the fence. As it is shown in Fig. 6, the rectangular extended area, Δ (which is the strip of new grass) was not entirely covered by the gateway connectivity area which implies that time spent in that area was not the same as actual pasture time in Δ. Assuming that the connectivity between the sensor nodes and the gateway was uniform in all directions, the area of connectivity would be a half circle area with radius r_0. By registering packet arrival time (time stamp) at the gateway followed by moving average filtering and threshold classification, the pasture time in the gateway connectivity area could be monitored.

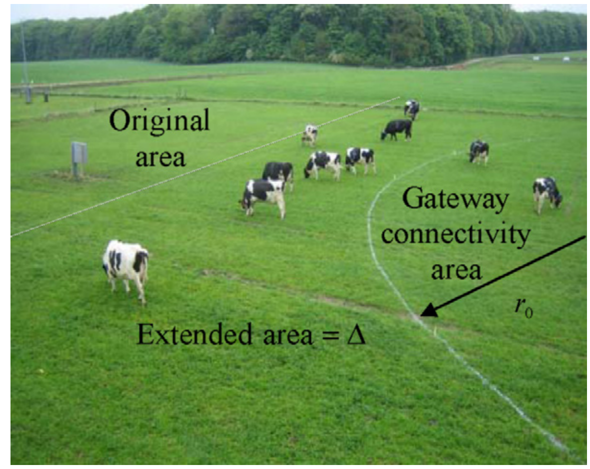

Fig. 6 – Extended strips of new grass and the connectivity area.

Assuming a uniform distribution over Δ, the pasture time T_E in the extended area as a function of pasture time T_C in the gateway connectivity area is given by:

$$T_E = KT_C \quad (5)$$

where constant K is the ratio of the extended area (Δ) to the gateway connectivity area ($\pi r_0^2/2$).

Since the constant K depended only on fixed geometrical quantities, it was the same constant for all individuals in the herd.

2.5. *Testing the hypothesis that monitored cows could represent the entire herd*

To estimate pasture time for the entire herd of cows based on data from the monitored subset of the herd (23%), it was assumed that the monitored cows could represent the whole herd. The validity of this assumption was examined by statistical tests that involved two random variables: D_1, the number of monitored cows in the extended area, and D_2, the total number of cows in that area. Both of these variables were sampled each minute over 3 h by manual observation. First, it was tested if the distributions of D_1 and D_2 could be approximated by normal distributions. Then it was tested if the ratio between the means of D_2 and D_1 was equal to the ratio k between the total number of cows and the number of monitored cows ($k = 30/7 = 4.3$). If the result of this test was positive, it would indicate that the monitored cows could represent all cows in the herd (with respect to pasture time in the extended area).

2.5.1. *Test of the normal approximation*

Pearson's chi-square test (χ^2) is one of a variety of chi-square statistical procedures whose results are evaluated by reference to the chi-square distribution (Chernoff and Lehmann, 1954). Pearson's chi-square is used to assess tests of goodness of fit which establishes whether or not an observed distribution differs from a theoretical distribution. Pearson chi-square tests were applied to the samples of D_1 and D_2 to evaluate whether the variables were normally distributed.

2.5.2. *Testing if the monitored cows was a representative sample of the herd*

In order to evaluate if the monitored subset of the herd could represent the whole herd, the following null hypothesis H_0 and alternative hypothesis H_1 were set up:

$$\begin{aligned} H_0 &: \mu_2 = k\mu_1 \\ H_1 &: \mu_2 \neq k\mu_1 \end{aligned} \quad \text{s.t.} \quad \sigma_1^2 \neq \sigma_2^2 \tag{6}$$

where μ_1 and μ_2 are the theoretical and unknown mean values of D_1 and D_2, respectively. σ_1 and σ_2 are the theoretical and unknown standard deviations of D_1 and D_2 and k is a constant representing the ratio of the number of monitored cows to the total number of cows. In order to define significance level (α), the probability function has been introduced:

$$\alpha = p(\text{reject } H_0 | H_0 \text{ is true}) \tag{7}$$

To reject the null hypothesis, modifications to the standard test were required to incorporate the ratio k. A modified version of the two-sample t-test was applied. The result is a criterion for rejection, $|t_0| > t_{1-(\alpha/2),\upsilon}$ where $t_{1-(\alpha/2),\upsilon}$ is the $1-(\alpha/2)$ quantile in the Student's t-distribution with υ degrees of freedom. The t-statistics t_0 and the degrees of freedom υ are defined as:

$$t_0 = \frac{\bar{y}_2 - k\bar{y}_1}{\sqrt{(s_2^2/n_2) + (k^2 s_1^2/n_1)}} \tag{8}$$

$$\upsilon = \frac{((s_2^2/n_2) + (k^2 s_1^2/n_1))^2}{((s_2^2/n_2)^2/(n_2-1)) + ((k^2 s_1^2/n_1)^2/(n_1-1))} \tag{9}$$

In Eqs. (8) and (9), $\bar{y}_1$ and $\bar{y}_2$ are the sample means of D_1 and D_2, respectively, while s_1 and s_2 are the sample standard deviations of D_1 and D_2. The sample sizes for D_1 and D_2 are n_1 and n_2, respectively.

3. Experimental setup and results

3.1. *Experimental setup*

The case study in the presented experiment was a group of dairy cows. The experiment was carried out during 6 days with 30 cows 6 h per day on average. Data from 3 days of the experiment were used for estimation of the optimal window length and threshold of the filter used for the packet arrival sequence. The packet arrival sequence from the remaining 3 days were filtered and applied for estimation of pasture time in the strip of new grass. During the calibration process, the nodes were placed at fixed distances (1–30 m far from the gateway) for 5 min at each distance. The sampling time was set to 1 s and it was expected to receive 300 samples per distance while the real number of packets received at each distance is presented by packet reception rate in Fig. 5. The experimental data in Fig. 3 represents the mean value of the readings taken at each distance.

Seven out of 30 cows were equipped with wireless nodes around the neck. The node on the collar as well as collar itself was fixed very well to prohibit any slide to right or left side. The antenna pointed the sky in order to have better communication between nodes and the gateway. The antenna was ½ wave dipole antenna, with an MMCX connector. The gateway was installed 1.2 m above the ground in a location as indicated in Fig. 1.

The shape of the extended area was rectangular and the area was 60 m by 40 m while the shape of the gateway connectivity area was a half circle with a radius of 25 m. Each day, a new extended area covered by new grass was provided for the cows. Manual registrations of absolute time of day when each of the 30 cows was in the extended area and the connectivity area were carried out 3 h per day during 3 days. Furthermore, the number of cows roaming in the extended area was registered manually with a sampling interval of 1 min during different grazing periods (e.g. first grazing period starts when the animal enters the field and second grazing period starts after first lying down period).

3.2. *Results and discussion*

3.2.1. *Pasture time monitoring*

Minimizing the cost function in Eq. (3) resulted in an optimal window length and threshold of 155 s and 0.388, respectively. The result of applying the moving average window with the optimal window length and the threshold to the packet delivery performance (Fig. 2) is presented in Fig. 7. While the delivery rate in the packet delivery performance was 68.8% (31.2% packet loss), applying the optimal window improved the results of packet delivery to 92%. Moving average filtering and subsequent classification of presence inside or outside the connectivity area were compared to manual registrations and resulted in errors as exemplified in Figs. 8 and 9 for two different nodes. Values of 0 and 1 indicate correct and incorrect classification, respectively.

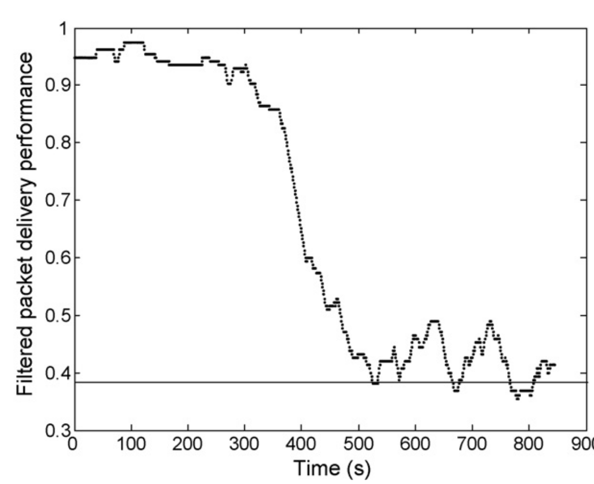

Fig. 7 – Filtered packet delivery performance by the optimal moving average window. The threshold (0.388) is presented by the horizontal line.

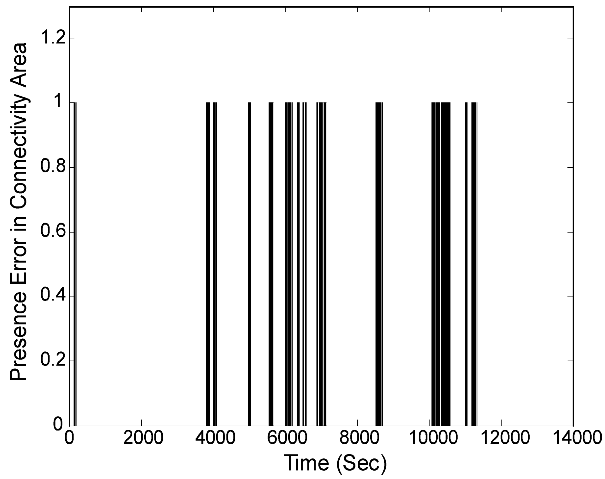

Fig. 8 – Classification error for one of the nodes, example 1.

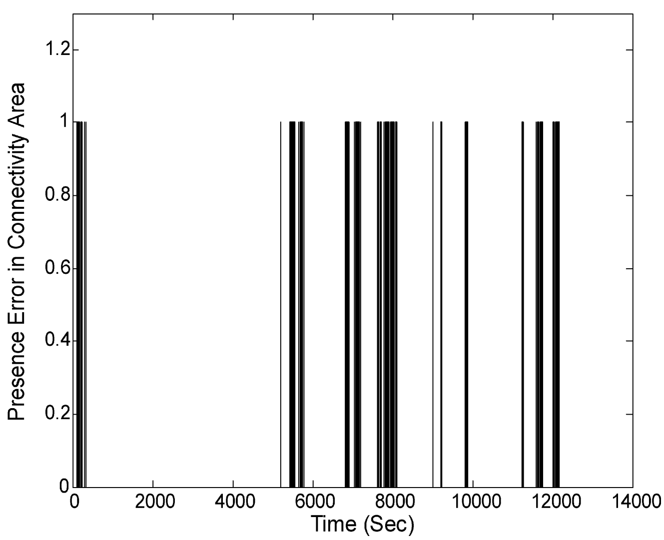

Fig. 9 – Classification error for one of the nodes, example 2.

The percentages of successful classification for the entire experiment are presented in Table 1.

In order to estimate the total pasture time length for each of the monitored cows in the extended area using estimated total pasture time length in the gateway connectivity area,

Table 1 – Classification success rate

Sensor node	Success rate (%)			Average per node (%)
	Day 1	Day 2	Day 3	
Node #1	84	82	78	81.3
Node #2	78	75	83	78.6
Node #3	88	76	90	84.6
Node #4	72	74	75	73.6
Node #5	75	78	75	76
Node #6	83	88	84	85
Node #7	68	72	71	70.3
Average per day	78.3	77.8	79.4	78.5[a]

[a] Overall average.

Table 2 – Estimated pasture time as percentage of true pasture time for all the monitored cows during the experiment

Sensor node	Success rate (%)			Average per node (%)
	Day 1	Day 2	Day 3	
Node #1	89	83	74	82
Node #2	92	95	86	91
Node #3	91	66	89	82
Node #4	67	70	72	69.7
Node #5	94	81	85	86.7
Node #6	80	77	70	75.7
Node #7	77	68	73	72.7
Average per day	84.3	77.1	78.4	79.9[a]

[a] Overall average.

Eq. (5) was applied with $K = 2.44$ as the extended area was $2400\,\mathrm{m}^2$ and gateway connectivity area was $981.74\,\mathrm{m}^2$. Statistical analysis of the cows' GPS positions confirmed that they were uniformly distributed over the extended area.

Table 2 shows the estimated pasture time as percentage of true pasture time for all the monitored cows during the experiment. As it can be concluded from Table 2, the estimated total pasture time using Eq. (5) was always an underestimation of the real total pasture time. One reason for this could be that the packet loss rate was generally higher in this experiment than it was in the experiment used for finding the optimal window length and classification threshold. Apart from the proposed moving average window, a potential solution for the problem of packet loss could be as introduced by Guo et al. (2006) in which an onboard flash memory was used in their designed nodes to store considerable amounts of data. Stored packets would then be sent to the gateway as soon as nodes could communicate to the gateway. However, this solution to the problem of packet loss would require extra hardware facilities and causes delay in classifying the presence or absence in the communication range of the gateway, this relatively short delay will not have a critical influence on the performance of the monitoring system.

3.2.2. *Normality test for number of cows in the extended area*

In order to evaluate whether the total number of cows (D_2) in the extended area was normally distributed using Pearson chi-square test, contingency tables were constructed for data from all days. As an example, Table 3 presents the contingency tables for two different datasets containing total number of cows ($D_2^{(1)}$ and $D_2^{(2)}$) in two different grazing periods from the first day. It was concluded from the Pearson chi-square test of goodness of fit with two degrees of freedom that D_2 was normally distributed (hypothesis accepted at a significance level of 0.2).

As presented in Table 3, the estimated mean value of $D_2^{(2)}$ (total number of cows in the extended area during second grazing period) is smaller while the estimated variance is larger compared to the estimated mean value and variance of $D_2^{(1)}$ (total number of cows in the extended area during first grazing period) with grass offer reduction.

Table 3 – Contingency table for evaluating Pearson chi-square test of goodness of fit to normal distributions for two different datasets containing total numbers of cows in the extended area

Data set #1 ($D_2^{(1)}$) intervals	Observed counts (O)	Expected counts (E)	$\frac{(O-E)^2}{E}$	Data set #2 ($D_2^{(2)}$) intervals	Observed counts (O)	Expected counts (E)	$\frac{(O-E)^2}{E}$
$D_2^{(1)} < 7.5$	6	5.21	0.12	$D_2^{(2)} < 4.5$	6	4.53	0.47
$7.5 < D_2^{(1)} < 10.5$	8	8.58	0.03	$4.5 < D_2^{(2)} < 7.5$	7	7.14	0.00
$10.5 < D_2^{(1)} < 13.5$	9	11.31	0.47	$7.5 < D_2^{(2)} < 9.5$	5	6.56	0.37
$13.5 < D_2^{(1)} < 16.5$	9	9.02	0	$9.5 < D_2^{(2)} < 11.5$	7	6.80	0.00
$D_2^{(1)} > 16.5$	8	5.85	0.78	$D_2^{(2)} > 11.5$	14	13.94	0
Sum	$n_2^{(1)} = 40$	$n_2^{(1)} = 40$	1.4	Sum	$n_2^{(2)} = 39$	$n_2^{(2)} = 39$	0.84
Chi-square test = 0.25				Chi-square test = 0.33			
$\bar{y}_2^{(1)} = 12.15, s_2^{(1)} = 4.13$				$\bar{y}_2^{(2)} = 9.86, s_2^{(2)} = 4.49$			

The 80% quantile in the chi-square distribution with 2 degrees of freedom is 3.21 so the test statistic of 1.4 and 0.84 is less and therefore the hypothesis of a normal distribution can be accepted.

Table 4 – Results of testing if the monitored cows could represent the entire herd

Statistics	Dataset			
	Grazing period 1		Grazing period 2	
	Total number of cows, $D_2^{(1)}$	Number of cows carrying sensors, $D_1^{(1)}$	Total number of cows, $D_2^{(2)}$	Number of cows carrying sensors, $D_1^{(2)}$
Average	$\bar{y}_2^{(1)} = 12.15$	$\bar{y}_1^{(1)} = 2.6$	$\bar{y}_2^{(2)} = 9.86$	$\bar{y}_1^{(2)} = 2.4$
S.D.	$s_2^{(1)} = 4.13$	$s_1^{(1)} = 1.31$	$s_2^{(2)} = 4.49$	$s_1^{(2)} = 1.28$
Observations	$n_2^{(1)} = 40$	$n_1^{(1)} = 40$	$n_2^{(2)} = 39$	$n_1^{(2)} = 39$
υ	71		73	
k	4.3 (=30/7)		4.3 (=30/7)	
t_0	1.24		−0.36	
$t_{1-(\alpha/2),\upsilon}$	1.289		1.287	
$\|t_0\| < t_{1-(\alpha/2),\upsilon}$	True		True	

The hypothesis was tested at significance level $\alpha = 0.2$.

Pearson chi-square test of goodness of fit was also applied to D_1 and the results demonstrated that the distribution of D_1 during first and second grazing period ($D_1^{(1)}$ and $D_1^{(2)}$) was Gaussian (hypothesis accepted at a significance level of 0.2) and the estimated mean value of $D_1^{(1)}$ and $D_1^{(2)}$ was 2.6 and 2.4, respectively.

3.2.3. *Testing the hypothesis that monitored cows could represent the entire herd*

The results of testing the null hypothesis in Eq. (6) are shown in Table 4. The significance level is chosen equal to 0.2 (Montgomery, 1996). Based on the results of the last row in Table 4, it is concluded that the introduced null hypothesis cannot be rejected.

4. Conclusion

The problem of online monitoring of cows' presence and pasture time in an extended area in the field with new grass has been addressed and solved by using wireless sensor networks. The total pasture time in the extended area was estimated by measuring the pasture time in the gateway connectivity area where the sensor nodes could communicate directly to the gateway. However, as the measured time in the connectivity area underestimated the total pasture time in the extended area, an area-based correction factor was applied and the results showed 79.9% success rate (21.1% error) on average.

As only 23% of the animals were equipped to be monitored by sensor nodes, investigations to evaluate whether the monitored animals could represent the whole herd were carried out. Pearson chi-square test of goodness of fit has been successfully applied to the datasets containing the number of cows roaming in the extended area and the number of cows carrying sensor nodes in the extended area. The results of statistical analysis indicated that the datasets were normally distributed. The hypothesis was that cows with and without sensor nodes would spend the same relative amount of time in the extended area. This hypothesis was confirmed by a modified two-sample t-test for the expected relation between the mean number of cows with sensor nodes in the extended area and the mean number of cows totally in the extended area.

Applying a moving average window with optimal window length and optimal threshold could successfully compensate for packet loss between sensor nodes and gateway and thereby improve the result of classification as being within or outside communication range of the gateway.

References

Bishop-Hurley, G.J., Swain, D.L., Anderson, D.M., Sikka, P., Crossman, C., Corke, P., 2007. Virtual fencing applications:

implementing and testing an automated cattle control system. J. Comput. Electron. Agric. 56 (1), 14–22.
Butler, Z., Corke, P., Peterson, R., Rus, D., 2004. Networked cows: virtual fences for controlling cows. In: Proceedings of ICRA. IEEE Conference on Robotics and Automation, vol. 5, New Orleans, LA, pp. 4429–4436.
Chernoff, H., Lehmann, E.L., 1954. The use of maximum likelihood estimation in chi-square test of goodness of fit. Ann. Math. Stat..
Guo, Y., Corke, P., Poulton, G., Wark, T., Bishop-Hurley, G., Swain, D., 2006. Animal behavior understanding using wireless sensor networks. In: Proceedings of the 1st IEEE International Workshop on Practical Issues in Building Sensor Network Applications (SenseApp 2006), IEEE, Tampa, Fl, pp. 608–614.
Kotanen, A., Hannikainen, M., Leppakoski, H., Hamalainen, T.D., 2003. Experiment on local positioning with bluetooth. In: Information Technology: Coding and Computing [Computers and Communications], 2003. Proceedings of ITCC. International Conference, pp. 297–303.
Lewis, F.L., 2004. In: Cook, D.J., Das, S.K. (Eds.), Wireless Sensor Networks. Smart Environments: Technologies, Protocols and Applications. John Wiley, New York, ISBN 978-0-471-54448-7.
Montgomery, D.C., 1996. Design and Analysis of Experiments, 4th ed. John Wiley & Sons, United States of America.
Munksgaard, L., Jensen, M.B., Herskin, M.S., Levendahl, P., 2005. The need for lying time in high producing dairy cows. In: Proceedings of the 39th International Congress of the ISAE, Kanagawa, Japan, pp. 38–42.
Nadimi, E.S., Bak, T., Izadi-Zamanabadi, R., 2006. Monitoring animals and herd behavior parameters using a wireless sensor network. In: Proceedings of XVI CIGR World Congress. Book of Abstracts, pp. 415–416.
Nadimi, E.S., Søgaard, H.T., Oudshoorn, F.W., Blanes-Vidal, V., Bak, T., 2007. Monitoring cow behavior parameters based on received signal strength using wireless sensor networks. In: Cox, S. (Ed.), Proceedings of 3rd European Conference on Precision Livestock Farming (ECPLF)., ISBN 978-90-8686-023-4, pp. 95–103.
Ng, M.L., Leong, K.S., Hall, D.M., Cole, P.H., 2005. A simple passive UHF RFID tag for livestock identification. In: Proceedings of IEEE International Symposium on Microwave, Antenna, Propagation and EMC Technologies for Wireless Communication, pp. 67–70.
Oudshoorn, F.W., Nadimi, E.S., 2007. Intelligent grazing management using wireless sensor networks. In: Cox, S. (Ed.), Proceedings of 3rd European Conference on Precision Livestock Farming (ECPLF)., ISBN 978-90-8686-023-4, pp. 111–116.
Oudshoorn, F.W., Kristensen, T., Nadimi, E.S., 2008. Dairy cow defecation and urination frequency and spatial distribution related to time limited grazing. Livestock Sci. 113, 62–73.
Schwager, M., Anderson, D.M., Butler, Z., Rus, D., 2007. Robust classification of animal tracking data. J. Comput. Electron. Agric. 56 (1), 46–59.
Sinopoli, B., Schenato, L., Franceschetti, M., Poolla, K., Jordan, M.I., Sastry, S.S., 2004. Kalman filtering with intermittent observation. IEEE Trans. Automatic Control 9, 1453–1464.
Szewczyk, R., Osterweil, E., Polastre, J., Hamilton, M., Mainwaring, A., Estrin, D., 2004. Habitat monitoring with sensor networks. J. Commun. ACM 47 (6), 34–40.
Torjusen, H., Lieblein, G., Wandel, M., Francis, C.A., 2001. Food system orientation and quality perception among consumers and producers of organic food in Hedmark County, Norway. J. Food Quality Preference 12 (3), 207–216.
Wang, N., Zhang, N., Wang, M., 2006. Wireless sensors in agriculture and food industry—recent development and future perspective. J. Comput. Electron. Agric. 50 (1), 1–14.
White, S.L., Sheffield, R.E., Washburn, S.P., King, L.D., Green, J.T., 2001. Spatial and time distribution of dairy cattle excreta in an intensive pasture system. J. Environ. Quality 30, 2180–2187.
Wilson, S.C., Dobos, R.C., Fell, L.R., 2005. Spectral analysis of feeding behavior of cattle kept under different feedlot conditions. J. Appl. Anim. Welfare Sci. 8 (1), 13–24.