Automated Estrus Detection

Karmella A. Dolecheck and Jeffrey M. Bewley[1]

[1]Associate Extension Professor
University of Kentucky Department of Animal and Food Sciences
jbewley@uky.edu; 859-257-7543

Take Home Messages

- Examples of automatically measured parameters related to behavioural estrus include mounting events, activity, rumination, body temperature, and progesterone (P4) levels.
- Correctly identified estrus events are true positives (TP), non-alerted estrus events are false negatives (FN), non-alerted non-estrus events are true negatives (TN), and alerted non-estrus events are false positives.
- Detection is a balance of sensitivity and specificity.
- Reasons producers may consider adopting automated technologies include current reductions in availability of skilled labor, greater opportunities to meet production goals, and increased electronic record keeping opportunities.
- Producers may reject automated technologies because of lack of confidence in technology and uncertainty in payback period.

Introduction

Automated estrus detection (AED) technologies supplement a producer's ability to collect information about their cows without increasing cow stress through disturbances or handling (Wathes et al., 2008). Examples of automatically measured parameters related to behavioural estrus include mounting events, activity, rumination, body temperature, and progesterone (P4) levels (Senger, 1994; Saint-Dizier and Chastant-Maillard, 2012; Fricke et al., 2014b).

Algorithms

A common problem with all information, except mounting behaviour, collected using automated technologies is that some changes in cow behaviour and physiology are not exclusive to estrus. As a result, software specific algorithms (sets of rules to follow during calculations) must be used to compare an animal's current behaviour with a cow-specific reference period, creating an estrus alert when a set threshold is exceeded (Saint-Dizier and Chastant-Maillard, 2012). To determine usefulness of a technology, comparisons are made between estrus events identified by the technology and a gold standard such as visual observation (VO), ultrasonography, blood or milk P4 levels, or a combination of these. Correctly identified estrus events are true positives (TP), non-alerted estrus events are false negatives (FN), non-alerted non-estrus events are true negatives (TN), and alerted non-estrus events are false positives (FP; Firk et al., 2002). Detection is a balance of sensitivity and specificity. Sensitivity, the probability that an event is alerted, is equal to TP/(TP+FN)*100 (Hogeveen et al., 2010). Specificity, the probability that when an event does not occur no alert is generated, is equal to TN/(TN+FP)*100. Because neither sensitivity nor specificity account for the prevalence of the event, other comparative measurements are also useful. These include positive predictive value [PPV = TP/(TP+FP)*100] and negative predictive value [NPV = TN/(TN+FN)*100]. Other common measures of detection ability include error rate [FP/(TP+FP)*100] and accuracy [(TP+TN)/(TP+TN+FP+FN)].

The algorithms used for any technology alert will greatly influence success (Saint-Dizier and Chastant-Maillard, 2012). An Australian study testing 5 different algorithms for AED using automated activity monitoring (AAM) found a variation in sensitivity ranging from 79.4 to 94.1% and a variation in specificity between 90.0 and 98.2% (Hockey et al., 2010). Similar studies testing different algorithms for AED using AAM reported sensitivities from 51 to 87% (Roelofs et al., 2005a; Lovendahl and Chagunda, 2010).

▪ Mounting

Standing estrus is the most definitive sign of estrus because it occurs almost exclusively in animals experiencing estrus. This behaviour has been monitored automatically using pressure-sensitive technologies glued to the tailhead of the cow (Xu et al., 1998; At-Taras and Spahr, 2001; Cavalieri et al., 2003a; Saint-Dizier and Chastant-Maillard, 2012). When activated by a standing event, cow ID, date, time, and duration of mount are sent to a computer to be reviewed. Standing events per estrus and length of standing estrus have been recorded using these devices in multiple studies (Stevenson et al., 1996; Dransfield et al., 1998; Xu et al., 1998; At-Taras and Spahr, 2001; Cavalieri et al., 2003b). The most recent study, conducted by Johnson et al. (2012), found 18.4 ± 8.9 standing events per 6.0 ± 4.9 h estrus period. Each standing event can last 2.3 to 3.8 s (Xu et al., 1998; At-Taras and Spahr, 2001). Season can affect results, with hot weather decreasing duration of estrus, but not number or duration of individual mounts (At-Taras and Spahr, 2001). Number of mounts can be affected by both parity and days in milk (DIM), with primiparous cows and cows <80 DIM having increased occurrence (Xu et al., 1998; At-Taras and Spahr, 2001; Peralta et al., 2005).

Cavalieri et al. (2003b) compared VO of estrus length and number and duration of mounts to rump-mounted pressure-sensitive technologies and found low correlations in synchronized cows. Still, both methods were successful at detecting estrus with sensitivity rates of 97.5% and 93.8% for VO and rump-mounted pressure-sensitive technologies, respectively (Cavalieri et al., 2003b). Additional studies agree that rump-mounted pressure-sensitive technology sensitivity is comparable to or better than VO in both cows (Xu et al., 1998; At-Taras and Spahr, 2001; Saumande, 2002; Peralta et al., 2005) and heifers (Stevenson et al., 1996), with sensitivity and PPV as high as 91.7 and 100%, respectively, using milk P4 as a comparison. Rump-mounted pressure-sensitive technologies have also shown results comparable to tail paint and pedometers (Cavalieri et al., 2003a).

One limitation of rump-mounted pressure-sensitive technologies is the labor required to attach and remove them because they are not left on the animal for an entire lactation like other automated technologies can be (Rorie et al., 2002). Additionally, some studies have reported estrus detection trouble because of lost or displaced monitors (Dohi et al., 1993; Xu et al., 1998). Researchers have considered a subcutaneous implantable device for measuring pressure from mounting, but concerns of animal welfare, consumer perception, and potential residue issues have limited development (Senger, 1994).

Recently, an alternative method of automated mounting detection has shown potential (Homer et al., 2013). An ultra-wideband radio technology captured 3-dimensional positioning of animals to determine height changes associated with cows mounting others or standing to be mounted. Although promising, further commercial demonstration of this method is necessary.

A restraint of both of these mounting behaviour monitors is that mounting behaviour must occur for them to work (Saint-Dizier and Chastant-Maillard, 2012). Multiple studies have reported standing estrus occurrence in fewer than 50% of estrus events (Van Eerdenburg et al., 1996; Heres et al., 2000). Modern facilities, especially concrete, limit mounting behaviour (De Silva et al., 1981; Britt et al., 1986). Additionally, most pressure sensitive systems only detect mounts lasting ≥ 2 s, but 40% of mounts may last < 2 s (Walker et al., 1996).

■ Activity

An increase in activity associated with estrus was first observed in rats in 1923 (Wang, 1923). Additional research showed this response in other female mammals, including humans, swine, and cattle (Altmann, 1941; Farris, 1944;

Farris, 1954). One of the first activity monitoring studies in cattle found that number of steps per hour increased 2 to 4 times in cows experiencing estrus compared with cows not in estrus (Kiddy, 1977). Duration of the activity increase associated with estrus is 16.1 ± 4.7 hours (Valenza et al., 2012) and multiple studies from a recent review estimated current AAM systems can accurately detect 70% of cows in estrus (Fricke et al., 2014b). Two types of AAM systems are currently available: 1) pedometers, usually attached to the leg and 2) accelerometers, which have been attached to the neck, leg, or ear (Saint-Dizier and Chastant-Maillard, 2012). Pedometers measure the number of steps taken and accelerometers measure three-dimensional movement, estimating overall activity (Fricke et al., 2014b).

In a recent comparison between AAM and VO conducted by Michaelis et al. (2014), no difference in estrus detection rate existed (42.1 vs. 37.3%, respectively). The sensitivity and PPV of AAM (35.6 and 83.3%, respectively) were numerically, but not significantly, greater than VO (34.3 and 75.1%). The ability of AAM to produce similar or better results than VO has also been shown in other research (Peter and Bosu, 1986; Liu and Spahr, 1993; At-Taras and Spahr, 2001). Automated activity monitoring can also be useful in heifers under a variety of housing systems, including pasture, dry lot, and tie stall (Sakaguchi et al., 2007). Comparisons between AAM and other estrus detection methods also exist. Cavalieri et al. (2003a) compared estrus detection of a pedometer, a rump-mounted pressure-sensitive mounting detector, and tail paint using milk P4 levels and pregnancy diagnosis and found no differences in sensitivities (81.4, 88.4, and 91.3%, respectively).

Reports concerning the percent of estrus events identified using AAM vary between 51 and 84% in both confinement and pasture situations (Lewis and Newman, 1984; Redden et al., 1993; Roelofs et al., 2005a; McGowan et al., 2007; Hockey et al., 2010; Kamphuis et al., 2012; Valenza et al., 2012). Yaniz et al. (2006) stated that a reduction in physical activity occurs with increased milk production, parity, and temperature humidity index. Holman et al. (2011) agreed that high milk yield and low BCS may negatively affect AAM sensitivity, additionally adding that lameness can affect results from leg mounted technologies. However, synchronization, parity, cow age, milk yield, season, DIM, and weather have been found in other studies to have no effect on physical activity (At-Taras and Spahr, 2001; Yaniz et al., 2006).

Recently, studies have focused on comparing AAM to timed artificial insemination (TAI). In 2010, Galon (2010) found no difference in first service conception rate (CR) between Ovsynch (17.6%) and pedometers (22.6%). A more comprehensive study compared TAI to AAM using over 900 animals from 3 herds (Neves et al., 2012). Time to pregnancy was shorter (82 vs. 125 days) for cows bred using the AAM.

▪ Rumination

Automated rumination monitoring can use a microphone system that lies on the cow's neck to identify the regurgitation and re-chewing of cud (Burfeind et al., 2011) or an accelerometer to identify motions associated with rumination (Bikker et al., 2014). Schirmann et al. (2009) validated a commercial, microphone-based rumination monitoring device, finding high correlations to VO of 51 cows (r = 0.93). Because of the decrease in feed intake during estrus (Maltz et al., 1997), the resulting decrease in rumination provides another possible method for AED (Reith and Hoy, 2012). Reith and Hoy (2012) showed a reduction in rumination on the day of estrus from a baseline of 429 min/day to 355 min/day. Overall, mean decrease in rumination during 265 estrus events was 17% (74 min), but with high variation (-71 to +16%). In a follow-up study that looked at 453 estrous cycles, rumination time decreased 19.6% (83 min/day) on the day of estrus (Reith et al., 2014). Pahl et al. (2015) also found a decrease in rumination on the day of (19.3%) and the day before (19.8%) inseminations leading to pregnancy.

▪ Temperature

Cow temperature fluctuates throughout the estrous cycle, being lowest just before estrus, high on the day of estrus, and low again at the time of ovulation in comparison to the high temperatures seen throughout the luteal phase of the cycle (Wrenn et al., 1958; Lewis and Newman, 1984; Suthar et al., 2011). The decrease before estrus may result from lowered P4 levels after luteolysis (Wrenn et al., 1958; Kyle et al., 1998), though Suthar et al. (2011) identified no correlation between body temperature and serum P4 concentrations (r = 0.018). The increase in temperature during estrus could be associated with the increase in activity during behavioural estrus (Walton and King, 1986; Redden et al., 1993), yet tie stall cows, whose movement is constricted, have also experienced increases in vaginal temperature during estrus (Suthar et al., 2011). Other hypotheses for increased vaginal temperature surrounding estrus are enhanced blood flow to the area (Suthar et al., 2011) and correlation with the luteinizing hormone (LH) surge (Clapper et al., 1990).

Regardless of reasoning, reticulorumen boluses, vaginal inserts, temperature monitoring ear tags, and milk temperature sensing technologies originally designed for disease detection could provide an additional method of estrus detection. Vaginal temperature increases between 0.10 and 1.02ºC (Lewis and Newman, 1984; Redden et al., 1993; Kyle et al., 1998; Fisher et al., 2008; Suthar et al., 2011). Milk temperature increases of 0.3ºC (Maatje and Rossing, 1976; McArthur et al., 1992) have been recorded during estrus. Rectal temperatures, though non-automated, have even greater reported increases during estrus (1.3ºC; Piccione et al., 2003). These temperature increases last for 6.8 ± 4.6 hours in dairy cows and 6.5 ± 2.7 hours in beef cows (Redden et al., 1993; Kyle et al., 1998).

Maatje and Rossing (1976) found 84% of visually observed estrus events were identifiable using twice-daily milk temperature monitoring. A follow-up study by McArthur et al. (1992) introduced skepticism after only 50% of estrus events were identified via milk temperature monitoring compared to P4 concentrations in the milk. Other studies have focused on vaginal temperature monitoring, finding sensitivities ranging from 69 to 86% compared to P4 concentrations, making them similar to VO (Redden et al., 1993; Kyle et al., 1998). Overall, temperature monitoring as a tool for estrus detection has both potential and difficulties (Ball et al., 1978; Schlünsen et al., 1987; Fordham et al., 1988; Cooper-Prado et al., 2011; Culmer, 2012). Past challenges have included large daily fluctuations in temperature, variability in temperature rises, seasonal variation, and problems with data recovery from reticulorumen temperature boluses. Many studies agree that temperature alone may not be specific enough to use for estrus detection because of the variety of factors (sickness, ambient temperature, water intake, etc.) that may also affect it (Walton and King, 1986; Fordham et al., 1988).

A newly proposed tool for automated temperature monitoring is measurements of body surface temperature using infrared technology (Talukder et al., 2014). Although originally discredited for high rates of FP and FN (Hurnik et al., 1985), new technology has been developed that is much more promising. Talukder et al. (2014) measured surface temperature on the vulva and muzzle of 20 cows and identified a significant decrease in temperature 48 h before, an increase 24 h before, and another decrease at ovulation as determined by ultrasound evaluation. The sensitivity and specificity of this method for estrus detection compared to plasma P4 varied from 58 to 92% and 29 to 57%, respectively, depending on the algorithm used. Creation of an accurate algorithm and automation of vulval temperature monitoring is challenging because of fecal contamination and tail placement. Alternative locations for infrared temperature monitoring such as the eye and back of the ear may be more appropriate (Hoffmann et al., 2013).

▪ Progesterone

Progesterone measurements can be estimated through both blood and milk sampling and are often used as the gold standard comparison when testing other estrus detection methods (Firk et al., 2002). Roelofs et al. (2006) demonstrated that milk P4 concentrations decline to < 5 ng/ml 80 hours before and < 2 ng/ml 71 hours before ovulation, with blood P4 following a similar pattern. Multiple reproductive parameters can be gained from measuring P4, including identification of estrus and estrus detection errors, likelihood of insemination success, pregnancy diagnosis or loss, ovarian cyst diagnosis, anestrus identification, and evaluation of responses to hormone intervention (Nebel, 1988; Blom and Ridder, 2010; Mazeris, 2010; Saint-Dizier and Chastant-Maillard, 2012). On-farm individual milk P4 tests have been developed (Marcus and Hackett, 1986; Worsfold et al., 1987; Nebel, 1988), but are not automated.

An alternative is automated detection through inline milk sampling systems (Pemberton et al., 2001; Gillis et al., 2002; Saint-Dizier and Chastant-Maillard, 2012). The only commercially available system of this kind is Herd Navigator (DeLaval, Tumba, Sweden), which collects milk at specific time points throughout the estrous cycle to determine a P4 curve for each cow (Friggens and Chagunda, 2005; Mazeris, 2010). An algorithm in the system then determines if the cow receives an estrus alert depending on her point in the estrous cycle. A group of Danish herds using the Herd Navigator system have reported CR between 40 and 63% and a mean reduction in days open (DO) of 22 d since adoption (Blom and Ridder, 2010). A separate survey reported that pregnancy rate of Herd Navigator test farms changed from 22.8% pre-installation to 40% two years later (Durkin, 2010).

Compared to inseminations resulting in pregnancy, the high sensitivity (93.3%) and specificity (93.7%) for estrus detection has identified the usefulness of the Herd Navigator as an AED tool (Friggens et al., 2008). Furthermore, the Herd Navigator can also conduct measurements of lactate dehydrogenase, urea, and β-hydroxybutyrate to detect metabolic diseases and mastitis. Regardless, high cost of the system has limited its adoption.

▪ Others

Lewis and Newman (1984) found vaginal pH to be lowest on the day of estrus, milk yield to be decreased surrounding estrus, and heart rate to be slowest during estrus. However, these variations were small and repeated measurements (because of lack of automation) are not yet feasible for commercial dairies. Similar, inability to automate has reduced interest in other areas, including monitoring electrical resistance of vaginal mucus, dry matter concentration and crystallization patterns of vaginal mucus, and blood P4 around estrus (Noonan et al., 1975; Leidl and Stolla, 1976; Heckman et al., 1979).

▪ Technology Combinations

According to de Mol et al. (1997), the missing link in automated technology monitoring is merging all available data. Combinations of multiple parameters would improve estrus detection rate when certain conditions (environmental temperature, pen changes, etc.) interfere with one monitoring method (Firk et al., 2002). Maatje et al. (1997a) considered the combination of activity, milk yield, and milk temperature for estrus detection, finding sensitivity improvements of 10 to 20% over activity alone. Peralta et al. (2005) also tested three parameters, finding the sensitivity of VO, activity monitoring, and mounting detection alone was 49.3%, 37.2% and 48%, respectively. The combination of all three systems increased estrus detection sensitivity to 80.2%. Additional studies have shown the usefulness of combining multiple variables for estrus detection (Redden et al., 1993; de Mol and Woldt, 2001b; Brehme et al., 2008; O'Connell et al., 2011).

Merging automatically collected data (activity, rumination, etc.) with an individual cow's history can also improve estrus detection algorithms. Firk et al. (2003) demonstrated that including information about the length of time since a cow's last estrus period decreased sensitivity from 91.7 to 87.9% but improved error rate from 34.6 to 12.5%.

The potential for multiple parameter combinations in estrus detection requires improved data analysis compared to univariate scenarios. Some multivariate evaluation techniques include statistical process control, fuzzy logic, neural networks, and machine learning.

Statistical process control monitors and detects changes in data over time. Control limits are set through calculations of the mean variation between observations, and when an observation goes outside of those control limits, an alert is triggered (De Vries and Conlin, 2003b). This allows the model to distinguish between natural variation and real change. Statistical process control has been used to manage mastitis (Niza-Ribeiro et al., 2004; Lukas et al., 2005) and reproductive performance (De Vries and Conlin, 2003a, b).

Fuzzy logic analysis involves 3 steps: fuzzification, fuzzy inference, and defuzzification (Firk et al., 2002). Fuzzification is the process of transforming real variables into linguistic variables. Fuzzy inference then applies rules to the transformed variables in a fashion similar to "if, then" statements to classify them. Defuzzification returns the values created by fuzzification and fuzzy inference back to readable values. In the dairy industry, fuzzy logic has been applied to mastitis (De Mol and Woldt, 2001a; Cavero et al., 2006; Kramer et al., 2009), lameness (Kramer et al., 2009), and estrus detection (De Mol and Woldt, 2001a).

Neural networks do not require a specific algorithm to work (Grzesiak et al., 2006). Instead, they learn how to make associations and adapt when presented with new data. Although most commonly used in engineering, business, and medicine, some models can predict milk production (Sanzogni and Kerr, 2001; Grzesiak et al., 2006; Sharma et al., 2007) and mastitis occurrence (Heald et al., 2000; Hassan et al., 2009).

Machine learning is another method of programming that allows for constant algorithm improvement through experience and data analysis (Alpaydin, 2004). Machine learning is applicable to retailers who track customer behaviour, financial institutions when identifying risk, and manufacturing scenarios to help minimize resource consumption. Reproductive performance in the dairy industry has also been evaluated using machine learning (Mitchell et al., 1996; Caraviello et al., 2006a; Shahinfar et al., 2013).

▪ Technology Effect on Timing of Insemination

Pregnancy outcome is dependent on timing of AI relative to ovulation (Nebel et al., 1994). Automated monitoring technologies' ability to predict ovulation may help maximize CR by determining ideal AI time (Senger, 1994). Dransfield et al. (1998) evaluated 2,661 inseminations in 17 herds and reported the highest CR when cows underwent AI 4 to 12 hours after the onset of standing activity as measured by an automated rump-mounted pressure-sensitive technology. A similar study using pedometer readings showed AI 6 to 17 hours after increased activity levels resulted in the highest CR, with no effect of disease, inseminator, or bull on the results (Maatje et al., 1997b). Vaginal temperature has also shown a high correlation ($r = 0.74$) with ovulation (Rajamahendran et al., 1989), and strong relationships with the LH peak (Clapper et al., 1990; Mosher et al., 1990; Fisher et al., 2008).

Automated technologies' ability to measure intensity and duration of estrus may further improve CR. Dransfield et al. (1998) reported that the probability of pregnancy increased with an increased number of standing events. Cows that stood for mounting less than 3 times experienced a 41% lower chance of becoming pregnant compared to cows that stood to be mounted 3 or more times before AI. Stevenson et al. (1983) agreed that increased estrus intensity resulted in a significant positive effect on CR.

▪ Technology Adoption

Technology adoption on dairy farms has been slow (Russell and Bewley, 2013). In 2007, the USDA estimated dairy herds using pedometers and pressure sensing technologies for estrus detection at 1.4 and 5.7%, respectively (USDA, 2007). Nevertheless, Borchers and Bewley (2014) recently conducted a producer survey and identified high adoption interest in mounting and cow activity monitoring technologies.

Reasons producers may consider adopting automated technologies include current reductions in availability of skilled labor, greater opportunities to meet production goals, and increased electronic record keeping opportunities (Wathes et al., 2008). Producers may reject automated technologies because of lack of confidence in technology and uncertainty in payback period. Russell and Bewley (2013) conducted a survey to identify reasons for slow technology adoption in Kentucky herds, and 42% and 30% of producers identified undesirable cost to benefit ratio and no economic value, respectively. Borchers and Bewley (2014) also identified economics (benefit to cost ratio and investment cost) as the two biggest factors influencing technology adoption. These results highlight the importance of evaluating economic feasibility of automated technologies.

■ References Available Upon Request