Journal of Water Supply: Research and Technology—AQUA | 63.1 | 2014

Optimal sensor placement for event detection and source identification in water distribution networks

Shuming Liu and Pierre Auckenthaler

ABSTRACT

This study focuses on the optimization of sensor placement with respect to the source identification and event detection. A multi-objective algorithm is used to solve the optimization problem. The numbers of possible source nodes for contamination events associated with the solutions on the Pareto fronts from the proposed method and benchmark method are calculated under the same configuration and compared. The comparison showed that the proposed method performs better than the benchmark method in detecting a contamination event and identifying its possible source.
Key words | multi-objective genetic algorithm, nodal demand uncertainties, sensor placement, source identification

Shuming Liu (corresponding author)
Pierre Auckenthaler
School of Environment,
Tsinghua University,
Beijing 100084,
China
E-mail: *shumingliu@tsinghua.edu.cn*

Shuming Liu
Tsinghua University-VEOLIA Environment Joint Research Center for Advanced Environmental Technology,
Tsinghua University,
Beijing 100084,
China

INTRODUCTION

Accidental contamination or malicious attack on a water distribution network can have severe effects on the health of the population relying on this water source. These events can potentially be detected by a network of online sensors. In order to reduce the exposure to contaminated water, it is important to design a methodology to efficiently place the sensors (Ostfeld *et al.* 2008). Once a contamination event is detected, the next important step is to identify the source of this event.

The sensor placement problem is concerned with locating an array of sensors in a network capable of detecting the presence of a contaminant if an event occurs. Since the first article by Lee & Deininger (1992), the sensor placement problem sparked significant research, most notably, leading to the Battle of the Water Sensor Networks (BWSNs) (Ostfeld *et al.* 2008). Due to its complexity, it is often treated as a multi-objective optimization problem. Optimization criteria may include population exposed, time to detection, volume of contaminated water consumed, number of failed detections, and length of pipe contaminated (Berry *et al.* 2005). These objectives are mostly related to the sensor's ability in event detection.

The source identification problem is concerned with identifying the contaminant source location (Tryby *et al.* 2010). It also attracted much interest and several methods have been developed to solve it, especially in recent BWSNs. For example, Cristo & Leopardi (2008) formulated it as an optimization problem, linearized using the water fraction matrix concept. Liu *et al.* (2011) solved this problem using an adaptive dynamic optimization technique. Wang & Harrison (2013, 2014) applied Bayesian analysis to solve the source identification problem, especially in the condition of stochastic variation of water demand. Shen & McBean (2012) investigated the false negative/positive issues in contaminant source identification.

Generally, these two problems were solved separately. Only a few researchers have attempted to investigate them jointly. In 2006, Preis & Ostfeld observed this dependency and proposed a method to optimize the sensor layout with respect to source identification (Preis & Ostfeld 2006). More recently, Tryby *et al.* (2010) represented the water distribution system as a linear system and optimized the positions of the sensors in order to improve the conditioning of the matrix. This paper presents a new method to focus on the optimization of sensor placement with respect to source identification and event detection. The proposed method utilizes different objectives compared with conventional methods.

doi: 10.2166/aqua.2013.106

METHODOLOGY

Nodal demand uncertainty

In the operation of a network, uncertainties might come from nodal demand, sudden loss of electricity, etc. For simplification, only the nodal demand uncertainty is considered in this research, which is realized by multiplying the base demand with a random demand multiplier. A truncated Gaussian probability function (PDF) is employed to generate the random demand multiplier (Babayan *et al.* 2005). This is expressed as:

$$\text{pdf} = \frac{1}{\sigma\sqrt{2\pi}} \exp^{-((x-E)^2/2\sigma^2)} \cdot x \geq -1 \tag{1}$$

where E and σ are given mean and standard deviation. In this research, the values of σ are 0, 0.1 and 0.2, and $E = 0$.

The random nodal demand is calculated as:

$$w(k, j) \rightarrow w_{\text{b}}(k, j) \cdot M(k, j) \tag{2}$$

where $w(k, j)$ is the demand at network junction k at time step j, $w_{\text{b}}(k, j)$ is the base demand at network junction k at time step j, $M(k, j)$ is the demand multiplier at network junction k at time step j.

Simulation of contamination data

Features of contamination events, including contaminant type, duration, contaminant magnitude and starting time, may have impacts on source identification. For simplification, it is assumed that the contamination is a single conservative injection event with specified unit duration and could happen at any point of the discretized timeline in the first 12 hours. Such an event is noted e and the set of all possible events is noted E. The type of sensors considered is with an alarm threshold d. If the indicator concentration (for example, dissolvedoxygen) is below d, no alarm will be triggered. If the concentration goes above d, the sensor immediately reports an abnormal situation. The impact of alarm threshold on model performance is examined by adding a shift Δ to the original alarm threshold. The values of Δ adopted are -10% and $+10\%$.

Ideally, all concentration values from all nodes at every time step, for all simulations, should be recorded. However, such records would lead to a large amount of data, much too large to be used efficiently, even for small networks (Krause *et al.* 2008). Instead, it can record the time of first detection of the simulated event by each node to reduce computing load (Krause *et al.* 2008), which is the principle used in this research. The time at which a sensor s (at node s) detects the event e under the nodal demand configuration i is called the detection time $t(e, s, i)$. $t(e, s, i)$. is calculated to be the shortest travel time of water parcels from the contamination event node to the sensor node, which is used as an indicator to identify the contamination source of event e for sensor at node s. In addition, the interval of detection time is denoted $[T_{\min}(e, s); T_{\max}(e, s)]$, in which $T_{\min}(e, s) = \min t(e, s, i)$ $T_{\max}(e, s) = \max t(e, s, i)$. In the case where the contaminant concentration is lower than the sensor's alarm threshold, it is considered 'non-detection'. If the event goes undetected during the whole simulation, it is defined as $t(e, s, i) = -1$.

Optimization

The sensor placement problem is configured to be a multi-objective optimization problem. The objectives are to (1) maximize the probability of detection and (2) minimize the overlay of interval of detection time. To solve the optimization problem, a NSGA-II method is employed. NSGA-II is an algorithm proposed by Deb *et al.* (2002), which has been broadly used in water distribution area (Preis & Ostfeld 2008; Liu *et al.* 2012).

First objective: maximize the probability of detection

The first objective is to maximize the probability of detection, which is defined as the ratio of detected events relative to all simulated events:

$$\text{fitness}_1 = \frac{\text{number of detected events}}{\text{total number of events}} \tag{3}$$

Second objective: minimize the overlay of interval of detection time

A single event will give a deterministic value of the detection time for a sensor. In this case, it is possible to distinguish two events by comparing the detection times. In the case where the detection times are identical, the two events can be distinguished by further comparing the sensor measurements. In the situation of detection, times and sensor measurements are both identical, the two events are the same from the sensor's point of view and will not be distinguished.

This simple dichotomy disappears when nodal demand uncertainties are considered, because the detection time of an event by a sensor becomes a set of possible times. The set of possible detection times of two events can be identical, completely different or partially overlapped. The partial overlap means that under some nodal demand configurations, the two events can be distinguished by the sensors. While under other nodal demand configurations, they cannot be distinguished. Therefore, the second objective is to minimize the overlap of internal of detection time. A value $v(s, e_1, e_2)$ is first calculated, corresponding to the ability for a single sensor s to distinguish between two events e_1 and e_2:

$$\begin{aligned} v(s, e_1, e_2) &= \frac{1}{2}\left(\frac{\text{overlap}}{l_1} + \frac{\text{overlap}}{l_2}\right) \\ l_1 &= T_{\max}(e_1, s) - T_{\min}(e_1, s) \\ l_2 &= T_{\max}(e_2, s) - T_{\min}(e_2, s) \end{aligned} \tag{4}$$

where l_1 and l_2 are the length of the interval of detection time for the first and second event, $l_1 \neq 0$, $l_2 \neq 0$, $T_{\max} > 0$, $T_{\min} > 0$, *overlap* is the length of the overlap between the interval of detection time.

The value attributed to a single sensor can be extended to a set of sensors by taking the minimum value from each sensor:

$$v(S, e_1, e_2) = \min_{s \in S} v(s, e_1, e_2) \tag{5}$$

By summary, the optimization problem for optimal sensor placement is formulated as follows:

$$\begin{aligned} &\text{Maximize fitness}_1 = \frac{\text{number of detected events}}{\text{total number of events}} \\ &\text{Minimize fitness}_2(S) = \sum \sum v(S, e_1, e_2) \\ &\qquad (e_1 \neq e_2; \quad e_1, e_2 \in E) \end{aligned} \tag{6}$$

A greater value of fitness_1 suggests an event e can be detected with high probability, while a greater value of fitness means a higher overlay of detection time and lower accuracy of source identification. In the case of non-detection, a large number, for example 10^{10} in this research, is set to fitness_2 as a penalty.

Comparison benchmark

For comparison, another set of solutions is yielded through optimization with respect to probability of detection (the first objective defined above) and time to detection (objective used in conventional methods). Time to detection is defined as the value of the difference between the contamination time (T_{event}) and the median value of the detection interval for the first sensor to detect the event. Small fitness values suggest small detection times:

$$\text{fitness}_3(S) = \min \sum (T_{\min} + 0.5 * (T_{\max} - T_{\min}) - T_{\text{event}}) \tag{7}$$

Therefore, the benchmark optimal sensor placement problem has two objectives: maximization of detection probability and minimization of detection time.

These two multi-optimization problems share one same objective. However, the second objective is different. To facilitate the comparison, after the optimization problem is solved, the numbers of possible source nodes for contamination events associated with the solutions on the Pareto fronts from both methods are calculated under the same configuration as the one used in the optimization processes.

Source identification

At time t, if sensor s_1 is triggered off and sensor s_2 is not, and $0 < T_{\min}(e_x, s_2) \leq T_{\max}(e_x, s_2) \leq T_{\min}(e_x, s_1)$, node x is excluded from the potential contamination source. Otherwise, x is a possible contamination source. At time $t + t'$, if sensor s_2 is triggered off and $0 < T_{\min}(e_x, s_2) \leq T_{\max}(e_x, s_2) \leq t'$, node x is excluded from the potential contamination source. Otherwise, x is a possible contamination source.

CASE STUDY

The methodology discussed above was implemented using Matlab. The hydraulic and water quality simulations were performed using the EPANET Program's toolkit (Rossman 1999). Example network Net 3 from EPANET was utilized to demonstrate the proposed method. For simplification, a base demand pattern is applied to all nodes.

Simulations were carried out with different levels of uncertainties of the nodal demand (σ: 0, 0.1 and 0.2) and alarm threshold (Δ: -10%, 0%, $+10\%$). For each level of uncertainties, the optimization process was conducted for three, six and 12 sensors, respectively. The injection duration is set to 1 minute. The nodal demand w is calculated using Equation (2). Based on the initial calculation of the flow velocity, water quality simulation length was set to 24 hours to ensure all contaminant is spread over the whole network.

RESULTS

Figure 1 presents the Pareto fronts obtained using the proposed method. It reveals that nodal demand uncertainty does not affect detection probability significantly. However, for a given number of sensors, with increasing of uncertainty, for the same detection probability, the identification fitness increases, which suggests a decrease of accuracy of source identification. Meanwhile, as shown in Figure 2, when the alarm threshold decreases, the detection probability and the overlay of interval of detection time increase, which means a contamination event can be easier to be detected, but with lower identification accuracy. In contrast, for a greater alarm threshold value, a contamination event can be more difficult to be detected, but with higher identification accuracy. These two findings suggest the proposed method is sensitive to nodal demand and alarm threshold.

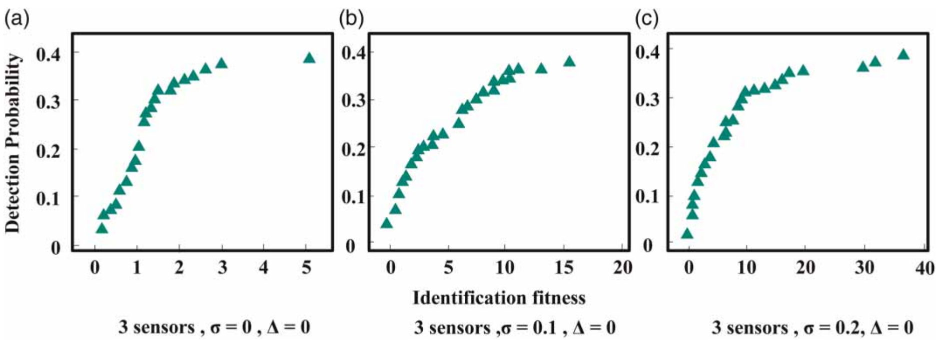

Figure 1 | Pareto front from the proposed method (nodal demand).

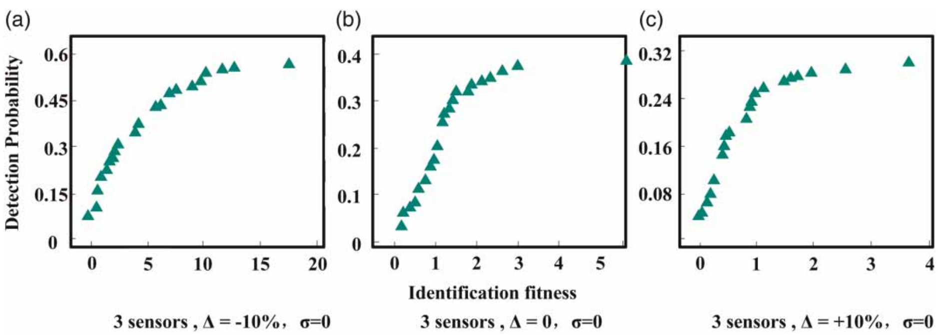

Figure 2 | Pareto front from the proposed method (alarm threshold).

The possible source nodes for contamination events associated with the solutions on the Pareto fronts from the proposed and benchmark methods are calculated and presented in Figure 3. Generally, for the same number of possible contamination sources, the results of the proposed method show higher detection probability than the one from the benchmark method. Meanwhile, for a given contamination event, the proposed method reports a lower number of possible contamination sources than the benchmark method with the same detection probability. This suggests that the proposed method performs better than the benchmark method in event detection and source identification.

Figure 4 presents the Pareto front obtained with six sensors, as well as the explicit placement of the sensors for four different sensors sets. The solutions present a gradual evolution going from the most compact sensor placement (Figure 4(a)), giving a better identification accuracy with low detection probability, to the most scattered one, giving a worse identification accuracy with high detection probability (Figure 4(d)). A compact sensor placement can identify more accurately the source of a contamination event but will miss many events. On the contrary, a well spread out sensor network can detect an event with higher probability, while accuracy of source identification is low.

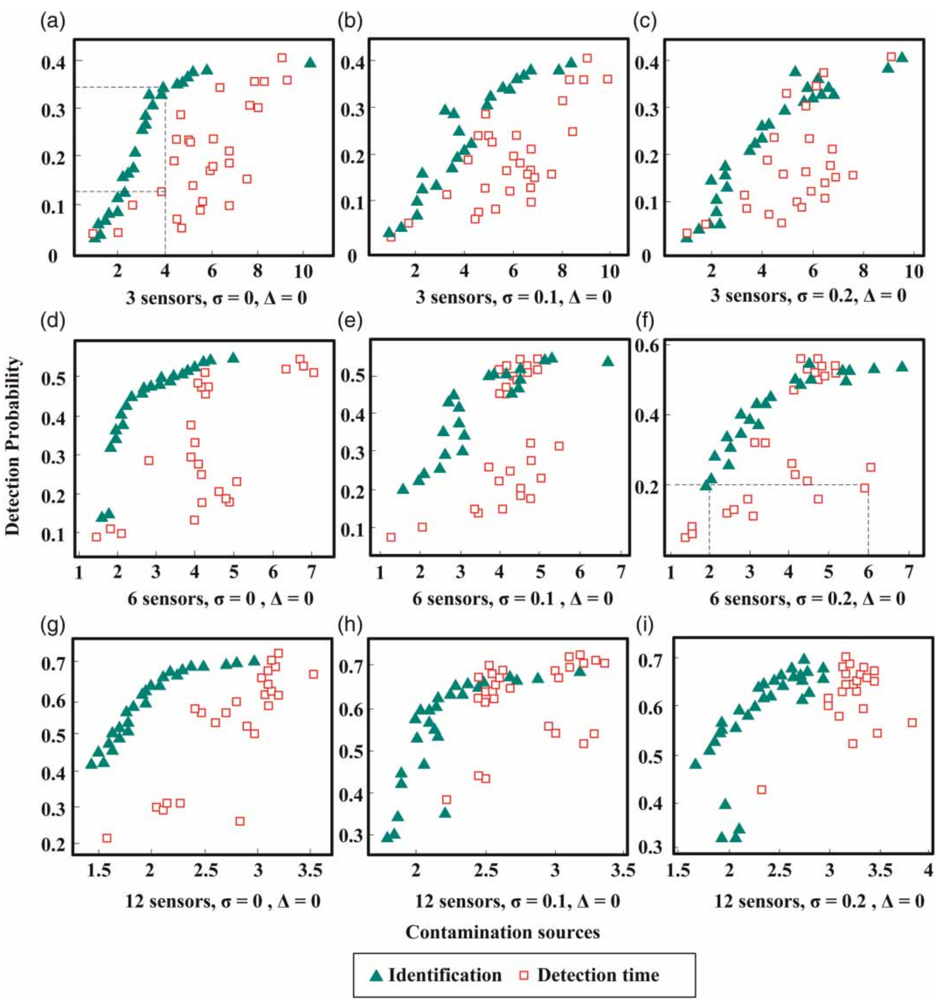

Figure 3 | Possible contamination sources from proposed method and benchmark method.

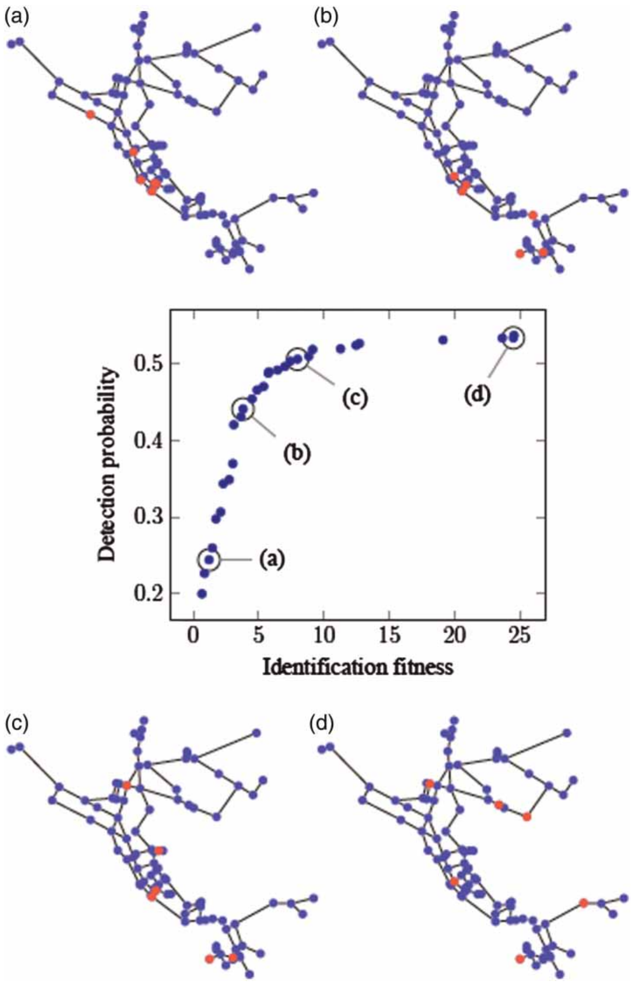

Figure 4 | Four sample solutions of the Pareto front obtained for six sensors with $\sigma = 0.1$.

CONCLUSIONS

1. The proposed method employs detection probability and overlap of detection time as objectives, which differentiates it from the conventional methods. By taking the overlap of detection time into consideration, the proposed method performs well, both in detecting a contamination event and identifying its possible sources.
2. The proposed method is sensitive to nodal demand and sensor's alarm threshold. Increase of nodal demand uncertainty and decrease of alarm threshold both could blur the accuracy of source identification.
3. A compact sensor placement will make use of the redundancy between the sensors to identify more accurately the source of a contamination event. On the contrary, a well spread out sensor network can detect an event with higher probability, while accuracy of source identification is low.
4. Many factors may have impacts on the model's performance. This research only simply examined nodal demand and alarm threshold. The impact of the profile of contamination event on method performance, including injection duration and contaminant magnitude, etc., and the use of detection time and sensor measurement to differentiate events should be investigated in future research.

ACKNOWLEDGEMENTS

The authors express their thanks to the China Major Water Project (2012ZX07408-002) and Tsinghua Independent Research Grant for their financial support. The authors would like to thank the two anonymous reviewers for their valuable comments on this manuscript.

REFERENCES

Babayan, A., Kapelan, Z., Savic, D. & Walters, G. 2005 Least-cost design of water distribution networks under demand uncertainty. *J. Water Resour. Plann. Manage.* **131** (5), 375–382.

Berry, J. W., Fleischer, L., Hart, W. E., Phillips, C. A. & Watson, J. P. 2005 Sensor placement in municipal water networks. *J. Water Resour. Plann. Manage.* **131** (3), 237–243.

Cristo, C. D. & Leopardi, A. 2008 Pollution source identification of accidental contamination in water distribution networks. *J. Water Resour. Plann. Manage.* **134** (2), 197–202.

Deb, K., Pratao, A., Agarwal, S. & Meyarivan, T. 2002 A fast and elitist multiobjective genetic algorithm: NSGA-II. *IEEE Trans. Evol. Comput.* **6** (2), 182–197.

Krause, A., Leskovec, J., Guestrin, C., VanBriesen, J. & Faloutsos, C. 2008 Efficient sensor placement optimization for securing large water distribution networks. *J. Water Resour. Plann. Manage.* **134** (6), 516–526.

Lee, B. H. & Deininger, R. A. 1992 Optimal locations of monitoring stations in water distribution system. *J. Environ. Eng.* **118** (1), 4–16.

Liu, L., Ranjithan, R. S. & Mahinthakumar, G. 2011 Contamination source identification in water distribution systems using an adaptive dynamic optimization procedure. *J. Water Resour. Plann. Manage.* **137** (2), 183–192.

Liu, S., Liu, W., Chen, J. & Wang, Q. 2012 Optimal locations of monitoring stations in water distribution systems under multiple demand patterns: a flaw of demand coverage method and modification. *Front. Environ. Sci. Eng.* **6** (2), 204–212.

Ostfeld, A., Uber, J. G., Salomons, E., Berry, J. W., Hart, W. E., Phillips, C. A., Watson, P.-J., Dorini, G., Jonkergouw, P., Kapelan, Z., Pierro, F., Khu, S.-T., Savic, D., Eliades, D., Polycarpou, M., Ghimire, S. R., Barkdoll, B. D., Gueli, R., Huang, J. J., McBean, E. A., James, W., Krause, A., Leskkovec, J., Isovitsch, S., Xu, J., Guestrin, C., VanBriesen, J., Small, M., Fischbeck, P., Preis, A., Propato, M., Piller, O., Trachtman, G. B., Wu, Z. Y. & Walski, T. 2008 The battle of the water sensor networks (BWSN): a design challenge for engineers and algorithms. *J. Water Resour. Plann. Manage.* **134** (6), 556–568.

Preis, A. & Ostfeld, A. 2006 Optimal sensors layout for contamination source identification in water distribution systems. In: *Proc. 8th Annual Water Distribution Systems Analysis Symp*. ASCE, Reston, VA.

Preis, A. & Ostfeld, A. 2008 A genetic algorithm for contaminant source characterization using imperfect sensors. *Civil Eng. Environ. Syst.* **25** (1), 29–39.

Rossman, L. A. 1999 The EPANET programmer's toolkit for analysis of water distribution systems. In: *Proc. Annual Water Resources Planning and Management Conf*. ASCE, Reston, VA.

Shen, H. & McBean, E. 2012 False negative/positive issues in contaminant source identification for water-distribution systems. *J. Water Resour. Plann. Manage.* **138** (3), 230–236.

Tryby, M. E., Propato, M. & Ranjithan, S. R. 2010 Monitoring design for source identification in water distribution systems. *J. Water Resour. Plann. Manage.* **136** (6), 637–646.

Wang, H. & Harrison, K. 2013 Bayesian update method for contaminant source characterization in water distribution systems. *J. Water Resour. Plann. Manage.* **139** (1), 13–22.

Wang, H. & Harrison, K. 2014 Improving efficiency of the Bayesian approach for water distribution contaminant source characterization with support vector regression. *J. Water Resour. Plann. Manage.* **140** (1), 3–11. doi:10.1061/(ASCE) WR.1943-5452.0000323.

First received 5 June 2013; accepted in revised form 19 September 2013. Available online 31 October 2013