Statistical modelling AD for process optimization and bench-marking – A case study of \textit{E. coli} inactivation across all Thames Water conventional sewage sludge treatment sites

J. Liu\textsuperscript{1}, Y. Gao\textsuperscript{1}, P. Pearce\textsuperscript{2}, A. Shana\textsuperscript{2} and S.R. Smith\textsuperscript{1}

\textsuperscript{1}Department of Civil and Environmental Engineering, Imperial College London, UK and 
\textsuperscript{2}Thames Water Utilities Ltd, Reading Sewage Treatment Works, Reading, UK

Correspondence

Professor Stephen R. Smith, Department of Civil and Environmental Engineering, Imperial College London, London SW7 2AZ UK. Email: s.r.smith@imperial.ac.uk

Presented at Sludge Tech conference 2015
Abstract

Untreated sewage sludge potentially contains a wide range of enteric pathogens that present a risk to human health. Mesophilic anaerobic digestion (MAD) is the most-favoured process for sewage sludge treatment in the UK. It is a well-established approach to sludge stabilization, but the mechanisms responsible for pathogen removal are poorly understood. Operational data collected by Thames Water from conventional MAD sites were statistically scrutinised to examine the effects of primary and secondary digestion on the removal of the enteric indicator bacteria, *Escherichia coli*, by using the IBM SPSS statistical software package for ANOVA, post-hoc and multiple regression analysis. The results showed that the process temperature conditions at the AD plants were equivalent to or exceeded the minimum estimated by the analysis necessary to comply with the $2 \log_{10}$ removal standard for *E. coli*. The results also showed that primary digestion conditions (specifically temperature) sublethally damaged *E. coli* and increased decay in secondary digestion and therefore over the whole digestion process.

Key words

*Escherichia coli*; mesophilic anaerobic digestion; operational parameters; pathogen inactivation; sewage sludge; statistical analysis
**Introduction**

Sewage sludge is an inevitable by-product of wastewater treatment and contains organic matter and essential plant nutrient resources which can be recycled for crop production and to improve soil structural properties. Recycling treated sewage sludge (biosolids) to agricultural land is generally recognized as the Best Practicable Environmental Option for sludge management (Crathorne *et al*., 2002). In the UK, the use of biosolids on agricultural land is regulated by The Sludge (use in agriculture) Regulations (SI 1989/1263, amended in 1990), and is supported by a Code of Practise for the Agricultural Use of Sewage Sludge (DoE, 1996), which provides additional control measures to ensure environmental safety and protect human health. Furthermore, the Safe Sludge Matrix (ADAS, 2001) prescribes numerical microbiological standards to be attained by sludge treatment processes, as well as providing greater clarity on land-use restrictions for biosolids depending on their microbiological classification.

Anaerobic digestion is the process of decomposition of organic matter by a complex community of symbiotic microorganisms under an oxygen-free environment (Ward *et al*., 2008). The DoE Code of Practice (DoE, 1996) recommends that primary mesophilic anaerobic digestion (MAD) is followed by a secondary digestion stage to significantly reduce the pathogen content of sewage sludge. Sludge is heated to mesophilic temperatures during the primary treatment step and secondary digestion is a separate treatment process performed at ambient temperature. Laboratory-scale simulations (Horan *et al*., 2004), to determine the inactivation rate of the enteric indicator bacteria, *Escherichia coli*, during sewage sludge treatment, showed that primary MAD achieved 1.66 $\log_{10}$ removal and a further 1.70 $\log_{10}$ removal was obtained in the following secondary digestion stage. The typical $E.\ coli$ removal rate for full-scale digestion sites in north-west England was 1.5 $\log_{10}$ by primary digestion, with secondary digestion contributing a further 1.0 $\log_{10}$ reduction (Le *et al*., 2002).

The mechanisms of infectious enteric pathogen and indicator bacteria inactivation during sewage sludge treatment at thermophilic or pasteurisation temperatures (>55°C) are relatively well defined and linked to thermal decay, the intrinsic heat resistance of microbial isolates, stress induced by the sewage sludge matrix environment and process retention time (Smith *et al*., 2005; Lang & Smith, 2008). However, the fundamental processes responsible for pathogen inactivation in lower temperature ranges, experienced, for example, during MAD of sludge, are poorly understood. Further investigation is also required to increase the reliability of compliance of conventional MAD treatment of sludge to consistently obtain a 2 $\log_{10}$ removal of *Escherichia coli* required for agricultural use by the Safe Sludge Matrix. This would
avoid the additional cost and logistical management of withholding sludge and providing for additional storage time to reduce *E. coli* numbers when digested sludge fails to meet the microbiological criteria.

The effects of operational and physico-chemical parameters on the removal of *E. coli* by primary MAD were examined by Gao (2013) from a statistical analysis of operational data collected by Thames Water for conventional digestion sites in the region. This showed that *E. coli* reduction was independent of primary digestion conditions; the only statistically significant parameter influencing *E. coli* removal across the primary digestion process was the number of indicator bacteria initially present in the raw sludge feed with higher numbers increasing the overall removal rate.

The aim of the study reported here was therefore to extend this work to quantify the operational and physico-chemical parameters controlling *E. coli* reduction during the secondary digestion stage and cross the entire primary MAD + secondary digestion process. The approach was based on a statistical analysis of operational data for 18 wastewater treatment plant (WWTP) operated by Thames Water with conventional sludge MAD.

**Methodology**

**Site Information**

Eighteen WWTP operating conventional MAD sludge treatment facilities within Thames Water were selected for analysis. The conventional MAD sites were divided into two groups according to their secondary digestion type and included secondary digesters connected in series or as a batch process.

**Overview and general statistical analysis approach**

A series of statistical analysis procedures were applied to a dataset compiled from several databases, holding different parameter types potentially relevant to *E. coli* removal, provided by Thames Water (Hazard Assessment Critical Control Point (HACCP), Compliance and Final *E. coli* Data), and developed previously by Gao (2013) from Thames Water databases (Combined Cockpit and Compliance Data), for WWTP operating conventional sludge MAD in the region. The overall approach to data merger and statistical analysis is shown in Figure 1. A descriptive analysis and screening process was applied to HACCP (providing information on primary MAD treatment conditions and physico-chemical data) and Compliance (primary MAD output *E. coli* numbers and primary MAD treatment conditions) datasets to remove extreme
outliers and to convert the variables into a consistent format based on monthly average values. These datasets were combined with the Final E. coli Data (providing final E. coli numbers in digested sludge). The time period represented by the combined data sets was 2007 – 2012.

Monthly average values were calculated for the range of operational parameters routinely measured at each WWTP and potentially important to E. coli removal during sewage sludge digestion. These included: hydraulic retention time (HRT), temperature, dry solids (DS) and volatile solids (VS) input, and volatile solids reduction (VSR). Escherichia coli reduction was calculated based on monthly average E. coli input and output values, to provide a representative measure of the indicator removal rate. This approach was preferred to averaging daily reduction values to minimise the effects of process retention time and the potential variation observed in input and output E. coli numbers measured at specific sampling times. Secondary and overall E. coli reductions were derived from the input and output E. coli results provided by the merged dataset. Actual primary HRT was calculated and used in the statistical analysis based on the effective digester volume and nominal HRT data recorded at each site. The average effective digester volume for each site was calculated by Gao (2013) based on tracer test data provided for 2011.

Post-hoc tests were applied following analysis of variance (ANOVA) to determine the significance of specific parameter comparisons. A correlation analysis was also completed to test the significance of the relationships between process and sludge physico-chemical conditions and E. coli reduction. Two multiple regression models were developed incorporating the statistically significant operational variables explaining Secondary and Total E. coli reductions in conventional MAD sludge treatment processes.

The IBM SPSS Statistics 21 programme was used to complete the statistical analysis calculations. Backward elimination and forward selection methods were used to identify predictor variables in the multiple regression models of Secondary and Total E. coli reduction. The tests of agreement with statistical assumptions were divided into three main components: (1) significant outliers, high leverage points or highly influential points; (2) independence of observations, linearity and multicollinearity; and (3) normality of residuals and homoscedasticity. Boxplots were used to identify and removed extreme outliers. The points that were recorded as high leverage (leverage value>0.2) and highly influential points (Cook's Distance values>1) were recorded and removed.
Step1

- HACCP data
- Final E. coli data
- Compliance data

Screening process
Descriptive analysis

Calculate monthly average values for each variable

Combine HACCP, Final E. coli, and Compliance data
- Calculate Secondary and Total E. coli reduction
- Screen E. coli reduction data

ANOVA and post-hoc analysis of Season, Site, and Secondary digestion type effects on E. coli reduction
Correlation analysis between physico-chemical parameters and E. coli reduction
Two-way ANOVA of Year, Site and interaction effects on Secondary E. coli reduction

Combined E. coli, HACCP and Compliance dataset
Step 2

- Combined *E. coli*, HACCP and Compliance dataset
- Combined Cockpit and Compliance dataset

Check consistency of duplicated data types

Final Combined dataset

Correlation analysis (dependent variables: Secondary and Total log₁₀ *E. coli* reduction)

Multiple regression analysis (dependent variables: Secondary and Total log₁₀ *E. coli* reduction)

Evaluate final multiple regression models

**Fig. 1** Statistical analysis procedures
Results

Overview of the dataset

Site average values for each operational variable were calculated from the monthly data and these are summarised as input and output variables in Table 1 and Table 2, respectively.

The mean total *E. coli* reduction for 18 sites was $2.9 \log_{10}$; the mean *E. coli* input was $6.9 \log_{10} \text{ g}^{-1} \text{ DS}$ and final *E. coli* mean output value was $4.0 \log_{10} \text{ g}^{-1} \text{ DS}$. The 18 sites had a mean primary VSR of 45.2% with a mean VS input of 76.7%, and mean primary VS output of 64.2%. Site CRW had the highest average primary digestion temperature (37.7 °C). However, it had the same total *E. coli* reduction rate ($3.0 \log_{10}$) as site BRA, with the lowest average digestion temperature (34.8 °C). The average actual primary HRTs for sites MAP, CAM, BRA, CRW, DEE were below the overall 18 site average value of 12.7 d, nevertheless all of these sites had an average total *E. coli* reduction larger than the overall mean rate for the 18 sites of $2.9 \log_{10}$. However, four sites with the poorest total *E. coli* reduction performance (AYS, BAN, EHY and OXF) had actual primary HRTs above the overall average value. Sites: CAM, BRA and CRW, had significantly longer secondary retention times compared to the other MAD sites examined, and in these cases the total *E. coli* reduction rate was above the overall average value recorded for all 18 MAD sites. The best overall total *E. coli* reduction performance was recorded for site RMD, with a mean total *E. coli* reduction of $4.61 \log_{10}$. Site RMD approximated to the overall 18 site average primary digestion temperature, DS input and VS input values. By contrast, the primary *E. coli* input and primary actual HRT for site RMD were slightly larger than the overall average values. Site WRG had the poorest *E. coli* reduction performance among 18 sites examined with an average total *E. coli* reduction of $2.1 \log_{10}$. Nevertheless, the results showed that the smallest site mean log$_{10}$ reduction value at Thames Water conventional MAD sludge treatment facilities achieved the minimum 2 log$_{10}$ kill required for agricultural use specified by the Safe Sludge Matrix (ADAS, 2001).
<table>
<thead>
<tr>
<th>Site</th>
<th>Site</th>
<th>Feed Volume (m³ d⁻¹)</th>
<th>P_Nominal_HRT (d)</th>
<th>Actual_HRT (d)</th>
<th>Temperature (°C)</th>
<th>DS Input (%)</th>
<th>VS Input (%)</th>
<th>E. coli Input (log₁₀ cfu g⁻¹ DS)</th>
<th>S_HRT (d)</th>
<th>Secondary digestion type</th>
</tr>
</thead>
<tbody>
<tr>
<td>AYS</td>
<td>99.8</td>
<td>17.7</td>
<td>14.9</td>
<td>37.1</td>
<td>4.8</td>
<td>76.2</td>
<td>6.2</td>
<td>13.6</td>
<td>B</td>
<td></td>
</tr>
<tr>
<td>BAN</td>
<td>70.7</td>
<td>20.2</td>
<td>19.0</td>
<td>36.2</td>
<td>3.6</td>
<td>77.7</td>
<td>6.7</td>
<td>1.8</td>
<td>B</td>
<td></td>
</tr>
<tr>
<td>BAS</td>
<td>134.1</td>
<td>17.5</td>
<td>13.6</td>
<td>36.7</td>
<td>4.4</td>
<td>74.9</td>
<td>7.0</td>
<td>16.4</td>
<td>S</td>
<td></td>
</tr>
<tr>
<td>BRA</td>
<td>57.8</td>
<td>15.3</td>
<td>9.5</td>
<td>34.8</td>
<td>4.5</td>
<td>77.6</td>
<td>7.3</td>
<td>89.6</td>
<td>S</td>
<td></td>
</tr>
<tr>
<td>BSD</td>
<td>48.7</td>
<td>22.2</td>
<td>18.0</td>
<td>36.8</td>
<td>4.6</td>
<td>80.3</td>
<td>7.3</td>
<td>15.3</td>
<td>B</td>
<td></td>
</tr>
<tr>
<td>CAM</td>
<td>81.4</td>
<td>16.2</td>
<td>12.0</td>
<td>36.0</td>
<td>5.2</td>
<td>76.4</td>
<td>6.5</td>
<td>68.6</td>
<td>S</td>
<td></td>
</tr>
<tr>
<td>CRW</td>
<td>106.7</td>
<td>18.0</td>
<td>11.8</td>
<td>37.7</td>
<td>4.9</td>
<td>78.1</td>
<td>6.9</td>
<td>71.4</td>
<td>S</td>
<td></td>
</tr>
<tr>
<td>DEE</td>
<td>148.1</td>
<td>18.5</td>
<td>10.6</td>
<td>34.8</td>
<td>4.0</td>
<td>80.3</td>
<td>7.2</td>
<td>9.8</td>
<td>S</td>
<td></td>
</tr>
<tr>
<td>DID</td>
<td>57.1</td>
<td>21.7</td>
<td>17.9</td>
<td>35.8</td>
<td>4.1</td>
<td>75.0</td>
<td>6.8</td>
<td>6.4</td>
<td>B</td>
<td></td>
</tr>
<tr>
<td>EHY</td>
<td>88.5</td>
<td>16.8</td>
<td>14.7</td>
<td>35.6</td>
<td>4.6</td>
<td>72.1</td>
<td>6.5</td>
<td>6.7</td>
<td>S</td>
<td></td>
</tr>
<tr>
<td>HOG</td>
<td>235.6</td>
<td>14.7</td>
<td>7.6</td>
<td>35.2</td>
<td>4.7</td>
<td>75.8</td>
<td>7.0</td>
<td>9.2</td>
<td>S</td>
<td></td>
</tr>
<tr>
<td>LOR</td>
<td>186.5</td>
<td>14.7</td>
<td>8.4</td>
<td>35.8</td>
<td>4.1</td>
<td>82.2</td>
<td>7.4</td>
<td>13.4</td>
<td>S</td>
<td></td>
</tr>
<tr>
<td>MAP</td>
<td>172.3</td>
<td>20.6</td>
<td>9.0</td>
<td>35.9</td>
<td>4.6</td>
<td>77.3</td>
<td>7.3</td>
<td>8.0</td>
<td>B</td>
<td></td>
</tr>
<tr>
<td>OXF</td>
<td>134.4</td>
<td>22.2</td>
<td>15.9</td>
<td>35.0</td>
<td>4.0</td>
<td>73.6</td>
<td>6.7</td>
<td>4.0</td>
<td>B</td>
<td></td>
</tr>
<tr>
<td>RMD</td>
<td>173.9</td>
<td>21.7</td>
<td>14.2</td>
<td>36.4</td>
<td>4.5</td>
<td>74.5</td>
<td>7.3</td>
<td>11.5</td>
<td>B</td>
<td></td>
</tr>
<tr>
<td>SLO</td>
<td>131.8</td>
<td>17.0</td>
<td>9.7</td>
<td>35.9</td>
<td>3.3</td>
<td>78.5</td>
<td>7.4</td>
<td>10.2</td>
<td>S</td>
<td></td>
</tr>
<tr>
<td>SWN</td>
<td>136.9</td>
<td>12.8</td>
<td>9.6</td>
<td>36.1</td>
<td>4.0</td>
<td>73.6</td>
<td>6.7</td>
<td>3.3</td>
<td>B</td>
<td></td>
</tr>
<tr>
<td>WRG</td>
<td>97.1</td>
<td>25.3</td>
<td>12.6</td>
<td>36.9</td>
<td>4.9</td>
<td>77.3</td>
<td>6.6</td>
<td>4.4</td>
<td>B</td>
<td></td>
</tr>
<tr>
<td>Overall</td>
<td>120.1</td>
<td>18.5</td>
<td>12.7</td>
<td>36.0</td>
<td>4.4</td>
<td>76.7</td>
<td>6.9</td>
<td>20.2</td>
<td>/</td>
<td></td>
</tr>
</tbody>
</table>

P: Primary; S: Secondary; DS: Dry solids; VS: Volatile solids; HRT: Hydraulic retention time; cfu: Colony forming units; d: Days; B: Batch; S: Series
<table>
<thead>
<tr>
<th>Site</th>
<th>P_DS Output (%)</th>
<th>P_VS Output (%)</th>
<th>P_E. coli Output (log&lt;sub&gt;10&lt;/sub&gt; cfu g&lt;sup&gt;-1&lt;/sup&gt; DS)</th>
<th>F_E. coli Output (log&lt;sub&gt;10&lt;/sub&gt; cfu g&lt;sup&gt;-1&lt;/sup&gt; DS)</th>
<th>P_VSR (%)</th>
<th>P_log&lt;sub&gt;10&lt;/sub&gt; kill (log&lt;sub&gt;10&lt;/sub&gt; cfu g&lt;sup&gt;-1&lt;/sup&gt; DS)</th>
<th>S_log&lt;sub&gt;10&lt;/sub&gt; kill (log&lt;sub&gt;10&lt;/sub&gt; cfu g&lt;sup&gt;-1&lt;/sup&gt; DS)</th>
<th>T_log&lt;sub&gt;10&lt;/sub&gt; kill (log&lt;sub&gt;10&lt;/sub&gt; cfu g&lt;sup&gt;-1&lt;/sup&gt; DS)</th>
</tr>
</thead>
<tbody>
<tr>
<td>AYS</td>
<td>3.0</td>
<td>64.2</td>
<td>5.5</td>
<td>3.9</td>
<td>43.4</td>
<td>0.7</td>
<td>1.5</td>
<td>2.4</td>
</tr>
<tr>
<td>BAN</td>
<td>2.2</td>
<td>66.4</td>
<td>5.4</td>
<td>4.5</td>
<td>37.2</td>
<td>1.3</td>
<td>0.9</td>
<td>2.3</td>
</tr>
<tr>
<td>BAS</td>
<td>3.5</td>
<td>61.6</td>
<td>4.7</td>
<td>3.6</td>
<td>44.9</td>
<td>2.7</td>
<td>1.4</td>
<td>3.6</td>
</tr>
<tr>
<td>BRA</td>
<td>4.1</td>
<td>61.2</td>
<td>5.4</td>
<td>4.2</td>
<td>55.9</td>
<td>1.4</td>
<td>1.0</td>
<td>3.0</td>
</tr>
<tr>
<td>BSD</td>
<td>2.7</td>
<td>67.8</td>
<td>5.5</td>
<td>4.7</td>
<td>49.7</td>
<td>1.8</td>
<td>0.8</td>
<td>2.6</td>
</tr>
<tr>
<td>CAM</td>
<td>3.2</td>
<td>65.0</td>
<td>6.0</td>
<td>3.2</td>
<td>41.4</td>
<td>0.5</td>
<td>2.7</td>
<td>3.4</td>
</tr>
<tr>
<td>CRW</td>
<td>2.9</td>
<td>63.2</td>
<td>5.4</td>
<td>4.1</td>
<td>52.8</td>
<td>1.5</td>
<td>1.6</td>
<td>3.0</td>
</tr>
<tr>
<td>DEE</td>
<td>2.9</td>
<td>66.7</td>
<td>4.1</td>
<td>4.1</td>
<td>49.9</td>
<td>3.0</td>
<td>0.0</td>
<td>3.0</td>
</tr>
<tr>
<td>DID</td>
<td>2.7</td>
<td>64.1</td>
<td>5.5</td>
<td>3.9</td>
<td>39.0</td>
<td>1.4</td>
<td>1.6</td>
<td>3.0</td>
</tr>
<tr>
<td>EHY</td>
<td>3.6</td>
<td>63.1</td>
<td>4.8</td>
<td>4.3</td>
<td>32.3</td>
<td>1.7</td>
<td>0.5</td>
<td>2.1</td>
</tr>
<tr>
<td>HOG</td>
<td>2.8</td>
<td>63.9</td>
<td>5.5</td>
<td>4.4</td>
<td>44.4</td>
<td>1.4</td>
<td>1.1</td>
<td>2.5</td>
</tr>
<tr>
<td>LOR</td>
<td>2.4</td>
<td>66.6</td>
<td>4.7</td>
<td>4.7</td>
<td>57.3</td>
<td>2.8</td>
<td>0.1</td>
<td>2.7</td>
</tr>
<tr>
<td>MAP</td>
<td>2.8</td>
<td>64.3</td>
<td>5.8</td>
<td>3.7</td>
<td>46.3</td>
<td>1.4</td>
<td>2.1</td>
<td>3.5</td>
</tr>
<tr>
<td>OXF</td>
<td>3.0</td>
<td>64.3</td>
<td>5.5</td>
<td>4.3</td>
<td>37.0</td>
<td>1.3</td>
<td>1.3</td>
<td>2.3</td>
</tr>
<tr>
<td>RMD</td>
<td>2.9</td>
<td>61.4</td>
<td>4.9</td>
<td>2.8</td>
<td>46.3</td>
<td>2.2</td>
<td>2.1</td>
<td>4.6</td>
</tr>
<tr>
<td>SLO</td>
<td>2.1</td>
<td>66.0</td>
<td>5.9</td>
<td>4.6</td>
<td>46.1</td>
<td>1.5</td>
<td>1.3</td>
<td>2.8</td>
</tr>
<tr>
<td>SWN</td>
<td>4.0</td>
<td>61.3</td>
<td>5.4</td>
<td>3.7</td>
<td>44.9</td>
<td>1.3</td>
<td>1.7</td>
<td>2.9</td>
</tr>
<tr>
<td>WRG</td>
<td>2.9</td>
<td>64.6</td>
<td>4.5</td>
<td>4.1</td>
<td>44.0</td>
<td>1.9</td>
<td>0.3</td>
<td>2.1</td>
</tr>
<tr>
<td>Overall</td>
<td>3.0</td>
<td>64.2</td>
<td>5.3</td>
<td>4.0</td>
<td>45.2</td>
<td>1.7</td>
<td>1.2</td>
<td>2.9</td>
</tr>
</tbody>
</table>

P: Primary; S: Secondary; F: Final; T: Total; DS: Dry solids; VS: Volatile solids; VSR: Volatile solids reduction; cfu: Colony forming units
Patterns in *E. coli* reduction with time (year)

ANOVA and the plot of annual mean data for *E. coli* reduction performance indicated the overall process removal was consistent and not significantly different between years (*P*=0.161), with an overall mean value equivalent to 2.9 log<sub>10</sub> (Figure 2). However, a statistically significant effect of time was detected for primary and secondary digestion when these process steps were considered separately (*P*=0.002 and <0.001, respectively). Thus, *E. coli* removal by secondary digestion increased significantly with time with decreasing primary removal (Figure 2). Thus, secondary digestion maintained the overall *E. coli* reduction performance, which, on average achieved compliance with the Safe Sludge Matrix, irrespective of the variations observed in removal rate for primary digestion.

![Graph showing annual primary, secondary, and total E. coli removals](image)

**Fig. 2** Mean annual primary, secondary and total *E. coli* removals

Effect of secondary digestion type on *E. coli* reduction

ANOVA indicated there was a statistically significant effect (*P*<0.001) of secondary digestion type on the *E. coli* reduction rate across this process step (Figure 3). The results showed that the mean *E. coli* reduction rate was significantly increased from log<sub>10</sub> 0.99 +/-0.054 (standard error) in series to 1.34 +/-0.049 in batch processes. However, there was no statistically significant effect (*P*>0.05) of secondary digestion type on total *E. coli* removal across the overall digestion process.
Seasonal effects on *E. coli* reduction

Mesophilic anaerobic digestion is a managed process operated above ambient temperatures. Nevertheless, seasonal differences were apparent in the *E. coli* reduction rate (Figure 4) and showed that secondary and overall removals increased significantly during the summer, spring and autumn compared to the cooler winter season, suggesting that higher ambient temperatures increased the overall *E. coli* removal. However, no statistically significant effect (\(P>0.05\)) of season was detected on pathogen reduction by primary MAD (Figure 5). Secondary digesters are unheated tanks and are therefore influenced by ambient temperature conditions, in contrast to heated primary digesters. Hence, seasonal changes in ambient temperature had a greater influence on secondary *E. coli* removal compared to primary removal and, consequently, also on the overall total removal rate. Further scrutiny of the data showed that this response was controlled by the secondary digestion conditions rather than any apparent variation in seasonal inputs of *E. coli* numbers, as these remained consistent and were not significantly different (\(P>0.05\)) between seasons (The average *E. coli* input for spring, summer, autumn and winter were 6.93, 6.94, 6.99 and 6.82 log\(_{10}\) cfu g\(^{-1}\) DS, respectively).
Fig. 4 Mean primary, secondary and total *E. coli* reduction in relation to season
Multiple regression modelling

**Secondary E. coli reduction**

Backward and forward predictor selection methods generated equivalent multiple regression models of *E. coli* reduction by secondary digestion (note: the significance threshold used for parameter selection was $P<0.1$). The final model had an adjusted $R^2$ of 72.7%. Primary *E. coli* log$_{10}$ reduction and primary digestion temperature were both identified as statistically significant explanatory variables (both $P<0.001$) of secondary *E. coli* reduction. Primary *E. coli* log$_{10}$ reduction explained 4.9% of the total variance in the data and had a negative coefficient. As suggested by the mean annual primary and secondary *E. coli* removals presented in Figure 2, secondary *E. coli* reduction performance was negatively correlated with the degree of removal achieved in the primary digestion process. Primary digestion temperature explained 2.8% of the total variance in *E. coli* numbers in secondary digested sludge and the regression coefficient describing the decay response was positive in this case.

This evidence suggested that primary digestion conditions (specifically temperature)
sublethally damage *E. coli*, increasing subsequent die-off of the indicator bacteria in secondary digestion.

**Total *E. coli* reduction**

The Forward selection multiple regression modelling procedure (with *P*<0.1 as the significance threshold for parameter selection) identified three statistically significant covariates (Primary *E. coli* Input, Primary Temperature, and Secondary Retention Time) that explained the Total *E. coli* reduction across the combined primary and secondary digestion processes. However, the diagnostic test for colinearity showed that the tolerance of secondary retention time was <0.1, indicating a multicollinearity conflict within the model. Multicollinearity occurs when two or more explanatory factors in a multiple regression model are not only correlated to the dependent variable, but also to each other (Pallant, 2013). Multicollinearity thus reduces the stability of estimates, increases the standard error of the coefficients, and most importantly, reduces the ability to identify the effects of correlated predictors (Mansfield and Helms, 1982).

The HACCP management system for microbiological quality of sewage sludge for agricultural use in Thames Water specifies the minimum secondary retention time depending on the primary digestion temperature. Thus, if the primary temperature falls below the HACCP set point, the secondary retention time is increased to produce treated sludge, compliant with the microbiological criteria (2 log₁₀ reduction of *E. coli*) specified by the Safe Sludge Matrix (ADAS, 2001). Therefore, as would be expected, the primary digestion temperature and secondary digestion retention time were negatively correlated with each other. However, this multicollinearity behaviour had no influence on the overall regression model or the associated statistics, such as adjusted *R*² (0.73) and *P* (*P*(Temperature)<0.001, *P*(Primary *E. coli* Input)<0.001, *P*(Secondary Retention Time)=0.001) values. Furthermore, the model provided stable predictions suggesting that these explanatory variables were correlated in a similar pattern. The multiple regression analysis of operational information held in the Thames Water databases showed that, in practice, generally, for every 1°C rise in primary digestion temperature, the secondary retention time was reduced on average by 0.6 day. Consequently, primary temperature and secondary retention were both important explanatory variables of total *E. coli* reduction, but their relative contributions could not be accurately quantified. This model could be useful to predict total pathogen reduction if primary temperature and secondary retention time are correlated in the same pattern as indicated by the original dataset. However, the partial *η*² and *P* values for primary temperature and secondary retention time were, respectively: 9.4 %, <0.001 and 2.3 %, 0.001. Consequently, primary temperature explained a larger proportion of
the total variance in *E. coli* numbers compared to secondary retention time. Therefore, secondary retention time was taken out of the model in the subsequent analysis to obtain a more accurate prediction and assessment of individual predictor effects on total *E. coli* reduction.

The multiple regression model showed that the statistically significant, independent factors controlling total *E. coli* removal across the combined primary and secondary digestion process were: (1) input numbers of *E. coli* in the raw feed sludge, and (2) primary digestion temperature. Thus, on average across all the conventional MAD treatment sites in Thames Water, a 2 log_{10} reduction in *E. coli* was achieved when the input numbers of *E. coli* in raw sludge were ≥5 log_{10} cfu g^{-1} DS (Figure 6) (the *P* value for this parameter in the model was <0.001 and the η^2 was 21.3 %), irrespective of any other operational parameters within the range of conditions observed at Thames Water sites. The model also detected a statistically significant effect (*P*<0.001, η^2=3.0%) of primary digestion temperature and showed that, on average, the overall log_{10} reduction in *E. coli* was raised by 0.0124 per 0.1°C lift in primary digestion temperature (note: this was linked to the increased rate of decay in secondary digestion with increasing primary digestion temperature). Therefore, the statistical analysis of monitoring data from Thames Water sludge treatment sites indicated that warmer conditions were beneficial for pathogen removal by conventional MAD and that increasing the digestion temperature from 35°C to 40°C would improve the overall average log_{10} *E. coli* removal by 0.62.

One operational application of the model was to estimate the minimum primary MAD operating temperature necessary to achieve compliance with the minimum 2 log_{10} removal standard for *E. coli* for each of the WWTP considered in the statistical analysis. The minimum operating temperatures estimated by this analysis were in the range 14.6-36.9°C based on average *E. coli* input values for each site.

Whilst statistically significant, the independent variables in the multiple regression model only accounted for 24.3% of the total variance observed in the *E. coli* removal data. Indeed, a large proportion of the statistical variation (42.1%) observed in *E. coli* reduction was attributable to major differences between sites and year, which could not be explained by the measured operational and sludge digestion parameters. Consequently, there is a considerable degree of uncertainty about other important process factors and the mechanisms controlling pathogen inactivation, for instance, these may include, but are not limited to: digester mixing, asset age, sludge bypass and microbiological effects of the dewatering process.
Fig. 6 Total *E. coli* reduction (log$_{10}$) in relation to primary MAD temperature (°C) ($P<0.001$, $\eta^2=3.0\%$) and input *E. coli* numbers in raw sludge (log$_{10}$ cfu g$^{-1}$ DS) ($P<0.001$, $\eta^2=21.3\%$)

**Discussion**

Secondary digestion type

Extended process retention times are usually reported to increase pathogen inactivation during sludge treatment across different temperature ranges (Baert et al., 2010; Massé and Topp, 2011; Pandey and Soupir, 2011). However, the statistical analysis of data collected from full-scale sludge treatment sites showed that secondary digestion processes operating in batch mode, with shorter average retention times of 7.6 d, had larger *E. coli* reductions (mean = 1.34 log$_{10}$) compared to series operation (mean = 0.99 log$_{10}$) with longer mean hydraulic retention equivalent to 32.4 d. Therefore, factors other than retention time were important in defining the effect of secondary digester type on *E. coli* reduction. For example, the greater variability in HRTs of mixed sludge in series systems compared to batch operation, where all the sludge receives an equivalent minimum retention period, could potentially explain the lower average *E. coli* reduction rate for series sites. Secondary HRTs were calculated based on the total secondary digester volume. However, information on the effective volume of secondary digesters was incomplete and it is possible that the HRT of series sites was
overestimated (especially for sites: BRA, CAM and CRW with apparently very long secondary 
HRT), since the effective volume is smaller than the overall total volume. Finally, an ANOVA 
test of secondary \textit{E. coli} inputs for batch and series sites indicated that there was a statistically 
significant difference \((P=0.004)\) between the two groups and sites operating batch secondary 
digesters had slightly more \textit{E. coli} in the input sludge (Table 3). This could further contribute to 
the larger removal observed for batch digesters if the behaviour followed the same patterns 
observed in primary digestion between the extent of \textit{E. coli} input and decay (Gao, 2013). However, given the small difference in input numbers the effect is only likely to be marginal 
compared to the other operational management factors responsible for pathogen decay in 
these secondary digester systems.

Batch operated sites had better secondary \textit{E. coli} reduction performance compared to those 
operated in series, nevertheless, there was no significant effect \((P>0.05)\) of secondary 
digester type on overall \textit{E. coli} removal. Thus, secondary digesters have a similar action by 
balancing the overall \textit{E. coli} reduction performance across the entire process, irrespective of 
the variations observed in the removal rate in primary digestion. Thus, increased reduction 
was observed in the secondary digestion stage (irrespective of the type) when removal rates 
were reduced in the primary stage and \textit{vice versa}, performed in a generally consistent manner.
The patterns observed in primary, secondary and total \textit{E. coli} reduction data indicated that the 
secondary digestion stage provided a final assurance barrier to ensure total \textit{E. coli} reductions 
were compliant with the Safe Sludge Matrix for agricultural utilisation of sewage sludge (ADAS, 
2001).

<table>
<thead>
<tr>
<th>Type</th>
<th>Secondary \textit{E. coli} input ((\log_{10} \text{cfu g}^{-1} \text{DS}))</th>
<th>Secondary \textit{log}<em>{10} reduction ((\log</em>{10} \text{cfu g}^{-1} \text{DS}))</th>
</tr>
</thead>
<tbody>
<tr>
<td>Batch</td>
<td>5.31</td>
<td>1.34</td>
</tr>
<tr>
<td>Series</td>
<td>5.14</td>
<td>0.98</td>
</tr>
</tbody>
</table>

Digestion and ambient temperature

The statistical analysis showed that primary digestion temperature had no statistically 
significant effect on \textit{E. coli} reduction across the primary digestion stage for the range of 
process temperatures applied at Thames Water operational MAD sites (Gao, 2013). This
agrees with other evidence that mesophilic conditions have only a marginal direct impact on pathogen reduction during primary mesophilic anaerobic digestion of sewage sludge (Smith et al., 2005; Lang and Smith, 2008). Nevertheless, seasonal temperature patterns in temperate climates, such as the UK, may influence the removal of E. coli during unheated secondary digestion as the reduction rate increased significantly during warmer ambient conditions compared to cooler seasonal periods (Fig. 4). An analysis of historic data for a group of United Utility’s secondary digesters also showed a similar seasonal influence on secondary E. coli removals (Le et al., 2002). The effect of temperature on the removal rate by secondary digestion increased from 0.5 to 1.0 log10 as secondary digestion temperature was raised from 10°C to 25°C (Le et al., 2002). Pathogen inactivation in MAD (35°C) is only weakly influenced by temperature, and is mainly attributable to other complex and interacting physical and biochemical factors such as the substrate availability, microbial competition and inhibition (Smith et al., 2005). However, the multiple regression analysis showed that primary digestion temperature had an indirect effect on final E. coli inactivation, which increased as the temperature of primary digestion was raised. This evidence suggested that the primary digestion conditions, and specifically temperature, appear to sublethally damage E. coli, increasing decay in secondary digestion, which is also a function of rising seasonal ambient temperature, and consequently across the entire process. A practical application of the multiple regression model was to predict the minimum primary digestion temperature required at specific sites to provide at least a 2 log10 removal of E. coli (data not shown) and the minimum temperatures estimated by the model were in the range: 14.6 – 36.9°C. This analysis showed that all of the conventional MAD treatment sites were operating at or above the necessary minimum threshold value as actual operating temperatures were in the range: 34.8 - 37.1°C.

Hydraulic retention time

No statistically significant effect of primary HRT on secondary or total E. coli reduction was detected for the range of HRT periods applied at the operational primary MAD sludge treatment sites (15 – 25 d, Table 1) considered in the statistical analysis. The retention time of a sludge treatment process is an important factor controlling pathogen decay (Smith et al., 2005), but this depends upon the operating range in practice. Thus, the data for full-scale sludge digestion processes were consistent with the comparison of E. coli reduction during MAD at three different retention times (11 d, 16 d, 25 d) by Chen et al. (2012), which showed that the removal rate significantly improved as the retention time increased from 11 to 16 d,
and 11 to 25 d, but not from 16 to 25 d. Similarly, Lee et al. (1989) found no significant effect of HRT between 10 and 20 d on pathogen reduction during MAD of sewage sludge. Consequently, the results reported here suggest that the HRT of primary digestion could be reduced to a minimum period of 15 d without impacting the \( E. coli \) removal efficiency of the sludge digestion process.

Although secondary retention time was identified as a significant explanatory variable of \( E. coli \) removal, multicollinearity between the secondary retention time and primary temperature was detected by the statistical analysis. In general, a 1 °C rise in primary digestion temperature was associated with a reduction in the secondary retention time equivalent to 0.6 days, and this relationship was a direct response to the HACCP management plan for microbiological quality of sludge implemented by Thames Water. Consequently, secondary retention time was removed from the set of predictor variables in the multiple regression model to avoid problems of multicollinearity.

Primary digestion temperature and primary \( E. coli \) input were two highly significant explanatory variables (both \( P < 0.001 \)) of total \( E. coli \) reduction in the multiple regression model. However, major differences remain between sites that cannot be reconciled or explained by the principal process control parameters, such as temperature or HRT (no statistically significant effect of HRT was detected at the range of HRTs applied at Thames Water sites), often assumed to be the major factors responsible for \( E. coli \) removal. For instance, the main effect of Site and the interaction effect with Year explained 42.1% of the total variance in total \( E. coli \) reduction, compared to 24.3% for \( E. coli \) Input and Primary Temperature. Nevertheless, despite the uncertainty in the mechanisms and operational factors controlling \( E. coli \) removal, conventional MAD processes with a standard configuration of primary and secondary liquid digestion readily achieve the microbiological quality criteria for agricultural use of sludge.

**Conclusions**

1. A series of statistical tests and multiple regression modelling techniques were applied to large process monitoring data sets collected for conventional MAD sludge treatment sites in the Thames Water region to quantify the significance of operational parameters controlling the removal of \( E. coli \) by conventional sewage sludge digestion processes.
2. One of the major factors responsible for increasing *E. coli* decay during MAD of sewage sludge was outside the control of the process operator and conditions and was related to the size of the bacterial population in the influent raw sludge.

3. Primary digestion conditions (specifically temperature) had no significant effect on *E. coli* numbers across the primary digestion stage of the process, but increased the decay during secondary digestion and consequently across the entire process. This may be explained because *E. coli* were sublethally damaged during primary digestion and the susceptibility to decay during secondary digestion increased with the temperature of primary digestion (and also with increasing seasonal temperature – see below). The average increase in overall log_{10} *E. coli* removal was equivalent to 0.0124 for each 0.1°C lift in primary digestion temperature. A practical application of the developed multiple regression model for *E. coli* removal was to predict the minimum primary digestion temperatures necessary to achieve an overall 2 log_{10} reduction of *E. coli* in sludge for specific conventional MAD treatment sites. This showed that all the sludge treatment sites were operating at or above the minimum temperature necessary to meet the conventional microbiological criteria for agricultural use.

4. Secondary digester type had a significant effect on secondary *E. coli* removal; batch operation had shorter retention times but was more effective at reducing *E. coli* numbers than series operation with longer average retention times. Nevertheless, there was no significant effect of secondary digester type on total *E. coli* removal and secondary digesters balanced the overall *E. coli* reduction across the process, irrespective of the removal efficiency in primary digestion.

5. A statistically significant effect of season was detected on secondary and total *E. coli* removal, and decay was increased by warmer ambient temperatures compared to cooler winter conditions.

6. Although major uncertainties remain about the factors and mechanisms responsible for pathogen inactivation by MAD processes between different sewage sludge treatment sites, a combination of primary and secondary digestion stages consistently provided treatment conditions that assured compliance with conventional microbiological criteria for agricultural use of sewage sludge.
Acknowledgements

The authors thank Aurelien Perrault, Nick Mills and Peter Winter and all other staff at Thames Water for their expert assistance with compiling and quality assuring the database information.
References


Word count: 5200