A Novel Method to Determine Buffer Strip Effectiveness on Deep Soils

Marius Heinen,* Ignatius G. A. M. Noij, Hanneke I. M. Heesmans, Jan Willem van Groenigen, Piet Groenendijk, and Jac T. N. M. Thissen

Unfertilized buffer strips (BS) generally improve surface water quality. High buffer strip effectiveness (BSE) has been reported for sloping shallow aquifers, but experimental data for plain landscapes with deeply permeable soils is lacking. We tested a novel method to determine BSE on a 20-m-deep, permeable sandy soil. Discharge from soil to ditch was temporarily collected in an in-stream reservoir to measure its quantity and quality, for both a BS and a reference (REF) treatment. Treatments were replicated once for the first, and three times for the next three leaching seasons. No significant BSE was obtained for nitrogen and phosphorus species in the reservoirs. Additionally, water samples were taken from the upper groundwater below the treatments. The effect of BS for nitrate was much bigger in upper groundwater than in the reservoirs that also collected groundwater from greater depths that were not influenced by the treatments. We conclude that measuring changes in upper groundwater to assess BSE is only valid under specific hydrogeological conditions. We propose an alternative experimental set-up for future research, including extra measurements before installing the BS and REF treatments to deal with spatial and temporal variability. The use of such data as covariates will increase the power of statistical tests by decreasing between-reservoir variability.

Improving surface water quality in rural areas with intensive agriculture is a major challenge. In the Netherlands, agriculture has not been able to keep up with the other economic sectors in terms of nutrient load reduction (Hoogervorst, 2009). Rural nonpoint nutrient sources from agricultural fields are more difficult to abate than the predominant point sources from other sectors in the main surface waters. To reduce nutrient loads from agriculture to surface water throughout the European Union (EU), the Nitrates Directive (EU, 1991) and the Water Framework Directive (EU, 2000) have been implemented.

In many countries, both within and outside the EU, riparian buffer zones or unfertilized buffer strips (BS) are accepted as a mitigation measure against nutrient loads, based on extensive evidence for their effectiveness (e.g., Barling and Moore, 1994; Dorioz et al., 2006; Dosskey, 2002; Mayer et al., 2005, 2007; Parkyn, 2004; Polyakov et al., 2005; Wenger, 1999). However, reported data for buffer strip effectiveness (BSE) is highly variable, ranging from below zero up to almost 100%, depending on the nutrient considered, input load, local hydrogeological conditions, BS width, BS vegetation, BS maintenance (removal of sediment and biomass), and time period after installation. According to Dosskey (2002), BSE tends to be overestimated due to ideal experimental circumstances and often a lack of a proper reference treatment.

Landscape and hydrology are recognized as important factors affecting BSE (Burt et al., 2002; Hill, 1996; Polyakov et al., 2005; Puckett, 2004; Sabater et al., 2003; Vidon and Hill, 2004). In particular, the distribution of discharge between surface runoff, subsurface runoff, and deeper groundwater plays an important role (Dorioz et al., 2006; Dosskey, 2002; Hoffmann et al., 2009; Mayer et al., 2005, 2007; Ranalli and Macalady, 2010; Rassam et al., 2008; Seibert et al., 2009). Surface runoff and shallow subsurface flow will travel through the cropped soil surface and the rooted, (micro)biologically active layer of soil.
the BS, thus allowing retention processes in the BS to occur, while deeper flow paths (including drain pipes) bypass the BS below the active layer.

For phosphorus (P) and solids, surface runoff is the key transport route that determines BSE (Dorioz et al., 2006; Hoffmann et al., 2009). For nitrate (NO$_3$–N), highest BSE was found with shallow lateral groundwater flow (Balestrini et al., 2008; Borin and Bigon, 2002; Dhondt et al., 2002, 2006; Hefting, 2003; Hoffmann et al., 2009; Schilling et al., 2006; Young and Briggs, 2005).

In the majority of BS studies, the depth of groundwater transport is limited by impermeable or less-permeable subsoil layers, and lateral surface and subsurface runoff is often driven by slope. According to Hill (1996), the ideal depth of the confining layer below the aquifer lies between 1 and 3 m below soil surface (bss). Situations with flat topography and predominantly discharge of deeper groundwater to surface water, such as in river deltas, have not been covered in the literature (Doskey, 2002). Phillips et al. (1993) studied a several meters thick aquifer, but with dunes of a few meters high. For the Po Valley, high BSE for nitrate was found (69–99%; Borin and Bigon, 2002), which was partly attributed to the 1-m-wide tree border between the 5-m grass BS and the stream. More important, a reference treatment was lacking, leading to possible overestimation of BSE (Doskey, 2002). The high groundwater table in the field border made denitrification highly likely to occur regardless of the presence of a BS.

Several methods to assess BSE have been proposed in the literature. There are two major approaches: (i) comparison of input into the BS with output from the BS (e.g., Borin and Bigon, 2002; Kuusemets and Mander, 1999; Sabater et al., 2003; Schmitt et al., 1999; Young and Briggs, 2005) and (ii) comparison of output from a field with BS with output from a reference field (e.g., Borin et al., 2005; Dillaha et al., 1989; Duchemin and Hogue, 2009; Uusi-Kämppä and Jauhiainen, 2010). However, a method to assess BSE on deeply permeable flat soils where a mixture of shallow and deep flow paths discharges to a ditch is still lacking. Concentration measurements cannot be restricted to one or two flow paths under these conditions. In other words, it is hard to tell what groundwater depth is representative for total lateral flow.

The main objective of this study was to measure BSE with a novel method that accounts for all discharge routes in a deeply permeable soil. We compared a 5-m-wide unfertilized grassed BS with a fertilized maize (Zea mays L.) reference treatment to grow forage maize. Prior to that, it was a meadow for horses.

### Materials and Methods

To measure BSE on a deeply permeable soil, we installed in-stream reservoirs in the ditch to collect the discharging water from the adjacent soil, to determine its quantity and quality. Treatments included both a 5-m-wide unfertilized BS and a reference treatment (REF) with fertilization up to the ditch bank. Reservoir observations were used for the assessment of BSE, using accumulated loads, flow-weighted concentrations, and statistical analysis of individual samplings.

To compare BSE based on reservoirs with BSE based on upper groundwater, upper groundwater concentration was monitored underneath both treatments.

#### Site Description

The experimental site was located in the eastern part of the Netherlands, near Beltrum (52°04′56″ N, 06°32′11″ E; 17 m above sea level). The landscape in the area undulates gently with a slope <1%. The sand layer is 20 m thick and rests on an impermeable base. The mean highest groundwater level is 0.3 to 0.5 m bss, and mean lowest groundwater is 1.2 to 1.6 m bss. The soil is a gleic podzol (FAO, 2002) in sand of periglacial aeolian origin (0–5 m bss). The Ap horizon extends to a depth of 0.3 m, followed by a Bh horizon down to 0.5 m bss, and the C horizon starts at 0.5 m bss. Some selected soil properties are presented in Table 1. The lower 15 m of the aquifer is sand of fluvial origin. The soil is layered, resulting in anisotropy in hydraulic conductivity. The field is bordered by ditches (bottom at 1.3 m bss), which function as drains during winter (the leaching season); during summer, they fall dry. Infiltration from the ditch into the soil is negligible. Since 2000 and during the experiment, the field has been used for growing forage maize. Prior to that, it was a meadow for horses.

#### Experimental Set-up and Treatments

**In-Stream Collector Reservoirs**

In February 2006, the first replicate (A) with two treatments was installed (Fig. 1). Along the ditch, a 5-m-wide (from the

### Table 1. Average soil properties for four horizons (± standard deviation based on three replicates): organic matter content (OM), soil texture on mineral basis, dry bulk density (ρ), hydraulic conductivity at saturation ($K_{sat}$), phosphorus saturation degree (PSD), and pH (water).

<table>
<thead>
<tr>
<th>Horizon</th>
<th>Depth (m)</th>
<th>OM (%)</th>
<th>Texture</th>
<th>$&lt;2 \mu m$</th>
<th>$&lt;16 \mu m$</th>
<th>$&lt;50 \mu m$</th>
<th>$&gt;50 \mu m$</th>
<th>$\rho$ (g cm$^{-3}$)</th>
<th>$K_{sat}$ (cm h$^{-1}$)</th>
<th>PSD (mol kg$^{-1}$)</th>
<th>pH-H$_2$O</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ap</td>
<td>0–0.3</td>
<td>0.054 ± 0.007</td>
<td>0.032 ± 0.016</td>
<td>0.050 ± 0.013</td>
<td>0.081 ± 0.014</td>
<td>0.070 ± 0.014</td>
<td>0.089 ± 0.014</td>
<td>1.174 ± 0.051</td>
<td>0.842 ± 0.053</td>
<td>0.507 ± 0.065</td>
<td>6.0 ± 0.2</td>
</tr>
<tr>
<td>Bh</td>
<td>0.5–0.8</td>
<td>0.027 ± 0.014</td>
<td>0.034 ± 0.028</td>
<td>0.043 ± 0.030</td>
<td>0.068 ± 0.040</td>
<td>0.055 ± 0.014</td>
<td>0.073 ± 0.014</td>
<td>1.712 ± 0.053</td>
<td>0.357 ± 0.057</td>
<td>0.116 ± 0.060</td>
<td>5.8 ± 0.1</td>
</tr>
<tr>
<td>Cg</td>
<td>0.5–1.5</td>
<td>0.011 ± 0.009</td>
<td>0.023 ± 0.016</td>
<td>0.025 ± 0.016</td>
<td>0.049 ± 0.018</td>
<td>0.047 ± 0.018</td>
<td>0.052 ± 0.018</td>
<td>1.687 ± 0.053</td>
<td>0.547 ± 0.057</td>
<td>0.083 ± 0.042</td>
<td>5.7 ± 0.3</td>
</tr>
<tr>
<td>Cr</td>
<td>1.5+</td>
<td>0.004 ± 0.001</td>
<td>0.030 ± 0.008</td>
<td>0.030 ± 0.007</td>
<td>0.053 ± 0.010</td>
<td>0.047 ± 0.010</td>
<td>0.058 ± 0.010</td>
<td>0.206 ± 0.060</td>
<td>0.206 ± 0.060</td>
<td>0.54 ± 0.3</td>
<td></td>
</tr>
</tbody>
</table>

† $\rho$ and $K_{sat}$ were only determined for replicate A for the first 3 horizons.

‡ PSD = $P_{tot}/(0.5(Al_{sat} + Fe_{sat}))$; molar concentrations of P, aluminum (Al), and iron (Fe) in oxalate extract.
edge of the ditch bank into the field) and 15-m-long (along the ditch) grass BS was installed. Next to the BS, a 5-m-long reservoir (reaching to the center of the ditch) made of wooden walls was installed in the ditch to collect water flowing from the soil into the ditch (Fig. 1). Walls consisted of tongue-and-groove planks (0.045 m thick) that were driven down to 1.5 m below the bottom of the ditch. To prevent any leakage, we mounted an additional wall of composite wood boards and filled the space between the two walls with bentonite. Yearly, the reservoir was pumped empty for visual inspection of leakage through the walls. We did not observe any leakage.

Beside the BS, a reference strip (REF) with maize planted up to the ditch is in the back, the 5-m-wide buffer strip (BS) is in the middle, and a 3-m-wide access grass strip is in the front.

Fig. 1. (a) Lay-out (in-set is not at scale) and (b) photograph of the experimental site. In (b), the ditch with reservoirs is located at the right, the reference plot (REF) with maize planted up to the ditch is in the back, the 5-m-wide buffer strip (BS) is in the middle, and a 3-m-wide access grass strip is in the front.
The REF was treated like the rest of the field, including tillage, slurry application, fertilizers, and pest control, except for a small obligatory uncultivated strip of 0.5 m from the edge of the ditch bank. We registered fertilizer rates, measured crop yield and nutrient uptake, and calculated nutrient surpluses of the treatments strips (Table 2). The average significant difference in nutrient surplus between the treatments was 100 kg N ha\(^{-1}\) yr\(^{-1}\) and 26 kg P ha\(^{-1}\) yr\(^{-1}\).

The water level in the reservoir was maintained at the outside ditch level (maximum difference 0.01 m) by pumping out excess water. Discharge (\(Q\), m\(^3\)) from the reservoir was measured at the pump outlet with a flow meter and logged by a programmable data taker that activated an automatic sampler at fixed discharge amounts to take water samples from the reservoir (Fig. 1a). Samples need to be taken proportional to discharge for establishing loads or flow-averaged concentrations (e.g., de Vos, 2001; Rozemeijer et al., 2010). The sampling bottles were filled in five steps, each step corresponding to 300 L of discharge. Based on an estimated contributing area of 300 m\(^2\) per reservoir, a full bottle corresponded to 5 mm rainfall excess. Water samples were immediately stored in an on-site refrigerator (max. 4°C) and transported to the laboratory once a week (including partly filled bottles) to determine nutrient concentrations (\(C\), g m\(^{-3}\)); if no water sample was present, a sample from the reservoir was taken manually (if water was present). The actual number of samples differed between the leaching seasons and the treatments; on average 30 (± 6) samples per reservoir per leaching season were taken. Water samples were split into three subsamples after thorough mixing. The first subsample was not filtered and analyzed for total nitrogen (N\(_t\)) and total phosphorus (P\(_t\)) with a segmented flow analyzer (SFA) after persulfate-borate destruction (based on NEMI methods I-4650-03 and I-2650-03; www.nemi.gov). The second subsample was analyzed in the same way, but after filtering over 0.45 μm to measure total soluble N (N\(_s\)) and total soluble P (P\(_s\)). The third subsample was filtered over a 0.45-μm membrane and analyzed for NO\(_3\)–N (+NO\(_2\)\(_\)–N), NH\(_4\)–N, PO\(_4\)–P (all in 0.01 M CaCl\(_2\) with SFA), dissolved organic carbon (DOC) (SFA), and Cl (flow injection analyzer). Organic nitrogen concentrations N\(_{org}\) were calculated as N\(_{org}\) = N\(_s\) – NO\(_3\)–N – NH\(_4\)–N. Here, we focus on N\(_s\), P\(_s\), NO\(_3\)–N, PO\(_4\)–P and Cl only. Results for N\(_s\), P\(_s\), NH\(_4\)–N, N\(_{org}\), and DOC are given in the Supplemental Material.

Loads (\(L\), g) from soil to reservoirs were computed according to the following equation:

\[
L = \sum QC
\]

In Eq. [1], the summation sign refers to a period of either equal time or equal discharge. Dutch leaching seasons typically run from 1 October until 1 April, but actual start and end dates were used. Loads based on equal time periods were computed for the leaching period per reservoir. Loads based on equal discharge were computed for the least cumulative discharge observed for the paired reservoirs.

The flow-averaged leaching concentration, \(\bar{C}\) (g m\(^{-3}\)), was computed according to (e.g., Chaubey et al., 1994; 1995)

\[
\bar{C} = \frac{\sum QC}{\sum Q}
\]

### Suction Cups

Suction cups (polyester acrylate, porosity ~65%, pore diameter ~0.45 μm, inert to N and P) were permanently installed in a transect perpendicular to the ditch at five distances from the center of the ditch (2, 4, 6, 8, and 52 m), and at five depths bss, covering the range between mean highest and mean lowest groundwater levels (0.5, 0.75, 1.0, 1.25 and 1.50 m). For replicate A, two transects per treatment were installed (Fig. 1a); for replicates B and C, only one transect per treatment was installed. The first 2 samplings were performed in December 2006, 10 samplings in 2007 (2 for replicates B and C), 9 samplings in 2008, 9 samplings in 2009, and 2 samplings in 2010. The suction cup just below the groundwater level was sampled at every sampling date (upper groundwater). After April 2008, suction cups were sampled at all depths at every other sampling time. In June 2009, two extra series of suction cups were installed per treatment at 2 and 8 m from the center of the ditch at depths 2 and 3 m bss. The deep suction cups were sampled in June, October, and December 2009 and March and April 2010. Water samples from the suction cups were analyzed for N\(_s\), NO\(_3\)–N, NH\(_4\)–N and PO\(_4\)–P, DOC, and Cl. Data for NH\(_4\)–N, DOC and N\(_{org}\) are presented in the Supplemental Material. Groundwater concentrations are denoted as \(C_{gw}\).

### Hydrology

Rainfall was measured hourly with a resolution of 0.2 mm. Based on an analysis of measured groundwater heads in observation wells from the DINO database (Data and Information of the Dutch Subsurface, TNO-NITG; www.TNO.nl), the major direction of regional groundwater flow is expected to be nearly perpendicular to our experimental ditch. We measured the local

### Table 2. Nitrogen (N) and phosphorus (P) fertilizer rate, uptake, and surplus (fertilizer rate – uptake) per individual growing season for the reference (REF) and buffer strip (BS) treatments. Averages (± standard deviation) for the three replicates in 2007–2009.

<table>
<thead>
<tr>
<th>Year</th>
<th>REF N</th>
<th>BS N</th>
<th>REF P</th>
<th>BS P</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Fertilizer†</td>
<td>Uptake‡</td>
<td>Surplus</td>
<td>Fertilizer†</td>
</tr>
<tr>
<td>2006</td>
<td>137</td>
<td>158</td>
<td>–20</td>
<td>37</td>
</tr>
<tr>
<td>2007</td>
<td>155</td>
<td>84 ± 22</td>
<td>72 ± 22</td>
<td>40</td>
</tr>
<tr>
<td>2008</td>
<td>185</td>
<td>187 ± 5</td>
<td>–2 ± 5</td>
<td>29</td>
</tr>
<tr>
<td>2009</td>
<td>221</td>
<td>126 ± 22</td>
<td>94 ± 22</td>
<td>35</td>
</tr>
</tbody>
</table>

† Fertilizer rate is the average rate for the whole field, except for the BS that were not fertilized.
‡ Uptake was determined in the 5-m-wide strips next to the ditch of each treatment.
groundwater table in two transects with groundwater wells at 2, 5, 15, 30, 60, 90, and 120 m from the center of the ditch (Fig. 1a) to confirm that the groundwater table was sloping (perpendicular) toward the ditch. The average stream line pattern for this location was simulated using the steady state model FLONET (Molson and Frind, 2010), based on groundwater divide, the averaged rainfall surplus, and the observed water discharges to the field ditch during the experimental period.

Deuterium Tracer Experiment

To assess the hydrological lag time of the treatment response, we applied a conservative tracer at the field edge of the treatment strips (6.5 m from ditch center; Fig. 1) before the first leaching season (replicate A: 17 Oct. 2006; replicates B and C: 2 Nov. 2007). Deuterated water (\(^{2}\text{H}_2\text{O}\)) is an effective tracer for this purpose because of its chemical stability, nonreactivity, ease of handling and sampling, relatively neutral buoyancy, and reasonable price (Becker and Coplen, 2001). These authors concluded that transport of deuterated water is conservative and produces almost identical breakthrough curves to that of other soluble tracers. Deuterated water (80% molar basis) was added at a depth of approximately 0.1 m and at a rate of 0.98 kg m\(^{-1}\) (replicates B and C: 1.50 kg m\(^{-1}\)), corresponding to 0.80 kg \(^{2}\text{H}_2\text{O}\) per m (for B and C: 1.22 kg \(^{2}\text{H}_2\text{O}\) m\(^{-1}\)). Ideally—that is, with flow exactly perpendicular to the ditch—5 m of the applied deuterium can be captured by the 5-m-long reservoirs. Theoretically, the maximum recovery would be 4.00 kg \(^{2}\text{H}_2\text{O}\) for replicate A and 6.12 kg \(^{2}\text{H}_2\text{O}\) for replicates B and C.

The \(^{2}\text{H}_2\text{O}\) concentration was measured in the filtered water samples of the reservoirs and in separate ditch samples, taken 30 m upstream of the experimental site. Later samples were taken for local background \(^{2}\text{H}_2\text{O}\) concentration. The contents, in isotopic research commonly expressed as the molar ratio to an international standard (expressed in pro-mille, and presented in the common delta notation), were transformed to a \(^{2}\text{H}_2\text{O}\) concentration (g m\(^{-3}\)) according to

\[
C = \frac{R_{\text{res}}}{R_{\text{res}} + \frac{18}{20}} \rho_w
\]

with

\[
R_{\text{res}} = R_{\text{res}} - R_{\text{beg}} = R_{\text{VSMOW}} \frac{\delta^2 H_{\text{res}} - \delta^2 H_{\text{beg}}}{1000}
\]

where \(C\) is the \(^{2}\text{H}_2\text{O}\) concentration above the background concentration (g m\(^{-3}\)), \(R\) is the molar ratio \(^{2}\text{H}/^{1}\text{H}\) (mol mol\(^{-1}\)), the subscripts “res” and “beg” refer to reservoir and background, \(\delta^2 H\) is the measured deuterium value (%), \(R_{\text{VSMOW}}\) is the Vienna Standard Mean Ocean Water molar ratio \([R_{\text{VSMOW}} = (155.76 \pm 0.1)10^{-6} \text{ mol mol}^{-1}]\), 18 and 20 are the molar weights (g mol\(^{-1}\)) of \(^{1}\text{H}_2\text{O}\) and \(^{2}\text{H}_2\text{O}\), respectively, and \(\rho_w\) is the density of water (here taken to be 10\(^{6}\) g m\(^{-3}\)). The \(^{2}\text{H}_2\text{O}\) concentration is used frequently instead of measured delta values, for example, in modeling studies (e.g., Braud et al., 2005a; Koarashi et al., 2002). The accumulated \(^{2}\text{H}_2\text{O}\) load per reservoir follows from Eq. [1] with \(C\) from Eq. [3]. The recovery follows from the load divided by the anticipated maximum recovery given above.

Buffer Strip Effectiveness

In Appendix A, we present four alternative definitions of buffer strip effectiveness (BSE) for surface water quality improvement, including references to definitions in the literature. The following two definitions were used in this study.

For the first definition, the difference between the inlet (in) and the outlet (out) of a BS determines BSE\(_I\):

\[
BSE_I = 1 - \frac{Y_{\text{BS, out}}}{Y_{\text{BS, in}}}
\]

where \(Y\) stands for either concentration \(C\) or load \(L\). The drawback of BSE\(_I\) is that even without a BS, \(Y\) might be reduced between inlet and outlet, in which case, BSE\(_I\) overestimates the true effect (Dosskey, 2002). We applied BSE\(_I\) to \(C_{pw}\) only.

For the second definition, the difference between paired BS and REF treatments determines BSE\(_{II}\):

\[
BSE_{II} = 1 - \frac{Y_{\text{BS, out}}}{Y_{\text{REF, out}}}
\]

We applied BSE\(_{II}\) to both \(\overline{C}\) and \(L\) in the reservoirs and to \(C_{pw}\) closest to the ditch.

The quantity \(Y\) is generally determined for a fixed period of time (leaching season). However, the accumulated discharge \(\sum Q\) for this time period may be different between BS and REF due to spatial variability. Therefore, we also considered alternative periods of equal discharge since the start of each leaching season. For instance, the average flow weighted concentration for a certain amount of discharge \(Q\) is computed according to the following (cf. Eq. [2]):

\[
\overline{C} = \frac{\sum Q_{\text{out}} C_{\text{out}}}{\sum Q_{\text{in}}}
\]

Note that if the period is based on equal \(\sum Q\), Eq. [6] yields identical results for \(Y = L\) and \(Y = \overline{C}\).

Besides the individual leaching seasons, we also considered the total period of three (replicates B and C) or four (replicate A) leaching seasons, both based on equal time and on equal discharge.

We calculated BSE\(_{II}\) for both separate replicates and their average (Appendix D).

Statistical Analysis

We also determined BSE from a statistical analysis (see also Schmitt et al., 1999; Spruill, 2000; Veum et al., 2009; Young and Briggs, 2005). We applied restricted (or residual) maximum likelihood analysis (VSNI, 2010; directive REML in GenStat) to the reservoir data \(Q, C,\) and \(L\) (only for \(N, P, \text{NO}_3\), and \(\text{Cl}\)). This type of analysis (REML) is appropriate for unbalanced data sets. The fixed model in REML was:

- constant + treatment (T) + leaching season (LS) + their interaction T×LS. The random model was: replicate (R) + interactions R×T + R×LS + R×T×LS. The treatments were BS and REF, the replicates were A, B and C, and the leaching seasons were 2006–2007, 2007–2008, 2008–2009 and 2009–2010. We
tested the null hypothesis that there is no difference between REF and BS, as well as the effects of LS and the interaction T×LS. We will present the $F$ probability ($P$ value) obtained for the fixed model terms, with significance considered as $P < 0.05$.

Inspection of the distribution of the residuals of $P$, (load and concentration) revealed that the residuals were not normally distributed. Therefore, the REML analysis was repeated for $P$, with log-transformed data for which better residuals were obtained.

**Results and Discussion**

**Hydrology and Discharge**

The measured groundwater levels yielded a sloping groundwater plane toward the ditch with the maximum groundwater level located 60 m from the center of the ditch (Fig. 2).

The first leaching season was the wettest and the third the driest, which is also reflected in the discharge volumes (Table 3). On average, discharge was proportional to rainfall surplus during the discharge period. The discharge volume per reservoir in liters was converted to millimeters in Table 3 by assuming a recharge land surface of 300 m$^2$ (5 m reservoir width times groundwater divide at 60 m field inwards, Fig. 1).

Except for replicate B, differences in discharge between pairs of BS and REF reservoirs were not consistent. Because no influence of the BS treatment on water discharge can be expected, this difference is attributed to spatial variability in hydrology.

A comparison of observed water discharges to the field ditch with rainfall excess rates revealed that only the rainfall excess within 30 m from the center of the ditch drained into the ditch (Table 3). So, only half of the expected recharge area (30×5 in stead of 60×5) drained into the ditch. The predicted stream lines according to FLONET are depicted in Fig. 3. The first streamline starting just outside the BS, where the deuterium tracer was added (Fig. 3: 6.5 m from ditch center), remained shallow (<1.5 m bss), whereas the furthest streamline reaching the ditch reached a depth of about 7 m bss. Although in reality the position of the stream lines differs from the average steady state pattern, it seems clear that many streamlines bypassed the BS at relatively great depths.

**Deuterium Tracer Experiment**

The recovery of deuterium in the reservoirs is presented in Fig. 4. Except for REF B, breakthrough starts after approximately 25 m$^3$ (167 mm) of discharge into the reservoirs. From the initial soil survey, we expected to have a relatively homogeneous soil. However, the breakthrough curves were not similar, implying the presence of significant spatial hydrological variability. The extremely early breakthrough for REF B was likely due to preferential flow paths.

Incomplete recovery of deuterium is likely as deuterated water was subject to removal with time due to uptake by the roots (transpiration) and evaporation from the soil (Braud et al., 2005b). Since we were only interested in the time of breakthrough, it is not necessary to quantify these losses.

Four treatments (REF A, BS A, REF B, and REF C) appeared to have reached maximum recovery, but there could still be some deuterated water left in the soil for treatments BS B and BS C. Assuming that the point of maximum recovery was reached at the end of the experiment, the average travel time of the water leaching from the edge of the treatment strip to the ditch was derived from Fig. 4 as the time period after which half of the maximum recovery was observed. For REF A, BS A, REF B, BS B, REF C, and BS C, this was 146, 411, 136, 790, 496, and 778 d, respectively. On average, residence time was one to two leaching seasons, so after one to two leaching seasons, groundwater below the treatment strips

**Table 3. Rainfall, reference grass evapotranspiration according to Makkink (de Bruin, 1987), rainfall excess, individual and average (± standard deviation) discharges into reservoirs for the four successive leaching seasons (1 October–1 April). The numbers in parentheses after the discharges represent the number of water samples taken.**

<table>
<thead>
<tr>
<th>Replicate</th>
<th>Rainfall</th>
<th>Reference evaporation</th>
<th>Rainfall excess</th>
<th>Discharge†</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>REF, A</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>BS, A</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>REF, B</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>BS, B</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>REF, C</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>BS, C</td>
</tr>
<tr>
<td>Average</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

| 2006–2007‡ | 489      | 107                   | 382            | 167 (36)   |
| 2007–2008  | 391      | 106                   | 285            | 131 (41)   |
| 2008–2009  | 273      | 97                    | 176            | 58 (27)    |
| 2009–2010  | 363      | 98                    | 265            | 83 (21)    |

† The measured discharges in liters were converted to millimeters by assuming a recharge land surface of 300 m$^2$ (5 m reservoir width × groundwater divide at 60 m field inwards). REF = reference; BS = buffer strip.

‡ In 2006, replicates B and C were not yet present.
was replaced by new water affected by the treatment. A treatment effect on leaching toward the ditch should be visible within our study period.

**Concentrations in Reservoirs**

The total number of flow proportional samples taken per replicate per leaching season is presented in Table 3. Concentrations of $N_t$ in the reservoirs fluctuated between 20 g m$^{-3}$ (leaching season) and 2 g m$^{-3}$ (growing season; Fig. 5), with an average of 13 g m$^{-3}$. Seasonal fluctuation was primarily due to fluctuating NO$_3$–N concentration, which comprised 85 to 90% of $N_t$. Such seasonal fluctuation of NO$_3$–N was also found by Jarvie et al. (2010), Quinn et al. (2008), and Van der Velde et al. (2010), who explained it with travel time distribution dynamics. During winter with higher groundwater table, NO$_3$–N travels via shorter stream lines, leaving less time for denitrification, thus causing higher NO$_3$–N concentration. Furthermore, during winter, temperatures are not favorable for denitrification.

The median $N_t$ and $P_t$ concentration ($13.2$ g m$^{-3}$ and $0.05$ g m$^{-3}$) corresponded well with the median $N_t$ and $P_t$ concentration ($14.0$ and $<0.06$ g m$^{-3}$) in ditch water of 11 sandy soil farms and the median upper groundwater concentration ($14.2$ and $<0.06$ g m$^{-3}$) below 148 sandy soil farms in the Netherlands (Fraters et al., 2008).

In >99% of our reservoir samples, PO$_4$–P concentrations were below the detection limit of 0.02 g P m$^{-3}$. There are two reasons for these low reservoir P concentrations. No surface runoff was detected, which is generally considered to be the key transport route for P loads (Dorioz et al., 2006; Hoffmann et al., 2009). Next, compared with phosphate-saturated soils in the Netherlands (Schoumans, 2004), the average groundwater level was relatively low (1 m bss). Lateral groundwater flow PO$_4$–P concentration is determined by the phosphorus saturation degree (PSD) of the soil layer at the depth of discharge (Dorioz et al., 2006; Hoffmann et al., 2009; Schoumans et al., 2009; Seibert et al., 2009). In Beltrum, PSD below 0.3 m bss was on average 0.09 (Table 1), far below the environmental limit of 0.25 corresponding with a concentration limit of 0.10 g P m$^{-3}$ for Dutch sandy soils (Van der Zee et al., 1990). The low P values at Beltrum can be explained by the small contribution of shallow flow (Fig. 3).

For later BSE assessment, we calculated flow-averaged $N_t$ and $P_t$ concentration per leaching season and for the entire experimental period (Fig. 6; see Supplemental Material for more data on discharges, loads, and other species). The flow-averaged concentrations of $N_t$, $P_t$, NO$_3$–N, and Cl differed between seasons, replicates, and treatments. However, there was no consistent difference between BS and REF.

We expect evaporation of water from the reservoirs will not have affected the results because it occurs at the same rate in both the REF and BS reservoirs and will not change the $C_{BS}/C_{REF}$ ratio (see Eq. [5] and [6]). Moreover, in practice, evaporation...
also occurs in the ditch. We did not observe growth of and nutrient uptake by water plants and algae during the leaching periods.

Concentration in Upper Groundwater

The NO$_3$–N concentration of the upper groundwater (mostly at 1 m bss) as a function of distance to the ditch resulted in similar patterns for all samplings (from December 2006 until April 2010) and for all three replicates. For that reason, only the averages are shown in Fig. 7. As may be expected, there was no difference between REF and BS outside the treatment strips (>6.5 m). Average $C_{gw}$ for NO$_3$–N was 31 g m$^{-3}$, which is higher than the median upper groundwater concentration of 11.5 g m$^{-3}$ found below 148 sandy soil farms in the Netherlands (Fraters et al., 2008). van Beek et al. (2007) reported NO$_3$–N concentrations outside a BS on a sandy soil in the range 15 to 29 g m$^{-3}$.

Underneath the BS, however, upper groundwater NO$_3$–N concentration was significantly lower than underneath REF, although there was no difference in reservoir concentration at the same sampling moments (Fig. 7). Especially for REF, the concentration in the reservoir does not resemble the concentration in the upper groundwater next to the ditch. Down to 1.5 m bss, NO$_3$–N underneath BS was always lower than underneath REF. However, no difference was recorded at 3 m bss, and at 2 m bss the concentration below BS was even higher than below REF (Fig. 8; left graph). In the treatment strips below 1.5 m bss, almost constant NO$_3$–N concentrations <30 g m$^{-3}$ were measured, i.e., somewhat lower than the upper groundwater concentration outside the treatment strips. Outside the treatment strips, no differences in NO$_3$–N concentrations were observed between the treatments (8 m from ditch center; Fig. 8 right graph).

At all times, the PO$_4$–P concentrations were below the detection limit (0.02 g P m$^{-3}$), so no BS effect could be determined. Fraters et al. (2008) also found such low upper groundwater PO$_4$–P concentrations below 148 sandy soil farms in the Netherlands (median value less than their detection limit of 0.06 g P m$^{-3}$).

Buffer Strip Effectiveness: BSE

The observations in the upper groundwater (Fig. 7) revealed a large drop in NO$_3$–N concentration underneath the BS. We used average $C_{gw}$ at 1 m depth in the BS (distance 2 m) and outside the BS (distance 8 m) to calculate a BSEI of 67% (Eq. [5], Table 4), indicating a clear BS effect in reducing upper groundwater NO$_3$–N concentrations. A similar result of 66% was obtained for BSEII with Eq. [6], using $C_{gw}$ in BS and REF, both at 2 m distance from the center of the ditch (Table 4). Positive BSE values of the same magnitude were also obtained for N$_w$ and Cl. Phosphate concentrations were too low to calculate BSE. Results for other species are presented in the Supplemental Material.

A different picture emerged when BSE for NO$_3$–N was computed from Eq. [6] based on reservoir data of either $L$ or $C$ for periods of equal times or equal discharges (Table 5; Fig. 9). The average BSE$_L$ for N$_w$, N$_w$, and NO$_3$–N were all negative and mutually comparable, due to similar flow-averaged concentrations. A negative BSE means that $L$ in the BS reservoir was higher than in the REF reservoir. Higher concentration in lateral flow at the BS position could have already been the case before installing the treatment due to spatial variability. We could not correct for this because no reservoir measurements were available of the period before treatment.
For Pt, a positive BSE was obtained, but this was based on very low concentrations, which makes BSE less relevant and also more sensitive to measurement errors. No data for PO₄–P are given since its concentration in most cases was below the detection limit. Results for other species are presented in the Supplemental Material.

Comparing the BSE values obtained from the upper groundwater C (Table 4) with those obtained from the reservoir C (Table 5) revealed no correspondence between the two. The BSE value of 66% obtained from Cgw NO₃–N falls in the range of values between 46 and 99% reported in literature for shallow groundwater flow under grassed BS (e.g., Dosskey, 2002; Mayer et al., 2007; Wenger, 1999). Our study showed that this effect does not necessarily equal the effect in the surface water system as determined from reservoir data. A better correspondence may be expected under typical hydrological situations where upper groundwater is the only contributor to ditch discharge.

In many cases, BSE based on L was larger than that based on C, and for replicate B there was a relevant difference between periods of equal time and equal discharge. As L is dependent on Q (see Eq. [1]) and Q is not influenced by treatment but by spatial variation, we prefer BSE based on C and equal discharge because this excludes the influence of Q on BSE.

For many species, there was large variation in BSE between replicates; sometimes BSE even changed from positive to negative. This is attributed to spatial variability. Also temporal variability played a role, as L and C differed between the seasons.

**Statistical Evaluation**

Table 6 presents the differences between the treatments for water discharge, and concentrations and loads of N, P, NO₃–N, and Cl, as well as the significance levels (P values) for the terms treatment (T), leaching season (LS), and their interaction (T × LS). The reported differences for Pt in Table 6 are the differences between the treatments for the back-transformed means.

The only significant treatment effect (P < 0.05) was obtained for the load of Pt (Table 6). However, the effect on the load (0.011 g) was very small. As mentioned above, we prefer to use concentration instead of load to assess BSE, to avoid influence of differences in discharge. The corresponding difference in C was only 0.006 g m⁻³, and not significant.

For many cases in Table 6, a significant effect was found for leaching season (LS). This is not surprising as the processes that determine concentration and load are strongly influenced by weather, especially the precipitation surplus, which differed between the seasons (Table 3). No interaction effect between treatment and leaching seasons (T × LS) was found. This means that the difference between BS and REF was not increasing over time, based on the available data. Although the tracer results show that the hydrological lag time lies in the order of 1 or 2 yr, this was not reflected by an increasing treatment effect over time.

The relative differences in Table 6 were of the same magnitude and sign as the BSE values reported in Table 5 (results for all leaching seasons). According
to the statistical analysis, the null hypothesis is not rejected, i.e., there was no significant effect of the BS ($P < 0.05$) on the improvement of the quality ($C$) of the ditch water.

The statistical analysis requires normally distributed input data, which is almost never completely true in experimental research on water quality. Furthermore, the power of the analysis was low because of the limited number of degrees of freedom. Given the small effects, (many) more replicates would be needed to show significant effects. For this type of experiments, however, the number of replicates is generally limited to two (e.g., Usi-Kämppä and Jauhiainen, 2010), three (as in our case), or four (e.g., Duchemin and Hogue, 2009; Hay et al., 2006; Magette et al., 1989).

In this study, there was a significant treatment effect on N (and P) surplus at the soil surface (Table 2). Subsequently, we observed a significant decrease in NO$_3$–N in the upper groundwater below the BS compared with the REF (Fig. 7), which is attributed to the negative N surplus of this treatment. From the measurements in the in-stream reservoirs, however, we did not find any effect of the BS treatment in the quality of the leaching groundwater. We concluded that most water entering the ditch originated from outside the treatment strips, based on the average stream line pattern (Fig. 3). Upper groundwater with reduced NO$_3$–N concentration hardly affected the concentration of the in-stream reservoirs. Therefore, BSE based on upper groundwater measurements is not valid for deeply permeable soils as in this study. Groundwater measurements, although frequently applied in literature, are only valid for BSE assessment under specific hydrological conditions, where shallow groundwater flow is the only (or predominant) contributor to lateral flow.

Spatial variability in hydrology and chemistry was the most likely cause of the observed higher reservoir N concentrations for the BS treatments than for the REF treatments, yielding a negative estimate of the BSE with respect to water quality improvement. For P, we did not see any treatment effect due to low concentrations, both in upper groundwater and reservoirs, because of low PSD below 0.3 m bss. The null hypothesis, stating that in the reservoirs there is no difference in concentration between the BS and REF treatments, was not rejected. Following Dosskey (2002), we suggest that BSE should be determined by monitoring losses before and after installation of a BS (BSE$_{in}$ in Appendix A). We believe that this approach may account for temporal variability but still not for spatial variability. Therefore, control treatments (REF) also need to be measured before and after implementation of the BS (see Appendix A, BSE$_{in}$). Initial variability either can then be accounted for in the definition of BSE or the initial measurements can be used to increase the power of the statistical test by using them as covariates to reduce between-reservoir variability.

We tested a novel method to determine BSE on a 20-m-deep, permeable sandy soil. Discharge from soil to ditch was temporarily collected in a reservoir to measure its quantity and quality, both for a BS and a REF treatment, in triplicate. Although a positive BS effect was observed in upper groundwater for N, no significant BSE was observed for N or P concentrations in the reservoirs during the experimental period of 4 yr. Reservoirs also collected groundwater from greater depths that were not influenced by the treatments. We conclude that measuring changes in upper groundwater to assess BSE is only valid under specific hydrogeological conditions with predominant shallow flow. A REF treatment is imperative if BS need to be evaluated for application on agricultural fields. Without REF, reduction of N loads or concentrations may be abusively attributed to BS treatment. We propose a

---

### Table 5. Buffer strip effectiveness (BSE) according to Eq. [6] for the three replicates at Beltrum and their averages ($\pm$ standard deviation) based on flow-averaged concentration ($C$) of total nitrogen (N$_t$), total phosphorus (P$_t$), NO$_3$–N, and Cl in the reservoirs for the complete experimental period of either equal time (t) or equal discharge (Q).

<table>
<thead>
<tr>
<th>Replicate</th>
<th>Period</th>
<th>$N_t$ ($L$)</th>
<th>$N_t$ ($C$)</th>
<th>$P_t$ ($L$)</th>
<th>$P_t$ ($C$)</th>
<th>NO$_3$–N ($L$)</th>
<th>NO$_3$–N ($C$)</th>
<th>Cl ($L$)</th>
<th>Cl ($C$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>Q</td>
<td>−17.21 ± 5.25</td>
<td>−17.21 ± 6.44</td>
<td>15.45 ± 3.22</td>
<td>15.45 ± 4.64</td>
<td>−17.15 ± 5.27</td>
<td>−17.15 ± 6.45</td>
<td>11.49 ± 2.34</td>
<td>11.49 ± 3.41</td>
</tr>
<tr>
<td></td>
<td>t</td>
<td>−5.94 ± 2.41</td>
<td>−14.41 ± 6.34</td>
<td>20.88 ± 4.48</td>
<td>14.55 ± 3.97</td>
<td>−5.15 ± 2.16</td>
<td>−13.57 ± 6.14</td>
<td>17.79 ± 2.06</td>
<td>11.21 ± 2.29</td>
</tr>
</tbody>
</table>

---

![Fig. 9. The buffer strip effectiveness (BSE) computed with Eq. [6] for total nitrogen (N$_t$), total phosphorus (P$_t$), NO$_3$–N, and Cl based on flow-averaged concentration and periods of equal discharge in the reservoirs. The error bars represent two times the sample standard deviation.](image-url)
superior experimental set-up for future research, including extra measurements before installing the BS and REF treatments to deal with spatial and temporal variability.

Appendix A: Expressions for Buffer Strip Effectiveness

The majority of the expressions for BSE in the literature can be given by

\[
BSE = \frac{Y_{\text{out}} - Y_{\text{in}}}{Y_{\beta}} = 1 - \frac{Y_{\alpha}}{Y_{\beta}} \quad [A1]
\]

where \( Y \) stands for either load \( L \) (kg, kg ha\(^{-1}\), or kg ha\(^{-1}\) yr\(^{-1}\)) or concentration \( C \) (g m\(^{-3}\)), and \( \alpha \) and \( \beta \) for the location where \( Y \) is measured with or without indication for the moment in time of measurement. Thus, BSE represents the (dimensionless) relative difference between \( Y_{\alpha} \) and some reference \( Y_{\beta} \). Both \( L \) and \( C \) may refer to subsurface or surface runoff or total loads. The maximum value for BSE is 1 (or 100%), when \( Y_{\alpha} \) equals zero. In case \( Y_{\beta} \) reaches zero BSE \( \rightarrow -\infty \). BSE becomes negative when \( Y_{\alpha} > Y_{\beta} \), which may occur if the buffer strip releases extra nutrients or due to spatial variability. Several references given by Dosskey (2001) reported (extreme) negative BSE values. To overcome the unbalanced ranges between positive and negative BSE, Eq. [A1] could be redefined as given in Appendix B. Below, we consider four cases for \( \alpha \) and \( \beta \), which are illustrated in Fig. 10.

The first approach considers the difference between the inlet (in) and the outlet (out) of a BS (Fig. 10, Method I). BSE\(_I\) is given by

\[
BSE_I = 1 - \frac{Y_{\text{BS,out}}}{Y_{\text{BS,in}}} \quad [A2]
\]

Examples for BSE\(_I\) based on \( C_{\text{out/in}} \) can be found in Borin and Bigon (2002), Sabater et al. (2003), Schmitt et al. (1999), and Young and Briggs (2005). Examples for BSE, based on \( L_{\text{out/in}} \) can be found in Barfield et al. (1998), Chaubey et al. (1994; 1995), Kuusemets and Mander (1999), Mander et al. (1997), and Patty et al. (1997). The drawback of BSE\(_I\) is that without a BS, \( Y \) might be reduced between inlet and outlet, so that BSE\(_I\) overestimates the true effect (Dosskey, 2002).

To account for this possible change, measurements outside the BS can be used (Fig. 10, Method II). The quantity \( Y \) leaving the BS is compared to \( Y \) measured at the outlet of a separate reference strip (REF). BSE\(_{II}\) follows from

\[
BSE_{II} = 1 - \frac{Y_{\text{REF,out}}}{Y_{\text{BS,out}}} \quad [A3]
\]

In this case, it is no longer necessary to measure \( Y \) at the inlet. Examples can be found in Dillaha et al. (1989), Duchemin and Hogue (2009), and Magette et al. (1989). The drawback of BSE\(_{II}\) is that \( Y_{\text{REF}} \) is not measured at the same location as \( Y_{\text{BS}} \). BSE\(_{II}\) is, therefore, influenced by spatial variability in soil physical, chemical, and biological properties.

To exclude spatial variability from BSE assessments, one could measure \( Y \) for a certain period before (subscript b) installing the BS and then continue measuring \( Y \) after (subscript a) installing the BS (Fig. 10 Method III). BSE\(_{III}\) is given by

\[
BSE_{III} = 1 - \frac{Y_{\text{BS,b,out}}}{Y_{\text{BS,a,out}}} \quad [A4]
\]

Table 6. Average difference (reference [REF] – buffer strip [BS]) and relative difference with respect to the reference treatment [100%(REF – BS)/REF] for water discharge, and concentrations and loads of total nitrogen (N\(_t\)), total phosphorus (P\(_t\)), NO\(_3\)–N, and Cl, and the significance level (P value) for the fixed restricted maximum likelihood (REML) model terms treatment (T), leaching seasons (LS), and their interaction (T×LS).

<table>
<thead>
<tr>
<th>Constituent</th>
<th>Quantity†</th>
<th>Units</th>
<th>Difference</th>
<th>Relative %</th>
<th>T</th>
<th>LS</th>
<th>T×LS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Water</td>
<td>Q</td>
<td>m(^3)</td>
<td>-0.048</td>
<td>3.49</td>
<td>0.721</td>
<td>0.029</td>
<td>0.360</td>
</tr>
<tr>
<td>N(_t)</td>
<td>C</td>
<td>g m(^{-3})</td>
<td>-2.34</td>
<td>-14.80</td>
<td>0.583</td>
<td>0.023</td>
<td>0.649</td>
</tr>
<tr>
<td>L</td>
<td>g</td>
<td></td>
<td>-4.18</td>
<td>-18.30</td>
<td>0.390</td>
<td>0.051</td>
<td>0.861</td>
</tr>
<tr>
<td>P(_t)†</td>
<td>C</td>
<td>g m(^{-3})</td>
<td>0.006</td>
<td>11.82</td>
<td>0.319</td>
<td>0.063</td>
<td>0.745</td>
</tr>
<tr>
<td>NO(_3)–N</td>
<td>C</td>
<td>g m(^{-3})</td>
<td>-2.13</td>
<td>-14.82</td>
<td>0.578</td>
<td>0.001</td>
<td>0.573</td>
</tr>
<tr>
<td>L</td>
<td>g</td>
<td></td>
<td>-3.635</td>
<td>-17.23</td>
<td>0.405</td>
<td>0.041</td>
<td>0.922</td>
</tr>
<tr>
<td>Cl</td>
<td>C</td>
<td>g m(^{-3})</td>
<td>2.34</td>
<td>9.09</td>
<td>0.262</td>
<td>0.007</td>
<td>0.649</td>
</tr>
<tr>
<td>L</td>
<td>g</td>
<td></td>
<td>2.565</td>
<td>7.24</td>
<td>0.605</td>
<td>0.024</td>
<td>0.926</td>
</tr>
</tbody>
</table>

† Water discharge and loads refer to quantities between two sampling times.
‡ For Pt, log-transformed data were used in the REML analysis; the difference for Pt is based on the back-transformed means of the treatments.
According to Dosskey (2002), this would yield the most direct estimate of the impact that buffer strip installation would have, but no study of this kind had been reported in the 11 review papers he considered. However, BSE\textsubscript{III} also has a disadvantage because it may be affected by autonomous changes and temporal variability, especially differing weather conditions, between the before and after periods.

The fourth approach is an attempt to overcome the drawbacks of the former ones by combining them and has not been proposed before (Fig. 10, Method IV). In Appendix C, BSE\textsubscript{IV} for this situation is derived and reads

\[
BSE_{IV} = 1 - \frac{Y_{BS,a,out}}{Y_{REF,a,out}} Y_{BS,b,out} \frac{Y_{REF,b}}{Y_{REF,a}} \tag{A5}
\]

Note that if \( Y \) in the reference plot is constant (no autonomous changes), i.e., \( Y_{REF,a,out} = Y_{REF,b,out} \), Eq. \((A5)\) reduces to Eq. \((A4)\). If \( Y \) in the before period does not differ between BS and REF, i.e., \( Y_{BS,b,out} = Y_{REF,b,out} \), Eq. \((A5)\) reduces to Eq. \((A3)\). Also, BSE\textsubscript{IV} can be redefined to yield values bounded in the range \([-1, 1]\) (Appendix C).

Both BSE\textsubscript{III} and BSE\textsubscript{IV} require measurements for two prolonged periods (before, after). To the best of our knowledge, no studies are known that have taken measurements for a long period before installing BS.

**Appendix B: Definition of BSE with Bounded Lower Value**

Instead of using \( Y_{a} \) in the denominator of Eq. \((A-1)\), one could use \( Y_{a} + Y_{b} \), resulting in

\[
BSE_{new} = \frac{Y_{a} - Y_{a}}{Y_{a} + Y_{a}} = \frac{1 - y}{1 + y} = \frac{BSE}{2 + BSE} \tag{B1}
\]

where \( y = Y_{a}/Y_{a} \). For \( Y_{a} = 0 \), BSE\textsubscript{new} = 1, and for \( Y_{a} = 0 \), BSE\textsubscript{new} = -1, so that BSE\textsubscript{new} is bounded by \([-1, 1]\). The denominator in Eq. \((B-1)\) uses \( Y_{a} + Y_{b} \), rather than the average \( 0.5(Y_{a} + Y_{b}) \), as then BSE\textsubscript{new} would be bounded by \([-2, 2]\).

For \( Y_{a} < Y_{b} \), BSE\textsubscript{new} < BSE. The largest difference occurs for \( y = 1+\sqrt{2} \) and equals \(-3+2\sqrt{2}\), where BSE = \( 2-\sqrt{2} \) and BSE\textsubscript{new} = \(-1+\sqrt{2}\). For any \( y \), the difference BSE\textsubscript{new} - BSE is given by \(-yBSE_{new} \) or \(-yBSE/(1+y)\).

**Appendix C: Derivation of BSE for the Before-After Reference-Buffer Strip Situation**

There are several ways to come to an expression for BSE for Method IV (Appendix A, Fig. 10). First, we start from Eq. \((A4)\). The newly measured after data \( Y_{BS,b} \), can be corrected for the observed change in a reference strip by multiplying it with the ratio \( Y_{REF,b}/Y_{REF,a} \), yielding

\[
Y_{BS,b,new} = Y_{BS,b} \frac{Y_{REF,b}}{Y_{REF,a}} \tag{C1}
\]

For convenience, we have omitted the subscript “out.” If in the reference strip, \( Y \) decreases so that \( Y_{REF,a} > Y_{REF,b} \) then one would expect the same autonomous decrease to occur in the BS. If then we multiply \( Y_{BS,a} \) by \( Y_{REF,a}/Y_{REF,b} \) then the difference \( Y_{BS,b,new} - Y_{BS,a} \) represents the change in the BS that can be truly attributed to the presence of the BS.

Alternatively, BSE\textsubscript{IV} can be calculated by subtracting an analogous expression obtained for the reference strip from Eq. \((A4)\), according to

\[
BSE = \frac{Y_{BS,b,new} - Y_{BS,a}}{Y_{BS,b}} \frac{Y_{REF,b}}{Y_{REF,a}} \tag{C2}
\]

or

\[
\frac{Y_{REF,a}}{Y_{REF,b}} \frac{Y_{BS,a}}{Y_{BS,b}} \frac{1}{Y_{REF,b}} \frac{Y_{REF,a}}{Y_{BS,b}} \tag{C3}
\]

Equivalently, BSE can be calculated by subtracting an analogous expression obtained for the reference strip from Eq. \((A3)\), according to

\[
BSE_{IV} = 1 - \frac{Y_{BS,a}}{Y_{BS,b}} \frac{Y_{REF,b}}{Y_{REF,a}} \tag{C4}
\]

For \( Y_{BS,a} = 0 \), BSE according to Eq. \((C4)\) has a maximum value of \( Y_{REF,b}/Y_{REF,a} \), which is not equal to the expected value of one. This can be overcome by normalizing Eq. \((C2)\) through division by \( Y_{REF,b}/Y_{REF,a} \), yielding

\[
BSE_{IV} = 1 - \frac{Y_{BS,a}}{Y_{BS,b}} \frac{Y_{REF,b}}{Y_{REF,a}} \tag{C5}
\]

Equations \((C1)\), \((C3)\), and \((C5)\) are identical. Note that if \( Y_{REF,b} > Y_{REF,a} \), Eq. \((C1)\) reduces to Eq. \((A4)\). If \( Y \) in the before period does not differ between BS and REF, i.e., \( Y_{BS,b,new} = Y_{REF,b,new} \), Eq. \((C1)\) reduces to Eq. \((A3)\).

The values for BSE\textsubscript{IV} fall in the range \([-\infty, 1]\). Analogous to the procedure in Appendix B, BSE\textsubscript{IV} can be rewritten such that the lower boundary is bounded as well, according to

\[
BSE_{IV,new} = \frac{Y_{BS,b,new} - Y_{BS,a}}{Y_{BS,b}} \frac{Y_{REF,b}}{Y_{REF,a}} = \frac{y_{0} - y}{2 - BSE_{IV}} \tag{C6}
\]

where \( y = Y_{a}/Y_{b} \) and \( y_{0} = Y_{0}/Y_{0} \). In case \( y_{0} \) has realistic values (not equal to zero or infinity), BSE\textsubscript{IV,new} = 1 for \( y_{0} = 0 \), and BSE\textsubscript{IV,new} = -1 for \( y_{0} \rightarrow 0 \) or \( y \rightarrow \infty \), so that BSE\textsubscript{IV,new} is bounded by the range \([-1, 1]\).

For \( y_{0} \) given by \( y \), then the difference BSE\textsubscript{IV,new} - BSE\textsubscript{IV} for any \( y \) is given by \(-yBSE_{IV,new} \) or \(-yBSE/(1+y)\).
Appendix D: Average BSE and Standard Deviation

The average BSE for all replicates is not the arithmetic average of the three individual BSE values as BSE is non-linearly related to $Y_{REF}$. For loads, the average BSE is obtained from

$$BSE_{LL, L} = \frac{L_{REF} - L_{BS}}{L_{REF}} ,$$  \hspace{1cm} \text{[D1]}$$

where $\bar{L}$ is the arithmetic average load of the three replicates. The corresponding standard deviation follows from applying the Delta method (assuming no correlation between data from the REF and BS reservoirs) (Cox, 1990):

$$t_{BSE, L} = \sqrt{BSE_{LL, L} \left[ \frac{L_{REF}}{L_{BS}} \right]^2 + \frac{L_{BS}}{L_{BS}}^2} \hspace{1cm} \text{[D2]}$$

where $s$ is the sample standard deviation. For concentrations, the average BSE is obtained from

$$BSE_{IL, C} = \frac{C_{REF} - C_{BS}}{C_{REF}} = \frac{L_{REF} - L_{BS}}{L_{REF}}$$

$$\frac{Q_{BS}}{Q_{BS}} $\hspace{1cm} \text{[D3]}$$

The corresponding standard deviation is given by

$$t_{BSE, C} = \sqrt{BSE_{ILL, C} \left[ \frac{L_{REF}}{L_{BS}} \right]^2 + \frac{L_{BS}}{L_{BS}}^2 + \frac{Q_{REF}}{Q_{REF}}^2 + \frac{Q_{BS}}{Q_{BS}}^2} \hspace{1cm} \text{[D4]}$$

Equations [D1] and [D3] can both be computed for periods of equal time or equal discharge. In case periods of equal discharge are considered, it easily follows that Eq. [D3] is equal to Eq. [D1] as $(Q_{REF} + Q_{REF} + Q_{REC}) = (Q_{BSA} + Q_{BSB} + Q_{BSC})$. However, the standard deviations differ.

Acknowledgments

This research was funded by the Dutch Ministry of Agriculture, Nature and Food Quality (research program BO-12-07-009-002; project “Effectiveness of buffer strips in the Netherlands”) and the Dutch Ministry of Housing, Spatial Planning and the Environment. Han te Beest, Jan van Kleef, Antonie van den Toorn, Meint Veninga, and Gerben Bakker were responsible for the construction and field work. We thank the Chemical and Biological Soil Laboratory (Alterra, Wageningen-UR) for performing all the chemical analyses. Eduard Hummelink carried out the deuterium tracer experiment. The deuterium contents by laser-absorption spectroscopy were done by the Davis Stable Isotope Facility of the University of California (USA). Special thanks are due to the farmers Huijnik and Ribbers for making available the field and electricity. We are grateful for the suggestions and critical comments by the scientific members of the support committee: Olga Clevering (Ministry of Transport, Public Works, and Water Management), Mariet Hefting (University of Utrecht), Oene Oenema (Alterra, Wageningen-UR), and Jaap Willems (Netherlands Environment Assessment Agency). We thank the three anonymous reviewers for their useful comments. This work was presented during a workshop organized within COST Action 869 in Ballater, UK.

References


