Nutrient source attribution: Quantitative typology distinction of active and legacy source contributions to waterborne loads

Yuanying Chen1,2 | Georgia Destouni1 | Romain Goldenberg1 | Carmen Prieto1

1Department of Physical Geography, Stockholm University, Stockholm, Sweden
2Department of Sustainable Development, Environmental Science and Engineering, KTH Royal Institute of Technology, Stockholm, Sweden

Abstract
Distinction between active and legacy sources of nutrients is needed for effective reduction of waterborne nutrient loads and associated eutrophication. This study quantifies main typological differences in nutrient load behaviour versus water discharge for active and legacy sources. This quantitative typology is used for source attribution based on monitoring data for water discharge and concentrations of total nitrogen (TN) and total phosphorous (TP) from 37 catchments draining into the Baltic Sea along the coastline of Sweden over the period 2003–2013. Results indicate dominant legacy source contributions to the monitored loads of TN and TP in most (33 of the total 37) study catchments. Dominant active sources are indicated in 1 catchment for TN, and mixed sources are indicated in 3 catchments for TN, and 4 catchments for TP. The TN and TP concentration contributions are quantified to be overall higher from the legacy than the active sources. Legacy concentrations also correlate well with key indicators of human activity in the catchments, agricultural land share for TN ($R^2 = 0.65$) and population density for TP ($R^2 = 0.56$). Legacy-dominated nutrient concentrations also change more slowly than in catchments with dominant active or mixed sources. Various data-based results and indications converge in indicating legacy source contributions as largely dominant, mainly anthropogenic, and with near-zero average change trends in the present study of catchments draining into the Baltic Sea along the coastline of Sweden, as in other parts of the world. These convergent indications emphasize needs to identify and map the different types of sources in each catchment, and differentiate strategies and measures to target each source type for possible achievement of shorter- and longer-term goals of water quality improvement.

KEYWORDS
Baltic, legacy sources, multi-catchment analysis, nutrients, source type attribution, water quality
1 | INTRODUCTION

A variety of sources contribute to stream concentrations and loads of nutrients (Destouni et al., 2010) that lead to the common environmental problem of eutrophication (Conley et al., 2009; Diaz & Rosenberg, 2011; Nixon, 1995; Schuwirth et al., 2018; Wurtsbaugh et al., 2019). A range of policies, strategies and management measures have been tried to mitigate nutrient loads and combat eutrophication (D’Elia et al., 2019; European Commission, 2000; Helsinki Commission, 2007; Iho et al., 2015; Le Moal et al., 2019). Yet, mitigation results remain unsatisfactory (Destouni et al., 2017; Van Meter et al., 2018). Since more than a decade ago, nutrient legacies have been repeatedly suggested as a possible key reason for frustrated efforts to reduce nutrient loads and eutrophication (Ardón et al., 2017; Ascott et al., 2017; Darracq et al., 2008; Destouni & Jarsjö, 2018; McCrackin et al., 2018; Poffenbarger et al., 2018; Puckett et al., 2011; Van Meter et al., 2017).

Accumulation of nutrient legacies in soil, groundwater and sediments is a consequence of both slow advective transport and various physical and biogeochemical immobilisation-remobilisation processes that have since long been quantified to occur in the subsurface parts of hydrological catchments (Cvetkovic et al., 2012; Destouni & Cvetkovic, 1991). The remobilisation (desorption, dissolution, diffusive mass transfer) components of these delay processes imply that nutrients are not just irreversibly retained in immobile water/soil/sediment legacy zones, but can also be released back to mobile water and continued, delayed nutrient transport. For mitigation efforts, the long time-lags involved in such delayed transport imply that nutrient decreases and water quality improvements in recipient surface and coastal waters are slow after implementation of mitigation measures (Darracq et al., 2008; Meals et al., 2010; Sharpley et al., 2013; Van Meter & Basu, 2015). This has made goal achievement in water quality and eutrophication management a major sustainability challenge requiring a long-term perspective (Destouni & Jarsjö, 2018; Haygarth et al., 2014; Murray et al., 2019).

To meet this important challenge, mitigation measures need to be selected and placed so they can be effective for the different types of nutrient sources. For currently active surface sources, relevant measures may be regulations and incentives for enhanced nutrient removal in wastewater treatment, reduced fertilizer use, changes in manure spreading time, cultivation of catch crops. For legacy sources in soil, groundwater and sediments, other types of measures are needed, such as well-placed reactive buffer zones, wetlands, pump-and-treat technologies. To effectively select and locate mitigation measures, different types of nutrient sources need to be distinguished, identified, and appropriately targeted (Chanat & Yang, 2018; Levi et al., 2018).

In this paper, a quantitative typology approach to source attribution is outlined that distinguishes between currently active and legacy sources of nutrients, based on commonly available stream monitoring data for water discharges and nutrient concentrations. For concrete case study quantification and testing of this approach we use data for discharge and concentrations of total nitrogen (TN) and total phosphorous (TP) representing 37 hydrological catchments that drain into the Baltic Sea along the coastline of Sweden over the time period 2003–2013. In a parallel study, a similar basic source attribution approach is applied to data for chloride and metals (Destouni et al., 2021). In addition to considering different hydrochemical constituents (nutrients), the present study develops the basic source attribution approach further to a discrete quantitative source typology (whereas the parallel study considers continuous degrees of active and legacy source contributions). The present approach application also regards a greater number of catchments over a larger geographic scale (along the whole coastline of Sweden) and with order-of-magnitude larger study catchments (on the order of $10^2 - 10^4$ km$^2$) than those in the parallel study (covering a local geographic area around a major Swedish lake, and including 19 monitored catchments for chloride and fewer for metals, with study catchment size on the order of $10^2 - 10^5$ km$^2$). In combination, the parallel studies test the applicability, generality and transferability of the underlying source attribution approach for different hydrochemical constituents in and across various hydrological catchment settings.

2 | MATERIALS AND METHODS

2.1 | Quantitative source typology

This section outlines the general quantitative typology of active, legacy, or mixed source dominance in a catchment. This typology is based on first-order quantification of some key characteristics that should, in general, mechanistically differ in the hydrochemical behaviour versus discharge between the contributions from these different types of sources. This difference can be used for relatively simple source type distinction based on commonly available stream monitoring data for water discharge and hydrochemical concentrations. At each catchment outlet (i.e., monitoring point), these monitored water flow and quality quantities integrate the effects of all source inputs, and subsequent physical and biogeochemical transport processes that occur over the catchment and along all hydrological transport pathways to the catchment outlet. In the following outline, we aim to quantify the general first-order characteristic source type differences that may emerge in the integrated hydrochemical signal at the outlet of a catchment, not to model the details and specifics of all sources and transport processes under various catchment conditions. For concreteness, we also focus the outline text on nutrients, since they are the focus of the present study application, but note that the source attribution approach is more general and applicable to various hydrochemical constituents, such as the chloride and metals focused on in the parallel study by Destouni et al. (2021).

2.1.1 | Currently active sources

Consider a nutrient source (or collection of sources) spread over a catchment and delivering a relatively stable average input mass flow...
rate, $I_{in-A}$, over time into the catchment. In the transport to the outlet/monitoring point, some fraction $(1 - \alpha_A)$ of $I_{in-A}$ may be retained and delayed within the catchment (Levi et al., 2018; Quin et al., 2015), in zones of low conductivity and slow flow, and/or by biogeochemical or physical sorption processes (adsorption, chemical precipitation, diffusive mass transfer to immobile water zones), and to some degree even be lost from the catchment (e.g., by denitrification for TN). For this to be considered currently active type of source, delivery of non-retained nutrient mass $(\alpha_A I_{in-A})$ from each point in time of the continuous source input flow $I_{in-A}$ should on average reach the outlet within one or just a few years after that input time. Conservation of mass for a stable input mass flow $(I_{in-A})$ with temporally constant relative retention $(1 - \alpha_A)$ then implies an average outlet load that is also stable, as:

$$L_{out-A} = C_{out-A} Q_{out} \approx \alpha_A I_{in-A}$$

where $C_{out-A}$ is the flux-averaged nutrient concentration contribution from the active sources in the water discharge ($Q_{out}$) at the outlet. Equation (1) implies chemodynamic concentration behaviour, where the concentration ($C_{out-A}$) varies with $Q_{out}$ in order for $L_{out-A}$ to remain stable on average (Basu et al., 2010; Levi et al., 2018; Selroos & Destouni, 2015).

Daily, monthly, and individual annual values of $I_{in-A}$ and $L_{out-A}$ can of course fluctuate around their respective stable longer-term average levels (c.f. Selroos & Destouni, 2015: Equation (1) with resulting stable average $L_{out-A}$ in Figure 3; Destouni & Jarsjö, 2018: Equation (1a) and blue line in Figure 1(c)). For active surface sources with stable average nutrient input rate $(I_{in-A})$, the expected characteristic type of $L_{out-A}$ behaviour versus $Q_{out}$ is a line with zero slope and positive intercept quantifying the stable average $L_{out}$ level, as illustrated by the green line in Figure 1(a). This expectation can be checked against data-given $L-Q$ regression lines, as outlined in the next section for the case study of 37 Swedish monitoring stations (Figure 1(b)).

### 2.1.2 | Legacy sources

For the typology of legacy sources, consider the nutrient mass fraction $(1 - \alpha_A)$ that has been retained within a catchment in soil, groundwater, and sediments, from continuous nutrient input $(I_{in-A})$ for decades. Some part of that immobilized nutrient mass may be continuously released back into mobile water (by dispersive/diffusive mass transfer, desorption, dissolution from slow/immobile water and/or sorption zones). The resulting outlet load contribution ($L_{out-L}$) can then be quantified based on Equation (2) of Destouni and Jarsjö (2018) as:

$$L_{out-L} = C_{out-L} Q_{out} \approx \left(\frac{C_0}{\kappa T}\right) Q_{out}$$

where $C_{out-L}$ is the flux-averaged nutrient concentration contribution from the legacy sources, $n$ is average volumetric water content in the soil/aquifer/sediment zones containing the legacy source ($n$ is average porosity if the pore volume in these is largely water-filled), $C_0$ is average bulk concentration in the legacy zones (i.e., average nutrient mass per unit bulk soil/aquifer/sediment volume) at the start of the study period, $T$ (unit: time) is average advective transport time from the legacy zones to catchment outlet, and $k$ is average relative release rate (unit: per time; with $1/k$ (unit: time) quantifying a characteristic time

![FIGURE 1](image-url) (a) Schematic illustration of regression line types for different types of (active, legacy, or mixed) nutrient sources, and (b) locations and numbering of the 37 most near-coastal monitoring points over Sweden included in this study. In (b), nutrient concentration and water discharge are measured in close proximity for the ‘load set’ (red circles), while for the ‘load-Est set’, discharge is measured at a more upstream station (blue stars) than the concentration measurement (green triangles)
scale until total source depletion under assumed zeroth-order release kinetics $C^* = C_0(1 - kt)$ where $t$ is running time).

With $C_0$, $n$, $k$ and $T$ assumed not to vary much in time around their respective average value, Equation (2) implies relatively stable average $C_{out-L}$ under variable discharge $Q_{out}$. The expected type of average $L_{out-L}$ behaviour versus $Q_{out}$ would then be a line with slope $(C_0kT/n)$ and intercept near zero (case A, brown line, Figure 1(a)) or negative (case B, blue line, Figure 1(a)). These legacy cases can also be checked against data-given $L$-$Q$ regression lines, and if the B case emerges from this, it would indicate that legacy nutrient release occurs first at, and above a minimum threshold $Q_{out}$ value (solid part of blue line in Figure 1(a)).

2.1.3 | Mixed sources

The typology of mixed active and legacy sources can finally be quantified in terms of output load as:

$$L_{out-M} = L_{out-L} + L_{out-A} = C_{out-L}Q_{out} + C_{out-A}Q_{out} = (1 - \gamma)L_{out-M} + \gamma L_{out-M}$$ (3)

where $\gamma = L_{out-A}/L_{out-M}$ is a dimensionless fraction ($0 \leq \gamma \leq 1$) quantifying the relative active source contribution to total load ($L_{out-M}$). The expected type of $L_{out-M}$ behaviour versus $Q_{out}$ is thereby also a line (purple in Figure 1(a)), with slope $S_M = C_{out-L}$ and intercept $I_M = C_{out-A}Q_{out}$ for $I_M \geq 0$. This expectation can also be checked against data-given $L$-$Q$ regression lines for various monitoring stations (Figure 1(b)), as outlined in the following section.

2.2 | Data

As a concrete case study for quantification and testing of the source typology outlined above, we use a dataset of coastal loads from 37 Swedish catchments draining to the Baltic Sea along the whole coastline of Sweden (Figure 1(b)). The associated 37 near-coastal monitoring stations are obtained from official Swedish environmental monitoring of streams and selected for having continuous data availability for TN and TP concentrations ($C_{out}$) over the study period 2003–2013 (see Data Availability Statement at the end of the main text for all data sources). We refer to this set of coastal measurement points as the ‘Total set’, of which only 19 have also closely associated measurements of water discharge ($Q_{out}$). By this we mean a $Q_{out}$ measurement close enough to the concentration $C_{out}$ data point to allow direct quantification of nutrient load as $L_{out} = Q_{out}C_{out}$. The set of only 19 coastal data points that fulfils this condition is referred to as the ‘Load set’. For the remaining 18 data points, we apply discharge-upscaling to extend the number of data points to the total 37 concentration points, and refer to the additional set of 18 data points as the ‘Load-est set’. For the ‘Load-est set’, we upscale discharge $Q^*$ measured upstream of the catchment outlet (where $C_{out}$ is measured) as $Q_{out} \approx Q^* A_c /A^*$, where $Q_{out}$ is the estimated discharge at the outlet

with total contributing catchment area $A_c$ and $A^*$ the smaller (sub) catchment area contributing to $Q^*$. The locations of the ‘Load set’, ‘Load-est set’ and ‘Total set’ data points are shown in Figure 1(b).

2.3 | Typology testing and quantification based on data

Based on the available monitoring data for water discharge ($Q_{out}$), and nutrient concentrations ($C_{out}$) and loads ($L_{out} = Q_{out}C_{out}$) for TN and TP, we test the expected type of $L_{out}$ behaviour versus $Q_{out}$ and assess the possible source dominance for each nutrient within each catchment (Figure 1(b)). This is done based on the regression line of $L_{out}$ versus $Q_{out}$ for TN and TP in each catchment, by quantifying the deviation of each regression line intercept from zero. A near-zero or negative intercept implies dominant legacy sources, and the deviation of a negative intercept from zero distinguishes legacy case B from legacy case A (blue line and brown line, respectively, Figure 1(a)). A large enough positive intercept implies instead dominant active sources if the line slope is negligible (green line, Figure 1(a)) or mixed sources if the line slope is considerable (purple line, Figure 1(a)).

Three alternative approaches are used to quantify the regression line deviation from zero intercept. These consider the shortest normalized line distance from the origin along: (i) both the vertical axis (for load $L$) and the horizontal axis (for discharge $Q$) (Figure 2); (ii) only the vertical ($L$) axis; and (iii) only the horizontal ($Q$) axis. An $L$-$Q$ regression line is then classified as legacy-dominated type A (requiring zero intercept, Figure 1(a)) only if all three approaches indicate it as such, and as one of the other source types if indicated as such by any one of the three approaches.
FIGURE 3  Regression lines for total nitrogen (TN) load ($L_{out}$, vertical axis) versus water discharge ($Q_{out}$, horizontal axis) and associated source attribution based on data for the “Total set” of stations (Figure 1(b)). Both $L_{out}$ and $Q_{out}$ are normalized with their respective average values over the study period 2003–2013; the value 1 on either axis thus represents the average value of that variable.

FIGURE 4  Regression lines for total phosphorus (TP) load ($L_{out}$, vertical axis) versus water discharge ($Q_{out}$, horizontal axis) and associated source attribution based on data for the “Total set” of stations (Figure 1(b)). Both $L_{out}$ and $Q_{out}$ are normalized with their respective average values over the study period 2003–2013; the value 1 on either axis thus represents the average value of that variable.
For the catchments that the data indicate as having dominant active sources or mixed sources with considerable active contributions, Equation (1) quantifies the associated average load level ($L_{out}/C_0A$), from which the flux-average concentration ($C_{out}A/C_0 = L_{out}/C_0A$) is also estimated. For the catchments that the data indicate as having dominant legacy sources or mixed sources with considerable legacy contributions, Equation (2) quantifies the associated average concentration level ($C_{out-\lambda}$). For catchments indicated by data to have mixed sources, the active contribution fraction ($\gamma$) can be estimated from the measured total load ($L_{out-M}$) and the regression line intercept ($I_\lambda$) as:

$$\gamma = \frac{I_M}{L_{out-M}} \text{ for } I_M \geq 0$$  (4)
### TABLE 1  Summary of source type classification for total nitrogen (TN) and total phosphorus (TP) based on the results from the three approaches shown in Figure 5

| Classification based on regression line distance from origin: | TN | TP |
|-------------------------------------------------------------|----|----|
| Along both the L axis and the Q axis (Figure 2)             | 4, 5, 13, 36 | 9, 12, 35, 36 |
| Along just the L axis                                       | 36 | 4, 9, 11, 12, 15, 25, 35, 36, 37 |
| Along just the Q axis                                       | 4, 5, 13, 33 | 5, 14, 19, 33 |
| Union of the above, final source type classification         | 36 | 4, 9, 11, 12, 15, 25, 35, 36, 37 |

### FIGURE 6  Statistics of the coefficient of determination ($R^2$) for the best fit regression lines of nutrient load versus discharge for the different sources of: (a) total nitrogen (TN; Figure 3); and (b) total phosphorus (TP; Figure 4). The boxplots show the median (line) and associated interquartile (box) and total (whiskers) ranges, and the red + symbol in (b) shows an outlier value.

### FIGURE 7  (a) Spatial distribution of different source types for total nitrogen (TN). Calculated TN source concentrations for (b) legacy sources, and (c) currently active sources. (d) Contribution fractions ($\gamma$) for the active sources in the mixed-source catchments.
The corresponding legacy load contribution fraction \((1-\gamma)\) can also be estimated from the regression line slope \(S_M\) as:

\[
1-\gamma = \frac{S_M Q_{out}}{L_{out} M}
\]  

For all data points and their associated catchments (Figure 1 (b)), we also map and analyse the spatial distribution of the nutrient sources of different types and their quantified contributions to the monitored concentrations over the study period 2003–2013.

To assess possible distinct characteristics of catchments with different source typology, we further quantify and compare the statistics of catchment area and main hydro-climatic conditions for the different types of catchments. To assess possible main drivers of the
concentration contributions from the different types of sources, we quantify the correlation of the $C_{\text{out}}$ contributions with agricultural land share and population density as key indicators of nutrient-related human activity in the catchments (Levi et al., 2018).

3 | RESULTS AND DISCUSSION

3.1 | Source attribution

For each coastal measurement station (Figure 1(b)), regression lines are fitted to the available $L_{\text{out}}$ and $Q_{\text{out}}$ data. Figures 3 and 4 show the data plots and resulting regression lines for TN and TP, respectively; the $L_{\text{out}}$ and $Q_{\text{out}}$ data are normalized with their respective average values over the study period 2003–2013. The deviation of each regression line intercept from zero is further quantified to determine the implied source type. A near-zero or negative intercept implies dominant legacy sources, and the deviation of a negative intercept from zero distinguishes case B from case A of this source type (blue line and brown line, respectively, Figure 1(a)). A large enough positive intercept instead implies dominant active sources if the line slope is negligible (green line, Figure 1(a)) or mixed sources if the line slope is considerable enough (purple line, Figure 1(a)).

For TN, the combined results from the three alternative approaches used to quantify the regression line deviation from zero intercept identify 33 catchments as legacy-dominated (32 as case A and 1 as case B), 1 catchment as active source-dominated, and 3 catchments as having mixed sources (Figure 5, Table 1). For TP, 33 catchments emerge as legacy-dominated (24 as case A and 9 as case B). No catchment exhibits active source dominance for TP, and
4 catchments emerge as having mixed sources. For both TN and TP, Figure 6 further shows that the coefficient of determination ($R^2$) for the regression lines that indicate dominant legacy sources is overall considerably greater (mostly in the interval 0.7–0.9) than for the other source types ($R^2$ near zero for active source lines, and <0.5 for mixed source lines).

### 3.2 | Source quantification and mapping

Figures 7 and 8 show the spatial distributions of the different source types, and the quantification of the associated nutrient concentrations and active source fraction $\gamma$ for TN and TP, respectively. The legacy source concentrations differ greatly among stations, but largely increase from the northern to the southern parts of Sweden for both TN and TP (Figures 7(b) and 8(b)). The southern parts, with greater legacy source concentrations for both TN and TP also have greater agriculture share and population density.

For both TN and TP, the active source concentrations and their spatial variations are smaller overall than for the legacy sources (Figures 7 and 8). For the catchments with mixed sources, the active source fractions ($\gamma$) are in the range 0.32–0.45 for TN and 0.33–0.59 for TP. Even in mixed-source catchments (3 for TN, Figure 7(d); 4 for TP, Figure 8(d)), the currently active sources thus contribute on average less than the legacy sources to the total monitored concentrations ($C_{\text{out}}$).

Change trends over the study period 2003–2013 (Figure 9) show that the TN concentrations are decreasing in all active-source and mixed-source catchments, and they do so at higher relative rates than in the legacy-source catchments with decreasing trends (19 of the total 33 legacy-source catchments). The TN concentration in the single legacy-source catchment with dominant active sources exhibits the greatest relative decrease rate, while the change trends for the legacy TN concentrations are on average near zero. For TP, 3 of the 4 catchments with mixed sources have decreasing trends while 1 catchment has an increasing trend. Overall, trend results show that nutrient concentration contributions from active sources mostly decrease and do so relatively fast, while those from legacy sources have relatively weak decreasing or increasing trends that are, on average, near zero. This result supports the source attribution, since concentration contributions from active sources should respond faster than those from legacy sources to mitigation measures, which have been taken in Sweden to mitigate nutrient loads to the Baltic Sea, targeting known active sources with insufficient results (Destouni et al., 2017).

### 3.3 | Source relationships with hydro-climatic and human activity conditions

In terms of hydro-climatic conditions, the catchments with dominant legacy sources have on average somewhat higher temperature and lower precipitation and runoff (the latter two for TN, while the different catchment types are on average more similar in these respects for TP) than those with active and mixed sources (Figure 10). These differences are mainly due to the spreading of the many legacy-source catchments over the whole of Sweden, and their relatively greater prevalence in the warmest and driest southern parts, where agricultural land share and population density are also greater than in the north. Overall, the legacy-source catchments are on average smaller than the catchments with active and mixed sources. This size

![Figure 11](https://example.com/figure11.png)

**Figure 11** Legacy source concentration ($C_{\text{out}}/C_0$, Equation (1)) of total nitrogen (TN) and total phosphorus (TP) versus: (a) agricultural land share; and (b) population density in the catchments with dominant legacy sources. Regression lines are fitted to all data points of case a legacy sources. Table 2 shows the associated coefficients of determination ($R^2$) for the case a legacy data points in each of the sub-datasets ‘load set’ and ‘load-Est set’ (Figure 1(b))
The results of dominant legacy sources, which are diffuse with regard to both time and space, and as such difficult to identify and mitigate, are consistent with previous reports of difficulties to achieve water quality goals (Destouni et al., 2017; Van Meter et al., 2018) in spite of many actions taken for such improvement in the Baltic region (Iho et al., 2015; Linke et al., 2014) and other parts of the world (D’Elia et al., 2019; Karydis & Kitsiou, 2012; Le Moal et al., 2019; Rabotyagov et al., 2014). Our results add to and support other research suggesting legacy stores of nutrients as a key reason for common failure to improve water quality after implementation of mitigation measures (Ardón et al., 2017; Ascott et al., 2017; Darraç et al., 2008; Destouni & Jarsjö, 2018; Poffenbarger et al., 2018; Puckett et al., 2011; Van Meter et al., 2017).

The correlation of legacy concentrations with agricultural land share and population density indicates the nutrient legacies found in the present multi-catchment analysis as largely anthropogenic. The essentially unchanging legacy concentrations are consistent with nutrient mitigation measures commonly focusing on active sources, and leaving legacy sources largely unmitigated. The overall greater legacy than active concentration contributions are also consistent with the legacy sources reflecting an unmitigated nutrient accumulation from past to present, while the active sources represent only current annual inputs. Various data-based indications thus converge in indicating legacy sources as often dominant, mainly anthropogenic, and relatively unchanging.

These converging indications imply a general need for differentiated mitigation strategies and measures to effectively target different source types. Identification and targeting of active sources, where water quality responses to local mitigation measures can be relatively fast, is needed to achieve relatively fast water quality improvements and meet shorter-term regulatory goals. For longer-term, large-scale water quality improvements, the more impactful legacy sources must also be identified and targeted with appropriate mitigation strategies and measures. The quantitative typology approach developed and tested in this study is general, transferable, and can help identify, map, and target active and legacy sources of nutrients (and other hydrochemical constituents) for both shorter- and longer-term water quality improvement.

### ACKNOWLEDGEMENTS

This research was supported by the European Commission (project COASTAL, 773782).

### CONFLICT OF INTEREST

The authors declare no conflict of interest.

### AUTHOR CONTRIBUTIONS

Conceptualization, methodology development, and writing of the paper by Georgia Destouni and Yuanying Chen; data compilations by

---

#### TABLE 2

Average coefficient of determination ($R^2$) for regression lines between legacy source concentration of total nitrogen (TN) and total phosphorus (TP) versus agricultural land share or population density in the legacy catchments of the total data set (Figure 11)

| Set of stations from Figure 1(b) | TN | TP |
|----------------------------------|----|----|
|                                 | Agricultural land share | Pop density | Agricultural land share | Pop density |
| Load set                        | 0.88\*                     | 0.57\*       | 0.60\*                     | 0.57\*       |
| Load-est set                    | 0.10                         | 0.33\*       | 0.28                         | 0.56\*       |
| Total                           | 0.65\*                     | 0.50\*       | 0.52\*                     | 0.56\*       |

*significant at 5% significance level.

---

The difference is consistent with parallel source attribution findings for metals and chloride (Destouni et al., 2021), and may be due to larger catchments allowing for greater source variation within them.

Figure 11 finally shows relatively strong correlation between the quantified legacy source concentrations ($\text{C}_{\text{est.}}$) of TN and TP and the human activity indicators of agricultural land share and population density in the contributing catchments. The catchments with mixed sources or dominant active sources are too few to evaluate corresponding correlations also for those types of sources for each nutrient. The correlation results for the legacy sources indicate that they are largely anthropogenic rather than natural backgrounds. This is consistent with a high current human activity level commonly reflecting a relatively high human activity level also at earlier times, such as several decades previously. The strongest legacy concentration correlations (greatest $R^2$ values) are with agricultural land share for TN (Figure 11(a)), and with population density for TP (Figure 11(b)). The correlations with population density are relatively robust for both TN and TP when considering the different sets of stations from Figure 1(b), the ‘Load set’ with more reliable data, the more approximate “Load-est set”, or the combined total set of stations (Table 2).

The legacy concentration correlations with agricultural land share are more sensitive to the choice of station set, in particular for TN with $R^2$ increasing to 0.88 for the ‘Load set’ from 0.65 for the total set of stations. For TP, $R^2$ for the legacy concentration correlation with population density increases to 0.60 for the ‘Load set’ from 0.52 for the total set of stations.

---

### 4 CONCLUSIONS

This study has developed a nutrient source typology based on load versus discharge behaviour, which can be used to distinguish and quantify active and legacy source contributions to monitored nutrient loads. The case study results for 37 Swedish catchments indicate dominant TN and TP legacy sources in 33 of the catchments. Dominant active sources are indicated in 1 catchment for TN, while mixed sources are indicated in 3 catchments for TN, and 4 catchments for TP. The quantified TN and TP concentration contributions are higher for legacy than for active and mixed sources. The legacy source contributions also correlate well with human-activity indicators of agricultural land share for TN ($R^2 = 0.65$) and population density for TP ($R^2 = 0.56$), and they change less over time than active and mixed source contributions.

---

**CONFLICT OF INTEREST**

The authors declare no conflict of interest.

**AUTHOR CONTRIBUTIONS**

Conceptualization, methodology development, and writing of the paper by Georgia Destouni and Yuanying Chen; data compilations by...
Yuanying Chen, Carmen Prieto and Romain Goldenberg; load calculations by Yuanying Chen; data analysis by all coauthors; visualization by Yuanying Chen and Georgia Destouni; supervision by Georgia Destouni.

**DATA AVAILABILITY STATEMENT**

The data for water discharges are available from the Swedish Meteorological and Hydrological Institute (SMHI), https://vattenweb.smhi.se/station/, as are also the data for precipitation, https://www.smhi.se/data/meteorologi/nederbord, and air temperature, https://www.smhi.se/data/meteorologi/temperatur. The data for nutrient concentrations are available from the Swedish University of Agricultural Sciences (SLU), Department of Water and Environment, Data host for inland waters. https://www.slu.se/institutioner/vatten-miljo/datavardskap/.

The data for agricultural land cover are available from CORINE Land Cover (CLC), https://land.copernicus.eu/pan-european/corine-land-cover. The data for population density are available from NASA Socio-economic Data and Applications Center (SEDAc), https://sedac.ciesin.columbia.edu/data/set/gpw-v4-population-density-adjusted-to-2015-unwpp-country-totals-rev11.

**ORCID**

Yuanying Chen https://orcid.org/0000-0002-2856-5535

Georgia Destouni https://orcid.org/0000-0001-9408-4425

Romain Goldenberg https://orcid.org/0000-0002-3549-3453

Carmen Prieto https://orcid.org/0000-0002-9073-8185

**REFERENCES**

Ardón, M., Helton, A. M., Scheurell, M. D., & Bernhardt, E. S. (2017). Fertilizer legacies meet saltwater incursion: Challenges and constraints for coastal plain wetland restoration. *Elementa: Science of the Anthropocene*, 5, 41.

Ascott, M. J., Goody, D. C., Wang, L., Stuart, M. E., Lewis, M. A., Ward, R. S., & Binley, A. M. (2017). Global patterns of nitrate storage in the vadose zone. *Nature Communications*, 8(1), 1–7.

Basu, N. B., Destouni, G., Jawitz, J. W., Thompson, S. E., Loukinova, N. V., Darraç, A., Zanardo, S., Maier, M., Sivapalan, M., Rinaldo, A., & Rao, P. S. C. (2010). Nutrient loads exported from managed catchments reveal emergent biogeochemical stationarity. *Geophysical Research Letters*, 37, L23404.

Chant, J. G., & Yang, G. (2018). Exploring drivers of regional water-quality change using differential spatially referenced regression—A pilot study in the Chesapeake Bay catchment. *Water Resources Research*, 54, 8120–8145.

Conley, D. J., Björck, S., Bonsdorff, E., Carstensen, J., Destouni, G., Gustafsson, B. G., & Müller-Karulis, B. (2009). Hypoxia-related processes in the Baltic Sea. *Environmental Science & Technology*, 43(10), 3412–3420.

Cvetkovic, V., Carstens, C., Selroos, J.-O., & Destouni, G. (2012). Water and solute transport along hydrological pathways. *Water Resources Research*, 48, W06537.

Darraç, A., Lindgren, G., & Destouni, G. (2008). Long-term development of phosphorus and nitrogen loads through the subsurface and surface water systems of drainage basins. *Global Biogeochemical Cycles*, 22, GB3022.

D’Elia, C. F., Bidjerano, M., & Wheeler, T. B. (2019). Population growth, nutrient enrichment, and science-based policy in the Chesapeake Bay watershed. In *Coasts and estuaries* (pp. 293–310). Elsevier.
Puckett, L. J., Tesoriero, A. J., & Dubrovsky, N. M. (2011). Nitrogen contamination of surficial aquifers—A growing legacy. *Environmental Science & Technology, 45*, 839–844.

Quin, A., Jaramillo, F., & Destouni, G. (2015). Dissecting the ecosystem service of large-scale pollutant retention. *Ambio, 44*(1), 127–137.

Rabotyagov, S. S., Kling, C. L., Gassman, P. W., Rabalais, N. N., & Turner, R. E. (2014). The economics of dead zones: Causes, impacts, policy challenges, and a model of the Gulf of Mexico hypoxic zone. *Review of Environmental Economics and Policy, 8*(1), 58–79.

Schuwirth, N., Honti, M., Logar, I., & Stamm, C. (2018). Multi-criteria decision analysis for integrated water quality assessment and management support. *Water Research X, 1*, 100010.

Selroos, J. O., & Destouni, G. (2015). Influence of spatial and temporal flow variability on solute transport in catchments. *Hydrological Processes, 29*(16), 3592–3603.

Sharpley, A., Jarvie, H. P., Buda, A., May, L., Spears, B., & Kleinman, P. (2013). Phosphorus legacy: Overcoming the effects of past management practices to mitigate future water quality impairment. *Journal of Environmental Quality, 42*(5), 1308–1326.

Van Meter, K. J., & Basu, N. B. (2015). Catchment legacies and time lags: A parsimonious catchment model to predict the effects of legacy storage on nitrogen export. *PLoS One, 10*(5), e0125971.

Van Meter, K. J., Basu, N. B., & Van Cappellen, P. (2017). Two centuries of nitrogen dynamics: Legacy sources and sinks in the Mississippi and Susquehanna River basins. *Global Biogeochemical Cycles, 31*(1), 2–23.

Van Meter, K. J., Van Cappellen, P., & Basu, N. B. (2018). Legacy nitrogen may prevent achievement of water quality goals in the Gulf of Mexico. *Science, 360*(6387), 427–430.

Wurtsbaugh, W. A., Pael, H. W., & Dodds, W. K. (2019). Nutrients, eutrophication and harmful algal blooms along the freshwater to marine continuum. *Wiley Interdisciplinary Reviews: Water, 6*, 706.