An Updated MANAGE: a database for exploring factors affecting nutrient loss from agriculture fields

# Stephanie A. Nummer1, Daren R. Harmel2, Song S. Qian1\*

1Department of Environmental Sciences, The University of Toledo, Toledo, OH 43606

2 Grassland Soil and Water Research Laboratory, USDA-ARS, Temple, TX 76502

# Core Ideas

# The October 2014 edition of MANAGE is a compilation of 65 publications – the largest dataset since it’s creation.

# The MANAGE database is comprised of observational data and this should be considered when applying statistics.

# There are numerous different field characteristics included in the Manage Database for analysis.

# Abstract

Fertilizers are used to increase crop yields, but they also increase pollution in nearby water bodies through agricultural runoff, putting water bodies at higher risk for eutrophication. The Measured Annual Nutrient loads from AGriclutural Environments (MANAGE) database provides field scale observational data about nutrient losses and conservation practices. We summarize the October 2014 update to MANAGE and identify variables in need of additional updates. This edition includes 65 publications that come from 20 different states and 6 Canadian provinces. Two variables in MANAGE have limited sample sizes available for analysis: crop yields (16% of entries have data) and conservation practices (19% of entries).

Keywords: MANAGE, nutrient loss, agriculture, conservation practice, nutrient loading

# Introduction

A variety of watershed models are used to track nutrient transportation and rely on data about land-use practices that promote or inhibit nutrient export to natural water bodies (Sharpley et al. 2002). Nutrient tracking watershed models need detailed site-level characteristic descriptions of soil type, crop type, and precipitation, which influence nutrient transportation (Robertson and Saad 2011). The USDA created the Measured Annual Nutrient loads from AGricultural Environments (MANAGE) database to provide this needed field scale data for such models (Harmel et al. 2006, Harmel et al. 2008).

MANAGE includes data from published studies with land use, rainfall quantities, soil loss data, and other site characteristics. Since its creation, there have been 5 updates to the database and 3 publications reporting the database and its updates. These publications showcase the new additions made to MANAGE and subsequent areas needing further research (Harmel et al. 2006). In addition, the MANAGE database, specifically the Drain Load table in MANAGE, has been used to review load data from fields with different drainage types (Christianson and Harmel 2015). One goal of the MANAGE project is to provide measured nutrient loading data from agriculture fields to the public (Harmel et al. 2006, Harmel et al. 2008). Additionally, MANAGE can be used to assist farmers and managers in determining the field management practices that will have the greatest benefit for stakeholders (Harmel et al. 2006, Harmel et al. 2008).

The current version of the MANAGE database, updated October 2014, includes 65 studies, increased from 55 in the previous version. The 65 studies in MANAGE represent 1,980 watershed years; the watershed year variable is the total number of years monitored for each individual entry – for example a field monitored for 9 years would have 9 watershed years. Conversely, an entry in the MANAGE database represents an observation from a single field. Thus, one entry is a single field while the watershed years for that field is the number of years the field was observed for. A publication in MANAGE may contribute more than one entry because the study may have monitored more than one field. Similar to the third version of MANAGE, the October 2014 update also has the nutrient concentration data for each entry. The 10 added studies in the October 2014 update increase the available sample size of concentration data and other variables with a large percentage of missing values. The goal of this communication is to provide a detailed summary of the October 2014 edition of MANAGE, including the improvements over the previous edition and the limitations, so that future users are aware of the MANAGE database and the advantages and restrictions that come with the data available.

# Methods

We used exploratory data analysis (Tukey, 1978) methods to describe the data through various summary statistics and exploratory graphs. Summary statistics, such as sample size, sample mean, median, variance and number of missing values for each variable, provide an overall picture of data availability and distribution. In addition, we also calculated area-weighted mean as measure of the total loads (with a unit of kg/ha/yr). All analyses were performed in R (R Core Team 2015),especially functionalities provided by the R package reshape2 (Wickham 2007).

Spatial attributes were mapped using ArcGIS 10.2 (ESRI 2013) to provide a visual display of the geographic distribution of the data in MANAGE.

# Results

The October 2014 edition of MANAGE includes data from 65 publications (10 more than the previous edition), a total of 1,980 observations (watershed years). The Load table in the MANAGE database, the table containing a majority of the variables considered in this analysis, includes 99 different categories of information (variables).

The area-weighted average of total nitrogen (TN) loading is 12.8 kg ha-1yr-1 and 2.1 kg ha-1 yr-1 for total phosphorus (TP). Both calculations were taken after missing values of TN, TP, and field size were removed. Compared to the area-weighted TN and TP loadings calculated from the 2007 edition of MANAGE (14.2 and 2.1 kg ha-1yr-1, respectively), the current update included fields with slightly lower TN loading and comparable TP loading.

The geographic balance of all editions of MANAGE, including the October 2014 edition, is not uniform. The data is concentrated in the central region of United States, where a large portion of the agriculture industry is located. Because agriculture land distribution in the U.S. is spatially uneven, data included in all versions of MANAGE are also spatially uneven (Table 1). Although the current MANAGE has the widest spatial coverage among all versions of MANAGE, data are still primarily from middle part of the continental U.S., along the Mississippi River basin (Figure 1).

 Many variables in MANAGE have significant amount of missing values (Table 2). These calculations show variables that have a suitable samples size for analysis and others that may act to limit analyses. Two variables with limited sample size that can raise concerns when using them in analyses include conservation practices and crop yields.

The number of entries containing crop yield data in the MANAGE database are limited and can limit the sample size when using crop yields in analyses. This information is important when considering agricultural economics and nutrient uptake by plants (Vagstad et al. 1997). To better understand the crop yield information that is available the average crop yield for the primary crop on each field was calculated (Table 3). Of the entries with crop yield data, potato has the highest average crop yield. We note that the interpretation of yield is crop-specific, as well as region-specific. In the current form, crop yield cannot be directly used to assess the effect of management practices on yield.

Considering the limited data available for crop yields, the next logical step was to review the fertilizer information included in MANAGE. Fertilizer application type and timing are important to consider because fertilizer application can impact nitrogen and phosphorus losses from agricultural fields. Precipitation and irrigation events carry nutrients from the field when they drain and make it important to consider the method and the timing of application. The fertilizer application timing (Table 4) and methods (Table 5) in MANAGE are summarized below. For both the fertilizer application timing and methods there is over half of the watershed years and entries with data available for analysis.

Data regarding conservation practices are also limited and can thus limit sample size when performing analyses on conservation practice. Conservation practices can impact the nutrient loads. The primary conservation practices considered in the MANAGE database include contour farming, filter strip, terrace, and grassed waterways. These conservation practices account for 17 percent of watershed years or 19 percent of entries in MANAGE. The other 83 percent of watershed years or 81 percent of entries do not have a conservation practice listed, which can mean either that the entries do not have a conservation practice applied to the field or that the publication did not list whether or not a conservation practice was implemented. While the former is unambiguous that there was no conservation practice, the latter is uncertain on the presence of a conservation practice. Ideally, a blank cell on conservation practice should have two separate categories when additional information is available (Duriancik et al. 2008).

Current and past editions of the MANAGE database have provided field scale information on agricultural fields to public, but a number of the variables within the database are correlated with each other and can influence analyses (Harmel et al. 2006, Harmel et al. 2008). One particular of variable that exemplifies the trend is the amount of phosphorus or nitrogen applied to a field as it relates to whether the field has a conservation practices. Fields with high fertilizer applications are more likely to have a conservation practice implemented (Figure 2). Given that a high fertilizer application rate is also linked to nutrient loss from the field, when studying the effect of conservation practice on nutrient loadings, fertilizer application rate is a confounding factor that must be controlled. Failure to control confounding factor may lead to misleading results. For example, for both phosphorus and nitrogen, the means of nutrient loads are greater for fields with conservation practices than for fields without conservation practices (Figure 3).

The current edition has additional attributes, including crop yield, that have a low percentage of available data and can thus limit the sample size available for analyses. Additionally, current and past editions have variables that need to be consider carefully because they can act as confounding factors. Although this is the case, the most recent update also has the highest number of publications and watershed years included in any edition of MANAGE thus far.

# Discussion

Past analyses of MANAGE have yielded counterintuitive results that can be partially attributed to confounding factors that were not accounted for (Harmel et al. 2006, Harmel et al. 2008). These confounding factors, which are factors that are not the independent variable but influence the outcome of the dependent variable, include field characteristics such as land use, soil characteristics, precipitation, and seasonal variations. In addition to confounding factors not considered previously, there are a number of attributes in the MANAGE database that have limited data available. As noted by previous publications, there is a definite gap in the amount of concentration data versus the quantity of nutrient load data (Harmel et al. 2008). Additionally, crop yield data are limited. Only 16 percent of the data in the October 2014 edition of MANAGE possess crop yield quantities. Furthermore, available yield data cannot be compared directly as yield varies by crop in the same region and by region for the same crop. As a result, the current MANAGE may not be appropriate for assessing the effect of various best management or conservation practices on crop yield.

 The MANAGE database includes only data from peer-reviewed publications. Although publication bias, as a result of selective publication of studies with statistically significant results, can be a concern, the studies included in MANAGE were not designed for hypothesis testing. As a result, there is not a selective publication process based on statistical significance. Nevertheless, we feel that the inclusion of grey literature (e.g., internal reports from relevant government and industrial research institutes) can not only increase sample size for subsequent statistical analysis, but also assess the potential of a publication bias (Easterbrook et al. 1991).

 The MANAGE database was initially designed to provide field scale data for assessing watershed model parameters. Support for these models are important because they evaluate and quantify the nutrients input to receiving water bodies (Sharpley 1995, Vagstad et al. 1997, Crain et al. 2006). However, MANAGE is an “observational data,’’ where random assignment of treatment (e.g., conservation practice, best management practice) is absent (Cochran and Rubin 1973) . In a randomized experiment, factors, other than the treatment, that may influence the response (confounding factors, e.g., nutrient loss) are balanced through randomization. The process of randomization ensures that the estimated treatment effect is unbiased. Without the randomization process, the estimated treatment effect can be biased. As a result, analysis of observational data for causal inference should use additional method (e.g., multilevel modeling and sub-setting data using propensity scores, (Qian and Harmel 2015).

 Overall, the MANAGE database is a valuable tool available to the public to aid decision makers and managers. The public availability and ease of use associated with MANAGE makes it ideal for a wide variety of uses including model development and field scale queries. Previous analyses of the database have noted that the confounding factors that vary from field to field need to be accounted for when finding causal effects. Accounting for these will allow for a more comprehensive analysis and can influence decisions made by farmers and policy makers.

# Acknowledgments

This work was supported by a grant from International Plant Nutrition Institute to SSQ. Comments and suggestions from Jonathan Bossenbroek are appreciated.

# References

Christianson, L., and R. Harmel. 2015. The MANAGE Drain Load database: Review and compilation of more than fifty years of North American drainage nutrient studies. Agricultural Water Management **159**:277-289.

Cochran, W. G., and D. B. Rubin. 1973. Controlling bias in observational studies: A review. The Indian Journal of Statistics **35**:417-446.

Crain, A. S., K. D. o. Agriculture, and G. Survey. 2006. Concentrations, and Estimated Loads and Yields of Nutrients and Suspended Sediment in the Little River Basin, Kentucky, 2003-04. U.S. Geological Survey.

Duriancik, L. F., D. Bucks, J. P. Dobrowolski, T. Drewes, S. D. Eckles, L. Jolley, R. L. Kellogg, D. Lund, J. R. Makuch, M. P. O'Neill, C. A. Rewa, M. R. Walbridge, R. Parry, and M. A. Weltz. 2008. The first five years of the Conservation Effects Assessment Project. Journal of Soil and Water Conservation **63**:185A-197A.

Easterbrook, P. J., R. Gopalan, J. Berlin, and D. R. Matthews. 1991. Publication bias in clinical research. The Lancet **337**:867-872.

ESRI. 2013. ArcGIS Desktop. Environmenal Systems Research Institute, Redlands, CA.

Harmel, D., S. Potter, P. Casebolt, K. Reckhow, C. Green, and R. Haney. 2006. Compilation of Measured Nutrient Load Data for Agricultural Land Uses in the United States JAWRA Journal of the American Water Resources Association **42**:1163-1178.

Harmel, D., S. Qian, K. Reckhow, and P. Casebolt. 2008. The MANAGE Database: Nutrient Load and Site Characteristic Updates and Runoff Concentration Data All rights reserved. No part of this periodical may be reproduced or transmitted in any form or by any means, electronic or mechanical, including photocopying, recording, or any information storage and retrieval system, without permission in writing from the publisher. J. Environ. Qual. **37**:2403-2406.

Qian, S. S., and R. D. Harmel. 2015. Applying Statistical causal Analyses to Agricultural Conservation: A Case Study Examining P Loss Impacts. Journal of American Water Resources Association (in press).

Robertson, D. M., and D. A. Saad. 2011. Nutrient Inputs to the Laurentian Great Lakes by Source and Watershed Estimated Using SPARROW Watershed Models1. JAWRA Journal of the American Water Resources Association **47**:1011-1033.

Sharpley, A. 1995. Identifying Sites Vulnerable to Phosphorus Loss in Agricultural Runoff. J. Environ. Qual. **24**:947-951.

Sharpley, A., P. Kleinman, R. McDowell, M. Gitau, and R. Bryant. 2002. Modeling phosphorus transport in agricultural watersheds: Processes and possibilities. Journal of Soil and Water Conservation **57**:425-439.

Team, R. C. 2015. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Vagstad, N., H. O. Eggestad, and T. R. Høyås. 1997. Mineral Nitrogen in Agricultural Soils and Nitrogen Losses: Relation to Soil Properties, Weather Conditions, and Farm Practices. Ambio **26**:266-272.

Wickham, H. 2007. Reshaping Data with the Reshape Package. Journal of Statistical Software **21**:1-20.

# Table and Figure Captions

**Table 1.** The top 3 states for percent of total watershed years and total entries in the MANAGE database.

**Table 2.** The number and percentage of missing entries and the associated watershed years for eight variables of interest.

**Table 3.** Average crop yield by crop type in the October 2014 edition of MANAGE.

**Table 4.** The number and percentage of total watershed years and entries per fertilizer application timings included in the October 2014 edition of MANAGE.

**Table 5.** Fertilizer Application Methods in the October 2014 update of MANAGE broken down by the number of percentage of the total watershed years and entries.

**Figure 1.** The number of watershed years per state in the October, 2014 edition of MANAGE.

**Figure 2.** Side by Side box plots of fertilizer application (Phosphorus Applied and Nitrogen Applied) for fields without a conservation practice and fields with a conservation practice. Both the left plot (Phosphorus Applied) and right plot (Nitrogen Applied) show a high mean fertilizer application in fields with conservation practices.

**Figure 3.** Side by side box plots showing the average total loads of Phosphorus (left plot) and Nitrogen (right plot) for fields with and without conservation practices.

# Tables

|  |  |  |
| --- | --- | --- |
| **State** | **% of Total Watershed Years** | **% of Total Entries** |
| OK | 25% | 23% |
| TX | 16% | 20% |
| OH | 14% | *not in top 3* |
| GA | *not in top 3* | 8% |

**Table 1.**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Variable** | **Number of Missing Entries** | **Percent of Missing Entries** | **Number of Missing Watershed Years** | **Percent of Missing Watershed Years** |
| First Fertilizer Application Method  | 127 | 38.5% | 803 | 40.6% |
| First Fertilizer Application Timing  | 136 | 41.2% | 868 | 43.8% |
| First Conservation Practice  | 268 | 81.2% | 1640 | 82.8% |
| Crop Yield | 277 | 83.9% | 1673 | 84.5% |
| Land Use | 0 | 0.0% | 0 | 0.0% |
| Tillage  | 3 | 0.9% | 9 | 0.5% |
| Average Nitrogen Applied | 114 | 34.5% | 842 | 42.5% |
| Average Phosphorus Applied | 59 | 17.9% | 466 | 23.5% |

**Table 2.**

|  |  |
| --- | --- |
| **Crop Type** | **Average Crop Yield (Mg/ha)** |
| Alfalfa | 12.4 |
| Coastal Bermuda Grass | 11.4 |
| Corn | 7.151538 |
| Cotton | 2.045 |
| Pasture | 5 |
| Potato | 27.485714 |
| Soybeans | 3.392 |
| Wheat | 2.64 |

**Table 3.**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **First Fertilizer Application Timing** | **Watershed Years** | **% of Total Watershed Years** | **Entries** | **% of Total Entries** |
| At Planting, within 1 week of plant | 245 | 12% | 50 | 15% |
| Grass at Establishment  | 38 | 2% | 6 | 2% |
| Grass in Dormant Season | 21 | 1% | 3 | 1% |
| Grass in Growing Season  | 335 | 17% | 46 | 14% |
| Other  | 14 | 1% | 7 | 2% |
| Out of Season, > 2 months before plant | 184 | 9% | 34 | 10% |
| Pre-Plant, 2 months-1 week before plant | 260 | 13% | 45 | 14% |
| Side/Top Dress, > 1 week after plant | 15 | 1% | 3 | 1% |
| Missing Value | 868 | 44% | 136 | 41% |

**Table 4.**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **First Fertilizer Application Method** | **Watershed Years** | **% of Total Watershed Years** | **Entries** | **% of Total Entries**  |
| Incorporated | 372 | 19% | 67 | 20% |
| Injected | 183 | 9% | 31 | 9% |
| Other | 18 | 1% | 3 | 1% |
| Surface Applied | 604 | 31% | 102 | 31% |
| Missing Values | 803 | 41% | 127 | 38% |

**Table 5.**

# Figures



**Figure 1.**



**Figure 2.**



**Figure 3.**