

Design and calibration of a nitrate decision support tool for groundwater wells in Wisconsin, USA

Paul F. Juckem^{a,*}, Nicholas T. Corson-Dosch^a, Laura A. Schachter^a, Christopher T. Green^b, Kelsie M. Ferin^c, Eric G. Booth^c, Christopher J. Kucharik^c, Brian P. Austin^d, Leon J. Kauffman^e

^a U.S. Geological Survey, Upper Midwest Water Science Center, 1 Gifford Pinchot Dr., Madison, WI, 53726, USA

^b U.S. Geological Survey, Water Resources Mission Area, P.O. Box 158, Moffett Field, California, 94035, USA

^c Department of Plant and Agroecosystem Sciences, University of Wisconsin – Madison, 457 Moore Hall, 1575 Linden Dr., Madison, WI, 53706, USA

^d Bureau of Drinking Water and Groundwater, Wisconsin Department of Natural Resources, 101 S. Webster St., P.O. Box 7921, Madison, Wisconsin, USA

^e U.S. Geological Survey, New Jersey Water Science Center, 3450 Princeton Pike Suite 110, Lawrenceville, NJ, 08648, USA

ARTICLE INFO

Handling Editor: Daniel P. Ames

Keywords:

Nitrate
Decision support
Groundwater
Uncertainty

ABSTRACT

This paper describes development of a nitrate decision support tool for groundwater wells (GW-NDST) that combines nitrate leaching and groundwater lag-times to compute well concentrations. The GW-NDST uses output from support models that simulate leached nitrate, groundwater age distributions, and nitrate reduction rates. The support models are linked through convolution to simulate nitrate transport to wells. Spatially distributed parameters were adjusted through calibration to 34,255 nitrate sample targets. Prediction uncertainty is illustrated via Monte Carlo realizations informed during calibration. Over 78% of target concentrations were within the simulated range of results from 450 realizations. An example forecasting scenario illustrates that a range of feasible outcomes exist and should be considered when interpreting forecasts for decision making. Uncertainty in forecasting is unavoidable; the intent of characterizing uncertainty in the GW-NDST is to facilitate decision making by increasing insight into the response of nitrate contamination to physical and chemical processes.

1. Introduction

The concentration of nitrate-N in a well is dependent upon many factors including the rate of nitrate leaching below the land surface, travel time of water through the vadose zone and aquifer material, and geochemical conditions in the aquifer. Major sources of groundwater nitrogen historically include human and animal waste (manure) and synthetic fertilizer (Byrnes et al., 2020). Nitrogen is a nutrient that promotes plant growth and is therefore an important component of crop fertilization. Excessive nitrogen fertilizer use or inefficient application timing of fertilizer and manure, however, can result in nitrate becoming a groundwater pollutant. Moreover, nitrate does not form strong bonds with soil particles, meaning that soils have limited capacity to store nitrate for subsequent plant use, especially during periods of precipitation (or irrigation) and associated infiltration (Masarik et al., 2014). Similarly, once leached below the root zone, nitrate is easily transported with groundwater over long distances through aquifers. Due to

heterogeneity of soil and aquifer properties that results in flow paths of varying rates and lengths, a single well screen or water sample contains many “parcels” of water with varying travel times. This time range can be estimated or modeled using a travel time distribution (TTD; Vogel, 1967b; Zuber, 1986; Varni and Carrera, 1998; Cook and Böhlke, 2000; Green et al., 2014). Nitrate concentrations in aquifers and wells can also be influenced by denitrification, which tends to occur at faster rates in zones with electron donors such as organic carbon or reduced inorganic iron and sulfur species (Korom, 1992).

While nitrate pollution of aquifers is a wide-spread problem, there are few readily available tools for rapid assessment of current and future nitrate concentrations in wells that can be applied across large regions (e.g.: Kourakos and Harter, 2014; León et al., 2000). Variable land use and geological heterogeneity create high degrees of spatiotemporal variability that are often difficult to capture even with detailed numerical models over small areas and pose significant challenges for simulating groundwater nitrate at large spatial scales. Regional and

* Corresponding author.

E-mail addresses: pjuckem@usgs.gov (P.F. Juckem), ncorson-dosch@usgs.gov (N.T. Corson-Dosch), lschachter@usgs.gov (L.A. Schachter), ctgreen@usgs.gov (C.T. Green), kferin@wisc.edu (K.M. Ferin), egbooth@wisc.edu (E.G. Booth), kucharik@wisc.edu (C.J. Kucharik), brian.austin@wisconsin.gov (B.P. Austin), lkauff@usgs.gov (L.J. Kauffman).

<https://doi.org/10.1016/j.envsoft.2024.105999>

Received 6 November 2023; Received in revised form 16 January 2024; Accepted 27 February 2024

Available online 12 March 2024

1364-8152/© 2024 The US Geological Survey. Published by Elsevier Ltd.

national-scale statistical and machine learning models have been used to successfully generate maps of nitrate occurrence (Nolan et al., 2018; Nolan and Hitt, 2006; Ransom et al., 2022; Mechenich and Johnson, 2022), but maps are static with limited application for evaluating future management scenarios. Fully deterministic models, such as SWAT (Neitsch et al., 2011), Agro-IBIS (Kucharik and Brye, 2003), and MODFLOW (Niswonger et al., 2011) for example, are often well suited for scenario simulations and can incorporate small-scale detail, but these models often focus on only one component of the process, such as either the nitrogen cycle or groundwater flow. Linking multiple deterministic models (Wei et al., 2019) can yield highly detailed simulations capable of incorporating complex forecast scenarios, yet the resources involved are often substantial, particularly on a “per area” basis. The problem is well suited to reduced complexity modeling, which involves distillation of a complex system into key features and processes that meet the needs for predictive capability (Sarofim et al., 2021). That is, reduced complexity models are typically more computationally efficient than complex numerical models yet incorporate adjustable model parameters that facilitate sensitivity evaluations of scenario outcomes (e.g., forecasts) involving important processes and conditions (Sarofim et al., 2021) – valuable characteristics of a decision support tool.

This paper presents the development of a tool designed to assist resource managers with assessing the potential for achieving future improvements in well water quality by reducing nitrate leaching rates near wells. The U.S. Geological Survey partnered with the Wisconsin Dept. Of Natural Resources and the University of Wisconsin-Madison to

develop a groundwater Nitrate Decision Support Tool (GW-NDST; Schachter et al., 2024a) for the state. The GW-NDST, also referred to as “the tool” in this paper, is designed to assist with the challenges of nitrate management by providing a flexible framework for improving the understanding of how changes in the amount of nitrate introduced to aquifers (nitrate leaching) translate to nitrate concentrations in individual wells. The relationship is influenced by the rate of leaching, time required for the nitrate to be transported through the unsaturated zone and the aquifer as it flows to wells, mixing of water with differing ages and nitrate concentrations captured by wells, and chemical transformation by denitrification along the flow path. The tool was developed using output from a dynamic ecosystem model (Agro-IBIS; Kucharik and Brye, 2003; Motew et al., 2017; Lark et al., 2022), analytical groundwater fate and transport methods (Maloszewski and Zuber, 1982; Green et al., 2018; Green et al., 2021), and oxygen and nitrate reduction rate estimates (Green et al., 2018; Juckem and Green, 2024) to address questions such as “how much nitrate leaching reduction would be needed in order to reduce nitrate below a specified concentration in a well of interest?” and “how long would it take before improvements from reduced nitrate leaching rates would be observed in the well?” The tool also quantifies forecast uncertainty by incorporating an ensemble of calibration parameters, spanning probable ranges, that were conditioned to over 34,000 nitrate sample targets collected across the state over multiple decades.

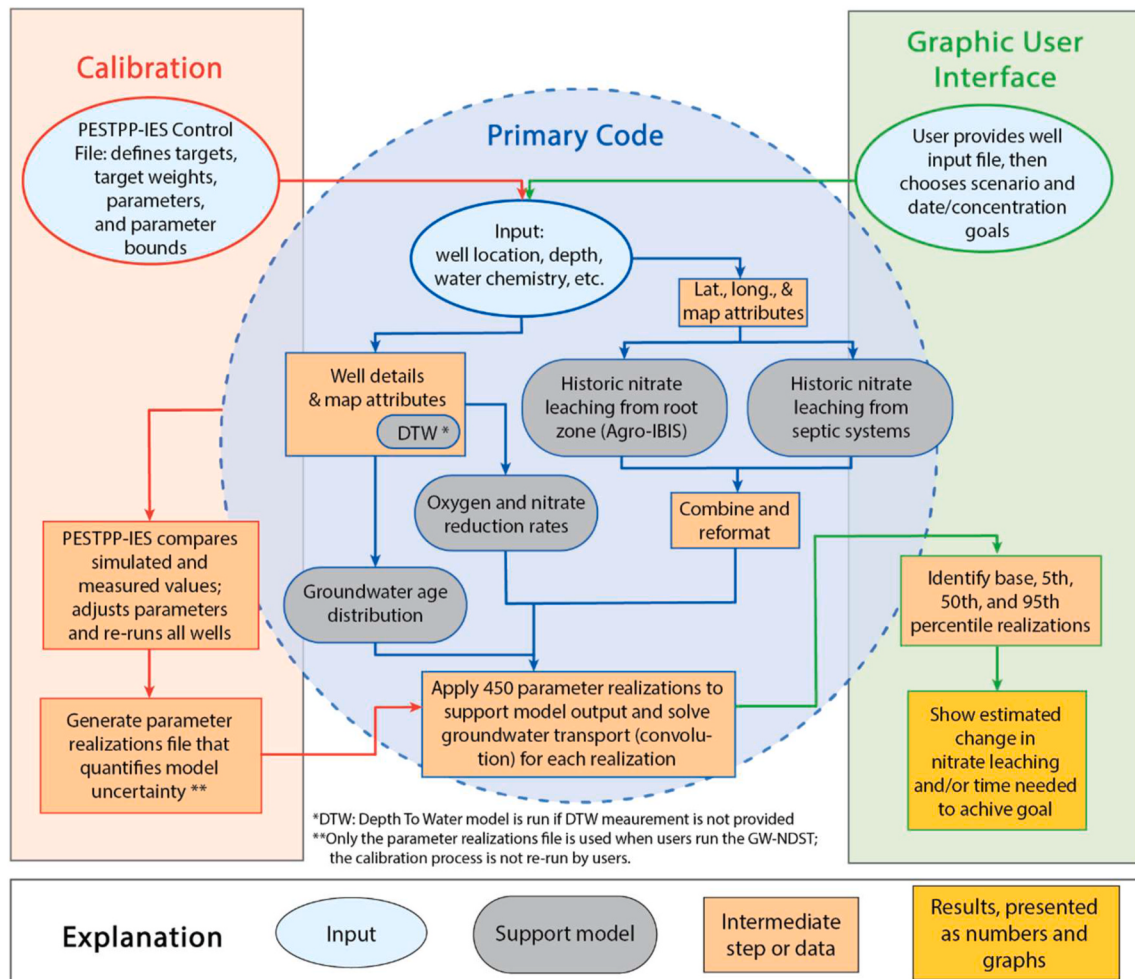


Fig. 1. Flow chart diagram of the GW-NDST for Wisconsin, highlighting the three major processes (calibration, primary code, Graphical User Interface) and how the input, support models, intermediate steps, and results fit into and connect those major processes.

2. Methods

This paper focuses on the implementation of a groundwater transport and forecasting method for nitrate concentrations in groundwater supplied wells located in the state of Wisconsin, USA. Fig. 1 illustrates in broad terms how the tool is organized and the major components of the tool's workflow. The method ingests user-supplied information for a single well, and then computes historical and future nitrate concentrations in the well using mathematical convolution (Maloszewski and Zuber, 1982) and location-specific input generated by multiple supporting models designed to simulate 1. Historical nitrate leaching, 2. Groundwater age distributions, and 3. Oxygen reduction and denitrification. The convolution equation is solved for numerous dates over time to graphically illustrate historical nitrate concentration trends in the user-specified well. An ensemble of reasonable parameter values generated from model calibration are used to assess model uncertainty in simulated results. Finally, the user can choose among six forecasting scenarios to assess potential nitrate leaching reductions needed to meet user-defined nitrate concentration goals by a user-specified future date. The GW-NDST is coded in Python (Van Rossum and Drake, 2009) and is made accessible to users in a simple Jupyter Notebook (Kluyver et al., 2016) Graphical User Interface (GUI). Readers should refer to the *Software availability* section for instructions on how to obtain and install the tool, and to Juckem et al. (2024) for access to all additional data that supported development of the GW-NDST. The following sections describe in greater detail the supporting models, the convolution method for contaminant transport, the GW-NDST calibration and uncertainty assessment, and demonstration of one forecasting scenario in the GW-NDST.

2.1. Nitrate sources

Nitrate loading to the water table in the GW-NDST is derived from two estimates of annual nitrate leaching below the root zone. One source is a dynamic ecosystem model, called Agro-IBIS (Kucharik and Brye, 2003; Motew et al., 2017), that incorporates nutrient application data for fertilizer, manure, and other sources to simulate plant growth and nutrient balances in the soil over space and time. Agro-IBIS has the ability to represent forest and grass plant functional types (PFTs) and corn, soybean, and wheat cropping systems of the Midwest USA. The other source of nitrate leaching is a decadal census-based estimate derived from septic systems that was developed for this tool (Schachter et al., 2024b).

Nitrate leached through the root zone was generated from a 116-year (1901–2016) spatially refined version of a regional-scale Agro-IBIS simulation by Lark et al. (2022). The regional study used the Agro-IBIS ecosystem model to, among other aims, simulate nitrate leaching below the root zone in agricultural areas across the Midwest USA based on historical land cover and associated fertilizer and manure applications (Lark et al., 2022). Agro-IBIS results were evaluated by comparing simulated nitrate leaching with data from a meta-analysis by Shrestha et al. (2023) and by comparing simulated crop yields with county-scale data. Historical nitrate leaching results from Agro-IBIS were desirable for use with the GW-NDST because the physically and biologically based Agro-IBIS model directly simulates crop growth (yield) using soil, meteorological, and nutrient application drivers, while tracking nitrogen flux through the physical and biological systems. Moreover, Agro-IBIS simulates these processes at hourly to daily time steps (subsequently summarized to annual time steps for the GW-NDST application) and at a 1 km by 1.5 km spatial resolution (0.1° latitude/longitude). Thus, the Agro-IBIS results are generated at a higher spatial and temporal resolution than many other potential data sources, such as county-scale fertilizer and manure estimates (Brakebill and Gronberg, 2017; Gronberg and Arnold, 2017), which were used as inputs to the model. These features of the refined Agro-IBIS model facilitated generation of detailed nitrate leaching rates over space and time

scales appropriate for a state-wide application.

A slight modification of the Agro-IBIS model results was required for application with the GW-NDST framework. Agro-IBIS is a one-dimensional model whereby computations within an individual cell of the model have no effect on adjacent cells. Additionally, open water is not simulated in the Agro-IBIS model because it is designed to simulate terrestrial systems only. The refined Agro-IBIS model of Lark et al. (2022) used a liberal method to identify open water areas in order to minimize computational problems due to expanding and contracting waterbody areas over time. That is, Lark et al. (2022) integrated multiple land use and land cover datasets that estimated changes in land cover since 1938. Individual model cells were defined as open water if the primary land classification for that cell was ever identified as open water in any of the historical land classification maps. This eliminated problems caused by land cover potentially flipping between open water and dry land during the simulation period. However, numerous homes with domestic wells are located along shorelines of lakes and other open water areas in Wisconsin. Thus, for application with the GW-NDST, nitrate leaching values were linearly interpolated across waterbodies for each individual year of the Agro-IBIS simulation using the *scipy.interpolate.griddata* method (Jones et al., n.d.) to ensure that any individual Agro-IBIS cell in which a user-supplied well is located contains a reasonable nitrate leaching history, with the assumption that leaching rates interpolated from adjacent cells adequately represent leaching rates along shorelines.

Nitrate leaching from septic systems was estimated using decadal census data and a per person nitrate excretion rate. In 1990, the Census Bureau asked respondents to indicate whether their primary place of residence was connected to a municipal sewer system or utilized a septic system for waste disposal. This is the last year in which respondents were directly asked about septic system use, thus the 1990 census data was used to estimate the percent of an area's population that relied upon septic systems; individual residence location information is not included in census data. Prior to 1990, counties were the finest resolution area with consistency between decadal censuses, so the fraction of residents using septic systems was computed at the county scale for 1990 and then interpolated backward in time, increasing to 100% of the population using septic systems in 1850 or the earliest decade for which a non-zero population was reported in each county. This approach relies upon the assumption that few if any modern sewage systems had been built prior to state incorporation (<https://www.mmsd.com/about-us/history>). The fraction of the population on septic systems was then multiplied by the county population to estimate the population using septic systems per county for each decade from 1850 to 1980. The population using septic systems was then multiplied by 4.1 kg of nitrate-nitrogen per person per year (11.2 g/person/day; Lusk et al., 2017) and normalized by county area to estimate the total leached nitrate in kg/ha from septic systems for each county. A similar approach for estimating septic nitrate leaching per area (kg/ha) was applied with 1990, 2000, and 2010 census data, but leveraged the finer resolution census block groups rather than counties. The results were mapped to shapefiles and then extracted to the 1 km by 1.5 km Agro-IBIS grid for each decade. The shapefiles, the python code used to generate the shapefiles, and a netCDF file that aggregates the 1 km by 1.5 km grids for each decade are available from (Schachter et al., 2024b). For implementation with the GW-NDST, the decadal leached septic nitrate data are extracted for the user-specified well location, and annual values are computed via linear interpolation between each decade. Septic leaching is also extrapolated to the current year (based on the date the user runs the tool) for the user-specified well location based on the average slope of annual leached nitrate between 1990 and 2000 and between 2000 and 2010.

2.2. Groundwater age model

The machine learning model of groundwater age (ageML) developed by Green et al. (2021) for the Great Lakes region using the R statistical

software program (R Core Team, 2019) was converted into an equivalent model (Kauffman, et al., 2024) in Python (Van Rossum and Drake, 2009) to allow for greater software compatibility and maintainability as part of the GW-NDST. The Python version of the ageML model was re-trained using the same training (80%) and testing (20%) dataset as Green et al. (2021). The ageML model uses predictor variables to estimate the mean groundwater age for a well, including landscape characteristics within a 500 m radius around the user-specified well, well construction information, well water level data, and water chemistry data for the well. The GW-NDST aggregates these predictor variables for the user-specified well and uses the ageML model to estimate a mean groundwater age. At a minimum, the well location (latitude and longitude) and well depth are required for running the tool; additional well-specific data, as described next, improves the ageML results but are not required for obtaining a solution. Well depth is the single most important predictor of groundwater age for the ageML model (Green et al., 2021), as has been documented for wells and aquifers based on first principles described in other studies (Vogel, 1967a; Maloszewski and Zuber, 1982; Haitjema, 1995; Luther and Haitjema, 1998; Green et al., 2018). The sampling depth, or the distance from the water table to the center of the saturated open interval of the well, was identified as the second most important predictor of groundwater age for the ageML model but requires knowledge of the depth to water and well casing or well screen length, either of which can be missing from some well records. If no depth to water information is provided by the user, a value is estimated using a separate machine learning model that was developed for this purpose (Smith et al., 2024). Finally, water chemistry information including nitrate, pH, chloride, and iron concentrations among others, are used by the ageML model (Green et al., 2021) to refine the estimate of mean age. Water level and chemistry information, which could be measured and reported repeatedly over time, are summarized in the GW-NDST as the mean of each year's average value prior to input to the ageML model. If water chemistry data is incomplete or not provided, the ageML model ignores that input and generates an estimated age without it, albeit with less predictive accuracy, as described by Green et al. (2021). In this way, the GW-NDST can be applied to nearly any well in the state but is expected to have improved accuracy when applied to wells with extensive well construction and water chemistry data.

The ageML model estimates a mean age for a given well location and depth; however, actual well water contains a distribution of ages from young to old water. Indeed, a distribution of groundwater ages is needed for the convolution solution implemented in the GW-NDST (see Convolution section). The mean age computed by the ageML model was converted to a travel time distribution $g(\tau)$ by applying a dispersion ratio to an advection-dispersion equation (Kreft and Zuber, 1978; Maloszewski and Zuber, 1982),

$$g(\tau) = \frac{d\sqrt{\bar{\tau}}}{\sqrt{4\pi d\bar{\tau}^3}} \exp\left(-\frac{d\bar{\tau}}{4\tau}\left(1 - \frac{\tau}{\bar{\tau}}\right)^2\right)$$

where d is the dimensionless dispersion ratio, τ is the travel time from infiltration to discharge from the well for a parcel of water sampled from the well, and $\bar{\tau}$ is the mean travel time of water discharged from the well. The mean travel time, $\bar{\tau}$, (incorporating both unsaturated and saturated flow) is generated from the ageML model, d is estimated from prior studies (commonly around 20; Green et al., 2018) and spatially calibrated across Wisconsin for the GW-NDST (see Calibration & uncertainty section), and τ is a travel time, which is represented in the GW-NDST with a series of timesteps of variable duration that are most refined around the mean travel time for each individual well.

2.3. Depth to water machine-learning model

As described in the Groundwater age model section, the depth to water in the well is a component of the “sample depth” predictor

variable of the groundwater ageML model (Green et al., 2021). For wells in which the user does not provide this information, the GW-NDST applies a depth to water machine-learning model to estimate the depth to water and “sample depth”, which is then supplied to the ageML model. The depth to water model (Smith et al., 2024) leveraged 48 GIS-based predictor variables and was trained on 61,692 water level measurements from wells and 53,307 stream elevations assumed to represent the water table during baseflow conditions. For implementation with the GW-NDST, the machine learning model parameter values that provided the lowest root mean squared error (RMSE), in terms of matching the depth to water model to the training dataset, were combined with GIS input data from within 500 m of the user-supplied well location. Again, model-estimated depth to water is only used in the GW-NDST workflow when measured depth to water is not provided by the user (Fig. 1).

2.4. Oxygen and nitrate reduction rate model

Nitrate dissolved in groundwater can be converted to inert nitrogen gas (N_2) through the denitrification process. However, denitrification within aquifers occurs only after most of the oxygen dissolved in the water has been depleted, with an effective oxygen cut-off threshold value (see equation 2 below) of approximately 2 mg/L (Green et al., 2008; Tesoriero and Puckett, 2011). Oxygen inhibition of denitrification creates non-linear reaction kinetics that are not captured by single parameter (zero- or first-order decay coefficient) kinetic models. The GW-NDST realistically represents the reaction kinetics by accounting for oxygen reduction and denitrification separately. Oxygen reduction rates were estimated for the GW-NDST using a multi-variate regression model (Juckem and Green, 2024) trained on dissolved oxygen concentrations from wells that had been sampled as part of numerous USGS and non-USGS studies (Baker et al., 2024). The wells used for training and validating the multi-variate regression model had at least one dissolved oxygen concentration value and a computed mean groundwater age informed by age tracer samples (Juckem and Green, 2024). This data included publicly available USGS data (U.S. Geological Survey, 2016) and sample data from 10 historical reports published by the Wisconsin Geological and Natural History Survey and the University of Wisconsin at Stevens Point (Baker et al., 2024). The multi-variate regression model was trained to match dissolved oxygen concentrations for 461 wells in the dataset using predictor variables of land cover, soil, and aquifer lithology and hydrogeologic position, plus other mappable variables. Mappable predictor variables were computed as the average values within a 500 m radius surrounding each well. Well specific metrics, including the thickness of the unsaturated zone and the sampling depth were also used as predictor variables.

Denitrification was implemented in the GW-NDST by computing an oxygen reduction rate from the multivariate regression model using the parameter values that best matched the training dataset, along with estimates of the dissolved oxygen concentration at which denitrification begins ($C_{O,cut}$) and estimated denitrification rates that were scaled to the computed oxygen reduction rate via a calibrated multiplier ($k_{O2_to_kNO3_mult}$). That is, denitrification is simulated in the GW-NDST only after estimated dissolved oxygen concentrations drop below a cut-off threshold value ($C_{O,cut}$). Thus, the groundwater age at which denitrification begins (τ_N^*) for an individual parcel of water is expressed as

$$2\tau_N^* = \tau - \tau_u - (C_{0,O} - C_{O,cut}) / k_{0,O}$$

where τ_u is the unsaturated zone travel time, $C_{0,O}$ is the input concentration of dissolved oxygen at the water table (assumed 9.8 mg/L everywhere for the GW-NDST), and $k_{0,O}$ is the zero-order oxygen reduction rate computed for the individual well by the multi-variate regression model. The denitrification rate ($k_{0,N}$) is computed as the product of the oxygen reduction rate ($k_{0,O}$) and a multiplier value ($k_{O2_to_kNO3_mult}$). Data were insufficient to independently estimate $C_{O,cut}$ and the denitrification rate multiplier ($k_{O2_to_kNO3_mult}$) values,

which were instead parameterized and estimated during a state-wide calibration of the GW-NDST to nitrate concentrations by assigning initial estimates with upper and lower bounds informed from Green et al. (2018, Fig. 7) and other studies (see Calibration & uncertainty section).

2.5. Convolution

Nitrate transport from below the root zone to wells is simulated in the GW-NDST via mathematical convolution of historical nitrate leaching over time and groundwater age distributions, with zero-order oxygen and denitrification rates. The method is described in detail in section 2.3 of (Green et al., 2018), but adapted and expanded here for use with any well location in Wisconsin and alternate sources of nitrate leaching data, in addition to a separate calibration scheme (Calibration & uncertainty). Briefly, nitrate concentration in a well is simulated over a range of travel times using the convolution equation,

$$3 C_s(t) = \int_{-\infty}^t [C_0(t') - (t - t')k_{0,N}]g(t - t')dt'$$

where C_s is the concentration [M/L³] of nitrate on the sample date (t); C_0 is the nitrate concentration on the date of infiltration at time t' ; $k_{0,N}$ is the zero-order denitrification rate [M/L³/T] between the date t^* , at which the reaction begins and the date of sampling (t); and g is the travel time distribution between the date of infiltration and the time of sampling ($\tau = t - t'$) as described in equation 1.

2.6. Calibration & uncertainty

The GW-NDST was calibrated to nitrate samples across Wisconsin by using the iterative ensemble smoother technique (IES; White, 2018) implemented in PESTPP (PESTPP-IES; White et al., 2020). Model calibration (also known as history matching or parameter estimation) using PESTPP-IES involves assigning weights to nitrate sample targets that reflect their accuracy and importance, and defining model parameters that can be adjusted to improve the match between simulated and target values. PESTPP-IES incrementally improves the model calibration by reducing the sum of squared weighted residuals, or phi, over a series of iterations that are informed by the response of simulated target values to perturbations of the parameter values in the ensemble of realizations. An important benefit of using the PESTPP-IES method for calibration is that the method can be used to generate hundreds of realizations of model parameters for which all parameter values remain within realistically defined bounds and the simulated match to target values remain “in calibration” (each retained realization has a phi value within 2 standard deviations from the mean phi of all realizations). When evaluated as an ensemble, the range of model parameter realizations provides insights about the range of probable forecast outcomes. The sections below describe details related to the targets, their weights, the model parameters, and implementation of the PESTPP-IES method for calibrating and quantifying uncertainty for the GW-NDST.

2.7. Targets

Calibration of the GW-NDST was performed by comparing 27,195 measured nitrate concentration targets from 16,979 wells with equivalent simulated values for each target. In addition, nitrate trends in wells with multiple samples were compared with equivalent simulated trends by computing differences among both 1. consecutive sample concentrations (9783 targets), and 2. concentrations for the first and last samples from a well (1600 targets). Thus, a total of 38,578 targets were used (27,195 measured concentrations and 11,383 concentration differences) to initialize the calibration process. Nitrate concentration and well construction data were obtained from the USGS NWIS database (U.

S. Geological Survey, 2016) and the WI-DNR’s Groundwater Retrieval Network (Wisconsin Department of Natural Resources, 2023).

An initial total of 53,708 candidate targets from the USGS and WI-DNR databases were reduced to the 27,195 concentration targets listed above via three filtering steps. First, for wells with more than 4 samples and having at least one pair of consecutive samples that differed by more than 10 mg-N/L (milligrams per liter as N), all samples with concentrations that exceeded 1.5× the interquartile range of samples from the well were removed. This step minimized the likelihood of errant or spurious samples. Second, all wells with measured concentrations greater than 40 mg-N/L were manually vetted and samples were removed if they were found to exhibit an unreasonably large change in concentration over a short time. While some wells with very high concentrations may be influenced by local processes that are not represented in the GW-NDST, such wells were not automatically removed as targets because the actual source was not documented. For example, wells potentially screened across a neighboring septic plume, which the GW-NDST is not designed to represent (septic is estimated at county and census block-group scales), are not identified in the source datasets and would require supplemental water chemistry sampling to identify. Third, all samples less than a background concentration of 1.0 mg-N/L (Nolan and Hitt, 2003) were removed as targets because the tool is intended to focus on wells with nitrate contamination, and early calibration efforts illustrated that the large number of low concentration samples biased calibration results.

After the 38,578 concentration and difference targets were identified and weighted (next paragraph), a fourth filtering step occurred during the calibration process itself. As part of the PESTPP-IES methodology (White et al., 2020), simulated outputs from prior parameter ensembles (evaluated after each iteration) were used to identify target values that were in prior-data conflict (Evans and Moshonov, 2006; Nott et al., 2020). That is, targets with values beyond 4 standard deviations of the mean of the distribution of simulated values for that target were considered to be in conflict and thus removed from the parameter adjustment process (1170 concentration and 3153 concentration difference targets). This resulted in 26,025 concentration targets and 8230 concentration difference targets, for a total of 34,255 targets used for assessment of the final calibration metrics. Removal of outlier targets (those in prior-data conflict) helps the calibration algorithm avoid unrealistic parameter values that can result in biased model forecasts (Nott et al., 2020; White 2018).

Weights for the nitrate concentration targets were initially estimated by assessing the variability of sampled concentrations among wells containing multiple samples within any 31-day period (429 samples). The approach was designed to provide an indication of the level of measurement uncertainty attributable to natural variability within aquifers and the sampling process. Thus, the initial weight assigned to each of the 27,195 concentration targets was computed as the inverse of the standard deviation (2.6) of concentrations within any 31-day period from this set of 429 samples, or a weight of 0.38. Similarly, initial weights assigned to the first-to-last and consecutive difference targets were computed as the inverse of the standard deviation (3.73 and 2.08, respectively) of the target values, or weights of 0.27 and 0.48, respectively. After assigning measurement-uncertainty-based weights, targets were grouped by their measured concentration (1.0–4.9 mg-N/L, 5.0 to 9.9 mg-N/L, and greater or equal to 10.0 mg-N/L) and the reported locational accuracy of the well (GPS or aerial photograph versus quarter-section). The first-to-last concentration difference and consecutive difference targets were also assigned as individual target groups. The contribution of each group to the sum of the squared-weighted-residuals (phi) was computed using initial parameter values and measurement-uncertainty based weights. Then target weights were re-balanced (Doherty and Hunt, 2010; Fienen et al., 2022) to amplify target groups deemed to be more accurate or more important for the calibration objectives, as illustrated in Table 1. For example, the target weight re-balancing process assigned greater weight to wells with higher

Table 1
Description of target groups, number of targets per group, initial measurement error-based weights, and contribution of target groups to phi before and after re-balancing.

Target group	Description	Number of targets	Initial weight	Initial contribution to phi (%)	Rebalanced contribution to phi (%)
PLSS_1to5	Nitrate sample concentrations from WI-DNR databases for wells located to within a quarter-section based on the Public Land Survey System. "1to5", "5to10", and "10plus" refer to the target concentration of samples within each group.	9244	0.38	26.4%	10%
PLSS_5to10		4777	0.38	16.2%	10%
PLSS_10plus		2799	0.38	17.1%	10%
GPS_1to5	Nitrate sample concentrations from USGS and WI-DNR databases for wells located via GPS or digitized on-screen from plat books or aerial photographs. "1to5", "5to10", and "10plus" refer to the target concentration of samples within each group.	6442	0.38	20.7%	20%
GPS_5to10		2770	0.38	10.9%	18%
GPS_10plus		1163	0.38	5.6%	17%
Diff_FL	Differences in nitrate concentration between the first (earliest) and latest sample from the same well.	1600	0.27	0.7%	5%
Diff_CS	Differences in nitrate concentration between consecutive samples from the same well.	9783	0.48	2.4%	10%

locational accuracy and samples with higher concentrations because the calibration dataset had relatively fewer targets with high concentrations. Thus, final weights for each target were assigned based on the original measurement uncertainty assessment and the manual re-balancing step illustrated in Table 1.

2.8. Parameters

Eleven parameter groups were used for calibrating the GW-NDST and consisted primarily of multiplier values applied to outputs from the support models, such as multipliers applied to the nitrate leaching flux from the Agro-IBIS model, a multiplier on the mean groundwater age produced from the ageML model, and others. Input values that were not available from support models or other sources were parameterized and estimated directly as native parameter values rather than as multipliers. Each parameter was allowed to vary spatially across the state using pilot points (Doherty, 2003; Doherty et al., 2010). These pilot points were conceptualized as geostatistically correlated fields, which were described by an exponential variogram and a 200,000 m range; pilot points were spaced 20,000 m apart. This approach resulted in 408 pilot points for each parameter group, for a total of 4488 unique pilot-point parameters. A maximum of 50 singular values were used to regularize, or limit, parameter variability. Readers are referred to White et al. (2020) for detailed descriptions on the application of pilot points, singular values, regularization, and other advanced calibration topics.

Setting realistic lower and upper parameter bounds was an important step in setting up the calibration because the bounds, in conjunction with *par_sigma_range* (discussed next), are a primary control over the ensemble of parameter values generated for the PESTPP-IES calibration process. Bounds for parameters that were not informed by output from support models (*disp_ratio*, *uz_mobile*, *O2_cut*, and *kO2_to_kNO3_mult*) were informed by the fifth and 95th percentile of values reported by Green et al. (2018). A similar approach was used to assess the starting multiplier value and bounds for recharge simulated from the Agro-IBIS model (*flux_mult*) by comparing the Agro-IBIS results with a baseflow analysis by Gebert et al. (2011). Bounds for support model multiplier parameters were informed from cross-validation of the individual support model results, as well as from prior experience and expert judgement. For example, the *age_mult* parameter's bounds were set based on cross-validation of the ageML model results (Green et al., 2021) in that, the lower and upper bounds of $0.5\times$ and $2.0\times$ (Table 2) are similar to the mean ratio of errors from the ageML model holdout (testing) dataset. This approach ensured that the groundwater ages used within the GW-NDST (after calibration) didn't deviate too much from the ageML model results, which were trained on groundwater ages informed by age tracer samples, yet also allowed flexibility for the ages used by the GW-NDST to improve the tool's match to measured nitrate concentrations while staying within the confidence interval of the ageML model.

The prior parameter ensemble was constructed for PESTPP-IES using a geostatistical draw and a prior covariance matrix that was informed by the variogram, parameter bounds, and the PESTPP-IES setting *par_sigma_range*. A value of 4 was used for *par_sigma_range*, which assumed that the parameter bounds approximated a 95-percent confidence interval (i.e., four standard deviations between the lower and upper bounds) on the prior parameter probability distribution. This setting provided a broad sampling of parameter values across the lower and upper bounds during the geostatistical draws that ultimately produced the ensemble of realizations used to characterize model uncertainty in the GW-NDST. Table 2 summarizes the parameters used to calibrate the GW-NDST, along with lower and upper bounds that were used to generate the parameter ensemble for PESTPP-IES.

The calibration process used by PESTPP-IES involves modifying all parameter values for all realizations comprising the ensemble (in incremental steps called iterations) to generate associated response changes for all simulated results (nitrate concentrations) to generate an approximation to the Jacobian matrix. This approximate Jacobian

Table 2

Calibration parameter group names, starting values, lower and upper bounds, description, and the source for justification of the starting parameter values and bounds.

Parameter group name	Starting Value	Lower bound (multiplier on starting value)	Upper bound (multiplier on starting value)	Actual lower bound	Actual upper bound	Description	Source and justification
age_mult	1	0.5	2	0.5	2	Multiplier on the mean age (τ) generated by the groundwater age machine learning model (ageML).	Lower and upper bounds approximately match computed average errors for holdout and cross validation with the age model by Green et al. (2021). From Green et al. (2018) ; starting and actual bounds equate to the median, 5th, and 95th percentiles.
disp_ratio	21.32	0.22	1.90	4.73	40.44	Value for the dispersion ratio (d) in equation 1.	
uz_mobile	0.16	0.63	1.94	0.1	0.31	Value for the unsaturated mobile water content used to estimate lag time in the unsaturated zone.	From Green et al. (2018) ; starting and actual bounds equate to the median, 5th, and 95th percentiles.
IBIS_mult	1	0.2	3	0.2	3	Multiplier on the nitrate leaching concentration generated by the Agro-IBIS model. Applies equally to all dates.	The bounds attempt to account for the coarse representation of variable land cover and associated nutrient application variability within the approx. 1 km by 1.5 km cells, while recognizing observed ranges in measured nitrate leaching rates below common midwestern agricultural fields (Shrestha et al., 2023).
IBIS_start_mult	1	0.2	3	0.2	3	Multiplier on the nitrate leaching concentration generated by the Agro-IBIS model. This multiplier is interpolated from the calibrated value in 1850 to 1.0 in 2016.	This parameter facilitates calibration of nitrate leaching trends. The bounds consider the coarse representation of the Agro-IBIS model cells and observed ranges in measured leaching rates (Shrestha et al., 2023).
IBIS_end_mult	1	0.2	3	0.2	3	Multiplier on the nitrate leaching concentration generated by the Agro-IBIS model. This multiplier is interpolated from 1.0 in 1850 to the calibrated value in 2016.	This parameter facilitates calibration of nitrate leaching trends. The bounds consider the coarse representation of the Agro-IBIS model cells and observed ranges in measured leaching rates (Shrestha et al., 2023).
flux_mult	0.55	0.25	2.50	0.14	1.4	Multiplier on the recharge flux simulated by the Agro-IBIS model. Applies equally to all dates.	Starting and actual bounds equate to the mean ratio of baseflow estimated recharge (Gebert et al., 2011) to IBIS simulated recharge for gages and partial record basin, and the 5th and 95th percentile of the ratio for partial record basins.
septic_mult	1	0.1	10	0.1	10	Multiplier on the nitrate leaching mass generated by the septic system algorithm. Applies equally to all dates.	The two orders of magnitude range across the bounds attempt to account for the coarse estimation of septic leaching at county to census block groups, as per Schachter et al. (2024b) .
kO2_mult	1	0.5	20	0.5	20	Multiplier on the oxygen reduction rate (k_{O_2}) in equation 2.	Lower and upper bounds are approximated as the ratio of the 5th and 95th percentiles, respectively, from Fig. 7 of Green et al. (2018) , divided by the median oxygen reduction rate computed with the multivariate regression model used for the GW-NDST (Juckem and Green, 2024).
O2_cut	0.87	0.45	8.25	0.39	7.18	Value for the oxygen cutoff threshold ($C_{O_{cut}}$) in equation 2.	From Green et al. (2018) table S3 after removing insensitive parameter values that were unchanged from the initial value of 2.1; starting and actual bounds equate to the median, 5th, and 95th percentiles. The bounds are within those observed by Tesoriero and Puckett (2011) .
kO2_to_kNO3_mult	2	0.25	10	0.5	20	Multiplier on the oxygen reduction rate (k_{O_2}) to generate a denitrification rate (k_{O_2N}) in equation 3.	Starting and actual bounds are approximated from Green et al. (2018) , Fig. 7 .

matrix is similar to the Jacobian, or “sensitivity” matrix generated by more traditional gradient-based methods in PESTPP that increment each parameter in isolation. Successive IES iterations are performed until the difference between simulated results and target values (ϕ) are judged to be sufficiently small to be considered “calibrated”. The result of this process is an estimate of the posterior parameter probability distribution, from which a posterior parameter ensemble can be generated and used to quantify model uncertainty. The approach requires a means for determining which parameter realizations to consider as being “in calibration” versus those with excessively large composite residuals, or ϕ . For calibrating the GW-NDST, realizations were excluded if their ϕ value was more than two standard deviations from the mean ϕ value of that iteration’s suite of realizations.

Unlike the traditional calibration methods in PEST (Doherty, 2018) and PESTPP (White et al., 2020), which focus on a single “best” parameter set, the idea behind the IES method (White, 2018) is to improve the overall fit of an ensemble of model parameters by reducing parameter uncertainty using the information contained in the observation data. That is, no single model result exactly replicates the real world, but an ensemble of results can be assessed for their potential to produce a range of results that likely encapsulate the “true” (though unknown) result. The calibration process for PESTPP-IES improves the fit between each ensemble member (a unique set of parameter values) and the target values during each iteration. This correspondence between simulated and target values improves during each iteration, sometimes dramatically; however, improved fit often comes at the expense of reduced variability of model predictions (ensemble collapse). The favored approach, therefore, is viewed as a balance between improving the match to measurements while retaining a reasonably broad range of results from the ensemble of realizations. This approach is similarly designed to reduce the chance of over-fitting the calibration dataset, which can often result in degraded forecasts (Kuhn and Johnson, 2013; Anderson et al., 2015).

2.9. Scenario implementation

The GW-NDST design focuses on simulating current and historical nitrate concentrations and providing users with flexible forecasting scenarios. The “historical_simulation” function computes historical nitrate concentrations up to a specified prior date or the current date of use, and generates three plots: historical nitrate leaching rates, groundwater age distributions, and historical nitrate concentrations in the well of interest. The visualization of historical leaching rates and groundwater ages can aid with understanding patterns and trends in the nitrate concentration history for each well. Another use of the historical plots is to assist the user with anticipating reasonable goal concentrations for the future scenarios, which are accessible via the

“future_scenarios” function. The GW-NDST incorporates six future scenario options: 1. A constant future leaching rate based on the latest Agro-IBIS and septic leaching rates, 2. A user-specified constant future leaching rate, 3. A user-specified percent change (from rates computed for a specified date or the date the tool is run) for future leaching rates, 4. An optimization method whereby the user specifies a goal concentration and date, and the GW-NDST computes the required immediate and constant nitrate leaching reduction required to meet this goal (if possible), 5. A similar optimization method based on a goal concentration and date, whereby the GW-NDST computes an annual nitrate leaching reduction rate (a per-year reduction rather than immediate reduction) that would be required to meet the goal (if possible), and 6. An optimization method whereby the user identifies a goal concentration and nitrate leaching rate, and the GW-NDST computes the date at which the goal concentration is expected to be met (if possible). Mean annual measured nitrate concentrations based on sample results provided in the user’s input file are also plotted in all cases for comparison with simulated results.

Uncertainty in the simulated results is illustrated by leveraging the 450-realization posterior parameter ensemble generated during the calibration process. Each realization includes a suite of reasonable parameter values that resulted in a similar match to nitrate concentration and trend targets as the “calibrated” or “base” realization. Thus, both the historical_simulation and future_scenarios functions display results from the base realization, as well as results from near-minimum, median, and near-maximum realizations. The specific realizations plotted for each user-identified well will differ and are selected based on their computed nitrate concentration for the date that the tool is used unless a past date is specified by the user. That is, computed concentrations for the date of use (or specified date) are ranked from lowest to highest from all 450 realizations, and the median realization is identified as the realization parameter set that produced the median concentration on that date. Similarly, the near-minimum realization parameter set is associated with the concentration that exceeds five percent of all computed concentrations on the date, and the near-maximum realization is associated with the concentration that exceeds 95 percent of computed concentrations. The median realization is used as a complement to the base realization with the goal of these two model results illustrating the central tendency of simulated historical, current, and future nitrate concentrations. The near-minimum and near-maximum realizations bracket a reasonable range in the simulated historical, current, and forecasted future nitrate concentrations.

3. Results

This section discusses results of the calibration process and uses an example run of the Graphical User Interface (GUI) to illustrate how

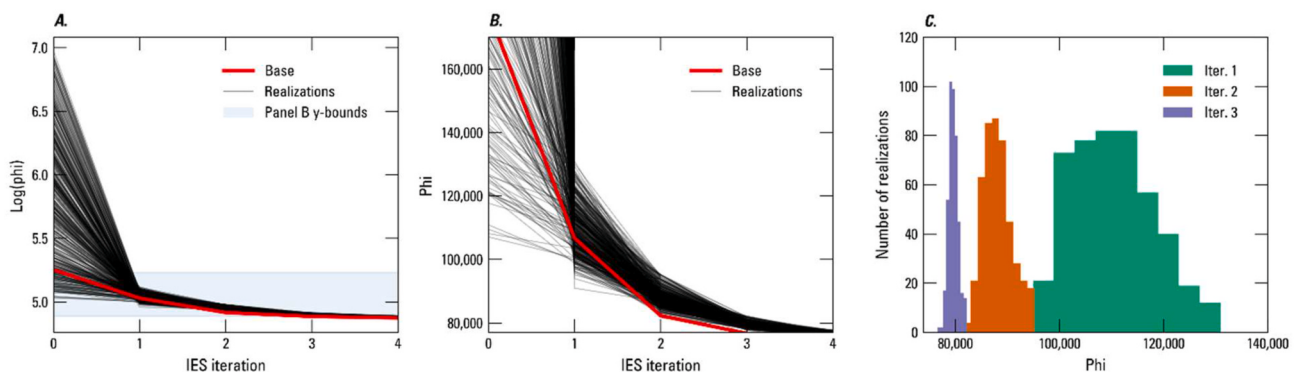


Fig. 2. Calibration metric (ϕ) characteristics of the PESTPP-IES calibration of the GW-NDST tool, showing (a) the reduction in $\log(\phi)$ for the “base” simulation (red line) and ensemble of the other 449 realizations (black lines) over 5 iterations including the initial parameter values (iteration 0), (b) a zoomed-in depiction of the reduction in ϕ for all 450 realizations across all iterations, and (c) histograms of ϕ for all realizations for iterations 1, 2, and 3 (iterations 0 and 4 omitted for improved X-axis scaling).

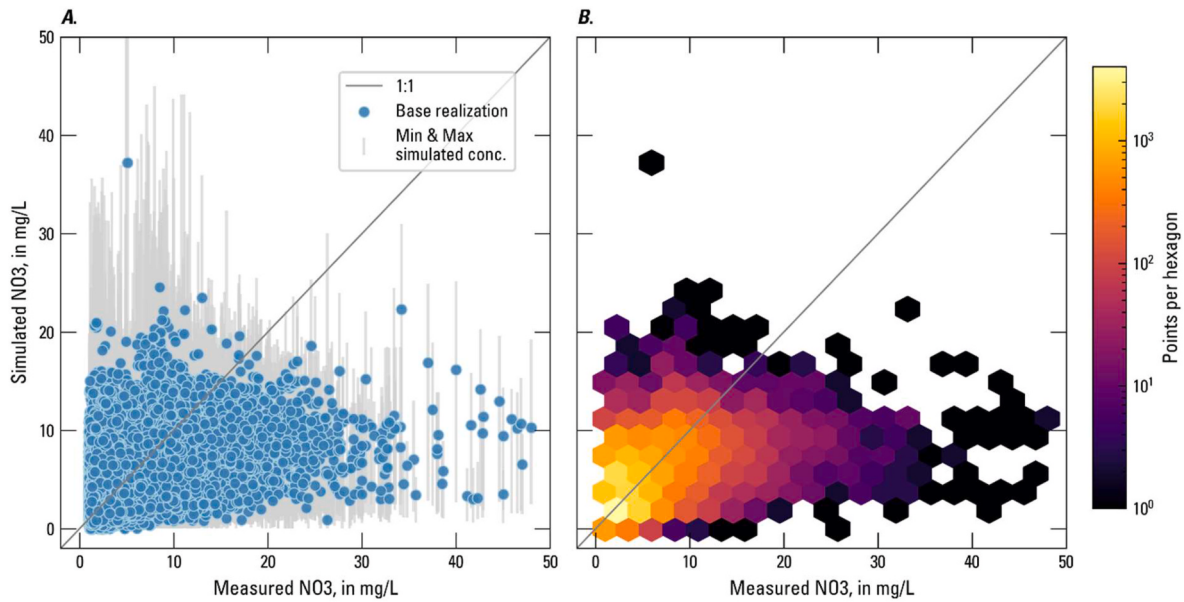


Fig. 3. One-to-one plots of measured versus simulated nitrate concentration for the base realization for PESTPP-IES iteration number 2, (a) with the minimum and maximum simulated range from the other 449 realizations shown for each target as a vertical gray line, and (b) shown as a hex-bin plot in which warmer colors illustrate a higher number of paired measured:simulated values.

results of the calibration and uncertainty analysis inform the forecasting scenario results. The calibration process and results are highlighted because they represent the means by which outputs from support models were adjusted in order to tune their collective results for working within the context of this tool, highlighting the model capabilities and limitations, and incorporating simulated uncertainty and forecasting outcomes in results.

3.1. Calibration results and model uncertainty

Iteration number 2 of the PESTPP-IES (White, 2018) calibration was selected and judged to be the best ensemble of parameter realizations based on the calibration metrics (ϕ , RMSE, etc.), the number of concentration targets contained within the range of ensemble results, the range of parameter values contained in the ensembles, and the limited number of parameters hitting the upper or lower bounds. Fig. 2

illustrates this balance of minimizing ϕ without collapsing the parameter set (narrow range of ϕ) across five PESTPP-IES iterations (only iterations 1–3 shown in 2c for greater clarity). Also, the results are shown for the final ensemble of 450 realizations (including the base realization) out of an initial 500 realizations because 50 realizations generated excessively high ϕ values (greater than 2 standard deviations from the mean ϕ of the 500 realizations) and were automatically removed by PESTPP-IES during the calibration process.

Results from iteration number 2 were selected partially because they adequately matched measured target concentrations and concentration trends, while minimizing biases and capturing over 78 percent of target concentrations within the ensemble of realizations. This is shown in Fig. 3 for iteration 2, and illustrates the GW-NDST's ability to match measured nitrate concentrations across the state that ranged from 1 mg-N/L to nearly 50 mg-N/L. For the base realization, the mean absolute error (MAE) for all concentration targets was 3.1 mg-N/L and the RMSE

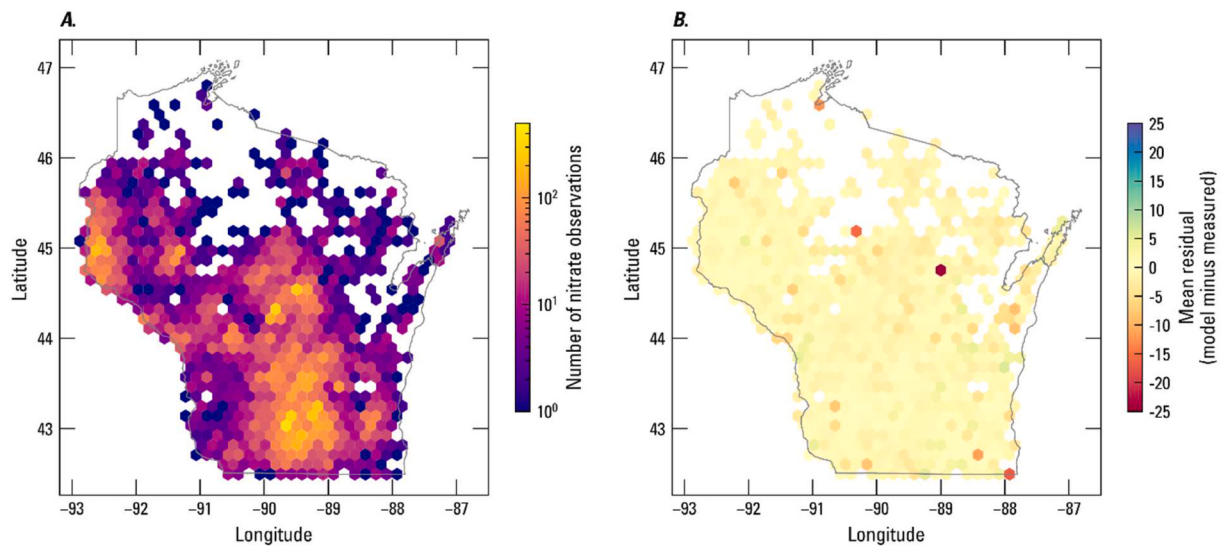


Fig. 4. Maps showing (a) the number of nitrate observations per hexbin area used to calibrate the GW-NDST across Wisconsin, and (b) the mean residual for each hexbin. Each hexbin covers approximately 160 square kilometers. White areas without hexbins had no nitrate observations.

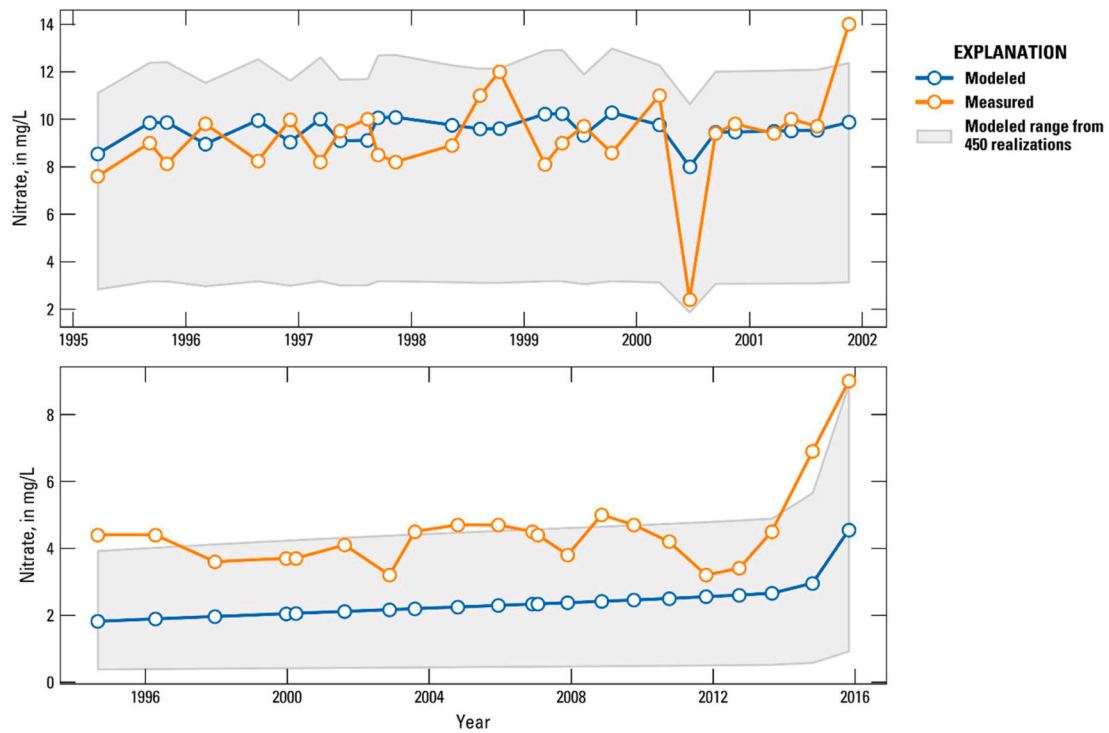


Fig. 5. Time-series of modeled (base realization) versus measured concentrations for two example wells.

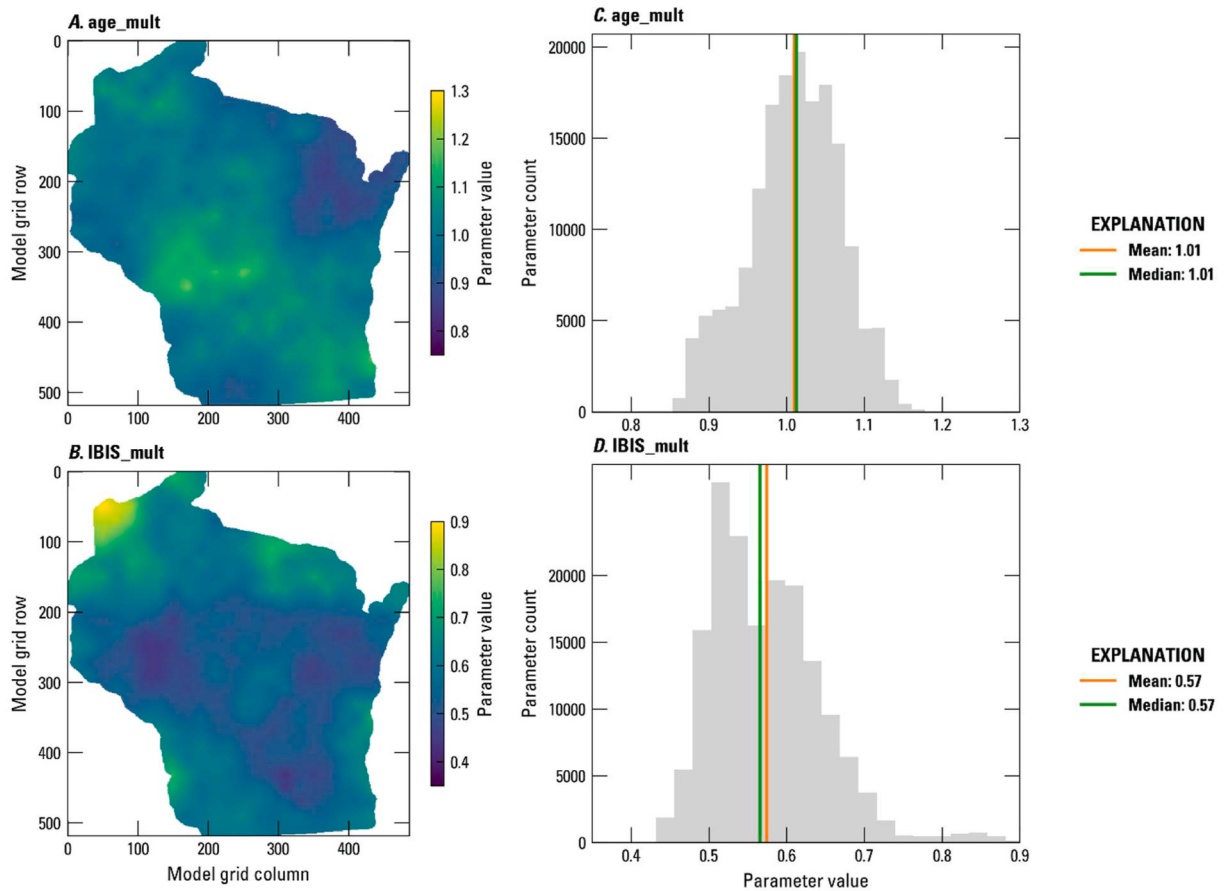


Fig. 6. Range of parameter values shown (a, b) spatially and (c, d) as distribution histograms for the age_mult (a, c) and IBIS_mult (b, d) parameter groups, respectively, for the base realization.

was 4.5 mg-N/L, with 61% of samples simulated within 3 mg-N/L and 81% simulated within 5 mg-N/L. The calibration exhibits minimal clustering of residuals up to about 20 mg-N/L, but systematically under-simulates concentrations above about 20 mg-N/L. The tool's performance above 20 mg-N/L likely results from a lack of site-specific information, such as septic system locations and farm-specific management data (manure and fertilizer application, crops, cover crop use, irrigation, etc.) related to potential intensive leaching areas and events that aren't captured by the grid resolution of the Agro-IBIS model and its underlying input datasets; preferential flow due to natural features or well construction, not captured by the ageML model (eq. 1), may also contribute. Nonetheless, improved correspondence among relatively high concentrations is achieved when considering the full range of results (gray vertical lines in Fig. 3a) from the 450 realizations. Indeed, the ensemble of results from the 450 realizations overlapped 78% of the target concentrations, as per the calibration objectives described above, highlighting the importance of incorporating model uncertainty in simulated outcomes and decision making.

Spatially, simulated residuals across the state (Fig. 4) exhibit minimal bias. The highest density of nitrate concentration targets (Fig. 4a) occurs in a north-south line through the center of the state from the Illinois border in the south to about 45° latitude, north of which much of the land is covered by forest and lakes. Smaller high-density areas of nitrate sample targets also occur throughout the state, such as in the north-west and south-east. Mean residuals (Fig. 4b) in these high-sample-density areas are consistently near zero mg-N/L, demonstrating the effectiveness of the spatially varying parameters (pilot points; see the Parameters section). Large mean residuals (greater than 10 mg-N/L) tend to occur where target density is low (less than about 10 wells within each hexbin in Fig. 4).

Trends in nitrate concentration were represented in the calibration by the first-last and the consecutive difference target groups, in which concentration differences were computed from wells that included multiple nitrate samples over multiple years. Two nutrient leaching calibration parameters in particular, IBIS_start_mult and IBIS_end_mult, provided added flexibility to the calibration process for influencing long-term trends (from 1850 to 2020). However, these parameters modify the leaching rate trends over the full 170-year duration of the Agro-IBIS results, and therefore have minimal influence on inter-annual variability. That is, much of the simulated temporal variability and trends observed in the GW-NDST, especially inter-annual variability, is controlled by the variability and trends in annual leaching rates directly simulated by the Agro-IBIS model. Two example wells illustrate characteristic timeseries data and simulation results (Fig. 5). The simulated pattern for these wells generally matched the time-series data, although the GW-NDST under-simulated the magnitude of change in temporal

fluctuations. This limitation is not surprising since farm-specific land management practices are not incorporated into the Agro-IBIS model, nor are transient stresses such as rapid infiltration or dynamic pumping incorporated into the groundwater transport processes. Nonetheless, simulated results for both wells generally mimic both short-term variability as well as longer-term trends exhibited by the data. Interestingly, the state-wide mean calibrated values for IBIS_start_mult and IBIS_end_mult were 1.14× and 0.90×, respectively, indicating that the actual nitrate leaching over time may have decreased relative to that simulated directly by the Agro-IBIS model, although the suite of 450 realizations included numerous individual realizations in which the parameters yielded increasing leaching rates over time compared with the raw Agro-IBIS output.

Parameter values varied spatially across the state (Fig. 6) as well as across the ensemble of realizations (Fig. 7; Table 3). A global sensitivity analysis (Table 3) using the Method of Morris (White et al., 2020) indicated that the IBIS_mult and age_mult parameter groups were the most sensitive for the calibration (51.0% and 38.5% of total sensitivity, respectively). For the base realization, each parameter field (interpolated between pilot points) varied smoothly across the state with local variations that improved the simulated match to targets in that area of the state. The maps in Fig. 6 further illustrate the geostatistical draw method's advantage of generating smoothly varying values from regional averages to local patterns without generating "bullseye" patterns that are common with least-squares interpolation methods. The associated histograms for the base realization (Fig. 6c and d) illustrate the distribution of values across the entire state for these sensitive parameters. Parameters also varied spatially for all the other 449 realizations through the same pilot point design, although the spatial patterns differed for every realization. For example, all values for the age_mult parameter across the state for all realizations (not just the base case) are shown in Fig. 7. This cumulative histogram (Fig. 7) illustrates a relatively dense distribution of values between the 5th and 95th percentiles (red vertical lines) of the parameter range, with the highest density of values centered around the mean and median values. Notably, the value of the age_mult parameter rarely reached the lower or upper bound anywhere in the state throughout all realizations, with the 5th and 95th percentile of values ranging from 0.71 to 1.45 times the mean age predicted by the ageML model (Table 3). Note that the histogram for age_mult in the base realization (Fig. 6c) is only one of the 450 realization histograms that make up Fig. 7.

3.2. Example scenario results

Example scenarios from the GW-NDST serve to clarify the relationships between nitrate leaching rates, groundwater ages, and associated

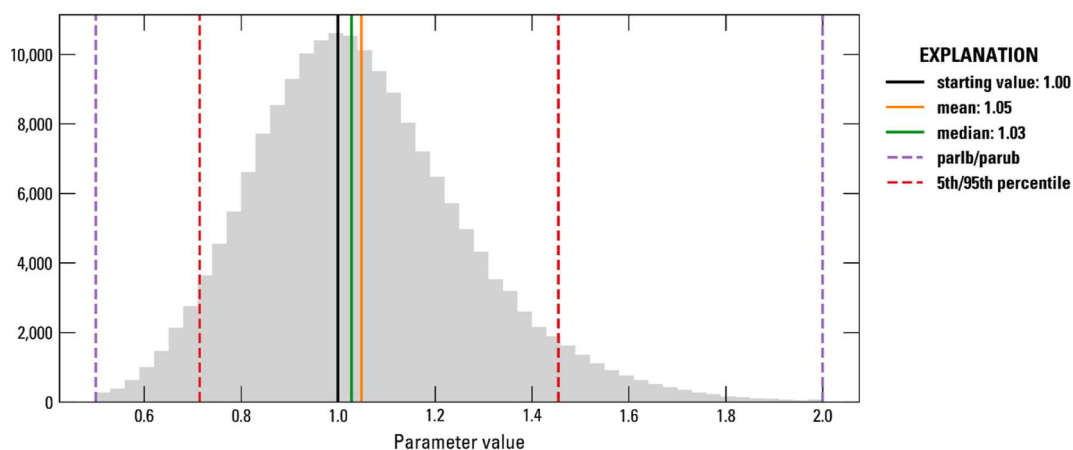


Fig. 7. Range of parameter values in the age_mult parameter group shown as a cumulative histogram for all 450 realizations across Wisconsin (parlb and parub are the lower bounds and upper bounds, respectively, for the parameter group, as per Table 2).

Table 3

Calibrated parameter group results, including mean and median values across the state, 5th to 95th percentile ranges, and percent of sensitivity computed from 450 realizations.

Parameter group name	Starting Value	Actual lower bound	Actual upper bound	Description	Mean calibrated value across WI	Median calibrated value across WI	5th percentile value	95th percentile value	Percent of total sensitivity
age_mult	1	0.5	2	Multiplier on the mean age (τ) generated by the groundwater age machine learning model (ageML).	1.05	1.03	0.71	1.45	38.5
disp_ratio	21.32	4.73	40.44	Value for the dispersion ratio (d) in equation 1.	16.64	15.52	7.25	29.6	3.4
uz_mobile	0.16	0.1	0.31	Value for the unsaturated mobile water content used to estimate lag time in the unsaturated zone.	0.17	0.17	0.11	0.25	0.4
IBIS_mult	1	0.2	3	Multiplier on the nitrate leaching concentration generated by the Agro-IBIS model. Applies equally to all dates.	0.64	0.57	0.26	1.27	51.0
IBIS_start_mult	1	0.2	3	Multiplier on the nitrate leaching concentration generated by the Agro-IBIS model. This multiplier is interpolated from the calibrated value in 1850 to 1.0 in 2016.	1.27	1.14	0.46	2.59	1.2
IBIS_end_mult	1	0.2	3	Multiplier on the nitrate leaching concentration generated by the Agro-IBIS model. This multiplier is interpolated from 1.0 in 1850 to the calibrated value in 2016.	0.91	0.8	0.34	1.86	2.5
flux_mult	0.55	0.14	1.4	Multiplier on the recharge flux simulated by the Agro-IBIS model. Applies equally to all dates.	0.64	0.59	0.28	1.17	1.3
septic_mult	1	0.1	10	Multiplier on the nitrate leaching mass generated by the septic system algorithm. Applies equally to all dates.	4.71	3.89	0.8	10	0.1
ko2_mult	1	0.5	20	Multiplier on the oxygen reduction rate ($k_{o,o}$) in equation 2.	1.07	0.83	0.5	2.46	1.4
O2_cut	0.87	0.39	7.18	Value for the oxygen cutoff threshold ($C_{o,cut}$) in equation 2.	1.1	0.86	0.4	2.61	0.1
ko2_to_kNO3_mult	2	0.5	20	Multiplier on the oxygen reduction rate ($k_{o,o}$) to generate a denitrification rate ($k_{o,n}$) in equation 3.	3.64	2.58	0.74	10.44	0.1

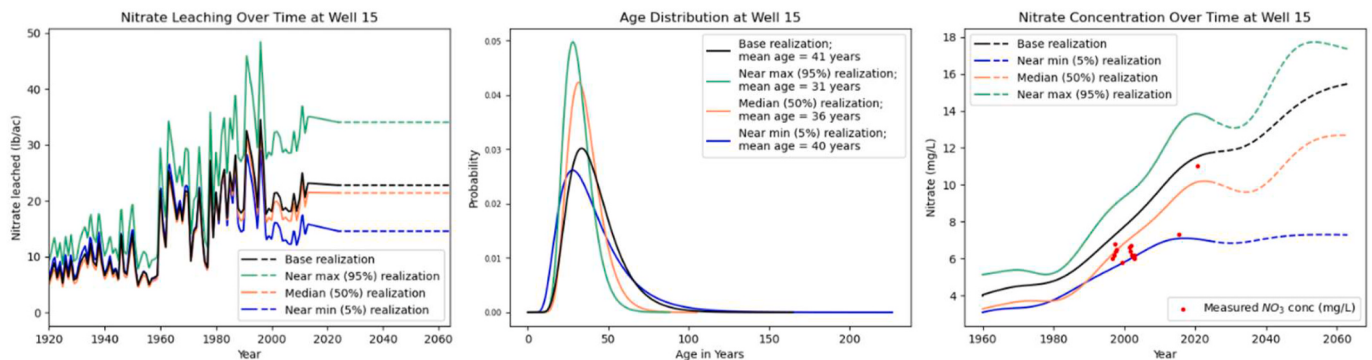


Fig. 8. Graphical results for an example well using the forecasting scenario option 1 (constant future leaching rate, with ranked realizations as of Aug. 17, 2023) to illustrate (a) historical and forecasted nitrate leaching rates, (b) groundwater age distributions, and (c) forecasted nitrate concentrations in the well for the base, 5th, 50th (median), and 95th percentile realizations.

estimates of nitrate concentrations in wells in Wisconsin. The tool includes a hindcasting scenario and six forecasting scenarios for assisting users with assessing potential interactions among possible future nitrate leaching rates and well concentrations. A brief overview is provided here to aid readers in understanding how the design and calibration of the tool influences application and results of forecasting scenarios. All scenarios include plots that show results from the base parameter realization as well as results from three additional parameter realizations

that yield near-minimum (5th percentile), median (50th percentile), and near-maximum (95th percentile) simulated concentrations for the date the GW-NDST is run (or a user-specified date). These additional realizations are displayed for comparison and assessment of model uncertainty. Historical nitrate leaching, groundwater age distributions, and oxygen and nitrogen reduction parameters contained in the suite of realizations are not altered for the forecasting scenarios; only future nitrate leaching rates are modified in the scenarios.

The most basic forecast scenario, scenario number 1, applies the latest estimated nitrate leaching rate forward in time to the user's chosen planning date. Scenario 1 addresses the question of "how will estimated well concentrations change in the future if current leaching rates do not change?" Results for an example well (Fig. 8) illustrate that most of the historically measured concentrations fall between the 5th percentile (near-minimum) and the 95th percentile (near-maximum) realizations, as described in the Calibration results and model uncertainty section. The example results also illustrate that maintaining current (the date the GW-NDST was run; Aug. 17, 2023 for this example) nitrate leaching rates will produce nearly steady nitrate concentrations for the 5th percentile realization, but increased nitrate concentrations of about 20%–35% for the 50th percentile (median), base, and 95th percentile realizations over the next four decades, with the base realization still appearing to increase beyond 2063. These patterns are likely driven by the differing historical leaching trends and groundwater age distributions among the four plotted realizations. Similarly, while the historical and forecasted leaching rates are similar for the base and 50th percentile realizations, the forecasted concentrations show greater divergence due to their differing groundwater age distributions (lag times) and potentially due to differing oxygen and nitrate reduction rates.

The dynamic variability among the individual realizations highlights the value of using multiple realizations to illustrate uncertainty of historical trends and forecasted concentrations due to the range of feasible parameter values. For example, the 5th and 50th percentile realizations nearly equally match the concentrations measured around the year 2000, and both realizations nearly match one of the measurements collected over a decade later. However, the forecasted concentrations for the 5th and 50th percentile realizations diverge substantially after about 2010, highlighting the limitations of historical measurements for informing future concentrations. Thus, users of the GW-NDST are encouraged to consider the full range of forecasted nitrate concentrations when contemplating decisions informed by the tool, and potentially use the results to guide targeted study and data collection to further facilitate the decision-making process.

4. Discussion

The GW-NDST was built by integrating multiple independently developed support models to simulate the key processes affecting nitrate concentrations in wells: historical leaching rates, travel time distributions, and biogeochemical reactions. The calibration leveraged spatially distributed parameters to modify, or tune, the support model output as necessary to better match 34,255 historical nitrate concentration and trend targets. Judgements made during the calibration process, such as target weights and parameterization, are not unassailable but were informed by data, results, literature, and experience (Thompson, 2022). The calibration improved model performance, reducing ϕ of the base realization by 60 percent from about 200,000 for the initial iteration to about 80,000 on iteration 2, while also generating 450 unique parameter realizations with broadly sampled but reasonable parameter combinations that allowed 78 percent of all target concentrations to fall within the ensemble of model results. Nonetheless, the calibration illustrates that matching some individual samples was difficult, especially those above 20 mg-N/L. The reduction of the mean IBIS mult parameter value from 1.0 to about 0.6 likely degraded the match to samples above 20 mg-N/L, despite the authors preferentially increasing the target weight for the most accurate targets with concentrations greater than 10 mg-N/L by a factor of three (Table 1). This result suggests that simulated leached nitrate from the Agro-IBIS model is on average too high, while peak leaching rates may be too low. This dichotomy is likely a consequence of spatial averaging and limited site-specific data. That is, the contributing area for most domestic and non-community wells is likely on the scale of a few kilometers or less; similar to or smaller than the grid resolution of the Agro-IBIS model. This coarseness in the model grid

results in the combining of sources with very high leaching rates, such as individual fields, with areas having lower leaching rates, and results in rates that are on average both too high and too similar (reduced variability). Shrestha et al. (2023) illustrate large differences in nitrate leaching rates due to crop types and crop rotations, soils, precipitation events, and fertilizer application rates, as well as natural variability among individual sites. Future work to refine the Agro-IBIS grid resolution and nutrient application inputs, along with improved estimates of septic system leaching variability, might improve calibration results but could also be limited by a general lack of site-specific data, such as nutrient application rates on individual fields or even at individual farms; such information is generally not publicly available, if recorded. However, emerging research that uses machine learning and remote sensing to map livestock facilities may aid in providing better estimates of spatially explicit manure application rates (Robinson et al., 2022; Shea et al., 2022).

Calibration of the GW-NDST also included concentration difference targets intended to provide insight into trends and interannual variability. Example comparisons (Fig. 5) illustrated that the tool properly reproduced long-term trends and the direction of inter-annual change at some wells, but generally under-estimates the magnitude of temporal variability. The coarse resolution of the Agro-IBIS model, combined with the steady-state, unimodal (lacking a second peak for preferential flow) groundwater age distribution formula (eq. 1), may influence this limited inter-annual variability simulated by the tool. Smaller Agro-IBIS cells that better match with individual agricultural fields would allow improved estimates of field-specific practices affecting temporal variability, compared with averaging across larger areas that requires blending of land cover, nutrient and water management, and soil characteristics. However, publicly available field-specific datasets on crop management and nutrient application rates and timing are lacking, which limits the extent to which finer resolution modeling can improve results.

Similarly, the ageML model of Green et al. (2021; Kauffman et al., 2024) that was used for the GW-NDST does not account for certain factors that may produce large variations of concentrations. A potential factor in concentration variations is preferential flow due to natural (karst, fractured rock) or well construction factors. The ageML model was not trained on datasets of wells with obvious preferential flow. The unimodal form of the age distribution, based on Fickian transport, can simulate rapid response to inputs in samples with a mean age close to zero (wells near the water table in shallow groundwater), but does not account for scenarios with both preferential and diffuse transport. While Green et al. (2014) showed that several age distribution models, including multimodal distributions, often yield similar results for nitrate predictions, extreme differences in expected and actual age distributions can occur. Users interested in applying the GW-NDST for wells in areas well-known for having karst or fractured bedrock aquifers in Wisconsin will be informed by the GUI that the tool may have limited validity for their well, but users will not be prevented from applying the tool. Future efforts to enhance the GW-NDST may include introducing alternative or more flexible forms of the age distribution model, and including methods to assess the source area of nutrients captured by wells. Recognizing the limitations of the tool to reproduce observed interannual variability in concentrations, users are encouraged to consider the full range of possible outcomes depicted by all four of the realizations illustrated in the results when evaluating decisions and to also delay judgement of management change effectiveness based on only a few measurements collected shortly after management implementation.

The support model that estimates oxygen and nitrate reduction rates had less influence on predictions than the nitrate leaching and groundwater age support models based on the estimated sensitivity of parameters (Table 3). Reactions can, however, greatly affect nitrate concentrations in reduced groundwater (Böhlke and Denver, 1995), which does occur in Wisconsin aquifers (Kraft et al., 2008; Erickson et al., 2021; Tesoriero et al., 2017). Emphasis on wells with high

concentrations of nitrate and removal of target concentrations less than 1 mg-N/L may have selected for relatively non-reactive areas where denitrification has less influence on concentrations. Assessment of reactivity at the scale of the GW-NDST remains challenging because data to characterize oxygen and nitrate reduction conditions in aquifers are scarce. Future efforts may leverage growing datasets related to redox conditions (Erickson et al., 2021) and machine learning methods.

Demonstrating application of the tool for an example well and forecasting scenario (Fig. 8) highlights the value of translating the results of the calibration into terms of model uncertainty in the GW-NDST results. For instance, while the base realization (the calibration metrics highlighted in Fig. 3) over-simulated all measured concentrations from the example well, the four realizations bracketed all measured values. Moreover, while simulated nitrate from the 5th and 50th percentile realizations had arguably similar matches to the measurements, their forecasted concentrations diverged substantially, with concentrations deviating by nearly a factor of two (7 mg-N/L vs 13 mg-N/L, respectively) over the 40-year forecast. These observations highlight the difficulty in attempting to identify a single “best” forecast of an uncertain future. Part of the challenge lies with known uncertainty around parameters that are insufficiently informed by measured data. That is, most of the 11 calibration parameter groups exhibited relatively low sensitivity for calibrating the model to concentration and concentration difference targets (Table 3). However, the Iterative Ensemble Smoother’s ability to sample a diverse combination of parameter values across reasonable bounds enabled the tool to incorporate that uncertainty into visualizations of alternative possible outcomes. Additional data collection, such as N₂ and noble gas samples to inform denitrification rates or bromide tracer studies to inform leaching rates and groundwater ages (Cardiff et al., 2022), along with agronomic data, could potentially better constrain future calibration efforts by allowing for tighter starting bounds on parameters. Yet regardless of hindcasting performance of the GW-NDST (or other modeling tools), caution in interpreting forecast accuracy is prudent when used for decision making (Thompson, 2022). That is, while high numerical accuracy is beneficial and has been pursued in the development of this GW-NDST, the tool may also be effective for testing assumptions and limitations related to relative concentration changes stemming from nutrient management decisions. Indeed, many nitrate leaching reduction practices are themselves highly variable and uncertain in their effectiveness over the large range of natural conditions (Dinnes et al., 2002; Masarik et al., 2014; Esmaeili et al., 2020). For scenarios in which forecasts indicate that large leaching reductions will be required to achieve well concentration goals, multiple nitrate leaching reduction strategies may need to be applied, with every individual management practice incrementally increasing the potential for success (Dinnes et al., 2002; Hajhamad and Almasri, 2009).

5. Conclusions

The GW-NDST presented in this paper was developed as a framework to assist users with understanding relationships between nitrate leaching rates, groundwater age and lag times, geochemical reactions, and nitrate concentrations in wells. The tool is novel in its use of multiple support models as input sources, quantification and display of uncertainty, and ease of use with individual wells over a broad region (Wisconsin, USA). Results from the tool are displayed as graphs of historical and forecasted nitrate leaching rates, groundwater age distributions (travel time lags), and computed historical and forecasted nitrate concentrations in the well of interest. Each graph includes results from a set of parameter realizations that characterize the range of results informed by model uncertainty, as informed by the calibration. The “base” and 50th percentile realizations represent central tendencies, and the 5th (near-minimum) and 95th (near-maximum) percentile realizations illustrate the broader range of possible results.

This study focused on the design and calibration of the GW-NDST, including interpretation of how key processes influence the forecasted results from the ensemble of parameter realizations sampled from the posterior parameter probability distribution. Information on model uncertainty was incorporated into the tool via an ensemble of parameter realizations generated through the calibration processes, with an emphasis placed on avoiding over-fitting of model parameters and acceptance of a range of model variability within the bounds of reasonable parameter values. The calibration process involved adjusting eleven spatially variable parameter groups, eight of which represented multipliers applied to support model outputs to improve the simulated match to 34,255 nitrate sample targets. Application of an ensemble calibration method, PESTPP-IES (White et al., 2020), facilitated simultaneous improvement in model fit and quantification of model uncertainty. Targets were weighted based on their measurement and locational accuracy plus their concentration, such that the fewest high-accuracy and high-concentration targets were weighted most heavily. Nonetheless, the calibration analysis illustrates that the tool may under-estimate concentrations greater than about 20 mg-N/L, which is likely a function of the coarse resolution (1 km × 1.5 km) and limited site-specific nutrient application information available to the Agro-IBIS ecosystem model (Lark et al., 2022; Kucharik and Brye, 2003) that generated historical nitrate leaching rates for the tool, and simplification of groundwater age distributions that represent nutrient transport through aquifers. Despite the bias, 78% of all target concentrations, including concentrations above 20 mg-N/L, are shown to be captured in the range of results simulated by the ensemble of realizations. This high percent of overlap is one of the important benefits for including results from the realizations in the plots generated by the tool.

The value of illustrating model uncertainty was further evidenced through an example application of a forecasting scenario. The example demonstrates that long term forecasts can differ among equally likely realizations. Thus, identification of a “single best” forecast can lead to misinterpretations. Consideration of the uncertainty illustrated by a range of realizations is an important component of the decision support aspect of the tool. Nonetheless, subsequent data collection and enhancements to the tool’s support models along with targeted data collection for future re-calibration efforts could reduce uncertainty in future versions of the GW-NDST. It is hoped that use of this tool will lead to additional data and methods refinement, which will in-turn drive prediction improvements over time.

6. Software availability

The Groundwater Nitrate Decision Support Tool (GW-NDST) for Wisconsin is available at no cost from: <https://doi.org/10.5066/P13ETB4Q>. The tool requires the user to have an operating system capable of running Python 3.10 and can be run on Windows, Mac OS, or Linux. The software was written in Python, and an YAML installation file is available from the above website for installing a Python environment containing the specific Python libraries and versions known to properly execute the code. Full installation of the software involves: 1. Cloning the computer code from the above website, 2. Installing a Python environment based on the YAML file included at that site, and 3. Downloading input GIS data from Corson-Dosch and Juckem (2024a; <https://doi.org/10.5066/P9Q1X606>; 1.1 GB compressed; 4.9 GB uncompressed) and parameter files from Corson-Dosch and Juckem (2024b; <https://doi.org/10.5066/P9QHPVU3>; 2.8 GB) that are required by the tool’s machine learning and statistical support models. The downloaded GIS data and parameter files must be extracted from compressed files and saved to specific sub-directories within the directory structure of the cloned/downloaded software. Full instructions for downloading the software and data files, extracting and saving the data files, and running the software can be found with the source code at: <https://doi.org/10.5066/P13ETB4Q>.

Software name: GW-NDST, version 1.1.0.

Developers: Variable; authorship will increase with subsequent version updates. Please refer to the software repository for a current list of authors.

Contact information: pjuckem@usgs.gov, lschachter@usgs.gov, ncorson-dosch@usgs.gov, ctgreen@usgs.gov.

First year available: 2024.

Program language: Python; a YAML file is available from the software repository to aid with installing an environment known to work with the software code.

Program size: 112 MB; Total including downloaded data: 4.0 GB compressed; 7.8 GB uncompressed.

Cost: Free.

Disclaimer

No warranty, expressed or implied, is made by the USGS or the U.S. Government as to the functionality of the software and related material nor shall the fact of release constitute any such warranty. The software is provided on the condition that neither the USGS nor the U.S. Government shall be held liable for any damages resulting from the authorized or unauthorized use of the software.

Any use of trade, firm, or product names is for descriptive purposes only and does not imply endorsement by the U.S. Government.

CRediT authorship contribution statement

Paul F. Juckem: Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Software, Supervision, Validation, Visualization, Writing – original draft, Writing – review & editing. **Nicholas T. Corson-Dosch:** Data curation, Formal analysis, Investigation, Methodology, Software, Validation, Visualization, Writing – review & editing. **Laura A. Schachter:** Data curation, Formal analysis, Investigation, Methodology, Software, Visualization, Writing – review & editing. **Christopher T. Green:** Conceptualization, Formal analysis, Investigation, Methodology, Software, Validation, Writing – review & editing. **Kelsie M. Ferin:** Data curation, Formal analysis, Investigation, Writing – review & editing. **Eric G. Booth:** Data curation, Formal analysis, Investigation, Writing – review & editing. **Christopher J. Kucharik:** Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Supervision, Validation, Writing – review & editing. **Brian P. Austin:** Conceptualization, Funding acquisition, Investigation, Writing – review & editing. **Leon J. Kauffman:** Data curation, Formal analysis, Investigation, Software, Validation, Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The Groundwater Nitrate Decision Support Tool for Wisconsin (version 1.1.0) is available at <https://doi.org/10.5066/P13ETB4Q>.

Acknowledgements

Reviewers, including Brent Heerspink and Chanse Ford (both USGS), are thanked for their thoughtful review and comments on the manuscript and software code. This work was supported by the Wisconsin Department of Natural Resources, the U.S. Geological Survey's Integrated Water Availability Assessments and Cooperative Matching Funds programs, the University of Wisconsin-Madison, and the National Science Foundation Innovations at the Nexus of Food Energy, and Water

Systems (INFEWS) program (award number 1855996). Melinda Erickson provided numerous references cited in the introduction. Michael Fienen provided guidance on calibration with PESTPP-IES. Erik Smith developed an R-based version of the depth-to-water machine learning model that was subsequently converted to Python. Anna Baker, Erik Smith, Wonsook Ha, and James Kennedy helped with publishing of associated data releases. Wonsook Ha assisted with Fig. 1.

References

- Anderson, M.P., Woessner, W.W., Hunt, R.J., 2015. *Applied Groundwater Modeling: Simulation of Flow and Advective Transport*, second ed. Academic Press Inc., London, p. 564p.
- Baker, A.C., Juckem, P.F., Kennedy, J.L., Green, C.T., 2024. Compiled Age Tracer and Redox Chemistry Data for the State of Wisconsin, 1987–2009. U.S. Geological Survey Data Release. <https://doi.org/10.5066/P9FJ26BV>.
- Böhlke, J.K., Denver, J.M., 1995. Combined use of groundwater dating, chemical, and isotopic analyses to resolve the history and fate of nitrate contamination in two agricultural watersheds, atlantic coastal plain, Maryland. *Water Resour. Res.* 31 (9), 2319–2339. <https://doi.org/10.1029/95WR01584>.
- Brakebill, J.W., Gronberg, J.A., 2017. County-Level Estimates of Nitrogen and Phosphorus from Commercial Fertilizer for the Conterminous United States, 1987–2012. <https://doi.org/10.5066/F7H41PKX>.
- Byrnes, D.K., Van Meter, K.J., Basu, N.B., 2020. Long-term shifts in U.S. Nitrogen sources and sinks revealed by the new TREND-nitrogen data set (1930–2017). *Global Biogeochem. Cycles* 34 (9), 1–16. <https://doi.org/10.1029/2020GB006626>.
- Cardiff, M., Schachter, L.A., Krause, J., Gotkowitz, M.B., Austin, B.P., 2022. Quantifying annual nitrogen loss to groundwater via edge-of-field monitoring: method and application. *Groundwater* 61 (1), 21–34. <https://doi.org/10.1111/gwat.13217>.
- Cook, P.G., Böhlke, J.K., 2000. Determining timescales for groundwater flow and solute transport. In: Cook, P.G., L Herczeg, A. (Eds.), *Environmental Tracers in Subsurface Hydrology*, vols. 1–30. Springer US, Boston, MA. https://doi.org/10.1007/978-1-4615-4557-6_1.
- Corson-Dosch, N.T., Juckem, P.F., 2024a. GIS Files Required to Run the Groundwater Nitrate Decision Support Tool for Wisconsin. U.S. Geological Survey Data Release. <https://doi.org/10.5066/P9Q1X606>.
- Corson-Dosch, N.T., Juckem, P.F., 2024b. Parameter Ensemble Files Required to Run the Groundwater Nitrate Decision Support Tool for Wisconsin. U.S. Geological Survey Data Release. <https://doi.org/10.5066/P9QHPVU3>.
- Dinnes, D.L., Karlen, D.L., Jaynes, D.B., Kaspar, T.C., Hatfield, J.L., Colvin, T.S., Cambardella, C.A., 2002. Nitrogen management strategies to reduce nitrate leaching in tile-drained midwestern soils. *Agron. J.* 94 (1), 153–171. <https://doi.org/10.2134/agronj2002.0153>.
- Doherty, J.E., 2018. *PEST: Model-independent Parameter Estimation User Manual*, seventh ed. Watermark Numerical Computing, Addendum. Brisbane, Queensland, Australia.
- Doherty, J.E., Fienen, M.N., Hunt, R.J., 2010. Approaches to Highly Parameterized Inversion: Pilot-Point Theory, Guidelines, and Research Directions. U.S. Geological Survey Scientific Investigations. <https://doi.org/10.3133/sir20105168>. Report 2010-5168.
- Doherty, J.E., Hunt, R.J., 2010. Approaches to Highly Parameterized Inversion: A Guide to Using PEST for Groundwater-Model Calibration. *U. S. Geological Survey Scientific Investigations Report 2010-5169*. <http://pubs.usgs.gov/sir/2010/5169/>.
- Doherty, J.E., 2003. Ground water model calibration using pilot points and regularization. *Ground Water* 41 (2), 170–177. <https://doi.org/10.1111/j.1745-6584.2003.tb02580.x>.
- Erickson, M.L., Elliott, S.M., Brown, C.J., Stackelberg, P.E., Ransom, K.M., Reddy, J.E., 2021. Machine learning predicted redox conditions in the glacial aquifer system, northern continental United States. *Water Resour. Res.* 57 (4), 1–19. <https://doi.org/10.1029/2020WR028207>.
- Esmaili, S., Thomson, N.R., Rudolph, D.L., 2020. Evaluation of nutrient beneficial management practices on nitrate loading to groundwater in a southern ontario agricultural landscape. *Can. Water Resour. J.* 45 (1), 90–107. <https://doi.org/10.1080/07011784.2019.1692697>.
- Evans, M., Moshonov, H., 2006. Checking for prior-data conflict. *Bayesian Analysis* 1 (4), 893–914. <https://doi.org/10.1214/06-BA129>.
- Fienen, M.N., Corson-Dosch, N.T., White, J.T., Leaf, A.T., Hunt, R.J., 2022. Risk-based wellhead protection decision support: a repeatable workflow approach. *Ground Water* 60 (1), 71–86. <https://doi.org/10.1111/gwat.13129>.
- Gebert, W.A., Walker, J.F., Kennedy, J.L., 2011. “Estimating 1970–99 Average Annual Groundwater Recharge in Wisconsin Using Streamflow Data: U.S. Geological Survey Open-File Report 2009–1210” Open-File (2009–1210): 118. <http://pubs.usgs.gov/ofr/2009/1210/>.
- Green, C.T., Ransom, K.M., Nolan, B.T., Liao, L., Harter, T., 2021. Machine learning predictions of mean ages of shallow well samples in the Great lakes basin, USA. *J. Hydrol.* 126908 <https://doi.org/10.1016/j.jhydrol.2021.126908>.
- Green, C.T., Liao, L., Nolan, B.T., Juckem, P.F., Shope, C.L., Tesoriero, A.J., Jurgens, B.C., 2018. Regional variability of nitrate fluxes in the unsaturated zone and groundwater, Wisconsin, USA. *Water Resour. Res.* 54 (1), 301–322. <https://doi.org/10.1002/2017WR022012>.
- Green, C.T., Puckett, L.J., Böhlke, J.K., Bekins, B.A., Phillips, S.P., Kauffman, L.J., Denver, J.M., Johnson, H.M., 2008. Limited occurrence of denitrification in four

- shallow aquifers in agricultural areas of the United States. *J. Environ. Qual.* 37 (3), 994–1009. <https://doi.org/10.2134/jeq2006.0419>.
- Green, C.T., Zhang, Y., Jurgens, B.C., Starn, J.J., Landon, M.K., 2014. Accuracy of travel time distribution (TTD) models as affected by TTD complexity, observation errors, and model and tracer selection. *Water Resour. Res.* 50, 6191–6213. <https://doi.org/10.1002/2014WR015625>.
- Gronberg, J.A., Arnold, T.L., 2017. County-Level Estimates of Nitrogen and Phosphorus from Animal Manure for the Conterminous United States, 2007 and 2012." *Open-File Report*. <https://doi.org/10.3133/ofr20171021>.
- Haitjema, H.M., 1995. On the residence time distribution in idealized groundwatersheds. *J. Hydrol.* 172 (1), 127–146. [https://doi.org/10.1016/0022-1694\(95\)02732-5](https://doi.org/10.1016/0022-1694(95)02732-5).
- Hajhamad, L., Almasri, M.N., 2009. Assessment of nitrate contamination of groundwater using lumped-parameter models. *Environ. Model. Software* 24 (9), 1073–1087. <https://doi.org/10.1016/j.envsoft.2009.02.014>.
- Jones, E., Oliphant, P., Peterson, and And Others. n.d. "Scipy–Open Source Scientific Tools for Python." <http://www.scipy.org>.
- Juckem, P.F., Baker, A.C., Corson-Dosch, N.T., Smith, E.A., Schachter, L.A., Kauffman, L. J., Green, C.T., Ha, W.S., 2024. Data to Support a Groundwater Nitrate Decision Support Tool for Wisconsin. U.S. Geological Survey Data Release. <https://doi.org/10.5066/P9TTAQ18>.
- Juckem, P.F., Green, C.T., 2024. Multivariate Regression Model for Predicting Oxygen Reduction Rates in Groundwater for the State of Wisconsin. U.S. Geological Survey Data Release. <https://doi.org/10.5066/P9NPR21>.
- Kauffman, L.J., Green, C.T., Ransom, K.M., Ha, W.S., 2024. Histogram-based Gradient Boosted Regression Tree Model of Mean Ages of Shallow Well Samples in the Great Lakes Basin. U.S. Geological Survey Data Release, USA. <https://doi.org/10.5066/P9LFX0XP>.
- Kluyver, T., Ragan-Kelley, B., Fernando, P., Granger, B., Bussonnier, M., Frederic, J., Kelley, K., et al., 2016. Jupyter notebooks – a publishing format for reproducible computational workflows. In: *Positioning and Power in Academic Publishing: Players, Agents and Agendas*. F. Loizides and B. Schmidt, pp. 87–90.
- Korom, S.F., 1992. Natural denitrification in the saturated zone: a review. *Water Resour. Res.* 28 (6), 1657–1668.
- Kourakos, G., Harter, T., 2014. Vectorized Simulation of Groundwater Flow and Streamline Transport, vol. 52. *Environmental modelling & software*, pp. 207–221. <https://doi.org/10.1016/j.envsoft.2013.10.029>.
- Kraft, G.J., Browne, B.A., DeVita, W.M., Mechenich, D.J., 2008. Agricultural pollutant penetration and steady state in thick aquifers. *Ground Water* 46 (1), 41–50. <https://doi.org/10.1111/j.1745-6584.2007.00378.x>.
- Kreft, A., Zuber, A., 1978. On the physical meaning of the dispersion equation and its solutions for different initial and boundary conditions. *Chem. Eng. Sci.* 33, 1471–1480.
- Kucharik, C.J., Brye, K.R., 2003. Integrated Biosphere simulator (IBIS) yield and nitrate loss predictions for Wisconsin maize receiving varied amounts of nitrogen fertilizer. *J. Environ. Qual.* 32, 247–268. <https://doi.org/10.2134/jeq2003.2470>.
- Kuhn, M., Johnson, K., 2013. *Applied Predictive Modeling*, vol. 26. Springer. <https://doi.org/10.1007/978-1-4614-6849-3>. Springer Science+Business Media B.V.
- Lark, T.J., Hendricks, N.P., Smith, A., Pates, N., Spawn-lee, S.A., 2022. Environmental outcomes of the US renewable fuel standard. *Proc. Natl. Acad. Sci. USA* 119 (9), 8. <https://doi.org/10.1073/pnas.2101084119>.
- León, L.F., Lam, D.C., Swayne, D.A., Farquhar, G.J., Soulis, E.D., 2000. Integration of a nonpoint source pollution model with a decision support system. *Environ. Model. Software* 15 (3), 249–255. [https://doi.org/10.1016/S1364-8152\(00\)00011-6](https://doi.org/10.1016/S1364-8152(00)00011-6).
- Lusk, M.G., Toor, G.S., Yang, Y.Y., Mechtensimer, S., De, M., Obreja, T.A., 2017. A review of the fate and transport of nitrogen, phosphorus, pathogens, and trace organic chemicals in septic systems. *Crit. Rev. Environ. Sci. Technol.* 47 (7), 455–541. <https://doi.org/10.1080/10643389.2017.1327787>.
- Luther, K.H., Haitjema, H.M., 1998. Numerical experiments on the residence time distributions of heterogeneous groundwatersheds. *J. Hydrol.* 207 (1–2), 1–17. [https://doi.org/10.1016/S0022-1694\(98\)00112-7](https://doi.org/10.1016/S0022-1694(98)00112-7).
- Maloszewski, P., Zuber, A., 1982. Determining the turnover time of groundwater systems with the aid of environmental tracers: 1. Models and their applicability. *J. Hydrol.* 57 (3), 207–231. [https://doi.org/10.1016/0022-1694\(82\)90147-0](https://doi.org/10.1016/0022-1694(82)90147-0).
- Masarik, K.C., Norman, J.M., Brye, K.R., 2014. Long-term drainage and nitrate leaching below well-drained continuous corn agroecosystems and a prairie. *J. Environ. Protect.* 5, 240–254. <https://doi.org/10.4236/jep.2014.54028>.
- Mechenich, D., Johnson, A., 2022. *Interactive Well Water Quality Viewer Version 4.1*. Center for Watershed Science and Education, University of Wisconsin-Stevens Point/University of Wisconsin-Madison Division of Extension, 2022.
- Motew, M., Chen, X., Booth, E.G., Carpenter, S.R., Pinkas, P., Zipper, S.C., Loheide, S.P., Donner, S.D., Tsuruta, K., Vadas, P.A., Kucharik, C.J., 2017. The influence of legacy P on lake water quality in a midwestern agricultural watershed. *Ecosystems* 20 (8), 1468–1482. <https://doi.org/10.1007/s10021-017-0125-0>.
- Neitsch, S.L., Arnold, J.G., Kiniry, J.R., Williams, J.R., 2011. Soil & Water Assessment Tool Theoretical Documentation Version 2009. Texas Water Resources Institute, pp. 1–647. <https://doi.org/10.1016/j.scitotenv.2015.11.063>.
- Niswonger, R.G., Panday, S., Motomu, I., 2011. MODFLOW-NWT, A Newton formulation for MODFLOW-2005. In: *U.S. Geological Survey Techniques And Methods Book 6-A37*, vol. 44. <https://pubs.usgs.gov/tm/tm6a37/>.
- Nolan, B.T., Hitt, K.J., 2003. "Nutrients in Shallow Ground Waters beneath Relatively Undeveloped Areas in the Conterminous United States," U.S. Geological Survey Water-Resources Investigations Report 02-4289, p. 17p. <http://pubs.usgs.gov/wri/wri024289/>.
- Nolan, B.T., Hitt, K.J., 2006. Vulnerability of shallow groundwater and drinking-water wells to nitrate in the United States. *Environ. Sci. Technol.* 40 (24), 7834–7840. <https://doi.org/10.1021/es060911u>.
- Nolan, B.T., Green, C.T., Juckem, P.F., Liao, L., Reddy, J.E., 2018. Metamodeling and mapping of nitrate flux in the unsaturated zone and groundwater, Wisconsin, USA. *J. Hydrol.* 559, 428–441. <https://doi.org/10.1016/j.jhydrol.2018.02.029>.
- Nott, D.J., Wang, X., Evans, M., Englert, B.G., 2020. Checking for prior-data conflict using prior-to-posterior divergences. *Stat. Sci.* 35 (2), 234–253. <https://doi.org/10.1214/19-STS731>.
- R Core Team, 2019. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. <https://www.r-project.org/>.
- Ransom, K.M., Nolan, B.T., Stackelberg, P.E., Belitz, K., Fram, M.S., 2022. Machine learning predictions of nitrate in groundwater used for drinking supply in the conterminous United States. *Sci. Total Environ.* 807 (3), 151065. <https://doi.org/10.1016/j.scitotenv.2021.151065>.
- Robinson, C., Chugg, B., Anderson, B., Ferres, J.M.L., Ho, D.E., 2022. Mapping industrial poultry operations at scale with deep learning and aerial imagery. *IEEE J. Sel. Top. Appl. Earth Obs. Rem. Sens.* 15, 7458–7471. <https://doi.org/10.1109/JSTARS.2022.3191544>.
- Sarofim, M.C., Smith, J.B., Juliana, A.St, Hartin, C., 2021. Improving reduced complexity model assessment and usability. *Nat. Clim. Change* 11 (1), 9–11. <https://doi.org/10.1038/s41558-020-00973-9>.
- Schachter, L.A., Juckem, P.F., Corson-Dosch, N.T., Green, C.T., 2024a. A Groundwater Nitrate Decision Support Tool (GW-NDST) for the State of Wisconsin, Version 1.1.0. U.S. Geological Survey Software Release. <https://doi.org/10.5066/P13ETB4Q>.
- Schachter, L.A., Juckem, P.F., Baker, A.C., 2024b. Calculated Leached Nitrogen from Septic Systems in Wisconsin, 1850–2010. U.S. Geological Survey Data Release. <https://doi.org/10.5066/P9LT7536>.
- Shea, K., Schaffer-Smith, D., Muenich, R.L., 2022. Using remote sensing to identify liquid manure applications in eastern North Carolina. *J. Environ. Manag.* 317 (May) <https://doi.org/10.1016/j.jenvman.2022.115334>.
- Shrestha, D., Masarik, K., Kucharik, C.J., 2023. Nitrate losses from midwest US agroecosystems: impacts of varied management and precipitation. *J. Soil Water Conserv.* 78 (3), 141–153. <https://doi.org/10.2489/jswc.2023.00048>.
- Smith, E.A., Kauffman, L.J., Ha, W.S., Juckem, P.F., 2024. Python-HBRT Model and Groundwater Levels Used for Estimating the Static, Shallow Water Table Depth for the State of Wisconsin. U.S. Geological Survey Data Release. <https://doi.org/10.5066/P9942AHY>.
- Tesoriero, A.J., Gronberg, J.A., Juckem, P.F., Miller, M.P., Austin, B.P., 2017. Predicting redox-sensitive contaminant concentrations in groundwater using random forest classification. *Water Resour. Res.* 53 (8), 7316–7331. <https://doi.org/10.1002/2016WR020197>.
- Tesoriero, A.J., Puckett, L.J., 2011. O₂ reduction and denitrification rates in shallow aquifers. *Water Resour. Res.* 47 (12), 1–17. <https://doi.org/10.1029/2011WR010471>.
- Thompson, E., 2022. *Escape from Model Land: How Mathematical Models Can Lead Us Astray and what We Can Do about it*. Hachette Book Group, New York, NY, p. 247p.
- U.S. Geological Survey, 2016. National Water Information System Data Available on the World Wide Web (USGS Water Data for the Nation). <https://doi.org/10.5066/F7P5KJN>, 2016.
- Van Rossum, G., Drake, F.L., 2009. *Python 3 Reference Manual*. CreateSpace, Scotts Valley, CA.
- Varni, M., Carrera, J., 1998. Simulation of groundwater age distributions. *Water Resour. Res.* 34 (12), 3271–3281. <https://doi.org/10.1029/98WR02536>.
- Vogel, J.C., 1967a. Isotopes in hydrology. In: *Proceedings. International Atomic Energy Agency, Vienna, Austria*.
- Vogel, J.C., 1967b. Investigation of groundwater flow with radiocarbon. In: *Isotopes in Hydrology*, vols. 355–68. International Atomic Energy Agency, Vienna, Austria.
- Wei, X., Bailey, R.T., Records, R.M., Wible, T.C., Arabi, M., 2019. Comprehensive Simulation of Nitrate Transport in Coupled Surface-Subsurface Hydrologic Systems Using the Linked SWAT-MODFLOW-Rt3d Model, vol. 122. *Environmental Modelling & Software*, 104242. <https://doi.org/10.1016/j.envsoft.2018.06.012>.
- White, J.T., 2018. A model-independent iterative ensemble smoother for efficient history-matching and uncertainty quantification in very high dimensions. *Environ. Model. Software* 109 (June), 191–201. <https://doi.org/10.1016/j.envsoft.2018.06.009>.
- White, J.T., Hunt, R.J., Fienen, M.N., Doherty, J.E., 2020. Approaches to Highly Parameterized Inversion: PEST++ Version 5, a Software Suite for Parameter Estimation, Uncertainty Analysis, Management Optimization and Sensitivity Analysis. <https://doi.org/10.3133/tm7C26>.
- Wisconsin Department of Natural Resources, 2023. Groundwater Retrieval Network. <https://dnr.wisconsin.gov/topic/Groundwater/GRN.html>.
- Zuber, A., 1986. Mathematical models for the interpretation of environmental radioisotopes in groundwater systems. In: Fritz, P., Fontes, J.C. (Eds.), *Handbook of Environmental Geochemistry*, vol. 2. Elsevier, Amsterdam, pp. 1–59. *The Terrestrial Environment*.