Edinburgh Research Explorer

Evaluation of four modelling approaches to estimate nitrous oxide emissions in China’s cropland

Citation for published version:

Digital Object Identifier (DOI):
10.1016/j.scitotenv.2018.10.336

Link:
Link to publication record in Edinburgh Research Explorer

Document Version:
Peer reviewed version

Published In:
Science of the Total Environment

General rights
Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy
The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact openaccess@ed.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.
Evaluation of four modelling approaches to estimate nitrous oxide emissions in China’s cropland

Abstract: Process-based models are useful tools to integrate the effects of detailed agricultural practices, soil characteristics, mass balance, and climate change on soil N\textsubscript{2}O emissions from soil - plant ecosystems, whereas static, seasonal or annual models often exist to estimate cumulative N\textsubscript{2}O emissions under data-limited conditions. A study was carried out to compare the capability of four models to estimate seasonal cumulative N\textsubscript{2}O fluxes from 419 field measurements representing 65 studies across China’s croplands. The models were 1) the DAYCENT model, 2) the DNDC model, 3) the linear regression model (YLRM) of Yue et al. (2018), and 4) IPCC Tier 1 emission factors. The DAYCENT and DNDC models estimated crop yields with R\textsuperscript{2} values of 0.60 and 0.66 respectively, but both models showed significant underestimation for all measurements. The estimated seasonal N\textsubscript{2}O emissions with R\textsuperscript{2} of 0.31, 0.30, 0.21 and 0.17 for DAYCENT, DNDC, YLRM, and IPCC, respectively. Based on RMSE, modelling efficiency and bias analysis, YLRM performed well on N\textsubscript{2}O emission prediction under no fertilization though bias still existed, while IPCC performed well for cotton and rapeseed and DNDC for soybean. The DAYCENT model accurately predicted the emissions with no bias across other crop and fertilization types whereas the DNDC model underestimated seasonal N\textsubscript{2}O emissions by 0.42 kg N\textsubscript{2}O-N ha\textsuperscript{-1} for all observed values. Model evaluation indicated that the DAYCENT and DNDC models simulated temporal patterns of daily N\textsubscript{2}O emissions effectively, but both models had difficulty in simulating the timing of the N\textsubscript{2}O fluxes following some events such as fertilization and water regime. According to this evaluation, algorithms for crop production and N\textsubscript{2}O emission should be improved to increase the accuracy in the prediction of unfertilized fields both for DAYCENT and DNDC. The effects of crop types and management modes such as fertilizations should also be further refined for YLRM.
Key words: nitrous oxide; model simulation; cropland; DAYCENT; DNDC; linear regression model
1 Introduction

Anthropogenic greenhouse gas (GHG) emissions, a major contributor to climate change (IPCC, 2013), have increased rapidly across the world by 41% from 38.2 Pg CO$_2$ equivalent (CO$_2$-eq) in 1990 to 53.9 Pg CO$_2$-eq in 2012 (ESSD, 2017). However, increasing food demand and overuse of resources could accelerate total emissions in coming decades (Reay et al., 2012). Meanwhile, the Paris Agreement aims to limit global warming to “well below” 2 degrees Celsius, with an ambition to pursue efforts to limit warming to below 1.5 degrees Celsius, and many countries have already made commitments to participate towards achieving these goals.

As one of the world's most populous countries, with 29.3% of the world’s total emissions (Janssens-Maenhout et al., 2017), China is of key importance for mitigating global emissions, and has recently pledged “no-increase” in chemical fertilizer and pesticide in order to achieving peak GHG emissions by the year 2030 (UNFCCC, 2015).

Nitrous oxide (N$_2$O) has a global warming potential (GWP) of approximately 265-310 times that of carbon dioxide (CO$_2$) over a 100-year timescale (Watson et al., 1996; IPCC, 2007; IPCC, 2013) with an atmospheric lifetime of approximately 120 years (Prather, 1998). Global N$_2$O emissions increased to 9.2 Tg N$_2$O in 2012 from 5.4 Tg N$_2$O in 1970. N$_2$O contributes to secondary inorganic aerosol formation and thus haze pollution in addition to climate change (Liu et al., 2017; Lu and Tian, 2013; Tian et al., 2011, 2012). For China, N$_2$O emissions accounted for 16.4% of total national GHG emissions (Janssens-Maenhout et al., 2017). The most significant source of N$_2$O emissions was agriculture, accounting for 51% of total national N$_2$O emissions (FAO, 2015). Emissions from agriculture tripled from 0.36 to 1.21 Tg N$_2$O in China between 1970 and 2014 (FAO, 2015). Given the importance of this source of emissions, reducing uncertainty in its estimation is an important issue for China to effectively identify mitigate priorities.
The Intergovernmental Panel on Climate Change (IPCC) provided a default global N$_2$O emission factor intended for use in national inventories of 1.25% with the confidence interval of 0.25-2.25% for fertilizer-induced emission (FIE) from all cropland in 1997 (IPCC, 1997). That is, that 1.25% of nitrogen applied in crop systems is released as N$_2$O. This factor was subsequently updated to 1% with the confidence interval of 0.3-3.0% and 0.3% with the confidence interval of 0.0-0.6% (Tier 1 approach) from upland crops and paddy rice cultivation, respectively (IPCC, 2006). Generally, the emission factor approach makes it easy to calculate the FIE using applied N rate, but also leads to large uncertainties. Therefore, as recommended by the IPCC (2006), higher Tier methods should be developed to obtain more representative, country specific emission rates or spatially disaggregated emission factors that are region and crop-specific.

Linear or nonlinear regression models can be developed to estimate N$_2$O emissions from croplands as a function of field and management variables based on field measurements (Bouwman et al., 2002a; Gerber et al., 2016; Albanito et al., 2017). For example, Yue et al. (2018) published a China-specific multi-variate empirical model for N$_2$O emissions to identify the spatial variability caused by the major drivers. On the other hand, process-based models have been widely used to estimate N$_2$O emissions and potential effects of global climate change on the terrestrial ecosystems. Several dynamic process-based models have been developed to predict N$_2$O emissions informed by an understanding of key soil processes and mechanisms (e.g. SUNDIAL by Smith et al., 1997; DNDC by Li et al., 1992; DAYCENT by Ogle et al., 2010). Compared to regression models, most process-based models simulate the emissions of several GHGs (CO$_2$, CH$_4$, N$_2$O) considering environmental and management related factors, such as crop growth, soil properties, fertilization and climate (Li et al., 1992; Ogle et al., 2010). DAYCENT and DNDC models are both widely-used ecosystem biogeochemistry models adopted to simulate N$_2$O emissions all over the world (Abdalla et al., 2010). DAYCENT, the
daily step version of CENTURY, simulates C, N, P, K and S dynamics of agricultural and
natural systems among vegetation and soil pools (Parton et al., 1998; Del Grosso et al., 2001).

DNDC was originally developed to predict N\textsubscript{2}O and CO\textsubscript{2} emissions from arable soils and has
been extended to estimate C and N cycles for other ecosystems (Li et al., 1992; Li, 2000).

There are limitations and uncertainties in estimating N\textsubscript{2}O fluxes from process-based model
simulations, associated with the representation of the mechanistic processes. Frolking et al.
(1998) found that DNDC simulated very low N\textsubscript{2}O fluxes for a dry site in Colorado. In contrast,
Smith et al. (2008) produced accurate predictions of average seasonal N\textsubscript{2}O emissions from the
DNDC model for two sites in Eastern Canada, while the DAYCENT model underestimated
N\textsubscript{2}O emissions. This variability in performance implies that model inter-comparisons are useful
to determine the most appropriate tool for a specific region or cropping system. For many
countries, including China, model inter-comparisons are especially important since many
process-based models, in spite of their intent to be generic, were originally calibrated on data
from North-American or European cropping systems. The objective of this study is to compare
the results of four models, namely DAYCENT, DNDC, YLRM, IPCC default method, by
calibrating and evaluating the N\textsubscript{2}O emission estimates under different cropping systems and N
fertilizer application across the major agricultural regions of China.
2 Materials and methods

2.1 Model descriptions

DAYCENT, the daily time-step version of CENTURY, is a process-based ecosystem model developed to simulate carbon (C), N, P, K and S dynamics in plant-soil systems (Parton et al., 1998; Del Grosso et al., 2001). The nitrogen fluxes through the plant, residue and soil organic matter pools are coupled with C and estimated based on the C transfer between conceptual soil C pools, and the C:N ratio of organic matter. The model considers symbiotic and asymbiotic N fixation, and fertilizer additions. Losses of N occur through removal of vegetation, nitrification, denitrification, NH$_3$ volatilization, leaching and run-off. Daily weather data, essential management events, and soil texture data are needed as model inputs (Table 1).

The DeNitrification - DeComposition model (DNDC), contains four main sub-models as follows: the soil climate sub-model calculating hourly and daily soil temperature and moisture fluxes in one dimension; the crop growth sub-model simulating crop biomass accumulation and partitioning; the decomposition sub-model calculates decomposition, nitrification, NH$_3$ volatilization and CO$_2$ production; and the denitrification sub-model tracking the sequential biochemical reduction from nitrate (NO$_3$) to NO$_2^-$, NO, N$_2$O and N$_2$ (Li et al., 1992; Li, 2000; Abdalla et al., 2010). The DNDC model v.9.5 was applied in the present study (http://www.dndc.sr.unh.edu/) using the input data in Table 1.

A linear regression model approach has also been applied (Yue et al., 2018), named as YLRM, fitting cumulative N$_2$O emissions ($Cum$ N$_2$O) in kg N ha$^{-1}$ season$^{-1}$ based on the following equation:

$$Cum\ N_2O = Exp(-2.709 + 0.004 \times N\ rate + 0.074 \times Temp + 0.013 \times Clay + \beta_1 \times crop\ type + \beta_2 \times N\ rate \otimes fert\ type + \epsilon)$$  \hspace{1cm} (1)
where $N \ rate$ represents the nitrogen fertilizer application rate in kg N ha$^{-1}$; $Temp$ is the annual average air temperature ($^\circ$C); $Clay$ indicates the fraction of clay (%); values of $\beta_1$ for the different crop type classes are 0 for legume, 0.700 for upland crops, -0.188 for rice; and values of $\beta_2$ for the different base fertilizer are 0 for mineral fertilizer and -0.002 for organic fertilizer, and 0 for no fertilizer applied. The required data are N fertilizer application rates, annual average air temperature, soil clay content, crop type, and fertilizer type (Table 1).

Finally, using the IPCC tier 1 emission factor method (IPCC$_{EF}$), annual cumulative N$_2$O emissions ($Cum \ N_2O$) in kg N ha$^{-1}$ year$^{-1}$ are calculated using the following equation:

$$Cum \ N_2O = N \ rate \times EF$$  (2)

where $N \ rate$ represents the nitrogen fertilizer application rate in kg N ha$^{-1}$; and values of $EF$ are 0.01 and 0.003 for upland crops and paddy rice cultivation, respectively. The only required data are N fertilizer rates for this model (Table 1).

2.2 Data sources and model setup

N$_2$O emissions data were collected during the crop growing season at the experimental sites (kg N ha$^{-1}$ season$^{-1}$) - defined as the period from planting to harvest for a given crop. We conducted a literature search in the databases: CNKI, ISI-Web of Knowledge and Google Scholar, with the search words “nitrous oxide”, “emission”, “chamber”, and “China”. A total of 134 papers were found and processed according to the publication date, journal category and data integrity. For these papers, a dataset of 65 studies with a total of 419 field N$_2$O emission measurements were compiled. The dataset included the following information: cumulative N$_2$O emissions; grain yields; geographic information; soil characteristics including clay content, C and N content, bulk density, and pH; cropping system; management practices; crop types - maize (MA), wheat (WH), rice paddy (RP), other crops (including soybean, cotton, rape, and fallow); and fertilizer types classified into 5 broad categories - Control, Mineral, Organic,
Mineral & Organic (M_O), Controlled-release fertilizer or Nitrification inhibitor (CRF_IN).

More detailed information is provided in Table S1. All data were used to test DAYCENT, DNDC, YLRM and the IPCC_EF. It should be noted that 264 N2O field emission measurements of the whole database (419 measurements) were used to derive the linear regression model of Yue et al. (2018), with all the measurements used to test the YLRM. Therefore, the YLRM model is not entirely independent of the evaluation data.

We used consistent model driving datasets for all the models to minimize any uncertainty arising from using different input datasets (Tian et al., 2018). Most of the soil, crop, and cultivation management data were obtained from the dataset. Missing soils data that were not provided in the papers were extracted from China Soil Scientific Database (http://www.soil.csdb.cn/) based on the soil type documented for the experimental site. Climate data, including daily maximum/minimum temperature and precipitation, were obtained from the China Meteorological Data Sharing Service System (http://new-cdc.cma.gov.cn:8081/home.do) for the station nearest to the reported site. Regional nitrogen deposition data were based on Xu et al. (2015).

For the process-based models, most of the parameters were based on prior research (DAYCENT with information from Cheng et al., 2014; DNDC from Abdalla et al., 2010). To initialize the relative size of each SOC pool (five pools with a specific turnover rate associated with each pool), historical runs were performed at the beginning of the DAYCENT simulations. The process of performing the historical run developed by Cheng et al. (2014) was used in this study. However, historical runs were not needed for DNDC, and the relative size of each SOC pool was set to default values according to the soil properties (Jiang et al., 2017).

Crop growth directly controls soil water and C and N regimes, and hence is crucial for a biogeochemical model to correctly simulate trace gas flues, such as N2O (Hu et al., 2012). PRDX (the maximum potential production parameter) is a coefficient (a dimensionless
constant) representing the maximum production of different crop varieties based on genetic
variation, which can be optimized by simulating crop yields in the range of 1-3 for DAYCENT.
Similarly, the indices of maximum biomass production, biomass fraction, and biomass C/N
ratio of grain, leaf, stem, and root distributions were optimized for yield simulations of field
conditions for DNDC.

2.3 Model validation

2.3.1 Evaluation of daily N$_2$O emissions

DAYCENT and DNDC simulation results were evaluated against field measurements of N$_2$O
emissions by comparing the association between observed and simulated temporal patterns of
N$_2$O fluxes, as well as comparing the coincidence between observed and simulated emission
values. Five representative benchmark sites were selected from the major regions of China to
conduct the model evaluation of daily N$_2$O emissions under typical cropping systems (Table
2). Daily observed emission values for model evaluation were extracted either directly from
tables, or were extracted from the figures using Getdata Graph Digitizer software
(http://www.getdata-graph-digitizer.com/).

2.3.2 Model accuracy

The performance accuracy of the models was assessed for simulation of grain yield and
seasonal cumulative N$_2$O fluxes. Cumulative N$_2$O fluxes were estimated as the sum of
simulated daily fluxes for DAYCENT and DNDC models, and directly by YLRM and
IPCC_EF. Based on quantitative methods in Smith et al. (1997), the total difference between
observed and simulated values was assessed by calculating the root mean squared error (RMSE)
and relative root mean squared error ($r$RMSE) using equation (3) and (4).

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(S_i - O_i)^2}{n}} \quad (3)$$
where, $S_i$ and $O_i$ represent the simulated value from the models and the observed value from the original studies; $i$ is a single observation, $n$ is the number of target values; and $\bar{O}$ represents the mean value of the observed data. The $rRMSE$ can be compared between different models whose errors are measured in the different units, and a low $rRMSE$ often indicates a strong predictive power.

Second, the accuracy of the model simulation was determined based on a calculation of modelling efficiency ($ME$) as shown in equation (5) (Nash and Sutcliffe, 1970). $ME$ provides a comparison of the efficiency of the chosen model compared to describing the data as the mean of the observations.

$$ME = \frac{\sum_{i=1}^{n}(O_i-\bar{O})^2 - \sum_{i=1}^{n}(S_i-\bar{O})^2}{\sum_{i=1}^{n}(O_i-\bar{O})^2}$$  \hspace{1cm} (5)$$

Values for $ME$ can be positive and negative with a maximum value of 1. Specifically, a positive value shows that the simulated values describe the trend in the observed data better than the mean of the observations, and closer to 1 suggests a better modelling efficiency. A negative value indicates that the simulated values describe the data less well than a mean of the observations.

Third, the bias in the total difference between simulated and observed values was determined by calculating the mean difference ($M$) with Equation (6) (Addiscott & Whitmore, 1987; Smith et al., 1997). With the estimated $M$, the $t$ statistic was used to show a significant difference between the simulated estimates and measurements, using Equation (7).

$$M = \frac{\sum_{i=1}^{n}(O_i - S_i)}{n}$$  \hspace{1cm} (6)$$

$$t = \frac{M \times \sqrt{n}}{\sqrt{\sum_{i=1}^{n}[(O_i - S_i) - M]^2 / (n-1)}}$$  \hspace{1cm} (7)$$
A t-value greater than the critical two-tailed 2.5% t-value indicates that the simulated values have a significant bias towards over- ($M < 0$) or underestimation ($M > 0$) compared to observed values.

Correlation between simulated and observed values was used to assess whether simulated values follow the same pattern as observed values, and an R-squared value was used to determine the strength of the correlation (Smith et al., 1997).

All the statistical analyses were calculated to evaluate model performance for each fertilizer and crop type individually. Note that the IPCC_EF was not used for control (no fertilizer) practice because the emissions with no N fertilization were excluded in emission factor development. All the statistical analyses were conducted in R version 3.4.0 (R Core Team, 2017) or Microsoft Excel 2013.
3 Results

3.1 Yield simulation

The observed grain yield data included 288 individual observations, which ranged from 400 to 15700 kg ha$^{-1}$ with an average value of 7524 kg ha$^{-1}$ (Table S2). The simulated yields ranged from 537 to 16657 kg ha$^{-1}$ for DAYCENT, and from 608 to 14731 kg ha$^{-1}$ for DNDC. As shown in Fig. 1, the correlation of simulated against observed yields of maize, wheat, and paddy rice had R² values of 0.52, 0.33, and 0.58 for DAYCENT, which were lower than the values of 0.59, 0.35, and 0.71 for DNDC, respectively. DNDC had the lower rRMSE and higher ME for maize and paddy rice, but higher rRMSE and lower ME for wheat than that of DAYCENT (Table 3). According to the M and t statistic, the two process-based models significantly underestimated the maize yields by 1439 kg ha$^{-1}$ for DAYCENT and 1131 kg ha$^{-1}$ for DNDC. Generally, DNDC model had the better performance on simulation of crop yields with the lower rRMSE and higher ME and R² than DAYCENT though both models had relatively high bias estimates of 814.2 and 621.3 kg ha$^{-1}$ for DAYCENT and DNDC, respectively.

3.2 Validation of daily N₂O emissions

Seasonal patterns of daily N₂O emissions were analysed for 5 sites with latitudes between 28.6° to 47.4° and longitudes from 113.3° to 133.3° (Table 2), representing several different climate regions and most common cropping systems in China. Seasonal emission patterns simulated by the DNDC and DAYCENT models were generally similar to the observed values for most of the experimental period. Also, a significant increasing trend in N₂O emissions was simulated with increasing N application rates, corresponding with experimental observations. Both DAYCENT and DNDC models failed to model the specific timing and magnitude of daily N₂O emission peaks. Overall, the DNDC model overestimated emissions on days with high precipitation by a factor of around 2, particularly at the upland sites (Fig. 2c and 2d). The
DAYCENT model overestimated the fluxes upon drainage of rice cultivation systems (Fig. 2a, 2b and 2e). According to Fig. S1 and Table S3, the correlation of simulated versus observed daily emission values had $R^2$ values of 0.37, 0.21, 0.60 and 0.27 for site a, b, c and d for DAYCENT, respectively, though a positive ME of 0.23 was only found in site a. DNDC had an $R^2$ of 0.43 in site c, but all the values of ME were negative for DNDC.

3.3 Validation of cumulative N$_2$O emission

The observed emissions from 419 field N$_2$O emission measurements across 65 studies ranged from 0.02 to 7.39 kg N ha$^{-1}$ (1.18 kg N ha$^{-1}$ on average) with N fertilizer applied in the range of 0-600 kg N ha$^{-1}$ (Table S4). The correlation of simulated versus observed emission values had $R^2$ values of 0.31, 0.30, 0.21, and 0.17 for DAYCENT, DNDC, YLRM, and IPCC_EF, respectively (Fig. 3). DAYCENT had the lowest RMSE and rRMSE with the values of 0.97 kg N ha$^{-1}$ and 82.4%, and the highest ME of 0.28 compared to other approaches (Table 4). The IPCC default method had the highest RMSE and rRMSE with the values of 1.24 kg N ha$^{-1}$ and 104.9%, and a negative ME of -0.16. According to the M and the t test of bias, the DNDC simulated values were significantly different from the observed values with a bias of 0.42 kg N ha$^{-1}$, but the other three models showed no significant bias between the simulated and observed values (Table 4).

3.4 Model performance for fertilizer and crop effects on N$_2$O emission

We also assessed the four models’ performance with respect to estimating seasonal cumulative N$_2$O emissions for different fertilizer and crop types (Table 5 and 7). YLRM estimated N$_2$O emissions under no fertilizer input with the lowest rRMSE (102%) and positive ME (0.03) though with a significant overestimation of 0.16 kg N ha$^{-1}$. DAYCENT had the best values of rRMSE and ME for other fertilizer types including mineral, organic, M_O and CRF_IN with no bias. In fact, only the DNDC model significantly underestimated the emissions under the
mineral fertilizer treatments. The YLRM and IPCC_EF did not perform well for organic fertilizer treatment and combined mineral and organic fertilization treatment with the negative ME values and significant bias.

N management, and particularly additions, are the most important drivers of soil N$_2$O emissions (Del Grosso et al., 2009). Given this fact, we further compared the correlations of N addition rates with observed emission values for the models (Fig. 4). Both the simulated and observed values had a similar response to fertilizer application rate. The DNDC and YLRM simulated values were lower than observed values in rice cropping system (Fig. 4a); in addition, the values simulated by DNDC were all lower than observed values in the upland cropping system (Fig. 4b). The range of the slopes were 0.002~0.004 and 0.004~0.005 for paddy rice and upland cropping systems, which were very similar with the observed slopes of 0.003 and 0.004 (Table 6). For IPCC default method, the value of 0.003 for paddy rice fields was the same as the observed slope, while the emission factor of 0.01 for upland from IPCC_EF was much higher than the values for observed emissions.

For crop types, DAYCENT had the lowest RMSE and rRMSE values for maize, wheat, and paddy rice; and had the highest ME values for maize and wheat (Table 7). All the models showed no significant bias between the simulated and observed values for maize. The IPCC default method performed well for cotton and rapeseed, as did DNDC for soybean. However, none model performed well for fallow treatments.
4 Discussion

Given the recognised difficulty in estimating N$_2$O emissions precisely and the ongoing challenge of developing models which perform over a wide range of conditions, model inter-comparisons are an important way to determine a best candidate model for a given region and to identify potential ways to reduce the uncertainties. Model inter-comparisons to estimate N$_2$O emissions have previously been carried out in several studies (Frolking et al., 1998; Smith et al., 2008; Brilli et al., 2017; Erhardt et al., 2018).

Reasonable simulation of crop yield is of key importance to accurately predict N$_2$O emissions for process-based models of plant-soil systems. The correlation of simulated against all observed values had R$^2$ values of 0.60 and 0.66 for DAYCENT and DNDC models, and both DAYCENT and DNDC models significantly underestimated yields by 10.8% and 8.2% of the average value (7522 kg ha$^{-1}$) for all yield measurements. For maize yields, the bias was 16.4% and 12.9% of the average value (8750 kg ha$^{-1}$) for DAYCENT and DNDC, respectively. The bias for maize yields simulations with DAYCENT model were higher than the value of 521.59 kg ha$^{-1}$ (underestimation) reported from Cheng et al. (2014). Similar levels biases for wheat grain yield (Ludwig et al., 2011) and cotton plant biomass (Cui et al., 2014) were also previously reported for DNDC. In this study, the bias could be due to the limitation associated with climate data which included only the maximum/minimum temperature and precipitation for the two process-based models, whilst other climate parameters e.g. humidity and wind speed were not available. This limitation may result in large uncertainties in the model simulations (e.g. influence of humidity on transpiration rates and water stress). The model’s performance could be improved if those variables were available in the climate dataset. We found that the simulated yields under control treatment were low for DAYCENT and DNDC, which resulted in large biases being 1331 and 2246 kg ha$^{-1}$ compared with observed values, respectively (Table 3). Production algorithms in DAYCENT and DNDC may be too sensitive
to N availability. Sansoulet et al. (2014) also found that DAYCENT was not effective in predicting biomass under limited N rates.

The models were able to simulate daily N$_2$O flux over time with less bias compared to the grain yields; however, there were some abnormal peak periods of emissions simulated by both models, compared to the observed emissions. Specifically, peaks simulated by the DNDC for maize and wheat after heavy rainfall events were large (Fig. 2c-2e), indicating N$_2$O emissions are highly sensitive to soil moisture dynamics in the models (Lessard et al., 1996; Frolking et al. 1998; Smith et al., 2002). In addition, Smith et al. (2008) observed that DAYCENT and DNDC models both had difficulty in capturing soil water content accurately, which are linked to soil texture. Groffman and Tiedje (1989) suggested that the smaller average pore size in finer textured soils leads to greater soil water retention and greater opportunity to create anaerobiosis, while denitrification occurs at lower rates in a well-drained coarse-textured soil (Bouwman et al., 2002a, b). Thus, there may be an opportunity to further resolve the relationship between soil texture and water-filled pore space, and improve model predictions. Also, the accuracy of capturing N$_2$O emission peaks is associated with the frequency of sampling; with low frequency sampling (e.g., once a week or month) implying that some of the peaks simulated by the models could be missed in measurement time series.

The R$^2$ between simulated and observed seasonal cumulative N$_2$O emissions with the four models ranged from 0.17 to 0.31, which is lower than that of yield simulated values. N$_2$O emissions are inherently difficult to predict precisely for the reasons stated above; however, this does suggest considerable opportunity for improvement. No significant bias was identified for DAYCENT, YLRM and IPCC_EF except for the significant bias of 0.42 kg N ha$^{-1}$ (35.6%) for the DNDC model (Table 4 and S3). Beheydt et al. (2007) reported an overestimation of 7.4 kg N$_2$O-N ha$^{-1}$ for DNDC based on 22 long-term N$_2$O field experiments. In addition, other research found that DAYCENT performed well with an R$^2$ of 0.78 (Cheng et al., 2014), which
was much higher than the value we found in this study. However, this study used more field measurements than Cheng et al. (2014), which may have added heterogeneity and uncertainty in model simulation. Abdalla et al. (2010) indicated that DAYCENT performed poorly when simulating control plots with no fertilizer application an N\textsubscript{2}O flux of 57%, less than the observed values compared to the value of 27% estimated in the study. Given these discrepancies for N\textsubscript{2}O simulation from China’s cropping system under no N input condition, the model needs to be further parameterized. Additionally, several studies have indicated that model accuracy could vary for different fertilizer and cropping types (Smith et al., 2002; Cheng et al., 2014; Necpalova et al., 2018). Regardless, DAYCENT had the lowest rRMSE value for all the fertilizer types and major crops in China (maize, wheat and rice). DNDC did not accurately simulate N\textsubscript{2}O emissions under the control with no fertilization, mineral N fertilizer or combined mineral and organic N fertilization, but performed relative well under organic fertilization and with nitrification inhibitor addition (Table 5). In contrast, Smith et al. (2002) found the DNDC model predictions of N\textsubscript{2}O fluxes from control, manure, and mineral N fertilization corresponded well with observed measurements from maize in Canada. Regardless, Li et al. (2017) reported that DNDC may be not suitable for China as it lacks a number of features which are crucial for representing Chinese agro-ecosystems, especially paddy rice cultivation, complex and multiple cropping systems, and intensive management practices.

The four models are used for different purposes, and are different in scope and function. The predictions of YLRM showed better performance than IPCC\_EF (Table 4, 5 and 7). While the YLRM model was used only to calculate fertilizer-induced N\textsubscript{2}O emissions based on the underlying datasets that were used to derive these functions, this does indicate that if a reasonably comprehensive dataset of N\textsubscript{2}O emissions exists for a given region, then better predictions might be obtained from a linear regression model than by using IPCC default
method. In fact, the IPCC guidelines (2006) recommends using regional datasets for deriving country-specific emission factors for key sources rather than using the default factors. The two process-based models, in theory, should be able to capture more heterogeneity and be applied across a broader range of croplands in China. As indicated by this evaluation, the process-based models (especially DAYCENT) performed better than YLRM and IPCC_EF for some crop types and under some fertilization modes (Table 4, 5 and 7). One of the key strengths of DAYCENT is the initialization of SOM pools to accurately represent soil carbon stocks, and the linkage between C and N flows through the plant-soil system. DNDC also has strengths related to fertilizer applications at varying depth, and a more mechanistic representation of the N cycle with Michaelis-Menten dynamics (Li et al., 2006). Process-based models, such as DAYCENT and DNDC, can also represent more management impacts than empirical functions, particularly if data are limited for fitting a statistical model. For example, Xu et al. (2000) showed a significant effect of splitting fertilizer into three or more applications in DNDC, reducing N$_2$O emissions by 25%. Field practices of irrigation and tillage also influence N$_2$O fluxes, and their impacts can be represented in these simulation models. Our results indicate that the accuracy of model simulations differs across a range of N rates. Cheng et al. (2014) showed DAYCENT tended to underestimate N$_2$O emissions at higher observed emission rates, which were also seen for paddy rice in Fig. 4a. Albanito et al. (2017) studied the emission factors of N$_2$O and found that empirically derived factors tended to decrease with the N application rates approaching 1% in crops fertilized above 300 kg N ha$^{-1}$, and the IPCC_EF do tend to underestimate N$_2$O emissions by approximately 21% below a fertilization of 200 kg N ha$^{-1}$. Similarly, Shcherbak et al. (2014) indicated that the IPCC_EF underestimate and overestimate N$_2$O emissions in croplands fertilized above and below the threshold of approximately 150 kg N ha$^{-1}$. Sansoulet et al. (2014) also showed the different sensitivity under limited and high N rates. The negative intercept for DNDC might indicate
that emissions are under-estimated with no fertilizer applied, which was consistent with the bias of 0.49 kg N ha⁻¹ shown in Table 5.

Environmental factors (especially climate) and management activities (e.g. N fertilizer, tillage, straw return, irrigation) influence N₂O producing processes over both temporal and spatial scales, resulting in heterogeneous N₂O emissions at field level (Flessa et al., 2002). Cumulative seasonal N₂O emissions based on the closed static chamber method were used in most of the experiments at monthly or weekly intervals, which may lead to high inherent variability of N₂O fluxes. Ju et al. (2011) showed that a sampling frequency of 3 or 6 days led to 112-236% overestimation of total N₂O emissions. Process-based models may predict a flux peak during times, such as after a rainfall event, which is not represented in observational datasets with a low sampling frequency. Hence, an overestimation or underestimation of N₂O fluxes from upland soils may occur with static chambers, and more continuous measurements will likely reduce uncertainties in evaluating models (Ju et al. 2011).

As indicated in Materials and Methods Section, 264 of 419 observations was used to fit the linear regression, and the YLRM model is not entirely independent of the evaluation data. This could favour a better model performance of the YLRM compared to other models, but the degree to which this impacts the findings is not clear. To address this, a correlation analysis was conducted of simulated N₂O emission with observed values using the 264 observations used to fit the YRLM model. As shown in Fig. S2, the R² was only 0.27 for YLRM, compared to the R² of 0.31 for both DAYCENT and DNDC in Fig. 3. This indicated that the YLRM was not an improvement over the process based models even though the regression model was based on these data. Therefore, the non-independence of evaluation data did not significantly affect our conclusion. Nevertheless, there are potential improvements for the process based models (Nečpalová et al., 2018). According to our results, algorithms for crop production and N₂O emission should be improved to increase the accuracy in the prediction of unfertilized
field both for DAYCENT and DNDC. While both the two process based models did not simulate the N$_2$O emissions for cotton and rapeseed well, the performance of DNDC was better than DAYCENT for soybean. Parameters related to these crop types could be optimized using more local data as a potential improvement for applying these models in China. Given the abnormal peaks of N$_2$O fluxes induced by heavy rain event simulated by DNDC (Fig. 2c and 2d), there may be an opportunity to further refine the algorithms and optimize parameters based on the relationship between soil moisture and N$_2$O emissions. In addition, the effects of crop types and management modes such as organic fertilization should also be further refined for YLRM.
5 Conclusion

The performance of the four models to estimate crop yields and N₂O emissions in Chinese croplands varied with types of cropping systems and N fertilization management practices. From the four models investigated, the DAYCENT model emerged as the best performing model for estimating N₂O emissions under Chinese condition. Process-based models, at least in theory, should produce more accurate and precise estimates of emission with much detailed data for soil and crop properties. However, it may be feasible to adopt the methods based on linear regression models to calculate the seasonal N₂O emissions with limited data according to the compared to the IPCC tier 1 method if there is not sufficient data or capacity to simulate N₂O emissions with DAYCENT or DNDC. Both of these models require considerably more information to estimate emissions compared to the linear regression approach. In order to refine empirical models and improve the suitability of process-based models for Chinese croplands, we recommend further development of these models to represent regional conditions associated with climate, soil properties, cropping systems and local agricultural practices.

Acknowledgments

This work was financially supported by “the Fundamental Research Funds for the Central Universities” under a grant number KJQN201673, and China Natural Science Foundation under a grant number 41877546 and 41501569. This work also contributes to the activities of NCircle - a BBSRC-Newton Funded project (BB/N013484/1), and the project “Comparison study of models in agroecosystems for simulating the low-emission development strategies” supported by Environmental Defense Fund. The first author thanks the China Scholarship Council (CSC) for funding to support study at University of Aberdeen, UK.


Janssens-Maenhout, G., Crippa, M., Guizzardi, D., Muntean, M., Schaaf, E., Olivier, J. G. J., ...
& Schure, K. M. (2017). Fossil CO₂ and GHG emissions of all world countries. EUR


their contributions to global climate warming. Journal of Geophysical Research


Convention on Climate Change (UNFCCC), Paris.

emissions from cold paddy field in the Sanjiang Plain. Transactions of the Chinese Society
of Agricultural Engineering, 24(10):170-176. (Chinese in English Abstract)

Watson, R.T., Zinyowera, M.C., Moss, R.H., Dokken, D.J. (1996). Climate change 1995,
impacts, adaptation and mitigation of climate change. Scientific Technical Report
Analysis. In: Watson, R.T., Zinyowera, M.C., Moss, R.H. (Eds.), Contribution of
Working Group II to the Second Assessment Report of the Intergovernmental Panel on

Xu, W., Hong, Y., Chen, X., & Li, C. S. (2000). N2O emissions from regional agricultural

Xu, W., Luo, X.S., Pan, Y.P., Zhang, L., Tang, A.H., Shen, J.L., Zhang, Y., Li, K.H., Wu, Q.H.,
Yang, D.W. and Zhang, Y.Y. (2015). Quantifying atmospheric nitrogen deposition
through a nationwide monitoring network across China. Atmospheric Chemistry and
Physics, 15(21), 12345-12360.

from croplands across mainland China. Agriculture Ecosystems & Environment, 268, 70-
78.

Zhou, Z., Xu, X., Bi, Z., Li, L., Li, B., & Xiong, Z. (2016). Soil concentration profiles and
diffusion and emission of nitrous oxide influenced by the application of biochar in a rice-
wheat annual rotation system. Environmental Science and Pollution Research, 23(8), 7949-7961.
Table captions

Table 1 Data inputs for model simulations

Table 2 Information of sites selected for simulating daily N₂O emissions

Table 3 Statistics describing the performance of DAYCENT and DNDC models for grain yield simulations

Table 4 Statistics describing the performance of four models for N₂O emissions simulations

Table 5 The performance of four models for estimating N₂O emissions under different fertilizer management

Table 6 Statistics describing the correlations of observed or simulated N₂O emissions with nitrogen fertilizer application rates shown in Fig. 4.

Table 7 The performance of four models for estimating N₂O emissions associated with crop types
Figure captions

Fig. 1 Comparisons of observed and simulated crop yields for experimental sites across China. Dotted lines represent the correlation lines and solid lines represent 1:1 (y=x) line.

Fig. 2 Comparison of observed and simulated daily patterns of N₂O emissions for selected sites. Solid arrows show the times of management practices: FL for flooding in rice-growing stage; FE for fertilizer application; D for drainage. (a, single rice in Northeast; b, double rice in South-Central; c, maize in Northeast; d, maize in Northeast; e, rice-wheat rotation in East).

Fig. 3 Comparison of observed and simulated seasonal cumulative N₂O emissions for four models. Dotted lines represent the regression lines and solid lines represent 1:1 (y=x) line.

Fig. 4 Regression relationships between the observed and simulated growing season N₂O emissions from a range of nitrogen fertilizer application rates for DAYCENT, DNDC, and YLRM models (a, rice paddy; b, upland).