Tentative title: Molecular diversity informed modeling of litter decomposition

Arjun Chakrawal1, Stefano Manzoni2, Emily Graham1, Satish Karra1, Odeta Qafoku1, John Bargar1?

1Environmental Molecular Sciences Laboratory, Pacific Northwest National Laboratory, Richland, WA, USA

2Department of Physical Geography and Bolin Centre for Climate Research, Stockholm University, 10691 Stockholm, Sweden

**Abstract:**

**Keywords:**

chemodiversity, decomposition, 13C NMR, plant litter, degree of reduction, carbon use efficiency

# Introduction

Soil biogeochemical models are struggling to couple the dynamics of multiple elements (carbon–C, nitrogen-N, phosphorous-P) and of the chemical composition of organic matter as informed by emerging molecular-scale data (Kothawala et al., 2021). The application of high-resolution mass spectrometry techniques such as Nuclear Magnetic Resonance (NMR) and Fourier Transform Ion Cyclotron Resonance mass spectroscopy (FTICR-MS) is elucidating the chemical composition at molecular scale (?) of organic matter in soils (Boye et al., 2017; Dignac et al., 2002; Ding et al., 2020; Hall et al., 2020; Normand et al., 2021; PRESTON et al., 1987), plant litter (G. Bonanomi et al., 2013; Preston et al., 2000), and microbial biomass (Hedges et al., 2002; Knicker and Lüdemann, 1995). However, despite the insights gained from molecular observations, there remains a critical gap in ecosystem-scale models simulating biogeochemical exchanges informed by molecular-scale processes. For instance, while solid-state 13C NMR has been extensively utilized to explore the influence of plant litter quality on its degradation, the direct integration of NMR data into litter decomposition models remains limited. In this study, we present a novel approach to model litter decomposition model that is informed and constrained using solid-state 13C NMR data.

Capturing changes in chemical composition is key for predicting decomposition rates. In fact, the complex and heterogeneous structure of lignin in plant cell walls provides a protective barrier for high-energy unbranched carbohydrates, such as cellulose, and cross-linked polysaccharide chains, like hemicellulose, that are housed within plant cells. These lignin compounds shield carbohydrates as well as proteins from microbial decomposition. Chemical constraints on access to high-energy substrates have been modeled using a rate modifier that decreases the uptake rate of carbohydrates and proteins with increasing lignin content in plant litter. Certain specialized decomposer organisms, including white-rot fungi and Agaricomycetes, have oxidative enzymatic capabilities and thus can break down lignocellulosic bonds in plant cell walls, releasing polysaccharides for microbial growth and respiration (Alcalde, 2015; Mattila et al., 2022). However, cost of oxidative enzyme production requires significant resource investment by microbes that would otherwise be used for growth (Moorhead et al., 2013; Shimizu et al., 2005). The costs of producing and maintaining oxidative enzymes can be modeled as reduction in C use efficiency of microbial community (Manzoni et al., 2021; Moorhead et al., 2013). These competing processes generate a trade-off between microbial capacity to access resources and capacity to convert such resources into biomass.

To describe this trade-off and its consequences for soil carbon budgets, a detailed chemical characterization of litter is needed. Such a characterization would allow describing in the model the chemical compounds that are directly involved in the shielding effect, as well as the enzymatic reactions that break them down. Moorhead et al. (2013) found that the decay rate of holocellulose (cellulose and hemicellulose) decreases while that of lignin increases with the lignocellulose index (lignin/(lignin + holocellulose)) during litter decomposition. This led to development of a rate modifier as a function of lignocellulose index and has been applied in LIDEL (Campbell et al., 2016) and MEMS models (Robertson et al., 2019). Building on Moorhead et al. (2013), Manzoni et al. (2021) employed a power law function to incorporate the diminishing rates of carbohydrates and protein pools in their model. However, these and other decomposition models (e.g., Liski et al. 2005) rely on a coarse characterization of litter chemistry based on proximate analysis, which does not reflect litter chemical composition and lacks the resolution for mechanistic understanding. The acid unhydrolysable and hydrolysable fractions from proximate analysis of plant material have been used as proxies for lignin and carbohydrates, respectively, despite their potential inaccuracies in representing actual carbohydrate and lignin fractions (Preston and Trofymow, 2015). As a consequence, using proximate analysis for parameterization of the rate modifier function has posed significant challenges particularly due to measurement errors inherent in quantifying lignin and carbohydrates.

To overcome these parameterization challenges, some models are now turning to NMR data to gain mechanistic insights on the chemical constraints on decomposition. Recently, Chakrawal et al. (2024) utilized an exponential function as a rate modifier that was constrained using lignin fraction derived from 13C NMR spectra data. We are aware of only one other study by Incerti et al. (2017) that directly utilizes 13C NMR data to inform and constrain a litter decomposition model, termed OMDY (organic matter dynamics). The rate modifier in the OMDY model is referred as an intramolecular protection function that decreases the rate of simulated litter pools with increasing fractions of pools considered to have a protective effect. OMDY considered four litter pools in based on four molecular group identified from 13C NMR spectra chemical shift regions—alkyl C (0–45), methoxy and N-alkyl C (46–60), O-alkyl C (61–90) and di-O-alkyl C (91–110)—of which alkyl C, O-alkyl C and di-O-alkyl C were found to have protective effect.

As detailed chemical data become available, we can also ask whether such empirical rate modifier and associated cost of oxidative enzyme production are still required for model parameterization. To address these multiple challenges (matching modelled and measured quantities; describing the lignin shielding mechanism; modeling enzymatic reactions), there is a need to develop a litter decomposition model that can be informed using detailed molecular scale chemical composition together with mass and nutrient loss data.

In this contribution, we demonstrate the integration of molecular-scale chemical composition of plant litter, specifically utilizing solid-state 13C NMR, to constrain coupled C and N litter decomposition models. Specifically, we aim to assess the efficacy of 13C NMR data in calibrating four model scenarios simulating litter decomposition with and without the protection of effect of lignin on carbohydrates and proteins decomposition, and with trade-offs in microbial C use efficiency. Our specific research question are as follows,

1. Can we use 13C NMR data to constrain litter decomposition model parameters?
2. Does lignin rate modifier improve calibration of the model?
3. How do estimated parameters vary across four model scenarios constraints using same dataset?

# Methods

## Data collation and preparation

We searched published literature for litter incubation studies either in lab or field condition reporting litter total C and N mass loss and spectra from solid state 13C NMR over time. In total, we included 17 studies spanning 89 litter samples collected from warm and cold temperate and Mediterranean climate. The incubation length in studies varied between 6 months (Mediterranean sites (G. Bonanomi et al., 2013)) to 6 years (Canadian intersite decomposition experiment (CIDET) sites, (Preston et al., 2009)), accounting for total mass loss in the range of 20-98% of starting litter mass. Furthermore, the entire dataset covered wide variety of litter types such as broadleaves, needles, roots, wood, crop residues, and other grasses. The details of initial litter chemical composition, duration of field exposure or incubation length, mean annual temperature and precipitation, and initial fraction of organic compound classes obtained from NMR data are summarized in Error! Reference source not found..

Mass loss data for C and N were digitized from published studies or provided by authors. The 13C NMR data are usually reported as integrated values of seven chemical shift regions (alkyl, methoxy, o-alkyl, di-o-alky, aromatic, phenolic, carbonyl, see Figure 1A) and directly digitized from the tables of published studies We considered only those studies that reported NMR data of integrated chemical shift in a tabular format. The integrated values of chemical shift regions are a quantitative measure of various functional groups of organic C present in the litter sample (Figure 1A), and can be used to estimate molecular scale chemical composition of litter. For example, the sum of di-O-alkyl and O-alkyl is indicative of carbohydrates, while aromatic and phenolics are indicatives of lignin like compounds (Kögel-Knabner, 2002). The integrated values of seven chemical shift regions are often normalized with the total area under the spectra thus representing the fraction of C of each functional group in total C of litter. The integrated chemical shift data can be transformed into fraction of five distinct molecular classes of compounds comprising litter sample using a molecular mixing model (Nelson and Baldock, 2005). These classes are carbohydrates, proteins, lignins, lipids, and carbonyls. The molecular mixing model is described in detailed in the following section.

## Molecular mixing model

Assuming plant litter sample is consisting of carbohydrates, proteins, lignins, lipids, and carbonyls organic compounds, molecular mixing model estimates fractions of each class such that the observed solid state 13C NMR spectra of whole litter sample is a weighted sum of the spectra of the organic compounds (Figure 1A). The elemental composition of these organic compounds is fixed (see Table X). Thus, the input data required for molecular mixing model is the observed integrated values of chemical shift regions of litter sample () and the organic compounds (). Let us denote the C fractions of molecular classes as in the units of gC per gC of litter, where is carbohydrates, proteins, lignins, lipids, or carbonyls. Mathematically, the molecular mixing model can be written as,

|  |  |  |
| --- | --- | --- |
|  |  | () |

where is a matrix of size [] whose rows are the NMR spectra of seven chemical shift regions of five organic compounds, is column vector of size [] containing fraction of five organic compounds, and is a column vector of size [] containing observed integrated values of seven chemical shift regions of litter sample. Equation (1) is solved for using total C and N concentration of litter as constraints to conserve mass balance. If CN ratio (gN /gC) of plant litter was reported, then protein fraction is constrained using N mass balance i.e., where is the CN ratio of proteins—indicating total N content of litter is in proteins. For more details on molecular mixing model see Chakrawal et al 2024 (in preparation). The molecular mixing model is used at each time point to convert NMR chemical shift data to C fraction of carbohydrates, proteins, lignin, lipids, and carbonyls which can be converted into mass unit (gC) multiplying by the total litter C (gC). We used scipy.optimize.minimize function with mass balance constraints to solve for . If the CN ratio of litter sample was not reported then the N mass balance constraint was dropped and was estimated simultaneously with other fraction using optimization. Toassess molecular mixing model performance, we calculated Nash–Sutcliffe modeling efficiency coefficient (NSE) and root mean squared (RMSE) values as goodness of fit metrices between observed NMR of litter and estimated NMR calculated using estimated (Figure 2B).

## Litter decomposition model

We developed a litter decomposition model for simulating the dynamics of C and nitrogen (N) within five distinct pools representing organic compounds as identified from the molecular mixing model using NMR data (Figure 1B). We assumed first order kinetics for the uptake rate () of each pool, which is reasonable when using coarse resolution data spanning months-to-years. The protection effect of lignin on carbohydrates and proteins is implemented using a sigmoidal function () that decreases with increasing fraction of lignin similar to Chakrawal et al. (2024). Further, we assumed that assimilated substrates can be used for growth and maintenance at a maximum C use efficiency (). The maximum CUE decreases with increasing decay constant for the lignin fraction in the case of model accounting for the investment into oxidative enzyme production (Manzoni et al., 2021).

Under N-limited conditions, we consider two alternative strategies: i) microorganisms preferentially retain N when they senesce (Manzoni et al., 2021) or ii) microorganisms regulate their CUE resulting in overflow respiration, resulting in decreased CUE (Schimel and Weintraub, 2003). Furthermore, we assume that microorganisms grow in a quasi-steady state condition, meaning their growth rate equals mortality rate. The necromass is recycled into various organic compound classes according to its composition, i.e., a fixed fractions of necromass is distributed across pools under C limited conditions, or with variable fractions under N limited conditions when strategy i) is adopted.

Based on these, we can write the mass balance equation for each organic compound as follows,

|  |  |  |
| --- | --- | --- |
|  |  | () |

where, is the mass of C in each pool, = or for carbohydrates, proteins, lignin, lipids, and carbonyls, respectively. The is the rate modifier affecting the rate constant of different pools. For carbohydrates and proteins, (with ), capturing the protection effect of lignin; for lignin capturing the effect of delayed decomposition of lignin in lignin-poor litter; for lipids and carbonyls (with ) indicating time invariant rate constants. The coefficients are the fractions of microbial mortality rate , recycling into respective substrate pools (). The uptake rate of each pool is prescribed using first-order kinetics with as rate constant, and written as follows,

|  |  |  |
| --- | --- | --- |
|  |  | (3) |

Next, the mass balance for microbial C () is written as follows,

|  |  |  |
| --- | --- | --- |
|  |  | () |

where, the first term on the right-hand side is the microbial growth rate , with as the C use efficiency (ratio of growth over total uptake rate) under C limited conditions. The cost of oxidative enzyme production is modeled as linear decrease in CUE with the lignin decay rate constant ( as follows,

|  |  |  |
| --- | --- | --- |
|  |  | (5) |

where is the cost factor. Based on Manzoni et al. (2021), we assume an inverse relation between cost factor and oxidative capacity, i.e., higher cost for low oxidative capacity, and formulated it as . Substituting, in eq (5), we obtain CUE as,

|  |  |  |
| --- | --- | --- |
|  |  | (6) |

Following Chakrawal et al. (2024), we formulated the rate modifier as decreasing function of lignin fraction, and is given as , where is the fraction of lignin C and is the scaling coefficient. For brevity we refer to this rate modifier as -function. Further, we adopted a quasi-steady state assumption for the microbial biomass (i.e., , so that .

Assuming that necromass recycling into the protein pool has the same CN ratio of the protein pool, we can write the N mass balance for protein pool,

|  |  |  |
| --- | --- | --- |
|  |  | () |

Note that eq (7) is redundant as it can be written as Next, the mass balance for microbial N () is written as follows (recall that N is only coming from protein pool),

|  |  |  |
| --- | --- | --- |
|  |  | () |

where, is the CN ratio of microbes and is the net N exchange rate from inorganic pool. The coefficient is the N retention factor. Imposing the homeostatic condition for microbial growth, i.e., , we calculate as follow,

|  |  |  |
| --- | --- | --- |
|  |  | () |

Replacing with from the quasi-steady state assumption for microbial growth, we obtain as

|  |  |  |
| --- | --- | --- |
|  |  | () |

The first term on the right-hand side is supply of N from protein pool and the second term is the N demand for microbial growth. If net N mineralization occurs and if then net N immobilization from inorganic N pool occurs. We define N limited microbial growth when net immobilization rate is higher than supply of N from inorganic pool, i.e., . Thus, under N limited condition, N uptake from the inorganic pool is constrained by the supply rate of inorganic N, i.e., .

We assume that microorganisms selectively retain N on turnover under N limited conditions by reducing . Following, Manzoni et al (2021), we use as N retention factor that reduces . Imposing the constraint , we can calculate the value of , as follows,

|  |  |  |
| --- | --- | --- |
|  |  | () |

The N retention factor varies between minimum under C limited condition to a theoretical maximum . The rate of necromass recycling into the protein N pool (i.e., ) must be the same as the rate of N loss in the form of necromass (i.e., ). From this equality we can estimate ,

|  |  |  |
| --- | --- | --- |
|  |  | () |

As and in C limited conditions , we can see from eq (10) that <1. This means that most of the C from necromass will be recycled in the C-only pools, while all the N is recycled in the protein pool. The decrease in under N limited conditions is associated with increased microbial turnover in carbohydrate pool i.e., .

As an alternate microbial adaptation strategy (referred to as flexible CUE) to N limited conditions, we assumed that microbes may regulate their maximum CUE to reduce N demand which will effectively lead to increased respiration. In this alternative strategy, we consider no preferential N retention, thus, . As above, imposing the constraint , we can calculate a under N limited conditions as follows,

|  |  |  |
| --- | --- | --- |
|  |  | () |

## Model parameterization, implementation and least-square model-data fitting

The final litter decomposition model only consisted of five mas balance equations (Eq (2)) for five organic compounds that needs to be solved. These systems of ordinary differential equations were solved using an iterative solver. The model has five rate constants , five mortality fractions , inorganic N supply rate, , oxidative enzyme cost factor the scaling coefficient , , and initial conditions of five pools as unknown parameters.

The maximum CUE, is constrained using oxidation state of litter C (Chakrawal et al., 2022) and varies dynamically; however, under N limited condition and when considering flexible CUE strategy, maximum CUE is computed using Eq (13). The oxidation state of litter is estimated as weighted sum of oxidation state of each organic compound (see Table (S2)). Following, Manzoni et al. (2021), we parametrizes the cost factor assuming higher cost for related to low oxidative capacity, thus, . Furthermore, Chakrawal et al. (2024) also found an inverse relationship between the cost factor and lignin rate constant (see supplementary Figure S3 from Chakrawal et al. (2024)).

The C:N ratio of microbes is assumed to be 16 (Zhang and Elser, 2017) and the C:N of proteins 3.2. The fraction of necromass recycling into the protein pool using eq (12), yields = 0.32, and other fractions recycling into lignin and lipid, carbonyl pools were estimated using the composition of fungal necromass from Beidler et al. (2020) as ,, and . Finally, the necromass fraction recycling into the carbohydrate pool yields . The scaling coefficient was determined by adapting the value of , previously established for aromatic carbon, to lignin carbon (Chakrawal et al., 2024). By scaling aromatic C to lignin C in the -function (approximating, 55% of lignin is aromatic C, see supplementary Fig XX), we estimated .

The initial conditions for the organic compounds were directly set from observed data using initial fractions and initial mass of litter sample. The inorganic N supply rate, was estimated as maximum rate of N accumulation for each litter incubation. The remaining five rate constants were estimated as best-fitted parameter by least-square method using observed time series data of carbohydrates, proteins, lignins, lipids, and carbonyls pools obtained from the molecular mixing model by processing NMR data. We used scipy.integrate.solve\_ivp with RK45 ode solver to solve the system of differential equations, and scipy.optimize.least\_squares for fitting the model to data. The least-square solver minimized the mean squared error was computed using augmented observation vector which containing all five organic compounds and a corresponding vector of model simulated values at observation time points. For faster convergence of least-square solver, we normalized the simulated and observed values by the maximum observed mass remaining in the respective organic compound class.

Due to time consuming and expansive NMR measurements, not all studies reported total C and N loss, and NMR data at the same time resolution. For instance, if there was a greater number of data points in total N than proteins estimated from NMR then total N was also included in the observation vector during model calibration. Moreover, we assumed 50% C content of litter when not reported.

We calculated Nash–Sutcliffe modeling efficiency coefficient (NSE) and root mean squared error (RMSE) as model performance metrices (Janssen and Heuberger, 1995). NSE values varies between negative infinity to 1, where NSE close to 1 reflect a good match between model simulation and observation. Negative and NSE<1 represent model cannot capture observations or in other words, the model parameters cannot be reliably obtained from the observations.

## Four model scenarios

We defined four model scenarios to explore the interaction between lignin protection of carbohydrate and protein pools and its impact on carbon use efficiency (CUE) resulting from investments in oxidative enzymes.

The first model scenario, "NPNE : no protection no oxidative enzyme cost" posits that the degradation rates of carbohydrates and proteins remain unaffected by lignin content, while microorganisms operate at maximum CUE without incurring additional costs from enzymatic activities. This is achieved by setting the value of and equal to 1.

The second model scenario, the “NPWE: no protection with oxidative enzyme cost” model maintains unaltered decomposition rates for carbohydrates and proteins despite the presence of lignin, but assumes an investment towards oxidative enzyme production, leading to a proportional reduction in CUE. This scenario is motivated from the production of ligninolytic enzymes for uptake of lignin like compound for growth and maintenance (del Cerro et al., 2021). In this scenario of the model, and are set to 1; however, in eq (6) is not set to 1 but allowed to vary as a function of fraction of lignin.

The third model scenario " PWOE: protection with oxidative enzyme cost (time invariant )" integrates both lignin's protective effects and the accompanying costs of oxidative enzymatic activity. In this scenario of the model no constraints are imposed on , but is set to 1.

In the first three model scenarios we assumed that the rate constant of lignin is assumed invariant, as . We relax this assumption in the fourth model scenario, "PWOV: protection with oxidative enzyme cost (time varying )", similar to the third model, but with time varying lignin rate constant by setting .

Should we do a statistical test to see if we can find negative correlation among carbohydrate, protein rates with lignin fraction using estimated parameters from different model fits. Idea would be to test if estimated parameters account for the protection effect lignin when model is explicitly accounting for such effects vs when model is not.

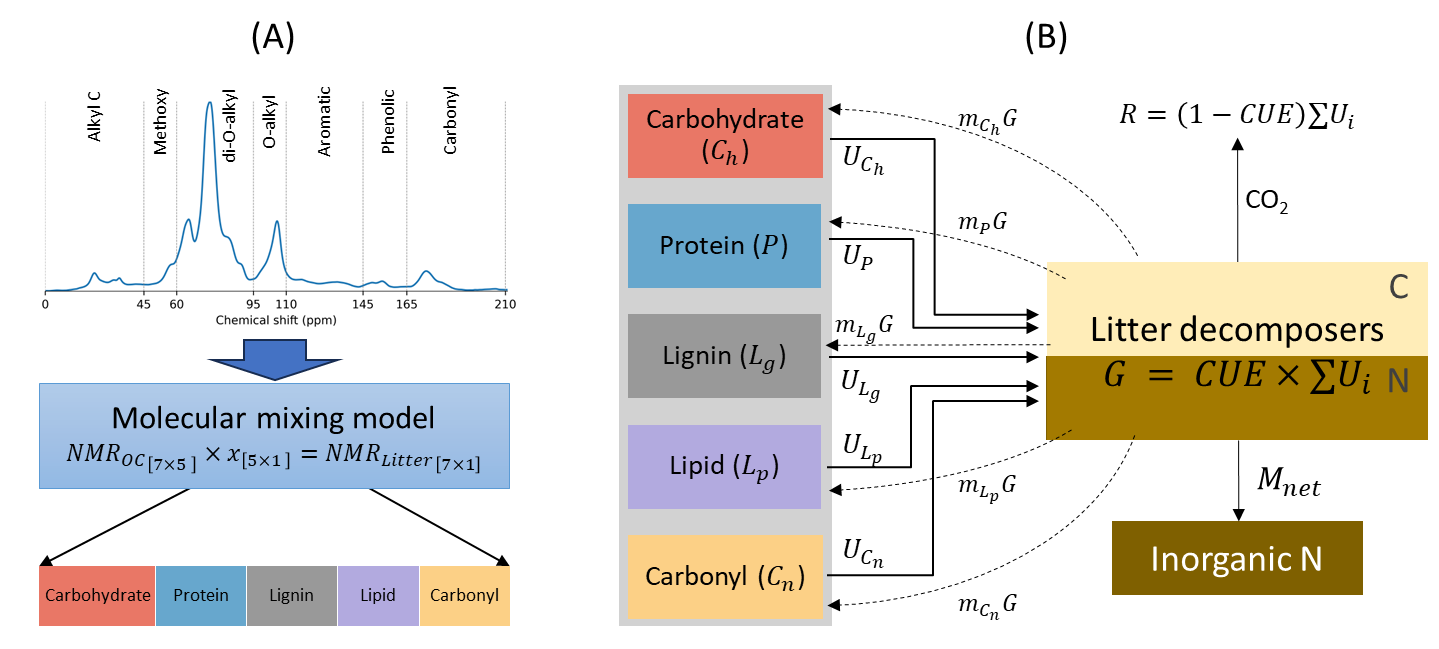


Figure 1 (A) Observation model schematic illustrating the preprocessing of solid-state 13C NMR spectra through a molecular mixing model to derive fractions of five major components of litter, namely, carbohydrate, protein, lignin, lipid and carbonyl. (B) Litter decomposition model schematic wherein microbes decompose various litter components with an overall carbon use efficiency (CUE). Solid and dashed lines denote substrate uptake rates () and microbial mortality rates () of respective pools, where = ( or ), is the microbial growth rate, is the fraction of necromass recycled into pools, and in denote carbon in mass units (gC). is the net N exchange rate from inorganic N pool, i.e., the difference between rates of N supply and microbial N demand.

# Results

We focus mainly on Nretention strategy and CUEregulation is used for model performance comparison

MMM result- violin plot: Xu et al not include NSE<0

Pool dynamic across models

Example fit

Model performance: scatter plot and NSE/RMSE figure

Comparison of estimated parameters from other studies

Variation in mean CUE with bulk chemistry with and without CUE regulation

Chart

Description automatically generated

Figure 2. (A)The range of variation of the fraction of organic compounds in litter estimated using molecular mixing model and 13C NMR data. (B) Nash–Sutcliffe modeling efficiency coefficient (NSE) and (C) root mean square error (RMSE in gC/gC litter) from the predicted NMR chemical shift using molecular mixing model and observed NMR chemical shift for the litter samples. The horizontal lines in the violin plots indicate quartiles: the median (middle line), upper quartile (Q3) and lower quartile (Q1). The box in boxplot shows the interquartile range (IQR), with the median marked by a horizontal line. Whiskers extend to values within 1.5 times the IQR from the quartiles and outliers beyond the whiskers are indicated by individual markers.

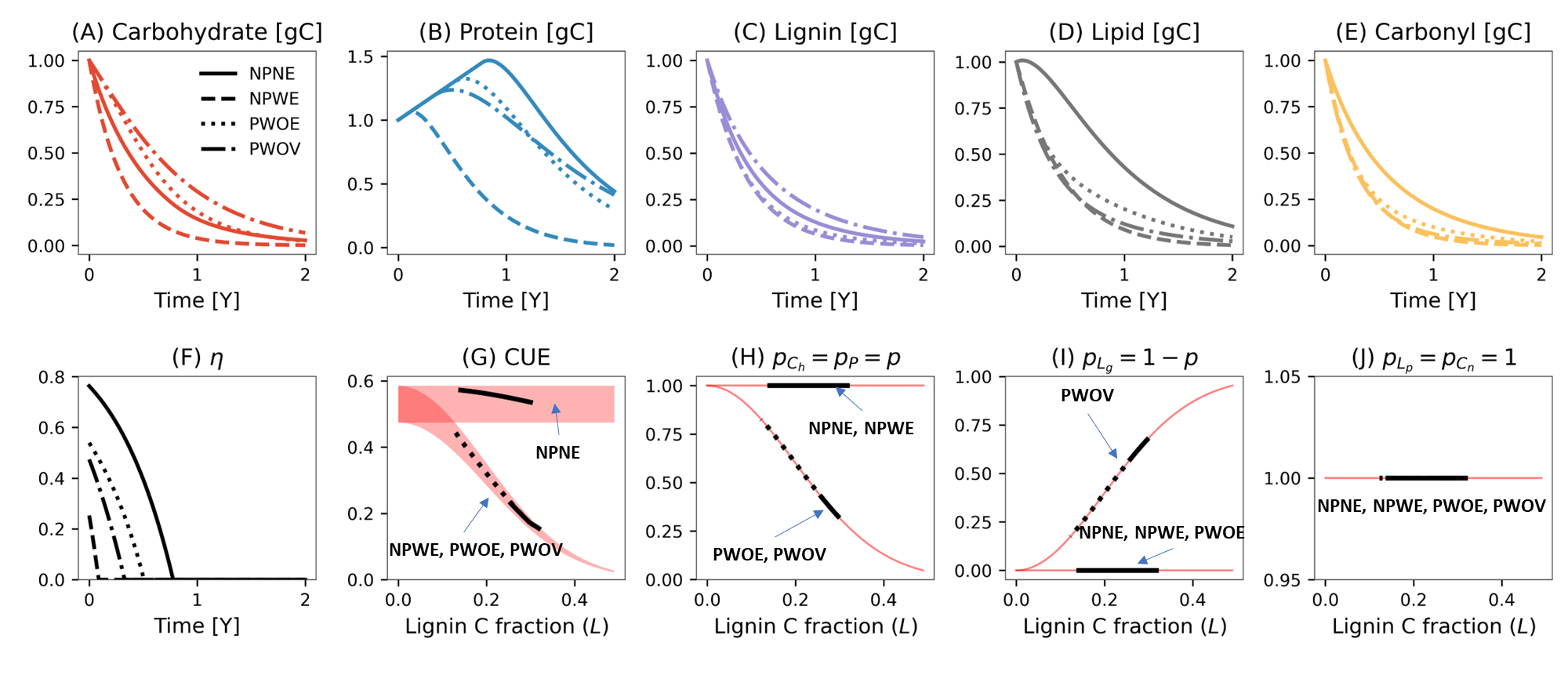


Figure 3 Simulated temporal variation in carbohydrate (A), protein (B), lignin (C) lipid (D), and carbonyl (E) pools, and N retention factor (F) in four model scenarios using N-retention strategy. Different model scenarios are illustrated using different line styles. The bottom panels (G-J) show the variation with lignin fraction of C use efficiency (CUE) (G), rate modifier for carbohydrate () and protein () (H), lignin rate constant ( (I), and rate modifiers for lipid () and carbonyl pools () (J). In panels G–J, the red areas or lines extend the range of variation of rate modifiers with lignin fraction, and black lines are the simulated range of p vs. L in each model scenario. In the panel G, the red area illustrates the plausible range of variation in CUE under different model scenarios—calculated based on a minimum and maximum degree of reduction of whole litter, set at 3.8 and 4.5, respectively, to determine . Model legends are as follows, **NPNE**: **n**o **p**rotection of carbohydrate and protein, and **n**o oxidative **e**nzyme cost, **NPWE**: **n**o **p**rotection carbohydrate and protein but **w**ith oxidative **e**nzyme cost, **PWOE**: **p**rotection of carbohydrate and protein **w**ith **o**xidative **e**nzyme cost but time invariant lignin rate constant, **PWOV**: **p**rotection of carbohydrate and protein **w**ith **o**xidative enzyme cost but time **v**arying lignin rate constant. Model parameters and initial conditions used in simulation were [0.01, 0.01, 0.008, 0.009, 0.01] d-1, and =1e-5 gN d-1; initial fraction of carbohydrate, protein, lignin, lipid, and carbonyl pools were 0.43,0.02,0.3,0.2,0.05, respectively; and the initial mass of litter was considered to be 1g on dry weight basis.

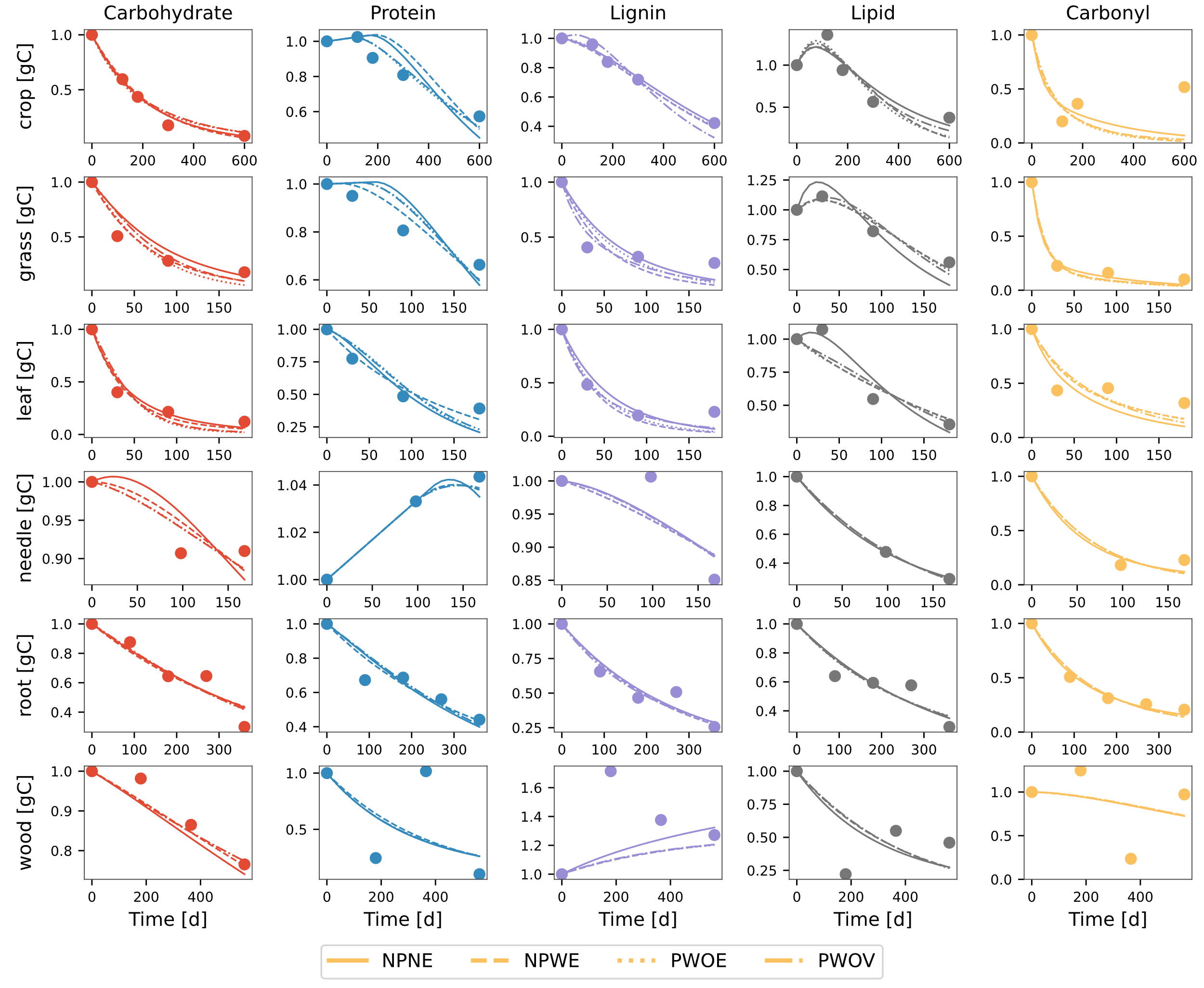


Figure 4 Comparison of modeled (different line styles) and observed (circles) changes in five litter pools (carbohydrate, protein, lignin, lipid, and carbonyl) for various litter types. The data points for the crop litter is from wheat straw buried at 15cm depth in sandy loam soil from Li et al. (2020); grass and leaf litter samples are of A. mauritanicus and A. unedo species, respectively, from Bonanomi et al. (2013); needle litter is from P. radiata from Almendros et al. (2000); root litter samples are fine roots of M. macclurei from Wang et al. (2013); and wood litter of Mulga twigs from Mathers et al. (2007). Different line styles are for four model variants. Model legends NPNE, NPNW, PWOE and PWOV are the same as in Figure 3**.**

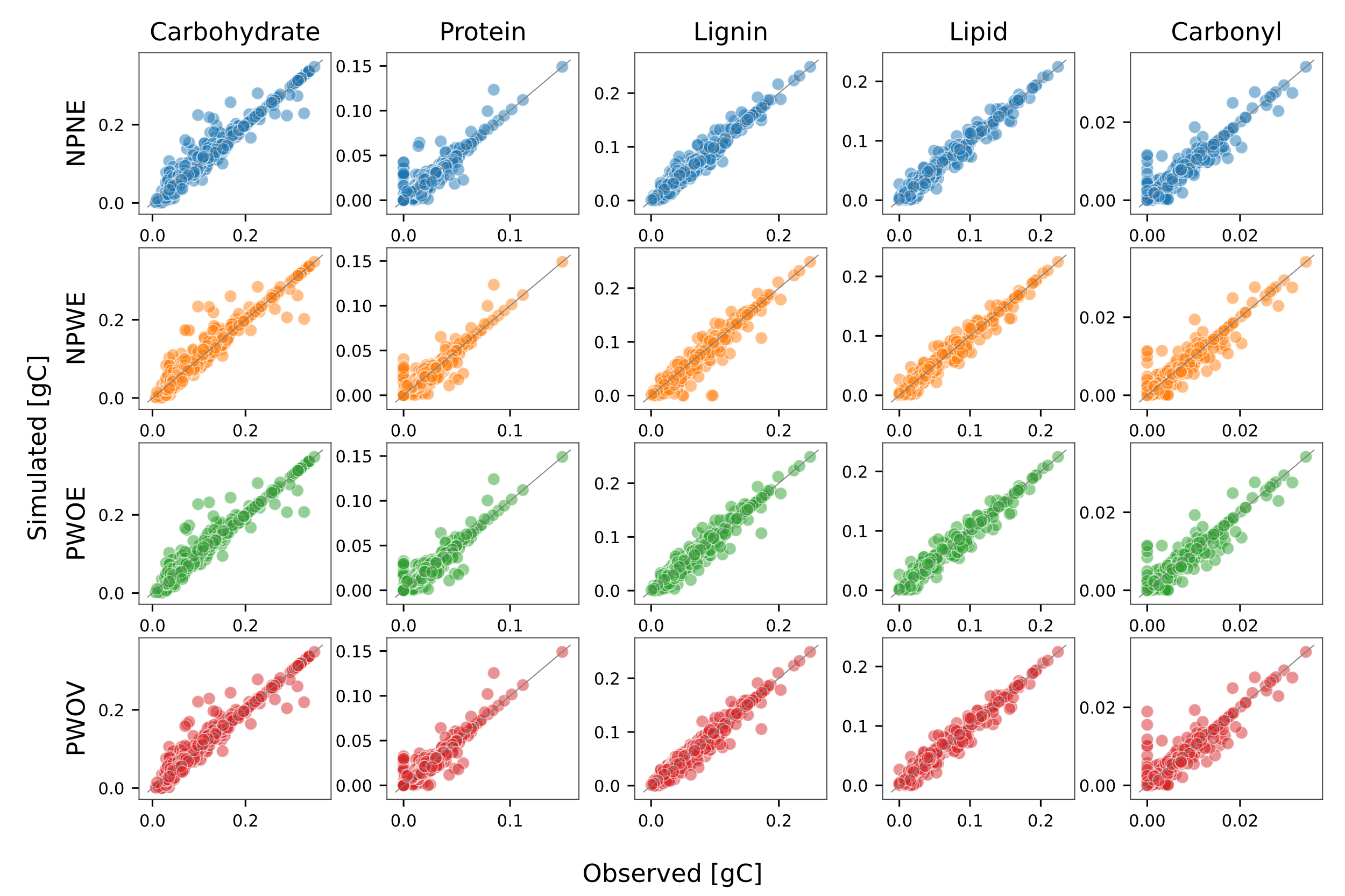


Figure 5 (A) Scatter plot of modeled and observed mass remaining of carbohydrate, protein, lignin, lipid and carbonyl among four model scenarios illustrate using different colors. The grey line represents 1:1 line. Model legends are the same as in Figure 3**.**

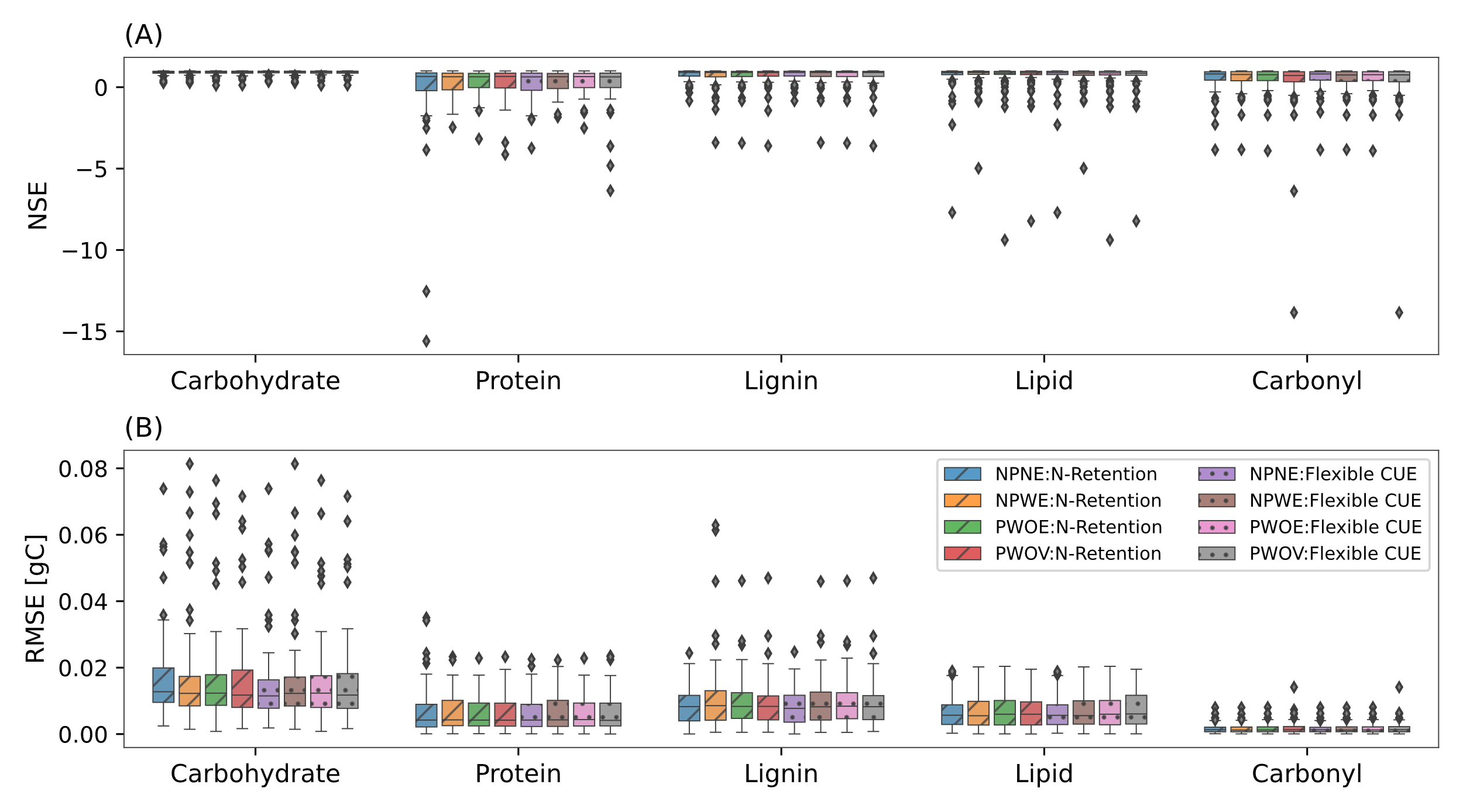
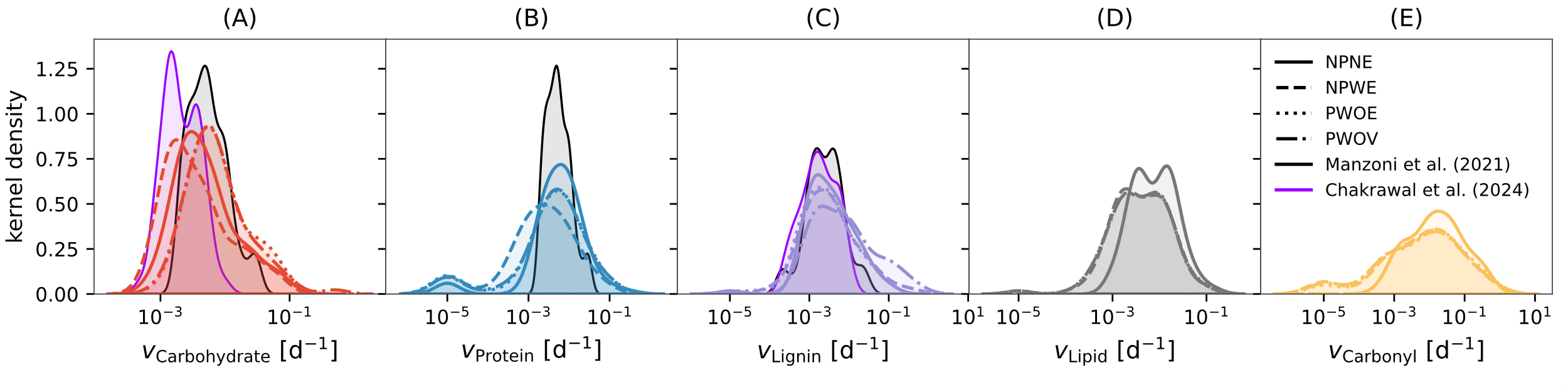


Figure 6 Boxplot of Nash–Sutcliffe modeling efficiency coefficient (NSE) (A) and root mean square error (RMSE in gC) (B) for five pools for four model scenarios. Diagonal lines and dots within boxes denote NSE and RMSE for two N adaptation strategies, N-retention and Flexible CUE, respectively. The box shows the interquartile range (IQR), with the median marked by a horizontal line. Whiskers extend to values within 1.5 times the IQR from the quartiles and outliers beyond the whiskers are indicated by individual markers.



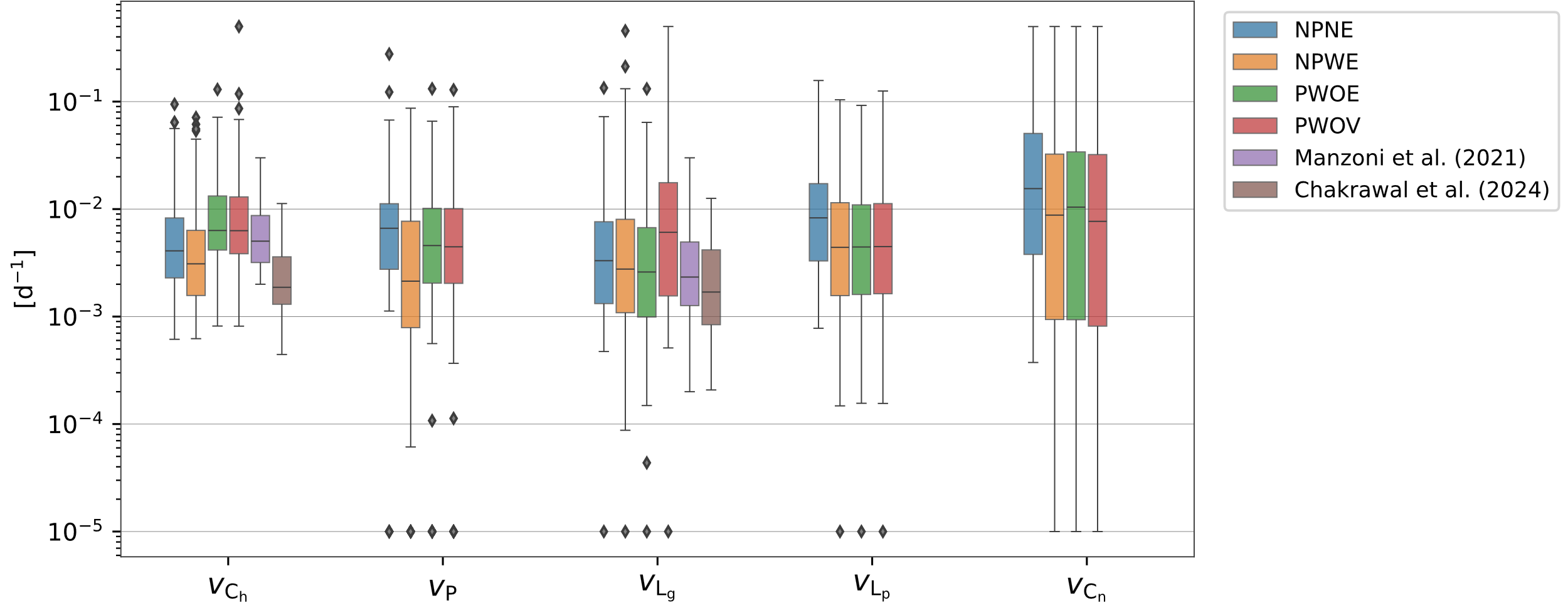
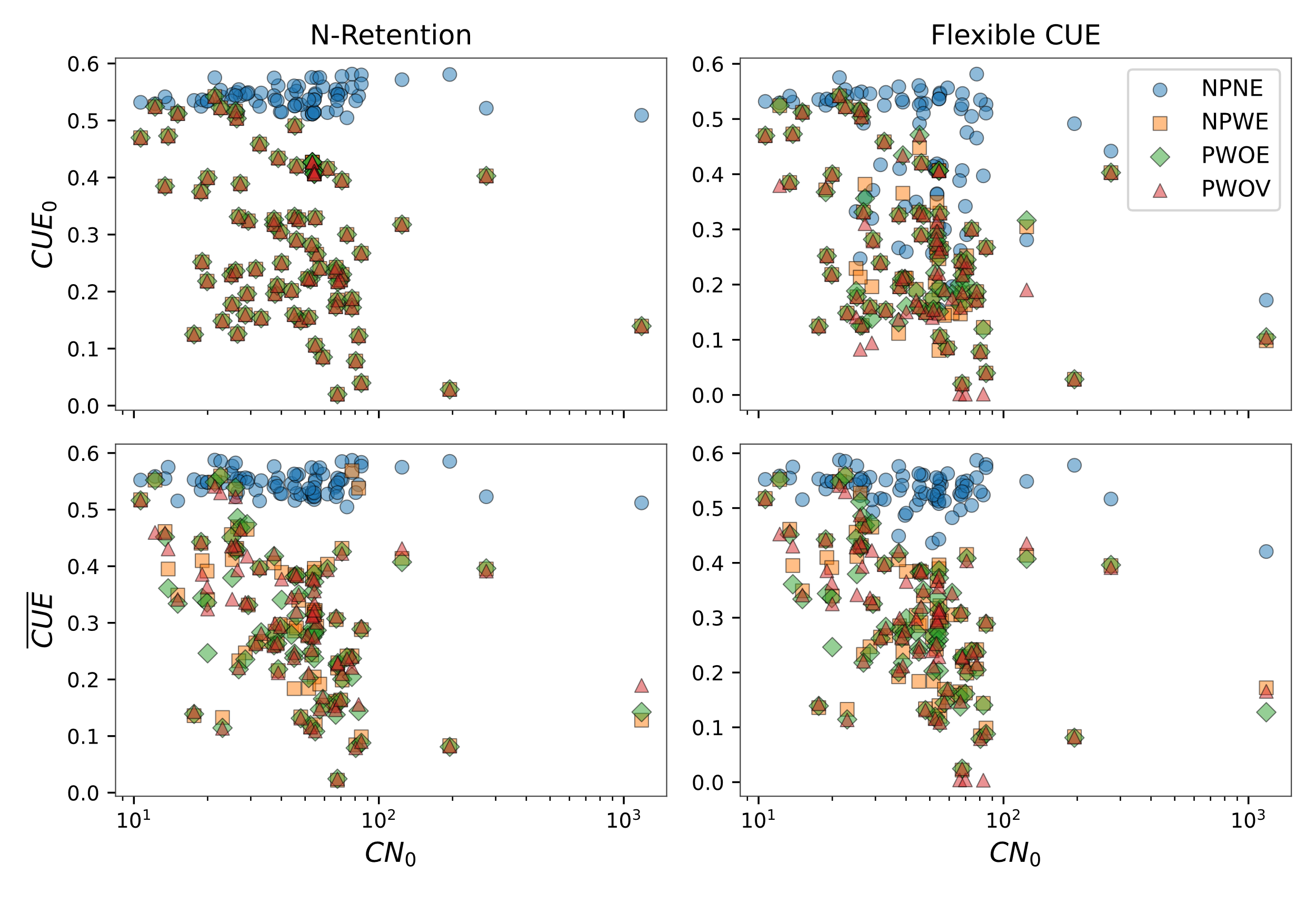


Figure 7 Distribution plot of estimated model parameters, i.e., the uptake rate constant for (A) carbohydrate, (B) protein, (C) lignin, (D) lipid, and (E) carbonyl for three model scenarios as different line styles. Note the log transformed value on X-axis for each variable. Distribution plots illustrated by black and purple lines in panels A, B and C are the rates of carbohydrates, proteins, and oxidizable (representative of lignin) pools, respectively, taken from Manzoni et al. (2021), and nonaromatic (representative of carbohydrates and proteins) and aromatic (representing lignin and other aromatic compounds) pools from Chakrawal et al. (2024). Model legends are the same as in Figure 3**.**



Chart, box and whisker chart

Description automatically generated

Figure 8 Variation in initial and temporal average carbon use efficiency (A) and degree of reduction (B) for three model scenarios.

# Discussion:

chemodiversity

How did we answer the research questions:

1. Can we use NMR data to constrain litter decomposition model parameters?

Coupled CN litter decomposition models have not been parametrized using NMR data. As microbes decompose litter, easily degradable food is lost faster, and litter is left with higher proportion of lignin like compounds which require costly oxidative enzyme that microbes can’t afford because high energy food is not available to compensate for anymore. This can be thought of as reducing return of framework.

1. Does lignin rate modifier improve calibration of the model?

If not, then why worry so much about parametrizing it? There are papers arguing that lignin does not decreases carbohydrate decomposition. Lignin is not the bottleneck. (but it depends) Yes, there are resources invested into oxidative enzyme but the presence of lignin per se may not exert decomposition limitation on carbohydrates. And if there are not enough labile high energy substrate to fuel those resource investments then microbial growth will be limited which will reduce the decomposition of all pool not just lignin. This is same as the priming effect in soils. Conceptually the rate modifier is same as return-on-investment principle.

Conceptually, the rate modifiers are similar to return-on-investment function expressed as the total substrate uptake capacity per unit cost of enzyme investment (Chakrawal et al., 2024; Wutzler et al., 2017). For example, in soil ecosystems, the energetic return on investment diminishes with depth as the energy content of organic matter decreases, while the activation energy of decomposition, in the form of exoenzymes, increases with depth. Use (Rovira et al., 2008) for litter energetics. Link to priming?

1. How does estimated parameters vary across three different models when constraints using same dataset?

Bring the issue of equifinality in model, models may fit the data equally well but perhaps for very different reasons.

## Model performance with NMR data vs proximate

Model performance, discuss OMDY

## Parametrizing protection effect of lignin/lipids on sugars and proteins

Does rates of sugars and protein decreases with lignin, lipid content?

## Variation in overall NOSC of litter during its degradation

Ideas from Gunina and Kuzyakov 2021, that as microbes decompose litter their nosc decreases

# Similarities of model developed here to be used in soils

With lipids being explicitly modeled, it will be easier to track storage part of microbial biomass?

# Conclusions

Accurately implementing such microbial controls on decomposition processes is paramount for improving partitioning of litter into soil organic matter. NMR data provide detailed and quantitative information on litter chemical composition opposed to traditional macrochemical indices use to describe litter quality such as carbon to nitrogen, nitrogen to lignin ratios.

# References

Table 1 Description of

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Study | Litter types | Climate | Incubation length  (day) | MAT  (oC) | MAP (mm) | C  (gC/g d.w. litter) | N\_conc  (gN/g d.w. litter) | C:N  (gC/gN) | Carbohydrate  (gC/g d.w. litter) | Protein  (gC/gC d.w. litter) | Lignin  (gC/gC d.w. litter) | Lipid  (gC/gC d.w. litter) | Carbonyl  (gC/gC d.w. litter) |
| Almendros et al. 2000 | needle, leaf | Mediterranean | 168 | 14.500 | 415 | 410 - 544 | 7 - 28 | 17.607 - 77.714 | 0.293 - 0.624 | 0.041 - 0.180 | 0.068 - 0.338 | 0.069 - 0.393 | 0 - 0.072 |
| Bonanomi et al 2013 | grass, leaf, needle | Mediterranean | 180 | 16.500 | 1080 | 354 - 487 | 11.297 - 35.290 | 12.200 - 40.100 | 0.357 - 0.600 | 0.079 - 0.260 | 0.028 - 0.332 | 0.059 - 0.298 | 0.003 - 0.052 |
| Certini et al 2023 | grass | Mediterranean | 720 | 9.200 | 1273 | 500 | 11.043 - 46.946 | 10.651 - 45.278 | 0.403 - 0.619 | 0.070 - 0.298 | 0.099 - 0.190 | 0.106 - 0.179 | 0.015 - 0.021 |
| De Marco et al. 2021 | leaf | Mediterranean | 403 - 810 | 16 | 756 | 487.830 - 615.090 | 4.300 - 7.250 | 70.700 - 124.310 | 0.337 - 0.430 | 0.026 - 0.045 | 0.173 - 0.246 | 0.324 - 0.365 | 0.003 - 0.014 |
| Gao et al 2016 | crop residue | Warm Temperate | 360 | 15.500 | 985 | 479.800 | 8.800 | 54.523 | 0.652 | 0.058 | 0.148 | 0.142 | 0 |
| Li et al 2020 | crop residue | Warm Temperate | 600 | 13.900 | 597 | 450.100 - 469.100 | 8.410 - 8.610 | 53.520 - 54.483 | 0.681 - 0.747 | 0.058 - 0.059 | 0.119 - 0.132 | 0.057 - 0.098 | 0.017 - 0.031 |
| Mathers et al., 2007 | crop residue, leaf, grass, wood, roots | Warm Temperate | 562 | 21 | 516 | 409.300 - 500 | 1.680 - 29.755 | 15.100 - 274.600 | 0.510 - 0.752 | 0.012 - 0.210 | 0.015 - 0.202 | 0.035 - 0.266 | 0.009 - 0.039 |
| McKee et al., 2016 | grass | Warm Temperate | 1096 | 12.800 | 835 | 500 |  |  | 0.696 | 0.061 | 0.042 | 0.171 | 0.030 |
| Ono et al 2009 | leaf | Warm Temperate | 1095 | 21 | 516 | 516.200 - 522.800 | 11.197 - 11.592 | 45.100 - 46.100 | 0.394 - 0.508 | 0.069 - 0.070 | 0.225 - 0.314 | 0.199 - 0.222 |  |
| Ono et al 2011 | leaf | Warm Temperate | 1095 | 21 | 516 | 592.700 - 609.400 | 10.617 - 11.141 | 53.200 - 57.400 | 0.373 - 0.387 | 0.055 - 0.060 | 0.237 - 0.261 | 0.310 - 0.317 |  |
| Ono et al 2013 | leaf | Warm Temperate | 1095 | 21 | 516 | 500 |  |  | 0.347 - 0.513 | 0 - 0.074 | 0.196 - 0.347 | 0.224 - 0.340 | 0 - 0.002 |
| Pastorelli et al 2021 | leaf | Mediterranean | 365 | 9.200 | 1273 | 500 | 10.417 | 48 | 0.465 | 0.066 | 0.315 | 0.122 | 0.033 |
| Preston et al. 2009 | needle, leaf, grass, wood | Cold Temperate | 2190 | 6.700 | 978 | 438 - 497 | 0.400 - 12.800 | 38.828 - 1182.500 | 0.374 - 0.698 | 0.003 - 0.082 | 0.129 - 0.342 | 0 - 0.272 | 0.002 - 0.039 |
| Quideau et al 2005 | leaf | Warm Temperate | 733 | 14.400 | 678 | 467 - 499 | 6.970 - 8.836 | 55 - 68 | 0.356 - 0.461 | 0.047 - 0.058 | 0.199 - 0.269 | 0.175 - 0.290 | 0.052 - 0.073 |
| Sjöberg et al 2004 | needle | Cold Temperate | 559 | 7.600 | 1440 | 500 | 9.091 - 21.739 | 23 - 55 | 0.412 - 0.431 | 0.058 - 0.138 | 0.318 - 0.359 | 0.129 - 0.145 | 0.003 - 0.007 |
| Wang et al 2013 | leaf, roots | Warm Temperate | 360 | 21 |  | 392.600 - 574.200 | 6 - 15.700 | 33.083 - 80.483 | 0.258 - 0.443 | 0.039 - 0.096 | 0.206 - 0.508 | 0.103 - 0.390 | 0.020 - 0.054 |
| Wang et al 2019 | leaf | Warm Temperate | 450 | 19.880 | 1000 | 509 - 512 | 2.629 - 6 | 84.833 - 194.729 | 0.188 - 0.219 | 0.016 - 0.037 | 0.456 - 0.486 | 0.259 - 0.299 | 0.011 - 0.028 |

# Supplementary information

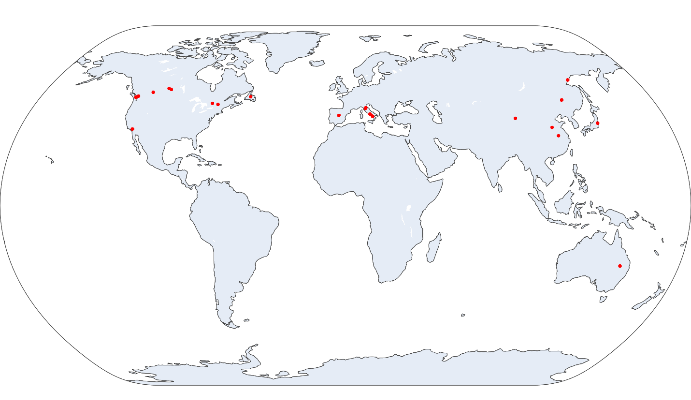


Figure 9: Geolocations of litter bag incubation sites

Add figure on performance of molecular mixing model

Scatter plot of response and predictors

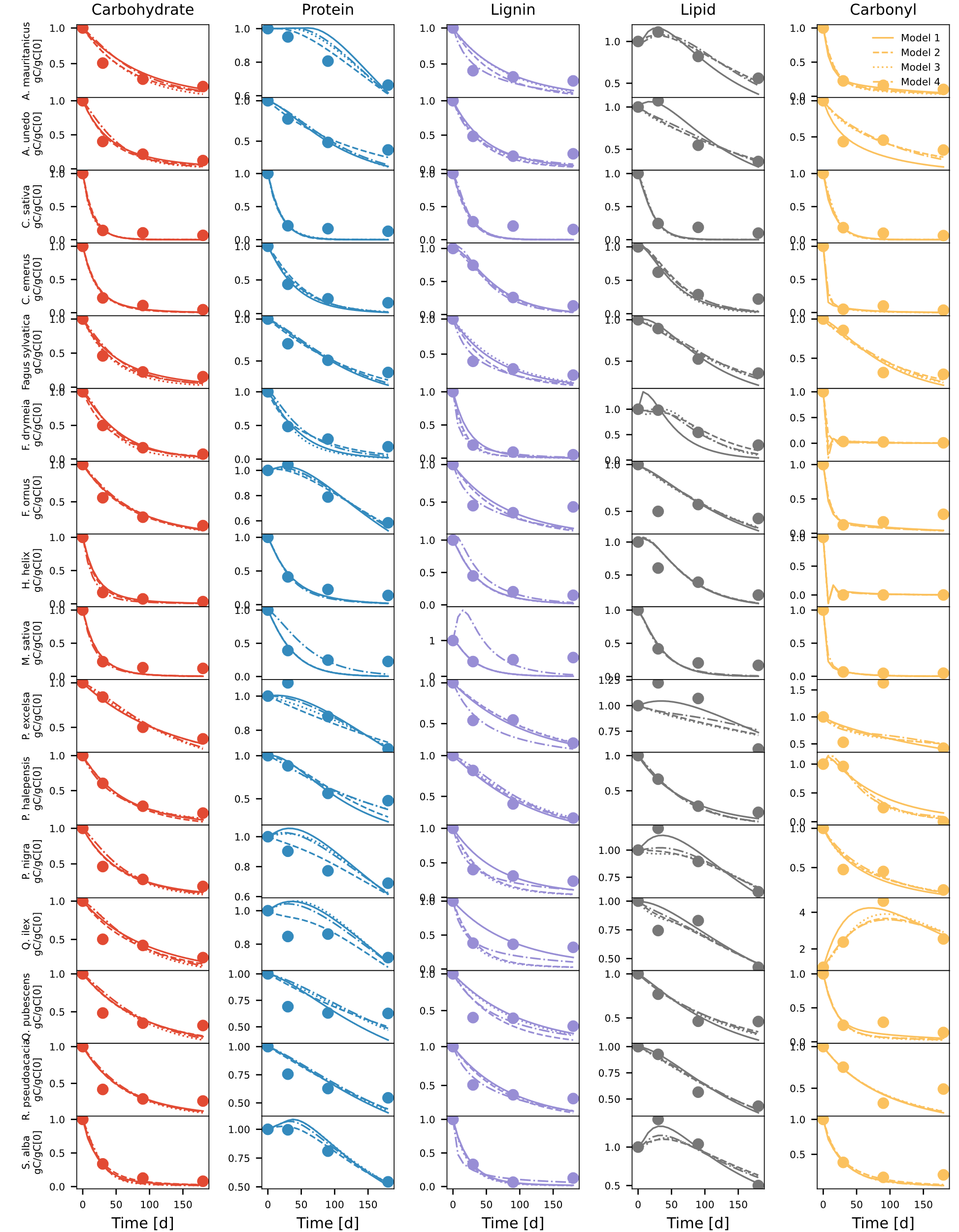


Figure 10 Comparison of modeled (different line styles) and observed (circles) changes in five litter pools (carbohydrate, protein, lignin, lipid, and carbonyl) for various litter samples compiled from Bonanomi et al. (2011; 2013). Different line styles are for four model variants. Model legends are the same as in Figure 3**.**

Chart, scatter chart

Description automatically generated

Figure 11 collinearity\_Residual\_Bonanomi et al 2011\_P. halepensis (rate constants identifiable)

Chart, scatter chart

Description automatically generated

Figure 12 collinearity\_Residual\_Preston et al. 2009\_Douglas-fir (rate constants and mLp identifiable)

Chart, scatter chart

Description automatically generated

Figure 13 Relationship between fraction of lignin C obtained from molecular mixing model and aromatic C as the sum of aromatic (110-145 ppm) and phenolic (145-165 ppm) regions across all litter samples.

Table S1 Chemical shift regions of five classes of organic compounds taken from Nelson and Baldock (2005)

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Chemical shift region (ppm) | Carbohydrate | Protein | Lignin | Lipid | Carbonyl |
| Alkyl C (0–45 ppm) | 0 | 39.6 | 10.5 | 75.6 | 0 |
| Methoxy (45–60 ppm) | 4.3 | 21.9 | 13.8 | 4.5 | 0 |
| O-alkyl (60-95 ppm) | 79 | 2.1 | 12.5 | 9 | 0 |
| Di-O-alkyl (95-110 ppm) | 15.7 | 0 | 8.6 | 0 | 0 |
| Aromatic (110-145 ppm) | 1 | 7.5 | 30.6 | 3.6 | 0 |
| Phenolic (145-165 ppm) | 0 | 2.5 | 19.5 | 0.7 | 0 |
| Carbonyl (165-210 ppm) | 0 | 26.4 | 4.6 | 6.6 | 100 |

Table S2 Elemental composition and nominal oxidation state of five classes of organic compounds from Baldock et al. (2004)

|  |  |  |
| --- | --- | --- |
| Organic compound | Elemental formula | NOSC |
| Carbohydrate |  | 0 |
| Protein |  | 0.034 |
| Lignin |  | -0.381 |
| Lipid |  | -1.47 |
| Carbonyl |  | 3 |