Predicting Grass Growth for Sustainable Dairy Farming: A CBR System Using Bayesian Case-Exclusion and *Post-Hoc*, Personalized Explanation-by-Example (XAI)

Eoin M. Kenny[1,2,4] (✉), Elodie Ruelle[3,4], Anne Geoghegan[3,4], Laurence Shalloo[3,4], Micheál O'Leary[3,4], Michael O'Donovan[3,4] and Mark T. Keane[1,2,4]

[1] School of Computer Science, University College Dublin, Dublin, Ireland
[2] Insight Centre for Data Analytics, UCD, Ireland
[3] Teagasc, Animal and Grassland Research, Ireland
[4] VistaMilk SFI Centre, Ireland
{eoin.kenny, mark.keane}@insight-centre.org
{elodie.ruelle, anne.geoghegan, laurence.shalloo, michael.oleary, michael.odonovan}@teagasc.ie

Abstract. Smart agriculture has emerged as a rich application domain for AI-driven decision support systems (DSS) that support sustainable and responsible agriculture, by improving resource-utilization through better on-farm, management decisions. However, smart agriculture's promise is often challenged by the high barriers to user adoption. This paper develops a case-based reasoning (CBR) system called PBI-CBR to predict grass growth for dairy farmers, that combines predictive accuracy and explanation capabilities designed to improve user adoption. The system provides *post-hoc, personalized explanation-by-example* for its predictions, by using explanatory cases from the same farm or county. A key novelty of PBI-CBR is its use of Bayesian methods for case exclusion in this regression domain. Experiments report the tradeoff that occurs between predictive accuracy and explanatory adequacy for different parametric variants of PBI-CBR, and how updating Bayesian priors each year reduces error.

Keywords: CBR, Bayesian Analysis, Smart Agriculture, Case Exclusion, XAI

1 Introduction

Although the promise of artificial intelligence (AI) in smart agriculture is usually advertised as increasing productivity, in the future it may become increasingly about improving sustainability [1, 2]. As climate change accelerates, what AI may actually deliver is a precision agriculture that allows farmers to measure, balance, and predict the outcomes of farm management-decisions in ways that mitigate the environmental impact of these activities. However, this future depends on the development of AI-enabled decision support systems (DSS) that are both predictively accurate (e.g., in predicting grass growth), and explainable to the end user (i.e., farmers) to encourage adoption and usage. In this paper, an existing DSS called PastureBase Ireland (PBI) is extended by using case-based reasoning (CBR) techniques; the so-called PBI-CBR system. This new DSS predicts grass growth for dairy farmers and offers explanations designed to improve user adoption. As such, the system is an instance of eXplainable AI (XAI), providing *post-hoc, personalized explanation-by-example* for its predictions, based on location (using cases from the same or nearby farms). One key novelty of PBI-CBR is its use of what we refer to as *Bayesian Case-Exclusion,* which excludes outlier cases from the prediction process using prior beliefs about data distribution(s), reducing error and improving explanations. In the remainder of this introduction, the sustainability context for this work is briefly described, before outlining the structure of the paper.

1.1 Context: Agriculture, Sustainability and AI

Concerns about the impact of agriculture on climate change and the development of sustainable models are growing [2]. The agricultural sector and consumers are faced with varying views from climate change denial, to proposals that animal agriculture is responsible for 18-51% of greenhouse grass emissions [29, 30]. However, there is perhaps a middle ground that is exemplified by the work here.

Recently, an argument has emerged arguing for a quick move to sustainable farming systems [5]; the so-called *agroecology* perspective. For example, in the dairy sector this agroecology view has proposed a move to pasture-based systems, where animals are predominantly fed on grass outdoors rather than on meal and supplements indoors. The pasture-based proposal has the potential to be sustainable, in part, by using grass as a carbon sink and extending the grazing season (reducing slurry emissions) [11]. Furthermore, humans have limited capacity to digest grass, as it is a non-edible protein, so it is not consuming a food people could eat [28]. However, these initiatives depend on precision technology, using AI, to monitor variables such as climate and grass growth.

This paper considers a CBR system[1] that supplements an existing DSS used by several thousand Irish dairy farmers (i.e., PBI), which predicts grass growth in the coming week for a specific farm and offers personalized explanations (see Sections 4 and 5). However, as we shall see, there are significant challenges in handling the data noise which arises in this domain, especially against the backdrop of increasing climate disruption. Finally, for the sake of brevity, note that we only consider the *retrieval* step of CBR for this current iteration of PBI-CBR.

1.2 PastureBase Ireland for Dairy Farmers (PBI)

Smart agriculture often depends upon providing new DSSs for farmers to aid them in making complex decisions about how to manage their farm for productivity and sustainability [2]. These systems have three main challenges. First, they must be predictively accurate. Second, they need to be easy to use and interpretable for end users, to encourage adoption and continued use. Third, they need to be able to support decision making in the context of increasing climatic disruption, where the climate in past years

[1] Several other approaches such as linear regression, neural networks, SVMs and tree algorithms were also tested alongside this CBR system. The CBR system's accuracy equalled or bettered these other systems.

may not be indicative of climate in future years. The present work extends an existing DSS called PBI used in the grass-fed, pasture-based dairy farming systems in Ireland.

PastureBase Ireland. Since 2013, Ireland's national agricultural research organization *Teagasc*, have provided PastureBase Ireland (PBI, https://pasturebase.teagasc.ie) as a grassland management system to provide information and advice for Irish dairy farmers. PBI has 6,000+ users out of ~18,000 dairy farmers in Ireland. Among other features, the PBI database has weekly records of *grass covers* for individual farms from 2013 to present. A *grass cover* for a farm is principally, the amount of grass available on that farm for cows to eat; formally, it is a measure of biomass in grass on the farm above ground level or a height of 4cm. PBI allows farmers to enter this grass cover data for each field/paddock of their farm in a given week, using their own measurements/estimates, thus allowing them to budget grass-availability for their herd. Our system, PBI-CBR, uses the grass growth rates calculated by PBI from this grass cover data to predict grass growth rates on a farm from one week to the next, a critical part of the grass budgeting process. Note, farmers vary in how regularly they use PBI; there are ~2,000 active users defined as those entering > 20 grass covers a year.

Fig. 1. A grass wedge as seen by farmer-users of PBI: The green columns represent each field/paddock on a farm, and the red line the *target pre-grazing yield* each paddock should be at before beginning rotational grazing. The y-axis is kilograms of dry matter per hectare, and the x-axis shows the farm's paddocks. The width of each paddock's green bar represents its total area. The *Days Last Event* number refers to when the paddock was last grazed.

Feed Forecasting and Grass Wedges. Among other variables, the feed needs for a dairy herd depends on the size of one's farm, the size of the herd, and the status of the herd (e.g., lactating animals). PBI takes these variables and forecasts the feed needs for a farm. PBI accounts for both *rotational grazing* and *set stocking*, in which the farmer grazes certain paddocks while resting others (which may be grazed later or cut for silage). PBI allows farmers to modify variables such as rotation length and paddock status (e.g., is it currently being grazed), while producing a number of reports to show the effect of changing variables. Fig. 1 shows all paddocks on a farm and the grass available in each paddock, measured in kilograms of dry matter per hectare (kg DM/ha; grass weight changes with moisture content, so dry weight is used). The red-line shows the *target pre-grazing yield* for each paddock, which can move up and down as the farmer changes variables (e.g., size of herd). If the red-line is below the top of a green-bar, then more grass is available than is currently needed (it could be cut for silage or meal supplementation reduced). However, if the red-line is above a green-bar then there is not enough grass to begin rotational grazing, and some meal supplementation may be required. These calculations are critical to the sustainability of the farming enterprise; stated simply, grass is inexpensive and meal supplements are the opposite. Also, meal requires transportation and possibly importation, so it entails increased carbon costs.

Grass Growth Prediction. Management decisions are largely based on grass growth, which varies based on soil/grass type, farming practices, climatic factors etc. In PBI, the farmer estimates a grass cover in paddocks and a calculation is done to determine the average growth rate since the previous grass cover. PBI-CBR aims to predict growth rates using machine learning (ML), by forecasting the growth-rate in the coming week using previous cases. Note, Teagasc currently uses a *mechanistic model* (a.k.a. a *first-principles* model) called the *Moorepark St. Gilles Grass Growth* model (MoST) that can predict growth-rates and continues to be tested [4]. However, key parameters of this model are not available for all farms (e.g., soil maps). A future system may combine PBI-CBR and MoST to make predictions, alternating between both models.

The PBI Dataset. We used the PBI dataset recorded from thousands of private farms in Ireland between 2013-2017. The primary feature of concern is the average grass growth rate for a farm since the last grass cover recorded, but location features (Farm ID-anonymized and County) are also important for explanation purposes. Ideally, to explain a prediction our system aims to provide an explanatory case from the same farm, but a case from a nearby farm in the same county is also acceptable. This was the advice given by the domain experts running the current system, although ultimately this proposal needs to be user tested.

1.3 Outline of Paper

Section 2 discusses noise in the PBI dataset used here and how a Bayesian approach is both useful and intuitive for case exclusion. Section 3 describes Experiment 1 (Expt. 1) which compares four systems on accuracy and explanatory success, and the tradeoff between both measures. Section 4 describes Expt. 2 which shows how updating priors using Bayesian analysis can improve prediction accuracy, possibly providing a means to deal with climate change in DSSs for this domain. Section 5 reviews relevant previous work in the area before final conclusions in Section 6.

2 Noise: The Gold-Standard and Working-Farm Datasets

We believe that this grass-growth domain is representative of the datasets and problems that AI will face in many smart agriculture contexts, especially in being *highly noisy*.

The data is gathered by end users (farmers) and, as such, is understood to contain errors, miss-recordings, adjustments, and estimates. For example, some of the recordings in the dataset are based on physical measurements with a device, whereas others are estimates from visual inspections. This inherent noise has profound implications for how prediction and explanation need to be handled in this domain. On the one hand, we need a systematic way to remove possibly-poor cases. On the other hand, we need to keep as many cases as possible, because each additional case has potential to improve the system's accuracy and interpretability. Indeed, case exclusion could also affect the tradeoff between predictive accuracy and explanatory adequacy. Our solution to these noise issues is to use one dataset to clean another; to use a gold-standard dataset gathered under controlled conditions by researchers (with is idealized but noise-free) to clean a working-farm dataset gathered by farmers as part of their daily work (which is noisy). Technically, we use a gold-standard set of historical grass-growth measurements to give a prior belief about the distribution of grass growth each week, which in turn allows us to exclude cases from the working-farm case-base that may contain errors; what we call

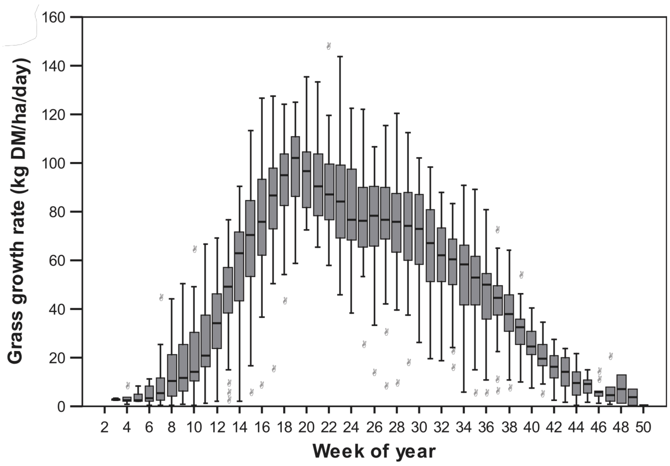

Fig. 2. The gold-standard dataset of grass growth measurements from 1982–2010 at Teagasc, Animal and Grassland Research and Innovation Centre, Moorepark, Fermoy, Co. Cork, Ireland [7], where the distribution of grass growth each week of the year is given as box plots.

Bayesian Case-Exclusion. As we shall see, this solution seems to exclude noisy cases while retaining enough high-quality ones to maintain accurate predictions and explanations. Next, we describe these two datasets and how the working-farm case-base was built.

2.1 A Gold-Standard Dataset: Teagasc Grass Growth (1982-2010)

The gold-standard dataset of grass-growth measurements we used covers 28 years of carefully-controlled, weekly measurements in which samples taken by researchers from the same pasture were cut, dried and weighted on a weekly basis at the Teagasc Moorepark Dairy Research Centre, Fermoy, Co. Cork (a major location for dairy farming). These measurements are somewhat idealized as they come from one location, which was not grazed (i.e., the livestock's impact – such as urine and trampling – on grass growth was excluded). However, they are very accurate and can thus serve as a good benchmark for determining outlier cases in the PBI dataset.

2.2 Case Definition and Case-Base Construction

The dataset used to construct the working-farm case-base came from the weekly grass covers entered by farmers in PBI. This dataset's growth rates were calculated using the grass covers recorded by farmers showing the estimate of grass available on a given farm for a given day. Some of these grass covers are known to be in-error; for example, often multiple entries are made on the same day, where the last entry of the day was the intended record. For the years 2013-2017, this dataset had 99,087 grass cover-records, that reduced to 92,635 when these same-day entries were removed. These grass cover-records are the raw data from which the cases used in PBI-CBR were generated to create the working-farm case-base.

Case Generation. Let a farm's data be $f = \{x_1, x_2, \ldots x_n\}$, where x_i is a grass cover-record for a single day, and n the total number of grass covers recorded (note the grass covers are in chronological order). The features of x_i used to generate a case (C_i) in the case-base are the average growth rate since the previous grass cover (gr), the week (wk), month (mth), and season ($seas$) in which the grass cover was recorded. Weather data (w_i) at the county level was scraped from Met Éireann (www.met.ie), and added as an average from x_i until x_{i+1}. The weather information in w_i is the maximum daily temperature ($maxt$), the average soil temperature 10cm below the surface ($soilt$) on a given day, and the average global radiation ($grad$) in a given day. Finally, gr from x_{i+1} is also added to C_i as the target feature for prediction. Thus, a case is represented as:

$$C_i(x_i, w_i, x_{i+1}) = \langle x_i(gr, wk, mth, seas), w_i(maxt, soilt, grad), x_{i+1}(gr) \rangle \quad (1)$$

Case Base Construction. Taking the raw-data grass cover-records (N=92,635) the cases as defined in (1) were constructed. However, given that the system has to predict one week ahead, only those cases where the target $x_{i+1}(gr)$ was recorded 5-9 days after x_i were included in the case base. Also, cases from January and December were excluded (as they tend to show zero growth), though they might be appropriate in a final deployed system. Finally, only those cases with accurate historical weather information until the next grass cover were considered (weather is a crucial factor in growth predictions). These steps resulted in a *working-farm case-base* of N=20,760 cases for use in experimental tests. Note, in each system variant (except for the *Control*) the number of cases in this *working-farm case-base* is reduced further by the respective method(s) used.

2.3 The Current Experiments

In the remainder of this paper, two experiments are reported that test several variants of the *Bayesian Case-Exclusion* idea. In Expt. 1, we examine what happens in this predictive CBR-system when cases are not excluded (*Control*), versus experimental systems in which we use the gold-standard dataset's distributions in different ways to modify or exclude cases (the *Exclude-2sd, Exclude-3sd and Transform-3sd* systems; see Section 3). These experimental system-variants examine performance when cases are *transformed* with reference to the gold-standard distributions or when cases are *excluded* a pre-defined number of standard deviations away from the means in the gold-standard distributions. The transformation system enables greater retention of cases, in turn helping with explanations. In Expt. 2, we explore *Adaptive Bayesian Case-Exclusion*, where priors derived from the gold-standard distributions are updated year-on-year, to see if performance improves (see Section 4).

3 Experiment 1: Bayesian Case-Exclusion

PBI-CBR is a CBR system for predicting grass growth, using the growth rates calculated from each farm. Two different datasets are used in the experiments, the gold-standard Teagasc data (1982-2010) and the PBI dataset (2013-2017), where the former is used to transform or exclude cases from the latter when making predictions for a particular farm in a given week of a given year. Hence, the gold-standard dataset is our "prior" belief (in Bayesian parlance), which is used to make probabilistic inferences in how to handle noise. In the working-farm case-base, the current week is used to predict one week ahead, allowing a farmer to make informed management decisions. In general, for this prediction, a mean squared error (MAE) of $\leq$ 10 kg DM/ha/day is sufficient. The main problem is the noise in the working-farm case-base, hence we use *Bayesian Case-Exclusion* to exclude outlier cases when making predictions. PBI-CBR also explains predictions using *post-hoc, personalized explanation-by-example* by referencing nearest neighboring cases from the same farm or county. So, the tests involve two measures: (i) *predictive accuracy*, as MAE for the growth-rate prediction measured in kg DM/ha/day, (ii) *explanatory success*, as the percentage of times nearest-neighbor cases are found from either the same-farm or same-county to the test-cases in the k nearest neighbors retrieved (a measure recommended by experts). However, it should be noted again that the "success" of these explanations is dependent on future user testing. Crucially, we tested four variants of the system:

- *Control*. A basic system that uses all the cases in the *working-farm case-base* (N=20,760; see Section 2.2); this case base was built mostly from the PBI dataset (from 2013-17) and, accordingly, is quite noisy and has many outliers.
- *Exclude-2sd*. A Bayesian system that excludes cases two-standard deviations away from the weekly, mean growth-rates of the gold-standard dataset (see Fig. 2). The rationale being that grass growth in a given week approximates a normal distribution (verified by plotting thousands of growth rates in histograms) and using the properties of such a distribution can aid in making probabilistic assumptions for how to exclude cases. Formally, the data for growth rate (GR) in a given week across all years in the gold-standard dataset approximates $GR \sim N(\mu_g, \sigma_g^2)$, where N is a normal distribution with parameters μ_g and σ_g for the mean and standard deviation,

respectively. All cases outside $\mu_g \pm 2\sigma_g$ are excluded (as well as other query-cases), thus excluding cases with ~5% probability of occurring. This step reduces the *working-farm case-base* by 42% (N=12,042 cases).

- *Exclude-3sd*. This is identical to the *Exclude-2sd* system but $\mu_g \pm 3\sigma_g$ is used to exclude cases, thus excluding cases with ~0.3% probability of occurring. This reduces the *working-farm case-base* by 21% (N=16,443 cases).
- *Transform-3sd*. This is a Bayesian system that transforms the growth-rates of cases using the gold-standard distributions. That is, the distribution of growth in a given week from the gold-standard dataset $[GR \sim N(\mu_g, \sigma_g^2)]$ is used to transform the growth-rate values of cases for the same week in the *working-farm case-base*, to fit to the parameters μ_g and σ_g^2. Formally, to transform the growth-rate (gr) in a grass cover x in any given week of the year we use:

$$y_{gr} = \left(x_{gr} - \mu\right) \times \frac{\sigma_g}{\sigma} + \mu_g \tag{2}$$

where x_{gr} is the growth rate in grass cover x, y_{gr} is the transformed growth rate of x_{gr}, μ and σ are the mean and standard deviation for the overall growth rate in that week in the *working-farm case-base,* respectively, and μ_g and σ_g are the mean and standard deviation for the overall growth rate in that week in the gold-standard dataset, respectively. The intuition being that the gold-standard dataset is closer to the ground-truth, hence if it is used to transform the growth rates (in the *working-farm case-base*), the overall deviation from the ground truth will reduce. Note, in this system cases that fall outside $\mu_g \pm 3\sigma_g$ after the transformation are still excluded, and, so, the *working-farm case-base* is reduced by 2% (N=20,282 cases).

As we shall see, exclusion methods improve prediction accuracy, with varying levels of explanatory success. The transform system retains as many cases as possible, aiding accuracy and explanatory success. Indeed, there are indications that the transformed case-base is closer to the ground truth as the correlation of Pearson's r between $maxt$ and *growth-rate* across all cases increases from $r = 3.92$ to $r = 5.11$ after transformation, reflecting known dependencies between temperature and grass-growth ($<$ 5°C grass does not grow, from 5-10°C it grows with temperature [4]).

3.1 Method: Procedure and Measures

For each system variant Monte Carlo cross-validation was used with 30 resampling iterations, each time taking 80/20% data for training and testing, respectively. An unweighted *k*-NN algorithm with Euclidean distance was used for case retrieval, with the averaged value of all nearest neighbors' target-growth-rates used as the prediction. Selected values of *k* ranging from 5-1000 were tested for each system variant to observe effects on prediction and explanatory outcomes. For each evaluation of *k* for each system, three measures were taken: (i) the MAE (ii) the *%Farm-Retrieval-Success*, the percentage of times the *k*-nearest-neighbors contained a case from the same farm as the query, and (iii) the *%County-Retrieval-Success*, the percentage of times the *k*-nearest-neighbors contained a case from the same county as the query.

3.2 Results and Discussion

Fig. 3a shows the results of running the system variants – *Control, Exclude-2sd, Exclude-3sd, Transform-3sd* – for all values of *k* in three graphs, one for each measure: MAE, %Farm-Retrieval-Success (%FRS), and %County-Retrieval-Success (%CRS). Across all systems, MAE is worst for the lowest and highest *k* with some improvement in between (*k* = 20-35). With regard to %FRS all system variants are very similar, though success does change for different values of *k*. For all systems, %FRS is very poor for low values of *k,* but beyond *k* = 50 it rises to ~80%; showing that only higher values of *k* deliver enough cases from the same farm to explain the predictions made. For all systems, %CRS is much better, as it starts high (~80%) for low values of *k* and rapidly reaches ~100%; showing that finding explanatory cases for a prediction from the same county is a common occurrence. However, the differences between the system variants are, perhaps, more interesting.

Overall, the *Control* system, which includes all cases, does the worst; it never gets lower than a MAE of 15 kg DM/ha/day (recall, acceptable error is ≤ 10 kg DM/ha/day). Similarly, the two exclusion-systems – *Exclude-2sd* and *Exclude-3sd* – do not reach the acceptability threshold. Overall, *Bayesian Case-Exclusion* does much better than the *Control*, but only *Exclude-2sd* with *k* = 35 has the somewhat acceptable MAE of ~10.01 kg DM/ha/day. Overall, the *Transform-3sd* system is the best with a MAE < 10 kg DM/ha/day for all values of *k* (note, many current mechanistic models have MAEs of ~10-20 kg DM/ha/day, showing the potential for AI solutions in this domain).

Finally, the best system is *Transform-3sd*; in Fig. 3a, comparing the 1st and 2nd graphs, we can see the tradeoff between MAE and %FRS for all values of *k*. The 1st graph shows that the lowest error (MAE = 8.6 kg DM/ha/day) occurs at *k* ~ 35, but at this level %FRS is poor at ~7% (see 2nd graph). Accordingly, *k* = 1000 is required to improve %FRS to ~85%. However, even at this value for *k*, an acceptable MAE is achieved (~9.8 kg DM/ha/day), making *Transform-3sd* the only system that successfully balances the tradeoff between accuracy and explanation. Note, with additional data from a given farm, it should be possible to improve this tradeoff even further.

4 Experiment 2: Updating Priors Year-on-Year

In Expt. 1 Bayesian exclusion or transformation of cases from the *working-farm case-base* gave improved performance. However, these systems exclude cases using parameters from the gold-standard dataset, gathered between 1982 and 2010. Recently climate change appears to be impacting the distribution for grass growth. For example, in the hot Irish summer of 2018 grass-growth stopped during July (normally it is ~100 kg DM/ha/day). In Expt. 1, this not considered, but Expt. 2 rectifies this by combining the two datasets to update Bayesian priors year-on-year by Bayesian analysis (e.g., see [3]) to estimate the unknown distributions of grass growth each week with a view to making predictions in 2017[2] (the final year's data). Hence, Expt. 2 has six versions of PBI-

[2] Predictions could only be made for 2017 because the earlier years of the PBI dataset (2013-2016) have too few cases, as the DSS was in its early years of adoption.

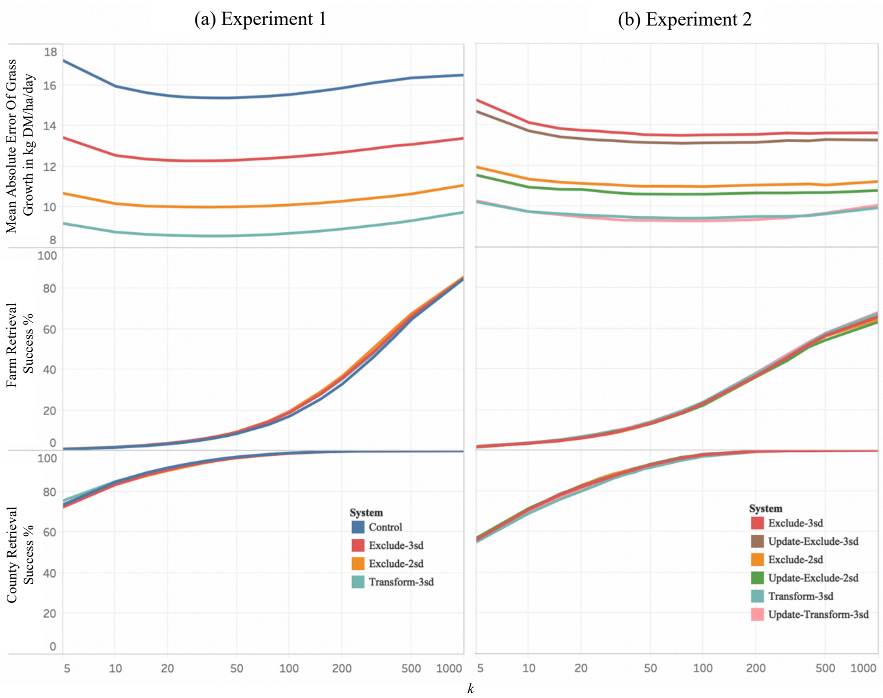

Fig. 3. The tradeoff between error and explanation. (a) Expt. 1 shows that as the value for *k* approaches 1000, more explanatory cases are retrieved, but the MAE for all systems also increases. *Transform-3sd* has the best MAE of ~8.6 kg DM/ha/day at *k* ~ 35, but same-farm explanatory success is low at ~7%; however, at *k* = 1000, the tradeoff is balanced, with the MAE still acceptable and %FRS at ~85%. (b) Expt. 2 shows MAE is improved for almost every update-variant, although the improvement in the transform-system is minimal; explanatory success and MAE are similar to Expt. 1, but poorer, likely due to less training data. Finally, note the log scale on the x-axis.

CBR, three systems from in Expt. 1 (*Exclude-2sd, Exclude-3sd, Transform-3sd*) and three variants of these in which priors were updated (*Update-Exclude-2sd, Update-Exclude-3sd, Update-Transform-3sd*). The updating procedure used is described next.

4.1 Updating Priors in Exclusion and Transformation Systems

To perform Bayesian updating, we take priors from the gold-standard dataset and then progressively use each year's data from the PBI-dataset to update them. First, take the gold-standard dataset and, binning all its data into weeks, for any given week, let the growth rate (GR) approximate a normal distribution $GR \sim N(\mu, \sigma^2)$, where μ and σ^2 are its mean and variance, respectively. In 2013, all the data for this week was processed into cases (see Section 3). Then, we proceed with *transformation* or *exclusion* methods on these cases depending on the system variant (as in Expt. 1), which gives the new data $D = \{C_1, C_2 \ldots C_n\}$ where n is the number of cases. Take the prior to be

$\mu \sim N(\mu_0, \sigma_0^2)$, where the value σ_0 is initially chosen as 4[3] and μ_0 is initially chosen as μ. Here the value for σ^2 is assumed to remain fixed[4]. Bayes rule shows that the posterior (for a given week) is proportional to the *likelihood* times the *prior*, in addition, because σ^2 and σ_0^2 are known we can ignore the constant of proportionality and derive that the posterior μ_p is:

$$\mu_p \sim N\left(\frac{\sigma^2}{\sigma^2+\sigma_0^2 n}\mu_0 + \frac{\sigma_0^2}{\sigma^2+\sigma_0^2 n} n\overline{x},\ \ \frac{\sigma^2\sigma_0^2}{\sigma^2+\sigma_0^2 n}\right) \tag{3}$$

where $\overline{x}$ is the empirical mean of the growth rates in the cases of D, for a full derivation and explanation the reader is referred to [3]. Although in CBR the word "Bayesian" usually infers the use of Bayesian networks, in this experiment it is used in a more traditional sense and refers to the estimation of an unknown distribution (a.k.a. the *posterior*) of grass growth using a prior belief (a.k.a. the *prior*) and a sample of data from the new year (i.e., the *likelihood*).

Using equation (3) we update values for μ_0 and σ_0^2, the new value of μ_0 was then used to update the original μ from the gold-standard dataset, which was used with σ^2 (the fixed variance from the gold-standard dataset) to repeat the whole process in 2014 for the same week. This process is repeated for all weeks of each year until the end of 2016 when all training data was collected. The latest priors in each week were again used to exclude or transform cases in 2017 for evaluation[5]. All evaluations were carried out on 2017 because there was insufficient training data in previous years to ensure adequate evaluations (2017 has ~40% of usable cases), though the years prior to 2017 were all used in the year-on-year updating to acquire the training data.

4.2 Case-Base Sizes after Transformation or Exclusion

Expt. 2 has six system variants, the *Exclude-2sd, Exclude-3sd* and *Transform-3sd* systems from Expt. 1, and matched versions of these systems, which used the updating methods described above called *Update-Exclude-2sd, Update-Exclude-3sd and Update-Transform-3sd*. In the updated variants, several aspects change, so the number of cases after transformation or exclusion vary slightly: *Exclude-2sd* (N=12,042), *Update-Exclude-2sd* (N=12,183), *Exclude-3sd* (N=16,443), *Update-Exclude-3sd* (N=16,379), *Transform-3sd* (N=20,282), and *Update-Transform-3sd* (N=20,120).

4.3 Method: Procedure and Measures

For each system variant the respective case base was split in a ~60/40% ratio of training and testing cases, respectively; the former coming from the PBI data from 2013-2016

---

3 The relatively large value of 4 was chosen to represent that we are not highly certain of the validity of the gold standard prior mean when compared to a typical dairy farm pasture.

4 The variance σ^2 wasn't adapted; if it changes it could lead to an unfair evaluation as *updated*-variants may differ a lot in the amount of data excluded compared to non-updated variants.

5 Note, for *transform* methods some knowledge about a given week's data distribution would need to be inferred if we were doing this in a live-system for formula (2) to be used.

and the latter from 2017. Crucially, note that results will be different from identical systems in Expt. 1 because of the different ratio for splits. For case retrieval, an unweighted *k*-NN was again used with Euclidean distance for selected values for *k* ranging from 5-1000. The same three measures were used as in Expt. 1: Mean Absolute Error (MAE), %Farm-Retrieval-Success (%FRS), and %County-Retrieval-Success (%CRS).

4.4 Results and Discussion

Fig. 3b shows the results of running the six system variants – *Exclude-2sd, Exclude-3sd, Transform-3sd, Update-Exclude-2sd, Update-Exclude-3sd and Update-Transform-3sd* – for all values of *k* in three graphs, one for each measure: MAE, %FRS, and %CRS. In general, the shape of the results replicates many of the findings of Expt. 1.

Regarding MAE, as before the transformation-versions do better than the exclusion-versions, the error decreases in order from *exclude-3sd* to *exclude-2sd* to *transform-3sd*; $k = 75$ is optimal for all systems, doing better than the lower and higher values of *k*. Overall the MAE scores (and explanation-success scores) are not as good as in Expt. 1, perhaps, reflecting the different ratios in the training and testing splits (i.e., they were 80/20% in Expt. 1 and ~60/40% in Expt. 2); note, the evaluation dataset is reduced to one-year in Expt. 2 (i.e., 2017), whereas it is across all 5 years in Expt. 1 (2013-17).

Expt. 2 shows that systems with Bayesian updating (*Update-Exclude-2sd, Update-Exclude-3sd and Update-Transform-3sd*) do better than systems without updating (*Exclude-2sd, Exclude-3sd, Transform-3sd)* at nearly every value of *k*, though the improvements are relatively modest, particularly in the transform version (see Fig. 3b).

Regarding explanation measures (%FRS, %CRS) the overall curve-shapes are similar to those in Expt. 1, with maximum values being %FRS=68% and %CRS=100%, in contrast to %FRS=85.94% and %CRS=99.98% in Expt. 1. Acceptable tradeoffs for accuracy and explanation are achieved for both of the transform systems (*Transform-3sd, Update-Transform-3sd*) in that at $k = 1000$ the MAE is ~9.95 kg DM/ha/day with ~67.5% explanatory-success rate for same-farm cases in both systems. These systems would likely improve if training and testing splits were more favorable as in Expt. 1.

5 Related Work

This work impinges on many areas, though the most relevant literatures are arguably in case-based maintenance, Bayesian CBR, and explanation in CBR DSSs for smart agriculture. Here we review the relevant literature and discuss its relevance to this work.

Case base maintenance (i.e., case base editing/deleting/exclusion/inclusion etc.) is a notable area of research for the CBR community [19]. However, the most popular methods have tended to focus on classification [16, 20-25], as opposed to regression [17]. Redmond and Highley [17] did try to convert Edited Nearest Neighbors [22] to handle regression by assigning two hyperparameters for *agree* and *accept* thresholds, but they acknowledge that applying the classification algorithms to regression is difficult. Our method requires no hyperparameters, though it does require the specification of a prior(s). Furthermore, most of the literature on case base maintenance is concerned with deleting cases to optimize case-bases; here we have used the phrase "case exclusion" rather than "case deletion" because we believe it is important to retain cases for future use. For instance, cases deemed outliers with extreme environmental conditions may be useful if climate change results in these extreme conditions becoming common or more data becomes available (e.g., soil type) identifying them as non-outlier data.

Much work has been done using Bayesian methods in CBR systems. Nikpour *et al.* [8] used Bayesian posterior distributions to modify case descriptions and dependencies in a model, showing the capability of such an approach to increase similarity assessment. Moreover, the vast majority of work combining CBR and Bayesian methods has involved combining Bayesian Networks with CBR systems, for which there are many architectures and approaches [27]. However, beyond the combination of Bayesian methods and CBR, these systems have little in common with the present work, which uses prior distributions for case exclusion. The best algorithm for a particular problem regarding case base maintenance will likely always depend on the domain in question [19], but here we present a novel option.

XAI within CBR has been shown to be important in intelligent systems [9, 31, 32], with some consideration of smart agriculture [10]. Additionally, it has been argued that recommender systems should play a central part in smart agriculture [12], and CBR is a popular approach for such systems [6]. Pu and Chen [26] have conducted user studies showing that designers should build trusted interfaces into recommender systems due to the high likelihood users will return. As smart agriculture arguably requires a recommender component [12], and it suffers from a user retention issue; this is of particular relevance. Moreover, in understanding the effects of environmental changes, Cho *et al.* [12] note that global warming and pollution have made environmental and agricultural modelling difficult, thus suggesting the use of a recommender system to support users, but no specific instances are described. Moreover, Holt [14] suggested that CBR could be used to help farm management decisions. CBR gives a unique ability to offer intuitive exemplar-based explanations, and user studies have shown it potentially superior to rule-based explanations [13], frameworks have been proposed for CBR XAI [18], but to the best of our knowledge no instance in smart agriculture has been proposed until the present paper. Branting *et al.* [10] did use CBR in the agricultural advisory system CARMA (which also produces explanations), but it only forms part of the consultation process, whilst our solution appears to be the first pure CBR approach.

6 Conclusions and Future Research

We have shown that a CBR system can be used for decision support in dairy farming to predict a key aspect of the enterprise accurately, while also providing case-based explanations that are personalized for a specific farm. To deal with noise in the dataset, we have used historical distributions based on accurate research measurements to determine what cases should or should not be included in the predictive model (i.e., our Bayesian exclusion approach). Furthermore, we have shown that transforming key-attributes of cases based on a goal-standard distribution (that is closer to a ground truth) can improve accuracy, and that using Bayesian analysis for updating priors year-on-year also improves performance. By our knowledge, all of this work is novel.

These systems have the ability to improve the sustainability of grasslands for dairy farming into the future. Accordingly, for us, the key question for future research is whether these techniques can continue to deliver accurate predictions in the face of climate change. One would hope that these CBR systems can maintain predictive accuracy by selectively picking useful cases from historical datasets (e.g., as soon as the data is available, we plan to test PBI-CBR against the extreme weather of 2018). So, though we may experience significant climate shifts, there will always be a case somewhere in the historical record that can provide accurate predictions.

Acknowledgements. This publication has emanated from research conducted with the financial support of (i) Science Foundation Ireland (SFI) to the Insight Centre for Data Analytics under Grant Number 12/RC/2289 and (ii) SFI and the Department of Agriculture, Food and Marine on behalf of the Government of Ireland to the VistaMilk SFI Research Centre under Grant Number 16/RC/3835.

References

1. Gafsi, M., Legagneux, B., Nguyen, G. and Robin, P.: Towards sustainable farming systems: Effectiveness and deficiency of the French procedure of sustainable agriculture. Agricultural Systems, 90(1-3), 226-242 (2006)
2. Lindblom, J., Lundström, C., Ljung, M. and Jonsson, A.: Promoting sustainable intensification in precision agriculture: review of decision support systems development and strategies. Precision Agriculture, 18(3), 309-331 (2017)
3. Murphy, K.P.: Conjugate Bayesian analysis of the Gaussian distribution. def, 1(2σ2), p.16 (2007)
4. Ruelle, E., Hennessy, D. and Delaby, L.: Development of the Moorepark St Gilles grass growth model (MoSt GG model): A predictive model for grass growth for pasture based systems. European Journal of Agronomy, 99, 80-91 (2018)
5. Poux, X. and Aubert, P.M.: An agroecological Europe in 2050: multifunctional agriculture for healthy eating (2018)
6. Tintarev, N. and Masthoff, J.: A survey of explanations in recommender systems. In: 2007 IEEE 23rd international conference on data engineering workshop, pp. 801-810. IEEE (2007)
7. Hurtado-Uria, C., Hennessy, D., Shalloo, L., O'Connor, D. and Delaby, L.: Relationships between meteorological data and grass growth over time in the south of Ireland. Irish Geography, 46(3), 175-201 (2013)
8. Nikpour, H., Aamodt, A. and Bach, K.: Bayesian Analysis in a Knowledge-Intensive CBR System. In: 22nd UK Symposium on Case-Based Reasoning. Miltos Petridis (2017)
9. Leake, D.B.: Evaluating explanations: A content theory. Psychology Press (2014)
10. Branting, K., Hastings, J.D. and Lockwood, J.A.: CARMA: A Case-Based Range Management Advisor. In *IAAI* (pp. 3-10) August (2001)
11. Teagasc - 8/2/19: The Dairy Carbon Navigator. Improving Carbon Efficiency on Irish Dairy Farms, https://www.teagasc.ie/publications/2019/the-dairy-carbon-navigator.php.
12. Cho, Y., Cho, K., Shin, C., Park, J. and Lee, E.S.: A Recommend Service Based on Expert Knowledge Model in Agricultural Environments. In: International Conference on Hybrid Information Technology, pp. 189-194. Springer, Berlin, Heidelberg (2012)

13. Cunningham, P., Doyle, D. and Loughrey, J.: An evaluation of the usefulness of case-based explanation. In International Conference on Case-Based Reasoning (pp. 122-130). Springer, Berlin, Heidelberg (2003)
14. Holt, A.: Understanding environmental and geographical complexities through similarity matching. Complexity International, 7, 1-16 (2000)
15. Bridge, D., Göker, M.H., McGinty, L. and Smyth, B.: Case-based recommender systems. The Knowledge Engineering Review, 20(3), 315-320 (2005)
16. Hart, P.: The condensed nearest neighbor rule (Corresp.). IEEE transactions on information theory, 14(3), 515-516 (1968)
17. Redmond, M.A. and Highley, T.: Empirical analysis of case-editing approaches for numeric prediction. In Innovations in Computing Sciences and Software Engineering (79-84). Springer, Dordrecht (2010)
18. Sørmo, F., Cassens, J. and Aamodt, A.: Explanation in case-based reasoning–perspectives and goals. Artificial Intelligence Review, 24(2), 109-143 (2005)
19. Smiti, A. and Elouedi, Z.: Overview of Maintenance for Case based Reasoning Systems. International Journal of Computer Applications, 975, 8887 (2011)
20. W. Gates: The Reduced Nearest Neighbor Rule, IEEE Transactions on Information Theory, 18, 3, 431-433 (1972)
21. G. Ritter and H. Woodruff and S. Lowry and T.Isenhour: An algorithm for a selective nearest neighbor decision rule, IEEE Transactions on Information Theory, 21, 665-669 (1975)
22. D. L. Wilson: Asymptotic Properties of Nearest Neighbor Rules Using Edited Data, IEEE Transactions on Systems Man and Cybernetics, 2, 408-421 (1972)
23. D. Guan, W. Yuan, Lee, Y. K. Lee, S. Lee: Nearest neighbor editing aided by unlabeled data, Inf. Sci., 179, 13, 2273-2282 (2009)
24. D. W. Aha, and D.Kibler, and M. K. Albert: Instance-based learning algorithms, Machine Learning, Springer Netherlands, 6, 37-66 (1991)
25. S. Markovitch and P. D. Scott: The Role of Forgetting in Learning, In Proceedings of the Fifth International Conference on Machine Learning, pp.459-465 (1988)
26. Pu, P. and Chen, L.: Trust-inspiring explanation interfaces for recommender systems. Knowledge-Based Systems, 20(6), 542-556 (2007)
27. Bruland, T., Aamodt, A. and Langseth, H.: Architectures integrating case-based reasoning and bayesian networks for clinical decision support. In International Conference on Intelligent Information Processing (pp. 82-91). Springer, Berlin, Heidelberg (2010)
28. Johns, T.: Phytochemicals as evolutionary mediators of human nutritional physiology. International Journal of Pharmacognosy, 34(5), 327-334 (1996)
29. Steinfeld, H., Gerber, P., Wassenaar, T.D., Castel, V., Rosales, M., Rosales, M. and de Haan, C.: Livestock's long shadow: environmental issues and options. Food & Agriculture Org. (2006)
30. Goodland, R. and Anhang, J.: Livestock and climate change: What if the key actors in climate change are... cows, pigs, and chickens? (2009)
31. Keane, Mark T., and Kenny, Eoin M.: How Case Based Reasoning Explained Neural Networks: An XAI Survey of *Post-Hoc* Explanation-by-Example in ANN-CBR Twins. arXiv preprint arXiv:1905.07186 (2019)
32. Kenny, Eoin M. and Keane, Mark T.: Twin-Systems to Explain Artificial Neural Networks using Case-Based Reasoning: Comparative Tests of Feature-Weighting Methods in ANN-CBR Twins for XAI. In: Proceedings of the Twenty-Eight International Joint Conference on Artificial Intelligence (2019)