From chemical fingerprints to environmental footprints: Advancing feed production through near-infrared spectroscopy

Maria Cairoli[1,4,5,*], Anne Ottenbros[2,5], Sin Yong Teng[1], Steef Hanssen[2], Mark Schoot[3], Christiaan Kapper[3], Rosalie van Zelm[2], Lutgarde Buydens[1], Mark Huijbregts[2,6] and Jeroen Jansen[1,6]

[1] *Radboud University, Institute for Molecules and Materials, Heyendaalseweg 135, 6525 AJ, Nijmegen, The Netherlands*

[2] *Radboud University, Radboud Institute for Biology and Environmental Science, Heyendaalseweg 135, 6525 AJ, Nijmegen, The Netherlands*

[3] *Nutricontrol, N.C.B. Laan 52, 5462 GE, Veghel, The Netherlands*

[4] *Lead contact*

[5] *These authors contributed equally*

[6] *These authors contributed equally*

[*]Correspondence: maria.cairoli@ru.nl

Abstract

Animal feed production involves balancing nutritional quality, profitability and environmental sustainability. Although near-infrared spectroscopy (NIRS) is currently used for real-time quality control of feed ingredients, we demonstrate that NIRS can also predict their environmental sustainability in a resource-efficient way. We use NIRS to determine ingredient origins and combine these with global spatially-explicit life cycle assessment (LCA) to estimate environmental footprints. By incorporating ingredient prices and transport, we then optimize feeds towards the triple goals of quality, profitability and sustainability. We show 3.3-39% reductions in climate change and land stress impacts on biodiversity while reducing profitability by only 0.82-2.4% over current production and ensuring quality. Our approach provides a suite of optimal feed ratios and identifies footprint-profitability trade-offs, aiding decision-makers in moving towards more environmentally sustainable feed. We conclude that NIRS-LCA is a powerful combination for enhancing sustainability that can be extended beyond feed to food, fiber and other biobased commodities.

Keywords

Livestock feed production; agri-food system; environmental sustainability; climate change; biodiversity loss; life cycle assessment; near-infrared spectroscopy; multi-objective optimization; decision-making

Introduction

Feed is vitally important in the livestock sector to sustain animal health and ensure the production of safe and high-quality products of animal origin[1]. The production of livestock feed is under continuous pressure from food-feed competition, disruptions in feed ingredient supply chains, contamination episodes, variations in nutrient quality of feedstocks and demand for sustainable agricultural practices[1–3]. Globally, feed production is responsible for an estimated 45% of the greenhouse gas (GHG) emissions of the livestock sector and uses 33% of the total arable land[2]. The projected rise in the demand for animal products[4] increases the urgency of drastically reducing the environmental footprint of livestock feed production[5,6].

Animal feed is produced by selecting and combining feed ingredients to create nutritionally optimal mixtures that meet market demands[7–9]. Increasing sustainability in feed production requires finding a trade-off between several objectives, including economic profitability, socially acceptable practices, and reduced environmental footprints[10]. However, current production primarily focuses on minimizing costs within a quality range[7,11]; impacts on the environment are typically not considered. Several studies have shown that environmental footprints may vary greatly across feed ingredients, due to differences in cultivation, processing, and geographical origin[12–14]. Castonguay et al. recently showed that trade-offs between environmental impacts and monetary costs may improve feed sustainability within global beef production[10]. Although the environmental impacts of livestock feeds are increasingly assessed through life cycle assessment (LCA)[15], these impacts are used mainly for regulatory purposes and not integrated into feed optimization. Transparent integration of LCA results in real-time production is therefore necessary to enable decision-makers to design more environmentally sustainable and high-quality feeds.

Ensuring feed quality in industrial agriculture requires detailed chemical knowledge of the nutritional value of the feed ingredients[16]. Ideally, such information is obtained quickly and noninvasively during production and is available at product release. Most laboratory analyses are very costly and too time-consuming for use in real time[17,18]. Therefore, to determine the nutritional composition of ingredients, feed manufacturers often resort to off-line laboratory measurements through occasional wet analysis[18] or available databases[19–21]. Process analytical technologies (PATs) based on near-infrared spectroscopy (NIRS) fingerprinting are becoming increasingly available to accurately identify feed ingredients according to their nutritional content in rapid, non-destructive and cost-effective ways[16,22–24]. NIRS-based information is often available in (near) real time, which allows for controlling and improving the ongoing process to enable quality assurance even during processing[24,25]. NIRS fingerprints, however, provide more than just nutritional value information; they can also be used to predict several parameters, such as geographical origin[26], that are closely related to the environmental footprint of the ingredients. However, such a link between NIRS and quantifying, controlling and improving the environmental sustainability of feed production has not yet been explored.

In this study, we show how NIRS and LCA can be combined to integrate environmental impacts in a transparent optimization framework that can be used to formulate feed mixtures that meet the desired livestock feed quality while minimizing monetary and environmental costs. Predictive machine learning models can be used to link chemical information to relevant feed ingredients properties, such as ingredient-specific nutritional compositions[17,27,28] and geographical origin[26,29,30]. Having both quality information and origin as intrinsic properties of the feed ingredients allows us to optimize the performance of industrial feed production in economic, environmental and quality terms in real time.

The focus of this study is on feed optimization for pigs and broilers; these feeds are among the top feeds produced in Europe[31]. Optimizing feeds for price and environmental footprint requires three main steps (**Fig. 1**). First, we predict the environmental footprints of feed ingredients by combining multivariate classification and global spatially explicit LCA. In classification, NIRS fingerprints are used to predict the country of origin of the ingredients. These predictions are the basis on which the LCA determines the environmental footprints of crop production. We focus on land stress (i.e., the effect of land occupation and transformation) and climate change, as they are the two key environmental impacts of agricultural production[32,33]. In our study, the geographical origin is directly linked to the location of the factory where the spectra were measured. This specification is included both in the classification model and in the LCA analysis by considering the transport from the country of origin to the factory location and, if needed, processing into miscible feed ingredients. Second, we predict accurate nutritional compositions of feed ingredients from NIRS fingerprints by using multivariate regression. By incorporating the nutritional composition, environmental footprint, and price of feed ingredients into a multi-objective optimization, we obtain quality-compliant mixture ratios that minimize the environmental and monetary costs of feed production in real time. The optimization step produces Pareto fronts that reveal the effect of nutritional variation on the final feed costs. The technique for order preference by similarity to ideal solution (TOPSIS)[34] then provides trade-off mixture ratios for each front that leverage optimality for the environmental footprint and monetary price under the constraint of quality compliance.

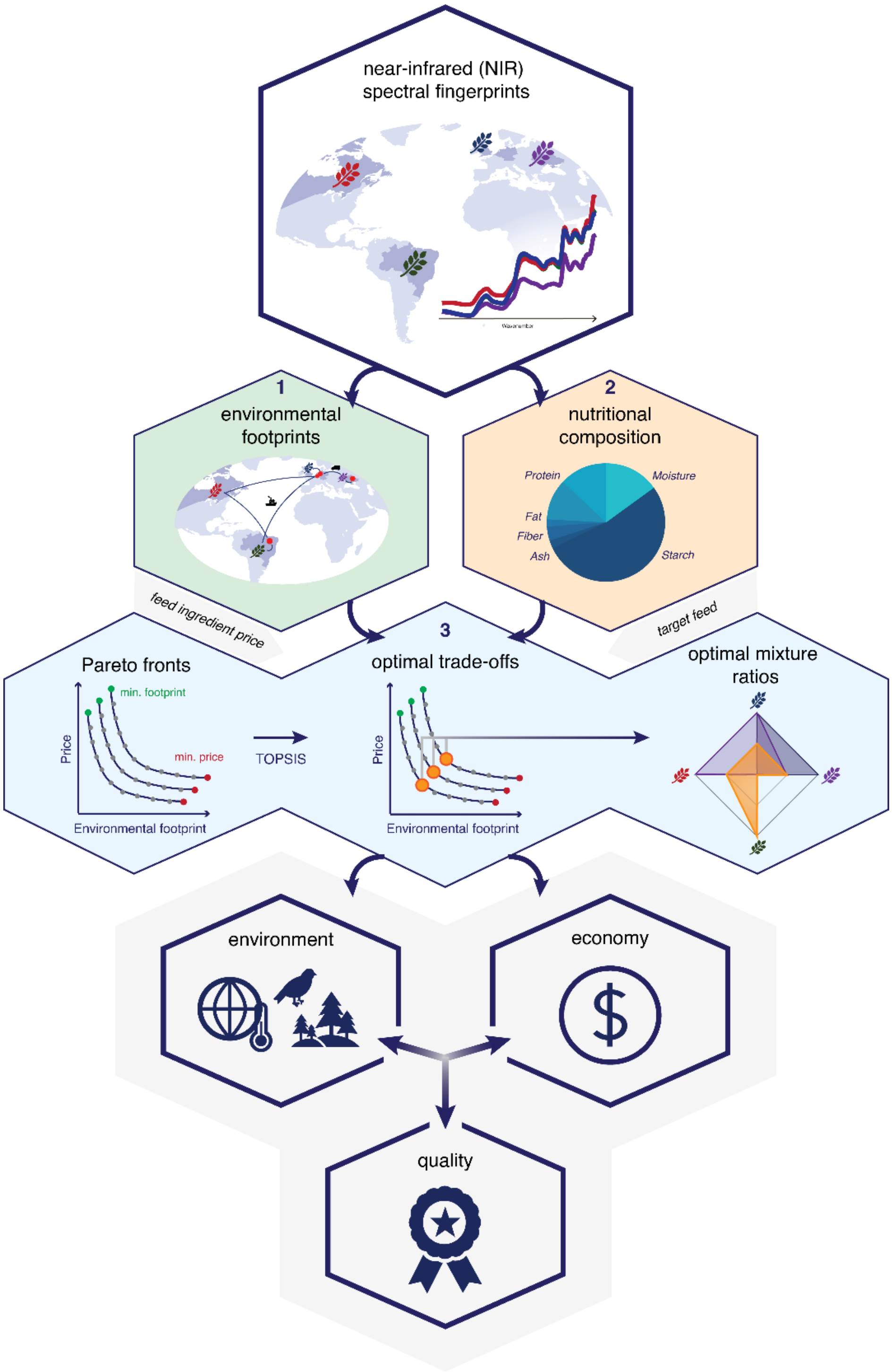

Fig. 1. Workflow of the approach presented in this study. The approach uses the near-infrared (NIR) spectra of eight feed ingredients used in feed production, which were harvested in six different countries of origin and whose spectra were measured at four factory locations. After measurement, the feed ingredients were transported to a common production location where they were mixed into feed. The exact production location was unknown and hence excluded from the study. NIR spectra were assigned to 18 different classes, characterized by feed ingredient, country of origin and measurement location. We predicted environmental footprints by combining multivariate classification with LCA and nutritional content via multivariate regression of each class of feed ingredients from the spectra. The predicted nutritional content was used for 1000 simulations with varying nutritional compositions for each class of feed ingredients. These simulations, together with the predicted environmental footprints, the target feed, and the price of feed ingredients, comprising commodity price[77] and transport cost, became the input for the multi-objective optimization framework. This framework aims to find, for each simulation, the trade-off mixture ratio that minimizes environmental footprints and monetary costs while meeting the quality standards. The figure was created using an existing world map[92].

Results and discussion

The use of NIRS fingerprints for consistent feed quality and authenticity

The geographical origin and nutritional compositions of feed ingredients determined from NIRS fingerprints provide the basis for (near) real-time quality and authenticity compliance within feed optimization. Accurately predicting such geographical origins ensures transparent estimations of the environmental footprint of feed ingredients, which are then readily available for feed optimization. NIRS fingerprints alone were able to discriminate feed ingredient samples according to their country of origin with high prediction accuracy (**Table S7** and **Fig. S5,** balanced accuracy = 0.94). These predictions were successfully linked to ingredient- and country-specific environmental footprints, i.e., land stress and climate change (**Figs. S6-S7**). NIRS fingerprints could also predict nutritional variations among and within ingredient groups with generally high accuracy (**Fig. S9**, root mean squared error = 1.7–5.5 g/kg), with varying performances depending on the ingredient and nutrient analyzed and the sample size (**Figs. S10-S15**). These predictions served as the basis for thousands of simulations that captured the nutritional variability among and within feed ingredients during feed optimization.

Including nutritional variability among and within feed ingredients in multi-objective optimization leads to mixture ratios that always meet the quality constraints within each Pareto front, for each target feed and environmental indicator (**Fig. 2**). This is essential because using occasional off-line measurements as an indicator of the average ingredient nutritional composition may render the produced feed unsuitable for meeting the animal's nutritional requirements. For example, a mixture ratio optimized from accurate nutritional compositions of feed ingredients was compared with that optimized from off-line measurements (**Fig. 2**). The two optimizations selected similar feed ingredients, but in different ratios and from different countries of origin. The extent to which nutritional requirements are not met when using off-line measurements was dependent on the target feeds and environmental indicators. For instance, broilers need more protein and fat than do pigs (**Table S2**); thus, there is a higher preference for soybean than for barley in broiler feed (**Fig. 2**). Off-line determination of ingredient quality generally failed to meet the protein, fat, and starch requirements for pigs or the fat and ash requirements for broilers (**Fig. S17a-d**). The nutritional requirements for pigs were not met for 92% of the simulations, with a median sum of absolute deviations of 8.3 g/kg. For broilers, 74-84% of the simulations did not meet the requirements, with a median sum of absolute deviations of 1.8-3.1 g/kg for land stress and climate change, respectively.

Our findings show that on-line determination of feed ingredient quality is necessary for consistent quality compliance in continuous production. NIRS holds promise for predicting feed compositions that consistently meet the quality in real time and is a viable alternative to more time- and costconsuming traditional methods based on wet chemical analysis. The advantages of NIRS are further enhanced by its ability to readily authenticate the origin of feed ingredients. Feed authentication is essential for ensuring correct labeling and safety in production[35] and for increasing transparency, traceability and accountability throughout the supply chain[36]. Accurate origin determination ultimately ensures transparent environmental assessment during production, allowing decision-makers to include environmental considerations in feed optimization.

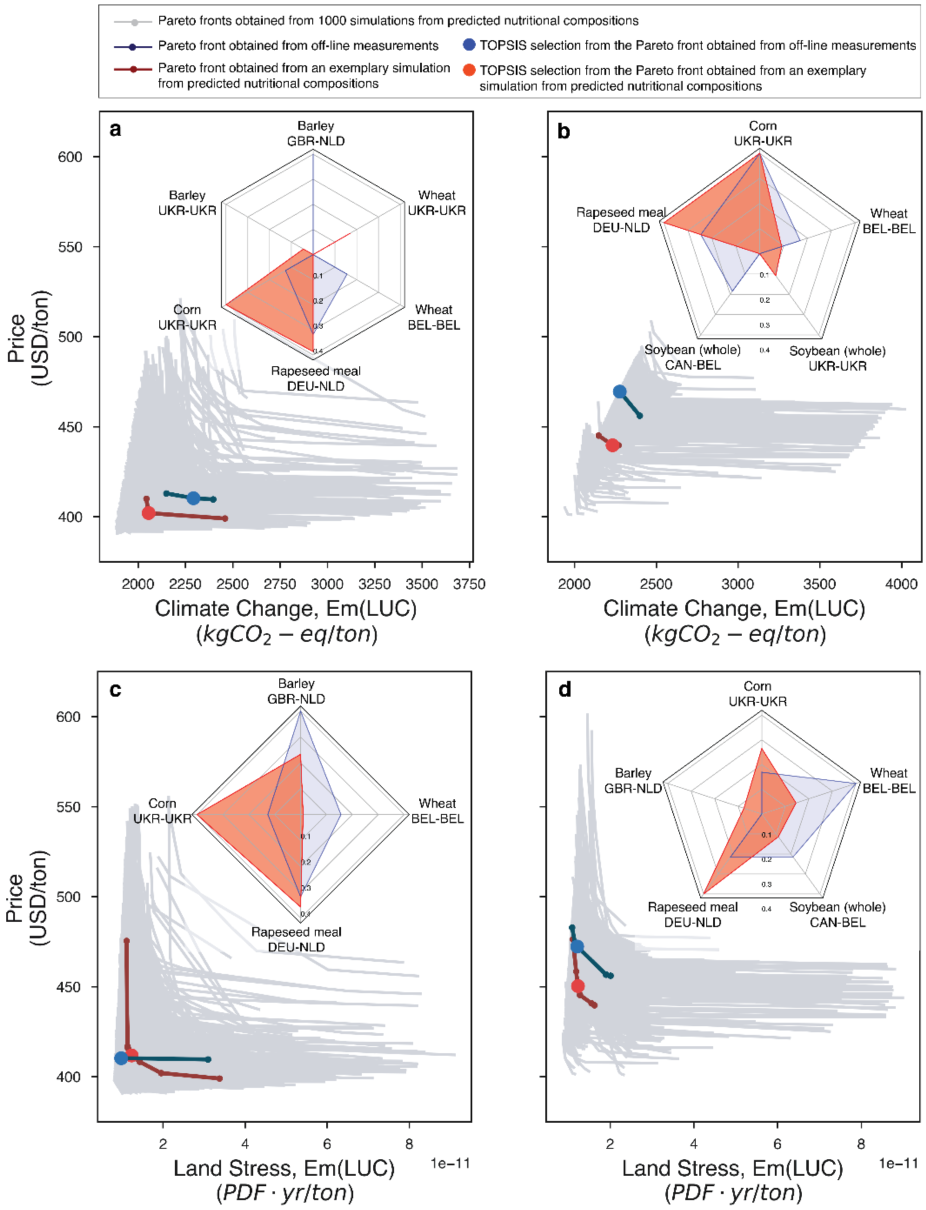

Fig. 2. Optimized monetary and environmental costs considering climate change and land stress with land-use change emissions. The gray lines show the stochastic Pareto fronts resulting from the optimization of the 1000 simulations built from predicted nutritional compositions of feed ingredients. From each Pareto front, a trade-off mixture ratio is selected with TOPSIS. The red line highlights an exemplary Pareto front, with the marked dot indicating the trade-off mixture ratio selected with TOPSIS. The blue line indicates the Pareto front obtained by optimizing the nutritional composition from off-line measurements, with the marked dot indicating the trade-off mixture ratio selected with TOPSIS. The right upper corner of each panel displays the mixture ratios corresponding to these trade-offs, where the labels indicate the selected feed ingredient, country of origin and measurement location. The results are for **a)** pig feed, climate change; **b)** broiler feed, climate change; **c)** pig feed, land stress; and **d)** broiler feed, land stress.

Influence of environmental footprints on feed ingredient mixture ratios

Considering different environmental footprints in feed optimization is crucial for comprehensively evaluating the environmental impact of the produced feed. Our analysis revealed that the choice of environmental indicator, *e.g.*, the impact of land stress on biodiversity or climate change, results in the selection of different feed ingredients from distinct countries of origin (**Fig. 2**). Specifically, the impacts of land stress on biodiversity are decisive for origin selection, particularly for barley and wheat for pig feed (**Fig. 2a,c**), and for soybean for broiler feed (**Fig. 2b,d**). While barley from Great Britain was primarily chosen for optimizing pig feed when considering land stress, barley from Ukraine was also selected as a viable option when considering climate change. The climate change impacts for barley are similar for both countries; however, the impacts of land stress on biodiversity are roughly seven times greater for barley from Ukraine than for that from Great Britain (**Figs. S6-S7**). For broiler feed, soybean was selected more often from Canada when optimizing for land stress impacts due to the roughly two times greater impacts of land stress on biodiversity for Ukrainian soybean.

Remarkably, the availability of environmental impact information during optimization allows feed ingredients to be selected from those with similar prices and nutritional compositions, while ensuring the lowest environmental footprint for the considered feed. For instance, in our framework, corn from Brazil was rarely selected due to both the associated high impact of land stress on biodiversity and climate change (**Fig. S18**). Despite having similar prices, the footprint for land stress is nine times greater and for climate change is three times greater for corn harvested in Brazil than for that harvested in Ukraine. Hence, due to its lower environmental impact, Ukrainian corn is a more profitable choice for both environmental and monetary costs than Brazilian corn (**Fig. S18**).

Our findings show the importance of including environmental impacts in selecting feed ingredients for production and suggest that various footprints should be considered to avoid burden shifting, *e.g.*, when a mixture ratio with low climate change but high land stress impacts is selected. For a more comprehensive evaluation, uncertainty in the footprint calculations may also be considered. For example, the effect of including the loss of carbon in agricultural land compared to natural vegetation was also evaluated (**Figs. S16-S17, S19-S20**). Including carbon loss from land-use change resulted in feed ingredients being selected more often from certain origins for climate change (*e.g.*, rapeseed meal from Germany instead of Ukraine). However, when optimizing for land stress, this inclusion did not result in substantially different ingredient selection (**Figs. S18-S19**).

Including environmental costs in feed optimization to increase the environmental sustainability of industrial production

The extent to which footprint reductions are possible by considering environmental costs during optimization can be understood by evaluating the obtained trade-off mixture ratios against feed ingredient mixtures that minimize only the feed price within each Pareto front (**Fig. 1**). **Fig. 3** shows that accounting for trade-offs between environmental and monetary costs results in relatively large reductions in the environmental footprint at only marginally increased feed prices. The degree to which this occurs varies depending on the target feed and environmental indicator.

The largest footprint reductions were observed when optimizing for the impacts of land stress on biodiversity loss, with median reductions of 39% and 34% against a median increased price of 1.1% and 2.4% for pig and broiler feed, respectively (**Fig. 3c-d**). Optimizing for climate change resulted in lower median environmental reductions of 5.7% and 3.3%, with lower median price increases of 0.82% and 0.92% for pigs and broilers, respectively (**Fig. 3a-b**). Compared with climate change, land stress impacts showed more variance within and among the countries of origin (**Figs. S6-S7**). For this reason, greater potential footprint reductions are expected with land stress than with climate change. The optimization of pig feed resulted in a larger footprint reduction compared to that of broiler feed. This may be attributed to the higher protein and fat requirements for broilers than for pigs (**Table S2**). Quality-compliant broiler feed requires selecting ingredients such as soybean, which has a greater environmental impact than barley or corn, which are more often selected for pigs (**Fig. S18**).

Our findings reveal that factoring environmental costs into optimization is essential for increasing the environmental sustainability of feed production. Specifically, the large observed reductions in land stress impacts suggest that including this impact category in optimization is crucial for reducing the impact of livestock feed production on biodiversity loss. This finding is remarkable considering that the impacts on biodiversity are often underappreciated and unaccounted for in industrial livestock systems[37], even compared to the more frequently estimated carbon emissions[15]. Making these estimates available through combining NIRS and LCA in feed optimization therefore offers the opportunity for multifaceted environmental value creation in the livestock business model.

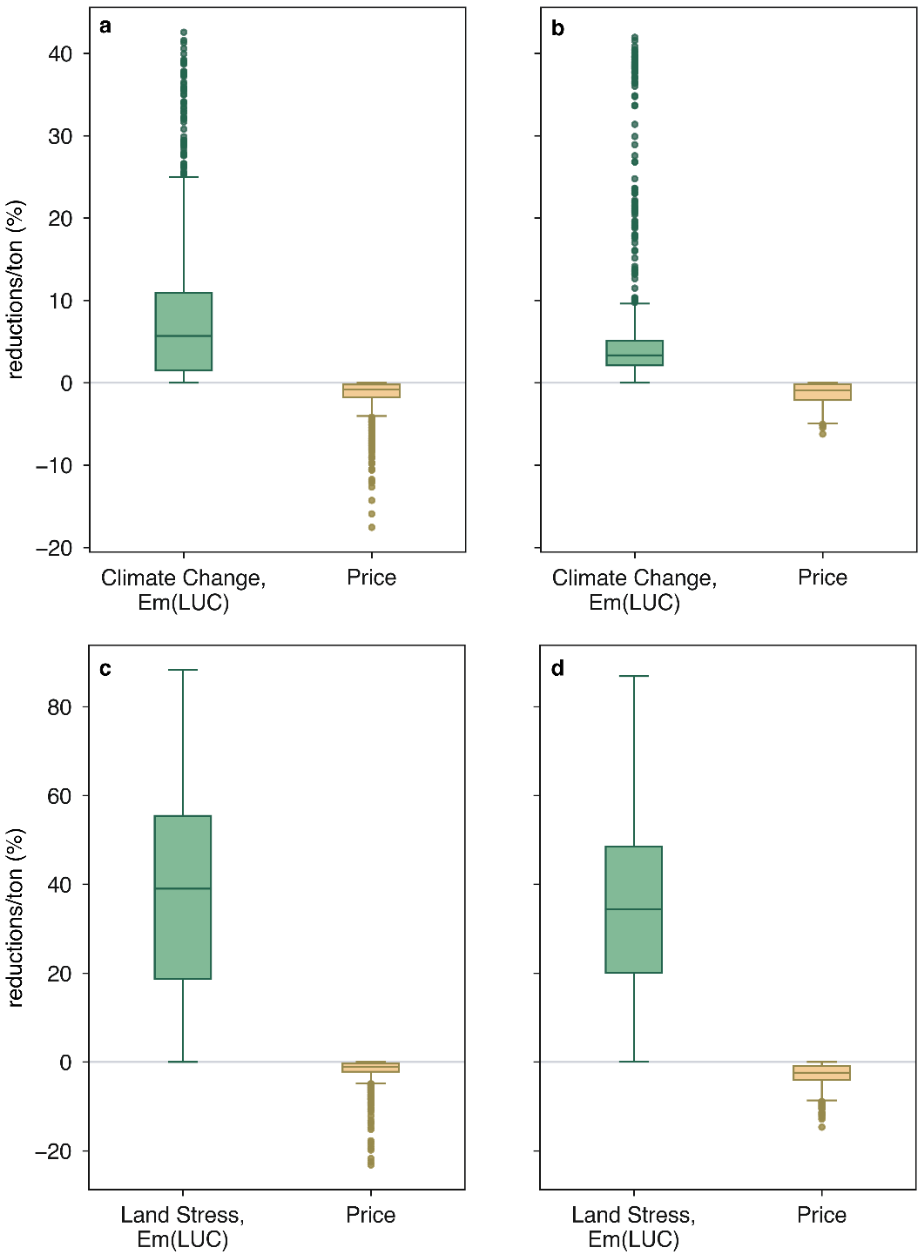

Fig. 3. Footprint reduction and price increase from choosing trade-off mixture ratios, considering climate change and land stress with land-use change emissions. The box and whisker plots illustrate the range of reductions (or increases) resulting from comparing the environmental and monetary costs of the trade-off mixture ratios selected by TOPSIS with the mixture ratio obtained by solely minimizing the feed price. These results are based on the 1000 Pareto fronts shown in **Fig. 2**. The box extends from the lower to upper quartile values of the data, with the line indicating the median reduction (or increase), considered in this study as the point estimate. The whiskers indicate the range of the data, and the dots represent outliers that extend beyond the ends of the whiskers. Footprint reductions and price increases are expressed as relative percentages of the environmental and monetary costs obtained by optimizing the feed price alone, respectively. The results are for: **a)** pig feed, climate change; **b)** broiler feed, climate change; **c)** pig feed, land stress; and **d)** broiler feed, land stress.

Opportunities and challenges related to combining NIRS and LCA to enhance the sustainability of industrial production

Uncovering the *true* impacts of production for decision-makers has recently been advocated by the Food and Agriculture Organization (FAO) of the United Nations as one of the major steps for increasing the sustainability of agri-food systems and realizing the 2030 Agenda of Sustainable Development[38]. By revealing the – as yet hidden – environmental costs during feed optimization, our approach provides decision-makers with an understanding of sustainability as a quantifiable property that can be controlled in real-time production, thus promoting increased environmental awareness and more responsible production patterns. By extracting valuable information on ingredient quality, geographical origin and other relevant attributes, such as ingredient shelf life, chemical hazards, and agronomic practices[17,39,40], NIRS data can be used to control and optimize industrial processes towards increased safety and sustainability with consistent quality. Merging NIRS technology and LCA therefore has high potential for reducing environmental impacts throughout the processing and manufacturing industry, especially when the effects of process control on aspects such as energy usage, pollution and feedstock use are transparent.

Creating industrial value from NIRS fingerprints, however, requires standardized spectroscopic procedures, robust multivariate calibration models, and regular model maintenance[23,41–44]. Incorporating *ad hoc* spectral pre-processing strategies and a sufficiently large number of samples is necessary to cover the high variability observed in industrial processes[43,45] due to measurement changes, seasonal variability and feedstock changes. An insufficient number of samples can increase the possibility of spectral artefacts interfering with NIRS fingerprints[45], thereby reducing the predictive power when modeling certain ingredients and nutrients. In our study, this was noted, *e.g.*, for predicted fiber from soybean (**Fig. S13**). This finding emphasizes the need to include large sample sizes in industrial settings. Such operational expenditures in model building and maintenance are needed to attain the ability to predict valuable process information.

Expanding the current coverage of the study to include more constraints and ingredients, many of which have been studied by NIRS [46–48], is possible with our approach. To create a more diverse and variable ingredient portfolio, the inclusion of the local availability of ingredients at the production site could be added as a model constraint; however, such information was unavailable at the time of analysis. The feed ratios shown in this study are thus possible only if comparable ingredients are available and if they are produced in a sufficiently large amount to meet the demand. Analogously, our approach may better quantify the footprint and thereby further optimize it through trade-off mixture ratios, with greater diversification of ingredient provenances. Data harmonization from different measurement locations does require robust analytical quality control, such that the difference in measurement location can be unambiguously distinguished from the geographical origin for all the feed ingredients. Greater transparency could be achieved through a minor addition in the data collection to integrate more feed ingredients from different origins for each location (*e.g.*, as for sunflower meal, **Table S1, Fig. S5**). Furthermore, increasing the geographical resolution of the LCA from country to region would better include regional agricultural practices involved in crop cultivation, which is particularly relevant for large and heterogeneous countries such as Brazil and Ukraine. Increasing this resolution, however, requires much more transparency in the value chain from field to factory than is currently available.

Combining environmental and monetary goals in feed optimization offers the opportunity to retrospectively identify those ingredients or origins that have never or hardly ever been selected due to their costs and/or nutritional compositions. These spectra can ultimately be used to develop procurement guidelines regarding which ingredients enable environmentally and economically sustainable production and which are so seldomly selected that they may be excluded from purchase. An essential prerequisite for using this approach in procurement is the integration of the variabilities and uncertainties in the international commodities market: ingredient pricing will vary greatly, yet it may be integrated as a source of variability in addition to nutritional quality and geographical origin. This addition would extend the scope of the proposed approach from a process control advisory tool to the procurement stages of the value chain.

Reporting environmental footprints is becoming equally important as part of traditional financial reporting for accessing feed markets in the European Union due to directives such as the Corporate Sustainability Reporting Directive (CSRD)[49]. In the future, large footprint reductions may be further encouraged by initiatives such as true pricing, environmental impact labeling, green public procurement or carbon pricing[50–52]. Environmental impact assessment is, however, a resource-intensive task for every company, especially for small and medium-sized enterprises (SMEs)[53]. Our approach enables the repurposing of the required sustainability data on feed ingredients, which are generally available at the time of processing, for active value creation through feed optimization, thereby closing the gap between real-time operational data and value-driven managerial decisions towards environmentally sustainable choices that are also economically sound.

We proposed a modeling framework that combines NIRS and LCA to improve the environmental sustainability of feed production. Our framework overcomes the drawbacks of seasonal and other variability in agricultural ingredients when designing feeds, as it enables the optimization of mixture ratios for ingredients under real-time variability. Additional goals, such as social sustainability or customer demand, may be further implemented. The approach presented here may be ultimately leveraged for diverse commodities, including food and other biobased commodities, providing a unique opportunity to increase sustainability throughout the agri-food system.

Experimental procedures

The key idea of our approach is to combine NIRS and LCA in an optimization framework to find optimal mixtures of feed ingredients that minimize the environmental and monetary costs of feed production while meeting the quality requirements. **Fig. S2** in the Supplemental Information shows a detailed workflow of the proposed strategy, which consists of three main steps. In the first step, we combine NIRS fingerprints and life impact assessment of feed ingredients in a classification model that allows predicting the ingredient environmental footprints. In the second step, we employ NIRS to predict accurate nutritional compositions of feed ingredients. The predicted information is the input, with the ingredient price, of a multi-objective optimization that aims at finding the optimal ingredient mixture ratios that allow for achieving trade-offs between environmental and monetary costs while meeting the quality standards.

NIRS dataset

The study dataset consists of 863 near-infrared (NIR) spectra of eight different feed ingredients, namely barley, corn, rapeseed meal, soybean expeller, soybean meal, soybean (whole bean), sunflower meal, and wheat, which are employed to obtain two compound feeds: pig feed and broiler feed. For all the spectra, reference nutritional values were obtained with reference methods for wet-chemical quality analysis from accredited laboratories. The considered feed ingredients were harvested from six countries of origin and transported to factories located in four different countries, where the spectra were measured. After measurement, the ingredients were transported to a common production location where they were mixed into feed products. The exact production location was unknown and hence excluded in the study. According to this specification, the NIR spectra belong to 18 classes, characterized by feed ingredient, country of origin and measurement location, as specified in **Table S1** (Supplemental Information). **Fig. S1** shows, as an example, representative NIR spectra for each class.

Chemometric analysis of NIR spectra

Predicting the nutritional composition and environmental footprint from NIR spectra requires employing multivariate chemometric techniques to remove spectral artefacts and enhance the model's predictive accuracy[54,55]. Finding the appropriate techniques enables testing the possibility of employing NIRS in (near) real-time feed optimization. Chemometric prediction aims at maximizing the relationship between the spectral data matrix $\mathbf{X}$ and the response to predict $\mathbf{y}$; this can be achieved by selecting the optimal pre-processing technique that extracts the information in the data matrix $\mathbf{X}$ which is relevant to the response[54–56]. We employed a classification approach to assign each feed ingredient to the respective country of origin, predicting the environmental impact when associated with LCA. In this study, the country of origin was related to the location of the factory where the spectra were measured: this information was included both in the classification model and in the LCA. We employed a regression approach to predict the nutritional composition of feed ingredients, which was measured as a quantitative response (**Table S1**). A preliminary step of data cleaning and train-test splitting was needed to accurately assess model performances, as described in the Supplemental Information.

Pre-processing strategy selection: removing unwanted variation to extract relevant information from NIR spectra

Spectra pre-processing consists of removing unwanted variations and artefacts that hinder relevant information in the raw NIR spectra[56]. Selecting the appropriate pre-processing strategy is crucial to enhance the predictive power of the chemometric model; however, this procedure may be time-consuming and subjective[54,56]. We therefore adopted a supervised pre-processing selection strategy based on exhaustive search[57], similar to that proposed by Gerretzen et al.[54]. This strategy enables testing selected pre-processing techniques suitable for NIRS[54,56] in combination with selected predictive estimators. We employed similar pre-processing techniques for regression and classification, including baseline correction, multiplicative scatter correction, smoothing, and variable scaling. Details on the selected techniques are provided in the Supplemental Information (**Table S3-S4**). We selected the optimal pre-processing techniques in cross-validation by minimizing the root mean squared error ($RMSE$) in regression and maximizing the weighted balanced accuracy ($wbAcc$) in classification. We evaluated the model predictive ability on the test set considering $RMSE$ for regression, and balanced accuracy ($bAcc$) for classification. We here report a short description of these metrics, referring the reader to the Supplemental Information for a more detailed explanation.

The $RMSE$ is defined as (**equation (1)**):

$$RMSE = \sqrt{\frac{\sum_{i=0}^{I}(y_i - \hat{y}_i)}{I}} \quad \textbf{(1)}$$

where $\mathbf{y}$ and $\hat{\mathbf{y}}$ are the vectors holding the observed and calculated nutritional values, respectively, and I the number of samples. Low $RMSE$ values indicate that the model is statistically appropriate to predict the nutritional composition of feed ingredients.

$bAcc$ allows for the assessment of classification performances accounting for class imbalance and is defined as the arithmetic mean of the class sensitivity Sn_g (**equation (2)**)[58]:

$$bAcc = \frac{\sum_{g=1}^{G} Sn_g}{G} \quad \textbf{(2)}$$

where the class sensitivity Sn_g is the ability of the classifier to correctly identify the samples of the $g-th$ class (**equation (3)**), and G is the total number of classes[58]:

$$Sn_g = \frac{c_{gg}}{n_g},\ where\ n_g = \sum_{k=1}^{G} c_{gk}. \quad \textbf{(3)}$$

In **equation (3)**, c_{gg} indicates the number of samples correctly classified. To select the most accurate model accounting for the environmental footprint of misclassification, we used a weighted version of balanced accuracy ($wbAcc$), defined as the average of the weighted sensitivity (wSn_g):

$$wbACC = \frac{\sum_{g=1}^{G} wSn_g}{G} \quad \textbf{(4)}$$

where the weighted sensitivity is defined as:

$$wSn_g = \frac{w_g * c_{gg}}{\sum_{k=1}^{G} c_{gk} * e_{gk} * w_g} \quad \textbf{(5)}$$

where c_{gk} indicates the number of samples belonging to class g and predicted to be in class k, w_g is the class weight to correct for class imbalance (defined in **equation (3)** of the Supplemental Information), and e_{gk} is the environmental footprint associated with predicting a sample belonging to the class g to the class k, computed as the absolute difference in environmental footprint among the classes. A weighted balanced accuracy and balanced accuracy close to or equal 1 indicate that the model correctly assigns the samples to their actual class.

Predicting environmental footprints from NIRS fingerprints and LCA

Multivariate classification

We developed a classification approach to correctly predict the environmental footprints of feed ingredients while penalizing the misclassification of the classes with the highest environmental footprint. We considered linear discriminant analysis (LDA) as a classifier, which is a well-established method in chemometrics to analyze spectral data[59,60]. We computed a classification model for each considered environmental indicator (i.e., climate change and land stress with and without including the effect of land-use change, defined in **equations (7), (10)** and in **equations (9), (12)** in the Supplemental Information), to discriminate feed ingredients coming from different countries of origin and measured in different locations. Accurate predictions (i.e. $bAcc$ close to 1) indicate that the models can be associated with life cycle impact assessment to predict the environmental footprint of feed ingredients.

Life cycle assessment

The environmental impact, expressed as land stress impacts on biodiversity (potentially disappeared fraction year, PDF·yr) and climate change (kg CO_2-eq), was calculated with life cycle assessment. For each indicator, two scenarios were calculated; with and without carbon stock loss due to land-use change (LUC). The carbon stock loss includes the initial carbon loss when land is transformed into agricultural land and the lost sequestration capacity of agricultural land compared to natural vegetation (i.e. foregone sequestration)[61]. These two scenarios were investigated because it was not known from the used database for how long agricultural land already existed and hence how much of this transformation effect should be attributed to crop production. The scenario without LUC emissions represents the situation where the area was already used as agricultural land and hence the effect of transforming land from agricultural to agricultural is negligible. The scenario with LUC emissions represents the situation where natural vegetation is transformed into agricultural land with an evaluation period of 30 years reflecting a typical plantation lifetime[61] and is used in the main results of the article. The formulas used to calculate the impact excluding land-use change can be found in the Supplemental Information.

Climate change

The impact on climate change including land-use change was calculated by **equations (6)-(7).**

$$CC_{LUC,g} = CF_{CC} * Em_{SC,g} + \frac{\sum_q^{Q_g} Em_{LUC,h,q}}{Q_g} \quad \textbf{(6)}$$

$$\mathbf{f}_{CC,LUC} = [CC_{LUC,1}, \dots, CC_{LUC,G}] \quad \textbf{(7)}$$

where $CC_{LUC,g}$ represents the total impact on climate change including LUC emissions for each class g (in kg CO_2-eq./ton feed ingredient), which is specified by feed ingredient h, grown in origin country o and transported to the location country l where the NIR spectra were measured. CF_{CC} represents the climate change characterization factor used to express the emissions in kg CO_2-eq. $Em_{SC,g}$ are the emissions from the supply chain (in ton feed ingredient) for each feed ingredient h, country of origin o and transport to location l, belonging to class g. The supply chain includes the material and energy requirements for agricultural practices and processing into animal feed ingredient and transport to the measurement location. $Em_{LUC,h,q}$ are the land-use change emissions over a period of 30 years, for each feed ingredient h and location in 30x30 arcminute[61] raster q. Q_g represents the maximum grid level for class g. $\mathbf{f}_{CC,LUC}$ is the vector containing the average impact on climate change in kg CO_2-eq/ton feed ingredient including LUC emissions for each class g. G is the total number of classes.

Supply chain emissions, including crop cultivation, harvesting and pre-processing into feed ingredients and transport to production location were based on background processes from Agri-footprint v6[62] and processed in SimaPro 9.4.0.2[63]. The processes were adjusted by removing the land-use, to avoid double counting with LUC emissions. The impact for climate change was calculated with ReCiPe 2016, midpoint, H[64]. The geographical resolution of Agri-footprint processes is per country, in line with the geographical resolution of the NIR spectra. The economic allocation of side products obtained during pre-processing was based on the economic allocation used in Agri-footprint v6.

Land-use change (LUC) emissions from changing carbon stocks were estimated based on the LPJml global vegetation and hydrological model[61,65,66], coupled with the IMAGE integrated assessment model[67]. Following the approach used by Hanssen et al.[61], carbon stocks after 30 years of growing feed crops were compared to carbon stocks under a simulated counterfactual of natural vegetation growth in the same location. The difference in carbon stocks was assumed to be emitted to the atmosphere as CO_2. These emissions were allocated to the cumulative feed crop production over 30 years, which was determined per location using crop yield data in MapSpam with a 5-minute resolution for 2010[68,69], as shown in **equation (8)**.

$$Em_{LUC,h,q} = \frac{(\Delta C_{h,q\in 0,30}) * r}{Y_{h,q\in o} * t} \quad \textbf{(8)}$$

where $Em_{LUC,i,q}$ are the LUC emissions of feed crop production (in kg CO_2-eq./ton feed ingredient) for each feed ingredient h in gridcel q. ΔC is the difference between carbon stocks under feed crop cultivation and natural vegetation (in tonne C) for crop h and origin country o; r is the molar mass ratio between CO_2 and C of 44.01/12.01. Y is the feed crop yield (in ton feed ingredient/year); and t is the 30-year time period considered (in years)[69,70].

Land stress

The impact on land stress, including land-use change, is the combined effect of land occupation and transformation and was calculated by **equations (9)-(10).**

$$LS_{LUC,g} = CF_{ED} * Em_{transport,g} + \frac{\sum_z^{Q_g} \frac{CF_{occ,z\in o} + CF_{trans,z\in o}}{yield_{h,o}}}{Q_g} \quad \textbf{(9)}$$

$$\mathbf{f}_{LS,LUC} = [LS_{LUC,1}, \dots, LS_{LUC,G}] \quad \textbf{(10)}$$

Where $LS_{LUC,g}$ represents the total impact on land stress including LUC emissions for each class g (in PDF·yr per ton feed ingredient), which is specified by feed ingredient h, grown in origin country o and transported to location country l. CF_{ED} represents the ecosystem damage characterization factor used to express the emissions from transport to from country of origin o to measurement location l ($Em_{transport,g}$) in PDF·yr. $CF_{occ,z}$ and $CF_{trans,z}$ are the ecosystem specific characterization factors for land occupation and land transformation (in PDF·yr/m^2) for each ecoregion z, respectively. $\mathbf{f}_{LS,LUC}$ is the vector containing the average impact on land stress including LUC emissions for each class g (in PDF·yr/ton feed ingredient). The yield was from MapSpam and the characterization factors for occupation and transformation were from Chaudhary et al.[70]. These characterization factors are based on the different ecoregions across the world[71]. For each country of origin, the ecoregions were identified together with their corresponding characterization factor. Depending on the areas where agricultural practices took place, based on yield, the effect of land occupation and transformation was calculated per PDF·yr/ton feed ingredient. A time of 30 years and economic allocation to by-products was used to calculate the impact on land stress per ton feed ingredient, as was also done with climate change. All generated maps were added in R, using the lowest map resolution (i.e., on grid cell level q).

Predicting the nutritional composition of feed ingredients from NIRS fingerprints

To predict the nutritional composition of feed ingredients, we compared the performances of two different regressors: partial least squares (PLS) and random forest regression. PLS regression is commonly used for NIR spectra analyses due to its ability to handle numerous correlated spectral features[72]. PLS identifies a set of new variables (latent variables, LVs), and finds the LVs' direction that explains the highest variance in the $\mathbf{X}$ matrix and is most correlated to the response vector $\mathbf{y}$[73]. We selected the number of LVs employed by the model with internal cross-validation to avoid overfitting. Random forest[74] regression has been recently demonstrated to be a powerful technique in multivariate calibration to deal with spectral complexity and possible non-linearity[75,76]. Therefore, we also tested this estimator to evaluate whether the final predictive accuracy would have been improved compared to the most commonly used PLS. Training a random forest model required tuning the model hyperparameters. We selected the optimal hyperparameters with genetic algorithms in cross-validation. PLS and random forest regression models were run independently for each feed ingredient and nutritional value, within the pre-processing optimization framework. Their performance was evaluated in cross-validation in combination with the tested pre-processing strategies: the combination with the lowest $RMSE$ in cross-validation was selected as the most optimal to predict the ingredients' nutritional composition. Details about the optimized hyperparameters and cross-validation schemes are provided in the Supplemental Information (**Table S5-S6**).

Multi-objective stochastic optimization of animal feed

We employed multi-objective optimization to find the optimal mixture ratio of feed ingredients that minimizes the environmental and monetary costs of animal feed production while meeting the feed quality requirements. We employed the weighted sum method to deal with the multi-objective problem as a single-objective linear optimization. The objective function is the sum of environmental footprint ($\mathbf{f}^{\mathrm{T}}\mathbf{x}$) and monetary price ($\mathbf{p}^{\mathrm{T}}\mathbf{x}$), multiplied by their weighting coefficients w_f and w_p (**equation (11)**).

$$\min\ w_f * \mathbf{f}^{\mathrm{T}}\mathbf{x} + w_p * \ \mathbf{p}^{\mathrm{T}}\mathbf{x} \tag{11}$$

$subject\ to$:

$$\sum_{g=1}^{G} x_g = 1$$

$$0 \leq \sum_{i=1}^{N_{crop}} x_i \leq \rho$$

$$\mathbf{b}^{\mathbf{lb}} \leq \mathbf{Ax} \leq \mathbf{b}^{\mathbf{ub}}$$

$$w_f + w_p = 1$$

$$w_f \geq 0,\ w_p \geq 0$$

with G optimization classes, M nutrients, and N_{crop} classes for each feed ingredient group. $\mathbf{x} \in \mathbb{R}^{\mathrm{G}\times 1}$is the decision vector to be optimized, holding the feed ingredient mixture ratio, $\mathbf{f} \in \mathbb{R}^{\mathrm{G}\times 1}$ holds the environmental footprint and $\mathbf{p} \in \mathbb{R}^{\mathrm{G}\times 1}$ the monetary price for each optimization class. The constraints defined in **equation (11)** ensure that the optimized mixture ratio meets the quality of the target feed: $\mathbf{A} \in \mathbb{R}^{\mathrm{M}\times \mathrm{G}}$ holds the nutritional composition of each optimization class, $\mathbf{b}^{\mathbf{lb}} \in \mathbb{R}^{\mathrm{M}\times 1}$ and $\mathbf{b}^{\mathbf{ub}} \in \mathbb{R}^{\mathrm{M}\times 1}$ hold the nutritional composition of the considered target feed, including the maximum allowed variation from the nominal value on the lower ($\mathbf{b}^{\mathbf{lb}}$) and upper ($\mathbf{b}^{\mathbf{ub}}$) boundaries. Nominal values and maximum allowed variation are reported in **Table S2** of the Supplemental Information. A maximum ρ value was applied to restrict the ratio of each optimization class in our study, thereby ensuring the diversity of ingredients in the feed.

Objective function

To estimate $\mathbf{f}$, we considered climate change and land stress in separate models as environmental indicators to minimize the environmental cost, with and without including the effect of land-use change, as defined in **equations (7), (10)** and in **equations (9), (12)** in the Supplemental Information. To estimate $\mathbf{p}$, we retrieved the feed ingredients prices from the international market with the Food Prices Monitoring and Analysis (FPMA) Tool[77] of the Food and Agriculture Organization (FAO) of the United Nations. We retrieved commodity prices for one selected month, March 2022, which was the most recent month at the time of the analysis. Prices were not available for each feed ingredient and country of origin considered in our study: Supplemental Information reports detailed information on all considered assumptions for price estimation (**equations (20)-(21)**, Supplemental Information).

Implementation of predicted nutritional compositions into optimization

Constraints in multi-objective optimization (**equation (11)**, **equations (17)-(19)** in the Supplemental Information) ensure that the optimized feed ingredient mixture ratios meet the nutritional constraint of the target feed. Predicting accurate nutritional compositions from NIRS fingerprints entails having diverse compositions due to nutritional variation. In our dataset, the number of samples for each feed ingredient class was too small to reproduce all the nutritional variability that might be observed in industrial production. Hence, we employed Monte Carlo sampling to obtain 1000 simulations with varying nutritional compositions among and within feed ingredients groups. Monte Carlo is a sampling-based methodology[78] that allows us to generate random scenarios based on the probability distributions of the predicted nutrients to observe the effect of nutritional variability on the final optimization. This methodology is extensively used in stochastic optimization in quantitative applications such as in science, engineering and economics, to optimize the process performances explicitly accounting for uncertainty[78,79]. We utilized the predicted nutritional compositions for each feed ingredient to build multivariate t-distributions to sample from. We considered multivariate t-distributions to account for the correlation among nutrients, ensuring that the simulated formulations respect nutritional constraints. A random error, derived from univariate normal distributions of $RMSE$ values, was finally added to the sampled nutritional values to account for the model error in prediction. The optimization model was run independently for each simulation in a stochastic optimization framework.

Selecting optimal trade-off mixture ratios

Varying the weights w_f and w_p for each simulation allows for obtaining stochastic Pareto fronts of feasible mixture ratios. To select the optimal trade-off mixture ratio, we employed the technique for order preference by similarity to ideal solution (TOPSIS)[34], weighting the objectives with the Shannon's entropy method. TOPSIS selects the optimal trade-off by finding alternatives with the shortest distance from the positive ideal solution and with the longest distance from the negative ideal solution[80,81]. TOPSIS is widely used in many research areas, such as supply chain management, manufacturing systems, or energy management, for its simplicity in concept and application, and for being able to find trade-off solutions among several objectives[80–82].

Comparison with off-line measurements

We performed a separate multi-objective optimization (**equation (11)**) to evaluate quality deviations resulting from utilizing nutritional values from occasional off-line measurements. To compensate for the fact that these values were not available in our study, we considered average compositions of feed ingredients from available measurements. Nutritional compositions for each of the 863 samples measured with NIRS were available from chemical wet analyses: we estimated average nutritional values for each ingredient to replicate the situation where measurements are not performed for each of the incoming ingredient batches. Since this optimization does not consider nutritional variability within feed ingredient groups, only one Pareto front is obtained; we selected the optimal trade-off mixture ratio on this Pareto front with TOPSIS. We used the selected mixture ratio to calculate the deviation in quality considering off-line measurements for each of the 1000 simulations, as detailed in the Supplemental Information (**equation (22)**, Supplemental Information).

Evaluation of footprint and price reductions from trade-off mixture ratios

We evaluated footprint and price reductions by comparing the trade-off mixture ratios selected with multi-objective optimization against conventional least-cost feed optimization practices (i.e. when w_f = 0 and $w_p = 1$ in **equation (11)**). For each of the 1000 quality-compliant simulations, we estimated percentage reductions as:

$$footprint\ reduction\ (\%) = \frac{\mathbf{f}^{\mathrm{T}}\mathbf{x}_{min\ price} - \mathbf{f}^{\mathrm{T}}\mathbf{x}_{trade-off}}{\mathbf{f}^{\mathrm{T}}\mathbf{x}_{min\ price}} * 100 \qquad \textbf{(12)}$$

$$price\ reduction\ (\%) = \frac{\mathbf{p}^{\mathrm{T}}\mathbf{x}_{min\ price} - \mathbf{p}^{\mathrm{T}}\mathbf{x}_{trade-off}}{\mathbf{p}^{\mathrm{T}}\mathbf{x}_{min\ price}} * 100 \qquad \textbf{(13)}$$

where $\mathbf{x}_{min\ price}$ and $\mathbf{x}_{trade-off}$ are the decision vectors obtained with the single-objective optimization of price and the TOPSIS selection, respectively. Note that for price negative reductions are expected, which consequently correspond to price increases. We report median values as point estimates to facilitate the comparison among different environmental indicators and target feeds.

Software

Python v.3.9.7 was employed to develop the classification, regression, and multi-objective optimization models. The following Python packages were used: sklearn[83], pandas[84], numpy[85], chemsy[57] for supervised pre-processing optimization, ILOG CPLEX Optimization Studio v.22.1.0[86]

(Python API) for multi-objective optimization, sklearn-genetic-opt[87] for hyperparameter optimization for random forest, pymcdm[88] for TOPSIS. Matplotlib[89] and seaborn[90] libraries were employed for visualization. SimaPro 9.4.0.2[63]with impact assessment method ReCiPe2016 (H) was used to calculate the environmental impact of the supply chain of the feed classes. Rstudio (2022.02.2+485) was used to calculate the total impact of the classes per grid cell in the country of origin, using the following packages: Terra, sp, readxl and sf.

Data and code availability

The data used to generate the results reported in this study, i.e. predictions from NIR spectra and LCA data, will be available with the codes supporting the analysis in a public repository[91]. The raw NIR spectra data will be available upon request to the lead author with permission from NutriControl.

Acknowledgments

The authors thank dr. ir. Jack Peerlings for the fruitful discussion on the research, and the members of the ISPT "Measurement 4 Management" consortium for their financial and in-kind contribution. This consortium consists of the following organizations: DSM, ISPT, Kraft Heinz, Magion, Nexperia, Nouryon, Nutricontrol, Radboud University, RIWA Rijn, Unilever. This project received funding from TKI E&I with the supplementary grant 'TKI-Toeslag' for Topconsortia for Knowledge and Innovation (TKI's) of the Ministry of Economic Affairs and Climate Policy. The manuscript was edited for proper English language, grammar, punctuation, spelling, and overall style by one or more of the highly qualified native English-speaking editors at Springer Nature Author Services.

Author contributions

M.C.: conceptualization, methodology, software, validation, formal analysis, investigation, data curation, writing-original draft, visualization; **A.O.**: conceptualization, methodology, software, validation, formal analysis, investigation, data curation, writing-original draft; **S.Y.T.**: methodology, software, writing-review & editing; **S.H.**: methodology, writing-review & editing; **M.S.**: methodology, resources, writing-review & editing, funding acquisition, project administration; **C.K.:** resources, funding acquisition, project administration; **R.v.Z.**: methodology, supervision, writing-review & editing, funding acquisition, project administration; **L.B.:** funding acquisition, project administration, **M.H.:** conceptualization, methodology, supervision, writing-review & editing, funding acquisition, project administration; **J.J.:** conceptualization, methodology, supervision, writing-review & editing, funding acquisition, project administration. **M.C.** and **A.O.** contributed equally as first authors of the manuscript. **J.J** and **M.H.** jointly supervised this work.

Declaration of interests

The authors declare no competing interests.

References

1. FAO. Codex Alimentarius. https://www.fao.org/fao-who-codexalimentarius/thematic-areas/animal-feed/en/ (2023).

2. Makkar, H. P. S. Review : Feed demand landscape and implications of food-not feed strategy for food security and climate change. *Animal, The International Journal of Animal Biosciences* **12**, 1744–1754 (2018).

3. European Commission. Commission adopts antitrust Guidelines for sustainability agreements in agriculture. https://ec.europa.eu/commission/presscorner/detail/en/ip_23_6370 (2023).

4. Henchion, M., Moloney, A. P., Hyland, J., Zimmermann, J. & McCarthy, S. Review: Trends for meat, milk and egg consumption for the next decades and the role played by livestock systems in the global production of proteins. *Animal* **15**, 100287 (2021).

5. European Green Deal. https://commission.europa.eu/strategy-and-policy/priorities-2019-2024/european-green-deal_en.

6. United Nations. Sustainable Development Goals. https://sdgs.un.org/goals (2023).

7. Kebreab, E., Hansen, A. V. & Strathe, A. B. Animal production for efficient phosphate utilization: From optimized feed to high efficiency livestock. *Curr Opin Biotechnol* **23**, 872–877 (2012).

8. Altun, A. A. & Şahman, M. A. Cost optimization of mixed feeds with the particle swarm optimization method. *Neural Comput Appl* **22**, 383–390 (2013).

9. Uyeh, D. D. *et al.* Precision animal feed formulation: An evolutionary multi-objective approach. *Anim Feed Sci Technol* **256**, 114211 (2019).

10. Castonguay, A. C. *et al.* Navigating sustainability trade-offs in global beef production. *Nat Sustain* **6**, 284–294 (2023).

11. Mallick, P., Muduli, K., Biswal, J. N. & Pumwa, J. Broiler Poultry Feed Cost Optimization Using Linear Programming Technique. *Journal of Operations and Strategic Planning* **3**, 31–57 (2020).

12. Nicholson, C. C., Emery, B. F. & Niles, M. T. Global relationships between crop diversity and nutritional stability. *Nat Commun* **12**, (2021).

13. St.Clair, S. B. & Lynch, J. P. The opening of Pandora's Box: Climate change impacts on soil fertility and crop nutrition in developing countries. *Plant Soil* **335**, 101–115 (2010).

14. de Vries, M. & de Boer, I. J. M. Comparing environmental impacts for livestock products: A review of life cycle assessments. *Livest Sci* **128**, 1–11 (2010).

15. McClelland, S. C., Arndt, C., Gordon, D. R. & Thoma, G. Type and number of environmental impact categories used in livestock life cycle assessment: A systematic review. *Livest Sci* **209**, 39–45 (2018).

16. der Poel, A. F. B. van *et al.* Future directions of animal feed technology research to meet the challenges of a changing world. *Anim Feed Sci Technol* **270**, 1–14 (2020).

17. Garnsworthy, P. C., Wiseman, J. & Fegeros, K. Prediction of chemical, nutritive and agronomic characteristics of wheat by near infrared spectroscopy. *Journal of Agricultural Science* **135**, 409–417 (2000).

18. Givens, D. I., De Boever, J. L. & Deaville, E. R. The principles, practices and some future applications of near infrared spectroscopy for predicting the nutritive value of foods for animals and humans. *Nutr Res Rev* **10**, 83–114 (1997).

19. CVB Veevoedertabel 2011, Productschap Diervoeder. https://edepot.wur.nl/367874 (2011).

20. Tran, H. *et al.* Development of feed composition tables using a statistical screening procedure. *J Dairy Sci* **103**, 3786–3803 (2020).

21. Sauvant, D., Perez, J.-M. & Tran, G. *Tables of Composition and Nutritional Value of Feed Materials: Pigs, Poultry, Cattle, Sheep, Goats, Rabbits, Horses and Fish*. (Wageningen Academic Publishers, 2004). doi:10.3920/978-90-8686-668-7.

22. Karoui, R. & De Baerdemaeker, J. A review of the analytical methods coupled with chemometric tools for the determination of the quality and identity of dairy products. *Food Chem* **102**, 621–640 (2007).

23. Cozzolino, D. The ability of near infrared (NIR) spectroscopy to predict functional properties in foods: Challenges and opportunities. *Molecules* **26**, (2021).

24. Berg, F. Van Den, Christian, B., Sørensen, K. M. & Engelsen, S. B. Process Analytical Technology in the food industry *. *Trends Food Sci Technol* **31**, 27–35 (2013).

25. Vanarase, A. U., Alcalà, M., Jerez Rozo, J. I., Muzzio, F. J. & Romañach, R. J. Real-time monitoring of drug concentration in a continuous powder mixing process using NIR spectroscopy. *Chem Eng Sci* **65**, 5728–5733 (2010).

26. Zhao, H., Guo, B., Wei, Y. & Zhang, B. Near infrared reflectance spectroscopy for determination of the geographical origin of wheat. *Food Chem* **138**, 1902–1907 (2013).

27. Kays, S. E., Barton, F. E. & Windham, W. R. Predicting protein content by near infrared reflectance spectroscopy in diverse cereal food products. *J Near Infrared Spectrosc* **8**, 35–43 (2000).

28. Lin, C. *et al.* Determination of grain protein content by near-infrared spectrometry and multivariate calibration in barley. *Food Chem* **162**, 10–15 (2014).

29. Lee, J. H. *et al.* Rapid Discrimination of the Country Origin of Soybeans Based on FT-NIR Spectroscopy and Data Expansion. *Food Anal Methods* **15**, 3322–3333 (2022).

30. Schütz, D., Riedl, J., Achten, E. & Fischer, M. Fourier-transform near-infrared spectroscopy as a fast screening tool for the verification of the geographical origin of grain maize (Zea mays L.). *Food Control* **136**, 108892 (2022).

31. Europe Animal Feed Market. https://www.feedandadditive.com/europe-animal-feed-market/ (2023).

32. Steinmann, Z. J. N., Schipper, A. M., Hauck, M. & Huijbregts, M. A. J. How Many Environmental Impact Indicators Are Needed in the Evaluation of Product Life Cycles? *Environ Sci Technol* **50**, 3913–3919 (2016).

33. Verones, F. *et al.* LC-IMPACT: A regionalized life cycle damage assessment method. *J Ind Ecol* **24**, 1201–1219 (2020).

34. Hwang, C.-L. & Yoon, K. Basic Concepts and Foundations. in *Multiple Attribute Decision Making: Methods and Applications A State-of-the-Art Survey* 16–57 (Springer Berlin Heidelberg, Berlin, Heidelberg, 1981). doi:10.1007/978-3-642-48318-9_2.

35. Pinotti, L. *et al.* Feed authentication as an essential component of food safety and control. *Outlook Agric* **34**, 243–248 (2005).

36. World Animal Protection. The Hidden Health Impacts of Industrial Livestock Systems. 1–78 https://www.worldanimalprotection.org/sites/default/files/2022-04/HealthImpactsofIndustrialLivestockSystemsFINALWEB.pdf (2022).

37. Naylor, R. *et al.* Losing the links between livestock and land. *Science (1979)* **310**, 1621–1622 (2005).

38. FAO. Revealing the true cost of food to transform agrifood systems. 1–150 https://www.fao.org/3/cc7724en/cc7724en.pdf (2023).

39. Porep, J. U., Kammerer, D. R. & Carle, R. On-line application of near infrared (NIR) spectroscopy in food production. *Trends Food Sci Technol* **46**, 211–230 (2015).

40. Fu, X. & Ying, Y. Food Safety Evaluation Based on Near Infrared Spectroscopy and Imaging: A Review. *Crit Rev Food Sci Nutr* **56**, 1913–1924 (2016).

41. Schaefer, C. *et al.* A Process Analytical Technology (PAT) approach to control a new API manufacturing process: Development, validation and implementation. *Talanta* **120**, 114–125 (2014).

42. Pu, Y. Y., O'Donnell, C., Tobin, J. T. & O'Shea, N. Review of near-infrared spectroscopy as a process analytical technology for real-time product monitoring in dairy processing. *Int Dairy J* **103**, 104623 (2020).

43. Schoot, M. *et al.* Cost-benefit analysis of calibration model maintenance strategies for process monitoring. *Anal Chim Acta* **1180**, 338890 (2021).

44. Wise, B. M. & Roginski, R. T. A calibration model maintenance roadmap. *IFAC-PapersOnLine* **28**, 260–265 (2015).

45. Schoot, M. *et al.* Investigating the need for preprocessing of near-infrared spectroscopic data as a function of sample size. *Chemometrics and Intelligent Laboratory Systems* **204**, 104105 (2020).

46. Li, J. *et al.* Use of near-infrared reflectance spectroscopy for the rapid determination of the digestible energy and metabolizable energy content of corn fed to growing pigs. *J Anim Sci Biotechnol* **7**, 1–9 (2016).

47. Mundhada, S., Chaudhry, M. M. A., Erkinbaev, C. & Paliwal, J. Non-Destructive Quality Monitoring of Flaxseed During Storage. *Journal of Food Measurement and Characterization* **16**, 3640–3650 (2022).

48. Bhatty, R. S. Measurement of oil in whole flaxseed by near-infrared reflectance spectroscopy. *J Am Oil Chem Soc* **68**, 34–38 (1991).

49. European Commission. Corporate Sustainability Reporting. https://finance.ec.europa.eu/capital-markets-union-and-financial-markets/company-reporting-and-auditing/company-reporting/corporate-sustainability-reporting_en (2023).

50. OECD Global Forum on Agriculture 2023. Background note. https://www.oecd.org/agriculture/events/oecd-global-forum-on-agriculture/global-forum-agriculture-2023-background-note.pdf (2023).

51. OECD Environment Directorate. Environmental Labelling and Information Schemes: Policy Perspectives. 24 https://www.oecd.org/env/policy-persectives-environmental-labelling-and-information-schemes.pdf (2016).

52. European Commission. Green Public Procurement. https://green-business.ec.europa.eu/green-public-procurement_en (2023).

53. Ortiz-Martínez, E., Marín-Hernández, S. & Santos-Jaén, J. M. Sustainability, corporate social responsibility, non-financial reporting and company performance: Relationships and mediating effects in Spanish small and medium sized enterprises. *Sustain Prod Consum* **35**, 349–364 (2023).

54. Gerretzen, J. *et al.* Simple and Effective Way for Data Preprocessing Selection Based on Design of Experiments. *Anal Chem* **87**, 12096–12103 (2015).

55. Blanco, M. & Villarroya, I. NIR spectroscopy: A rapid-response analytical tool. *TrAC - Trends in Analytical Chemistry* **21**, 240–250 (2002).

56. Engel, J. *et al.* Breaking with trends in pre-processing? *TrAC - Trends in Analytical Chemistry* **50**, 96–106 (2013).

57. S.Y., Teng., M., Dingemans, M., Cairoli, J., Jansen. tsyet12/Chemsy: Chemsy v1.0b. Zenodo. https://doi.org/10.5281/zenodo.5793315 (2021).

58. Ballabio, D., Grisoni, F. & Todeschini, R. Multivariate comparison of classification performance measures. *Chemometrics and Intelligent Laboratory Systems* **174**, 33–44 (2018).

59. Wu, W. *et al.* Comparison of regularized discriminant analysis, linear discriminant analysis and quadratic discriminant analysis, applied to NIR data. *Anal Chim Acta* **329**, 257–265 (1996).

60. Krzanowski, W. J., Jonathan, P., McCarthy, W. V & Thomas, M. R. Discriminant analysis with singular covariance matrices: methods and applications to spectroscopic data. *J R Stat Soc Ser C Appl Stat* **44**, 101–115 (1995).

61. Hanssen, S. V. *et al.* The climate change mitigation potential of bioenergy with carbon capture and storage. *Nat Clim Chang* **10**, 1023–1029 (2020).

62. Agri-footprint v6. https://www.cabdirect.org/cabdirect/abstract/20153221334.

63. PRé Consultants B.v. SimaPro (v9.4.0.2). (2022).

64. Huijbregts, M. A. J. *et al.* ReCiPe2016: a harmonised life cycle impact assessment method at midpoint and endpoint level. *Int J Life Cycle Assess* **22**, 138–147 (2017).

65. Beringer, T., Lucht, W. & Schaphoff, S. Bioenergy production potential of global biomass plantations under environmental and agricultural constraints. *GCB Bioenergy* **3**, 299–312 (2011).

66. Müller, C. *et al.* Drivers and patterns of land biosphere carbon balance reversal. *Environmental Research Letters* **11**, (2016).

67. Stehfest, E., van Vuuren, D., Bouwman, L. & Kram, T. *Integrated Assessment of Global Environmental Change with IMAGE 3.0: Model Description and Policy Applications*. (Netherlands Environmental Assessment Agency (PBL), 2014).

68. Mapsmap. https://mapspam.info/index.php/methodology/ (2023).

69. Yu, Q. *et al.* A cultivated planet in 2010-Part 2: The global gridded agricultural-production maps. *Earth Syst Sci Data* **12**, 3545–3572 (2020).

70. Chaudhary, A. & Brooks, T. M. Land Use Intensity-Specific Global Characterization Factors to Assess Product Biodiversity Footprints. *Environ Sci Technol* **52**, 5094–5104 (2018).

71. Olson, D. M. *et al.* Terrestrial ecoregions of the world: A new map of life on Earth. *Bioscience* **51**, 933–938 (2001).

72. Wold, S., Sjöström, M. & Eriksson, L. PLS-regression: A basic tool of chemometrics. *Chemometrics and Intelligent Laboratory Systems* **58**, 109–130 (2001).

73. Consonni, V., Baccolo, G., Gosetti, F., Todeschini, R. & Ballabio, D. A MATLAB toolbox for multivariate regression coupled with variable selection. *Chemometrics and Intelligent Laboratory Systems* **213**, 104313 (2021).

74. Breiman, L. Random Forests. *Mach Learn* 5–32 (2001).

75. de Santana, F. B., de Souza, A. M. & Poppi, R. J. Visible and near infrared spectroscopy coupled to random forest to quantify some soil quality parameters. *Spectrochim Acta A Mol Biomol Spectrosc* **191**, 454–462 (2018).

76. Albuquerque, C. D. L., Nogueira, R. B. & Poppi, R. J. Determination of 17β-estradiol and noradrenaline in dog serum using surface-enhanced Raman spectroscopy and random Forest. *Microchemical Journal* **128**, 95–101 (2016).

77. FAO.Food Price Monitoring and Analysis Tool (FPMA) v4. License: CC BY-NC-SA 3.0 IGO. Extracted from: https://fpma.fao.org. Date of Access: 08-06-2022.

78. Homem-de-Mello, T. & Bayraksan, G. Monte Carlo sampling-based methods for stochastic optimization. *Surveys in Operations Research and Management Science* **19**, 56–85 (2014).

79. Kroese, D. P. & Rubinstein, R. Y. Monte Carlo methods. *Wiley Interdiscip Rev Comput Stat* **4**, 48–58 (2012).

80. Wang, J. J., Jing, Y. Y., Zhang, C. F. & Zhao, J. H. Review on multi-criteria decision analysis aid in sustainable energy decision-making. *Renewable and Sustainable Energy Reviews* **13**, 2263–2278 (2009).

81. Pandey, V., Komal & Dincer, H. A review on TOPSIS method and its extensions for different applications with recent development. *Soft comput* **27**, 18011–18039 (2023).

82. Zyoud, S. H. & Fuchs-Hanusch, D. A bibliometric-based survey on AHP and TOPSIS techniques. *Expert Syst Appl* **78**, 158–181 (2017).

83. Pedregosa, F. *et al.* Scikit-learn: Machine learning in Python. *the Journal of machine Learning research* **12**, 2825–2830 (2011).

84. McKinney, W. Data Structures for Statistical Computing in Python. *Proceedings of the 9th Python in Science Conference* **1**, 56–61 (2010).

85. Harris, C. R. *et al.* Array programming with NumPy. *Nature* **585**, 357–362 (2020).

86. ILOG CPLEX Optimization Studio 22.1.0 https://www.ibm.com/docs/es/icos/22.1.0.

87. sklearn-genetic-opt 0.10.1. https://sklearn-genetic-opt.readthedocs.io/en/stable/.

88. Kizielewicz, B., Shekhovtsov, A. & Sałabun, W. pymcdm—The universal library for solving multi-criteria decision-making problems. *SoftwareX* **22**, 101368 (2023).

89. Hunter, J. D. Matplotlib: A 2D graphics environment. *Comput Sci Eng* **9**, 90–95 (2007).

90. Waskom, M. Seaborn: Statistical Data Visualization. *J Open Source Softw* **6**, 3021 (2021).

91. Cairoli, M. & Ottenbros, A. Fingerprints to footprints: a modelling framework based on NIRS fingerprints and life cycle assessment for sustainable feed production. https://doi.org/10.5281/zenodo.10726831 (2024).

92. Vecteezy. World Map With Borders. https://www.vecteezy.com/free-vector/world-map-with-borders (2023).