Article

Visually Explaining Uncertain Price Predictions in Agrifood: A User-Centred Case-Study

Jeroen Ooge *[ID] and Katrien Verbert [ID]

Department of Computer Science, KU Leuven, 3001 Leuven, Belgium; katrien.verbert@kuleuven.be
* Correspondence: jeroen.ooge@kuleuven.be; Tel.: +32-16-32-42-86

Abstract: The rise of 'big data' in agrifood has increased the need for decision support systems that harvest the power of artificial intelligence. While many such systems have been proposed, their uptake is limited, for example because they often lack uncertainty representations and are rarely designed in a user-centred way. We present a prototypical visual decision support system that incorporates price prediction, uncertainty, and visual analytics techniques. We evaluated our prototype with 10 participants who are active in different parts of agrifood. Through semi-structured interviews and questionnaires, we collected quantitative and qualitative data about four metrics: usability, usefulness and needs, model understanding, and trust. Our results reveal that the first three metrics can directly and indirectly affect appropriate trust, and that perception differences exist between people with diverging experience levels in predictive modelling. Overall, this suggests that user-centred approaches are key for increasing uptake of visual decision support systems in agrifood.

Keywords: visual analytics; visualisation; uncertainty; explainable artificial intelligence; decision support systems; mixed-methods; thematic analysis

1. Introduction

Under the impulse of success stories in other domains, artificial intelligence and 'big data' are on the rise in agrifood [1], leading to promising research directions such as *Agriculture 4.0* [2] and the broader *Agrifood 4.0* [3], *precision agriculture* [4–6], and *smart farming* [7–9]. While the adoption of such technologies is still modest in real-life agrifood applications [10], it is expected that the wide availability of cloud computing and remote sensing [11] will further boost their spread [12]. To process the explosive amount of information in this era of growing digitisation and to make data-grounded decisions, agrifood stakeholders increasingly need the assistance of *decision support systems* (DSSs) [2] that facilitate learning and allow to modify decision processes by integrating domain knowledge, rather than systems that merely prescribe actions [13,14].

Yet, even though the need for DSSs in agrifood has been acknowledged for over two decades [13] and many prototypes have been proposed [2,15], the uptake of these systems has been limited so far. Parker et al. [16,17], Zhai et al. [2], and Rose et al. [18] discussed several reasons for this low uptake: user interfaces of DSSs are not always user-friendly and lack visualisations, DSSs are not necessarily relevant when they do not meet end users' needs or decision-making styles, outputs often miss uncertainty representations, and end users often distrust DSSs with opaque underlying algorithms. In other words, developers of DSSs for agrifood face important design challenges such as increasing usability, guarding usefulness for end users, and raising appropriate trust in underlying decision models.

Tackling these challenges requires human-centred approaches, which lie at the core of *human–computer interaction* (HCI), an interdisciplinary field that connects computer science, social sciences, and technology-applying domains such as agrifood. Specifically, HCI studies how interfaces can be designed and tailored to specific end users or application contexts to improve user experience, for example [19–21]. Two subdomains of HCI specialise in visualising complex information and explaining artificial intelligence, respectively.

The first subdomain, *visual analytics*, fosters analytical reasoning with visual dashboards that support advanced interaction and visual exploration to discover hidden patterns in data [22–24]. The second subdomain, *explainable artificial intelligence* (XAI), seeks techniques that give insights into outcomes of artificial intelligence models, and studies interrelated topics such as trust, fairness, bias, causality, accountability, privacy, and reasoning [25].

Visual analytics and XAI are relevant in agrifood because DSSs increasingly include predictive models and benefit from visualising information. Yet, current DSSs in agrifood often lack uncertainty representations and are rarely designed in a user-centred way [26]. To enable informed decision-making by different end users, researchers and practitioners have called for adopting more user-centred and HCI practices in agrifood [26–28].

We address this call by presenting a visual DSS that shows predicted food product prices and uncertainty in the predictions. We evaluated our prototype with 10 participants who are active in different parts of agrifood; collecting and analysing both qualitative and quantitative data. In particular, we focused on the following research questions:

RQ1 Usability : How user-friendly are the interaction functionalities and the visualisation in our visual DSS?
RQ2 Usefulness and needs : How useful is our visual DSS and how does it accommodate the needs of people active in agrifood?
RQ3 Model understanding : How does visualising uncertain predictions affect people's understanding of the prediction model underlying our visual DSS?
RQ4 Trust : How does visualising uncertain predictions affect people's trust in the prediction model underlying our visual DSS?

Our research contribution consists of extensively evaluating our visual DSS from two perspectives. First, considering our prototype as a product, we assessed its usability and usefulness. Sections 4.1 and 4.2 show that participants were generally very positive about our prototype's usability (RQ1) and expressed needs regarding control, comparison, and explanations (RQ2). Second, considering our prototype as an XAI research tool, we dived deeper into what affected participants' understanding of and trust in the prediction model underlying our DSS, and the relation with uncertainty visualisation. Sections 4.3 and 4.4 show that participants' understanding was affected on an algorithmic and an outcome level (RQ3), and that trust in the prediction model evolved under several factors (RQ4). In both perspectives, we considered the impact of participants' experience with predictive modelling, observing different responses for different experience levels. Finally, we made our prototypical visual DSS open-source so that the community can use it as a flexible basis for more advanced dashboards tailored to specific contexts.

2. Background and Related Work

To contextualise our research, we first discuss visualisation for DSSs and uncertainty representation. Then, we turn towards XAI and focus on trust.

2.1. Visualisation for Decision Support Systems

Visualising information augments people's abilities to get insights into complex data and more effectively fulfil tasks that cannot be automated [29]. Presenting decision-making information visually has also been found to make DSSs more user-friendly [18]. Hence, it is no surprise that DSSs often incorporate visualisations to facilitate decision-making across application domains, e.g., healthcare [30–32], learning analytics [33,34], finance [35], and supply chain analytics [36,37]. In many of these domains, decision-making is supported by visual analytics, which combines powerful visualisations with advanced interaction techniques [38] and automated data analysis. This allows people to iteratively generate and test hypotheses [22–24,39]. In healthcare, for example, visual analytics has been applied to personalise medical treatments by analysing electronic health records, modelling diseases and medical prediction, optimising care pathways, and so on [40,41].

In agrifood, many visual DSSs have been proposed too, for example in dairy farming [42], crop control [43,44], land assessment [45], irrigation management [46], and climate monitoring [47]. Yet, Gutiérrez et al. [15] found that most visual DSSs include maps, contain a single visualisation, and are intended for farmers to manage crops or assess land suitability. This suggests room for dashboards with multiple visualisations in other application areas such as livestock monitoring and sales. In addition, it suggests that current visual DSSs in agrifood are less advanced than visual analytics approaches in terms of varied visualisations and interaction possibilities.

2.2. Uncertainty Visualisation

Visual DSSs are subject to uncertainties in the data and uncertainties propagated during the data processing, modelling, and visualisation [48,49]. These uncertainties can be visualised in many ways [50,51], but there are two challenges. First, visualising uncertainty entails a trade-off: showing too much uncertainty may overload or confuse people, whereas showing too little uncertainty feigns accuracy and may mislead people [48]. Second, some approaches for uncertainty visualisation may be clearer or less misleading than others.

Tackling these challenges is hard, which unfortunately often results in simply omitting uncertainty [52,53]. This is currently the case in agriculture: visual DSSs rarely consider uncertainty [2,15]. One exception, for example, is CropGIS [44], which predicts produced biomass of maize under different meteorological conditions. CropGIS then visualises the mean prediction in a line chart, together with the minimum, maximum, and 1σ-confidence interval, resembling a *fan chart* [54] with a single fan.

Researchers in information visualisation face the above two challenges by studying the pros and cons of different uncertainty visualisation techniques. For example, in the case of predicted time series, studies have shown that (a) similar to fan charts, uncertainty intervals around a prediction line are best distinguished with different opacity levels [55]; (b) fan charts are a good compromise between accuracy and uncertainty [56]; and (c) compared to ensemble charts, fan charts lead to higher acceptance of predictions [57].

2.3. Visualisation for Explainable Artificial Intelligence

As visual DSSs often incorporate complex algorithms, end users typically need explanations to understand the algorithmic decision-making, appropriately trust it, and detect potential biases [58]. There is no one-size-fits-all explanation, however. Human-centred XAI researchers therefore study how explanations can be effectively designed, considering factors such as the application context [59–61], human reasoning processes [62], and end users' goals [63] or personal characteristics [61,64].

XAI and visual analytics largely intersect. Visualisations can namely serve as explanations when people get visual insight in model outcomes and model behaviour, actively interact with them, and steer the underlying algorithms [65]. Given the wide interest in visualisation for XAI, many surveys have discussed the state-of-the-art in visual analytics for machine learning [66,67], deep learning [68], predictive modelling [69], and enhancing trust in machine learning [70] from different perspectives. A meta-analysis of all these surveys confirmed the key role of visualisation in interpreting machine learning [71].

2.4. Trust in Intelligent Systems

Many application domains call for increasing end users' trust in algorithmic decision-making of DSSs, including agrifood [15,18]. In the scope of explaining black-box algorithms, trust is thus heavily studied in XAI and visual analytics. However, trust is a slippery concept for at least two reasons. First, there is no widely accepted definition for trust in intelligent systems, although many definitions have been proposed [72–74]. Second, measuring trust is very challenging because it evolves [75–77] and is affected by many factors [78], for example, domain expertise [75,77], visualised information and uncertainty [48,79], model accuracy [80,81], and level of transparency [82]. In addition, there is growing consensus among XAI researchers that optimising trust is not always desirable; rather, the stress

should lie on *appropriate* trust [58] and trust calibration [83,84]. Some researchers even argue that XAI research should move away from trust and focus on utility instead [85].

3. Materials and Methods

This section presents how we conducted our user-centred study. We first describe our visual DSS, study rationale, and overall study design. Then, we provide more details on how we measured usability, trust, and experience with predictive regression.

3.1. Visual Decision Support System

We developed a prototypical visual DSS for exploring product prices in various countries. Besides visualising historical price evolutions, our system visualises predicted future prices and the prediction model's uncertainty. Rather than building an advanced standalone interface with an accurate prediction model, we aimed to create a simple and flexible proof of concept for which the underlying dataset and prediction model could easily be replaced. To encourage future adaptations, we built our prototype with the open-source Meteor, React, and D3 frameworks, and made our code publicly available at https://github.com/JeroenOoge/explaining-predictions-agrifood (accessed on 9 July 2022).

In our proof of concept, the dataset contained price evolutions in European countries over the past 3 decades for over 400 food products, including fruits, vegetables, dairy, meat, and cereals. For each country separately, price predictions were generated by fitting a third-degree polynomial to the country's past price data with linear regression and least-squares estimation, extrapolating the fit for five years from the last known data point on. Uncertainty consisted of 55–99%-prediction intervals with increments of 5%.

Figure 1 shows our dashboard. At the top, two search fields with dropdown menus allow selecting a desired food product and countries available for that product. In the middle, the price evolution for selected countries is visualised in a line graph; each country is represented by a differently coloured full line. At the bottom, five checkboxes allow to enable or disable visual components: the first is enabled by default (*Past data*); the others are related to the prediction outcome and model (*Future prediction*, *Future uncertainty*, *Past fit*, and *Past uncertainty*). The future prediction and past fit are visualised as dashed lines, and the prediction intervals as stacked bands (i.e., fans), where larger intervals gradually become lighter. Finally, hovering over the chart and its visual components shows details-on-demand in the form of a tooltip with the exact price values or additional information.

Figure 1. Screenshots of our responsive visual DSS during interaction. **Left**: selecting a food product in the upper left search field and getting details about the price and date upon hovering over the line chart. **Right**: selecting countries in the upper right search field and getting a description of the hovered fan ("*In 80 out of 100 occasions, the product price lies between A and B*". where A and B are the lower and upper bounds of the prediction interval at the indicated date, respectively).

3.2. Study Rationale

Adapting to economic uncertainty and predicting market fluctuations are important challenges in Agrifood 4.0 [2]. To meet these challenges, we framed our study in the context of predicting food product prices and built upon an earlier pilot study [77], which showed that four people experienced with predictive modelling had different trust evolutions while using our visual DSS. To investigate the transferability of our preliminary results, we recruited via email 10 end users who are active in agrifood or finance. Then, we evaluated our prototypical visual DSS according to four metrics: usability, usefulness and user needs, model understanding, and trust. With the former two, we considered our prototype as a *product*: we wanted to identify issues with the visualisation and the interaction possibilities and find out whether our prototype matches participants' needs. With the latter two, we considered our prototype as an *XAI research tool*: we set out to discover how the visual components in our visual DSS impact participants' understanding of the prediction model and what affects participants' trust in the model. For all four metrics, we also considered the effect of participants' profession and experience with predictive modelling.

In addition, we were interested in whether our visual DSS would allow participants to identify the limitations of our simple prediction model. We assumed that obvious prediction failures, for example, an almost flat regression line for clearly periodic price evolutions, would not evoke lively discussions. Therefore, we deliberately built our study around a specific case of butter prices in France (data available for 1991–2011) and the Netherlands (data available for 1991–2019), with two not too obvious shortcomings. First, the model fit the past data rather poorly (high RMSEA). Second, even though France and the Netherlands had historically similar prices, the prediction for France largely diverged from the real data in the Netherlands, suggesting poor prediction performance.

3.3. Study Design

In July–October 2020, we collected qualitative data on our four evaluation metrics with online semi-structured interviews, quantitative data from Likert-type questions on trust, and observational data on how participants interacted with our visual DSS (participants shared their screen during the study). Figure 2 shows the overall structure of our study.

First, participants introduced themselves and we familiarised them with our visual DSS: we explained how they could compare past butter prices in France and the Netherlands and see details-on-demand in the visualisation, and we introduced the price prediction functionality without revealing details about the underlying prediction model.

Next, participants went through eight scenarios, enabling the *Future prediction*, *Future uncertainty*, *Past fit*, and *Past uncertainty* checkboxes one by one, first for a setting with one country (France; Scenarios 1–4) and then for a setting with two countries (France and the Netherlands; Scenarios 5–8). Figure 3 shows some representative screenshots. Each scenario consisted of three phases: (1) we asked participants to explore the visualisation while thinking out loud (*Explore the new component in the visualisation. Explain what you see. What grabs your attention?*); (2) we asked them about their trust and model understanding (*Do you trust the prediction model? Do you understand how the prediction model works? Which parts of the visualisation made you say that?*); and (3) we quantitatively measured their trust.

Finally, after completing all scenarios, participants reported their experience with four concepts related to predictive modelling and answered additional questions about model understanding and usefulness (*Which combination(s) of components do you find most useful to get insights into the prediction model? Would you like to investigate or explore other things to get insights into the prediction model? Would you use this visualisation for your job activities?*). In the post-study discussion, we asked participants how they experienced the study and stressed that our prediction model was not meant for making real-life decisions.
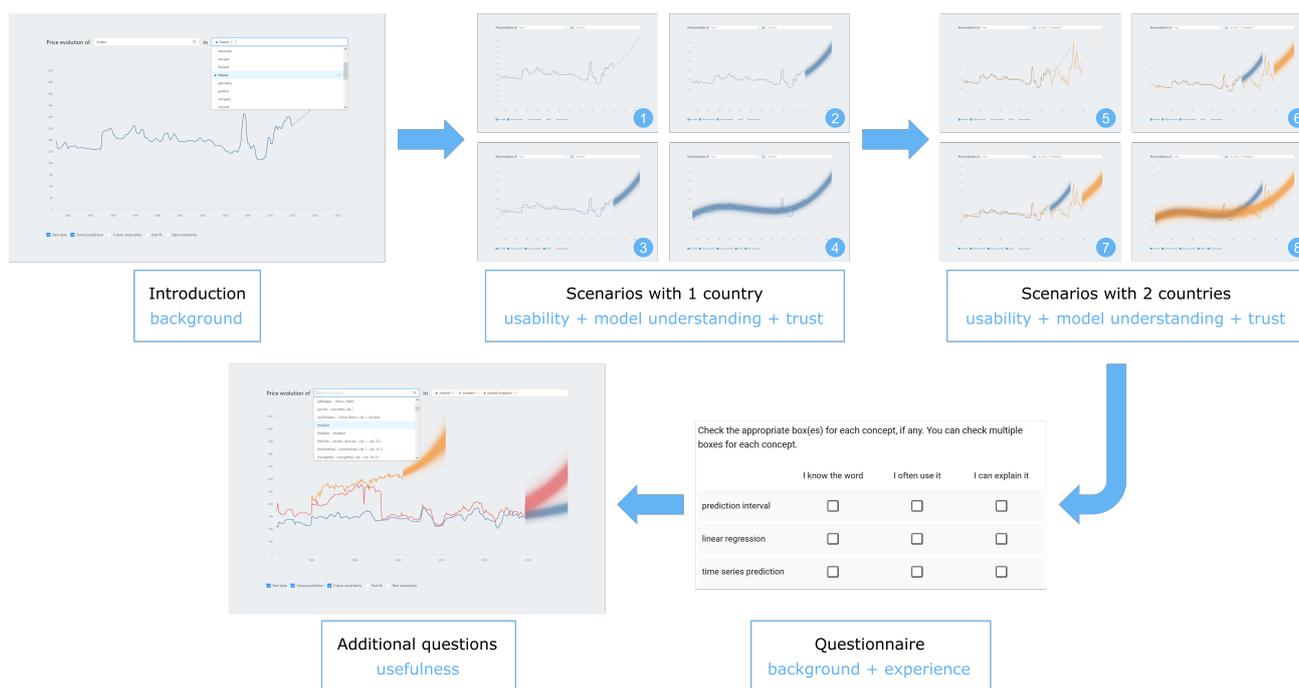
Figure 2. The flow of our study, including 5 phases: an introduction, four scenarios with one country, four scenarios with two countries, a questionnaire, and additional questions.

3.4. Measurement Instruments and Qualitative Analysis

To assess **usability** , we observed participants' interactions with our visual DSS and analysed their think-aloud feedback during exploration. As such, we could study whether participants easily found the information they were looking for; understood filtering, clicking and hovering functionalities; and had further suggestions. In contrast to Likert scales for overall usability [86,87], this approach gives concrete insights into how, why, and which parts of visualisations should be adapted to improve usability.

To quantitatively measure **trust** in each scenario, we averaged responses to four Likert-type questions rated on a 7-point range (0–not at all to 6–extremely). These questions were inspired by a widely-used scale for trust in automated systems by Jian et al. [88]. Yet, as we considered it unfeasible for participants to answer all 12 items in this scale 8 times, we selected and adapted the 4 items that seemed most relevant for prediction models:

1. *I am suspicious of the prediction model's outputs* (reverse-scored);
2. *I am confident in the prediction model*;
3. *I can trust the prediction model*;
4. *The prediction model is deceptive* (reverse-scored).

To measure participants' experience with predictive regression, we combined self-reported data and indirect experience indicators. First, participants self-reported their experience with the concepts *prediction interval*, *linear regression*, and *time series prediction* through checkboxes *I know the word* (K), *I often use it* (U) and *I can explain it* (E). For each concept, we assigned a score between 0 (very inexperienced) and 5 (very experienced) based on their answers ($K = 1$, $K \& U = 3$, $K \& E = 4$, $K \& U \& E = 5$); the average E_s served as a final estimate for self-reported experience. Second, we scored participants' experience between 0 and 5 based on their background (E_b) and use of jargon related to statistics or predictive modelling during the interview (E_j). Then, we used the average of E_s, E_b and E_j as an estimate for experience with predictive regression.

Finally, to qualitatively analyse participants' feedback, we recorded the interviews, which lasted 70–130 min, depending on the amount of feedback. We then thematically analysed 120 pages of transcription, following the 6 phases from Braun and Clarke [89]. Specifically, we first coded our data deductively (i.e., starting from our four metrics) and

then inductively for each metric (i.e., driven by the data instead of preset topics). To guard the originality of participants' feedback and respect participants' efforts to speak English, we only corrected language mistakes in quotes below when clarification was needed.
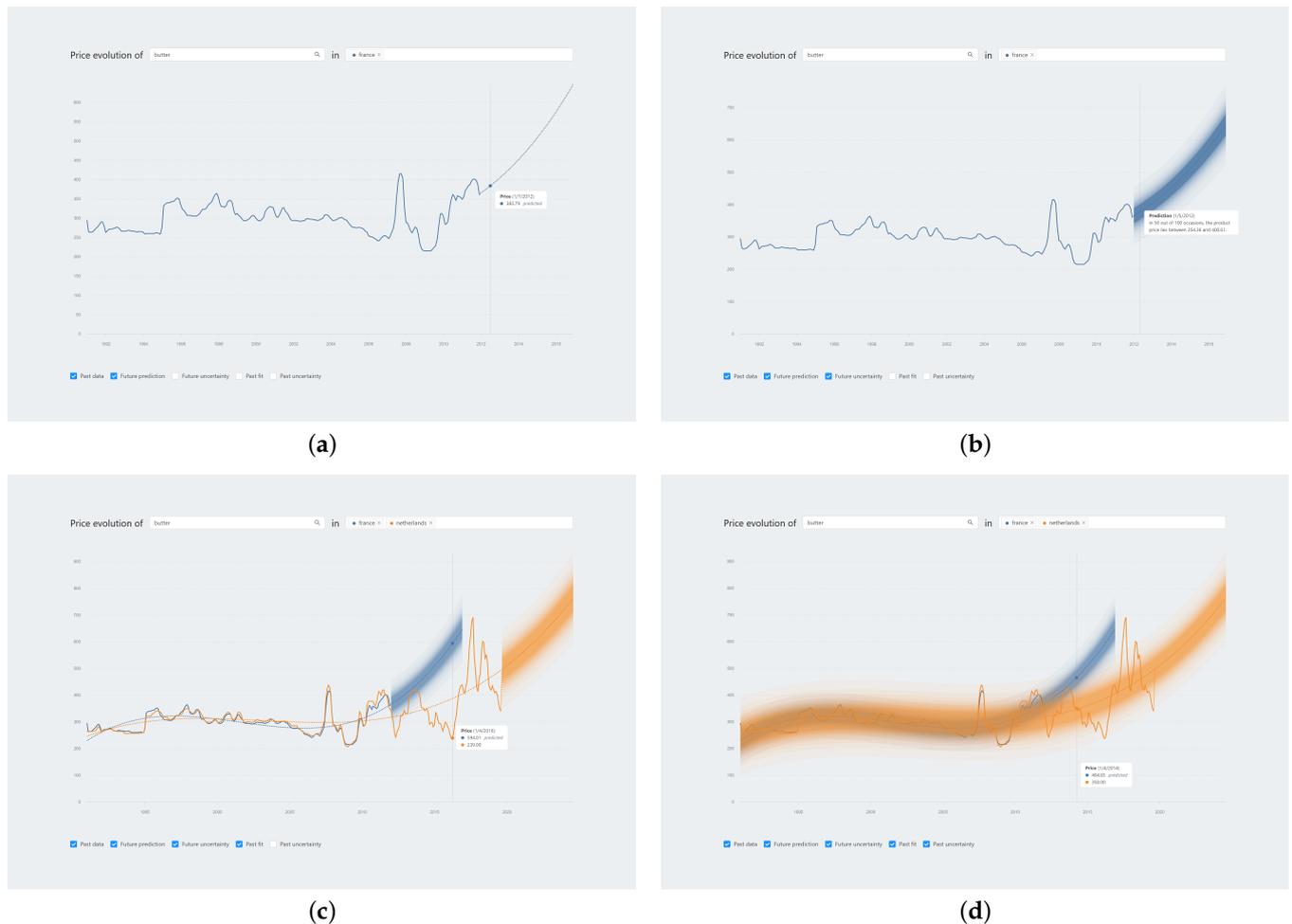

(**a**)

(**b**)

(**c**)

(**d**)

Figure 3. Our visual DSS with different sets of enabled visual components. (**a**) Scenario 1: the future prediction for France is visualised as a dashed line. (**b**) Scenario 2: the future uncertainty for France is visualised as fans. (**c**) Scenario 7: the past fit for France and the Netherlands is visualised as dashed lines. (**d**) Scenario 8: the past uncertainty for France and the Netherlands is visualised as fans.

4. Results

This section presents the findings of our study with 10 participants whose specifics are shown in Table 1. First, approaching our visual DSS as a product, we focus on usability and usefulness. Then, taking an XAI research perspective, we turn towards model understanding and trust. Throughout, as summarised in Table 2, we also highlight differences between participants who have low, medium, and high experience with predictive regression.

4.1. Usability

Our semi-structured interviews brought up four themes on usability: *Understanding the visualisation*, *Visual encoding of information*, *Interacting with the visualisation*, and *Workflow*.

> **Understanding the visualisation**: most participants understood the overall goal, but some visual components need clarification.

Table 1. Participants' background information, including their experience with predictive regression (l low, m medium, h high) as an average of self-reported experience (E_s), background (E_b), and jargon use (E_j). All participants identified as male and had a post-graduate education level.

ID	Profession	Country	Age	Experience (E_s, E_b, E_j)
P1	*Industry*: quality manager in a biscuit factory; deals with food safety issues, supply simulations	Greece	45–54	h 4.7 (4, 5, 5)
P2	*Industry*: food safety auditor for a certification body; audits companies on food safety and fraud	Greece	35–44	l 0.6 (0.3, 1, 0.5)
P3	*Industry*: quality manager in a biscuit factory; deals with food safety issues, supply simulations	Greece	35–44	m 2.9 (2.7, 3, 3)
P4	*Academia*: professor in mechanical engineering; expertise in food quality and life cycle assessment	Italy	45–54	h 4.8 (5, 5, 4.5)
P5	*Academia*: agricultural economist; expertise in value chains, food security and consumption	Italy	35–44	h 3.9 (2.3, 5, 4.5)
P6	*Industry*: sales manager for a refrigeration manufacturer; buys raw materials and sells products	Greece	35–44	h 3.8 (4.3, 4, 3)
P7	*Industry*: raw materials manager in a food company; recruits agriculturalists and keeps bees	Greece	18–34	l 0.2 (0, 0.5, 0)
P8	*Industry*: settlements coordinator in a mortgages company; verifying and approving mortgages *	Australia	35–44	h 3.7 (1, 5, 5)
P9	*Industry (Academia)*: researcher in agriculture; expertise in food chemistry and -microbiology	Greece	35–44	h 4.6 (3.7, 5, 5)
P10	*Academia (Industry)*: researcher in natural cosmetics; expertise in food science	Tunesia	18–34	h 4.3 (3, 5, 5)

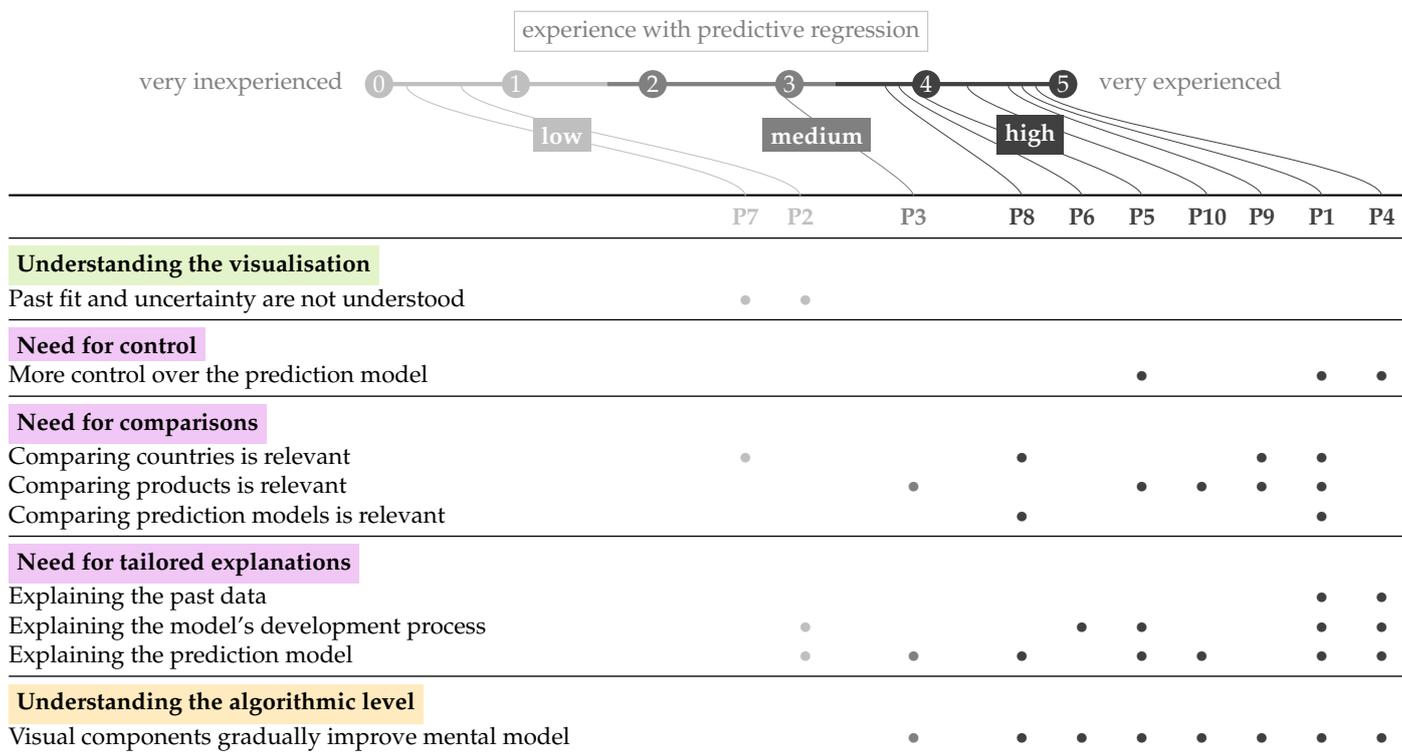* active in finance, no experience in agrifood.

Overall, participants were very positive about the visualisation and understood its main goal. For example, P4 found the visualisation *"very readable"* and complimented it for being a *"very simple instrument"* with a clear aim; P5 described the visualisation as *"very easy, simple, clear, and [without] any frills"*; and P8 stated: *"The dashboard I like. It's very simple and easy to use, so it's not too complex or anything like this. [. . .] It's just easy to use, gives you all the information [. . .] in a very sort of simple way"*. Most participants understood the visual components sufficiently and could use them without further clarification.

Specifically, participants described the future uncertainty fans as *"area[s] in which the price is statistically expected"* (P1), which *"shows the spread of [. . .] the predicted values around the [prediction] line"* (P9). In more economical terms, P5 talked about *"buffer points, which [indicate] the minimum and maximum of the variation of the future price"* and considered the fans' percentages to be *"the likelihood to be in these buffers"*. Many participants furthermore observed that uncertainty fans enlarge for larger percentages, entailing a trade-off between precision and correctness: *"[If you restrict a 90%-fan to a 50%-fan, then] you have more accuracy but you don't have a good prediction"*.

In addition, participants correctly interpreted the past fit as the *"fit between the model and the real data"* (P5), *"normalization of the slope"* (P3), *"average trend"* (P3, P6), *"natural evolution of the curve"* (P4), or *"total, general shape of the price evolution"* (P10). However, P2 and P7 did not understand the past fit line and P10 expected details when hovering over it.

Finally, while most participants seemed to intuitively understand the past uncertainty, they often lapsed into vague descriptions or were unsure how it was computed; e.g., *"it's the same like before: [. . .] the uncertainty factor"* (P3) or *"I think that you used your future model, whatever the model, and you tr[ied] to predict the past, I don't know"* (P6; *you* refers to the interviewer). Especially P2 and P7 could not get their head around the past uncertainty, with P2 questioning what others perhaps did not ask out loud: *"If you have the real numbers from the past, what's important about the uncertainty?"* Furthermore, P10 seemed to misinterpret the prediction intervals for showing accuracy: *"past uncertainty, it gives us like our model is most of the time, 85% accurate, let's say, in this point, and at the same point here it's 90%. I mean it gives us a better understanding of the model and if it's accurate or not"*.

Table 2. Some topics raised by the participants, ordered by their experience with predictive regression (P2 and P7 have low experience; P3 has medium experience; others have high experience).

	P7	P2	P3	P8	P6	P5	P10	P9	P1	P4
Understanding the visualisation										
Past fit and uncertainty are not understood	●	●								
Need for control										
More control over the prediction model						●			●	●
Need for comparisons										
Comparing countries is relevant	●			●				●	●	
Comparing products is relevant			●			●	●	●	●	
Comparing prediction models is relevant				●					●	
Need for tailored explanations										
Explaining the past data									●	●
Explaining the model's development process		●			●	●			●	●
Explaining the prediction model		●	●	●		●	●		●	●
Understanding the algorithmic level										
Visual components gradually improve mental model			●	●	●	●	●	●	●	●

In conclusion, it would be helpful to clarify the past fit and uncertainty components, especially for participants with low experience in predictive regression (see Table 2). To clarify the uncertainty, adapting the fans' tooltip could be a start because P6 pointed out that currently, some might confuse the word 'occasions' with 'iterations' and therefore misinterpret the X%-fan as representing *"X out of 100 calculations."*

> **Visual encoding of information**: visually encoding uncertain price evolutions as a line graph with fans was clear yet limited.

All participants understood the visual encoding of price evolution as a line chart, and also the visual encoding of uncertainty as fans did not seem to cause confusion. Regarding the latter, P1 and P3 discussed the different shades explicitly: *"The more prices you get scattering around the line, the more, the deeper the shadow becomes [and vice versa]. So statistically, more prices are expected to be falling in a short distance above or below the line"*. (P1) and *"as it goes [from the prediction line] to the borders, [...] the possibility it goes down"* (P3).

Yet, the visual encoding has two limitations. First, when uncertainty components are enabled, simultaneously plotting multiple countries can be *"a little bit confusing"* (P2) or *"a little bit disturbing"* (P10) because of the many different colours and the overlapping graphical elements that hamper hovering specific fans. For example, when P8 plotted about 15 countries simultaneously, he said bluntly: *"Oof. [...] Yeah, I'm not really gonna get much out of that"*. Fortunately, participants realised that the trade-off between completeness and overplotting is their own responsibility: *"you cannot compare, I don't know, 10 different commodities in 10 different countries, otherwise no one can understand what is shown in the graph"* (P5). Second, although participants understood that the Y-axis unit was not important for the study, they frequently mentioned that it should be clarified in real-life applications. For example, P6 joked: *"I mean, what is this 300? 300 cows or what?"*

> **Interacting with the visualisation**: participants did not experience major filtering or hovering issues; zooming might be handy.

The filtering functionality was clear for all participants. Regarding the hovering functionality, getting details-on-demand through hovering seemed natural for both the line chart and the uncertainty fans. One minor remark here is that P5, P6, and P7 did not spontaneously hover over the fans when they first saw them, which suggests that a real-life fan chart might need to stress this possibility. Two participants found the highlighting of hovered uncertainty fans suboptimal. First, P8 regretted that he could not simultaneously highlight a fan and see price details (*"as soon as I move my mouse out, I lose it [the fan tooltip], so it's very fiddly"*); and he proposed to allow pinning the fans. Second, P10 agreed that highlighted fans obscure other details and suggested altering their visual encoding from fans to lines that indicate standard deviations along with the corresponding probabilities.

In addition, P10's interactions in Scenario 7 demonstrated that a zooming feature could improve usability: P10 disabled the future uncertainty to reduce the Y-axis' length and thus artificially zoom in on the past fit lines to better see small-scale changes.

> **Workflow**: the current workflow for selecting products and countries was clear, but alternative workflows might be more efficient.

All participants understood the current workflow of first choosing a product and then selecting one or more countries. Yet, P5 and P9 proposed alternative workflows that could improve usability when focusing on a fixed set of countries. Tapping into the idea of focusing on a single country, P9 found it *"a bit annoying that anytime we are choosing a product [we need] to select again a country; [...] if you choose a product, you can play with the countries, but if you choose a country you cannot play with the products"*. Thus, to make the process of comparing different products for the same country less *"time-consuming"*, he would reverse the current selection order. Generalising this idea, P5 suggested a two-step selection workflow: an initial step to *"include all I want in the analysis–for example, different products for the same country or different countries for the same product"*, followed by visualising the selected information. Then, *"a sort of matrix with all the countries I have selected"* instead of dropdown lists would allow to quickly (de)select countries or products, which is, for example, convenient to remove overlap in the visualisation.

4.2. Usefulness and Needs

Participants raised two themes on usefulness (*Overall usefulness of the visualisation* and *Usefulness of the visual components*) and three themes on their needs (*Need for control*, *Need for comparisons*, and *Need for tailored explanations*).

> **Overall usefulness of the visualisation**: a visual DSS similar to ours was deemed useful for different tasks in agrifood or finance.

All participants agreed that visual DSSs similar to ours can be useful for different tasks in agrifood or finance. Generally speaking, P2 said that *"it's a very good tool for everyone in the food industry"* and P5 expected that *"a lot of people are looking for something similar"*.

More concretely, participants indicated that visualising predicted product prices can benefit industrial and academical agrifood parties. For agrifood companies, our visual DSS could be *"useful mainly in order to make future schedules"* (P9) such that *"people who make decisions [and] who need insights in future price evolutions [...] can make contracts [with suppliers] for the coming years in order to avoid to pay too much"* instead of reacting to the market (P3). In addition, P2 saw a link with food fraud detection: *"the food price many times affects the food fraud cases [so] it helps companies to predict [the number of] food fraud cases"*. In agrifood research, P4 explained that researchers often study economical aspects such as demand and logistics, so he found our visualisation *"very interesting [...] to make some evaluation about the importance of some particular market and which is the prospective of that market"*.

Participants also saw more general applications for our visual DSS. For example, P10 stated that exporting companies would be interested in predicting demand in foreign countries, and P8 indicated that financial companies would be interested in predicting interest rates because *"this sort of helps you make better business decisions [. . . and] be better prepared"*. Thus, our visual DSS could be more useful when people can upload and visualise their own data. Furthermore, our visualisation is not bound to be a standalone tool: P1 *"would expect to see this dashboard attached in [a full analysis of the prediction model]; a text, showing, explaining how it works"* and P3, anticipating that the prediction model could consider climate change and geopolitics, saw the opportunity to extend our dashboard with additional visualisations of, for example, temperature and carbon emissions.

> **Usefulness of the visual components**: how useful visual components were depended on the context, but uncertainty was a natural requirement for many.

Participants often mentioned that the usefulness of the visual components depends on the desired insight. For example, while P5 found all components *"very useful"* to analyse a single time series, he would probably hide the past fit and past uncertainty when comparing multiple time series: *"It depends in my opinion on what you want to visualise"*. In addition, P9 distinguished between obtaining precise values and drawing overall conclusions about the trend: *"You need [. . .] the future prediction to have an exact number [. . .] but just to make conclusions, you don't need it. You just need the [future] uncertainty and the fit"*. Last, P6 noted that he did not need an explicit dotted line to get a feeling about the general past trend. Given these considerations, the flexibility to enable and disable visual components in our visual DSS seems very useful.

Regarding the uncertainty components, most participants considered them a natural requirement because of the predictive context. For example, P1 said *"Whenever we need to predict something, there is always an uncertainty in our prediction. So it's more something that I would expect"*. and P8 agreed *"There are always going to be [macro level] factors that sort of change the prediction"*. Some participants even asked for future uncertainty representations right in Scenario 1: *"It could be interesting [. . .] to have the minimum and maximum value in that prediction period. A sort of standard value. [. . .] I expect [. . .] a sort of uncertain value [instead of] a precise value"*. (P4) and *"Maybe you should add some best cases and worse cases"* (P10). While discussing uncertainty, participants also touched upon a fundamental trade-off: *"It's like a double-shaped blade, you know. It gives you more liberty in choosing which kind of occasions you will be having, and at the same time, it gives you like not accurate results"*. (P10), and *"The thicker the lines [fans] become, the more useless the data because [. . .] everything is within specs, but you see you have a huge variation"* (P1). P4 added that, instead of multiple uncertainty levels, he only required a 1σ-interval. Overall, it thus seems essential to visualise the uncertainty in predictions, potentially allowing to modify the number of shown uncertainty levels.

> **Need for control**: some participants requested additional control over the visualisation or the prediction model.

Some participants proposed additional features to explore the visualisation. Specifically, P5 suggested to allow filtering on specific time intervals; P5 and P10 proposed to allow changing the currency such that end users can better relate to the price evolutions; and P8 was looking for more in-depth pricing details such as the price per unit, retail price, and trading indicators such as the moving average convergence divergence.

In addition, some participants highly experienced with predictive regression voiced a need for more control over the prediction model (see Table 2). For example, P1 explained that he wants absolute control over prediction models: *"I use quite often the regression, the data analysis function in Excel. So I use the data in the way I want. I fit the models that I consider to fit best for the case. [. . .] The visualisation [. . .] would be quite helpful but based on what I have seen until now, I wouldn't [. . .] consider very much the prediction values. I would only use it for historical data acquisition"*. P5 also seemed to allude to this by stating that our visual DSS would be *"extremely useful"* if scientists and practitioners could download the available data

and graphs for further analysis. Other requests for control were changing the predicted time span (P2, P4, P8) and the time frame used for training the prediction model (P1).

> **Need for comparisons**: participants found it important to simultaneously compare countries, products, and prediction models.

Participants across all levels of experience with predictive regression stressed the relevance of comparing countries (see Table 2). For example, P9 said: *"Of course comparing different countries is really useful because we are talking about [...] a unite Europe [and] you might have incoming products from different countries. [...You] might have a purchaser from Italy and one from Germany, so you have both as an alternative to buy materials"*. Given this united European market, P8 added that he liked comparing prices with the European average.

Regarding the need to compare products, two ideas to extend our visual DSS arose. First, P3, P5, and P9 suggested to compare *similar* products (e.g., cereals, sweeteners, vegetal oils) in the same graph to understand potential relations between them. Such insights could, for example, be useful for farmers and regulatory bodies: *"the decision for farmers to produce rice instead of maize, or wheat instead of barley and so on, could be strongly conditioned by the provision [...], and regulatory bod[ies] for the market can provide specific support for specific farmers"*. (P5). Second, P10 suggested to simultaneously compare *different* products: *"For instance, if you want to make a muffin, you would have like flour, wheat, some milk, some eggs, flavour vanilla or chocolate. So you wanna keep each ingredient into consideration. [...] Maybe you can have like a [curve] for each ingredient [...and see the total] cost [for] the final product"*.

Last, participants experienced with predictive modelling would find comparing different prediction models useful to get an idea about how well they agree on their predictions and to, as P8 mentioned, follow the most frequent prediction, giving more weight to sophisticated models. Still, P1 emphasised: *"[I] would expect each model to be discussed: why does this model predict different values from another one and the reasoning behind that"*.

> **Need for tailored explanations**: participants required tailored explanations about different aspects with different levels of detail.

Participants brought up four different aspects for which they needed explanations, and, interestingly, Table 2 shows that these participants had low to high experience with predictive modelling. First, P1 and P4 required a discussion of the past data and sudden peaks or troughs, backed by economical factors. Both P9 and P10, however, suspected that people active in industry would be most interested in explanations regarding the future, rather than the past. Second, participants wanted to know more about the provenance and accuracy of the raw price data, the model developers, the data processing, and the training of the prediction model. Third, P2 and P6 wanted to know how reliable the predictions were: *"The [end user] needs to feel that the model is predicting OK without knowing though what the model is doing. [...] You need somehow to explain to the end user what could be the prediction capability"*. (P6). Fourth, typically triggered by the steep predicted price increases in Scenarios 1–8, many participants requested explanations about the prediction model itself. For example, P3 asked about the model's input factors: *"For me, it's very critical to understand what factors this model takes into account to predict such a high rise of the butter [price]."*, and P4 wanted *"a basic idea on how the prediction model works rather than going with something sort of blindly, [to see] evidence that this all works"*.

Furthermore, two participants had opposite views on the required level of detail in explanations. On the one hand, P1 requested full transparency of the prediction model: *"If it is a regression, I would be interest[ed] to see the equation that comes from the model. I would expect to see a discussion on the price variation, the reasoning"*. On the other hand, P6 vividly argued that he did not need this amount of detail: *" I don't believe you need to give it to a third party, to a user, when [they are] looking at data, the mathematics behind the model. [...] In my job, for example, one of the most important things is to know raw material prices [...] and I need to have a good prediction. Now how the prediction works? I really don't care"*.

The two observations above seemed to be part of a more general phenomenon: many participants alluded to tailoring explanations, i.e., adapting them to different contexts and to the people that need them. For example, P4 attributed his need for a description of the model to his *"research mind"*, but added that seeing uncertainty already filled part of that need, while economists would probably require more details: *"After see[ing . . .] the statistical evaluation [uncertainty], in my opinion, my need [for a more detailed explanation] is lower because I of course consider the fact that perhaps they derived from some economical model that are at the basis of this evaluation. [. . .] Perhaps for economist[s . . .] it would be more interesting to know something more about the model. But of course, this is not my topic so for me it's sufficient what I see in the graph"*. Similarly, when P1 asked for *"a very thorough discussion"* of the prediction model, he added: *"But this is me, OK. I'm an engineer, I'm quite experienced in mathematics and statistics and you understand, I know how it can work. I don't know if the same discussion was done with somebody who is not quite good in maths or in statistics, what his[/their] perspective would be."* Finally, while P5 found our visual DSS useful for educational purposes, he acknowledged that he would require more a detailed explanation when using it in high-stakes contexts: *"If I need to use it for a practical or a professional use, like the support for the country or the region, for a specific policy, and so on, I think I have to give them, to guarantee them about the quality of the data. And if I don't know exactly the model, what you have included and so on, and I couldn't replicate your analysis, it's quite impossible to use it as a standard or a benchmark"*.

4.3. Model Understanding

This section uncovers how the visual components and functionalities in our visual DSS impacted participants' understanding of the prediction model. Three themes, *Understanding the algorithmic level*, *Understanding the outcome level*, and *Understanding by comparing countries*, reveal that understanding manifested itself on an algorithmic and an outcome level.

> **Understanding the algorithmic level**: the visual components improved participants' understanding of the prediction model's technicalities, but only gradually.

In Scenarios 1 and 5, all participants indicated that simply plotting predictions does not invoke model understanding. For example, P5 stated: *"I have no idea which kind of variables you included in the model, if the model is based on different variables, I don't know, so the general international market or a policy decision, a local decision in France, or climate change or climate information. [. . .] and the technological evolution or [. . .] macroeconomic data"*. This lack of understanding was typically followed by a request for an explanation.

Yet, the stepwise introduction of extra visual components improved many participants' mental model of the prediction model, ranging from a better intuition to identifying the true modelling technique (see Table 2). To P3, P4, and P8, the future uncertainty suggested the model to be a statistical technique: *"It was more clear to me that we're not talking about, let's say, absolute values, but talking about the statistical model, so there you can see the possibility of the price evolution of the butter to be inside this space"* (P3). After enabling the past fit, P4 and P10 noticed that the past fit and future prediction formed a continuous curve, which gave them a better idea about how the prediction was constructed: *"[I] know in a better way the model [. . .] the evolution of the future is more clear [. . .] Of course, I don't know which is the mathematical model but I know that this is in a sort of curve, fit that you obtain, and so the model, I see the input from this evolution of the data"* (P4). Visualising the past uncertainty sometimes further reinforced understanding the link between past fit and future prediction: P4 noted that *"with this representation [. . .] the future prediction is completely integrated in the previously data"* and also P6, while unsure about how the past uncertainty was generated, got the feeling that the prediction was based on the trend line. After seeing the uncertainty and fit components, P1, P8, and P9 even strongly suspected that the prediction model was a regression. For example, P1 correctly identified the prediction as a third-degree polynomial, but he admitted that the visual components did not reveal the precise mathematical equation.

P6 explained how the *"step by step approach"* allowed him to *"understand parts of how the model works"* without revealing the technicalities: *"If you would only show me the first picture,*

no, I would not be able to tell you how the model works, but going to the future uncertainty and past uncertainty, and presenting also the trend line, then OK, you get a clearer picture of how the model probably works. But still, the details, it's not something that I think you can get with these simple steps". Furthermore, he added that none of the visual components was all-enlightening: *"Obviously it had to do with the whole sequence. [...] Step by step then you can get it. But it's not like you go like you know 'wow, wow, this is clear now'. [...] it is a gradual, let's say, picture*". Interestingly, to improve the mental model faster, P4 suggested an alternative *"more logical"* order for enabling the visual components: he would first show the past data, past fit, and past uncertainty to explain *"that you have a statistical consideration"* and only then show the future prediction and future uncertainty to clarify that they are *"derived from the past fit"*.

> **Understanding the outcome level**: the visual components allowed participants to interpret model outcomes and assess their accuracy.

Participants often commented that uncertainty did not explain the prediction's upward trend. For example, P5 said that *"uncertainty doesn't explain the prediction, it's just more inclusive"* and P1 added that he did not know whether he *"should expect that the price would increase or would decrease sharply"* after the prediction horizon.

However, the uncertainty and past fit components gave participants insights into model performance. In particular, P5 explained that the uncertainty revealed how well the model fits the data: narrow uncertainty meant a good fit; wide uncertainty meant a worse fit. The past uncertainty was also *"a sort of measure of the robustness of the model"* (P5), which gave a *"better understanding of the model and if it's accurate or not"* (P10) and indicated whether *"the past performance might repeat itself in the future, providing that the trend remains the same"* (P8). The past fit, then, allowed participants to detect outliers due to exceptional market events. For example, when P1 enabled the past fit, he said: *"[The] model has explained reasonably, reasonably, the variation of butter price throughout the decades. Could not predict the peak that occurred in 2008. Might have been an issue due to the financial crisis [...] We don't have this information but something has happened there that could not be predicted"*.

Finally, P6 proposed to assess model performance by comparing a country's past data with what the model predicts without that data: *"why you don't [...] compare what the model told us and what actually happened? Then you can evaluate also the effectiveness of your model"*.

> **Understanding by comparing countries**: comparing countries had cons for understanding the algorithmic level, but pros for understanding the outcome level.

On the algorithmic level, the feature to compare countries sometimes led to misunderstanding the model's technicalities. This was illustrated by P3, who in Scenario 5 wrongly assumed that, to predict product prices in one given country, the prediction model also considered data from other countries: *"This model probably took into account what happened in the region, I mean in Europe, during this period of time. So that's probably why the price of the butter in France is going to rise so much. [...] now I can understand, let's say the reasoning behind this slope, why this slope is very steep [and] goes up"*.

On the outcome level, however, comparing countries allowed participants to better understand the model's performance. P4 and P10, for example, were especially interested in the model's consistency and expected that countries with similar price evolutions in the past would have similar price predictions. More importantly, in our experiment, showing data from France and the Netherlands allowed participants to compare the model's prediction for France with real data from the Netherlands. For P1, *"that was what actually convinced [him] that the model is quite unreliable"* because *"we can see that the actual data of Netherlands are far away from the prediction of the forecasted data for France"*. Similarly, P5, P6, and P8 emphasised that the divergence in price was accentuated by the fact that a large portion of the real data for the Netherlands did not lie inside the future uncertainty fans for France: *"the prediction buffer which should include all the data, more or less because it's 99% of the variation, doesn't include, doesn't encompass the real data [...] If we assume that [...] data of the Netherlands would be a reliable prediction [for France...] there is a big problem with the prediction model"* (P6).

4.4. Trust

Our results on trust consist of two parts. First, we present participants' quantitative trust evolution over the eight scenarios to spot differences and similarities. Next, we contextualise observed trends with the thematically analysed qualitative feedback.

4.4.1. Quantitative Results on Trust

Figure 4 shows the evolution of participants' reported trust scores over all scenarios. Overall, participants had very different trust evolutions. Yet, there is a clear distinction between two groups: P1, P5, and P6 converged to low trust, whereas the other participants converged to at least rather trusting the prediction model. The level of experience with predictive regression did not explain this distinction because, for example, while P1 and P4 had the highest experience scores, they were both on different extremes of the trust scale. Another observation is that few participants reported dramatic changes in trust: only P6 and P10 have a difference of at least 2.5 between their minimal and maximal trust scores.
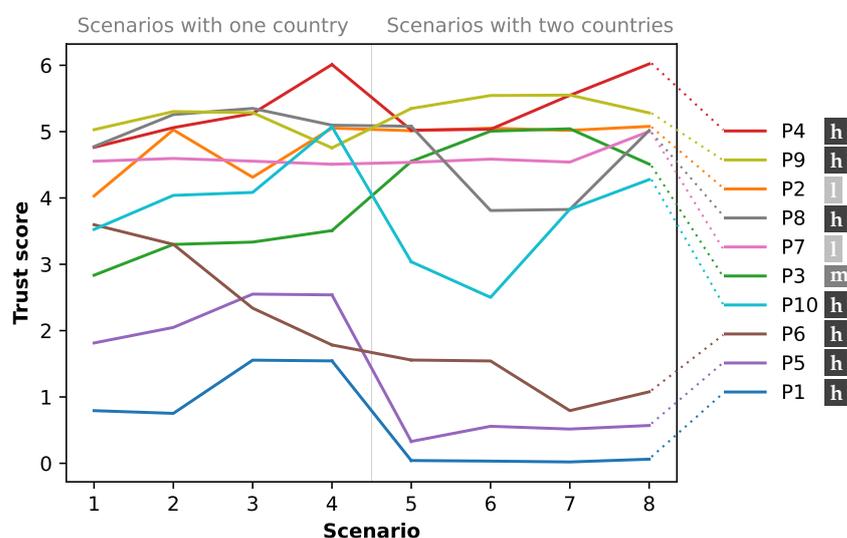
Figure 4. Participants' trust in the prediction model over eight scenarios. Scenarios 1–4 showed data for one country; Scenarios 5–8 showed data for two countries. Lines are slightly jittered for clarity. The legend includes the level of experience with predictive regression (l low, m medium, h high).

4.4.2. Qualitative Results on Trust

Four themes impacted trust in the prediction model. The first two themes, *Model performance* and *Model understanding*, were heavily impacted by expectation violation and expectation agreement: when participants encountered things that did not meet their expectations, their trust typically decreased, and vice versa. The other two themes, *Presence of uncertainty* and *Explanations*, tapped into what participants required for growing trust.

> **Model performance**: seeing the model performance affected how participants assessed the model's trustworthiness; seeing model failures had a negative impact.

In Scenarios 1–4, participants assessed the prediction model based on the past fit and past uncertainty. The past fit did not decrease most participants' trust because it seemed to fit *"reasonably good the price variation, though in quite some [. . .] rough estimation"* (P1) and thus *"gives more robustness to the model"* (P5). Yet, for P6, the past fit highlighted that specific outliers were not foreseen by the model, which made him more unconfident: *"Why does not predict that it will have a peak and then go down again. [. . .] It does not persuade me. [. . .] I'm losing my confidence with a trend line"*.

Likewise, the past uncertainty led to mixed trust responses. On the one hand, some participants indicated aspects that increased their trust. For example, for P4, the option to

do an *"evaluation of the data during the past"* increased his trust in *"the correctness of the model"*. P10 made a similar argument based on the fans showing the model's accuracy: *"I think this past uncertainty will add more credibility to our prediction model [. . .] I think I'm better trusting this model, [. . .] I can have a better understanding [. . .] of the prediction model as it goes over the years"*. Furthermore, P8 observed that most of the data points lay inside the uncertainty fans, but found it reassuring that some lay outside: *"[the price] falls out every now and again, which I mean, it does happen with everything. [. . .] I guess it increases my trust because if it was too perfect, you'd be like, you know, I mean, nothing in life is 100% certain, so why would this thing be?"* On the other hand, P6 actually became more hesitant when seeing a peak outside the 99%-fan: *"So there is a problem there, right? I know that it is only for a small period of time, like few months that the model fails over whatever I see here, like 25 years. But still, it fails. Is it acceptable? I don't know. I mean, if it was inside the band that I see here, maybe I would be happy"*.

In Scenarios 5–8, participants often assessed the prediction model by comparing the prediction for France with the real data or the past fit for the Netherlands. Many participants noticed a divergence between both: *"the butter price in France historically was closely linked to the butter price in Netherlands and we can see that the actual data of Netherlands are far away from the prediction of the forecasted data for France"* (P1). As Figure 4 shows, this resulted in a huge drop in trust for P1, P5, P6, P8, and P10 because they expected a prediction for France similar to the data for the Netherlands. For example, P1 said that he *"would not trust this model at all"* because it *"convinced me that the model is quite unreliable"*, and P6 motivated: *"I don't trust the model. You see, the real data was totally different than the prediction. [. . .] obviously, you prove that in a sense there are flaws in the model prediction"*.

Yet, not all participants experienced the divergence as an expectation violation. For example, P4 pointed out that the long-term performance seemed good and hypothesised that market events might have caused the divergence: *"in 2010 you have a differentiation. [. . .] the events that you have in Netherland are perhaps due to particular events that you had there, which I don't know, of course. [. . .] in the extrapolation, [. . .] the values are different, but the behaviour is very similar. [. . .] But I consider that at the end, five years later or 10 years later, also in Netherland you have the same price"*. P8 make a similar remark in Scenario 6, restoring his trust afterwards: *"if you look at around 2016, the price prediction is way off. Way way off, but it sort of meets the further it goes along. So I think [. . .] if I was trying to make a price prediction like five years in the future or something, I trust it more, rather than, I would if it was one or two years in the future"*. However, P5 called these observations of 'good' long-term performance a *"bias in the visualisation"* caused by the prediction for France coincidentally stopping at the real peaks of the Netherlands.

> **Model understanding**: participants' trust reactions differed depending on how they understood the prediction model on an outcome or algorithmic level.

Participants' model understanding on an outcome level influenced their trust. A typical example was Scenario 1, where the prediction line caused a lot of scepticism because it violated many participants' expectations for two reasons. First, participants could not understand its steep slope. For example, P10 joked *"it can't be like this: it goes like higher up in the sky. [chuckles]"* and P6 added: *"The trend of the previous ten years, no 20 years, does not imply that you're gonna have this rapid index increase"*. Instead, some expected a price behaviour *"similar like the last 10 years, let's say"* (P3). Second, participants noticed that it did not have peaks or troughs like the past data: *"The thing that I'm worried about it that the curved line is like so decent, so perfect, so shaped"*. (P10); and *"that peak that I see on October of 2007 and that trough that I see on March of 2009 is not what I see in the model prediction, comparing five years to five years"*. (P6). However, most participants still reported a trust score above neutral because of mitigating considerations that agreed with their expectations. For example, P6 noted that in the last few plotted years, *"there has been an increasing rate which does not look too different toward what the prediction model has there"*. Furthermore, due to *"the global inflation and the economic crisis etcetera, and a lot of pressure on the market places"* (P10) increasing prices

seemed plausible: *"usually we have increase prices, not decreases [laughs], so that's why I'm more in the part that I'm trusting the prediction"* (P9).

Participants' trust was also affected by their model understanding on an algorithmic level. First, understanding decreased trust under expectation violation. For example, P6 understood that predictions were based on the past fit, but observed several unexpected things, which is why he insisted: *"I have a better understanding how the model works, but I don't trust it, I insist"*. Second, understanding increased trust under expectation agreement. For example, in Scenario 5, P3 gained trust because he built a (wrong) mental model that met his observations: *"this model probably took into account what happened [...] in Europe [...]. So that's probably why the price of the butter in France is going to rise so much. [...] I would say that I'm not suspicious anymore. [...] Because now I can understand, let's say the reasoning behind this slope"*. Furthermore, P9 reported high trust scores because he saw *"nothing strange. It's just what I was expecting to see. [...] it's just a regression [...] for me that I'm understanding how the models are working now, it looks normal"*. One comment here is that P9, upon seeing the diverging behaviour of France and the Netherlands in Scenario 5, also mentioned: *"it might, change my trust for the model as a model, OK, and how you incorporate the model in your data set but not for the prediction that we are generating for the future. Maybe a better model will give you better [results]"*. This suggests that P9 based his trust on *how* the prediction outcomes were computed, rather than whether regression was a suitable technique.

> **Presence of uncertainty**: seeing that the prediction model accounted for uncertainty did not decrease participants' trust.

In Scenario 2, none of the participants indicated that their trust in the prediction model decreased because of the presence of future uncertainty. On the contrary, most participants' trust increased. P9, for example, explained why: *"the more descriptive the model becomes, and the more alternatives that it gives you, it makes you trust more. When you have just a line, you more or less, you cannot believe that things in real life are so accurate, right? [chuckles ...] I would say that future prediction without future uncertainty is not much trustful"*. While P1 and P3 agreed with this, they both stressed that the uncertainty did not increase their trust dramatically because it did not take away their need for an explanation: *"it's a model that takes some reasonable uncertainty, but still I cannot trust it because I don't know how it was developed"*. (P1); and *"I'm more, let's say, confident about this prediction model. But still, I want to know the reason why the butter has to go up"*. (P3).

For P4, the uncertainty overall generated more trust because it suggested the prediction model to be the product of scientific studies: *"I trust in a more–I suppose that behind this value you have some studies, some studies that come from your research for your ability"*. Furthermore, related to algorithmic model understanding, P4 believed that the uncertainty suggested the prediction model to be of a statistical nature: *"I prefer the fact that the model works in a statistical way because with some consideration, I suppose this is more right in a model that works in the future. [...] I'm more and more trusting, trust about the correctness of the model"*.

> **Explanations**: participants considered explanations about the development process and the prediction model requisites for building trust.

To trust the prediction model, participants mentioned that they needed an explanation about the development process and data provenance. For example, P1 said that *"in order to trust a prediction model, I need to know how it was developed"*. Furthermore, P4 and P5 alluded to the importance of who developed the prediction model. P4 referred to trusting the model developers' competence: *"when I approach information that come from an organisation or something like you, I suppose, my behaviour is to accept this evolution because I suppose that you have the competence to develop a model. [...] I have to believe in you with some [...] suspicious behaviour"*. In turn, P5 argued that a model stemming from an official institution might be more reliable: *"if such a prediction comes from an official body like FAO or World Bank or so on, could be more reliable, I can say. If come from a university [...] it's not an official body and it's more difficult to understand. So I just can imagine that [...] when a World Bank provide prediction, it's*

the fruit of the convergent opinion of different practitioners and scientists". Concerning the data provenance, P1 asked about the accuracy of the given historical data because *"in order to trust a prediction model, I need to know [. . .] what is the raw data [in]put"*.

Participants also considered an explanation about the prediction model itself key for building trust. For example, P5 did not trust the prediction in Scenario 1 because *"I have no idea how you provide this prediction, how you calculate it and the model behind. [. . .] there is no explanation of the model, and it's quite difficult to trust in the model without any description"*. P1 agreed: *"whenever I have a prediction model, I always try to find the physics and engineering behind that. If there is no physics explanation or engineering explanation, I'm quite sceptical"*.

5. Discussion

This section answers our research questions by discussing our quantitative and qualitative results. Then, based on our observations, it underlines the need for user-centred approaches in agrifood to increase the uptake of visual DSSs.

5.1. A User-Friendly and Useful Visual DSS

Our results show that participants were generally very positive about our prototypical visual DSS in terms of usability (RQ1): the visualisation, its interaction possibilities, and the general workflow were clear overall. In addition, participants imagined that a visual DSS similar to ours would be useful as support in several decision-making contexts, including food fraud detection, business scheduling, and market evaluation (RQ2). They also highly appreciated that our visual DSS fulfilled their need to compare countries and that visual components could be restricted to those relevant for desired insights. Thus, our prototype seems to be a user-friendly flexible basis for more advanced visual DSSs that extend our interface, and could be embedded in (dynamic) analytics reports.

Yet, we recognise two points of attention related to people's experience with predictive modelling. First, while many participants stressed the usefulness of uncertainty, our prototype could not remove all confusion around past uncertainty and past fit. Thus, especially for people who are less experienced with predictive modelling, it seems necessary to elaborate on the past fit and uncertainty components when used in a visual DSS. This could be realised with more detailed tooltips, a brief information screen, or—as suggested by Sacha et al. [48]—a simple tutorial with some exemplar usage scenarios. Second, especially people with high predictive modelling experience could have a need for controlling and comparing different prediction models. To meet this need, visual DSS in agrifood could draw inspiration from visual analytics systems evaluated in other domains [90–92].

5.2. Tailoring, Tailoring, Tailoring: Different End Users, Different Needs

Participants covered three important needs (RQ2): controlling the visualisation and prediction model; comparing countries, products and prediction models; and getting explanations about the past data, data processing, prediction reliability, and prediction model. Interestingly, other studies on predictive DSSs also revealed a need for comparison. For example, comparing cows' milk production allowed animal researchers to identify trends, clusters, and anomalies [42]; and product demand analysts expressed the need to compare prediction performance for similar products [93].

Overall, participants' needs seemed heavily subject to their personal background and job activities. This shows the importance of tailoring visual DSSs and explanations on at least three levels. First, *tailoring towards the application context*: the specific agrifood subdomain and the overall goal of the visual DSS determine which functionalities and visual components are useful. Second, *tailoring towards experience with predictive modelling*: for people with low experience, an intuitive understanding of the prediction model and little control over the prediction model might suffice, whereas people with high experience might require mathematical explanations and control over the prediction model. Third, *tailoring towards tasks*: different tasks and desired insights might require different visual explanations, similar to what Gutiérrez et al. [56] argued for.

5.3. Gradual Model Understanding through Visual Analysis

The visual components and comparison functionality in our visual DSS affected participants' model understanding on two levels (RQ3). On an *algorithmic level*, many participants gradually grew a better intuition of the model's technicalities. In XAI terms, the visual components thus served as explanations that fostered their mental model. On an *outcome level*, participants could better interpret predictions and assess their accuracy.

However, some participants created mental models that did not stroke with the real regression model. For example, they assumed that the model based its predictions on price evolutions in multiple countries or considered additional input variables such as climate and geopolitics. This suggests that complementary explanations are necessary to avoid wrong assumptions, bearing in mind that these explanations should balance soundness and completeness [94]: simply adding more information does not necessarily spark useful mental models. Other participants' model understanding did not improve because they could not analyse the visualised information thoroughly, most likely due to low experience with predictive regression or time series analysis overall. To grow correct model understanding, such end users seem to require more guidance in the data analysis process; it is unclear whether the current exploratory nature of our visual DSS fits this need.

5.4. Trust Is Multi-Faceted and Evolves

Our results subscribe to the multi-faceted and evolving nature of people's trust in a prediction model (RQ4), similar to many previous studies [75–78]. We identified four themes that influenced people's trust: the model's performance, understanding the model, uncertainty in the model's outcomes, and explanations about the development process or the prediction model itself. The former two themes were strongly coloured by whether participants' expectations were violated or met; the negative impact of expectation violation is in line with findings from Kizilcec et al. [82]. The latter two themes covered what participants deemed necessary to grow trust. The fact that participants required the presence of uncertainty for building trust reinforces the call for incorporating uncertainty in visual DSSs for agrifood.

We observed clear evidence of trust calibration [48]: participants' trust was based on a continuous trade-off between the aforementioned four themes. The direction in which their trust evolved then depended on which theme was most dominant. For example, most participants initially focused on requiring explanations. Some then evolved to distrusting the prediction model due to low performance, whereas others developed more trust due to observations that matched their model understanding. This explains the different trust evolutions in our quantitative measurements. An important note here is that the quantitative scores are hard to compare directly because participants typically have different calibrations for scoring. On an individual level, though, we found that most participants' trust scores did not change drastically over the eight scenarios. For participants with low experience in predictive modelling, this was most likely due to their inability to fully analyse the visualised information. Why these participants trusted the prediction model nevertheless is unclear. Potentially, factors such as good usability fostered their trust, or the participants reported what they conceived as desirable.

5.5. Fostering Appropriate Trust through Usefulness and Meeting Needs

While our results presented four evaluation metrics and their corresponding themes separately, some themes are connected or partially overlap. Figure 5 summarises all themes together with their most relevant relations grounded in our qualitative data. The relations clearly link usefulness to trust, either directly, or indirectly via model understanding.
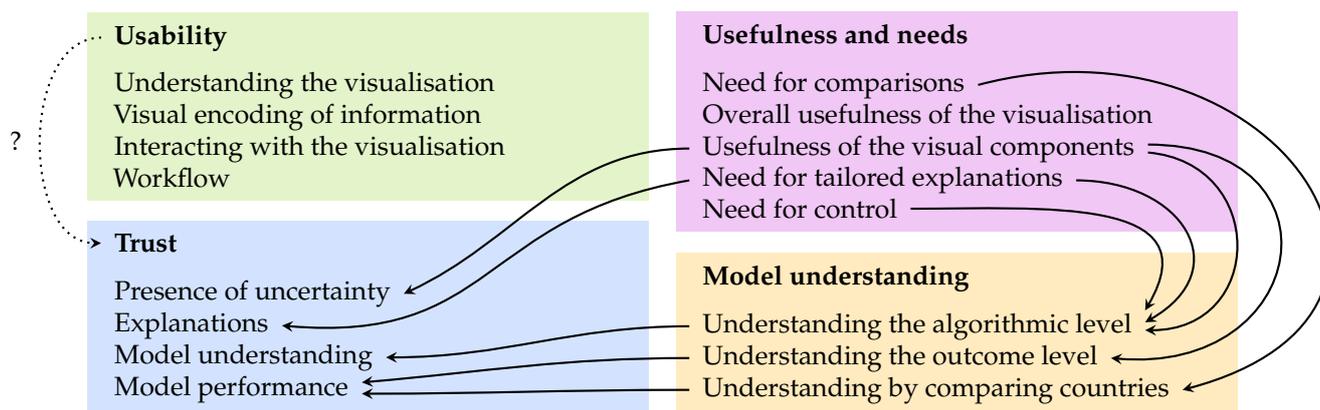
Figure 5. Summary of the themes on usability, usefulness and needs, model understanding, and trust. Some relations between themes are indicated with arrows; themes are reordered to avoid overlap.

Two direct relations concern uncertainty and explanations. First, while uncertainty was considered a natural and useful requirement for bringing nuance to predictions, participants also considered it a requisite for building trust. There exist interesting parallels in other domains: for example, people tend to discount weather forecasts without uncertainty [52]. Second, participants often stressed a need for explanations about the prediction model and its development process, adding that they could not build trust without them. This illustrates the relevance of XAI research into the utility of explanations [85].

Two indirect relations link usefulness to trust through model understanding. First, the visual components in our DSS were deemed useful for understanding the model on an algorithmic level. Control over the prediction model and tailored explanations about the prediction model were expected to facilitate the same. In turn, observing things that agree with model understanding led to increased trust. This suggests that improving model transparency with tailored explanations, for example carefully designed visualisations, can foster appropriate trust, which is in line with common beliefs in the XAI community [58]. Second, the visual components and the functionality to compare countries in our DSS allowed participants to better understand model outcomes, which in turn revealed model performance. Seeing the prediction model's performance allows assessing its trustworthiness, which is essential for appropriate trust [83,84].

5.6. Taking a Step Back: Increasing Uptake of DSSs in Agrifood with User-Centred Approaches

Before concluding, we reflect upon the broader impact of our findings for agrifood. Central in our overall story was the lacking uptake of (visual) DSSs in agrifood. Rose et al. [18] pointed out that trust is a key factor for increasing uptake. Quotes from our interviews such as *"I think that for a scientist I can use prediction data only if my trust on this data is full"* (P5) and *"you don't have the time to [...] explore if the model works or does not work. [...] I just want to believe what I have in front of me"* (P6) indeed seem to confirm that people will not use applications they distrust. From this point of view, it seems reasonable that scholars and practitioners in agrifood and other domains often advocate for designing DSSs that increase trust.

However, simply designing for increasing trust is not always desirable and should not be the final goal because trust eventually manifests itself when applications prove to be reliable and useful over time [85]. Our results, summarised in Figure 5, support this claim: the relations between usefulness and trust suggest that useful and tailored visual DSSs may eventually foster appropriate trust. Therefore, it seems recommended to apply user-centred approaches to design useful DSSs that meet end users' needs. In the long run, this can foster appropriate trust and in turn uptake. Furthermore, user-centred approaches have the additional asset of exposing people to new technologies [27], which can also

stimulate trust [18]. Thus, user-centred approaches seem vital for ameliorating the current low uptake of visual DSSs in agrifood.

5.7. Limitations and Transferability

Our research is subject to some limitations. Most importantly, our sample of 10 participants is most likely too small to achieve full data saturation in our qualitative results. Yet, it is encouraging that our trust themes largely correspond to those found in our pilot study [77]. Larger studies could investigate whether more themes emerge concerning trust as well as the other evaluation metrics. To further validate our observed differences between people with different levels of experience in predictive regression, it would be particularly interesting to include more people with low or medium experience. Furthermore, future work can investigate the transferability of our results to other domains such as finance and healthcare, where predictive models play an important role too. Since our sample contained only one participant active in finance, we cannot draw strong conclusions on potential differences with agrifood yet.

Finally, as good performance is a core factor for uptake of DSSs [18], real-life applications based on our prototypical visual DSS should include suitable models for forecasting time series, for example, exponential smoothing or LSTM [95,96].

6. Conclusions

We presented a prototypical visual DSS for agrifood that incorporates price prediction, uncertainty and visual analytics techniques. An elaborate evaluation with 10 participants active in agrifood or finance revealed many insights concerning usability, usefulness and needs, model understanding, and trust. For example, participants were generally very positive about our prototype's usability and discussed needs regarding control, comparison, and explanations. Our results also show that usefulness and trust are related, either directly, or indirectly through model understanding. Moreover, we observed that participants' job activities and experience with predictive modelling influenced their perceptions and needs. Combining all these findings illustrates that user-centred approaches are vital for increasing the uptake of visual DSSs in agrifood.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

References

1. Kamilaris, A.; Kartakoullis, A.; Prenafeta-Boldú, F.X. A Review on the Practice of Big Data Analysis in Agriculture. *Comput. Electron. Agric.* **2017**, *143*, 23–37. [CrossRef]
2. Zhai, Z.; Martínez, J.F.; Beltran, V.; Martínez, N.L. Decision Support Systems for Agriculture 4.0: Survey and Challenges. *Comput. Electron. Agric.* **2020**, *170*, 105256. [CrossRef]
3. Lezoche, M.; Hernandez, J.E.; Alemany Díaz, M.d.M.E.; Panetto, H.; Kacprzyk, J. Agri-Food 4.0: A Survey of the Supply Chains and Technologies for the Future Agriculture. *Comput. Ind.* **2020**, *117*, 103187. [CrossRef]
4. Cisternas, I.; Velásquez, I.; Caro, A.; Rodríguez, A. Systematic Literature Review of Implementations of Precision Agriculture. *Comput. Electron. Agric.* **2020**, *176*, 105626. [CrossRef]
5. Linaza, M.T.; Posada, J.; Bund, J.; Eisert, P.; Quartulli, M.; Döllner, J.; Pagani, A.; Olaizola, I.G.; Barriguinha, A.; Moysiadis, T.; et al. Data-Driven Artificial Intelligence Applications for Sustainable Precision Agriculture. *Agronomy* **2021**, *11*, 1227. [CrossRef]
6. Wachowiak, M.P.; Walters, D.F.; Kovacs, J.M.; Wachowiak-Smolíková, R.; James, A.L. Visual Analytics and Remote Sensing Imagery to Support Community-Based Research for Precision Agriculture in Emerging Areas. *Comput. Electron. Agric.* **2017**, *143*, 149–164. [CrossRef]
7. Wolfert, S.; Ge, L.; Verdouw, C.; Bogaardt, M.J. Big Data in Smart Farming—A Review. *Agric. Syst.* **2017**, *153*, 69–80. [CrossRef]
8. Moysiadis, V.; Sarigiannidis, P.; Vitsas, V.; Khelifi, A. Smart Farming in Europe. *Comput. Sci. Rev.* **2021**, *39*, 100345. [CrossRef]
9. Ayoub Shaikh, T.; Rasool, T.; Rasheed Lone, F. Towards Leveraging the Role of Machine Learning and Artificial Intelligence in Precision Agriculture and Smart Farming. *Comput. Electron. Agric.* **2022**, *198*, 107119. [CrossRef]
10. Osinga, S.A.; Paudel, D.; Mouzakitis, S.A.; Athanasiadis, I.N. Big Data in Agriculture: Between Opportunity and Solution. *Agric. Syst.* **2022**, *195*, 103298. [CrossRef]
11. Navarro, E.; Costa, N.; Pereira, A. A Systematic Review of IoT Solutions for Smart Farming. *Sensors* **2020**, *20*, 4231. [CrossRef] [PubMed]
12. Liakos, K.; Busato, P.; Moshou, D.; Pearson, S.; Bochtis, D. Machine Learning in Agriculture: A Review. *Sensors* **2018**, *18*, 2674. [CrossRef] [PubMed]
13. McCown, R. Changing Systems for Supporting Farmers' Decisions: Problems, Paradigms, and Prospects. *Agric. Syst.* **2002**, *74*, 179–220. [CrossRef]
14. Rojo, D.; Htun, N.N.; Parra, D.; De Croon, R.; Verbert, K. AHMoSe: A Knowledge-Based Visual Support System for Selecting Regression Machine Learning Models. *Comput. Electron. Agric.* **2021**, *187*, 106183. [CrossRef]
15. Gutiérrez, F.; Htun, N.N.; Schlenz, F.; Kasimati, A.; Verbert, K. A Review of Visualisations in Agricultural Decision Support Systems: An HCI Perspective. *Comput. Electron. Agric.* **2019**, *163*, 104844. [CrossRef]
16. Parker, C.; Campion, S. Improving the Uptake of Decision Support Systems in Agriculture. In Proceedings of the First European Conference for Information Technology in Agriculture, Copenhagen, Denmark, 15–18 June 1997; pp. 129–134.
17. Parker, C. A User-Centred Design Method for Agricultural DSS. In Proceedings of the EFITA-99: Proceedings of the Second European Conference for Information Technology in Agriculture, Bonn, Germany, 27–30 September 1999; pp. 27–30.
18. Rose, D.C.; Sutherland, W.J.; Parker, C.; Lobley, M.; Winter, M.; Morris, C.; Twining, S.; Ffoulkes, C.; Amano, T.; Dicks, L.V. Decision Support Tools for Agriculture: Towards Effective Design and Delivery. *Agric. Syst.* **2016**, *149*, 165–174. [CrossRef]
19. Carroll, J.M. Human–Computer Interaction: Psychology as a Science of Design. *Int. J. Hum.-Comput. Stud.* **1997**, *46*, 501–522. [CrossRef]
20. Shneiderman, B.; Plaisant, C.; Cohen, M.; Jacobs, S.; Elmqvist, N.; Diakopoulos, N. *Designing the User Interface: Strategies for Effective Human-Computer Interaction*, 6th ed.; Pearson: Hoboken, NJ, USA, 2016.
21. Olson, G.M.; Olson, J.S. Human-Computer Interaction: Psychological Aspects of the Human Use of Computing. *Annu. Rev. Psychol.* **2003**, *54*, 491–516. [CrossRef]
22. Keim, D.A.; Mansmann, F.; Schneidewind, J.; Thomas, J.; Ziegler, H. Visual Analytics: Scope and Challenges. In *Visual Data Mining*; Simoff, S.J., Böhlen, M.H., Mazeika, A., Eds.; Springer: Berlin/Heidelberg, Germany, 2008; Volume 4404, pp. 76–90. [CrossRef]
23. Cui, W. Visual Analytics: A Comprehensive Overview. *IEEE Access* **2019**, *7*, 81555–81573. [CrossRef]
24. Ham, D.H. The State of the Art of Visual Analytics. In Proceedings of the EKC 2009 Proceedings of the EU-Korea Conference on Science and Technology, Reading, UK, 5–7 August 2009; Springer Proceedings in Physics; Lee, J.H., Lee, H., Kim, J.S., Eds.; Springer: Berlin/Heidelberg, Germany, 2010; pp. 213–222. [CrossRef]
25. Abdul, A.; Vermeulen, J.; Wang, D.; Lim, B.Y.; Kankanhalli, M. Trends and Trajectories for Explainable, Accountable and Intelligible Systems: An HCI Research Agenda. In Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems, Montreal, QC, Canada, 21–26 April 2018; Association for Computing Machinery: New York, NY, USA, 2018; pp. 1–18. [CrossRef]
26. Rose, D.C.; Parker, C.; Fodey, J.; Park, C.; Sutherland, W.J.; Dicks, L.V. Involving Stakeholders in Agricultural Decision Support Systems: Improving User-Centred Design. *Int. J. Agric. Manag.* **2017**, *6*, 10.

27. Parker, C.; Sinclair, M. User-Centred Design Does Make a Difference. The Case of Decision Support Systems in Crop Production. *Behav. Inf. Technol.* **2001**, *20*, 449–460. [CrossRef]

28. Lindblom, J.; Lundström, C.; Ljung, M.; Jonsson, A. Promoting Sustainable Intensification in Precision Agriculture: Review of Decision Support Systems Development and Strategies. *Precis. Agric.* **2017**, *18*, 309–331. [CrossRef]

29. Munzner, T. *Visualization Analysis and Design*; A K Peters/CRC Press: Boca Raton, FL, USA, 2014. [CrossRef]

30. Rind, A. Interactive Information Visualization to Explore and Query Electronic Health Records. *Found. Trends Hum. Comput. Interact.* **2013**, *5*, 207–298. [CrossRef]

31. Botha, C.P.; Preim, B.; Kaufman, A.; Takahashi, S.; Ynnerman, A. From Individual to Population: Challenges in Medical Visualization. *arXiv* **2012**, arXiv:1206.1148.

32. West, V.; Borland, D.; Hammond, W. Innovative Information Visualization of Electronic Health Record Data: A Systematic Review. *J. Am. Med. Inf. Assoc.* **2015**, *22*, 330–339. [CrossRef]

33. Verbert, K.; Govaerts, S.; Duval, E.; Santos, J.L.; Van Assche, F.; Parra, G.; Klerkx, J. Learning Dashboards: An Overview and Future Research Opportunities. *Pers. Ubiquitous Comput.* **2013**, *18*, 1499–1514. [CrossRef]

34. Vieira, C.; Parsons, P.; Byrd, V. Visual Learning Analytics of Educational Data: A Systematic Literature Review and Research Agenda. *Comput. Educ.* **2018**, *122*, 119–135. [CrossRef]

35. Savikhin, A.; Lam, H.C.; Fisher, B.; Ebert, D.S. An Experimental Study of Financial Portfolio Selection with Visual Analytics for Decision Support. In Proceedings of the 2011 44th Hawaii International Conference on System Sciences, Kauai, HI, USA, 4–7 January 2011; pp. 1–10. [CrossRef]

36. Khakpour, A.; Colomo-Palacios, R.; Martini, A. Visual Analytics for Decision Support: A Supply Chain Perspective. *IEEE Access* **2021**, *9*, 81326–81344. [CrossRef]

37. Basole, R.C.; Bellamy, M.A.; Park, H. Visualization of Innovation in Global Supply Chain Networks. *Decis. Sci.* **2017**, *48*, 288–306. [CrossRef]

38. Yi, J.S.; ah Kang, Y.; Stasko, J.; Jacko, J. Toward a Deeper Understanding of the Role of Interaction in Information Visualization. *IEEE Trans. Vis. Comput. Graph.* **2007**, *13*, 1224–1231. [CrossRef]

39. Keim, D.A.; Mansmann, F.; Thomas, J. Visual Analytics: How Much Visualization and How Much Analytics? *ACM Sigkdd Explor. Newsl.* **2010**, *11*, 5–8. [CrossRef]

40. Hu, J.; Perer, A.; Wang, F. Data Driven Analytics for Personalized Healthcare. In *Healthcare Information Management Systems*; Weaver, C.A., Ball, M.J., Kim, G.R., Kiel, J.M., Eds.; Springer International Publishing: Cham, Switzerland, 2016; pp. 529–554. [CrossRef]

41. Preim, B.; Lawonn, K. A Survey of Visual Analytics for Public Health. *Comput. Graph. Forum* **2020**, *39*, 543–580. [CrossRef]

42. Di Silvestro, L.; Burch, M.; Caccamo, M.; Weiskopf, D.; Beck, F.; Gallo, G. Visual Analysis of Time-Dependent Multivariate Data from Dairy Farming Industry. In Proceedings of the 2014 International Conference on Information Visualization Theory and Applications (IVAPP) Lisbon, Portugal, 5–8 January 2014; pp. 99–106.

43. Armstrong, L.J.; Nallan, S.A. Agricultural Decision Support Framework for Visualisation and Prediction of Western Australian Crop Production. In Proceedings of the 2016 3rd International Conference on Computing for Sustainable Global Development (INDIACom), New Delhi, India, 16–18 March 2016; pp. 1907–1912.

44. Machwitz, M.; Hass, E.; Junk, J.; Udelhoven, T.; Schlerf, M. CropGIS—A Web Application for the Spatial and Temporal Visualization of Past, Present and Future Crop Biomass Development. *Comput. Electron. Agric.* **2019**, *161*, 185–193. [CrossRef]

45. Ochola, W.O.; Kerkides, P. An Integrated Indicator-Based Spatial Decision Support System for Land Quality Assessment in Kenya. *Comput. Electron. Agric.* **2004**, *45*, 3–26. [CrossRef]

46. Accorsi, P.; Lalande, N.; Fabrègue, M.; Braud, A.; Poncelet, P.; Sallaberry, A.; Bringay, S.; Teisseire, M.; Cernesson, F.; Le Ber, F. HydroQual: Visual Analysis of River Water Quality. In Proceedings of the 2014 IEEE Conference on Visual Analytics Science and Technology (VAST), Paris, France, 25–31 October 2014; pp. 123–132. [CrossRef]

47. Jarvis, D.H.; Wachowiak, M.P.; Walters, D.F.; Kovacs, J.M. Adoption of Web-Based Spatial Tools by Agricultural Producers: Conversations with Seven Northeastern Ontario Farmers Using the GeoVisage Decision Support System. *Agriculture* **2017**, *7*, 69. [CrossRef]

48. Sacha, D.; Senaratne, H.; Kwon, B.C.; Ellis, G.; Keim, D.A. The Role of Uncertainty, Awareness, and Trust in Visual Analytics. *IEEE Trans. Vis. Comput. Graph.* **2016**, *22*, 240–249. [CrossRef]

49. Skeels, M.; Lee, B.; Smith, G.; Robertson, G.G. Revealing Uncertainty for Information Visualization. *Inf. Vis.* **2010**, *9*, 70–81. [CrossRef]

50. Demmans Epp, C.; Bull, S. Uncertainty Representation in Visualizations of Learning Analytics for Learners: Current Approaches and Opportunities. *IEEE Trans. Learn. Technol.* **2015**, *8*, 242–260. [CrossRef]

51. Spiegelhalter, D.; Pearson, M.; Short, I. Visualizing Uncertainty About the Future. *Science* **2011**, *333*, 1393–1400. [CrossRef]

52. Franconeri, S.L.; Padilla, L.M.; Shah, P.; Zacks, J.M.; Hullman, J. The Science of Visual Data Communication: What Works. *Psychol. Sci. Public Interest* **2021**, *22*, 110–161. [CrossRef]

53. Hullman, J. Why Authors Don't Visualize Uncertainty. *IEEE Trans. Vis. Comput. Graph.* **2020**, *26*, 130–139. [CrossRef] [PubMed]

54. Britton, E.; Fisher, P.; Whitley, J. *Quarterly Bulletin February 1998*; Technical Report; Bank of England: London, UK, 1998.

55. Seipp, K.; Gutiérrez, F.; Ochoa, X.; Verbert, K. Towards a Visual Guide for Communicating Uncertainty in Visual Analytics. *J. Comput. Lang.* **2019**, *50*, 1–18. [CrossRef]

56. Gutiérrez, F.; Ochoa, X.; Seipp, K.; Broos, T.; Verbert, K. Benefits and Trade-Offs of Different Model Representations in Decision Support Systems for Non-expert Users. In *Human-Computer Interaction–INTERACT 2019*; Lecture Notes in Computer Science; Lamas, D., Loizides, F., Nacke, L., Petrie, H., Winckler, M., Zaphiris, P., Eds.; Springer International Publishing: Cham, Switzerland, 2019; pp. 576–597. [CrossRef]

57. Leffrang, D.; Müller, O. Should I Follow This Model? The Effect of Uncertainty Visualization on the Acceptance of Time Series Forecasts. In Proceedings of the 2021 IEEE Workshop on TRust and EXpertise in Visual Analytics (TREX), Virtual, 24 October 2021; pp. 20–26. [CrossRef]

58. Gunning, D.; Aha, D. DARPA's Explainable Artificial Intelligence (XAI) Program. *AI Mag.* **2019**, *40*, 44–58.

59. Vellido, A. The Importance of Interpretability and Visualization in Machine Learning for Applications in Medicine and Health Care. *Neural Comput. Appl.* **2020**, *32*, 18069–18083. [CrossRef]

60. Dhanorkar, S.; Wolf, C.T.; Qian, K.; Xu, A.; Popa, L.; Li, Y. Who Needs to Know What, When?: Broadening the Explainable AI (XAI) Design Space by Looking at Explanations Across the AI Lifecycle. In Proceedings of the Designing Interactive Systems Conference 2021, Virtual, 28 June–2 July 2021; Association for Computing Machinery: New York, NY, USA, 2021; pp. 1591–1602.

61. Suresh, H.; Gomez, S.R.; Nam, K.K.; Satyanarayan, A. Beyond Expertise and Roles: A Framework to Characterize the Stakeholders of Interpretable Machine Learning and Their Needs. In Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems, Virtual, 8–13 May 2021; Association for Computing Machinery: New York, NY, USA, 2021; pp. 1–16. [CrossRef]

62. Wang, D.; Yang, Q.; Abdul, A.; Lim, B.Y. Designing Theory-Driven User-Centric Explainable AI. In Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems, Glasgow, UK, 4–9 May 2019; Association for Computing Machinery: New York, NY, USA, 2019; pp. 1–15.

63. Mohseni, S.; Zarei, N.; Ragan, E.D. A Multidisciplinary Survey and Framework for Design and Evaluation of Explainable AI Systems. *ACM Trans. Interact. Intell. Syst.* **2021**, *11*, 24:1–24:45. [CrossRef]

64. Millecamp, M.; Htun, N.N.; Conati, C.; Verbert, K. To Explain or Not to Explain: The Effects of Personal Characteristics When Explaining Music Recommendations. In Proceedings of the 24th International Conference on Intelligent User Interfaces, Marina del Ray, CA, USA, 17–20 March 2019; ACM: New York, NY, USA, 2019; pp. 397–407. [CrossRef]

65. Ooge, J.; Stiglic, G.; Verbert, K. Explaining Artificial Intelligence with Visual Analytics in Healthcare. *Wires Data Min. Knowl. Discov.* **2021**, 12, e1427. [CrossRef]

66. Endert, A.; Ribarsky, W.; Turkay, C.; Wong, B.W.; Nabney, I.; Blanco, I.D.; Rossi, F. The State of the Art in Integrating Machine Learning into Visual Analytics: Integrating Machine Learning into Visual Analytics. *Comput. Graph. Forum* **2017**, *36*, 458–486. [CrossRef]

67. Liu, S.; Wang, X.; Liu, M.; Zhu, J. Towards Better Analysis of Machine Learning Models: A Visual Analytics Perspective. *Vis. Inf.* **2017**, *1*, 48–56. [CrossRef]

68. Hohman, F.; Kahng, M.; Pienta, R.; Chau, D.H. Visual Analytics in Deep Learning: An Interrogative Survey for the Next Frontiers. *IEEE Trans. Vis. Comput. Graph.* **2019**, *25*, 2674–2693. [CrossRef]

69. Lu, Y.; Garcia, R.; Hansen, B.; Gleicher, M.; Maciejewski, R. The State-of-the-Art in Predictive Visual Analytics. *Comput. Graph. Forum* **2017**, *36*, 539–562. [CrossRef]

70. Chatzimparmpas, A.; Martins, R.; Jusufi, I.; Kucher, K.; Rossi, F.; Kerren, A. The State of the Art in Enhancing Trust in Machine Learning Models with the Use of Visualizations. *Comput. Graph. Forum* **2020**, *39*, 713–756. [CrossRef]

71. Chatzimparmpas, A.; Martins, R.M.; Jusufi, I.; Kerren, A. A Survey of Surveys on the Use of Visualization for Interpreting Machine Learning Models. *Inf. Vis.* **2020**, *19*, 207–233. [CrossRef]

72. Jacovi, A.; Marasović, A.; Miller, T.; Goldberg, Y. Formalizing Trust in Artificial Intelligence: Prerequisites, Causes and Goals of Human Trust in AI. *arXiv* **2021**, arXiv:2010.07487.

73. Madsen, M.; Gregor, S. Measuring Human-Computer Trust. In Proceedings of the 11th Australasian Conference on Information Systems, Brisbane, Australia, 6–8 December 2000; Volume 53, pp. 6–8.

74. Vereschak, O.; Bailly, G.; Caramiaux, B. How to Evaluate Trust in AI-Assisted Decision Making? A Survey of Empirical Methodologies. *Proc. ACM Hum.-Comput. Interact.* **2021**, *5*, 327:1–327:39. [CrossRef]

75. Nourani, M.; King, J.; Ragan, E. The Role of Domain Expertise in User Trust and the Impact of First Impressions with Intelligent Systems. In Proceedings of the AAAI Conference on Human Computation and Crowdsourcing, Virtual, 25–29 October 2020; Volume 8, pp. 112–121.

76. Holliday, D.; Wilson, S.; Stumpf, S. User Trust in Intelligent Systems: A Journey Over Time. In Proceedings of the 21st International Conference on Intelligent User Interfaces, Sonoma, CA, USA, 7–10 March 2016; ACM: New York, NY, USA, 2016; pp. 164–168.

77. Ooge, J.; Verbert, K. Trust in Prediction Models: A Mixed-Methods Pilot Study on the Impact of Domain Expertise. In Proceedings of the 2021 IEEE Workshop on TRust and EXpertise in Visual Analytics (TREX), Virtual, 24 October 2021; IEEE: New Orleans, LA, USA, 2021; pp. 8–13. [CrossRef]

78. Hoff, K.A.; Bashir, M. Trust in Automation: Integrating Empirical Evidence on Factors That Influence Trust. *Hum. Factors J. Hum. Factors Ergon. Soc.* **2015**, *57*, 407–434. [CrossRef]

79. Mayr, E.; Hynek, N.; Salisu, S.; Windhager, F. Trust in Information Visualization. In Proceedings of the EuroVis Workshop on Trustworthy Visualization (TrustVis), Porto, Portugal, 3 June 2019; p. 5. [CrossRef]

80. Yin, M.; Wortman Vaughan, J.; Wallach, H. Understanding the Effect of Accuracy on Trust in Machine Learning Models. In Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems, Glasgow, UK, 4–9 May 2019; ACM: New York, NY, USA, 2019; pp. 1–12. [CrossRef]

81. Papenmeier, A.; Kern, D.; Englebienne, G.; Seifert, C. It's Complicated: The Relationship between User Trust, Model Accuracy and Explanations in AI. *ACM Trans. Comput. Hum. Interact.* **2022**, *29*, 35:1–35:33. [CrossRef]

82. Kizilcec, R.F. How Much Information?: Effects of Transparency on Trust in an Algorithmic Interface. In Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems, San Jose, CA, USA, 7–12 May 2016; ACM: New York, NY, USA, 2016; pp. 2390–2395. [CrossRef]

83. Han, W.; Schulz, H.J. Beyond Trust Building—Calibrating Trust in Visual Analytics. In Proceedings of the 2020 IEEE Workshop on TRust and EXpertise in Visual Analytics (TREX), Virtual, 25 October 2020; IEEE: Salt Lake City, UT, USA, 2020; pp. 9–15. [CrossRef]

84. Solhaug, B.; Elgesem, D.; Stolen, K. Why Trust Is Not Proportional to Risk. In Proceedings of the The Second International Conference on Availability, Reliability and Security (ARES'07), Vienna, Austria, 10–13 April 2007; pp. 11–18. [CrossRef]

85. Davis, B.; Glenski, M.; Sealy, W.; Arendt, D. Measure Utility, Gain Trust: Practical Advice for XAI Researchers. In Proceedings of the 2020 IEEE Workshop on TRust and EXpertise in Visual Analytics (TREX), Virtual, 25 October 2020; IEEE: Salt Lake City, UT, USA, 2020; pp. 1–8. [CrossRef]

86. Brooke, J. SUS: A 'quick and Dirty' Usability Scale. In *Usability Evaluation in Industry*; Jordan, P.W., Thomas, B., Weerdmeester, B.A., McClelland, I.L., Eds.; Taylor & Francis: London, UK, 1996; Volume 189.

87. Bangor, A.; Kortum, P.T.; Miller, J.T. An Empirical Evaluation of the System Usability Scale. *Int. J. Hum. Comput. Interact.* **2008**, *24*, 574–594. [CrossRef]

88. Jian, J.Y.; Bisantz, A.M.; Drury, C.G. Foundations for an Empirically Determined Scale of Trust in Automated Systems. *Int. J. Cogn. Ergon.* **2000**, *4*, 53–71. [CrossRef]

89. Braun, V.; Clarke, V. Thematic Analysis. In *APA Handbook of Research Methods in Psychology, Vol 2: Research Designs: Quantitative, Qualitative, Neuropsychological, and Biological*; APA Handbooks in Psychology®, American Psychological Association: Washington, DC, USA, 2012; pp. 57–71. [CrossRef]

90. Badam, S.K.; Zhao, J.; Sen, S.; Elmqvist, N.; Ebert, D. TimeFork: Interactive Prediction of Time Series. In Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems, San Jose, CA, USA, 7–12 May 2016; ACM: New York, NY, USA, 2016; pp. 5409–5420. [CrossRef]

91. Bögl, M.; Aigner, W.; Filzmoser, P.; Gschwandtner, T.; Lammarsch, T.; Miksch, S.; Rind, A. Visual Analytics Methods to Guide Diagnostics for Time Series Model Predictions. In Proceedings of the 2014 IEEE VIS Workshop on Visualization for Predictive Analytics, Paris, France, 9 November 2014; Volume 1.

92. Ali, M.; Alqahtani, A.; Jones, M.; Xie, X. Clustering and Classification for Time Series Data in Visual Analytics: A Survey. *IEEE Access* **2019**, *7*, 181314–181338. [CrossRef]

93. Sun, D.; Feng, Z.; Chen, Y.; Wang, Y.; Zeng, J.; Yuan, M.; Pong, T.C.; Qu, H. DFSeer: A Visual Analytics Approach to Facilitate Model Selection for Demand Forecasting. In Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems, Honolulu, HI, USA, 25–30 April 2020; ACM: New York, NY, USA, 2020; pp. 1–13. [CrossRef]

94. Kulesza, T.; Stumpf, S.; Burnett, M.; Yang, S.; Kwan, I.; Wong, W.K. Too Much, Too Little, or Just Right? Ways Explanations Impact End Users' Mental Models. In Proceedings of the 2013 IEEE Symposium on Visual Languages and Human Centric Computing, San Jose, CA, USA, 15–19 September 2013; IEEE: San Jose, CA, USA, 2013; pp. 3–10. [CrossRef]

95. Brockwell, P.J.; Davis, R.A. *Introduction to Time Series and Forecasting*; Springer Texts in Statistics; Springer International Publishing: Cham, Switzerland, 2016. [CrossRef]

96. Hyndman, R.J.; Athanasopoulos, G. Forecasting: Principles and Practice; OTexts, 2018. Available online: https://otexts.com/fpp2 (accessed on 8 July 2022).