Received: 29 November 2019 | Revised: 23 December 2019 | Accepted: 23 January 2020

DOI: 10.1002/wics.1503

OVERVIEW

WIREs COMPUTATIONAL STATISTICS

WILEY

Model exploration using conditional visualization

Catherine B. Hurley

Department of Mathematics and Statistics, Maynooth University, Maynooth, Ireland

Correspondence
Catherine B. Hurley, Department of Mathematics and Statistics, Maynooth University, Maynooth Ireland.
Email: catherine.hurley@mu.ie

Abstract

Ideally, statistical parametric model fitting is followed by various summary tables which show predictor contributions, visualizations which assess model assumptions and goodness of fit, and test statistics which compare models. In contrast, modern machine-learning fits are usually black box in nature, offer high-performing predictions but suffer from an interpretability deficit. We examine how the paradigm of conditional visualization can be used to address this, specifically to explain predictor contributions, assess goodness of fit, and compare multiple, competing fits. We compare visualizations from techniques including trellis, condvis, visreg, lime, partial dependence, and ice plots. Our examples use random forest fits, but all techniques presented are model agnostic.

This article is categorized under:

Statistical and Graphical Methods of Data Analysis > Statistical Graphics and Visualization
Statistical Learning and Exploratory Methods of the Data Sciences > Exploratory Data Analysis
Statistical Learning and Exploratory Methods of the Data Sciences > Modeling Methods

KEYWORDS

black box, interaction, machine learning, visualization

1 | INTRODUCTION

Ideally, statistical parametric model fitting is followed by various summary tables which show predictor contributions, visualizations which assess model assumptions and goodness of fit, and test statistics which compare models. Most machine-learning models fit complex algorithms to arbitrarily large datasets. These algorithms are well known to be high on performance and low on interpretability. There are no simple model summary tables and no assumptions to be assessed. While such models are primarily used for prediction, it is important for reasons of trust to explain predictor contributions, assess goodness of fit, and compare competing fits. In this article, we discuss how the technique of conditional visualization can address these issues.

First, we present an overview of conditional visualization techniques for model exploration, including the use of facets, partial dependence (pd) plots (Friedman, 2001), ice plots (Goldstein, Kapelner, Bleich, & Pitkin, 2015), and visualizing regression functions with visreg (Breheny & Burchett, 2017). Then, we move on to the condvis paradigm (O'Connell, Hurley, & Domijan, 2017), which incorporates observations local to the conditioning fit and harnesses the power of interaction. We follow that with a discussion of additive explanations, including nomograms and lime plots. We conclude with a discussion comparing the techniques presented.

All of the methods presented can be described as showing *models in data space* as opposed to more familiar plots of *data in model space* such as residual plots (Wickham, Cook, & Hofmann, 2015). The visualization techniques are essentially model agnostic in that they do not assume any particular model structure, just the availability of a predict function.

Throughout this paper, we use the following notation. Consider a dataset of n observations $\{\mathrm{x}_i, y_i\}$, where $\mathrm{x}_i = (x_{i,1}, \ldots, x_{i,p})$ is a vector of predictors and y_i is the response. Let $\hat{f}$ denote a fitted model. One or two predictors of primary interest are designated as *section predictor(s)* x^S. Remaining predictors are termed *conditioning predictor(s)* x^C.

For comparison purposes, techniques are illustrated using regression random forest fits to the RailTrail dataset available from the R package mosaic (Pruim, Kaplan, & Horton, 2017), though all methods apply also to classification problems. The data has observations which are counts of users passing a sensor located on a trail in Massachusetts measured on 90 days from April to November in 2005. The response is volume of trail users, predictors are daily high and low temperatures (hightemp and lowtemp), season (fall, spring, summer), cloudcover, and dayType (weekday or weekend).

2 | CONDITIONAL VISUALIZATION STRATEGIES

In this section, we present an overview of various conditional visualization techniques for model exploration, namely the use of facets, pd plots (Friedman, 2001), ice plots (Goldstein et al., 2015), and visreg (Breheny & Burchett, 2017). We use random forest fits for the RailTrail dataset to illustrate the concepts, though techniques are model agnostic.

2.1 | Faceting and trellis

Consider a random forest relating volume to hightemp, season, and dayType. As there is one continuous predictor and two factors, a trellis display (Becker, Cleveland, & Shyu, 1996) or faceting plot (Wickham, 2016; Wilkinson, 2005) will show how the fit varies with hightemp, for each level of season and dayType. Figure 1 shows the random forest fit displayed using facets. Spring and summer volume levels are higher than in fall, weekend volumes are higher than weekday. Except in summer, volume increases with hightemp, but volume decreases with hightemp for summer weekends, indicating a three-way interaction.

This is an example of conditional visualization. Figure 1 shows the fit versus hightemp *conditional* on season and dayType. Here, hightemp is the section predictor and season and dayType are the conditioning predictors.

The fitted curves show predictor contributions. As the panels show the observed values of hightemp and volume for each combination of season and dayType, we can see that broadly speaking, the fit captures the patterns evident in the raw data, though the spring fits in particular could be improved. By overlaying a second fit, we could compare its performance to the random forest fit. The trellis concept has also been used to construct displays where each panel is a heatmap visualizing the dependence of a fit on two quantitative predictors, conditional on two further categorical predictors (Nason, Emerson, & LeBlanc, 2004).

Trellis visualizations generalize the faceting concept to the setting where the conditioning variables are quantitative which are binned into overlapping intervals called *shingles*. Trellis and its R implementation Lattice (Sarkar, 2008) focus on displaying data rather than fits. In principle, the fits could be displayed by conditioning on the shingle midpoint or mode. As the dimension of the conditioning space increases, faceting layouts become cumbersome and most of the facets or shingles will be empty.

2.2 | Partial dependence and ice plots

Partial dependence (Friedman, 2001) and ice plots (Goldstein et al., 2015) both use a single plot to visualize how $\hat{f}$ depends on a chosen section predictor x^S. Strictly speaking, partial dependence (pd) plots are not conditional visualizations in the sense that the effects of other predictors, those in x^C, are averaged out.

Ice plots are closely related to pd plots, omitting the averaging step. In the case of a single section predictor, an ice plot displays the curves $\hat{f}(\mathrm{x}^S, \mathrm{x}_i^C)$ versus x^S, letting x^S vary over its range, for each observation i. Observations (x_i^S, y_i) for $i = 1, 2, \ldots, n$ are also plotted.

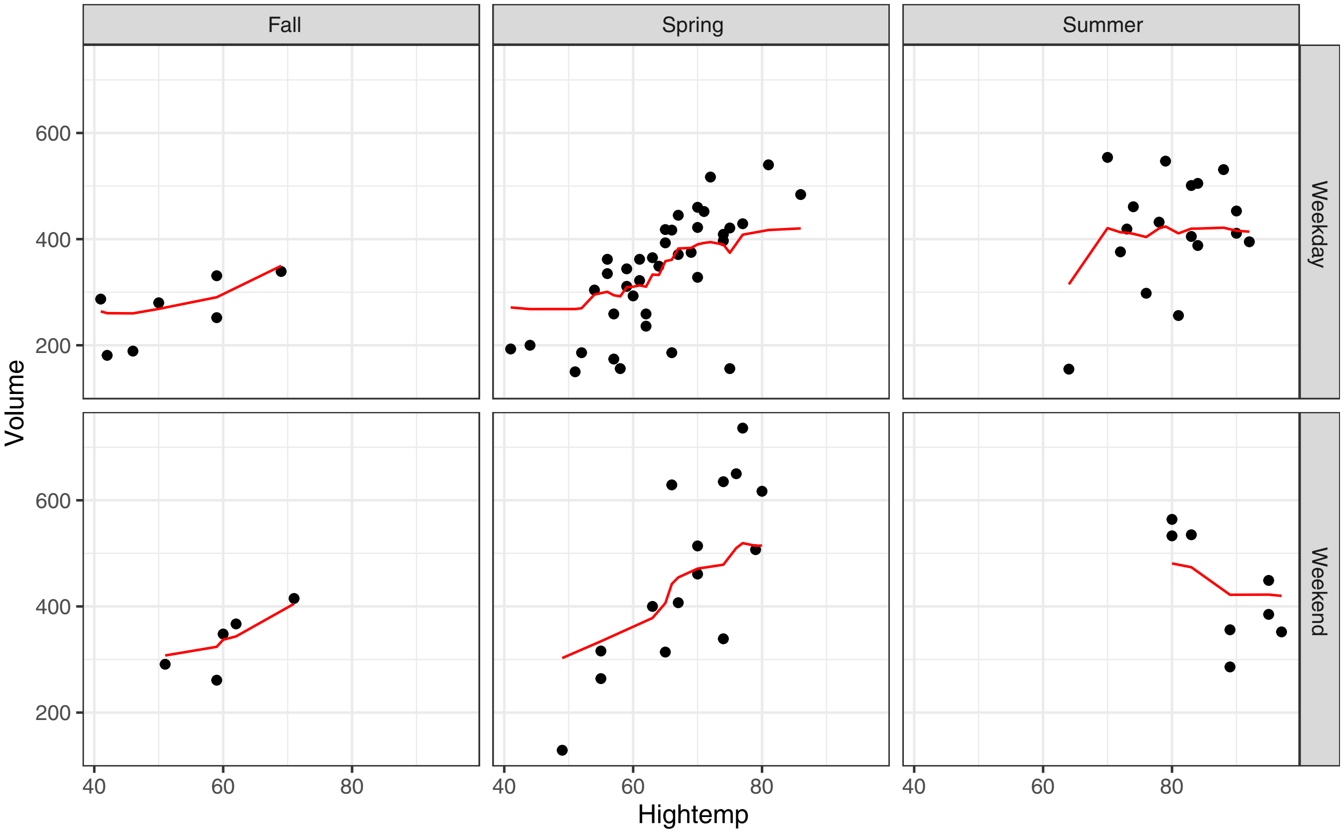

FIGURE 1 Random forest fit relating volume to hightemp, season, and dayType for the RailTrail dataset

Partial dependence plots show the average ice plot curve, displaying $\frac{1}{n}\sum_{i=1}^{n}\hat{f}\left(\mathrm{x}^{S},\mathrm{x}_{i}^{C}\right)$ versus x^{S}. Both pd and ice plots are implemented in the R package ICEbox (Goldstein et al., 2015).

As an example, we fit a random forest to the RailTrail dataset, relating volume to three quantitative predictors hightemp, lowtemp, and precip. Figure 2 displays an ice plot with hightemp as the section variable. Here, each of the n curves is conditional on observed values of lowtemp and precip. The pd curve is shown with a yellow outline. The pd curve captures the increasing volume up to temperatures of 75°F and shows a gentle decrease after that. There is a lack of homogeneity among the ice curves: for example, some flatten out at temperatures of about 65°F. The benefit of ice curves over the summarizing pd curve is apparent here, illustrating the value of conditional methods.

Discovering the explanation for varying patterns of ice curves is more challenging. Does the effect of hightemp on volume vary with lowtemp, precip, or both? Embedding ice plots in an interactive, multiplot display with brushing would help. In fact, in the next section, we discuss how interactive exploration indicates a three-way interaction (see Figure 4).

It is not apparent from the ice plot whether or not the fit is an adequate summary of the data. Also, many of the ice curves involve extrapolation over unrealistic or even impossible values of hightemp. For example, the maximum hightemp value for this dataset is 97°F, and this observation has a lowtemp value of 71°F. The ice curve for this observation is calculated by varying hightemp from 40 to 97°F; clearly values below 71°F are impossible. Generally, correlated predictors cause pd plots to be biased.

Apley (2016) suggested accumulated local effects (ale) plots as a correction for this bias, where local effects are averaged over the conditional distribution of x^{C} given a value for x^{S}. Like pd plots, ale plots summarize the overall pattern of dependence between $\hat{f}$ and x^{S}, but there are no individual curves to show variation in this pattern.

2.3 | Visreg

The visreg package (Breheny & Burchett, 2017) is primarily for displaying regression functions which are additive models, though it is also useful for random forests and support vector machines. Visreg plots display how $\hat{f}$ varies with

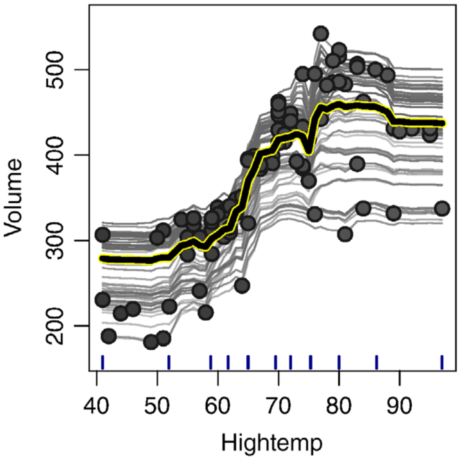

FIGURE 2 An ice plot with partial dependence curve outlined in yellow for hightemp, for the random forest fit to the RailTrail dataset, relating volume to hightemp, lowtemp, and precip

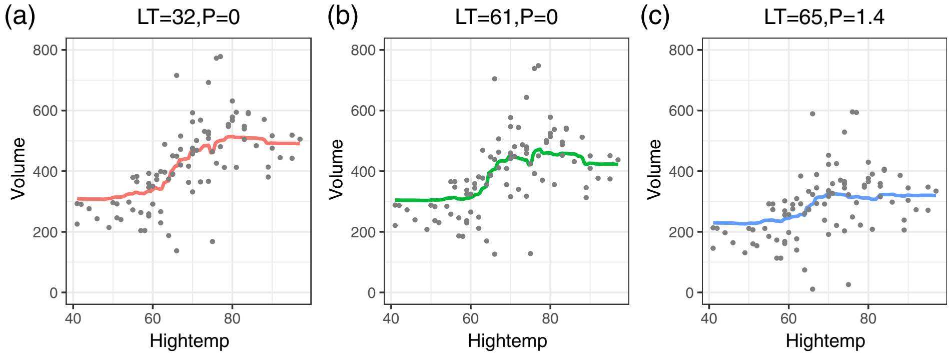

FIGURE 3 Visreg plots of the RailTrail random forest fit relating volume to hightemp, lowtemp (LT), and precip (P). The section variable is hightemp and the conditioning values of (LT,P) are (a) (32,0), (b) (61,0), and (c) (65,1.4)

x^S, for fixed values of $x^C = u$. Specifically, the plot shows $\hat{f}(x^S, x^C = u)$ versus x^S, where x^S varies over its range. The default setting of u is the median for quantitative and mode for factors, though there may be no observations at or near this location. Overlaid point coordinates are partial residuals, which for non-additive fits are defined as $\hat{f}\left(x_i^S, x^C = u\right) + e_i$, where $e_i = y_i - \hat{y}_i$ are residuals, $i = 1, 2, \ldots, n$.

Figure 3 shows visreg plots relating the fit to hightemp, for three different values for the conditioning predictors lowtemp and precip. The plotted partial residuals do not relate to the conditioning values of lowtemp and precip and are the same in all three plots, which gives a misleading picture of goodness of fit. For instance, the plots suggest the fitted curves are biased for hightemp below 60°F, but this is incorrect as we will see later in Figure 4.

Visreg also offers a faceting option, using just one conditioning predictor.

3 | INTERACTIVE CONDITIONAL VISUALIZATION

In O'Connell et al. (2017), the authors described a paradigm for interactive conditional visualization of statistical models. The condvis paradigm is a generalization of the faceting plots in Figure 1 to quantitative as well as categorical conditioning predictors and is appropriate for any fitting algorithm offering predictions. These ideas are implemented for a broad range of supervised and unsupervised learning fits in the shiny-based R package condvis2 (Hurley, O'Connell, & Domijan, 2019). (There is also an earlier, non-shiny-based implementation O'Connell, Hurley, and Domijan (2016)). Here, we give an overview of the basic concept.

3.1 | Condvis concepts

A fixed value of the conditioning predictors $x^C = u$ specifies a section. A section plot then visualizes how the fit $\hat{f}$ varies with x^S, for $x^C = u$. For a single section predictor, the plot shows $\hat{f}(x^S, x^C = u)$ versus x^S, where x^S varies over its range. (Two section predictors are also permitted). The plot also shows observations $\left(x_i^S, y_i\right)$ whose conditioning values x_i^C are near u. This is done by calculating a distance d_i between x_i^C and u, using a Euclidean or maxnorm distance for standardized quantitative predictors. This is transformed into a similarity score

$$s_i = \max\left(0, 1 - \frac{d_i}{\sigma}\right)$$

where $\sigma > 0$ is a threshold parameter set by the user. For categorical predictors, similarity is set to zero when the factor levels of x_i^C and u do not match. Alternatively, a Gower-type distance (Gower, 1971) is available which combines distances for quantitative and categorical predictors into a single value. The color of points is faded in proportion to the similarity score, where points with a zero score are not shown.

In the interactive implementations in O'Connell et al. (2016) and Hurley et al. (2019), the section points u are chosen by clicking on one- or two-dimensional displays of the conditioning predictors.

3.2 | One section predictor

Figure 4 shows how the fit relates to hightemp, for three different values for the conditioning predictors lowtemp and precip, shown in Figure 5.

The plots in Figure 4 show that volume increases for hightemp from 60°F. In each panel, the volume drops off or flattens for higher temperatures. The third panel has only a few points because it represents the fit for days where the weather conditions are unusually hot and wet. The predicted fit is lower than in Figure 4a,b, though as there is very little data, the fit is largely based on extrapolation. There appears to be evidence of lack of fit in the first panel for hightemp <60°F, as points lie predominantly below the curve. However, the light point color indicates these points are not that close to the section point shown as the red cross in Figure 5. Interactive exploration verifies that these points correspond to days with some rain. Generally, moving the section point by clicking on a plot of precip and lowtemp will show predicted volume reduces as rain increases. Comparing Figure 4a,b, the differing curve shapes indicate an interaction between hightemp and lowtemp, while a comparison of b and c suggest a

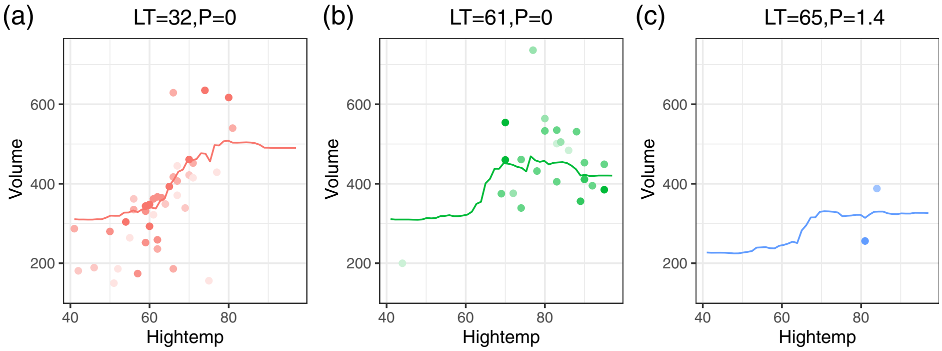

FIGURE 4 Condvis plots of the RailTrail random forest fit relating volume to hightemp, lowtemp (LT), and precip (P). The section variable is hightemp and the conditioning values of (LT,P) are (a) (32,0), (b) (61,0), and (c) (65,1.4)

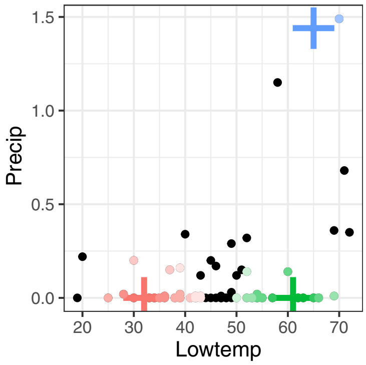

FIGURE 5 The conditioning values of lowtemp and precip used in Figure 4, marked with a cross. Red (green, blue) points near the red (green, blue) cross are visible in Figure 4a–c

hightemp, precip interaction. Interactive exploration suggests three-way interaction, but there may not be enough data so support this conclusion.

We note that the curves displayed in Figure 4 are the same as those in Figure 3, but the latter shows (misleading) partial residuals instead of observations near the section point $\mathrm{x}^C = \mathrm{u}$. The ice plots of Figure 2 are also closely related to condvis displays, in that ice plots show fits where the section points are the n observations, overlayed in a single display.

3.3 | Two section predictors

There can be two section variables, which if both are numerical are displayed as a heatmap or surface plot. If one of the section predictors is a factor, the curves can be overlaid.

Again for the RailTrail dataset, we fit a random forest relating volume to hightemp, lowtemp, precip, cloudcover, dayType, and season. Figure 6 shows how the fit relates to two section predictors hightemp and dayType for three different values for the conditioning predictors lowtemp, precip cloudcover, and season. The conditioning predictors (not shown) could be plotted as a parallel coordinate plot or two pairwise plots.

The most interesting pattern here is that the volumes are higher at weekends (purple) for the first panel only, where the condition is low lowtemp, low cloudcover, and spring. This weekend bonus reduces as either lowtemp or precip increase.

The condvis strategy also permits multiple fits which can be overlaid or positioned on a grid. Confidence and prediction intervals can be shown if offered by the fitting algorithm.

As the dimension of the conditioning space increases, interactively selecting section points avoiding near-empty sections is tricky. One strategy limits interactive exploration to predictors identified as important by some measure of variable importance. Another strategy as proposed in O'Connell (2017) designs tours of conditioning space to present *interesting* section plots. These tours consist of visualizing the fit for a prechosen sequence of sections $\mathrm{x}^C = \{\mathrm{u}_1, \mathrm{u}_2, ..., \mathrm{u}_t\}$ where t is the length of the tour. Tours are implemented in Hurley et al. (2019), where tours on offer include u_j randomly chosen from observed values of the conditioning predictors, formed as centroids (or medoids) from k-means (or k-medoids) of the conditioning predictors. Alternatively, tours may be chosen to highlight lack of fit, or differences between two or more fits.

3.4 | Core plots

Cook (1995) proposed an idea related to condvis which he called CORE (for conditional regression plots). These plots display $\left(\mathrm{x}_i^S, y_i\right)$ for points i such that $\mathrm{x}_i^C \in \mathcal{N}(u)$, though did not include plots of $\hat{f}$. Regions $\mathcal{N}$ were to be formed by brushing plots of at most two predictors, which limited the utility of CORE methods to low-dimensional settings. To overcome this limitation, he proposed using sufficient dimension reduction methods to reduce the dimension of the conditioning space. These concepts are further developed in Zhang (2013). This would seem to be an alternative to tours for exploring interesting areas of predictor space.

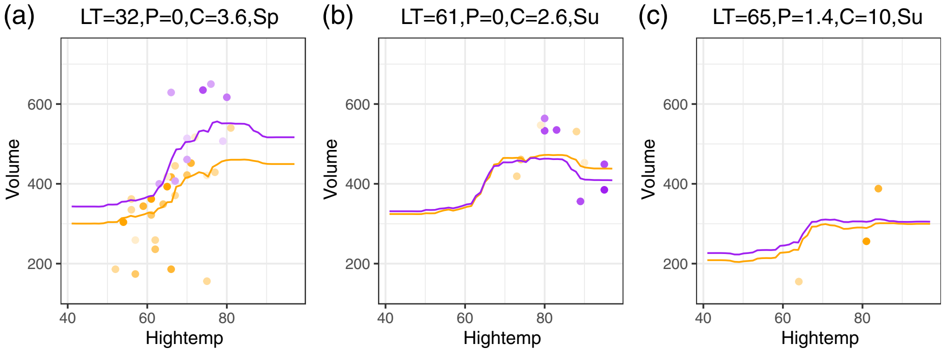

FIGURE 6 Condvis plots of the RailTrail random forest fit relating volume to hightemp, dayType, lowtemp (LT), precip (P), cloudcover (C), and season (S). The section variables are hightemp and dayType and the conditioning values of (LT,P, C,S) are (a) (32,0, 3.6, spring), (b) (61,0, 2.6, summer), and (c) (65,1.4, 10, summer). Weekend points and fits are in purple, weekday points are in orange

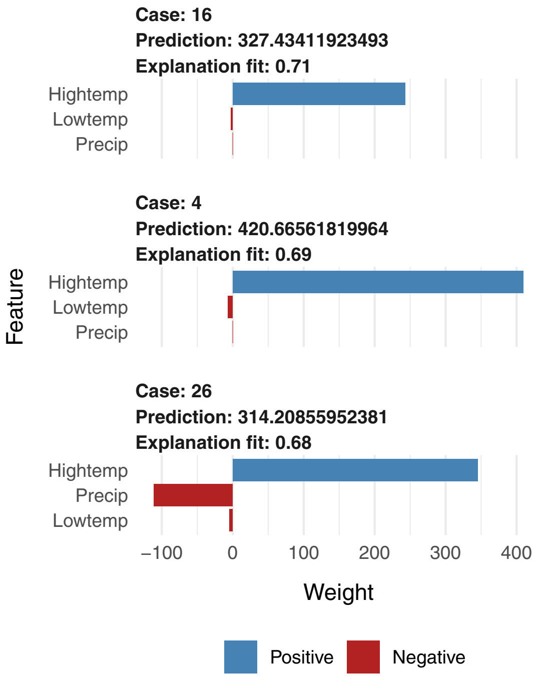

FIGURE 7 A lime explainer plot for the RailTrail random forest fit relating volume to hightemp, lowtemp, and precip. Cases 16, 4, and 26 have predictor values (hightemp, lowtemp, and precip) of (a) (54, 32,0), (b) (96, 61,0), and (c) (81, 65,1.4)

4 | ADDITIVE EXPLANATIONS

4.1 | Nomograms

In linear regression problems, nomograms have been used as a chart to assist in hand calculation of predictions, assigning a score to each predictor setting, and then obtaining the prediction by manually totaling the scores. See for example the nomogram function in the R package rms (Harrell Jr, 2019). Also, Jakulin, Možina, Demšar, Bratko, and Zupan (2005) give a nomogram construction for support vector machines.

The R package DynNom (Jalali, Alvarez-Iglesias, Roshan, & Newell, 2019; Jalali, Roshan, Alvarez-Iglesias, & Newell, 2019) offers dynamic nomograms for a range of additive regression models. Using the terminology from the earlier section, all predictors are designated as conditioning predictors whose values $\mathrm{x}^C = \mathrm{u}$ are interactively set by the user, causing a plot to show the corresponding prediction and confidence interval. As DynNom does not show individual predictor contributions to the overall prediction as in a standard nomogram, there is no reason why these dynamic nomograms could not be extended to (nonadditive) arbitrary black-box models. This would be useful for calculating and comparing predictions for different values of $\mathrm{x}^C = \mathrm{u}$. But, such nomograms would not show how a fit depends on section variables, as was shown by condvis in Figure 4 and visreg in Figure 3, so they do not assist in explaining or understanding a fitted model.

4.2 | Additive approximations with lime

Some authors (Nugent and Cunningham (2005) and Ribeiro, Singh, and Guestrin (2016)) have proposed using locally linear fits to derive explanations for fits from machine-learning models. In their setup, all predictors are designated as conditioning predictors. The lime algorithm of Ribeiro et al. (2016) fits a local ridge regression $\hat{\hat{f}}$ at $\mathrm{x}^C = \mathrm{u}$ relating $\hat{f}$ to predictors using nearby sampled data weighted by a similarity score. In R, this algorithm is provided by package lime (Pedersen & Benesty, 2019).

In Figure 7 we show explanations given by lime for the fit to three observations. As before, our fit is a random forest relating volume to hightemp, lowtemp, and precip, for the RailTrail dataset. Each panel of the plot shows the local fit $\hat{\hat{f}}$ in a nomogram-type display. The bars represent $\hat{\beta}_j u_j$, the local contribution of each predictor, and the local R^2 is also reported.

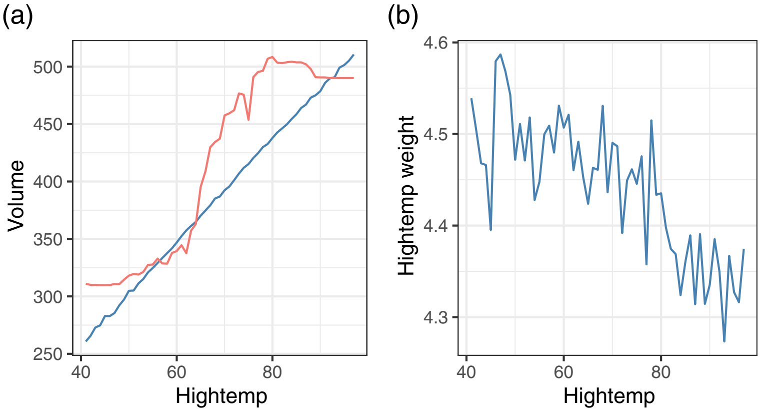

FIGURE 8 For the RailTrail random forest fit relating volume to hightemp, lowtemp, and precip, plot (a) compares the lime explainer fit in blue to the random forest fit in red, varying hightemp. Plot (b) shows the weights for hightemp in the lime local linear approximation. Both plots fix lowtemp = 32 and precip = 0

Note that the predictor settings chosen have identical lowtemp and precip values to those used for Figures 4 and 3. For the three conditions, the lime explainer shows the positive effect of hightemp on predicted volume. The bar length for hightemp represents how much the explainer fit changes if hightemp is dropped from its current value to zero, keeping lowtemp and precip unchanged. But from the three panels of Figure 4, we see that the chosen values for hightemp occur in locations where the predicted hightemp is fairly flat, so the lime explainer is misleading. Furthermore, a superficial interpretation of Figure 7 might suggest that precip is important only for case 26, but the apparent noncontribution for cases 16 and 4 in Figure 7 occurs because they both have precip values of zero. Again, exploration of the fit with condvis shows that increasing precip generally reduces predicted volume.

The lime algorithm has quite a few parameters whose settings affect the given explanations, including a choice of feature selection algorithms. The methodology is not concerned with validation of the fitted model $\hat{f}$ or with uncovering its deficiencies. Lime sets out to give simple explanations, but these may be based on poor approximations $\hat{\hat{f}}$ as in Figure 8a, and the explanations themselves may be unstable, as in Figure 8b.

Lime explanations also do not allow for local interactions. A recently proposed alternative (Ribeiro, Singh, & Guestrin, 2018) uses rule-based instead of linear fits which have the benefit of incorporating feature interactions.

4.3 | Shapley methods

Like lime, Shapley methods construct an explainer $\hat{\hat{f}}$ that is a sum of feature contributions. But these are not based on regression, rather equations from competitive game theory. Shapley values use a contribution for feature j that measures the effect on prediction of adding feature j to all 2^{p-1} subsets of other features, combining these in a weighted sum. Štrumbelj and Kononenko (2014) give details and describe a sampling algorithm which reduces the exponential time complexity. Visualizations of Shapley values are similar to those for lime. Lundberg and Lee (2017) and Lundberg et al. (2020) propose other approximations and some interesting new model visualizations which are beyond the scope of the current paper.

5 | DISCUSSION

All of the methods presented here apply to any black-box fitting algorithm, in the classification and regression setting. All methods attempt to show how $\hat{f}$ relates to predictors. Additive explainer methods do not at present use a designated section predictor or predictors, and so cannot show directly how a fit varies as a predictor changes. However, lime visualizations such as that in Figure 8 could be constructed varying hightemp say, and keeping other predictors fixed.

Only condvis and DynNom use the power of interaction to facilitate model exploration, but interactive control of section points could be added to other methods.

Condvis visualizations facilitate goodness-of-fit assessments using overlaid nearby weighted observations. Other methods do not support goodness-of-fit assessments. Condvis also supports the comparison of multiple fits, either via overlaid or side-by-side curves or surfaces which is not available in implementations of other methods discussed, though this feature could be added. Confidence and prediction intervals for the response are available for condvis and visreg displays and also for nomograms. Lime displays could be enhanced with an uncertainty measure, but these are for the approximate fit rather than the fit itself.

In principle, all of the methods presented here are suitable for big n datasets. Condvis section plots show just high-similarity observations, plots of conditioning predictors can use subsets of $n' < < n$ observations. Similarly, visreg could just use subsets, and ice plots could show a selection of n' curves.

Large p datasets present a bigger challenge for conditional visualization techniques. A partial solution is to use variable importance measures to identify features that drive predictions. The most important predictor is the natural choice of section predictor for condvis, visreg, and pdp/ice plots, though in condvis the choice is under interactive control. For condvis, interactive exploration can be focused on the important predictor subset, while the conditioning values for predictors deemed to be of little importance could be fixed at the median for numeric predictors and mode for factors. Lime offers built-in feature selection algorithms, based on standard linear regression techniques. As observations are sparse in high dimensions, unless section points are carefully selected sections will be empty. Condvis software includes tours of conditioning space to locate interesting section plots. As these tours simply identify a set of relevant section points, these could also be used to construct visreg or ice plot displays.

CONFLICT OF INTEREST

The author has declared no conflicts of interest for this article.

ORCID

Catherine B. Hurley https://orcid.org/0000-0003-2758-5531

RELATED WIREs ARTICLE

New developments for net-effect plots

REFERENCES

Apley, D. W. (2016). Visualizing the effects of predictor variables in black box supervised learning models. *arXiv e-prints*, arXiv:1612.08468.

Becker, R. A., Cleveland, W. S., & Shyu, M.-J. (1996). The visual design and control of trellis display. *Journal of Computational and Graphical Statistics*, *5*(2), 123–155. https://doi.org/10.1080/10618600.1996.10474701

Breheny, P., & Burchett, W. (2017). Visualization of regression models using visreg. *The R Journal*, *9*(2), 56–71. https://doi.org/10.32614/RJ-2017-046

Cook, R. D. (1995). Graphics for studying net effects of regression predictors. *Statistica Sinica*, *5*(2), 689–708 Retrieved from http://www.jstor.org/stable/24305064

Friedman, J. H. (2001). Greedy function approximation: A gradient boosting machine. *The Annals of Statistics*, *29*(5), 1189–1232 Retrieved from http://www.jstor.org/stable/2699986

Goldstein, A., Kapelner, A., Bleich, J., & Pitkin, E. (2015). Peeking inside the black box: Visualizing statistical learning with plots of individual conditional expectation. *Journal of Computational and Graphical Statistics*, *24*(1), 44–65. https://doi.org/10.1080/10618600.2014.907095

Gower, J. C. (1971). A general coefficient of similarity and some of its properties. *Biometrics*, *27*(4), 857–871 Retrieved from http://www.jstor.org/stable/2528823

Harrell Jr, F. E. (2019). rms: Regression modeling strategies [Computer software manual]. Retrieved from https://CRAN.R-project.org/package=rms (R package version 5.1–3.1)

Hurley, C., O'Connell, M., & Domijan, K. (2019). *Condvis2: Conditional visualization for supervised and unsupervised models in shiny [Computer software manual]* (R package version 0.1.1). Retrieved from https://CRAN.R-project.org/package=condvis

Jakulin, A., Možina, M., Demšar, J., Bratko, I., & Zupan, B. (2005). Nomograms for visualizing support vector machines. In *Proceedings of the eleventh acm sigkdd international conference on knowledge discovery in data mining* (pp. 108–117). New York, NY: ACM.

Jalali, A., Alvarez-Iglesias, A., Roshan, D., & Newell, J. (2019, 11). Visualising statistical models using dynamic nomograms. *PLoS One*, *14*(11), 1–15. https://doi.org/10.1371/journal.pone.0225253

Jalali, A., Roshan, D., Alvarez-Iglesias, A., & Newell, J. (2019). *Dynnom: Visualising statistical models using dynamic nomograms [Computer software manual]* (R package version 5.0.1). Retrieved from https://CRAN.R-project.org/package=DynNom

Lundberg, S. M., Erion, G., Chen, H., DeGrave, A., Prutkin, J. M., Nair, B., … Lee, S.-I. (2020). From local explanations to global understanding with explainable AI for trees. *Nature Machine Intelligence*, *2*(1), 56–67.

Lundberg, S. M., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. In I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, & R. Garnett (Eds.), *Advances in neural information processing systems* (Vol. 30, pp. 4765–4774). Curran Associates, Inc Retrieved from http://papers.nips.cc/paper/7062-a-unified-approach-to-interpreting-model-predictions.pdf

Nason, M., Emerson, S., & LeBlanc, M. (2004). Cartscans: A tool for visualizing complex models. *Journal of Computational and Graphical Statistics*, *13*(4), 807–825. Retrieved from https://doi.org/10.1198/106186004X11417

Nugent, C., & Cunningham, P. (2005, October). A case-based explanation system for black-box systems. *Artificial Intelligence Review*, *24*(2), 163–178. https://doi.org/10.1007/s10462-005-4609-5

O'Connell, M. (2017). *Conditional visualisation for statistical models* (Doctoral dissertation). National University of Ireland Maynooth. Retrieved from http://mural.maynoothuniversity.ie/8141/

O'Connell, M., Hurley, C., & Domijan, K. (2016). *Condvis: Conditional visualization for statistical models [Computer software manual]* (R package version 0.1.1). Retrieved from https://CRAN.R-project.org/package=condvis

O'Connell, M., Hurley, C., & Domijan, K. (2017). Conditional visualization for statistical models: An introduction to the condvis package in R. *Journal of Statistical Software, Articles*, *81*(5), 1–20. https://doi.org/10.18637/jss.v081.i05

Pedersen, T. L., & Benesty, M. (2019). *Lime: Local interpretable model-agnostic explanations [Computer software manual]* (R package version 0.5.0). Retrieved from https://CRAN.R-project.org/package=lime

Pruim, R., Kaplan, D. T., & Horton, N. J. (2017). The mosaic package: Helping students to 'think with data' using R. *The R Journal*, *9*(1), 77–102 Retrieved from https://journal.r-project.org/archive/2017/RJ-2017-024/index.html

Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). Why should I trust you? Explaining the predictions of any classifier. In *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining* (pp. 1135–1144). San Francisco, CA, August 13–17, 2016.

Ribeiro, M. T., Singh, S., & Guestrin, C. (2018). *Anchors: High-precision model-agnostic explanations*. Retrieved from https://www.aaai.org/ocs/index.php/AAAI/AAAI18/paper/view/16982

Sarkar, D. (2008). *Lattice: Multivariate data visualization with R*. New York, NY: Springer Retrieved from http://lmdvr.r-forge.r-project.org (ISBN 978-0-387-75968-5)

Štrumbelj, E., & Kononenko, I. (2014). Explaining prediction models and individual predictions with feature contributions. *Knowledge and Information Systems*, *41*(3), 647–665. Retrieved from https://doi.org/10.1007/s10115-013-0679-x

Wickham, H. (2016). *ggplot2: Elegant graphics for data analysis*. New York, NY: Springer-Verlag Retrieved from https://ggplot2.tidyverse.org

Wickham, H., Cook, D., & Hofmann, H. (2015). Visualizing statistical models: Removing the blindfold. *Statistical Analysis and Data Mining: The ASA Data Science Journal*, *8*(4), 203–225. https://doi.org/10.1002/sam.11271

Wilkinson, L. (2005). *The grammar of graphics (statistics and computing)*. Secaucus, NJ: Springer-Verlag New York, Inc.

Zhang, X. (2013). New developments for net-effect plots. *WIREs: Computational Statistics*, *5*(2), 105–113. https://doi.org/10.1002/wics.1247

How to cite this article: Hurley CB. Model exploration using conditional visualization. *WIREs Comput Stat*. 2021;13:e1503. https://doi.org/10.1002/wics.1503