Interpretation of dam deformation and leakage with boosted regression trees

Fernando Salazar[1], Miguel Á. Toledo[2], Eugenio Oñate[1], Benjamín Suárez[1]

Abstract

Predictive models are essential in dam safety assessment. They have been traditionally based on simple statistical tools such as the hydrostatic-season-time (HST) model. These tools are well known to have limitations in terms of accuracy and reliability. In the recent years, the examples of application of machine learning and related techniques are becoming more frequent as an alternative to HST. While they proved to feature higher flexibility and prediction accuracy, they are also more difficult to interpret. As a consequence, the vast majority of the research is limited to prediction accuracy estimation. In this work, one of the most popular machine learning techniques (boosted regression trees), was applied to model 8 radial displacements and 4 leakage flows at La Baells Dam. The possibilities of model interpretation were explored: the relative influence of each predictor was computed, and the partial dependence plots were obtained. Both results were analysed to draw conclusions on dam response to environmental variables, and its evolution over time. The results show that this technique can efficiently identify dam performance changes with higher flexibility and reliability than simple regression models.

Keywords:
machine learning, dam safety, dam monitoring, boosted regression trees

1. Introduction

Dam monitoring is essential to ensure its proper operation and its long-term safety [1]. One of the main tasks to be carried out is the comparison between the expected response and that registered by the monitoring system, to understand the dam behaviour and to detect potential anomalies. In this context, predictive models are necessary to estimate the dam response in a given situation.

Data-based tools allow building predictive models based on monitoring data, i.e., without explicitly considering the physical properties of the dam and the foundation. The hydrostatic-season-time (HST) model [2] is the most widely applied, and the only generally accepted by practitioners.

*Corresponding author: F. Salazar
Email addresses: fsalazar@cimne.upc.edu (Fernando Salazar), matoledo@caminos.upm.es (Miguel Á. Toledo), onate@cimne.upc.edu (Eugenio Oñate), benjamin.suarez@upc.edu (Benjamín Suárez)
[1]International Center for Numerical Methods in Engineering (CIMNE). Campus Norte UPC. Gran Capitán s/n. 08034. Barcelona, Spain
[2]Technical University of Madrid (UPM). Civil Engineering Department: Hydraulics, Energy and Environment. Profesor Aranguren s/n, 28040, Madrid, Spain

HST is based on multiple linear regression considering the three most influential external variables: hydrostatic load, air temperature and time. The main advantages of HST are:

1. It frequently provides useful estimations of displacements in concrete dams [3].
2. It is simple and thus easily interpretable: the effect of each external variable can be isolated in a straightforward manner, since they are cumulative.
3. Since the thermal effect is considered as a periodic function, the time series of air temperature are not required. This widens the possibilities of application, as only the reservoir level variation is needs to be available to build an HST model.
4. It is well known by practitioners and frequently applied in several countries [3].

Nonetheless, HST also features conceptual limitations that damage the prediction accuracy [3] and may lead to misinterpretation of the results [4]. For example, it is based on the assumption that the hydrostatic load and the temperature are independent, whereas it is obviously not the case: the thermal field in the dam body, especially in the vicinity of the water surface, is strongly dependant on the water temperature in the upstream face [5]. In turn, the thermal load influences the stress and displacement fields.

Several modifications to the original HST model have been proposed to overcome these drawbacks. They focus on improving the consideration of the thermal load, by taking into account the actual air temperature instead of the historical mean [6], or the effect of the water temperature on the upstream face [3], [7].

In the recent years, non-parametric techniques have emerged as an alternative to HST for building data-based behaviour models [8], e.g. support vector machines (SVN) [9], neural networks (NN) [10], adaptive neuro-fuzzy systems (ANFIS) [11], among others [8]. In general, these tools are more suitable to model non-linear cause-effect relations, as well as interaction among external variables, as that previously mentioned between hydrostatic load and temperature. On the contrary, they are typically more difficult to interpret, what led them to be termed as "black box" models (e.g. [12]).

Most of the published works focused on building predictive models whose accuracy was generally higher than that offered by HST (e.g. [10], [13], [14]). Since the resulting model was seldom analysed, little information was provided for dam safety assessment. Some exceptions worth mentioning, though simple, were due to Santillán *et al.* [15], Mata [10] and Cheng and Zheng [16].

Therefore, dam engineers face a dilemma: the HST model is widely known and used and easily interpretable. However, it is based on some incorrect assumptions, and its accuracy can be increased. On the other hand, more flexible and accurate models are available, but they are more difficult to implement and analyse. The same problem arose in the field of statistics [17].

The objective of this work is to investigate the possibilities of interpretation of one of these black box models to:

1. Identify the effect of each external variable on the dam behaviour
2. Detect the temporal evolution of the dam response
3. Provide meaningful information to draw conclusions about dam safety

Among the plethora of machine learning techniques available [18], a previous comparative study [13] showed boosted regression trees (BRT) as one of the more appropriate tools for

the prediction of dam response. In this paper, the technique was further explored, with focus on the interpretation of the results for dam behaviour identification. In particular, the partial dependence plots were examined to isolate the effect of each action, and the relative influence (RI) was computed to identify the strength of each input-output relation. Furthermore, the results were interpreted from an overall viewpoint to draw conclusions on the dam behaviour.

The method was applied to the analysis of La Baells Dam, as compared to the conventional HST model.

The rest of the paper is organised as follows. A brief introduction to BRT is presented, including the methods for interpretation. Then, the case study and the HST version taken as reference are described. The results are included and interpreted in terms of the dam behaviour, and the differences between both methods are discussed.

2. Methods

2.1. Boosted regression trees

The objective of a predictive model is to estimate the value of an output variable $Y \in \mathbb{R}$ (i.e. radial displacement or leakage), based on a set of predictors (reservoir level, air temperature, etc.) $X \in \mathbb{R}^p$, i.e. $Y \approx \hat{Y} = F(X)$. The observed values are denoted as $(x_i, y_i), i = 1, ..., N$, where N is the number of observations. Note that each x_i is a vector with p components, each of which is referred to as x_i^j, when necessary. Similarly, $X^j, j = 1, ..., p$ stands for each dimension of the input space.

BRT models are built by combining two algorithms: a set of single models are fitted by means of decision trees [19], and their output is combined to compute the overall prediction using boosting [20]. For the sake of completeness, a short description of both techniques follow, although excellent introductions can be found in [21], [22], [23], [12].

2.1.1. Regression trees

Regression trees were first proposed as statistical models by Breiman *et al.* [19]. They are based on the recursive division of the training data in groups of "similar" cases. The output of a regression tree is the mean of the output variable for the observations within each group.

When more than one predictor is considered (as usual), the best split point for each is computed, and the one which results in greater error reduction is selected. As a consequence, non-relevant predictors are automatically discarded by the algorithm, as the error reduction for a split in a low relevant predictor will generally be lower than that in an informative one.

Other interesting properties of regression trees are:

- They are robust against outliers.
- They require little data pre-processing.
- They can handle numerical and categorical predictors.
- They are appropriate to model non-lineal relations, as well as interaction among predictors.

By contrast, regression trees are unstable, i. e., small variations in the training data lead to notably different results. Also, they are not appropriate for certain input-output relations, such as a straight 45º line [23].

2.1.2. Boosting

Boosting is a general scheme to build ensemble prediction models [20]. It is based on the generation of a (frequently high) number of simple models (also referred to as "weak learners") on altered versions of the training data. The overall prediction is computed as a weighted sum of the output of each model in the ensemble. The rationale behind the method is that the average of the prediction of many simple learners can outperform that from a complex one [24].

The idea is to fit each learner to the residual of the previous ensemble. The main steps of the original boosting algorithm when using regression trees and the squared-error loss function can be summarised as follows [25]:

1. Start predicting with the average of the observations (constant):

$$F_0(X) = f_0(X) = \bar{y}_i$$

2. For $m = 1$ to M
   (a) Compute the prediction error on the training set:

$$\tilde{y}_i = y_i - F_{m-1}(x_i)$$

   (b) Draw a random sub-sample of the training set (S_m)
   (c) Consider S_m and fit a new regression tree to the residuals of the previous ensemble:

$$\tilde{y}_i \approx f_m(X), i \in S_m$$

   (d) Update the ensemble:

$$F_m(X) \Leftarrow F_{m-1}(X) + f_m(X)$$

3. F_M is the final model

It is generally accepted that this procedure is prone to over-fitting, because the training error decreases with each iteration [25]. To overcome this problem, it is convenient to add a regularization parameter $\nu \in (0, 1)$, so that step (d) turns into:

$$F_m(X) \Leftarrow F_{m-1}(X) + \nu \cdot f_m(X)$$

Some empirical analyses showed that relatively low values of ν (below 0.1) greatly improve generalisation capability [20]. In practice, it is common to set the regularisation parameter and consider a number of trees such that the training error stabilises [21]. Subsequently, a certain number of terms are pruned using for example cross-validation. This is the approach employed in this work, with $\nu = 0.001$ and a maximum of 10,000 trees. It was verified that the training error reached the minimum before adding the maximum number of trees.

Five-fold cross-validation was applied to determine the amount of trees in the final ensemble. The process was repeated using trees of depth 1 and 2 (*interaction.depth*), and the most accurate for each target was selected. The rest of the parameters were set to their default values [26].

All the calculations were performed in the R environment [27].

2.2. Model interpretation

2.2.1. Relative influence (RI)

BRT models are robust against the presence of uninformative predictors, as they are discarded during the selection of the best split. Moreover, it seems reasonable to think that the most relevant predictors are more frequently selected during training. In other words, the relative influence (RI) of each input is proportional to the frequency with which they appear in the ensemble. Friedman [20] proposed a formulation to compute a measure of RI for BRT models based on this intuition. Both the relative presence and the error reduction achieved are considered in the computation. The results are normalised so that they add up to 100.

Based on this measurement, the most influential variables were identified for each output, and the results were interpreted in relation to dam behaviour. In order to facilitate the analysis, the RI was plotted as word clouds [28]. These plots resemble histograms, with the advantage of being more appropriate to visualise a greater set of variables. The code representing each predictor was displayed with a font size proportional to its relative influence with the library "wordcloud" [29].

Furthermore, two degrees of variable selection were applied, based on the RI of each predictor. First, a BRT model (M1) was trained with all the variables considered (section 2.4). Second, the inputs with $RI\left(X^j\right) > min\left(RI\left(X^j\right)\right) + sd\left(RI\left(X^j\right)\right)$ were selected to build a new model (M2). This criteria is heuristic and based on the *1-SE rule* proposed by Breiman *et al.* [19]. Finally, a model with three predictors was generated (M3), featuring the more relevant variables of each group: temperature, time and reservoir level for radial displacements, and rainfall, time and level for leakage flows.

These three versions were generated to analyse the effect of the presence of uninformative variables in the predictor set. Moreover, the simplest model facilitates the analysis, as the effect of each action is concentrated in one single predictor.

In this sense, the temporal evolution is particularly relevant for dam safety evaluation, as it can help to identify a progressive deterioration of the dam or the foundation, which might result in a serious fault if not corrected.

2.2.2. Partial dependence plots

Multi-linear regression models and HST in particular are based on the assumption that the input variables are statistically independent, so the prediction is computed as the sum of their contributions. As a result, the effect of each predictor in the response can be easily identified, by plotting $f(X^j), \forall j = 1...p$.

This method is not appropriate for BRT models, as interactions among predictors are accounted for. While this results in more flexibility, it also implies that the identification of the relation between predictors and response is not straightforward.

Nonetheless, it is possible to examine the predictor-response relationship by means of the partial dependence plots [20]. This tool can be applied to any black box model, as it is based on the marginal effect of each predictor on the output, as learned by the model. Let X^j be the variable of interest. A set of equally spaced values are defined along its range: $X^j = x_k^j$. For each of those values, the average of the model predictions is computed:

$$\bar{F}\left(x_k^j\right) = \frac{1}{N}\sum_{i=1}^{N} F\left(x_k^j, x_i^{jc}\right) \tag{1}$$

where x_i^{jc} is the value for all inputs other than X^j for the observation i.

Similar plots can be obtained for interactions among inputs: the average prediction is computed for couples of fixed x_k^j, where j takes two different values. Hence, the results can be plotted as a three-dimensional surface (section 3.3). In this work, partial dependence plots were restricted to the simplest model, which considered three predictors. Therefore, three 3D plots allowed investigating the pairwise interactions for all the inputs considered in the simplified model.

2.3. HST model

A conventional HST model was fitted for comparison purposes:

$$\begin{aligned} \hat{Y} = F(t,h,s) &= a_0 + a_1 h + a_2 h^2 + a_3 h^3 \\ &\quad + a_4 h^4 + a_5 h^5 + a_6 e^{-t} \\ &\quad + a_7 t + a_8 cos(s) + a_9 sin(s) \\ &\quad + a_{10} sin^2(s) + a_{11} sin(s) cos(s) \end{aligned} \tag{2}$$

where

$$s = \frac{d}{365,25} 2\pi \tag{3}$$

where d is the number of days since 1 January, t is the elapsed time (years), h is the reservoir level, and $a_1, a_2, ..., a_{11}$ are the coefficients to fit.

The contribution of each action can be computed by adding the correspondent terms:

$$\hat{Y}_h = a_1 h + a_2 h^2 + a_3 h^3 + a_4 h^4 + a_5 h^5 \tag{4}$$

$$\begin{aligned} \hat{Y}_s &= a_8 cos(s) + a_9 sin(s) \\ &\quad + a_{10} sin^2(s) + a_{11} sin(s) cos(s) \end{aligned} \tag{5}$$

$$\hat{Y}_t = a_5 e^{-t} + a_6 t \tag{6}$$

This model was also employed to check the reliability of the temporal behaviour identified by BRT models for some devices. After an HST model was fitted to the training data, a modified version of the time series of the target variable was generated by removing the temporal term ($\hat{Y}_t$) and adding random noise of zero mean and a standard deviation equal to 0.5 (mm for displacements; l/min for leakage):

$$Y_{mod} = \hat{Y}_h + \hat{Y}_s + N(0, 0.5) \tag{7}$$

The result is a time series whose dependence from the temperature and the level approximates that of the actual displacement, while being totally time-independent.

2.4. Case study

The data used for the study correspond to La Baells Dam. It is a double curvature arch dam, with a height of 102 m, which entered into service in 1976. Among the available records, the study focused on 12 variables: 8 corresponded to radial displacements measured

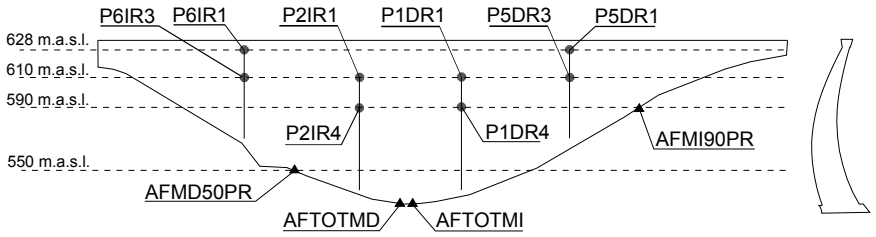

Figure 1: Geometry and location of the monitoring devices in La Baells Dam. Left: view from downstream. Right: highest cross-section.

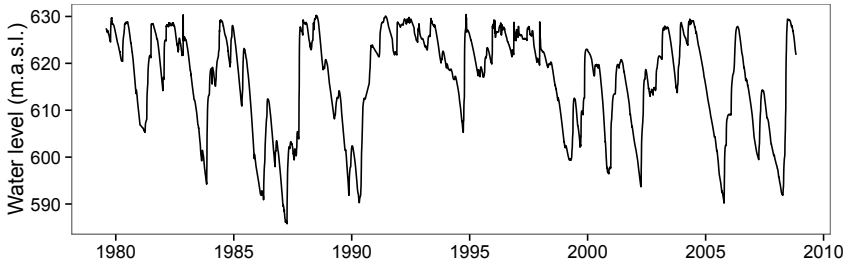

Figure 2: Time series of the reservoir level at La Baells Dam.

by pendulums (along the upstream-downstream direction), and four to leakage flow. The location of each monitoring device is depicted in Figure 1.

As for the environmental variables, the mean daily air temperature, the reservoir level and the daily rainfall were available. Figure 2 depicts the reservoir level variation in the period considered, whereas the other two are included in the Appendix (Figures A1 and A2).

Since BRT models automatically discard those predictors not associated with the output [30], the initial model considered a relatively large set of inputs. The objective was to test that property (by introducing obviously unimportant predictors), as well as to explore the rate of influence of several variables whose relevance was not so obvious (e. g. the rate of variation of the reservoir level). The complete list of predictors is included in table 1.

All the calculations were performed on a training set covering the period 1980-1997, when weekly records were available. The relative influence and the partial dependence were computed with this data set. The model accuracy was assessed for a validation set covering the period 1998-2008.

The goodness of fit was computed in terms of the mean absolute error (MAE):

$$MAE = \frac{\sum_{i=1}^{N} |y_i - F(x_i)|}{N} \tag{8}$$

where N is the size of the training (or validation) set, y_i are the observed outputs and $F(x_i)$ the predicted values. As MAE is measured in the same units as the variable to predict, it is an intuitive measure of accuracy. However, it is not appropriate to compare models for different targets, as it does not account for the standard deviation of the output.

To overcome this limitation, the average relative variance (ARV) [31] was also computed:

Table 1: Predictor variables considered for the initial BRT model (M1).

Code	Group	Type	Period (days)
Level	Hydrostatic load	Original	-
Lev007	Hydrostatic load	Moving average	7
Lev014			14
Lev030			30
Lev060			60
Lev090			90
Lev180			180
Tair	Air temperature	Moving average	1
Tair007			7
Tair014			14
Tair030			30
Tair060			60
Tair090			90
Tair180			180
Rain	Rainfall	Accumulated	1
Rain030			30
Rain060			60
Rain090			90
Rain180			180
NDay	Time	Original	-
Year			-
Month	Season	Original	-
n010	Hydrostatic load	Rate of variation	10
n020			20
n030			30

$$ARV = \frac{\sum_{i=1}^{N}(y_i - F(x_i))^2}{\sum_{i=1}^{N}(y_i - \bar{y})^2} = \frac{MSE}{\sigma^2} \tag{9}$$

where $\bar{y}$ is the output mean. Given that ARV denotes the ratio between the mean squared error (MSE) and the variance (σ^2), it accounts both for the magnitude and the deviation of the target variable.

2.5. Overall procedure

For each target, the complete process comprised the following steps:

1. Fit a BRT model on the training data with the variables in table 1 (M1).
2. Compute the RI and generate the word cloud.
3. Select the most relevant predictors with the 1-SE rule (see section 2.2.1) and fit a new BRT model (M2).
4. Build a simple BRT model (M3) with the most influential variable of each group (temperature, level and time for displacements, and rainfall, level and time for leakage).
5. Generate the univariate and bivariate partial dependence plots for the simplest model.
6. Compute the goodness of fit for each model in both the training and the validation sets.

3. Results and discussion

3.1. Model accuracy

Although the work focused on model interpretation and its implications on dam safety, the goodness of fit was also checked in order to a) observe the effect of variable selection, and b) check the prediction accuracy of the model used for interpretation (M3).

Table 2 contains the error indices for each target, while more detailed results are included in the Appendix. For those models with variable selection, the predictors are also listed. The results show that BRT efficiently discarded irrelevant inputs, since the fitting accuracy was similar for each version in most cases (i.e., the presence of uninformative predictors did not damage the fitting accuracy).

The residuals were higher for the validation period, what reveals some degree of overfitting. A probable reason is that time was considered as any other predictor, and thus extrapolation over time was required to calculate the response in a more recent period. It is well known that non-parametric models lose much of their accuracy when predictions are made outside the range of variation of the input variables [32]. The increase in prediction error is greater for those targets for which time influence is more important, as is the case of the leakage in the left margin (up to four times larger). In these cases (AFMI90PR and AFTOTMI), the usefulness of the ARV is clearly observed: while the MAE is similar for the training and validation periods, the ARV is notably greater in the latter case, because the variance is lower in the most recent period (leakage flow decreased significantly over time).

Table 2: Accuracy of each model and target for the training and validation sets. The results and inputs considered by the most accurate version are highlighted in bold.

	Train		Validation		
Target	MAE	ARV	MAE	ARV	Inputs
P1DR1	0,64	0,03	0,91	0,08	All
	0,68	0,03	0,81	0,06	Tair090,Level,NDay,Lev007,Lev014
	0,69	**0,03**	**0,78**	**0,06**	**NDay,Tair090,Level**
P1DR4	**0,46**	**0,03**	**0,65**	**0,08**	**All**
	0,50	0,03	0,66	0,08	Level,Tair090,NDay,Lev007,Lev014,Lev030
	0,51	0,03	0,67	0,08	NDay,Tair090,Level
P2IR1	0,66	0,03	1,03	0,09	All
	0,85	0,05	1,09	0,09	Tair090,Level,Lev007,Lev014
	0,71	**0,04**	**0,98**	**0,08**	**NDay,Tair090,Level**
P2IR4	**0,48**	**0,05**	**0,90**	**0,14**	**All**
	0,61	0,06	0,93	0,14	Level,Tair090,Lev007,Lev014,Lev030
	0,53	0,06	0,94	0,16	NDay,Tair090,Level
P5DR1	**0,66**	**0,05**	**0,82**	**0,08**	**All**
	0,64	0,05	0,87	0,10	Tair060,Level,Tair030
	0,83	0,08	0,93	0,11	NDay,Tair060,Level
P5DR3	**0,25**	**0,03**	**0,47**	**0,21**	**All**
	0,33	0,05	0,55	0,22	Tair060,Level,Tair030
	0,31	0,04	0,52	0,24	NDay,Tair060,Level
P6IR1	0,60	0,04	0,80	0,09	All
	0,65	**0,05**	**0,78**	**0,08**	**Tair060,Tair030,Level,NDay**
	0,83	0,08	0,85	0,1	NDay,Tair060,Level
P6IR3	**0,23**	**0,02**	**0,40**	**0,08**	**All**
	0,37	0,05	0,67	0,17	Tair060,Level,Tair030
	0,29	0,03	0,43	0,09	NDay,Tair060,Level
AFMD50PR	**1,28**	**0,16**	**0,93**	**0,19**	**All**
	1,45	0,17	1,36	0,28	Level,Lev014,Lev007
	1,16	0,14	1,23	0,48	NDay,Rain090,Level
AFMI90PR	0,08	0,09	0,15	0,51	All
	0,08	0,10	0,12	0,45	Lev007,NDay,Level,Lev014,Lev030
	0,08	**0,10**	**0,12**	**0,46**	**NDay,Rain030,Lev007**
AFTOTMD	**1,64**	**0,15**	**1,67**	**0,37**	**All**
	1,87	0,19	1,73	0,45	Level,Lev007,Lev014
	1,69	0,18	1,97	0,52	NDay,Rain180,Level
AFTOTMI	**0,41**	**0,11**	**0,44**	**0,40**	**All**
	0,44	0,12	0,44	0,42	NDay,Lev060,Lev014,Lev007,Lev030,Lev180,Lev090,Level
	0,54	0,18	0,46	0,60	NDay,Rain180,Lev060

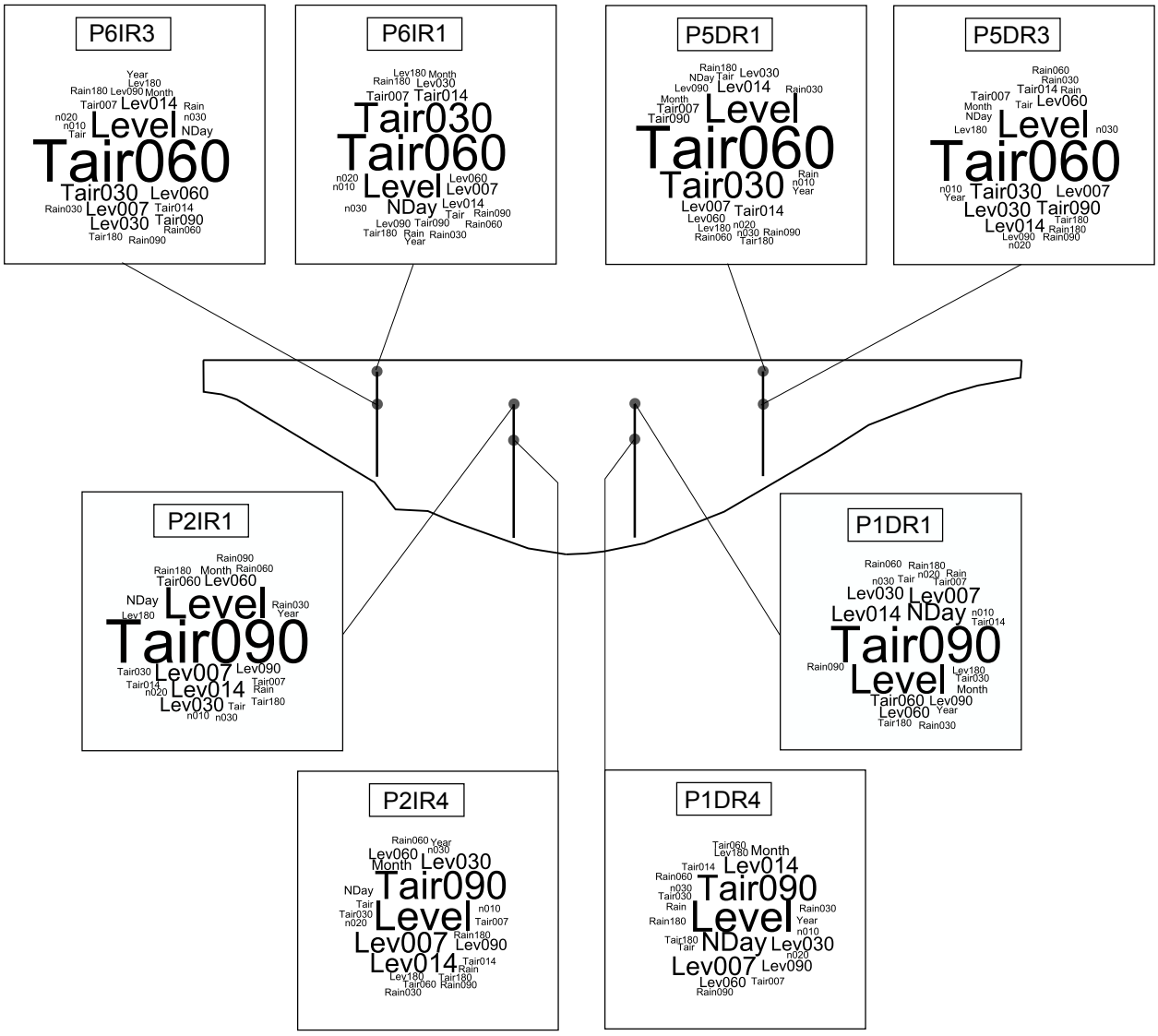

Figure 3: Word clouds for the radial displacements analysed.

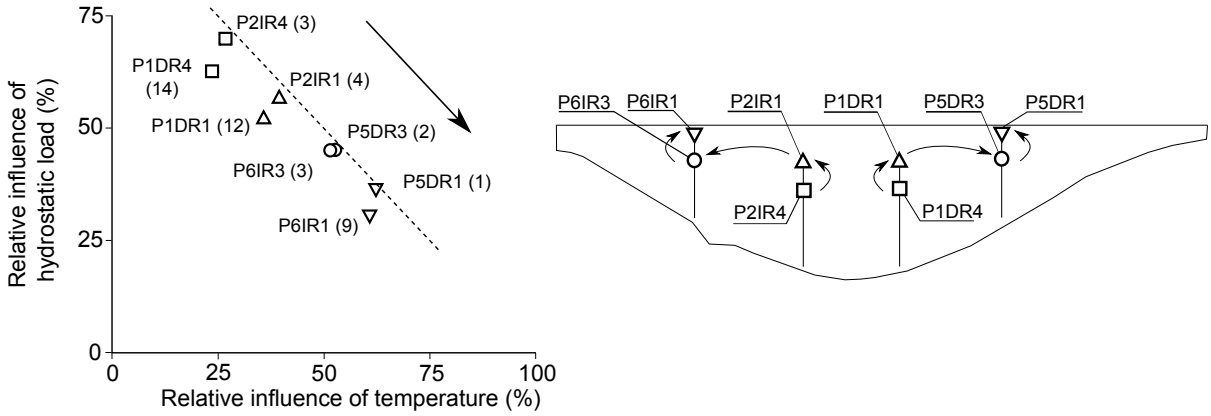

Figure 4: Relative influence of inputs in radial displacements, aggregated by type. The values in parenthesis correspond to the relative influence of time. Since the result is normalised, they sum 100 for each location. Hence, the distance to the $x + y = 100$ line (dashed) is proportional to the importance of the time effect. It should be noted that the devices in symmetrical locations with respect to the dam axis are grouped (i.e. P6IR3 and P5DR3). The arrows highlight the path of increasing influence of temperature in both plots, also symmetrical.

3.2. Variable importance

3.2.1. Radial displacements

Figure 3 depicts the RI of the predictors for each radial displacement considered. While Tair90 was the most relevant thermal input for the central sections (P1DR and P2IR), Tair060 took its place for those in the vicinity of the abutments (P5DR and P6IR). The higher thermal inertia of the central blocks might be due to their greater average thickness.

As for the hydrostatic load, the reservoir level at the date of the record was always more influential than all the moving averages, what reveals an immediate response of the dam to this load.

The RI of the rate of reservoir level variation was similar to that of rainfall, hence negligible.

From an overall viewpoint, a high degree of symmetry was observed, with the remarkable exception of the greater influence of NDay for P1DR1 and P1DR4 (Figure 3). This issue was further investigated by aggregating the relative influence of inputs by type: hydrostatic load, air temperature and time (Table 1). Figure 4 shows the result for each location considered. The symmetry is neatly observed, as well as the increasing RI of the temperature with respect to that of the hydrostatic load, from the foundation towards the crown, and from the centre to the abutments.

3.2.2. Leakage

The RI of the inputs for the leakage flows revealed a clear different behaviour between the right (AFMD50PR and AFTOTMD) and the left margins (AFMI90PR and AFTOTMI). While the former responded mainly to the hydrostatic load, with little inertia, the latter showed a remarkable dependence on time, as well as a greater relevance of several rolling means of reservoir level. Figure 5 shows the word clouds for the leakage flows.

The low inertia with respect to the hydrostatic load suggests that most of the leakage flow comes from the reservoir, while the effect of rainfall is negligible.

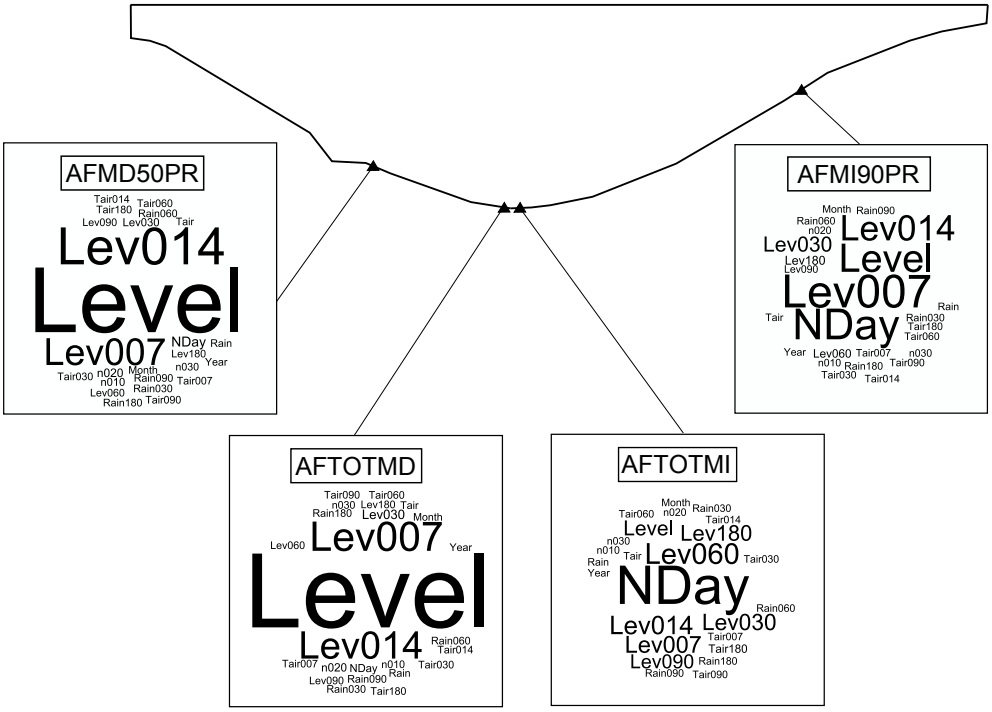

Figure 5: Word clouds for the leakage measurement locations analysed

Finally, it should be noted that temperature was irrelevant.

3.3. Partial dependence

3.3.1. Radial displacements

Figure 6 displays the univariate partial dependence plots for the radial displacements. The association between the air temperature and the reservoir level is intuitive: high temperature generates displacements towards upstream (positive direction in the vertical axis), due to concrete expansion, coerced by the abutments. The effect of the hydrostatic load is the opposite: high levels imply greater load and displacement towards downstream, and vice-versa.

These plots also show that the air temperature had greater influence than reservoir level for P5D and P6I (steeper curves with greater range of variation). The inverse effect was observed for P1DR4 and P2IR4, whereas both had similar relevance for P1DR1 and P2IR1. These results are coherent with the computed RI (Figures 3 and 4).

A deeper analysis was performed for P1DR1, which featured the greatest influence of time. The correspondent plot in Figure 6 shows a small step at the beginning of the period, followed by a sensibly constant behaviour until 1990, and a second larger step that stabilised afterwards. The bivariate plots for P1DR1 in Figure 7 show the step around 1991-1992 for the whole ranges of level and temperature.

This qualitative behaviour was observed for all displacements with lower magnitude, except for P5DR3. The effect was more clearly registered for P1DR1 and P1DR4, which in turn showed a greater influence of time (see Figure 3 and the Appendix).

Figure 8 shows the contribution of each action as obtained with the HST model. Unlike the previous case, the influence of time is linear, equivalent to a constant-rate displacement towards downstream.

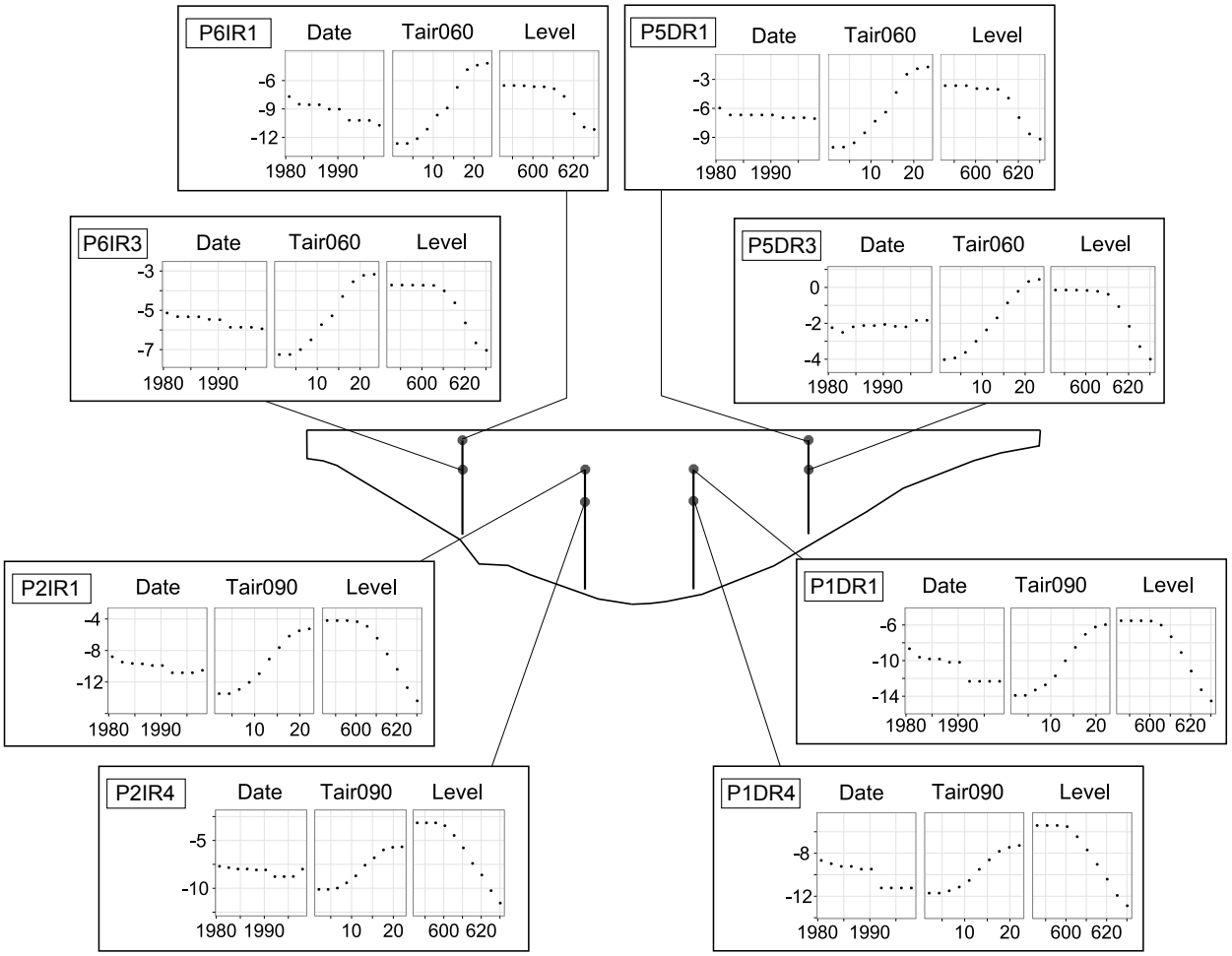

Figure 6: Partial dependence plot for the radial displacements analysed. Movement towards downstream correspond to lower values in the vertical axis, and vice-versa.

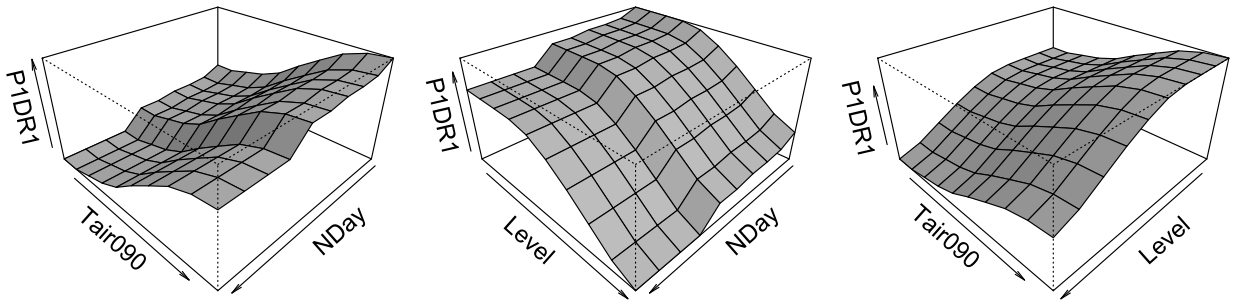

Figure 7: Interaction plots for P1DR1. It should be noted that the step along the temporal axis is observed for all the range of temperature and level.

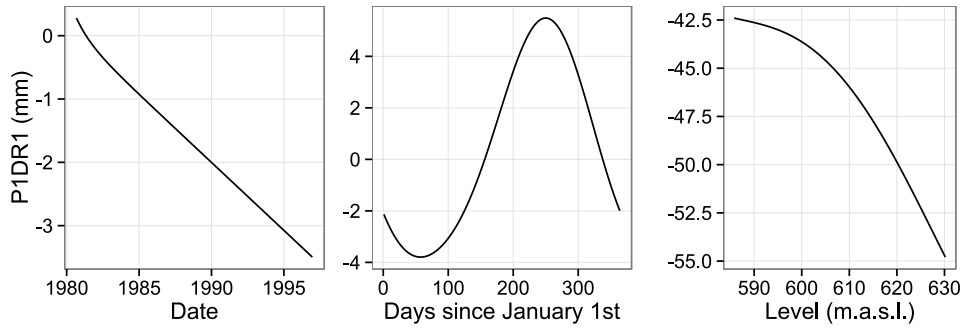

Figure 8: P1DR1. Contribution of the air temperature, the hydrostatic load and the time as drawn from the HST model.

This different model interpretation deserved a further verification. Not only because the results were substantially different, but also because the time effect is essential for the early detection of dam deterioration. In view of the temporal variation as captured by the BRT model, it could be concluded that some anomaly happened around 1991, which stabilised in the later years. On the contrary, the HST model interpreted a constant drift towards downstream of roughly 0.2 mm/year which might be serious in terms of the dam safety.

As mentioned above, the shape of the output-time dependency must be defined a priori for HST (in this case, a combination of exponential and negative linear functions was chosen), while in principle it can take any form for BRT. Therefore, it could be concluded that the actual behaviour of the dam was that showed by the BRT model, and that the result of the HST was due to the previously imposed restriction.

However, the average reservoir level in the period 1991-1997 was significantly higher than before 1991 (Figure 2), and might be the cause of the step registered in Figure 6: it represents a greater displacement towards downstream in the most recent period, which is consistent with the higher average hydrostatic load.

The verification was performed by fitting a new BRT model to the artificial data generated ($\hat{Y}_{mod}$) without time variation (eq. 7). It should be recalled that the artificial time series data maintains the original reservoir level variation, and thus the higher load in the 1991-1997 period. Figure 9 contains the partial dependence plot for this BRT model, which clearly shows that the independence of the artificial data with respect to time was correctly captured. This result confirms that the step in the time dependence captured by BRT is not a consequence of the higher hydrostatic load in 1991-1997.

As regards the HST model, it can be concluded that the linear trend is the best least squares fit that can be obtained to the observed behaviour (constant-step-constant) with a linear function. This might lead to a wrong interpretation of dam performance, not supported by the observed data.

It should be mentioned that more sophisticated versions of the HST model can be employed, and in particular a step can be considered, as Carrère and Noret-Duchêne showed in their analysis of the Schlegeiss Dam [33]. However, they only decided to try a step after observing a previous linear fit, where the sudden change in dam behaviour could be noted. In general, it can be difficult to identify a change in dam behaviour by simple data exploration, as is the case of La Baells Dam (see Figure A3). In this sense, the non-parametric nature of

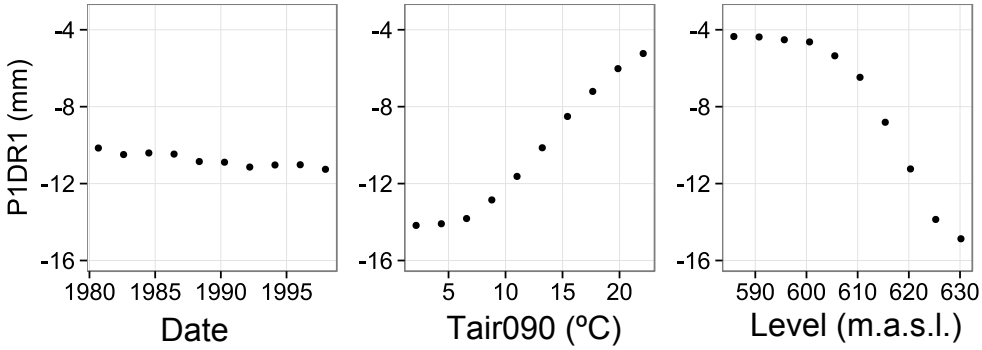

Figure 9: Partial dependence plot for the artificial time-independent data. P1DR1. It should be noted that time influence is negligible.

BRT models helps to identify performance changes of any type.

3.3.2. Leakage flows

Although the word clouds showed that neither rainfall nor temperature were influential on the leakage flow, partial dependence plots were generated as a further verification for the simplest model (M3 model; section 2.5).

Figure 10 contains the results, which confirm the conclusion of the word clouds: the time effect is irrelevant in the right abutment, except by certain erratic behaviour in the first two years and in the last three. On the contrary, a sharp decrease in leakage flow is revealed around 1983 for both locations in the left abutment. A lower decrease is observed in later years.

These results might be due to the colmatation of the cracking network in the left abutment, which would have led to lower permeability and leakage flow.

The shape of the effect of the hydrostatic load is sensibly exponential, with low influence for reservoir level below 610 m.a.s.l.

4. Summary and Conclusions

BRT models with different degree of variable selection were fitted to 8 radial displacements and 4 leakage flows at La Baells Dam. The relative influence of each input was computed and depicted via word clouds, which offered an efficient visualisation of the overall response of the dam. These graphs, together with the univariate and bivariate partial dependence plots, allowed interpretation of the BRT models: useful information regarding dam behaviour was obtained, such as the thermal inertia, the variation over time, and the performance of each area of the dam body.

The results showed a symmetrical behaviour of the dam in terms of displacements, as well as some interesting patterns, which will be the subject of future research:

- the thermal inertia was higher near the abutments.
- the RI of the temperature with respect to that of the hydrostatic load increased from the foundation towards the crown, and from the centre to the abutments.

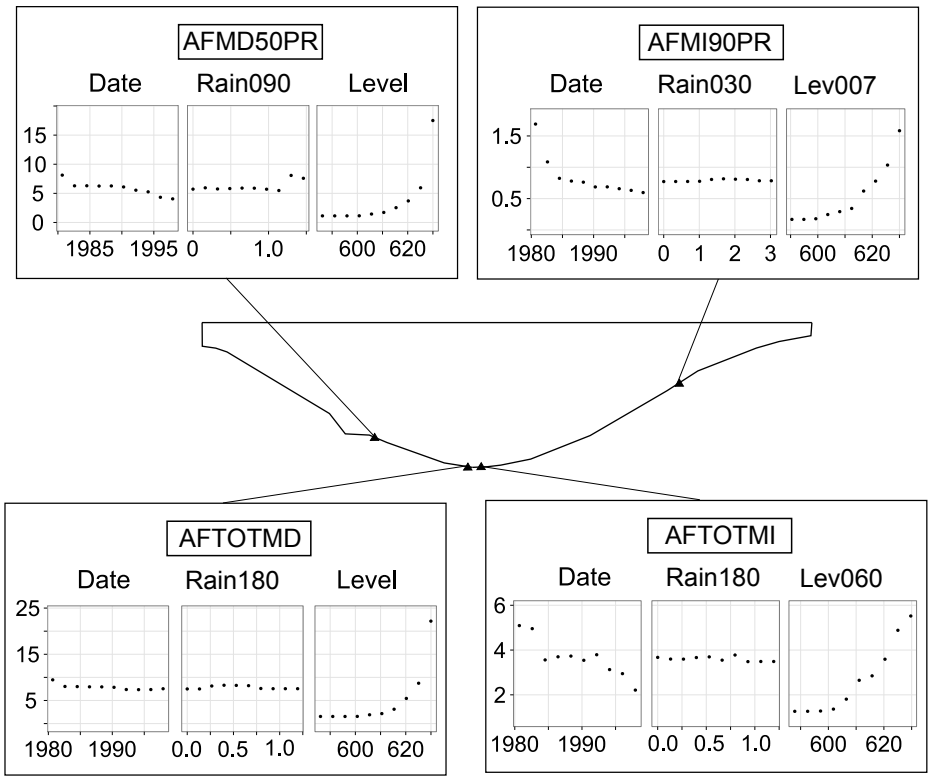

Figure 10: Partial dependence plot for leakage flows.

As regards the leakage flows, the different behaviour over time of each abutment was identified.

The amount of predictors considered in the BRT model did not significantly affect the prediction accuracy: the results confirm that the algorithm efficiently discard the less relevant inputs.

The application of BRT models to make predictions for a more recent period than that used for training involves extrapolation over time (provided that some time dependent predictor is considered). Hence, results should be analysed carefully, in particular if the time effect seems relevant. This applies to any data-based model considering time as input, including HST.

A sudden change in radial displacements was identified by the BRT model, especially for P1DR1. By contrast, the HST model suggested a constant-rate drift in the downstream direction. It was verified that the step towards downstream captured by the BRT model was not due to the higher average hydrostatic load actually registered for the 1991-1997 period. This suggests that partial dependence plots based on BRT models are more effective to identify performance changes, as they are not coerced by the shape of the regression functions that need to be defined a priori for HST.

The flexibility and robustness of BRT models make them suitable to model any output variable, as well as to identify changes in dam behaviour. Nevertheless, data-based models should never be the only source of information to make decisions on dam safety. Their results need to be checked against those provided by other means, such as deterministic models. Also, all available information about the dam behaviour should be taken into account, especially that obtained by visual inspection.

5. Acknowledgements

The authors thank Carlos Barbero, dam safety manager at the Catalan Water Agency, for providing the monitoring data.
The research was supported by the Spanish Ministry of Economy and Competitiveness (*Ministerio de Economía y Competitividad*, MINECO) through the projects iComplex (IPT-2012-0813-390000) and AIDA (BIA2013-49018-C2-1-R and BIA2013- 49018-C2-2-R).

References

[1] G. Lombardi, Advanced data interpretation for diagnosis of concrete dams, Tech. rep., CISM (2004).

[2] G. Willm, N. Beaujoint, Les méthodes de surveillance des barrages au service de la production hydraulique d'Electricité de France-Problèmes ancients et solutions nouvelles, in: 9th ICOLD Congres, 1967, pp. 529–550, q34-R30. [in French].

[3] M. Tatin, M. Briffaut, F. Dufour, A. Simon, J.-P. Fabre, Thermal displacements of concrete dams: Accounting for water temperature in statistical models, Engineering Structures 91 (2015) 26–39.

[4] F. Amberg, Interpretative models for concrete dam displacements, in: 23th ICOLD Congress, 2009, q91-R43.

[5] D. Santillán, E. Salete, D. Vicente, M. Toledo, Treatment of solar radiation by spatial and temporal discretization for modeling the thermal response of arch dams, Journal of Engineering Mechanics 140 (11).

[6] I. Penot, B. Daumas, J. Fabre, Monitoring behaviour, Water Power and Dam Construction.

[7] F. Salazar, M. Toledo, Discussion on "Thermal displacements of concrete dams: Accounting for water temperature in statistical models", Engineering Structures (2015) –doi:http://dx.doi.org/10.1016/j.engstruct.2015.08.001.

[8] F. Salazar, R. Morán, M. Á. Toledo, E. Oñate, Data-based models for the prediction of dam behaviour: A review and some methodological considerations, Archives of Computational Methods in Engineering (2015) 1–21.

[9] V. Ranković, N. Grujović, D. Divac, N. Milivojević, Development of support vector regression identification model for prediction of dam structural behaviour, Structural Safety 48 (2014) 33–39.

[10] J. Mata, Interpretation of concrete dam behaviour with artificial neural network and multiple linear regression models, Engineering Structures 3 (3) (2011) 903 – 910. doi:10.1016/j.engstruct.2010.12.011.

[11] S. Demirkaya, Deformation analysis of an arch dam using ANFIS, in: Proceedings of the second international workshop on application of artificial intelligence and innovations in engineering geodesy. Braunschweig, Germany, 2010, p. 21–31.

[12] L. Auret, C. Aldrich, Empirical comparison of tree ensemble variable importance measures, Chemometrics and Intelligent Laboratory Systems 105 (2) (2011) 157–170.

[13] F. Salazar, M. Toledo, E. Oñate, R. Morán, An empirical comparison of machine learning techniques for dam behaviour modelling, Structural Safety 56 (2015) 9–17.

[14] F. Li, Z. Wang, G. Liu, Towards an error correction model for dam monitoring data analysis based on cointegration theory, Structural Safety 43 (2013) 12–20.

[15] D. Santillán, J. Fraile-Ardanuy, M. Toledo, Seepage prediction in arch dams by means of artificial neural networks, Water Technology and Science V (3).

[16] L. Cheng, D. Zheng, Two online dam safety monitoring models based on the process of extracting environmental effect, Advances in Engineering Software 57 (2013) 48–56.

[17] L. Breiman, et al., Statistical modeling: The two cultures (with comments and a rejoinder by the author), Statistical Science 16 (3) (2001) 199–231.

[18] T. Hastie, R. Tibshirani, J. Firedman, The Elements of Statistical Learning - Data Mining, Inference, and Prediction, 2nd Edition, Springer, New York, 2009.

[19] L. Breiman, J. H. Friedman, R. A. Olshen, C. J. Stone, Classification and regression trees, Wadsworth & Brooks, Monterrey, CA, 1984.

[20] J. Friedman, Greedy function approximation: a gradient boosting machine, Annals of Statistics (2001) 1189 – 1232.

[21] G. Ridgeway, Generalized Boosted Models: A guide to the gbm package, r package vignette (2007).
URL `http://CRAN.R-project.org/package=gbm`

[22] J. Leathwick, J. Elith, M. Francis, T. Hastie, P. Taylor, Variation in demersal fish species richness in the oceans surrounding new zealand: an analysis using boosted regression trees, Marine Ecology Progress Series 321 (2006) 267–281.

[23] J. Elith, J. R. Leathwick, T. Hastie, A working guide to boosted regression trees, Journal of Animal Ecology 77 (4) (2008) 802–813.

[24] R. E. Schapire, The boosting approach to machine learning: An overview, in: Nonlinear estimation and classification, Springer, 2003, pp. 149–171.

[25] A. Michelis, Traditional versus non-traditional boosting algorithms, Master's thesis, University of Manchester (2012).

[26] G. R. with contributions from others, gbm: Generalized Boosted Regression Models, r package version 2.1 (2013).

[27] R Core Team, R: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing, Vienna, Austria (2013).
URL `http://www.R-project.org/`

[28] O. Kaser, D. Lemire, Tag-cloud drawing: Algorithms for cloud visualization, arXiv preprint cs/0703109.

[29] I. Fellows, wordcloud: Word Clouds, r package version 2.5 (2014).
URL `http://CRAN.R-project.org/package=wordcloud`

[30] J. H. Friedman, J. J. Meulman, Multiple additive regression trees with application in epidemiology, Statistics in medicine 22 (9) (2003) 1365–1381.

[31] A. S. Weigend, B. A. Huberman, D. E. Rumelhart, Predicting sunspots and exchange rates with connectionist networks, in: S. Eubank, M. Casdagli (Eds.), Proc. of the 1990 NATO Workshop on Nonlinear Modeling and Forecasting (Santa Fe, NM), Vol. 12, Addison-Wesley, Redwood, CA, 1992, pp. 395–432.

[32] G. Hooker, Diagnosing extrapolation: Tree-based density estimation, in: Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining, ACM, 2004, pp. 569–574.

[33] A. Carrère, C. Noret-Duchêne, Interpretation of an arch dam behaviour using enhanced statistical models, in: Proceedings of the Sixth ICOLD Benchmark Workshop on Numerical Analysis of Dams, Salzburg, Austria, 2001.

Appendix A.

All the plots generated during the analysis are included herein: the time series of mean air temperature and daily rainfall, and a set of plots for each target variable:

- The 2D and 3D partial dependence plots for BRT Model 3 fitted to the original data
- The location of each device within the dam body
- The 2D partial dependence plot for BRT Model 3 fitted on the altered version of the target (independent of time)
- The word cloud for Model 1
- Observations versus BRT model predictions for the training and validation sets, together with the model residuals.

The partial dependence for the artificial data was included to highlight that the BRT models correctly captured the time independence when it was imposed in the time series of the target variable.

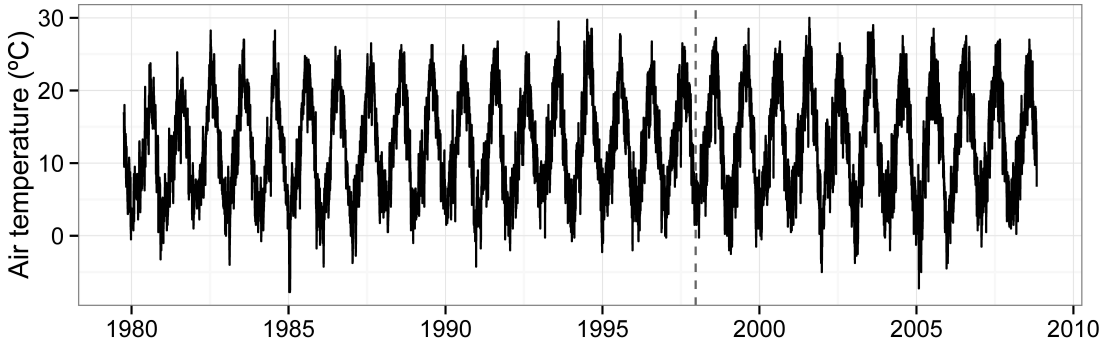

Figure A1: Time series of the mean air temperature at La Baells dam site.

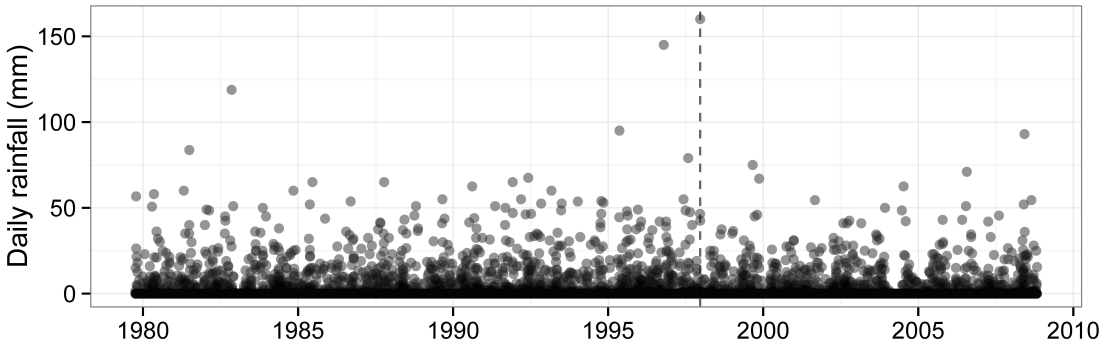

Figure A2: Time series of the daily rainfall at La Baells dam site.

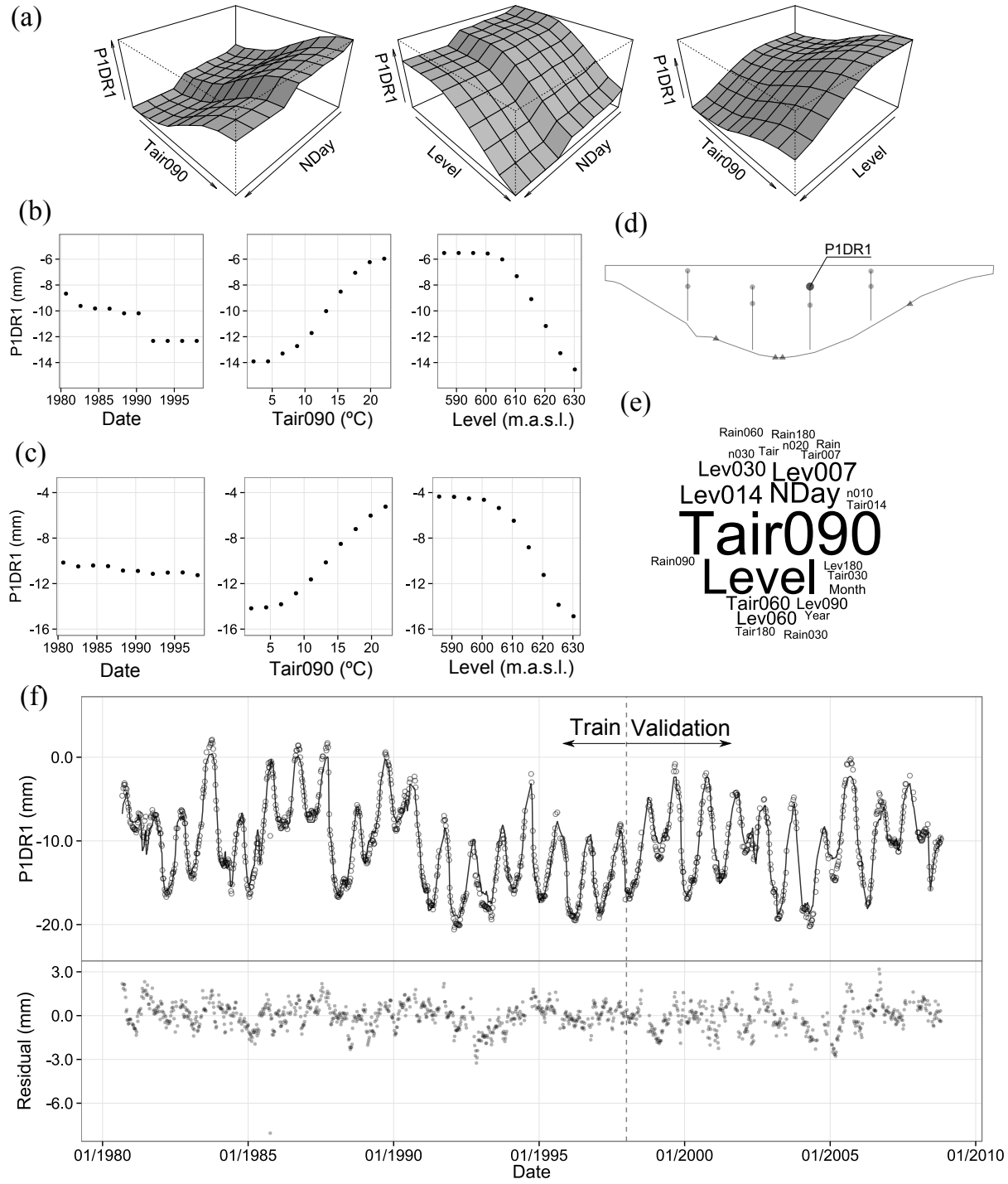

Figure A3: P1DR1. (a) 3D partial dependence plot; (b) 2D Partial dependence plot; (c) Idem for artificial data (time-independent); (d) Device location; (e) Word cloud of relative influence; (f) Model fit and residuals for the train and the validation sets.

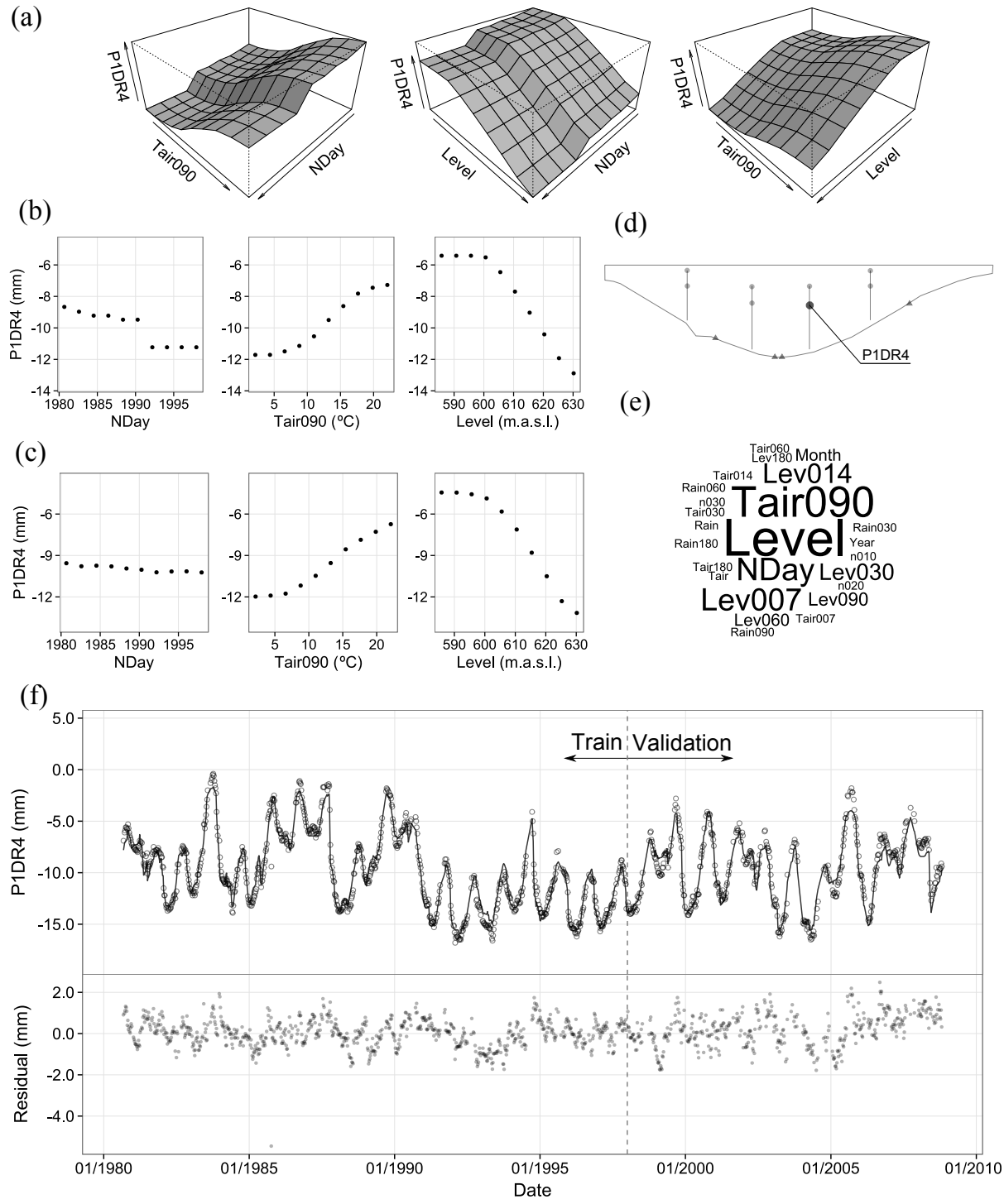

Figure A4: P1DR4. (a) 3D partial dependence plot; (b) 2D Partial dependence plot; (c) Idem for artificial data (time-independent); (d) Device location; (e) Word cloud of relative influence; (f) Model fit and residuals for the train and the validation sets.

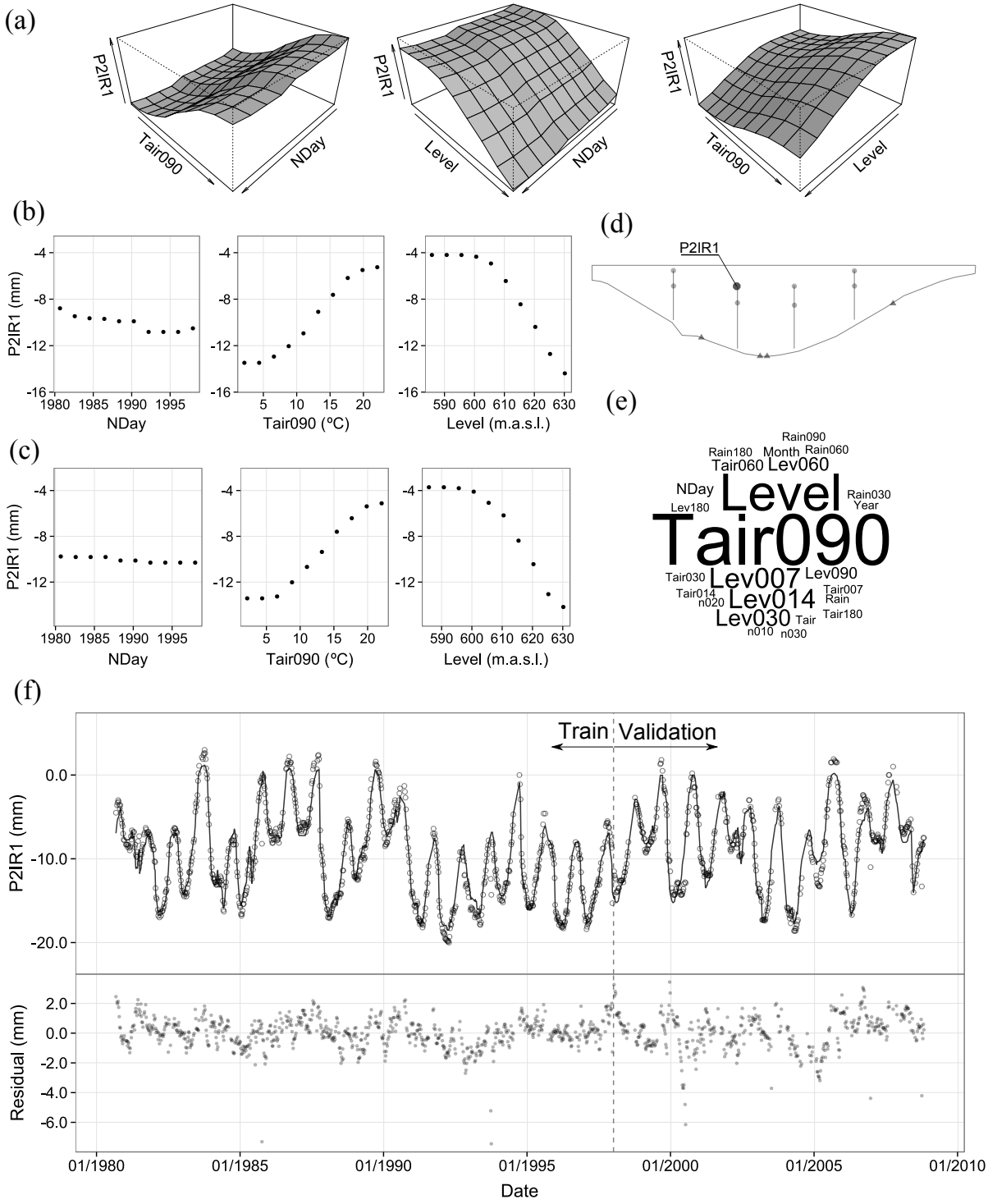

Figure A5: P2IR1. (a) 3D partial dependence plot; (b) 2D Partial dependence plot; (c) Idem for artificial data (time-independent); (d) Device location; (e) Word cloud of relative influence; (f) Model fit and residuals for the train and the validation sets.

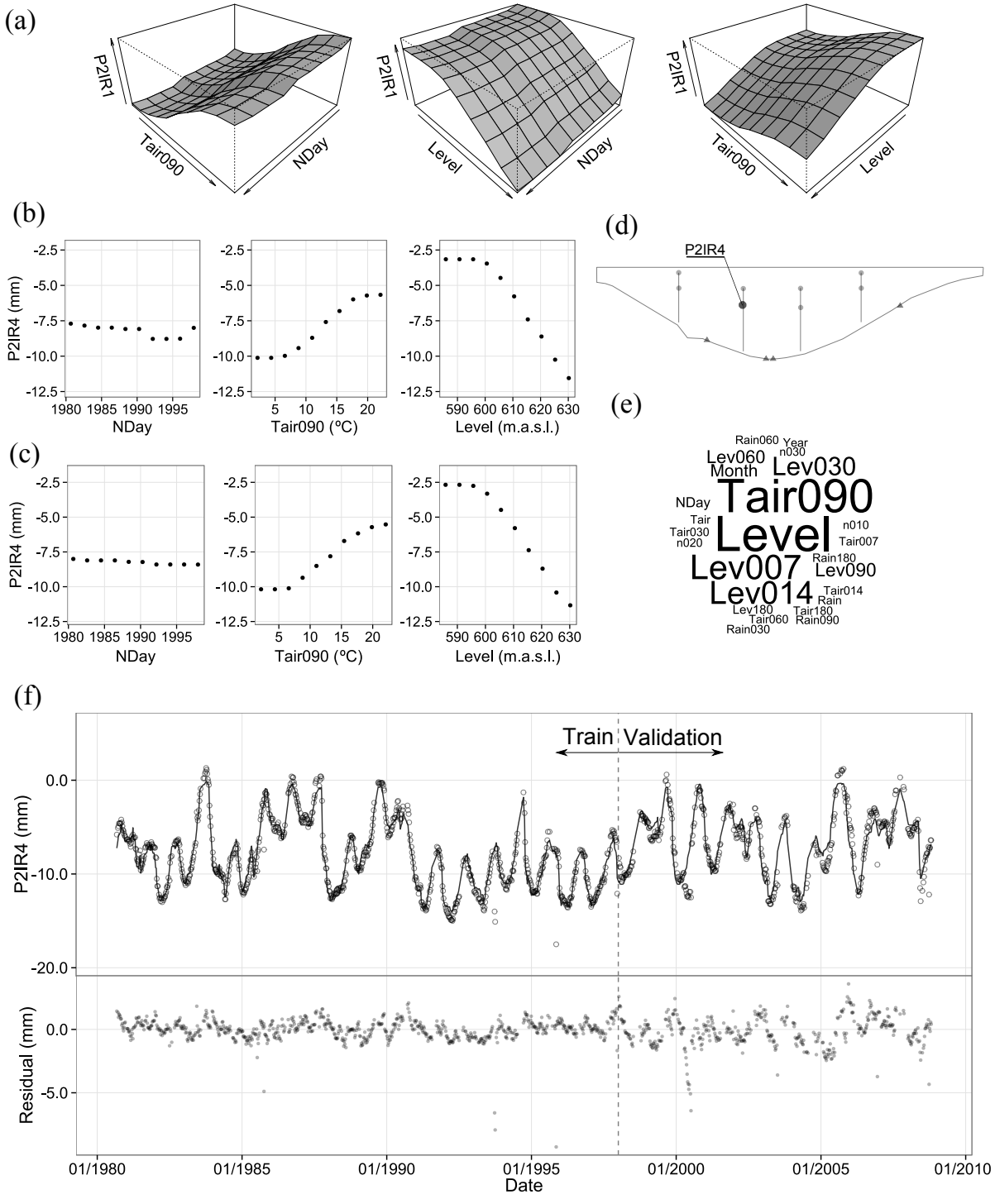

Figure A6: P2IR4. (a) 3D partial dependence plot; (b) 2D Partial dependence plot; (c) Idem for artificial data (time-independent); (d) Device location; (e) Word cloud of relative influence; (f) Model fit and residuals for the train and the validation sets.

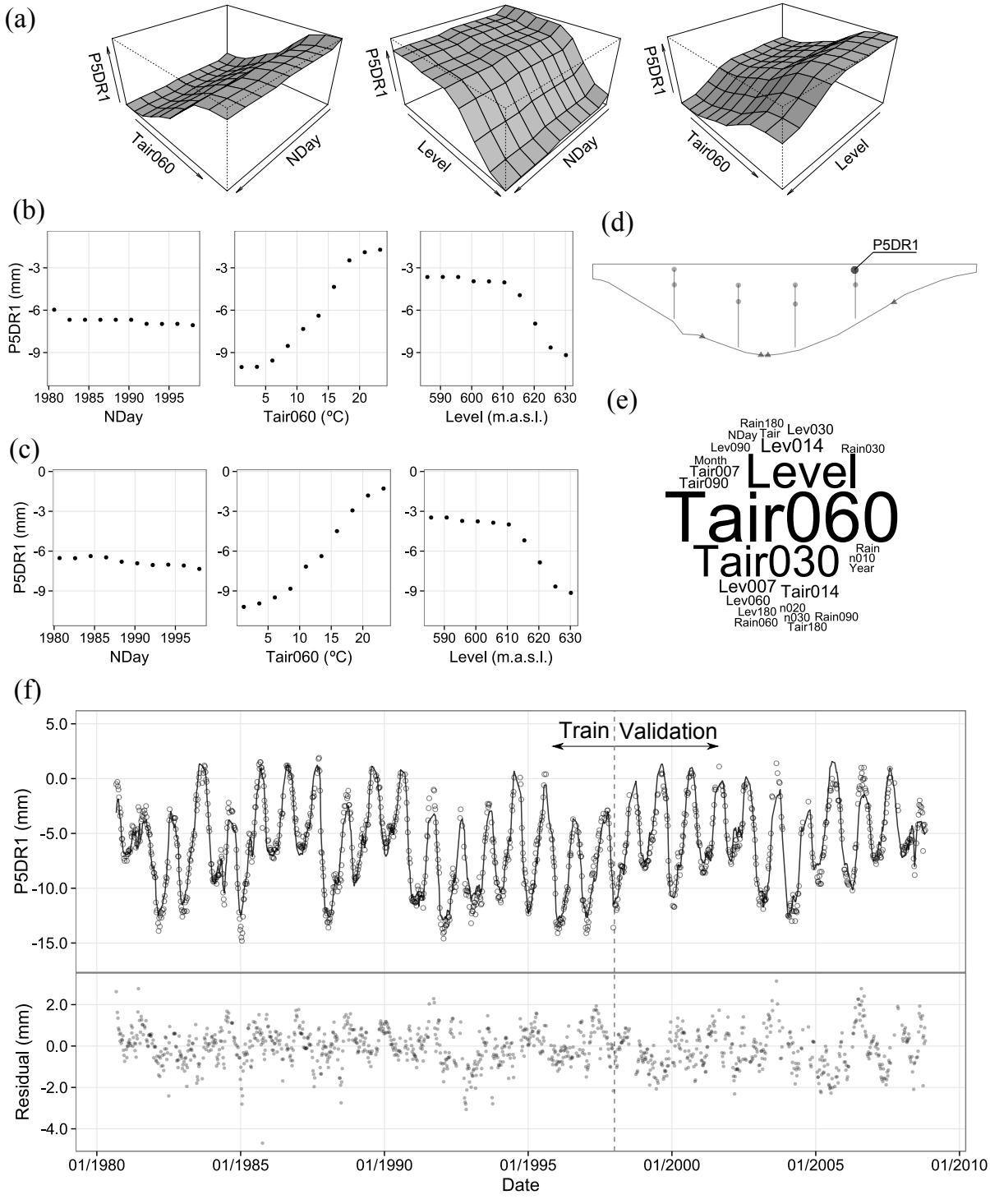

Figure A7: P5DR1. (a) 3D partial dependence plot; (b) 2D Partial dependence plot; (c) Idem for artificial data (time-independent); (d) Device location; (e) Word cloud of relative influence; (f) Model fit and residuals for the train and the validation sets.

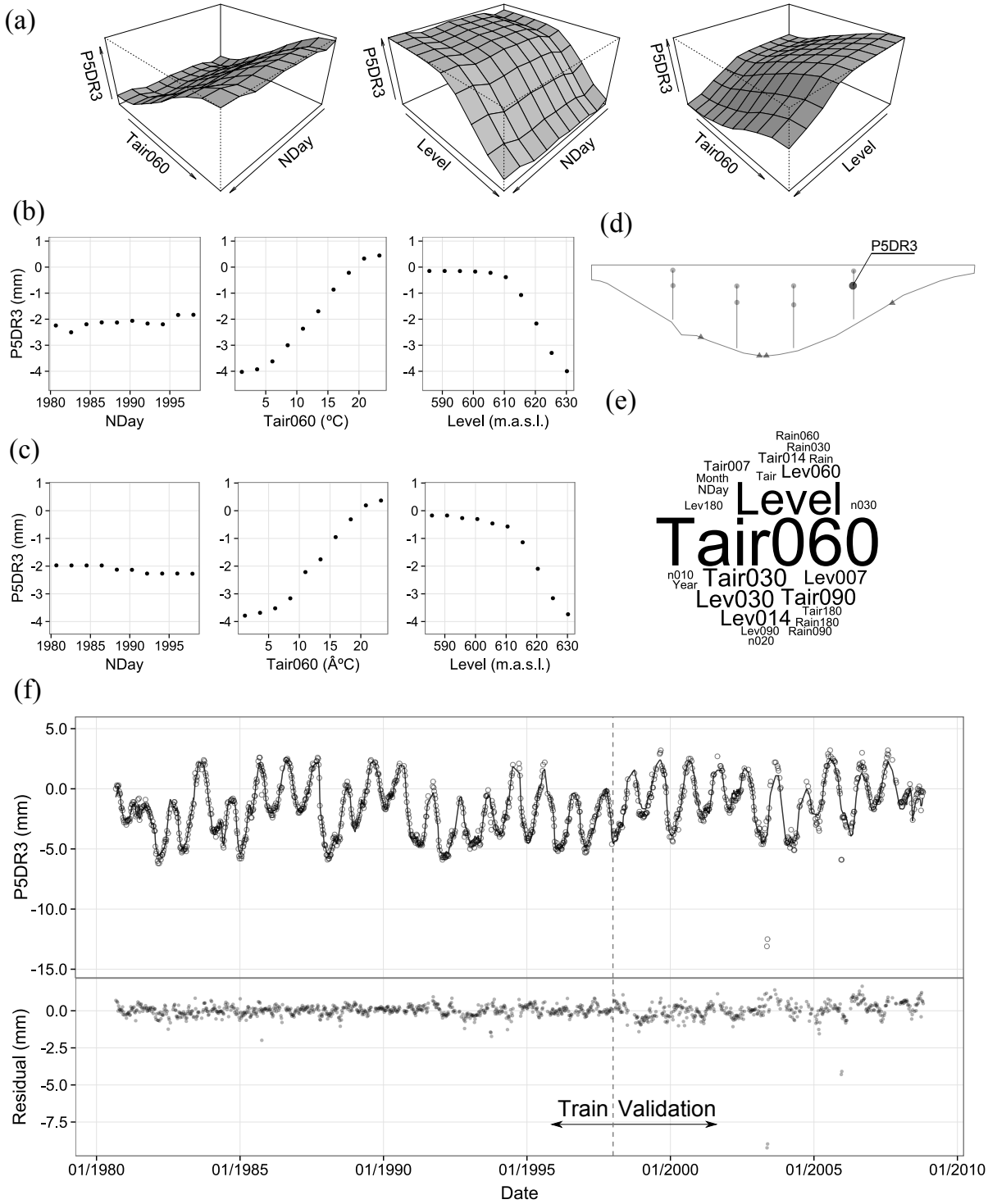

Figure A8: P5DR3. (a) 3D partial dependence plot; (b) 2D Partial dependence plot; (c) Idem for artificial data (time-independent); (d) Device location; (e) Word cloud of relative influence; (f) Model fit and residuals for the train and the validation sets.

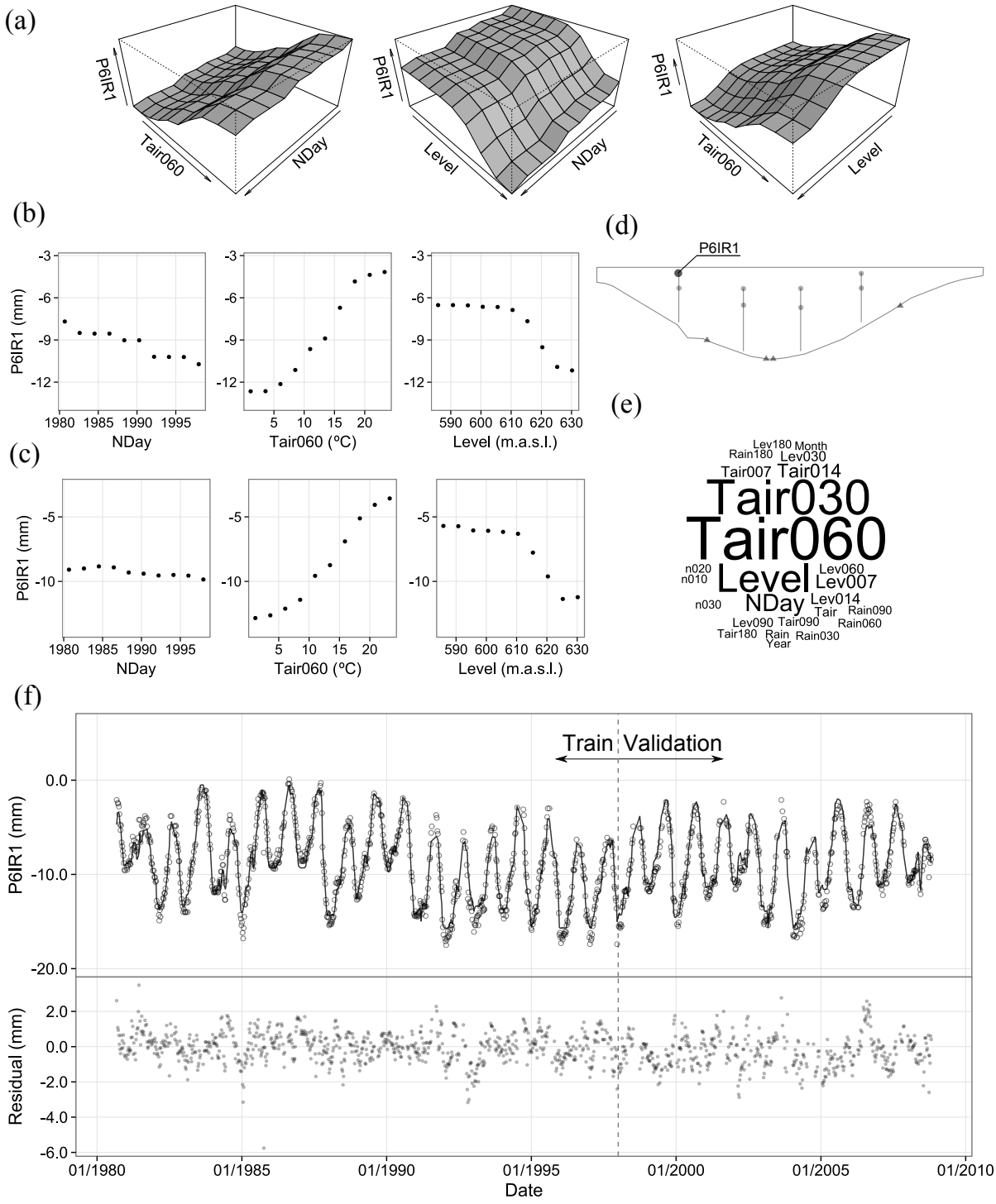

Figure A9: P6IR1. (a) 3D partial dependence plot; (b) 2D Partial dependence plot; (c) Idem for artificial data (time-independent); (d) Device location; (e) Word cloud of relative influence; (f) Model fit and residuals for the train and the validation sets.

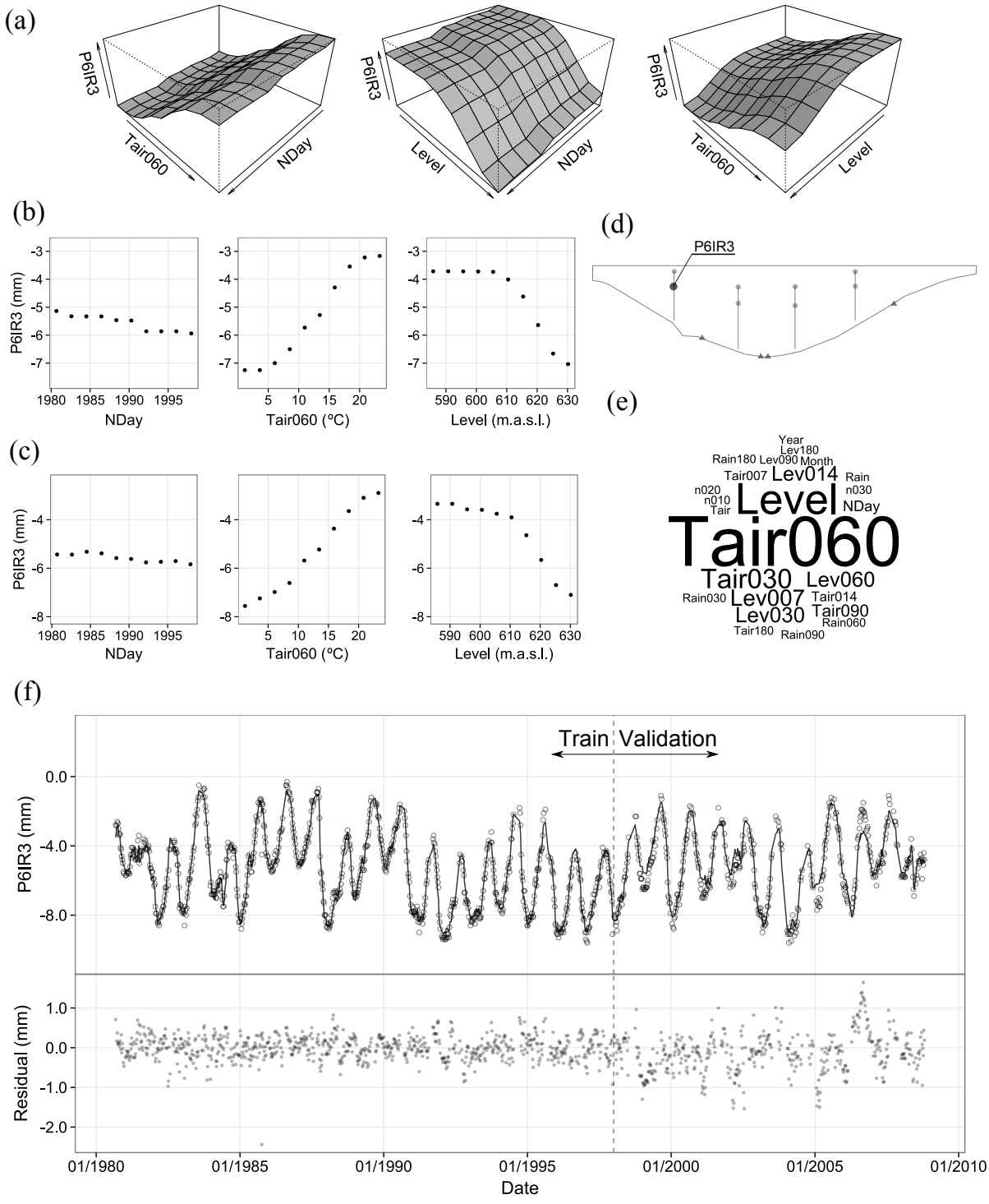

Figure A10: P6IR3. (a) 3D partial dependence plot; (b) 2D Partial dependence plot; (c) Idem for artificial data (time-independent); (d) Device location; (e) Word cloud of relative influence; (f) Model fit and residuals for the train and the validation sets.

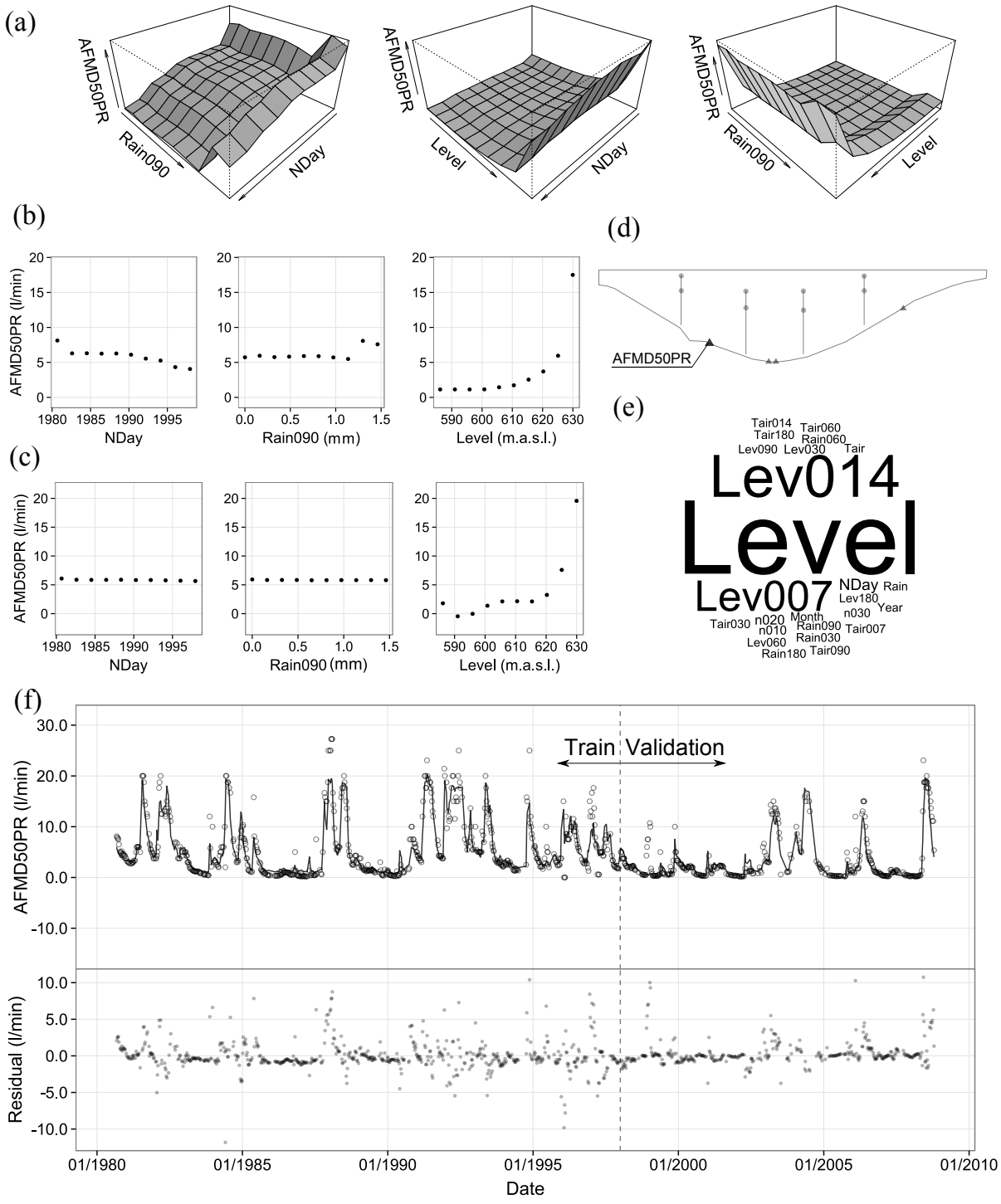

Figure A11: AFMD50PR. (a) 3D partial dependence plot; (b) 2D Partial dependence plot; (c) Idem for artificial data (time-independent); (d) Device location; (e) Word cloud of relative influence; (f) Model fit and residuals for the train and the validation sets.

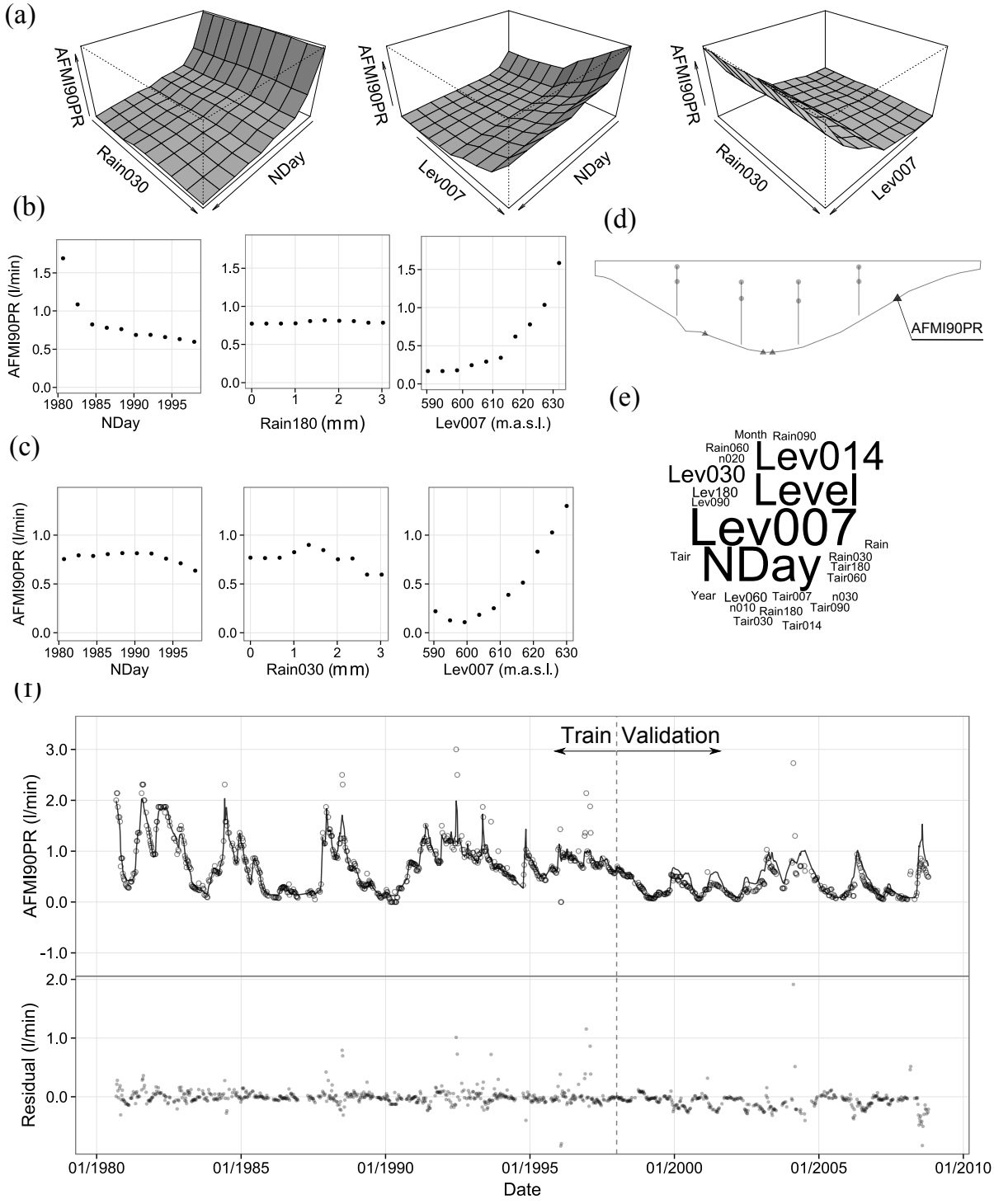

Figure A12: AFMI90PR. (a) 3D partial dependence plot; (b) 2D Partial dependence plot; (c) Idem for artificial data (time-independent); (d) Device location; (e) Word cloud of relative influence; (f) Model fit and residuals for the train and the validation sets.

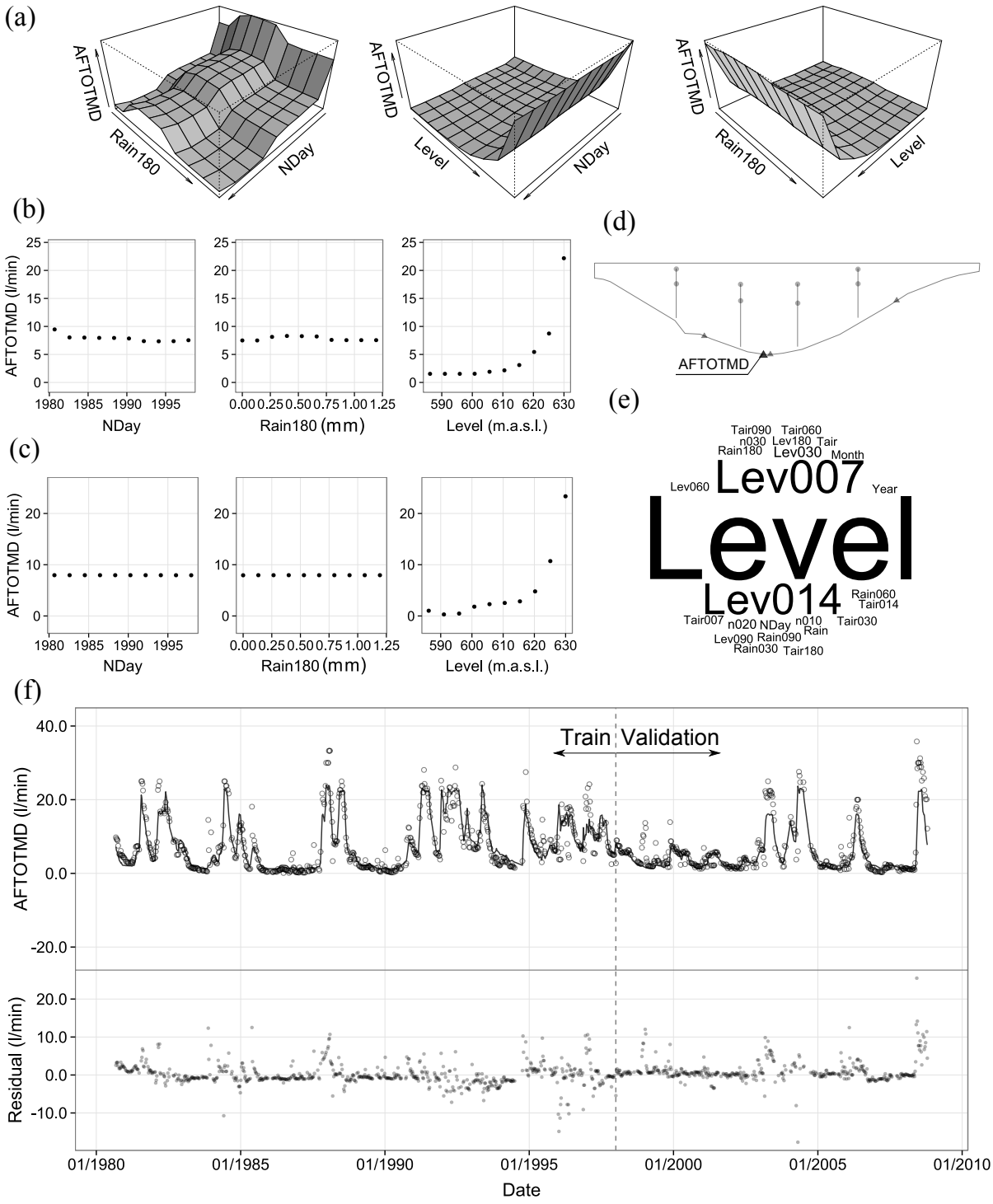

Figure A13: AFTOTMD. (a) 3D partial dependence plot; (b) 2D Partial dependence plot; (c) Idem for artificial data (time-independent); (d) Device location; (e) Word cloud of relative influence; (f) Model fit and residuals for the train and the validation sets.

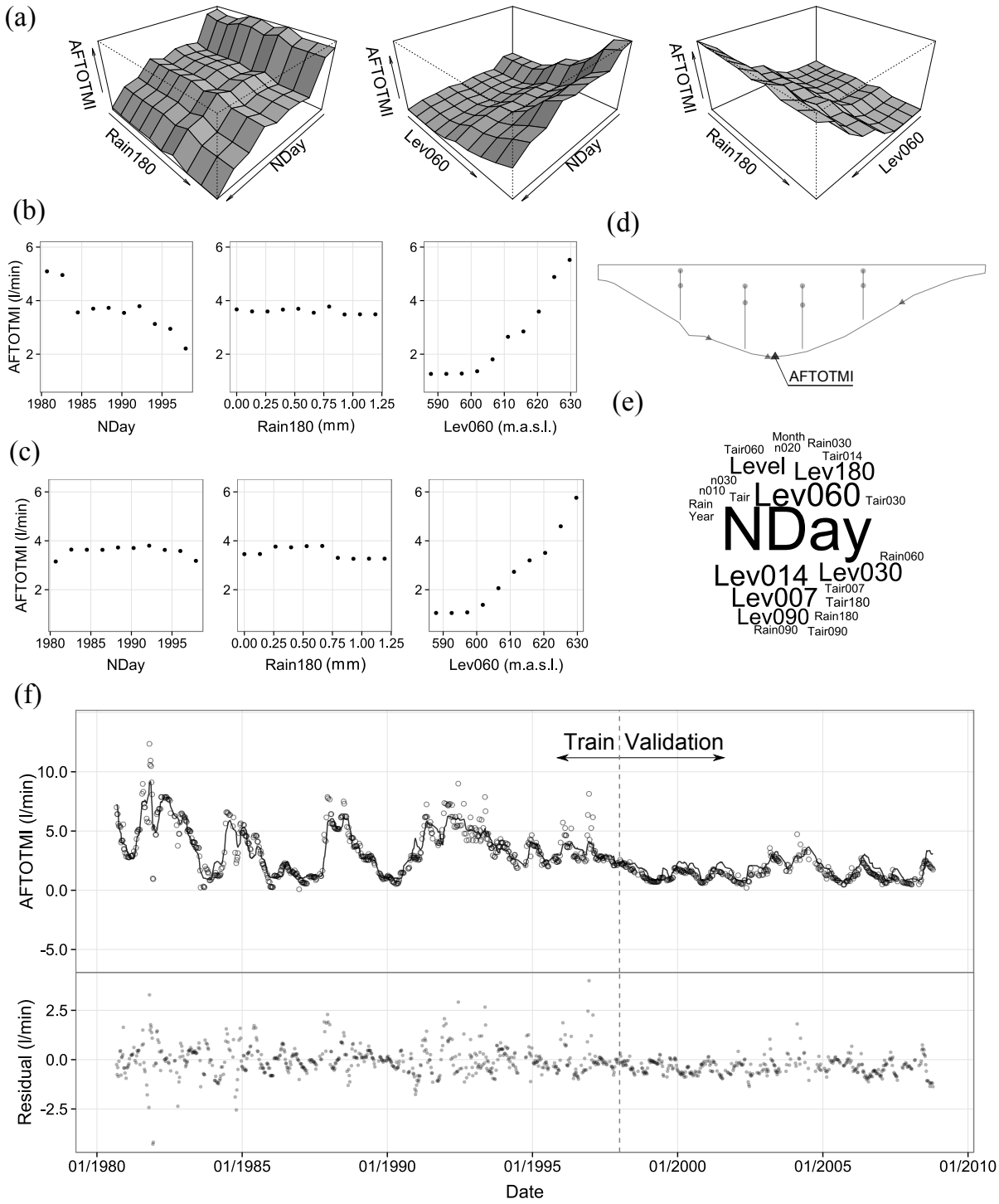

Figure A14: AFTOTMI. (a) 3D partial dependence plot; (b) 2D Partial dependence plot; (c) Idem for artificial data (time-independent); (d) Device location; (e) Word cloud of relative influence; (f) Model fit and residuals for the train and the validation sets.