Accepted: 30 May 2017
DOI: 10.1111/jac.12225

CLIMATE CHANGE

WILEY Journal of Agronomy and Crop Science

Designing an experiment with quantitative treatment factors to study the effects of climate change

H.-P. Piepho[1] | M. Herndl[2] | E. M. Pötsch[2] | M. Bahn[3]

[1]Biostatistics Unit, Institute of Crop Science, Universität Hohenheim, Stuttgart, Germany

[2]Institute of Plant Production and Cultural Landscape, Agricultural Research and Education Centre, Raumberg-Gumpenstein, Austria

[3]Institute of Ecology, University of Innsbruck, Innsbruck, Austria

Correspondence
H.-P. Piepho, Biostatistics Unit, Institute of Crop Science, Universität Hohenheim, Stuttgart, Germany.
Email: piepho@uni-hohenheim.de

Funding information
Austrian Ministry of Agriculture, Forestry, Environment and Water Management; Austrian Science Fund, Grant Number: P28572-B22

Abstract

Experiments for studying the effects of climatic change on ecosystems often involve manipulation of one or several quantitative treatment factors of interest. Response surface regression is the method of choice for these types of experiment. Here, we describe the development of a design of a free air CO_2 enrichment experiment with two quantitative treatment factors, that is, elevated temperature and CO_2 enrichment. The design strategy takes account of budget constraints imposing limitations on the number of plots with elevated temperature and CO_2 levels. The approach is based on polynomial regression models and is focussed on an efficient estimation of interaction between the two treatment factors. Extension to more than two factors is straightforward. An analysis of soil moisture data demonstrates the overall suitability of the proposed design to analyse non-linear interactions of two (or more) global change factors.

KEYWORDS
experimental design, field experiment, free air CO_2 enrichment, multifactorial experiment, polynomial regression, response surface regression

1 | INTRODUCTION

Experiments for studying the effects of climatic change on ecosystems often involve manipulation of one or several treatment factors of interest, such as the level of CO_2 concentration, air and soil temperature or changes in precipitation amounts or patterns (Beier et al., 2012; Knapp, Harper, Danner, & Lett, 2002; Morgan et al., 2011; Roy et al., 2016). A popular example is free air CO_2 enrichment (FACE) experiments (McLeod & Long, 1999). Such outdoor experiments can be used, for example, to inform crop management decisions (Roth & Gardner, 1989; Sanchez, 2000), or evaluate the performance of simulation models (Tubiello et al., 1999). Quite often the treatment factors are quantitative in nature, meaning that regression is the preferred method for statistical analysis (Edmondson, 1991; Piepho & Edmondson, 2017), rather than pairwise comparison of treatment means, by far the most commonly used method of analysis for designed experiments in agriculture and ecology. When ecosystem responses to environmental treatment factors are nonlinear, polynomial regression provides a convenient modelling framework. With more than one such treatment factor, response surface regression is a particularly appealing method of analysis, a method that is widely used in experiments for process optimization in various industrial branches (Box & Draper, 2007; Dean & Voss, 1999). Design choices concern the placement and number of levels for each treatment factor as well as the number of replications for each factorial treatment combination (Atkinson, Donev, & Tobias, 2007). Designs for response surface regression used in industrial applications, where the primary objective is output maximization, typically differ from those commonly used for factorial field and greenhouse experiments. To date, only few ecosystem experiments have considered the necessity to account for nonlinear responses at the design stage (Zhou, Weng, & Luo, 2008) and have largely been restricted to single-factor experiments (e.g., Gill et al., 2002). The realization of multistep experiments accounting for interactive effects between multiple environmental drivers (Dieleman et al., 2012; Leuzinger et al., 2011; Luo et al., 2011) has typically been constrained by the inevitable trade-off between the number of treatments imposed and the replication required for ANOVA-based analyses comparing treatment means. Designs for response surface regression are an attractive alternative approach to optimize statistical power and capture nonlinear responses with a smaller number of plots, which can be implemented cost-effectively.

The Lysi-T-FACE experiment (Herndl, Pötsch, Bohner, & Kandolf, 2011) was set up in 2011 at the Agricultural Research and Education Centre Raumberg-Gumpenstein (Austria) to assess the effects of global warming, that is elevated CO_2 and temperature on the productivity of a managed mountainous grassland system. The Lysi-T-FACE experiment entails a multifactorial approach for different levels of warming and elevated CO_2 concentration. In this paper we describe our strategy to design the Lysi-T-FACE experiment based on response surface regression models, enabling an efficient estimation of treatment effects. We believe that this strategy is of general interest to researchers conducting global change experiments. It involves some differences to approaches used for response surface regression in industrial contexts, but also to designs traditionally used for factorial experiments in agriculture and ecology, where treatment levels are mainly varied on a regular grid with equal replication. In the following sections, we (i) briefly review polynomial regression modelling for factorial experiments, (ii) consider criteria for evaluating design efficiency as relevant to our purpose, (iii) turn to the design of the Lysi-T-FACE experiment exemplifying the analysis using response surface regression and (iv) finally discuss the chosen approach.

2 | MATERIALS AND METHODS

2.1 | Polynomial regression models for quantitative treatment factors

Response surface regression is a special kind of multiple regression (Box & Draper, 2007; Dean & Voss, 1999, p. 547–592; Lenth, 2009), in which the response to changes in one or several input variables is assessed. Response surface regression was developed mainly in manufacturing and chemical engineering. In these applications, the objective is to optimize a production process with respect to the levels of a number of input variables, and this is achieved using a step-wise approach. The process is started with a good and realistic estimate of the optimal production point for the relevant input variables. A few points around this expected optimum are then explored, and ultimately, a response surface model is fitted that allows analytical determination of the optimum. The most commonly used models for this purpose are second-degree polynomials.

In agriculture and ecology, such models can also be used to study the joint effect of several quantitative factors on a response (Edmondson, 1991). If the primary aim is an optimization of a production system, the same designs and analytical methods primarily developed for industrial applications can be used, even though a sequential approach as used in engineering is not usually a feasible option. By contrast, in the experiment considered here, the objective is not the optimization of a system, but rather an assessment of the effect of elevated temperature and CO_2 on relevant outcome variables. While the objectives (and the design) for such trials may be different, the regression models suitable for analysis are the same.

For the two input variables "temperature elevation" (x_T) and "CO_2 enhancement" (x_C), a linear response surface can be represented by the following first-order model:

$$E(y) = \alpha + \beta_T x_T + \beta_C x_C, \quad (1)$$

where α is an intercept and β_T and β_C are regression coefficients. This model can also be tested for lack-of-fit, provided that at least some of the design points (x_T, x_C) are replicated (Dean & Voss, 1999). To implement this analysis, each design point (x_T, x_C) receives its own level for a lack-of-fit variable, denoted here as "lackfit." This is used as a qualitative treatment factor to fit an effect δ for each individual design point (x_T, x_C):

$$E(y) = \alpha + \beta_T x_T + \beta_C x_C + \delta. \quad (2)$$

When there is no significant lack-of-fit, one may proceed by reporting the fit of model (1). Otherwise, the degree of the polynomial may be increased, leading to a second-order model:

$$E(y) = \alpha + \beta_T x_T + \beta_{TT} x_{TT}^2 + \beta_C x_C + \beta_{CC} x_C^2 + \beta_{TC} x_T x_C, \quad (3)$$

where α is an intercept and β_T, β_{TT}, β_C, β_{CC}, and β_{TC} are regression coefficients. This second-order model may have different shapes depending on the sign and magnitude of the regression coefficients. For example, the response surface can either have a maximum or a minimum. Moreover, it is possible that the surface has no optimum, but takes the shape of a saddle having a "saddle point," at which the slope of the surface is zero in all direction (stationary point). Which of these three cases applies, can be checked using the so-called canonical coefficients. The second-order model (3) can also be tested for lack-of-fit by adding an effect δ as with the first-order model in Equation 2. For details, the reader is referred to pertinent textbooks such as Box & Draper (2007) and Dean & Voss (1999).

2.2 | Optimality criteria for experimental designs and common designs for response surface regression

Common measures of design efficiency are based on the information matrix $\boldsymbol{X}^T\boldsymbol{X}$, where $\boldsymbol{X}$ is the design matrix and $\boldsymbol{X}^T$ its transpose (Gilmour & Trinca, 2012). For the second-order model, this matrix has six columns corresponding to the six regression parameters in Equation 3, and the rows correspond to the plots of the design. The information matrix is important because it is proportional to the inverse of the variance–covariance matrix of the least squares estimates of the regression parameters. Loosely speaking, a good design minimizes the variance of the parameter estimates, as quantified by $(\boldsymbol{X}^T\boldsymbol{X})^{-1}$, and this maximizes the information about the parameters, as quantified by the information matrix $\boldsymbol{X}^T\boldsymbol{X}$. Two commonly used efficiency measures are

$$\text{D-efficiency} = \frac{[\det(\boldsymbol{X}^T\boldsymbol{X})]^{1/p}}{n} \times 100$$

and

$$\text{A-efficiency} = \frac{p/n}{\text{trace}[(\boldsymbol{X}^T\boldsymbol{X})^{-1}]} \times 100,$$

where p is the number of columns in $\boldsymbol{X}$ and n is the number of plots (Atkinson et al., 2007). The D- and A-efficiencies can be viewed as measures quantifying how large $\boldsymbol{X}^T\boldsymbol{X}$ and $(\boldsymbol{X}^T\boldsymbol{X})^{-1}$ are. They have an interpretation as the relative numbers of design points, in per cent, for a hypothetical orthogonal design to achieve the same values of det($\boldsymbol{X}^T\boldsymbol{X}$) and trace[($\boldsymbol{X}^T\boldsymbol{X}$)$^{-1}$], respectively, as the candidate design (Mitchell, 1974). Large values of D-efficiency and A-efficiency are desirable. Both of these measures are integral measures of the precision of the regression coefficients. D-optimality is a criterion which has proven useful in the context of regression analysis, where it has an interpretation in terms of its equivalence with minimizing the maximum variance of predicted responses (John & Williams, 1995, p. 32).

As one of our objectives is to study the interaction, we will also consider the variance of the estimate of the interaction effect β_{TC}, which is proportional to the corresponding diagonal element of $(\boldsymbol{X}^T\boldsymbol{X})^{-1}$. Small values of the variance (or its square root, the standard error) are desirable. A convenient method to compute all of these measures is to perform a dummy analysis for a given design $\boldsymbol{X}$, in which we set the data to arbitrary values and perform an analysis for the linear model corresponding to the design matrix $\boldsymbol{X}$, fixing the residual variance σ^2 at a pre-specified value. This can be done, for example using the MIXED procedure of SAS (Stroup, 2002). Without loss of generality we here fix the variance at $\sigma^2 = 10^6$.

When the objective of an experiment is to find the optimal production point, classical designs for estimating first-order and second-order models are available. For example, the central composite design and the Box-Behnken design (Box & Draper, 2007; Cochran & Cox, 1957, p. 342 ff.; Dean & Voss, 1999; Lenth, 2009; Roth & Gardner, 1989) are the most commonly used designs based on the second-order model. We will not go into detail here because optimization is not our objective. However, we mention here that the central composite design has design points not on a rectangular grid for the input variables, but the design points are placed on an ellipsoid. Also, the centre point of the design is usually replicated more often than are the other design points because this improves precision of the estimate of the point optimizing the response. The design we develop here will not have this property because our objective is different and there are economic constraints as will be explained in the next section.

2.3 | Designing the Lysi-T-FACE experiment in Raumberg-Gumpenstein

The Lysi-T-FACE experiment aims to analyse effects of warming and elevated CO_2 on the productivity of managed grassland. The experiment was to follow a multifactor approach combining two levels of warming and two levels of elevated atmospheric CO_2, including control levels for each of the two factors. It was determined that temperature elevation should be varied between 0°C and 3°C, whereas CO_2 elevation should be varied between 0 and 300 ppm.

We decided to use a completely randomized design. Thus, the design problem boils down to selecting the treatment combinations and their replications. The design can also be blocked in principle (Atkinson et al., 2007), but we decided against this option because large spatial trends were not expected for the experimental field at hand and because blocking costs error *df* which would be limiting with a small number of plots.

The traditional experimental approach for characterizing responses to two interacting factors is to use a full factorial design. For example, a 3 × 3 factorial design with three equally spaced levels for each factor and four replications per treatment would require a total of 36 plots (Table 1). This design would have 24 plots with elevated temperature and 24 plots with elevated CO_2. Due to the fact that the FACE equipment used for elevated treatment levels is very expensive and field space and labour capacity pose constraints, the use of a design with such a large number of plots with elevated temperature or CO_2 was not feasible.

The size of the experimental facility was limited to 24 plots due to budget constraints. Moreover, the budget allowed for a total of 20 plots with temperature or CO_2 elevation. We specifically considered design scenarios for 10 plots with elevated temperature and 10 plots with elevated CO_2. An important consequence of the budget constraints is that the design must have 14 plots with no elevation in temperature and 14 plots with no elevation in CO_2. This, in turn, constrains the number of plots for the control (no elevation for either factor) to lie between 4 and 8.

If we use 4 plots for the control, we must have 10 further plots with no elevation in temperature and 10 further plot with no elevation in CO_2, leaving no plots for treatments where both temperature and CO_2 are elevated. An example of such a design in a 3 × 3 layout is given in Table 2. This is not a useful design because it does not permit interactions to be tested. For this reason, we only considered designs with more than 4 plots for the control. It is convenient to use the design in Table 2 as a starting point for developing such designs, keeping track of the constraints as the start design is modified.

When the control has the maximum of eight replications, we can have only four plots where both temperature and CO_2 are elevated. This, in turn, limits the total number of treatment combinations that can be accommodated in the overall design. We therefore

TABLE 1 An equally replicated design for a 3 × 3 factorial experiment with 36 plots

	CO_2 (ppm)		
Temperature (°C)	**0**	**150**	**300**
0	4	4	4
1.5	4	4	4
3	4	4	4

TABLE 2 Start configuration for a 3 × 3 design with 24 plots

	CO_2 (ppm)		
Temperature (°C)	**0**	**150**	**300**
0	4	5	5
1.5	5		
3	5		

specifically considered 3 × 3 and 4 × 4 factorial designs with an equal spacing of factor levels. As will be detailed below, we finally settled for a 3 × 3 design, so we will primarily focus on designs of this dimension.

The allocation of treatments to plots was to be chosen so that treatment effects can be efficiently estimated. There were several effects and treatment comparisons of interest:

1. The design should permit an efficient estimation of treatment effects via a response surface regression approach.
2. The design should allow for testing the lack-of-fit of regression models.
3. Comparisons with the control serve to establish that there are treatment effects at all, so the control treatment plays a key role in the design and should have extra replication.
4. The design should be optimized to detect the interaction between the two treatment factors (warming and CO_2 elevation).

To set the stage, we will first consider a D-optimal design generated with the OPTEX procedure of SAS (Atkinson et al., 2007). Candidate runs (design points) were placed on a 3 × 3 grid with $x_T \in (0, 1.5, 3)$ and $x_C \in (0, 150, 300)$ (Table 3). A D-optimal design would place 14.6% of the observations at each of the four vertices, 8.0% at each of the mid-points of the four edges, and 9.6% at the origin (Box & Draper, 2007, p. 475; citing Fedorov 1972). Of course these proportions are not realizable in small samples. The design in Table 3 comes reasonably close to this optimum, however, and it also produces the smallest standard error for interaction among all 3 × 3 designs considered (see below). But it has 14 plots with elevated temperature and 15 plots with elevated CO_2. While the design violates the budget constraints, it is a reasonable benchmark for alternative designs.

We now consider designs observing the budget constraints allowing 10 plots with elevated CO_2 and 10 plots with elevated temperature. The start design in Table 2 does not allow estimating a second-order model because it only has five design points, but there are six parameters. To identify all parameters, at least six design points are needed. For testing the lack-of-fit, at least one additional design point is needed, amounting to a minimum of seven design points. To accommodate additional design points, but at the same time meet the budget constraints, we need to increase the number of plots for the control treatment and at the same time reduce the number of treatments with elevated temperature only and elevated CO_2 only. Five such designs are shown in Tables 4–8. The design in Table 8 is the most efficient design, followed by that in Table 5. Table 8 comes closest to the D-optimal design in Table 3 as regards the dominance of the four vertices. We prefer the design in Table 5, however, because the two treatments with both factors at elevated level are equally replicated and therefore will have a better power for lack-of-fit. We also note that this design has two tetrads that allow testing interaction between the two factors using a cell means model. A tetrad is a 2 × 2 subtable with all cells filled. Thus, one tetrad is given by the cells (0, 0), (3, 0), (0, 300), (3, 300) and the other tetrad is represented by the cells (0, 0), (1.5, 0), (0, 150), (1.5, 150). Moreover, the design has a third of the plots allocated to the control, which is desirable because comparisons with the control are important to establish any effects due to elevated temperature or CO_2 (Dunnett, 1955). Also note in this regard that the control is the only cell that is involved in both tetrads.

We now turn to 4 × 4 designs. Three examples are given in Tables 9–11. For comparison, we also report a D-optimal design in Table 12. This provides better D-efficiency and a smaller standard error of β_{TC} than the designs in Tables 9–11, but it does not obey the budget constraints. On the whole, the designs in Tables 9–11 are not as D-efficient as the 3 × 3 designs. In some cases, the

TABLE 3 D-optimal 3 × 3 design for 24 plots (generated with the SAS procedure OPTEX)

	CO_2 (ppm)		
Temperature (°C)	0	150	300
0	4/24 = 0.167	2	4
1.5	2/24 = 0.083	2	2
3	3/24 = 0.125	2	3

D-efficiency = 102.1, A-efficiency = 99.0, $SE(\beta_{TC})$ = 1.1974 for $\sigma^2 = 10^6$. Note that there are four possible D-optimal 3 × 3 designs with equal D-efficiency, involving to permutations of the number of replications at the four vertices, and different runs of OPTEX can produce different D-optimal designs.

TABLE 4 A 3 × 3 design for 24 plots

	CO_2 (ppm)		
Temperature (°C)	0	150	300
0	6	4	4
1.5	4	1	
3	4		1

D-efficiency = 38.9, A-efficiency = 19.6, $SE(\beta_{TC})$ = 1.4184 for $\sigma^2 = 10^6$.

TABLE 5 A 3 × 3 design for 24 plots

	CO_2 (ppm)		
Temperature (°C)	0	150	300
0	8	3	3
1.5	3	2	
3	3		2

D-efficiency = 40.7, A-efficiency = 1.4, $SE(\beta_{TC})$ = 1.2481 for $\sigma^2 = 10^6$.

TABLE 6 A 3 × 3 design for 24 plots

	CO_2 (ppm)		
Temperature (°C)	0	150	300
0	8	3	3
1.5	3	1	1
3	3	1	1

D-efficiency = 40.3, A-efficiency = 21.5, $SE(\beta_{TC})$ = 1.3987 for $\sigma^2 = 10^6$.

TABLE 7 A 3 × 3 design for 24 plots

	CO_2 (ppm)		
Temperature (°C)	0	150	300
0	8	3	3
1.5	3		2
3	3	2	

D-efficiency = 35.6, A-efficiency = 15.3, $SE(\beta_{TC})$ = 1.8703 for $\sigma^2 = 10^6$.

TABLE 8 A 3 × 3 design for 24 plots

	CO_2 (ppm)		
Temperature (°C)	0	150	300
0	8	3	3
1.5	3	1	
3	3		3

D-efficiency = 40.9, A-efficiency = 18.5, $SE(\beta_{TC})$ = 1.1714 for $\sigma^2 = 10^6$.

TABLE 9 A 4 × 4 design for 24 plots

	CO_2 (ppm)			
Temperature (°C)	0	100	200	300
0	8	2	2	2
1	2	1		
2	2		1	
3	2			2

D-efficiency = 36.6, A-efficiency = 18.2, $SE(\beta_{TC})$ = 1.3309 for $\sigma^2 = 10^6$.

TABLE 10 A 4 × 4 design for 24 plots

	CO_2 (ppm)			
Temperature (°C)	0	100	200	300
0	8	1	2	3
1	1	1		
2	2		1	
3	3			2

D-efficiency = 38.3, A-efficiency = 17.5, $SE(\beta_{TC})$ = 1.2149 for $\sigma^2 = 10^6$.

TABLE 11 A 4 × 4 design for 24 plots

	CO_2 (ppm)			
Temperature (°C)	0	100	200	300
0	8	1	2	3
1	1			
2	2		1	
3	3			3

D-efficiency = 38.3, A-efficiency = 15.1, $SE(\beta_{TC})$ = 1.1417 for $\sigma^2 = 10^6$.

precision for the interaction effect is slightly better than for a similar 3 × 3 design. A downside of the 4 × 4 designs is that some cells only have one replication and so there is an increased risk of losing

TABLE 12 D-optimal 4 × 4 design for 24

	CO_2 (ppm)			
Temperature (°C)	0	100	200	300
0	3	1	1	3
1	2			2
2	1		2	
3	3	2		4

D-efficiency = 113.6, A-efficiency = 98.5, $SE(\beta_{TC})$ = 1.1897 for $\sigma^2 = 10^6$. Note that this design does not observe the budget constraints.

a whole treatment in case a plot fails to produce results, so they are expected to be less robust against missing plots. Our final decision therefore was to go with the 3 × 3 design in Table 5, because it has a minimum of two replications for the chosen design points. It is acknowledged that design choice is not clear-cut, because there are several factors to balance. We do think that the design in Table 5 provides a viable compromise. The computer code for generating and evaluating the designs in Tables 3–12 is provided as Appendix S1.

2.4 | Analysis options, exemplified using preliminary results

2.4.1 | Polynomial regression and lack-of-fit testing

To illustrate the analysis for the chosen design, we here consider data on soil moisture (in per cent) obtained from the Lysi-T-FACE experiment on 6 May 2016. Three repeat measurements were taken per plot and averaged for further analysis. The second-order model shows no indication of lack-of-fit (p = .1176; Table 13), but also the second-order terms were not significant. Thus, a first-order model was tried, which showed no lack-of-fit either (Table 14). It turned out, however, that only the linear effect for temperature was significant, so the model may be further reduced accordingly (Table 15). The fitted linear model is shown in Figure 1 and had an R^2 of .329. Inspection of the regression plot suggests that a quadratic model might fit better. The plot also suggests that we could drop the linear term; that is, the model is $\alpha + \beta_{TT}x_T^2$ such that a maximum occurs at the control level of temperature ($x_T = 0$). This model indeed fitted

TABLE 13 Second-order model for soil moisture data (6 May 2016) with test for lack-of-fit

Source	*df*	Sum of squares	*F*-value	*p*-Value
CO_2 (C)	1	5.34	2.34	.1444
Temperature (T)	1	26.08	11.43	.0036
C^2	1	2.56	1.12	.3042
T^2	1	0.52	0.23	.6393
C × T	1	2.53	1.11	.3069
Lack-of-fit	1	6.21	2.72	.1176
Error	17	38.81		

TABLE 14 First-order model for soil moisture data (6 May 2016) with test for lack-of-fit

Source	*df*	Sum of squares	*F*-value	*p*-Value
CO_2 (C)	1	5.34	2.34	.1444
Temperature (T)	1	26.08	11.43	.0036
Lack-of-fit	4	11.82	1.29	.3114
Error	17	38.81		

TABLE 15 Linear model in temperature for soil moisture data (6 May 2016) with test for lack-of-fit

Source	*df*	Sum of squares	*F*-value	*p*-Value
Temperature (T)	1	26.99	11.82	.0031
Lack-of-fit	5	16.25	1.42	.2657
Error	17	38.81		

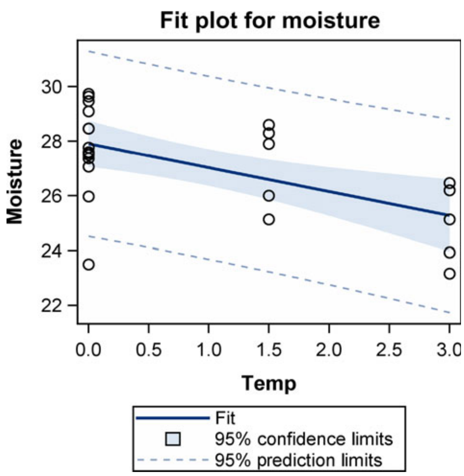

FIGURE 1 Plot of the fitted linear regression on temperature (R^2 = .329)

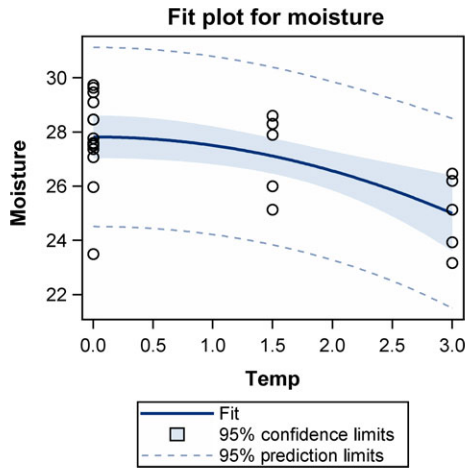

FIGURE 2 Plot of a quadratic regression on temperature that omits the linear term $(\alpha + \beta_{TT}x_T^2)$ (R^2 = .357)

TABLE 16 Quadratic model in temperature for soil moisture data (6 May 2016) with test for lack-of-fit. The model does not include a linear term

Source	*df*	Sum of squares	*F*-value	*p*-Value
Temperature (T^2)	1	29.33	12.85	.0023
Lack-of-fit	5	13.91	1.22	.3425
Error	17	38.81		

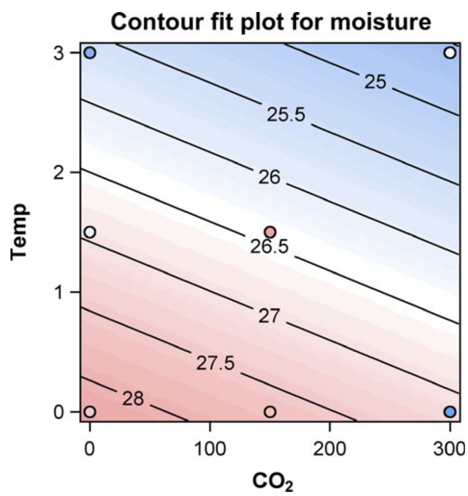

FIGURE 3 Contour plot for the fitted first-order model

better than the linear one in terms of the coefficient of determination (Figure 2; R^2 = .357) and the significance for the regression term (Table 16). Note that when the linear term is added $(\alpha + \beta_T x_T + \beta_{TT}x_T^2)$, the quadratic model will have an optimum at $x_T = -\beta_T/(2\beta_{TT})$, which is not at the control level of temperature ($x_T = 0$). We also note that the quadratic term was not significant (p = .3422), when the linear term was included as well.

For illustration of the output that can be produced with these types of model, we also show a contour plot for the first-order model (Figure 3) as well as a contour plot and 3D plot for the second-order model (Figure 4), but these should not be over-interpreted because the second-order terms as well as the first-order term for CO_2 were non-significant (Tables 13 and 14). The contour plots (Figures 3 and 4, panel a) show the seven design points in the CO_2-temperature plane and depict the moisture response as contours of equal moisture.

2.4.2 | Two-way analysis of variance and test of contrasts (tetrads, comparison with control)

We here analyse the experiment using a two-way analysis of variance (ANOVA) regarding both treatment factors as qualitative and show how to test treatment contrasts of interest. The ANOVA is shown in Table 17. Neither the interaction nor the main effect for CO_2 are significant at the 5% level, but the main effect for temperature is significant (p = .0179). Table 18 shows the seven treatment

TABLE 17 Two-way ANOVA for soil moisture data (6 May 2016) using sequential (Type I) sums of squares (Nelder, 1994)

Source	*df*	Sum of squares	*F*-value	*p*-Value
Temperature (T)	2	29.36	6.43	.0083
CO_2 (C)	2	5.16	1.13	.3470
C × T	2	8.74	1.91	.1780
Error	17	38.81		

TABLE 18 Treatment means for soil moisture data (6 May 2016)

	CO_2 (ppm)			
Temperature (°C)	0	150	300	Marginal means
0	28.43	27.34	26.58	27.80
1.5	26.34	28.45		27.19
3	25.18		24.68	24.98

None of the pairwise comparisons are significant at the 5% level by a *t* test.

means for the CO_2-by-temperature combinations. In addition, we reported the marginal means for temperature, because only the temperature main effect was significant (Table 17). The treatment means suggest that a relevant effect of elevated temperature only sets in above a 1.5°C increase, and this agrees with the fit of the quadratic model in Figure 2.

For illustration, we tested the two tetrad contrasts for interaction and the contrast of all treatments versus the control. Note that the tetrads provide a very lucid way to inspect the interaction of the two treatment factors when these are modelled as qualitative. Table 19 shows the contrasts coefficients along with the results. For example, the comparison of the control versus the other six treatments has a coefficient of 1 for the control and of −1/6 for all other treatments and is significant (p = .074). The two tetrad contrasts are not significant, coinciding with the non-significant *F* test of the two-way ANOVA. If the two interaction contrasts are tested jointly using an *F* test, the *p*-value coincides with that of the *F* test for interaction in Table 17.

3 | DISCUSSION AND CONCLUDING REMARKS

We believe that response surface regression methods hold much merit in global change experiments, because the treatment factors involved are typically quantitative in nature. Such regressions methods are preferable to classical approaches based on mean comparisons because they allow a more efficient estimation of treatment effects, due to more parsimonious modelling. Also, regression approaches allow interpolation to treatment levels that have not been tested but are well within the space spanned by the quantitative treatment factors. The design for the Lysi-T-FACE experiment comprises two quantitative factors, that is temperature elevation and CO_2 elevation, each with three levels. The approach is easily extended to accommodate more than two treatment factors and

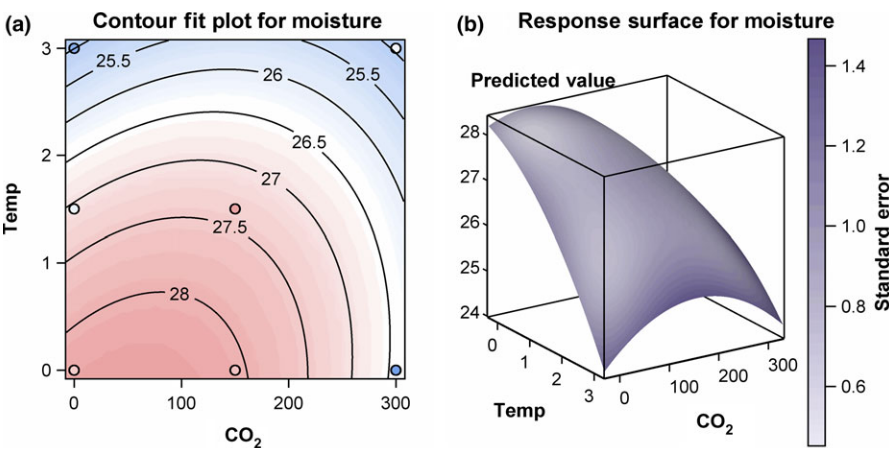

FIGURE 4 Contour plot (a) and 3D-plot (b) for the fitted second-order model

TABLE 19 Tests of tetrad contrasts for interaction and contrast of control versus all other treatments for soil moisture data (6 May 2016)

Treatment	1	2	3	4	5	6	7			
Temperature	0	1.5	3	0	1.5	0	3			
CO_2	0	0	0	150	150	300	300			
	Contrast coefficients							Estimate	*SE*	*p*-Value
Control versus rest	1	−1/6	−1/6	−1/6	−1/6	−1/6	−1/6	1.9995	0.6583	.0074
Tetrad 1	1	−1	0	−1	1	0	0	3.1902	1.7172	.0806
Tetrad 2	1	0	−1	0	0	−1	1	1.3569	1.7172	.4403

more than three treatment levels. Our strategy, which accounts for budget constraints, thus limiting the replication for certain treatment combinations in favour of others, is readily adapted to other global change experiments.

We designed the experiment for a given total number of experimental units (plot) that was pre-set based on budget constraints. For a given total number of plots, the design can be optimized with respect to the allocation of treatments as was shown in this paper. It should be emphasized, however, that total sample size can be adapted to meet a pre-set precision requirement, which can be defined based on the variances and covariances of the parameter estimates (Atkinson et al., 2007). For example, in the Lysi-T-FACE experiment the sample size could have been further increased to achieve a pre-determined value for the standard error of the parameter estimate for the interaction effect β_{CT}. This option was not explored here, however, due to budget constraints on the total number of plots that could be realized.

In the illustration of statistical analysis options we restricted attention to the case where there is a single observation of the response per plot. Often there are repeated measurements taken at different time points on the same plot, and an objective of the analysis of such data is to compare the time courses between different treatments. In this case, serial correlation among repeated measurements taken on the same plot needs to be taken into account. For an illustration using polynomial regression with repeated measures data, see Piepho & Edmondson (2017).

We should also mention here that measurements of response variable before start of the treatments can be considered as covariate in an analysis of covariance (Milliken & Johnson, 2002). If a covariate can explain some part of the total experimental error, analysis of covariance can lead to substantial gains in precision (Mühleisen, Reif, Maurer, Möhring, & Piepho, 2013).

Currently, the experimental site Lysi-T-FACE has been extended further including additional plots to account also for interactions between the strongest warming and CO_2 treatment with extreme weather events. Several research projects, representing and connecting scientists from different Universities and Research Institutes, are currently carried out at the facility to study the impacts of future climate conditions and elevated CO_2 on the productivity, species abundance and forage quality, and on the biogeochemistry and related greenhouse gas fluxes.

ACKNOWLEDGEMENTS

The Lysi-T-FACE facility at Raumberg-Gumpenstein has been financially supported by the Austrian Ministry of Agriculture, Forestry, Environment and Water Management, the Federal Provinces of Styria, Tyrol, Vorarlberg, Salzburg and Lower Austria, and by the Austrian Science Fund (FWF-project P28572-B22).

REFERENCES

Atkinson, A. C., Donev, A., & Tobias, R. (2007). *Optimum experimental design, with SAS*. Oxford: Oxford University Press.

Beier, C., Beierkuhnlein, C., Wohlgemuth, T., Penuelas, J., Emmett, B., Körner, C., … Hansen, K. (2012): Precipitation manipulation experiments – challenges and recommendations for the future. *Ecology Letters*, *15*, 899–911.

Box, G. E. P., & Draper, N. R. (2007). *Response surfaces, mixtures, and ridge analysis*. New York, NY: Wiley.

Cochran, W. G., & Cox, G. M. (1957). *Experimental Designs*. New York: Wiley.

Dean, A., & Voss, D. (1999). *Design and analysis of experiments*. New York, NY: Springer.

Dieleman, W., Vicca, S., Dijkstra, F., Hagedorn, F., Hovenden, M., Larsen, K., … Janssens, I. A. (2012). Simple additive effects are rare: A quantitative review of plant biomass and soil process responses to combined manipulations of CO_2 and temperature. *Global Change Biology*, *18*, 2681–2693.

Dunnett, C. W. (1955). A multiple comparison procedure for comparing several treatments with a control. *Journal of American Statistical Association*, *50*, 1096–1121.

Edmondson, R. N. (1991). Agricultural response surface experiments based on four-level factorial designs. *Biometrics*, *47*, 1435–1448.

Gill, R., Polley, H., Johnson, H., Anderson, L., Maherali, H., & Jackson, R. (2002). Nonlinear grassland responses to past and future atmospheric CO_2. *Nature*, *417*, 279–282.

Gilmour, S. G., & Trinca, L. A. (2012). Optimum design of experiments for statistical inference. *Applied Statistics*, *61*, 345–401.

Herndl, M., Pötsch, E. M., Bohner, A., & Kandolf, M. (2011). Lysimeter als Bestandteil eines technischen Versuchskonzeptes zur Simulation der Erderwärmung im Grünland. 14. Gumpensteiner Lysimetertagung, 3.-4. Mai 2011, Bericht HBFLA Raumberg-Gumpenstein, 119–126.

John, J. A., & Williams, E. R. (1995). *Computer generated cyclic designs* (2nd ed.). London: Chapman and Hall.

Knapp, A. K., Harper, C. W., Danner, B. T., & Lett, M. S. (2002). Rainfall variability, carbon cycling, and plant species diversity in a mesic grassland. *Science*, *298*, 2202–2205.

Lenth, R. V. (2009). Response-surface methods in R, using rsm. *Journal of Statistical Software*, *32*(7).

Leuzinger, S., Luo, Y., Beier, C., Dielemann, W., Vicca, S., & Körner, C. (2011). Do global change experiments overestimate impacts on terrestrial ecosystems? *Trends in Ecology and Evolution*, *26*, 236–241.

Luo, Y., Melillo, J., Niu, S., Beier, C., Clark, J., Classen, A., … Torn, M. (2011). Coordinated approaches to quantify long-term ecosystem dynamics in response to global change. *Global Change Biology*, *17*, 843–854.

McLeod, A. R., & Long, S.P. (1999). Free-air carbon dioxide enrichment (FACE) in global change research: a review. *Adv. Ecol. Res. 28*, 1–56.

Milliken, G. A., & Johnson, D. E. (2002). *Analysis of messy data. Volume III: Analysis of covariance*. Boca Raton, FL, CRC Press.

Mitchell, T. J. (1974). Computer construction of 'D-optimal' first-order designs. *Technometrics*, *20*, 211–220.

Morgan, J., LeCain, D., Pendall, E., Blumenthal, D., Kimball, B., Carrillo, Y., … West, M. (2011). C4 grasses prosper as carbon dioxide eliminates desiccation in warmed semi-arid grassland. *Nature*, *476*, 202–206.

Mühleisen, J., Reif, J., Maurer, H. P., Möhring, J., & Piepho, H. P. (2013). Visual scorings of drought stress intensity as covariates for improved variety trial analysis. *Journal of Agronomy and Crop Science*, *199*, 321–330.

Nelder, J. A. (1994). The statistics of linear models: Back to basics. *Statistics and Computing*, *4*, 221–234.

Piepho, H. P., & Edmondson, R. N. (2017). A tutorial on polynomial regression for factorial experiments with quantitative treatment levels. *Journal of Agricultural Science*, (in revision).

Roth, R. L., & Gardner, B. R. (1989). Modified self-moving irrigation system for water-nitrogen crop production studies. *Applied Engineering in Agriculture*, *5*, 175–179.

Roy, J., Picon-Cochard, C., Augusti, A., Benot, M. L., Thiery, L., Darsonville, O., ... Soussana, J. F. (2016). Elevated CO_2 maintains grassland net carbon uptake under a future heat and drought extreme. *Proceedings of the National Academy of Sciences USA*, *113*, 6224–6229.

Sanchez, C. A. (2000). Response of lettuce to water and nitrogen on sand and the potential for leaching of nitrate-N. *HortScience*, *35*, 73–77.

Stroup, W. W. (2002). Power analysis based on spatial effects mixed models: A tool for comparing design and analysis strategies in the presence of spatial variability. *Journal of Agricultural, Biological and Environmental Statistics*, *7*, 491–511.

Tubiello, F. N., Rosenzweig, C., Kimball, B. A., Pinter, P. J. Jr, Wall, G. W., Hunsaker, D. J., ... Garcia, R. L. (1999). Testing CERES-wheat with free-air carbon dioxide enrichment (FACE) experiment data: CO_2 and water interactions. *Agronomy Journal*, *91*, 247–255.

Zhou, X., Weng, E., & Luo, Y. (2008). Modeling patterns of nonlinearity ecosystem responses to temperature, CO_2, and precipitation changes. *Ecological Applications*, *18*, 453–466.

SUPPORTING INFORMATION

Additional Supporting Information may be found online in the supporting information tab for this article.

How to cite this article: Piepho H-P, Herndl M, Pötsch EM, Bahn M. Designing an experiment with quantitative treatment factors to study the effects of climate change. *J Agro Crop Sci*. 2017;00:1–9. https://doi.org/10.1111/jac.12225