Evolutionary Visual Exploration: Evaluation With Expert Users

Nadia Boukhelifa, Waldo Gonzalo Cancino Ticona, Anastasia Bezerianos, Evelyne Lutton

▶ **To cite this version:**

Evolutionary Visual Exploration: Evaluation With Expert Users

N. Boukhelifa[1], W. Cancino[1], A. Bezerianos[2,1] and E. Lutton[1,3]

[1]INRIA Saclay - Île-de-France, France   [2]Univ Paris-Sud & CNRS, Orsay, France   [3]INRA, Grignon, France

Abstract

We present an Evolutionary Visual Exploration (EVE) *system that combines visual analytics with stochastic optimisation to aid the exploration of multidimensional datasets characterised by a large number of possible views or projections. Starting from dimensions whose values are automatically calculated by a PCA, an interactive evolutionary algorithm progressively builds (or evolves) non-trivial viewpoints in the form of linear and non-linear dimension combinations, to help users discover new interesting views and relationships in their data. The criteria for evolving new dimensions is not known a priori and are partially specified by the user via an interactive interface: (i) The user selects views with meaningful or interesting visual patterns and provides a satisfaction score. (ii) The system calibrates a fitness function (optimised by the evolutionary algorithm) to take into account the user input, and then calculates new views. Our method leverages automatic tools to detect interesting visual features and human interpretation to derive meaning, validate the findings and guide the exploration without having to grasp advanced statistical concepts. To validate our method, we built a prototype tool (EvoGraphDice) as an extension of an existing scatterplot matrix inspection tool, and conducted an observational study with five domain experts. Our results show that EvoGraphDice can help users quantify qualitative hypotheses and try out different scenarios to dynamically transform their data. Importantly, it allowed our experts to think laterally, better formulate their research questions and build new hypotheses for further investigation.*

1. Introduction

The purpose of visual exploration is to find meaningful patterns in the data which can lead to insight. In a high-dimensionality context, this task becomes rather challenging as viewers may be faced with a large space of alternative views on the data. One way to help navigate such a space is the "grand tour" method [Asi85] which offers a complete view of the search space in a smooth sequence of projections showing various viewpoints of the data. However, the time required to inspect all these views may be prohibitive [Hub85]. A related approach that improves on this is "projection pursuit" [Fri87] where the aim is to visit only the most interesting views; interesting referring to projections that deviate more from a normal distribution. The criteria for deciding whether a projection is interesting have mostly been defined prior to user exploration, using objective measures such as the quality metrics surveyed in [BTK11].

We present a novel visual analysis tool to explore multidimensional datasets where the system proposes interesting views based on both objective measures, such as different vi-sual patterns in the two-dimensional projections of the data, and subjective measures corresponding to user satisfaction with the presented view. These subjective measures are not known prior to user exploration. To demonstrate our ideas, we built a prototype (*EvoGraphDice*) as an extension of an existing scatterplot matrix inspection tool. We use low dimension projection to handle data multi-dimensionality, and linear and non-linear combinations of dimensions for an axis of the projection plane to propose alternative views. User exploration is guided by an Interactive Evolutionary Algorithm (IEA) which can both generate new views and adapt to user interest. Below, we provide background for the topic of evolutionary computation before listing our contributions.

Evolutionary Algorithms (EAs) are stochastic optimisation heuristics that copy, in a very abstract manner, the principles of natural evolution that let a population of individuals be adapted to its environment [Gol89]. They have the major advantage over other optimisation techniques of making only few assumptions on the function to be optimised. In short, an EA considers populations of potential so-
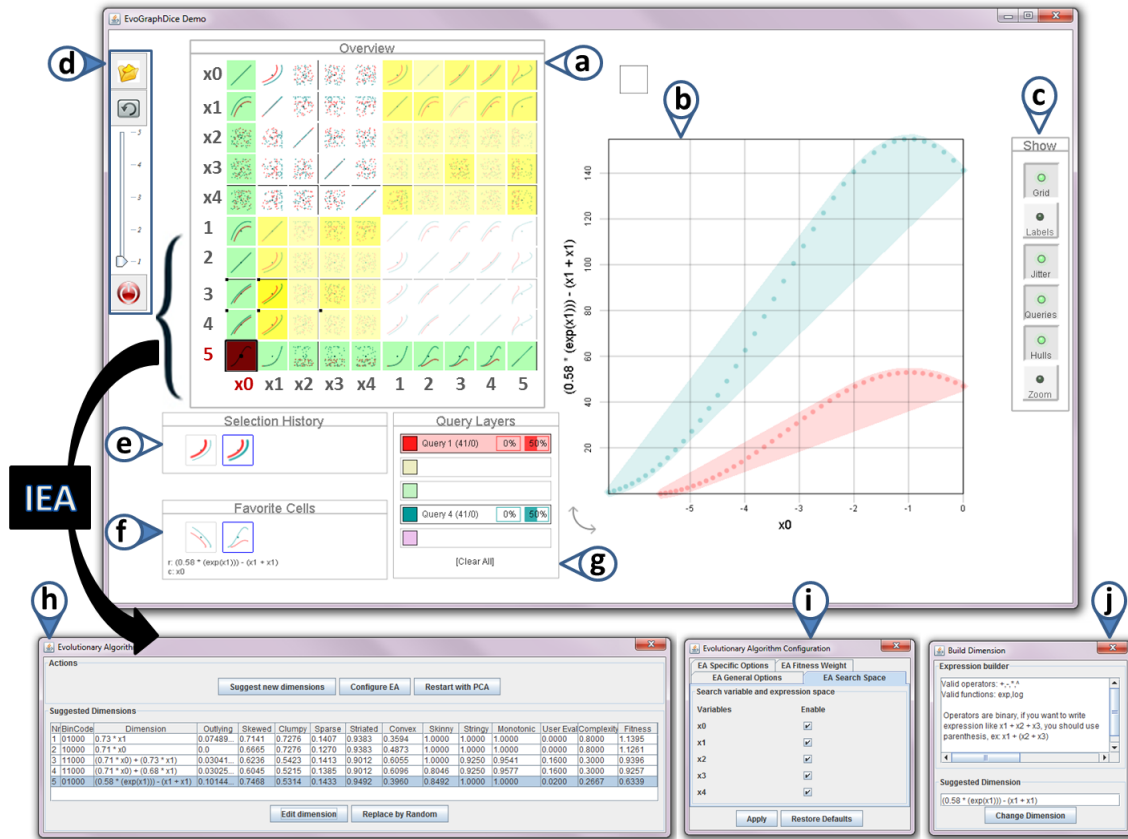
Figure 1: *EvoGraphDice prototype showing an exploration session of a synthetic dataset. New extensions to the GraphDice system are indicated by coloured label arrows. Widgets: (a) an overview scatterplot matrix showing the original data set of 5 dimensions (x0..x4) and the new dimensions (1..5) as suggested by the evolutionary algorithm. (b) main plot view. (c) tool bar for main plot view. (d) a tool bar with (top to bottom)"favorite" toggle button, "evolve" button , a slider to evaluate cells and a restart (PCA) button. (e) the selection history tool. (f) the favorite cells window. (g) the selection query window. (h) IEA main control window. (i) window to limit the search space. (j) dimension editor.*

lutions exactly like a natural population of individuals that live, fight, and reproduce, but the natural environment pressure is replaced by an "optimisation" pressure. In this way, individuals that reproduce are the best ones with respect to the problem to be solved. Reproduction consists of generating new solutions via variation schemes (the genetic operators), that, by analogy with nature, are called mutation if they involve one individual, or crossover if they involve two parent solutions. A *fitness function*, computed for each individual, is optimised by the EA. Evolutionary optimisation techniques are particularly efficient to address complex problems (irregular, discontinuous) where classical deterministic methods fail [Ban97, PLM08], but they can also deal with varying environments [JB05], or non computable quantities [Tak08]. More specifically, Interactive Evolutionary Algorithms (IEAs) are focussed on the optimisation of subjective quantities captured via a user interface.

Evolutionary Visual Exploration (EVE): we feel that

Interactive Evolutionary Algorithms (IEA) are convenient for guiding the user in exploring complex datasets. This opinion is founded by the following characteristics of EAs: *(i) focus:* an IEA performs an optimisation, i.e. it drives the exploration towards "interesting" areas of the search space (areas of high fitness function and good user satisfaction), *(ii) diversity:* by nature, an IEA has a stochastic behaviour, and its population-based scheme allows to display a variety of solutions to the user at any time, *(iii) adaptation:* EAs are able to deal with time varying environments and are able to follow changes of user interest and focus [Lut06].

The contributions of this paper are: (1) a framework for Evolutionary Visual Exploration (EVE) that marries techniques from visual analysis and evolutionary computation to guide user exploration towards interesting views on the data; (2) a prototype tool (*EvoGraphDice*) [Evo] to demonstrate our framework; and (3) an observational study with five domain expert users to evaluate *EvoGraphDice*.

2. Related Work

Related work is organised as follows; (1) a brief overview of quality metrics used to describe specific properties of data projections; (2) description of quality metrics we use in this work as part of the automatic evaluation of scatterplots; and (3) a summary of related work to IEA.

Quality Metrics: faced with the overwhelming possibilities of exploration paths in multidimensional visualization, researchers in the field designed quality metrics that evaluate the various projections of the data, in the hope of focusing user search on the most promising views. In a recent survey, Bertini et al. [BTK11] used the data flow model to classify quality metrics into three types: metrics that draw information from the data space, from the image space or from both.

Amongst metrics calculated at the data space are clustering and outliers. The rank-by-feature framework [SS05], for instance, visualises an optimal set of features according to a user selected quality metric such as correlation or uniformity. They use axis-parallel projections to produce 1D or 2D views and color brightness to denote ranking scores. Amongst image based metrics are scagnostics [WW08] which describe measures of interest for pairs of dimensions based on their geometrical appearance on a scatterplot. The mixed metrics combine information from the data and image spaces at the same time. Peng et al. [PWR04], for example, combine data features such as correlation information with view features such as axes adjacency to measure clutter as a result of reordering visualization axes [BTK11].

When interaction with quality metrics is available, it is either to select a metric amongst others, or to set threshold values [BTK11]. Having to specify the type of ranking criteria requires users to be familiar with advanced statistical concepts. In our case, the quality metric is pre-defined as a vector of nine image-based measures (described in the next section) and the threshold values are adapted according to user feedback. Thus, our method leverages automatic tools to detect interesting features and human interpretation to derive meaning, validate the findings and guide the exploration without having to grasp advanced statistical concepts.

Scagnostics[†] are based on geometric graphs which are calculated from areas, perimeters and lengths of these graphs. They include nine measures to characterise scatterplots (Fig. 2) and are useful for quickly discovering regularities and anomalies in scatterplot matrices. The underlying algorithm detects different types of point distributions including multivariate normal, log normal, multinomial, sparse, dense, convex and clusters. It does so by binning, detecting outliers and computing measures based on the following three statistical properties: *shape* for convex, skinny and stringy distributions; *trend* for monotonic distributions; and

---

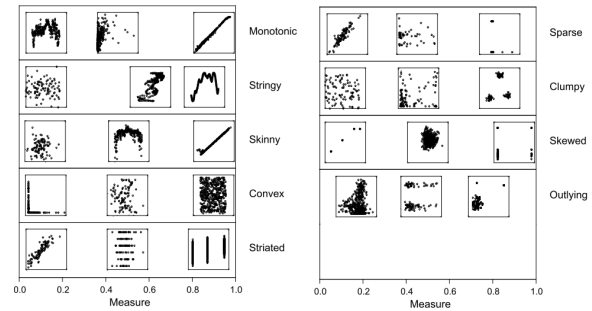[†] Available as a free downloadable package in R from http://www.rforge.net/scagnostics/

Figure 2: *Nine scagnostics measures from [WW08].*

density for skewed, clumpy, outlying, sparse and striated. These measures have proven statistical properties and are computable for moderately large data sets [WW08].

Visualization and IEA: visualization tools have been used in IEA both as representation and exploration tools to help users better evaluate the output of interactive evolutionary algorithms [HT00, LSA*06]. Despite efforts to design good user interfaces for IEA, human interaction with these systems usually raises several problems, mainly linked to the "user bottleneck" [PC97], human fatigue and slowness. Various solutions have been considered [PC97, Tak98, Ban97] such as reducing the population size (micro-EAs), constraining the search space to focus on *a priori* "interesting" areas, and deploying approximated user models (also called *surrogate functions*) to filter obvious bad solutions [LPLV05]. In the visualization community, work on parameter space exploration and optimisation relates to ours. Matkovic et al. [MGJH11], for instance, tried to interactively find an optimal combination of input parameters for a complex diesel engine injection system using visual analysis techniques. However, to our knowledge, we are the first to propose using IEA as optimisation tools to help navigate large search spaces.

3. EvoGraphDice

Since our main contribution in this work does not lie in a novel visualization system, but in enabling an IEA to guide user exploration, we used an existing visualization tool (GraphDice [EDF08, BCD*10]) to manage the various projections of the data. Views are organised in a scatterplot matrix (SPLOM) of 2D projections, Fig. 1(a). Users can do brushing and linking using a lasso tool. *EvoGraphDice* displays the dimensions proposed by the IEA as additional rows (and columns) in the SPLOM. The system initially displays dimensions returned by a PCA, after which the user can evolve new dimensions by pressing the "evolve" button, Fig. 1(d). The proposed views are displayed in yellow background; the darker the color the more interesting the view. The system provides an initial score (1 to 5) for each new view but the user can adapt this score using the slider in Fig. 1(d). User evaluated cells are flagged (small black square) to distinguish them from system evaluated

cells. *EvoGraphDice* can be initialised at any time using the "restart" button which resets parameters of the IEA. Users can save views (Fig. 1(f)) and bring them back into the SPLOM if they have been replaced during the exploration.

The current population is also displayed as a table (Fig. 1(h)) where each row corresponds to a combined dimension described by a mathematical expression and various components of the fitness function such as the scagnostics measures. The user can edit an individual using the "dimension editor" in Fig. 1(j), and limit the dimension search space Fig. 1(i), which results in a system reset similar to precessing the "restart" button. Note that many EA parameters can be tuned, such as the fitness threshold and crossover/mutation/replacement rates (see [CBL12]).

Our prototype has been developed from a first version [CBL12] based on an IEA that only manipulated linear combinations of dimensions. Our new extensions are: (i) a Genetic Programming (GP) algorithm allowing the manipulation of non-linear combinations of dimensions as variable size mathematical formula, (ii) user assessment of proposed views is explicitly captured via a slider, (iii) a surrogate function based on scagnostics measurements is used to predict and simplify the interactions of the user with the IEA, (iv) color highlighting of cells is used to draw user attention to the most interesting views.

Search Space: The space searched by the evolutionary process is the set of all dimensions that can be built by combining the initial dimensions with operators and constants, encoded as trees according to the Genetic Programming (GP) framework [Koz92]. These combinations can be complex mathematical expressions containing quadratic, exponential or logarithmic terms (evolved expressions can be any combination using $+, -, *, /, (.)^{(.)}, exp$ and *log* operators).

Genetic Engine: We have chosen to evolve a small set of combined dimensions, in order to let the user see all individuals of the population at a glance: if n is the number of initial dimensions, a population of another n combined dimensions is evolved. At each iteration, that is each time the user clicks on the "evolve" button, a new generation is produced by application of selection/crossover/mutation operators and then presented to the user whose judgment (evaluation) is explicitly collected via a slider.

Initialisation: A set of *a priori* interesting dimensions has been chosen as a starting point. A PCA analysis is performed [Smi02] on the original data and the corresponding n linear combinations form the initial population.

The fitness function, that is optimised by the genetic engine, is a sum of three terms:

1. **A surrogate function** f_{sc}, that plays the role of a predictor, and helps the system to better adapt to user needs. It is based on scagnostics measurements computed for every cell of each dimension y_i, the corresponding fitness term is a linear combination of the highest values of the scagnostics ($SC_k(y_i, x_j)$) of each scatterplot cell (y_i, x_j):

$$f_{sc}(y_i) = \sum_{k=1..9} w_k(\max_j SC_k(y_i, x_j)) \qquad (1)$$

The weights w_k that govern the relative importance of each scagnostic measurement are initialised to a uniform weight ($1/9$). Then, as soon as enough interactions are recorded (n, the number of variables), w_k are updated via a simple multilinear regression on the m past interactions ($m \geq n$ corresponds to the length of the "memory" of the system).

2. **A Complexity term** that favours dimensions made of a small number of variables and simple mathematical expressions :

$$f_c(y_i) = \left(1 - \frac{nvars(y_i)}{n}\right) \times \frac{1}{depth(y_i)}, \qquad (2)$$

$nvars(y_i)$ is the number of original variables involved in the mathematical expression of y_i, and $depth(y_i)$ is the depth of the GP tree representing y_i.

3. **A user evaluation term,** $f_u(y_i)$, that is an average of the user evaluation for each cell corresponding to y_i (range of 1 to 5 from "bad" to "excellent").

Diversity management : The evolutionary mechanisms naturally tend to concentrate the population around good solutions. So for small populations sizes, there is a risk of premature convergence if no diversity preservation mechanism exists. In *EvoGraphDice*, each time a new dimension y_i' is generated, its Euclidean distance to the current population is computed. If y_i' is too close to one of the individuals of the current population, it is replaced by a random individual.

4. Case Studies with Expert Users

We conducted an observational study with five domain experts. During the study sessions, we encouraged participants to think-aloud and share their findings with the study facilitator. We wrote observations, conducted semi-structured interviews and questionnaires, video-recorded the sessions and logged user interactions. The following sections describe the study setup, observations and findings for each expert.

4.1. Method

Due to the open-ended style of exploration using *Evo-GraphDice*, and the subjective nature of our fitness function, we chose a qualitative observational study methodology [Car08, MSM12, SMM12] that better suits our evaluation needs. We wanted to evaluate the usability and utility of our tool. In particular, we attempted to answer the following three questions: (i) is our tool understandable and can it be learnt; (ii) are experts able to confirm known insight in their data; and (iii) are experts able to evolve views that contain new insight or allow them to generate a new hypothesis, and if so how easy or difficult is it to reach those findings.

4.2. Participants and Apparatus

We evaluated our prototype with 5 domain expert users (2 female), ages $27 - 42$ (mean 34.2). Experts were academics and practitioners who had multidimensional datasets related to their domain of expertise (scientific simulation, medicine and geography) and were interested in further exploration. They consisted of one graduate student, four senior researchers and one medical surgeon. Participants had previously explored their datasets using graphical tools (e.g. Excel and JMP) or used statistical methods (PCA and regression analysis) but felt there is more to discover in their data than their current tools allowed them to. Experience with advanced multidimensional visualization tools varied from none, to experts who already used GraphDice or other SPLOM-based tools (two experts). None of our participants previously used dimension combination to analyse their data but three performed PCA-type analysis. The first three case studies ran at our research lab on an HP Z800 workstation PC with a $19''$ dual monitor (1280 x 1024 screen resolution). The last two case studies ran on a similar setup at the experts' institutions. Each session lasted on average 2.5 hours.

4.3. Tasks and Procedure

Participants were asked to carry out two main tasks: (**T1**) show in the tool what they already know about their data, hypothesis and questions they wanted answered; and (**T2**) explore their data in light of these hypotheses and research questions. The first task (and a training game) was designed to test if the tool is understandable, easy to learn, and can help experts rediscover known findings. The second open-ended one explores how domain experts use our tool to answer questions about their data and gain new insights.

Prior to the actual study, participants filled by e-mail a pre-questionnaire to elicit their background, knowledge about the dataset they want to explore, and experience with multidimensional data visualizations. In particular, they were asked to describe the dimensions of the data sets they provided, known relationships between variables, and hypotheses they wanted to investigate. The main study ran in two parts; first training then open exploration as follows:

Training: participants played a game designed to teach them how to operate the tool. A 5D dataset was synthesised with two enclosed curvilinear dependencies between two variables (x0 and x1) and random data for the rest of the dimensions. Participants were asked to evolve a scatterplot where it is possible to separate the two curves in Fig. 3 (left) with a straight line and were given around 20 minutes to complete the task (this task is equivalent to separating the two convex hulls in Fig. 3). Two participants successfully separated the two curves, while the remaining experts evolved views very close to a correct solution within the allocated time.

Open Exploration: the second part of the study ended after about one hour of exploration (a maximum limit of two

hours was set based on a pilot to avoid user fatigue), and participants were encouraged to take breaks. A facilitator was present to answer experts' questions and discuss their findings. Throughout the study, a second screen with an open text editor and pen and paper were provided to the experts as means of writing down their exploration findings. At the end, participants filled in a short questionnaire rating aspects of the tool (5-point Likert scales), such as the ease of performing the two main tasks, and open ended questions regarding their exploration strategy and helpful features of the tool.
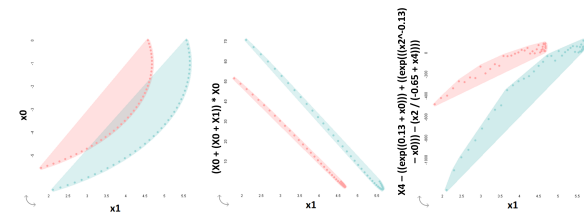

Figure 3: *Two different solutions (screenshots of plots) for the training game problem (left) that involve a simple dimension combination (middle) and a complex formula (right).*

4.4. Data Collection and Analysis

Participants were video-taped and log data of user interactions was gathered for further analysis (table 1). Live and video observations, the results of the questionnaire, and the log analysis are described separately for each case study.

4.5. Expert 1: Electrical Consumption Profiles

Dataset: (9D) describes the electrical consumption of 900 anonymised businesses during non-peak (*npk*) and non-plan (*npn*) hours ('plan' refers to an agreed unit rate for a defined period of time) for winter (*W*) and summer seasons (*S*), their geographical altitude and the total consumption cost.

Goal: the expert wanted to investigate electricity consumption patterns of these businesses and their impact on the total cost of consumption. The expert noted in the pre-questionnaire that he would like to sum-up some dimensions in twos in order to focus on one aspect of consumption (e.g. non-plan and non-peak for summer), and therefore had a clear motivation for combining dimensions. This, he argued, may allow him to see interesting consumption profiles.

Observations: the expert hypothesised that altitude has an influence on electricity consumption during both summer and winter seasons. He also had some prior knowledge about existing outliers in the data. During the study he was able to quickly verify both of these hypothesis.

The most important new finding made by the expert, which was not part of the original search space, is a view showing a linear combination of the four parameters of interest to the expert (*npn* and *npk* for summer and winter) which brought to evidence in a quantitative manner that *npnW* consumption is the more correlated to the total consumption.

Expert	G	T1	T2	Q	Data	Size	D	LimitSearch	Evolve	Eval	OVisits	NVisits	Insight
1	-	4	4	3	business	9x900	1:10	3	3	16	40	105	2(1)
2	-	4	4	3	timeseries	7x78	1:33	4	3	8	114	115	4(3)
3	9	5	5	3	geometrical	12x67	0:49	4	21	90	99	344	2(1)
4	7	5	4	3	statistical	10x200	2:23	7	13	83	110	309	6(1)
5	-	3	2	4	geospatial	11x653	1:27	5	5	20	64	229	-

Table 1: *Log data showing: (G) the generation when a solution for the game was found, (T1&T2) experts' scores for ease of completing tasks T1&T2 on a 5-point Likert scale, 5 signifies "very easy", (Q) score for user agreement with EvoGraphDice cell evaluations on 5-point Likert scale, 5 indicates strong agreement, (Data&Size) type and size of dataset, (D) duration (hh:mm) of T2, (LimitSearch) breath of exploration indicated by the number of times the expert limited the search space, (Evolve) depth of exploration indicated by the maximum reached generation, (Eval) how many new cells were evaluated by the user, (OVisits&NVisits) number of times the expert visited the original cells and the new cells respectively, and (Insight) number of times the expert limited the search space and the generation (between parenthesis) where the insight was found.*

In the user's own words: "we always talk about this qualitatively. This is the first time I see concrete weights ... To understand what is a better fare, it is necessary to find a good approximation of the consumption profile", like the one found in Fig. 4. Thus this insight can lead to electricity plans that are better suited to clients' consumption needs.
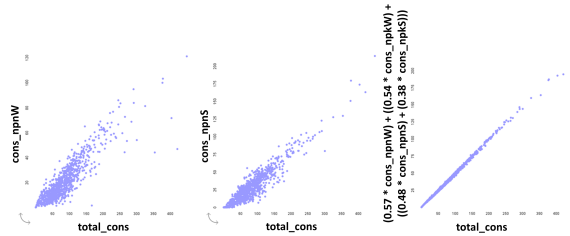

Figure 4: *Confirmed findings (left and centre) and new insight found by the expert (right): a linear combination of four parameters that approximates customer consumption.*

According to this participant, his exploration strategy was to look at propositions in detail along a row, e.g., to examine proposed dimensions plotted against the total consumption. Overall, the expert did not evolve many generations (depth of exploration was three generations at most), but used "limit the search space" facility three times, indicating that he was trying to formulate an interesting hypothesis more than he investigated one in depth. The solution he found was after limiting the search space for the second time.

The expert liked the ability to limit the search space and to enter formulae for the combined dimensions using the dimension editor, e.g. to invert a weight.

4.6. Expert 2: Biscuit Baking Process

Dataset: (7D) describes 78 data points recorded from several industrial biscuit training processes taken by experts in the industry. There are two input parameters relating to temperature settings and three output parameters relating to biscuits (weight loss, height and colour) and a timestep.

Goal: the expert wanted to visualise dependencies in the data between input and output parameters (intuitively such correlation should exist but its exact nature is not clear).

Observations: the expert was able to quickly verify known profiles in the data, for instance the influence of temperature on height and color of the biscuit.

The more general profile of the relationships between input and output parameters was not evident from the original dimensions, thus the expert looked at a wider space using combined dimensions. He observed that there might be some exponential factors that link outputs and inputs, in particular an exponential dimension of one of the input parameters (proposed by the GP) was linearly correlated between all output variables: "we would probably not have considered looking at exponential relationships" indicating a surprising finding and, thus, the ability of the tool to encourage lateral thinking. Further investigation showed that the exponential of temperatures has a specific meaning in thermally activated processes (explained by the Arrhenius law [Wik]).

To look for new relationships, the expert's strategy was to evolve a few generations and choose a visualization that showed linear or quadratic relationships. Like the first expert, he edited promising views using the "dimension editor" to see if this had better or worse impact on the relationship between variables. Most importantly he tried to reduce formula complexity to make better sense of the relationship.

The expert found the evolution functionality and the preview matrix useful to reach interesting cells. However, he wanted to examine scores per dimension as well as per views, e.g. to see if a dimension is always highly ranked; and for the tool to highlight new dimensions that have not been visited before.

4.7. Expert 3: Anatomical Planning for Surgery

Dataset: (12D) of which half describe anatomical and geometrical values related to a 3D planning of a surgical operation (total hip arthroplasty) for 67 patients, and the other half represent values for the same parameters after surgery.

Goal: the expert wanted to investigate whether there is a correlation between the planned values and the final values for each of the investigated parameters, and, if it exists, how strong is this correlation. Since there are many parameters to examine with potentially many interactions between them,

the expert focussed on examining the offsets for the cup anteversion parameter (AntvCupSupine) which corresponds to the orientation angle of the cup of the hip prosthesis.

Observations: the expert already knew that there is a relationship between the planned and real values for the AntvCupSupine parameter. This was easily verified in the original dimension space. To explore this further, the expert examined the view showing the before and after values, then made three lasso selections corresponding to over-fit, best-fit and under-fit values (Fig. 5 left), then examined brushed cells in the original search space.

In terms of new insight, the expert found a new cell where the two problematic groups (in red and blue) were separated from the well-restored group (in green) with the exception of one data point. The proposed dimension had a simple formula that involved two original dimensions. Views showing such separation may correspond to special geometrical settings or anatomical features for the observed patients. The expert noted that he needs to examine these patients more carefully with special attention to the selected parameters which can then lead to better pre and post-surgery results.

The expert followed the training game example as his exploration strategy, which may explain the big depth of exploration (21 generations): he made lasso selections of groups of data points, and evolved views that he scored highly depending on whether the overlap between the clusters is minimised. He examined the proposed dimensions in relation to the AntvCupSupine parameter and made use of the "favorite" facility to compare interesting cells that were replaced in the next generations. Notably, he evaluated more than 26% of new visited cells. This seemed to be an important part of his exploration strategy.
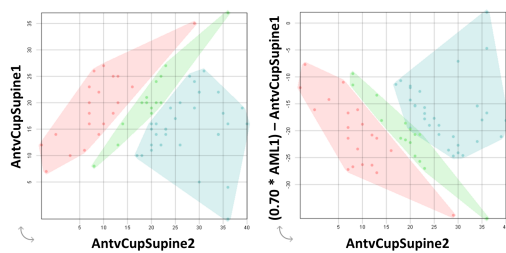

Figure 5: *Selections of over-fit (red), best-fit (green) and under-fit (blue) parameter values (left), and (right) a finding by the expert showing a separation between the two groups of interest in relation to a new parameter.*

The expert commented that he liked the direct visual interaction with the data but he did not like the uncertainty in whether a solution existed and whether the tool will find it. For example, he was interested to see if the degree of separation between his data groups became smaller between evolutions. He suggested adding more adapted tools for selecting data clusters and including statistical information.

4.8. Expert 4: Pareto Front Exploration

Dataset: (10D) describes the outpt of a genetic algorithm that was used to calibrate a city growth and emergence model. The data represents a set of parameter values (7 dimensions) and their objective fitness scores (3 dimensions). The explored dataset only includes the first best 200 parameter values that the algorithm found according to the three objectives of the calibration model (i.e. the Pareto front of the global parameter space exploration).

Goal: the expert wanted to explore the dataset from the two different perspectives (parameter and objective space) as well as the interaction between the two spaces, e.g. does a special profile in the parameter space correspond to a special profile in the objective space?

Observations: prior to the study, the expert had an idea about some characteristics of the data, e.g. there are two large clusters that can be differentiated by the value of one parameter (pAdoption). This type of calibration was also known to produce a characterisable response in the objective space. This hypothesis was easily verified using *Evo-GraphDice* via brushing and linking between cells in the parameter space and the objective space.
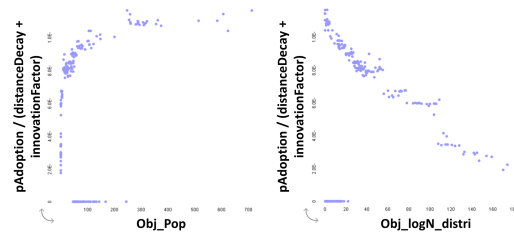

Figure 6: *An interesting combined dimension from the parameter space and its impact on two objective dimensions.*

In terms of new insight, the expert was able to find an interesting combined dimension that gave a good correlation for two parameters of the objective space (Fig. 6). The expert commented that this combination may be an important finding because it involves parameters that affect only one part of the simulation model. This indicates that those parameters, at least for these two output indicators, work together; and that this linear combination could be one way to reduce the complexity of the model.

As for strategies, the expert mentioned primarily limiting the search space (7 times), evolving (13 generations) and examining cells that had monotonic or striated distributions. She also made good use of visual queries and cell evaluation; 27% of new visited cells were evaluated by the participant.

The expert liked the ability to limit the search space and the freedom the tool offers to explore the data by "evolution". She added that the tool was helpful in reaching insight because of its ability to visualise and suggest combinations of dimensions that actually had a visual pattern: "in Excel it is difficult to find a formula that would give a nice pattern".

4.9. Expert 5: Urban Organisation and Perception

Dataset: (11D) describes geo-spatial information about 653 inhabitants and their urban environments including their profiles (e.g. gender, age), perception of their neighbourhood and objective variables describing their street such as the distance to the nearest metro station and type of district.

Goal: as with other experts, this participant was interested in finding relationships between variables; specifically she wanted to identify groups of inhabitants having similar profiles and to find the most discriminating variables for these individuals in order to make sense of the formed groups.

Observations: the expert had already found interesting correlations between the different original variables using her own statistical tools. She was able to confirm these findings. For instance, that an individual's perception of the size of their neighbourhood was dependent on the distance to the next metro station, but this was only true if context (type of district) was taken into account. However, the expert was aware of correlations requiring an interaction between two variables against a third but found it difficult to see them using *EvoGraphDice*. The expert noted two major difficulties that may have hindered the exploration and thus lack of early insight or hypothesis generation: (a) difficulty in determining the criteria for scoring patterns without knowing what a good pattern is in advance; and (b) the nature of data about human behaviour and perception has high variability, thus examining averages, for instance, is more appropriate.

Despite the aforementioned difficulties, the expert found two interesting views where clusters and outlier groups seem to correspond to a known profile (Fig. 7). However, she was not able to fully interpret the proposed combined dimensions as the choice of variables made sense, but the overall interpretation of the pattern was not clear to the participant.
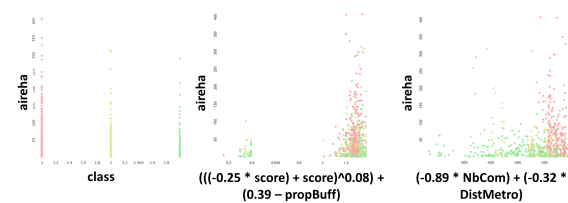

Figure 7: *Two interesting combined dimensions (centre and right) found by the system and their impact on one objective dimension (aireha). Brushing and linking to an original view (left) shows interesting profiles.*

This expert's exploration strategy was to limit the search space to $3 - 4$ variables and examine their interaction with one original dimension (e.g. perception of space). She also made selections and examined the brushed views in the original space. Since the expert did not evolve many generation (5 max) and only evaluated a few cells (20 overall)–due to the aforementioned difficulties– the system did not learn well the type of distributions the expert was looking for.

The expert tended to agree with the system's proposed scores (Table 1 Q), which she found interesting because of the choice of variables and the simplicity of the proposed formulae. As interpretation of results was difficult using the current point-based presentation, the expert noted that showing aggregated values and variance would help her better understand the views.

5. Summary of Results

Almost all participants were able to easily confirm prior knowledge about their datasets (2 x 'very easy', 2 x 'easy', 1 x 'neutral'). One expert found this task challenging because of the lack of data aggregation that her type of analysis requires. Overall, participants confirmed known correlation, clusters or outliers in their data. In the remainder of this section, we summarise our study findings concerning new found insight, successful tasks and exploration strategies.

5.1. Insight Generation and Tasks

If we include hypothesis formation as part of insight generation, similar to work by Saraiya et al. [SND05], *EvoGraphDice* helped our participants generate new insight in the form of distinct observations about the data (4 experts), new hypothesis (1 experts) and better formulation of research questions (4 experts). Distinct observations found by the experts were either clustering, linear or non-linear relationships, and similarly to generated hypotheses, they always linked a dimension in the original data set and a new proposed dimension. The subjective evaluation of ease of task T2 (table 1) shows most experts found it easy to find new insight: 1 x 'very easy', 3 x 'easy' and 1 x 'not easy'. Not surprisingly, those who reached a concrete new finding scored the tool highly in comparison to those who did not.

The found solutions were regarded by the experts as interesting because they had one or more of the following properties: (i) a visual pattern such as those modeled by the scagnostics measures; (ii) a simple formula involving few dimensions; (iii) a selective choice of dimensions (corresponding to an unformulated hypothesis or an inherent aspect of their data model); and (iv) a domain value. Regarding the latter point, not all participants were able to state the immediate domain value but in general, our participants stated that *EvoGraphDice* helped them:

- interact visually with data (experts 3)
- try out alternative scenarios by editing dimensions (experts 1,2)
- think laterally (expert 2)
- quantify a qualitative hypothesis (expert 1)
- formulate a new hypothesis or refine an existing one (1-4)

5.2. Exploration Strategies

Overall, participants followed the same exploration pattern consisting of first examining the original dimensions then inspecting and evaluating the first generation of the proposed dimensions (returned by the PCA) followed by one

or more iterations of the following steps: (i) limit the search space; (ii) select and rank cells; (iii) evolve; and (iv) interpret and verify. However, the frequency of using some tools (e.g. "evolve" vs. "limit the search space") varied depending on whether the expert had an *a priori* focused hypothesis (i.e. a research question involving typically $3-4$ dimensions). We observed that the looser the initial hypothesis, the more often they tried to change the search space; and the more focused the hypothesis the more generations they inspected. Indeed, these two strategies of *exploration* and *exploitation* are supported by EAs [Ban97] where on the one hand the user wants to visit new regions of the search space and on the other hand they want to explore solutions (combined dimensions) close to one region of the search space.

6. Discussion

Most of our experts were able to formulate interesting hypothesis or reach new insight requiring looking at data in terms of a combination of dimensions. Our approach consists of proposing new views based on automatically calculated metrics and user feedback. On the one hand, our method is complimentary to PCA, clustering and regression analysis that automatically find data patterns, and optimise a fit. On the other hand, we allow users to interactively select examples of visual pattern types they are interested in, and that may not be easy to express mathematically. Users can then verify the new relationships they find in *EvoGraphDice* by using the dedicated automatic data analysis tools.

In comparison to automatic analysis such as in statistics and data mining, our approach offers: (i) **Intuitiveness:** a visual approach to interact with data requiring no prior statistical knowledge; (ii) **Interactivity:** rather than fitting the data to pre-defined shapes in a static manner, using an IEA the user can dynamically steer the exploration process towards a pattern of interest. These patterns can involve dimension concatenations that are not obvious at the outset of the exploration; (iii) **Flexibility:** ability to edit and try out alternative dimension combination scenarios, or limit the search space. (iv) and **Adaptability:** the system can adjust to user change of interest over time.

There are limitations to using our tool, such as the types of datasets to explore and issues related to the interpretation of combined dimensions which we discuss below. Moreover, the issue of the convergence of the genetic algorithm is an interesting one given that the IEA deals with optimisation. However, this is not easy to study in our case as there is no unique solution to converge to, rather the optimisation is dynamically adapted to follow user interest over time.

First, we are constrained by the SPLOM representation of *EvoGraphDice* which does not provide a natural way to interact with some dataset types such as timeseries. Data with high variability provides additional challenges that we do not currently address, such as detecting and evolving aggregated patterns. In addition, we tested our prototype with user-provided datasets that are small to medium sized, having dimensions between $7-12$. Although our algorithm can deal with a large number of data points, it may not handle well larger number of dimensions as complex combined dimensions may be difficult to avoid. In this case, a dimension reduction technique can be applied to the dataset before feeding the results to *EvoGraphDice*.

Second, not all variables can be combined, therefore the user should as soon as possible limit the search space to "combinable" dimensions. This in a sense requires the user to have some domain knowledge and to make an initial hypothesis about the data. The proposed dimensions can involve complex or unforeseen combinations yielding a visual pattern but one that can be difficult to interpret. To help address this issue, we used "complexity" of a combined dimension as a component of the IEA fitness function. Nonetheless, our method can still yield complex dimensions that are difficult to interpret. We noticed that our participants only looked at the combined dimensions in relation to an original dimension, most likely to ground the observation. This problem of interpretation, however, is common to all tools that offer dimension combination.

7. Conclusion and Future Work

We presented a prototype tool (*EvoGraphDice*) for supporting Evolutionary Visual Exploration (EVE) that combines visual analysis with interactive evolutionary computation to help steer the exploration towards interesting views on the data. Our method complements PCA, clustering and regression types of analysis, offering additional features such as interactivity and adaptability. We conducted an observational study with domain experts and found that our tool allowed users to evolve characteristics that are not visible in the original dimensions space. Our experts were able to try out different scenarios, think laterally, quantify qualitative hypotheses and formulate new ones.

Future work for our tool includes longitudinal user studies to explore in detail the long term evolution of user focus, as well as addressing issues such as improving the IEA to detect more complex visual patterns (beyond those currently detected by Scagnostics), handling data with high-dimensionality and bridging *EvoGraphDice* and existing statistical packages to combine powerful statistical analysis with flexible and intuitive visual exploration.

Our work demonstrated that tightly combining visualization and optimisation techniques can yield exciting results in data analysis and opens new venues for research, but also highlights challenges such as monitoring algorithm convergence, history visualization of diverging exploration paths, and appropriate methodologies for evaluation.

References

[Asi85] ASIMOV D.: The grand tour: a tool for viewing multi-dimensional data. *SIAM J. Sci. Stat. Comput. 6*, 1 (Jan. 1985), 128–143. 1

[Ban97] BANZHAF W.: *Handbook of Evolutionary Computation*. Oxford University Press, 1997, ch. Interactive Evolution. 2, 3, 9

[BCD*10] BEZERIANOS A., CHEVALIER F., DRAGICEVIC P., ELMQVIST N., FEKETE J.-D.: GraphDice: A System for Exploring Multivariate Social Networks. *Computer Graphics Forum (Proc. EuroVis 2010) 29*, 3 (2010), 863–872. 3

[BTK11] BERTINI E., TATU A., KEIM D.: Quality metrics in high-dimensional data visualization: An overview and systematization. *IEEE Transactions on Visualization and Computer Graphics 17*, 12 (Dec. 2011), 2203–2212. 1, 3

[Car08] CARPENDALE S.: Evaluating information visualizations. In *Information Visualization*, Kerren A., Stasko J. T., Fekete J.-D., North C., (Eds.). Springer-Verlag, Berlin, Heidelberg, 2008, pp. 19–45. 4

[CBL12] CANCINO W., BOUKHELIFA N., LUTTON E.: Evographdice: Interactive evolution for visual analytics. In *IEEE Congress on Evolutionary Computation, June 10-15* (2012). June 10-15, Brisbane, Australia. 4

[EDF08] ELMQVIST N., DRAGICEVIC P., FEKETE J.-D.: Rolling the Dice: Multidimensional Visual Exploration using Scatterplot Matrix Navigation. *IEEE Transactions on Visualization and Computer Graphics (Proc. InfoVis 2008) 14*, 6 (2008), 1141–1148. 3

[Evo] EVOGRAPHDICE:. http://www.aviz.fr/Research/EVE. Accessed: 10/03/2013. 2

[Fri87] FRIEDMAN J. H.: Exploratory projection pursuit. *J. Am. Statistical Assoc 82*, 397 (1987), 249–266. 1

[Gol89] GOLDBERG D. E.: *Genetic Algorithms in Search, Optimization and Machine Learning*, 1st ed. Addison-Wesley Longman Publishing Co., Inc., Boston, MA, USA, 1989. 1

[HT00] HAYASHIDA N., TAKAGI H.: Visualized IEC: interactive evolutionary computation with multidimensional data visualization. In *IECON 2000. 26th Annual Conference of the IEEE* (2000), vol. 4, pp. 2738–2743. 3

[Hub85] HUBER P. J.: Projection Pursuit. *The Annals of Statistics 13*, 2 (1985), 435–475. 1

[JB05] JIN Y., BRANKE J.: Evolutionary optimization in uncertain environments-a survey. *IEEE Trans. Evolutionary Computation 9*, 3 (2005), 303–317. 2

[Koz92] KOZA J. R.: *Genetic Programming* . MIT Press, 1992. 4

[LPLV05] LUTTON E., PILZ M., LÉVY VÉHEL J.: The fitness map scheme. application to interactive multifractal image denoising. In *CEC2005* (Edinburgh, UK, September, 2-5 2005), IEEE Congress on Evolutionary Computation. 3

[LSA*06] LLORÀ X., SASTRY K., ALÍAS F., GOLDBERG D. E., WELGE M.: Analyzing active interactive genetic algorithms using visual analytics. In *Proceedings of the 8th annual conference on Genetic and evolutionary computation* (New York, NY, USA, 2006), GECCO '06, ACM, pp. 1417–1418. 3

[Lut06] LUTTON E.: Evolution of fractal shapes for artists and designers. *IJAIT, International Journal of Artificial Intelligence Tools 15*, 4 (2006), 651–672. Special Issue on AI in Music and Art. 2

[MGJH11] MATKOVIC K., GRACANIN D., JELOVIC M., HAUSER H.: Interactive visual analysis supporting design, tuning, and optimization of diesel engine injection. *Proceedings of IEEE Visualization 2011 (Discovery Exhibition)* (2011). 3

[MSM12] MEYER M., SEDLMAIR M., MUNZNER T.: The Four-Level Nested Model Revisited: Blocks and Guidelines. In *Proceedings of the VisWeek Workshop Beyond Time and Errors: Novel Evaluation Methods for Information Visualization (BELIV)* (2012), ACM Press. 4

[PC97] POLI R., CAGNONI S.: Genetic programming with user-driven selection: Experiments on the evolution of algorithms for image enhancement. In *Genetic Programming 1997: Proceedings of the Second Annual Conference* (1997), Morgan Kaufmann, pp. 269–277. 3

[PLM08] POLI R., LANGDON W. B., MCPHEE N. F.: *A field guide to genetic programming*. Published via http://lulu.com and freely available at http://www.gp-field-guide.org.uk, 2008. (With contributions by J. R. Koza). 2

[PWR04] PENG W., WARD M. O., RUNDENSTEINER E. A.: Clutter reduction in multi-dimensional data visualization using dimension reordering. In *Proceedings of the IEEE Symposium on Information Visualization* (Washington, DC, USA, 2004), IN-FOVIS '04, IEEE Computer Society, pp. 89–96. 3

[Smi02] SMITH I.: A tutorial on principal component analysis, 2002. 4

[SMM12] SEDLMAIR M., MEYER M., MUNZNER T.: Design Study Methodology: Reflections from the Trenches and the Stacks. *IEEE Trans. Visualization and Computer Graphics (Proc. InfoVis) 18*, 12 (2012), 2431–2440. 4

[SND05] SARAIYA P., NORTH C., DUCA K.: An insight-based methodology for evaluating bioinformatics visualizations. *IEEE Transactions on Visualization and Computer Graphics 11*, 4 (July 2005), 443–456. 8

[SS05] SEO J., SHNEIDERMAN B.: Rank-by-feature framework for interactive exploration of multidimensional data. *Information Visualization 4*, 2 (2005), 99–113. 3

[Tak98] TAKAGI H.: Interactive Evolutionary Computation: System Optimization Based on Human Subjective Evaluation. *INES'98* (1998). 3

[Tak08] TAKAGI H.: New topics from recent interactive evolutionary computation researches. In *Knowledge-Based Intelligent Information and Engineering Systems* (2008), p. 14. 2

[Wik] WIKIPEDIA: Arrhenius equation. http://en.wikipedia.org/wikiArrhenius_equation. Accessed: 10/03/2013. 6

[WW08] WILKINSON L., WILLS G.: Scagnostics Distributions. *Journal of Computational and Graphical Statistics 17*, 2 (2008), 473–491. 3