MATHEMATICAL CLUSTERING INTEGRATED WITH SWOT ANALYSIS AS A TOOL FOR DESIGN OF SUSTAINABLE DEVELOPMENT STRATEGIES

L. NONDEK & M. SMUTNÝ
Integra Consulting Services Ltd., Prague, Czech Republic.

ABSTRACT

The paper describes planning procedures used originally for drafting the Sustainable Development Strategic Framework (SDSF) of the Czech Republic in 2008–2009 and compares them with the use of a self-organizing map (SOM), combined with the Ward's clustering method. Clustering followed by a series of SWOT analyses is a new technique that allows for the development of large-scope strategic documents covering many sectors and accommodating various economic, social, and environmental issues. Large initial sets of statements (problems, trends, forecasts, etc.) can be converted by multiple SWOT analyses into a consistent set of interventions. An optimal structure of the clustered statements has to be found experimentally. Use of SOM makes searching for the optimal structure (information model of the strategy) efficient. Such information model can be broadly discussed by stakeholders and purposefully modified (generation of strategic alternatives) before the best alternative is transformed into a final strategy text.
Keywords: cluster analysis, strategic planning, sustainability, SWOT.

1 INTRODUCTION

Sustainable development is a general concept integrating social, economic, and environmental dimensions [1, 2]. Discussions on the subject often take place at a general level, being based upon various definitions of sustainability [3]. If preparation of strategic documents is carried out without a strong relationship to the national or regional economic, social, environmental, and political context, then the practical value of such exercises is limited. This can be documented by example: the National Strategy of Sustainable Development of the Czech Republic which was drafted by an academic team in 2007 has not been accepted by the Czech Government. It was, therefore, decided by the Governmental Council for Sustainable Development late in 2007 to modify the draft document in order to accommodate key problems and reflect the Sustainable Development Strategy of the European Union [4].

Concrete approaches to sustainability proposed by Faber *et al*. [5] were adopted by the consultancy team assigned by the Governmental Council for Sustainable Development to coordinate elaboration of a modified Sustainable Development Strategy of the Czech Republic (adopted as the Sustainable Development Strategic Framework (SDSF) in 2010) as described further in this article. The objectives, priorities, and measures toward sustainable development had to be defined according to the given planning time horizon (2020), the national context (institutional environment, resources, societal processes, implementing organizations, etc.), and the interests of stakeholders as influenced likely by the planned interventions (e.g. Czech Republic, political representation, economic operators, NGOs, etc.).

There are two basic practical approaches to strategy making: procedural and interactive [6]. In the case of the Sustainable Development Strategy of the Czech Republic, an inter-ministerial working group (WG) was established which included representatives of all ministries as well as the Office of the Government. The Consultant combined a procedural approach (rules and procedures adopted by the WG) with an interactive one (openness, creativity). The WG's decision-making was facilitated through the use of the AHP (analytic hierarchy process) technique introduced by Saaty [7] in a modification for groups [8].

This paper describes the procedure, which was used previously by Nondek and Kosíková-Šulcová [9] for the purpose of strategic environmental assessment (SEA). The authors have adopted and further modified a planning and assessment approach proposed by Trochim [10], which is based on generation and structuring of statements into a conceptual map used as a planning or assessment tool. The visualization of the ordered set of statements can be achieved by multidimensional scaling (MDS) as proposed originally by Trochim and Linton [13] and Trochim [11]. Instead of MDS, Nondek and Kosíková-Šulcová [9] used fuzzy clustering, or Ward's clustering, in combination with SWOT analysis, which was applied to transform the structured statements into planning interventions [12]. Large sets of statements (100 or more) require application of multiple SWOT analysis to the clusters of statements. This approach makes it possible to identify critical relationships (synergic or antagonistic) between the outcomes of partial SWOT analyses. Another reason for the use of SWOT analysis is that the method has been recommended by the European Commission as a part of preparation of regional and sectoral operational programs. The above approach had been tested and modified by the authors in several regional projects and recently used in drafting of the SDSF of the Czech Republic in 2008–2009 [14].

This paper presents lessons learned through the development of the Czech SDSF. A major obstacle was difficulties that the WG had in understanding the structures of clustered statements in high-dimensional space. We, therefore, tested the use of self-organizing maps (SOM) combined with Ward's clustering method. SOM is an unsupervised neural network introduced by Kohonen [15], which enables the presentation and visualization of high-dimensional data as two-dimensional (2-D) maps. This makes the ordering of statements (high-dimensional data) via clustering more understandable and easier to use as a base for planning as compared with the originally used approach by Nondek and Kosíková-Šulcová [9]. The testing of SOM applicability was carried out as an ex-post experiment with the original set of statements used during preparation of SDSF in 2008–2009.

2 PLANNING PROCEDURE

The planning process was supervised by the Governmental Council for Sustainable Development – the Managing Committee was established for this purpose within the Council. High officials (usually deputy ministers) were involved in the Managing Committee; its role was to review and control the overall approach to preparation of the strategy and to endorse draft outputs before submitting them for discussion within the Governmental Council for Sustainable Development, or to public consultation.

The strategy content was developed by the WG, coordinated by the Ministry of Environment in close cooperation with the consultants, and involving representatives of all ministries and the Office of Government (i.e. usually about 15 persons participated at the meetings). The group worked by regular meetings, where background materials were discussed (usually after they had been provided for comments well before the WG meeting) and their content was modified according to discussions. In later stages (after the draft strategic framework was prepared), small 'micro-teams' were established within the WG in line with priority axes (i.e. each micro-team was responsible for elaboration of one priority axis). The composition of micro-teams reflected interests and responsibilities of ministries (e.g. the Ministry of Industry and Trade participated in all micro-teams, dealing with energy issues, business, innovations, etc.). The whole process involved intensive e-communication among WG members.

The process' milestones can be summarized as follows:

1. May 2008: Initiation of the process, establishment of the Managing Committee, establishing WG, methodology development, draft vision, and reference framework.
2. June 2008: public consultations on the draft vision and reference framework.
3. July 2008: situation analysis and list of statements for SWOT analyses.
4. September 2008: setting up an information model (clustering) and conducting SWOT analyses.
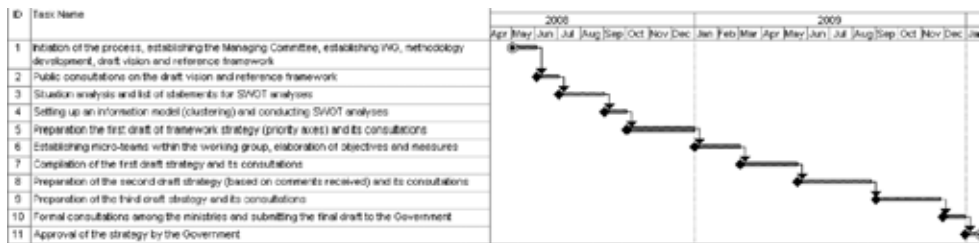
Figure 1: Overview of the strategy preparation timeline.

5. October–December 2008: preparation of the first draft of the framework strategy (priority axes) and its consultations.
6. January–February 2009: establishing micro-teams within the WG, elaboration of objectives and measures.
7. March–May 2009: compilation of the first draft strategy and consultations.
8. May–June 2009: preparation of the second draft strategy (based on comments received) and consultations.
9. September–December 2009: preparation of the third draft strategy and consultations.
10. November 2009: formal consultations among the ministries and submission of the final draft to the Government.
11. January 2010: approval of the strategy by the Government.

Figure 1 below provides an overview of the strategy preparation timeline. It illustrates the fact that the first 6 months of the process (May–October 2008) were dedicated to analytical works needed for preparation of the first draft of the strategy.

In order to involve all relevant stakeholders in the preparation of the strategy, a 'Platform for sustainability' was established at the beginning of the planning process. Various institutions, organizations (both governmental – e.g. regional authorities, and non-governmental, universities, research institutes, etc.), MPs and parliamentary committees, political parties, etc. were invited to participate in the Platform. The aim was to create a wide-open structure, where all important issues relevant to the strategy could be discussed. Two meetings of the Platform were organized during the strategy preparation – the first one was focused on key issues and problems which should be included in the strategy, the second meeting was dedicated to discussion on the draft strategy. The Platform members were regularly informed about progress, and draft outputs were sent to them with requests for comments.

The overall process took over one-and-a-half years and required about 150 working days from the external consultancy team, and approximately the same amount of time from WG members. The time necessary for carrying out the activities described here (reference framework, structuring statements (information model), and carrying out the SWOT analyses) was approximately 20 working days, including discussions and approvals by the WG.

The scope of the SDSF was defined by means of a reference framework at the national level. This framework was created in the first stage of planning when the WG proposed a strategic vision and subsequently derived a set of national sustainability themes. After this inception stage, the planning procedure continued according to the following steps:

1. Formulation of the draft vision (max. 150 words), and derivation of a set of sustainability themes (T_1–T_n, n = 3–8).
2. Analysis of the relevant planning and analytical documents (including previous development strategy, background papers, sectoral documents, etc.) with respect to the above

sustainability themes, extraction of the key statements related to sustainable development (list of statements).

3. Categorization (ranking) of statements using the reference system formed by T_1–T_n, cluster analysis of statements (e.g. fuzzy clustering).
4. Carrying out consistency assessment of clusters, and revising the ranking of inconsistent statements followed by new cluster analysis.
5. Conversion of the clusters into SWOT tables and carrying out a series of partial SWOT analyses.
6. Assessment of the whole structure created by the clusters and their SWOT analyses, definition of the strategic axes (global objectives) and checking whether the structure supports synergies and enables conflict solving.
7. Modification of the strategic vision, adaptation of the T_1–T_n (e.g. via expert AHP weighting) is carried out and steps 3–6 are repeated to obtain variants of the strategy. Specific analytical studies are prepared accordingly.
8. Selection of the best strategic variant by AHP. T_1–T_n previously defined can be used as criteria for AHP pair comparison of the variants generated.
9. Completing the selected variant in a full-text format written in non-expert language.

The procedure described above first creates an information model of the strategy (steps 1–6), which is optimized (steps 7 and 8) and finally converted into a full text of the strategic document (step 9).

2.1 Sustainability reference framework

The themes of sustainability derived from the strategic vision and finally adopted by the WG are listed below. The themes were specified by bracketed sets of key words:

T_1 – **Economic development in harmony with the environment** (industrial emissions, dematerialization of the economy, services, knowledge-based economy, decoupling, etc.).

T_2 – **Rational consumer and producer** (responsibility of producers, consumption patterns, reduced environmental pressures, waste management, recycling, etc.).

T_3 – **Sustainable mobility** (energy intensity of transport, logistic systems, e-society, etc.).

T_4 – **Balanced spatial and regional development, sustainable landscape** (landscape functions, regional development, spatial planning, biodiversity, climate change, etc.).

T_5 – **Healthy citizen, healthy society** (public health, aging population, open and coherent society, social inequality and exclusion, pathological social behavior, etc.).

T_6 – **Flexible society of knowledge and education** (education, research, and development, environmental awareness, brain drain, e-commuting, etc.).

T_7 – **Effective government** (balanced budget, competent administrators, transparency, public administration, good governance, stable legal environment, corruption, etc.).

T_8 – **Responsible partnership** (international cooperation, stable and prosperous Europe, good relationships with neighbors, development aid, etc.).

Relevant documents as selected by the WG, i.e. existing national strategies, plans and important background papers were summarized in order to prepare the set of the statements describing key critical problems (step 2). This activity was carried out by the consultancy team. After revision of the list by the WG, the consultancy team ranked the strength of relations (relevance) between statements

and themes (T_1–T_8) derived from the strategic vision. The strength (relevance) was ranked using the following scale:

0 = no relevance
1 = indirect or weak relationship
2 = direct relationship, and
3 = strong and direct relationship.

In addition, rankings 0.5, 1.5, and 2.5 were allowed. No difference was made between positive or negative relationships. The ranking was carried out independently by three experts of the consultancy team, and median values were then taken for clustering. The individual statements were in this way positioned into eight-dimension space with coordinates T_1–T_8. The Spearman correlation coefficients between pair rankings were calculated to detect positive correlations which may weaken the clustering. Correlations were found between pairs T_1–T_3, T_1–T_4, T_2–T_3, T_2–T_6, T_2–T_8, and T_5–T_7 (ρ_{crit} = 0.3) and the clustering (fuzzy method) experiments were alternatively carried out using T_1, T_2, T_5, T_8 or T_3, T_4, $T_5(T_7)$, T_8 rankings.

2.2 Cluster analysis

Cluster analysis was applied to identify aggregates of similar statements and to review their content. The software NCSS 2007 (power analysis software, version 07.1.21) [16] was used for the original clustering experiments (2008–2009), when a list of 170 statements was generated by the WG. It is important to stress that SWOT analysis was originally designed for business companies and it does not manage more than 12–16 items analyzed in one SWOT table (Weihrich [12]). In the case of SDSF, the list included 170 items to be dealt with (after preliminary aggregation of similar statements and deletion of several items with very low overall relevance to T_1–T_8). In such case, it was necessary to work with about 20 or more clusters, thereby enabling a series of internally consistent SWOT analyses.

By use of fuzzy clustering, several structures having 20–44 clusters were found as being suitable when Dunn's index was used as a validity criterion (Kaufmann and Rousseeuw [17]). Statements with low membership in the fuzzy clusters were checked and eventually shifted (re-ranked) into the nearest clusters. At this stage the following difficulties were identified:

a] Presentation and explanation of the structures created by fuzzy clustering (numeric results available only) to the WG.
b] Laborious preparation of strategic variants.

As an alternative to fuzzy clustering, Ward's clustering method (again using the software NCSS) was tested and a structure with 27 clusters was empirically chosen by means of a dendrogram used for visualization of the result. This structure was adopted by the WG as a starting point for defining the contents of the final document. Strategic variants were developed by modifying the set of T_1–T_8 (by elimination of correlating T's) and repeating steps 5 and 6.

2.3 Kohonen's self-organizing maps

As mentioned above, the consultancy team found that the visual representation by dendrograms is not sufficiently understandable for non-specialists (WG members), which might present a

serious problem in practical application of the above approach in strategic planning. It is, however, possible to make the above clustering procedures and their visualization more efficient and transparent by means of SOM, designed as a special category of neural networks by Kohonen [15]. SOMs are useful for visualizing high-dimensional data in the form of 2-D maps. In an ex-post experiment with the set of statements used originally for drafting the SDSF, the commercial software Viscovery SOMine, version 5.2, [18] was tested.

A set of comparative ex-post experiments was carried out with the originally developed Model 1 (T_1–T_8 with equal weights) using the SOM-Ward method. Three large segments, corresponding to the three main parts of the dendrogram obtained originally by NCSS, were initially formed (Model 1a), see Table 1. The content of segments was characterized using their mean T_1–T_8 coordinates, which indicate their relevance to T_1–T_8.

It is evident that the segments are still heterogeneous and their size does not allow SWOT analysis. Therefore, using results of the original Ward's clustering, the structure of three segments was broken into 27 clusters (Fig. 2). The sizes of the 27 clusters are between 2 and 12 statements per cluster (six items is a median).

An alternative Model 2 was generated, where the correlations between rankings were eliminated and the clustering was therefore carried out accordingly T_3, T_4, T_7, T_8 yielding 4 segments. SOM of Model 2 with 21 clusters is shown in Fig. 3. Purposeful modification of T_1–T_n is therefore a convenient way to easily generate variants of the basic structure.

Table 1: Prevailing contents, size (number of statements), and T_1–T_8 coordinates of segments, Model 1a.

Segment	Contents	Size	T_1	T_2	T_3	T_4	T_5	T_6	T_7	T_8
S_1	Economy and globalization	83	2.41	2.66	2.41	1.86	1.21	1.99	2.54	2.48
S_2	Economy, environment, and regional development	51	2.55	0.78	1.41	2.31	1.14	0.78	1.94	1.02
S_3	Social stability	36	0.56	0.53	0.69	0.31	3.00	2.33	2.89	1.33

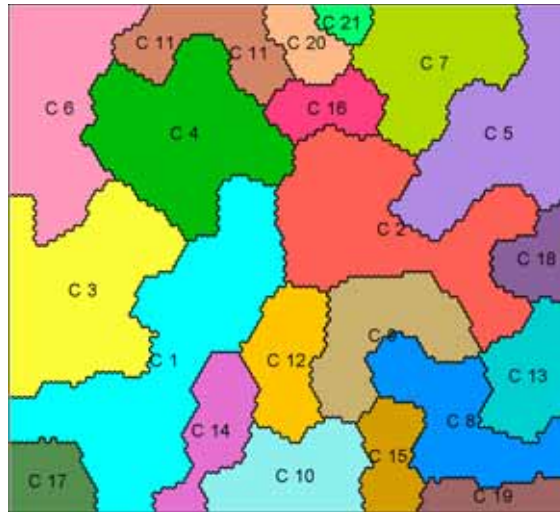Figure 2: Kohonen's map of 27 clusters, SOM-Ward method (Model 1a).

Figure 3: Kohonen's map of 21 clusters, SOM-Ward method (Model 2).

2.4 Transformation of clusters into SWOT analysis

In the first step, statements in each cluster were reformulated as strengths (S), weaknesses (W), opportunities (O), and threats (T). As an example, cluster C1 (Model 2) is converted into SWOT. Cluster C1 is obviously composed of two unrelated groups of statements (see Table 2) C1/1 and C1/2.

The C1/1 sub-cluster is related to transport (economy, emissions). By merging some statements, the following SWOT table was developed (Table 3).

Finally, the following **interventions** were proposed, being based upon selected combinations of S, W, O, and T:

- Investment into infrastructure via EU Funds (W_1, W_2, W_3-T_1).
- Railway reconstruction (S_1-T_1).
- Well-designed main roads and highways, application of strict EIA and land use planning procedures (S_1-T_1, W_3).
- Economic instruments to limit road transport, e.g. taxes, charges (W_1, W_2, W_3, T_1-O_2, O_3).
- Fleet improvement (emissions, fuel consumption, natural gas in municipal transport), technical checks of fleet, taxes, charges, subsidies (T_1, W_3-O_1).

Quantifiable objectives can be derived from the interventions, e.g. decrease of transport emissions, decrease of energy consumption in the transport sector, new highways and railroads, etc. There are likely conflicts with neighboring clusters C3 and C4 (see Fig. 3), e.g. landscape fragmentation, ecosystem protection, dissemination of invasive species and urban sprawl. Synergies are detected with cluster C2, e.g. better functioning of rescue systems and mobility of workforce. Moreover cluster C16, which consists of only two statements 'Growing transport increases road accidents' and 'Jammed highways make impossible rescue operations', can be merged with C2 thereby strengthening the synergies between investment, road safety, and better functioning of rapid rescue systems. In this way, the whole SOM can be converted into a single set of interventions and objectives (basic structure of the strategic document). SOM therefore plays a role of information model of the strategy.

Table 2: List of statements in C1 and proposed SWOT categories.

Sub-cluster	Statement	SWOT category
C1/1	Frequent conflicts between nature protection and economic interests	W
C1/1	Dense transport infrastructure (road, railway)	S
C1/1	Obsolete transport infrastructure (road, railway)	W
C1/1	Protection of climate causes pressure to transport emissions	O
C1/1	Negative impacts of growing cargo transport (incl. transit), emissions, noise, accidents	T
C1/1	Growing transport emissions incl. CO_2	T
C1/1	Binding emissions ceilings (SO_2, NO_x, VOC, NH_3)	O
C1/1	Post-Kyoto process obligations	O
C1/1	Insufficient traffic inter-connection between regional cities (centers)	S
C1/1	Dense network of communication to regional centers	S
C1/1	Roads in regional centers in bad condition, no by-passes	S
C1/1	Capital Prague is the engine of economic growth	S
C1/1	Impaired environment in Prague, weak transport infrastructure	W
C1/2	Prevention of wastes, reduction of wastes per unit of GDP	O
C1/1, C1/2	Dependence on imports of raw materials	O
C1/1, C1/2	High potential for economy dematerialization	O
C1/2	Further grow of energy consumption and its necessary coverage	T
C1/2	Large energy consumption per GDP	W
C1/2	Increase of the material efficiency	O

Table 3: Example of a SWOT table created from sub-cluster C1/1.

S_1: Dense network (road, railway)	W_1: Missing infrastructure
S_2: Binding emission ceilings	W_2: Obsolete infrastructure
	W_3: Conflict with nature and landscape protection
O_1: post-Kyoto process	T_1: Growing trend of road transport
O_2: Modal shift from road to railway	
O_3: Dematerialization of economy	

3 RESULTS AND DISCUSSION

Our previous work [9] used the above-mentioned approach developed by Trochim [10] as a starting point. We would now like to compare the characteristic features of concept mapping with the presented information model of the strategic document. Both methods have strengths and weaknesses. Concept mapping is suitable for more informal WGs, a narrower spectrum of issues addressed by the plan, and less time required for the planning exercise. Both methods are based upon structured statements (bottom-up approach, open planning) and no obligatory structure or content of the document is set at the beginning. There are, however, several substantial differences between both procedures (Table 4).

Table 4: Comparison of concept map with information model of strategic plan.

Concept mapping (Trochim [10])	Information model	Remarks
Statements are generated in brain-storming sessions	Verified statements are used only, e.g. excerpted from documents, studies, forecasts, etc. Subjective opinion is not included	Brainstorming saves time but leads to subjectively biased statements and important problems might be overlooked. The information model is based on verified information
List of statements related to single institutions or sectors	List of statements related to several sectors and to international contexts (e.g. EU, multilateral, or bilateral agreements)	Internal communication between the delegated members of the WG and their institutions (sectors) takes time to get approval
Statements are sorted according to their overall similarity on a binary scale (0, 1). Sorting is carried out by each participant ('card sorting' method)	Statements are related to external references (sustainability themes) via an expert ranking (0–3). Ranks are approved by the WG. Ranking is carried out independently by a group of experts	Concept mapping works with direct comparisons, which are difficult to make over large sets of N statements. In such a case, $N \times (N - 1)/2$ of pair matching has to be done. An alternative simpler sorting method can be used
Statements are positioned in the concept map by means of MDS, clustering has been found as too rigid	Statements are clustered accordingly of their average ranking (relationship to T_1–T_n). Number of Ts used for clustering can be reduced by correlation analysis	Ranking of statements accordingly to their relationship to T_1–T_n can be supervised or modified by WG members. Weights of ranks can be adjusted (AHP for groups)
Map is visualized using graphical 2-D outputs from MDS. It is possible to create several maps. The selection of the proper map is a subject of discussion	Statements are visualized by a single SOM	SOM is easy to construct. MDS may require recalculation of $(N \times N/2)$ ranking matrix. In both cases, commercial software has to be used
Maps based upon MDS present a concept of planning process (tool for cognition and structured communication)	SOM of N statements can be transferred into a set of interventions created by SWOT. The new structure is analyzed for conflicts between interventions (objectives)	Concept mapping is more focused on cognitive processes. The information model is an intermediate between list of statements and the final planning document
Authors of concept mapping found clustering to be a too rigid method, which tends to yield inconsistent clusters. MDS with no boundaries between groups of statements is 'softer'	Clusters are viewed as a starting point for the information model of the strategy. Shifts of statements to neighboring clusters or multiple use of key statements in neighboring clusters are expected	Ranking of statements is in both cases a subjective process. MDS maps or SOM clusters depend upon selection of statements and their ranking. In both cases, maps are valuable tools mediating structured planning

Continued

Table 4: *Continued*

| Concept mapping (Trochim [10])	Information model	Remarks
Optimal structure is reached via mediated discussion of participants during an interpretation of the map	Optimal structure of clusters is validated by the ability of the planning team to carry out a set of SWOT analyses	Clustering enables optimization of the information model with respect to weights of T_1–T_n. The optimal number of clusters, D, is N/6 to N/10 (relates to the size of an optimal SWOT table)
Alternatives are not generated	Alternatives can be easily generated by changing weights of T_1–T_n	WG can adjust weights via AHP. SOM-Ward model corrects mutually correlated Ts
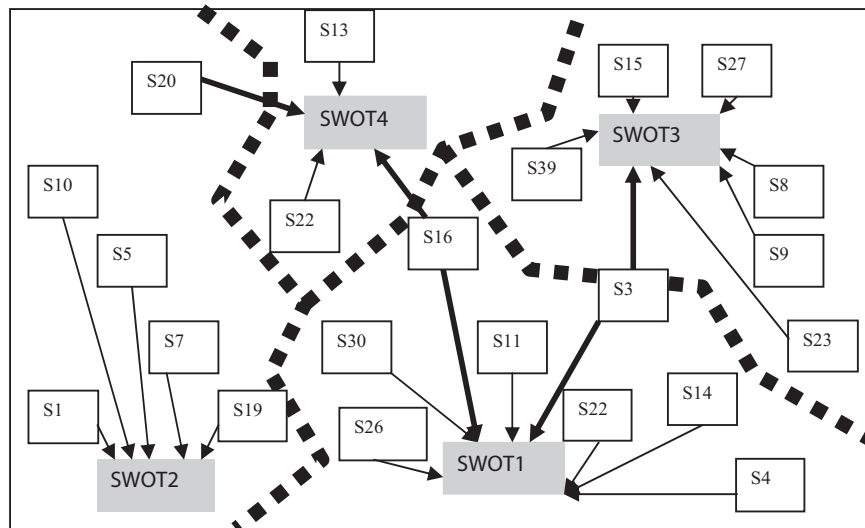
Figure 4: A schematic diagram of part of a SOM with four clusters. Arrows show to which SWOT each statement belongs: statements S3 and S16 are used simultaneously in two SWOT tables and statement S20 is coherent only with SWOT4.

Figure 4 presents a schematic diagram of SOM, in which there are four clusters containing statements ordered in accordance to their T_1–T_n coordinates.

The statements can be converted into the SWOT categories: S: strength, W: weakness, O: opportunity, and T: threat. Therefore, the validity of each cluster can be tested only by formulating a SWOT analysis. There is a certain flexibility to formulate a statement as S or O, or as W or T. Some statements close to a cluster boundary can be used in neighboring SWOT analysis, such as S3 or S16, or completely shifted to other clusters such as S20 (Fig. 4). The initial SOM is therefore a rough framework to be further modified and discussed. In some cases, the clusters are evidently heterogeneous to be transformed into consistent SWOT tables.

3.1 Validity of clusters

There are mathematical tests of validity and quality of clusters based upon intra-cluster similarity or dissimilarity [19–21]. However in our case, the quality of clusters does not depend upon the clustering procedure only but also upon:

- Design of the T_1–T_n in working steps 1 and 7 and
- Correct categorization of statements accordingly to their relationship to T_1–T_n, which can later be revised for individual statements (working steps 3 and 4).

The structure of clusters and their contents used for SWOT analyses can be viewed as an information model of the strategic document drafted. Since the clusters are organized at a higher level in segments according to their mutual similarity, the inconsistencies can be fixed by shifting statements (change of T_1–T_n) to neighboring clusters (working step 4). Sometimes, the structure is labile and a shift of a single statement can cause substantial reordering. However, we are looking for a robust structure of our information model.

Logical consistence of clusters in Model 1 and Model 2 have been tested by our ability to carry out a partial SWOT analysis within each cluster. If the cluster is composed of statements which are not logically related, it has to be split accordingly into two or more fully consistent sub-clusters (in some cases even up to 'single-object clusters'). A cluster with n objects divided into sub-clusters with n_i, n_j, n_k, … items ($n = n_i + n_j + n_k + …$) has therefore an empirical consistency index, ci.

$$ci = (n_i^2 + n_j^2 + n_k^2 + …)/n^2 \tag{1}$$

This means that ci < 1, in comparison with fully consistent clusters having unit consistency. Overall consistency (CI) of a set of clusters is an average ci calculated over all clusters. If all N statements are arranged in D perfectly consistent clusters (CI = 1) then the information based upon relationships between statements inside clusters is fully suitable for SWOT analyses and relationships between neighboring clusters are evident. In case of inconsistent clusters, a part of the information is lost. In any case, there should exist an optimal structure achievable through application of a clustering procedure; growing number of more consistent clusters disturbs the principle of parsimony (lowest possible D is preferred). Behind an optimum it is difficult to grasp the intra-cluster relationships. On the other hand, the decreasing number of clusters, D, decreases their consistency leading to large but inconsistent clusters.

We tried to express this overall information with respect to the number of statements N, consistency CI, and model dimensionality D (number of clusters) by means of a regularization function r(d,n). Regularization functions R(p) are described in the literature related generally to modeling [22–24]. The following regularization functions R(p), where p = D/N, penalize the information loss due to violation of the parsimony principle, and therefore they are used for model selection (eqns 2–5):

$$\text{Akaike final prediction error, FPE} = (1 + p)/(1 – p) \tag{2}$$

$$\text{Bayesian information criterion, BIC} = 1 + \ln(n)p/(2 – 2p) \tag{3}$$

$$\text{Generalized cross-validation, GCV} = 1/(1 – p)^2 \tag{4}$$

$$\text{Shibata's model selector, SMS} = 1 + 2p \tag{5}$$

Table 5: Average consistency index, CI, and its penalized values CI(FPE), CI(GCV), CI(SMS), and CI(BIC) related to model dimensionality D.

Model	CI	N	D	p	CI(FPE)	CI(GCV)	CI(SMS)	CI(BIC)
Hierarchical clustering, Model 1	0.624	170	27	0.16	0.45	0.44	0.47	0.58
Fuzzy, Model 1	0.821	170	44	0.26	0.48	0.45	0.54	0.52
SOM-Ward, Model 1a	0.802	170	27	0.16	0.58	0.57	0.61	0.74
Ward, Model 2	0.72	170	21	0.12	0.56	0.55	0.58	0.77
SOM-Ward, Model 2	0.79	170	21	0.12	0.62	0.61	0.63	0.85
Set of unstructured statements	0.006	170	1	0.01	0.01	0.01	0.01	0.01

In Table 5, the overall CI is penalized (multiplied by each R(p), see eqns 2–5) with respect to the growing dimensionality (D, number of clusters) of the information model. An ideal model is derived from an optimal set of sufficiently consistent clusters, which enables the carrying out of partial SWOT analyses and simultaneously establishes a logical structure across the clusters. As shown in Table 5, Model 1a (SOM-Ward) would be a better alternative to the original Model 1 (hierarchical clustering, NCSS). Model 1a (SOM-Ward) requires only about 15% of shifts to get the final structure used by the WG, which was derived from Model 1 when about one-third of statements had to be shifted. Model 2 (SOM-Ward) presents a planning variant to the structure used originally by the WG. Bezdek and Pal [19] proposed to use several different clustering models and varied the parameters of each of them. If the results across various trials lead to similar and consistent outcomes, the user may assume that a meaningful structure of data has been found.

3.2 Validity of the information model

Validity tests (Table 5) do not take into account the inter-cluster relations, which means the consistency of the whole map. The clusters are positioned in SOM according to their similarity/dissimilarity, which means that thematically similar clusters should be adjacent. It is evident for Model 2, with 21 clusters, that the inter-cluster similarity is higher inside the segments (Table 6) and decreases over the boundary of segments.

Comparing the major groups of Model 2 with four segments of SOM (Table 7), G_1 relates to S_4, G_2 to S_4 and S_1, G_3 to S_1 and S_3, and G_4 to S_2. The differences are caused by statements and clusters which are close to the boundary between segments. If fuzzy clustering is used, these statements have a comparable membership in several clusters which may belong to different segments. It shows that the structure is partially diffuse or fuzzy and this makes it possible to include some statements into two or more SWOT analyses (Fig. 4). Initial SOM is a first blueprint of the final strategic document. Sometimes it helps to exclude statements with lower relevance, e.g. summary ranking lower than 30% of maximum ranking (8 points in case of T_1–T_8). In the case of SDSF, there were only five statements with the total relevance lower than 8 and their exclusion has a slightly positive effect on SOM. Excluded statements can be added back to the resulting structure so that any information is not lost.

In the field of strategic planning, the presentation and adoption of the structure of planning documents is important. An advantage of SOM-Ward method is the combination of a flexible clustering procedure generating variants (via change of weights of T_1–T_n) and additional 2-D visualization of clusters.

Table 6: Groups of clusters with thematic similarity (Model 2).

Group of similar clusters	Thematic content and key clusters	Related clusters
G_1 (C8, C13, C15, and C19)	R&D, innovation, education, energy, and material efficiency	C10: tourism and services; C2: labor force
G_2 (C1, C3, C14, and C17)	Climate change, air emissions and sources, energy, economy, global threats	C12: efficient governance, regulatory reform, agriculture
G_3 (C3, C4, C6, and C11)	Rural areas, spatial planning and regional development, hydrologic regime, floods, soil protection, biodiversity, forest, mining	C20: blackouts, disasters; C16: access to transport services, road accidents
G_4 (C2, C5, C7, and C18)	Family, education, aging society, health, culture, social cohesion, social exclusion, life in cities	C9: floods; C16: access to transport services, road accidents; C20: blackouts, disasters; C21: small and medium enterprises

Table 7: Prevailing contents, size (number of statements), and T_1–T_8 coordinates of segments, Model 2.

Segment	Contents	Size	T_4	T_7	T_8	T_1	T_5	T_6	T_2	T_3
S_1	Sustainable regional development	64	2.75	2.66	1.97	2.80	1.27	1.11	1.61	1.84
S_2	Stable and educated society, efficient governance	66	0.48	2.92	2.06	1.27	2.14	2.53	1.67	1.59
S_3	Transport and regional economy	26	1.88	1.46	0.11	2.08	1.70	0.88	1.08	1.77
S_4	Globalization challenge	14	1.86	0.93	2.93	2.36	0.00	2.00	2.79	2.00

4 CONCLUSIONS

Clustering followed by series of SWOT analyses is a novel technique, enabling the development of large-scale strategic documents covering a wide spectrum of sectors and accommodating various economic, social, and environmental issues. In this way, an initial set of statements (problems, trends, expectations, etc.) can be converted by multiple SWOT analyses into a consistent set of interventions. An optimal structure of the clustered statements has to be found experimentally. In this way, the model allows for the derivation of a structured set of interventions and objectives (strategy) which are relevant to the strategic vision. In our experience, the result of clustering is never perfect due to several semi-objective steps (excerption and formulation of statements, development of T_1–T_n, classification of statements accordingly). It can be further improved by shifting inconsistent statements into thematically related (preferably neighboring) clusters.

Use of SOM in combination with a proper clustering method, e.g. Ward's clustering procedure not only makes searching of the optimal structure faster, but also enables visualization of the result. The resulting structure (information model of the strategy) can be discussed and purposefully modified before it is transformed into a final full text of the plan.

ACKNOWLEDGMENT

REFERENCES

[1] *Our Common Future, Report of the World Commission on Environment and Development*, UN World Commission on Environment and Development, August 2 1987, Published as Annex to General Assembly document A/42/427, Development and International Co-operation: Environment, http://www.un-documents.net/a42-427.htm.

[2] *World Summit Outcome*, UN Document A/60/L.1, 15 September 2005, http://www.un.org/summit2005/documents.html.

[3] Pezzoli K., Sustainable Development: A Trans-disciplinary Overview of the Literature. *Journal of Environmental Planning and Management*, 40(5), pp. 549–574, 1997. doi: http://dx.doi.org/10.1080/09640569711949

[4] Council of the EU, *The Sustainable Development Strategy of the European Union*, Document No. 10917/06, 2006, http://register.consilium.europa.eu/pdf/en/06/st10/ st10917.en06.pdf.

[5] Faber, R., Jorna, N. & van Engelen J., The Sustainability of "Sustainability" - A study into the conceptual foundations of the notion of "sustainability". *Journal of Environment Assessment Policy and Management*, **7(1)**, pp. 1–33, 2005. doi: http://dx.doi.org/10.1142/S1464333205001955

[6] Hendry, K.P., Kiel, G.C. & Nicholson, G J., How boards strategise: a strategy as practice view. *Long Range Planning*, **43(1)**, pp. 33–56, 2010. doi: http://dx.doi.org/10.1016/j.lrp.2009.09.005

[7] Saaty, T.L., A scaling method for priorities in hierarchical structures. *Journal of Mathematical Psychology*, **15(3)**, pp. 234–281, 1977. doi: http://dx.doi.org/10.1016/0022-2496(77)90033-5

[8] Ramanathan, R. & Ganesh, L.S., Group preference aggregation methods employed in AHP: An evaluation and an intrinsic process for deriving members' weightages. *European Journal of Operational Research*, **79(2)**, pp. 249–265, 1994. doi: http://dx.doi.org/10.1016/0377-2217(94)90356-5

[9] Nondek, L. & Kosíková-Šulcová, S., Sustainability assessment of National Strategic Reference Framework of the Czech Republic by means of key indicators and cluster analysis (Project report in Czech), Regional Environmental Centre, Prague 2006.

[10] Trochim, W.M.K., An introduction to concept mapping for planning and evaluation. *Evaluation and Program Planning*, **12**, pp. 1–16, 1989. doi: http://dx.doi.org/10.1016/0149-7189(89)90016-5

[11] Trochim, W.M.K., An Introduction to Concept Mapping for Planning and Evaluation, 2007, http://www.socialresearchmethods.net/research/epp1/epp1.htm.

[12] Weihrich, H., The TOWS matrix – A tool for situational analysis. *Long Range Planning*, **15(2)**, pp. 54–66, 1982. doi: http://dx.doi.org/10.1016/0024-6301(82)90120-0

[13] Trochim, W.M.K. & Linton, R., Conceptualization for planning and evaluation. *Evaluation and Program Planning*, **9(4)**, pp. 289–308, 1986. doi: http://dx.doi.org/10.1016/0149-7189(86)90044-3

[14] The Strategic Framework for Sustainable Development in the Czech Republic, Ministry of the Environment, Prague 2010, http://www.mzp.cz/C125750E003B698B/en/czech_republic_strategy_sd/ $FILE/KM-SFSD_CR_EN-20100317.pdf.

[15] Kohonen, T., *Self-Organizing Maps*. Third, extended edition, Springer, 2001. doi: http://dx.doi.org/10.1007/978-3-642-56927-2

[16] NCSS. Statistical & Power Analysis Software, Kaysville, Utah, 2007.

[17] Kaufmann, L. & Rousseeuw, P.J., *Finding Groups in Data*, John Wiley & Sons, Inc., New York, 1990.

[18] Viscovery SOMine 5.2 software, Viscovery Software GmbH, Vienna, http://www.viscovery.net/somine/.

[19] Bezdek, J.C. & Pal, N.R., Some new indexes of cluster validity. *IEEE Transactions on System, Man and Cybernetics*, **28(3)**, pp. 301–315, 1998. doi: http://dx.doi.org/10.1109/3477.678624

[20] Chou, C.H., Su M.C. & Lai, E., A new cluster validity measure and its application to image compression. *Pattern Analysis and Applications*, **7(2)**, pp. 205–220, 2004. doi: http://dx.doi.org/10.1007/s10044-004-0218-1

[21] Kim, M. & Ramakrishna, R.S., New indices for cluster validity assessment. *Pattern Recognition Letters*, **26(15)**, pp. 2353–2363, 2005. doi: http://dx.doi.org/10.1016/j.patrec.2005.04.007

[22] Burnham, K.P. & Anderson D.R., *Model Selection and Multimodel Inference: A Practical Information - Theoretic Approach*. 2nd edn, Springer Verlag: Berlin, 2002.

[23] Pereira, C.A.B. & Stern, J.M., Model selection: Full Bayesian approach. *Environmetrics*, **12(6)**, pp. 559–568, 2001. doi: http://dx.doi.org/10.1002/env.482

[24] Schwarz, G., Estimating the Dimension of a Model. *Annals of Statistics*, **6(2)**, pp. 461–464, 1978. doi: http://dx.doi.org/10.1214/aos/1176344136