Alexey Poyda
Geophysical Center RAS
Moscow State University

Environmental Scenario Search Engine

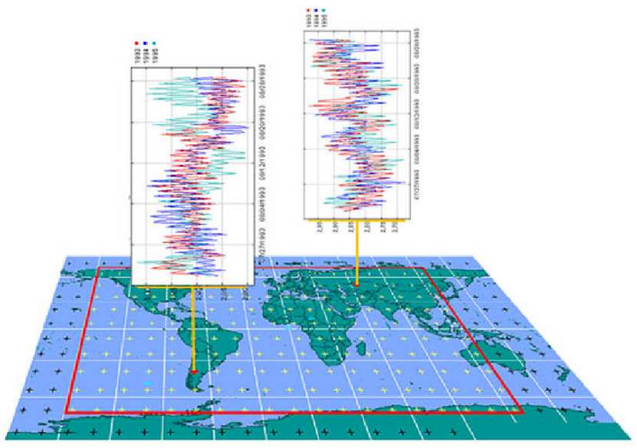


Figure 1: Basic data element is a time series, i.e. an array of values of a parameter at different times at a specific grid point, observatory location, or on specific satellite trajectory.

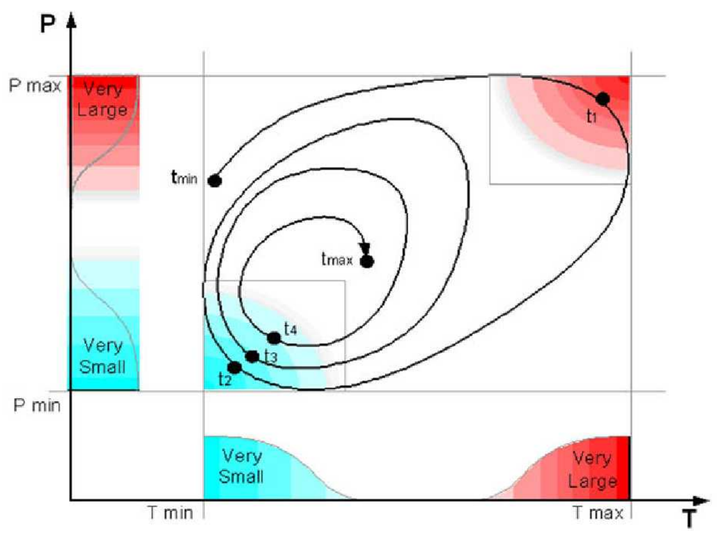


Figure 2: Time-series as a trajectory in the two-dimensional phase space (P – pressure, T – temperature)

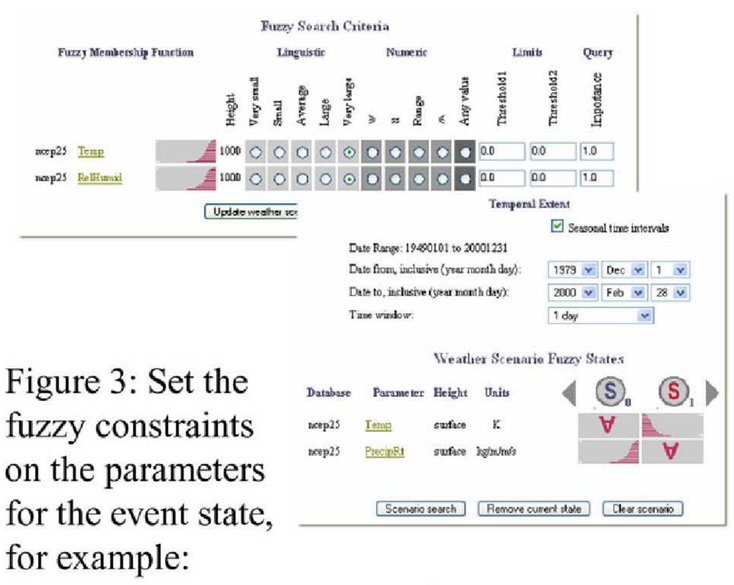
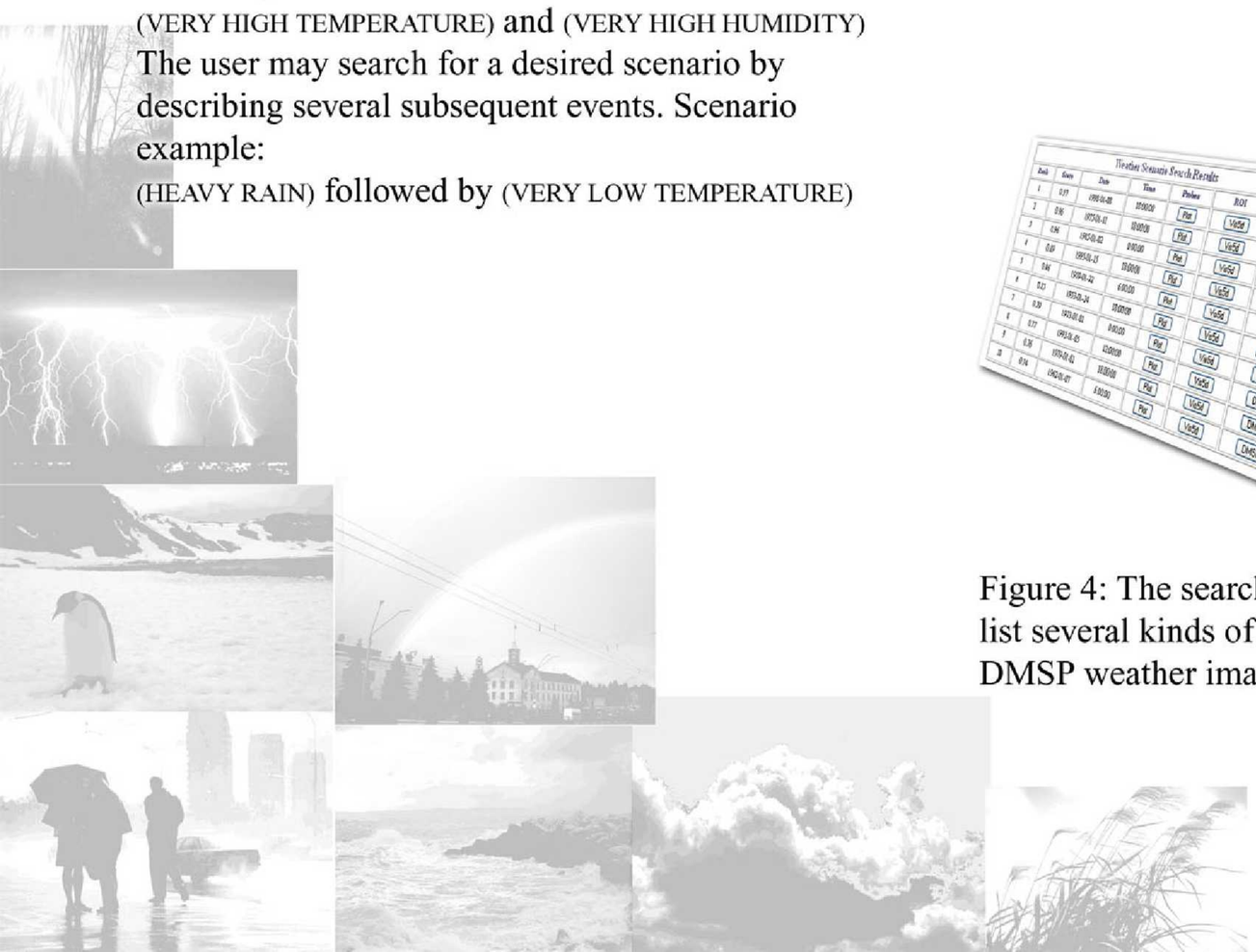


Figure 3: Set the fuzzy constraints on the parameters for the event state, for example:
(VERY HIGH TEMPERATURE) and (VERY HIGH HUMIDITY)
The user may search for a desired scenario by describing several subsequent events. Scenario example:
(HEAVY RAIN) followed by (VERY LOW TEMPERATURE)



We present algorithms and software toolbox for the parallel mining for a set of conditions inside distributed very large databases from multiple environmental domains. The software toolbox is called Environmental Scenario Search Engine (ESSE). The prime requirement of the ESSE system design is to allow the user to query the environmental data archives in human linguistic terms. The natural environment includes elements from multiple domains such as space, terrestrial weather, oceans and terrain. The mapping between human language and computer systems involves fuzzy logic. We use Data Resource Grid-service abstraction layer to virtualize "sequential databases" providing time-series for our search engine. The Data Resource interface is implemented as an OGSA-DAI component with a simple XML output schema. Time-series selected from the Data Resource in XML format can be mined directly by the ESSE or transformed into another format with XSLT to be used by another client, Microsoft Excel being one of the examples.

The base data model in our study is vector-valued time-series.

$$\mathbf{X} = \{\mathbf{x}(t_i)\},\ i = 1...N,\ \mathbf{x}(t_i) = (x_1(t_i),...,x_M(t_i))$$

where N is the number of time samples, and M is the number of observed parameters. It can be represented as a trajectory in the M-dimensional phase space . For example, in Figure 2 we have two-dimensional trajectory in the pressure-temperature (P-T) space. A (fuzzy) state S in a phase space is a fuzzy set which can be described by fuzzy logic expression composed of predicates describing in numerical or linguistic terms the parameter values in each of M dimensions. For example, the state S1 corresponding to the "red" region in Figure 2 can be described by the fuzzy expression:

$$S_1 = \text{(Very Large(P))AND(Very Large(T))},$$

where the linguistic term Very Large() is a predicate, and the operator AND stands for the fuzzy logic conjunction. In the same way, the state S2 corresponding to the "blue" region is:

$$S_2 = \text{(Very Small(P))AND(Very Small(T))},$$

Now, combining the descriptions of the states with the "time shift" operator Shift(dT,) to describe transitions between the states, we can write the following symbolic expression for the environmental scenario "very low temperature and pressure after very high temperature and pressure":

$$\text{Scenario} = \text{(Shift(dT=1,S1))AND(S2)}.$$

The only pair of observations in Figure 2 which fit the above scenario is the pair (t1, t2). Our environmental scenario search engine, ESSE, is designed to mine for the phase space transitions like that in very large scientific databases.

The applications of the ESSE systems are broad. As more and more data archives become available through projects like ESG (DOD), CLASS (NOAA), EOSDIS (NASA), DODS (Univ. RI) and other network accessible data systems, the tools to extract information from them become more valuable.

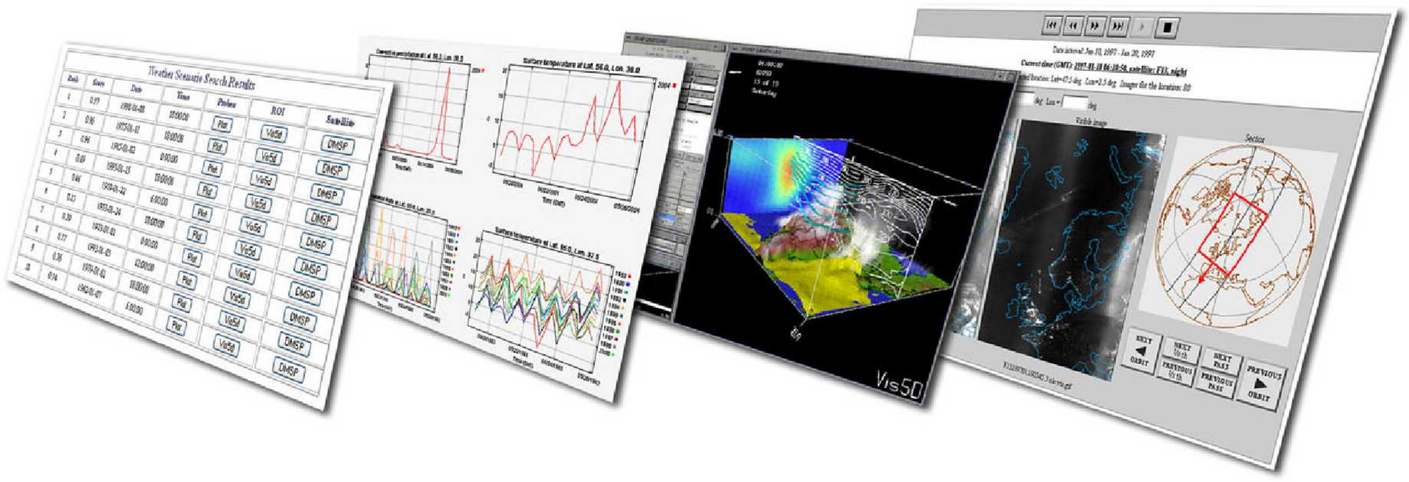


Figure 4: The search results are represented as a list of fuzzy scores. For each element in the list several kinds of visualization are available: time series plots, Vis5D volume animations, DMSP weather images.

http://esse.wdcb.ru
http://teos1.wdcb.ru/esse/