Spatial Statistics in Analysis and Prediction: Methods, Algorithms and Applications

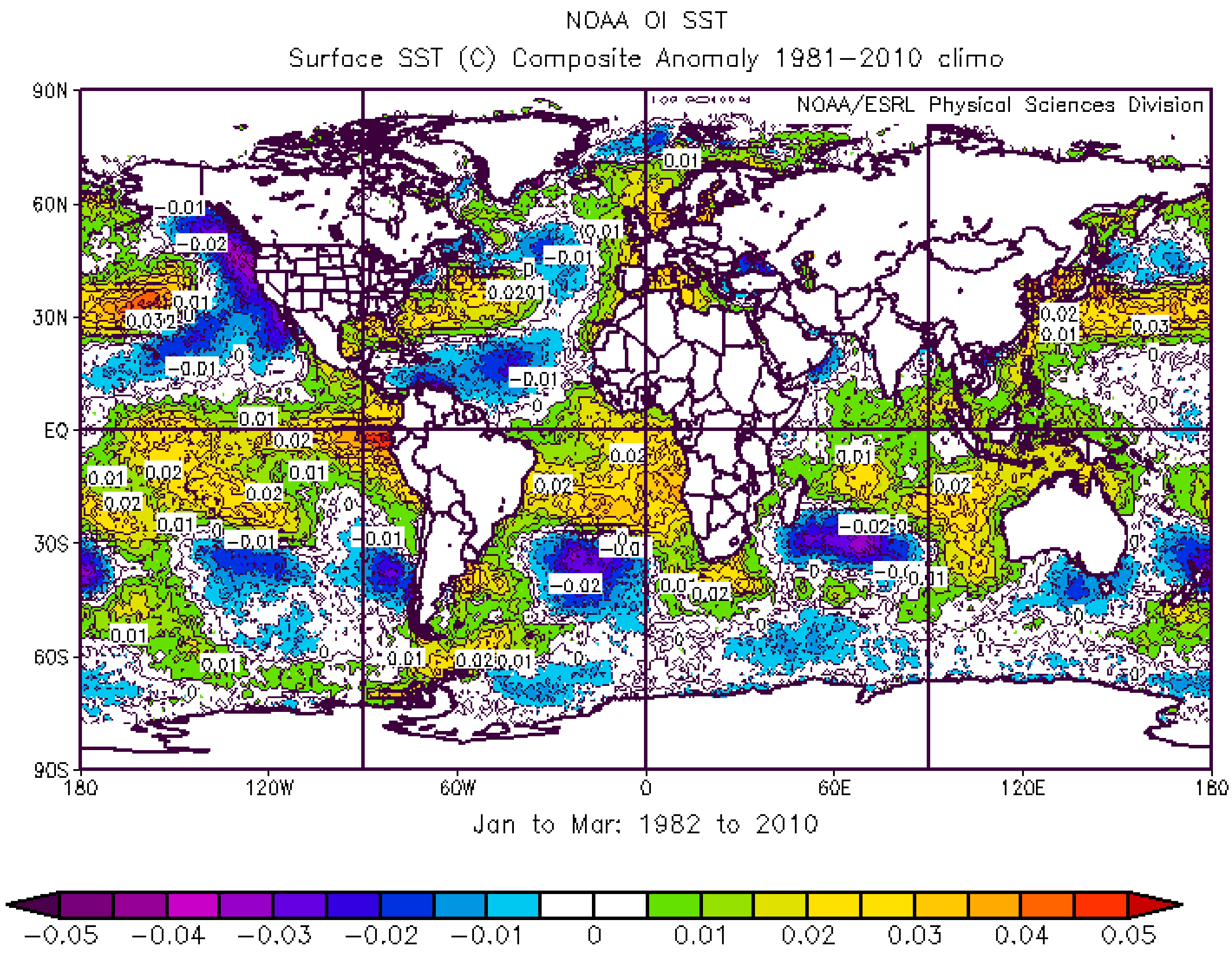
Oleg M.Pokrovsky

Main Geophysical Observatory,

St. Petersburg, Russia

Outlines

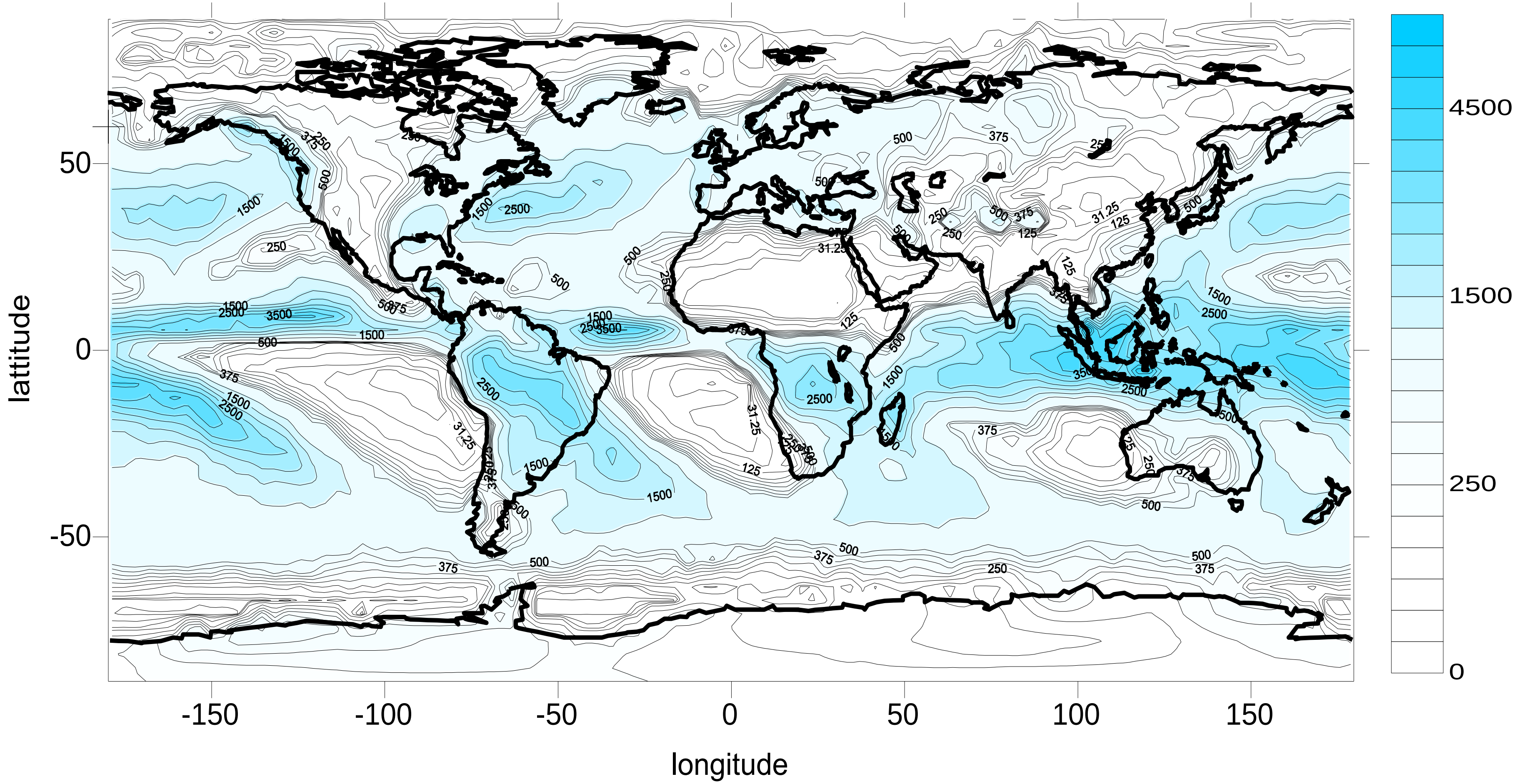
- Data Analysis (3-D and 4-D Assimilation concept)
- Kriging
- Spectral Analysis
- Fuzzy classification
- Fuzzy-Neural prediction model



Circulation to Climate Change

Precipitation field statistics

Global Mean field for precipitation annual sums (mm per year) for 32 years (1979-2009, CMAP data)



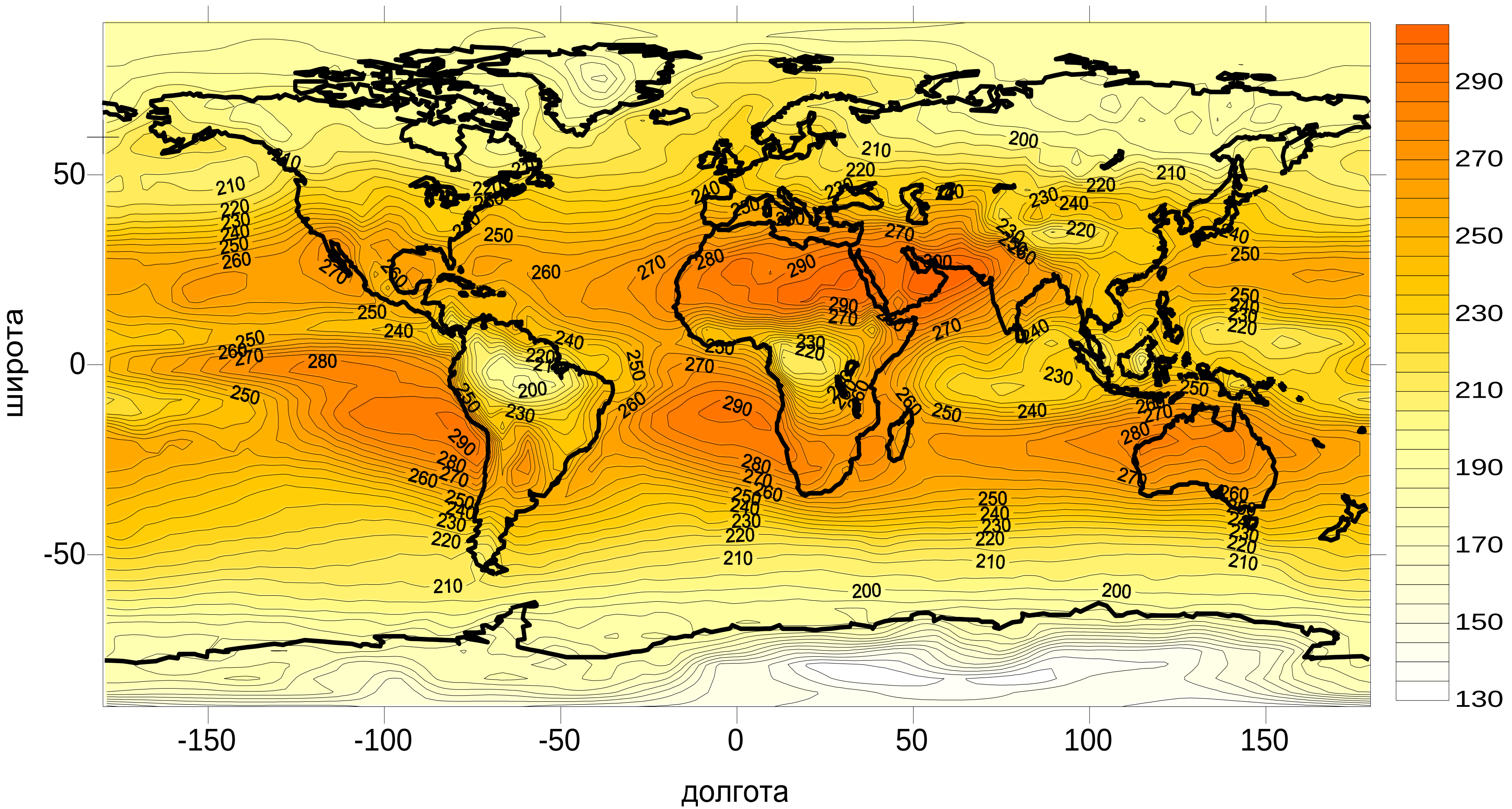
29-30 July 2014

Remote Sensing for Global Water Circulation to Climate Change

Outgoing Longwave Radiation

OLR

Global mean field of OLR (w/m^2) for 1979-2010

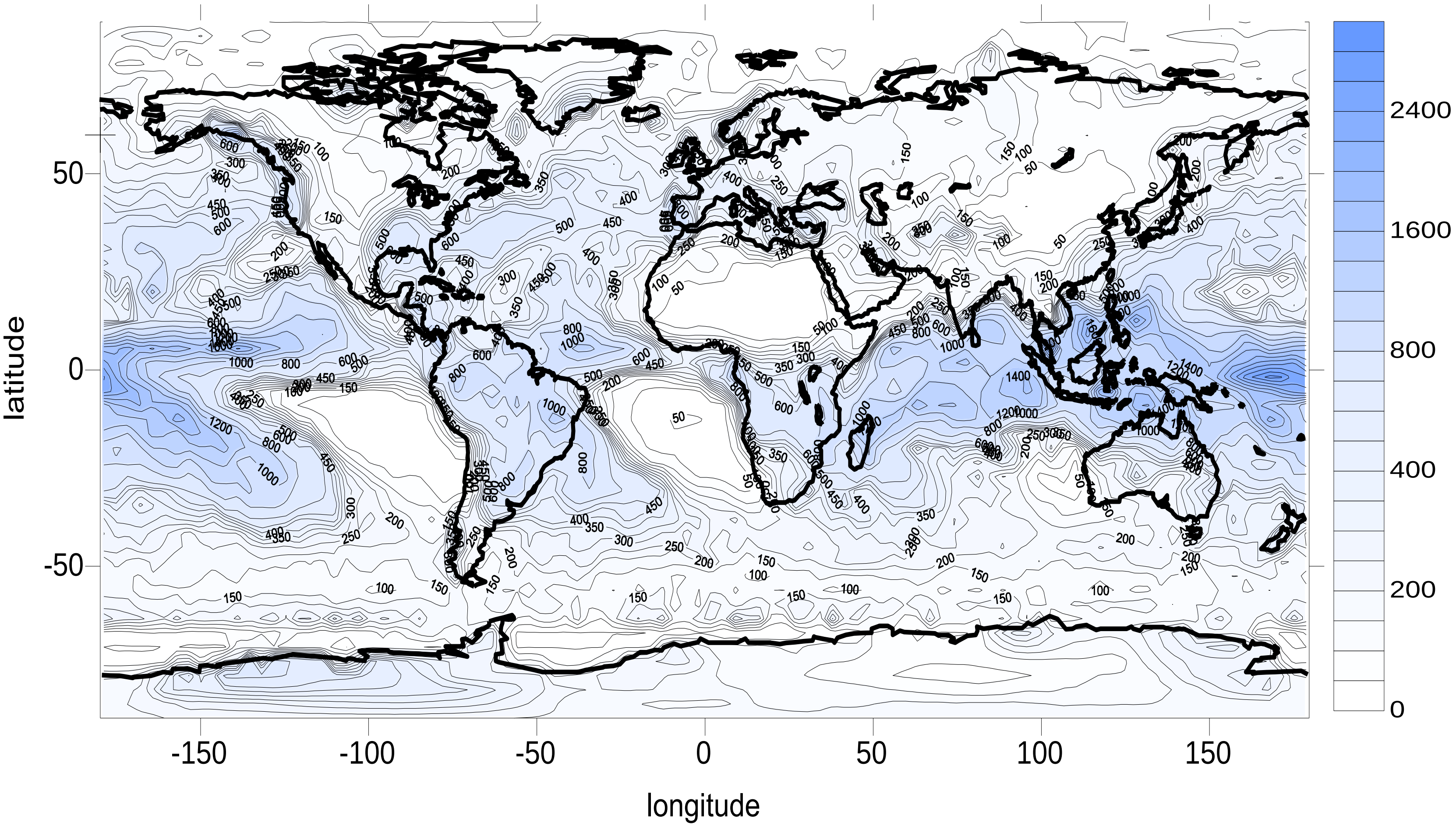


29-30 July 2014

Remote Sensing for Global Water
Circulation to Climate Change

Variability Field

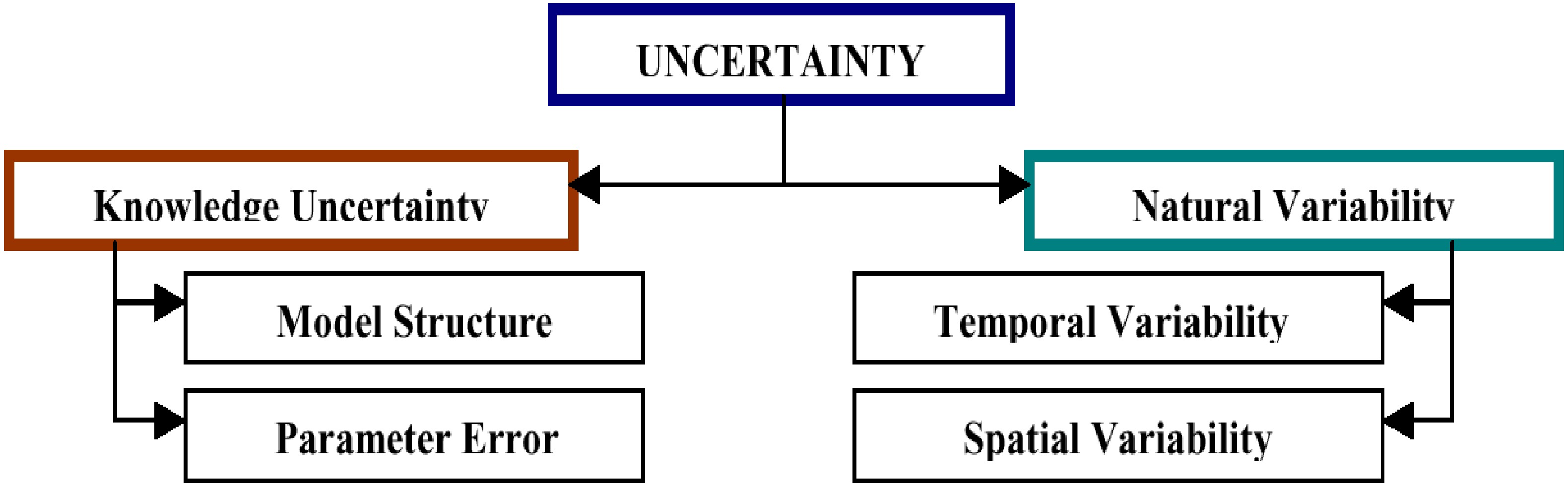
Global field for precipitation variability (STD) annual sums (mm per year) for 32 years (1979-2009, CMAP data)



29-30 July 2014

Remote Sensing for Global Water Circulation to Climate Change

Types of Uncertainties



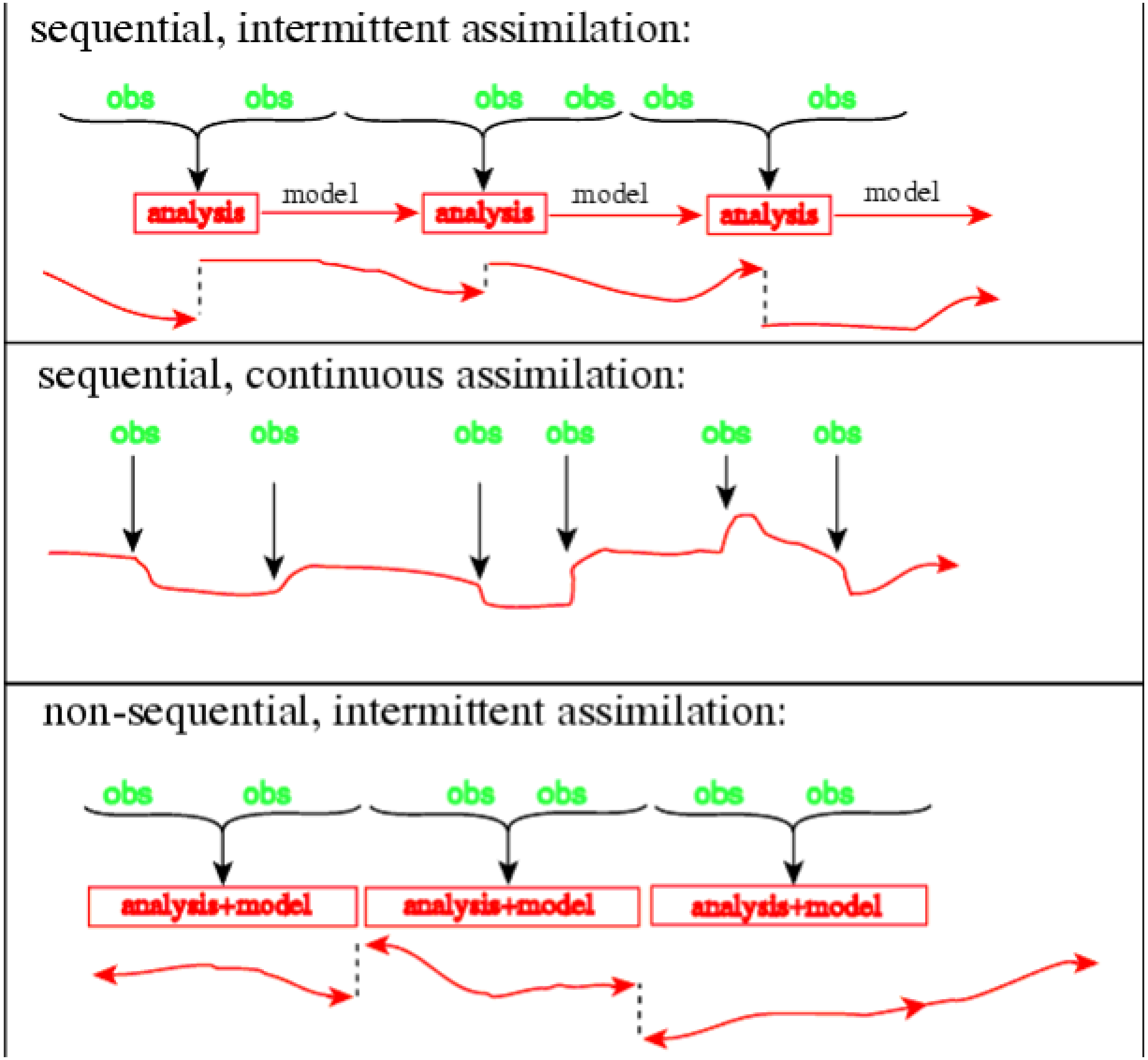
Analysis.

An analysis is the production of an accurate image of the true state of the atmosphere, ocean and surface at a given time, represented in a model as a collection of numbers.

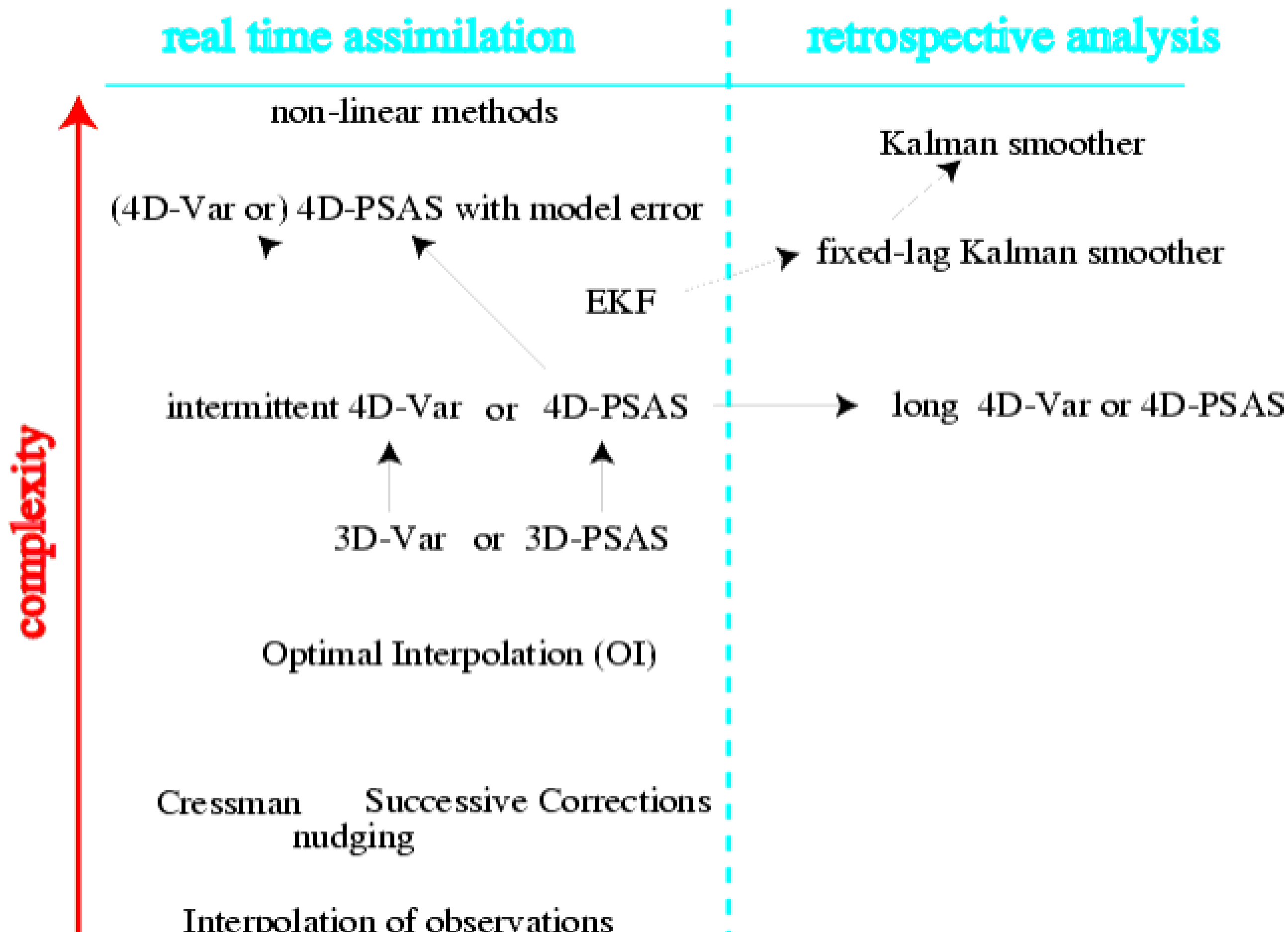
Concept of Data Assimilation

- The background information can be a climatology or a trivial state; it can also be generated from the output of a previous analysis, using some assumptions of consistency in time of the model state, like stationarity (hypothesis of persistence) or the evolution predicted by a forecast model. In a well-behaved system, one expects that this allows the information to be accumulated in time into the model state, and to propagate to all variables of the model.

Concept of Data Assimilation



A summarized history of the main data assimilation algorithms used in meteorology and oceanography, roughly classified according to their complexity (and cost) of implementation, and their applicability to real-time problems. Currently, the most commonly used for operational applications are OI, 3D-Var and 4D-Var.



Cressman analysis

The model state is assumed to be univariate and represented as grid-point values. If we denote by $\mathbf{x}_b$ a previous estimate of the model state (*background*) provided by climatology, persistence or a previous forecast, and by $\mathbf{y}(i)$, a set of $i = 1 \ldots\ldots n$ observations of the same parameter, a simple kind of Cressman analysis is provided by the model state $\mathbf{x}_a$ defined at each grid point *j* according to the following *update* equation:

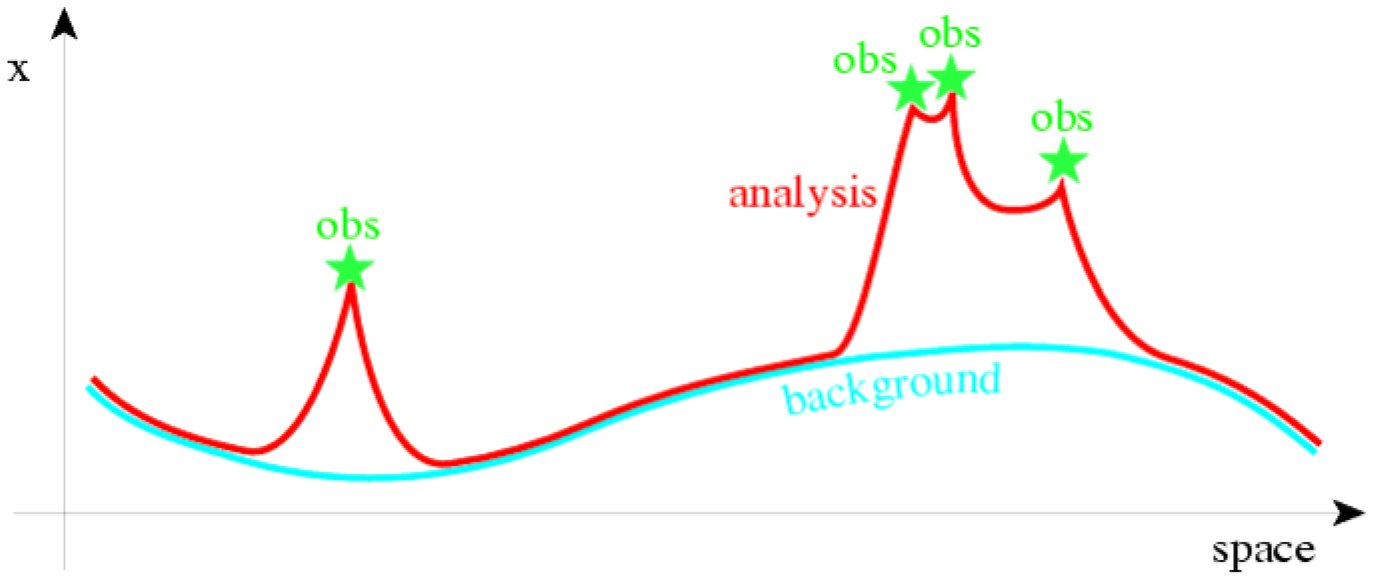
$$\mathbf{x}_a(j) = \mathbf{x}_b(j) + \frac{\sum_{i=1}^{n} w(i,j)\{\mathbf{y}(i) - \mathbf{x}_b(i)\}}{\sum_{i=1}^{n} w(i,j)}$$

$$w(i,j) = \max\left(0, \frac{R^2 - d_{i,j}^2}{R^2 + d_{i,j}^2}\right)$$

where $d_{i,j}$ is a measure of the distance between points i and j . $\mathbf{x}_b(i)$ is the background state interpolated to point i . The weight function $w(i,j)$ equals one if the grid point j is collocated with observation i . It is a decreasing function of distance which is zero if $d_{i,j} > R$, where *R* is a user-defined constant (the "influence radius") beyond which the

observations have no weight

An example of Cressman analysis of a one-dimensional field. The background field   is represented as the blue function, and the observations in green. The analysis (black curve) is produced by interpolating between the background (grey curve) and the observed value, in the vicinity of each observation; the closer the observation, the larger its weight.



The Cressman method is not satisfactory in practice for the following reasons:

- if we have a preliminary estimate of the analysis with a good quality, we do not want to replace it by values provided from poor quality observations.
- when going away from an observation, it is not clear how to relax the analysis toward the arbitrary state, i.e. how to decide on the shape of the function   .
- an analysis should respect some basic known properties of the true system, like smoothness of the fields, or relationship between the variables (e.g. hydrostatic balance, or saturation constraints). This is not guaranteed by the Cressman method: random observation errors could generate unphysical features in the analysis.

Kriging

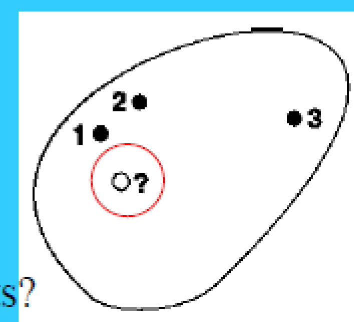
- Weighted Linear Estimators
- Some Definitions
- Derivation of the Kriging Equations
- Some Examples
- Different Types of Kriging
- How / Where is Kriging Used

Weighted Linear Estimators

- The basic idea is to estimate the attribute value (say, porosity) at a location where we do not know the true value

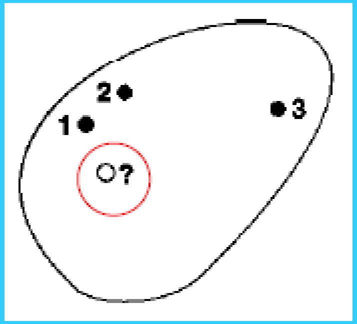
$$\mathbf{Z^*(u) = \sum_{i=1}^{n} \lambda_i \cdot Z(u_i)}$$

- where **u** refers to a location, $Z^*(\mathbf{u})$ is an estimate at location **u**, there are n data values $Z(\boldsymbol{u}_i)$, i=1,...,n, and λ_i refer to weights.



- What factors could be considered in assigning the weights?
  - –closeness to the location being estimated
  - –redundancy between the data values
  - –anisotropic continuity (preferential direction)
  - –magnitude of continuity / variability

Weighted Linear Estimators



- Assign all of the weight to the nearest data (polygonal-type estimate)
- Assign the weights inversely proportional to the distance from the location being estimated (inverse distance schemes)

$$\lambda_i = \frac{\frac{1}{c + d_i^w}}{\sum_{i=1}^{n} \frac{1}{c + d_i^w}}$$

where d_i is the distance between data i and the location being estimated, c is a small constant, and ω is a power (usually between 1 to 3).

- How about using the variogram? ↦ that is kriging

Some Definitions

- Consider the residual data values:

$$Y(\mathbf{u}_i) = Z(\mathbf{u}_i) - m(\mathbf{u}_i),\ i=1,\ldots,n$$

- where m(**u**) could be constant, locally varying, or considered constant but unknown.
- Variogram is defined as:

$$2\,\gamma(\mathbf{h}) = \mathbf{E}\{[\ Y(\mathbf{u}\})\ -\ Y(\mathbf{u}+\mathbf{h}]^2\}$$

- Covariance is defined as:

$$C(\mathbf{h}) = \mathbf{E}\{\ Y(\mathbf{u}) \cdot Y(\mathbf{u}+\mathbf{h})\}$$

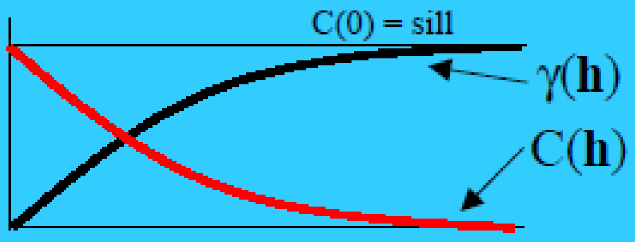
- Link between the Variogram and Covariance:

$$2\,\gamma(\mathbf{h}) = [\,E\{\ Y^2(\mathbf{u}) + [\,E\{\ Y^2(\mathbf{u}+\mathbf{h})\}]\ -\ 2 \cdot [\,E\{\ Y(\mathbf{u}) \cdot Y(\mathbf{u}+\mathbf{h})]$$

$$= \mathrm{Var}\{Y(\mathbf{u})\} + \mathrm{Var}\{Y(\mathbf{u}+\mathbf{h})\} - 2 \cdot C(\mathbf{h}\})$$

$$= 2\,[\,C(0) - C(\mathbf{h})]$$

So, $C(\mathbf{h}) = C(0) - \gamma(\mathbf{h})$



Simple Kriging (1)

- Consider a linear estimator:

$$Y^*(\mathbf{u}) = \sum_{i=1}^{n} \lambda_i \cdot Y(\mathbf{u}_i)$$

- where $Y(\mathbf{u}_i)$ are the residual data (data values minus the mean) and $Y^*(\mathbf{u})$ is the estimate (add the mean back in)
- The error variance is defined as

A^2-2ab+b^2

$$E\{[Y^*(u) - Y(u)]^2\}$$

$$E\{[Y^*(u)]^2\} - 2 \cdot E\{Y^*(u) \cdot Y(u)\} + E\{[Y(u)]^2\}$$

$$\sum_{i=1}^{n}\sum_{j=1}^{n} \lambda_i \lambda_j E\{Y(u_i) \cdot Y(u_j)\} - 2 \cdot \sum_{i=1}^{n} \lambda_i E\{Y(u) \cdot Y(u_i)\} + C(0)$$

$$\sum_{i=1}^{n}\sum_{j=1}^{n} \lambda_i \lambda_j C(u_i, u_j) - 2 \cdot \sum_{i=1}^{n} \lambda_i C(u, u_i) + C(0)$$

Simple Kriging (2)

- Optimal weights $\lambda_{i},_{i}$=1,…,n may be determined by taking partial derivatives of the error variance w.r.t. the weights

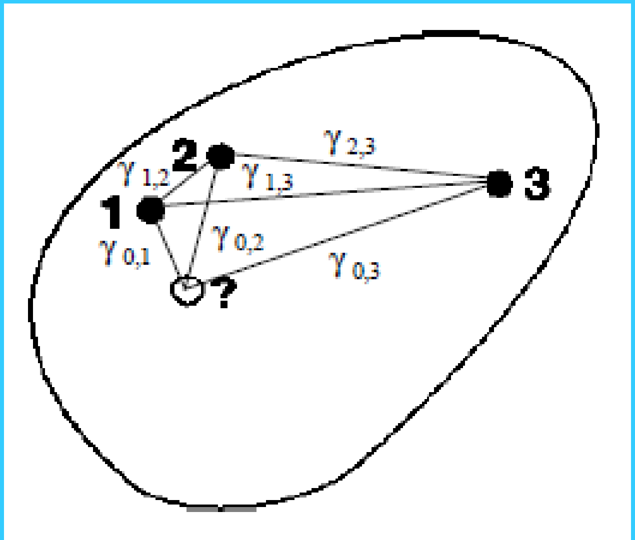
$$\frac{\partial[\quad]}{\partial\lambda_i} = 2 \cdot \sum_{j=1}^{n} \lambda_j C(u_i, u_j) - 2 \cdot C(u, u_i), \quad i = 1,\dots,n$$

- and setting them to zero

$$\sum_{j=1}^{n} \lambda_j C(u_i, u_j) = C(u, u_i), \quad i = 1,\dots,n$$

- This system of n equations with n unknown weights is the simple kriging (SK) system

Simple Kriging: Some Details



- There are three equations to determine the three weights:

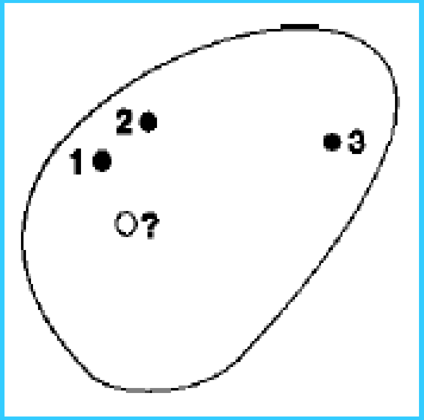
$$\lambda_1 \cdot C(1,1) + \lambda_2 \cdot C(1,2) + \lambda_3 \cdot C(1,3) = C(0,1)$$
$$\lambda_1 \cdot C(2,1) + \lambda_2 \cdot C(2,2) + \lambda_3 \cdot C(2,3) = C(0,2)$$
$$\lambda_1 \cdot C(3,1) + \lambda_2 \cdot C(3,2) + \lambda_3 \cdot C(3,3) = C(0,3)$$

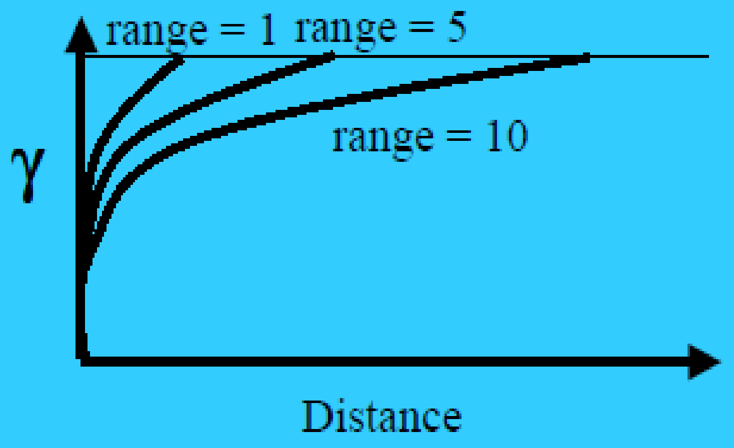
- In matrix notation: (Recall that C(**h**}) = C(0) - γ(**h**))

$$\begin{bmatrix} C(1,1) & C(1,2) & C(1,3) \\ C(2,1) & C(2,2) & C(2,3) \\ C(3,1) & C(3,2) & C(3,3) \end{bmatrix} \begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \lambda_3 \end{bmatrix} = \begin{bmatrix} C(0,1) \\ C(0,2) \\ C(0,3) \end{bmatrix}$$

Simple Kriging

Changing the Range





Simple kriging with a zero nugget effect and an isotropic spherical variogram with three different ranges:

	λ_1	λ_2	λ_3
range = 10	0.781	0.012	0.065
5	0.648	-0.027	0.001
1	0.000	0.000	0.000

- Kriging is a procedure for constructing a minimum error variance linear estimate at a location where the true value is unknown
- The main controls on the kriging weights are:
  - closeness of the data to the location being estimated
  - redundancy between the data
  - the variogram
- Simple Kriging (SK) does not constrain the weights and works with the residual from the mean
- Ordinary Kriging (OK) constrains the sum of the weights to be 1.0, therefore, the mean does not need to be known
- There are many different types of kriging
- Two implicit assumptions are stationarity (work around with different types of kriging) and ergodicity (more slippery)
- Kriging is not used directly for mapping the spatial distribution of an attribute (sometimes when the attribute is smooth). It is used, however, for building conditional distributions for stochastic simulation

The ingredients of a good analysis:

- 1) one should start from a good-quality first guess, i.e. a previous analysis or forecast that gives an overview of the situation,
- 2) if observations are dense, then one assumes that the truth probably lies near their average. One must make a compromise between the first guess and the observed values. The analysis should be closest to the data we trust most, whereas suspicious data will be given little weight (e.g. for less accurate remote sensing data).
- 3) the analysis should be smooth, because we know that the true field is. When going away from an observation, the analysis will relax smoothly to the first guess on scales known to be typical of the usual physical phenomena.
- 4) the analysis should also try to respect the known physical features of the system. Of course, it is possible in exceptional cases that unusual scales and imbalances happen, and a good analyst must be able to recognize this, because exceptional cases are usually important too.

State vector

- a column matrix called the *state vector* **x**
- a state vector , the *true* state at the time of the analysis $\mathbf{x_t}$
- $\mathbf{x_b}$ *background* estimate of the true state before the analysis is carried out
- the*analysis* is denoted $\mathbf{x_a}$,

which is what we are looking for

The key to data analysis is the use of the discrepancies between observations and state vector. According to the previous paragraph, this is given by the vector of departures at the observation points:

$$\mathbf{y} - \mathbf{H}(\mathbf{x})$$

Error analysis1

Given a background field $\mathbf{x}_b$ just before doing an analysis, there is one and only one vector of errors that separates it from the true state:

$$\varepsilon_b = \mathbf{x}_b - \mathbf{x}_t$$

If we were able to repeat each analysis experiment a large number of times, under exactly the same conditions, but with different realizations of errors generated by unknown causes, ε_b would be different each time. We can calculate statistics such as averages, variances and histograms of frequencies of ε_b . In the limit of a very large number of realizations, we expect the statistics to converge to values which depend only on the physical processes responsible for the errors, not on any particular realization of these errors. When we do another analysis under the same conditions, we do not expect to know what will be the error ε_b , but at least

Error analysis2

The errors in the background and in the observations[2] are modelled as follows:

- **background errors**: $\varepsilon_b = \mathbf{x}_b - \mathbf{x}_t$, of average $\bar{\varepsilon}_b$ and covariances $\mathbf{B} = \overline{(\varepsilon_b - \bar{\varepsilon}_b)(\varepsilon_b - \bar{\varepsilon}_b)^T}$. They are the estimation errors of the background state, i.e. the difference between the background state vector and its true value. They do not include discretization errors.
- **observation errors**: $\varepsilon_o = \mathbf{y} - H(\mathbf{x}_t)$, of average $\bar{\varepsilon}_o$ and covariances $\mathbf{R} = \overline{(\varepsilon_o - \bar{\varepsilon}_o)(\varepsilon_o - \bar{\varepsilon}_o)^T}$. They contain errors in the observation process (instrumental errors, because the reported value is not a perfect image of reality), errors in the design of the operator H , and representativeness errors i.e. discretization errors which prevent $\mathbf{x}_t$ from being a perfect image of the true state[3].
- **analysis errors**: $\varepsilon_a = \mathbf{x}_a - \mathbf{x}_t$, of average $\bar{\varepsilon}_a$. A measure $\overline{\|\varepsilon_a - \bar{\varepsilon}_a\|}$ of these errors is given by the trace of the analysis error covariance matrix $\mathbf{A}$,

$$\mathrm{Tr}(\mathbf{A}) = \overline{\|\varepsilon_a - \bar{\varepsilon}_a\|^2} .$$

Notations:

The dimension of the model state is n and the dimension of the observation vector is p. We will denote:

- $\mathbf{x}_t$ true model state (dimension n)
- $\mathbf{x}_b$ background model state (dimension n)
- $\mathbf{x}_a$ analysis model state (dimension n)
- $\mathbf{y}$ vector of observations (dimension p)
- H observation operator (from dimension n to p)
- $\mathbf{B}$ covariance matrix of the background errors $(\mathbf{x}_b - \mathbf{x}_t)$ (dimension $n \times n$)
- $\mathbf{R}$ covariance matrix of observation errors $(\mathbf{y} - H[\mathbf{x}_t])$ (dimension $p \times p$)
- $\mathbf{A}$ covariance matrix of the analysis errors $(\mathbf{x}_a - \mathbf{x}_t)$ (dimension $n \times n$)

Hypotheses

The following hypotheses are assumed:

- **Linearized observation operator**: the variations of the observation operator in the vicinity of the background state are linear: for any $\mathbf{x}$ close enough to $\mathbf{x}_b$, $H(\mathbf{x}) - H(\mathbf{x}_b) = \mathbf{H}(\mathbf{x} - \mathbf{x}_b)$ where $\mathbf{H}$ is a linear operator.
- **Non-trivial errors**: $\mathbf{B}$ and $\mathbf{R}$ are positive definite matrices.
- **Unbiased errors**: the expectation of the background and observation errors is zero i.e. $\overline{\mathbf{x}_b - \mathbf{x}_t} = \overline{\mathbf{y} - H(\mathbf{x}_t)} = 0$
- **Uncorrelated errors**: observation and background errors are mutually uncorrelated i.e. $\overline{(\mathbf{x}_b - \mathbf{x}_t)(\mathbf{y} - H[\mathbf{y}_t])^T} = 0$
- **Linear analysis**: we look for an analysis defined by corrections to the background which depend linearly on background observation departures.
- **Optimal analysis**: we look for an analysis state which is as close as possible to the true state in an r.m.s. sense (i.e. it is a minimum variance estimate).

Least-squares analysis equations

(a) The *optimal least-squares estimator*, or *BLUE analysis*, is defined by the following interpolation equations:

$$\mathbf{x}_a = \mathbf{x}_b + \mathbf{K}(\mathbf{y} - H[\mathbf{x}_b]) \quad \text{(A1)}$$

$$\mathbf{K} = \mathbf{B}\mathbf{H}^T(\mathbf{H}\mathbf{B}\mathbf{H}^T + \mathbf{R})^{-1} \quad \text{(A2)}$$

where the linear operator $\mathbf{K}$ is called the *gain*, or *weight matrix*, of the analysis.

(a) The *analysis error covariance matrix* is, for any $\mathbf{K}$:

$$\mathbf{A} = (\mathbf{I} - \mathbf{K}\mathbf{H})\mathbf{B}(\mathbf{I} - \mathbf{K}\mathbf{H})^T + \mathbf{K}\mathbf{R}\mathbf{K}^T \quad \text{(A3)}$$

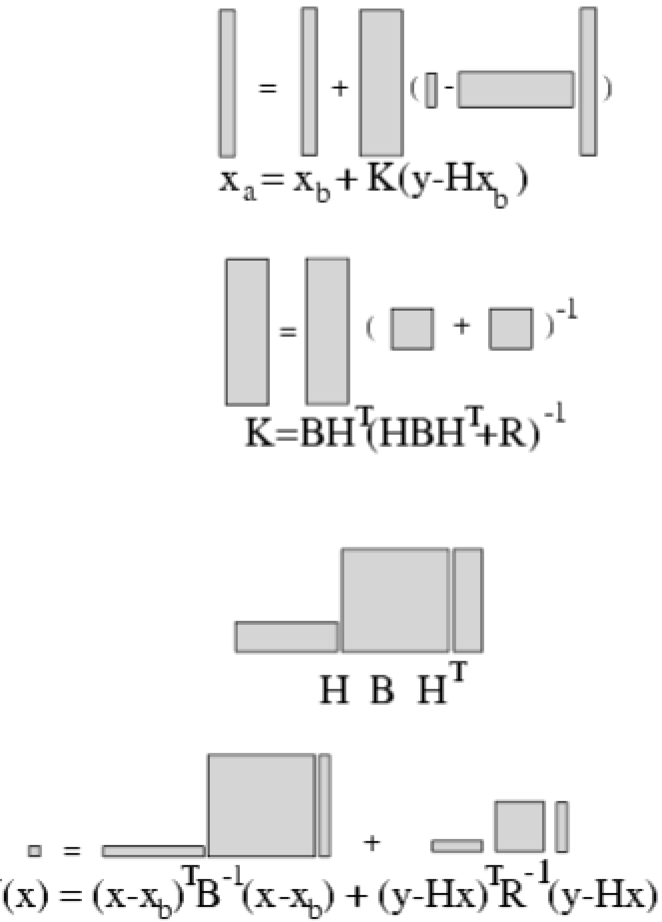
If $\mathbf{K}$ is the optimal least-squares gain, the expression becomes

$$\mathbf{A} = (\mathbf{I} - \mathbf{K}\mathbf{H})\mathbf{B} \quad \text{(A4)}$$

(a) The BLUE analysis is equivalently obtained as a solution to the *variational optimization problem*:

$$\begin{aligned} \mathbf{x}_a &= \operatorname{Arg\,min} J \\ J(\mathbf{x}) &= (\mathbf{x} - \mathbf{x}_b)^T \mathbf{B}^{-1}(\mathbf{x} - \mathbf{x}_b) + (\mathbf{y} - H[\mathbf{x}])^T \mathbf{R}^{-1}(\mathbf{y} - H[\mathbf{x}]) \\ &= J_b(\mathbf{x}) + J_o(\mathbf{x}) \end{aligned} \quad \text{(A5)}$$

Sketches of the shapes of the matrices and vector dimensions involved in an Optimal assimilation analysis



Conclusion

We have seen that there are **two main ways** of defining the statistical analysis problem:

- either assume that the background and error covariances are known, and derive the analysis equations by requiring that the total analysis error variances are minimum,
- or assume that the background and observation error PDFs are Gaussian, and derive the analysis equations by looking for the state with the maximum probability.

Both approaches lead to two mathematically equivalent algorithms:

- the direct determination of the analysis gain matrix ,
- the minimization of a quadratic cost function.

A simple scalar illustration of least-squares estimation

Let us assume that we need to estimate the temperature T_t of a room.

We have a thermometer of known accuracy σ_o (the standard deviation of measurement error) and we observe T_o , which is considered to have expectation T_t (i.e. we assume that the observation is unbiased) and variance σ_o^2 . In the absence of any other information the best estimate we can provide of the temperature is T_o , with accuracy σ_o .

However we may have some additional information about the temperature of the room. We may have a reading from another, independent thermometer, perhaps with a different accuracy. We may notice that everyone in the room is wearing a jumper-another timely piece of information from which we can derive an estimate, although with a rather large associated error. We may have an accurate observation from an earlier date, which can be treated as an estimate for the current time, with an error suitably inflated to account for the separation in time. Any of these observations could be treated as *a priori* or *background* information, to be used with T_o in estimating the room temperature. Let our background estimate be T_b , of expectation T_t (i.e. it is unbiased) and of accuracy σ_b . Intuitively T_o and T_b can be combined to provide a better estimate (or *analysis*) of T_t than any of these taken alone. We are going to look for a linear weighted average of the form:

$$T_a = kT_o + (1-k)T_b$$

which can be rewritten as $T_a = T_b + k(T_o - T_b)$, i.e. we look for a correction to the background which is a linear function of the difference between the observation and the background.

A simple scalar illustration of least-squares estimation 2

The error variance of the estimate is:

$$\sigma_a^2 = (1-k)^2\sigma_b^2 + k^2\sigma_o^2$$

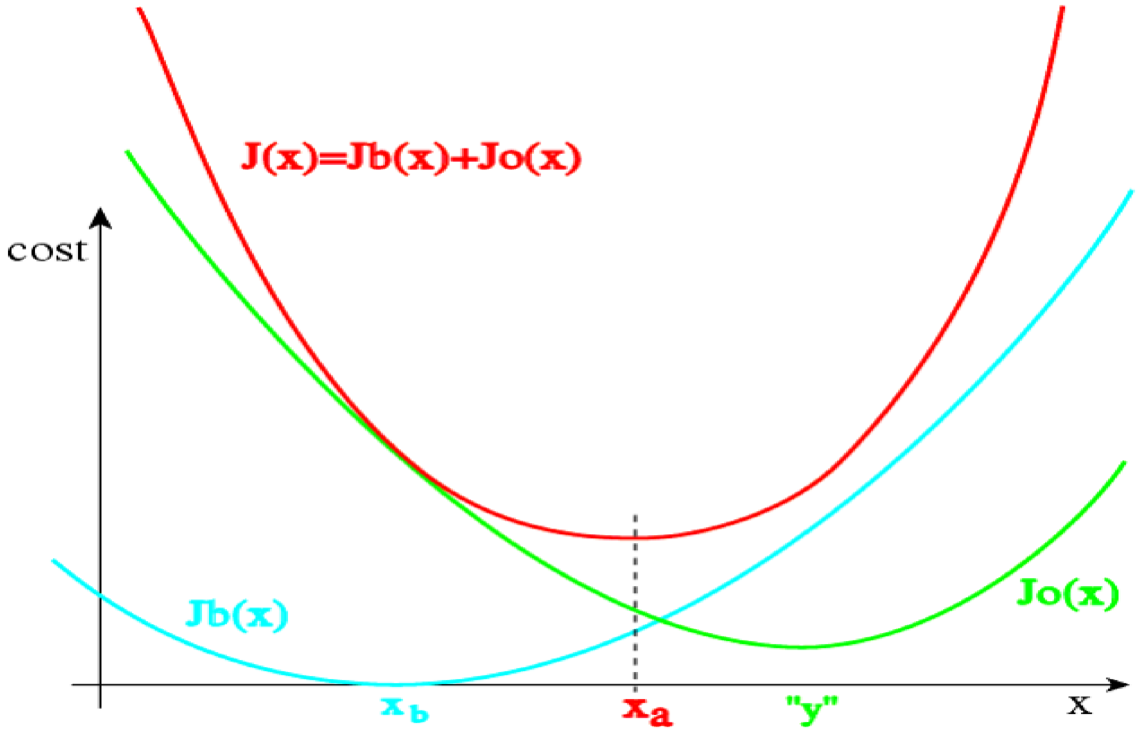
where we have assumed that the observation and background errors are uncorrelated. We choose the optimal value of k that minimizes the analysis error variance:

$$k = \frac{\sigma_b^2}{\sigma_b^2 + \sigma_o^2}$$

which is equivalent to minimizing (Fig. 5)

$$J(T) = J_b(T) + J_o(T) = \frac{(T-T_b)^2}{\sigma_b^2} + \frac{(T-T_o)^2}{\sigma_b^2}$$

Figure 5 . Schematic representation of the variational form of the least-squares analysis, in a scalar system where the observation y is in the same space as the model x : the cost-function terms J_b and J_o are both convex and tend to "pull" the analysis towards the background x_b and the observation y , respectively. The minimum of their sum is somewhere between x_b and y , and is the optimal least-squares analysis.



Error analysis

It is interesting to look at the variance of analysis error for the optimal k :

$$\frac{1}{\sigma_a^2} = \frac{1}{\sigma_o^2} + \frac{1}{\sigma_b^2}$$

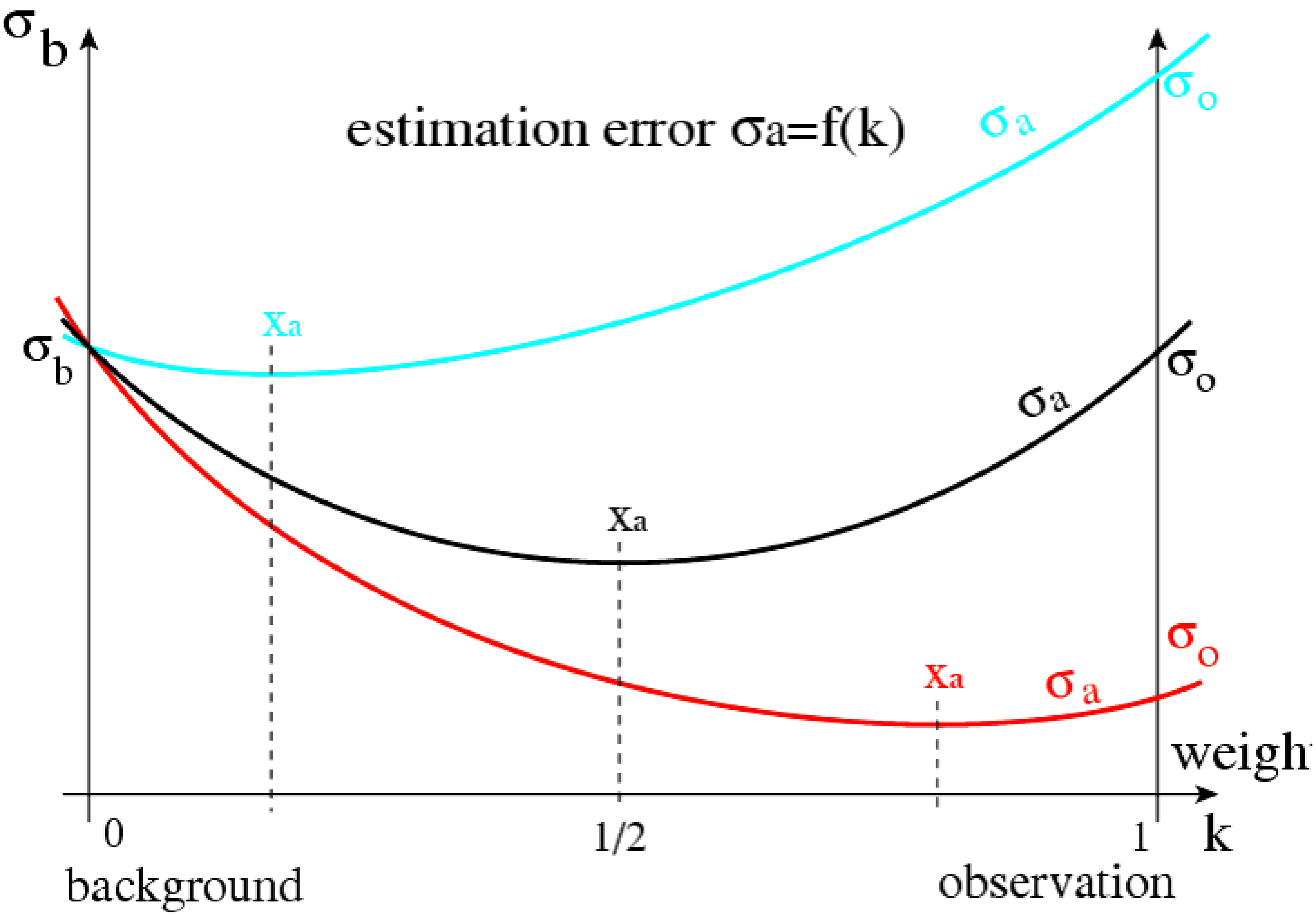
or

$$\sigma_a^2 = \frac{\sigma_o^2}{1+(\sigma_o/\sigma_b)^2} = \frac{\sigma_b^2}{1+(\sigma_b/\sigma_o)^2} = (1-k)\sigma_b^2$$

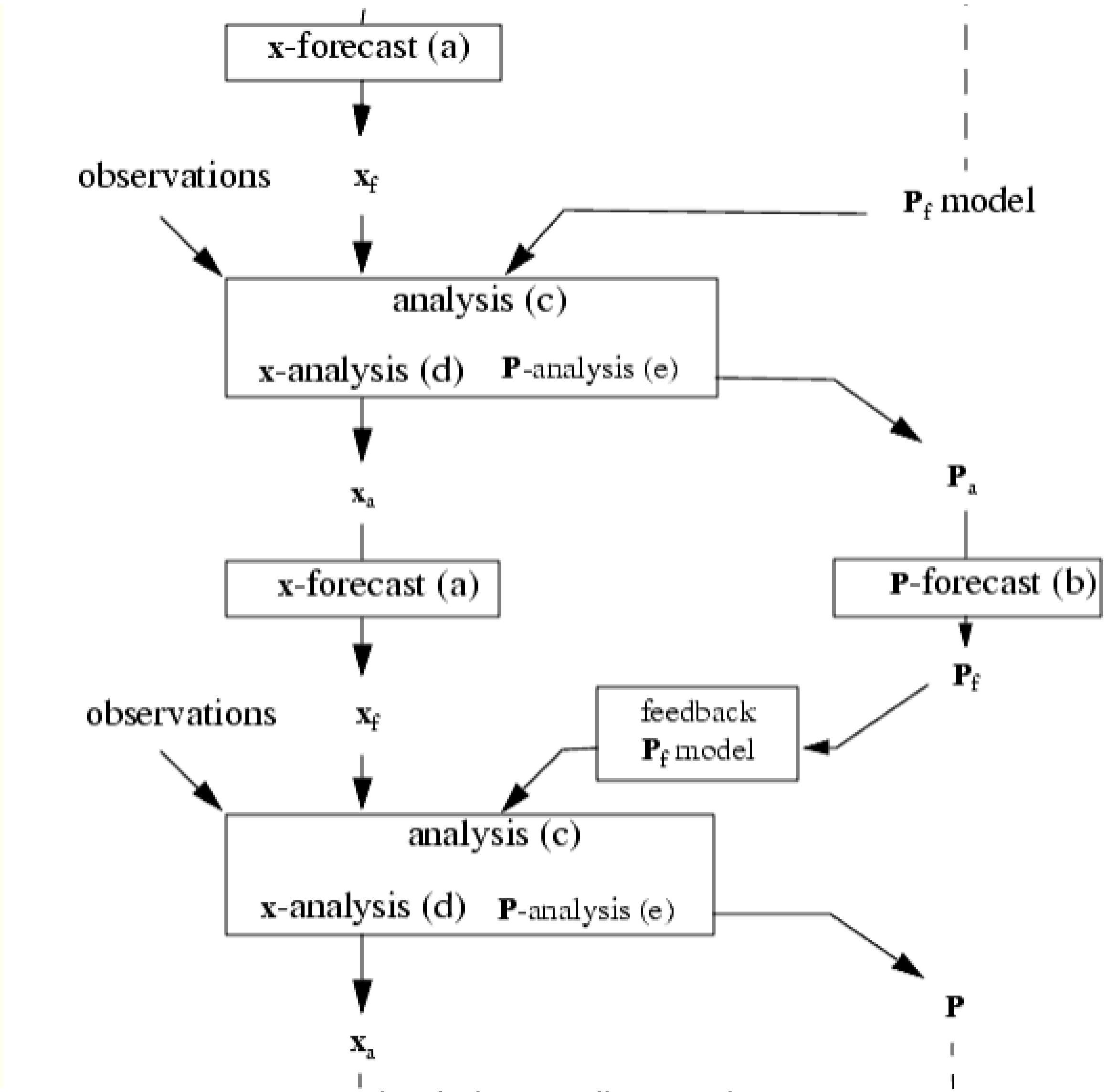
which shows that the analysis error variance is always smaller than both the background and observation error variances, and it is smallest if both are equal, in which case the analysis error variance is half the background error variance.

Error analysis 2

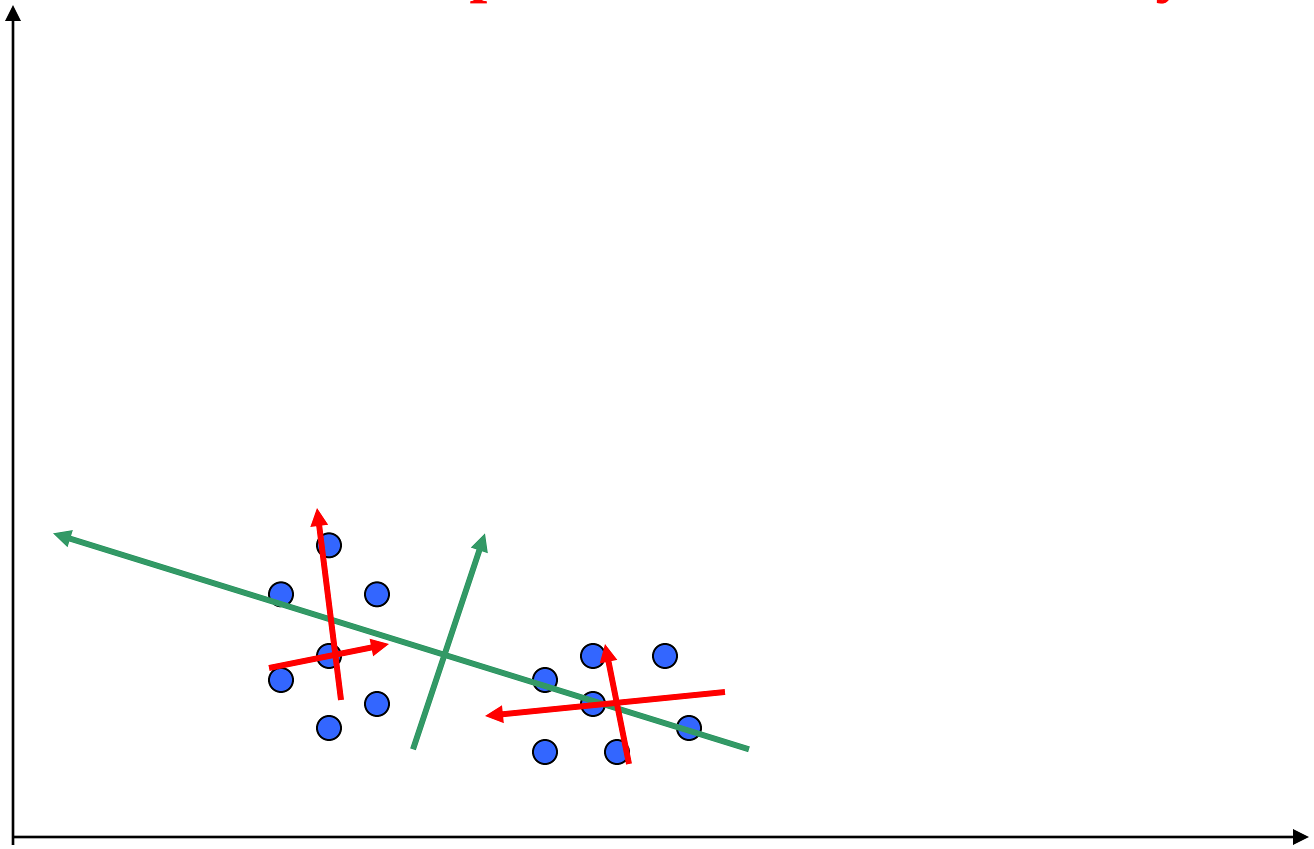
- In the limiting case of a very low quality measurement ($\sigma_o >> \sigma_b$), $k = 0$ and the analysis remains equal to the background.
- On the other hand, if the observation has a very high quality ($\sigma_o >> \sigma_b$), $k = 1$ and the analysis is equal to the observation.
- If both have the same accuracy, $\sigma_o = \sigma_b$, $k = 1/2$ and the analysis is simply the average of T_o and T_b , which reflects the fact that we trust as much the observation as the background, so we make a compromise.
- In all cases, $0 \le k \le 1$, which means that the analysis is a weighted average of the background and the observation.


σb
estimation error σa=f(k)
σa
σo
Xa
σb
σa
σo
Xa
σo
Xa
σa
weigh
0
1/2
1
k
background
observation

The analysis/forecast cycle of 4-D assimilation scheme



Relationship between EOF and Fuzzy sets



Enhancement of correlation within fuzzy clusters:

The correlation between the variable couples:

air temperature, humidity, pressure and humidity

0.4-0.6 - without clussification

0.8-0.9 - within fuzzy clusters:

Motivation for this study:

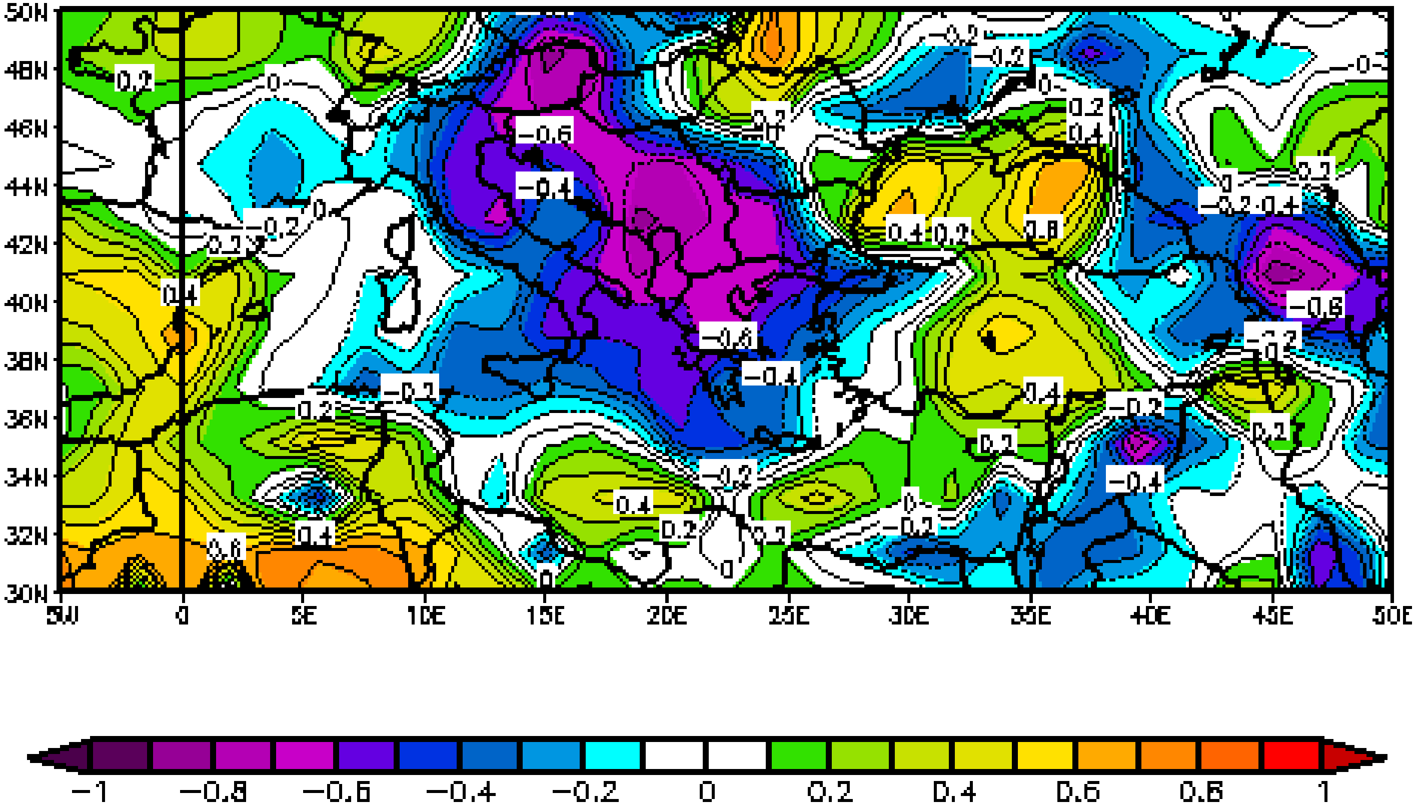
1. High spatial variability of precipitation field – it is hard to describe it by means of differential equation with a low resolution
2. Precipitations follow to changes in atmospheric circulation regimes - rapid rain rate changes are occurred after transition from one regime to another
3. There is no considerable rain rate correlation to other meteorological variables, but there are feedback linkages to be revealed

Data sets:

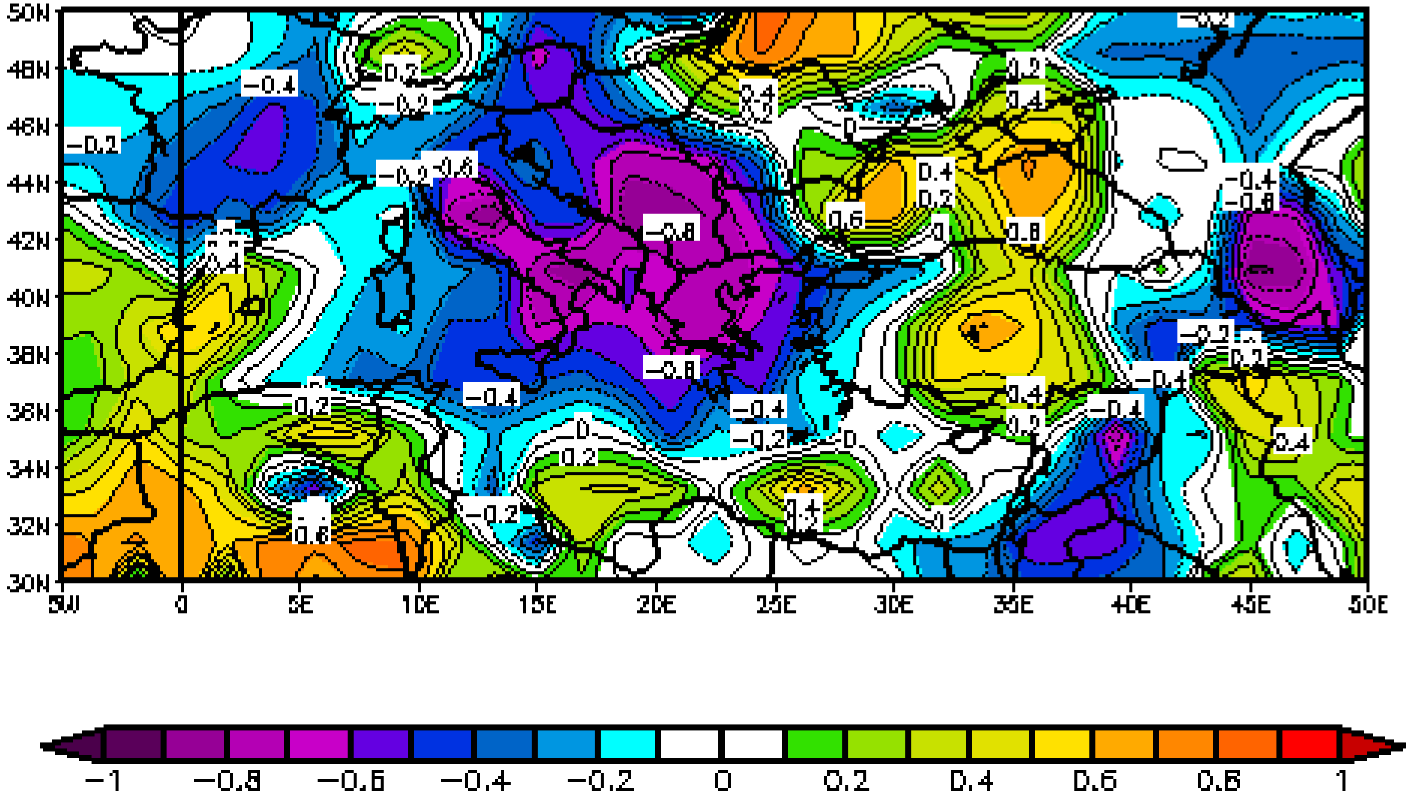
NCEP- NCAR Reanalysis data:

Daily and Monthly

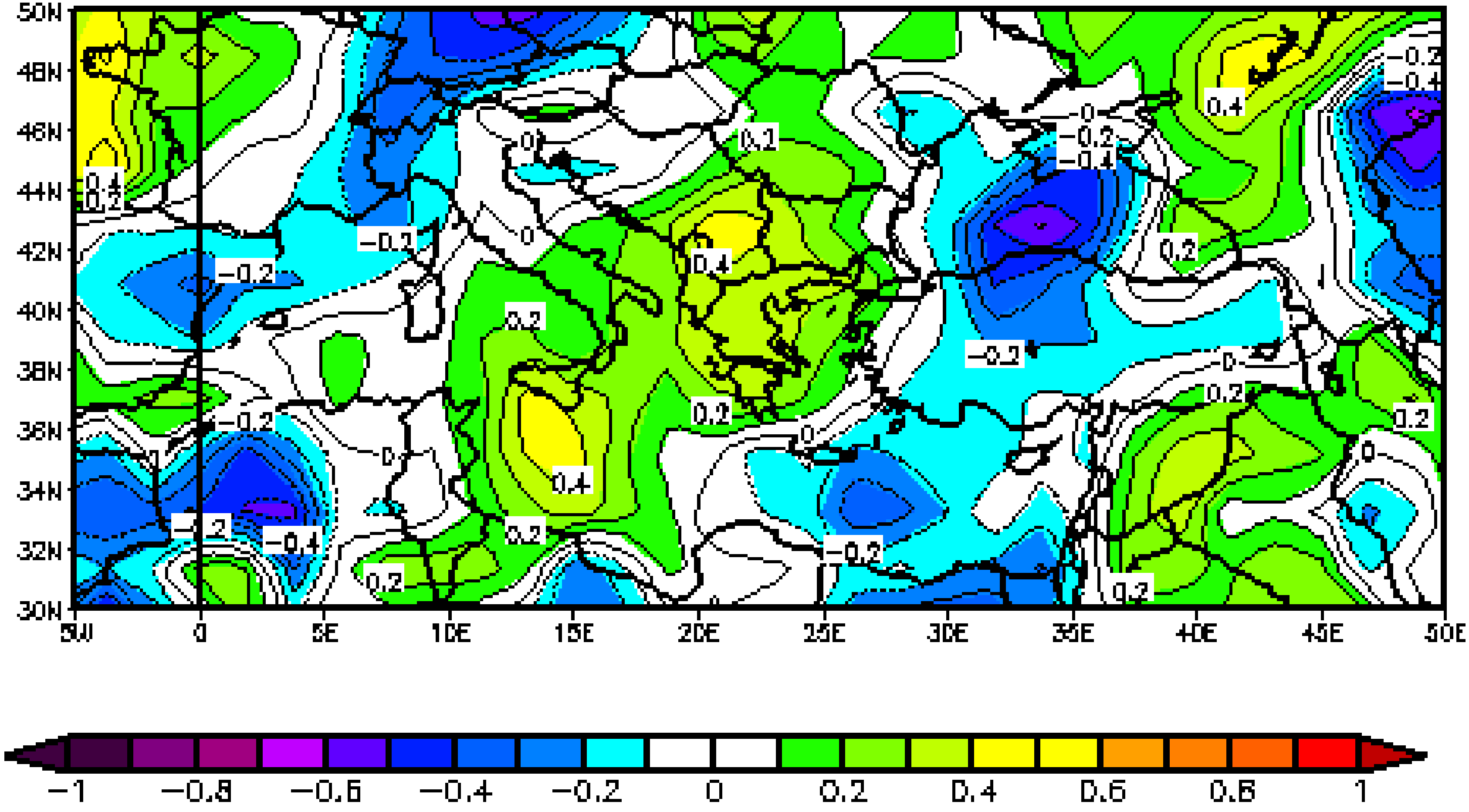
Seasonal correlation of **April-March (1995-2005)** precipitation rate with February-March **NAO** (index leads by 2 months) in Mediterranean area



Seasonal correlation of **April-March (1995-2005)** **precipitation rate** with February-March **AO** (index leads by 2 months) in Mediterranean area

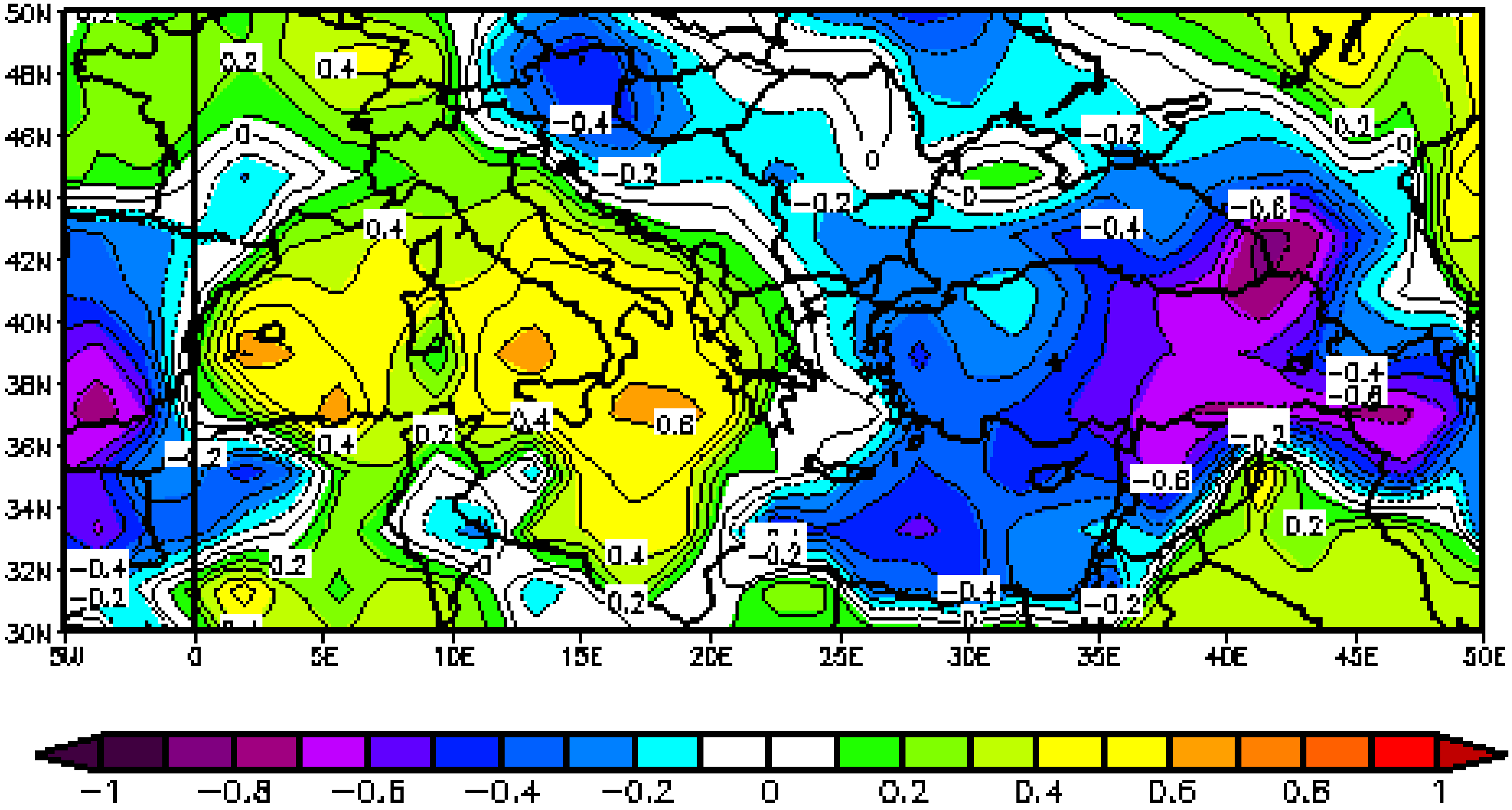


- Seasonal correlation of **April-May (1965-1975) precipitation rate** with February-March **NAO** (index leads by 2 months) in Mediterranean area



Circulation to Climate Change

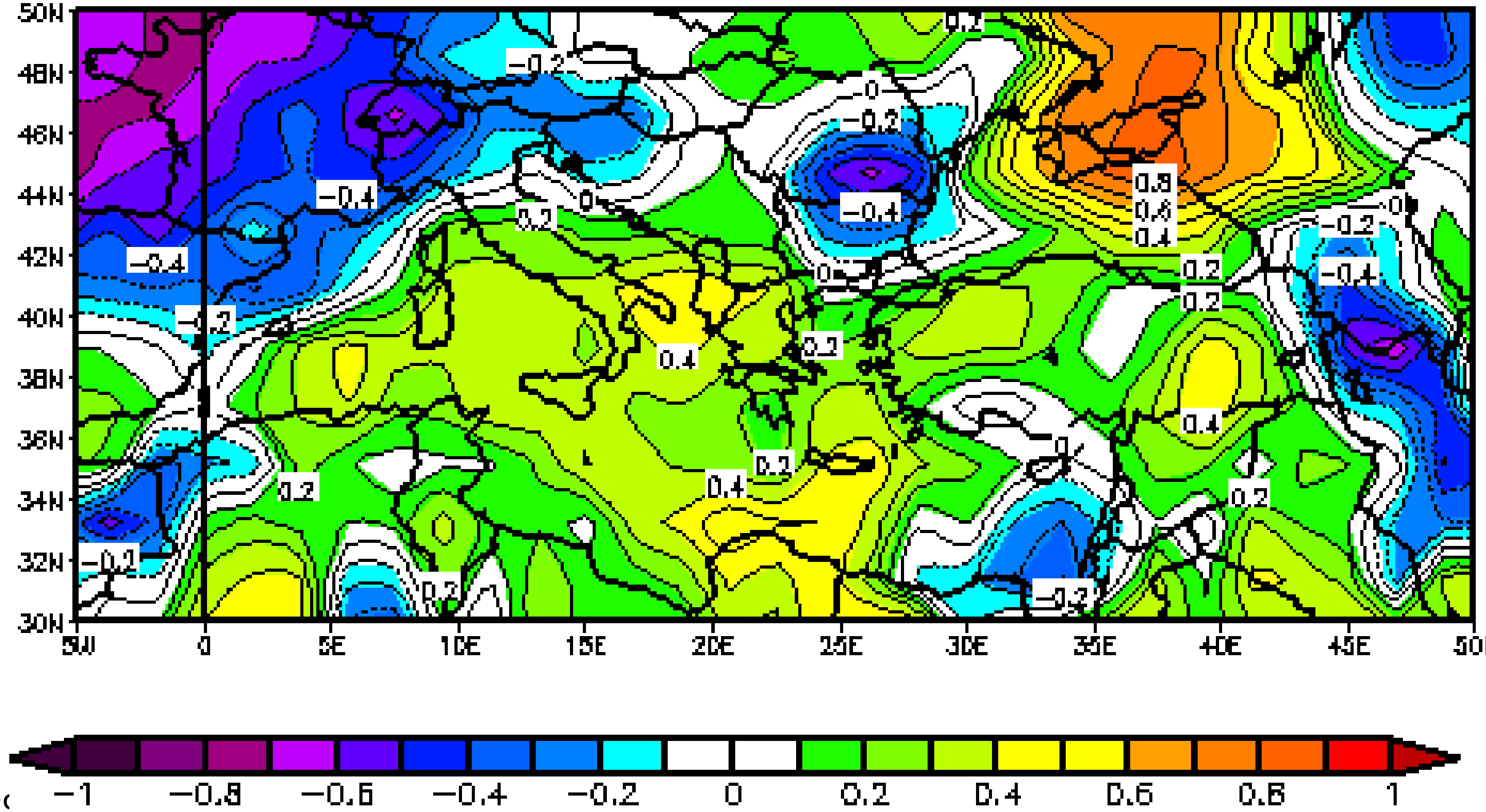
Seasonal correlation of **June-July (1995-2005)** **precipitation rate** with **April-May NAO** (index leads by 2 months) in Mediterranean area



29-30 July 2014

Remote Sensing for Global Water Circulation to Climate Change

Seasonal correlation of **September-October (1995-2005) precipitation rate** with **July-August NAO** (index leads by 2 months) in Mediterranean area



FUZZY ALGORITHM

- Clustering analysis is a fundamental but important tool in statistical data analysis. In the past, the clustering techniques have been widely applied in interdisciplinary scientific areas such as pattern recognition, information retrieval, clinical diagnosis, and microbiological analysis. In the literature, the k-means is a typical clustering algorithm, which partitions the input data set that generally forms k* true clusters into k categories (also simply called clusters without further distinction) with each represented by its center (Pokrovsky *et al*., 2002)).

References on Fuzzy sets

- S. Wu, M. J. Er, and Y. Gao, “A fast approach for automatic generation
- of fuzzy rules by generalized dynamic fuzzy neural networks,” *IEEE*
- *Trans. Fuzzy Syst.*, vol. 9, pp. 578–594, Aug. 2001.
- J. Yen, “Fuzzy logic—a modern perspective,” *IEEE Trans. Knowledge*
- *Data Eng.*, vol. 11, pp. 153–165, Jan. 1999.
- J. Yen and L. Wang, “Application of statistical information criteria for
- optimal fuzzy model construction,” *IEEE Trans. Fuzzy Syst.*, vol. 6, pp.
- 362–372, June 1998.
- H. Ying, “General SISO Takagi-Sugeno fuzzy systems with linear rule
- consequent are universal approximators,” *IEEE Trans. Fuzzy Syst.*, vol.
- 6, pp. 582–587, Aug. 1998.
- L. A. Zadeh, “Fuzzy sets,” *Inform. Control*, vol. 8, pp. 338–353, 1965.
- H.-J. Zimmermann, *Fuzzy Set Theory and its Applications*. Norwell,
- MA: Kluwer, 1991.

Introduction

- Fuzzy Logic was initiated in 1965, by Dr. Lotfi A. Zadeh, professor for computer science at the university of California in Berkley.

- Basically, Fuzzy Logic is a multivalued logic, that allows intermediate values to be defined between conventional evaluations like true/false, yes/no, high/low, etc.

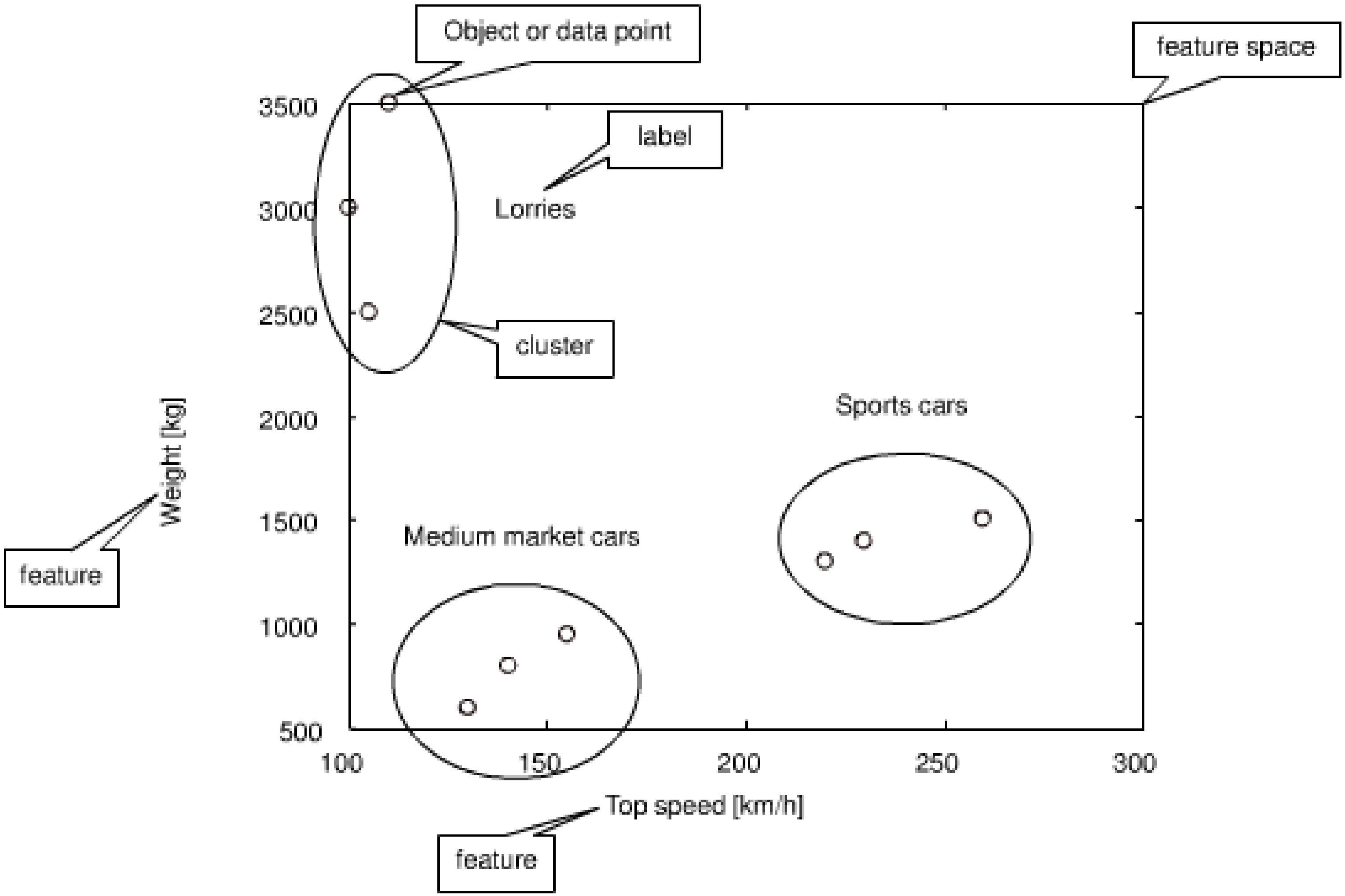
- Fuzzy Logic starts with and builds on a set of user–supplied human language rules.

- Fuzzy Systems convert these rules to their mathematical equivalents.

- This simplifies the job of the system designer and the computer, and results in much more accurate representations of the way system behaves in real world.

- Fuzzy Logic provides a simple way to arrive at a definite conclusion based upon vague, ambiguous, imprecise, noisy, or missing input information.

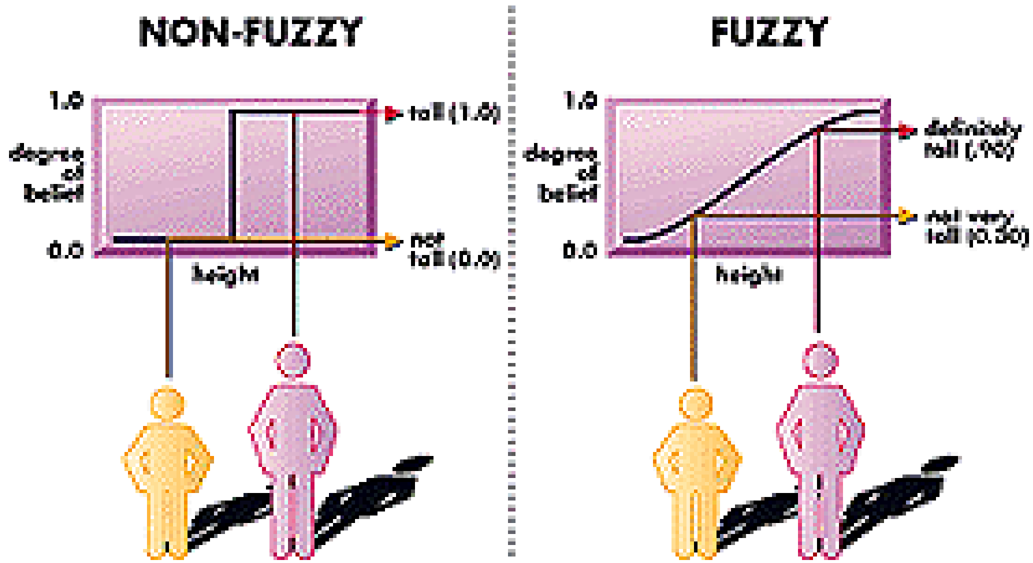
Terminology



Fuzzy Logic

➢ What is Fuzzy Logic?

Fuzzy Logic is a superset of conventional (Boolean) logic that has been extended to handle the concept of partial truth, i.e. truth values between “completely true” and “completely false”.



Fuzzy Logic

➢ How Fuzzy Logic works?

- In Fuzzy Logic, unlike standard conditional logic, the truth of any statement is a matter of degree. (e.g How cold is it? How high shall we set the heat?)

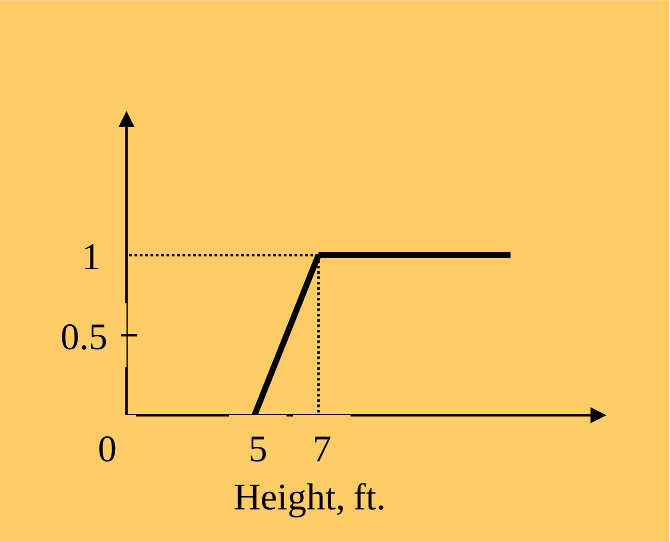
- The degree to which any Fuzzy statement is true is denoted by a value between 0 and 1.

- Fuzzy Logic needs to be able to manipulate degrees of "may be" in addition to true and false.

➢ Example:

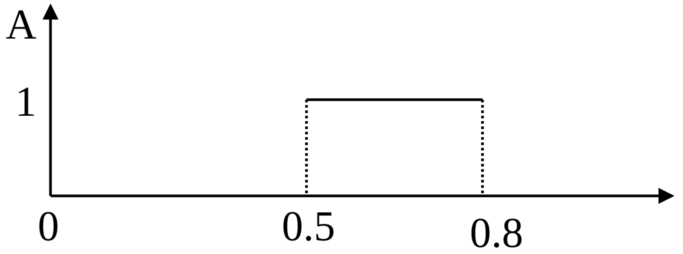
tall(x) = {
0, if height(x) < 5 ft.,
(height(x)-5ft.)/2ft., if 5 ft. <= height (x) <= 7 ft.,
1, if height(x) > 7 ft.
}

U: universe of discourse (i.e. set of people)
TALL: Fuzzy Subset



Fuzzy Sets

➢ In classical mathematics we are familiar with what we call *crisp* sets. In this method, the characteristic function assigns a number 1 or 0 to each element in the set, depending on whether the element is in the subset A or not.

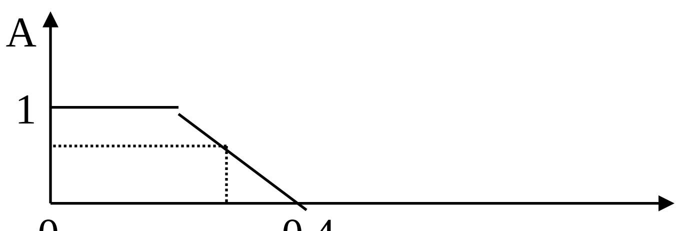


1 → In set A

0 → Not in set A

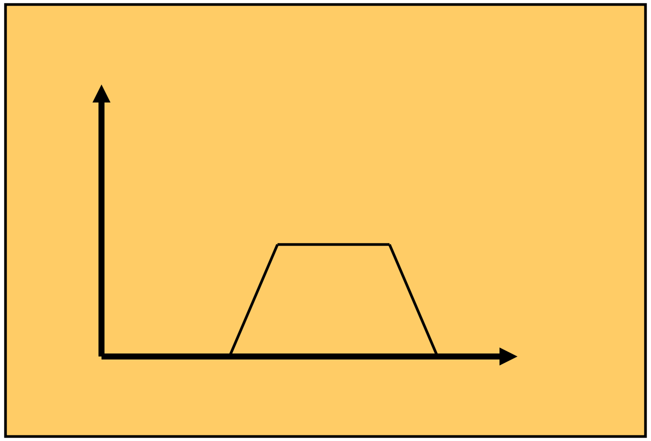
➢ This concept is sufficient for many areas of application, but it lacks flexibility for some applications like classification of remotely sensed data analysis.

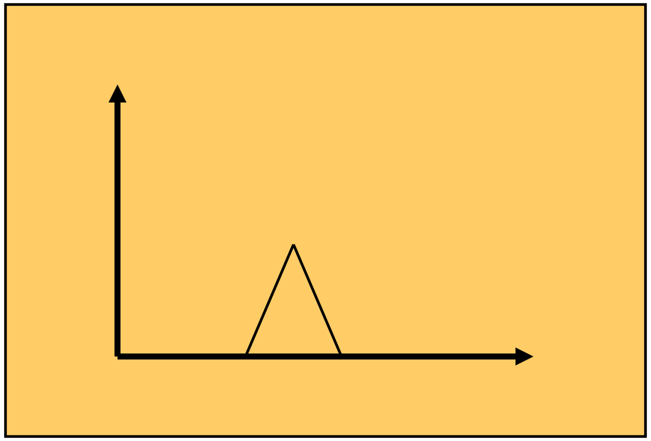
➢ The membership function is a graphical representation of the magnitude of participation of each input. It associates weighting with each of the inputs that are processed.

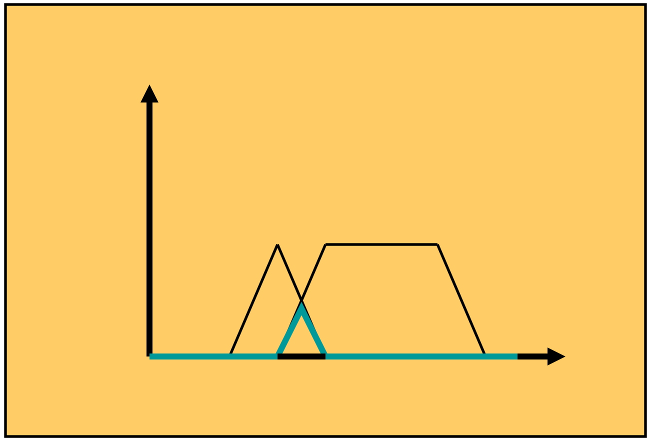


Operations on Fuzzy Sets

Fuzzy *AND*:

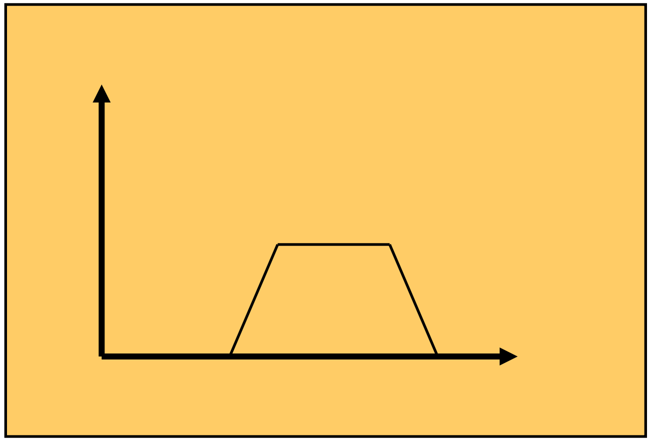


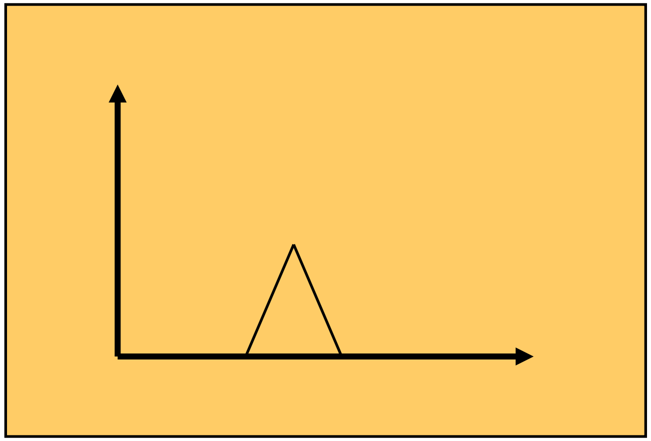


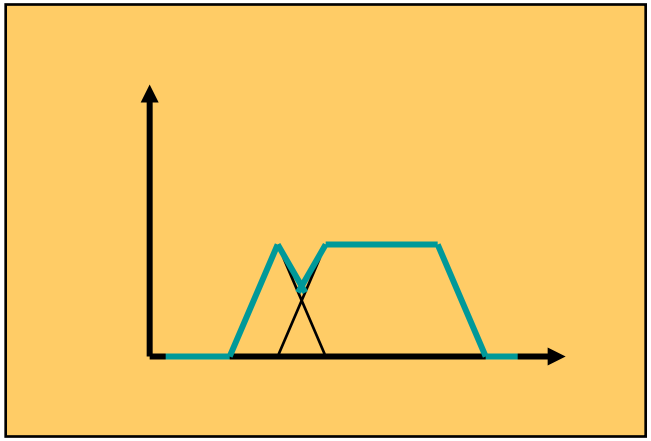


Operations on Fuzzy Sets (contd.)

Fuzzy *OR*:

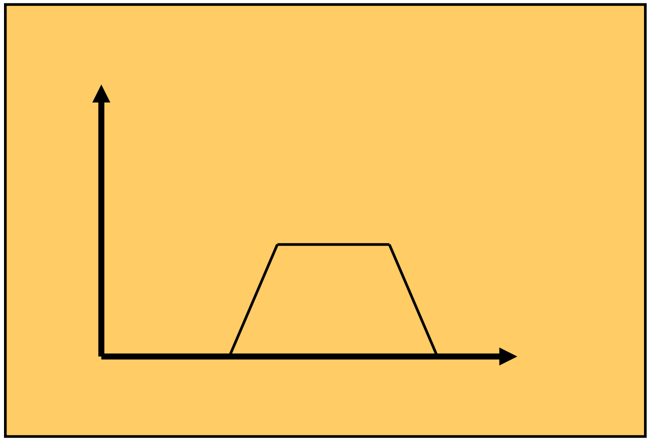


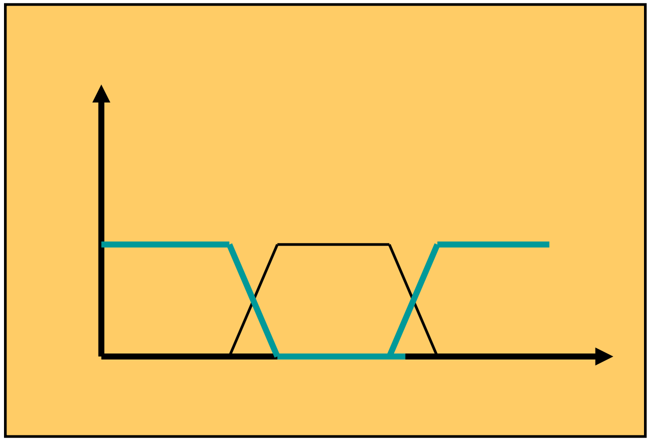




Operations on Fuzzy Sets (contd.)

Fuzzy *NOT*:





Probability Vs Fuzzy Logic

Probability	Fuzzy Logic
Probability Measure	Membership Function
Before an event happens	After it happened
Measure Theory	Set Theory
Domain is 2U (Boolean Algebra)	Domain is [0,1]U (Cannot be a Boolean Algebra)

Properties

The following rules which are common in classical set theory also apply to Fuzzy Logic.

- De Morgan's Law:

$$\overline{(A \cap B)} = \overline{A} \cap \overline{B} \qquad \overline{(A \cup B)} = \overline{A} \cap \overline{B}$$

- Associativity:

$$(A \cap B) \cap C = A \cap (B \cap C)$$
$$(A \cup B) \cup C = A \cup (B \cup C)$$

- Commutativity:

$$A \cap B = B \cap A \,,\; A \cup B = B \cup A$$

- Distributivity:

$$A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$$
$$A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$$

Fuzzy Systems

Conclusions

- Fuzzy Logic provides a different way to approach a control or classification problem. This method focuses on what the system should do rather than trying to model how it works.

- Fuzzy approach requires a sufficient expert knowledge for the formulation of the rule base, the combination of the sets and the defuzzification.

- Fuzzy Logic might be helpful, for very complex processes, when there is no simple mathematical model.

Pros & Cons

- Advantages:

  - Helpful for very complex or highly nonlinear processes.
  - Allows use of "fuzzy" concepts like medium, low, etc.
  - Biggest impact is for control problems.
  - Help avoid discontinuities in behavior.

- Disadvantages:

  - Sometimes results are unexpected and hard to debug.
  - Computationally complicated.
  - According to literature, Fuzzy Logic is not recommendable, if conventional approach yields a satisfying result.

Fuzzy System Applications

1. Pattern Recognition and Classification
2. Fuzzy Clustering
3. Image and Speech Processing
4. Fuzzy Systems for Predictions
5. Fuzzy Control
6. Monitoring
7. Diagnosis
8. Optimization and Decision Making
9. Group Decision Making

K-means Algorithm

Although the k-means technique has been widely used due to its easy implementation, it has **two major drawbacks**:
(1) It implies that the data clusters are spherical because it performs clustering based on the Euclidean distance only;
(2) It needs to pre-assign the number, k, of clusters. Many experiments have shown that the k-means algorithm can work well when k is equal to k*. However, in many practical cases, it is impossible to know the exact cluster number in advance. Under the circumstances, the k-means algorithm often leads to a poor clustering performance.

Clustering based on k-means is closely related to a number of other clustering and location problems. A **k-means algorithm** is measured by two criteria: intra-cluster criterion and inter-cluster criterion. These include the Euclidean k-medians, in which the objective is to minimize the sum of distances to the nearest center, and the geometric k-center problem, in which the objective is to minimize the maximum distance from every point to its closest center. K-means is the most popular iterative centroid-based divisive algorithm. The specific fuzzy classification algorithm considered herein is now recalled and briefly discussed (Matoušek, 2000).

In such algorithms the definition of centroid will be used extensively; specifically, the **centroid** of M, say *w* , is given by

$$w = \frac{1}{N} \sum_{i=1}^{N} x_i \qquad (10.1)$$

where X_i is the i-th column of matrix **X**. Similarly, the centroids of the sub-clusters X_l and X_r , say w_l and w_r , are given by:

$$w_l = \frac{1}{N_l} \sum_{i=1}^{N_l} X_{l,i} \qquad w_r = \sum_{i=1}^{N_r} X_{r,i} \qquad (10.2)$$

where $X_{l,i}$ and , $X_{r,i}$ are the i-th columns of X_l and , X_r respectively

k-means algorithm:

Step 1. (Initialization). Randomly select a point, say $c_l \in R^p$; then compute the centroid w of M (see Eq. (10.1)), and compute $c_r = w - (c_l - w)$

Step 2. Divide a set M=$\{x_1, x_2, ..., x_N\}$ into two sub-clusters M_l and M_r, according to the following rule:

$$\begin{cases} x_i \in M_l, if \|x_i - c_l\| \leq \|x_i - c_r\| \\ x_i \in M_r, if \|x_i - c_r\| \leq \|x_i - c_l\| \end{cases}$$

Step 3. Compute the centroids of M_l and M_r: w_l and w_r, as in Eq. (10.2).

Step 4. If $w_l = c_l$ and $w_r = c_r$, stop, else, let $c_l = w_l$; $c_r = w_r$ and go to Step 2.

Fuzzy membership matrix **M**

Point *k*'s membership of cluster *i*

Fuzziness exponent

$$m_{ik} = \frac{1}{\sum_{j=1}^{c} \left(\frac{d_{ik}}{d_{jk}} \right)^{2/(q-1)}}$$

Distance from point *k* to current cluster centre *i*

Distance from point *k* to other cluster centres *j*

$$d_{ik} = \left\| \mathbf{u}_k - \mathbf{c}_i \right\|$$

Fuzzy membership matrix **M**

$$
\begin{aligned}
m_{ik} &= \frac{1}{\sum_{j=1}^{c} \left(\frac{d_{ik}}{d_{jk}} \right)^{2/(q-1)}} \\
&= \frac{1}{\left(\frac{d_{ik}}{d_{1k}} \right)^{2/(q-1)} + \left(\frac{d_{ik}}{d_{2k}} \right)^{2/(q-1)} + \cdots + \left(\frac{d_{ik}}{d_{ck}} \right)^{2/(q-1)}} \\
&= \frac{\frac{1}{d_{ik}^{\;2/(q-1)}}}{\frac{1}{d_{1k}^{\;2/(q-1)}} + \frac{1}{d_{2k}^{\;2/(q-1)}} + \cdots + \frac{1}{d_{ck}^{\;2/(q-1)}}}
\end{aligned}
$$

Gravitation to cluster *i* relative to total gravitation

Fuzzy c-partition

All clusters *C* together fill the whole universe *U*.
Remark: The sum of memberships for a data point is 1, and the total for all points is K

$$\bigcup_{i=1}^{c} C_i = U$$

Not valid: Clusters do overlap

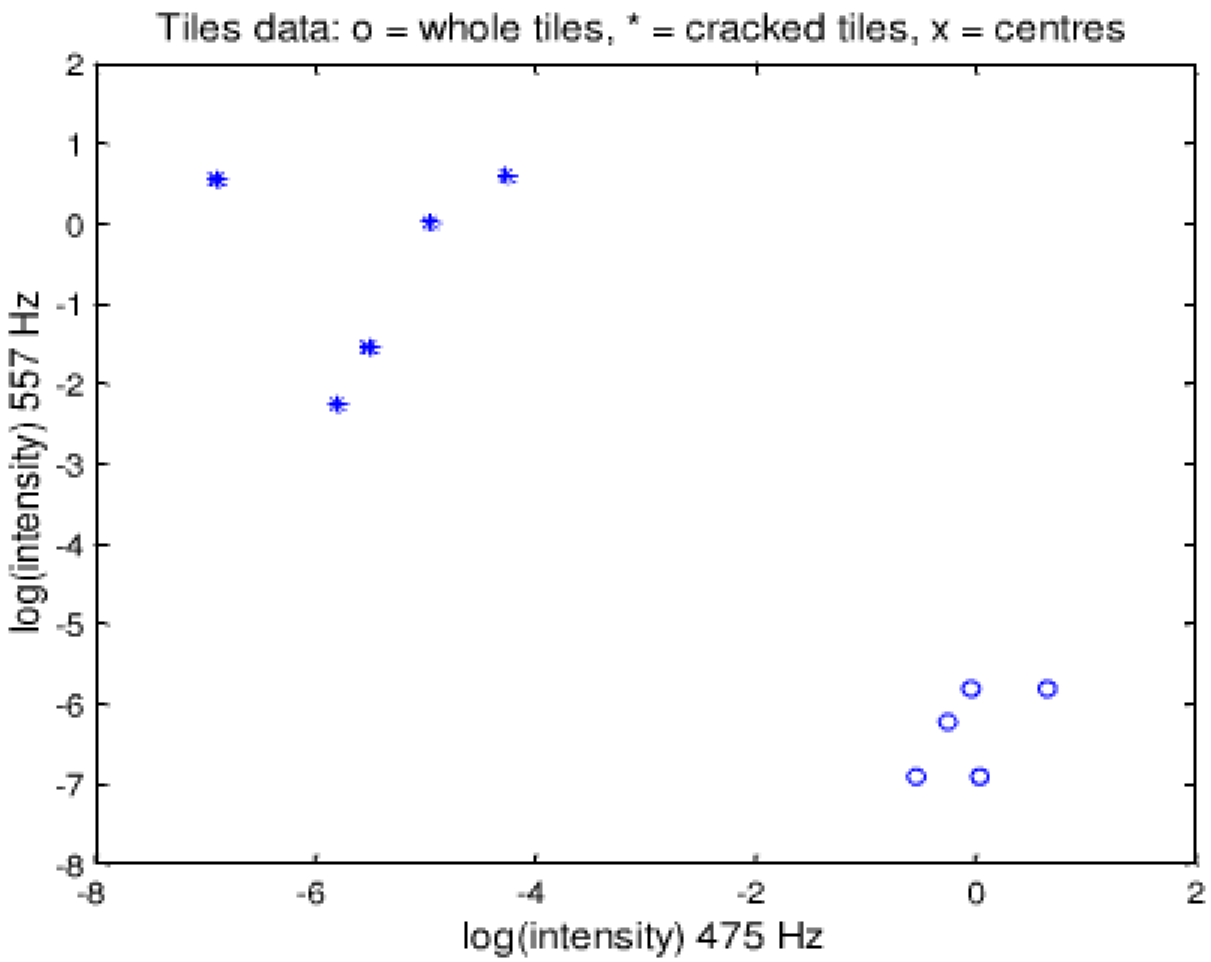
$$C_i \cap C_j = \emptyset \quad \text{for all } i \neq j$$

A cluster *C* is never empty and it is smaller than the whole universe *U*

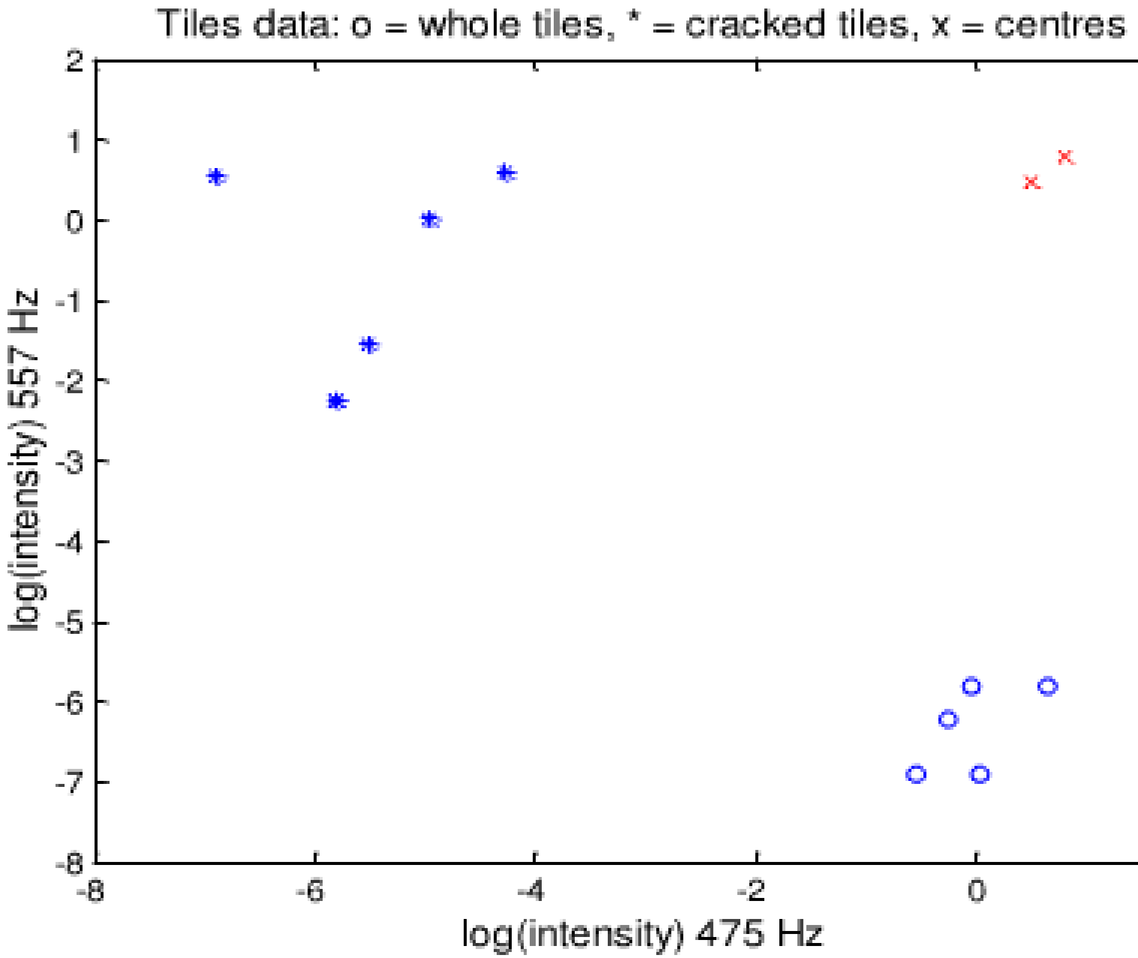
$$\emptyset \subset C_i \subset U \quad \text{for all } i$$

$$2 \leq c \leq K$$

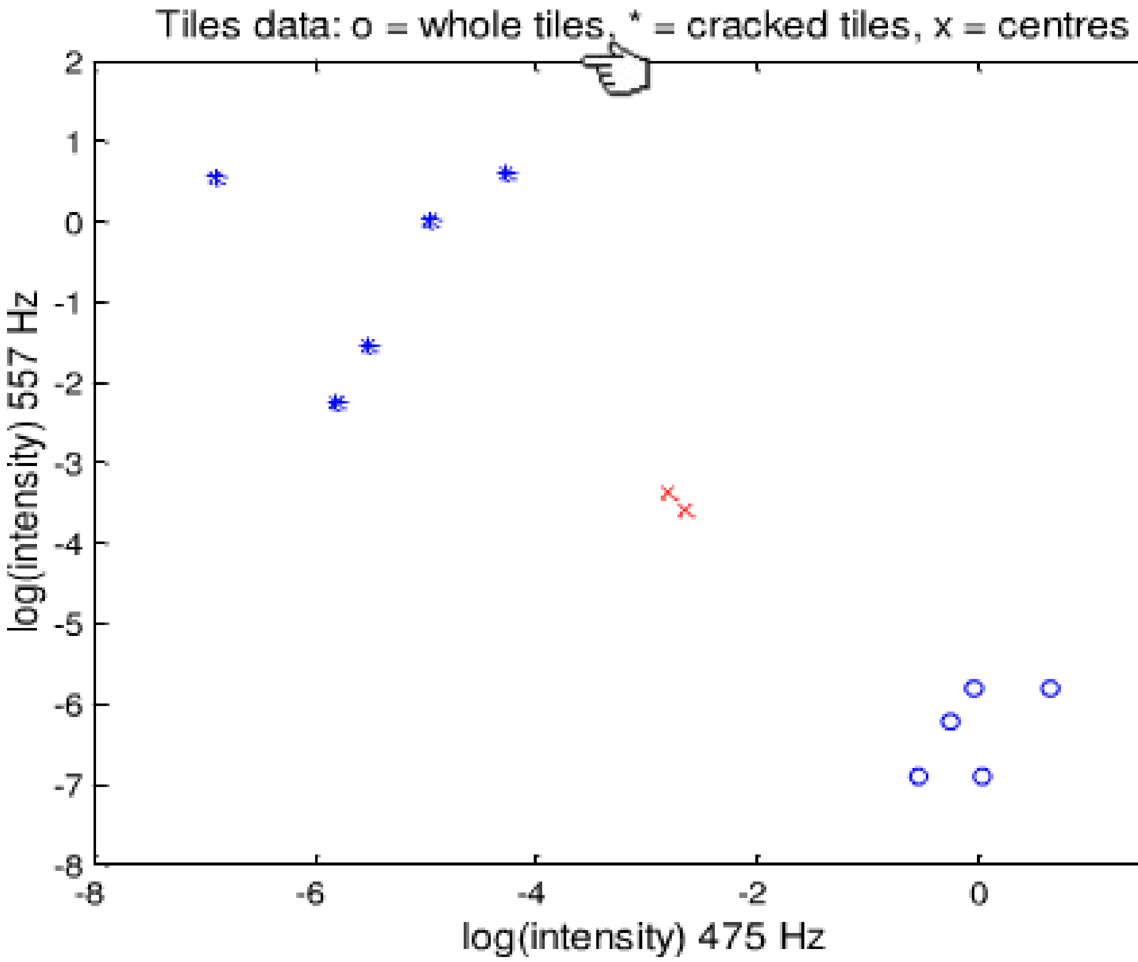
There must be at least 2 clusters in a c-partition and at most as many as the number of data points *K*



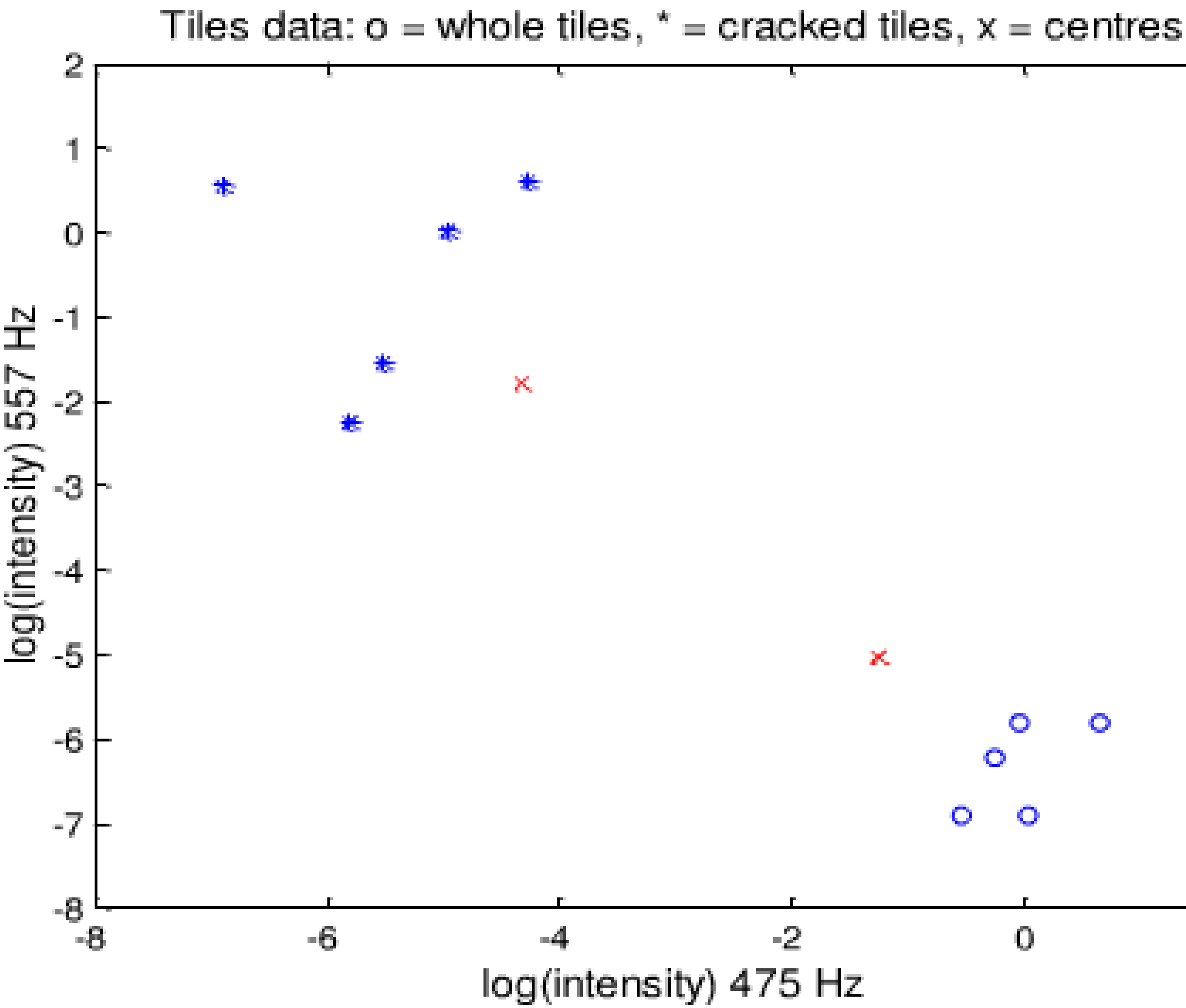
Plot of tiles by frequencies (logarithms). The whole tiles (o) seem well separated from the cracked tiles (*). The **objective** is to find the two clusters.



1. Place two cluster centres (x) at random.
2. Assign each data point (* and o) to the nearest cluster centre (x)

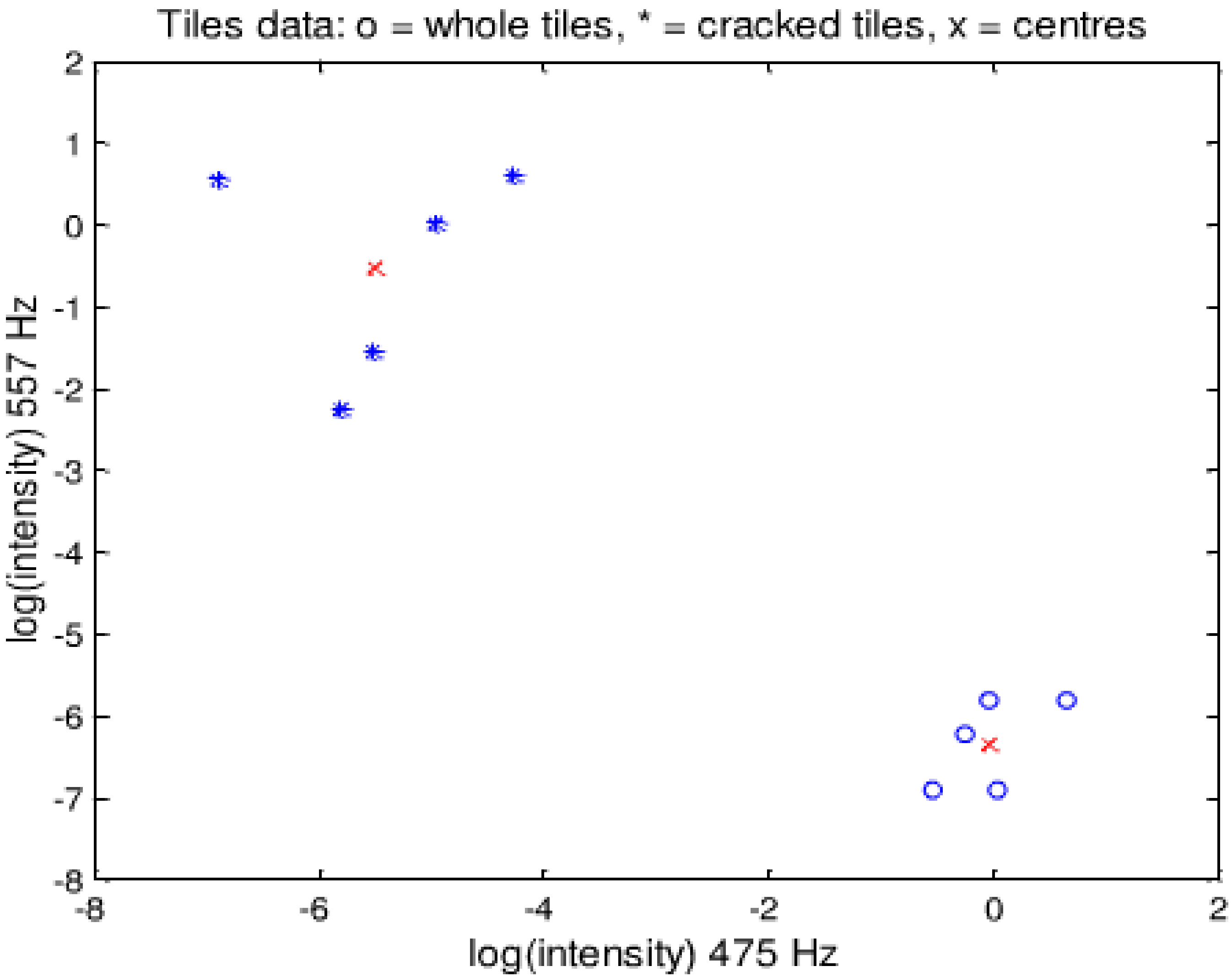


1. Compute the new centre of each class
2. Move the crosses (x)

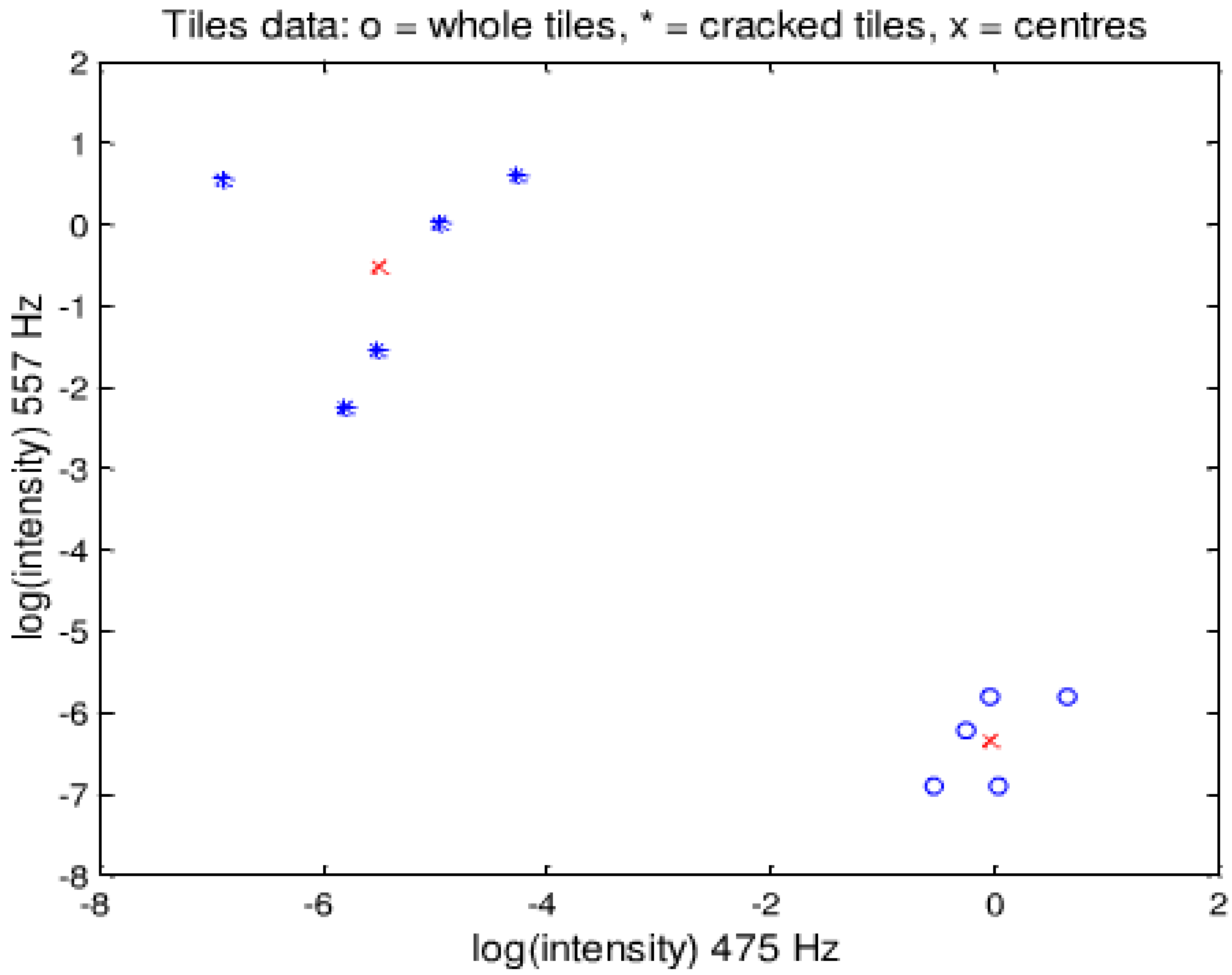


Tiles data: o = whole tiles, * = cracked tiles, x = centres
2
1
0
-1
-2
-3
-4
-5
-6
-7
-8
log(intensity) 557 Hz
-8
-6
-4
-2
0
2
log(intensity) 475 Hz

Iteration 2


Tiles data: o = whole tiles, * = cracked tiles, x = centres
log(intensity) 557 Hz
log(intensity) 475 Hz
2
1
0
-1
-2
-3
-4
-5
-6
-7
-8
-8
-6
-4
-2
0
2

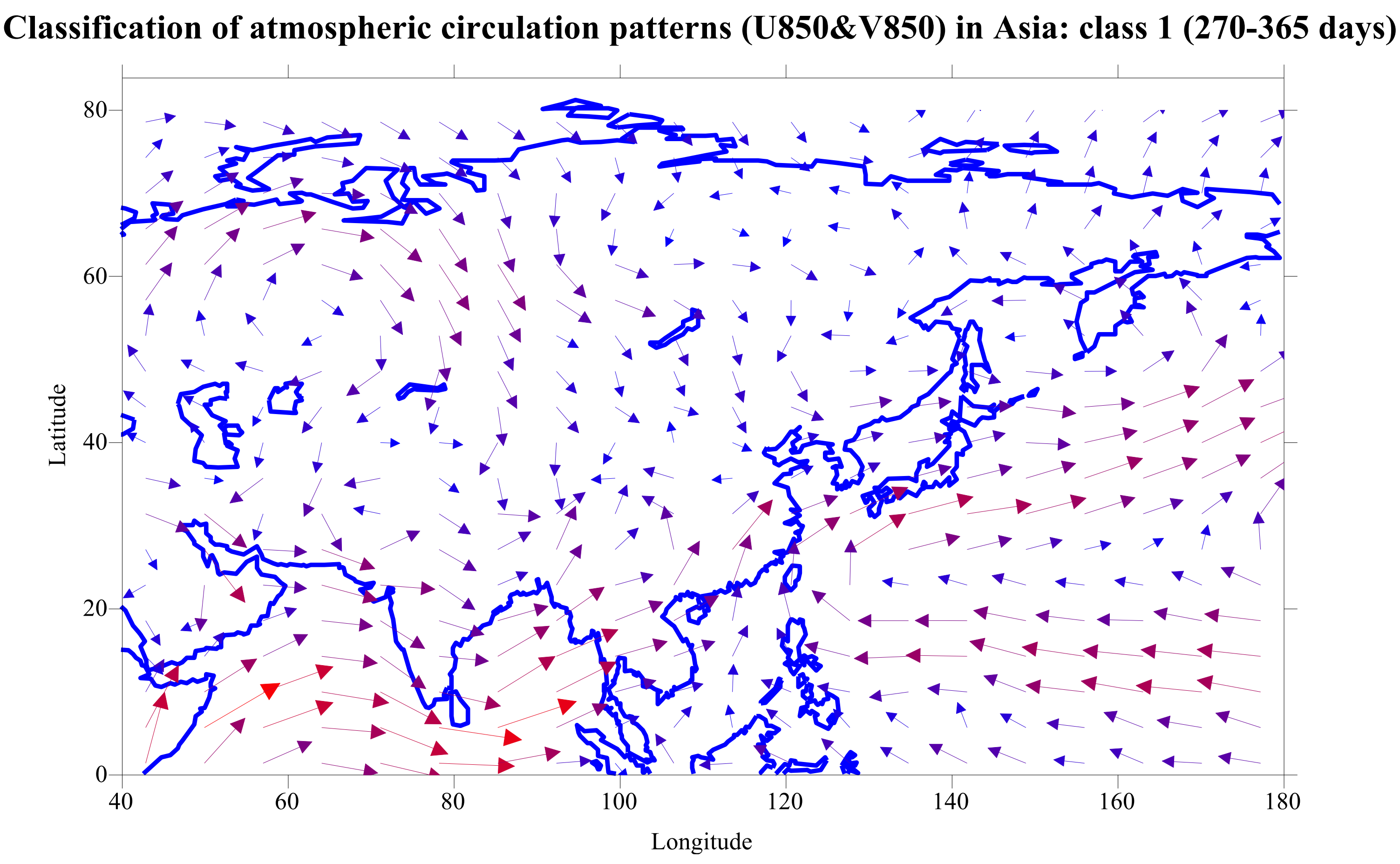
Iteration 3



Iteration 4 (then stop, because no visible change)
Each data point belongs to the cluster defined by the nearest centre

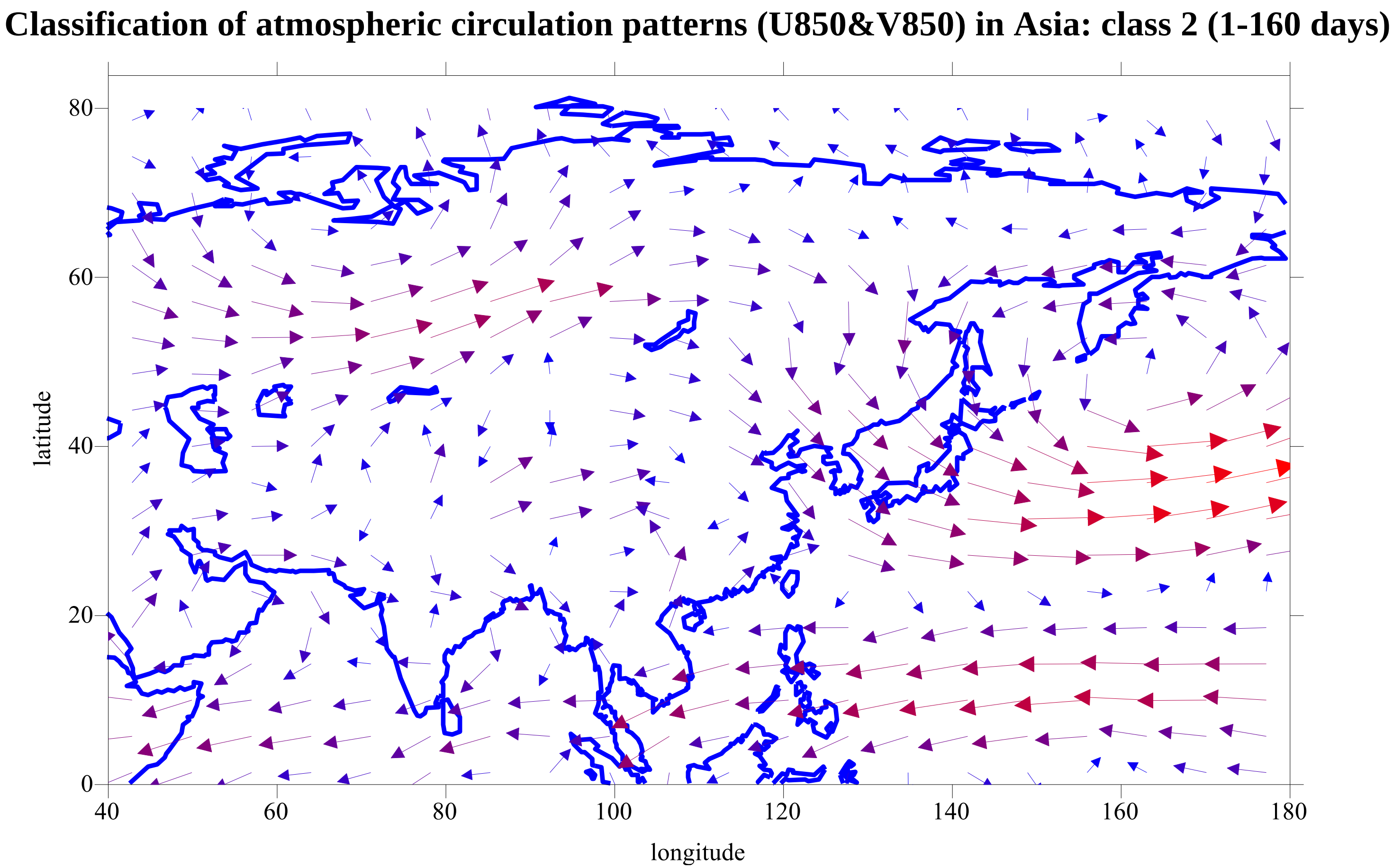
Fuzzy Classification of atmospheric circulation regimes in Asia

1991



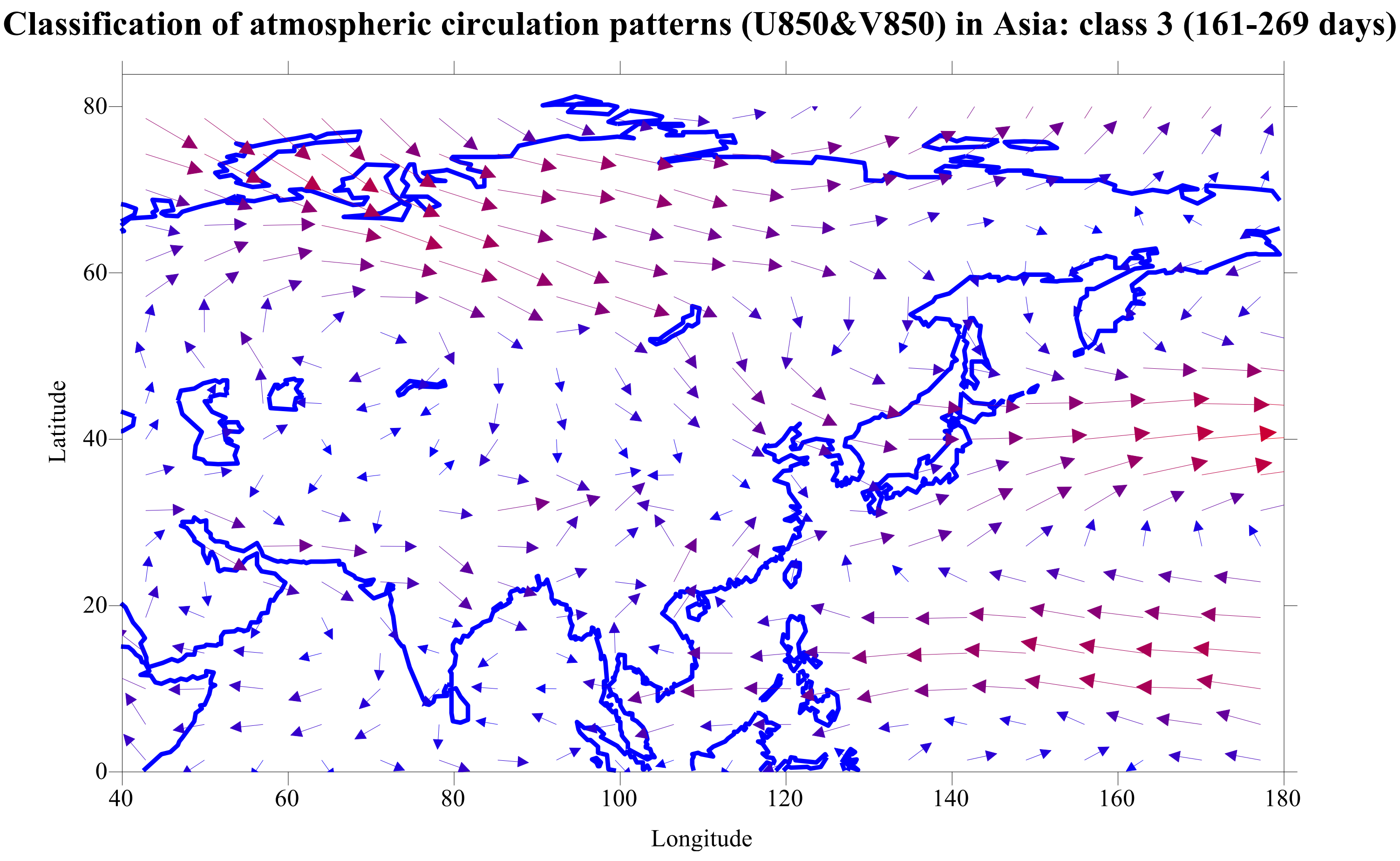
Classification of atmospheric circulation patterns (U850&V850) in Asia: class 1 (270-365 days)
80
60
40
20
0
Latitude
40
60
80
100
120
140
160
180
Longitude

1991

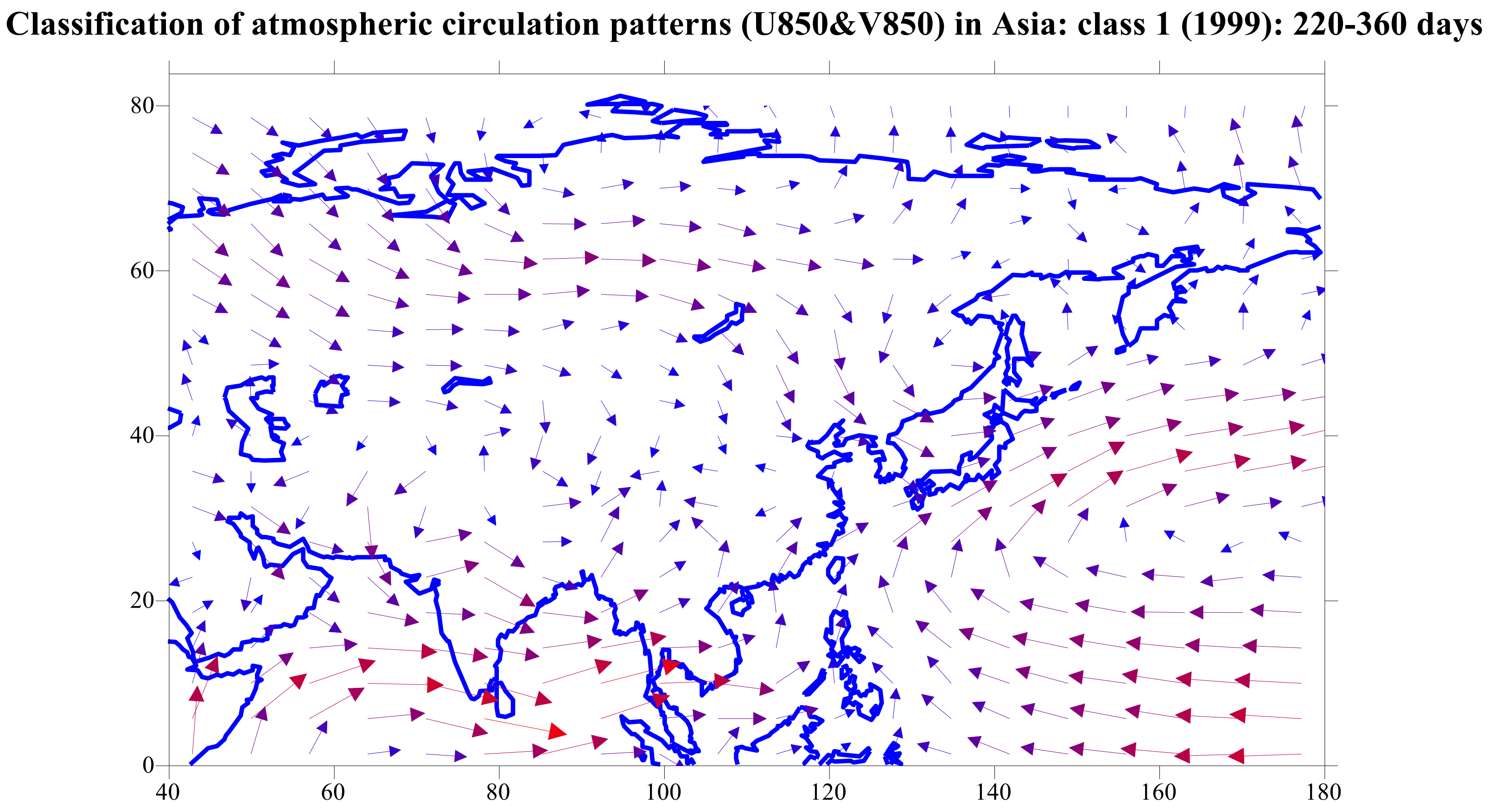


Classification of atmospheric circulation patterns (U850&V850) in Asia: class 2 (1-160 days)
80
60
40
20
0
latitude
40
60
80
100
120
140
160
180
longitude

1991

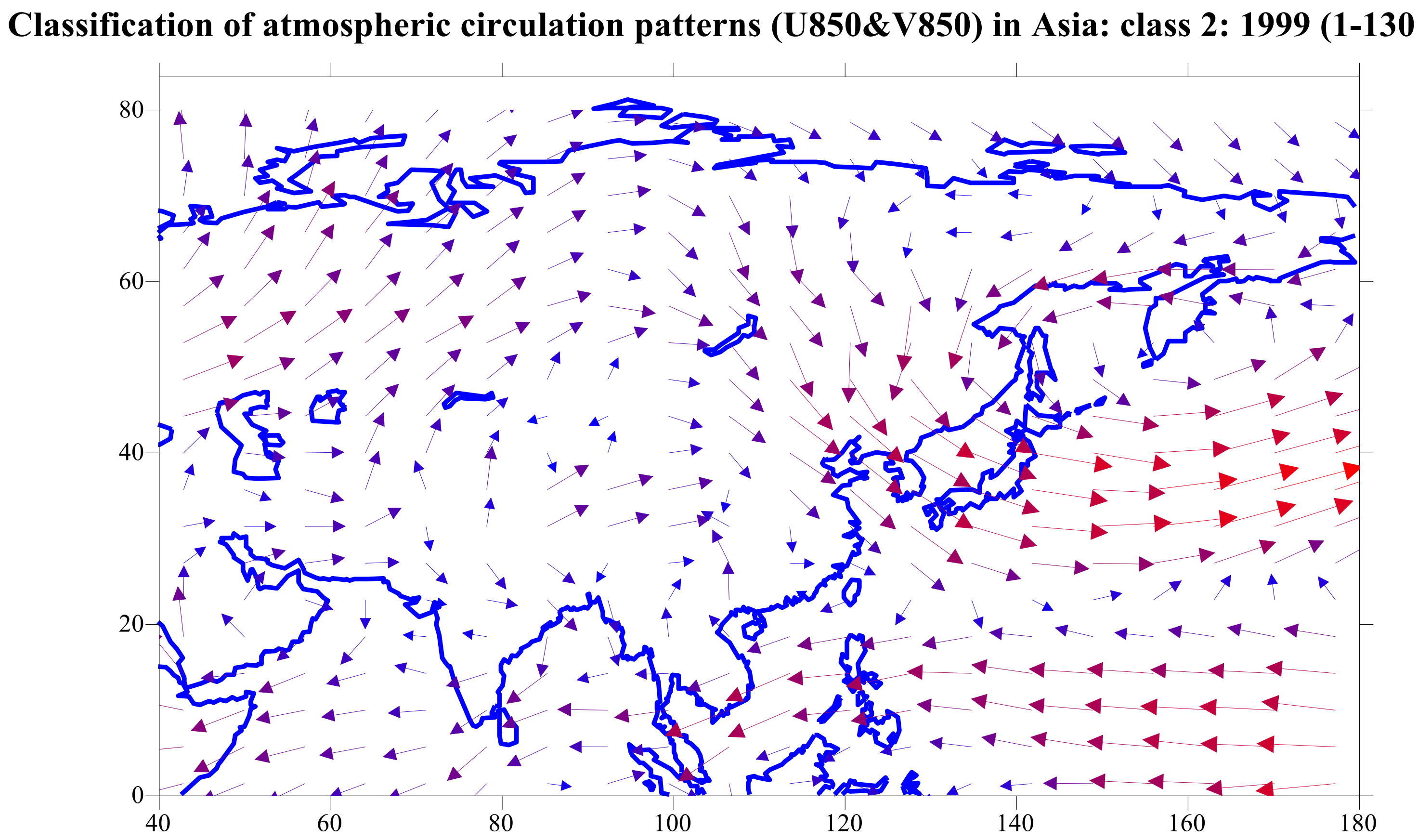

Classification of atmospheric circulation patterns (U850&V850) in Asia: class 3 (161-269 days)
80
60
40
20
0
Latitude
40
60
80
100
120
140
160
180
Longitude

1999



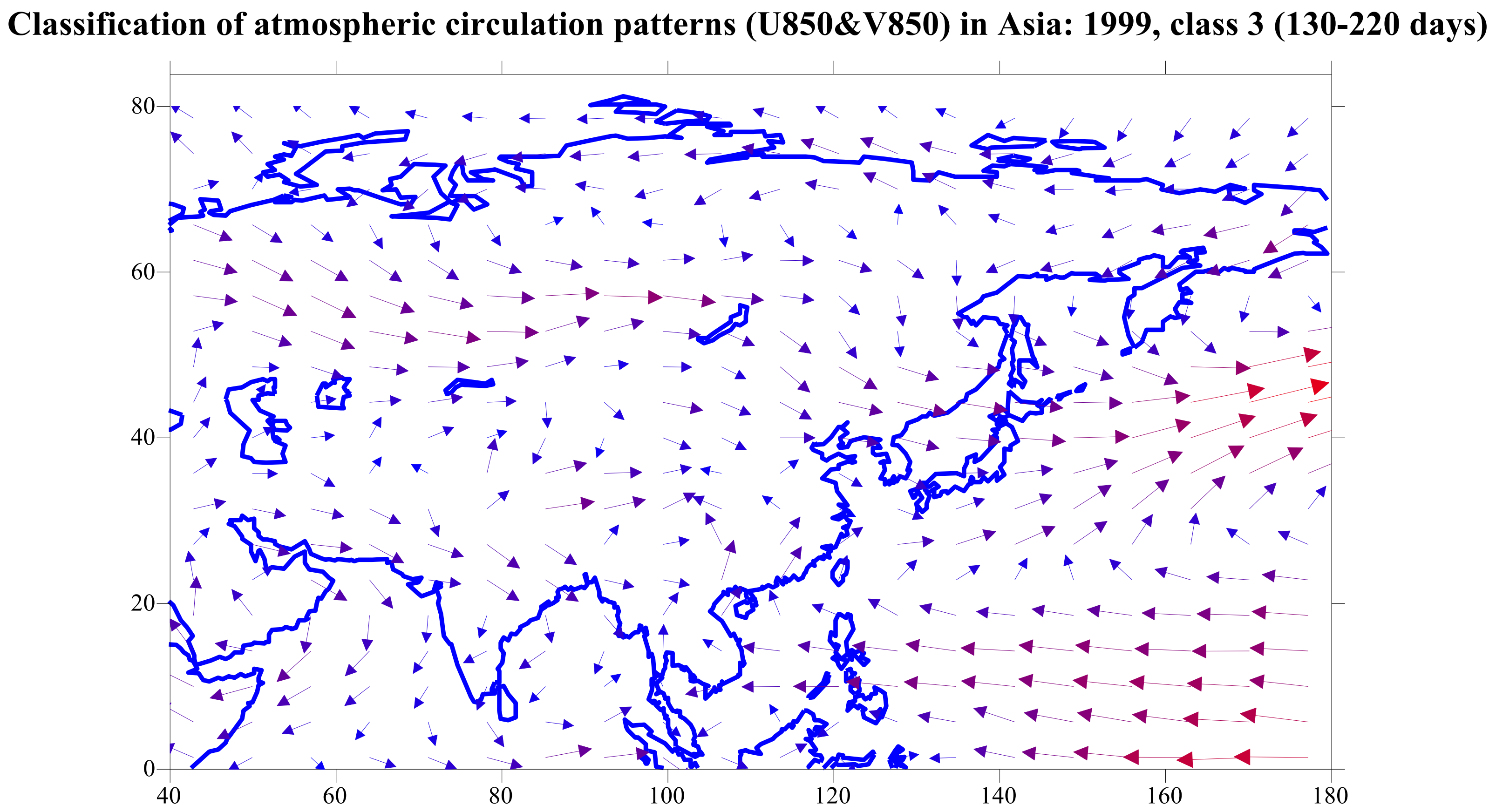
Classification of atmospheric circulation patterns (U850&V850) in Asia: class 1 (1999): 220-360 days
80
60
40
20
0
40
60
80
100
120
140
160
180

1999



Classification of atmospheric circulation patterns (U850&V850) in Asia: class 2: 1999 (1-130 days)
80
60
40
20
0
40
60
80
100
120
140
160
180

1999

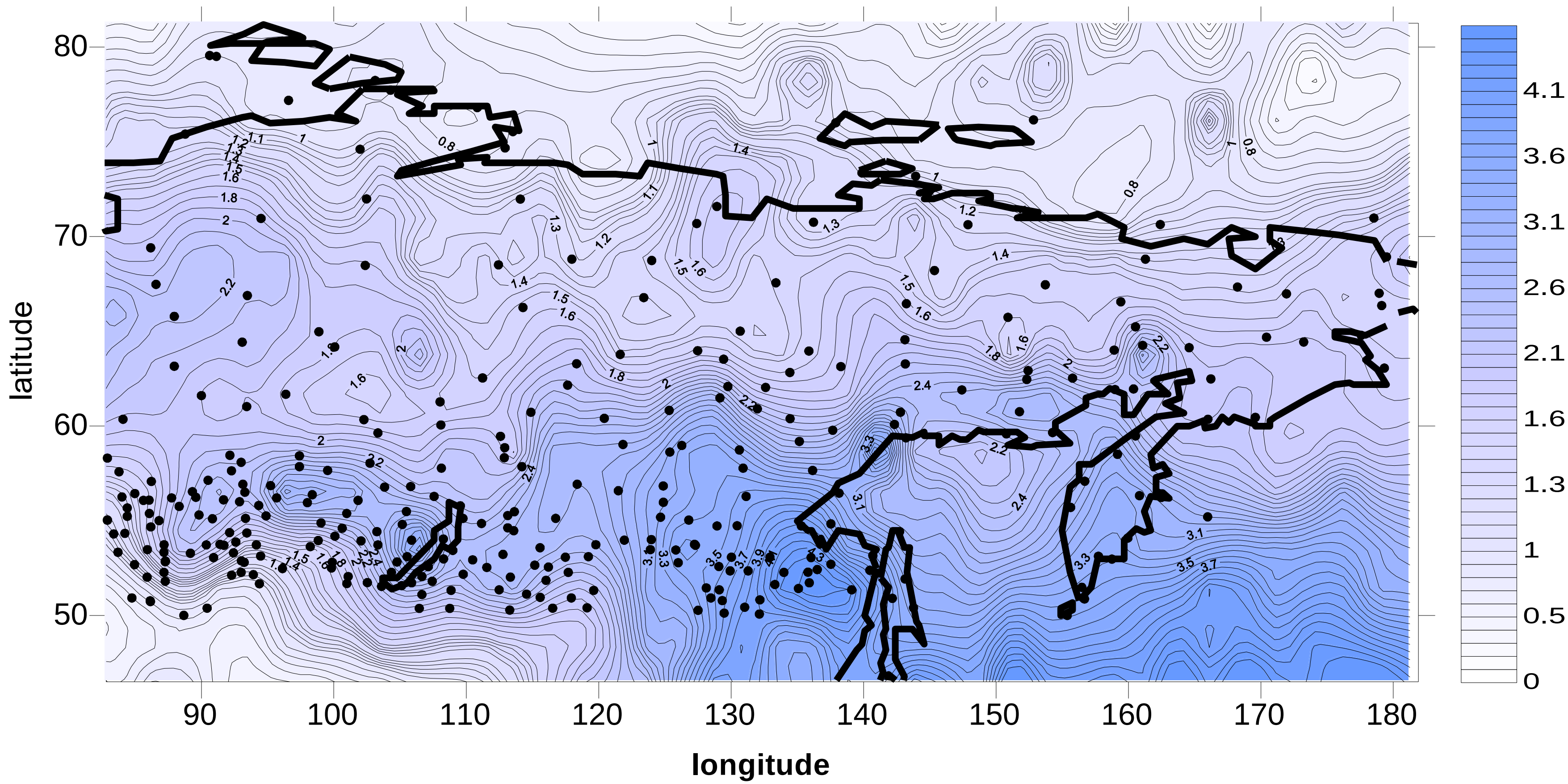


Classification of atmospheric circulation patterns (U850&V850) in Asia: 1999, class 3 (130-220 days)
80
60
40
20
0
40
60
80
100
120
140
160
180

Precipitation Field Statistics in Northern Asia

CMAP Data Set for 1979-2011

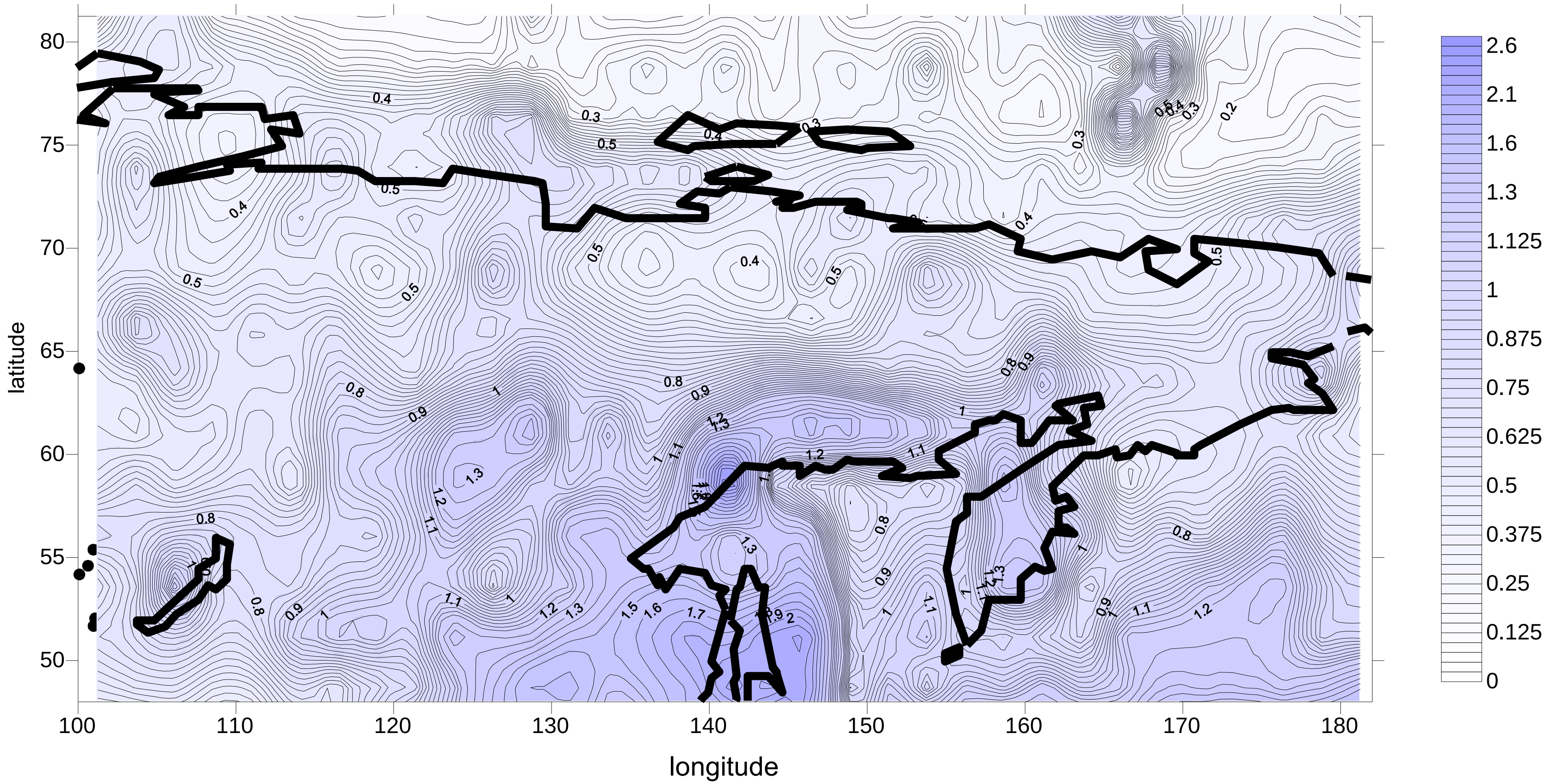
Mean Precipitation Rate (mm/day) for August, 1979-2011

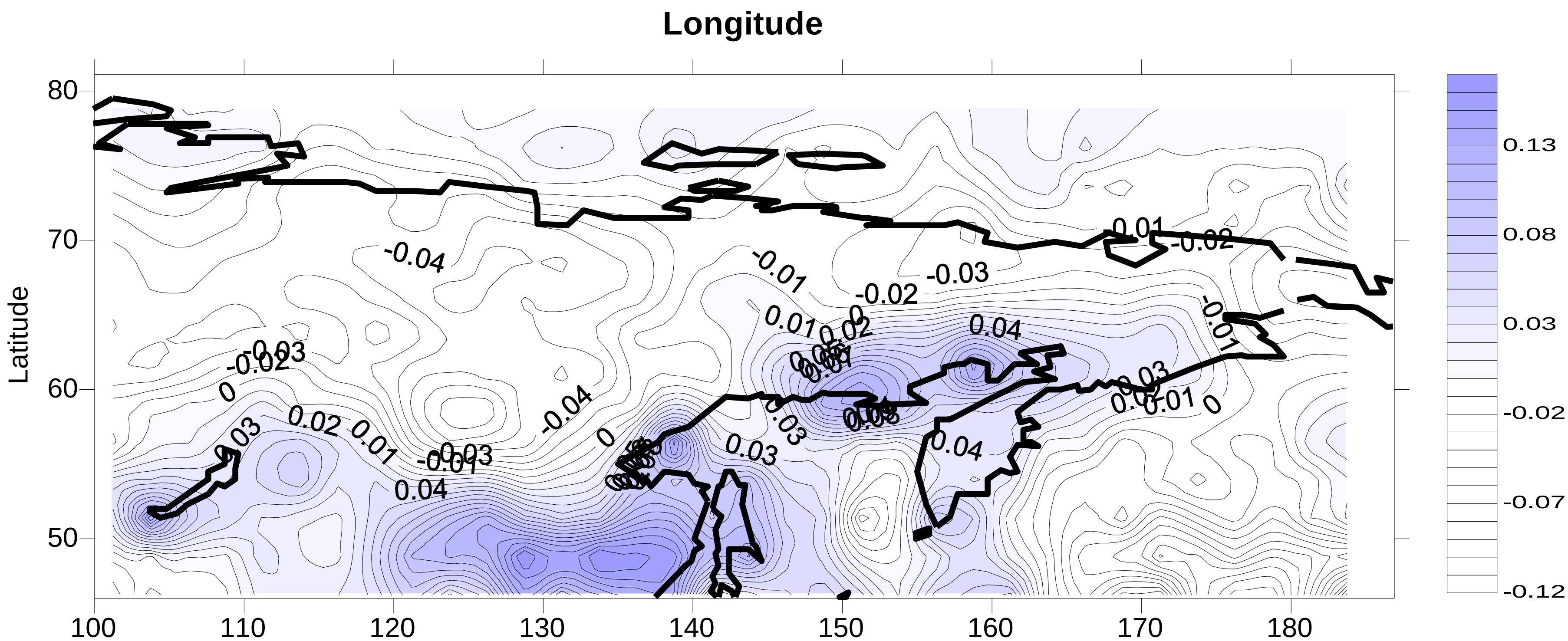


29-30 July 2014

Remote Sensing for Global Water
Circulation to Climate Change

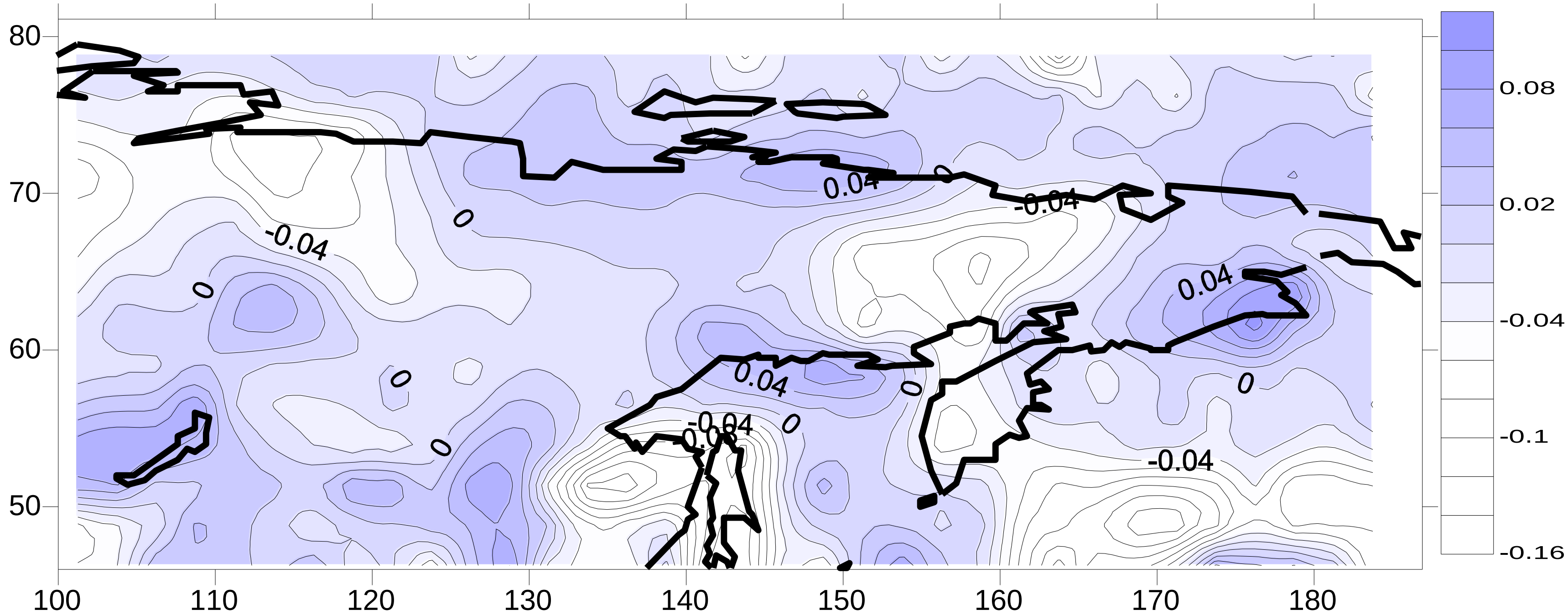
Rain Rate Variability (mm/day) - STD - in Northern Asia: August, 1989-2010



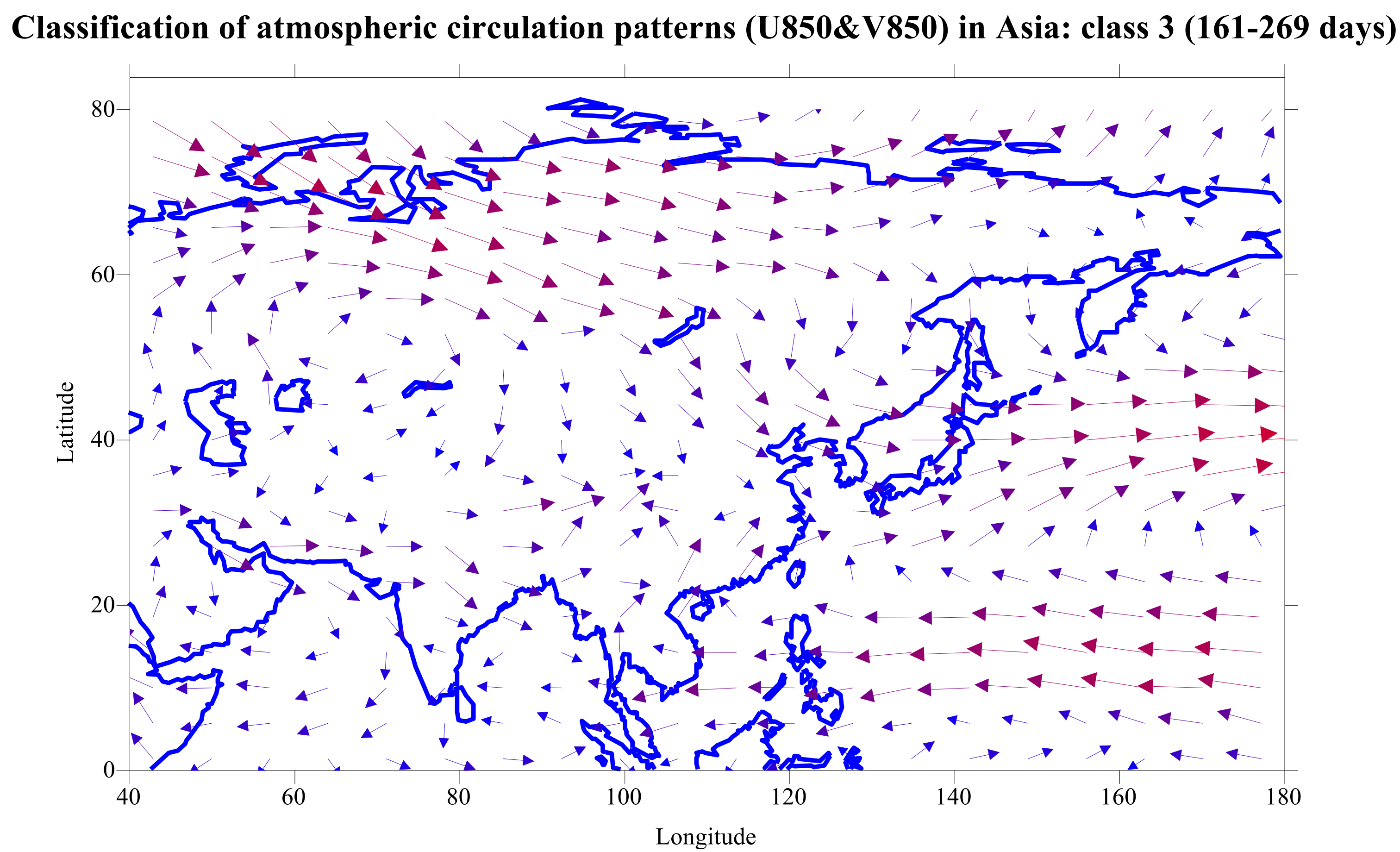


Rain Rate Variability Field - in Northern Asia for August, 1989-2010: EOF N1. (Pokrovsky, 2012)

Longitude

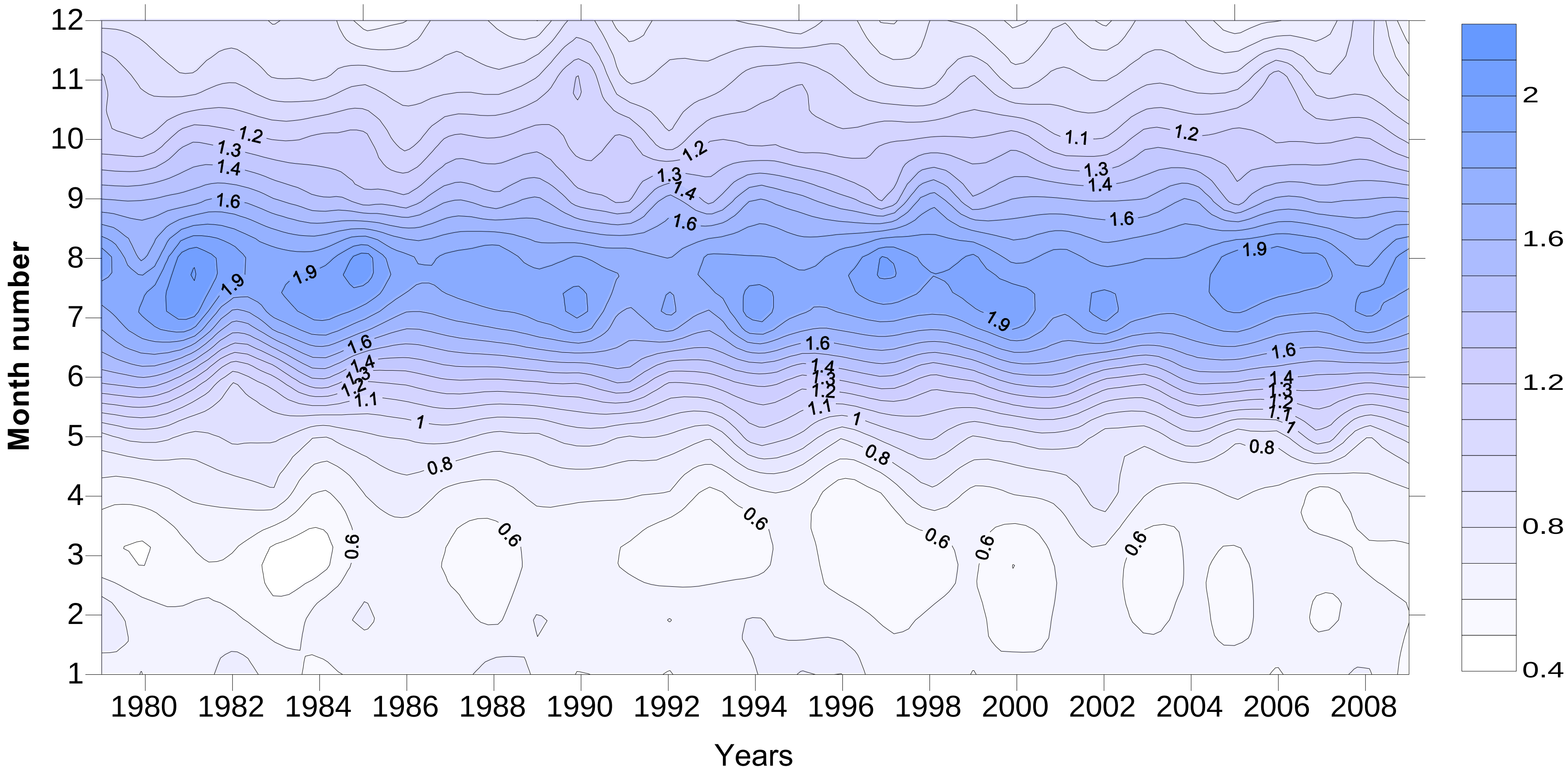


Rain Rate Variability Field - in Northern Asia for August, 1989-2010: EOF N2. (Pokrovsky, 2012)



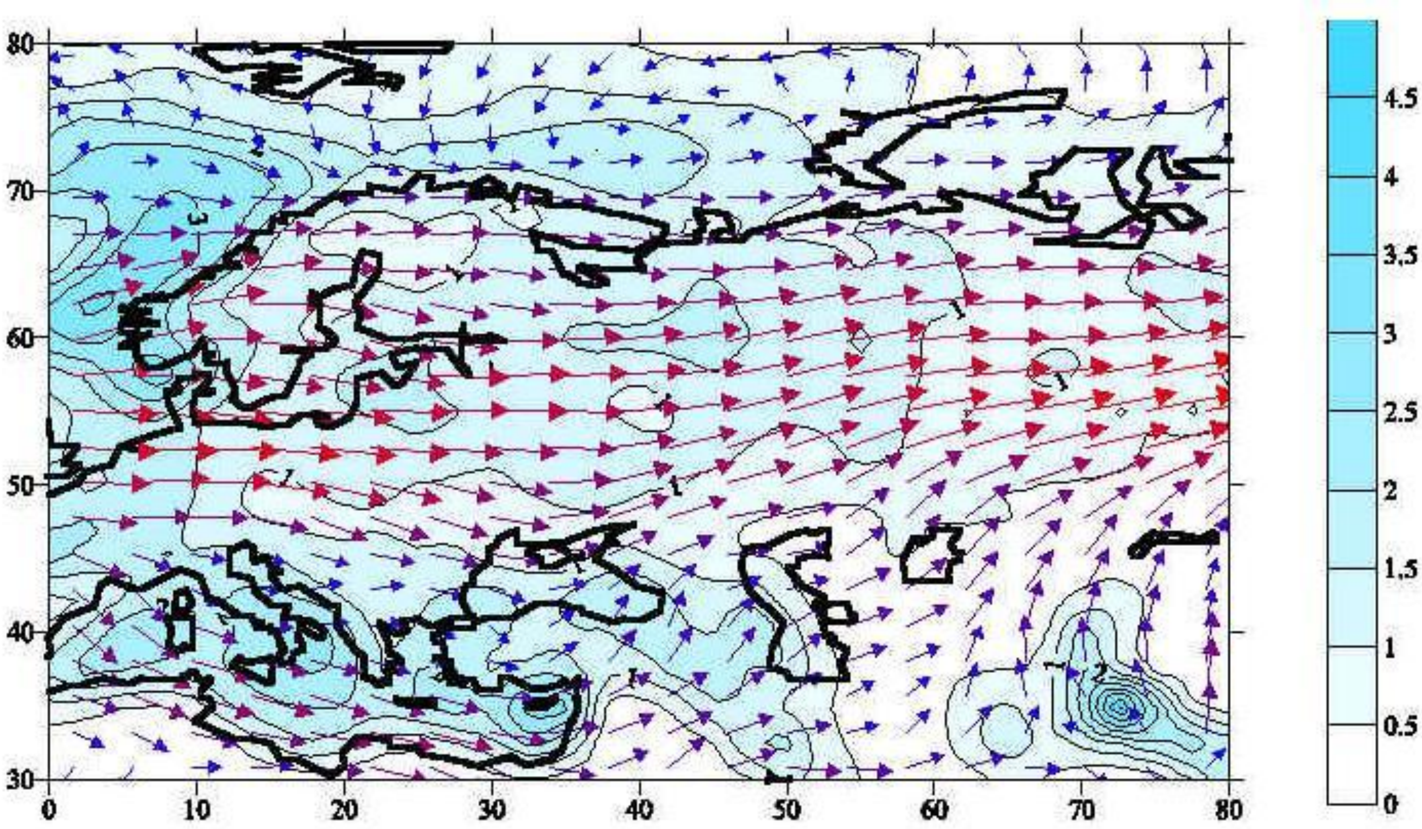
Classification of atmospheric circulation patterns (U850&V850) in Asia: class 3 (161-269 days)
Latitude
Longitude
80
60
40
20
0
40
60
80
100
120
140
160
180

Average over Northern Asia Precipitation Rate (mm/day) distribution by years and monthes



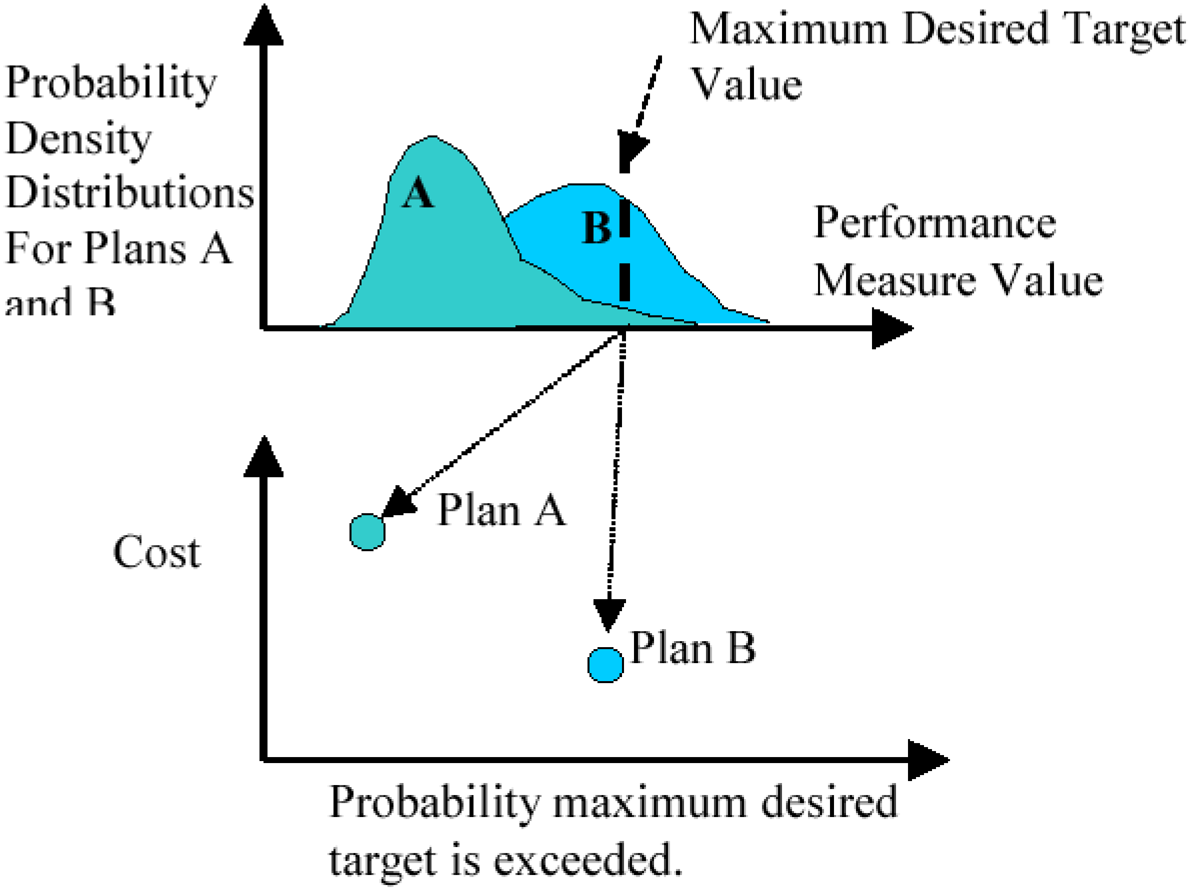
Fuzzy Classification of atmospheric circulation regimes in Europe

Monthly circulation regime 1 (winter and early spring): the joint pattern for U850-V850 and responding precipitation rate (mm/day) fields marked by blue color

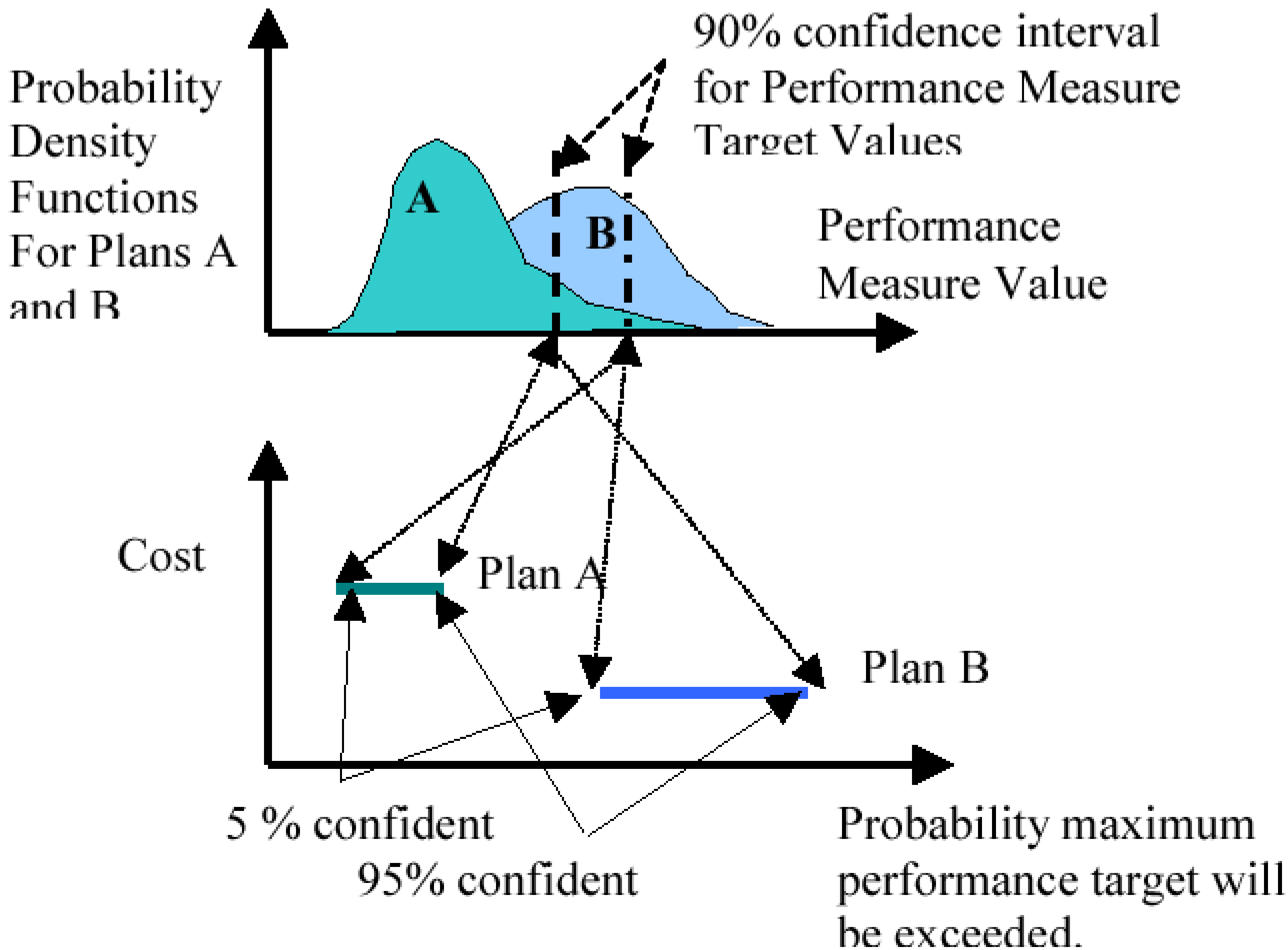


Decision Making

Probabilities and Costs of Alternative Projects: A and B



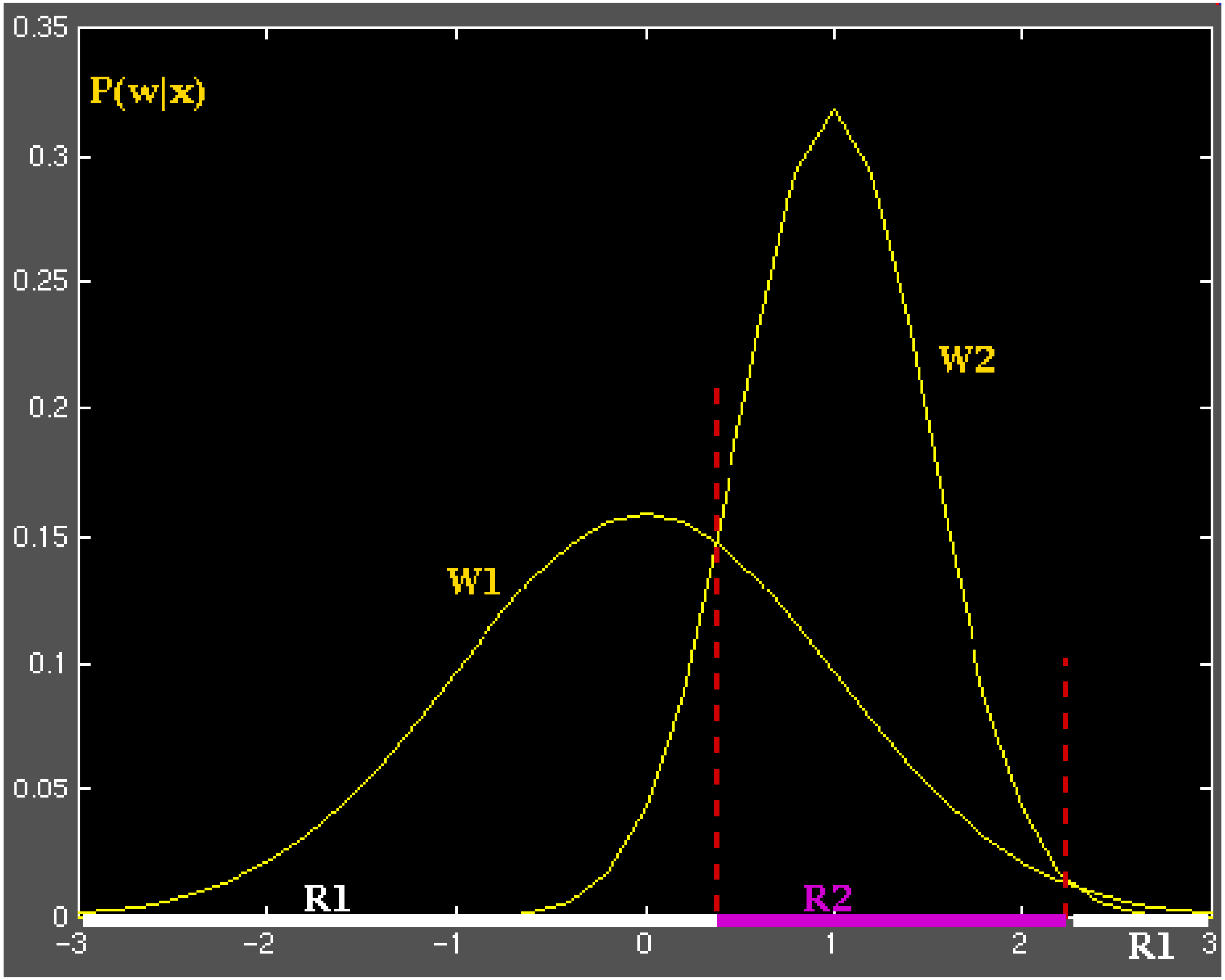
The 95% confidence levels are associated with the higher probabilities of exceeding the desired maximum target and the 5% confident levels are associated with the more desirable lower probabilities of exceeding the desired maximum target.

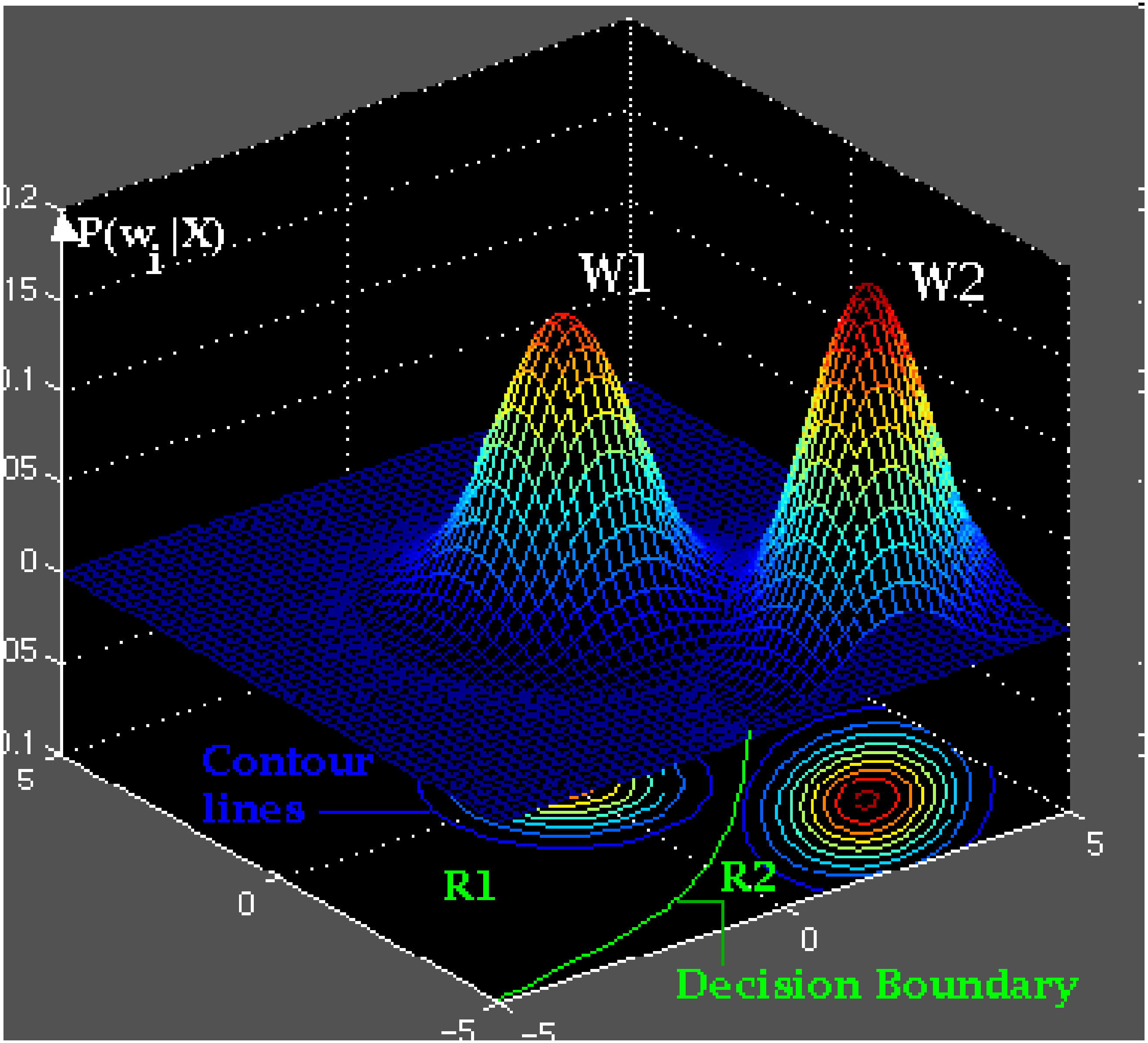


Bayesian decision theory (4):
Bayesian Theorem

$$P(W \mid X) = \frac{P(X \mid W)P(W)}{P(X)} \qquad (1)$$

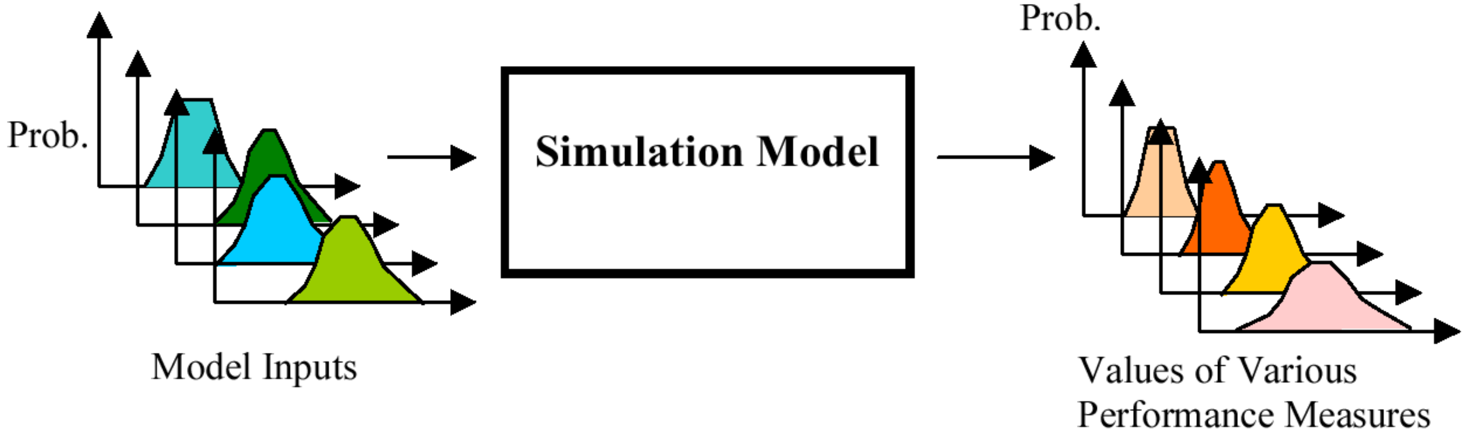
$$P(X) = \sum_i P(X \mid W_i)P(W_i) \qquad (2)$$


P(w|x)
W2
W1
R1
R2
R1
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
−3
−2
−1
0
1
2
3


P(w
i
|X)
W1
W2
Contour
lines
R1
R2
Decision Boundary
0.2
0.15
0.1
0.05
0
-0.05
-0.1
5
0
-5
-5
0
5

Droughts and Floods are related to long-lived extreme precipitation events

Fuzzy-Neural Forecasting methodology



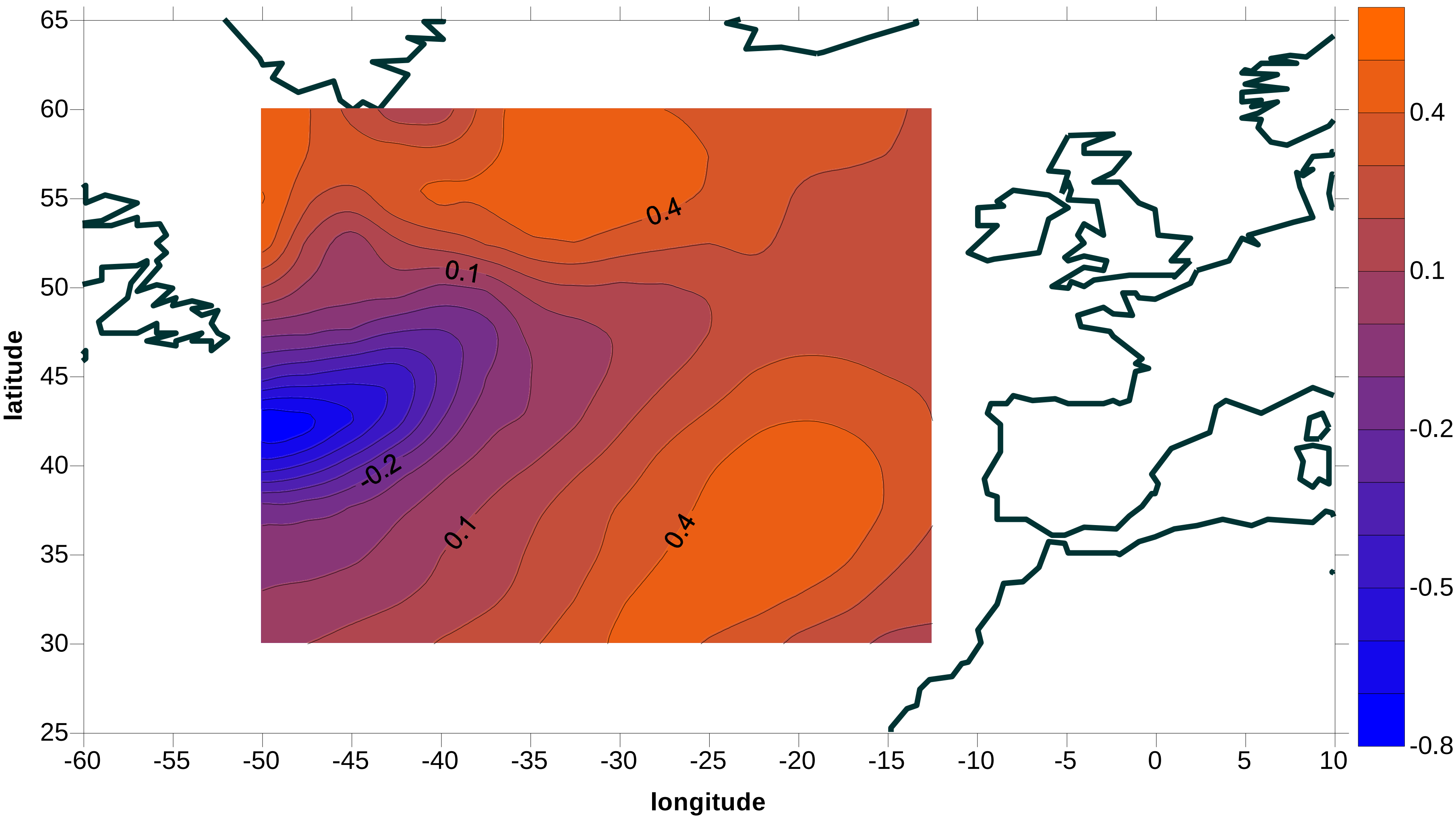
Fuzzy classification of precipitation anomalies in Europe (July)

Three most frequent rain rate distribution in July

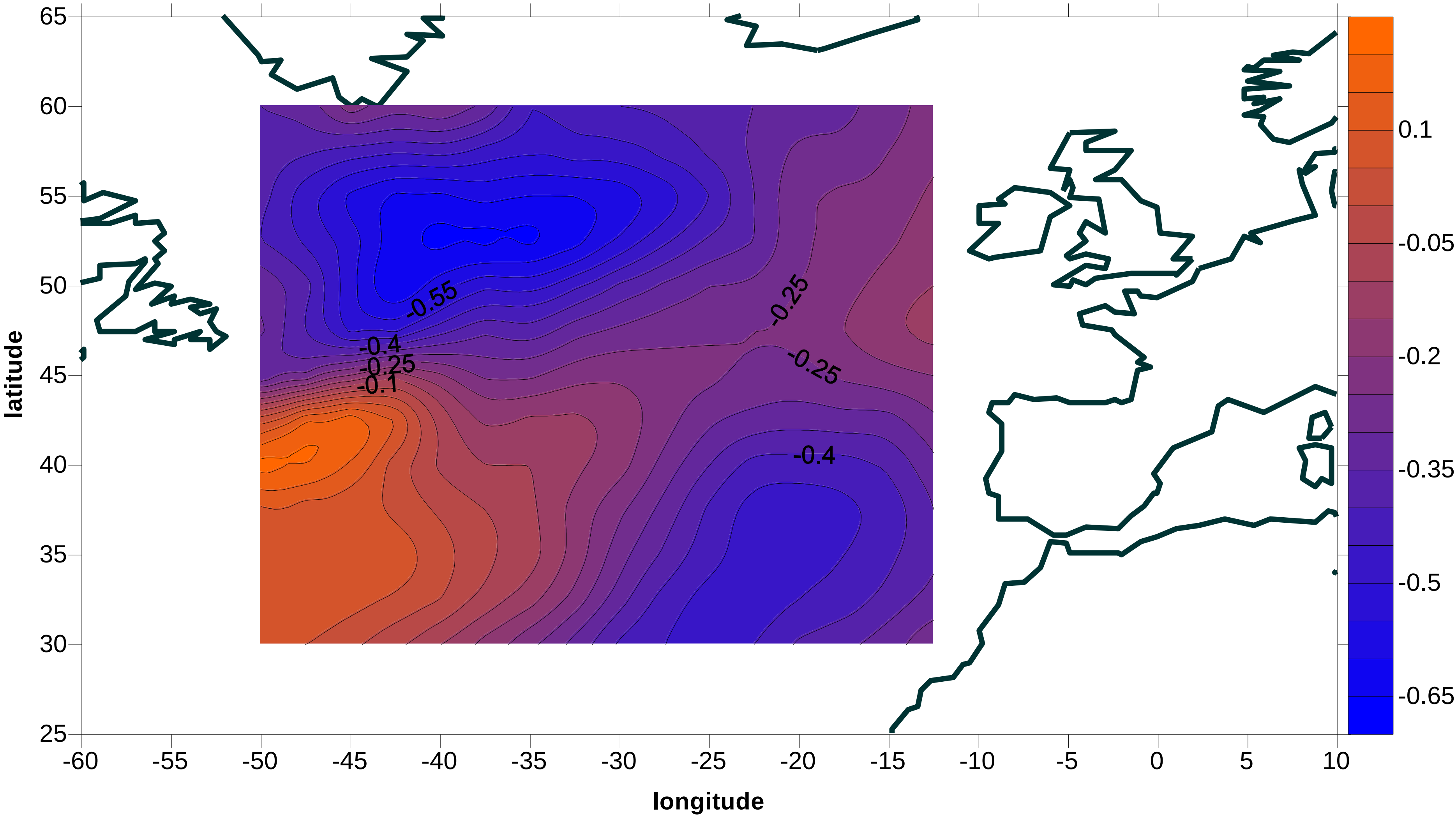
1. **Wet** weather in Northern Europe (Scandinavia and Russia) – high NAO
2. Rainy weather in Western Europe and dry or even **drought** conditions in Eastern part of European Russia
3. Extreme rainy in countries of Central and Eastern Europe (e.g. **floods**)

Fuzzy classification of the SST in North Atlantic in May

SST in North Atlantic: Fuzzy set N 1 - May (1958-98)



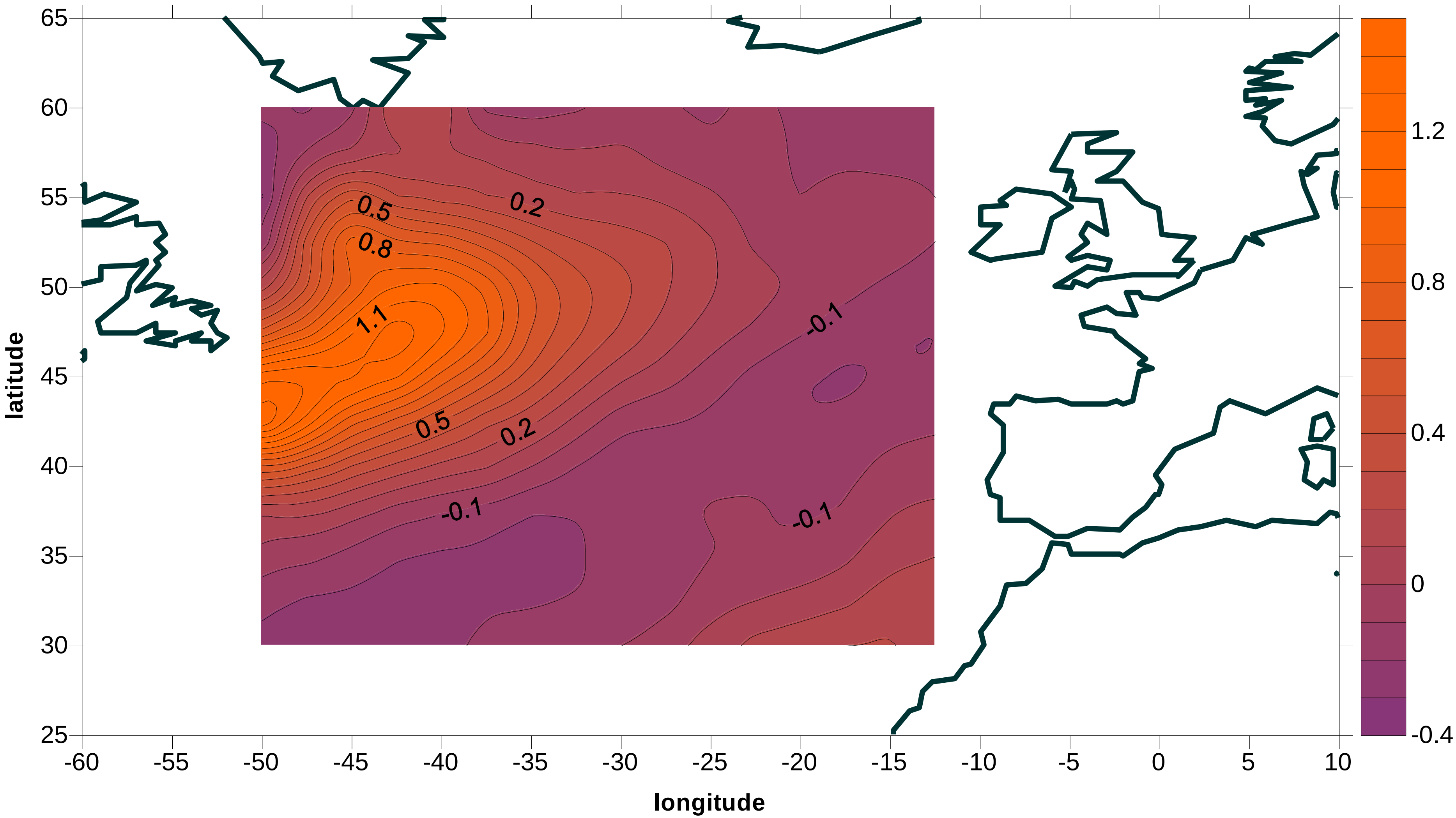
SST in North Atlantic: Fuzzy set N 2 - May (1958-98)



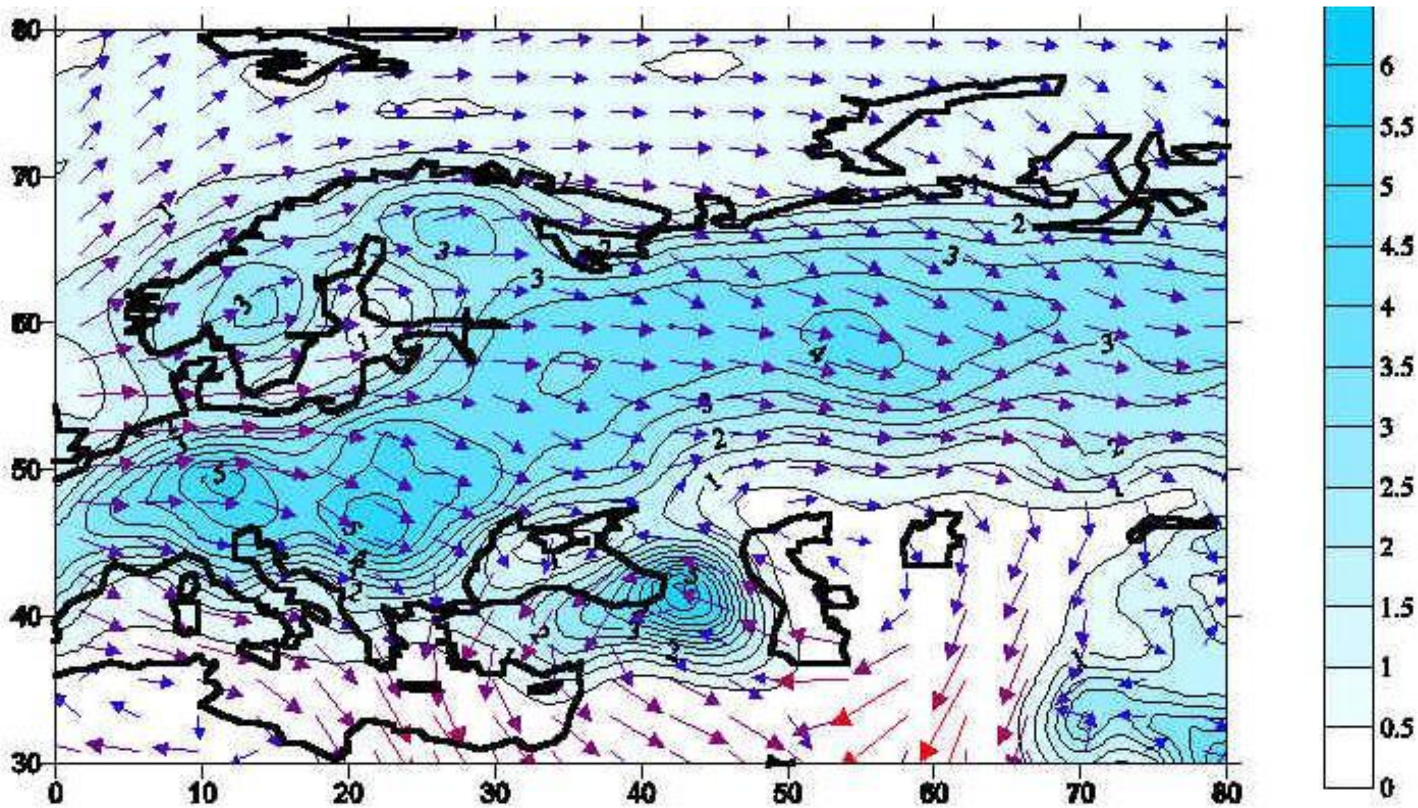
29-30 July 2014

Remote Sensing for Global Water Circulation to Climate Change

SST in North Atlantic: Fuzzy set N 3- May (1958-98)



Monthly circulation regime 2 (summer): the joint pattern for U850-V850 and responding **precipitation rate (mm/day) fields marked by blue color**

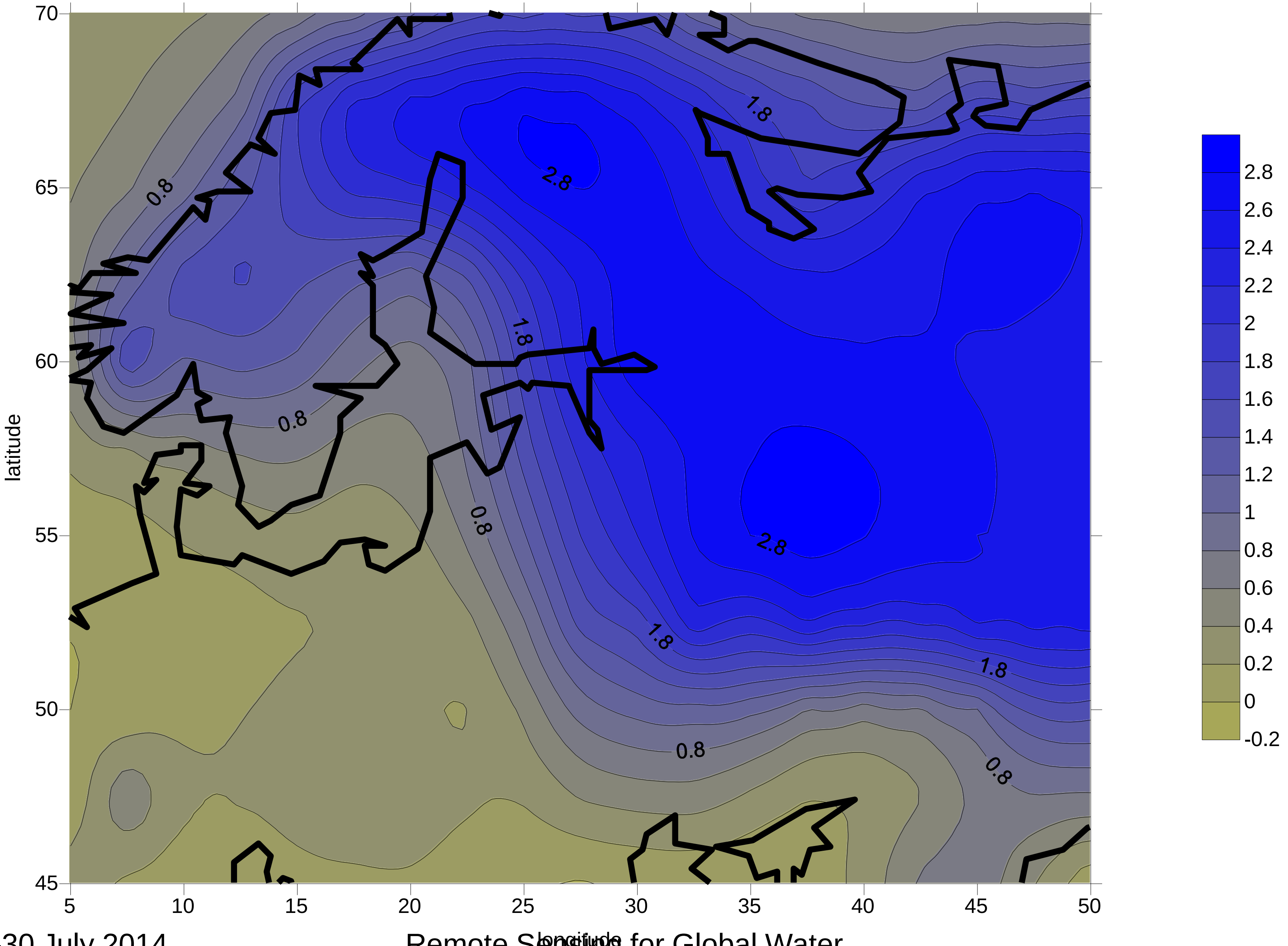


Major circulation regimes over eastern North Atlantic and Europe in summer by three vorticity poles:

- North-Western (Scandinavia)
- Western Mediterranean
- Caucasian

Fuzzy classification of precipitation regimes in July

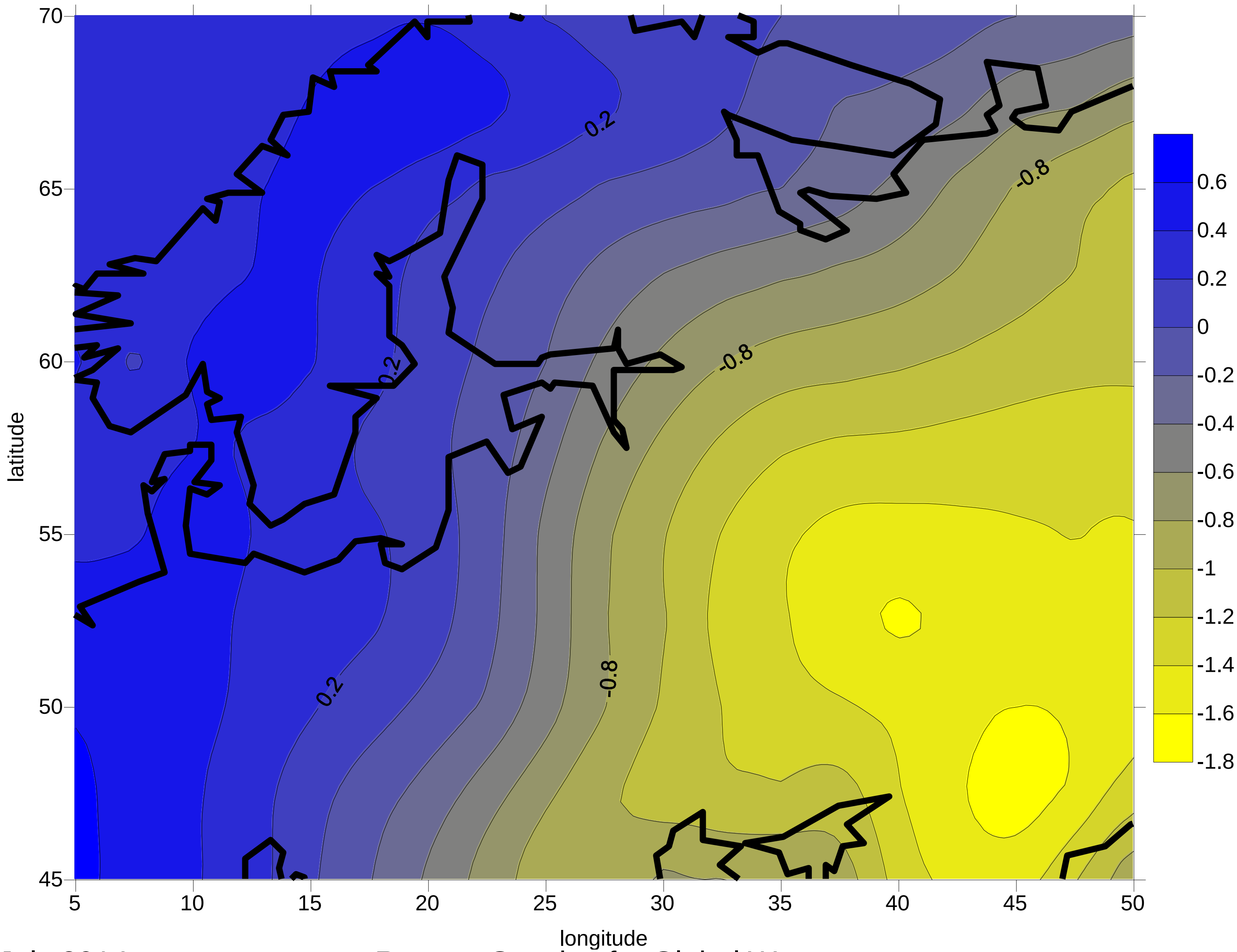
Rain field in Europe: July, fuzzy set N 1 (mm/day)



29-30 July 2014

Remote Sensing for Global Water Circulation to Climate Change

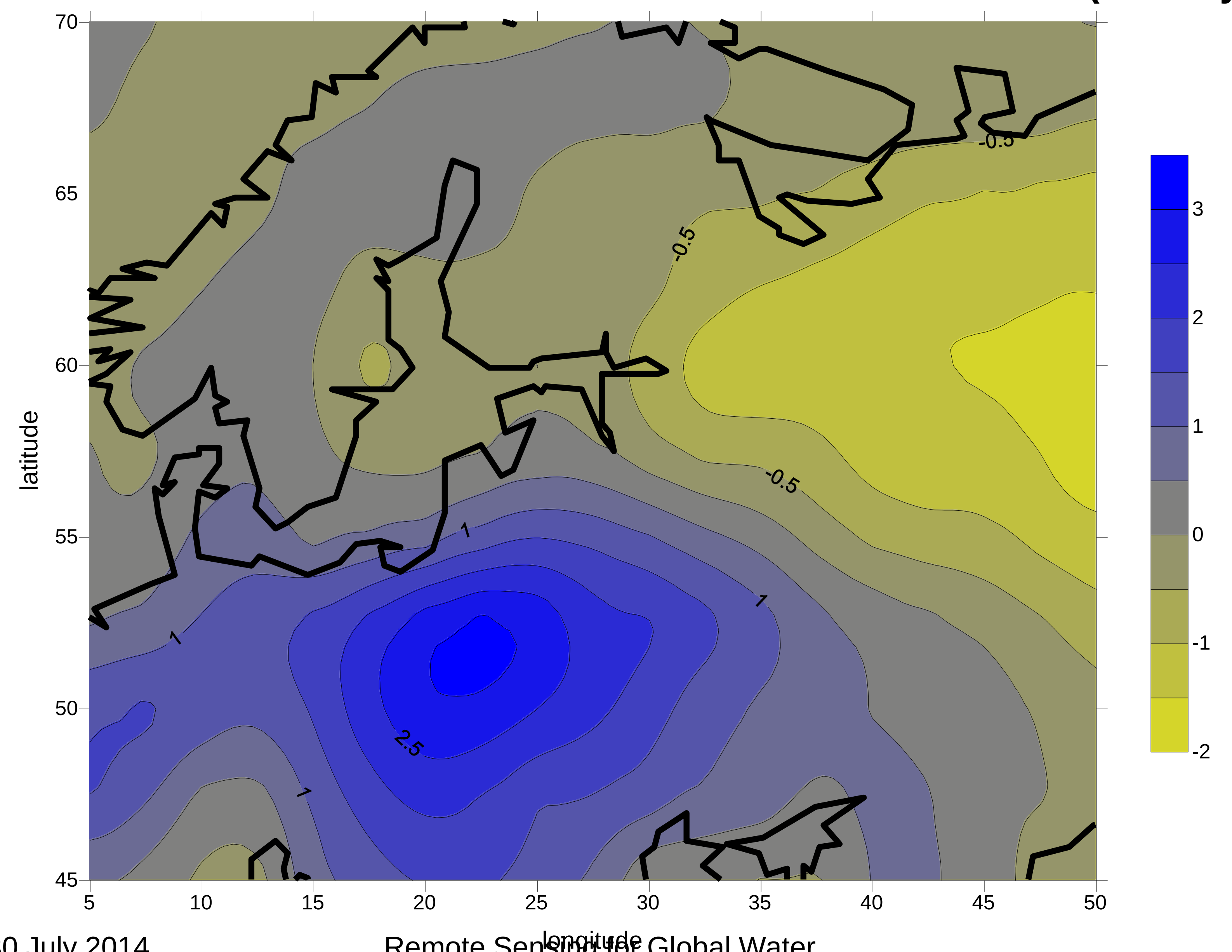
Rain field in Europe: July, fuzzy set N 2 (mm/day)



29-30 July 2014

Remote Sensing for Global Water Circulation to Climate Change

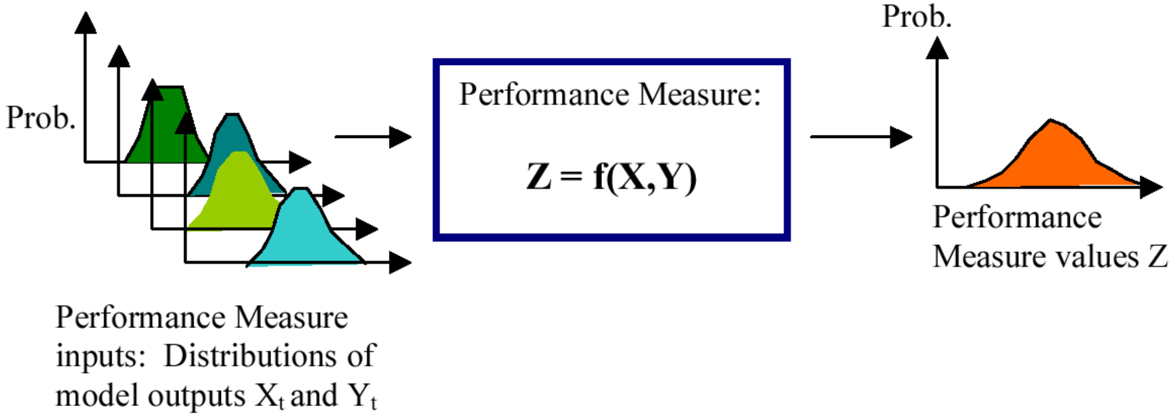
Precipitation rate (anomaly field Europe) for July (1958-98): fuzzy sample N3 (mm/day)

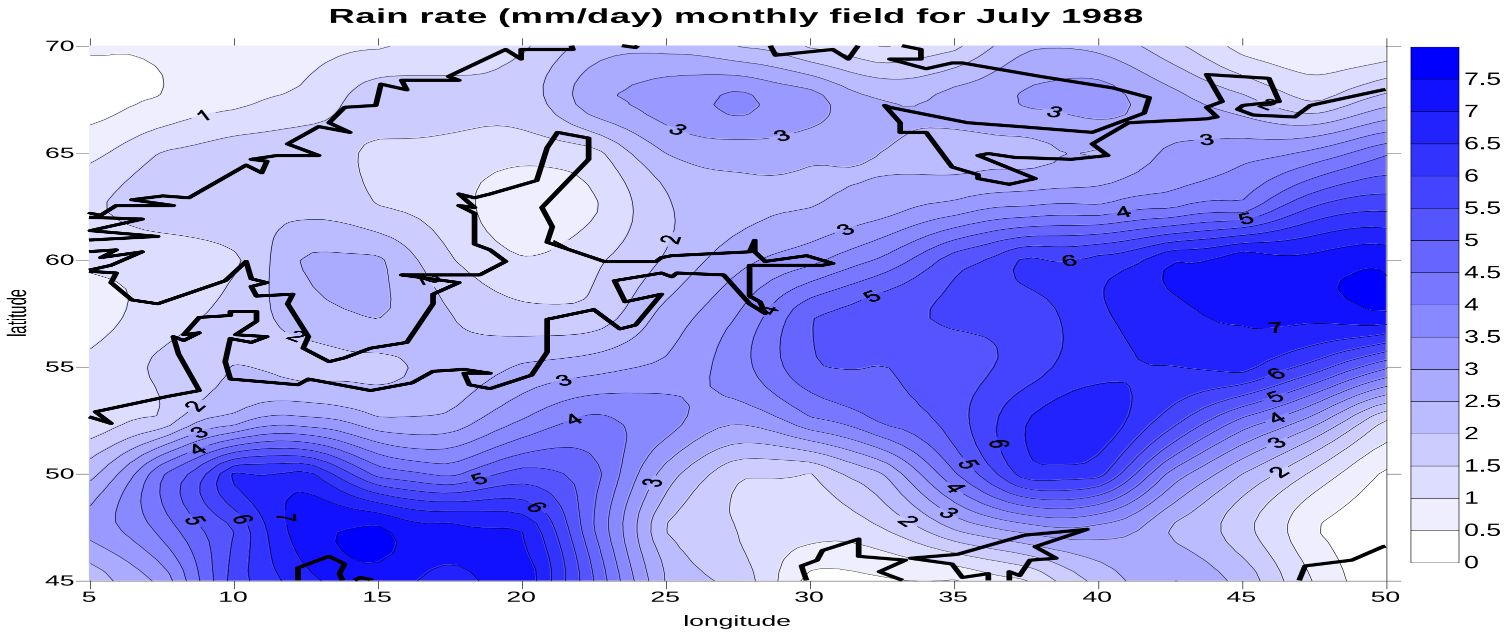


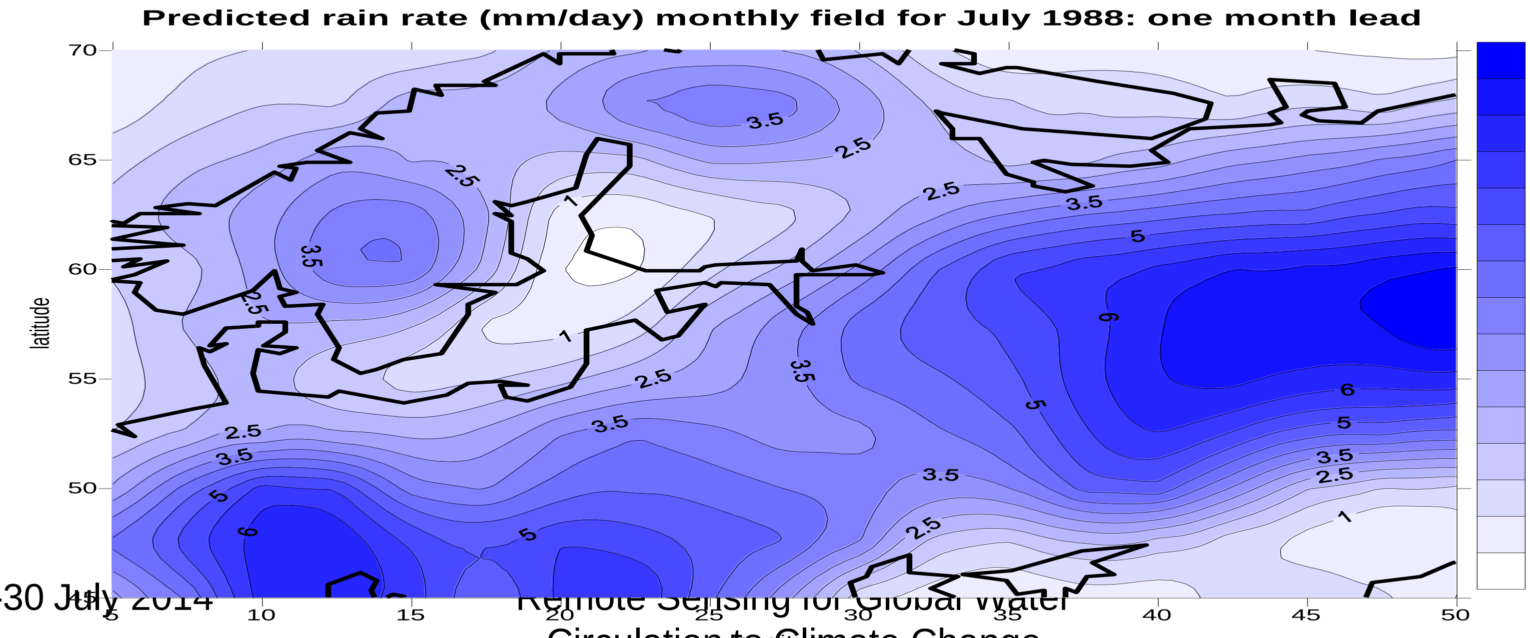
29-30 July 2014

Remote Sensing for Global Water Circulation to Climate Change

Calculating the distribution of a performance measure from the distributions of its input X(t) and predicted Y(t) variables




Rain rate (mm/day) monthly field for July 1988
latitude
longitude

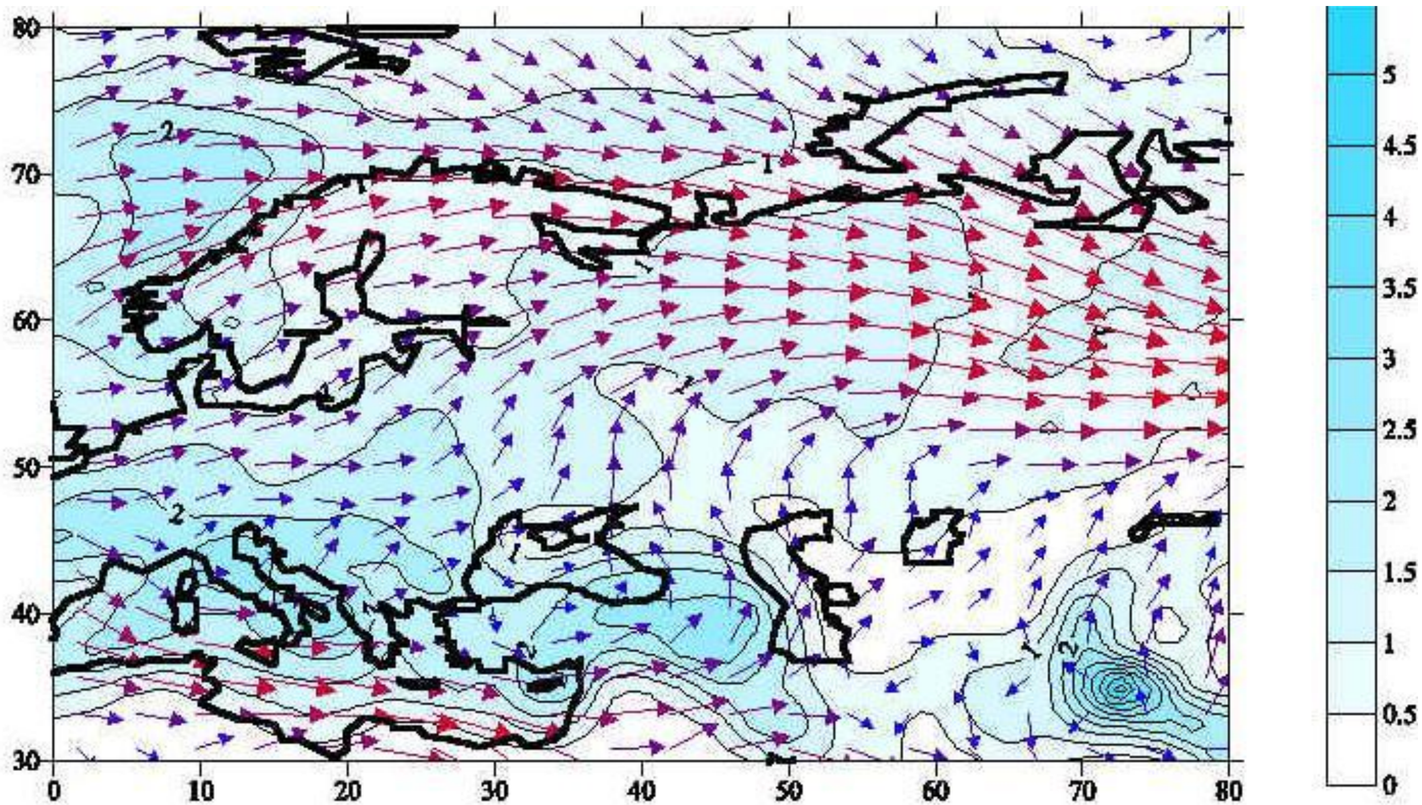

Predicted rain rate (mm/day) monthly field for July 1988: one month lead
latitude
longitude

Three most frequent rain rate distribution in July

- **Wet** weather in Northern Europe (Scandinavia and Russia) – high NAO
- Rainy weather in Western Europe and dry or even **drought** conditions in Eastern part of European Russia
- Extreme rainy in countries of Central and Eastern Europe (e.g. **floods**)

Monthly circulation regime 3 (**autumn and early winter**): the joint pattern for U850-V850 and responding precipitation rate (mm/day) fields marked by blue color

(mm/day)



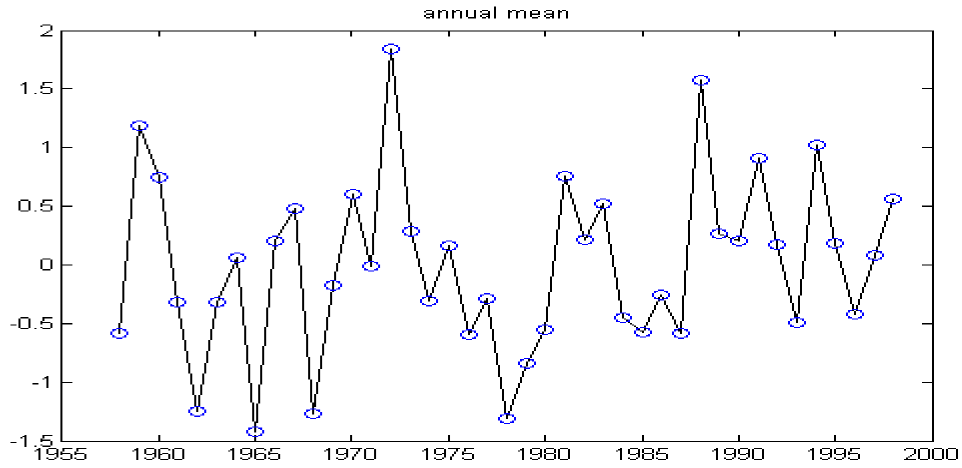
In late fall and winter the vorticity system consists of three other poles:

- North-Western
- Northern Africa
- Arctic Russia (Kara Sea area)

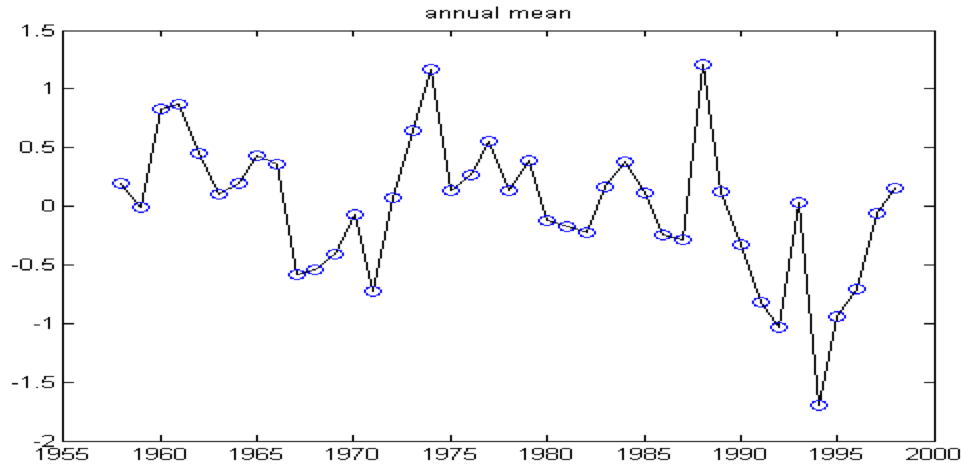
Climatology (anomalies) for July 1958-98:

averaged over 45-70 N; 5-50 E

Surface air temperature (C)



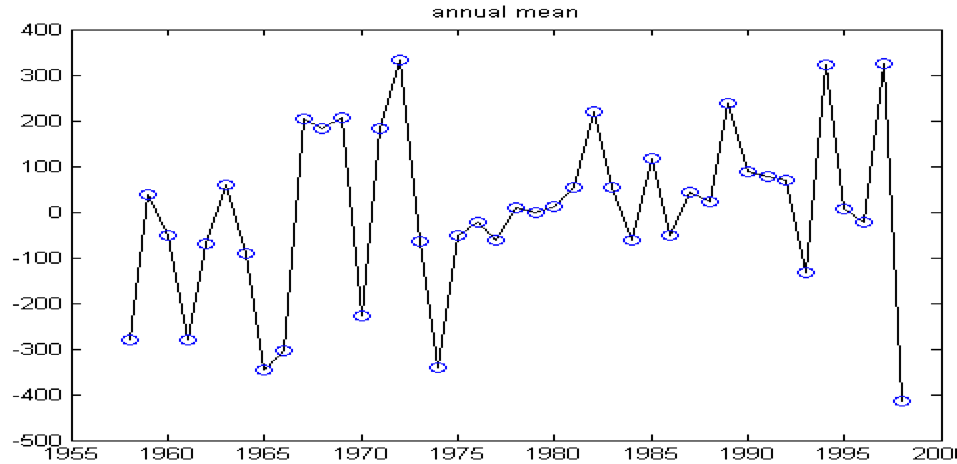
Precipitation ((mm/day))



Climatology (anomalies) for July 1958-98:

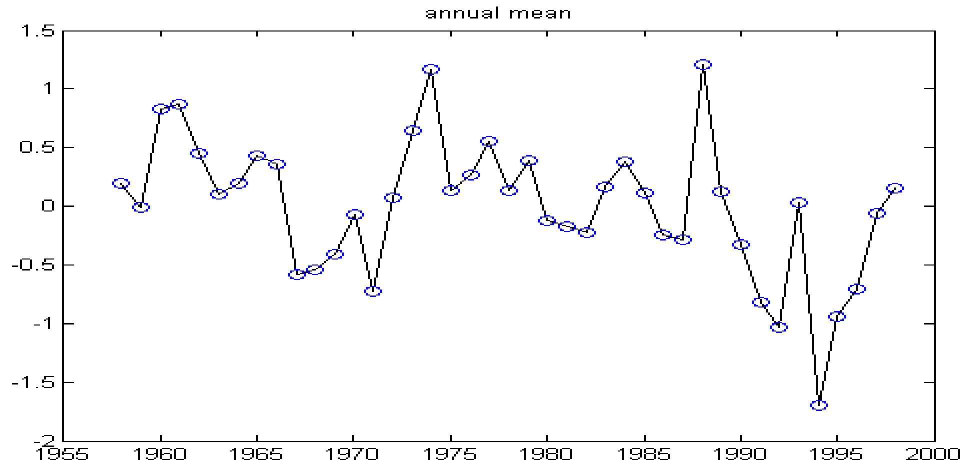
averaged over 45-70 N; 5-50 E

Sea Level Pressure (HPa)



Precipitation

(mm/day)



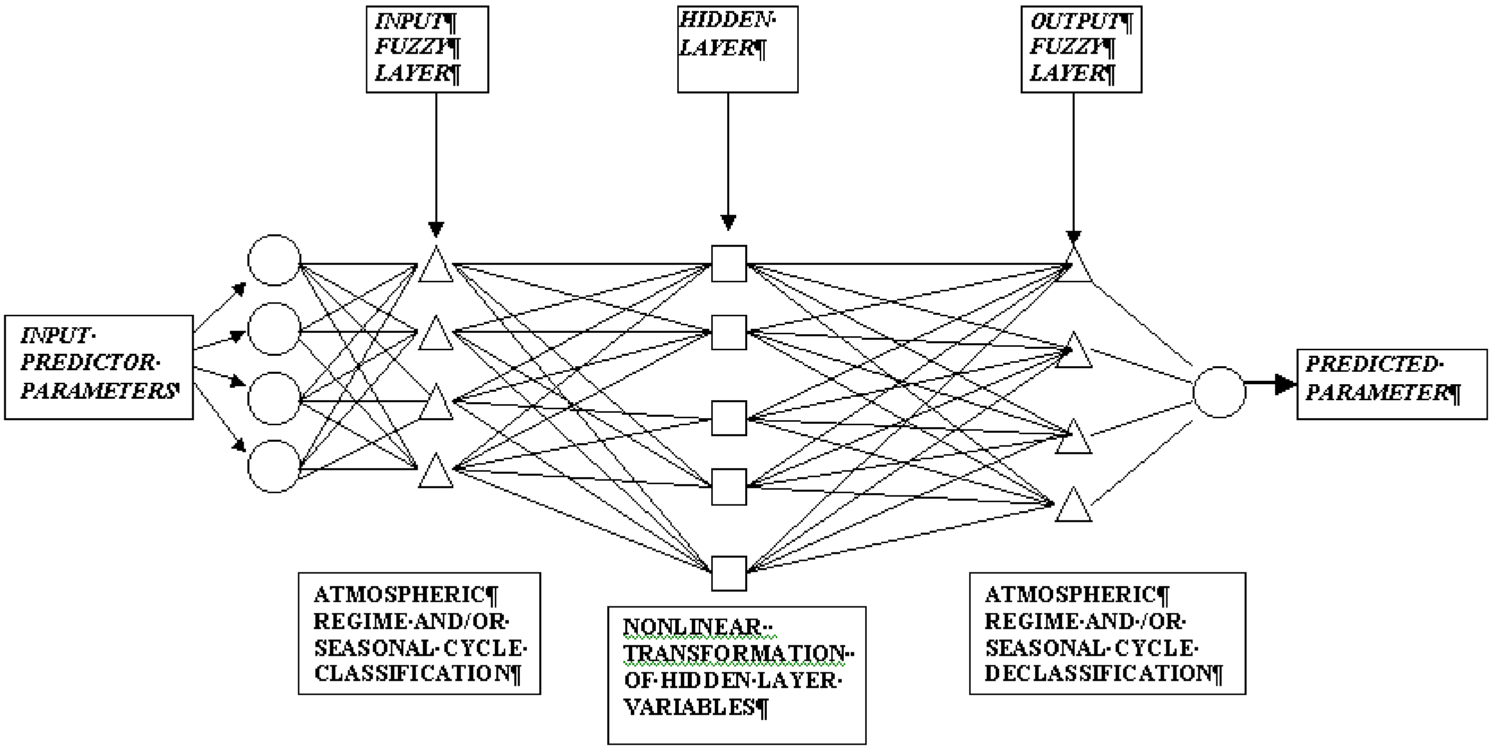
Preliminary conclusions

- Rain rate is more uniformly distributed in the winter in various latitude belts across Europe than in summer
- A zonal circulation type dominates in this case and more precipitation is delivered from the Atlantic.
- More intensive precipitations are occurred in Southern Europe because of strong moisture transport into this area from Atlantic

Neural network methodology advantages:

- **Flexibility in description of the arbitrary non-linear dependencies**
- **Easy incorporation of the additional and renewal feedbacks**
- **Fast self-learning feature**
- **High approximation fit to measurement data achieved by NN model**

Fuzzy-Neural Model

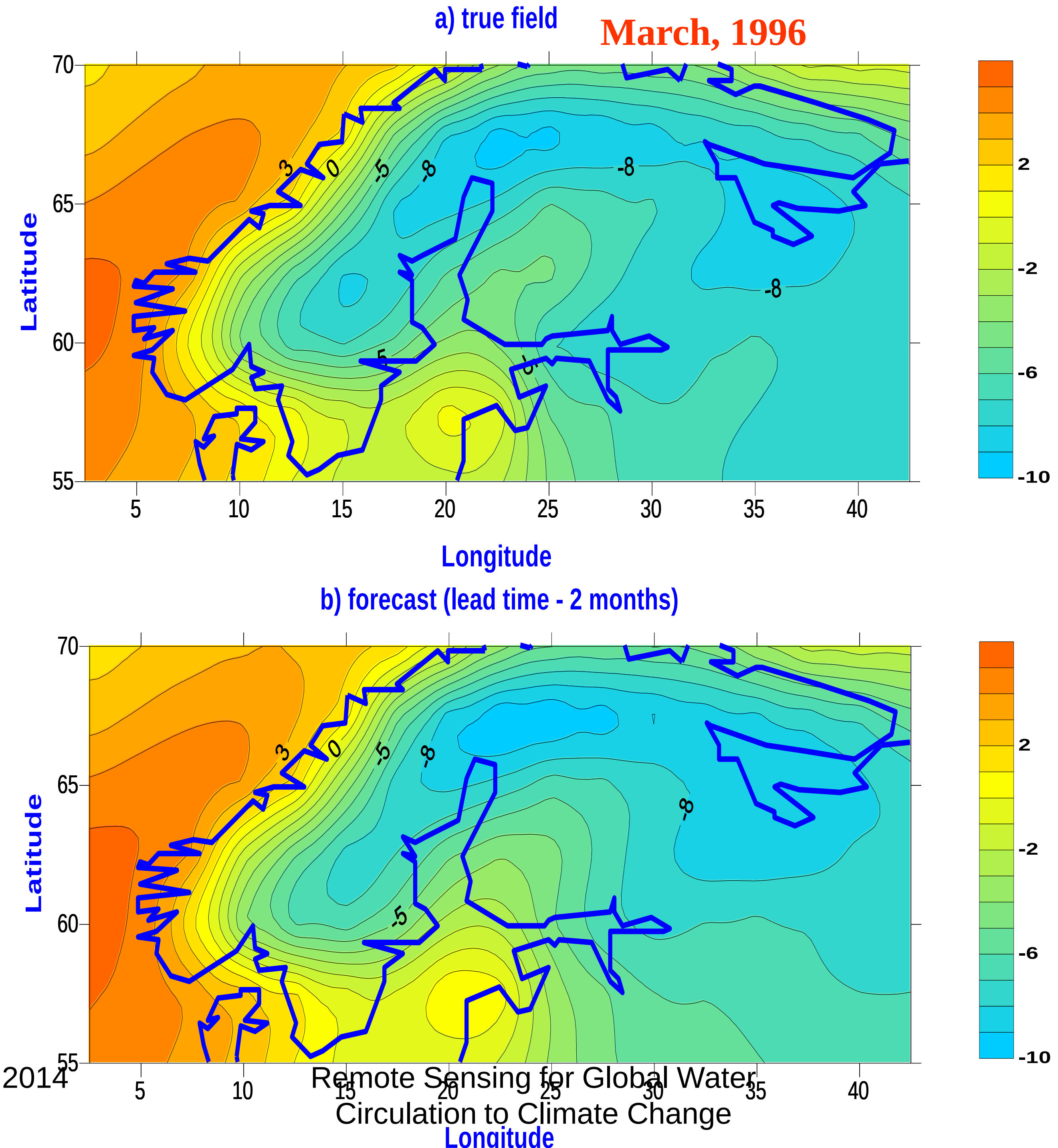


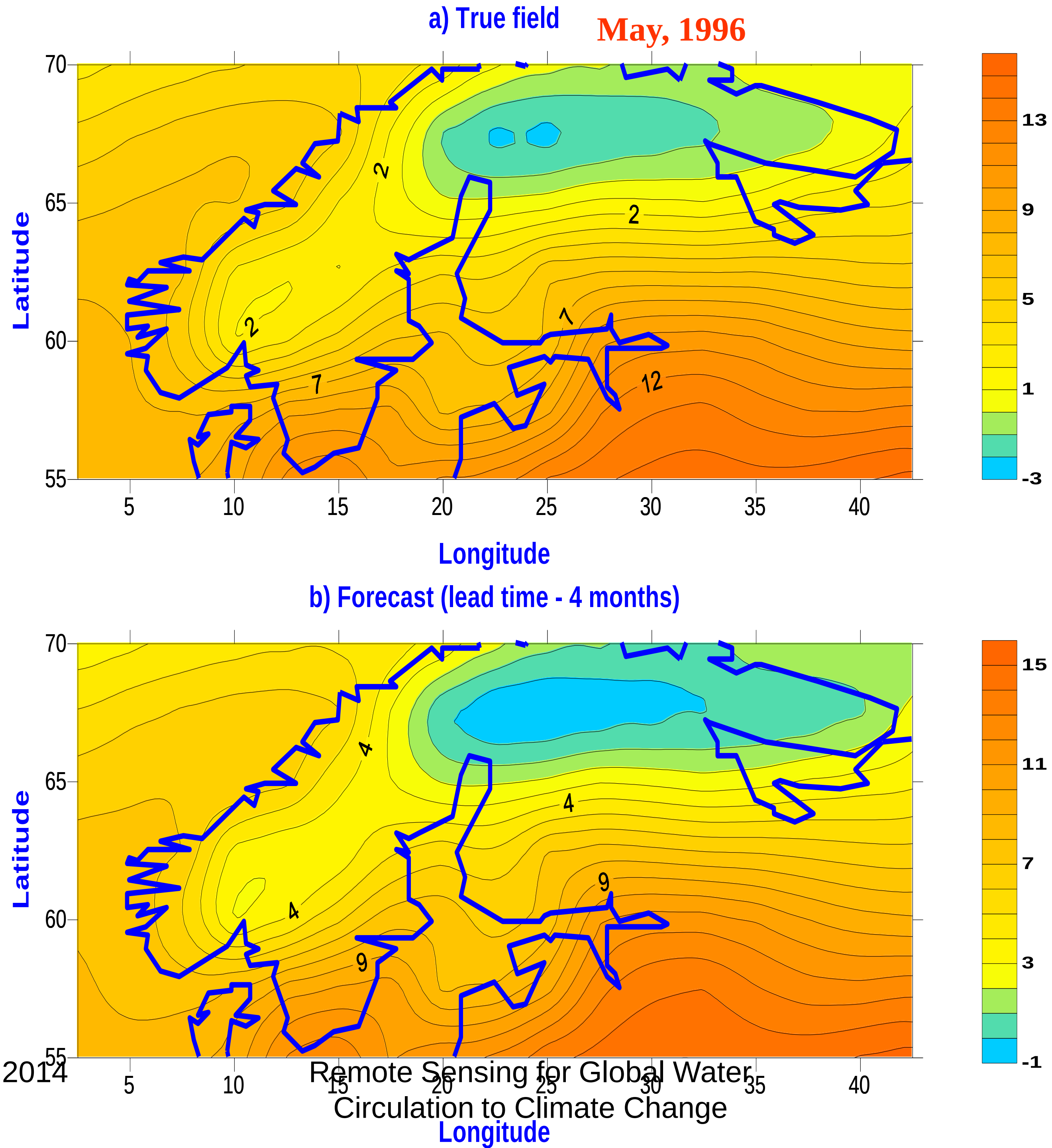
Main low frequency oscillation areas (ENSO, NAO, AO) are included in predictor analysis

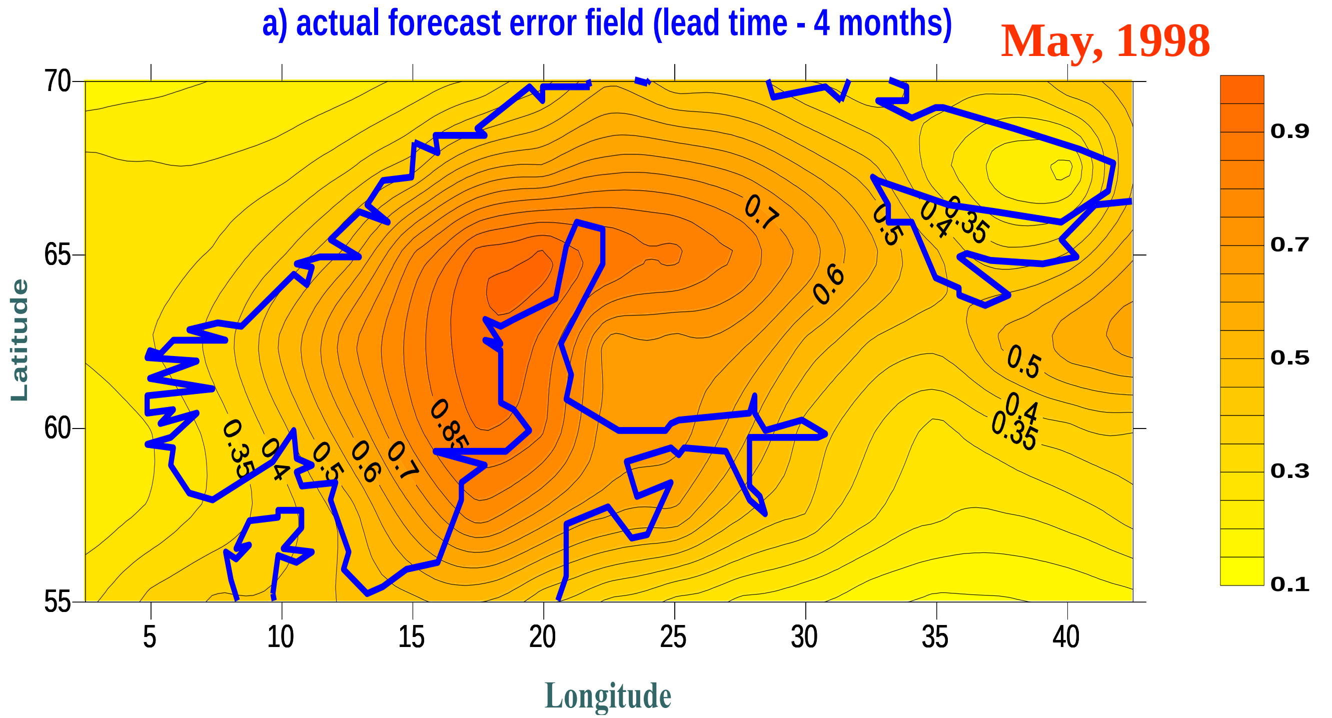
Why “slow oscillations”were selected?

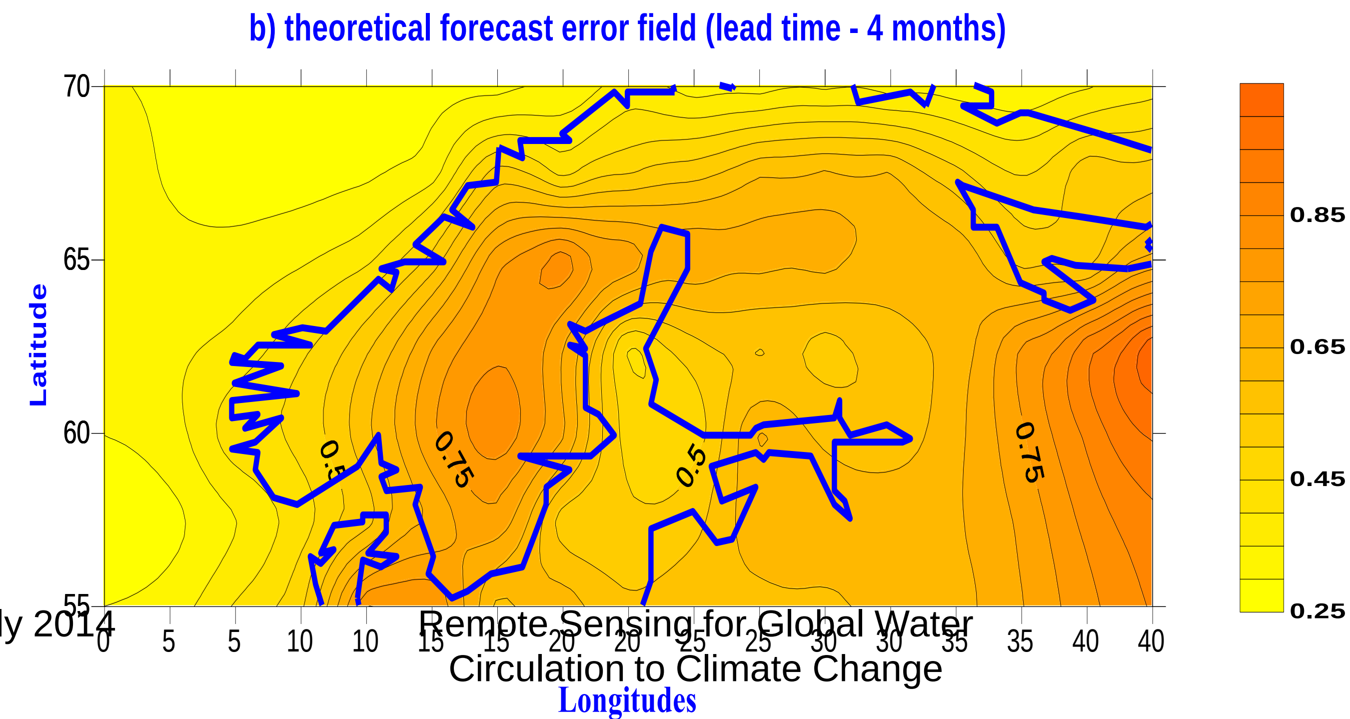
- Slow oscillations exhibit the highest amplitudes among other frequencies
- Slow oscillations explain the most part of total variability
- Slow oscillations are most stable to perturbation in initial conditions for numerical weather models based on differential equation solution

Temperature Field Seasonal Prediction
(Anomalies)

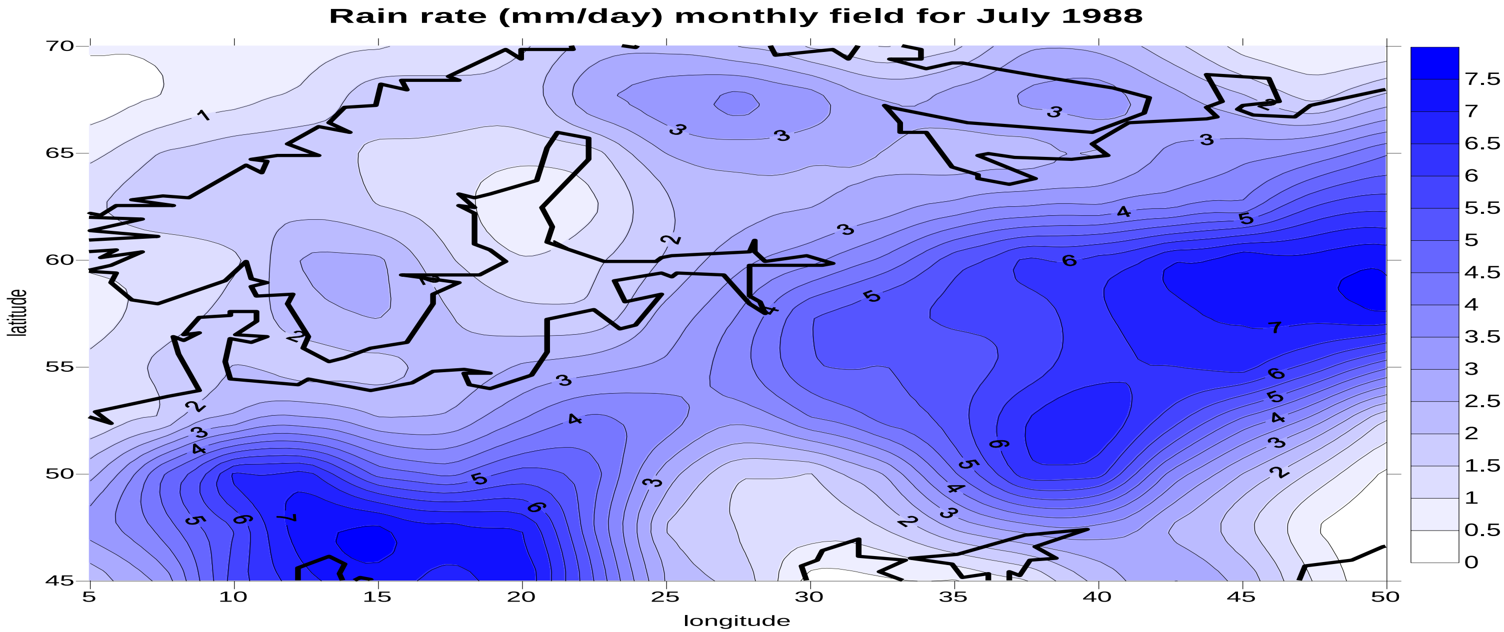

a) true field
March, 1996
70
65
60
55
Latitude
5
10
15
20
25
30
35
40
Longitude
3
0
-5
-8
-8
-8
2
-2
-6
-10
b) forecast (lead time - 2 months)
70
65
60
55
Latitude
5
10
15
20
25
30
35
40
Longitude
3
0
-5
-8
-8
-5
2
-2
-6
-10

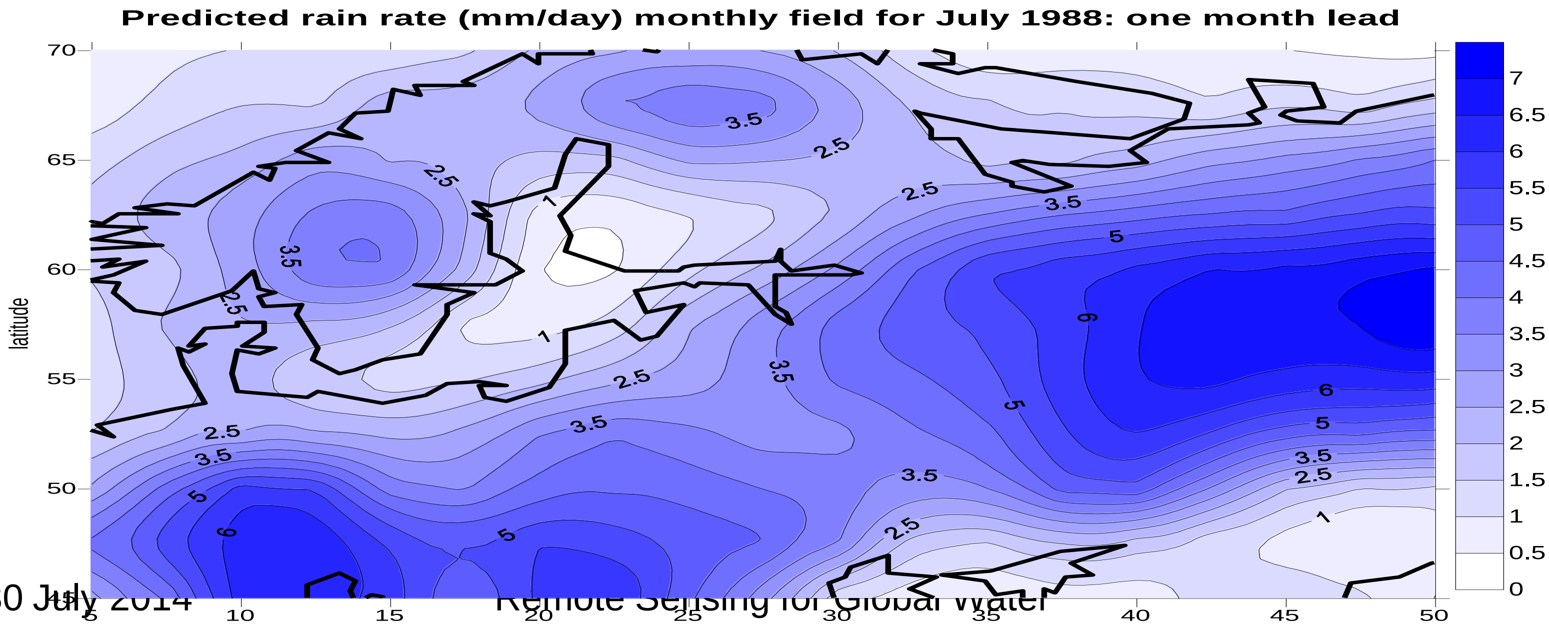

a) True field
May, 1996
Latitude
Longitude
b) Forecast (lead time - 4 months)
Latitude
Longitude
13
9
5
1
-3
15
11
7
3
-1

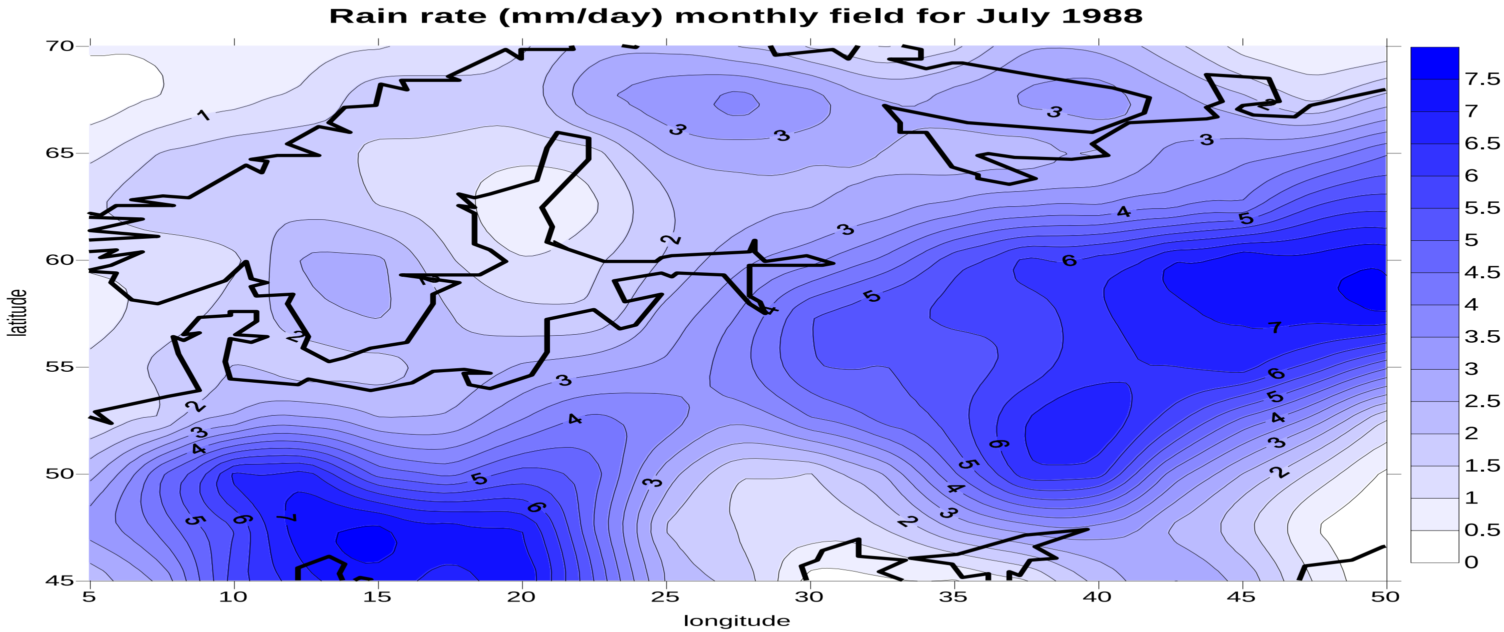

a) actual forecast error field (lead time - 4 months)
May, 1998
Latitude
Longitude
70
65
60
55
5
10
15
20
25
30
35
40
0.9
0.7
0.5
0.3
0.1

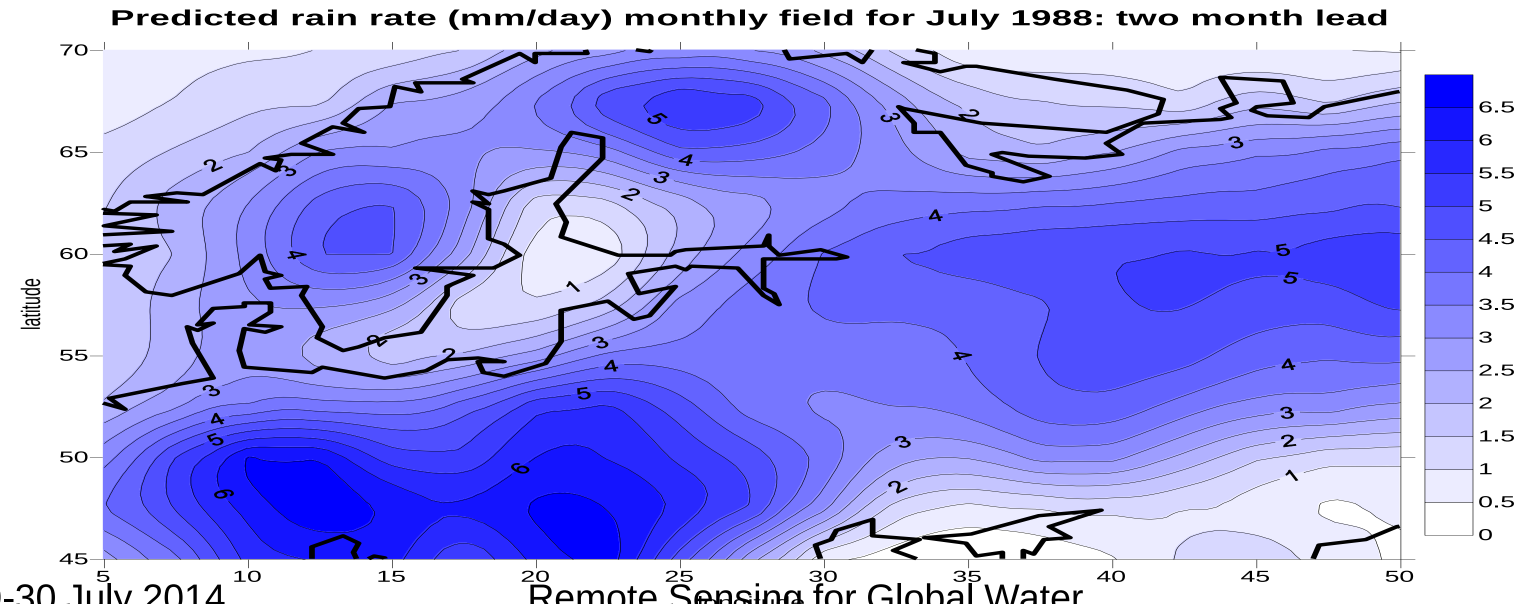

b) theoretical forecast error field (lead time - 4 months)
Latitude
Longitudes
0.85
0.65
0.45
0.25

Monthly Rain Rate Field Seasonal Prediction for July 1988


Rain rate (mm/day) monthly field for July 1988
latitude
longitude


Predicted rain rate (mm/day) monthly field for July 1988: one month lead
latitude
longitude


Rain rate (mm/day) monthly field for July 1988
latitude
longitude
70
65
60
55
50
45
5
10
15
20
25
30
35
40
45
50
7.5
7
6.5
6
5.5
5
4.5
4
3.5
3
2.5
2
1.5
1
0.5
0


Predicted rain rate (mm/day) monthly field for July 1988: two month lead
latitude
longitude
70
65
60
55
50
45
5
10
15
20
25
30
35
40
45
50
6.5
6
5.5
5
4.5
4
3.5
3
2.5
2
1.5
1
0.5
0

Conclusions

- **First attempt to develop an intellectual climate model has been carried out**
- **Its principal distinctive features are as following:**

1. **Self-learning ability**
2. **Accumulation of past observing information with saving all changes in inter-parameter links**
3. **Adaptation of model feedbacks to changes in observing samples and its trends**

THANK YOU
FOR YOUR ATTENTION

St. Petersburg

THANK YOU
FOR YOUR ATTENTION

St. Petersburg