Climate classification in the Northern Hemisphere using envelopes of temperature signals

Y V Volkov1,2, V A Tartakovsky1, N N Cheredko1, V G Maximov1, M M Kabanov1 and S N Kapustin1
1Institute of Monitoring of Climatic and Ecological Systems SB RAS
Akademichesky Ave., 10/3, Tomsk, 634055, Russia
2Tomsk Polytechnic University, Lenina Ave., 30, 634050, Tomsk, Russia
Tel. +7 382 249 2265. Fax +7 382 249 1950. sun-effect@list.ru

Abstract. Based on an analytic signal theory, the authors have developed an iterative algorithm for climate clustering using a temperature oscillation analysis. Surface temperature is selected as an integrated climate change indicator. Temperature series are studied as modulated signals. The algorithm enables signal grouping on various spatio-temporal scales using the available information on the synchronicity of envelopes of the signals.

1. Introduction
The climate changes accompanied by a serious transformation of geophysical fields are in the focus of attention all over the world. The consequences of these changes are the changing environmental conditions influencing the development of the modern civilization. It is known that climate changes are characterized by high spatial heterogeneity, especially outside the tropical zones. This fact complicates climate modeling and prediction.

The significance of the research of time-space shifts of the climate parameter fields during global natural changes increases [1-4]. The major source of all transformations in the geosphere is the energy flux of the Sun. As a result of a combination of the contributing factors, climate zones are formed on the Earth’s surface. Classification is considered to be one of the most recognized ways to reveal the structure of any system. It would be natural to assume that the climate system structure should undergo transformations with the global climate changes. Accelerated climate changes over the last decades increase the requirements for credibility of climate classifications.

The quality of any classification is defined by the level of separation between the structures obtained, to what extent the groups distinguished reflect the inner logic and patterns of the problem being solved, contribute to achieving analysis goals, and help to discover new properties of the studied objects.

Methods that are most frequently used to solve problems like this are cluster and factor analysis [3, 5-7]. Approaches to the selection of clustering criteria mostly come down to determining a maximum likelihood of parameters within a class and maximum discrimination between the classes. An overview of major quantitative methods for estimation of the quality of clusterization can be found in [8]. A shortcoming of these methods is the lack of a rigid formalized approach to determination and estimation of the classifications criteria. For example, a certain level of arbitrariness is introduced by predefining the resulting number of classes. In the method we suggest the number of classes is not predefined. The result reflects a true picture of congruence of oscillations in the parameter being studied. A researcher obtains natural structures. By varying the separation criteria it is possible to
obtain a field structure according to the required level of temporal or spatial scale of the contributing factor.

There are works on determining the bounds of climate clusters based on distinguishing uniform average annual temperature oscillations, taking into account the sign and phase of an oscillation [9]. These works do not elaborate the concept of uniformity and do not mention explicit calculation of the amplitude and phase.

Earlier we have developed a climate clustering algorithm using the phase of temperature oscillations [10]. This paper describes clustering using envelopes of temporal series, which are, much like phases, introduced by constructing analytic signals from the source data. A crucial advantage of the developed approach is its objectivity and the capability to account for the specifics of variability using the parameters of time series oscillations. Notably, there is no need to know the inner workings forming the oscillation date patterns. The result is non-overlapping geographic areas, where the change dynamics of the studied parameter is most congruent within the area, while there is a high degree of discrimination between different areas.

2. Data series and calculations

The calculations used monthly average surface temperature values measured at 818 meteorological stations in the Northern Hemisphere from 1955 to 2010, i.e. 56 years (https://crudata.uea.ac.uk/cru/data/temperature).

The surface temperature is an integrated indicator of climate changes. Due to Earth’s revolution around the Sun, temperature values averaged over short time periods are changing with a period equal to one calendar year. These oscillations can be described in terms of the amplitude and phase [11].

Figure 1(a) demonstrates the original temperature signal $T_e(t)$ obtained at the weather station «Tomsk» (Russia). Figure 1(b) demonstrates the Fourier spectrum $|FT|$ of this signal. Numerical implementation of the Fourier transform operator was made using the FFT algorithm.

A consistent definition of the envelope and phase is given based on the analytic signal (AS) introduced in 1946 by D. Gabor [11]. The analytic signal $W(t)$ is a complex function representing the temperature oscillation $T(t)$ in time as a natural generalization of harmonic oscillations:

$$ W(t) = T(t) + iV(t), \quad V(t) = \frac{1}{\pi} \text{v.p.} \int_{-\infty}^{\infty} \frac{T(s)}{t-s} ds, $$

where the improper integral is defined in the sense of the Cauchy principal value (v.p.) in cases when $s \to \pm \infty$ and for $t=s$. The imaginary component of the analytic signal $V(t)$ is the Hilbert transform [12-14] of its real component $T(t)$. Introduction of $W(t)$ allows defining the amplitude and phase in the following well-known way:

$$ a(t) = \sqrt{T^2(t) + V^2(t)}, \quad \Phi(t) = \arctg \frac{V(t)}{T(t)}, \quad \Phi(t) = 2\pi v_t + \varphi(t). $$

Here the function $a(t)$ describes the law of signal envelope change which will be used for analysis, and the function $\Phi(t)$ is a full phase determined as a principal value of the arctg function and can be sewn by continuity. For a narrow-band signal the continuous phase $\Phi(t)$ is always realized and changed monotonically.

From a computational viewpoint, the Hilbert transform is equivalent to multiplication in the frequency area ($\nu$) by the function $i\text{sgn} \nu$ and, therefore, AS corresponding to this real-valued function $T(t)$ is obtained by zeroing one half of this function’s Fourier spectrum. Such one-sided spectrum is called casual.
Figure 1. Temperature signal and its properties: (a) original temperature signal $T(x)$ obtained at the weather station «Tomsk» (Russia; synoptic index: 29430; latitude: 56.5°N; longitude: 84.9°O; altitude: 139 m); (b) normalized modulus of signal $|FT|$ from 1955 to 2010. Spectral frequencies are shown in observation interval fractions, first 224 readings of 2048 are shown. Spectral interval $\nu$ is shown by dark color, (c) filtered temperature signal $T(x)$ and envelope $a(x)$.

This operation is valid not for narrow-band signals only, but in more general cases too, when $T(x)$ is essentially a double-sideband signal, i. e. when the carrier frequency $\nu_c$ is only so high that the spectra of the functions exp$i2\pi\nu_c x$ and $a(x)$exp$i\phi(x)$ do not overlap. For a real signal this condition is fulfilled approximately. In our case the signal $T(x)$ is composed artificially, and so it is always possible to pick out a spectral window $\Omega(\nu)$ so that the phase $\Phi(x)$ will be monotonous. However, to retain useful information the spectral window size should be picked as wide as possible. The result will depend on specific meteodata, and the criterion of effectiveness of performed operations will be reasonableness of an obtained climate classification.

According to formula (2), envelopes $a(x)$ are calculated for frequency filtered signals (Figure 1b). Observation data represented by the envelope functions $a(x)$ form a set $L$ whose elements correspond to meteo-stations $j = 1, 2, ..., N$ or $k = 1, 2, ..., N$; $N = 818$. The procedure of classification consists in calculating the correlation matrix $r_{j,k}$ of $L$ set elements and consequent separation of this set into groups $\hat{G}$ with numbers $\nu$ such that the correlation coefficient of its elements within the group
\[ r_{j,k} = \langle L_j, L_k \rangle \geq \tilde{r}, \text{ where } \tilde{r} \text{ is the set correlation level}. \]

For the elements of the set \( L \) in the group \( n \Gamma \) the current value is calculated by replacing the original value by the average value within the group; this is performed for each group. Then the calculation of the \( \| r_{j,k} \| \) matrix is repeated, and groups are formed again. Iterations stop once the current values and the average values in groups differ less than by a predefined small value \( \epsilon \). In the limit, a certain number of groups consisting of complex elements are formed. Thus, for each group a limit value is determined, which we will call a typical element of a group. The described climate clustering algorithm is general (3) and, hence, the typical element is complex by definition, it consists of a typical envelope and a typical phase:

\[ \chi^n + i \phi^n \in \tilde{G}. \] (3)

The iterative process convergence follows from two conditions: the sample finiteness (the number of meteo-stations equals \( N \)) and validity of the inequality \( \| L_j \| \leq \max \| L_i \| \). Therefore, in the iterative process the current values of elements from any sample subset tend to a limit, to typical elements. Typical element values are determined by the correlation of the original and current elements involved in the iterative process and by the predefined correlation level \( \tilde{r} \) within a group.

In the climate clustering algorithm the value of the correlation coefficient \( \tilde{r} \) defines the level of granularity of clusters and, hence, their total number. The maximum number of clusters corresponds to the number of original estimates of the envelopes. Typical estimation signifies some consistent pattern in the behavior of the processes being studied and is an attribute of a separate climate cluster.

3. Climate classification
The key feature of the proposed classification method is the fact that the result has the form of a set of structure states of the field being studied obtained at different stages, for different tasks, depending on the research goals. Minimal levels of the threshold correlation coefficient \( \tilde{r} \) allow analyzing the large-scale structure of the studied field and estimating the global climate forcing and the contributing factors. Increased threshold level results in a more complicated structure, but the higher the threshold, the more influence of the local specifics on climate is reflected in the result. Nevertheless, the coherence and discrimination for the classes obtained remain high.

The proposed algorithm was used to obtain a climate classification in two spatial scales corresponding to the correlation coefficient intervals of \([0.8; 1]\) and \([0.4; 1]\). As a result of grouping of the surface temperature data in the correlation coefficient interval of \([0.8; 1]\), the original space reduced from 818 to 16 groups (Figure 2b). In the interval of \([0.4; 1]\), the original space reduced to 4 groups (Figure 2a). Comparing the classification schemes in Figure 2, it can be noticed that the part of the groups formed with a higher value of the \( \tilde{r} \) criterion could be fully or partially included into groups formed with a lower value of \( \tilde{r} \). This likelihood suggests a common influence and local specifics affecting the temperature signals. The cluster number in Figure 2a is defined by the cluster number 2b containing the largest number of elements (and taking up more territory) and closest in location.

In the scheme of surface temperature clustering in the Northern Hemisphere with \( \tilde{r} = 0.4 \) we can see a latitudinal distribution of the climate clusters (transfer from West to East), with a fairly distinctive separation of the borders (Figure 2(a)). Finer groups are formed with \( \tilde{r} = 0.8 \) (Figure 2(b)).

With a predefined envelope correlation level \( \tilde{r} \) the iterative process implemented in the climate clustering algorithm converges to a finite number of “typical” envelopes. Each “typical” function is unique. The correlation values between “typical” envelopes do not exceed 0.53 with \( \tilde{r} = 0.8 \) (Table 1) and 0.04 with \( \tilde{r} = 0.4 \) (Table 2). The least correlation value between “typical” functions provides for a better discrimination between the elements of different climate clusters.
Figure 2. Clustering scheme of a surface temperature field in Northern Hemisphere (1955-2010): (a) calculations with correlation coefficient interval from 0.4 to 1; (b) calculations with correlation coefficient interval from 0.8 to 1. Each climate class is labeled by a number from 1 to 18. Weather stations not assigned to any class are denoted by the “+” symbol.

Table 1. Correlation coefficients of “typical” envelopes obtained with $\hat{r} = 0.8$.

|   | 1   | 2  | 3  | 4  | 5  | 6  | 7  | 8  | 9  | 10 | 11 | 12 | 13 | 14 | 15 | 16 |
|---|-----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|
| 1 | 1   | -0.24 | 1 | 0.46 | -0.38 | 1 | 0.08 | -0.02 | 0.15 | 1 | -0.11 | 0.35 | -0.43 | 0.03 | 1 |
| 2 | 0.06 | -0.24 | 0.19 | -0.09 | -0.38 | 1 | 0.13 | -0.30 | 0.16 | 0.28 | 0.01 | -0.10 | 1 |
| 3 | -0.07 | 0.52 | -0.29 | 0.34 | 0.08 | -0.06 | -0.28 | 1 | 0.02 | 0.17 | 0.04 | -0.08 | 0.13 | -0.05 | 0.15 | -0.12 | 1 |
| 4 | -0.27 | 0.20 | -0.12 | 0.39 | 0.18 | -0.08 | 0.03 | 0.36 | -0.13 | 1 | 0.20 | -0.12 | 0.10 | 0.03 | 0.39 | -0.20 | 0.02 | -0.13 | -0.06 | 0.03 | 1 |
| 5 | 0.03 | 0.10 | 0.06 | -0.08 | 0.09 | 0.12 | -0.25 | 0.12 | 0.29 | -0.05 | 0.00 | 1 |
| 6 | 0.03 | 0.10 | 0.06 | -0.08 | 0.09 | 0.12 | -0.25 | 0.12 | 0.29 | -0.05 | 0.00 | 1 |
| 7 | -0.11 | 0.35 | -0.43 | 0.03 | 1 | 0.08 | -0.41 | 0.13 | 0.36 | -0.24 | 0.06 | 0.22 | 0.06 | 0.28 | -0.21 | -0.06 | 1 |
| 8 | 0.28 | -0.32 | 0.22 | -0.10 | -0.41 | 0.53 | 0.08 | -0.14 | -0.03 | -0.41 | -0.27 | -0.08 | -0.32 | 1 |
| 9 | 0.24 | -0.01 | 0.06 | -0.01 | -0.24 | 0.10 | -0.24 | 0.08 | -0.37 | -0.03 | 0.10 | 0.08 | -0.32 | -0.02 | 1 |
| 10 | 0.36 | 0.40 | -0.05 | 0.09 | 0.39 | -0.18 | 0.09 | 0.20 | 0.25 | -0.02 | -0.01 | 0.09 | 0.03 | -0.08 | -0.02 | 1 |
| # | 1   | 2   | 3   | 4   | 5   | 6   | 7   | 8   | 9   | 10 | 11 | 12 | 13 | 14 | 15 | 16 |
Table 2. Correlation coefficients of “typical” envelopes obtained with $\tilde{r} = 0.4$.

| # | 7   | 1   | 6    | -0.44 | 1   | 11   | -0.01 | -0.05 | 1   | 18 | 0.04 | -0.32 | -0.09 | 1 |
|---|------|------|------|-------|------|------|-------|-------|------|----|------|-------|-------|----|

The obtained envelope functions allow performing SD calculation for each of the 818 signals. The SD values are distributed among groups and shown in Figure 3(a) for $\tilde{r} = 0.8$ and in Figure 3(b) for $\tilde{r} = 0.4$.

The minimal absolute value of envelope standard deviation sufficient for climate clustering is less than 0.01°C.

Figure 3. Standard deviations of 818 temperature signal envelopes distributed by climate cluster numbers obtained for: (a) $\tilde{r} = 0.8$, (b) $\tilde{r} = 0.4$.

4. Conclusions
The above-proposed approach can be used as an analytical basis for climate research in any spatial scale based on surface temperature data.
An envelope as a characteristic of a temperature oscillation allows us to objectively distinguish the climate clusters and enables further dynamic climate clusterization, which is important in a changing geosphere. The above application of the climate clustering algorithm resulted in a distinctive geographical allocation of the climate clusters. The climate clusters correspond to the well-known concepts of climate geography, hence confirming the physical validity of this new approach.

Acknowledgements
This research was supported by the Russian Academy of Sciences (project IX.138.2.4).

References
[1] Peel M C, Finlayson B L and McMahon T A 2007 Updated world map of the Köppen-Geiger climate classification Hydrocl. and Earth Sys. Sc. 11 1633–44
[2] Franziska H, Körper J, Spangehl T and Cubasch U 2012 Shifts of climate zones in multi-model climate change experiments using the Köppen climate classification Met. Zeitsch. 21 (2) 111–23
[3] Zhang X and Yan X 2014 Spatiotemporal change in geographical distribution of global climate types in the context of climate warming Clim. Dyn. 43 (3-4) 595–605
[4] Rohli R V, Joyner T A, Reynolds St J, Shaw C and Vazquez J R 2015 Globally Extended Köppen-Geiger climate classification and temporal shifts in terrestrial climatic types Phys. Geogr. 36 (2) 142–57
[5] DeGaetano A T 2001 Spatial grouping of United States climate stations using a hybrid clustering approach Int. J. Climatol. 21 791–807
[6] Shergertukov B G 2008 Regionalnye i sezonnye zakonomernosti izmenenij sovremennogo klimata (Obninsk: GU VNIIGMI-MCD Press) p 247
[7] Zakusilov V P and Zakusilov P V 2009 Ispolzovanie komponentnogo analiza dlya harakteristik atmosfernoj cirkulyacii nad zadannym geograficheskim rajonom Vest. VSU. Geogr. Geokol. 2 67–71
[8] Sivogolovko E V 2011 Metody ocenki kachestva chetkoj klasterizacii Comp. Tools in Educ. J. 4 (96) 14-31
[9] Salugashvili R S 2012 Climate fluctuations within the First Natural Synoptic Area and climatic zoning Uch. Zapiski KSU Estest. Nauki 154 (3) 216–27
[10] Cheredko N N, Tartakovsky V A, Krutikov V A and Volkov Y V 2017 Climate classification in the Northern Hemisphere Using Phases of Temperature Signals Atmos. Ocean. Opt. 30 (1) 63–9 doi: 10.1134/S1024856017010043
[11] Gabor D 1946 Theory of communication J. IEE. 3 (3) 429-41
[12] Vakman D E and Vajnshtejn L A 1977 Amplituda faza chastota osnovnye ponyatiya teorii Kolebanij Uspekhi fiz. Nauk. 123 (4) 657–82
[13] Gonorovskij I S 1986 Radiotekhnicheskie cepi i signaly (Moscow: Radio i svyaz) p 512
[14] Tartakovsky V A 2002 Causality and demodulation of optical monotone-phase signal Atmos. Ocean. Opt. 15 (1) 78–86