Clustering of small watersheds over annual precipitation data reveals sounding correspondence to the cluster pattern determined by annual temperature course

M G Erunova¹,⁴ and M G Sadovsky²,³

¹Federal Research Center Krasnoyarsk Science Center of the SB RAS, Akademgorodok, 50, Krasnoyarsk, 660036, Russia
²Institute of Computational Modelling SB RAS, Akademgorodok, 50/44, Krasnoyarsk, 660036, Russia
³V.F.Voino-Yasenetsky Krasnoyarsk State Medical University, Partizana Zheleznyaka str., 1, Krasnoyarsk, 660022, Russia
⁴E-mail: marina.erunova@gmail.com

Abstract. Previously, highly informative clustering of small watersheds provided over the annual temperature course has been reported. Here we report on the interplay between the pattern and the distribution of small watersheds developed over the annual precipitation data. A sounding correlation between these two cluster structures is found. To reveal the correspondence, smart methods of both spatial and comprehensive analysis have been implemented. We used advanced GIS tools to detect the spatial behavior of the data under consideration, and elastic map technique to reveal the cluster patterns in the space of environmental data (temperature and precipitation). Some further applications and perspective studies in this area are discussed.

1. Introduction

A comprehensive analysis of spatially distributed environmental processes through Earth remote sensing (ERS) data implementation requires a determination of an elementary entity to be taken into consideration. The approach based on pixel resolution analysis of spatially distributed environmental processes presented in [1] exhibits some disadvantages; in particular, a pixel is not a naturally identified element of a geographical site, thus providing a noise and artificially generated false signaling together with the loss of comprehensive data. Alternatively, the approach based on naturally defined entities that are the smallest watersheds observed in a river basin system is reported in [2]; this approach shows high efficiency and feasibility, for a number of geographical tasks.

Temperature and precipitations are among the key abiotic factors in dynamics of an ecosystem. The impact of them may not be excluded from any study, not it might be controlled by a researcher. Hence, these two factors must be incorporated into any study and are stipulated to be a kind of “environmental template” determining a number of other related processes. To do that, one ought to reveal a kind of an order in an ensemble of considerably small territory units: whether such units comprise greater consortia and if yes, then whether the consortia are homogeneous from the point of view of the environmental properties of those small units, is a key question in any research of that type. Previously [2], small watersheds have been shown to be very suitable and valuable entities to form a pattern in terms of their...
geographical and environmental properties, if clustered over the annual temperature course. This approach opposes the pixel-based clustering [1], unambiguously approving the advantages of that former.

Here we present some further results in modeling of land classification based on basin identification approach, for Krasnoyarsk region. MERIT Hydrologically Adjusted Elevations digital relief model accompanied with the corresponding spatial database implementation [3 – 5] was used to do the research.

We identified four levels in the tree pattern, and annual course of the surface temperature has been determined, over ERS data, for each watershed of the lowest (the fourth one) level. The temperature records have 7 days gaps between the measurements so that each basin was supplied with the mean day temperature averaged over the basin. Simultaneously, the watersheds were supplied with the precipitation records. That latter follows the records of temperature: it also has 7 days gaps in observations. Water basins identified within the territory were arranged hierarchically: the nodes correspond to basins of rivers, and edges correspond to the tributary structure. This classification follows the morphometric features of relief in river geosystem. To reveal the inner structuredness, an advanced statistical analysis of the multidimensional data was implemented. The aim of this paper is to compare two patterns provided by clustering: the former is provided by annual temperature course, and the latter is provided by annual precipitation course observed over the same watersheds.

2. Material and methods

Everywhere further, for the purposes of the study we used MODIS data, as well as precipitation records, were gathered in 2019, only. For each watershed enlisted in the database, the annual course of the surface temperature has been determined, over ERS data. These data are provided by MOD11A2 product presenting MODIS averaged data of land surface temperature and emissivity over 8 days. Firstly, daily average temperature figures were obtained from these ERS data, for each pixel falling into a watershed. Next, the eight-day average temperature has been calculated over a watershed. Thus, so we studied the distribution of the watersheds in the metric space provided by this annual data record, and compared it to other land use features. Formally, an annual temperature course record must comprise 49 entries; however, we have to exclude four dated from the analysis since there is a lot of lacunae in the data records obtained for these days. The following dates are excluded from the record: 25-th, 201-th, 297-th and 345-th days; they correspond to January 25, 2019, July 20, 2019, October 24, 2019 and December 11, 2019.

Elastic map technique [6 – 8] has been used to reveal cluster pattern. The method decreases data dimension, clusters and visualizes them consisting in approximation of the multidimensional data with a manifold of two-dimensional manifolds (a square and a sphere). An implementation of the method is described in [6 – 8]; the specific version of that latter is also presented in [2, 7].

3. Results

Previously [9], the cluster pattern observed through the annual temperature course data recorded for considerably small watersheds, for two middle-range revers has been reported. Figure 1 shows the new one pattern developed over the data of annual course of precipitations. Six clusters are apparent in this figure. Two clusters are subdivided into two subclusters each. These are the clusters shown in figure 1 in red and pink (left upper corner of the elastic map), and in sapphire and cerulean (central part of the bottom of the map, in figure 1). Finally, small white circles correspond to the watersheds that fall beyond any cluster.

The left subfigure in figure 1 shows the distribution of the smaller watersheds belonging to the basins of the river of Kan (blue triangles), the river of Mana (red diamonds), the river of Tuba (yellow circles) and the river of Oya (parakeet diamonds), respectively.
Figure 1. Cluster pattern of the distribution of 573 watersheds (triangles) in the 49-dimensional metric space of annual course of precipitation. The distribution of the watersheds of four main rivers is shown in left and the clusters are shown in right, see text for details.

Figure 2. Correspondence between clustering developed over annual precipitation data (A) and clustering developed over the annual temperature data; see explanation in text.

We studied cluster pattern to be revealed through the analysis of annual precipitation record, for 2019. Previously, similar pattern developed over the annual temperature course measured with 7 days gaps, 45-dimensional metric space also has been developed [2, 9]. The inconsistency of the space
dimensionality, for these two patterns, results from the severe lacunae observed in the temperature annual course record coming from the cloudiness, in a day of observation.

Figure 2 shows the precipitation-based clustering pattern (figure 2A) vs. the temperature-based clustering pattern (figure 2B). Let us focus on these two patterns. First of all, the color labelling system is absolutely independent, in these two figures, and one should not derive any correspondence based on the color attribution; same is true for the number of clusters observed on elastic map. First of all, the temperature yields significantly more detailed clustering, if compared to that one developed over precipitations: ten clusters against six ones. It should be stressed that in spite of the map legend indicates formally eight clusters for precipitation data, we deliberately split two clusters into subclusters: theses are the first and the fifth clusters labelled in the legend as 11 and 12 subclusters (51 and 52 subclusters, respectively). An idea standing behind such split is rather clear; let now change for figure 1. To cluster the watersheds, we used local density clustering technique. Figure 1 shows the elastic map with local density charts drawn out with contrast parameter $\mu = 0.15$. A growth of that parameter to $\mu = 0.20$ results in a merge of these subclusters into a single one each. Keeping in mind that we have used very high contrast parameter value, we split the clusters into a couple of subclusters each.

4. Discussion

Even a short glance at figure 2 reveals a considerable interrelation between the clusters identified over the temperature, and those identified over precipitations. Obviously, the detail level for temperature data exceeds that one for precipitation data. A sounding smoothness in the precipitation clustering pattern is another essential feature differing that latter from the temperature clustering. This fact strongly correlates to various observations towards the location and density of whether observatories necessary to gather the temperature and precipitation data with proximal accuracy [10 – 12]; indeed, there is a well-known fact that one needs less number (and density) of precipitation observatories, in comparison to those for temperature records, to get the same accuracy in observations. Figure 2 brings another evidence of that fact.

Evidently, these two cluster patterns yield a kind of “two dimensions” where the interplay between various abiotic factors takes place, in the dynamics of such environmental systems. In such capacity, more detailed investigation of the interrelation between the patterns is needed, to reveal some further issues in the dynamics. In particular, one should investigate the composition of the clustering pattern developed over temperature, in comparison to that one developed over precipitation, and vice versa. Yet, this point falls beyond the scope of this paper. Similarly, various additional figures might be involved in such analysis, including physical geography ones (slope, aspect, elevation, etc.).

A smoothness of the cluster patterns observed for precipitation data poses a problem in clustering technique to be used for the study: indeed, the local density approach is very descriptive; however, it may bring few false signals. To be exact, it may mimicry some real “signal”; this effect is observed when there are two (or few) nearly located clusters of approximate local density, while different in abundance of the points to be comprised in them. A decrease of the sharpness (contrast) in elastic map in a cluster identification results in a loss of smaller cluster, not to the absorption of that latter. This discrepancy could be eliminated due to implementation of some other methods of clustering; DBSCAN seems to be the best candidate for that, if applied to inner coordinates of the elastic maps.

References

[1] Sardooi E R, Azareh A, Choubin B, Barkhori S, Singh V P and Shamshirban Sh 2019 Applied Geography 111 102071
[2] Erunova M and Sadovsky M 2020 IOP Conference Series: Earth and Environmental Science 548 032022
[3] Erunova M and Yakubailik O 2020 Zoning of the territory on the basis of morphometric analysis of basin geosystems IOP Conf. Ser.: Earth Environ. Sci. 421 062039
[4] Yamazaki D, Ikeshima D, Tawatari R, Yamaguchi T, O'Loughlin F, Neal J, Sampson C, Kanae S and Bates P 2017 A high accuracy map of global terrain elevations Geophysical Research
Letters 44 5844

[5] Yamazaki D, Ikeshima D, Sosa J, Bates P, Allen G and Pavelsky T 2019 MERIT Hydro: A high resolution global hydrography map based on latest topography datasets Water Resources Research 55 5053

[6] Gorban A and Zinovyev A 2015 Fast and user-friendly non-linear principal manifold learning by method of elastic maps IEEE International Conference on Data Science and Advanced (Campus des Cordeliers Paris) 1-9

[7] Akinduko A and Gorban A 2014 Multiscale principal component analysis Journal of Physics: Conference Series 490 012081

[8] Mirkes E, Zinovyev A and Gorban A 2013 Lecture Notes in Computer Science 7902 500

[9] Erunova M, Yakubailik O and Sadovsky M 2020 Annual temperature variation reliably identifies different sites in a large water basin IOP Conference series 862 062060

[10] Lei Jiang, Liqing Zhao and Zihao Zhao 2017 Advances in meteorology 2017 5761275

[11] Colston J M, Ahmed T, Mahopo C, Kang G, Kosek M, de Sousa Junior F, Shrestha P S, Svensen E, Turab A and Zaitchik B 2018 Environmental Research 165 91-109

[12] Knerr I, Trachte K, Garel E, Huneau F, Santoni S and Bendix J 2020 Atmosphere 11 417