Journal of Global Change Data & Discovery. 2017, 1(1): 66-73
DOI:10.3974/geodp.2017.01.10
www.geodoi.ac.cn

Global Change Research Data Publishing & Repository

Monthly Temperature Dataset of China at 1 km Resolution

Jing, W. L. Yang, Y. P.* Yue, X. F.

1. State Key Laboratory of Resources and Environmental Information System, Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences, Beijing 100101, China;
2. University of the Chinese Academy of Sciences, Beijing 100049, China

Abstract: Surface air temperature is an important parameter which controls land surface processes, and is crucial to ecological, environmental and hydrological modelling. Temperature records at meteorological stations have been widely used in modelling, but there has been an increasing use of temperature data in grid form, especially for remote sensing data. Although grid temperature can be estimated from in-situ temperature records using interpolation algorithms, low accuracy has been found in the estimations, due to limited distribution of ground stations, especially when there are insufficient sites to represent all land-cover types and terrain conditions in the area. The NCEP/NCAR reanalysis project has produced climate variables by using a "frozen" state-of-the-art global data assimilation system and a database as complete as possible. Although resolution of the NCEP data has a coarse resolution (0.5 degree), it provides global, consistent, and long-term estimation of climate variables. This dataset presents a downscaling approach for deriving monthly temperatures at 1 km resolution from the NCEP by utilizing derived relationships between monthly aggregated NCEP temperatures and other ground elements, i.e., terrain, vegetation and geographic locations. A regression tree model was chosen for detecting the possible relationships. Monthly temperatures at 1 km resolution during 2000 to 2010 of China`s land area have been produced using this approach. The final predicted temperatures were compared with observed records at 380 meteorological stations in China. The results indicated that the downscaled estimations can represent the spatial distribution and trends and the magnitude of temperatures on an inter-month basis, with R^2 values ranging from 0.861 to 0.948, and RMSE values from 1.88ºC to 2.68ºC.

Keywords: **monthly average temperature; downscaling, NCEP/NCAR, NDVI**

1 Introduction

Surface air temperature is an important parameter reflecting surface environmental conditions, and is a basic observational parameter in the meteorological dataset [1,2]. Most land surface processes are controlled by near-surface air temperatures, which makes the air sur-

Received: 20-12-2014; **Accepted:** 15-03-2015; **Published:** 25-03-2017
Corresponding author: Yang Y. P. D-4110-2017, Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences, yangyp@igsnrr.ac.cn
Article Citation: Jing, W. L., Yang, Y. P., Yue, X. F.. Monthly temperature dataset of China at 1 km resolution [J]. *Journal of Global Change Data and Discovery*, 2017, 1(1): 66-73. DOI:10.3974/geodp.2017.01.10.
Dataset Citation: Jing, W. L., Yang, Y. P., Feng, M., et al.. Historical Monthly Temperature Data of China at 1 km resolution [DB], *Global Change Research Data Publishing & Repository*, DOI:10.3974/geodb.2015.01.02.V1.

face temperature an important surface meteorological input factor in various kinds of models of plant physiology, ecology, environment and hydrology [1]. It would thus be expected that achieving highly accurate high-resolution distribution data of air temperature near the surface would contribute to gaining an improved understanding of land surface processes and global climate change[1-4].

Air temperature distribution near the surface can be realized through temperature records at meteorological stations and using remote sensing data. Temperature records at meteorological stations have been widely used, but rasterized temperature data has increasingly been required for modeling, especially for those studies using remote sensing data. Although grid temperature can be estimated from observed temperatures at stations using interpolation algorithms[5,6], the estimations typically have low accuracy, because this approach depends heavily on the quantity and distribution of available meteorological stations; also, the distribution density of meteorological station is not regular, most stations being located in flat and inhabited regions, there being few stations in sparsely populated regions or with complex topography[5-8]. Another issue is that the observation data at meteorological stations may only represent temperatures near the location of the stations, and therefore in complex terrain or regions with different landscapes, the extent that the meteorological station data can be representative is very limited [1].

Remote sensing technology has greatly improved the resolution and quality of temperature observation[2]. However, due to technical constraints, satellite-based thermal sensors are unable to provide both spatially and temporally dense near-surface air temperature imaging data. The NCEP/NCAR global reanalysis products produced by the National Centers for Environmental Prediction (NCEP) and the National Center for Atmospheric Research (NCAR) use a "frozen" state-of-the-art global data assimilation system and a database as comprehensive as possible, which includes land surface, ship, rawinsonde, aircraft, satellite and other temperature datasets. The resolution of the global reanalysis datasets is 0.5 degree for longitude and latitude and results are available at 6-hour intervals [9-11]. The NCEP/NCAR monthly average reanalysis products are reliable and the long-time-series product is widely applied in climate diagnostics, hydrological cycles and many other research studies[11-15]. However, resolution of the NCEP/NCAR reanalysis product can hardly meet the requirement of many regional researches or those requiring high spatial resolution. Downscaling methods have been used for producing high-resolution climate data from coarse data[16-21].

Surface air temperature is related to many other natural factors, such as topography and vegetation. A close correlation between inter-month NDVI and temperature has been recognized in the northern mid- to high-latitude areas. In this study, on the basis of there being close relationships between monthly average temperature, topography, vegetation factors and geographical locations, statistical downscaling models were built based on the NCEP/NCAR monthly average temperatures, MODIS NDVI and GLS DEM, and applied to the land area of China to achieve a time series from 2000 to 2010. In this way, monthly average temperatures at 1 km resolution were obtained [22].

2 Metadata of the dataset

The descriptions of the "ChinaTemp" dataset are recorded. This information includes the full name of the dataset, the abbreviated name, the authors and the corresponding author, the geographical regions for the dataset content, the year of the dataset, the spatial and temporal resolution, the number of the dataset files, the data format and size, the data files, the foundations, the data publisher, and the data sharing policy. Table 1 below summarizes the main metadata elements of the dataset.

Table 1 Metadata summary of the ChinaTemp dataset

Items	Description		
Dataset full name	Monthly temperature dataset of China at 1 km resolution		
Dataset short name	ChinaTemp		
Authors	Jing, W. L. D-2691-2017, IGSNRR/CAS, jingwl@lreis.ac.cn Yang, Y. P. D-4110-2017, IGSNRR/CAS, yangyp@igsnrr.ac.cn Yue, X. F. D-4113-2017, IGSNRR/CAS, lexf@igsnrr.ac.cn		
Geographical region	Mainland of China		
Time	2000 to 2010; Temporal resolution: Monthly		
Spatial resolution	1 km	Number of tiles	132
Data format	.tif	Data Size	4096MB
Data files	The dataset consists of 132 files and one attached dataset. These consist of the monthly temperature data file and the validation dataset. 1. tmp_yyyy_mm.tif, this is the monthly temperature data. Data size is 30720 KB. 2. tmp_new.xls, this is validation results using observation of 380 stations. 3. sthinfo.csv, this the name and locations of the 380 stations.		
Foundation(s)	Ministry of Science and Technology of P. R. China (STSN-04-08)		
Data Publisher	Global Change Research Data Publishing & Repository, http://www.geodoi.ac.cn/.		
Address	No.11A, Datun Road, Chaoyang District, Beijing 100101, China		
Data sharing policy	***Data*** from the Global Change Research Data Publishing & Repository includes metadata, datasets (data products), and publications (in this case, in the *Journal of Global Change Data & Discovery*). ***Data*** sharing policy includes: (1) ***Data*** are openly available and can be free downloaded via the Internet; (2) End users are encouraged to use ***Data*** subject to citation; (3) Users, who are by definition also value-added service providers, are welcome to redistribute ***Data*** subject to written permission from the GCdataPR Editorial Office and the issuance of a ***Data*** redistribution license, and; (4) If ***Data*** are used to compile new datasets, the 'ten per cent principal' should be followed such that ***Data*** records utilized should not surpass 10% of the new dataset contents, while sources should be clearly noted in suitable places in the new dataset [23].		

3 Data resources and methods

3.1 Data resources

The NCEP/NCAR global reanalysis products are produced by the NCEP and the NCAR. The NCEP/NCAR reanalysis products are widely used for research on climate diagnostics, hydrological cycle, etc. The data used in this study were the NCEP/NCAR reanalysis monthly surface temperatures and precipitation data, covering the time series from 2000 to 2010 and featuring a spatial resolution of 0.5 degree in geographical (latitude and longitude) projection. The original data were re-projected to the Albers equal-area conic projection using the nearest neighbor resampling algorithm.

The Normalized Difference Vegetation Index (NDVI) is a measure of the vegetation activity and biomass[24]. A close correlation between the inter-month NDVI and temperature has been recognized for the northern mid- to high-latitude areas[25-29]. NDVI products have now been extended to give global coverage with high resolution. MODIS NDVI data used in this research and covering the time series from 2000 to 2010 with sinusoidal projection were re-projected to the Albers equal-area conic projection, and the nearest neighbor resampling algorithm was used to resample MODIS NDVI images and with a pixel size of 1 km by 1 km being retained during the re-projection procedure. The MODIS NDVI data have some noticeable regular bands in some very low NDVI regions. The land cover type of these regions is mainly desert, differences of the NDVI having little impact on the distribution of temperature and precipitation. To eliminate the effect of the bands when sampling and modeling, the NDVI values that were less than 0.1 were replaced with the value of 0.1 to constrain the NDVI value to the range 0.1 to 1.0.

The Global Land Survey Digital Elevation Model (GLSDEM) is a spatial data collection model that is crucial to new global imagery products from NASA and the USGS. The GLS collection was made possible by the development of more accurate elevation models with global or regional coverage, including the National Elevation Dataset (NED) of the USA, the Canadian Digital Elevation Dataset (CDED) and the Shuttle Radar Topography Mission (SRTM) which covers most of the globe. The elevation data were incorporated into the GLSDEM product, which has a resolution of 3 arc second (90 m) and in geographic coordinates the WGS-84 datum (http://glcf.umiacs.umd.edu/data/glsdem/).

3.2 Method

The method applied for this dataset is based on the assumption that the air temperatures near the surface have a close relationship with the inter-month NDVI, topography and geographical locations. A regression tree model was chosen for detecting the possible relationships between NCEP temperature and DEM, NDVI, and the geographical locations at the 50 km scale every month. The models were applied to DEM and NDVI data with 1 km resolution so that downscaled monthly mean temperatures were obtained. The specific downscaling process was as follows: (1) The NCEP surface air temperature was aggregated every 6 hours to a monthly temporal resolution. (2) The DEM (1 km) and the NDVI (1 km) were aggregated to the 50 km scale. (3) The DEM, the monthly NDVI at 50 km and the temperature derived from the NCEP were extracted to enable regression tree models to be established between temperature and NDVI, DEM, and geographical coordinates at the 50 km scale. (4) The models established at 50 km resolution were applied to the NDVI and the DEM at a resolution of 1 km, so that monthly predicted temperatures at 1 km resolution were achieved. (5) The predicted temperature was aggregated to the 50 km scale, by subtracting the data from the NCEP temperature at 50 km, to obtain the residual of the regression tree models. (6) The simple spline tension interpolation method was applied to obtain the residual at a resolution of 1 km; (7) by adding this residual to the predictive temperature at the 1 km scale, the final predictive temperature at 1 km resolution was obtained. A flow chart for the process is shown in Figure 1.

4 Dataset compositions, quality control and validation

4.1 Downscaled temperature data

The regression tree model was applied to the corresponding NDVI and DEM data to obtain the monthly temperature data with 1 km spatial resolution. The data were aggregated to the

NCEP/NCAR scale (50 km) using a simple averaging method. The averaged data were then subtracted from the NCEP/NCAR data. The residuals were interpolated using the thin spline technique to the 1 km scale, by adding this high-resolution residual to the predicted temperature and precipitation at 1 km resolution, the final downscaled monthly average temperature and monthly total precipitation of China at 1 km resolution being obtained. Figure 3 shows the downscaled temperature for May 2005.

To validate the accuracy of the final downscaled estimations predicted by the regression models established at the 50 km scale, the R^2, the MAE and the RMSE were calculated. Table 2 shows the validation results for the temperature derived from the NCEP data and the final predictive temperature at a resolution of 50 km. Figure 2 presents the scatter plots for

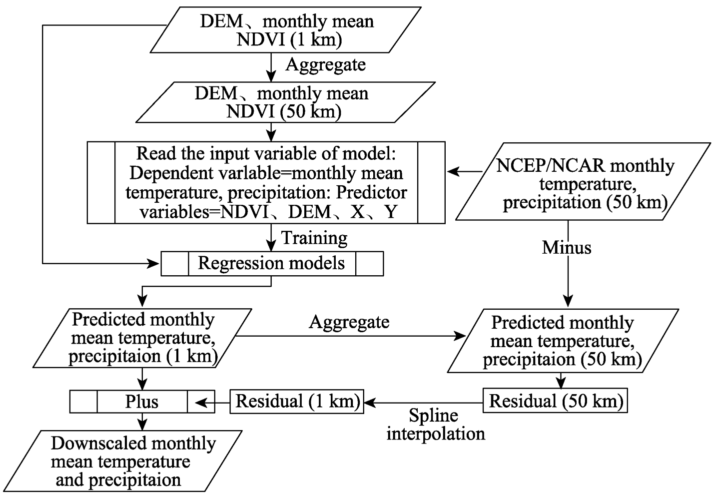

Figure 1 Flow chart of the temperature downscaling algorithm

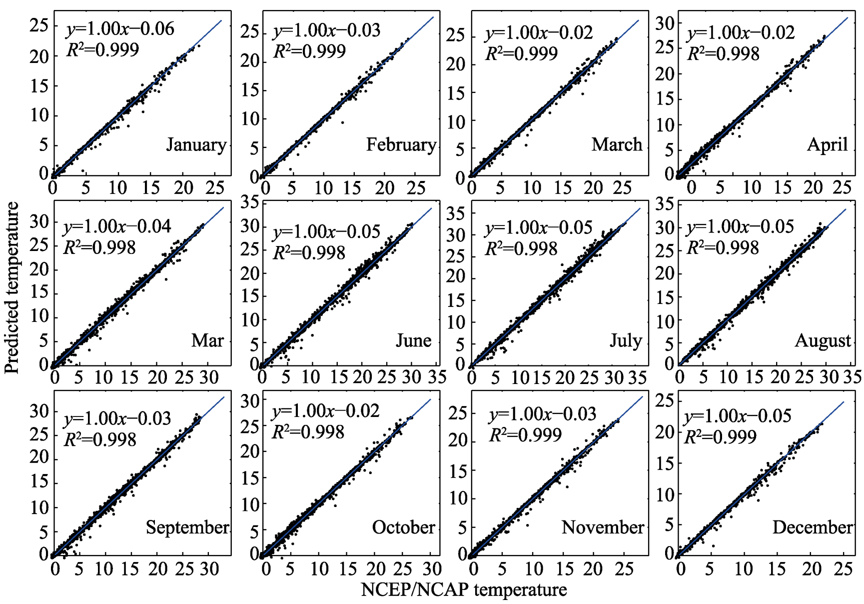

Figure 2 Scatter plots for the original NCEP temperature and predicted temperature at 50 km resolution for 2010

the original NCEP temperature and the predicted temperature at 50 km resolution for 2010. As shown in Table 2, the R^2 values for every month are all higher than 0.99, the RMSE range is from 0.315 to 0.391, and the MAE values are all less than 0.25ºC. Furthermore, the downscale temperatures are higher in relation to the original NCEP data.

Table 2 Comparison of results with the original NCEP/NCAR temperature data

	January	February	March	April	May	June	July	August	September	October	November	December
R^2	0.999	0.999	0.999	0.998	0.998	0.998	0.998	0.998	0.998	0.998	0.999	0.999
RMSE (°C)	0.349	0.330	0.362	0.381	0.391	0.381	0.354	0.331	0.315	0.333	0.366	0.341
MAE (°C)	0.181	0.173	0.191	0.204	0.220	0.217	0.202	0.189	0.178	0.180	0.195	0.180

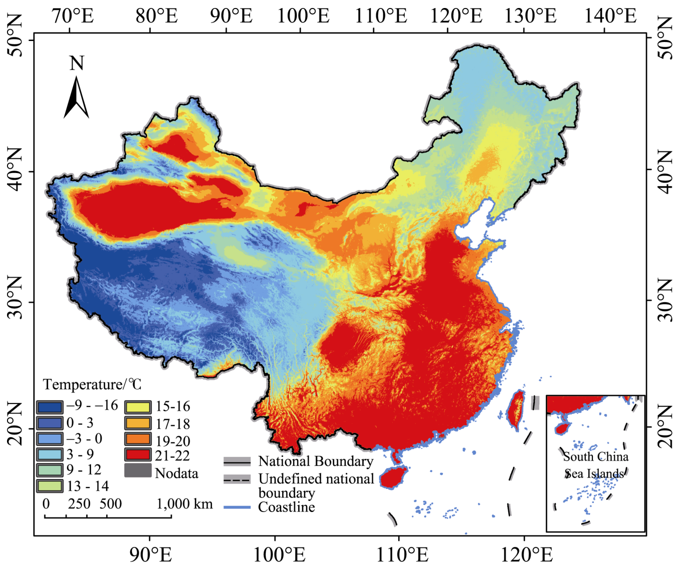

Figure 3 Downscaled temperature for May 2005

4.2 Validation of the original NCEP/NCAR temperature data

Although judging from the validation results presented in Table 2, the downscaled estimations are highly accurate for the original NCEP temperature at a resolution of 50 km, the observed temperatures from the meteorological stations were taken into account for further validation of estimations at 1 km resolution. The validation results for the downscaled temperatures at 380 stations for a 10-year period are shown in Table 3. The R^2 values range from 0.865 to 0.948. The RMSE and MAE values were higher during January to April than in other months, and values were relatively lower from May to December. Thus, the downscaled results are considered to be accurate and can effectively capture the trends of monthly average temperature distribution of China`s land area.

Table 3 Validation results for the downscaled results using observations

	January	February	March	April	May	June	July	August	September	October	November	December
R^2	0.941	0.938	0.922	0.902	0.890	0.865	0.861	0.869	0.886	0.925	0.948	0.950
RMSE (°C)	2.681	2.660	2.411	2.219	2.021	1.979	1.973	1.880	1.964	1.938	2.033	2.415
MAE (°C)	2.031	1.879	1.673	1.517	1.426	1.463	1.500	1.432	1.458	1.378	1.419	1.821

5 Discussion and conclusion

The surface air temperature is an important parameter having great impact on land surface processes. Moreover, temperature data with high resolution is crucial in many fields of research such as ecology, environment and hydrology. This study has investigated the downscaling algorithm for NCEP temperature by using a time series of NCEP 6-hour intervals of surface air temperature and monthly MODIS NDVI data from 2000 to 2010. DEM, NDVI and geographical locations were taken into account in the establishment of the regression tree models, such that interpolation resulted in the production of residual, monthly temperatures for the land area of China from 2000 to 2010 at 1 km resolution.

The downscaled results were validated based on the NCEP temperatures at the 50 km scale and meteorological stations throughout China. The results showed that the regression tree downscaling model performed very well and validation results based on the meteorological stations` data also showed strong agreement with the downscaled results; however, weaker agreement was found for the western part of China, where the terrain is more complex. Three points might explain this issue: (1) regression tree models, based on the 50 km scale in complex terrain areas, may not effectively reflect the characteristics of the terrain and the environmental vegetation factors in these regions; (2) the statistical downscaling model, the NCEP/NCAR temperature data and the final downscaled results were based on the regional temperatures, while meteorological stations represent just the temperature near the locations of the stations; in regions of complex terrain where stations are sparse, the uncertainties associated with such comparisons could potentially bias the validation results; and (3) given that NCEP/NCAR products involve reanalysis based on integrated multiple types of data sources, complex topography and physical process, combined with a lack of surface or upper air observations in complex terrains, the better accuracy of the NCEP/NCAR reanalysis products would inevitably be compromised [30, 31].

The final downscaled results can capture the trends in inter-month variability and the magnitude of monthly average temperatures. However, the reliability of the downscaling algorithm appeared to be influenced by both the accuracy of the NCEP/NCAR temperature fields and the complex terrain; hence there is a need for refinement and further improvement of the downscaling algorithms.

Author contributions

Yang, Y. P. designed the study. Jing, W. L. analyzed the results and wrote the paper. Yue, X. F. contributed to the data processing.

References

[1] Zhu, S. Y., Zhang, G. X. Progress in near surface air temperature retrieved by remote sensing technology. *Advances in Earth Science*, 2011, 26(7): 724-730. (In Chinese)

[2] Qi, S. H., Wang, J. B., Zhang, Q. Y., Luo, C. F., Zheng, L. Study on the estimation of air temperature from MODIS data. *Journal of Remote Sensing*, 2005, 9(5): 570-575.

[3] Cresswell, M. P., Morse, A. P., Thomson, M. C., Connor, S. J. Estimation surface air temperatures from Meteosta land surface temperatures, using an empirical solar zenith angle model. *International Journal of Remote Sensing*, 2010, 20(6): 1125-1132.

[4] Mao, K. B., Tang, H. J., Wang, X. F., Zhou, Q. B., Wang, D. L. Near-surface air temperature estimation from ASTER data based on neural network algorithm. *International Journal of Remote Sensing*, 2008, 29(20): 6021-6028.

[5] Liao, S. B., Li, Z. H., You, S. C. Comparison on methods for rasterization of air temperature data. *Resources*

Science, 2003, 25(6): 83-88. (In Chinese)
[6] Liao, S. B., Li, Z. H. Some practical problems ralated to rasterization of air temperature. *Meteorological Science and Technology*, 2004, 32(5): 352-356. (In Chinese)
[7] Cai, D. H., Guo, N., Li, C. W. Interpolation of air temperature based on DEM over eastern region of Gansu. *Journal of Arid Meteorology*, 2009, 27(1): 10-17. (In Chinese)
[8] Li, J., You, S. C., Huang, J. F. Spatial interpolation method and spatial distribution characteristics of monthly mean temperature in China during 1961-2000. *Ecology and Environment*, 2006, 15(1): 109-114. (In Chinese)
[9] Kalnay, E., Kanamitsu, M., Kistler, R., Collins, W., Deaven, D., Gandin, L., et al. The NCEP/NCAR 40-year reanalysis project. *Bulletin of American Meteorological Society*, 1996, 77(3): 437-471.
[10] Zhao, T. B., Fu, C. B. Preliminary comparison and analysis between ERA-40, NCEP-2 reanalysis and observations over China. *Climatic and Environmental Research*, 2006, 11(1): 14-32. (In Chinese)
[11] Huang, G. The Assessment and difference of the interdecadal variations of climate change in northern part of China with the NCEP/NCAR and ERA-40 reanalysis data. *Climatic and Envorinmental Research*, 2006, 11(3): 310-320. (In Chinese)
[12] Shi, X. H., Xu, X. D., Xie, L. A. Reliability analyses of anomalies of NCEP/NCAR reanalysis wind speed and surface temperature in climate change research in China. *Acta Meteorologica Sinica*, 2006, 64(6): 709-722. (In Chinese)
[13] Zhou, Q., Zhao, F. S., Gao, W. H. Comparison and analysisi between NCEP/NCAR every 6 hours analysis land surface and air temperature and 753 Chinese stations' observation in 2005. *Meteorology*, 2008, 34(2): 83-91. (In Chinese)
[14] Su, Z. X., Lu, S. H., Luo, S. W. The examinations and analysis of NCEP/NCAR 40 years gloal reanalysis data in China. *Plateau Meteorology*, 1999, 18(2): 209-218. (In Chinese)
[15] Xu, Y., Ding, Y. H., Zhao, Z. C. Confidence analysis of NCEP/NCAR 50-year global reanalyzed data in climate change research in China. *Quarterly Journal of Applied Meteorology*, 2001, 12(3): 337-347. (In Chinese)
[16] Liang, S. Quantitative remote sensing of land surfaces. Wiley-Interscience, 2004.
[17] Liu, Y. H., Guo, W. D., Feng, J. M. Zhang, K. X. A summary of methods for statistical downscaling of meteorological data. *Advances in Earth Science*, 2011, 26(8): 837-847. (In Chinese)
[18] Immerzeel, W. W., Rutten, M. M., Droogers, P. Spatial downscaling of TRMM precipitation using vegetative response on the Iberian Peninsula. *Remote Sensing of Environment*, 2009, 113: 362-370.
[19] Carone, G. J., Bramente, P. D. Translating monthly temperature from regional to local scale in the southeastern United States. *Climate Research*, 1995, 5: 229-242.
[20] Wigley, T. M. L., Jones, P. D., Briffa, K. R., et al. Obtaining sub-grid-scale information from coarse-resolution general circulation model output. *Journal of Geophysical Research*, 1990, 95(D2): 1943-1953.
[21] Tang, G. L., Ding, Y. H. Impacts of the average air temperature derived from maximum and minimum temperatures on annual mean air temperatures series of china. *Journal of Applied Meteorological Science*, 2007, 18(2): 187-192. (In Chinese)
[22] Jing, W. L., Yang, Y. P., Yue, X. F. 2015.Historical Monthly Temperature Data of China in 1km [DB], *Global Change Research Data Publishing & Repository*, DOI:10.3974/geodb.2015.01.02.V1.
[23] GCdataPR Editorial Office. GCdataPR Data Sharing Policy. DOI:10.3974/dp.policy.2014.05.
[24] Fan, N., Xie, G. D., Zhang, C. S., Chen, L., Li, W. H., Cheng, S. K. Spatial-temporal dynamic changes of vegetation cover in Lancang River Basin during 2001-2010. *Resources Science*, 2012, 34(7): 1222-1231. (In Chinese)
[25] Luo, L., Wang, Z. M., Song, K. S. Zhang, B., Liu, D. W., Ren, C. Y., Zhang, S. M. Research on the correlation between NDVI and climatic factors of different vegetations in the northeast China. *Acta Botanica Boreali-Occidentalia Sinica*, 2009, 29(4): 800-808. (In Chinese)
[26] Wang, Z. M., Guo, Z. X., Song K. S. Responses of vegetation NDVI in northeast China to climate change. *Chinese Journal of Ecology*, 2009, 28(6): 1041-1048. (In Chinese)
[27] Cui, L. L., Shi, J., Yang, Y. M., Fan, W. Y. Ten-day response of vegetation NDVI to the variations of temperature and precipitaton in eastern China. *Acta Geographica Sinica*, 2009, 64(7): 850-860. (In Chinese)
[28] Wang, Y. L., Fan, G. Z., Zhou, D. W., Hua, W., Huang, X. L. The study of the relationship between normalized difference vegetation index and both temperature and precipitation in east China. *Journal of Tropical Meteorology*, 2009, 25(6): 725-732. (In Chinese)
[29] Dai, S. P., Zhang, B., Wang, H. J. Spatio-temporal change of vegetation index NDVI in northwest China and its influencing facotors. *Journal of Geo-information Science*, 2010, 12(3): 315-321. (In Chinese)
[30] Chao, Z. H. The reliability evaluation of three types of air temperature reanalysis data in west China. *Transactions of Atmospheric Sciences*, 2011, 34(2): 162-169. (In Chinese)
[31] Li, C., Zhang, T. J., Chen, J. Climatic change of Qinghai-Xizang Plateau region in recent 40-year reanlysis and surface observation data—contrast of observational data and NCEP, ECMWF surface air temperature and precipitation. *Plateau Meteorology*, 2004, 23(S1): 97-103. (In Chinese)