Cloud Publications
International Journal of Advanced Remote Sensing and GIS
2018, Volume 7, Issue 1, pp. 2478-2517
ISSN 2320 – 0243, Crossref: 10.23953/cloud.ijarsg.337

Research Article

Estimation of the Near-Surface Air Temperature during the Day and Nighttime from MODIS in Berlin, Germany

Forough Marzban[1], Tim Conrad[2], Pouria Marzban[3], Sahar Sodoudi[1]

[1]Institut für Meteorologie, Freie Universität Berlin, Carl-Heinrich-Becker-Weg 6-10, 12165 Berlin, Germany
[2]Department of Mathematics and Computer Science, Freie Universität, Berlin, Germany
[3]Department of Computer Engineering, Islamic Azad University, Sari, Iran

Publication Date: 22 February 2018

DOI: https://doi.org/10.23953/cloud.ijarsg.337

Abstract Air temperature (T_{air} or T_{2m}) is an important climatological variable for forest biosphere processes and climate change research. Due to the low density and the uneven distribution of weather stations, traditional ground-based observations cannot accurately capture the spatial distribution of T_{air}. In this study, T_{air} in Berlin is estimated during the day and night time over six land cover/land use (LC/LU) types by satellite remote sensing data over a large domain and a relatively long period (7 years). Aqua and Terra MODIS (Moderate Resolution Imaging Spectroradiometer) data and meteorological data for the period from 2007 to 2013 were collected to estimate T_{air}. Twelve environmental variables (land surface temperature (LST), normalized difference vegetation index (NDVI), Julian day, latitude, longitude, Emissivity31, Emissivity32, altitude, albedo, wind speed, wind direction and air pressure) were selected as predictors. Moreover, a comparison between LST from MODIS Terra and Aqua with daytime and night time air temperatures (T_{day}, T_{night}) was done respectively and in addition, the spatial variability of LST and T_{air} relationship by applying a varying window size on the MODIS LST grid was examined. An analysis of the relationship between the observed T_{air} and the spatially averaged remotely sensed LST, indicated that 3 × 3 and 1 × 1 pixel size was the optimal window size for the statistical model estimating T_{air} from MODIS data during the day and night time, respectively. Three supervised learning methods (Adaptive Neuro Fuzzy Inference system (ANFIS), Artificial Neural Network (ANN) and Support vector machine (SVR)) were used to estimate T_{air} during the day and night time, and their performances were validated by cross-validation for each LC/LU. Moreover, tuning the hyper parameters of some models like SVR and ANN were investigated. For tuning the hyper parameters of SVR, Simulated Annealing (SA) was applied (SA-SVR model) and a multiple-layer feed-forward (MLF) neural networks with three layers and different nodes in hidden layers are used with Levenber-Marquardt back-propagation (LM-BP), in order to achieve higher accuracy in the estimation of T_{air}. Results indicated that the ANN model achieved better accuracy (RMSE= 2.16°C, MAE = 1.69°C, R^2 = 0.95) than SA_SVR model (RMSE= 2.50°C, MAE = 1.92°C, R^2 = 0.91) and ANFIS model (RMSE= 2.88°C, MAE= 2.2°C, R^2 = 0.89) over six LC/LU during the day and night time. The Q-Q diagram of SA-SVR, ANFIS and NN show that all three models slightly tended to underestimate and overestimate the extreme and low temperatures for all LC/LU classes during the day and night time. The weak performance in the extreme and low temperatures are a consequence of the small number of data in these temperatures. These satisfactory results indicate that this approach is proper for estimating air temperature and spatial window size is an important factor that should be considered in the estimation of air temperature.

Keywords *ANFIS; ANN; Cross-validation; MODIS; Simulated annealing; SVR*

1. Introduction

The standard meteorological T_{air} is measured in a shelter at 2m height (Brunel, 1989; Jin and Dickinson, 2010). It is an important indicator of terrestrial environmental conditions across the earth (Prihodko and Goward (1997); Peón et al., 2014) and one of the most widely used climatic variables in global change studies. It plays an important role in multiple biological and physical processes among the hydrosphere, atmosphere and biosphere (Stisen et al., 2007; Shamir et al., 2014; Benali et al., 2012). Regarding ecosystem, it influences the distribution of plant species (Cabrera 2002) and affects the dynamics of the soil-plant-water system (Chartzoulakis and Psarras, 2005; Zavala, 2004), being included in evapotranspiration models (Allen et al., 2006; Carlson et al., 1995) as well as hydrological models (Purkey et al., 2007; Yates et al., 2005). At the individual level, temperature affects plant growth and net primary productivity since photosynthetic and respiration rates depend on it. Moreover, T_{air} plays a critical role in vegetation distributions, phenology, and growth (Benavides et al., 2007; Stahl et al., 2006). The maximum temperature also shows significant relationship with the occurrence of wildfire on hot and sunny days (Aldersley et al., 2011; Litschert et al., 2012). Therefore, detailed knowledge of the spatial variability of air temperature is of interest for many research and management.

In addition, T_{air} plays an important role in energy balance and is a key input in various environmental models and applications, such as crop evapotranspiration estimation (De Bruin et al., 2010), distributed hydrology (Gao et al., 2014) and climate change models (Lofgren et al., 2011). Moreover, the importance of temperature in urban area are related to heat stress and human health. Meteorological measurements provide accurate discrete T_{air} information for specific locations but have limited ability to describe its spatial heterogeneity over large areas (Benali et al., 2012; Willmott and Robeson, 1995). The non-uniform spatial distribution of weather station locations within most networks and the complexity of the land surface conditions and patterns make it a challenge to get spatial-continuous T_{air} data.

However, weather stations are usually sparsely distributed in mountainous regions, especially in high-elevation areas, and thus may not optimally represent all environments (Rolland, 2003). Given the large spatial heterogeneity of T_{air} in complex terrain (Holden et al., 2011), it is difficult to accurately characterize the distribution of T_{air} over mountainous areas (Carrega, 1995). Different interpolation methods have been used to generate spatially continuous Tair from point station measurements (Benavides et al., 2007; Dodson and Marks, 1997; Duhan et al., 2013; Kurtzman and Kadmon, 1999; Stahl et al., 2006). However, the performance of interpolation methods is highly dependent on the spatial density and distribution of weather stations (Chan and Paelinckx, 2008; Vogt et al., 1997), which is not considered satisfactory in mountainous areas.

Satellite remote sensing observations from global imaging sensors, such as the Advanced Very High Resolution Radiometer (AVHRR) and Moderate Resolution Imaging Spectroradiometer (MODIS), represent a potentially valuable alternative to characterize spatially-detailed Tair patterns across large areas. A split window technique was applied to AVHRR (Pinheiro et al. 2006), MODIS (Wan et al., 2002), and Meteosat (Atitar and Sobrino, 2009) thermal data to estimate Land Surface Temperature (LST). The science-grade quality of the LST data collected by MODIS has proven valuable for monitoring land surface dynamics over large areas (Benali et al., 2012, Mostovoy et al., 2006, Lin et al., 2012).

The earth's surface is heated by solar radiation, while the atmosphere is mainly heated from the ground up through longwave infrared radiation (Frederick et al., 2006). The relationship between Land

Surface Temperature and T_{air} may vary with time and location, because the land surface energy balance is a complex phenomenon that depends on multiple factors (e.g., cloud cover, surface roughness, wind speed and soil moisture), whereas some of them (e.g., wind speed) are usually not available from satellite (Goward et al., 1997; Prince et al., 1998; Stisen et al., 2007).

An accurate estimation of T_{air} and the mapping of its spatial distribution are useful for predicting ecological consequences of climate change. For example, climate warming will lead to higher temperatures and an increase of extreme weather conditions, which are associated with changes in wildfire regime (Westerling et al., 2006; Chen et al., 2011; Manzo-Delgado et al., 2009), forest biomass distribution (Reich et al., 2014) and crop yield (Ruane et al., 2014; Rosenzweig et al., 2014). The demand for accurate spatial T_{air} data over a large scale has continued to rise (Oyler et al., 2015; Beier et al., 2012). However, the spatial distribution of the weather stations in many parts of the world, is often limited which restricts the use of T_{air} measurements over a large spatial domain (Vancutsem et al., 2010). LST, but on the other hand, is measured in a global extent with significant higher spatial coverage (Jin and Dickinson, 2010). The US National Research Council and the Intergovernmental Panel on Climate Change (IPCC) expressed the need for long-term remotely sensed LST data in global warming studies to overcome the limits of conventional surface T_{air} measurements (IPCC, 2007, Jin, 2004). Remote sensing data has great potential to estimate spatial-temporal patterns of T_{air} which can further our knowledge, on both the climate and terrestrial biological processes at regional and global scales (Benaliet al., 2012). Monitoring and understanding the trends of T_{air} and LST are crucial in the study of regional and global climate changes (Yoo et al., 2011). LST can be monitored and modelled from multiple daily satellite observations, such as the MODIS LST. Studies have shown that LST can be used for linear regression estimates of daily minimum and maximum T_{air} on a local scale (Mostovoy et al., 2006; Vancutsem et al., 2010; Zhang et al., 2011a; Yoo et al., 2011; Evrendilek et al., 2012; Benali et al., 2012; Zhu et al., 2013). Cresswell et al. (1999) found an over and underestimation of T_{air} during the day and at night, respectively, from Meteosat LST observations. They attempted to correct these errors and produce a proxy of T_{air} T_{air} by applying a solar zenith angle correction on the Meteosat geostationary observations. They achieved an accuracy of 3°C for over 70% of the Meteosat temperatures. Similarly, Jin and Dickinson (2010) have studied the differences in the diurnal cycles of LST and T_{air} over a single site. Some studies (Florio et al., 2004) have used several statistical approaches that combined a simple AVHRR Spilt-Window Technique (SWT) with ground meteorological station measurements in the prediction of T_{air}. Other studies (Wloczyk et al., 2011) have used the Landsat LST data to derive T_{air}. They have attempted to assign the satellite-derived T_{air} to a certain height above the ground and have investigated the possibility of a simple correction for reference height. They also considered the link between T_{air} spatial pattern and the window-size of the Landsat LST pixels. Xu et al. (2012) used four empirical regression models to estimate the relationship between T_{air} measurements and the MODIS-Aqua LST and found different relationships between the two different LC types in their study. They also assessed the effect of the MODIS LST window-size on the agreement between the two variables and found that spatial averaging over multiple pixels improves the accuracy of T_{air} estimates. Zaksek and Schroedter-Homscheidt (2009) reviewed the types of methods commonly used to estimate T_{air} based on LST, dividing them into three distinct groups:

1) Statistical approaches which are based on regression techniques, can be simple if only based on LST and T_{air} (e.g. Mostovoy et al., 2006; Vogt et al., 1997) or advanced, when more than one independent variable is used such as solar zenith angle (SZA), elevation, altitude, Julian day among others (Lin et al., 2012; Cresswell et al., 1999; Jang et al., 2004). Lin et al. (2012) used stepwise linear regression method to estimate daily maximum air temperature (T_{max}) and daily minimum air temperature (T_{min}) with MAE = 1.9, agreement index = 0.79 and MAE = 1.9 °C, agreement index = 0.92, respectively, over east Africa. Fu et al. (2011) used linear regression between MODIS LST and

T_{max} from stations on the northern Tibetan Plateau. In general, these methods perform well within the spatial and time frame they were developed, but the accuracy might decrease when extended in time and space (Stisen et al., 2007). Statistical methods generally perform well within the spatial and time frame they were derived in, but have limited generalization and require large amounts of data to train the algorithms (Stisen et al., 2007).

2) The second category is index-based such as Temperature-Vegetation index (TVX). It is based on the assumption that for an infinitely thick canopy, the top-of-canopy temperature is the same as within the canopy (Czajkowski et al., 2000; Prihodko and Goward (1997), Nemani and Running et al., 1989; Nieto et al., 2011) and uses the Normalized Difference Vegetation Index (NDVI) as a key input variable. However, the assumption of linear and negative slope between LST and NDVI is not always applicable and is influenced by the seasons, the type of ecosystem and soil moisture variability (Sandholt et al., 2002; Vancutsem et al., 2010). Zhu et al. (2013) used the TVX method to estimate daily T_{max} with RMSE (the root mean square error) =3.709 °C, MAE (the mean absolute error) = 3.03 °C and r (correlation coefficient) = 0.83 in Xiangride River Basin of China. However, Vancutsem et al. (2010) found that TVX method did not adapt to different ecosystems over Africa because non-significant relationship between LST and NDVI in their study. Karnieli et al. (2003) found that the approaches based on this negative NDVI/LST relationship have minimal utility in energy-limited environments (e.g., high latitude and elevations) compared to moisture-limited environments because vegetation-expressed NDVI response is more related to available solar radiation than land surface conditions (e.g. soil moisture).

3) The final approach uses surface energy balance parameterizations based on physically-based models (Sun et al., 2005). The sum of incoming net radiation is considered equal to the sum of the soil heat flux, sensible flux and latent heat flux (Zakšek and Schroedter-Homscheidt, 2009; Meteotest 2010; Sun et al., 2005). However these methods require large amounts of information that are usually not only from remote sensing (e.g., roughness, soil physical properties) (Benali et al., 2012, Mostovoy et al., 2006, Prince et al., 1998).

Most of the previous studies have focused on daily estimations or instantaneous T_{air}. The TVX method has been widely used for T_{air} estimation. Czajkowski et al. (2000) estimated Tavg for a weekly period with associated RMSE between 1.72 and 3.48 °C and R2=0.64. Stisen et al. (2007) and Prihodko and Goward (1997) estimated T_{air} with RMSE higher than 2.5 °C and R^2 between 0.64 and 0.86. Cresswell et al. (1999) used a statistical method to derive instantaneous T_{air} with an associated RMSE below 3 °C for more than 70% of the sampled data. Zaksek and Schroedter-Homscheidt (2009) used a more sophisticated method, which was based on the energy balance to estimate instantaneous T_{air} with an RMSE of 2°C. Vancutsem et al. (2010) used 1 km MODIS data to estimate weekly T_{min} and T_{max}. They reported correlations between LST and T_{min} ranging from 0.01 to 0.96 for several stations and T_{max} was estimated with an R^2=0.92 and RMSE=1.83 °C.

Moreover, in previous studies, several variables were employed to estimate air temperature. For example, the variables used by Benali et al. (2012) included LST, Julian Day, elevation, and the distance to coast. Benali et al. (2012) used both weekly daytime LST data (LST_{day}) and night time LST data (LST_{night}) to estimate the average, maximum and minimum weekly temperature. They found that there was a higher correlation between average weekly temperature and averaged weekly LST_{night}, which indicates the potential of LST_{night} in estimating averaged weekly temperature. The variables used by Kim and Han (2013) included LST, NDVI, altitude, and solar zenith angle. The variables used by Cristóbal, Ninyerola and Pons (2008) included LST, NDVI, and albedo. The variables used by Zakšek and Schroedter-Homscheidt (2009) included LST, NDVI, solar zenith, albedo, solar radiation, and altitude. After comprehensive consideration of these variables, twelve variables were selected as the predictors for the modelling of air temperature during the day and night time: LST, NDVI, Julian day, latitude, longitude, Emissivity31, Emissivity32, altitude, albedo, wind speed, wind direction and air pressure.

The main objective of this study was to estimate the air temperature during day and night time with high spatial resolution in Berlin from Moderate Resolution Imaging Spectroradiometer (MODIS) data by for different land cover types.

First, this research presents the comparison of state-of-the-art remote sensing-based LST data from MODIS with T_{air} for the six LC/LU. Within this study, we compared the relationship between T_{air} and the Four LST products of MODIS over Berlin in order to analyze the agreement between LST from MODIS Terra and Aqua and T_{air} for the period of 2007 to 2013 based on different land cover classes. Specifically, the spatial scale effects of the relationship between T_{air} and LST were first analysed to determine the best window size to retrieve T_{air} in the study area. The comparison is done by using statistical parameters such as the correlation coefficient, the slope and the intercept with the y-axis of the regression line, mean bias error (MBE), and normalized mean bias which known as bias. The MBE is calculated by the difference between LST and T_{air} divided by the amount of observed time steps. If the MBE is positive, the LST detects warmer temperatures than the measured T_{air}, and vice versa (Hachem et al., 2012). Then Adaptive neuro fuzzy system (ANFIS), Artificial Neural Network (ANN) and support vector machine (SVR) models were developed to estimate T_{air} , and the accuracy of these models were assessed by comparison with the observed air temperature data from weather stations and the cross validation (CV) approach, in order to find the best model with high accuracy during the day and night time. The errors associated with T_{air} estimation based on remote sensing data are often large and strongly limit its applicability (e.g. Czajkowski et al., 2000; Vazquez et al., 1997; Vogt et al., 1997). One of the objectives of this work is to provide T_{air} estimations with an accuracy, which will potentiate the future applications. Moreover, tuning the hyper parameters of some models like SVR and ANN were investigated. In order to select the hyper parameters of SVR, Simulated Annealing (SA) was applied and a multiple-layer feed-forward (MLF) neural networks with three layers and different nodes in hidden layers are used with Levenberg–Marquardt back-propagation (LM-BP) in order to achieve higher accuracy in the estimation of T_{air} during the day and night time over six LC/LU.

2. Materials and Methods

2.1. The Study Area

Berlin is the capital city of Germany. It is located in the northeast of the country, covers an area of 892 km^2. Berlin is located on a mostly flat topography. Regarding land use patterns, Berlin is characterized by a significant amount of green areas and water bodies. Outside the inner city, there is relatively low buildings and population density, with many allotment gardens for private cultivation and recreation. There are a considerable number of urban brownfield sites, despite the trend of population growth in the last decade. Berlin consists of 45% water bodies and urban green spaces (forested and unforested, allotment gardens), almost 20% transport and infrastructural areas (streets and railways), and around 35% built-up areas (e.g. for residential use). Table 1 shows the location and the related land use of the weather stations in Berlin which are used in this study.

2.2. Data Description

Three main datasets for the period of 2007-2013 according to the availability of meteorological station record and MODIS data were used:

1. Ground measurements from 20 meteorological stations in Berlin.
2. Remotely sensed data.
3. Digital elevation model (Berlin Digital Environmental Atlas).

Table 1: *Information about weather stations over Berlin including their LC/LU, latitude, longitude and elevation*

Station	LC/LU	Lat	Long	Elevation(m)
Botanischer-Garten	Green urban area	52.45	13.30	46.88
Fasanen	Industrial, commercial, public, military	52.51	13.33	34.08
Tegel-Forstamt	Forest	52.60	13.27	39.58
Gatow	Industrial, commercial, public, military	52.47	13.13	47.09
Marzahn1	Green urban area	52.54	13.58	50.61
Pichelsdorf	Evergreen needle leaf tree	52.50	13.19	29.66
Wannsee	Evergreen needle leaf tree	52.43	13.18	40.77
Dahlem-FU	Industrial, commercial, public, military	52.45	13.31	67.50
Tegel	Airport	52.56	13.30	35.25
Schonefeld	Airport	52.38	13.53	45
Buch	Industrial, commercial, public, military	52.63	13.50	65.45
Marzahn2	Green urban area	52.54	13.55	63.29
Kaniswall	Agriculture, semi-natural and wet area	52.40	13.73	32.57
Tempelhof	Airport	52.46	13.40	47.74
Eiskeller	Agriculture, semi-natural and wet area	52.58	13.13	31.78
Kreuzberg	Industrial, commercial, public, military	52.49	13.40	34.91
Wannsee-meteo	Evergreen needle leaf tree	52.43	13.18	43.49
Adlershof	Industrial, commercial, public, military	52.42	13.52	35.15
Potsdam	Industrial, commercial, public, military	52.38	13.11	33.79
Insulaner	Green urban area	52.45	13.35	43.75

2.2.1. Meteorological Data

Air temperature observations were obtained from 20 different meteorological ground stations in this study area. The measurements included daily T_{air}, wind speed, wind direction, air pressure and Julian day. The meteorological station records were obtained from the Deutscher Wetterdienst (ftp://ftp-cdc.dwd.de/pub/CDC) and from the Freie university Berlin meteorological station (http://mevis-www.met.fu-berlin.de/devel/mevis). The accuracy of observation data in meteorological stations is as following:

1. 2m air temperature: ± 0.2 K
2. Wind speed: ± 0.3 of measured value
3. Wind direction: ± 5C^{o}
4. Relative humidity: ± 0.3% up to 0.5%
5. Air pressure: ± 0.1 hpa

2.2.2. MODIS Data

The second source of data is the satellite data. MODIS sensors were launched on board the National Aerodynamics and Space Administration (NASA) Observing System (EOS) Terra and Aqua satellites in December 1999 and May 2002, respectively (Zhu et al., 2013). Both sensors are on board sun-synchronous polar orbiting satellites. MODIS Terra data is available during 10:30–12:00 a.m. and p.m. (daytime/night time) local time, while MODIS Aqua sensor collects the imagery during 1:00–3:00 a.m. and p.m. (daytime/night time). In this study, the following products of MODIS were used:

1. MODIS daily land surface temperature at 1km resolution, from Terra (MOD11A1.005)
2. MODIS daily land surface temperature at 1km resolution, from Aqua (MYD11A1.005)
3. MODIS monthly vegetation index at 1km resolution from Terra (MOD13A2.005)
4. MODIS monthly vegetation index at 1km resolution from Aqua (MYD13A2.005) product

Daily MODIS LST and monthly NDVI from Aqua and Terra were extracted at the nearest points to the stations. The LST product from MODIS has been used in previous studies to derive Tair (Benali et al., 2012, Vancutsem et al., 2010, Zhu et al., 2013). All MODIS LST data used in this study were acquired from the U.S. Geological Survey (USGS) website (Piao et al., 2009). We used two MODIS LST products MOD11A1 and MYD11A1 from Terra and Aqua satellites, respectively. The MODIS LST consists of daytime and night time data at a spatial resolution of 1 km. Thus, in total there are four LST datasets: Aqua daytime, Aqua night time, Terra daytime and Terra night time.

2.2.2.1. *Vegetation Index*

Normalized difference vegetation index is the most common remote sensing index used to parameterize vegetation status (Zhu et al., 2013, Stow et al., 2004, Raynolds et al., 2008). The absorption and reflectivity of the vegetation cover are correlated with their structural properties, such as leaf area index (LAI), fractional vegetation cover (FVC), and their physiological condition (Bustos et al.2014, Raynolds et al., 2006). The values of NDVI vary between −1 and 1, where the range between 0.2 and 0.9 is mostly common in continuous vegetation cover (Bustos et al.2014). In this study, the NDVI was extracted from Terra (MOD13A2.005) and Aqua (MYD13A2.005) products with 16-day temporal and 1km resolution as mentioned in Table 2.

Table 2: *Data source and variables*

Variable	Source	Explanation
Land surface temperature	MODIS	Land Surface temperature derived over the 2007–2013 time period using MYD11A1, MOD11A1 product
Julian Date	Meteorological data	The continuous count of days was from 1 January to the last day every year
Emissivity31	MODIS	Emissivity31 derived over the 2007–2013 time period using MYD11A2, MOD11A1 product
Emissivity32	MODIS	Emissivity32 derived over the 2007–2013 time period using MYD11A2, MOD11A1 product
Normalized Difference Vegetation Index	MODIS	vegetation index at 1km resolution from Terra (MOD13A2.005) and Aqua (MYD13A2.005) products with 16day temporal resolution
Albedo	MODIS	Albedo at 1Km resolution from Terra (MCD43B3.005) product with 16 day temporal resolution
Relative humidity	Meteorological Data	The RH was extracted for each station during the year of 2007–2013
Altitude	DEM	The altitude extracted from a 5m resolution digital elevation model (DEM) according to the location of meteorological stations
Latitude	Meteorological data	The geographical location of meteorological stations was extracted from meteorological metadata
Wind direction	Meteorological data	The WD was extracted for each station during the year of 2007–2013
Wind speed	Meteorological data	The WS was extracted for each station during the year of 2007–2013
Air pressure	Meteorological data	The AP was extracted for each station during the year of 2007–2013

2.2.3. Auxiliary Data

In addition to MODIS products (LST, emissivity31, emissivity32), Albedo and NDVI, some auxiliary variables were used, including latitude, altitude, Julian day, air pressure, wind speed, wind direction and relative humidity which is presented in Table 2. These auxiliary variables either have a known impact on T_{air} and LST or influence the relationship between T_{air} and LST. Latitude, LC/LU and altitude were derived from the location of meteorological stations. Altitude was obtained from a 5m resolution digital elevation model (DEM) (downloaded from https://www.eea.europa.eu/data-and-maps/data/urban-atlas). Moreover, Julian day was also considered as proxies for the fraction of solar energy absorption during the day and emission during the night, influencing the diurnal amplitude of T_{air} throughout the year. Julian day is the continuous count of days from 1 January every year. In addition, the LC/LU of each meteorological station was extracted in terms of its position, and reclassified into urban, industrial, forest, airport, needle leaf trees and agriculture based on a 5m resolution map of LC/LU, which was downloaded from https://www.eea.europa.eu/data-and-maps/data/urban-atlas.

In addition, all data (Auxiliary and MODIS data) were combined to create a single dataset for each LC/LU for day and night time. The collinearity of independent variables was detected using variance inflation factor (VIF > 10) and pair wise correlation (r > 0.75) (Zurr et al., 2010, Dormann et al., 2013).

2.3. LST Pre-Processing

A certain number of pre-processing steps were required to convert the original LST product in HDF format to raster layers with a versatile projected coordinate system. Firstly, raster subsets of the LST product were extracted based on the boundary extent of the study area. LST L3 product is gridded in the global Sinusoidal projection, and the grid containing data for the study area is located at column 18 (h18) and line 03 (v03). It is important to eliminate low quality data in the MODIS LST data because remote sensing based T_{air} estimates are strongly influenced by errors (e.g., errors caused by clouds and large sensor viewing angles, uncertainties in surface emissivity (Wan et al., 2004). Validation studies of MODIS LST show that under clear sky conditions the precision is approximately 1 K or less, but higher errors would be observed at large viewing angles and in semiarid regions (Wan et al., 2008). So only the pixels of the targeted land cover types that were flagged in the MODIS quality assurance data as cloud free and of high quality were retained.

2.3.1. Calculating LST of Weather Station Location

LST data under clear sky conditions at weather stations are retrieved by the following steps:

- A total of 5110 MODIS HDF format (MOD11A1 and MYD11A1, h18v03, Collection 5, from 1 January 2007 to 31 December 2013 over Berlin) in HDF (Hierarchical Data Format) format were re-projected to WGS_1984_UTM_ zone_33N using the nearest neighbor resampling method. The corresponding layers (LST_Day_1km, LST_Night_1km, Daytime LST observation time, and Night time LST observation time) were extracted. However, Daytime and Night time LST observation time were used in order to identify the approximate overpass time of MODIS at local time.
- MODIS LST data for the pixels in which the weather stations are located are extracted from MODIS using nearest neighbor algorithm.
- All these LST data (DN value) were converted to Celsius temperature using the following equation:

$$T\ (C^{o}) = 0.02 * DN - 273.15 \ldots (1)$$

Where is the C^{o} is the Celsius temperature and 0.02 is the scale factor of the MODIS LST product.

- Removing low quality data: MODIS LST products are not available for a location (pixel) if clouds are present (Wan 2008). However, there are some pixels that are lightly covered or contaminated by clouds. These pixels are not removed because the contamination is very small and cannot be detected by the cloud-removing mask algorithm (Ackerman et al., 2008, Williamson et al., 2013; Xu and Shen 2013). To avoid this kind of data, only the pixels of the targeted land cover types that were flagged in the MODIS quality assurance data as cloud-free and of high quality were retained.

2.4. The Relationship between Observed T_{air} and the Four LST Products from MODIS

The influence of the time of observation on the estimation of T_{air} has been discussed in several studies, which resulted in different conclusions. Benali et al. (2012) stated that the use of both aqua LST_{day} and LST_{night} could improve the estimation of T_{day} and T_{night} , (T_{day} and T_{night} are not the maximum and minimum temperature of a day and night time)respectively, because the MODIS Aqua overpass time is closer to the time of both T_{day} and T_{night} than Terra's. In contrast, Zhu et al. (2013) showed that both terra LST_{day} and LST_{night} were better than aqua LST_{day} and LST_{night} for T_{air} estimations in Xiangride River basin of China. In another study, Mostovoy et al. (2006) found that the difference between the satellite overpass (Terra and Aqua) had little impact on the estimation accuracy of T_{air}.

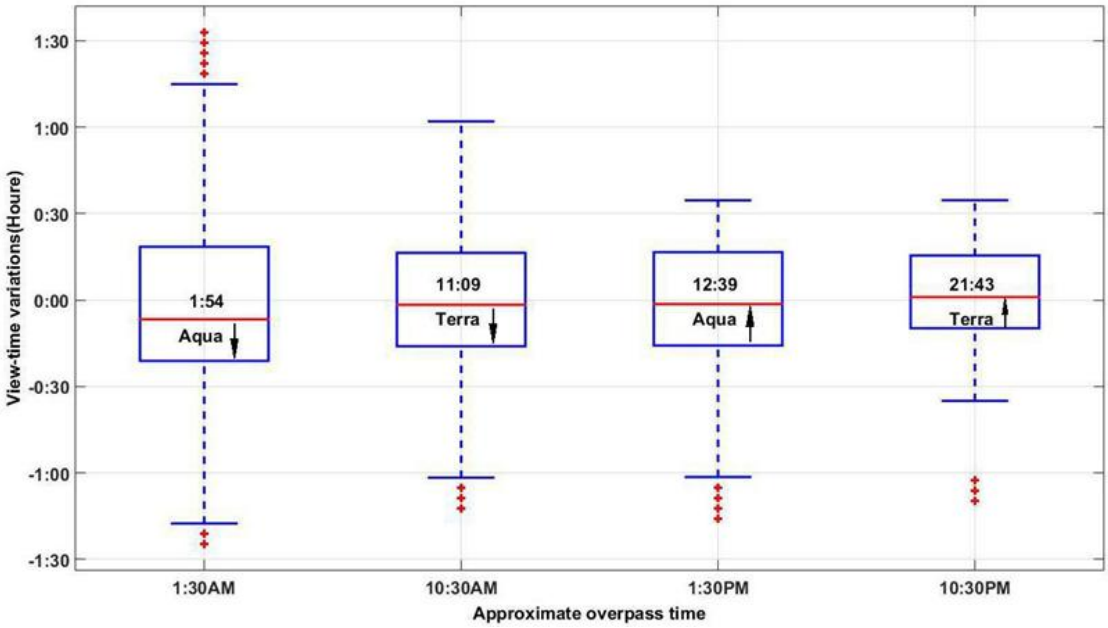

Figure 1: *Average viewing times (local solar) & overpass nodes (shown as labels and arrows), maximum variations from the mean observation times (in hour shown by lower and upper caps of whiskers), median times (middle line), lower (25th) and upper (75th) quartiles of all observation times (lower and upper edges of boxes) of four overpasses of MODIS (onboard Terra and Aqua, two overpasses each) over the study area of 7 years (2007–2013). The mean local solar observation time of each overpass is subtracted from the series (scaled to 0) but is labeled on each box*

2.4.1. Temporal Matching of T_{air} to LST Observations

The temporal frequency of the MODIS LST L3 product is four observations per day (Terra passes over the equator at approximately 10:30 am, 10:30 pm each day, Aqua satellite passes over the equator at approximately 1:30 pm and 1:30 am) in cloud free conditions, which are derived from a composite of several MODIS overpasses with different view angles (Wan 1999; Zhu et al., 2013). Depending on the local longitude (which results in changes in the sensor's viewing angle) and latitude, the local solar observation times at each pixel can vary up to 120 minutes or more over a repeating cycle (16 days) of the MODIS twin sensors (Figure 1). Other than that, overpass times do not follow a regular period during the day and over the sensor's repeat cycle. On the other hand, T_{air} data from the weather stations are provided at an hourly and by the minute frequency in Berlin in standard time (MEZ and UTC). This complicates the matching of the MODIS observation times with T_{air} time-series. To overcome this issue, for those stations that are in minute temporal resolution, we only need to convert from MEZ to UTC, but for other stations that are in hourly temporal resolution, a linear equation was considered for the synchronizing of T_{air} with LST form HDF file. For the creation of a data set for use during the day and night time for each LC/LU separately, we need to consider the overpass time of MODIS over Berlin. Another point is that the data has a significant number of missing points due to clouds in our study area, therefore in order to increase the maximum of usable observations, as did Alcantara et al. (2013) in his research, the terra and aqua data were considered. As shown in Figure 1, the LST day from terra and aqua in descending and ascending orbit were considered as a day time series, respectively and only terra in ascending orbit was considered as night time series because of higher correlation which was observed between LST and T_{air} in this time.

2.4.2. Day/Night analysis

Apart from spatial variations, the observation time can affect the relationship between LST and T_{air} time-series. To identify any variability in LST and T_{air} relationship in a diurnal basis, time-series of both variables were separated, based on the MODIS overpass times to produce four series over a single pixel window from MODIS-Terra and MODIS-Aqua day and night overpasses (four in total) were used in this analysis.

2.4.3. Statistical methods

A simple linear relationship is often assumed between LST and T_{air} in literature (Brunel, 1989; Mostovoy et al., 2006). In view of this, a univariate linear regression analysis with the MODIS LST as the independent (or explanatory) and T_{air} as the dependent (or response) variable was applied to analyze LST and T_{air} relationship. The correlation coefficient, r, is reported as a quantitative measure to evaluate the strength of the agreement between LST and T_{air} time series in different steps of the analysis. Significance levels (*p-values*) are reported in the results to express how unlikely the given *r* values would occur if no relationship between the explanatory and response variables did exist. The smaller the *p-level* is, the more significant the relationship. Moreover, two other statistical measurements such as RMSD and Normalized mean Bias (Bias) were considered as following:

$$Normalized\ mean\ bias = \frac{\sum_{i=1}^{n}(M-O)}{\sum_{i=1}^{n} O} \quad \ldots (2)$$

Where n is the number of data, M is LST value and O is the temperature.

3. Theory and Methodology

In this section, a brief overview of a SVR, Simulated annealing (SA), ANN and ANFIS will be discussed.

3.1. Support Vector Regression

Support vector machine (SVM) is a very promising artificial intelligence method applied extensively for solving the classification problems. Support Vector Regression (SVR) method is derived from the SVM, which is a powerful technique to solve a nonlinear regression problem, but it has received less attention, due to the fact that SVR algorithm is sensitive to users' defined free parameters. The involved hyper parameters of the SVR model consist of penalty parameter, insensitive loss function parameter, and the parameter for kernel function. Inappropriate parameters in SVR can lead to over fitting or under fitting problems.How to properly use the hyper parameters is a major task, which has a significant impact on the optimal generalization performance and the SVR regression accuracy (Schölkopf and Smola 1998). Recently, a number of new algorithms like genetic algorithm, grid search optimizing, cross-validation and particle swarm optimization (PSO) have been proposed for the optimization of the SVR parameters (Sartakhti et al., 2011; Ustün et al., 2005; Wang et al., 2016; Chen and Wang, 2007; Hu et al., 2010; Keerthi, 2002; Ito and Nakano, 2005). In this work, the SA algorithm was applied for tuning the parameters of SVR.

3.1.1. Brief Overview of SVR

In this section, the basic SVR concept is concisely described; for detailed description, please see (Cristianini and Taylor, 2000; Smola and Schölkopf, 2004; Ito and Nakano, 2005; Keerthi, 2002). Suppose a given training data of elements $\{(x_i, y_i), = 1, 2, \dots N\}$, where x_i denotes the ith element in n-dimensional space; that is $x_i = \{x_1 i, \dots x_n i\} \in R^n$, and $y_i \in \mathrm{R}$ is the output value corresponding to x_i. According to mathematical notation, the SVR algorithm builds the linear regression function as follows:

$$f\,(x, \omega) = (\omega| \cdot\ \varphi(x) + b)\,,$$
$$: R^{n} \rightarrow F,\ \omega \in F, \qquad \dots (3)$$

Where w and b are the slope and offset coefficients and x denotes the high-dimensional feature space, which is nonlinearly mapped from the input space x. The previous regression problem is equivalent to minimizing the following convex optimization problem shown in equation (4):

$$Min\ \ \frac{1}{2}||\omega||^2$$

Subject to
$$y_i - (x_i) - b \leq \varepsilon$$
$$\omega\varphi\ (x_i) + b - y_i \leq \varepsilon. \qquad \dots (4)$$

In this equation, an implicit assumption is that a function f essentially approximates all pairs (xi, yi) with ε precision, but sometimes this may not be the case. Therefore, by introducing two additional positive slack variables ξi and ξ_i^*, the minimization is reformulated as the following constrained optimization problem shown in equation 5:

$$\min\ R(\omega, \xi, \xi^*) = \frac{1}{2}\|\omega\|^2 + C\sum_{i=1}^{N}(\xi_i + \xi_i^*)$$
$$y_i - (\omega, \varphi(x)) - b \leq \varepsilon + \xi_i^*$$
$$(\omega, \varphi(x)) + b - y_i \leq \varepsilon + \xi_i$$
$$\xi_i, \xi_i^* \geq 0, i = 1,2,\ldots,N,\ \ \varepsilon \geq 0, \qquad \ldots (5)$$

Where the parameter c is the regulator, which is determined by the user and it influences a tradeoff between an approximation error and the weights vector norm and slack variables that represent the distance from actual values to the corresponding boundary values of tube. According to the strategy outlined by Scholkopf and Smola (1998), by applying Lagrangian theory and the KKT condition, the constrained optimization problem can be further restated as the following equation: was applied in the study, which has the ability to universally approximate any distribution in the feature space. With an appropriate parameter, the radial basis function (RBF) usually provides a better prediction performance, so it is adopted in this study as shown in the following formula:

$$f(x, \omega) = \sum_{i=1}^{N}(\alpha_i - \alpha_i^*)\, K(x_i, x) + b$$
$$s.t.\ \ \sum_{i=1}^{N}(\alpha_i - \alpha_i^*) = 0. \qquad \ldots (6)$$

Here αi and αi* are the Lagrange multipliers. The term (xi, x) is defined as the kernel function. The nonlinear separable cases could be easily transformed to linear cases by mapping the original variable into a new feature space of high dimension using (xi, x). The RBF was applied in the study, which has the ability to universally approximate any distribution in the feature space. With an appropriate parameter, RBF usually provides a better prediction performance, so it is adopted in this study as shown in equation 7:

$$K(x_i, x_j) = \exp\left(-\frac{\|x_i - x_j\|^2}{2\sigma^2}\right) \qquad \ldots (7)$$

where xi and xj are input vector spaces and σ^2 is the bandwidth of the kernel function. In the above equations, there exist three hyper-parameters to be determined in advance, that is, the penalty parameter c, insensitive parameter Ɛ, and the related kernel function parameters σ^2. They heavily affect the regression accuracy and computation complexity of SVR. The penalty parameter c controls the degree of punishing the samples whose errors go beyond the given value. The insensitive parameter Ɛ controls the width of the Ɛ-insensitive zone used to fit the training data. The value of Ɛ can enhance the generalization capability; with the increase of Ɛ, the number of support vectors will decrease, and the algorithmic computation complexity will also reduce. The bandwidth of σ the kernel function has a great influence on the performance of the learning machine. In this study, one optimization method, that is, simulated annealing (SA), is presented to determine the optimal hyper parameters of the SVR model. According to research of Ustün and Melssen (2005), the general range of c, σ^2, and Ɛ has been given. In the trial operation, we narrowed it to avoid blindness in the optimization process. In this study, the set of hyper parameter (c, σ^2, Ɛ) is initialized in the given range c ∈ [0, 1000], σ^2 ∈ [0, 2], and Ɛ ∈ [0, 0.0001], where optimization method (SA) is to seek the global optimal solutions.

3.1.2. Simulated Annealing Optimization Method

Simulated annealing is a local search algorithm capable of escaping from local optima. Its ease of implementation and convergence properties and its use of hill climbing moves to escape local optima have made it a popular technique over the past two decades. Survey articles that provide a good overview of simulated annealing's theoretical development and domains of application include (Eglese, 1990; Fleischer, 1995; Henderson et al., 2003; Koulamas et al., 1994; Romeo et al., 1991; Anily and Federgruen, 1987; Suman and Kumar, 2006; Abramson et al., 1999; Ben-Ameur, 2004; Aarts and Korst, 1989; van Laarhoven and Aarts, 1988; Aarts and Lenstra, 1997). This study proposed an SA-based approach for parameter tuning in the SVR. For convenience, the SVR model with SA is referred to as a SA-SVR method. The idea, is to find the parameters that minimize the generalization error of the algorithm at hand. This error can be estimated on some data which has not been used for learning. To achieve this aim, the three basic decision variables as mentioned before must be tuned in proper manner. We propose here a methodology for automatically tuning multiple parameters for the SVR. The process of SA-SVR algorithm approach is briefly summarized as follows:

Algorithm: Simulated annealing algorithm.

Step 1: Solution space X
Object function F
Neighborhood structure N

Step 2: Current =An initial solution, among all possible state (X)
$S_{optimal}$=Current
T0=INFINITY
T=T0
Iteration=MAX_Iter
Epoch=1
Select temperature reduction function alpha, 0.8≤ alpha ≤0.99

Step 3: Repeat
Next= randomly selected from N (Current)
ΔF=F [Next] – F [Current]
If ΔF> 0:
Current=Next
else
r=rand (0, 1) % Generate a random number r ϵ (0, 1)
if $r < e^{(-\Delta f/T)}$:
Current=Next
Until Epoch <= Iteration
T=alpha*T
If F [Current] < F [$S_{optimal}$] :
$S_{optimal}$ =Current
Until stop condition is met

Step 4: Return $S_{optimal}$ as an approximation to the global minimum solution

The proposed parameter values of SA-SVR approach were set as follows: Iteration = 200, T_0 was set to a sufficiently large number, while the set of hyper parameters (C, σ^2, ε) is initialized in the given range $C \in [0, 10000]$, $\sigma^2 \in [0, 2]$, and $\varepsilon \in [0, 0.0001]$, where optimization method (SA) is to seek the global optimal solutions. The best solution among these possible solutions is then selected as the optimal solution in the SA-SVR.

3.2. An Adaptive Neuro-Fuzzy Inference System

An adaptive neuro-fuzzy inference system or adaptive network-based fuzzy inference system (ANFIS) is a kind of artificial neural network that is based on Takagi–Sugeno fuzzy inference system (Sugeno and Tanaka, 1992; Takagi and Sugeno, 1985). The technique was developed in the early 1990s (Jang and Shing (1991, 1993)). Since it integrates both neural networks and fuzzy logic principles, it has a potential to capture the benefits of both in a single framework. Its inference system corresponds to a set of fuzzy IF–THEN rules that have learning capability to approximate nonlinear functions (Abraham, 2005). Hence, ANFIS is considered to be a universal estimator (Jang, Sun and Mizutani, 1997).

3.3. Neural Network

Artificial neural network models are universal approximations with the ability to generalize through learning non-linear relationships between provided variables of input(s) and output(s) (Hájek and Olej 2012). ANN are organized and interconnected collections of processing units (neurons or nodes), whose operation is analogue to a neural structure (Müller and Fill, 2003). ANN extract its computational power from its solid parallel distribution structure and ability to learn/generalize, allowing the resolution of complex propositions in many known areas (Haykin 2001). ANN execution is inspired on the human brain (Haykin, 2001) and has been used in many applications with success. In agreement with Galvão et al. (1999), by the reason of its nonlinear structure, the ANN can acquire more complex data characteristics, which are not always possible using traditional statistical techniques (Maier et al., 2010; Razavi and Tolson, 2011). ANN is a robust computational technique which is primarily used for pattern recognition, classification, and prediction (Bose and Liang, 1996; Haykin, 1999; Panchal et al., 2011). The use of ANNs in meteorological applications includes the prediction of ozone concentration, sulfur dioxide concentration, tornadoes, storms, solar radiation, carbon dioxide, pollutants, and monsoon rainfall (Gardner and Dorling, 1998), monthly and year precipitation levels (Bodri and Cermak, 2000), tide charts (Steidley et al. 2005), wave heights (Wedge et al. 2005), flash floods (Luk et al., 2000), and air temperature (Jain et al., 2003; Smith et al., 2006; Maqsood et al. 2004), estimation of dew point temperature (Mittal and Zhang, 2003; Shank et al., 2008). Bilgili and Sahin (2010) used ANN for predicting long-term monthly temperature and rainfall in Turkey. Kisi and Shiri (2011) introduced new hybrid wavelet-AI models for precipitation forecasting. Smith et al. (2005) developed an enhanced ANN for air temperature prediction by including information on seasonality and modifying parameters of an existing ANN model.

3.3.1. Determining Hidden Node

Many researchers put their best effort in analyzing the solution to the problem that how many neurons are kept in hidden layers in order to get the best results (Rivals I. and Personnaz L. 2000; F. Fnaiech.et al. 2001; Kortmann-Unbehauen 1988; Onoda 1995; Md. Islam and Murase (2001); Stuti Asthana and Rakesh K Bhujade (2011); Kazuhiro Shinike 2010; Doukim et al. 2010; Yuan et al. 2003; Wu and Hong 2010; Panchal et al. 2011; Hunter et al. 2012; Shuxiang et al. 2008; Ke and Liu (2008)), but unfortunately no one succeeded in finding the optimal formula for calculating the number of neurons that the neural network training time can be reduced and also accuracy in determining the target output can be increased. Usually some rule-of-thumb methods are used for determining the number of neurons in the hidden nodes.

1. The number of hidden layer neurons are 2/3 (or 70% to 90%) of the size of the input layered. If this is insufficient then the number of output layer neurons can be added later on (Boger and Guterman, 1997).

2. The number of hidden layer neurons should be less than twice of the number of neurons in input layer (Berry and Linoff,1997).
3. The size of the hidden layer neurons is between the input layer size and the output layer size (Blum, 1992).

But the above three methods are not considered to be always true because not only the input layer and the output layer decide the size of the hidden layer neurons, but also the complexity of the activation function applied on the neurons, the neural network architecture, the training algorithm, and most important the training samples of the database on which the neural network is designed to execute. In this work, we decided to use the cross validating approach in the 3-layers MLP in the following simulations, in order to select the number of hidden nodes in the second layer. The 3-layer MLP contains an input layer, one hidden layer with nonlinear transfer functions and an output layer with linear transfer functions. The training algorithm is Back Propagation (BP) in order to get the configuration that minimizes the RMSE in the test phase while keeping an eye on over fitting and the train set error.

3.3.2. Assess Predictive Performance of Models

In a real application, cross-validation is a model assessment technique (Allen, 1974; Stone, 1974; Geisser, 1975) used to evaluate a machine learning algorithm's performance in making predictions on new datasets which has not been trained on. This is done by partitioning a dataset and using a subset to train the algorithm and the remaining data for testing. Because cross-validation does not use all of the data to build a model, it is a commonly used method to prevent over fitting during training. Each round of cross-validation involves, randomly partitioning the original dataset into a training set and a testing set. The training set is then used to train a supervised learning algorithm and the testing set is used to evaluate its performance. This process is repeated several times and the average cross-validation error is used as a performance indicator (Hastie et al., 2009; Yang, 2007b). Common CV techniques include, k-fold, Holdout, Leave out, repeated random sub-sampling, Stratify, Substituting. In this work, we apply K-fold CV (with k=4) techniques, in order to test how well our model is able to be trained by some data and then to estimate the data it hasn't seen before and then to select the best model.

3.4. Data Normalization

Before computing, data of both input and output variables were normalized. In this study, data of all variables used were normalized into the range [0, 1] with:

$$X_{norm} = \frac{X_i - X_{min}}{X_{max} - X_{min}} \quad \dots(8)$$

where X_{norm} is the normalized value, Xi is the original value, and X_{min} and X_{max} are the minimum and maximum values out of the sample of Xi. This was due to the eliminating influence of different dimensions of data and to the avoidance of overflows of the model during calculations, as a result of very large or small weights towards a maximization of model parsimony with considering computational effort. After the computation, output values were transformed back to the real prediction data.

3.5. Model Calibration and Validation

Cross-validation was used to evaluate the generalizability of a model for estimating the air temperature with the LST data. The observations were randomly divided into two parts. 70% of the observations were used for model calibration, and the rest were used as test dataset for model validation. The accuracy of the estimated air temperature obtained from three estimating models, ANFIS, NN and SA-SVR, have been assessed by a set of statistic measures, including: Root Mean-square Error (RMSE), coefficient of determination R-squared (R2), Mean Bias Error (MBE) and Mean Absolute Error (MAE), respectively. The RMSE (was mainly used in the development process of the model and represents residual errors, which gives a global perspective of the differences between the observed and estimated values (Sousa et al., 2007; Zheng et al., 2013; Willmott et al., 2005). The RMSD is calculated similarly to RMSE. These goodness of fit criteria are expressed as equations (9-12):

$$R = \frac{\sum_{i=1}^{M}(O_i - \bar{O})(S_i - \bar{S})}{\sqrt{\sum_{i=1}^{M}(O_i - \bar{O})^2} \times \sqrt{\sum_{i=1}^{M}(S_i - \bar{S})^2}} \quad (9)$$

$$MBE = \frac{1}{M}\sum_{i=1}^{M}(O_i - S_i) \quad (10)$$

$$MAE = \frac{1}{M}\sum_{i=1}^{M}|O_i - S_i| \quad (11)$$

$$RMSE = \sqrt{\frac{1}{M}\sum_{i=1}^{M}(O_i - S_i)^2} \quad (12)$$

where, M is the total number of the observation data, O and S are the average of the observed and estimated T_{2m}, and O_i and S_i are the observed and estimated T_{2m} of the i^{th} data, respectively. In addition, graphical goodness-of-fit criteria such as quantile-quantile (Q-Q) diagram, bar plot of RMSE in train and test phases were applied for the comprehensive evaluation of simulation results. Although, the R^2 criteria is a measure of goodness-of-fit of the model and higher values are indicative that the predictive model fits the data in a better way. By definition, R^2 is the proportional measure of variance of one variable that can be predicted from the other variable. Thus, ideally the values of R^2 to approach one is always desirable. However, a high R^2 tells you that the curve came very close to the points, but in reality, it does not always indicate the model quality (Maddala, 2001). In order to have a reliable statistical comparison between the models, both the MAE and RMSE can be used together to ascertain the variation in errors in a given set of estimation. It should be noted that in MAE, all the individual errors have equal weight on the average, making it a linear score, but the RMSE has a quadratic error rule, where the errors are squared before being averaged. As a result, a relatively high weight is given to large errors. This could be useful when large errors are undesirable in a statistical model (Chai and Draxler, 2014; Armstrong, 2002).

4. Theoretical Concepts for Selecting Input Parameters

Spatial and temporal variation in temperature are governed by physical processes. For example, land surface temperature at some ‘locations’ in space and time ($s0, t_0 | s \in S, t \in T$) is a function of incoming solar radiation, cooling factor by wind, land cover, temperature inversion and other effects. The temperature patterns differ between day and nighttime also; during the night temperature patterns are mainly determined by land cover, air humidity and proximity to water bodies and/or soil moisture (van Leeuwen et al., 2011). In urban and industrial areas, temperature is often locally somewhat higher due to heat emissions from industrial activities or heating (e.g. Cheval and Dumitrescu 2009). Moreover, near-surface air temperature is driven more by land surface temperature than by direct solar radiation (Zakšek and Schroedter-Homscheidt 2009), making LST an important variable for estimating T_{air}.

Other parameters, such as vegetation cover, soil moisture, solar radiation, and albedo also have some influence on air temperature. In previous studies, several variables were employed to estimate air temperature. For example, the variables used by Benali et al. (2012) included LST, Julian Day, elevation, and distance to coast. The variables used by Kim and Han (2013) included LST, NDVI, altitude, and solar zenith angle. The variables used by Cristóbal, Ninyerola, and Pons (2008) included LST, NDVI, and albedo. The variables used by Zakšek and Schroedter-Homscheidt (2009) included LST, NDVI, solar zenith, albedo, solar radiation, and altitude. After comprehensive consideration of these variables, twelve variables were selected as the predictors for modelling air temperature: LST, NDVI, latitude, longitude, altitude, albedo, wind speed, wind direction, emissivity31, emissivity32, relative humidity and Julian day. The reasons for selecting these variables as input to our model for estimating the air temperature are summarized as follows:

1 The latitude, longitude and elevation were selected as an input parameter to model because the incoming solar radiation can be globally derived as a function of this factors. Moreover, latitude, longitudes and elevation are always the underlying effect relative to temperature (Zhao and Cheng, 2005; Samanta et al., 2012; Stahl et al., 2006).

2 Emissivity is important, because all objects at temperatures above absolute zero emit thermal radiation. However, for any particular wavelength and temperature the amount of thermal radiation emitted depends on the emissivity of the object's surface. Emissivity is defined as the ratio of the energy radiated from a material's surface to that radiated from a blackbody (a perfect emitter) at the same temperature and wavelength and under the same viewing conditions. The emissivity of a surface depends not only on the material but also on the nature of the surface. The emissivity also depends on the temperature of the surface as well as wavelength and angle. Knowledge of surface emissivity is important both for accurate non-contact temperature measurement and for heat transfer calculations. Moreover, Surface emissivity is a measure of inherent efficiency of the surface in converting heat energy into radiant energy above the surface (Sobrino et al., 2001). Therefore, land surface emissivity is critical for determining the thermal radiation of the land surface (Caselles et al., 1995). The emissivity of a surface is controlled by some factors such as water content, chemical composition, structure, roughness, and the observation conditions (i.e. wavelength, pixel resolution and observation angle) (Snyder et al., 1998). For these reasons, in our study, due to considering six different LC/LU, the land surface emissivity also considered as an input parameter.

3 LST is the radiative temperature of the land surface (Ghent et al., 2010). It is influenced by albedo, vegetation cover and soil moisture (Land Surface Temperature Copernicus Global Land Service). The “surface” can include snow and ice, bare soil, grass, or the roofs of buildings (Land Surface Temperature: Global Maps, 2016). Near-surface air temperature “is a measurement of the average kinetic energy of the air near the surface of the Earth” (Near Surface Air Temperature - GES DISC-Goddard Earth Sciences Data and Information Services Center, 2016). Usually LST is measured by remote sensing whereas air temperature is measured 1-2 m above the ground. Near-surface air temperature is a consequence of complex effects of the turbulent heat transports produced by nearby heated surfaces (Unger, et al., 2009). The advantage of using MODIS LST is that, they account for small differences in

temperature that are due to different land cover, moisture content which cannot modeled with constant physical parameters such as elevation, latitude, longitudes.

4 The Julian day is proxies for the fraction of solar energy absorption during the day and emission during the night, influencing the diurnal amplitude of T_{air} throughout the year. The Julian day included the information of vegetation cover changes with seasons. Julian day is the continuous counting of days from 1st January every year.

5 The NDVI and Albedo reflect the seasonal variation of land cover.

6 The Relative humidity (RH) is the ratio of the partial pressure of water vapor to the equilibrium vapor pressure of water at a given temperature. Relative humidity depends on temperature and the pressure of the system of interest. It requires less water vapor to attain high relative humidity at low temperatures; more water vapor is required to attain high relative humidity in warm or hot air (Perry and Green, 2007).

7 Moreover, Seasonal variation in some parameters such as relative humidity, wind speed, wind direction and air pressure contribute to explaining seasonal variation air temperature over six LC/LU.

8 The MODIS LST can be used to improve spatial prediction of ground-measured values.

5. Results and Discussion

5.1. MODIS LST versus T_{air} time-series over a single pixel

Before analyzing the effects of MODIS window size, the daily variability of LST and T_{air} relationship was examined by using separate LST series (over 1x1 window) (Diurnal differences). In this section, LST series used in this analysis is a composite time series which includes four daily LST observations (except for cloudy days) from both the MODIS Terra and Aqua day and night overpasses (approximately at 1:30, 10:30, 13:30, 22:30) supplied in the LST L3 product.

Table 3: *Statistical analyses between MODIS LST products and T_{air} observation from automatic meteorological stations. MOD_{day}, MOD_{night}, MYD_{day} and MYD_{night} are representative of MOD11A1 LST_{day}, MOD11A1 LST_{night}, MYD11A1 LST_{day} and MYD11A1 LST_{night} from Terra and Aqua respectively for urban and industrial LCT*

Dataset	Urban				Industrial			
	R^2	RMSD	MBE	Bias	R^2	RMSD	MBE	Bias
^{MOD}day' ^{T}day	0.88	3.72	0.25	0.01	0.86	3.69	-0.58	-0.03
^{MOD}nigt, ^{T}night	0.87	3.63	-1.89	-0.21	0.80	4.57	-2.65	-0.28
^{MYD}day' ^{T}day	0.87	4.22	1.57	0.09	0.86	3.81	0.23	0.01
^{MYD}nig ^{T}night	0.88	2.94	-1.50	-0.21	0.80	4.21	-2.55	-0.35

Table 4: *Statistical analyses between MODIS LST products and T_{air} observation from automatic meteorological stations. MOD_{day}, MOD_{night} , MYD_{day} and MYD_{night} are representative of MOD11A1 LST_{day}, MOD11A1 LST_{night}, MYD11A1 LST_{day} and MYD11A1 LST_{night} from Terra and Aqua respectively for agriculture and needle leaf trees LCT*

Dataset	Agriculture				Needleleaftrees			
	R^2	RMSD	MBE	Bias	R^2	RMSD	MBE	Bias
^{MOD}day' ^{T}day	0.91	2.89	-1.42	-0.08	0.85	4.21	0.49	0.02

^{MOD}night' ^{T}night	0.85	3.23	-0.82	-0.11	0.87	3.61	-1.93	-0.20
^{MYD}day' ^{T}day	0.92	3.05	-0.22	-0.01	0.83	4.41	1.11	0.06
^{MYD}night' ^{T}night	0.85	2.83	-0.43	-0.09	0.87	3.15	-1.59	-0.21

The comparison between MODIS LST data and the T_{air} observations shows that LST_{day} and LST_{night} from both Terra and Aqua, with the mean relative bias above and under zero tended to overestimate T_{day} and underestimate T_{night} (Table 3-5) respectively as Cresswell et al. (1999) found the same result.

Table 5: *Statistical analyses between MODIS LST products and T_{air} observation from automatic meteorological stations. MOD_{day}, MOD_{night} , MYD_{day} and MYD_{night} are representative of MOD11A1 LST_{day}, MOD11A1 LST_{night}, MYD11A1 LST_{day} and MYD11A1 LST_{night} from Terra and Aqua respectively for airport and forest LCT*

Dataset	Airport				Forest			
	R^2	RMSD	MBE	Bias	R^2	RMSD	MBE	Bias
^{MOD}day' ^{T}day	0.90	3.99	1.65	0.10	0.89	4.43	2.46	0.13
^{MOD}night' ^{T}night	0.86	4.32	-2.92	-0.31	0.83	3.69	0.73	0.08
^{MYD}day' ^{T}day	0.89	4.80	2.72	0.15	0.88	4.82	3.54	0.17
^{MYD}night' ^{T}night	0.86	3.71	-2.41	-0.32	0.88	2.85	1.15	0.18

As shown in the table, a higher relative RMSD and bias values were seen for the Aqua $LST_{daytime}$ than the Terra $LST_{daytime}$ which might be given to the fact that more solar radiation has been received at the time of the Aqua MODIS overpass later in the day. Considering the scatterplots of LST_{night} and T_{night} from Aqua for the industrial LC type, has higher scattering than daytime observations which are more spread around the 1:1 line (Figure 5). This is undeniable evidence of the negative impact of urbanization on a surface urban heat island (UHI) and global warming This indicates the urban heat island with RMSD=4.21°C (Nguyen et al., 2015).

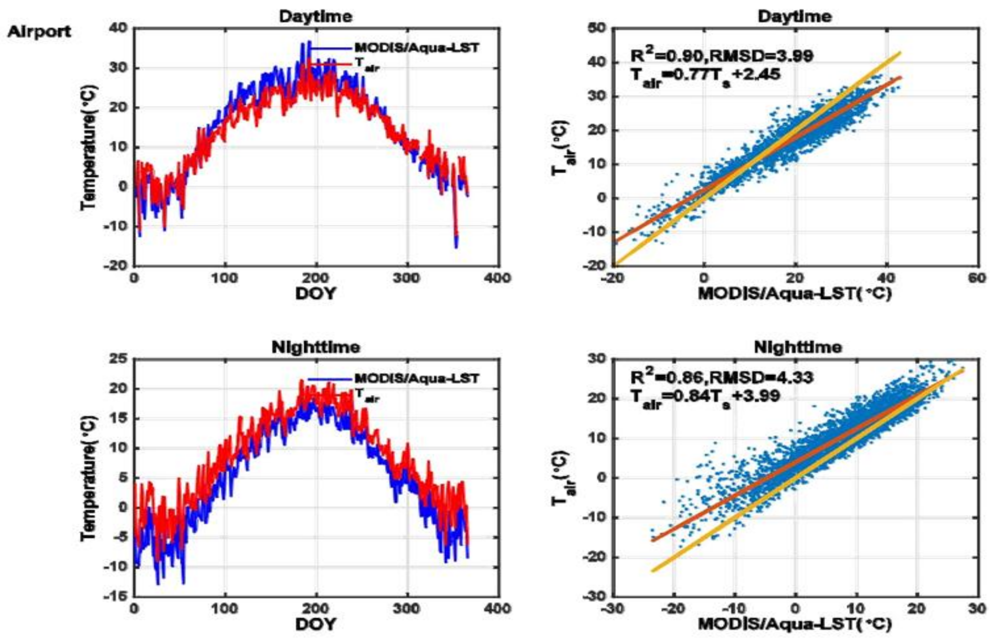

Figure 2: *Correlations between LST and Tair time-series separated based on approximate overpass times of MODIS-Aqua for seven years between 2007 to 2013, where each scatterplot shows MODIS-Aqua daytime (right-up) and MODIS-Aqua night time (right-down) and also MODIS-Aqua day and nighttime observations (left up and down plots) plotted against T_{air} measurements at the corresponding times for airport LCT with P-value<0.01. DOY means day of year*

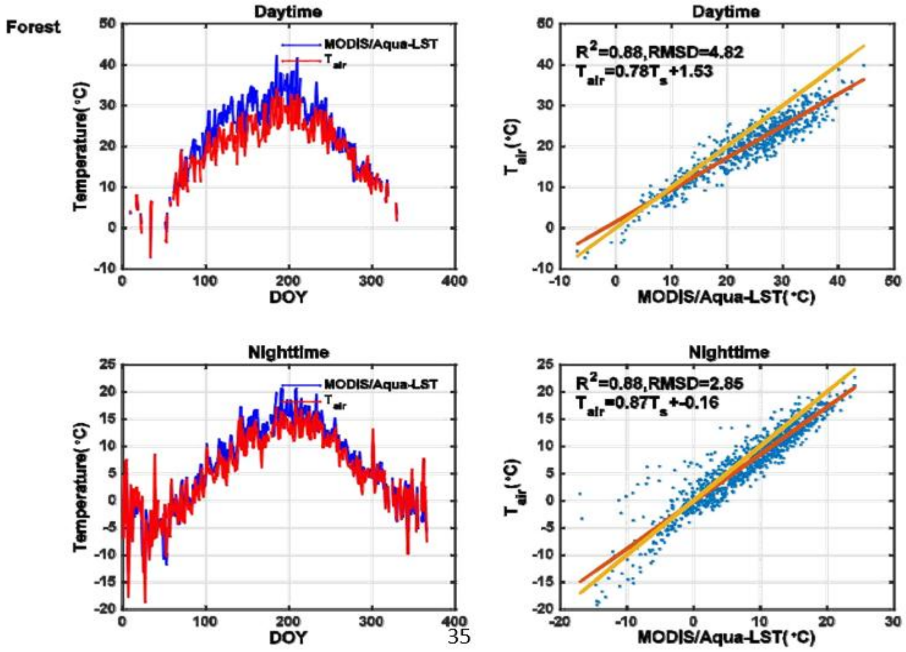

Figure 3: *Correlations between LST and Tair time-series separated based on approximate overpass times of MODIS-Aqua for seven years between 2007 to 2013, where each scatterplot shows MODIS-Aqua daytime (right-up) and MODIS-Aqua night time (right-down) and also MODIS-Aqua day and nighttime observations (left up and down plots) plotted against T_{air} measurements at the corresponding times for forest LCT with P-value<0.01. DOY means day of year*

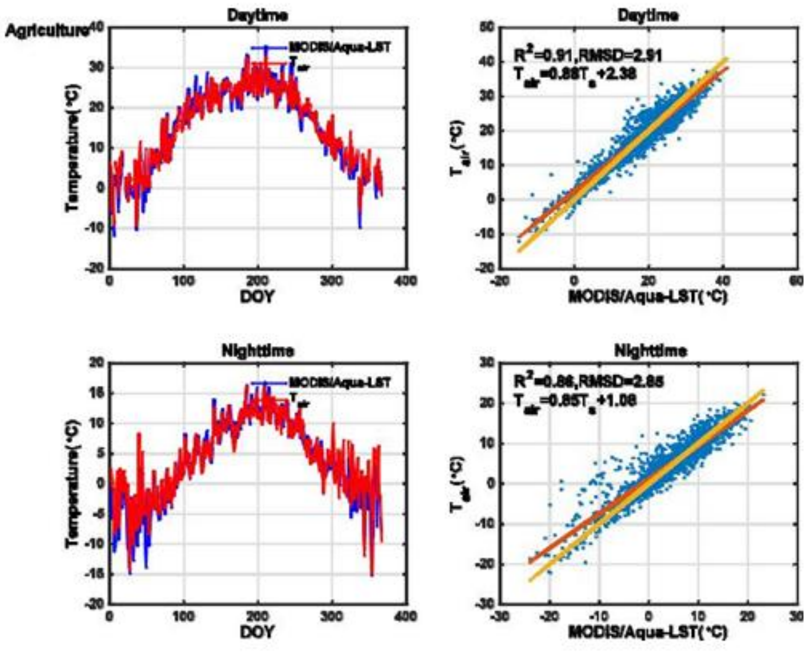

Figure 4: *Correlations between LST and T_{air} time-series separated based on approximate overpass times of MODIS-Aqua for seven years between 2007 to 2013, where each scatterplot shows MODIS-Aqua daytime (right-up) and MODIS-Aqua night time (right-down) and also MODIS-Aqua day and nighttime observations (left up and down plots) plotted against T_{air} measurements at the corresponding times for agriculture LCT with P-value<0.01*

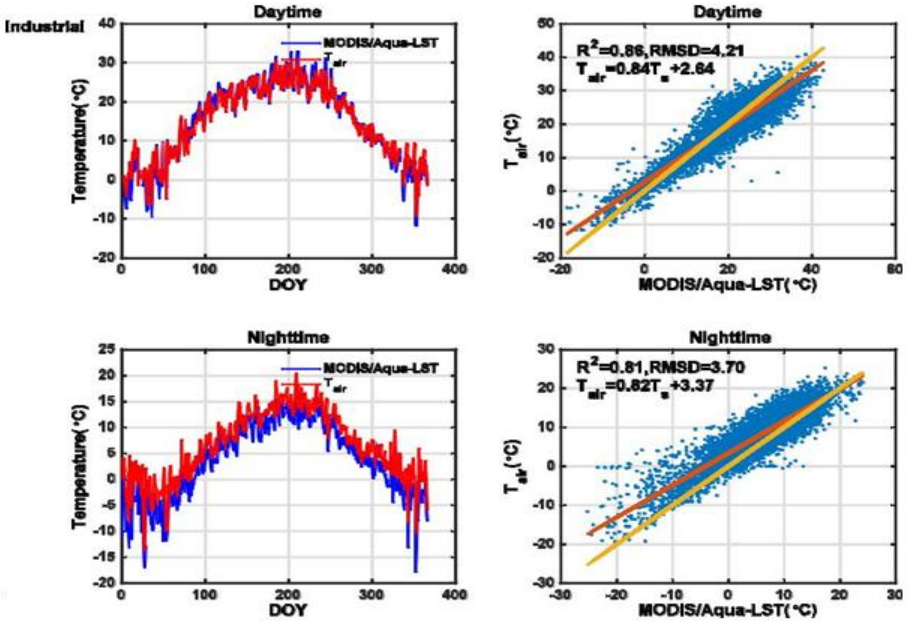

Figure 5: *Correlations between LST and T_{air} time-series separated based on approximate overpass times of MODIS-Aqua for seven years between 2007 to 2013, where each scatterplot shows MODIS-Aqua daytime (right-up) and MODIS-Aqua night time (right-down) and also MODIS-Aqua day and nighttime observations (left up and down plots) plotted against T_{air} measurements at the corresponding times for industrial LCT with P-value <0.01. DOY means day of year*

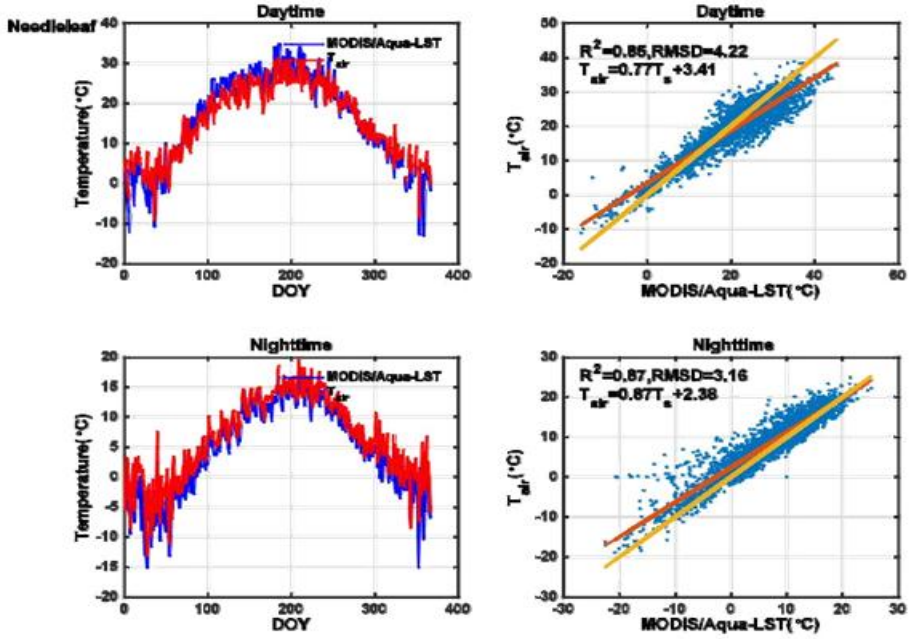

Figure 6: *Correlations between LST and T_{air} time-series separated based on approximate overpass times of MODIS-Aqua for seven years between 2007 to 2013, where each scatterplot shows MODIS-Aqua daytime (right-up) and MODIS-Aqua night time (right-down) and also MODIS-Aqua day and nighttime observations (left up and down plots) plotted against T_{air} measurements at the corresponding times for needle leaf trees LCT with P-value <0.01. DOY means day of year*

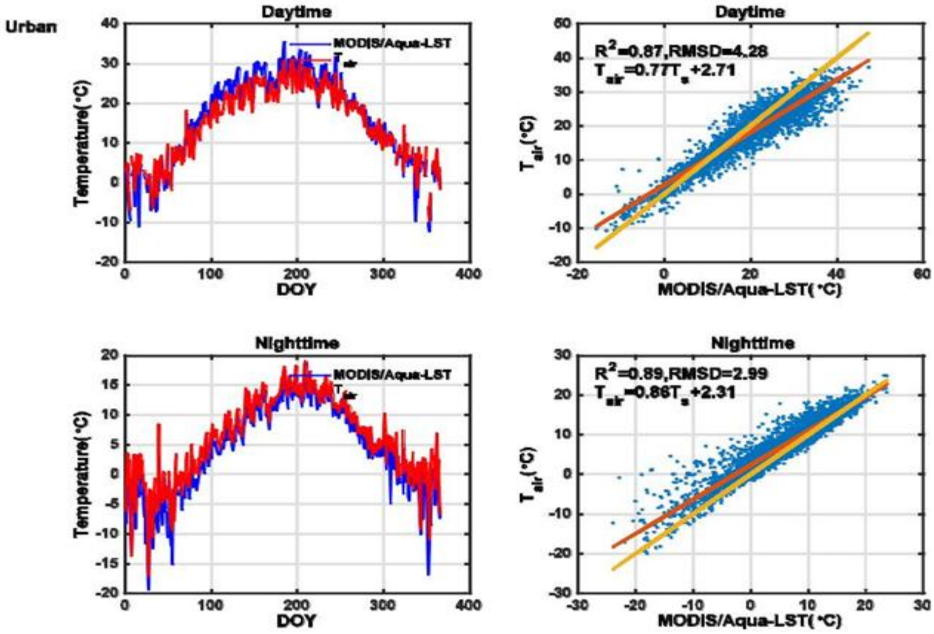

Figure 7: *Correlations between LST and T_{air} time-series separated based on approximate overpass times of MODIS-Aqua for seven years between 2007 to 2013, where each scatterplot shows MODIS-Aqua daytime (right-up) and MODIS-Aqua night time (right-down) and also MODIS-Aqua day and nighttime observations (left up and down plots) plotted against T_{air} measurements at the corresponding times for urban LCT with P-value <0.01. DOY means day of year*

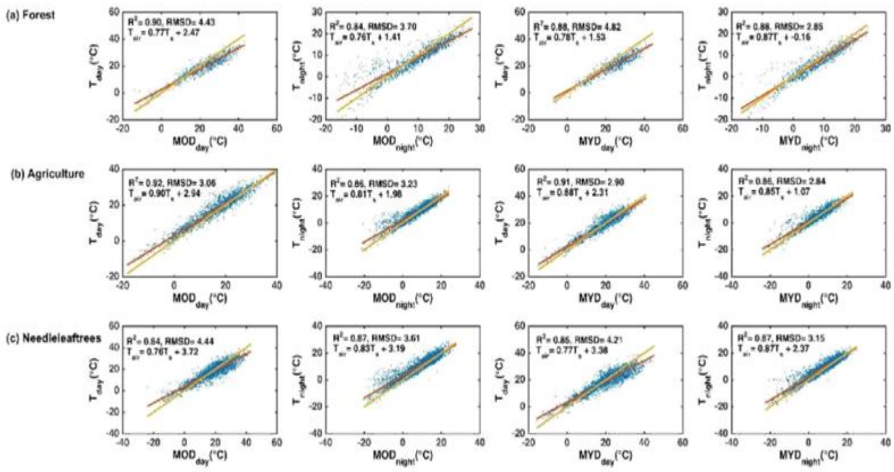

Figure 8: *Scatter plots between observed T_{air} (T_{day} and T_{night}) and LST from Four MODIS products (MOD_{day}, MoD_{night}, MYDday, MYD_{night}) which are represented for three different LC/LU for considered year (2007 to 2013). RMSD is the Root-Mean_Squar Deviation and it's calculated as the same as RMSE. R^2 is adjusted correlation coefficient between T_{air} and T_S. Tsis land surface temperature*

Both Aqua and Terra LST_{night} underestimated the T_{night} as well except for forest. Moreover, according to RMSD from Table 3-5 and MODIS LST from Terra, a higher RMSDs is found for industrial and airport LC types during night time which indicates the UHI phenomena (with RMSD = 4.57°C and 4.32°C respectively). Moreover Table 3-5 show that, correlations between the MODIS LST from Terra data are generally stronger from the daytime series compared with those from the night series, except for needle leaf trees. The needle leaf tree type showed more complex correlation patterns from day and night observations. The possible reason for this, is that the values of LST recorded by MODIS observation on this particular LC type is not exactly a representative of the skin temperature of the soil, but rather affected by the temperature near the top of the trees (canopy temperature). In addition, LST and T_{air} are correlated to a certain degree, with some drawbacks depending on factors, such as land cover type (Jin et al., 2010; Mildrexler et al., 2011). In general, Figure 2-7 show that the time-series of the MODIS LST over six LC/ LU classes were correlated individually during the day and night time. They are highly correlated with R^2 > 0.80. Moreover, Figure 2-7 show that, during the warm months the LST_{day} is higher than T_{day} due to strong radiation, while as expected during the cold months LST_{day} is lower than T_{day} for almost all LC/LU. Moreover, almost for all LC/LU, the LST_{night} is close to T_{night}. As due to long wave, radiation from surface LST and T_{air} at night are closer.

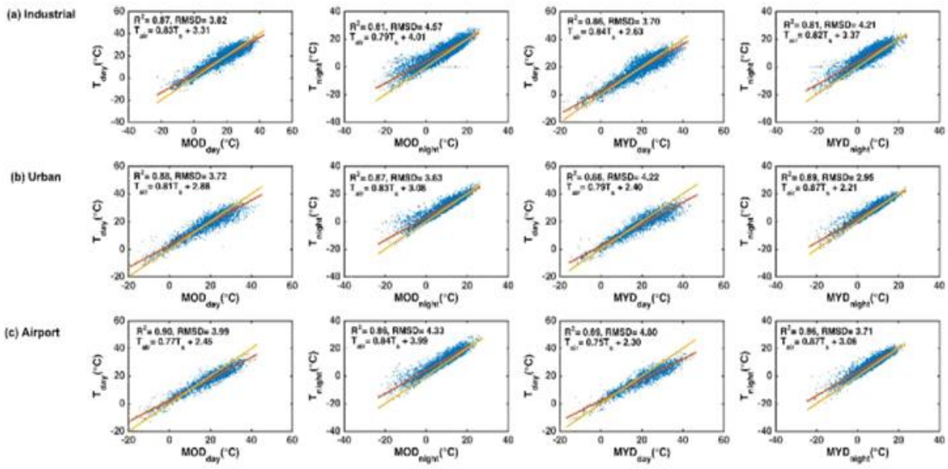

Figure 9: *Scatter plots between observed Tair (T_{day} and T_{night}) and LST from Four MODIS products (MOD_{day}, MoD_{night}, MYD_{day}, MYD_{night}) which are represented for three different LC/LU for considered year (2007 to 2013). RMSD is the Root-Mean_Squar Deviation and it's calculated as the same as RMSE. R2 is adjusted correlation coefficient between T_{air} and TS. Tsis land surface temperature*

Both the Terra and Aqua LST products were compared with the ground-based T_{air} as shown in the Figure 8 and 9, the night time LST datasets (MOD_{night} and MYD_{night}) and the observed T_{air} are more linearly concentrated along the fitting line than the daytime datasets. Strong correlations were observed between the night time LST and T_{night} with minimal bias (0.81<R^2 <0.89, RMSE< 4.80 and MBE < 2.91 °C). Specifically, the MYD_{night} tends to be more accurate for the estimation of T_{air} with lower intercepts, smaller RMSD and MBE than MOD_{night}. For T_{day}, the MYD_{day} had good agreement than MOD_{day} with lower intercept. This is most likely because the Aqua overpass time (1:30 the time when maximum temperature was recorded). However, LST from Aqua and Terra seems to be best for estimation of T_{day} among the LST products.

To sum up, the relationship between LST and T_{air} may vary with time and location because the land surface energy balance is a complex phenomenon that depends on multiple factors (e.g., cloud cover, surface roughness, wind speed and soil moisture). In addition, the LST and T_{air} are different in principle. The satellite remotely sensed LST is a measure of the surface radiation. LST was calculated from the emissivity's surface, which is sensitive to LC, especially during daytime and another reason is the heat capacity or specific heat of LC. However, the specific heat varies significantly from one LC to another. The variation of the difference between LST_{day} and T_{day} may be due to the different heat capacities or specific heats of LC types (Marzban et al., 2017, Voogt and Oke, 2003). The heat capacity changes with temperature, which may result in different relations at the different times even over the same LC. Our results showed that the MODIS LST correlates best with T_{air} measurement during the daytime. To some extent, this outcome was contradictory to the other works in the literature (e.g., Zhang et al., 2011a; Benali et al., 2012, Marzban et al., 2017) where they have reported a stronger correlation at night time compared to daytime. It must be noted, though, that Benali et al. (2012) used MODIS-Terra but not MODIS-Aqua observations. Variations in the MODIS, overpasses time in its 16-day repeated cycle which enabled us to reconstruct the diurnal LST profile over a 7 years period. Although, many studies have shown a higher agreement between LST and T_{air} at night (Zhang et al., 2011a; Benali et al., 2012), this is not the case for all hours of the day or night. During some hours of the night the LST - T_{air} relationship is weaker than some hours during the day. These differences could be understood, as not only being the time of observation, but also geographical location affecting the relationship between LST product and T_{air} and therefore, affecting the accuracy estimation of T_{air} based on LST products.

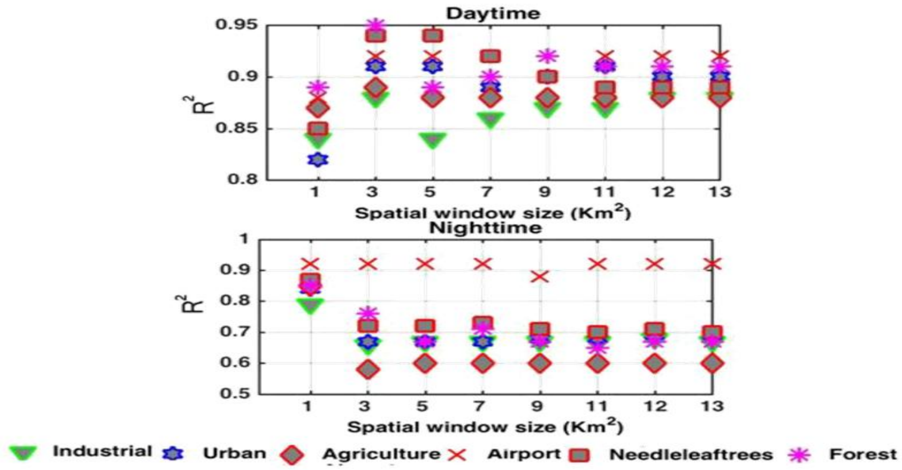

Figure 10: *Variation of the correlation coefficient between T_{day}, T_{night} and LST with the varying spatial window size over six LCT*

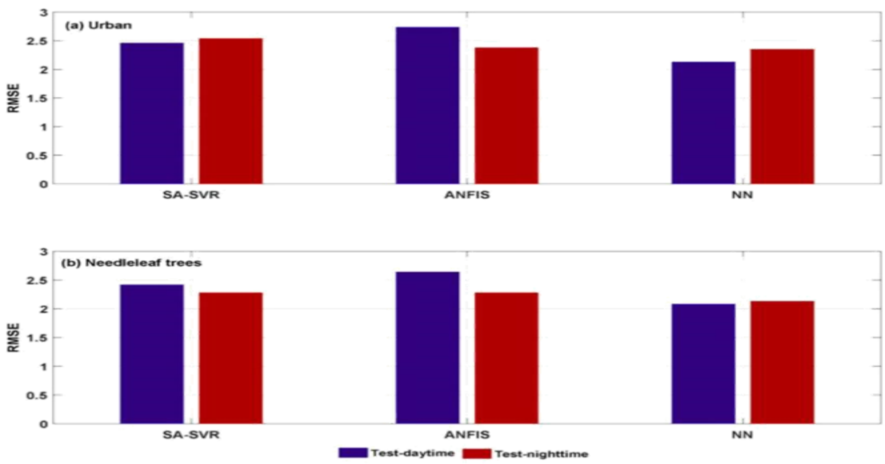

Figure 11: *Bar plot of estimated T2m versus measured temperature during day and nighttime in test phases using SA-SVR, ANFIS and NN for (a) urban and (b) Needle leaf trees LCT*

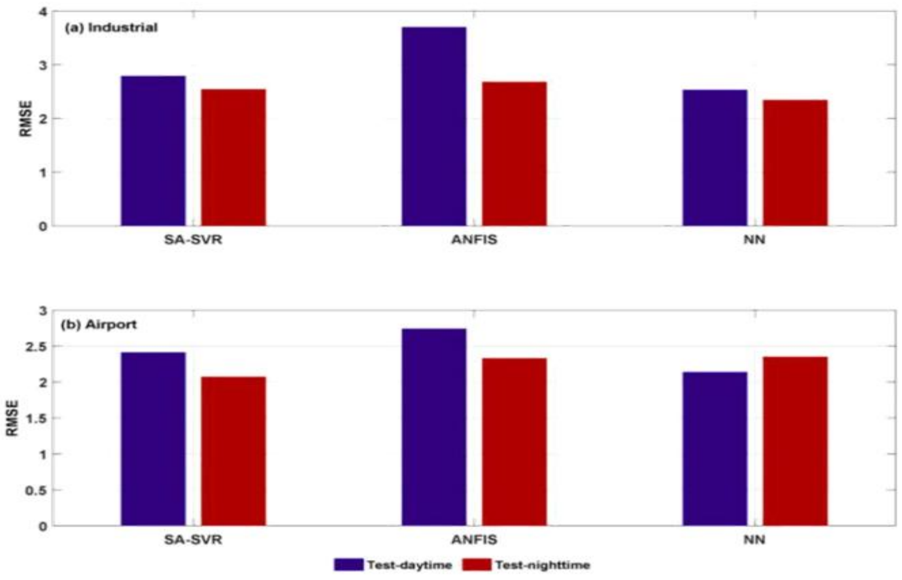

Figure 12: *Bar plot of estimated T2m versus measured temperature during day and nighttime in test phases using SA-SVR,ANFIS and NN for (a) Industrial and (b) Airport LCT*

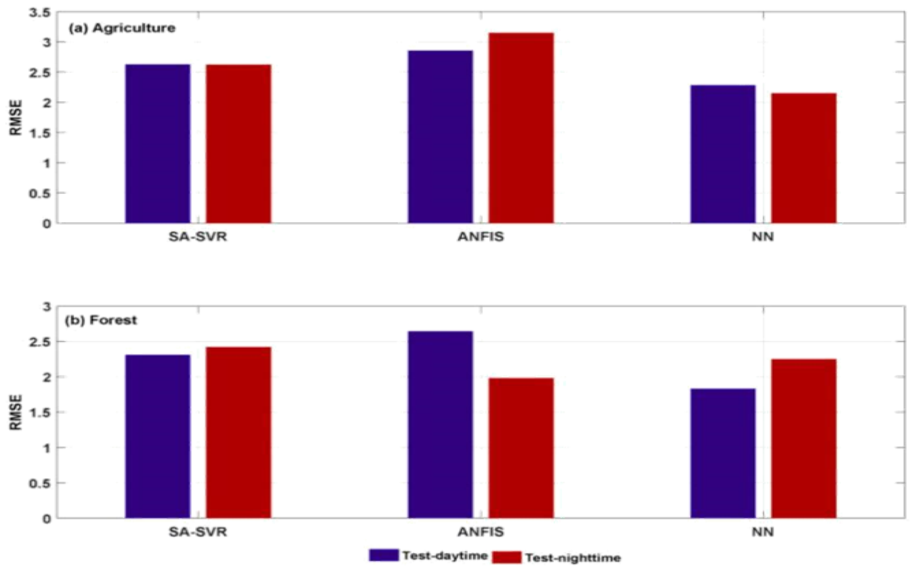

Figure 13: *Bar plot of estimated T2m versus measured temperature during day and nighttime in test phases using SA-SVR, ANFIS and NN for (a) Agriculture and (b) Forest LCT*

5.2. Multiple LST window size

The relationship between the observed T_{night} –T_{day} and LST is not limited to a single pixel, because the temperature of the near-surface air mass in a given area, is influenced by many factors such as energy exchanges with the land surface over a larger area. On the other hand, the T_{air} is impressed by both the local radiation budget and air advection from the surrounding areas, thus, for better understanding of the spatial variability in LST- T_{air} relationship, a spatial window with a varying size is examined to discover the optimal spatial extent over which LST agrees best with the T_{air}

measurements. In order to describe the effects of LST window-size on the LST and T_{air} relationship better, firstly, the time-series of LST from a single pixel (1x1 window size) overlapping each weather station were retrieved from the MODIS LST grid and then the LST of 3x3, 5x5, 7x7, 9x9, 11x11, 13x13 and 15x15 pixels were generated, respectively. Secondly, in order to determine the proper spatial window size for estimating air temperature, correlation coefficient analysis was made for different LC/LU. As shown in figure 8, the correlations were improved very slightly when the window size was increased from 1 to 3 pixels for daytime. The highest correlation values were achieved with 3x3 window for all LC/LU during the daytime and at the 1x1 during the night time. Significance levels of all correlations were found to be at which can be interpreted from p-values (all *p-values* were <0.01). According to these results, the window size was selected for all LC/LU prior to model development for day and night time data set.

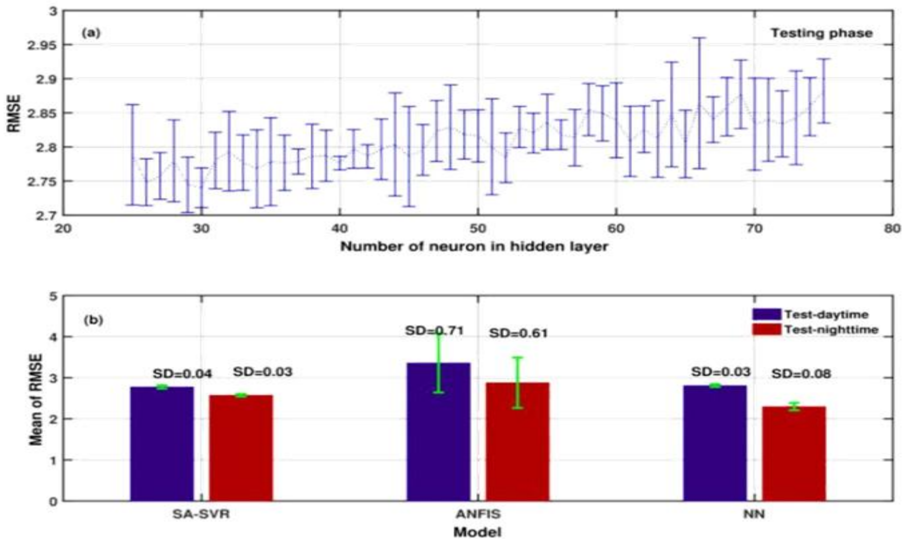

Figure 14: *This two subplots show the effect of K-fold cross validation (with k=4) in three models. In subplot (a) x and y-axes show the average of cross validation error (RMSE) and number of nodes in hidden layer in testing phase respectively. In subplot (b), x and y-axes show the type of model and the average of cross validation error (RMSE) in three models respectively*

5.3. Discussion

Three different methods namely SA_SVR, ANN and ANFIS were employed to estimate T_{air} during the day and night time in Berlin by using the twelve variables as predictors. The performance of the three models was assessed using cross-validation with k=4 fold over different LC/LU during the day and night time. All samples from each LC/LU were used in turn as the validation data set to test the model, while the remaining samples were used as the training data set to fit the model. RMSE, R^2, MBE and MAE were calculated from the measured and estimated T_{air} values to assess model performance. As shown in Table 6-11, ANN model with three layers structure, has higher adjusted R^2 value ranged from 0.93 to 0.97, RMSE ranged from 1.83 °C to 2.53 °C and MAE ranged from 1.53 °C to 1.94 °C in test phases for all LC/LU for estimating T_{day}. The results showed that all models have similar capability in the training phase for estimating T_{night} but the ANN has a higher adjusted R^2 which ranged from 0.89 to 0.93, RMSE and ranged from 2.13 °C to 2.35°C and also MAE ranged from 1.54 °C to 1.84 °C values in the test phase in comparison to ANFIS and SA-SVR. The bar plots of RMSE for the three methods on testing data for each LC/LU are shown in Figure 11-13, respectively. As shown in Table 6-11 and Figure 11-13, the three models SA-SVR, ANFIS and ANN have satisfactory been able to capture the relationship between the process variables. The bar plots depicted the performance of the NN model on the testing data which was better than those of ANFIS and SA-SVR models for the whole of LC/LU during the day and night time, but again we applied a CV approach to assess the models performance in the test phase for the three mentioned models. As can be seen from Figure 14 (b), the SA-SVR and NN models are more robust and stable than ANFIS model regarding their SD values (ranged from 0.03 to 0.08) during the day and night time and we can say that, these two models are more reliable than the ANFIS model.

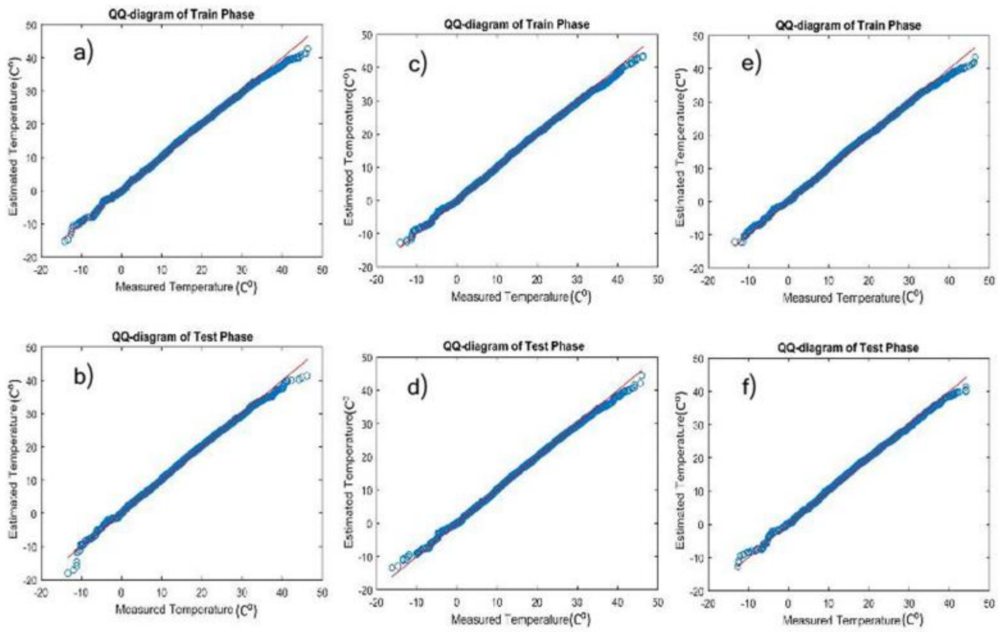

Figure 15: *Q-Q diagram of estimated T2m versus measured temperature during daytime for Industrial LCT using SA-SVR, ANFIS and ANN in testing phase*

Table 6: *Statistic indices between estimated T_{day} values obtained by SA-SVR and measured value from meteorological stationover six LCT in test phase*

LCT	RMSE	MAE	MBE	R2
Agriculture	2.62	2	0.13	0.92
Forest	2.31	1.84	0.06	0.91
Industrial	2.79	2.12	0.09	0.91
Urban	2.46	1.89	0.13	0.92
Airport	2.41	1.87	0.02	0.92
Needleleaf trees	2.42	1.85	0.09	0.93

Moreover, in order to find the optimum number of neurons in hidden layer, various numbers of neurons are used in the MLP and the optimum number of hidden neurons is determined using the CV approach to get the configuration that minimizes the RMSE in the test phase. Figure 14 (a) shows that after a certain number of hidden neurons are added, the model will start over fitting our data and give bad estimates on the test set. This indicates that over fitting starts to occur when the number of neurons is greater than 30, and in this point the model has lowest RMSE, and obviously we can conclude that the optimal number of hidden neurons should be 30, but if we consider the error bar which is the indicator of standard deviation, the less variation was observed at point 40, and then we can say that the model is more stable at this point as compared to point 30 (which is the number of neurons).

Table 7: *Statistic indices between estimated T_{day} values obtained by SA-SVR and measured value from meteorological station over six LCT in test phase*

LCT	RMSE	MAE	MBE	R2
Agriculture	2.62	1.87	0.26	0.88
Forest	2.42	1.67	0.20	0.88
Industrial	2.54	1.84	0.16	0.89
Urban	2.56	1.87	0.16	0.89
Airport	2.07	1.49	0.16	0.92
Needleleaf trees	2.28	1.61	0.18	0.91

Table 8: *Statistic indices between estimated T_{day} values obtained by ANFIS and measured value from meteorological station over six LCT in test phase*

LCT	RMSE	MAE	MBE	R2
Agriculture	2.85	2.21	0.12	0.91
Forest	2.64	2.08	0.69	0.90
Industrial	3.70	2.78	0.17	0.88
Urban	2.74	2.03	0.35	0.90
Airport	2.75	2.08	0.36	0.90
Needleleaf trees	2.64	2.06	0.24	0.90

Table 9: *Statistic indices between estimated T_{night} values obtained by ANFIS and measured value from meteorological station over six LCT in test phase*

LCT	RMSE	MAE	MBE	R2
Agriculture	3.15	2.40	-0.06	0.84
Forest	1.98	1.45	-0.04	0.92
Industrial	2.68	1.91	0.19	0.88
Urban	2.38	1.72	-0.18	0.90
Airport	2.33	1.69	-0.16	0.90
Needleleaf trees	2.28	1.65	0.50	0.92

Table 10: *Statistic indices between estimated T_{day} values obtained by ANN and measured value from meteorological station over six LCT in test phase*

LCT	RMSE	MAE	MBE	R2
Agriculture	2.28	1.82	0.05	0.97
Forest	1.83	1.54	-0.29	0.97
Industrial	2.53	1.94	-0.102	0.93
Urban	2.13	1.62	-0.07	0.97
Airport	2.14	1.69	0.06	0.95
Needleleaf trees	2.08	1.57	0.14	0.95

Moreover, Figure 15 shows Q-Q diagram of SA-SVR (left), ANFIS (middle) and ANN (right) models. Q-Q diagrams are often used to determine whether the model could extract the behavior of the observed data (Chambers et al., 1983). As shown in Figure 15, the models cannot estimate the high temperature for all LC/LU during the day and night time. The weak performance of all models at high temperature are a consequence of a small number of data in these temperatures, and this is also highly related to the study area condition (Berlin) which has a short summer and has only a few number of high temperatures. In these cases, the learning algorithm of the three mentioned models have the tendency to underestimate the temperature. Therefore, the generalization of these models for the high temperature is reduced.

Table 11: *Statistic indices between estimated T_{night} values obtained by ANN and measured value from meteorological station over six LCT in test phase*

LCT	RMSE	MAE	MBE	R2
Agriculture	2.15	1.59	0.16	0.90
Forest	2.15	1.54	-0.01	0.89
Industrial	2.34	1.73	0.03	0.91
Urban	2.35	1.81	-0.06	0.93
Airport	2.35	1.84	0.10	0.92
Needleleaf trees	2.13	1.55	0.06	0.92

6. Conclusions

In this study, the comparison between the LST and T_{air} observations was done. The comparison shows that LST_{day} and LST_{night} from both Terra and Aqua, with the mean relative bias above and under zero tended to overestimate T_{day} and underestimate T_{night} respectively, and also a higher relative RMSD and bias values were seen for the Aqua $LST_{daytime}$ than the Terra $LST_{daytime}$ which might be given the fact that more solar radiation has been received at the time of the aqua MODIS overpass later in the day. The scatterplots of LST_{night} and T_{night} from Aqua for industrial LC/LU has higher scattering than daytime observations which are more spread around the 1:1 line (Figure 4). This indicates UHI phenomena with RMSD = 4.21°C. Moreover, according to RMSD from Table 3-5 and MODIS LST from Terra, a higher RMSDs is found for industrial and airport LC/LU types during the night time which indicated the UHI phenomena (with RMSD= 4.57°C and 4.32°C respectively). The results show that, the correlations between the MODIS LST from Terra data are generally stronger from the daytime series compared with those from the night time series except for needle leaf trees. The needle leaf tree type showed a more complex correlation pattern from day and night observations. The reason is that the values of LST recorded by MODIS observation on this particular LC type is not exactly a representative of the skin temperature of the soil, but rather affected by the temperature near the top of the trees. In general, the results showed that the time-series of the MODIS LST over six LC/LU classes were correlated individually during the day and night time. They are highly correlated with $r > 0.80$. In addition, the results indicate that, during the warm months the LST_{day} is higher than T_{day} while as expected during cold months LST_{day} is lower than T_{day} for almost all LC/LU. Moreover, for almost all LC/LU, the LST_{night} is close to T_{night}. Overall, the relationship between LST and T_{air} is varied with time and location because the land surface energy balance is a complex phenomenon that depends on multiple factors (e.g., cloud cover, surface roughness, wind speed and soil moisture). In the other words, The LST–T_{air} relationship is mainly controlled by the surface energy balance, but it also depends on factors that are closely linked to energy processes (Prince et al., 1998; Zhang et al., 2015).

Moreover, in this study, the air temperature during the day and night time in the period from 2007 to 2013 was estimated for the Berlin area over six LC/LU, using 1 km Aqua and Terra/MODIS data. The correlation coefficient between observed T_{air} and remotely sensed LST shows an increasing trend, with a spatial window size increasing from 1 km × 1 km to 3 km × 3 km, and subsequently decreasing slightly at window sizes larger than 3 km × 3 km for the daytime, but for the night time this correlation coefficient between observed T_{air} and LST showed a decreasing trend, with spatial window size from 1 km × 1 km to 13 km × 13 km, and subsequently decreasing slightly at window sizes larger than 1 km × 1 km. These window sizes were therefore used to spatially average five satellite-derived environmental variables, (NDVI, Albedo, Emissivity31, Emissivity32) which were used as predictors of T_{air} in the three models.

In addition, a difficult task with ANN involves choosing the hidden nodes' number. Here, the ANN with one layer was used and the hidden nodes' number was determined using error and trials. For the ANFIS model, Gaussian membership function (MF) and 250 iterations were used. Different number of membership functions were tested and the best of which gave the minimum RMSE and was selected, which was 4 MFs for each variable. For the adjustment, the parameter in SVR model, the simulated annealing was applied. The ANN, ANFIs and SA-SVR models are compared in the test phase based on Table 6-11. The ANN model, among the six LC/LU during the day and night time performed better than the two other models with RMSE which ranged from 1.83°C to 2.53°C and from 2.13°C to 2.35°C during the day and night time respectively. The RMSE of SA-SVR model is ranged from 2.07 to 2.79 °C during the day and night time over six LC/LU and also the highest RMSE was observed in the ANFIS model with a range from 2.64°C to 3.70°C during the day and night time over six LC/LU. These results indicated that the ANN model out performs the SA-SVR and ANFIS models for almost whole LC/LU during the day and night time, but based on Figure 14 and the cross-validation results, the SA-SVR and ANN models out performs the ANFIS model. Moreover, the results showed that there was a high similarity between the training and testing table which demonstrates that the over-fitting has not been occurred in the SA-SVR, ANFIS and ANN. The Q-Q diagram of SA-SVR, ANFIS and ANN shows that all three models slightly tended to underestimate and overestimate the extreme and low temperature for all LC/LU during the day and night time. The weak performance in the extreme and low temperature are a consequence of a small number of data in these temperatures. In these cases, the generalization of these models reduce for estimating the high and low temperature. In addition, despite moderate to high correlations between LST and T_{air}, LST cannot be directly used for estimating air temperature due to the large difference in MBE (Table 3-5), while by applying some additional parameters, in three models (Table 6-11), It can be seen that the MBE was reduced notably, in all LC/LU during day and night time.

Moreover, prediction of long-term monthly air temperature using ANFIS and ANN had been done in the study of Kisi and Shiri (2014). They applied station latitude, longitude and altitude values as input variable to predict the long-term monthly temperature values. They found that the ANN models generally performed better than the ANFIS model in the test period. The ANN models generally performed better than the ANFIS model in the test period and they found that for the ANN model, the maximum and minimum determination coefficient values were between 0.921 and 0.995. The maximum and minimum determination coefficient values were found as 0.99 and 0.876 for the ANFIS model in different stations. Testing results of the ANN and ANFIS models in the study of Kisi and Shiri (2014) show the RMSE values range from 0.1.53 to 4.20◦C and 1.18 ◦C to 9.25◦C for each station, respectively.

Moreover, in the study of Xu. et al. (2014), they applied spatially averaged values of LST, NDVI, modified normalized difference water index (MNDWI), latitude, longitude, distance to ocean, altitude, albedo and solar radiation as predictors of T_{air} in linear regression and random forest models for estimating T_{air} in summer periods from 2003 to 2012. In their study, prior to model development, they also investigated the window size effect on the relationship between LST and T_{air}. The Cross-validation results of their study show that the random forest model (MAE = 2.02°C, R^2 = 0.74) outperforms the linear regression model (MAE = 2.41°C, R^2 = 0.64) and the distribution of residuals from the random forest model slightly overestimates T_{air}, with a mean residual value of 0.09°C.

To sum up, in our study, instead of estimation monthly air temperature and only using the geographical input data, we estimate air temperature during day and night. Moreover, different parameters such as NDVI, Albedo, relative humidity, wind speed, wind direction and Julian day have been take into consideration, which are representative of seasonal changes. The satisfactory results suggested that this modelling approach is appropriate for estimating air temperature in Berlin over six different LC/LU. In addition, the results indicate that MODIS time series of LST can be successfully combined with ground measurements of temperature to produce accurate and more detailed predications of temperature during day and night time. Although the air temperature estimated from satellites tends to be higher than ground-based measurement, the use of satellite remote sensing data can help to overcome the spatial problem of estimating T_{air} particularly in areas with low station density using satellite-based land surface temperature estimation and ground-based relationship between LST and T_{air}. To reduce the biases in satellite-estimated air temperature, it can be effective to use retrieval method based on land surface heat budget (e.g. Kato and Yamaguchi, 2005) in future work.

Acknowledgments

This research was supported by Helmholtz graduate research school GeoSim. We would like thank Dr. Klaus Müller and Thomas Dümmel from the Institute of Meteorology, Freie Universität Berlin, who provided meteorological data in this research. We would also like thank Patricia Margerison for proofreading this manuscript. Our thanks also go to anonymous reviewers for their very thorough and constructive comments.

References

Abraham, A. 2005. *Adaptation of fuzzy inference system using neural learning*. Nedjah, Nadia; de Macedo Mourelle, Luiza, Fuzzy Systems Engineering. Theory and Practice, Studies in Fuzziness and Soft Computing, Springer Verlag, Germany, pp.53-83.

Aarts, E.H.L. and Korst, J. 1989. *Simulated annealing and boltzmann machines: a stochastic approach to combinatorial optimization and neural computing*. Wiley.

Aarts, E.H.L. and Lenstra, J.K. 1997. *Local search in combinatorial optimization*. Chichester, Wiley.

Abramson, D., Krishnamoorthy, M. and Dang, H. 1999. Simulated annealing cooling schedules for the school timetabling problem. *Asia-Pac. J. Oper. Res.*, 16, pp.1-22.

Anily, S. and Federgruen, A. 1987. Simulated annealing methods with general acceptance probabilities. *J. Appl. Probab.*, 24, pp.657-667.

Allen, D.M. 1974. The relationship between variable selection and data augmentation and a method for prediction. *Technometrics*, 16, pp.125-127.

Armstrong, J.S. and Collopy, F. 2002. Error measures for generalizing about forecasting methods: empirical comparisons. *Int. J. Forecast*, 8(1), pp.69-80.

Asthana, S. and Bhujade, R.K. 2011. Handwritten Multiscript Pin Code Recognition System having Multiple hidden layers using Back Propagation Neural Network. *International Journal of Electronics Communication and Computer Engineering*.

Ben-Ameur, W. 2004. Computing the initial temperature of simulated annealing. *Comput. Optim. Appl.*, 29, pp.369-385.

Bose, N.K. and Liang, P., 1996. *Neural network fundamentals with graphs, algorithms and applications*. McGRaw-Hill Series in Electrical and Computer Engineering. New York: McGraw-Hill.

Bodri, L. and Cermak, V. 2000. Prediction of extreme precipitation using a neural network: application to summer flood occurrence in Moravia. *Advances in Engineering Software*, 31, pp.311-321.

Boger, Z. and Guterman, H. 1997. *Knowledge extraction from artificial neural network models*. IEEE Systems, Man, and Cybernetics Conference, Orlando, FL, USA.

Berry, M.J.A. and Linoff, G. 1997. *Data Mining Techniques*. NY: John Wiley & Sons.

Blum, A. 1992. *Neural Networks in C++*. NY: Wiley.

Benali, A., Carvalho, A., Nunes, J., Carvalhais, N., and Santos, A. 2012. Estimating air surface temperature in Portugal using MODIS LST data. *Remote Sensing of Environment*, 124, pp.108-121.

Brunel, J. 1989. Estimation of sensible heat flux from measurements of surface radiative temperature and air temperature at two meters: application to determine actual evaporation rate. *Agricultural and Forest Meteorology*, 46(3), pp.179-191.

Beier, C.M., Signell, S.A., Luttman, A. and DeGaetano, A.T. 2012. High-Resolution Climate Change Mapping with Gridded Historical Climate Products. *Landsc. Ecol.*, 27, pp.327-342.

Bilgili, M. and Sahin, B. 2010. Prediction of long-term monthly temperature and rainfall in Turkey. *Energy Sources, Part A*, 32, pp.60-71.

Chai, T. and Draxler, R.R. 2014. Root mean square error (RMSE) or mean absolute error (MAE) – Arguments against avoiding RMSE in the literature. *Geosci Model Dev.*, 7(3), pp.1247-1250.

Czajkowski, K.P., Goward, S.N., Stadler, S. and Walz, A. 2000. Thermal remote sensing of near surface environmental variables: Application over the Oklahoma Mesonet. *The Professional Geographer*, 52(2), pp.345-357.

Cresswell, M.P., Morse, A.P., Thomson, M.C., and Connor, S.J. 1999. Estimating surface air temperatures, from Meteosat land surface temperatures, using an empirical solar zenith angle model. *International Journal of Remote Sensing*, 20(6), pp.1125-1132.

Chen, Y., Randerson, J.T., Morton, D.C., DeFries, R.S., Collatz, Y. and Marlier, M.E. 2011. Forecasting fire season severity in south america using sea surface temperature anomalies. *Science*, 334, pp.787-791.

Cristóbal, J. Ninyerola, M. and Pons, X. 2008. Modeling air temperature through a combination of remote sensing and GIS Data. *Journal of Geophysical Research*, 113, D13106.

Cheval, S. and Dumitrescu, A. 2009. The July urban heat island of Bucharest as derived from MODIS images. *Theor Appl Cli-matol*, 96(1), pp.145-153.

Caselles, V., Coll, C., Valor, E. and Rubio, E. 1995. Mapping land surface emissivity using AVHRR data: Application to La Mancha, Spain. *Remote Sensing Reviews*, 12, pp.311-330.

Chen, K.Y. and Wang, C.H. 2007. Support vector regression with Genetic algorithms in forecasting tourism demand. *Tourism Management*, 28, pp.215-226.

Doukim, C.A., Dargham, J.A. and Chekima, A. 2010. *Finding the number of hidden neurons for an MLP neural network using coarse to fine search technique*. Information Sciences Signal Processing and their Applications (ISSPA), pp.606-609.

Doukim, C.A. and Chekima, A. 2010. *Finding the number of hidden neurons for an MLP neural network using coarse to fine search technique*. International Conference on Information Sciences, Signal Processing and their Applications (ISSPA '10), pp.606-609.

Eglese, R.W. 2014. Simulated annealing: a tool for operational research. *Eur. J. Oper. Res.*, 46, pp.271-281.

Evrendilek, F., Karakaya, N., Gungor, K. and Aslan, G. 2012. Satellite-based and mesoscale regression modeling of monthly air and soil temperatures over complex terrain in Turkey. *Expert Systems with Applications*, 39(2), pp.2059-2066.

Florio, E.N., Lele, S.R., Chi Chang, Y., Sterner, R. and Glass, G E. 2004. Integrating AVHRR satellite data and NOAA ground observations to predict surface air temperature: a statistical approach. *International Journal of Remote Sensing*, 25(15), pp.2979-2994.

Fleischer, M.A. 1995. *Simulated annealing: Past, present, and future*. In: Alexopoulos, C., Kang, K., Lilegdon, W.R. and Goldsman, D., (eds.) Proceedings of the 1995 Winter Simulation Conference, pp. 155-161. IEEE Press, Arlington, Virginia.

Fnaiech, F., Fnaiech, N. and Najim, M. 2001. *A new feed forward neural network hidden layer neuron pruning algorithm*. IEEE International Conference on Acoustics, Speech, and Signal Processing, Proceedings. (ICASSP '01).

Geisser, S. 1975. The predictive sample reuse method with applications. *Journal of the American Statistical Association*, 70, pp.320-328.

Goward, S.N., Cruickshanks, G.D., and Hope, A.S. 1985. Observed relation between thermal emission and reflected spectral radiance of a complex vegetated landscape. *Remote Sensing of Environment*, 18(2), pp.137-146.

Galvão, C., Valença, M., Vieira, V., Diniz, L., Lacerda, E., Carvalho, A. and Ludermir, T. 1999. *Sistemas inteligentes: Aplicações a recursos hídricos e ciências ambientais*. Porto Alegre, UFRGS/ABRH, 246 pages.

Gardner, M.W. and Dorling, S.R. 1998. Artificial neural networks (the multilayer perceptron) - A review of applications in the atmospheric sciences. *Atmospheric Environment*, 32(14-15), pp.2627-2636.

Ghent, D., Kaduk, J. and Balzter, H. 2010. Assimilation of land surface temperature into the land surface model JULES with an ensemble Kalman filter. *J. Geophys. Res. Atmos*.

Hájek, P. and Olej, V. 2012. Ozone prediction on the basis of neural networks, support vector regression and methods with uncertainty. *Ecol. Inform.*, 12, pp.31-42.

Henderson, D., Jacobson, S.H. and Johnson, A.W. 2003. *Handbook of Metaheuristics*. Kluwer, Boston, MA.

Haykin, S. 1999. *Neural networks: a comprehensive foundation*. Prentice Hall, Upper Saddle River, NJ.

Haykin, S. 2001. *Redes Neurais: princípios e prática*. Porto Alegre: Editora Bookman, 900 pages.

Hastie, T., Tibshirani, R. and Friedman, J. 2009. *The elements of statistical learning: data mining*. Inference and Prediction, Springer, New York.

Hu, G., Hu, L. and Liu, W. 2010. *Grid resources Prediction with support vector regression and particle swarm optimization*. International Joint Conference on Computational Sciences and Optimization (CSO'10), 1, pp.417-422.

Hong, J.S. 2007. *A literature review of wind forecasting technology in the world*. IEEE Lausanne Power Tech, pp.504-509.

Hunter, D., Hao, Y. and Wilamowski, B.M. 2012. Selection of proper neural network sizes and architecture- A comparative study. *IEEE Trans Indust Inf.*, 8(2), pp.228-240.

IPCC Climate Change. 2007. *The Physical Science Basis*. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change. In: Solomon, S. and Miller, H.L. (Eds.), United Kingdom: Cambridge University Press, Cambridge.

Ito, K. and Nakano, R. 2005. Optimizing Support Vector regression hyper-parameters based on cross-validation. *International Joint Conference on Neural Networks*, 3, pp.871-876.

Islam, M.M. and Murase, K. 2001. A new algorithm to design compact two-hidden-layer artificial neural networks. *Neural Networks*, 14, , pp.1265-1278.

Jang, Jyh-Shing R. 1991. *Fuzzy modeling using generalized Neural Networks and Kalman Filter Algorithm (PDF)*. National Conference on Artificial Intelligence, Anaheim, CA, USA, pp.762-767.

Jang, J.S.R. 1993. ANFIS: adaptive-network-based fuzzy inference system. *IEEE Transactions on Systems, Man and Cybernetics*, 23(3).

Jang, Sun, Mizutani. 1997. *Neuro-fuzzy and soft computing*. Prentice Hall, pp.335-368.

Jang, J.D., Viau, A.A. and Anctil, F. 2004. Neural network estimation of air temperatures from AVHRR data. *International Journal of Remote Sensing*, 25(21), pp.4541-4554.

Jin, M. and Dickinson, R.E. 2010. Land surface skin temperature climatology: benefitting from the strengths of satellite observations. *Environmental Research Letters*, 5(4), p.044004.

Jin, M. 2004. Analysis of land skin temperature using AVHRR observations. *Bulletin of American Meteorological Society*, 85, pp.587-600.

Jain, A., McClendon, R.W., Hoogenboom, G. and Ramyaa, R. 2003. *Prediction of frost for fruit protection using artificial neural networks*. American Society of Agricultural Engineers, St Joseph, MI, ASAE Paper 03-3075.

Korrmann, M. and Unbehauen, H. 1988. Structure detection in the identification of nonlinear systems. *Automatique productive informatique Industrielle (APH)*, 22, pp.5-25.

Keerthi, S. 2002. Efficient tuning of SVM hyper parameters using radius/margin bound and iterative algorithms. *IEEE Transactions on Neural Networks*, 13, pp.1225-1229.

Koulamas, C., Antony, S.R. and Jaen, R. 1994. A survey of simulated annealing applications to operations research problems. *OMEGA-Int. J. Manage. Sci.*, 22, pp.41-56.

Kazuhiro Shin-ike. 2010. *A two phase method for determining the number of neurons in the hidden layer of a 3-layer neural network*. SICE Annual Conference, The Grand Hotel, Taipei, Taiwan.

Ke, J. and Liu, X. 2008. Empirical analysis of optimal hidden neurons in neural network modeling for stock prediction. *Pac Asia Work Comput Intell Indust Appl.*, 2, pp.828-832.

Kisi, O. and Shiri, J. 2011. Precipitation forecasting using wavelet-genetic programming and wavelet-neuro-fuzzy conjunction models. *Water Resource Management*, 25(13), pp.3135-3152.

Luk, K.C., Ball, J.E. and Sharma, A. 2000. A study of optimal model lag and spatial inputs for artificial neural network for rainfall forecasting. *Journal of Hydrology*, 227, pp.56-65.

Land Surface Temperature - Copernicus Global Land Service. 2016. Available online: http://land.copernicus. eu/global/products/lst (accessed on 1 March 2016).

Land Surface Temperature: Global Maps. 2016. Available online:http://earthobservatory.nasa.gov/GlobalMaps/view.php?d1=MOD11C1_M_LSTDA. Accessed on 1 March 2016.

Maier, H.R., Jain, A., Dandy, G.C. and Sudheer, K.P. 2010. Methods used for the development of neural networks for the prediction of water resource variables in river systems: Current status and future directions. *Environ. Model. Softw.*, 25, pp.891-909.

Muller, M. and Fill, H. 2003. *Redes Neurais aplicadas na propagação de vazões, in: Simpósio Brasileiro de Recursos Hídricos*. Curitiba, Brazil, unpaginated CD-Rom Proceedings.

Mostovoy, G.V., King, R.L., Reddy, K.R., Kakani, V.G. and Filippova, M.G. 2006. Statistical Estimation of Daily Maximum and Minimum Air Temperatures from MODIS LST Data over the State of Mississippi. *GI Science & Remote Sensing*, 43(1), pp.78-110.

Mittal, G.S. and Zhang, J. 2003. Artificial neural network-based psychrometric predictor. *Biosystems Engineering*, 85(3), pp.283-289.

Maqsood, I., Khan, M.R. and Abraham, A. 2004. An ensemble of neural networks for weather forecasting. *Neural Computing & Applications*, 13, pp.112-122.

Maddala, G.S. 2011. *Introduction to econometrics*. 3rd ed. John Wiley & Sons.

Meteotest. 2010. Meteonorm handbook, Part III: Theory Part 2. Accessed online in 9th February 2011. Availabe from: http://www.meteonorm.com/media/pdf/mn6_software.pdf.

Manzo-Delgado, L., Sánchez-Colón, S. and Álvarez, R. 2009. Assessment of Seasonal Forest Fire Risk Using Noaa-Avhrr: A Case Study in Central Mexico. *Int. J. Remote Sens.*, 30, pp.4991-5013.

Marzban, F., Sodoudi, S. and Preusker, R. 2017. The influence of land cover type on the relationship between NDVI-LST and LST-T_{Air}. *International Journal of Remote Sensing*.

Cristianini, N. and Taylor, L. 2000. *An introduction to support vector machines*. Cambridge University Press, Cambridge, UK.

Nieto, H., Sandholt, I., Aguado, I., Chuvieco, E. and Stisen, S. 2011. Air temperature estimation with MSG-SEVIRI data: Calibration and validation of the TVX algorithm for the Iberian Peninsula. *Remote Sensing of Environment*, 115(1), pp.107-116.

Near Surface Air Temperature - GES DISC-Goddard Earth Sciences Data and Information Services Center. 2016. Available online: http://disc.gsfc.nasa.gov/hydrology/dataholdings/parameters/near_surf_air_temp.shtml. Accessed on 17 January 2016.

Nguyen, O. and Suwandana, E. 2015. Temporal change and its spatial variety on land surface temperature and land use changes in the Red River Delta, Vietnam, using MODIS time series imagery. *Environ Monit Assess*.

Oyler, J.W., Ballantyne, A. and Running, S.W. 2015. Creating a topoclimatic daily air temperature dataset for the conterminous united states using homogenized station data and remotely sensed land skin temperature. Int. J. Climatol, 35, pp.2258-2279.

Onoda, T. 1995. *Neural network information criterion for the optimal number of hidden units*. iIEEE International Conference on Neural Networks, pp.275-280.

Panchal, G., Ganatra, A. and Panchal, D. 2011. Behaviour analysis of multilayer perceptrons with multiple hidden neurons and hidden layers. *Int J Comput Theory Eng.*, 3(2), pp.332-337.

Prihodko, L. and Goward, S.N. 1997. Estimation of air temperature from remotely sensed surface observations. *Remote Sensing of Environment*, 60(3), pp.335-346.

Prince, S., Goetz, S., Dubayah, R., Czajkowski, K. and Thawley, M. 1998. Inference of surface and air temperature, atmospheric precipitable water and vapor pressure deficit using Advanced Very High-Resolution Radiometer satellite observations: comparison with field observations. *Journal of Hydrology*, 212-213, pp.230-249.

Peón, J., Carmen, R. and Javier, F.C. 2014. Improvements in the estimation of daily minimum air temperature in peninsular spain using MODIS land surface temperature. *Int. J. Remote Sens.*, 35, pp.5148-5166.

Panchal, G. and Panchal, D. 2011. Behaviour analysis of multilayer perceptrons with multiple hidden neurons and hidden layers. *International Journal of Computer Theory and Engineering*, 3, pp.332-337.

Razavi, S. and Tolson, B.A. 2011. A New Formulation for Feedforward Neural Networks. *IEEE Trans. Neural Netw.*, 22, pp.1588-1598.

Romeo, F. and Sangiovanni-Vincentelli, A. 1991. A theoretical framework for simulated annealing. *Algorithmic*, 6, pp.302-345.

Rivals, I. and Personnaz, L. 2000. *A statistical procedure for determining the optimal number of hidden neurons of a neural model*. Second International Symposium on NeuralComputation (NC'2000), Berlin.

Reich, P.B., Luo, Y. and Oleksyn, J. 2014. Temperature drives global patterns in forest biomass distribution in Leaves, Stems, and Roots. *Proc. Natl. Acad. Sci. USA*, 111, pp.13721-13726.

Ruane, A.C., McDermid, S. and Cecil, L.D. 2014. Carbon-temperature-water change analysis for peanut production under climate change: a prototype for the agmip coordinated climate-crop modeling project (C3mp). *Glob. Chang. Biol.*, 20, pp.394-407.

Rosenzweig, C., Elliott, J. and Khabarov, N. 2014. Assessing Agricultural Risks of Climate Change in the 21st Century in a Global Gridded Crop Model Intercomparison. *Proc. Natl. Acad. Sci. USA*, 111, pp.3268-3273.

Snyder, W.C., Wan, Z., Zhang, Y. and Feng, Y.Z. 1998. Classification based emissivity for land surface temperature measurement from space. *International Journal of Remote Sensing*, 19, pp.2753-2774.

Sobrino, J.A., Raissouni, N. and Li, Z.L. 2001. A comparative study of land surface emissivity retrieval from NOAA data. *Remote Sensing of Environment*, 75, pp.256-266.

Shuxiang, X. and Chen, L. 2008. *A novel approach for determining the optimal number of hidden layer neurons for FNN's and its application in data mining*. International Conference on Information Technology and Application, pp.683-686.

Sousa, S., Martins, F., Alvimferraz, M. and Pereira, M. 2007. Multiple linear regression and artificial neural networks based on principal components to predict ozone concentrations. *Environ. Model. Softw.*, 22, pp.97-103.

Sartakhti, J.S. and Mozafari, K. 2011. Hepatitis disease diagnosis using a novel hybridmethod based on support vector machine and simulated annealing (SVM-SA). *Computer Methods and Programs in Biomedicine*, 108, pp.570-579.

Stisen, S., Sandholt, I. and Eklundh, L. 2007. Estimation of diurnal air temperature using MSG SEVIRI data in West Africa. *Remote Sensing of Environment*, 110(2), pp.262-274.

Steidley, C., Sadovski, A., Tissot, P. and Bachnak, R. 2005. *Using an artificial neural network to improve predictions of water level where tide charts fail*. In: Ali, M. (Editors), Innovations in Applied Artificial Intelligence, Springer, Bari, Italy.

Suman, B. and Kumar, P. 2006. A survey of simulated annealing as a tool for single and multi objective optimization. *J. Oper. Res. Soc.*, 57, pp.1143-1160.

Sugeno, M. and Tanaka, K. 1992. Successive identification of a fuzzy model and its applications to prediction of a complex system. *Proc Fuzzy Sets Syst.*, 42, pp.315-334.

Smith, B.A., McClendon, R.W. and Hoogenboom, G. 2006. Improving Air Temperature Prediction with Artificial Neural Networks. *International Journal of Computational Intelligence*, 3(3), pp.179-186.

Stone, M., 1974. Cross-validation choice and assessment of statistical predictions. *Journal of the Royal Statistical Society*, 36, pp.111-147.

Sandholt, I., Rasmussen, K. and Andersen, J. 2002. A simple interpretation of the surface temperature/vegetation index space for assessment of surface moisture status. *Remote Sensing of Environment*, 79(2-3), pp.213-224.

Smola, A.J. and Schölkopf, B. 2004. A tutorial on support vector regression. *Statistics and Computing*, 14, pp.199-222.

Sun, Y.J., Wang, J.F., Zhang, R.H., Gillies, R.R., Xue, Y. and Bo, Y.C. 2005. Air temperature retrieval from remote sensing data based on thermodynamics. *Theoretical and Applied Climatology*, 80(1), pp.37-48.

Shamir, E. and Georgakakos, K.P. 2014. MODIS land surface temperature as an index of surface air temperature for operational snowpack estimation. *Remote Sens. Environ.*, 152, pp.83-98.

Smith, B.A., McClendon, R.W. and Hoogenboom, G. 2005. An enhanced artificial neural network for air temperature prediction. *World Academy of Science, Engineering and Technology (PWASET)*, 7, pp.7-12.

Shank, D.B., Hoogenboom, G. and McClendon, R.W. 2008. Dewpoint temperature prediction using artificial neural networks. *Journal of Applied Meteorology and Climatology*, 47, pp.1757-1769.

Stahl, K.R.D., Moore, J.A. Floyer, M.G. and Mckendry, I.G. 2006. Comparison of approaches for spatial interpolation of daily air temperature in a large region with complex topography and highly variable station density. *Agricultural and Forest Meteorology*, 139, pp.224-236.

Schölkopf, B. 1998. Learning with kernels, Ph.D. Thesis, GMD, Birlinghoven, Germany.

Takagi, T. and Sugeno, M. 1985. Fuzzy identification of systems and its applications to modeling and control. *IEEE Trans Syst Man Cybern*, 15(1), pp.116-32.

Unger, J., Gál, T. and Tobak, Z. 2009. *Air temperature versus surface temperature in urban environment*. International Conference on Urban Climate, Yokohama, Japan.

Ustün, B., Melssen, W.J. and Oudenhuijzen, M. 2005. Determination of optimal support vector regression parameters by genetic algorithms and simplex optimization. *Analytica Chimica Acta*, 544, pp.292-305.

Van Laarhoven, P.J.M. 1988. *Theoretical and computational aspects of simulated annealing*. Centrum voor Wiskunde en Informatica, Amsterdam, Netherlands.

Vancutsem, C., Ceccato, P., Dinku, T., and Connor, S.J. 2010. Evaluation of MODIS land surface temperature data to estimate air temperature in different ecosystems over Africa. *Remote Sensing of Environment*, 114(2), pp.449-465.

Van Laarhoven, P.J.M. and Aarts, E.H.L. 1987. *Simulated annealing: Theory and applications*. D. Reidel; Kluwer, Dordrecht, Boston, Norwell, MA.

Van Leeuwen, T., Frank, D., Jin, Y., Smyth, P. and Randerson, J. 2011. Optimal use of land surface temperature data to detect changes in tropical forest cover. *J Geophys Res*., 116, p.G02002.

Vogt, J., Viau, A.A. and Paquet, F. 1997. Mapping regional air temperature fields using satellite derived surface skin temperatures. *International Journal of Climatology*, 17, pp.1559-1579.

Vazquez, D.P., Reyes, F.J.O. and Arboledas, L.A. 1997. A comparative study of algorithms for estimating land surface temperature from AVHRR data. *Remote Sensing of Environment*, 62, pp. 215-222.

Wloczyk, C., Borg, E., Richter, R., and Miegel, K. 2011. Estimation of instantaneous air temperature above vegetation and soil surfaces from Landsat 7 ETM+ data in northern Germany. *International Journal of Remote Sensing*, 32(24), pp.9119-9136.

Wedge, D., Ingram, D., McLean, D., Mingham, C. and Bandar, Z. 2005. *A global-local artificial neural network with application to wave overtopping prediction*. In: Duch, W. and Kacprzyk, J. (Editors), Artificial Neural Networks: Formal Models and Their Applications – ICANN, Springer, Warsaw, Poland.

Willmott, C.J. and Matsuura, K. 2005. Advantages of the Mean Absolute Error (MAE) over the Root Mean Square Error (RMSE) in Assessing Average Model Performance. *Clim. Res.*, 30, pp.79-82.

Wan, Z., Zhang, Y. and Zhang, Q. 2004. Quality assessment and validation of the MODIS global land surface temperature. *Int. J. Remote Sens.*, 25, pp.261-274.

Wan, Z.. 2008. Radiance-based validation of the V5 MODIS land-surface temperature product. *Int. J. Remote Sens.*, 29, pp.5373-5395.

Westerling, A.L., Hidalgo, H.G., Cayan, D.R. and Swetnam, T.W. 2006. Warming and Earlier Spring Increase Western U.S. *Forest Wildfire Activity*, 313, pp.940-943.

Wang, J. and Tan, Z. 2012. An annual load forecasting model based on support vector regression with differential evolution algorithm. *Applied Energy*, 94, pp.65-70.

Xu, Y., Qin, Z. and Shen, Y. 2012. Study on the estimation of near-surface air temperature from MODIS data by statistical methods. *International Journal of Remote Sensing*, 33(24), pp.7629-7643.

Xu, Y. and Shen, Y. 2013. Recunstruction of the land surface temperature time series using harmonic analysis. *Computers & Geosciences*, 61, pp.126-132.

Yoo, J.M., Won, Y.I., Cho, Y.J., Jeong, M.J., Shin, D.B., Lee, S.J., Lee, Y.R., Oh, S.M., and Ban, S.J. 2011. Temperature trends in the skin/surface, midtroposphere and low stratosphere near Korea from satellite and ground measurements. *Asia-Pacific Journal of Atmospheric Sciences*, 47, pp.439-455.

Yang, Y. 2007b. Consistency of cross validation for comparing regression procedures. *The Annals of Statistics*, 35, pp.2450-2473.

Yuan, H.C. and Huai, X.Y. 2003. Amethod for estimating the number of hidden neurons in feed-forward neural networks based on information entropy. *Computers and Electronics in Agriculture*, 40, pp.57-64.

Zhu, W. and Jia, S. 2013. Estimation of daily maximum and minimum air temperature using MODIS land surface temperature products. *Remote Sensing of Environment*, 130, pp.62-73.

Zhang, W., Huang, Y., Yu, Y. and Sun, W. 2011a. Empirical models for estimating daily maximum, minimum and mean air temperatures with MODIS land surface temperatures. *International Journal of Remote Sensing*, 32(1), pp.1-26.

Zaksek, K. and Schroedter-Homscheidt, M. 2009. Parameterization of air temperature in high temporal and spatial resolution from a combination of the SEVIRI and MODIS instruments. *ISPRS Journal of Photogrammetry and Remote Sensing*, 64(4), pp.414-421.

Zheng, X., Zhu, J. and Yan, Q. 2013. Monthly air temperatures over Northern China estimated by integrating MODIS Data with Gis Techniques. *J. Appl. Meteorol. Climatol*, 52, pp.1987-2000.