Meteorologische Zeitschrift, Vol. 14, No. 5, 677-686 (October 2005)

Article

The dependence of the urban heat island intensity on latitude – A statistical approach

Dedicated to the founder of urban climatology in Germany, Father Albert KRATZER (13.08.1905–13.04.1975) in memory of his 100^{th} birthday.

UWE WIENERT*[1] and WILHELM KUTTLER[2]

[1]Bad Münder, Germany
[2]Institute of Geography, University of Duisburg-Essen, Essen, Germany

(Manuscript received July 27, 2004; in revised form June 6, 2005; accepted August 1, 2005)

Abstract

The question about a possible dependence of the Urban Heat Island (UHI) on latitude and its causes has not been solved satisfactorily up to now. This paper tries to give an answer to the problem by extracting the UHI of cities on a global scale from the available international scientific literature and treating them with various statistical methods. In addition parameters like the city population, use of primary energy, topographic features, height above sea level and the energy balance of the earth's surface are included into the statistical calculations in order to substantiate a latitudinal variation of the UHI. Three main results could be derived from the statistical examinations, namely: 1. The UHI shows a trend to increase from low to high latitudes 2. The part of the observed variance of the UHI explained by its latitudinal variation, however, remains relatively small with about 6 %. 3. A statistically significant dependence of the UHI on latitude is based mainly on the latitudinal variation of anthropogenic heat production and radiation balance.

Zusammenfassung

Die Frage nach einer möglichen Breitenabhängigkeit der städtischen Wärmeinsel (UHI) und deren Ursachen ist bis heute nicht zufriedenstellend beantwortet. Die vorliegende Publikation versucht zur Lösung des Problems beizutragen, indem aus der frei zugänglichen internationalen Fachliteratur Daten zur UHI von Städten im globalen Umfang entnommen und anschließend mit unterschiedlichen statistischen Methoden bearbeitet wurden. Parameter wie die Größe der Stadtpopulation, der Verbrauch an Primärenergie, topographische Merkmale, die Geländehöhe über dem Meeresspiegel und die Energiebilanz der Erdoberfläche werden in die statistischen Berechnungen miteinbezogen, um eine Breitenabhängigkeit der UHI zu untermauern. Die statistischen Untersuchungen führten zu drei wesentlichen Ergebnissen: 1. Die UHI zeigt die Tendenz einer Zunahme von niederen zu hohen Breiten. 2. Der Anteil der beobachteten Varianz der UHI, der durch ihre Breitenabhängigkeit erklärt wird, bleibt allerdings mit etwa 6 % relativ klein. 3. Eine statistisch signifikante Breitenabhängigkeit der UHI wird hauptsächlich durch eine Breitenvariation der anthropogenen Wärmeproduktion und der Strahlungsbilanz bedingt.

1 Introduction

Developing higher air and surface temperatures compared to the open surrounding countryside that is defined as "urban heat island" (UHI) represent the most well-known and probably best examined phenomenon of the urban climate. This concept comprises the occurring island-like temperature surplus in contrast to the natural environment of the city being observed usually during the night, but in less frequent cases also during daylight hours (GARSTANG et al., 1975). Finally these effects are caused by changes in the urban radiation and heat balance of the earth's surface, in the urban canopy layer or the urban boundary layer being most pronounced in weather situations with low winds and few or no clouds (OKE, 1973, 1995). Considering these weather conditions the UHI might become relevant in the area of human biometeorology concerning thermal comfort (MAYER, 1993) and the supply of cooler air to the cities (KUTTLER et al., 1998; WEBER and KUTTLER, 2004). The intensity and the temporal occurrence of the UHI are mainly controlled by local factors (KUTTLER, 2004a, b). On a larger scale impacts, such as the latitudinal location of the city, could be responsible for modifications of the UHI, too.

Numerical simulations performed by ATWATER (1977) for example considering hypothetical cities in tropical, subtropical, middle and polar latitudes hint towards a dependence of the UHI on radiative processes, day and night lengths and anthropogenic heat produc-

*Corresponding author: Uwe Wienert, Steinkreuzer Weg 8, 31848 Bad Münder, Germany, e-mail: u.wienert@t-online.de

0941-2948/2005/0069 $ 4.50
DOI: 10.1127/0941-2948/2005/0069

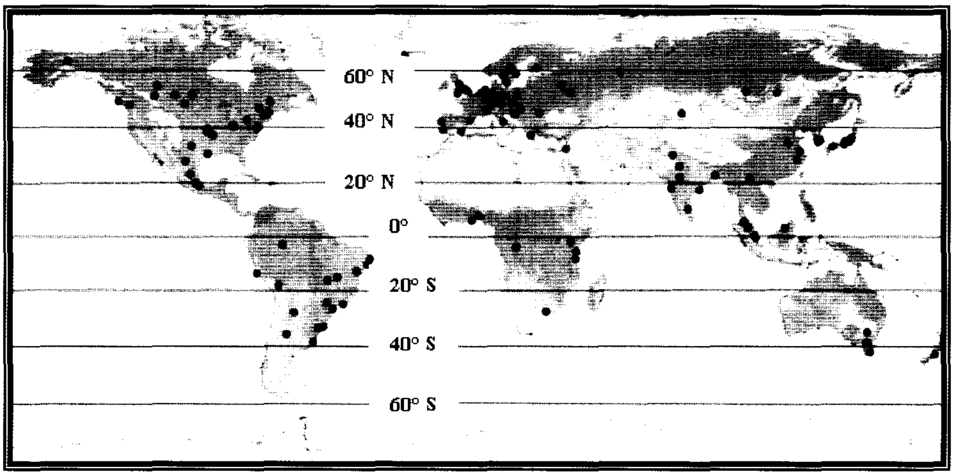

Figure 1: Geographical position of 150 towns and cities used in statistical calculations (due to the size of symbols locations partly covered by each other).

tion. First assumptions about the intensity of the UHI possibly being altered by the latitudinal location of the city already came up a long time ago. The Benedictine monk Albert KRATZER found anomalies by comparing the horizontal temperature gradients of Calcutta (22° 39' N) and Berlin (52° 30' N). The annual mean UHI of Calcutta had a value of only 0.4 K whereas the one of Berlin reached 1.0 K meaning more than a twofold increase of its intensity (KRATZER, 1937).

Since there has not been available a comprehensive evaluation of that problem, this study examines a possible variation of the UHI with latitude on a broader global database using different statistical methods. The objective of this study, however, is not limited to just finding a possible correlation of the UHI with latitude. If there exists such a correlation, some important physical and climatological causes for the UHI depending on latitude will be pointed out, too.

2 Data

For evaluating a latitudinal dependence of the UHI a statistical approach was favored. OKE (1986) proposed to analyse available urban climate studies to investigate the spatial transferability of urban climate information. Following that proposal the published international scientific literature over a period from 1929 to 2000 concerning the UHI in different cities around the world served as the data source. Before 1929 there were very few investigations concerning the UHI of mainly European cities. In addition more current measurements of these cities are available now. In most cases no older urban studies than 1929 were used here. 223 cities between latitudes 43° S and 65° N with measurements of the UHI could be extracted from those publications. Altogether 510 quotations of the urban air temperature modification were evaluated. The major part of this literature comprising about 224 papers deals with the UHI in mid-latitudes (between 40° N or S and 60° N or S) whereas 117 publications describe the UHI in the tropics between 20° S and 20° N. The UHI in subtropical cities (between 20° N or S and 40° N or S) is represented in about 169 contributions (WIENERT, 2002).

The expression UHI, however, stands for a variety of terms describing the urban air temperature surplus. There are yearly, monthly, weekly or seasonal means of the UHI as well as climatic trends illustrating the development of the urban warming with time. The different time scales for determining the mean values of the characteristic warming of a city or town ranging from a few weeks to several decades or sometimes a century make it even more difficult to compare all these values with each other let alone using them in statistical calculations. So a term had to be found for the UHI being quite independent from the period of time. The maximum Urban Heat Island intensity abbreviated as UHI_{max} represents this term. UHI_{max} stands for the highest daily warming of an urban area under most favorable weather conditions, such as few or no clouds and low winds. Usually the measuring period extends over a day or at least over the time of day when the most intensive urban heat island is expected.

Fortunately, the UHI_{max} represents the most extensive partial data set of the evaluated scientific literature. Altogether 150 out of 223 cities mentioned in the papers showed measurements of the daily maximum urban heat island intensity. Fig. 1 contains the geographical distribution of these locations. Referring to WIENERT (2002) the data of these cities together with the appropriate literature are not listed here because this table would exceed the available space for this paper.

Different statistical methods such as regression, factorial and cluster analysis were used to prove a possible variation of the UHI_{max} with latitude. In order to find

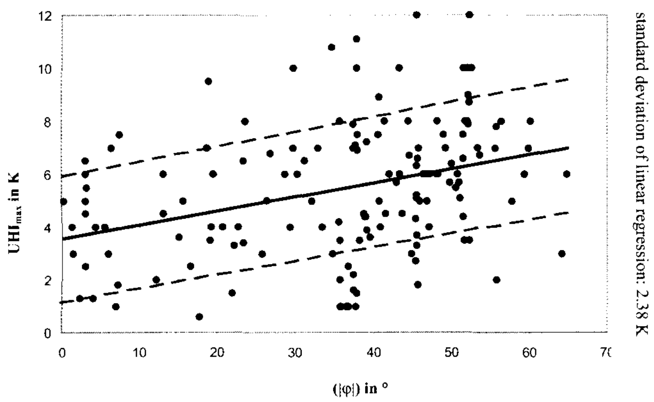

Figure 2: Bivariate regression of the maximum urban heat island intensity (UHI_{max}) with absolute values of latitude $|\varphi|$ (northern and southern hemisphere combined). Solid line designates linear regression, dashed lines designate the standard deviation of linear regression.

the causes for a dependency of the UHI_{max} on latitude further parameters controlling the formation of the UHI were included in the statistical calculations.

These parameters consisted of the size of the city population, anthropogenic heat production in form of the annual per capita consumption of primary energy, height above sea level and topographic features (coast, plain, valley). In addition the climate classification by TERJUNG and LOUIE (1972) based on the energy balance of the earth's surface was inserted into the statistical calculations. The reason for choosing this climate classification consists in the fact that the radiation and heat balance form the energy balance which controls the air temperature near the earth's surface. So this seems to be the most promising climatic factor for proving a connection with the UHI_{max} because differences in the energy balance of the earth's surface mainly determine the formation of an urban air temperature surplus. In the following text the climate classification by TERJUNG and LOUIE will be cited without its year of publication.

3 Statistical calculations and results

The cities with measurements of UHI_{max} according to chapter 2 are now treated by several statistical methods. For further explanations of the statistical calculations used here it is referred to the special literature such as BAHRENBERG et al. (1990a, b); SCHÖNWIESE (1985); WINSTAT (1998).

Bivariate, partial and multiple regression analysis are implemented in chapter 3.1 to find a significant correlation between the UHI_{max} and latitude. The independent variables used in the regressional calculations are briefly discussed especially concerning their preparation for statistical treatment. Partial regression tries to show the importance of a certain variable for the UHI_{max}, excluding the influence of other parameters employed in these calculations.

Factorial analysis is going to investigate in chapter 3.2 to what extent a latitudinal dependence possibly determines the UHI_{max}. For that purpose factors being independent from each other are extracted from the data set. The factors can be characterized by those variables that show the highest correlation with the factor in question. Bivariate regression between one factor as independent and the UHI_{max} as dependent variable explains if this factor has a statistically significant influence on the degree of the urban heat island intensity.

Cluster analysis was used in chapter 3.3 to show whether parameters responsible for the formation of urban excess warming could be combined in typical groups being attached to certain latitudinal zones.

3.1 Bivariate, partial and multiple regressions

The simple linear regression (3.1) of the 150 cities shows an increase of the UHI_{max} with latitude. In these calculations the so-called absolute latitude was used. This means that northern and southern latitudes by convention defined as positive resp. negative values are not differentiated from each other. The reason for combining northern and southern latitudes is based on the minor number of cities from the southern hemisphere. Just 26 of the total 150 cities examined in this paper lie south of the equator.

$$UHI_{\max} = 0.052972 \cdot |\varphi| + 3.5585 \qquad (3.1)$$

with:

- UHI_{max}: daily maximum heat island intensity in K
- $|\varphi|$: absolute latitude in decimal degrees (northern and southern latitudes combined)

The correlation (3.1) explains just 12 % of the observed variance of the UHI_{max} on a significance level of more than 99 %. The regression line and the standard deviation and distinct data points are to be seen in Fig. 2.

This simple bivariate regression, however, does not allow any conclusions concerning those parameters that mainly might cause such a latitudinal dependence. Therefore several multiple regression calculations were performed using the UHI_{max} as the dependent variable and size of city population, annual energy use (indicator for anthropogenic heat production), height above sea level, topography (coast, plain, valley) and the climate classification after TERJUNG and LOUIE as independent variables. The principle equation of multiple regression is outlined in (3.2).

$$\begin{aligned} UHI_{\max} &= b_1 \cdot |\varphi| + b_2 \cdot \log(POP) \\ &+ b_3 \cdot H + b_4 \cdot W_a + b_5 \cdot IN \\ &+ b_6 \cdot OUT + c_1 \cdot T1 + c_2 \cdot T_2 + IT1 + IT2 \end{aligned} \qquad (3.2)$$

with:

- UHI_{max}: daily maximum heat island intensity in K
- $|\varphi|$: absolute latitude in decimal degrees
- POP: city population number
- H: interval of height above sea level
- W_a: interval of primary energy consumption
- IN: input number according to climate classification after TERJUNG and LOUIE
- OUT: output number according to climate classification after TERJUNG and LOUIE
- T1: topographic feature "plain" (for further explanation see below)
- T2; topographic feature "valley" (for further explanation see below)
- (T3): topographic feature "coast" is not necessary to specify in (3.2) since in that case T1 and T2 are zero (for further explanation see below)
- IT1: interactive term with T1
- IT2: interactive term with T2
- b_n, c_n: regression coefficients

The intervals for height of terrain above sea level (H) and primary energy consumption (W_a) of the country assigned to the city under investigation can be taken from Table 1 and Table 2. Both intervals have the dimension [1]. Some of the variables in (3.2) are categorized on an ordinal scale as those in Table 1 and Table 2. If the dependent variable has a metric scale, as is the case with UHI_{max} in (3.2) and the independent variables are of a mixed character containing metric and discrete variables a multiple regression might be performed. The results, however, have to be interpreted carefully, since differences between discrete values of variables categorized on an ordinal scale cannot be directly connected to differences in the original data, for example the height above sea level in Table 1. For this reason the multiple regression has to be tested by other statistical methods such as the variance analysis applied here in form of a discriminance analysis together with cluster calculations in chapter 3.3 and 3.4.

Table 1: Definition of height intervals according to height of terrain above mean sea level (H).

Height interval	Height of terrain above mean sea level in m
0	$0 \leq$ H < 100
1	$100 \leq$ H < 200
2	$200 \leq$ H < 500
3	$500 \leq$ H < 1000
4	H ≥ 1000

Among the independent variables the terms "IN" and "OUT" in (3.2) designate the climate classification after TERJUNG and LOUIE. These input (IN) and output (OUT) numbers describe the energy balance of the earth's surface as defined in (3.3).

$$\underset{Input}{Q + q_{a\downarrow}} = \underset{Output}{q_l + q_s + q_{a\uparrow}} \tag{3.3}$$

with:

- Q: radiation balance in W/m^2
- $q_a \downarrow$: horizontal import of sensible heat (for example by ocean currents) in W/m^2
- q_l: latent heat flux in W/m^2
- qs: sensible heat flux in W/m^2
- $q_a \uparrow$: horizontal export of sensible heat (for example by ocean currents) in W/m^2

The left side of (3.3) indicates the energy gain mainly by solar radiation thus being called the input side. The right side of (3.3) stands for the energy loss of the earth's surface mainly by latent and sensible heat fluxes and is called the output side. TERJUNG and LOUIE defined a special classification scheme considering the maximum monthly mean and the average yearly variation of the energy balance of the earth's surface. In addition the authors compare these values to the maxima occurring on earth.

The classification scheme as used in the statistical calculations is shown in Table 3. Hence it becomes evident that the authors utilize a two-digit number system to characterize the energy balance split up into the input side (radiation balance and horizontal import of sensible heat) and the output side (energy loss by latent and sensible heat fluxes and horizontal export of sensible heat). The first digit stands for the relation between the maximum monthly mean of the energy input or output and their highest values on earth (1: 100 % to 90 %, 2: 89 % to 80 %,..., 9: $\leq$ 19 % of the earth's maximum energy input/output). The second digit describes the yearly variation (range) of the energy input/output compared to

Table 2: Definition of energy intervals according to annual consumption of primary energy (W_a) per capita and country from ATLAS INTERNATIONAL (1989).

energy interval	Annual consumption of primary energy (W_a) in MJ/a
0	$W_a < 14650$
1	$14650 \leq W_a < 73250$
2	$73250 \leq W_a < 146500$
3	$146500 \leq W_a < 293000$
4	$W_a \geq 293000$

Table 3: Climate classification scheme after TERJUNG and LOUIE (1972) based on the energy balance of the earth's surface and modified for use in statistical calculations. Input: radiation balance plus horizontal import of sensible heat. Output: energy loss by vertical fluxes of latent and sensible heat and horizontal export of sensible heat. Range: yearly variation of input and output parameters. Two digit numbers: first digit for input/output, second digit for range in percent of maximum occurring on earth referring to mean monthly values 1: 100 %–90 %, 2: 89 %–80 %......9: 19–0 % of maximum on earth. Examples: Kuala Lumpur: input-number: 28 / output-number: 58. Calgary input-number: 21/ output-number: 55.

Input- / Output →

Range ↑	1	2	3	4	5	6	7	8	9
9	19	29	39	49	59	69	79	89	99
8	18	A	38	48	D	68	78	G	98
7	17	27	37	47	57	67	77	87	97
6	16	26	36	46	56	66	76	86	96
5	15	B	35	45	E	65	75	H	95
4	14	24	34	44	54	64	74	84	94
3	13	23	33	43	53	63	73	83	93
2	12	C	32	42	F	62	72	I	92
1	11	21	31	41	51	61	71	81	91

the maximum range encountered on earth (1: 100 % to 90 %, 2: 89 % to 80 %,..., 9: $\leq$ 19 % of the earth's maximum yearly range).

Both, the energy gain [input side in (3.3)] and the energy loss [output side in (3.3)] of the earth's surface, were now put into numbers according to the classification scheme after TERJUNG and LOUIE. So every city included in the regression analysis was marked by an input and an output number consisting of two digits each.

The topographical features (coast, plain, valley) are defined as dummy variables as listed below:

- T1 = 1: topographic feature "plain" proves true
- T1 = 0: topographic feature "non-plain"
- T2 = 1: topographic feature "valley" proves true
- T2 = 0: topographic feature "non-valley"
- T1 = 0 **and** T2 = 0: feature "coast" proves true

Of the three only two topographic features are necessary in the multiple regression equation (3.2). The third feature "coast" automatically proves true if T1 and T2 are set to zero. The dummy variables concerning topographic features require so-called interactive terms (IT1, IT2) in (3.2). These interactive terms consider the possibility that the dependence of the UHI_{max} on the variables used in the multiple regression might be different for example for cities in valleys and for cities on plains.

In (3.4) the interactive term IT1 with the topographic feature "plain" is shown. The interactive term IT2 with the topographic feature "valley" (T2) is calculated in accordance to (3.4) by replacing IT1 and T1 by IT2 and T2 respectively and the regression coefficients becoming b_{n2}.

$$IT1 = b_{11} \cdot (T1 \cdot |\varphi|) + b_{21} \cdot (T1 \cdot \log(POP)) + b_{31} \cdot (T1 \cdot H) + b_{41} \cdot (T1 \cdot W_a) + b_{51} \cdot (T1 \cdot IN) + b_{61} \cdot (T_1 \cdot OUT) \tag{3.4}$$

with:

- IT1: interactive term with geographic feature "plain"
- $|\varphi|$: absolute latitude in decimal degrees
- POP: city population number
- H: interval of height above sea level
- W_a: interval of primary energy consumption
- IN: input number after TERJUNG and LOUIE
- OUT: output number after TERJUNG and LOUIE
- T1: topographic feature "plain"
- b_{n1}: regression coefficients

With the set of variables mentioned above the multiple regression calculation according to (3.2) was carried out. As a result one obtains a coefficient of determination of 0.54 thus explaining 54 % of the observed variance of the UHI_{max}. The interesting question is which variables contribute most to the statistical explanation of the UHI_{max}. This might be answered by performing different partial regression evaluations.

In the course of partial regression analysis one tries to find the correlation of a parameter, for example UHI_{max}, with a certain independent variable excluding the influence of the other independent variables. In other words: a correlation, for example between the UHI_{max} and the energy interval (indicator for anthropogenic heat production), does not explain the possible causes behind such a dependence. So one might think of the size of the city population or latitude with which anthropogenic heat production could be correlated itself. Excluding the influence of these two variables on the energy interval in a partial regression, might clarify if they have an important impact on how anthropogenic heat production controls the UHI_{max}.

Partial regression analysis shows that of the 20 variables including the interactive terms employed in the multiple regression using the climate classification after TERJUNG and LOUIE only three variables correlate significantly with the UHI_{max}; namely the size of city population, the energy interval as an indicator for anthropogenic heat production and the input number according to the climate classification after TERJUNG and LOUIE as an indicator for the radiation balance of the earth's

Table 4: Independent variables correlating significantly with UHI_{max} in partial regression analysis.
*) logarithmic value of city population number, **) input number after TERJUNG and LOUIE as indicator for radiation balance at the earth's surface, r: correlation coefficient, P: probability of chance.

Independent variable	r	r^2	P
size of city population *)	0.249	0.062	0.004
energy interval	0.155	0.024	0.082
Input number**)	–0.251	0.063	0.004
latitude	0.068	0.005	0.444

surface. Table 4 contains these variables with some statistical parameters.

In this table latitude is listed in addition to be compared with the other independent variables. As can be seen latitude has no significant connection with the UHI_{max} because the partial regression only gives a probability around 44 %. This result might be obtained just by chance. Throughout this study a statistically significant correlation is assumed for P-values (probability of chance) less than 10 %.

The input number after TERJUNG and LOUIE standing for the radiation balance of the earth's surface is defined as the sum of the maximum monthly mean value and the maximum annual variation of the radiation balance. Low input numbers are attached to high values of that sum and vice versa according to Table 3. Since the annual range of the radiation balance exhibits a marked increase towards higher latitudes, the special definition of the input number causes a negative correlation coefficient. So UHI_{max} shows lower values in climates with higher input numbers describing reduced annual ranges and/or reduced maxima of the monthly radiation balance that most frequently occur in tropical zones.

If there exists a dependence of the UHI_{max} on latitude it should be based on a latitudinal variation of the variables cited in Table 4 significantly correlating with the UHI_{max}. The results of partial regression calculations examining the latitudinal dependence of city population, energy interval and input number are listed in Table 5. There it is shown that latitude explains almost 6 % of the observed variance concerning the input number after TERJUNG and LOUIE (indicator for radiation balance) but 43 % of the observed variance concerning the energy interval as an indicator for anthropogenic heat production.

Table 5 hints towards a negative correlation between the size of the city population and latitude meaning higher city populations in lower latitudes. Anthropogenic heat production in form of the energy interval increases with latitude. According to the specific definition of the input number this parameter shows a negative correlation coefficient meaning, however, an increase especially in the annual variability of the radiation balance of the earth's surface with latitude.

Table 5: Independent variables correlating significantly with latitude (φ) in partial regression analysis.
*) logarithmic value of city population number, **) input number after TERJUNG and LOUIE as indicator for radiation balance at the earth's surface, r(φ): correlation coefficient from partial regression with latitude, P(φ): probability of chance from partial regression with latitude.

Independent variable	r(φ)	$r(\varphi)^2$	P(φ)
size of city population *)	–0.21	0.044	0.01
energy interval	0.66	0.430	$8.5 \cdot 10^{-20}$
Input number**)	–0.23	0.056	0.006

All the results discussed above give reason for the assumption that a latitudinal variation of the UHI_{max} is most probably caused by a latitudinal dependence of anthropogenic heat production and to a lesser degree by the radiation balance of the earth's surface. This means higher values of UHI_{max} with growing anthropogenic heat production and increasing variability of the radiation balance.

3.2 Factorial analysis

In order to substantiate the assumed connection between the UHI_{max} and latitude being controlled by a latitudinal dependence of variables like anthropogenic heat production and radiation balance a factorial analysis was performed with the variables used in the multiple regression calculation explained in chapter 3.1. Accordingly these are latitude, city population, energy interval (indicator for anthropogenic heat production), height interval (indicator for height of terrain above mean sea level), topography (coast, plain, valley) and the climate classification based on the energy balance of the earth's surface.

The factorial analysis tries to reduce the number of variables to a few factors that can be described by certain characteristics due to their correlation with the variables involved. In these factorial calculations the variables are assumed to be standardized and orthogonal to each other. So equation (3.5) shows the adequate model of factorial analysis.

$$X_i - \gamma_i \cdot U_i = a_{i1} \cdot F_1 + \ldots + a_{il} \cdot F_l + \ldots + a_{iq} \cdot F_q \quad (3.5)$$

with:

- X_i: i^{th} variable
- U_i: "single rest factor" of X_i standing for the part of the variance not being explained by the factors F_1 to F_q
- γ_i: partial regression coefficient of U_i
- a_{il}: correlation coefficient of factor F_l with variable X_i
- F_l: l^{th} factor with l = 1,...,q

The problem in applying factorial analysis exists in the way how the part of the variance that the variable X_i shares with the other variables can be determined. This question is dealt with by the so-called estimation of communalities. The communality of X_i is formed by the sum of the squares of the loadings of all factors F_1,... ,F_q. The discrete factor loading is defined as the correlation coefficient of this factor with X_i. The multiple coefficient of determination of one variable with the other variables is chosen as the most frequent method for estimating the communalities.

This criterion can be considered as meaningful since the multiple coefficient of determination stands for the part of the variance the variable has in common with the other ones. In the factorial calculations performed, the method using the multiple coefficients of determination for estimating the communalities was applied.

To achieve an evident result the factorial analysis ought to deliver a so-called "simple structure". The criterion of the "simple structure" means that a factor can be interpreted best if one part of the variables has loadings near one while the other part of the variables shows loadings near zero. So the factor F_l is to be calculated according to (3.6) obtaining a maximum variance of its loadings.

$$\sum_{l=1}^{q} s_l^2 \rightarrow MAX \qquad (3.6)$$

with:

- s_l^2: variance of the loadings of factor F_l
- q: number of variables

The relation (3.6) represents the variance maximization criterion for the orthogonal rotation of the factors abbreviated as "varimax-rotation".

With the factorial analysis using a varimax rotation two factors could be extracted. Fig. 3 shows the correlation coefficients of the independent variables from the multiple regression calculations including the climate classification. Factor 1 yields the highest correlation coefficients with the energy interval as an indicator for anthropogenic heat production, latitude and input number standing for the radiation balance.

The output number in Fig. 3 describes the energy loss of the earth's surface mainly in form of the latent and sensible heat fluxes. This variable, however, has a considerably lower correlation coefficient with factor 1 especially compared to the energy interval and latitude.

Factor 2 exhibits the highest correlation coefficients with the topographic features (coast, plain, valley) and the height interval representing the height of terrain above mean sea level. So factor 1 was called "latitudinal energy"-factor and factor 2 was named "topographic"-factor. The factors extracted by factorial analysis have the advantage of being independent from each other.

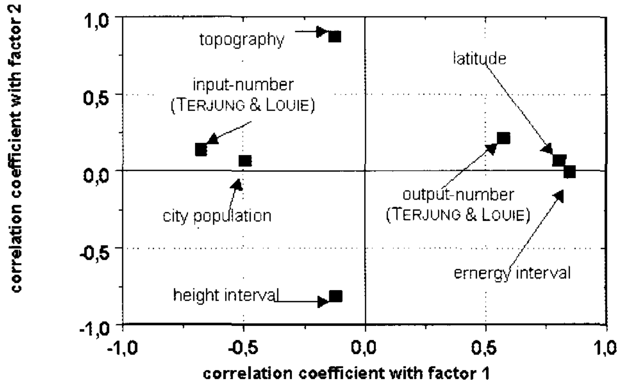

Figure 3: Correlation of independent variables including the climate classification after TERJUNG and LOUIE with factors extracted by factorial analysis. Factor 1 is called "latitudinal energy factor" since it correlates best with the variables "latitude", "energy interval" (anthropogenic heat production) and "input number" (radiation balance). Factor 2 designates the "topographic factor" according to its closest correlation with the variables "topography" (topographic features coast, plain, valley) and height interval (height above sea level). (see text for more details).

Table 6: Bivariate regression of UHI_{max} with factors calculated by factorial analysis.
r: correlation coefficient, P: probability of chance.

independent variable	r	r^2	p
factor 1 (latitudinal energy factor)	0.255	0.065	0.0016
factor 2 (topographic factor)	0.113	0.013	0.17

This means that their correlation coefficients lie close to zero. So the connection of the UHI_{max} with the two factors can be calculated with a simple bivariate regression. The results of this bivariate regression are arranged in Table 6.

Linking the two factors to the UHI_{max} with bivariate regressions gave a highly significant correlation for factor 1 having a significance level of more than 99 % and a probability of chance well below 1 % (see Table 6). The "latitudinal energy"-factor 1 describes around 6 % of the observed variance of the UHI_{max}. The "topographic"-factor 2, however, shows only an accidental correlation with the UHI_{max} with a probability of chance around 17 %.

This result hints towards a latitudinal dependence of the UHI_{max} being caused mainly by anthropogenic heat production and radiation balance with their own specific dependences on latitude. In accordance with the findings in chapter 3.1 dealing with multiple and partial regression it might also be assumed, that the magnitude of the UHI_{max} is determined to a lesser degree by latitudinal variation of the radiation balance at the earth's surface than by latitudinal differences in anthropogenic

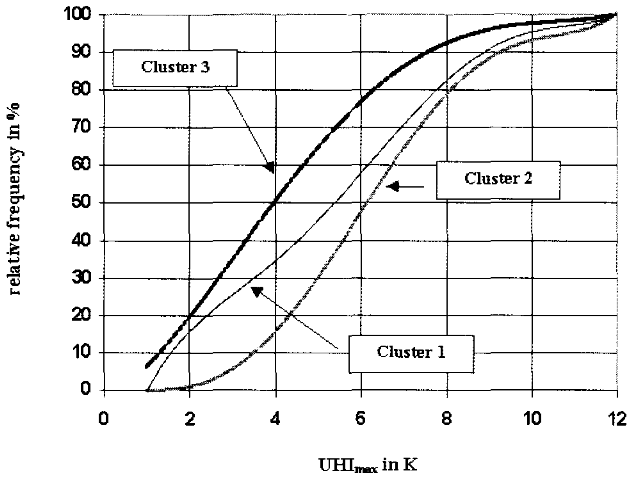

Figure 4: Cumulated frequency distribution of UHI_{max} in different clusters. Cluster 1 stands for subtropical, Cluster 2 for tropical and Cluster 3 for mid-latitudes.

heat production. So the variable "energy interval" as indicator for anthropogenic heat production shows a correlation coefficient with factor 1 of around 0.8 whereas the variable "input-number" as indicator for the radiation balance at the earth's surface yields a correlation coefficient with factor 1 of around –0.7 (see Fig. 3).

3.3 Cluster analysis

Cluster analysis was employed as a quite different statistical method to confirm a latitudinal variation of the UHI_{max} as found by multiple regression as well as by factorial analysis. The parameters being responsible for the formation of the UHI_{max} such as city population, energy interval (indicator for anthropogenic heat production), height interval (indicator for height of terrain above mean sea level), topographic features (coast, plain, valley) and climate classification after TERJUNG and LOUIE (indicator for energy balance at the earth's surface) were used for cluster calculations.

Cluster analysis basically tries to arrange cities with similar characteristics of the parameters mentioned above in one group or cluster so that the variance **within** the cluster becomes a minimum and the variance **between** the clusters becomes a maximum. Altogether three typical clusters could be extracted from the data set. Table 7 contains the main results of the cluster definition.

76 % of the cities in cluster 3 stem from latitudes between 20° S and 20° N. So this cluster was characterized as tropical. In cluster 1 76 % of the member cities belong to latitudes between 20° N or S and 40° N or S, thus being called subtropical. Cluster 2 mainly contains cities from middle latitudes. In this cluster 69 % of the member cities belong to latitudes between 40° N or S and 60° N or S. Table 7 illustrates an increase in the UHI_{max} from the tropical cluster 3 over the subtropical cluster 1 to the middle latitude cluster 2. The average UHI_{max} reaches a

Table 7: Typical cluster values for latitude and UHI_{max}. $\Delta|\varphi|$: most frequent latitude interval (northern and southern hemisphere combined), $rf(\Delta|\varphi|)$: relative frequency for $\Delta|\varphi|$ in the cluster, $(UHI_{max})_m$: mean value of UHI_{max} of all cluster members.

$\Delta\|\varphi\|$	$rf(\Delta\|\varphi\|)$	$(UHI_{max})_m$	Cluster
0°– 20°	76 %	4.0 K	3
20°–40°	76 %	5.0 K	1
40°–60°	69 %	6.1 K	2

value of 4.0 K in the tropical cluster 3 compared to 5.0 K in the subtropical cluster 1 and 6.1 K in the middle latitude cluster 2.

That also becomes evident in Fig. 4 containing the cumulated frequency distribution of the UHI_{max} per cluster. Here the tropical and subtropical clusters 3 and 1 show a majority of cities with values of the UHI_{max} **below** 6.0 K. In the subtropical cluster 1, however, there are more cities with higher values of the UHI_{max} compared to the tropical cluster 3. Cities with an UHI_{max} of more than for example 4 K make up 50 % of the cluster members in the tropical group whereas in the subtropical cluster 1 around 65 % of the member cities show values of the UHI_{max} **above** 4 K. Around 85 % of the cities in the middle latitude cluster 2, however, have values of the UHI_{max} above 4 K. Consequently cluster analysis confirms the latitudinal dependence of the UHI_{max} found by multiple regression and factorial calculations as discussed in chapters 3.1 and 3.2.

3.4 Discriminance analysis

With the help of a discriminance analysis it was evaluated which characteristic parameters used for the cluster classification contribute most to the separation of the groups. The principle of discriminance analysis is to find a function Y, the so-called discriminance function, that separates the clusters in the best possible way. The discriminance function can be written according to (3.7) as a linear combination of the k variables X_i.

$$Y = \nu_1 \cdot X_1 + \nu_2 \cdot X_2 + \ldots\ldots + \nu_k \cdot X_k \tag{3.7}$$

with:

- Y: discriminance function
- X_k: characterizing variable (k = 1, 2, 3,. . .k)
- ν_k: dicriminance coefficient of variable X_k

In the case of G Clusters at most G-1 discriminance functions can be found. The discriminance coefficients are calculated according to the discriminance criterion in (3.8).

$$\Gamma = \frac{\sum_{g=1}^{G} n_g \cdot (Ym_g - Ym)^2}{\sum_{g=1}^{G} \sum_{j=1}^{n_g} (Y_{gj} - Ym_g)^2} \rightarrow MAX \tag{3.8}$$

with:

- Γ : discriminance criterion
- G: number of clusters
- Ym_g: mean value of discriminance functions in cluster g
- Ym: mean value of discriminance functions of all N elements
- n_g: number of elements in cluster g
- Y_{gj}: value of discriminance function of element j in cluster g
- MAX: maximum value of Γ

In this study the number of elements (n) is formed by the cities which are arranged in clusters according to their characterizing variables such as size of city population, energy interval, height interval, energy balance of the earth's surface and topographic features. The discriminance criterion in (3.8), however, corresponds to the quotient made up by the variance between the clusters (denominator in (3.8)) and the variance within the clusters (numerator in (3.8)). Performing a discriminance analysis for three clusters as defined in chapter 3.3 yielded two discriminance functions.

Of these two discriminance functions only the first one is considered here since it already explains 97.5 % of the variance between the clusters. The coefficients for the first discriminance function are listed in Table 8. In terms of absolute values the input number after TERJUNG and LOUIE as an indicator of the radiation balance of the earth's surface and the energy interval as an indicator for anthropogenic heat production show the highest discriminance coefficients in Table 8. By far these variables contribute most to the separation of the clusters thus proving their importance for a latitudinal dependence of the UHI_{max}.

Table 8: Values of coefficients for the first discriminance function in case of three clusters.

characteristic variable	**discriminance coefficient**
size of city population	0.18
energy interval (indicator for anthropogenic heat production)	−1.07
height interval (height of terrain above mean sea level)	0.08
topography (coast, plain, valley)	0.02
Input number after TERJUNG AND LOUIE (radiation balance)	−2.15
Output number after TERJUNG AND LOUIE (sensible and latent heat fluxes)	0.28
UHI_{max}	−0.14

4 Conclusion

The statistical calculations performed here are based on parameters like the number of the city population, anthropogenic heat production, height above sea level, topographic features and energy balance of the earth's surface most probably controlling the magnitude of the UHI_{max}. As indicator for the energy balance of the earth's surface the climate classification after TERJUNG and LOUIE was applied. Anthropogenic heat production was parameterized by the annual per capita consumption of primary energy.

Simple bivariate regression led to a positive correlation between the UHI_{max} and latitude, e.g. UHI_{max} increases with latitude. This correlation was found to be highly significant with a probability of chance less than 1 % explaining 12 % of the observed variance of the UHI_{max}. Using all parameters mentioned above most probably having an influence on the magnitude of the UHI_{max} as independent variables in multiple-regression-calculations yielded a coefficient of determination around 0.54. So 54 % of the observed variance of the UHI_{max} is described by those variables.

Partial regression analysis calculating the influence of one variable on the UHI_{max} among all independent variables employed showed that anthropogenic heat production and radiation balance at the earth's surface mostly determine the urban heat island intensity. The partial coefficients of determination reach values up to around 6 % for the input number after TERJUNG and LOUIE as an indicator for the radiation balance of the earth's surface. Anthropogenic heat production and radiation balance exhibit a strong latitudinal dependence being statistically significant. Therefore it is assumed that a variation of the UHI_{max} with latitude is most probably caused by the appropriate latitudinal dependence of anthropogenic heat production and radiation balance.

Factorial analysis states the findings mentioned above. Two factors could be extracted. Factor 1 shows the highest correlations with latitude, anthropogenic heat production and radiation balance thus being called "latitudinal energy factor". Factor 2 correlates highest with the topographic features (coast, plain, valley) and height above mean sea level. So factor 2 was labelled as "topographic factor". Since both factors are independent from each other simple bivariate regression can be applied to calculate their influence on the UHI_{max}. Of these two factors only the latitudinal energy factor (factor 1) has a statistically significant correlation with the UHI_{max} showing a coefficient of determination around 6 %.

Cluster analysis also proved a latitudinal dependence of the UHI_{max}. Altogether three clusters were calculated. 76 % of the cities in cluster 3 and 1 stem from latitudes between 20° S and 20° N and between 20° N(S) and 40° N(S) respectively. 69 % of the cities in cluster 2 lie in latitudes between 40° N (S) and 60° N (S). So cluster 3 and 1 might be characterized as "tropical" and as "subtropical" whereas cluster 2 represents the middle latitudes.

Looking at the mean maximum heat island intensity the tropical cluster shows the lowest value with 4.0 K. The subtropical and the middle latitude cluster have a mean maximum urban heat island intensity of 5.0 K and 6.1 K respectively. So an increase of the urban air temperature surplus from tropical to middle latitudes becomes evident. According to discriminance analysis the variables that contribute most to the separation of the clusters are the input number after TERJUNG and LOUIE as an indicator for the radiation balance and anthropogenic heat production.

Resuming there are three main results from the statistical examinations:

1. The UHI_{max} tends to increase from low to high latitudes.
2. The part of the observed variance of the UHI_{max} explained by its latitudinal variation, however, remains relatively small with about 6 %.
3. A statistically significant dependence of the maximum urban heat island intensity (UHI_{max}) on latitude is based mainly on the latitudinal variation of anthropogenic heat production and radiation balance.

Thus an increasing urban heat island intensity with latitude already assumed more than 60 years ago by KRATZER (1937) could be stated now. According to the actual findings this latitudinal dependence, however, explains only a small part of about 6 % of the observed variation concerning the urban heat island intensity.

The present study has shown a slight variation of the urban heat island intensity with latitude. But the local characteristics of urban areas (f.e. building density, kind of building material, extent of artificially sealed soil surface), however, obviously seem to be the dominating factors in the formation of the urban heat island. Future research on this subject should take that into consideration if the necessary informations are available in the appropriate literature.

References

ATLAS INTERNATIONAL, 1989: Verlagsgruppe Bertelsmann GmbH/RV. Reise- und Verkehrsverlag GmbH. – Berlin, Gütersloh, München, Stuttgart.

ATWATER, M. A., 1977: Urbanization and pollutant effects on the thermal structure in four climate regions. – J. Appl. Meteor. **16**, 9, 888–895.

BAHRENBERG, G., E. GIESE, J. NIPPER, 1990a: Statistische Methoden in der Geographie, Band 1: Univariate und bivariate Statistik. – B.G. Teubner, Stuttgart.

—, —, —, 1990b: Statistische Methoden in der Geographie, Band 2: Multivariate Statistik. – B.G. Teubner, Stuttgart.

GARSTANG, P.D., G. TYSON, D. EMMITT, 1975: The structure of heat islands. – Rev. Geophys. and Space Phys. **13**, 139–155.

KRATZER, P.A., 1937: Das Stadtklima, 2nd ed. 1956. – Vieweg, Braunschweig.

KUTTLER, W., 2004a: Stadtklima, Teil 1: Grundzüge und Ursachen. Urban Climate, Part 1: Basics and Causes. – UWSF - Z. Umweltchem. Ökotox. **16** (3), 187–199.

—, 2004b: Stadtklima, Teil 2: Phänomene und Wirkungen. Urban Climate, Part 2: Phenomena and Effects. – UWSF - Z. Umweltchem. Ökotox. **16** (4), 263–274.

KUTTLER, W., D. DÜTEMEYER, A.-B. BARLAG, 1998: Influence of regional and local winds on urban ventilation in Cologne, Germany. – Meteorol. Z. N.F. **7**, 77–87.

MAYER, H., 1993: Urban Bioclimatology. – Experientia **49** (11), 957–963.

OKE, T.R., 1973: City size and the urban heat island. – Atmos. Environ. **7**, 769–779.

—, 1986: Urban climatology and the tropical city – an introduction. – In: Urban climatology and its applications with special regard to tropical areas, Proc. of the Technical Conference, WMO 652, 1–25.

—, 1995: The Heat Island of the Urban Boundary Layer: Characteristics, Causes and Effects. – In: CERMAK, J.E., A.G. DAVENPORT, E.J. PLATE, D.X. VIEGAS (Eds.): Wind Climate in Cities. Ser. E, Appl. Sciences **277**, Kluwer Academic Publishers, Dordrecht, The Netherlands, 81–107.

SCHÖNWIESE, C.D., 1985: Praktische Statistik für Meteorologen und Geowissenschaftler. – Gebrüder Borntraeger, Berlin, Stuttgart.

TERJUNG, H.W., S.S.F. LOUIE, 1972: Energy input-output climates of the world. – Arch. Met. Geoph. Biokl. Ser. B **20**, 129–166.

WEBER, S., W. KUTTLER, 2004: Cold-air ventilation and the nocturnal boundary layer structure above an urban ballast facet. – Meteorol. Z. **13**, 413–420.

WIENERT, U., 2002: Untersuchungen zur Breiten- und Klimazonenabhängigkeit der urbanen Wärmeinsel – eine statistische Analyse. – Essener Ökologische Schriften **16**, 211 pp.

WINSTAT, 1998: Das Statistik-Programm. – Kalmia Inc., Cambridge, MA, USA, Springer electronic media, Springer Verlag, Berlin, Heidelberg, New York, ISBN 3-540-14684-9.