Mausam, (1991), **42**, 2, 161-166

551.509.53 : 551.509.313

Minimum temperature forecasting by stochastic techniques : An evidence of the heat island effect

E. HERNANDEZ, R. GARCIA and M. T. TESO

Dep. de Fisica del Aire y Geofisica, Universidad Complutense De Madrid, MADRID

(*Received 7 June 1985*)

सार — यह शोध-पत्र मुख्यत: प्रसंभाव्य दृष्टिकोण से न्यूनतम तापमानों के पूर्वानुमान से संबंधित है। कुछ उपयुक्त सैद्धान्तिक विचार उन परिवर्तनों के चुनाव की ओर ले गए जो T_m के लिए अधिक प्रासंगिक है। निदर्शता दो वेधशालाओं के लिए की गई जो कि एक-दूसर के काफी नजदीक थी। इनमें से एक मैंडरिड के बीच में और दूसरी सीमा (बराजस हवाई अड्डे) पर थी। प्राप्त किए गए निदर्शों ने यह दर्शाया कि दोनों स्थानों पर मुख्य परिवर्तन तो समान है किन्तु बाहरी क्षेत्र में निष्पादन अपेक्षाकृत कुछ निम्न कोटि का है। यह दर्शाया गया है कि दोनों स्थानों की विशेषताओं को ध्यान में रखते हुए इन के बीच के अन्तर को मैंडरिड के केन्द्र में ऊष्मा द्वीप प्रभाव के लिए आबंटित किया जा सकता है।

ABSTRACT. This paper deals mainly with the forecasting of minimum temperatures (T_m) from an stochastic viewpoint. Some appropriate theoretical considerations lead to a choice of those variables significant connected to T_m. Modelling has been carried out for two nearby observatories, one in the centre of Madrid, the other one in the border (Barajas airport). The obtained models allow to show that the significant variables are the same for both locations, but performance in the peripheral area is of a rather inferior quality. It is shown, taking into account the characteristics of both places, that the difference between them can be allotted to the heat-island effect in the centre of Madrid.

Key words — ARMAX, heat island effect, persistence, partial anto-correlation, stochastic models, white noise.

1. Introduction

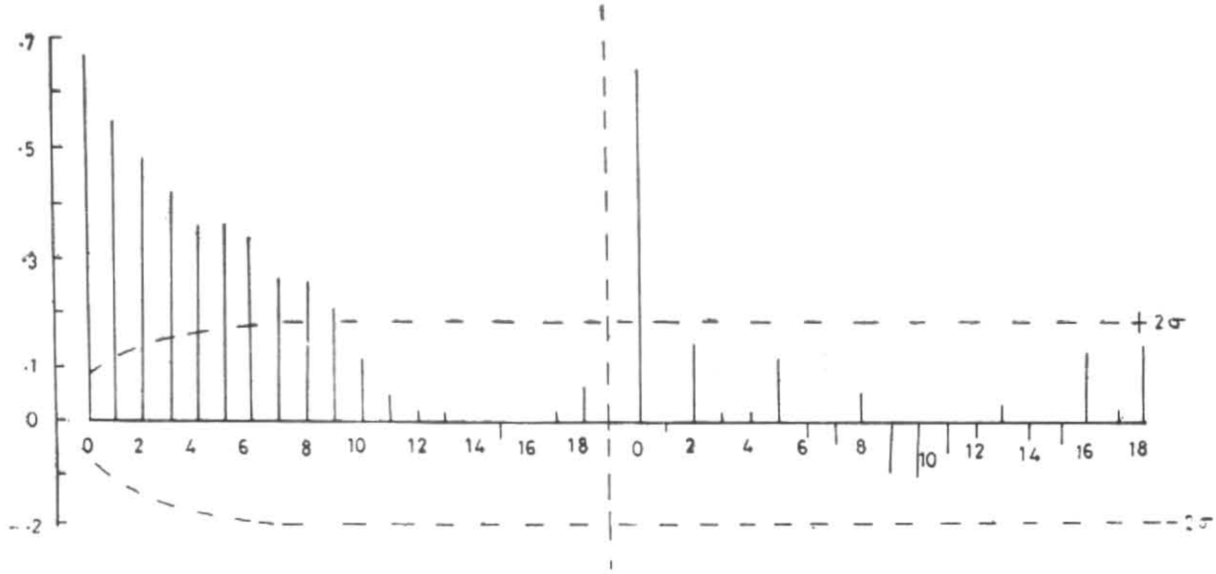
The development of theories in order to analyse time series from a stochastic viewpoint has proved an useful technique whenever one has to analyse and to predict certain atmospheric variables. Thus, a wide spectrum of works ranges from the first papers of Tiao (1975), Chock (1975) in the environmental sciences to the development of models suitable for purely meteorological variables, *e.g.*, wind (Bacci & Finzi 1980), temperature (Madden 1977), rainfall (Delleur 1978) etc. The employed models vary from the simply Auto Regressive Moving Average (ARMA) model on climatological or real time basis to those of ARMAX type including exogenous inputs (Finzi *et al.* 1984).

This paper deals with the introduction of stochastic models in the framework of prediction of minimum temperatures (T_m). Special attention will be paid to the influence of the heat island effect due to a large city (Madrid) as special problems of real time day-to-day forecasting. With this purpose in mind a comparison between two models, one for the centre of the city, the other for Barajas airport (15 km away from the said city centre), is carried out. These observatories have been chosen because orographic, climatological and conditions of any other type happen to be highly homogeneous except for the urban nature of the first station.

2. Methodology

The modelling for the a variable T_m has been done up to now, from several standpoints (Klein 1970, Hernandez 1972, Sanchez 1976, Parton 1981) in view of its interest. In many of the previous approaches, their validity and range of application are highly restricted by certain meteorological conditions, especially those related to wind and cloudiness. Thus, the building of models independent from the above limitations and therefore directly applicable, appears to be the foremost interest. Here an ARMA modelling has been selected (Box and Jenkins 1972, Katz and Skaggs 1980) with the purpose of forecasting and. above all, of characterising the behaviour of the variable only with its previous history. In this case, a complete set of recorded observations was available, so the following step was to build a more complex model allowing the inclusion of exogenous inputs, *i. e.*, an ARMAX model.

Stochastic modelling is a technique where one obtains a mathematical model of the behaviour of a time series in terms of the past history of the series (the so-called auto regressive part) and of a mixture of other stochastic influences (the moving average part) in the form of a white noise. This is suitable whenever only information about the studied variable is available. In the case where several possible inputs are considered, whose



Figs. 1(a & b). (a) Auto-correlation function for T_{mt} Retiro, & (b) Partial auto-correlation function for T_{mt} Retiro. [Abscisse give the time lags and ordinates give the auto-correlation. Dashed lines give the $\pm 2\sigma$ limits. The same holds for the rest of the figures]

influence upon the studied variable (the output variable) is suspected, the underlying philosophy remains the same, but now the white noise term becomes less important due to the presence of some term where the explicit influence of the inputs is reflected. It is quite evident that the identified models will now have three main influences : A purely temporal part, a physically based term and some stochastic component. Thus, the interesting study of these models aims not only at the forecasting performance. Rather, a most important problem will be how to decipher the weight of physical effects on the modelling. In this paper, a comparison of two different situations shows the existence of distinct physical influences for each case.

A Transfer Function Model (TFM) (Box and Jenkins 1970) has been selected. Here one has a wide variety of inputs but only one output, where this one has no influence whatsoever upon the said inputs. A multiple linear filter with an additional noise term is allowed to exist between them. In any case, the general expression for such a model is :

$$T_{mt} = \sum_{i=1}^{n} \frac{W_i(B)}{\delta_i(B)} Y_{it} + N_t$$

where,

n = Number of selected inputs,

B = The usual backward shift operator,

$W_i(B)$, $\delta_i(B)$ = Polynomials in B,

Y_{it} = Inputs,

N_t = Noise term.

Identification techniques include pre-whitening and the estimation techniques make use of hindcasting (Jenkins 1978).

The variables have been chosen from those known exactly in the moment of carrying over the forecast, concretely at 18 h (local time), a time where routine weather reports from the "Instituto Nacional de Meteorologia" can be obtained.

An AR (1) model was identified, with the expression :

$$(1 - 0.67B)\, T_{m_t} = a_t.$$

Its performance was evaluated by way of the correlation coefficient between predicted an observed values and the error percentages below 1° C and 2° C. The values obtained were

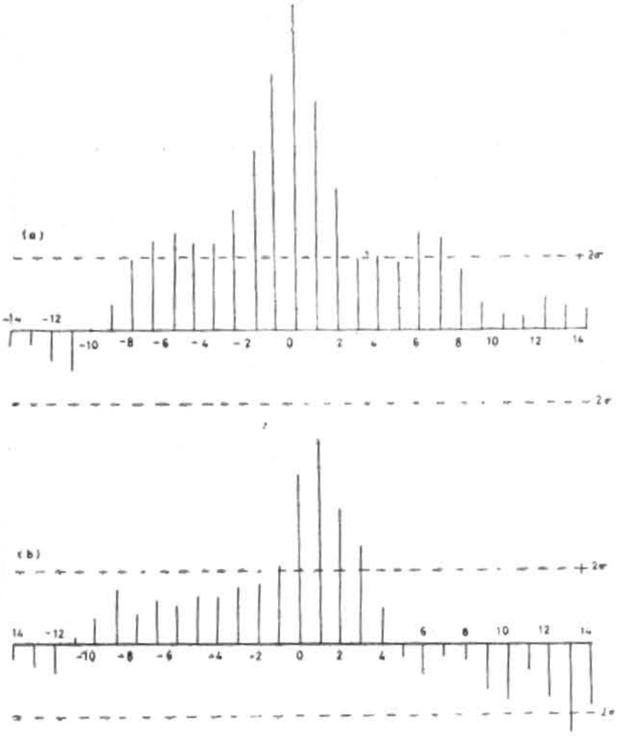
$$r = 0.67;\ (\% < 1°\text{C}) = 38\%;\ (\%\ 2°\text{ C}) = 79\%$$

According to the considerations made in the above section, a model was constructed within the framework of TFM; this model was designed to take into account the variables mentioned in that section. Appropriate models were developed for those variables, all of them recorded at 18 h (local time) of the previous day, but ϕ, corresponding to the day where forecasting was made. Univariate models for them are reflected below:

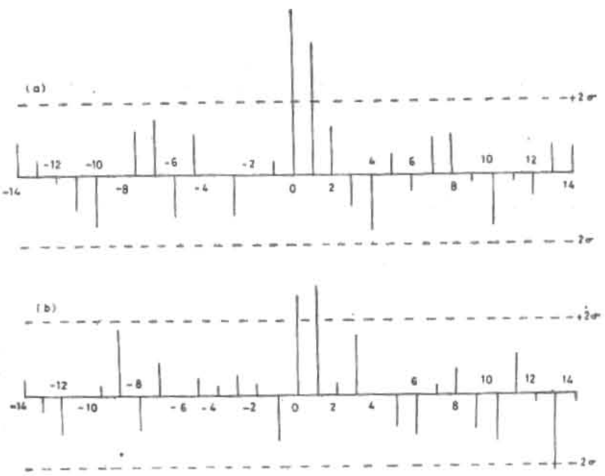
T_w	$(1-.66B)\ T_{wt} = a_t$
T	$(1-.77B)\ T_t = a_t$
ϕ	$(1-.50B)\ \phi_t = a_t$
T_d	$(1-.56B)\ T_{dt} = a_t$
V	$V_t = a_t$
C	$C_t = a_t$

where, T_w = Wet thermometer temperature,

T = Air temperature.



Figs. 2(a & b). (a) Cross-correlation function between T_m and T & (b) Cross-correlation function between T_m and T_d



Figs. 3(a & b). (a) Prewhitened cross-correlation function between T_m and T & (b) Prewhitened cross-correlation function between T_m and T_d

It must be remarked that all variables depending directly on temperature are described by AR (1) models, *i. e.*, of simple persistence, whereas cloudiness and wind (measured at 18 h) show only random properties, because they are white noise.

With the aim of discriminating those really significant variables among the set of all influencing ones, a prewhitening procedure was carried over. Simultaneously, a screening procedure was obtained as a sub-product. The main idea was to avoid the inclusion of variables whose only relationship with T_m lies on similar temporal behaviour, with no other actual physical relationship (Jenkins 1987).

3. Models for the "Retiro" (near a park in the centre of the city)

The choice of the variables has been made taking into account that they are the ones most easily available from normally equipped observatories. What is new, is that with these simple inputs very good results can be achieved through a model and this one can be implemented in any normal computing device.

The variables employed are : Air temperature (T) wet thermometer temperature T_w, dew point temperature T_d, wind speed V and cloudiness C. In addition to these, the thermic oscillation $\phi = T_{max} - T_{min}$ of the previous day is also included. It is found that T_w is always the most influencing quantity (Klein 1970).

T_m at the Retiro observatory has been initially modelled from an univariate (ARMA) viewpoint with the aim of obtaining the minimum reachable level of forecasting, thus giving way to an adequate modelling of the noise for the transfer function model (TFM). The basic period for the modelization ranges from 1 November 1979 to 31 March 1980, which roughly corresponds to the coldest months in Madrid, but values from the periods 1977-1978 and 1978-1979 have been also used in order to test all models. Since very similar results were achieved, only those of the 1979-1980 period are presented.

The auto-correlation and partial auto-correlation functions (acf, pacf) shown in Figs. 1 (a & b) were computed, after testing that no further transformations were needed.

Following this line of thought, cross-correlation functions (CCF) between the residuals were obtained by applications of the model for each input variable, including the T_m itself. Summing up, one computes the cross-correlation between a white noise (input series) and the residuals for the output series.

Once the above procedure was carried over, it was found that except for wind and cloudiness, all other variables proved significant. Non-significance of wind can be explained (in contrast with other models where it has significance) because recorded values are very small ones and their global influence, but for very rare cases, can be eliminated. This fact shows the strength of the proposed stochastic model. It is strong enough to derive T_m only on temperature grounds. This agrees with meteorological evidences

TABLE 1

Retiro observatory

	Model	r	$<1°$ (%)	$<2°$ (%)	$>2°$ (%)	FIF (%)
79–80	Stochastic	.96	84	99.2	.8	.8
(winter)	Linear	.91	75	75	15	5
77–78	Stochastic	.95	82	97	3	2.1
	Linear	.90	72	83	17	6

FIF — Frost incorrectly forecasted

TABLE 2

Barajas airport

	Model	r	$<1°$ (%)	$<2°$ (%)	$>2°$ (%)	FIP (%)
79–80	Stochastic	.95	57.1	85.5	14.5	10
	Linear	.86	50	79.1	21	12
77–78	Stochastic	.93	55	84	16	11.5
	Linear	.82	47	79	21	12.5

FIP — Frost incorrectly predicted

in Madrid, where long anticyclonic episodes with clear skies are predominant in winter months (Hernandez 1984). Thus, the effects of wind speed and cloudiness for the whole series were assumed to be negligible.

It has been shown in previous models (Sanchez 1976, Zikoev 1953) that the roles of T_d and T are interchangeable, because they explain a similar percentage of the variance and inclusion of both does not yield sensible improvements of the models. Therefore, special interest has been put in the study of CCF for T and T_d. Figs. 2 (a & b) show the CCF without pre-whitening. Figs. 3 (a & b) show them afterwards. Contrast between both is clear — for T, the prewhitening leaves only two significant or relation (lag 0 and 1), whereas for T_d there are no significant ones left — no values outside the width of $\pm 2\ \sigma$ are observable.

This has led to give up T_d as a significant input, thus implying that the relationship between T_m and moisture is given only through T_w.

Thus, the only significant variables for the behaviour of T_m are T_w, ϕ and T, which were indeed, included as inputs of the TFM model built. With the aid of hindcasting and maximum likelihood methods, an evaluation of the model was obtained. The resulting model was :

$$Tm_t = -0.1\ Tw_{t-1} + 0.17\ Tw_{t-2} - 0.66\ \phi_{t-1} + 0.77\ T_{t-1} - 0.1\ T_{t-2} + a_t - 0.23\ a_{t-1}.$$

Forecasting yielded a correlation coefficient $r = 0.96$, where the error <1 °C was 84%, and $<2°$ C, 99.2% of the cases studied. Whenever frost appeared (temperature lower than 0°C at the shelter), it was correctly predicted, and only in .8% of the cases where frost was predicted did it not happen. The Table 1 shows a comparison between the proposed stochastic models and the best performing model of linear type; both for the year studied and a test year.

The linear model does include wind speed and cloudiness, thus effecting the very high performances of the proposed model, especially the very narrow bandwidth of residuals and the very accurate forecasting of frost, for both, study and test years. It is also shown that the obtained results are of a higher quality than those furnished by any linear model.

Therefore this model gives highly satisfactory results depending only on the inputs, thus showing the great utility and usability of stochastic models for these specific purposes. It must be remarked that their applicability cannot be restricted by any meteorological conditions, *viz.*, calm days, clear ones, etc, which usually alter the behaviour of most models. Thus, an expression valid for any winter day, with results in real time better than any quality criteria, has been obtained.

4. Models for Barajas airport (15 km away from the city centre)

A procedure similar to the one developed for the Retiro has been followed. The same significative variables were obtained, but a number of remarkable differences appeared. Now the model

$$(1 - 0.47B - 0.15B^2 - 0.22B^3)\ Tm_t = a_t,$$

i. e., the persistence is of three days backwards, whereas for the Retiro the persistence was a Markovian one, including only the day before. For this case, the explained variance amounted to 33; 64%, or 12% less than the one explained by the Retiro model. This shows a persistent situation more intense, but less extended in time for the city centre. It seems that an explanation for this behaviour lies in the heat island effect. This effect exerts a sheltering activity over the centre of the city, so the temperature changes within are smoother than in peripheral areas, thus showing a more regular behaviour. Evidences of the heat island effect are quite apparent when studying the differences between temperature regime and thermodynamic soundings.

The models corresponding to the mentioned inputs are given below :

T_w	$(1-.73B)\quad T_{wt} = a_t$
T	$(1-.67B)\quad T_t = a_t$
ϕ	$(1-.51B-.15B^3)\quad \phi_t = a_t$
T_d	$(1-.63B)\quad T_{dt} = a_t$
V	$V_t = a_t$
C	$C_t = a_t$

Now T_W shows a lesser persistence as compared with the one in the Retiro model, though of the same order. The thermal oscillation shows also a third-order term not present in the Retiro model. Finally, the transfer function model (TFM) for Barajas is expressed as :

$$T_{m_t} = -0.38\,T_{W_{t-1}} - 0.69\,\phi_{t-1} + 0.79T_{t-1} + \alpha_t - 0.12\,\alpha_{t-1} - 0.20\,\alpha_{t-2} - 0.5\,\alpha_{t-3}.$$

It can be observed that the second-order terms have disappeared, whereas the noise term (this indicating the dynamic part of the filter) appears with two additional terms. Table 2 summarizes the results similar to Table 1, but for Barajas airport.

Again, it appears that the stochastic model is better than any other models, though forecasting quality is not as good as in the city centre.

For both cases, the TFM models reach their maximum effectiveness, they do not allow for new significant variables, so the reasons for discrepancies cannot lie on incomplete or incorrect modellings. An explanation can be found in the different behaviour of T_m within the heat island, where more persistence and a more regular behaviour can be found. Thus, the models employed reach maximum efficiency at the city centre; nevertheless, they give highly satisfactory results for peripheral areas outside the heat island. The prediction of T_m is, therefore, easier within the heat island effect area. This effect is nothing but a temperature smoother, and its behaviour is best described by stochastic models. A thorough research in order to explain the different predictive ability of the AR (1) Retiro model and the AR (3) Barajas model has been undertaken. Orography, climatology and other possible influences are similar for both places. Only the localisation of the Retiro in the city centre is highly significant. Previous papers (Hernandez 1977) show that the temperature difference between the city centre and rural peripheral areas amounts to some 5° C, showing a moderate heat island effect of urban nature. This effect is produced by several causes : anthropogenic heat, decreased loss of sensible heat, decrease in net long wave radiations loss. Intensity of the effect depends on wind speed, cloudiness and city size (Oke 1978). Any way, its main influence seems to be that of preventing the city from strong thermal oscillations due to radiation, a kind of thermostatic action. In this way, in a rural observatory, a lesser persistence of the series is to be expected. The model accounts for this in the amount of explained variance, which is indeed, less in this case. The persistences are, also, of different nature : Markovian for the city centre, three-term for Barajas. From a pure forecasting viewpoint, rural areas require more terms and, obviously, bigger efforts. On the other hand, the lack of thermal sheltering yields a more irregular behaviour of the rural temperature, so it seems logical to admit that more erratic behaviour is the cause of the systematic lower performances given by rural models unlike to urban area models, as shown in Tables 1 and 2.

5. Summary and conclusions

The ability of stochastic models for the forecasting of T_m has been satisfactorily proved. Development of T_m models has shown that the only significant variables influencing the behaviour of T_m are : the daily thermic oscillation ϕ, wet thermometer termperature T_w and the air temperature T at 18 h, all of them recorded the day previous to the one of forecasting. Comparison of the stochastic approach and the linear models is highly favourable to the first ones, the quality levels reached are much better : very narrow bandwidth for residues (99.9 % of the days an error $\leqslant 2$°C) in the Retiro.

To be remarked are the different temporal characteristics of the modelled variable if one compares the centre of the city (within the heat island) and peripheral areas thereof, outside the heat island. In Barajas airport one has totally acceptable results : an increase of over 16% in the explained variable when one uses the stochastic model. In interior areas, as more persistent behaviour of Markovian type is present, thus yielding better results in stochastic forecasting.

For peripheral areas (Madrid airport), the persistence is not of Markovian type, includes three terms and is much less intense. The error <2°C appeared 84.5% of the days under study. This shows the very decisive influence of the heat island effect upon the stochastic forecasting method when applied to restricted areas.

Further, it must be remarked that stochastic models are not only better predictors, they give also information about the underlying physical phenomena, which linear models do not account for.

References

Bacci, G. and Finzi, G., 1980, "A Statistical Predictor of Surface Wind Speed in Milan City", WMO RA, VI Meeting on Forecasting of conditions leading to high levels of air pollution occurrence, Leningrad, USSR, Nov. 17-20.

Box, G.E.P. and Jenkins, G.M., 1970, "Time Series Analysis Forecasting and Control", Holden Day, San Francisco.

Chock, D.P., Terrel, T.R. and Levitt, S.B., 1975, "Time series analysis of Riverside, California", Air Quality Data, *Atm. Env.*, **9**, 978-989.

Delleur, J.M. and Karras, M.L., 1978, "Stochastic Models for monthly rainfall forecasting and synthetic generation", *J. Appl. Meteor.*, **17**, 1528-1536.

Finzi, G., Garcia, R. and Hernandez, E., 1984, "The SO_2 pollution in Madrid — Part II : A comparison between two stochastic models for real-time forecasting purposes", *Nuovo Cimento* (in press).

Hernandez, E. and Garmendia ,J., 1972, "Prediction de temperaturas minimas nocturnas en noches despejadas y en calma", *Rev. Geof.*, **XXXIII**, 1, 25-31.

Hernandez, E. and Soler, A., 1977, "Relacion estadistica entre indices micro y macroclimaticos", *Rev. Geof.*, **XXXV**, 267-273.

Hernandez, E. and Garcia, R., 1984, "The SO_2 pollution in Madrid–Part I : Statistical and Meteorological Aspects", *Nuovo Cimento* (in press).

Jenkins, G.M., 1978, "Practical Experiences with modelling and forecasting series", *G.J.P. Time Series Library*, Lancaster.

Katz, R.W. and Skaggs, R.M., 1980, "On the Use of ARMA Processes to Model Meteorological Time Series", *Mon. Weath. Rev.*, **109**, 479-484.

Klein, W.M., 1970, "Computer forecasts of minimum temperature", *J. Appl. Meteor.*, **9**, 350-359.

Madden, R., 1977, "Estimates of the auto-correlations and spectra of seasonal near temperatures over North America", *Mon. Weath Rev.*, **105**, 9-18.

Oke, T.R., 1978, Review of Urban Climatology, WMO Tech. Note No. 169, p. 67.

Parton, J.W., 1981, "Model for diurnal variation in soil and air temperature", *Agr. Met.*, **23**, 205-216.

Sanchez, E., Hernandez, E. and Garmendia, J., 1976, "Uso de la oscilacion termica para predecir temperaturas minimas", *Rev. Geof.*, **XXXIV**, 3y 4, 38-45.

Tiao, G.C., Box, G.E.P. and Hamming, W.J., 1975, "A statistical analysis of the Los Angeles Ambient Carbon Monoxide Data 1955-1972", *J. Air Pollut. Contr. Ass*, **25** 1129-1136.

Zikoev, N.T., 1953, "Selective annotated bibliography on frost and frost forecasting", *Meteorol. Abs. and Bibl.*, **4**, 337-397.