5.1

METHODICAL AND PREDICTABILITY ASPECTS OF MOS SYSTEMS

Klaus Knüpffer *

Meteo Service GmbH
Berlin, Germany

Preprints of the 13th Conference on Probability & Statistics in the Atmospheric Sciences, San Francisco 19-23 February 1996

1. INTRODUCTION

Meteo Service was founded in 1994. The company is specialized in development of statistical weather forecasting systems for interested weather services. The methodical basis are advanced versions of common interpretation methods like MOS or Kalman filtering.

A review of the projects implemented so far is given in section 2. The problem of statistical overfitting is described in section 3. It is closely connected with the question of determining the optimum number of predictors in regression equations. Results of a predictability study are presented in section 4. Predictability is expressed there in terms of reduction of variance of the automatic forecast as compared to an optimum combination of persistency and climate as reference forecast. This predictability measure allows for an estimation of the usefulness of the whole automatic forecasting system consisting of the numerical model and its statistical interpretation for local weather element forecasting.

2. PROJECTS

2.1 *Overview*

The following statistical weather forecasting systems have been developed and implemented during the last years with the methods outlined below:

1992: MOS system based on ECMWF for Meteo Consult Wageningen (the Netherlands),

1993: Kalman filter based on the German European Model (EM) for the Deutscher Wetterdienst (DWD),

1995: MOS system based on the EM for the automatic production of TAF guidance for the DWD (Auto-TAF),

1995: Comprehensive MOS system based on the ECMWF model for the Swedish Meteorological and Hydrological Institute (SMHI).

Some information on these systems can be found in Knüpffer (1993 and 1995).

* *Corresponding author address:* Klaus Knüpffer, METEO SERVICE weather research GmbH, Teltower Damm 25, 14169 Berlin, Germany

Generally, Meteo Service recommends to apply MOS systems instead of Kalman filters if there is a situation of choice. Reasons for this are as follows:

(1) Kalman applications can only take into account a very limited number of predictors for stability reasons. No predictor selection is possible.

(2) A Kalman filter forgets quickly important information from the past. The higher the adaptivity the quicker it forgets. But most of the information regarding the relationships between MOS predictors and predictands is model independent and should not be forgotten by the forecasting system.

(3) There are indications that the effect of model changes on MOS equations is sometimes overestimated: Different attempts to Kalman filter results of statistical forecasting systems did not lead to any significant improvement.

2.2 *MOS Systems*

According to the experiences gained so far the design of a MOS system is first and foremost a synoptical challange: The better the preparation of the predictors and predictands the better are the results. Preparation comprises all aspects of data quality control and linear and non-linear transformations. They allow to design a system based on linear regression as non-linear as desired (Glahn, 1989). Non-linear transformations range from the application of simple analytical functions up to designing independent modules which are based on knowledge of physics which are statistically optimized. An example for this are modules for snow accumulation and snow melt developed for the SMHI project (see predictor group G in table 2). It is a non-linear combination of QPF/24h and spot time values of T2m, Abs_Err_T2m,P(ww_RR),P(ww_Solid). Empirical constants like snow melting rate depending on temperature are part of the definition of these predictors and are optimized empirically. The quality of snow forecasts has increased remarkably after the introduction of these modules.

The idea of including statistical forecasts into the predictor set is based on investigations by Saha and van den Dool (1988). They found that the skill of numerically produced forecasts can be improved in the extended medium range by using persistency of the numerical forecast for day X-1 instead of the forecast for day X. Persistencies of statistical forecasts which are valid one time step (3, 6 or 12 h) or 24 h before the valid time of the forecast are the most frequently used predictors of our MOS systems. Furthermore, statistical forecasts of selected other predictands are included into the predictor set. E.g. in the Auto-TAF project the statistical forecast of P(Cb) is used as predictor for P(Ths).

Temperature forecasts are very important in the SMHI system. Therefore 4 additional temperature predictands have been defined: DMO_T2m - T2m, T_850 - T2m, ThW_850 - T2m and Tv_1000_925 - T2m (ThW_850: wet bulb potential temperature in 850 hPa, Tv_1000_925: Averaged virtual temperature of the layer between 1000 and 925 hPa). These are differences between reasonable first guess temperature predictors and observed temperatures. The final temperature forecast is a combination of the forecasts of the 5 temperature predictands, weighted according to the RMSE statistics at the developmental sample. For most flat land stations the summer temperature forecasts are mainly based on the predictand Tv_1000_925 - T2m because they show the lowest RMSE. Temperature forecasts are supplemented by a forecast of the forecast error. This has been done by defining the absolute error of the T2m forecast as a new predictand which then is statistically forecasted. Thus, an error estimation can be given which dependends on the station, the season and the current and forecasted weather situation.

Changes of the numerical model in the developmental sample are handled using binary predictors which are equal to 1 before the date of the model change and zero thereafter. This removes the consequences of changes of (non-conditional) systematical errors of predictors. There are several other special features in the Meteo Service MOS system which are not described here. Table 1 provides a general description of both the SMHI and the Auto-TAF MOS system. Table 2 contains predictands and predictors of both projects.

	DWD: Auto-TAF	SMHI
Numerical model	EM (00 and 12 UTC run)	ECMWF (00 and 12 UTC run)
Horizontal resolution	55 km	ca. 100 km (T+3,..., 48 h) ca. 200 km (T+54,...,180 h)
Development sample	01/92 - 06/95	01/92 - 02/95
Number of stations	ca. 130 in Germany	ca. 150: 100 in Sweden 50 in other Europe
Issue times T	0,3,...18,21 UTC (Short TAF) 4,10,16,22 UTC (Long TAF)	00 UTC: 12 UTC Run 06 UTC: 00 UTC Run
Forecasting lead times 1)	T+3,6,9,12 h (Short TAF) T+8,11...23,26 h (Long TAF)	T+3,6,...,33,36,42,...,90,96, 96,108,...,168,180 h (00 UTC) T+3,6...33,36,39..60,66 h (06 UTC)
Number of predictands	ca. 50	ca. 50
Number of potential predictors	ca. 200	ca. 200
Further statistical post-processing	No	Kalman filter combination with other statistical forecasts based on ECMWF and HIRLAM

Table 1: Characteristics of the MOS Projects Auto-TAF and SMHI(ECMWF)

1) Lead times are related to issue times T. For the Auto-TAF (DWD) project, the forecasts issued between 8 and 20 UTCare based on the 00 UTC run, the others on the 12 UTC run of the EM.

Predictands Auto-TAF

T2m
Td

DD
FF
FX_Max_Gust
FX_Max_Mean
P(FX>FF+10kt)

CldCov< 200ft
CldCov< 500ft
CldCov<1000ft
CldCov<1500ft
CldCov<5000ft
CldCov Total

P(BKN< 200ft)
P(BKN< 500ft)
P(BKN<1000ft)
P(BKN<1500ft)
P(BKN< 5000ft)

P(OVC< 200ft)
P(OVC< 500ft)
P(OVC<1000ft)
P(OVC<1500ft)
P(OVC<5000ft)

P(Cb)

P(ww_RR)
P(ww_Strat)
P(ww_Conv)
P(ww_Ths)

P(ww_Liq)
P(ww_Fz)
P(ww_Sol)

RR/6h
P(RR/6h>.1mm)
P(RR/6h>5 mm)

Vis
P(Vis< 200m)
P(Vis< 500m)
P(Vis< 800m)
P(Vis<1.5km)
P(Vis<3 km)
P(Vis<5 km)

Predictors and predictor types both projects

A: Original model predictors

- Temperature, geopotential height and rel. humidity (RH) at 1000, 925, 850, 700, 500, 300 hPa
- DMO temperature, wind, clouds, precipitation, EM: long and short wave radiation

B: Derived model predictors

- Geostrophic wind (u,v,ff)
- Vorticity index derived from geopotential heights in 1000 and 500 hPa
- Gradients of ThW_850, DMO_T2m (u,v,absolute value)
- Advection of ThW_850, DMO_T2m and vorticity in 500 hPa
- Grid binary predictors derived from DMO wind and precipitation for probabilistic predictands
- Non-linearily transformed, vertically integrated RH values (maximum, average, certain products)
- Thunderstorm indices

C: Predictors of Valid time $VT\pm\Delta t$ and $VT\pm 2\Delta t$

- Δt=Time interval between neighbouring Vts
- A few predictand specific pre-selected predictors only

D: Persistency of the predictand:

- two most recent observations of the predictand,
- most recent predictand value of the valid hour of the forecast

E: Other statistical forecasts:

- Persistency of the statistical forecast,
- Statistical forecast of predictand specific pre-selected other predictands

F: Date dependent predictors:

- Harmonic functions: Sin(k*f(day)) and Cos(k*f(day) k=1,2,3; f(day)=2π*DayInYear/365
- Binary predictors: equal to 1 before the date of a model change and equal to 0 thereafter

G: Special predictors

- Auto-TAF: Conditional climatologies of colour states (Markov chain forecasts) based on 27 years of observations; used for categorical and probabilistic ceiling and visibility predictands.
- SMHI: snow accumulation and snow melting modules for snow cover forecasts.

Predictands SMHI

T2m
Abs_Err_T2m
T_Min
T_Max

DD
FF
FX_Max_Gust
P(FX>14m/s)
P(FX>24m/s)

CldCov (N)
CCLow (NL)

P(N>2)
P(N>5)
P(N=8)

Cld-Base (ClB)
P(ClB<2000m)
P(ClB<1000m)
P(ClB<300m)
P(ClB<100m)

P(W_RR_Any)
P(W_RR_Strat)
P(W_RR_Conv)
P(W_RR_Ths)

P(ww_RR)
P(ww_RR_Solid)
RR_Intensity_ww
P(Int>1mm/h)

QPF/6h
P(RR>0mm/6h)
P(RR>1mm/6h)
P(RR>5mm/6h)

QPF/12h
P(RR>0mm/12h)
P(RR>1mm/12h)
P(RR>5mm/1[illegible]

QPF/24h)
P(RR>0mm/24h)
P(RR>1mm/24h)
P(RR>5mm/24h)

Vis
P(Vis<1000m)

Table 2: Predictors and Predictands of the Auto-TAF and SMHI MOS Projects

their averaged predictability. Temperature and wind show the highest predictability followed by precipitation, clouds and visibility. No predictability is obtained for the prediction of rare events. Figure 1 shows the RV(StF,Ref) values for the groups above. Only temperature forecasts show significant skill as compared to the reference forecast in the medium range. Wind forecasts show about equally good skill in the short range but it decreases more rapidly with increasing lead time. Thus, a 6 days temperature forecast has about the same skill as a 3 days wind forecast and a 1 days precipitation forecast. Forecasts for rare events are on the other extreme: No significant skill can be observed at any lead time on the average.

Figures 2 and 3 show similar RV curves for the predictands T2m and total cloud cover. Comparisons with pure DMO forecasts are made. The resolution in lead time is increased as compared to figure 1.

Abbrev.	Predictands	Predictability
temp	T2m, Tmax, Tmin	> 8 days
wind	u,v,ff	6.7 days
RR/12h	QPF/12h, all POP/12h	5.0 days
P(RR_ww)	Hourly precipitation predictands based on ww code	5.0 days
Cld, Vis	All cloud and visibility predictands except for last row	5.2 days
Cld, Vis rare events	P(ClB<100m) and P(Vis<1000m)	no

Table 7: Groups of predictands and their predictability

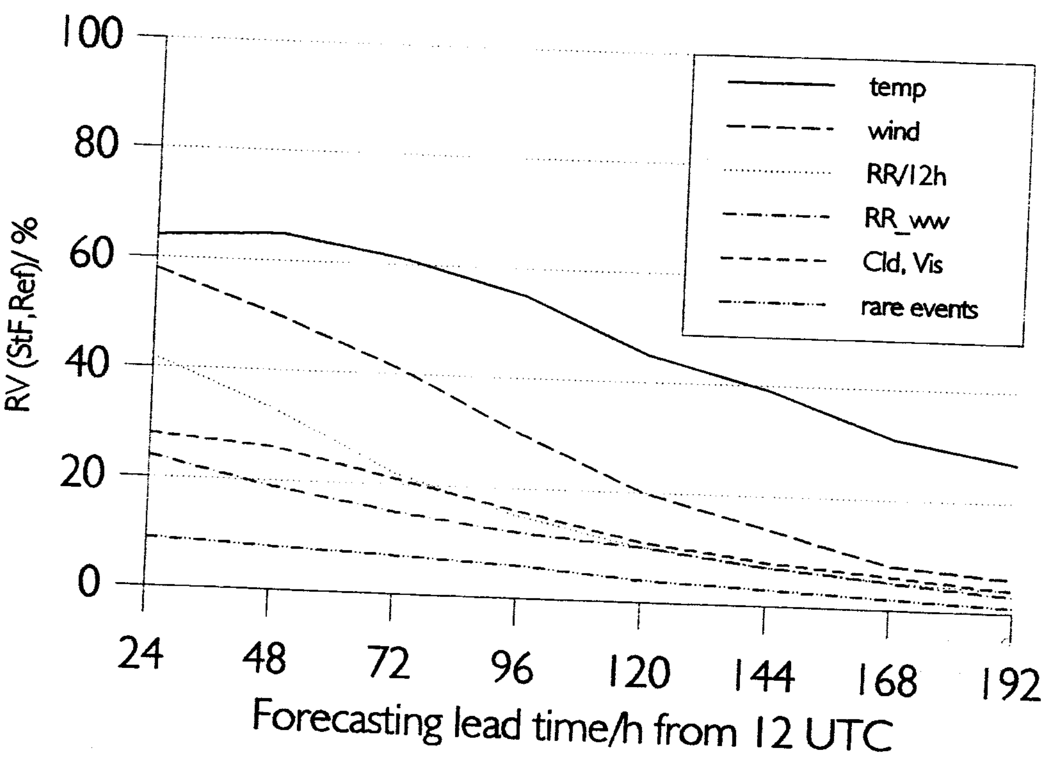

Figure 1: Predictability study: Reduction of variance of statistical forecasts as compared to an optimum combination of persistency and climatological expectance.

2.3 *Kalman Filter*

The Meteo Service Kalman filter is an extension of the standard Kalman filter algorithm as described in Simonson (1991). Investigations have shown that it seems to be impossible to find an analytical solution for determining an appropriate adaptivity of the filter coefficients a priori. Therefore a developmental sample of about 100 days is used for an empirical estimation of the optimum adaption speed of the filter coefficients. This optimization is done for each station, element and lead time separately. The elements to be forecasted by the DWD application are T2m, Tmin, Tmax, Td, dd, ff and N (total cloud cover). The structure of the Kalman filter equations is:

$$Kalman\_Fc = Const(t) + Coeff(t) * DMO\_Fc \qquad (1)$$

The Kalman filter was extended with a method which allows to apply the filter coefficients at places without observations. For this, a representativity study was performed. It resulted in an algorithm which assigns the five most representative stations to any given place in the forecast area. Representativity between two places is a composition of the following measures:

(1) geographical distance (1km),
(2) the difference between the deviation in model height and real height (1m),
(3) their difference in geographical height (3m),
(4) a factor gained from a station combination matrix.

The length units in brackets indicate equivalent influences. Measure (2) is important for the distinction between hill stations and valley stations. The matrix (4) contains factors which are applied to the result of the combination of (1) to (3) in dependence on the surface characteristics: sea points, coastal places or inland places.

For operational application at any place a weighted average of the filter coefficients of the five most representative stations to this place is applied. Systematical verifications of this approach were performed. For this purpose each station for which Kalman filter coefficients were developed was considered as a place without filter coefficients. The result of the Kalman Filter application was compared with the application of the mixed filter coefficients of the representative stations. This verification, based on 5 months and 60 inland stations, showed that there is no loss in Kalman filter forecast quality for temperature and cloud cover forecasts. The effect on wind speed was more remarkably: The reduction of variance of the Kalman filter forecasts as compared to pure DMO decreased to about half the original amount when interpolated coefficients were applied instead of the original ones.

Anyway, this approach allowed for improving DMO forecasts at any location in the forecast area.

3. REGRESSION: METHODICAL ASPECTS

3.1 *Artificial Skill*

Any predictor in a regression equation is (better: may be) a source of useful information and is (for sure) a source of statistical overfitting. Statistical overfitting of a regression equation can be expressed as (expected) artificial skill AS:

$$AS = (RMSE^{I} - RMSE^{D}) / RMSE^{D} \qquad (2)$$

$RMSE^{I}$ is the expected RMSE of the predictand at independent data. $RMSE^{D}$ is the RMSE of the predictand at the developmental sample.

Different methods to quantify AS are derived and discussed in this section. The results are used for practical design of the regression algorithm. The artificial skill depends on

- the sample size (number of cases n),
- the number p of predictors (including the regression constant) in the regression equation,
- the number pp of potential predictors to choose from.

First, the influence of p is examined in a situation without predictor selection: all potential predictors are used in the regression (p=pp). Next, the influence of pp in a situation of choosing one predictor out of a predictor set of pp potential predictors is investigated. A combination of the results allows for determining when to stop including additional predictors into a regression equation and allows for important conclusions for MOS system design.

3.2 *Monte Carlo Simulations: p=pp*

Monte Carlo simulations have been performed in order to quantify the artificial skill under the condition p=pp. Regression equations based on random numbers (Ran) in the range [0,1] have been derived for combinations of n and p. The true regression equation for $n \to \infty$ is known in this case : y =E(Ran) =0.5. The artificial skill can be decomposed into two parts:

1) AS^{D}: Artificial Skill of the developmental equation as compared to the true equation:

$$AS^{D} = (RMSE^{T} - RMSE^{D}) / RMSE^{D} \qquad (3)$$

$RMSE^{T}$: RMSE of the true regression equation.

n pp	20	50	100	200	500	1000
1	.23	.14	.10	.07	.05	.03
5	.42	.26	.18	.13	.08	.06
10	.48	.30	.21	.15	.09	.07
100	.66	.41	.29	.20	.13	.09
1000	.82	.51	.36	.25	.16	.11

Table 5: critical correlation R in dependence on the sample size n and the number of potential predictors pp according to equation (9) (Enke, 1988) with confidence level S=0.18

3.4 *Example and Conclusion*

An example for the application of the critical correlation R for determining the optimum number of predictors in a forward selection algorithm is shown in table 6. The predictors are listed in the order of their selection. From the left to the right the following values are listed: Mean value (MV), standard deviation (SD), correlation of the predictor to the predictand (r(Pd), correlation of the predictor to the residual of the equation with p-1 predictors (r(Res), predictor names and units, : expected reduction of variance at independent data of the equation with p predictors as compared to the equationwithp-1 predictors (ΔRV^I). $RMSE^D$ and $RMSE^I$: as in equation (2). Coef: 0-5: Regression coefficients of the 0-5 predictor equations. The predictor selection stopped when r(Res) < R. Calculations for expected RMSE and RV values at independent data are based on R. E.g. ΔRV^I can be derived from ΔRV^D which is part of the development statistics as follows:

$$\Delta RV^I = (\Delta RV^D - R^2) / (1 - R^2) \qquad (10)$$

The number of selected predictors depends much on the information value of the predictor set and thus on the predictability of the predictand. There are equations with more than 10 predictors e.g. for short range temperature forecasts as well as many equations with only one predictor for predictands with little predictability.

It is interesting to note that selection of a certain number of predictors leads to a higher total artificial skill of a regression equation than an a priori determination of the same number of predictors. Therefore a mixed approach is applied: First a regression equation with a few fixed predictors is calculated with $RMSE^I$ according to (7). Then the forward algorithm described above is applied.

4. PREDICTABILITY STUDY

A predictability study based on the ECMWF model has been carried out. The performance of statistical forecasts (StF) is compared to a simple reference forecast (Ref) which is an optimized combination of persistency and climatological expectance of the predictand. This allows for a qualitative estimation of the benefit of the whole process of producing automatically local weather forecasts consisting of numerical weather forecast and statistical intepretation. A second purpose is to discuss the quality of pure DMO forecasts in this context. For forecast evaluation the RV measure is used:

$$RV(StF,Ref) = 100\% * (1 - MSEI(StF) \;/\; MSE(Ref))$$
$$RV(DMO,Ref) = 100\% * (1 - MSE(DMO) / MSE(Ref)) \qquad (11)$$

Predictability is defined as the lead time when RV(StF,Ref) is 10%.The predictability study ist based on a set of 32 flat land stations with minimum distance of 100 km to the nearest sea coast or mountain area. Most of them are Swedish stations. This selection has been done in order to allow for a fair evaluation of DMO forecasts which are interpolated from grid points to stations. Results of lead times around 24, 48, ...192 hours and of summer and winter equations are averaged. 6 groups of predictands with similar predictability characteristics are considered. Table 7 contains the predictands belonging to the groups and

Station: 10381 (Berlin-Dahlem) Issue Time: 01z | Predictand: T2m / 0.1 K Valid Time: 3z | ECMWF-Run: 12z Season: Summer | n: 546 R: 0.13 | pp: 105

MV	SD	r(Pd)	r(Res)	Name	Unit	ΔRV^I	$RMSE^D$	$RMSE^I$	Coef: 0	1	2	3	4	5
1.0	0.0	0.00	-	Constant	0.1K	-	45.34	45.35	115.5	-4.50	-20.63	-17.22	-18.68	-22.93
127.8	46.8	0.97	0.97	Pers_Pd_0z	0.1K	93.79	11.26	11.30		0.94	0.95	0.93	0.80	0.78
58.7	17.1	-0.01	0.38	RH_Vert_Av_3z	%	13.20	10.40	10.53			0.25	0.24	0.19	0.15
-0.6	70.8	-0.39	-0.21	Sin_Day	0.01	3.41	10.14	10.34				-0.04	-0.01	-0.01
108.3	39.4	0.88	0.17	ThW_925_6z	0.1K	10.90	9.49	9.76					0.19	0.23
25.5	10.2	-0.01	0.17	DMO_FF_10m_3z	0.1 m/s	1.91	9.32	9.67						0.20

Table 6: Example for a regression equation. Explanations: see text above

2) AS^I: Artificial skill appearing as damage due to the application of these equations instead of using the true equation at independent data:

$$AS^I = (RMSE^I - RMSE^T) / RMSE^T \qquad (4)$$

The results of the Monte Carlo simulations for AS^D and AS^I are shown in table 3.

n \ p	4 AS^D	4 AS^I	8 AS^D	8 AS^I	16 AS^D	16 AS^I	32 AS^D	32 AS^I
32 MV	5.5	7.3	12	13	40	41	∞	>10³
SD	3.9	7.5	7	11	13	24	?	>10³
64 MV	2.5	3.5	6.2	7.0	15	16	40	44
SD	1.8	3.8	3.0	5.3	6	9	8	17
128 MV	1.1	1.5	2.7	3.2	6.7	6.9	15	15
SD	1.0	2.2	1.4	2.7	2.3	3.8	2	4
256 MV	.59	.87	1.4	1.6	3.2	3.3	6.5	6.5
SD	.47	.95	0.7	1.4	1.2	2.0	1.5	2.5
512 MV	.30	.51	.70	.89	1.6	1.6	3.2	3.2
SD	.25	.52	.34	.64	0.5	1.0	0.7	1.3
1024MV	.16	.36	.36	.44	.77	.78	1.5	1.5
SD	.15	.30	.22	.35	.27	.48	0.4	0.4

Table 3: Decomposed Artificial skills AS^D and AS^I (in %) at the dependent and independent sample: Mean values (MV) and standard deviation (SD) of these skills gained from 160 individual Monte-Carlo regression experiments per table field.

Important results are:

- AS^D and AS^I are of the same magnitude.
- There is a strong correlation between AS and p (positive) and AS and n (negative).
- High SD in the magnitude of MV indicates that the results of the individual experiments differ a lot.

An analytical description of AS in terms of p and n can be derived based on conclusions drawn by Carr (1988). Referring to research of Lorenz (1956, 1977) on non-correlated variables he found:

$$MSE^D = (1 - p / n) * MSE^T \qquad (5)$$
$$MSE^I = (1 + p / (n-p-1)) * MSE^T \qquad (6)$$

where MSE is equal to $RMSE^2$ and p+1 in Carr is replaced by p here because the constant is not considered as predictor by Carr. Equation (6) does not provide an interpretable result for n=p. Therefore and for reasons of simplicity the term n-p-1 is replaced by n-p. This has little effect on typical MOS applications where n»p. Using this slightly modified equation (6) together with (5) results in the following formula for the estimation of the RMSE at independent data based on the RMSE at the developmental sample, sample size and number of predictors:

$$RMSE^I = RMSE^D * n / (n- p) \qquad (7)$$

The quantified artificial skill is according to (2) and (7):

$$AS=p / (n - p) \qquad (8)$$

Table 4 compares the results of the Monte-Carlo simulations for AS (as can be derived from AS^D and AS^I in table 3 with respect to equations (2), (3) and (4)) with the slightly modified Lorenz approximation (8) . The results are similar. This should allow for practical application of equation (7).

n \ p	4 MC	4 Lo	8 MC	8 Lo	16 MC	16 Lo	32 MC	32 Lo
32	13	14	27	33	99	100	∞	∞
64	6.0	6.6	14	14	33	33	102	100
128	2.7	3.2	6.0	6.6	14	14	32	33
256	1.5	1.5	3.1	3.2	6.6	6.6	13	14
512	.82	.79	1.6	1.5	3.2	3.2	6.5	6.6
1024	.52	.39	.81	.79	1.6	1.5	3.0	3.2

Table 4: artificial skill values AS (in %) of the Monte Carlo simulations (MC) and according to slightly modified Lorenz approximation (Lo) , equation (8).

3.3 *Predictor Selection: pp>p*

Enke (1988) investigated the effect of selecting one predictor out of a pool of potential predictors on artificial skill. It is based on modifications of classical significance testing with respect to the situation of predictor choice in a foreward regression scheme. He derived an approximative formula for a critical correlation coefficient R in dependence on a given confidence level S for predictor selection:

$$R = \frac{\left(-\ln 2\frac{S}{pp}\right)^{0,6135}}{\sqrt{n-1}} \qquad (9)$$

R can be interpreted as threshold for predictor selection or rejection in a forward regression scheme. The application of standard values (0.01 or 0.05) for S resulted in values for R which appeared to be rather high. The question of the optimum value for S is an interesting subject of further research. Currently, a value of 0.18 is used in our regression scheme. Table 5 shows the critical correlation coefficients R which controls the number of predictors in the regression equation in dependence on n and pp.