

A Statistical Analysis of the Lake Levels at Lake Neusiedl

Johannes Ledolter

University of Iowa and Vienna University of Economics and Business
Administration

Abstract: A long record of daily data is used to study the lake levels of Lake Neusiedl, a large steppe lake at the eastern border of Austria. Daily lake level changes are modeled as functions of precipitation, temperature, and wind conditions. The occurrence and the amount of daily precipitation are modeled with logistic regressions and generalized linear models.

Zusammenfassung: Der Wasserstand des Neusiedlersees, ein grosser Step-pensee an der Grenze zwischen Österreich und Ungarn, wird mittels Tages-daten über einen sehr langen Zeitraum hinweg beschrieben. Tägliche Än-derungen des Wasserstandes werden als Funktion von Niederschlag, Lufttem-peratur, und Windverhältnissen modelliert. Das Auftreten und die Menge des Niederschlages werden mit logistischen Regressionsmodellen und verallge-meinerten linearen Modellen (GLM) erklärt.

Keywords: Circular Data, Generalized Linear Model (GLM), Logistic Re-gression, Rainfall-Runoff Relationship, Time Series Regression Model.

1 Introduction

Lake Neusiedl (“Neusiedlersee”), the second largest steppe lake in Central Europe, covers 315 km², of which 240 km² is on the Austrian side and 75 km² on the Hungarian side. The lake’s drainage basin has an area of about 1,120 km². From north to south, the lake is 36 km long, and from east to west, it is between 6 km and 12 km wide. On average, the lake’s surface is roughly 115.45 meters above the Adriatic Sea and the lake is no deeper than 1.8 meters. A man-made sluice (“Einser Kanal”) in the south-east corner of the lake provides the only (and also controllable) outflow. Because of the channel’s very shallow nature, the outflow into the river Raab (which feeds into the Danube) is largest if winds are from the northwest, while flood waters from the Raab and the Danube can back up into the lake. Lake levels fluctuate widely. Fluctuations are due to meteorological conditions such as precipitation and temperature. Due to the shallow nature of the lake, the local lake level is also affected by wind conditions.

In this paper we analyze daily data on lake levels and meteorological conditions from January 1971 through December 2004. Our objectives are to:

- study the behavior of daily average lake levels over this 34-year period,
- establish relationships between yearly averages of average lake level and meteorological conditions,
- model daily changes of average lake levels as a function of rainfall and meteorological conditions,
- compare lake levels at an individual station (Neusiedl am See) to average lake levels, and investigate the impact of wind,

- construct models for the occurrence and the magnitude of daily rainfall.

Lake level changes have a great impact on tourism and agriculture, and they affect the economic well-being of the whole area. Our paper illustrates how carefully selected graphs and relatively simple appropriate statistical methods can shed light on a problem of practical interest. In our analysis of daily lake level changes we employ time series regression models relating the daily changes to meteorological conditions while also taking account of the serial correlation in the data. In our modeling of the lake level at a specific location we explain how to best deal with the circular nature of wind direction. In our modeling of daily precipitation we develop models for both the occurrence and the magnitude of daily rainfall, and we show how these models can be estimated within the generalized linear models (GLM) framework.

2 Data

The Hydrology Office of the state of Burgenland (Hydrographischer Dienst Burgenland) provided

- daily data on average lake level (representing the overall level of the lake) and lake levels at various measurement stations including Neusiedl am See (in meters above Adriatic sea level),
- daily data on precipitation (in mm of rainfall) at eight measurement stations surrounding the lake (Pöttching, Steinbrunn, Margarethen, Donnerskirchen, Oggau, Rust, Podersdorf, Apleton); the sum of daily measurements from these eight stations is taken as a measure of overall daily precipitation,
- daily data on water temperature (in degree C), measured at the station Neusiedl am See (from 1976),
- bimonthly data on water outflow (million of meter³) through the Einser Kanal (from 1971 through 2000).

The Austrian Central Institute for Meteorology and Geodynamics provided daily data on air temperature, wind speed and wind direction for its station at Neusiedl am See. Wind speed and wind direction was measured at 7 am, 2 pm, and 7 pm. Wind speed, originally measured in Beaufort units, is converted into km/h for some of our analyses. Wind direction is coded as 0 (no wind), 4 (winds from north-east), 8 (east), 12, 16 (south), 20, 24 (west), 28, and 32 (north). Average daily air temperature is expressed in degrees C.

3 Descriptive Analysis of Daily Average Lake Levels and an Analysis of Yearly Averages

Figure 1 shows time series graphs of the daily average lake level; each panel represents a single year. Water evaporation on hot summer days tends to decrease the lake level during the summer months, while precipitation during the spring and fall/winter periods keeps

the lake level high. However, one notices considerable variability in these patterns from year to year. One also observes large differences among the lake levels between years.

Yearly averages of the lake level and related meteorological conditions are shown in Figure 2. The large reduction in the average lake level (about 50 cm) from 1996 through 2004 is particularly noteworthy. A decline in precipitation occurs during the same period, suggesting that the decline in the lake level during these years is due to the dry weather conditions. A scatter plot of yearly average lake level against yearly average precipitation is shown in Figure 3. The fitted regression line explains 18.7 percent of the variability. Scatter plots of yearly average lake level against yearly average water (and also air) temperature indicate only modest relationships, and are not shown. The only other noteworthy relationship is between yearly average lake level and yearly outflow through the artificial channel. The outflow increases with increasing lake level. This can be explained by the channel's operation which restricts the outflow if the lake level is low.

Another obvious relationship is between yearly average water and air temperatures; a one degree increase in average air temperature results in an increase of average water temperature of about the same magnitude (slope = 0.9634).

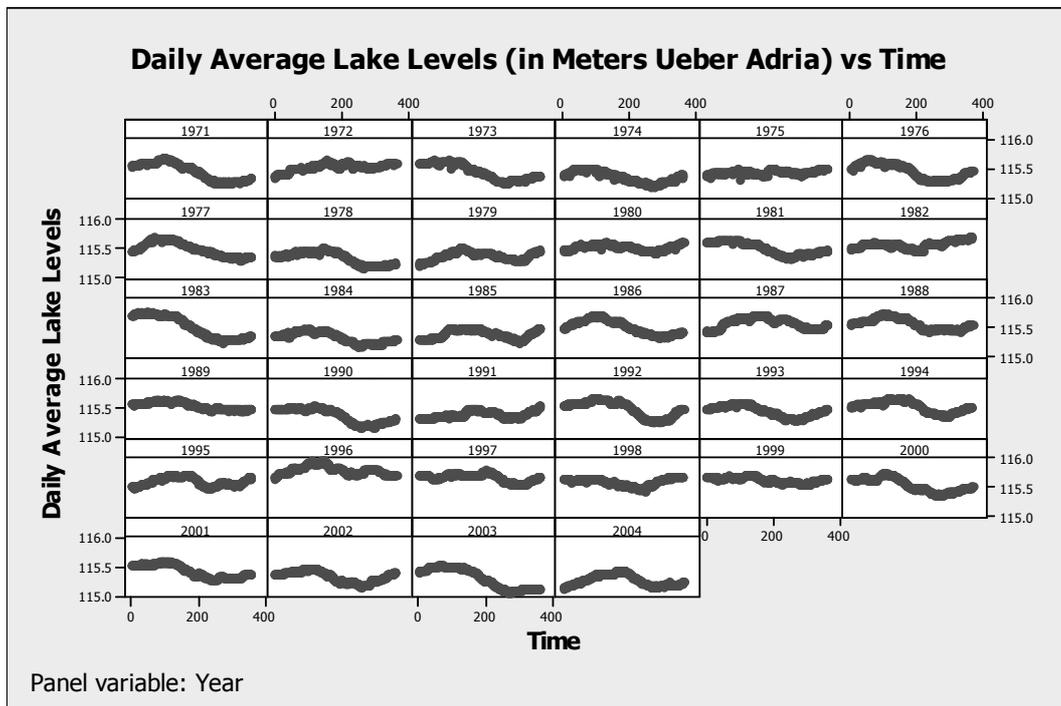


Figure 1: Time series plots of daily average lake levels.

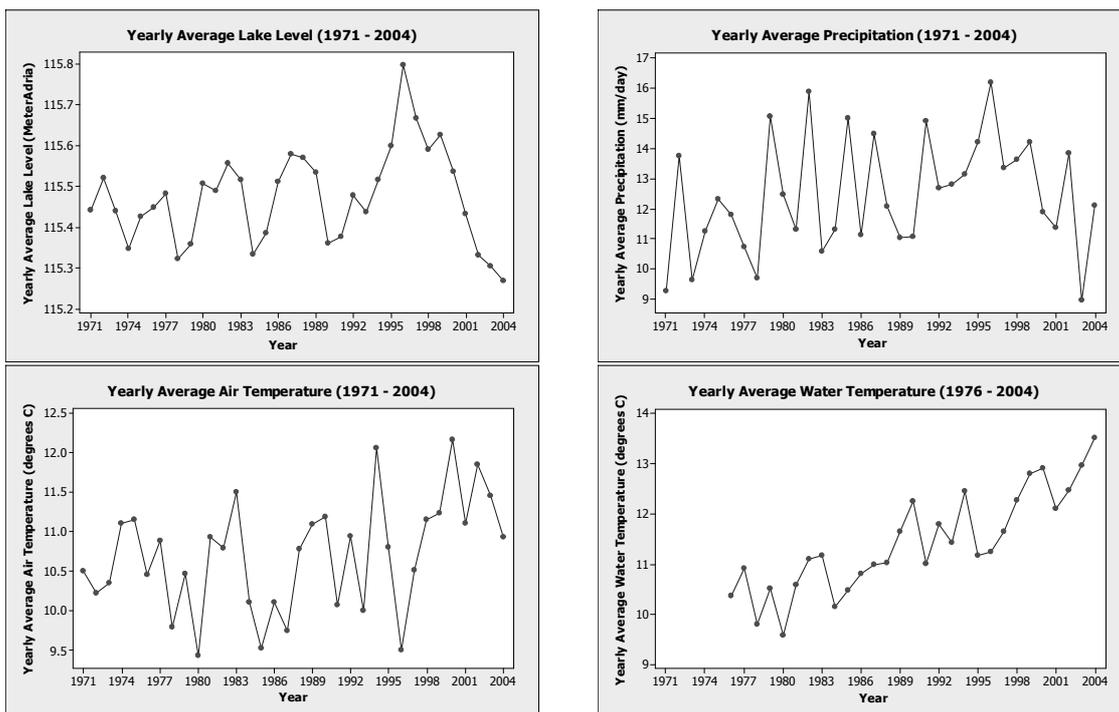


Figure 2: Time series plots of yearly averages: average lake level, precipitation, air temperature and water temperature.

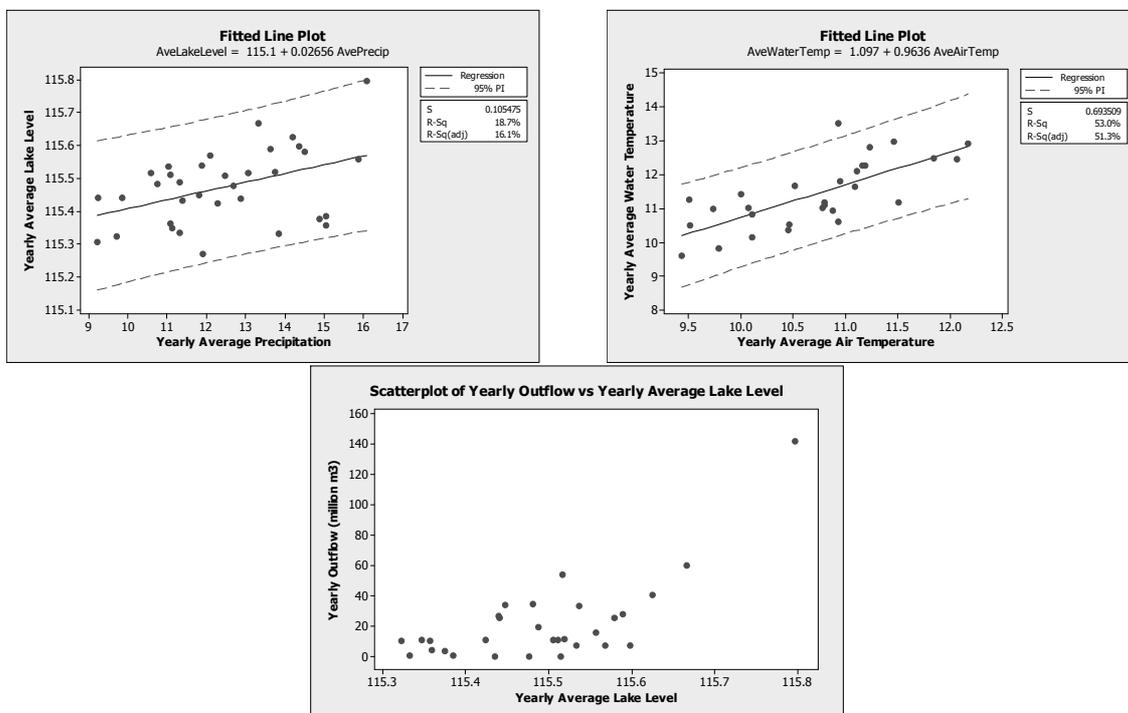


Figure 3: Scatter plots: Yearly average lake levels against yearly average precipitation. Yearly average water temperatures against yearly average air temperature. Yearly average outflow against yearly average lake level.

4 Modeling the Daily Changes of Average Lake Levels as a Function of Precipitation and Air Temperature

4.1 Preliminary Exploration

Daily changes of average lake levels expressed in cm have a mean of about zero (-0.0023), a standard deviation of roughly 1 (0.93 cm), minimum -11 cm, and maximum 14 cm. The distribution is symmetric. Since large negative and positive differences occur rarely, we combine the smallest changes from -11 cm to -4 cm into a single group and the largest changes from 4 cm to 14 cm into another. For each of the nine groups of daily lake level changes (that is, daily changes of -4 cm or less, -3 cm, ..., 3 cm, and 4 cm or more) we construct box plots of the contemporaneous precipitation during the 24-hour interval, as well as of the precipitation at previous (lag) and future (lead) days. While we expect contemporaneous and lagged effects, we expect no relationship between the changes in water level and future precipitation. Nevertheless we have included the lead effect as a check of the reasonableness of the analysis. The results in Figure 4, especially the third quartiles of the box plots, show that current and past precipitation are tied to increases in the lake level. It is especially the precipitation of the previous day (lag 1) that is most influential. It is reassuring that there is no evidence of a relationship with future rainfalls.

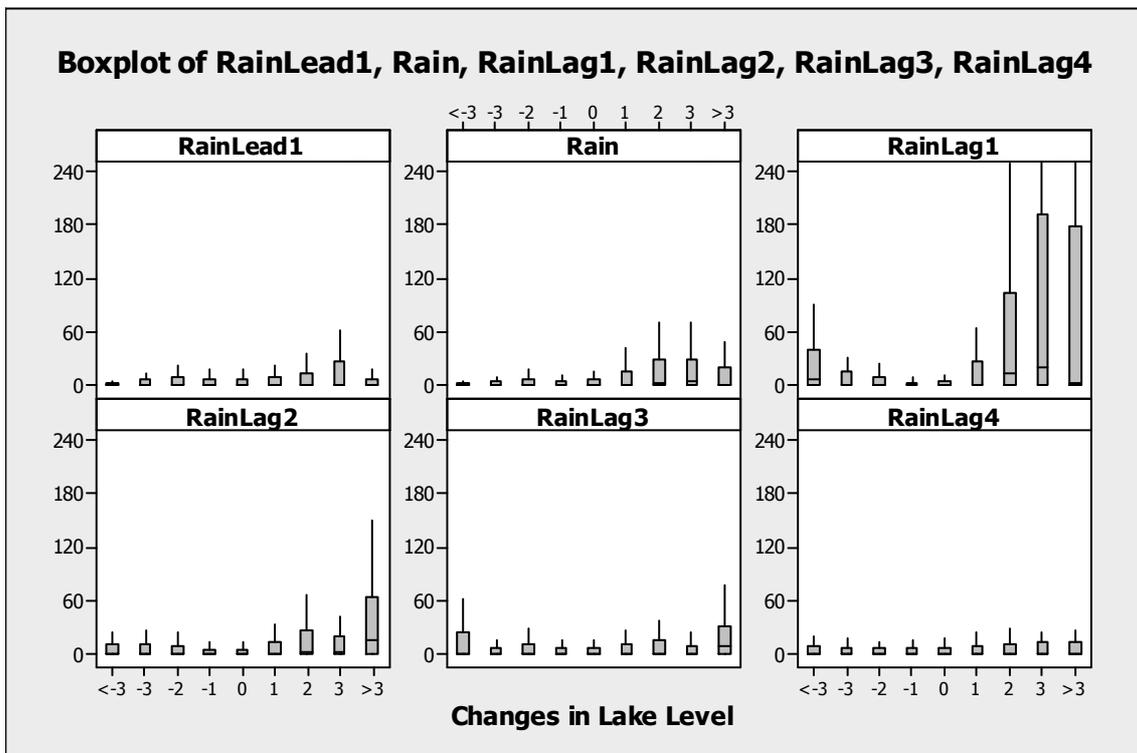


Figure 4: Box plots of precipitation and its lags/leads, for specified changes in water level.

4.2 Time Series Regression Model

Instead of using rainfall in its original metric, which has a distribution with a very long right tail (see Section 6), we consider the logarithm of precipitation, $P_t^* = \log(1 + P_t)$. The addition of “one” is necessary because of days without rain. A regression model relating the changes in lake levels to precipitation during the concurrent 24-hour period as well as prior precipitation is considered. We also add concurrent air temperature to the model as we expect evaporation to play a role; more water is evaporated on warm days. The diagnostics of the regression model with uncorrelated errors shows evidence of serial correlation. We correct for that by introducing a second order moving average model for the error components. Table 1 shows the fitting results of the time series regression model

$$Level_t - Level_{t-1} = \beta_0 + \beta_1 P_t^* + \beta_2 P_{t-1}^* + \beta_3 P_{t-2}^* + \beta_4 Temp_t + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2}.$$

The model explains 8.2 percent of the variability; the largest rain impact comes from precipitation during the previous day. The negative coefficient of air temperature expresses the effect of evaporation. While not much importance should be attached to the very large t-statistics (which is largely due to the extremely large sample size), the t-ratios express the fact that the effect of the lag 1 precipitation is more significant than that of the contemporaneous and the lag 2 precipitations.

Table 1: Maximum likelihood estimation results for the time series regression model assuming normal errors.

Variable	Coefficient	Std.Error	t-Ratio
Constant β_0	-0.0204	0.0112	-1.83
$\log(1 + P_t)$ β_1	0.0404	0.0055	7.28
$\log(1 + P_{t-1})$ β_2	0.0888	0.0061	14.57
$\log(1 + P_{t-2})$ β_3	0.0244	0.0055	4.41
$Temp_t$ β_4	-0.0143	0.0007	-21.07
MA(1) θ_1	-0.1534	0.0089	-17.29
MA(2) θ_2	-0.1511	0.0098	-17.02
R-square	8.22 %		
Durbin-Watson	1.98		

4.3 Exploring the Presence of Nonlinearity in the Relationship

Many models for the rainfall-runoff relationship have been developed in hydrology, and the range of models extends from simple linear models to very elaborate nonlinear formulations that take account of catchment descriptors of the area such as terrain, soil conditions, and vegetation properties. See for example, Jakeman and Hornberger (1993) who investigate how much complexity is warranted in a rainfall-runoff model; Sivapalan et al. (2003) who discuss the importance of developing accurate predictive models; and Rodriguez-Iturbe and Rinaldo (1997) who describe recent developments in nonlinear rainfall-runoff modeling.

The saturation of the soil from previous rains may affect the runoff of precipitation, and hence the impact of rain on the lake level. We use an exponentially weighted moving average of past rainfalls to measure the saturation of the soil. With smoothing constant $\alpha = 0.05$, the saturation is given by

$$\begin{aligned} S_t &= EWMA(P_t^*) \\ &= \alpha [P_t^* + (1 - \alpha)P_{t-1}^* + (1 - \alpha)^2 P_{t-2}^* + (1 - \alpha)^3 P_{t-3}^* + \dots] \\ &= \alpha \sum_{i \geq 0} (1 - \alpha)^i P_{t-i}^*. \end{aligned}$$

The exponentially weighted moving average S_t is the fitted value that results from an ARIMA(0, 1, 1) model with moving average parameter $\theta = 0.95$; see Abraham and Ledolter (1983).

The coplot in Figure 5 (using the statistical software R) assesses the nonlinearity graphically. There we plot changes in the lake level against the logarithmically transformed precipitation (here at lag 1, as our earlier analysis has shown that the effect is strongest at this lag), and we condition the graph on the magnitude of prior saturation, $S_{t-1} = EWMA(P_{t-1}^*)$. The scatter plots in the bottom panel, from left to right, are for days t with prior day saturation levels S_{t-1} within the three lower groupings shown on top of the coplot; the bottom left graph is the scatter plot when S_{t-1} is between 0 and about 0.85. The scatter plots in the top panel, from left to right, are for days t with prior day saturation levels S_{t-1} within the three top groupings; the upper right graph is the scatter plot when S_{t-1} is between 1.2 and about 2.3. We add to these graphs the nonparametric “lowess” smoother (see Cleveland, 1979, 1981) that provides a robust regression fit of lake level changes on prior precipitation. We observe that the relationship changes with the level of saturation. The effect of rainfall becomes stronger if the soil is saturated, suggesting the presence of some nonlinearity. As expected, the amount of nonlinearity is rather minor as we are dealing with a very large lake, and not a river. Generalized additive models for the difference in the lake levels, $Level_t - Level_{t-1}$, with mean function $\mu_t = f_0(S_{t-1}) + f_1(S_{t-1})P_t^* + f_2(S_{t-1})P_{t-1}^* + f_3(S_{t-1})P_{t-2}^* + \beta Temp_t$ and normal errors, failed to provide interpretable results. Generalized additive models have all the features of generalized linear models (including standard link functions, parent distributions, and linear model components), but include in their mean specification also sums of smooth nonparametric functions of some regressors. The functions $f_i(S_{t-1})$ are represented through penalized regression splines, and the necessary smoothing parameters are selected through cross validation; see Wood (2006) and the available R-computer software (mgcv).

5 Analysis of Daily Lake Levels at the Station Neusiedl am See: Modeling the Impact of Wind

Due to the shallow nature of the lake, the local lake level is also affected by wind conditions. We investigate the role of wind by studying the difference between the daily lake level at Neusiedl am See, a station at the northern tip of the lake, and the daily average

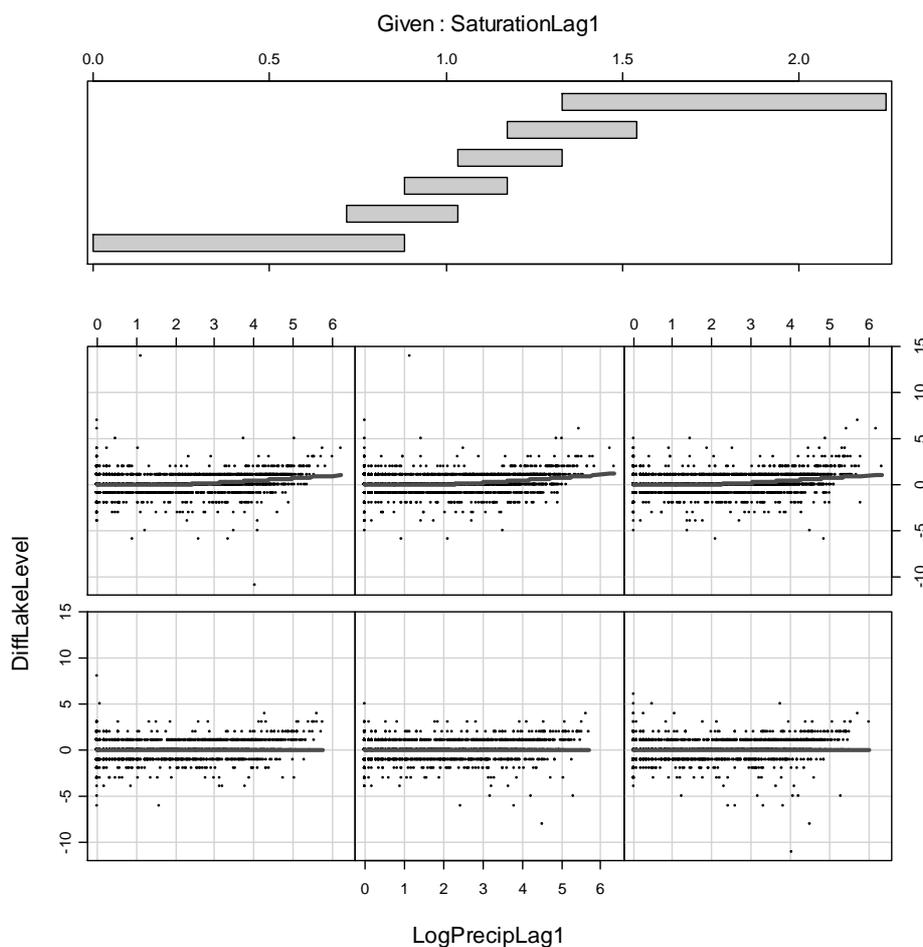


Figure 5: Coplot of successive lake level differences against lag 1 transformed precipitation, conditioned on the saturation $S_{t-1} = EWMA(P_{t-1}^*)$.

lake level. Relative to the overall lake level, the lake level at Neusiedl can be expected to decrease with strong winds from the north (as northerly winds transport water south) and increase with strong winds from the south (as southerly winds push water towards the northern edge of the lake).

A graph of the averages of these lake level differences, for non-overlapping categories of wind speed at 2 pm (measured in Beaufort, where 1 represents a wind speed between 1 and 5 km/h, 2 [6–11], 3 [12–19], 4 [20–28], 5 [29–38], 6 [39–49] and 7 [50–61 km/h]) and 2 pm wind direction (given in 45 degree angle increments), is shown in Figure 6. This graph shows quite convincingly that winds from the south increase and winds from the north decrease the lake level at Neusiedl, relative to the average lake level, and that the effects increase with wind speed.

Figure 6 relates the average of daily lake level differences to the wind conditions at 2 pm. Similar patterns arise for wind conditions at 7 am and 7 pm, and they are not shown here. Next, we average wind speed and wind direction across the three daily measurement times (7 am, 2 pm, 7 pm). Since wind direction is circular, it would be incorrect to compute the arithmetic average of the three wind directions. Instead, we construct

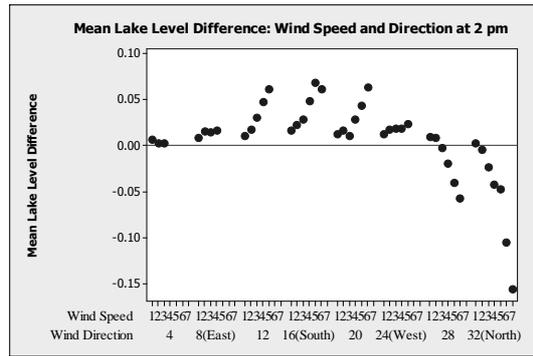


Figure 6: Plot of average daily lake level differences (Neusiedl minus overall), for non-overlapping categories of wind speed and wind direction at 2 pm.

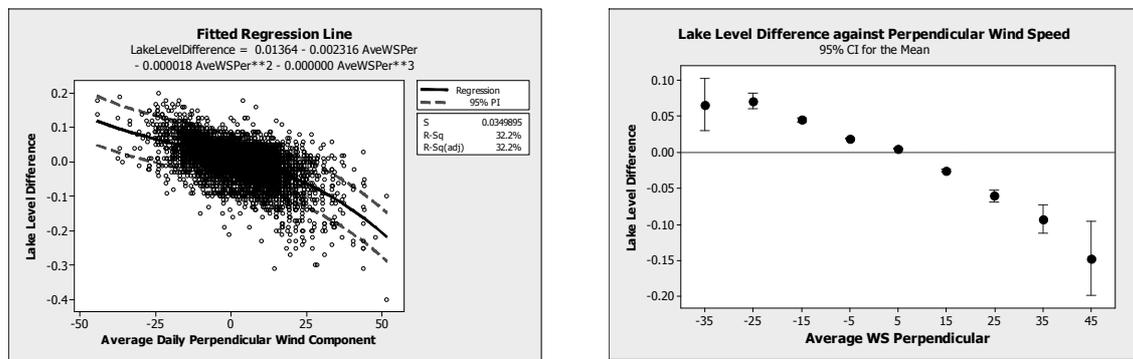


Figure 7: Plot of daily lake level differences (Neusiedl minus overall) against the average daily perpendicular (due north) wind component (raw data and smoothed version).

the perpendicular wind speed component (due north) at each of the three time periods, $WS_p = \text{wind speed} \times \cos(\text{wind direction from North})$, and average the three resulting components. In Figure 7a we plot the daily difference in the lake levels against the daily average perpendicular wind speed component. A smoothed version of this graph is shown in Figure 7b. There we divide daily average perpendicular wind speed components into non-overlapping groups and calculate, for each group, the average difference of the lake levels. The intervals shown on this graph represent 95 percent confidence for the means. Figure 7b shows an approximate linear relationship. The push-up effect of strong winds from the south is somewhat smaller (7.5 cm) than the diffusion effect of strong winds from the north (about 10 cm). Apparently it is more difficult for winds from the south to build up water at the northern edge of the lake than it is for northern winds to push the water out of the way.

6 Statistical Models for Daily Rainfall: Modeling its Occurrence and its Magnitude

Our model for the daily rainfall, as expressed as the sum of rainfall at the eight measurement stations in the vicinity of the lake, consists of two components (see Coe and Stern,

1982): (a) A model for the occurrence of daily rainfall (0 for no rain, and 1 for rain), specified as a logistic regression with covariates for monthly seasonality and the presence or absence of rain during the previous day. Trace rainfall amounts (those with magnitude of 0.5 mm or less) are coded as zero. (b) A generalized linear model for the amount of daily rainfall, for days with rain, specified as a Gamma distribution with a log link function to monthly seasonality and previous rainfall.

6.1 Modeling the Occurrence of Rain

Precipitation occurs on 42.75 percent of the days. The graph of the monthly proportions of days with rain in Figure 8 exhibits seasonal variation, with June and November showing the largest fraction of rainy days. A simple cross tabulation shows that the chance of rain after a day with no rain is 29.5 percent, whereas the probability of rain is 60.5 after a day with rain. This leads us to consider a logistic regression model that relates the logarithm of the odds (for rain) ratio to monthly indicators and the presence of rain during the previous day. Alternatively, seasonal harmonics could be used. The fitting results are shown in Table 2. Prior presence of rain increases the odds for rain more than three-fold.

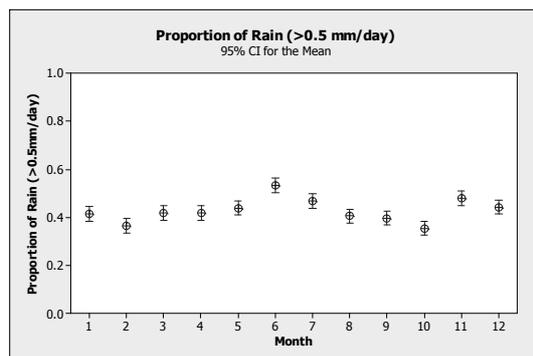


Figure 8: Proportion of days with rain.

Ignoring the seasonal components, the logistic regression model is equivalent to a first-order Markov chain model with separate transition probabilities for the two different states that describe the presence of prior rain (no rain or rain).

A Markov chain model with seasonal indicators amounts to dividing the data set $\{(\text{presence of rain on day } t - 1, \text{ presence of rain on day } t); t = 2, \dots, n\}$ into two non-overlapping segments according to the rain state on the previous day (no rain and rain), and fitting to each segment a logistic regression with monthly indicators as covariates. For example, the no rain/rain sequence “0010110”, resulting in the six pairs (0, 0), (0, 1), (1, 0), (0, 1), (1, 1), (1, 0) in the logistic regression on lag rain as covariate, would be divided into two data sets: One where previous rain is 0 [that is, (0, 0), (0, 1), (0, 1)], and the other where previous rain is 1 [that is, (1, 0), (1, 1), (1, 0)]. The approach of fitting two separate logistic models on seasonal indicators is more general than our logistic model with seasonal indicators and lag rain in Table 2 as it allows for different seasonal effects. In fact, fitting two separate logistic models provides the saturated model, and hence the deviance 20.9439, with 11 degrees of freedom and probability value 0.034, in

Table 2: Occurrence model: Logistic regression model relating the log odds of rain to monthly indicators and the presence of rain on the previous day (Minitab output): $\log\{P[Rain]/P[NoRain]\} = \beta_0 + \beta_1x_1 + \dots + \beta_px_p$.

Logistic Regression Table

Predictor	Coef	SE(Coef)	Z	P	Odds	95% CI	
					Ratio	Lower	Upper
Constant	-0.897	0.068	-13.21	0.000			
LagRainOcc	1.276	0.038	33.21	0.000	3.58	3.32	3.86
FEBRUARY	-0.180	0.096	-1.87	0.061	0.84	0.69	1.01
MARCH	0.014	0.093	0.15	0.882	1.01	0.85	1.22
APRIL	0.008	0.094	0.08	0.936	1.01	0.84	1.21
MAY	0.071	0.093	0.77	0.444	1.07	0.90	1.29
JUNE	0.359	0.093	3.85	0.000	1.43	1.19	1.72
JULY	0.158	0.092	1.71	0.087	1.17	0.98	1.40
AUGUST	-0.029	0.093	-0.32	0.750	0.97	0.81	1.16
SEPTEMBER	-0.065	0.094	-0.69	0.488	0.94	0.78	1.13
OCTOBER	-0.210	0.094	-2.23	0.026	0.81	0.67	0.98
NOVEMBER	0.201	0.093	2.16	0.031	1.22	1.02	1.47
DECEMBER	0.074	0.093	0.80	0.426	1.08	0.90	1.29

Log-Likelihood = -7842.999

Test that all slopes are zero: G = 1266.058, DF = 12, P-Value = 0.000

Goodness-of-Fit Tests

Method	Chi-Square	DF	P
Pearson	20.9646	11	0.034
Deviance	20.9439	11	0.034
Hosmer-Lemeshow	9.8752	7	0.196

Table 2 is the log-likelihood ratio statistic for testing the hypothesis that the coefficients of the seasonal indicators are the same for both segments. The differences turn out to be minor and only moderately significant (probability value 0.034).

We can extend the approach by considering higher-order models; that is adding the presence of rain two days prior to the logistic regression model or dividing the data set into four different segments depending on the state of rainfall during the two previous periods. It turns out that a first-order model is sufficient.

6.2 Modeling the Amount of Rainfall

A histogram of the amount of rainfall for the 5309 days with rainfall is given in Figure 9. The 3-parameter Gamma distribution with p.d.f.

$$f(x; \alpha, \beta, \lambda) = \frac{1}{\beta^\alpha \Gamma(\alpha)} (x - \gamma)^{\alpha-1} \exp[-(x - \gamma)/\beta]$$

is fit to the data. The fitted model in Figure 9, with parameter estimates $\hat{\alpha} = 0.509$ (shape), $\hat{\beta} = 56.17$ (scale), and $\hat{\gamma} = 0.594$ (threshold), shows excellent agreement with the data. This is confirmed by the gamma probability plot (not shown).

Figure 10 shows that there is a seasonal pattern to the amount of rainfall, with rainfall amounts highest during the months of May through October. A generalized linear model (GLM; see Nelder and Wedderburn, 1972, McCullagh and Nelder, 1983) with Gamma parent distribution and log link function that relates the logarithm of the mean to seasonal indicators and the amount of rain from the previous day is estimated. The “glm” routine of the R Statistical Software is used for the estimation. The estimation results in Table 3 confirm the seasonal pattern and a significant carry-over effect of the rainfall amount (expressed in logs) of the previous day. The estimated shape parameter of the gamma distribution, $1/\text{Dispersion} = 1/2.182 = 0.46$, is similar to our earlier estimate of the shape parameter (0.507) in the model without covariates.

We estimate the same model, but replace the amount of precipitation during the previous day with an indicator variable that signals the presence of precipitation. The results in the second part of the table indicate that presence of rainfall during the previous day increases the mean amount of rainfall by $100[\exp(0.204) - 1] = 23$ percent. The seasonal pattern is practically unchanged.

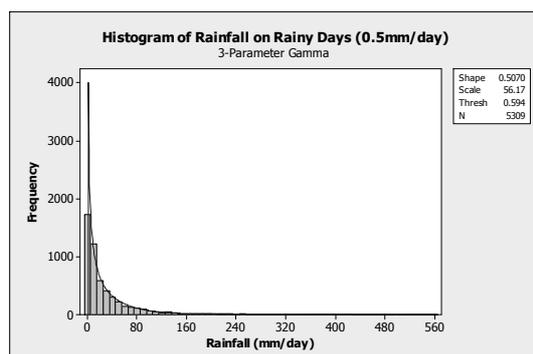


Figure 9: Histogram of the amount of rainfall, with fitted gamma distribution.

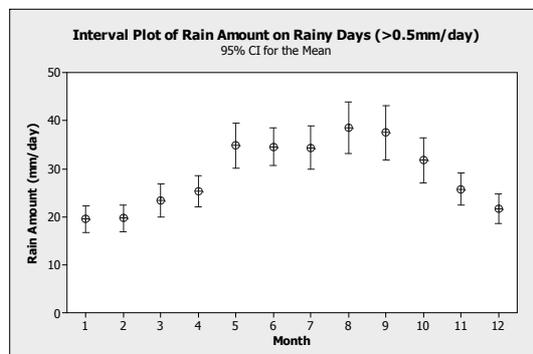


Figure 10: Average rainfall amounts for different months.

Table 3: Output of the generalized linear model with a logarithmic link function and a Gamma response distribution (R output): $Y \sim \text{Gamma}$, with the logarithm of its mean, $\log \mu = \beta_0 + \beta_1 x_1 + \dots + \beta_p x_p$, related to the explanatory variables.

(a) Using the log amount of rain during the previous day as covariate

Coefficients:

	Est	SE(est)	t-value	Pr(> t)	
(Intercept)	2.834	0.073	39.041	< 2e-16	***
{Log(LagAmount)	0.086	0.012	7.109	1.32e-12	***
FEBRUARY	0.004	0.106	0.039	0.9687	
MARCH	0.167	0.100	1.671	0.0947	
APRIL	0.231	0.101	2.291	0.0220	*
MAY	0.570	0.099	5.771	8.33e-09	***
JUNE	0.545	0.095	5.728	1.07e-08	***
JULY	0.533	0.097	5.482	4.39e-08	***
AUGUST	0.676	0.101	6.712	2.11e-11	***
SEPTEMBER	0.656	0.102	6.429	1.40e-10	***
OCTOBER	0.473	0.104	4.540	5.74e-06	***
NOVEMBER	0.255	0.097	2.613	0.0090	**
DECEMBER	0.076	0.098	0.772	0.4399	

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05

(Dispersion parameter for Gamma family taken to be 2.181875)

(b) Using an indicator for the presence of rain during the previous day as covariate

Coefficients:

	Est	SE(est)	t-value	Pr(> t)	
(Intercept)	2.846	0.074	38.514	< 2e-16	***
Ind(LagAmount)	0.204	0.041	4.975	6.74e-07	***
FEBRUARY	0.008	0.106	0.075	0.94027	
MARCH	0.182	0.099	1.825	0.06800	
APRIL	0.256	0.100	2.557	0.01058	*
MAY	0.589	0.098	5.989	2.24e-09	***
JUNE	0.566	0.095	5.977	2.41e-09	***
JULY	0.563	0.097	5.821	6.19e-09	***
AUGUST	0.700	0.100	6.981	3.29e-12	***
SEPTEMBER	0.670	0.102	6.590	4.83e-11	***
OCTOBER	0.498	0.104	4.791	1.70e-06	***
NOVEMBER	0.278	0.097	2.867	0.00415	**
DECEMBER	0.092	0.098	0.940	0.34733	

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05

(Dispersion parameter for Gamma family taken to be 2.166655)

Acknowledgments

I would like to thank Karl Maracek of the Hydrology Office of the state of Burgenland and Roland Potzmann of the Austrian Central Institute for Meteorology and Geodynamics for providing the data for this study, and my colleagues Peter Hackl and Kung-Sik Chan for many helpful discussions during the preparation of this manuscript.

References

- Abraham, B., and Ledolter, J. (1983). *Statistical Methods for Forecasting*. New York: John Wiley and Sons.
- Cleveland, W. S. (1979). Robust locally weighted regression and smoothing scatterplots. *Journal of the American Statistical Association*, 74, 829-836.
- Cleveland, W. S. (1981). LOWESS: A program for smoothing scatterplots by robust locally weighted regression. *The American Statistician*, 35, 54.
- Coe, R., and Stern, R. D. (1982). Fitting models to daily rainfall. *Journal of Applied Meteorology*, 21, 1024-1031.
- Jakeman, A. J., and Hornberger, G. M. (1993). How much complexity is warranted in a rainfall-runoff model. *Water Resources Research*, 29, 2637-2649.
- McCullagh, P., and Nelder, J. A. (1983). *Generalized Linear Models*. London: Chapman & Hall.
- Nelder, J. A., and Wedderburn, R. W. M. (1972). Generalized linear models. *Journal of the Royal Statistical Society, Series A*, 135, 379-394.
- Rodriguez-Iturbe, I., and Rinaldo, A. (1997). *Fractal River Basins: Chance and Self-Organization*. Cambridge University Press.
- Sivapalan, M., Takeuchi, K., Franks, S., Gupta, V., Karambiri, H., Lakshmi, V., et al. (2003). IAHS decade on predictions in ungauged basins (PUB) 2003-2012: Shaping an exciting future for the hydrologic sciences. *Hydrological Sciences Journal*, 48, 857-880.
- Wood, S. N. (2006). *Generalized Additive Models: An Introduction with R*. Chapman and Hall/CRC Press.

Author's Address:

Johannes Ledolter
Department of Management Sciences
S352 Pappajohn Business Bldg
The University of Iowa
Iowa City, IA 52242-1994
USA
E-mail: johannes-ledolter@uiowa.edu