An Application of the Autoregressive Conditional Poisson (ACP) model

Presented at SASA 2010 Conference

Jenny Holloway, Linda Haines*, Kerry Leask*

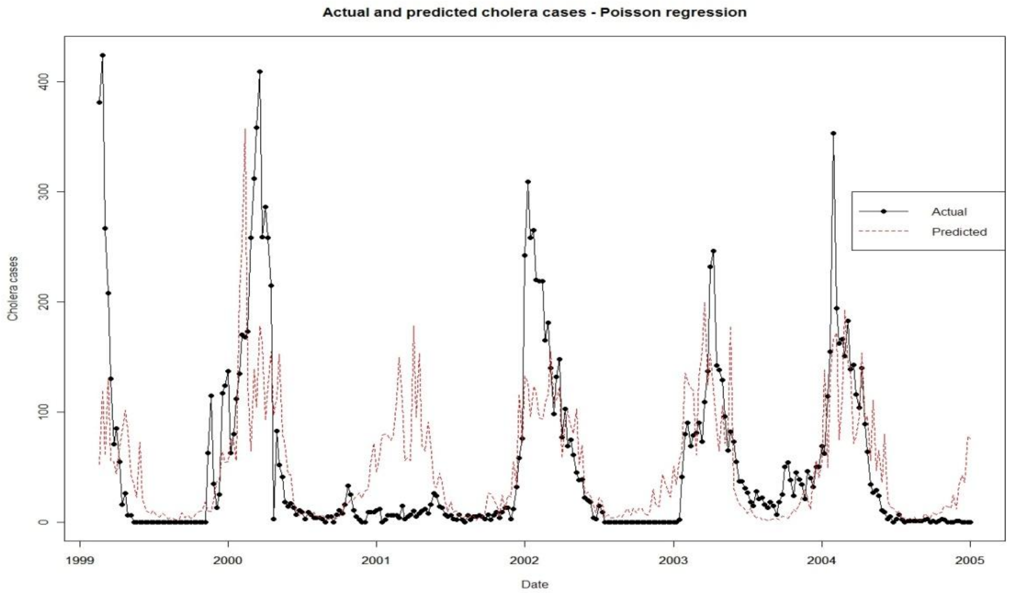
November 2010

* University of Cape Town

our future through science

Problem statement

- Time series data of counts
  - Discreteness
    - Positive counts
    - Tends to be over-dispersed
  - Time series properties
    - Typically contains serial correlation



- Cholera example
  - Static Poisson or negative binomial models with constant mean do not perform well
  - Time series models for continuous data result in negative values

- Need a model that can handle:
  - Positive counts
  - Over-dispersion
  - Serial correlation

our future through science

Brief overview

- Many count data time series models can be characterised as either observation-driven models or parameter-driven models (Cox (1981))

- Observation-driven models
  - Generic form

$$y_t \sim Poisson(\mu_t)$$

    where the equation for the mean, μ_t, includes lagged values of the observed variable, y_t

  - Easy to compute

- Parameter-driven models
  - Generic form

$$y_t \sim Poisson(\mu_t)$$

    where the equation for the mean, μ_t, contains some random variable which is independent of past observations

  - Computationally intensive

ACP (Autoregressive Conditional Poisson)

- Observation-driven model developed by Heinen (2003)

- Model handles:
  - Discreteness
  - Over-dispersion
  - Serial correlation

- Easy to estimate using maximum likelihood techniques

- ML estimation means that the usual diagnostic tests can be used.

- Can easily incorporate explanatory variables

our future through science

Description of the ACP model

Given a time series of counts, $y_1,\ldots,y_T$,where $Y_{t\,-1}$ denotes the information on the time series up to time *t* - 1, then for the **ACP(1,1)** model, the counts, conditional on past observations, are modelled as

$$y_t | Y_{t-1} \sim Poisson(\mu_t)$$

with an autoregressive conditional mean given as

$$\mu_t = \omega + \alpha y_{t-1} + \beta \mu_{t-1}$$

for $\omega > 0$ and $\alpha, \beta \geq 0$.

Note: This can be extended to include additional lags.

ACP - properties of unconditional moments

Provided $\alpha + \beta < 1$, the ACP(1,1) is stationary and has an unconditional mean and variance given by

$$E[y_t] = \mu = \frac{\omega}{1 - (\alpha + \beta)}$$

$$Var[y_t] = \frac{\mu(1 - (\alpha + \beta)^2 + \alpha^2)}{1 - (\alpha + \beta)^2}$$

So for $\alpha \neq 0$, the variance is always greater than the mean.

Hence, the ACP model is **over-dispersed**, even though the conditional distribution is equi-dispersed.

DACP (Double Autoregressive Conditional Poisson) model

- Observation-driven model developed by Heinen (2003)
- Uses ACP framework but replaces the Poisson distribution with the double Poisson distribution of Efron (1986)
- Additional to the characteristics of the ACP model, the DACP model allows the conditional variance to be larger or smaller than the mean and therefore accommodates
  - both under-dispersion and over-dispersion; and
  - more extreme cases of over-dispersion.

Description of the DACP model

The **double Poisson** density can be written as

$$f(y|\mu,\gamma) = \left(\gamma^{\frac{1}{2}}e^{-\gamma\mu}\right)\left(\frac{e^{-y}y^{y}}{y!}\right)\left(\frac{e\mu}{y}\right)^{\gamma y}$$

for $\mu > 0$ and $\gamma > 0$. Requires a multiplicative constant to make it into a true density with probabilities summing to 1.

For the **DACP(1,1)** model, the counts, conditional on past observations, are modelled as

$$y_t|Y_{t-1} \sim Double\ Poisson(\mu_t, \gamma)$$

with an autoregressive conditional mean given as

$$\mu_t = \omega + \alpha y_{t-1} + \beta\mu_{t-1}$$

for $\omega > 0$ and $\alpha, \beta \geq 0$.

our future through science

DACP - properties of moments

The conditional mean and variance for the DACP(1,1) model are

$$E[y_t|Y_{t-1}] = \mu_t$$

$$Var[y_t|Y_{t-1}] = \frac{\mu_t}{\gamma}$$

Provided $\alpha + \beta < 1$, the DACP(1,1) is stationary and has an unconditional mean and variance given by

$$E[y_t] = \mu = \frac{\omega}{1-(\alpha+\beta)}$$

$$Var[y_t] = \frac{1}{\gamma}\frac{\mu(1-(\alpha+\beta)^2+\alpha^2)}{1-(\alpha+\beta)^2}$$

So for $\gamma < 1$, the variance is always greater than the mean and the model exhibits over-dispersion.

our future through science

Cholera example

- Data of cholera outbreaks in Beira, Mozambique
  - Weekly data containing cholera counts, average air temperature, cumulated rainfall, and other variables obtained from remote sensing.
- Test the relationships between cholera outbreaks and environmental factors
  - Do climatic conditions drive the proliferation of cholera cases?

Map from: http://kids.yahoo.com/directory/Around-the-World/Countries/Mozambique/Maps

our future through science

Results: ACP & DACP vs Poisson

- Cholera cases modelled using
  - Lag 6 air temperature

Parameters	ACP	DACP	Poisson
ω	0.0991	0.1028	
α	0.0213	0.0222	
β	0.1724	0.1764	
γ		0.0825	
Intercept			-9.0965
Lag6 temp	0.1300	0.1284	0.4961

	ACP	DACP	Poisson
RMSE	32.7	32.7	61.8
MAE	15.9	15.9	37.2

All parameters shown are significant in the models

RMSE – Root mean squared error

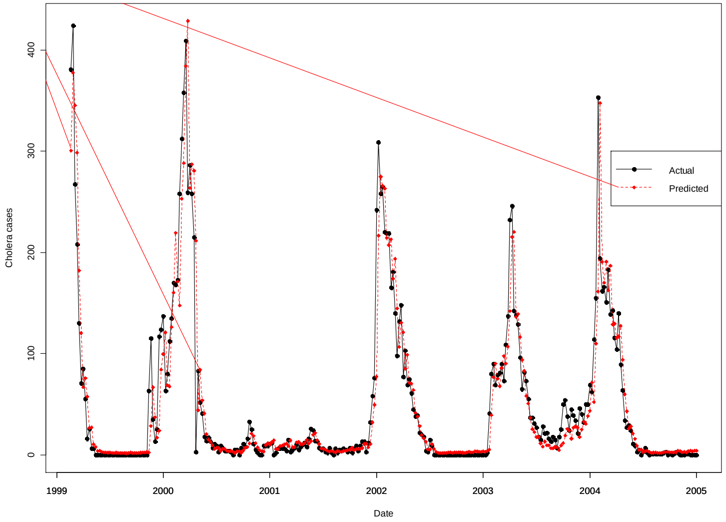
MAE – Mean absolute error

our future through science

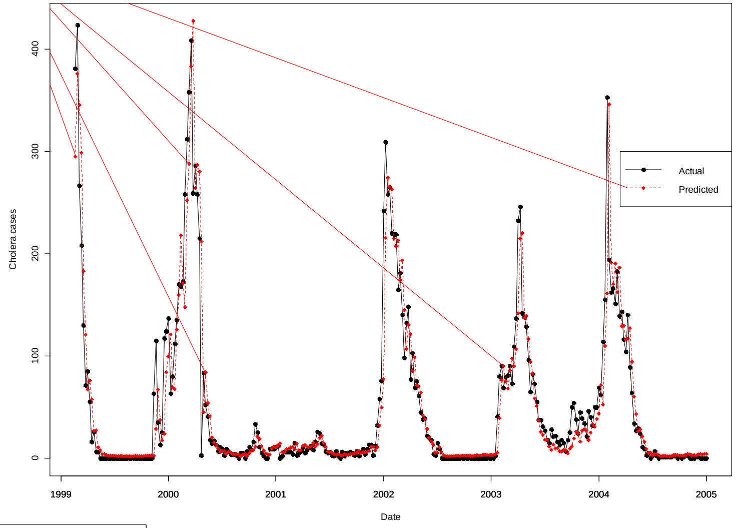
Results: ACP & DACP vs Poisson

- Plots of actual vs predicted – models using lag6 air temperature

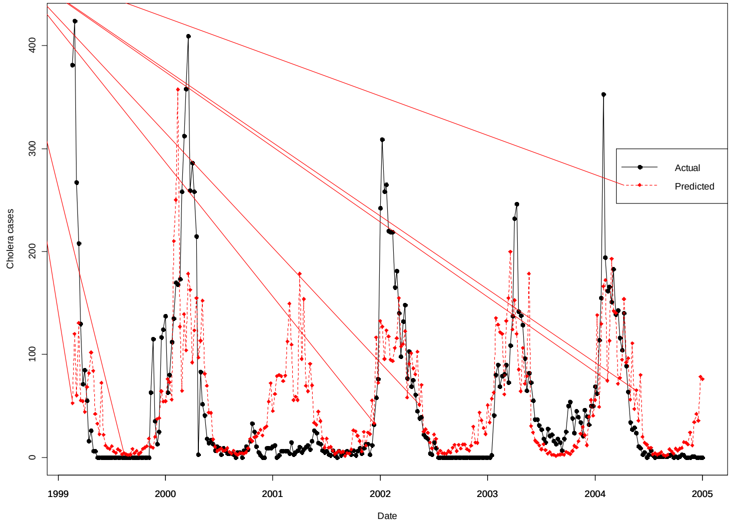
ACP model



DACP model



Poisson regression

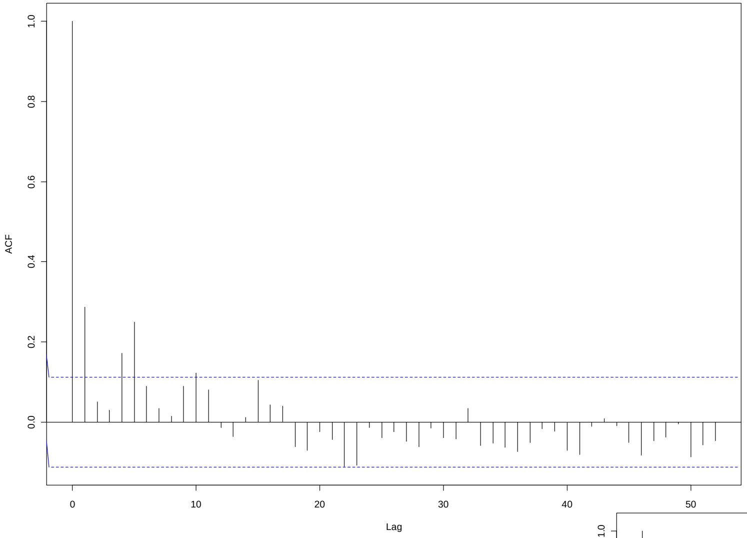


our future through science

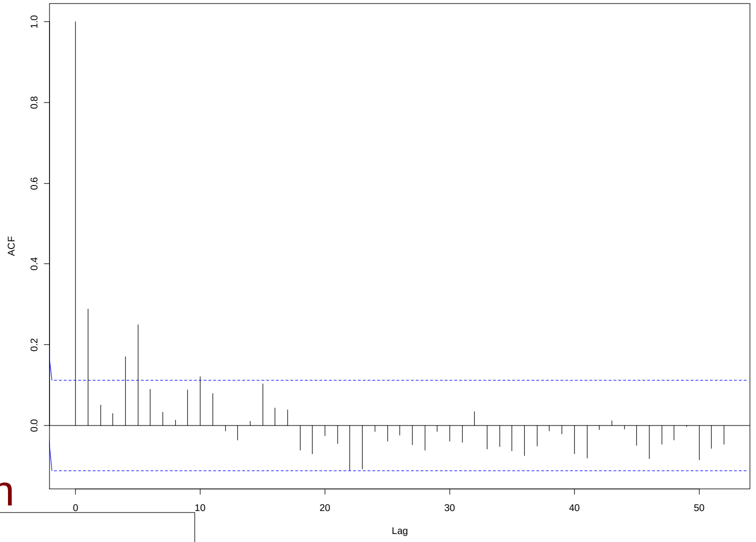
Results: ACP & DACP vs Poisson

- Autocorrelation function plots – models using lag6 air temperature

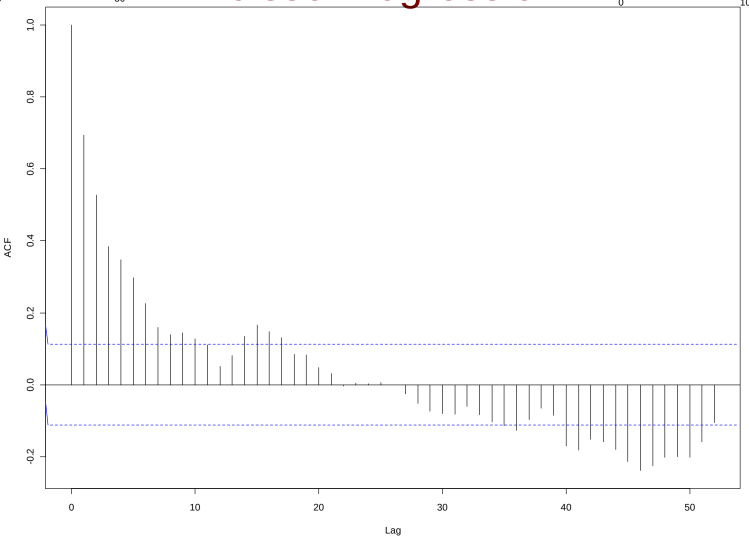
ACP model



DACP model



Poisson regression



our future through science

Results: ACP & DACP vs Poisson

- Cholera cases modelled using
  - Seasonal variables and lag 6 air temperature

Parameters	ACP	DACP	Poisson
ω	0.0438	0.0530	
α	0.0214	0.0247	
β	0.2291	0.2628	
γ		0.0951	
Intercept			-0.8624
Lag6 temp	0.1368	0.1278	0.1595
Cos(2πt/52)	0.1338	0.1902	0.5925
Sin(2πt/52)	-0.3932	-0.3217	1.2030

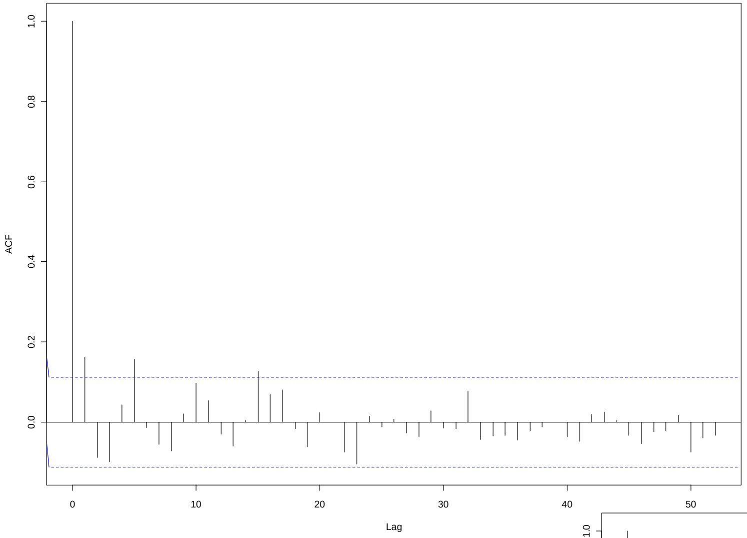
All parameters shown are significant in the models

	ACP	DACP	Poisson
RMSE	31.0	31.2	57.0
MAE	15.1	15.3	32.8

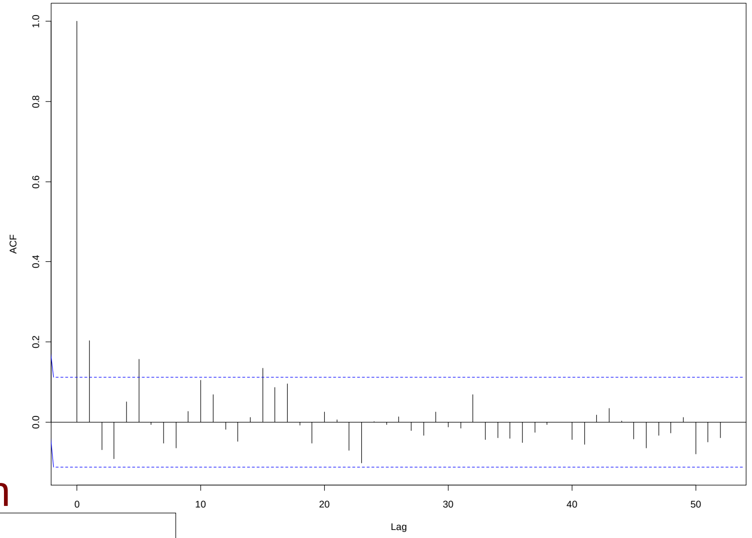
Results: ACP & DACP vs Poisson

- ACF plots – models using seasonal variables and lag6 air temperature

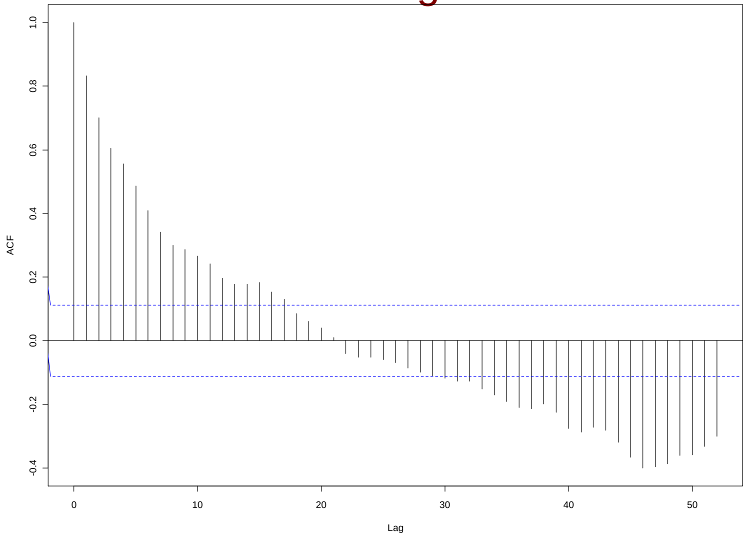
ACP model



DACP model



Poisson regression



CSIR
our future through science

Results: Poisson vs ACP vs DACP

- Likelihood ratio (LR) tests
  - LR test can be:
    - computed as twice the difference between the restricted and unrestricted log-likelihoods
    - Tested against χ^2 distribution

- LR tests - models using seasonal variables and lag6 temperature
  - LR test for autocorrelation in data using ACP model
    i.e. testing $\alpha = \beta = 0$ (equivalent to static Poisson with constant mean)
    - LR test is highly significant
    - Therefore reject static Poisson in favour of ACP model

  - LR test for over-dispersion in data using DACP model
    i.e. testing $\gamma = 1$ (equivalent to ACP model)
    - LR test is highly significant
    - Therefore reject ACP in favour of DACP

Final remarks

- Static Poisson regression not suited to data with high serial correlation
- ACP and DACP models
  - Can handle serial correlation
  - Have a similar fit
- For better estimation of standard errors and log-likelihoods
  - ACP more suited to data with small amounts of overdispersion
  - DACP model can accommodate large amounts of overdispersion
- ACP and DACP model are easy to implement and estimate

Acknowledgements

- Data supplied by government departments in Beira and Mozambique
  - Special thanks to Department of Health (Beira city and Sofala province) and CHAEM lab

Questions?

our future through science

References

- Cox, DR (1981), *Statistical analysis of time series: Some recent developments*, Scandinavian Journal of Statistics (8)
- Heinen, A. (2003), *Modelling time series count data: An autoregressive conditional Poisson model,* Discussion paper
- Efron, B. (1986), *Double exponential families and their use in generalized linear regression*, Journal of the American statistical Association 81, 709-721.