Modeling of seasonal baseline in influenza data using HMMs

Al Ozonoff*, Paola Sebastiani

Boston University School of Public Health
Department of Biostatistics

aozonoff@bu.edu

1/27/06

Motivation
- Old and new
- National P+I mortality

Approach

Results

Motivation

Old motivation

Motivation
- Old and new
- National P+I mortality

Approach

Results

- Originally motivated to improve performance of “syndromic surveillance” to detect outbreaks of disease, e.g. bioterrorist attack.
- Paradigm: Establish what is “normal”, then be vigilant for deviations from normal behavior. Some model is used for baseline; one-step-ahead prediction tells us what is expected; departure from this prediction (one-step-ahead residual) forms basis for test statistic.
- Typical approach is to model respiratory illness as sinusoid (i.e. Serfling’s method) and look for additional outbreak signal on top of baseline.
- Problem with this approach: sinusoid fits data poorly during influenza epidemic periods. Implication for prospective surveillance is decreased performance (i.e. lower power for detection of outbreaks) during epidemic periods.

Old motivation

Motivation
- Old and new
- National P+I mortality

Approach

Results

- Originally motivated to improve performance of “syndromic surveillance” to detect outbreaks of disease, e.g. bioterrorist attack.
- Paradigm: Establish what is “normal”, then be vigilant for deviations from normal behavior. Some model is used for baseline; one-step-ahead prediction tells us what is expected; departure from this prediction (one-step-ahead residual) forms basis for test statistic.
- Typical approach is to model respiratory illness as sinusoid (i.e. Serfling’s method) and look for additional outbreak signal on top of baseline.
- Problem with this approach: sinusoid fits data poorly during influenza epidemic periods. Implication for prospective surveillance is decreased performance (i.e. lower power for detection of outbreaks) during epidemic periods.

Old motivation

Motivation
- Old and new
- National P+I mortality

Approach

Results

- Originally motivated to improve performance of “syndromic surveillance” to detect outbreaks of disease, e.g. bioterrorist attack.
- Paradigm: Establish what is “normal”, then be vigilant for deviations from normal behavior. Some model is used for baseline; one-step-ahead prediction tells us what is expected; departure from this prediction (one-step-ahead residual) forms basis for test statistic.
- Typical approach is to model respiratory illness as sinusoid (i.e. Serfling’s method) and look for additional outbreak signal on top of baseline.
- Problem with this approach: sinusoid fits data poorly during influenza epidemic periods. Implication for prospective surveillance is decreased performance (i.e. lower power for detection of outbreaks) during epidemic periods.

Old motivation

Motivation
- Old and new
- National P+I mortality

Approach

Results

- Originally motivated to improve performance of “syndromic surveillance” to detect outbreaks of disease, e.g. bioterrorist attack.
- Paradigm: Establish what is “normal”, then be vigilant for deviations from normal behavior. Some model is used for baseline; one-step-ahead prediction tells us what is expected; departure from this prediction (one-step-ahead residual) forms basis for test statistic.
- Typical approach is to model respiratory illness as sinusoid (i.e. Serfling’s method) and look for additional outbreak signal on top of baseline.
- Problem with this approach: sinusoid fits data poorly during influenza epidemic periods. Implication for prospective surveillance is decreased performance (i.e. lower power for detection of outbreaks) during epidemic periods.

New motivation

Motivation
- Old and new
- National P+I mortality

Approach

Results

- Recent interest in influenza spurred by prospects of novel strain emerging to cause pandemic illness. Renewed effort to understand historical record of influenza epidemics; to model spread of disease in space and time; to prepare for possibility (eventuality?) of pandemic.
- Seasonality of influenza not completely understood. Difficult to model spatio-temporal patterns of disease. Data sources beyond traditional influenza surveillance data are increasingly becoming available.
- Improved modeling of several time series (dispersed across a geographic area) may start with model for a single time series. Better temporal models $\Rightarrow$ better spatio-temporal models.

New motivation

Motivation
- Old and new
- National P+I mortality

Approach

Results

- Recent interest in influenza spurred by prospects of novel strain emerging to cause pandemic illness. Renewed effort to understand historical record of influenza epidemics; to model spread of disease in space and time; to prepare for possibility (eventuality?) of pandemic.
- Seasonality of influenza not completely understood. Difficult to model spatio-temporal patterns of disease. Data sources beyond traditional influenza surveillance data are increasingly becoming available.
- Improved modeling of several time series (dispersed across a geographic area) may start with model for a single time series. Better temporal models $\Rightarrow$ better spatio-temporal models.

New motivation

Motivation
- Old and new
- National P+I mortality

Approach

Results

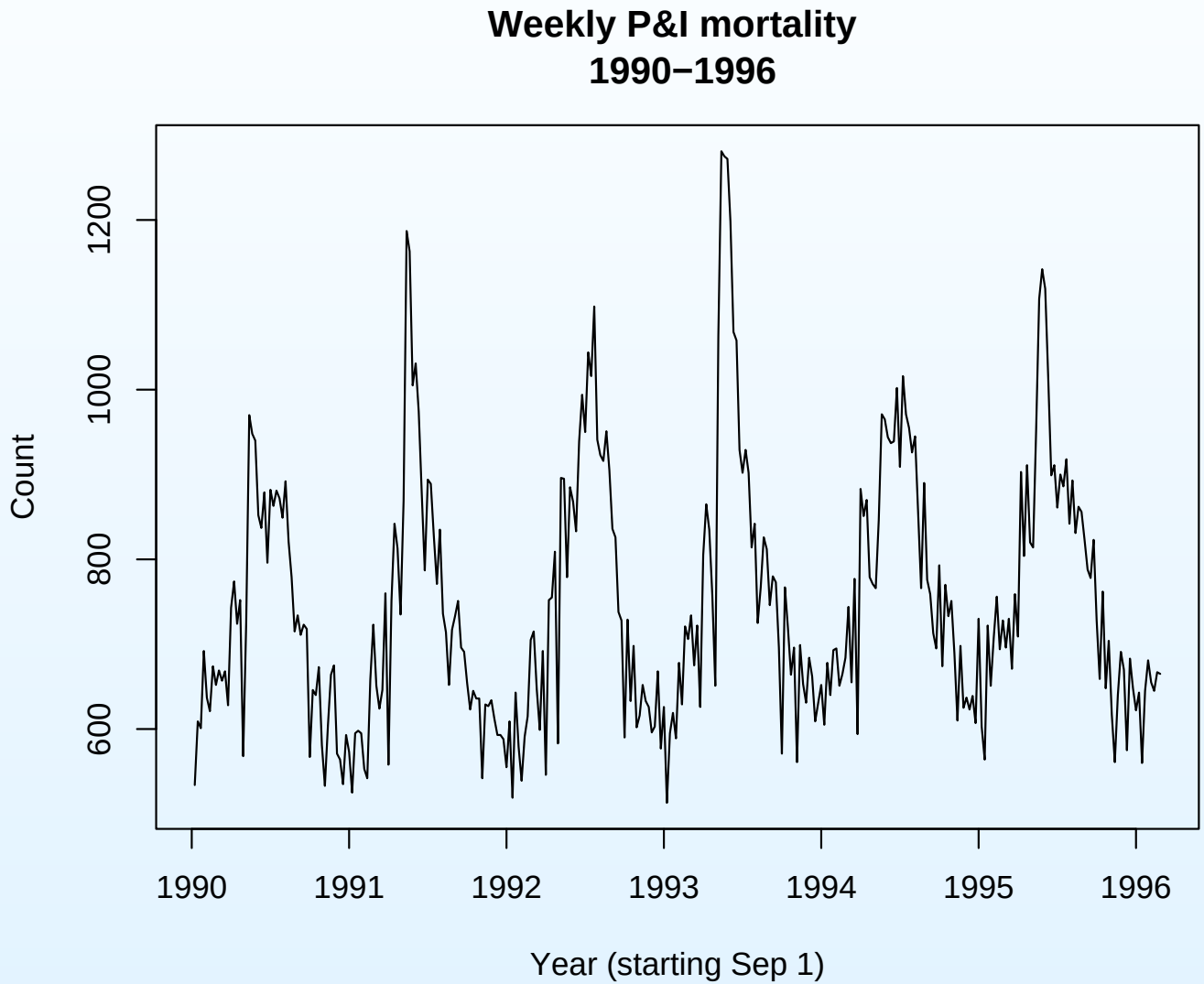
- Recent interest in influenza spurred by prospects of novel strain emerging to cause pandemic illness. Renewed effort to understand historical record of influenza epidemics; to model spread of disease in space and time; to prepare for possibility (eventuality?) of pandemic.
- Seasonality of influenza not completely understood. Difficult to model spatio-temporal patterns of disease. Data sources beyond traditional influenza surveillance data are increasingly becoming available.
- Improved modeling of several time series (dispersed across a geographic area) may start with model for a single time series. Better temporal models $\Rightarrow$ better spatio-temporal models.

National P+I mortality

Motivation
- Old and new
- National P+I mortality

Approach

Results

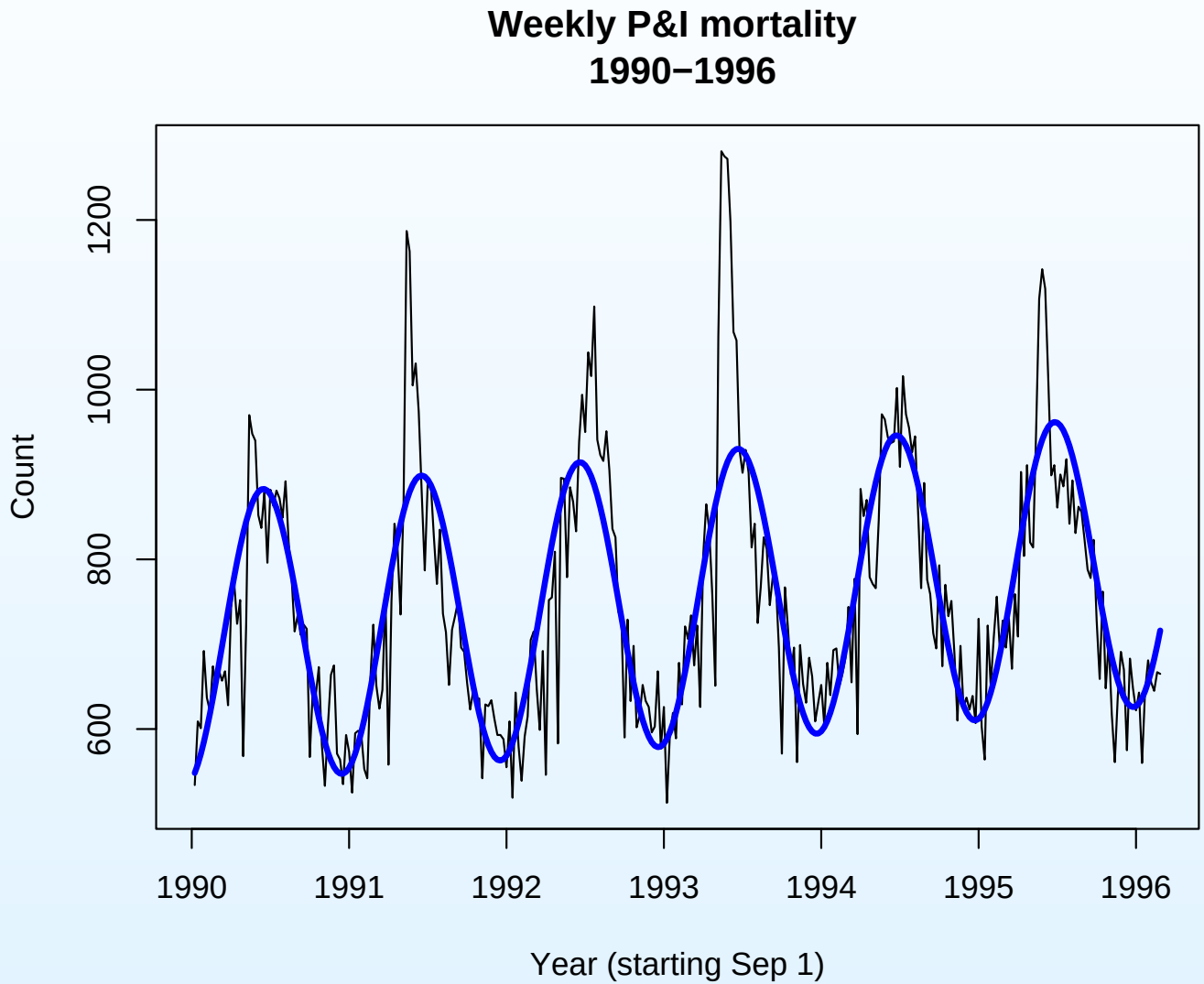


National P+I mortality

Motivation
- Old and new
- National P+I mortality

Approach

Results

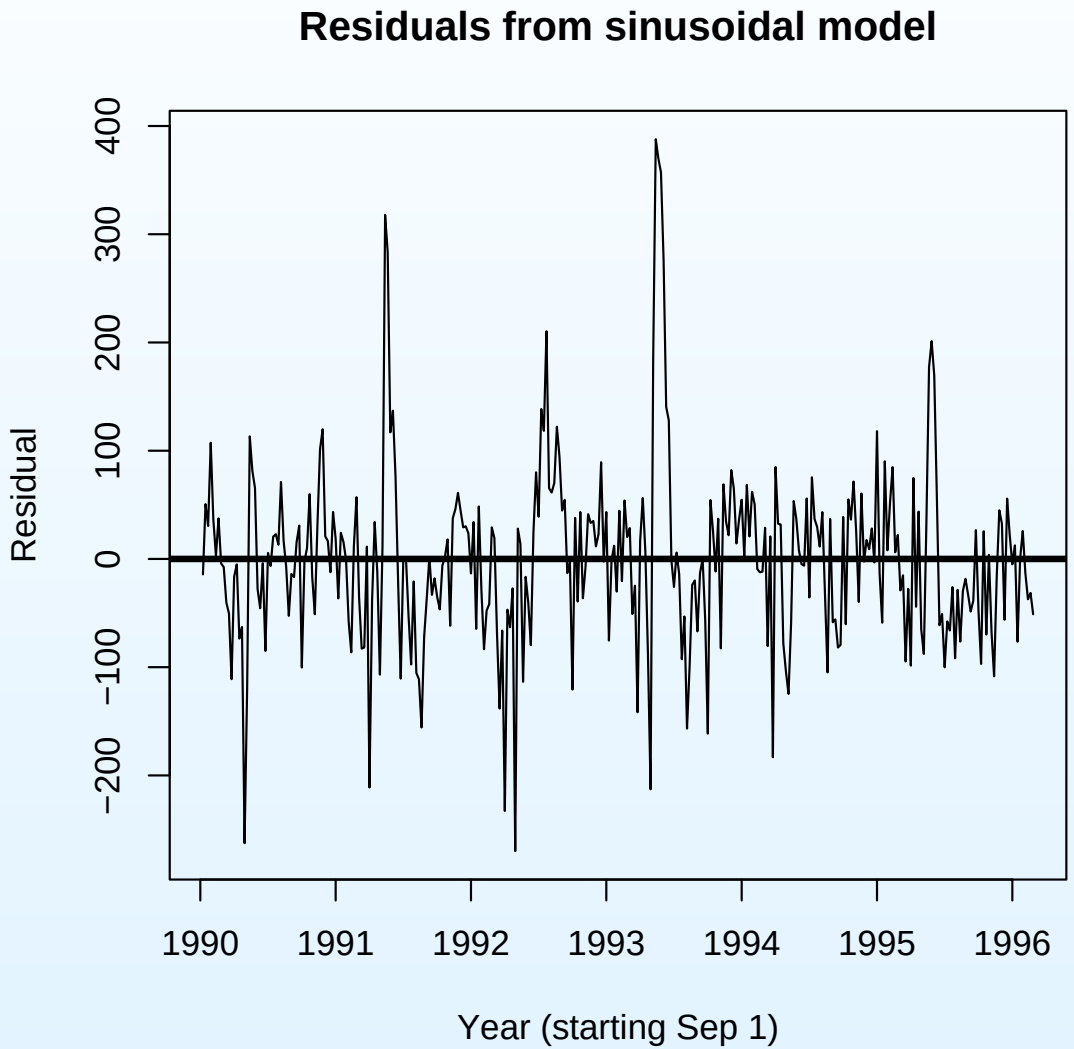


National P+I mortality

Motivation
- Old and new
- National P+I mortality

Approach

Results



Motivation

Approach

- Classical approach
- Other approaches
- HMMs
- Evaluation

Results

Approach

Classical approach

Motivation
Approach
- Classical approach
- Other approaches
- HMMs
- Evaluation

Results

- Serfling's model based upon observation that underlying seasonal baseline is roughly sinusoidal (also true for mortality of some diseases besides influenza). May be driven by temp; annual patterns (e.g. school year); dynamics of disease.

$$Y_t = \alpha_0 + \alpha_1 t + \beta_1 \sin\left(\frac{2\pi t}{52}\right) + \beta_2 \cos\left(\frac{2\pi t}{52}\right) + \epsilon_t$$

- Because Serfling's model reflects seasonal baseline, large deviations above this baseline indicate epidemic state. Integrating residuals allows calculation of "excess mortality" i.e. mortality attributed to influenza above what would be expected, accounting for seasonal variation.
- Model performs well for what it is asked to do. However, not well suited to making one-step-ahead predictions, since model fit is poor during epidemic state.

Classical approach

Motivation
Approach
- Classical approach
- Other approaches
- HMMs
- Evaluation

Results

- Serfling's model based upon observation that underlying seasonal baseline is roughly sinusoidal (also true for mortality of some diseases besides influenza). May be driven by temp; annual patterns (e.g. school year); dynamics of disease.

$$Y_t = \alpha_0 + \alpha_1 t + \beta_1 \sin(\frac{2\pi t}{52}) + \beta_2 \cos(\frac{2\pi t}{52}) + \epsilon_t$$

- Because Serfling's model reflects seasonal baseline, large deviations above this baseline indicate epidemic state. Integrating residuals allows calculation of "excess mortality" i.e. mortality attributed to influenza above what would be expected, accounting for seasonal variation.
- Model performs well for what it is asked to do. However, not well suited to making one-step-ahead predictions, since model fit is poor during epidemic state.

Classical approach

Motivation
Approach
- Classical approach
- Other approaches
- HMMs
- Evaluation

Results

- Serfling's model based upon observation that underlying seasonal baseline is roughly sinusoidal (also true for mortality of some diseases besides influenza). May be driven by temp; annual patterns (e.g. school year); dynamics of disease.

$$Y_t = \alpha_0 + \alpha_1 t + \beta_1 \sin(\frac{2\pi t}{52}) + \beta_2 \cos(\frac{2\pi t}{52}) + \epsilon_t$$

- Because Serfling's model reflects seasonal baseline, large deviations above this baseline indicate epidemic state. Integrating residuals allows calculation of "excess mortality" i.e. mortality attributed to influenza above what would be expected, accounting for seasonal variation.
- Model performs well for what it is asked to do. However, not well suited to making one-step-ahead predictions, since model fit is poor during epidemic state.

Other approaches

Motivation
Approach
- Classical approach
- Other approaches
- HMMs
- Evaluation

Results

- Periodic regression with auto-regressive component (PARMA) has been used in syndromic surveillance settings. Model fit improved during epidemic periods thanks to auto-regression. Problematic for surveillance, since AR component may in fact model the outbreaks instead of detecting them.
- “Method of analogues” is a non-parametric forecasting technique with roots in meteorology. Shown by Viboud et al. (AJE 2003) to significantly outperform other methods in one-step-ahead (and many-step-ahead) prediction. Because it is a non-parametric procedure, it ignores and obscures any knowledge about mechanism of disease.
- Nuño and Pagano developing mixed models approach using annual Gaussian to achieve better fit, as well as phase shift treated as random effect to allow for flexibility in timing of epidemic state. Bimodal Gaussian also considered to accomodate occasional dual-wave behavior.

Other approaches

Motivation
Approach
- Classical approach
- Other approaches
- HMMs
- Evaluation

Results

- Periodic regression with auto-regressive component (PARMA) has been used in syndromic surveillance settings. Model fit improved during epidemic periods thanks to auto-regression. Problematic for surveillance, since AR component may in fact model the outbreaks instead of detecting them.
- “Method of analogues” is a non-parametric forecasting technique with roots in meteorology. Shown by Viboud et al. (AJE 2003) to significantly outperform other methods in one-step-ahead (and many-step-ahead) prediction. Because it is a non-parametric procedure, it ignores and obscures any knowledge about mechanism of disease.
- Nuño and Pagano developing mixed models approach using annual Gaussian to achieve better fit, as well as phase shift treated as random effect to allow for flexibility in timing of epidemic state. Bimodal Gaussian also considered to accomodate occasional dual-wave behavior.

Other approaches

Motivation
Approach
- Classical approach
- Other approaches
- HMMs
- Evaluation

Results

- Periodic regression with auto-regressive component (PARMA) has been used in syndromic surveillance settings. Model fit improved during epidemic periods thanks to auto-regression. Problematic for surveillance, since AR component may in fact model the outbreaks instead of detecting them.
- “Method of analogues” is a non-parametric forecasting technique with roots in meteorology. Shown by Viboud et al. (AJE 2003) to significantly outperform other methods in one-step-ahead (and many-step-ahead) prediction. Because it is a non-parametric procedure, it ignores and obscures any knowledge about mechanism of disease.
- Nuño and Pagano developing mixed models approach using annual Gaussian to achieve better fit, as well as phase shift treated as random effect to allow for flexibility in timing of epidemic state. Bimodal Gaussian also considered to accomodate occasional dual-wave behavior.

Hidden Markov Models (HMMs)

Motivation
Approach
- Classical approach
- Other approaches
- HMMs
- Evaluation

Results

- Idea behind HMMs: there is a ‘hidden” (latent, unobserved) discrete random variable, representing some part of the disease process. Observed variables are modeled, conditional upon the hidden state. Thus, if we know the state we also know the distribution of observed random variable.
- Markov property: conditional probability of state change (transition probability) depends only on the value of latent state at previous time point. Thus specify the Markov model for k states with a $k \times k$ matrix of transition probabilities, and the distributions of the observed data conditional on the hidden state.
- Parameter estimation accomplished with Bayesian inference Using Gibbs Sampling (BUGS). Freeware available, e.g. WinBUGS, OpenBUGS, etc.

Hidden Markov Models (HMMs)

Motivation
Approach
- Classical approach
- Other approaches
- HMMs
- Evaluation

Results

- Idea behind HMMs: there is a ‘hidden” (latent, unobserved) discrete random variable, representing some part of the disease process. Observed variables are modeled, conditional upon the hidden state. Thus, if we know the state we also know the distribution of observed random variable.
- Markov property: conditional probability of state change (transition probability) depends only on the value of latent state at previous time point. Thus specify the Markov model for k states with a $k \times k$ matrix of transition probabilities, and the distributions of the observed data conditional on the hidden state.
- Parameter estimation accomplished with Bayesian inference Using Gibbs Sampling (BUGS). Freeware available, e.g. WinBUGS, OpenBUGS, etc.

Hidden Markov Models (HMMs)

Motivation
Approach
- Classical approach
- Other approaches
- HMMs
- Evaluation

Results

- Idea behind HMMs: there is a ‘hidden” (latent, unobserved) discrete random variable, representing some part of the disease process. Observed variables are modeled, conditional upon the hidden state. Thus, if we know the state we also know the distribution of observed random variable.
- Markov property: conditional probability of state change (transition probability) depends only on the value of latent state at previous time point. Thus specify the Markov model for k states with a $k \times k$ matrix of transition probabilities, and the distributions of the observed data conditional on the hidden state.
- Parameter estimation accomplished with Bayesian inference Using Gibbs Sampling (BUGS). Freeware available, e.g. WinBUGS, OpenBUGS, etc.

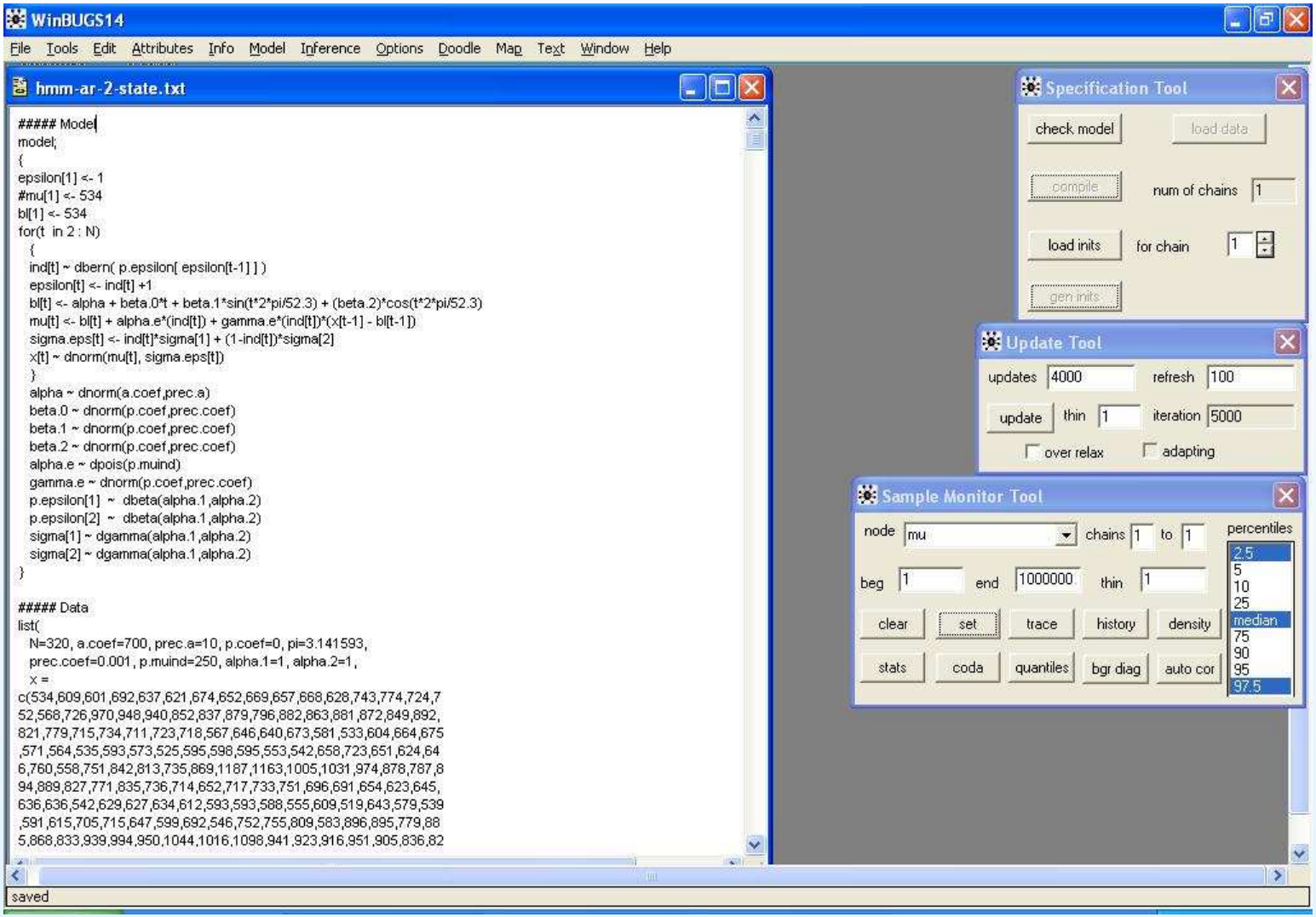
WinBUGS screen shot

Motivation

Approach

- Classical approach
- Other approaches
- HMMs
- Evaluation

Results



Hidden Markov Models (HMMs)

Motivation
Approach
- Classical approach
- Other approaches
- HMMs
- Evaluation

Results

- Computationally demanding part of model fitting is algorithmic search for the most likely sequence of hidden states, given the observed data. Other parameters (e.g. distributional models for observed variables) estimated simultaneously via Gibbs sampling.
- HMMs used previously for sentinel ILI data from France by Le Strat and Carrat (Stat Med 1999) as well as Rath, Carreras, Sebastiani (Proc IDA 2003). Cooper and Lipsitch (Biostat 2004) applied HMMs to nosocomial infections in hospitals. Various other applications to disease data.
- Latent variable provides information about mechanism of disease. Epidemic and non-epidemic behavior are modeled separately.

Hidden Markov Models (HMMs)

Motivation

Approach
- Classical approach
- Other approaches
- HMMs
- Evaluation

Results

- Computationally demanding part of model fitting is algorithmic search for the most likely sequence of hidden states, given the observed data. Other parameters (e.g. distributional models for observed variables) estimated simultaneously via Gibbs sampling.
- HMMs used previously for sentinel ILI data from France by Le Strat and Carrat (Stat Med 1999) as well as Rath, Carreras, Sebastiani (Proc IDA 2003). Cooper and Lipsitch (Biostat 2004) applied HMMs to nosocomial infections in hospitals. Various other applications to disease data.
- Latent variable provides information about mechanism of disease. Epidemic and non-epidemic behavior are modeled separately.

Hidden Markov Models (HMMs)

Motivation
Approach
- Classical approach
- Other approaches
- HMMs
- Evaluation

Results

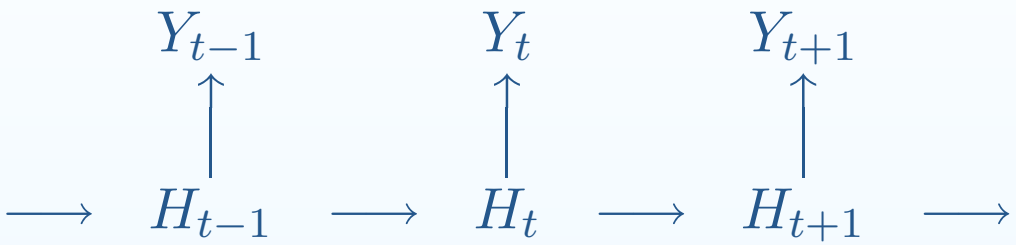
- Computationally demanding part of model fitting is algorithmic search for the most likely sequence of hidden states, given the observed data. Other parameters (e.g. distributional models for observed variables) estimated simultaneously via Gibbs sampling.
- HMMs used previously for sentinel ILI data from France by Le Strat and Carrat (Stat Med 1999) as well as Rath, Carreras, Sebastiani (Proc IDA 2003). Cooper and Lipsitch (Biostat 2004) applied HMMs to nosocomial infections in hospitals. Various other applications to disease data.
- Latent variable provides information about mechanism of disease. Epidemic and non-epidemic behavior are modeled separately.

Hidden Markov Models (HMMs)

Motivation

Approach
- Classical approach
- Other approaches
- HMMs
- Evaluation

Results



Y_t are observed data i.e. weekly P&I counts.
H_t are the hidden states (for us, 2-state model).
Arrows indicate conditional dependencies.

$$Y_t \sim \alpha_0 + \alpha_1 t + \beta_1 \sin\left(\frac{2\pi t}{52}\right) + \beta_2 \cos\left(\frac{2\pi t}{52}\right) \Big| H_t = 0$$

$$Y_t \sim \left(\alpha_0 + \alpha_{\mathbf{e}}\right) + \alpha_1 t + \beta_1 \sin\left(\frac{2\pi t}{52}\right) + \beta_2 \cos\left(\frac{2\pi t}{52}\right) \Big| H_t = 1$$

Evaluation

Motivation
Approach
- Classical approach
- Other approaches
- HMMs
- Evaluation

Results

- Our approach: systematically investigate various HMMs and evaluate to improve univariate time series models for influenza. Test bed data are P&I mortality figures from CDC 122 Cities surveillance system.
- Straightforward evaluation scheme to compare models: use fixed period of mortality data (e.g. 1990-1994) to fit all models. Use subsequent year (1995) to simulate prospective surveillance and calculate one-step-ahead residuals.
- Change time periods and average to ensure evaluation is not dependent on particular years chosen for model fitting and predictions.
- Compare several HMMs; Serfling's method; PARMA; perhaps other methods? Consider many-step-ahead predictive power.

Evaluation

Motivation
Approach
- Classical approach
- Other approaches
- HMMs
- Evaluation

Results

- Our approach: systematically investigate various HMMs and evaluate to improve univariate time series models for influenza. Test bed data are P&I mortality figures from CDC 122 Cities surveillance system.
- Straightforward evaluation scheme to compare models: use fixed period of mortality data (e.g. 1990-1994) to fit all models. Use subsequent year (1995) to simulate prospective surveillance and calculate one-step-ahead residuals.
- Change time periods and average to ensure evaluation is not dependent on particular years chosen for model fitting and predictions.
- Compare several HMMs; Serfling's method; PARMA; perhaps other methods? Consider many-step-ahead predictive power.

Evaluation

Motivation
Approach
- Classical approach
- Other approaches
- HMMs
- Evaluation

Results

- Our approach: systematically investigate various HMMs and evaluate to improve univariate time series models for influenza. Test bed data are P&I mortality figures from CDC 122 Cities surveillance system.
- Straightforward evaluation scheme to compare models: use fixed period of mortality data (e.g. 1990-1994) to fit all models. Use subsequent year (1995) to simulate prospective surveillance and calculate one-step-ahead residuals.
- Change time periods and average to ensure evaluation is not dependent on particular years chosen for model fitting and predictions.
- Compare several HMMs; Serfling's method; PARMA; perhaps other methods? Consider many-step-ahead predictive power.

Evaluation

Motivation
Approach
- Classical approach
- Other approaches
- HMMs
- Evaluation

Results

- Our approach: systematically investigate various HMMs and evaluate to improve univariate time series models for influenza. Test bed data are P&I mortality figures from CDC 122 Cities surveillance system.
- Straightforward evaluation scheme to compare models: use fixed period of mortality data (e.g. 1990-1994) to fit all models. Use subsequent year (1995) to simulate prospective surveillance and calculate one-step-ahead residuals.
- Change time periods and average to ensure evaluation is not dependent on particular years chosen for model fitting and predictions.
- Compare several HMMs; Serfling's method; PARMA; perhaps other methods? Consider many-step-ahead predictive power.

Motivation

Approach

Results
- Preliminary results
- Data
- Models
- Model fits
- Residuals
- Conclusions

Results

Preliminary results

Motivation
Approach
Results
- Preliminary results
- Data
- Models
- Model fits
- Residuals
- Conclusions

- Research supported by pilot funds from the Blood Center of Wisconsin. Second month of a 10 month funding period; results are preliminary.
- Presenting goodness-of-fit evaluation only; prospective evaluation in progress.
- First step in research program: evaluate HMMs on national mortality data. Future work will incorporate results of univariate modeling into spatio-temporal models at the regional/city levels, e.g. using dynamic Bayesian networks as in Sebastiani, Mandl et al. (Stat Med, in press).
- Eventually, follow similar approach with influenza-like illness (ILI) data. Allows for predictive spatio-temporal models of influenza morbidity.

Preliminary results

Motivation
Approach
Results
- Preliminary results
- Data
- Models
- Model fits
- Residuals
- Conclusions

- Research supported by pilot funds from the Blood Center of Wisconsin. Second month of a 10 month funding period; results are preliminary.
- Presenting goodness-of-fit evaluation only; prospective evaluation in progress.
- First step in research program: evaluate HMMs on national mortality data. Future work will incorporate results of univariate modeling into spatio-temporal models at the regional/city levels, e.g. using dynamic Bayesian networks as in Sebastiani, Mandl et al. (Stat Med, in press).
- Eventually, follow similar approach with influenza-like illness (ILI) data. Allows for predictive spatio-temporal models of influenza morbidity.

Preliminary results

Motivation
Approach
Results
- Preliminary results
- Data
- Models
- Model fits
- Residuals
- Conclusions

- Research supported by pilot funds from the Blood Center of Wisconsin. Second month of a 10 month funding period; results are preliminary.
- Presenting goodness-of-fit evaluation only; prospective evaluation in progress.
- First step in research program: evaluate HMMs on national mortality data. Future work will incorporate results of univariate modeling into spatio-temporal models at the regional/city levels, e.g. using dynamic Bayesian networks as in Sebastiani, Mandl et al. (Stat Med, in press).
- Eventually, follow similar approach with influenza-like illness (ILI) data. Allows for predictive spatio-temporal models of influenza morbidity.

Preliminary results

Motivation
Approach
Results
- Preliminary results
- Data
- Models
- Model fits
- Residuals
- Conclusions

- Research supported by pilot funds from the Blood Center of Wisconsin. Second month of a 10 month funding period; results are preliminary.
- Presenting goodness-of-fit evaluation only; prospective evaluation in progress.
- First step in research program: evaluate HMMs on national mortality data. Future work will incorporate results of univariate modeling into spatio-temporal models at the regional/city levels, e.g. using dynamic Bayesian networks as in Sebastiani, Mandl et al. (Stat Med, in press).
- Eventually, follow similar approach with influenza-like illness (ILI) data. Allows for predictive spatio-temporal models of influenza morbidity.

Data

Motivation
Approach
Results
- Preliminary results
- Data
- Models
- Model fits
- Residuals
- Conclusions

- CDC has been operating 122 Cities program continuously since (circa) 1960. Weekly counts of deaths attributed to pneumonia and influenza (P&I) reported to CDC by each of the participating cities within 2-3 weeks, as well as total deaths for week.
- Covers approx. 25% of the U.S. pop'n. Basis for CDC determination of epidemic influenza (Serfling).
- Age-specific counts available. 122 cities divided into 9 administrative regions, roughly 14 cities per region.
- Limitations of data: difficult to accurately attribute deaths to influenza; mortality known to lag morbidity (e.g. ILI activity) by 2-4 weeks or more; behavior of mortality curve may differ from that of influenza morbidity.

Data

Motivation
Approach
Results
- Preliminary results
- Data
- Models
- Model fits
- Residuals
- Conclusions

- CDC has been operating 122 Cities program continuously since (circa) 1960. Weekly counts of deaths attributed to pneumonia and influenza (P&I) reported to CDC by each of the participating cities within 2-3 weeks, as well as total deaths for week.
- Covers approx. 25% of the U.S. pop'n. Basis for CDC determination of epidemic influenza (Serfling).
- Age-specific counts available. 122 cities divided into 9 administrative regions, roughly 14 cities per region.
- Limitations of data: difficult to accurately attribute deaths to influenza; mortality known to lag morbidity (e.g. ILI activity) by 2-4 weeks or more; behavior of mortality curve may differ from that of influenza morbidity.

Data

Motivation

Approach

Results
- Preliminary results
- Data
- Models
- Model fits
- Residuals
- Conclusions

- CDC has been operating 122 Cities program continuously since (circa) 1960. Weekly counts of deaths attributed to pneumonia and influenza (P&I) reported to CDC by each of the participating cities within 2-3 weeks, as well as total deaths for week.
- Covers approx. 25% of the U.S. pop'n. Basis for CDC determination of epidemic influenza (Serfling).
- Age-specific counts available. 122 cities divided into 9 administrative regions, roughly 14 cities per region.
- Limitations of data: difficult to accurately attribute deaths to influenza; mortality known to lag morbidity (e.g. ILI activity) by 2-4 weeks or more; behavior of mortality curve may differ from that of influenza morbidity.

Data

Motivation
Approach
Results
- Preliminary results
- Data
- Models
- Model fits
- Residuals
- Conclusions

- CDC has been operating 122 Cities program continuously since (circa) 1960. Weekly counts of deaths attributed to pneumonia and influenza (P&I) reported to CDC by each of the participating cities within 2-3 weeks, as well as total deaths for week.
- Covers approx. 25% of the U.S. pop'n. Basis for CDC determination of epidemic influenza (Serfling).
- Age-specific counts available. 122 cities divided into 9 administrative regions, roughly 14 cities per region.
- Limitations of data: difficult to accurately attribute deaths to influenza; mortality known to lag morbidity (e.g. ILI activity) by 2-4 weeks or more; behavior of mortality curve may differ from that of influenza morbidity.

Models

Motivation
Approach
Results
- Preliminary results
- Data
- Models
- Model fits
- Residuals
- Conclusions

1. Traditional cyclic model (Serfling). OLS regression with terms for intercept, linear trend, two periodic terms for sinusoid with phase shift.
2. Periodic auto-regression (PARMA) fits cyclic model plus additional ARMA terms. Fixed order of ARMA model at (1,0).
3. Naive 2-state HMM. Non-epidemic state, data follow Serfling's model. Epidemic state involves a simple mean shift.
4. 2-state AR-HMM. Non-epidemic state, data follow PARMA. Epidemic state auto-regresses deviation from cyclic baseline.

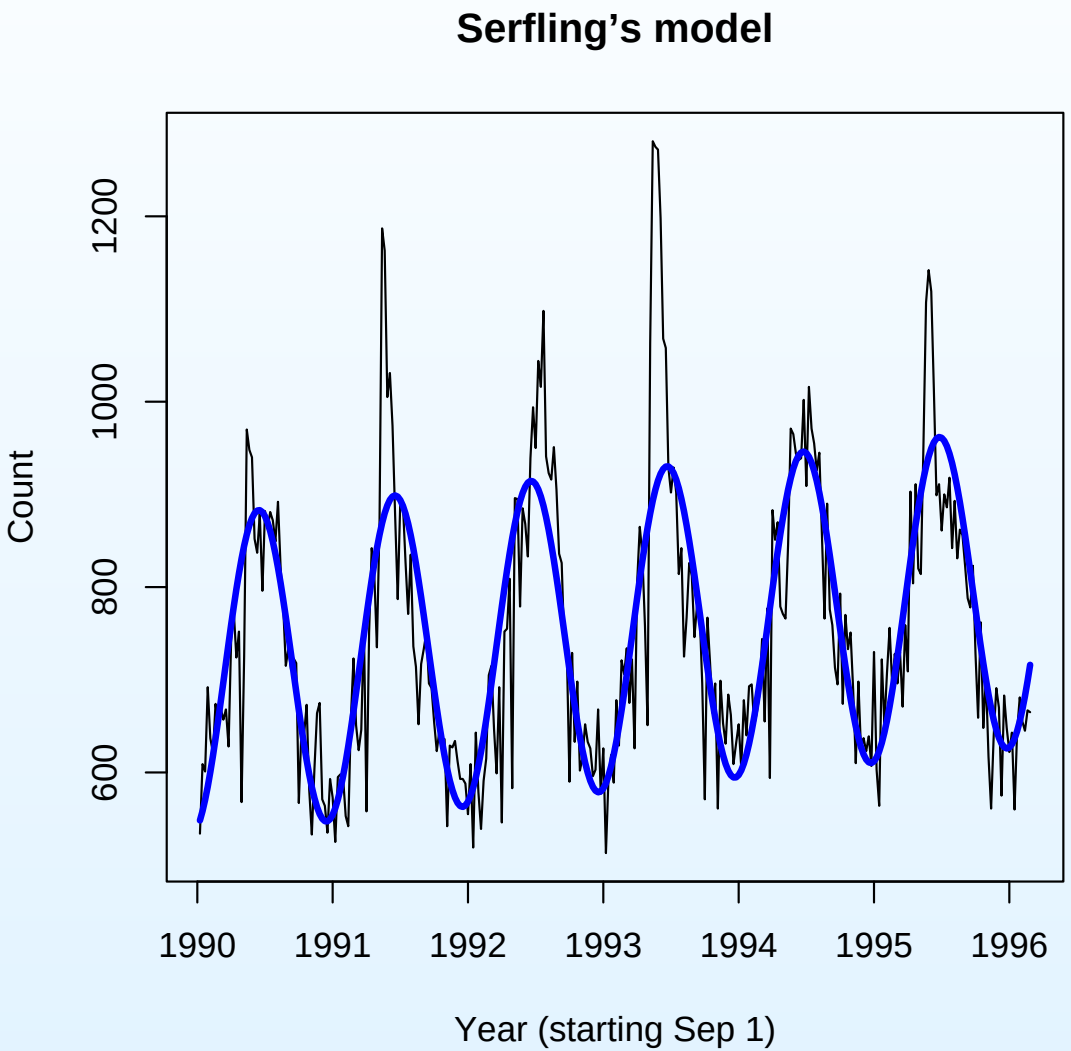
Serfling

Motivation

Approach

Results

- Preliminary results
- Data
- Models
- Model fits
- Residuals
- Conclusions



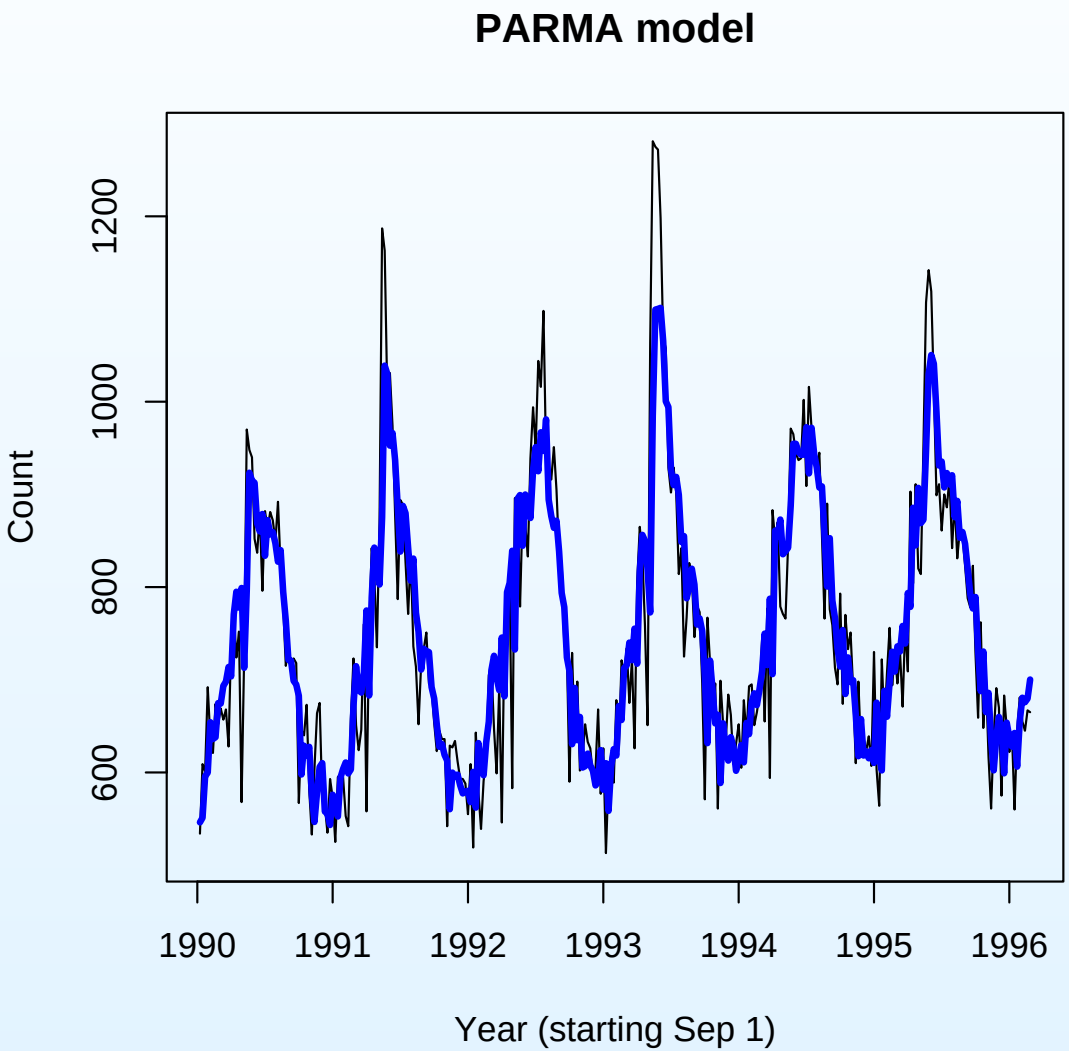
PARMA

Motivation

Approach

Results

- Preliminary results
- Data
- Models
- Model fits
- Residuals
- Conclusions



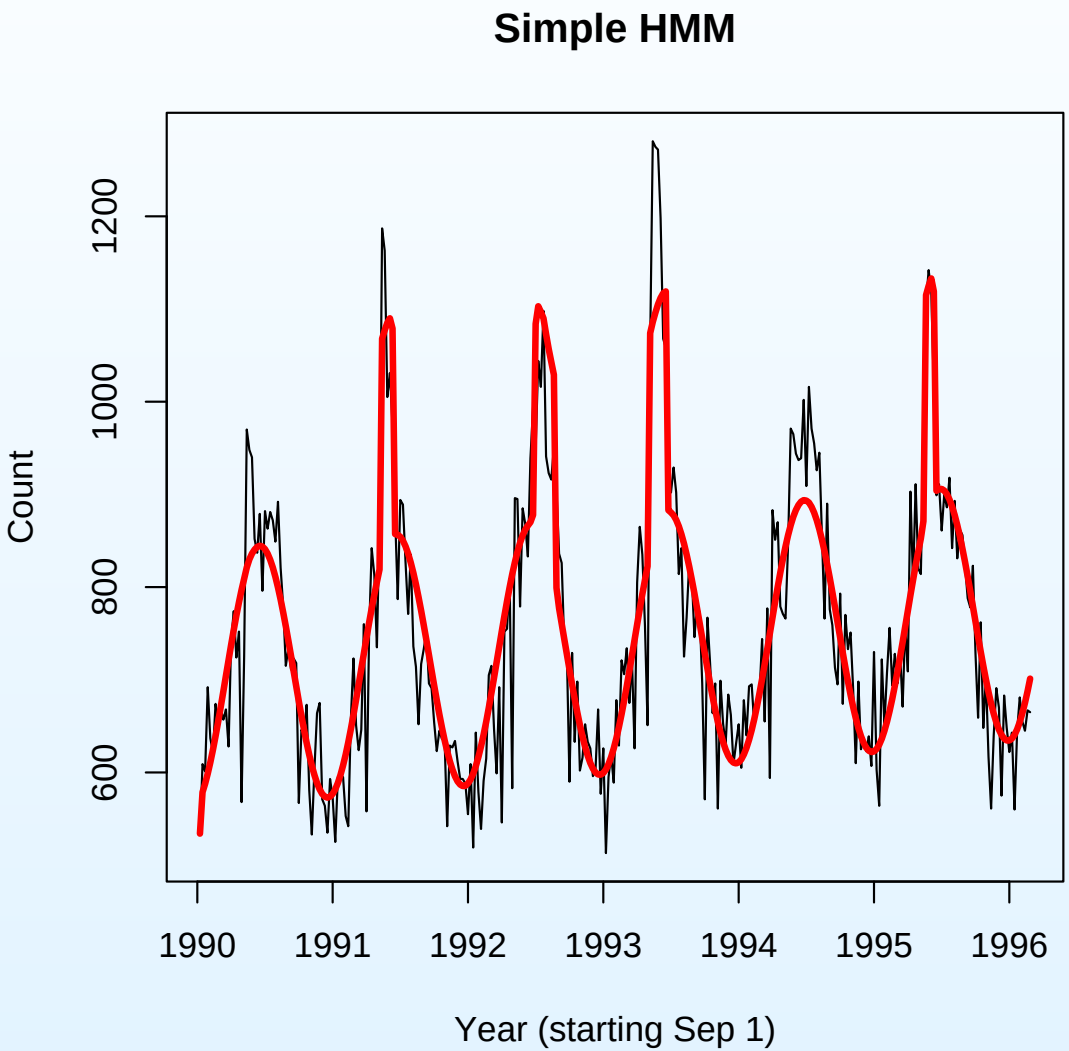
Simple HMM

Motivation

Approach

Results

- Preliminary results
- Data
- Models
- Model fits
- Residuals
- Conclusions



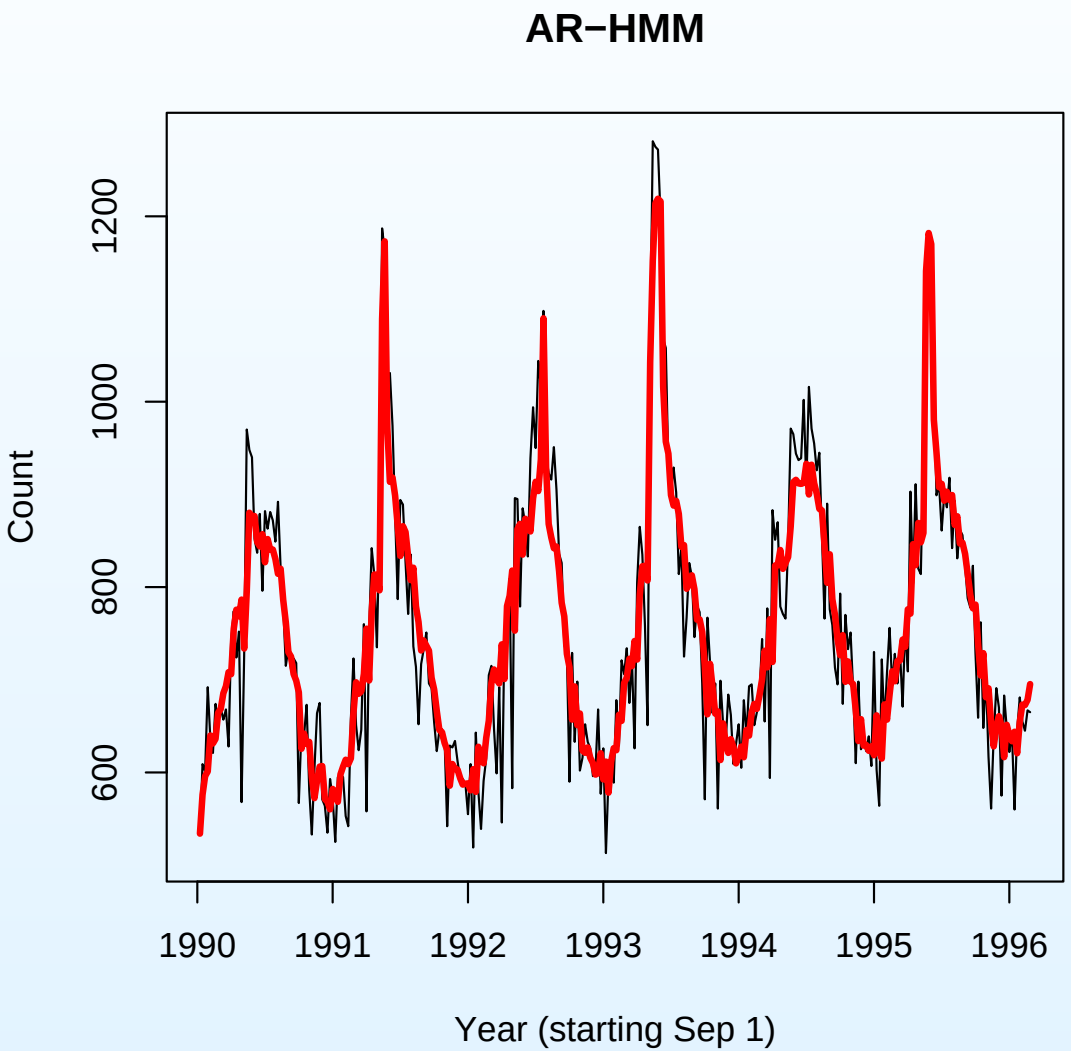
AR-HMM

Motivation

Approach

Results

- Preliminary results
- Data
- Models
- Model fits
- Residuals
- Conclusions



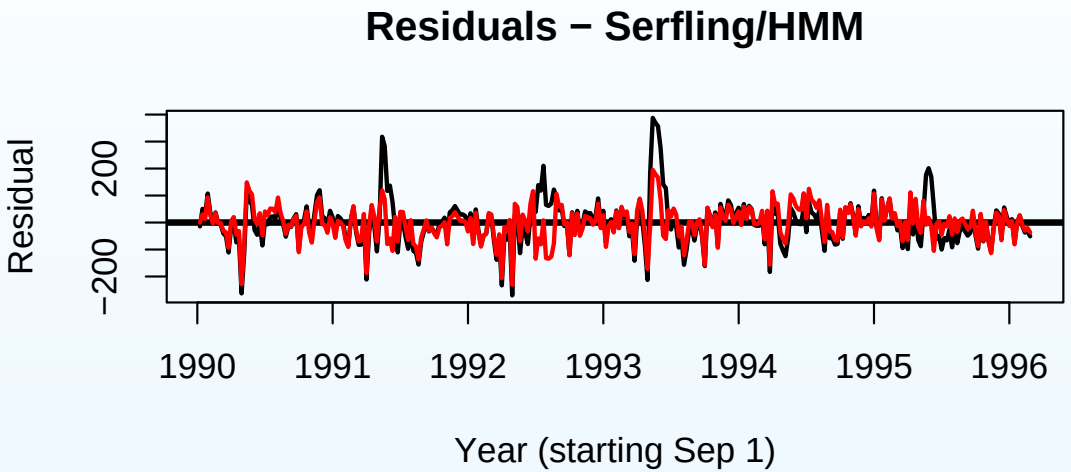
Residuals

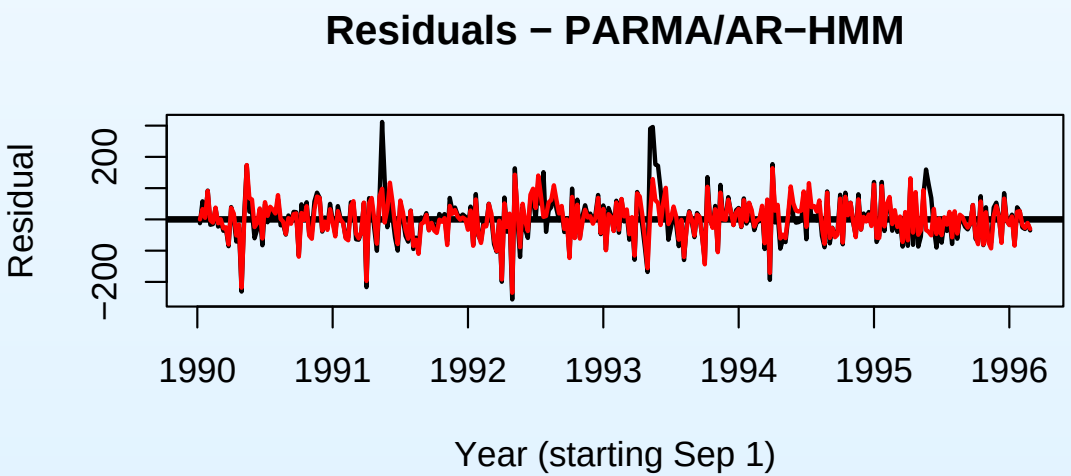
Motivation

Approach

Results

- Preliminary results
- Data
- Models
- Model fits
- Residuals
- Conclusions





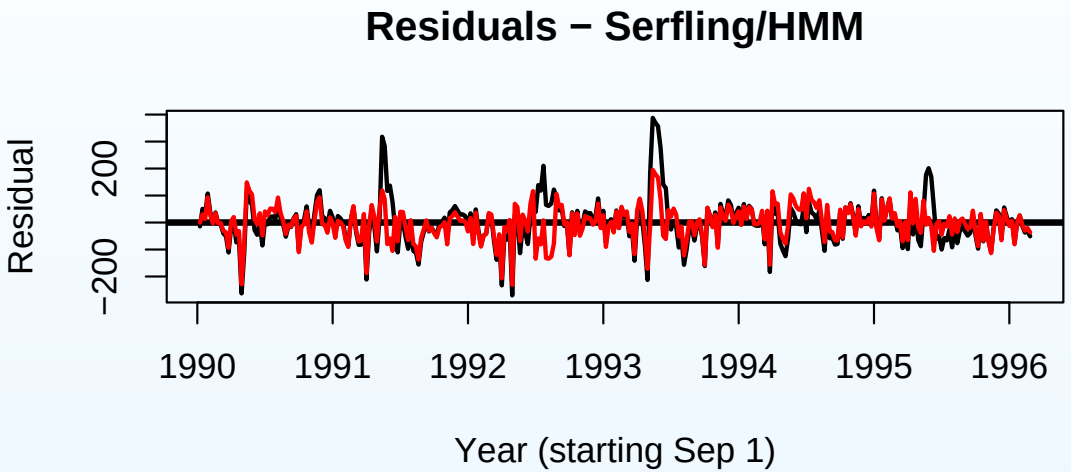
Residuals

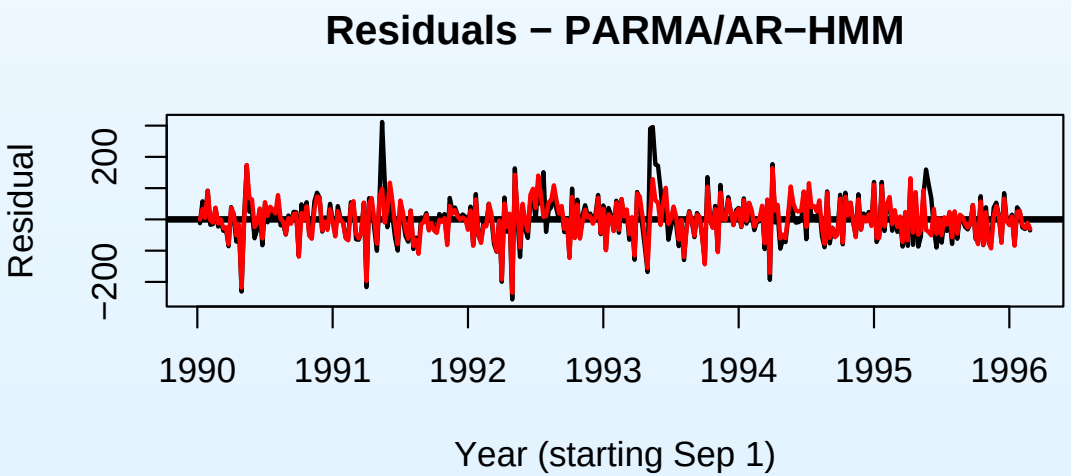
Motivation

Approach

Results

- Preliminary results
- Data
- Models
- Model fits
- Residuals
- Conclusions





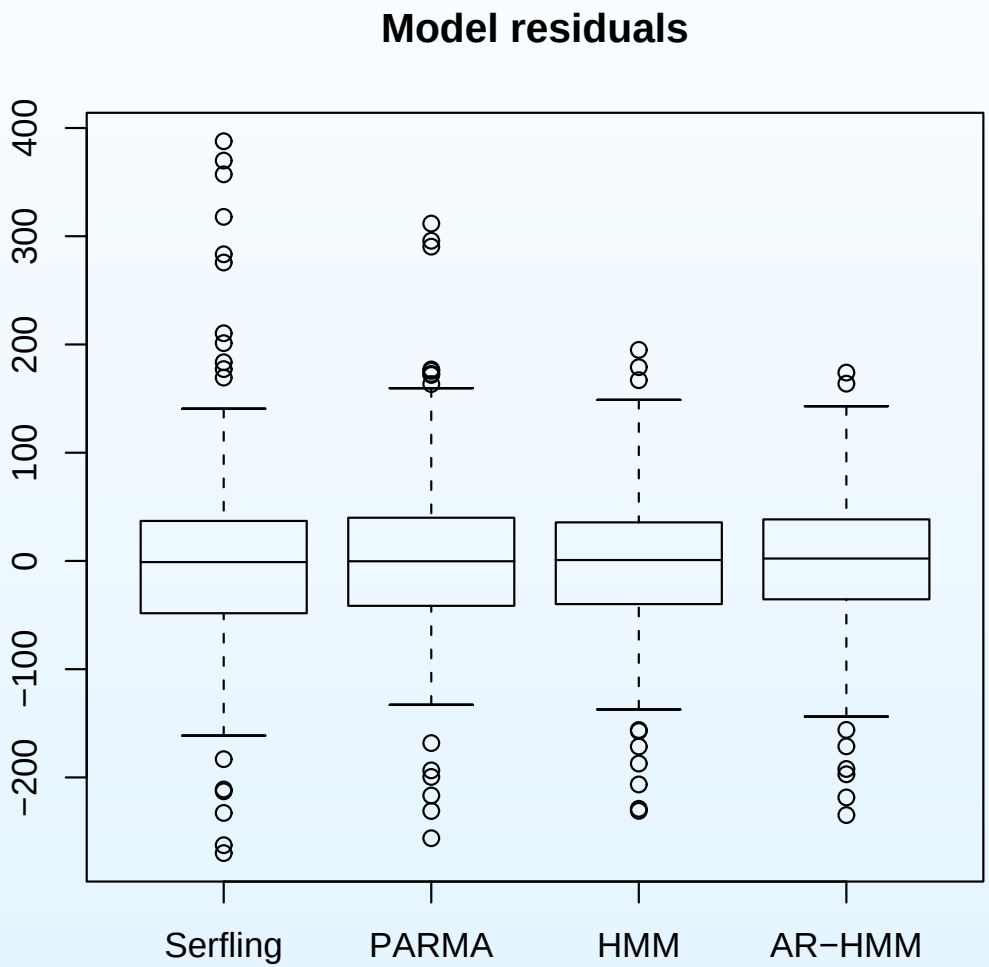
Residuals

Motivation

Approach

Results

- Preliminary results
- Data
- Models
- Model fits
- Residuals
- Conclusions



Residuals

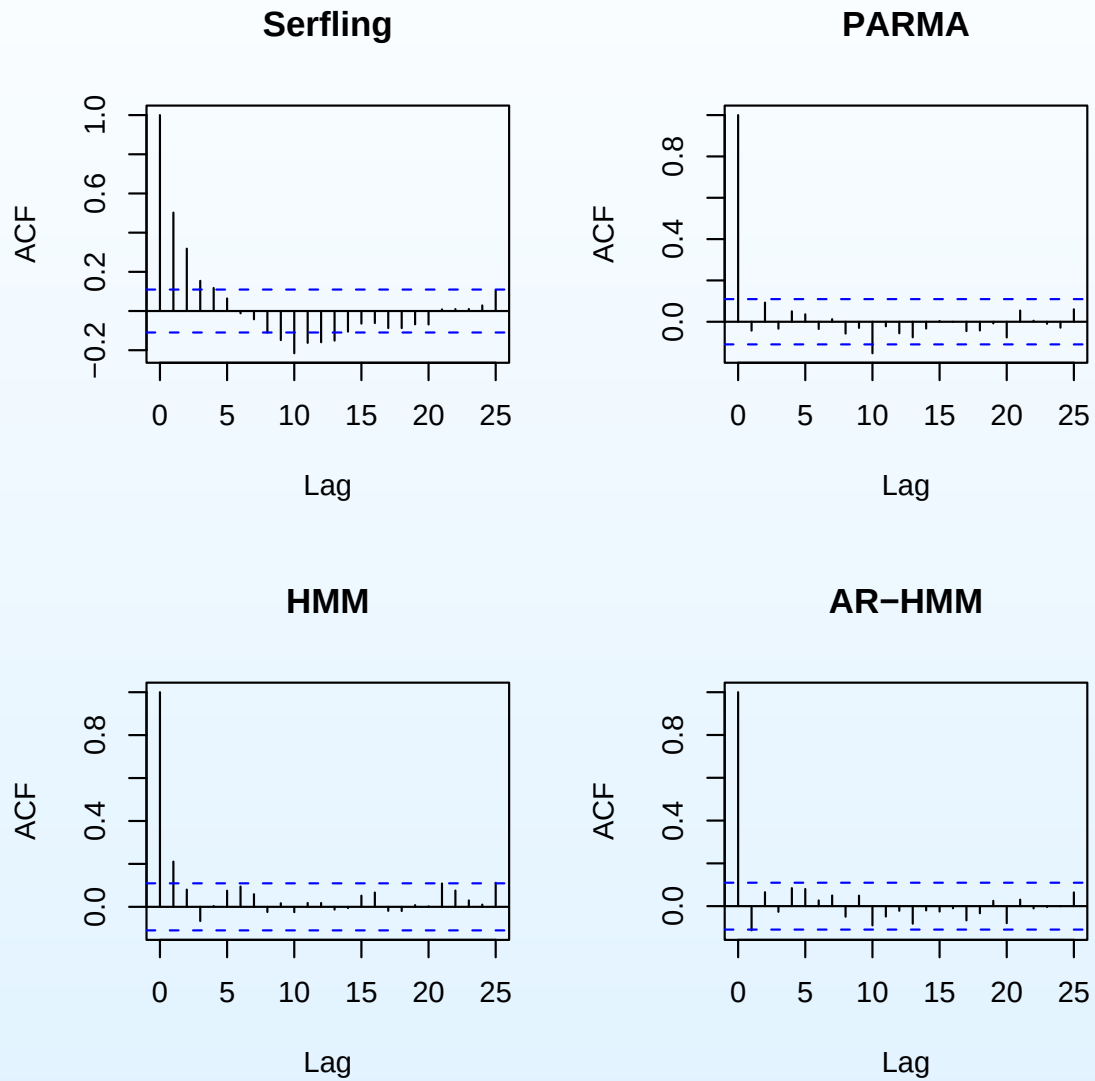
Motivation
Approach
Results
- Preliminary results
- Data
- Models
- Model fits
- Residuals
- Conclusions

Both HMMs provide a roughly 25% reduction in RMSE from Serfling, roughly 10% reduction for PARMA.

Model	RMSE
Serfling	83.3
PARMA	72.0
Simple HMM	63.7
AR-HMM	60.4

ACF of residuals

Motivation
Approach
Results
- Preliminary results
- Data
- Models
- Model fits
- Residuals
- Conclusions



Conclusions

Motivation
Approach
Results
- Preliminary results
- Data
- Models
- Model fits
- Residuals
- Conclusions

- Temporal modeling of influenza surveillance data can be substantially improved by implementing straightforward time series methods.
- HMMs are a natural choice for modeling influenza data. Maintain some information about mechanism of disease and allow for explicit modeling of epidemic and non-epidemic phases.
- Further evaluation should be followed by efforts to integrate several time series across spatial regions. Continue to work towards predictive spatio-temporal models of influenza.

Conclusions

Motivation
Approach
Results
- Preliminary results
- Data
- Models
- Model fits
- Residuals
- Conclusions

- Temporal modeling of influenza surveillance data can be substantially improved by implementing straightforward time series methods.
- HMMs are a natural choice for modeling influenza data. Maintain some information about mechanism of disease and allow for explicit modeling of epidemic and non-epidemic phases.
- Further evaluation should be followed by efforts to integrate several time series across spatial regions. Continue to work towards predictive spatio-temporal models of influenza.

Conclusions

Motivation
Approach
Results
- Preliminary results
- Data
- Models
- Model fits
- Residuals
- Conclusions

- Temporal modeling of influenza surveillance data can be substantially improved by implementing straightforward time series methods.
- HMMs are a natural choice for modeling influenza data. Maintain some information about mechanism of disease and allow for explicit modeling of epidemic and non-epidemic phases.
- Further evaluation should be followed by efforts to integrate several time series across spatial regions. Continue to work towards predictive spatio-temporal models of influenza.

Acknowledgements

Motivation
Approach
Results
- Preliminary results
- Data
- Models
- Model fits
- Residuals
- Conclusions

- Many thanks to DIMACS, the workshop organizers, and participants.
- The author gratefully acknowledges the contributions of his co-author Dr. Sebastiani and research assistant Suporn Sukpraprut.
- Thanks also to collaborators at the Harvard School of Public Health: Marcello Pagano, Laura Forsberg, Caroline Jeffery, and Miriam Nuño.
- Willie Anderson of CDC (NCHS) graciously offered his assistance in acquiring historical data in electronic format.
- Research partially supported by a pilot grant originating from the Blood Center of Wisconsin, via NIAID grant U19 AI62627-02.