Journal of Statistical Distributions
and Applications
a SpringerOpen Journal

**RESEARCH**                                                                 **Open Access**

# Modelling changes in Arctic Sea Ice Cover: an application of generalized and inflated beta and gamma densities

Peter Congdon

Correspondence:
p.congdon@qmul.ac.uk
School of Geography, Queen Mary
University of London, London E1
4NS, UK

**Abstract**

A modelling framework for changing Arctic sea ice extent is developed reflecting different trends and seasonal extremes in nine Arctic sub-regions. Core sub-regions retain partial ice cover throughout the year, and in winter show complete ice cover, while in peripheral sub-regions, winter coverage is not complete, and there is no ice cover at all in the summer. A generalized beta representation is developed for monthly ice extents in core sub-regions, with inflation to model maximum winter extents. For peripheral sub-regions, a gamma time series with excess zeroes (representing summer sea ice absence) is developed. Different trend representations (deterministic vs. stochastic) are compared for non-extreme observations. Other potential applications of the generalized beta density allowing zero or maximum inflation are discussed.

**Keywords:** Sea ice extent; Seasonal extremes; Inflated generalized beta; Inflated gamma; Bayesian

## 1 Introduction

In recent years, Arctic sea ice has been declining with wider climatic implications. The latter are multi-faceted and subject to discussion and uncertainty; see, for example, (Screen et al. 2013). However, one implication follows from the role of sea ice in regulating the global temperature via its ability (compared to the ocean surface) to reflect the sun's radiation. The albedo of snow-covered sea ice is 0.90, meaning it reflects 90% of the sun's radiation, whereas the ocean surface reflects only 10%. Less sea ice and more ocean surface will lead to a warmer Arctic, and contribute to global warming. There is also evidence that decline in Arctic sea ice is influencing atmospheric circulation (including the jet stream) within and beyond the Arctic, with impacts including winter cold surges in Europe and North America (Liu et al. 2012).

Satellite records of Arctic sea ice extent (available with full annual coverage since 1979) show a downward trend, albeit with fluctuations such as unusually low points in 1990 (Serreze et al. 1995) and in 2007 (Comiso et al. 2008). For the Arctic as a whole, the rate of decline in summer ice extent has been greater (Stroeve et al. 2012), with an 8.2% linear rate of decline (per decade) in September ice extent between 1979 and 2011, as against a 3.9% linear rate of decline for March ice extent (see Table 1).

**Table 1 Arctic sea ice extent**

|  | (millions of km$^2$) | |
|---|---|---|
|  | **March** | **September** |
| 1979 | 16.4 | 7.2 |
| 1980 | 16.1 | 7.8 |
| 1985 | 16.1 | 6.9 |
| 1990 | 15.9 | 6.2 |
| 1995 | 15.3 | 6.1 |
| 2000 | 15.3 | 6.3 |
| 2005 | 14.8 | 5.6 |
| 2006 | 14.5 | 5.9 |
| 2007 | 14.7 | 4.3 |
| 2008 | 15.2 | 4.7 |
| 2009 | 15.2 | 5.4 |
| 2010 | 15.1 | 4.9 |
| 2011 | 14.3 | 4.6 |
| 2012 | 15.2 | 3.6 |

However, the Arctic extent totals in Table 1 aggregate over sub-regions with dissimilar trends and seasonal extremes; for example, some sub-regions show substantial falls in extent in both winter and summer months. The present study accordingly reviews issues in modelling sea ice trends for Arctic sub-regions, and includes short term projections three years beyond the range of the observed data. While a number of forecasts of sea ice cover have been developed (e.g. Stroeve et al. 2007), these generally consider the entire Arctic, and do not consider sub-regional differences. The observed data are monthly NASA satellite records from 1979 to 2011 (396 observations), disaggregated to nine Arctic sub-regions. These are the central Arctic Ocean, the Canadian Archipelago, Hudson Bay, Sea of Okhotsk, Bering Sea, Baffin Bay, Greenland Sea, Kara-Barents Sea, and Gulf of St. Lawrence (see Figure 1). The observed data were extracted in June 2013 at http://neptune.gsfc.nasa.gov/csb/index.php?section=59.

There are important regional differences in trends and seasonal patterns according to location (e.g. Arctic core vs. periphery) that are relevant to statistical analysis and forecasts. These include modelling seasonal extremes. For central regions there is still some summer cover but a reversion to complete cover (maximum possible extent) in winter months. To represent this pattern, a generalized beta representation is applied, including "maximum inflation" to account for winter total coverage.

By contrast, in some non-core regions such as the Sea of Okhotsk and Bering Sea, winter extents do not cover the entire sea region, and there is no ice cover in the summer. For these regions, a gamma density with excess zeroes (representing summer sea ice absence) is developed. In both situations, regression models including trend and seasonal components are applied to represent the change in mean sea ice extent, while a separate logit regression is applied to the probability of inflation. Contrasting stochastic and deterministic trend representations are evaluated. Estimation is based on monthly observations in the first 31 years of satellite observations (1979-2009) with cross-validation using the remaining two years of observed data as a test sample. Bayesian estimation and inference are implemented using the Winbugs package via Markov chain Monte Carlo (MCMC) methods (Lunn et al. 2009).

**Figure 1 Map of case study sub-regions (image courtesy of National Snow and Ice Data Center (2012), University of Colorado, Boulder).**

## 2 Generalized beta time series regression

### 2.1 Application of the generalized beta to sea ice extents

As mentioned above, while there is decline in Arctic sea ice overall, there are considerable regional differences in trends and seasonal ice extent. What are here denoted "core" Arctic sea regions (namely the central Arctic Ocean, the Canadian Archipelago, and Hudson Bay) retain partial ice cover throughout the year, and in winter months show complete ice cover. For example, in 2011, the central Arctic had readings of 7.158mn $km^2$ throughout January to March while the Canadian Archipelago had readings of 0.751 mn $km^2$ for January through to April. Figure 2 shows monthly extent totals for the central Arctic ocean during 2007-11 and illustrates maximum inflation in winter. The trend in these three regions is similar to that in the Arctic ocean considered as an aggregate, namely stronger declines in summer extent.

**Figure 2 Observed monthly extents, millions km², 2007-11, central Arctic ocean.**
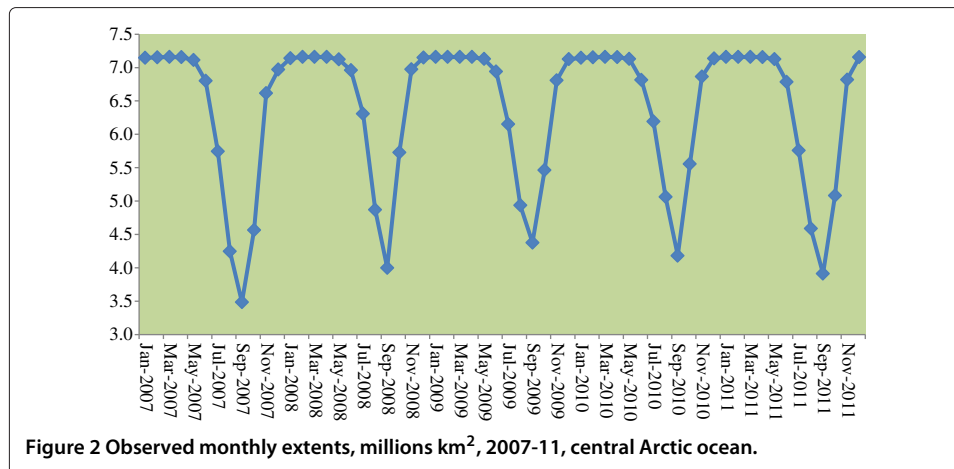
A time series representation needs to express the annual reversion to complete cover (maximum recurrence) in winter months, together with the irregular trend for declining extent in non-winter months. The application of a generalized form of the beta density is motivated by the fact that the observed ice extents $y$ can be seen as ratios $r = y/d$ to a maximum possible extent $d$, though substantive interest is in trends in extents $y$. It is important that the bounded nature of the response is included in any model. Another possibility might be some form of truncated sampling mechanism (e.g. a log-normal for extent readings $y$ with ceiling $d$) but this precludes any analysis of the factors producing seasonal extremes.

Consider the beta distribution on $(0, 1)$, with density function given by

$$f(z|a, b) = \frac{1}{B(a, b)} z^{a-1}(1 - z)^{b-1},$$

with $a > 0$, $b > 0$. An alternative representation (Ospina and Ferrari 2010) involves mean and precision parameters $(\mu, \phi)$, where $a = \mu\phi$, $b = (1 - \mu)\phi$, namely

$$f(z|\mu, \phi) = \frac{1}{B(\mu\phi, \phi - \phi\mu)} z^{\mu\phi-1}(1 - z)^{(1-\mu)\phi-1},$$

with $\phi > 0$, and $0 < \mu < 1$. This form facilitates separate modelling of mean and variance trends (Huang and Oosterlee 2008). For values $(a, b)$ apart from $a = b = 1$, the *Beta*$(a, b)$ density has zero mass at the extreme values 0 and 1, and zero-inflated or one-inflated versions of the beta need to be applied (Ospina and Ferrari 2010). Let $g = 0$ or 1, then inflation at either boundary is achieved by the mechanism

$$f(z|\alpha_g, \mu, \phi) = \begin{cases} \alpha_g, & z = g \\ (1 - \alpha_g)f(z|\mu, \phi), & z \in (0, 1), \end{cases}$$

where $\alpha_g$ is an inflation probability.

The generalized beta is obtained by extending the support interval to an arbitrary bounded interval $(c, d)$ (with $d > 0$) via a linear transformation $y = c + z(d - c)$ (Pham-Gia and Duong 1989), so that

$$f(y|a, b, c, d) = \frac{1}{B(a, b)(d - c)^{a+b-1}} (y - c)^{a-1}(d - y)^{b-1},$$

with equivalent representation

$$f(y|\mu,\phi,c,d) = \frac{1}{B(\mu\phi,\phi-\phi\mu)(d-c)^{\phi-1}}(y-c)^{\mu\phi-1}(d-y)^{(1-\mu)\phi-1},\tag{1}$$

with mean $c + (d-c)\mu$, and variance $(d-c)^2\mu(1-\mu)/(\phi+1)$.

For the generalized beta in (1), inflation will need to be applied for values occurring at the boundary points, when $y = c$ or $y = d$ (this may be termed minimum and maximum inflation). The maximum inflated version of the generalized beta is particularly relevant to the sea ice application and has the form

$$f(y|\alpha_d,\mu,\phi,c,d) = \begin{cases} \alpha_d, & y = d \\ (1-\alpha_d)f(y|\mu,\phi,c,d), & y \in (c,d). \end{cases}$$

In the generalized beta applied to core Arctic regions, $c = 0$ while $d$ is the maximum winter extent (namely $d_1$=7.158 mn $km^2$ for the central Arctic ocean, $d_2$=0.751mn $km^2$ for the Canadian Archipelago, and $d_3$=1.233 mn $km^2$ for Hudson Bay).

While summer minimum extents in the central Arctic and Canadian Archipelago remain well in excess of zero, those in Hudson Bay are becoming relatively small, e.g. $y = 0.025$ mn $km^2$ in September 2010. This raises the possibility of needing to represent both maximum and minimum inflation in the generalized beta. This can be handled by the mechanism

$$f(y|\alpha_c,\alpha_d,\mu,\phi,c,d) = \begin{cases} \alpha_c, & y = c \\ \alpha_d, & y = d \\ (1-\alpha_c-\alpha_d)f(y|\mu,\phi,c,d), & y \in (c,d), \end{cases}$$

where the vector of probabilities $(\alpha_c,\alpha_c,1-\alpha_c-\alpha_d)$ should be modelled using a multinomial logistic.

## 2.2 Generalized beta time series regression with maximum inflation for sea ice extents

Let $\{\mu_t,\phi_t,\alpha_{dt}\}$ denote the series of parameters underlying the $y_t$ series, $m = m_t$ represent the month that observation $t$ corresponds to, and $s = s_t$ represent the year corresponding to observation $t$ (e.g. $s = 2$ in 1980 for observations $t = 13,..,24$ and $s = 31$ for observations $t = 361,..,372$). A parsimonious time series model is sought (Ledolter and Abraham 1981), combining close fit with low predictive variability, especially for cross-validatory and out-of-sample predictions. As discussed in Section 4, these aspects of fit are assessed using a posterior predictive fit criterion (Laud and Ibrahim 1995). A parsimonious model for the level of the series is expressed by a logit regression in $\mu_t$,

$logit(\mu_t) = \Delta_{ms} + \vartheta_m$, where $\Delta_{ms}$ represents trend for each combination of month $m$ and year $s$, and $\vartheta_m$ represents seasonal effects.

Two options for the trend are considered. One option is a stochastic trend, with random variation around a central linear trend. To allow for steeper declines in some months, a discrete mixture is implemented via

$$\begin{aligned} \Delta_{ms} &= \rho_m\varphi_{1s} + (1-\rho_m)\varphi_{2s}, \\ \rho_m &\sim Bern(\pi_m), \\ \varphi_{0s} &= \delta_{00} + \delta_{01}s + u_s, \\ \varphi_{1s} &= \delta_{10} + \delta_{11}s + u_s, \\ u_s &\sim N\left(u_{s-1},\sigma_u^2\right), \end{aligned}\tag{2}$$

with $\rho_m \sim Bern(\pi_m)$ being binary indicators. With the constraint $\delta_{11} < \delta_{01}$, $\varphi_{1s}$ represents the stronger downward trend.

The other form of trend assumption (deterministic) involves a linear trend in years $s_t$ combined with a short term AR1 lag effect in extents $y_t$. The latter represents carry over between successive months; for example, if September extent is relatively low in a particular year, then October extent may also be relatively low. The linear trend may vary between months and by broad sub-period. For example, (Comiso et al. 2008) report a stronger decline during 1996-2007 than 1979-96. Here we consider three sub-periods $p = 1, ..., 3$ of 12 years, including out-of-sample years (2012-14), namely 1979-1990, 1991-2002, and 2003-2014. The trend parameter for a particular time point is chosen by a monthly specific discrete mixture between guide linear trend parameters, specific to broad period, $\Gamma_{0p}$ and $\Gamma_{1p}$, with $\Gamma_{1p} < \Gamma_{0p}$. The AR1 lag effect is also taken to vary by month, with normal priors for each monthly lag parameter. Thus for month $m = m_t$, and year $s_t$,

$$logit(\mu_t) = \Delta_t + \vartheta_m,$$
$$\Delta_t = \gamma_0 + \gamma_{1m}s_t + \gamma_{2m}y_{t-1},$$

where the linear trend $\gamma_{1m}$ for month $m = m_t$ in period $p$ is chosen using a discrete mixture

$$\gamma_{1m} = \rho_m\Gamma_{0p} + (1 - \rho_m)\Gamma_{1p}; \qquad \rho_m \sim Bern(\pi_m).$$

Remaining aspects of the model are applicable across different representations of trend. Seasonal (monthly) effects are represented by a Fourier series (Höhle and Paul 2008),

$$\vartheta_m = \sum_{j=1}^{J_1}(\beta_{2j-1}\cos(\omega jm) + \beta_{2j}\sin(\omega jm)), \tag{3}$$

where $\omega = 2\pi/M$, with $M = 12$, and $J_1$ is the number of harmonics. To allow for changing precision it is assumed that

$$log(\phi_t) = \eta_1 + \eta_{2m}s, \tag{4}$$

namely a linear trend (varying by month) in year units. For example, (Stroeve et al. 2012) find evidence of increased variability in overall Arctic sea ice extent, especially in summer months.

To represent extreme data (complete winter coverage), a logit regression is used to model the probabilities $\alpha_{dt}$ of maximum inflation, with form

$$logit(\alpha_{dt}) = \xi_0 + \sum_{j=1}^{J_2}(\xi_{2j-1}\cos(\omega jm_t) + \xi_{2j}\sin(\omega jm_t)).$$

A trend element in the inflation probability is not included as it would be confounded with the trend model in the mean.

### 2.3 Other generalized beta applications

While the application here focuses on sea ice extent and a time series application, the generalized inflated beta with mechanisms or regressions for both extreme and non-extreme observations has potential applications in other settings where the observations can be regarded as ratios $r_i = y_i/d_i$ of actual extents to a maximum extent $d_i$, but substantive interest is in the extents $y_i$. The extents may be, inter alia, expressed in spatial units (e.g.

areas in millions of square kilometres) or time units (e.g. durations in hours). As an example with time extents, one might analyse hours with cloud cover $y_i$ in relation to daylight hours $d_i$, while a spatial application might consider desertified extents $y_i$ in relation to total area extents $d_i$.

Data of this form can be considered as a form of compositional data, and widely used methods (Aitchison and Egozcue 2005, Butler and Glasbey 2008) focus on the ratios $r_i$, or more specifically the log ratios. Spatial applications involving compositional data and zero inflation have been described in (Leininger et al. 2013), but also focus on the ratios.

However, for policy purposes, the interest may be in trends or patterns in the extents themselves (i.e. the y-data), rather than in the ratios, as is the case with the sea ice extents. Alternatively stated, "compositions provide information only about the relative magnitudes of the compositional components and so interpretations involving absolute values ... cannot be justified" (Aitchison and Egozcue 2005) p. 839. Thus in the case of desertification (e.g. Zhao et al. 2010), the substantive focus may be on spreading desertification, implying analysis of desertified extents $y_i$. Some areas may be totally desertified with $y_i = d_i$ (maximum inflation). Regression modelling of desertification extents would then need to include a mechanism or regression describing maximum inflation, as would spatial forecasting (or interpolation) of desertified extents in situations where comprehensive assessment of desertification status is only available for some area units.

The generalized beta density with maximum or zero inflation might be potentially extended to Dirichlet density applications, and to generalized inflated Dirichlet densities parallel to equation (1), where there are more than two categories and where extreme observations can occur. For example, with three categories the observations would be $(y_{1i}, y_{2i}, y_{3i})$, with $\sum_k y_{ki} = d_i$, and maximum inflation occurring when any $y_{ki} = d_i$. The inflation probabilities can be modelled using a multiple logistic. For example, in the sea ice application, one may distinguish by sea-ice type (e.g. Fissel et al. 2011) between perennial multi-year sea ice and first-year ice, so that sub-region observations become $(y_1, y_2, y_3)$ for area covered by multi-year ice, area covered by first-year ice, and area without ice cover respectively.

## 3 Inflated gamma time series regression

In what are here denoted "peripheral" Arctic regions (Sea of Okhotsk, Bering Sea, Gulf of St. Lawrence), the winter ice extents do not provide complete ice cover of the sea region, while in at least one peak summer month (July, August, September) there is no ice cover at all. For example, Figure 3 shows the extent series for the Bering Sea for 2007-11. Let $y_t$ denote the observed extents (mn $km^2$) for these regions following a gamma density,

$$f(y|\mu, \phi) = \frac{(\phi/\mu)^\phi}{\Gamma(\phi)} y^{\phi-1} e(-\phi y/\mu),$$

with mean $\mu$ and variance $\mu^2/\phi$. A zero inflated version of the gamma is motivated by the need to represent absence of ice cover in summer months, as in the mechanism

$$f(y|\alpha_0, \mu, \phi) = \begin{cases} \alpha_0, & y = 0 \\ (1-\alpha_0)f(y|\mu, \phi), & y \in (0, \infty). \end{cases}$$
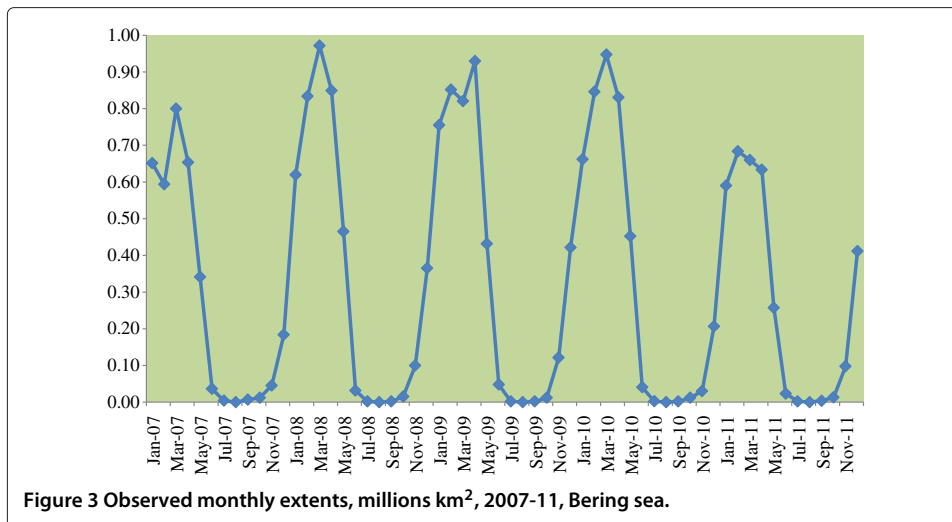
**Figure 3 Observed monthly extents, millions km$^2$, 2007-11, Bering sea.**

Let $\{\mu_t, \phi_t, \alpha_{0t}\}$ denote the series of parameters underlying the observations $y_t$. The link for $\mu_t$ depends on which of two options for trend is applied.

One option is a stochastic trend, with log-link

$$log(\mu_t) = \Delta_{ms} + \vartheta_m, \tag{5}$$

where $\Delta_{ms}$ is as in (2) above. The other option, a deterministic trend, combines a linear trend in years $s$ with a short term lag effect in the response. The linear trend may vary between months, and also between sub-period $p$. As above, the trend parameter for a particular time point is chosen according to a monthly specific discrete mixture between guide linear trends, specific to broad period, $\Gamma_{0p}$ and $\Gamma_{1p}$, with $\Gamma_{1p} < \Gamma_{0p}$. Lag effects are also taken to vary by month. To avoid an explosive impact of the lag effect (Blundell et al. 2002, Jung et al. 2006), the following link scheme provides the deterministic trend option,

$$\mu_t = \exp(\zeta_0 + \zeta_{1m}s_t + \vartheta_m) + \zeta_{2m}y_{t-1}, \tag{6}$$

with the constraint $\zeta_{2m} \geq 0$. For example, one may take $\zeta_{2m}$ to have exponential or gamma priors.

Remaining aspects of the model are applicable across different representations of trend. Seasonal effects $\vartheta_m$ are represented by a Fourier series, as in (3), and precision parameters modelled using the regression form (4). To represent summer extremes (no ice cover at all), a logit regression for the probability of zero inflation is used, with

$$logit(\alpha_{0t}) = \xi_0 + \sum_{j=1}^{J_2}(\xi_{2j-1}\cos(\omega j m_t) + \xi_{2j}\sin(\omega j m_t)).$$

For the remaining three Arctic sub-regions (Baffin Sea, Greenland Sea, Kara-Barents), the data series demonstrate neither maximum recurrence in winter, nor complete summer ice disappearance as yet. The maximum values observed for these three series (in millions $km^2$) are 1.766 (Baffin Sea, March 1993), 1.115 (Greenland Sea, January 1982) and 2.168 (Kara-Barents, April 1979), whereas recent winter maxima are below these levels. Recent summer extents might be taken to indicate incipient ice disappearance in summer, such as a reading of 0.04 mn $km^2$ in the Baffin Sea in August 2011. However, for

these regions a gamma regression may be adopted, without any inflation mechanism. The time series regression therefore involves $\{\mu_t, \phi_t\}$, and stochastic and deterministic trend options, as in (5) and (6) respectively, may be compared.

## 4 Model fit and comparison

### 4.1 Analysis framework

To illustrate a full comparative analysis, three sub-regions are considered as representative of the three sets of regions described above: the central Arctic, the Sea of Okhotsk and the Greenland Sea. Different models are applied to monthly observations from January 1979 to December 2009 (namely for $t = 1, .., T$ with $T = 372$), with in-sample cross-validation for the two year period 2010-2011, namely for $t = T + 1, ..., T_1$ with $T_1 = 396$. Out of sample predictions are made for a further 36 months, January 2012 to December 2014. Posterior inferences are based on the second halves of two chain runs of 50,000 iterations, with convergence assessed using Brooks-Gelman-Rubin statistics (Brooks and Gelman 1998).

### 4.2 Generalized beta regression with maximum inflation

The central Arctic is one of three sea regions with recurrent winter maxima and partial summer ice cover, and Figure 4 shows the trend in yearly averages for the central Arctic, with the anomalous low cover years (1990, 2007) apparent. In applying the generalized beta regression, normal priors with mean zero and variance 1000 are adopted on fixed effect parameters $\{\beta, \xi, \eta, \Gamma, \delta, u_1, \gamma_0, \gamma_1\}$, while the lag parameters $\gamma_{2m}$ are assigned $N(0,1)$ priors. A gamma prior with shape 1 and scale 0.001 is adopted for the precision $1/\sigma_u^2$, while $Beta(1, 1)$ priors are adopted for the $\pi_m$. The Fourier seasonal representations for the means $\mu_t$ and inflation probabilities $\alpha_{dt}$ have $J_1 = J_2 = 3$ harmonics, as insignificant regression effects $\{\beta, \xi\}$ were obtained at higher numbers of harmonics.

Predictions are obtained by generating replicate inflation indicators $D_{new,t} \sim Bern(\alpha_{dt})$, and replicate scaled beta values $dz_{new,t}$, where $z_{new,t} \sim Beta(\mu_t\phi_t, \phi_t - \phi_t\mu_t)$. Predictions are then

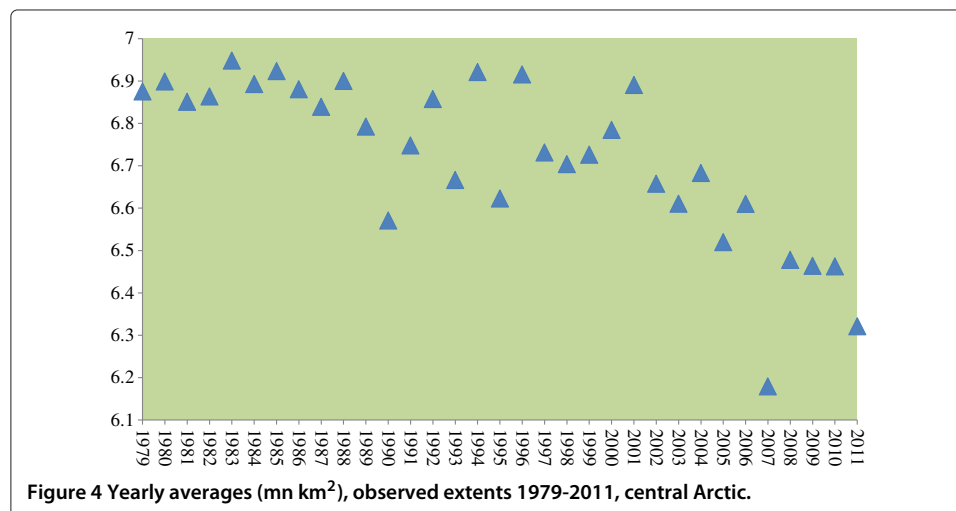$$y_{new,t} = D_{new,t}d + (1 - D_{new,t})dz_{new,t}.$$



**Figure 4 Yearly averages (mn km$^2$), observed extents 1979-2011, central Arctic.**

Let $M_{new,t}$ and $V_{new,t}$ be the posterior means and variances of $y_{new,t}$. In-sample predictive fit (within the training dataset from January 1979 through to the end of 2009) is based on the L-criterion (Laud and Ibrahim 1995), namely the square root of

$$L^2 = \sum_{t=1}^{T} V_{new,t} + \sum_{t=1}^{T} (M_{new,t} - y_t)^2.$$

Cross-validatory predictions beyond $T = 372$ (for the 24 months in 2010 and 2011) are made by treating $\{y_{T+1}, ..., y_{T_1}\}$ as missing data, and generating predictions $\{y_{new,T+1}, ..., y_{new,T_1}\}$. These can be compared with observed test data $y_{obs,t}$ for 2010 and 2011. Cross validatory fit using the L criterion, denoted CV-L, is based on

$$\text{CV-L}^2 = \sum_{t=T+1}^{T+24} V_{new,t} + \sum_{t=T+1}^{T+24} \left(M_{new,t} - y_{obs,t}\right)^2 = P_1 + P_2,$$

which will reflect the precision of future predictions ($P_1$) as well as the fit ($P_2$). To further assess predictive performance, a check is made whether observed extents are within 95% intervals of $y_{new}$. In a model that effectively reproduces the data, predictive coverage is at or above 95% (Gelfand 1996, p. 158).

Table 2 shows posterior means and variances during the cross-validation period, while Table 3 summarises the L-criteria for both cross-validation periods (test data) and the in-sample estimation (training data). It is apparent that the deterministic trend has a better cross-validation performance but a worse in-sample performance. In particular, the deterministic trend is less adaptable to anomalous observations such as occurred during the summers of 1990 and 2007. Both models are effective in reproducing the original data satisfactorily in terms of coverage obtained by $y_{new,t}$.

The better out-of-sample cross validation under the deterministic trend is due to lower predictive variability ($P_1 = 2.06$ under the deterministic trend, whereas $P_1 = 6.13$ under the stochastic trend); pure fit is broadly similar between the two options ($P_2 = 0.41$ under deterministic trend, as against $P_2 = 0.34$ under stochastic trend). Figure 5 shows observations and predictions for 2010-11, as well as out-of-sample predictions through to December 2014, under the deterministic trend option. Posterior mean predictions for September extent in 2012, 2013 and 2014 are respectively 3.88, 3.78 and 3.68mn $km^2$. Table 4 summarizes selected parameters from the deterministic trend model. The strongest downward trends in mean extent, as represented by the guide parameters $\Gamma_{0p}$ and $\Gamma_{1p}$, are for 1979-90 and for the partially observed third period (2003-14), with estimates based on observations for 2003-09. Parameters $\eta_{2m}$ for trends in precision are nonsignificant in summer months (i.e the 95% credible intervals straddle zero), though the 95% credible intervals for $\eta_{2m}$ in July and August are biased to negative values, namely (-0.052,0.009) and (-0.058,0.009).

### 4.3 Gamma regression time series

The Sea of Okhotsk is one of three regions with recurrent summer disappearance of ice cover. Table 5 shows that unlike the trend for the whole Arctic as an aggregate (e.g. see Table 1), the most pronounced declines in average extent in the Sea of Okhotsk have been in the first quarter months (January, February, March). In applying gamma time series regression with zero inflation, priors are as in Section 4.2, except that exponential $E(1)$ priors are adopted for the lag parameters $\zeta_{2m}$, in order to ensure that $\mu_t$ is positive - see

**Table 2 In-sample cross validation (for 2010 and 2011)**

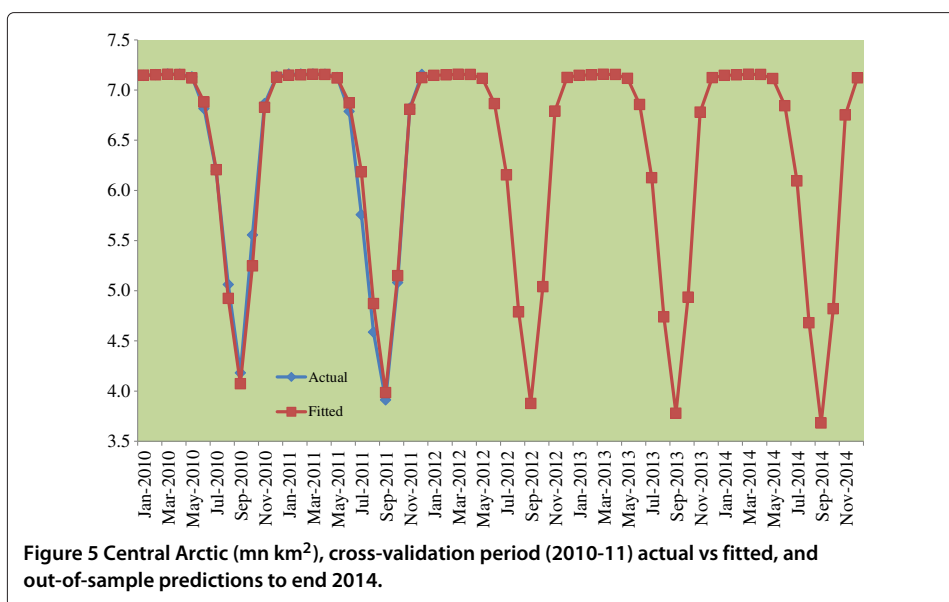| Month | Actual (Test data) | Stochastic trend | | | Deterministic trend | | |
|---|---|---|---|---|---|---|---|
| | | Posterior mean | Posterior median | Posterior variance | Posterior mean | Posterior median | Posterior variance |
| Jan-2010 | 7.145 | 7.148 | 7.158 | 0.0004 | 7.148 | 7.158 | 0.0004 |
| Feb-2010 | 7.154 | 7.154 | 7.158 | 0.0001 | 7.151 | 7.158 | 0.0006 |
| Mar-2010 | 7.158 | 7.157 | 7.158 | 0.0000 | 7.157 | 7.158 | 0.0000 |
| Apr-2010 | 7.157 | 7.155 | 7.158 | 0.0001 | 7.155 | 7.158 | 0.0000 |
| May-2010 | 7.130 | 7.112 | 7.119 | 0.0013 | 7.120 | 7.122 | 0.0003 |
| Jun-2010 | 6.814 | 6.823 | 6.865 | 0.0419 | 6.881 | 6.901 | 0.0173 |
| Jul-2010 | 6.192 | 6.077 | 6.147 | 0.2322 | 6.206 | 6.234 | 0.1054 |
| Aug-2010 | 5.061 | 4.886 | 4.944 | 0.5817 | 4.923 | 4.917 | 0.2896 |
| Sep-2010 | 4.181 | 4.174 | 4.200 | 0.7522 | 4.072 | 4.074 | 0.2886 |
| Oct-2010 | 5.555 | 5.586 | 5.667 | 0.4638 | 5.248 | 5.278 | 0.2480 |
| Nov-2010 | 6.864 | 6.808 | 6.863 | 0.0693 | 6.828 | 6.863 | 0.0406 |
| Dec-2010 | 7.137 | 7.108 | 7.138 | 0.0061 | 7.127 | 7.144 | 0.0022 |
| Jan-2011 | 7.158 | 7.145 | 7.158 | 0.0007 | 7.148 | 7.158 | 0.0004 |
| Feb-2011 | 7.158 | 7.153 | 7.158 | 0.0001 | 7.151 | 7.158 | 0.0003 |
| Mar-2011 | 7.158 | 7.157 | 7.158 | 0.0000 | 7.157 | 7.158 | 0.0000 |
| Apr-2011 | 7.157 | 7.154 | 7.158 | 0.0001 | 7.155 | 7.158 | 0.0000 |
| May-2011 | 7.126 | 7.105 | 7.118 | 0.0029 | 7.119 | 7.121 | 0.0003 |
| Jun-2011 | 6.786 | 6.777 | 6.856 | 0.0962 | 6.872 | 6.889 | 0.0173 |
| Jul-2011 | 5.758 | 5.983 | 6.140 | 0.4925 | 6.185 | 6.208 | 0.1069 |
| Aug-2011 | 4.587 | 4.792 | 4.907 | 1.0733 | 4.872 | 4.867 | 0.3136 |
| Sep-2011 | 3.911 | 4.091 | 4.130 | 1.3271 | 3.984 | 3.981 | 0.3035 |
| Oct-2011 | 5.082 | 5.479 | 5.628 | 0.8492 | 5.150 | 5.181 | 0.2711 |
| Nov-2011 | 6.819 | 6.756 | 6.848 | 0.1338 | 6.808 | 6.845 | 0.0478 |
| Dec-2011 | 7.155 | 7.100 | 7.139 | 0.0097 | 7.124 | 7.142 | 0.0024 |

equation (6). The Fourier seasonal effects series included in representations for $\mu_t$ and the inflation probabilities $\alpha_{0t}$ have $J_1 = J_2 = 3$ harmonics. Predictions are obtained by generating replicate inflation indicators $D_{new,t} \sim Bern(\alpha_{0t})$, and replicate gamma values $w_{new,t} \sim Gamma(\phi_t, \phi_t/\mu_t)$. Predictions are then obtained as

$$y_{new,t} = 0 \times D_{new,t} + (1 - D_{new,t})w_{new,t}.$$

Table 6 summarises the L-criteria for the cross-validation period (when predictions are compared to test data) and for the in-sample estimation period (when predictions compared to training data). The deterministic trend has better performance (see Figure 6 for

**Table 3 Predictive fit (L-criterion) and predictive checks, sea ice extent models, central Arctic ocean**

| | Trend model | |
|---|---|---|
| | Stochastic | Deterministic |
| In-sample fit (1979-2009) | 2.77 | 3.96 |
| | Stochastic | Deterministic |
| Cross validation fit (2010-11) | 2.54 | 1.57 |
| **% of observations within 95% credible intervals of posterior predictions** | | |
| | Stochastic | Deterministic |
| In-sample period | 97.8 | 97.6 |
| Cross-validation period | 100.0 | 100.0 |

**Figure 5 Central Arctic (mn km$^2$), cross-validation period (2010-11) actual vs fitted, and out-of-sample predictions to end 2014.**

actual and fitted series). March extents (in mn $km^2$) are predicted to fall to 0.895 in 2012, 0.889 in 2013 and 0.879 in 2014 as compared to 0.964 in 2011. The strongest downward trend in mean extent as shown by the deterministic trend guide parameters $\{\Gamma_{0p}, \Gamma_{1p}\}$ is actually in the first sub-period (1979-1990). It may be noted that the actual data for the Sea of Okhotsk show a fall in annual average extent (mn $km^2$) from 0.58 in 1979 to 0.37 in 1990. The posterior means (sd's) for $\Gamma_{1p}$ are respectively $-0.016$ (0.010), $-0.0085$ (0.0044), and $-0.014$ (0.005).

For the three remaining sea regions (e.g. the Greenland Sea) with partial summer cover and no winter maximum recurrence, a gamma regression may be applied. In fact, like the Sea of Okhotsk, the Greenland Sea has declines in extent across both winter and summer months. For example, the average March extent (in mn $km^2$) fell from 0.98 in 1979-89 to 0.82 in 1990-2000 and 0.78 in 2001-11, while the average September extent fell from 0.37 (1979-89) to 0.34 (1990-2000) and 0.25 (2001-11). For this sub-region, a deterministic trend model has better in-sample fit and cross-validatory fit (see Figure 7). March extents are predicted as 0.767 in 2012, 0.766 in 2013 and 0.762 in 2014 as compared to 0.752 in 2011, and 0.764 in 2010. September extents are predicted as 0.294 in 2012, 0.290 in 2013 and 0.290 in 2014 as compared to 0.332 in 2011, and 0.264 in 2010.

### 4.4 Implications for prediction over entire Arctic

It is relevant to assess how far aggregated predictions (totalling over models for each of the nine regions) compare with predictions from a single model applied to the extent time series for the entire Arctic sea (i.e. to the data summarised in Table 1). The aggregated predictions are based on training data for 1979-2009 and combine

a) generalized beta regressions (with maximum inflation) applied to extent data for the Central Arctic, Canadian Archipelago, and Hudson Bay;

b) zero inflated gamma regressions applied to extent data for the Sea of Okhotsk, Bering Sea, and Gulf of St.Lawrence;

**Table 4 Deterministic trend parameters**

| | Posterior summary | | |
|---|---|---|---|
| Short term lag parameters | Mean | 2.5% | 97.5% |
| $\gamma_{21}$ | -0.475 | -1.971 | 0.833 |
| $\gamma_{22}$ | -0.455 | -2.074 | 0.960 |
| $\gamma_{23}$ | 0.344 | -0.310 | 0.947 |
| $\gamma_{24}$ | 0.166 | -0.227 | 0.668 |
| $\gamma_{25}$ | 0.823 | 0.305 | 1.303 |
| $\gamma_{26}$ | 0.516 | 0.191 | 0.939 |
| $\gamma_{27}$ | 0.907 | 0.348 | 1.821 |
| $\gamma_{28}$ | 2.004 | 1.453 | 2.431 |
| $\gamma_{29}$ | 3.222 | 2.580 | 3.926 |
| $\gamma_{2,10}$ | 2.999 | 1.552 | 4.394 |
| $\gamma_{2,11}$ | 1.110 | -0.186 | 2.390 |
| $\gamma_{2,12}$ | 0.545 | -1.110 | 1.875 |
| Linear trend guide parameters by period | Mean | 2.5% | 97.5% |
| $\Gamma_{11}$ | -0.028 | -0.048 | -0.007 |
| $\Gamma_{12}$ | -0.019 | -0.029 | -0.008 |
| $\Gamma_{13}$ | -0.028 | -0.036 | -0.019 |
| $\Gamma_{21}$ | -0.037 | -0.061 | -0.017 |
| $\Gamma_{22}$ | -0.025 | -0.036 | -0.016 |
| $\Gamma_{23}$ | -0.034 | -0.043 | -0.026 |
| Precision trends $\eta_{2m}$ | Mean | 2.5% | 97.5% |
| $\eta_{21}$ | 0.028 | -0.024 | 0.072 |
| $\eta_{22}$ | 0.092 | 0.033 | 0.134 |
| $\eta_{23}$ | 0.233 | 0.122 | 0.308 |
| $\eta_{24}$ | 0.068 | 0.011 | 0.113 |
| $\eta_{25}$ | 0.064 | 0.030 | 0.097 |
| $\eta_{26}$ | 0.000 | -0.034 | 0.031 |
| $\eta_{27}$ | -0.021 | -0.052 | 0.009 |
| $\eta_{28}$ | -0.023 | -0.058 | 0.009 |
| $\eta_{29}$ | 0.005 | -0.035 | 0.041 |
| $\eta_{2,10}$ | -0.017 | -0.059 | 0.016 |
| $\eta_{2,11}$ | -0.026 | -0.060 | 0.005 |
| $\eta_{2,12}$ | 0.001 | -0.033 | 0.031 |

**Table 5 Trends in extent by month, sea of Okhotsk**

| | Period | | |
|---|---|---|---|
| Month | 1979-89 | 1990-2000 | 2001-2011 |
| January | 0.91 | 0.73 | 0.73 |
| February | 1.22 | 1.03 | 1.05 |
| March | 1.24 | 1.10 | 1.06 |
| April | 0.91 | 0.73 | 0.71 |
| May | 0.34 | 0.27 | 0.29 |
| June | 0.11 | 0.11 | 0.10 |
| July | 0 | 0 | 0 |
| August | 0 | 0 | 0 |
| September | 0 | 0 | 0 |
| October | 0.05 | 0.05 | 0.04 |
| November | 0.08 | 0.08 | 0.06 |
| December | 0.35 | 0.36 | 0.30 |
| All Year | 0.44 | 0.37 | 0.36 |

**Table 6 Posterior predictive loss, sea ice extent model, sea of Okhotsk**

|  | Trend model | |
| --- | --- | --- |
|  | Stochastic | Deterministic |
| In-sample fit (1979-2009) | 2.95 | 2.43 |
|  | Stochastic | Deterministic |
| Cross validation fit (2010-11) | 1.11 | 0.68 |
| **% of observations within 95% credible intervals of posterior predictions** | | |
|  | Stochastic | Deterministic |
| In-sample period | 97.6 | 97.6 |
| Cross-validation period | 100.0 | 95.8 |

c) gamma regressions without inflation applied to extent data for the Baffin Sea, Greenland Sea, and Kara-Barents.

The deterministic trend model is adopted, and MCMC iterations 25,000-26,000 are cumulated over the nine models, providing posterior summary statistics (mean, variance, 2.5% percentile, 97.5% percentile) for cross-validatory predicted extents in 2010-11 across the entire Arctic region. By contrast, the single region approach (applied to the extent series for 1979-2009 encompassing the entire Arctic region) involves a gamma regression with deterministic trend - since the stochastic trend option provided much less precise predictions to 2010-11.

The comparative L-criteria for cross-validatory fit (over the 24 months in 2010-11) are 2.43 (aggregated predictions over sub-regions) and 3.70 (single region approach). Figure 8 shows the cross-validatory predictions (2010-11) and out-of-sample predictions (2012-14) for the combined forecast aggregating over sub-region models. March extents (in mn km$^2$) are predicted as 14.42 in 2012, 14.35 in 2013 and 14.31 in 2014, as compared to actual extents of 14.34 in 2011, and 14.88 in 2010. September extents (in mn km$^2$) are predicted as 4.97 in 2012, 4.84 in 2013 and 4.71 in 2014, as compared to actual totals of 4.60 in 2011, and 4.90 in 2010.

Although sub-region data are not (at the time of writing) available beyond 2011, totals for the entire Arctic are available from ftp://sidads.colorado.edu/DATASETS/ NOAA/G02135/. (Note that the latter differ slightly before 2012 from entire Arctic



**Figure 6 Sea of Okhotsk (mn km$^2$), cross-validation period (2010-11) actual vs fitted, and out-of-sample predictions to end 2014.**

**Figure 7 Greenland Sea (mn km$^2$), cross-validation period (2010-11) actual vs fitted, and out-of-sample predictions to end 2014.**
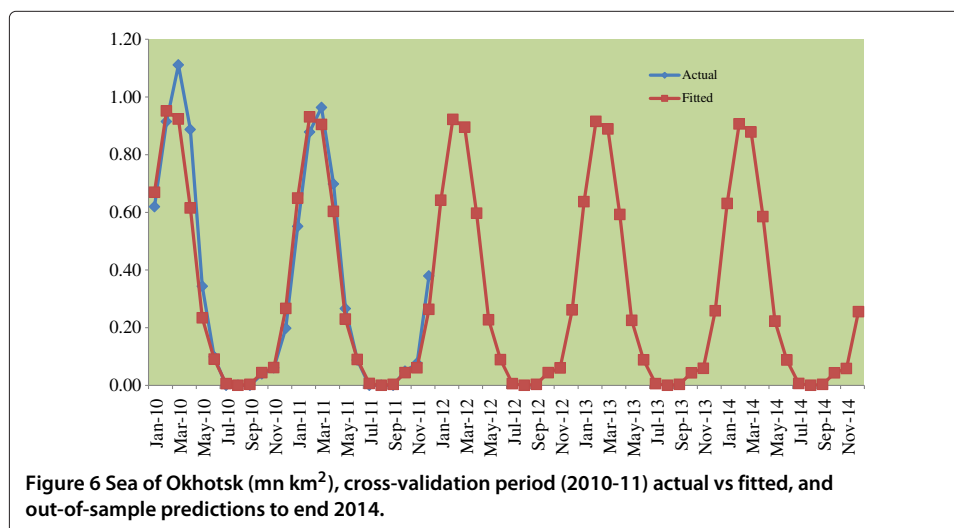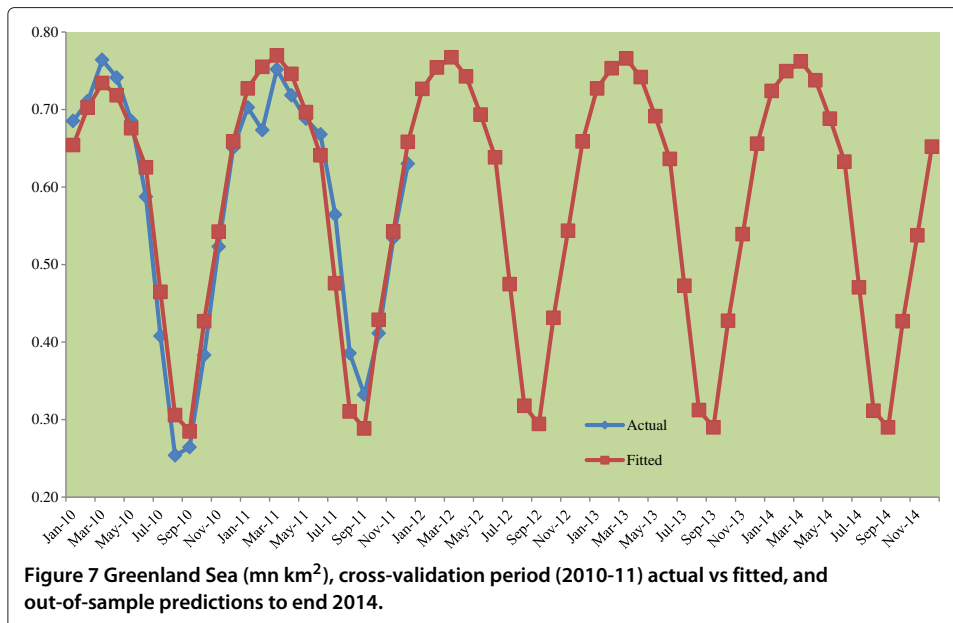
extents based on totalling the sub-regional series at http://neptune.gsfc.nasa.gov/csb/index.php?section=59). In fact, the March 2012 figure of 15.21 mn $km^2$ (NSIDC, 2012) was the highest March average ice extent since 2008. By contrast, the September 2012 figure was anomalously low at 3.61 mn $km^2$.

## 5    Summary and conclusions

This paper has considered sub-regional aspects of the observed downward trend in Arctic ice sea cover. Arctic sub-regions differ in observed trends and also in seasonal extremes. Thus some regions still have total winter cover but partial summer cover, while other regions have partial winter cover and no ice cover in summer. While downward trends



**Figure 8 Entire Arctic combined forest (mn km$^2$), cross-validation period (2010-11) actual vs fitted, and out-of-sample predictions to end 2014.**
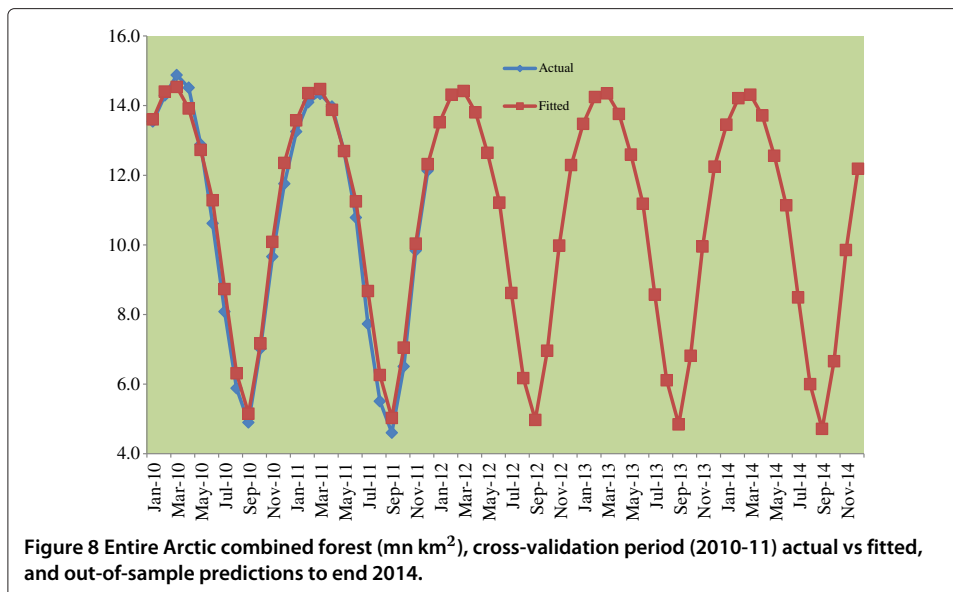
in sea ice extent across the Arctic as a whole show a stronger summer decline, this is not necessarily the case when sub-regions are considered.

These differences are important for choosing the appropriate distribution, and inflation mechanism. Sub-regional differences in ice loss may also be important for assessing relatively localized impacts on climate or economic activities (e.g. Fissel et al. 2011). There may also be benefits in prediction (see Section 4.4) through considering such sub-regional differences and in aggregating over region-specific models.

Here generalized beta densities (with maximum inflation) and gamma regression (with zero inflation) have been used to represent recurrent winter maxima and summer ice disappearance respectively. Other applications of this methodology may be envisaged, outside time series applications. Possible applications of the generalized inflated beta density are discussed in Section 2.3. It may be noted that transformations of ice extent such as $w_t = log(1 + y_t)$ might be envisaged as ways of avoiding beta or gamma regressions, and instead leading possibly to modelling using normal or skewnormal likelihoods. However, the problem of seasonal extreme inflation at particular values remains, and the methodology proposed in the paper provides a suitable representation for such extremes or for explaining them.

Another approach, as in compositional data analysis, is to focus on the ratios $r_t = y_t/d_t$ to the maximum, in the case when the data has two categories (e.g. area covered by sea ice $y_t$, remaining area $d_t - y_t$). Generally compositional data analysis involves Gaussian likelihoods applied to log-transformed ratios. This method can adjust for zero inflation (e.g. Butler and Glasbey, 2008), but does not generate direct inferences or out-of-sample predictions for the extent data themselves.

**Competing interests**
The authors declare that they have no competing interests.

**References**
Aitchison, J, Egozcue, J: Compositional data analysis: Where are we and where should we be heading? Math. Geol. **37**, 829-850 (2005)
Blundell, R, Griffith, R, Windmeijer, F: Individual effects and dynamics in count data models. Journal of Econometrics. **108**, 113–131 (2002)
Brooks, S, Gelman, A: General methods for monitoring convergence of iterative simulations. J. Comput. Graph.l Stat. **7**, 434–455 (1998)
Butler, A, Glasbey, C: A latent Gaussian model for compositional data with zeros. J. R. Stat. Soc. Series C. **57**, 505–520 (2008)
Comiso, J, Parkinson, C, Gersten, R, Stock, L: Accelerated decline in the Arctic sea ice cover. Geophys. Res. Lett. **35**, L01703 (2008). doi:10.1029/2007GL031972
Fissel, D, de Saavedra Álvarez, M, Kulan, N, Mudge, T, Marko, J: Long-term trends for Sea Ice in the Western Arctic Ocean: implications for shipping and offshore oil and gas activities. In: *Proceedings of the Twenty-first 2011 International Offshore and, Polar Engineering Conference*. Maui, Hawaii, USA, (2011). International Society of Offshore and Polar Engineers (ISOPE). http://www.isope.org/publications/ISOPEproceedinglist.htm
Gelfand, A: Model determination using sampling based methods. In: W. Gilks, S. Richardson, D. Spiegelhalter (eds.) Markov Chain Monte Carlo in Practice, Chapman and Hall/CRC, Boca Raton, pp. 145–157 (1996)
Höhle, M, Paul, M: Count data regression charts for the monitoring of surveillance time series. Comput. Stat. & Data Anal. **52**, 4357–4368 (2008)
Huang, X, Oosterlee, C: Generalized beta regression models for random loss-given-default. Department of Applied Mathematical Analysis, Delft University of Technology, Report 08–10 (2008)
Jung, R, Kukuk, M, Liesenfeld, R: Time series of count data: modeling, estimation and diagnostics. Comput. Stat. Data Anal. **51**, 2350–2364 (2006)
Laud, P, Ibrahim, J: Predictive model selection. J R Stat Soc. **57B**, 247–262 (1995)
Ledolter, J, Abraham, B: Parsimony and its importance in time series forecasting. Technometrics. **23**, 411–414 (1981)
Leininger, T, Gelfand, A, Allen, J, Silander, J: Spatial regression modeling for compositional data with many zeros. J. Agricultural Biol. Environ. Stat. **18**(3), 314–334 (2013)
Liu, J, Curry, J, Wang, H, Song, M, Horton, R: Impact of declining Arctic sea ice on winter snowfall. Proc. Nat. Acad. Sci. **109**, 4074–4079 (2012)

Lunn, D, Spiegelhalter, D, Thomas, A, Best, N: The BUGS project: Evolution, critique and future directions. Stat. Med. **28**, 3049–3067 (2009)

National Snow and Ice Data Center (2012). http://nsidc.org/arcticseaicenews/2012/04/arctic-sea-ice-enters-the-spring-melt-season/

Ospina, R, Ferrari, S: Inflated beta distributions. Stat. Papers. **51**, 111–126 (2010)

Pham-Gia, T, Duong, Q: The generalized beta- and F-distributions in statistical modelling. Math. Comput. Modell. **12**, 1613–1625 (1989)

Screen, J, Deser, C, Simmonds, I, Tomas, R: Atmospheric impacts of Arctic sea-ice loss 1979–2009: separating forced change from atmospheric internal variability. Climate Dynamics (2013). doi:10.1007/s00382-013-1830-9

Serreze, M, Maslanik, J, Key, J, Kokaly, R, Robinson, D: Diagnosis of the record minimum in arctic sea ice area during 1990 and associated snow cover extremes. Geophys. Res. Lett. **22**, 2183–2186 (1995)

Stroeve, J, Holland, M, Meier, W, Scambos, T, Serreze, M: Arctic sea ice decline: faster than forecast. Geophys. Res. Lett. **34**, L09501 (2007)

Stroeve, J, Serreze, M, Holland, M, Kay, J, Malanik, J, Barrett, A: The Arctic's rapidly shrinking sea ice cover: a research synthesis. Climatic Change. **110**, 1005–1027 (2012)

Zhao, X, Luo, Y, Wang, S, Huang, W, Lian, J: Is desertification reversion sustainable in northern China: a case study in Naiman county, part of a typical agro-pastoral transitional zone in Inner-Mongolia, China. Global Environ. Res. **14**, 63–70 (2010)