RMetS

Royal Meteorological Society

Probabilistic methods for seasonal forecasting in a changing climate: Cox-type regression models

Aline H. N. Maia[a]* and Holger Meinke[b]

[a] *Embrapa Meio Ambiente, Jaguariúna, São Paulo, Brazil*
[b] *Centre for Crop Systems Analysis, Wageningen University and Research Centre, Wageningen, The Netherlands*

ABSTRACT: For climate risk management, cumulative distribution functions (CDFs) are an important source of information. They are ideally suited to compare probabilistic forecasts of primary (e.g. rainfall) or secondary data (e.g. crop yields). Summarised as CDFs, such forecasts allow an easy quantitative assessment of possible, alternative actions. Although the degree of uncertainty associated with CDF estimation could influence decisions, such information is rarely provided. Hence, we propose Cox-type regression models (CRMs) as a statistical framework for making inferences on CDFs in climate science. CRMs were designed for modelling probability distributions rather than just mean or median values. This makes the approach appealing for risk assessments where probabilities of extremes are often more informative than central tendency measures. CRMs are semi-parametric approaches originally designed for modelling risks arising from time-to-event data. Here we extend this original concept to other positive variables of interest beyond the time domain. We also provide tools for estimating CDFs and surrounding uncertainty envelopes from empirical data. These statistical techniques intrinsically account for non-stationarities in time series that might be the result of climate change. This feature makes CRMs attractive candidates to investigate the feasibility of developing rigorous global circulation model (GCM)–CRM interfaces for provision of user-relevant forecasts. To demonstrate the applicability of CRMs, we present two examples for El Niño/Southern Oscillation (ENSO)-based forecasts: the onset date of the wet season (Cairns, Australia) and total wet season rainfall (Quixeramobim, Brazil). This study emphasises the methodological aspects of CRMs rather than discussing merits or limitations of the ENSO-based predictors. Copyright © 2009 Royal Meteorological Society

KEY WORDS   survival analysis; semi-parametric models; seasonal forecasts; ENSO; uncertainty assessment; climate change; probabilistic risk assessment

Received 9 February 2009; Revised 18 September 2009; Accepted 20 September 2009

1. Introduction

Managers of climate-sensitive industries can incorporate probabilistic forecasts into their decision-making as long as the associated uncertainties are clearly spelled out (Nelson *et al*., 2007). This is particularly true for agriculture and related sectors where proactive adaptation to climate risk is becoming increasingly important (Hansen *et al*., 2006; Meinke *et al*., 2009). Such operational climate risk management requires knowledge about the likely consequences of the future state of the climate systems. Often variables of interest (Y), such as time of onset of the wet season (Lo *et al*., 2007), rainfall, crop yields (Meinke *et al*., 1996) or return on investment (Twomlow *et al*., 2008) are provided as cumulative distribution functions (CDFs) $[P(Y \leq y)]$ or probability of exceedance functions (PEFs) $[P(Y > y)]$. Such probabilistic representation of decision variables helps risk managers to conduct rapid assessments of management options. CDFs/PEFs are particularly convenient to summarise time series that are not or only weakly auto-correlated (if time series are moderately to strongly auto-correlated, a CDF/PEF summary will result in the loss of some information). The decision variables in our examples (time to wet season onset and seasonal rainfall amounts) are at most weakly auto-correlated, thus allowing the CDF/PEF representation to convey seasonal climate information (Dunn, 2001; Maia *et al*., 2007). As there is a biunivoque correspondence between PEFs and CDFs, we use the terms interchangeably, depending on the context.

A range of statistical methods are used for estimating and comparing CDFs for climate risk assessment such as nonparametric approaches (Maia and Meinke, 1999; Stone *et al*., 2000; Hansen, 2005; Maia *et al*., 2007) or methods based on logistic approach (Nicholls *et al*., 1982; Nicholls, 1984) whereby smooth functions are composed from estimates obtained via logistic regression, a parametric method used for modelling probabilities of exceeding single thresholds (Lo *et al*., 2007).

Perfectly accurate CDFs cannot be obtained for noisy systems (Smith, 2000). Although it appears intuitively obvious that the degree of uncertainty that is associated with any estimated CDF could impact decision-making, such information usually is not provided. Yet, omitting

* Correspondence to: Aline H. N. Maia, Embrapa Meio Ambiente, PO Box 69, Jaguariúna, CEP: 13.820-000, São Paulo, Brazil.
E-mail: ahmaia@cnpma.embrapa.br

this uncertainty implies scientific precision that does not exist. The economist, John Maynard Keynes, supposedly captured this in his statement '*It is more important to be roughly right, than precisely wrong*' (Corazza, 1999). A colleague, therefore, often quips that probability curves should be drawn with a thick felt pen (Hayman, personal communication). Here we provide the statistical equivalent to such a felt pen, a method that provides confidence bands for probability estimates by accounting for underlying uncertainties.

CDFs are often used to summarise information in biomedical research, social sciences or engineering where the objective is modelling the time until the occurrence of a certain event (time-to-event variables) such as death (Finkelstein *et al*., 1993), divorce (Mazur and Michalek, 1998) or equipment/component failure (Elfaki *et al*., 2007).

Statistical approaches for making inferences about PEFs and associated risk measures [relative risks, hazard ratios (HRs)] are referred to as survival analysis, reliability analysis or event history analysis in the fields of medicine (Collett, 2003), engineering (Crowder *et al*., 1994) and social sciences (Yamaguchi, 1991), respectively. In spite of their wide use for biomedical risk assessments, survival analysis is rarely applied to climate risks (e.g. Maia *et al*., 1999; Maia *et al*., 2007). Recently, considerable improvements have been made in both classical (e.g. Therneau and Grambsch, 2000) and Bayesian survival analysis (Ibrahim *et al*., 2004), allowing for a high degree of flexibility. The family of survival analysis methods includes, for instance, (1) nonparametric approaches for estimating (Kaplan and Meier, 1958) and comparing PEFs (e.g. Log-rank test, Mantel, 1966); (2) parametric survival models based on theoretical underlying distributions (e.g. exponential, log-normal, Weibull, gamma, also known as accelerate time-to-failure models) and (3) semi-parametric models that do not require specification about underlying distributions for the time-to-event variable, referred to as Cox-type regression models (CRMs) (Kalbfleisch and Prentice, 1980; Lawless, 1982). CRMs comprise the classic CRM (Cox, 1972) and a wide range of generalisations, including Bayesian approaches for CRM estimation with potential application for construction of credibility intervals for CDFs (Ibrahim *et al*., 2004). For reasons of clarity, we restrict our examples to the simple, classical CRM approach as this satisfies our key objectives (see below). However, our analyses could easily be extended to account for non-proportional hazards (NPHs) or covariance structures in the data, should this be required.

For estimating risks and associated uncertainties in environmental sciences, there are, of course, a wide range of statistical techniques available that could be used routinely (e.g. quantile regression or neural networks, Koenker and Geling, 2001; Cawley *et al*., 2007; nonparametric conditional CDF kernel estimation, Li and Racine, 2008). We are not advocating the use of CRMs as a replacement for conventional techniques. Instead,

we argue that they could value-add under certain circumstances. The adequacy of any statistical method always depends on intrinsic characteristics of data structures, the problem domain and the intended applications. A comparative analysis between the Cox approach and any of these techniques is therefore extremely context specific and probably not very useful. Instead, we aim to advance the use of CRMs to estimate and compare CDFs as an additional tool for risk assessments and scenario development. Specifically, we

(1) introduce CRMs to climate scientists as a potential alternative to commonly used approaches;
(2) demonstrate the applicability of CRMs for statistical, PEF-based climate risk forecasting and suggest its use for downscaling outputs of global circulation models (GCMs,);
(3) provide methods for quantifying the degree of uncertainty surrounding probabilistic forecasts and
(4) extend the use of CRMs by replacing time-to-event variables with other quantities of interest (e.g. rainfall).

It is particularly the last objective we regard as a key innovation of our work. The principal tools that form the family of 'survival analysis methods' were developed for time-to-event data, yet they can equally be applied to other variables of interest. However, due to their heritage, extension beyond the time domain is still rare (e.g. Pajek and Kubala-Kukus, 2004). Their so far nearly exclusive use for time-to-event applications is one of the key obstacles for a wider knowledge, acceptance and application of this approach. Based on data from two locations (Cairns, Australia and Quixeramobim, Brazil), we outline the use and validity of CRMs as risk assessment tools in climate research and demonstrate an extension beyond the time domain by using CRM to estimate rainfall probabilities.

2. Background

Operational seasonal forecast systems based on the analogue years approach (AYA) are used globally to provide information for decision makers managing climate-sensitive systems (e.g. Sivakumar *et al*., 2000; Ferreyra *et al*., 2001; Selvaraju *et al*., 2004; Meinke and Stone, 2005). As long as AYA forecast systems are based on scientifically well-understood mechanisms, time series can be partitioned into 'year or season types' (analogue years) based on prevailing ocean and/or atmospheric conditions, resulting for instance, in Southern Oscillation Index (SOI) or El Niño/Southern Oscillation (ENSO) phases (e.g. Meinke and Hammer, 1997; Adams *et al*., 2003; Travasso *et al*.,2009). These time series are usually represented by their respective PEFs: a conditional PEF_K for each class k and an unconditional PEF (the baseline, also referred to as 'climatology'). Prevailing oceanic and atmospheric conditions can then be assigned to a particular category

k. The correspondent PEF$_K$ is then adopted for providing the quantitative basis for probabilistic assessments (Maia *et al*., 2007). In spite of its utility and robustness, the use of AYA has limitations, particularly when time series are short and/or there are many classes, which can lead to poor estimates. Although AYA provides critically important information for climate risk management, confidence bands for empirical PEFs arising from AYA are rarely reported. Therefore, Dunn (2001) proposed the use of bootstrap for providing confidence bands for quantiles of seasonal rainfall distribution, for given fixed probabilities of exceedance. However, analytical methods, based on nonparametric likelihood ratio, developed specifically for assessing uncertainties of quantiles of survival distributions (Li *et al*., 1996) could also be applied to rainfall or any other prognostic variable. The lack of information on PEF or quantiles' uncertainties in probabilistic climate forecast assessments contrasts with other domains (e.g. health care, maintenance and repair of components/equipments), where such uncertainty assessments are routinely provided (Finkelstein *et al*., 1993; De Lorgeril *et al*., 1999, Smith *et al*., 2001; Gibbons *et al*., 2002).

Most statistical software packages have special modules to perform survival analysis; for example, the procedures LIFETEST, LIFEREG and PHREG of the SAS/STAT® Software (SAS Institute Inc., 2004a) and the packages Survival, Survbayes and Coxrobust of the free software R (Language and environment for statistical computing, R Development Core Team, 2006). This greatly facilitates the operational assessment of risks and uncertainties, especially for spatial studies.

Based on theoretically founded likelihood-based methods, CRMs allow for estimating PEFs and their uncertainties, ranking and selecting multiple risk factors and quantifying their impacts on probabilistic outputs. These methods can also handle incomplete information, referred to as censored data (in the context of survival analysis, 'censored data' means that some units of observation have incomplete information regarding time-to-event, e.g. when the 'event of interest' such as death does not occur before the pre-established end of a study). Recently, some authors have proposed the use of Cox models as an innovative tool for modelling time-to-event variables in natural sciences. Anthony *et al*. (2007) used a CRM to assess the risk of coral mortality in response to temperature, light and sediment regime, whereas Angilletta *et al*. (2007) used the same tools for modelling thermal tolerance of leaf-cutter ants from São Paulo city (São Paulo, Brazil). Gienapp *et al*. (2005) and Visser and Both (2005) proposed the use of Cox approaches for predicting phenology of various populations under climate change scenarios. In the field of international relations, studies on factors influencing time to signature of treaties on carbon emissions (Fredriksson and Gaston, 2000) or duration of international conflicts (Box-Steffensmeier *et al*., 2003) also employed CRMs as a methodological reference.

The original Cox model assumes proportional hazards (PH), i.e. the absence of interaction between predictor and predictand. This constitutes the simplest Cox-type model

and will be hereafter referred to as CoxPH model. In the absence of an appropriate time-dependent covariate, Cox models simply assume an average effect over the range of observed data (Allison, 1995; Scheike and Zhang, 2003). However, a wide range of generalisations of the CoxPH model are available (Hess, 1994; Fisher and Lin, 1999; Quantin *et al*., 1999; Van Houwelingen and Eilers, 2000; Ahmed *et al*., 2007), allowing for adequate modelling of NPHs, if necessary.

In summary CRMs have some attractive properties for probabilistic seasonal climate forecasts such as the following:

(1) CRMs are designed for modelling probabilities instead of mean values. This makes the approach appealing for risk assessments where probabilities of extremes are often much more informative than central tendency measures alone (e.g. mean, median).
(2) CRMs do not require assumptions regarding the type of underlying probability distributions of the climate-related response variable being modelled (in contrast to, for instance, ordinary least squares multiple linear regression, logistic regression or parametric survival analysis).
(3) The validity of PHs assumption can be tested and, if needed, CoxPH models can be generalised to more flexible Cox-type NPH models.
(4) Estimates of probabilities of exceedance $[P(Y > y)]$ can be simultaneously obtained for multiple thresholds (y), via PEF modelling approaches, an advantage over the use of multiple logistic regressions (e.g. Lo *et al*., 2007).
(5) Quantiles of the target distribution are jointly estimated, in contrast to the quantile regression approach where individual quantile estimation provides an useful but only coarse approximation of the distribution shape (Cawley *et al*., 2007).
(6) The influence of many predictors on climate risks can be investigated simultaneously – the contribution of each potential predictor can objectively be evaluated via likelihood tests (Cox, 1975).
(7) Theoretically sound methods for assessing uncertainties surrounding PEFs are readily available (Kalbfleisch and Prentice, 1980).
(8) Censored data can be accommodated without the need to balance the data set and discarding potentially valuable information in the process (e.g. trimmed measures for representing central tendency, calculated after exclusion of extreme censored values).

Therefore, the use of CRMs in climate studies should be considered in line with other methods that are used routinely.

To demonstrate some potential applications, we use CRMs firstly for a probabilistic risk assessment of time to wet season onset in the Northern Australia (October–March) based on the state of well-established oceanographic [large-scale patterns of sea surface temperature (SST) anomalies] and atmospheric (average SOI)

indices prior to the commencement of the wet season. We then extend the method beyond its application to time-to-event data by assessing the probabilities of exceeding threshold values of rainfall amounts for the wet season in Northeastern Brazil (March–June) based on similar predictors (preceding average SOI values and tropical Pacific SST anomalies). Further, we provide associated uncertainties (confidence bands) for estimated PEFs that might guide decision makers in their choice between alternative decisions. To our knowledge, this study is the first using the CRMs to analyse the linkage among oceanic/atmospheric indexes and climate risks, thereby extending the methods to the domain of seasonal climate risk assessments.

3. Data and methods

The Cox-regression model, originally designed for modelling survival functions under the hypothesis of PHs, is given by

$$S(t; \; \boldsymbol{x_i}) = [S_0(t)]^{\exp(\boldsymbol{x_i}' \cdot \; \boldsymbol{\beta})} \tag{1}$$

where t is the time-to-event, $S(t, \; \boldsymbol{x_i})$ is the survival function for a particular set of predictor values, $S_0(t)$ is the baseline survival function, $\boldsymbol{x'_i}$ is the transposed vector of predictors (covariates) and $\boldsymbol{\beta}$ is the vector of unknown model parameters corresponding to each predictor variable. The derivative of $S(t; \; \boldsymbol{x'_i})$ with respect to t is the so-called hazard function $h(t; \; \boldsymbol{x'_i})$ that represents the instantaneous failure rate at each time t, as function of the covariates.

For our examples, we use the simplest member of the Cox-type family as an introduction to the nonparametric survival analysis approaches. Such a simple model is subject to restrictions as outlined by Koenker and Geling (2001); in the CoxPH model the covariates exert a pure location shift on the baseline hazard function, which corresponds to the underlying assumption that the covariates only affect the location and not the shape of the distribution of transformed survival times. However, these restrictions can be overcome by using generalisations of Cox models, once diagnostics show evidences against PHs assumption.

Here we investigate the influence of predictors based on oceanic/atmospheric phenomena such as the Southern Oscillation and warming/cooling of surface water in the Pacific basin (ENSO-based predictors).

(1) For Cairns, Northern Australia (16.93 °S, 145.78 °E), we quantify the influence of the state of the ocean/atmosphere continuum on time to onset of the monsoonal wet season, as suggested by Lo *et al.* (2007). The chosen predictors were the SOI (mean of JJA monthly SOI values) and the first rotated principal component (SST1; mean of JJA monthly SST1 values) of large-scale SST anomalies (Drosdowsky and Chambers, 2001).
(2) For Quixeramobim, Northeastern Brazil, we quantify the influence of two potential predictors, namely,

SST anomalies (October–February) in the Niño3.4 region (SST3.4, 5°N–5 °S; 170–120 °W) and mean SOI (October–February) on wet season (MAMJ) rainfall amounts at one location in the State of Ceará, Northeastern Brazil (Quixeramobim, 5.08 °S, 38.06 °W).

We quantify the influence of potential predictors on PEFs for the respective climatological variables of interest. Further, we quantify uncertainties associated with PEF estimates for some specific predictor values. Finally, we search for evidence of non-proportionality by including an interaction term as proposed by Allison (1995).

For both the examples, we used the PHREG procedure of the SAS/STAT® Software (SAS Institute Inc., 2004a) for fitting Cox models and estimating PEFs and associated uncertainties. To construct the graphs, we used the GPLOT Procedure of the SAS/GRAPH® Software (SAS Institute Inc., 2004b).

3.1. Examples

3.1.1. Example I: onset of wet season at Cairns, Northern Australia

For Cairns (16.93 °S, 145.78 °E), we used a high-quality, daily rainfall data set (Haylock and Nicholls, 2000) to calculate the time to wet season onset for each year (1948–2004). Two oceanic/atmospheric indices as described by Lo *et al.* (2007) were used as potential predictors: JJA SOI and JJA SST1. Following Lo *et al.* (2007), we defined the onset of wet season as the date at which 15% of the long-term mean of total summer rainfall (September–April) is first reached (after 1 September and before 31 March).

The SOI is calculated using the barometric pressure difference between Tahiti and Darwin. It serves as a useful atmospheric indicator of ENSO and can therefore be used to probabilistically forecast seasonal rainfall amounts for all locations influenced by ENSO (Stone *et al.*, 1996). We have adopted the SOI version from the Australian Bureau of Meteorology, where the commonly used SOI index is multiplied by ten. SOI data are available at http://www.bom.gov.au/climate/current/soihtm1.shtml.

The first rotated principal component of the large-scale pattern of SST anomalies (SST1) was derived from monthly Indo-Pacific SSTs (Drosdowsky and Chambers, 2001). The spatial pattern shows strong loadings in the Pacific indicating the strong influence of ENSO. The time series associated with SST1 correlates well with other typical ENSO indexes such as SSTs in the Niño3.4 region. Data are available at www.bom.gov.au/climate/ahead/sst_data_table.html.

3.1.2. Example II: monthly rainfall at Quixeramobim, Ceará, Northeastern Brazil

This part of Brazil is known to be strongly impacted by ENSO (Rao and Hada, 1990; Coelho *et al.*, 2002; Souza *et al.*, 2004). To investigate the impact of SST3.4

and SOI on wet season rainfall, we used monthly rainfall data (1951–2007) for Quixeramobim (5.2 °S, 39.5 °W) from Funceme (Ceará's state meteorological agency; www.funceme.br), the SST anomaly data set for Nino3.4 for the same period (available from NOAA at www.cpc.noaa.gov/data/indices/sstoi.indices) and the SOI series from BoM (www.bom.gov.au/climate/current/soihtm1.shtml).

3.2. Summary description of the Cox-regression model (CRM)

The term Cox regression refers to the combination of (1) the PHs model itself and (2) a method for model parameter estimation from empirical data known as the Cox maximum partial likelihood (Allison, 1995). The CRM assumes a parametric form for the effects of the predictors, but allows for an unspecified form for the underlying survivor function, i.e. a form that does not require the underlying probability density function of (in this case) the time-to-event to be specified (e.g. normal or gamma). For the onset analysis at Cairns, the time-to-event variable corresponds to time to onset (T days) and the survival function ($S(t) = P(T > t)|X = x$) corresponds to the probability of T exceeding a particular threshold $t (t > 0)$, in response to SOI or SST anomalies. In the case of Quixeramobim, the survival function describes the probability of the seasonal MAMJ rainfall (Y) exceeding a particular threshold amount (y, $y > 0$) in response to SOI or SST3.4 anomalies.

Similar to the Cox model described in Equation (1) for modelling survival functions, an extended CRM version for modelling PEFs in the context of climate risk assessments can be expressed as

$$\text{PEF}(y; \ \boldsymbol{x_i}) = [\text{PEF}_0(t)]^{\exp(\boldsymbol{x_i}' \cdot \ \boldsymbol{\beta})} \quad (2)$$

where y is the response variable, $\text{PEF}_0(y)$ is the baseline PEF (i.e. climatology), $\boldsymbol{x'_i} = (x_{i1}, \ x_{i2}, \ \dots x_{ij}, \ \dots x_{ik})$ is a vector of climate indexes used as predictors (covariates) and $\boldsymbol{\beta} = (\beta_1, \ \beta_2, \ \dots \ \beta_i, \dots \ \beta_k)'$ is the vector of unknown model parameters corresponding to each predictor variable (the baseline PEF, PEF_0 corresponds to the PEF in the absence of predictors; in climate science this baseline is generally referred to as 'climatology').

Here we adopt a terminology more appropriate for climate studies, by using $\text{PEF}(y, \boldsymbol{x_i})$ instead of the classical notation previously described for $S(t, \boldsymbol{x_i})$ (Equation (1)). Replacing the time variable $T = t$ by any other response variable $Y = y$ (as in the case of Quixeramobim, where rainfall is the predicant), we can therefore use the CRM for modelling PEFs for important climate-dependent variables like water stress, crop yields or even economic measures like farm income. For such variables, the hazard function [derivative of $\text{PEF}(y; \boldsymbol{x'_i})$ with respect to y] cannot be interpreted as an instantaneous failure rate, but the methods for estimating PEFs and associated uncertainties remain valid.

To estimate $\boldsymbol{\beta}$, Cox (1975) proposed a method based on partial likelihood that accounts for incomplete or censored data. In example I, y is time to onset (days), the elements of $\boldsymbol{x'_i}$ correspond to oceanic/atmospheric indexes used as candidate predictors (SOI and SST1) for explaining variability of onset PEFs and the elements of $\boldsymbol{\beta}$ are the model parameters that quantify the influence of each individual predictor on the PEF.

3.3. Diagnostics of non-proportionality

The assumption of proportionality of hazards of the CoxPH might not hold for climate-related CRMs. For Northern Australia, for example, Power *et al.* (2006) showed that the ENSO impacts on the rainfall amount during wet season are stronger during La Niña years, characterizing a non-proportional pattern. Such cases require more flexible CRMs that allow for adequate modelling of non-proportionality. For this purpose, an interaction term can be included into the model (e.g. the interaction between SOI and time to onset). This provides a simple method to generalise CPH models (Allison, 1995). Violation of the assumption of proportionality can be tested using a likelihood test. Should such a test reveal no evidence of interactions, we may conclude that the assumption of proportionality is not violated and disregard the covariate. If, on the other hand, the evidence is strong, Allison (1995) suggests incorporation of the interaction term into the model.

However, in many cases a simple interaction term might not be sufficient to model the real nature of non-proportionality; it might require more flexible survival models (Therneau and Grambsch, 2000) such as the so-called relative survival regression models (Giorgi *et al.*, 2003) and methods based on partition trees or polynomial splines for estimating time-varying model parameters (Hess, 1994; Intrator and Kooperberg, 1995; Xu and Adak, 2002). A comparison of several NPH modelling approaches can be found in Quantin *et al.* (1999); Van Houwelingen and Eilers (2000) and Ahmed *et al.* (2007).

Beyond conducting likelihood tests for the interaction terms, we also applied descriptive techniques for quantifying evidences of NPH (Allison, 1995; Hess, 1995) as follows.

(1) Plots of $g(t) = \log(-\log(\text{PEF}(t))$ *versus* time to onset: the predictor was categorised into three classes: C1 (SOI < -10), C2 ($-10 <$ SOI < 10) and C3 (SOI > 10). The corresponding $g(t)$ values for each class were plotted against time to onset (t). As explained in Allison (1995), under PHs, $h_a(t) = k. \ h_b \ (t)$ for any pair (a,b) of classes. This is equivalent to $\text{PEF}_a(t) = [h_b \ (t)]^k$. Taking the logarithm, multiplying by -1 and taking the logarithm again, we obtain $\log[-\log(\text{PEF}_a(t))] = \log(k) + \log[-log(\text{PEF}_b(t))]$, i.e. the lines $g(t)$ *versus* t for any pair of classes. These should be parallel if the assumption of proportionality holds.

(2) Plots of Schoenfeld (1982) residuals against time to onset: a randomly scattered pattern indicates no violation of the PHs assumption; a linear regression of

residuals on time to onset can be used for a descriptive assessment of trends. Low p-values associated with the t-test for the linear coefficient would then constitute evidence against the PH assumption.

Further guidelines on quantifying evidences against proportionality of hazards can be found in Hess (1994, 1995) and N'gandu (1997).

3.4. Confidence bands for probabilities of exceedance curves

Following Kalbfleisch and Prentice (1980), we assess uncertainties around PEFs using the following procedures: let PEF$(y; \varepsilon)$ correspond to the empirical cumulative hazard function estimate of the $P(Y > y|x_i)$ for $x_i = \varepsilon$. The standard error of $\log(\text{PEF}(y; \varepsilon))$ is approximated by $s_0(y; \varepsilon)$ which is the square root of the variance estimate of the hazard function. Based on the delta method, the standard error of PEF$(y; \varepsilon)$ is approximated by

$$s_1(y; \varepsilon) \approx \text{PEF}(y; \varepsilon) \, . s_0(y; \varepsilon) \qquad (3)$$

Let $z_{\alpha/2}$ be the upper $100.(1 - \alpha/2)$ percentile point of the standard normal distribution. By using a normal approximation, a $100(1 - \alpha/2)\%$ confidence interval for the 'true' unknown $P(Y > y|x_i = \varepsilon)$ is then given by

$$\text{PEF}(y; \varepsilon) \pm z_{\alpha/2} . s(y; \varepsilon) \qquad (4)$$

Alternative methods for calculating such confidence bands for survival functions (PEF) based on likelihood ratio are described by Murphy (1995) and Hollander *et al.* (1997).

4. Results and discussion

Our motivation for this work is simple – we call it 'interdisciplinarity in action'; we want to show that CRMs, which are commonly used in other domains of scientific risk assessment, can also play a useful role in climate research; a comprehensive review of alternative methods is not our goal. The examples discussed here were chosen to highlight a multitude of features that CRMs can account for. In the process, we extend the use of CRMs beyond their traditional domains of application. Although CRMs were initially designed for modelling time-to-event variables (in our case, time to wet season onset date at Cairns), they can readily be applied to account for other types of positive response variables (e.g. seasonal rainfall at Quixeramobim) of interest in probabilistic risk assessments. For instance, Pajek and Kubala-Kukus (2004) used survival analysis to model CDFs of concentrations of trace elements in human tissues. Such an approach allows assessments of the so-called non-detected concentrations by using the statistical concept of censoring. For operational climate risk management, CRMs can be extended to any variable of interest such as crop yields or economic value of a commodity derived from production models or rainfall from GCMs. Further, if the PH assumption does not hold, NPH models can be applied. Once an adequate survival model has been selected, PEF estimates can be provided for any combination of predictors and PEF confidence limits can be readily provided for any desired confidence level.

4.1. Example I: time to onset of the wet season, Cairns, Australia

We quantified the influence of SOI and SST on the time to wet season onset for Cairns, Australia, using CoxPH models (Figure 1). Estimated PEFs derived from models using SOI (model A, Figure 1(a)) or SST1 (model B, Figure 1(b)) as predictors clearly indicate the impact of ENSO on this variable as shown by Lo *et al.* (2007). We only show PEF estimates for some specific SOI (-15, 0 and $+15$) or SST anomaly values (-1.5, 0 and 1.5), but estimates can be obtained for any predictor value. Table I shows parameter estimates and respective results of likelihood tests for the CoxPH models used
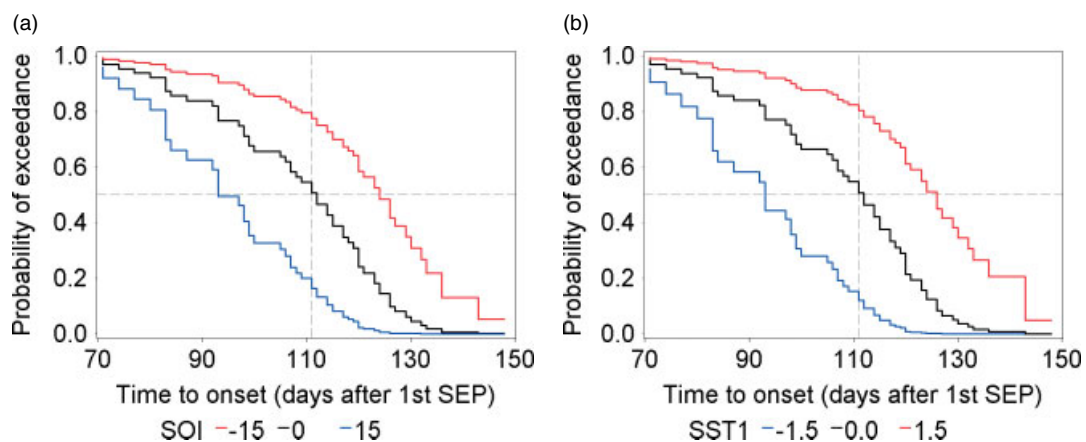

Figure 1. PEFs for time to wet season onset at Cairns, estimated using CoxPH models for selected values of both predictors: (a) June–August SOI and (b) June–August SST. Onset date was determined by accumulating 15% total, average wet season rainfall after 1 September. The vertical dashed lines indicate median time to onset (111 days after 1 September, i.e. 20 December); the solid black lines indicate empirical climatology (baseline PEF).

Table I. Parameter estimates of CPH models fitted for testing the influence of atmospheric/oceanic predictors (JJA SOI, model A1 and JJA SST1, model A2) on the PEFs for time to onset of the wet season at Cairns, Australia.

Models	Predictor	β Estimate (b)	Standard error	χ^2	Estimate of the hazard ratio ($\exp(b)$)	p-value*
A1	SOI	0.0652	0.01742	14.02	1.067	0.0002
A2	SST1	−0.7570	0.19868	14.51	0.469	0.0001

* Nominal significance levels arising from the maximum likelihood chi-square (χ^2) tests for the hypotheses $\beta_I = 0$ (no predictor influence).
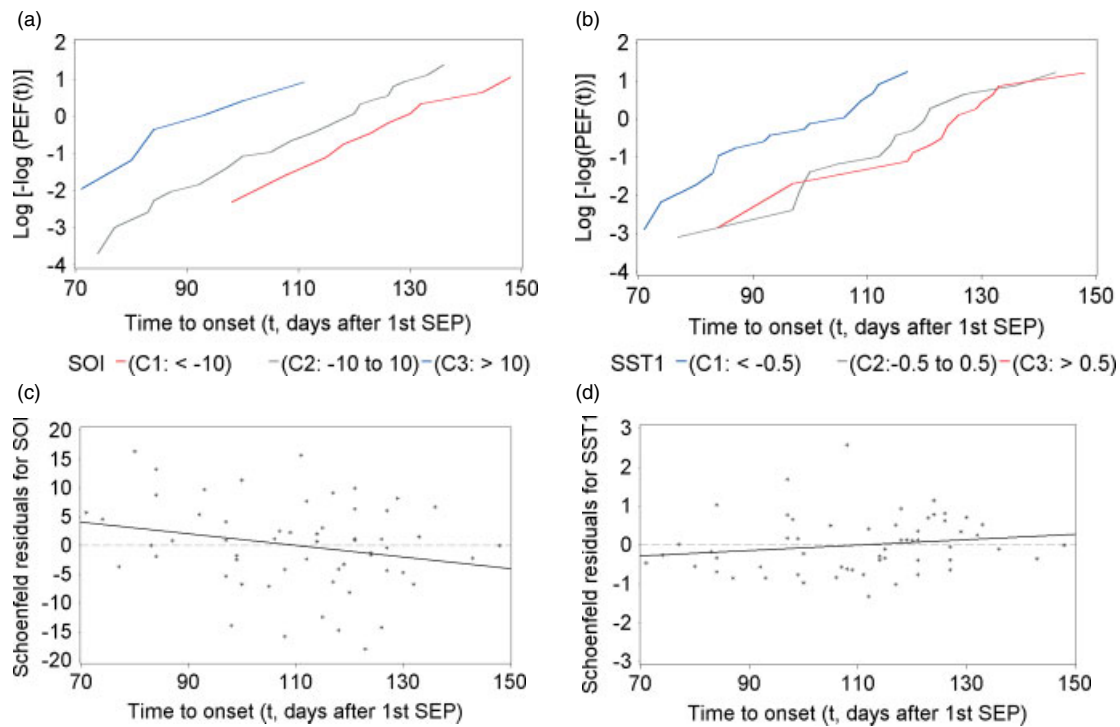

Figure 2. Diagnostics of non-proportionality. Parallelism of $\log(-\log(\text{PEF}(t))$ *versus* time to onset (t) for each class (C1, C2 and C3) of predictor values: (a) SOI as predictor (model A1) and (b) SST1 as predictor (model A2). Schoenfeld residuals *versus* time to onset, showing estimated linear trend (solid black line) (c) for model A1 and (d) for model A2.

to quantify the average (over time) influence of JJA SOI and SST1 on the probability of late wet season onset [PEF(t) = Prob($T > t$)]. Results indicate that the SST1 influence on time to onset, based on unitary changes in predictors, is stronger than SOI influence, with respective p-values of 0.0001 (model B) and 0.0002 (model A1).

Using the HR estimates ($\exp(b)$) from Table I, we objectively quantify the influence of predictors on PEFs for the climate-related variables. When HR estimate is greater than one (positive b), for each unitary increase in the predictor, the baseline PEF, $\text{PEF}_0(y)$ is powered to the corresponding HR, that is $\text{PEF}(t) = [\text{PEF}_0(t)]^{\exp(b)}$. As $\text{PEF}_0(t)$ is a value between 0 and 1, it results in a decrease in Prob($Y > y$).

In model A1, for each unitary increase in SOI, the baseline PEF, $\text{PEF}_0(t)$ is powered to 1.067 (Table I, positive $b = 0.0652$), resulting in a decrease in the probability of the time to onset exceeding a particular threshold t. Conversely, in model A2 (negative $b = -0.757$), unitary increases in SST1 leads to an

increase in the PEF(t). These results are consistent with the well-known influence of ENSO on North Australia, where La Niña conditions (indicated by positive SOI and negative SST anomalies, respectively) favour an earlier than normal start of wet season (Lo *et al.*, 2007).

In Figure 2, we show graphical diagnostics to investigate possible violations of PH assumption for both predictors (SOI and SST1), as suggested by Allison (1995) and Hess (1995). Tests based on the inclusion of interaction terms proposed in Allison (1995) (SOI*time for model A1 and SST1*time for Model A2) did not reveal strong evidence against the proportionality assumption of hazards: p-values arising from likelihood tests for the 'no interaction' hypothesis were 0.07 and 0.15 for SOI and SST1, respectively. Also, graphical diagnostics (Figure 2) also did not indicate strong violation of PH assumption.

t-Test for linear trends in Schoenfeld (1982) residuals resulted in p-values of 0.08 for SOI and 0.20 for SST1. When we applied the function $g(t) = \log(-\log(\text{PEF}(t))$
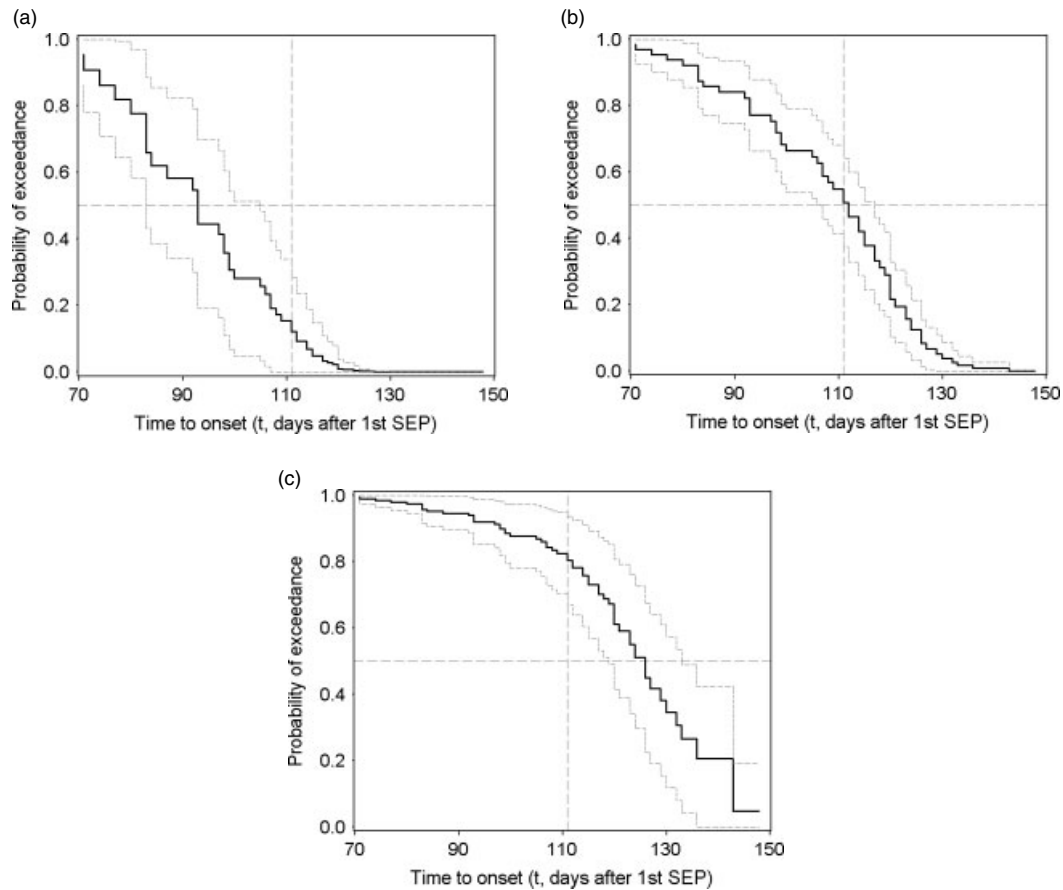
Figure 3. PEFs for time to wet season onset (days after 1 September) at Cairns, Northern Australia, with respective 95% confidence limits, for three selected JJA SST1 values: (a) SST1 = −1.5, (b) SST1 = 0 and (c) SST1 = 1.5. PEFs and associated uncertainties were estimated using CoxPH model. The vertical dashed lines indicate median time to onset based on climatology.

to three different predictor values, results revealed a tendency for higher predictor values exerting a stronger influence on time to onset in the value range corresponding to La Niña years (classes C3 and C1 for SOI and SST, respectively), a finding consistent with known non-proportional patterns of ENSO influence on rainfall (Power *et al*., 2006). However, for our objectives (to demonstrate the value of the overall general framework), in spite of moderate evidence against the PH assumption, we applied CPH models for quantifying an average multiplicative effect (over time) of the respective predictors on the instantaneous failure rates (HRs) (Allison, 1995; Scheike and Zhang, 2003).

So far, the PEFs in Figure 1 have no uncertainty estimates associated. In a final step we provide the 95% confidence bands for PEFs arising from model A2, based on Equations (3) and (4) (Figure 3). The width of the confidence limits depends on the series length and the signal strength of the predictor values at which the PEF was evaluated.

4.2. Example II: seasonal rainfall at Quixeramobim, Ceará, Northeastern Brazil

Wet season rainfall in the northeast of Brazil is impacted by ENSO, although other factors such as the gradient

mode of Atlantic SSTs are also important (Rao and Hada, 1990; Coelho *et al*., 2002; Souza *et al*., 2004). We use a very simple approach to quantify ENSO impact in this region by applying the Cox approach to model the seasonal rainfall PEF as function of SOI (model B1) and SST3.4 (model B2).

The likelihood test for non-proportionality did not reveal any evidence for the violation of the proportionality assumption for model B2 ($p = 0.2858$). For model B1, some evidence was found ($p = 0.0123$), but as already discussed, even when PH is violated, the CoxPH model can be applied and the results then interpreted as the average predictor's influence over the range of the response variable (Allison, 1985). Here we applied the CPH model to estimate PEFs for some specific SST3.4 values (Figure 4) and their associated uncertainties (Figure 5).

For model B1, the HR estimate ($\exp(b) = 0.968$) is lower than 1 (Table II, $b = -0.033$). Thus, unitary increases in SOI lead to higher probabilities of rainfall exceeding a threshold y. The opposite occurs for SST3.4 where b is positive (Table II, $b = 0.330$) and consequently, the HR estimate is higher than 1 ($\exp(b) = 1.391$). Again, these results are fully consistent with the well-known influences of ENSO on Northeastern Brazil, where El Niño-like conditions (positive SST3.4, negative
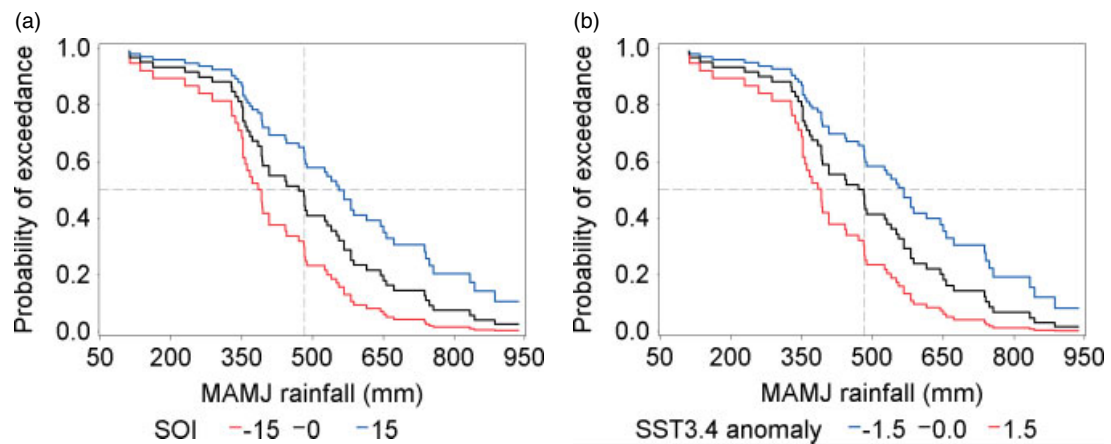
Figure 4. PEFs for seasonal rainfall (MAMJ) at Quixeramobim (Ceará, Brazil) estimated via CoxPH model using two alternative predictors: (a) October–February SOI or (b) October–February SST anomalies in Nino3.4 (SST3.4). The vertical dashed line indicates median the wet season rainfall (482 mm), the solid grey line indicates the empirical climatology (baseline PEF). Data from 1951 to 2007.
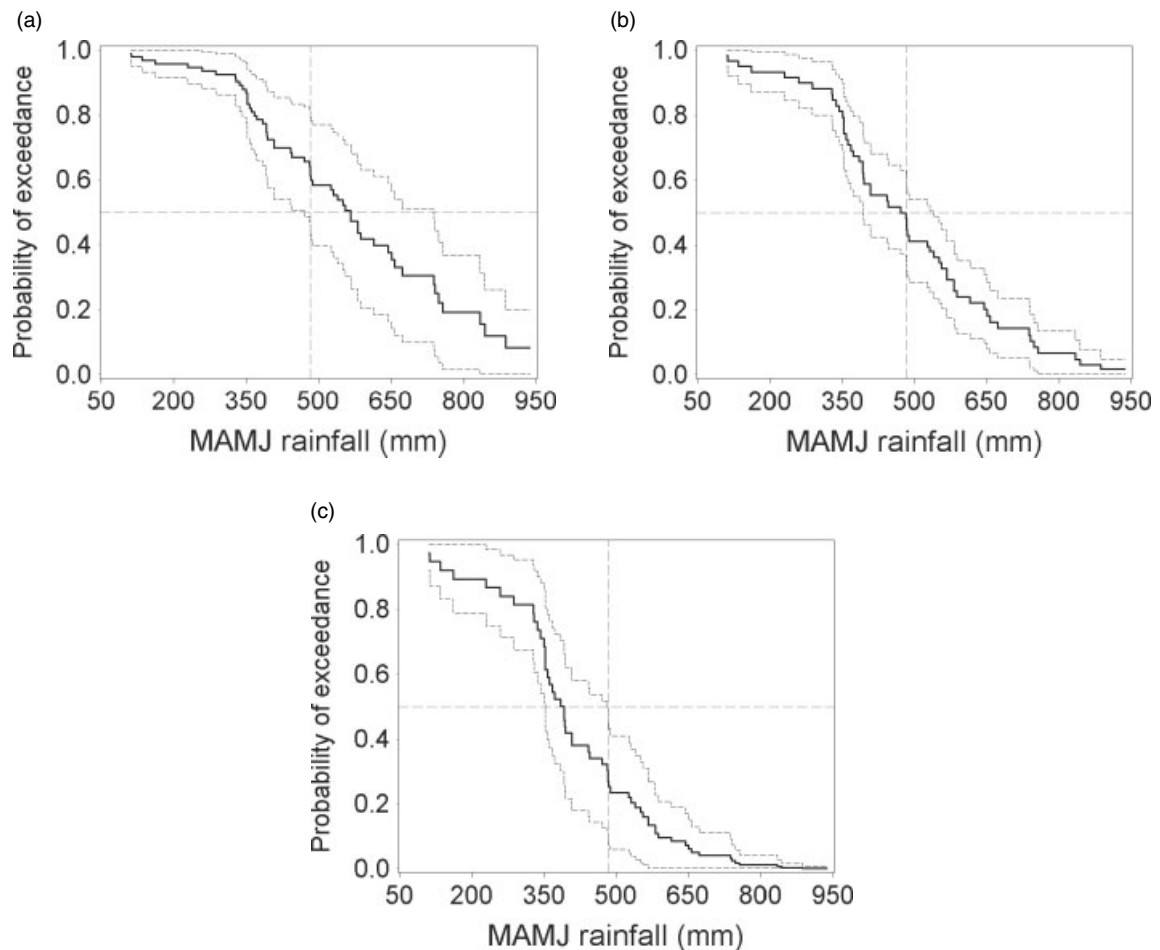

Figure 5. PEFs for seasonal rainfall (MAMJ) at Quixeramobim (Ceará, Brazil) with respective 95% confidence limits (dashed grey step functions), estimated via CPH models, for selected October–February SST3.4 values: (a) SST = −1.5, (b) SST = 0 and (c) SST = 1.5. The vertical dashed line indicates median seasonal rainfall (482 mm).

SOI) leads to low probability of above median rainfall during the wet season (Coelho *et al.*, 2002; Grimm, 2003).

Our results (Figure 5) confirm the known patterns of ENSO signal in the region, whereby at this particular location (Quixeramobim), during the La Niña event of 2008 (at the time of writing; March 2008) the probability of exceeding median rainfall (450 mm MAMJ total rainfall) is 0.7 based on five-monthly SST average of −1.7 (October–February).

Table II. Parameter estimates for CoxPH models used for quantifying the influence the October–February SOI (model B1) and October–February SST3.4 (model B2) on the PEFs for wet season rainfall at Quixeramobim, Ceará, Brazil.

Models	Predictor	β Estimate (b)	Standard error	Estimate of the hazard ratio ($\exp(\beta)$)	χ^2	p-value*
B1	SOI	−0.033	0.020	0.968	2.71	0.0999
B2	SST3.4	0.330	0.148	1.391	5.01	0.0253

* Nominal significance levels arising from chi-square (χ^2) likelihood test for the hypothesis $\beta_I = 0$ (no predictor influence).

Here we used the simplest model of the Cox-type family (assuming proportionality of hazards), which might not adequately reproduce the influence of some climate predictors on risks (see Section 3). This limitation could be overcome by using the more flexible NPH models that are able to capture nonlinear and disproportional influences of predictors.

CRMs intrinsically account for trends in empirical data. This means that non-stationarity in primary variables such as SSTs or SOI, which might be a consequence of climate change, are accounted for. This feature opens the possibility to develop GCM–CRM interfaces that could seamlessly connect with existing risk management by providing the necessary inputs for GCM-CRM connections could be achieved by using projections of oceanic/atmospheric predictors as input into CRMs resulting in estimated PEFs that link to biological models (Meinke and Stone, 2005).

When applied to gridded data, CRMs allow objective spatial assessment of either individual or joint influences of predictors on risks. Mapping the coefficients of the Cox model (which express the magnitude of the predictor's influence) and p-values resulting from likelihood tests provides a complete descriptive and inferential assessment of predictor influence on the risks under investigation (Maia *et al.*, 2007). Once the strongest predictors are selected, PEFs can be estimated for any combination of predictor values. The resulting risk estimates and their respective uncertainties can provide valuable information for decision-making in climate-sensitive industries.

Our examples are not intended to serve as rigorous seasonal climate forecasts; they were deliberately designed to serve as an easy introduction of CRMs to the statistical toolkit for climate research and applications. Yet, in spite of this simplicity, our results confirm current understanding of ENSO influences on rainfall and show how through the use of CRMs valuable information can be extracted from historical time series and provided to decision makers.

Acknowledgements

References

Adams RM, Houston LL, McCarl BA, Tiscareño LM, Matus GJ, Weiher RF. 2003. The benefits to Mexican agriculture of an El Niño-southern oscillation (ENSO) early warning system. *Agricultural and Forest Meteorology* **115**: 183–194, DOI: 10.1016/S0168-1923 (02) 00201-0.

Ahmed FE, Vos PW, Holbert D. 2007. Modeling survival in colon cancer: a methodological review. *Molecular Cancer* **6**: 15, DOI: 10.1186/1476-4598-6-15.

Allison PD. 1995. *Survival Analysis Using the SAS® System: A Practical Guide*. SAS Institute Inc.: Cary, USA; 292.

Angilletta MJ, Wilson RS, Niehaus AC, Sears MW, Navas CA, Ribeiro PL. 2007. Urban physiology: city ants possess high heat tolerance. *PLoS ONE* **2**(2): e258, DOI: 10.1371/journal.pone.0000258.

Anthony KRN, Connolly SR, Hoegh-Guldberg O. 2007. Bleaching, energetics, and coral mortality risk: Effects of temperature, light, and sediment regime. *Limnology and Oceanography* **52**: 716–726. [Available online at http://aslo.org/lo/toc/vol_52/issue_2/0716.pdf].

Box-Steffensmeier JM, Reiter D, Zorn C. 2003. Nonproportional hazards and event history Analysis in international relations. *Journal of Conflict Resolution* **47**: 33–53, DOI: 10.1177/0022002702239510. [Available online at http://www.jstor.org/stable/3176181].

Cawley GC, Janacek GJ, Haylock MR, Dorling SR. 2007. Predictive uncertainty in environmental modelling. *Neural Networks* **20**(4): 537–549.

Coelho CAS, Uvo CB, Ambrizzi T. 2002. Exploring the impacts of the tropical Pacific SST on precipitation patterns over South America during ENSO periods. *Theoretical and applied climatology* **71**: 185–197, DOI: 10.1007/s007040200004. [Available online at http://www.springerlink.com/content/7mn84kmuac0ngkkx/fulltext.pdf].

Collett D. 2003. *Modelling Survival Data in Medical Research*, 2nd edn. Chapman and Hall: London; 391.

Corazza M. 1999. Merton-like theoretical frame for fractional Brownian motion in finance. In *Current Topics in Quantitative Finance*, Canestrelli E (ed.). Physica Verlag: Heidelberg; 37.

Cox DR. 1972. Regression models and life-tables (with discussion). *Journal of the Royal Statistical Society* **34B**: 187–220. [Available online at: http://www.stat.rutgers.edu/~rebecka/Stat687/cox.pdf].

Cox DR. 1975. Partial likelihood. *Biometrika* **62**: 269–276. [Available online at http://biomet.oxfordjournals.org/cgi/reprint/62/2/269].

Crowder MJ, Kimber AC, Sweeting J. 1994. *Statistical Analysis of Reliability Data*. Chapman and Hall: London.

De Lorgeril M, Salen P, Martin J-L, Monjaud I, Delaye J, Mamelle N, Final Report of the Lyon Diet Heart Study. 1999. Mediterranean diet, traditional risk factors, and the rate of cardiovascular complications after myocardial infarction. *Circulation* **99**: 779–785. [Available online at http://circ.ahajournals.org/cgi/reprint/99/6/779].

Drosdowsky W, Chambers LE. 2001. Near-global sea surface temperature anomalies as predictors of Australian seasonal rainfall. *Journal of Climate* **14**: 1677–1687, DOI: 10.1175/1520-0442 (2001)014<1677:NACNGS>2.0.CO;2.

Dunn PK. 2001. Bootstrap confidence intervals for predicted rainfall quantiles. *International Journal of Climatology* **21**: 89–94, DOI: 10.1002/joc.596.

Elfaki FAM, Bin Daud I, Ibrahim NA, Abdullah MY, Usman M. 2007. Competing risks for reliability analysis using Cox's model. *Engineering Computations* **24**: 373–383, DOI: 10.1108/02644400710748698.

Ferreyra RA, Podestá GP, Messina CD, Letson D, Dardanelli J, Guevara E, Meira S. 2001. A linked-modeling framework to estimate maize production risk associated with ENSO-related climate variability in Argentina. *Agricultural and Forest Meteorology* **107**: 177–192, DOI: 10.1016/S0168-1923 (00) 00240-9.

Finkelstein DM, Moore DF, Schoenfeld DA. 1993. A proportional hazards model for truncated AIDS data. *Biometrics* **49**: 731–740. [Available online at http://www.jstor.org/stable/2532194].

Fisher LD, Lin DY. 1999. Time-dependent covariates in the Cox proportional-hazards regression model. *Annual Review of Public Health* **20**: 145–157, DOI: 10.1146/annurev.publhealth.20.1.145.

Fredriksson PG, Gaston N. 2000. Environmental governance in federal systems: the effects of capital competition and lobby groups. *Economic Inquiry* **38**: 501–514, DOI: 10.1111/j. 1465–7295.2000.tb00032.x.

Gibbons LE, Teri L, Logsdon R, McCurry SM, Kukull W, Bowen J, McCormick W, Larson E. 2002. Anxiety symptoms as predictors of nursing home placement in patients with Alzheimer's disease. *Journal of Clinical Geropsychology* **8**: 335–342, DOI: 10.1023/A:1019635525375.

Gienapp P, Hemerik L, Visser ME. 2005. A new statistical tool to predict phenology under climate change scenarios. *Global Change Biology* **11**: 600–606, DOI: 10.1111/j. 1365–2486.2005.00925.x.

Giorgi R, Abrahamowicz M, Quantin C, Bolard P, Esteve J, Gouvernet J, Faivre J. 2003. A relative survival regression model using B-spline functions to model non-proportional hazards. *Statistics in Medicine* **22**: 2767–2784, DOI: 10.1002/sim.1484.

Grimm AM. 2003. The El Niño impact on the summer monsoon in Brazil: regional processes versus remote influences. *Journal of Climate* **16**: 263–280, DOI: 10.1175/1520-0442(2003)016<0263: TENIOT>2.0.CO;2.

Hansen JW. 2005. Integrating seasonal climate prediction and agricultural models for insights into agricultural practice. *Philosophical Transactions of the Royal Society of London Series B: Biological Sciences* **360**: 2037–2047, DOI: 10.1098/rstb.2005.1747.

Hansen JW, Challinor A, Ines AVM, Wheeler T, Moron V. 2006. Translating climate forecasts into agricultural terms: advances and challenges. *Climate Research* **33**: 27–41, DOI: 10.3354/cr033027.

Haylock M, Nicholls N. 2000. Trends in extreme rainfall indices for an updated high quality data set for Australia, 1910–1998. *International Journal of Climatology* **20**: 1533–1541, DOI: 10.1002/1097-0088 (20001115)20:13<1533::AID-JOC586>3.0.CO;2-J.

Hess KR. 1994. Assessing time-by-covariate interactions in proportional hazards regression models using cubic spline functions. *Statistics in Medicine* **13**: 1045–1062, DOI: 10.1002/sim.4780131007.

Hess KR. 1995. Graphical methods for assessing violations of the proportional hazards assumption in cox regression. *Statistics in Medicine* **14**: 1707–1723, DOI: 10.1002/sim.4780141510.

Hollander M, McKeague IW, Yang J. 1997. Likelihood ratio-based confidence bands for survival functions. *Journal of the American Statistical Association* **92**(437): 215–226. [Available at: http://www.jstor.org/stable/2291466].

Ibrahim JG, Chen M-H, Sinha D. 2004. *Bayesian Survival Analysis*. Springer-Verlag: New York, USA; 479.

Intrator O, Kooperberg C. 1995. Trees and splines in survival analysis. *Statistical Methods in Medical Research* **4**: 237–261, DOI: 10.1177/096228029500400305.

Kalbfleisch JD, Prentice RL. 1980. *The Statistical Analysis of Failure Time Data*. John Wiley and Sons: New York, USA; 321.

Kaplan EL, Meier P. 1958. Nonparametric estimation from incomplete observations. *Journal of the American Statistical Association* **53**: 457–481.

Koenker R, Geling O. 2001. Reappraising medfly longevity: a quantile regression survival analysis. *Journal of the American Statistical Association* **96**(454): 458–468.

Lawless JE. 1982. *Statistical Models and Methods for Lifetime Data*. John Wiley and Sons: New York, USA; 580.

Li Q, Racine JS. 2008.. Nonparametric estimation of conditional CDF and quantile functions with mixed categorical and continuous data. *Journal of Business and Economic Statistics* **26**: 423–434, DOI: 10.1198/073500107000000250.

Li G, Tiwari RC, Wells MT. 1996. Quantile comparison functions in two-sample problems, with application to comparisons of diagnostic markers. *Journal of the American Statistical Association* **91**(434): 689–698.

Lo F, Wheeler M, Meinke H, Donald A. 2007. Probabilistic forecasts of the onset of the north Australian wet season. *Monthly Weather Review* **135**: 3506–3520, DOI: 10.1175/MWR3473.1.

Maia A, de HN, Meinke H. 1999. Non-parametrical survival analysis as a statistical tool in agricultural decision making. In *International Symposium, Modeling Cropping Systems, Proceedings of the European Society for Agronomy. Div. Agroclimatology and Agronomic Modeling*, Lleida, Spain, 21–23 June 1999; 103–104.

Maia AHN, Meinke H, Lennox S, Stone RC. 2007. Inferential, non-parametric statistics to assess quality of probabilistic forecast systems. *Monthly Weather Review* **135**: 351–362, DOI: 10.1175/MWR3291.1.

Mantel N. 1966. Evaluation of survival data and two new rank order statistics arising in its consideration. *Cancer Chemotherapy Reports* **50**: 163–170.

Mazur A, Michalek J. 1998. Marriage, divorce, and male testosterone. *Social Forces* **77**: 315–330. [Available online at http://cogprints.org/632/0/Joel.html].

Meinke H, Hammer GL. 1997. Forecasting regional crop production using SOI phases: a case study for the Australian peanut industry. *Australian Journal of Agricultural Research* **48**: 789–793, DOI: 10.1071/AR96155.

Meinke H, Howden SM, Struik PC, Nelson R, Rodriguez D, Chapman SC. 2009. Adaptation science for agricultural and natural resource management – Urgency and theoretical basis. *Current Opinion in Environmental Sustainability* **1**: 69–76, DOI: 10.1016/j.cosust.2009.07.007.

Meinke H, Stone RC. 2005. Seasonal and inter-annual climate forecasting: the new tool for increasing preparedness to climate variability and change in agricultural planning and operations. *Climatic Change* **70**: 221–253, DOI: 10.1007/s10584-005-5948-6.

Meinke H, Stone RC, Hammer GL. 1996. SOI phases and climatic risk to peanut production: a case study for Northern Australia. *International Journal of Climatology* **16**: 783–789, DOI: 10.1002/(SICI)1097-0088(199607)16:7<783::AID-JOC58>3.0.CO; 2-D.

Murphy SA. 1995. Likelihood ratio-based confidence intervals in survival analysis. *Journal of the American Statistical Association* **90**(432): 1399–1405. [Available at: http://www.jstor.org/stable/2291531].

Nelson R, Kokic P, Meinke H. 2007. From rainfall to farm incomes – transforming advice for Australian drought policy: Part II – forecasting farm incomes. *Australian Journal of Agricultural Research* **58**(10): 1004–1012, DOI: 10.1071/AR06195.

N'gandu NH. 1997. An empirical comparison of statistical tests for assessing the proportional hazards assumption of Cox's model. *Statistics in Medicine* **16**: 611–626, DOI: 10.1002/(SICI)1097-0258(19970330)16:6<611::AID-SIM437>3.0.CO;2-T.

Nicholls N. 1984. A system for predicting the onset of the north Australian wet-season. *Journal of Climatology* **4**: 425–435, DOI: 10.1002/joc.3370040407.

Nicholls N, McBride JL, Ormerod RJ. 1982. On predicting the onset of the Australian wet-season at Darwin. *Monthly Weather Review* **110**: 14–17. [Available online at: http://www.bom.gov.au/bmrc/clfor/cfstaff/nnn/pubs/onset.pdf].

Pajek M, Kubala-Kukus A. 2004. Censoring approach to the detection limits in X-ray fluorescence analysis. *Spectrochimica Acta Part B-Atomic Spectroscopy* **59**: 1739–1746, DOI: 10.1016/j.sab.2004.07.014.

Power S, Haylock M, Colman R, Wang X. 2006. The predictability of interdecadal changes in ENSO activity and ENSO teleconnections. *Journal of Climate* **19**: 4755–4771, DOI: 10.1175/JCLI3868.1.

Quantin C, Abrahamowicz M, Moreau T, Bartlett TG, MacKenzie T, Tazi MA, Lalonde L, Faivre J. 1999. Variation over time of the effects of prognostic factors in a population-based study of colon cancer: comparison of statistical models. *American Journal of Epidemiology* **150**: 1188–1200. [Available online at http://aje.oxfordjournals.org/cgi/reprint/150/11/1188].

Rao VB, Hada K. 1990. Characteristics of rainfall over Brazil: annual and variations and connections with the Southern Oscillation. *Theoretical and Applied Climatology* **42**: 81–91, DOI: 10.1007/BF00868215.

R Development Core Team. 2006. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing. [Available online at http://www.R-project.org].

SAS Institute Inc. 2004a. *SAS/STAT® User's Guide, Version 9.1, v. 1-7*. SAS Institute Inc.: Cary, NC; 5180.

SAS Institute Inc. 2004b. *SAS/GRAPH Software: Reference, Version 8*. SAS Institute Inc.: Cary, NC; 801–858.

Scheike TH, Zhang MJ. 2003. Extensions and applications of the Cox-Aalen survival model. *Biometrics* **59**: 1036–1045, DOI: 10.1111/j.0006-341X.2003.00119.x.

Schoenfeld D. 1982. Partial residuals for the proportional hazards regression model. *Biometrika* **69**: 239–241, DOI: 10.1093/biomet/69.1.239.

Selvaraju R, Meinke H, Hansen J. 2004. Climate information contributes to better water management of irrigated cropping systems in southern India. In *Proceedings Fourth International Crop Science Congress*. The Regional Institute Ltd.: Brisbane; CD-ROM, 2.6. [Available online at http://www.cropscience.org.au.].

Sivakumar MVK, Gommes R, Baier W. 2000. Agrometeorology and sustainable agriculture. *Agricultural and Forest Meteorology* **103**: 11–26, DOI: 10.1016/S0168-1923 (00) 00115-5.

Smith LA. 2000. Disentangling uncertainty and error: on the predictability of nonlinear systems. In *Nonlinear Dynamics and Statistics*, Mees A (ed.). Birkhauser: Boston, USA; 31–64.

Smith GE, O'Brien PC, Ivnik RJ, Kokmen A, Tangalos EG. 2001. Prospective analysis of risk factors for nursing home placement of dementia patients. *Neurology* **57**: 1467–1473. [Available online at: http://www.ncbi.nlm.nih.gov/pubmed/11673591].

Souza EB, Kayano MT, Ambrizzi T. 2004. The Eastern Amazon/Northeast Brazil regional precipitation in a weekly timescale modulated by tropical Pacific and Atlantic SST anomalies during austral autumn. *Revista Brasileira de Meteorologia* **19**: 113–122.

Stone RC, Hammer GL, Marcussen T. 1996. Prediction of global rainfall probabilities using phases of the Southern Oscillation index. *Nature* **384**: 252–255, DOI: 10.1038/384252a0.

Stone RC, Smith I, McIntosh P. 2000. Statistical methods for deriving seasonal climate forecasts from GCM's. In *Applications of Seasonal Climate Forecasting in Agricultural and Natural Ecosystems*, Hammer GL, Nichols N, Mitchell C (eds). Kluwer Academic: Dordrecht, The Netherlands; 135–147.

Therneau TM, Grambsch PM. 2000. *Modeling Survival Data: Extending the Cox Model*. Springer: New York; 350.

Travasso MI, Magrin GO, Grondona MO, Rodriguez GR. 2009. The use of SST and SOI anomalies as indicators of crop yield variability. *International Journal of Climatology* **29**: 23–29, DOI: 10.1002/joc.1701.

Twomlow S, Mugabe FT, Mwale M, Delve R, Nanja D, Carberry P, Howden M. 2008. Building adaptive capacity to cope with increasing vulnerability due to climatic change in Africa – A new approach. *Physics and Chemistry of the Earth* **33**: 780–787, DOI: 10.1016/j.pce.2008.06.048.

Van Houwelingen HC, Eilers PHC. 2000. Non-proportional hazards models in survival analysis, COMPSTAT 2000. In *Proceedings in Computational Statistics*, Bethlehem JG, Van der Heijden PGM (eds). Physica-Verlag: Heidelberg, Germany; 151–160.

Visser ME, Both C. 2005. Shifts in phenology due to global climate change: the need for a yardstick. *Philosophical Transactions of the Royal Society of London Series B: Biological Sciences* **272**: 2561–2569, DOI: 10.1098/rspb.2005.3356.

Xu R, Adak S. 2002. Survival analysis with time-varying regression effects using a tree-based approach. *Biometrics* **58**: 305–315, DOI: 10.1111/j.0006-341X.2002.00305.

Yamaguchi K. 1991. Event history analysis. *Applied Social Research Methods*, Vol. 28. Sage Publications: Newbury Park, USA.