

## Manuscript # bg-2015-296: Response to referees

We thank the three referees for taking time to review our manuscript "*Marine regime shifts in ocean biogeochemical models: a case study in the Gulf of Alaska*" and for providing thoughtful and thorough comments. We agree with the referees' main suggestions, and have revised the manuscript to address those points.

Referees comments are in bold and our responses in plain text.

### Response to anonymous referee 1:

**This paper explores the response of five different ocean biogeochemical models averaged over the Gulf of Alaska (GoA) to the same physical forcing over the period 1950-2007. Time series of physical (SST and MLD) chemical (DIN, SI, FE) and biological (surface chlorophyll, integrated primary productivity, and surface phytoplankton and zooplankton biomass) from the models are analyzed using a change point detection scheme. The method consists of a set of regression models that can classify the time series as a constant mean, mean shift, trend, shift in the intercept of the trend and shift in both the intercept and trend. The method is able to detect one change throughout the time series. A downward trend in GoA SST is identified prior to 1976 followed by a weak upward trend afterwards with a slight upward trend in MLD over the period. Most of the simulated biogeochemistry time series indicate a change around 1976 but show a mix of behavior with the simpler models exhibiting more regime-like behavior than the more complex models. A comparison of how different ocean biogeochemical models with the same physical state simulate time series of key quantities in different parts of the globe, including the Gulf of Alaska, is a useful endeavor. The same can be said for a fairly rigorous evaluation of change points in these time series. Thus I accept this paper for publication in Biogeosciences, but I think the manuscript can be improved in several ways and thus I recommend a major revision.**

We thank the referee for this positive view of our manuscript and for his/her constructive suggestions, which significantly helped improve our manuscript.

1) While regime analysis has become very popular especially in climate and marine ecosystem analysis, it may lead to a misinterpretation of the underlying dynamics of a system, especially for relatively short time series. For example, regimes are often linked to phases of the Pacific Decadal Oscillation (PDO). The transition in the North Pacific around 1976-77 has been linked to a change in the PDO (which extends into the tropics), while the one around 1998 is more associated with the second EOF sometimes termed the North Pacific Gyre Oscillation (NPGO, also relevant for the discussion in the Introduction, page 5, lines 22-33). Rather than regimes these just might be periods where one pattern is more prevalent than another, where both of these patterns impact the GoA. In addition, evaluating the time series as single AR1 process (for the model of the mean and no change) may not be the best null hypothesis. Anomalies in the state of the GoA are strongly influenced by ENSO and other factors. If these processes fluctuate, they could cause rapid changes in the remote time series even if these processes are linear and add to each other. The authors should discuss these

**complicating factors (or perhaps even try and incorporate them as one of their models. See the following papers:**

**Bond, N. A., J. E. Overland, M. Spillane, and P. Stabeno (2003), Recent shifts in the state of the North Pacific, *Geophys. Res. Lett.*, 30(23), 2183, doi:10.1029/2003GL018597. B**

**Di Lorenzo, E. et al. North Pacific Gyre Oscillation links ocean climate and ecosystem change. *Geophys. Res. Lett.* 35, L08607 (2008).**

**Matthew Newman, 2007: Interannual to Decadal Predictability of Tropical and North Pacific Sea Surface Temperatures. *J. Climate*, 20, 2333–2356. doi: <http://dx.doi.org/10.1175/JCLI4165.1>**

**Newman, M., G. P. Compo, M. A. Alexander, 2003: ENSO-forced variability of the**

**Pacific Decadal Oscillation. *J. Climate*, 16, 3853-3857.**

**Schneider, N., and B. D. Cornuelle (2005), The forcing of the Pacific Decadal Oscillation, *J. Clim.*, 18, 4355 – 4373.**

We highly appreciate this suggestion. We added a paragraph in the discussion and conclusions section where we discuss potential contribution of the PDO, NPGO and ENSO to the underlying dynamics based on the papers suggested above. If both the PDO and NPGO fluctuations drive changes in the North Pacific climate and ecosystem functioning (and implicitly ENSO which influences the PDO), the question arises whether either of these indices exhibit a shift at a similar time. We have included these three large-scale oscillations in our analysis and verify whether they also exhibit a change-point in the late 1970s (Fig. 3, presented at the end of our response). As mentioned in previous work, the PDO index exhibits a significant shift in 1976/77, but we find no significant shifts in the multivariate ENSO index or the NPGO index. Clearly, by detecting a shift in the late 1970s in PDO only we cannot conclusively tie the PDO and untie the NPGO and ENSO to the shift in climate and ecosystem dynamics of the Gulf of Alaska, but this provides useful information for future work aiming at determining causal mechanisms, teleconnections and physical/biogeochemical dynamics linking global climate patterns to ocean productivity in this region.

As for using an AR(1) as a null model, we think that it is an appropriate null hypothesis at the annual time scale and should act as a parameter which roughly comprises these complex factors. The need to add an autoregressive model to a statistical model usually indicates some external factors have not been taken into account in the fit. Here, we summarized these external factors through an AR(1), as our objective is not to attribute specific external factors influencing physical/biogeochemical variables in the Gulf of Alaska, but rather assess whether they exhibit significant abrupt changes. This point is now clarified in the methodology section.

**2) How would the method used here classify a pure sine wave with (one) zero value some where in the time series? Would it classify the zero crossing as a change point or regime shift? (Same goes for the changes in the “trends” when the amplitude of the waves switches sign.) Clearly the dynamics behind an oscillatory signal would likely be quite different than the dynamics for a regime shift or trend.**

Great question. Yes, the dynamics behind an oscillatory signal would be different than the dynamics for a regime shift or trend. However, given that the models take into account the autocorrelation, we think the oscillatory signal would be interpreted as temporal correlation, which the autocorrelation component should pick up.

**3) The authors should probably reference other modeling studies of decadal physical and biogeochemical changes in the northeast Pacific:**

**Alexander, M., A. Capotondi, A. Miller, F. Chai, R. Brodeur and C. Deser, 2008: Decadal variability in the Northeast Pacific in a physical-ecosystem model: The role of mixed layer depth and trophic interactions. *Journal of Geophysical Research - Oceans*, 113, C02017, doi:10.1029/2007JC004359.**

**Capotondi, A., M. A. Alexander, C. Deser, and A. J. Miller (2005), Low frequency pycnocline variability in the northeast Pacific, *J. Phys. Oceanogr.*, 35, 1403 – 1420**

**Haigh, S. P., K. L. Denman, and W. W. Hsieh (2001), Simulation of the planktonic ecosystem response to pre- and post-1976 forcing in an isopycnic model of the North**

**Pacific, *Can. J. Fish. Aquat. Sci.*, 58, 703 – 722.**

We thank the referee for bringing these studies to our attention. We have included a new paragraph in the discussion and conclusions section in which we compare our results to these modelling studies. We feel this suggestion really improved our discussion. Our results are in agreement with Haigh et al. (2001), who are suggesting a year-round deepening of the mixed layer depth in the Gulf of Alaska after 1976, which led to a decrease in nutrient, phytoplankton and zooplankton after 1976. Alexander et al. (2008) rather simulate a shoaling in the winter mixed layer depth in the late 1970s, giving rise to a early-spring increase in primary production, phytoplankton and zooplankton biomass followed by a late-spring decline in both phytoplankton and zooplankton biomass. Despite the caveat that we are analysing annual mean time series it is important to point out the contradictory direction of change in mixed layer depth. Possibly reconciling this discrepancy, Capotondi et al. (2005) suggest a deepening trend in a broad band along the coast and shoaling in the central part of the Gulf of Alaska. Thus, the comparison of the various attempts to simulate the late 1970s regime shift of the Gulf of Alaska raises the possibility that the abrupt and spatially coherent ecosystem change is actually caused by a previously unappreciated heterogeneous set of environmental changes with distinct spatial pattern and timing in the annual cycle. Further analysis would be required to investigate changes at the seasonal scale and at a finer spatial resolution and is beyond the scope of this study.

**4) The reference for Wunsch 1999 on page 4 line 20 is missing from the reference list.**

Reference added – we thank the referee for pointing this out.

**5) The authors should probably include a discussion of the physical model simulation here even if it is described in other papers and/or on line. Are the BGC models driven offline where the ocean model (NEMO) is run first and then the values are fed into the BGC models? Note this does not allow for feedback of the biology on the physics (e.g. changes in solar absorption by phytoplankton). If**

**the surface sensible and latent heat flux are computed using the observed air temperature and model SST, the results will be to strongly relax the model SST towards the observed SST (as the observed air temperature & SST are highly correlated. If this is the case then the model will likely obtain the correct SST regardless of if it is a good simulation or not, e.g. see: Seager, R., Y. Kushnir, and M. A. Cane (1995), On heat flux boundary conditions for ocean models, *J. Phys. Oceanogr.*, 25, 3219 – 3230.**

We thank the reviewer for this suggestion that clarifies the simulations that were used. The simulations were "online", in that physics and biogeochemistry were both formally simulated simultaneously. Feedbacks between the model biology and ocean physics (e.g. by the absorption of downwelling solar radiation) were disabled so that all of the biogeochemical models experienced consistent simulated physics. In addition, sea surface salinity was weakly relaxed (characteristic timescale of 180 days) towards observations to minimise drift, while model SST was not relaxed to observed SST. These details have been added to section 2.2.

**5) During late winter in the subarctic North Pacific the mixed layer extends to the upper portion of the halocline, located between depths of approximately 70 and 120 m (Roden, 1964; Freeland et al., 1997; de Boyer Montegut et al., 2004) and the MLD is mainly controlled by salinity not by temperature (this would have impacted the Polovina et al. [1995] finding). Low-frequency changes in the Ekman pumping in the Gulf of Alaska, which vertically displaces the halocline, may impact the wintertime MLD by moving a layer with strong density gradients toward or away from the surface. After the mid-1970s the pycnocline was shallower in the central part of the Gulf of Alaska and deeper in a broad band along the coast, primarily due to the local response to Ekman pumping (Cummins and Lagerloef, 2002; Capotondi et al., 2005). This impact should be included using a density definition for MLD although a change in MLD about 1976-77 seen in other studies is not found here.**

**de Boyer Montegut, C., G. Madec, A. S. Fischer, A. Lazar, and D. Iudicone (2004), Mixed layer depth over the global ocean: An examination of profile data and a profilebased climatology, *J. Geophys. Res.*, 109, C12003, doi:10.1029/2004JC002378.**

**Capotondi, A., M. A. Alexander, C. Deser, and A. J. Miller (2005), Low frequency pycnocline variability in the northeast Pacific, *J. Phys. Oceanogr.*, 35, 1403 – 1420. Cummins, P. F., and G. S. Lagerloef (2002), Low frequency pycnocline depth variability at station P in the northeast Pacific, *J. Phys. Oceanogr.*, 32, 3207 – 3215.**

**Freeland, H., K. Denman, C. S. Wong, F. Whitney, and R. Jacques (1997), Evidence of change in the winter mixed layer in the northeast Pacific Ocean, *Deep Sea Res., Part I*, 44, 2117 – 2129.**

**Roden, G. I. (1964), Shallow temperature inversions in the Pacific Ocean, *J. Geophys. Res.*, 69, 2899 – 2914.**

We indeed used a density dependent mixed layer depth, which should take account of these effects. However, we cannot separate central and coast MLD as we use a basin-wide mean time series of MLD (shallower in the middle and deeper along the coast). We added a section in the discussion and conclusion to highlight this difference. Please also see response to comment 3.

**7) Page 15. Paragraph 12-20. The authors indicate that several of the models depict a regime shift in the Gulf of Alaska in late 1980s (instead of the mid 1970s) and that this shift seems to be mainly forced by changes in MLD. However, the change detection method and Figs. 2-4 do not appear to show much of a change in MLD around 1989 either in the full time series or in the PCs.**

We agree and have now removed discussion about a late 1980s shift as this is not detected in the individual time series neither in the PC1 of the five OBGC models. We feel the manuscript is more focused now.

**8) Give the correlation values for the curves in Fig. 7 & 8. Are the correlation values during the different epochs significantly different from each other (rather than significant – i.e. different than zero). The lines in Figs. 7 do not seem significantly different from each other, especially given the large spread (see 7 above).**

In the revised version of the manuscript, previous Figs. 7-8 have been combined into one figure for less repetition and more efficient use of space (Fig. 6, presented below). The corresponding slope values are presented in Table 5, which we now also mention in the figure caption for clarity. We are actually testing whether the relationships before and after the shift are different from each other, not whether they are different from zero. They are mostly not significantly different (see Table 5, we pasted it at the end of our response).

**9) Bottom of page 16 top of 17 (also in the abstract). The authors indicate that all models simulate a decrease in nutrients and biological productivity after 1976. Perhaps, this statement is based on Fig. 3; however, an examination of Figs. A1-A5 indicates more complex behavior. For example, the change point analysis suggests a downward trend for PHY & ZOO for the DiatHadOCC and PlankTOM10 models and an upward trend in FE in the ERSEM model over the entire record.**

We reworded this part to reflect that this statement is true only for the models simulating the shift: “A shift in model SST occurred in 1976 and matched a shift in observed SST. This abrupt change was accompanied by a smooth deepening of the mixed layer depth followed by an overall decrease in nutrients and productivity. The three OBGC models simulating a shift in 1977 are consistent in the direction of change, but the abruptness of the change varies among them (Fig. 5).” We also made this distinction in the abstract.

**10) Bottom of page 17 top of 18. A point of clarification . . . “20th century” simulations from the CMIP5 archive (the simulations that are referenced the most from the archive) do not produce a climate shift in the mid 1970s. These models are initialized in the mid 19th century and due to chaotic interactions values during a given time in the model do not directly correspond to those in nature (although the idea is that the models have the correct sensitivity to climate change and have the basic statistics of climate variability correct.) The Meehl and Teng studies (including the one referenced here) are based on initialized hindcast model runs just within a few years (up to 10) prior to the**

**period examined.**

We agree that this part of the discussion was overoptimistic and vague and decided to remove this paragraph from the discussion section in the revised version of the manuscript.

**11) The authors note that the simpler models tend to produce more regime-like behavior. Are there references from other fields, e.g. systems theory, which can support this from a more general perspective?**

See response to comment 12 below.

**12) While the authors note that the models produce different change points, they don't comment much on the difference between models. Indeed one is struck by how different the simulations are especially given that the physical forcing is identical. What does this say about the state of ocean BGC modeling? Are there observations say at OWS P that could support one model over another? Are the BGC models highly nonlinear in that one should perform multiple ensembles (based on different initial BGC conditions) as is done for climate system forecasts – i.e. one would get different results from individual ensemble members. If the model results are so different, does that suggest caution in using change point analyses?**

We thank the reviewer for this comment, which helped improve the discussion. The performance of the models in terms of a fit to observations has been assessed globally and was published in Kwiatkowski et al. (2014). All models showed skills in simulating some variables, but simpler models were broadly closer to observations overall. This was added to the discussion section.

As for the optimal level of complexity, this is an unsettled question in the field of marine ecosystem modelling (e.g. Allen et al., 2010). Extremely simple models are easy to interpret but may not be able to reproduce realistic behaviour, while too much complexity will lead to uncertainty and problems in interpretation of the model (Allen et al., 2010). Given differences we observed in the studied region, our results suggest caution on relying on a single “ultimate” model for understanding regime shifts behaviour and rather favour multiple lower to intermediate complexity models, as also recommended by Fulton et al. (2003). However, one should be careful transferring these results to other regions. More complex models could outperform simple models in different ecosystems, and have been suggested to be generally more portable (i.e. ability to perform well in diverse regions and physical settings) in a modelling comparative study focusing on the equatorial Pacific and Arabian Sea (Friedrichs et al., 2007). In future work, an ensemble approach to quantify the effects of model and internal variability uncertainty in regime shift detection would be beneficial. We added a section in the third paragraph of the discussion and conclusion section to discuss these points.

References:

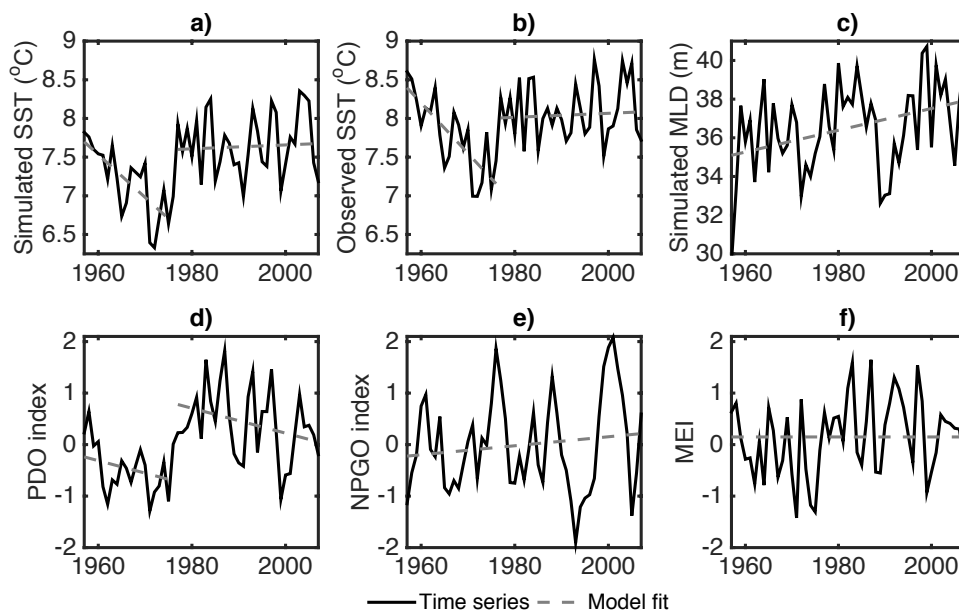
Allen, J. I., Aiken, J., Anderson, T. R., Buitenhuis, E., Cornell, S., Geider, R., Haines, K., Hirata, T., Holt, J., Le Quéré, C., Hardman-Mountford, N., Ross, O. N., Sinha, B.,

and While, J.: Marine ecosystem models for earth systems applications: The MarQUEST experience. *Journal of Marine Systems*, 81, 19-33, 2010.

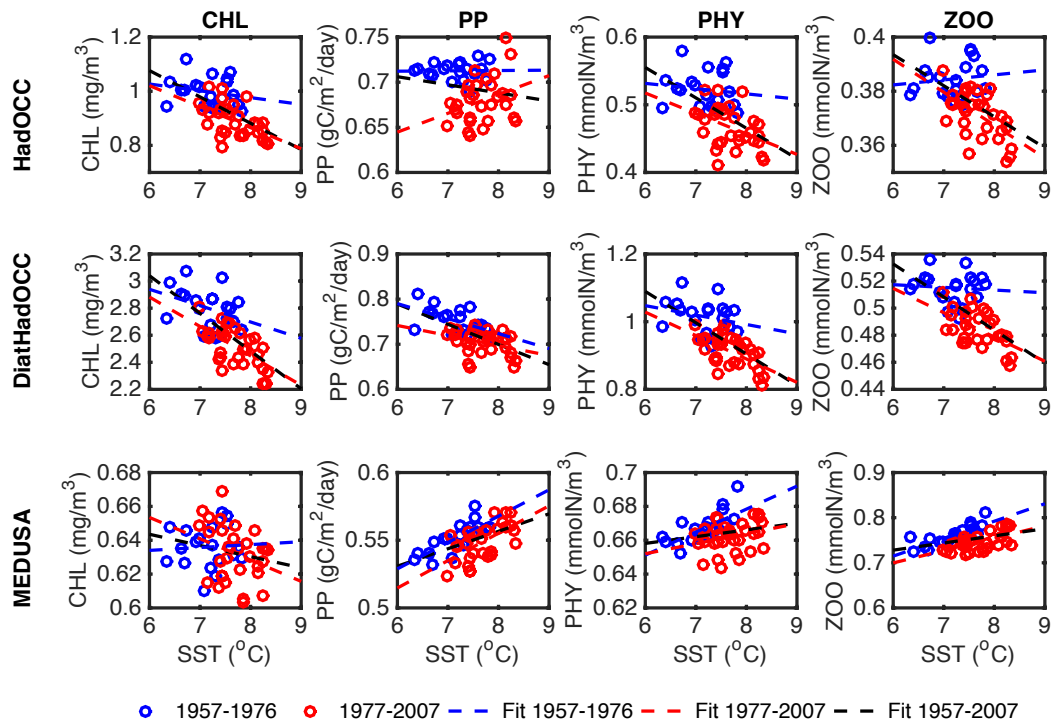
Friedrichs, M. A., Dusenberry, M. J., Anderson, L., Armstrong, R., Chai, F., Christian, J., Doney, S., Dunne, J., Fujii, M., Hood, R., McGillicuddy, D., Moore, M., Schartau, M., Spitz, Y., and Wiggert, J.: Assessment of skill and portability in regional marine biogeochemical models: the role of multiple plankton groups. *Journal of Geophysical Research*, 112, C08001, 2007.

Fulton, E. A., Smith, A. D. M., and Johnson, C. R.: Effect of complexity on marine ecosystem models. *Marine Ecology Progress Series*, 253, 1-16, 2003.

Kwiatkowski, L., Yool, A., Allen, J. I., Anderson, T. R., Barciela, R., Buitenhuis, E. T., Butenschön, M., Enright, C., Halloran, P. R., Le Quéré, C., de Mora, L., Racault, M.-F., Sinha, B., Totterdell, I. J., and Cox, P. M.: iMarNet: an ocean biogeochemistry model inter-comparison project within a common physical ocean modeling framework. *Biogeosciences*, 11, 7291-7394, 2014.



**Figure 3.** Time series of (a) simulated sea surface temperature (SST), (b) observed SST and (c) simulated mixed layer depth (MLD) for the Gulf of Alaska. The simulated time series of SST and MLD are the same in all 5 ocean models used. Time series of large-scale oscillations representing the climate in the Gulf of Alaska: (d) Pacific Decadal Oscillation (PDO) index, (e) North Pacific Gyre Oscillation (NPGO) index and (f) Multivariate El Niño Southern Oscillation index (MEI). The grey dotted lines represent the statistical model chosen (see Table A1) to fit these time series. Both the simulated SST and observed SST exhibit a significant shift in intercept and trend occurring in 1976 ( $p$ -value  $< 0.05$ , see Table A1). The MLD time series does not exhibit a significant shift and is best represented by a linear trend. Among the large-scale oscillations, only the PDO exhibits a significant shift in 1976.



**Figure 6.** Relationships matrix between simulated sea surface temperature (SST) and the biological variables over the Gulf of Alaska region. Rows represent different models (HadOCC, DiatHadOCC and MEDUSA) and columns represent different biological variables (surface chlorophyll (CHL), integrated primary production (PP), total surface phytoplankton (PHY) and zooplankton biomass (ZOO)). Linear relationships are inferred for the periods 1957-1976, 1977-2007 and 1957-2007 using least square regression. Table 5 presents test results on the similarity of these relationships.



**Table 5.** Forcing-response regressions in HadOCC, Diat-HadOCC and MEDUSA with sea surface temperature (SST) as the physical forcing and surface chlorophyll (CHL), integrated primary production (PP), total surface phytoplankton (PHY) and zooplankton biomass (ZOO) as the responses. The slopes of the linear regressions between the forcing and response before and after the shift are compared using a test of equality of two regression slopes. Bold indicates significant slope differences (p-value < 0.05).

<b>HadOCC</b>					
Forcing	Response	Slope 1957-1976 (standard error)	Slope 1977-2007 (standard error)	Test statistic	p-value
SST	CHL	-0.025 (0.028)	-0.008 (0.024)	1.407	0.166
	PP	0.000 (0.005)	0.021 (0.011)	-1.703	0.095
	TPHY	-0.008 (0.014)	-0.030 (0.013)	1.179	0.245
	TZOO	0.002 (0.004)	-0.012 (0.003)	2.823	<b>0.007</b>
<b>Diat-HadOCC</b>					
Forcing	Response	Slope 1957-1976 (standard error)	Slope 1977-2007 (standard error)	Test statistic	p-value
SST	CHL	-0.121 (0.071)	-0.217 (0.052)	1.095	0.279
	PP	-0.033 (0.012) <sup>b</sup>	-0.022 (0.012)	-0.666	0.508
	TPHY	-0.028 (0.025)	-0.069 (0.018)	1.345	0.185
	TZOO	-0.002 (0.006)	-0.018 (0.005)	2.034	<b>0.048</b>
<b>MEDUSA</b>					
Forcing	Response	Slope 1957-1976 (standard error)	Slope 1977-2007 (standard error)	Test statistic	p-value
SST	CHL	0.002 (0.006)	-0.013 (0.007)	1.476	0.146
	PP	0.019 (0.004)	0.020 (0.005)	-0.129	0.898
	TPHY	0.014 (0.004)	0.006 (0.004)	1.458	0.151
	TZOO	0.039 (0.007)	0.027 (0.007) <sup>b</sup>	1.132	0.263

<sup>a</sup> residuals not normally distributed (Lilliefors test, 5% critical level)

<sup>b</sup> residual variance not constant (Breusch Pagan test, 5% critical level)