When Fish are not Poisson:

Modelling the Migration of Atlantic Salmon (*Salmo salar*) and Sea Trout (*Salmo trutta*) at Multiple Time Scales.

Zelda van der Waal

A thesis submitted to Newcastle University in accordance with the requirements of the degree of Doctor of Philosophy in the Faculty of Science, School of Biology, Newcastle-upon-Tyne, UK.

February 2014

Abstract.

Migratory species undertake prolonged seasonal journeys; monitoring these movements is challenging but can sometimes be achieved by observations that taken locally and, ideally, using remote methods.

Amongst the best known examples of migrating fish in Europe, are Atlantic salmon (*Salmo salar*) and sea trout (*Salmo trutta*) that migrate between river and seawater. Characteristics of habitat suitability, feeding opportunities, predation, as well as salmonid sensitivity and needs, vary throughout successive stages of their anadromous life cycle.

Since the marine stage is the longest but is also challenging to monitor, in-river fish counters are of increasing importance in understanding salmonid patterns in abundance. The original contribution of this thesis lies in the use of modelling techniques to investigate salmonid migration, based on temporal observations produced by an electronic fish counter triggered by salmonid passage, as they return to spawn in the River Tyne.

Small scale observation revealed seasonal differences; aggregation behaviour intensified during the middle of the migration season, and explanatory covariates varied in both their effect size and relevance to salmonid abundance. At the population scale, migration was highly driven by annual periodicity, abundance increased with river temperature and there was an NAO effect with a four year lag, underlining the importance of marine conditions to parent population and/or post-smolts. Differences between distinct populations of *S. salar* and *S. trutta* appeared related to a species-specific annual periodicity and oceanic conditions as salmonids return (more so for *S. salar*). State-space models suggested a complex demographic structure for the two species. There was a species identification learning curve that affected the data by 2007. A classification algorithm determined that observations are more likely to be *S. salar* for larger signal amplitude, within a higher river flow and earlier in the year; characteristics were too similar between the two species to reach a useful classification success rate (69%). The project overall suggests specificities relating to both species and age-class that cannot be addressed in depth with the collected data; emerging limitations and recommendations are discussed.

Dedication.

J'ai la chance d'avoir une famille qui m'inspire et me donne ce sentiment de ce quelquechose d'inconditionel, qui m'a à la fois permis et rendue capable de faire ce que j'ai voulu. Ma Mère, Mon père, Ma Mémé, mon Pépé, ma Oma, mon Opa, et Axel: cette thèse est un des fruits de cette philosophie de vie que j'ai adoptée avec vous.

Thank you to my partner Jared who was there throughout this journey and its many ups and downs. These have been very happy years.

Acknowledgements.

This thesis represents countless cycles of exploration, inquiry, enlightenment, doubt, confusion, uncertainty, many mistakes, precious 'yes' moments, necessary wanders and perseverance.

I received much encouragement and motivation from both of my supervisors and feel lucky to have spent these three years under their guidance. Steve continually and convincingly conveyed a spirit of adventure in regard to research and an excitement in regard to learning. Aileen provided solid advice and continual support even after receiving the same piece of writing edited for the 12th time in a day. I thank both of you for the teaching I received, which went far beyond the academic context and will have lasting impacts.

I also express my gratitude to all the research team in the Modelling Suite who answered every one of my many questions, in particular Andrew, and who smoothened my research progress with their humour and support.

This research uses data routinely collected by the Environment Agency, who provided financial contribution towards the University fees. Thanks are expressed to Jon Shelley, Niall Cook and Philip Rippon, from the Environment Agency, who retrieved data and supplied pragmatic perspectives.

List of abbreviations.

AIC	Akaike information criterion
CAT score	Correlation-adjusted t-score
df	Degree of freedom
GLM	Generalized linear model
GLS	Generalized least squares
LDA	Linear discriminant analysis
LME	Linear mixed-effect
μ	Count process of a zero-altered or zero-inflated model
Neg. Bin.	Negative binomial distribution
P	Poisson distribution
π	Zero mass of a zero-altered or zero-inflated model
sig	Statistical significance (Wald test unless specified otherwise)
TP, TN, FP, FN	True positive, true negative, false positive, false negative
ZA	Zero altered model
ZI	Zero inflated model

Table of contents.

List of figures.

List of tables.

List of equations

“Home - is where I want to be,
But I guess I'm already there?
I come home -she lifted up her wings
Guess that this must be the place.
I can't tell one from the other,
Did I find you, or you find me?
There was a time before we were born
If someone asks, this where I'll be . . . “

Talking Heads, 1983
This must be the place (Naive Melody)

Chapter 1. .

"It cannot but affect our philosophy favorably to be reminded of these shoals of migratory fishes, of salmon, shad, alewives, marsh-bankers, and others, which penetrate up the innumerable rivers of our coast in the spring, even to the interior lakes, their scales gleaming in the sun; and again, of the fry which in still greater numbers wend their way downward to the sea."

H.D. Thoreau, 1849.
A week on the Concord and Merimack Rivers

Chapter 1. Introduction.

1.1. The Salmonids

For thousands of years, salmon and trout have had cultural connections to societies; the few records of fish in prehistoric iconography contain illustrations of both (Desse-Berset, 2011). In Irish Celtic mythology, salmon and trout are often associated with symbols of mysterious knowledge, wisdom and power; the Salmon of Knowledge gains all the knowledge of the world after eating nine hazelnuts that fell into the Well of Wisdom (Rolleston, 1910). The ability of salmonids to swim in both fresh and saltwater (two distinct worlds in mythology), leap over waterfalls and swim against the flow, also inspired the "salmon leap", a Celtic combat strategy (McKillop, 2004).

During the Middle Ages, salmon and trout were plentiful in Northern Europe; the fish were eaten fresh, smoked or salted and considered a food of the masses, poor and rich alike (McLeod *et al.*, 2006). In 1787, a local statute stipulated that "*local apprentices should not be obliged to dine on salmon off'ner than twice a week*" (in Gloucester, John Byng 1787 cited by McLeod *et al.*, 2006). Trout was also common in medieval cookbooks, as an alternative to salmon or boiled in wine (Adamson, 2004).

Here, the Atlantic salmon, *Salmo salar* L., and the migratory brown trout (also called sea trout), *Salmo trutta* L., are of particular interest.

The genus name *Salmo* (in latin, salmon) is of uncertain origin; possibly linked to *salir* (in latin, to leap; Andrews, 1955). Atlantic salmon owes its species name *salar* to its spectacular jumps; the name *trutta* is less obvious and possibly originating from *trōgein* (in Greek, gnaw) (Jonsson and Jonsson, 2011).

The species *S. salar* and *S. trutta* are sibling species. Their morphologies are so resemblant that an expert eye is often required to distinguish between the two; in addition, genetic and phenotypic variability between the populations of each species is high (Jonsson and Jonsson, 2011). This led to several changes in the

taxonomic classification of both species over the last centuries, before both were eventually separated and contained in the same genus *Salmo* (Behnke and Tomelleri, 2002) and are often referred to as *salmonids*.

Both species are of high importance in the UK, where they are exploited for sport and food. They both demand very high water quality and specific habitat criteria (Heggenes *et al.,* 2002) and their management is often combined as they tend to share similar streams.

The two species are anadromous; they breed in freshwater, yet have a marine stage to their lifecycles, where most of the growth and development occurs prior to returning to their natal river to breed (Hansen and Quinn, 1998; Gross, 1991).

The recognition of the return migratory route relies on the detection of olfactory cues. Salmonids release olfactory cues in the water such as urine, pheromones and the overall secretions produced by each fish; these are used subsequently by the salmonids to identify the natal river (Moore *et al.*, 1998) as well as conspecifics (Jaensson *et al.*, 2007; Moore *et al.*, 1994). It is currently not known whether olfactory cues are released and/or learnt in a specific order or not (*sequence* versus *imprinting* theory, McCormick *et al.*, 1998). Individual recognition implies awareness of the social rank, which allows reduced agnostic interactions (Moore *et al.*, 1994) and shorter fights (O'Connor *et al.*, 1999). Kins tend to group together and form a stable social system where fish are at-ease (Hojesjo *et al.*, 1998; Brown and Brown, 1992).

Despite all individuals of *S. trutta* being physiologically able to adapt to seawater, there are residents (Brown trout), who do not leave the freshwater stream and migratory individuals (Sea trout). Most females are migratory while up to 70% of males are residents (Bekkevold *et al.*, 2004). Residency is an alternative life strategy from which smaller fish result (Tanguy *et al.*, 1994). Here, only migratory *S. trutta* L. will be considered.

The particular life cycle of anadromous salmonids has long sparked the interest of many studies and is widely described in the scientific literature (the major works reviewed here are Eatherley *et al.*, 2005; Klemetsen *et al.*, 2003; Garcia-Vazquez *et al.*, 2001; Bardonnet and Baglinière, 2000; O'Connor *et al.*, 1999; McCormick *et al.*, 1998 and Brobbel *et al.*, 1996).

Anadromous life-history demands two important migrations. The first takes salmonids to the ocean where they feed abundantly on seawater organisms (small crustaceans and fish), grow, and become mature; after which they migrate back to their natal rivers to breed. In the UK, the spawning season takes place over the spring and summer of each year (Figure 1.1(a)); fry emerge from a pebble substrate after a few weeks (Figure 1.1(b)).

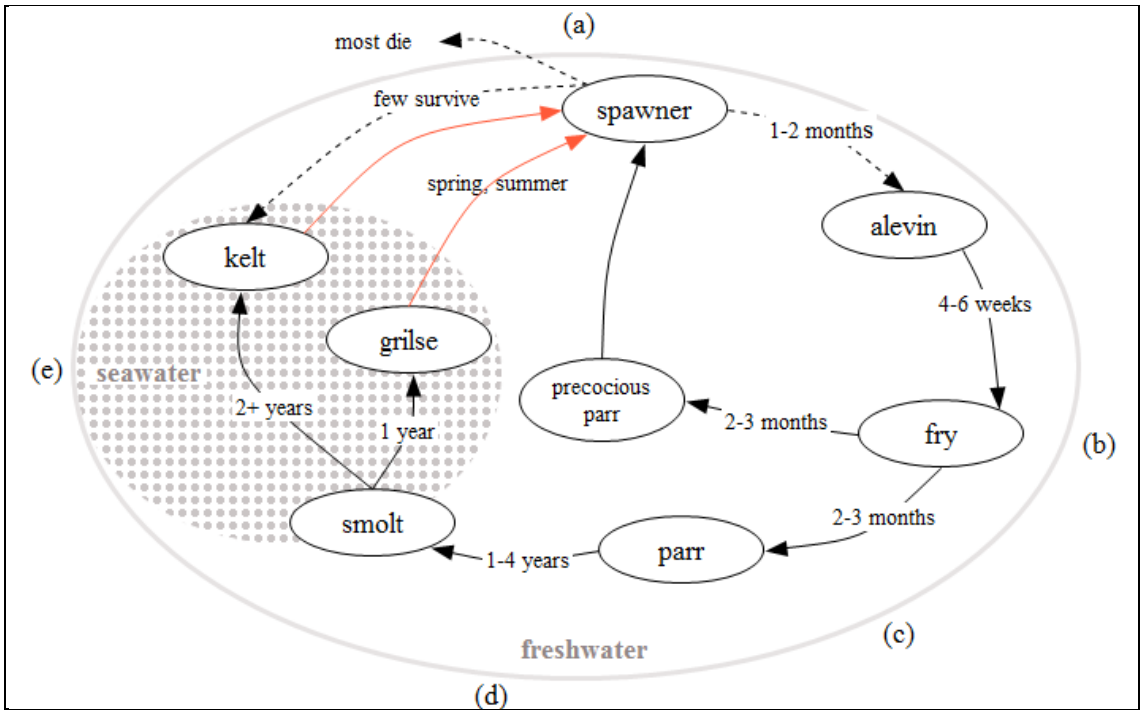


Figure 1.1: The major stages of the life cycle of anadromous salmonids.
The orange arrows represent the annual upstream migration by spawners, which are either kelt or grilse depending on the amount of time spent at sea (e). Most salmonids die after spawning, the few survivors return to sea (a).

After 2-3 months, fry become parr (Figure 1.1(c)). Parr may be precocious and breed without having left the river (*i.e.* as all resident *S. trutta* and 5-20% of *S. salar,* in Scotland, SNH, 2007). In both species, these mature parr contribute to a considerable part of the population growth and genetic flow. In *S. salar* they have high mortality rates.

All migratory parr (including precocious) mature in the river for 1 to 4 years during which they become smolt. Smoltification is likely triggered by seasonal variation in temperature and photoperiod and involves physiological and behavioural changes relevant to the transition into seawater (Figure 1.1(d), McCormick *et al.*, 1998).

Physical adaptations to seawater and breeding migration include the development of better olfactory and visual aptitudes, higher scope for growth

and an increased tolerance to salinity (McCormick *et al.*, 1998). Smolts become of a silver colour and lose the dark lateral bars typical of parr; they also have a leaner figure more suited to swimming in seawater (Folmar and Dickhoff, 1980), organize into schools, and lose their sense of territoriality (Keenleyside and Yamamoto, 1962).

Smolts migrate towards the sea, where they spend one year (grilse) or more (kelt) (Figure 1.1(e)). Growth rate increases with feeding availability and the high nutritive value of seawater plankton. In spring, breeders return to the natal river for the annual breeding migration after which most migratory salmonids do not survive.

The marine phase challenges survivorship as the quality of the habitat directly conditions growth, fitness, and survival (Friedland *et al.*, 1993). Global monitoring of fish populations is however difficult in vast marine environment, and is linked to a lack of knowledge regarding the marine life stage of salmonids.

There is a world-wide need to understand, optimise, and manage fishing, but monitoring oceanic phenomena is difficult. Conservationists are concerned with the preservation of species and function; world fish catch appeared stable for a considerable period of time in the 1980s but the decline that followed the 1989 peak is still progressing today (Beaugrand and Reid, 2003; Anderson, 1996).

Under the UN legislation, anadromous salmonids are straddling stocks; these refer to fish that occupy both international waters and economic exclusion zone of one or more states (200 nautical miles from the coast, McKelvey *et al.*, 2002). Populations of straddling fish all over the world are thought to be "*undermanaged and overfished*" (Stokke, 2000). World fish stocks are globally poorly managed; policies do not prevent the destruction of populations (fully to heavily exploited, over-exploited, depleted or slowly recovering, Anderson, 1996), and there are not enough practical answers to improving fisheries managements proposed to managers by scientifics (Pitcher and Cheung, 2013; Fulton *et al.*, 2010).

S. salar is more abundant than ever but 98% of it is in aquaculture for human consumption (Parrish *et al.*, 1998). The remaining 2% of *S. salar* are wild fish and are in global decline (McGinnity *et al.*, 2003). The high abundance of

farmed fish also impacts on the wild populations (*e.g*. escapees and interactions with the wild populations) with both immediate consequences on behaviour and competition, and long-term consequences on the wild genetic patrimony (Houde *et al*., 2010).

In the UK, *S. salar* is in decline; Scotland holds 80% of the UK stocks and these are stable but England's populations are declining (JNCC, 2007). According to the Environment Agency , approximately half of the salmon rivers of England and Wales are at risk or "probably at risk" regarding the local conservation limits of salmon stocks; the trend is slowly moving towards compliance with management objectives since 2004 (Environment Agency and Cefas, 2012).

Declining trends in migratory salmonids are mainly attributed to the anthropogenic use of the freshwater environment, which has led to overharvesting, water pollution, dewatering and "main stem dams" construction. Factors often co-occur so individual impact is hard to quantify (Parrish *et al*., 1998).

Construction of dams without fish passages has reduced anadromous species around the world (Parrish *et al*., 1998). Water pollution also had a destructive effect in many rivers in Europe. Polluting discharges such as domestic effluent and industrial waste (Parrish *et al.,* 1998) engendered critical levels of water pollution and pesticides that affect fish reproductive systems, for example the pyrethroid pesticide cypermethrin inhibits the olfactory response of males to female pheromone release after ovulation, which compromises reproduction (Moore and Waring, 2001).

Little is known about salmonid epidemiology, partly because sick fish are hard to track through time (Bakke and Harris, 1998). Hormonal and immune systems are closely related. Stress (cortisol) may lead to immunosuppression (Felvoden *et al*., 1993); growth hormone is associated with increased immune response; and reproductive hormones lower antibody productivity so that spawners are generally more sensitive to infestation than younger individuals (Harris and Bird, 2000). There is a risk of homogenization of the gene pool by breeding of wild and cultured fish (Bakke and Harris, 1998).

Common salmonid parasites include sea lice, of which the species *Lepeophtheirus salmonis* is the most studied. Sea lice cause a general physical

weakening due to skin lesions, stress, increased leaping behaviour (Costello, 2006), and lower swimming speed (Wagner *et al.*, 2003). Without engendering direct mortality, sea lice are one of the major threats to salmonids and in particular to *S. trutta* (Freyhof and Kottelat, 2008). As *L. Salmonis* populations increase exponentially at higher temperatures, there are increasing concerns over the effect of global warming (Costello, 2006).

Another stressor for *S. salar* is infestation by *Anisakis simplex* which causes red vent syndrome, specific skin lesions engendering red, swollen, bloody vent (Noguera *et al.*, 2009). *A. simplex* is a round worm commonly observed in the body cavity of wild oceanic fish in which it lives harmlessly. Examples of bacterial diseases in wild salmonids include furunculosis, bacterial kidney disease, and ulcerative dermal necrosis, none of which have been a cause of concern since the 1980s (Noguera *et al.*, 2009).

Over the last century, industrialized landscapes have pressured salmonids to either retreat to streams free of anthropogenic influence, or disappear (Montgomery, 2004). This argument is rejected by several instances of return of salmonids after recent improvements of river quality. For instance, Atlantic salmon were caught in the Seine (Paris) in 2008 for the first time in 70 years (*Le Monde*, 2008). In the UK, while national trends are declining (150,000 fewer *S. salar* recorded between 1994 and 2006), local increases are also recorded, in particular in industrialised rivers of the North-East England region (JNCC, 2007).

Freshwater ecosystems services include water, food and waste assimilation (Folke *et al.*, 1997). Urban streams also provide a higher quality of life to city inhabitants (Bolund and Hunhammar, 1999). It is estimated that ecosystem support areas require to be 500-1000 times larger than the city they support (Baltic cities, Folke *et al.*, 1997) and by 2030, more than 60% of the world's population is expected to be occupying urban areas (UN, 1997). The societal need for urban stream ecosystem services will increase accordingly: there is a need to insure the sustainability of urban streams and the ecosystem services they offer.

The scientific literature offers several ideas to manage salmonid stocks sustainably. Highly migrating fish are particularly prone to overexploitation. In

1995, an agreement by the United Unions aimed to create a world-wide management plan in order to protect salmonid species and limit the disputes originating from the rivalries over fishing stock (Stokke, 2000); the agreement required the control of harvest levels to be jointly applied by all fishing countries.

Management tactics of *S. salar* may however require important local adaptations. Great variations of habitat use and demand within species and complexity of the aquatic ecosystems (freshwater and oceans) imply a high level of specificity in management requirements (Armstrong *et al.*, 1998); effective management strategies likely require both large and small scale considerations.

It may be argued that the conservation of salmonid stocks should be limited to the wild genetic pool of genes, making the relevance of hatcheries and breeding programs controversial (Nielsen, 1998).

The precautionary approach is recognised nearly unanimously by the scientific community (Dodson *et al.*, 1998); for salmonids, it means that the protection of existing salmonid stocks is the priority for their conservation. Cooperation of scientists and managers is a necessity to link conservation issues to conservation plans (Morishita, 2008; Armstrong, 2006).

In UK, the Salmon and Freshwater Fisheries Act (1975) regulates most matters regarding salmonids. Regulations include: the attribution of fishing licences (Part IV), regulations of fishing methods (Part I), the delimitation of fishing seasons (Schedule I), the limits of catch size (Part V, F71, release of salmonids smaller than 12 inches) and the preservation of the salmonid migratory path from anthropogenic constructions and actions (Part II).

1.2. The River Tyne

The River Tyne, in the North of England, flows through the cities of Newcastle-upon-Tyne and Gateshead. It has two major tributaries of which one is natural (South Tyne) and one is regulated (North Tyne) by releases of Kielder Water (linked to reservoir management, Figure 1.2).

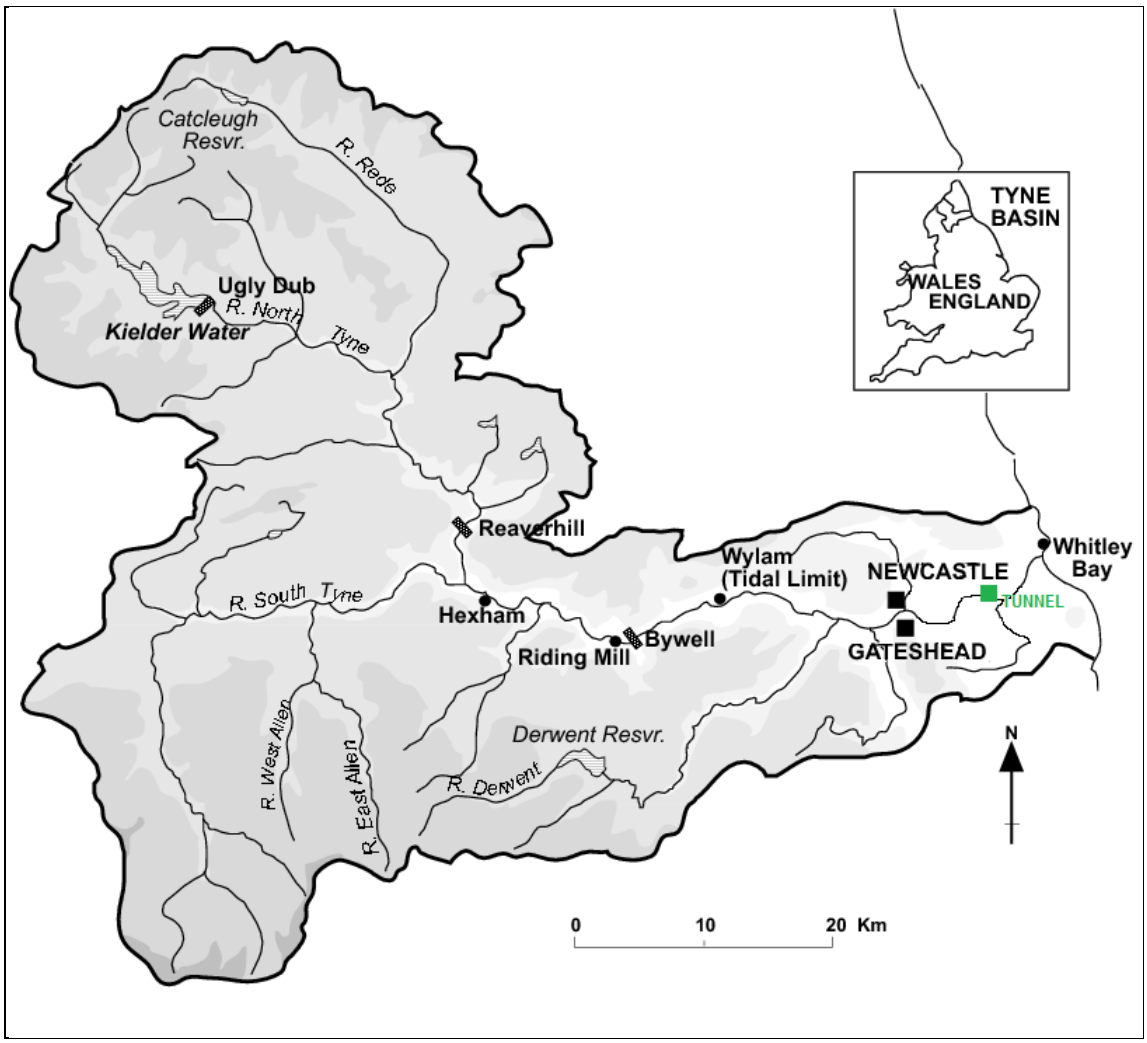


Figure 1.2: The River Tyne catchment.
The Kielder Water reservoir is responsible for the increase in minimum flow level. The Riding Mill station is located 11km upstream from the tidal limit at Wylam and 61km downstream from Kielder Water. (Map courtesy of Archer, 2013).

Since at least the Middle Ages, the River Tyne has been known as a salmon river. However as for many European streams, industrialisation rendered the Tyne too polluted to support migratory salmonids (Jonsson and Jonsson, 2011). The river became saturated with domestic, industrial, and pharmaceutical pollutions (Roberts and Thomas, 2006; Archer, 2003) and was soon one of the rivers of North East England the most affected by anthropogenic activity (Hall *et al.*, 1997). By the late 1950's, no salmon was to be caught and numbers remained under 400 a year until the 1970's (Archer *et al.*, 2008).

In the 1980's, major changes occurred in the industrial landscape of Newcastle-upon-Tyne and Gateshead, which greatly benefitted the quality of the River Tyne. The Tyneside interceptor Sewerage scheme allowed the interception of over 250 effluent discharges that subsequently ceased flowing directly into the

estuary; minimum freshwater flow was increased by releases from Kielder Water that reduced conditions of drought (Haile *et al.*, 1989 and Figure 1.2).

At the same time, Kielder salmonid hatchery was created in an attempt to restock the river with *S. salar* (Haile *et al.*, 1989). The improvement of water quality was soon matched by a strong return of the migratory salmonid populations. Today, the River Tyne is amongst the best salmon rod fishing rivers in England and Wales (Archer, 2008) with an estimated 21,000 to 49,000 annual returns since 2004 (*S. salar* and *S. trutta* combined, EA, 2013).

In 2009, Newcastle city council began to construct a second underwater crossing of the River Tyne between East Howden and Jarrow (Figure 1.2). As these works disturbed the river bottom through dredging and removal of sediments, the New Tyne Crossing Scheme funded an enhanced fishery monitoring programme to be undertaken by the Environment Agency.

The Environment Agency is an Executive Non-departmental Public Body responsible to the Secretary of State for Environment, Food and Rural Affairs and an Assembly Sponsored Public Body responsible to the National Assembly for Wales. The actions undertaken by the agency affect a wide range of environmental issues. This study makes use of these data routinely collected by the Environment Agency.

1.3. Counting salmonids in the River Tyne

Salmonid count data were generated by a resistivity counter installed at the Riding Mill Station (located on Figure 1.2). Resistivity counters are an array of electrodes measuring water electrical resistance. The counter relies on the principle that fish have a lower electrical resistance than the water in which the device is placed: the passage of salmonids over the electrodes provokes an increase in conductivity between them (Forbes *et al.*, 2000) indicating passage of a fish. Each movement of salmonid travelling across the counter generates a signal in the shape of a waveform, without affecting the fish (Figure 1.3).

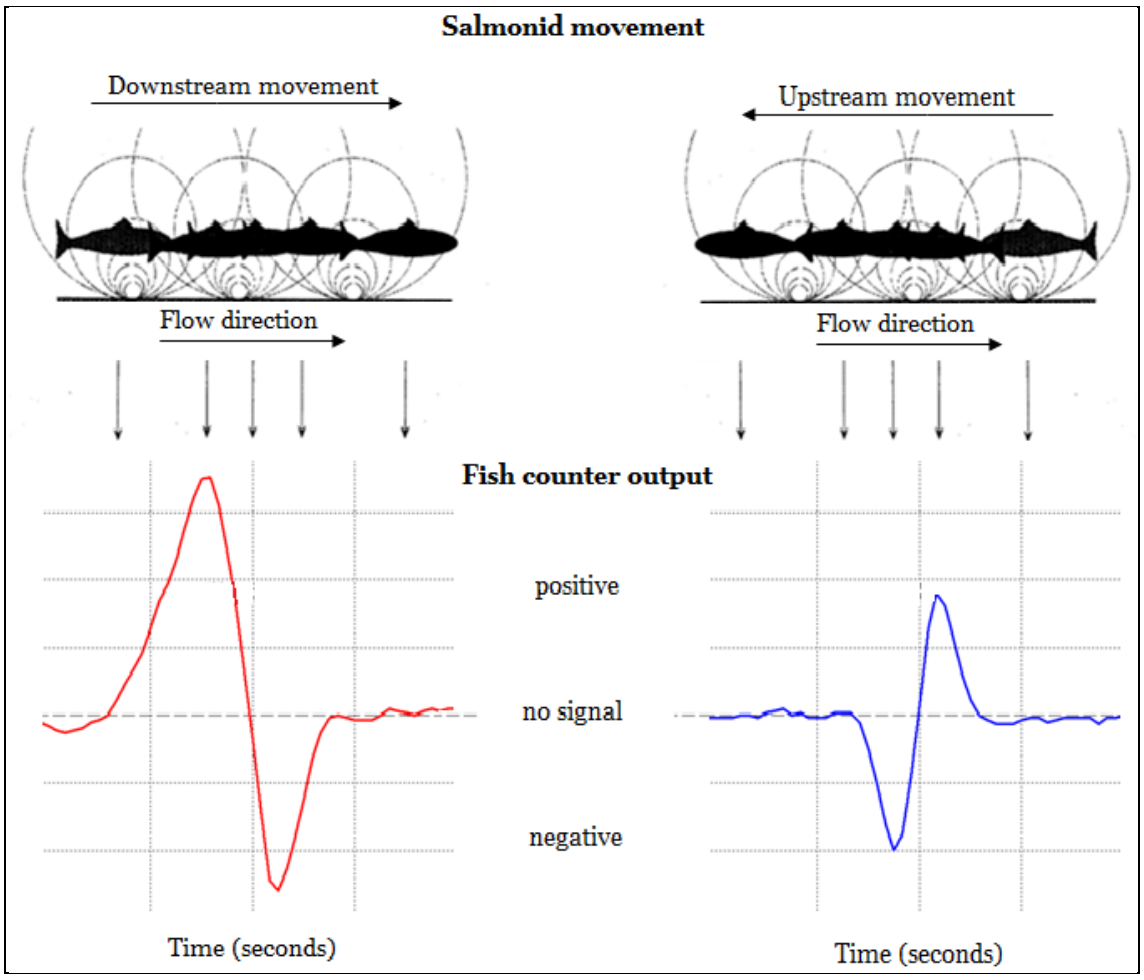


Figure 1.3: The passage of salmonids over counter electrodes and the associated waveforms showing the change in resistance.
The order of peaks describes the direction of the movement and the magnitude of the waveform is proportional to the size of the fish (fish counter illustration: Fewings, 1994, reproduced with permission of the Atlantic Salmon Trust, Andrews, 2013; signal data: EA, February 2013).

There are three noticeable moments of the waveform that correspond to the passage of salmonid across three electrodes on the face of the weir at Riding Mill station (Figure 1.3). Electrodes at both ends of the array are fed an anti-

phase sinusoidal signal (3.5 kHz). The central electrode, in the middle of the other two, emits no signal and serves as a fulcrum that detects the difference in signal (Aquantic, *per. com.*, 2013).

As a fish swims across the first electrode it reduces the resistance between this electrode and the central electrode. The central electrode detects a signal which is in phase with the third electrode, and it registers a peak. When a fish is above the central electrode, the signals of the other two electrodes are balanced so the central electrode signal returns a null value, which corresponds to the zero between peak and trough. As the fish travels further through the third electrode, the unbalanced signal at the central electrode is in phase with the upstream drive, which produces a trough (Aquantic, *per. com.*, 2013).

There are two distinct uses of the signal data produced. In most instances, signals are used to count salmonids passing over the counter. These count data are binned at a minimum of 15 minutes intervals. Signals can also be studied in terms of amplitude of the waveform. A positive relationship between signal size and salmonid size has been previously described (Nicholson *et al.*, 1994) however no formal analysis has investigated this relationship for populations of the River Tyne.

The counter is divided into four parallel channels. A video camera was installed on one of the channels (channel 4) that allows validating fish presence by visual examination of the video records and waveforms (by EA staff members). Also for this channel, it is possible to link signal size data to either *S. salar* or *S. trutta*.

The resistivity counter was installed in 2003 and the video recordings started in 2004. Both count data are available from the start until 2011.

1.4. Hierarchy and scale of observation

> *"Four blind men are let into a courtyard to experience an elephant for the first time. The first grasps the trunk and declares that elephants are fire hoses. The second touches an ear and maintains that elephants are rug. The third walks into its side and believes elephants are a kind of wall."* O'Neil *et al.*, 1986

Like elephants, animal populations may be studied at different scales and provide the observer with different conclusions. A given scale addresses a specific problem; a theory contrasts the conclusions. A point of view neither encompasses all possible observations nor can be considered more fundamental.

Spatial and temporal scale is an essential characteristic of an observation set and is defined by the ecological question (MacNeil *et al.*, 2009; Levin, 1992). An observation set includes "phenomena of interest, specific measurements taken and the techniques used to analyse the data" (O'Neil *et al.*, 1986). Thus, scale may be regulated by the number of measurements and intervals between them over a given observation period. The study of a single individual behaviour in a determined area may require description at a small scale, while a contrasting observation set at a larger scale may describe the species migration pattern.

Grain is the minimum resolvable time period in the dataset, *i.e.* the measurement interval (also called inner scale or resolution). The observation period is the scope in time (also called outer scale, or extent). Fine scales refer to fine grain and small extent; large scales refer to coarser grain and larger extent (Urban, 2005; Schneider, 2001).

This thesis seeks to investigate the ecology of salmonid movement and migration across a range of ecological questions at different temporal scales, using count data and a set of environmental covariates hypothesised as being important in determining the movement of fish.

1.5. Environmental covariates

River parameters monitored by the Environmental Agency include flow, temperature, tidal state, turbidity, colour, pH, conductivity (Table 1.1). Measurements were taken over four sites (Riding Mill, Bywell, North Shields and Northumbria Water's Horsley treatment works station) at different time intervals. In addition, data on solar irradiance, NAO and AMO were retrieved or calculated (respectively from NOAA, 2005, JRC, 2012 and Enfield *et al.*, 2001). The time ranges covered by the data collection varied (Table 1.2); counter data were available from 2003 to 2012 and video records were available from 2004 to 2012.

Table 1.1: The parameters monitored on the River Tyne, with location, format and source.

Parameter (unit)	Location	From counter	Measure interval	Source	Used in Chapters	Scale of effect studied
Flow ($m^3.s^{-1}$)	Bywell	1km down	15 min	EA	2 to 6	Both small and broad
Temperature (°C)	Bywell	1km down	15 min	EA	2 to 6	Both small and broad
Tidal state (m)	North Shields	42km down	15 min	EA	2 and 6	Both small and broad
Turbidity (NTU)	Horsley	8km down	4 hours	EA	2 and 6	Small
Colour (HU)	Horsley	8km down	4 hours	EA	2 and 6	Small
pH (NA)	Riding Mill	0km	15 min	EA	2 and 6	Both small and broad
Conductivity ($\mu S.cm^{-1}$)	Riding Mill	0km	15 min	EA	2 and 6	Small
Solar irradiance ($W.m^{-2}$)	calculated	0km	15 min	JRC, 2012	2 to 6	Both small and broad
NAO (index)	Retrieved	NA	1 month	NOAA, 2005	3 to 5	Broad
AMO (index)	Retrieved	NA	1 month	Enfield *et al.*, 2001	3 to 5	Broad

Table 1.2: The time ranges covered for the covariates in this study.
Grey cells indicate that the data were at least partly available for the year in question.

Parameter	2003	2004	205	2006	2007	2008	2009	2010	2011
Flow									
Temperature									
Tidal state									
Turbidity	NA								NA
Colour			NA						NA
pH		NA							NA
Conductivity		NA							NA
Solar irradiance									
NAO									
AMO									

The values for solar irradiance were calculated for the specific geographic coordinates of Newcastle-upon-Tyne. They are values for global irradiance, which combines diffuse irradiance (amount of radiation received by scattered atmospheric particles) and direct irradiance (direct from the sun) and were corrected by the average monthly cloud cover (JRC, 2012).

North Atlantic oscillation (NAO) index values refer to the difference in atmospheric pressure at sea level between the Icelandic low and the Azores high. This difference modulates the strength and direction of westerly winds across the Atlantic Ocean; both have an impact on the climate of the whole Northern hemispheric (NOAA, 2005). Fluctuations in NAO are not periodic (Greatbatch, 2000).

The index for the Atlantic multi-decadal oscillations (AMO) refers to the variations in the North Atlantic sea surface temperatures. AMO is correlated with air temperatures and rainfall over the Northern hemisphere and linked to climatic conditions over a broad range (from summer rainfall in India to Atlantic hurricanes, Ting *et al.*, 2008). The periodicity of AMO is uncertain given the number of years of data available; it is assumed to be of a quasi-cycle of approximately 70 years, which would imply a peak in 2020 (Enfield *et al.*, 2001).

1.6. Scales of observation of salmonids in the River Tyne

Salmonids of the River Tyne and influential factors associated with their migration can be studied at several temporal resolutions migration (Figure 1.4). Hypotheses are formulated accordingly as follows.

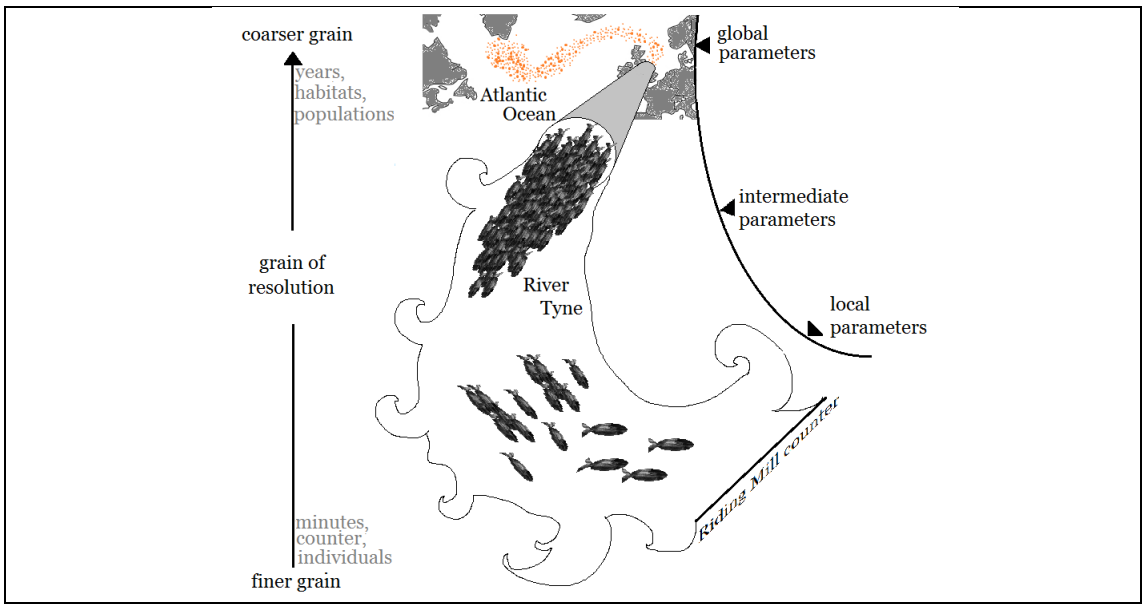


Figure 1.4: The relationship between the grain of the resolution, the spatial scale of observation of the salmonid population and the level of parameters relevant to the observation.

For this project, smallest levels of observation related to individual behaviour; it is the count of the individual fish as it moves through the counter and the factors that determine this movement. These are likely to be stochastic, involving intrinsic characteristics of the individual, and challenging to model as a response to environmental covariates. The intermediate level is the group. Salmonids form loose groups that may or may not be determined by their arrival and queuing. There may therefore be group strategies and dynamics interacting with the environment in a way that is somewhat better understood at the group scale. At the highest level, salmonid movements are observed at the population level over their entire territory of occurrence, effectively integrating all of the individual behaviour and interactions between individuals and groups.

Relevant temporal scales of observation of salmonids of the River Tyne involve specific combinations of time intervals between measurements and duration of the observation period. In Chapter 2, salmonids are studied as individuals interacting with others. The small timescale of observation uses intervals between counts of 15 minutes and observation periods of 4 days. In Chapters 3

to 5, a large timescale of observation is used as salmonids are studied as populations. Counts are binned into 14 day intervals and all available years are used. Chapter 6 uses instant counts to investigate individual characteristics. At each scale, hypotheses are tested concerning the role of environmental covariates measured at the same temporal scale. Hence, the influence of conductivity, turbidity and colour was considered for the smallest time interval since it was hypothesised that this would determine individual movement decision by fish as punctually experienced local conditions. Flow, temperature, solar irradiance, and tidal state were hypothesised to be concerned with small and medium temporal scales of observation. The three former parameters were also concerned with a large scale and a coarse grain of resolution; they follow an annual cycle which may consequently impact on the salmonids at the large annual scale, as well as values of NAO and AMO and the annual seasonal cycle.

Local parameters were by definition only relevant for the study of small scale phenomena; these parameters had a high short-term variability which suggested that the potential response of salmonids to the variations in these parameters was fast and simultaneous with the environmental changes. Intervals between measurements were consequently small and their aggregation to several days made little ecological sense. Fluctuations in these parameters were not cyclical.

Intermediate parameters had a seasonal component; daily cycles (temperature and solar irradiance), or cycles of 28 days (tide) and/or annual cycles (temperature, solar irradiance and flow).

The strength and direction of tidal currents regulate the level of effort experienced by the salmonids on their upstream movement. Salmonids may consequently show preferences for specific moments of the tidal cycle to swim through the estuary (Smith and Smith, 1997) and the swimming speed is affected by the strength of the current (Tang *et al.*, 2000). In turn, the time of entrance and swimming speed partly decide of the time of arrival at the counter. The tide consequently influences salmonid movements on the short-term, and its influence can be detected at the counter level; tidal currents were considered local parameters.

NAO and AMO are linked to many aspects of the climate in the northern hemisphere; they were the highest temporal parameter available for this dataset. These indices capture combined information on temperature, rainfall (and related flow rate), as well as plankton density and location, *i.e.* productivity in the Atlantic Ocean (Beaugrand and Reid, 2003; Attrill and Power, 2002). As salmonids spend several months depending on the oceanic productivity, it can be hypothesised that migration routes of both species was likely affected by conditions represented by the NAO and AMO indices (Dadswell *et al.*, 2010). These values were to some extent, a summary of climatic and productivity conditions met during their oceanic phase.

Additionally at the largest scale, harmonics were used to account the annual seasonal periodicity. The year count was also available as a surrogate for between-year variations and recruitment occurring during the time spent out at sea. Plankton data were not readily available (a dataset was obtained and tested externally but the reliability of the data could not be verified).

1.7. Rationale

Chapter 2

Salmonids swim up the River Tyne and are seen either alone or as part of a group. Can we identify the factors that make individual salmonids aggregate or not? Can individual presence and grouping behaviour be described by environmental parameters?

Chapter 3

The resistivity counter at Riding Mill aims to improve conservation. Can the data produced by the counter be used to describe the migration and so be used for management? What drives the salmonid migration? As an application, can we determine whether the tunnel construction in the River Tyne impacted on salmonid populations?

Chapter 4

The camera at Riding Mill provides species-specific data. Do video recordings constitute an exploitable source of information? Can the migration of *S. salar* and *S. trutta* be modelled separately based on these records?

Chapter 5

Environmental conditions experienced during the migration were described using static measurements and time-invariant relationships. Yet, each salmonid species covers a specific and considerable geographic range and their relation to parameter and parameter measurements may change throughout the year and between years. Is time-dependency a tool for improving the description of the migration model for each species?

Chapter 6

Can we discriminate between the species *S. salar* and *S. trutta* in the River Tyne based upon a combination of signal data and environmental parameters?

Chapter 7

From these observations at various scales, concluding reflections are formulated regarding how salmonids are linked to each other and the surrounding environment.

Chapter 2. .

About counting on fish being Poisson.

Chapter 2. A temporal study of fine scale salmonid movements.

2.1. Introduction

There is a lack of understanding of salmonid aggregation behaviour and the factors driving it, especially in river. Fish commonly form social aggregations that are determined socially rather than solely by environmental conditions, although abiotic factors contribute to shaping the shoal structure (Tien *et al.,* 2004). In presence of a threat, a common mechanism of aggregation consists in approaching the nearest neighbour in time, which involves consideration of movement time, fish position and angle (*e.g.* in sticklebacks, *Gasterosteus aculeatus*, Krause and Tegeder, 1994). Furthermore, foraging may be optimized via transfer of information amongst individuals of the same shoal in habitat where food resources are patchy (Tien *et al.*, 2004) (Pitcher, 1985).

It is a common belief shared by fishermen and scientists that, when at sea, anadromous fish organize as shoals; albeit little research exists to confirm the theory (Palm *et al.*, 2008). Anadromous salmonids are known to aggregate during their seaward "spring" migration (Riley *et al.*, 2002; Krause *et al.*, 2000). The spring migration is thought to be triggered by both the physiological readiness of the fish to the transition between fresh and saltwater (*i.e.* smoltification, McCormick *et al.*, 1998) and external environmental releasing factors of which onset of darkness is thought to be the major factor (Riley *et al.*, 2002).

Spring migration aggregation behaviour varies with extrinsic and intrinsic factors. For instance in the River Conway (Wales), movements of *S. trutta* were

observed at all times in the lower reaches of the estuary but in the upstream part of the river, *S. trutta* were resident in the river during the day and moved actively and higher in the water column at night (Moore *et al.*, 1998). The duration of residency may also change seasonally for both *S. salar* and *S. trutta*, and tends to become shorter towards the end of the season (Moore *et al.*, 1998).

Shoals are cohesive groups, as opposed to random aggregations or individuals (Palm *et al.*, 2008). During spring migration salmonids go from being solitary to forming shoals further downstream (Riley *et al.*, 2002). The cues initiating the downstream migration vary throughout season with, for example, downstream movement of *S. salar* being correlated with the onset of darkness in April but independent from abiotic conditions during the rest of the year (Riley *et al.*, 2002).

Antagonistic behaviour is reduced amongst groups of *S. trutta* composed by familiar individuals (Hosjesjos *et al.*, 1998) and a preference for conspecifics over heterospecifics has been demonstrated for each species (Brown and Brown, 1992). Recognition mechanisms remain unclear as much experimentation on kin recognition used salmonids reared in groups; recognition may consequently be linked to familiar environmental conditions rather than the recognition of kin and it has not been demonstrated whether or not the olfactive preferences as fry persist until adulthood (Krause *et al.*, 2000).

There is no evidence that individual *S. salar* group with kin while in the Baltic Sea, but post-smolts *S. trutta* tend to gather with individuals originating from the same English river (Palm *et al.*, 2008). In natal rivers, spawners on their return migration are observed either alone or in groups. This study aims to determine the extent to which variation in riparian conditions may be used to explain levels of aggregation in salmonids of the River Tyne.

Upstream migration from an estuary to spawning grounds is a succession of positive rheotaxis and prolonged stops that are river and species-specific (Bazarov and Golovanov, 2003). Wild salmonids tend to travel towards spawning grounds without deviating their route, even more for *S. trutta* than for *S. salar* (Bazarov and Golovanov, 2003). Swimming speed fluctuates from a population to another, but there is a general tendency to slow down when

approaching spawning grounds and speed is often different for *S. salar* and *S. trutta* (Bazarov and Golovanov, 2003).

Benefits for salmonids spending a long period of time in freshwater are unclear, especially for *S. salar* which is both fasting and missing on rich feeding opportunities from the sea environment during this part of the migration (Fleming, 1996). Residency can be a mechanism to cope with adverse environmental conditions. For instance, *S. salar* displays 'sit-and-wait' behaviour by marking a pause behind an obstacle to shelter from turbulent flow, often downstream of an obstacle such as a rock (*i.e.* home rock, Haro *et al.*, 2004). In this instance, the presence of suitable rock and turbulent flow may induce residency periods that can last hours.

In north Norway three phases of movement have been differentiated (Okland *et al.*, 2001); direct migration until near the spawning grounds during which salmonids may pause or not, followed by seemingly inconsistent movements around the spawning territory (possibly a searching phase*)* and a final period of residency before spawning (Okland *et al.*, 2001). In south-west England, *S. salar* also start by migrating up close to their spawning grounds, then rest for a considerable period of time (*i.e.* residency) before migrating to the spawning grounds (Heggenes *et al.*, 2002).

The reasons for residence during upstream migration relate to habitat suitability, rather than territoriality and ownership as it may be the case for parr of both species (Johnsson and Forser, 2002). The causes of the duration and occurrence of residency have not always been determined. During downstream migration, smolts of both species are physiologically capable to progress into saline environment when they are located in the estuary, so smoltification does not appear to play a role (Dempson *et al.*, 2011). In *S. salar,* residency time relates to thermal acclimation, food availability, and tidal currents (Dempson *et al.*, 2011).

The difference between grouping and random behaviour, and the parameters influencing the time spent in residence, may be accounted for using different generalized linear model (GLM) approaches and through the use of selected error distributions.

GLMs express the dependence of a response variable to a vector of explanatory

regressors (Zeileis *et al.*, 2008). GLMs permit analysis of continuous and discrete data in a framework similar to the one developed for normal theory linear regression, applied to data not normally distributed. Furthermore, additional specifications of the mean (link function) and the variance (variance function) enable a much wider spectrum of application than with linear regression (Breslow, 1996). A GLM is developed in three steps: assuming a distribution for the error, specifying the predictor function, and stating the link between expected values and the predictor function (Zuur *et al.*, 2009). These strong assumptions about the structure of the data require the deviance-based residuals of the models to be examined in order to verify that the statistical conditions are met (Pierce and Schafer, 1986). In case of a lack of fit, residuals may help understanding the relationship between observed values and the candidate model (Fox, 2009).

The Poisson and negative binomial distributions of the error both consider discrete positive or null values, which make them suitable to the study of salmonid counts. The two distributions handle the data differently. The Poisson distribution is skewed to the right and implies that the probability of an event occurring is proportional to the interval considered. The events are hence considered independent, so that when modelled with a Poisson distribution of the error, salmonids are considered individually and their abundance is proportional to the duration of the period studied. The negative binomial distribution implies a relation between events, here the salmonid counts. The abundance in salmonid is then defined as a ratio: the abundance at a given time over the number of observation required before a certain abundance is reached. In other words, this distribution of the error implies that salmonid counts are a fraction of a larger group.

Equation (2.1)	$E(Y_t) = \mu = g^{-1}(a + b\,X_t) + e_t$	
Equation (2.2)	Poisson error:	$e_t \sim P(0, v^2 = \mu)$
Equation (2.3)	Negative binomial error:	$e_t \sim NB(0, v^2 = \mu + \mu^2/\theta)$

GLMs assume response ($E(Y_t)$) was generated by a given distribution, whose mean (μ) relates linearly to independent variables ($a + b\,X_t$) via a link function (g^{-1}).

Overdispersion in count data is defined as the variance being greater than the mean (Zuur *et al.*, 2009). The Poisson distribution implies mean and variance being equal (Equation 2.2), while negative binomial distribution allows the variance to be larger than the mean via an extra parameter within the mathematical description of the variance (Equation 2.3).

Consequently, the negative binomial distribution can be used to model processes where overdispersion occurs while the Poisson distribution cannot. The parameter Theta (θ) of the variance function is the dispersion parameter (Zuur *et al.,* 2009); it defines how far from the Poisson distribution the distribution is (and hence, how much closer the process corresponds to a group behaviour). The Poisson distribution compares to the value of θ becoming infinitely high. Counts following the statistical assumptions of the negative binomial distribution (Equation 2.3) may be grouped in clusters, with the individuals of one cluster either having a common random term, or being each associated with a dispersion parameter also dependent on covariates (Cox, 1983). So, the level of aggregation increases inversely to the value of Theta, which here is assumed to describe the schooling behaviour.

In this study, salmonids of the River Tyne are observed over a short time, allowing the description of individual behaviour. At this reduced scale, null count values will be both abundant and meaningful, and may reflect distinct aspects of the local behaviour of a salmonid. The values are referred to as *false* zeros *or true, structural* zeros. Generalized linear models are appropriate to model count data; however the ordinary Poisson and negative binomial distributions produce biased estimates and standard errors if the abundance of zeros is extremely high (Zuur *et al.*, 2010). To more accurately model these data, mixture and two-part models will be developed; these models account for the abundance of null values and their potentially different origins (Figure 2.1).

In ecology, a true zero occurs when an absence is due to ecological processes (Martin *et al.*, 2005). In reality, null values are also falsely recorded (sampling error) and a suitable habitat may not be saturated by chance; both conditions lead to what are called here false zeros (Ridout *et al.*, 1998). False zeros are attributed to chance, sampling design and sampling or observer errors (Zuur *et al.*, 2009). The underlying assumptions of the ZI and ZA designs in this study are illustrated on Figure 2.1.

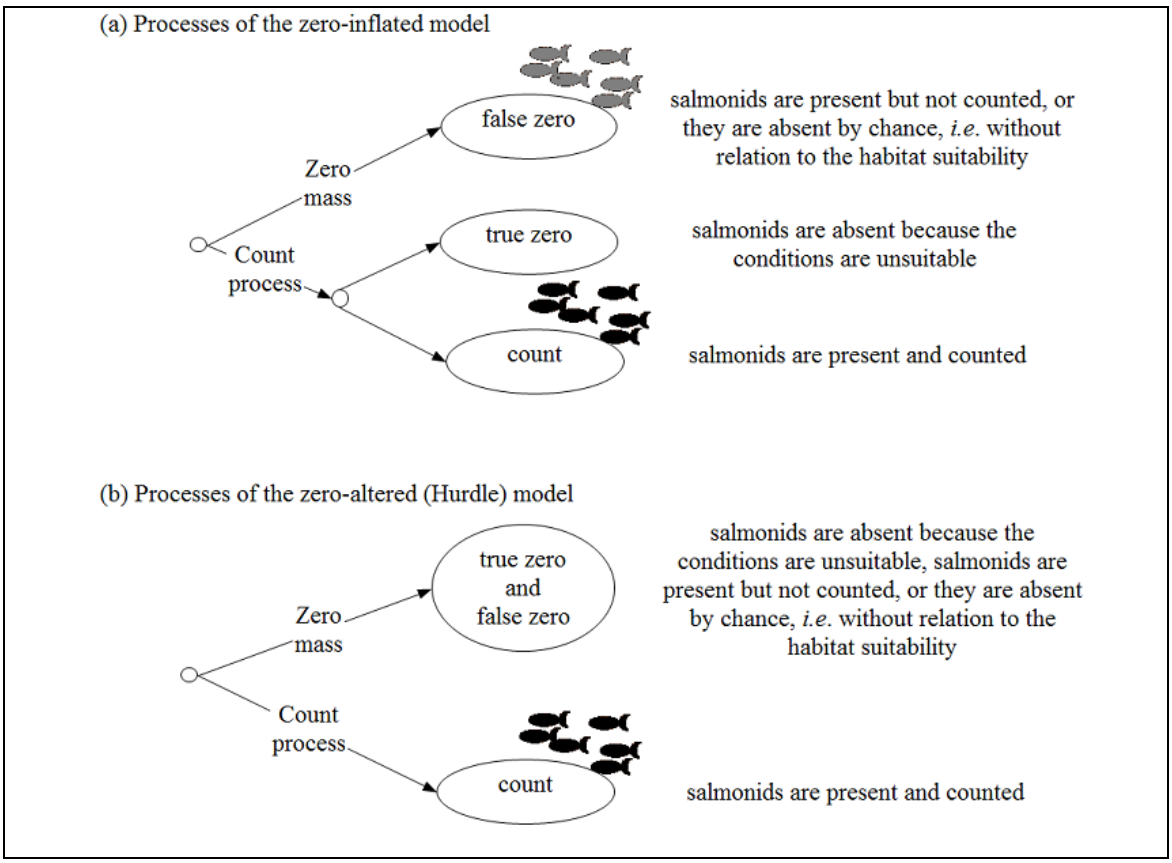


Figure 2.1: Illustration of the underlying principles of (a) mixture models (ZI) and (b) two-part models (ZA).
Based on the sketches by Zuur *et al.,* 2012 and Zuur *et al.*, 2009. The ZI design considers that a null value may correspond to the habitat being unsuitable or due to the experimental design; there is a dual process producing null vales. The ZA design does not make a distinction between the two types of zeros and considers one process is causing null values and another is causing positive counts.

Explicit modelling of the dual process generating null values may impact on the inferences of a study and greatly improves those (Martin *et al.,* 2005). For instance, inferences differ whether the non-occurrence of an event relates to the unsuitability of a habitat or missed detection (Zuur *et al.,* 2009) or else by sampling at-risk or not at-risk populations (Rose *et al.,* 2006). Mixture and two-part models have recently become popular (Zuur *et al.*, 2009) and are applied in varied fields. Agarwal *et al.* (2002) apply a ZIP model to spatial data and demonstrated that zero-inflation in a survey of isopod species is explained by the sampling location. Carrivick *et al.* (2003) analysed counts of injuries in a hospital; ZI models allow determining the factors affecting the risk of injury (*i.e* the variations in *zero* injury). Minami *et al.* (2007) use ZI models to account for the fluctuations in shark catchability responsible for null records not necessarily related to the fluctuations in the population density. Sheu *et al.* (2004) use ZI

models to distinguish smokers from non-smokers and reducing smokers who have not smoked during the observation period.

The aim of this chapter is to investigate salmonid movement behaviour at the counter located at Riding Mill, with a view to identifying whether movement is individual or grouped in nature; and to identify the processes influencing the movement of either type as appropriate.

As they return to spawn, the movement behaviour of salmonids and the conditions leading the different types of zeros may be addressed with GLM, mixture and two-part models. Seasonal variations in the migratory behaviour of salmonids during their upstream migration in the River Tyne, as observed during the spring migration in other rivers (Riley *et al.*, 2002; Moore *et al.*, 1998), may be investigated by defining multiple observation periods along the year.

2.2. Material and methods

2.2.1. Data

Data were recorded at 15 minutes interval during the year 2008. Salmonid counts were recorded by the electronic fish counter at Riding Mill and generated a univariate time-series that did not discriminate between *S. salar* and *S. trutta*. Seven observation periods (*a* to *g*) of four days duration (*i.e.* 384 measurement points each) were distributed throughout the migration period (Figure 2.2).

The first observation period occurred before the start of the migration season and the other six observation periods were two replicates of three moments of the salmonid migration in the River Tyne. The samples reflected three phases of the migration (initial increase in counts, middle and ending periods) and one prior to the migration. There were no missing values in the datasets.

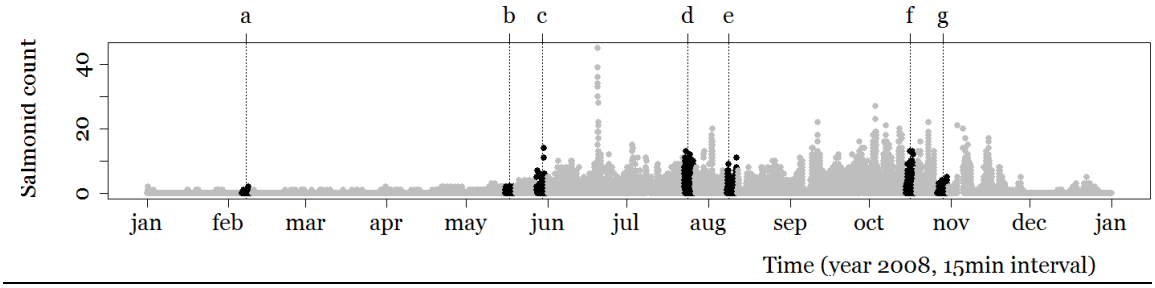


Figure 2.2: The salmonid counts recorded at Riding Mill during the year 2008. The seven observation periods of 4 days appear in black.

The covariates of interest for this study were of local influence: flow, temperature, tidal state, turbidity, colour, pH, conductivity and solar irradiance. The river parameters that were measured at a distance from the counter (flow and temperature at Bywell, turbidity and colour at Horsely) were lagged according to the distance from the counter (respectively 1.36km and 9.27km) and the flow regime at the time of measurement before being incorporated in the models. Measurement lags were determined by the distance between the measurement device and the counter (Appendix A.i, Table A.1) and the dominant flow regime occurring during the observation period in relation to the annual average (Appendix A.i, Table A.2, temporal variations of flow Appendix A.i, Figure A.7.1).

The counter at Riding Mill is located 41.9km upstream from the Tyne estuary and 11km upstream from the tidal limit at Wylam (Figure 1.2, page 11). To determine the amount of time between the salmonids reaching the counter and the time at which the tidal state may have influenced counts, the fit of hourly lags of the tidal values were compared through a complete tidal cycle (12 hours). The lags reflected various durations prior to the salmonid being counted, augmented by an undetermined number of tidal cycles. The most relevant lag was first determined by AICc comparison of different lags in the full model. Then, only the relevant lag only was used in the model for each observation period. The travel time between the estuary and the counter was likely influenced by the river velocity and the tide. However the length of the river along which the tidal state was in influential was not determined: so, the travel time was not approximated further beyond the relevant tidal lag.

2.2.2.Models

Initially, a classic data exploration was undertaken, which was concerned with the description of the data distribution and correlations for each observation period as for the whole year 2008 (after Zuur *et al.*, 2009).

Following this, the relationship between the observed salmonid counts and the selected environmental covariates was investigated for each observation period (*a* to *g*). The first step of the analysis consisted in selecting the most parsimonious model type for each observation period. The model types were Generalized Linear Models (GLM), Zero-Altered (ZA, also called two-part-) models and Zero-Inflated (ZI, also called mixture-) models.

GLM are suited for modelling count data; over-abundant zeros may however lead to a lack of fit due to the incompatibility of many zeros with the error distributions available for this type of model (Zeileis *et al.*, 2008). In this case, ZA and ZI models may be of better fit. The two-part (ZA) models consider a process responsible for the generation of zeros, and another distinct process generating counts. The zero mass is modelled as a binomial model of the probability that a null count is recorded and the count part is similar to a GLM on positive counts only (Zuur *et al.*, 2009).

The mixture (ZI) models also contain two parts. However this time, the null values are modelled by both part of the model. The count part accounts for the error distribution except for the excess zeros: it contains the counts corresponding to a Poisson or negative binomial distribution, including null values. The process generating the excess zeros, which are not explained by the count part distribution, is modelled by a binomial model, *i.e.* the zero mass. For this reason, mixture models remain applicable if there are no false zeros (Zuur *et al.*, 2012). The zero mass hence models the probability of recording a false positive against other types of data (Zuur *et al.*, 2009). For both mixture and two-part models, the count part was referred to as "μ" and the zero mass as "π".

The error distribution was assumed either Poisson or negative binomial, the systemic part of the model was specified by the aforementioned environmental covariates, and there was a log-link function between the mean and the predictor function, which ensured that fitted values are non negative. The aggregation coefficient (θ) of the negative binomial distribution quantified the level of aggregation contained in the count data (Zuur *et al.*, 2009; Ridout *et al.*, 2001).

Full models, that contained all covariates, were initially developed for the six model types (GLM, ZA and ZI with either Poisson or negative binomial distribution of the error for each). The models were compared using the Akaike Information Criterion (AIC, Greaves *et al.*, 2006; Burnham and Anderson, 2002; Sakamoto *et al.*, 1986). AIC quantifies the relative loss of information between a set of candidate models (Buckley *et al.*, 2003). AICc was corrected for small samples so as to penalize more for extra parameters than AIC, and avoid overfitting. Considering n the number of points in each time series and k the number of parameters in the model:

$$AIC = -2.log - likelihood + 2k \qquad AICc = AIC + \frac{2k\,(k+1)}{n-k-1}$$

The difference in AICc scores between models allowed the selection of the best approximating model type amongst the six, for each observation period (Rushton *et al.*, 2004; Burnham and Anderson, 2002).

The influence of covariates was then investigated. First, the optimal lag to apply to the tidal state covariate was determined. The Riding Mill station is located on the non-tidal part of the River Tyne; the tidal state was consequently of

influence downstream from the counter. The distance or time between the recording of a given salmonid and location or moment at which the tide was influential is undetermined. In case of the tidal state being influential of salmonid counts at Riding Mill, it is assumed that the travel time between the location where the tide is influential and the counter is a combination of the moment of the tidal state cycle that the salmonids selected for initiating upstream movement and an additional amount of hours, related to the travel conditions within the river which modulate the travel speed of the fish. The optimal moment of the tidal state cycle was determined for each observation period. The tidal state data were lagged for every hour of a complete tidal cycle. Each lag was tested within the full model for every observation period and the optimal tidal lag was determined using AICc ranking.

Second, step-wise deletion was then used to remove non-significant covariates and produce a minimal model, that only contained covariates assessed as significant by the Wald test (*P*-value<0.05). The process generated a set of candidate models for each observation period. AIC was again computed for each candidate model, and used to select the most parsimonious model, or set of models, based on a combination of log-likelihood and number of parameters within the model (Rushton *et al.*, 2004; Burnham and Anderson, 2002). A Pearson X^2 test allowed quantification of the contribution of each covariate to the fit of the model (Breslow, 1996).

Goodness of fit was further described by the Pearson correlation coefficient and the coefficients of the linear regression of the observed versus predicted values (Zuur *et al.*, 2009). Graphical tools were used to validate the underlying model assumptions (scatterplots of variance residuals with a view to assess their normality, temporal patterns, homoscedasticity and independence) in addition to the statistical tests, Zuur *et al.*, 2010; Zuur *et al.*, 2009).

2.2.3.Software

The GLM, ZI and ZA models were generated in the MASS (Venables and Ripley, 2002) and pscl (Zeileis *et al.*, 2008) packages operated in the R.2.15.1 environment (R Core Team, 2012). Underlying parameter estimation was similar for both packages (maximum likelihood).

2.3. Results

2.3.1. Data exploration

The seven observation periods contained high proportions of zeros (20.3% to 98.7%) and the count data for the observation periods *c* to *f* were overdispersed (Table 2.1).

Table 2.1: A summary of the count data for each observation period.
The ratio of the variance over the observed mean indicated overdispersion when greater than 2.

Observation period	a 6-10 Feb	b 16-20 May	c 28-31 May	d 23-26 July	e 08-11 Aug.	f 15-18 Oct.	g 27-30 Oct.
Range of counts	0-2	0-2	0-14	0-11	0-11	0-13	0-5
Total number of fish	6	39	285	1044	547	805	166
Count mean	0.02	0.10	0.74	2.72	1.42	2.09	0.43
Count variance	0.02	0.12	2.27	5.58	3.38	6.16	0.74
Mean / variance	1.00	1.26	3.09	2.05	2.38	2.94	1.74
% Null counts	98.70	91.67	64.84	20.31	42.45	35.16	73.70
% Single counts	1.04	6.51	18.49	18.22	23.70	18.22	15.35
% Counts>1	0.26	1.82	16.67	61.47	33.85	46.62	10.95

From *a* to *d*, the number of salmonids increased and the proportion of null counts decreased then the opposite pattern occurred until October (except for the period *e* in August which departed slightly from this trend). During the middle of the migration season, count data were overdispersed, had a broader range and more records of two or more salmonids.

Over the whole year 2008, strong correlations existed between the covariates flow and turbidity (0.7), and pH and conductivity (0.7); a lower correlation existed between colour and pH (0.6), and colour and conductivity (0.6). The covariates flow and conductivity were slightly correlated (0.5) and all other correlations were smaller than 0.5 (Appendix A.i, Figure A.7.2). All parameters were included in the full models; if correlation occurred to the point of preventing model convergence, then the covariate conductivity was omitted. Conductivity was less interpretable and independent in comparison with pH and colour (conductivity may represent particles density that may change water colour and pH depending on the nature of the particles).

2.3.2.Selection of the model types

The full models for each observation period (GLMs Poisson and negative binomial, ZAP and ZANB, and ZIP and ZINB) were ranked according to their AICc score (Appendix A.i, Table A.3). The models generating the lowest AICc scores are presented in Table 2.2 along with their AICc scores and differences with the second-ranked models.

Table 2.2: The optimal model type for each observation period, based on model ranking according to AICc scores.

	a	b	c	d	e	F	g
Selected model type	ZIP	ZIP	ZINB	ZINB	ZINB	ZANB	ZIP
AICc	47.4	228.2	801.2	1469.9	1022.5	1239.0	487.5
ΔAICc with second-ranked model	2.9	4.1	29.4	11.6	42.7	0.3	5.1
AICc weight	0.728	0.887	1.000	0.997	1.000	0.534	0.928

Given the high AICc weights and high ΔAICc values between first and second-ranked models, model selection uncertainty was low for most observation periods. For the observation periods *a, b* and *g*, the full ZIP model was of better fit amongst the candidate models; for the periods *c, d* and *e*, the ZINB model was of better fit. The ZAP and ZANB models could not be computed for the period *e*, explaining the very high ΔAICc with the second-ranked model.

Model selection uncertainty was high for the observation period *f* due to inconclusive ΔAICc and AICc weight. The first-ranked model was a ZAP model and both distributions of the error led to a comparable fit (Appendix A.i, p. 182). The ZANB model was retained (the ZAP model was nested in the ZANB model; in case of a better fit of the ZAP model the dispersion parameter θ becomes infinitely high or insignificant since the negative binomial variance $v_{NB}=\mu+\mu^2/\theta$ becomes closer to Poisson variance $v_P=\mu$).

2.3.3.Lag of the tidal state

The optimal tidal lag was determined using AICc ranking of full models containing one of the 12 lags (Table 2.3, full list and Figure A.7.4 in Appendix A.i).

Table 2.3: The optimal lag of the tidal state covariate for each observation period.

Observation period	a	b	C	d	e	F	g
Selected lag	2h	4h	0h	8h	5h	6h	9h

2.4. Analysis of the models

2.4.1. Comparative analysis of observation periods

Selected outputs of the most parsimonious models are presented in Table 2.4; the comparison of fitted and observed values indicated a good fit of the trend.

Observation / Fit assessment	a 6-10 Feb.	b 16-20 May	c 28-31 May	d 23-26 July	e 08-11 Aug.	F 15-18 Oct.	g 27-30 Oct.
Obs. μ	0.02	0.10	0.74	2.72	1.43	2.09	0.43
Fitted μ	0.03	0.10	0.76	2.71	1.42	2.08	0.43
Obs. variance	0.02	0.13	2.27	5.58	3.39	6.16	0.75
Fitted variance	0.03	0.03	0.49	1.85	1.46	3.12	0.27
Adj-R^2 lm (observed~fitted)	0.43	0.27	0.13	0.35	0.46	0.56	0.37

The variance was however underestimated, explaining the failure of the model to detect peaks, and the outliers (Table 2.4). The observation period *f* had the highest coefficient of determination (Adj.R^2=0.56).

Table 2.4: The comparison of selected outputs from the most parsimonious models for each of the seven observation periods.

The temporal fluctuations in observed and fitted values of salmonid counts are displayed in Figure 2.3 for each of the seven observation periods. The models detected the global trends but failed to account for peaks, as the difference between observed and fitted variance suggested.

The temporal fluctuations in observed and fitted values of salmonid counts are displayed on Figure 2.3 for each of the seven observation periods. The models detected the global trends.

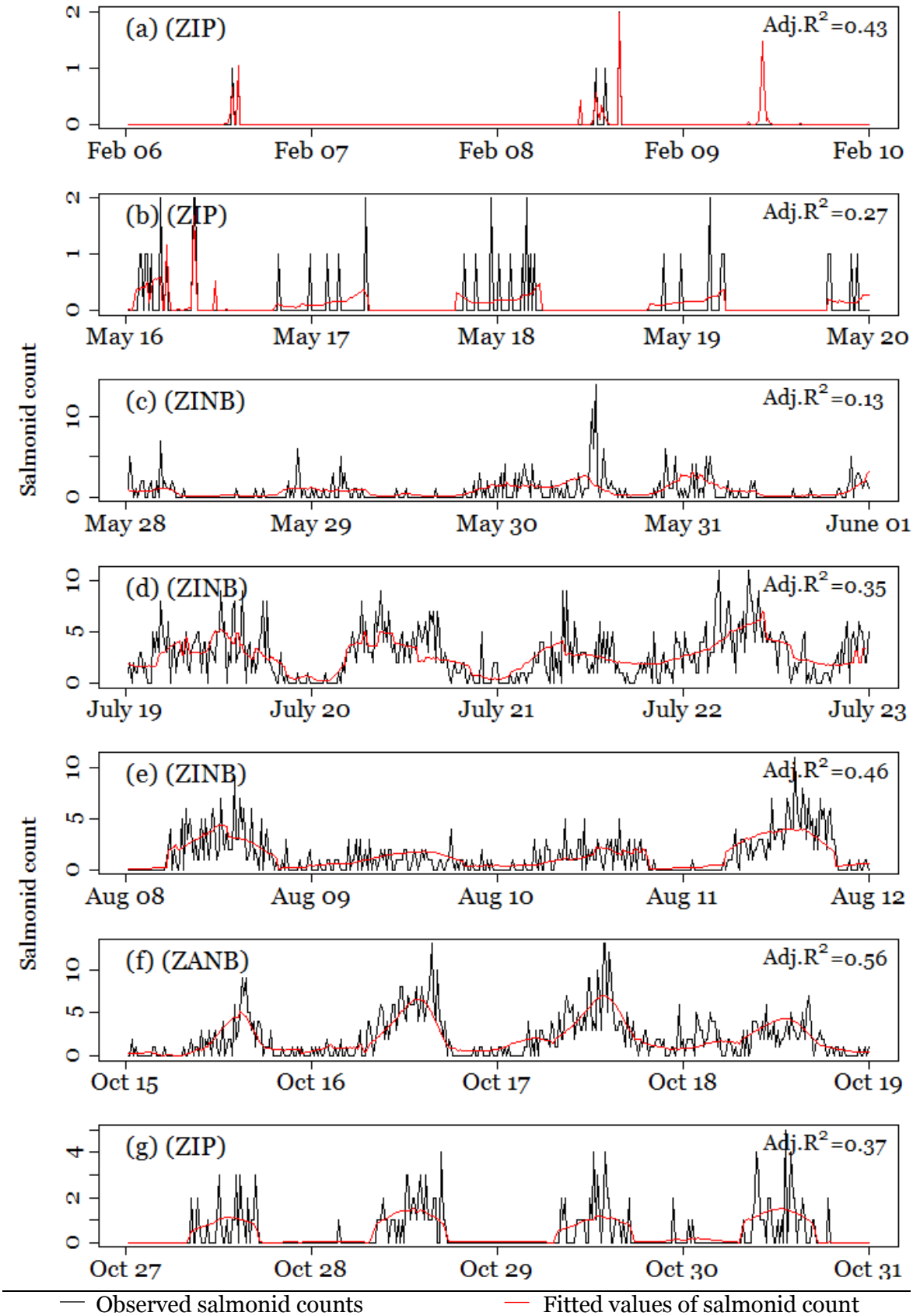


Figure 2.3: The observed and fitted salmonid counts for each observation period.

The interval between counts is 15 minutes. Note the difference in scale (y axis), with fewer fish at the beginning and end of the migration season. Adjusted R^2 values are reported for the model of each observation period.

The sign and significance of the parameter estimates are presented in Table 2.5. The values and associated level of significance and the likelihood tests quantifying the significance of the contribution of each parameter to the models are presented in (Appendix A.i, Table A.4 and Table A.5).

Table 2.5: The parameters estimated by the most parsimonious models for the seven observation periods.
The sign of the parameter estimates is reported along with the associated level of significance (** for $P<0.05$, * for $0.05<P<0.10$, others have P-value>0.10). The model type is the type of the most parsimonious model selected for each observation period. ZI stands for zero-inflated, ZA for zero-altered, P for Poisson and NB for negative binomial.

	Observation / Model period / Parameters	a 6-10 Feb.	b 16-20 May	c 28-31 May	d 23-26 July	e 08-11 Aug.	F 15-18 Oct.	g 27-30 Oct.
	Model type	ZIP	ZIP	ZINB	ZINB	ZINB	ZANB	ZIP
	Parameter estimates							
Count Mass (μ)	Flow	-	.	.	+**	.	-**	-**
	Temperature	.	.	+**	-**	+**	.	+*
	Tide	+	+	+**	.	.	-**	.
	Turbidity	.	.	.	+**	.	.	.
	Colour	+	-**	.	-**	.	.	.
	pH	.	-*	-**	-**	-**	.	.
	Conductivity	.	-**	.	+**	.	-**	.
	Solar irradiance	+	+**	-**	+**	+**	.	+**
Zero Mass (π)	Flow	-	+*	-**	-*	.	-**	.
	Temperature	-	-*	.	.	+**	-**	-*
	Tide	+	+*	+**	+**	+*	-**	.
	Turbidity	.	+	+**	.	.	+**	+**
	Colour	+	+	.	.	.	.	.
	pH	.	+*	.	.	-**	+**	.
	Conductivity	.	+	.	-*	.	.	.
	Solar irradiance	+	+*	+**	-**	-**	-**	-**
	Theta	.	.	0.98	8.87**	9.87**	18.64**	.

The optimal error distribution differed throughout the year. For the observation periods *a, b, c* and *g*, the calculation of Theta was either not applicable (Poisson distribution for periods *a*, *b* and *g*) or not estimated with significance (period *c*). The models for the observation periods *d, e* and *f* produced varying and highly significant Theta values. The ratios of the mean squared counts over the Theta values quantified the intensity of the aggregation of the salmonids crossing the counter. For the period *d* the product of the ratio was 1.21 $\left(\frac{\mu^2}{\theta} = {}^{7.34}/_{8.8}\right)$ and for the periods *e* and *f* they were 0.21 and 0.23 (respectively ${}^{2.04}/_{9.87}$ and ${}^{4.37}/_{18.64}$); the aggregation was the most important for period *d*, and was lower for the periods *e* and *f*.

The period *f* was the only period when the ZA model type was the most parsimonious. The ZA model did not include null values within the count process so the overdispersion was assumed to be due to high values only. This was confirmed by the mean and the range of counts (μ=2.09, 0-13, Table 2.4) and the high proportion of groups of salmonids (49.62%) during this period.

The number of significant predictors also varied along the year and more predictors were significant towards the middle of the migration season.

The amount of solar irradiance was a recurrent predictor of both the counts and the probability of excess zeros. During observation *a*, both the number of salmonids and the probability of excess zeros increase with more daylight (*i.e.* salmonids were in higher number during the day but were recorded less often than during the night). Temporal variations in salmonid counts (Figure 2.3b) confirmed this assumption. During observation periods *c*, higher solar irradiance was associated with fewer counts and a higher probability of excess zeros (*i.e.* the activity was mostly nocturnal and most null counts were recorded during the day). Increased solar irradiance was associated with higher salmonid counts and lower probability of excess zeros for the observation periods *b, d, e* and *g* (*i.e.* the salmonids moved through the counter during the day and most null values are recorded at night). The solar irradiance was not influential of the salmonid counts during observation *f* indicating that the activity was likely occurring all day long at comparable numbers, with fewer null counts recorded during the night. Shorter days in relation to nights also may limit the ability to detect a trend related to daylight.

Tidal state was an important predictor of excess zeros. In ZI models, excess zeros were all false zeros: for the observation periods *a* to *f*, a higher tidal state was associated with higher probabilities of excess zeros within the count data.

Turbidity was associated with more excess zeros for all observation periods except *d* and *e*, in the middle of the migration season. Higher turbidity was associated with higher counts for observation *d*.

The covariate flow did not affect the probability of excess zeros only in the two instances for which the flow regime was high; flow likely only had an effect at lower values. Flow values for these periods (*e* and *g*) were not extreme; the

respective ranges were 37.5m³.s⁻¹ to 153.0m³.s⁻¹ and 57.9m³.s⁻¹ to 191.0m³.s⁻¹, with 9.1% and 22.6% of the values above 120m³.s⁻¹.

2.4.2. Comparative analysis of selected covariates

Fluctuations in pH values were positively correlated with salmonid counts for the observation periods *b, c, d* and *e* (Table 2.5). During these observation periods the pH values fluctuated daily (Figure 2.4) and reached the daily peak in the afternoon (except for period *e*, Figure 2.4c, no decrease on two nights).

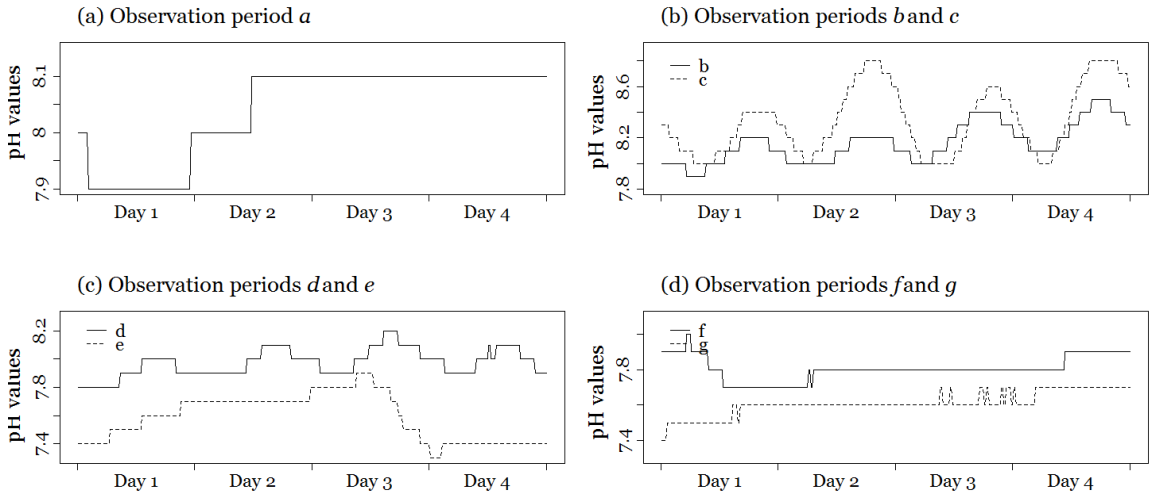


Figure 2.4: The temporal variations of the pH values during each of the seven observation periods.
The scales are different and the observation periods *b* to *d* show daily oscillations.

For period *e*, higher pH values also led to a lower probability of false zeros; the counts on the second and third days did not present the same pattern as the first and fourth day (Figure 2.3e). Higher pH was associated with an increased probability of null counts (both types of zeros confounded) during period *f*. The small range of the fluctuations may be linked to a bias due to the short observation period. The ranges of pH values (Figure 2.5 (a)) indicated that the observation *f* presented a relatively small range of values; the pH range and fluctuations within it may hence be influential in itself, rather than the pH values alone.

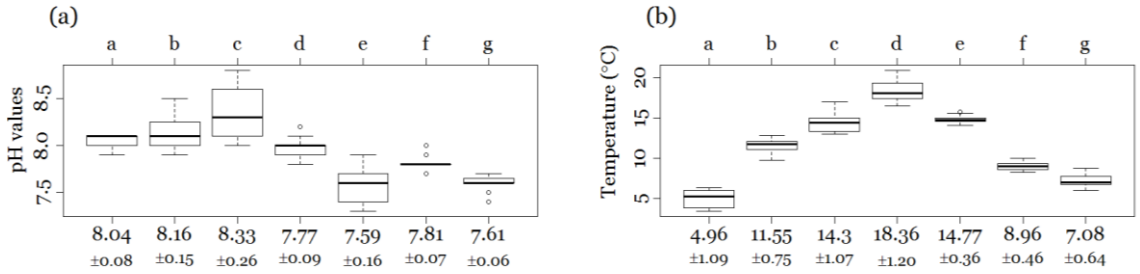


Figure 2.5: The range of pH and temperature values for each of the seven observation periods.
The mean ± standard deviation is given for each observation period.

Higher temperatures were associated with significantly more salmonid counts for observations *c* and *e* and significantly fewer counts for observation *d* (Table 2.5). The mean temperature increased until July and decreased afterwards; the temperatures were highest during the observation *d* (Figure 2.5 (b)). The thermal ranges for periods *c* and *e* were lower than for the period *d*, with all temperatures above 13°C. During observations *b, e*, and *f,* temperature was linked to the probability of excess zeros.

2.4.3. Goodness of fit of the most parsimonious models

Table 2.6 shows the coefficients of the linear regression of the predicted versus the observed values and the resulting adjusted R^2 for the most parsimonious model of each observation period.

Table 2.6: The assessment of the fit of the models for each observation period.

Observation period	a	b	c	d	e	f	g
Adjusted R^2	0.4331	0.2720	0.1335	0.3452	0.4603	0.5621	0.3725
Slope	0.55	1.03	0.80	1.02	1.03	1.05	1.01
Intercept	0.00	-0.05	0.14	-0.06	-0.05	-0.09	-0.00

Values for the regression slope were overall close to 1, except for the observation periods *a* and *c* for which it was indicative of an underestimation (0.55 and 0.80). Intercept values were close to null except for the observation period *c* (0.14). The Pearson variance residuals showed temporal patterns (Appendix A.i, Figure A.7.4), they were not normally distributed and presented outliers (Appendix A.i, Figure A.7.5).

2.5. Discussion

Models were developed to describe the relationship between the detection of salmonids at the counter located at Riding Mill at different stages in the migration season, and the conditions occurring simultaneously in the River Tyne. The models underlined the role of several riparian covariates in the salmonid behaviour within the River Tyne; the comparison of seven observation periods along a year indicated temporal variations in the migration behaviour of salmonids and their response to covariates.

During the initial observation periods in early spring (pre-migration season), the river covariates provided a poor description of the fluctuations in the number and presence of salmonids in the River Tyne; there were few fish to model. A better description was obtained for the following observation periods occurring during the migration season (end of May until October). The initial arrival of salmonids was likely mostly influenced by marine covariates (Jonsson and Jonsson, 2011) while the rest of the migration had more influence from river and coastal water conditions which may fluctuate jointly.

On downstream migration of both *S. salar* and *S. trutta*, the duration of residency varies spatially (distance from the estuary) and temporally (seasonally, shortens towards the end of the migration season) (Moore *et al.*, 1998). Here, the duration of residency was not estimated directly. Rather, the probability of excess zeros was assumed to reflect the river covariates that engendered a down period for migration upstream towards the counter at Riding Mill. Recurrent parameters involved in the generation of excess zeros, and presumably causing residency, were solar irradiance, tidal state and temperature. Salmonid movement behaviour is described first in relation to the environmental parameters, second in terms of the processes involved.

2.5.1. The salmonid movement behaviour in relation to the environmental parameters

Horizontal movements of water are a combination of currents: the river flow, the oscillating tidal currents, the general circulatory system of the oceans and currents generated by varying meteorological conditions (Bodwitch, 1802).

Previous studies suggested that the effect of tide on salmonids is due to fluctuations in water rate (Salveit *et al.*, 2001), however this information was not available for this study. The direction of the tidal current is not directly related to tidal state and flood and ebb do not necessarily occur at the same time as the rise and fall of the tide; inferences from the present models can only concern the variations in water level associated with the tide. The tidal state likely influences the ability for salmonids to access shallow parts of the river however it may not affect the speed or position of the fish (as currents do) and, conversely, the covariate flow measured for this study affected only the speed or position of the salmonids (not the accessibility of shallow waters).

Flow speed is a recurrent factor associated with salmonid migration. Fast natural flow triggers downstream migration (Jonsson *et al.*, 2007; McCormick *et al.*, 1998); flow speed directly impacts energy demands as salmonids adapt their swimming behaviour (Haro *et al.*, 2004), and a minimum flow rate conditions the movement of salmonids (Heggenes *et al.*, 1996). In this study, the flow regime was high for two of the seven observation periods. Values were not extreme (periods *e* and *g* 37.5-153.0$m^3.s^{-1}$ and 57.9-191.0$m^3.s^{-1}$) and respectively 9.1% and 22.6% of the salmonid counts occurred at a flow rate greater than 120$m^3.s^{-1}$ (thought to be the high limiting flow value in the River Tyne, Environment Agency, *per. com.*, 2011). This covariate did not affect the probability of excess zeros during these observation periods. This may indicate that flow values were within suitable monitoring conditions; speed was not so low as to limit the salmonid presence and not so high so as to hinder movements (false zeros) or allow bypassing the counter. The only observation period to be modelled with a ZA design (*f*, mid-October) presented particular conditions in the combination of low flow and high amount of null values (over a third of the count data) and provided the best-fitting model (R^2=0.56). The ZA design did not differentiate between types of zeros (Zuur *et al.*, 2009). This possibly indicated that the observation period involved null values due to low flow, which were properly accounted for by the model.

For most observation periods (*a* to *f*), increasing tidal state was associated with higher probabilities of excess, excess zeros within the salmonid count data. The tidal state is not a covariate commonly measured for the study of salmonids; as opposed to tidal currents for which various stages are known to be preferred

among specific populations of *S. salar* (Smith and Smith, 1997; Potter, 1988) and *S. trutta* (Moore *et al.*, 1998). The duration of observation periods (4 days) implied that the fundamental periodicity of tidal cycle (12hours, 25minutes) was very similar to the half-periodicity of diurnal oscillations and the cycles of temperature and solar irradiance. The significance of tide as a predictor of salmonid movements suggests that tide influenced the upstream movement of salmonids within the portion of the river that is influenced by the tide (*i.e.* tidal river). The tidal river was too broad and undefined to estimate the salmonid travel time between the estuary and the counter, which was likely related to both the solar cycle and the tidal cycle.

Temperatures play a key role in energy management and the spatio-temporal distribution of ectotherms (Hutchinson and Maness, 1979), generating species-specific requirements of thermal habitats. Higher temperatures associated with higher salmonid counts during observation periods *c* and *e* which presented thermal ranges that were relatively high (14.31°C±1.07 and 14.77°C±0.36); it can be assumed that these thermal ranges constituted a suitable thermal habitat. During the warmest observation period (*d*), an increase in temperature was associated with fewer salmonid counts. It was determined that for *S. salar,* a thermal habitat of 16.0°C is comfortable and higher temperatures constitute a stress; beyond 17˚C heart rate at rest and general metabolism increase and resistance to stress is reduced; at 18.4°C *S. salar* seeks cooler areas, *i.e.* thermal refuges (Anderson *et al.*, 1998). For the warmest observation period *d*, higher probability of false zeros occurred with higher temperatures. This suggests that the thermal range (18.36°C±1.2) was beyond the thermal limit typical to *S. salar*. The high temperature range may have cause a period of residency downstream from the counter, either within a thermal refuge or to wait for thermal conditions to become more suitable. Temperature is also considered a major controlling factor in the migration of *S. trutta* (Byrne *et al.*, 2004). Over the seven observation periods, the optimal thermal habitat for salmonids in the River Tyne was likely between 13°C and 18°C.

River turbidity was directly linked to the visibility of salmonids by predators (Thorpe, 1994). High turbidity levels limit visual contact, which may cause *S. salar* to struggle to aggregate, as they cannot distinguish one another (Jepsen *et al.*, 1998). *S. trutta* may encounter similar difficulties and also modify their

foraging behaviour as they are visual feeders (Klemetsen *et al.*, 2003). Overall, turbidity may constitute a constraint as salmonids adapt their vertical distribution to the level of turbidity, *i.e.* deeper waters tend to be under-exploited if visibility is lower. This may represent a disadvantage when coping with predation (Thorpe, 1994) and may force the salmonids to disperse in a manner that is not the most energy efficient, and hence cause a time of residence.

pH was a recurrent predictor of migration, and negatively correlated with salmonid counts. Extreme pH values may induce problems with olfaction and behaviour (pH=5.1, Royce-Malmgren and Watson, 1987) but the ranges observed were likely not extreme; so, it appears that the variations in pH contributed to variation in counts rather than the pH values alone.

Respiration produces carbonic acid and photosynthesis absorbs it, causing the water to become more acidic or more alkaline (Wurts, 2003). This suggests that temporal variation in pH was likely linked to biological activity and representative of the interaction between respiration and photosynthesis that varyoccur in the River Tyne daily and seasonally. During the first observation period in February, pH was relatively high and no temporal trend was observed. This observation represents the few live species of occurrence during winter. During the following periods (*b*, *c, d,* and *e*) pH fluctuated daily. Carbonic acid released during respiration and photosynthesis likely caused the river pH to increase during the day and, as the photosynthesis stops at night, the pH was reduced every evening until the next sunrise when the cycle started again (Wurts, 2003).

pH values were erratic for the rest of the season (*f* and *g*), which may relate to the high amount of biological activity occurring at the late stage of the migration season. At this stage, salmonids were recorded both at night and during the day. Daylight was also the longest of the year, suggesting that photosynthesis took place during a longer part of the day. The global pH values were low and reflective of the advanced stage of the production season (high biomass and respiration, Wurts, 2003). Overall, it is likely that the pH values in the River Tyne were both explanatory of the salmonid activity and a by-product of the salmonids presence.

Salmonid activity took place at different times of the day during the observation periods. Recorded movements were initially solely nocturnal (*c*), then mostly diurnal (*d* and *e*), then recorded all day long, with daylight being irrelevant to variations in numbers (*f*); movements were solely diurnal during the last observation period (*g*). It is known that the seasonal downstream migration of *S. salar* is initiated by the light (Riley *et al.*, 2002; McCormick *et al.*, 1998). This was also likely the case for the upstream migration in the River Tyne, with a tendency for diurnal activity for the first part of the migration then for nocturnal activity. Upstream migration preferably takes place under minimal predation pressure (Smith and Smith, 1997; Potter, 1988). Diurnal migration during warm conditions may be an attempt to minimize predation such as birds (McCormick *et al.*, 1998). Conversely nocturnal movements may be a strategy to avoid reduced flow (Metcalfe *et al.*, 1997).

The preference for a given time of day for salmonids to be active matched the river flow in other studies, *i.e.* nocturnal activity in summer and diurnal otherwise corresponded to respectively low and high flow periods , suggesting that preferred conditions consequently related to the optimum visibility (Smith and Smith, 1997; Potter, 1988). Absolute water level and change in water level were found to be major controlling factors in the migration process of *S. trutta* (Byrne *et al.*, 2004). This pattern may not be applicable to the present study due to the minimal river flow is maintained by water releases from Kielder Water (Haile *et al.*, 1989).

2.5.2.Processes of the salmonid movement behaviour

Within the ZI and ZA models, the negative binomial distribution allowed for overdispersion within both the count part of the model and excess null counts. In contrast, the ZIP model (Poisson distribution) implied that overdispersion was restricted to the zeros (both true and false) and in the ZAP model, overdispersion existed in the zero mass only (Zuur *et al.*, 2009).

The goodness of fit of the negative binomial distribution to the datasets *d, e* and *f* was emphasized by the relevance of the θ values to model these observation periods. The importance of the aggregation was given by the value of the ratio; a larger ratio value representing a higer difference with a Poisson distribution and

a higher aggregation. The negative binomial distribution in ZI and ZA models implied that overdispersion was present. This implied a number of both null and high counts values, which corresponded to a grouping phenomenon.

Theta fluctuated along the migration season of 2008. This heterogeneity may indicate instability in the aggregation, potentially because either not all salmonids aggregate and several distinct grouping behaviours may occur and cannot be quantified at once. For instance, the level of aggregation may be species-specific (*i.e.* two species have a different rate of aggregation and *S. trutta* may displace *S. salar* during the months of July to October, when the counts are estimated most aggregated, E.A., *per. com.*, 2012). Age is discriminatory of the start time in salmon (*e.g.* grilse start before kelt in the Dee, Scotland, Jonsson *et al.*, 1990); aggregation may vary with the age-class.

It is also possible that salmonids do not aggregate constantly and at all times. During the third period (*c)* salmonids may be transitioning from solitary to grouping behaviour (given the Poisson distribution prior to *c* and the negative binomial after it). As the salmonid population is composed of two species, each composed of several age groups. Grouping patterns may also depend on the photoperiod such as seen in caged *S. salar* (circular schools during the day, slower and shallow swim at dusk and interruption of the school, Juell and Fosseidengen, 2004).

The sampling design does not allow definite inferences regarding aggregation behaviour. This is because the ability of the models to differentiate between random aggregation and shoals may be questioned, as the aggregated counts may not be representative of natural salmonid groups. For instance, two salmonids located at each end of the counter array, passing the counter at close interval, may be recorded as a pair, while their common occurrence may not have an ecological meaning.

Fluctuations in salmonid counts were mostly modelled with a ZI model design which implied a dual process behind the generation of null values: it was hypothesised that excess zeros were related to the inability of salmonids to reach the proximity of the sampling area (*i.e.* the counter) and that the true zeros described unsuitable conditions in the vicinity of the counter.

Counts of both species of salmonids were modelled jointly: null values corresponded to records of both species simultaneously absent. Excess zeros may consequently imply that the ecological characteristics of the river habitat considered unfavourable by the model are in fact the conditions that are unfavourable to both species (Zuur *et al.*, 2010). If the habitat preferences common to both species were modelled, the species-specific requirements likely induced false zeros in the sense that the description of the river conditions were more suited to a species than the other. During the spring migration in the Kumijoki River (Finland) *S. salar* moves slower than *S. trutta* (Bazarov and Golovanov, 2003). It may be assumed that the response of *S. salar* to river conditions is hence delayed in relation to the response of *S. trutta*.

In this study, the proportion of excess zeros varied with observation period, associating the notion of ecological site with the sampling time of year. Outside the breeding season, conditions in the River Tyne are unfavourable in that they are not associated with the salmonid migration. The anadromous salmonids concerned in this study migrate long distances in the North Sea for *S. trutta* and even further for *S. salar*. Excess zeros are consequently linked to the nature of the sampling period. This is confirmed by the non-significance of environmental covariates in February, when the conditions in the River Tyne likely do not affect the salmonids. Excess zeros were not suitably modelled (by covariates as by a constant intercept), indicating that marine parameters or time may have been relevant to detection probability. The high resolution of the observation however rendered the incorporation of broad range climate indices such as NAO or AMO rather imprecise.

River parameters only start having a significant influence of the excess zeros from the end of May. It can be assumed that from this time of year onwards, environmental conditions brought that salmonids to the river, and the river conditions do play a role in whether or not the fish are counted at Riding Mill. Consistent with literature about the downstream migration, recurring covariates of influence regarding the excess zeros are tide, flow, and solar irradiance. Temperature also intervenes but was linked to solar irradiance in previous studies, and is investigated further later in this discussion.

The study of individual movement required measurement intervals to be small. It was argued that the duality of the count process (one generating false zeros,

the other generating counts and true zeros) should not rest on the selection of the scale of the analysis (Lord *et al.*, 2005). For instance, counting car crashes over 1/10 of a mile will result in a greater proportion of zeros than over a mile (Lord *et al.*, 2005).

Here, the spatial scale was imposed and the temporal scale was intentionally restricted in order to capture local and individual movement (if salmonid counts were considered at longer intervals the dataset would contain a smaller proportion of null values). Salmonid movement across the counter are dependent upon the downstream conditions experienced by the individuals. In this regards, the process is dual state; being counted by the counter is conditional of experiencing conditions allowing individuals to reach the counter. For this reason, downstream conditions were associated with the generation of false zeros. All salmonids on their way to the spawning grounds in the River Tyne necessarily pass through the counter, so the location of the counter on the River Tyne was considered relevant to the study without a bias relating to the distance (as opposed to the study by Lord *et al.*, 2005).

The spatial scale however implied that the conditions recorded corresponded to the habitat description by the counter only, not downstream from it. The values were not lagged because the travel time of salmonids is within the River Tyne is not documented and the level at which the covariates have influenced the salmonids behaviour is also undetermined (what level of the river for what parameter etc, would result in numerous possible models). It was assumed that the measurements were representative of the conditions in the whole river, with a more or less extended lag; if this lag was of less than 4 days then the model would have picked the relationship between the false zeros and the covariates, even if the relationship is lagged.

False zeros implied the non-detection of the fish by the counter, or the presence of fish belonging to an "always zero" group. In the latter case, the fish did not get the opportunity to reach the sampled area (*i.e.* the counter). This may either be because the conditions were unsuitable beforehand (but they may be suitable within the sampled area) or because the fish remained in the river for a period of residency.

The opportunities for the production of excess zeros existed at many levels of the upstream migration, long before reaching the counter (Figure 2.6).

As salmonids return to the Tyne estuary, sea conditions dictate whether the fish will reach the estuary and when it will occur (Figure 2.6 (I)). No data were available to account for this stage, which would also involve lags hardly determinable. The lack of information regarding sea conditions may explain the contribution of parameters to the model's fit despite not being significant for the first two periods, as these are essential particularly to the first periods studied (migration triggers occurring at sea) but not included in the model.

So, salmonids enter the river immediately or after a residency of variable duration (Figure 2.6 (II)), river covariates that potentially influenced this behaviour were measured upstream from the estuary and consequently occurred before the salmonids experienced them. However, cyclical covariates such tidal state and solar irradiance remain relevant. Preferable conditions for the entering the river through the estuary were likely linked to minimal predation pressure (Smith and Smith, 1997; Potter, 1988). So, low visibility in the water may have played a role: flow and light, turbidity and colour may be relevant.

Once in the river, salmonids were subject to freshwater conditions and progressed upstream according to how suitable the resulting riparian habitat was. Unsuitable conditions may generate two types of zeros. True zeros correspond to environmental that lack suitability at the sampling site. False zeros correspond to environmental conditions preventing fish from reaching the sampling site. Consequently, the influence of the habitat suitability may be in the shape of true or false zeros according to the distance to the counter. The closer to the counter, the more likely that zeros are true, as they correspond to the conditions of the sampling area. However unsuitable conditions responsible for the fish to not pursue its travel at a location downstream from the counter (in a chronological order of occurrence that may be different from the one illustrated on Figure 2.6) engendered false zeros (Figure 2.6 (III)).

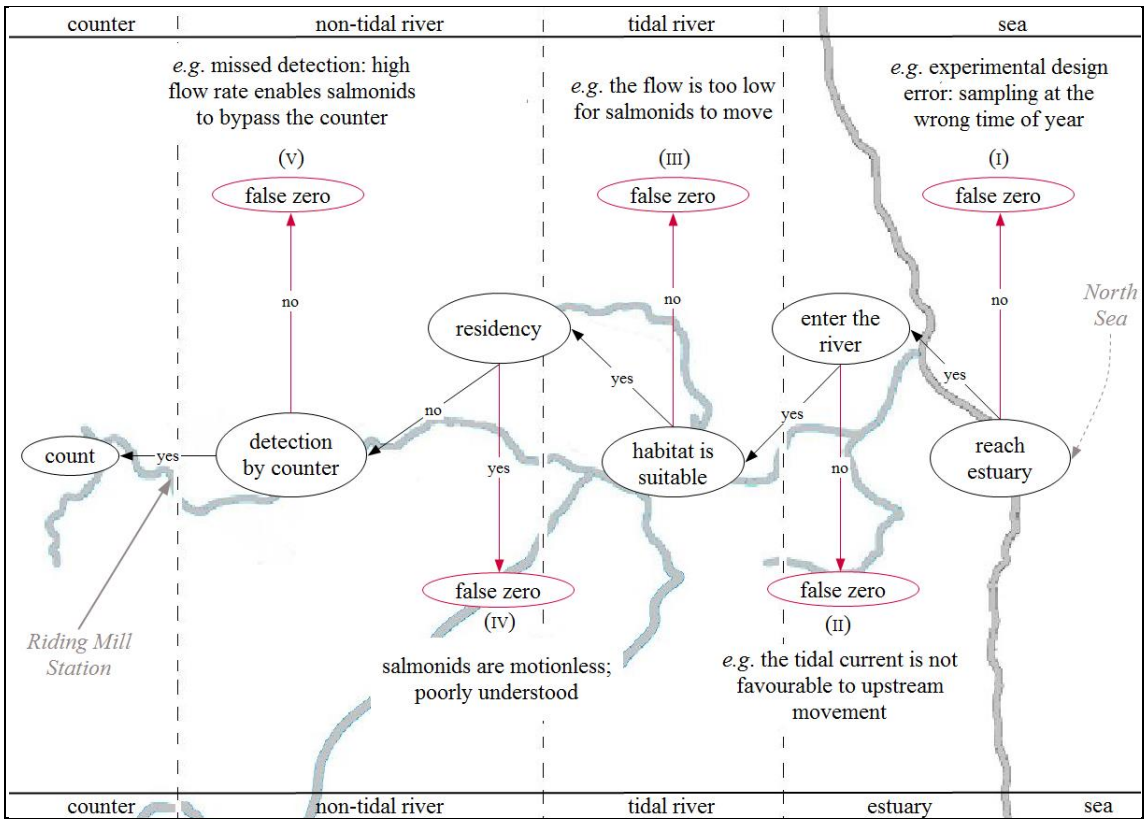


Figure 2.6: The opportunities that may generate false zeros along the upstream migration of salmonids in the River Tyne.
The scenarios that engender false zeros are in red. For a count to be recorded within an observation period, the fish must reach the estuary, enter the river, find the habitat suitable, not stay in residency and be detected by the counter; otherwise a false zero will be generated.

Within suitable river conditions, the salmonids may either progress upstream or remain resident in the river, in which case false zeros were produced as the fish does not reach the counter regardless of the suitability of conditions (Figure 2.6 (IV)). Residency periods are prolonged stops that can occur several times during upstream migration. Residency patterns appear river-specific and also differ between individuals. In South-West England salmonids travel close to the spawning grounds and rest for a considerable period of residency before completing the last part of their journey (McCormick *et al.*, 1998). In North Norway, salmonids migrate upstream to spawning grounds with several pauses, then move without apparent consistency around the spawning area, with a considerable residency period before spawning (Okland *et al.*, 2001). Benefits and influences on the residency period are still unclear (Fleming, 1996).

Travel speed is different between species, with *S. salar* being generally slower (Bazarov and Golovanov, 2003). As both species counts were recorded jointly, the reduced speed of *S. salar* may generate excess zeros, given that the time unit per salmonid was implicitly considered identical by the models for all

individuals, and consequently for both species. Slower individuals may be considered absent while conditions are favourable.

Finally, salmonids that reached the counter may have missed detection (Figure 2.6 (V)). High flow values potentially influenced the non-detection of fish due to their ability to by-pass the counter when water levels rise; also, the counter may occasionally underestimate the number of salmonids by recording single counts when two fish or more pass the electrodes simultaneously, or in a manner that may be associated with a larger fish instead of several smaller fish (Environment Agency, *per. com.*, 2010). This situation may be more likely when the salmonids are more abundant, and so the impact on the analysis may be negligible.

2.6. Conclusions

The study used joint counts of two species of salmonid *S. salar* and *S. trutta* to investigate the behaviour of both species.

Tide, solar irradiance, and temperature were cyclical variables whose periods likely became confounded considering the duration of the observations; their effects were significant albeit potentially intricate, mingled and consequently difficult to contrast within the models.

The aggregation behaviour and the response of salmonids to the river covariates varied throughout the year. The level of aggregation of individuals increased towards the middle of the season. This increase was associated with either a higher abundance in aggregating salmonids, or with a higher frequency of the aggregating behaviour in the salmonids.

Whilst this study described the short-term relationship between individual salmonids and the River Tyne conditions, the drivers of the annual salmonid migration remain to be identified and will be investigated in the next chapter.

Chapter 3. .

"The end was contained in the beginning."

G. Orwell, 1949.
1984

Chapter 3. Describing the River Tyne Salmonid population movements and the impact of the New Tyne Tunnel construction works.

3.1. Introduction

The Salmonidae Family is characterised by its salinity tolerance, regardless of whether the species is anadromous (Tanguy *et al.*, 1994). *Salmonids* are part of the *Salmo* taxon, which is one of the 11 genuses of the Family (Integrated Taxonomic Information System, 2010). Of the *Salmo* taxon, two of the 30 species are of particular interest here: *Salmo salar* and *Salmo trutta*.

Important intra-specific variations exist amongst geographically isolated salmonid populations. Differences in *S. salar* populations have been attributed regarding phenotype, behaviour, development, biology, among other life history trait variations (Taylor, 1991). *S. trutta* possess a trait variation from which a great aptitude for colonisation results. Habitat range and use, body size, growth rate, migration patterns, differ within and amongst populations (Klemetsen *et al.*, 2003). Genetic differentiation is possible between some populations and has been used as a tool to demonstrate the genetic evolution within the species that can be a result of exposure to specific local conditions (in *S. trutta*, Nielsen, 1998). *S. salar* is often cited as an example of *local adaptation engendering genetic variation* (Elliott *et al.*, 1998; Hutchings and Jones, 1998).

Both *S. salar* and *S. trutta* are thought to have been geographically isolated in the past but, due to a global decrease in habitat quality, were forced to share an increasing proportion of spawning ground over the last decades. Natural hybridization between *S. salar* and *S. trutta* now happens fairly commonly wherever the species are sympatric (Garcia-Vazquez *et al.*, 2001; Jansson and Ost, 1997) such as in the River Tyne. The hybrids are believed to be the result of either error in mating choice because of the large female trout body (Jansson and Ost, 1997) or inter-specific competition, where up to 65% of trout eggs may be fertilized by salmon parr (Garcia-Vazquez *et al.*, 2001).

Considerable differences between populations have an established genetic origin related to the local population isolation and its adaptability to a given environment. Salmonid populations over a specific geographic range may hence be studied as a whole biological entity. This chapter considers salmonids of the River Tyne as a single population.

Anadromous salmonid populations undergo two important migrations. A first migration, after 1 to 4 years in the river, is to the ocean where they become mature and fit enough to migrate a second time 1 to 3 years later to their natal river and breed. Triggers of salmonid migration remain incompletely understood. The most influential parameter may be photoperiod, guiding the periodicity of migration, but conditions of water temperature, turbidity and flow also intervene and are called "*releasing factors*" (McCormick *et al.*, 1998). Salmonid migration back to the River Tyne for spawning occurs annually, over several summer months. Many peaks in salmonid abundance are observed within a migration season; these vary slightly between years both in intensity and time of occurrence but always according to annual periodicity.

The annual migration of anadromous salmonids involves significant energetic costs in term of swimming, osmo-regulation, and increased mortality due to adaptation to different habitats (McCormick *et al.*, 1998). The periodic migration is consequently expected to present compensatory biological advantages that imply a population dependency on periodic natural parameters.

The growth of Atlantic salmon follows a light-pituitary axis, whereby the endocrine system is regulated photo-periodically (Bjørnsson, 1997). Longer daylight increases growth hormones level while the periodicity of daylight regulates growth and optimises relationship between body size and length (*i.e.* a "condition factor"). Variation in daylight also influences behavioural and physiological seasonal changes and adaptations throughout the year. Photoperiod, via its effects on growth hormone, is also directly linked to osmotic pressure adaptability, increased dominant feeding behaviour, inhibited anti-predator behaviour in juveniles, stimulation of swimming, coping with starvation, sexual maturation (smoltification, secondary sexual characters), and healthy metabolism (Bjørnsson, 1997). In sea-run *S. trutta*, lower light levels are thought to trigger upward swimming and downstream migration (Moore *et al.,* 1998).

The periodicity(ies) contained in River Tyne salmonid counts over time will be investigated using the wavelet analysis. This relatively new spectral method from the 1980's (Cho and Chon, 2006) detects scale-dependent regularities contained in a time-series. The result is a quantification of the periodicity relative to each data point over the study period. The process compares to a Fourier transform new ability to represent multi-scale, multi-frequency and non-stationary patterns (Grinsted *et al.*, 2004). These properties make the wavelet method suited to model patterns presenting transient and/or short-lived periodicities accross many fields of application (*e.g.* climatic data, Bradshaw and McIntosh, 1994; geophysics, Grinsted *et al.*, 2004; ecology, Cazelles *et al.*, 2008).

Trends in abundance of salmonids are strongly linked to mortality at sea; marine conditions as post-smolts enter the Baltic Basin are particularly important (Friedland *et al.*, 1998). In *S. salar*, the highest mortality occurs during the first two weeks at sea (28%) after which it is thought to decrease quickly (Frieland, 1998).

As ectothermic fish, the marine habitat requirements of salmonids are largely described by the thermohaline circulation; post-smolts disperse as a function of sea surface currents, temperature, and salinity. Atlantic and coastal waters dominate the surface currents over the North of the UK. Atlantic water moves eastwards then south along the coast and coastal waters move south along the east coast (Holm *et al.*, 2000). In the Norwegian Sea, the distribution of post-smolts is determined by a combination of high salinities (>35ppt) and warmer temperatures (9-11°C) that correspond to saline Atlantic currents (Holm *et al.*, 2000). These parameters condition the optimum habitat for post-smolts in May, both at the downstream migration stage and as they return to the natal river to breed (Holm *et al.*, 2000; Frieland, 1998). The speed and direction of the surface currents also influence the dispersal of post-smolts; high velocities enable *S. salar* to reach northern latitudes quickly; salmonids also swim actively and may be held by feeding opportunities (Holm *et al.*, 2000).

Marine temperatures regulate the growth and maturation of salmonids, that in turn condition the ability to cope with competition within and between species (Friedland, 1998). Survival during the first year at sea is predominantly linked to broad scale ocean conditions. Suitability also implies precise timing; warm or

cold thermal habitats are more influential depending on the month of the year which is directly related to the age, stage of migration and location of *S. salar* (Friedland, 1998) and *S. trutta*. As the return migration starts many months before the start of the spawning season (Jonsson and Jonsson, 2011), the demands in habitat suitability extend over several months.

The ability to predict salmonid migrations requires an understanding of the basic biology and the factors potentially disturbing it. Delays in migration due to anthropogenic disturbances can have a high impact on the survival of *S. salar* smolts during the downstream migration (Marschall *et al.*, 2011). The construction of a tunnel crossing the River Tyne required specific works in the river during the end of the year 2009. The construction was likely the source for a temporary modification of the river as a habitat, for instance via increased sound levels, movement in sediments, vibrations, turbidity, and physical hindrance. The impact of this potentially disruptive period will be investigated at two levels. Firstly, the long-term population impact will be addressed with the broadest available time frame (2004-2011). Secondly, a more local study will be carried in order to describe the possible day-to-day consequences of the construction works on a more local and individual behaviour of the fish.

This chapter follows a progressive model building approach. First, the wavelet transform will be used to describe the periodicity of signal contained in River Tyne salmonid population time-series. The relevant periodicity will be incorporated into generalized linear models via harmonics. Generalized linear models involve a linear relationship between environmental variables and a response that is independently distributed. Mixed-effect models also contained fixed effects, comparable to the GLM components in that they were associated with the average effect of predictors on the response variable (Pinheiro, 1994). In addition, LME models contain variance-covariance components wthat allow for modelling of the covariance structure of the random effect terms and changing the intercept and slope of the global trend according to specific clusters. This feature makes LME models suitable for modelling dependent data, whether it is over time or among the subjects sampled (Fox, 2002). A simple generalized linear model design will be the basis for further modelling and potential addition of random effects. Investigation of the long-term effects of the tunnel construction works will use an intervention variable and an

interrupted time-series analysis. Short-term effects will be concerned with a limited period of time and use a model design to account for the numerous null values observed at this small scale of observation analysis.

3.2. Material and Methods

3.2.1. Data

Response variable and transformations

Salmonid counts were obtained from an electronic fish counter installed at the Riding Mill station. The study uses data collected over eight years (2004 to 2012); the response variable is a univariate time series.

A Box-Cox process was used to estimate the best normalizing power transformation of the salmonid count data, improving normality and equalizing variances. Such transformations are recognized useful for the robustness of the analyses, even non-parametric (Osborne, 2010). The count data were augmented by 0.5 in order to avoid infinities when taking logarithms (Breslow, 1996) and due to the impossibility of applying Box-Cox transformation to null values (after recommendations of Yamamura, 1999). Time-series of salmonid counts were also log- and Anscombe transformed for the GLMs.

Covariates and format

Models concerned with the study of the general migration pattern used covariates related to broad time scale. These were: temperature, flow, year, NAO values at various lags, and annual harmonics *(see Introduction chapter for theoretical justification*). Additional AMO data were also tested (Enfield *et al.*, 2001). All data were pooled into 2 week period bins. Data for the period 2004-2010 were used to develop the models and data for the year 2011 were used to test the forecasting ability of the models.

Models concerned with the short-term effects of tunnel construction used all available covariates (*i.e.* similar covariates to aforementioned, and turbidity, colour, conductivity) and data were binned at a finer grain (one day interval

between measurements). The dataset was a subset from 20th September 2009 to 10th February 2010. The dredging of the trench to accommodate the tunnel sections took place between the 9th November 2009 and 15th December 2009. This period potentially of greatest impact (EA, 2013) was incorporated to the dataset by a covariate with two categorical levels. A second covariate was incremented by one every day from the start date of the construction works.

3.2.2.Analysis

Wavelet analysis and periodicity

The wavelet transform provided a representation of signal periodicity within both frequency and time domains. Spectrum plots generated straightforward visualization of the fluctuations in periodicity over time. The harmonic analysis was based on a continuous wavelet transform that considered all possible scales (as opposed to the discrete transform). Morlet wavelets were used to detect periodic behaviour present in the response variables. The Morlet wavelet is defined as:

Equation (3.1) $$\psi o(\eta) = \pi^{-1/4} e^{i\omega_0 \eta} e^{-1\eta^* \eta/2}$$

ω_0 is frequency (dimensionless) and η is time (dimensionless). The wavelet is stretched in time by varying its scale: η= scale * time. The wavelet analysis was hence not restricted to stationary periodic events but also allowed the visualization of localized intermittent periodicities (Grinsted *et al.*, 2004).

Harmonic regression

Harmonic analysis used a Fourier series to decompose the signal contained in the time-series into harmonics made of sine and cosine waves. The resulting component was in the form of a trigonometric transformation that included the relevant period outlined by the wavelets analysis such as (period expressed in number of weeks):

Equation (3.2) $$\text{period of } x \text{ weeks} = a * \cos(2\pi * x/\text{week}) + b * \sin(2\pi * x/\text{week})$$

Covariates were then included and tested in the models. The coefficients *a* and *b* of the Fourier series were estimated and used to determine the amplitude *c* and

phase angle *d* of the harmonic by being combined into a single cosine function (Jakubauskas *et al.*, 2001):

Equation (3.3) $c = \sqrt{a^2 + b^2}$ $d = \arctan\left(\frac{b}{a}\right)$

$(+\pi \text{ if a} < 0;\ +0\ if a > 0)$

The value of the phase angle was then divided by a full period length (2π) and multiplied by 52 (*i.e.* the number of weeks in a year) to locate the peak predicted by the model after the start of each year.

Models and interrupted time-series analysis

GLMs were developed using stepwise removal of non-significant covariates from a full model containing all covariates, aiming to obtain a most parsimonious models containing only covariates for which *P*-value<0.05. AIC was calculated via the log-likelihood value for each model, with *n* being the number of points in each time series and *k* the number of parameters contained in the model: $AIC = -2*log\text{-}likelihood + 2k$

Models were assessed in term of overdispersion by the ratio of residual deviance over degree of freedom (indicative of overdispersion when above 1, Zuur *et al.*, 2009). Goodness of fit was quantified by the Pearson correlation coefficient and residuals were examined visually for homogeneity, normality and independence.

In LME models, all explanatory variables were initially contained as mixed effects and a minimal adequate model was designed by stepwise deletion of non-significant variables one at a time. An *F*-test of the likelihood ratio test and the Akaike information criterion (AIC, Sakamoto and Kitigawa, 1986) compared the minimal model to the initial full model. An additive forward stepwise approach was then followed to choose the covariates associate with the random effects explaining the cluster-to-cluster variability. This strategy followed previous studies (Buckley *et al.*, 2003; Pinheiro, 1994). Values of AIC were again used to decide whether or not to include the covariate. Maximum likelihood was applied in order to compare the models and eventually restricted maximum likelihood estimates were used as they included fixed effects in the calculation of the degree of freedom (MLE does not, Pinheiro, 1994).

Two-part and mixture models were designed according to a backward stepwise approach; AIC was used for model selection and the significance of each model component was quantified via a χ^2 test. A binary variable accounting for the tunnel construction works period was created and allowed applying an interrupted time-series analysis approach by testing the significance of the effect of the variable when included in the models (Wagner *et al.*, 2002).

3.2.3.Software

The wavelet analysis, GLM, LME and zero-inflated and zero-altered models were developed using, respectively, biwavelet (Gouhier and Grinsted, 2012), MASS (Venables and Ripley, 2002), nlme (Pinheiro and Bates, 2000) and pscl (Jackman, 2012) packages, operated in the R.15.1 environment (R Core Team, 2012).

3.3. Results

3.3.1. Harmonic analysis

Wavelet power spectrums and matching average wavelet power spectrums were generated for the time series of salmonid counts (Figure 3.1).

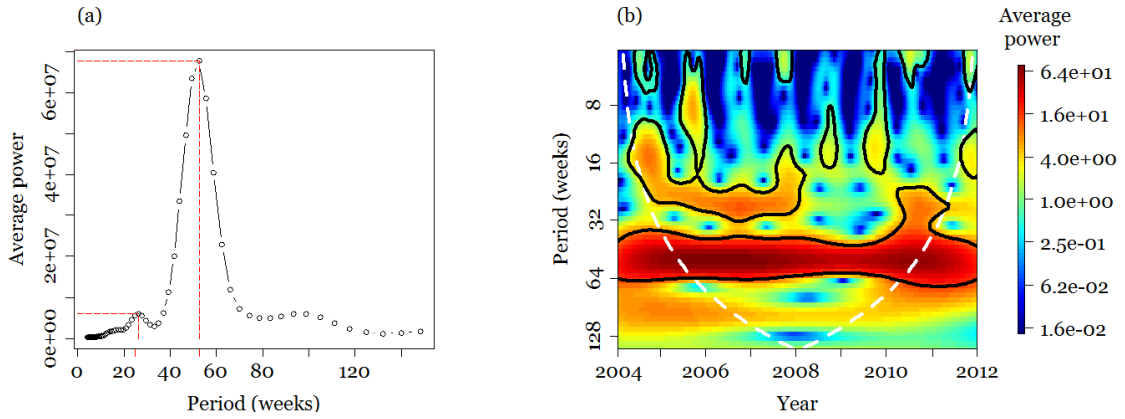


Figure 3.1: The wavelet analysis of the River Tyne salmonid counts over the observation period 2004-2011.
With (a) wavelet power spectrum and (b) associated average power spectrum. The interpretation was restricted to the cone of influence; highest power appeared in yellow through red on the spectrum and the lowest power was in blue. A thick black line defined the contours for significance at the 95% level for all power of all points; the *P*-value corresponded to the average power over a given period, which may vary in value and significance over years. The primary period was of 52.48 weeks and the secondary period was of 24.77 weeks.

Significant values were extracted (Table 3.1). The highest power average was for a period of 52.48 weeks (*P*-value<0.001) and appeared constant and strong throughout the years. Secondary periodicities were also observed (yellow, outlined regions on the average power spectrums). These were transitory periodicities, occurring more particularly during the first half of the whole observation period.

Table 3.1: Periodicity values of the time-series of salmonid counts, extracted from the wavelet transform for the observation period 2004-2011.

	Duration (weeks)	power	*Z*-value	*P*-value
Primary period	52.48	194.74	7.29	<0.001
Secondary period	24.77	7.94	1.06	0.14457

Consequently the cyclic variations observed in the time-series were likely not perfectly sinusoidal during the first half of the observation period (Diggle, 1990).

3.3.2.Response transformation

Computation of the maximum likelihood function and posterior distribution of Box-Cox transformed response variable indicated that a power transformation of λ=0.1818 was optimal (after Box and Cox, 1964, Appendix A.ii, Figure A.7.6). This power corresponded to count values incremented by 0.5 and was located closer to λ=0.5 when this value was lowered. Consequently, the more usual square root transformation was also applied.

3.3.3. Description of the River Tyne Salmonid migration

Generalized linear models

Outputs from the first-ranked GLMs according to AIC scores are presented in Table 3.2. The ratio of the residual deviance over the residual degree of freedom indicated a high level of overdispersion in the models, 200 to 400 folds for all but the log- and Box-Cox transformed time-series.

Table 3.2: Statistics of goodness-of-fit and overdispersion for the optimal GLM obtained for of salmonid counts according for transformations of the response.

	Residual Deviance	df	Ratio	Pearson Cor.	Normality test W,P	df
Count, Poisson	87721.0	196	447.6	0.67	1,<0.001	13
Count, neg. bin.	87679.7	196	447.3	0.68	1,<0.001	14
Log(count+0.5)	310.2	203	1.5	0.81	1,<0.001	7
Anscombe(count)	44470.0	205	216.9	0.64	1,<0.05	5
$\sqrt{count}$	44590.5	205	217.5	0.64	1,<0.001	5
$count^{0.182}$	70.8	204	0.35	0.81	1,<0.001	6

Transformations reduced overdispersion, with the log- and Box-Cox being more acceptable. The non-normality of the residuals distribution however indicated a failure in the model type (Figure 3.2). No parsimonious model was obtained using the GLMs design.

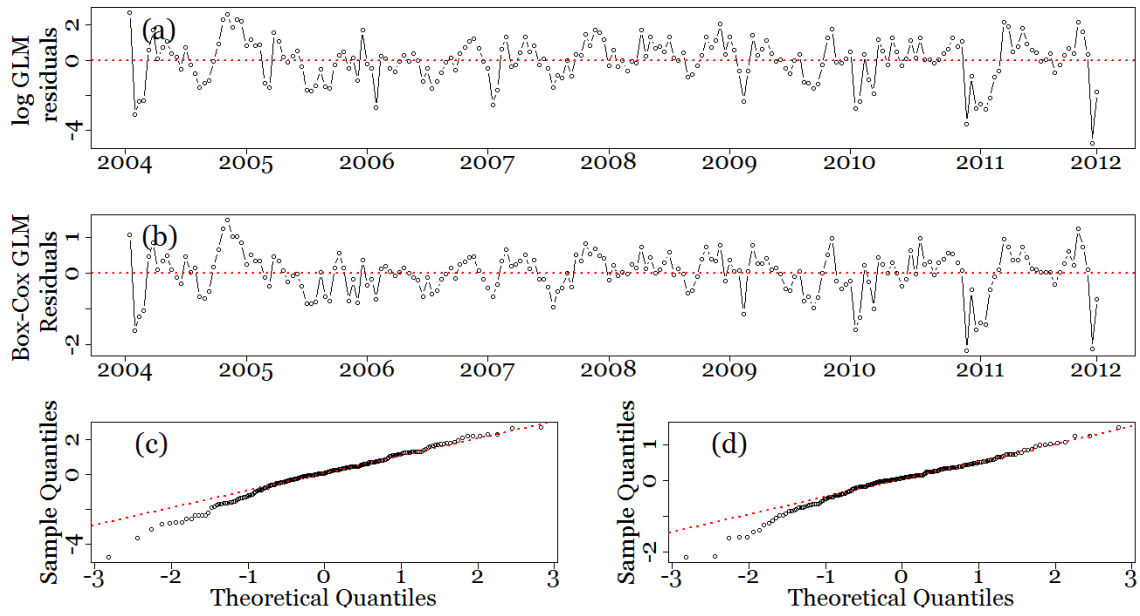


Figure 3.2: Temporal variations (a&b) and QQ-plots (c&d) of the residuals of the first-ranked GLMs, for log- and Box-Cox transformed salmonid counts.
The residuals were non-normally distributed for both transformation (c&d).

Linear mixed effect model, Box-Cox transformed counts

The most parsimonious linear mixed effects model included random effects of temperature and year (AIC= 238.37, ΔAIC<64.45 with the full model, with 9 parameters against 18). The random effects provided with a much better description of the data than a linear regression (F=13.99, P-value<0.005). The coefficients of random effects for the temperature covariate of each year (Table 3.3) indicated a difference in rate of increase of salmonid counts with temperature on an annual basis, in relation to the whole period of study.

Table 3.3: The random effects coefficients of the most parsimonious LME model of the Box-Cox transformed salmonid counts.
Values are given for intercept and temperature for each year.

	2004	2005	2006	2007	2008	2009	2010
Intercept	0.3109	0.145	-0.066	0.101	0.014	-0.011	-0.493
temperature	-0.015	-0.023	0.001	-0.010	0.004	-0.009	0.051

Likelihood ratio tests indicated significant relationships between the salmonid counts and annual periodicity (χ=175.00, df=2, P-value<0.001), temperature (χ=3.69, df=1, P-value<0.055), and NAO values with a lag of 4 years (χ=8.849, df=1, P-value<0.005). Coefficients of the fixed effects are presented in Table 3.4.

Table 3.4: Restricted maximum likelihood coefficients of the most parsimonious LME model of the Box-Cox transformed salmonid counts.

	Value	Std. Error	df	t-value	P-value
Intercept	2.160	0.238	172	9.081	0.0000
cos (1year)	-0.421	0.152	172	-2.764	0.0063
sin (1year)	-1.030	0.066	172	-15.637	0.0000
temperature	0.0779	0.023	172	3.384	0.0009
NAO4yr	-0.205	0.056	172	-3.682	0.0003

Fitted values for the period 2004-2011 along with the forecast of salmonid counts for the year 2011 by the same model are presented on Figure 3.3 for the back-transformed values. The examination of residuals distribution per year pointed a few outliers (Appendix A.ii, Figure A.7.7 and Figure A.7.8) which were identified on the plot of residuals over time (Figure 3.3).

Residuals values were centred around zero (Appendix A.ii, Figure A.7.7). The distribution of the model's residuals was not significantly different from normal (Figure 3.4a&b, Shapiro test W=0.99, P-value=0.568). A pattern remained when plotting the residuals against fitted values (Figure 3.4c). Autocorrelation was present (Figure 3.4d) and there was a strong negative correlation between the random intercept and slope, indicating that for an increase by one unit of

standard deviation of the intercept, the slope would decrease by 0.923 unit of standard deviation (Appendix A.ii, Table A.13). The annual periodicity was also closely correlated to other fixed effects, which was expected due to the cyclical nature of all the parameters. As high levels of correlation may be the sign of an ill-conditioned variance matrix, a serial correlation structure was added to the model. It accounted for autoregressive correlation without decay with greater temporal distance (compound symmetry). The component increased the AIC value (ΔAIC=2.00) and did not affect the coefficients nor the level of significance of the estimates, so it was not retained.

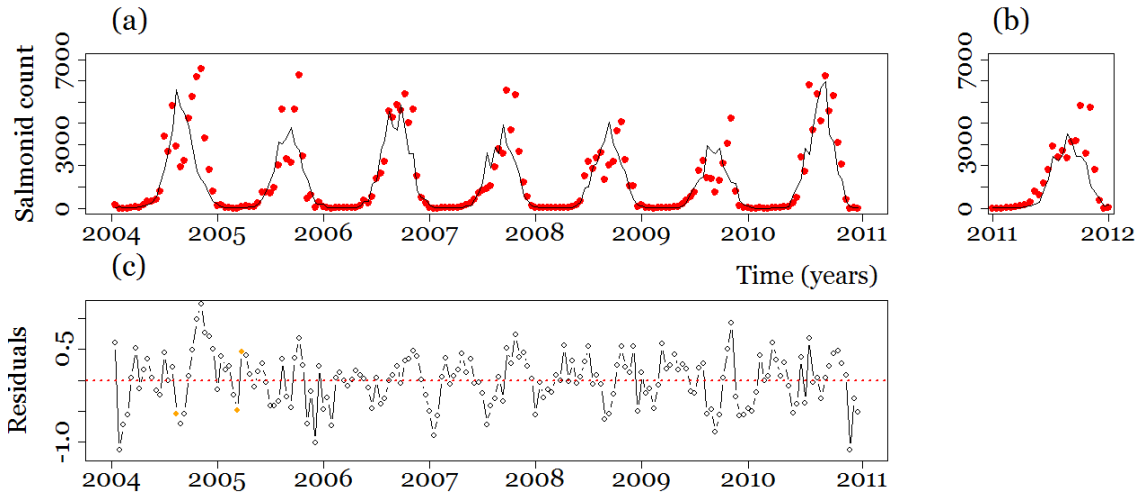


Figure 3.3: Predictions from the most parsimonious LME model of the Box-Cox transformed salmonid counts.
(a) Back-transformed fitted values based on the period 2004-2011, (b) forecast for the year 2011 based on the same model and (c) corresponding residuals of the model over the years 2004-2011.

The model was of good fit for the transformed values (Figure 3.5a) but the variance increased for higher values of the raw data (Figure 3.5b). The source of this variability could not be accounted for by the model. The forecasting ability was tested for year 2011 and appeared of similar goodness of fit (Figure 3.5c&d).

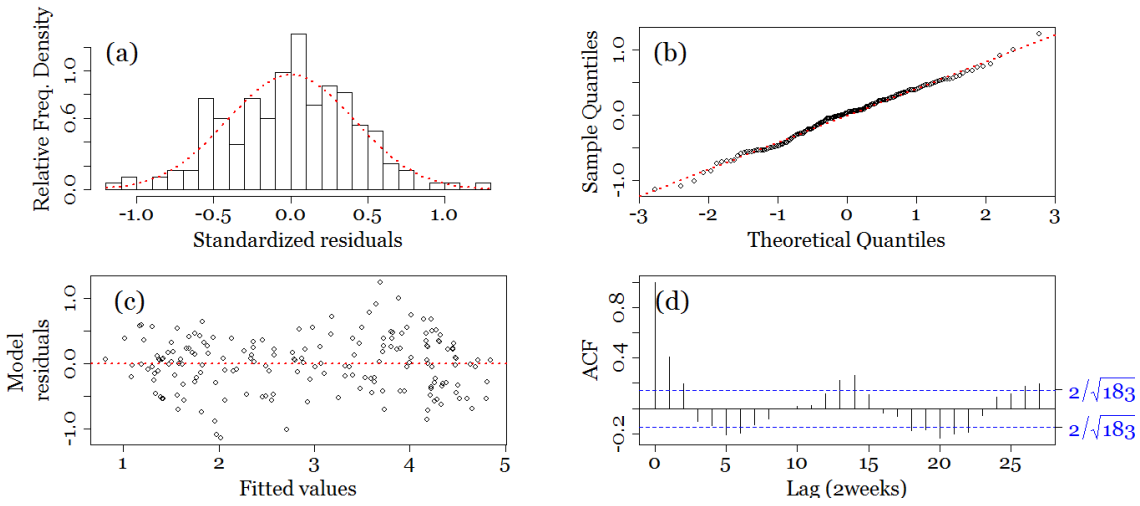


Figure 3.4: Diagnostic plots for the optimal LME model for the Box-Cox transformed salmonid counts for the observation period 2004-2011.
The residuals were non-normally distributed (a&b), and there was (c) some evidence of heteroscedasticity and (d) fading cyclical autocorrelation.

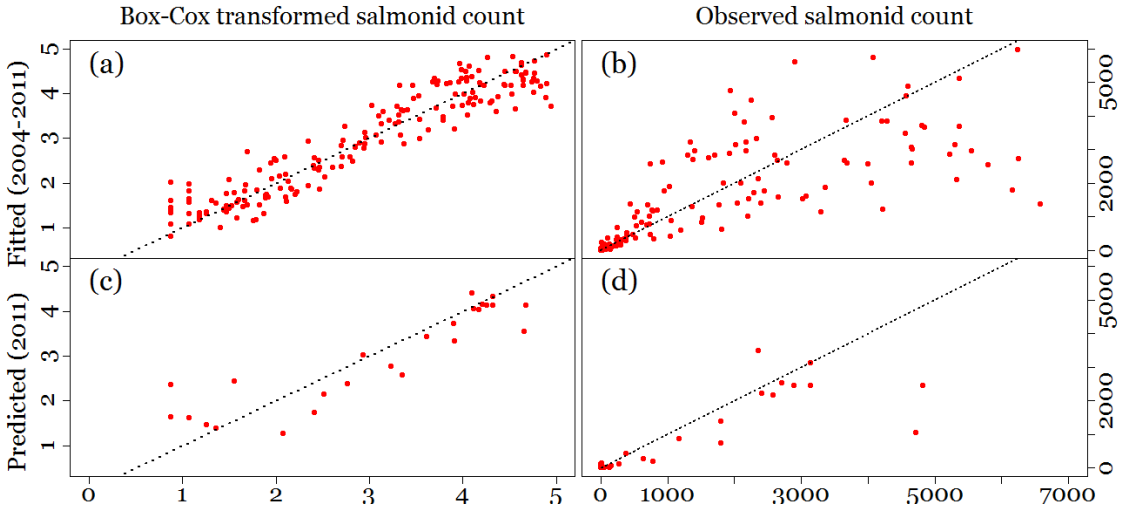


Figure 3.5: Fitted (2004-2011) and predicted (2011) values against corresponding observed values for the model of the Box-Cox transformed salmonid counts.
The figure shows the relationship between predicted and fitted values and Box-Cox transformed salmonid counts (a&c) and unchanged salmonid counts (b&d).

Linear mixed effect model, square root transformed counts

The most parsimonious model included random effects of temperature and year (AIC=1579.97, ΔAIC>26.54 with the full model). Again, random effects allowed achieving a better description of the data than a linear regression (*P*-value<0.005, *F*=15.50). The coefficients of random effects (Table 3.5) again indicated an annual difference in rate of increase of salmonid counts with temperature; estimated differences were more important than when using the Box-Cox transformation.

Table 3.5: The random effects coefficients of the most parsimonious LME model of the square root transofrmed salmonid counts for the whole observation period.
Values are given for intercept and temperature for each year.

	2004	2005	2006	2007	2008	2009	2010	2011
Intercept	-0.711	3.419	-0.039	1.868	-0.225	4.050	-5.621	-2.742
temperature	0.184	-0.582	-0.006	-0.307	0.018	-0.673	0.926	0.440

Comparable relationships existed between salmonid counts and annual periodicity (χ=223.05, df=2, *P*-value<0.001) and temperature (χ=6.5361, df=1, *P*-value<0.05). Estimates had lower values than for the previous model, in accordance with the scales of transformation as the power of the Box-Cox transformation was a smaller fraction than square root power. The tunnel construction works period appeared related to variations in salmonid counts (χ=13.7, df=1, *P*-value<0.001). The coefficients of these fixed effects are presented in Table 3.6.

Table 3.6: The coefficients for the fixed effects of the optimal linear mixed effects model of the square root transformed salmonid counts for the whole observation period.

	Value	Std Error	df	*t*-value	*P*-value
Intercept	15.279	4.558	197	3.352	0.0010
cos (1year)	-7.714	3.169	197	-2.434	0.0158
sin (1year)	-22.924	1.373	197	-16.698	0.0000
temperature	1.254	0.478	197	2.627	0.0093
tunnel	-18.363	6.176	197	-2.973	0.0033

The residuals distribution indicated a considerable lack of conformity to model assumptions. Residuals were non-normally distributed (Shapiro test W=0.9854, *P*-value<0.05, Figure 3.6a&b and Appendix A.ii, Figure A.7.12) and presented temporal patterns (Appendix A.ii, Figure A.7.13). A horn-shaped distribution pointed to a lack of independence and linearity of the variance (Figure 3.6c, outlining the relevance of a different power for the transformation); heteroscedascticity and autocorrelation were still present Figure 3.6d).

Correlation between effects remained (Appendix A.ii, Table A.14). The higher outliers on Figure 3.6b were identified as mostly underestimations of peaks during the migration season (Figure 3.7).

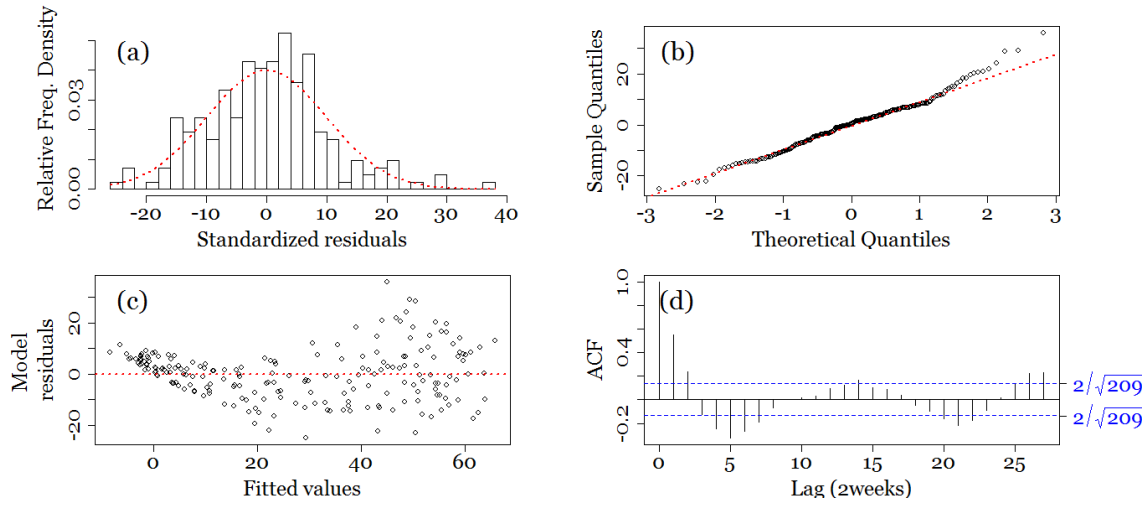


Figure 3.6: Diagnostic plots for the optimal LME model for the square-root transformed salmonid counts for the whole observation period.
The residuals were left skewed (a&b), and there was (c) a strong horn-shaped heteroscedasticity and (d) fading cyclical autocorrelation.

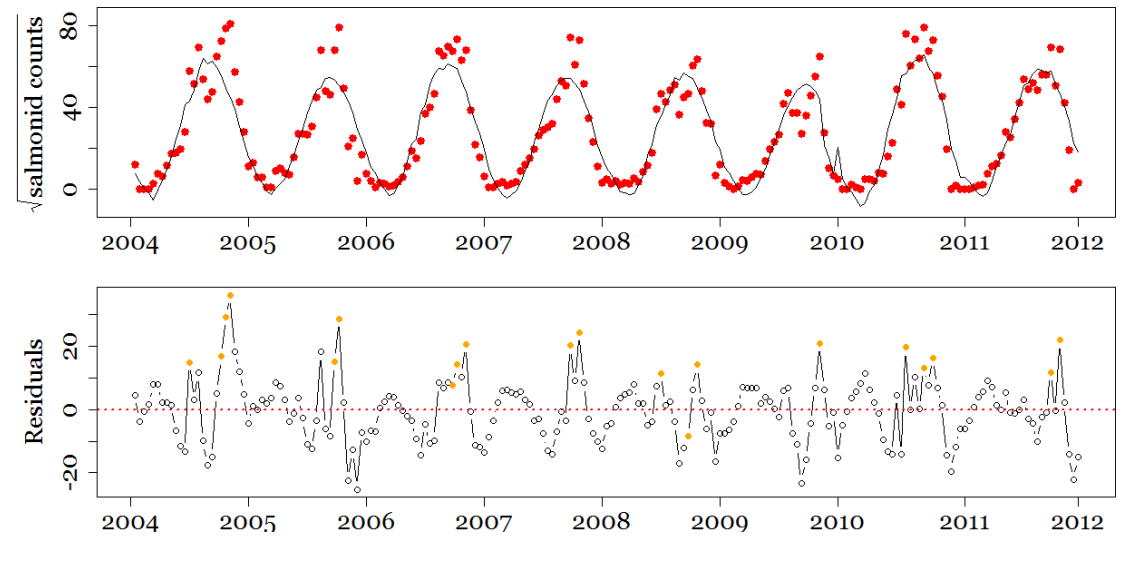


Figure 3.7: Predictions from the most parsimonious LME model of the root square transformed salmonid counts for the whole observation period.
Fitted values based on the period 2004-2012 and corresponding residuals over time.

3.3.4. Impact of tunnel construction on individual salmonid

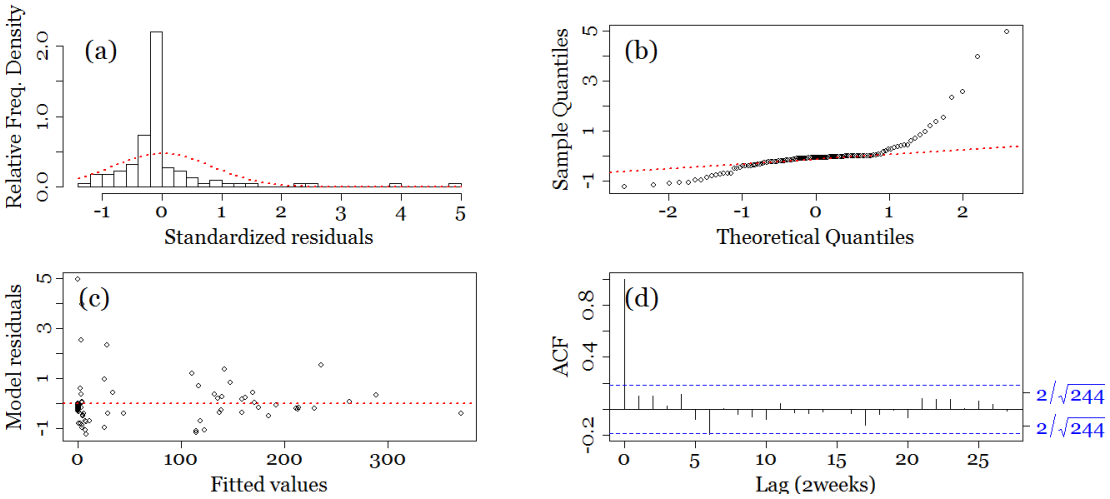
The hurdle model with a binomial distribution of the error was the most parsimonious (θ=1.962, *P*-value<0.01 for Log(θ)). The model did not differentiate between the types of zeros. The AIC between the full and the most parsimonious model was ΔAIC=7.57 (AIC=528.05, df=11).

The model coefficients are presented in Table 3.7 for both the count and zero hurdle parts of the model. Variations in salmonid numbers were negatively related to the values of NAO with a lag of 5 years (χ=5.71, df=1, *P*-value<0.05). The tunnel presented a small negative effect (χ=8.45, df=1, *P*-value<0.005) corresponding to a decrease of a third of a fish during the tunnel construction works (exp(-1.05)=0.35). The time since the tunnel construction works was linked to a lower count of salmonids (χ=41.41, df=1, *P*-value<0.001). The zero-part of the model estimated the ratio of probabilities for a non-null count (implied by the zero part) divided by the probability of a non-null count (implied by the whole model distribution). The odds of zero salmonids at the Riding Mill station increased with higher tidal state (χ=6.77, df=1, *P*-value<0.01), and during the tunnel construction (χ=3.994, df=1, *P*-value<0.05), with 95.5% more chances for null counts during tunnel period. Turbidity was associated with less null counts (χ=33.30, df=1, *P*-value<0.001) and the absence was strongly related to the annual periodicity (χ=79.17, df=2, *P*-value<0.001).

Table 3.7: The count and zero hurdle coefficients for the model of salmonid counts during the period 20th September 2009 to 10th February 2010.

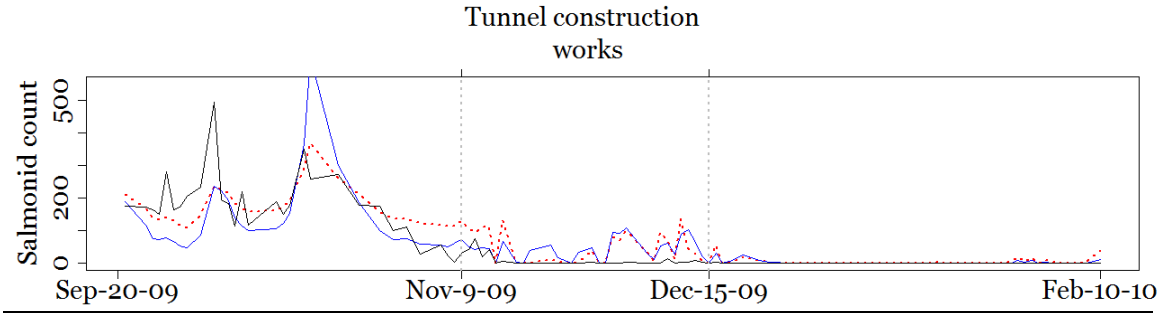
Estimate		Value	Std Error	Z-value	*P*-value
Count	Intercept	5.0261	0.1371	36.652	< 2e-16
	NAO, 5year	-0.4168	0.1737	-2.400	0.01641
	Time since	-0.0835	0.0117	-7.164	7.81e-13
	Tunnel	-1.0492	0.3303	-3.176	0.00149
Zero	Intercept	24.0384	10.1064	2.379	0.0173
	Tide	12.0448	5.4851	2.196	0.0281
	Turbidity	-1.0563	0.3452	-3.060	0.0022
	Cos(1year)	-31.4329	-31.4329	10.842	0.0037
	Sin(1year)	-7.5045	-7.5045	2.818	0.0078
	tunnel	3.0560	1.5502	1.971	0.0487
	Log(theta)	0.6740	0.2407	2.800	0.0051

Correlations between covariates of significant effects are presented in Table 3.8. Tunnel construction and turbidity were correlated. Temporal covariates appeared correlated, whether these were cyclical (sine and cosine waves, tide) or a linear measure of time (time since tunnel construction works).



° Residuals of hurdle model of salmonid counts ····· Normal distribution

Figure 3.8 : Diagnostic plots for the optimal hurdle model for salmonid counts for the period 20th September 2009 to 10th February 2010.
The residuals were left skewed and strongly non-normally distributed with an important departure of the right tail (a&b), and there was (c) heteroscedasticity and (d) autocorrelation.



— Observed salmonid counts
— Predicted salmonid counts in observed conditions
····· Predicted salmonid counts supposing no tunnel construction

Figure 3.9: Observed and fitted values according to the most parsimonious Hurdle model for the period 20th September 2009 to 10th February 2010.
The salmonid counts are predicted according to observed conditions and also supposing that no tunnel construction works took place.

Table 3.8: Correlation value and significance between the hurdle model components for the period 20th September 2009 to 10th February 2010.
The correlation coefficients are displayed in the lower panel and corresponding *P*-values are displayed in the upper panel. Values on grey background are non-significant.

	NAO	Time since	Tunnel	Tide	Turbidity	cos (1 year)	sin (1 year)
NAO	.	<0.001	<0.05	0.762	<0.001	<0.001	<0.005
Time since	0.473	.	<0.001	<0.005	<0.001	<0.001	0.047
Tunnel	-0.096	-0.452	.	0.585	<0.001	<0.001	<0.001
Tide	-0.261	-0.546	0.272	.	<0.001	<0.001	<0.05
Turbidity	-0.066	-0.407	0.676	0.209	.	<0.001	<0.001
cos(1year)	0.553	0.640	0.351	-0.357	-0.001	.	<0.001
sin(1 year)	0.505	0.996	-0.369	-0.539	-0.348	0.705	.

3.4. Discussion

Fluctuations in River Tyne salmonid numbers during the years 2004-2012 were defined by an annual periodicity, NAO values with a lag of four years and river temperature. This indicates that migration seasons with high salmonid counts were characterized by a synchronized increase in river temperature and lower NAO values four years prior to the observations.

It was hypothesised that other covariates would be of importance but were not selected by the models, including the Atlantic Multidecadal Oscillation. AMO values induce fluctuations of sea surface temperatures; at sea, *S. salar* appears to be found in lower quantities when the AMO is in a warm phase rather than a cold phase (Condron *et al.*, 2005). Likewise, salmon stock survival is correlated with spring water temperatures in the Atlantic Ocean, extending into the North Sea (Friedland *et al.*, 1998).

Cycles in sea temperature and fish communities are closely related. Global ocean warming has increased plant biomass and turnover in plankton communities (Beaugrand and Reid, 2003) and fish (MacNeil *et al.*, 2010) towards a community structure more similar to warmer waters. The latter are not suitable for salmonids to feed on and the community turnover becomes associated to shift in prey base (Beaugrand *et al.*, 2008) that leads to food limitations for *S. salar* and *S. trutta*. For several fish species in the North Sea (*e.g.* cod *Gadus morhua* and anglerfish *Lophius piscatorius)* this turnover has also been associated with changes in the spatial distribution and deepening of fish assemblages (MacNeil *et al.*, 2010). These processes suggest a possible global and persistent decrease in energetic gain from the marine stage of salmonids.

NAO effect on salmonid counts was significant only when lagged. The four year lag implies that counts of breeding salmonids were, on average, related to temperature at earlier life stages four years previous to being recorded; the stage in question depends upon the combination of years spent in river and at sea. The duration of the maturation in freshwater varies considerably between individuals of the same species (1 to 4 years for *S. salar*, Klemetsen *et al.*, 2003 and 1 to 3 years in *S. trutta*, Bekkevold *et al.*, 2004); time spent at sea also fluctuates from one to more years (Friedland *et al.*, 1993). When averaging these

conditions over the several indiscriminate age-classes of each species, this indicates that the lag of four years in NAO values likely represented the marine conditions when the recorded salmonids were either smolts in the river (in which case they could not have experienced sea conditions so these are not concerned), or post-smolt entering the sea, or the parent population of the recoded salmonids as they returned to spawn. NAO values were linked to thermal estuarine conditions, which influence numerous fish species either at a juvenile stage or by influencing conditions for the parent stock (Attrill and Power, 2002).

The importance of parental stock on abundance of *S. salar* has been shown by stochastic life history models (*e.g.* Dumas and Prouzet, 2003; Jonsson *et al.*, 1998) that make use of demographic data and require representative sampling to quantify all life stages within a population. Demographic features were not incorporated here as demographic data were not available for this study. If the freshwater stage is an important source of mortality in *S. salar* as they return to spawn in the River Tyne, it is possible that the lagged NAO value, relating to the freshwater stage of *S. salar*, was selected in the model partly as a representation of the spawning population.

Long-term dependency in environmental conditions has been demonstrated at the fry stage for both salmonid species: abundant food leads to a well represented age class (match-mismatch, Jonsson and Jonsson, 2004). Scarce food at this life stage is associated with low productivity regardless of whether conditions improve. Smoltification involves high energetic requirements for growth and maturation and is associated with increased food intake (Jorgensen and Jobling, 1992). It can be hypothesised that lagged values of NAO are linked to feeding conditions in the estuary at the smolt stage, and that these conditions significantly influenced how many salmonids of their age class concerned return to the River Tyne as breeders.

In addition, anadromous fish are believed to be highly sensitive to mortality when entering the oceanic habitat. Causes of high mortality are poorly known, however this totally new environment implies new predation, diseases and food; all of which challenge physical and behavioural characteristics (Friedland, 1998; McCormick *et al.*, 1998). Anadromous salmonids have a physiological and ecological *smolt window*, periods of time during which smolts survival is

heavily dependent upon their readiness for the osmotic change inherent in entering the marine environment and habitat suitability (McCormick *et al.*, 1998). This set of requirements makes the first period of the marine phase most challenging for up to 2 months (Friedland *et al.*, 1993). European Atlantic salmon populations are habitat dependent beginning at the post-smolt stage: associated growth and fitness are consequently compromised if not provided with adequate habitat requirements (Friedland *et al.*, 1993). Other authors argue that predation is the main factor responsible for marine mortality (Hansen and Quinn, 1998).

The relationship between the sea surface temperature, thermohaline circulation and fluctuations in AMO (Mingfang *et al.*, 2009) and NAO (Wang *et al.*, 2010) allowed incorporation of indices representative of the oceanic climate in the models of salmonid migration. The un-lagged values of NAO and AMO were not relevant to the salmonid migration model, implying that marine conditions as the salmonids were returning towards the estuary were not significant predictors of the temporal abundance patterns in salmonids observed in the river.

It may be hypothesised that because *S. salar* and *S. trutta* follow a different migration route when at sea, actual marine influences may not be accounted for in a single model as movements are species-specific. Post-smolts of both species were however exposed to similar conditions as they occupy the Baltic Basin simultaneously for the first months of the post-smolt stage. Post smolts of *S. salar* from the River Tyne, like the Scottish natives, likely use near-shore areas at the beginning of the marine migration (Malcom *et al.*, 2010) and *S. trutta* tend to remain along the coast (Jonsson and Jonsson, 2011). A single model representing the two species jointly was hence relevant to describe this stage of the migration and, equally, condition for the parent counted at Riding Mill.

The relevance of year as a random effect indicated a difference in salmonid counts between the years, possibly reflecting differences in marine survival rates and/or annual variation in maturation rate (Friedland, 1998). The rate of change in counts also varied with temperature, indicating that the relationship to temperature is year-specific. Autocorrelation within the residuals persisted in the most parsimonious model. Whilst salmonid behaviour is likely to have been autocorrelated, many of the environmental predictors used in the models would

also show some level of autocorrelation, including the random effects. But, allowing correlation in temperature and time would equate to correlated random effects for the intercept and slope on the data. This would cause the variance of the intercept to approach zero while the covariance of intercept and slope stay non-zero, which would be a mathematical impossibility (Bates, 2007). Several sources advise against refining correlation structures despite lowering AIC scores (Zuur *et al.*, 2009). Attempts to improve the correlation structure did not succeed and were considered not suited for the model.

The fit of models was in most cases improved when the count data were transformed. Box-Cox transformations are often applied by using a close value of lambda particular to the field of study. Here the power transformation used the exact lambda value recommended by the Box-Cox transformation (λ=0.182) and models became linear. Stepwise selection of covariates assuming a linear relationship when it is not the case, risks retaining a covariate only because it carries some level of non-linearity (Royston and Altman, 1994). By transforming the response variable, the chances of such erroneous selection are minimized as the relationship is closer to linear. Here, the exact lambda value optimized modelling the linrst relationship between response and explanatory variables. The ecological meaning of lambda may consequently not be obvious initially but it is in order to define it as well as possible that this level precision was maintained.

Response transformations help issues related to both heteroscedasticity and skewness: lambda informs on the shape of the relationship between response and explanatory variables. Selection of the value of lambda is recommended to be accomplished in light of the data (Box and Cox, 1964). However here, this value can be considered one of the parameters investigated by the model as no ecologically meaningful value was apparent due to the complexity of the processes involved and the spread in time. In some instances the power relationship may simply be defined by the function in question, or may be representative of one or several mechanisms that are not described by the model. Given that salmonid life history is multiannual, covers a range of marine and freshwater habitats and that little data are available about migration for this study, there were likely missing covariates.

The shape of the response transformation may also be a product of the conversion of local data to a more general perspective. The aggregation of the dataset allowed matching the scale of the study to the scale of the measurements. Counts were standardized to represent 2 weeks but the initial data were collected individually and locally. In that regard, this study imposed an indirect scaling on the salmonid ecology. Data aggregation may raise mathematical concerns for such scaling exercises, including the danger of generating nonlinearity by averaging (Marquet *et al.*, 2005). Ecological processes may change when moving from a small to larger scale made of aggregates of the small scale observations; this phenomenon called transmutation has been related to non-linear relationships across scales due to spatial (landscape) heterogeneity (King *et al.*, 1991) and salmonid numbers were summed over the time intervals but this issue may also be due to averaging of explanatory data such as flow, which take on different meanings depending on the way they were aggregated (mean or sum).

A response transformation with a more approximate and traditional (λ=0.5) lambda value than the one advocated by Box-Cox methodology (λ=0.182) resulted in different sets of predictors being selected for variation in salmonid counts. Annual seasonal cycle was retained and higher counts were again related to an increase in temperature. Values of NAO were this time not retained and the tunnel construction appeared to lower fish counts considerably during the period in question. The estimated differences associated with (similar) random effects were more important in this model than when using the Box-Cox transformation. This possibly indicated that the general trend was not optimally accounted for since the model attributed more variability to random effects without the residuals being appropriately distributed. This confirmed the assumption that a more precise Box-Cox transformation accounts for an amount of variability that allows selection that may not be explained with the available model components (in regard to linearity and unless there is an ecologically meaningful power value, Royston and Altman, 1994); and allows the selection of covariate with a reduced bias.

The Fourier series, containing a trigonometric expression of annual periodicity provided were a basic representation the annual seasonal cycle, a controlling factor of the temporal variability of many marine ecosystems (Bertram *et al.,*

2001). The River Tyne salmonid population was part of a complex marine community involving other organisms, themselves responding to external parameters according to their own cycles. Plankton dispersal, for instance, responds to the annual seasonal cycle and in turn influences salmonid distribution (Beaugrand and Reid, 2003). Annual seasonal cycle was admittedly a factor in itself, but the significance of the annual periodicity as a covariate also indicated that an additional factor, or factors, correlated with annual periodicity may affect the salmonid life cycle.

This becomes obvious in the transient periodicity observed in the wavelet transform of the time-series. The presence of periodicities other than the main one indicates that the waveform of the salmonid counts may not be perfectly sinusoidal (Diggle, 1990). The waveforms synthesised by the most parsimonious model are presented on Figure 3.10, along with a simple annual periodicity, the secondary periodicity (24.7 weeks) and the sum of both.

The waveform made of both periods on Figure 3.10d illustrates the imperfect sinusoidal shape relevant to the first half of the observation period (2004-2008). Such a shape was not represented by the stationary models. The double hump may represent a change in behaviour of one of both species and/or a lack of synchronicity in the return numbers during the years in question. Explanations as to why both species were not in perfect phase or how the reaction to external parameters may have been time-varying, is addressed separately in the following chapter.

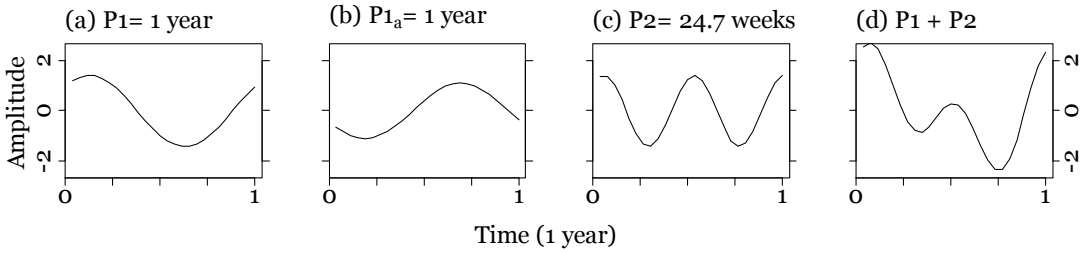


Figure 3.10: Illustration of the waveforms relevant to the description of periodicity contained in the time-series of salmonid counts.
The waveforms contain (a) an annual period P1, (b) an annual period with coefficients as estimated by the most parsimonious LME model of salmonid counts $P1_a$, (c) a 24.7 weeks period P2 and (d) the sum of both periods, over a year time.

The model investigating the tunnel construction works aimed to take into account the many null counts recorded during this period. Zero-inflated and zero-altered models were both used and only the latter type appeared suitable in light of Chapter 2. In this instance, the main difference between the two designs

was the ability of the zero-inflated models to separate between true and false zeros (Zuur *et al.*, 2009). True zeros implied the absence of salmonid as a response to the river parameters rendering the habitat unsuitable; excess zeros were null records in conditions where the habitat was good. Winter conditions were bound to induce null counts due to unsuitable conditions (*i.e.* true zero). Null counts due to the construction works, whether true (the tunnel creates an unsuitable habitat) or false (salmonid are queuing downstream from the counter, present but not counter), were consequently confounded and not differentiable by the model. This is implied by the selection of a zero-altered model which confounds all types of zeros. This also indicates that the tunnel construction works took place at a time that was optimal for the non-disruption of the salmonid migration pattern.

Delayed downstream migration of *S. salar* smolts is associated with lower survival because of the mismatch with the fish physiology and environmental conditions *(*Marschall *et al.*, 2011). Descending smolts must reach the estuary before temperatures become too high; if a delay prevents the fish to do so then mortality increases. The construction works in the Tyne took place in December. Natural conditions were either unsuitable (*i.e.* naturally associated with rare counts) or becoming more suitable for migration. A delay was unlikely to be associated with a mismatch.

Strong correlations between covariates existed for this period, some coincidental with construction. For instance, turbidity values jumped during the tunnel construction period (Appendix A.ii, Figure A.7.14).The measure was taken upstream from the tunnel works, so were correlated but unrelated. The number of days since construction appeared to lower salmonid counts but was again confounded with the natural migration cycle. The period considered (September 2009 to February 2010) was considered not representative enough to extend the inferences from the model estimates to a broader scale and hence not suited to test potential long-term effects. The long-term LME model suggested no subsequent impact to be recorded on the available period. However, the link with NAO values four years before implies that a long-term effect, however unlikely, may be investigated again on a time-series including four to five years following the construction works (*i.e.* with data up to 2014-2015).

3.5. Conclusions

Salmonid counts in the River Tyne during the years 2004-2012 were modelled using GLMs and LME models which described the characteristics shared by the two species constituting the salmonid population: *S. salar* and *S. trutta*.

A positive relationship was established between salmonid counts and the river temperature. Values of NAO with a four year lag also appeared influential, underlining the importance of sea conditions for the parent population and/or at the juvenile stage as they share territory in the North Sea. Marine factors consequently controlled annual survivorship in a similar manner for both species. Annual seasonal cycle and/or a parameter with an associated cycle, was also a predictor of the migration pattern. No covariate reflected the importance of marine conditions during migration towards the natal river.

The tunnel construction appeared to have taken place at an optimal time regarding the salmonid migration, engendering no significant disruption to the migration pattern.

The failure of the GLMs, the imperfect diagnostics from the LME models and the fractional power transformation required, all pointed towards a potentially non-stationary relationship between the counts and the available environmental parameters and the dependence of the response. A relationship between salmonid counts and time was made explicit in terms of both autocorrelation and periodicity; an imperfect sinusoid with a transient component was not accounted for by the models. These elements suggested a high level of complexity contained in the patterns of salmonid counts, which likely implies the existence of several confounded ecological processes.

This complexity is likely partly due to species-specific traits contained in a single time-series. In this case, complexity may be reduced and processes described by describing *S. salar* and *S. trutta* as distinct populations. In the next chapter, the patterns specific to the count of each species will be investigated individually.

Chapter 4.

"Everything alters me, but nothing changes me".

S. Dali
Quoted in *The Secret Life of Salvador Dali*
S. Dali and H.M. Chevalier, 1942.

Chapter 4. An analysis of the migration patterns of *S. salar* and *S. trutta* in the River Tyne.

4.1. Introduction

By definition, migratory species are mobile and adjust their behaviour so as to pursue suitable environments throughout the progression of their lifecycle (Robinson *et al.*, 2009). Geographic locations for breeding, migrating and non-breeding stages of their life-cycle vary and meet different criteria of suitability (Robinson *et al.*, 2009) so that reproduction and growth are synchronized with peaks in specific resources that satisfy the demands of critical life stages (Gill *et al.*, 2001).

The management and conservation of migratory species is particularly challenging. Poor knowledge of migration patterns and associated demands are a common problem in the conservation of migrants, as well as the variety of pressures linked to movements, fluctuations in home range, and the need for both local and broad-scale protection that can be difficult over national boundaries (Epstein *et al.*, 2009).

An estimated 3% of fish species are migratory, and little to no knowledge exists about their migration conditions (Robinson *et al.*, 2009). *S. salar* and *S. trutta* are no exception and their differences between the migration add to the difficulty in monitoring them. *S. salar* and *S. trutta* tend to overlap in time and space when migrating both downstream and upstream along natal rivers (Jonsson and Jonsson, 2011). However during the marine phase of their lifecycle, these salmonids follow distinct routes. Anadromous *S. trutta* tend to feed and remain within coastal waters and estuaries, and rarely venture further than 100km from the shore (Jonsson and Jonsson, 2011) but still cover appreciable distances (Pratten and Shearer, 1983; commonly 100 to 500km, Swain *et al.*, 1960) that are specific to each population and increase with the length of the natal river (Jonsson and Jonsson, 2011). In several studies, *S. trutta* was observed in the reaches of non-natal rivers, emphasizing their

extensive coastal migration (*e.g.* recaptures of natives from Devon in River Tweed, England, Swain *et al.*, 1960; natives from Moray Firth moving between the Beauly and Ness Rivers, Scotland, Anonymous, 1930). The migration route of *S. salar* extends well beyond the North Sea basin. It is likely that *S. salar* travel to the North Atlantic sub-polar gyre directly after their downstream migration, moving counter-clockwise within these rich feeding grounds until migrating back towards Europe (Dadswell *et al.*, 2010); North American and European salmon meet on these feeding grounds but the latter move south (Windsor *et al.*, 2012).

Groups also exist within each species; the influence and function of these groups on their migratory route is poorly described. The timing of migration sometimes differs slightly with regards to age class and/or fish size, in both species (JNCC, 2007; Birkeland, 1996; Jonsson *et al.*, 1990). The phase difference may or may not reflect coverage of a different migration route, but it does imply that conditions are not lived concurrently for all groups within a population. Dispersal behaviour in *S. trutta* is potentially sex-biased (Bekkevold *et al.*, 2004) and the distance covered is highly variable between individuals (Malcom *et al.*, 2010). Kelt and grilse of *S. salar* reach for feeding grounds that are likely geographically distinct and not fully described (Gauthier-Ouellet *et al.*, 2009).

Environmental requirements are however shared by whole populations. Notably, the suitability of thermal habitat greatly influences in the abundance of both species as they enter the marine environment (Friedland, 1998). Thermal conditions, in particular saline Atlantic currents, regulate the dispersal of *S. salar* in May as they return towards their natal river (Holm *et al.*, 2000). These conditions are assumed to be partly reflected by the values of AMO and NAO (Visbeck *et al.*, 2001).

In both species, survival at sea is generally estimated by the difference in numbers of smolts emigrating towards the sea and breeders returning to the natal rivers; the factors responsible for mortality at sea are however poorly understood (Windsor *et al.*, 2012). The survival of *S. salar* is in decline and recent regulations of fisheries suggest that other oceanic factors are responsible for this trend (Hansen *et al.*, 2012), such as the turnover in communities due to ocean warming (MacNeil *et al.*, 2010; Beaugrand *et al.*, 2008). Complete sets of

marine data would allow identifying the factors driving the trends in salmonid survival during the marine stage and quantify their influence.

Collaboration between fishermen and scientists is often suggested as a solution for improved marine data (Pauly *et al.*, 2013). Such collaboration was tested for the monitoring of the Pacific salmon (species of the Subfamily *Oncorhynchus*) with an aim to collect real-time data at sea, and underlined operational issues. The harsh and remote sea environment engendered power demand and connectivity issues resulting in loss of data and there was an overall difficulty to familiarise with the scientific software (Lavrakas *et al.,* 2012).

In response to this lack of understanding and need for data, the International Atlantic salmon research board (IASRB) was established in 2002 as a part of the North Atlantic salmon conservation organization (NASCO). The IASRB investigates the causes for marine mortality in *S. salar* and prioritises the funding of related projects. In 2012, long-term monitoring of survival, distribution, and migration of *S. salar* in relation to predation and feeding opportunities at sea were the highest research priorities (£4.3M spent mostly by the European Union and Norway, IASRB, 2012).

Some of the major findings of IASRB (Hansen *et al.*, 2012) were that productivity and abundance of *S. salar* are influenced by broad-scale factors and that the decline in their abundance decline is stronger in multi-sea-winter individuals; long-term changes are linked to northern hemisphere sea temperatures and the NAO, and there is a Northwards movement of fish and *S. salar* prey in the North Atlantic. Ocean-scale dynamics constitute the factors impacting on oceanic organisms (*e.g.* ocean warming is global but has spatially varying effects on bleaching of coral reefs, Graham *et al.*, 2008), however, local environmental dynamics may be affected differently by the climate and the resulting impacts on age classes of *S. salar* may hence vary geographically (Hansen *et al.*, 2012). A complete understanding of salmonid migration consequently requires a description of process in both oceanic conditions and local riparian environments.

Several methods may be used to monitor salmonids in rivers, including electrofishing, trapping, in-situ counters, and angling records. In the UK, electrofishing is commonly used by the Environment Agency (*e.g.* Rivers

Thames and Tyne, EA, 2013). Electric fields are generated in the water with two electrodes that attract the fish and render it numb momentarily, whereby the fish are effortlessly caught with a net (EA, 2013). The technique is considered uniquely harmful to salmonids as the convulsions induced by the electrical shock may cause spinal injuries, internal haemorrhages, and asphyxiation (Snyder, 2003).

The abundance of *S. salar* and *S. trutta* in the Rivers Dee, Lune, Tamar, and Tyne (EA, 2013) is also monitored by fish traps. Permanent traps temporarily capture breeding salmonids travelling upstream and data are collected about their physical condition (species, weight, size, injuries, diseases, parasites); the fish are also tagged, from which data may contribute to studies of movement and distribution. Often, monitoring devices do not cover the whole width of the rivers (*e.g.* River Tees and River Tamar) and abundance is estimated from a combination of sources (*e.g.* fish counter data and trap catches, River Tamar, Cornwall).

Both salmonid species are widely exploited by recreational angling (Curtis, 2002) and anglers and owners of fisheries contribute to evaluating the health of fisheries by declaring catches in a logbook (EA, 2013). Anglers may also choose to provide scale samples that are can be analyzed to determine the age of spawners and timing of the migration. In Scotland, a questionnaire is sent to owners of fisheries annually (section 64 of the Salmon and Freshwater Fisheries, Scotland, Act 2003).

At Riding Mill Station, a video camera installed with a fish counter (Figure 4.1a) has recorded upstream movements since 2004 (Figure 4.1b), providing abundance data for the populations of *S. salar* and *S. trutta* in the River Tyne. The dataset was assembled using a non-invasive device, independent of the fishing industry, and it includes associated measurements of river parameters (flow, temperature, turbidity, colour, pH and conductivity).

As migratory range differs between *S. salar* and *S. trutta*, the way environmental parameters influence fish development during the marine phase is species-specific. Consequently the study of migration for the two species benefits from being considered individually. The main objective of this study is to investigate the extent to which it is possible to use the video records from

Riding Mill station to develop parsimonious models that adequately describe large scale migration patterns of *S. salar* and *S. trutta*, according to variation in known environmental parameters.

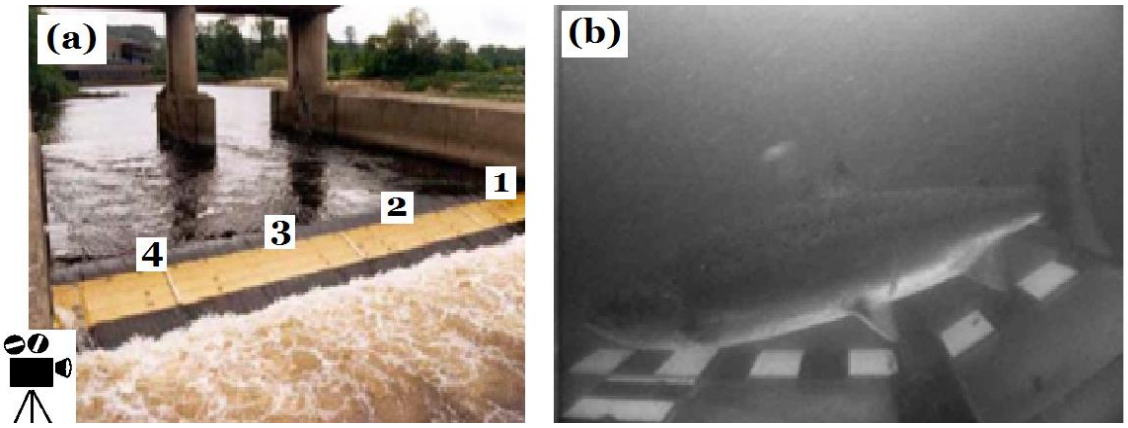


Figure 4.1: (a) The four channels of the resistivity counter at Riding Mill and (b) an example of video recording produced by the camera at Riding Mill.
Channels are numbered (1 to 4) and the camera is located at channel 4.

To investigate the regularities specific to each species over the whole observation period, wavelet analysis will be undertaken independently on time-series observations (Grinsted *et al.*, 2004). Then the synchronicity between fluctuations in the abundance of both species will be tested using the wavelet coherence (Vadrevu and Choi, 2011). Relevant periodicities will be incorporated in GLM, LME, or GLS models if evidence suggests dependence within the data (Fox, 2002). The relative proportion of each species count will be studied over time.

4.2. Material and Methods

4.2.1. Data

Response variables, transformations

Three time series were obtained from video records on channel 4 (Figure 4.1) of the fish counter at Riding Mill station from 2004 to 2011, resulting in three univariate time series counts of *S. salar*, counts of *S. trutta,* and a total sum of counts.

The response data were non-normal so the Box-Cox process was used to estimate the best normalizing power transformation for the identified count data (Osborne, 2010). The two time-series of identified counts were also Anscombe- and log-transformed. For the latter transformation, the effect of predictors was back-transformed by:

$$Back\text{-}transformed\ estimate = 100.e^{model\ estimate\ -1}$$

The back-transformed estimate was expressed as the percentage change in the response variable for each unit of increase in the predictor. The count data was incremented by 0.5 for the Box-Cox and log-transformation (Yamamura, 1999).

Covariates and format

As this study was concerned with general migration patterns, the covariates used in the models were related to their higher level of observation. These were: temperature, flow, year, NAO values, AMO values at various lags, and annual harmonics *(see Introduction chapter for theoretical justification*) with bins of 2weeks. Data for the period 2004-2010 were used to develop the models and data for the year 2011 were used to test the forecasting ability of the models.

4.2.2. Analysis

Wavelet analysis and periodicity

As for Chapter 3, the harmonic analysis was based on a continuous wavelet transform, which detected stationary periodic events as well as localized intermittent periodicities (Grinsted *et al.*, 2004). Initially this was done with each time series separately, then a cross wavelet transform was constructed with the two time-series of video counts of *S. salar* and *S. trutta*. The analysis generated a spectrum showing wavelet coherence between the two time-series as well as their common power and phase relationship in the time-frequency space (*i.e.* over time).

A wavelet coherence function quantified the linear relationship between frequencies contained in counts of *S. salar* and *S. trutta*. Because the wavelet process considered all scales, the coherence function assessed the strength of correlation between both species counts for varied frequencies and times (*i.e.* as opposed to calculating individual product moment correlations that may be biased or hidden by potential non-significant relationships at certain scales, Vadrevu and Choi, 2011). Wavelet coherence is the local correlation between two continuous wavelet transforms, but in time frequency space (Grinsted *et al.*, 2004).

The spectrum of wavelet coherence illustrated the phase between oscillations contained in the two time series via the direction of arrows. The phase was a time difference expressed as angles in order to normalize time regardless of time period. The time-series were in phase when pointing right (90°), in anti-phase when pointing left (270°) and there was evidence of a phase lead (*i.e.* time delay) when pointing up or down (0° or 180°) (Gouhier and Grinsted, 2012).

Harmonic regression

As for Chapter 3, a harmonic analysis used the Fourier series to decompose the signal contained in the three time-series into harmonics made of a sine wave and a cosine wave that were included and tested in the models.

Models

The relationship between counts of each species and environmental covariates was investigated using GLMs. In case of evidence of residuals autocorrelation or violation of homogeneity in the variance structure, mixed-effect models were investigated. Full linear mixed effect models were run on the same methodology as Chapter 3, containing all covariates as fixed effects and using REML in order for the models to be compared to a full GLS model. If evidence suggested that the use of random effect was redundant, GLS was used.

Initially, the variance structure of the model was selected, based on visual examination of the GLM outputs. Relevant structures were tested on a full model so as to maximise the number of explanatory variables in the fixed part of the model before complicating the random structure (*i.e.* random effect parameter in LME models, variance in GLS models). Then, the correlation structure ARMA(p,q) was selected amongst all combinations of p=0,1,2,3 and q=0,1,2,3. The start value was set at 0.2 for both parameters (as in Zuur *et al.*, 2009).

As for Chapter 3, all candidate models were assessed in term of overdispersion (ratio of residual deviance over degree of freedom), goodness of fit (Pearson correlation coefficient), distribution of residuals (homogeneity, normality and independence), and AIC score (calculated via the log-likelihood value for each model).

Loess regressions were fitted for the two response time-series in order to smooth counts over time. Loess function was computed for all values of counts (Cleveland *et al.*, 1990). The loess regression aimed to identify global trends of the time-series over the whole observation period. The polynomial degree was 1 so that the fitting was locally linear, which was generally suitable for data with gentle curvature (Cleveland *et al.*, 1990). The value of the loess span determined the number of points to be included in the local regression at every point; each included point had a neighbourhood weight that decreased with its distance to the point being fitted (Cleveland *et al.*, 1990). For instance, a span value of 0.2 implied that 20% of the data points were used for the local fitting: 10% located before and 10% after the fitted point. For increased values of span, the locally fitted linear regression became smoother and the weights tended to one;

consequently, the value of span was selected so as to keep it as low as possible, identical for the two time-series compared, and with reduced confidence intervals (as these implied potential null values which had to be minimized in order for the loess regression to be usable as components of ratios).

4.2.3. Software

The wavelet analysis, GLM, GLS, LME, and loess regression were developed using, respectively, biwavelet (Gouhier and Grinsted, 2012), MASS (Venables and Ripley, 2002), nlme (Pinheiro and Bates, 2000), and gppois (Hogg *et al.*, 2012) packages, in the R.15.1 environment (R Core Team, 2012).

4.3. Results

4.3.1. Harmonic analysis

Individual periodicity

The wavelet power spectrums and matching average wavelet power spectrums were generated for the time series for each species (Figure 4.2).

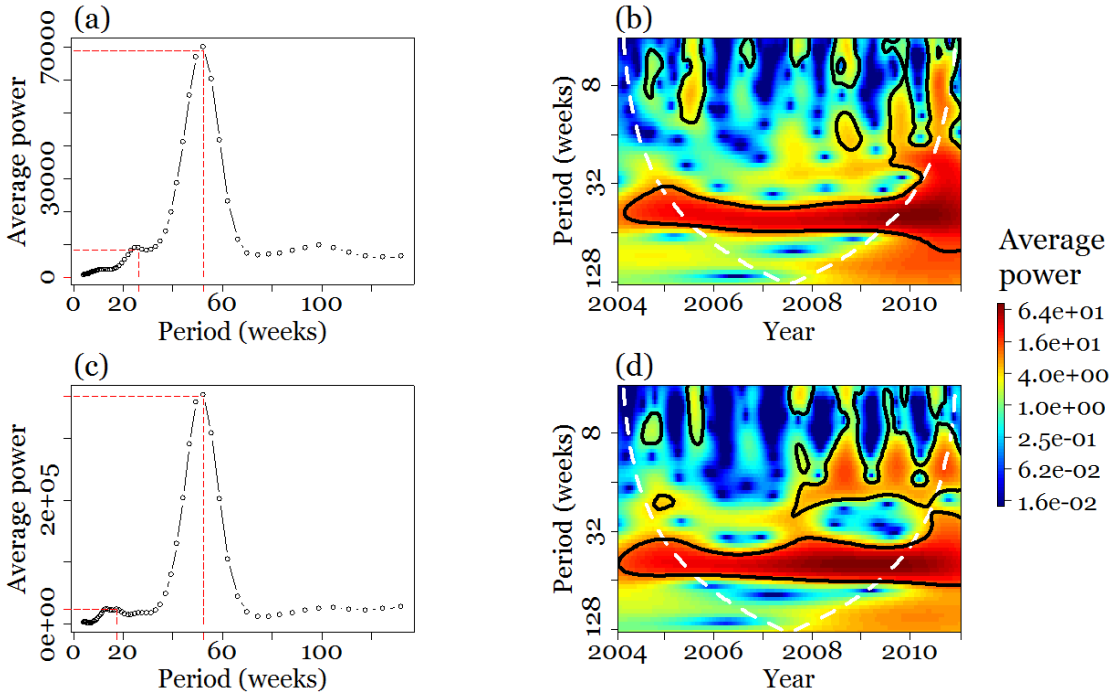


Figure 4.2: The wavelet analysis of non-transformed counts of *S. salar* and *S. trutta* over the observation period 2004-2011.
With (a&c) wavelet power spectrum and (b&d) resulting average power spectrum over the whole observation period. The interpretation was restricted to the cone of influence; highest power appeared in yellow through red on the spectrum and the lowest power was in blue. A thick black line defined the contours for significance at the 95% level for all power of all points; the *P*-value corresponded to the average power over a given period, which may vary in value and significance over years. The primary period was of 52.48 weeks for both species and the secondary period was of 26.24 weeks for *S. salar* and 17.51 weeks for *S. trutta*.

Values of interest were extracted and are presented in Table 4.1. Unsurprisingly, for both species the highest power average was for a period of 52.48 weeks (*P*-value<0.001); this periodicity was constant and strong throughout the years for both species.

Table 4.1: Periodicity values of counts of *S. salar* and *S. trutta*.
Values were extracted from the wavelet transforms for the observation period 2004-2011.

	S. salar				*S. trutta*			
	Duration (weeks)	power	*Z*-value	*P*-value	Duration (weeks)	power	*Z*-value	*P*-value
Primary period	52.48	$6.9.10^4$	3.34	<0.001	52.48	$3.7.10^5$	4.08	<0.001
Secondary period	26.24	$8.6.10^3$	0.79	0.215	17.51	$2.4.10^4$	1.06	0.145

Secondary periodicities were also observed (yellow, outlined regions on the average power spectrums over time, Figure 4.2b&d) and had different durations and significance (Table 4.1). They were transitory periodicities, occurring more particularly during the second half of the whole observation period. This indicated that the cyclic variations observed in both time-series were likely not perfectly sinusoidal during the second half of the observation period (Diggle, 1990). These secondary periods presented low significance level and were discontinuous so they were not expected to be well represented by the non-stationary models. The significance and power of the transient periodicities were higher for *S. trutta* than for *S. salar*.

Synchronicity of the two species counts

The wavelet coherence between counts of *S. salar* and *S. trutta* (Figure 4.3) displayed highly significant sections with high coherence, high common power and high synchronicity (plots showing common power, wavelet coefficients and phase relationship in Appendix A.iii, Figure A.7.15). These sections were of high significant power on the wavelet power spectrums for distinct counts of *S. salar* and *S. trutta*.

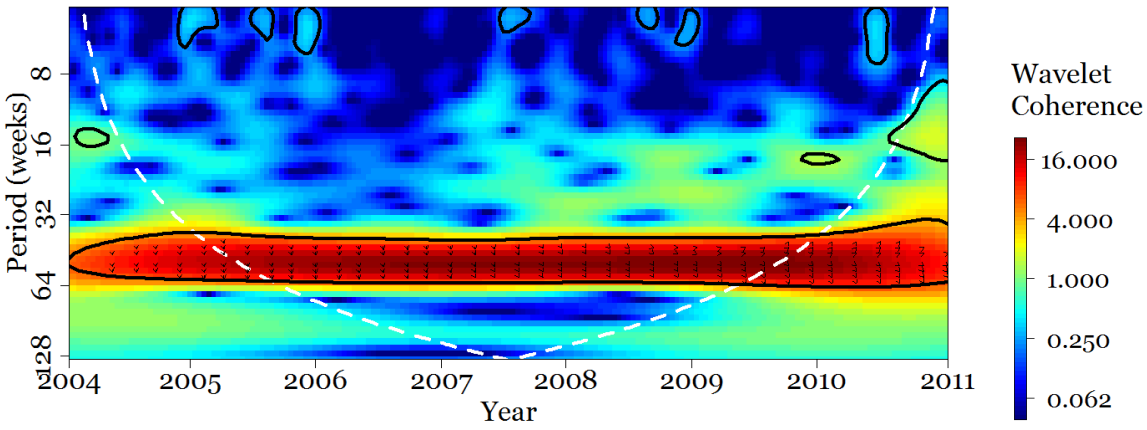


Figure 4.3: Wavelet coherence between counts of *S. salar* and *S. trutta* for the whole observation period.
The regions in time frequency space where the oscillations of the two time series have a coherent phase appear in red. The relative phase relationship is shown by arrows. In this instance, the significant regions coincide with high power.

There was a significant phase relationship between the abundance of *S. salar* and *S. trutta* in the River Tyne at periodicities of 32-64 weeks. The most important and significant maximum common power was for a period of 52.47 weeks, equivalent to a year (Z-value=3.26, P-value<0.001).

The significant regions displayed the evolution of the phase over the whole observation period. From mid-2005 until mid-2007, the arrows pointing down indicated that the oscillations in counts of *S. salar* led the oscillations in counts of *S. salar* (by 90°, *i.e.* one quarter of a period). From mid-2007 until 2008, the two time-series were in phase (arrows pointing to the right). From 2008 to mid-2009 the relative phase relationship became mostly reversed: the oscillations in counts of *S. trutta* led the oscillations in counts of *S. salar* (arrows pointing upwards, again by a quarter of a period). Thus, oscillations in the two time-series were not phase locked (as the direction of the arrows varied): the oscillations in both time-series were not synchronized throughout the

observation period and the lead was first in *S. salar* and started shifting towards a lead for *S. trutta* in 2008.

4.3.2. Generalized linear models

Outputs from optimal GLMs according to AICc scores are presented in Table 4.2 and Table 4.3, for different transformation of the response variable of, respectively, *S. salar* and *S. trutta*. The ratio of residual deviance over the residual degrees of freedom was used to quantify the level of overdispersion contained in the models.

Residuals were non-normally distributed so the response were subjected to Box-Cox transformation. The computation of the maximum likelihood function and posterior distribution of Box-Cox transformed response variable indicated that a power transformation of λ=0.14 was optimal for *S. salar* and a power transformation of λ=-0.02 was optimal for *S. trutta* (after Box and Cox, 1964, Appendix A.iii, Figure A.7.16). These powers corresponded to incremented counts values (+0.5).

In *S. salar*, the GLM of the Box-Cox transformed time-series was underdispersed and other models were overdispersed 7 to 27 fold (ratio in Table 4.2), except when using a negative binomial distribution of the error and for the log-transformed counts (Table 4.2).

Table 4.2: The statistics of goodness-of-fit and overdispersion for the optimal GLMs of several transformations of counts of *S. salar*.

	Residual Deviance	df	Ratio	Pearson Cor.	Normality test W , P	AICc	ΔAIC with full model (+)	Model df
Count, Poisson	4774.21	172	27.76	0.63	0.99, 0.43	1582.7	0.9	14
Count, neg. bin.	215.46	177	1.22	0.66	0.99, 0.67	1571.4	12.3	7
Log(count+0.5)	221.85	178	1.25	0.82	0.98,<0.05	567.0	14.1	6
Anscombe(count)	1343.16	178	7.55	0.79	0.97,<0.01	896.6	11.9	6
$\sqrt{count}$	1397.08	178	7.85	0.79	0.98,<0.01	903.8	11.4	6
$Count^{0.14}$	8.53	178	0.05	0.77	0.98, <0.05	-29.2	14.8	6

For the model of log-transformed counts, the low *P*-value for the normality test of the residuals suggested against a normal distribution. The GLM with a

negative binomial error structure presented residuals which were homogenous and normally distributed (Figure 4.4a&b) but auto-correlated (Figure 4.4c).

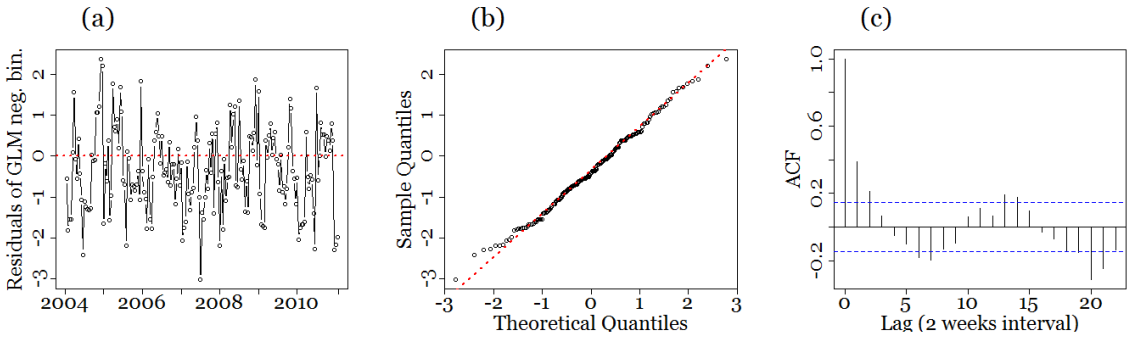


Figure 4.4: Diagnostic plots of the residuals of the GLM of counts of *S. salar* with a negative binomial distribution of the error.
The (a) temporal pattern, (b) Q-Q plot and (c) autocorrelation function of the residuals are displayed. On (c), the dotted blue represent the 2√n white noise limit with n=183. The residuals are homogenous and normally distributed but present autocorrelation.

In *S. trutta,* the GLM of the Box-Cox transformed time-series was again underdispersed and other models were overdispersed 15 to 61 folds, except for the log-transformed counts (Table 4.3).

Table 4.3: The statistics of goodness-of-fit and overdispersion for the optimal GLMs of several transformations of counts of *S. trutta*.

	Residual Deviance	df	Ratio	Pearson Cor.	Normality test W , P	AICc	ΔAIC with full model	Model df
Count, Poisson	10369.85	171	60.64	0.74	0.90, <0.001	11113.5	1.7	12
Count, neg. bin.	10578.36	171	61.87	0.73	0.89, <0.001	1589.8	9734.6	15
Log(count+0.5)	223.11	178	1.25	0.91	0.99, 0.20	568.1	2.9	6
Anscombe(count)	2777.53	179	15.52	0.84	0.99, 0.21	1027.4	11.7	5
$\sqrt{count}$	2825.82	179	15.79	0.84	0.99, 0.22	1030.6	11.5	5
$Count^{0.14}$	0.08	176	0.00	-0.90	0.99,0.22	-881.9	1.9	8

The residuals of the GLM of the log-transformed counts were homogenous and normally distributed (Figure 4.5a&b) but autocorrelated (Figure 4.5c).

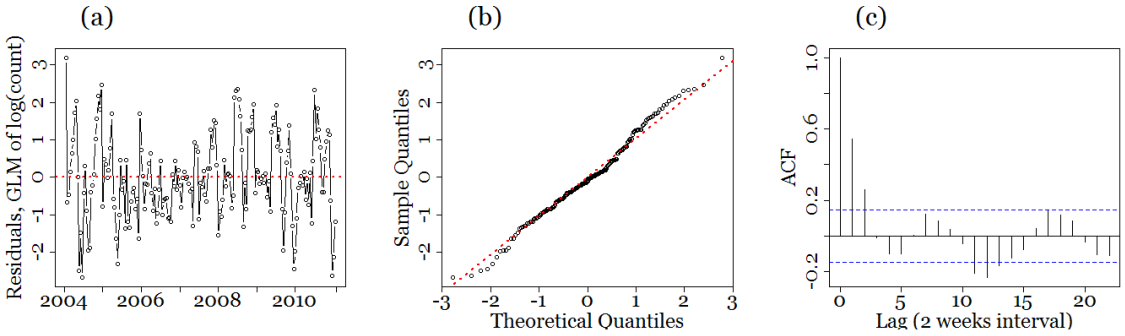


Figure 4.5: Diagnostic plot of the residuals of the GLM of log-transformed counts of *S. trutta* with a Gaussian distribution of the error.
The (a) temporal pattern, (b) Q-Q plot and (c) autocorrelation function of the residuals are displayed. On (c), the dotted blue lines represent the white noise limit. The residuals are homogenous and normally distributed but present autocorrelation.

4.3.3. Linear mixed effect models

Full LME models were developed, that contained all available covariates as fixed effects and year as random effect (Appendix A.iii, Table A.15 and Table A.16); they were compared to the full GLS models for *S. salar* and *S. trutta (*Table 4.4).

Table 4.4: The comparison of the full GLS and full LME models for the counts of *S. salar* and *S. trutta*.
The comparison uses ANOVA; and variance between and within the random effect term (year) of the LME models (*i.e.* intercept and residuals).

	Model	AIC	df	Log-likelihood	χ^2	*P*-value	Variance (intercept)	Variance (residuals)
S. salar	GLS	1918.37	15	-944.18	3.28	0.07	NA	NA
	LME	1917.09	16	-942.54			182.64	2653.87
S. trutta	GLS	2141.34	15	-1055.67	1.90	0 .168	NA	NA
	LME	2141.43	16	-1054.72			631.10	10027.79

For both species, the AICc values of both models were comparable and the likelihood-ratio tests were inconclusive, indicating that the GLS and LME models were of comparable fit.

However because the variability within year was over 14 times the variability between years (approximately 14.5 for *S. salar* and 15.9 for *S. trutta*), the parameter year was not productive as a random effect so the simpler GLS models were preferred over the LME models. The high variability between years indicated high heterogeneity and consequently the requirement for an explicit correlation structure in the GLS model. Log-transformed values were used after the outputs of various response transformations used in the GLMs.

4.3.4. General least squares models

Variance structure

The LME models indicated a strong variability between years so a year-specific variance structure was selected to account for heteroscedasticity in the GLS models. Compared to the GLM, this improved the models significantly with a likelihood over 20 fold for *S. salar* (χ^2= 20.15, *P*-value<0.005) and over 26 fold for *S. trutta* (χ^2= 26.06, *P*-value<0.001).

Error structure

For each species, selected error structures were investigated and compared using the full GLS models; the outputs are presented in Table 4.5.

Table 4.5: The AICc values and ranking for the full GLS models of *S. salar* and *S. trutta*, for various error structures.
The models are developed with AR(1) error structure and ARMA(*p,q*) error structure with each combination of *p*=0,1,2,3 and *q*=0,1,2,3.

S. salar				*S. trutta*			
Candidate error structure	AICc	df	ΔAICc	Candidate error structure	AICc	df	ΔAICc
ARMA(0,2)	564.9	25	0	ARMA(3,2)	533.8	28	0
AR(1)	565.1	24	0.3	ARMA(3,0)	550.7	26	17.0
ARMA(2,0)	566.1	25	1.3	ARMA(0,2)	552.1	25	18.3
ARMA(1,1)	566.6	25	1.7	ARMA(2,2)	552.4	27	18.7
ARMA(0,3)	567.5	26	2.7	AR(1)	553.0	24	19.2
ARMA(1,2)	567.5	26	2.7	ARMA(1,2)	553.4	26	19.6
ARMA(3,2)	567.6	28	2.8	ARMA(3,1)	553.5	27	19.7
ARMA(3,0)	568.3	26	3.4	ARMA(0,3)	553.6	26	19.8
ARMA(2,1)	568.3	26	3.5	ARMA(2,0)	554.0	26	20.2
ARMA(0,1)	569.2	24	4.3	ARMA(2,1)	554.0	25	20.3
ARMA(3,3)	570.1	29	5.2	ARMA(1,1)	554.8	25	21.0
ARMA(2,2)	570.1	27	5.2	ARMA(2,3)	556.0	28	22.2
ARMA(1,3)	570.3	27	5.4	ARMA(1,3)	556.6	27	22.8
ARMA(3,1)	570.9	27	6.1	ARMA(3,3)	557.8	29	24.0
ARMA(2,3)	573.0	28	8.1	ARMA(0,1)	568.5	24	34.7

For counts of *S. salar*, comparable AIC scores were obtained for the error structures ARMA(0,2) and AR(1) (ΔAIC<0.5, Table 4.5). The simplest structure AR(1) was selected. For counts of *S. trutta*, the lowest AIC score was obtained for the error structure ARMA(3,2) (Table 4.5). The ΔAICc with the second AIC-ranked model was too high to justify a simpler structure; the model contained a

complex combination of three autoregressive parameters and two moving average parameters.

Parameters estimation

The estimated parameters are presented in Table 4.6 for the most parsimonious GLS models for the counts of each species.

Table 4.6: Output from the optimal GLS models of the log-transformed counts of *S. salar* and *S. trutta*.

	S. salar	*S. trutta*
	Covariance function estimates	
Year 2004	1.0000	1.0000
Year 2005	0.9534	0.8196
Year 2006	0.4628	0.3773
Year 2007	1.0353	0.5636
Year 2008	0.7860	0.7376
Year 2009	0.6030	0.5959
Year 2010	0.9443	0.5844
	Error structure estimates	
AR	Φ= 0.41	Φ_1 =2.09, Φ_2=-1.89, Φ_3=0.89
MA	.	θ_1=-1.52, θ_2=0.97
	Diagnostic criteria	
log-likelihood	-230.196	-196.815
Residual SE	1.1716	1.5770
df	183 total; 179 residual	183 total; 179 residual

For *S. salar*, the covariance function indicated a difference in particular for years 2006 and 2009. The estimated auto-regressive parameter for the AR(1) GLS model implied a 0.41 correlation between two lags of counts of *S. salar*.

For *S. trutta,* the autocorrelation between two given lags of count of *S. trutta* was high. The alternation of positive and negative values, for both the autoregressive and moving average coefficients, was associated with the oscillating decay in the autocorrelation function of the time-series (Appendix A.iii, Figure A.7.17).

Coefficients

A plot of the fitted and observed values for each species counts (Figure 4.6) showed that the fit of both models was high (for *S. salar* Adj.R^2=0.6664, F=364.6 on 1 and 181 degree of freedom, *P*-value<0.001 and for *S. trutta* Adj.R^2=0.8177, F=817.4 on 1 and 181 degree of freedom, *P*-value<0.001).

The backwards selection using likelihood ratio tests between nested models provided the coefficients in Table 4.7.

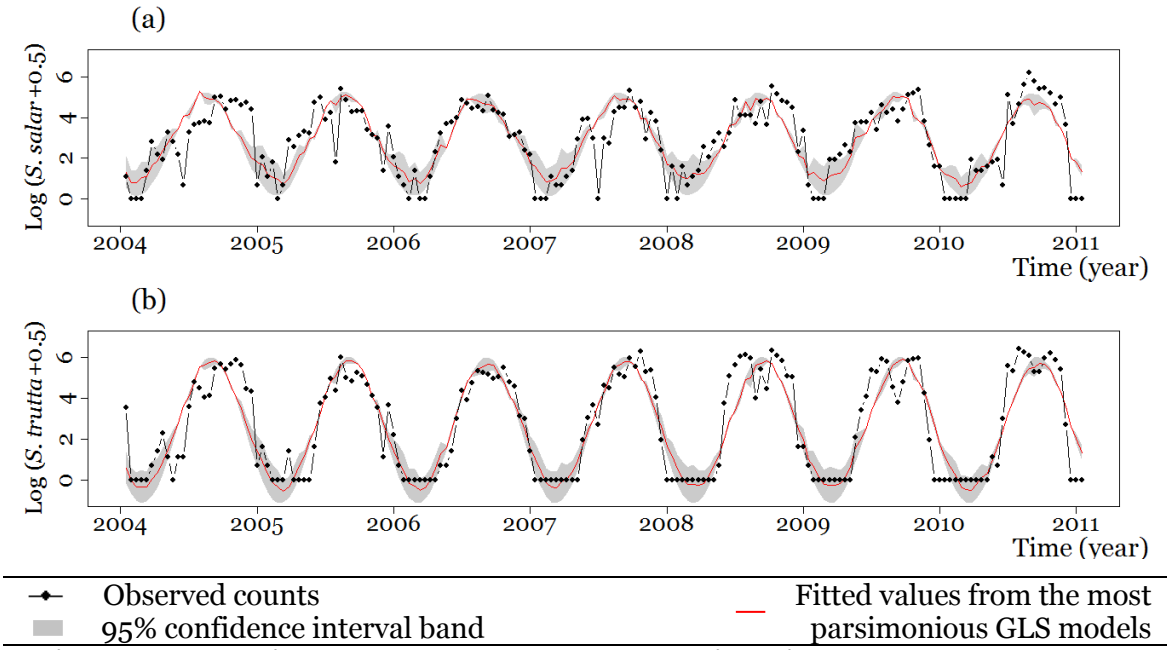


Figure 4.6: The fitted values from the most parsimonious GLS models and the observed counts of *S. salar* and *S. trutta* over the whole observation period.

The variations in counts of both species were largely explained by the annual periodicity parameters (in *S. salar* χ^2=70.32, *P*-value<0.0001 and in *S. trutta* χ^2=72.13029, *P*-value<0.0001). All fish counts also increased with every unit of NAO (increase of 35.01% in *S. salar*, χ^2=10.09, *P*-value=0.0015 and increase of 18.51% in *S. trutta*, χ^2=4.94776, *P*-value=0.0261).

Table 4.7: The estimation of the coefficients for each parameter in the most parsimonious GLS models for the count of each species.

	S. salar		*S. trutta*	
intercept	2.9379	*P*=0.0000	2.6888	*P*=0.0000
NAO	0.3002	*P*=0.0013	0.1699	*P*=0.0243
cos(period 1 year) = *a*	-1.5162	*P*=0.0000	-1.7080	*P*=0.0000
sin(period 1 year) = *b*	-1.4131	*P*=0.0000	-2.6314	*P*=0.0000

The models estimated the coefficients *a* and *b* of the Fourier series, which were used to determine the amplitude *c* and phase angle *d* of the harmonic (Table 4.8).

Table 4.8: The estimation of the amplitude *c*, the phase angle *d* and occurence of peak *e* for the time-series of count of each species.
The coefficients are estimated from the fourier coefficients a and b estimated by the most parsimonious GLS models. The value π was added to the value d, as $a<0$ for both species; the week during which the peak was predicted to occur was obtained by multiplying the value e by 52 (weeks in a year).

	a	b	$c = \sqrt{a^2 + b^2}$	$d = arctan\left(\frac{b}{a}\right)$	$e = \frac{d}{2\pi}$	Peak (week)
S. salar	-1.5162	-1.4131	2.05	3.89	0.6195	32.2
S. trutta	-1.7080	-2.6314	3.13	4.13	0.6587	34.2

As expected, the amplitude of the waveform was smaller for *S. salar* than for *S. trutta*. The phase angle values suggested an earlier peak in *S. salar*, as illustrated in Figure 4.7 (2weeks earlier in *S. salar*). The waveforms based on the periodicity of the transient periods, detected by the harmonic analysis, are also represented. The sum of both waveforms illustrated the shape of a hypothetical waveform containing both the fundamental and transient periodicities.

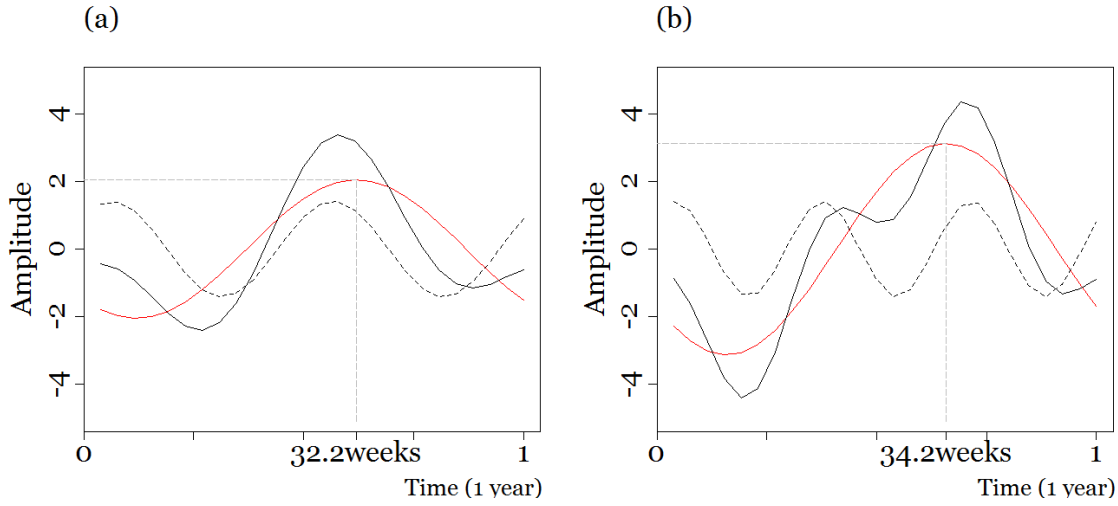


Figure 4.7: The synthesis of hypothetical waveforms relevant to the description of the time-series of count of *S. salar* and *S. trutta*.
The hypothetical waveforms are based on the sum of the fundamental harmonics (of period one year) and secondary periodicities as detected on the wavelet transforms for (a) *S. salar* and (b) *S. trutta*. The models predict peaks occurring at 32.2 (*i.e.* first-second week of August) and 34.2 weeks (*i.e.* mid to last week of August).

The shape of the waveforms was different for both species and altered by the addition of the second sinusoid. The relevance of transient periodicities may indicate a better suitability of these waveforms over the more simple waveforms defined by the fundamental periodicity only.

Analysis of residuals

The normalised residuals were centered around zero across years and independent (Appendix A.iii, Figure A.7.18) but outliers occurred (low outliers in the middle of most years, and the spread was wider at the beginning and end of each year on several occasions.

For the model of counts of *S. salar*, the residuals were distributed asymmetrically, left skewed and clearly bimodal. The two distinct peaks existed on each side of the zero value (Figure 4.8a) and parted from the normal distribution (Figure 4.8b). The shape was characterized by two modes and heavy lower and upper tails.

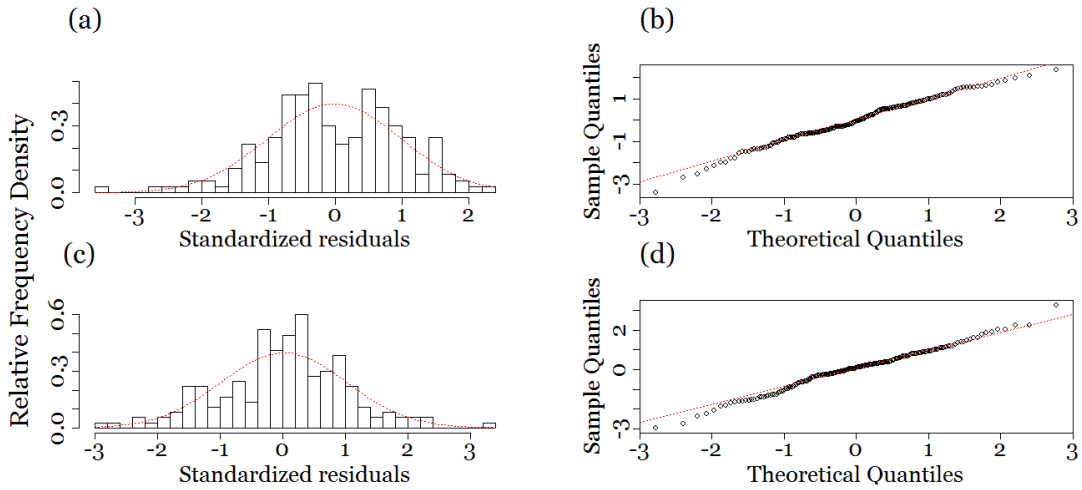


Figure 4.8: For *S. salar* and *S. trutta:* respectively, (a&c) frequency histogram and (b&d) Q-Q plot against normal distribution, for the residuals of the most parsimonious GLS models.

For *S. trutta*, the residuals distribution was asymmetrical, skewed to the right with events being mostly typical but with a plateau on the negative side in relation to the normal distribution (Figure 4.8c&d), rendering the frequency distribution seemingly bimodal and with heavy lower and upper tails.

4.3.5.Forecast accuracy of selected models

The forecasting accuracy of the most parsimonious GLS models was tested on the data for the year 2011, by predicting the counts of each species based on the variations of the river covariates (Figure 4.9).

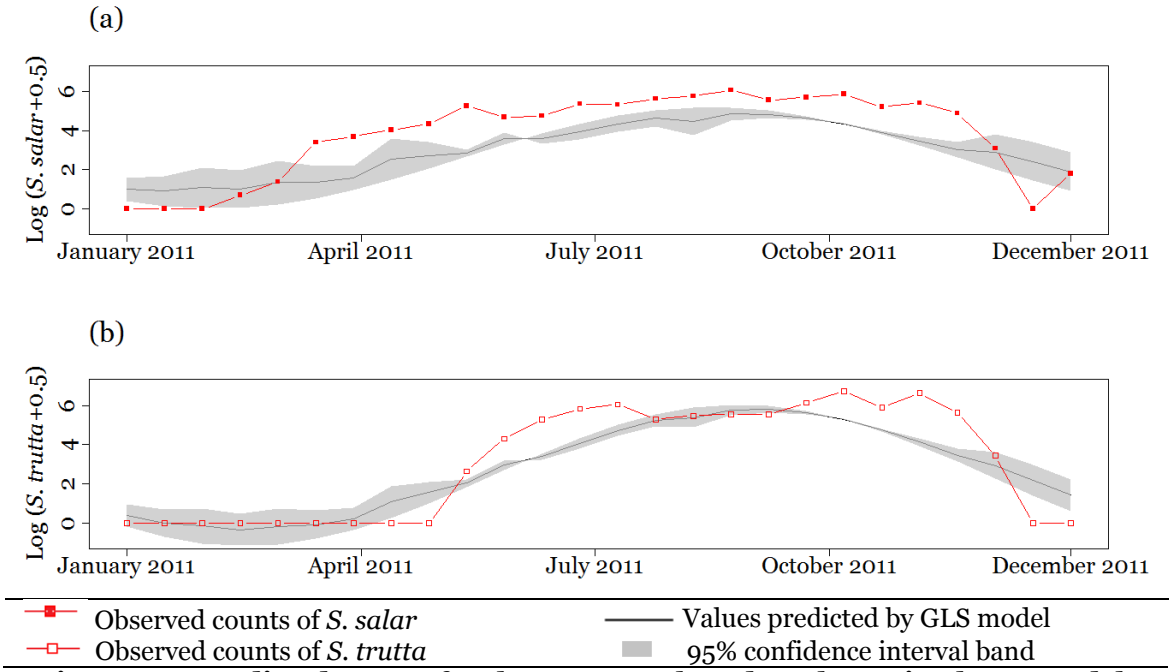


Figure 4.9: Predicted counts for the year 2011 based on the optimal GLS models and the environmental data for the year 2011, with observed counts of *S. salar* and *S. trutta*.

The correlation between observed and predicted values was high for both species, with the models explaining 75 to 84% of the variations in counts (Table 4.9).

Table 4.9: The intercept and slope of the linear regression of observed versus predicted counts and the Adjusted R^2 values, for *S. salar* and *S. trutta* for the year 2011.

	S. salar	*S. trutta*
Intercept	-0.37 P=0.49	-0.20 P=0.58
Slope	1.39 P<0.0001	1.19 P<0.0001
Adjusted R^2	0.7518	0.8499
	F=76.74, df=24 , P<0.0001	F=142.5, df=24, P<0.0001

The models lacked accuracy, reflected by the negative and non-significant intercept values of the linear regression. The significant slope values indicated a good representation of the trend, with some level underestimation (71% to 84% underestimation of observed counts), but the prediction was only over one year.

4.3.6. *Relative proportions of species*

The counts of *S. salar* and *S. trutta* (Figure 4.10a&b) were represented by local regressions loess degree 1 and span 0.3 (Appendix A.iii, Figure A.7.19 and Figure A.7.20). This span value implied that each local regression was fitted using 30% of the time series (*i.e.* 54 points or 6 months). The ratio of the two regressions presented a fluctuating temporal pattern (Figure 4.10c).

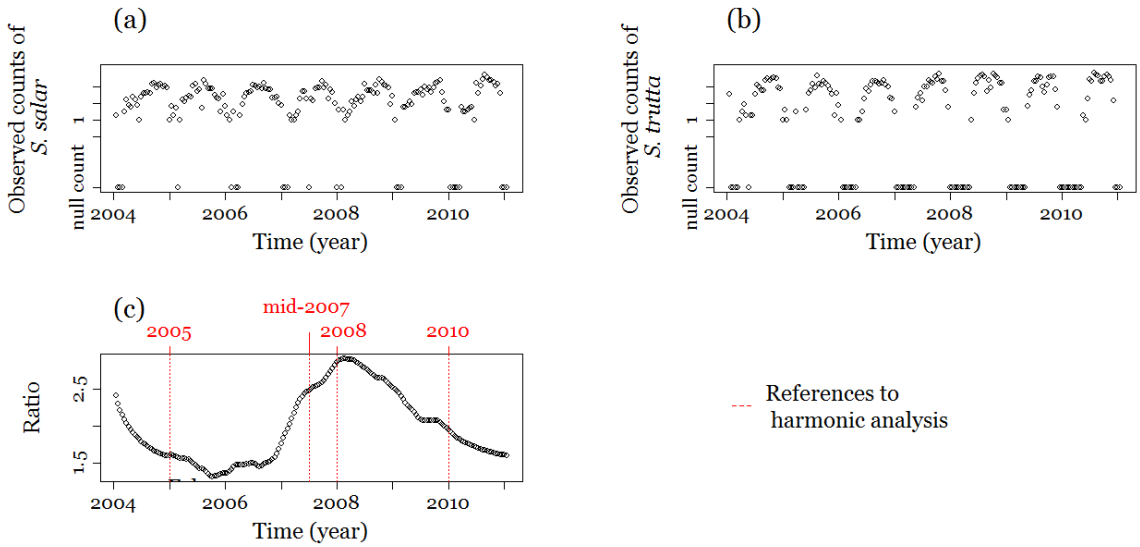


Figure 4.10: The counts of (a) *S. salar* and (b) *S. trutta* and (c) the ratio of the loess function for counts of *S. trutta* over the loess function for counts of *S. salar*.
The values are plotted over the whole observation period.

To a certain extent, the temporal pattern may be matched to shifts in synchronicity in the fluctuations of both species counts, observed on the coherence wavelet analysis. Outputs of the wavelet coherence function were not interpretable before 2005 and after 2010. Between 2005 and mid-2007, peaks in counts of *S. salar* led peaks in *S. trutta*; from mid-2007 to 2008 the species were in phase; *S. trutta* led *S. salar* between 2008 and 2010. Also, in *S. trutta* the transient periodicity started from 2008 onwards. Both the shift in lead species and the appearance of a transient periodicity in *S. trutta* were synchronized with fluctuations in the relative proportion of each species.

4.4. Discussion

In this chapter, models were developed to investigate the extent to which counts of *S. salar* and *S. trutta* can be modelled separately in the River Tyne, based on video records collected at Riding Mill weir and a set of environmental covariates. The study relied on information collected over the four channels in the weir. As species identification was not undertaken over the entire the weir, the extent to which salmonid movements in one channel reflect movements of the whole population cannot be determined. Records were assumed representative of the population throughout the observation period.

For both species GLMs were overdispersed and did not respect the assumptions of normality, homogeneity and independence of residuals (Breslow, 1996). This lack of conformity underlined a disparity between the ecology of both species and the principles involved in GLMs. This type of model assumes a linear relationship between the predictors and the time-series and independently distributed errors (Breslow, 1996), implying that fish counts do not influence the next counts, *i.e.* fish are solely responding to external conditions. Failure to comply with this condition may reflect a behavioural feature of the migration; fluctuations in salmonid numbers observed during the year may be influenced by both environmental variation and population behaviour.

Mixed-effect models allowed the relationship between response and covariate to vary according to predetermined groups (Pinheiro, 1994). For both species, LME models proved to be unproductively constraining in comparison to GLS models, as a random effect did not provide a better model. The latter were consequently used and explicitly accounted for characteristics of heteroscedasticity and autocorrelation contained in the response data (Fox, 2002).

High standard deviations in relation to the means, and evidence of a tendency of residuals to increase with higher predicted values, suggested the relevance of the log-transformation to stabilise the variance (Leydesdorff and Bensman, 2006). The most parsimonious GLS models underlined the importance of annual periodicity in the prediction of counts of each species in the River Tyne. Values of NAO (no lag) were also important predictors. For both species, the year count was used to modulate the variance function, possibly as a surrogate

for between-year variations due to recruitment during the time spent out at sea. The GLS models were of good fit and the prediction of year 2011 displayed good forecasting properties.

The correlation structure of the optimal GLS model for counts of *S. trutta* was complex (5 components for AR and MA parameters) indicating a complex and/or changing population structure with interdependence (Zuur *et al.*, 2009). Also, the residuals of the most parsimonious models followed a bimodal distribution, indicating that these models failed to account for a process inducing bimodality for at least some of the observation period.

The stationary process described by the GLS models implied a constant distribution of the error and variance over time (Fox, 2002), which was a possible cause for bimodality in the residuals. The Fourier series may, for instance, attempt to account for a non-stationary process. If the seasonality was incorporated as a constant component while it was in fact changing over time, then the change in question may have engendered a second distinct mode in the residuals, showing that the process is not being accounted for by the model.

The harmonic analysis provided evidence of a shift in periodicity during the year 2008 especially for counts of *S. trutta*. For this species, transient periodicities were increasingly significant from 2008, indicating that the waveform progressed towards an imperfect sinusoid (Diggle, 1990). The Fourier coefficients, as estimated by the most parsimonious model (Jakubauskas *et al.*, 2001), confirmed that the assimilation of the transient periodicity to the fundamental periodicity generated waveforms with double peaks, the first being of lower amplitude than the second. The plot of the fitted values and observed counts of *S. trutta* over time exacerbated this pattern (two peaks in counts from year 2008 onwards on Figure 4.6, page 98). As *S. salar* did not exhibit this trend, a change in the phase of the two time-series also occurred in 2008 onwards.

The wavelet coherence function underlined that the phase shift varied throughout the observation period (Grinsted *et al.*, 2004). The lead in arrivals was first for *S. salar* and started shifting towards a lead for *S. trutta* in 2008. This implied that the waveform corresponding to counts of *S. salar* had a tendency to develop earlier than the waveform of counts of *S. trutta* until 2008,

when the opposite trend started. However, it is likely that the double peak from year 2008 onwards implied that the first peak in counts of *S. trutta* occurred earlier than for a single annual peak as in the years prior to 2008. This explained the reversal tendency in lead species. The proportion of *S. trutta* was always higher than *S. salar* but for the same reason, this proportion increased until 2008 when it reached a peak after which it decreased again. From 2008 onwards, the peak in *S. salar* was likely synchronised with a decrease in *S. trutta* (due to a peak before and after) hence decreasing the value of the ratio.

Geographic ranges of *S. salar* and *S. trutta* are distinct over the winter season: in summer both species hatch, grow and later breed in the River Tyne but in winter, *S. salar* migrates towards the West of Greenland (Dadswell *et al.*, 2010) while *S. trutta* remains along the coast of Northern Europe (Jonsson and Jonsson, 2011). This has implications for the origin of the phase shift reversal observed in 2008.

The analysis suggested that a change occurred in the temporal distribution of counts of *S. trutta* from 2008 onwards, which did not affect *S. salar*; or at least, not to the same degree.

First, the two species may respond differently to a change that occurred on their common ground, *i.e.* a change occurred in a resource used by *S. trutta* which *S. salar* does not depend on. For instance, *S. salar* fast in the River Tyne during upstream migration while *S. trutta* continue to feed (Klemetsen *et al.*, 2003). Therefore, fluctuations in feeding opportunities may alter numbers of *S. trutta* without affecting *S. salar*. However, as feeding opportunities in the river are likely reflected by the available covariates used in the model, this hypothesis was unlikely as it considered the difference between species without explaining the change occurring in 2008.

Secondly, environmental factors not experienced by both species, where and when the two species occupy different grounds, such as the North Sea, where *S. trutta* remains over winter while *S. salar* migrates further into the Atlantic (Jonsson and Jonsson, 2011). Thus, *S. trutta* spends a large part of the year independently in the North Sea basin. During the first years of the observation period, the wavelet spectrum showed secondary periodicities, however weak and of low significance. It may be suggested that the secondary periodicity may have

in fact been the amplification of an existing trend which was initially weak and intensified over time until becoming significant in 2008. Similarly in *S. salar*, the secondary periodicity was weak throughout the whole observation period. An explanation may be that the reversal observed over time was linked to fluctuations of a resource that, again, *S. trutta* relied on and which is not as important to *S. salar,* or for not as long; which was consistent with this resource being located in the North Sea. Thus, the covariate likely affected the fish as they were about to come upstream, at the estuary level. The reversal observed in 2008 may be due to a change in the behaviour of *S. trutta* due to change in the North Sea (*i.e.* abiotic parameters or abundance of an interacting species), or a change in a covariate in the North Sea for which measurements were not available for this study.

Thirdly, the switch from a sinusoidal waveform towards a bimodal trend in 2008 may be anevidence of the existence of two distinct subpopulations of *S. trutta* in the River Tyne. A comparable temporal bimodality has been observed in returns of sockeye salmon (*Oncorhynchus nerka*) with an early and later run responsible for two peaks every year (Canada, Fillatre, 2002). These two runs were genetically distinct despite the absence of geographical or physical barriers such as the reproductive isolation characteristic of the early stage of speciation. Ecological speciation occurs when populations becomes reproductively isolated due to evolutionary processes driven by ecological factors (Rundle and Nosil, 2005). Spatial isolation may occur without barrier and at microgeographic scales (Fillatre, 2002). If the two peaks in counts of *S. trutta* were to represent two genetically isolated populations, the mechanisms behind the division may be both spatial (*i.e.* occupying different areas in the North Sea or spawning grounds) and/or temporal (*i.e.* difference in return timing).

This hypothesis requires genetic analyses and a comparative study of the life history characteristics of both runs (body size, sex ratio, age at maturity, time in river and sea) in order to confirm the ecological speciation argument (Hendry *et al.*, 2000). However if the population was considered to be two distinct runs, two distinct management strategies may be required. Non-stationarity under this hypothesis would be due to the two runs. The two runs may have been synchronised during the first half of the observation period, and then have

become temporally distinct. If both runs were spatially distinct, a marine covariate may have affected one run but not the other.

The fastest settling of isolated reproduction documented in *S. salar* was fewer than 13 generations (Hendry *et al.*, 2000). Speciation may manifest as differences in phenotype (*e.g.* size difference in Sockeye salmon, Hendry *et al.*, 2000), but genotype adaptations have also been observed that had behavioural implications (*e.g.* changes in migratory orientation and timing that may alter mating strategies and select against hybrids in European blackcaps *Sylvia atricapilla*, Hendry *et al.*, 2007).

Over an observation period of seven years, the present study does not pretend to demonstrate reproductive isolation. Rather, it suggests that the population of *S. trutta* of the River Tyne may have consisted of two reproductively isolated populations over many years, that possibly became temporally or geographically distinct; a difference between the two groups may exist in their migration grounds and/or timing.

The most parsimonious models underlined the importance of annual periodicity in predicting temporal variation in counts of each species in the River Tyne. Annual seasonal cycle is a known controlling factor of temporal variability of many marine ecosystems (Bertram *et al.,* 2001), several natural process occur with annual periodicity such as northern hemisphere sea temperature (Hansen *et al.*, 2012).

Values of NAO were also important predictors. The values of the NAO index are linked to climatic conditions over a considerable part of the Northern hemisphere (Wang *et al.*, 2010) and relate to sea surface temperature and circulation in the North Atlantic (Visbeck *et al.*, 2001). The NAO is consequently linked to thermohaline conditions that determine habitat suitability for it during return migration (Holm *et al.*, 2000).

The importance of marine conditions four years prior to returns was demonstrated for the salmonids of the River Tyne in Chapter 3; then, as the two species were modelled jointly, it was their common requirements and traits that were modelled, *i.e.* strong habitat requirements as the post-smolt stage and/or parent population. In this chapter, traits modelled for each population.

Conditions for marine dispersal of *S. trutta* are not well known (Malcom *et al.*, 2010). Some evidence of behavioural differences between males and females was found, with the males migrating further (Bekkevold *et al.*, 2004). Also, there is a 3:1 ratio in the number of females to males migrating as only 25-30% of males migrate (Rasmussen, 1986). Females migrate towards the sea after 1 to 3 years spent maturing in freshwater, implying multiple age-classes (Bekkevold *et al.*, 2004). It is also possible that a part of the population of anadromous *S. trutta* of the River Tyne does not migrate as far as the Baltic Sea, but rather remains in the Tyne estuary (*i.e.* slob trout, commonly observed in Scotland) and may re-enter freshwater frequently (Malcom *et al.*, 2010). Overall there is a high variability in the dispersal of *S. trutta* that go to sea, (Malcom *et al.*, 2010). Also geographically, adult *S. trutta* display a wide ranging migration which is often local, mostly coastal, but sometimes offshore, for undetermined reasons (Jonsson and Jonsson, 2011; Malcom *et al.*, 2010)

Kelt account for the majority of *S. salar* caught in west Greenland, *i.e.* multi-sea winter individuals (over 90% of catches, Gauthier-Ouellet *et al.*, 2009). There is currently no certainty regarding the winter feeding grounds used by grilse of the same species, as the Greenland and Faroese fisheries are the main providers of data concerned with winter oceanic counts of *S. salar* (Malcom *et al.*, 2010). Two sampling regions may however not be representative of the distribution *S. salar* in the whole Atlantic. In fact, several studies suggest that *S. salar* grilse may use a marine habitat that is much broader than the west Greenland region; this habitat is currently poorly defined but it may be because its definition involves large scale ocean currents (Dadswell *et al.*, 2010; Malcom *et al.*, 2010). This theory has implications for the marine migration route of *S. salar*. It can be assumed that the marine migration towards the River Tyne starts from a wide range of locations and that *S. salar* undergo a broad spectrum of marine conditions during their journey and level of their influence may depend on the age class of the individuals.

Although important, differences exist within species, both shared common requirements of habitat suitability, as described by the models. A possible mechanism for the link between NAO values and estuary conditions is the difference in temperature between marine and estuary the NAO is linked to; these differences condition the exploitation of optimal thermal habitat (Attrill

and Power, 2002). The commonality between species may lie in the estuarine conditions, occurs when approaching the estuary for the spawning migration, reflected by the relevance of the simultaneous NAO values in the most parsimonious models for both species.

The relationship with the NAO values was twice greater for counts of *S. salar* than for *S. trutta* (0.30, *P*-value <0.01 and 0.17, *P*-value <0.05), indicating that temporal fluctuations in counts of *S. salar* respond to more broad-scale marine conditions, such as the NAO index, than *S. trutta,* which remains more locally distributed.

Simultaneous values of NAO are likely not to correspond to mortality, but rather to a varying gradient of habitat suitability. The distribution of thermal habitats moves with the NAO in the Atlantic Ocean; isotherms undulate and draw a suitable thermal path for *S. salar*, that is required to reach the targeted estuary of the natal river (Friedland, 1998). When, in May, the warm (8-10°C) isotherm extends towards the Norwegian coast in May, survival is better for both grilse and kelt (Friedland, 1998).

For both species, the year was used to modulate the variance function. Annual change of intercept was possibly a surrogate for between-year variation due to survival during the time spent out at sea, potentially accounting for the maturation rate known to differ between years for unclear reasons (Friedland, 1998).

For *S. salar*, these conclusions are in accordance with the major findings of I.A.S.R.B. (Hansen *et al.*, 2012) that broad-scale factors regulate the abundance of *S. salar* and that long-term changes are linked to the northern hemisphere sea temperature and NAO. Further I.A.S.R.B. studies suggest that the decline is stronger in multi-sea-winter individuals and that the impact of marine parameters varies between age classes of *S. salar*. These factors were not accounted for by models in this study and may help explain the underestimation of abundance by the GLS model.

Finally, counts from Channel 4 may not represent trends occurring on the whole width of the weir, nor for entire populations. In *S. trutta*, this may imply that the peaks observed were indicative of a decrease in the use of this channel in mid-summer, rather than a decrease in the overall counts of *S. trutta*. A quick

visual comparison with the counts provided by the fish counter confirmed that the double peaks were not observed consistently and did not match the video records of *S. trutta*.

The fish passing through Channel 4 may be a non-representative sample of the populations; they may be smaller fish that prefer to travel up the edges of the weir in order to avoid high water velocity (Tang *et al.,* 2000). Under this assumption, the decrease in counts of *S. trutta* in channel 4 may indicate that the population contained a larger proportion of large fish during this time; and that the large fish used the central channels

However, this shift would still account for a change in the global population structure. Consequently even if the video records were not representative of the global population trends, the change observed in 2008 was likely meaningful at the population level. There were no species-specific data relating fish size and channel use; comparing size between channels would not be informative as both species were confounded. This hypothesis was consequently not verifiable with the available data.

However, both species were shown to have distinct responses to the covariates. A model is needed, that is able to account for the potential non-stationary and species-specific variation suggested by this study; state-space models will be used in the next chapter to investigate this issue.

4.5. Conclusions

Whether video records reflected a behaviour specific to Channel 4 or of the whole populations of *S. salar* and *S. trutta*, is lekely that these observations had implications concerning the whole populations.

A shift was observed from 2008 onwards in several trends and hypotheses such as ecologically driven speciation in *S. trutta* may be investigated by a study of the genetics of the population as well as life history traits and would justify telemetry studies to investigate dispersal and time partitioning within both the River Tyne and the North Sea environments.

The most parsimonious models included temporal patterns in abundance of *S. salar* and *S. trutta*. The models accounted for characteristics shared by all individuals in the River Tyne. The relevance of simultaneous NAO values reflected that near-shore conditions during return migration mostly explained fluctuations in abundance of the two species.

Differences in migration patterns are also known to occur within species, that were not accounted for by a single model and likely explained the occasional lack of fit of the models. A better understanding of the age groups contained in the species counts may help improve the fit of the models.

Autocorrelation in population movement, changing periodicity, and transient imperfect sinusoids, all suggested that incorporation of time-variation in estimates and non-stationarity in the model may improve description of fluctuations in counts of each species. In the next chapter, state-space models will be used to attempt and describe the time-dependency of the response for each species.

Chapter 5. .

"History does not repeat itself, but it does rhyme".

M. Twain
Quoted in *Diplomacy and its discontents*
J.G. Eayrs, 1971.

Chapter 5. Modelling time dependency in the abundance of the River Tyne salmonid population.

5.1. Introduction

Classical linear regression models assume that a linear combination of explanatory variables can describe the mean of a response variable by including constant coefficients estimated by the model (Diggle, 1990). The model coefficients remain constant over time so the processes described are concerned with the entire observation period: model components such as harmonics describe seasonality with constant frequency and amplitude and non-linear trends can be expressed by a polynomial.

Dynamic generalized linear models differ in that the regression coefficients may fluctuate over time; this allows, for instance, description of changes and evolutions contained within seasonal patterns (Lundbye-Christensen *et al.*, 2009). Observed data may be the result of such dynamic processes. State space models may be used to represent both the relationship between seasonality and time, and the relationship to response data (Chow *et al.*, 2009). The expression "state space" refers to all the possible unobserved states of a dynamic system, which these models aim to describe (Commandeur and Koopman, 2007). Integration of this information in a model may greatly refine the description of seasonality, thus improving the model precision and predictive ability.

In the preceding chapters, the applications of classical linear regression models suggested that the use of dynamic modelling might explain the temporal fluctuations in the abundance of salmonids in the River Tyne more completely. For instance, the assumption of stationarity in the preceding models (Chapter 4) may not represent the time-series accurately. Stationarity implies that the relationship between mean and variance is stable over time; it is an underlying condition of GLMs, the failure of which tells us that stationarity may not be present. GLS models allowed for the adjustment of the variance structure for a parameter (*i.e.* annually) but a lack of fit in the previous chapter suggested

room for improvement. The annual peaks that characterize temporal fluctuations in salmonid counts are a seasonal phenomenon but the variations they contains are may not be fully described by a stationary process. Additionally, transient periodicities have been shown to arise within the count data during the second half of the observation period, which classical models did not address. State space models are dynamic and do not require the time-series to be stationary (Commandeur and Koopman, 2007).

The use of state-space models in this chapter tests the hypothesis that the time-series of salmonid counts contain dynamic properties that evolve through time and were not directly measured (Commandeur and Koopman, 2007; Pichler, 2007; Zivot, 2006).

State space models quantify a process using a minimum of two separate mathematical models, each described by a distinct equation: an observation and a transition equation(s). The observation (or measurement) equation describes the relationship between state variables and the response; it contains the sum of all states at time *t,* and accounts for the error linked to measurement noise and inaccuracy (Orderud, 2005). The transition equation is dynamic; it defines the vector of state variables. The state variables are possibly unmeasured; this equation represents the evolution of the hidden stochastic processes through time, assumed to be driving the state variables (Pichler, 2007).

The time-dependency is explicitly expressed in the unobservable states by a Markovian process (x_{t+1} depends on x_t) and are associated with state noise vectors, *i.e.* error for each innovation (Orderud, 2005). The transition equation consequently describes the probability density function of a hidden process.

An observation includes both a signal and a noise; considering that the process described by the state variables is reflected by the response data but not directly observable, the signal contained in the response variable corresponds to the state variables, gathered in a state vector at time *t* (Chatfield, 2009). The signal (*i.e.* the state vector) is the observation *minus* the noise; it equals a filtered version of the observation. So, the state equation estimation uses a filtered version of the data.

Several filtering methods may be used for the state estimation protocol. The Kalman filter (Kalman, 1960) is robust and widely used for the estimation of both linear and non-linear state dynamics (Orderud, 2005).

The Kalman filter operates in two-steps: (1) forecasting and (2) updating. Firstly, the observation at time *t* is calculated as a forecast of the observation at *t-1* (the series consequently lags one time point in comparison to the input data, Commandeur and Kooper, 2007). The predicted means are augmented by the corresponding variance and the sum, constituting the Kalman filtered state. Secondly, the filtered state is updated sequentially by including all available observations (*i.e.* point by point considering all data, not only the preceding point) and this updating process adds new information to the model by describing the relationship between data points in reverse temporal order compared to the initial forecast process (Zivot, 2006). The resulting vector is augmented with the estimated variance, yielding the smoothed state.

Kalman filtering of x implies that state x_t evolved from state at x_{t-1} via state transition model G_t, augmented by error w_t; the state equation is:

Equation (5.1)	$x_t = G_t * x_{t-1} + w_t$	$w_t \sim N(0, W_t)$	(state equation)

The observation equation of y at time t contains the state x_t, the measurement error v_t, and an observation model H_t that maps the state space into the observed space:

Equation (5.2)	(2) $y_t = H_t * x_t + v_t$	$v_t \sim N(0, V_t)$	(obs. equation)

In brief, state space models describe the observation state, transition process, and initial observation value, as schematized on Figure 5.1.

The resulting (possible) time-variation in the state-space model coefficients is a powerful feature of state-space models (Aoki, 1987) and justifies a cautiously determined set of candidate models in order to avoid overfitting of the covariate estimates. State-space models will be developed to investigate whether modelling salmonid counts in the River Tyne can benefit from time-varying estimation of environmental parameters.

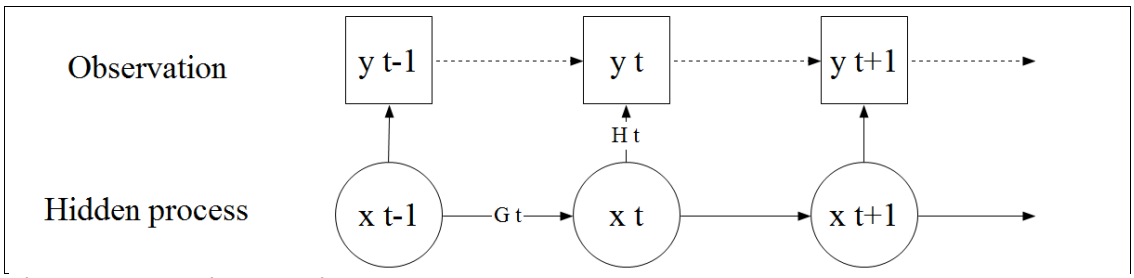


Figure 5.1: An illustration of the structure of state-space models.
The value of x_t depends on the value of x_{t-1} and the value of y_t depends on the value of x_t. The hidden process is estimated with the Kalman filter G_t. The observation is measured (response and measured explanatory variables).

Here, the covariates included in the candidate models were based on inferences from preceding chapters. Covariates were described in relation to salmonid counts at time *t* (Figure 5.2). Whether a time-varying coefficient was assigned to a given covariate was established before model fitting.

Model development involved several time frames and related time-dependent processes occurring simultaneously: life history, migration, climate, and monitoring (Figure 5.2). On the one hand, processes influencing the salmonid abundance at time *t* likely took place over several years preceding the observation (Figure 5.2, relevance of NAO with a 4 year lag, Chapter 3). On the other hand, the static nature of covariate measurements used for modelling a migratory population justified the investigation of time-variation in most environmental covariates. Time-variation in this case was a substitute for the varying distance from the point of covariate measurement as the salmonids migrate.

Population dynamics were described by periodic fluctuations with potential annual variation in amplitude and phase. Time-varying coefficients allow for explicit description of the year-by-year variations in the frequency and phase of both the fundamental and transient seasonal patterns (Fanshawe *et al.*, 2008). Population dynamics may be described explicitly in order to disentangle them from climatic and ecological drivers, especially as these drivers are often spread over different scales in studies of migratory species; the separation of these population dynamics and environmental processes can be achieved via state space modelling (*e.g.* migratory wader species, Robinson *et al.*, 2009) leading to a more complete description of the salmonid migration process.

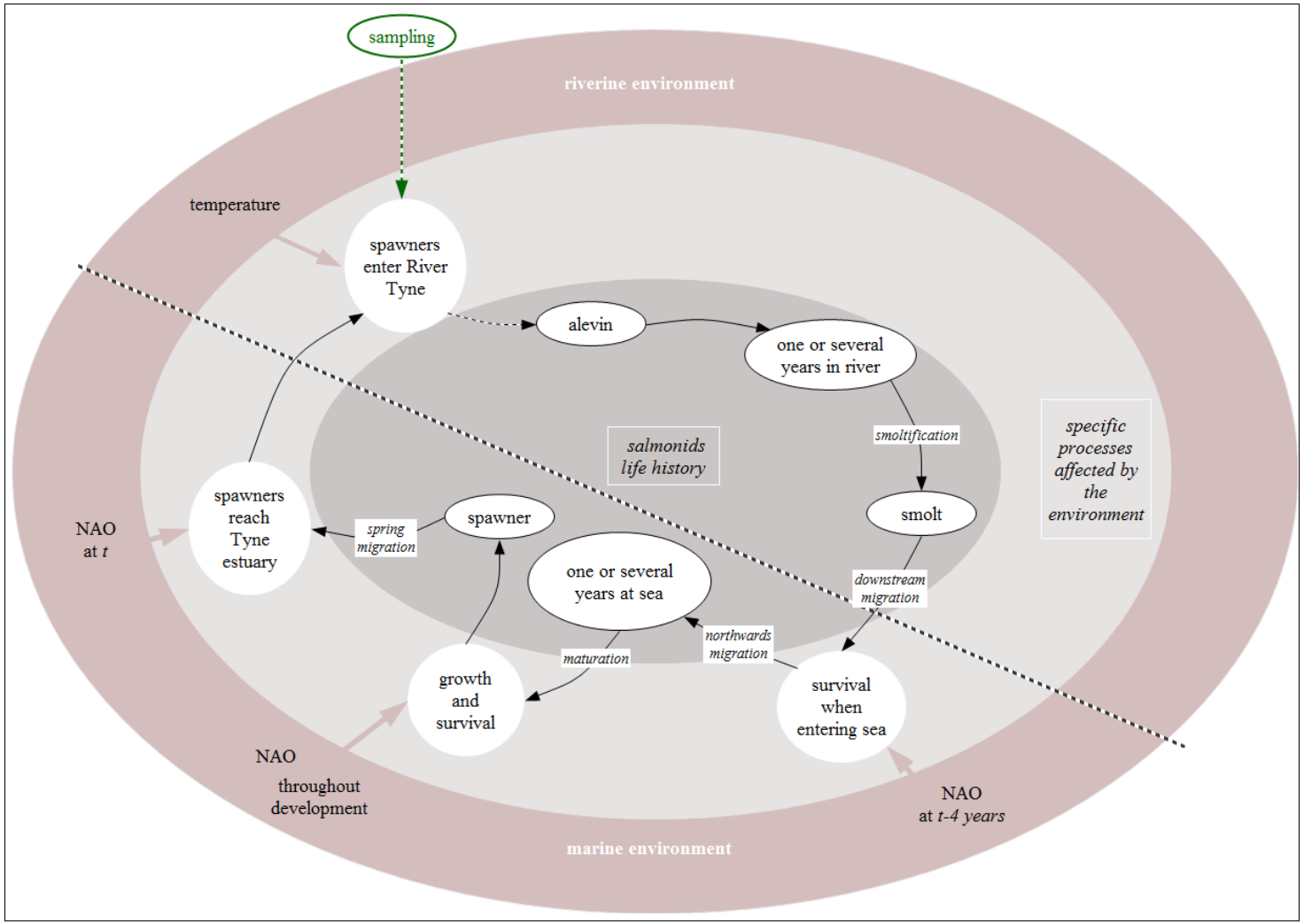


Figure 5.2: The life history and selected biological processes relevant to the salmonid migration, and relationship with the salmonid data and the environmental covariates available.

At a given sampling time, the observed spawning population count is the combined result of life history characteristics of the salmonids and the influence of environment on the processes investigated.

NAO conditions four years prior to being recorded in the River Tyne likely influence mortality as smolts enter the sea. NAO conditions throughout maturation can also influence growth and survival. Current NAO conditions likely influence the habitat suitability for salmonids as they reach the River Tyne estuary.

Seasonal fluctuations in river temperature likely condition the suitability of spawning habitat and salmonid movements from the estuary to the breeding sites.

Monitoring of salmonid counts and temperature in the River Tyne is continuous but represents only a small part of salmonid life history.

The last section in this chapter is concerned with the possibility that temporal variations partly emerged from problems identifying salmonids as *S. salar* or *S. trutta*. This is implied by a species identification rate increasing throughout the observation period, indicating learning. The learning process was modelled by a Michaelis-Menten equation (Menten and Michaelis, 1913, translated by Johnson and Goody, 2011) that traditionally describes enzymatic reaction rates dependent on substrate concentration and nutrient availability, forming a rectangular hyperbola with a plateau at the maximum reaction rate (Bertolazzi, 2005). It is suggested that a similar relationship can describe learning rate as a function of the time spent observing salmonids. Inverse modelling confronts a mathematical model with the data and allows modelling unmeasurable processes by applying their specific mathematical definition to a process (Soetaert and Petzoldt, 2010); species identification learning rate according to the Michaelis-Menten equation will be confronted with the observed data in order to investigate this hypothesis.

5.2. Material and Methods

5.2.1. Data

Video camera records from Channel 4 at the Riding Mill counter (Figure 4.1, page 85) produced two univariate time-series counts of *S. salar* and *S. trutta* during the data collection period (2004 to 2012). As for analyses at the population level (Chapters 3 and 4), all data were pooled into 14 day bins and log-transformed. All values contained in the dataset were determined (no NAs). The covariates included were previously identified as influentialat the population level. Specifically, these covariates were: NAO (without lag and with a 4 year lag), temperature at the Riding Mill station, the fundamental annual periodicity, and the transient periodicity specific to fluctuations in each species.

5.2.2.Analysis

As the standard approach to develop state-space models (Pichler, 2007), the prior distribution of the error was assumed normal. The state-space model calibration was diffuse for the static parameters of the system (*i.e.* coefficients were estimated by maximum likelihood within the model). Estimation of the state equation components used the recursive and sequential Kalman filter (Kalman, 1960) applied to the two log-transformed time-series, followed by smoothing. The model parameters were augmented as part of the state vector with the smoothed series; the ensemble constituted the values fitted by the state-space model.

Parsimonious state-space models were selected according to their ranking based on an adaptation of AIC, as recommended by Commandeur and Koopman (2007):

Equation (5.3)

$$AIC = \frac{1}{n * (-2n * \log(likelihood) + (2(q+w)))}$$

Values q and w were the number of hyperparameters in the model; q was the number of diffuse initial values in the state (*i.e.* the parameters); and w was the total number of estimated error variances (*i.e.* the number of state equation or coefficients to be time-varying). Indices of goodness of fit were the Pearson correlation coefficient, the variance residuals and the log-likelihood of the model.

The normality of the distribution of forecast errors was assessed using standardized predictions; these were obtained by dividing the forecast error by the square root of the forecast variances for the observations (Petris, 2011).

5.2.3. Candidate models

Models were fitted without intercepts as the dynamics modelled by the periodicity components and the intercept were similar (*i.e.* variations in amplitude and phase).

The equations for the state space models were so that:

Equation (5.4)	$y_t = x_t$	$+ \beta.temperature_t + \delta.NAO_t + \zeta.(4\ year\ lagged\ NAO)_t + a.cos(\omega_1 t_t) + b.sin(\omega_1 t_t) + c.cos(\omega_2 t_t) + d.sin(\omega_2 t_t) + v_t$ $\quad v_t \sim N(0, V_t)$
Equation (5.5)	$x_t = \gamma\, x_{t-1}$	$+ w_t \quad w_t \sim N(0, W_t)$

The estimated coefficients (β, δ, ζ, a, b, c and d) were constrained to null, or constant or time-varying values according to the model assumptions. A time-varying coefficient implied an additional equation described the transition relating to the estimated covariate (Commandeur and Koopman, 2007). Omitted covariates implied that the corresponding coefficients were set to null; constant estimates held similar assumptions as in the classic regressions formerly used. Estimates were allowed to vary with time based on logical assumptions (Table 5.1).

Candidate models implied underlying hypotheses depending on the justification for applying a time-varying coefficient to a given covariate. These justifications were determined prior to model fitting (Table 5.1).

Table 5.1: The justifications for applying time-variation or not to the coefficients of the covariates of candidate models.

Parameters	Time-varying coefficient	Underlying hypothesis regarding time-variation	In-thesis reference
Temperature	Yes (β)	The response to temperature varies through the year	Chapter 2
NAO with a lag of 4 years	Yes (ζ)	Measurement of the parameter is static while populations move; a parameter may have different implications depending on the time of influence (*e.g.* as smolt or as spawner)	Chapter 3
NAO	Yes (δ)		Chapter 4
Fundamental period	Yes (a,b)	May account for population dynamics from a year to another (*e.g.* survival)	Chapter 3
Transient period	Yes (c,d)	By definition	Chapter 4
Intercept	No	It would account for dynamics similar to the ones in time-varying harmonics.	.

The estimate β for temperature was allowed to vary with time as it was demonstrated that the response to temperature varies throughout the year (Chapter 2). The underlying hypothesis was that the description of fluctuations in salmonid abundance may be improved by including a time-varying effect of temperature on salmonid numbers throughout the year.

NAO values were also included as potential predictors. Since the data were pooled into bins of 14 days, it was assumed that the salmonids responded to NAO conditions while in the estuary, where they would have spent most of their time. NAO is a broad scale index that describes a large part of the North Atlantic Ocean (Hansen *et al.*, 2012), hence it was assumed representative of the conditions experienced by the salmonids throughout the whole migration route. However the NAO index is also static and the salmonids travel long distances which are not fully determined. The time-variation allowed in the coefficient δ associated with NAO values, aims to reflect the changing distance of the salmonids with regards to the measurement of the NAO. Incorporation of time-variation in the NAO values with a 4 year lag held similar assumptions.

Fundamental periodicity was annual in the counts of both species in that they fluctuated seasonally, with an increase in spring and summer and rare counts over winter (described during the harmonic analysis in Chapters 3 and 4). The estimates (a&b) were allowed to vary in time to account for varying annual amplitude and phase. The transient periodicities (26.4 and 17.5 weeks for *S. salar* and *S. trutta*) were also included and the corresponding estimates (c&d) were allowed to vary with time as they appeared transient in the harmonic analysis of each time-series (Chapter 4). The values for the transient periodicities were also estimated by constant coefficients in case they were in fact constant but confounded with the fundamental periodicity if it was itself time-varying.

The simplest model had constant estimates only and parameter selection was based on AIC. Estimates were made time-varying individually then by pairs, until a full saturated model that contained all time-varying estimates was obtained.

5.2.4.Software

The state space models (including Kalman filtering) were generated in the sspir package (Dethlefsen *et al.*, 2012) operated in the R.15.1 environment (R Core Team, 2012). The package allowed model specification of the dynamic generalized linear models given normal or Poisson errors. The Michaelis-Menten equation to fit the learning curve was developed post-hoc in the FME package (Stoetaert and Petzoldt, 2010).

5.3. Results

5.3.1. Time-varying estimates

State-space models of the log-transformed counts of *S. salar* and *S. trutta* were developed and summarized in Table 5.2 and Table 5.3.

For *S. salar* (Table 5.2), stepwise removal of individual covariates led to the most parsimonious model having temperature (constant estimate) and a time-varying estimate associated with the annual periodicity (model 14). The temporal fluctuations of the estimates are displayed on Figure 5.3.

For *S. trutta* (Table 5.3), the most parsimonious model contained all covariates except for both NAO related covariates. All estimates were constant through time (model 7).

Table 5.2: Candidate state-space model comparisons for time-series of log-transformed counts of *S. salar*.
For each combination, the estimates are either (.) not included, (1) constant, or (tvar) time-varying. The number of state disturbance variances to be calculated (w) and the number of diffuse initial values (q) are the hyperparameters which intervened in the AIC score calculation.

Model	β (Temperature)	δ (NAO)	ζ (NAO, lag 4 yr)	a, b (1 year p.)	c, d (26.24wks p.)	Var (res)	R^2	Log-likelihood	q	w	AIC	AIC rank
1	1	1	1	1	1	2.04	0.6110	-422.55	7	0	845.17	4
2	.	1	1	1	1	1.56	0.6196	-1253.36	6	0	2506.77	6
3	1	.	1	1	1	2.03	0.6023	-416.59	6	0	833.24	3
4	1	1	.	1	1	2.04	0.6041	-416.25	6	0	832.55	2
5	1	1	1	.	1	1.97	0.5178	-443.94	5	0	887.92	5
6	**1**	**1**	**1**	**1**	**.**	**2.05**	**0.6051**	**-406.77**	**5**	**0**	**813.60**	**1**
7	1	1	.	1	.	2.03	0.5986	-400.33	4	0	800.69	2
8	1	.	1	1	.	2.01	0.5968	-400.64	4	0	801.32	3
9	**1**	**.**	**.**	**1**	**.**	**2.03**	**0.5906**	**-394.06**	**3**	**0**	**788.15**	**1**
10	1	tvar	.	1	.	1.71	0.7563	-408.65	4	1	817.34	4
11	1	.	tvar	1	.	1.67	0.7342	-409.74	4	1	819.54	5
12	1	.	.	1	tvar	0.74	0.9456	-414.68	5	1	829.42	6
13	tvar	.	.	1	.	0.06	0.9935	-666.81	3	1	1333.65	3
14	**1**	**.**	**.**	**tvar**	**.**	**1.24**	**0.9374**	**-379.40**	**3**	**1**	**758.84**	**1**
15	tvar	.	.	tvar	.	0.05	0.9970	-676.73	3	2	1353.5	2
16	tvar	tvar	tvar	tvar	tvar	0.03	1.000	-732.84	7	5	1465.80	-

Single step covariate removal with fixed effects was developed in models 1 to 6. Transient periodicity was the first covariate removed as its removal decreased the AIC score of the model (model 6). NAO with a 4 years lag and current NAO were removed individually (models 7 and 8) and jointly (model 9), based on the lower AIC obtained after their removal in models 3 and 4. Model 9 had the most support given the data, so the three covariates of this model were tested with a time-varying estimate (models 10 to 12) before being permanently omitted due to lower parsimony (increased AIC scores).

The covariates in the top AIC-ranked model (temperature and fundamental periodicity, model 9) were augmented by a time-varying estimate individually (models 13 and 14) and jointly (model 15). Model 14 produced the lowest AIC score. The model 16 was saturated in that it contains all available covariates, all augmented by a time-varying estimate.

Table 5.3: Candidate state-space models comparison for time-series of the log-transformed counts of *S. trutta*.
For each combination, the estimates are either (.) not included, (1) constant, or (tvar) time-varying. The number of state disturbance variances to be calculated (w) and the number of diffuse initial values (q) are the hyperparameters which intervened in the AIC score calculation.

Model	β (Temperature)	δ (NAO)	ζ (NAO, lag 4 yr)	a, b (1 year p.)	c, d (17.51wks p.)	Var (res)	R^2	Log-likelihood	q	w	AIC	AIC rank
1	1	1	1	1	1	4.71	0.8410	-383.42	7	0	766.91	3
2	.	1	1	1	1	4.26	0.8355	-1025.21	6	0	2050.48	6
3	**1**	**.**	**1**	**1**	**1**	**4.79**	**0.8404**	**-374.72**	**6**	**0**	**749.51**	**1**
4	1	1	.	1	1	4.79	0.8400	-375.26	6	0	750.58	2
5	1	1	1	.	1	2.31	0.5105	-631.83	5	0	1263.70	5
6	1	1	1	1	.	4.73	0.8163	-384.12	5	0	768.28	4
7	**1**	**.**	**.**	**1**	**1**	**4.88**	**0.8392**	**-366.54**	**5**	**0**	**733.12**	**1**
8	1	tvar	.	1	1	3.88	0.8162	-390.14	6	1	780.34	2
9	1	.	tvar	1	1	3.24	0.8163	-393.19	6	1	786.45	3
10	1	tvar	tvar	1	1	2.53	0.8163	-421.02	6	2	842.12	4
11	.	.	.	1	1	4.10	0.8438	-1008.96	4	0	2017.96	4
12	1	.	.	.	1	1.71	0.4916	-636.27	3	0	1272.58	3
13	1	.	.	1	.	4.81	0.8143	-367.39	3	0	734.81	2
14	tvar	.	.	1	1	0.10	0.9982	-678.17	5	1	1356.39	5
15	1	.	.	tvar	1	2.39	0.9773	-383.55	5	1	767.15	2
16	1	.	.	1	tvar	1.43	0.9813	-398.85	5	1	797.75	3
17	tvar	.	.	tvar	1	0.08	0.9995	-688.44	5	2	1376.94	6
18	tvar	.	.	1	tvar	0.07	0.9997	-694.69	5	2	1389.45	7
19	1	.	.	tvar	tvar	1.14	0.9912	-438.63	5	2	877.32	4
20	tvar	tvar	tvar	tvar	tvar	0.05	1.0000	-731.82	7	5	1463.76	-

Models 1 to 6 show single step covariate removal with fixed effects. NAO was the first covariate removed as its removal decreased the AIC score substantially (model 3). Then NAO with a 4 years lag was removed (model 7). Both NAO covariates were tested with a time-varying estimate individually (models 8 and 9) and jointly (model 10).

Single step covariate removal was pursued (after model 7 of lowest AIC score) but no covariate was removed without decrease the parsimony of the model (models 11 to 13). Covariates in model 7 were augmented by a time-varying estimate one by one (models 14 to 16) then by pairs (models 17 to 19); all

corresponding AIC scores were higher than for model 7. Model 20 was saturated (all available covariates augmented by a time-varying estimate).

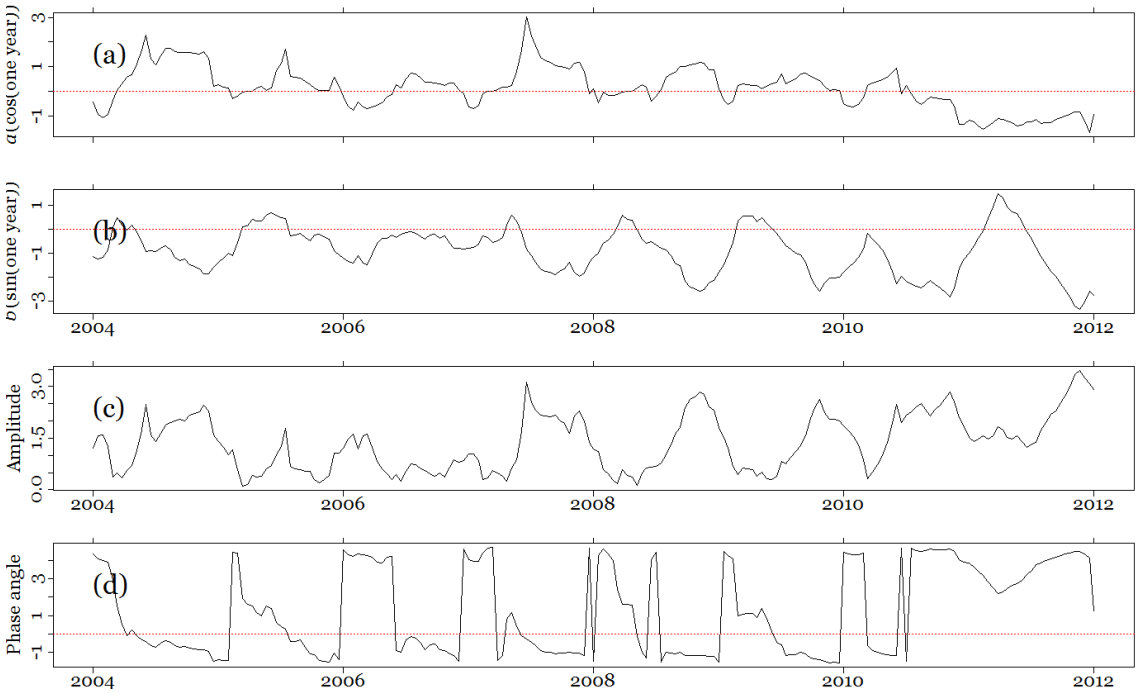


Figure 5.3: Temporal variation of time-varying estimates for the fundamental annual periodicity (most parsimonious model 14, Table 5.2).
Parameters *a* and *b* (a and b) are the coefficients of the annual dynamic harmonic regression (*a*cos(wt)* and *b*sin(wt)*). Amplitude was calculated from these coefficients ($\sqrt{a^2 + b^2}$) as well as the phase angle ($arctan\left(\frac{b}{a}\right)$, see Chapter 4 for details). The temperature estimate was constant over time (β=0.2574).

In the top AIC-ranked model for log-transformed counts of *S. salar* (model 14, Table 5.2), the coefficients *a* and *b* associated with the annual cycle changed dynamically (Figure 5.3a&b). Additional characteristics of the sinusoid were derived from these coefficients (as it was completed in Chapter 4, page 99); the amplitude and phase showed substantial temporal variation (Figure 5.3c&d).

5.3.2. Time-varying identification; learning curve and positive identification

The proportion of salmonids positively identified over all video records was described as a function of learning rate and environmental conditions. The proportion of identified salmonids was first modelled (GLM) as a function of environmental covariates (*i.e.* water visibility conditions and fish density). The most parsimonious model (Figure 5.4a and Table 5.4) indicated that identification was higher at night and lower salmonid density; the covariates

were important predictors of proportion of identified salmonids (solar irradiance, χ^2=29.27, df=1, *P*-value<0.001, counts in channel 4, χ^2=29.42, df=1, *P*-value<0.001).

The residuals of this model were in turn expressed as a function of the hypothetical species identification learning curve as defined by the Michaelis-Menten equation (Figure 5.4b and Table 5.4).

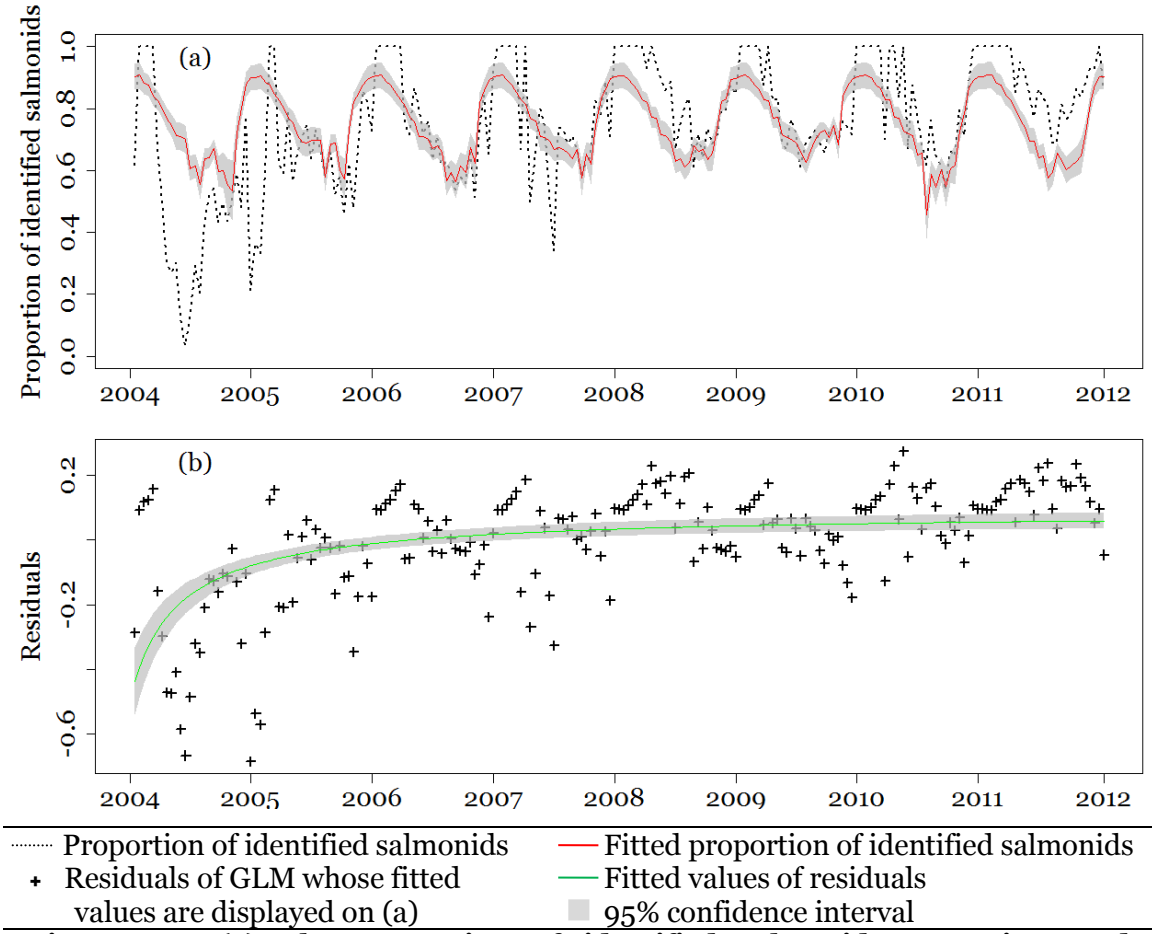


Figure 5.4: (a) The proportion of identified salmonids over time and corresponding fitted values based on fish density, visibility and the learning curve and (b) the resulting residuals.

The species identification learning rate curve was characterized by a decelerating rate. Proportion of identified salmonid became limited after overcoming the learning curve (Figure 5.4b).

Both GLMs explained a relatively low amount of variation (Table 5.4), indicating that the covariates included were likely important predictors of the proportion of salmonid identified but additional explanatory covariates were missing.

Table 5.4: Statistics of the GLMs of the proportion of identified salmonids and of the resulting the residuals.

GLM of proportion identified		GLM of residuals	
Predictor	Estimate (s.e.), *P*-value	Predictor	Estimate (s.e.), *P*-value
Intercept	$9.39e^{-01}$ ($2.59e^{-02}$), *P*<0.001	Intercept	-0.49 (0.06), *P*<0.001
Solar irradiance	$-1.13e^{-03}$ ($2.04e^{-04}$), *P*<0.001	Values fitted by M.-M. equation	0.66 (0.08), *P*<0.001
Counts in Channel 4	$-7.27e^{-05}$ ($1.30e^{-05}$), *P*<0.001		
Deviance explained	Adj.R^2=0.2928, F=87.11 on 1 and 207 df *P*<0.001	Deviance explained	Adj.R^2= 0.2506, F= 70.55 on 1 and 207 df *P*<0.001

The sum of the fitted values of both models provided a representation of the proportion of identified salmonids over time which was of better fit (Adj. R^2=0.4729, F=187.6 on 1 and 207 df, *P*-value<0.001, Figure 5.5). The residuals of this last model showed a heavy lower tail associated with an approximate fit to the first year of the data (Appendix A.iv, Figure A.7.21).

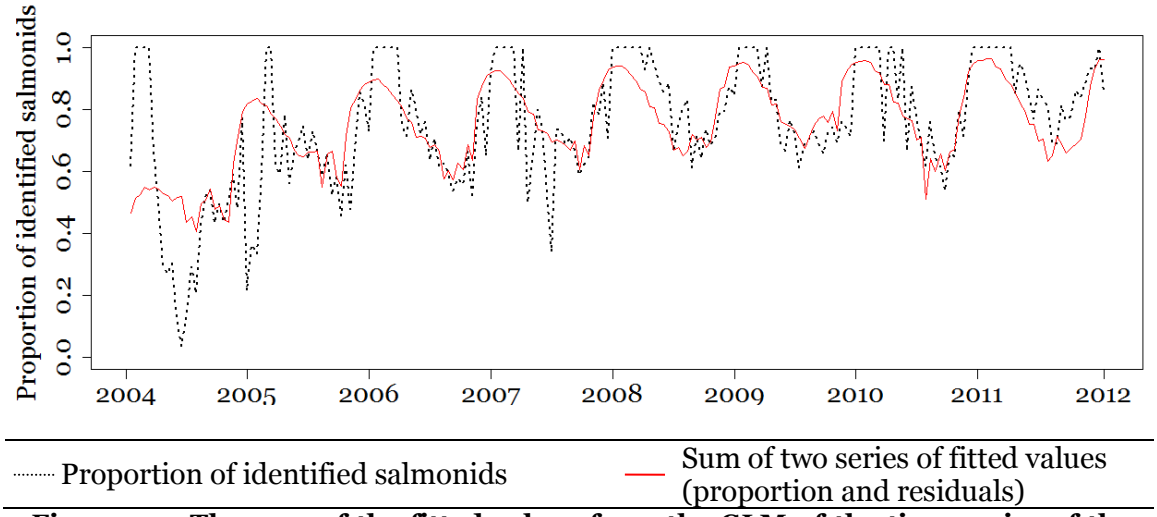


Figure 5.5: The sum of the fitted values from the GLM of the time-series of the proportion of identified salmonids, and the fitted values of the GLM of the resulting residuals.

The addition of the learning rate curve to the most parsimonious SSM and most parsimonious GLS models (in Chapter 4) did not improve either model (increased AIC score). Temporal variation in the hypothetical learning rate was significantly correlated with the variations in the time-varying estimates of all covariates, in both species (Table 5.5). The strongest correlations occured with the NAO values with a lag of 4 years and with the transient periodicity for each species.

Table 5.5: The Pearson correlation between the rate of the learning curve and the fluctuations in time-varying estimates.
The estimates were in the saturated models of the log-transformed counts of each species; all are on on 1 and 207 degrees of freedom.

	S. salar			*S. trutta*		
	Adj. R^2	F	*P*-value	Adj. R^2	F	*P*-value
β (Temperature)	0.0340	8.331	*P*<0.001	0.1358	33.69	*P*<0.001
δ (NAO)	0.3012	90.63	*P*<0.001	0.2082	55.71	*P*<0.001
ζ (NAO, 4 year lag)	0.4076	144.1	*P*<0.001	0.7229	543.8	*P*<0.001
a (P=fundamental)	0.0635	15.1	*P*<0.001	0.6147	332.9	*P*<0.001
b (P= fundamental)	0.0363	8.827	*P*<0.001	0.1389	34.54	*P*<0.001
c (P=transient)	0.5118	219.1	*P*<0.001	0.5365	241.8	*P*<0.001
d (P=transient)	0.4041	142	*P*<0.001	0.6774	437.9	*P*<0.001

The initial acceleration in the learning rate curve lasted until approximately 2007 after which the proportion of positively identified salmonids plateaued (Figure 5.4b). Goodness of fit for the GLS models developed in Chapter 4 was contrasted for the periods before and after 2007 (Table 5.6).

Table 5.6: Contrasting the goodness of fit of the GLS models (Chapter 4) according to temporal changes observed on the learning curve.

Observation period	*S. salar*	*S. trutta*
2004 - 2007	Adj. R^2= 0.6191, F= 126.2, on 1 and 76 df, *P*< $2.2 . 10^{-16}$	Adj. R^2=0.8390, F= 402.2, on 1 and 76 df, *P*< $2.2 . 10^{-16}$
2007 - 2011	Adj. R^2= 0.7053, F= 249.9, on 1 and 103 df, *P*< $2.2 . 10^{-16}$	Adj. R^2=0.8663, F= 706.5, on 1 and 103 df, *P*< $2.2 . 10^{-16}$
2004 - 2011	Adj. R^2= 0.6664, F= 364.6, on 1 and 181 df, *P*< $2.2 . 10^{-16}$	Adj. R^2=0.8177, F= 817.4, on 1 and 181 df, *P*< $2.2 . 10^{-16}$

In both species, goodness of fit was higher for 2007-2011 than for 2004-2007 and the whole observation period 2004-2011 (Table 5.6).

There were also differences in the residuals of the GLS models fitted before and after 2007. In *S. trutta* the autocorrelation was stronger and more cyclical

during the observation 2007-2011 (Appendix A.iv, Figure A.7.22f). The distribution of residuals was further from normal during the period 2007-2011 for the counts of both species and there was a heavier lower tail compared to the normal distribution in *S. trutta* (Appendix A.iv, Figure A.7.23 and Figure A.7.24).

5.4. Discussion

The analyses undertaken in the preceding chapters led to the hypothesis that the complex and dynamic behaviour exhibited by *S. salar* and *S. trutta* may the source of patterns in the fluctuations of salmonid abundance that are not stationary. This suggested the use of dynamic modelling that can account for temporal variation linked to the static measurements of covariates which vary in their distance to the riding Mill station, and are likely specific to *S. salar* and *S. trutta*. In this chapter, the use of state-space models to describe the counts of each species allows two major inferences with regards to time-dependency.

The dynamic harmonics of the state space models (cosine and sine coefficients), are not easily interpretable on their own (Chow *et al.*, 2009). Rather, they constituted an efficient tool to identify and describe the changes in amplitude and phase of a cycle initially included in the model of the the log-transformed counts of *S. salar*.

The relevance of the dynamic annual harmonic implied that the harmonic coefficients *a* and *b* appeared in the vector of state variables (since one state equation is formulated per time-varying component) whilst the cosine and sine functions were part of the measurement matrix (*i.e.* with a determined and fixed annual period). So, the dynamic harmonic regression was similar to a factor analysis model with non-linear constraints on the measured processes (*i.e* the fixed annual period). On one hand, the relevance of time-varying factors may imply that the cycle changed in amplitude and phase throughout time. There is an alternative type of inference possible (in both stochastic cyclic models and dynamic harmonic regression models, Chow *et al.*, 2009), that is reached by examining temporal changes in the cosine and sine weights of the cycle (and other resulting sinusoid characteristics). By definition, these cyclical

specifications require the cycle frequency to be time-invariant. Variation in the phase angle imply that the time at which the cycle is estimated to reach its peak varies. So, in cases where the frequency contained in the time-series actually changes throughout time, these changes become reflected by variation in the phase angle value (Chow *et al.*, 2009); variations in the phase may either reflect variation in the frequency of the time-series, or, it may imply that multiple frequencies contained in the time-series are not all described by the observed components of the harmonic regression.

The first inference from this study is that multiple time-dependencies exist within the recorded spawning population of each species. It is likely that each age-class represents a specific time-dependency and possibly a specific response to the covariates available in this study (and may also be described by a specific harmonic, Klemetsen *et al.,* 2003; McCormick *et al.*, 1998). The sampled spawning population on year *t* may be composed of kelts and grilses being born up to 5 years prior to the year of record (Figure 5.6). As a result, the youngest individuals counted at Riding Mill may be 2 years old (*i.e.* one year of smoltification and one year at sea, not including precocious parr) and the oldest may be over 7 (*i.e.* 4 years of smoltification and 3 years at sea). Therefore, spawning population may be composed of five age classes in unknown fluctuating proportions.

This source of heterogeneity in the population supports the use of time-dependent estimates. In *S. salar*, kelt and grilse are known for seeking different feeding grounds (Jonsson and Jonsson, 2011), so over winter conditions may be different for the two age classes. Time-dependent estimates, however, accounted for a single pattern and individual age-class dependent characteristics were not described. Variation between age-classes was likely too high to be modelled without explicitly distinguishing between the natural groupings contained in the samples at a given time. The models developed in Chapter 4 for the counts of each species, indicated that counts were partly a function of the previous counts, which implies an auto-regressive dependency which may be modelled as such (Chapter 4), but which may also fluctuate over time as a function of variation in age class composition throughout the year.

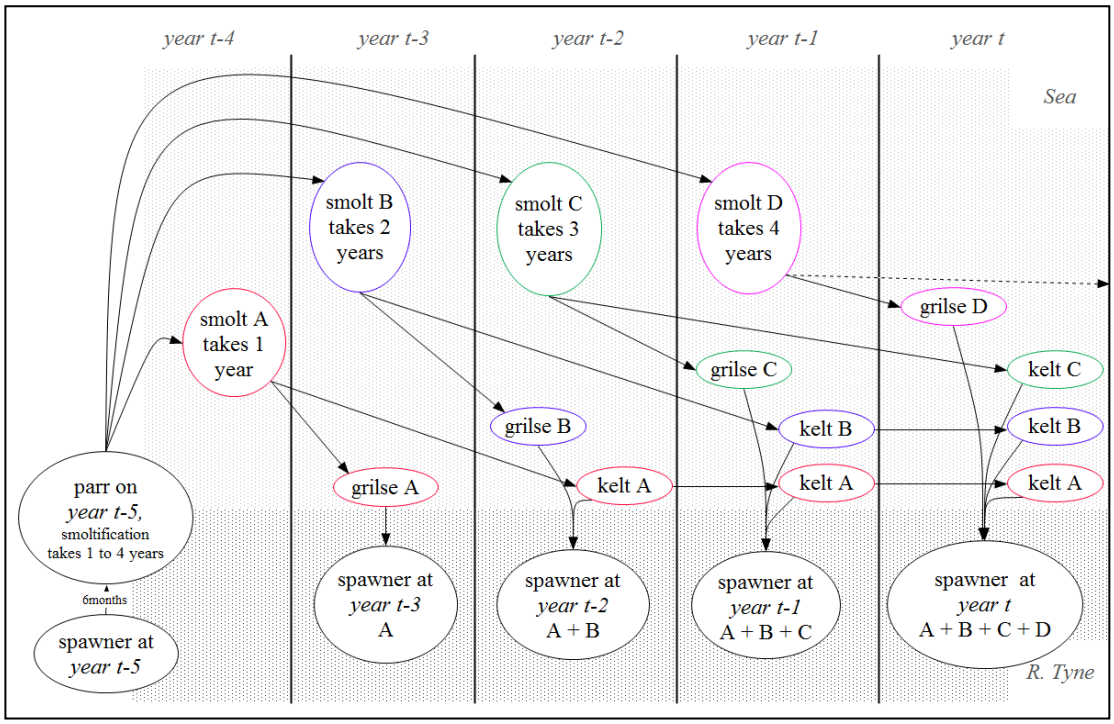


Figure 5.6: An illustration of age-class composition of the spawning salmonid population on year *t*.
This illustration is based on average durations for *S. salar*,illustrated for year *t*. Smoltification takes 1 to 4 years so that salmonids may develop into grilse or kelt throughout this period. Grilse spend one year at sea and kelt migrate back after multiple winters. On average, the spawning population is estimated to contain individuals born 1 to 5 years prior to being recorded in the River Tyne.

Capture-mark-recapture experiments have evaluated site-specific demographic features of localised populations of *S. salar (e.g.* recruitment success from egg to smolt depending on spawner age class, and between-stage transition rate; Rivot *et al.*, 2004; Dumas and Prouzet, 2003; Jonsson *et al.*, 1998). Rivot *et al.* (2004) used Bayesian state-space modelling to derive the dynamics between the unobserved stage-dependent population dynamics and the observation process using observed counts of *S. salar* and random sampling of demographic features. This study described selected traits of the life history of *S. salar*, but conclusions were limited due to simplification of the stage transitions (*i.e.* combinations of time and rate of transition from smolt to spawner stage), the absence of environmental data in the model, and the model's limited ability to account for the evolution of dynamics over time.

In contrast, the state-space models in the present study made use of environmental data and time-variation. It is suggested that the incorporation of species-specific demographic data would complete the model. Demographic processes appear to be essential to a comprehensive description of the fluctuations in salmonid population abundance.

The second inference from this study concerned the influence of increasing species identification learning rate over time. Ideally, the identification of each species would have been assessed (*i.e.* how identifying individual *S. salar* compares with identifying *S. trutta* in term of time required to achieving stable proficiency over time). The identifiability would for instance compare the number of identified *S. salar* against the true count of *S. salar*; data on true number of *S. salar* were not available for this study so the identifiability was not assessed.

Several methods were applied to incorporate the influence of identification within the state-space models (*e.g.* a given minimum proportion of identified salmonids, duration spent identifiying a given minimum proportion, both at several fixed amounts for a given proportion). However the processes involved in the biology of both species are complex and vary through time, up to 4 years according to the analysis in Chapter 3 (*i.e.* NAO values 4 years preceding the count). Data refinement to exclude subsets of the observation periods during which a minimum proportion of identified salmonids was not reached (*e.g.* years 2004-2005 displayed low proportion of identified salmonids) meant that the remaining time-series was likely not long enough to reflect the dynamics of the salmonid movements which took place over 4 years so this method was not retained (several proportions were tested under 50% of identified counts). Instead, state-space models indicated that when allowed to vary in time, the coefficients of all parameters showed a clear drift in 2008, with a transition period (2007-2010) that delimited two distinct trends each time, with a different trend before and after the transition period. This indicates that the count data were not accurate during the period 2004-2008. The time-varying parameter estimation likely attempted to fit a non-representative time-series (until 2008) followed by a representative time series (until 2012).

Identification skills are acquired with experience, a process described by a learning curve. The slope of a learning curve may influence the outcome of a process. For instance, surgical experience improves the outcome of many kinds of procedures; a given amount of time or number of events, specific to the procedure, leads to sufficient experience to overcome the learning curve (Atug *et al.*, 2006). When it forms a plateau indicating that the outcome from a given process is stable in time, the learning curve is overcome. The window of

experience required to overcome the learning curve is specific to the process and subject.

Here the proportion of salmonids identified from video data was assumed to be a function of proficiency at identification, an unmeasured variable, and the observation conditions, which are dependent on both water clarity and the density of salmonids (Table 5.7, *Eq. (5.6)*). First, the proportion of salmonids identified was modelled as a function of observation conditions (Table 5.7, *Eq. (5.7)*). Water clarity was represented by solar irradiance and fish density by the salmonid counts in Channel 4. The residual variation in this model (Table 5.7, e_1 in *Eq. (5.7)*) was assumed to be related to the learning curve of an observer. The learning rate was described by the Michaelis-Menten equation (Table 5.7, *Eq. (5.8)*) and the residuals were compared to the output of the equation (Table 5.7, *Eq. (5.9)*).

Table 5.7: Hypotheses considered when modelling the temporal fluctuations in the proportion of salmonids identified to the species level.
The constants *e, e1* and *e2* are error terms; *a* is the maximum proportion of salmonids that can be identified and *b* is the Michaelis constant; here it is the time at which the learning rate is half of the maximum.

Eq. (5.6) Proportion identified ~ Learning rate + Water clarity + Fish density + e
Eq. (5.7) Proportion identified ~ Water clarity + Fish density + e_1
Eq. (5.8) Learning rate ~ $\frac{a * \text{Time}}{b + \text{Time}}$ *Eq. (5.9)* e_1 ~ Learning rate+ e_2

The Michaelis-Menten equation estimated the term *a* as the maximum learning rate, defined as the maximum proportion of salmonids (*i.e.* considering an infinite observation period). The model assumed that learning rate increased asymptotically, until approaching a maximum rate proportion of salmonids identifiable, likely limited by conditions of visibility.

The relatively low goodness of fit of the total model indicated that additional covariates were required to describe the conditions of identifiability of salmonids. There was an impact of the learning on the goodness of fit of the model that became apparent when comparing the goodness of fit of the GLS models from Chapter 4 to subsets of the count data before and after the year 2008.

The learning curve reached a plateau at approximately 2007. After this initial phase, any effect of the salmonid species identification learning rate no longer influences the data. However, visibility conditions also appeared to be important predictors.

The proportion of identified salmonids was higher during hours of darkness, however the salmonids passing through channel 4 are visible at night due to lamps are used in low light conditions and the salmonids are considered easier to identify during the day than at night (EA, *per. com,* 2013).

When full occupation of the channel occurs, the proportion of salmonids identified cannot be maximal as salmonids hide each other and likely hinder the detection of classifying features. There was a limitation associated with the space available for salmonids and the reduced visibility resulting from higher salmonid density, reflected by the influence of the number of salmonids in the channel 4 of the fish counter.

Originally, the Michelis–Menten equation assumes single substrate enzyme kinetics (Johnson and Goody, 2011). When applying the equation to the learning rate at identification of species on video records, a single learning time-frame was assumed. However, during the first two years of the video camera being in place, there was a considerable turn-over of observers responsible for the species identification (EA, *per. com.,* 2013). In a model, this turn-over was compared to several time-frames of learning rate being in place during the period in question (*i.e.* one learning rate per observer). The Michaelis-Menten equation considered a single observer with a single learning rate, which helps explain the lower goodness of fit in the curve at the beginning of the observation period (2004-2005). After this initial multi-observer period, a single observer produced the video data and the learning rate curve showed a better fit.

The learning process may also in fact have contained two distinct processes: learning to identify *S. salar* and learning to identify *S. trutta*. It can be assumed that the proportion of unidentified fish varied with the ease with which either species was identified. Although the characteristic features of both species are thought to be equally discernible, particular features suggest that *S. trutta* may require less observation time to be identified (*e.g. S. trutta* has spots under the lateral line and *S. salar* does not; E.A., *per. com.*, 2013). This may mean that

identification time require for the two species occurs on two observation time scales, in which case, the learning rate of identification of *S. salar*, specifically, may have reduced the global learning rate of species identification in this study.

When variation in observed data are partly due learning, a firm uncertainty will exist, that cannot be accurately estimated and is particularly strong with regards to predictions (Nemet, 2006). The influence of learning rate here was supported by strong correlation with transient periodicities in the video count time-series that were not apparent when considering the salmonids as a whole.

5.5. Conclusions

The hypothesis that stochasticity was an intrinsic feature of the temporal fluctuations in the abundance of salmonids in the River Tyne was tested. State-space models were developed that aimed to describe time-variation in the response of *S. salar* and *S. trutta* to covariates in explaining the migration of both species, according to prior theoretical justifications (Chapters 2, 3 and 4).

Fluctuations in the abundance of *S. salar* were described using a dynamic harmonic regression that reflected changes in the amplitude and phase of a seasonal cycle over time. It also indicated that several natural groups likely exist within the counts associated with the different age-classes represented in the dynamic harmonic regression.

A second hypothesis was investigated that considered the existence of a species identification learning curve, *i.e.* an increase in the reliability of the species identification data over time. It was shown that the proportion of identified salmonids increased and became more stable over time according to the progression of a learning rate. Video count data were not sufficiently representative of the natural phenomenon occurring during the years 2004-2008. The learning process was overcome and the video recordings produced now display a stable temporal accuracy.

The age-class composition and the natural groupings contained in the populations of each species need to be more precisely described to achieve full understanding of the processes involved in the migration of those species. In the next chapter, an algorithm will be investigated, that aims to rely on signal size and environmental conditions to classify salmonids into *S. salar* or *S. trutta* and potentially into natural sub-groups within each species.

Chapter 6. .

”A practical botanist can usually at the first glance distinguish the plants of Africa, Asia, America, and the Alps; but it is not easy to tell how he is able to do this”.

C. Linnaeus (1707-1778)
Quoted in *Familiar lectures on botany*
M.L. Phelps, 1856.

Chapter 6. Discriminating between *S. salar* and *S. trutta* when recorded at Riding Mill station.

6.1. Introduction

Non-invasive techniques of species identification are a long-term goal in ecology; this is particularly true for marine organisms, challenging to monitor and identify at sea (Horne, 2000). Ideally, the processes of detection, identification, and enumeration of animals are undertaken without disrupting their natural passage and survival, thereby using techniques said to be "stealthy" (Parrish, 1999). Marine species are monitored for conservation and/or fishing purpose and absence of disturbance is especially necessary for species at risk (Eatherley *et al.*, 2005).

S. salar is of conservation concern (Bardonnet and Baglinière, 2000) and both *S. salar* and *S. trutta* have a high commercial value (Aprahamian *et al.*, 2010) with a need to be managed sustainably. Stock assessment for each species is central to their management and this can only be achieved through accurate monitoring studies. In comparison with active methods, remote census approaches imply reduced costs and may be more accurate (the bias relating to the sampling technique is null or smaller than for an active sampling), which are both important benefits to resource managers (Horne, 2000).

Resources required for fish monitoring are proportional to their behaviour and habitat size (Côté and Perrow, 2006). Sampling in the marine environment incurs high costs in equipment, trained staff and time. Many techniques are also destructive (*e.g.* netting, trawling, hook and lining; Côté and Perrow, 2006), with prohibitively detrimental effects on salmonids. This results in a limited knowledge of the trends and behaviour of salmonid populations when at sea (Friedland, 1998; Smith *et al.,* 1993).

However while in freshwater, salmonids may easily be caught, trapped, observed, or counted remotely (Côté and Perrow, 2006). The automatic fish-

counting device at Riding Mill weir, on the River Tyne, is an example of remote sampling. The counter reports abundance of *S. salar* and *S. trutta* combined.

The array of available remote, non-invasive sampling methods is growing. Acoustic sounders are an example of remote sensing techniques used at sea: sound is emitted and the responding echoes are used to detect aquatic organisms (Kracker, 1999). Soon after the debut this technique, fishermen joined their knowledge of marine species habitat characteristics to sounders echograms; combinations of habitat characteristics and echogram properties can now allow identifying numerous marine species remotely. The process developed species reference libraries, containing information discriminating between species based on their habitat characteristics and the type of signal they produce (Horne, 2000).

In rivers, resistivity counters monitor the difference in voltage across an array of electrodes generated by the passage of salmonids (Coyle and Reed, 2012; Forbes *et al.*, 2000), without known incidence. The amplitude of the signal varies with the size of the fish (Nicholson *et al.*, 1994) but the counter does not readily identify species.

In some instances, the amplitude of the signal may suffice to assign a fish to a predefined category. This is the case for distinguishing rainbow trout from steelhead trout, two morphs of *Oncorhynchus mykiss*. The former is a freshwater resident and does not grow as large as its migratory co-specific; the size difference is discriminating (McCubbin and Ignace, 1999). No analysis has investigated the relationship between signal amplitude, fish size, and salmonid species identification in the River Tyne.

The movement of salmonid across a four-channel resistivity counter generates a signal in the shape of a waveform. At Riding Mill, the counter has been supplemented by a video camera covering one of four channels (Figure 6.1a), enabling the distinction between *S. salar* or *S. trutta* on this channel (Figure 6.1b). In these instances, signal amplitude is linked to the species, time of observation, and river covariates. There may be differences between species that discriminate between them when counted at Riding Mill.

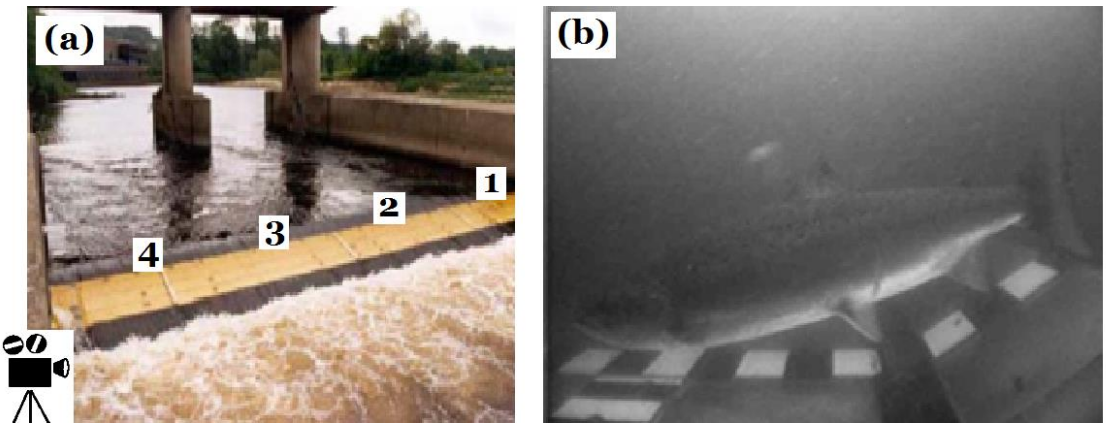


Figure 6.1: (a) The four channels of the resistivity counter at Riding Mill and (b) an example of video recording produced by the camera at Riding Mill.

Scientific literature suggests several parameters as classifiers of salmonids to the species level. For instance, time of year may be discriminatory as sympatric salmonid populations often present slight variation in migration timing (*S. salar* commonly ascend natal rivers before *S. trutta,* Jensen *et al*., 2012).

Flow regime is a recurrent factor associated with the upstream migration rate in anadromous salmonids (Lilja and Romakkaniemi, 2002). High flow regimes may involve turbulence, which increases the energy cost of swimming considerably (by up to 1.5, Enders *et al*., 2003). *S. salar* individuals cope with high velocities and turbulence better than *S. trutta* by "sit-and-wait" behind a "home rock" for up to 80% of the time (Haro *et al*., 2004; Enders *et al*., 2003), and rheotaxis positioning (Heggenes *et al*., 1996); high velocities may select against *S. trutta*.

Thermal limits in salmonids are also species-specific. *S. salar* has a higher upper critical range (22-33°C) than *S. trutta* (20-30°C) and a slightly higher lower critical range (0-7°C versus 0-4°C, Jonsson and Jonsson, 2009); *S. salar* may be more tolerant to higher temperatures.

Various stages of the tidal cycle may also differ amongst populations of *S. salar* (*e.g*. preference for the Ebb tide in the River Dee, Aberdeenshire, Smith and Smith, 1997; Potter, 1988) and between species.

Parameters of conductivity, pH, and turbidity relate to the concentration of dissolved particles in suspension in water. The conditioning of water is made via urine, pheromones, and other secretions produced by each species; it can be assumed that these factors alter water chemistry and potentially impact on the detection of olfactory cues (Jaensson *et al*., 2007). These cues allow recognition

of a natal stream during return migration and the identification of conspecifics (Moore *et al.*, 1994; Brown and Brown, 1992). Both species follow a strategy of inclusive fitness; individuals benefit from including kin in the optimisation of fitness, and fish that fail to detect homing cues may become stranded. Homing cues are species-specific so parameters interacting with their detection may be discriminating. In particular, olfactory cues are affected by pH level. Response to olfactory cues is altered in acidic water in *S. salar* (pH=5.1, Royce-Malmgren and Watson, 1987) and overall, *S. salar* is more sensitive to low pH values than *S. trutta* (disturbance at pH=6.3 versus pH=5.5 in trout; Jonsson and Jonsson, 2011). This suggests that acidic conditions may select against *S. salar*.

The social structure of *S. salar* relies on visual cues, where body and eye colour distinguish dominant individuals (O'Connor *et al.*, 1999). Strong social hierarchy in *S. salar* limits aggressive interactions to rewarding instances (Suter and Huntingford, 2002; O'Connor *et al.*, 1999). In *S. trutta*, dominants find themselves fighting continuously in order to preserve their position in the global hierarchy (Petersson and Jarvi, 1997) but when kins are present, *S. trutta* spend less time being alert for potential competition and more time feeding (Hojesjo *et al.*, 1998). Low visibility may yet be disruptive as they are visual feeders (Klemetsen *et al.*, 2003).

The two salmonid species display slight differences in habitat requirements and preferences. This study aims to formulate an algorithm discriminating between *S. salar* and *S. trutta* in the River Tyne, on the basis of signal amplitude, time of passage, and habitat preferences for each species. Parameters used to distinguish between species will be tested based on observations in other salmonid populations.

In a second time, the study will be concerned with the river-specific population structure of each species. In several UK rivers (such as Rivers Dee and South Esk), grilse exceed kelts by up 7 times (JNCC, 2007); grilse are younger and smaller than kelts due to the shorter time they spend at sea. Breeding success is directly related to body size (Fleming, 1998), the age ratio for these groups is therefore likely to influence the fitness of the population. Estimating the typical proportion of smaller individuals for the River Tyne would be useful for local management of each species as the information can be integrated in fishing

regulations. A classification algorithm to discriminate between categories within each species (grilse and kelts or other natural groups) will be investigated.

Grilse and kelts tend to be structured temporally so that timing of migration may discriminate between them. However the order of appearance varies among rivers (*e.g*. kelt first in Norway, grilse first in the River Dee, Scotland; Jonsson *et al*., 1990). Fast growers may migrate younger and smaller, than slow growers (Jonsson and Jonsson, 2011). In *S. trutta*, premature returns (only a few weeks after their seaward migration) were observed mostly in smaller fish (In Ireland, Birkeland, 1996). In *S. salar* the size and sex may also affect the temporal structure of the migration population (Norway, Jonsson *et al*., 1990).

Body size may also be a discriminating factor within species counts. In anadromous fish, proficiency in osmotic regulation increases with size and age (Jonsson *et al.,* 1990). Lower temperatures render osmotic regulation more energy-demanding, so that winters may be more challenging to the least resistant, smallest, and youngest fish. The duration spent at sea in spring may be limited as a coping strategy and imply early returns of small fish. In addition, the relationship between body weight and swimming cost (Tang *et al*., 2000) suggests that small fish may be more likely to avoid high flow and take advantage of flood tide than large fish.

Covariates interfering with the organization of the social structure will also be investigated, under the hypothesis that compromised conditions may select against the least dominant fish.

Two algorithms will be investigated, that each assigns each count of *S. salar* or *S. trutta* to a predefined category specific to each species based upon time of passage and river conditions. Groups will be defined based upon the signal amplitude.

Multiple parameters used to describe the habitats suggests the use of a multivariate classification technique. Classification analyses are suited to detect and qualify the factors determining a population structure containing predefined groups (James, 1985) and have been used in the past in ecology (*e.g*. spider communities, Rushton and Eyre, 1992).

Here, a linear discriminant analysis (LDA) will be used to classify predictively. The LDA will assess the extent to which it is possible to identify salmonid

species based on environmental parameters, time, and signal amplitude. Secondly, potential groups within each species (grilse and kelts) will be investigated; again, LDA will be used to assess whether it is possible to classify observations of a given species into one of these groups, based upon the environmental parameters and time.

6.2. Material and methods

6.2.1. Signal and environmental parameters

The passage of salmonids across the Riding Mill electrode array generated sinusoidal signals (Figure 6.2). These waveforms are made of a peak and trough corresponding to a difference in resistance between the first and second electrodes and the third and second electrodes (description in Chapter 1, page 13). Here, signal data are concerned with the amplitude of these waveforms.

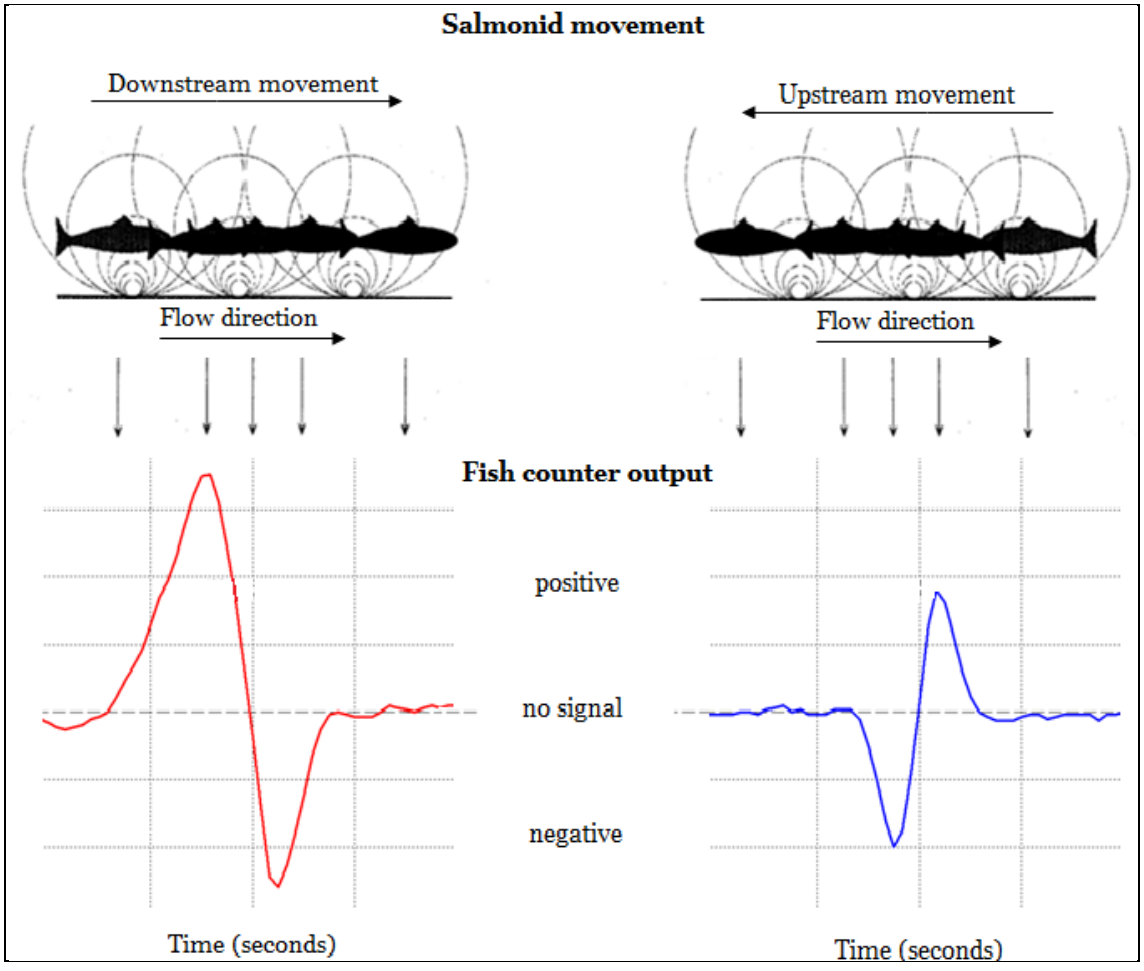


Figure 6.2: The passage of salmonids over anarray of 3 electrodes and the associated waveforms showing change in resistance.
Fish counter illustration: Fewings, 1994, reproduced with permission of the Atlantic Salmon Trust (Andrews, 2013); signal data: Environment Agency (February 2013).

Signal data were instant recordings of the passage of salmonids. Environmental data, originally recorded at a fixed time intervals, were merged and aligned to correspond with the signal amplitude data. The resulting dataset contained synchronized information on the conditions of flow, temperature, tide, conductivity, pH, and turbidity, as well as signal amplitude and species (*S. salar*

or *S. trutta*) for each count. In addition, time indices were included in the form of both linear (continuous count of time) and cyclical (trigonometric expression of time) predictors for day, week, and month in relation to the year. Data were available for years 2004 to 2010 and included a total of 21490 count records of fish movement (Appendix A.v, Figure A.7.25).

The stability in the efficiency of species identification was assessed by studying the progression of the proportion of unidentified fish over the years. Loess regressions were fitted for the time-series of unidentified fish and the time-series of total counts per video records. Loess smoothing allowed investigating the global trends of the time-series over the whole observation period (Cleveland *et al.*, 1990).

As for Chapter 4, the polynomial degree was 1 so that the fitting was locally linear, and the value of span was selected so as to keep it as low as possible, identical for both time-series and with reduced confidence intervals (Cleveland *et al.*, 1990).

6.2.2. Discriminant analysis

Linear discriminant analysis is a special case of the Fisher's linear discriminant with added assumptions of homogeneity of variance and normality (Rao, 1948). Using the two species as groups and the environmental parameters as group discriminators, linear stepwise discriminant analysis (James, 1985) was used to determine the extent to which the species could be separated in terms of measured river parameters and signal amplitude. It was assumed that parameters predicting the most accurate separation of groups was also the most likely to be of greatest significance in determining the species of salmonid counted.

Predictors were ranked and selected according to their local CAT scores (correlation-adjusted t-scores) which quantify the relationship between each parameter centroid and pooled mean (Zuber and Strimmer, 2009; Strimmer, 2008). Predictors were trained on a representative subset of the data and tested on the remaining data. A comparison of observations with predicted class assessed prediction success. The training dataset contained data for years 2004, 2006, 2008 and 2010; the test data contained the years in between. Alternation

of years circumvented a potential source of bias linked to the improvement of visual identification of the species through learning.

River covariates were used in the model based on hypothetical discriminating properties as presented in Table 6.1. The use of signal amplitude assumed a size difference between the species, detected via the relationship between fish size and signal amplitude (Nicholson *et al.*, 1994). For the covariate time, various expressions and scale of time were used (hour of the day, and day, week, month of the year: each linearly and as a trigonometric expression).

Table 6.1: The parameters included in the LDA for the classification of salmonid observations as *S. salar* or *S. trutta*.
With associated hypothetical discriminative property and reference in literature.

Covariate	Hypothesis	In literature
Signal amplitude	*S. salar* tends to be larger than *S. trutta*	Maitland, 1965
Time	Timing of migration differs between species	Jensen *et al.*, 2012
Flow	*S. salar* is more tolerant to higher velocities than *S. trutta*	Armstrong *et al.*, 2003
Temperature	Thermal limits are species-specific	Jonsson and Jonsson, 2009
Turbidity, Conductivity, pH	Suspension particles interfere with detection of olfactory cues differently in each species, limit visual communication in *S. salar* and limit visual feeding in *S. trutta*	O'Connor *et al.*, 1999, Klemetsen *et al.*, 2003
pH	*S. salar* is more sensitive to acidic water than *S. trutta*	Jonsson and Jonsson, 2011
Tide	There is a preference for a given stage of the tidal cycle for one or both species	Smith and Smith, 1997

Potential natural groupings within counts of *S. salar* and *S. trutta* were assessed with a cluster analysis of the signal amplitude data for each species (multimodal distribution with finite normal mixture models and selection by BIC value, after Fraley and Raftery, 2002). The hypothetical discriminatory properties of each parameter for this second classification are presented in Table 6.2.

Table 6.2: The parameters included in the LDA for the classification of *S. salar* or *S. trutta* into subgroups of each species.
With associated hypothetical discriminative property and reference in literature.

Covariate	Hypothesis	In literature
Time of year	Migration timing is age-class specific	Jonsson *et al.*, 1990
Flow	Swimming cost increase with flow, small fish do not perform as well as large fish within high water velocity	Tang *et al.,* 2000
Temperature	Larger fish tolerate colder temperatures	Jonsson *et al.,*

	than small fish due to their better osmoregulation ability	1990 Parry, 1958
Conductivity, Turbidity, pH	Suspension particles interfere with detection of visual and olfactory cues differently in dominant and others	O'Connor *et al.*, 1999 Jaensson *et al.*, 2007
Tide	Swimming cost increase with flow, small fish may take advantage of flood tide	Tang *et al.*, 2000

6.2.3. Software

The LDA, the calculation of false discovery rates and the cluster analysis were generated, respectively, in the sda (Ahdesmaki *et al.*, 2012), fdrtool (Klaus and Strimmer, 2012) and mclust (Fraley *et al.*, 2012) packages operated in the R.15.1 environment (R Core Team, 2012).

6.3. Results

6.3.1. Discriminating between the two species

After excluding data with missing explanatory parameters, there remained 13565 training and 7925 testing observations. The top parameters that differentiated between *S. salar* and *S. trutta* were volumetric flow rate at Bywell, signal amplitude, and week of the year in both the linear and trigonometric forms (Appendix A.v, Table A.17).

The final classification algorithm contained only the trigonometric function of the week as it lead to higher prediction success than the linear function. Classifiers of flow rate at Bywell, signal amplitude, and week of the year, generated 64% of correct classification predictions of species on the test data (2836 errors out of 7925 predictions).

Values of signal amplitude and flow rate are presented on Figure 6.3 for each species and the trigonometric expression of week is illustrated on Figure 6.4. The maximum probability for a count to be assigned to *S. salar* was reached on the 7th week and the minimum on the 33th week of the year.

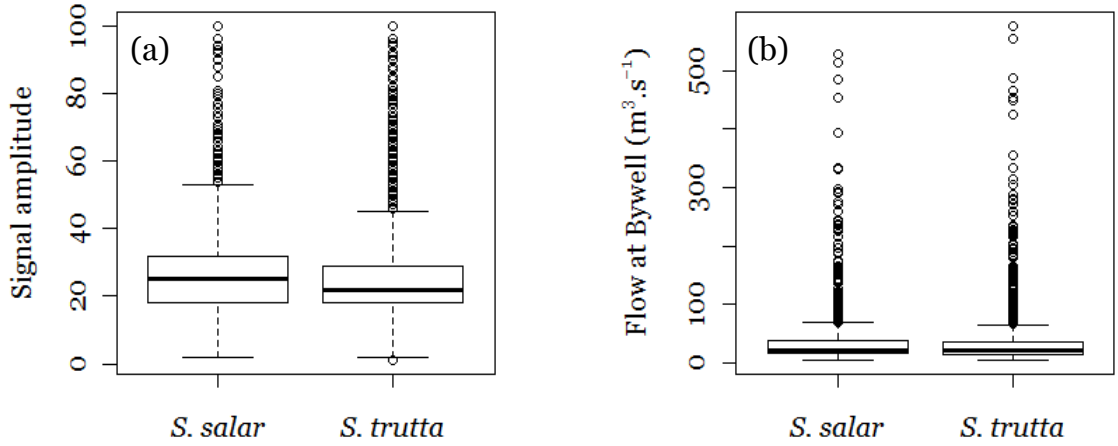


Figure 6.3: Distributions of (a) signal amplitude and (b) flow rate experienced at Bywell, for each species, in the River Tyne.
For *S. salar* the mean signal amplitude was of 26.59(±10.64) which was higher than for *S. trutta* (24.58 ±9.75). Mean flow rate values associated with *S. salar* were of 30.01(±26.81), also higher than for *S. trutta* (27.63±24.84). High standard deviations underline the important overlap in the characteristics of the two species.

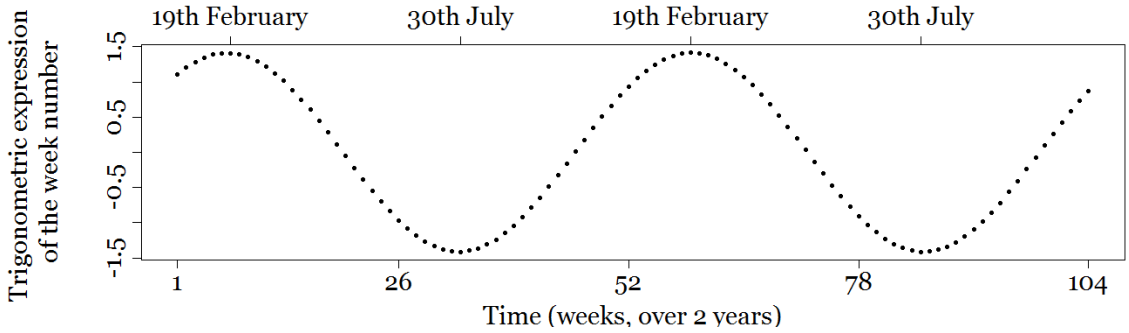


Figure 6.4: An illustration of the progression of the trigonometric expression of the week of the year.
The waveform is the result of the equation

$$T=\cos(\text{week number}*2*\pi/52.5)+\sin(\text{week number}*2*\pi/52.5)$$

Illustration over two years; the maximum value is reached on the third week of February, the lowest on the first week of August.

The use of flow, signal size, and week of the year as classifiers in the algorithm generated true positives only (Appendix A.v, Figure A.7.26) and individual CAT scores indicated that conditions of higher flow rate combined with greater signal amplitude, were indicative of a higher probability of a count being *S. salar*. Conversely, lower flow rates, smaller signal amplitudes, and proximity to the end of July, indicated a higher probability of the fish being *S. trutta*.

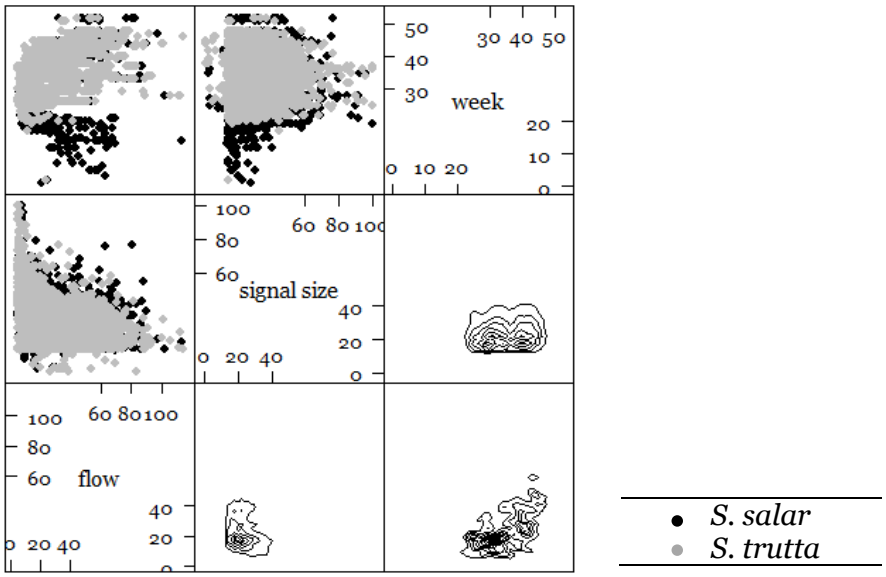


Figure 6.5: Scatter-plot matrix showing the pairwise relationships between the three classifiers of the two species and the bivariate density estimates.
Data points are in black for *S. salar* and in grey for *S. trutta*. Pairwise relationships between classifiers appear in the upper panel, and the bivariate density estimates in the lower panel. On the upper panel, data points for *S. trutta* do not cover any trend in points for *S. salar*. On the lower panel, the absence of two distinct modes in the representation of the bivariate densities reflects the important overlap of the values for the two species.

6.3.2. *Discriminating within counts of each species*

The results of the cluster analysis indicated that observations of each species may be divided into seven groups according to their signal amplitude (Appendix A.v, Figure A.7.27). The suggested multimodal distribution is presented on Figure 6.6 with the corresponding range of the signal amplitude per cluster on Figure 6.7.

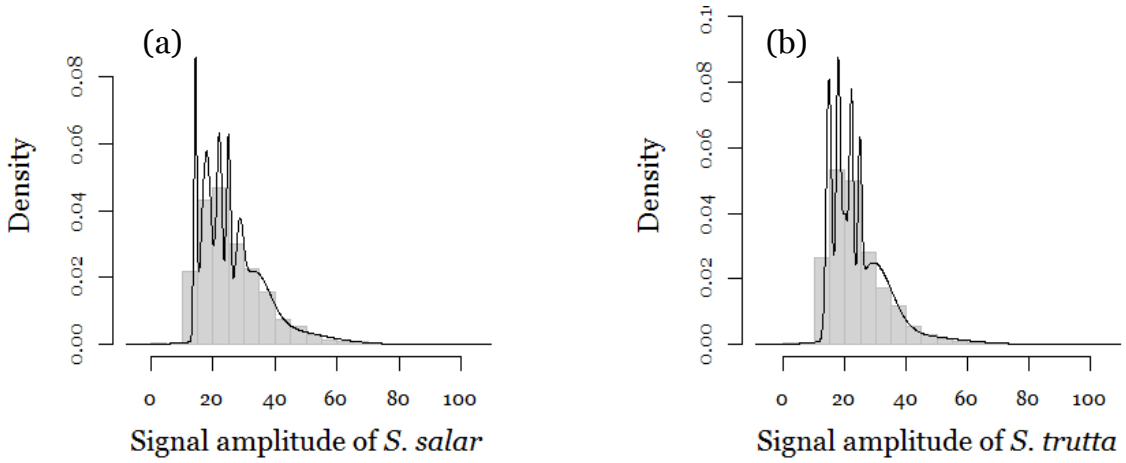


Figure 6.6: The frequency density histograms of the signal amplitude data generated by the passage of (a) *S. salar* and (b) *S. trutta*.
Original data feature as a histogram and the multimodal distributions suggested by the cluster analysis are represented by a plain line.

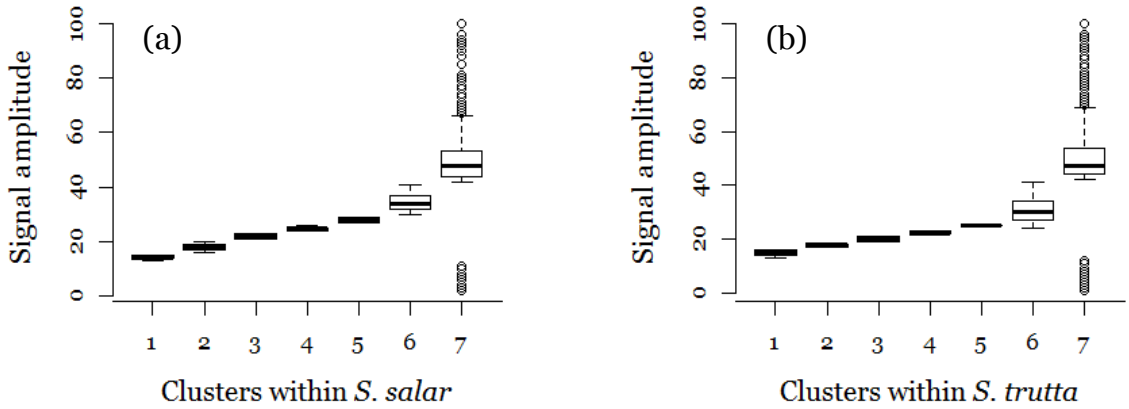


Figure 6.7: The range of signal amplitude contained in each cluster suggested by the cluster analysis for (a) *S. salar* and (b) *S. trutta*.

In *S. salar,* of all the parameters considered, hour of the day was the single discriminating classifier to generate true positives (Appendix A.v, Figure A.7.28a and Table A.18). The maximum classification success rate was only 27.02% (613 errors out of 840 predictions, Appendix A.v, Figure A.7.28b and Table A.19).

In *S. trutta*, no river parameter significantly discriminated between the predefined groups (Appendix A.v, Figure A.7.28b). The top classifier generated only 27.77% classification success (1302 errors out of 1800 predictions).

6.4. Discussion

It was determined in Chapter 5 that the efficiency of the visual species identification on the video records by the data collectors appears to have improved throughout time. More particularly, the two first years of observation likely suffered from low efficiency that may have rendered the records unsuitable for interpretation. Records provided by the fish counter and covering all 4 channels were considered likely to be more representative of fluctuations in both populations than video records.

In this study, several hypotheses were formulated based on observations found in the literature, such as:

- time of year and time of day (Jensen *et al.*, 2012),
- flow velocity (Armstrong, 2003),
- river temperature (Jonsson and Jonsson, 2009),
- conductivity, pH, and turbidity (Jonsson and Jonsson, 2011; Jaensson *et al.*, 2007; Klemetsen *et al.*, 2003; Suter and Huntingford, 2002; Royce-Malmgren and Watson, 1987; O'Connor *et al.*, 1999),
- and tide (Smith and Smith, 1997; Potter, 1988).

The final classification algorithm used a combination of three river parameters and generated a classification success of approximately 64%. Values of flow at Bywell were the most discriminating river feature, supporting the ability of *S. salar* to cope with strong flow better than *S. trutta* (Armstrong *et al.*, 2003).

The second most discriminating feature was the signal amplitude generated by the passage of fish across the counter, supporting a positive relationship between signal amplitude and fish size (Nicholson *et al.*, 1994), with *S. salar* being larger than *S. trutta*.

Third, time of year played a role in predicting which of the two species a new count was more likely to belong to. The probability of a count belonging to *S.*

salar was highest on the 7th week of each year and it was lowest on the 33rd week (when *S. trutta* becomes the most likely species). The selection of a trigonometric expression of time indicated a progressive evolution of the probabilities from these dates before and after these weeks.

For the second part of the study, cluster analyses were used to detect potential subgroups within each species based on signal amplitudes. The relevance of a clustering based on fish size relied on the following observations. Kelts and grilse may migrate at distinct times of year in the River Tyne (Jonsson *et al.*, 1990). A preference of small fish for lower flows than larger fish may exist given the link between body weight and swimming effort (Tang *et al.*, 2000). Similarly, small fish may benefit from moving during flood tide. Osmoregulation is more difficult for small fish (Jonsson *et al.,* 1990) and becomes more demanding at lower temperatures (Parry, 1958), therefore small fish may migrate earlier in order to cope with cold winter conditions. Covariates interfering with the conditions favourable to the organization social structure (*i.e.* with olfactory cues, Jaensson *et al.*, 2007, and visibility, Suter and Huntingford, 2002; O'Connor *et al.*, 1999) may result in the selection of more dominant individuals as their individual fitness is likely less dependent on social structure.

In previous studies showing age and size to be discriminatory of migration timing, the salmonid population included two sub-groups with distinct age and/or size during the return migration (Jonsson *et al.*, 1990). Here, and for both species, the analysis selected an optimal separation into seven clusters.

The records of counts in Channel 4 may not represent the whole populations nor the characteristics across the whole weir and salmonids passing through Channel 4 may be a non-representative sample of the populations. For instance, there may be smaller fish of each species that prefer edges in order to avoid high water velocity (Tang *et al.,* 2000); large fish may have used the central channels and not seek shelter from high velocity. Also, although the characteristic features of both species are thought to be equally discernible, the observer reviewing the video records had more opportunities to discern classifying features of salmonids located closer to the video camera. Smaller fish are thought to prefer the edges, in particular *S. trutta* (EA, *per. com.*, 2013), so it is possible that a higher proportion of *S. trutta* was identified over *S. salar*

amongst the salmonids present in Channel 4. This potential bias would affect sample representativeness and may have compromised the possibility for an effective classification algorithm.

Assumptions based on equal potential selection of dominant individuals were likely also compromised. Dominant fish are not necessarily the largest ones, but rather those with a faster metabolism (Metcalfe *et al.*, 1992). A fast metabolism implies a relatively high growth potential but more importantly an earlier migration age, and vivacity, and aptitude for competition (Fleming, 1996). Dominants may hence migrate upstream earlier than the more submissive fish, but body size may not be a good hypothesis.

Visual verification of signals minimized false positives by elimination of signals corresponding to, for instance, groups of fish, unsuccessful attempts to cross the electrode array, wave action due to wind, and non-fish objects that engender block-like signals (EA, *per. com.*, 2013). As a positive relationship has been found between counter efficiency and fish size (Nicholson *et al.*, 1994), the main source of bias among signal amplitude data was non-detection of fish and most likely non-detection of smaller fish.

In the River Tyne, signals are of reduced amplitude in comparison to other sites due to weaker detection conditions from a combination of water depth over the electrode array and low water conductivity (EA, *per. com.*, 2013). As a consequence, smaller fish may not be counted as regularly as larger ones. A threshold appeared around a signal of 26 in amplitude. Signals with amplitude lower than this value were rare and considered outliers by the cluster analysis, which grouped them with the highest signal values (seventh cluster). When clustering according to the signal amplitude (and, indirectly, fish size), some essential information related to smaller fish was possibly lacking. The threshold (Appendix A.v, Figure A.7.25) questions the efficiency of the counter at detecting smaller fish, which in turn compromises the analysis as this bias may have impacted differently on the classification between two species and between the groups within species.

The failure of the classification algorithm that divided observations of either *S. salar* or *S. trutta* into groups suggests that: first, signal amplitude data may be too biased for the study of partitioning within each species and second, the

underlying process responsible for the temporal structure in the populations may be dependent upon environmental parameters not included in the analysis. Information linked to climatic cycles were however contained in the time indices as all were ordered throughout the year.

The monitoring of parameters known to affect salmonid species differently may improve the performance of the LDA. The installation of a camera covering the whole width of the passage may also provide with an additional classifier based on the potential preference for a given channel.

The difference between small and larger fish may be improved by adding electrodes spaced more narrowly or using two resistivity counters operating side by side. A suggested setting for the two salmonid species of interest here, is an electrode array made of 5 electrodes instead of 3; separated by 20cm instead of 45 cm (*e.g*. such the one at Morar, Mucomir, Eatherley *et al*., 2005).

A confirmation of species-specific habitat preferences of salmonids in the River Tyne might further improve the precision of the conditional probabilities used in the classification including the channel specificity and depth of occurrence. As channel of occurrence was always reported, the classifier could potentially be applied to past data and improve the knowledge of a dataset already collected.

Ultimately, sensory systems used for the identification of species must rely on reference libraries containing recognition features; these libraries must be created and curated by scientists (Parrish, 1999). By testing the discriminating ability of the available parameters in the River Tyne, this study supplies information for the reference library of the recognition of salmonid species *S. salar* and *S. trutta*.

6.5. Conclusions

A classification algorithm using flow, signal amplitude and time of year allowed modest discrimination between the two species (approximately 64% prediction success). The two species displayed comparable preferences and characteristics in the covariates measured.

The assignment of counts *S. salar* and *S. trutta* to natural sub-groups was however not possible using linear discriminant analysis.

A better description of local habitat preferences of the River Tyne salmonids would allow a more accurate discrimination between *S. salar* and *S. trutta* and may lead to a classification into natural sub-groups within each species. In turn, age-class information may be used to feed the models developed in the previous chapters of this thesis to help achieve a more complete description of the salmonid migration and the factors influencing it.

Chapter 7. .

"Stories are webs, interconnected strand to strand, and you follow each story to the center, because the center is the end. Each person is a strand of the story."

N. Gaiman, 2005.
Anansi Boys

"There is no *one* thing that's true. It's all true".

E. Hemingway, 1940.
For Whom the Bell Tolls

Chapter 7. Conclusions.

Salmonid fish are valued as iconic target species around the world and represented by Atlantic salmon (*Salmo salar*) and the brown trout (*Salmo trutta*) in the UK. These fish are anadromous in that they breed in freshwater and yet have a marine stage to their lives. Most of their growth, development, and natural mortality take place in the North Atlantic, with individuals returning to their natal river annually to spawn (O'Connell, 2003; Gross, 1991). Both species demand high water quality and specific habitat (Heggenes *et al.*, 2002). In the past, pollution and intensive use of the UK river system have led to substantial declines, with few rivers supporting both species. Salmonids are recovering following improvements of the riparian habitats across the UK; however, fine-scale assessments are needed to quantify trends in abundance and their extent (Eatherley *et al.*, 2005). The River Tyne currently supports both species. The present research aimed to improve understanding of salmonid migration patterns in the River Tyne by focusing on the analysis of trends in abundance as salmonids return to reproduce.

On the River Tyne, salmonids are observed individually and in groups. The first study in this thesis (Chapter 2) attempted to describe a day in the life of individual salmonids on the last length of their return journey to spawn in the River Tyne. At the individual level, salmonid behaviour was complex, changing response to environmental parameters. These salmonids were most abundant between July and November, with seasonal characteristics of environment and abundance appeared to be involved in their abundance and aggregation behaviour.

Early in the year, salmonids moved upstream individually but became more aggregated as the spawning season progressed. It was hypothesised that increased abundance likely became a parameter itself in that it influenced the response to extrinsic forces occurring in the river. Visibility from predators can be less relevant for individuals when surrounded by many conspecifics (Aukema and Raffa, 2004; Kozak and Boughman, 2012; McClure and Despland, 2011), which suggested that the response to visibility parameters weakened as

abundance increased; also because of the reduced volume of water available, high densities may have forced the fish to expose themselves. The proximity to the limits of the thermal habitat naturally became reduced over the summer, which led to a reduced stimulation of salmonid movement by increases in temperature.

In this study, models were limited by the small scale required to the investigation of individual fish movements. Cyclical variables (temperature, flow, daylight, pH) became confounded because of the short sampling duration. The high amount of null values was partly due to the short time interval between observations (every 15minutes). By comparing the effect of parameters included in the model of each observation period, short-term relationships between individual salmonid movements and river conditions were contrasted throughout the year.

Chapter 3 was concerned with longer-term drivers of annual migration. Every year in the River Tyne, salmonids start arriving in spring according to a seasonal annual periodicity. *S. salar* and *S. trutta* were considered jointly with the features and drivers common to the migration of both species modelled together. Variation in river temperature had a major influence on salmonid movements, suggesting that habitat suitability was defined by thermal criteria over and above other available measurements. Oceanic conditions were partly represented by North Atlantic Oscillation values and NAO conditions four years prior to the salmonids return were essential in predicting magnitude of fluctuations in abundance.

The impact of the construction of the River Tyne tunnel was considered minimal; the model used to reach this conclusion was a direct demonstration that these count data may be used as a tool to monitor River Tyne salmonids.

Salmonids were monitored in the River Tyne as they return to spawn, *i.e.* during a point of the migration journey that was momentous but momentary. The monitoring at Riding Mill provided data which contributed to the understanding of the ecology of salmonids as a population. However, the co-existence of two species within the population advocated a cautionary approach with regards to the model developed as intrinsic differences between the two species implies that some level of prediction error is likely to occur if the

environmental conditions happened to become different from the ones measured during the sampling period.

This issue was addressed in the study in Chapter 4, in which salmonids were considered as two distinct populations of *S. salar* and *S. trutta*. Annual periodicity remained a fundamental feature of the fluctuations observed in the abundance of both species. Seasonal cycles happened to be species-specific, suggesting changes in the synchronicity and sinusoidal shape of counts for both species. The NAO was relevant in a quasi-instantaneous manner (*i.e* relevant without lag but considering time bins of 14 days). The different migratory routes followed by the two species concurred with a species-specific magnitude of the influence of the Atlantic conditions (NAO effect of 0.3002, $P<0.05$ for *S. salar* and 0.1699, $P<0.05$ for *S. trutta*).

Temporal fluctuations in the abundance of Tyne salmonids were sufficiently independent to be modelled separately. It was suggested that the video data recorded at the Riding Mill station constituted a source of information that was not complete but nonetheless informative of the drivers specific to migration patterns of each species.

Chapter 5 suggested learning in the identification of salmonids as *S. salar* or *S. trutta* by data collectors that may have influenced goodness of fit of models considering the species separately (Chapter 4) early in the time series; this reached a plateau during the last years of observation. Temporal variation in the response of salmonid abundance to environmental parameters was suggested by theoretical arguments such as a large geographic range and its description via static measurements, as well as empirical arguments such as the seasonal variation in response to temperature (Chapter 2), relevance of a complex Box-Cox transformation of the response data (suspected to be representative of several processes with superposed time-scales (Chapter 3) and evidence of transient periodicities in fluctuations of salmonid abundance (Chapter 4).

Fluctuations in counts of *S. salar* varied periodically. State-space models suggested a high complexity within the population demographics of both species, via time-varying patterns likely related to age-class specificity in response to environmental conditions, suggesting that the records contain complex information on population demographics. Understanding of the whole

population by modelling is likely to greatly benefit from a detailed partitioning of the salmonid counts into age-classes, that would allow accounting for an age-class specific response to environmental conditions.

Inferences from the Chapters 3 to 5 demonstrated the importance of oceanic conditions as drivers in the observed patterns, for the post-smolts that enter the North Sea Basin and as grilse and kelt return to spawn, reflected by the relevance of NAO values in predicting abundance.

Highly nutritious marine organisms eaten by salmonids at sea justify the energy put into an anadromous mode of life (Hansen and Quinn, 1998); the growth of salmonids is directly dependent upon their access and use of abundant marine resources (Thorpe, 1984).

Temporal variation in plankton index (which reflects quality and quantity of plankton for larval cod *Gadus morhua*, Beaugrand *et al.*, 2004) were compared to net catches of *S. trutta* in the River Tweed for the period 1958-1998 (Bendall, The Living North Sea Project and the Rivers Trust). A strong link between the abundance of *C. Finmarchicus* and NAO values was demonstrated for the period 1962-1992 (Fromentin and Planque, 1996). This suggests a common response to a marine covariate, and so an indirect but still causal relationship between plankton and salmonids, as salmonids feed on larger crustaceans and are likely not influenced by plankton directly. Increases in phytoplankton production are associated with decreases in zooplankton production and *S. salar,* an increase in temperature and NAO values (Windsor *et al.*, 2012).

Instead, the influence of NAO values in describing temporal patterns of abundance in salmonids was hypothesized to be a partial representation of the importance of other components of the oceanic environment. Namely, NAO values likely reflected variations in the surface currents, temperature, and salinity known to shape the distribution of post-smolts (post-smolt initial movements at sea rely on these parameters, Holm *et al.*, 2000). Temperature is one recurrent axis of the multidimensional niche of fish (Magnuson *et al.,* 1979); movements of thermal optima shape fish communities in terms of distribution and diversity (MacNeil *et al.*, 2010; Beaugrand *et al.,* 2008). Characteristics and importance of thermal habitat may vary according to the sensitivity of the fish and hence its life stage. As salmonid post-smolts enter the

North Sea Basin, unsuitable thermal conditions may force the fish to swim harder and further to reach improved conditions; this creates an energy deficit at a crucial life stage that may be lethal (Friedland, 1998). Shifts in oceanic conditions may also lead to shift in predators and prey that may lead to the post-smolts missing on resources and not reach the body size necessary to face the predation associated with the diverse regions covered during migrating (Beaugrand, 2004). The conditions at this precise life-stage consequently have important and lasting consequences.

Chapters 3 to 5 concurred with the hypothesis that broad-scale factors are of high influence and that long-term changes in the population of salmonids are linked to conditions experienced while overwintering in the northern Atlantic Ocean (Hansen *et al.*, 2012 about *S. salar*). It was shown that oceanic thermohaline conditions define the extent to which grilse and kelt can detect the return route, and are strongly related to NAO that partly draws the required thermal path for salmonids up to their natal river (Holm *et al.*, 2000; Friedland, 1998), and that abundance of spawners on a given year naturally determines the abundance of offspring years later (in the River Tyne, the average time appeared to be of 4 years).

In chapter 6, differences between river conditions associated with presence of *S. salar* and *S. trutta* were investigated. The two species were shown to be distinct with regards to extreme values of signal size and flow velocity; time of year also contributed to classifying salmonids more reliably. Larger fish found within higher flows earlier in the year were most likely *S. salar*, whilst the opposite conditions were associated with a higher chance of being *S. trutta*. Both species were similar in their response to many of the covariates measured in the River Tyne, so classification rate success remained low.

Salmonid migration patterns were described based on models of continuous count observations at one point in the River Tyne. Much statistical modelling in ecology assumes that the data collected are made up of a large number of independent and identical components (Weinberg, 1975). However, ecological data are often highly correlated, with dependence between phenomena over different scales, driven by environmental parameters which themselves may be correlated. This means that data collected longitudinally are typically autocorrelated (Roy *et al.*, 2005) and must be modelled appropriately.

Variation in response may originate from phenomena that vary over several time scales. When changes are rapid relative to the temporal domain, resulting in largely independent values (*e.g*. year-to-year changes produce independent values from one year to the other as in individual fish survival rates, Pyper and Peterman, 1998), this variability is said to occur at high frequency. In these instances, serial correlation is unlikely to interfere with frequentist inference. Low frequency variability characterises slow changes. These occur on a long-term scale and lead to closely related values, *e.g*. values from one year to the other are this time very similar (Pyper and Peterman, 1998). Low-frequency variability constitutes a statistical challenge in that it leads to serial dependence.

Various approaches are available to address the issue of serial dependence explicitly. Often, response variable may be transformed so that a particular frequency of variability dominates the signal contained in the data. This amounts to selecting a type of transformation in accordance with the time-scale of the variability that is investigated. For instance, the first differences of a response variable remove low-frequency variability (slow changes) contained in the original pattern so that high-frequency (rapid changes) may be better detected; a smoothed version of a time-series may emphasize low-frequency variability (slow changes) that may otherwise obstructed by high autocorrelation (rapid changes) (Pyper and Peterman, 1998).

Transformations may lead to an improved fit of the models but this may be combined with a lowered interpretability if the models do not improve the understanding of the processes modelled. The distorting effects of a transformation on the signal means that they are not always ecologically meaningful, which renders the outputs of models challenging to interpret and may lead to missing important relationships between different pace of change in the processes involved (Pyper and Peterman, 1998).

In this study, an alternative approach was used to address the issue of serial correlation. The scale of observation was adjusted by considering various resolutions of the samples. By doing so, variability was investigated at several levels; the smallest scale (and highest resolution) investigated the lowest frequency variability (Chapter 2, individual behaviour) and the largest scale (and lowest resolution) investigated the highest frequency variability (Chapter 3, salmonid population behaviour). In between, partitioning of salmonid counts

into two distinct species groups allowed for refinement of the description of high frequency variability (Chapters 4 and 5, species-specific population behaviour). Limitations due to assumptions of independence and randomness that rarely apply to living systems were avoided by using different and complementary modelling techniques, considering small and broad scales and investigating different partitioning of the population (*i.e.* as individuals, as a whole, as species).

Ecological systems contain different levels of complexity that form an ”organized complexity” (O'Neill *et al.*, 1986). By analysing the complexity of a living system, one in fact seeks to define the organization underlying the complexity; the latter may be studied and understood with mathematical tools. The study of living organisms often requires partitioning observational data into arbitrary groups, and many studies consider a single scale made up of closed and isolated communities to get around such problems (Leibold *et al.*, 2004). Whether considering individuals or communities, ecologically defined entities may be thought of as responding individually to temporal and spatial variation (Levin, 1992).

The assumption of an underlying organization reflected by the salmonid count data lead to the study of time scales of variability in the processes driving salmonid abundance. The scope of ecology contains assemblages of species, environmental conditions and processes across scales (Wittaker, 1975). Arbitrary structures within a living system amounts to the division of ecological processes according to the scale over which they take place; particular scales are more important to the description of particular processes (Cash *et al.*, 2006). Here, the use and contrast of several scales of resolution amounted to studying of different levels of organization within the salmonid population, as if observing salmonids from different distances and within with changing fields of observation. Resolution of observation related to the size of minimal sufficient detail to describe the living system (Levin, 1992).

The problem of the correspondence between scale and pattern is considered a central problem in ecology due to numerous and confounded scales of space, time, and ecological organization as well as the fact that apparent characteristics of a living system may be altered by the resolution used to study it (Levin, 1992). Short and long-term studies are therefore complementary in that they provide

answers to questions posed at different scales, and the combination is necessary to a complete understanding of the mechanisms determining the dynamics in a system.

This approach (*e.g*. MacNeil *et al.,* 2009; Connell *et al*., 1997) relates to complex systems science, which uses tools from both philosophy and mathematics to investigate how the relationships between parts of a system can shape a collective behaviour, and how in turn the system interacts with the environment (Anand *et al*., 2010). This emerging scientific approach has contributed to developing ecological concepts in a new light, including the understanding of fluctuations in abundance of natural populations (Anand *et al*., 2010). Here, the signal contained in the observational data was not a perfect sinusoid, but the signal was modelled which implies that it was not random. This indicates that the living system that was studied was not ordered, nor disordered; the living system was a complex system.

In a study by Cohen (1995), hypothetical population fluctuations were simulated using classic population models in which the environmental parameters were set as constant white noise. As a result, the population size was set to be dependent upon internal dynamics only (*i.e*. $N_{t+1} \sim$ *function* (N_t), using several types of function based on empirical studies). The simulated time-series were dominated by high frequency variability, which suggested that the intrinsic population dynamics caused rapid changes in the population abundance. However, natural time-series are commonly described by models based on low-frequency variability (*i.e*. the variability induced by environmental parameters). These models may consequently not be accurate for the description of the high frequency variability contained in time-series, which is caused by intrinsic population dynamics (Cohen, 1995). For this reason, the persistence of autocorrelation in the models of ecological data may be considered a witness of the importance of the intrinsic population dynamics in describing the fluctuations in abundance (as here, in salmonids), rather be considered as an diagnostic elements against model selection.

This concept is a direct product of complex systems science approach to the analysis of population time-series (Anand *et al*., 2010) in that it seeks to complement the understanding of a global ecological process by applying a concept of a given scale to observational data (*i.e*. slow response to

environmental parameter) and pursue the investigation of the process by interpreting the model output using principles emerging from a smaller scale (*i.e.* fast changes due to intrinsic population dynamics), rather than discrediting a model function based on a lack of fit to the underlying statistical assumptions.

In that idea, either there are no scaling laws in ecology or each and every scale serve the description of a living system equally well (Levin, 1992). The latter suggests that no scale is more relevant to the description of a living system than another. So, the scale at which models should be developed depends on the question asked. However, using various scales allows the description of different ecological processes because the scale either suppresses or emphasizes detail. By generating a set of models at different levels of complexity, the internal heterogeneity of a system may be understood as well as the essential forces driving the broad scale dynamics. Analyses over multiple question-specific scales are a powerful tool to understand the development and maintenance of living systems.

Appendices. .

"My favourite piece of information is that Branwell Brontë, brother of Emily and Charlotte, died standing up leaning against a mantel piece, in order to prove it could be done.

This is not quite true, in fact. My absolute favourite piece of information is the fact that young sloths are so inept that they frequently grab their own arms and legs instead of tree limbs, and fall out of trees.

However, this is not relevant to what is currently on my mind because it concerns sloths, whereas the Branwell Brontë piece of information concerns writers and feeling like death and doing things to prove they can be done, all of which are pertinent to my current situation to a degree that is, frankly, spooky."

D. Adams, 2002.
The Salmon of Doubt

Appendices.

Appendix A.i

Table A.1: The riparian covariates and relevant measurement lags according to the flow regime.

Parameter	Location (N.G.R.)	Distance to counter (NZ 027619)	Measurement lag	
			Low regime	High regime
Flow Temperature	Bywell (NZ038617)	1.36km	45minutes	0
Turbidity Colour	Horsely (NZ100643)	9.27km	4 hours	45minutes

Table A.2: The characteristic values of the covariate flow for each observation period.

All flow values expressed in $m^3.s^{-1}$, the mean in 2008 was 68.04$m^3.s^{-1}$.

Observation period	Lowest flow rate	Highest flow rate	Mean	Measurements above 2008 mean	Selected regime
a	51.80	133.00	51.80	27.6%	Low
b	11.00	17.70	11.92	0%	Low
c	7.73	10.70	9.42	0%	Low
d	15.40	88.30	39.84	21.6%	Low
e	37.50	153.00	70.25	63.0%	High
f	35.50	129.00	54.05	17.7%	Low
g	57.90	191.00	100.52	76.6%	High

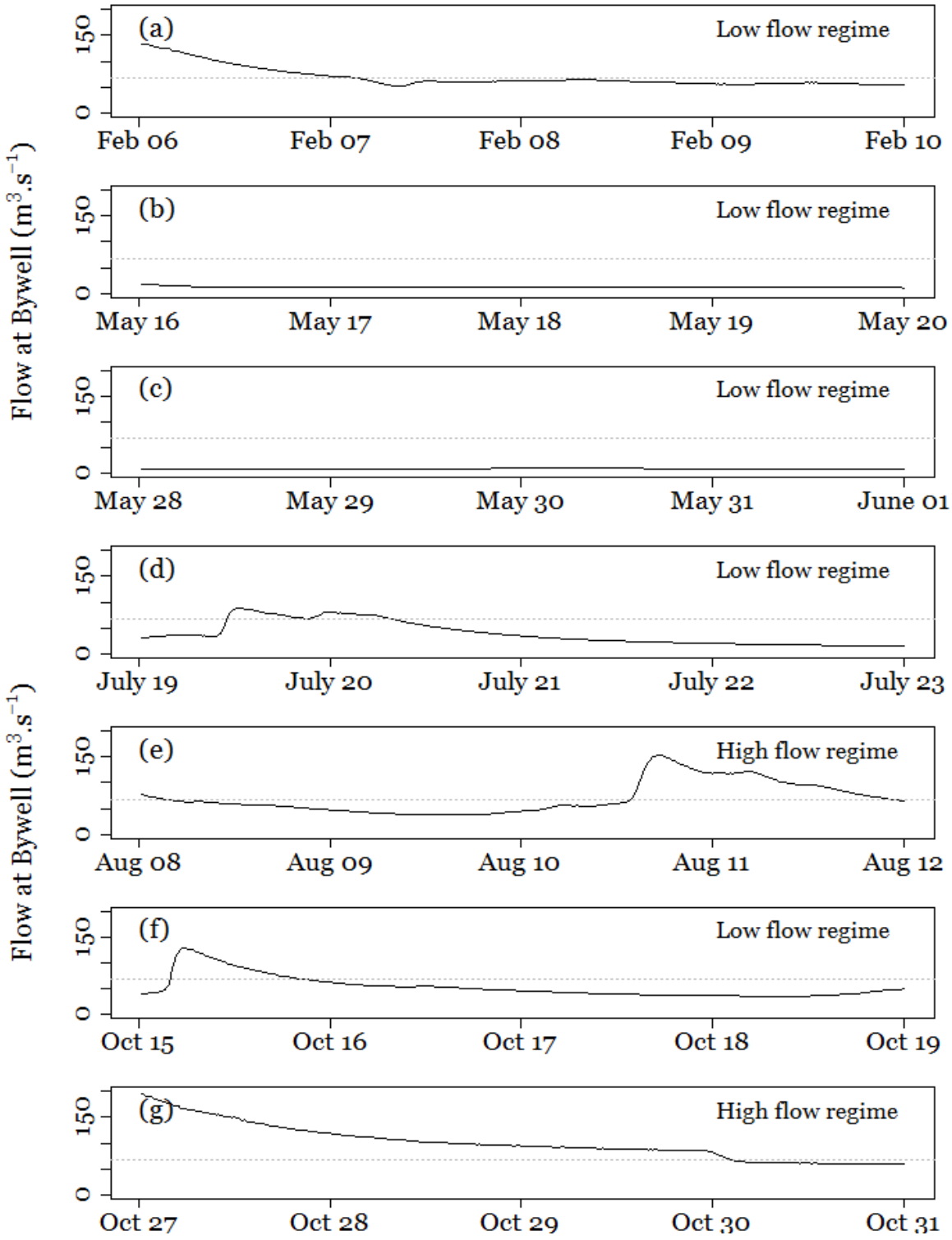


Figure A.7.1: The values for flow rate as measured at Bywell, for each observation period.
The flow regime was selected based on a comparison with the annual mean flow rate value; two periods presented high regimes (e and g) and the others presented low flow regimes.

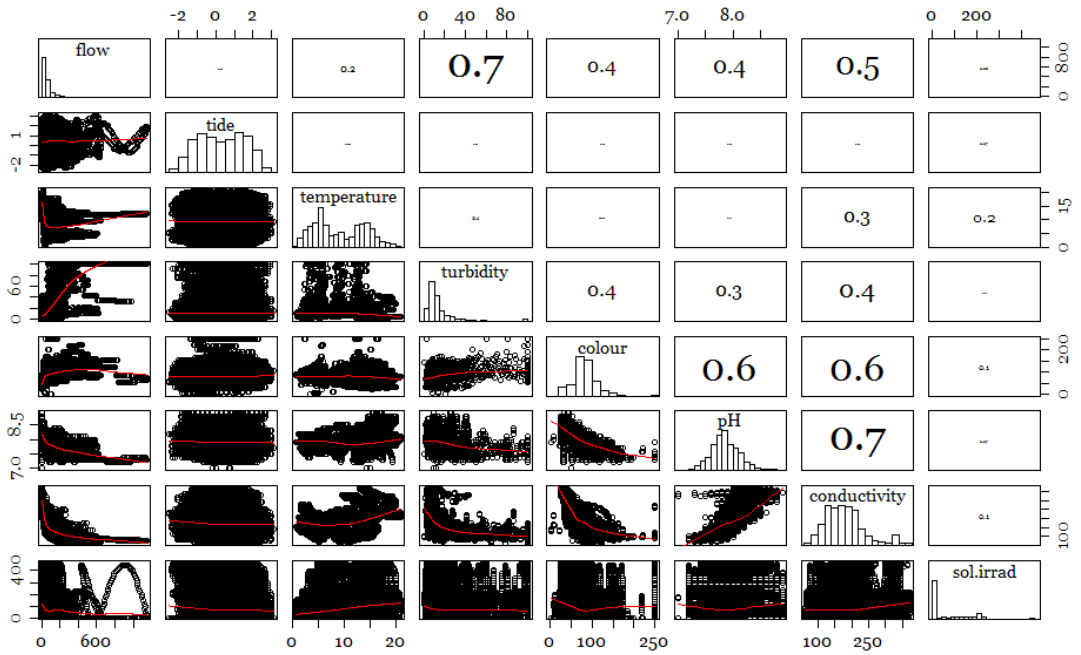


Figure A.7.2: Pair-wise correlations considering the whole year 2008.
Correlation values are in the upper panel and the associated pair-wise scatter-plots are in the lower panel, with the distribution of each covariate in the diagonal.

Table A.3: The full model of each type for each observation period (a to g) ranked according to their AICc score.

Period a	AICc	df	ΔAICc	weight
Full zero inflated Poisson	47.4	12	0.0	0.7275
Full zero inflated negative binomial	50.3	13	2.9	0.1749
Full negative binomial GLM	51.9	9	4.5	0.0777
Full Poisson GLM	54.6	7	7.2	0.0198
Full zero altered negative binomial	64.3	13	16.9	<0.001
Period b	AICc	df	ΔAICc	weight
Full zero inflated Poisson	228.2	18	0.0	0.887
Full zero inflated negative binomial	232.3	19	4.1	0.113
Full negative binomial GLM	246.0	10	17.8	<0.001
Full Poisson GLM	246.5	9	18.3	<0.001
Full zero altered Poisson	252.6	18	24.4	<0.001
Full zero altered negative binomial	254.9	19	26.6	<0.001
Period c	AICc	df	ΔAICc	weight
Full zero inflated negative binomial	801.2	19	0.0	1
Full negative binomial GLM	830.6	10	29.4	<0.001
Full zero altered negative binomial	836.5	19	35.4	<0.001
Full zero inflated Poisson	877.6	18	76.4	<0.001
Full zero altered Poisson	879.5	18	78.4	<0.001
Full Poisson GLM	947.0	9	145.8	<0.001
Period d	AICc	df	ΔAICc	weight
Full zero inflated negative binomial	1469.9	19	0.0	0.99701
Full zero inflated Poisson	1481.5	18	11.6	0.00299
Full zero altered negative binomial	1494.9	19	25.0	< 0.001
Full zero altered Poisson	1500.2	18	30.3	< 0.001
Full negative binomial GLM	1520.2	10	50.3	< 0.001
Full Poisson GLM	1553.5	9	83.6	< 0.001
Period e	AICc	df	ΔAICc	weight
Full zero inflated negative binomial	1022.5	19	0.0	1
Full zero altered negative binomial	1065.1	19	42.7	<0.001
Full negative binomial GLM	1082.0	10	59.6	<0.001
Full Poisson GLM	1092.7	9	70.2	<0.001
Period f	AICc	df	ΔAICc	weight
Full zero altered Poisson	1239.0	20	0.0	0.534
Full zero altered negative binomial	1239.3	21	0.3	0.466
Full negative binomial GLM	1258.1	11	19.1	<0.001
Full Poisson GLM	1271.5	10	32.5	<0.001
Period g	AICc	df	ΔAICc	weight
Full zero inflated Poisson	487.5	18	0.0	0.9281
Full zero inflated negative binomial	492.6	21	5.1	0.0719
Full zero altered Poisson	523.2	18	35.7	<0.001
Full zero altered negative binomial	525.8	20	38.3	<0.001
Full Poisson GLM	536.7	10	49.2	<0.001
Full negative binomial GLM	536.8	11	49.2	<0.001

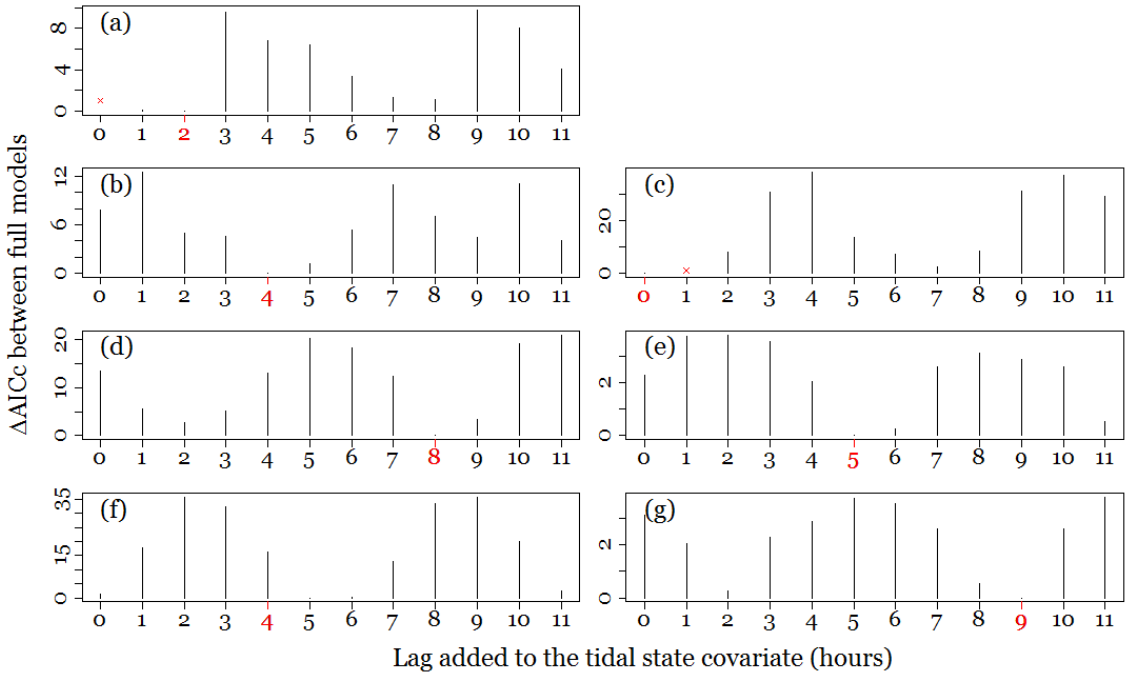


Figure A.7.3: The differences in AICc scores compared to the best-ranked model, for the tidal lags covering one tidal cycle.
The full models were used for each time period (*a* to *g*), using the model type selected beforehand. The cyclical properties of the tide engendered a cyclical fluctuation of the ΔAICc scores for increasing tidal values. The numbers in red were the lags generating the lowest AICc scores, symbols (x) represent no model convergence.

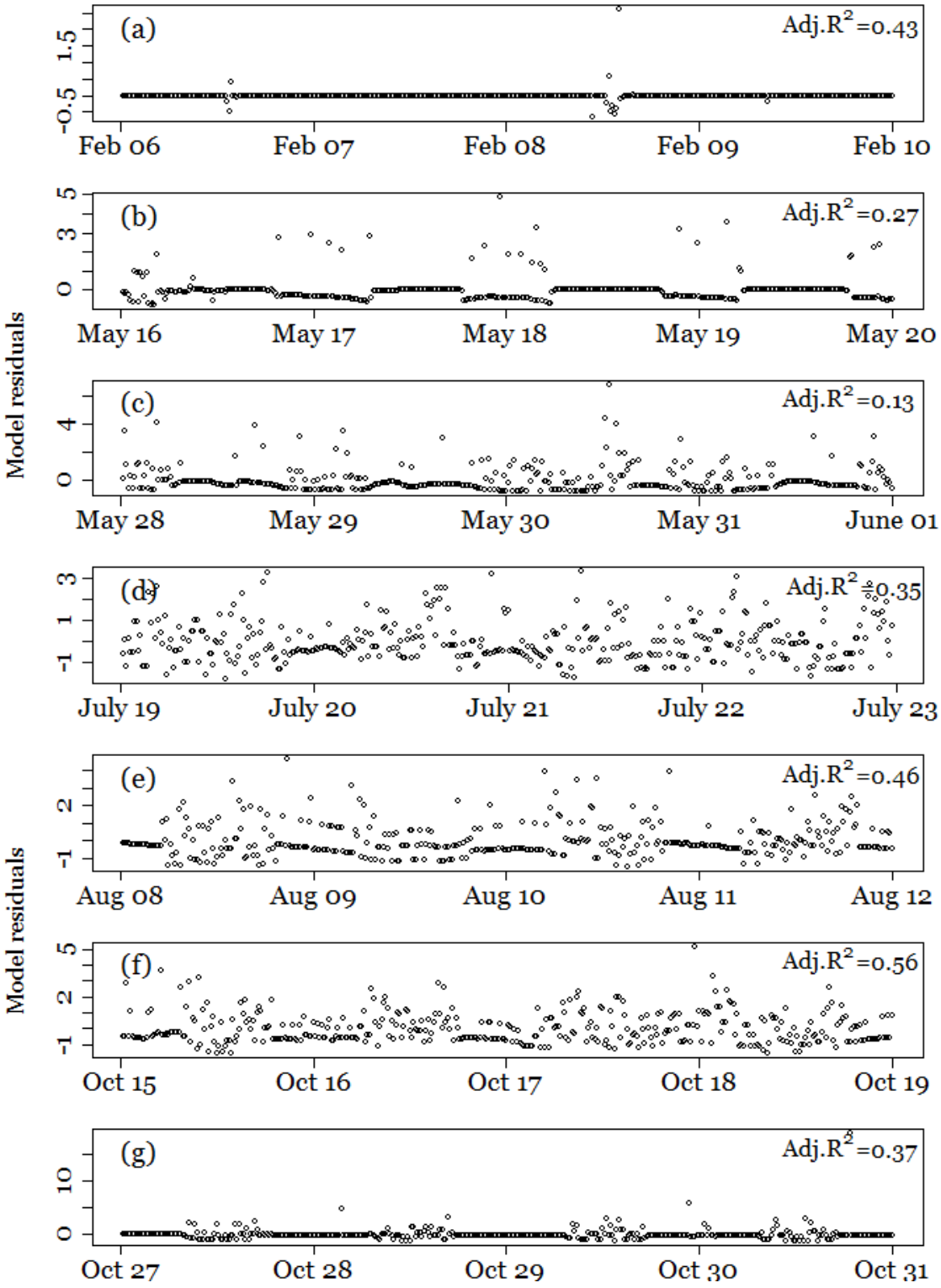


Figure A.7.4: The temporal variations of the residuals of each model throughout each observation period.
There were temporal patterns in the residuals. Note the change in scale.

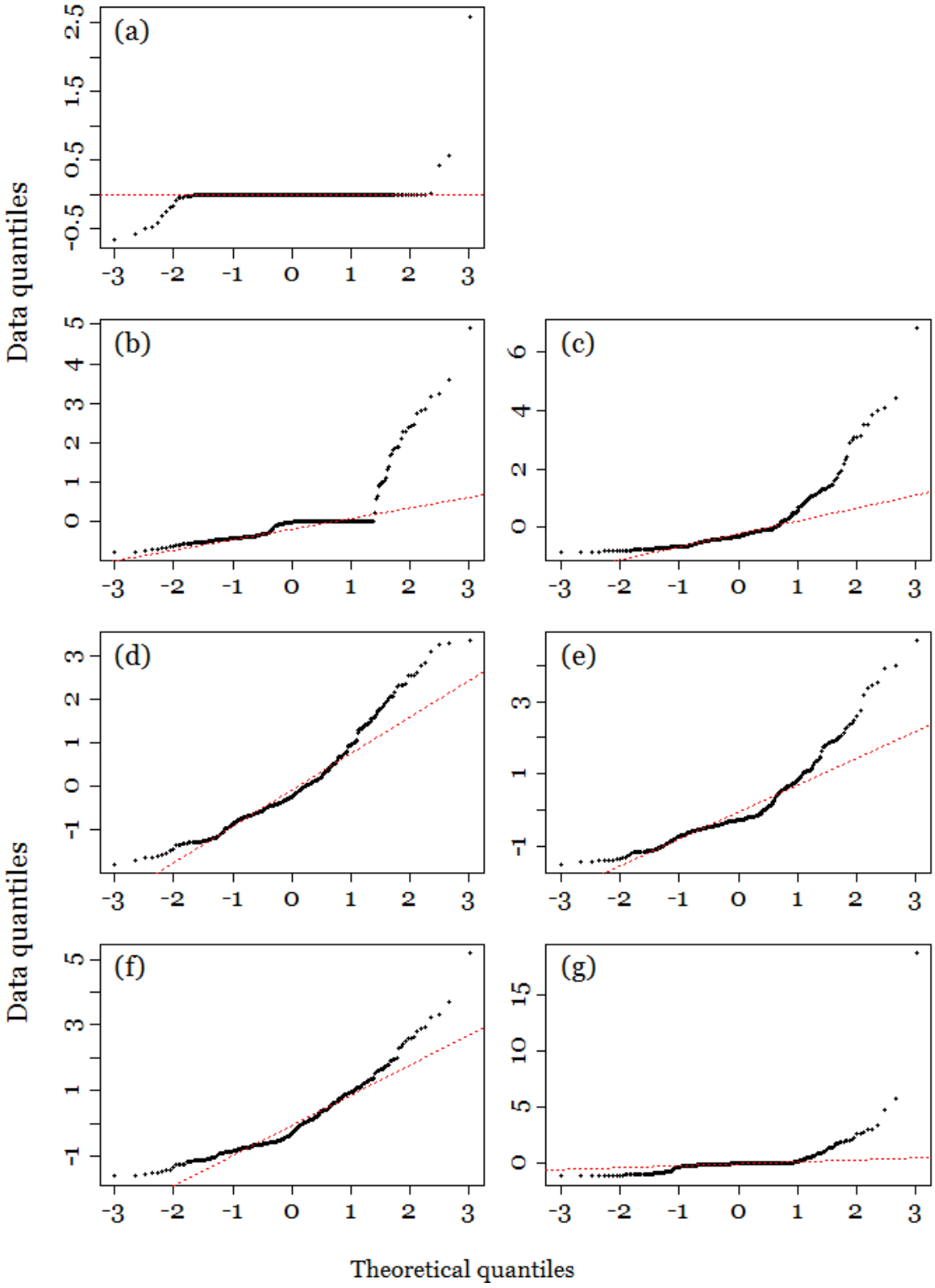


Figure A.7.5: The normal Q-Q plots of the residuals of the most parsimonious model for each observation period.
The fitted values (vertical axis) were compared to the normal distribution on the horizontal axis. The red dotted line (x=y) represented the normal distribution. For all observation periods, the points followed a strongly nonlinear pattern, suggesting that the residuals were not normally distributed.

Table A.4: The estimates for the most parsimonious model for each observation period.
Note that the tide has a different lag depending on the observation period.

		a	b	c	d	e	f	g
Logistic part	Intercept	-1717	-1352.44	46.32	57.42	-28.49	-67.52	24.46
	Flow	-51.90	20.89	-7.76	-0.1	.	-0.03	.
	Temperature	-452.78	-36.90	.	.	5.95	-1.85	-4.61
	Tide	332.14	10.05	2.55	1.04	0.90	-0.21	.
	Turbidity	.	14.14	1.10	.	.	0.10	0.85
	Colour	50.67	1.96	.	.	.	.	.
	pH	.	143.85	.	.	-7.83	10.78	.
	Conductivity	.	0.33	.	-0.37	.	.	.
	Solar irradiance	28.98	0.43	0.04	-0.06	-0.19	0.02	-0.14
Count part	Intercept	-229.53	35.24	17.25	34.92	14.44	6.930	-4.42
	Flow	-1.34	.	.	0.02	.	-0.038	-0.02
	Temperature		.	0.35	-0.25	0.393	.	0.86
	Tide	24.66	0.06	0.58	.	.	.	.
	Turbidity	.	.	.	0.04	.	.	.
	Colour	2.31	-0.09	.	-0.01	.	.	.
	pH	.	-3.08	-2.67	-4.77	-2.678	.	.
	Conductivity	.	-0.02	.	0.04	.	-0.028	.
	Solar irradiance	1.82	0.01	-0.00	0.00	0.005	0.007	0.00
	Log(Theta)	.	.	-0.02	2.18	2.290	.	.
	Theta	.	.	0.98	8.87	9.872	.	.

Table A.5: The significance corresponding to the estimates for the most parsimonious model for each observation period.
P-values are reported. Note that the tide has a different lag depending on the observation period.

		a	b	c	d	e	f	g
Logistic part	Intercept	1.00	0.07	0.01	0.06	0.271	0.020	0.068
	Flow	0.37	0.08	0.01	0.09	.	0.001	.
	Temperature	0.69	0.07	.	.	0.003	0.000	0.064
	Tide	NA	0.06	0.02	0.04	0.066	0.012	.
	Turbidity	.	0.21	0.02	.	.	0.001	0.024
	Colour	NA	0.14	.	.	.	.	.
	pH	.	0.05	.	.	0.002	0.007	.
	Conductivity	.	0.33	.	0.06	.	.	.
	Solar irradiance	NA	0.08	0.00	0.02	0.005	0.000	-0.000
Count part	Intercept	0.16	0.02	0.00	0.00	0.000	0.000	0.057
	Flow	0.19	.	.	0.00	.	0.000	0.023
	Temperature	.	.	0.01	0.00	0.002	.	0.053
	Tide	0.18	0.67	0.00	.	.	.	.
	Turbidity	.	.	.	0.00	.	.	.
	Colour	0.15	0.01	.	0.00	.	.	.
	pH	.	0.07	0.00	0.00	0.000	.	.
	Conductivity	.	0.01	.	0.00	.	0.000	.
	Solar irradiance	0.18	0.03	0.00	0.00	0.000	0.000	0.028
	Log(Theta)	.	.	0.93	0.00	0.000	.	.

The stepwise removal of each parameter contained in the optimal model (lowest AICc value) for each observation period (a to g) produced the model outputs below. Parameters were contained in the count part (μ), the logistic part (π) or both. The differences in AICc and likelihood after each removal are reported. Parameters are ranked from the most to the least significant.

Table A.6: The output of stepwise removal of each parameter for period (a).

Dropped term	Model part	ΔAICc	df	Likelihood ratio test	
Tide	μ	15.6	10	χ^2= 17.719	df=1, *P*=2.561e-05
Temperature	π	15.3	10	χ^2= 17.462	df=1, *P* =2.93e-05
Colour	π	15.0	10	χ^2= 17.088	df=1, *P*= 3.568e-05
Solar irradiance	μ	14.4	10	χ^2= 16.530	df=1, *P*=4.788e-05
Flow	π	12.5	10	χ^2= 14.668	df=1, *P*= 0.0001282
Solar irradiance	π	12.0	10	χ^2= 14.105	df=1, *P*=0.0001729
Tide	π	10.9	10	χ^2= 12.974	df=1, *P*= 0.0003159
Flow	μ	9.1	10	χ^2= 11.242	df=1, *P*=0.0007995
Colour	μ	8.8	10	χ^2= 10.9639	df=1, P=0.0009639
None	.	0.0	11	.	.

Table A.7: The output of stepwise removal of each parameter for period (b).

Dropped term	Model part	ΔAICc	df	Likelihood ratio test	
Solar irradiance	π	33.0	14	χ^2=35.132	df=1, *P*=3.081e-09
Temperature	π	21.9	14	χ^2=24.061	df=1, *P*=9.332e-07
pH	π	18.5	14	χ^2=20.693	df=1, *P*=5.391e-06
Conductivity	π	16.9	14	χ^2=19.039	df=1, *P*=1.281e-05
Colour	π	15.9	14	χ^2=18.035	df=1, *P*=2.169e-05
Tide	π	15.6	14	χ^2=17.808	df=1, *P*=2.444e-05
Conductivity	μ	14.9	14	χ^2=17.102	df=1, *P*=3.543e-05
Solar irradiance	μ	13.9	14	χ^2=16.020	df=1, *P*=6.266e-05
Turbidity)	π	12.3	14	χ^2=14.486	df=1, *P*=0.0001412
Colour	μ	9.6	14	χ^2=11.815	df=1, *P*=0.0005877
Flow	π	8.1	14	χ^2=10.275	df=1, *P*=0.001349
Tide	μ	6.8	14	χ^2=8.9787	df=1, *P*=0.002731
pH	μ	5.3	14	χ^2=7.4997	df=1, *P*=0.006171
None	.	0.0	15	.	.

Table A.8: The output of stepwise removal of each parameter for period (c).

Dropped term	Model part	ΔAICc	df	Likelihood ratio test	
Tide	μ	28.8	10	χ^2=30.887	df=1, *P*=2.735e-08
pH	μ	23.1	10	χ^2=25.266	df=1, *P*=4.995e-07
Solar irradiance	π	19.7	10	χ^2=21.846	df=1, *P*=2.954e-06
Flow	π	19.7	10	χ^2=21.821	df=1, *P*=2.993e-06
Solar irradiance	μ	13.4	10	χ^2=15.502	df=1, *P*=8.242e-05
Tide	π	8.9	10	χ^2=10.977	df=1, *P*=0.0009224
Turbidity	π	7.8	10	χ^2=9.8865	df=1, *P*=0.001665
Temperature	μ	5.2	10	χ^2=7.3565	df=1, *P*=0.006682
None	.	0.0	11	.	.

Table A.9: The output of stepwise removal of each parameter for period (d).

Dropped term	Model part	ΔAICc	Df	Likelihood ratio test	
Conductivity	μ	53.8	13	χ^2=55.958	df=1, *P*=7.403e-14
Solar irradiance	π	36.4	13	χ^2=38.523	df=1, *P*=5.411e-10
pH	μ	28.4	13	χ^2=30.512	df=1, *P*= 3.318e-08
Flow	μ	19.8	13	χ^2=21.922	df=1, *P*=2.84e-06
Conductivity	π	18.0	13	χ^2=20.174	df=1, *P*=7.071e-06
Temperature	μ	16.4	13	χ^2=18.583	df=1, *P*=1.626e-05
Turbidity	μ	11.6	13	χ^2=13.713	df=1, *P*=0.000213)
Solar irradiance	μ	10.1	13	χ^2=12.258	df=1, *P*=0.0004634
Colour	μ	6.1	13	χ^2=8.2885	df=1, *P*=0.00399)
Tide	π	3.3	13	χ^2=5.444	df=1, *P*=0.01964
Flow	π	2.8	13	χ^2=4.9783	df=1, *P*=0.02567
None	.	0.0	14	.	.

Table A.10: The output of stepwise removal of each parameter for period (e).

Dropped term	Model part	ΔAICc	df	Likelihood ratio test	
Solar irradiance	π	68.6	9	χ^2=70.695	df=1, *P*<2.2e-16
pH	μ	58.9	9	χ^2=60.976	df=1, *P*=5.777e-15
Solar irradiance	μ	39.3	9	χ^2=41.425	df=1, *P*=1.225e-10
*p*H	π	14.9	9	χ^2=16.992	df=1, *P*=3.753e-05
Temperature	π	12.0	9	χ^2=14.067	df=1, *P*=0.0001764
Temperature	μ	6.9	9	χ^2=8.9885	df=1, *P*=0.002717
Tide	π	2.3	9	χ^2=4.4276	df=1, *P*=0.03536
None	.	0.0	10	.	.

Table A.11: The output of stepwise removal of each parameter for period (f).

Dropped term	Model part	ΔAICc	df	Likelihood ratio test	
Solar irradiance	μ	115.5	12	χ^2=117.67	df=1, *P*< 2.2e-16
Solar irradiance	π	97.5	12	χ^2=99.62	df=1, *P*< 2.2e-16
Flow	μ	40.4	12	χ^2=42.557	df=1, *P*=6.864e-11
Tide	μ	31.3	12	χ^2=33.416	df=1, *P*=7.44e-09
Conductivity	μ	26.6	12	χ^2=28.703	df=1, *P*=8.438e-08
Temperature	π	15.0	12	χ^2=17.136	df=1, *P*=3.48e-05
Flow	π	10.1	12	χ^2= 12.929	df=1, *P*=0.0004549
Turbidity	π	8.8	12	χ^2=10.99	df=1, *P*=0.0009161
pH	π	5.2	12	χ^2=7.2991	df=1, *P*=0.006899
Tide	π	4.3	12	χ^2=6.4057	df=1, *P*=0.01138
none	.	0.0	13	.	.

Table A.12: The output of stepwise removal of each parameter for period (g).

Dropped term	Model part	ΔAICc	df	Likelihood ratio test	
Solar irradiance	π	61.8	7	χ^2=63.885	df=1, *P*=1.319e-15
Turbidity	π	5.3	7	χ^2= 7.3476	df=1, *P*=0.006715
Flow	μ	3.4	7	χ^2=5.4591	df=1, *P*=0.01947
Solar irradiance	μ	2.9	7	χ^2=4.9679	df=1, *P*=0.02582
Temperature	π	2.7	7	χ^2=4.7799	df=1, *P*=0.02879
Temperature	μ	1.8	7	χ^2=3.9224	df=1, *P*=0.04765
None	.	0.0	8	.	.

Appendix A.ii

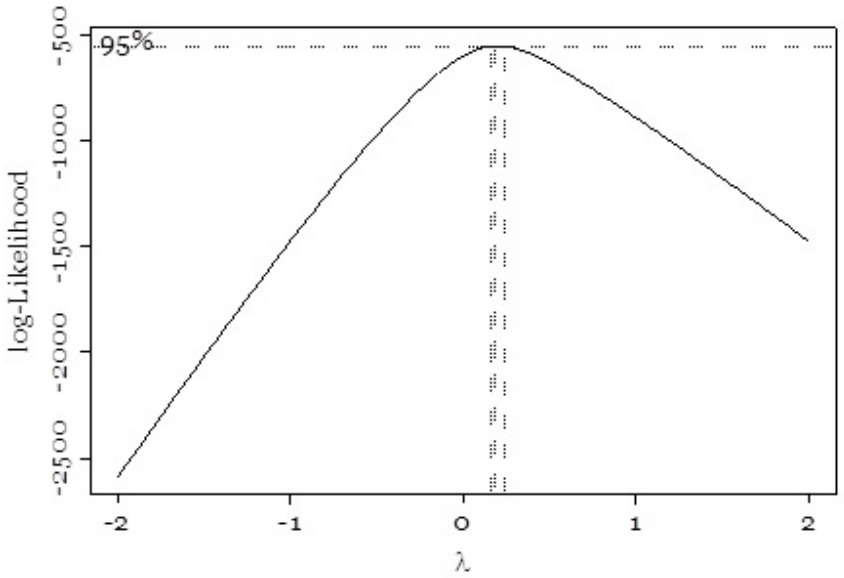


Figure A.7.6: The profile log-likelihoods for the parameter λ of the Box-Cox power transformation of the salmonid count data.

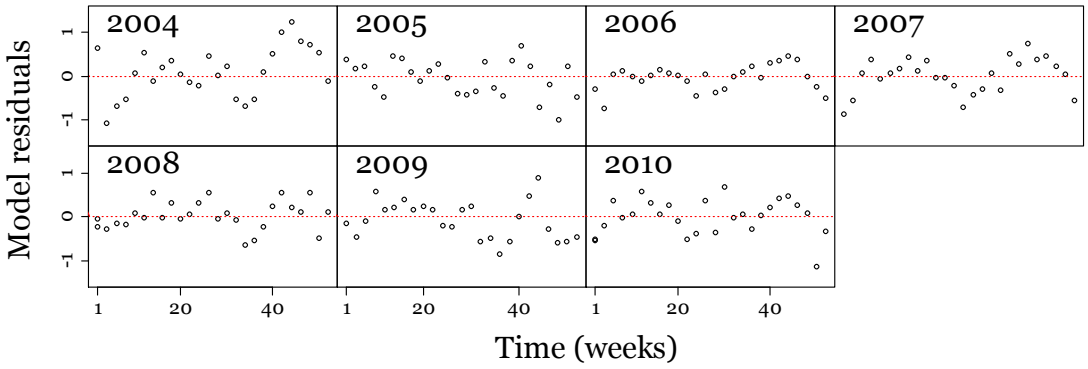


Figure A.7.7: LME model residuals for Box-Cox transformed salmonid counts per year for the observation period 2004-2011.

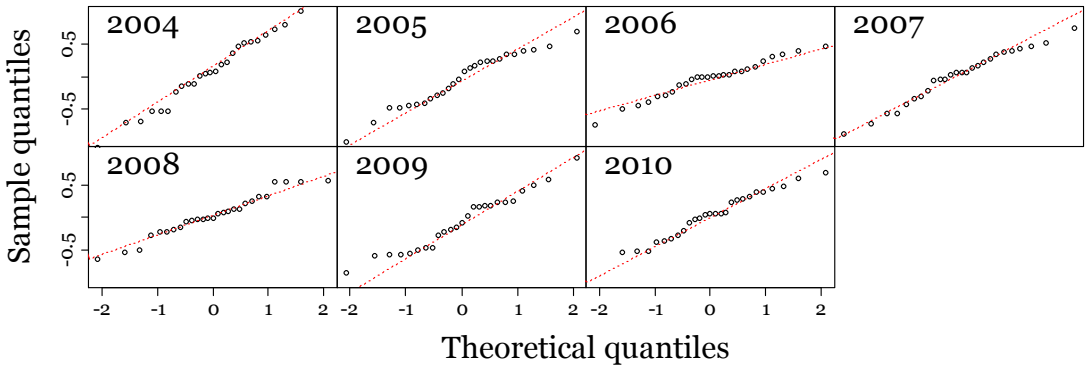


Figure A.7.8: Q-Q plots of LME model residuals for Box-Cox transformed salmonid counts per year for the observation period 2004-2011.

Table A.13: Correlations for random and fixed effects and within group residuals for the optimal LME model of Box-Cox transformed salmonid counts for the observation period 2004-2011.

Random effect	St. Dev	correlation	Fixed effects	intercept	correlation Cos (1year)	Sin (1year)	Temperature
Intercept	0.2932	.	Cos(1year)	-0.834	.	.	.
Temperature	0.0277	-0.923	Sin(1year)	-0.625	0.683	.	.
Residuals	0.4271	.	Temperature	-0.974	0.850	0.634	.
			NAO, lag 4years	-0.079	0.063	-0.107	0.079

Standardized Within-Group Residuals				
Min	Q1	Med	Q3	Max
-2.6406913	-0.6576961	0.0810359	0.6460973	2.8867025

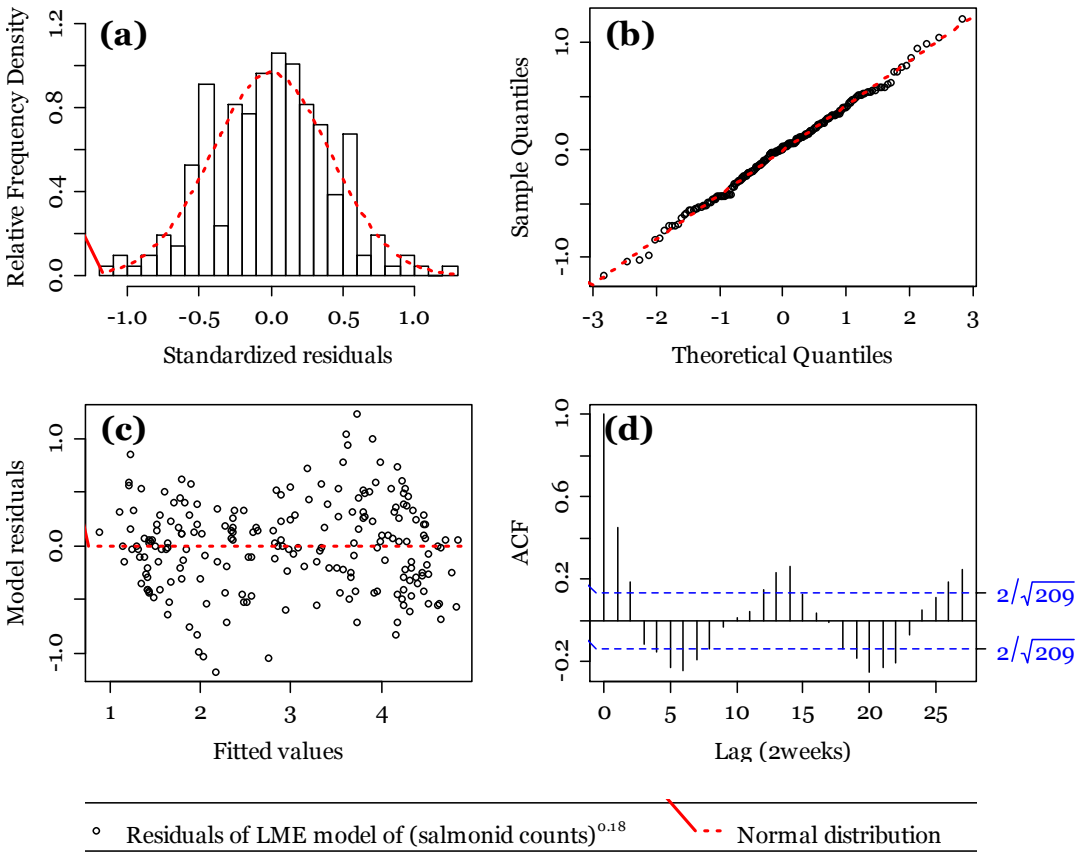


Figure A.7.9: Diagnostic plots for the optimal LME model for the Box-Cox transformed salmonid counts for the whole observation period.
The residuals were normally distributed (a&b) (Shapiro test: W=0.9963, P=0.8963), there was some evidence of (c) heteroscedasticity and (d) autocorrelation.

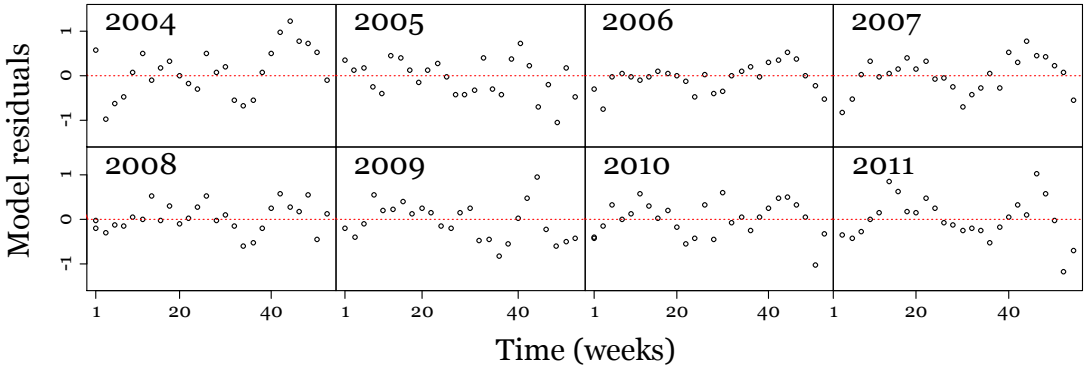


Figure A.7.10: LME model residuals for Box-Cox transformed salmonid counts per year for the whole observation period.

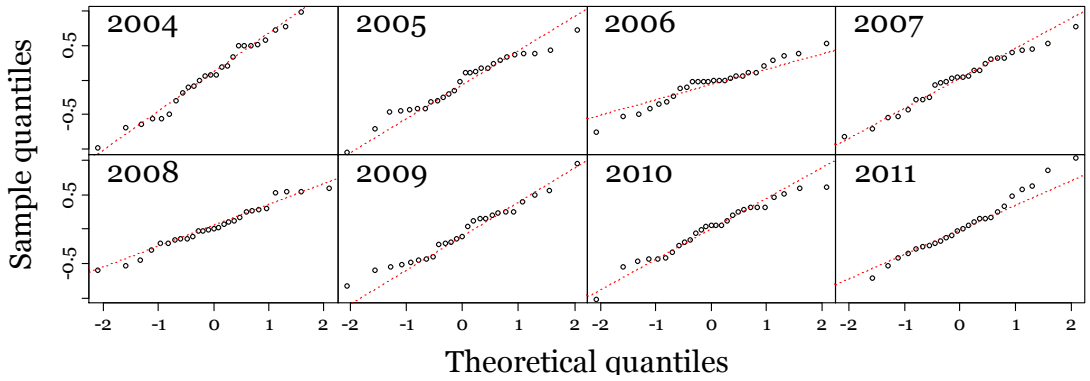


Figure A.7.11: Q-Q plots of LME model residuals for Box-Cox transformed salmonid counts per year for the whole observation period.

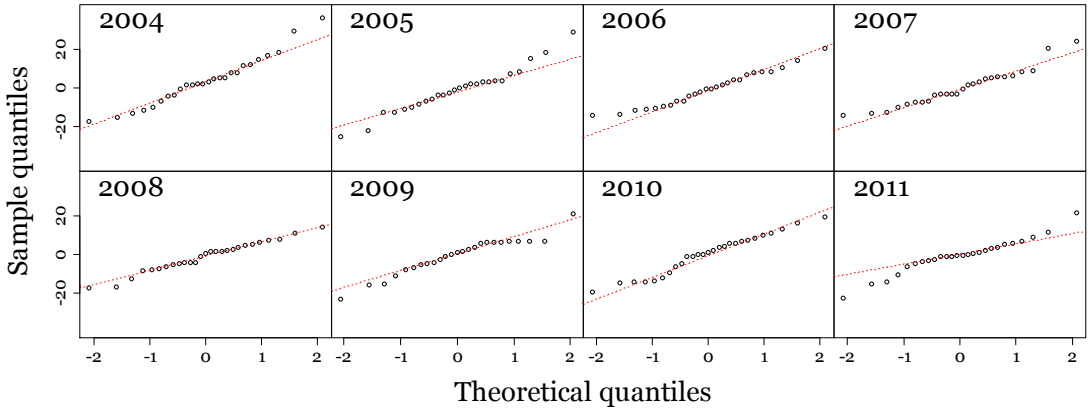


Figure A.7.12: Q-Q plots of LME model residuals for square root transformed salmonid counts per year for the whole observation period.

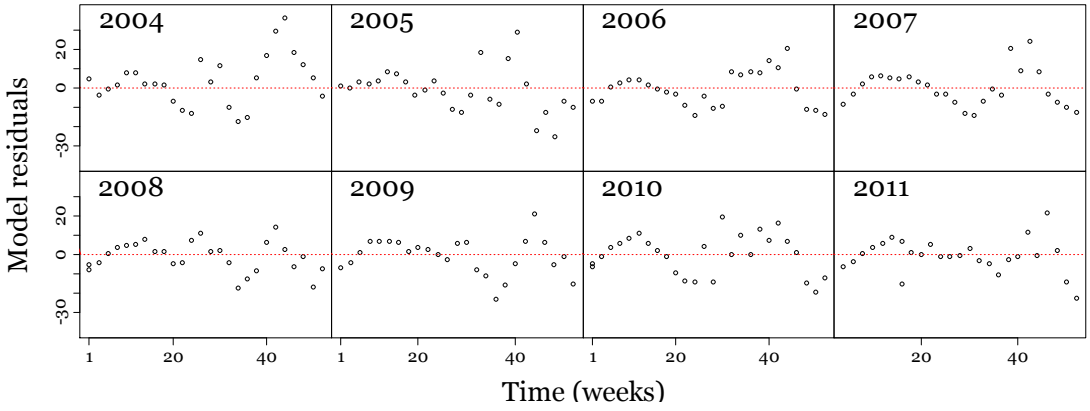


Figure A.7.13: LME model residuals for square root transformed salmonid counts per year for the whole observation period.

Table A.14: Correlations for random and fixed effects and within group residuals for the optimal LME model of square root transformed salmonid counts for the whole observation period.

Random effect	St. Dev	correl ation	Fixed effects	intercept	correlation Cos (1year)	Sin (1year)	Temper ature
Intercept	3.7073	.	Cos(1year)	-0.893	.	.	.
Temperature	0.5974	-0.984	Sin(1year)	-0.642	0.637	.	.
Residuals	10.2394	.	Temperature	-0.972	0.850	0.607	.
			Tunnel	-0.173	0.117	0.201	0.146

Standardized Within-Group Residuals				
Min	Q1	Med	Q3	Max
-2.45880254	-0.65749752	0.05155325	0.57114916	3.52157691

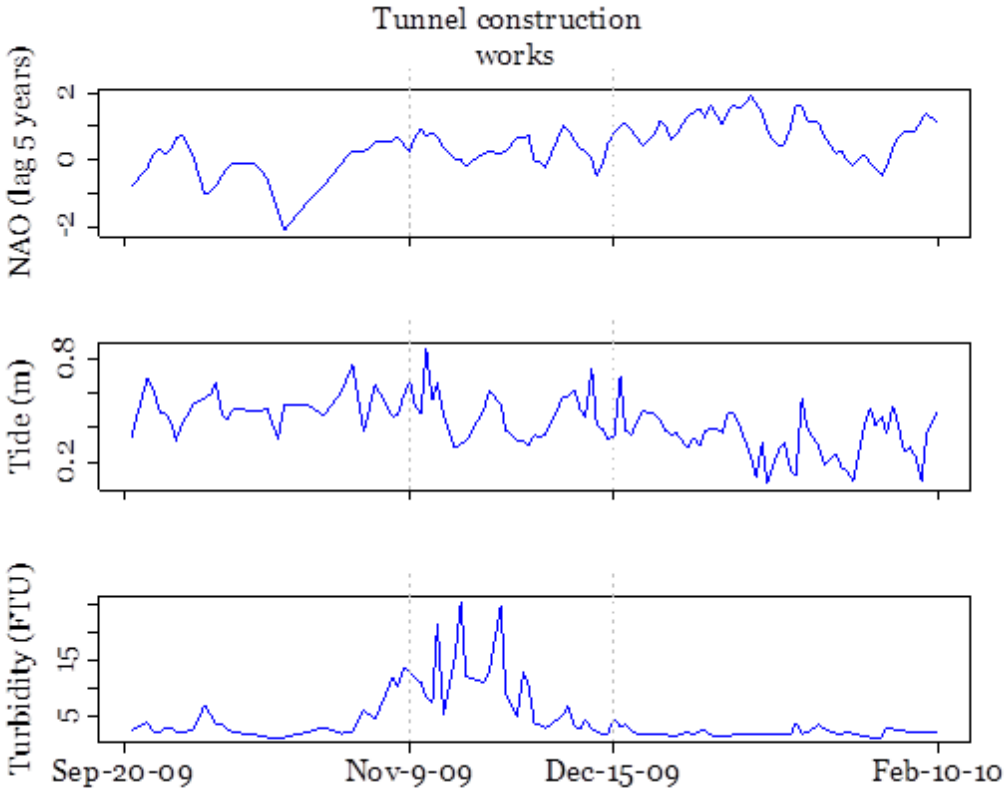


Figure A.7.14: The temporal variations of NAO (with a lag of 5 years), tidal state and turbidity over the tunnel construction period.
The components were included in the most parsimonious hurdle model for this period. The interval between measure is one day.

Appendix A.iii

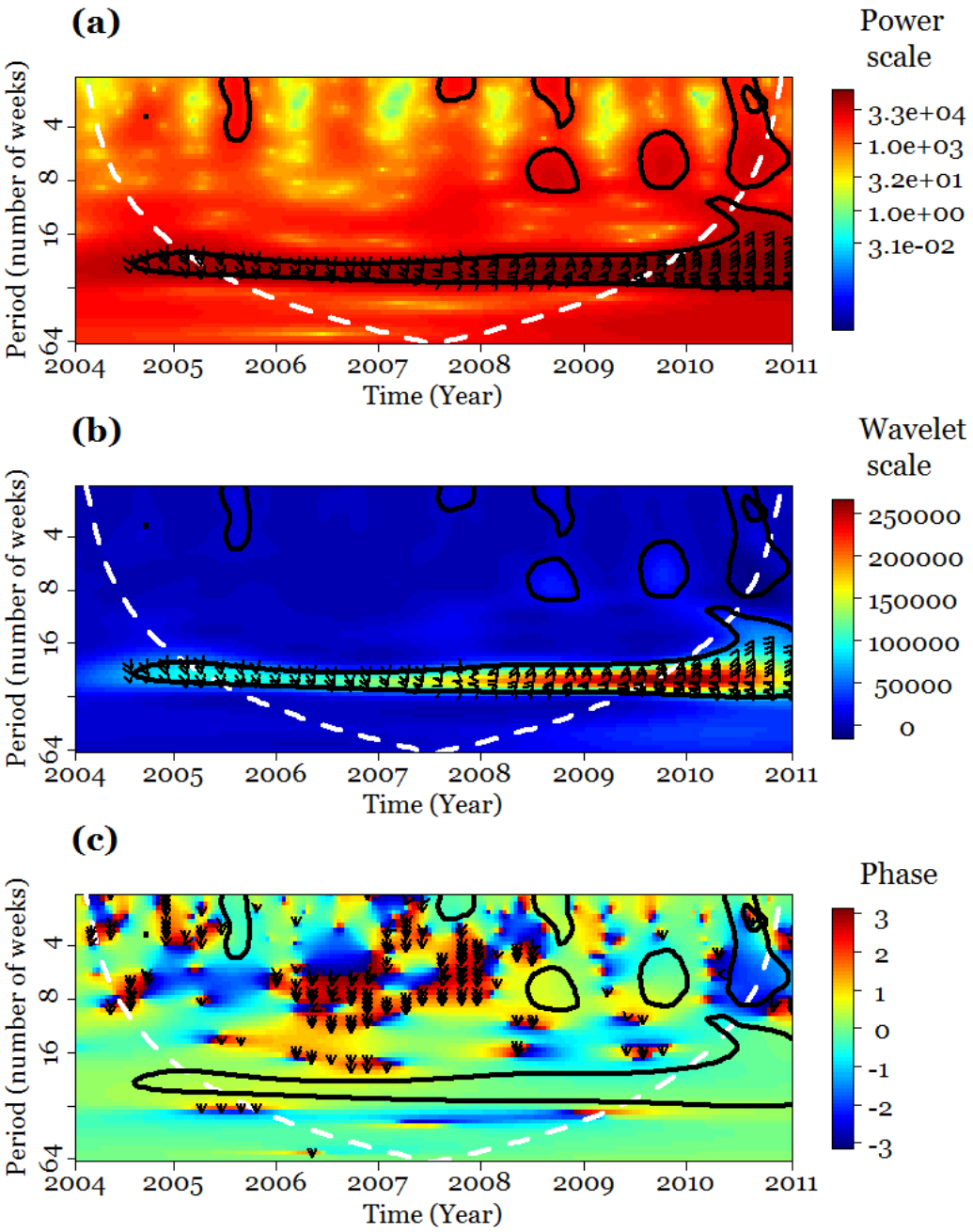


Figure A.7.15: Additional outputs related to the wavelet coherence between counts of *S. salar* and *S. trutta* for the whole observation period.
(a) power of the wavelet coherence, (b) scale of the wavelet and (d) phase.

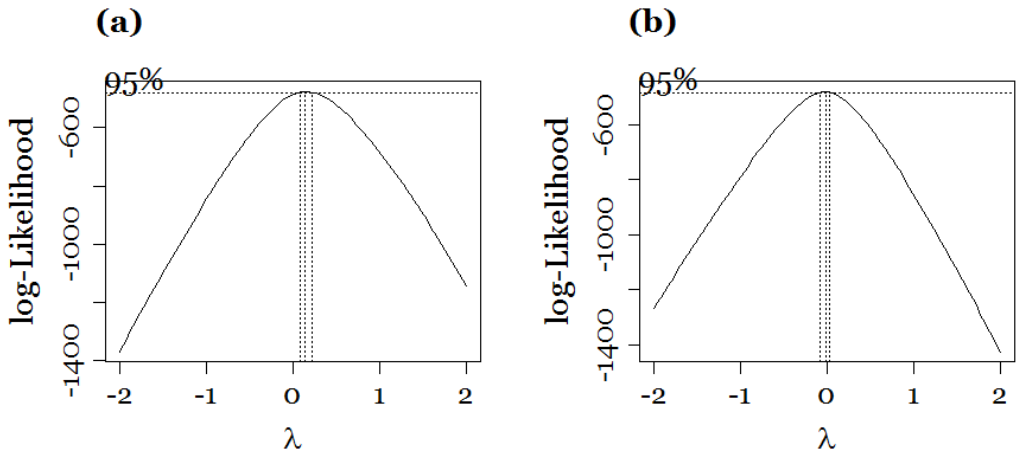


Figure A.7.16: The Log-likelihoods profile for the parameter λ of the Box-Cox transformation of counts of (a) *S. salar* and (b) *S. trutta*.

Table A.15: Summary of outputs from full LME model of count of *S. salar*.

Linear mixed-effects model fit by REML, containing all available parameters as fixed effects and year as random effect; 183 observations, 7 groups, AICc=1917.085, Log-likelihood=-942.5427

Random effects estimates:
1| year, Standard deviation intercept=13.5143, Residual=51.5157

Fixed effects estimates:

	Value	Std.Error	DF	*t*-value	*P*-value
Intercept	106.4316	27.2801	163	3.9014	0.0001
cos(p1)	-64.9074	19.1240	163	-3.3940	0.0009
sin(p1)	-65.3075	8.3605	163	-7.8114	0.0000
cos(p2)	9.6437	5.6770	163	1.6987	0.0913
sin(p2)	5.1413	5.9206	163	0.8684	0.3865
flow	-0.1492	0.1221	163	-1.2219	0.2235
temperature	-4.5299	2.5064	163	-1.8073	0.0726
NAO	11.8349	6.4722	163	1.8286	0.0693
NAO, lag=6m	-7.0406	6.5507	163	-1.0748	0.2841
NAO, lag=1yr	6.7540	6.4884	163	1.04094	0.2994
NAO, lag=2yr	-8.5473	6.7279	163	-1.2705	0.2057
NAO, lag=4yr	-5.3402	6.9292	163	-0.7707	0.4420
NAO, lag=5yr	3.0219	7.1814	163	0.4208	0.6745
tunnel	-42.5420	33.0309	163	-1.2879	0.1996

Standardized Within-Group Residuals:

Min	Q1	Med	Q3	Max
-1.6992	-0.5882	-0.0470	0.3961	6.1402

Table A.16: Summary of outputs from full LME model of count of *S. trutta*.

Linear mixed-effects model fit by REML, containing all available parameters as fixed effects and year as random effect; 183 observations, 7 groups, AICc=2141.43, Log-likelihood=-1054.715

Random effects estimates:
1 | year, Standard deviation intercept=25.1217, Residual=100.1389

Fixed effects estimates:

	Value	Std.Error	DF	*t*-value	*P*-value
Intercept	176.6288	52.63733	163	3.3555	0.0010
cos(p1)	-102.3189	36.9470	163	-2.7693	0.0063
sin(p1)	-138.5657	16.2013	163	-8.5528	0.0000
cos(p2)	-0.1876	11.0332	163	-0.0170	0.9865
sin(p2)	-12.4725	11.5076	163	-1.0838	0.2800
flow	-0.1791	0.2371	163	-0.7553	0.4512
temperature	-6.2223	4.8386	163	-1.2860	0.2003
NAO	14.6117	12.5627	163	1.1631	0.2465
NAO6m	6.8974	12.7240	163	0.5421	0.5885
NAO1yr	-12.3936	12.6079	163	-0.9830	0.3271
NAO2yr	-4.5704	13.0711	163	-0.3497	0.7270
NAO4yr	-0.2180	13.4652	163	-0.0162	0.9871
NAO5yr	-11.3392	13.9545	163	-0.8126	0.4176
tunnel	-141.2721	64.1374	163	-2.2026	0.0290

Standardized Within-Group Residuals:

Min	Q1	Med	Q3	Max
-2.3400	-0.6086	-0.0589	0.3579	3.7315

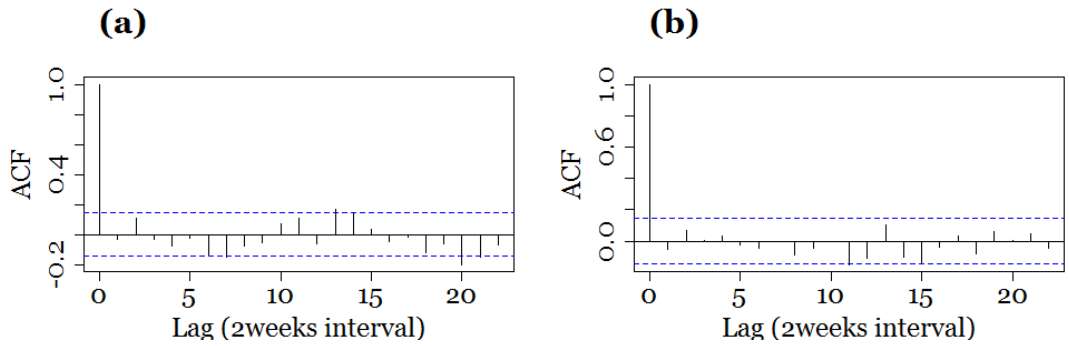


Figure A.7.17: The autocorrelation function of the residuals of the most parsimonious models of log-transformed counts of (a) *S. salar* and (b) *S. trutta*.

Figure A.7.18: The temporal fluctuations in the residuals of the most parsimonious GLS model of log-transformed counts of (a) *S. salar* and (b) *S. trutta*.

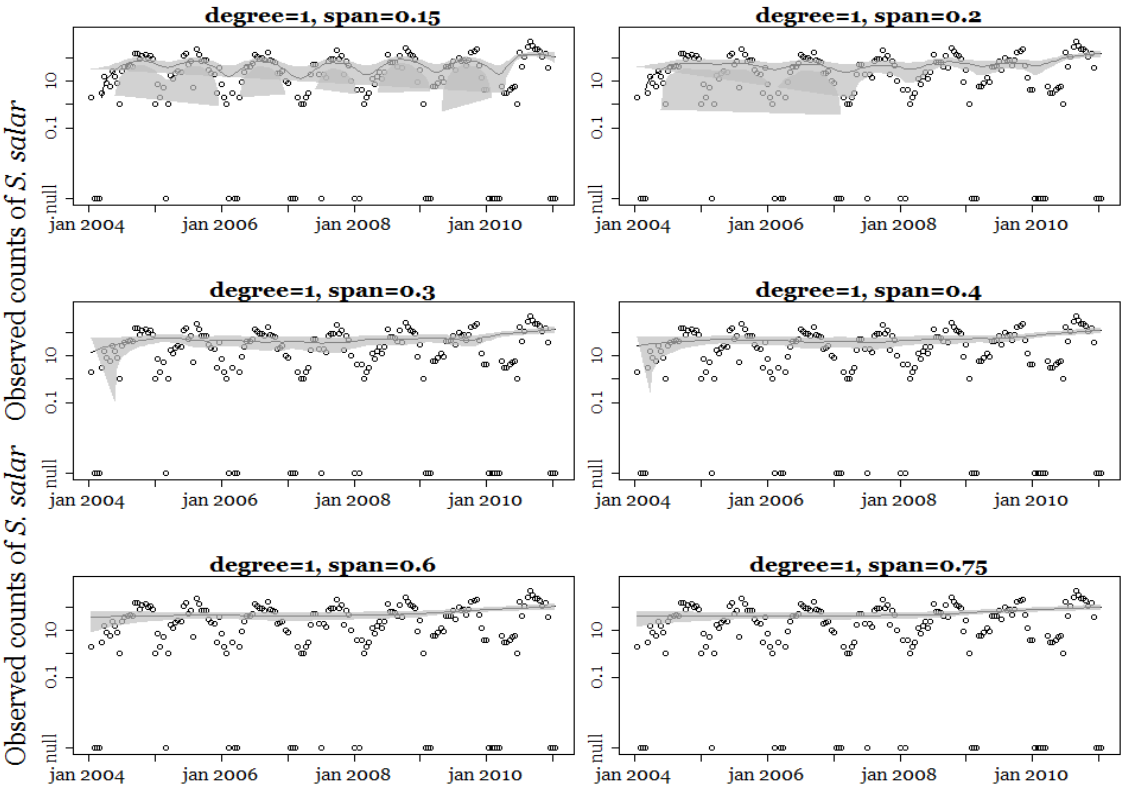


Figure A.7.19: Loess fit for counts of *S. salar*.

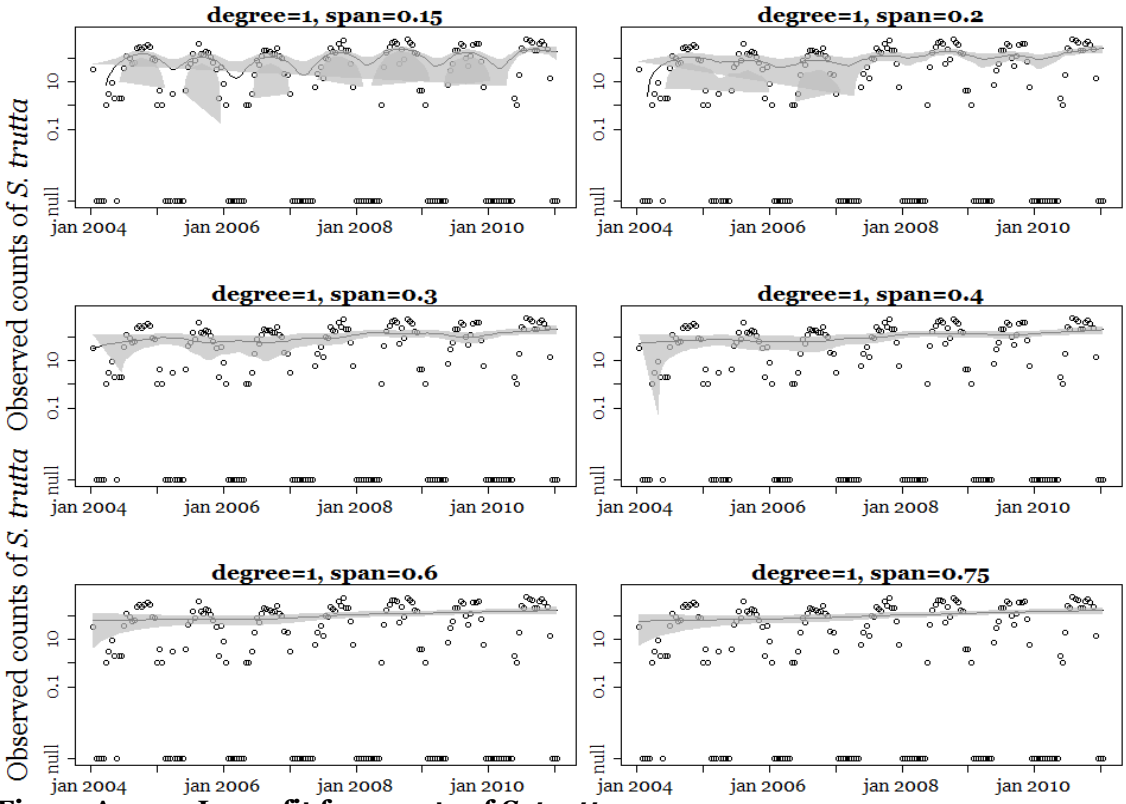


Figure A.7.20: Loess fit for counts of *S. trutta*.

Appendix A.iv

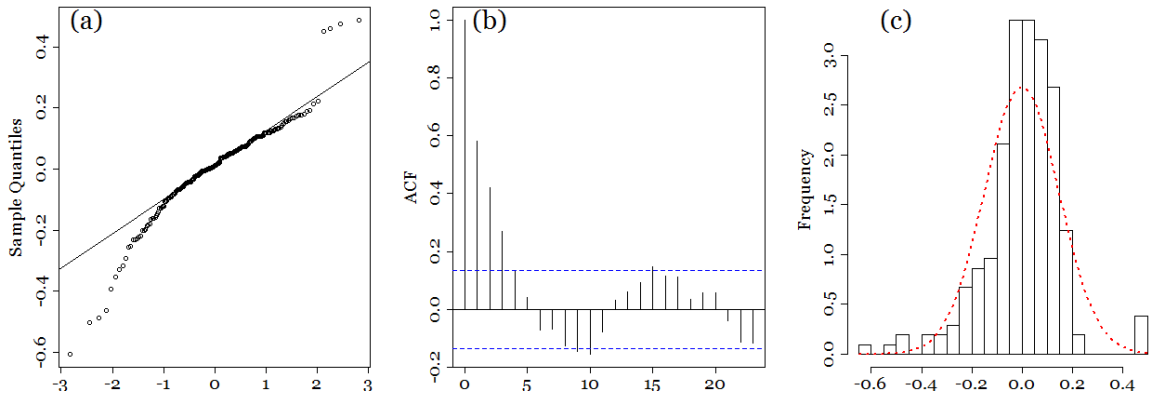


Figure A.7.21: Diagnostic plots of the residuals of the sum of the fitted values of the GLM of the proportion of identified counts and the GLM of the resulting residuals.
With (a) QQ-plot, (b) the autocorrelation function and (c) the frequency distribution, the normal distribution is represented by the red dotted line.

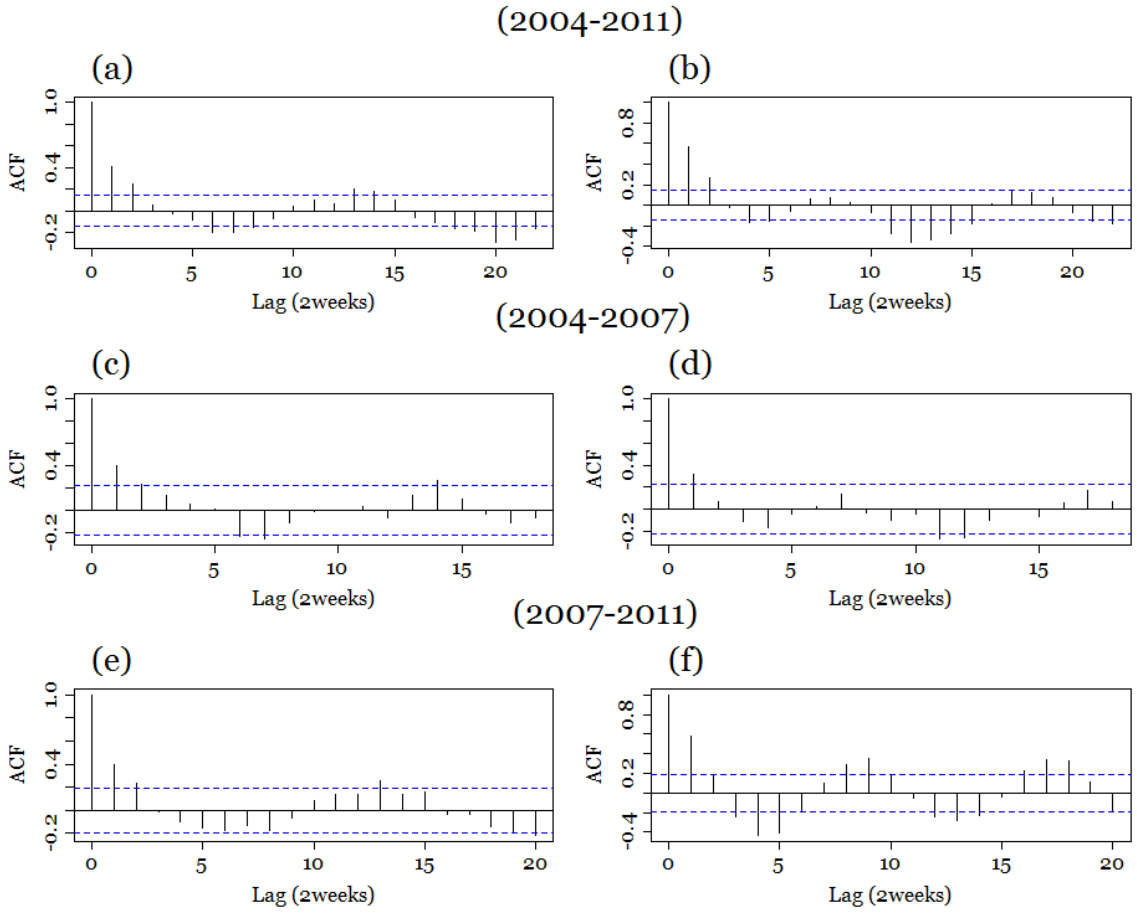


Figure A.7.22: The autocorrelation functions of the residuals of the most parsimonious GLS models, contrasting before and after 2007.
In *S. salar* (a, c, e) and *S. trutta* (b, d, f).

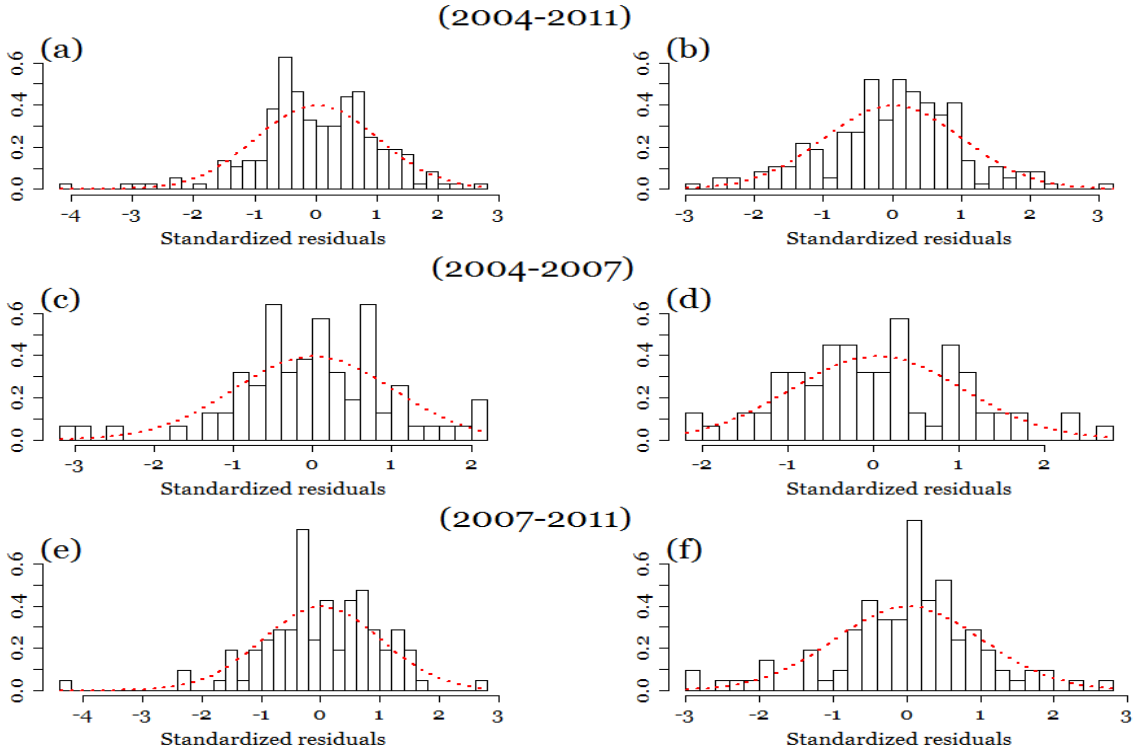


Figure A.7.23: The distribution of the residuals of the most parsimonious GLS models, contrasting before and after 2007.
In *S. salar* (a, c, e) and *S. trutta* (b, d, f). The red dotted line represents the normal distribution.

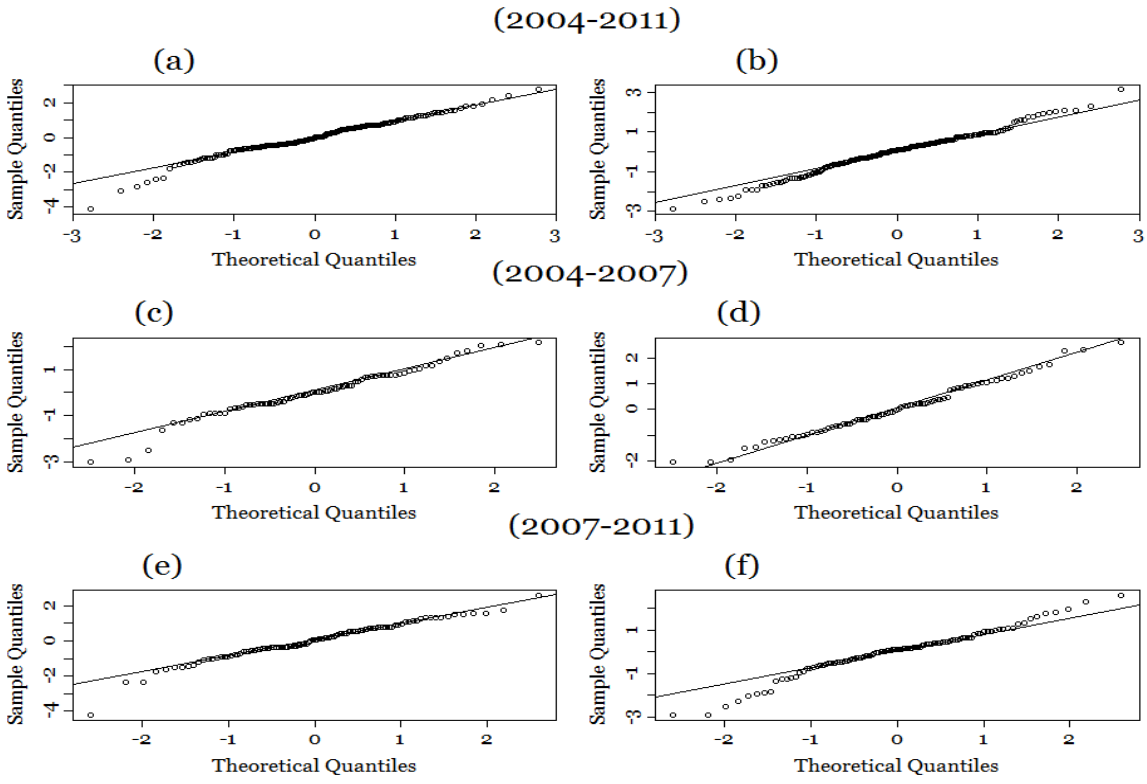


Figure A.7.24: The distribution of the residuals of the most parsimonious GLS models, contrasting before and after 2007.
QQ-plots, in *S. salar* (a, c, e) and *S. trutta* (b, d, f).

Appendix A.v

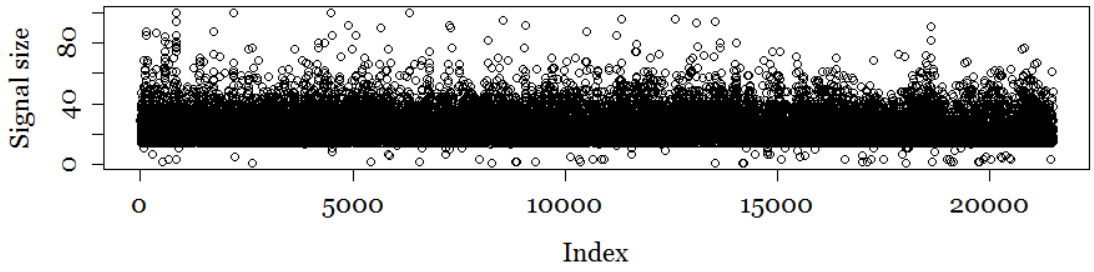


Figure A.7.25: The available signal amplitude data, both species confounded. There is a visible threshold at the amplitude of 26.

Table A.17: Summary of square CAT scores.
CAT scores for the first fourteen river parameters tested as potential classifiers between *S. salar* and *S. trutta*, ranked from highest to lowest value. Minimal shrinkage was required (λ=0.0014).

River feature	CAT score	River feature	CAT score
Flow (at Bywell)	464.77	Conductivity	49.55
Signal amplitude	365.82	Day of the year (linear)	48.43
Week of the year (linear)	165.57	Hour of the day (trigo)	47.54
Week of the year (trigo)	144.82	Tidal state	35.51
Hour of the day (linear)	130.48	Flow at Reaverhill	35.29
Turbidity at Riding Mill	67.45	Month of the year (linear)	34.09
pH at Riding Mill	63.50	Flow (at Ugly dub)	14.73

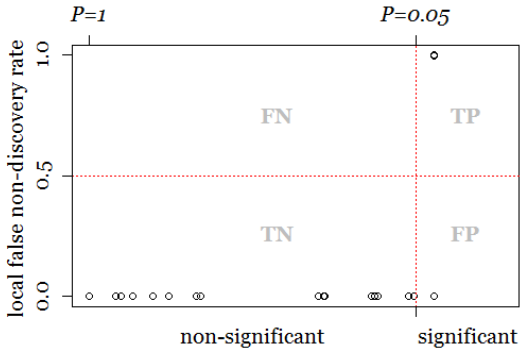


Figure A.7.26: The classification properties associated with each river parameters based on their local false non-discovery rate and associated *P*-value. False negatives (FN), true positives (TP), true negatives (TN) or false positives (FP) are generated by the parameter. Several points overlapping in TP (corresponding to flow, signal amplitude, week, hour).

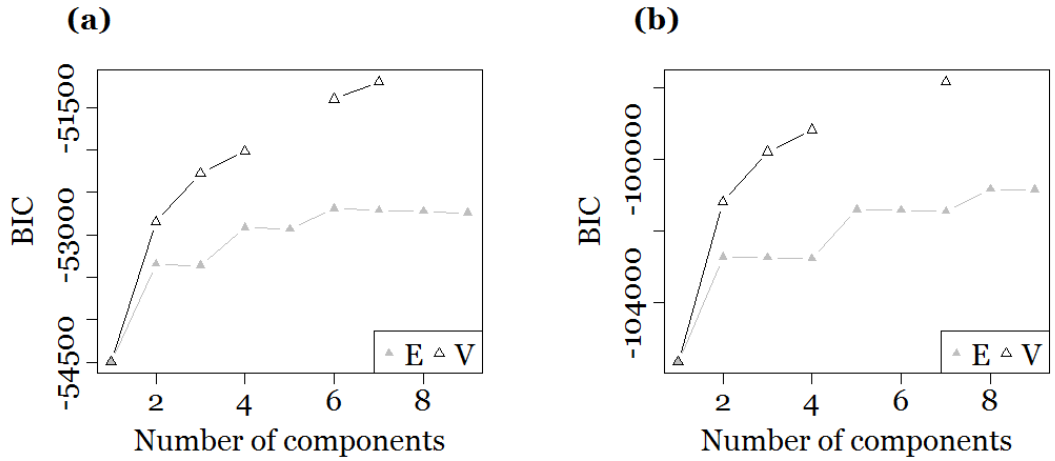


Figure A.7.27: The BIC values according to the number of clusters for equal (E) and variable (V) variances models, for (a) *S. salar* and (b) *S. trutta*.
The highest BIC value is for 7 clusters and variable variance, for both species.

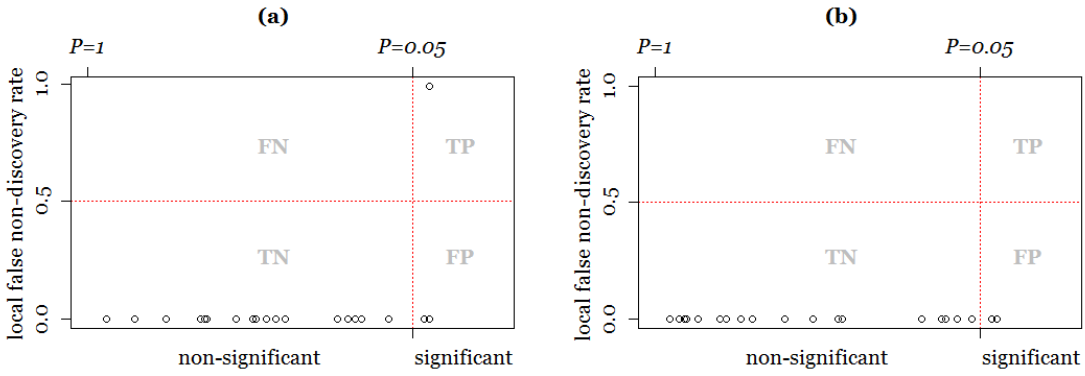


Figure A.7.28: The classification properties associated with each river parameters for the discrimination of sub-groups within (a) *S. salar* and (b) *S. trutta*.
Classification properties were based on their local false non-discovery rate and associated *P*-value. False negatives (FN), true positives (TP), true negatives (TN) or false positives (FP) were generated by each parameter.

Table A.18: The summary of square CAT scores for the first fourteen river parameters tested as potential classifiers between clusters identified within counts of *S. salar*, ranked from highest to lowest value.
Minimal shrinkage was required (λ= 0.0183).

River feature	CAT score	River feature	CAT score
Conductivity (EA)	117.76	Flow (at Reiver)	27.66
Conductivity (NW)	88.33	Hour (linear)	27.00
Flow (at Bywell)	66.24	Flow (at Otterburn)	22.69
Day (linear)	58.15	Temperature (at U. Dub)	13.49
Hour (trigo)	53.00	Month (linear)	11.80
Week (linear)	51.51	Day (trigo)	9.77
Temperature (at Bywell)	45.11	Flow (at U. Dub)	8.89

Table A.19 The summary of square CAT scores for the first fourteen river parameters tested as potential classifiers between clusters identified within counts of *S. trutta*, ranked from highest to lowest value.
Minimal shrinkage was required (λ= 0.0004).

River feature	CAT score	River feature	CAT score
Hour (trigo)	38.31	Turbidity (NW)	12.50
Conductivity (EA)	37.05	Temperature (at Bywell)	9.12
Flow (at Bywell)	27.68	Day (linear)	8.57
Hour (linear)	17.69	Temperature (at U. Dub)	8.06
Flow (at Ugly Dub)	14.60	Month (linear)	7.50
Week(linear)	13.94	Flow (at Otterburn)	7.32
Conductivity (NW)	13.37	Tide	6.50

Appendix A.vi Published collaborative work

Soto E., K. McGovern-Hopkins, R. Klinger-Bowen, B.K. Fox, J. Brook, N. Antonio, Z. van der Waal, S. Rushton, A. Mill and C.S. Tamaru. 2013. Prevalence of *Francisella noatunensis* subsp. *Orientialis* in cultured Tilapia on the Island of Oahu, Hawaii. *Journal of Aquatic Animal Health*, **25**: 104-109.

The author contributed to this paper by designing and undertaking the analysis, writing the description of the modelling approach in the Method section and producing the text and plots for the Results section (excluding Bacterial identification).

Journal of Aquatic Animal Health 25:104–109, 2013

ISSN: 0899-7659 print / 1548-8667 online
DOI: 10.1080/08997659.2013.781554

ARTICLE

Prevalence of *Francisella noatunensis* subsp. *orientalis* in Cultured Tilapia on the Island of Oahu, Hawaii

Esteban Soto*
Department of Pathobiology, School of Veterinary Medicine, Ross University, Post Office Box 334, Basseterre, St. Kitts, West Indies

Kathleen McGovern-Hopkins, Ruth Klinger-Bowen, and Bradley K. Fox
Department of Molecular Biosciences and Bioengineering, College of Tropical Agriculture and Human Resources, University of Hawaii–Manoa, Agricultural Science 218, 1955 East-West Road, Honolulu, Hawaii 96822, USA

James Brock and Nathene Antonio
Moana Technologies, LLC, Halawa Research Station, 99-1255 Waiua Place 4, Aiea, Hawaii 96701, USA

Zelda van der Waal, Stephen Rushton, and Aileen Mill
School of Biology, Newcastle University, Newcastle-upon-Tyne NE1 7RU, UK

Clyde S. Tamaru
Department of Molecular Biosciences and Bioengineering, College of Tropical Agriculture and Human Resources, University of Hawaii–Manoa, Agricultural Science 218, 1955 East-West Road, Honolulu, Hawaii 96822, USA

Abstract

Francisellosis is an emergent disease in cultured and wild aquatic animals. The causative agent, *Francisella noatunensis* subsp. *orientalis* (*Fno*), is a gram-negative bacterium recognized as one of the most virulent pathogens of warmwater fish. The main objective of this project was to investigate the prevalence of *Fno* in cultured tilapia (specifically, Mozambique Tilapia *Oreochromis mossambicus*, Koilapia [also known as Wami Tilapia] *O. hornorum*, Blue Tilapia *O. aureus*, and Nile Tilapia *O. niloticus* hybrids) on the island of Oahu, Hawaii, using conventional and real-time PCR assays followed by statistical modeling to compare the different diagnostic methods and identify potential risk factors. During 2010 and 2012, 827 fish were collected from different geographical locations throughout the island of Oahu. Upon collection of fish, the water temperature in the rearing system and the length of individual fish were measured. Extraction of DNA from different tissues collected aseptically during necropsy served as a template for molecular diagnosis. High correlation between both molecular methods was observed. Moreover, the bacterium was isolated from infected tilapia on selective media and confirmed to be *Fno* utilizing a species-specific Taqman-based real-time PCR assay. Although a direct comparison of the prevalence of *Fno* between the different geographical areas was not possible, the results indicate a high prevalence of *Fno* DNA in cultured tilapia throughout the farm sites located on Oahu. Of the different tilapia species and hybrids currently cultured in Hawaii, Mozambique Tilapia were more susceptible to infection than Koilapia. Water temperature in the rearing systems and fish size also had a strong effect on the predicted level of infection, with fish held at lower temperatures and smaller fish being more susceptible to piscine francisellosis.

*Corresponding author: esoto@rossvet.edu.kn
Received November 30, 2012; accepted February 22, 2013

Aquaculture on the islands of Hawaii has increased in the past two decades. In 2010, Hawaiian grown aquaculture products totaled US$30.0 million (NASS 2011). However, as in many agricultural practices, infectious diseases pose a substantial constraint to the potential of fish farming on the islands.

In Hawaiian cultured fish, an emergent pathogen was recognized in Mozambique Tilapia *Oreochromis mossambicus* and Blackchin Tilapia *Sarotherodon melanotheron* in wild and farmed populations on Oahu, Hawaii, in 1994 (Mauel et al. 2003). Isolation and complete characterization of the pathogen was not possible at the time, but the mortality events, clinical signs, and gross and histopathological findings were later found to be consistent with francisellosis in fish (Mauel et al. 2003; Birkbeck et al. 2011; Colquhoun and Duodu 2011). Francisellosis in fish is caused by members of the genus *Francisella*, namely *F. noatunensis* subsp. *noatunensis* and *F. noatunensis* subsp. *orientalis* (*Fno*; synonym, *F. asiatica*) (Birkbeck et al. 2011; Colquhoun and Duodu 2011). Recently, the use of molecular and serological methods has confirmed the involvement of *Fno* as the causative agent of mortality events of cultured fish on Oahu (Soto et al. 2012b).

Francisella noatunensis subsp. *orientalis* is an emergent pathogen of wild and cultured warmwater fish (Birkbeck et al. 2011; Colquhoun and Duodu 2011). During the past decade, the pathogen has been found as the causative agent of mortality events and disease outbreaks in many different geographical locations, including Asia, Europe, North and Central America, the Caribbean islands, and the Pacific islands including Hawaii (Birkbeck et al. 2011; Colquhoun and Duodu 2011; Soto et al. 2012b). Given the increasing importance of the tilapia industry in the Pacific islands and globally, the main objectives of this project were to study the prevalence of *Fno* in cultured tilapia *Oreochromis* spp. on the island of Oahu using molecular diagnostic methods (end-point and real-time PCR) and to use statistical modeling to investigate potential risk factors for disease.

METHODS

Fish collection.—A total of 827 cultured fish were collected from December 2010 to August 2011 from 20 different locations on the island of Oahu, Hawaii, specifically from the districts of Manoa, Waimanalo, Kaneohe, Mililani, Waianae, Kunia, Aiea, Kaimuki, Kailua, and Honolulu (Figure 1). Different species of tilapia were cultured in a variety of freshwater systems that can be characterized as being closed recirculating, static, or flow through. The different species of tilapia and their hybrids included Mozambique Tilapia *Oreochromis mossambicus*, Koilapia (also known as Wami Tilapia) *O. hornorum*, Blue Tilapia *O. aureus*, and Nile Tilapia *O. niloticus* hybrids. Fish were collected and total length, culture method, and water temperature of the rearing system were recorded. Fish were euthanized with an overdose of 2-Phenoxyethanol (Sigma) and analyzed for external and internal clinical signs of disease.

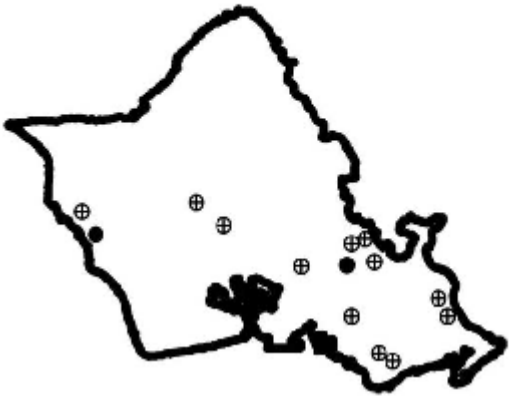


FIGURE 1. Distribution of aquaculture farms sampled on Oahu during the project. Open circles with a cross are positive for *Fno* and solid black circles are negative for *Fno*.

At this point, clinical signs consistent with previous reports of francisellosis in tilapia (splenomegaly with multifocal white nodular lesions in the spleen) and possession of granulomas in splenic tissue were recorded. After a complete necropsy, 552 spleen samples and 235 gill samples were collected and stored in ethanol until DNA extractions were performed. Due to the cost of the molecular diagnosis, 127 splenic samples were preserved individually, while 425 were pooled in groups of two to five fish per pool. Similarly, 179 gill samples were processed as pools of four and five fish per pool, while 56 were processed individually.

Bacterial isolation.—Splenic tissue obtained from cultured Mozambique Tilapia ($n = 12$) and Koilapia ($n = 8$) presenting similar mortality events and clinical findings as previously reported (Soto et al. 2009) were aseptically collected at the University of Hawaii–Manoa. The tissues were individually placed in sterile 1.5-mL microcentrifuge tubes and sent in ice to the Ross University–School of Veterinary Medicine Marine Laboratory for further analysis. Fish tissues were used for bacteriological analysis by streaking on Modified Thayer-Martin Agar, (BD BBL, Sparks, Maryland) following previous reports (Soto et al. 2009) and were incubated for 5 d at 27°C. Colonies observed from primary isolation agar plates were replated for purity of culture and were frozen following previously reported methods (Soto et al. 2009).

Molecular diagnosis.—Extraction of DNA from collected tissues was performed using a Roche High Pure PCR Template Kit (Roche Applied Science, Indianapolis, Indiana), following protocol suggested by the manufacturer in a laminar flow biosafety cabinet. All isolates recovered were used for molecular analysis. A loop of the bacterium was suspended in 500 µL of sterile 1 × PBS and subjected to DNA extraction using the DNeasy Blood and Tissue Kit (Qiagen, Valencia, California)

TABLE 1. Primers and probes utilized in the study.

Primer or probe	PCR method	Sequence (5′–3′)	Reference
Primer			
F11	Conventional	taccagttggaaacgactgt	Forsman et al. (1994)
F5	Conventional	ccttttttgagtttcgctcc	Forsman et al. (1994)
iglC forward	Real-time	gggcgtatctaaggatggtatgag	Soto et al. (2010)
iglC reverse	Real-time	agcacagcatacaggcaagcta	Soto et al. (2010)
Probe			
iglC probe	Real-time	FAM atctattgatgggctcacaacttcacaa BHQ-1	Soto et al. (2010)

following the manufacturer's suggested protocol for gram-negative bacteria. Extracted DNA was stored at –20 C until further use.

Conventional PCR and real-time PCR analysis were performed following previously published methods. The primers and probes have been previously validated and were found to be specific for the detection of members of the genus *Francisella* (Forsman et al. 1994) and for *Fno* (Soto et al. 2010) (Table 1).

Statistical analysis.—The level of infection by *Fno* was determined using PCR and real-time PCR performed on splenic or gill tissues from individual or pooled tissues at each location. Samples obtained varied with location and fish species. When the spleen was sampled, the presence or absence of white nodular lesions in the spleen (granulomas) was recorded. Both PCR and real-time PCR variables were expressed as a binomial response (positive or negative) and are related to the Cycle Threshold (CT) value. The CT value was defined as the value 40-CT for which the values beyond 40 were considered null; due to the inverse relationship between DNA quantity and CT, higher 40-CT values indicated increased bacterial burdens.

A generalized linear modeling (GLM) approach was used to investigate the relationship between the level of infection in fish, species, fish length, and water temperature. The presence of granulomas and the presence of bacterial DNA as measured by PCR and real-time PCR were modeled using a binomial error structure. Infection as recorded by the 40-CT was modeled as a simple GLM with normal errors.

Full models containing all covariates available were run and a parsimonious model was identified using step-wise deletion of nonsignificant parameters. Model fit was assessed using an information theoretic approach (Burnham and Anderson 2002). Underlying model assumptions were tested visually using scatter plots of variance residuals, which attribute a measure of the deviance from the fit of the model contributed by each observation. The distribution of residuals was tested for normality and independence to all parameters. The area under the curve of the receiver operating characteristics (ROC) function was used to assess the overall accuracy of the models (Fielding and Bell 1997). All models were analyzed using the software R.2.15.1 with the packages MASS (Venables and Ripley 2002) and statistics (R Development Core Team 2012).

RESULTS

Bacterial Identification

Eight *Fno* isolates were recovered from splenic tissues of Koilapia and Mozambique Tilapia submitted for bacterial isolation. Seven isolates were recovered from Mozambique Tilapia and one from Koilapia. Growth of *Fno* was visible on the agar media, 48–72 h post inoculation and colonies were gray, smooth, and convex. Additionally, all of them were positively identified as *Fno* utilizing a real-time PCR assay for the detection of the *Fno iglC* gene (Soto et al. 2010).

Data Exploration

Detection of *Fno* DNA by conventional and real-time PCR were strongly correlated (Pearson correlation coefficient = 0.8). The values for the area under the ROC curves were 0.848 and 0.846, respectively for PCR and real-time PCR. This implied that in both cases, over 84% of the time, a random selection from the positive group will have a score greater than a random selection from the negative class (Fielding and Bell 1997). Both diagnostic methods consequently gave high and similar levels of general accuracy.

Koilapia and Mozambique Tilapia accounted for approximately 86% of the total number of samples analyzed in the project. Over 93% of fish samples originated from 8 of the 20 farms sampled during the project, so a geographical comparison was not possible. Since not all tissues were analyzed for all the fish, there was insufficient data to analyze the effects of tissue type on the level of infection. The use of temperature and fish length as independent variables was validated by the nonsignificant correlation between them (df = 179, $P = 0.17985$). There was also no significant difference between species in terms of size or the temperatures at which they were reared ($F = 1.1284$, df = 145, $P = 0.4636$ for fish length and $F = 0.7523$, df = 137, $P = 0.09597$ for temperature).

PCR and Real-Time PCR Models

Temperature and fish length were identified as significant risk factors of infection as estimated by PCR and real-time PCR diagnosis. Infection was more likely in colder water (–0.38, df = 179, $P < 0.001$ for PCR and –0.36, df = 179, $P < 0.001$for real-time PCR) and in smaller fish (–0.13, df = 179, $P < 0.001$

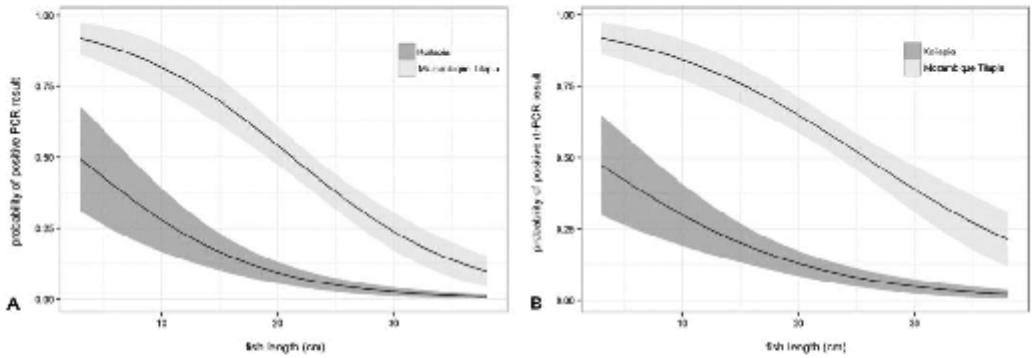


FIGURE 2. Prediction of the probability of a positive (A) conventional PCR or (B) real-time PCR result for increasing values of temperature for the Koilapia and Mozambique Tilapia species culture in Hawai (predicted mean ± 95% CI).

for PCR and –0.11, df = 179, $P < 0.001$ for real-time PCR). From the two principal cultured tilapia species, Mozambique Tilapia were more prone to infection than Koilapia (2.45, df = 179, $P < 0.001$ for PCR and 2.53, df = 179, $P < 0.001$ for real-time PCR). Curves showing predicted probability of infection for each species in relation to temperature and body size are illustrated on Figure 2 and Figure 3.

Threshold Cycle CT Value Model

A lower level of infection was predicted by the model in conditions of higher temperature (–1.19, df = 179, $P < 0.001$) and in larger fish (–0.31, df = 179, $P < 0.001$). Mozambique Tilapia appeared prone to higher infectious level than Koilapia (5.61, df = 179, $P < 0.001$). The model predictions of the value CT-40 (increasing with the level of infection) according to values of temperature and fish length are illustrated in Figure 4 for both tilapia species.

Presence of Granulomas in Spleen

The presence of granulomas in the spleen samples was related to species, with Mozambique Tilapia more likely to display granulomas than Koilapia (3.49, df = 98, $P < 0.001$). Granulomas also appeared more likely to be present during high infective events (identified for lower CT values [0.17, df = 98, $P < 0.05$] when analyzed by real-time PCR).

DISCUSSION

Due to the increasing importance of disease outbreaks in the aquaculture industry and the impacts disease can have on both feral and farmed fish, monitoring and surveillance of diseases have been considered to be of socioeconomic importance in

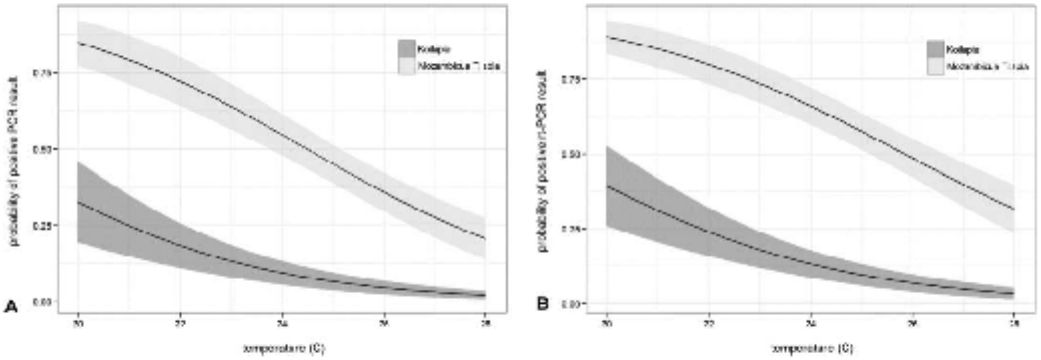


FIGURE 3. Prediction of the probability of positive (A) conventional PCR or (B) real-time PCR results for increasing values of fish length for the Koilapia and Mozambique Tilapia species culture in Hawai (predicted mean ± 95% CI).

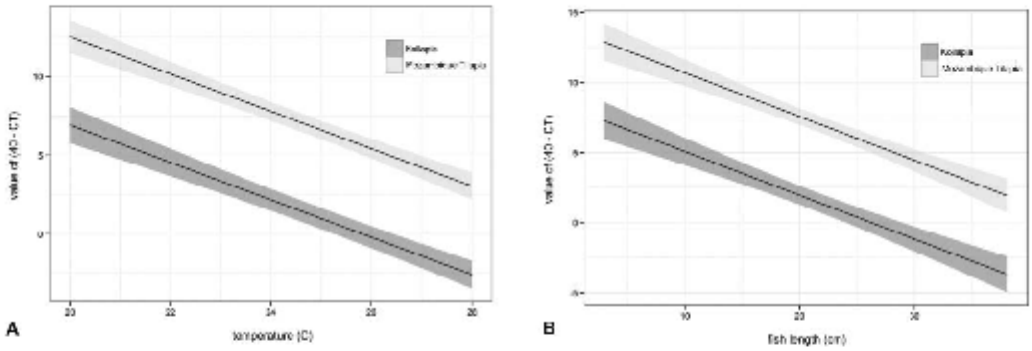


FIGURE 4. Prediction of the CT value (40-CT) for increasing values of (A) temperature and (B) fish length for the Koilapia and Mozambique Tilapia species (predicted mean ± 95% CI).

many regions of the world (USFWS and AFS-FHS 2005) for many years. Since the early 1990s, different tilapia species, including Mozambique Tilapia and Blackchin Tilapia, in wild and farmed populations on Oahu have been affected by what was recognized as a rickettsia-like organism or piscirickettsiosis-like syndrome, not diagnosed until recent years as francisellosis, caused by the emergent pathogen *Fno* (Mauel et al. 2003; Soto et al. 2012b). Significant losses experienced in the early 1990s resulted in restricted movement of tilapia between Oahu and the rest of the Hawaiian Islands (PQ Policy 98-09 Section 150A-8, HRS, November 5, 1998) and this restriction remains in effect. Due to the rapidly expanding fish aquaculture industry in Hawaii, there is a need to develop and incorporate rapid, sensitive, accurate, and cost effective methods for identifying specific pathogens to aid in managing the disease threat. Molecular diagnostic methodologies are considered to be the state of the art with regard to detection and surveillance of pathogens and for these reasons are being investigated for use in monitoring for the presence of *Fno* in cultured and wild tilapia on and around the island of Oahu. The results from these activities will provide the necessary evidence that decision makers and stakeholders need in order to make informed decisions on the necessary steps to be taken for the continued expansion and diversification of this sector of Hawaii's aquaculture industry.

The use of conventional and real-time PCR were chosen over conventional bacteriological methods (agar and broth culture) since molecular diagnostic methods enhance the diagnosis of this fastidious organism that could be present at low levels in fish tissue, requires specialized media to grow, and may possibly be overgrown by secondary contaminants following attempts at primary isolation (Soto et al. 2009). However, as with many other diagnostic techniques, the molecular technologies like conventional PCR also have some limitations. The method is time consuming, results are based on band-size discrimination, results are measured at End-Point (plateau), and there is often lower sensitivity, lower resolution, and no quantification when compared to real-time PCR. For these reasons, real-time PCR was also utilized in the project, since it collects data in the exponential phase of the amplification process, has an increased dynamic range of detection, and has reduced time in post-PCR processing (Soto et al. 2010). The disadvantages of real-time PCR include the cost of the assay, which is substantially higher than that of either microscopy or conventional PCR, and the need for specialized real-time PCR analyzers, which are currently beyond the means of many laboratories. As indicated by the results obtained in the present study, it is confirmed by isolation and molecular methods that *Fno* is prevalent in cultured tilapia obtained from several geographical locations throughout the island of Oahu. Moreover, Mozambique Tilapia cocultured with Koilapia appeared more prone to infection by *Fno*, and to develop granulomas in the spleen (classical finding for francisellosis in fish), than the Koilapia species. Water temperature and fish size also had a strong effect on the predicted level of infection, with lower temperatures and smaller tilapia being more susceptible to infection of *Fno*. This finding needs further investigation and confirmation in laboratory confined experimental challenges, but it highlights the importance of the development of efficacious vaccines or other prophylactic methodologies given at an early stage of production. Similar findings were reported by Mauel et al. (2003), where cultured and wild tilapia mortalities on the island of Oahu occurred only during the cooler months of the year and were not recorded during the warmer months. Recent research has demonstrated that temperature significantly influenced the development of francisellosis in tilapia. Fish maintained at 25°C developed francisellosis and had considerably higher mortality and splenic bacterial concentrations than control fish and fish maintained at 30°C (Soto et al. 2012a). These findings are consistent with the current investigation. Since water temperature is well known to play a critical role in the development of both specific and nonspecific

immunity in cold-blooded animals, such as fish, further research in regards to the immune response of tilapia at different temperatures is necessary (Collins et al. 1976; Hrubec et al. 1996; Watts et al. 2001).

When and how *Fno* reached Hawaiian waters is currently unknown. Interestingly, wild populations of Blackchin Tilapia obtained from several coastal systems were also found to be infected with *Fno*, indicating that not only cultured fish populations are affected by this emergent pathogen and investigation of local fish populations are clearly warranted (authors, unpublished). This is particularly true for Blackchin Tilapia, which is one of the most harmful introduced fish in low-elevation areas of Hawaiian streams, wetlands, and estuaries, and likely has caused more negative ecosystem effects in Hawaiian streams and wetlands than any other introduced aquatic species (Englund et al. 2000). Blackchin Tilapia, first introduced into Hawaii in 1951 for tuna bait and weed control (Randall 1987), is now the dominant estuarine fish species on all the major Hawaiian islands. Not only do introductions impact native populations through predation, competition, and degradation of habitat, but they also can be vectors for transferring parasites to native fish populations (Font 2007). Surveillance of native fish populations that could also be affected by this bacterium has been recommended, and it is being undertaken. Moreover, due to fish movements from Oahu to other Hawaiian islands, the prevalence of *Fno* in other Hawaiian and Pacific islands should be investigated, and if not present, regulation of fish movement should be enforced to avoid the spread of this pathogen to new geographical locations. These activities will all become part of the development of a disease management program for this specific pathogen that should result in an overall improvement in the tilapia production in Hawaii.

ACKNOWLEDGMENTS

The authors kindly thank the Center for Tropical and Subtropical Agriculture and the National Institute of Food and Agriculture of the U.S. Department of Agriculture for their support under Award Numbers 2007-38500-18471, 2008-38500-19435, and 2010-38500-20948. Partial support was also provided by U.S. Department of Agriculture National Institute of Food and Agriculture, Smith-Lever funds for Cooperative Extension Project ID 12–506, Strengthen aquaculture research and extension at CTAHR.

REFERENCES

Birkbeck, T. H., S. W. Feist, and D. W. Verner-Jeffreys. 2011. *Francisella* infections in fish and shellfish. Journal of Fish Diseases 34: 173–187.

Burnham, K. P., and D. R. Anderson. 2002. Model selection and multimodel inference: a practical information-theoretic approach, 2nd edition. Springer, New York.

Collins, M. T., D. L. Dawe, and J. B. Gratzek. 1976. Immune response of Channel Catfish under different environmental conditions. Journal of the American Veterinary Medical Association 169:991–994.

Colquhoun, D. J., and S. Duodu. 2011. *Francisella* infections in farmed and wild aquatic organisms. Veterinary Research 42:47–63.

Englund, R. A., K. Arakaki, D. J. Preston, S. L. Coles, and L. G. Eldredge. 2000. Nonindigenous freshwater and estuarine species introductions and their potential to affect sportfishing in the lower stream and estuarine regions of the south and west shores of Oahu, Hawaii. Final Report to the Hawaii Department of Land and Natural Resources, Hawaii Biological Survey, Bishop Museum Technical Report 17, Honolulu.

Fielding, A. H., and J. F. Bell. 1997. A review of methods for the assessment of prediction errors in conservation presence/absence models. Environmental Conservation 24:38–49.

Font, W. F. 2007. Parasites of Hawaiian stream fishes: sources and impacts. Pages 157–169 *in* N. L. Evenhuis and J. M. Fitzsimons, editors. Biology of Hawaiian streams and estuaries: proceedings of the symposium on the biology of Hawaiian streams and estuaries, April 2005. Bishop Museum Bulletin in Cultural and Environmental Studies 3, Bishop Museum Press, Honolulu.

Forsman, M., G. Sandström, and A. Sjöstedt. 1994. Analysis of 16S ribosomal DNA sequences of *Francisella* strains and utilization for determination of the phylogeny of the genus and for identification of strains by PCR. International Journal of Systematic Bacteriology 44:38–46.

Hrubec, T. C., J. L. Robertson, S. A. Smith, and M. K. Tinker. 1996. The effect of temperature and water quality on antibody response to *Aeromonas salmonicida* in sunshine bass (*Morone chrysops* × *Morone saxatilis*). Veterinary Immunology and Immunopathology 50:157–166.

Mauel, M. J., D. L. Miller, K. Frazier, A. D. Liggett, L. Styer, D. Montgomery-Brock, and J. Brock. 2003. Characterization of a piscirickettsiosis-like disease in Hawaiian tilapia. Diseases of Aquatic Organisms 53:249–255.

NASS (National Agricultural Statistics Service). 2011. Hawaii aquaculture. U.S. Department of Agriculture, NASS, Washington, D.C. Available: www.nass.usda.gov/Statistics_by_State/Hawaii/index.asp. (November 2012).

R Development Core Team. 2012. R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna. Available: www.R-project.org/.(November 2012).

Randall, J. E. 1987. Introductions of marine fishes to the Hawaiian Islands. Bulletin of Marine Science 41:490–502.

Soto, E., S. B. Abrams, and F. Revan. 2012a. Effects of temperature and salt concentration on *Francisella noatunensis* subsp. *orientalis* infections in Nile Tilapia *Oreochromis niloticus*. Diseases of Aquatic Organisms 101:217–223.

Soto, E., K. Bowles, D. Fernandez, and J. P. Hawke. 2010. Development of a real-time PCR assay for identification and quantification of the fish pathogen *Francisella noatunensis* subsp. *orientalis*. Diseases of Aquatic Organisms 89:199–207.

Soto, E., J. P. Hawke, D. Fernandez, and J. A. Morales. 2009. *Francisella* sp., an emerging pathogen of tilapia, *Oreochromis niloticus* (L.), in Costa Rica. Journal of Fish Diseases 32:713–722.

Soto, E., O. Illanes, D. Hilchie, J. A. Morales, P. Sunyakumthorn, J. P. Hawke, A. E. Goodwin, A. Riggs, R. P. Yanong, D. B. Pouder, R. Francis-Floyd, M. Arauz, L. Bogdanovic, and F. Castillo-Alcala. 2012b. Molecular and immunohistochemical diagnosis of *Francisella noatunensis* subsp. *orientalis* from formalin-fixed, paraffin-embedded tissues. Journal of Veterinary Diagnostic Investigation 24:840–845.

USFWS and AFS-FHS (U.S. Fish and Wildlife Service and American Fisheries Society–Fish Health Section). 2005. Standard procedures for aquatic animal health inspections. *In* AFS-FHS. FHS blue book: suggested procedures for the detection and identification of certain finfish and shellfish pathogens, 2005 edition. AFS-FHS, Bethesda, Maryland.

Venables, W. N., and B. D. Ripley. 2002. Modern applied statistics with S, 4th edition. Springer Science, New York.

Watts, M., B. L. Munday, and C. M. Burke. 2001. Immune responses of teleost fish. Australian Veterinary Journal 79:570–574.

References. .

References.

Adamson M.W. 2004. *Food through history: food in medieval times*. Westport, CT, USA : Greenwood Publishing Group.

Agarwal D.K., A.E. Gelfand and S. Citron-Pousty. 2002. Zero-inflated models with applications to spatial count data. *Environmental and Ecological Statistics*, **9**: 341-355.

Ahdesmaki M. and Strimmer. 2010. Feature selection in omics prediction problems using CAT scores and false nondiscovery rate control. *The annals of Applied Statistics*, **4** (1): 503-519.

Ahdesmaki M., V. Zuber and K. Strimmer. 2012. Sda: Shrinkage Discriminant Analysis and CAT Scores Variable Selection. R Package version 1.2.3. Available at <http://CRAN.R-project.org/package=sda>.

Alagaraja, K. 2011. Simple methods for estimation of parameters for assessing exploited fish stocks. *Indian Journal of Fisheries*. **31** (2): 177-208.

Anand M., A. Gonzalez, F. Guichard, J. Kolasa and L. Parrott. 2010. Ecological systems as complex systems: challenges for an emerging science. *Diversity*, **2**: 395-410.

Anderson C.N.K., C. Hsieh, S.A. Sandin, R. Hewitt, A. Hollowed, J. Beddington, R.M. May and G. Sugihara. 2008. Why fishing magnifies fluctuations in fish abundance. *Nature,* **452**: 835-839.

Anderson D.H. 1996. The straddling stocks agreement of 1995: an initial assessment. *The International and Comparative Law Quaterly*, **45** (2): 463-475.

Anderson W.G., R. Booth, T.A. Beddow, R.S. McKinley, B. Finstad, F. Okland and D. Scruton. 1998. Remote monitoring of heart rate as a measure of recovery in angled Atlantic salmon, *Salmo salar* (L.). *Hydrobiologia*, **371-372**: 233-240.

Andrews A.C. 1955. Greek and Latin terms for Salmon and Trout. *Transactions and Proceedings of the American Philological Association*, **86**: 308-318.

Anonymous. 1930. Sea Trout from the Moray Firth. *Nature*, **314** (125): 218-219.

Aoki M. 1987. State Space Modeling of Time Series. Berlin, Germany: Springer-Verlag.

Aprahamian M.W., P. Hickley, B.A. Shields and G.W. Mawle. 2010. Examining changes in participation in recreational fisheries in England and Wales. *Fisheries Management and Ecology*, **17**: 93-105.

Archer D., *P.* Rippon, R. Inverarity and R. Merrix. 2008. The role of regulatory releases and natural spates on salmonid migration in the River Tyne North-east England. *Proceedings of the 10th BHS National Hydrology Symposium (Exeter, 15–18 September 2008):* 439–444.

Archer D.R. 2003. *Tyne and Tide: a Celebration of the River Tyne*. Northumberland, UK: Daryan Press.

Armstrong J.D. and K.H. Nislow. 2006. Critical habitat during the transition from maternal provisioning in freshwater fish, with emphasis on Atlantic salmon (*Salmo salar*)and brown trout (*Salmo trutta*). *Journal of Zoology*, **269** (4): 403-413.

Armstong J.D., P.S. Kemp, G.J.A Kennedy, M. Ladle and N.J. Milner. 2003. Habitat requirements of Atlantic salmon and brown trout in rivers and streams. *Fisheries Research*, **62** (2): 143-170.

Armstrong J.D., J.W.A. Grant, H.L. Forgsen, K.D. Fausch, R.M DeGraaf, I.A. Fleming, T.D. Prowse and I.J Schlosser. 1998. The application of science to the management of Atlantic salmon (*Salmo salar*): integration across scales. *Canadian Journal of Fisheries and Aquatic Sciences*, **55** (1): 303-311.

Attrill M.J. and M. Power. 2002. Climatic influence on a marine fish assemblage. *Nature*, **417**: 275-278.

Atug F., E.P. Castle, S.K. Srivastav, S.V. Burgess, R. Thomas, R. Davis. 2006. Positive surgical margins in robotic-assisted radical prostatectomy: impact of learning curve on oncologic outcomes. *European Urology*, **49** (5): 866-872.

Aukema B.H. and *K.F.* Raffa. 2004. Does aggregation benefit bark beetles by diluting predation? Links between a group-colonisation strategy and the absence of emergent multiple predator effects. *Ecological Entomology,* **29**: 129–138.

Bakke T.A. and P.D. Harris. 1998. Diseases and parasites in wild Atlantic salmon (*Salmo salar*) populations. *Canadian Journal of Fish and Aquatic Sciences*, **55** (S1): 247-266.

Bardonnet A. and J.-L. Bagliniere. 2000. Freshwater habitat of Atlantic salmon (*Salmo salar*). *Canadian Journal of Fisheries and Aquatic Sciences*, **57**: 497-506.

Bates D. 2007. [R] Convergence error code in mixed effects models. R-help list, run by maechler@stat.math.ethz.ch and bates@r-project.org. Available at <https://stat.ethz.ch/pipermail/r-help/2007-December/148391.html>. (Accessed February 2013).

Bazarov M.I. and V.K. Golovanov. 2003. Migratory behaviour of Atlantic salmon and sea trout in the middle Kumijoki River, southern Finland (based on data from biotelemetry tracking). Atlantic salmon: biology, conservation and restoration. Petrozavodsk , Russia.

Beaugrand G. and R.R. Kirby. 2010. Climate, plankton and cod. *Global Change Biology*, **16**: 1268-1280.

Beaugrand G., M. Edwards, K. Brander, C. Luczak and F. Ibanez. 2008. Causes and projections of abrupt climate-driven ecosystem shifts in the North Atlantic. *Ecology Letters*, **11**: 1157-1168.

Beaugrand G. 2004. The North Sea regime shift: evidence, causes, mechanisms and consequences. *Progress in Oceanography*, **60**: 245-262.

Beaugrand G., K.M. Brander, J.A. Lindley, S. Souissi and P.C. Reid. 2003. Plankton effect on cod recruitment in the North Sea. *Nature*, **426**: 661-664.

Beaugrand G. and P. Reid. 2003. Long-term changes in phytoplankton, zooplankton and salmon related to climate. Global Change Biology, **9**: 801-817.

Bekkevold D., M.H. Hansen and K.-L.D. Mensberg. 2004. Genetic detection of sex-specific dispersal in historical and contemporary populations of anadromous brown trout *Salmo trutta*. *Molecular Ecology*, **13**: 1707-1712.

Benhke R. And J.R. Tomelleri. 2002. *Trout and salmon of North America.* New York: Simon and Schuster.

Bertolazzi E. 2005. A combination formula of Michaelis-Menten-Monod type. *Computers and Mathematics with Applications*, **50**: 201-215.

Bertram D.F., D.L. Mackas and S.M. McKinnell. 2001. The seasonal cycle revisited: interannual variation and ecosystem consequences. *Progress in Oceanography*, **49**: 283-307.

Birkeland K. 1996. Consequences of premature return by sea trout (*Salmo trutta*) infested with the salmon louse (*Lepophtheirus salmonis* Krøyer): migration, growth and mortality. *Canadian Journal of Fisheries and Aquatic Sciences,* **53**: 2808-2813.

Bjørnsson B.T. 1997. The biology of salmon growth hormone: from daylight to dominance. *Fish Physiology and biochemistry*, **17:** 9-24.

Bowditch N. 1802. *The American practical navigator: an epitome of navigation*. National Imagery and Mapping Agency, U.S. Government 2002 Bicentennial Edition. Bethesda: Maryland, USA. Available at <ftp://ftp.flaterco.com/xtide/Bowditch.pdf>.

Bolund P. and S. Hunhammar. 1999. Ecosystem services in urban areas. *Ecological Economics*, **29** (2): 293-301.

Box G.E.P. and D.R Cox. 1964. An analysis of transformations. *Journal of the Royal Statistical Society, Series B (Methodological*), **26** (2): 211-252.

Bradshaw G.A. and B.A. McIntosh. 1994. Detecting climate-induced patterns using wavelet analysis. *Environmental Pollution*, **83**: 135-142.

Breslow N.E. 1996. Generalized linear models: checking assumptions and strengthening conclusions. *Statistica Applicata*, **8**: 23-41.

Brix A. and P.J. Diggle. 2001. Spatiotemporal prediction for log-Gaussian Cox processes. *Journal of the Royal Statistical Society, Series B*, **63** (4): *823-841*.

Brobbel M.A., M.P. Wilkie, K. Davidson, J.D. Kieffer, A.T. Bielak and B.L. Tufts. 1996. Physiological effects of catch and release angling in Atlantic salmon (*Salmo salar*) at different stages of freshwater migration. *Canadian Journal of Fisheries and Aquatic Sciences*, **53**: 2036-2043.

Brown G.E. and J.A. Brown and A.M. Crosbie. 1992. Do rainbow trout and Atlantic salmon discriminate kin? *Canadian Journal of Zoology*, **70**: 1636-1640.

Buckley Y.M., D.T. Briese and M. Rees. 2003. Demography and management of the invasive plant species Hypericum perforatum. I. Using multi-level mixed-effects models for characterizing growth, survival and fecundity in a long-term data set. *Journal of Applied Ecology*, **40**: 481-493.

Byrne C.J., R. Poole, M. Dillane, G. Rogan and K.F. Whelan. 2004. Temporal and environmental influences on the variation in sea trout (*Salmo trutta* L.) smolt migration in the Burrishoole system in the west of Ireland from 1971 to 2000. *Fisheries Research*, **66**: 85-94.

Carrivick P.J.W., A.H. Lee and K.K.W. Yau. 2003. Zero-inflated Poisson modelling to evaluate occupational safety interventions. *Safety Science*, **41**: 53-63.

Cash D.W., W.N. Adger, F. Berkes, P. Garden, L. Lebel, P. Olson, L. Pritchard and O. Young. 2006. Scale and cross-scale dynamics: Governance and information in a multilevel world. *Ecology and Society*, **11** (2): 8.

Cazelles B., M. Chavez, D. Berteaux, F. Ménard, J.O. Vik, S. Jenouvrier and N.C. Stenseth. 2008. Wavelet analysis of ecological time series. *Oecologia*, (Population Ecology, Original Paper), **156**: 287-304.

Chatfield C. 2003. The analysis of time series: an introduction. Sixth Edition. Chapman and Hall/CRC Texts in Statistical Science series.

Cho E. and T.-S. Chon. 2006 Application of wavelet analysis to ecological data. *Ecological Informatics*, **1**: 229-233.

Chow S.-M., E.L. Hamaker, F. Fujita and S.M. Boker. 2009. Representing time-varying cyclic dynamics using multiple-subject state-space models. *British Journal of Mathematical and Statistical Psychology*, **62**: 683-716.

Cleveland R.B., W.S. Cleveland, S.E. McRae and I. Terpenning. 1990. STL: a seasonal-trend decomposition procedure based on Loess. *Journal of Official Statistics*, **6** (1): 3-34.

Cohen J.E. 1995. Unexpected dominance of high frequencies in chaotic nonlinear population models, *Nature*, **378**: 610-613.

Commandeur J.J and S.J. Koopman. 2007. *An introduction to state space time series analysis*. New York: Oxford University Press Inc.

Connell J.H., T.P. Hugues and C.C. Wallace. 1997. A 30-year study of coral abundance, recruitment, and disturbance at several scales in space and time. *Ecological Monographs*, **67** (4): 461-488.

Costello M.J. 2006. Ecology of sea lice parasitic on farmed and wild fish. *TRENDS in Parasitology*, **22** (10): 476-483.

Côté I.M. and Perrow M.R. 2006. *Fish* in: Sutherland W.J. (ed) *Ecological census techniques*, 2nd Edition: 250-277. Cambridge University Press, Cambridge.

Cox D.R. 1983. Some remarks on overdispersion. *Biometrika*, **70** (1): 269-274.

Coyle C.L. and D. Reed. 2012. Assessment of the performance of a flat panel resistivity fish counter at Peterson Creek, 2007 and 2008. Alaska Department of Fish and Game, *Fishery Data Series*, **12-78** (Anchorage).

Curtis J.A. 2002. Estimating the demand for salmon angling in Ireland. *The Economic and Social Review*, **33** (3): 319-332.

Dadswell M.J., A.D. Spares, J.M. Reader and M.J.W Stokesbury. 2010. The North Atlantic subpolar gyre and the marine migration of Atlantic salmon Salmo salar: the 'Merry-go-Round' hypothesis. *Journal of Fish Biology*, **77** (3): 435-467.

Dempson J.B., M.J. Robertson, C.J. Pennell, G. Furey, M. Bloom, M. Shears, L.M.N. Ollerhead, K.D. Clarke, R. Hinks and G.J. Robertson. 2011. Residency time, migration route and survival of Atlantic salmon *Salmo salar* smolts in a Canadian fjord. *Journal o Fish Biology*, **78** (7): 1976-1992.

Desse-Berset N. 2011. *Biology and Conservation of the European Sturgeon Acipenser sturio L 1758*. Chapter 7: Ancient Sturgean populations in France through archaeozoological remains, from Prehistoric time until the eighteenth century. Williot P., E. Rochard, N. Desse-Berset, N. Kirschbaum and F. Gessner (Eds).

Dethlefsen C., S. Lundbye-Christensen and A.L. Christensen. 2012. sspir: State Space Models in R. R package version 0.2.10. Available at <http://CRAN.R-project.org/package=sspir>.

Diggle, P.J. 1990. *Time series: a biostatistical introduction*. Oxford, UK: Oxford Statistical Science Series 5, Clarendon Press.

Dodson J.J., R.J. Gibson, R.A. Cunjak, K.D. Friedland, C.G. de Leaniz, M.R. Gross, R. Newbury, J.L. Nielsen, M.E. Power and S. Roy. 1998. Elements in the development of conservation plans for Atlantic salmon (*Salmo salar*). *Canadian Journal of Fisheries and Aquatic Sciences*, **55** (1): 312-323.

Dowd M. 2011. Estimating parameters for a stochastic dynamic marine ecological system. *Environmetrics*: **22** (4): 501-515.

Dumas J. and P. Prouzet. 2003. Variability of demographic parameters and population dynamics of Atlantic salmon (*Salmo salar* L.) in a southwest French river. *ICES Journal of Marine Science*, **60**: 356-370.

Eatherley D.M.R., J.L. Thorley, A.B. Stephen, I. Simpson, J.C. MacLean and A.F Youngston. 2005. Trends in Atlantic salmon: the role of automatic fish counter data in their recording. *Scottish Natural Heritage Commissioned Report No. 100* (ROAME No. F01NB02).

Ellies J. and R. Davies. 2001. *Statistical interpretation and species apportionment of acoustic fish counter data*. Bristol, UK. Environment Agency Wales. WRc Report No. EA 5012.

Elliott S.R., T.A. Coe, J.M. Helfield and R.J. Naiman. 1998. Spatial variation in environmental characteristics of Atlantic salmon (*Salmo salar*) rivers. *Canadian Journal of Fisheries and Aquatic Sciences*, **55** (1): 267-280.

Elliott J.M. 1985. Population regulation for different life-stages of migratory trout *Salmo trutta* in a Lake District stream, 1966-83. *Journal of Animal Ecology*, **54** (2): 617-638.

Enfield D.B., A.M. Mestas-Nunez, and P.J. Trimble, 2001: The Atlantic Multidecadal Oscillation and its relationship to rainfall and river flows in the continental U.S., *Geophysical Research Letters*, **28**: 2077-2080.

EA (Environment Agency). 2013. Available at <http://www.environment-agency.gov.uk>. [Accessed April 2013].

Enquist B.J., E.P. Economo, T.E. Huxman, A.P. Allen, D.D Ignace and J.F. Gillooly. 2003. Scaling metabolism from organisms to ecosystems. *Nature*, **423**: 639-642.

Environment Agency and Cefas. 2012. *Annual assessment of salmon stocks and fisheries in England and Wales, 2011*. Preliminary assessment prepared for ICES, March 2012.

Epstein J.H., K.J. Olival, J.R.D. Pulliam, C. Smith, J. Westrum, T. Hughes, A.P. Dobson, A. Zubaid, S.A. Rahman, M.M. Basir, H.E. Field and P. Daszak. 2009. *Pteropus vampyrus*, a hunted migratory species with a multinational home-range and a need for regional management. *Journal of Applied Ecology*, **46**: 991-1002.

Fanshawe T.R., P.J. Diggle, S. Rushton, R. Sanderson, P.W.W. Lurz, S.V. Glinianaia, M.S. Pearce, L. Parker, M. Charlton and T. Pless-Mulloli. 2008. Modelling spatio-temporal variation in exposure to particulate matter: a two-stage approach. *Environmetrics*, **19**: 549-566.

FAO. 2013. Appendix I – Fish and fishery products – apparent consumption. Available at <ftp://ftp.fao.org/fi/stat/summary/appIybc.pdf>. [Accessed April 2013].

Fevolden S.E., R. Nordmo, T. Refstie and K.H. Roed. 1993. Disease resistance in Atlantic salmon (*Salmo salar*) selected for high or low responses to stress. *Aquaculture*, **109**: 215-224.

Fewings G.A. 1994. *Automatic salmon counting technologies – A contemporary review*. Pitlochry, Perthshire: Atlantic Salmon Trust.

Fillatre E.K. 2002. *Bimodal return distribution in a northern population of salmon: Genetic, life history, and habitat analysis of adult and juvenile sockeye salmon* (*Oncorhynchus nerka*). Ph.D. University of Windsor, Available at : <http://scholar.uwindsor.ca/cgi/viewcontent.cgi?article=2296&context=etd>. [Accessed April 2013].

Fleming I.A. 1996. Reproductive strategies of Atlantic salmon: ecology and evolution. *Reviews in Fish Biology and Fisheries*, **6**: 379-416.

Folke C., A. Jansson, J. Larsson and R. Costanza. 1997. Ecosystem appropriation by cities. *Ambio*, **26** (3): 167-172.

Folmar L.C. and W.W. Dickhoff. 1980. The parr-smolt transformation (smoltification) and seawater adaptation in Salmonids. *Aquaculture*, **21**: 1-37.

H E Forbes H.A., G.W. Smith, A.D.F. Johnstone and A.B. Stephen. 2000. An assessment of the performance of the resistivity fish counter in the Borland lift fish pass at Dundreggan dam on the River Moriston. Fisheries Research Services Report No 02/00.

Fox J. 2009. Regression Diagnostics. Lecture notes November 2009, Department of Sociology, McMaster University Canada. Available at <http://socserv.socsci.mcmaster.ca/jfox/Courses/Brazil-2009/>. [Accessed April 2013].

Fox J. 2002. Linear mixed models: Appendix to An R and S-PLUS companion to applied regression. On R home page available at <http://www.r-project.org/>. [Accessed April 2013].

Fraley C., A.E. Raftery, T.B. Murphy and L. Scrucca. 2012. Mclust. Version 4 for R: Normal Mixture Modeling for Model-Based Clustering, Classification, and Density Estimation. Technical Report No. 597. Department of Statistics, University of Washington.

Fraley C. and A.E. Raftery. 2002. Model-based Clustering, Discriminant Analysis and Density Estimation. *Journal of the American Statistical Association*, **97**: 611-631.

Freyhof J. and M. Kottelat. 2008. *Salmo trutta*. In: IUCN 2010. IUCN Red List of Threatened Species. Version 2010.4. Available at <www.iucnredlist.org>. [Accessed November 2010].

Friedland K.D. 1998. Ocean climate influences on critical Atlantic salmon (*Salmo salar*) life history events. *Canadian Journal of Fisheries and Aquatic Sciences*, **55** (1): 119-130.

Friedland K.D., L.P. Hansen and D.A. Dunkley. 1998. Marine temperatures experiences by postmolts and the survival of Atlantic salmon, *Salmo salar* L., in the North Sea area. *Fisheries Oceanography*, **7** (1): 22-34.

Friedland K.D., D.G. Reddin and J.F. Kocik. 1993. Marine survival of North American and European Atlantic salmon: effects of growth and environment. *ICES Journal of Marine Science*, **50**: 481-492.

Fulton E.E., A.D.M. Smith, D.C. Smith and I.E. van Putten. 2010. Human behaviour: the key source of uncertainty in fisheries management. *Fish and Fisheries*, **12** (1): 2-17.

Garcia-Vazquez E., P. Mornan, J.L. Martinez, B. de Gaudemar and E. Beall. 2001. Alternative mating strategies in Atlantic salmon and brown trout. *The Journal of Heredity*, **92** (2): 146-149.

Gauthier-Ouellet M., M. Dionne, F. Caron, T.L. King and L. Bernatchez. 2009. Spatiotemporal dynamics of the Atlantic salmon (*Salmo salar*) Greenland fishery inferred from mixed-stock analysis. *Canadian Journal of Fisheries and Aquatic Sciences*, **66**: 2040-2051.

Gill J.A., K. Norris, P.M. Potts, T.G. Gunnarsson, P.W. Atkinson and W.J. Sutherland. 2001. The buffer effect and large-scale population in migratory birds. *Nature*, **412**: 436-438.

Goodenough A.E., A.G. Hart and R. Stafford. 2012. Regression with Empirical Variable Selection: Description of a New Method and Application to Ecological Datasets. *PLoS ONE,* **7** (3): e34338.

Gouhier T.C. and A. Grinsted. 2012. biwavelet: Conduct univariate and bivariate wavelet analyses. R package version 0.12. Available at <http://CRAN.R-project.org/package=biwavelet>.

Graham N.A.J., T.R. McClanahan, M.A. MacNeil, S.K. Wilson, N.V.C. Polunin, S. Jennings, P. Chabanet, S. Clark, M.D. Spalding, Y. Letourneur, L. Bigot, R. Galzin, M.C. Öhman, K.C. Garpe, A.J. Edwards and C.R.C. Shappard. 2008. Climate warming, marine protected areas and the ocean-scale integrity of coral reef ecosystems. *PLoS ONE*, **3** (8): e3039.

Greatbatch R. 2000. The North Atlantic Oscillation. *Stochastic Environmental Research and Risk Assessment*, **14**: 213-242.

Greaves R.K., R.A. Sanderson and S.P. Rushton. 2006. Predicting species occurrence using information-theoretic approaches and significance testing: an example of dormouse distribution in Cumbria, UK. *Biological Conservation*, **130** (2): 239-250.

Grinsted A., J.C. Moore and S.Jevrejeva. 2004. Application of the cross wavelet transform and wavelet coherence to geophysical time series. *Nonlinear Processes in Geophysics*, **11**: 561-566.

Grønvik S. and A. Klemetsen. 1987 Marine food and diet overlap of co-ocurring Artic Charr *Salvelinus alpinus* (L.), Brown Trout *Salmo trutta* L. and Atlantic Salmon *S. salar* L. off Senja, N. Norway. *Polar Biology*, **7**: 173-177.

Gross M.R. 1998. One species with two biologies: Atlantic salmon (*Salmo salar*) in the wild and in aquaculture. *Canadian Journal of Fisheries and Aquatic Sciences*, **55** (1): 131-144.

Gross M. R. 1991. Salmon breeding behaviour and life history evolution in changing environments. *Ecology*, **72** (4): 1180-1186.

Gutierrez-Osuna R. 2011. Pattern Analysis; L10: Linear discriminant analysis. Fall 2011 (Lectures). College Station: Texas Agricultural and Mechanical University. Available at <http://courses.cs.tamu.edu/rgutier/csce666_f11/>. [Accessed June 2013].

Haile S.M., A. James and D. Sear. 1989. *The effects of Kielder reservoir on the ecology of the River North Tyne*. Report to the Northumbria Water Authority, 1989. University of Newcastle-upon-Tyne, Department of Civil Engineering.

Hall J.A., C.L.J. Frid and M.E. Gill. 1997. The response of estuarine fish and benthos to an increasing discharge of sewage effluent. *Marine Pollution Bulletin*, **34** (7): 527-535.

Hansen L.P., P. Hutchinson, D.G. Reddin and M.L. Winsor. 2012. Salmon at sea: scientific advances and their implications for management: an introduction. *ICES Journal of Marine Science*, **69** (9): 1533-1537.

Hansen L.P. and T.P. Quinn. 1998. The marine phase of the Atlantic salmon (*Salmo salar*) life cycle, with comparisons to Pacific salmon. *Canadian Journal of Fisheries and Aquatic Sciences*, **55** (1): 104-118.

Haro A., M. Odeh, J. Noreika and T. Castro-Santos. 2004. Effect of water acceleration on downstream migratory behaviour and passage of Atlantic salmon smolts and juvenile American shad at surface bypasses. *Transactions of the American Fisheries Society*, **127** (1): 118-127.

Harris J. and J.D. Bird. 2000. Modulation of the fish immune system by hormones, mini review. *Veterinary Immunology and Immunopathology*, **77**: 163-176.

Haury J., *J.L.* Bagliniére, *A.I.* Cassie and G. Maisse. 1994. Analysis of spatial and temporal organisation in a salmonid brook in relation to physical factors and macrophytic vegetation. *Hydrobiologia*, **300/301**: 269–277.

Heggenes J., S.J. Salveit, D. Bird and R.Grew. 2002. Static habitat partitioning and dynamic selection by sympatric young Atlantic salmon and brown trout in south-west England streams. *Journal of Fish biology*, **60**: 72-86.

Heggenes J., S.J. Saltveit and O. Lingaas. 1996. Predicting fish habitat use to changes in water flow: modelling critical minimum flows for Atlantic salmon, *Salmo salar*, and brown trout, *S. trutta*. *Regulated Rivers: Research and Management*, **12**: 331-344.

Hendry P.H., P. Nosil and L.H. Riesberg. 2007. The speed of ecological speciation. *Functional Ecology*, **21** (3): 455-464.

Hendry A.P., J.K. Wenburg, P. Bentzen, E.C. Volk and T.P. Quinn. 2000. Rapid evolution of reproductive isolation in the wild: evidence from introduced salmon. *Science*, **290**: 516-518.

Hesthagen T. 1988. Movement of brown trout, *Salmo trutta*, and juvenile Atlantic salmon, *Salmo salar*, in a coastal stream in northern Norway. *Journal of Fish Biology*, **32** (5): 639 – 653.

Hilborn R., C.J. Walters and D. Ludwig. 1995. Sustainable Exploitation of Renewable Resources. *Annual Review of Ecology and Systematics*. **26**: 45-67.

Hogg III C.H., K. Mullen and I. Levin. 2012. gppois: A Bayesian approach for denoising one-dimensional data.R package version 0.2-1. *Journal of Applied Crystallography*, **45** (3): 471-481. Available at <http://dx.doi.org/10.1107/S0021889812015154>.

Holm M., J.C. Holst and L.P. Hansen. 2000. Spatial and temporal distribution of post-smolts of Atlantic salmon (*Salmo salar* L.) in the Norwegian Sea and adjacent areas. *ICES Journal of Marine Science*, **57**: 955-964.

Horne J.K. 2000. Acoustic approaches to remote species identification: a review. *Fisheries Oceanography*, **9** (4): 356-371.

Houde A.L.S., D.J. Fraser and J.A Hutchings. 2010, Fitness-related consequences of competitive interactions between farmed and wild Atlantic salmon at different proportional representations of wild-farmed hybrids. *ICES Journal of Marine Science*, **67**: 657-667.

Hughes, P. J. 2006. North Atlantic Decadal Variability of Ocean Surface Fluxes. Ph.D. thesis, Florida State University, USA.

Hughes N.F. 2004. The wave-drag hypothesis: an explanation for size-based lateral segregation during the upstream migration of salmonids. *Canadian Journal of Aquatic Sciences*, **61**: 103-109.

Huidong T and B. Cazelles. 2011. WaveletCo: Wavelet Coherence Analysis. R package version 1.0. Available at <http://CRAN.R-project.org/package=WaveletCo>.

Hutchings J.A and J.D. Reynolds. 2004. Marine fish population collapses: consequences for recovery and extinction risk. *BioScience*, **54** (4): 297-309.

Hutchings J.A. and M.E.B. Jones. 1998. Life history variation and growth rate thresholds for maturity in Atlantic salmon, *Salmo salar*. *Canadian Journal of Fisheries and Aquatic Sciences*. **55** (1): 22-47.

Hutchinson V.H. and J.D. Maness. 1979. The role of behavior in temperature acclimation and tolerance in ectotherms. *American Zoologist*, **19**: 367-384.

I.A.S.R.B., International Atlantic salmon research board. 2012. Inventory of Research Relating to Salmon Mortality in the Sea. ICR(12)3. Available at <http://www.nasco.int/sas/research.htm>.

IPCC, Intergovernmental Panel on Climate Change, IPCC TAR WG1. 2001. *9.3.4.3: Thermohaline circulation changes* in J.T. Houghton, Y. Ding, D.J. Griggs, M. Noguer, P.J. van der Linden, X. Dai, K. Maskell and C.A. Johnson. Climate Change 2001: The Scientific Basis, Contribution of Working Group I to the Third Assessment Report of the IPCC, Cambridge University Press. Available at <http://www.grida.no/publications/other/ipcc_tar/?src=/climate/ipcc_tar/wg1/357.htm>.

Jackman S. 2012. pscl: Classes and Methods for R Developed in the Political Science Computational Laboratory, Stanford University. Department of Political Science, Stanford University. Stanford, California. R package version 1.04.4. Available at <http://pscl.stanford.edu/>.

Jaensson A., A.P. Scott, A. Moore, H. Kylin and K.H. Olsen. 2007. Effects of a pyrethroid pesticide on endocrine responses to female odours and reproductive behaviour in male parr of brown trout (*Salmo trutta* L.). *Aquatic Toxicology*, **81**: 1-9.

Jakubauskas M.E., D.R. Legates and J.H. Kastens. 2001. Harmonic analysis of time-series AVHRR NDVI data. *Photogrammetric Engineering and Remote Sensing*, **67** (4): 461-470.

James M. 1985. *Classification algorithms*. Collins. New York, NY, USA: Wiley-Interscience.

Jansson H. and T. Ost. 1997. Hybridization between Atlantic salmon (*Salmo salar*) and brown trout (*S. trutta*) in a restored section of the River Dalalven, Sweden. *Canadian Journal of Fisheries and Aquatic Sciences*, **54**: 2033-2039.

Jennings A., S.P.R. Greenstreet and J.D. Reynolds. 1999. Structural changes in an exploited fish community: a consequence of differential fishing effects in species with contrasting life histories. *Journal of Animal Ecology*, **68**: 617-627.

Jensen A.J, B. Finstad, P. Fiske, N.A. Hvidsten, A.H. Rikardsen and L. Saksgård. 2012. Timing of smolt migration in sympatric populations of Atlantic salmon (*Salmo salar*), brown trout (*Salmo trutta*), and Arctic charr (*Salvelinus alpinus*). *Canadian Journal of Fisheries and Aquatic Sciences*, **69**: 711-723.

Jepsen N., K. Aarestrup, F Okland and G. Rasmussen, 1998. Survival of radio-tagged Atlantic salmon (*Salmo salar* L.) and trout (*Salmo trutta* L.) smolts passing a reservoir during seaward migration. *Hydrobiologia*, **371-372**: 347-353.

Joanes D.N. and C.A. Gill. 1998. Comparing measures of sample skewness and kurtosis. *The Statistician,* **47** (1): 183-189.

Johnson K.A. and R.S. Goody. 2011. The Original Michaelis Constant: Translation of the 1913 Michaelis-Menten Paper. *Biochemistry*, **50** (39): 8264-8269.

Johnsson J.I. and A. Forser. 2002. Residence duration influences the outcome of territorial conflicts in brown trout (*Salmo trutta*). *Behavioural Ecology and Sociobiology,* **51** (3): 282-286.

Joint Nature Conservation Committee. 2007. Second report by the UK under Article 17 on the implementation of the habitats directive from January 2001 to December 2006. Peterborough: JNCC. Available at <www.jncc.gov.uk/article17>. [Accessed November 2010].

Jonsen I.D., J.M. Flemming and R.A. Myers. 2005. Robust state-space modelling of animal movement data. *Ecology*, **86** (11): 2874-2880.

Jonsson B. and N. Jonsson. 2011. Ecology of Atlantic Salmon and Brown Trout: Habitat as a Template for Life Histories. D.L.G. Noakes, Fisheries & Wildlife Department, Oregon State University, Corvallis, USA (Eds.). Dordrecht, Heidelberg, London, New York: Springer.

Jonsson B. and N. Jonsson. 2009. A review of the likely effects of climate change on anadromous Atlantic salmon *Salmo salar* and brown trout *Salmo trutta*, with particular reference to water temperature and flow. *Journal of Fish Biology*, **75**: 2381-2447.

Jonsson B., N. Jonsson and L.P. Hansen. 2007. Factors affecting river entry of adult Atlantic salmon in a small river. *Journal of Fish Biology*, **71**: 943-956.

Jonsson B. and N. Jonsson. 2004. Factors affecting marine production of Atlantic salmon (*Salmo salar*). *Canadian Journal of Fisheries and Aquatic Sciences*, **61**: 2369-2383.

Jonsson N., B. Jonsson and L.P. Hansen. 1998. The relative role of density-dependent and density-independent survival in the life cycle of Atlantic salmon *Salmo salar*. *Journal of Animal Ecology*, **67** (5): 751-762.

Jonsson N., B. Jonsson and L.P. Hansen. 1990. Partial segregation in the timing of migration of Atlantic salmon of different ages. *Animal Behaviour*, **40**: 313-321.

Jorgensen E.H. and M. Jobling. 1992. Feeding behaviour and effect of feeding regime on growth of Atlantic salmon, *Salmo salar*. *Aquaculture*, **101**: 135-146.

JRC: European commission Joint Research Committee. Calculation page Available at <http://re.jrc.ec.europa.eu/pvgis/apps/radday.php>. [Accessed March 2013].

Juell J.-E. and J.E. Fosseidengen. 2004. Use of artificial light to control swimming depth and fish density of Atlantic salmon (*Salmo salar*) in production cages. *Aquaculture*, **233**: 269-282.

Kaitala V., E. Ranta and J. Lindström. 1996. Cyclic population dynamics and random perturbations. *Journal of Animal Ecology*, **65**: 249-251.

Kalman R.E. 1960. A new approach to linear filtering and prediction problems. *Transactions of the ASME-Journal of Basic engineering*, **82** (D): 35-45.

Keenleyside M.H.A. and F.T. Yamamoto. 1962. Territorial behaviour of juvenile Atlantic salmon. *Behaviour*, **19** (1/2): 139-169.

Kennelly S.J. and M.K. Broadhurst. 2002. By-catch begone: changes in the philosophy of fishing technology. *Fish and Fisheries*, **3**: 340-355.

King A.W., A.R. Johnson and R.V. O'Neill. 1991. Transmutation and functional representation of heterogeneous landscapes. *Landscape Ecology*, **5** (4): 239-253.

Klaus B. and K. Strimmer. 2012. fdrtool: Estimation of (Local) False Discovery Rates and Higher Criticism. R package version 1.2.10. Available at <http://CRAN.R-project.org/package=fdrtool>.

Klemetsen A., P.-A. Amundsen, J.B. Dempson, B. Jonsson, N. Jonsson, M.F. O'Connell and E. Mortensen. 2003. Atlantic salmon *Salmo salar* L., brown trout *Salmo trutta* L. and Arctic charr *Salvelinus alpinus* (L.): a review of aspects of their life histories. *Ecology of Freshwater Fish*, **12**: 1-59.

Knight J.R., R.J. Allan, C.K. Folland and M. Vellinga. 2005. A signature of persistent natural thermohaline circulation cycles in observed climate. *Geophysical Research Letters*, **32**: 20.

Knudsen, M. F., M.S. Seidenkrantz, B.H. Jacobsen and A. Kuijpers. 2011. Tracking the Atlantic Multidecadal Oscillation through the last 8,000 years. *Nature communications*, **2**: 178.

Kozak G.M. and J.W. Boughman. 2012. Plastic responses to parents and predators lead to divergent shoaling behaviour in sticklebacks. *Journal of Evolutionary biology*, **25** (4): 759-769.

Kracker L.M. 1999. The geography of fish: the use of remote sensing and spatial analysis tools in fisheries research. *The Professional Geographer*, **51** (3): 440-450.

Krause J., R.K. Butlin, N. Peuhkuri and V.L. Pritchard. 2000. The social organization of fish shoals: a test of the predictive power of laboratory experiments for the field. *Biological Reviews*, **75**: 477-501.

Krause J. and R.W. Tegeder. 1994. The mechanism of aggregation behaviour in fish shoals: individuals minimize approach time to neighbours. *Animal Behaviour*, **48**: 353-359.

Lavrakas J.W. W. Black, A. Lawson and P. Lawson. *At-sea data collection in the salmon fisheries using GPS-enabled Android.* Organization of Fish and Wildlife Information Managers, 2012 Conference and Annual Meeting, Austin, Texas.

Leibold M.A., M. Holyoak, N. Mouquet, P. Amarasekare, J.M. Chase, M.F. Hoopes, R.D. Holt, J.B. Shurin, R. Law, D. Tilman, M. Loreau and A. Gonzalez. 2004. The metacommunity concept: a framework for multi-scale ecology. *Ecology Letters*, **7**: 601-613.

Le Monde. 2008. Un saumon atlantique de 7 kg pêché aux portes de Paris. *Le Monde*, October 09th 2008.

Levin S.A. 1989. The problem of pattern and scale in ecology. *Ecology*, **73** (6): 1943-1967.

Leydesdorff L. and S. Bensman. 2006. Classification and powerlas: the logarithmic transformation. *Journal of the American Society for Information Science and Technology*, **57** (11): 1470-1486.

Lord D., S.P. Washington abd J.N. Ivan. 2005. Poisson, Poisson-Gamma and zero-inflated regression models of motor vehicle crashes: balancing statistical fit and theory. *Accident analysis and prevention*, **37** (1): 35-46.

Lundbye-Christensen S., C. Dethlefsen, A. Gorst-Rasmussen, T. Fischer, H.C. Schonheyder, K.J Rothman and H.T. Sorensen. 2009. Examining secular trends and seasonality in count data using dynamic generalized linear modelling: a new methodological approach illustrated with hospital discharge data on myocardial infarction. 2009. *European Journal of Epidemiology*, **24**: 225-230.

MacNeil M.A., N.A.J. Graham, J.E. Cinner, N.K. Dulvy, P.A. Loring, S. Jennings, N.V.C. Polunin, A.T. Fisk and T.R. McClanahan. 2010. Transitional states in marine fisheries: adapting to predicted global change. *Philosophical Transactions of the Royal Society B*, **365**: 3753-3763.

MacNeil M.A., N.A.J. Graham, N.V.C. Polunin, M. Kulbicki, R. Galzin, M. Harmelin-Vivien and S.P. Rushton. 2009. Hierarchical drivers of reef-fish metacommunity structure. *Ecology*, **90** (1): 252-264.

Magnuson J.J., L.B. Crowder and P.A. Medvick. 1979. Temperature as an ecological resource. *American Zoologist,* **19** (1): 331-343.

Maitland P.S. 1965. The feeding relationship of Salmon, Trout, Minnows, Stone Loach and Three-Spined Stickle-Backs in the River Endrick, Scotland. *Journal of Animal Ecology*. **34** (1): 109-133.

Malcolm I. A., J. Godfrey and A. F. Youngson. 2010. Review of migratory routes and behaviour of Atlantic salmon, sea trout and European eel in Scotland's

coastal environment: implications for the development of marine renewables. *Scottish Marine and* Freshwater *Science,* **1** (14).

Marquet P.A., R.A. Quinones, S. Abades, F. Labra, M. Tognelli, M. Arim and M. Rivadeneira. 2005. Scaling and power-laws in ecological systems. *The Journal of Experimental Biology*, **208**: 1749-1769.

Marschall E.A., M.E. Mather, D.L. Parrish, G.W. Allison and J.R. McMenemy. 2011. Migration delays caused by anthropogenic barriers: modeling dams, temperature, and success of migrating salmon smolts. *Ecological Applications*, **21** (8): 3014–3031.

Martin T.G., B.A. Wintle, J.R. Rhodes, P.M. Kuhnert, S.A. Field, A.J. Low-Choy, A.J. Tyre and H.P. Possingham. 2005. Zero tolerance ecology: improving ecological inferences by modelling the source of zero observations. *Ecology Letters*, **8** (11): 1235-1246.

Masood E. 1996. Scientific caution 'blunts efforts' to conserve fish stocks. *Nature*, **379**: 481.

McCormick S.D., L.P. Hansen, T.P. Quinn and R.L. Saunders. 1998. Movement, migration and smelting of Atlantic salmon (*Salmo salar*). *Canadian Journal of Fisheries and Aquatic Sciences*, **55** (1): 77-92.

McCubbing D.J.F. and D. Ignace. 1999. Salmonid escapement estimates on the Deadman River, resistivity counter video validation and escapement estimates. MEOLP Project Report No. 2000, Vancouver.

McGinnity P., P. Prodohl, A. Ferguson, R. Hynes, N. O Maoileidigh, N. Baker, D. Cotter, B. O'Hea, D. Cooke, G. Rogan, J. Taggart and T. Cross. 2003. Fitness reduction and potential extinction of wild populations of Atlantic salmon, *Salmo salar*, as a result of interactions with escaped farm salmon. *Proceedings of the Royal Society, Biological Sciences*, **270**: 2443-2450.

McKelvey R.W., L.K. Sandal and S.I. Steinshamn. 2002. Fish wars on the high seas: a straddling stock competition model. *International Game Theory Review*, **4** (1): 53-69.

McKillop J. 2004. "Salmon" in: *A Dictionary of Celtic Mythology*. Oxford: Oxford University Press.

McLeod C., J. Grice, H. Campbell and T. Herleth. 2006. *Super salmon: the industrialisation of fish farming and the drive towards GM technologies in salmon production*. Centre of the Study of Agriculture, Food and Environment, University of Otago, New Zealand. CSAFE Discussion Paper No. 5.

McClure M. and E. Despland. 2011. Defensive responses by a social caterpillar are tailored to different predators and change with larval instar and group size. *Naturwissenschaften*, **98** (5): 425-434.

Mendelssohn R., F.B. Schwing and S.J. Bograd. 2004. Nonstationary seasonality of upper ocean temperature in the California Current. *Journal of Geophysical Research*, **109** (C10015).

Menten L. and M.I. Michaelis. 1913. Die Kinetik des Invertinwirkung. *Biochemische Zeitschrift*, **49**: 333-369.

Metcalfe N.B., S.K. Valdimarsson and N.H.C. Fraser. 1997. Habitat profitability and choice in a sit-and wait predator: juvenile salmon prefer slower currents on darker nights. *Journal of Animal Ecology*, **66**: 866-875.

Metcalfe N.B., P.J. Wright and J.E. Thorpe. 1992. Relationships between social status, otolith size at first feeding and subsequent growth in Atlantic salmon *(Salmo salar*). *Journal of Animal Ecology,* **61** (2): 585-589.

Middlemas S.J., D.C. Stewart, S. Mackay and J.D. Armstrong. 2009. Habitat use and dispersal of post-smolt sea trout *Salmo trutta* in a Scottish sea loch system. *Journal of Fish Biology*, **74**: 639-651.

Minami M., C.E. Lennert-Cody, W. Gao and M. Román-Verdesoto. 2007. Modelling shark bycatch: the zero-inflated negative binomial regression model with smoothing. *Fisheries Research*, **84**: 210-221.

Mingfang T., Y. Kushnir, R. Seagerand and C. Li. 2009. Forced and Internal Twentieth-Century SST Trends in the North Atlantic. *Journal of Climate*, **22** (6): 1469–1481.

Montgomery D.R. 2004. *King of fish: the thousand-year run of salmon*. Boulder: Westview Press.

Moore A., M. Ives, M. Scott and S. Bamber. 1998. The migratory behaviour of wild sea trout (*Salmo trutta* L.) smolts in the estuary of the River Conwy, North Wales. *Aquaculture*, **168**: 57-68.

Moore A., M.J. Ives and L. Kell. 1994. The role of urine in sibling recognition in Atlantic salmon *Salmo salar* (L.) parr. The *Royal Society Proceedings: Biological Sciences*, **255** (1343): 173-180.

Morishita J. 2008. What is the ecosystem approach for fisheries management? *Marine Policy,* **32**: 19-26.

Nemet G.F. 2006. Beyond the learning curve: factors influencing cost reductions in photovoltaics. *Energy Policy*, **34** (17): 3218-3232.

Nicholson S.A., Gallagher R., Aprahamian M.W and Best P. 1994. *Validation of the performance of the Aquantic 2100A fish counter*. Warrington, UK: Environment Agency North West. (NRA/NW/FTR /94/10).

Nielsen J.L. 1998. Population genetics and the conservation and management of Atlantic salmon (*Salmo salar*). *Canadian Journal of Fisheries and Aquatic Sciences*. **55** (1): 145-152.

NOAA. 2005. National Weather Service, National Centres for Environmental Prediction, Climate Prediction Center, USA. Available at <www.cpc.ncep.noaa.gov>.

O'Connell M.F. 2003. An examination of the use of angling data to estimate total returns of Atlantic salmon, Salmo salar, to two rivers in Newfoudland, Canada. *Fisheries Management and Ecology*, **10**: 201-208.

O'Connor K.I., N.B. Metcalfe and A.C. Taylor. 1999. Does darkening signal submission in territorial contests between juvenile Atlantic salmon, *Salmo salar*? *Animal Behaviour*, **58**: 1269-1276.

Okland F., J.Erkinaro, K. Moen, E. Niemela, P. Fiske, R.S. McKinley and E.B. Thorstad. 2001. Return migration of Atlantic salmon in the river Tana: phases of migratory behaviour. *Journal of Fish Biology*, **59**: 862-874.

O'Neil R.V., D.L. DeAngelis, J.B. Waide and T.F.H. Allen. 1986. A Hierarchical Concept of Ecosystems. Volume 23 of Monographs in population biology. Princeton, New Jersey : Princeton University Press.

Orderud F. 2005. Comparison of Kalman filter estimation approaches for state space models with nonlinear measurements. In: *Proceedings of Scandinavian conference on simulation and modelling*.

Osborne, J.W. 2010. Improving your data transformations: applying the Box-Cox transformation. *Practical Assessment, Research & Evaluation*, **15** (12): 1-9.

Palm S., J. Dannewitz, T. Järvi, M. Koljonen, T. Prestegaard and K.H. Olsén. 2008. *Canadian Journal of Fisheries and Aquatic Sciences*, **56**: 1738-1748.

Parrish J.K. 1999. Toward remote species identification. *Oceanography*, **12** (3): 30-32.

Parrish D.L., R. J. Behnke, S.R. Gephard, S.D McCormick and G.H. Reeves. 1998. Why aren't there more Atlantic salmon (*Salmo salar*)? *Canadian Journal of Fisheries and Aquatic Sciences*, **55** (1): 281-187.

Parry G. 1958. Size and osmoregulation in salmonid fishes. *Nature*, **181**: 1218-1219.

Pauly D., R. Hilborn and T.A. Branch. 2013. Fisheries: does catch reflect abundance? *Nature*, **494**: 303-316.

Pauly D., V. Christensen, S. Guénette, T.J. Pitcher, U.R. Sumaila, C.J. Walters, R. Watson and D. Zeller. 2002. Towards sustainability in world fisheries. *Nature*, **418**: 689-696.

Perriere G. and J. Thioulouse. 2003. Use of Correspondence Discriminant Analysis to predict the subcellular location of bacterial proteins. *Computer Methods and Programs in Biomedicine*, **70**: 99–105.

Petersson E. and T. Jarvi. 1997. Reproductive behaviour of sea trout (*Salmo trutta*) – the consequences of sea-ranching. *Behaviour*, **134**: 1-22.

Pichler P. 2007. State space models and the Kalman filter. *Seminar paper for Vektorautoregressive Methoden*, Universität Wien.

Pierce D.A. and D.W. Schafer. 1986. Residuals in Generalized Linear Models. *Journal of the American Statistical Association*, **81** (396): 977-986.

Pinheiro J., D. Bates, S. DebRoy, D. Sarkar and the R Development Core Team. 2012. nlme: Linear and Nonlinear Mixed Effects Models. R package version 3.1-104. Available at <http://cran.r-project.org/web/packages/nlme/index.html>. [Accessed June 2013].

Pinheiro J., D. Bates, S. DebRoy, D. Sarkar and the R Development Core Team. 2012. nlme: Linear and Nonlinear Mixed Effects Models. R package version 3.1-104. Available at <http://cran.r-project.org/web/packages/nlme/index.html>.

Pinheiro J.C. 1994. Topics in Mixed Effects Models. A thesis submitted in partial fulfilment of the requirements for the degree of Doctor of Philosophy (Statistics), at the University of Wisconsin, Madison.

Pitcher T.J. and W.W.L. Cheung. 2013. Fisheries: Hope or despair? *Marine Pollution Bulletin*, **74** (2): 206-516.

Pitcher T.J. 1985. Heuristic definitions of fish shoaling behaviour. *Animal Behaviour*, **31** (2): 611-613.

Potter E.C.E. 1988. Movements of Atlantic salmon, *Salmo salar* L., in an estuary in South-west England, *Journal of Fish Biology*, **33** (sA): 153-159.

Pratten D.J. and W.M. Shearer. 1983. The migrations of North Esk Sea Trout. *Aquaculture Research*, **14** (3): 99-113.

Pyper B.J. and R.M. Peterman. 1998. Comparison of methods to account for autocorrelation in correlation analyses of fish data. *Canadian Journal of Fisheries and Aquatic Sciences,* **55**: 2127-2140.

Quinn T.P., P. McGinnity and T.F. Gross. 2006. Long-term declines in body size and shifts in run timing of Atlantic salmon in Ireland. *Journal of Fish Biology,* **68**: 1713-1730.

R Core Team. 2012. R: A language and environement for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Available at <www.R-project.org/>.

Randle H.D. and P. Nosil. 2005. Ecological speciation. *Ecology Letters*, **8**: 336-352.

Rao R.C. 1948. The utilization of multiple measurements in problems of biological classification. *Journal of the Royal Statistical Society – Series B: Statistical Methodology*, **10** (2): 159-203.

Rasmussen G. 1986. The population dynamics of brown trout (*Salmo trutta* L.) in the realtion to year-class size. *Polskie Archiwum Hydrobiologii*, **33**: 489-508.

Ridout M., J. Hinde and C.G.B. Demétrio. 2001. A score test for testing a zero-inflated Poisson regression model against zero-inflated negative binomial alternatives. *Biometrics*, **57**: 219-223.

Ridout M., C.G.B. Demétrio and J. Hinde. 1998. Models for count data with many zeros. In: *The XIXth International Biometric Conference* (Invited paper), Cape Town, South Africa, December 14 to 18, 1998: 179-192.

Rijnsdorp A.D., P.I. van Leeuwen, N. Daan and H.J.L. Heessen. 1996. Changes in abundance of demersal fish species in the North Sea between 1906-1909 and 1990-1995. *ICES Journal of Marine Sciences*, **53**: 1054-1062.

Riley W.D. 2007. Seasonal downstream movements of juvenile Atlantic salmon, *Salmo salar* L., with evidence of solitary migration of smolts. *Aquaculture*, **273**: 194-199.

Riley W.D., M.O. Eagle and S.J. Ives. 2002. The onset of downstream movement of juvenile Atlantic salmon, *Salmo salar* L., in a chalk stream. *Fisheries Management and Ecology*, **8**: 87-94.

Rivot E., E. Prévost, E. Parent and J.L. Baglinière. 2004. A Bayesian state-space modelling framework for fitting a salmon stage-structured population dynamic model to multiple series of field data. *Ecological Modelling*, **179**: 463-485.

Roberts P.H. and K.V. Thomas. 2006. The occurrence of selected pharmaceuticals in wastewater effluent and surface waters of the lower Tyne catchment. *Science of the Total Environment*, **356** (1-3): 143-153.

Robinson R.A., H.Q.P. Crick, J.A. Learmonth, I.M.D. Maclean, C.D. Thomas, F. Bairlein, M.C. Forchhammer, C.M. Francis, J.A. Gill, B.J. Godley, J. Harwood, G.C. Hays, B. Huntley, A.M. Hutson, G.J. Pierce, M.M. Rehfisch, D.W. Sims, M.B. Santos, T.H. Sparks, D.A. Stroud and M.E. Visser. 2009. Travelling through a warming world: climate change and migratory species. *Endangered Species Research,* **7**: 87-99.

Rolleston T.W. 1910. The High Deeds of Finn and other Bardic Romances of Ancient Ireland. London: G.G. Harrap & Co.

Rose C.E., S.W. Martin, K.A. Wannemuehler and B.D. Plikaytis. 2006. On the use of zero-inflated and hurdle models for modelling vaccine adverse event count data. *Journal of Biopharmaceutical Statistics*, **16**: 463-481.

Roy M., R.D. Holt and M. Barfield. 2005. Temporal autocorrelation can enhance the persistence and abundance of metapopulations comprised of coupled sinks. *The American Naturalist*, **166** (2): 246-261.

Royston P. And D.G. Altman. 1994. Regression using fractional polynomials of continuous covariates: parsimonious parametric modelling. *Applied Statistics*, **43** (3): 429-467.

Rushton S.P., S.J. Ormerod and G. Kerby. 2004. New paradigms for modelling species distributions? *Journal of Applied Ecology*, **41**: 193-200.

Rushton S.P. and M.D Eyre. 1992. Grassland spider habitat in north-east England. *Journal of biogeography*, **19**: 99-109.

Sakamoto Y. and G. Kitigawa. 1986. Akaike Information Criterion Statistics. Tokyo : KTK Scientific Publishers ; Dordrecht and Boston : Reidel and Hingham. Sold and distributed in the U.S.A. and Canada by Kluwer Academic Publishers.

Salmon and Freshwater Fisheries Act 1975. (c.51). Available at <http://www.legislation.gov.uk/ukpga/1975/51/introduction>. (Accessed October 2010).

Salveit S.J, J.H. Halleraker, J.V. Arnekleiv and A. Harby. 2001. Field experiment on stranding in juvenile Atlantic salmon (*Salmo salar*) and brown trout (*Salmo trutta*) during rapid flow decreases caused by hydropeaking. *Regulated Rivers: Research and Management*, **17**: 609-622.

Schneider D.C. 2001. The rise of the concept of scale in ecology. *BioScience*, **51** (7): 545-553.

Scottish Natural Heritage. 2007. Atlantic salmon catchment management plan 2007. Dumfries and Galloway : Scottish Natural Heritage.

Sheu M.-l., T.w. Hu, T.E. Keeler, M. Ong and H.-Y. Sung. 2004. The 3effect of a major cigarette price change on smoking behaviour in California: a zero-inflated negative binomial model. *Health Economics*, **13**: 781-791.

Smith I.P. and G.W. Smith. 1997. Tidal and diel timing of river entry by adult Atlantic salmon returning to the Aberdeenshire Dee, Scotland. *Journal of Fish Biology*, **50**: 463-474.

Snyder d.e. 2003. Invited overview: conclusions from a review of electrofishing and its harmful effects on fish. *Reviews in Fish Biology and Fisheries*, **13**: 445-453.

Stoetaert K. and T. Petzoldt. 2010. Inverse Modelling, Sensitivity and Monte Carlo Analysis Using Package FME. *Journal of Statistical Software*, **33** (3): 1-28. Available at <http://www.jstatsoft.org/v33/i03/>.

Stokke O.S. 2000. Managing straddling stocks: the interplay of global and regional regimes. *Ocean & Coastal Management*, **43**: 205-234.

Stolte L. 1981. The forgotten salmon of the Merrimack. *U.S. Fish and Wildlife Service, Department of the Interior, Region 5*. Washington D.C: U.S Government.

Strimmer K. 2008. A unified approach to false discovery rate estimation. *BMC Bioinformatics*, **9** (303).

Suter H.C. and F.A. Huntingford. 2002. Eye colour in juvenile Atlantic salmon: effects of social status, aggression and foraging success. *Journal of Fish Biology*, **61**: 606-614.

Swain A., I.R.H. Allan and M.J. Bulleid. 1960. Recapture in the River Tweed of a Sea-trout marked in Devonshire. *Nature*, **4740**: 877.

Tang M., D. Boisclair, C. Ménard and J.A. Downing. 2000. Influence of body weight, swimming characteristics, and water temperature on the cost of swimming in brook trout (*Salvelinus fontinalis*). *Canadian Journal of Fisheries and Aquatic Sciences,* **57**: 1482-1488.

Tanguy J.M., D. Ombredane, J.L. Bagliniere and P. Prunet. 1994. Aspects of parr-smolt transformation in anadromous and resident forms of brown trout (*Salmo trutta*) in comparison with Atlantic salmon (*Salmo salar*). *Aquaculture*, **121**: 51-63.

Tarik C. Gouhier and Aslak Grinsted. 2012. biwavelet: Conduct univariate and bivariate wavelet analyses. R package version 0.12. Available at <http://CRAN.R-project.org/package=biwavelet>.

Taylor E.B. 1991. A review of local adaptation in Salmonidae, with particular reference to Pacific and Atlantic salmon. *Aquaculture*, **98**: 185-207

Thoreau H.D. 1849. *A week on the Concord and Merimack Rivers*. Edited by C.F. Hovde, W.L. Howarth and E.H. Witherell. Princeton: Princeton University Press.

Thorpe J.E. 1994. Salmonid fishes and the estuarine environment. Estuaries, Part A: Adaptation to the Estuarine Environment. *Symposium Papers from the 31st Annual Conference, Canadian Society of Zoologist*, **17** (1A): 76-93.

Thorpe J.E. 1984. *Downstream movements of juvenile salmonids: a forward speculative view*. Mechanisms of migration in fishes. Springer US, 1984: 387-396.

Tien J.H., S.A. Levin and D.I. Rubenstein. 2004. Dynamics of fish shoals: identifying key decision rules. *Evolutionary Ecology Research*, **6**: 555-565.

Ting M., Y. Kushnir, R. Seager and C. Li. 2009. Forced and Internal Twentieth-Century SST Trends in the North Atlantic. *Journal of Climate,* **22** (6): 1469–1481.

Trunk G.V. 1979. A problem of dimensionality: a simple example. *IEEE Transactions on Pattern Analysis and Machine Intelligence,* **1** (3): 306-307.

UN. 1997. Urban and Rural Areas 1996 UN, New York (1997). *United Nations publications* (ST/ESA/SER.a/166), Sales No. E97.XIII.3.

Urban D.L. 2005. Modeling ecological processes across scales. *Ecology*, **86** (8): 1996-2006.

Vadrevu K. P. and Y. Choi, 2011. Wavelet analysis of airborne CO_2 measurements and related meteorological parameters over heterogeneous landscapes. *Atmospheric Research*, **102**: 77-90.

Visbeck M.H., J.W. Hurrell, L. Polvani and H.M. Cullen. 2001. The North Atlantic Oscillation: past, present, and future. *PNAS*, **98** (23): 12876-12877.

Wagner A.K., S.B. Soumerai, F. Zhang and D. Ross-Degnan. 2002. Segmented regression analysis of interrupted time series studies in medication use research, *Journal of Clinical Pharmacology and Therapeutics*, **27** (4): 299-309.

Wang C, H. Liu and S.-K. Lee. 2010. The record-breaking cold temperatures during the winter of 2009/2010 in the Northern Hemisphere. *Atmospheric Science Letters*, **11**: 161-168.

Wang W., B.T. Anderson, R.K. Kaufman and R.B. Myneni. 2004. The relation between the North Atlantic oscillation and SSTs in the North Atlantic Basin. *Journal of Climate*, **17**: 4752-4759.

Welton J.S., W.R.C. Beaumont and M. Ladle. 1999. Timing of migration and changes in age structure of Atlantic salmon, *Salmo salar* L., in the River Frome, a Dorset stream, over a 24-year period. *Fisheries Management and Ecology*, **6** (6): 437-458.

Weinberg G.M. 1975. *An introduction to general systems thinking*. New York, USA: Dorset House Publishing.

Whalen K.G., D.L. Parrish and S.D. McCormick. 1999. Migration timing of Atlantic salmon smolts relative to environmental and physiological factors. *Transactions of the American Fisheries Society*, **128**: 289-301.

Windsor M.L., P. Hutchinson, L.P. Hansen and D.G. Reddin. 2012. *Atlantic salmon at sea: Findings from recent research and their implications for*

management. NASCO document CNL(12)60. Edinburgh, UK: North Atlantic Salmon Conservation Organization (NASCO).

Wittaker R.H. 1975. Communities and ecosystems. New York, USA: Macmillan, New York.

Wurts W.A. 2003. Daily pH cycle and Ammonia toxicity. *World Aquaculture*, **34** (2): 20-21.

Yamamura K. 1999. Transformation using (x+0.5) to stabilize the variance of populations. *Researches on Population Ecology*, **41**: 229-234.

Youngston A.F., W.C. Jordan, E. Verspoor, P. McGinnity, T. Cross and A. Fergusson. 2003. Management of salmonid fisheries in the British Isles: towards a practical approach based on population genetics. *Fisheries Research*, **62**: 193-209.

Zeileis A., C. Kleiber and S. Jackman. 2008. Regression models for count data in R. *Journal of Statistical Software*, **27** (8). Available at <http://www.jstatsoft.org/v27/i08/>.

Zivot E. 2006. State space models and the Kalman filter (Lecture). Available at <http://faculty.washington.edu>.

Zuber V. and K. Strimmer. 2009. Gene ranking and biomarker discovery under correlation. *Bioinformatics*, **25**: 2700-2707.

Zuur A.F., A.A. Saveliev and E.N. Ieno. 2012. *Zero inflated and generalized linear mixed models with R*. Newburgh, United Kingdom: Highland Statistics Ltd.

Zuur A.F., E.I. Ieno and C.S. Elphick. 2010. A protocol for data exploration to avoid common statistical problems. *Methods in Ecology and Evolution*, **1** (1): 3-14.

Zuur A.F., E.N. Ieno, N.J. Walker, A.A. Saleviev and G.M. Smith. 2009. Mixed Effects Models and Extensions in Ecology with R. New York, NY 10013, USA: Springer Science + Business Media.