Bayesian arrival model for Atlantic salmon smolt counts powered by environmental covariates and expert knowledge

Henni Pulkkinen, Panu Orell, Jaakko Erkinaro, Samu Mäntyniemi

H. Pulkkinen.\* Natural Resources Institute Finland, Paavo Havaksen tie 3 P.O. Box 413, 90014-Oulu. Finland. henni.pulkkinen@luke.fi

P. Orell. Natural Resources Institute Finland, Paavo Havaksen tie 3 P.O. Box 413, 90014-Oulu. Finland. panu.orell@luke.fi

J. Erkinaro. Natural Resources Institute Finland, Paavo Havaksen tie 3 P.O. Box 413, 90014-Oulu. Finland. jaakko.erkinaro@luke.fi

S. Mäntyniemi. Natural Resources Institute Finland, . Finland. samu.mantyniemi@luke.fi

Corresponding author: Henni Pulkkinen (phone: +358 40 485 0088)

# Abstract

Not more than 175 words

Keywords: salmon, arrival model, passage count, hierarchical Bayesian, environmental covariates, expert knowledge, biological realism, missing data

# Introduction

Limit the Introduction largely to the scope, purpose, and rationale of the study. Restrict the literature review and other background information to that needed in defining the problem or setting the work in perspective. Try beginning with the purpose or scope of the work, defining the problem next, and adding guideposts to orient the reader. An introduction generally need not exceed 375–500 words.

Migratory fish species are often monitored along their migration routes at different life stages to collect fishery-independent data for stock assessment. Monitoring methods include e.g. counting fences, video, sonar or snorkeling counts and common to all is that they rarely provide perfect information about the number of individuals passing the system (Dempson et al. 1991; Romakkaniemi et al 2000; Orell et al. 2007; 2011). Part of the fish run may be missed because of difficult environmental conditions or device failures, there may be double counting, or the data achieved may be otherwise partial or biased (e.g. Holmes et al. 2006). Thus, the assumptions made when interpreting the data can have a great influence on the estimated stock abundance.

Increasing the information beyond the data collected can improve the trustworthiness of stock assessments (Kuparinen et al. 2012): such information can be based on the resources available in other fields of biological research such as ecology or life history theory, but also on the physical and technical details of the observation processes. Moreover, methods of Bayesian inference enable both estimation of uncertainty and combination of various sources of information. Often statistical methods are used, however, without considering the biological realism and just focusing on the data analysis. Such procedure easily results in making awkward model assumptions that cannot be biologically interpreted and the potential effects of those decisions may pass undiscussed.

During the course of time, various methods have been used to estimate run dynamics and passage counts. The simplest methods are based on “connect-the-dots” type of linear interpolation (Gewin and WanHatten 2005, Johnson et al. 2007) but such approach requires an observation before and after the missing time point and thus missing tails cannot be estimated. By assuming constant proportion of run passing on a given date enables estimation of missing tails using expectation-maximization algorithms (Van Alen 2000), but considering the variability, for example, in annual environmental conditions (e.g. Orell et al. 2007; Otero et al. 2014), such assumption is problematic.

Hilborn et al. (1999) used maximum likelihood method to estimate number of salmon with ground-based stream survey, accounting also for estimate of uncertainty. Su et al. (2001) continued their work with hierarchical Bayesian approach enabling learning from years with more data to those with missing data. Furthermore, Sethi and Bradley (2016) introduced Bayesian approach to estimate missing passage at weirs with run curve model to account for arrival dynamics and process variation model to describe the observed data. While these studies account for uncertainty, they do not discuss their model assumptions against biological knowledge. This together with consideration of difference between the observation and model prediction as “noise” instead of proportion unobserved (as in Sethi and Bradley 2016), may cast some doubt on the meaningfulness of the estimated uncertainty.

In our study, we introduce a Bayesian model framework estimating the annual number of Atlantic salmon (*Salmo salar*) smolts passing the video monitoring site in river Utsjoki. The framework is built in pieces of separate sub-models that cover the processes of departing, travelling and observing. Environmental covariates are used as a source of information for key parameters of the sub-models, and expert knowledge is utilized in providing biologically meaningful model structures and prior distributions. Bayesian structure is assumed over the study years making it possible to learn from the processes, borrowing strength from data rich datasets to those with missing data. Predictive studies are made to illustrate the model performance in situations with missing data. We show that a biologically meaningful passage model can be built without assuming any specific shape for the arrival distribution and that expert knowledge is a key source of information in estimating the total uncertainty in the annual total passages especially when conditions in nature have a great influence on the probability to observe the passing individuals.

# Materials and Methods

Limit the information on materials and methods to what is needed in judging whether the findings are valid. To facilitate assessment, give all the information in one section when possible. Refer to the literature concerning descriptions of equipment or techniques already published, detailing only adaptations. Often, it helps to begin statements on procedures with a phrase indicating the purpose, such as “To determine … we …”. If the section is long, consider using subheadings corresponding to headings for the findings.

Data

River Utsjoki is a tributary of the large River Teno in the northernmost border of Finland and Norway. Each spring, the monitoring system of 8 video cameras is set at the river bed under a bridge close to the tributary outlet (see Orell et al. 2011; Fig 1). The video footage provides data on both smolts decending and adults ascending the river during the course of summer. In this study we analyze data from years 2005-2009 and 2014. These years were chosen to demonstrate the highly variable environmental conditions, especially as flow velocity varied strongly in those years (Figure 2). The data are aggregated in daily counts over 61 days of June-July.

Data on environmental covariates contain daily air temperature and flow velocity measurements. These datasets are aggregated into daily averages. Both covariates are measured near the video site, although conditions may differ between the video site and upper parts of the stream from where the smolts begin their migration. These datasets are, however, considered as reasonable proxies for the environmental conditions affecting the migration behavior of the smolts.

Data from the first 23 days of June in 2005 is missing because the high level of water delayed the setup of the monitoring system in that year. Thus, the missing counts are predicted for those dates. Predictive studies are performed by treating data in 2007 and 2014 as partly missing: For 2007 the first 17% of the run strength and for 2014 the peak +-2 days count is removed from the analysis. The model predicted counts for those dates are compared with the observed data to evaluate the predictive performance of the model.

Expert elicitation as a source of informative priors

Model structure and informative priors were elicited from an expert that is most familiar with the behavior of salmon smolts and the video monitoring system at Utsjoki. The expert was asked to base his views on his background knowledge about the processes in question instead of considering the particular data that will included in the study. Elicitation was carried out iteratively, first by asking preliminary questions on, e.g. biological process in question, and then providing graphical illustration of the parameters in question. Changes were made in the prior distributions until the expert agreed that the illustration is in line with his views on the topic. Some examples of questions elicited from the expert are listed in table 1.

Statistical models

Next, the main pieces of the model framework are described, consisting of the processes of:

1. Smolts making the decision to depart while daily temperature affects the probability to depart in a given day;
2. Time it takes in days for an average smolt to arrive at the video site after migration decision has been made, daily flow velocity affecting the travel time;
3. Observation process in which flow (/level of water) influences visibility and the probability that an individual smolt passing the site is observed;

Process of departing

**An individual smolt’s probability to begin the migration in day given that the smolt has not departed yet is considered to depend on the temperature on that day. We assume a logit-normally linear relationship for the smolt’s probability to depart and temperature :**

**Informative prior distributions are given to parameters , and according to expert view (see following section about expert elicitation and Tables 1 and 2).**

**The daily probabilities of departing (given those individuals have not departed yet; ) can be transformed into proportions of the total smolt run that depart each day () as**

1. **.**

Process of travelling

According to the expert view, each smolt should arrive at the video site within 14 days after departing. Thus we assume that the smolts that depart on day pass the video site in one of the next 14 days according to cumulative distribution function of lognormal distribution:

1. .

Here is the cumulative density function of the standard normal distribution, is the mean of log(travel time) to video site of smolts that depart on day , and is the standard deviation of log(travel time) (Schwarz & Dempson 1994, Mäntyniemi & Romakkaniemi 2002). Thus, is the proportion of smolts that depart on day and arrive on day , when goes from to

To ensure that all smolts arrive within 14 days of departing, ’s are standardized:

1. ,

and thus

It is further considered that increasing flow velocity increases the speed at which the smolts travel, shortening their travel time. The expected travelling time on real scale is assumed to follow a lognormal distribution with expected value and coefficient of variation and it is assumed to depend on the flow velocity ():

1. **).**

Again, informative prior distributions are given for parameters **, and according to expert view (see tables 1 and 2).**

Arrival distribution

**Total annual size of the smolt run passing the video site on is assumed to follow a uniform distribution on a log scale:**

1. **.**

**Parameterisation at the log-scale allows for a wide, minimally informative prior distribution for the annual total abundance. The annual run size is further considered to be distributed over 61 days of June-July according to Dirichlet-multinomial distribution. Dirichlet-multinomial is a multivariate version of a beta-binomial distribution, meaning that instead of two potential outcomes (as with binomial distribution) there can be any finite number of potential outcomes, number being 61 in this case. For computational simplicity, we approximate multinomial distribution as**



**Here ’s are the** proportions of smolts passing the video site each day, assumed to follow a Dirichlet distribution:

1. **,**

**where is the expected proportion of smolts passing the video site in day and is the overdispersion parameter. To ease the computation, however, we approximate the Dirichlet-distribution with a set of lognormal** distributions **(see appendix A for details).**

**When combining the processes of departing and travelling described earlier in this chapter, we achieve the joint distribution:**

1. **.**

**With the set of parameters we can finally formulate the arrival distribution, i.e. the expected proportion of smolts passing the video site each day:**

Observation process

In previous sections we have covered the processes assumed to affect the timing of the smolt passage to the video site. To combine the observed number of smolts at the video monitoring experiment, process of observing deeds to be included. It seems natural to assume that some of the individuals pass the site unobserved, and thus this proportion must be acknowledged when total run size is estimated. Next, we introduce two versions for observation process.

The first version is simple yet not very realistic. It assumes that the number of smolts observed at the video site on day in year ( follows Beta-binomial distribution with fixed proportion of unobserved:

1. .

Here is the total number of smolts passing the video site and numerical constants indicate that the probability that an individual smolt is observed follows distribution having mean 0.91 and standard deviation 0.03. Thus, 9% of smolts would be assumed to pass the video site unobserved on average, with 95% PI of [5%, 14%]. Albeit an observation model with fixed probability for observing is not realistic, such can be useful for illustrating how the processes of departing and travelling affect the model results yet acknowledging that the process of observing is not perfect.

The second version for observation process aims at higher realism, accounting for expert views on how high the probability to observe a smolt may be in excellent vs. poor circumstances and how the level of water in the stream would affect this probability. Again the number observed follows a Beta-binomial distribution:

1. .

Now the probability to observe a smolt follows a distribution, where is the expected probability for a smolt to be observed and is the overdispersion parameter. The expected value of the process is linked to flow velocity following the expert view that in very good visibility (low flow velocity) at maximum 90% of the smolts are observed. As the flow velocity increases the visibility decreases and again, according to the expert, observation probability decreases gradually at minimum of 30%. The expected probability is considered to follow logit-normally linear relationship on the interval (0.3,0.9):

1. and

where , and have prior distributions according to the expert view.

Later in this paper we will refer to the model with simple, fixed, observation process as model 1 and to the observation model with expert knowledge and environmental covariate as model 2.

Models 1 and 2 (with simple vs. realistic observation process, respectively) were fitted using MCMC sampling with JAGS 4.2.0 software (Just Another Gibbs Sampler; Plummer 2003) and R runjags package (Denwood 2016). Both models were run with 2 chains for 1 500 000 iterations this taking around 3 days per model. Convergence was diagnosed using Gelman-Rubin statistics (REFXX). The complete JAGS code for the model structure can be found from appendix B.

# Results

Limit the results to answers to the questions posed in the purpose of the work and condense them as comprehensively as possible. Give the findings as nearly as possible in the terms in which the observations or measurements were made so as to avoid confusion between facts and inferences. State noteworthy findings to be noted in each table and figure, and avoid restating in the text what is clear from the captions. Material supplementary to the text can be archived in the report literature or a recognized data depository and referenced in the text (see Supplementary material section)

Posterior estimates of model 2 indicate that annually on average 25-75% of smolts pass the video site unobserved (Figure 3). Considering the estimated uncertainty among the years without missing data, proportion of unobserved is 90% at most (in 2008 upper limit for 90% PI is 16000 smolts vs observed data of 8400 smolts). Posterior estimates of daily arrivals are illustrated in figures 4 (model 1) and 5 (model 2).

Table 3 shows the predicted posterior distributions for the time periods with missing counts as well as the observed annual counts for comparison when available. Model 2 predicts 24-33% higher number of smolts compared to model 1 as a result of more realistic observation model. Scenarios to study predictive performance of the model contain ‘first 17% of missing’ -scenario for 2007 and ‘peak day +- 2 days missing’ –scenario for 2014. In both cases, all daily counts fall in the predicted 90% PI indicating good predictive performance (Table 3). The percentage of predicted out of the estimated annual total is close to identical between models in all 3 cases (Table 3).

Figures 6 and 7 illustrate the prior and posterior distributions from model 2 for expected processes of departing and travelling. Posterior distributions from model 1 are nearly identical and thus not illustrated here. The expected probability to begin the migration at given temperature updates quite heavily from the prior distributions, supporting 4-5 degrees higher temperature at which the migration sets off (Figure 6). The posterior cumulative distribution of the expected travel time to the video site (Figure 7) supports somewhat longer travel times than expected a priori both for low (10m3/s) and high (100m3/s) flow velocities. Note that posterior cumulative distributions at these specific flow velocities are chosen for illustrative purposes only and that in practice the models allow any positive value for flow velocity.

In model 2, also the observation probability depends on the flow velocity. Figure 8 illustrates the prior and posterior distributions of the observation process given the flow velocity. Priors and posteriors are close the same, indicating that this model framework does not enable further learning about the process of observing.

# Discussion

Limit the Discussion to giving the main contributions of the study and interpreting particular findings, comparing them with those of other workers. Emphasis should be maintained on synthesis and interpretation and exposition of broadly applicable generalizations and principles. If these are exceptions or unsettled points, note them and show how the findings agree or contrast with previously published work. Limit speculation to what can be supported with reasonable evidence. End the Discussion with a short summary of the significance of the work and conclusions drawn. If the Discussion is brief and straightforward, it can be combined with the Results section.

In our study, we have broken down an arrival model for Atlantic salmon smolts separating it to departing, travelling and observation processes. This distinction makes it possible to go beyond assuming that fish count data as such follows mathematical rules and instead concentrating on biologically meaningful mathematical assumptions that are related to each sub-process. The model framework aims at biological realism by accounting for expert knowledge in the model formulation (including choices for likelihood functions and prior distributions) and in the way environmental covariates are utilised in providing information about the timing of the migration. The estimated total uncertainty accounts for the uncertainty in the sub-processes, including the observation process in which variability in the environmental conditions (i.e. level of water/flow velocity) are known to have a large impact on the probability to observe passing smolts.

One of the most important advantages of our approach is that the need to assume a specific shape for the arrival distribution is avoided. Past studies (Hilborn et al. 1999, Su et al. 2001, Sethi & Bradley 2016) have assumed unimodal shape (e.g. normal, skew-normal, Student’s t) for the arrival curve although when taking a look at the shape of (any) arrival time series it seems clear that the nature does not care to follow such rules and thus the assumptions are tolerable simplifications at best (see e.g. Figure 2 in Sethi & Bradley 2016).

Rather little attention is paid in past studies to the role of observation models in estimating the total uncertainty. Hillborn et al. 1999 distinguishes process and observation errors, assuming observer efficiency either as known or unchanging in time. They also suggest that weir studies in index streams should be used to estimate observer bias in other streams. Su et al. 2001 discuss the issue of error in the observation process and mainly conclude that such can cause convergence problems in sampling. Sethi and Bradly (2016) assume either normal or negative Binomial models for the observed passage counts, both of which allow for the “true” passage to be greater or smaller than the observed. In our opinion, however, it would seem more realistic to consider the number of individuals observed as a minimum estimate for the true number especially if movement of individuals downstream is negligible. Depending on the approach chosen, the estimate of total annual passage can be very different. As Hilborn et al. (1999) put it, ‘when we admit uncertainty in observer efficiency we become much less certain about the actual escapement’.

Sethi and Bradley (2016) fitted datasets in their study using several combinations of arrival and process error models. They utilized information-theoretic model selection (DIC, Spiegelhalter et al. 2002) to find out the best fitting model for each dataset. However, no discussion took place whether some of these models would be biologically more reasonable than another. We argue further that the Bayesian Model averaging (BMA, Hoeting et al. 1999) is more suitable method to consider model uncertainty compared to DIC. Instead of calculating information criteria separately for a dataset under alternative models, these alternatives can be defined simultaneously under a single model framework and the posterior probabilities for those alternatives calculated (Pulkkinen and Mäntyniemi 2013). Inclusion of hierarchical structure over exchangeable datasets studying which of the alternative models supports the whole dataset the best minimizes the chance that a selected model only fits a set of data by chance.

Su et al. (2001) used hierarchical structure in their study, assuming that years in which timing of the run is well known can be used for passing information to years with missing data. We argue that learning should rather take place over the (biological) process, assuming exchangeability over year specific parameters of those processes. However, our model structure is not hierarchical, as we simply assume that the underlying processes are the same from year to year allowing between- and within-years variation. If there were a need for a meta-analysis of similar studies over several stocks of the same species, exchangeability in biological parameters could be considered between the stocks. In fact, such approach could allow meta-analysis of data from different types of monitoring systems if detailed observation processes were tailored for each system and exchangeability was considered as reasonable assumption for the biological processes.

Our study could be fine-tuned and extended further. Six years of data were chosen to illustrate the approach, but for the needs of stock assessment full time series of data (15 years) will be analysed. Because of a lack of several experts from which to elicit from, prior understanding of the processes is based only on expert knowledge of one person, although it would be desirable to combine knowledge from several experts as the views of uncertainty tend to be greater between experts than within them (Uusitalo et al. 2005). In an ideal situation, environmental covariates would be available from upstream locations from where the smolts originate. Data on air temperature would be replaced with water temperature to better explain the physiological processes the smolts encounter. To increase the realism even more, the schooling behavior of smolts could be accounted for. Salmon smolts have a tendency to school during their migration (e.g. Bakshtanskiy et al. 1988) and this creates more variation in the monitoring data compared to if those would move independently (Mäntyniemi and Romakkaniemi, 2002). For Utsjoki video survey, daily average school sizes of smolts are recorded and incorporating those into the observation model framework could be relatively simple.

Despite these issues, we claim that our model framework is reasonable and most of all, contains the elements of realism that are required to provide reliable estimates for the strength of total annual smolt run. This approach can be easily applied to similar monitoring surveys, those being, for example, sonar, trap, or weir studies. Environmental covariates are utilized to estimate both the dynamics of the passage and the observation process and as far as we know, such has not been included in the passage models so far.

# Acknowledgements

Acknowledgements should be written in the third person. We strongly urge authors to limit acknowledgments to those who contributed substantially to scientific and technical aspects of the paper, gave financial support, or improved the quality of the presentation. Avoid acknowledging those whose contribution was clerical only.

# References

Bakshtanskiy, E., Nesterov, V., and Neklyodov, M. 1988. Development of schooling behaviour in juvenile Atlantic salmon, Salmo salar, during seaward migration. J. Ichthyol. 28: 91–101.

Dempson, J. B, and Stansbury, D. E. 1991. Using partial counting fences and atwo-sample stratified design for mark–recapture estimation of an Atlantic salmonsmolt population. North American Journal of Fisheries Management 11, 27–37.

Denwood, M.J. 2016. runjags: An R Package Providing Interface Utilities, Model Templates, Parallel Computing Methods and Additional Distributions for MCMC Models in JAGS. Journal of Statistical Software, 71(9), 1-25. doi:10.18637/jss.v071.i09.

Hilborn, R., Bue, B.G., and Sharr, S. 1999. Estimating spawning escapements from periodic counts: a comparison of methods. Can. J. Fish. Aquat. Sci. 56(5): 888–896. doi:10.1139/f99-013.

Holmes, J.A, Cronkite, G.M.W., Enzenhofer, H.J., and Mulligan, T.J. 2006. Accuracy and precision of fish-count data from a “dual-frequency identification sonar” (DIDSON) imaging system, ICES Journal of Marine Science. 63: 543–555, https://doi.org/10.1016/j.icesjms.2005.08.015

Kuparinen, A., Mäntyniemi, S., Hutchings, J.A., and Kuikka, S. 2012. Increasing biological realism of fisheries stock assessment: towards hierarchical Bayesian methods. Environ. Rev. 20: 135–151. doi:10.1139/A2012-006.

Mäntyniemi, S., and Romakkaniemi, A. 2002. Bayesian mark-recapture estimation with an application to a salmonid smolt population. Can. J. Fish. Aquat. Sci. 59: 1748-1758. Doi:10.1139/F02-146.

Orell, P., Erkinaro, J., Svenning, M., Davidsen, J., and Niemelä, E. 2007. Synchrony in the downstream migration of smolts and upstream migration of adult Atlantic salmon in the sub-Arctic River Utsjoki. Journal of Fish Biology 71: 1735–1750.

Orell, P., Erkinaro, J., and Karppinen, P. 2011. Accuracy of snorkelling counts in assessing spawning stock of Atlantic salmon, Salmo salar, as verified by radio-tagged fish and underwater video monitoring. – Fisheries Management and Ecology 18: 392–399.

Otero, J., L'Abée-Lund, J.H., Castro-Santos, T., Leonardsson, K., Storvik, G.O., Jonsson, B., Dempson, B., Russell, I.C., Jensen, A.J., Baglinière, J.-L., Dionne, M., Armstrong, J.D., Romakkaniemi, A., Letcher, B.H., Kocik, J.F., Erkinaro, J., Poole, R., Rogan, G., Lundqvist, H., MacLean, J.C., Jokikokko, E., Arnekleiv, J.V., Kennedy, R.J., Niemelä, E., Caballero, P., Music, P.A., Antonsson, T., Gudjonsson, S., Veselov, A.E. Lamberg, A., Groom, S., Taylor, B.H., Taberner, M., Dillane, M., Arnason, F., Horton, G., Hvidsten, N.A., Jonsson, I.R., Jonsson, N., McKelvey, S., Næsje, T., Skaala, Ø., Smith, G.W., Sægrov, H., Stenseth, N.C. and Vøllestad, L.A. 2014. Basin-scale phenology and effects of climate variability on global timing of initial seaward migration of Atlantic salmon (Salmo salar). Global Change Biology, 20: 61-75.

Plummer, M. 2003. JAGS: A program for analysis of Bayesian graphical models using Gibbs sampling. In proceedings of the 3rd International Workshop on Distributed Statistical Computing (DSC 2003), 20-22 March, Vienna, Austrua. ISSN 1609-395X.

Pulkkinen, H., and Mäntyniemi, S. 2013. Maximum survival of eggs as the key parameter of stock–recruit meta-analysis: accounting for parameter and structural uncertainty. Can. J. Fish. Aquat. Sci. 70: 527–533 (2013) dx.doi.org/10.1139/cjfas-2012-0268.

Romakkaniemi, A., Lilja, J., Nykänen, M., Marjomäki, T.J. and Jurvelius, J. 2000. Spawning run of Atlantic Salmon (Salmo salar) in the River Tornionjoki monitored by horizontal split-beam echosounding. Aquatic Living Resources 13: 349-354,

Schwarz, C., and Dempson, J. 1994. Mark-recapture estimation of a salmon smolt population. Biometrics, 50: 98-108.

Sethi, A.S., and Bradley, C. 2016. Statistical arrival models to estimate missed passage counts at

fish weirs. Can. J. Fish. Aquat. Sci. 73: 1251–1260. dx.doi.org/10.1139/cjfas-2015-0318.

Spiegelhalter, D.J., Best, N.G., Carlin, B.P., and van der Linde, A. 2002. Bayesian measures of model complexity and fit (with discussion). J. R. Stat. Soc. Ser. B, 64: 583–639.

Su, Z., Adkison, M.D., and Van Alen, B.W. 2001. A hierarchical Bayesian model for estimating historical salmon escapement and escapement timing. Can. J. Fish. Aquat. Sci. 58(8): 1648–1662. doi:10.1139/f01-099.

Uusitalo, L., Kuikka, S., and Romakkaniemi, A. 2005. Estimation of Atlantic salmon smolt carrying capacity of rivers using expert knowledge. ICES J. Mar. Sci. 62: 708-722. doi:10.1016/j.icesjms.2005.02.005.

Table 1. Examples of questions elicited from the expert

|  |  |
| --- | --- |
| Question | Expert view |
| Process of departing |  |
| * **At what range of temperature (air temperature at Utsjoki logger) first smolts are considered to begin their migration?** | * 1. °C |
| * **What level of temperature would be considered so high that the process of migration not be influenced by further increase in temperature (i.e. logit curve levels out)?** | 12 °C |
| Process of travelling |  |
| * **What is the the maximum traveling time (in days) in which a smolt can be considered to arrive at the video site?** | 14 days |
| * **What is the minimum travelling time in which a smolt can be considered to arrive at the video site?** | 1-2 days |
| Process of observing |  |
| * **What is the upper (/lower) limit for the expected probability at which an average smolt is considered to be observed in best (/worst) possible environmental conditions, i.e. at minimum (/maximum) flow velocity?** | 90% (30%) |
| * **At what level of flow visibility can be considered optimal?** | 20m3/s |
| * **At what level of flow visibility is considered to be poorest?** | 50-60m3/s |
|  |  |

Table 2: List of symbols, descriptions and prior distributions

|  |  |  |
| --- | --- | --- |
| Symbol | Description | Prior distribution |
|  | Total annual number of smolts on log scale | **Uniform(7,15)** |
|  | Total annual number of smolts on real scale | - |
|  | Daily proportion of smolts arriving at video site | - |
|  | Expected proportion of smolts arriving at video site | - |
|  | Overdispersion in the arrival process | **Uniform(0.001,100000)** |
|  | Probability to begin migration (if haven’t done so earlier) | - |
|  | Proportion of smolts departing each day | - |
|  | Intercept for prob to depart given temperature | N(-20,1) |
|  | Regression coefficient for prob to depart given temperature | logN(0.6,0.1) |
|  | Standard deviation of probability to depart | logN(0,1) |
|  | Unstandardised proportion for travel time | - |
|  | Standardised proportion for travel time | - |
|  | Mean of log(travel time) | - |
|  | Standard deviation of log(travel time) |  |
|  | Coefficient of variation of log(travel time) | **Uniform(0.001,2)** |
|  | Intercept for expected travel time given flow | logN(0.52,0.07) |
|  | Regression coefficient for expected travel time given flow | logN(-4.6,0.04) |
|  | Coefficient of variation of expected travel time | **Uniform(0.001,1)** |
|  |  |  |
| Data |  |  |
|  | Daily (air) temperature | - |
|  | Daily flow | - |
|  | Daily number of smolts observed | - |
|  |  |  |

Table 3. Results from the predictive study in 2005 (real missing data), 2007 (‘first 17% missed’ -scenario) and 2014 (‘peak of the migration +-2 days missing’ -scenario).

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | Sum of missing counts | % of missing counts out of total observed |  | model 1 | |  | model 2 | |
|  |  | Number of predicted | % of predicted out of total |  | Number of predicted | % of predicted out of total |
| Year |  |  |
| 2005 | NA | NA |  | 1620 (340,3930) | 10% (2%,22%) |  | 2020 (450,4850) | 10% (3%,22%) |
| 2007 | 2539 | 17% |  | 2500 (510,6220) | 14% (4%,31%) |  | 3220 (700,7990) | 15% (4%,32%) |
| 2014 | 4085 | 21% |  | 3790 (2030,6040) | 18% (11%,26%) |  | 4700 (2620,7390) | 18% (11%,26%) |

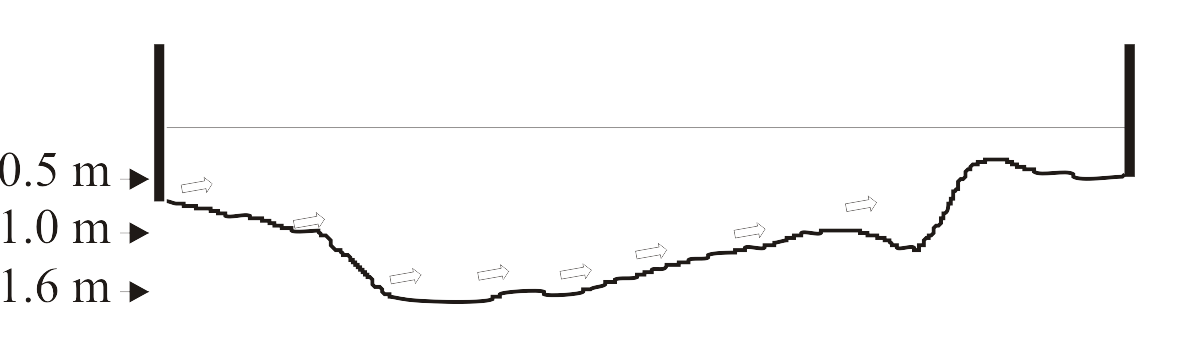
Figure 1: Setup of the video cameras at the river bed in Utsjoki.

Figure 2. Data on daily mean air temperature and mean flow velocity at Utsjoki on study years

 Figure 3. Annual number of smolts passing the video site. Posterior distributions of model 1 (grey boxplots), model 2 (black boxplots) and count data (black dots).

 Figure 4. Daily number of smolts that pass the video site. Posterior distributions of model 1 (boxplots) and count data (dots). Data points included in the study are marked with black and the ones excluded (counts treated as missing) marked with grey.

 Figure 5. Daily number of smolts that pass the video site. Posterior distributions of model 2 (boxplots) and count data (dots). Data points included in the study are marked with black and the ones excluded (counts treated as missing) marked with grey.



Figure 6. Expected probability that a smolt begins the migration at given temperature. Prior (grey boxplots) and posterior distributions (black boxplots) from model 2.

 Figure 7. Cumulative distribution of expected travel time in days from the point of departure to the video site. Above: Cumulative distribution at low flow velocity (10 m3/s). Below: Cumulative distribution at high flow velocity (100 m3/s). Prior (grey boxplots) and posterior distributions (black boxplots) from model 2.



Figure 8. The expected observation probability at given flow velocity. Prior (grey boxplots) and posterior distributions (black boxplots) from model 2.

# Appendix A: Lognormal approximation for Dirichlet-distribution

Let’s consider a set of Dirichlet-distributed parameters:

(A.1) **,**

**Where is the expected probability for slot and is the overdispersion parameter. Such Dirichlet-distribution can be approximated with a set of lognormally distributed by assuming**

(A.2)

**(A.3)**

**(A.4) and**

**(A.5)**

# Appendix B: JAGS code for the arrival model