

가우시안 프로세스 회귀를 통한 열 성장 계수 기반의 어류 성장 예측 모델[☆]

TGC-based Fish Growth Estimation Model using Gaussian Process Regression Approach

성 주 형¹ 조 성 윤¹ 정 다 은¹ 김 중 원² 박 정 환³ 권 기 원^{1*} 고 영 명^{2*}
Juhyoung Sung Sungyoon Cho Da-Eun Jung Jongwon Kim Jeonghwan Park Kiwon Kwon Young Myoung Ko

요 약

최근 수산 자원의 고갈에 따른 육상 양식장에서의 '기르는 어업'에 의한 생산성 향상에 대한 기대가 크게 고조되고 있다. 육상 양식장의 경우, 해상 환경과 달리 환경 및 양성 요소에 대한 제어와 관리가 용이하며, 출하 계획에 따른 생산량 조정이 가능한 이점이 있다. 반면, 자연 환경에서와 달리 어류 성장을 위한 인위적인 관리가 필요하기 때문에 운영에 따른 비용이 크게 증가할 수 있는 단점이 있다. 따라서, 계획된 목표 출하량에 맞추어 효율적으로 양식장을 운영함으로써 이윤 극대화를 추구할 수 있다. 이러한 효율적인 양식장 운영 및 어류 양성을 위해서는 대상 어종에 따른 정확한 성장 예측 모델이 반드시 요구된다. 현재까지 대부분의 성장 예측 모델은 양식장 수집 데이터를 활용하여 통계적 분석 기반의 수치 해석적인 결과들이 주를 이룬다. 본 논문에서는 기존의 통계적 관점에 의한 성장 예측 모델이 가질 수 있는 데이터 확보의 어려움 및 낮은 정확도에 대한 정량적 수치를 제공하기 어려운 단점을 극복하기 위해 확률적 관점에서의 성장 예측 모델을 제시한다. 확률적 접근을 위하여 양성에 가장 중요한 요소인 수온을 기반으로 한 가우시안 프로세스 회귀 방식을 도입하여 모델링을 수행한다. 이를 통해, 특정 시점에서의 성장 예측값에 대한 평균치와 해당 값에 대한 신뢰구간을 동시에 제공함으로써 보다 효율적인 양식장 운영을 위한 참고 수치를 제공할 수 있을 것으로 기대한다.

☞ 주제어 : 가우시안 프로세스, 열 성장 계수, 성장 예측 모델, 성장 데이터, 수온

ABSTRACT

Recently, as the fishery resources are depleted, expectations for productivity improvement by 'rearing fishery' in land farms are greatly rising. In the case of land farms, unlike ocean environments, it is easy to control and manage environmental and breeding factors, and has the advantage of being able to adjust production according to the production plan. On the other hand, unlike in the natural environment, there is a disadvantage in that operation costs may significantly increase due to the artificial management for fish growth. Therefore, profit maximization can be pursued by efficiently operating the farm in accordance with the planned target shipment. In order to operate such an efficient farm and nurture fish, an accurate growth prediction model according to the target fish species is absolutely required. Most of the growth prediction models are mainly numerical results based on statistical analysis using farm data. In this paper, we present a growth prediction model from a stochastic point of view to overcome the difficulties in securing data and the difficulty in providing quantitative expected values for inaccuracies that existing growth prediction models from a statistical point of view may have. For a stochastic approach, modeling is performed by introducing a Gaussian process regression method based on water temperature, which is the most important factor in positive growth. From the corresponding results, it is expected that it will be able to provide reference values for more efficient farm operation by simultaneously providing the average value of the predicted growth value at a specific point in time and the confidence interval for that value.

☞ keyword : Gaussian process, thermal growth coefficient, growth prediction model, growth data, water temperature

¹ Smart Network Research Center, Korea Electronics Technology Institute, Seoul, 03924, Korea.

² Pohang University of Science and Technology, Pohang, Gyeongbuk, 37673, Korea.

³ Pukyong National University, Busan, 48513, Korea.

* Corresponding authors (kwonkw@keti.re.kr, youngko@postech.ac.kr)

[Received 22 December 2022, Reviewed 06 January 2023, Accepted 1 February 2023]

☆ 이 논문은 2022년도 정부(해양수산부)의 재원으로 해양수산 과학 기술진흥원의 지원을 받아 수행된 연구임(No.20200448, 어류양식장용 생산성 안정과 자동화 장비 위협요소 관리를 위한 생물/장비 예지보전 기술 및 교구/교재 개발)

1. 서 론

최근 세계적인 인구 증가 및 남획에 따른 수산 자원 고갈의 우려가 심화되는 환경에서 기존의 ‘잡는 어업’에서 ‘기르는 어업’으로의 변화에 대한 기대가 크게 증가하고 있다. 특히, 해양 양식 대비 양식 환경 제어 및 사료와 배설물 등에 의한 오염 요소 관리에 용이한 육상 양식 산업을 통한 어업 생산성 향상에 대한 기대가 증대되고 있으며 이를 위한 스마트 양식의 필요성이 대두되고 있다 [1, 2].

스마트 양식의 고도화를 위해서는 양식장 환경 최적화를 위한 제어 기반의 ICT 기술 뿐만 아니라 수산 관점에서 양식 어종에 대한 정확한 성장률 예측 기술 연계 개발이 핵심적인 요구사항이다 [3-5]. 본 논문에서는 어류 성장률 예측 모델링에 대하여 기존의 통계학적 데이터를 바탕으로 한 모델링과 대비하여 확률적 접근을 통한 새로운 모델링 기법을 제시한다.

어류의 성장에 영향을 끼치는 변수는 크게 다음과 같은 2가지로 구성된다. 양식 대상 종에 대한 생물학적인 변수와 양식장에 대한 물리적인 변수이다. 따라서, 양식 기반의 어류 생산을 위해, 급격하게 변화하는 양식장 환경 조건을 고려하여 안정적인 수산물 생산량 확보를 위한 주요한 접근 방식을 통한 연구가 필요하다.

어류의 양식에 있어 육상 양식 방법은 주변 환경을 조절할 수 있다는 장점이 있지만 어류 생산에 사용되는 비용이 높게 책정된다는 단점이 있다. 예를 들어, 유수식 양식 환경에서는 해수를 사용하여 양식장에 대한 물리적인 변수인 온도, 수질 등이 지속적으로 변화한다는 특징을 고려해야 한다. 따라서, 육상 양식 환경에서 다양한 변수 조건에 따른 어류의 성장 예측 모델을 생성하게 되면 수요 예측에 따른 최적의 양식장 제어 및 운영 관리를 통하여 비용 절감 및 효율성 향상을 기대할 수 있다. 일반적인 어류 성장 예측 모델은 특정 시점에서의 기대되는 예상 어류의 무게를 예측하는 모델로, 다양한 생물학적 변수 및 물리적인 변수가 복합적으로 연계되어 있다 [6, 7]. 한편, 기존의 통계 분석 기반 어류 성장 예측 모델은 모델이 도출한 예측값에 대한 신뢰도를 측정할 수 없다는 문제점이 있다. 특히, 노이즈가 포함된 부정확한 수집 데이터를 이용할 경우, 예측값의 정확도가 현저하게 떨어지게 된다는 단점이 있다. 본 논문에서는 다양한 양식 환경 변수 중, 적정 사료 공급 및 생육 환경이 제공되는 환경에서 성장에 가장 중요한 환경 변수인 수온을 주요 변수로 중심으로 한 입의의 시점에서의 성장 예측 모

델을 제안한다. 특히, 기존의 통계 분석 기반 어류 성장 모델의 단점과 부족한 현장 데이터에 의한 모델링의 정확도의 문제를 개선하고자 가우시안 프로세스(Gaussian Process, GP) 회귀(regression) 기반의 변동성을 반영하여 확률적 접근(probabilistic approach)을 통한 어류 성장 예측 모델링 방법을 제안한다.

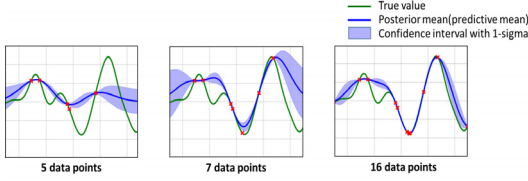
본 논문에서는 열 성장 계수(Thermal Growth Coefficient, TGC) 기반의 수온을 주요 변수로 한 GP 회귀 기반 어류 성장 예측 모델 설계 방법을 제안한다. 추가로, 어류 성장 예측 모델에 대한 성능 검증을 수행한다. 본 논문의 구성은 다음과 같다. 2절에서는 GP 회귀의 배경 이론 및 GP 회귀 기반 어류 성장 예측 모델의 구조에 대해 소개한다. 3절에서는 GP 회귀 기반의 어류 성장 예측 모델 설계 방법을 제안한다. 4절에서는 본 논문에서 제안하는 어류 성장 예측 모델에 대하여 어류의 1년 주기 동안의 증량을 통한 성능 분석을 수행한다. 5절에서는 본 논문에 대한 결론 및 향후 연구개발 방향을 제시한다.

2. GP 회귀 기반 어류 성장 예측 모델 개요

본 절에서는 GP 회귀의 배경 이론 및 GP 회귀 기반의 어류 성장 예측 모델의 입력력 구조 및 동작 흐름에 대해 소개한다.

2.1 GP 회귀 배경 이론

GP란, 확률 과정의 일종으로 유한개의 확률 변수를 선택하였을 때, 그 확률 변수들이 다변량 정규분포(multivariate normal distribution)를 따르는 확률 과정이다. GP 회귀는 GP를 사전 분포(prior distribution)로 하여 주어진 데이터에 따라 사후 분포(posterior distribution)를 계산하는 회귀 분석 방법이다 [8]. GP 회귀를 통해 학습된 모델은 각 입력값에 따른 예측값과 이에 대한 신뢰 구간(confidence interval)의 정보를 제공한다. 그림 1에 훈련 데이터의 숫자에 따른 GP 회귀 수행결과를 도시하였고, 빨간 점은 훈련 데이터이며 훈련 데이터의 수가 증가할수록 파란색으로 표시된 사후 분포의 평균이 녹색으로 표시된 실제 값에 가까워지며 신뢰구간 내에 실제 값이 분포하는 것을 확인할 수 있다.



(그림 1) GP 회귀에 의한 예측 모델 예시

(Figure 1) Example of GP regression based prediction model

GP 회귀를 위한 입력 변수는 벡터의 형태로 $X = (x_1, x_2, \dots, x_n)^T$ 으로 표기한다. 이에 대한 출력 변수로는 Y 로 표기하며 이에 해당하는 입력출력 관계식은 다음과 같이 표기한다.

$$Y = f(X) + \epsilon \quad (1)$$

여기서, $f(X) \sim \mathcal{N}(\mu(X), K(X, X))$ 의 분포를 따르고, $\epsilon \sim \mathcal{N}(0, \lambda^2 I)$ 의 분포를 따른다. 여기서 사용자가 지정하는 파라미터로는 μ, K, λ 이며, 임의의 값을 정의할 수 있고, 최대우도법(Maximum Likelihood Estimation, MLE)을 통해 값을 추정할 수 있다. 각 파라미터에 따른 세부 정의는 다음과 같다.

1) 평균 함수(Mean function, μ) : GP 회귀 모델을 훈련하기 이전 확률 변수 분포의 평균 함수이다. 본 논문에서는 어류 성장 예측 모델 개발 과정을 위해, 훈련 출력값의 평균값을 갖는 평균 함수를 사용하였다.

2) 잡음 분산(Noise variance, λ^2) : 출력값이 갖는 잡음 분산의 크기를 의미한다.

3) 커널 함수(Kernel function, K) : GP에서 커널 함수(혹은 covariance function)은 두 개의 입력 변수에 대해 양의 정부호(positive definite, PD)이며 대칭 함수(symmetric function) 조건을 만족해야 하며, 다음과 같이 계산된다.

$$K(X, X) = \begin{pmatrix} K(x_1, x_1) & \dots & K(x_n, x_1) \\ \vdots & \ddots & \vdots \\ K(x_1, x_n) & \dots & K(x_n, x_n) \end{pmatrix} \quad (2)$$

본 논문에서 어류 성장 예측 모델 개발을 위해 정상성(stationary), 비퇴화(non-degenerative) 등의 안정적인 특징을 갖는 방사형 기저 함수(Radial basis function, RBF) 커널을 사용하였다. 크기 m 의 벡터 x 에 대해,

RBF 커널은 파라미터로 $\sigma_f^2, l_1, \dots, l_m$ 를 가지며 식은 다음과 같다.

$$K(x, x') = \sigma_f^2 \exp\left(-\frac{1}{2} (x - x')^T \begin{pmatrix} l_1 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & l_m \end{pmatrix}^{-1} (x - x')\right) \quad (3)$$

따라서, 평균 함수를 특정하였을 경우, GP 회귀에서 결정해야 하는 파라미터 중 최적화를 통해 도출이 필요한 파라미터는 커널 함수의 파라미터와 잡음 분산을 의미하는 $\sigma_f^2, l_1, \dots, l_m, \lambda$ 이다. 본 논문에서는 기계학습 기법을 사용하여 [9,10] $\sigma_f^2, l_1, \dots, l_m, \lambda$ 의 값을 MLE를 통해 추정하고, 경사 하강법(Gradient descent)을 사용하여 GP 회귀 모델에 대한 최적화를 진행한다. 이를 위한 MLE 식은 다음과 같다.

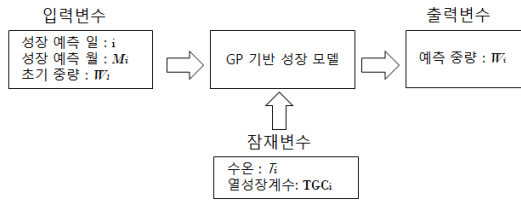
$$\begin{aligned} & \sigma_f^2, l_1, \dots, l_m, \lambda \\ & = \max_{\sigma_f^2, l_1, \dots, l_m, \lambda} P(Y_N | X_N, \sigma_f^2, l_1, \dots, l_m, \lambda) \\ & = -\frac{1}{2} Y_N^T (K(X_N, X_N) + \lambda^2 I)^{-1} Y_N \\ & \quad - \frac{1}{2} \log \| (K(X_N, X_N) + \lambda^2 I) \| - \frac{N}{2} 2\pi \end{aligned} \quad (4)$$

2.2 GP 회귀 기반 어류 성장 예측 모델 구조

부족한 현장 데이터 및 수집 센서 데이터에 포함된 잡음 오차 등을 고려하였을 때, 어떠한 시점에서 중량값을 특정하여 어류의 성장을 예측하는 것은 정확도가 낮다고 볼 수 있다. 한편, 양식장 현장에서 수집한 수치 데이터는 모집단에서 추출된 표본의 분포라고 볼 수 있다. 이 경우, 큰 수의 법칙(law of large numbers, LLN)에 따라 해당 표본 평균이 모집단의 평균값과 근접해지므로 수집된 수치 데이터를 잡음이 포함된 평균값이라고 가정하였고, 자연 현상에서 잡음은 가우시안 분포를 따르므로 수식 (1)의 GP 회귀의 적용이 가능하다. 따라서, 본 논문에서는 GP 회귀를 이용하여 어류의 중량을 특징하는 것이 아니라 어류의 중량의 대표값을 예측하고, 예측값의 신뢰구간을 출력하도록 하여 기존의 어류 성장 예측의 낮은 정확도를 보완할 수 있는 방법을 제시한다.

GP 회귀 기반 어류 성장 예측 모델에 대한 입출력 구조는 그림 2와 같다. 모델의 입력 변수는 임의의 성장 예

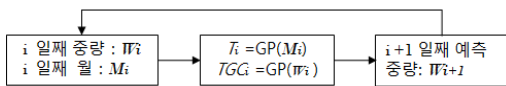
측일 $i(i=0, \dots, 364)$, 성장 예측일에 따른 월 M_i 그리고 초기 입식 어류 중량 W_0 으로 구성된다. 모델의 출력 변수 W_i 는 i 번째 날짜에서의 어류 중량을 의미한다. 모델이 어류 중량을 예측하기 위해 이용하는 파라미터는 i 번째 날짜에서의 수온 T_i 및 열 성장 계수 TGC_i 이다.



(그림 2) 어류 성장 예측 모델 입출력 구조

(Figure 2) Input-output structure in growth prediction model

GP 회귀 모델이 $i+1$ 번째 날짜에서 어류 중량을 예측하기 위한 절차는 그림 3에 도시하였다. 우선, i 번째 날짜에서의 T_i 를 GP 기반 모델링을 통해 예측하고, W_i 를 이용하여 TGC_i 값을 구한다. 그 이후, W_i, T_i, TGC_i 를 이용하여, 다음날 어류의 중량 W_{i+1} 을 예측한다. 그림 3의 구조에서 확인할 수 있듯이 GP 회귀 기반 어류 성장 예측 모델은 여러 서브 모델의 출력값을 입력으로 사용하는 계층적(hierarchical) 구조를 지니고 있다는 특징이 있다.



(그림 3) 어류 성장 예측 모델 동작 구조

(Figure 3) Operation structure in growth prediction model

3. GP 회귀 기반 어류 성장 예측 모델 설계

2절에서 제시한 GP 회귀 기반 어류 성장 예측 모델은 기존의 통계적 분석 기반 어류 성장 예측 모델과 비교하였을 때, 모델의 예측값에 대한 신뢰구간을 제시하여, 사용자가 예측값의 정확도를 판단할 수 있다는 장점이 있다. 본 절에서는 2절에서 기술된 성장 예측 모델 설계를 위한 세부 모델인 GP 회귀 기반 수온 예측 모델 및 TGC

예측 모델, 어류 중량 증가량 예측 모델을 설계를 제안한다.

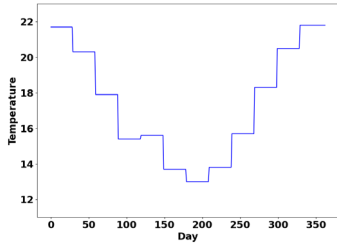
3.1 GP 회귀 기반 수온 예측 모델

본 절에서는 일별 수온 예측을 위한 GP 회귀 기반 TGC 모델을 제안한다. 수온 예측 모델 학습을 위하여 월별 국립수산물과학원 및 제주도 양식장 제공 데이터를 기초로 하여 다음과 같은 월에 따른 평균 수온 계산 수식을 사용한다.

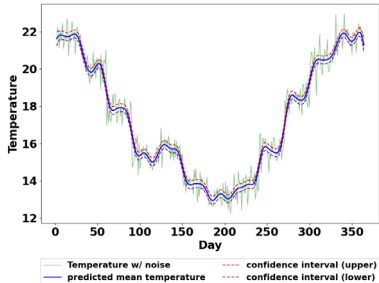
$$T = 17.51 + 4.468 * \sin \left(\frac{2\pi M}{10.91} + 2.998 \right) \quad (5)$$

식 (5)에서 계산된 수온을 그대로 이용하면 동일한 월에 해당하는 날짜에 대해 모두 동일한 수온이 계산되므로 월별 수온 계산 결과값에 $N(0, \sigma_T^2)$ 를 따르는 잡음을 추가하여 날짜별로 수온이 변동되도록 설계하였다. 이렇게 잡음이 추가된 수온 데이터를 GP 회귀 모델 학습을 위한 훈련 데이터로 사용하였다. 수온에 대한 GP 회귀는 모델의 가중치를 계산하는 과정과 계산된 가중치를 대입하여 GP 회귀에 의한 출력값 및 분포를 계산하는 과정으로 구성된다. 식 (4)를 이용하여 수온 예측 모델 가중치의 업데이트를 수행한다. GP 회귀 모델은 2절에서 소개하였듯이 예측값과 예측값에 대한 신뢰구간을 출력하고, 출력 모델의 예측값은 평균값으로 출력하도록 하였다.

그림 4은 9월을 시작점으로 하였을 때, 기존 통계적 수온 모델과 GP 회귀 기반 수온 모델의 날짜별 수온 결과를 도시한 그래프이다. 날짜별 변동을 주기 위해 생성한 잡음의 분산을 $\sigma_T^2 = 0.5$ 로 설정하였다. 그림 4(a)에서 파란색 실선은 기존 통계적 수온 예측 모델에 의한 수온을 나타낸다. 이를 통해, 기존 통계적 수온 모델을 이용할 경우, 수온의 변동성을 고려하지 못하고 월별로 동일한 수온을 사용해야 하는 한계가 있는 점을 확인할 수 있다. 그림 4(b)에서 파란색 실선은 GP 회귀 모델의 수온 예측값의 평균값을 나타내고, 초록색 실선은 모델의 학습 과정에서 변동성을 주기 위해 생성한 잡음이 있는 경우의 수온을 나타낸다. 추가로, 빨간색 점선과 보라색 점선은 각각 GP 회귀 모델의 수온 예측값에 대한 신뢰구간의 상한과 하한을 의미한다. 기존 통계 모델과 달리 GP 회귀 모델은 날짜별 변동성을 고려한 수온을 반영할 수 있을 뿐만 아니라 예측값에 대한 신뢰도가 출력되는 것을 확인할 수 있다.



(a) 통계적 수온 예측 모델을 적용한 일별 수온



(b) GP 회귀 기반 일별 수온 예측 평균 및 신뢰 구간

(그림 4) 통계 및 GP 회귀 기반 수온 예측 결과

(Figure 4) Water temperature prediction results based on statistics and GP regression

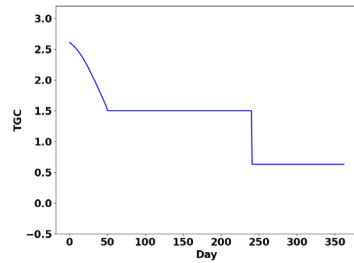
3.2 GP 회귀 기반 TGC 모델

본 절에서는 일별 수온을 바탕으로 어류 중량 증가분 예측을 위한 GP 회귀 기반 TGC 모델을 제안한다. 수온 예측 모델과 마찬가지로, 어류의 중량 구간별 통계적 TGC 모델을 기준으로 하며 통계적 데이터 기반의 TGC 모델은 다음 수식과 같이 어류의 중량값과의 관계식으로 표현할 수 있다.

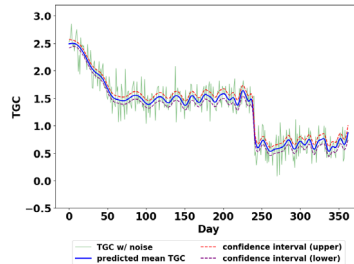
$$TGC_i = \begin{cases} 2.6984 - 0.01899 W_i & (W_i \leq 60) \\ 1.5 & (60 < W_i \leq 570) \\ 0.63 & (W_i > 570) \end{cases} \quad (6)$$

수온 모델과 유사하게 중량에 따른 TGC 출력값에 $N(0, \sigma_{TGC}^2)$ 를 따르는 잡음을 추가하여 변동성을 반영할 수 있도록 GP 회귀 모델을 설계하였다. 이렇게 잡음이 추가된 TGC 데이터를 GP 회귀 모델 학습을 위해 사용한 훈련 데이터로 사용하였다. TGC GP 회귀 모델의 학습 절차 또한 수온 예측 모델의 학습 절차와 동일하다. TGC 예측 모델은 예측값과 예측값에 대한 신뢰구간을 출력하고, 예측값은 평균값으로 출력하도록 하였다.

그림 5는 초기 어류의 무게 $W_0 = 5g$ 시작점으로 하였을 때, 기존 통계적 TGC 모델과 GP 회귀 기반 TGC 모델의 날짜별 TGC 결과를 도시한 그래프이다. 어류의 중량에 따른 TGC 값에 변동을 주기 위해 생성한 잡음의 분산을 $\sigma_{TGC}^2 = 0.2$ 로 설정하였다. 그림 5-(a)에서 파란색 실선은 기존 통계적 TGC 모델에 의해 계산된 TGC를 나타낸다. 수온에서의 결과와 유사하게 통계적 모델을 이용할 경우, 변동성을 고려하지 못하고, 오직 어류의 무게에 의해서만 TGC가 결정된다는 한계를 확인할 수 있다. 그림 5-(b)에서 파란색 실선은 GP 회귀 모델의 TGC 예측값의 평균값을 나타내고, 초록색 실선은 모델의 학습 과정에서 변동성을 주기 위해 생성한 잡음이 있는 경우의 TGC를 나타낸다. 추가로, 빨간색 점선과 보라색 점선은 각각 GP 회귀 모델의 TGC 예측값에 대한 신뢰구간의 상한과 하한을 의미한다. 기존 통계 모델과 달리 GP 회귀 모델은 변동성을 반영할 수 있을 뿐만 아니라 예측값에 대한 신뢰도가 출력되는 것을 확인할 수 있다.



(a) 통계적 TGC 모델을 적용한 일별 TGC



(b) GP 회귀 기반 일별 TGC 예측 평균 및 신뢰 구간

(그림 5) 통계 및 GP 회귀 기반 TGC 예측 결과

(Figure 5) TGC prediction results based on statistics and GP regression

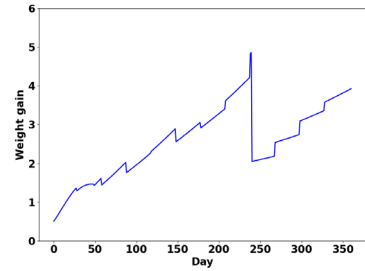
3.3 GP 회귀 기반 어류 중량 증가량 예측 모델

본 절에서는 GP 회귀 기반 일별 수온 예측 모델 및 TGC 예측 모델을 기반으로 i 번째 날짜에서의 어류 중량에 대한 $i+1$ 번째 날짜에서의 어류 중량 증가량 예측 모델을 제안한다. 전통적인 TGC 통계적 데이터 기반의 어류 중량 증가량은 다음 수식과 같이 수온, TGC, 어류의 중량값 사이의 관계식으로 표현된다 [6].

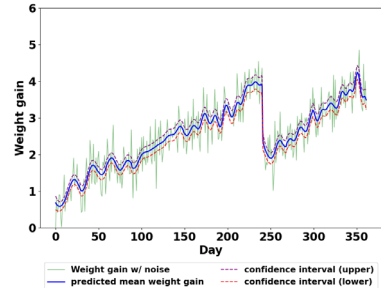
$$\Delta W_i = W_{i+1} - W_i = \left(\frac{TGC_i * T_i}{1000} + \sqrt[3]{W_i} \right)^3 \quad (7)$$

앞서 설계한 수온 및 TGC GP 회귀 모델과 유사하게 식 (7)을 그대로 사용하지 않고, $N(0, \sigma_{\Delta W}^2)$ 를 따르는 잡음을 추가하여 변동성을 반영할 수 있도록 GP 회귀 모델을 설계하였다. 이 때, 어류의 중량 증가분은 항상 음이 아닌 값이라고 가정하여 중량 증가분에 대한 잡음 결과가 음수가 나올 경우, 0이 되도록 보정하였다. 어류 성장 예측 모델의 학습 절차 또한 수온 및 TGC 예측 모델의 학습 절차와 동일하다. 어류 중량 증가량 예측 모델은 예측값과 예측값에 대한 신뢰구간을 출력하고, 예측값은 평균값으로 출력하도록 하였다.

그림 6는 9월을 어류 양성 시작점으로 하고, 초기 어류의 무게 $W_0 = 5g$ 일 때, 기존 통계적 어류 중량 증가량 모델과 GP 회귀 기반 어류 중량 증가량 모델의 날짜별 어류 중량의 증가분에 대한 결과를 도시한 그래프이다. 어류 중량 증가분에 변동성을 주기 위해 생성한 잡음의 분산을 $\sigma_{\Delta W}^2 = 0.4$ 로 설정하였다. 그림 6-(a)에서 파란색 실선은 기존 통계적 모델에 의해 계산된 날짜별 어류의 중량 증가분을 나타낸다. 앞선 수온과 TGC 결과와 유사하게 기존 통계적 모델은 변동성을 고려하지 못한다는 한계를 확인할 수 있다. 그림 6-(b)에서 파란색 실선은 GP 회귀 모델의 어류 중량 증가분 예측값의 평균값을 나타내고, 초록색 실선은 모델의 학습 과정에서 변동성을 주기 위해 생성한 잡음이 있는 경우에 대한 어류의 중량 증가분을 나타낸다. 추가로, 빨간색 점선과 보라색 점선은 각각 GP 회귀 모델의 어류 중량 증가분 예측값에 대한 신뢰구간의 상한과 하한을 의미한다. 기존 통계 모델과 달리 GP 회귀 모델은 변동성을 반영할 수 있을 뿐만 아니라 예측값에 대한 신뢰도가 출력되는 것을 확인할 수 있다.



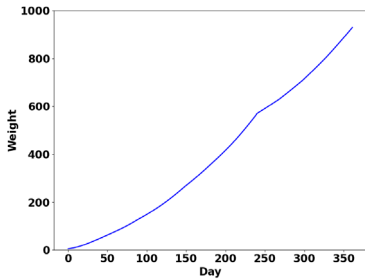
(a) 통계적 어류 중량 증가 모델을 적용한 일별 중량 증가량



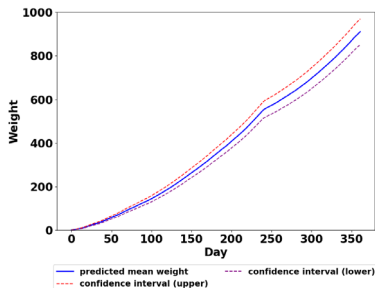
(b) GP 회귀 기반 일별 중량 증가량 예측 평균 및 신뢰 구간 (그림 6) 통계 및 GP 회귀 기반 중량 증가량 예측 결과 (Figure 6) Weight gain results based on statistics and GP regression

4. 실험 결과

본 절에서는 3절에서 설계한 계층 구조의 GP 회귀 기반 예측 모델을 모두 합쳐서 날짜별 어류의 성장 예측 결과를 분석한다. 초기 조건으로는 9월을 어류 양성 시작점으로 하고, 초기 어류의 중량 $W_0 = 5g$ 을 고려하였다. 그림 7은 위의 초기 조건을 바탕으로 기존 통계적 어류 성장 모델과 GP 회귀 기반 예측 모델을 통한 1년 양성 주기 내에서 어류의 최종 중량까지의 예측값을 도시한 그래프이다. 그림 7-(a)에서 파란색 실선은 기존 통계적 모델에 의해 계산된 날짜별 어류의 중량 예측 결과를 나타낸다. 그림 7-(b)에서 파란색 실선은 GP 회귀 모델의 어류 중량 예측값의 평균값을 나타낸다. 추가로, 빨간색 점선과 보라색 점선은 각각 GP 회귀 모델의 어류 중량 예측값에 대한 신뢰구간의 상한과 하한을 의미한다. 한편, GP 회귀 모델의 중량 예측값 평균과 최대, 최소값에 대한 신뢰구간을 비교하였을 때, 6.3%~10%의 변동 폭을 나타낸다. 이를 통해, 기존 통계적 모델과 달리 제안하는 GP 회귀 모델은 어류의 양식 주기 내에서 계산된 중량값이 어느 정도의 신뢰도를 나타내는지 정량적으로 알 수 있다.



(a) 통계적 모델을 적용한 1년 주기 어류의 성장도



(b) GP 회귀 기반 1년 주기 어류 예측 성장도 평균 및 신뢰구간 (그림 7) 통계 및 GP 회귀 기반 어류 성장도 예측 결과 (Figure 7) Weight prediction results based on statistics and GP regression

5. 결 론

본 논문에서는 GP 회귀 기반의 어류 성장 예측 모델을 통하여 특정 시점에서의 어류 중량에 대한 평균값 및 해당 값에 대한 신뢰 정도를 정량적으로 획득할 수 있는 방식을 제안하였다. 종래의 통계적 수치에 의한 성장 예측 모델의 경우 특정 값을 제공함에 따라 해당 예측치에 대한 확실성의 측정이 어렵고 신뢰도가 높은 수집 데이터 확보가 어려운 수산 양식장 환경에서의 잡음의 영향을 반영하기 어려운 단점이 존재하였으나 GP 회귀 기반의 어류 성장 예측 모델은 신뢰 구간의 정량적 수치도 함께 제공함으로써, 수산 양식장 관리 및 출하 관정을 위한 생산량 예측에 있어 현장에서 효율적인 대응이 가능할 것으로 예상된다. 향후 연구 계획으로는 기존의 통계적 데이터의 변동성을 제공하기 위한 가우시안 잡음 생성에 대하여 현장 수집 데이터를 통한 적합성 여부 및 정교화 과정을 수행하고 성장 주요 변수로써 수온 외에 산소 및 먹이 공급량 등에 따른 GP 회귀 기반 성장 예측 모델링에 대한 연구를 수행할 예정이다.

참고문헌(Reference)

- [1] Z. Hu, R. Li, X. Xia, C. Yu, X. Fan and Y. Zhao, "A method overview in smart aquaculture," *Environ. Monit. Assess.*, vol. 192, no. 8, pp. 1-25, Jul. 2020. <https://doi.org/10.1007/s10661-020-08409-9>
- [2] D. Lee, B. Bae, J. Lee, S. Kim and H. Kim, "Development of the smart aquaculture technology and industrialization plan," *Journal of Fisheries and Marine Science Education*, vol. 33, no. 2, pp. 412-420, Apr. 2021. <https://doi.org/10.13000/JFMSE.2021.4.33.2.412>
- [3] H. AlZubi, W. Al-Nuaimy, J. Buckley and I. Young, "An intelligent behavior-based fish feeding system," *13th Int. Multi-Conference Syst. Signals. Devices, SSD*, pp. 22-29, 2016. <https://doi.org/10.1109/SSD.2016.7473754>
- [4] R. Schraml *et al.*, "Towards fish individuality-based aquaculture," *IEEE Trans. Instrum. Inform.*, vol. 17, no. 6, pp. 4356-4366, Jun. 2021. <https://doi.org/10.1109/TII.2020.3006933>
- [5] Y. Eh, "Productivity of aquaculture facility utilization," *The Journal of Fisheries Business Administration*, vol. 45, no. 2, pp. 85-95, 2014. <https://doi.org/10.12939/FBA.2014.45.2.085>
- [6] M. Jobling, "The thermal growth coefficient (TGC) model of fish growth: A cautionary note," *Aquaculture Research*, pp. 581-584, May 2003. <https://doi.org/10.1046/j.1365-2109.2003.00859.x>
- [7] H. Thorarensen and A. P. Farrell, "The biological requirements for post-smolt Atlantic salmon in closed-containment systems," *Aquacult.*, vol. 312, pp. 1-14, Feb. 2011. <https://doi.org/10.1016/j.aquaculture.2010.11.043>
- [8] C. K. Williams and C. E. Rasmussen, "Gaussian processes for regression," in *Adv. Neural Inform. Processing Syst.* 8, pp. 514-520, 1996. <https://dl.acm.org/doi/10.5555/2998828.2998901>
- [9] C. E. Rasmussen and C. K. Williams, *Gaussian Processes in Machine learning*. Cambridge, U.K.: Cambridge Univ. Press, 2006.

- [10] H. Liu, Y.-S. Ong, X. Shen, and J. Cai, "When Gaussian process meets Big Data: A review of scalable GPs," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 31, no. 11, pp. 4405-4423, Nov. 2020.
<https://doi.org/10.1109/TNNLS.2019.2957109>

◎ 저 자 소 개 ◎



성 주 형(Juhyoung Sung)

2015년 고려대학교 컴퓨터통신공학부(공학사)
2019년 한국과학기술원(KAIST) 전기및전자공학과(공학석사)
2015년~2017년 대한민국 육군 정보통신장교
2019년~2021년 삼성전자 네트워크사업부 연구원
2021년~현재 한국전자기술연구원 스마트네트워크연구센터 선임연구원
E-mail: jh.sung@keti.re.kr
관심분야: 무선 통신, 시스템 최적화, 강화학습



조 성 윤(Sungyoon Cho)

2006년 연세대학교 전기전자공학과(공학사)
2008년 연세대학교 전기전자공학과(공학석사)
2013년 연세대학교 전기전자공학과(공학박사)
2013년~2020년 삼성전자 시스템 LSI사업부 책임연구원
2020년~현재 한국전자기술연구원 스마트네트워크연구센터 책임연구원
관심분야: 5G/6G 통신시스템, 적응적 신호처리, 딥러닝
E-mail: sycho@keti.re.kr



정 다 은(Da-eun Jung)

2018년 호서대학교 해양IT공학과(공학사)
2020년 호서대학교 정보통신공학과(공학석사)
2020년~2021년 산엔지니어링 선임연구원
2021년~현재 한국전자기술연구원 스마트네트워크연구센터 연구원
E-mail: dejung@keti.re.kr
관심분야: 수산기술, 해양 ICT, 통신시스템, 임베디드 시스템, 딥러닝



김 종 원(Jongwon Kim)

2020년 포항공과대학교 산업경영공학과 및 수학과(공학사 및 이학사)
2022년 포항공과대학교 산업경영공학과(공학석사)
2022년~현재 포항공과대학교 산업경영공학과 박사과정
E-mail: pioneer0517@postech.ac.kr
관심분야: 확률 모델링, 페이지안 최적화

◎ 저 자 소 개 ◎



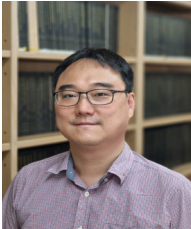
박 정 환(Jeonghwan Park)

1998년 국립부경대학교 양식학과(수산학사)
2000년 국립부경대학교 수산생물학과(수산학석사)
2005년 국립부경대학교 수산생물학과(수산학박사)
20012년 North Carolina State University, Dept. of Biology(선임연구원)
2016년 University of Arkansas at Pine Bluff, Dept. of Fisheries & Aquaculture(조교수)
2016년~현재 국립부경대학교 수산생명과학부 양식응용생명과학전공(부교수)
2021년~현재 국립부경대학교 수산과학기술센터(센터장)
관심분야: 어류양식, 어류성장모델, 수질환경관리, 순환여과양식시스템
E-mail : parkj@pknu.ac.kr



권 기 원(Kiwon Kwon)

1997년 광운대학교 컴퓨터공학과(공학사)
1999년 광운대학교 대학원 컴퓨터공학과(공학석사)
2011년 중앙대학교 대학원 전자전기공학부(공학박사)
1999년~현재 한국전자기술연구원 스마트네트워크연구센터 센터장
관심분야 : 디지털트윈, 유무선디지털통신시스템, 해양수산ICT융합
E-mail : kwonkw@keti.re.kr



고 영 명(Young Myoung Ko)

1998년 서울대학교 산업공학과(공학사)
2000년 서울대학교 대학원 산업공학과(공학석사)
2011년 Texas A&M 대학교 대학원 산업시스템공학과(공학박사)
2013년~현재 포항공과대학교 산업경영공학과 부교수
관심분야 : 확률모형, 응용통계, 최적화
E-mail : youngko@postech.ac.kr