Article

https://doi.org/10.7850/jkso.2024.29.1.001
pISSN : 1226-2978 eISSN : 2671-8820

기계학습법을 이용한 동해 울릉분지의 봄과 여름 순군집생산 추정

함도식[1]* · 이인희[2] · 추민기[3]
[1]부산대학교 해양학과 및 해양연구소 부교수, [2]부산대학교 해양학과 대학원생, [3]부산대학교 해양학과 학부생

Estimation of the Spring and Summer Net Community Production in the Ulleung Basin using Machine Learning Methods

DOSHIK HAHM[1]*, INHEE LEE[2] AND MINKI CHOO[3]
[1]Associate Professor, Department of Oceanography and Marine Research Institute, Pusan National University, Busan 46241, Korea
[2]Graduate Student, Department of Oceanography, Pusan National University, Busan 46241, Korea
[3]Undergraduate Student, Department of Oceanography, Pusan National University, Busan 46241, Korea

*Corresponding author: hahm@pusan.ac.kr
Editor Jeomshik Hwang
Received 23 November 2023; Revised 5 January 2024; Accepted 6 January 2024

ABSTRACT

동해 남서부해역은 대마난류나 연안 용승에 의한 영양염 공급 등으로 동해 북부나 동부에 비해 일차생산력이 높은 것으로 알려져 있지만, 이 해역의 생물 펌프에 관한 연구는 제한적이다. 본 연구에서는 O_2/Ar 측정으로 산출한 고해상도 순군집생산 현장 관측 결과와 기계학습 모형을 결합하여 시공간 해상도가 8일 간격, 4 km인 봄과 여름 순군집생산 시계열 자료를 추정하였다. 기계 모형의 예측과 실측의 평균 제곱근 오차는 6 mmol O_2 m^{-2} d^{-1}로 관측값 평균의 15%에 해당했다. 울릉분지 중앙부의 순군집생산은 3월에 49 mmol O_2 m^{-2} d^{-1}로 가장 높았고, 6월과 7월에 18 mmol O_2 m^{-2} d^{-1}로 가장 낮았다. 이 같은 계절 변화는 ^{3}He 기체교환율로 추정한 질산염 공급률이나 ^{234}Th 비평형법으로 추정한 입자유기탄소 방출률과 유사하였다. 봄과 여름의 순군집생산 추정으로 한정된 이 연구 방법을 가을과 겨울로 확대하기 위해서는 아표층수의 표층 혼입에 따른 O_2/Ar 순군집생산의 오차를 보정하는 연구가 필요하다.

The southwestern part of the East Sea is known to have a high primary productivity compared to those in the northern and eastern parts, which is attributed to nutrients supplies either by Tsushima Warm Current or by coastal upwelling. However, research on the biological pump in this area is limited. We developed machine learning models to estimate net community production (NCP), a measure of biological pump, with high spatial and time scales of 4 km and 8 days, respectively. The models were fed with the input parameters of sea surface temperature, chlorophyll-a, mixed layer depths, and photosynthetically active radiation and trained with observed NCP derived from high resolution measurements of surface O_2/Ar. The root mean square error between the predicted values by the best performing machine model and the observed NCP was 6 mmol O_2 m^{-2} d^{-1}, corresponding to 15% of the average of observed NCP. The NCP in the central part of the Ulleung Basin was highest in March at 49 mmol O_2 m^{-2} d^{-1} and lowest in June and July at 18 mmol O_2 m^{-2} d^{-1}. These seasonal variations were similar to the vertical nitrate flux based on the ^{3}He gas exchange rate and to the particulate organic carbon flux estimated by the ^{234}Th disequilibrium method. To expand this method, which produces NCP estimate for spring and summer, to autumn and winter, it is necessary to devise a way to correct bias in NCP by the entrainment of subsurface waters during the seasons.

Keywords: Net community production, Biological pump, Machine learning, East sea, Ulleung basin

1. 서 론

울릉분지를 포함하고 있는 동해 남서부해역의 연간 일차생산력은 220 − 260 g C m^{-2} yr^{-1}로 북쪽의 일본 분지나 동쪽의 야마토 분지에 비해 10 - 20% 높은 것으로 알려져 있다(Yamada *et al*., 2005; Joo *et al*., 2016). 이처럼 높은 생산력을 지닌 까닭

Copyright ⓒ 2024 Journal of the Korean Society of Oceanography.

으로는 여름철 한반도 남동해안을 따라 발생하는 연안 용승에 의한 영양염 공급(Hyun *et al*., 2009; Yoo and Park, 2009; Hahm *et al*., 2019a), 수온약층에 걸친 소용돌이에 의한 영양염의 수직 공급(Kim *et al*., 2012), 대마난류를 통한 영양염 공급(Onitsuka *et al*., 2007) 등이 제안되었다.

해양 표층에서 일차생산으로 만들어진 유기물의 일부는 심층으로 침강하여 심해 생물에게 에너지를 공급할 뿐만 아니라 대기 이산화탄소를 수십 년에서 수백 년 동안 심해에 격리시킨다. '생물펌프(biological pump; Volk and Hoffert, 1985)'라 불리는 이 과정이 기후 조절에 중요한 역할을 하는 까닭에, 생물펌프의 척도가 되는 신생산(new production), 방출생산(export production), 순군집생산(net community production, NCP) 등을 측정하기 위한 다양한 실험 방법들이 제안되어 전 대양에서 관측이 이루어지고 있다. 하지만, 높은 일차생산력으로 주목받고 있는 울릉분지의 생물 펌프를 가늠해볼 수 있는 연구 결과는 매우 제한적이다. Kwak *et al*.(2013b)은 매월 실시한 ^{13}C-^{15}N 배양 실험을 근거로 연간 신생산이 일차생산의 54%에 달하는 146 g C m^{-2} yr^{-1}으로 보고하였다. 반면에, ^{3}He의 해양-대기 기체교환율과 수온약층 구간의 ^{3}He과 NO_3^- 상관관계로 추정한 연간 신생산(질산염의 수직공급률)은 64 g C m^{-2} yr^{-1}(Hahm and Kim, 2001), ^{238}U-^{234}Th 비평형법으로 추정한 입자성유기입자 방출률은 59 g C m^{-2} yr^{-1}(Kim *et al*., 2011)로, 표층 수온과 엽록소로 추정한 표층 질산염 농도의 겨울과 여름 차이로 계산한 신생산 66 g Cm^{-2} yr^{-1}(Joo *et al*., 2018)과 유사하였다. 배양법에 의한 울릉분지 신생산 추정(Kwak *et al*., 2013b)과 추적자 기반 생물 펌프 추정 결과(Hahm and Kim, 2001; Kim *et al*., 2011) 간의 차이에 대한 체계적인 해석은 아직 이루어지지 않았다.

동해 남서부해역 생물 펌프 추정 결과들이 이처럼 상이한 것은 극히 제한적인 현장 관측과 이를 보완할 광역 생물펌프 추정 모형이 부재하기 때문이다. 평형기주입식 질량분석법(Cassar *et al*., 2009)은 연구선 이동 중에 1분 이내의 짧은 시간 간격으로 연속하여 용존 O_2와 Ar를 관측하여 순군집생산을 추정(O_2/Ar-NCP)하는 기법으로, 광역 관측에 제한이 따르는 ^{15}N 배양법이나 ^{238}U-^{234}Th 비평형법 등의 대안이 될 수 있다. 이에 더해 표층 광역 관측으로 얻어진 고해상도 O_2/Ar-NCP와 이의 조절 인자로 작용할 수 있는 표층 수온, 엽록소, 광량 등의 정보를 결합하여기계학습 모형의 개발과 검증에도 활용함으로써 순군집생산의 광역 시계열 추정이 가능하다(Li and Cassar, 2016; Park *et al*., 2019).

본 연구에서는 2013년부터 2021년까지 7회의 독립된 탐사를 통해 획득한 O_2/Ar-NCP 결과와 주요 기계학습 모형(Random Forest, Extra Trees, Extreme Gradient Boosting)을 결합하여 동해 남서부해역의 순군집생산을 추정하였다. 기계학습 모형의 입력 자료로는 시공간 해상도 8일, 4 km인 수온(sea surface temperature, SST), 엽록소(chlorophyll-*a*, CHL), 혼합층 수심(mixed layer depth, MLD), 광합성 광량(photosynthetically active radiation, PAR)을 이용하였다. 앞 서 소개한 ^{15}N 배양법(Kwak *et al*., 2013b)과 ^{238}U-^{234}Th 비평형법(Kim *et al*., 2011)에 의한 울릉분지 관측 결과의 최고 시간 해상도가 각각 월별과 계절별로 제한되어 있는데 비해, 본 기계학습 모형을 이용한 추정 결과는 입력 자료와 동일하게 8일, 4 km의 높은 시공간 해상도를 갖고 있다. 순군집생산 추정 기간은 혼합층 수심이 점차 얕아져 혼합층 하부 해수의 유입을 무시할 수 있는 3월부터 8월까지로 하였다. 본 연구는 향후 혼합층 하부 해수의 유입에 따른 O_2/Ar-NCP 오차를 보정하는 기법을 개발하여 기계학습 모형으로 연간 순군집생산을 추정하기 위한 사전 연구에 해당한다.

2. 현장 관측과 자료 준비

2.1 순군집생산 현장 관측과 격자화

이번 연구에 이용된 순군집생산은 동해 울릉분지에서 평형기주입식 질량분석기(equilibrator inlet mass spectrometer, EIMS)를 이용하여 탐사선의 이동 중에 연속 관측한 표층 해수의 O_2/Ar로부터 계산되었다(Cassar *et al*., 2009). 7회의 독립탐

사에는 동일한 대기-해수 평형기(SKU-C900732/900733, General Oceanics)와 사중극자질량분석기(QMG 220, Pfeiffer)로 구성된 EIMS를 이용하여 O_2/Ar을 관측하였다(Hahm and Lee, 2018; Hahm *et al*., 2019a).

비활성기체인 Ar은 수온에 따른 용해도와 확산계수 등이 O_2와 매우 유사하여 수온과 염분의 변화, 기체교환 등 물리현상에 따른 산소 농도 변화를 잘 대변한다. O_2/Ar의 변화는 광합성이나 호흡과 같은 생물학적 현상에 따른 산소 변화를 구별하는 도구로 활용되고 있다(Craig and Hayward, 1987).

생물학적 현상에 따른 과포화도(biological O_2 saturation anomaly)로 불리는 ΔO_2/Ar은 다음과 같이 정의된다.

$$\triangle \mathrm{O}_2/\mathrm{Ar} = \frac{(\mathrm{O}_2/\mathrm{Ar})_{sample}}{(\mathrm{O}_2/\mathrm{Ar})_{sat}} - 1 \tag{1}$$

여기서 $(O_2/Ar)_{sample}$와 $(O_2/Ar)_{sat}$는 해수 시료와 대기에 포화된 해수의 O_2/Ar비를 각각 나타낸다(Kaiser *et al*., 2005). 수평 이류와 수직 혼합이 무시할 만큼 작으면, 생물학적 과포화 산소의 기체 교환율은 산소의 혼합층 체류시간(1 - 2 주)의 수 배에 해당하는 과거의 평균적인 순군집생산(NCP)을 대변한다(Teeter *et al*., 2018):

$$NCP = k_w \cdot \rho_w \cdot [\mathrm{O}_2]_{sat} \cdot \triangle \mathrm{O}_2/\mathrm{Ar} \tag{2}$$

여기서 ρ_w는 해수의 밀도, $[O_2]_{sat}$는 기체 포화상태의 산소 농도이다. k_w는 일별 가중치를 적용한 산소의 기체이동속도로, 관측 시기의 산소 농도가 과거 기체교환의 누적된 결과임을 반영하기 위한 것이다(Teeter *et al*., 2018). 즉,

$$k_w = \frac{\Sigma_{t=1}^{30} k_t w_t}{\Sigma_{t=1}^{30} w_t} \tag{3}$$

$$w_1 = 1, w_t = w_{t-1}(1 - f_{t-1}) \tag{4}$$

$$f_t = \frac{k_t \cdot 1day}{MLD} \tag{5}$$

관측일의 가중치(w_1)는 1이고, 특정일 기체교환에 의해 환기된 혼합층 비율(f_t)이 클 수록 전 날 가중치가 빠르게 감소한다. 일별 기체이동속도(k) 계산을 위한 풍속에는 ERA5 바람(https://www.ecmwf.int/en/forecasts/dataset/ecmwf-reanalysis-v5 (ECMWF, 2023))을, 풍속에서 기체이동속도를 계산하기 위해서는 Wanninkhof (2014)가 제시한 관계식을 이용하였다. 과거 일별 k에서 k_w를 산출하는데 필요한 혼합층 수심은 'Ocean Productivity' 홈페이지(https://sites.science.oregonstate.edu/ocean.productivity/index.php(Ocean Productivity, 2023))에서 얻은 8 km, 8일 해상도의 자료를 일별로 내삽하여 사용하였다. 이 자료의 혼합층 수심은 HYCOM 모형 결과를 이용하여 10 m에 비해 밀도가 0.03 kg m^{-3} 증가한 수심으로 결정되었다. 자세한 k_w 산출 방식은 Teeter *et al*.(2018)를 참고하기 바란다.

2013년부터 2021년에 사이 봄과 여름(3월부터 8월까지)에 획득한 순군집생산 자료 28500개를 기계학습에 이용하였다.

Table 1. Counts of the NCP data in each grid with a temporal and spatial resolutions of 8 days and 4 km

Month	Date	Data count
3	2018-03-06	84
	2018-03-14	85
5	2021-05-25	50
6	2016-06-25	85
	2017-06-26	120
	2019-06-02	205
	2019-06-10	18
7	2013-07-12	128
	2013-07-20	435
	2017-07-20	115

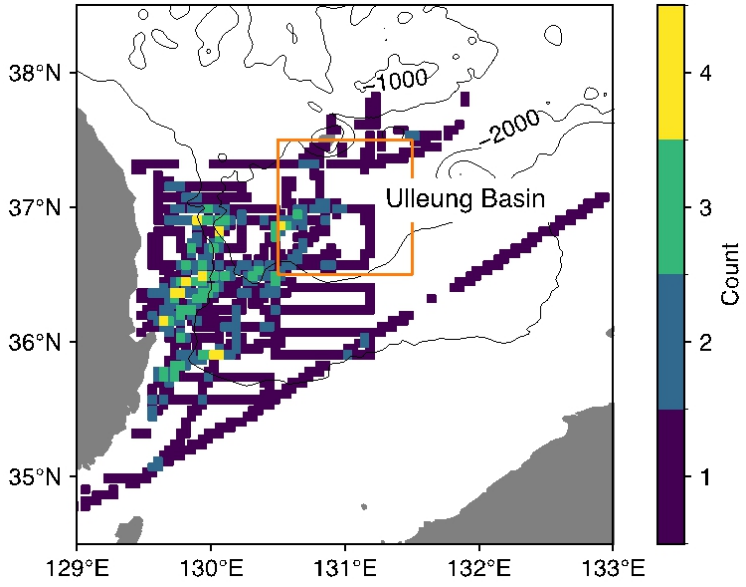

Fig. 1. Numbers of observed data in each grid with a temporal and spatial resolutions of 8 days and 4 km, respectively. The orange rectangle (36.5 — 37.5 °N, 130.5 — 131.5 °E) represents the center of the Ulleung Basin, which roughly follows 2000 m contour line.

가을과 겨울에는 혼합층 수심이 점차 깊어짐에 따라 산소 농도가 낮은 혼합층 하부 해수의 표층 유입(entrainment)에 의해 순군집생산이 과소평가될 수 있기에, 가을과 겨울 자료는 기계학습 모형 구축에서 제외하였다(Lee *et al*., 2021). 1분 간격으로 얻어진 순군집생산은 기계학습 입력 자료의 해상도에 맞추어 공간 간격 4 km, 시간 간격 8일로 격자화하였다(Table 1). 격자화 후 순군집생산 자료는 총 1325개로 주로 6월(32%)과 7월(51%)에 집중되었다. 관측 빈도는 울릉 분지의 남서부 해역에서 3 - 4회로 높았다(Fig. 1).

2.2 위성 및 재해석 자료

기계학습 모형의 입력 자료로 기계학습으로 순군집생산을 추정한 선행 연구(Chang *et al*., 2014; Park *et al*., 2019)를 참고하여 표층 수온(SST), 엽록소(CHL), 광합성 광량(PAR), 혼합층 수심(MLD)으로 선정하였다(Table 2). NASA Ocean Color

Table 2. List of input data used in this study

Data	Website	Resolution	Transformation
SST (Aqua MODIS)			-
CHL (Aqua MODIS)	NASA Ocean Color	4 km, 8-day	$\log_{10}$CHL
PAR (Aqua MODIS)			-
MLD (HYCOM)	Ocean Productivity	8 km, 8-day	interpolated to 4 km

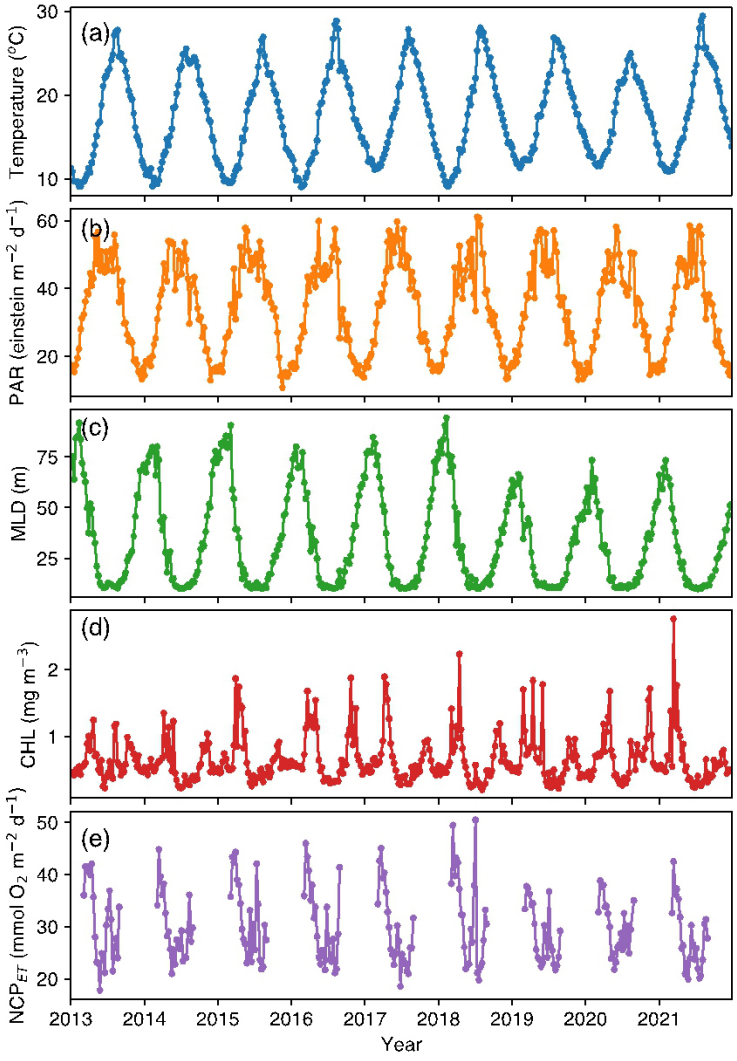

Fig. 2. Time series of the 8-day averages of the input data in the study area (a-d) and spring and summer net community production predicted by the Extra Trees model (e).

(https://oceancolor.gsfc.nasa.gov/l3/(NASA, 2023))에서 4 km, 8일 해상도의 MODIS SST, CHL과 PAR 자료를 얻었다. MLD는 'Ocean Productivity' 홈페이지에서 얻은 8 km, 8일 해상도의 자료를 4 km 간격으로 내삽하여 사용하였다(https://sites.science.oregonstate.edu/ocean.productivity/index.php(Ocean Productivity, 2023.)). CHL 자료는 상용로그 변환으로 비대칭도(skewness)를 줄였다. 입력 자료의 시간 범위는 처음 관측이 있었던 2013년부터 마지막 관측이 있었던 2021년까지로 하였다(Fig. 2).

위에서 언급한 위성 및 재해석 자료에 더해 경도(LON)와 위도(LAT)의 위치 정보와 시간 정보를 입력 자료에 포함시켰다. 동해 남서부해역의 순군집생산을 증가시키는 주요 요인으로 간주되는 연안 용승을 포함하여, 주요 입력 인자(SST, CHL, PAR, MLD)로 충분히 예측되지 않는 요소를 보완하기 위해 위치 정보를 포함시켰다(Nakaoka *et al*., 2013; Zeng *et al*., 2014). 또, 자료의 계절 변화를 재현하기 위해 날짜 순서(day of year; 이후 DOY)의 사인함수와 코사인 함수를 추가 변수로 활용하였다(Gregor *et al*., 2017; Hahm *et al*., 2019b).

3. 기계학습 모형

격자화한 순군집생산 관측 결과와 시공간이 일치하는 입력 자료를 결합하여 기계학습 모형 개발에 필요한 자료를 준비하였다. 격자화한 순군집생산 자료는 1325개였지만, CHL 자료의 제한으로 인해 최종 모형 개발에 이용 가능한 입력자료과 순군집생산 자료 짝은 모두 833개였다. 이 자료의 80%, 10%, 10%를 모형의 훈련, 검증, 시험 용도에 사용하였다.

동해 남서부해역의 순군집생산을 가장 잘 설명하는 기계학습 모형의 탐색에 TPOT (Tree-based Pipeline Optimization Tool)을 사용하였다. TPOT은 Python으로 쓰여진 자동기계학습(Automated Machine Learning) 도구로 유전 알고리즘을 이용하여, 다양한 기계학습 모형과 모형 변수 조합을 효율적으로 탐색하여 최적의 기계학습 모형을 만들어낸다(Le *et al*., 2020). TPOT의 기본 탐색 대상 모형에는 scikit-learn 기계학습 패키지(https://scikit-learn.org(Scikit Learn, 2023))에 포함된 주요 모형(Extra Trees, Support Vector Machine, Random Forest 등)과 xgboost (https://xgboost.readthedocs.io(XGBoost, 2023)) 패키지의 Extreme Gradient Boosting 모형 등이 있다. TPOT의 모형 성능 판정 기준은 모형이 예측한 값과 현장관측 결과(시험 자료)의 최소 평균제곱근오차(root mean square error, RMSE)로 정하였다.

TPOT이 최적화한 기계학습 모형인 Extra Trees (ET)에 더해 해양의 fCO_2 모형 구축에 활용되고 있는 Random Forest (RF)와 Extreme Gradient Boosting (XGB) (Gregor *et al*., 2017, 2019) 모형을 대상으로 TPOT을 이용하여 각 모형별 최적 변수를 탐색하였다. RF 모형은 Decision Tree (DT) 구성에 필요한 입력 자료 표본과 입력 변수(feature)를 무작위로 선택하여 서로 다른 다수의 DT를 만들고 평균함으로써 DT의 과적합을 줄인다(Breiman, 2001). ET는 RF와 유사하게 다수의 DT를 만들고 평균하지만, DT의 경계값을 무작위로 선택하여 모형에 무작위성을 증가시킴으로써, 모형의 과적합과 모형의 특정 변수에 대한 의존도를 줄일 수 있다. 또, RF와 달리 DT의 최적 경계값 탐색 시간을 줄임으로써 RF에 비해 연산효율이 높다(Geurts *et al*., 2006).

XGB도 앞에서 언급한 RF, ET와 동일하게 다수의 DT를 결합하여 최종학습 모형을 구축하는 공통점이 있지만, 예측 능력이 떨어지는 DT (weak learner)의 오차를 줄이는 새로운 DT를 반복하여 생성함으로써 강력한 예측 모형을 구축(Gradient Boosting)하는 차이점이 있다. XGB는 효과적인 regularization 기법을 이용하여 모형의 과적합을 줄이고 일반화를 강화할 뿐만 아니라, 분산 병렬 연산이 가능하여 규모가 큰 자료를 효과적으로 취급하는 장점도 갖고 있다(Chen and Guestrin, 2016).

4. 결과 및 토의

4.1 기계학습 모형의 성능

TPOT을 활용한 초기 모형 성능 탐색에서 가장 우수한 성능을 보인 Extra Trees (ET) 모형과 이와 유사한 성능을 보인

Random Forest (RF), Extreme Gradient Boosting (XGB) 모형을 대상으로 추가로 TPOT을 이용해 개별 모형 별 변수 최적화를 진행하였다. 최종적으로 동해 남서부 해역의 순군집생산에 활용한 모형별 순군집생산 추정치와 현장 관측 결과(시험 자료)를 비교하여 Table 3와 Fig. 3에 나타냈다. ET 모형의 RMSE가 6.1 mmol O_2 m^{-2} d^{-1}로 RF나 XGB의 8 mmol O_2 m^{-2} d^{-1}에 비해 낮았고, R^2는 0.96으로 가장 높았다.

순군집생산에 중요한 역할을 하는 입력 변수를 파악하기 위해 가장 우수한 성능을 보였던 ET 모형에서 출력하는 상대적 중요도(feature importance)를 살펴보았다(Fig. 4(a)). 각 변수의 중요도는 0과 1 사이의 값을 가지며, 변수별 중요도의 합은 1이다. 위도와 경도의 공간 정보가 가장 중요도가 높았고, 시간 정보는 가장 중요도가 낮았다. 모형 입력 변수의 상대 중요도는

Table 3. Parameters of the models developed for the prediction of NCP. RMSE stands for root mean square errors between predicted values and observed data reserved for testing (n=80)

Model	Parameter	RMSE
Extra Trees	n_estimators=300, max_features=1.0, min_samples_leaf=1, min_samples_split=2, bootstrap=False	6.1
Random Forest	n_estimators=100, max_features=0.2, min_samples_leaf=1, min_samples_split=2, bootstrap=False	8.1
Extreme Gradient Boosting	n_estimators=300, learning_rate=0.1, max_depth=7, min_child_weight=3, subsample=0.7	8.0

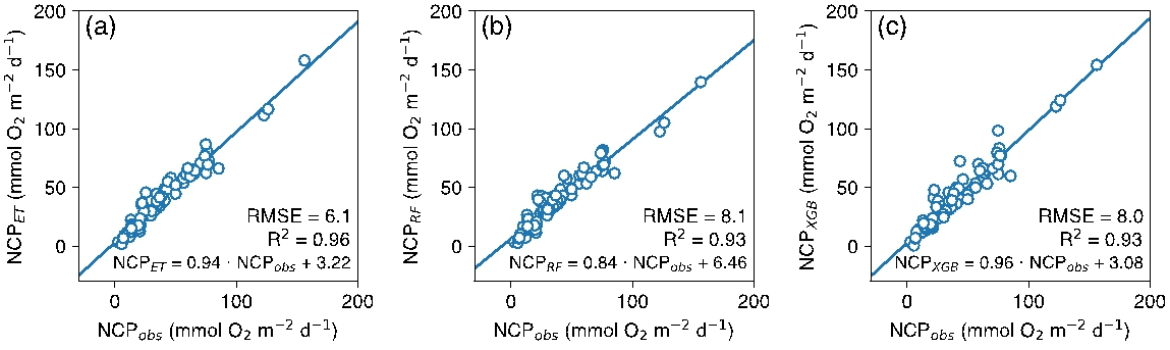

Fig. 3. Comparison between observed and predicted net community production by (a) Extra Trees (ET), (b) Random Forest (RF), and (c) Extreme Gradient Boosting (XGB) models.

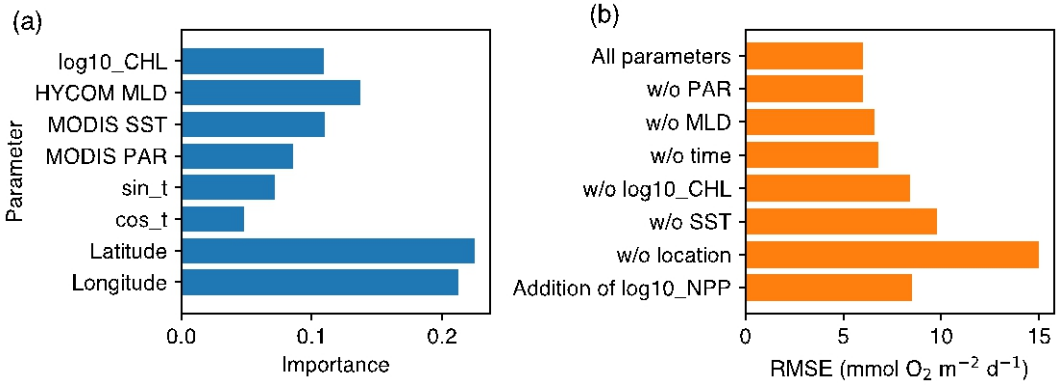

Fig. 4. (a) Relative importance of the parameter for the best performing ET model. (b) Root mean square errors (RMSE) of ET models with different sets of input parameters. 'All parameters' indicates the parameters in (a). The bar at the bottom shows RMSE when net primary production is added to 'all parameters' for the ET model.

변수 조합에 따라 달라질 수 있기에, 입력 변수를 한 가지씩 제외하였을 때 ET모형의 RMSE 변화를 추가로 살펴보았다(Fig. 4(b)). ET 모형에서 제공한 상대중요도에서 예측한 것과 같이 위경도 변수를 제외하게 되면 RMSE가 15 mmol O_2 m^{-1} d^{-1}로 가장 크게 증가하였다. 이처럼 위치 정보가 중요하게 작용하는 까닭은 대마 난류, 여름철 울산, 경주 연안에 종종 일어나는 연안 용승 등으로 인해 동해 연안과 외해의 큰 생산력 차이를 상당부분 위치 정보가 설명해주기 때문으로 여겨진다(Hahm *et al*., 2019a, 2019b). SST와 CHL을 제외한 경우 RMSE가 각각 9.8 mmol O_2 m^{-1} d^{-1}과 8.4 mmol O_2 m^{-1} d^{-1}로 PAR나 MLD에 비해 순군집생산 예측에 상대적으로 중요한 역할을 하는 것으로 나타났다. 입력 변수의 추가가 반드시 기계학습 모형의 성능을 증가시키지는 않았다. 예를 들어, Fig. 4(a)의 변수를 포함한 모형의 RMSE가 6.1 mmol O_2 m^{-1} d^{-1}이었음에 반해 일차생산력(Vertically Generalized Production Model for net primary production, VGPM-NPP)을 추가한 모형에서는 RMSE가 8.5 mmol O_2 m^{-1} d^{-1}로 오히려 증가하였다(Fig. 4(b)).

기계학습 모형의 시공간 예측 성능을 살펴보기 위해 7월 중 뚜렷한 연안 용승이 관측되었던 2013년의 월별 순군집생산 평균을 ET 모형으로 산출하였다(Fig. 5). 2013년 순군집생산은 울릉분지 중앙부(36.5 ― 37.5 °N, 130.5 ― 131.5 °E)를 기준으로 3, 4월에 50 mmol O_2 m^{-2} d^{-1} 내외로 가장 높았고, 여름에는 15 mmol O_2 m^{-2} d^{-1} 정도로 낮게 예측되었다. 7월에는 연안 용승으로 인해 울산, 경주 연안의 순군집생산이 100 mmol O_2 m^{-2} d^{-1} 이상으로 높았고, 용승한 해수가 울릉분지 쪽으로 이동함에 따라 순군집생산도 용승 해역 북동쪽으로 증가하는 것으로 예측했다. 그 결과 2013년 7월 중순에는 30 mmol O_2 m^{-2} d^{-1} 가량으로 순군집생산이 20 mmol O_2 m^{-2} d^{-1} 미만이었던 6월과 8월보다 높았다. 이 같은 결과는 연안 용승에 의한 일차생산력의 증가를 보고하고 있는 기존 연구들과도 일치하는 것이다(Hyun *et al*., 2009; Hahm *et al*., 2019a).

4.2 울릉분지 봄과 여름의 생물펌프

기계학습모형이 순군집생산의 계절 변화를 잘 반영하는가 살펴보기 위해 울릉 분지 중앙부에서 얻어진 다른 생물 펌프 척도(신생산, 입자성유기탄소 플럭스 등)와 본 연구에서 8일 간격으로 예측한 순군집생산을 비교하였다(Fig. 6). 질소 단위로 보고된 선행 연구와 비교하기 위해 순군집생산 자료를 O_2:N의 비를 170:16으로 가정하여 변환하였다(Anderson *et al.*, 1994). ET, RF, XGB 세 모형 모두에서 순군집생산은 3월 초(73일)에 최대(4 ― 5 mmol N m^{-2} d^{-1}), 8월 초(217일)에 최소(1

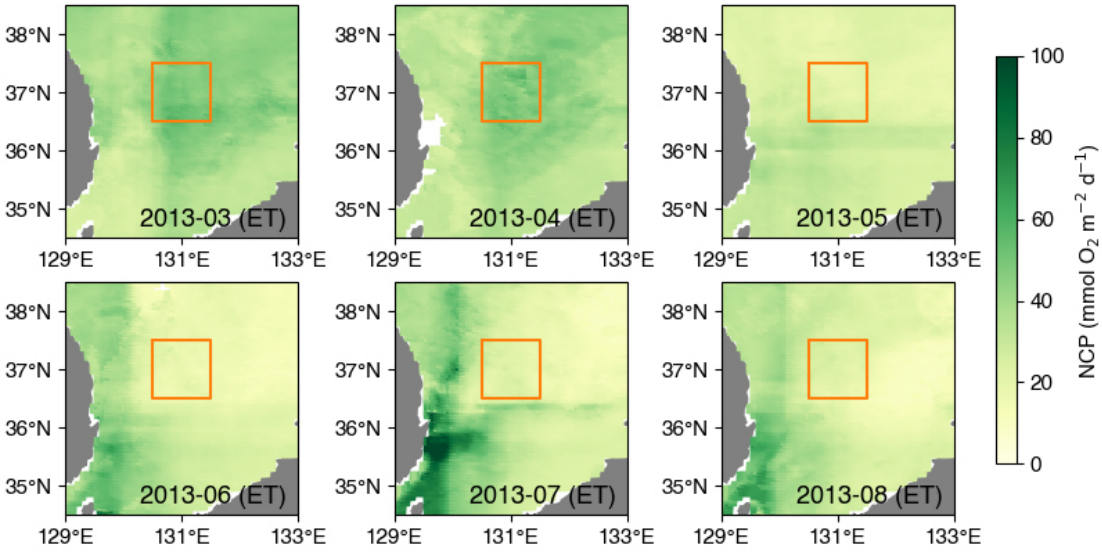

Fig. 5. Monthly net community production in 2013 predicted by the Extra Trees model.

— 2 mmol N m^{-2} d^{-1})로 나타났다. 순군집생산이 봄에서 여름으로 가며 점차 감소하는 경향은 Hahm and Kim (2001)이 ^{3}He의 기체교환율과 수온약층의 ^{3}He와 질산염의 관계식으로 추정한 질산염 수직 유입률과 유사하였다. ^{238}U-^{234}Th비평형법으로 추정한 입자성 유기 탄소 침강률도 4월(105일)과 8월(227일)에 각각 2.7 mmol N m^{-2} d^{-1}과 2.3 mmol N m^{-2} d^{-1}(Kim *et al*., 2011)로 순군집생산과 유사하였다.

Yang(1997, 1998)이 ^{15}N배양법으로 추정한 신생산도 봄에서 여름으로 가며 2.4 mmol N m^{-2} d^{-1}에서 0.2 — 0.7 mmol N m^{-2} d^{-1}로 감소하여 유사한 계절 변화를 보였다. 절대값은 순군집생산이 신생산에 비해 다소 높은 것으로 나타나는데, 이는 지화학 추적자(여기서는 O_2/Ar)를 이용한 현장(*in situ*) 측정 방법이 아중규모 혼합에 의한 영양염 공급을 포함하는 긴 시간과 넓은 공간의 생산력을 반영하기 때문에, 배양 용기의 부피로 공간 규모가 제한되어 있고 수 시간 내외의 짧은 시간 동안 이루어지는 배양법에 비해 생산력을 높게 평가한다는 선행 연구와 일치하는 것이다(Lipschultz *et al*., 2002; Stanley *et al*., 2015). 하지만 두 생물펌프 추정이 큰 시간 간격을 두고 독립적으로 이루어져 울릉분지의 특성을 반영하는 구체적인 원인을 살펴보기는 어렵다. Kwak *et al*.(2013b)은 월별 ^{15}N 배양법으로 측정한 신생산을 근거로 연평균 신생산을 12 ± 5 mmol N m^{-2} d^{-1}로 보고하였다. 이는 앞서 소개한 선형 연구나 본 연구에서 추정한 평균 순군집생산 2.5 mmol N m^{-2} d^{-1}에 비해 현저히 큰 값이다.

Hahm and Kim(2001)에서 유의한 질산염의 수직 공급이 없는 것으로 간주한 7-8월(180-240일)에도 기계학습으로 추정한 순군집생산과 입자성유기탄소 방출률은 모두 최대 2 mmol N m^{-2} d^{-1} 가량 존재하는 것으로 나타났다. 이는 Hahm and Kim(2001)이 추정한 질산염 수직 공급이 혼합층 수심이 증가함에 따른 질산염 혼입(entrainment)의 영향이 지배적인 점을 감안하면, 연안 용승 해수의 유입(Hyun *et al*., 2009; Hahm *et al*., 2019a)이나 대마 난류의 질산염 약층이 얕아지는 현상(Kim *et al*., 2013; Kwak *et al*., 2013a)에 따른 수평 이류에 의한 순군집생산으로 추정된다. 또, 동해 남부해역의 연평균 일차생산을 5 - 9% 증가시키는 것으로 알려져 있는 대기 침적도 여름철 순군집생산에 기여할 것으로 여겨진다(Onitsuka *et al*., 2009; Lee and Yoo, 2016).

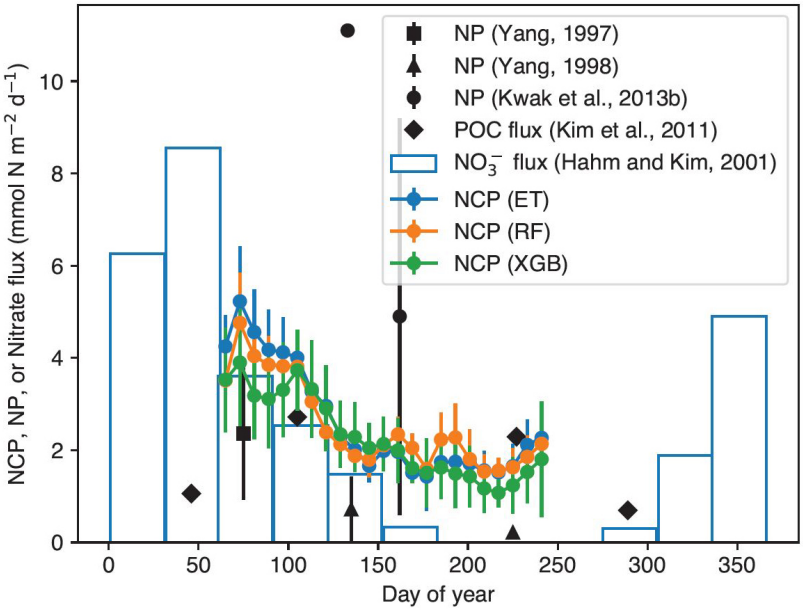

Fig. 6. Time series of net community production in the Ulleung Basin (36.5 — 37.5 °N, 130.5 — 131.5 °E) estimated by the machine learning algorithms. The circles and error bars indicate the averages and standard deviations of the values predicted for the years between 2013 and 2021. Previous estimates of the biological pump are shown for comparison. NP and POC stand for new production and particulate organic carbon.

4.3 모형 개선 방향

ET 기계학습 모형의 RMSE 6.1 mmol O_2 m^{-2} d^{-1}은 관측값 평균의 15%에 해당하는 것이다. 이는 O_2/Ar- NCP 추정법이 갖는 오차 30%(Hamme and Emerson, 2006)나 배양법의 오차 20%(Duarte *et al*., 2013)에 비해 작아 대체로 모형의 성능은 받아들일 만하다. 하지만 여름에 비해 순군집생산이 높은 봄철 현장 관측 비중이 17%로 제한되어 있어 향후 연간 순군집생산 추정의 오차를 줄이기 위해서는 봄철에 추가적인 현장 관측 자료를 축적할 필요가 있다.

Li and Cassar(2016)에서 일차생산(net primary production, NPP)을 대양의 NCP 추정에 활용한 점을 고려하면, 본 연구에서 활용한 입력 변수(SST, CHL, MLD, PAR) 에 NPP를 추가하였을 때 RMSE 증가한 점(Fig. 4(b))은 추가적인 탐색이 필요하다. 본 연구에서 이용하였던 전 대양 알고리즘(VGPM-NPP; Behrenfeld and Falkowski, 1997) 대신 동해에 최적화된 알고리즘(Yamada *et al*., 2005)을 적용한 NPP를 입력 변수 조합에 포함시켰을 때 RMSE를 더 줄일 수 있나 추가적인 연구가 가능하겠다.

본 연구의 기계학습법을 적용하여 연간 순군집생산을 추정하기 위해서는 가을과 겨울에 혼합층 수심이 깊어짐에 따라 산소 농도가 낮은 저층수의 표층 유입으로 표층의 순군집생산이 저평가되는 것을 보정하기 위한 방법이 필요하다. 울릉분지 중앙부, 연안 용승해역 등 수직 혼입의 양상이 서로 다른 해역별로 O_2/Ar(혹은 O_2) 수직 분포를 측정하고, 동시에 수직 확산 계수를 측정함으로써 저층수의 혼입을 보정(Haskell *et al*., 2016; Lee *et al*., 2021)할 수 있겠다. 추가적으로 수직 혼합에 영향을 받지 않는 ^{18}O 배양법(Ferrón *et al*., 2016)으로 독립적으로 얻은 순군집생산과 표층 O_2/Ar을 이용한 순군집생산 결과를 비교하여 혼입 영향을 보정하는 방법 등을 향후 연구에 포함시킬 계획이다.

5. 결 론

2013년부터 2021년 사이 7회의 독립적인 항해를 통해 표층 해수 O_2/Ar 측정하여 동해 남서부해역의 순군집생산(O_2/Ar-NCP)을 추정하였다. 고해상도(<1 km)로 얻어진 순군집생산 현장 관측 자료와 기계학습 모형을 결합하여 4 km 공간 해상도와 8일 시간 해상도를 갖는 봄과 여름의 순군집생산을 추정하였다. 기계학습 모형의 입력 자료로는 수온, 염록소, 혼합층 수심, 광합성 광량을 사용하였다. 3월 중 순군집생산은 50 mmol O_2 m^{-2} d^{-1} 가량으로 가장 높았고, 여름에는 15 mmol O_2 m^{-2} d^{-1}로 가장 낮았다. 이는 ^3He의 기체교환율로 추정한 질산염의 수직 공급률과 비슷한 정도이고, ^{238}U-^{234}Th 비평형법으로 추정한 입자 유기 탄소의 침강률과도 유사한 것이었다. 이 연구에서 표층 O_2/Ar 연속 측정으로 산출한 고해상도 순군집생산 자료와 기계학습 모형이 동해 남서부해역의 생물 펌프 추정에 효과적인 방법이 될 수 있음을 보였다. 앞 서 언급한 대로 혼합층 수심이 깊어짐에 따라 산소 농도가 낮은 저층수의 표층 유입을 보정하기 위한 추가적인 연구가 이루어진다면, 기계학습 모형의 예측 범위를 가을과 겨울로 넓혀 연간 순군집생산을 추정할 수 있을 것이다.

사 사

이 과제는 부산대학교 기본연구지원사업(2년)에 의하여 연구되었음.

참고문헌(References)

Anderson, L., A. Urence and J.L. Sarmiento, 1994. Redfield ratios of remineralization determined by nutrient data analysis. Global Biogeochemical Cycles, **8**(1): 65-80.

Behrenfeld, M.J. and P.G. Falkowski, 1997. Photosynthetic rates derived from satellite-based chlorophyll concentration, Limnology And Oceanography, **42**(1): 1-20.

Breiman, L., 2001. Random Forests. Machine Learning, **45**(1): 5-32.

Cassar, N., B.A. Barnett, M.L. Bender, J. Kaiser, R.C. Hamme and B. Tilbrook, 2009. Continuous High-Frequency Dissolved O_2/Ar Measurements by Equilibrator Inlet Mass Spectrometry. Analytical Chemistry, **81**: 1855-1864.

Chang, C.H., N.C. Johnson and N. Cassar, 2014. Neural network-based estimates of Southern Ocean net community production from in-situ O_2 / Ar and satellite observation: a methodological study. Biogeosciences, **11**: 3279-3297.

Chen, T. and C. Guestrin, 2016. Xgboost: A scalable tree boosting system. Proceedings of the 22nd Acm Sigkdd International Conference on Knowledge Discovery and Data Mining, 785-794.

Craig, H. and T. Hayward, 1987. Oxygen supersaturation in the ocean: Biological versus physical contributions. Science, **235**(4785), 199-202.

Duarte, C.M., A. Regaudie-de Gioux, J.M. Arrieta, A. Delgado-Huertas, and S. Agust´ı, 2013. The Oligotrophic Ocean is Heterotrophic. Annual Review Of Marine Science, **5**(1): 551-569.

European Centre for Medium-Range Weather Forecasts (ECMWF), 2023. Available at: https://www.ecmwf.int/en/forecasts/dataset/ecmwf-reanalysis-v5. [last accessed: October 10, 2023].

Ferrón, S., D.A. Valle, K.M. del, Björkman, P.D. Quay, M.J. Church and D.M. Karl, 2016. Application of membrane inlet mass spectrometry to measure aquatic gross primary production by the 18O in vitro method. Limnology and Oceanography: Methods, **14**(9): 610-622. DOI: https://doi.org/10.1002/lom3.10116.

Geurts, P., D. Ernst and L. Wehenkel, 2006. Extremely randomized trees. Machine Learning, **63**(1): 3-42. DOI: https://doi.org/10.1007/s10994-006-6226-1.

Gregor, L., S. Kok and P. Monteiro, 2017. Empirical methods for the estimation of Southern Ocean CO_2: support vector and random forest regression. Biogeosciences, **14**.

Gregor, L., T.J. Ryan-Keogh, S.-A. Nicholson, M. Plessis, I. du, Giddy and S. Swart, 2019. GliderTools: A Python Toolbox for Processing Underwater Glider Data. Frontiers In Marine Science, **6**.

Hahm, D. and I. Lee, 2018. Estimation of Net Community Production Based on O_2/Ar Measurements. Journal of The Korean Society of Oceanography, **23**(1): 49-62.

Hahm, D. and K.-R. Kim, 2001. An estimation of the new production in the southern East Sea using helium isotopes. Journal Of The Korean Society Of Oceanography, **36**(1): 19-26.

Hahm, D., S. Park, S.-H. Choi, D.-J. Kang, T. Rho and T. Lee, 2019b. Estimation of surface fCO_2 in the southwest east sea using machine learning techniques. The Sea, **24**(3): 375-388.

Hahm, D., T.S. Rhee, H.-C. Kim, C.J. Jang, Y.S. Kim and J.-H. Park, 2019a. An observation of primary production enhanced by coastal upwelling in the southwest east/japan sea. Journal of Marine Systems, **195**: 30-37.

Hamme, R. and S. Emerson, 2006. Constraining bubble dynamics and mixing with dissolved gases: Implications for productivity measurements by oxygen mass balance. Journal of Marine Research, **64**(1): 73.

Haskell, W.Z., M.G. Prokopenko, R.H.R. Stanley and A.N. Knapp, 2016. Estimates of vertical turbulent mixing used to determine a vertical gradient in net and gross oxygen production in the oligotrophic South Pacific Gyre. Geophysical Research Letters, **43**(14): 7590-7599.

Hyun, J., D. Kim, C. Shin, J. Noh, E. Yang, J. Mok, S. Kim, H. Kim and S. Yoo, 2009. Enhanced phytoplankton and bacterioplankton production coupled to coastal upwelling and an anticyclonic eddy in the Ulleung basin, East Sea. Aquatic Microbial Ecology, **54**: 45-54.

Joo, H., D. Lee, S.H. Son and S.H. Lee, 2018. Annual new production of phytoplankton estimated from MODIS-derived nitrate concentration in the East/Japan Sea. Remote Sensing, **10**(5), DOI: 10.3390/rs10050806.

Joo, H.T., S.H. Son, J.W. Park, J.J. Kang, J.-Y. Jeong, C.I. Lee, C.-K. Kang and S.H. Lee, 2016. Long-term pattern of primary productivity in the East/Japan sea based on ocean color data derived from MODIS-Aqua. Remote Sensing, **8**(1): 25.

Kaiser, J., M.K. Reuer, B. Barnett and M.L. Bender, 2005. Marine productivity estimates from continuous O_2/Ar ratio measurements by membrane inlet mass spectrometry. Geophysical Research Letters, **32**(19).

Kim, D., E.J. Yang, K.H. Kim, C.W. Shin, J. Park, S. Yoo and J.H. Hyun, 2012. Impact of an anticyclonic eddy on the summer nutrient and chlorophyll a distributions in the Ulleung Basin, East Sea (Japan Sea). Ices Journal of Marine Science, **69**(1): 23-29.

Kim, D., M.-S. Choi, H.-Y. Oh, Y.-H. Song, J.-H. Noh and K.H. Kim, 2011. Seasonal export fluxes of particulate organic carbon from 234Th/238U disequilibrium measurements in the ulleung Basin1 (tsushima basin) of the east Sea1 (sea of japan). Journal of Oceanography, **67**(5): 577-588. DOI: https://doi.org/10.1007/s10872-011-0058-8.

Kim, S.-K., K.-I. Chang, B. Kim and Y.-K. Cho, 2013. Contribution of ocean current to the increase in N abundance in the Northwestern Pacific marginal seas. Geophysical Research Letters, **40**(1): 143-148.

Kwak, J.H., J. Hwang, E.J. Choy, H.J. Park, D.-J. Kang, T. Lee, K.-I. Chang, K.-R. Kim and C.-K. Kang, 2013a. High primary productivity and f-ratio in summer in the Ulleung basin of the East/Japan Sea. Deep Sea Research I, **79**: 74-85.

Kwak, J.H., S.H. Lee, H.J. Park, E.J. Choy, H.D. Jeong, K.R. Kim and C.K. Kang, 2013b. Monthly measured primary and new productivities in the Ulleung Basin as a biological "hot spot" in the East/Japan Sea. Biogeosciences, **10**(7): 4405-4417.

Le, T.T., W. Fu and J.H. Moore, 2020. Scaling tree-based automated machine learning to biomedical big data with a feature set selector. Bioinformatics, **36**(1): 250-256. DOI: https://doi.org/10.1093/bioinformatics/btz470.

Lee, I., D. Hahm, D. Shin, C.-S. Hong, S. Nam, G. Kim and T. Lee, 2021. Determination and uncertainty of spring net community production estimated from O_2/Ar measurements in the northern East China Sea and southern Yellow Sea. Continental Shelf Research, **230**: 104570.

Lee, S. and S. Yoo, 2016. Interannual variability of the phytoplankton community by the changes in vertical mixing and atmospheric deposition in the ulleung basin, east sea: A modelling study. Ecological Modelling, **322**: 31-47. DOI: https://doi. org/https://doi.org/10.1016/j.ecolmodel.2015.11.012.

Li, Z. and N. Cassar, 2016. Satellite estimates of net community production based on O 2/Ar observations and comparison to other estimates. Global Biogeochemical Cycles, **30**(5): 735-752.

Lipschultz, F., N.R. Bates, C.A. Carlson and D.A. Hansell, 2002. New production in the Sargasso Sea: History and current status. Global Biogeochemical Cycles, **16**(1): 1-1-1-17, DOI: https://doi.org/10.1029/2000GB001319.

Nakaoka, S., M. Telszewski, Y. Nojiri, S. Yasunaka, C. Miyazaki, H. Mukai and N. Usui, 2013. Estimating temporal and spatial variation of ocean surface pCO_2 in the North Pacific using a self-organizing map neural network technique. Biogeosciences, **10**(9): 6093-6106.

NASA, 2023. Available at: https://oceancolor.gsfc.nasa.gov/l3/. [last accessed: October 10, 2023].

Ocean Productivity, 2023. Available at: https://sites.science.oregonstate.edu/ocean.productivity/index.php. [last accessed: October 10, 2023].

Onitsuka, G., I. Uno, T. Yanagi and J.-H. Yoon, 2009. Modeling the effects of atmospheric nitrogen input on biological production in the japan sea. Journal of Oceanography, **65**(3): 433-438. DOI: https://doi.org/10.1007/s10872-009-0038-4.

Onitsuka, G., T. Yanagi and J.-H. Yoon, 2007. A numerical study on nutrient sources in the surface layer of the Japan Sea using a coupled physical-ecosystem model. Journal Of Geophysical Research-Oceans, **112**(C5): C05042.

Park, K., D. Hahm, J.O. Choi, S. Xu, H.-C. Kim and S. Lee, 2019. Spatiotemporal variation in summer net community production in the amundsen sea polynya: A self-organizing map analysis approach. Continental Shelf Research, **184**: 21-29.

Scikit Learn, 2023. Available at: https://scikit-learn.org. [last accessed: October 10, 2023].

Stanley, R.H.R., W.J. Jenkins, S.C. Doney and D.E. Lott III, 2015. The ^{3}He flux gauge in the Sargasso Sea: a determination of physical nutrient fluxes to the euphotic zone at the Bermuda Atlantic Time-series Site. Biogeosciences, **12**(17): 5199-5210.

Teeter, L., R.C. Hamme, D. Ianson and L. Bianucci, 2018. Accurate estimation of net community production from O2/ar measurements. Global Biogeochemical Cycles, **32**(8): 1163-1181.

Volk, T. and M.I. Hoffert, 1985. Ocean Carbon Pumps: Analysis of Relative Strengths and Efficiencies in Ocean-Driven Atmospheric CO_2 Changes, in The Carbon Cycle and Atmospheric CO_2: Natural Variations Archean to Present, pp. 99-110, American Geophysical Union, Washington, D.C.

Wanninkhof, R. 2014. Relationship between wind speed and gas exchange over the ocean revisited. Limnol. Oceanogr.: Methods, **12**(6): 351-362.

XGBoost, 2023. Available at: https://xgboost.readthedocs.io. [last accessed: October 10, 2023].

Yamada, K., J. Ishizaka and H. Nagata, 2005. Spatial and Temporal Variability of Satellite Primary Production in the Japan Sea from 1998 to 2002. Journal Of Oceanography, **61**(5): 857-869.

Yang, S.R., 1997. Primary production in the ocean-waste disposal area in: Development of monitoring technology for the wastes disposal sea areas. 2nd KORDI report. Korea Ocean Research; Development Institute.

Yang, S.R., 1998. Primary production in the ocean-waste disposal area in: Development of monitoring technology for the wastes disposal sea areas. 3rd KORDI report,. Korea Ocean Research; Development Institute.

Yoo, S. and J. Park, 2009. Why is the southwest the most productive region of the East Sea/Sea of Japan? Journal Of Marine Systems, **78**(2): 15-15.

Zeng, J., Y. Nojiri, P. Landschützer, M. Telszewski and S. Nakaoka, 2014. A Global Surface Ocean fCO_2 Climatology Based on a Feed-Forward Neural Network. Journal Of Atmospheric And Oceanic Technology, **31**(8): 1838-1849.