
인공신경망모형을 이용한 양식수산물 단위 전망에 관한 연구 - 김 양식을 중심으로 -

A Study on the Forecasting Unit Production of Aquacultural Product
- Focusing on Seaweed Farming-

천성훈 · 김철현 · 노아현 · 허수진



한국해양수산개발원
KOREA MARITIME INSTITUTE

저자	천성훈, 김철현, 노아현, 허수진			
내부연구진	연구책임자	천성훈	한국해양수산개발원	수산정책사업본부 연구원
	공동연구원	김철현	한국해양수산개발원	수산정책사업본부 전문연구원
	공동연구원	노아현	한국해양수산개발원	수산정책사업본부 전문연구원
	공동연구원	허수진	한국해양수산개발원	수산정책사업본부 전문연구원

연구기간 2021. 4. 7. ~ 2021. 10. 31.

보고서 집필내역

연구책임자	천성훈 연구총괄, 제1장, 제3장, 제4장, 제5장, 제6장
내부연구진	김철현 제3장, 제4장
	노아현 제2장, 제6장
	허수진 제4장, 제5장

산·학·연·정 연구자문위원	김동우	포항공과대학교	인공지능대학원	조교수
	이진면	산업연구원	동향·통계분석본부	선임연구위원
	박은정	국립수산과학원	수산식품품질관리센터장	
	황미숙	국립수산과학원	수산종자육종연구소장	

※ 순서는 산·학·연·정 순임

발간사

지속적인 경제성장에도 불구하고 우리나라의 식량 자급률은 하락하는 추세를 나타내고 있으며 식량 안보지수는 경제협력개발기구(OECD) 가입국 중 최하위권 수준입니다. 유엔 식량농업기구(FAO)와 세계식량계획(WFP)은 코로나19 사태 이후 세계적인 식량 위기를 경고하였고, 주요 식량 생산국은 자국의 식량안보 강화를 명목으로 식량 수출을 제한하거나 중단하기에 이르렀습니다. 국제 곡물가격이 급등하면서 식량 위기에 대한 우려가 커지고 있으며 세계 각국은 자국의 식량 자급률을 높이기 위해 만전을 기하고 있습니다. 이러한 가운데 세계 최고 수준의 1인당 수산물 소비량을 기록 중인 우리나라에 있어 수산업은 식량안보의 한 축으로서 필수 불가결한 요소입니다.

우리나라는 1970년대 세계 3대 원양어업국으로 손꼽힐 정도로 수산업의 지위가 대단했으나, 오늘날에 이르러서는 어족자원의 남획과 기후변화, 그리고 국제적인 규제 강화로 잡는 어업의 성장이 한계에 다다랐습니다. 그럼에도 불구하고 고부가가치 어종 중심의 기르는 어업의 발전과 함께 우리나라 수산업의 성장은 여전히 지속되고 있습니다. 또한 세계의 저명한 학자들은 양식산업을 21세기 주요 산업으로 지목하였으며 세계적인 수산물 소비는 증가세를 보이고 있습니다. 코로나 펜데믹 이후 식량안보에 대한 경각심이 더욱 커지고 있는 만큼, 국가 차원의 정책지원을 통해 수산업을 육성하고 보호해야 할 필요성이 높아지고 있습니다.

올바른 정책 의사결정을 위해서는 산업에 대한 진단과 전망이 뒷받침되어야 합니다. 본 연구는 우리나라 수산업 분야의 사회과학 연구에는 아직 보편적이지 않은 인공 신경망을 활용하여 수산 정책 의사결정에 필요한 수급 전망 수단과 활용방안을 제시하였습니다. 해양수산 학연 협력사업의 지원을 받았으며, 포항공과대학교 인공지능대학원 김동우 교수와 긴밀히 협력하였습니다. 익숙하지 않은 분야임에도 연구를 제안하고 무사히 수행한 4인의 우리원 연구진, 코로나19로 인해 어려운 현실 속에서도 높은 수준의 자문과 아낌없는 협조를 제공해주신 김동우 교수님, 그리고 연구 수행에 도움을 주신 모든 분께 이 지면을 빌어 감사의 말씀을 전합니다.

2021년 12월
한국해양수산개발원
원장 김 종 덕

목차

요약 _ i

Executive Summary _ ix

01	서론 _1	
	제1절 연구 배경 및 목적	1
	1. 연구 배경 및 필요성	1
	2. 연구 목적	3
	제2절 연구 내용 및 방법	5
	1. 연구 내용	5
	2. 연구 방법	7
	제3절 선행연구 검토	9
	1. 국내 선행연구	9
	2. 해외 선행연구	13
	3. 선행연구와의 차별성	17
02	김 수급 동향 및 김 양식 방법 _19	
	제1절 김 수급 동향	19
	1. 생산 동향	20
	2. 가격 동향	24
	3. 수출 동향	26
	제2절 김 양식 방법 및 지역별 양식 환경	28
	1. 김의 종류 및 생육 조건	28
	2. 김 양식 방법 및 지역별 양식 환경	33
03	인공신경망모형의 이론적 고찰 _53	
	제1절 인공신경망모형 개요	53

1. 인공신경망모형	53
2. 인공신경망모형의 종류	55
제2절 인공신경망모형의 학습	68
1. 손실함수	68
2. 오차역전파와 경사 하강법	69
3. 과적합	73

04 예측모형 설계_77

제1절 분석자료	77
1. 분석자료별 기초통계량	78
2. 상관관계 분석	84
제2절 모형 설계	93
1. 지역별 DB	93
2. 인공신경망 예측모형 및 학습조건 설계	96

05 예측모형 추정 결과_101

제1절 인공신경망모형	101
1. 전남	102
2. 완도	108
3. 고흥	114
4. 종합	120
제2절 비교모형(다중회귀모형)	141
1. 개별방정식 예측력 검정	143
2. 종합	145

06 결론_153

제1절 요약 및 결론	153
-------------	-----

1. 요약	153
2. 결론	156
제2절 제언	158
1. 본 연구의 기여	158
2. 제언	159

참고문헌_163

부록_171

표 목차

〈표 1-1〉 주요 연구 내용	6
〈표 1-2〉 국내 선행연구(자연과학 분야)	9
〈표 1-3〉 국내 선행연구(농업 분야)	10
〈표 1-4〉 국내 선행연구(수산업 분야)	12
〈표 1-5〉 해외 선행연구	13
〈표 2-1〉 연도별 김 양식 면허 및 어가 수	20
〈표 2-2〉 연산별 김 생산량	22
〈표 2-3〉 연산별·지역별 김 생산량	23
〈표 2-4〉 연산별 김 산지가격	24
〈표 2-5〉 연도별 김 수출량 추이	27
〈표 2-6〉 주요 김 품종과 양식법	35
〈표 2-7〉 서해권 김 양식 어장 표층 수온(2021. 3. 기준)	39
〈표 2-8〉 서해권 김 양식 어장 용존무기인(DIP) (2021. 3. 기준)	40
〈표 2-9〉 서해권 김 양식 어장 용존무기질소(DIN) (2021. 3. 기준)	41
〈표 2-10〉 남서해권 김 양식 어장 표층 수온(2021. 2. 기준)	43
〈표 2-11〉 남서해권 김 양식 어장 용존무기인(DIP) (2021. 2. 기준)	44
〈표 2-12〉 남서해권 김 양식 어장 용존무기질소(DIN) (2021. 2. 기준)	45
〈표 2-13〉 동해권 김 양식 어장 표층 수온(2021. 2. 기준)	46
〈표 2-14〉 동해권 김 양식 어장 용존무기인(DIP) (2021. 2. 기준)	47
〈표 2-15〉 동해권 김 양식 어장 용존무기질소(DIN) (2021. 2. 기준)	48
〈표 2-16〉 2017~18년산 해조류 양식피해 조사 목록	49
〈표 4-1〉 단수 기초통계량	78
〈표 4-2〉 기상변수(완도) 기초통계량	79
〈표 4-3〉 기상변수(해남) 기초통계량	80
〈표 4-4〉 기상변수(고흥) 기초통계량	81
〈표 4-5〉 해양변수(완도) 기초통계량	82
〈표 4-6〉 해양변수(고흥) 기초통계량	83
〈표 4-7〉 전라남도 단수와 완도 기상변수 상관계수	85
〈표 4-8〉 전라남도 단수와 해남 기상변수 상관계수	85
〈표 4-9〉 전라남도 단수와 고흥 기상변수 상관계수	86

〈표 4-10〉 완도 단수와 기상변수 상관계수	87
〈표 4-11〉 고흥 단수와 기상변수 상관계수	87
〈표 4-12〉 전라남도 단수와 완도 해황변수 상관계수	88
〈표 4-13〉 전라남도 단수와 고흥 해황변수 상관계수	89
〈표 4-14〉 완도 단수와 해황변수 상관계수	89
〈표 4-15〉 고흥 단수와 해황변수 상관관계	90
〈표 4-16〉 완도 기상변수와 해황변수의 상관계수	91
〈표 4-17〉 고흥 기상변수와 해황변수의 상관계수	91
〈표 4-18〉 지역별 단수 예측모형에 사용된 독립변수	95
〈표 4-19〉 학습 조건별 세부 설정값	99
〈표 5-1〉 예측모형별 구조	101
〈표 5-2〉 다층 퍼셉트론 추정 결과(전남)	103
〈표 5-3〉 장단기 기억신경망 추정 결과(전남)	105
〈표 5-4〉 양방향 장단기 기억신경망 추정 결과(전남)	107
〈표 5-5〉 다층 퍼셉트론 추정 결과(완도)	109
〈표 5-6〉 장단기 기억신경망 추정 결과(완도)	111
〈표 5-7〉 양방향 장단기 기억신경망 추정 결과(완도)	113
〈표 5-8〉 다층 퍼셉트론 추정 결과(고흥)	115
〈표 5-9〉 장단기 기억신경망 추정 결과(고흥)	117
〈표 5-10〉 양방향 장단기 기억신경망 추정 결과(고흥)	119
〈표 5-11〉 전남지역 단기 예측모형 김 단수 추정 결과(인공신경망)	123
〈표 5-12〉 전남지역 장기 예측모형 김 단수 추정 결과(인공신경망)	125
〈표 5-13〉 전남지역 어기 예측모형1 김 단수 추정 결과(인공신경망)	126
〈표 5-14〉 전남지역 어기 예측모형2 김 단수 추정 결과(인공신경망)	128
〈표 5-15〉 완도지역 단기 예측모형 김 단수 추정 결과(인공신경망)	129
〈표 5-16〉 완도지역 장기 예측모형 김 단수 추정 결과(인공신경망)	130
〈표 5-17〉 완도지역 어기 예측모형1 김 단수 추정 결과(인공신경망)	132
〈표 5-18〉 완도지역 어기 예측모형2 김 단수 추정 결과(인공신경망)	133
〈표 5-19〉 고흥지역 단기 예측모형 김 단수 추정 결과(인공신경망)	135
〈표 5-20〉 고흥지역 장기 예측모형 김 단수 추정 결과(인공신경망)	136

〈표 5-21〉 고흥지역 어기 예측모형1 김 단수 추정 결과(인공신경망)	138
〈표 5-22〉 고흥지역 어기 예측모형2 김 단수 추정 결과(인공신경망)	140
〈표 5-23〉 비교모형 개별방정식 구조	141
〈표 5-24〉 개별방정식 예측력 검정 결과	144
〈표 5-25〉 전남지역 단기 예측모형 김 단수 추정 결과(다중회귀)	145
〈표 5-26〉 전남지역 장기 예측모형 김 단수 추정 결과(다중회귀)	146
〈표 5-27〉 완도지역 단기 예측모형 김 단수 추정 결과(다중회귀)	148
〈표 5-28〉 완도지역 장기 예측모형 김 단수 추정 결과(다중회귀)	149
〈표 5-29〉 고흥지역 단기 예측모형 김 단수 추정 결과(다중회귀)	150
〈표 5-30〉 고흥지역 장기 예측모형 김 단수 추정 결과(다중회귀)	152
〈표 6-1〉 지역별 2021년산 김 단수 추정모형 비교(1개월 추정모형)	154
〈표 6-2〉 지역별 2021년산 김 단수 추정모형 비교(어기 추정모형, 인공신경망) ...	155
〈표 6-3〉 연구 개선방안	160

그림 목차

〈그림 1-1〉 연구 목적 및 목표	4
〈그림 1-2〉 연구 수행 체계	7
〈그림 2-1〉 연도별 김 시설량 추이	21
〈그림 2-2〉 연도별 김 생산량 추이	22
〈그림 2-3〉 연산별·지역별 김 생산량 추이	23
〈그림 2-4〉 물김 월별 산지가격 추이	25
〈그림 2-5〉 마른김 월별 도매가격 추이(김밥용김 중품 기준)	26
〈그림 2-6〉 월별 김 수출량 추이	27
〈그림 2-7〉 주요 양식 김의 종류	29
〈그림 2-8〉 지주식 김 양식장의 만조 시 김발 부동과 간조 시 김발 노출	33
〈그림 2-9〉 부류식의 뒤집기에 의한 노출(왼쪽 : 수면부동, 오른쪽 : 노출)	34
〈그림 2-10〉 서천지역 김 양식시설 평면도	36
〈그림 2-11〉 부산지역 김 양식시설 평면도	37
〈그림 2-12〉 서해권 수온 분포도	39
〈그림 2-13〉 서해권 용존무기인(DIP) 분포도	40
〈그림 2-14〉 서해권 용존무기질소(DIN) 분포도	42
〈그림 2-15〉 남서해권 수온 분포도	43
〈그림 2-16〉 남서해권 용존무기인(DIP) 분포도	44
〈그림 2-17〉 남서해권 용존무기질소(DIN) 분포도	45
〈그림 2-18〉 동해권 수온 분포도	46
〈그림 2-19〉 동해권 용존무기인(DIP) 분포도	47
〈그림 2-20〉 동해권 용존무기질소(DIN) 분포도	48
〈그림 3-1〉 인공신경망모형의 구조	55
〈그림 3-2〉 활성화함수의 종류	58
〈그림 3-3〉 다층 퍼셉트론모형 학습을 위한 행렬데이터 변환	61
〈그림 3-4〉 순환신경망의 구조	62
〈그림 3-5〉 순환신경망의 입출력 형태	63
〈그림 3-6〉 장단기 기억신경망의 구조	64
〈그림 3-7〉 양방향 순환신경망의 구조	66
〈그림 3-8〉 심층신경망의 구조	67

〈그림 3-9〉 과적합된 모형의 손실함수	73
〈그림 3-10〉 Dropout 적용 후 신경망 구조의 변화	75
〈그림 4-1〉 김 단수 예측을 위한 인공신경망모형의 구조	97
〈그림 4-2〉 예측 시차별 종속변수 및 독립변수(예시)	98
〈그림 5-1〉 교차검증데이터의 순환	120
〈그림 5-2〉 예측시차별 시험오차와 검증오차의 산포도	121
〈그림 5-3〉 전남지역 단기 예측모형 김 단수 추정 결과(인공신경망)	124
〈그림 5-4〉 전남지역 장기 예측모형 김 단수 추정 결과(인공신경망)	125
〈그림 5-5〉 전남지역 어기 예측모형1 김 단수 추정 결과(인공신경망)	127
〈그림 5-6〉 전남지역 어기 예측모형2 김 단수 추정 결과(인공신경망)	128
〈그림 5-7〉 완도지역 단기 예측모형 김 단수 추정 결과(인공신경망)	129
〈그림 5-8〉 완도지역 장기 예측모형 김 단수 추정 결과(인공신경망)	131
〈그림 5-9〉 완도지역 어기 예측모형1 김 단수 추정 결과(인공신경망)	132
〈그림 5-10〉 완도지역 어기 예측모형2 김 단수 추정 결과(인공신경망)	134
〈그림 5-11〉 고흥지역 단기 예측모형 김 단수 추정 결과(인공신경망)	135
〈그림 5-12〉 고흥지역 장기 예측모형 김 단수 추정 결과(인공신경망)	137
〈그림 5-13〉 고흥지역 어기 예측모형1 김 단수 추정 결과(인공신경망)	139
〈그림 5-14〉 고흥지역 어기 예측모형2 김 단수 추정 결과(인공신경망)	140
〈그림 5-15〉 전남지역 단기 예측모형 김 단수 추정 결과(다중회귀)	146
〈그림 5-16〉 전남지역 장기 예측모형 김 단수 추정 결과(다중회귀)	147
〈그림 5-17〉 완도지역 단기 예측모형 김 단수 추정 결과(다중회귀)	148
〈그림 5-18〉 완도지역 장기 예측모형 김 단수 추정 결과(다중회귀)	149
〈그림 5-19〉 고흥지역 단기 예측모형 김 단수 추정 결과(다중회귀)	151
〈그림 5-20〉 고흥지역 장기 예측모형 김 단수 추정 결과(다중회귀)	152
〈그림 6-1〉 수급 전망체계 구축 및 활용방안	161

요약

1. 연구 목적

- 본 연구에서는 인공신경망모형을 이용하여 양식어장의 환경(기상 및 수질)을 반영한 정량적 단수 예측모형을 구축하였음
 - 양식수산물의 단수(단위생산량)는 어장의 생산성 및 작황을 나타내는 지표로서, 양식어장의 기상 및 수질 변동에 직·간접적 영향을 받음
 - 양식수산물의 생산량 변화로 인한 정책적 의사결정이 필요할 경우 정량적인 생산량 예측이 필요하지만, 어장의 환경조건(기상, 수질 등)을 이용한 양식수산물 생산 예측을 위한 연구는 미비한 실정임
 - 따라서 본 연구에서는 양식수산물 단수 전망에 있어 환경변수의 활용 및 예측력 제고를 위해 인공신경망모형을 이용하였음
- 여러 가지 조건의 인공신경망모형과 다중회귀모형의 결과를 비교하고 이를 바탕으로 정책적 활용방안을 제시하고자 함
 - MLP, LSTM, 양방향 LSTM 모형과 투입변수 및 세부 파라미터 설정을 달리한 여러 가지 모형을 추정하고 다중회귀모형과 결과를 비교하였음
 - 비교 결과를 바탕으로 결론을 도출하고 정책적 활용방안 및 실무적 활용을 위한 향후 과제를 도출하고자 하였음

2. 연구 방법 및 특징

1) 연구 방법

- 인공신경망모형으로 기상 및 수질 데이터를 반영한 지역별 양식 김 단수 예측모형을 구축하였음
 - 인공신경망모형을 이용한 김 단수 전망에는 지역, 예측 시차, 독립변수의 구분(기상/수질), 인공신경망 알고리즘의 종류(MLP, LSTM, 양방향 LSTM)에 따라 총 216가지 모형을 학습하고 결과를 비교하였음
 - 모형의 예측력 평가지표로는 MSE, MAPE, Theil's U 계수를 사용함
- 인공신경망모형의 평가를 위한 비교모형으로 다중회귀모형을 설정함
 - 앞서 제시된 예측력 평가지표 외에 구축된 인공신경망모형을 평가하기 위한 수단으로서 다중회귀모형을 설정함
 - 비교를 위해 동등한 조건의 변수를 독립변수로 이용하거나 유의성이 높은 변수를 선별하여 사용한 모형을 각각 도출하였음

〈요약 표 1〉 분석 방법

분석 DB 구축	• 기상변수(강수량, 풍속, 일조량 등) • 수질변수(수온, DO, DIN, DIP 등)
인공신경망모형 구축	• MLP, LSTM, 양방향 LSTM • 변수 및 파라미터 설정 다양화
다중회귀모형 구축 (비교모형)	• 인공신경망모형 평가를 위해 동일 조건 및 최적 조건 모형 분석
모형 비교 및 평가	• 인공신경망모형 및 다중회귀모형 예측 결과 비교 • 예측력 평가지표 비교
결론 제시	• 분석 결과를 바탕으로 결론 도출

2) 연구의 특징

- 본 연구의 차별성은 두 가지로 요약할 수 있는데, 첫째는 양식수산물
단수 예측을 위해 기상 및 수질변수를 사용했다는 점임
- 현재 수산 분야의 생산량 예측 연구는 대부분 어획노력량, 시설량 등의
인위적 투입 요소나 비용, 가격 등의 경제적 요소를 변수로 활용하였음
- 본 연구에서는 수산물의 생산량을 결정짓는 요인 중 작황에 해당하는 단수
(단위생산량)에 집중하였으며, 단수에 직·간접적 영향을 미치는 기상 및
수질과 관련한 데이터를 독립변수로 활용하였음
- 둘째 특징은 조건을 다양한 조건의 인공신경망모형을 이용해 양식수
산물 단수 예측모형을 구축하고 비교하였다는 점임
- 기존의 수산 분야 연구는 계량경제 이론에 입각한 경제모형 및 생물경제모
형 등 회귀분석모형 기반의 예측 연구가 대부분이었음
- 본 연구에서는 양식수산물 단수 예측을 위해 세 가지 인공신경망 알고리즘
(MLP, LSTM, 양방향 LSTM)과 다양한 조건에서의 예측 결과를 비교하여
인공신경망모형의 유효성을 검토하였음

3. 연구 결과

1) 결과 요약

- 인공신경망모형 훈련 결과 지역별, 예측 시차별로 가장 예측력이 우수
한 6개 모형을 선별하여 다중회귀모형과 비교한 결과 3개 모형이 다중
회귀모형보다 우수한 예측력을 나타내었음
- 예측 시차는 단기(1개월)와 장기(5개월)로 설정하였음

- 변수 및 파라미터를 다르게 설정한 108개 모형을 훈련하였으며, 각 모형을 이용해 지역별, 예측 시차별 2021년산 김 단수를 추정하고 MSE가 가장 낮은 모형을 선택하였음
- 지역별, 예측 시차별 6개 모형을 동일한 변수를 사용한 다중회귀모형과 비교한 결과 절반인 3개 모형의 예측력이 다중회귀모형보다 우수했음

〈요약 표 2〉 지역별 2021년산 김 단수 추정모형 비교(1개월 추정모형)

지역	구분	모형	MSE	Theil's U
전남	단기	인공신경망	11.44	0.09
		다중회귀	14.21	0.10
	장기	인공신경망	11.15	0.09
		다중회귀	10.73	0.08
완도	단기	인공신경망	33.82	0.10
		다중회귀	97.11	0.16
	장기	인공신경망	49.82	0.12
		다중회귀	76.24	0.15
고흥	단기	인공신경망	37.81	0.09
		다중회귀	7.57	0.04
	장기	인공신경망	16.26	0.06
		다중회귀	8.12	0.04

주: 음영으로 표시된 인공신경망모형은 다중회귀모형에 비해 예측력 지표가 우수한 모형을 나타냄

- 일정 기간의 데이터를 바탕으로 어기 전체의 생산량을 예측하는 어기 추정 모형을 학습한 결과에서도 일부 모형의 예측력이 우수한 것으로 나타났음
- 어기 모형은 어기 전체 단수 예측을 위해 8, 9월 데이터를 이용하는 어기에측모형1과, 상반기(1월~6월) 데이터를 사용하는 어기에측모형2로 구분함
- 앞서 1개월 예측에서보다 불리한 조건(시차)의 데이터를 사용하였으나, 일부 모형에서 오히려 더 나은 결과를 도출하였음
- 8, 9월 데이터를 이용해 전남지역 김 단수를 예측한 모형이 〈요약 표 2〉의 전남지역 단기 다중회귀모형보다 낮은 MSE와 Theil's U 계수를 나타냄

- 또한 완도지역 예측모형의 경우 입력 데이터에 관계없이 1개월 예측 다중 회귀모형에 비해 예측력이 우수하게 나타났음

〈요약 표 3〉 지역별 2021년산 김 단수 추정모형 비교(어기 추정모형)

구분	모형	MSE	Theil's U
전남	어기예측모형1	13.44	0.09
	어기예측모형2	17.94	0.11
완도	어기예측모형1	15.95	0.08
	어기예측모형2	41.85	0.11
고흥	어기예측모형1	12.15	0.05
	어기예측모형2	14.27	0.06

2) 결론

- 인공신경망모형은 다중회귀모형과 동일한 조건의 변수를 사용한 6개 모형 중 3개 모형에서 예측력이 더 우수했으며, 더 불리한 조건의 변수를 사용한 경우에도 일부 우수한 예측력 지표를 나타냈음
- 인공신경망모형은 변수 사용의 자유도가 높고 다중회귀모형보다 복잡한 상관관계의 표현이 가능하므로 이 같은 결과가 도출되었다고 판단됨
- 이러한 장점으로 인해 향후 수산분야의 연구에서도 인공신경망의 활용성이 높아질 수 있을 것으로 기대됨
- 본 연구에는 다음과 같은 한계점이 존재함
 - 본 연구에 사용한 기상 및 수질 데이터는 도/군의 넓은 지자체 단위로 구분된 지역의 대푯값을 사용하였으나, 실제 개별 양식여장을 모두 대표한다고 단정할 수는 없으며 충분한 양의 시계열을 확보했다고 보기 어려움
 - 변수 선정 시 환경 요인에 집중하였기 때문에 실제 김 생산에 주요한 영향을 미치는 채묘, 관리 노동력 활용 등에 관한 변수는 활용되지 않았음

- 인공신경망모형에서는 개별 변수의 유의성을 정량적으로 확인할 수 없어 변수의 적합성이나 예측력을 저해한 요인을 판단하기 어려움
- 모형의 예측력에만 집중하여 독립변수와 종속변수의 명확한 상관관계에 대한 이론적, 실증적 근거에 대한 분석은 부족했음
- 따라서 양식수산물 작황 예측을 위한 실질적 활용을 위해서는 심화된 후속 연구를 통해 앞의 단점을 극복해야 할 것임

3) 제언

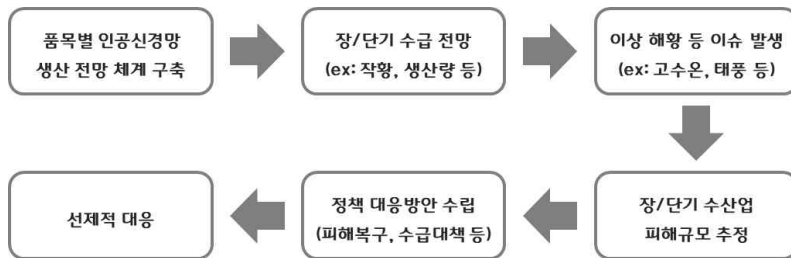
- 본 연구의 학술적 기여는 기존의 활용도가 낮았던 방법론과 자료를 활용했다는 점에서 찾을 수 있음
- 향후 진행될 연구에서는 데이터베이스 구축, 모형 설계, 평가에 이르는 과정에서 앞서 제시된 한계점을 보완해야 함

〈요약 표 4〉 연구 개선방안

DB 구축
- 양식어장의 위치에 따라 어장의 환경을 대표할 수 있는 기상 또는 수질 데이터 탐색 - 작황에 영향을 줄 수 있는 인위적 투입요소(시설 종류, 양성 밀도, 채묘 등) DB화 - 위성사진 등 비수치적 자료 활용
↓
모형설계
- 연구 목적에 맞는 인공신경망모형 설계 ▷ 정량 전망이 필요한 경우 - 회귀모형 ▷ 증감 예측이 필요한 경우 - 분류모형
↓
모형평가
- 변수 조합의 다양화 및 모델 해석 기법 등을 이용한 변수의 유의성 분석 - 시나리오 분석을 통한 변수의 영향력 추정 - MSE, MAPE, Theil's U, RMSPE 등 예측력 검정지표를 통한 평가
↓
실질적 운용방안 수립
- DB 구축 및 관리체계 구축 - 모형 관리 및 운용 매뉴얼 수립 - 정책 대응체계 구축

- 심화 연구를 통해 인공신경망모형 수급 전망체계를 구축한다면 수산 정책 의사결정을 위한 도구로 활용할 수 있음
- 태풍, 고수온과 같은 기상, 수질 변화에 따른 수산업 피해가 예상될 경우 인공신경망모형을 이용해 즉각 예상 피해 규모를 정량적으로 도출함으로써 향후 구체적 정책 의사결정에 기여할 수 있을 것으로 판단됨

〈요약 그림 1〉 정책적 활용방안



- 정밀한 예측모형 수립과 어장 환경 DB 활용도 제고를 위해서는 현재 보다 정밀한 범위에서 장기적인 관측이 이루어져야 함
- 현재 국립수산물과학원 등에서 제공하는 어장 환경 데이터는 특정 지점에 한하여 관측이 이루어지고 있으며 연속성이 떨어지는 자료가 많음
- 세분화된 어장 환경 DB가 구축되고 장기적으로 유지·관리된다면 현재보다 정밀한 수준의 예측이 가능할 것이며, 관련 연구에서의 활용도도 향상될 것으로 판단됨

EXECUTIVE SUMMARY

1. Purpose

- This study constructed a quantitative model for forecasting unit yield reflecting the environment (weather and water quality) of aquaculture farms by utilizing an artificial neural network model.
- As an indicator representing the productivity and yield of aquacultured seafood, unit yield is both directly and indirectly affected by a change in weather and water quality of fishing grounds.
- Forecasting a quantitative production volume is necessary when policy decisions should be made due to changes in the production of farmed seafood. However, studies for forecasting the production of farmed seafood utilizing environmental conditions (weather, water quality) of fishing grounds have been insufficient.
- Therefore, this study utilized an artificial neural network model for forecasting the unit yield of farmed seafood in an effort to

apply environmental variables and improve the predictability.

- After comparing the results of the analysis carried out by an artificial neural network model and a multiple regression model with a variety of conditions, this study aims to present a number of policy-wise application measures.
- This study estimated MLP, LSTM, bidirectional LSTM model as well as a variety of models with different settings of input variables and detailed parameters, and then compared the result with that of a multiple regression model.
- Having concluded the result based on the comparison, this study intends to extract future tasks for policy-wise application measures as well as its practical utilization.

2. Methodology and Features

1) Methodology

- This study established a forecasting model for the unit yield of farmed seaweed by region based on an artificial neural network model, reflecting weather and water quality data.
- A total of 216 models were learned and then compared the result for forecasting the unit yield of seaweed based on an artificial neural network model depending on variables such as region, forecast horizon, classification of dependent variables (weather/ water quality) and type of algorithms (MLP, LSTM and bidirectional LSTM).

- For an indicator of assessing the predictability of the model, MSE, MAPE and Theil's U coefficients were used.
- A multiple regression model was selected as a comparison model for evaluating the artificial neural network model.
- In addition to indicators for assessing the predictability of the model, this study selected a multiple regression model as a means to evaluate the established artificial neural network model.
- For comparison purposes, this study extracted models respectively by setting variables with equal conditions as dependent variables as well as by selecting variables with a high statistical significance.

〈Summary table 1〉 Analysis Methods

Establishing an analysis DB	• Weather variables (precipitation, wind speed, amount of sunshine) • Water quality variables (water temperature, DO, DIN, DIP etc.)
Building an artificial neural network model	• MLP, LSTM, Bidirectional LSTM • Diversifying variables and parameter setting
Constructing a multiple regression model (comparison model)	• Analyzing models with equal conditions as well as models with optimal conditions for assessing the artificial neural network model
Comparing and assessing models	• Comparing the results of forecasts based on an artificial neural network model and a multiple regression model • Comparing indicators for assessing the predictability
Presenting conclusion	• Extracting conclusion based on analysis results

2) Features

- This study differentiates itself from preceding studies for the two following reasons. First, variables such as weather and water quality were utilized for forecasting the unit yield of farmed seafood.
- Existing studies for forecasting the production in the fishery sector have used either artificial input factors such as fishing effort and the number of facilities or economic factors including cost and price as variables.
- This study, however, concentrates on the unit yield (production per unit) which represents harvest among the factors that determine the production volume of fishery products. Following this, the study utilized dependent variables such as weather and water quality-related data which pose direct and indirect effects on unit yield.
- The second distinctive feature of this study was to develop a forecasting model for the unit yield of farmed seafood by utilizing an artificial neural network model with a variety of conditions.
- Most of the existing studies in the fishery sector have included a forecasting study based on regression analysis based on an economic model grounded on econometric theories and a bio-economic model.
- However, this study compared the results of forecasts carried out with three artificial neural network algorithms (MLP, LSTM,

bidirectional LSTM) as well as with a variety of conditions, verifying the effectiveness of the artificial neural network model.

3. Results

1) Summary

- Based on the result of training an artificial neural network model, this study selected 6 models that exhibit excellent predictability by region and by forecast horizon. Following this process, this study compared the 6 selected models with a multiple regression model, resulting in 3 models presenting more superb predictability than the multiple regression models.
- Forecast horizon was established as short-term (1 month) and long-term (5 months).
- This study conducted training for 108 models with different settings of variables and parameters. While estimating the unit yield of seaweed produced in 2021 by region and forecast horizon, this study selected the model possessing the lowest MSE.
- This study compared 6 artificial neural network models by region as well as by forecast horizon with multiple regression models which use the same variables. The result showed that half of the 6 models presented superior predictability to that of multiple regression models.

〈Summary table 2〉 Comparison of models for estimating the unit yield of seaweed produced in 2021 by region (1month estimation model)

Region	Category	Model	MSE	Theil's U
Jaolla-namdo	Short-term	Artificial neural network	11.44	0.09
		Multiple regression	14.21	0.10
	Long-term	Artificial neural network	11.15	0.09
		Multiple regression	10.73	0.08
Wando	Short-term	Artificial neural network	33.82	0.10
		Multiple regression	97.11	0.16
	Long-term	Artificial neural network	49.82	0.12
		Multiple regression	76.24	0.15
Goheung	Short-term	Artificial neural network	37.81	0.09
		Multiple regression	7.57	0.04
	Long-term	Artificial neural network	16.26	0.06
		Multiple regression	8.12	0.04

Footnote: Artificial neural networks colored green present models that showed superior predictability than that of multiple regression models.

- This study also includes learning of a fishing season estimation model which forecasts the production of the whole fishing season based on a certain period of data. The result showed that some of the artificial neural network models showed superb predictability.
- In order to forecast the unit yield of the whole fishing season, this study classifies fishing seasons models into two; Fishing Season Forecasting Model 1 which utilizes August and September data and Fishing Season Forecasting Model 2 which uses the first half of the year data (January ~ June).
- The data used in this testing contained more disadvantageous conditions than those previously used in 1-month forecasts. However, some models produced rather a better predictability.
- The forecasting model that predicted the unit yield of seaweed

in Jeonnam province with August and September data showed lower MSE and Theil's U coefficients than a short-term multiple regression model for Jeonnam province as illustrated in <Summary table 2>.

- Moreover, The forecasting model for Wando showed an excellent predictability than 1-month predicting multiple regression models regardless of input data.

<Summary table 3> Comparison of estimation models for the unit yield of seaweed produced in 2021 by region (Fishing Season Estimation Model)

Category	Model	MSE	Theil's U
Jaolla-namdo	Fishing season forecasting model1	13.44	0.09
	Fishing season forecasting model2	17.94	0.11
Wando	Fishing season forecasting model1	15.95	0.08
	Fishing season forecasting model2	41.85	0.11
Goheung	Fishing season forecasting model1	12.15	0.05
	Fishing season forecasting model2	14.27	0.06

2) Conclusion

- Among 6 artificial neural network models that use the variables with the same conditions as the multiple regression models, 3 artificial neural network models showed better predictability. Even with variables containing more disadvantageous conditions, some models indicated excellent predictability.
- Since degrees of freedom in artificial neural network models are relatively high in terms of using variables as well as the

models are capable of demonstrating a more complex correlation than multiple regression models, these results can be extracted.

- Due to these advantages, the applicability of artificial neural network models is expected to increase in further fisheries-related studies.
- This study, however, possesses limitations as follows;
 - Weather and water quality data used in this study came from representative values of a region classified as a unit of broad local government such as Do/Gun. However, that doesn't mean that these data can represent individual aquaculture farms.
 - While concentrating on environmental factors in selecting variables, this study did not utilize variables that affect the production of seaweed in reality such as seedling and use of managerial labor.
 - Since the statistical significance of individual variables cannot be identified in an artificial neural network model, it is difficult to evaluate the suitability of variables or find factors that deteriorate the predictability.
 - The fact that mathematical interpretation for parameters and weighted values is difficult can be another weakness of an artificial neural network model.
 - Therefore, a number of in-depth follow-up studies should be carried out to overcome the shortcomings mentioned earlier, as well as practically utilize the model for forecasting the yield of farmed seafood.

3) Suggestions

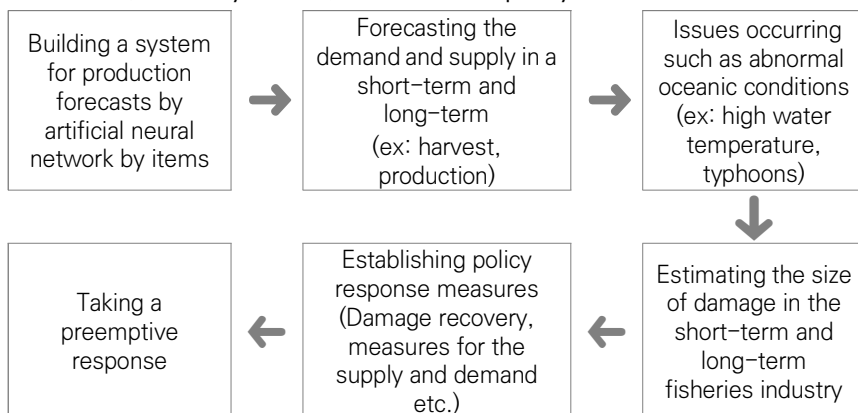
- The academic contribution of this study can be found in utilization of methodologies and data that have not been used so far.
- Follow-up studies should overcome the limitations presented earlier in the overall stages from the establishment of a database, model design to assessment.

〈Summary table 4〉 Measures to implement in-depth follow-up studies by sector

DB Establishment
• Searching weather and water quality-related data which can represent the environment of fishing grounds depending on the location of aquaculture farms • Building a database for artificial input factors that can affect the yield (type of facilities, density of cultivation and seedling) • Utilizing non-numeric data such as satellite images etc.
↓
Model design
• Combining a variety of artificial neural network algorithms befitting to data type (ex: time series data, LSTM, Image data – CNN) • Searching optimal parameter setting such as hidden layer, perceptron and optimizer • Design an artificial neural network model befitting to the purpose of the study ▷ When forecasting a fixed quantity is required – regression models ▷ When forecasting an increase and(or) decrease is required – classification models
↓
Model evaluation
• Extracting the significance of individual variables by diversifying the combination of variables • Estimating the impact of variables through scenario analysis • Assessing the predictability with testing indicators such as MSE, MAPE, Theil's U and RMSPE
↓
Establishment of practical operation measures
• Building a DB as well as a management system • Establishing a manual for model management and operation • Setting up a policy response system

- If in-depth studies contribute to the establishment of an artificial neural network model for forecasting the supply and demand, it can be utilized as an effective tool for the decision making process of fisheries policies.
- Especially when changes of water quality and weather including typhoons and high water temperature are expected to damage the fishery industry, such a system will be able to immediately calculate the scale of estimated damage utilizing an artificial neural network model, further contributing to a specific policy-making process.

〈Summary table 1〉 Measures for policy-wise utilization



- In order to establish a more precise forecasting model and improve the utilization of DB regarding the environment of fishing grounds, it is necessary to conduct long-term monitoring in a narrower scope.

- The data surrounding the environment of fishing grounds offered by the National Institute of Fisheries Science and others only provide observatory data after monitoring specific sites, while falling short of consistency.
- If a more segmented database on the environment of fishing grounds is established and regularly maintained, a more precise level of forecasting will be available and enhance the utilization in relevant studies.

01

서론

제1절 연구 배경 및 목적

1. 연구 배경 및 필요성

수산업에서는 인간이 통제할 수 없는 자연적 요인으로 인해 생산의 불확실성이 크게 나타난다. 어업은 형태에 따라 자연에 서식하는 수산 동식물을 채취하거나 포획하는 어로어업과, 일정한 공간에서 수산생물을 번식 및 육성시켜 수확하는 양식어업으로 구분할 수 있다.¹⁾ 양식어업은 수산물의 양성 단계에서부터 인위적인 요소를 투입해 관리하므로 어로어업보다는 장기적인 계획에 따라 안정적인 생산이 가능하다. 하지만 해수의 성분, 영양염류, 수온 등 양식어업이 성립되기 위한 전제조건이 되는 수중 환경은 사실상 인간의 노력으로는 통제 불가능하며, 양식어장의 환경은 어장의 지리적 위치, 해수의 흐름, 수질, 기상 조건 등에 영향을 받는다.

기상 조건과 수질은 불규칙적 변동이 잦으며 이로 인한 해양 환경의 변화는 양식어업에 일시적이거나 지속적인 영향을 미친다. 예를 들어 멧게

1) 조영제 외(2012), p. 109, 110.

양식의 경우 2018년 하절기 고수온 피해로 익년도 출하 예정이던 2년산 명게가 대량 폐사하면서 2019년 경남지역의 명게 생산량이 40% 이상 감소하기도 했다. 또한, 지난 7월 전남에서 집중 호우가 발생하며 대량의 민물이 바다로 유입되었고, 이로 인해 강진지역 32개 전복 해사가두리 양식장에서 2,000만 마리 이상의 전복이 폐사하는 등의 피해가 발생했다.²⁾ 양식 전복과 명게 모두 2년 이상의 양성 기간이 소요되기 때문에, 이러한 기상 변화에 따른 피해는 장기간에 걸쳐 수급에 영향을 미치게 된다.

이처럼 양식어장 일대의 기상 및 환경 조건은 수산물 생산에 상당한 수준의 영향을 미치므로, 일정 수준 이상의 변화가 관찰될 경우 미래의 생산에도 영향을 미치리라는 것을 어렵지 않게 추론할 수 있다. 우리나라 양식산업의 규모와 중요성은 계속해서 커지고 있음에도, 기후 변화와 이상 해황 등으로 인해 생산의 불안정성은 심화되는 실정이다. 이로 인해 기상 및 해황에 따른 수급 불안정 문제에 대한 정책적인 대안 마련의 필요성이 높아지고 있다.

앞서 서술한 집중 호우에 의한 전복 폐사 사례와 같이, 급격한 폐사 발생으로 피해 규모에 대한 정량적인 측정이 이루어졌을 때에는 정책적 의사결정에 있어 구체적·정량적 대책 수립이 가능하다. 그러나 일시적인 환경의 변화라고 할지라도 오랜 기간을 두고 서서히 영향을 미치는 경우, 그것이 수산물 생산에 어느 정도의 변동을 초래하는지 명확한 상관관계를 정량적으로 표현하기는 어렵다. 이와 같은 이유로 정량적인 전망이 수반되지 않는다면 특정 사안에 대한 정책적인 접근에는 구체성이 떨어질 수밖에 없다.

양식 어장의 환경 변화에 의한 생산 변동을 정량적으로 예측하기 위해서는 기상이나 수질 등의 요인을 모형에 반영해야 한다. 양식수산물의 생산 전망과 관련한 연구는 활발히 이루어지고 있지만, 대부분 경제적인 요소를

2) 연합뉴스(2021.7.12)

고려하고 있으며 환경과 관련한 변수의 사용은 제한적으로 이루어지고 있다. 이는 대부분 계량경제학 이론에 근거한 방법론이 사용되었기 때문인데, 수온, 풍속 등 환경변수가 생산에 미치는 영향이 불명확한 경우가 많으며, 환경변수를 내생화 할 수 없어 외생변수로 반영하는 경우가 많다. 따라서 계량경제모형의 제약을 벗어나 인공신경망모형을 포함한 다양한 연구 방법론의 적용이 필요하다. 해외 사례와 국내 농업 분야의 경우 인공신경망 이론 기반의 연구 비중이 커지고 있으며, 유의미한 결과를 도출하였다. 수산 분야 역시 수급 전망에 있어 중요한 요소인 환경변수를 사용하고, 이를 위해 기존의 틀을 벗어난 예측 방법론의 적용이 필요한 시점이라고 판단된다.

우리나라의 주요 양식품종 중 김은 생산량과 금액이 가장 크다. 생산 규모가 계속해서 확대되고 있으며, 최근에는 전 세계적으로 소비가 늘면서 우리나라 김의 수출 또한 빠른 속도로 증가하는 등 산업적 중요성이 커지고 있다. 그러나 기후변화 등의 요인으로 황백화와 갯병 등이 발생함에 따라 생산의 불안정성은 여전히 크게 나타나고 있다. 따라서 본 연구에서는 양식수산물 중 생산 규모가 가장 크고 환경 요인에 따른 불안정성이 크게 나타나는 김을 실증분석 대상으로 하여 연구를 수행하였다.

2. 연구 목적

본 연구의 목적은 양식수산물 단수 전망에 있어 환경변수의 활용도를 높이고 예측력을 제고하기 위한 인공신경망모형의 유효성 검토와 활용방안 제시이다. 이를 위한 목표는 크게 세 가지로 구분할 수 있다.

우선 첫째는 분석 대상 어업과 데이터의 특징에 부합하는 인공신경망모형의 설계와 구축이다. 즉 분석 대상인 김 양식업의 수급 동향과 양식 방법 및 특징에 대해서 살펴보고, 분석자료에 적합한 형태의 인공신경망모형을

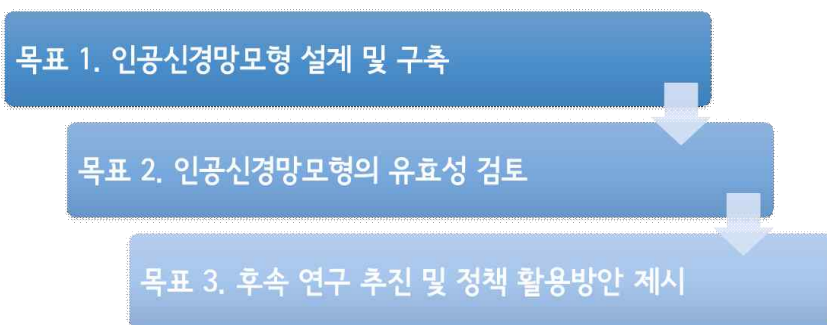
구축하는 것이다.

둘째는 실증분석을 통해 인공신경망모형의 유효성을 검토하는 것이다. 현재까지 인공신경망모형의 경우 다양한 분야에서 우수한 예측력을 나타냈으나, 수산업 분야 연구에서는 활용 사례가 드물고 전통적인 계량경제 분석 방법론과 같이 통계적 유의성 검증체계가 두텁지 않다. 따라서 실증분석에서는 인공신경망모형의 예측력 검정과 더불어 다중회귀모형과의 비교를 통해 유효성을 검토하고자 하였다.

셋째는 후속 연구의 추진방안 및 정책 활용방안 제시이다. 다시 말해 실증분석 결과에 따라 연구의 한계점과 시사점을 도출하고, 이를 바탕으로 한계점 극복 및 실질적 활용을 위한 후속 연구 추진방안과 정책적 활용방안을 제안하고자 하였다.

〈그림 1-1〉 연구 목적 및 목표

양식수산물 단수 전망을 위한 인공신경망모형의 유효성 검토 및 활용방안 제시



자료: 저자 작성

제2절 연구 내용 및 방법

1. 연구 내용

본 연구는 양식수산물 단수 전망을 위한 인공신경망모형의 유효성을 검토하는 것을 목적으로 한다. 이를 위해 통계학 및 계량경제 이론에 기초한 다중회귀모형과 인공신경망모형을 살펴보고자 한다. 특히 환경변수 반영에 따른 모형의 예측 성능 향상 여부와 각 모형의 장단점을 이론적·실증적으로 연구하고자 하였다.

김은 우리나라 양식수산물 중 생산량 및 생산금액이 가장 많으며, 최근에는 세계적으로 김 소비가 늘어나면서 우리나라 김 수출도 증가하는 추세에 있어 산업적 중요성이 나날이 커지고 있는 품목이다. 하지만 다른 수산물과 마찬가지로 어장 환경의 영향을 많이 받으므로 고수온, 갯병, 황백화 등 다양한 요인으로 인해 생산량 변동 폭이 커 생산의 불안정성은 해소되지 않고 있다. 본 연구에서는 산업적 중요성과 더불어 생산 불안정성이 뚜렷하게 나타나는 김을 중심으로 실증분석을 실시하였다.

본 연구의 구성은 총 6장으로, 제1장에서는 연구 배경 및 목적, 연구 방법 및 주요 내용, 선행연구를 검토하였다. 선행연구 검토에서는 예측모형과 관련한 국내외의 다양한 연구를 살펴보았으며, 주요 연구 결과 및 최근 연구 동향 등을 바탕으로 연구의 필요성과 연구 설계에 관한 내용을 서술하였다. 다음으로 제2장에서는 본 연구의 실증분석 대상인 김 양식의 특징을 알아보았다. 먼저 시기별·지역별 김 수급 동향과 주요 수급 이슈에 대해 조사하였다. 또한 주요 김 양식 방법과 지역별 양식 환경, 주요 환경 조건에 대해 알아보으로써 전망모형 설계를 위해 고려해야 할 요소들을 살펴보았다.

제3장에서는 본 연구의 실증분석 방법론을 소개하였다. 인공신경망모형의 이론적 배경 및 주요 인공신경망모형의 특징과 장단점 등을 알아보았다. 제4장에서는 양식수산물 단수 전망모형 설계에 대한 내용을 다루었다. 우선 제2장에서 살펴본 김양식의 특징과 환경 조건 등을 고려하여 구축한 생산 및 환경변수 등의 데이터베이스(자료)를 설명하고, 제3장에서 소개한 방법론을 기반으로 하여 지역별 김 단수 전망을 위한 계량경제모형 및 인공신경망모형을 설계하였다.

제5장에서는 제4장에서 설계한 모형을 이용해 인공신경망모형과 다중회귀모형의 김 단수 전망 추정 결과를 비교하였다. 각 모형의 객관적 비교를 위해 예측치와 관측치를 이용한 예측력 검정 수단과 함께 모형의 해석 및 전망, 시뮬레이션 등을 수행하여 각각의 장단점을 고찰하였다. 마지막으로 제6장에서는 본문의 요약과 함께 연구의 결론을 도출하였다. 연구 결과를 바탕으로 후속 연구의 추진방안과 정책적 활용방안을 도출하였다.

〈표 1-1〉 주요 연구 내용

순서	내용
제1장	서론
	- 연구 배경 및 목적, 내용 및 방법, 선행연구 검토
제2장	김 수급 동향 및 양식 현황
	- 수급 동향, 양식 방법 및 양식 환경
제3장	인공신경망모형의 이론적 고찰
	- 인공신경망모형의 개요, 인공신경망 학습 방법
제4장	예측모형 설계
	- 분석 자료 및 기초 분석, 예측모형 및 학습조건 설계
제5장	예측모형 추정 결과
	- 인공신경망모형 및 다중회귀모형 추정 결과
제6장	결론 및 제언
	- 요약 및 결론, 향후 과제 및 제언

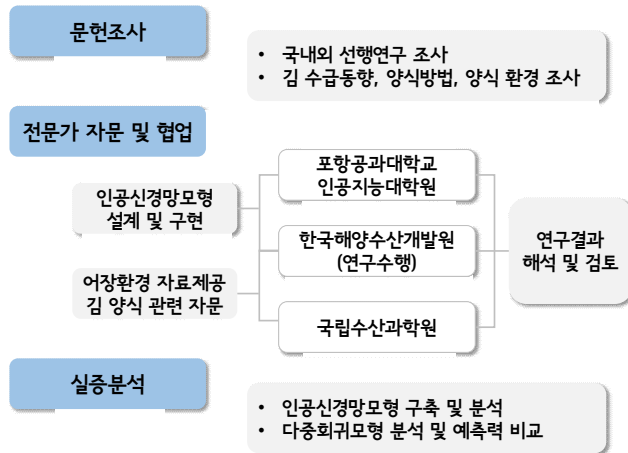
자료: 저자 작성

2. 연구 방법

1) 연구 수행 체계

본 연구의 추진에는 문헌 및 자료 조사, 전문가 자문 및 협업, 실증분석 등의 방법을 사용하였다. 문헌조사로는 연구 설계를 위해 선행연구를 살펴 보았고, 실증분석에 앞서 본 연구의 분석 대상인 김 수급 동향과 양식 방법, 지역별 양식 환경 등에 관한 현황을 알아보았다. 국립수산물학원의 김 양식 매뉴얼을 참고하여 김 양식 방법 및 환경 등을 살펴보았고, 동 기관의 어장 환경 모니터링 자료 등을 바탕으로 분석 데이터를 구축하였다.

〈그림 1-2〉 연구 수행 체계



자료: 저자 작성

다음으로 전문가 자문 및 협업을 통해 실증분석모형을 설계하고 본 연구의 실증분석 목적과 데이터의 특징에 부합하는 인공신경망모형을 구현하였다. 인공신경망모형의 구현에는 프로그래밍 언어(파이썬)를 이용하였다. 또

한 구현된 모형을 이용한 실증분석의 세부 내용과 더불어 연구 결과 검토를 위해 전문가 자문을 실시하였다.

실증분석에는 전문가 자문을 통해 구현한 인공신경망모형을 이용하였다. 본 연구의 목표 중 하나는 인공신경망모형의 유효성을 검토하는 것이다. 이를 위해 인공신경망모형의 예측치와 관측치의 오차를 이용하는 예측력 점정지표와 더불어 다중회귀모형과의 비교를 실시하였다.

2) 실증분석 방법: 인공신경망모형

계량경제모형의 경우 기본적으로 경제학적 이론에 근거하고 있으며, 현재까지의 수산 분야 생산 전망과 관련한 연구는 대부분 투입요소(양식 면적, 시설량 등)와 시장요소(가격, 생산량 등) 데이터를 이용해 전망치를 도출한다. 어장환경의 변화는 양식수산물의 작황에 직접적인 영향을 미치지만, 변수의 유의성이나 모형의 적합도 문제 등으로 인해 계량경제모형에 환경 변수를 반영하여 유의미한 결과를 도출하기는 어렵다.

본 연구에서는 위와 같은 한계점을 인식하여 대안으로서 인공신경망모형을 이용하고자 하였다. 인공신경망모형은 변수, 레이어, 퍼셉트론, 활성화 함수 등의 요소로 복잡하고 비선형적인 관계를 표현하며 기계학습을 통해 최적의 해를 찾는 방법이다. 계량경제모형에 비해 변수 사용에 있어 이론적 제약은 적고 예측성능에 집중한 방법으로 다양한 문제에 활용할 수 있다.

최근에는 산업을 막론하고 다양한 분야에서 각광을 받으며 활용 분야가 넓어지고 있지만, 수산 분야에서는 여전히 계량경제이론 기반의 연구가 주류를 차지하고 있으며 인공신경망모형을 이용한 연구는 활발하지 않다. 따라서 본 연구에서는 양식수산물 단수 전망에 있어 환경변수의 활용도와 예측력 제고를 위해 인공신경망모형을 이용했다.

제3절 선행연구 검토

1. 국내 선행연구

연구 설계를 위해 본 연구의 주제와 관련한 국내외 농수산 분야 선행연구를 검토하였다. 우선 우리나라 주요 양식수산물에 대한 생물학적 연구를 살펴보았다. 박민우 외(2013)는 특정 수온, 염분 조건에서 전복의 사망 및 탈락률, 전복 발의 조직학적 변화를 실험하였고, 김철원·오현주·신윤경(2013)는 환경이 다른 가막만 4개 해역의 수질에 따른 굴 생존율과 비만도 원인을 분석하였다.

〈표 1-2〉 국내 선행연구(자연과학 분야)

연구 주제(제목)	분석 방법	주요 내용
• 수온 및 염분 농도에 따른 북방전복의 생존율, 탈락률 및 발의 조직학적 변화(박민우 외, 2013)	실험을 통한 전복의 변화 관찰	- 특정 수온, 염분 조건에서 전복의 사망 및 탈락률, 발의 조직학적 변화 실험
• 가막만 양식 굴, <i>Crassostrea gigas</i> 폐사에 영향을 끼치는 수온의 영향(김철원·오현주·신윤경, 2013)	조사 해역별 수질환경 분석 및 양식 굴 성장 조사	- 물리 환경이 다른 가막만 4개 해역의 수질 분석 및 해역별 양식 굴 생존율, 비만도 원인 분석
• 양식 넙치의 생리조건에 미치는 연속적인 수온 급강하의 영향(장영진 외, 1999)	수온 급강하에 따른 스트레스 반응 실험	- 냉수대로 인한 양식넙치의 생리 반응 분석
• 양식김의 광합성에 미치는 수온, 염분 및 부니의 영향(장선덕·진평·박기영, 1983)	수온, 염분, 부니에 따른 양식 김 광합성 반응 실험	- 수온, 염분, 부니 농도 조건별 양식 김 광합성 반응 분석
• 고창 갯벌에서 중국산 및 총남산 바지락의 성장, 비만도 및 폐사 특성 비교(안현미 외, 2016)	바지락 종패의 갯벌 이식 시험 양식 실험	- 양식장 환경요인에 따른 품종별(중국산, 한국산) 바지락 폐사율, 성장 및 비만도 통계 분석
• 온도자극 및 수용밀도에 따른 우렁쉥이 폐사(홍정표·김영석·허성범, 2000)	실내 및 야외 실험	- 실내 수용밀도 및 온도 자극, 야외 시험어장 수심별 수질 관측 및 폐사율 조사 실험

자료:저자 작성

장영진 외(1999)는 수온 급강하 실험을 통해 냉수대로 인한 양식 넙치의 생리 반응을 분석하였으며, 장선덕·진평·박기영(1983)의 연구에서는 수온, 염분, 부니 농도에 따른 양식 김의 광합성 반응을 분석했다. 자연과학적 연구에서는 샘플을 채집하거나 조건을 통제하는 실험을 통해 환경이 수산생물 개체에 어떠한 영향을 미치는지를 연구한다. 따라서 이러한 선행연구를 통해 특정 품종의 생물학적인 적정 환경을 추정할 수 있다. 그러나 상기의 연구에서는 실제 양식어장 전체의 생산성보다는 수산생물 개체의 변화 관찰에 집중하였으며, 이를 통해 어장의 생산성 증감을 구체적으로 추정하지는 않았다.

국내 농업분야 연구 중 박성호(2020)는 식용작물 생산량 예측을 위해 회귀분석, 의사결정 트리이론 기반의 머신러닝모형과 DNN, RNN, LSTM과 같은 딥러닝모형을 비교하였다. 또한 식용작물별 최적의 머신러닝과 딥러닝 모델을 소개하였다.

〈표 1-3〉 국내 선행연구(농업 분야)

연구 주제(제목)	분석 방법	주요 내용
• 머신러닝 기반 식용작물별 생산량 예측 연구(박성호, 2020)	회귀분석, 트리 기반 머신러닝, 딥러닝 모델 성능 비교	- 통계청 및 기상청 공공데이터 활용 식용작물 생산량 예측모형 설계 - 예측모형별 성능 비교
• 인공 신경망을 이용한 채소 단위 예측 모형 개발: 고추를 중심으로(이춘수·양성범, 2017b)	인공신경망, 패널모형, 적응기대모형	- 기상변수를 이용한 인공신경망 고추 단위 예측모형 제시 - 적응 기대모형 및 패널모형과 비교 - 인공신경망모형의 예측력이 우수
• 역전파 신경망 모델을 이용한 기준 작물 증발산량 산정(김민영 외, 2019)	인공신경망, 다중 선형회귀모형	- 역전파 신경망 모형을 이용한 기준 작물 증발산량 추정 - 다중 선형회귀모형과 성능 비교 - 인공신경망모형의 예측력이 우수
• 기상 및 토양정보가 고령지 배추 단위예측에 미치는 영향(권태웅·윤상후·김래용, 2019)	회귀분석 및 랜덤포레스트 모형	- 고령지 배추 주산지 5개 지점의 토양 정보를 이용한 예측모형 구축 - 회귀모형 및 랜덤 포레스트 모형의 중요 설명변수 및 예측력 비교

연구 주제(제목)	분석 방법	주요 내용
• 생육정보를 이용한 가을배추와 가을무 단수 예측 모형 개발(이춘수양성범, 2017a)	pooled OLS	- 농작물 생육조사 현황 분석 - 생육조사 정보, 기상요인, 주산지 패널데이터를 이용한 pooled OLS모형 구축 - 생육정보 및 기상변수 시나리오 분석
• MODIS NDVI와 기상요인을 고려한 마늘·양파 주산지 단수예측모형 개발(나상일 외, 2017)	다중선형회귀모형	- 작물 생육지수 및 기상요인과 단수와의 상관관계 분석 - 마늘, 양파 시군별 단수 예측모형 분석
• 쌀 단수 예측모형 개발연구(한석호, 2016)	비선형 확률모형	- 패널자료의 고정효과모형분석을 통한 기상변수 계수 추정 - 기상 시나리오 분석을 통한 단수 예측

자료:저자 작성

이춘수양성범(2017b)은 기상변수를 이용한 고추 단수 예측을 위해 인공신경망모형을 이용했다. 또한 모형의 예측력을 적응기대모형 및 패널모형과 비교한 결과 인공신경망모형의 예측력이 더 우수한 것으로 나타났다. 김민영 외(2019)는 인공신경망모형을 이용해 기준 작물 증발산량을 추정하였으며, 다중 선형회귀모형과 비교한 결과 인공신경망모형이 다중 선형회귀모형보다 뛰어난 예측력을 보였다. 권태용·윤상후·김래용(2019), 이춘수양성범(2017a), 나상일 외(2017), 한석호(2016)의 연구에서는 농산물의 단수 예측을 위해 기상 및 토양 등의 환경 요인을 활용하였다. 농업 분야의 연구에서는 인공신경망모형의 활용도가 높아지고 있었으며, 경제적 변수와 더불어 환경변수를 반영하려는 시도가 활발했다.

다음으로 국내 수산업 분야의 선행연구를 살펴보았다. 허수진 외(2020)와 이현동·허수진·하현정(2020)의 연구에서는 구조방정식 모형을 이용해 단기 및 중기 김 양식 산업 수급 전망모형을 구축하였다. 이현동 외(2018)의 연구에서도 마찬가지로 구조방정식 모형을 이용하여 우리나라의 양식어업 전망모형을 구축하였다.

황강석·최정화·오택윤(2012)는 고등어 위판가격 변동 예측을 위해 인공신경망 예측모형을 설계하고 인공신경망모형의 활용 가능성을 제시했다. 최종두·최영준(2014)는 기후변화에 따른 굴 생산량 감소를 가정하여 굴 양식업의 경제성을 추정하기 위해 순현재가치, 비용편익 시나리오 분석을 실시하였다. 옥영수·김상태·고봉현(2007)는 양식넙치의 가격 예측을 위해 ARIMA, VAR 모형을 이용하였으며, 조용호(2006)의 연구에서도 ARIMA, VAR, SARIMA 등의 계량경제모형을 이용해 수산물 생산량 예측모형을 설계하였다. 이처럼 국내 수산 분야의 연구에서는 농업 분야보다 인공신경망모형의 활용도가 낮았으며, 생산성 예측에 환경변수를 반영한 사례도 드물었다.

〈표 1-4〉 국내 선행연구(수산업 분야)

연구 주제(제목)	분석 방법	주요 내용
• 김 중기 수급전망모형 Gim-MFoS 구축 연구(허수진 외, 2020)	다중회귀, 패널분석, 구조방정식형	- 김 중기 수급 전망모형 구축 및 정책 시뮬레이션
• 김 수급 전망모형 구축에 관한 연구(이현동·허수진·하현정, 2020)	다중회귀, 구조방정식모형	- 수산업관측센터 김 수급 전망모형 문제점 분석 및 신규 모형 구축
• 수산 부문 전망모형 KMI-FOSIM 구축 연구(이현동 외, 2018)	다중회귀, 구조방정식모형	- 양식어업 전망모형 구축, 양식어업 전망모형 추정 및 전망
• 기후변화로 인한 수온상승이 굴양식 본양성 생산방식의 경제성에 미치는 영향분석(최종두·최영준, 2014)	순현재가치, 비용편익 시나리오 분석	- 기후변화에 따른 굴 생산량 감소를 가정한 굴 양식 경제성 분석
• 인공신경망을 활용한 고등어의 위판가격 변동 예측(황강석·최정화·오택윤, 2012)	인공신경망모형	- 인공신경망모형을 통한 부산 공동 어시장의 위판단가 예측
• 양식넙치의 가격변동 및 예측에 관한 연구(옥영수·김상태·고봉현, 2007)	ARIMA, VAR	- 넙치 가격전망 시계열 분석
• Forecasting Total Marine Production through Multiple Time Series Model(조용호, 2006)	ARIMA, VAR, SARIMA	- 생산량 전망을 위한 시계열 예측모형 설계 및 전망

자료:저자 작성

2. 해외 선행연구

해외 선행연구 중 농수산 분야의 생산 예측모형에 관한 연구는 다음과 같다. Poorbagher *et al.*(2021)은 이란의 어로어업 및 양식어업 생산량 예측을 위해 ARIMA 모형을 이용했으며, Folorunso *et al.*(2019)은 탱크 순환여과 양식시스템의 용존산소량 예측을 위해 인공신경망모형과 다중회귀모형을 이용했다. Free *et al.*(2019)의 연구에서는 해양생물의 자원 평가를 위해 잉여생산량모형을 이용했으며, 해수면 온도 데이터를 활용했다. Torkashvand, Ahmadi, and Nikraves(2017)는 이란 Astaneh Ashrafieh 지역 키위 경도와 영양성분의 관계 분석을 위해 인공신경망모형을 이용했으며, 결과를 다중회귀모형과 비교하였다. 이 외에도 다수의 연구에서 ARIMA와 같은 시계열 분석모형과 다중회귀모형 등의 계량경제모형과 더불어 인공신경망모형을 이용했다. 또한 인공신경망모형을 이용한 연구에서는 모형 자체의 예측력 검정과 더불어 전통적인 계량경제모형의 결과와 비교한 경우가 많았다.

〈표 1-5〉 해외 선행연구

연구 주제(제목)	분석 방법	주요 내용
• Modelling the time series of capture fishery and aquaculture production in Iran(Poorbagher <i>et al.</i> , 2021)	ARIMA	- 이란의 어로어업 및 양식어업 생산량 시계열 분석
• Modelling the enviromental footprint of sea bram cage aquaculture in relation to spatial stocking design(Melita <i>et al.</i> , 2020)	모델링 및 시뮬레이션	- 돔 가두리 양식에서 성장, 유기 탄소 배출, 확산 및 퇴적 모델링 - 환경적 영향을 최소화하는 의사결정 및 생산 최적화를 위한 시뮬레이션
• Water quality index estimation model for aquaculture system using artificial neural network(Folorunso <i>et al.</i> , 2019)	인공신경망모형	- 인공 신경망 모형을 이용한 탱크순환여과(tank-cultured recirculatory) 양식 시스템의 용존산소량(f(수온, 알칼리도, pH, 전도)) 예측

연구 주제(제목)	분석 방법	주요 내용
• Impacts of historical warming on marine fisheries production(Free <i>et al.</i> , 2019)	수온 연계 잉여생산량모형	- 235종 해양생물의 자원평가 DB 및 해수면 온도 데이터를 이용한 잉여생산량 모형 분석
• Prediction of kiwifruit firmness using fruit mineral nutrient concentration by artificial neural network(ANN) and multiple linear regressions (MLR)(Torkashvand, Ahmadi, and Nikraves, 2017)	인공신경망, 다중회귀모형	- 이란 Astaneh Ashrafieh 지역 키 위 경도와 영양성분의 관계 분석
• Use of different approaches to model catch per unit effort abundance of fish(Vinod <i>et al.</i> , 2016)	로지스틱 회귀, 의사결정트리, 인공신경망모형	- 인도 구자라트 해안 트롤 및 자망 어업의 CPUE 예측을 위한 로지스틱 회귀, 의사결정트리, 인공신경망모형 접근 방법 연구
• Application of artificial neural network into the freshwater fish caught in Turkey(Benzer and Benzer, 2015)	인공신경망모형	- 터키 내수면어업 생산량 예측을 위한 인공신경망모형 분석
• Modelling and forecasting marine fish production in odisha using seasonal ARIMA model(Raman <i>et al.</i> , 2015)	ARIMA 모형	- 인도 동부 오디샤 지역의 어류 생산량 추정
• Crop Yield Forecasting Using Artificial Neural Networks: A Comparison between Spatial and Temporal Models(Guo and Xue, 2014)	인공신경망모형	- 다양한 인공신경망모형을 이용한 호주 퀸즐랜드 밀 생산량 분석, 곡물 생산량 예측 및 모형 비교
• Wheat yield prediction using artificial neural network and crop prediction techniques (Vikas and Dhaka, 2014)	인공신경망, 다중회귀, ARIMA 모형	- 밀 생산량 예측을 위한 인공신경망 모형 적용 기법, 원리 분석
• An analysis of the relationship between temperature variation and fish production in Lagos, Nigeria(Ezenwaji <i>et al.</i> , 2014)	피어슨 상관관계 분석	- 나이지리아 라고스(Lagos) 해안의 표층 수온과 어류 어획량의 상관관계 분석

연구 주제(제목)	분석 방법	주요 내용
- Weather Based Forecasting Model for Crops Yield Using Neural Network Approach. Statistics and Application (Laxmi and Kumar, 2011)	인공신경망모형	- 인공신경망모형을 이용한 인도의 지역별 기후변수 모델링, 쌀, 밀, 사탕수수 생산량 예측
A Review of Methods for Prediction of Potential Fish Production with Application to the Great Lakes and Lake Winnipeg(Leach <i>et al.</i>, 2011)	잠재 생산량 예측 연구 리뷰	- 잠재 생산량 추정 연구 리뷰 및 Great Lakes, Lake Winnipeg 의 잠재 생산량 추정 방식 적용
Statistical modelling of inland fish production in India(Paul and Das, 2010)	ARIMA	- 인도 내수면 어류 생산량 예측을 위한 모델링 및 전망, 모형 평가
Application of Vegetation Indices for Agricultural Crop Yield Prediction Using Neural Network Techniques(Panda <i>et al.</i>, 2010)	인공신경망모형	- 미국 노스다코타주 옥수수 연구현장 옥수수 수확량 예측
Neuralnet: Training of Neural Networks(Fritsch, 2010)	인공신경망모형 고찰	- 다층구조를 가진 지도학습 기반 인공신경망 알고리즘 리뷰 및 적용 방법에 대한 연구
Integrated dynamic aquaculture and wastewater tretment modelling for recirculating aquaculture system(Wik <i>et al.</i>, 2009)	순환여과 폐수처리 시스템 모델링 및 시뮬레이션	- 순환여과식 양식시스템 성능 분석, 예측을 위한 동적 양식 및 폐수 처리 모델링 프레임워크 제시
Bioeconomic modelling and salmon aquaculture: An overview of the literature(Pomeroy <i>et al.</i>, 2008)	생물경제모델링 연구 리뷰	- 생물경제모델링 수행 연구 리뷰, 칠레 연어생산의 생물경제모형 이슈 분석
Short-term forecasting of halibut CPUE: Linear and non-linear univariate approaches(Czerwinski, Estrada, and Hernando, 2007)	ARIMA, 인공신경망모형	- ARIMA 및 인공신경망모형을 이용한 베링해와 알래스카만 태평양 넙치(Pacific halibut)의 단위노력당 어획량 예측

연구 주제(제목)	분석 방법	주요 내용
• Neural network and its application in aquaculture(Balakrishnan <i>et al.</i> , 2007)	인공신경망모형 고찰	- 인공신경망모형의 특징 및 양식산업 적용 방안(수산물 질병 예측, 품질관리, 항생제 잔류량 예측, 수산물 생산 맵 모델링 등)
• Artificial neural networks for rice yield prediction in mountainous regions(Ji <i>et al.</i> , 2007)	인공신경망, 다중 회귀모형	- 중국 푸젠성 쌀 생산 예측을 위한 인공신경망모형 추정
• Climate variability and its effects on major fisheries in Korea(Kim <i>et al.</i> , 2007)	시계열 데이터 분석	- 기후변화 및 기후변화가 한국의 주요 어업에 미치는 영향분석
• The effects of water temperature in aquaculture management(Leon, Hernandez, and Leon-Santana, 2006)	어류 성장 모델링	- 카나리아 제도 해상 양식장의 수온이 적정 크기의 양식 돔(sea bream) 중량 및 크기 관리에 미치는 영향 분석 - 어류의 성장함수를 포함한 양식장 예상 수익률 예측 모델 연구
• Effect of rainfall as a component of climate change on estuarine fish production in Queensland, Australia(Meynecke <i>et al.</i> , 2006)	잉여생산량 모형	- 기상 변수와 하구의존성 (estuary- dependent) 어류의 관계 분석- 어획량, 강우량, 남방진동계수(SOI)의 상관관계분석
• Fish production and climate: Sprat in the Baltic sea(Mackenzie and Koster, 2004)	상관관계 분석	- 발틱해 청어 가입량과 수온, 빙하면적, 기후의 상관관계 분석
• Aquaculture Production Trends Analysis(Tacon, 2003)	시계열 분석	- 지역별, 수역별, 품종별 양식수산물 생산 동향 분석
• Predicting fish distribution in a mesotrophic lake by hydroacoustic survey and artificial neural networks(Brosse, Lek, and Dauba, 1999)	인공신경망, 다중 회귀모형, Generalized Additive Model	- 프랑스 페어루프 호수의 평균 수심, 바닥 경사, 샘플링 수층, 산소농도, 수온에 따른 어류 분포 예측

연구 주제(제목)	분석 방법	주요 내용
• Systems modelling and bioeconomic modelling I aquaculture(Cacho, 1997)	시스템모델링, 생물경제모델링 연구 리뷰	- 시스템 모델링, 생물경제 모델링 종류, 구조 분석 및 향후 전망
• Fish production correlated with primary productivity, not the morphoedaphic index(Downing, Plante, and Lalonde, 1990)	단순/다중회귀분석	- 일차생산(식물플랑크톤)과 생물량, 어업 생산량의 상관관계 분석
• Water quality and soil condition of fish ponds in some states of India in relation to fish production(Banerjea, 1967)	토양 및 물 샘플링, 성분 분석, 생산성 분석	- 인도 양어장의 물, 토양 샘플링 및 성분 분석 - 수질 및 토양 조건과 어류 생산성의 상관관계 분석

자료:저자 작성

3. 선행연구와의 차별성

본 연구의 차별성은 두 가지로 요약할 수 있다. 첫째는 양식수산물 단수 예측을 위해 기상 및 수질 변수를 활용했다는 점이다. 지금까지의 수산업 수급 예측 연구는 대부분 어획노력량, 시설량 등의 인위적 투입 요소나 비용, 가격 등의 경제적 요소를 변수로 활용하였다. 그러나 본 연구에서는 수산물의 생산량을 결정짓는 요인 중 작황에 해당하는 단수(단위생산량)에 대해 직·간접적 영향을 미치는 기상 및 수질 데이터를 독립변수로 활용했다.

둘째는 다양한 조건의 인공신경망 양식수산물 단수 예측모형을 구축하고 비교하였다는 점이다. 기존의 수산 분야 연구에서는 계량경제 이론에 입각한 경제모형 및 생물경제모형 등 회귀분석모형 기반의 예측이 대부분이었다. 본 연구에서는 양식수산물 단수 예측을 위해 세 가지 인공신경망 알고리즘(MLP, LSTM, 양방향 LSTM)과 변수 조합, 파라미터 설정을 달리하는 등 다양한 조건에서의 예측 결과를 비교하여 인공신경망모형의 유효성을 검토하였다.

02

김 수급 동향 및 김 양식 방법

제2장에서는 본 연구의 분석 대상인 김 양식산업의 현황에 대해 살펴보았다. 먼저 제1절에서는 우리나라의 김 양식 규모와 수급 동향을 파악하기 위해 면허 및 어가 수, 시설 동향, 생산 동향, 가격 동향, 그리고 수출 동향에 대해 살펴보았다. 제2절에서는 김의 종류와 생육 조건, 양식 방법과 지역별 양식 환경에 대해 알아보았다.

제1절 김 수급 동향

통계청 어업생산통계에 따르면, 우리나라의 김 양식 생산량은 1970년 3만 5,800톤을 기록한 이래 1984년에는 10만 톤을 넘어섰고, 2002년 20만 톤을 넘어 2019년에는 60만 6,000톤³⁾까지 증가했다. 이처럼 지난 50년 동안 국내 김 양식업은 크게 확대되었으며, 2014년 이후로 줄곧 수산물 수출 실적 1위를 차지하는 등 고부가가치 제품으로서의 중요성도 날로 커지고 있다.

3) 이는 천해양식수산물 해조류 전체 생산량의 약 33%에 달한다(자료: 통계청 어업생산동향조사, 각 연도).

그러나 이와 같은 산업적 성장에도 불구하고 아직도 생산 현장에서는 고수온 등 이상 해황에 따른 작황 부진, 반복적으로 발생하는 황백화 및 질병 등의 영향으로 수급 불안 상황이 빈번히 발생한다. 그리고 이렇게 나타나는 수급 불안정은 생산 어가의 경영 악화를 초래할 뿐만 아니라 산업 성장을 저해하는 걸림돌로 작용하고 있다.

1. 생산 동향

1) 면허 및 어가 수

통계청 어업생산통계에서는 국내 김 양식 면허 및 어가 수 관련 데이터를 2019년 자료까지 공표하고 있다. 2019년 기준 전국의 김 양식 면허 수는 806건으로 2015년 이후 증가세를 보였다. 이 수는 5년 전인 2014년과 비교하면 4.5% 증가한 것이다.

〈표 2-1〉 연도별 김 양식 면허 및 어가 수

단위: 건, ha, 개

구 분	2014	2015	2016	2017	2018	2019
면허(건)	771	758	778	781	792	806
면적(ha)	54,963	55,254	57,150	63,671	62,448	63,526
어가(개)	1,623	1,873	2,251	2,178	2,278	2,307

자료: 통계청, 천해양식어업권통계(검색일: 2021. 9. 24)

양식 면허 면적 또한 건수 증가와 함께 최근 5년간 늘었다. 2019년 김 양식 면허 면적은 6만 3,526ha로 전년 대비 소폭 늘었고, 2014년과 비교하면 8,563ha 증가했다. 김 수출 증가세에 힘입어 시설양 또한 지속해서 늘어나는 상황이다.

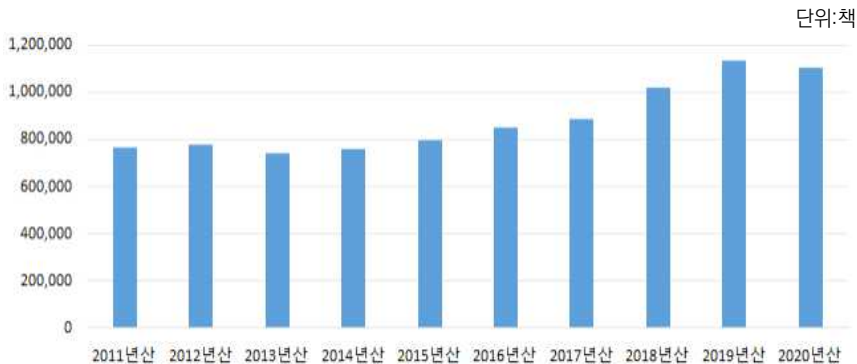
한편 어촌의 고령화 및 인력 감소 추세로 전복, 미역 등 기타 양식 품목

의 생산 여가가 줄어드는 것과 상반되게, 김의 경우 해외시장이 계속해서 확대됨에 따라 여가 수 또한 늘고 있다.

2) 시설 동향

수산업관측센터에 따르면 2021년산 기준 김 시설량은 110만 5,000여 책이었다. 이 양은 전년도인 2020년산 대비 소폭 줄어든 것이지만 2018년산 이후 3년째 100만 책을 넘어섰다. 10년 전인 2011년 기준 김 시설량은 76만 7,700여 책에 불과하였으나, 이후 전 세계로 우리나라 김이 수출됨에 따라 시설량은 큰 폭으로 증가했다.

〈그림 2-1〉 연도별 김 시설량 추이



자료: KMI 수산업관측센터 「연도별 김 시설량」(내부자료)

3) 생산 동향

2020년산 김 생산량은 2019년산 대비 21.2% 감소한 1억 3,979만 속이었다. 2020년산 김 생산량은 역대 최고로 많았던 2019년산보다 21.2% 적었고, 최근 5년 평균에 비해서도 5.7% 줄었다. 이는 어기 동안 수온이 예

년보다 높았고 여기 초 태풍을 시작으로 강풍이 자주 발생하는 등 해황 여건이 김 생장에 적합하지 못했기 때문이다.

최근 10년 동안의 김 생산량 추이를 살펴보면 2011년부터 2013년까지 늘다가 이후 3년간 생산이 감소했다. 이후에도 2018년산과 2019년산을 제외하고는 김 시설량이 대체로 증가세를 보여온 것과 달리, 김 생산량은 매년 수온, 갯병, 황백화 등의 해황 여건에 따라 생산량에 상당한 차이가 발생한다.

〈그림 2-2〉 연도별 김 생산량 추이



〈표 2-2〉 연산별 김 생산량

단위: 만 속

구 분	2011년산	2012년산	2013년산	2014년산	2015년산
생산량	12,791	12,498	13,970	13,471	12,783
구 분	2016년산	2017년산	2018년산	2019년산	2020년산
생산량	12,379	14,416	16,791	17,746	13,978

자료: KMI 수산업관측센터 「김 생산량」(내부자료)

지역별로 김 생산량을 살펴보면, 주산지인 전남의 2020년산 김 생산량은 전국의 약 73%를 차지한 1억 183만 속이었다. 다음으로 생산량이 많은

지역은 인천·경기로 1,123만 속의 김이 생산되었고, 전북 1,087만 속, 충남 969만 속이었다. 전남의 경우 2020년산 작황이 이례적으로 좋지 못해 전년산 및 최근 5년 평균에 비해서도 생산량이 매우 적었고, 이 외에 전북과 인천·경기지역의 김 생산은 다소 양호했다. 2020년산 어기 동안 황백화 및 갯병 등이 발생한 충남지역의 김 생산량은 전년산 대비 대폭 감소했다.

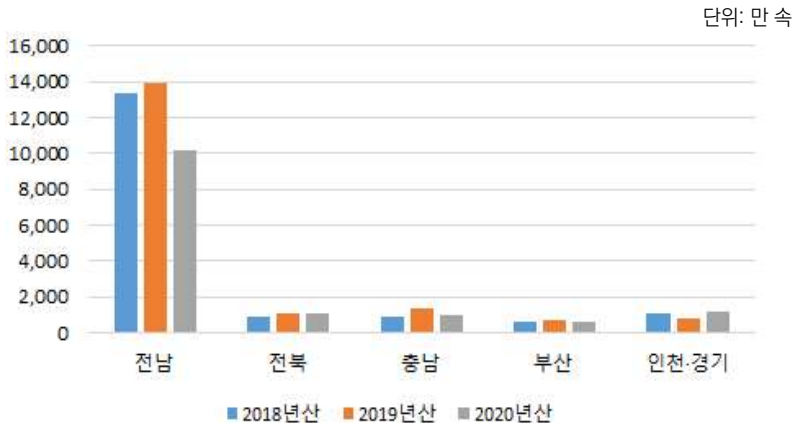
〈표 2-3〉 연산별·지역별 김 생산량

단위: 만 속

구 분	2016년산	2017년산	2018년산	2019년산	2020년산
전 국	12,379	14,416	16,791	17,746	13,978
전 남	9,436	10,468	13,381	13,887	10,183
전 북	830	1,235	890	1,033	1,087
충 남	1,061	1,583	869	1,399	969
부 산	351	487	561	677	616
인천·경기	700	643	1,089	749	1,123

자료: KMI 수산업관측센터 「김 생산량」(내부자료)

〈그림 2-3〉 연산별·지역별 김 생산량 추이



자료: KMI 수산업관측센터 「김 생산량」(내부자료)

2. 가격 동향

1) 산지가격

2020년산 물김 산지가격은 2019년산(866원/kg)보다 11.5% 높은 kg당 평균 966원이었다. 이는 평년과는 비슷한 수준이지만 10년 전인 2011년과 비교하면 50% 가까이 증가한 수준이다. 생산이 많아졌음에도 산지가격이 높은 수준을 유지하고 있는 이유는 비교적 단가가 높은 잇바디돌김 생산이 늘어났기 때문이다. 통상 물김 산지가격은 잇바디돌김이 주로 생산되는 10~11월 가격이 가장 높고 이후 여기 말까지 하락세를 보인다.

〈표 2-4〉 연산별 김 산지가격

단위: 원/kg					
구 분	2011년산	2012년산	2013년산	2014년산	2015년산
산지가격	650	752	661	728	784
구 분	2016년산	2017년산	2018년산	2019년산	2020년산
산지가격	1,014	1,103	1,074	866	966

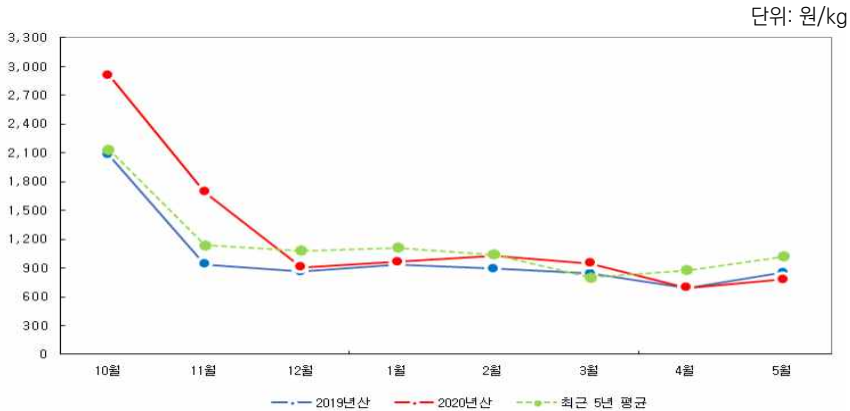
자료: KMI 수산업관측센터 「김 산지가격」(내부자료)

월별로는 5월을 제외한 모든 월에서 전년 동월 대비 산지가격이 비싸게 형성되었고, 특히 여기 초에 높았다. 10월과 11월에는 잇바디돌김 생산량 감소의 영향으로 전년 동기보다 산지가격이 크게 상승했다. 10월 평균 물김 산지가격은 최근 5년 내 가장 높은 kg당 2,907원에 형성되었고, 11월 가격 또한 전년산 대비 80.5% 증가한 1,691원이었다.

2020년산 물김의 경우 품질이 좋지 못해 12월 이후 산지가격 상승은 생산 감소에 비해 다소 제한적이었다. 12월 물김 평균 산지가격은 일반김 생산이 본격화됨에 따라 전월 대비 큰 폭으로 하락했다. 그러나 전반적인 작황 부진으로 전년 동월에 비해서는 5.3% 높았다. 1월⁴⁾에는 생산량이 전년

대비 20% 이상 감소했음에도 불구하고 품질 저하로 산지가격은 전년과 비슷한 kg당 평균 950원에 형성되었다.

〈그림 2-4〉 물김 월별 산지가격 추이



자료: KMI 수산업관측센터 「김 산지가격」(내부자료)

2) 도매가격

마른김 도매가격은 통상 당해 연산의 생산량에 따라 연초에 형성된 뒤 다음 연산의 햇김이 나오기까지 그 가격 수준에서 큰 변동이 없다. 2020년 마른김(김밥용김) 도매가격은 전년⁵⁾보다 높은 속당 4,671원에 형성되었다. 이는 2020년에 산지 물김 생산 부진에 따른 시장 내 마른김 반입 감소로 대부분 월에서 도매가격이 전년 동월 대비 높게 형성되었기 때문이다.

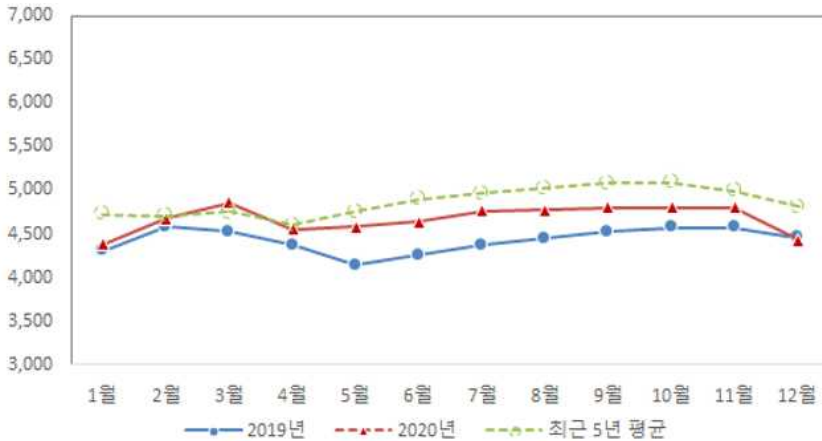
2020년 김 도매가격의 경우 전년에 비해 변동이 다소 작았는데, 이는 코로나19 장기화로 재래시장 내 거래가 부진하여 경매 거래 횟수가 줄었기 때문이다. 이에 햇김(일반김) 반입이 시작되는 12월에는 시장 내 많은 재고의 영향으로 도매가격이 전년 동월보다 낮은 수준으로 다시 내렸다.

4) 2020년 1월 김 생산량은 2,264만 속으로 전년 동월(2,959만 속) 대비 23.5% 감소했음

5) 2019년 연평균 마른김(김밥용김) 도매가격은 속당 4,426원이었음

〈그림 2-5〉 마른김 월별 도매가격 추이(김밥용김 중품 기준)

단위: 원/속



자료: KMI 수산업관측센터 「김 도매가격」(내부자료)

3. 수출 동향

1) 수출량

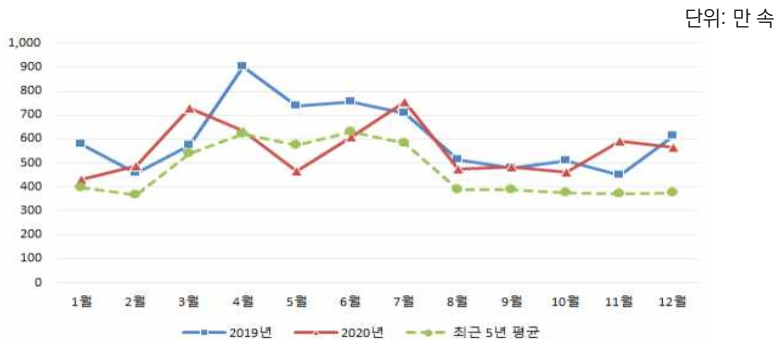
김 수출은 2011년에 1억 6,141억 달러를 나타낸 이후 10년 동안 시장이 꾸준히 성장하여 2020년에는 6억 달러 달성을 이룩하였다. 수출량 또한 2011년 2,807만 속에서 2020년에는 2.4배에 달하는 6,689만 속까지 늘었다. 특히 2020년 김 수출은 코로나19라는 악재 속에서도 최고 실적을 달성하여 매우 고무적이라고 할 수 있다.

수출금액을 연도별로 살펴보면 2011년에 1억 6,141만 달러였고, 2015년에는 약 두 배 증가하여 3억 달러를 달성하였다. 2017년에는 5억 달러, 2020년에는 6억 달러를 기록했으며, 정부는 김 수출금액 목표를 매년 새롭게 설정하며 다양한 방면에서 산업을 지원하고 있다.

2) 제품별·국가별

우리나라 김 수출에서 마른김과 조미김의 비중은 2020년 기준 각각 56.3%, 43.3%이다. 2020년의 경우 전년 대비 조미김(-6.4%) 및 마른김(-9.4%) 수출이 모두 줄었다. 주요 대상국별 비중을 살펴보면 수출 비중이 가장 높은 국가는 일본으로 연간 1만 4,009속이 수출되어 전체의 21.3%를 차지하였다. 그다음은 중국 16.8%, 미국 15.5%, 태국 15.4%순 이었다. 2020년에는 위의 주요 수출대상국으로의 수출이 대체로 부진했던 가운데, 이를 제외한 기타 국가로의 수출은 전년 대비 20% 이상 많았다. 오스트리아, 벨기에 등 유럽과 남미에서도 한국산 김 수요가 점차 늘고 있다.

〈그림 2-6〉 월별 김 수출량 추이



자료: KMI 수산업관측센터「김 수출량」(내부자료)

〈표 2-5〉 연도별 김 수출량 추이

단위: 속, 만 달러

구 분	2011년	2012년	2013년	2014년	2015년
수출량	2,807	3,806	3,914	3,798	4,407
수출금액	16,141	23,019	25,067	27,343	30,464
구 분	2016년	2017년	2018년	2019년	2020년
수출량	4,661	5,767	5,924	7,267	6,689
수출금액	35,292	51,278	52,544	57,911	60,155

자료: KMI 수산업관측센터「김 수출량」(내부자료)

제2절 김 양식 방법 및 지역별 양식 환경⁶⁾

1. 김의 종류 및 생육 조건

1) 김의 종류

(1) 참김(*Pyropia tenera*)

우리나라에 서식하는 김은 크게 일반김과 돌김의 두 가지로 구분하며, 일반김에는 참김과 방사무늬김, 돌김에는 잇바디돌김과 모무늬돌김 등이 있다.

첫째로 참김은 암수한몸이며 적갈색을 띤다. 서식 지역은 우리나라 남해안 일부 해안이며, 주로 바위나 굴 껍데기에 부착하여 자라고 파래와 혼생하여 자라기도 한다. 참김은 예로부터 우리나라와 일본에서 양식되어 온 품종이었으나, 인공채묘가 가능하고 환경적응성이 뛰어난 방사무늬김이 널리 보급되면서 참김을 양식장에서는 발견하기 어렵게 되었다. 또한 환경적 요인의 변화에 따라 자연 암반에서도 점차 사라지고 있어 일본에서는 참김을 멸종위기종으로 지정하기도 하였다.

(2) 방사무늬김(*P. yezonensis*)

방사무늬김은 일반김 종류 중 하나로 우리나라 전 연안, 주로 조간대의 중·하부에 서식하고 다른 해조류나 해산동물 등의 해양생물, 바위, 밧줄 및 인공 구조물 등 다양한 곳에 부착하여 자란다. 방사무늬김도 참김과 마찬가지로 암수한몸이며 색은 적갈색 또는 적록색을 띤다. 참김이 점차 양식장에서 사라진 것과 다르게, 방사무늬김은 인공채묘의 발전에 맞춰

6) 본 절의 내용은 국립수산물과학원(2018) 내용을 참고하여 작성하였다.

국내 김 양식의 주요 품종으로 자리 잡았다. 현재 국내에서 생산되는 김의 70% 이상이 일반김이라 불리는 방사무늬김이다.

〈그림 2-7〉 주요 양식 김의 종류



자료: 국립수산물과학원(2018), p. 19

(3) 잇바디돌김(*P. dentata*)

잇바디돌김은 앞서 두 종류의 김과는 달리 암수딴몸이며 엽체의 모양도 피침형으로 특이한 형태를 가진다. 색은 적갈색이며, 우리나라의 남·서해안, 제주도 조간대 상부 또는 비말대 등에서 주로 바위에 부착되어 자라는 특징이 있다. 잇바디돌김은 곱창김이라고도 불리며 주로 어기 초이른 시기에 전남 일부에서 양식이 이루어진다.

(4) 모무늬돌김(*P. seriata*)

모무늬돌김은 암수한몸이며 엽체의 형태는 원형, 신장형 또는 갈매기형으로 대체로 적갈색을 띤다. 서식지는 우리나라 남해안과 서해안 전역의 조간대 중부이며, 잇바디돌김과 달리 어기 말까지 전남 해역에서 양식된다.

(5) 기타

위의 네 가지 주요 김 종 이외에도 큰참김(*P. tenera* var. *tamstsuenensis*), 큰방사무늬김(*P. yezonensis* f. *narawaensis*), 둥근김(*P. kuniedai*), 긴잎돌김(*P. pseudolinearis*) 등이 국내에 분포해 있다.

2) 김의 생육 조건

김의 생육에 영향을 미치는 자연 환경적 요인에는 수온, 영양염, 염분 등이 있다. 다음에서는 각각의 요인들에 대해 기술하고, 최적의 조건을 살펴보고자 한다.

(1) 수온

김 생육에 영향을 미치는 여러 가지 요인 중 수온은 김 생육 전 단계에 걸쳐 상당한 영향을 미치기 때문에 가장 중요한 요인이다. 김 채묘기인 9~10월경에는 수온이 22℃ 이하가 되어야 종자가 잘 붙으며, 포자가 발아하기 위해서는 15~22℃의 수온이 유지되어야 한다. 또한 김 엽체가 생장하는 성육기인 11~4월까지의 수온이 5~15℃를 형성해야 한다. 우리나라 김 주산지인 남해 연안의 겨울철 수온은 12월에 접어들면서부터 15℃ 이하로 내려가며, 이후 4월경까지 김 생장에 적합한 수온대가 만들어진다.

우리나라에서 양식되는 김 종류 중 생산량이 가장 많은 일반김(방사무늬김)의 경우 수온이 15℃ 부근까지 내려가는 동안 영양번식(무성생식)을 반복하다, 수온이 15℃ 이하로 떨어지면 번식이 중지되고 엽상체의 길이 생장이 이루어진다.

(2) 영양염

김 생육에 영향을 미치는 두 번째 요인은 영양염이다. 영양염은 인화합물, 질소화합물 등의 염류로, 해수 및 육수에 포함되어 식물 플랑크톤이나 조류의 영양이 되는 물질을 말한다.⁷⁾

김이 바다에서 잘 자라기 위해서는 이 영양염이 일정 수준 이상 존재해야 한다. 보통 김 생산을 위해서는 질산염 0.1mg/L, 인산염 0.03~0.95 mg/L 이상의 염류가 필요하며, 이 외에도 철, 망가니즈, 구리, 코발트 등의 미량 원소도 필요하다.

해양에 적절한 수준의 영양염이 형성되지 못할 경우, 육안으로 관찰 시 김 색택이 황백색으로 바뀌고 현미경으로 관찰 시 병균에 대한 감염 없이 세포질 내 액포가 비대해지는 상태인 황백화 현상이 발생한다. 김에 황백화가 오면 염체가 퇴색하거나 탈락하여 물김 품질이 저하되고 생산량이 감소한다.

(3) 풍파(풍속)

바람은 김 생장에 영향을 미치는 해양 환경을 조성한다는 점에서 또 하나의 중요한 요인이라 할 수 있다. 일반적으로 양식시설이 파괴되지 않을 정도의 풍파는 필요하다. 바람이 불면 공기의 용해량이 많아져 이산화탄소, 산소가 김 양식장에 많이 공급될 수 있고, 양식장 해수의 소통이 원활해져 영양염 공급이 촉진된다.

(4) 조류

조류는 밀물과 썰물에 의해 주기적으로 일어나는 해수의 흐름을 말하

7) 네이버 지식백과(검색일: 2021. 9. 23)

는데, 바람이나 태풍 등 기타 자연적 환경에 의해 조류 소통이 변화할 수 있다. 또한 양식시설 밀집도 등 인위적인 요인에 의해서도 조류 소통은 달라진다.

일반적으로 김 양식장에 적합한 조류는 지주식 양식장의 경우 유속 20cm/sec 이상으로 조류가 빠르고 바람이 잘 부는 곳이 좋다. 영양화 정도의 차이에 따라 부영양화 어장은 10cm/sec, 빈영양화 어장은 30cm/sec의 유속이면 양호하다. 김 생육 단계에 따라서는 채묘 시기에는 7cm/sec, 발아기에는 7~25cm/sec의 다소 빠른 유속이 좋다.

(5) 햇빛

해상에서 양식이 이루어지는 김의 특성상 햇빛도 김 생육에 영향을 미치는 요인 중 하나이다. 이를 '조도'로 표현할 수 있는데, 조도란 단위 면적당 주어지는 빛의 양을 말한다. 빛은 김의 수직분포(서식대)를 결정짓는다.

김의 보상점은 300~500Lux로서 약한 광에서도 광합성이 가능하다는 특징이 있다. 이로 인해 일반적으로 햇빛은 김 생육에 있어 필수적이기는 하나, 김 양식장 선정 시 미리 고려되는 부분인 수심, 지질 등에 의해 결정 되기 때문에 특이사항이 없는 한 변화가 잦은 요인은 아니다.

(6) 기타

기타 김 생육에 영향을 미치는 요인으로는 김 양식장의 위치(가 작업 및 관리하기에 편한 곳에 있는지 여부) 등이 있으며, 최근에는 중국에서 떠내려온 팽생이모자반으로 인한 피해 등과 같은 잡조에 의한 피해 발생 여부 등이 새롭게 고려해야 할 요인이 되고 있다.

2. 김 양식 방법 및 지역별 양식 환경

1) 김 양식 방법

국내 김 양식 방법은 크게 두 가지, 지주식(말목식)과 부류식(뜬흙림발식)으로 나눌 수 있다. 그리고 이 중 부류식은 다시 노출 부류식과 무노출 부류식으로 나뉜다.

(1) 지주식

지주식은 김 양식 방법 중 가장 역사가 오래된 방식이다. 지주식 양식법은 주로 수심이 낮은 내만의 저질에 말목을 고정하여 김발을 시설하는 방법으로, 조석에 따라 하루 2번 김발이 공기 중에 노출된다. 김발을 공기 중에 노출하면 건조에 의한 규조류 및 기타 부착생물이 김발에 붙는 것을 억제할 수 있으며, 갯병 예방뿐 아니라 더 품질 좋은 김을 생산할 수 있다.

〈그림 2-8〉 지주식 김 양식장의 만조 시 김발 부동과 간조 시 김발 노출



자료: 국립수산물과학원(2018), p. 57

우리나라에서 지주식 김을 생산할 수 있는 지역은 많지 않다. 남서해 연안 중 조석 차이를 느낄 수 있는 해역에서만 지주식 김 양식이 이루어진다. 신안, 무안 등 일부에서 지주식을 사용하며, 최근에는 해당 지역에서도 점차 김 양식이 대형화, 자동화됨에 따라 지주식 양식을 하는 어가가 줄어들고 있다.

(2) 부류식

부류식은 김발을 해수면에 띄워서 양식하는 방법을 통칭하며 ‘뜬흠림밭’이라고 불리기도 한다. 지주식과 달리 김발을 공기 중에 노출하지 않을 뿐 아니라 말목을 사용하지 않고 양식을 할 수 있어서 내만뿐 아니라 외해에서도 이 방식을 통해 김 양식이 가능하다.

〈그림 2-9〉 부류식의 뒤집기에 의한 노출(왼쪽 : 수면부동, 오른쪽 : 노출)



자료: 국립수산물과학원(2018), p. 62

부류식은 무노출 상태이므로 지주식에 비해 생장이 빨라 생산성이 높다는 이점을 가지고 있는 반면, 갯병에 약하다는 단점이 있다. 최근에는 이를 개선하기 위해 스티로폼 뜬을 부류식 김발에 부착하고, 양식 밧을 인위적으로 뒤집어 일정 시간 동안 공기 중에 김발을 노출시키는 노출부류식 방법도 사용되고 있다. 부류식 시설은 단독으로 설치할 수도 있고, ‘세트밭’이라고 하여 여러 개의 김발을 연결하여 시설할 수도 있다.

(3) 지역별 품종 및 양식 방법

앞서 김 품종과 양식 방법을 살펴보았다. 김은 우리나라 서해, 남서해, 남해, 남동해까지 넓은 지역에서 양식되고 있으며, 지역별로 양식품종 및 방법이 모두 다르다. 다음의 표는 지역별 주요 양식 종과 양식법을 정리한 것이다.

먼저 충남과 전북지역은 합쳐서 서해권으로 묶을 수 있는데, 충남에서는 서산, 보령, 서천, 태안에서 김이 생산되며 전북에서는 군산, 고창, 부안 등에서 생산된다. 서해권 김 양식의 특징은 생산되는 김 종류가 일반 김에 해당하는 방사무늬김이고, 양식 방법은 부류식을 여러 개 묶어서 시설하는 세트발 방식을 주로 사용한다는 것이다.

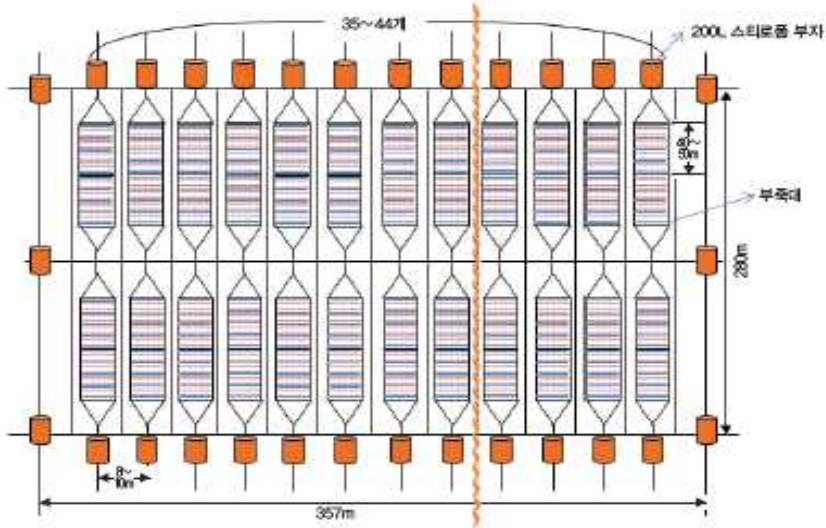
〈표 2-6〉 주요 김 품종과 양식법

지 역		주요 양식품종	양식법
서해권(충남, 전북)		방사무늬김	무노출부류식/셋트발
남서해 (저해)	신 안	방사무늬김, 모무늬돌김, 잇바 디돌김	지주식
	해 남		지주식, 노출부류식
	무 안		지주식
	완 도		노출부류식
남동해(부산)		방사무늬김	무노출부류식/셋트발

자료: 국립수산물과학원(2018), p. 65

김 시설 규격은 충남 서천의 경우 10ha 어장에 1세트를 시설 기준으로 하여 1세트 크기는 가로 357m, 세로 280m로 표준화하였다. 김발의 너비는 2m로 동일하나 길이는 40m, 50m, 60m 등 다양한데, 길이에 상관없이 1책이라 부른다.

〈그림 2-10〉 서천지역 김 양식시설 평면도

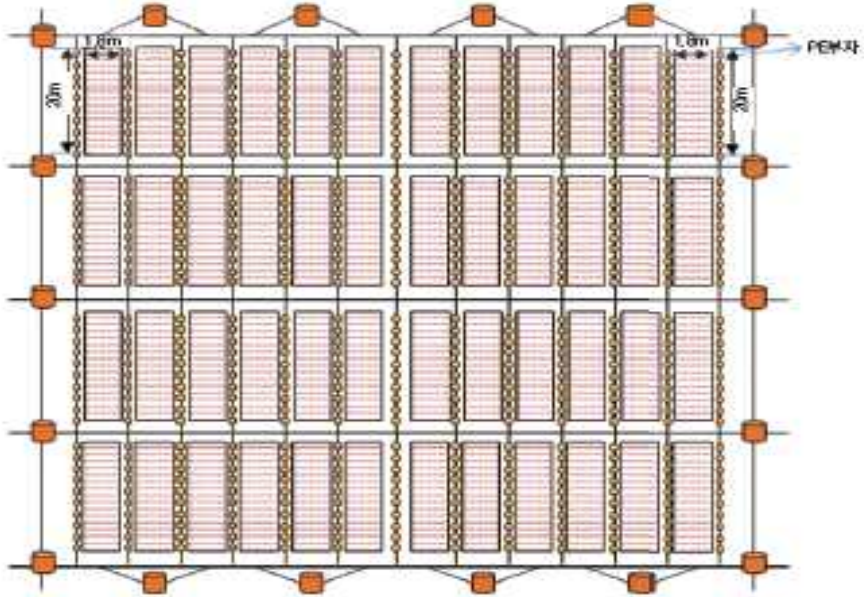


자료: 국립수산물과학원(2018), p. 66

다음으로 남서해권에 해당하는 지역은 김 생산의 주산지인 전남이다. 전남에서는 목포, 여수, 고흥, 장흥, 강진, 해남, 무안, 함평, 영광, 완도, 진도, 신안 등 바다를 인접하고 있는 지자체에서 대부분 김 양식을 한다. 이는 전남의 바다 환경이 다양한 양식 종, 양식 방법을 사용하기에 적합하기 때문이다. 전남에서는 방사무늬김, 잇바디돌김, 모무늬돌김 등 국내에서 유통되는 모든 종류의 김이 생산된다. 양식 방법 또한 다양한데, 신안과 무안에서는 지주식 방법으로 김을 양식하고, 해남에서는 지주식과 노출부류식으로, 완도, 진도, 고흥 등지에서도 대부분 노출부류식 방법을 사용한다.

전남의 김밭 규격은 일반적으로 2m×40m(너비×길이)를 1책으로 사용하고 있으며, 1ha당 시설할 수 있는 책 수는 13책으로 제한된다.

〈그림 2-11〉 부산지역 김 양식시설 평면도



자료: 국립수산물학원(2018), p. 67

마지막은 부산에 해당하는 남동해이다. 부산에서는 사하구, 강서구에 김 양식시설이 분포해 있다. 부산지역에서 생산되는 김은 대부분 일반김에 해당하는 방사무늬김이고, 양식 방법은 서해권과 같이 부류식을 여러 개 묶어서 시설하는 세트발이다.

부산의 세트발은 어장의 형태, 면적에 따라 정사각형 또는 직사각형 형태를 띠며 김발 1개의 크기는 1.8m×20m가 일반적인 규격으로, 1ha당 24책이 표준이다. 앞서 서해권과 시설 방법 중 차이를 보이는 부분은 김발을 띄우기 위해 시설 중간에 사용되는 부축대의 재질이 나무가 아닌 PE(polyethylene) 성질이라 충남이나 전북에 비해 부력이 다소 떨어진다는 점이다.

2) 지역별 양식 환경⁸⁾

김 양식 환경의 특징을 해역별로 나누어 살펴보기 위해 국립수산물과학원의 서해권, 남서해권, 남동해권의 어장 환경 모니터링 자료를 살펴보았다. 현재 국립수산물과학원에서는 어장관리법 제6조⁹⁾에 근거하여 우리나라 해역에 조사 거점을 두고 정기적으로 어장 환경 조사를 실시¹⁰⁾하고 있다. 조사 항목은 수온, 염분, 수소이온농도(PH), 투명도, 용존산소(DO), 화학적산소요구량(COD), 용존무기질소(DIN), 용존무기인(DIP) 등이다. 다음에서는 이 중 김 생장에 많은 영향을 미치는 요인인 수온, 용존무기질소(DIN), 용존무기인(DIP) 수치를 해역별로 살펴보고자 한다.

(1) 서해권(경기, 충남, 전북)

서해권 어장은 경기, 충남, 전북지역으로 볼 수 있다. 국립수산물과학원의 서해권 해조류 양식 어장 조사 거점 중 김 양식장이 있는 지역은 인천, 경기, 태안, 보령·서천, 군산, 고창·영광 연안이다.

가. 수온

서해권 김 양식 어장의 경우 국내 김 생산 해역 중 가장 북쪽에 있으며 특히 경기, 인천지역 수온은 충남보다 평균 1℃ 이상 낮은 분포를 보였다. 다만 비교적 외해에 해당하는 충남 태안의 경우 전체 지역 중 가장 낮은

8) 국립수산물과학원의 어장 환경모니터링 조사 결과를 바탕으로 작성하였다.

9) (어장 환경의 조사) ① 해양수산부장관은 어장의 효율적인 이용·보전과 어장 환경 상태 및 오염원의 측정·조사 등을 위하여 해양수산부령으로 정하는 바에 따라 어장 환경조사망을 구성·운영하고 정기적으로 어장 환경을 조사해야 한다. 어장관리법 제6조

10) 국립수산물과학원은 전국 연안 및 주요 양식어장 주변 해역의 정기적 조사를 통해 해양환경 현황을 파악하고, 양식어장의 지속적 활용을 위한 체계적인 환경관리 정책 수립을 위한 기초자료로 활용하기 위해 전국 269개(2021년 기준) 정점(59개 해역)에서 수질 및 퇴적물 오염에 영향을 미치는 성분을 조사하는 양식 어장 모니터링 사업을 실시하고 그 결과를 홈페이지에 공표하고 있음

수온대를 보였다. 반면 전북 고창과 영광은 서해권 김 양식 어장 중 수온이 가장 높게 형성되었다.

〈표 2-7〉은 2021년 3월 기준, 각 해역의 수온 분포를 나타낸 것이다. 2021년 연안 표층 수온은 보령·서천 등 일부 해역을 제외하고 대체로 평년보다 다소 낮게 형성되었다.

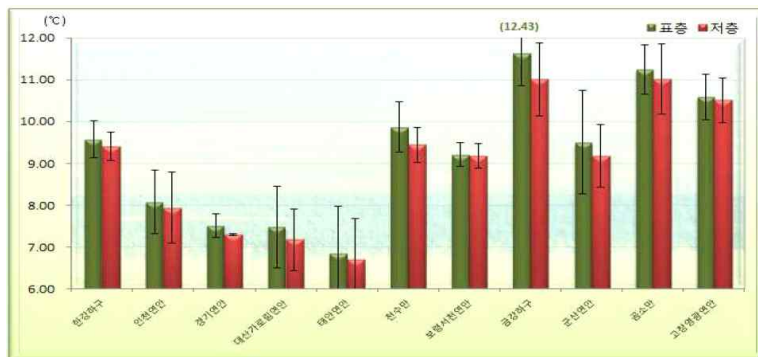
〈표 2-7〉 서해권 김 양식 어장 표층 수온(2021. 3. 기준)

지 역	수온	평년 대비
인 천	7.10~8.95(8.08±0.75)℃	0.46℃ ↓
경 기	7.31~7.72(7.52±0.29)℃	0.25℃ ↓
태 안	5.76~8.59(6.84±1.14)℃	0.23℃ ↓
보령·서천	8.89~9.40(9.21±0.28)℃	0.08℃ ↑
군 산	8.32~11.78(9.50±1.24)℃	0.59℃ ↓
고창·영광	10.27~11.21(10.59±0.54)℃	0.39℃ ↓

자료: 국립수산물과학원, 어장환경모니터링 통계검색(검색일: 2021. 9. 23)

2021년 3월 기준으로는 모든 어장의 수온이 김 생육기 적정 수온인 5~15℃ 범위를 보였으나, 수온이 낮은 12~1월에는 경기, 인천 및 태안 등 일부 지역에서 업체 냉해 피해가 발생하기도 하며, 수온이 높은 지역에서는 어기 초 고수온으로 인한 채묘 부진 등이 나타나기도 한다.

〈그림 2-12〉 서해권 수온 분포도



자료: 국립수산물과학원(2021c), p. 3

나. 용존무기인(DIP)

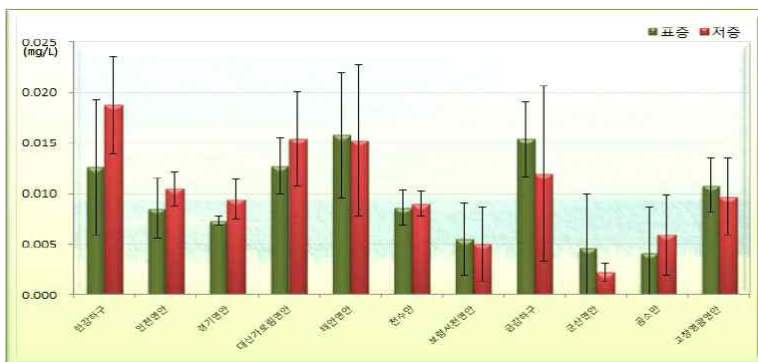
용존무기인(DIP)은 김 생장에 필요한 영양염 중 하나로, 김 양식어장 환경모니터링 대상 요소 중 하나이다. 2021년 3월 기준 서해권 김 양식장의 DIP 수치를 살펴본 결과, 대체로 평년보다 낮은 것으로 조사되었다. 일반적으로 영양염은 풍부할수록 김 생장에 도움을 주며, 표층의 DIP 농도 값은 6개 거점 해역 중 태안이 가장 높은 것으로 나타났다.

〈표 2-8〉 서해권 김 양식 어장 용존무기인(DIP) (2021. 3. 기준)

지 역	용존무기인(DIP)	평년 대비
인 천	0.003~0.012(0.009±0.003)mg/L	0.011mg/L ↓
경 기	0.007~0.008(0.007±0.000)mg/L	0.004mg/L ↓
태 안	0.007~0.023(0.016±0.006)mg/L	0.002mg/L ↓
보령·서천	0.003~0.010(0.005±0.004)mg/L,	0.001mg/L ↑
군 산	0.001~0.015(0.005±0.005)mg/L,	0.000mg/L-
고창·영광	0.008~0.014(0.011±0.003)mg/L,	0.003mg/L ↑

자료: 국립수산과학원, 어장환경모니터링 통계검색(검색일: 2021. 9. 23)

〈그림 2-13〉 서해권 용존무기인(DIP) 분포도



자료: 국립수산과학원(2021c), p. 7

다. 용존무기질소(DIN)

용존무기질소(DIN)도 앞서 DIP와 마찬가지로 김 생장에 필요한 영양염이자 김 양식어장 환경모니터링 대상 요소 중 하나이다. 통상적으로 DIN 농도는 민물 유입이 많은 강 하구와 가까운 해역에서 높게 나타난다. DIN 농도 값은 0.07mg/L 이상이면 양호한 것으로 볼 수 있다.

2021년 3월 기준 서해권 김 양식장의 DIN 수치를 살펴본 결과, 북쪽 연안인 인천, 경기 및 태안 해역은 평년보다 낮은 농도를 보였으며, 상대적으로 남쪽에 위치한 보령·서천, 군산, 고창·영광 연안의 농도는 평년보다 높은 농도를 보였다.

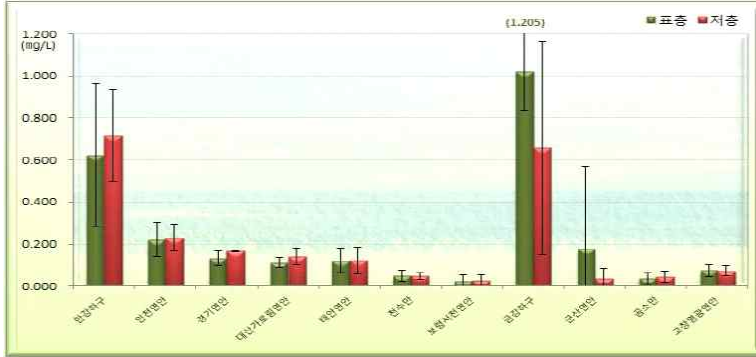
올해의 경우 표층 DIN 농도 값은 6개 거점 해역 중 인천이 평년 대비 다소 낮아졌으나 절대 수치는 가장 높은 것으로 나타났고, 보령·서천 연안이 가장 낮게 나타났다. 그러나 전반적으로 양호한 수치였다. 영양염의 농도가 낮아지면 김에 다양한 질병이 발생하게 된다.

〈표 2-9〉 서해권 김 양식 어장 용존무기질소(DIN) (2021. 3. 기준)

지 역	용존무기질소(DIN)	평년 대비
인 천	0.132~0.378(0.221±0.082)mg/L	0.030mg/L ↓
경 기	0.109~0.157(0.133±0.034)mg/L	0.003mg/L ↓
태 안	0.036~0.189(0.121±0.059)mg/L,	0.023mg/L ↓
보령·서천	0.008~0.058(0.025±0.029)mg/L,	0.001mg/L ↑
군 산	0.007~0.981(0.175±0.395)mg/L,	0.148mg/L ↑
고창·영광	0.046~0.104(0.075±0.029)mg/L	0.035mg/L ↑

자료: 국립수산과학원, 어장환경모니터링 통계검색(검색일: 2021. 9. 23)

〈그림 2-14〉 서해권 용존무기질소(DIN) 분포도



자료: 국립수산물과학원(2021c), p. 13

(2) 남서해권(전남)

남서해권 어장은 김 주산지인 전남에 해당한다. 국립수산물과학원의 남해 서부연안 해조류 양식 어장 조사 거점 중 김 양식장이 있는 지역은 목포, 진도, 완도, 도암만, 득량만, 고흥 연안이다.

가. 수온

남서해권 김 양식 어장은 국내에서 김 생산이 가장 많이 이루어지는 곳으로 다양한 품종의 김이 양식되며, 다양한 방식의 김 시설이 가능하다. 전국에서 가장 먼저 물김이 생산되고 가장 늦게 마무리되는 진도지역이 남서해권에 속한다.

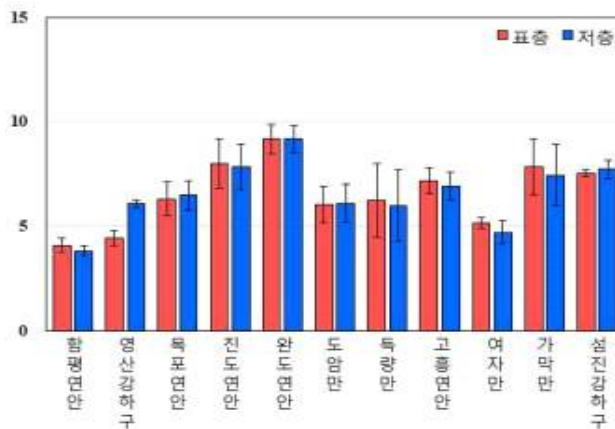
〈표 2-10〉은 2021년 2월 기준, 각 해역의 수온 분포를 나타낸 것이다. 2021년 연초에는 예년보다 날씨가 따뜻했고, 수온 역시 모든 해역에서 평년 대비 높게 형성되었다. 수온이 가장 낮게 형성되는 2월에도 남서해권 김 양식 거점 해역은 적정 김 성장 수온인 5~15℃를 유지하였다.

〈표 2-10〉 남서해권 김 양식 어장 표층 수온(2021. 2. 기준)

지 역	수 온	평년 대비
목 포	5.7~7.2(6.3 ± 0.8)℃	0.2℃ ↑
진 도	6.5~9.4(8.0 ± 1.2)℃	1.6℃ ↑
완 도	7.9~10.1(9.2 ± 0.7)℃	0.7℃ ↑
도암만	5.4~7.0(6.0 ± 0.9)℃	0.4℃ ↑
득량만	5.0~9.4(6.2 ± 1.7)℃	1.0℃ ↑
고흥	6.3~7.8(7.2 ± 0.6)℃	1.1℃ ↑

자료: 국립수산과학원, 어장환경모니터링 통계검색(검색일: 2021. 9. 23)

〈그림 2-15〉 남서해권 수온 분포도



자료: 국립수산과학원(2021a), p. 3

해역별 수온 분포를 살펴보면 남서해권에서 2월 수온이 가장 높은 곳은 완도 연안으로 평균 7.9~10.1℃였다. 그 다음으로 수온이 높은 곳은 진도 해역이었고, 김이 양식되는 해역 중에는 목포 연안이 가장 낮은 수온대를 보였다. 수온이 가장 낮은 해역도 겨울철 5℃ 이상으로, 남서해권에서는 냉해 피해가 거의 없어 최적화된 김 양식지라고 할 수 있다.

나. 용존무기인(DIP)

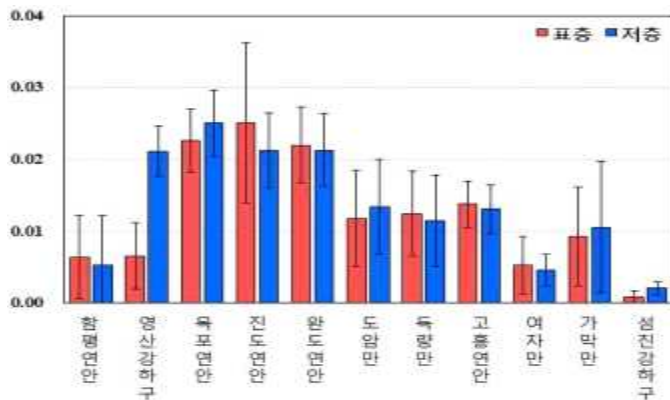
2021년 2월 기준 남서해권 김 양식장의 DIP 수치를 살펴본 결과, 모든 해역에서 평년보다 높았다. 김 양식장이 많이 분포해 있는 목포, 진도, 완도, 고흥 등의 연안에서는 표층의 DIP 농도 값이 높았다.

〈표 2-11〉 남서해권 김 양식 어장 용존무기인(DIP) (2021. 2. 기준)

지 역	용존무기인(DIP)	평년 대비
목 포	0.018~0.027(0.018~0.027)mg/L	0.004mg/L ↑
진 도	0.015~0.040(0.025±0.011)mg/L	0.009mg/L ↑
완 도	0.014~0.034(0.022±0.005)mg/L	0.007mg/L ↑
도암만	0.006~0.019(0.012±0.007)mg/L	0.005mg/L ↑
득량만	0.004~0.022(0.012±0.006)mg/L	0.008mg/L ↑
고흥	0.010~0.017(0.014±0.003)mg/L	0.008mg/L ↑

자료: 국립수산물과학원, 어장환경모니터링 통계검색

〈그림 2-16〉 남서해권 용존무기인(DIP) 분포도



자료: 국립수산물과학원(2021a), p. 7

6개 거점 해역 중 진도 연안이 가장 높았고, 그다음은 목포, 완도 순이었다. 영양염이 가장 풍부한 진도 연안에서는 여기 초 잇바디돌김, 이후에는 방사무늬김과 모무늬돌김 등 다양한 품종의 김이 생산된다.

다. 용존무기질소(DIN)

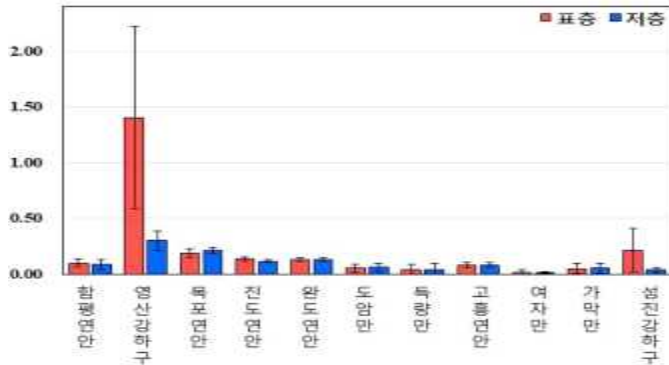
2021년 2월 기준 남서해권 김 양식장의 DIN 수치를 살펴본 결과, 표층 DIN 농도 값은 6개 거점 해역 중 모든 연안에서 평년 대비 높은 것으로 나타났다.

〈표 2-12〉 남서해권 김 양식 어장 용존무기질소(DIN) (2021. 2. 기준)

지 역	용존무기질소(DIN)	평년 대비
목 포	0.063~0.132(0.100±0.035)mg/L	0.063mg/L ↑
진 도	0.115~0.157(0.135±0.020)mg/L	0.017mg/L ↑
완 도	0.099~0.148(0.127±0.018)mg/L	0.009mg/L ↑
도암만	0.026~0.096(0.053±0.038)mg/L	0.008mg/L ↑
득량만	0.007~0.137(0.042±0.050)mg/L	0.011mg/L ↑
고흥	0.046~0.101(0.079±0.026)mg/L	0.029mg/L ↑

자료: 국립수산물과학원, 어장환경모니터링 통계검색

〈그림 2-17〉 남서해권 용존무기질소(DIN) 분포도



자료: 국립수산물과학원(2021a), p. 13

김 양식장이 많이 분포해 있는 목포, 진도, 완도 등의 연안에서는 표층 DIP 농도와 마찬가지로 DIN 농도도 높게 나타났다.

(3) 동해권(부산)

동해권 김 양식 어장의 경우 시설량과 생산량이 많지 않다. 동해권 어장 중 김이 생산되는 곳은 부산으로, 기초 행정구역으로는 강서구와 사하구가 해당된다. 동해권 양식어장 환경모니터링 거점 중 김 양식이 이루어지는 곳은 부산, 낙동강 하구 두 곳이다.

가. 수온

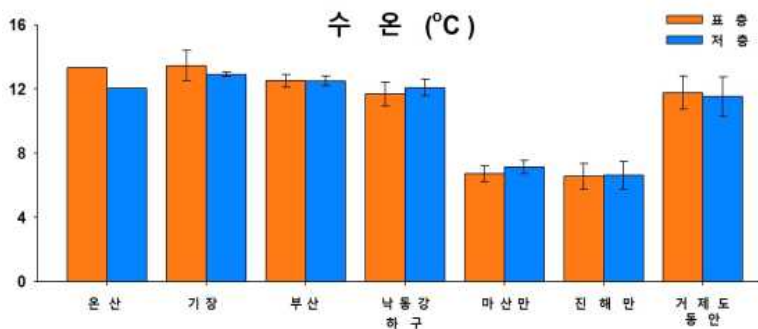
아래의 표는 2021년 2월 기준, 동해권 해역의 수온 분포를 나타낸 것이다. 연안 표층 수온은 부산의 경우 12.23~12.96℃이고 평균 12.51℃였다.

〈표 2-13〉 동해권 김 양식 어장 표층 수온(2021. 2. 기준)

지 역	수 온	평균 수온
부 산	12.23~12.96℃	12.51℃
낙동강 하구	10.88~12.36℃	11.68℃

자료: 국립수산과학원, 어장환경모니터링 통계검색(검색일: 2021. 9. 23)

〈그림 2-18〉 동해권 수온 분포도



자료: 국립수산과학원(2021b), p. 3

앞서 서해권, 남서해권 수온과 비교했을 때 겨울철임에도 수온이 매우 높은 것을 확인할 수 있다. 낙동강 하구의 2월 수온은 부산 연안보다는 다소 낮은 10.88~12.36℃이었고, 평균 11.68℃를 형성했다.

나. 용존무기인(DIP)

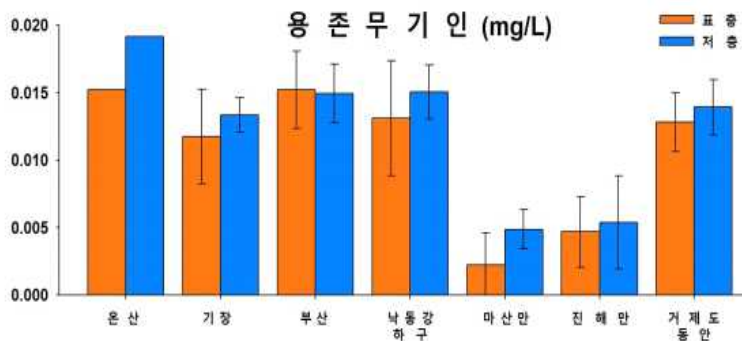
2021년 2월 기준 동해권 김 양식장의 DIP 수치를 살펴본 결과, 대체로 평년보다 높았다. 김 양식 적지인 남서해권의 DIP 농도가 높은 것과 마찬가지로 동해권에서 김 양식이 이루어지는 부산 연안, 낙동강 하구의 DIP 농도는 평균 0.013, 0.015mg/L로 비교적 높았다.

〈표 2-14〉 동해권 김 양식 어장 용존무기인(DIP) (2021. 2. 기준)

지 역	용존무기인(DIP)	평균 농도
부 산	0.012~0.01mg/L	0.015mg/L ↑
낙동강 하구	0.010~0.01mg/L	0.013mg/L ↑

자료: 국립수산과학원, 어장환경모니터링 통계검색(검색일: 2021. 9. 23)

〈그림 2-19〉 동해권 용존무기인(DIP) 분포도



자료: 국립수산과학원(2021b), p. 7

다. 용존무기질소(DIN)

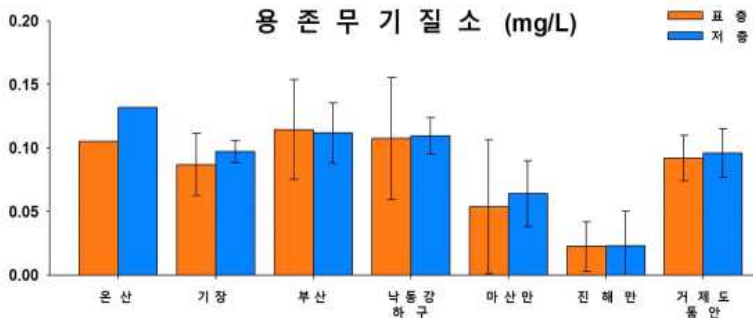
동해권 김 양식장의 용존무기질소(DIN)도 앞서 DIP와 마찬가지로 남서 해권만큼 높게 나타났다. 2021년 2월 기준 동해권 부산 연안의 DIN 농도는 0.074~0.15mg/L로 평균 0.114mg/L였고, 낙동강 하구의 농도는 0.078~0.16mg/L이며 평균 0.107mg/L로 조사되었다.

〈표 2-15〉 동해권 김 양식 어장 용존무기질소(DIN) (2021. 2. 기준)

지 역	용존무기질소(DIN)	평균 농도
부 산	0.074~0.15mg/L	0.114mg/L ↑
낙동강 하구	0.078~0.16mg/L	0.107mg/L ↑

자료: 국립수산과학원, 어장환경모니터링 통계검색(검색일: 2021. 9. 23)

〈그림 2-20〉 동해권 용존무기질소(DIN) 분포도



자료: 국립수산과학원(2021b), p. 13

김은 우리나라 서해, 남해, 동해에서 모두 생산되고 있지만, 대부분 연안이 김 양식 적지인 전남을 제외하고는 수온이 김 생장에 적합하고 영양염이 풍부한 곳에서 주로 김 생산이 이루어지고 있다.

3) 김 작황 감소 유형

앞서 김 작황, 즉 단수에 영향을 미치는 요인들을 살펴보았다. 해수에 직접적인 영향을 미치는 요인으로 수온, 영양염 등이 있고, 이 외에 햇빛, 풍파, 조류, 기타 부착생물 등이 매 어기 김 생산 단수에 영향을 준다. 이렇듯 여러 환경 요인 중 일부만 적정 수준을 유지하지 못해도 김 생장은 부진해진다. 지금부터는 앞서 살펴본 요인들이 김 생장에 부정적인 영향을 미칠 때 발생하는 몇 가지 작황 감소 유형에 대해 살펴보려고 한다.

(1) 영양염 부족

가장 대표적인 김 단수 감소 유형은 영양염 부족이다. 김 생장에 필수적인 영양염은 위에서 언급한 DIN, DIP 이외에도 다양한 종류가 해수에 존재한다. 이러한 영양염이 폭우, 담수 유입 등의 원인에 의해 부족해질 수 있는데, 영양염이 부족하면 김 생장이 더뎈다.

〈표 2-16〉 2017~18년산 해조류 양식피해 조사 목록

	지역	수 온	평년 대비
1	2016. 11. 4.	완도군 주요 미역 양식장 초기 업체 고사	고수온
2	2016. 11. 15.	고흥군 김 업체 선택 이상 및 생장 부진	영양염 부족
3	2016. 12. 7.	고흥군 해조류 피해 발생 합동조사	영양염 부족
4	2016. 12. 20.	해남군 해조류(김) 피해 발생 합동조사	영양염 부족
5	2017. 1. 10.	기장군 다시마 양식장 종자 폐사 원인 조사	고수온 등
6	2017. 2. 28.	울주군 양식 미역 피해 원인 규명 분석	고수온
7	2017. 3. 28.	세월호 관련 유류피해 조사	유류유출
8	2017. 9. 21.	양식생물 질병 모니터링 및 폐사 조사	영양염 부족
9	2017. 10. 20.	김 양식장 환경 조사(고흥, 해남, 완도, 진도)	영양염 부족

자료: 국립수산물과학원 내부자료인 '주요 양식품종 모니터링' 보고서를 토대로 저자가 재작성함

실제 2016~2017년 국립수산물과학원에서 실시한 ‘주요 양식품종 모니터링’ 자료에 따르면, 영양염 부족에 의한 피해 상황 조사가 전체 조사 분석 건수 9개 중 절반 이상인 5개에 해당하는 것으로 나타났다. 영양염 부족은 넓은 해역에 걸쳐 일회성에 그치지 않고 빈번히 발생한다는 점에서 김 작황을 변동시키는 가장 중요한 변수라고 할 수 있다. 또한 영양염 부족은 단순히 생장만을 지연시키는 것이 아니라 각종 질병 감염의 주요 원인이 된다는 점에서 더 큰 문제가 있다. 김 시설 초기 영양염이 결핍되면 김 엽체 색이 노랗게 변하는 황백화가 발생한다.

(2) 질병 감염

김 시설이 바다에 설치되어 있는 9월~4월 동안 김 엽체는 바다로부터 오는 다양한 질병에 노출된다. 앞서 언급한 황백화도 영양염 부족에 따른 질병 중 하나로 볼 수 있으며 갯병, 쪼그랑병, 녹반병 등 다양한 질병 위험이 김 양식에 존재한다.

양식장에서 나타나는 질병은 흔히 갯병이라고 한다. 갯병은 병원균의 침입으로 인해 나타나는 기생성 질병, 그리고 생리적인 장애로 인해 나타나는 생리적 갯병으로 나눌 수 있는데, 주된 발생 원인은 김의 양식이 이루어지는 기간 중 수온 상승, 조류 소통 문제, 밀식 등과 밀접한 관련이 있다. 특정 시기가 아닌 김 양식 기간 내내 발생 위험이 있으며 김 양식 초기 고수온기에는 싹갯병, 붉은갯병 등이 발생하기 쉽고, 이와 같은 갯병은 김 작황 부진에 막대한 영향을 미친다.

쪼그랑병은 갯병만큼 작황에 큰 영향을 미치는 것은 아니지만, 과거부터 계속 있어온 질병 중 하나다. 성엽의 엽체 양면에 돌기가 생기고, 증상이 심해지면 엽체에 섬세한 견직물상의 주름이 생기면서 흑색으로 변하여 광택을 잃게 된다.

다음으로 녹반병은 김 업체에 여러 형태의 병반부가 생겨 이 병반부가 최후에 유실되는 질병으로, 김 양식 기간 전반에 걸쳐 나타날 수 있으나 여기 초 11~12월에 피해가 가장 심하다. 녹반병 감염은 김 수확량과 품질 저하에 영향을 미친다.

해황 변동으로 김이 성장하는 동안 어떤 장애를 겪게 되면 김 업체가 약해질 수밖에 없고 이때 병원균 등의 침입을 받아 질병이 발생한다. 이 중 가장 흔하게 발생하는 붉은갯병의 주 발생 시기는 11~12월, 그리고 양식이 거의 끝나갈 무렵인 2~3월이다. 수온이 느리게 내려간 해에 갯병이 많이 발생한다는 속설이 있는데, 이처럼 김 업체 성장 단계의 적정 수온과 질병 사이에 밀접한 연관이 있음이 과거 사례를 통해 관찰되었다. 채묘 온도에 해당하는 시기인 수온 22℃가 되는 날부터 김 생장이 활발해지는 13℃ 이하로 수온이 내려가기까지 걸리는 기간이 55일 이상 되는 해에는 김 작황이 좋지 못했고, 이 기간이 60일을 넘어갔던 1985년에는 김 작황 부진이 매우 뚜렷하게 발생했다.

(3) 해적생물 부착

김 작황 감소 유형의 마지막은 해적생물 부착이다. 해상에 시설한 김 양식 밭에 붙어 김 성장을 방해하는 해적생물로는 파래, 매생이, 따개비, 규조류 등이 있으며 최근에는 해상에 떠다니는 팽생이모자반도 새로운 위협 요인으로 나타나고 있다.

파래는 수온이 22.5℃ 이상일 때 채묘를 한 경우, 또는 조류 소통이 나쁜 곳에서 주로 발생하기 때문에 밀식이 원인이 될 수 있다. 매생이는 수온이 18℃ 이하로 내려가면 김밭에 부착이 시작되는데, 부착 수위가 있으므로 매생이가 부착됐을 경우에는 밭의 수위를 내려서 며칠간 고정해 두는 것이 좋다. 규조류는 김 업체에 착생하여 김의 성장을 방해하며 세포를 파괴하

기도 해서 마른김으로 건조시켰을 때 김 품질이 나빠지게 하는 해적생물 중 하나이다. 규조류 발생 또한 조류 소통이 나쁘고 해수 비중이 낮은 해에 많이 발생하는 특징이 있다. 이와 같은 해적생물 부착을 예방하기 위해서는 적절한 시기에 채묘를 하고, 밀식을 지양하며 해황 여건에 따라 김발의 설치 수위를 다르게 하는 등 지속적인 양식장 관리가 필요하다.

그리고 최근 몇 년 사이에는 김 양식이 이루어지는 겨울~봄철에 중국에서 팽생이모자반이 해류를 타고 우리나라 남서해권, 즉 전남지역 양식장으로 떠내려오는 상황이 종종 발생했다. 팽생이모자반은 넓은 해역에 걸쳐 큰 덩어리로 몰려다니며 김발에 붙어 김 생장을 방해한다. 그뿐만 아니라 김 채취 작업 시기에 발생하면 생산자들이 김 채취 대신 팽생이모자반 제거 작업을 해야 하므로 김 생산 단수 감소에 막대한 영향을 미치게 된다. 팽생이모자반은 그 해의 수온, 풍향 등에 따라 발생 및 피해 규모가 결정되며, 김 양식장이 밀집한 해역으로 유입되기 전에 해상에서 선박 등을 이용해 제거하는 것이 중요하다.

03

인공신경망모형의 이론적 고찰

제3장에서는 본 연구에서 채택한 분석 기법인 인공신경망모형의 이론에 대해 고찰하였다. 제1절에서는 인공신경망모형의 기초 개념과 주요 인공신경망모형의 종류를 알아보았다. 제2절에서는 인공신경망모형의 학습에 관한 내용을 다루었으며, 이는 머신러닝을 위한 주요 알고리즘과 학습을 위한 설정 조건들을 포함한다.

제1절 인공신경망모형 개요

1. 인공신경망모형

인공신경망(artificial neural network)은 인공지능의 구현을 위해 인간의 뇌에서 일어나는 신경세포(뉴런)의 네트워크에서 나타나는 작용을 흉내 낸 모형이다. 뉴런은 세포핵이 있는 신경세포체, 다른 세포로부터 신호를 받아들이는 수상돌기, 다른 세포에 신호를 전달하는 축삭돌기로 구성된다. 다른 세포들로부터 신호를 전달받은 뉴런은 전달된 신호의 강도가 일정 수준 이상이 되면 다른 뉴런으로 신호를 전달하고, 그렇지 않을 때는 무시하

는 것으로 알려져 있다. 인간의 뇌는 이와 같은 뉴런의 네트워크에서 발생하는 신호 흐름에 따라 다양한 정보를 처리하게 된다.¹¹⁾

인공신경망에서는 위와 유사한 방식으로 단일 또는 복수의 은닉층에 존재하는 퍼셉트론(은닉노드)과 활성화함수라는 구성 요소를 이용해 수학적으로 정보 처리 과정을 표현한다. 퍼셉트론에서는 데이터(변수)를 입력받아 회귀분석모형의 연산과 유사하게 편향(bias, 상수)과 각 변수에 부여된 가중치를 이용해 출력값을 계산한다. 이때 출력값은 활성화함수에 따라 뒤쪽에 있는 퍼셉트론에 얼마나 전달될지가 결정된다. 각 퍼셉트론에서 이와 같은 연산이 끝나면 출력값은 그 다음 은닉층에 있는 퍼셉트론에 전달되며, 이 같은 과정을 반복한 끝에 최종 출력값이 결정된다.

회귀모형의 경우 비교적 단순한 구조를 가지므로 행렬 연산 등의 방법을 이용해 단시간 내에 모수(상수와 계수) 추정이 가능하지만, 인공신경망의 경우 고도로 복잡한 구조로 인해 회귀모형과 같은 방식의 추정을 해내기는 사실상 불가능에 가깝다. 따라서 컴퓨터를 이용해 임의의 가중치를 부여한 후 모형을 통해 예측된 값과 관측치의 오차를 이용해 지속적으로 가중치를 갱신하는 머신러닝 알고리즘으로 최적의 가중치를 구한다. 머신러닝은 기계, 즉 컴퓨터가 스스로 학습을 통해 미가공된 자료의 특징을 파악하고 지식을 취하는 것으로,¹²⁾ 깊은 층의 인공신경망을 학습하는 머신러닝 알고리즘을 딥러닝이라고 한다.

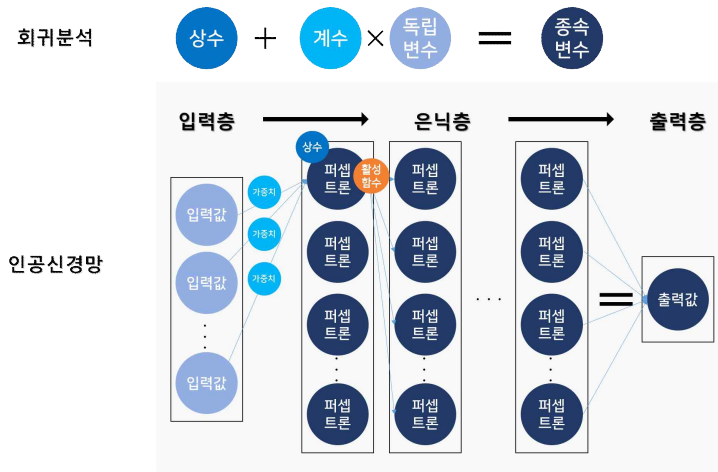
일반적으로 회귀분석에서는 독립변수와 종속변수의 선형 상관관계를 함수로 표현한다. 또한 이론적인 근거에 따라 독립변수와 종속변수를 가공함으로써 선형관계뿐만 아니라 고차함수, 로그함수 등 비선형적 관계의 표현도 가능하다. 하지만 회귀분석의 기본적인 구조는 선형함수이며 가설을 검

11) 와쿠이 요시유키와쿠이 사다미 지음, 박광수 옮김(2019), p. 24.

12) 이안 굿펠로·요슈아 벤지오·에런 쿠빌 지음, 류광 옮김(2020), p. 3.

정하기 위해 행해지는 분석인 만큼 철저하게 이론적으로 검증된 규칙적인 관계를 표현하는 데 그친다. 하지만 인공신경망모형은 <그림 3-1>과 같이 입력층의 크기(독립변수의 개수)와 상관없이 은닉층의 요소를 통해 많은 수의 고차원적이고 복잡한 비선형 관계를 표현하는 것이 가능하다.

<그림 3-1> 인공신경망모형의 구조



자료: 저자 작성

2. 인공신경망모형의 종류¹³⁾

1) 다층 퍼셉트론(Multi-layer Perceptron, MLP)

인공신경망모형의 가장 기본적인 형태는 여러 심층 신경망모형의 기초가 되는 다층퍼셉트론(MLP)이다. 퍼셉트론은 신경망을 구성하는 기본 요소로, 주어진 N 차원의 입력값 $x \in \mathbb{R}^N$ 에 대해 [식 3-1]과 같은 과정을 거쳐 출력값을 산출한다.

13) 이 내용은 본 과제의 자문보고서(김동우, 2021) 내용을 참고하여 작성하였다.

$$f(x) = \sigma(w^\top x + b) \quad [\text{식 3-1}]$$

여기서 $w \in \mathbb{R}^N$, b 는 각각 퍼셉트론의 가중치와 편향으로 이루어진 파라미터로 선형함수를 구성하며, 이는 비선형 활성화함수 σ 에 의해 퍼셉트론의 출력값이 결정된다. 다층 퍼셉트론모형은 복수의 퍼셉트론으로 구성된 여러 개의 은닉층으로 구성된다. 만약 첫 번째 은닉층에서 입력값 x 를 받는 k 개의 퍼셉트론을 갖는다고 할 때 이것은 아래와 같이 표현할 수 있다.

$$\begin{aligned} h_1 &= \sigma(w_1^\top x + b_1), \\ h_2 &= \sigma(w_2^\top x + b_2), \\ &\vdots \\ h_k &= \sigma(w_k^\top x + b_k), \end{aligned} \quad [\text{식 3-2}]$$

여기서 $w_1, \dots, w_k$ 는 각 입력값 x 에 대응하는 행렬로, $b_1, \dots, b_k$ 는 각 퍼셉트론의 선형 연산에 대응하는 벡터로 볼 수 있으므로 아래와 같이 행렬 W 와 벡터 B 로 간략하게 나타낼 수 있다.

$$H = \sigma(WX + B) \quad [\text{식 3-3}]$$

이와 같은 퍼셉트론으로 이루어진 은닉층을 여러 개로 구성한 것이 다층 퍼셉트론이다. 첫 번째 은닉층의 입력이 X 일 때 결과값을 $H^{(1)}$ 이라고 하면 이것은 두 번째 계층에서는 첫 번째의 X 대신 $H^{(1)}$ 이 퍼셉트론의 입력값이 되며, k 개의 층을 가진 모형이라면 이는 아래와 같이 나타낼 수 있다.

$$\begin{aligned} H^{(1)} &= \sigma(W^{(0)}X + B^{(0)}), \\ H^{(2)} &= \sigma(W^{(1)}H^{(1)} + B^{(1)}) \\ &\vdots \\ H^{(k)} &= \sigma(W^{(k)}H^{(k-1)} + B^{(k)}) \end{aligned} \quad [\text{식 3-4}]$$

여기서 활성화함수의 존재는 신경망모형에 비선형성을 부여하는 가장 중요한 요소이다. 만약 3계층의 다층 퍼셉트론을 [식 3-5]와 같이 정의하고 각 계층의 잠재 유닛의 수를 K_1, K_2, K_3 이라고 하면 각 변수 행렬 W 의 차원은 $W^{(1)} \in \mathbb{R}^{K_1 \times N}, W^{(2)} \in \mathbb{R}^{K_2 \times K_1}, W^{(3)} \in \mathbb{R}^{K_3 \times K_2}$, 이다.

$$\begin{aligned} H^{(1)} &= \sigma(W^{(1)}X), \\ H^{(2)} &= \sigma(W^{(2)}H^{(1)}) \\ H^{(3)} &= \sigma(W^{(3)}H^{(2)}) \end{aligned} \quad [\text{식 3-5}]$$

여기서 활성화함수를 없애면 위의 모형은 아래와 같이 나타낼 수 있다.

$$\begin{aligned} H^{(1)} &= W^{(1)}X \\ H^{(2)} &= W^{(2)}W^{(1)}X \\ H^{(3)} &= W^{(3)}W^{(2)}W^{(1)}X \end{aligned} \quad [\text{식 3-6}]$$

즉 $H^{(3)} = \mathbf{W}X$, $\mathbf{W} = W^{(1)} \times W^{(2)} \times W^{(3)}$ 이므로 이를 통해 인공신경망모형에서 은닉층의 개수를 아무리 많이 설정하더라도 활성화함수가 없다면 선형 모델로 나타낼 수밖에 없다는 것을 알 수 있다.

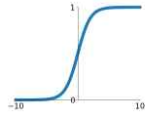
따라서 활성화함수가 어떤 형태인가에 따라 독립변수와 종속변수 간 상관관계의 형태가 결정되므로 어떤 활성화함수를 선택하느냐가 인공신경망모형의 설정에 매우 중요한 요소라고 할 수 있다. <그림 3-2>는 여섯 가지의 대표적인 활성화함수와, 그것을 그래프로 나타낸 모양이다.

우선 인공신경망모형에서 가장 오랫동안 보편적으로 사용된 Sigmoid 함수는 뉴런의 신경학적인 활성 패턴과 유사한 모양으로, 실수인 입력값을 0과 1사이의 출력으로 변환하는 역할을 한다.

〈그림 3-2〉 활성화함수의 종류

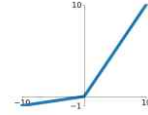
Sigmoid

$$\sigma(x) = \frac{1}{1+e^{-x}}$$



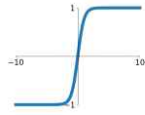
Leaky ReLU

$$\max(0.1x, x)$$



tanh

$$\tanh(x)$$

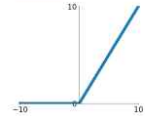


Maxout

$$\max(w_1^T x + b_1, w_2^T x + b_2)$$

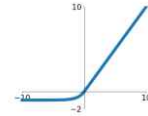
ReLU

$$\max(0, x)$$



ELU

$$\begin{cases} x & x \geq 0 \\ \alpha(e^x - 1) & x < 0 \end{cases}$$



자료: CS231n: Convolutional Neural Networks for Visual Recognition(검색일: 2021. 9. 24)

Sigmoid 함수는 아래의 [식 3-7]과 같이 계산되며 그 결과값이 0과 1 사이로 한정된다. 따라서 확률적으로 해석이 가능하다는 장점을 가지지만 다층의 신경망 학습에는 불리한 특징을 가진다.

$$\sigma(x) = \frac{1}{1 + \exp(x)}$$

[식 3-7]

다음으로 tanh 함수는 쌍곡탄젠트(hyperbolic tangent, tanh)라고 불리며, 그래프로 나타낸 모양은 Sigmoid 함수와 유사한 반환되는 출력값은 -1에서 1사이에서 결정된다. [식 3-8]과 같이 나타낼 수 있으며 sigmoid 함수와 마찬가지로 다계층 신경망 학습에는 불리하다.

$$\sigma(x) = \frac{\exp(x) - \exp(-x)}{\exp(x) + \exp(-x)} \quad [\text{식 3-8}]$$

앞의 두 함수는 출력값의 범위가 한정되어있지만 ReLU(Rectified Linear Unit) 함수의 경우 [식 3-9]와 같이 0보다 작은 입력값은 모두 0으로 제한하고, 0보다 큰 경우 변화 없이 그대로 출력하는 형태의 함수이다. 이는 1960년대에 처음 제안된 이후 크게 활용되지 않았으나 2011년 Glorot, Bordes, and Bengio(2011)¹⁴⁾에서 이것이 다층 신경망 학습에 유리하다는 사실이 밝혀지고 난 후 다양한 심층 신경망 모형에서 사용되고 있다.

$$\sigma(x) = \max(0, x) \quad [\text{식 3-9}]$$

ReLU 함수의 미분계수는 입력값의 범위에 따라 0 또는 1로 수렴한다는 특징이 있으며 지수함수가 포함되어 있지 않아 지수함수를 포함한 모형에 비해 더욱 빠른 학습이 가능하다는 장점을 가진다.

Leaky ReLU 함수는 <그림 3-2>에서와 같이 ReLU 함수와 유사한 모양을 보이지만 입력값이 음수일 경우를 0이 아닌 αx 로 변경한 형태이다. α 에는 일반적으로 1보다 작은 상수값을 사용하게 되는데, 이로 인해 음의 입력값에 대한 미분계수는 ReLU 함수와 달리 0이 아니다.

$$\sigma(x) = \max(\alpha x, x) \quad [\text{식 3-10}]$$

ELU(Exponential Linear Unit) 함수 역시 ReLU, Leaky ReLU 함수와 유사한 형태로 나타나지만 음의 입력값에 대한 출력이 지수 형태로 [식

14) Glorot, Bordes, and Bengio(2011), pp. 315-323.

3-11]과 같이 표현된다. 함수의 α 는 입력값이 음수일 경우 얼마나 빠르게 -1에 수렴하는지를 결정하게 된다.

$$\sigma(x) = \begin{cases} x & x \geq 0 \\ \alpha(\exp(x) - 1) & x < 0 \end{cases} \quad [\text{식 3-11}]$$

마지막으로 Maxout 함수는 앞의 활성화함수들과는 다르게 두 개의 뉴런으로부터 입력받은 값 중 상위의 값을 선택하는 함수이다. 이 경우 다른 함수에 비해 훨씬 더 다양한 관계를 표현하는 것이 가능하지만 다른 함수보다 두 배 많은 변수가 필요하므로 학습 속도가 느려진다는 단점이 있다.

$$\sigma(x) = \max(w_1^\top x, w_2^\top x) \quad [\text{식 3-12}]$$

한편 보편 근사 정리(Universal Approximation Theorem)¹⁵⁾에 따르면 2계층의 sigmoid 활성화함수를 사용한 퍼셉트론모형에서 충분한 수의 잠재변수를 사용한다면 어떠한 연속함수에 대해서도 오차 이내로 근사할 수 있다. 다시 말해 이는 신경망 모형은 어떠한 관계든지 함수로 표현할 수 있다는 근거를 제공하지만, 이는 어디까지나 모형의 표현력에 대한 것으로 실제 학습을 통해 원하는 함수를 얼마나 근사할 수 있을지에 대해서는 추가적인 증명이 필요하다.

2) 순환신경망(RNN: Recurrent neural network)

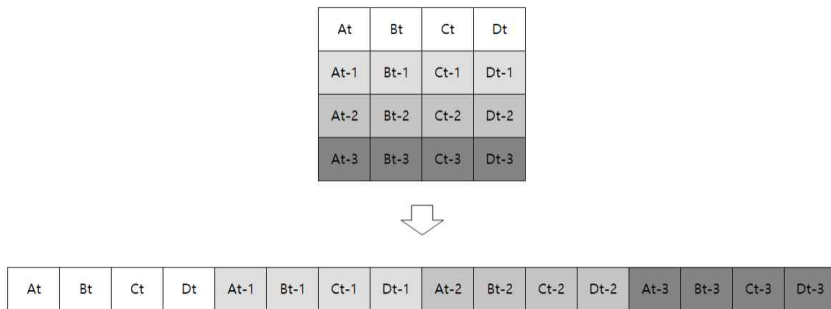
순환신경망모형은 순차 혹은 시계열 데이터를 처리하는 데 있어 데이터

15) Hornik, Stinchcombe, and White(1989), pp. 359-366.

의 순서 또는 시간에 해당하는 정보를 반영할 수 있도록 설계된 신경망모형이다. 앞서 살펴본 다층 퍼셉트론모형에서와는 다르게 입력값이 되는 데이터가 벡터가 아닌 행렬 형태로 입력하여 순서와 시간을 고려하여 연산을 수행한다.

다층 퍼셉트론모형의 순차데이터 학습을 위해서는 아래의 <그림 3-3>과 같이 행렬 데이터를 벡터 형태로 전환시켜야 한다. 이 경우 N 차원의 데이터 T 개가 $N \times T$ 차원으로 변하면서 기존의 행렬 형태에서 가지고 있던 데이터의 순서, 또는 시간에 대한 정보가 소실된다. 또한 벡터로 변환되며 늘어난 데이터의 길이만큼 첫 번째 은닉층에서 구해야 할 파라미터의 차원도 같이 길어진다. 예를 들어 아래의 그림과 같이 4×4 행렬의 입력값이 주어질 경우 첫 번째 계층의 각 퍼셉트론에 대응하는 가중치 w 의 수는 4개가 아닌 16개가 되므로, 첫 번째 층 전체의 퍼셉트론이 k 개라고 할 경우 파라미터 행렬의 차원은 $N \cdot T \times k$ 가 된다.

<그림 3-3> 다층 퍼셉트론모형 학습을 위한 행렬데이터 변환

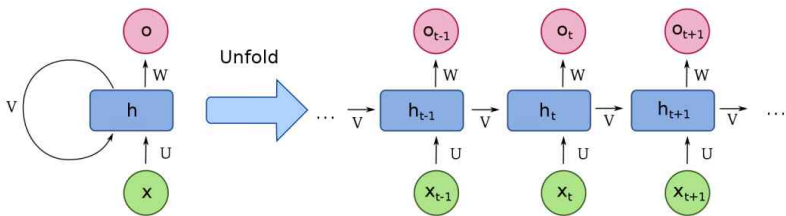


자료: 본 과제의 자문보고서(김동우, 2021)를 바탕으로 하여 저자 작성

이런 식의 구조가 가지는 문제점 중 하나는 입력 데이터의 길이가 바뀌면 이것을 처리하는 데 문제가 생긴다는 점이다. 만약 텍스트 데이터를 입력값으로 받는 경우, 문장의 구성에 따라 문장을 구성하는 단어의 수가 달

라지게 된다. 그런데 다층 퍼셉트론모형에서는 문장의 길이가 변할 경우 입력값의 처리를 위해 일정한 크기로 문장을 잘라서 재배열하는 등의 전처리가 필요하며, 이로 인해 정보 손실이 일어나게 된다. 위와 같이 다층 퍼셉트론모형에서 순차 또는 시계열 데이터를 학습할 때 발생하는 문제를 해결하기 위해 개발된 것이 순환신경망 모형이다.

〈그림 3-4〉 순환신경망의 구조



자료: 위키백과(검색일: 2021. 9. 24)

순환신경망은 변수 공유(parameter sharing) 기법을 사용해 다양한 길이의 입력 데이터를 처리하여 예측 문제를 해결하는 모형이다. 순차데이터 x 가 주어졌을 때 순환신경망은 [식 3-13]과 같이 정의된다.

$$\begin{aligned} h_t &= \sigma_h(Ux_t + Vh_{t-1} + b_h), \\ o_t &= \sigma_y(Wh_t + b_o) \end{aligned} \quad [\text{식 3-13}]$$

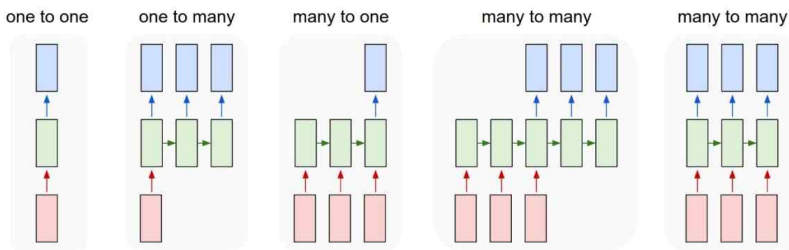
여기서 h_t 는 입력시간(순서) t 에서 현재 데이터의 상태를 나타내는 상태 벡터이며 o_t 는 최종 출력값이다. 상태 벡터 h_t 를 구하기 위해서는 이전 시간의 데이터를 입력받아 계산한 h_{t-1} 과 현재의 입력값 x_t 를 선형(U , V) 및 비선형(σ) 변환하는 과정을 거친다. 여기서 변수 행렬인 U , V , W 는 입력시간 t 에 관계없이 동일한 값이 사용된다는 특징이 있는데, 이를 변수 공유라고 한다. 만약 t 시점의 예측값을 도출하고자 하는 경우 상태 벡터 h_t 의 선

형-비선형 변환을 통해 예측값인 o_t 를 출력하는 데 사용된다.

상태 벡터 차원의 크기는 변수 행렬 U , V , W 의 차원 크기를 결정하게 되는데, 상태 벡터 h_t 가 K 차원, 출력값 o 가 O 차원이라고 가정하면 각 변수 행렬의 차원은 $U \in \mathbb{R}^{K \times D}$, $V \in \mathbb{R}^{K \times K}$, $W \in \mathbb{R}^{K \times O}$, $b_h \in \mathbb{R}^K$, $b_o \in \mathbb{R}^O$ 로 정의할 수 있다.

순환신경망의 이 같은 구조는 다양한 순차 및 시계열 데이터를 다룰 때 사용된다. 아래의 <그림 3-5>와 같이 입력과 출력 형태에 따라 일대일, 일대다, 다대일, 다대다 구조로 다양하게 나눌 수 있다.

<그림 3-5> 순환신경망의 입출력 형태



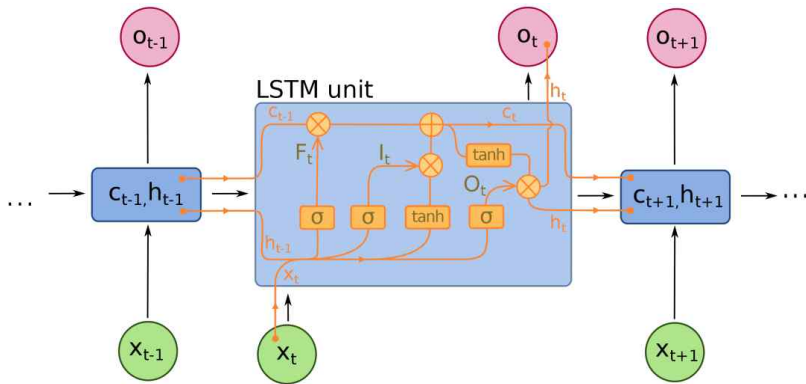
자료: Andrej Karpathy blog(2015. 5. 20)

3) 장단기 기억신경망(LSTM: Long short-term memory)

장단기 기억신경망(LSTM)은 순환신경망이 가지는 문제점을 해결하기 위해 개발된 모형이다. 순환신경망모형의 문제점으로 지적되는 것 중 하나는 순차데이터의 길이가 길어질 경우 발생한다. 특정 시점의 정보가 다른 시점에 비해 중요한 단서를 가지거나, 그 반대의 경우가 있을 수 있는데 기본적인 순환신경망 구조에서는 시간이 지남에 따라 벡터의 성분이 변하도록 설계되어있어 정보가 소실될 가능성이 크다는 것이다.

예를 들어 시간 t 의 예측을 위해 h_1 의 정보가 매우 중요할 수 있는데, 이것이 과거의 정보이고 입력 시계열의 길이가 매우 길다면 순환신경망의 연산 과정 중에서 h_1 의 정보가 필요한 만큼 보존될 확률이 희박해진다. LSTM은 이러한 문제를 해결하기 위해 기본적인 순환신경망 구조에 입력(input), 출력(output) 및 소거(forget) 모듈을 추가한 구조를 가지며, 이 세 가지 핵심 요소를 통해 두 가지 종류의 상태 벡터 c , h (장기/단기 기억)를 가진다.

〈그림 3-6〉 장단기 기억신경망의 구조



자료: Medium(2019. 7. 23)

LSTM모형에서 t 시점의 장기/단기 기억을 결정하는 c_t 와 h_t 는 [식 3-14]와 같이 나타낼 수 있다.

$$\begin{aligned} c_t &= f_t \circ c_{t-1} + i_t \circ \tilde{c}_t, \\ h_t &= o_t \circ \sigma_h(c_t) \end{aligned} \quad [\text{식 3-14}]$$

여기서 σ_h 는 tanh 또는 항등함수를 사용하며, i , o , f 는 각각 입력, 출력, 소거 게이트로 아래와 같은 계산 과정을 거친다. 입력 게이트는 [식

3-15]와 같으며 단기 기억 메모리와 입력값을 받아 장기 기억 메모리의 어떤 부분을 수정할지 결정한다.

$$i_t = \sigma_g(W_i x_t + U_i h_{t-1} + b_i) \quad [\text{식 3-15}]$$

출력게이트는 [식 3-16]과 같으며 여기서는 단기 기억 메모리와 입력값을 이용해 장기 기억 메모리의 어떤 부분을 이용해 단기 기억 메모리를 재구성 할지를 결정하게 된다.

$$o_t = \sigma_g(W_o x_t + U_o h_{t-1} + b_o) \quad [\text{식 3-16}]$$

소거게이트는 [식 3-17]과 같이 나타낼 수 있으며 단기 기억 메모리와 입력값으로 장기 기억 메모리의 어떤 부분을 제거할지 결정하게 된다.

$$f_t = \sigma_g(W_f x_t + U_f h_{t-1} + b_f) \quad [\text{식 3-17}]$$

입력/출력/소거게이트의 활성화함수로는 sigmoid 함수가 사용되며, 결과가 1에 가까울수록 입력/출력/소거 연산을 수행하는 쪽에 가까워진다. 최종적으로 선택된 장기 기억 메모리들은 [식 3-18] 같은 연산을 거쳐 갱신된다. 이때 활성화함수는 tanh 함수가 사용된다. 이는 sigmoid 함수의 경우 출력값이 0과 1사이로 한정되어 장기 기억 메모리 내부 값의 감소를 표현할 수 없지만, tanh 함수는 -1에서 1사이의 값을 반환하므로 감소에 대한 표현이 가능하기 때문이다.

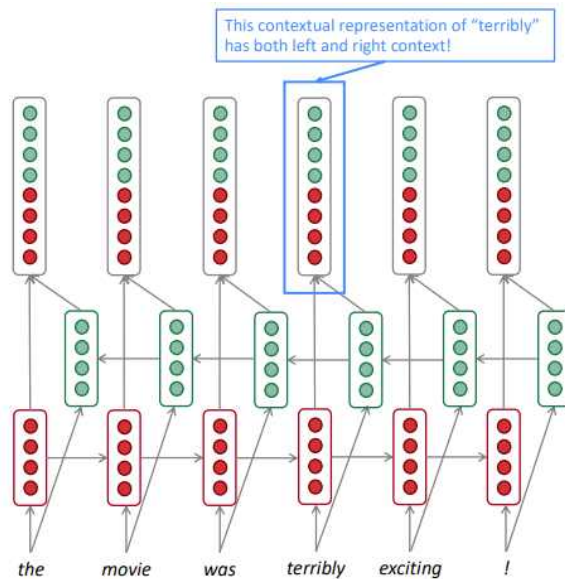
$$\tilde{c}_t = \sigma_c(W_c x_t + U_c h_{t-1} + b_c)$$

[식 3-18]

4) 양방향 순환신경망

앞서 살펴본 RNN 및 LSTM 모형은 시계열 데이터를 시간 순서에 따라 처리하는 구조를 가진다. 그러나 양방향 순환신경망은 아래의 그림과 같이 순환신경망을 이중으로 구성하여 하나의 순환신경망에서는 시계열 데이터의 연산을 순방향으로 처리하고, 남은 하나에서는 역방향으로 뒤에서부터 앞으로 데이터를 처리한다. 최종적으로는 양방향의 산출 결과 벡터를 합쳐 출력값을 결정하게 된다.

〈그림 3-7〉 양방향 순환신경망의 구조



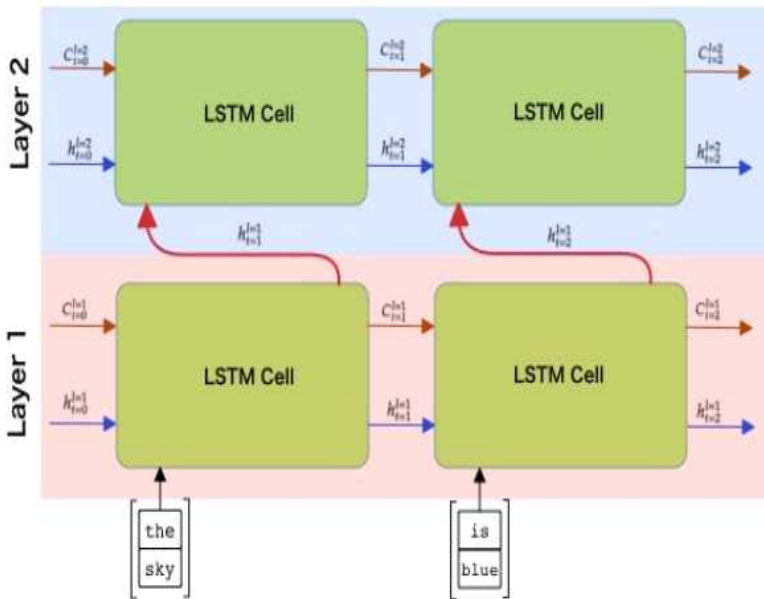
자료: Abigail(2019), p.37

5) 심층신경망(DNN: Deep Neural Network,)

심층신경망은 연구자의 목적에 맞는 여러 층의 신경망을 하나의 손실함수 내에서 구성한 모형이다. 다층 퍼셉트론과 같이 RNN, LSTM 모형도 여러 개의 층을 설정하고 각 층별 잠재변수의 수를 설정함으로써 심층신경망을 구성할 수 있다.

신경망의 층수와 잠재변수를 늘릴수록 더욱 다양한 형태의 함수 표현이 가능하지만 그만큼 많은 양의 파라미터를 추정해야 하므로 학습 시간이 오래 걸리고 과적합이 더 많이 나타나는 등의 어려움이 발생한다.

〈그림 3-8〉 심층신경망의 구조



자료: towards data science(2020. 7. 27)

제2절 인공신경망모형의 학습

1. 손실함수

제1절에서 살펴본 인공신경망모형은 주어진 입력값에 대한 예측값을 출력하는 구조로 이루어져 있다. 인공신경망모형 내부의 편의와 가중치 등의 파라미터를 최적의 값으로 설정함으로써 모형을 통해 도출한 예측치와 실제 관측치 사이의 오차를 최소화할 수 있는데, 이때 학습(머신러닝)을 통해 반복적으로 파라미터(가중치, 편향 등)를 갱신하는 과정이 필요하다. 학습 과정에서 모형의 예측이 얼마나 정확한지를 정량적으로 나타내는 지표를 손실함수라고 하며, 대표적인 손실함수로는 평균제곱오차(MSE: Mean Squared Error)와 교차 엔트로피(Cross Entropy) 등이 있다.

관측치를 y_i , 모형을 통한 예측치를 $\hat{y}_i$ 라고 할 때, 평균제곱오차(MSE)를 활용한 손실함수는 [식 3-19]와 같이 정의할 수 있다.

$$L = \sum_{i=1}^N l(y_i, \hat{y}_i) = \sum_{i=1}^N (y_i - \hat{y}_i)^2 \quad [\text{식 3-19}]$$

변수 W 를 가지는 신경망모형에 대해 예측값 $\hat{y}_i$ 를 $fw(x_i)$ 라고 하면 학습의 목표는 아래의 수식과 같이 전체 손실값을 최소화하는 변수 W 를 찾는 것이다. 최적의 W 를 찾기 위한 가장 대표적인 방법은 오차역전파(error back-propagation) 기법이다.

$$\arg \min_W \sum_{i=1}^N (y_i - fw(x_i))^2 \quad [\text{식 3-20}]$$

2. 오차역전파와 경사 하강법

1) 경사 하강법

위의 [식 3-20]에 나타난 최적화 문제를 풀기 위해서는 다양한 방법이 있지만 신경망모형에서 가장 효율적으로 알려진 것은 오차역전파 알고리즘이다. 오차 역전파 알고리즘은 시간 t 에서의 변수 W_t 를 기준으로 하는 목적함수의 미분계수를 구하고, 미분계수의 반대 방향으로 변수를 변화시켜 최적의 값을 찾아가는 경사 하강법(*gradient decent*)을 이용한다.

$$\begin{aligned} W_{t+1} &= W_t - \lambda \frac{\partial L}{\partial W} \\ &= W_t - \lambda \frac{\partial \sum_{i=1}^N (y_i - fw(x_i))^2}{\partial W} \end{aligned} \quad [\text{식 3-21}]$$

여기서 λ 는 학습률(*learning rate*)이라고 하며, 가중치를 한 번 갱신할 때마다 얼마나 크게 변경시킬지를 결정한다. 예를 들어 이것을 3계층 다층 퍼셉트론 학습에 적용하기 위해서는 [식 3-22]와 같이 세 변수 행렬에 대한 미분계수를 계산해야 한다.

$$\frac{\partial L}{\partial W^{(1)}}, \frac{\partial L}{\partial W^{(2)}}, \frac{\partial L}{\partial W^{(3)}} \quad [\text{식 3-22}]$$

이 같은 방식으로 모든 변수에 대해 미분계수를 구하려면 깊은 신경망일 수록 많은 시간이 필요하다. 역전파 알고리즘은 [식 3-23]과 같이 미분의 연쇄법칙을 사용함으로써 계산량을 줄여 학습시간을 단축시킬 수 있다.

$$\frac{\partial L}{\partial W^{(1)}} = \frac{\partial L}{\partial h^{(3)}} \frac{\partial h^{(3)}}{\partial h^{(2)}} \frac{\partial h^{(2)}}{\partial h^{(1)}} \frac{\partial h^{(1)}}{\partial W^{(1)}} \quad [\text{식 3-23}]$$

여기서 편미분계수는 $\frac{\partial L}{\partial W^{(2)}}$ 의 계산을 위해 다시 사용될 수 있다.

$$\frac{\partial L}{\partial h^{(2)}} = \frac{\partial L}{\partial h^{(3)}} \frac{\partial h^{(3)}}{\partial h^{(2)}} \quad [\text{식 3-24}]$$

따라서 $W^{(1)}$, $W^{(2)}$, $W^{(3)}$ 의 계산을 위해서는 $\frac{\partial L}{\partial h^{(3)}}$, $\frac{\partial L}{\partial h^{(2)}}$, $\frac{\partial L}{\partial h^{(1)}}$ 를 계산해야 하며, 각각의 결과는 다음 계산을 위해 재사용될 수 있음을 알 수 있다. 입력값 x 에 대한 손실을 구하는 과정을 순전파(forward propagation), 손실값을 기준으로 역방향으로 계산하며 가중치를 갱신하는 과정을 역전파(backward propagation)라고 한다.

2) 확률적 경사 하강법

경사 하강법은 전체 데이터에 대한 미분계수를 계산해야 하므로 계산량이 커져 학습에 많은 시간이 걸린다. 확률적 경사 하강법(Stochastic gradient descent)은 변수의 갱신 과정에서 전체 데이터가 아닌 일부 데이터만 추려서 사용하는 방법으로, 정확한 확률계수가 아닌 확률계수의 기댓값을 구하는 방식이다. 이는 [식 3-25]와 같이 표현할 수 있는데, 여기서 S 는 무작위로 선택된 데이터의 크기(batch size)이며, (x_i, y_i) 는 i 번째 무작위로 선택된 데이터이다.

$$\begin{aligned}
 W_{t+1} &= W_t - \lambda NE \left[\frac{\partial L}{\partial W} \right] & [\text{식 3-25}] \\
 &= W_t - \lambda \frac{N}{S} \frac{\partial \sum_{i=1}^S (y_i - fw(x_i))^2}{\partial W} \\
 &= W_t - \lambda' \frac{\partial \sum_{i=1}^S (y_i - fw(x_i))^2}{\partial W}
 \end{aligned}$$

3) 다른 최적화 기법

확률적 경사 하강법을 사용할 때에는 손실함수에 안정점(saddle point)가 있을 경우 그 구간을 쉽게 벗어나지 못하는 문제점이 있다. 이러한 문제 해결을 위해 고안된 방법으로 가속도를 고려한 확률적 경사 하강법 (momentum SGD) 또는 ADAM¹⁶⁾ 등이 있다.

가속도를 고려한 확률적 경사 하강법은 학습이 진행되는 동안의 갱신 패턴을 누적하여 특정 방향으로 변수들이 계속 갱신되고 있을 경우 그 방향으로 더 빠르게 변수를 갱신하는 방법이다. 이는 [식 3-26]과 같은 과정을 통해 변수의 갱신이 이루어진다.

$$\begin{aligned}
 v_t &= \gamma v_{t-1} + \eta \nabla w L & [\text{식 3-26}] \\
 W_t &= W_t - v_t
 \end{aligned}$$

v_t 의 갱신은 기존의 갱신 과정이 누적된 v_{t-1} 과 현재 변수 기반 미분계수의 조합으로 이루어지며, 이 값이 경사 하강법을 통해 현재 변수를 갱신하는 데 사용된다. 가속도 기반의 경사 하강법은 γ , η 가 적절하게 설정되었

16) Kingma and Ba(2014), arXiv preprint arXiv:1412.6980.

을 때에는 우수한 학습 성능을 보여주지만 그렇지 않을 경우 잘못된 방향으로 학습이 이루어질 가능성이 높다.

이와 같은 단점을 극복하기 위해 제안된 대표적인 방법이 ADAM이다. ADAM은 가속도를 활용하면서도 가속도의 증가가 클 경우 해당 변수가 충분히 많이 갱신되었다면 갱신된 양에 비례하여 갱신 속도를 줄이는 방법이다. 이를 수식으로 표현하면 [식 3-27]과 같다.

$$\begin{aligned}v_t &= \beta_1 v_{t-1} - (1 - \beta_1) \nabla wL \\s_t &= \beta_2 v_{t-1} - (1 - \beta_2) \nabla^2 wL \\W_{t+1} &= W_t - \eta \frac{v_t}{\sqrt{s_t + \epsilon}}\end{aligned}\tag{식 3-27}$$

갱신 과정에서 s_t 는 미분계수의 제곱에 비례하는 값을 누적시켜 해당 변수의 갱신량을 줄이는 역할을 하는데, 가속도를 활용하여 특정 변수를 상당량 갱신한 경우 누적된 s_t 값이 그만큼 커지게 되므로 해당 변수들의 갱신량이 줄어드는 효과가 있어 보다 안정적인 갱신이 가능하다.

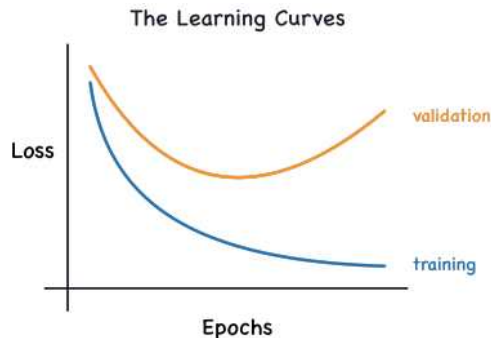
β_1 은 가속도 기반 경사 하강법에서 가속도 반영량을 결정하는데, β_1 이 1에 가까울수록 이전의 미분계수 사용 비중이 높아지고, 0에 가까울수록 새롭게 계산된 미분계수에 더 큰 가중치를 둔다. β_2 는 누적 미분계수를 조절하는 변수로 값이 1에 가까울수록 이전의 누적 미분계수를 적극적으로 사용하며, 0에 가까울수록 이전의 누적 미분계수 반영 비중을 줄이고 현재의 미분계수를 더 많이 반영한다는 의미로 해석할 수 있다.

3. 과적합

1) 과적합

일반적으로 인공신경망모형을 학습할 때에는 전체 데이터를 학습데이터(training), 검증데이터(validation), 시험데이터(test)로 나누어 학습 성능을 판단한다. 일반적으로 학습데이터는 실제 학습 과정에서 변수를 갱신하기 위해 입력값으로 사용되는 데이터를 의미하며, 검증데이터는 학습 과정이 올바르게 진행되고 있는지 모니터링 하기 위한 데이터이고, 마지막으로 시험데이터는 최종 학습된 모형의 성능을 평가를 위해 사용된다.

〈그림 3-9〉 과적합된 모형의 손실함수



자료: Kaggle(검색일: 2021. 9. 24)

머신러닝 알고리즘은 학습데이터에 대한 손실함수의 최소화를 목표로 학습이 진행되므로 학습이 진행됨에 따라 학습데이터를 이용한 예측값의 오차는 줄어들지만, 일정 수준 이상 학습이 진행되면 검증데이터의 손실값은 커지는 경향을 보인다. 이러한 현상을 과적합(overfitting)이라고 하는데, 〈그림 3-9〉는 전형적인 과적합 패턴을 보여준다. 일반적으로 과적합은 은닉층과 퍼셉트론이 많을수록, 학습 횟수가 많을수록 자주 발생한다. 과적합 방지를 위한 대표적인 방법으로는 변수의 정규화, Dropout, 조기종료(early stopping) 등이 있다.

2) 과적합 방지 기법

(1) 변수 정규화

변수 정규화는 일반적으로 모형에 과적합이 발생했을 경우 그렇지 않은 모형에 비해 변수들의 절대값이 크다는 데 기초하여 최적화 과정 중 변수의 값이 일정 수치 이상 커지지 않도록 손실함수에 [식 3-28]과 같은 정규화 항을 추가하는 것이다.

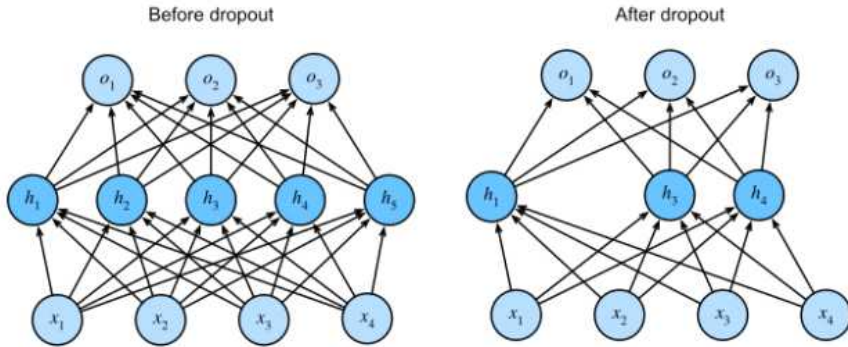
$$L = \sum_{i=1}^N l(y_i, \hat{y}_i) + \lambda R(W) \quad [\text{식 3-28}]$$

정규화하는 방법에는 크게 l-1, l-2 정규화 방식이 있는데, l-1 정규화의 경우 $R(W) = \sum |W_i|$ 와 같이 모든 변수의 절대값의 합을 목적함수에 추가하는 방법이며, l-2는 $R(W) = \sum (W_i)^2$ 과 같이 모든 변수의 제곱합을 목적함수에 추가하는 방법이다. 두 가지 방식 중 하나만 적용하거나 둘 다 한 번에 적용하여 학습하는 것도 가능하다.

(2) Dropout

Dropout은 순전파와 역전파 과정에서 구성된 신경망 전체를 활용하지 않고 무작위의 퍼셉트론을 제거한 후 남은 퍼셉트론만을 대상으로 학습을 진행하는 알고리즘이다. 일반적으로 학습이 진행되기 전 제거확률(또는 dropout 비율) p 를 지정하면 학습이 진행되면서 p 의 확률로 잠재변수가 소거되는 방식이다. 이는 <그림 3-10>과 같이 나타낼 수 있으며, Dropout이 적용된 모형은 일부 퍼셉트론이 제거된 상태에서 최종 결과를 출력하게 된다.

〈그림 3-10〉 Dropout 적용 후 신경망 구조의 변화



자료: Zhang *et al.*(2021), p.162

따라서 학습은 퍼셉트론이 일부만 사용되었을 경우에도 최적의 예측을 수행해야 하므로 부분적인 관찰이 일어난 상태에서도 유의미한 특성을 추출할 수 있는 방향으로 학습이 진행된다. 그러나 최종 학습이 끝난 이후의 모형에서는 시험데이터에 대해 Dropout을 적용하지 않은 전체 신경망을 이용하여 예측을 수행하는 것이 일반적이다.

(3) 조기종료

조기종료(Early stopping)는 검증데이터에 대한 예측 성능지표, 즉 손실 함수가 일정 횟수 이상의 학습이 이루어졌음에도 줄어들지 않을 경우 학습을 중단하는 방법이다. 일반적으로 검증데이터에 대한 손실함수가 커지기 시작하면 과적합이 진행될 확률이 높으므로 지속적인 학습이 모형의 예측력 향상에 도움이 되지 않기 때문이다. 추정해야 하는 파라미터의 수가 많고 복잡한 모형일수록 과적합 발생 확률이 높고 학습 시간도 오래 걸리기 때문에, 가용 시간 내에 최적의 모형을 찾아내기 위해서는 조기종료 사용이 도움이 될 수 있다.

04

예측모형 설계

제4장은 인공신경망 예측모형의 설계에 대한 내용이다. 제1절에서는 예측모형 구축에 사용한 분석자료와 분석자료별 기초통계 분석을 수행하였다. 제2절에서는 김 양식 단수 예측모형의 설계에 관한 내용을 다루었다. 분석자료를 바탕으로 한 DB 구축과 지역별 김 양식 단수 예측 모형의 설계, 그리고 인공신경망모형의 훈련을 위한 조건 설정 등의 내용을 포함한다.

제1절 분석자료

본 과제에서 구축하고자하는 예측모형의 대상은 지역별 김 단수이므로 종속변수(예측 대상)로는 수산업관측센터의 월별 김 단수를 사용하였으며, 환경변수로는 크게 기상변수와 수질변수 두 가지를 사용하였다. 기상변수와 해황변수는 앞서 살펴본 김의 생장에 영향을 미치는 환경적 요인을 기준으로 충분한 시계열이 확보된 변수들을 선정하여 분석에 활용하였다.

기상변수는 기상청에서 제공하는 2006년부터 2021년 6월까지 완도, 고흥, 해남지역의 월간 강수량과 일 최대 강수량, 월간 평균 풍속, 최대 순간풍속, 일조량(일조합)을 사용하였다. 해황변수는 2010년부터 국립수산과학원에서 제공하는 완도와 고흥 2개 지역의 표층수온, 표층용존산소량, 표층용존질소함유량, 표층무기인농도, 표층클로로필농도를 사용하였다.

1. 분석자료별 기초통계량

1) 생산변수(단수) 기초통계량

자료의 특성을 파악하기 위한 기초통계량을 분석하였다. 지역별 단수(속/책)의 평균을 살펴보면 전라남도는 19, 완도는 29, 고흥은 26으로 나타나 완도와 고흥지역이 전라남도 전체보다 평균 단수가 높은 것으로 나타났다. 중앙값은 전라남도가 18, 완도 26, 고흥 21로 전남 전체와 완도지역은 평균과 차이가 적으나 고흥은 평균과 중앙값의 차이가 컸다. 표준편차는 전라남도가 10.73으로 가장 낮으며, 완도와 고흥은 19.92 및 19.97로 유사했다.

〈표 4-1〉 단수 기초통계량

단위: 속/책			
구분	전남	완도	고흥
평균	18.5134	28.5805	26.4262
중앙값	17.9828	25.6315	21.0507
최대값	55.1623	76.5792	84.8737
최소값	2.1322	1.61475	0.2824
표준편차	10.7290	19.9197	19.9692
왜도	0.7787	0.47842	0.6318
첨도	3.7453	2.0943	2.7632
Jarque-Bera (P-value)	11.0548 (0.0040)	6.4373 (0.0400)	6.1281 (0.0467)
표본 수	89	89	89

자료: 저자 작성

왜도는 세 지역 모두 양수이지만 0에 가깝고, 오른쪽으로 다소 꼬리가 길지만 크게 치우쳐 있지는 않은 형태임을 알 수 있었다. 첨도는 전남 3.75, 완도 2.09, 고흥 2.76으로 세 지역 모두 분포가 평균을 중심으로 모여 있는 뾰족한 모습을 확인하였다. Jarque-Bera 검정결과 통계치가 모두 유의한 것으로 나타나 세 지역의 단수는 모두 정규분포와는 거리가 멀었다.

2) 기상변수 기초통계량

완도지역 기상변수의 평균은 월간 강수량 127, 일 최대 강수량 51, 월평균풍속 3, 최대순간풍속 20, 일조량 180이었다. 중앙값은 월간 강수량 100, 일 최대강수량 48, 월 평균풍속 3, 최대순간풍속 19, 일조량 178로 월간 강수량은 평균과 중앙값의 차이가 상대적으로 컸다. 표준편차는 월간 강수량 105.57, 일 최대강수량 35.96, 월 평균풍속 0.73, 최대순간풍속 5.05, 일조량 41.45로 월간 강수량의 표준편차가 가장 크게 나타났다.

〈표 4-2〉 기상변수(완도) 기초통계량

단위: mm, m/s

구분	월강수량	일최대강수량	월평균풍속	최대순간풍속	일조량
평균	126.5085	51.0244	3.0051	19.8750	180.0318
중앙값	99.9500	48.0000	2.9000	18.9000	178.3000
최댓값	569.5000	164.8000	5.1000	51.8000	300.5000
최솟값	0.5000	0.2000	1.4000	12.0000	79.7000
표준편차	105.5749	35.9641	0.7295	5.0475	41.4547
왜도	1.4518	0.7638	0.4653	1.9478	0.1157
첨도	5.6814	3.0606	2.6744	11.4303	2.8691
Jarque-Bera (P-value)	114.5491 (0.0000)	17.1395 (0.0001)	7.1286 (0.0283)	632.4593 (0.0000)	0.5180 (0.7718)
표본 수	176	176	176	176	176

자료: 저자 작성

왜도는 모든 변수들이 양의 값으로 오른쪽으로 다소 긴 꼬리를 가진 분포임을 확인할 수 있었으며, 월간 강수량과 최대순간풍속이 조금 더 왼쪽으로 치우친 분포를 가진 것으로 나타났다. 첨도는 월간 강수량 5.68, 일 최대강수량 3.06, 월평균풍속 2.67, 최대순간풍속 11.43, 일조량 2.87로 모든 변수들의 분포가 평균을 중심으로 뾰족한 모습을 보이는 것을 확인할 수 있었다. Jarque-Bera통계치는 일조량을 제외하고는 모두 유의적인 것으로 나타나 일조량만 정규분포 형태이며, 나머지 변수들은 정규분포와 거리가 먼 것으로 나타났다.

〈표 4-3〉 기상변수(해남) 기초통계량

단위: mm, m/s

구분	월강수량	일최대강수량	월평균풍속	최대순간풍속	일조량
평균	105.4489	44.0778	2.1966	16.2330	174.1494
중앙값	74.9500	30.0000	2.2000	16.0000	171.4500
최대값	549.1000	186.6000	3.2000	30.5000	282.7000
최소값	2.2000	1.5000	1.0000	6.3000	73.5000
표준편차	97.6090	39.6162	0.3819	3.9716	42.4873
왜도	1.7791	1.5416	0.1016	0.6130	0.1397
첨도	6.9487	5.1596	3.0353	3.7411	2.7307
Jarque-Bera (P-value)	207.1924 (0.0000)	103.9152 (0.0000)	0.3118 (0.8556)	15.0515 (0.0005)	1.1043 (0.5757)
표본 수	176	176	176	176	176

자료: 저자 작성

해남지역 기상변수들의 평균은 월간 강수량 105, 일 최대강수량 44, 월 평균풍속 2, 최대순간풍속 16, 일조량 174로 나타났다. 중앙값은 월간 강수량 75, 일 최대강수량 30, 월 평균풍속 2, 최대순간풍속 16, 일조량 171로 완도지역과 동일하게 다른 변수들은 평균과 큰 차이가 없었으나, 월간 강수량은 상대적으로 큰 차이를 보이고 있다. 표준편차는 월간 강수량 97.61, 일 최대강수량 39.62, 월 평균풍속 0.38, 최대순간풍속 3.97, 일조량 42.49로 월간 강수량의 표준편차가 가장 큰 것으로 분석되었다.

왜도는 모든 변수들이 양의 값으로 오른쪽으로 다소 긴 꼬리를 가진 분포임을 확인 할 수 있었으며 월간 강수량과 일 최대강수량이 조금 더 왼쪽으로 치우친 분포를 가진 것으로 나타났다. 첨도는 월간 강수량 6.95, 일 최대강수량 5.16, 월 평균풍속 3.04, 최대순간풍속 3.74, 일조량 2.73으로 모든 변수들의 분포가 평균을 중심으로 뾰족한 모습을 보이는 것을 확인할 수 있었다. Jarque-Bera통계치는 월 평균풍속과 일조량을 제외한 변수들은 유의적인 것으로 나타나 월 평균풍속과 일조량만 정규분포 형태이며, 나머지 변수들은 정규분포와 거리가 먼 것으로 나타났다.

〈표 4-4〉 기상변수(고흥) 기초통계량

단위: mm, m/s

구분	월강수량	일최대강수량	월평균풍속	최대순간풍속	일조량
평균	121.6358	53.9563	1.7278	16.1222	178.5670
중앙값	83.5000	38.5000	1.7000	16.1500	178.3500
최대값	656.1000	306.5000	3.1000	30.7000	277.1000
최소값	0.0000	0.0000	0.7000	8.9000	67.2000
표준편차	115.6675	51.7609	0.4450	3.5603	39.3519
왜도	1.7105	2.1515	0.4777	0.7304	-0.0945
첨도	6.5987	9.3677	2.8201	4.5649	2.9020
Jarque-Bera (P-value)	180.7957 (0.0000)	433.1261 (0.0000)	6.9299 (0.0313)	33.6068 (0.0000)	0.3322 (0.8470)
표본 수	176	176	176	176	176

자료: 저자 작성

고흥지역 기상변수들의 평균은 월간 강수량 122, 일 최대강수량 54, 월 평균풍속 2, 최대순간풍속 16, 일조량 179로 나타났다. 중앙값은 월간 강수량 84, 일 최대강수량 39, 월 평균풍속 2, 최대순간풍속 16, 일조량 178로 월간 강수량과 일 최대강수량은 평균과 중앙값이 어느 정도 차이가 있는 것으로 확인되었다. 표준편차는 월간 강수량 115.67, 일 최대강수량 51.76, 월 평균풍속 0.45, 최대순간풍속 3.56, 일조량 39.35로 월간 강수량의 표준편차가 가장 큰 것으로 분석되었다.

왜도는 모든 변수들이 양의 값으로 오른쪽으로 다소 긴 꼬리를 가진 분포임을 확인 할 수 있었으며 월간 강수량과 일 최대강수량이 조금 더 왼쪽으로 치우친 분포를 가진 것으로 나타났다. 첨도는 월간 강수량 6.60, 일 최대강수량 9.37, 월평균풍속 2.82, 최대순간풍속 4.56, 일조량 2.90으로 모든 변수들의 분포가 평균을 중심으로 뾰족한 모습을 보이는 것을 확인할 수 있었다. Jarque-Bera통계치는 일조량을 제외한 변수들은 유의적인 것으로 나타나 일조량만 정규분포 형태이며, 나머지 변수들은 정규분포와 거리가 먼 것으로 나타났다.

기상변수에 대한 기초통계량 분석 결과 완도는 다른 지역에 비해 월간 강수량이 많고 월 평균풍속과 최대순간풍속이 높으며 일조량이 많다는 특성을 확인할 수 있었다. 해남은 월간 강수량, 일 최대강수량, 일조량이 가장 적은 지역이며, 고흥은 월 평균풍속이 가장 낮은 지역으로 나타났다.

3) 해양변수 기초통계량

완도지역 해양변수들의 평균은 표층수온 16, 표층용존산소량 8, 표층용존질소함유량 0.09, 표층무기인농도 0.01, 표층클로로필농도 3으로 나타났다. 중앙값은 표층수온 15, 표층용존산소량 8, 표층용존질소함유량 0.09, 표층무기인농도 0.01, 표층클로로필농도 1로 표층클로로필농도는 평균과 중앙값이 어느 정도 차이가 있는 것으로 확인되었다. 표준편차는 표층수온 5.05, 표층용존산소량 1.12, 표층용존질소함유량 0.04, 표층무기인농도 0.01, 표층클로로필농도 3.22로 표층수온과 표층클로로필농도의 표준편차가 가장 큰 것으로 분석되었다.

〈표 4-5〉 해양변수(완도) 기초통계량

단위: °C, mg/L, µg/L

구분	표층수온	DO	DIN	DIP	CHL
평균	15.6873	7.6867	0.0941	0.0117	2.5491
중앙값	15.4713	7.6175	0.0917	0.0119	1.4535
최댓값	26.3692	10.6275	0.2637	0.0240	19.2683
최솟값	7.564	5.4890	0.0134	0.0004	0.2958
표준편차	5.0481	1.1182	0.0434	0.0056	3.2199
왜도	0.1170	0.1251	0.5746	0.0685	2.9306
첨도	1.7215	2.2859	3.9274	2.0660	11.9217
Jarque-Bera (P-value)	9.6428 (0.0081)	3.2682 (0.1951)	12.4478 (0.0002)	5.0873 (0.0786)	650.4744 (0.0000)
표본 수	137	137	137	137	137

자료: 저자 작성

왜도는 모든 변수들이 양의 값으로 오른쪽으로 다소 긴 꼬리를 가진 분포임을 확인할 수 있었으며 대부분의 변수는 0에서 크게 벗어나지 않았으나 표층클로로필농도는 평균을 중심으로 다소 왼쪽으로 치우친 분포를 가진 것으로 나타났다. 첨도는 표층수온 1.72, 표층용존산소량 2.29, 표층용존질소함유량 3.93, 표층무기인농도 2.07, 표층클로로필농도 11.92로 모든 변수들의 분포가 평균을 중심으로 뾰족한 모습을 보이는 것을 확인할 수 있었으며 표층클로로필농도가 가장 뾰족한 형태의 분포를 가지고 있는 것으로 나타났다. Jarque-Bera통계치는 표층용존산소량을 제외한 변수들은 유의적인 것으로 나타나 표층용존산소량만 정규분포 형태이며, 나머지 변수들은 정규분포와 거리가 먼 것으로 나타났다.

〈표 4-6〉 해황변수(고흥) 기초통계량

단위: °C, mg/L, µg/L

구분	표층수온	DO	DIN	DIP	CHL
평균	16.2262	7.6766	0.0594	0.0071	4.5992
중앙값	16.4025	7.5000	0.0511	0.0063	2.8502
최대값	27.9125	11.0025	0.2368	0.0188	28.5375
최소값	3.8600	4.7325	0.0075	0.0005	0.4516
표준편차	6.6891	1.3304	0.0398	0.0044	5.3115
왜도	-0.0206	0.3060	1.7785	0.6488	2.7710
첨도	1.6858	2.3906	7.5766	2.5379	10.9581
Jarque-Bera (P-value)	9.8693 (0.0072)	4.2579 (0.1190)	191.7913 (0.0000)	10.8314 (0.0044)	536.8412 (0.0000)
표본 수	137	137	137	137	137

자료: 저자 작성

고흥지역 해황변수들의 평균은 표층수온 16, 표층용존산소량 8, 표층용존질소함유량 0.06, 표층무기인농도 0.01, 표층클로로필농도 5로 나타났다. 중앙값은 표층수온 16, 표층용존산소량 8, 표층용존질소함유량 0.05, 표층무기인농도 0.01, 표층클로로필농도 3으로 표층클로로필농도는 완도 지역과 마찬가지로 평균과 중앙값이 어느 정도 차이가 있었다. 표준편차는

표층수온 6.69, 표층용존산소량 1.33, 표층용존질소함유량 0.04, 표층무기인농도 0.004, 표층클로로필농도 5.31로 표층수온과 표층클로로필농도의 표준편차가 가장 큰 것으로 분석되었다.

왜도는 표층수온은 음의 값으로 왼쪽으로 꼬리가 긴 형태를 가지고 있으며, 나머지 변수들은 양의 값으로 오른쪽으로 다소 긴 꼬리를 가진 분포임을 확인 할 수 있었다. 특히 표층용존질소함유량과 표층클로로필농도는 평균을 중심으로 다소 왼쪽으로 치우친 분포를 가진 것으로 나타났다. 첨도는 표층수온 1.69, 표층용존산소량 2.39, 표층용존질소함유량 7.58, 표층무기인농도 2.54, 표층클로로필농도 10.96으로 모든 변수들의 분포가 평균을 중심으로 뾰족한 모습을 보이는 것을 확인할 수 있었으며 표층클로로필농도가 가장 뾰족한 형태의 분포를 나타냈다. Jarque-Bera통계치는 표층용존산소량을 제외한 변수들은 유의적인 것으로 나타나 표층용존산소량을 제외한 변수들은 정규분포와 거리가 먼 것으로 나타났다.

2. 상관관계 분석

기초통계량 분석에 이어 변수 간 상관관계를 분석하였다. 상관관계 분석을 통해 도출되기를 기대하는 방향성은 단수와 강수량 관련 변수 사이에는 음의 상관관계, 풍속 관련 변수 사이에는 양의 상관관계, 일조량과는 양의 상관관계를 가지는 것으로 기대하고 있다. 이는 강수량은 염도를 저하시켜 단수에 악영향을 미치고, 풍속은 조류 소통을 원활하게 하여 김의 생장에 도움을 주기 때문이다. 다만 최대순간풍속은 너무 높을 경우 김이 탈락하고 시설에 피해가 있을 수도 있어 음의 상관관계를 가질 가능성도 있을 것으로 판단된다. 단수와 해황변수 사이에는 수온은 음의 상관관계, 나머지 변수들은 김의 생장을 도와주는 역할을 하기 때문에 양의 상관관계가 도출되는 것이 본 연구의 방향성에 부합할 것으로 판단된다.

1) 단수와 기상변수의 상관관계 분석

기상변수는 지역별로 분리되어 있기 때문에 전라남도의 김 단수와 각 지역별 기상변수 사이의 상관계수를 분석하였다. 전라남도 단수와 완도지역 기상변수 사이의 상관관계 분석 결과 단수와 유의미한 상관관계를 가진 변수는 존재하지 않았으며 월간 강수량과 일 최대강수량 사이, 월평균 풍속과 최대순간풍속 사이에만 높은 양의 상관관계가 있는 것으로 나타났다.

〈표 4-7〉 전라남도 단수와 완도 기상변수 상관계수

구분	단수	월간 강수량	일최대강수량	월평균풍속	최대순간풍속	일조량
단수	1.0000					
월강수량	-0.0071	1.0000				
일최대강수량	-0.0857	0.8386	1.0000			
월평균풍속	-0.0011	-0.0033	-0.0910	1.0000		
최대순간풍속	0.0095	0.3470	0.2220	0.5202	1.0000	
일조량	0.0263	-0.1778	0.0559	-0.0782	-0.2150	1.0000

자료: 저자 작성

전라남도 단수와 해남지역의 기상변수 사이의 상관관계를 분석해 본 결과 월평균 풍속이 0.2136, 최대순간풍속이 0.1560으로 풍속이 단수와 약한 양의 상관관계에 있음을 확인할 수 있었다. 반면 월간강수량과 일 최대강수량은 약한 음의 상관관계에 있는 것으로 나타났다.

〈표 4-8〉 전라남도 단수와 해남 기상변수 상관계수

구분	단수	월간 강수량	일최대강수량	월평균풍속	최대순간풍속	일조량
단수	1.0000					
월강수량	-0.1067	1.0000				
일최대강수량	-0.1383	0.8648	1.0000			
월평균풍속	0.2136	0.2924	0.1705	1.0000		
최대순간풍속	0.1560	0.2825	0.1551	0.3431	1.0000	
일조량	0.0004	-0.2621	-0.1106	-0.0767	-0.1371	1.0000

자료: 저자 작성

전라남도 단수와 고흥지역 기상변수의 상관관계는 월평균 풍속이 0.2679로 가장 높은 상관관계를 보이고 있었다. 나머지 변수들은 단수와 큰 상관관계가 존재하지 않는 것으로 확인되었다.

전라남도 전체의 단수와 지역별 기상변수 사이의 상관관계 분석 결과 대체로 풍속에 관한 변수들은 단수에 양의 영향을 미치며, 강수량과 관련된 변수들은 음의 영향을 미치는 것으로 나타났다. 이는 김의 생육조건 중 적당한 풍속이 있어야 공기의 용해량이 많아져 이산화탄소, 산소가 많이 공급될 수 있고, 해수의 소동이 원활해져 영양염 공급이 풍부해진다는 특성과 어느 정도 부합하는 결과인 것으로 판단된다.

강수량이 단수와 음의 상관관계에 있는 것은 강수량이 많을 때 바닷물의 염도를 낮추거나 육상의 물질들이 바다로 떠 내려와 환경적인 충격을 주어 김의 생육에 부정적인 영향을 끼치는 경우가 있기 때문으로 보인다.

〈표 4-9〉 전라남도 단수와 고흥 기상변수 상관계수

구분	단수	월간 강수량	일최대강수량	월평균풍속	최대순간풍속	일조량
단수	1.0000					
월강수량	-0.0420	1.0000				
일최대강수량	-0.0939	0.8858	1.0000			
월평균풍속	0.2679	-0.1818	-0.2032	1.0000		
최대순간풍속	0.0954	0.1874	0.0893	0.4676	1.0000	
일조량	0.0723	-0.3652	-0.2263	0.0579	-0.0696	1.0000

자료: 저자 작성

완도지역 단수와 기상변수 사이의 상관관계를 살펴보면 일조량이 0.2662로 다른 변수들에 비해 높은 상관관계를 보이고 있다. 월간 강수량, 일 최대 강수량, 일조량은 단수와 양의 상관관계에 있음을 확인할 수 있으며, 월 평균 풍속과 최대순간풍속은 약한 음의 상관관계에 있다. 따라서 완도지역에서는 일조량이 단수에 가장 밀접한 영향을 미치는 변수로 확인된다.

〈표 4-10〉 완도 단수와 기상변수 상관계수

구분	단수	월간 강수량	일최대강수량	월평균풍속	최대순간풍속	일조량
단수	1.0000					
월강수량	0.1724	1.0000				
일최대강수량	0.1733	0.8386	1.0000			
월평균풍속	-0.0196	-0.0033	-0.0910	1.0000		
최대순간풍속	-0.0401	0.3470	0.2220	0.5202	1.0000	
일조량	0.2662	-0.1778	0.0559	-0.0782	-0.2150	1.0000

자료: 저자 작성

고흥지역 단수와 기상변수 사이의 상관관계를 살펴보면, 월간 강수량이 -0.3296으로 가장 높은 상관관계를 보이고 있으며 다음으로는 일 최대강수량이 -0.2628의 상관계수를 가진 것으로 나타났다. 따라서 고흥지역도 강수량이 단수에 부정적 영향을 미칠 가능성이 높을 것으로 보인다. 최대 순간풍속과 일조량 역시 약한 음의 상관관계에 있으며, 월 평균풍속과는 양의 상관관계에 있지만 그 영향력이 매우 낮은 것으로 보인다.

〈표 4-11〉 고흥 단수와 기상변수 상관계수

구분	단수	월간 강수량	일최대강수량	월평균풍속	최대순간풍속	일조량
단수	1.0000					
월강수량	-0.3296	1.0000				
일최대강수량	-0.2628	0.8858	1.0000			
월평균풍속	0.0424	-0.1818	-0.2032	1.0000		
최대순간풍속	-0.1442	0.1874	0.0893	0.4676	1.0000	
일조량	-0.1554	-0.3652	-0.2263	0.0579	-0.0696	1.0000

자료: 저자 작성

2) 단수와 해황변수의 상관관계 분석

전라남도 단수와 완도지역 해황변수 사이의 상관관계를 살펴보면, 표층 수온이 -0.7268로 가장 높은 상관관계를 보이고 있으며, 다음으로 표층용

존산소량이 0.4000으로 둘째로 높은 상관관계를 보이고 있다. 표층클로로필농도는 약한 음의 상관관계, 표층무기인농도와 표층용존질소함유량¹⁷⁾은 약한 양의 상관관계를 가지고 있는 것으로 나타났다.

〈표 4-12〉 전라남도 단수와 완도 해황변수 상관계수

구분	단수	표층수온	DO	DIN	DIP	CHL
단수	1.0000					
표층수온	-0.7268	1.0000				
DO	0.4000	-0.5860	1.0000			
DIN	0.1460	-0.5429	0.1522	1.0000		
DIP	0.1827	-0.5523	0.4471	0.6885	1.0000	
CHL	-0.2482	0.3267	-0.3490	0.0820	-0.2246	1.0000

자료: 저자 작성

전라남도 단수와 고흥지역 해황변수 사이의 상관관계를 살펴보면, 표층수온이 -0.6934로 가장 높은 상관관계를 보이고 있으며, 다음으로 표층용존산소량이 0.4895로 둘째로 높은 상관관계를 보이고 있다. 나머지 변수인 표층클로로필농도, 표층무기인농도와 표층용존질소함유량은 단수와 유의적인 상관관계는 존재하지 않는 것으로 나타났다.

전라남도 전체의 단수와 특정 지역 해황변수와의 상관관계 분석 결과 공통적으로 표층수온과는 큰 음의 상관관계가 존재했고, 표층용존산소량과는 양의 상관관계가 어느 정도 있는 것으로 분석되었다. 전라남도 전체의 단수와 특정 지역 해황변수 사이의 상관관계를 분석하였다는 한계는 있으나 표층수온, 표층용존산소량과는 유의미한 상관관계가 있음을 확인할 수 있다.

완도지역의 단수와 해황변수 사이의 상관관계를 분석해보면, 표층수온이 -0.6198로 가장 높은 상관관계를 가지고 있었으며, 다음으로는 표층용존산소량이 0.4351로 둘째로 높은 상관관계를 가지고 있었다.

17) 이하의 표에서는 다음과 같은 표기를 사용한다. 표층용존산소량: DO, 표층용존질소함유량: DIN, 표층무기인농도: DIP, 표층클로로필농도: CHL

〈표 4-13〉 전라남도 단수와 고흥 해황변수 상관계수

구분	단수	표층수온	DO	DIN	DIP	CHL
단수	1.0000					
표층수온	-0.6934	1.0000				
DO	0.4895	-0.7392	1.0000			
DIN	-0.0832	-0.3226	0.1268	1.0000		
DIP	-0.0603	-0.3596	0.3253	0.7333	1.0000	
CHL	0.0360	0.1254	-0.1143	0.1648	-0.0708	1.0000

자료: 저자 작성

나머지 변수들 중에서는 표층클로로필농도가 약한 음의 상관관계에 있으며 표층용존질소함유량과 표층무기인농도는 단수와 상관관계가 거의 없는 것으로 나타났다.

〈표 4-14〉 완도 단수와 해황변수 상관계수

구분	단수	표층수온	DO	DIN	DIP	CHL
단수	1.0000					
표층수온	-0.6198	1.0000				
DO	0.4351	-0.5860	1.0000			
DIN	0.0155	-0.5429	0.1522	1.0000		
DIP	0.0239	-0.5523	0.4471	0.6885	1.0000	
CHL	-0.1492	0.3267	-0.3490	0.0820	-0.2246	1.0000

자료: 저자 작성

고흥지역 단수와 해황변수 사이의 상관관계를 분석해본 결과 완도지역과 동일하게 표층수온이 -0.5864로 가장 높은 상관관계를 가지고 있었으며, 다음으로 표층용존산소량이 0.2891로 둘째로 높은 상관관계를 가지고 있었다. 나머지 변수들 중에서는 표층무기인농도가 약한 양의 상관관계에 있었으며 표층용존질소함유량과 표층클로로필농도는 거의 상관관계가 없는 것으로 나타났다. 전라남도 단수와 지역별 해황변수의 상관관계 분석 결과와 마찬가지로 완도와 고흥지역에서도 표층수온과는 음의 상관관계가 존재하고 표층용존산소량과는 양의 상관관계가 존재했다.

해황변수의 상관관계 분석으로 확인할 수 있는 점은 김의 단수가 표층수온과 표층용존산소량에 의해 결정되는 부분이 크다는 점이다. 구체적으로 단수와 표층수온 사이에 높은 상관계수가 도출된 것은 김의 성장에 있어서 가장 중요한 요소 중 하나가 수온이기 때문인 것으로 판단된다. 또한 상관계수의 부호가 음으로 나온 것은 여름철 고수온에 약한 김의 특성이 어느 정도 반영된 결과인 것으로 보인다.

표층용존산소량은 많을수록 바다의 영양염류를 풍부하게 만들어서 김의 성장에 긍정적인 역할을 할 가능성이 높기 때문에 단수와는 매우 높은 양의 상관관계에 있는 것으로 판단된다.

〈표 4-15〉 고흥 단수와 해황변수 상관관계

구분	단수	표층수온	DO	DIN	DIP	CHL
단수	1.0000					
표층수온	-0.5864	1.0000				
DO	0.2891	-0.7392	1.0000			
DIN	0.0382	-0.3226	0.1268	1.0000		
DIP	0.1760	-0.3596	0.3253	0.7333	1.0000	
CHL	0.0705	0.1254	-0.1143	0.1648	-0.0708	1.0000

자료: 저자 작성

3) 기상변수와 해황변수의 상관관계 분석

단수에 영향을 미치는 기상변수와 해황변수 사이의 상관관계 분석 결과 완도지역에서는 대체로 강수량 관련 변수와 해황변수 사이의 상관관계가 높은 것으로 나타났다. 강수량 변수인 표층수온, 표층클로로필 농도와는 양의 상관관계를, 표층용존산소량과 표층용존질소함유량, 표층무기인농도와는 음의 상관관계를 가지고 있는 것으로 확인되었다.

고흥지역에서도 완도지역과 동일하게 대체로 강수량 관련 변수와 해황변수 사이의 상관관계가 높은 것으로 나타났다. 강수량 변수인 표층수온, 표

층클로로필 농도와는 양의 상관관계를, 표층용존산소량과 표층용존질소함량, 표층무기인농도와는 음의 상관관계를 나타냈다.

〈표 4-16〉 완도 기상변수와 해양변수의 상관계수

구분	월강수량	일최대강수량	월평균풍속	최대순간풍속	일조량	표층수온	DO	DIN	DI	CHL
월간강수량	1.00									
일최대강수량	0.84	1.00								
월평균풍속	0.00	-0.09	1.00							
최대순간풍속	0.35	0.22	0.52	1.00						
일조량	-0.18	0.06	-0.08	-0.21	1.00					
표층수온	0.44	0.43	-0.34	0.19	-0.01	1.00				
DO	-0.22	-0.22	0.30	-0.07	0.19	-0.59	1.00			
DIN	-0.44	-0.40	0.04	-0.17	-0.04	-0.54	0.15	1.00		
DIP	-0.43	-0.40	0.03	-0.19	0.00	-0.55	0.45	0.69	1.00	
CHL	0.32	0.29	-0.01	0.15	-0.01	0.33	-0.35	0.08	-0.22	1.00

자료: 저자 작성

그러나 고흥지역에서는 풍속 관련 변수가 완도지역과 다소 차이를 보였다. 수온과는 유의미한 음의 상관관계를 가지고 있는 것으로 나타났으며, 표층용존산소량과는 양의 상관관계가 높은 것으로 나타났다.

〈표 4-17〉 고흥 기상변수와 해양변수의 상관계수

구분	월간강수량	일최대강수량	월평균풍속	최대순간풍속	일조량	표층수온	DO	DIN	DIP	CHL
월간강수량	1.00									
일최대강수량	0.89	1.00								
월평균풍속	-0.18	-0.20	1.00							
최대순간풍속	0.19	0.09	0.47	1.00						
일조량	-0.37	-0.23	0.06	-0.07	1.00					
표층수온	0.55	0.49	-0.58	-0.20	-0.13	1.00				
DO	-0.40	-0.36	0.53	0.15	0.18	-0.74	1.00			
DIN	-0.22	-0.16	0.14	0.19	-0.06	-0.32	0.13	1.00		
DI	-0.38	-0.33	0.19	0.16	0.06	-0.36	0.33	0.73	1.00	
CHL	0.22	0.19	0.02	0.10	-0.20	0.13	-0.11	0.16	-0.07	1.00

자료: 저자 작성

4) 시사점

기초통계량 분석과 상관관계 분석을 한 결과 몇 가지 시사점을 도출할 수 있었다. 첫째, 기초통계량 분석 결과에서는 단수와 기상 및 해황변수들이 일부는 정규분포 형태를 보였으나 일부는 정규분포에서 벗어난 것으로 나타났다. 이는 자료의 왜도 및 첨도가 정규분포와 일치하지 않는 변수들이 많다는 의미이므로 이 변수들은 로그변환, 차분 등을 이용하여 분석하는 것이 더 이론에 부합할 수 있다.

둘째, 상관관계 분석에서 기상변수는 어느 정도 기대한 것과 일치하는 부호의 계수가 추정된 변수들이 많았으나 해황변수는 기대한 것과 반대되는 부호가 나오는 경우가 많았다. 기대치와 반대되는 부호로 계수가 추정된 경우 회귀분석 기반의 예측모형 구축 시 기본적인 가정을 위배하므로 일반적인 계량경제 분석을 진행하기에는 어려움이 있다.

셋째, 최대풍속과 수온 같은 변수들은 단수에 부정적인 영향을 미칠지 긍정적인 영향을 미칠지 판단하기 어려운 부분이 있다. 태풍과 같은 천재지변이 아닌 일반적인 기상 상태의 최대풍속이 김의 탈락률에 미치는 영향은 본 연구에서 가정을 설정하기가 굉장히 어렵기 때문이다. 수온 역시 높아질수록 단수에 부정적 영향을 미치는 것은 맞지만 지나치게 낮은 수온도 김 생장에 부정적 영향을 미치기 때문에 기대부호를 설정하기가 어렵다.

기상 및 해황변수들은 분명히 김의 단수에 영향을 미치고 있으나, 위에 제시한 한계로 인해 계량경제학적 연구방법론을 가지고는 분석하기에 어려움이 있었다. 따라서 본 연구에서는 자료의 제약이 없는 인공신경망모형을 이용하여 기상 및 해황변수들을 고려한 단수를 추정하고 예측력 검정에 초점을 맞추어 다중회귀모형과 비교분석하고자 한다.

제2절 모형 설계

본 연구는 양식수산물의 단수 예측에 있어 인공지능망모형 도입의 유효성을 확인하는 것이 목적이다. 현재까지의 연구에서 인공지능망모형 최적 파라미터(은닉층 및 퍼셉트론의 수 등)의 정량적 설정을 위한 기준이나 방법은 제시되지 않았다. 따라서 최적의 예측모형 도출을 위해서는 변수와 파라미터를 여러 가지 방법으로 변화시켜 보면서 각각의 예측 성능을 비교하여 우수한 모형을 선택하는 것이 최선이라고 할 수 있다. 따라서 이번 연구에서는 변수 간의 이론적 상관관계와 모형의 유의성을 제시하기보다는 가용한 자료를 이용하여 최대한 다양한 예측모형을 추정하고 비교하는 것을 목표로 모형을 설계하였다. 인공지능망모형을 이용한 김 단수 전망에는 예측 대상이 되는 지역, 예측 대상 시차, 입력 데이터, 모형의 종류, 은닉층의 퍼셉트론 수에 따라 총 216개의 모형을 학습하고 예측 결과를 비교하였다. 모형 구축에는 Python 3.7 및 Tensorflow 2.0 등을 활용하였다.

1. 지역별 DB

본 연구에서는 우리나라에서 김 생산 비중이 가장 큰 전남지역의 김 단수 예측모형을 설계하였으며, 전남지역 내에서도 가용자료의 폭이 가장 넓었던 완도와 고흥에 대해서는 세부 지역 DB를 별도로 구분하여 단수 예측모형을 구축하였다. 가용 데이터의 종류별 시계열 길이가 달라 독립변수를 각 지역별로 3개의 세트로 구분하였으며, 이에 따라 총 9개의 DB를 구축하였다. 또한 생산의 계절성 반영을 위해 11월부터 5월까지의 더미변수를 추가하여 분석했다.

1) 전남

전남의 김 단수 추정을 위해 사용된 데이터는 2006년 11월부터 2021년 6월까지의 전남지역의 평균 월별 김 단수와 동 기간의 월별 기상변수, 그리고 2010년 2월부터 2021년 6월까지의 수질 변수의 월별 데이터를 사용하였다.

기상변수에는 완도, 고흥, 해남 3개 지역의 월간 강수량과 일 최대강수량, 월간 평균풍속, 최대순간풍속, 일조량(일조합)을 사용하였으며, 수질변수에는 완도와 고흥 2개 해역 표층의 수온, 용존산소량, 용존질소함유량, 용존무기인함유량, 클로로필 농도를 사용하였다.

독립변수별 시계열의 길이는 1: 기상변수(2006. 11.~2021. 6.), 2: 수질변수(2010. 2.~2021. 6.), 3: 기상+수질변수(2010. 2.~2021. 6.)의 총 3개의 데이터셋으로 나누어 각각 추정하였다.

2) 완도

완도의 김 단수 추정을 위해 사용된 데이터는 2006년 11월부터 2021년 6월까지의 완도지역의 평균 월별 김 단수와 기상변수, 그리고 2010년 2월부터 2021년 6월까지의 수질변수 월별 데이터를 사용하였다.

기상변수에는 완도지역의 월간 강수량과 일 최대강수량, 월간 평균풍속, 최대순간풍속, 일조량(일조합)을 사용하였으며, 수질변수에는 완도 해역 표층의 수온, 용존산소량, 용존질소함유량, 용존무기인함유량, 클로로필 농도를 사용하였다.

전남과 마찬가지로 독립변수 시계열은 1: 기상변수(2006. 11.~2021. 6.), 2: 수질변수(2010. 2.~2021. 6.), 3: 기상+수질변수(2010. 2.~2021. 6.)의 총 3개 데이터셋으로 나누어 각각 추정하였다.

3) 고흥

고흥의 김 단수 추정을 위해 사용된 데이터는 2006년 11월부터 2021년 6월까지의 고흥지역의 평균 월별 김 단수와 기상변수, 그리고 2010년 2월부터 2021년 6월까지의 수질변수 월별 데이터를 사용하였다.

기상변수에는 고흥지역의 월간 강수량과 일 최대강수량, 월간 평균풍속, 최대순간풍속, 일조량(일조합)을 사용하였으며, 수질변수에는 고흥 해역 표층의 수온, 용존산소량, 용존질소함유량, 용존무기인함유량, 클로로필 농도를 사용하였다.

앞의 경우와 마찬가지로 데이터별 시계열 길이가 달라 독립변수는 1: 기상변수(2006. 11.~2021. 6.), 2: 수질변수(2010. 2.~2021. 6.), 3: 기상+수질변수(2010. 2.~2021. 6.)의 총 3개 데이터셋으로 나누어 각각 추정하였다.

〈표 4-18〉 지역별 단수 예측모형에 사용된 독립변수

독립변수			단수 예측 지역								
종류	지역	항목	전남			완도			고흥		
			1	2	3	1	2	3	1	2	3
기상 변수	완도	월간 강수량	◎		◎	◎		◎			
		일 최대강수량	◎		◎	◎		◎			
		월간 평균풍속	◎		◎	◎		◎			
		최대순간풍속	◎		◎	◎		◎			
		일조량	◎		◎	◎		◎			
	고흥	월간 강수량	◎		◎				◎		◎
		일 최대강수량	◎		◎				◎		◎
		월간 평균풍속	◎		◎				◎		◎
		최대순간풍속	◎		◎				◎		◎
		일조량	◎		◎				◎		◎

독립변수			단수 예측 지역								
종류	지역	항목	전남			완도			고흥		
			1	2	3	1	2	3	1	2	3
	해남	월간 강수량	◎		◎						
		일 최대강수량	◎		◎						
		월간 평균풍속	◎		◎						
		최대순간풍속	◎		◎						
		일조량	◎		◎						
수질 변수 (표층)	완도	수온		◎	◎		◎	◎			
		DO		◎	◎		◎	◎			
		DIN		◎	◎		◎	◎			
		DIP		◎	◎		◎	◎			
		CHL		◎	◎		◎	◎			
	해남	수온		◎	◎					◎	◎
		DO		◎	◎					◎	◎
		DIN		◎	◎					◎	◎
		DIP		◎	◎					◎	◎
		CHL		◎	◎					◎	◎

주: ◎는 각각의 독립변수가 지역별 1~4번 예측모형에 사용되었음을 의미함. 입력된 데이터의 시계열 길이는 1, 3은 2010년 2월부터 2021년 6월까지임
 자료: 저자 작성

2. 인공신경망 예측모형 및 학습조건 설계

1) 인공신경망 예측모형

지역별 김 단수 예측을 위해 사용한 인공신경망모형은 다층 퍼셉트론(MLP), 장단기 기억신경망(LSTM), 양방향 장단기 기억신경망(Bidirectional LSTM) 등 세 가지이다. 모형의 종류와 상관없이 모두 4개의 은닉층으로 구성하였으며, 노드(퍼셉트론)의 개수는 각 은닉층별 10개인 경우와 100개인 경우 두 가지로 구분하였다.

종속변수의 시차와 길이에 따라 1개월 예측모형과 어기 예측모형으로 구분하였으며, 각각은 다시 독립변수와 종속변수의 시차에 따라 단기 및 장기 예측모형으로 구분하였다. 1개월 예측모형의 경우 다중회귀모형과 동일한 변수를 사용하여 비교하기 위해 계절더미(11월~5월)를 추가하였으며, 계절더미는 입력 데이터의 시점이 아닌 예측 대상의 시점에 대응해야 하므로 별도의 입력층을 통한 선형 연산(퍼셉트론)을 거쳐 기상 및 수질변수의 인공지능경망모형(MLP, LSTM, 양방향 LSTM) 도출 결과와 결합한 이후 최종 예측값을 출력하는 구조로 설정하였다. 어기 예측모형은 출력층에서 8개 출력값을 위한 가중치가 각각 별도로 계산되므로 계절더미를 추가하지 않았다.

〈그림 4-1〉 김 단수 예측을 위한 인공지능경망모형의 구조



자료: 저자 작성

모형의 구분을 간략히 하기 위해 입력데이터의 시계열 길이-독립변수와 종속변수의 시차-예측 대상 시계열 길이로 표현하면, 1개월 예측모형은 1-1-1(단기), 1-5-1(장기)의 두 가지로 나뉜다. 1-1-1은 $t-1$ 시점의 데이터를 이용하여 t 시점의 단수를 예측하는 모형이고, 1-5-1은 $t-5$ 시점의 데이터를 이용하여 t 시점의 단수를 예측하는 모형이다. 또한 보다 장기적인 예측모형 실험을 위한 어기 예측모형은 8월부터 9월까지(입식시기)의 데이터를 이용하여 10월부터 이듬해 5월까지의 어기 전체 기간을 예측하기 위한 2-1-8, 그리고 상반기(1월~6월) 데이터를 이용해 4개월 후인 10월부터 어기 전체기간 예측을 위한 6-4-8의 두 가지 시차를 추가로 실험하였다.

〈그림 4-2〉 예측 시차별 종속변수 및 독립변수(예시)



자료: 저자 작성

따라서 세 가지 인공지능망모형, 2가지 은닉층 설정, 네 가지 시차, 그리고 앞서 구분한 아홉 가지의 데이터셋을 조합($3 \times 2 \times 4 \times 9$)하여 총 216가지의 모형을 훈련하였다.

2) 학습 조건

앞서 설정한 216가지 인공신경망모형의 비교를 위해 학습 조건은 모두 동일하게 설정하였다. 우선 데이터의 구분은 예측 대상(종속변수) 기준 2020년 9월 이전의 자료를 훈련데이터로, 2020년 10월부터의 데이터는 시험데이터로 구분하였으며, 훈련데이터는 다시 전체의 무작위 80%는 훈련데이터로, 나머지 20%는 검증데이터로 나누었다. 1회 학습에 투입되는 데이터의 크기(batch size)는 1개월 예측의 경우 4, 어기 전체 예측의 경우 1로 설정했으며, 손실함수로는 평균제곱오차(MSE)를 사용하였다.

활성화 함수로는 ReLU 함수를 사용하였으며, 과적합 방지를 위해 모든 은닉층의 dropout을 50%로 설정하였다. 최적화 방법은 ADAM을 사용하였고 학습률(learning)은 0.001, beta1은 0.9, beta2는 0.999로 설정하였다.

〈표 4-19〉 학습 조건별 세부 설정값

항목	설정값	항목	설정값
훈련:검증데이터	7.5:2.5	학습률	0.001
batch size	4, 1	beta1	0.9
손실함수	MSE	beta2	0.999
활성함수	ReLU	학습 횟수(데이터 집합별)	1,000회
dropout	50%	총학습 횟수	4,000회
최적화 방법	ADAM	교차검증데이터 집합	4개

자료: 저자 작성

비생산시기를 제외하고 실제 여기에 해당하는 훈련데이터에서 검증데이터를 제외하면, 훈련데이터의 시계열 길이는 1개월 예측모형의 경우 100개 미만, 어기 예측모형의 경우 10개~14개 수준에 불과하여 검증데이터의 분류에 따라 오차의 크기가 불안정적일 것으로 판단된다. 따라서 단일 검증데이터가 아닌 교차검증을 통해 편향을 최소화하고자 하였다. 교차검증은 훈련데

이터를 n 개의 집합으로 나눈 뒤 개별 집합을 순차적으로 검증데이터로 활용하는 방법이다. 본 연구에서는 교차검증 데이터를 총 4개 집합으로 구성하였으며, 각 집합별 학습 횟수는 1,000회, 이에 따른 총 학습 횟수 4,000회 중 검증오차가 가장 낮은 모형을 선택하도록 설정하였다.

05

예측모형 추정 결과

제5장에서는 예측모형의 추정 결과에 대해 살펴보았다. 제1절에서는 인공신경망모형의 추정 결과를 지역별로 나누어 예측력을 살펴보았다. 각 지역별·시차별 예측 모형 중 예측력이 높은 모형을 선별해 모형을 통한 예측치와 실제 관측치를 비교하였다. 제2절에서는 제1절의 인공신경망모형과 같은 변수를 사용한 다중회귀모형을 추정하고 예측력 검정 결과를 서술하였다.

제1절 인공신경망모형

제5장에서는 앞서 제4장에서 설계한 바와 같이, 각각의 지역별로 예측시차, 은닉노드 수, 데이터셋으로 구분한 216개의 예측모형을 훈련하였으며, 각각의 예측모형별 변수 및 구조는 <표 5-1>과 같다.

<표 5-1> 예측모형별 구조

구분	내용
기상변수	• 월간강수량, 일 최대강수량, 월 평균풍속, 최대순간풍속, 일조량
수질변수	• 표층수온, 표층용존산소, 표층용존질소, 표층무기인농도, 표층클로로필농도
데이터셋1	• 기상변수(2006년 11월~2021년 6월)
데이터셋2	• 수질변수(2010년 2월~2021년 6월)
데이터셋3	• 기상변수 및 수질변수(2010년 2월 이후)

모형	종속변수		독립변수	
	시점	설명	시점	설명
단기예측모형	t	1개월 단수 예측	t-1	전월 데이터
장기예측모형	t	1개월 단수 예측	t-5	5개월 전 데이터
어기예측모형1	t ~ t+7	10월~5월, 전체 어기	t-1 ~ t-2	8월, 9월 데이터(입식 시기)
어기예측모형2	t ~ t+7	10월~5월, 전체 어기	t-4 ~ t-9	1월~6월(상반기 데이터)

자료: 저자 작성

1. 전남

1) 다층 퍼셉트론

전남지역 단수 예측을 위한 다층 퍼셉트론모형의 학습 결과는 <표 5-2>와 같다. 대체로 훈련오차보다 검증오차 및 시험오차가 낮게 나타나, 과적합 문제는 없는 것으로 보인다. 단기 예측모형의 경우 기상변수와 수질변수를 모두 사용한 모형에서 오차가 작게 나타났다. 검증오차는 각 은닉층의 은닉노드를 100개로 설정한 모형에서 가장 작았으나, 시험오차는 은닉노드를 10개로 설정한 모형에서 더 작은 것으로 나타났다.

다음으로 장기 예측모형은 앞의 경우와 마찬가지로 기상변수와 수질변수를 모두 사용한 모형의 오차가 작게 나타났으며, 검증오차는 각층의 은닉노드를 100개로 설정했을 때 가장 작았다. 그러나 이 경우에도 시험오차는 은닉노드가 10개인 경우가 더 작은 것으로 나타났다. 어기 예측모형1의 경우에는 은닉노드의 수를 100개로 설정하고 기상변수와 수질변수를 모두 사용한 모형에서 검증오차와 시험오차가 모두 가장 작은 것으로 나타났다.

어기 예측모형 2의 경우에는 앞의 경우와 다르게 수질변수만을 사용한 모형에서 검증오차와 시험오차가 가장 작게 나타났으며, 이때 각 층의 은닉노드 수는 100개였다. 단기 및 장기 예측모형과 같은 1개월 예측모형에 비해 어기 예측모형에서 시험오차가 크게 나타났다. 이는 어기 전의 데이

터만을 사용하여 예측한 것으로, 여기 동안의 기상 및 수질 변화는 반영하지 못했기 때문에 풀이할 수 있다.

〈표 5-2〉 다층 퍼셉트론 추정 결과(전남)

모형 구분	은닉 노드	데이터셋	MSE		
			훈련오차	검증오차	시험오차
단기 예측 모형	10	1	37.422	12.510	16.357
		2	30.947	16.540	18.613
		3	38.128	13.028	12.380
	100	1	15.084	4.926	45.082
		2	16.168	4.799	30.720
		3	12.258	1.944	42.514
장기 예측 모형	10	1	30.345	10.779	11.878
		2	23.103	11.292	26.456
		3	29.744	9.526	11.146
	100	1	34.316	4.608	19.764
		2	13.113	5.680	22.394
		3	13.297	4.102	13.127
여기 예측 모형1	10	1	73.070	9.708	36.089
		2	102.368	5.826	20.717
		3	158.846	6.190	20.264
	100	1	23.799	3.505	28.890
		2	76.660	5.804	26.196
		3	86.194	3.390	13.657
여기 예측 모형2	10	1	62.303	10.227	32.317
		2	102.620	28.059	20.484
		3	68.508	28.023	67.596
	100	1	23.290	3.911	17.945
		2	25.563	5.124	20.780
		3	134.517	4.232	41.192

주1) 음영 칸은 모형의 예측시차 형태별, 데이터 분류(훈련/검증/시험)별로 가장 낮은 MSE를 표시한 것임

2) 은닉노드는 각 은닉층의 은닉노드 수를 의미함(모든 모형의 은닉층은 4개로 동일하게 설정하였음)

3) 데이터셋은 다음과 같이 분류됨 1: 기상변수(2006. 11~2021. 6), 2: 수질변수(2010. 2~2021. 6), 3: 기상+수질변수(2010. 2~2021. 6)

자료: 저자 작성

2) 장단기 기억신경망

전남지역 단수 예측을 위한 장단기 기억신경망모형의 학습 결과는 <표 5-3>과 같다. 다층 퍼셉트론모형에서와 마찬가지로 검증오차와 시험오차가 훈련오차보다 대체로 낮게 나타나므로 과적합되지 않은 것으로 판단된다.

단기 예측모형의 경우 검증오차가 가장 낮은 모형은 은닉노드를 100개로 설정하고 수질변수만을 사용한 경우였으며, 시험오차가 가장 낮은 모형은 은닉노드를 10개로 설정하고 기상변수만을 사용한 모형이었다.

장기 예측모형의 경우에도 은닉노드를 100개로 설정하고 수질변수만을 사용한 모형에서 검증오차가 가장 낮았다. 반면 시험오차는 은닉노드를 10개로 설정하고 기상변수와 수질변수를 모두 사용한 모형에서 가장 낮은 것으로 나타났다.

여기 예측모형1에서는 은닉노드를 100개로 설정하고 수질변수만을 사용했을 때 검증오차가 가장 낮았으며, 시험오차는 은닉노드를 100개로 설정하고 기상변수와 수질변수를 모두 사용한 모형에서 가장 낮았다.

여기 예측모형2의 경우에도 앞의 경우와 마찬가지로 은닉노드를 100개로 설정하고 수질변수만을 사용한 모형에서 검증오차가 가장 낮게 나타났다. 시험오차는 은닉노드를 100개로 설정하고 기상변수만을 사용한 경우에 가장 낮았다.

장단기 기억신경망을 학습한 결과 수질변수만을 사용한 경우의 훈련오차 및 검증오차가 기상변수를 사용하거나 기상변수와 수질변수를 둘 다 사용한 것에 비해 훨씬 낮은 오차를 나타내었다. 반면 시험오차는 기상변수가 포함된 모형에서 더 낮게 나타났다.

〈표 5-3〉 장단기 기억신경망 추정 결과(전남)

모형 구분	은닉 노드	데이 터셋	MSE		
			훈련오차	검증오차	시험오차
단기 예측 모형	10	1	39.219	17.596	15.590
		2	18.852	17.385	22.035
		3	35.807	16.964	22.536
	100	1	110.615	64.753	64.902
		2	14.940	3.490	16.731
		3	53.672	32.540	19.648
장기 예측 모형	10	1	41.234	15.397	20.148
		2	19.659	6.960	22.564
		3	30.356	20.100	19.611
	100	1	108.722	54.592	157.393
		2	19.988	5.747	22.483
		3	126.140	36.016	136.780
여기 예측 모형1	10	1	40.372	20.804	19.077
		2	24.343	5.717	27.526
		3	28.595	9.397	16.684
	100	1	821.441	9.914	42.919
		2	4.286	0.421	37.847
		3	19.146	1.157	13.439
여기 예측 모형2	10	1	30.115	18.268	20.715
		2	11.648	4.139	22.757
		3	17.782	3.263	37.533
	100	1	27.093	4.333	18.020
		2	7.038	0.254	72.378
		3	10.073	1.413	29.271

주1) 음영 칸은 모형의 예측시차 형태별, 데이터 분류(훈련/검증/시험)별로 가장 낮은 MSE를 표시한 것임
 2) 은닉노드는 각 은닉층의 은닉노드 수를 의미함(모든 모형의 은닉층은 4개로 동일하게 설정하였음)
 3) 데이터셋은 다음과 같이 분류됨 1: 기상변수(2006. 11~2021. 6), 2: 수질변수(2010. 2~2021. 6), 3: 기상+수질변수(2010. 2~2021. 6)
 자료: 저자 작성

3) 양방향 장단기 기억신경망

전남지역 단수 예측을 위한 양방향 장단기 기억신경망모형의 학습 결과는 <표 5-4>와 같다. 앞의 경우와 마찬가지로 대체로 훈련오차에 비해 검증오차와 시험오차가 낮은 것으로 나타났다.

단기 예측모형의 경우는 은닉층별 은닉노드를 각각 100개로 설정하고 수질변수를 독립변수로 활용한 모형의 검증오차가 가장 낮았고, 시험오차는 은닉노드 100개에 기상변수를 활용한 모형에서 가장 낮은 것으로 나타났다.

장기 예측모형의 경우에도 은닉노드를 100개로 수질변수를 사용한 모형의 검증오차가 가장 낮게 나타났으며, 시험오차가 가장 낮은 모형은 은닉노드 100개에 기상변수를 사용한 모형이었다.

여기 예측모형1에서는 은닉노드 100개에 수질변수만을 사용한 모형의 검증오차가 가장 낮았으며, 시험오차가 가장 낮았던 모형은 은닉노드 10개에 수질변수와 기상변수를 모두 사용한 모형이었다.

여기 예측모형2에서는 각 층의 은닉노드를 100개씩 설정하고 수질변수를 활용한 모형의 검증오차가 가장 낮았으며, 시험오차는 은닉노드를 10개로 설정하고 기상변수를 사용한 모형에서 가장 낮았다.

양방향 장단기 기억신경망모형의 훈련 결과 검증오차는 장단기 기억신경망과 같이 수질변수를 사용한 모형에서 월등히 낮게 나타났으나, 시험오차는 기상변수가 들어간 모형에서 낮게 나타났다. 훈련데이터와 검증데이터는 임의로 분류되지만, 시험데이터는 2021년산으로 한정되어 있다는 측면에서 생각해보면, 2020년산까지의 전남지역 김 단수 추정에는 수질변수를 사용한 장단기 기억신경망모형이 효과적이었지만 2021년산의 설명을 위해서는 기상요인을 반영할 필요가 있었다고 해석할 수 있다.

〈표 5-4〉 양방향 장단기 기억신경망 추정 결과(전남)

모형 구분	은닉 노드	데이 터셋	MSE		
			훈련오차	검증오차	시험오차
단기 예측 모형	10	1	79.813	38.575	51.267
		2	49.397	40.853	44.851
		3	51.774	39.004	51.407
	100	1	18.715	7.059	11.445
		2	16.141	4.318	16.021
		3	46.160	11.498	22.387
장기 예측 모형	10	1	99.849	37.930	71.642
		2	51.923	32.090	29.518
		3	90.121	74.583	54.744
	100	1	29.819	11.314	19.353
		2	11.247	4.048	36.765
		3	28.888	7.017	26.893
여기 예측 모형1	10	1	23.896	9.331	17.352
		2	7.838	2.305	25.563
		3	22.709	6.554	13.729
	100	1	26.386	13.549	51.733
		2	7.660	0.773	29.084
		3	93.946	14.251	70.273
여기 예측 모형2	10	1	15.099	8.942	20.126
		2	3.043	1.507	70.328
		3	13.668	3.822	27.476
	100	1	67.504	5.797	37.681
		2	1.429	0.175	57.126
		3	52.780	10.305	157.594

주1) 음영 칸은 모형의 예측시차 형태별, 데이터 분류(훈련/검증/시험)별로 가장 낮은 MSE를 표시한 것임

2) 은닉노드는 각 은닉층의 은닉노드 수를 의미함(모든 모형의 은닉층은 4개로 동일하게 설정하였음)

3) 데이터셋은 다음과 같이 분류됨 1: 기상변수(2006. 11~2021. 6), 2: 수질변수(2010. 2~2021. 6), 3: 기상+수질변수(2010. 2~2021. 6)

자료: 저자 작성

2. 완도

1) 다층 퍼셉트론

완도지역 단수 예측을 위한 다층 퍼셉트론모형의 학습 결과는 <표 5-5>와 같다. 훈련오차보다 검증오차가 낮게 나타나지만, 일부에서 시험오차가 훈련 오차나 검증오차보다 월등히 높게 나타나 다소 과적합된 경향을 보였다.

단기 예측모형에서 검증오차가 가장 낮았던 모형은 은닉노드 100개에 기상변수와 수질변수를 모두 이용한 경우였으며, 시험오차는 은닉노드 10개에 기상변수와 수질변수를 모두 투입한 모형이 가장 작았다.

장기 예측모형에서 검증오차는 은닉노드의 수를 100개로 하여 수질변수와 기상변수를 모두 사용한 모형이 가장 낮았으며, 시험오차는 은닉노드의 수를 10개로 설정하고 수질변수와 기상변수를 모두 사용한 모형이 가장 낮았다.

여기 예측모형1의 경우 검증오차가 가장 작았던 모형은 은닉노드의 수를 100개로 하여 기상변수를 투입변수로 사용한 경우였으며, 시험오차는 100개의 은닉노드가 설정되어 있고 수질변수를 사용한 모형에서 가장 작게 나타났다.

다음으로 여기 예측모형2의 경우에는 은닉노드의 수를 100개로 설정하고 기상변수만을 사용한 모형이 검증오차가 가장 낮았고, 시험오차는 은닉노드의 수를 10개로 하고 기상변수를 사용한 모형에서 가장 낮았다.

모든 모형에서 은닉노드 100개인 경우의 검증오차가 낮게 나타났으나, 시험오차는 대부분 높게 나타나는 경향이 있었다. 이는 완도지역의 단수 예측을 위해 현재의 데이터를 이용할 경우 은닉노드를 100개로 설정하는 것은 과적합을 유발할 수 있다고 해석할 수 있다.

〈표 5-5〉 다층 퍼셉트론 추정 결과(완도)

모형 구분	은닉 노드	데이 터셋	MSE		
			훈련오차	검증오차	시험오차
단기 예측 모형	10	1	142.214	81.807	92.874
		2	85.706	49.003	128.741
		3	98.431	73.585	78.306
	100	1	62.363	37.489	152.342
		2	71.689	34.349	128.120
		3	83.804	16.693	117.234
장기 예측 모형	10	1	138.372	81.513	44.960
		2	78.575	64.347	151.803
		3	73.956	64.045	98.155
	100	1	124.607	39.849	83.614
		2	65.608	35.628	84.974
		3	47.049	19.281	117.379
여기 예측 모형1	10	1	158.601	75.680	45.168
		2	126.240	60.313	80.156
		3	128.588	62.173	73.351
	100	1	120.528	15.982	54.163
		2	104.087	24.406	34.554
		3	128.030	25.737	77.263
여기 예측 모형2	10	1	196.221	89.518	39.065
		2	216.167	66.177	67.386
		3	114.541	66.979	80.615
	100	1	134.761	18.340	61.075
		2	103.323	22.843	84.817
		3	90.204	27.879	109.788

주1) 음영 칸은 모형의 예측시차 형태별, 데이터 분류(훈련/검증/시험)별로 가장 낮은 MSE를 표시한 것임

2) 은닉노드는 각 은닉층의 은닉노드 수를 의미함(모든 모형의 은닉층은 4개로 동일하게 설정하였음)

3) 데이터셋은 다음과 같이 분류됨 1: 기상변수(2006. 11~2021. 6), 2: 수질변수(2010. 2~2021. 6), 3: 기상+수질변수(2010. 2~2021. 6)

자료: 저자 작성

2) 장단기 기억신경망

완도지역 단수 예측을 위한 장단기 기억신경망모형의 학습 결과는 <표 5-6>과 같다. 앞의 경우와 마찬가지로 일부 모형에서 시험오차가 훈련오차 및 검증오차에 비해 월등히 높게 나타나 과적합이 의심된다.

단기 예측모형의 경우 은닉노드의 수를 100개로 설정하여 기상변수와 수질변수를 설명변수로 사용한 모형의 검증오차가 가장 작았다. 시험오차는 은닉노드를 100개로, 투입변수는 수질변수만을 사용한 모형에서 가장 낮았다.

장기 예측모형의 경우에는 은닉노드의 수가 100개이고 기상변수와 수질변수를 모두 사용한 모형의 검증오차가 가장 작았고, 시험오차는 은닉노드를 10개로 하여 기상변수만을 사용한 모형에서 가장 작았다.

여기 예측모형1에서는 은닉노드를 100개로 설정하고 수질변수만을 사용한 모형에서 검증오차가 0.809로 극히 작았으나, 시험오차는 53.226으로 훈련오차보다 2배 이상 크게 나타났다. 시험오차가 가장 작았던 모형은 은닉노드를 10개로 하여 기상변수만을 사용한 모형이었다. 해당 모형은 반대로 훈련오차와 검증오차에 비해 시험오차가 월등히 작았다.

여기 예측모형2는 앞의 경우와 유사하게 은닉노드를 100개로 설정하고 수질변수만을 사용한 모형에서 검증오차가 0.872로 가장 작았다. 시험오차는 은닉노드를 10개로 하여 기상변수를 사용한 모형에서 가장 작게 나타났다.

다층 퍼셉트론모형에서와 마찬가지로 은닉노드가 100개인 모형의 검증오차가 10개인 경우에 비해 대체로 작게 나타났으나, 시험오차는 오히려 더 큰 경우가 많았다.

〈표 5-6〉 장단기 기억신경망 추정 결과(완도)

모형 구분	은닉 노드	데이 터셋	MSE		
			훈련오차	검증오차	시험오차
단기 예측 모형	10	1	123.173	76.347	53.313
		2	105.603	79.921	134.101
		3	110.630	76.438	129.676
	100	1	515.609	201.124	242.956
		2	153.897	43.454	29.892
		3	70.414	20.659	97.013
장기 예측 모형	10	1	143.389	100.735	58.768
		2	79.621	64.514	95.144
		3	69.226	73.783	189.269
	100	1	119.693	73.851	82.843
		2	76.007	47.130	101.666
		3	82.931	27.608	182.848
여기 예측 모형1	10	1	112.839	84.684	15.372
		2	69.397	16.475	139.806
		3	134.358	38.496	56.432
	100	1	125.950	63.424	26.997
		2	23.959	0.809	53.226
		3	70.842	1.543	23.381
여기 예측 모형2	10	1	77.901	76.991	37.618
		2	49.247	27.867	189.047
		3	47.144	25.234	117.959
	100	1	64.962	4.742	75.830
		2	18.692	0.872	131.122
		3	40.039	5.055	139.516

주1) 음영 칸은 모형의 예측시차 형태별, 데이터 분류(훈련/검증/시험)별로 가장 낮은 MSE를 표시한 것임

2) 은닉노드는 각 은닉층의 은닉노드 수를 의미함(모든 모형의 은닉층은 4개로 동일하게 설정하였음)

3) 데이터셋은 다음과 같이 분류됨 1: 기상변수(2006. 11~2021. 6), 2: 수질변수(2010. 2~2021. 6), 3: 기상+수질변수(2010. 2~2021. 6)

자료: 저자 작성

3) 양방향 장단기 기억신경망

완도지역 단수 예측을 위한 양방향 장단기 기억신경망모형의 학습 결과는 <표 5-7>과 같다. 일부 모형에서 시험오차가 훈련오차 및 검증오차에 비해 월등히 높게 나타나 과적합이 나타난 것으로 보인다.

단기 예측모형의 경우 은닉노드가 10개인 모형보다 100개인 모형의 검증오차가 낮게 나타났다. 그러나 시험오차가 가장 낮은 모형은 은닉노드가 10개면서 기상데이터와 수질데이터를 모두 사용한 모형이었다.

장기 예측모형의 경우에도 은닉노드가 100개인 모형이 10개인 모형보다 검증오차가 작았고 이 중 기상변수만을 이용한 모형에서 가장 작은 것으로 나타났다. 그러나 시험오차는 크게 차이가 없었으며, 시험오차가 가장 낮은 모형은 은닉노드를 10개로 설정하고 기상변수만을 이용한 모형이었다.

여기 예측모형1의 경우 장단기 기억신경망과 비슷하게 수질변수만을 사용한 모형의 검증오차가 다른 모형에 비해 월등히 작게 나타났다. 그러나 시험오차가 가장 작았던 모형은 은닉노드를 100개로 설정하고 수질변수와 기상변수를 모두 사용한 모형이었다.

여기 예측모형2에서도 수질변수를 사용한 모형에서 검증오차가 매우 작게 나타났으나, 시험오차가 가장 작은 모형은 은닉노드를 100개로 설정하고 기상변수만을 이용한 모형이었다.

앞서 장단기 기억신경망에서와 마찬가지로, 여기 예측모형에서 수질변수만을 사용했을 때 훈련오차 및 검증오차는 매우 작아졌지만, 시험오차는 크게 늘어났다. 또한 기상변수만을 사용한 모형에서 시험오차가 작게 나타났는데, 장단기 기억신경망모형으로 완도지역 김 단수를 안정적으로 추정하기 위해서는 기상변수가 포함되어야 함을 유추할 수 있다.

〈표 5-7〉 양방향 장단기 기억신경망 추정 결과(완도)

모형 구분	은닉 노드	데이 터셋	MSE		
			훈련오차	검증오차	시험오차
단기 예측 모형	10	1	382.115	150.465	134.718
		2	222.241	148.776	215.276
		3	175.172	122.646	52.020
	100	1	153.421	39.307	99.795
		2	82.741	26.593	155.684
		3	65.253	26.525	78.122
장기 예측 모형	10	1	274.153	125.817	133.661
		2	346.814	196.205	137.422
		3	349.897	224.818	229.070
	100	1	129.131	49.902	143.422
		2	245.359	120.377	226.770
		3	169.422	144.859	229.506
여기 예측 모형1	10	1	129.675	69.261	61.911
		2	56.956	10.088	78.895
		3	70.169	31.153	77.035
	100	1	164.032	130.120	84.830
		2	21.354	2.131	60.347
		3	297.255	59.014	18.470
여기 예측 모형2	10	1	116.207	74.024	39.793
		2	38.106	2.313	157.048
		3	12.952	40.149	180.555
	100	1	140.551	71.486	39.623
		2	63.234	3.106	205.271
		3	347.008	62.281	47.557

주1) 음영 칸은 모형의 예측시차 형태별, 데이터 분류(훈련/검증/시험)별로 가장 낮은 MSE를 표시한 것임

2) 은닉노드는 각 은닉층의 은닉노드 수를 의미함(모든 모형의 은닉층은 4개로 동일하게 설정하였음)

3) 데이터셋은 다음과 같이 분류됨 1: 기상변수(2006. 11~2021. 6), 2: 수질변수(2010. 2~2021. 6), 3: 기상+수질변수(2010. 2~2021. 6)

자료: 저자 작성

3. 고흥

1) 다층 퍼셉트론

고흥지역 김 단수 예측을 위한 다층 퍼셉트론모형의 학습 결과는 <표 5-8>과 같다. 대체로 은닉노드의 수가 100개일 때의 검증오차가 작게 나타났다으나, 시험오차는 은닉노드 수에 따른 차이가 불규칙적으로 나타났다.

단기 예측모형의 경우 은닉노드가 100개일 때 기상변수와 수질변수를 모두 사용한 모형에서 검증오차가 가장 작게 나타났다. 반면 시험오차가 가장 작았던 모형은 은닉노드를 100개로 설정하여 수질변수만을 사용한 모형이었다.

장기 예측모형도 마찬가지로 은닉노드가 100개이고 기상변수와 수질변수를 모두 사용한 모형에서 검증오차가 가장 작았으나, 시험오차는 은닉노드를 10개로 설정하여 기상변수만을 사용한 모형에서 가장 작았다.

여기 예측모형1의 경우에서도 은닉노드를 100개로 설정하고 수질변수와 기상변수를 모두 사용한 모형의 검증오차가 작았다. 이 경우 시험오차는 은닉노드를 10개로 설정하고 수질변수만을 사용한 모형에서 가장 작은 것으로 나타났다.

여기 예측모형2의 경우에도 은닉노드 100개에 기상변수와 수질변수를 모두 사용한 모형의 검증오차가 가장 작았으며, 시험오차는 은닉노드 10개에 기상변수와 수질변수를 사용한 모형이 가장 작았다.

검증오차만을 보면 고흥지역의 김 단수 예측모형 훈련에서는 은닉노드를 100개로 하여 기상변수와 수질변수가 모두 사용되었을 때 가장 낮았으나, 시험오차는 오히려 은닉노드가 100개이거나 기상 또는 수질변수 중 한 가지만 선택한 모형에서 더 낮게 나타나는 경향이 있었다.

〈표 5-8〉 다층 퍼셉트론 추정 결과(고흥)

모형 구분	은닉 노드	데이 터셋	MSE		
			훈련오차	검증오차	시험오차
단기 예측 모형	10	1	208.732	90.002	52.327
		2	206.350	64.411	52.845
		3	113.999	65.750	105.737
	100	1	173.039	56.769	63.990
		2	132.552	50.106	43.914
		3	110.550	16.533	64.771
장기 예측 모형	10	1	180.827	56.973	51.923
		2	186.328	83.145	54.450
		3	173.238	75.435	64.006
	100	1	163.776	38.875	57.253
		2	255.387	81.238	59.616
		3	107.296	18.047	78.716
여기 예측 모형1	10	1	400.167	42.942	130.176
		2	228.666	58.161	33.117
		3	233.878	61.091	33.349
	100	1	143.687	17.245	65.898
		2	122.858	16.468	37.128
		3	55.131	7.596	50.296
여기 예측 모형2	10	1	318.822	66.610	69.359
		2	244.519	109.961	21.106
		3	319.526	103.955	20.201
	100	1	1097.637	8.680	1236.066
		2	10809.909	18.653	33.507
		3	96.311	3.756	269.319

주1) 음영 칸은 모형의 예측시차 형태별, 데이터 분류별로 가장 낮은 MSE를 표시한 것임

주2) 은닉노드는 각 은닉층의 은닉노드 수를 의미함(모든 모형의 은닉층은 4개로 동일하게 설정하였음)

주3) 데이터셋은 다음과 같이 분류됨 1: 기상변수(2006. 11.~2021. 6.), 2: 수질변수(2010. 2.~2021. 6.),

3: 기상+수질변수(2010. 2.~2021. 6.)

자료: 저자 작성

2) 장단기 기억신경망

고흥지역 김 단수 예측을 위한 장단기 기억신경망의 학습 결과는 <표 5-9>와 같다. 은닉노드가 10개일 때보다 100개일 때의 검증오차가 낮게 나타났다으며, 단기 및 장기 예측모형보다 어기 예측모형에서 이러한 경향이 두드러지게 나타났다.

단기 예측모형의 경우 검증오차는 은닉노드를 100개로 설정하여 기상변수와 수질변수를 모두 사용한 모형에서 가장 작았다. 시험오차가 가장 낮은 모형은 은닉노드 100개에 수질변수만을 이용한 모형이었다.

장기예측에서도 마찬가지로 은닉노드의 수를 100개로 설정하고 기상변수와 수질변수를 모두 사용한 모형에서 검증오차가 가장 낮았다. 그러나 시험오차가 가장 낮았던 모형은 은닉노드를 10개로 설정하고 기상변수만을 사용한 모형이었다.

어기 예측모형1의 경우 은닉노드를 100개로 설정하고 수질변수만을 사용한 모형의 검증오차가 가장 낮았으며, 시험오차는 은닉노드를 100개로 설정하고 기상변수만을 사용한 모형에서 가장 낮았다.

어기 예측모형2의 경우 은닉노드가 100개일 때 수질변수만을 사용한 모형의 검증오차가 가장 낮았으며, 은닉노드를 10개로 하여 기상변수만을 사용한 모형에서 시험오차가 가장 작게 나타났다.

다른 지역의 경우와 마찬가지로 고흥지역의 장단기 기억신경망모형 학습에서도 은닉노드의 수를 100개로 설정하고 수질변수만을 사용한 모형에서 검증오차가 매우 작았으나, 해당 모형의 시험오차는 비교적 크게 나타났다.

〈표 5-9〉 장단기 기억신경망 추정 결과(고흥)

모형 구분	은닉 노드	데이 터셋	MSE		
			훈련오차	검증오차	시험오차
단기 예측 모형	10	1	202.641	117.787	31.011
		2	180.925	102.043	77.743
		3	148.155	65.479	98.689
	100	1	196.736	81.811	58.297
		2	99.349	40.963	26.599
		3	104.006	15.439	59.203
장기 예측 모형	10	1	191.002	54.777	30.935
		2	187.849	88.133	48.654
		3	191.199	85.473	69.230
	100	1	182.763	55.689	58.734
		2	202.288	97.029	69.833
		3	123.266	14.607	87.547
여기 예측 모형1	10	1	197.200	99.390	39.025
		2	172.378	55.291	41.915
		3	111.662	33.552	60.596
	100	1	90.193	8.052	37.051
		2	49.366	1.396	214.063
		3	41.768	6.045	150.024
여기 예측 모형2	10	1	96.323	69.375	49.301
		2	868.338	8.594	128.512
		3	193.622	102.881	53.370
	100	1	68.971	12.667	52.152
		2	107.753	1.158	406.198
		3	25.815	2.848	427.166

주1) 음영 칸은 모형의 예측시차 형태별, 데이터 분류별로 가장 낮은 MSE를 표시한 것임

주2) 은닉노드는 각 은닉층의 은닉노드 수를 의미함(모든 모형의 은닉층은 4개로 동일하게 설정하였음)

주3) 데이터셋은 다음과 같이 분류됨 1: 기상변수(2006. 11.~2021. 6.), 2: 수질변수(2010. 2.~2021. 6.),

3: 기상+수질변수(2010. 2.~2021. 6.)

자료: 저자 작성

3) 양방향 장단기 기억신경망

고흥지역 김 단수 예측을 위한 양방향 장단기 기억신경망의 학습 결과는 <표 5-10>과 같다. 일부 모형에서 시험오차가 매우 크게 나타나 과적합된 것으로 보이지만, 대체로 훈련오차에 비해 검증오차와 시험오차가 작게 나타났다.

단기 예측모형 비교에서 검증오차가 가장 낮은 모형은 은닉노드의 수를 100개로 하고 투입변수는 기상변수와 수질변수를 모두 사용한 모형이었다. 다음으로 시험오차가 가장 낮았던 모형은 은닉노드의 수를 10개로 설정하고 수질변수만을 사용한 모형이었다.

장기 예측모형의 경우에도 은닉노드의 수를 100개로 설정하고 기상변수와 수질변수를 모두 사용한 모형의 검증오차가 가장 낮았으며, 시험오차 역시 동일한 모형에서 가장 작게 나타났다.

여기 예측모형1에서는 다른 지역의 여기 예측모형과 마찬가지로 은닉노드를 100개로 설정하고 수질변수만을 투입한 모형의 검증오차가 매우 낮게 나타났다으며, 시험오차는 은닉노드를 10개로 설정하고 기상변수만을 사용한 모형에서 가장 낮았다.

여기 예측모형2의 경우 검증오차가 가장 낮았던 모형은 은닉노드를 10개로 설정하고 수질변수를 사용한 모형이었으며, 시험오차가 가장 낮았던 모형은 은닉노드의 수를 10개로 하고 기상변수만을 사용한 모형이었다.

양방향 장단기 신경망에서는 은닉노드를 100개로 설정한 경우의 검증오차가 10개인 경우보다 낮게 나타나는 경향이 있었으나, 시험오차는 오히려 은닉노드가 10개일 때 적게 나타났다.

〈표 5-10〉 양방향 장단기 기억신경망 추정 결과(고흥)

모형 구분	은닉 노드	데이 터셋	MSE		
			훈련오차	검증오차	시험오차
단기 예측 모형	10	1	337.520	183.324	196.102
		2	85.184	59.331	37.967
		3	816.861	240.347	396.959
	100	1	140.755	56.870	110.328
		2	120.755	64.794	81.201
		3	136.816	11.907	118.308
장기 예측 모형	10	1	337.582	125.206	186.914
		2	226.628	205.951	159.325
		3	201.831	339.529	247.274
	100	1	152.082	69.229	83.189
		2	238.157	101.748	125.078
		3	109.966	20.263	53.232
여기 예측 모형1	10	1	195.442	34.005	40.955
		2	88.442	31.722	153.230
		3	94.962	41.078	293.926
	100	1	80.547	26.579	86.215
		2	31.968	4.998	289.169
		3	29.421	38.667	1405.759
여기 예측 모형2	10	1	98.060	79.662	54.074
		2	73.612	4.022	390.831
		3	61.321	39.004	170.801
	100	1	295.408	26.753	125.771
		2	159.977	10.338	141.669
		3	264.635	109.973	65.354

주1) 음영 칸은 모형의 예측시차 형태별, 데이터 분류별로 가장 낮은 MSE를 표시한 것임

주2) 은닉노드는 각 은닉층의 은닉노드 수를 의미함(모든 모형의 은닉층은 4개로 동일하게 설정하였음)

주3) 데이터셋은 다음과 같이 분류됨 1: 기상변수(2006. 11.~2021. 6.), 2: 수질변수(2010. 2.~2021. 6.), 3: 기상+수질변수(2010. 2.~2021. 6.)

자료: 저자 작성

4. 종합

앞서 실시한 지역별 김 단수 예측모형에서는 전반적으로 훈련오차보다 검증오차와 시험오차가 작게 나타났는데, 이는 <그림 5-1>과 같이 훈련데이터와 검증데이터를 교차하며 훈련에 사용하는 교차검증 방식을 사용했기 때문이다. 한 세트의 교차 검증데이터가 사용되고 나면 그 다음 훈련부터는 훈련데이터로 충분히 사용된 데이터가 검증데이터로 사용된다. 이때 가중치는 이미 훈련된 검증데이터로부터 얻은 가중치를 갱신하는 과정을 거치므로 검증오차가 낮아질 수밖에 없다.

<그림 5-1> 교차검증데이터의 순환



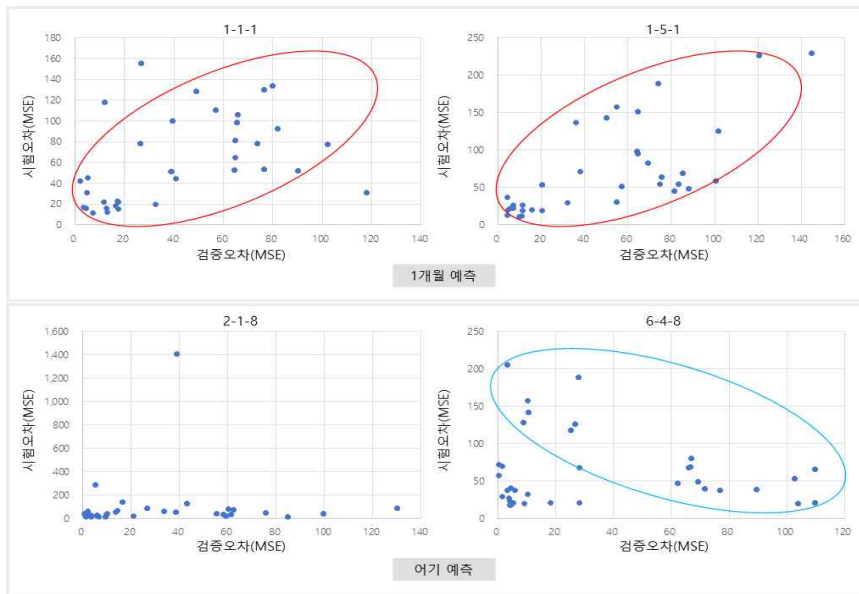
자료: 저자 작성

일반적으로 검증오차가 낮을수록 모형의 예측력이 높을 거라고 추정할 수 있으나, 본 연구에서는 교차검증을 이용함으로써 검증데이터 역시 과적합될 수 있다는 문제점이 있다. 특히 단기 및 장기 예측모형과 같은 1개월 예측모형보다 훨씬 더 적은 데이터로 훈련하는 단기 예측모형1과 2에서 이러한 현상이 뚜렷하게 나타났다.

1개월 예측모형의 경우 시차에 따라 100개 전후의 월 단위 훈련데이터가 사용되었으며, 전반적으로 검증오차가 커짐에 따라 시험오차도 커지는 경향

이 있었다. 그러나 여기 예측모형에서는 모형의 시차에 따라 10~14개의 비교적 적은 수의 어기(연도)별 자료가 훈련데이터로 사용되기 때문에, 은닉노드 10개 또는 100개 등 복잡한 모형의 구조에 비해 적은 데이터가 훈련에 사용되어 훈련데이터의 과적합이 나타나기 쉽다. 실제 검증오차와 시험오차의 상관관계가 1개월 예측모형과 달리 거의 나타나지 않았고, 오히려 부분적으로는 검증오차가 작을 때 시험오차가 더 커지는 추세를 나타내기도 했다.

〈그림 5-2〉 예측시차별 시험오차와 검증오차의 산포도



자료: 저자 작성

따라서 본 연구에서는 학습에 전혀 사용되지 않은 시험데이터와 관측치를 이용해 계산한 MSE(시험오차)를 검증지표로 사용하는 것이 적절하다고 판단하였다. 앞의 예측모형 추정 결과를 바탕으로 지역별 모형 구분별로 가장 시험오차가 낮은 모형을 추려내어 2021년산 김 단수를 추정하고 각각의 예측력을 비교하였다.

예측력 비교를 위한 지표로는 MSE(평균제곱오차), MAPE(평균절대비율오차), Theil's U(테일의 불균등 계수)값을 살펴보았다. MSE는 오차(관측치-예측치)의 크기, MAPE는 오차의 비율(오차/관측치)을 나타내며, Theil's U 계수 앞의 두 지표와 마찬가지로 예측의 정확도와 더불어 오차의 원인이 무엇인지 살펴보기 위한 지표로 활용할 수 있다.

$$MSE = \frac{1}{N} \sum_{i=1}^N (Y_i - f_i)^2 \quad [\text{식 5-1}]$$

$$MAPE = \frac{1}{N} \sum_{i=1}^N \left(\frac{Y_i - f_i}{Y_i} \right) \quad [\text{식 5-2}]$$

$$Theil's U = \frac{\sqrt{\frac{1}{N} \sum_{i=1}^N (Y_i - f_i)^2}}{\sqrt{\frac{1}{N} \sum_{i=1}^N Y_i^2} + \sqrt{\frac{1}{N} \sum_{i=1}^N f_i^2}} \quad [\text{식 5-3}]$$

이 중 Theil's U 계수는 아래와 같이 편의비율(U_M), 분산비율(U_S), 공분산비율(U_C)의 세 가지 편불균등계수로 나뉜다.

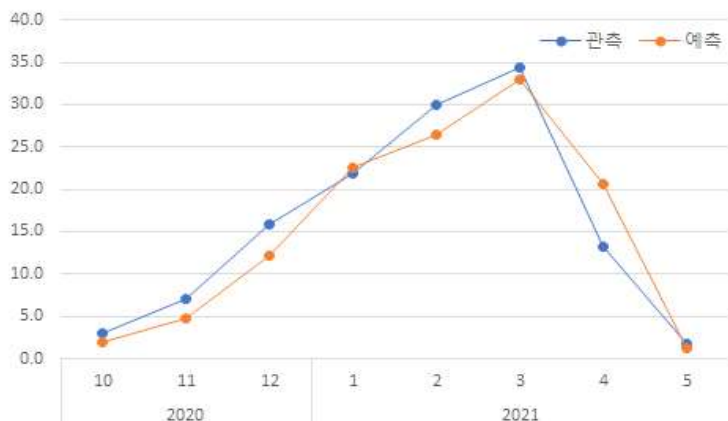
[식 5-4]

$$U_M = \frac{(\bar{Y} - \bar{f})^2}{\frac{1}{N} \sum_{i=1}^N (Y_i - f_i)^2}, U_S = \frac{(s_y - s_f)^2}{\frac{1}{N} \sum_{i=1}^N (Y_i - f_i)^2}, U_C = \frac{2(1 - \rho)s_y s_f}{\frac{1}{N} \sum_{i=1}^N (Y_i - f_i)^2}$$

U_M 은 편향(bias)에 의한 오차의 비중, U_S 는 분산의 불일치로 인한 오차의 비중, 그리고 U_C 는 불완전한 공분산으로 인한 오차의 비중이다. U_M 이나 U_S 의 비중이 상대적으로 크다면 향후 추가적 정보(시계열 확보 등)를 통해 모형의 예측력을 개선할 수 있으나, U_C 가 크게 나타나는 경우 추가 자료를 통한 모형의 오차를 줄이기 어렵다.¹⁸⁾

18) 이진면 외(2007), Cook(2019), Skaggs&Snyder(1991)

〈그림 5-3〉 전남지역 단기 예측모형 김 단수 추정 결과(인공신경망)



자료: 저자 작성

Theil's U는 0.09로 양호했고, 오차의 원인별 비중을 살펴보면 추정치 범위가 관측치와 비슷했던 만큼 U_M 과 U_S 는 각각 0.02, 0.00으로 작았고, 비체계적 요인을 뜻하는 U_C 가 0.97로 높게 나타나 현재 모형에 추가적인 정보 반영을 통한 개선이 어려울 것으로 추정되었다.

나. 장기 예측

전남지역의 장기 예측에서 가장 낮은 시험오차를 보였던 모형은 과거 기상 및 수질변수를 모두 학습에 사용한 다층 퍼셉트론모형(은닉노드 10개)으로, 모형의 시험오차는 11.146이었다.

해당 모형을 이용해 2021년산 전남지역의 김 단수를 예측한 결과는 〈표 5-12〉 및 〈그림 5-4〉와 같다. 본 모형은 예측 목표 시점을 t 라고 할 때, $t-5$ 시점(5개월 전)의 독립변수를 사용한다. 따라서 예측에 사용된 데이터는 2020년 5월부터 2020년 12월까지의 데이터이다. 예측값은 10월부터 단수가 증가하기 시작하여 2월에 정점에 달하고 3월부터 하락하여 실제 관측치와는 다소 상이한 결과를 나타soT다.

관측치의 범위는 1.74~34.36, 평균은 15.89였으며, 모형을 이용한 예측치의 범위는 0~32.89, 평균은 15.55로 분포는 비슷하게 나타났다. MSE는 11.15로 단기 모형에 비해 작았으나, MAPE값은 32.03으로 오히려 더 컸으며 Theil's U계수는 0.09로 같았다. 앞의 모형과 마찬가지로 U_C 값이 크게 나타나 추가적인 정보를 통한 모형 개선이 어려울 것으로 추정된다.

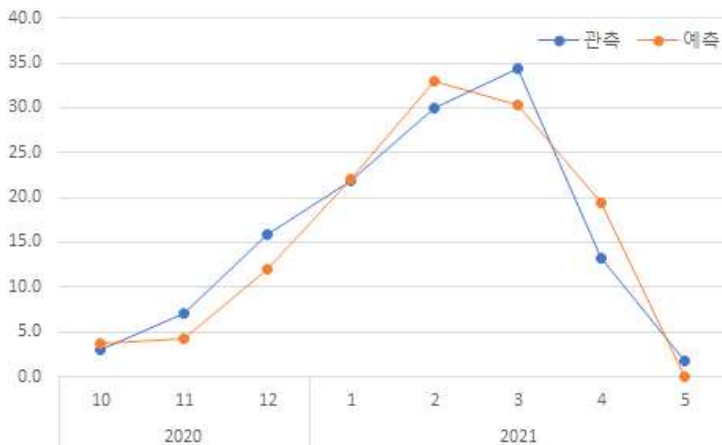
〈표 5-12〉 전남지역 장기 예측모형 김 단수 추정 결과(인공신경망)

단위: 속/책

	2020			2021				
	10	11	12	1	2	3	4	5
관측	2.97	7.07	15.95	21.83	29.99	34.36	13.24	1.74
예측	3.62	4.19	11.94	22.04	32.89	30.39	19.34	0.00
관측-예측	-0.66	2.87	4.01	-0.21	-2.90	3.96	-6.10	1.74
MSE	MAPE		Theil's U		U_M		U_S	
	11.15		32.03		0.01		0.02	
			0.09				0.97	

자료: 저자 작성

〈그림 5-4〉 전남지역 장기 예측모형 김 단수 추정 결과(인공신경망)



자료: 저자 작성

(2) 어기 예측모형

가. 어기 예측모형1

전남지역의 어기별 김 단수 예측을 위한 어기 예측모형1의 조건에서 가장 낮은 시험오차를 보였던 모형은 기상 및 수질변수를 모두 사용한 장단기 기역신경망모형(은닉노드 100개)이었다.

해당 모형을 이용해 2021년산 전남지역의 김 단수를 예측한 결과는 <표 5-13> 및 <그림 5-5>와 같다. 본 모형은 예측 목표 시점을 $t \sim t+7$ 이라고 할 때, $t-1 \sim t-2$ 시점까지의 데이터를 사용한다. 따라서 예측에 사용된 데이터는 2020년 8월부터 2020년 9월까지의 데이터이다. 예측값은 10월부터 단수가 증가하기 시작하여 3월에 정점에 달하고 4, 5월에 하락하는 실제 관측치와 비슷한 값과 방향을 보여주었다.

예측치는 최저 1.36에서 최고 36.6, 평균 16.45의 범위에서 추정되었다. MSE는 13.44로 1개월 예측모형에 비해 높았으며, MAPE는 27.60, Theil's U 계수는 0.09로 1개월 예측모형과 비슷했다. 2개월간의 데이터만을 이용하여 전체 어기를 예측했다는 것은 독립변수와 최대 9개월까지의 시차가 있음을 의미한다. 그럼에도 불구하고 예측값의 방향성이나 MAPE 측면에서 앞의 1개월 예측모형보다 나은 예측력을 나타내었다고 볼 수 있다.

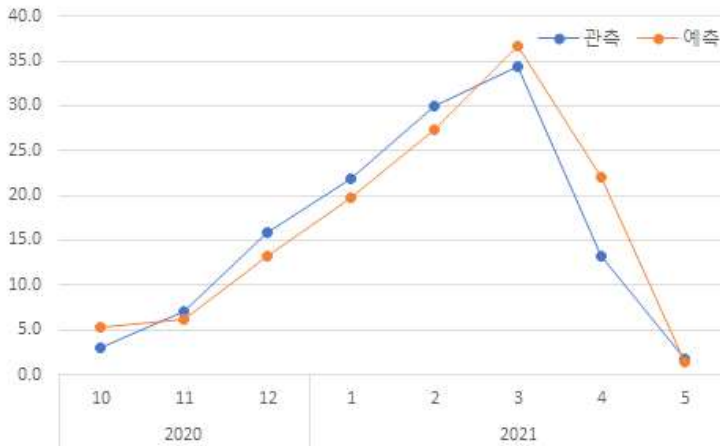
<표 5-13> 전남지역 어기 예측모형1 김 단수 추정 결과(인공신경망)

단위: 속/척

	2020			2021				
	10	11	12	1	2	3	4	5
관측	2.97	7.07	15.95	21.83	29.99	34.36	13.24	1.74
예측	5.26	6.12	13.24	19.71	27.28	36.60	22.02	1.36
관측-예측	-2.29	0.95	2.71	2.12	2.71	-2.24	-8.77	0.37
MSE	MAPE		Theil's U	U _M		U _S	U _C	
	13.44		27.60	0.09		0.02	0.00	

자료: 저자 작성

〈그림 5-5〉 전남지역 어기 예측모형1 김 단수 추정 결과(인공신경망)



자료: 저자 작성

나. 어기 예측모형2

전남지역 어기 예측모형2의 경우 가장 낮은 시험오차를 보였던 모형은 기상데이터를 학습에 사용한 다층 퍼셉트론모형(은닉노드 100개)이었다.

해당 모형을 이용해 2021년산 전남지역의 김 단수를 예측한 결과는 〈표 5-14〉 및 〈그림 5-6〉과 같다. 본 모형은 예측 목표 시점을 t 부터 $t+7$ 까지라고 할 때, $t-4 \sim t-9$ 까지의 6개월치 데이터를 사용한다. 따라서 예측에 사용된 데이터는 2020년 1월부터 2020년 6월까지의 데이터이다. 예측값은 10월부터 단수가 증가하기 시작하여 3월에 정점에 달하고 4, 5월에 하락하는 실제 관측치와 비슷한 값과 방향을 보인다.

예측치는 최저 1.71에서 최고 32.32, 평균은 14.45의 범위를 나타내었다. 관측치보다 평균이 1.44 낮았으며 4월, 10월을 제외하고는 전반적으로 과소 추정되었다. 이때 MSE는 17.74로 다른 모형에 비해 컸으며, Theil's U계수도 0.11로 비교적 높았고 MAPE는 22.45로 비교적 낮았다. 또한 앞의 경우와 마찬가지로 U_c 가 0.82로 높게 나타났다.

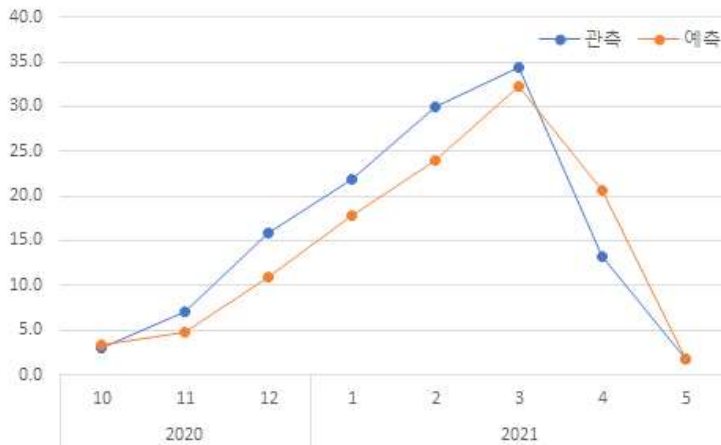
〈표 5-14〉 전남지역 어기 예측모형2 김 단수 추정 결과(인공신경망)

단위: 속/책

	2020			2021				
	10	11	12	1	2	3	4	5
관측	2.97	7.07	15.95	21.83	29.99	34.36	13.24	1.74
예측	3.41	4.84	10.98	17.72	23.92	32.32	20.72	1.71
관측-예측	-0.44	2.22	4.97	4.11	6.07	2.04	-7.47	0.02
MSE	MAPE		Theil's U	U _M		U _S	U _C	
	17.94		22.54	0.11		0.12	0.06	

자료: 저자 작성

〈그림 5-6〉 전남지역 어기 예측모형2 김 단수 추정 결과(인공신경망)



자료: 저자 작성

2) 완도

(1) 1개월 예측모형

가. 단기 예측

완도지역의 단기 예측에서 가장 낮은 시험오차를 보였던 모형은 수질변수를 학습에 사용한 장단기 기억신경망모형(은닉노드 100개)으로, 모형의 시험오차¹⁹⁾는 29.89였다.

해당 모형을 이용해 2021년산 전남지역의 김 단수를 예측한 결과는 <표 5-15> 및 <그림 5-7>과 같다. 본 모형은 예측 목표 시점을 t 라고 할 때, $t-1 \sim t-5$ 까지의 독립변수를 사용한다. 따라서 예측에 사용된 데이터는 2020년 10월부터 2021년 4월까지의 데이터이다. 전반적으로 전남지역 예측 모형에 비해 MSE, MAPE, Theil's U 계수 등이 크게 나타났다.

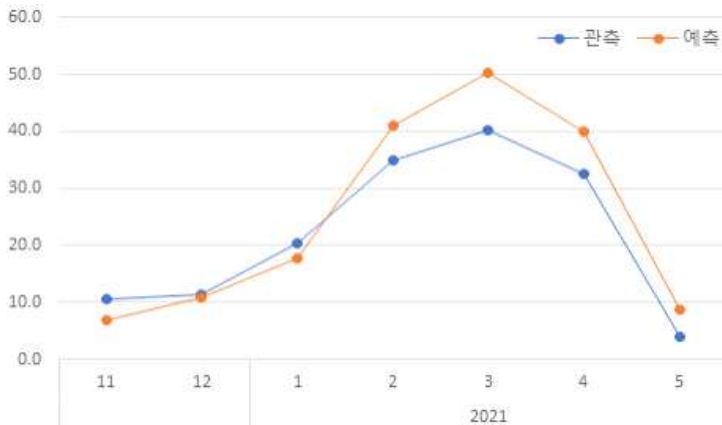
<표 5-15> 완도지역 단기 예측모형 김 단수 추정 결과(인공신경망)

단위: 속/책

	2020		2021				
	11	12	1	2	3	4	5
관측	10.54	11.24	20.24	34.86	40.31	32.41	4.01
예측	6.78	10.72	17.78	41.08	50.20	40.00	8.71
관측-예측	3.76	0.52	2.46	-6.22	-9.89	-7.59	-4.71
MSE	MAPE		Theil's U		U_M		U_S
	33.82		33.68		0.10		0.28
					0.42		0.29

자료: 저자 작성

<그림 5-7> 완도지역 단기 예측모형 김 단수 추정 결과(인공신경망)



자료: 저자 작성

19) 시험오차는 10월부터 5월까지의 단수 예측에 대한 오차였으며, 완도의 2021년산 김 생산 어기는 11월~5월로 <표 5-15>와는 차이가 있음

관측치는 최저 4.01, 최고 40.31, 평균 21.94의 분포를 보였으나, 예측치는 최저 6.78, 최고 50.20, 평균 25.04로 관측치에 비해 높게 나타났다. 특히 생산 비중이 높은 2월 이후 단수에 대한 과대추정이 두드러졌다. 이 때 MSE는 33.82, MAPE는 33.68, Theil's U 계수는 0.10이었다.

Theil' U 계수를 세부적으로 살펴보면, U_M 과 U_S 가 각각 0.28, 0.42로 전남지역 예측모형에서보다 크게 나타났다. 오차에서 비체계적 요인의 비중을 뜻하는 U_C 는 0.29로 비교적 작게 나타나 추가적인 자료를 통해 모형을 개선이 가능할 것임을 알 수 있다.

나. 장기 예측

완도지역의 장기 예측에서 가장 낮은 시험오차를 보였던 모형은 기상데이터만을 이용한 다층 퍼셉트론모형(은닉노드 10개)이며, 해당 모형의 시험오차는 44.96으로 단기 예측모형에 비해 컸다.

해당 모형을 이용해 2021년산 전남지역의 김 단수를 예측한 결과는 <표 5-16> 및 <그림 5-8>과 같다. 본 모형은 예측 목표 시점을 t 라고 할 때, $t-5$ 시점의 데이터를 사용한다. 따라서 예측에 사용된 데이터는 2020년 6월부터 2020년 12월까지의 데이터이다.

<표 5-16> 완도지역 장기 예측모형 김 단수 추정 결과(인공신경망)

단위: 속/책

	2020		2021				
	11	12	1	2	3	4	5
관측	10.54	11.24	20.24	34.86	40.31	32.41	4.01
예측	7.86	7.72	21.94	47.88	45.01	42.90	8.97
관측-예측	2.68	3.52	-1.70	-13.02	-4.71	-10.49	-4.96
MSE	MAPE		Theil's U	U_M	U_S	U_C	
49.82	38.64		0.12	0.34	0.37	0.29	

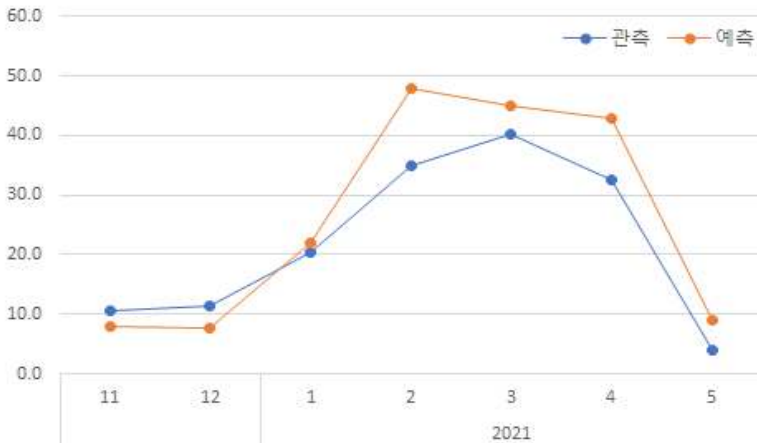
자료: 저자 작성

예측값은 10월부터 단수가 증가하기 시작하여 2월에 정점에 이른 후 하락하는 추세를 보여, 실제 관측치(3월 정점 이후 하락)와는 다소 상이한 값과 변동 추세를 보여주었다.

예측치는 최저 7.72~47.88, 평균은 26.04로 앞의 경우와 마찬가지로 11월 및 12월을 제외하고는 관측치에 비해 과대 추정된 경향을 보였다. 또한 MSE는 49.82, MAPE는 38.64, Theil'U 계수는 0.12로 단기 예측모형에 비해 부정확한 것을 알 수 있었다.

또한 Theil' U 계수를 세부적으로 살펴보면, U_M 과 U_S 가 각각 0.34, 0.37로 앞의 모형과 마찬가지로 크게 나타났다. 비체계적 요인의 비중을 뜻하는 U_C 는 0.29로 비교적 작게 나타나 추가적인 자료를 통해 모형을 개선이 필요하다고 판단되었다.

〈그림 5-8〉 완도지역 장기 예측모형 김 단수 추정 결과(인공신경망)



자료: 저자 작성

(2) 어기 예측모형

가. 어기 예측모형1

완도지역 어기 예측모형1 중 가장 낮은 시험오차를 보였던 모형은 기상 데이터를 학습에 사용한 장단기 기억신경망모형(은닉노드 10개)으로, 모형의 시험오차는 15.37이었다.

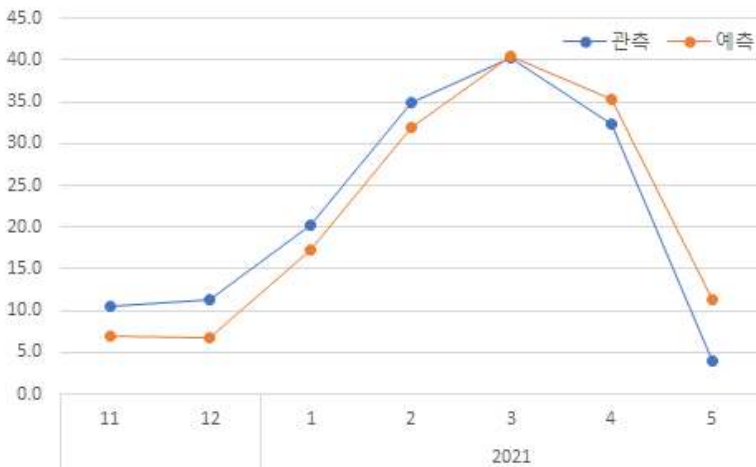
〈표 5-17〉 완도지역 어기 예측모형1 김 단수 추정 결과(인공신경망)

단위: 속/책

	2020		2021				
	11	12	1	2	3	4	5
관측	10.54	11.24	20.24	34.86	40.31	32.41	4.01
예측	6.96	6.82	17.18	32.02	40.47	35.30	11.31
관측-예측	3.59	4.42	3.06	2.84	-0.16	-2.89	-7.31
MSE	MAPE		Theil's U		U _M		U _S
	15.95		41.20		0.08		0.02
					0.00		0.98

자료: 저자 작성

〈그림 5-9〉 완도지역 어기 예측모형1 김 단수 추정 결과(인공신경망)



자료: 저자 작성

예측치의 범위는 최저 6.82, 최고 40.47, 평균 21.44로 앞서 추정한 1개월 모형에 비해 실제 관측치와 더 유사한 흐름을 보였다. MSE는 15.95, MAPE는 41.20, Theil's U계수는 0.08로 예측력 검정지표 역시 1개월 추정 모형에 비해 나은 것으로 나타났다. 또한 U_c 값이 0.98로 비체계적 오차 비중이 크게 나타나 추가 자료를 통한 모형 개선은 한계가 있을 것으로 보인다.

나. 어기 예측모형2

완도지역의 어기 예측모형2 중 가장 예측력이 좋았던 모형은 기상 데이터를 사용한 장단기 기억신경망모형(은닉노드 10개)으로, 시험오차는 37.62였다.

해당 모형을 이용해 2021년산 전남지역의 김 단수를 예측한 결과는 <표 5-18> 및 <그림 5-10>과 같다. 본 모형은 예측 목표 시점을 t 부터 $t+7$ 라고 할 때, $t-4 \sim t-9$ 시점까지의 기상변수를 사용한다. 따라서 예측에 사용된 데이터는 2020년 1월부터 2020년 6월까지의 데이터이다. 아래의 표는 완도 지역의 2021년산 어기에 해당하는 2020년 11월부터 2021년 5월까지의 예측치 및 이를 바탕으로 산출한 예측력 검정지표를 나타낸다.

<표 5-18> 완도지역 어기 예측모형2 김 단수 추정 결과(인공신경망)

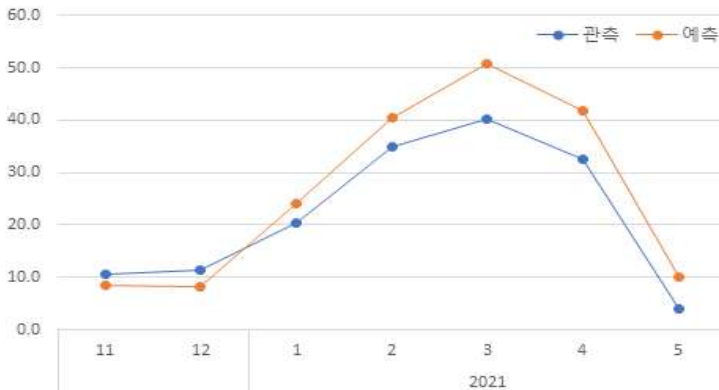
	2020		2021					단위: 속/책
	11	12	1	2	3	4	5	
관측	10.54	11.24	20.24	34.86	40.31	32.41	4.01	
예측	8.46	8.31	24.12	40.45	50.89	41.67	9.99	
관측-예측	2.08	2.94	-3.88	-5.59	-10.58	-9.26	-5.98	
MSE	MAPE		Theil's U		U_M		U_s	U_c
	41.85		40.76		0.11		0.45	0.33
								0.22

자료: 저자 작성

예측치는 3월에 정점에 달하고 이후 하락하는 실제 관측치와 비슷한 흐름을 보였으나, 최저 8.31, 최고 50.89, 평균 26.27로 관측치에 비해 전반적으로 과대 추정되었다. 1월부터 6월은 김 종자가 입식되기 전의 시점으로, 직접적으로 생산에 영향을 미치는 데이터가 고려되지 않았기 때문에

앞의 예측시차 2-1-8 모형에 비해 오차가 크게 나타난 것으로 볼 수 있다.

〈그림 5-10〉 완도지역 어기 예측모형2 김 단수 추정 결과(인공신경망)



자료: 저자 작성

MSE는 41.85, MAPE는 40.76, Theil's U 계수는 0.11로 높아 예측력이 낮게 나타났다. 세부적인 편불균등계수를 살펴보면, U_M 과 U_S 가 각각 0.45, 0.33으로 1개월 예측모형에서와 비슷하게 크게 나타났다. 비체계적 요인의 비중을 뜻하는 U_C 는 0.22로 나타나 새로운 자료를 추가하여 모형을 개선한다면 예측의 정확도를 높일 수 있을 것으로 보인다.

3) 고흥

(1) 1개월 예측모형

가. 단기 예측

고흥지역의 단기 예측에서 가장 낮은 시험오차를 보였던 모형은 수질 데이터를 학습에 사용한 장단기 기억 신경망모형(은닉노드 100개)으로, 모형의 시험오차²⁰⁾는 26.56이었다.

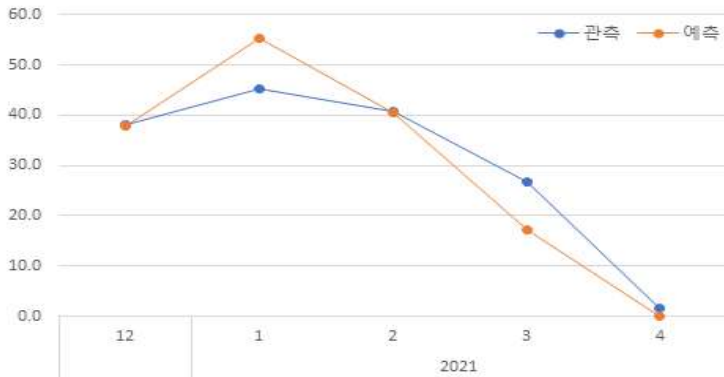
〈표 5-19〉 고흥지역 단기 예측모형 김 단수 추정 결과(인공신경망)

단위: 속/책

	2020	2021			
	12	1	2	3	4
관측	38.18	45.36	40.79	26.59	1.64
예측	37.94	55.24	40.56	17.18	0.00
관측-예측	0.23	-9.88	0.24	9.41	1.64
MSE	MAPE	Theil's U	U_M	U_s	U_c
37.81	31.67	0.09	0.00	0.35	0.64

자료: 저자 작성

〈그림 5-11〉 고흥지역 단기 예측모형 김 단수 추정 결과(인공신경망)



자료: 저자 작성

해당 모형을 이용해 2021년산 전남지역의 김 단수를 예측한 결과는 〈표 5-19〉 및 〈그림 5-11〉과 같다. 본 모형은 예측 목표 시점을 t 라고 할 때, $t-1$ 시점의 독립변수를 사용한다. 따라서 예측에 사용된 데이터는 2020년 11월부터 2021년 3월까지의 데이터이다. 예측값은 1월에 정점에 달하고 이후 하락하는 실제 관측치와 비슷한 추이를 보여주었다.

관측치는 최소 1.64, 최고 45.36, 평균 50.31의 범위를 보였으나, 예측치는 최소 0, 최고 55.24, 평균 30.19로 관측치에 비해 편차가 크게 나타났다.

20) 시험오차는 10월부터 5월까지의 단수 예측에 대한 오차였으며, 고흥의 2021년산 김 생산 어기는 12~4월로 〈표 5-19〉와는 차이가 있다.

1월과 3월의 오차가 크게 나타났는데, 2021년산에 예측에 대한 MSE는 37.81, MAPE는 31.67, Theil's U 계수는 0.09였다.

MSE와 MAPE가 컸던 것에 비해 Theil's U 계수는 양호하게 나타났는데, 편불균등계수를 살펴보면 U_s 가 0.35로 비교적 크게 나타나 오차에서 예측값의 분산 차이가 차지하는 비중이 큰 것으로 나타났다.

나. 장기 예측

고흥지역의 장기(1개월) 예측에서 가장 낮은 시험오차를 보였던 모형은 기상 데이터를 학습에 사용한 장단기 기억신경망모형(은닉노드 10개)으로, 모형의 시험오차²¹⁾는 30.94였다.

해당 모형을 이용해 2021년산 전남지역의 김 단수를 예측한 결과는 <표 5-20> 및 <그림 5-12>와 같다. 본 모형은 예측 목표 시점을 t 라고 할 때, $t-5$ 시점의 기상변수를 사용한다. 따라서 예측에 사용된 데이터는 2020년 7월부터 2020년 11월까지의 데이터이다.

<표 5-20> 고흥지역 장기 예측모형 김 단수 추정 결과(인공신경망)

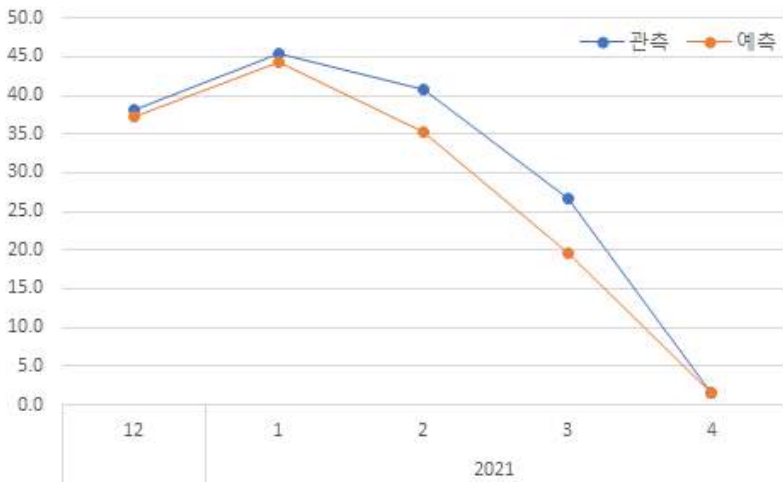
단위: 속/책

	2020	2021			
	12	1	2	3	4
관측	38.18	45.36	40.79	26.59	1.64
예측	37.28	44.28	35.24	19.63	1.49
관측-예측	0.90	1.08	5.56	6.96	0.15
MSE	MAPE	Theil's U	U_M	U_s	U_c
16.26	10.69	0.06	0.53	0.01	0.46

자료: 저자 작성

21) 시험오차는 10월부터 5월까지의 단수 예측에 대한 오차였으며, 고흥의 2021년산 김 생산 어기는 12~4월로 <표 5-19>와는 차이가 있다.

〈그림 5-12〉 고흥지역 장기 예측모형 김 단수 추정 결과(인공신경망)



자료: 저자 작성

예측값은 1월에 정점에 달하고 이후 하락하는 실제 관측치와 비슷한 값과 방향을 보여주었다. 이전 모형보다 직접적인 상관관계가 떨어진다고 볼 수 있는 5개월 전의 데이터를 사용하였음에도, 전반적인 예측력이 양호하게 나타났다.

예측치의 범위는 최소 1.49에서 최고 44.28, 평균 27.58로 단기 예측모형보다 관측치에 가깝고 편차가 작았으나, 전반적으로 과소추정이 나타났다. 이때 MSE는 16.26, MAPE는 10.69, Theil's U 계수는 0.06으로 단기 예측모형에 비해 좋은 예측력을 나타냈다.

편불균등계수를 살펴보면, U_M 이 0.53으로 가장 커 관측값과 예측값의 평균의 차이에서 오는 오차의 비중이 가장 큰 것으로 나타났으며, U_S 는 0.01로 분산의 차이는 작은 것으로 추정된다. 또한 U_C 는 0.46으로 나타나 다른 자료를 추가로 반영하여 모형을 개선시킬 수 있을 것으로 판단된다.

(2) 어기 예측모형

가. 어기 예측모형1

고흥지역의 어기 예측모형1에서 가장 낮은 시험오차를 보였던 모형은 수질데이터를 학습에 사용한 다층 퍼셉트론모형(은닉노드 10개)으로, 모형의 시험오차는 33.12였다.

해당 모형을 이용해 2021년산 고흥지역의 김 단수를 예측한 결과는 <표 5-21> 및 <그림 5-13>와 같다. 본 모형은 예측 목표 시점을 t 부터 $t+7$ 라고 할 때, $t-1 \sim t-2$ 시점까지의 독립변수를 사용한다. 따라서 예측에 사용된 데이터는 2020년 8월부터 2020년 9월까지의 데이터이다. 아래의 표는 전체(10월부터 5월) 예측값 중 고흥지역의 2021년산 어기에 해당하는 2020년 12월부터 2021년 4월까지의 예측치와 이에 대한 예측력 검정지표이다.

예측치의 범위는 최저 6.56에서 최고 42.74, 평균은 29.86으로 앞서 1개월 추정 모형에 비해 실제 관측치와 더 가까운 값이 도출되었다. 이때 MSE는 12.15, Theil's U 계수는 0.05로 양호하게 나타났으나, MAPE 값은 65.35로 다소 크게 추정되었다. 이는 생산량이 작은 4월(관측치 1.64)의 예측치가 6.56으로 오차 비율이 높았기 때문이다.

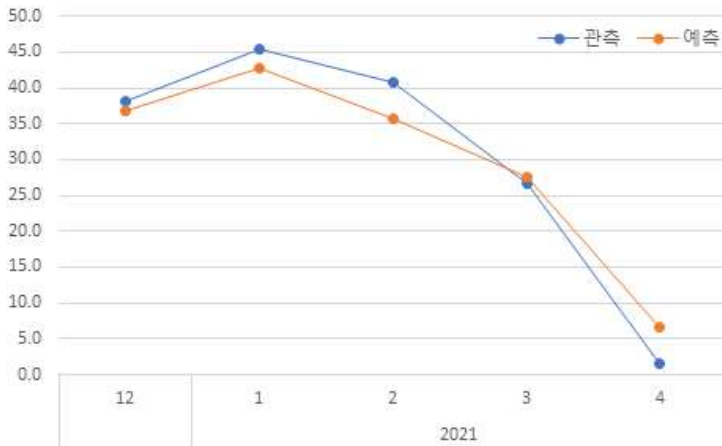
<표 5-21> 고흥지역 어기 예측모형1 김 단수 추정 결과(인공신경망)

단위: 속/책

	2020	2021			
	12	1	2	3	4
관측	38.18	45.36	40.79	26.59	1.64
예측	36.82	42.74	35.61	27.55	6.56
관측-예측	1.36	2.62	5.18	-0.96	-4.92
MSE	MAPE	Theil's U	U_M	U_S	U_C
12.15	65.35	0.05	0.04	0.79	0.18

자료: 저자 작성

〈그림 5-13〉 고흥지역 어기 예측모형1 김 단수 추정 결과(인공신경망)



자료: 저자 작성

편불균등계수를 살펴보면, U_M 이 0.04로 평균의 차이로 인한 오차는 작았으나, U_S 가 0.79로 분산의 차이가 크게 나타났다. 또한 U_C 는 0.18으로 나타나 추가 자료를 통해 모형을 상당 부분 개선할 수 있을 것으로 판단된다.

나. 어기 예측모형2

고흥지역의 어기 예측모형2에서 가장 낮은 시험오차를 보였던 모형은 기상 데이터 및 수질데이터를 모두 사용한 다층 퍼셉트론모형(은닉노드 10개)으로, 모형의 시험오차는 20.20이었다.

해당 모형을 이용해 2021년산 전남지역의 김 단수를 예측한 결과는 〈표 5-22〉 및 〈그림 5-14〉와 같다. 본 모형은 예측 목표 시점을 t 부터 $t+7$ 라고 할 때, $t-4 \sim t-9$ 시점까지의 기상변수를 사용한다. 따라서 예측에 사용된 데이터는 2020년 1월부터 2020년 6월까지의 데이터이다. 아래의 표는 전체 (10월부터 5월) 예측값 중 고흥지역의 2021년산 어기에 해당하는 2020년 12월부터 2021년 4월까지의 예측치와 이에 대한 예측력 검정지표이다.

〈표 5-22〉 고흥지역 어기 예측모형2 김 단수 추정 결과(인공신경망)

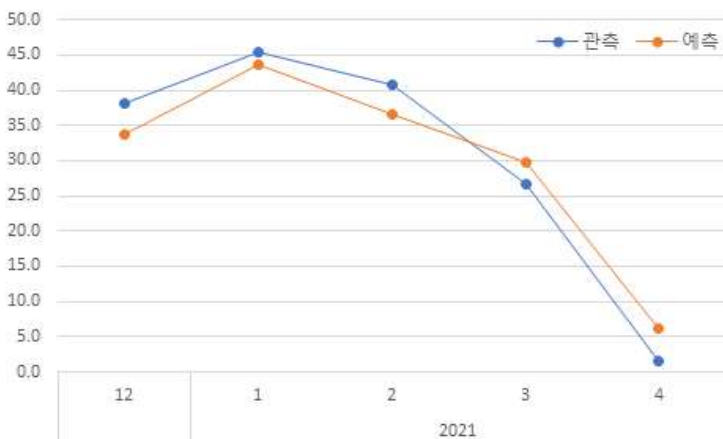
단위: 속/책

	2020	2021			
	12	1	2	3	4
관측	38.18	45.36	40.79	26.59	1.64
예측	33.72	43.66	36.53	29.81	6.11
관측-예측	4.46	1.70	4.26	-3.22	-4.48
MSE	MAPE	Theil's U	U_M	U_S	U_C
14.27	62.36	0.06	0.02	0.61	0.37

자료: 저자 작성

예측치의 범위는 최저 6.11에서 최고 43.66, 평균은 29.97로 1개월 추정 모형에 비해 실제 관측치와 더 가까운 값이 도출되었다. 이때 MSE는 14.27, Theil's U 계수는 0.06으로 양호하게 나타났으나, 앞의 경우와 마찬가지로 생산량이 작은 4월 추정치 절대값 비율 오차가 커 MAPE 값은 65.35로 다소 크게 추정되었다. 바로 앞의 모형과 마찬가지로 U_S 값이 0.61로 추가 자료 반영을 통해 분산 차이를 줄일 수 있을 것으로 보인다.

〈그림 5-14〉 고흥지역 어기 예측모형2 김 단수 추정 결과(인공신경망)



자료: 저자 작성

제2절 비교모형(다중회귀모형)

인공신경망모형과의 비교를 위해 전남, 완도, 고흥지역 김 단수에 대한 개별방정식을 다중회귀모형으로 추정하였다. 추정 방법은 통상최소자승법(OLS: Ordinary Least Square Method)을 이용하였으며, 전체 개별방정식에 대해 level-level 함수 형태의 방정식을 추정하였다. 개별방정식 추정에는 EVIEWS12 통계패키지를 이용하였다.

지역별 단수 개별방정식은 기상변수, 수질변수가 영향을 미치는 구조로 설정하였는데, 투입변수 및 투입변수의 시차에 따라 지역별로 6개 모형을 추정하였다. 비교모형의 구조는 <표 5-23>과 같다.

<표 5-23> 비교모형 개별방정식 구조

구분	투입변수
모형1	• 기상변수(월간강수량 _{t-1} , 일최대강수량 _{t-1} , 월평균풍속 _{t-1} , 최대순간풍속 _{t-1} , 일조량 _{t-1}), 계절더미(11월, 12월, 1월, 2월, 3월, 4월, 5월)
모형2	• 기상변수(월간강수량 _{t-5} , 일최대강수량 _{t-5} , 월평균풍속 _{t-5} , 최대순간풍속 _{t-5} , 일조량 _{t-5}), 계절더미(11월, 12월, 1월, 2월, 3월, 4월, 5월)
모형3	• 수질변수(표층수온 _{t-1} , 표층용존산소 _{t-1} , 표층용존질소 _{t-1} , 표층무기인농도 _{t-1} , 표층클로로필농도 _{t-1}), 계절더미(11월, 12월, 1월, 2월, 3월, 4월, 5월)
모형4	• 수질변수(표층수온 _{t-5} , 표층용존산소 _{t-5} , 표층용존질소 _{t-5} , 표층무기인농도 _{t-5} , 표층클로로필농도 _{t-5}), 계절더미(11월, 12월, 1월, 2월, 3월, 4월, 5월)
모형5	• 기상변수(월간강수량 _{t-1} , 일최대강수량 _{t-1} , 월평균풍속 _{t-1} , 최대순간풍속 _{t-1} , 일조량 _{t-1}), 수질변수(표층수온 _{t-1} , 표층용존산소 _{t-1} , 표층용존질소 _{t-1} , 표층무기인농도 _{t-1} , 표층클로로필농도 _{t-1}), 계절더미(11월, 12월, 1월, 2월, 3월, 4월, 5월)
모형6	• 기상변수(월간강수량 _{t-5} , 일최대강수량 _{t-5} , 월평균풍속 _{t-5} , 최대순간풍속 _{t-5} , 일조량 _{t-5}), 수질변수(표층수온 _{t-5} , 표층용존산소 _{t-5} , 표층용존질소 _{t-5} , 표층무기인농도 _{t-5} , 표층클로로필농도 _{t-5}), 계절더미(11월, 12월, 1월, 2월, 3월, 4월, 5월)

자료: 저자 작성

첫째 모형은 강수량, 풍속, 일조량 등 기상변수와 계절더미만을 고려하였으며, 1개월 이전의 기상변수가 단수에 영향을 미치는 것으로 설정하였

다. 둘째 모형은 첫째 모형과 동일한 변수이나, 5개월 이전의 기상변수가 영향을 미치는 구조이다. 셋째 모형은 1개월 이전의 수온, 용존산소, 용존 질소 등의 수질변수와 계절더미가 김 단수에 영향을 미치는 구조다. 동일한 변수지만 5개월 이전의 데이터가 영향을 주는 구조는 모형4로 설정하였다. 마지막으로 기상변수, 수질변수, 계절더미 전체가 단수에 영향을 미치는 구조로 모형5(시차 1개월)와 모형6(시차 5개월)을 설정하였다.

이처럼 동일한 변수가 투입되는 모형 6개를 지역별로 총 18개 추정하였는데, 이 18개 모형에 대해서는 인공신경망과 동일하게 비교하기 위해 추가적인 더미변수나 통계적 유의성을 고려하지 않았다. 다만 합리적인 비교를 위해 투입변수의 통계적 유의성을 고려하고 모형의 설명력을 높이기 위해 더미변수를 투입한 모형을 추가적으로 추정하였다.²²⁾ 추가 모형은 1개월 시차, 5개월 시차 변수를 투입하여 지역별로 2개의 개별방정식을 추정하였는데, 이를 모형7, 모형8로 설정하였다.

한편, 인공신경망 예측모형의 경우 여기 직전 2개월(8, 9월) 및 상반기(1월~6월) 데이터를 이용해 여기 전체(11월~5월)를 예측하기 위한 두 가지 여기 예측모형을 설정하였다. 해당 모형은 데이터의 시계열이 월별이 아닌 양식 김의 여기 단위인 연도별로 이루어져 변수의 시계열 길이가 15개로 짧아진 반면, 독립변수의 개수는 시계열 개수보다 많아 다중회귀 분석이 불가능하다. 따라서 다중회귀 분석에서는 여기 예측모형을 제외하였다.

비교모형 분석 목적은 최적의 예측모형 도출이 아닌 동등 변수 조건의 인공신경망모형과의 비교이므로, 본문에서는 세부 추정결과를 제시하지 않았다(부록참고).

22) 추정계수의 유의 수준이 최대 10% 이내이며, 반복적인 추정 결과에서 결정계수(R^2)가 가장 높고, 자기상관(Aurocorrelation)이 존재하지 않는 모형을 최종적으로 선정하였다.

1. 개별방정식 예측력 검증

앞서 전남, 완도, 고흥지역의 단수 예측모형 설명력을 결정계수를 통해 비교하였다. 그러나 모형의 실질적인 예측력을 판단하기 위해서는 실제값과 예측값을 비교하여 살펴볼 필요가 있는데, 본 연구에서는 2021년산(2020년 10월~2021년 5월) 김 단수에 대한 모형의 예측력 검정을 실시하였다.

모형 예측력 검증 방법은 MSE(Mean Squared Error), Theil's U 계수(Theil's inequality coefficient) 등 2가지를 이용하였다. MAE는 실제치와 예측치의 차이, 즉 오차의 평균을 제공한 것으로 이 값이 클수록 예측력이 낮으며, Theil's U 계수는 0에서 1 사이의 값을 가지는데 0에 가까울수록 예측력이 우수한 것으로 평가한다.

지역별 단수 예측모형에 대한 예측력 검증 결과, 모형의 결정계수가 낮았던 고흥지역의 예측력이 비교적 높은 것으로 나타났으며, 완도지역의 예측력이 가장 낮은 것으로 도출되었다. 세부적으로는 전남지역과 완도지역의 경우, 5개월 전 기상변수와 수질변수가 영향을 미치는 구조인 모형6의 예측력이 가장 좋게 나타났다. 고흥지역에서는 5개월 전 기상변수와 1개월 전 수질변수가 영향을 미치는 모형2와 모형3의 예측력이 높게 나타났다.

그러나 예측력이 좋은 것으로 나타난 전남지역 모형6과 고흥지역 모형3은 자기상관 문제가 있는 모형이므로, 해당 모형으로 단수를 예측하기는 어려울 것으로 판단된다. 각 지역의 모형7, 모형8의 경우 통계적 유의성과 자기상관 문제를 고려하여 최적으로 추정했음에도 불구하고 낮은 예측력을 보였는데, 이는 단수 추정에 투입되는 설명변수인 기상요인, 수질요인이 연도별 및 월별 편차가 존재하여, 예측력을 저하시키는 요인으로 작용하는 것으로 판단된다.

〈표 5-24〉 개별방정식 예측력 검정 결과

구분		MSE	Theil's U
전남	모형1	23.26	0.12
	모형2	18.47	0.11
	모형3	20.62	0.12
	모형4	11.27	0.08
	모형5	14.21	0.10
	모형6	10.73	0.08
	모형7	16.77	0.11
	모형8	26.73	0.13
완도	모형1	141.18	0.19
	모형2	95.46	0.17
	모형3	97.11	0.16
	모형4	170.92	0.21
	모형5	104.67	0.17
	모형6	76.24	0.15
	모형7	154.09	0.23
	모형8	102.97	0.17
고흥	모형1	31.32	0.08
	모형2	8.12	0.04
	모형3	7.57	0.04
	모형4	9.42	0.04
	모형5	18.56	0.06
	모형6	42.64	0.09
	모형7	25.58	0.07
	모형8	35.97	0.09

자료: 저자 작성

2. 종합

종합에서는 비교모형(다중회귀분석) 예측력 검정 결과를 바탕으로 지역별 단기 예측모형과 장기 예측모형 중 가장 시험오차가 낮은 모형을 추려내어 2021년산 김 단수를 추정하고 각각의 예측력을 검정 지표를 통해 비교하였다.

1) 전남

(1) 단기 예측

전남지역의 단기 예측에서 가장 예측력이 높았던 모형은 1개월 전의 기상 요인, 수질요인이 영향을 미치는 모형5로, 해당 모형을 이용해 2021년산 전남지역의 김 단수를 예측한 결과는 <표 5-25> 및 <그림 5-15>과 같다. 본 모형은 2010년 3월부터 2020년 5월까지의 데이터를 이용하여 2021년산 (2020년 10월~2021년 5월) 단수를 예측한 것이다.

생산량이 적은 어기 초와 어기 말의 오차 비율이 커 MAPE가 다소 크게 추정되었다. Theil's U 계수는 0.10으로 편불균등계수를 살펴보면 U_M 과 U_S 는 각각 0.05, 0.00이었으며 U_C 가 0.95로 크게 나타나 추가 자료를 이용하더라도 모형의 예측력 개선이 어려울 것으로 판단된다.

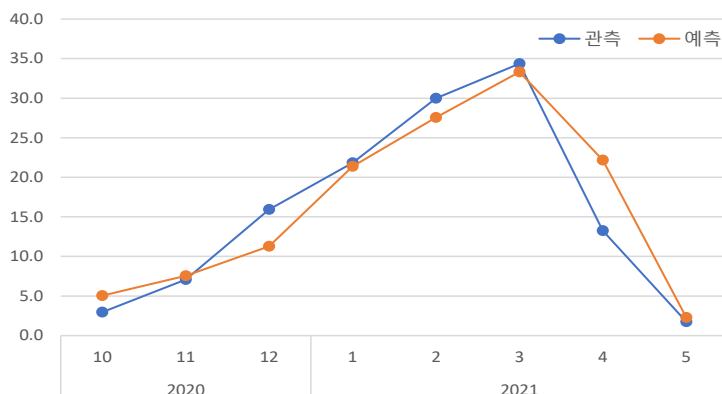
<표 5-25> 전남지역 단기 예측모형 김 단수 추정 결과(다중회귀)

단위: 속/책

	2020			2021				
	10	11	12	1	2	3	4	5
관측	2.97	7.07	15.95	21.83	29.99	34.36	13.24	1.74
예측	5.05	7.56	11.28	21.37	27.57	33.33	22.18	2.26
관측-예측	-2.08	-0.49	4.67	0.46	2.42	1.03	-8.94	-0.52
MSE	MAPE		Theil's U	U_M		U_S	U_C	
14.21	27.17		0.10	0.01		0.04	0.95	

자료: 저자 작성

〈그림 5-15〉 전남지역 단기 예측모형 김 단수 추정 결과(다중회귀)



자료: 저자 작성

(2) 장기 예측

전남지역의 장기 예측에서 가장 높은 예측력을 보였던 모형은 5개월 전의 기상변수와 수질변수를 모두 고려한 모형6으로 MSE는 10.73이었다. 본 모형은 2010년 3월부터 2020년 5월까지의 데이터를 이용하여 추정한 것으로, 단기 예측 결과와 유사하게 주 생산시기의 오차는 낮으나, 어기 초와 어기 말 오차는 큰 것으로 나타났다.

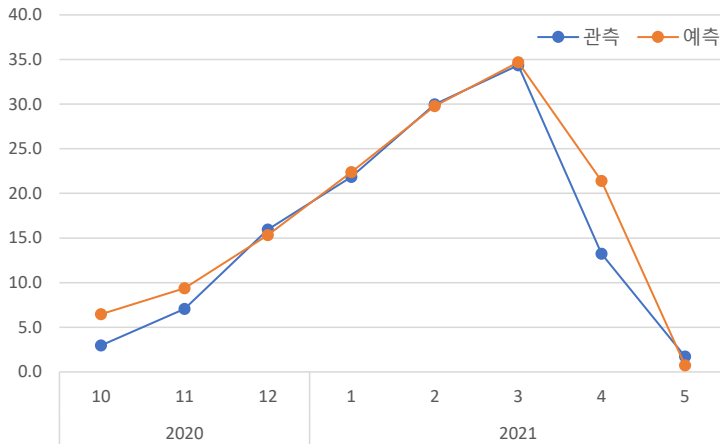
〈표 5-26〉 전남지역 장기 예측모형 김 단수 추정 결과(다중회귀)

단위: 속/책

	2020			2021				
	10	11	12	1	2	3	4	5
관측	2.97	7.07	15.95	21.83	29.99	34.36	13.24	1.74
예측	6.46	9.36	15.33	22.37	29.78	34.68	21.40	0.73
관측-예측	-3.49	-2.29	0.62	-0.54	0.21	-0.32	-8.16	1.01
MSE	MAPE		Theil's U		U _M		U _S	U _C
	10.73		34.74		0.08		0.25	0.01

자료: 저자 작성

〈그림 5-16〉 전남지역 장기 예측모형 김 단수 추정 결과(다중회귀)



자료: 저자 작성

MSE와 MAPE가 앞의 모형보다 크게 추정되었으며, Theil's U 계수도 0.14로 예측력이 다소 낮은 것으로 나타났다. 편불균등계수를 살펴보면 U_M 과 U_S 는 각각 0.25, 0.01이었으며 U_C 가 0.74로 크게 나타나 추가 자료를 이용하더라도 모형의 예측력 향상에는 한계가 있을 것으로 보인다.

2) 완도

(1) 단기 예측

완도지역의 단기 예측에서 가장 예측력이 높았던 모형은 1개월 전의 수질요인이 영향을 미치는 모형3으로, MSE는 97.11, Theil's U 계수는 0.16이었다.

완도지역의 모형3을 이용해 2021년산 완도지역 김 단수를 예측한 결과는 〈표 5-27〉 및 〈그림 5-17〉와 같다. 완도지역 단수 예측모형은 전남지역에 비해 다소 예측력이 낮은 것으로 나타났는데, 이는 단기예측 결과 그림을 통해서도 확연하게 확인할 수 있다. 본 모형은 2010년 3월부터 2020년 5월까

지의 데이터를 이용하여 2021년산(2020년 11월~2021년 5월) 단수를 예측한 것으로, 여기 초와 여기 말 오차가 매우 컸고 주 생산시기에도 전남지역보다 실제치와 예측치 차이가 전남지역보다 컸다.

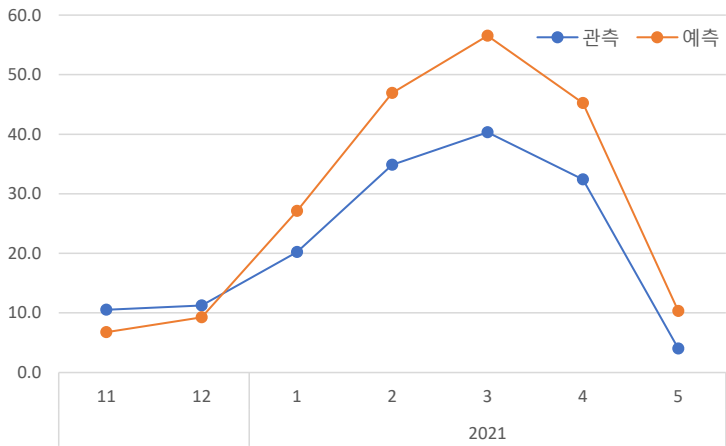
〈표 5-27〉 완도지역 단기 예측모형 김 단수 추정 결과(다중회귀)

단위: 속/책

	2020		2021				
	11	12	1	2	3	4	5
관측	10.54	11.24	20.24	34.86	40.31	32.41	4.01
예측	6.74	9.25	27.13	46.95	56.53	45.24	10.33
관측-예측	3.80	1.99	-6.89	-12.09	-16.22	-12.83	-6.32
MSE	MAPE		Theil's U		U _M		U _S
	97.11		51.46		0.16		0.50
					0.40		0.11

자료: 저자 작성

〈그림 5-17〉 완도지역 단기 예측모형 김 단수 추정 결과(다중회귀)



자료: 저자 작성

MSE와 MAPE는 각각 97.11, 51.46으로 크게 나타났으며, Theil's U 계수도 0.16으로 예측력이 낮았다. 편불균등계수를 살펴보면 전반적으로 과소 추정된만큼 U_M 값이 0.50으로 크게 나타났고, U_S 가 0.40으로 분산도 불일

치하는 것으로 나타났다. 따라서 완도의 경우 추가자료를 이용해 모형을 보완하여 예측력을 향상시킬 수 있을 것으로 기대된다.

(2) 장기 예측

완도지역의 장기 예측에서 가장 높은 예측력을 보였던 모형은 5개월 전의 기상변수와 수질변수를 모두 고려한 모형6으로 MSE는 76.24, Theil's U 계수는 0.15로 계산되었다.

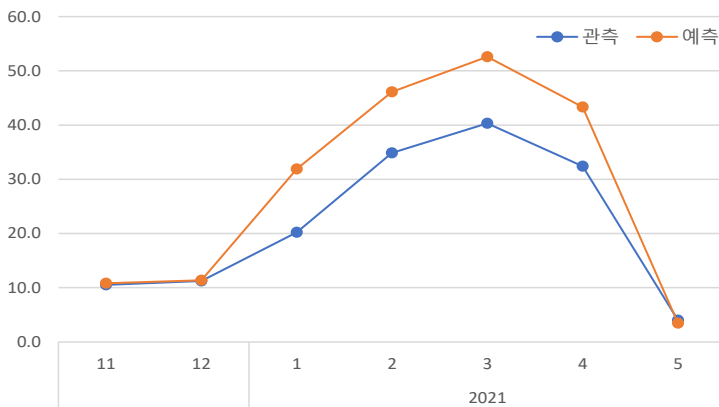
〈표 5-28〉 완도지역 장기 예측모형 김 단수 추정 결과(다중회귀)

단위: 속/책

	2020		2021				
	11	12	1	2	3	4	5
관측	10.54	11.24	20.24	34.86	40.31	32.41	4.01
예측	10.80	11.37	31.91	46.13	52.58	43.33	3.45
관측-예측	-0.26	-0.13	-11.67	-11.27	-12.27	-10.92	0.56
MSE	MAPE		Theil's U		U _M		U _S
	76.24		24.51		0.57		0.38
			0.15		0.06		0.06

자료: 저자 작성

〈그림 5-18〉 완도지역 장기 예측모형 김 단수 추정 결과(다중회귀)



자료: 저자 작성

본 모형은 2010년 11월부터 2020년 5월까지의 데이터를 이용하여 추정
한 것으로, 단기 모형보다 예측력이 높은 것으로 나타났다. <그림 5-18>을
통해 예측치와 실제치의 차이를 살펴보면, 주 생산시기보다 어기 초와 어기
말 오차가 더 낮은 모습을 보였다.

단기 예측모형과 마찬가지로 MSE와 MAPE, Theil's U계수도 크게 나타
나 예측력이 낮게 나타났다. 편불균등 계수에서도 U_M 값이 0.57로 컸고, U_S
가 0.38로 분산의 차이도 오차에서 큰 비중을 차지하는 것으로 나타났다. 따
라서 완도의 장기 예측모형의 경우도 마찬가지로 추가 자료를 이용해 모형을
보완하여 예측력을 높일 수 있을 것으로 보인다.

3) 고흥

(1) 단기 예측

고흥지역의 단기 예측에서 예측력이 가장 높았던 모형은 1개월 전의 수질
요인이 영향을 미치는 모형3이었다. 본 모형은 2010년 3월부터 2020년 4월
까지의 데이터를 이용하여 2021년산(2020년 12월~2021년 4월) 김 단수를
예측한 것으로, 어기 초인 10월, 11월과 어기 말인 5월에는 고흥의 생산이
없어 단수 관측치가 없었다. 이로 인해 주 생산시기에 대한 예측력 비교만
이루어지면서, 다른 지역에 비해 상대적으로 높은 예측력을 보였다.

<표 5-29> 고흥지역 단기 예측모형 김 단수 추정 결과(다중회귀)

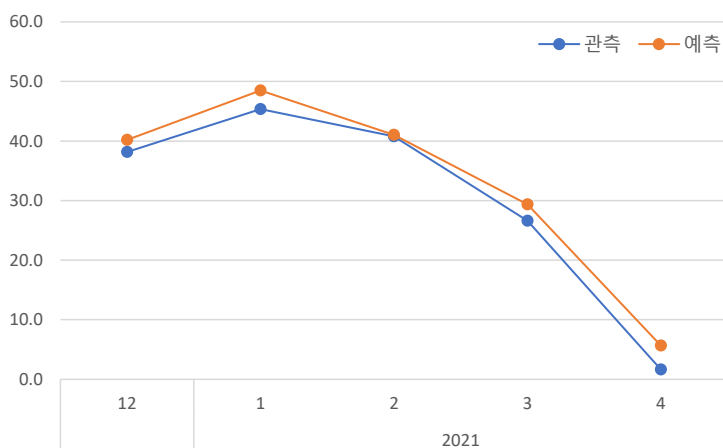
단위: 속/척

	2020	2021			
	12	1	2	3	4
관측	38.18	45.36	40.79	26.59	1.64
예측	40.22	48.51	41.06	29.35	5.65
관측-예측	-2.04	-3.15	-0.27	-2.76	-4.01
MSE	MAPE	Theil's U	U_M	U_S	U_C
7.57	53.75	0.04	0.79	0.07	0.14

자료: 저자 작성

MSE는 7.57로 양호했으나 여기 말인 4월의 절대비율오차가 커 MAPE값이 다소 높게 나타났다. Theil's U계수도 0.04로 예측력이 우수한 것으로 나타났다. 편상관계수에서는 U_M 이 0.79로 높아 관측치와 평균의 차이에서 오는 오차의 비중이 높게 나타났다. U_S 값은 0.07로 분산의 차이는 크지 않았다. U_C 는 0.14로, 추가적인 자료를 이용해 예측치와 관측치의 평균 차이를 줄일 수 있을 것으로 기대된다.

〈그림 5-19〉 고흥지역 단기 예측모형 김 단수 추정 결과(다중회귀)



자료: 저자 작성

(2) 장기 예측

고흥지역의 장기 예측에서 가장 높은 예측력을 보였던 모형은 5개월 전의 기상변수만 고려한 모형2였다. 고흥지역 모형2의 MSE는 8.12, Theil's U계수는 0.04로, 단기 예측 결과와 비슷한 것으로 나타났다.

본 모형은 2007년 4월부터 2020년 4월까지의 데이터를 이용하여 추정한 결과로, 해당 모형을 이용해 2021년산 전남지역의 김 단수를 예측한 결과는 〈표 5-30〉 및 〈그림 5-20〉과 같다.

단기 예측모형과 마찬가지로 MSE는 8.12로 양호했으나 생산량이 적은 3월 및 4월의 절대비율오차가 커 MAPE값이 다소 높게 나타났다. Theil's U계수도 0.04로 예측력이 우수한 것으로 나타났다. 편상관계수에서는 U_M 과 U_S 값이 각각 0.19, 0.23으로 관측값과 예측치의 평균과 분산 차이에서 오는 오차가 커 추가적인 자료를 이용해 모형을 개선할 수 있을 것으로 기대된다.

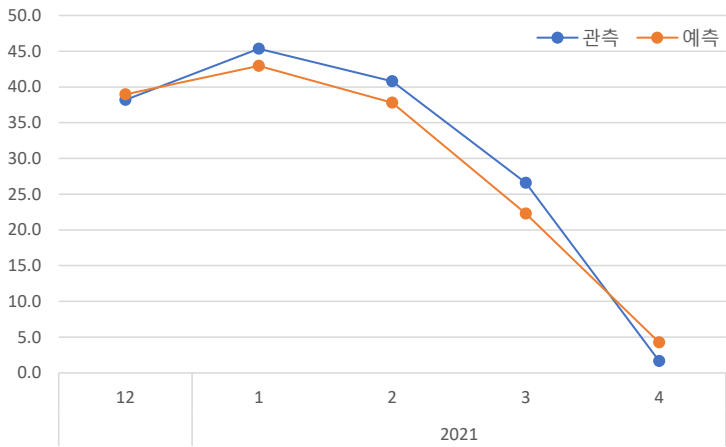
〈표 5-30〉 고흥지역 장기 예측모형 김 단수 추정 결과(다중회귀)

단위: 속/책

	2020	2021			
	12	1	2	3	4
관측	38.18	45.36	40.79	26.59	1.64
예측	38.98	42.97	37.81	22.29	4.25
관측-예측	-0.80	2.39	2.98	4.30	-2.61
MSE	MAPE	Theil's U	U_M	U_S	U_C
8.12	38.14	0.04	0.19	0.23	0.58

자료: 저자 작성

〈그림 5-20〉 고흥지역 장기 예측모형 김 단수 추정 결과(다중회귀)



자료: 저자 작성

06

결론

제1절 요약 및 결론

1. 요약

본 연구는 양식수산물의 단수 전망 시 인공신경망 전망모형 도입 타당성 검토를 위해 김 단수 전망을 위한 인공신경망모형을 구축하여 다중회귀모형과 비교하였다. 인공신경망모형의 가장 기본적인 모형인 MLP모형과 시계열자료의 분석에 유리하게 설계된 LSTM, 그리고 양방향 LSTM 모형을 이용해 전망모형을 구축하였다. 지역별 단수 예측을 위한 최적의 모형을 도출하기 위해 모형의 파라미터와 입력 데이터를 다르게 설정한 216개 인공신경망 모형을 훈련하고, 다중회귀모형과 비교하였다.

우선 전남지역 김 단수의 단기 예측모형의 경우 인공신경망모형이 MSE와 Theil's U 계수가 낮아 다중회귀모형에 비해 예측력이 높았으나, 장기 예측모형에서는 다중회귀모형이 더 나은 예측력을 보여주었다. 전남지역 예측모형의 경우 전체적으로 편불등균계수에서의 U_C 값이 높게 나타나 추가 자료를 통한 오차의 개선이 어려울 것으로 판단되었다.

완도지역 예측모형은 단기와 장기 모두 인공신경망모형의 오차가 더 작게 나타났다. Theil's U 계수는 모든 모형에서 0.1 이상이었으며, 편불균등계수를 살펴보면 인공신경망과 다중회귀모형 모두 U_M 과 U_S 값이 크게 나타나 추가 자료를 통해 예측력을 향상시킬 수 있을 것으로 기대된다.

고흥지역 예측모형의 경우 단기와 장기 모두에서 다중회귀모형의 예측 오차가 더 낮게 나타났다. 전반적으로 Theil's U 계수는 양호하게 추정되었으며, 편불균등계수 중 U_M 값이 높은 경우가 많았다. 특히 단기 예측을 위한 다중회귀모형의 경우 Theil's U 계수가 0.04로 낮았음에도 U_M 값이 0.79로 나타나 추가 자료를 통해 예측 오차를 더 줄일 수 있을 것으로 추정되었다.

〈표 6-1〉 지역별 2021년산 김 단수 추정모형 비교(1개월 추정모형)

지역	구분	모형	MSE	Theil's U	U_M	U_S	UC
전남	단기	인공신경망	11.44	0.09	0.02	0.00	0.97
		다중회귀	14.21	0.10	0.01	0.04	0.95
	장기	인공신경망	11.15	0.09	0.01	0.02	0.97
		다중회귀	10.73	0.08	0.25	0.01	0.74
완도	단기	인공신경망	33.82	0.10	0.28	0.42	0.29
		다중회귀	97.11	0.16	0.50	0.40	0.11
	장기	인공신경망	49.82	0.12	0.34	0.37	0.29
		다중회귀	76.24	0.15	0.57	0.38	0.06
고흥	단기	인공신경망	37.81	0.09	0.00	0.35	0.64
		다중회귀	7.57	0.04	0.79	0.07	0.14
	장기	인공신경망	16.26	0.06	0.53	0.01	0.46
		다중회귀	8.12	0.04	0.19	0.23	0.58

자료: 저자 작성

이상과 같이 단기 및 장기 1개월 예측모형에서는 총 6개 모형 중 3개 모형에서 인공신경망모형의 예측력이 다중회귀모형에 비해 높았다. 김은 채묘부터 수확까지 양성 기간이 짧으므로 1개월 전 데이터를 이용하는 단기모형은 해상 입식 후 작황에 직접적으로 영향을 미치는 기간의 자료를 반영한다고 볼 수 있다. 5개월 전 데이터를 이용한 장기모형은 입식 이전 변동 추세의 반영 비중이 커진다. 따라서 장기보다는 단기 예측모형에서 인공신경망이 더 나은 결과를 나타낸 경우가 많았다는 점에서 의미가 있다.

여기 예측모형의 경우 다중회귀모형에 비해 적은 독립변수를 사용하여 더 긴 시차를 예측했음에도 일부 사례에서 다중회귀모형보다 뛰어난 예측력을 보였다. 전남의 경우 8, 9월의 데이터를 이용해 전체 어기를 예측한 모형의 MSE 및 Theil's U 계수가 다중회귀(단기)모형에 비해 우수한 것으로 나타났으며, 완도의 경우 두 가지 어기 예측 시 모두 인공신경망모형이 다중회귀모형보다 예측력이 좋은 것으로 나타났다.

〈표 6-2〉 지역별 2021년산 김 단수 추정모형 비교(어기 추정모형, 인공신경망)

구분	모형	MSE	Theil's U	U_M	U_S	U_C
전남	어기예측모형1	13.44	0.09	0.02	0.00	0.98
	어기예측모형2	17.94	0.11	0.12	0.06	0.82
완도	어기예측모형1	15.95	0.08	0.02	0.00	0.98
	어기예측모형2	41.85	0.11	0.45	0.33	0.22
고흥	어기예측모형1	12.15	0.05	0.04	0.79	0.18
	어기예측모형2	14.27	0.06	0.02	0.61	0.37

자료: 저자 작성

인공신경망모형은 동일한 변수를 사용한 모형 6개 모형 중 절반에서 다중회귀모형보다 예측력이 높게 나타났으며, 일부 데이터를 사용하여 여기 전체를 예측한 경우에서도 좋은 예측력을 보였다. 그러나 다중회귀모형에 비해 예측력이 낮았던 모형도 있었으며 그 중 고흥지역 예측모형의 예측력 차이가 특히 크게 나타났다. 고흥지역 모형의 경우 편불균등계수 중 U_S 값이 높게 나타났으므로 추가 정보를 반영하여 분산 차이로 인한 오차를 개선할 수 있을 것으로 판단된다.

한편, 다중회귀모형은 대부분 독립변수의 유의성이 낮았고, 유의성이 낮았던 선별 모형에서도 기대부호가 반대로 나타나는 등 문제를 보였다. 이는 고수온, 폭우 등 악조건이 포함된 데이터를 입력했을 때 단수가 오히려 증가할 것으로 예측하므로, 예측모형으로서는 부적합한 것으로 보인다.

2. 결론

양식수산물 단수 예측을 위한 인공신경망모형 및 다중회귀모형 비교 결과, 지역별, 장기 및 단기 예측을 위해 같은 조건의 변수를 사용한 6가지 모형 중 절반인 3가지에서 인공신경망모형의 예측력이 더 높았다. 제한된 자료로 더 긴 시차를 예측하는 여기 예측모형에서도 일부에서 다중회귀모형보다 더 나은 예측력을 나타내기도 했다. 반면 다중회귀모형(비교모형)은 개별 변수의 유의성이 낮았고, 유의성이 높은 변수만을 활용한 경우에서도 기대부호가 반대로 나오는 등 예측모형으로서의 적합성이 떨어졌다.

인공신경망모형은 변수의 가공이 없이도 다중회귀모형이 표현하기 어려운 복잡한 상관관계의 표현이 가능하다. 또한, 변수 하나하나에 대한 유의성 검정이나 해석보다는 전체 모형의 예측 정확도에 집중하는 방법이므로 변수 사용의 측면에서 상대적으로 높은 자유도를 가진다. 위의 결과는 이

러한 특징이 반영된 것으로 판단되며, 이 같은 이점 때문에 수산업 전망 부문 연구에도 인공지능경망모형의 활용성이 높아질 것으로 기대된다.

본 연구의 한계점은 다음과 같다. 첫째, 모형 구축에 사용된 데이터의 세분화 정도와 자료의 양이 충분하지 않았다는 점이다. 개별 양식어장은 지리적으로 넓게 분포되어 있어 위치에 따라 관측된 수치가 다르게 나타날 수 있으나, 이 같은 세부적 차이는 고려하지 않았다. 또한 수질 관련 데이터의 경우 2010년 이후의 월별 자료만 사용하는 등 예측모형 구축에 활용된 데이터의 시계열이 짧아 충분한 양의 데이터를 사용했다고 보기 어렵다.

둘째, 양식수산물의 생산성에 크게 영향을 미치는 인위적 생산요소를 반영하지 않았다. 연구에 사용한 설명변수의 경우 자연적 요인에 국한하여 분석하였기 때문에 김 양식에 크게 영향을 미치는 다른 요소, 예를 들어 종자의 양과 품질, 노동력 등 인위적인 요인에 대한 데이터 확보가 부족했다.

셋째, 다음으로 예측모형 구축에 사용된 개별 변수의 유의성 또는 유의확률을 확인할 수 없어 변수의 적합성을 판단하기 어렵다. 다시 말해 만약 모형의 예측력이 떨어진다면 변수의 유의성이 낮았기 때문일 수도 있고, 부적절한 파라미터 설정으로 인해 과소적합 또는 과대적합이 발생했기 때문일 수 있으나 이에 대한 명확한 단서를 제시하여 단정하기는 어렵다.

넷째, 기상 및 수질 변수의 정확한 상관관계에 대한 고찰이 부족했다. 본 연구에서는 모형의 예측력에만 초점을 두고 연구를 수행하여 독립변수와 종속변수의 명확한 상관관계에 대한 분석과 검증이 충분하지는 않았다.

후속 연구에서는 예측모형의 정확도를 높이고 활용도 제고를 위해 앞서 제시된 바와 같이 분석자료를 충분히 확보하고, 인공지능경망모형이 가지는 한계점 등을 보완하는 방향의 접근이 필요할 것으로 판단된다.

제2절 제언

1. 본 연구의 기여

본 연구의 학술적 기여는 기존 활용도가 낮았던 방법론과 자료를 활용했다는 점에서 찾을 수 있다. 첫째, 기존의 수산업 분야 연구에서는 활용도가 낮았던 인공신경망모형을 통한 접근을 시도했다는 점이다. 우수한 예측성을 바탕으로 국내외 다양한 산업 및 연구 분야에 활용되고 있으나 유독 국내의 수산 분야에서는 활용 사례가 드물었다. 본 연구에서는 양식수산물 작황 전망에 있어서 계량경제이론 기반의 방법론 대신 인공신경망모형을 실증분석에 이용하여 유효성을 제시하였다.

둘째, 수산물 작황 전망을 위해 환경변수의 활용했다는 점이다. 기후 변화나 이상 해황 등은 양식수산물의 작황에 영향을 미치며, 수급 불균형을 유발한다. 일정 수준 이상의 환경 변화가 관찰되는 경우 직관적인 수준에서도 생산량 변화를 예측할 수 있음에도, 계량경제모형으로 이를 직접 반영하기는 어렵지만, 본 연구에서는 기존의 연구와 달리 환경변수를 직접 반영하여 예측모형을 구축했다는 점에서 의미를 가진다.

마지막으로 본 연구에서는 제언을 통해 한계점에 대한 개선방안, 그리고 정책적인 활용방안 등을 제시하고자 한다. 이를 바탕으로 후속 연구를 통해 기존 이론의 발전과 더불어 실질적 정책 의사결정 도구로 이용될 수 있을 것으로 기대한다.

2. 제언

1) 연구 개선방안

본 연구를 기초로 향후 실질적인 정책 의사결정 수단을 개발하기 위해서는 세부 품목별로 심층적인 후속 연구를 통해 다음과 같이 본 연구의 한계를 보완하고 실질적인 운용체계를 확립하기 위한 노력이 필요하다.

우선 정밀한 추정을 위해 세부 지역별 어장 환경을 대표할 수 있는 데이터베이스의 탐색과 구축이 필요하다. 기상 및 수질과 같은 환경변수와 함께 시설의 종류나 밀도, 채묘, 노동력 활용 여부와 같은 인위적인 투입요소에 대한 장기 시계열 데이터의 구축 및 활용방안도 고려해야 한다. 또한 정성적 설문조사 자료, 위성 이미지와 같은 비수치적 자료의 활용은 예측 모형의 합리성과 정밀도를 높일 수 있을 것이다.

다음으로 연구의 목적에 맞는 예측모형의 구축이 필요하다. 정책 의사결정의 도구로서 정확한 생산량을 예측하는 방법도 좋지만, 다양한 외부 요소로 인해 정확한 생산량 예측은 어려울 수 있다. 만약 특정한 정책 의사결정에 필요한 것이 정확한 생산량보다는 일정 수준의 증감 예측이라면 본 연구와 같은 회귀모형보다는 분류모형 기반의 분석이 더욱 효과적일 수 있다.

그리고 인공지능망모형으로는 변수의 유의성 검정이나 결과에 대한 해석이 어렵다. 따라서 다양한 변수 조합을 통해 변수의 유의성을 도출하고, 시나리오 분석 등을 통해서 간접적으로 변수의 영향력과 상관관계를 유추하는 방법 등을 시도해볼 수 있다. 또한 최근에는 인공지능망의 복잡한 예측 결과 해석을 위한 다양한 기법들이 개발되고 있으므로 이를 이용해 어떤 지표들이 의미있게 사용되었는지 분석할 수 있을 것으로 기대된다.

따라서 후속 연구에서는 데이터베이스 구축, 모형설계, 평가에 이르는 과정에서 제시된 한계점을 보완해 나가야 할 것이며, 실질적인 운용방안 도출을 통해 정책적 활용방안을 구체화할 수 있을 것이다.

〈표 6-3〉 연구 개선방안

DB 구축
- 양식어장의 위치에 따라 어장의 환경을 대표할 수 있는 기상 또는 수질 데이터 탐색 - 작황에 영향을 줄 수 있는 인위적 투입요소(시설 종류, 양성 밀도, 채묘 등) DB화 - 위성사진 등 비수치적 자료 활용
↓
모형설계
- 연구 목적에 맞는 인공지능망모형 설계 ▷ 정량 전망이 필요한 경우 - 회귀모형 ▷ 증감 예측이 필요한 경우 - 분류모형
↓
모형평가
- 변수 조합의 다양화 및 모델 해석 기법 등을 이용한 변수의 유의성 분석 - 시나리오 분석을 통한 변수의 영향력 추정 - MSE, MAPE, Theil's U, RMSPE 등 예측력 검정지표를 통한 평가
↓
실질적 운용방안 수립
- DB 구축 및 관리체계 구축 - 모형 관리 및 운용 매뉴얼 수립 - 정책 대응체계 구축

자료: 저자 작성

2) 인공지능망을 이용한 수급 전망체계 구축 및 활용방안

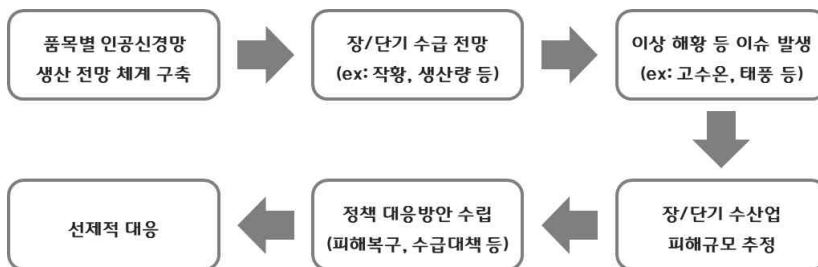
지금도 우리나라의 양식수산업 생산 규모는 확대되는 추세에 있으나, 기후변화와 더불어 기상과 해황의 불규칙성으로 인해 매년 크고 작은 피해가

발생하는 등 생산의 불안정성이 확대되고 있다. 이에 따라 업계 및 정부에서는 양식수산물의 수급 불안정성 해소를 위해 장기적인 전망과 선제적 대응 요구가 늘어나고 있다.

이러한 수요를 반영하기 위해 한국해양수산개발원에서는 수산부문 총량 모형 KMI-FOSiM을 구축하여 운용 중에 있으며, 수산업관측사업을 통해서도 주요 양식수산물에 대한 수급전망모형을 구축하여 양식수산물 수급 전망 및 정책연구 분야에 활용되고 있다. 상기 전망 수단은 계량경제 이론에 기반하여 우수한 예측력을 가지고 있으나, 고수온, 가을태풍 등 환경 변화를 즉각적으로 반영하는 데는 한계점이 존재한다. 따라서 인공신경망모형과 같은 방법을 보완적으로 활용할 수 있다.

기존의 전망체계에서 취약한 부분을 인공신경망모형을 통해 보완하는 방향으로 활용한다면 다양한 변화에 대한 정책 시뮬레이션이 가능할 것이다. 예를 들어 이상 해황 등 이슈가 발생했을 때, 즉각적으로 장·단기적인 수산업 피해 규모를 추정하고 이를 바탕으로 피해 복구 또는 수급대책을 수립하는 등 구체적인 정책 대응방안 도출에 기여할 수 있을 것으로 기대된다.

〈그림 6-1〉 수급 전망체계 구축 및 활용방안



자료: 저자 작성

3) 어장 환경 관측 DB 강화

어장 환경과 관련한 데이터는 국립수산물품질관리원의 한국해양자료센터, 실시간 해양환경 어장정보 시스템 등에서 제공한다. 대부분의 해양 데이터 관측은 특정 지점에 한정하여 이루어지고 있어 양식어장 환경 분석에 이를 활용하기 위해서는 특정 구역의 대푯값이나 가장 가까운 지점의 관측치를 이용해야 하는 실정이다.

지리적으로 인접한 양식 어장이라고 할지라도 주변 지형과 구조물, 해수의 흐름, 양식시설의 밀도 등에 따라 해수, 즉 어장 환경의 성질은 크게 달라질 수 있다. 또한 분석 대상이 되는 지역에 인접한 관측소가 존재하는 경우도 일부 시점의 공백이 존재하거나 관측이 중단되는 등 데이터의 연속성이 떨어지는 경우가 많았다. 이런 경우 동일한 모형에 투입되는 다른 자료의 활용 범위에까지 제한이 생기게 된다.

본연구에서도 유사한 문제점이 있었다. 해황변수의 관측 시계열이 2010년부터 활용 가능하여 해황 변수를 사용한 모형에서는 생산 데이터와 기상 변수 데이터의 활용도 제한적으로 이루어져 전반적인 훈련데이터의 시계열이 짧아졌다. 분석모형의 예측력 검정을 위해 Theil's U 계수와 편불균등계수를 살펴본 결과 추가적인 자료, 즉 충분한 시계열이 확보된다면 평균이나 분산의 차이에서 오는 모형의 오차를 줄일 수 있을 것으로 판단되었다.

따라서 향후 세분화된 양식어장 환경 데이터의 관측과 더불어 장기적인 유지·관리가 이루어진다면 연구양식수산물뿐만 아니라 수산업 전반의 생산 및 작황 예측의 정밀도, 그리고 관련 연구에서 자료의 활용도를 향상시킬 수 있을 것으로 보인다.

참고문헌

국내 문헌

- 국립수산물과학원(2010), 『김 양식 매뉴얼』
- _____(2017), 『주요 양식품종 모니터링』
- _____(2018), 『주요 양식품종 모니터링』
- _____(2021a), 『2021년 2월 남서해부 연안 어장 환경』
- _____(2021b), 『2021년 2월 동해권 연안 어장 환경』
- _____(2021c), 『2021년 3월 서해 연안 어장 환경』
- 권태용·윤상후·김래용(2019), 「기상 및 토양정보가 고랭지배추 단수예측에 미치는 영향」, 『한국환경과학회지』, 제28권 8호, 한국환경과학회, pp. 701~707.
- 김동우(2021), 「자문보고서: 시계열 데이터 분석용 인공신경망모형 프로그래밍」(내 부자료)
- 김민영 외(2019), 「역전파 신경망 모델을 이용한 기준 작물 증발산량 산정」, 『한국농공학회논문집』, 제61권 6호, 한국농공학회, pp. 111~121.
- 김철원·오현주·신윤경(2013), 「가막만 양식 굴, *Crassostrea gigas* 폐사에 영향을 끼치는 수온의 영향」, 『The Korean journal of malacology』, 제29권 3호, 한국패류학회, pp. 245~250.
- 나상일 외(2017), 「MODIS NDVI와 기상요인을 고려한 마늘·양파 주산단지 단수예측 모형 개발」, 『대한원격탐사학회지』, 제33권 5-2호, 한원격탐사학회, pp. 647~659.

- 박성호(2020), 「머신러닝 기반 식용 작물별 생산량 예측 연구」, 성균관대학교 대학원 석사학위 논문.
- 박민우 외(2013), 「수온 및 염분농도에 따른 북방전복, *Haliotis discus hannai*(Ino, 1952)의 생존율, 탈락율 및 발의 조직학적 변화」, 『The Korean journal of malacology』, 제29권 4호, 한국패류학회, pp. 303~311.
- 안현미 외(2016), 「고창 갯벌에서 중국산 및 충남산 바지락의 성장, 비만도 및 폐사 특성 비교」, 『The Korean journal of malacology』, 제32권 3호, 한국패류학회, pp. 175~184.
- 옥영수·김상태·고봉현(2007), 「양식 넙치의 가격변동 및 예측에 관한 연구」, 『수산경영론집』, 제38권 2호, 한국수산경영학회, pp. 41~62.
- 와쿠이 요시유키·와쿠이 사다미 지음, 『처음 배우는 딥러닝 수학』, 박광수 역(2018), 한빛미디어.
- 이안 굿펠로·요슈아 벤지오·러런 쿠빌 지음, 『심층 학습』, 류광 역(1998), 제이펍.
- 이진면 외(2007), 「KIET 산업경제계량모형」, 산업연구원, pp. 118~119.
- 이춘수·양성범(2017a), 「생육정보를 이용한 가을배추와 가을무 단수 예측 모형 개발」, 『한국유기농업학회지』, 제25권 2호, 한국유기농업학회, pp. 279~293.
- _____(2017b), 「인공 신경망을 이용한 채소 단수 예측 모형 개발 : 고추를 중심으로」, 『한국유기농업학회지』, 제25권 3호, 한국유기농업학회, pp. 555~567.
- 이현동 외 (2018), 『수산부문 전망모형 「KMI-FOSiM」 구축 연구(1/2) - 양식산업 전망 모형 구축을 중심으로 -』, 한국해양수산개발원.
- 이현동·허수진·하현정(2020), 「김 수급 전망모형 구축에 관한 연구」, 『한국도서연구』, 제32권 4호, 한국도서학회.
- 장선덕·진평·박기영(1983), 「양식김의 광합성에 미치는 수온, 염분 및 부이의 영향」, 『한국수산과학회지』, 제16권 4호, 한국수산과학회, pp. 335~340.
- 장영진 외(1999), 「양식넙치(*Paralichthys olivaceus*)의 생리조건에 미치는 연속적인 수온 급강하의 영향」, 『한국수산과학회지』, 제32권 5호, 한국수산과학회, pp. 601~606.
- 조영제 외(2012), 『수산의 이해』, 블루앤노트.
- 조용호(2006), 「Forecasting Total Marine Production through Multiple Time Series Model」, 『한국데이터정보과학회지』, 제17권 1호, 한국데이터정보과학회, p

p. 63~76.

- 최종두·최영준(2014), 「기후변화로 인한 수온상승이 굴양식 본양성 생산방식의 경제성에 미치는 영향분석」, 『Ocean and Polar Research』, 제36권 2호, 한국해양과학기술연구원, pp. 157~163.
- 한석호(2016), 「쌀 단수 예측모형 개발 연구」, 『농업생명과학연구』, 제50권 3호, 경상대학교 농업생명과학연구원, pp. 219~229.
- 허수진 외(2020), 『김 중기 수급전망모형 「Gim-MFoS」 구축 연구』, 한국해양수산개발원.
- 홍정표·김영섭·허성범(2000), 「온도자극 및 수용밀도에 따른 우렁챙이 폐사」, 『한국양식학회지』, 제13권 4호, 한국수산과학회, pp. 285~293.
- 황강석·최정화·오택운(2012), 「인공신경망을 활용한 고등어의 위판가격 변동 예측 : 어획량 제한이 없었던 TAC제도 시행 이전의 경우」, 『한국어업기술학회지』, 제48권 1호, 한국수산해양기술학회, pp. 72~81.

국외 문헌

- Abigail See(2019), “Vanishing Gradients and Fancy RNNs”, *Natural Language Processing with Deep Learning*, p. 37
- A. J. Tacon(2003), “Aquaculture production trends analysis”. *FAO Fisheries Circular*, 886, pp. 5-29.
- Balakrishnan, M. *et al.*(2007), “Neural network and its application in aquaculture”, *Bioinformatics and statistics in Fisheries Research*, 3, pp. 145-151.
- Banerjea, S. M.(1967), “Water Quality and Soil Condition of Fishponds in Some State s of India in Relation to Fish Production”, *India Journal of Fisheries*, Vol. 14(1-2), pp. 115-144.
- Benzer, R. & S. Benzer(2015), “Application of artificial neural network into the fresh water fish caught in Turkey”, *International of Fisheries and Aquatic Studies*, 2(5), pp. 341-346.
- Brosse, S., S. Lek, & F. Dauba(1999), “Predicting Fish Distribution in a Mesotrophic

-
- Lake by Hydroacoustic Survey and Artificial Neural Networks”, *American Society of Limnology and Oceanography*, Vol. 44(5), pp. 1293-1303.
- Cacho, O. J.(1997), “Systems Modelling and Bioeconomic Modelling in Aquaculture”, *Aquaculture Economics & Management*, Vol. 1(1), pp. 45-64.
- Cook, S. (2019). “Forecast Evaluation using Theil’s Inequality Coefficients”, School of Management, Swansea University
- Czerwinski, I. A., J. C. Gutierrez-Estrada & J. A. Hernando-Casal(2007). “Short-term forecasting of halibut CPUE: Linear and non-linear univariate approaches”, *FISHERIESRESEARCH*, 86, pp. 120-128.
- Downing, J. A., C. Plante & S. Lalonde(1990), “Fish Production Correlated with Primary Productivity, not the Morphoedaphic Index”, *Canadian Journal of Fisheries and Aquatic Sciences*, 47, pp. 1929-1936.
- Ezenwaji E. E. *et al.*(2014), “An Analysis of the Relationship between Temperature Variation and Fish Production in Lagos, Nigeria”, *IOSR Journal of Agriculture and Veterinary Science*, Vol.7(11), pp. 38-43.
- Folorunso, T. *et al.*(2019). “Water Quality Index Estimation Model for Aquaculture System Using Artificial Neural Network”, *Journal of Advanced Computer Engineering Technology*, 5(3), pp. 195-204.
- Free, C. M. *et al.*(2019), “Impacts of historical warming on marine fisheries production”, *Science*, Vol. 363(6430), pp. 979-983.
- Glorot, X., A. Bordes & Y. Bengio(2011), “Deep sparse rectifier neural networks”, *In Proceedings of the fourteenth international conference on artificial intelligence and statistics*, JMLR Workshop and Conference Proceedings, pp. 315-323.
- Günther, F. & Fritsch, S.(2010), “Neuralnet: training of neural networks”. *R.J.*, 2(1), pp. 30-38
- Guo, W.W. & H. Xue.(2014), “Crop Yield Forecasting Using Artificial Neural Networks”, *Mathematical Problems in Engineering*, Hindawi, Vol. 2014, Article ID 857865, pp. 1-7.
- Hornik, K., M. Stinchcombe & H. White(1989), “Multilayer feedforward networks

- are universal approximators", *Neural networks*, Vol. 2(5), pp. 359-366.
- Kim, S. *et al.*(2007), "Climate variability and its effects on major fisheries in Korea", *Ocean Science Journal*, 42(3), pp. 179-192.
- Kingma, D. P. & J. Ba(2014), "Adam: A method for stochastic optimization", .
- Ji, B., Y. Sun, S. Yang & J. Wan(2007), "Artificial neural networks for rice yield prediction in mountainous regions", *Journal of Agriculture Science*, 145, Cambridge University Press, pp. 249-261.
- Laxmi, R. R. & A. Kumar(2011), "Weather based forecasting model for crops yield using neural network approach", *Statistics and Applications*, Vol. 9(1&2), pp. 55-69.
- Leach, J. H. *et al.*(2011), "A Review of Methods for Prediction of Potential Fish Production with Application to the Great Lakes and Lake Winnipeg", *Canadian Journal of Fisheries and Aquatic Sciences*, 44, pp. 471-485.
- Leon C. J., J. M. Hernandez & M. Leon-Santana(2005), "The effects of water temperature in aquaculture", *Applied Economics*, pp. 2159-2168.
- Mackenzie, B. R. & F. W. Koster(2004), "Fish Production and climate: sprat in the Baltic sea", *Ecology*, Vol. 85(3), pp. 784-794.
- Melita, B., L. Bavcevic, S. Grguric, F. Vresnik, J. Krizan, and O. Antonic(2020), "Modeling the Environmental Footprint of Sea Bream Cage Aquaculture in Relation to Spatial Stocking Design", *Journal of Environmental Management*, 270, 110811.
- Meynecke, J. O. *et al.*(2006), "Effect of Rainfall as a Component of Climate Change on Estuarine Fish Production in Queensland, Australia", *Estuarine Coastal and Shelf Science*, 69, pp. 491-504.
- Panda, S. S., D. P. Ames & S. Panigrahi(2010), "Application of vegetation indices for agricultural crop yield prediction using neural network techniques". *Remote Sensing*, 2(3), pp. 673-696.
- Poorbagher, H., S. Eagderi, R. & Nahavandi(2021), "Modelling the time series of capture fishery and aquacultural production in Iran", *International Journal of Aquatic Biology*, Iranian Society of Ichthyology, pp. 200-206.

-
- Paul, R. K. & M. K. Das (2010), "Statistical modelling of inland fish production in India", *Journal of the Inland Fisheries Society of India*, Vol. 42(2), pp. 1-7.
- Pomeroy, R. *et al.* (2008), "Bioeconomic Modelling and Salmon Aquaculture: An Overview of the Literature", *International Journal of Environment and Pollution*, Vol. 33(4), pp. 485-500.
- Zhang, A. *et al.* (2019), "Dive Into Deep Learning (Tools for Engagement)", Corwin Publishers, p. 162.
- Raman, R. K., T. V. Sathianandan & A. P. Sharma (2017), "Modelling and forecasting marine fish production in Odisha using seasonal ARIMA model", *The National Academy of Sciences*, 40(6), pp. 393-397.
- Skaggs, R. K. & Snyder, D. L. (1991). "Accuracy in Forecasting: The Results of a Competition" *Economics Research Institute Study Paper*, 91(2), pp. 1-25.
- Torkashavand, A. M., A. Ahmadi & N. L. Nikravesh (2017), "Prediction of kiwifruit firmness using fruit mineral nutrient concentration by artificial neural network (ANN) and multiple linear regressions (MLR)", *Journal of Integrative Agriculture*, 16(7), Elsevier, pp. 1634-1644.
- Vikas, L. & V. S. Dhaka (2014), "Wheat yield prediction using artificial neural network and crop prediction techniques", *International Journal for Research*, Vol. 2(9), pp. 330-341.
- Wik, T. E. I., B. T. Linden, & P. I. Wramner (2009), "Integrated Dynamic Aquaculture and Wastewater Treatment Modelling for Recirculating Aquaculture Systems", *Aquaculture*, 287, pp. 361-370.
- Yadav, V. K., S. Jahageerdar & V. Ramasubramanian (2016), "Use of different approaches to model catch per unit effort (CPUE) abundance of fish", *Indian Journal of Geo Marine Sciences*, Vol. 45(12), pp. 1677-1687.
- Zhang, A., Lipton, Z. C., Li, M. & Smola, A. J. (2021). "Dive into deep learning". arXiv preprint arXiv:2106.11342.

인터넷 자료

- 국립수산물과학원, 어장 환경모니터링 통계검색, https://www.nifs.go.kr/femo/data_statistics.femo (검색일: 2021. 9. 23.)
- 네이버 지식백과, 「영양염」, <https://terms.naver.com/> (검색일: 2021. 9. 23.)
- 연합뉴스(2021. 7. 12.), 「민물 폭탄에 강진만 바다 전복 2천200만마리 모조리 폐사」, <https://www.yna.co.kr/view/AKR20210712123200054?input=1195m> (검색일: 2021. 9. 23.)
- 위키백과, 「순환신경망」, <http://www.wikipedia.org/> (검색일: 2021. 9. 24.)
- 통계청, 어업생산동향조사, <http://kosis.kr/> (검색일: 2021. 9. 24.)
- 통계청, 천해양식어업권통계, <http://kosis.kr/> (검색일: 2021. 9. 24.)
- Andrej Karpathy blog(2015. 5. 20.), 「The Unreasonable Effectiveness of Recurrent Neural Networks」, <http://karpathy.github.io/2015/05/21/rnn-effectiveness/> (검색일: 2021. 9. 24.)
- CS231n: Convolutional Neural Networks for Visual Recognition, 「Neural Networks and Backpropagation」, <http://cs231n.stanford.edu/schedule.html>(검색일: 2021. 9. 24.)
- Kaggle, 「Overfitting and Underfitting」, <https://www.kaggle.com> (검색일: 2021. 9. 24.)
- Medium(2019. 7. 23), 「How to work with Time Distributed data in a neural network」, <https://medium.com/> (검색일: 2021. 9. 24.)
- towards data science(2020. 7. 27), 「From a LSTM cell to a Multilayer LSTM Network with PyTorch」, <https://towardsdatascience.com> (검색일: 2021. 9. 24.)

법령 자료

- 「어장관리법」(2020. 8. 28., 법률 제16568호).

부록

비교모형(다중회귀모형) 추정 결과

1) 전남

(1) 전체 투입변수 추정 결과(모형1~모형6)

전남지역의 단수를 투입변수 및 시차에 따라 6개 모형으로 구분하여 추정하였다. 그 결과 모든 모형의 결정계수 값이 80% 이상으로 비교적 높게 나타났으며, 특히 기상변수와 수질변수, 계절더미를 모두 고려하고 5개월 시차를 적용한 모형6이 가장 높은 것으로 나타났다.

세부 추정 결과를 살펴보면, 기상요인만을 고려한 모형1, 모형2는 계절더미를 제외하면 월평균풍속의 영향력이 가장 큰 것으로 나타났다. 모형1의 경우 1개월 전 월평균풍속이 1m/s 증가하면, 김 단수는 6.45속, 모형2는 6.07속 감소하는 것으로 나타났다. 계절적 영향으로는 주 출하시기인 3월의 영향이 가장 큰 것으로 도출되었다.

수질변수만 고려한 모형 3, 4의 경우 모형4의 결정계수 값이 더 크게 나타나, 1개월 전의 수질보다는 5개월 전의 수질이 김 단수에 미치는 영향이 큰 것으로 분석된다. 모형4에서 표층수온과 표층용존산소가 통계적으로 유

의하게 나타났는데, 두 변수 모두 음(-)의 영향을 미치는 것으로 도출되었다. 계절더미의 경우 11월을 제외한 모든 월이 유의한 것으로 나타났다.

마지막으로 기상변수, 수질변수를 모두 고려한 모형5와 모형6 추정 결과, 모형5에서는 1개월 전의 표층수온을 제외한 모든 수질변수가 유의하지 않았으며, 계절더미도 모든 월이 유의하지 않은 것으로 나타났다. 모형6에서는 반대로 5개월 전의 기상변수 전체가 유의하지 않다는 결과가 도출되었으며, 계절더미는 11월과 5월을 제외한 모든 월이 유의한 것으로 나타났다.

〈전남지역 추정 결과(모형1~모형6)〉

구분	모형1	모형2	모형3	모형4	모형5	모형6
상수	6.03	13.76 **	35.30 *	41.60 **	30.37	38.05 **
월간강수량	0.02	-0.01	-	-	-0.01	-0.01
일최대강수량	-0.03	0.06 **	-	-	-0.01	0.05
월평균풍속	-6.45 ***	-6.07 ***	-	-	-5.06 **	-2.19
최대순간풍속	-0.09	0.14	-	-	0.33	0.08
일조량	0.06 ***	-0.00	-	-	0.05 *	-0.01
표층수온	-	-	-1.22 *	-1.36 **	-1.19 *	-1.01
표층무기인농도	-	-	-102.5 0	200.83	-192.8 4	222.64
표층용존질소	-	-	9.21	5.26	8.10	-5.44
표층용존산소	-	-	-0.41	-2.14 **	-0.30	-1.73 *
표층클로로필농도	-	-	-0.02	-0.16	0.04	-0.14
계절더미(11월)	1.17	0.62	-1.62	5.86	-3.26	3.65
계절더미(12월)	10.25 ***	5.06 **	-1.01	14.85 **	0.14	11.15 *
계절더미(1월)	19.34 ***	11.14 ***	1.20	25.44 **	4.26	20.57 ***
계절더미(2월)	26.14 ***	17.34 ***	6.52	29.61 **	8.96	25.96 ***
계절더미(3월)	34.75 ***	26.24 ***	12.40	36.03 **	15.63	32.94 ***
계절더미(4월)	20.50 ***	16.11 ***	3.02	16.40 **	4.28	16.18 ***
계절더미(5월)	1.41	-0.21	-14.56 *	-11.24 **	-12.74	-8.99
결정계수(R^2)	0.83	0.82	0.82	0.85	0.85	0.86
Durbin-Watson	1.33	0.94	0.85	0.93	1.42	0.93

주: ***, **, *는 각각 1% 5%, 10% 유의수준에서 통계적으로 유의함을 의미함. 해당 내용은 이어지는 비교 모형 추정 결과에서도 모두 동일하게 적용됨
자료: 저자 작성

(2) 선별 투입변수 추정 결과(모형7~모형8)

모형1부터 모형6까지의 추정 결과를 통해 전남지역의 1개월 및 5개월 전 기상변수, 수질변수가 김 단수에 미치는 영향을 살펴보았다. 그러나 모형1~모형6 결과 모두 Durbin-Watson 검정 결과, 자기상관이 존재하는 것으로 나타나 통계적으로 효율적인 추정량으로 보기는 어렵다.

〈전남지역 추정 결과(모형7~모형8)〉

구분	모형7	모형8
상수	49.32***	-13.14*
월간강수량	-0.07***	-
일최대강수량	0.08**	-
월평균풍속	-8.62***	-8.84***
최대순간풍속	-	0.38**
일조량	-0.09***	-
표층수온	-1.43***	1.59***
표층무기인농도	-	729.64***
표층용존질소	-31.87*	-
표층용존산소	3.19***	-
표층클로로필농도	0.48**	-
계절더미(3월)	-	18.49***
DUM11	39.32***	-
DUM12	-24.00***	-
DUM13	16.12***	-
DUM21	-	27.02***
DUM22	-	18.26***
DUM23	-	13.92***
결정계수(R^2)	0.85	0.83
Durbin-Watson	1.71	1.84

주: DUM11(2017. 3.=1, 그 외 0), DUM12(2019. .5=1, 그 외 0), DUM13(2014. 4, 2015. 10, 2017. 2, 2017. 4=1, 그 외 0), DUM21(2017. 3, 2017. 4=1, 그 외 0), DUM22(2012. 4, 2013. 4, 2015. 10, 2017. 2, 2018. 2=1, 그 외 0), DUM23(2011. 4, 2015. 4, 2018. 4=1, 그 외 0)

자료: 저자 작성

즉 이러한 추정 결과를 인공신경망 결과와 비교하기에는 한계가 있을 것으로 판단하여, 통계적 유의성, 자기상관 등을 모두 고려한 모형을 추가적으로 추정하였다. 추정 결과 1개월 시차를 준 모형7과 5개월 시차를 준 모형8 모두 결정계수 값이 80% 이상으로 양호하게 나타났으며, 자기상관 또한 없는 것으로 나타났다.

모형7을 살펴보면 결정계수가 0.85로 모형8보다 조금 더 높게 나타났으며, 기상변수와 수질변수 중 유의한 영향을 미치는 변수도 더 많은 것으로 도출되었다. 개별 변수 중에는 표층수온의 t 값이 가장 컸는데, 표층수온이 1도 상승하면 김 책당 생산량은 1.43속만큼 감소하는 것으로 나타났다.

모형8은 결정계수가 0.83으로, 5개월 전의 월평균풍속, 최대순간풍속, 표층수온, 표층무기인농도가 당월 김 단수에 유의한 영향을 미치는 것으로 나타났다. 모형7과 마찬가지로 표층수온의 t 값이 가장 컸는데, 계수 값이 양(+)의 영향을 미치는 것으로 나타나, 음(-)의 상관관계가 있던 모형7과 상이하였다.

2) 완도

(1) 전체 투입변수 추정 결과(모형1~모형6)

완도지역의 단수를 기상변수, 수질변수를 투입하여 추정한 결과, 기상변수만을 고려한 모형1, 모형2를 제외한 모든 모형의 결정계수가 0.8 이상으로 나타났다. 특히 Durbin-Watson 검정 결과 자기상관이 존재하는 것으로 나타난 다른 모형과 달리, 모형6은 자기상관도 존재하지 않는 것으로 도출되어 가장 우수한 것으로 평가되었다.

세부 모형 결과를 살펴보면, 모형1, 모형2는 유의한 영향을 미치는 변수가 동일했는데, 두 모형 모두 계절더미를 제외하면 일조량의 t 값이 가장 큰

것으로 나타났다. 수질변수만 고려한 모형 3, 4의 경우 대부분의 변수가 유의하지 않은 것으로 나타났는데, 모형3에서는 표층용존산소만 유의하였으며, 모형4는 계절더미만 유의한 영향을 미치는 것으로 나타났다.

〈완도지역 추정 결과(모형1~모형6)〉

구분	모형1	모형2	모형3	모형4	모형5	모형6
상수	-2.64	-5.34	4.45	0.75	-22.38	2.78
월간강수량	0.02	0.00	-	-	-0.01	-0.02
일최대강수량	0.03	0.09	-	-	0.06	0.10*
월평균풍속	-3.81*	-4.14*	-	-	0.70	9.06***
최대순간풍속	-0.60	0.20	-	-	0.01	-0.36
일조량	0.14***	0.07*	-	-	0.11**	0.00
표층수온	-	-	-0.87	-0.26	-0.53	-1.27
표층무기인농도	-	-	-298.65	509.38	-408.52	607.64*
표층용존질소	-	-	56.34	-18.23	43.45	-20.18
표층용존산소	-	-	3.06**	1.15	2.53*	-0.40
표층클로로필농도	-	-	0.32	0.60	0.35	0.37
계절더미(11월)	0.01	3.67	-3.66	-0.59	-3.48	6.14
계절더미(12월)	9.80	2.28	-6.40	1.53	1.75	10.67
계절더미(1월)	28.13***	14.67*	6.23	19.03	17.73	29.17**
계절더미(2월)	44.15***	34.69***	23.39	38.93***	34.01*	52.16***
계절더미(3월)	52.82***	43.70***	30.32	51.08***	41.23**	60.32***
계절더미(4월)	37.93***	40.91***	22.20	37.21***	27.36	42.87***
계절더미(5월)	7.79	13.60	-9.55	-0.05	-4.77	-0.58
결정계수(R^2)	0.76	0.73	0.81	0.85	0.83	0.88
Durbin-Watson	1.11	0.83	0.96	1.08	1.20	1.60

자료: 저자 작성

마지막으로 모형 5와 6은 유의한 영향을 미치는 변수가 달랐는데, 시차를 1개월로 설정한 모형5는 일조량과 표층용존산소가 유의한 것으로 나타

났고, 5개월의 시차가 적용된 모형6은 일최대강수량, 월평균풍속, 표층무기인농도가 유의한 것으로 도출되었다. 결정계수 값이 가장 큰 모형6의 경우, 기상변수와 수질변수 중 월평균풍속의 t값이 가장 큰 것으로 나타났는데, 월평균풍속이 1m/s만큼 강해지면 김 단수는 9.06속 증가하는 것으로 추정되었다.

(2) 선별 투입변수 추정 결과(모형7~모형8)

통계적 유의성, 자기상관 등을 고려하여 완도지역의 단수를 추정하는 최적의 모형을 추정한 결과, 5개월 이전의 기상, 수질 변수들이 김 단수에 영향을 미치는 모형8의 결정계수 값이 0.91로 높게 나타났다. 모형8은 일최대강수량, 표층수온, 표층무기인농도 등이 유의한 영향을 미치는 것으로 나타났는데, 표층무기인농도의 t값이 5.5로 기상변수와 수질변수 중 가장 높았다.

〈완도지역 추정 결과(모형7~모형8)〉

구분	모형7	모형8
상수	63.12 ***	-28.67 ***
월간강수량	-0.22 ***	-
일최대강수량	0.35 ***	0.13 ***
월평균풍속	-	-
최대순간풍속	-	-
일조량	-	-
표층수온	-4.15 ***	1.23 ***
표층무기인농도	-484.01 **	1388.74 ***
표층용존질소	-	-
표층용존산소	3.34 ***	-
표층클로로필농도	2.22 ***	-
계절더미(2월)	-	29.74 ***
계절더미(3월)	-	51.03 ***
계절더미(4월)	-	34.40 ***

구분	모형7	모형8
SD21	-	-9.31 ***
DUM11	31.70 ***	-
DUM12	-31.34 ***	-
DUM13	21.98 ***	-
DUM21	-	17.68 ***
DUM22	-	-27.70 ***
DUM23	-	-17.03 ***
결정계수(R^2)	0.84	0.91
Durbin-Watson	1.68	1.73

주: SD21(2018. 3. 이전 0, 이후1), DUM11(2014. 2, 2014. 4, 2015. 10=1, 그 외 0), DUM12(2017. 5, 2019.5, 2020. 5=1, 그 외 0), DUM13(2014. 3, 2018. 4=1, 그 외 0), DUM21(2011. 4, 2011. 5, 2014. 2, 2015. 1=1, 그 외 0), DUM22(2020. 3=1, 그 외 0), DUM23(2016. 3, 2019. 3=1, 그 외 0)

자료: 저자 작성

3) 고흥

(1) 전체 투입변수 추정 결과(모형1~모형6)

고흥지역의 단수 추정 결과, 대체적으로 모든 모형의 설명력이 낮은 것으로 나타났는데, 특히 기상요인만 고려한 모형1, 2는 결정계수가 0.5 이하로 매우 낮은 설명력을 보였다.

〈고흥지역 추정 결과(모형1~모형6)〉

구분	모형1	모형2	모형3	모형4	모형5	모형6
상수	-20.84	11.68	60.91	13.99	78.90	30.72
월간강수량	0.10	-0.02	-	-	0.06	-0.02
일최대강수량	-0.20	0.09	-	-	-0.26	0.04
월평균풍속	-0.30	-6.83	-	-	2.53	-1.00
최대순간풍속	-0.28	0.45	-	-	-0.10	0.20
일조량	0.14	-0.05	-	-	0.04	-0.16**
표층수온	-	-	-2.41	-0.41	-3.27*	0.84

구분	모형1	모형2	모형3	모형4	모형5	모형6
표층무기인농도	-	-	701.57	-563.88	622.60	72.43
표층용존질소	-	-	-67.06	89.94	-23.69	-14.76
표층용존산소	-	-	-0.47	-0.85	-0.81	-1.06
표층클로로필농도	-	-	-0.01	-0.17	-0.05	-0.30
계절더미(11월)	16.03	12.81	10.30	18.69	12.74	3.41
계절더미(12월)	40.51**	32.83*	20.02	42.75**	10.26	21.78
계절더미(1월)	46.87***	40.24**	15.93	52.76**	-1.43	32.08
계절더미(2월)	36.86**	30.53	3.12	42.80**	-16.47	23.98
계절더미(3월)	28.41	21.93	-13.97	30.82*	-30.09	19.68
계절더미(4월)	3.98	3.67	-28.48	4.26	-43.24	-3.88
계절더미(5월)	2.12	4.34	-27.08	-5.24	-28.10	-5.81
결정계수(R^2)	0.48	0.50	0.59	0.56	0.65	0.62
Durbin-Watson	0.92	0.92	1.20	1.15	1.43	1.35

자료: 저자 작성

결정계수만 봤을 때는 기상변수와 수질변수를 모두 고려한 모형 5, 6이 가장 높게 나타났는데, 통계적으로 유의한 변수는 1개만 존재하고 자기상관 또한 존재하여 전반적으로 고흥지역의 단수 추정에는 적합한 모형이 도출되지 않았다.

(2) 선별 투입변수 추정 결과(모형7~모형8)

앞의 모형1~모형6에서 고흥 단수를 추정하기에 적합한 모형이 존재하지 않아, 기상변수와 수질변수를 고려하여 설명력이 높은 모형을 재추정하였다. 그 결과 유의한 영향을 미치는 변수가 다르게 나타났는데, 모형7의 경우 1개월 이전 월간강수량, 일최대강수량, 일조량, 표층무기인농도가 유의한 것으로 나타났으며, 모형8은 5개월 이전의 최대순간풍속, 표층수온이 유의하게 추정되었다.

모형7과 모형8의 결정계수 값은 각각 0.81, 0.83으로 비슷하게 추정되었으며, 자기상관도 존재하지 않는 것으로 나타났다.²³⁾ 모형7에서는 1개월 이전의 월간강수량과 일조량이 비슷한 t값을 보였는데, 두 변수 모두 김 단수에 음(-)의 영향을 미치는 것으로 나타났다. 반면 모형 8에서는 5개월 이전의 표층수온 t값이 가장 컸는데, 5개월 이전 표층수온이 1도 상승하면 김 단수가 3.67속만큼 증가하는 것으로 추정되었다.

전남과 완도지역의 선별 투입변수 추정 결과와 비교하면 모형의 설명력도 다소 낮은 것으로 나타났고, 더미변수도 상대적으로 많이 투입되어²⁴⁾ 기상변수와 수질변수가 고흥지역의 김 단수를 설명하는 데는 다소 한계가 있는 것으로 판단되었다.

〈고흥지역 추정 결과(모형7~모형8)〉

구분	모형7	모형8
상수	62.90***	-71.88***
월간강수량	-0.27***	
일최대강수량	0.21**	
월평균풍속		
최대순간풍속		1.16***
일조량	-0.20***	
표층수온		3.67***
표층무기인농도	886.93***	
표층용존질소		
표층용존산소		
표층클로로필농도		
SD21		7.28**

23) Durbin-Watson 검정결과가 1.5로 자기상관 문제가 의심되어 추가적으로 자기상관 검정(Breusch-Godfrey LM Test)을 실시하였는데, 자기상관이 존재하지 않는 것으로 나타났다.

24) 더미변수는 독립변수로 설명할 수 없는 이상치를 고려하기 위해 투입되는 것으로, 더미변수가 많이 투입되었다는 것은 투입된 독립변수만으로 종속변수를 설명하기 어려움을 의미한다.

구분	모형7	모형8
DUM11	22.35***	
DUM12	52.35***	
DUM13	43.03***	
DUM14	30.68***	
DUM21		23.34***
DUM22		39.07***
DUM23		-35.04***
DUM24		31.05***
결정계수(R^2)	0.81	0.83
Durbin-Watson	1.52	1.50

주: SD21(2017. 2 이전 0, 이후1), DUM11(2010. 11, 2011. 11, 2019. 2=1, 그 외 0), DUM12(2011. 12, 2012. 12=1, 그 외 0), DUM13(2013. 1, 2018. 1=1, 그 외 0), DUM14(2010. 12, 2017. 3, 2018. 2, 2019. 1=1, 그 외 0), DUM21(2010. 11, 2010. 12=1, 그 외 0), DUM22(2012. 12=1, 그 외 0), DUM23(2016. 12, 2017. 1=1, 그 외 0), DUM24(2011. 12, 2017. 3, 2018. 1, 2018. 2=1, 그 외 0)

자료: 저자 작성

기본연구보고서 발간목록

Ⅰ 2021년

01	시민참여형 해안돌봄 정책 도입방안 연구	정지호
02	해양 플라스틱 쓰레기가 선박 운항에 미치는 영향분석 연구	김보람
03	여성어업인 노동정책 방향 연구	홍혜수
04	순환경제 시스템을 활용한 어업폐기물의 자원화 방안 연구	고동훈
05	1conomy 시대, HMR 시장 확대에 따른 수산부문 대응전략 수립 연구	마창모
06	AIS 데이터 기반 해상교통 안전 평가모델 개발 연구	황선일
07	항만 에너지 관리시스템 도입을 위한 로드맵 구축 연구 - 부산항 신항 전기에너지 소비를 중심으로 -	김근섭
08	해양분야 리빙랩 활성화 방안 연구	좌미라
09	빅데이터의 연계·활용을 통한 선박의 배출량 산정체계 고도화 방안 연구	안용성
10	해양공간계획 집행체계 고도화 방안 연구	최희정
11	수산식품산업 푸드테크 환경분석 및 적용전략 연구	이상건
12	AIS데이터기반 해상물동량 추정 연구	황수진
13	항만부문 정책의사결정 지원 빅데이터 플랫폼 구축 방안 연구	이기열
14	항만 컨테이너 물류 프로세스 디지털화 모형 연구	서정용
15	해양수산 분야 국제산업연관표 구축기반 연구	정수빈
16	해양수산 과학기술 정책평가모형 연구 - 해양수산 R&D를 중심으로 -	전형모
17	글로벌 경제위기와 해운산업 대응체계 연구	박성화
18	항만산업 경기진단체계 구축 연구 - 경기동행지수 개발을 중심으로 -	김성아
19	전국 무역항 부두 분류 체계 개선방안 연구	이수영
20	IMO 규제기반 해사산업의 글로벌 지속발전방안 연구 - 新해사산업의 육성 및 지원을 위한 법제화 연구(5차년도) -	박한선
21	수소에너지 거점 구축을 위한 항만의 대응방안 연구	신수용

2020년

01	해양수산업의 지역 간 연관구조 분석	장정인
02	해양수산 분야 기술 대외의존도 분석연구 - 스마트항만을 중심으로 -	전형모
03	지역 해양수산 오픈 플랫폼 구축방안 연구	최지연
04	갯벌 거버넌스 개선방안에 관한 연구	육근형
05	해양환경정책의 능동적 추진을 위한 재원체계 개선 방안 연구	박수진
06	항만 대기환경 관리 표준 및 평가모형 연구	안용성
07	해양수산분야 사회문제해결형 R&D 기반 구축 연구	좌미라
08	해양 유입 하천쓰레기 관리체계 개선방안	이윤정
09	수산식품 품질·안전관리 제도 개선방안 연구	이현동
10	국제법 변화에 대응한 어선원 안전 및 권리 제고방안 연구	한덕훈
11	스마트 양식 클러스터 추진 방안 수립 연구	이상철
12	해양포유류 보호에 관한 수산업 대응 방안 연구	정명화
13	수산물 수급통계 개선 방안 연구	김수현
14	IMO 온실가스 규제 대응 정책방향 연구	박한선
15	퇴직전문가 활용을 통한 해운업 경쟁력 강화방안 연구	안영균
16	글로벌 선사들의 물류통합화 전략에 대한 국적선사의 대응방안	전형진
17	내항여객운송항로 정책 발전방안 연구	김태일
18	블록체인 기술기반 식품콜드체인 체계 구축 연구	조지성
19	항만자동화 도입 관련 노무 갈등 해소 방안 연구	김찬호
20	스마트항만과 스마트도시 연계 발전 방안 연구	이연경
21	항만의 회복탄력성 측정 모형 구축에 관한 연구	김성기
22	IMO 규제기반 해사산업의 글로벌 지속발전방안 연구 - 新해사산업의경제적파급효과분석연구(4차년도) -	박한선
23	국내 항만연계 산업의 가치사슬 및 공급사슬 연계성 강화방안 - 자동차 산업을 중심으로	신수용

수시연구보고서 발간목록

Ⅰ 2021년

01	해양환경산업 육성 방안 연구	한기원
02	선원의 인권 및 근로여건 향상을 위한 선원근로감독관 제도 개선 방안 연구	허성례
03	포스트 코로나 시대를 대비하는 지역 해양축제 활성화 연구	최일선
04	회복탄력성 개념을 적용한 항만 위기관리 시스템 구축 연구	김성기
05	중대재해처벌법 시행에 따른 항만에서의 대응방안 연구	최상균
06	크루즈산업 COVID-19 방역체계 구축방안 연구	황진희
07	한국형 선주사의 최적 운영방안 연구	고병욱
08	어촌형 생활서비스 전달체계 개선 연구	이호림
09	원양산업의 ESG 도입 기초 연구	윤미경
10	수상레저활동 관리체계 개선방안 연구	홍장원
11	연안지역 자연성 회복을 위한 정책 방향_간척지·담수호 중심으로	정지호
12	스마트항만 구축에 따른 탄소저감 효과 연구 - 동력전환을 중심으로 -	김가현

Ⅰ 2020년

01	포스트 코로나 19 해양수산 분야 정책방안	박광서
02	생분해성 어구 사용 활성화 방안 연구	심성현
03	해양법 전문인력 양성 방안 연구	박영길
04	무인도서 해양주권 강화와 이용 활성화를 위한 제도 개선방안	정지호
05	지역 해양수산 재정분권 대응방향	황재희
06	데이터 3법 개정에 따른 항만·물류 데이터 활용도 제고 방안	이기열
07	양식 활어 유통 효율화 방안 연구	마창모
08	해양레저관광사업 추진을 위한 제도정비 방안	홍장원
09	해양바이오기업의 규제 정비 방안 연구	최석문
10	비상체제 시 선원의 안전을 위한 선박-항만-항공 이동경로 구축방안	이혜진

일반연구보고서 발간목록

Ⅰ 2021년

01	해양수산 정책영향평가를 위한 기초 연구	김주현
02	해양교육의 사회·경제적 가치 평가 연구	이슬기
03	해양수산업 경기진단체계 기초연구 - 해운업을 중심으로	권장한
04	선박 기술진보를 고려한 탄소 배출량 추정 연구	최건우
05	서포트벡터머신(SVM) 기법을 활용한 해운시황 예측 연구	김병주
06	크루즈 여객 수요 전망 모형 구축 연구	안승현
07	디지털 공급사슬 물류정보통합 구축전략 연구(Ⅲ)	서정용
08	수출용 전복의 유통경로 분석 연구	이정필
09	인공지능기반 해상운임 예측 연구(3차년도)	황수진
10	인공신경망모형을 이용한 양식수산물 단수 전망에 관한 연구 - 김 양식을 중심으로 -	천성훈
11	해양수산분야 미래 리스크 발굴 및 파급효과 분석 연구	박광서

Ⅰ 2020년

01	지역 해양관광 경쟁력 지수 체계화 연구	최일선
02	AIS 기반 글로벌 선박 배기가스 배출량 분석 연구	강무홍
03	김 중기 수급전망모형 「Gim-MFoS」 구축 연구	허수진
04	중앙 북극 공해 비규제어업방지협정 이행방안 연구	김민수
05	동해 평화관광구역 조성 방안 연구	윤인주
06	디지털 공급사슬 물류정보통합 구축전략 연구(Ⅱ) - 일반 수출입 컨테이너 정보교환방식 중심 -	이연경
07	스트레스 테스트를 통한 우리나라 해운·조선 기업의 안정성 분석	박성화
08	인공지능기반 해상운임예측 연구	황수진
09	국제물류주산업 실태분석 및 경쟁력 제고 방향 연구	최나영환
10	디지털화에 따른 개별 직업의 대체 가능성 추정 연구	박희대
11	시계열 분석을 통한 해운시장 분석 및 예측 연구	고병욱
12	해양수산분야 글로벌 경제효과 분석모형(KMI-GEM) 시범 구축 연구	임병호
13	해운항만물류 인력양성사업 관리운영 제도 개선 방안 연구	이자연

일반연구 2021-10

**인공신경망모형을 이용한
양식수산물 단수 전망에 관한 연구**
- 김 양식을 중심으로 -

인 쇄 2021년 12월 29일

발 행 2021년 12월 31일

발 행 인 김 종 덕

발 행 처 한국해양수산개발원

주 소 49111 부산시 영도구 해양로 301번길 26(동삼동)

연 락 처 051-797-4800 (FAX 051-797-4810)

등 록 1984년 8월 6일 제313-1984-1호

조판·인쇄 애드원플러스 (070-4390-3850)

판매 및 보급: 정부간행물판매센터 Tel: 02-394-0337

정가 6,000원

