

Article

이에도 해양과학기지 수온 시계열 자료의 이상값 검출을 위한
국제 품질검사의 성능 평가

민용침¹ · 전현정¹ · 정진용¹ · 박승환¹ · 이재익¹ · 정종민¹ · 민인기¹ · 김용선^{2,3*}

¹한국해양과학기술원 해양재난·재해연구센터
(49111) 부산광역시 영도구 해양로 385

²한국해양과학기술원 해양순환연구센터
(49111) 부산광역시 영도구 해양로 385

³한국해양대학교 해양과학기술전문대학원 해양과학기술융합학과
(49112) 부산광역시 영도구 태종로 727

Evaluation of International Quality Control Procedures for Detecting
Outliers in Water Temperature Time-series at Ieodo Ocean Research Station

Yongchim Min¹, Hyunjung Jun¹, Jin-Yong Jeong¹, Sung-Hwan Park¹, Jaeik Lee¹,
Jeongmin Jeong¹, Inki Min¹, and Yong Sun Kim^{2,3*}

¹Marine Disaster Research Center, Korea Institute of Ocean Science & Technology,
Busan 49111, Korea

²Ocean Circulation Research Center, Korea Institute of Ocean Science & Technology,
Busan 49111, Korea

³Department of Convergence Study on the Ocean Science and Technology,
Ocean Science and Technology School, Korea Maritime and Ocean University, Busan 49112, Korea

Abstract : Quality control (QC) to process observed time series has become more critical as the types and amount of observed data have increased along with the development of ocean observing sensors and communication technology. International ocean observing institutions have developed and operated automatic QC procedures for these observed time series. In this study, the performance of automated QC procedures proposed by U.S. IOOS (Integrated Ocean Observing System), NDBC (National Data Buoy Center), and OOI (Ocean Observing Initiative) were evaluated for observed time-series particularly from the Yellow and East China Seas by taking advantage of a confusion matrix. We focused on detecting additive outliers (AO) and temporary change outliers (TCO) based on ocean temperature observation from the Ieodo Ocean Research Station (I-ORS) in 2013. Our results present that the IOOS variability check procedure tends to classify normal data as AO or TCO. The NDBC variability check tracks outliers well but also tends to classify a lot of normal data as abnormal, particularly in the case of rapidly fluctuating time-series. The OOI procedure seems to detect the AO and TCO most effectively and the rate of classifying normal data as abnormal is also the lowest among the international checks. However, all three checks need additional scrutiny because they often fail to classify outliers when intermittent observations are performed or as a result of systematic errors, as well as tending to classify normal data as outliers in the case where there is abrupt change in the observed data due to a sensor being located within a sharp boundary between two water masses, which is a common feature in shallow water observations. Therefore, this study underlines the necessity of developing a new QC algorithm for time-series occurring in a shallow sea.

Key words : Korea Ocean Research Stations (KORS), U.S. IOOS, NDBC, OOI, outliers

1. 서론

해양관측자료는 국가 차원의 재원을 바탕으로 생산되어 공공재의 성격을 가지고 있으며 그 활용도 제고를 위해 전 세계의 다양한 사용자에게 제공되어야 한다. 이러한 관측 자료는 자연 현상 분석을 위한 연구에 직접적으로 사용될 뿐만 아니라 수치 모형의 자료 동화나 빅데이터 기반의 예측 및 예보에 활용되어 인명 및 재산피해 예방에 기여할 수 있다. 통계 분석 등의 가공 과정을 거쳐 생산된 해양과학 정보는 사회적 의사 결정이나 국가 정책 결정에 주요한 기초 과학 자료로 활용될 수도 있다. 최근 해양관측자료의 종류와 양이 비약적으로 증가하고, 어업 활동이나 해양 레저와 같은 해양 활동 인구가 증가하면서 관측 자료의 최종 사용자가 일반인으로 확대되고 있다 (Weller et al. 2019).

관측 자료에는 다양한 이상값이 존재한다. Hawkins (1980)는 이상값을 ‘다른 방법으로 생산된 것으로 보일 정도로 대부분의 자료와 통계적 특성이 다른 소수의 자료’로 정의하였다. 이상값은 오류(error)나 소음(noise)이 대부분이나 실제 자연 현상이 반영된 경우도 있다. 일반적으로 관측 자료의 오류는 센서의 불량이나 노후, 누적 오차, 시스템의 한계나 인간의 실수 등으로 인하여 발생한다(이 2015). 이상값이 포함된 자료의 분석에는 Rosner 검정 방법과 같은 로버스트(robust) 추정 방법을 사용하여 이상값으로 인한 통계값의 편향 효과를 최소화할 수 있다(조 등 2016). 품질관리는 관측자료의 이상과 정상을 명확하게 구분하는 과정(OOI 2013)을 의미하며, 관측 및 자료 관리에 있어 자료 품질개선에 대한 보다 본질적인 해결책으로 이와 같은 품질관리를 필요로 한다. 이 논문에서 ‘오류에 해당하는 이상값’을 통상 이상값으로, 이상값을 찾는 활동을 품질관리로 정의하였다.

품질관리는 과거 개인이나 연구 과제 내에서 자체적으로 개발하여 활용하는 것이 일반적이었으나, 최근에는 국제기관들이 생산하는 관측 자료의 효율적인 관리를 위해 표준화된 품질관리 절차를 구축하여 운영하고 있다. 정부간 해양학위원회(Intergovernmental Oceanographic Commission, IOC)는 여러 연구 그룹에서 사용하던 실시간 자료의 품질관리와 검증 방법을 정리하여 배포하고 있다(UNESCO 1993). 세계기상기구(World Meteorological Organization, WMO)는 전 세계적으로 공유되는 기상관측자료와 통계 자료의 균일한 배포를 위해 관측 절차와 자료를 표준화하고 품질관리를 위한 가이드를 만들어 운영하고 있다(WMO 2010). 미 통합해양관측시스템(U.S. Integrated Ocean Observing System, IOOS)과 국제자료부이센터(National Data Buoy Center, NDBC), 해양관측계획(Ocean Observatories Initiative, OOI) 등도 장기 시계열 자료의

품질 확보를 위해 표준화된 품질 관리 매뉴얼과 자동 품질관리 알고리즘을 제공하고 있다(IOOS 2014; NDBC 2009; OOI 2013). 특히 NDBC의 품질관리 절차는 국내 유관기관인 국립해양조사원, 기상청의 자체 품질관리에도 참고하여 운영 중이다(국립해양조사원 2019; 기상청 2019).

관측 시스템의 자동화와 정보통신 기술의 발달로 인해 고해상도의 해양-대기 시계열 자료가 실시간으로 생산되면서, 자료의 품질 개선을 위한 품질관리 시스템의 구축 혹은 품질검사 기법의 개선을 위한 연구가 다양하게 수행되고 있다. 한반도를 포함한 황해와 동중국해에서 생산되는 시계열 자료의 품질관리를 중심으로 살펴보면, 채 등(2014)은 복합기상센서에서 생산되는 시계열 자료에 대한 품질관리 시스템(Quality Control system for Integrated Meteorological Sensor Data, QCIMS)을 개발하여 그 결과를 기상청의 기상관측자료 실시간 품질관리시스템(real-time quality control system for meteorological observational data)과 비교하였다. Zhou et al. (2018)은 동중국해 해저 관측 시스템(East China Sea seafloor observatory system)에서 생산하는 층별 유황 및 유속, 수온, 염분, 수심, CO₂, pH, 클로로필, 탁도 등 시계열 자료의 품질개선을 위해 ARIMA (Auto Regressive Integrated Moving Average) 모델을 응용한 품질검사 알고리즘을 개발하였으며, 이 중 pH와 전기전도도의 시계열 자료에 품질관리 기법을 적용하여 그 성능을 평가하였다. 이 연구들은 각기 개발한 품질관리의 결과를 기존의 것과 비교하거나 단일 품질관리의 성능만을 평가하였으며 국제 사회에서 개발 및 활용하고 있는 품질관리에 대한 분석은 미비하였다. 최근, 민 등(2020)은 해양과학기 지 시계열 관측 자료의 품질개선을 위해 2013년 이어도 해양과학기 지 층별 수온 시계열 자료에 OOI에서 제공하는 품질관리를 적용하여 그 효과와 특성을 분석하였다. 이 연구는 OOI 품질관리 기법이 변동이 큰 천해 시계열 자료의 정상 자료를 이상으로 분류하였으며, 특히 이와 같은 오류는 주로 변화율 검사(gradient check)에서 발생하였음을 언급한 바 있다. 하지만 품질관리 기법에 대한 정성적인 분석만을 수행하였으며, 국제 품질관리 기법 간의 비교나 정량적인 성능평가는 수행되지 않았다.

이 연구는 변동성이 큰 천해 시계열 관측자료의 품질을 개선하기 위하여 국제 사회에서 통용되는 품질관리 기법에서 ‘가법적 이상값’과 ‘일시적 변동 이상값’과 같은 변동성 이상값을 검출하는 품질검사의 특성을 검토하고 그 성능을 정량적으로 평가하였다. 민 등 (2020)의 선행연구에서 사용한 2013년 층별 수온 관측자료로 검사 Case를 분류하고, IOOS와 NDBC, 그리고 OOI의 변동성 검사에 대한 정량적인 평가를 위해 혼돈행렬(confusion matrix, 혹은 오차행렬)과 성능평가지표를 이용하여 각 검사의 특성을 분석하였다. 분석 결과, 세 가지 국제 품질검사 중

NDBC와 OOI의 변동성 검사가 변동성 이상값을 가장 효과적으로 검출하였다. 그러나, 국제 품질검사 기법 모두 천해역과 같이 조석 변화에 따라 해수면 높이의 변동이 심한 해역에서 정상적으로 관측된 변동이 큰 시계열 자료를 이상으로 분류하는 것을 확인하였다. 특히 NDBC가 제공하는 변동성 검사에서 그 비율이 가장 높았다. 이 연구에서는 위와 같은 자동 품질검사와 함께 수동으로 수행되는 추가적인 검사가 반드시 수행되어야 하며, 천해역에 최적화된 새로운 품질검사의 개발이 필요함을 제시하였다.

2. 성능 평가 자료

이어도 해양과학기지(이하 기지)는 동중국해에 건설된 플랫폼 형태의 구조물로, 가거초 및 소청초 기지와 함께 관할 해역 내 단일 정점으로는 가장 많은 해양, 기상, 대기 환경 요소의 시계열 관측 자료를 생산하고 있는 정점 중 하나이다(Ha et al. 2019; Kim et al. 2019). 또한, 유관 기관의 관측 정점 중 유일하게 관할 해역의 경계에서 층별 수온 염분 관측을 수행하고 있다. 민 등 (2020)은 기지에서 생산하는 수온 시계열 자료의 품질을 개선하기 위해 OOI에서 제공하는 여섯 가지 품질관리(전 지구 및 지역범위 검사-튀는 값 검사-동일값 검사-기울기 검사-표류 검사)를 2013년 이어도 기지 층별 수온 시계열 자료에 시범 적용하고 그 성능을 분석한 바 있다. 이어도 층별 수온 시계열 자료는 연변동성이 매우 뚜렷하며, 표층과 저층에서 작은 변동을 보이는 반면 내부 조석파나 관성 주기파 등으로 인해 여름철 중층에서 매우 큰 변동을 보인다. 특히, 2013년에는 전력 케이블의 노후로 인하여 층별 시계열 자료에 다수의 변동성 이상값을 확인할 수 있다(민 등 2020). 이 연구는 민 등 (2020)이 사용한 동일한 수온 시계열 자료를 대상으로 하여 천해역에서 관측된 시계열 자료에서 빈번히 발생할 수 있는 이상값의 종류를 검토하고 이를 바탕으로 국제기관에서 제안한 품질관리의 성능 평가를 위한 Case를 분류하였다.

이어도 해양과학기지 수온 관측 자료

국립해양조사원은 이어도 기지에서 표층(3 m)과 중층(20.5 m), 저층(38 m)에 Aanderaa사의 CT3919를 설치하여 층별 수온과 염분 관측을 수행하고 있다. CT3919는 유도식 CT (Inductive Type Conductivity-Temperature) 센서로 크기가 작고 내구성이 강하여 태풍과 같은 악기상 관측에 유리하다. 또한, 케이블을 이용한 전력 공급 및 자료 전송으로 관측치를 실시간으로 확인할 수 있기 때문에 장비 파손 및 유실에도 결측을 최소화할 수 있다. 하지만 센서의 노후나 외력으로 인한 케이블 손상 등 다양한 원인으로 인해 이상값이 발생할 수 있기 때문에, 이에 대한 후처리가 필수적이다. 또한, 유도식 센서의 특성상 생물 오손(biofouling) 발생 시 전기 전도도에 영향을 주어 염분값의 하락을 유발하지만(이 등 2021). 수온의 경우 생물 오손의 영향이 미미하다. 센서는 정기적으로 점검 및 세척(연 4회)을 수행하고 있으므로, 이 연구에서 사용하는 수온 시계열 관측값에는 장기 표류와 같은 센서 오류는 없다고 가정하였다.

이어도 기지에서 10분 간격으로 관측한 2013년도 층별 수온 시계열 자료를 Fig. 1에 도시하였다. 그림에서 회색은 범위 검사(range check)를 통해 이상으로 분류된 자료로 6°C보다 낮고 33°C보다 높은 비정상적인 수온값을 보이는 관측 자료이다. 범위 검사는 수온 자료의 발생 빈도를 조사하여 전체의 96.5%가 분포하는 유효 범위를 기준으로 하였다(민 등 2020). 범위 검사 결과, 표층(3 m)의 경우는 다른 층과 비교하여 9월과 11월에 상대적으로 적은 개수의 이상값을 확인할 수 있었다. 중층(20.5 m)은 센서 복구 후 5월부터 관측이 수행되었으며 6월부터 10월까지 자료에서 다수의 이상값이 존재하였다. 이는 이 시기의 중층 수온 시계열 자료에서 큰 변동이 발생하기 때문이다. 구조물에 고정된 계류 라인에 센서를 설치하는 기지의 층별 수온 시계열 관측의 특성상 성층이 발달하는 시기에 조위 변화에 따라 센서가 수온약층의 상층과 하층에 번갈아 가며 위치하여 큰 수온 변화가 발생한다(민 등 2020).

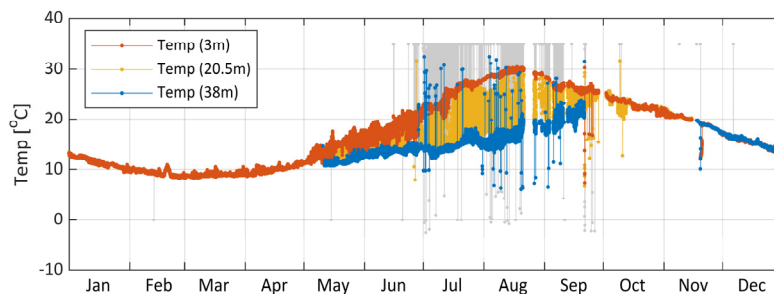


Fig. 1. Observed water temperatures at fixed depths (red line: 3 m, yellow: 20.5 m, blue: 38 m) from the Ieodo Ocean Research Station (I-ORS) in 2013. Gray dots and lines indicate data flagged based on a range check, i.e., colder than 6°C and warmer than 33°C

저층(38 m)에서의 수온 관측은 5월에 시작되었다. 관측 수행 후 4개월 후인 9월부터 약 2달간 결측이 발생하였으며, 센서 점검을 통해 11월부터 다시 복구되었다. 저층의 경우 관측 케이블 누수로 인하여 튀는 형태의 이상값이 다수 발생하였다(민 등 2020). 이러한 이상값은 범위 검사 후에도 상당 부분 남아있는 것을 확인할 수 있다(Fig. 1). 10월 중순 이후에는 세 개의 수층에서 거의 동일한 수온 값을 보이는데, 이는 전층이 혼합되었기 때문이다.

이상값과 성능 평가 CASE 분류

시계열 관측 자료에서 나타날 수 있는 이상값은 형태나 특성에 따라 Table 1과 같이 가법적 이상값(additive outliers), 일시적 변화 이상값(temporary change outliers), 혁신적 이상값(innovational outliers), 그리고 수준이동 이상값(level shift outliers)로 분류할 수 있다(유와 주 2019). 가법적 이상값은 시계열에서 지나치게 크거나 작게 관측된 단일 자료로, 주로 센서가 전기적 이상 신호를 수신하거나 자료가 전송 및 처리되는 과정에서 발생한다. 일시적 변화 이상값은 일시적으로 위나 아래로 이동하였으나 원래 상태로 빠

Table 1. The Types of outliers on the time-series observation data

Outlier Type	Content
AO (Additive Outliers)	Extremely large or small single value occurring in the data such as a spike
TCO (Temporary Change Outliers)	Time series move up or down temporarily and tend to get back to normal
IO (Innovational Outliers)	Data obtained ordinarily but having different statistical characteristics (e.g., trend, std, mean, and so on) owing to bad weather such as typhoon, storm surge, marine heat waves, etc.
LSO (Level Shift Outliers)	Entire time series moved up or down after a specific time

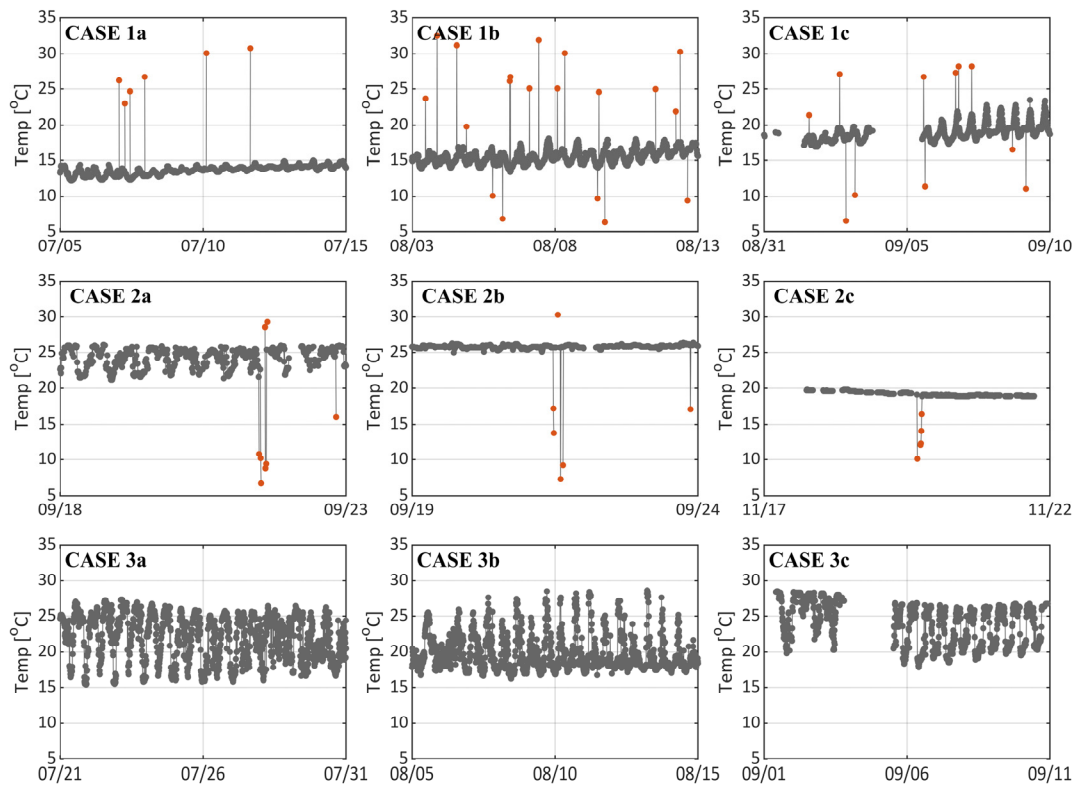


Fig. 2. Classifying cases from I-ORS water temperature observation in 2013 for evaluating international QC checks (gray is good data and red is outliers). The number of normal data and outliers for each case are summarized in Table 2

르게 회복하는 시계열로, 지속적인 이상 신호로 인한 자료의 오염이나 혹은 일시적으로 이물질이 센서에 부착하는 경우 등에서 나타난다. 혁신적 이상값의 경우는 태풍이나 악기상 발생에 따라 시계열의 양상이 기존과 다른 추세를 보이는 관측 자료를 말하며, 수준이동 이상값은 특정시간 이후 시계열 전체가 위나 아래로 이동한 경우로 기준면이나 장비의 위치를 변경하는 경우에 발생한다. 여기서 혁신적 이상값은 오류가 아닌 이상값이며, 수준이동 이상값은 오류이나 후보정을 통한 개선이 필요한 이상값을 의미한다. 가법적 이상값과 일시적 변화 이상값(이하 변동성 이상값)은 시계열 관측 자료에서 가장 빈번하게 발생하는 ‘오류에 해당하는 이상값’으로, 분석 자료인 2013년 이어도 기지 수온 시계열 자료에서도 가장 높은 빈도로 발생하였다(민 등 2020).

변동성 이상값이 발생한 기간의 수온 시계열 자료를 바탕으로 각 품질검사의 성능 평가를 위하여 검사 Case를 각각 가법적 이상값이 포함된 Case 1, 가법적 이상값과 일시적 변화 이상값이 포함된 Case 2, 이상값이 포함되지 않으며 변동이 큰 자료로 구성된 Case 3으로 분류하였다(Fig. 2). Case 1은 시계열의 변동이 작은 Case 1a과 변동이 큰 Case 1b, c 로 구성하였으며, 3156개 자료 중에서 이상값이 36개로 전체의 1.14%에 해당한다. Case 2는 1587개 자료 중에서 19개의 이상값이 발생하였으며, 이 중 가법적 이상값은 8개, 일시적 변화 이상값은 11개이다. Case 1에서 가법적 이상값은 분산되어 있는 반면 Case 2에서 가법적 이상값은 밀집되어 있고 인접 자료에서 결측이 다수 존재하는 특성을 보인다. 또한, Case 2에서 일시적 변화 이상값은 2개의 자료로 구성되어 있으나, Case 2f의 일시적 변화 이상값은 5개의 자료로 구성되어 있다. Case 3은 주로 중층 수온 시계열 자료로 10분간 변화량이 최대 7.3°C이며 3964개의 자료 모두 정상 자료로 사료된다. 관측 시기와 가법적 이상값, 일시적 변화 이상값, 그리고 정상 자료의 개수 등 Case에 따른 특성을 Table 2에 정리하였다.

3. 연구 방법

국제적으로 통용되는 IOOS, NDBC, OOI 품질관리를 분석하고 품질검사의 성능을 정량적으로 평가하기 위하여 혼돈행렬과 성능평가지표를 사용하였다. 민 등 (2020)은 국제 품질관리의 각 품질검사 기법들을 1) 센서의 한계 혹은 관측 자료의 통계 분석을 통하여 계산된 임계치를 정하여 검사자료가 이를 벗어나는지 평가하는 범위 검사, 2) 검사자료에 물리적으로 불가능한 변동이 발생하는 경우를 검출하는 변동성 검사류, 3) 관측 센서의 이상 유무를 측정하는 센서 검사류, 그리고 4) 배터리 전압 감소로 인한 오류 검사 혹은 관련 변수나 장비 간의 관계를 바탕으로 하는 검사 등의 기타 검사로 분류한 바 있다. 범위 검사류는 가장 기초적인 품질검사로 Case 분류 전 단계에서 자동으로 수행되었다. 센서 검사류에는 동일값 검사와 표류 검사 등이 있는데, 검출되는 자료의 발생 빈도가 낮거나 센서의 민감도를 함께 고려할 필요가 있으며 기관간 유사한 알고리즘을 사용하고 있다. 기타 검사에는 변동성 검사의 응용 검사로 내적 일치, 다중 변동, 신호 감소, 이웃 검사 등이 있으며 다른 관측 요소 혹은 장비를 복합적으로 판단해야 하여 분석에서 제외하였다(Fig. 3). 변동성 검사류는 튀는 값 검사, 변화율 검사, 시간연속성 검사 등이 있으며 주로 가법적 이상값, 일시적 변화 이상값과 같은 변동성 이상값을 검출한다. 변동성 검사류는 각 기관마다 다른 알고리즘을 사용하여 검사마다 다른 특성을 가진다. 따라서 각 품질관리의 성능과 특성을 대표할 수 있는 변동성 검사류에 집중하였 집중하였으며 상세한 설명은 3.1장에서 다룬다.

국제 품질검사의 성능 평가를 위하여 2장에서 자료의 정상과 이상을 구분한 검사 Case에 IOOS, NDBC, OOI의 변동성 검사류에 해당하는 품질검사를 적용하였다. 검사 Case에서 이상으로 분류한 자료와 국제 품질검사에서 이상값으로 분류한 자료를 비교하여 혼돈행렬을 작성하고

Table 2. Characteristics of water temperature cases from the I-ORS in 2013. See Fig 2 for these time series and outliers

Date and layer information for each case				Note
CASE 1	(a) 05 July–14 July Bottom Layer	(b) 03 Aug–12Aug Bottom Layer	(c) 31 Aug–08 Sep Bottom Layer	Only AO 3156 data (36 outliers)
CASE 2	(d) 18 Sep–22 Sep Middle Layer	(e) 19 Sep–28 Sep Surface Layer	(f) 17 Nov–21 Nov Bottom Layer	TCO, AO 1587 data (19 outliers)
CASE 3	(g) 21 July–30 July Middle Layer	(h) 05 Aug–14 Aug Middle Layer	(i) 01 Sep–10 Sep Middle Layer	No outliers but large fluctuation 3964 data (0 outliers)

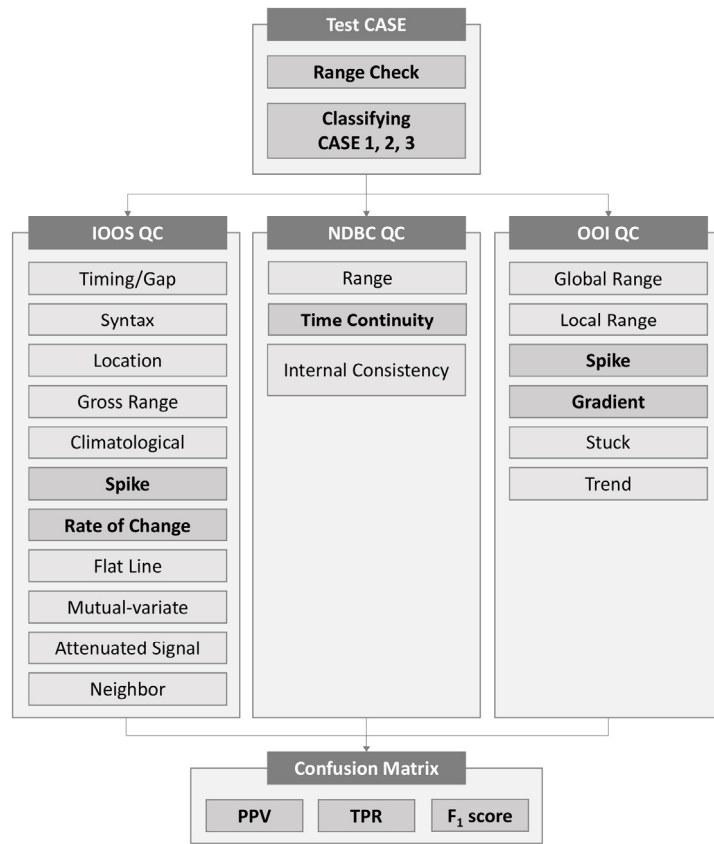


Fig. 3. The flow chart for this study procedure. The procedures that are normally used to detect Additive Outliers (AO) or Temporary Change Outliers (TCO) are boldfaced among the standard QCs from the IOOS, NDBC, and OOI

성능평가지표인 재현율과 정밀도, F_1 score를 산출하여 국제 품질검사의 특성과 성능을 정량적으로 분석하였다 (Fig. 3).

국제 품질검사 소개 및 입력값 설정

IOOS는 수온 시계열 자료의 품질관리를 위해 12단계의 품질검사를 수행하며(IOOS 2014), github를 통해 프로그램을 제공하고 있다(IOOS 2021). IOOS의 변동성 검사류에 해당하는 검사로 튀는 값(spike)과 변화율(rate of change) 검사(이하 IOOS 변동성 검사)가 있다. 먼저, IOOS의 튀는 값 검사는 검사자료 전후에 관측된 두 인접 자료의 평균과 검사자료의 차를 계산하여 이 값이 임계값을 초과할 경우 검사자료를 이상으로 분류한다(IOOS 2014). 튀는 값 검사에서 의심 임계값(suspect threshold)과 실패 임계값(fail threshold)을 입력하여 검사 강도를 결정할 수 있으나, 이 연구에서는 품질관리 매뉴얼(IOOS 2014)에서 제안한 3°C 를 임계값으로 사용하였다. IOOS의 변화율 검사는 검사자료의 변화율이 임계값을 초과할 경우 해당 자료를 이상값으로 분류한다(IOOS 2014). 통상 품질검사는 시계열

자료를 시간 순서에 따라 순차적으로 검사하여 이상 여부를 평가하지만, IOOS의 변화율 검사는 전 기간에 대한 시간 변화율을 한 번에 계산한다. 이 검사에서 입력값인 임계값은 2013년 이어도 수온 시계열 자료 전체 변화율의 평균인 $1.39^{\circ}\text{C}/10 \text{ min}$ 을 1시간 평균 변화율로 환산하여 사용하였다.

NDBC의 품질관리는 수온 관측에 대하여 총 3단계의 검사를 수행하며(NDBC 2009), 변동성 검사류에 해당하는 검사로 시간 연속성(time continuity) 검사가 있다. 이 검사는 알래스카만에서 멕시코만에 이르는 여러 정점의 통계값을 바탕으로 산출한 경험식(Eq. 1)을 기초로 구성되었으며 관련 품질검사 프로그램은 제공하고 있지 않다. 기상청과 국립해양조사원에서도 품질검사를 위하여 동일한 식을 적용한 프로그램을 개발하여 사용하고 있다. 이 연구에서는 NDBC가 제시한 Eq. 1을 바탕으로 시간 연속성 검사(이하 NDBC 변동성 검사)를 구현하여 성능 평가를 수행하였다. 이 방법은 검사자료와 이전 자료의 시간 변화량(T)을 기반으로 변화율 임계값(σ_T)를 산출하고(Eq. 1), 대상 검사자료와 이전 자료간 변화율을 계산하여 이

변화율 임계값을 초과할 경우 이상으로 분류한다(NDBC 2009). 이때, 시간 변화량이 3시간을 초과할 경우 T 는 3을 입력한다(NDBC 2009). Eq. 1은 12시간 이내에서 선형 관계를 보이는 조건에서 유도된 경험식이므로(NDBC 2009) 시간 변화량이 12시간을 초과할 경우 결측 구간으로 판단하였다. 결측이 발생하여 검사자료를 평가하기 위한 이전 자료가 없을 때 검사자료의 정상 여부를 판단하기 위해서 24시간동안 관측된 자료의 평균과 표준편차로 검사자료의 정상 여부를 평가하는 알고리즘을 추가하였다. 임계값 (σ_T) 계산에서 σ 는 관측 양상블의 표준 편차의 추정치로, 수온에 대하여 NDBC에서 제안한 8.6을 사용하였다.

$$\sigma_T = 0.58\sigma\sqrt{T} \quad (\text{Eq. 1})$$

OOI의 품질관리는 총 6단계의 품질검사로 구성되어 있으며(OOI 2013), 웹페이지에서 프로그램을 제공하고 있다(OOI 2021). OOI는 변동성 검사류에 튀는 값(spike)과 기울기(gradient) 검사(이하 OOI 변동성 검사)를 사용한다. OOI의 튀는 값 검사는 검사자료에 인접한 일정 기간(L) 자료의 평균과 검사자료 간의 차를 구하고, 이 차이가 동일 기간(L) 자료의 최대값과 최소값의 차(R)의 N배보다 클 경우 검사자료를 이상으로 분류한다(OOI 2012b). OOI의 기울기 검사는 관측 자료를 순차적으로 검토하여 검사자료의 변화율이 사전에 정의한 범위(DDATDX)를 초과할 경우 이상으로 분류한다(OOI 2012a). 만약 검사자료의 이전 자료가 없어 변화율을 구할 수 없는 경우, 사전에 정한 기준값(STARTDAT)과 수용 범위(TOLDAT)로 계산된 범위 내에 검사자료가 존재하는 경우에 해당 자료를 정상으로 분류한다(OOI 2012a). OOI 검사의 입력값은 민 등 (2020)의 기준을 참고하여 튀는 값 검사의 L은 2시간에 해당하는 13, N은 OOI에서 제시한 5를 사용하였고 DDATDX는

이어도 기지 수온 표준편차의 10배인 -5°C에서 5°C로, STARTDAT는 이어도 기지 전층 수온 시계열 자료의 평균인 5.5°C, TOLDAT는 10°C으로 설정하였다.

혼돈행렬과 성능평가 지표

혼돈행렬은 검출기에 해당하는 품질검사의 수행 성능을 평가하기 위한 표로, 기계학습 등 검출기의 성능 평가에 빈번히 활용된다. 혼돈행렬을 나타낸 Fig. 4에서 세로축은 실제 자료의 참(Positive, P)과 거짓(Negative, N)이며, 가로축은 품질검사가 검출한 참(Predicted Positive, PP)과 거짓(Predicted Negative, PN)이다. 품질검사는 이상값의 검출을 목적으로 하므로 여기서 실제 참은 이상 자료를, 실제 거짓은 정상 자료를 의미한다. 따라서, TP (True Positive)와 TN (True Negative)은 검출기가 실제 참과 거짓의 분류에 성공한 경우이며 FP (False Positive, Type I Error)와 FN (False Negative, Type II Error)은 분류에 실패한 경우이다.

혼돈행렬을 활용하여 검출기인 개별 품질검사의 성능 평가 지표를 산출할 수 있다. 성능평가 지표 중 PPV (Positive Predictive Value)는 검출기가 구분한 참 중에서 실제 참의 비율로 정밀도(precision)를 의미한다(Eq. 2). TPR (True Positive Rate)은 실제 참 중에서 검출된 참의 비율이며 재현율(recall)이라고도 한다(Eq. 3). 검출기의 정밀도를 높이고자 보수적으로 운용할 경우 PPV가 증가하지만 TPR은 감소할 수 있다. 한편, 실제 참을 최대한 검출하고자 할 경우 TPR이 증가하나 PPV가 감소할 수 있다. 따라서 PPV와 TPR은 상호 보완적으로 사용할 수 있으며 두 지표가 모두 높은 경우 좋은 검출기라 할 수 있다. 이 때, 품질검사와 같이 참과 거짓 분류 간 규모의 차이가 클 경우 PPV와 TPR의 조화 평균인 F1score (Eq. 4)로 그 정확도를 평가할 수 있다. 각각의 계산식은 다음과 같다.

Confusion Matrix		Actual Condition		
		P (Actual Positive)	N (Actual Negative)	
Predicted condition	PP (Predicted Positive)	TP (True Positive)	FP (False Positive, Type I Error)	PPV (Positive Predictive Value, Precision)
	PN (Predicted Negative)	FN (False Negative, Type II Error)	TN (True Negative)	
		TPR (True Positive Rate, Recall)	F1 score	

Fig. 4. Confusion matrix for evaluating each international check. PPV is the ratio of TP to PP; TPR is the ratio of TP to P. The F1 score is the harmony average of PPV and TPR. See text for detailed information

$$PPV = TP / (TP + FP) \tag{Eq. 2}$$

$$TPR = TP / (TP + FN) \tag{Eq. 3}$$

$$F_1\text{score} = 2 \times (PPV \times TPR) / (PPV + TPR) \tag{Eq. 4}$$

4. 결 과

IOOS 변동성 검사

IOOS에서 제시한 변동성 검사류의 성능평가 결과를 Fig. 5와 Table 3에 정리하였다. IOOS 변동성 검사의 경우 Case 1의 TPR이 1.00으로 재현율이 높았지만 Case 2에서 19개

이상값 중 11개만을 검출하여 0.58의 재현율을 보였다. PPV는 각각 0.38과 0.55로, 이상값보다 많거나 비슷한 개수의 정상 자료를 이상으로 분류하면서 낮은 정밀도를 보였다. 이는 IOOS의 튀는 값 검사에서 검사자료에 인접한 자료를 기준으로 검사자료의 정상 여부를 판단하기 때문이다. 즉, 정상 자료에 변동성 이상값이 인접한 경우 정상 자료도 임계값을 초과하는 것으로 판단한다. IOOS의 변화율 검사는 변화율의 크기만을 평가하므로 변동성 이상값으로 인해 크게 나타나는 시간 변화율을 모두 이상으로 평가한다. 다시 말해, 변동성 이상값뿐만 아니라 인접 자료 또한 이상으로 분류할 수 있다. 또한, 변동이 큰 정상 자료로 구성된 Case 3에서 전체의 2.43%에 해당하는 94개

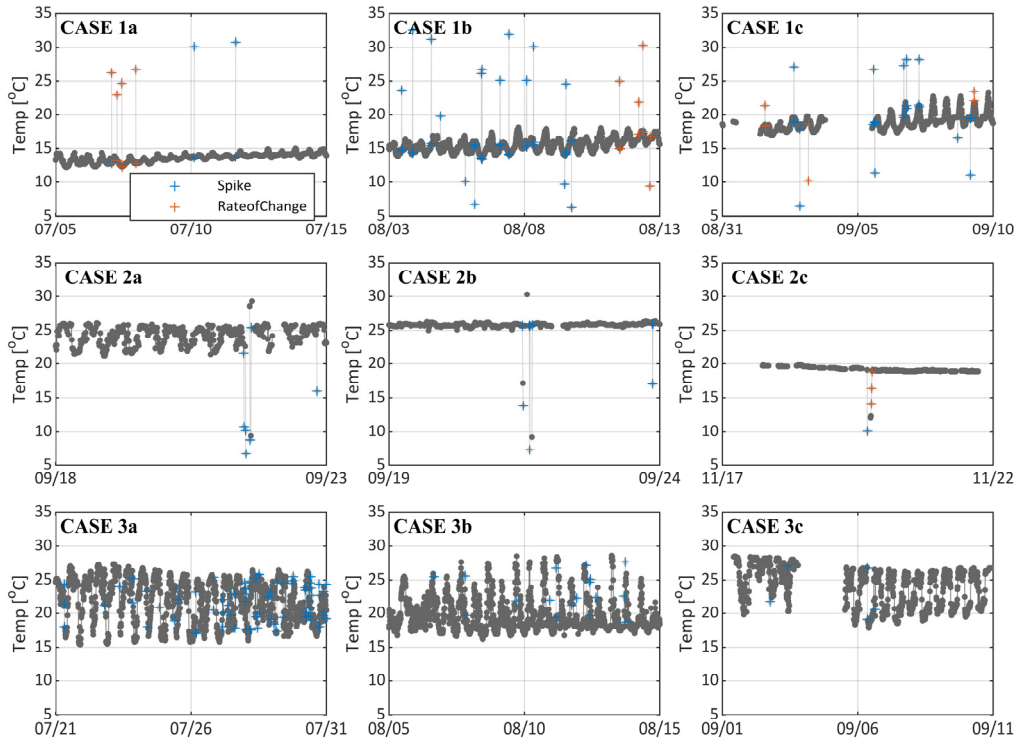


Fig. 5. Applying the spike, rate of change checks suggested by the IOOS to the cases from the I-ORS. A blue cross is flagged data from the spike check, and a red is based on the rate of change check

Table 3. Confusion matrix by applying the IOOS checks to the cases, which is corresponding to Fig. 5. For the detailed information on the confusion matrix, see Fig. 4

IOOS	CASE 1		CASE 2		CASE 3	
	P	N	P	N	P	N
PP	36	60	11	9	0	94
PN	0	3060	8	1559	0	3870
PPV	0.38		0.55		-	
TPR	1.00		0.58		-	
F ₁ score	0.55		0.56		-	

자료를 이상으로 분류하였다. IOOS 변동성 검사의 F_1 score는 Case 1과 2에서 0.55와 0.56으로 세 가지의 국제 품질검사 중 가장 낮은 정확도를 보였다.

NDBC 변동성 검사

NDBC 변동성 검사는 Case 1과 2에서 TPR이 각각 1.00과 0.89로, 시계열에 존재하는 가법적 이상값을 모두 검출하였다. 일시적 변화 이상값도 2개를 제외한 17개를 검출하여 높은 재현율을 보였다. Case 1에서 PPV는 1.00으로 검출기가 이상값만을 검출하였으나, Case 2에서 26개의 정상자료를 이상으로 분류하며 비교적 낮은 정밀도(0.40)를 보였다. 이는 급격한 변동을 가지는 정상 자료를 이상으로 분류하거나(Case 2d), 정상으로 분류된 이상값으로 인해

오히려 인접 정상 자료를 이상으로 분류하기 때문이다 (Case 2f). 이 검사의 F_1 score는 Case 1과 2에서 각각 1.00과 0.55로, 전체적으로 변동성 이상값을 검출하는 성능은 우수하지만 급격한 변동을 가지는 시계열에서 정상 자료를 이상 자료로 분류하는 특성을 보인다. 이와 같은 특성은 Case 3에서도 확인 할 수 있는데 정상 수온 시계열 자료에서 전체 자료의 21.71%에 달하는 707개의 자료를 이상값으로 분류하였다. 이 결과는 조석과 대기 기압장 변화에 의해 해양 변동이 큰 천해역에서 획득한 시계열 자료에 NDBC 변동성 검사를 적용할 때 유의해야함을 시사한다(Fig. 6, Table 4).

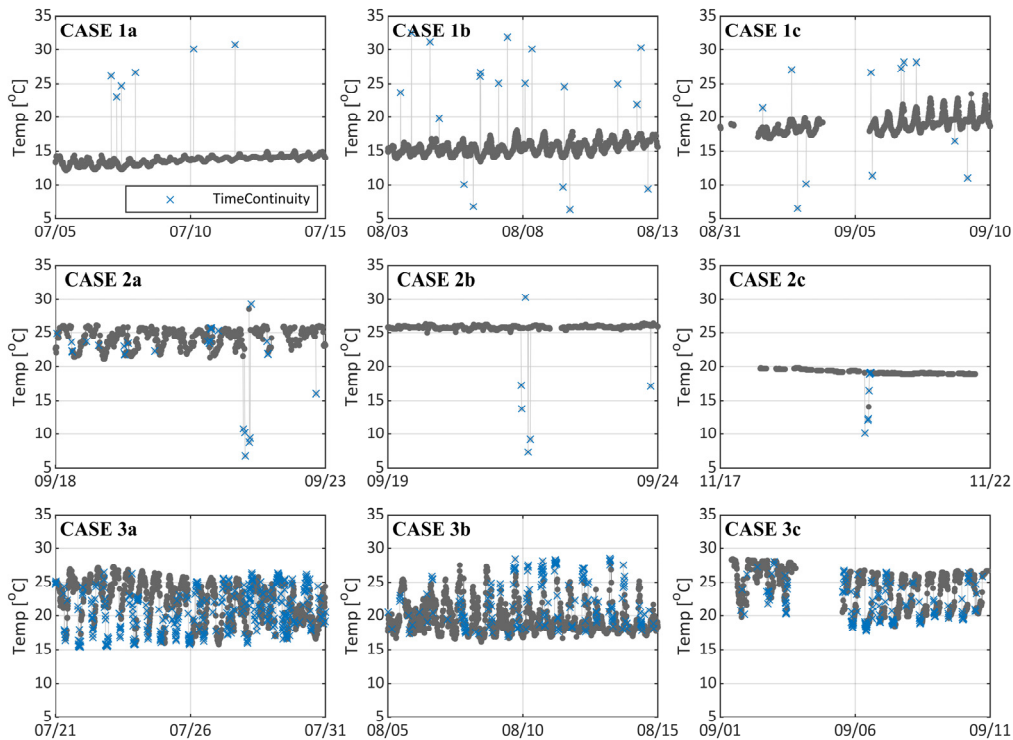


Fig. 6. Same as Fig 5 except for the NDBC. A blue x indicates flagged data based on the time continuity check suggested by the NDBC

Table 4. Same as Table 3 except for the NDBC checks. The table is corresponding to Fig. 6

NDBC	CASE 1		CASE 2		CASE 3	
	P	N	P	N	P	N
PP	36	0	17	26	0	707
PN	0	3120	2	1542	0	3257
PPV	1.00		0.40		-	
TPR	1.00		0.89		-	
F_1 score	1.00		0.55		-	

OOI 변동성 검사

OOI 변동성 검사는 Case 1과 2에서 PPV가 모두 1.00을 기록하였는데 이는 검출기의 정밀도가 매우 높으며 동시에 검출한 자료가 모두 실제 이상값임을 의미한다. TPR은 각각 0.94와 0.63으로, Case 1에서 실제 가법적 이상값 중 2개를 제외한 34개를 정확하게 검출하였으며 Case 2에서 실제 변동성 이상값 중 7개를 제외한 12개를 이상으로 분류하였다. Case 2d, e의 경우 변동성 이상값이 조밀하게 분포하는데, OOI의 튀는 값 검사는 검사자료 주변 관측 자료를 기준으로 검사를 수행하기 때문에 이처럼 변동성 이상값이 밀집된 경우 임계값의 범위가 증가하면서 이상값 검출 비율, 즉 재현율이 낮아지는 특징을 가진다. 또한, 변화율 검사에서 검사자료의 시간 변

화율 계산을 위한 이전 자료가 없는 경우 검사자료가 정상 범위에 있는지 판단하므로, Case 2c와 같이 다수 자료로 구성된 일시적 변화 이상값의 재현율이 낮아질 수 있다. 이 검사의 F₁score는 Case 1과 2에서 각 0.97과 0.73으로 세 검사 중 가장 높은 정확도를 보였다. 하지만 이 검사 또한, Case 3에서 전체 자료 중 1.1%에 해당하는 42개의 자료를 이상으로 분류하는 단점을 가진다(Fig. 7, Table 5).

종합 평가

위에서 수행한 성능 평가의 결과를 바탕으로 2013년 이어도 해양과학기지 수온 시계열 자료 중에서 변동이 큰 시계열이 포함된 중층과 다수의 변동성 이상값이 포함된

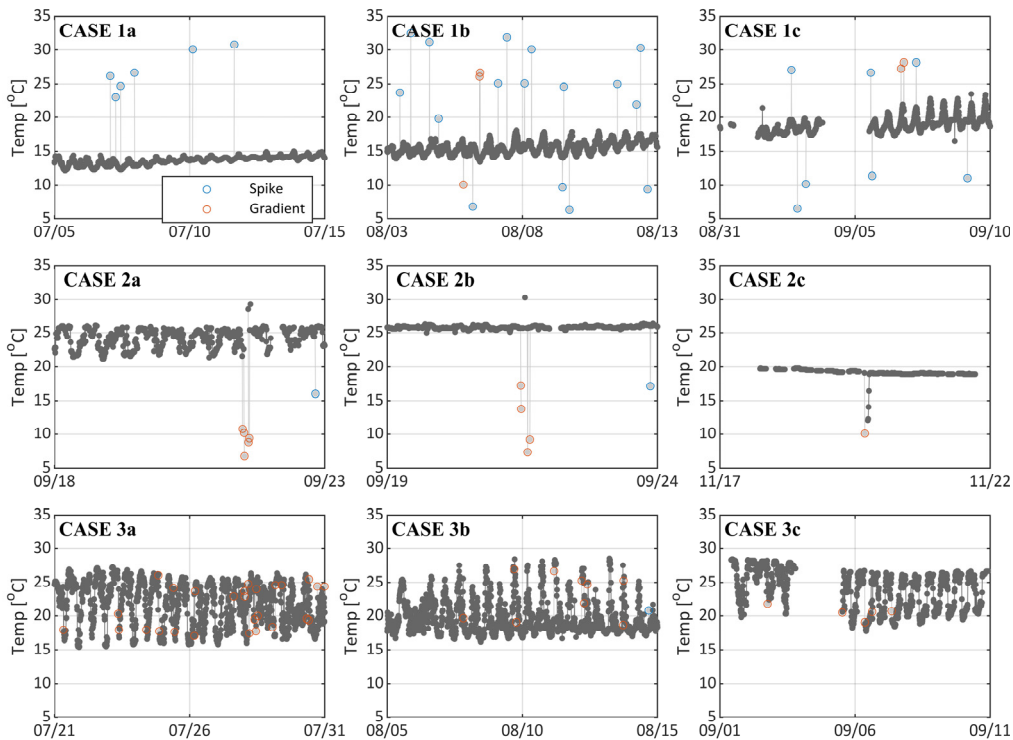


Fig. 7. Same as Fig 5. except for the OOI. A blue circle indicates a flagged data based on the spike check and a red one from the gradient check, which are suggested by the OOI to detect AO and TCO

Table 5. Same as Table 3 except for the OOI checks (see also Fig. 7)

OOI	CASE 1		CASE 2		CASE 3	
	P	N	P	N	P	N
PP	34	0	12	0	0	42
PN	2	3120	7	1568	0	3922
PPV	1.00		1.00		-	
TPR	0.94		0.63		-	
F ₁ score	0.97		0.73		-	

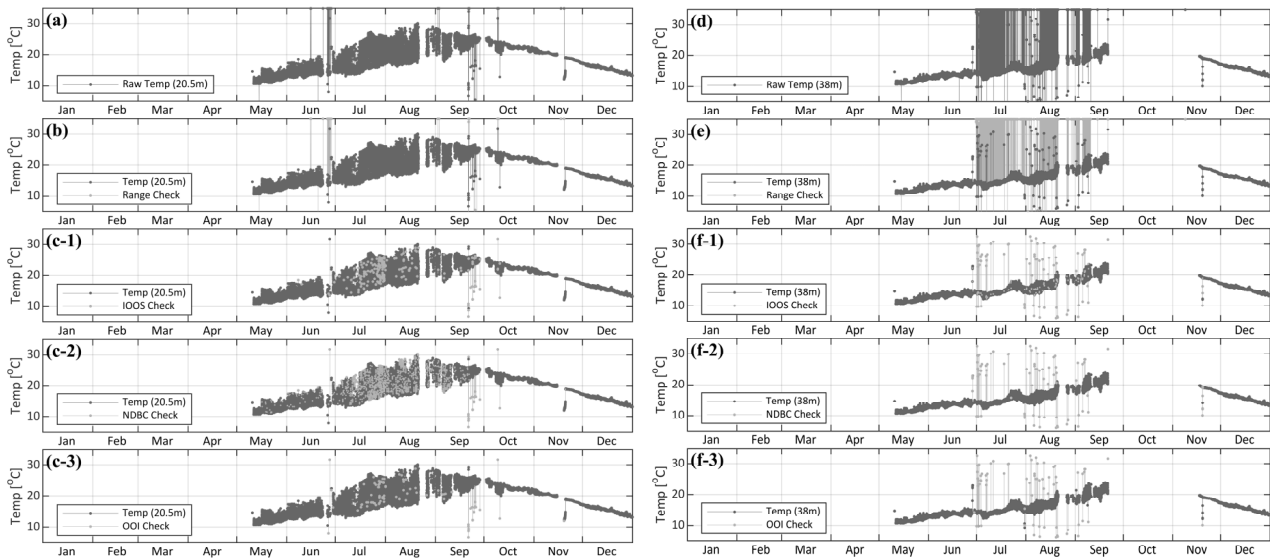


Fig. 8. Examples of applying the range check (b, d) and each of the variability checks (c-1, d-1 of IOOS check; c-2, f-2 of NDBC check; and c-3, f-3 of OOI check) to the I-ORS water temperature time-series observed in 2013. The left column indicates the one observed at a middle layer (20.5 m), and the right column for a bottom layer (38 m)

Table 6. Detected outliers and basic statistical information changes of the data by applying each international check to observed water temperature at the middle (20.5 m) and bottom (38 m) layers from the Ieodo Ocean Research Station in 2013

(a) Middle Layer (20.5 m)

data type	detected data	size	mean	SD	Median
raw data	-	27,937	18.50	4.74	17.8
data after range check	287	27,650	18.36	4.45	17.8
data after IOOS check	203 (203 of spike, 0 of rate of change)	27,447	18.34	4.45	17.7
data after NDBC check	1320	26,330	18.22	4.44	17.4
data after OOI check	84 (12 of spike, 72 of gradient)	27,566	18.35	4.45	17.8

(b) Bottom Layer (38 m)

data type	detected data	size	mean	SD	Median
raw data	-	21,782	16.26	5.26	14.9
data after range check	1272	20,510	15.18	2.64	14.7
data after IOOS check	194 (143 of spike, 51 of rate of change)	20,316	15.15	2.57	14.7
data after NDBC check	95	20,415	15.16	2.56	14.7
data after OOI check	73 (55 of spike, 18 of gradient)	20,437	15.16	2.57	14.7

저층 자료에 앞서 분석한 IOOS, NDBC, OOI의 변동성 검사를 적용하였다(Fig. 8). 먼저, 민 등 (2020)이 제안을 참고하여 범위검사를 수행한 후(Fig. 8a and d) 각 변동성 검사를 적용하고(Fig. 8c and d), 그 결과를 기초로 평균, 표준편차, 중간값 등의 통계정보를 비교하였다(Table 6). 품질검사 결과, 저층에서 IOOS, NDBC, OOI 변동성 검사가 수온 시계열 자료에서 각 194개, 95개, 73개의 자료를 변동성 이상값으로 검출하였다. IOOS 변동성 검사는 이상값에 인접한 다수의 정상 자료를 이상으로 분류하는 특성을 보인 반면 NDBC 변동성 검사는 변동성 이상값을 효과적으로 검출하였다. OOI 변동성 검사의 경우 대부분의 변동성 이상값을 검출하였으나, 변동성 이상값이 밀집되어 있는 경우 일부 이상값을 검출하지 못하는 것을 확인하였다. 하지만 변동이 큰 중층 수온 시계열에서 세 품질검사가 27,650개 관측 자료 중에서 각각 203(0.73%), 1320(4.77%), 84(0.30%)개의 관측 자료를 이상으로 분류하였다. 특히 NDBC 검사가 다른 검사에 비해 6-15배 더 많은 관측 자료를 변동성 이상값으로 분류하였으며 이 중 대부분은 정상적으로 관측된 자료로 사료된다(Table 6). 각 검사의 통계정보를 살펴보면, 먼저 범위검사를 적용할 경우 평균과 표준편차가 전반적으로 감소하는 추세를 보였으며 특히 저층의 평균과 표준편차가 각각 1.07°C , 2.62°C 감소하였는데(Table 6b), 이는 케이블 누수로 인해 저층에서 34.9°C 로 기록된 다수의 이상값이 제거되었기 때문이다. 각 변동성 검사 적용 시 통계정보 변화를 살펴보면 중층에서 가장 많은 자료를 검출한 NDBC 변동성 검사결과 수온 평균이 0.14°C 감소하였으며, 다른 변동성 검사의 경우 평균과 표준편차에서 $0.01\text{--}0.02^{\circ}\text{C}$ 의 미미한 감소를 보였다(Table 6).

품질검사는 보통 검사자료와 인접 자료의 관계를 이용하여 검사자료의 이상 여부를 판별하기 때문에 인접 자료의 분포 역시 중요하다. Fig. 9는 2013년 이어도 해양과학기지에서 관측한 수온 시계열 자료 중 국제 품질검사의

적용이 어려운 경우의 예시이다. Fig. 9a는 케이블 이상으로 인하여 간헐적 결측이 발생한 경우이다. 시간적으로 상당량의 결측이 발생하여 검사자료의 이상 여부를 평가하기 어렵다. 또한, 검사자료의 시간 변화율 계산 시 분모에 해당하는 시간 변화량이 커지면서 상대적으로 시간 변화율이 낮아지는 효과가 발생하여 이상값이 정상으로 평가될 수 있다. Fig. 9b는 현장 점검 시기 대기에 노출된 수온 센서의 관측값으로, 이처럼 많은 자료로 구성된 일시적 변화 이상값의 경우 품질검사에서 오히려 이상 자료를 정상 관측값으로 분류될 수 있어 사용에 유의해야 한다. 마지막으로, Fig. 9c는 시계열 자료의 변동이 큰 경우에 해당한다. 이와 같은 경우 품질검사 임계값의 범위가 커지기 때문에 시계열에서 발생한 변동성 이상값을 정상 자료로 분류하는 경우가 발생한다.

5. 결론

최근 해양관측자료의 종류와 양이 증가하고 자료의 최종 사용자의 범위가 연구자에서 기업, 일반인 등으로 확대되면서 관측 자료의 관리와 함께 품질관리의 중요성이 강조되고 있다. 국제 관측기관들은 표준화된 품질관리 절차를 구축하여 자동 품질검사 시스템을 운영하고 있으나 각 품질관리 기법의 특성에 대하여 정량적인 성능 평가가 부재하였다. 이 연구에서는 국제 품질관리의 변동성 이상값 검출 성능을 정량적으로 평가하기 위해 2013년 이어도 기지 수온 시계열 자료를 바탕으로 성능 검토를 위한 Case를 분류하고, 혼돈행렬과 성능 평가 지표를 이용하여 각 품질검사의 성능 및 특성을 분석하였다.

IOOS 변동성 검사는 가법적 이상값 검출에 대한 재현율이 높았으나 일시적 변화 이상값의 재현율이 낮았다. 특히 많은 정상 자료를 이상으로 분류하면서 낮은 정밀도를 보였다. 이 결과는 검사자료의 변화량 크기를 바탕으로 이상 여부를 평가하는 튀는 값 검사와 변화율 검사

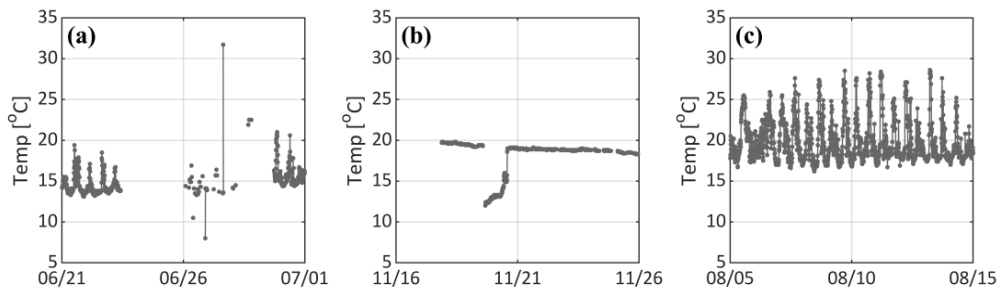


Fig. 9. Examples of observed temperature time-series characterized by extreme cases. (a) and (b) are the surface temperature observed during 21 June–1 July, 16 Nov–25 Nov 2013, respectively. The variability checks of the international QC procedures fail to detect these apparent outliers. (c) is the temperature at the middle level during 05 Aug–10 Aug 2013. The variability check tends not to detect outliers due to large fluctuations in the time series

의 특성상 변동성 이상값의 인접 자료도 함께 ‘변화가 큰 이상값’으로 판단하기 때문이다. IOOS 변동성 검사는 변동이 큰 정상 시계열 자료를 이상으로 분류하기 때문에 활용에 유의해야 한다. NDBC 변동성 검사는 시간 변화에 따른 임계값을 이용하여 관측 자료의 이상 여부를 평가하기 때문에 가법적 이상값의 검출 정확도가 매우 높았고, 일시적 변화 이상값의 재현율도 매우 높았다. 하지만 시계열이 단기간에 크게 변동하는 경우 타 기법에 비하여 5-6배 많은 정상 자료를 이상으로 분류하는 단점을 보였다. 따라서 검사 해역에서 본 검사의 적용 시 다년간의 관측 자료를 바탕으로 경험식의 상수 σ (양상블의 표준 편차의 추정치)를 조정할 필요가 있다. OOI 변동성 검사는 다른 검사와 비교하여 변동성 이상값의 검출 시 매우 높은 정밀도를 보였다. 튀는 값 검사의 경우 가법적 이상값의 검출에서 매우 높은 재현율을 보이지만 가법적 이상값이 밀집되거나 결측률이 높을 경우 낮은 재현율을 보였다. 전반적으로 국제품질검사의 경우 가법적 이상값의 검출 성능이 뛰어났고 일시적 변화 이상값의 검출의 경우 그 효과가 미비하였다. 세 국제 품질검사 중 변동성 이상값의 검출에 OOI 검사의 정확도가 가장 우수하다고 판단된다(F_1 score: 0.97, 0.73). 그러나, 천해와 같이 변동이 큰 해양환경에서 획득한 시계열 자료에서 정상 자료를 이상으로 분류하는 경우 또한 빈번하게 관찰되어 수동 검사와 같은 추가적인 검사가 반드시 필요하다. 또한, 황해 및 동중국해와 같은 천해에서 생산되며 시기별 변동이 크게 달라지는 시계열 자료의 품질검사를 위하여 가변 임계값을 사용하는 등 자료의 변동 특성을 고려한 품질검사 알고리즘의 개발이 요구된다.

이 논문은 수온 시계열 자료만을 대상으로 품질검사의 성능평가를 수행하였으나 향후 염분이나 파랑, 풍속, 기온, 습도 등 보다 다양한 관측 자료에 적용할 수 있다. 또한 혼돈행렬과 같은 성능평가지표를 활용한 정량적인 성능 평가 방법을 활용할 경우 각 정점에서 생산된 시계열 자료에 대한 품질관리를 보다 정량적이고 객관적으로 평가할 수 있다. 향후 이 성능평가 기법을 바탕으로 해양과학기지에서 사용하고 있는 품질관리 기법의 성능을 종합적으로 제시하고자 한다.

사 사

이 논문은 2021년 해양수산부의 재원으로 한국해양과학기술진흥원의 지원을 받아 수행된 연구입니다(관할해역 첨단 해양과학기지 구축 및 융합연구). 한국연구재단의 지역해 기후장 구축 하이브리드 기술 개발(NRF-2020R1F1A1070398)과 한국해양과학기술원(북서태평양 순환과 기후변동성이 한반도 주변 해역 변화와 물질 순환에 미치는

영향, PE99911)의 지원을 받았습니다. 연구에 사용한 자료를 획득하기 위하여 해양과학기지 내 체류 연구를 수행한 연구자분들과 과학기지 유지 보수를 도와주신 국립해양조사원과 관련 연구자 분들에게 깊은 감사를 드립니다.

References

- 국립해양조사원 (2019) 해양예보 의사결정 체계의 고도화를 위한 국가해양관측망개선연구 결과보고서. 국립해양조사원, 부산, 11-1192136-00507-01, 522 p
- 기상청 (2019) 기상관측데이터 품질검사 프로세스. 기상청 국가기후데이터센터, 서울, 11-1360000-001637-01, 62 p
- 민용침, 정진용, 장찬주, 이재익, 정종민, 민인기, 심재설, 김용선 (2020) 해양과학기지 시계열 관측 자료 품질관리 시스템 구축: 국제 관측자료 품질관리 방안 수온 관측 자료 시범적용과 문제점. *Ocean Polar Res* **42**(3):195-210. doi: 10.4217/OPR.2020.42.3.195
- 유재성, 주재걸 (2019) Unit Root Test를 기반으로 한 장기 시계열 데이터의 Non-Stationary 발생에 따른 구조 변화 검정 및 시각화 연구. *KIPS Trans Soft Data Eng* **8**(7):289-302. doi:10.3745/KTSDE.2019.8.7.289
- 이상호 (2015) R을 활용한 이상치 검출과 처리. 자유아카데미, 파주, 383 p
- 이재익, 이수찬, 정종민, 민용침, 정진용, 김용선 (2021) 해양과학기지에서 수행한 염분 표류 방지 실험 결과 및 염분 고품질 시계열 자료 생산을 위한 제언. *J Coast Prev* **8**(4): 323-333. doi:10.20481/ksmdp.2021.8.4.323
- 조호연, 이기섭, 안순모 (2016) 이상자료가 연안 환경자료의 통계 척도에 미치는 영향. *Ocean Polar Res* **38**(2):149-159. doi:10.4217/OPR.2016.38.2.149
- 채정훈, 박문수, 최영진 (2014) WISE 복합기상센서 관측 자료 품질관리시스템. *Kor Meteorol Soc* **24**(3):445-456. doi:10.14191/Atmos.2014.24.3.445
- Ha K-J, Nam S, Jeong J-Y, Moon I-J, Lee M, Yun J, Jang CJ, Kim YS, Byun D-S, Heo K-Y, Shim J-S (2019) Observations utilizing Korea ocean research stations and their applications for process studies. *B Am Meteorol Soc* **100**(10):2061-2075. doi:10.1175/BAMS-D-18-0305.1
- Hawkins DM (1980) Identification of outliers. Springer, Amsterdam, 188 p
- IOOS (2014) Manual for real-time quality control of in-situ temperature and salinity data Version 1.0. Integrated Ocean Observing System, Silver Spring, 55 p
- IOOS (2021) IOOS QC. <https://github.com/ioos/qartod> Accessed 1 Mar 2021
- Kim YS, Jang CJ, Noh J-H, Kim K-T, Kwon J-I, Min Y, Jeong J, Lee J, Min I-K, Shim J-S, Byun D-S, Jeong J-Y (2019) A Yellow Sea monitoring system and its scientific

- applications. *Front Mar Sci* **6**:601. doi:10.3389/fmars.2019.00601
- NDBC (2009) Handbook of automated data quality control checks and procedures. National Data Buoy Center, Mississippi, 78 p
- OOI (2012a) Data product specification for gradient test, version 1-01. Consortium for Ocean Leadership. https://oceanobservatories.org/wp-content/uploads/2015/09/1341-10004_Data_Product_SPEC_GLBLRNG_OOI.pdf Accessed 30 Sep 2019
- OOI (2012b) Data product specification for spike test, version 1-01. Consortium for Ocean Leadership. https://oceanobservatories.org/wp-content/uploads/2015/09/1341-10006_Data_Product_SPEC_SPKETST_OOI.pdf Accessed 30 Sep 2019
- OOI (2013) Protocols and procedures for OOI data products: QA, QC, calibration, physical samples, version 1-22. Consortium for Ocean Leadership, 1102-00300 https://oceanobservatories.org/wp-content/uploads/2015/09/1102-00300_Protocols_Procedures_Data_Products_QAQC_Cal_Physical_Samples_OOI Accessed 30 Sep 2019
- OOI (2021) Automated QC Algorithms. https://github.com/ooici/ion-functions/tree/master/ion_functions/qc Accessed 1 Mar 2021
- UNESCO (1993) Manual of quality control procedures for validation of oceanographic data. IOC, Paris, 25 p
- Weller RA, Baker DJ, Glackin MM, Roberts SJ, Schmitt RW, Twigg ES and Vimont DJ (2019) The challenge of sustaining ocean observations. *Front Mar Sci* **6**:105. doi:10.3389/fmars.2019.00105
- WMO (2010) Guide to the global observing system, WMO-NO. 488. World Meteorological Organization. <https://community.wmo.int/wmo-no-488-guide-global-observing-system> Accessed 1 Mar 2021
- Zhou Y, Qin R, Xu H, Sadiq S, Yu Y (2018) A data quality control method for seafloor observatories: the application of observed time series data in the East China Sea. *Sensors* **18**(8):2628. doi:10.3390/s18082628
- Min Y, Jeong J-Y, Jang CJ, Lee J, Jeong J, Min I-K, Shim J-S, Kim YS (2020) Quality control of observed temperature time series from the Korea ocean research stations: preliminary application of ocean observation initiative's approach and its limitation. *Ocean Polar Res* **42**(3):195–210. doi:10.4217/OPR.2020.42.3.195
- Yoo J, Choo J (2019) A study on the test and visualization of change in structures associated with the occurrence of non-stationary of long-term time series data based on unit root test. *KIPS Trans Soft Data Eng* **8**(7):289–302. doi:10.3745/KTSDE.2019.8.7.289
- Lee SH (2015) Outlier detection and treatment using R. Freeacademi, Paju, 383p
- Lee J, Lee S, Jeong J, Min Y, Jeong J-Y, Kim YS (2021) Salinity drift prevention experiments in the Korea ocean research stations and suggestions for high quality salinity observation. *J Coast Prev* **8**(4):323–333. doi:10.20481/kscdp.2021.8.4.323
- Cho H-Y, Lee K-S, Ahn S-M (2016) Impact of outliers on the statistical measures of the environmental monitoring. *Ocean Polar Res* **38**(2):149–159. doi:10.4217/OPR.2016.38.2.149
- Chae J-H, Park M-S, Choi Y-J (2014) The WISE quality control system for integrated meteorological sensor data. *Kor Meteorol Soc* **24**(3):445–456. doi:10.14191/Atmos.2014.24.3.445

Author's Information

Yongchim Min

Full-time Research Specialist, Korea Institute of Ocean Science & Technology

Hyunjung Jun

Post Master Scientist, Korea Institute of Ocean Science & Technology

Jin-Yong Jeong

Principal Research Scientist, Korea Institute of Ocean Science & Technology

Sung-Hwan Park

Senior Research Scientist, Korea Institute of Ocean Science & Technology

Jaek Lee

Full-time Research Specialist, Korea Institute of Ocean Science & Technology

국문 참고자료의 영문표기

English translation / Romanization of references originally written in Korean

- KHOA (2019) A report on the improvement of the national ocean observation network to advance the ocean forecast decision-making system, Korea Hydrographic and Oceanographic Agency, Busan, 11-1192136-00507-01, 522 p
- KMA (2019) Quality control of Meteorological Observations, KMA National Climate Data Center, 11-1360000-001637-01, Seoul, 62 p

Jongmin Jeong

Full-time Research Specialist, Korea Institute of Ocean
Science & Technology

Inki Min

Principal Research Specialist, Korea Institute of Ocean
Science & Technology

Yong Sun Kim

Senior Research Scientist, Korea Institute of Ocean
Science & Technology
Professor, Korea Maritime and Ocean University

Received Oct. 26, 2021

Revised Nov. 23, 2021

Accepted Dec. 7, 2021