

K-water연구원

2019-TE-RR-410-1984

*Interim Report*

# 물정보 표준 DB와 연계한 품질관리 알고리즘 적용기술 개발

Development of quality control algorithm for standard database of  
water information

2019. 12

---

**KOREA**

---

**WATER RESOURCES**

---

**CORPORATION**

---

# 물정보 표준 DB와 연계한 품질관리 알고리즘 적용기술 개발

Development of quality control algorithm for standard database of  
water information

2019. 12

# 제 출 문

한국수자원공사 사 장 귀하

이 보고서를 2019년도 K-water연구원 연구지원처에서 수행한 “물정보 표준 DB와 연계한 품질관리 알고리즘 적용기술 개발”의 최종 연구보고서로 제출합니다.

2019. 12. 31

연구책임자 : K-water연구원

연구수행자 : K-water연구원

전자통신3급 김 종 립

책 임 위 원 정 지 영

전 산 4 급 최 연 식

전자통신 4 급 김 대 욱

실 무 직 김 용 섭

# 연구개발결과 활용보고서

|                          |  |  |         |                      |
|--------------------------|--|--|---------|----------------------|
| [과제번호 : KIWE-WFRC-17-06] |  |  |         |                      |
| (1)연구개발과제명               | 물정보 표준 DB와 연계한 품질관리 알고리즘 적용기술 개발   |  |         |                      |
| (2)수행부서                  | 연구지원처  | (3)연구책임자   | 성명      | 김 종 립                |
|                          |  |  | 소속      | 연구지원처                |
|                          |  |  | 전화번호    | 042-870-7682         |
|                          |  |  | E-mail  | kjr8963@kwater.or.kr |
| (4)연 구 기 간               | 2019. 4. ~ 2019. 9.<br>(총 6개월)   | (5)총연구개발비<br>(천원)  | 총액      | 6,000                |
| (6)협 동 기 관               |  |  | 정부      |                      |
| (7)참 여 기 업               |  |  | 기업      |                      |
| (8)기 술 분 야               |  | (9)참여 연구원  | 5 명/ 1년 |                      |
| (10)연구개발목표               | <ul style="list-style-type: none"> <li>○ 물정보 분석 DB 연계 품질관리 소프트웨어 개발</li> <li>○ 알고리즘별 적정 매개변수 범위 선정 및 초기값 적용</li> <li>○ 범용 알고리즘 타분야 확대 방안 수립</li> </ul> |  |         |                      |
| 1. 연구 총괄                 | (11)주요연구내용   | <p><input type="checkbox"/> 물정보 분석 DB 연계 품질관리 소프트웨어 개발</p> <ul style="list-style-type: none"> <li>○ (범용성 검토) 기 개발 범용 알고리즘 적용성 검토 및 S/W 구축관련 전문가 회의 (4회)               <ul style="list-style-type: none"> <li>- 물정보 표준 DB와 연계한 설계 및 적용(↔물종합정보처)</li> <li>- GUI(C#)↔알고리즘(R) 연동 기능 구현 및 시각화 표출</li> <li>- S/W개발 시연을 통한 의견수렴 및 중간보고(1회)</li> </ul> </li> </ul> <p><input type="checkbox"/> 알고리즘별 적정 매개변수 범위 선정 및 초기값 적용</p> <ul style="list-style-type: none"> <li>○ (알고리즘 개선) 알고리즘 최적화 피드백을 통한 매개변수 최적화 분석               <ul style="list-style-type: none"> <li>- 대상 유역의 확대 및 관련 매개변수의 적용범위 제시</li> <li>- 서버기반의 자동화 구현을 위한 알고리즘 개선</li> </ul> </li> </ul> <p><input type="checkbox"/> 범용 알고리즘 타분야 확대 방안 수립</p> <ul style="list-style-type: none"> <li>○ 데이터 관리 분야/항목/기술에 대한 구체적 확대 방안 구축</li> </ul> |         |                      |

|                         |   |   |            |          |   |           |            |                   |    |
|-------------------------|---|---|------------|----------|---|-----------|------------|-------------------|----|
| 2. 연구성과 활용 유형 및 주요 연구성과 | (12) 연구결과의 활용   |   |            |          |   |           |            |                   |    |
|                         | 구분  | 기술실용화                                     |            |          |   | 교육 및 지도   | 정책활용       | 설계기준, 시방서, 지침에 활용 | 기타 |
| 실용화 완료                  |   | 실용화 추진중                                   | 실용화 보류(중단) | 선행 연구    |   |           |            |                   |    |
|                         |   |   |            | ○        |   |           |            |                   |    |
| 3. 경제적 효과               | (13) 주요 연구성과 및 기술 개발단계  |   |            |          |   |           |            |                   |    |
|                         | ◦기술명 : 물정보 표준 DB와 연계한 품질관리 알고리즘 적용기술 개발<br>◦기술특성 : 물종합정보처에서 개발중인 물정보 표준 DB에 연계 가능한 품질관리 알고리즘의 자체 개발 |   |            |          | ◦기술의 완성도<br>기초, 탐색연구단계 ( ), 응용연구단계(○), 개발연구단계 ( )<br>기업화 준비단계 ( ), 상품화 완료단계 ( )<br>◦기술 수명주기<br>기술개념 정립기 (○), 기술 실험기 ( )<br>기술적용 시작기 ( ), 기술적용 성장기 ( )<br>기술적용 성숙기 ( ), 기술적용 쇠퇴기 ( ) |           |            |                   |    |
| 3. 경제적 효과               | (14) 현장적용실적   |   |            |          |   |           |            |                   |    |
|                         | ①<br>공사현장 정보  | 공사명                                       | 공사기간       | 물량 (총물량) | 발주기관  | 시공기관      | R&D 참여자 역할 |                   |    |
|                         |   |   |            |          |   |           |            |                   |    |
|                         |   |   |            |          |   |           |            |                   |    |
|                         | ②<br>공사 현장별 적용효과  | 가. 현장명                                    |            |          |   |           |            |                   |    |
|                         |   | 구분  | 기존 (A)     | 개선 (B)   | 단위효과 (C=A-B)  | 현장총물량 (D) | 향상도        | 적용현장 총효과 (C×D)    |    |
|                         |   | 공사비 (백만원/m <sup>2</sup> ) <sup>1</sup>    |            |          |   |           |            |                   |    |
|                         |   | 유지관리비 (백만원/m <sup>2</sup> ) <sup>1</sup>  |            |          |   |           |            |                   |    |
|                         |   | 기타비용절감 (백만원/m <sup>2</sup> ) <sup>1</sup> |            |          |   |           |            |                   |    |
|                         |   | 사고율 (명/UC)                                |            |          |   |           |            |                   |    |
| 고용 (명/UC)               |   |   |            |          |   |           |            |                   |    |
| 공기 (일/UC)               |   |   |            |          |   |           |            |                   |    |
| 내구성 (年)                 |   |   |            |          |   |           |            |                   |    |

※ 개별 공사에 가장 적합한 척도 사용가능 ex) m<sup>2</sup>, Km

※ 적용 현장이 여러 곳일 경우 위 서식 복사해서 사용

|  |                                |             |             |                |               |              |               |      |
|--|--------------------------------|-------------|-------------|----------------|---------------|--------------|---------------|------|
| 4.<br>기술적<br>효과                        | (15) 기술개발목표 달성도                |             |             |                |               |              |               |      |
|  | 구분                             | 지표명<br>(단위) | 개발전<br>국내수준 | 당개발사업<br>최종목표치 | 개발결과<br>실제달성치 | 현재국내<br>기술수준 | 세계 최고<br>기술수준 |      |
|  | 기술<br>성능<br>지표                 |             |             |                |               |              |               |      |
|  |                                |             |             |                |               |              |               |      |
|  |                                |             |             |                |               |              |               |      |
|  | (16) 산업재산권 (발명특허·실용신안·의장·s/w ) |             |             |                |               |              |               |      |
|  | 구분                             | 명칭          | 출원국         | 출원 및<br>등록자    | 출원            |              | 등록            |      |
|  |                                |             |             |                | 출원일           | 출원번호         | 등록일           | 등록번호 |
|  |                                |             |             |                |               |              |               |      |
|  | (17) 건설신기술 지정                  |             |             |                |               |              |               |      |
|  | 명 칭                            |             |             | 출 원 일          | 고 시 일         | 보호기간         | 지정번호          |      |
|  |                                |             |             |                |               |              |               |      |
| (18) 기타 등록 및 지정현황(상품, 서비스 등록, KT 마크 등) |                                |             |             |                |               |              |               |      |
| 명 칭                                    |                                |             | 출 원 일       | 고 시 일          | 보호기간          | 기 타          |               |      |
|  |                                |             |             |                |               |              |               |      |
| 5.<br>학술적<br>효과                        | (19) 국내외 전문학술지 게재              |             |             |                |               |              |               |      |
|  | 구분                             | 논문제목        |             | 학술지명           | 국명            | 통권, 호        | 년,월           |      |
|  | 일반<br>학술지                      |             |             |                |               |              |               |      |
|  |                                |             |             |                |               |              |               |      |
|  | (20) 학술회의 발표                   |             |             |                |               |              |               |      |
|  | 논문제목                           |             |             |                | 학술회의 명칭       | 국 명          | 연 도 · 호       |      |
|  |                                |             |             | 한국통신학회         | 한국            | 2019         |               |      |
|  |                                |             |             | 외국             | 한국            | 2019         |               |      |

|                 |   |                       |        |         |         |      |      |
|-----------------|---|-----------------------|--------|---------|---------|------|------|
| 6.<br>공공적<br>효과 | (21) 법·제도 (설계기준, 지방서, 지침) 등에 활용   |                       |        |         |         |      |      |
|                 | 구분  | 명칭 및 조항 <sup>1)</sup> | 시행일시   | 기 존 내 용 | 신 규 내 용 |      |      |
|                 | 법   |                       |        |         |         |      |      |
|                 | 규정/지침   |                       |        |         |         |      |      |
|                 | 기준  |                       |        |         |         |      |      |
|                 | 지방서   |                       |        |         |         |      |      |
|                 | (22) 정책수립에 활용   |                       |        |         |         |      |      |
|                 | 시책명   | 시책추진 실적               | 시책추진일시 | 담당부처·팀  | 기대효과    |      |      |
|                 |   |                       |        |         |         |      |      |
|                 |   |                       |        |         |         |      |      |
|                 |   |                       |        |         |         |      |      |
|                 | (23) 교육 및 지도에 활용  |                       |        |         |         |      |      |
|                 | 구분  | 교육명                   | 교육내용   | 교육대상    | 교육시간    | 교육인원 | 교육회수 |
|                 | 교육훈련  |                       |        |         |         |      |      |
|                 |   |                       |        |         |         |      |      |
|                 | 기술상담  |                       |        |         |         |      |      |
|                 | 현장기술지도  |                       |        |         |         |      |      |
|                 | (24) 사회적 성과   |                       |        |         |         |      |      |
|                 | <ul style="list-style-type: none"> <li>○ (기술활용) 표준 DB 활용을 위한 사내 시스템 적용으로 전사적 품질관리에 활용</li> <li>○ (성과확산) 수자원,수도분야 데이터 품질관리기술 확보로 타분야(지하수, 발전, 수질, 설비 등) 확대적용(안) 및 자동화 알고리즘 제시</li> <li>○ (비용절감) 표준 DB의 품질관리에 대한 자체 기술력 활용</li> </ul> |                       |        |         |         |      |      |

# 요 약 문

## 1. 연구제목

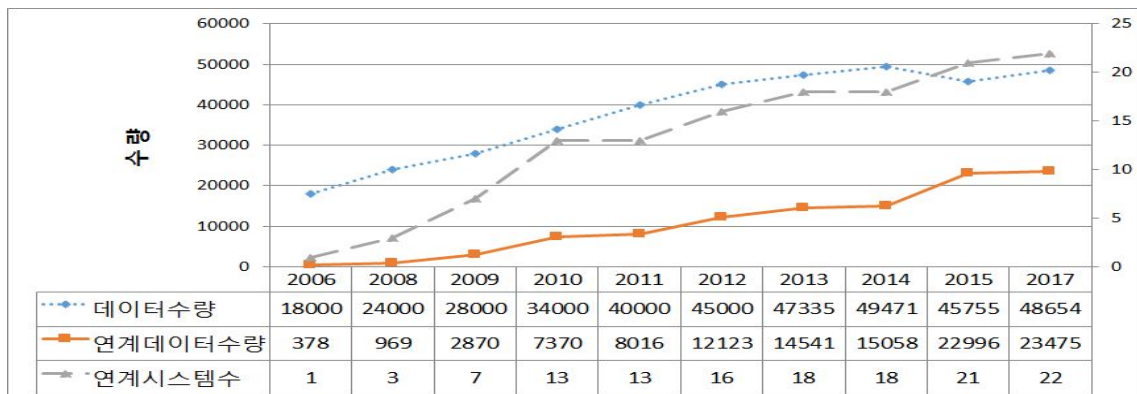
물정보 표준 DB와 연계한 품질관리 알고리즘 적용기술 개발

## 2. 연구기간

2019. 4. 1 ~ 2019. 09. 30

## 3. 연구목적

- 통합물정보 관리를 위한 데이터 표준도입과 고품질 데이터 확보를 통한 지능정보기술 적용기반확장으로 4차 산업혁명을 선도하기 위한 K-water 미션설정
- 수자원(수위, 우량) 및 수도(압력, 유량) 시계열 데이터의 품질관리를 위한 검·보정 알고리즘 확보
- 공간에 영향을 받지 않는 계측 데이터의 범용적인 품질관리 절차 도출 및 자동화
- 적용 대상별 적합 오·결측 기법과 매개변수 최적화 방안 도출 및 사내 표준화된 DB내 시스템 품질관리 자동화 지원



### 데이터 연계 급증

### 이상데이터 발생 증가

### 품질관리 필요성 증대

**2006년 ~ 현재까지 시스템간 수도데이터 연계 급증**

\* 2006년 (17개) → 2016년 (22개)

**수도데이터 양이 급증하면서 이상데이터 수지도 증가**

\* 15.7 ~ 12(6개월) 평균 0.5% 결측 발생

**정부 3.0 공공 데이터 공개 정책으로 고품질 데이터 확보 중요성 대두**

\* 실시간수도정보시간데이터개발

|               |   |
|---------------|---|
| 기술적<br>측 면    | ○ 계측 데이터별 품질관리 적합 방안(알고리즘, 매개변수 최적화) 및 표준화 된 DB에 연계 가능한 시스템의 자체 개발을 통한 기술력 확보 필요    |
| 경제·산업적<br>측 면 | ○ 4차 산업혁명의 기반이 되는 표준 데이터에 대한 품질관리를 통해 신뢰성 높은 데이터 제공으로 新부가가치 창출기반 확보 및 대국민 신뢰도 제고 필요 |

□ 국내외 관련(유사) 기술의 개발 현황

- 국내 데이터 오결측에 대한 검출 알고리즘 개발 및 정제알고리즘 연구개발
  - 오결측 관련 DB연계 프로그램 개발 및 사례분석, 정제방법론 개발
  - 상관관계 분석을 위한 데이터 추출 TB 구축 및 SW모델 개발
- \* 분석기법과 알고리즘 개발 및 적용이 본 연구와는 차이점이 있음

#### 4. 연구내용

| 연구목표                           | 연구 내용  | 세부 연구내용  |
|--------------------------------|--|--|
| ○ 알고리즘 적용성 검증 및 파일럿 S/W 개발     | - 알고리즘 적용성 검증 및 유형별 적용기법 분석<br>- Pilot S/W 개발          | - 데이터 생애주기별(FA-OA간) 적용성에 대한 평가<br>- 오결측 형태(부분/연속) CASE별 분석<br>- 실제 표준 DB와 연계하여 Site, Result, Value Table을 참조 가능한 S/W 개발<br>- 시계열 데이터의 품질관리 후 품질관리수준을 조정하여 표준DB에 반영 |
| ○ 표준 DB와 연계한 알고리즘 개발 및 최적기법 제시 | - 표준 DB 연계 자동화를 위한 알고리즘 개발<br>- 시범적용 수계에 대한 최적 보정기법 제시 | - 일정 주기별 품질관리 자동화 적용기술 개발<br>- 품질관리 알고리즘 시범적용 대상 수계선정 및 최적 보정기법  |
| ○ 타분야 확대적용 방안 마련               | - 타분야 확대적용 방안마련  | - 하수, 발전, 수질, 설비 분야의 품질관리 알고리즘 적용방안 마련   |

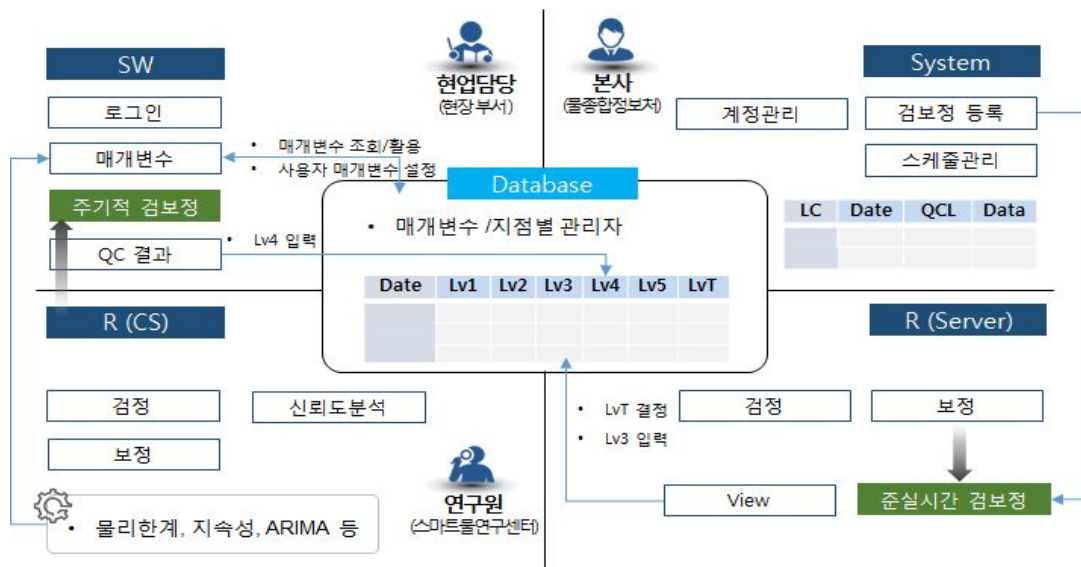
## 5. 연구결과

본 연구의 주요 결과는 표준 DB 연계 자동화를 위한 알고리즘 개발 및 구현과 품질관리 알고리즘 타분야 확대방안을 수립하여 알고리즘 구축에 대한 표준화를 마련하고 물종합 정보처에서 개발중인 표준DB KDM에 실용화를 통해서 검증하는데 큰 의의가 있다.

### 가. 표준 DB 연계 자동화를 위한 알고리즘 개발 및 구현

□ 품질관리 적용 대상 및 주기에 따른 알고리즘 개발 및 DB 연계 활용 체계 구축

- (로컬) 각 사이트 담당자별 비주기적 수동 품질관리 수행
  - (서버) 주기적 오결측 탐지 및 보정 스케줄을 적용한 자동화 품질관리 수행
- \* 검·보정 알고리즘은 범용적으로 개발된 R 형태로 로컬 및 서버에 공통 반영



□ 확장수계 대상 검·보정 알고리즘 적용에 따른 신뢰도 분석 및 적정성 검토 수행

- 공간적 확장을 통한 알고리즘 신뢰도 분석
- 수도,수자원 알고리즘에 사용된 원시데이터의 신뢰성을 위한 샘플링 분석 시행 (FA ↔ OA)

#### 1. 수위원수위, 우량 분야 데이터 분석

□ [본시DB 10분 데이터, TECH DB 10분 데이터]

| 구분 | 장소 | 구분 | 데이터 유형 | 구분 | 데이터 유형 | 데이터 유형 |
|----|----|----|--------|----|--------|--------|
| 1  | 원수 | 원수 | 원수     | 원수 | 원수     | 원수     |
| 2  | 원수 | 원수 | 원수     | 원수 | 원수     | 원수     |
| 3  | 원수 | 원수 | 원수     | 원수 | 원수     | 원수     |
| 4  | 원수 | 원수 | 원수     | 원수 | 원수     | 원수     |
| 5  | 원수 | 원수 | 원수     | 원수 | 원수     | 원수     |
| 6  | 원수 | 원수 | 원수     | 원수 | 원수     | 원수     |
| 7  | 원수 | 원수 | 원수     | 원수 | 원수     | 원수     |
| 8  | 원수 | 원수 | 원수     | 원수 | 원수     | 원수     |
| 9  | 원수 | 원수 | 원수     | 원수 | 원수     | 원수     |
| 10 | 원수 | 원수 | 원수     | 원수 | 원수     | 원수     |
| 11 | 원수 | 원수 | 원수     | 원수 | 원수     | 원수     |
| 12 | 원수 | 원수 | 원수     | 원수 | 원수     | 원수     |
| 13 | 원수 | 원수 | 원수     | 원수 | 원수     | 원수     |
| 14 | 원수 | 원수 | 원수     | 원수 | 원수     | 원수     |
| 15 | 원수 | 원수 | 원수     | 원수 | 원수     | 원수     |
| 16 | 원수 | 원수 | 원수     | 원수 | 원수     | 원수     |
| 17 | 원수 | 원수 | 원수     | 원수 | 원수     | 원수     |
| 18 | 원수 | 원수 | 원수     | 원수 | 원수     | 원수     |
| 19 | 원수 | 원수 | 원수     | 원수 | 원수     | 원수     |
| 20 | 원수 | 원수 | 원수     | 원수 | 원수     | 원수     |
| 21 | 원수 | 원수 | 원수     | 원수 | 원수     | 원수     |
| 22 | 원수 | 원수 | 원수     | 원수 | 원수     | 원수     |
| 23 | 원수 | 원수 | 원수     | 원수 | 원수     | 원수     |
| 24 | 원수 | 원수 | 원수     | 원수 | 원수     | 원수     |
| 25 | 원수 | 원수 | 원수     | 원수 | 원수     | 원수     |
| 26 | 원수 | 원수 | 원수     | 원수 | 원수     | 원수     |
| 27 | 원수 | 원수 | 원수     | 원수 | 원수     | 원수     |
| 28 | 원수 | 원수 | 원수     | 원수 | 원수     | 원수     |
| 29 | 원수 | 원수 | 원수     | 원수 | 원수     | 원수     |
| 30 | 원수 | 원수 | 원수     | 원수 | 원수     | 원수     |

#### 2. 수위(수량, 일량) 분야 데이터 분석(확장수계)에 활용하기 위한

□ [본시DB 10분 데이터, 본부DB 10분 데이터, TECH DB 10분 데이터]

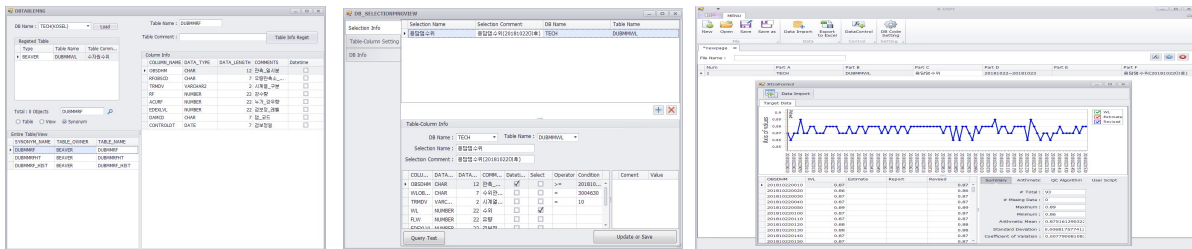
| 구분 | 장소 | 구분 | 데이터 유형 | 구분 | 데이터 유형 | 데이터 유형 |
|----|----|----|--------|----|--------|--------|
| 1  | 원수 | 원수 | 원수     | 원수 | 원수     | 원수     |
| 2  | 원수 | 원수 | 원수     | 원수 | 원수     | 원수     |
| 3  | 원수 | 원수 | 원수     | 원수 | 원수     | 원수     |
| 4  | 원수 | 원수 | 원수     | 원수 | 원수     | 원수     |
| 5  | 원수 | 원수 | 원수     | 원수 | 원수     | 원수     |
| 6  | 원수 | 원수 | 원수     | 원수 | 원수     | 원수     |
| 7  | 원수 | 원수 | 원수     | 원수 | 원수     | 원수     |
| 8  | 원수 | 원수 | 원수     | 원수 | 원수     | 원수     |
| 9  | 원수 | 원수 | 원수     | 원수 | 원수     | 원수     |
| 10 | 원수 | 원수 | 원수     | 원수 | 원수     | 원수     |
| 11 | 원수 | 원수 | 원수     | 원수 | 원수     | 원수     |
| 12 | 원수 | 원수 | 원수     | 원수 | 원수     | 원수     |
| 13 | 원수 | 원수 | 원수     | 원수 | 원수     | 원수     |
| 14 | 원수 | 원수 | 원수     | 원수 | 원수     | 원수     |
| 15 | 원수 | 원수 | 원수     | 원수 | 원수     | 원수     |
| 16 | 원수 | 원수 | 원수     | 원수 | 원수     | 원수     |
| 17 | 원수 | 원수 | 원수     | 원수 | 원수     | 원수     |
| 18 | 원수 | 원수 | 원수     | 원수 | 원수     | 원수     |
| 19 | 원수 | 원수 | 원수     | 원수 | 원수     | 원수     |
| 20 | 원수 | 원수 | 원수     | 원수 | 원수     | 원수     |
| 21 | 원수 | 원수 | 원수     | 원수 | 원수     | 원수     |
| 22 | 원수 | 원수 | 원수     | 원수 | 원수     | 원수     |
| 23 | 원수 | 원수 | 원수     | 원수 | 원수     | 원수     |
| 24 | 원수 | 원수 | 원수     | 원수 | 원수     | 원수     |
| 25 | 원수 | 원수 | 원수     | 원수 | 원수     | 원수     |
| 26 | 원수 | 원수 | 원수     | 원수 | 원수     | 원수     |
| 27 | 원수 | 원수 | 원수     | 원수 | 원수     | 원수     |
| 28 | 원수 | 원수 | 원수     | 원수 | 원수     | 원수     |
| 29 | 원수 | 원수 | 원수     | 원수 | 원수     | 원수     |
| 30 | 원수 | 원수 | 원수     | 원수 | 원수     | 원수     |

| 구분 | 장소 | 구분 | 데이터 유형 | 구분 | 데이터 유형 | 데이터 유형 |
|----|----|----|--------|----|--------|--------|
| 1  | 원수 | 원수 | 원수     | 원수 | 원수     | 원수     |
| 2  | 원수 | 원수 | 원수     | 원수 | 원수     | 원수     |
| 3  | 원수 | 원수 | 원수     | 원수 | 원수     | 원수     |
| 4  | 원수 | 원수 | 원수     | 원수 | 원수     | 원수     |
| 5  | 원수 | 원수 | 원수     | 원수 | 원수     | 원수     |
| 6  | 원수 | 원수 | 원수     | 원수 | 원수     | 원수     |
| 7  | 원수 | 원수 | 원수     | 원수 | 원수     | 원수     |
| 8  | 원수 | 원수 | 원수     | 원수 | 원수     | 원수     |
| 9  | 원수 | 원수 | 원수     | 원수 | 원수     | 원수     |
| 10 | 원수 | 원수 | 원수     | 원수 | 원수     | 원수     |
| 11 | 원수 | 원수 | 원수     | 원수 | 원수     | 원수     |
| 12 | 원수 | 원수 | 원수     | 원수 | 원수     | 원수     |
| 13 | 원수 | 원수 | 원수     | 원수 | 원수     | 원수     |
| 14 | 원수 | 원수 | 원수     | 원수 | 원수     | 원수     |
| 15 | 원수 | 원수 | 원수     | 원수 | 원수     | 원수     |
| 16 | 원수 | 원수 | 원수     | 원수 | 원수     | 원수     |
| 17 | 원수 | 원수 | 원수     | 원수 | 원수     | 원수     |
| 18 | 원수 | 원수 | 원수     | 원수 | 원수     | 원수     |
| 19 | 원수 | 원수 | 원수     | 원수 | 원수     | 원수     |
| 20 | 원수 | 원수 | 원수     | 원수 | 원수     | 원수     |
| 21 | 원수 | 원수 | 원수     | 원수 | 원수     | 원수     |
| 22 | 원수 | 원수 | 원수     | 원수 | 원수     | 원수     |
| 23 | 원수 | 원수 | 원수     | 원수 | 원수     | 원수     |
| 24 | 원수 | 원수 | 원수     | 원수 | 원수     | 원수     |
| 25 | 원수 | 원수 | 원수     | 원수 | 원수     | 원수     |
| 26 | 원수 | 원수 | 원수     | 원수 | 원수     | 원수     |
| 27 | 원수 | 원수 | 원수     | 원수 | 원수     | 원수     |
| 28 | 원수 | 원수 | 원수     | 원수 | 원수     | 원수     |
| 29 | 원수 | 원수 | 원수     | 원수 | 원수     | 원수     |
| 30 | 원수 | 원수 | 원수     | 원수 | 원수     | 원수     |

## 나. 표준 DB 연계 소프트웨어 개발 및 실용화

### □ 물정보DB와 연계한 알고리즘 개발완료

- DB 연계를 위한 Connection 및 Query 생성기능을 제공하여 다양한 부문의 업무 담당자가 운영 중인 DB에 접근하여 데이터를 조회할 수 있음
- 데이터 품질관리 검토정 알고리즘 적용한 품질활동 프로세스 구현
- DB조회 데이터와 및 Excel 데이터를 Import 하고, 자체적인 파일포맷(kdsm)으로 저장과 불러오기가 가능하며, Excel 형태로 Export를 지원함



## 다. 품질관리 알고리즘 타분야 확대방안 제시

- (분야조사) 지하수, 발전, 수질, 설비 등 실시간 또는 준실시간 데이터 오결측 품질관리 알고리즘 적용 분야 조사
  - 지하수(수위관측소, 600여개)와 같이 품질목표가 기 연구와 유사하여 QC 프로세스 적용이 수월함
  - \* 현황 : 관측데이터가 평균변화율보다 10m 이상 차이시 결측 처리, 오측 데이터는 1일 평균 대비 10% 오차율을 보이는 자료에 대해서 오측 인지는 하고 있으나 보정은 전무한 상황
- (기법조사) 통계 및 빅데이터 기법과 같은 최신의 기술의 적용 고려
  - 수도분야의 인공지능 기법(LSTM)이 적용된 K-PICK의 적용 내용을 파악하여 수자원 분야로 확대 고려

# 목 차

|                                    |      |
|------------------------------------|------|
| 제 출 문 .....                        | i    |
| 요 약 문 .....                        | vi   |
| 목 차 .....                          | xvii |
| 표 목 차 .....                        | xx   |
| 그림목차 .....                         | xii  |
| <br>                               |      |
| 제 1 장 서 론 .....                    | 1    |
| 1.1 연구배경 및 목적 .....                | 1    |
| 1.2 연구내용 .....                     | 2    |
| <br>                               |      |
| 제 2 장 통합물정보 품질관리 알고리즘 개선 .....     | 3    |
| 2.1 알고리즘 개선사항 .....                | 3    |
| 2.1.1 개선의견 도출 .....                | 3    |
| 2.1.2 전문가 회의 .....                 | 4    |
| 2.2 알고리즘 개선 적용 .....               | 8    |
| 2.3 프로세스 개선 적용 .....               | 9    |
| <br>                               |      |
| 제 3 장 통합물정보 품질관리 알고리즘 분석 .....     | 16   |
| 3.1 시계열 데이터 오결측 검출 분석기법 .....      | 16   |
| 3.1.1 오결측 데이터 검출 분석기법 .....        | 16   |
| 3.2 알고리즘 타분야 적용 .....              | 19   |
| 3.3.1 타분야 적용성 검토 .....             | 19   |
| 3.3.2 알고리즘 타분야 적용 및 확대 방안 제시 ..... | 25   |
| <br>                               |      |
| 제 4 장 K-WQMS 프로그램 .....            | 33   |
| 4.1 K-WQMS 개요 .....                | 33   |
| 4.1.1 K-WQMS 특징 .....              | 33   |
| 4.1.2 K-WQMS 구성 .....              | 34   |
| 4.2 K-WQMS 프로그램 기능 .....           | 35   |

|   |    |
|---|----|
| 4.2.1 자료조회 기능 .....                             | 35 |
| 4.2.2 품질관리 기능 .....                             | 37 |
| 4.2.3 검정결과 기능 .....                             | 39 |
| 4.2.4 보정결과 기능 .....                             | 40 |
| 4.2.5 신뢰도 평가 기능 .....                           | 41 |
| 4.2.6 보정 결과값 업로드 .....                          | 42 |
| 4.2.7 데이터 출력 .....                              | 42 |
| 4.3 물정보 분석 표준 DB .....                          | 43 |
| 4.3.1 ODM(Observation Data Model) .....         | 43 |
| 4.3.2 KDM(K-water Observation Data Model) ..... | 45 |
| 참 고 문 헌 .....                                   | 47 |

# 표 목 차

|  |    |
|--|----|
| 표 1 적용방안 수립 및 검·보정 알고리즘 개발 개선 의견수렴 .....       | 3  |
| 표 2 품질관리 가이드라인 수립 및 보고서 작성, 매뉴얼 분야 개선 의견수렴 · 4 | 4  |
| 표 3 품질관리 알고리즘 매개변수 적용 현황(수자원 우량) .....         | 5  |
| 표 4 품질관리 알고리즘 매개변수 적용 현황(수자원 수위) .....         | 5  |
| 표 5 품질관리 알고리즘 매개변수 적용 현황(수도 유량) .....          | 6  |
| 표 6 품질관리 알고리즘 매개변수 적용 현황(수도 압력) .....          | 6  |
| 표 7 테이블 명(VERIFY) .....                        | 8  |
| 표 8 시범 적용 결과 .....                             | 19 |
| 표 9 파주 월릉배수지 매개변수 .....                        | 20 |
| 표 10 파주 월릉배수지 탁도 매개변수 .....                    | 20 |
| 표 11 파주 월릉배수지 pH 매개변수 .....                    | 21 |
| 표 12 파주 월릉배수지 잔류염소 매개변수 .....                  | 21 |
| 표 13 파주 월릉배수지 온도 매개변수 .....                    | 22 |
| 표 14 파주 월릉배수지 온도 매개변수 .....                    | 22 |
| 표 15 데이터 품질관리 담당자 직무별 검·보정 대상 데이터(예시) .....    | 22 |
| 표 16 데이터 품질관리 계측항목 및 처리항목(예시) .....            | 22 |
| 표 17 물분야 데이터 관리 국가 물정보 Hub 구축 목표와 실행과제 .....   | 27 |
| 표 18 단계별 추진방안(예시) .....                        | 27 |
| 표 19 주요 법·규정·기준·매뉴얼 현황 .....                   | 30 |
| 표 20 물정보/경영기술 분야 규정 및 기준 현황 .....              | 30 |
| 표 21 수자원 분야 데이터 관리 현황 .....                    | 31 |
| 표 22 수도 분야 데이터 관리 현황 .....                     | 31 |
| 표 23 정보시스템 현황 .....                            | 32 |

# 그 립 목 차

|  |    |
|--|----|
| 그림 1 수도 데이터 연계 현황 .....  | 1  |
| 그림 2 데이터 관리 필요성 .....  | 1  |
| 그림 3 강우 데이터의 통계 검사 방법 모식도 예시 .....   | 7  |
| 그림 4 매개변수 적용안 제시(물리검사) .....   | 0  |
| 그림 5 매개변수 적용안 제시(지속검사) .....   | 0  |
| 그림 6 관측소 매개변수 자동 추출 코드 .....   | 11 |
| 그림 7 매개변수 적용안 제시(지속검사) .....   | 2  |
| 그림 8 추세검사 검정 결과(오류 상황) .....   | 2  |
| 그림 9 보정 결과(오류 상황) .....  | 3  |
| 그림 10 R 코딩 로직 수정 반영 .....  | 14 |
| 그림 11 추세분석 알고리즘 보정 결과(검정 결과) .....   | 4  |
| 그림 12 시스템 구성도 .....  | 15 |
| 그림 13 품질관리 프로세스 검정 계층의 알고리즘 유형 .....   | 16 |
| 그림 14 품질관리 프로세스 데이터 분야 및 검정 알고리즘별 적용 현황 .....                                | 17 |
| 그림 15 강우 데이터 검정 알고리즘의 매개변수 조사 및 적용 현황 .....                                  | 17 |
| 그림 16 수위 데이터 검정 알고리즘의 매개변수 조사 및 적용 현황 .....                                  | 18 |
| 그림 17 수도 데이터 검정 알고리즘의 매개변수 조사 및 적용 현황 .....                                  | 18 |
| 그림 18 품질관리 디렉토리 구조 .....   | 23 |
| 그림 19 R 코드 구조 .....  | 24 |
| 그림 20 확장 칼만 필터를 이용한 유량자료의 실시간 품질향상(2012, 유철상)82                              |    |
| 그림 21 Wavelet Transform 방법과 SVM 모형을 활용한 상수도 수요량 예측기<br>법 개발(2012, 권현한) ..... | 22 |
| 그림 22 SW 구성도 .....   | 3  |
| 그림 23 메뉴 구성도 .....   | 34 |
| 그림 24 검색 및 조회 화면 예시 .....  | 35 |
| 그림 25 품질 수행 기능 예시 화면 .....   | 37 |
| 그림 26 관측소 설정 화면 .....  | 38 |
| 그림 27 매개변수 설정 기능 .....   | 38 |
| 그림 28 검보정 실행 .....   | 39 |
| 그림 29 검정결과 예시 .....  | 39 |
| 그림 30 보정결과 예시 .....  | 40 |

|                           |    |
|---------------------------|----|
| 그림 31 보정 결과 선택 예시 .....   | 41 |
| 그림 32 신뢰도 평가 결과 예시 .....  | 41 |
| 그림 33 보정값 업로드 기능 예시 ..... | 42 |
| 그림 34 데이터 출력 예시 .....     | 42 |

# 제 1 장 서 론

## 1.1 연구배경 및 목적

과거에는 소규모 자료와 신속성 확보 위주의 데이터 서비스가 우선시 되었다면 최근 4차 산업 시대에 들어서는 데이터 고유의 특성을 고려한 대표성을 가질 수 있는 검증된 데이터를 서비스 하는 것에 초점이 맞춰지고 있다. 빅데이터 시대의 근간이 되는 표준 데이터에 대한 품질관리를 통해 신뢰성이 보장된 데이터를 제공함으로써 새로운 부가가치를 창출할 수 있는 기반을 확보하고 대국민을 대상으로 품질이 보증된 데이터를 서비스 할 수 있는 기반 제고가 필요한 시점이다. 최근 통합물정보 관리를 위한 데이터 표준도입과 고품질 데이터 확보를 통한 지능정보기술 적용기반확장으로 4차 산업혁명을 선도하기 위한 K-water 데이터 관리가 중요해짐에 따라 수자원(수위, 우량) 및 수도(압력, 유량) 등 현재의 데이터 관리수준을 분석하고 시계열 데이터의 품질관리를 위해 기개발된 검·보정 알고리즘의 계측 데이터의 범용적인 품질관리 절차 도출 및 현업에서 개발중인 물정보 분석DB에 실증하여 검증하는 연구가 필요한 실정이다.

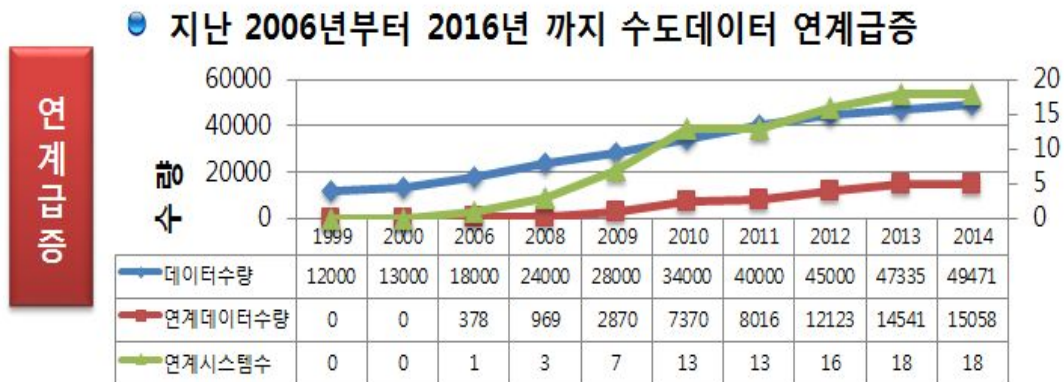


그림 1 수도 데이터 연계 현황



그림 2 데이터 관리 필요성

## □ 기술적 측면

- 계측 데이터별 품질관리 적합 방안(알고리즘, 매개변수 최적화) 및 표준화 된 DB에 연계 가능한 시스템의 자체 개발을 통한 기술력 확보 필요

## □ 경제·산업적 측면

- 4차 산업혁명의 기반이 되는 표준 데이터에 대한 품질관리를 통해 신뢰성 높은 데이터 제공으로 新부가가치 창출기반 확보 및 대국민 신뢰도 제고 필요

## □ 국내외 관련(유사) 기술의 개발 현황

- 국내 데이터 오결측에 대한 검출 알고리즘 개발 및 정제알고리즘 연구개발
  - 오결측 관련 DB연계 프로그램 개발 및 사례분석, 정제방법론 개발
  - 상관관계 분석을 위한 데이터 추출 TB 구축 및 SW모델 개발
  - \* 분석기법과 알고리즘 개발 및 적용이 본 연구와는 차이점이 있음

## 1.2 연구내용

### □ 알고리즘 적용성 검증 및 파일럿 S/W 개발

- 알고리즘 적용성 검증 및 유형별 적용기법 분석 및 Pilot S/W 개발
  - 데이터 생애주기별(FA→OA간) 적용성에 대한 평가
  - 오결측 형태(부분/연속) CASE별 분석
  - 실제 표준 DB와 연계하여 Site, Result, Value Table을 참조 가능한 S/W 개발
  - 시계열 데이터의 품질관리 후 품질관리수준을 조정하여 표준DB에 반영

### □ 표준 DB 연계 자동화를 위한 알고리즘 개발 및 최적 보정기법 제시

- 표준 DB 연계 자동화를 위한 알고리즘 개발 및 시범적용 수계에 대한 최적 보정 기법 제시
  - 일정 주기별 품질관리 자동화 적용기술 개발
  - 품질관리 알고리즘 시범적용 대상 수계선정 및 최적 보정기법

### □ 타분야 확대적용 방안 마련

- 하수, 발전, 수질, 설비 분야의 품질관리 알고리즘 적용방안 마련

# 제 2 장 통합물정보 품질관리 알고리즘 분석

## 2.1 알고리즘 개선사항

### 2.1.1 개선의견 도출

기존에 개발된 알고리즘에 대한 현업 요구사항을 수용하고 향후 물정보 DB(가칭 KDM)에 실용화를 구현하고자 데이터 품질관리 프로세스 구축 및 검보정 알고리즘 적용 방안 등 총 4개 분야에 대해 표1, 표2와 같이 개선의견을 도출하였으며, 해당 사항을 중점적으로 반영하였다.

표 1 적용방안 수립 및 검·보정 알고리즘 개발 개선 의견수렴

| ① 데이터 품질관리 프로세스 구축 및 적용방안 수립  | ② 프로세스 계층별 데이터 검·보정 알고리즘 개발   |
|---|---|
| <p>2.3.1 데이터 품질관리 방향성</p> <p>① 데이터 품질관리 절차</p> <p>가. 품질관리 검정 프로세스 제안</p> <p>나. 품질관리 보정 프로세스 제안</p> <p>2.3.2 데이터 품질관리 차별성</p> <p>① 공통적으로 적용할수 있는 품질관리 프로세스 제안(범용)</p> <p>(고도화) 운영 CASE별 평가마련<br/>(장기결측, 부분결측등)</p> <p>(확장성) 공간, 항목적 확장가능여부</p> <p>(좌표) 소유역(관측국 좌표 알고리즘) 적용여부</p> <p>- 통합물관리처 협업필요(조완희수석)</p> | <p>3.2.1 검정 알고리즘</p> <p>① 공간 검사에서 강우의 경우 인근관측소와 RDS 산정값 적용여부</p> <p>② 1시간 이상 동일강우, 무강우 제외 (알고리즘은 어떻게 구성)</p> <p>3.3 데이터 보정 알고리즘의 신뢰도 평가 방법</p> <p>① K-water내 데이터 신뢰도를 참고하여 2%로 설정(신뢰도?)</p> <p>4.1 수자원 분야 데이터</p> <p>① 표4.1에서 표추가(데이터 추출DB표현) 예) 00수위 : TECH DB</p> <p>4.1.1 강우</p> <p>① 표4.2에서 데이터 신뢰도(%)등 수식으로 표현 필요 신뢰도(C) = (A*B)/B</p> <p>② 오측(물리+통계)와 실제 게릴라성 강우의 판별</p> <p>③ 마항(P.60) 40% 이상이 되어 분한 적용성(앞뒤 문맥)</p> |

표 2 품질관리 가이드라인 수립 및 보고서 작성, 매뉴얼 분야 개선 의견수렴

| ③ 품질관리 가이드라인 수립   | ④ 보고서 작성 및 매뉴얼 작성  |
|---|--|
| <p>5장 제목 및 목차</p> <ul style="list-style-type: none"> <li>① 5.5, 5.6, 5.7의 제목 검토필요</li> <li>② 전반적으로 페이지 공백 많음</li> </ul> <p>5.3.3 K-water 품질관리시스템</p> <ul style="list-style-type: none"> <li>① 시스템 소개 정도이며 기존 시스템 분석 후 신규 알고리즘에 대한 차별성 표현이 필요</li> </ul> <p>5.4.1 오결측 검정 알고리즘</p> <ul style="list-style-type: none"> <li>① 가이드라인이므로 해설부문에 예제로 이해가 쉽게 될 수 있도록 간단한 설명이 필요</li> </ul> | <p>일반) 홍수통제소, 기상청, Kwater 에서 운영중인 DB시스템에 탑재된 통계학적 알고리즘 자료 및 분석 자료 필요</p> <p>일반) 수동품질관리 알고리즘에 대한 자동화 적용기법과 개선에 대한 방안제시 필요</p> <p>일반) 신규 적용된 알고리즘의 요약표 필요<br/>- 개발된 통계알고리즘에 대한 단계별 절차와 활용방법, 적용대상 기술필요</p> |

### 2.1.2 전문가 회의

관련 협업부서인 통합물관리처와 물종합정보처의 원활한 과제수행을 위해서 주기적 또는 비상시적 전문가 회의를 지속적으로 수행하였으며 주요 개선 의견 도출 내용은 다음과 같다.

#### 가. 매개변수

첫 번째로 품질관리 알고리즘 매개변수 적용 등에 관한 내용으로 1차 Kick Off 회의(19.327)시 제시된 안건으로 주요 내용은 다음과 같다. 품질관리 알고리즘 매개변수 적용을 위해 매개변수는 품질관리 알고리즘을 수행하기 위한 설정값으로 관측지점 또는 유역별로 지정하여 사용하며 초기 매개변수는 각 관측소 지점별로 DB 구축되어야 하며, 추후 운영을 통해 사용자가 매개변수를 조정하여 사용가능도록 구성한다. 적용을 위한 환경 구성은 협의 후 반영하며 업무분장은 현업은 DB 테이블 구성 및 매개변수 지정 입력을 하고 연구원은 통계 분석을 통한 매개변수 적용(안)을 제시하는 것으로 한다. 시공간 특성을 고려한 매개변수 설정은 품질관리 수준을 향상시키는 디테일한 부분이며 전국단위 관측소를 기준으로 개별 매개변수를 설정하는 것은 상당량의 기초자료와 분석 시간이 필요하다. 따라서 금회 반영한 내용으로는 전국단위 범위는 반영하면서 시그마 값이 적용되는 일반적인 기준의 항목을 제외한 물리검사 및 ARIMA 차수에 대해서 각 관측소별 분석값을 적용할 수 있도록 구성하는 것이다. 우선 기존 연구를 통해 반영된 각 항목별 매개변

수의 내용은 다음 표와 같다.

표 3 품질관리 알고리즘 매개변수 적용 현황(수자원 우량)

| 변수명   | 내용                | 적용값   | 기타   |
|-------|-------------------|-------|--|
| set1M | 물리검사 물리한계 분단위 상한값 | 10    | 하한 (0mm 고정)                                |
| set2P | 추세검사 단위시간 내 변동한계  | 6     | 변화량평균( $\bar{\chi}$ ) + 6×표준편차( $\sigma$ ) |
| set3  | 지속검사 한계 설정값       | 6     | 지속 기간수(1시간=10분 × 6)                        |
| set4  | 통계검사 오차범위 설정값     | 3     | 평균관측치( $\bar{\chi}$ ) ± 3×표준편차( $\sigma$ ) |
| set5L | 공간검사 설정 하한값       | 0.2   | 인근관측소 20% 미만                               |
| set5U | 공간검사 설정 상한값       | 2.8   | 인근관측소 280% 초과                              |
| P,D,Q | ARIMA 보정 차수 설정값   | 2,1,2 |  |

수자원(우량) 관측소 관련 매개변수는 검정5개, 보정1개 값을 사용하였다. 주목할 부분은 공간검사 설정 상·하한 기준은 홍수통제소 및 기타 논문에 적용된 기준치보다 한계 범위를 축소하여 상대적으로 엄격한 기준으로 설정하여 오측 판단의 유효성을 증가시켰다.

표 4 품질관리 알고리즘 매개변수 적용 현황(수자원 수위)

| 변수명   | 내용                 | 적용값   | 기타   |
|-------|--------------------|-------|--|
| set1L | 물리검사 물리한계 하한 설정값   | 0     | 단위시간 관측값의 하한                               |
| set1U | 물리검사 물리한계 상한 설정값   | 13.13 | 단위시간 관측값의 상한                               |
| set2M | 추세검사 10분내 변동한계 설정값 | 0.2   | 변동한계 초과                                    |
| set2N | 2시간 반향 기울기 설정값     | 1     | 평균기울기와 현재기울기 -1배                           |
| set2P | 2시간 동향 기울기 설정값     | 3     | 평균기울기와 현재기울기 3배                            |
| set3  | 지속검사 한계 설정값        | 6*24  | 지속기간수 (10분 × 6 × 24)                       |
| set4  | 통계검사 오차범위 설정값      | 3     | 평균관측치( $\bar{\chi}$ ) ± 3×표준편차( $\sigma$ ) |
| set5  | 공간검사 오차범위 설정값      | 3     | 직전 일주일과 현재 수위차의 오차평균 ± 3×표준편차( $\sigma$ )  |
| P,D,Q | ARIMA 보정 차수 설정값    | 2,1,2 |  |

위의 표에서 알 수 있듯이 수자원(수위) 관측소 관련 매개변수는 검정8개, 보정1개 값을 사용하였다.

표 5 품질관리 알고리즘 매개변수 적용 현황(수도 유량)

| 변수명   | 내용               | 적용값   | 기타                                       |
|-------|------------------|-------|--|
| set1L | 물리검사 물리한계 하한 설정값 | 0     | 단위시간 관측값의 하한                             |
| set1U | 물리검사 물리한계 상한 설정값 | 33200 | 단위시간 관측값의 상한                             |
| set2M |                  | 0     |  |
| set2H |                  | 0     |  |
| set3  | 지속검사 한계 설정값      | 20    | 지속기간수 (1분 × 20)                          |
| set4  | 통계검사 오차범위 설정값    | 3     | 평균관측치( $\bar{x}$ ) ± 3×표준편차( $\sigma$ )  |
| set5  | 공간검사 오차범위 설정값    | 3     | 직전 1일과 현재 유량차의 오차평균 ± 3×표준편차( $\sigma$ ) |
| P,D,Q | ARIMA 보정 차수 설정값  | 2,1,2 |  |

위의 표에서 알 수 있듯이 수도(유량) 관측소 관련 매개변수는 검정7개, 보정1개 값을 사용하였다.

표 6 품질관리 알고리즘 매개변수 적용 현황(수도 압력)

| 변수명   | 내용               | 적용값   | 기타                                       |
|-------|------------------|-------|--|
| set1L | 물리검사 물리한계 하한 설정값 | 0     | 단위시간 관측값의 하한                             |
| set1U | 물리검사 물리한계 상한 설정값 | 10.0  | 단위시간 관측값의 상한                             |
| set2M |                  | 0     |  |
| set2H |                  | 0     |  |
| set3  | 지속검사 한계 설정값      | 20    | 지속기간수 (1분 × 20)                          |
| set4  | 통계검사 오차범위 설정값    | 3     | 평균관측치( $\bar{x}$ ) ± 3×표준편차( $\sigma$ )  |
| set5  | 공간검사 오차범위 설정값    | 3     | 직전 1일과 현재 압력차의 오차평균 ± 3×표준편차( $\sigma$ ) |
| P,D,Q | ARIMA 보정 차수 설정값  | 2,1,2 |  |

위의 표에서 알 수 있듯이 수도(압력) 관측소 관련 매개변수는 검정7개, 보정1개 값을 사용하였다.

## 나. 알고리즘

두 번째 향후 알고리즘 개발에 있어서 여러 가지 중요한 사항은 아래와 같다. 먼저 강우 같은 경우 공간적 연계성을 적용하는 공간검사 및 보정 알고리즘에 있어 목적 관측소 기준으로 참조가 되는 주변 관측소 지점 선정이 중요한데 기존 연구에서는 거리에 따른 임의의 지점을 선정하여 반영으나 다수의 지점에 대한 수동 선정은 효율이 저하함에 따라 본 연구에서는 강우는 RDS 보정 기법을 적용하여 소유역 단위 포함된 관측소 중 4분위에 해당하는 관측소 정보를 적용하는 것으로 2개 이하일 경우는 적용하지 않는 것으로 하였으며 수위의 경우는 우선 1개의 인근 관측소로 선정하고, 상·하류 관계보다는 거리를 우선으로 하는 것으로 하였다. 여기서 향후 검토가 되어야 할 내용은 본류와 지류가 혼합

된 경우는 어떻게 처리할 것인가 그리고 수위의 경우 분류 선상에 있는 관측소라도 유의 거리가 있어 적용 여부를 판단할 필요는 없는지를 검토하는 것이 필요하였다.

강우 데이터의 경우 통계 검사시 강우는 확률적 오차에 따른 기준으로서 관측소간 데이터 편차의 평균을 의미하는 [평균관측치( $\bar{X}$ ) +(-) 3\*표준편차( $\sigma$ )]의 방법을 적용하여 사용하였고 과년도 데이터를 대상으로 상한 범위 이상의 값을 갖는 데이터를 대상으로 해당 데이터의 시계열(index)을 파악한다. 또한 품질관리 기간에 해당 하는 과년도 데이터를 대상으로, 하한 범위 이하의 값을 갖는 데이터를 대상으로 해당 데이터의 시계열(index)을 파악한다. 강우 데이터의 통계검사 방법 모식도 아래 그림과 같다.

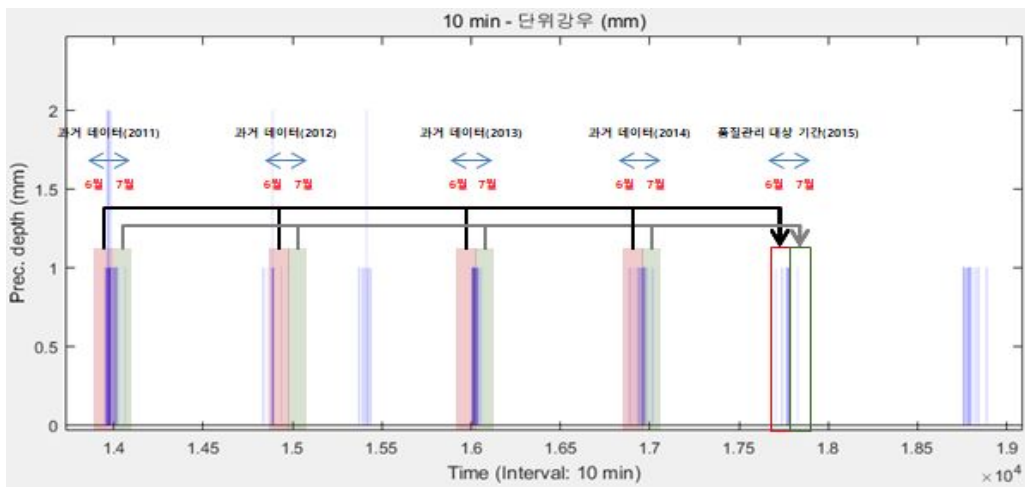


그림 3 강우 데이터의 통계 검사 방법 모식도 예시

우량 데이터는 공간검사의 경우 인근 관측소 RDS 값이란 의미가 평균 적용하는 것인지 검토가 필요하였으며 인근관측소의 경우 직선거리 가까운 순으로 2개 이상 3개 이하를 적용하는 것도 검토가 필요하였다. 관측소가 2개 이하일 경우 검정과 보정은 어떻게 할 것인가도 검토 대상 이었다. 수위데이터는 통계검사의 경우 수위 데이터의 품질관리 알고리즘은 확률적 오차에 따른 기준으로서, 관측소간 데이터 편차의 평균을 의미하는 [평균관측치( $\bar{X}$ ) +(-) 3\*표준편차( $\sigma$ )]의 방법을 적용하여 설정값으로 사용하는데 과거 대상 기간과 홍수기 / 갈수기의 구분은 적용되어 있는지 확인이 필요하며 공간검사의 경우 수위 데이터의 공간적 오차에 따른 기준으로써, 인근 관측소 값을 활용하여 직전 일주일에 대하여 대상 관측소와 참조 관측소간 수위 차의 [편차평균  $\pm$  3\* 표준편차( $\sigma$ )]의 방법을 적용하여 설정값으로 사용하는데 만약 공간 관측소 데이터가 없다면 어떤식으로 알고리즘을 적용을 해야 하는지 전문가 의견을 수렴하였다. 그리고 현재 구현된 Matlab 알고리즘은 1분 단위 분석이 가능한 구조로 되어 있는지 이것이 안되면 R언어로 컨버팅을 해야 하는지 여부도 제시되었다.

## 다. 프로세스

세 번째 전사 데이터 분석용 표준 DB의 품질관리 도구로 알고리즘을 도입함에 있어 용도에 따른 적용 방안에 대해서 논의 되었다. 데이터 생산에서 활용까지(FA→OA) 데이터 흐름 전체에 반영될 수 있는 방안으로 현업에서 직접 담당자가 사용할 수 있는 SW와 본사에서 주기적으로 수행하는 시스템으로 구별하여 동일한 알고리즘을 적용하는 것이다. 데이터 품질관리의 2중화 구조를 통해서 데이터의 레벨화 적용이 가능하고 최종적 판단의 기준을 다양화 할 수 있도록 체계를 구축하여 반영한다는 내용이다.

## 2.2 알고리즘 개선 적용

기존의 알고리즘은 유역(대청댐) 공통적으로 적용되어 있으며, 항목은 우량(6개), 수위(9개), 유량(7개), 압력(7개)이 각각 기준에 따라 아래 그림으로 적용되어 있어 이를 통해 개선사항을 발굴하고 검증하는 작업이 필요하다.

알고리즘 매개변수 적용방안에 있어서 알고리즘 초기 매개변수는 DB상 입력하여 관리하도록 하였고 매개변수 Table 구성안은 단일 테이블로 구성하여 반영(매개변수 최대치 기준 가변 테이블)하고 연구진은 과거 10년 데이터를 기반으로 통계 범위 제시하고 현업에서는 매개변수 결정 및 입력하는 것으로 정리하였다. 관측소 정보 테이블 구성 및 자료 입력을 요청하였고 이는 물종합정보처에서 지정 관측소를 기준으로 작업이 필요한 상황이며 관측소 기본정보와 필수항목인 좌표 및 고도 정보도 구축되어 제공 되어야 한다. 관측소 네트워크 관계성을 지정하여 관리하는 작업을 금회 수행 예정이며 강우(RDS), 수위(상류지점), 유량, 압력(상류 노드) 관계성을 정리하고 공간보정을 위한 기본 정보를 DB상에서 도출할 수 있음으로 매우 효율적이다. 매표에서 R로 컨버팅 하기 위해서 테이블을 정의하고 QC VERIFY를 아래 표 2.6과 같이 정의하고 지속적으로 컬럼 등에 대한 수정을 하였다.

표 7 테이블 명(VERIFY)

### QC\_VERIFY\_INFO

| 번호 | 컬럼한글명     | 컬럼명(변경)           | KEY | Datatype     |
|----|-----------|-------------------|-----|--------------|
| 1  | 검정알고리즘번호  | QC_VERIFY_INFO_ID | PK  | NUMBER       |
| 2  | 알고리즘 형식   | QC_TYPE           |     | VARCHAR2(10) |
| 3  | 물리검사하한    | SET1L             |     | NUMBER       |
| 4  | 물리검사상한    | SET1U             |     | NUMBER       |
| 5  | 추세검사_변동한계 | SET2M             |     | NUMBER       |
| 6  | 반향_기울기    | SET2N             |     | NUMBER       |
| 7  | 동향_기울기    | SET2P             |     | NUMBER       |
| 8  | 지속검사한계    | SET3              |     | NUMBER       |
| 9  | 통계검사_오차범위 | SET4              |     | NUMBER       |

|    |           |                |  |        |
|----|-----------|----------------|--|--------|
| 10 | 공간검사_오차범위 | SET5           |  | NUMBER |
| 11 | 공간검사_하한   | SET5L          |  | NUMBER |
| 12 | 공간검사_상한   | SET5U          |  | NUMBER |
| 13 | 사용승인여부    | USE_APPROVE_YN |  | BOOL   |

#### QC\_SCHE

| 번호 | 컬럼한글명     | 컬럼명(변경)             | KEY | Datatype     |
|----|-----------|---------------------|-----|--------------|
| 1  | 스케줄번호     | QC_SCHE_ID          | PK  | NUMBER       |
| 2  | 대상 RST_ID | RST_ID              | FK  | NUMBER       |
| 3  | 검정알고리즘번호  | QC_VERIFY_INFO_ID   | FK  | NUMBER       |
| 4  | 보정알고리즘번호  | QC_REVISION_INFO_ID | FK  | NUMBER       |
| 5  | 설정자       | USER_EMP_NO         |     | VARCHAR2(8)  |
| 6  | 설정일시      | SETTING_DATE        |     | DATETIME     |
| 7  | 승인자       | APPROVE_EMP_NO      |     | VARCHAR2(8)  |
| 8  | 승인일자      | APPROVE_DATE        |     | DATETIME     |
| 9  | 승인여부      | APPROVE_YN          |     | BOOL         |
| 10 | 시작일자      | START_DATE          |     | DATETIME     |
| 11 | 종료일자      | END_DATE            |     | DATETIME     |
| 12 | 수행주기      | EXC_INTRVL          |     | NUMBER       |
| 13 | 대상 레벨     | T_QC_LEVEL          |     | VARCHAR2(20) |
| 14 | 보정 레벨     | QC_LEVEL            |     | VARCHAR2(20) |

#### QC\_REVISION\_INFO

| 번호 | 컬럼한글명    | 컬럼명(변경)             | KEY | Datatype     |
|----|----------|---------------------|-----|--------------|
| 1  | 보정알고리즘번호 | QC_REVISION_INFO_ID | PK  | NUMBER       |
| 2  | 알고리즘형식   | REVISION_TYPE       |     | VARCHAR2(14) |
| 3  | P        | ARIMA_P             |     | NUMBER       |
| 4  | D        | ARIMA_D             |     | NUMBER       |
| 5  | Q        | ARIMA_Q             |     | NUMBER       |
| 6  | 사용승인여부   | USE_APPROVE_YN      |     | BOOL         |

#### QC\_REF\_SITE

| 번호 | 컬럼한글명       | 컬럼명(변경)    | KEY | Datatype |
|----|-------------|------------|-----|----------|
| 1  | 참조지역 RST_ID | RST_ID     | PK  | NUMBER   |
| 2  | 스케줄ID       | QC_SCHE_ID | FK  | NUMBER   |
| 3  | 사용여부        | USE_YN     | FK  | BOOL     |

#### QC\_SCHE\_LOG

| 번호 | 컬럼한글명     | 컬럼명(변경)             | KEY | Datatype     |
|----|-----------|---------------------|-----|--------------|
| 1  | 스케줄로그번호   | QC_SCHE_LOG_ID      | PK  | NUMBER       |
| 2  | 스케줄번호     | QC_SCHE_ID          | FK  | NUMBER       |
| 3  | 대상 RST_ID | RST_ID              | FK  | NUMBER       |
| 4  | 검정알고리즘번호  | QC_VERIFY_INFO_ID   | FK  | NUMBER       |
| 5  | 보정알고리즘번호  | QC_REVISION_INFO_ID | FK  | NUMBER       |
| 6  | 보정레벨      | QC_LEVEL            |     | VARCHAR2(20) |
| 7  | 수행일시      | RUN_DATE            |     | DATETIME     |
| 8  | 수행주기      | EXC_INTRVL          |     | NUMBER       |

|    |         |            |  |          |
|----|---------|------------|--|----------|
| 9  | 시작데이터일자 | START_DATE |  | DATETIME |
| 10 | 종료데이터일자 | END_DATE   |  | DATETIME |

### 가. 매개변수 분석

R로 개발된 매개변수 추출 기능실행파일은 총 4가지 매개변수 추출로 min, max, 지속 검사 개수, ARIMA(P,D,Q) 기존 알고리즘 호출방식과 동일하고 입력자료 엑셀구성과 output 파일 이름만 설정하면 결과(엑셀) 도출된다. 입력자료 파일을 보게 되면 날짜와 값으로 구성되어 있고 향후 데이터 추출시 요긴하게 활용될 것으로 판단이 된다. 그리고 출력파일을 보시면 순서대로 min, max, arima\_p, arima\_d, arima\_q와 per개수 10개로 구성되어 있습니다. 그리고 품질관리 알고리즘 매개변수 적용안에 있어서 샘플로 1999 ~ 2018, 청성 수위관측소를 위주로 물리검사는 과거 데이터를 기초로 최고치와 최저치 또는 상위 % 이상 기준값을 제시(10년 ~ 20년)하였다. 이 해당 제시값은 통계적인 기준치임으로 현장 계측기 한계 반영이 필요하며 내용은 아래 그림과 같다.

[조건]

청성 : 665929건

※ 2006부터 데이터 생성

```
> print(max1)
[1] 6.38
> print(min1)
[1] -10.02
> quantile(b, c(0.9999))
99.99%
5.288144
```

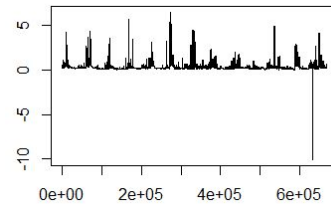
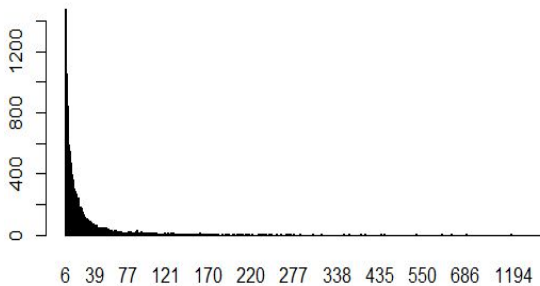


그림 4 매개변수 적용안 제시(물리검사)

지속검사의 경우 과거 데이터를 기초로 지속적 반복 회수를 검토하여 상한 범위 제시하였다. 그 내용은 아래 그림과 같다.



```
> y[1:10,]
  Var1 Freq
1     6 1475
2     7 1055
3     8  845
4     9  787
5    10  589
6    11  549
7    12  472
8    13  397
9    14  361
10   15  358

> y[120:130,]
  Var1 Freq
120  125   9
121  126   8
122  127  13
123  128   9
124  129  15
125  130   8
126  131   9
127  132   3
~
128  133   8
129  134   8
130  135   5
```

그림 5 매개변수 적용안 제시(지속검사)

그리고 ARIMA의 경우 P,D,Q 차수 설정에 있어 자동화를 고려하였으며 시그마 배수를 사용하였다. 그간 검증을 통해 알고리즘을 개선 후 다시 물종합정보처의 현업부서에서 시연을 수행한 후 현업부서의 니즈와 의견을 최종 다시 수렴해보았다. 그 내용은 수자원분야의 우량의 경우 공간검사 알고리즘은 RDS 적용을 요청하였고 이는 거리와 유역 단위

를 고려하여 반영하는 것이 필요하였다. 수자원 분야의 수위의 공간 알고리즘은 레퍼런스 지점 적용을 요청하였으며 이는 GIS 데이터를 가지고 수작업을 통한 상류 지점 선정 또는 분류 지점에 대해서는 상관관계 분석을 통해 유사성이 높은 지점으로 선정이 필요하다. 이는 매칭 테이블을 구성하여 시스템내 반영할 필요가 있다. 그리고 수도의 경우 공간 알고리즘 관련 네트워크 관계를 어떻게 구성해야 하는지의 대한 고찰이 필요하며 지점 확장에 있어 각 지점별 매개변수 적용방안에 있어 WRO 기준 관측소 코드 및 정보를 제공하면, R을 통해 DB 연계 데이터 수집 및 분석 후 매개변수 도출 반영하는 것이 필요하다. 그리고 지점확장의 경우 물리적 상하한과 ARIMA PDQ 반영/이외 변수는 연구 결과를 적용한다.

- 지역적 확대에 따른 적용 관측소는 현재 K-water에서 관리하고 있는 WRO DB에 존재하고 있는 사이트를 대상으로 하고 있으며, 이는 우량관측소 1982개, 수위관측소 1205개 지점으로 최근3년(2016.9 ~ 2019.9) 자료를 분석하여 품질관리 알고리즘 적용 매개변수인 최대/최소값 및 arima 계수값 등을 도출하여 반영하였다.
- 시·공간 등 변화하는 환경에 대응하기 위해서 매개변수 조정을 정기적으로 분석하여 반영할 필요가 있기에 매개변수 도출 기능으로 개발하여 S/W 반영하였다.
- 담당자는 해당 관측소의 경험값과 도출된 분석 결과를 판단하여 각 관측소 고유 매개변수로 설정하여 관리한다.
- DB 현황 : 결측값 데이터에 적용한 999 이상의 값도 존재하고 있었으며, 누적값과 실적값이 혼재되어 사용되는 경우도 있다. 999 이상값과 -값은 분석 대상에서 제외하여 계산하였으며 arima는 R패키지에서 제공하고 있는 auto.arima를 적용하여 차수를 결정하여 반영한다.

```

43 ~ if(nrow(data) != 0) {
44
45   #data frame -> frame sort
46   data <- data.frame(data)
47   ss <- data[order(data[,1], data[,2]),]
48   ss[which(ss[,2] >= 999),] <- NA
49   ss[which(ss[,2] < 0),] <- 0
50   ss <- na.omit(ss)
51
52 ~ if(nrow(ss) != 0) {
53   parameter_max <- max(ss[,2]) ; parameter_min <- min(ss[,2])
54   parameter_arima <- arimaorder(auto.arima(ss[,2], seasonal = TRUE))
55   parameter_result <- cbind(obs_code, obs_name, parameter_max, parameter_arima[1], parameter_arima[2], parameter_arima[3])
56 ~ } else {
57   parameter_result <- cbind(obs_code, obs_name, "", "", "", "")
58 ~ }
59
60 ~ } else {
61   parameter_result <- cbind(obs_code, obs_name, "", "", "", "")
62 ~ }

```

그림 6 관측소 매개변수 자동 추출 코드

## 나. 알고리즘 고도화

오측 유형의 데이터를 검출하기 위한 필수 알고리즘으로 적용된 추세검사는 지정 시간 내 변동값 추세적 변동한계를 초과하는 값을 추출하기 위한 방법으로 적용된 항목에 따라 세 개 또는 네 가지 변수를 입력받아 해석한다. 수위의 예를 들면 「과거 2시간 평균

기울기와 현재 기울기 차이 3배 이상」, 「과거 2시간 평균기울기와 현재 기울기 차이 -1 배 이하」의 경우를 설정값으로 적용하여 사용중에 있다.

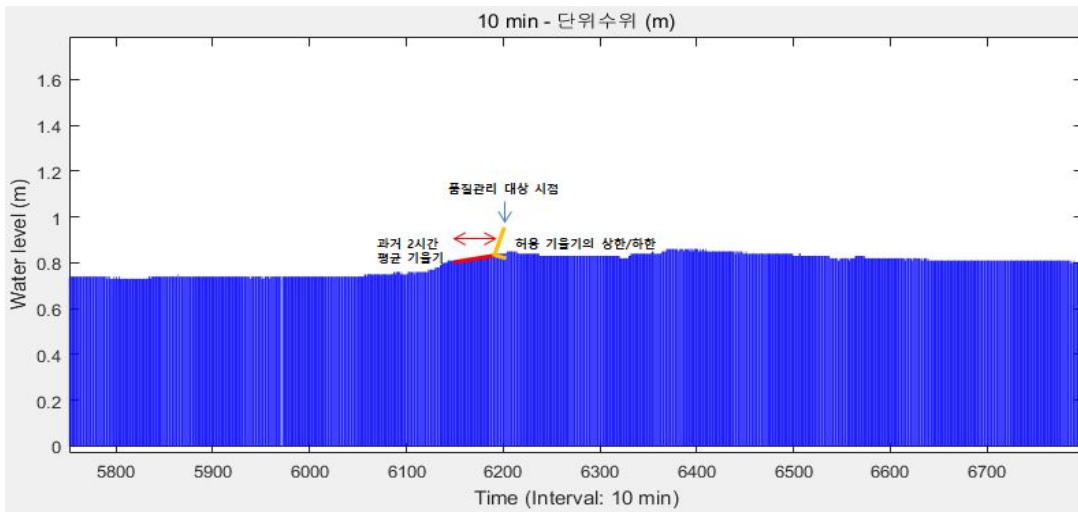


그림 7 매개변수 적용안 제시(지속검사)

그러나 다음과 같이 오측이 연속된 구간 범위에서 발생하는 추세검사 검측 시 구간 전체를 오측 범위로 판단하지 않고 시작점과 마지막 끝나는 지점에 대해서만 오류로 판단하는 경우가 있다. 마지막 끝나는 지점은 실제적으로 정상범위의 값임에도 불구하고 오측으로 판별하는 심각한 오류인 샘이다. 이러한 문제점은 판별에서만 끝난 것이 아니라 판별값을 기준으로 보정하기 때문에 보정 결과도 아래 그림과 같이 영향을 끼치게 된다.

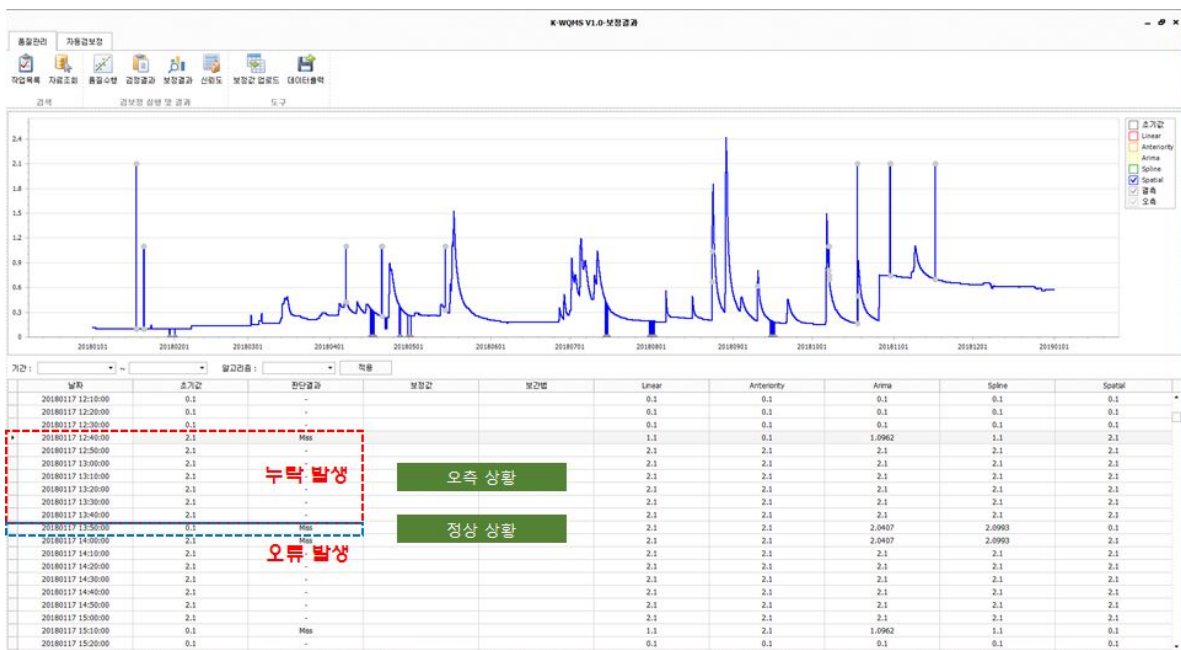


그림 8 추세검사 검정 결과(오류 상황)

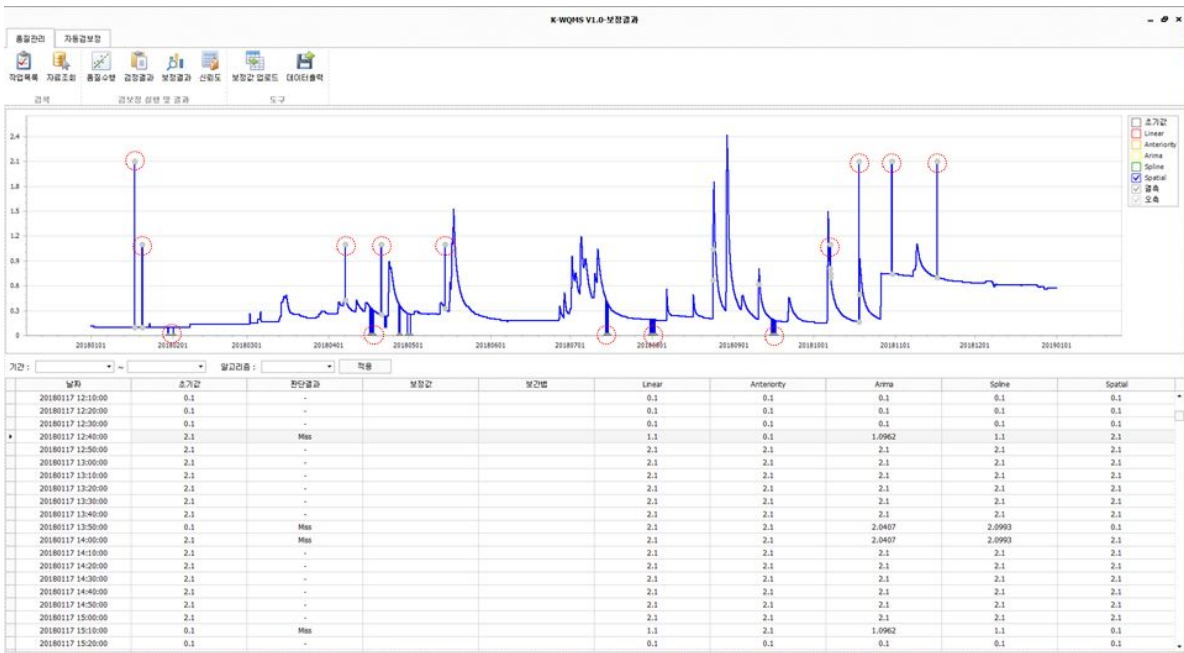


그림 9 보정 결과(오류 상황)

주된 원인은 판단 기준이 검측 기준 시점과 판단 바로 전 시점의 값을 비교하여 검증하는 알고리즘 특성에 따른 것으로 시작지점의 오류 발생 인지 후 같은 오류값이 발생하더라도 바로 전 값과의 기울기 검사를 수행하는 과정에서 오류의 지속적인 상태가 정상적인 상태로 판별되어 발생하는 것으로 검증 과정에서의 오류 발생뿐만 아니라 오류값을 기준으로 보정하는 프로세스에 따라 보정시에도 오류 상태를 그대로 반영하는 크리티컬한 문제점을 내포하고 있었다.

이와 같은 문제점을 해결할 수 있는 방안은 알고리즘 자체를 변경하는 방식이 아닌 R 코드상에 적용된 전산적인 로직을 변경하는 것으로 검증 시점값과 바로 전 값을 비교하는 방식에서 중간과정에 Tempdata 변수를 임의로 생성하여 이전값을 대체하여 비교하게 하는 방식이다. 추세검사 매개변수 적용 분석에 따른 정상적인 상황에서는 Tempdata값이 기존 전 값으로 변경되게 하여 지속적인 검사를 진행하게 되고 추세분석 한계 범위에 포함되는 검측 범위의 경우 Tempdata를 정상적인 상황의 마지막 기준치를 그대로 사용하게 됨으로써 구간 범위 오측 상황을 판별할 수 있게 된다.

```
tempdata <- unitFilledData[timestep+1]
```

Tempdata 생성

```

for (i in (timestep+2):length(unitFilledData)){
  searchType2set2N <- -(searchType2Nset)*mean(diff(unitFilledData[(i-timestep-1):(i-1)]))
  searchType2set2P <- (searchType2Pset)*mean(diff(unitFilledData[(i-timestep-1):(i-1)]))

  if (abs(unitFilledData[i]-tempdata) > searchType2set1M) {
    searchType2IndM <- c(searchType2IndM, i)
  } else {
    if (abs(mean(diff(unitFilledData[(i-timestep-1):(i-1)]))) >= 0.03){
      if (mean(diff(unitFilledData[(i-timestep-1):(i-1)])) > 0) {
        if((unitFilledData[i] - tempdata) <= searchType2set2N || ((unitFilledData[i] - tempdata) >= searchType2set2P)) {
          searchType2Ind <- c(searchType2Ind, i)
          unitFilledData[i] <- unitFilledData[i-1]
        }
      } else if (mean(diff(unitFilledData[(i-timestep-1):(i-1)])) < 0) {
        if((unitFilledData[i] - tempdata) >= searchType2set2N || ((unitFilledData[i] - tempdata) <= searchType2set2P)) {
          searchType2Ind <- c(searchType2Ind, i)
          unitFilledData[i] <- unitFilledData[i-1]
        }
      } else {
        tempdata <- unitFilledData[i]
      }
    } else {
      tempdata <- unitFilledData[i]
    }
  }
  searchType2Ind <- unique(c(searchType2IndM, searchType2Ind))
}

```

오류상황 : tempdata 그대로 사용하여 비교

정상상황 : tempdata 변경

그림 10 R 코딩 로직 수정 반영

해당 로직변경을 적용한 추세분석 알고리즘의 검보정 결과는 아래 그림과 같이 확실한 오측 판별과 보정이 적용된 상황을 확인할 수 있다.

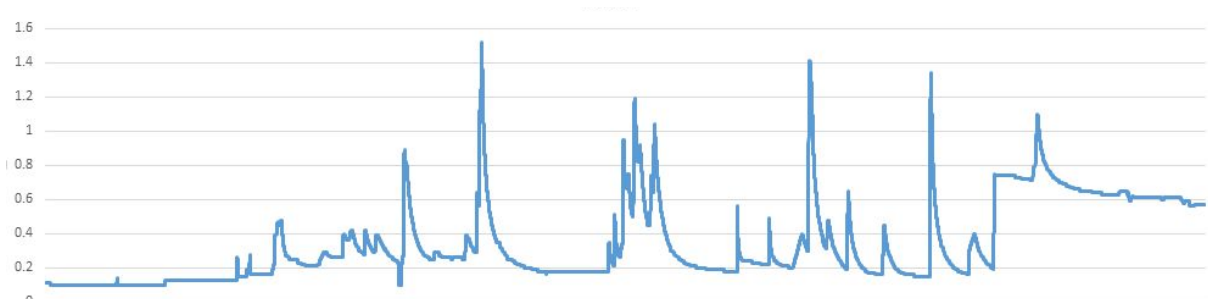


그림 11 추세분석 알고리즘 보정 결과(검정 결과)

현재 적용된 추세검사의 추가적인 한계 요소는 우량 측정항목은 갈수기시 지속적으로 0값이 존재할 수 있다는 점이다. 이와 같은 구간에서 아주 작은 1의 값이 발생하더라도 추세검사의 평균 기울기 적용시 오측으로 판별하는 경우가 발생한다. 향후 추세검사의 정확성 향상을 위해서는 과거 3년 또는 장기간의 홍수기, 갈수기 및 홍수 사상 등을 포함한 추세를 적용하여 기준한계를 먼저 검토하고 기간 평균 기울기를 적용하는 방식으로 오측 판단의 기준을 유연하게 확장 시킬 필요가 있을 것으로 판단된다.

### 2.3 프로세스 개선 적용

기존 품질관리 체계와 명확한 기준이 미흡한 상황에서 구축한 품질관리 프로세스를 통해 단순 병렬식 검정 방식을 탈피하여 오결측 발생 특성을 고려한 필수/세부 검사 수행과 보정을 실시 하였다. 본 연구에서는 로컬데이터를 대상으로 단일적인 수행하는 방식과 더불어 자동화 방식을 적용하여 주기적으로 스케줄링된 작업을 수행함으로써 다수의 관측 지점의 품질관리를 일괄 수행할 수 있도록 체계를 구성하였으며, 레벨화를 적용한 단계적 품질관리가 가능하도록 하였다. 자동화를 통한 1차적 관리와 현장단에서 수동으로 실시한 2차적 관리는 생애주기와 K-water 자료 흐름의 각 지점마다의 품질을 담당할 수 있는 계기로 보고 있다.

(로컬) 각 사이트 담당자별 비주기적 수동 품질관리 수행

- CS Stand alone으로 구성된 S/W 설치를 통한 개별적 품질관리 수행 방식

(서버) 주기적 스케줄링에 따른 오결측 탐지 및 보정을 적용한 자동화 품질관리 수행

- 운영 서버단에 품질관리 환경을 설치하여 자동으로 수행하는 방식
- 검보정 알고리즘은 범용적으로 개발된 R소스로 로컬 및 서버 파트에 공통 반영하였다.
- 대상 및 적용 알고리즘과 매개변수 설정은 시스템에서 기능으로 제공할 수 있도록 프로그램 개발 적용

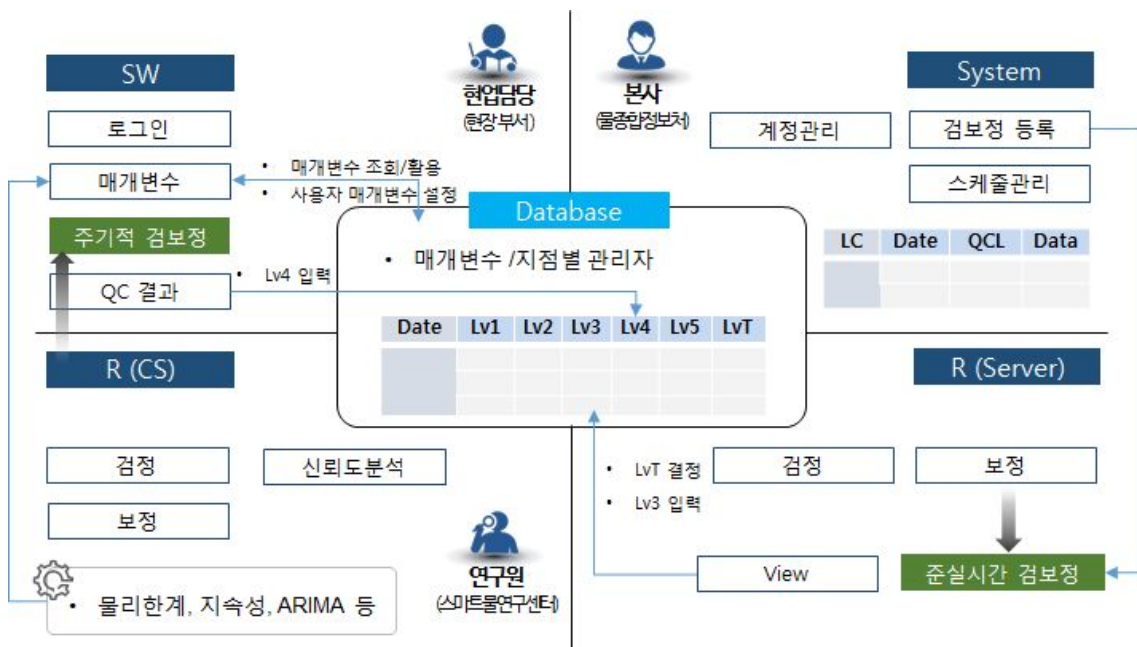


그림 12 시스템 구성도

# 제 3 장 통합물정보 품질관리 알고리즘 개선

## 3.1 시계열 데이터 오결측 검출 분석 기법

### 3.1.1 오결측 데이터 검출 분석 기법

R로 컨버팅한 알고리즘 역시 기존에 검출 알고리즘을 그대로 반영하고 현업에서 요청한 개선사항에 대해서 반영을 하였다. 본 연구에 적용된 품질관리 프로세스에서는 관측 데이터의 검정 유형을 여섯 가지 유형으로 분류하며, 검정 계층 알고리즘의 효율을 고려하여 오·결측 유형을 비교적 간단한 필수검사와 상대적으로 복잡한 선택검사 단계로 차등하여 적용할 수 있도록 구축하였다.

필수검사는 항목은 시계열 검사, 물리한계 검사, 변동한계 검사, 지속한계 검사 4가지로 구성되어 있으며, 선택검사는 통계범위 검사 및 공간일치성 검사로 아래 그림과 같이 구성되어 있다.

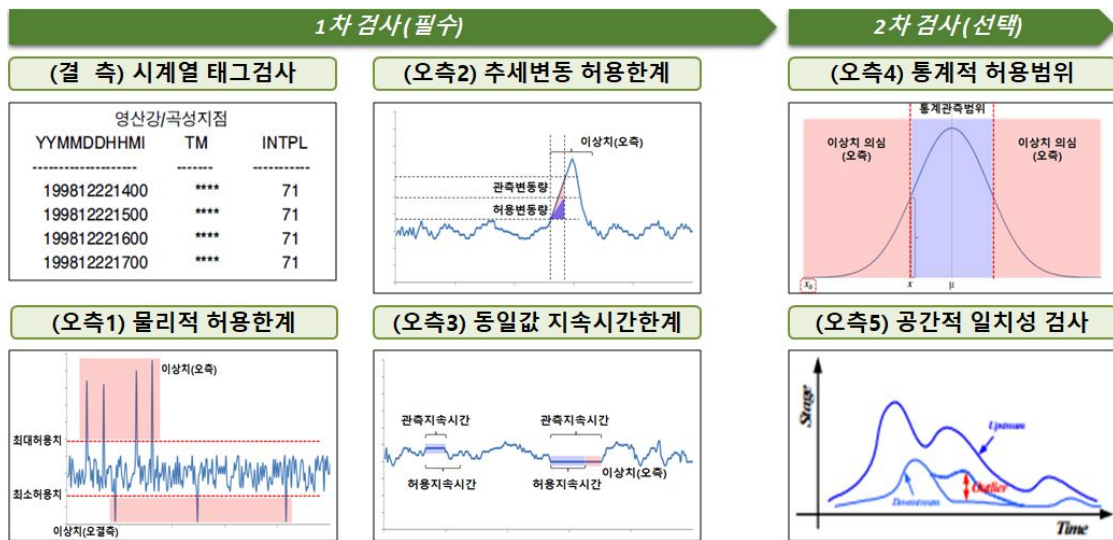
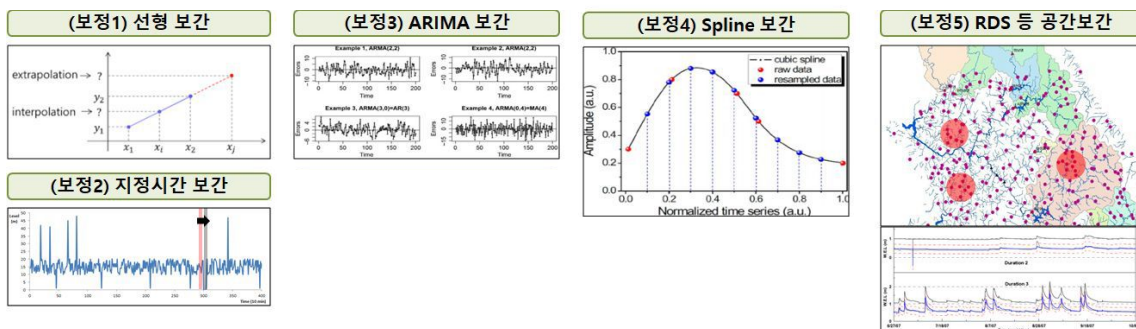


그림 13 품질관리 프로세스 검정 계층의 알고리즘 유형



1차 검사의 경우 태그검사, 물리검사, 추세검사, 지속검사가 있으며 2차 검사는 선택사항이고 통계검사, 공간검사가 있다. 강우, 수위 등 세부항목별 적용된 프로세스는 아래 그림과 같으며 강우 데이터 검정 알고리즘의 매개변수는 아래 그림과 같이 적용되어 있다. 세부적인 항목별 매개변수는 홍수통제소, 기상청, 논문 등의 레퍼런스를 참조하고 실 데이터 적용 테스트 후 K-water 매개변수를 결정하여 반영하였다.

| 1차 검사(필수) |           |  |   | 2차 검사(선택)              |  |  |
|-----------|-----------|--|---|------------------------|--|--|
|           | (결측) 태그검사 | (오측1) 물리검사   | (오측2) 추세검사  | (오측3) 지속검사             | (오측4) 통계검사   | (오측5) 공간검사   |
| 강우        | 시계열 검사    | 상한: 설정값 이상<br>하한: 0값 미만  | 경계: 품질관리 기간 내 관측 변동값들의 ( $\bar{x} + 6\sigma$ )  | 경계: 설정시간 이상<br>1) 60분  | 상한: 과거 같은 달(월) 관측 데이터들의 ( $\bar{x} + 3\sigma$ )<br>하한: 과거 같은 달(월) 관측 데이터들의 ( $\bar{x} - 3\sigma$ )   | 상한: RDS 산정값의 (+)설정배수 이상<br>하한: RDS 산정값의 (-)설정배수 이상<br>1) 280 %, 2) 20 %                  |
| 수위        | 시계열 검사    | 상한: 설정값 이상<br>하한: 설정값 이하<br>1) 관측 범위 최대/최소 한계<br>2) 기왕 최대/최소 관측        | 경계1: 설정값 이상<br>경계2: 2시간 기율기의 (+)설정배수 이상<br>경계3: 2시간 기율기의 (-)설정배수 이상<br>1) 20 cm, 2) 3배, 3) 1배 | 경계: 설정시간 이상<br>1) 24시간 | 상한: 과거 같은 달(월) 관측 데이터들의 ( $\bar{x} + 3\sigma$ )<br>하한: 과거 같은 달(월) 관측 데이터들의 ( $\bar{x} - 3\sigma$ )   | 상한: 7일간 인근 관측값 편차의 ( $\bar{x} + 3\sigma$ )<br>하한: 7일간 인근 관측값 편차의 ( $\bar{x} - 3\sigma$ ) |
| 유량        | 시계열 검사    | 상한: 설정값 이상<br>하한: 0값 이하<br>1) 시설 최대 용량<br>2) 기왕 최대/최소 관측               | 경계: 품질관리 기간 내 관측 변동값들의 ( $\bar{x} + 6\sigma$ )  | 경계: 설정시간 이상<br>1) 20분  | 상한: 7일간 같은 시(時) 관측 데이터들의 ( $\bar{x} + 3\sigma$ )<br>하한: 7일간 같은 시(時) 관측 데이터들의 ( $\bar{x} - 3\sigma$ ) | 상한: 1일간 인근 관측값 편차의 ( $\bar{x} + 3\sigma$ )<br>하한: 1일간 인근 관측값 편차의 ( $\bar{x} - 3\sigma$ ) |
| 압력        | 시계열 검사    | 상한: 설정값 이상<br>하한: 0값 이하<br>1) 동수경사선의 최고점과 최저 지반고의 편차<br>2) 기왕 최대/최소 관측 | 경계: 품질관리 기간 내 관측 변동값들의 ( $\bar{x} + 6\sigma$ )  | 경계: 설정시간 이상<br>1) 20분  | 상한: 7일간 같은 시(時) 관측 데이터들의 ( $\bar{x} + 3\sigma$ )<br>하한: 7일간 같은 시(時) 관측 데이터들의 ( $\bar{x} - 3\sigma$ ) | 상한: 1일간 인근 관측값 편차의 ( $\bar{x} + 3\sigma$ )<br>하한: 1일간 인근 관측값 편차의 ( $\bar{x} - 3\sigma$ ) |

그림 14 품질관리 프로세스 데이터 분야 및 검정 알고리즘별 적용 현황

| 1차 검사(필수)   |  | 2차 검사(선택)   |
|---|--|---|
| (결측) 시계열 태그검사   | (오측2) 추세변동 허용한계  | (오측4) 통계적 허용범위  |
| (K-water)<br>• 계측이 되지 않는 경우 (시계열 누락)<br><br>(Weather Planet <sup>1)</sup> )<br>• 'Null' 값 기록  | (K-water)<br>• 누계 관측값이 감소하는 경우<br><br>(Shulski et al., 2014)<br>• 단위시간 변화량의 통계 범위 이탈 ( $\bar{x} + 6\sigma$ ) | (기상청)<br>• 당해 월간, 동월 연간 데이터 비교<br>• 월평균 강수량의 범위 이탈<br>$(\bar{x} \pm 2.576\sigma / \bar{x} \pm 3.291\sigma, \sigma = 1.4826 \frac{P_{78} - P_{22}}{2})$<br><br>(Hubbard et al., 2005)<br>• 평균관측치의 통계 범위 이탈 ( $\bar{x} \pm 3\sigma$ )<br>(무강우 제외)           |
| (오측1) 물리적 허용한계  | (오측3) 동일값 지속시간한계   | (오측5) 공간적 일치성 검사  |
| (K-water)<br>• 10분 10 / 20 mm 이상<br>[실시간 품질관리기준: 10 mm]<br><br>(홍수통제소)<br>• 1시간 100 / 150 mm 이상<br>• 기왕 최대, 최소 관측값<br><br>(기상청)<br>• 10분 100 mm, 1시간 300 mm, 1일 1,000 mm 이상 | (금회 제한)<br>• 1시간 이상 동일 강우량 관측<br>(무강우 제외)  | (K-water)<br>• 인근관측소 RDS 값의 20% 미만 / 280% 초과<br><br>(홍수통제소)<br>• 인근관측소 RDS 값의 50% 미만 / 200% 초과<br><br>(기상청)<br>• 인근관측소 RI (공간강수확증, AMeDAS) 비교<br>• 인근관측소 비모수 공간성 (Madsen-Allerup)<br><br>(Weather Planet)<br>• 인근관측소 강우량의 범위 이탈 ( $\bar{x} \pm 3\sigma$ ) |

그림 15 강우 데이터 검정 알고리즘의 매개변수 조사 및 적용 현황

그리고 수위데이터 검출 알고리즘은 시계열 태그검사, 추세변동 허용한계검사, 통계적 허용범위의 매개변수가 있으며 상세내용은 아래 그림과 같다.

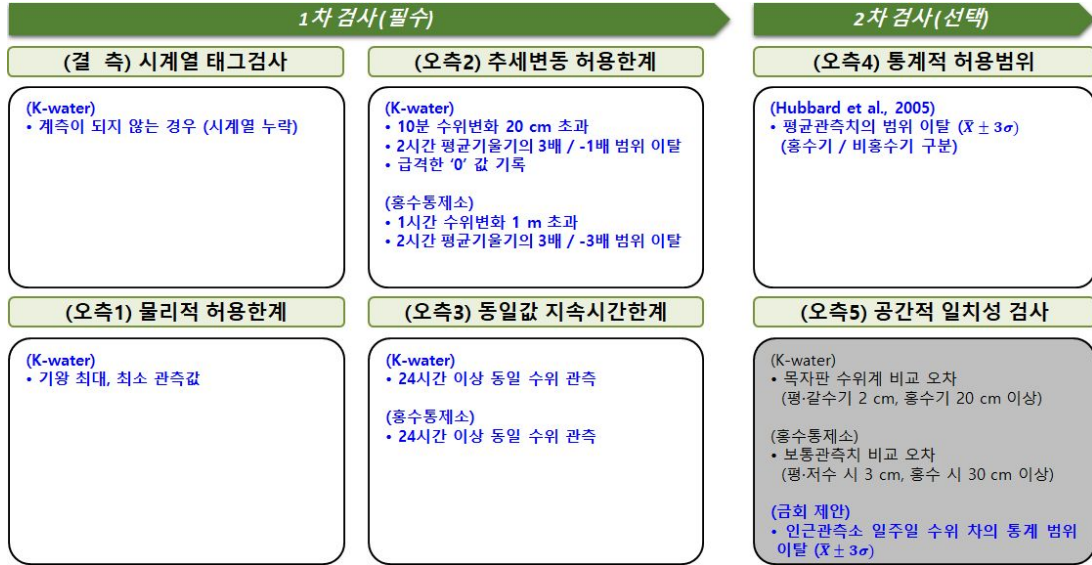


그림 16 수위 데이터 검정 알고리즘의 매개변수 조사 및 적용 현황

또한 수도 데이터 검출 알고리즘도 시계열 태그검사, 추세변동 허용한계검사, 통계적 허용범위 검사가 있으며 상세내용은 아래 그림과 같음을 알 수 있다.

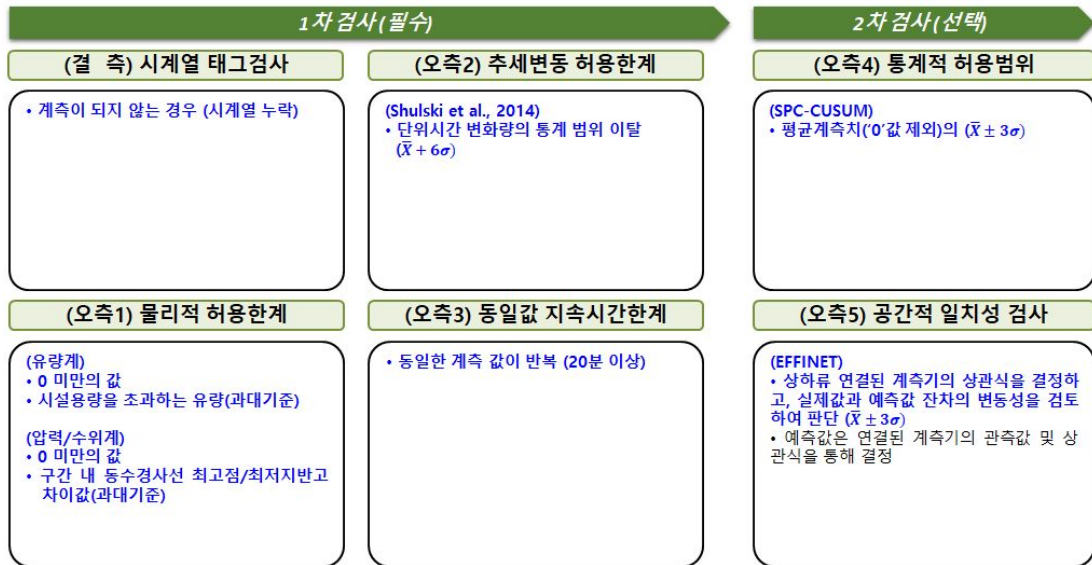


그림 17 수도 데이터 검정 알고리즘의 매개변수 조사 및 적용 현황

### 3.2 알고리즘 타분야 적용

범용적인 알고리즘은 물정보 분석DB에 시범적용을 통해 검증을 하고 있으며 이에 더하여 데이터체계개선전담반에서 요청한 water-NET 활용 중인 수도 데이터를 기반으로 적용성을 검토하였다. 이것은 수도(수질) 실 데이터를 기반으로 기 연구 성과인 데이터 품질관리 알고리즘 검·보정 테스트를 통한 적용성 검토로 분야 및 항목의 확대 적용의 의미를 가지고 OA단 계측 데이터의 주기적(준실시간) QC 수행을 위한 범용적 프로세스 및 S/W 개발 완료된 알고리즘을 파주 월릉계통 데이터에 적용하는 것을 연구하였다.

#### 3.2.1 타분야 적용성 검토

현재 K-water 수도 분야 1분 간격으로 측정된 데이터는 실시간 운영관리 시스템인 water-NET을 통해 모니터링 및 관리되고 있다. 품질관리 또한 해당 시스템내 적용되어 있지만 결측에 대한 구분과 단순 보정기법만 적용되고 있는 현실이다. 과거 오랜 기간 전부터 발생된 결측 및 누락된 데이터의 처리가 필요하며 수리, 수질 분야에 대한 보완이 필요함에 따라 금회 파주시 월릉배수지의 수리(유량/압력/수위), 수질(탁도/잔류염소) 항목 대상으로 1년간 자료를 바탕으로 범용 품질관리 프로세스와 알고리즘을 적용하여 검증해 보았다. 대상은 항목별 데이터셋 시작일과 종료일 적용은 1분 데이터이고 매개변수는 제공된 물리적 한계 기준치를 참조하였으며 기타 매개변수 항목은 범용적으로 적용된 기본값으로 설정하였다. 시범 적용 전체 결과인 아래 표를 참고하면 검정 알고리즘을 통해 검출된 결측 비율이 일반적인 상황에 비해 높은 비율로 발생하는 원인은 누락 값 범위가 상대적으로 높은 상황을 반영하고 있다.

표 8 시범 적용 결과

| 구 분  | 결측치     |       | 오측의심   |       |
|------|---------|-------|--------|-------|
|      | (개)     | (%)   | (개)    | (%)   |
| 유량   | 25,315  | 4.82  | 47,137 | 8.97  |
| 압력   | 31,859  | 15.42 | 4,457  | 2.16  |
| 잔류염소 | 147,480 | 29.83 | 15,190 | 3.07  |
| 탁도   | 25,315  | 4.82  | 60,810 | 11.57 |
| 온도   | 24,345  | 4.67  | 81,213 | 15.57 |
| pH   | 24,351  | 4.67  | 84,967 | 16.28 |

세부적인 항목별 결과로는 첫 번째 압력 데이터 관련하여 월릉1-1, 파주 문산(정) WR1-1 초음파 압력을 분석하였으며 분석 기간은 2018.08.01. 00:00 ~ 2019.07.10. 06:50으로 설정하였다. 적용된 매개변수는 아래 표와 같다.

표 9 파주 월릉배수지 매개변수

| 변수명   | 내용               | 적용값   | 기타  |
|-------|------------------|-------|---|
| set1L | 물리검사 물리한계 하한 설정값 | 0     | 단위시간 관측값의 하한                                  |
| set1U | 물리검사 물리한계 상한 설정값 | 15    | 단위시간 관측값의 상한                                  |
| set2P | 추세검사 단위시간 내 변동한계 | 6     | 변화량평균( $\bar{\chi}$ ) +<br>6×표준편차( $\sigma$ ) |
| set3  | 지속검사 한계 설정값      | 20    | 지속기간수 (1분 × 20)                               |
| set4  | 통계검사 오차범위 설정값    | 3     | 평균관측치( $\bar{\chi}$ ) ±<br>3×표준편차( $\sigma$ ) |
| set5  | 공간검사 오차범위 설정값    | 3     | 직전 1일과 현재 압력차의<br>오차평균 ± 3×표준편차( $\sigma$ )   |
| P,D,Q | ARIMA 보정 차수 설정값  | 2,1,2 |   |

"Loss Sensitive = 29.834261, count = 147480"  
 "Physical Sensitive = 0.036615, count = 181"  
 "Trend Sensitive = 0.411263, count = 2033"  
 "Persistence Sensitive = 0.392652, count = 1941"  
 "Statistics Sensitive = 2.612824, count = 12916"  
 "Spatial Sensitive = 0.000000, count = 0"  
 "Data Result Sensitive = 33.096245, count = 163605"  
 "-----"  
 "Data <Loss> Rate = 29.834261, count = 147480"  
 "Data <Miss> Rate = 0.189145, count = 935"  
 "Data <Miss doubt> Rate = 3.072840, count = 15190"  
 "Data <Normality> Rate = 66.903755, count = 330726"

(탁도) 월릉배수지 탁도의 경우 분석 기간은 2018.08.01. 00:00 ~ 2019.08.01. 00:00이고 매개변수는 아래 표와 같다.

표 10 파주 월릉배수지 탁도 매개변수

| 변수명   | 내용               | 적용값 | 기타           |
|-------|------------------|-----|--------------|
| set1L | 물리검사 물리한계 하한 설정값 | 0   | 단위시간 관측값의 하한 |
| set1U | 물리검사 물리한계 상한 설정값 | 10  | 단위시간 관측값의 상한 |

"Loss Sensitive = 4.816391, count = 25315"  
 "Physical Sensitive = 0.871003, count = 4578"  
 "Trend Sensitive = 0.022831, count = 120"  
 "Persistence Sensitive = 5.171033, count = 27179"  
 "Statistics Sensitive = 7.820000, count = 41102"  
 "Spatial Sensitive = 0.000000, count = 0"

"Data Result Sensitive = 17.542585, count = 92204"  
 "-----"  
 "Data <Loss> Rate = 4.816391, count = 25315"  
 "Data <Miss> Rate = 1.156581, count = 6079"  
 "Data <Miss doubt> Rate = 11.569613, count = 60810"  
 "Data <Normality> Rate = 82.457415, count = 433397"

(pH) 월롱배수지 pH의 경우 분석 기간은 2018.08.03. 14:27 ~ 2019.08.01. 00:00이고 매  
 개변수는 아래 표와 같다.

표 11 파주 월롱배수지 pH 매개변수

| 변수명   | 내용               | 적용값 | 기타           |
|-------|------------------|-----|--------------|
| set1L | 물리검사 물리한계 하한 설정값 | 0   | 단위시간 관측값의 하한 |
| set1U | 물리검사 물리한계 상한 설정값 | 14  | 단위시간 관측값의 상한 |

"Loss Sensitive = 4.666248, count = 24351"  
 "Physical Sensitive = 0.891629, count = 4653"  
 "Trend Sensitive = 0.017246, count = 90"  
 "Persistence Sensitive = 4.382452, count = 22870"  
 "Statistics Sensitive = 14.559628, count = 75980"  
 "Spatial Sensitive = 0.000000, count = 0"  
 "Data Result Sensitive = 22.720723, count = 118569"  
 "-----"  
 "Data <Loss> Rate = 4.666248, count = 24351"  
 "Data <Miss> Rate = 1.772718, count = 9251"  
 "Data <Miss doubt> Rate = 16.281757, count = 84967"  
 "Data <Normality> Rate = 77.279277, count = 403285"

(잔류염소) 월롱배수지 잔류염소의 경우 분석 기간은 2018.08.01. 00:00 ~ 2019.08.01.  
 00:00이고 매개변수는 아래 표와 같다.

표 12 파주 월롱배수지 잔류염소 매개변수

| 변수명   | 내용               | 적용값 | 기타           |
|-------|------------------|-----|--------------|
| set1L | 물리검사 물리한계 하한 설정값 | 0   | 단위시간 관측값의 하한 |
| set1U | 물리검사 물리한계 상한 설정값 | 1   | 단위시간 관측값의 상한 |

"Loss Sensitive = 4.816391, count = 25315"  
 "Physical Sensitive = 0.885463, count = 4654"  
 "Trend Sensitive = 0.075342, count = 396"

"Persistence Sensitive = 0.351027, count = 1845"  
 "Statistics Sensitive = 9.496938, count = 49916"  
 "Spatial Sensitive = 0.000000, count = 0"  
 "Data Result Sensitive = 14.700885, count = 77268"  
 "-----"  
 "Data <Loss> Rate = 4.816391, count = 25315"  
 "Data <Miss> Rate = 0.916284, count = 4816"  
 "Data <Miss doubt> Rate = 8.968210, count = 47137"  
 "Data <Normality> Rate = 85.299115, count = 448333"

(온도) 월롱배수지 온도의 경우 분석 기간은 2018.08.03. 14:29 ~ 2019.07.31. 20:00이고 매개변수는 아래 표와 같다.

표 13 파주 월롱배수지 온도 매개변수

| 변수명   | 내용               | 적용값 | 기타           |
|-------|------------------|-----|--------------|
| set1L | 물리검사 물리한계 하한 설정값 | 0   | 단위시간 관측값의 하한 |
| set1U | 물리검사 물리한계 상한 설정값 | 40  | 단위시간 관측값의 상한 |

"Loss Sensitive = 4.667262, count = 24345"  
 "Physical Sensitive = 1.636082, count = 8534"  
 "Trend Sensitive = 0.019363, count = 101"  
 "Persistence Sensitive = 4.637738, count = 24191"  
 "Statistics Sensitive = 13.329065, count = 69526"  
 "Spatial Sensitive = 0.000000, count = 0"  
 "Data Result Sensitive = 22.061801, count = 115077"  
 "-----"  
 "Data <Loss> Rate = 4.667262, count = 24345"  
 "Data <Miss> Rate = 1.824920, count = 9519"  
 "Data <Miss doubt> Rate = 15.569619, count = 81213"  
 "Data <Normality> Rate = 77.938199, count = 406535"

(유출순시) 월롱배수지 유출순시 경우 분석 기간은 2018.08.01. 00:50 ~ 2019.07.31. 12:25이고 매개변수는 아래 표와 같다.

표 14 파주 월롱배수지 온도 매개변수

| 변수명   | 내용               | 적용값  | 기타           |
|-------|------------------|------|--------------|
| set1L | 물리검사 물리한계 하한 설정값 | 0    | 단위시간 관측값의 하한 |
| set1U | 물리검사 물리한계 상한 설정값 | 2300 | 단위시간 관측값의 상한 |

```

"Loss Sensitive = 7.585509, count = 39813"
"Physical Sensitive = 1.831931, count = 9615"
"Trend Sensitive = 0.300654, count = 1578"
"Persistence Sensitive = 1.049621, count = 5509"
"Statistics Sensitive = 3.108281, count = 16314"
"Spatial Sensitive = 0.000000, count = 0"
>Data Result Sensitive = 11.666438, count = 61232"
"-----"
>Data <Loss> Rate = 7.585509, count = 39813"
>Data <Miss> Rate = 1.900140, count = 9973"
>Data <Miss doubt> Rate = 2.180789, count = 11446"
>Data <Normality> Rate = 88.333562, count = 463624"

```

월릉배수지의 분석에 적용된 품질관리 적용 코드의 디렉토리 구조와 코드 진행 프로세스는 아래 그림으로 확인할 수 있다. 시작점 호출 부분의 입력 변수들만 조정하여 실행하면 메인 프로세스에서 검·보정 및 결과 도출까지 실행하는 구조로 되어 있고 각 알고리즘은 모듈로 구성되어 있어 실행 여부를 사용자가 결정할 수 있도록 구성되어 있다. 결과 값은 그림과 같이 엑셀로 출력이 되고 있다.

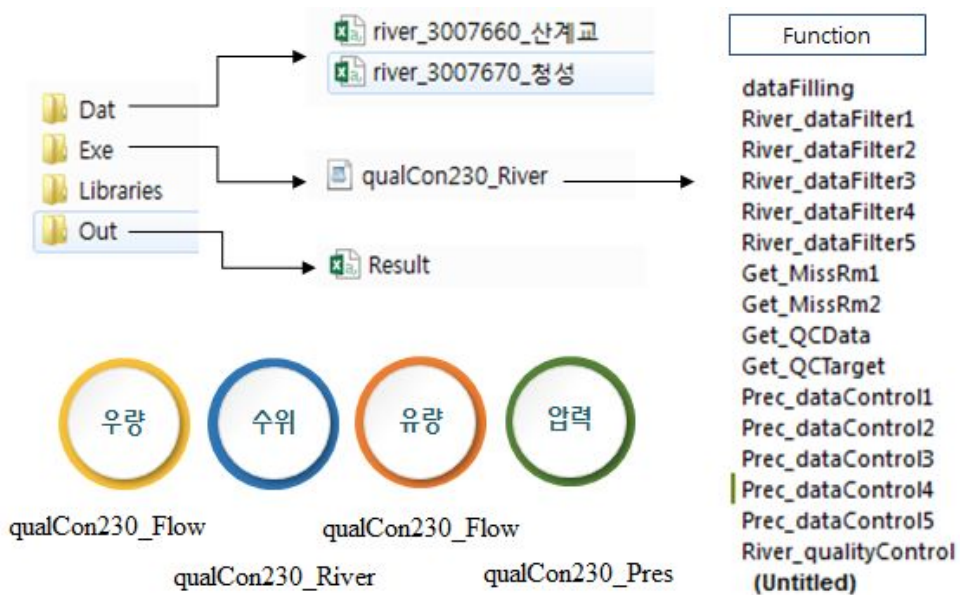


그림 18 품질관리 디렉토리 구조

```

setwd('D:/R/QC/Exe')

dataFilling <- function(unitTime, unitRawData, unitRawTime, fpart, ftime) {}

Prec_dataFilter1 <- function(unitFilledData, searchType1Mset) {}
Prec_dataFilter2 <- function(unitFilledData, searchType2Pset, searchType1Mset) {}
Prec_dataFilter3 <- function(unitFilledData, searchType3set) {}
Prec_dataFilter4 <- function(unitFilledTime, unitFilledData, dataS, dataP, searchType4set)
Prec_dataFilter5 <- function(unitFilledData, searchType5Lset, searchType5Uset, searchTypeData, searchTypeDist) {}

Prec_dataControl1 <- function(unitFilledData, QCTarget) {}
Prec_dataControl2 <- function(unitFilledData, QCTarget, unitFilledTime) {}
Prec_dataControl3 <- function(unitTime, unitFilledData, QCTarget, P, D, Q) {}
Prec_dataControl4 <- function(unitTime, unitFilledData, QCTarget) {}
Prec_dataControl5 <- function(unitFilledData, QCTarget, varargData, varargDist) {}

Pre_qualityControl <- function(unitTime, database, Tstart, Tend, set1M, set2P, set3, set4, set5L, set5U, Xorg, Yorg, varargin, Rstart, Rend,
setAP, setAD, setAQ, setOutNM, cKey1, cKey2, cKey3, cKey4, cKey5, cKeyR) {}

refdata <- c('prec_8000701.csv', 'prec_3003431.csv')
refx <- c('127.665', '127.661')
refy <- c('36.009', '35.893')

varargin1 <- cbind(refdata, refx, refy)
Pre_qualityControl(10, 'prec_3003420.csv', '201602010010', '201612312250', 10, 6, 6, 3, 0.2, 2.8, 127.659, 35.929, varargin1, '201607010010',
'201607312400', 2, 1, 2, 'Prec_Result', 1, 1, 1, 1, 0)

```

③ 결측

④ 검정

⑤ 보정

② 메인클래스

참조점 설정

① 시작점(호출)

그림 19 R 코드 구조

R코드 개별 적용을 통한 수질 항목을 적용한 결과들을 살펴보면 기존 알고리즘의 매개 변수 조정을 통한 오결측 판별의 능력은 탁월한 것으로 판단이 되고 보정값 또한 유의미한 결과를 도출하였음을 알 수 있지만, 데이터를 통한 사고 인지와 같은 특정 목적을 달성하기 위한 내용은 추가적인 연구를 통해 새로운 결과를 도출해야 한다.

### 3.2.2 알고리즘 타분야 적용 및 확대 방안 제시

품질관리 알고리즘 타분야 확대방안은 분야별 핵심품질항목(CTQ) 선정을 통해 품질관리 대상 범위를 선정하고, 데이터 중요도에 따라 우선순위 결정 추진하는 것이 필요하고 \* 핵심품질항목(Critical To Quality) : 주문(수위, 우량), 수도(유량, 압력) 이다. 핵심 분야 이외의 타 분야(지하수, 발전, 수질, 설비)와 항목 확대에 대한 데이터 분석 및 오·결측 적용방안 제시가 필요하다.

추진계획으로는 데이터 생애주기 전반에 대한 품질관리 프로세스(QA) 정립과 품질수준 향상을 위한 품질활동(QC), 품질관리 문화정착을 함께 추진하고 데이터의 표준관리 및 구조에 대한 사항은 데이터 표준화를 통해 별도 추진하는 것이 필요하다 간략히 용어에 대해 정의를 하면 품질보증(QA, Quality Assurance)은 산출물이나 서비스의 품질에 대한 요구사항이 충족되도록 하는 계획되고 체계화된 모든 활동을 의미하고 품질검사(QC, Quality Control)는 품질 요구사항을 충족하는데 사용되는 현업적인 모든 기술과 활동으로 합격 자료만 사용자에게 전달되도록 하는 절차를 의미한다. 또한 데이터 품질 수준을 보증하는 품질관리 프로세스 정립에 있어서 지침은 품질관리 목표, 적용 범위 및 대상, 품질진단 및 개선 활동을 의미하고 연보 등 데이터 발간에 대한 품질 수준 공인 절차를 마련하는 것이 필요하다. 또한, 품질진단 결과 및 현업 요구사항을 반영한 연간계획 수립을 하는 품질관리 프로세스 정립이 필요한 실정이다. 다음으로는 데이터 품질향상을 위한 기술·시스템 등 활동 기반 마련을 위해서 데이터 항목별 검·보정 알고리즘 검증·보완 및 신규 개발기술이 필요하고 지침·기준을 반영한 단계별 데이터 품질관리 시스템 개발, 품질관리 현황 모니터링 및 개선 활동 이행사항 점검하는 진단이 필요하다.

마지막으로 품질관리 역량 강화 및 데이터 품질 인식도 제고를 위해 데이터진흥원 품질 인증, 행안부 공공데이터 수준 평가를 위한 인증이 필요하고 품질관리 전문가 교육 개설 및 품질 우수부서 포상을 통해 품질관리 문화를 정착하고 이를 홍보하기 위해 연간 품질관리 우수부서·담당자 포상 및 우수사례를 홍보하는 것이 필요하다. 이를 정리하여 요약하면 데이터 품질관리의 최종목표는 데이터 기반 혁신을 통한 미래 50년 비전 및 전사 전략 달성을 위하여 전사 데이터 품질관리 체계 확립 및 실행, 문화 기반 마련

☞ 체계 확립(지침, 공인), 품질 활동(기술, 시스템), 문화 정착(인증, 교육)을 마련하는 것이 필요하다. 또한 품질관리 핵심 품질항목(CTQ) 선정을 통해 품질관리 대상 범위를 선정하고, 데이터 중요도에 따라 우선순위 결정, 단계별 추진하는 것이 필요하다. 핵심 품질항목(Critical To Quality)의 정의는 기업의 경영에 중요한 영향을 미치는 품질관리 대상 정보 항목으로 정의된다. 수도 분야 핵심 품질 항목을 예를 들어 표현하면 표와 같이 나타낼 수 있다.

표 15 데이터 품질관리 담당자 직무별 검·보정 대상 데이터(예시)

| 구분                          | 담당직무           | 검·보정 데이터 항목   |           | 주기  |
|-----------------------------|----------------|---|-----------|-----|
| 공공데이터<br>개방포털               | 수질관리           | 수질 - 정수지 유출 탁도, 잔류염소, pH  |           | 시간  |
|                             | 시설관리<br>(설비관리) | 유량 - 취수장, 정수장, 가압장 유출(순시, 적산차)<br>수위 - 정수지, 배수지 수위<br>압력 - 취수장, 정수장, 가압장 유출 |           |     |
| 수돗물 수질<br>목표 달성도<br>(적정성평가) | 수질관리           | 생활  | 여과지 통합 탁도 | 15분 |
|                             |                |   | 여과지 지별 탁도 | 시간  |
|                             | 공업             | 정수지 잔류염소  | 시간        |     |
|                             |                | 정수지 탁도<br>정수지 pH  | 시간        |     |
| 표준<br>운영일보<br>(적정성평가)       | 설비관리           | 정수장 공정별 유량,<br>약품투입량, 염소투입량, 펌프가동시간, 전력량                                    |           | 시간  |
|                             | 수질관리           | 공정별 수질(탁도, 잔류염소, pH, COD, SS)   |           |     |
| 관망<br>운영관리                  | 시설관리           | 계통별 메인 유량(순시, 적산차), 압력  |           | 시간  |
|                             | 수질관리           | 급배수 계통 탁도, 잔류염소   |           |     |

수자원분야 현황은 실시간 수문자료 품질관리 기준(별표 3. 주요 자료처리 항목)에 수문자료 처리 대상 데이터 항목을 제시하며 핵심품질항목을 예를 들어 표현하면 표와 같이 나타낼 수 있다.

표 16 데이터 품질관리 계측항목 및 처리항목(예시)

| 항 목 | 계측항목          | 처리항목                 |
|-----|---------------|----------------------|
| 수위  | 본댐 및 보 계측수위   | 평균수위                 |
|     |               | 저수량                  |
|     |               | 저수율                  |
|     |               | 공용량                  |
|     |               | 계획홍수위/상시만수위/제한수위 공용량 |
|     |               | 계획홍수위/상시만수위 소요강우     |
|     | 방수로 수위        | 정시수위                 |
|     | 하천 수위         | 유량                   |
| 우량  | 우량값           | 관측소 시간우량             |
|     |               | 관측소 금회누계             |
|     |               | 관측소 10분, 30분 강우      |
|     |               | 유역평균 시간우량            |
|     |               | 유역평균 금회/일 누계         |
| 여수로 | 수문개도          | 유역평균 실시간/년 누계 우량     |
|     |               | 열린 수문수               |
|     |               | 수문별/전체수문 방류량         |
| 기타  | 기타방류(취수량) 계측치 | 시간평균/일평균 수문방류량       |
|     |               | 기타(취수량) 방류량          |
| 발전  | 발전전력          | 시간평균/일평균 기타방류량       |
|     |               | 발전상태(발전시스템 자료입력)     |

| 항 목 | 계측항목        | 처리항목                       |
|-----|-------------|----------------------------|
|     | (자료생산:GIOS) | 호기별 발전방류량(발전시스템 자료입력)      |
|     |             | 전체 발전방류량(발전시스템 자료입력)       |
|     |             | 시간평균 발전방류량(발전시스템 자료입력)     |
|     |             | 일평균 발전방류량(발전시스템 자료입력)      |
| 방류량 |             | 전체/시간평균/일평균 방류량            |
| 유입량 |             | 1분/정시(10분, 30분, 1시간)/일 유입량 |

물분야 데이터 관리 선도를 통한 국가 물정보 Hub 구축을 위해 목표와 추진방향, 실행 과제 등은 아래 표와 같이 알 수 있으며 이전 2018년도부터 추진된 단계별 추진방안을 참고하고 장기적인 계획도 수립이 향후에 필요하다.

표 17 물분야 데이터 관리 국가 물정보 Hub 구축 목표와 실행과제

|       |   |  |   |
|-------|---|--|---|
| 목 표   | 신뢰도 높은 데이터 생산을 통한 물관리 효율화 및 新 가치 창출   |  |   |
| 추진 방향 | 프로세스(QA) 정립   | 품질활동(QC) 지원  | 품질문화 정착   |
| 실행 과제 | <ul style="list-style-type: none"> <li>① 품질관리 지침</li> <li>② 데이터 공인 절차</li> <li>③ 선순환 체계 마련</li> </ul> | <ul style="list-style-type: none"> <li>④ 기술 개발</li> <li>⑤ 시스템 구축</li> <li>⑥ 품질 진단</li> </ul> | <ul style="list-style-type: none"> <li>⑦ 외부 인증 획득</li> <li>⑧ 역량 강화 교육</li> <li>⑨ 포상 및 홍보</li> </ul> |

표 18 단계별 추진방안(예시)

| 구분    | 2018년   | 2019년  | 2020년  |
|-------|---|--|--|
| QA    | <ul style="list-style-type: none"> <li>- 품질관리 지침</li> <li>- 데이터 공인 절차</li> </ul>                        | <ul style="list-style-type: none"> <li>- 평가/요구사항 검토</li> <li>- 제도 개선</li> </ul>                        | <ul style="list-style-type: none"> <li>- 평가/요구사항 검토</li> <li>- 제도 개선</li> </ul>                        |
| QC    | <ul style="list-style-type: none"> <li>- 알고리즘 보완·개발</li> <li>- 시스템 구축 방안</li> <li>- 품질진단 컨설팅</li> </ul> | <ul style="list-style-type: none"> <li>- 알고리즘 고도화</li> <li>- 시스템 구축</li> <li>- 진단 수행</li> </ul>        | <ul style="list-style-type: none"> <li>- 시스템 고도화</li> <li>- 진단 수행</li> </ul>                           |
| 문화 정착 | <ul style="list-style-type: none"> <li>- 외부 인증 준비</li> <li>- 공감대 형성 워크숍</li> </ul>                      | <ul style="list-style-type: none"> <li>- 외부 인증 추진</li> <li>- 역량 강화 교육</li> <li>- 우수부서 포상·홍보</li> </ul> | <ul style="list-style-type: none"> <li>- 외부 인증 추진</li> <li>- 역량 강화 교육</li> <li>- 우수부서 포상·홍보</li> </ul> |

129개 태그에 대해 1년 치 데이터를 히스토리안에서 추출하고, 이를 이용하여 각 태그에 대한 분위수를 추출한다. 동일 패턴이 존재할 시 중복을 제거하며 이를 QUAN\_LIB\_TB 테이블에 저장한다. 위의 과정을 모든 태그에 적용하여 분위수를 도출하면 아래와 같다.

향후 확대방안시 중점 추진사항에 대해서 검토한 결과 분야별 대표적인 품질오류 사례(수자원 1건, 수도 1건) 개선을 위한 품질관리 알고리즘 개발 및 시범 시스템 개발로 확대 및 검증이 필요하며 수자원 분야 댐유입량에 대해 예를 들어보면 다음과 같다. 수자원의 경우 댐 유입량은 저수량 변화에 평균 방류량을 합산하여 산정하게 되는데 넓은 댐의 저수면적과 1cm 단위의 측정으로 인해 저수량 변화가 크게 발생하여 연속적으로 큰 진폭이 발생하는 경우가 빈번하며 때로는 음의(-) 값이 산정이 되는데 자료동화 기법 중 시계열 자료에 적용이 용이하고 강우-유출 현상과 같은 비선형 문제에 적용할 수 있는 칼만필터(Kalman Filter) 기법과 강우-유출 모형의 결합 검토가 필요하다. 아래 그림과 같이 예시로 알 수 있다.

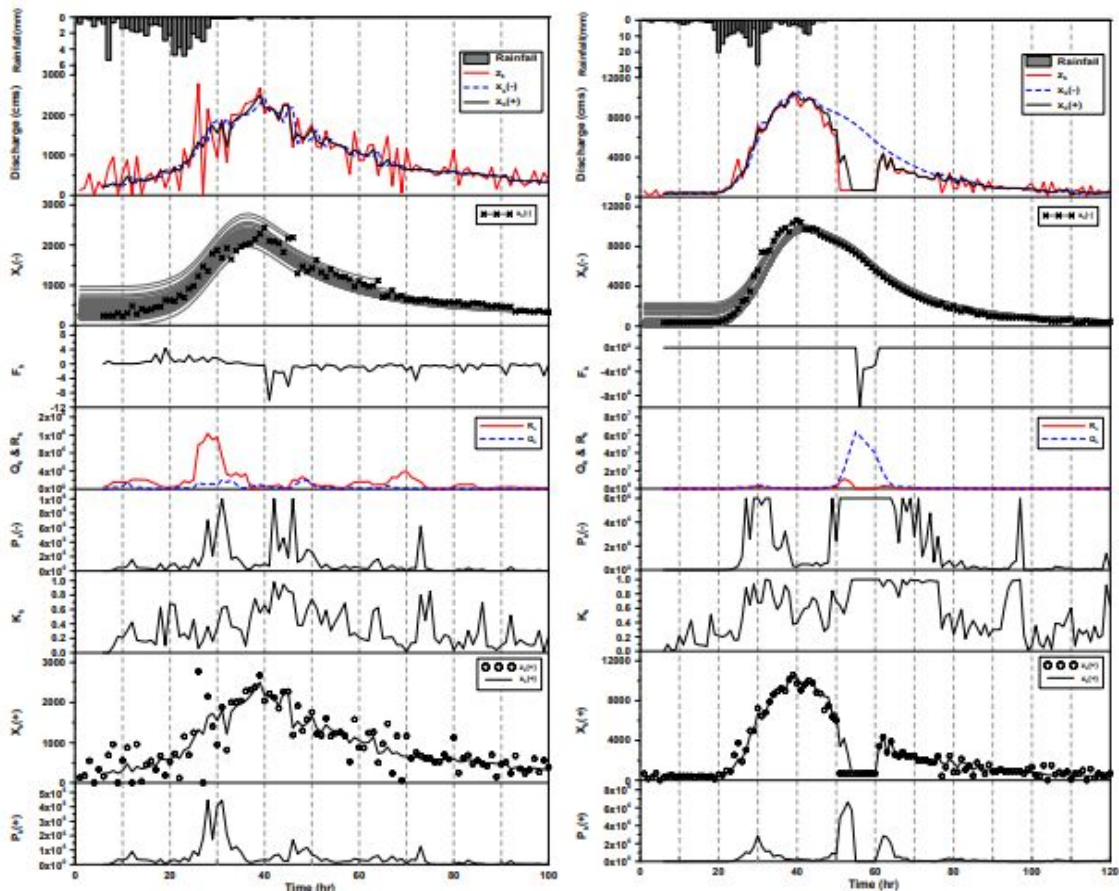
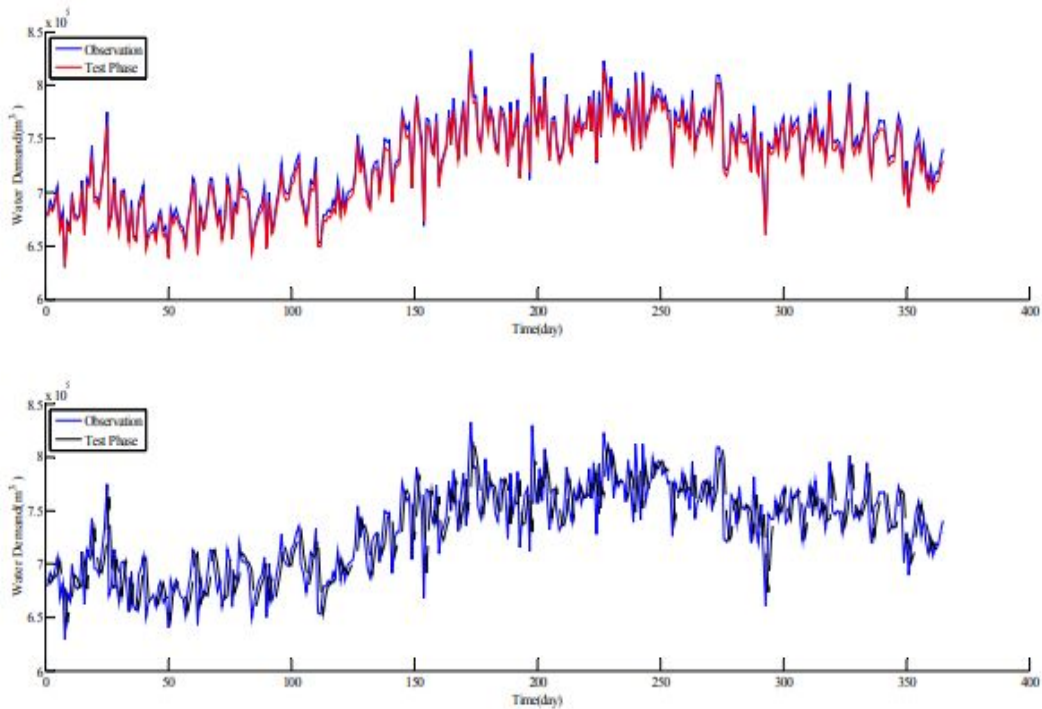


그림 20 확장 칼만 필터를 이용한 유량자료의 실시간 품질향상(2012, 유철상)

또한 수도분야에서 유량을 예를 들면 유량 데이터는 수도운영의 누수관리, 고객 수도요금 부과, 물공급 안정성 모니터링 등에 사용되는 중요 데이터이나 통신 및 기계적 오류로 인해 오·결측의 발생 빈도가 높다. 그러나 자연현상이 아닌 펌프, 수용가 사용량 변동, 누수발생 등 복합적이고 인위적인 요소의 영향을 받아 데이터 보정이 어려움이 있다 이를 보완하기 위해 머신러닝 기반의 예측 알고리즘을 개발하고 이를 활용하여 장기간 결측 데이터를 보정을 검토해야 하며 아래 그림과 같이 예시로 알 수 있다.



**Fig. 11. Improvement of Self Prediction Problem using the Proposed W SVM Model. The upper Panel Represents the W SVM Prediction Result and the Lower Panel Indicates the ARIMA Prediction Result**

그림 21 Wavelet Transform 방법과 SVM 모형을 활용한 상수도 수요량 예측기법 개발(2012, 권현한)

이상은 OA단에서의 품질관리에 대해서 검토하였으며 다음은 FA에서 품질관리에 대해서 고찰하고자 한다. FA단에서의 품질관리는 수동으로 반복적으로 진행되는 단순한 검·보정 사항에 대한 자동화 알고리즘 개발 및 시스템화 방안 마련이 필요하며 예를 들어 지하수 데이터는 별도의 품질관리 시스템이 없고 담당자가 수작업으로 데이터 보정 시행하고 있으며 데이터 품질평가 체계에 있어서는 현재 진행중인 품질관리의 현황 평가 및 개발된 알고리즘 적용에 따른 품질 향상을 검증하기 위한 평가방안 마련 필요하다. 또한 국내외 사례 및 분야별 데이터 특성을 반영한 데이터품질지수(DQI)를 개발하여 지속적인 품질 모니터링 및 진단 체계를 마련하는 것이 필요하다. DQI의 정의는 데이터품질지수(DQI, Data Quality Indicate)를 의미하며 데이터가 사용자의 목적에 맞게 저장되고 관리되고 있는지를 측정하고자 기준을 유형화하여 정의하고 있다.

참고적으로 품질관리 관련 국가와 K-water의 자료에 대해 설명하고자 한다. 국가에서는 국가 표준(참조 표준, 참조 데이터, KOLAS)과 같은 구조적 체계가 없어 목적, 항목·방법 등에 따라 기관별 품질관리 수행하고 있으며 이는 국가 참조표준 및 KOLAS 체계의 규제 범위(국가표준기본법)에 근거를 두고 데이터 생산 또는 생산기기의 검교정 기관에 대한 인정·인증 방법 및 절차, 데이터 생산 기관의 품질관리 조직, 절차서 등의 수립과 유지관리 방법, 생산된 데이터의 공인(인정)을 위한 현장 점검 및 내부 평가 절차, 방법 등을 정의하고 있다. 그러나 수문 데이터의 생산·관리·발간 전반에 대한 국가 체계

미흡하고 현 수문자료 국가 공인체계에서는 생산 및 품질관리 절차에 대한 구체적 검증·확인 없이, 수문자료공인심의위원회의 심의를 거쳐 공인 시행하고 있는 실정이다. 주요 법·규정·기준·매뉴얼 현황은 아래 표와 같으며 행안부는 공공데이터를 중심으로 기관별 데이터 품질관리 수준을 평가('18:시범, '19:본격시행)를 시행중이다.

표 19 주요 법·규정·기준·매뉴얼 현황

| 구분          | 국토부                         | K-water                                 | 기상청                                    | 환경부                                   | 기타  |
|-------------|-----------------------------|---|--|---------------------------------------|---|
| 법           | ·하천법                        | -                                       | ·기상법<br>·기상관측<br>표준화법                  | ·수질 및 수<br>생태계<br>보전에<br>관한 법률        | ·국가표준<br>기본법                              |
| 규정          | ·수문조사<br>업무규정               | -                                       | ·기상관측<br>업무규정<br>·기상관측<br>표준화<br>업무 규정 | ·환경시험<br>검사기관<br>정도관리<br>운영 규정        | -   |
| 기준          | ·공인 기준<br>·수문 품질<br>관리 기준   | -                                       | -                                      | ·수질오염<br>공정시험<br>방법                   | ·KS 기준<br>(기술표준원)                         |
| 지침<br>(매뉴얼) | ·수문관측<br>매뉴얼<br>·유역조사<br>지침 | ·수자원 환경<br>기초조사<br>업무기준<br>·수문조사<br>매뉴얼 | ·기상청<br>데이터 종합<br>품질관리<br>지침           | ·측정망<br>운영 계획<br>·수질측정망<br>정도관리<br>지침 | ·하천유량<br>측정지침<br>·ADCP<br>활용지침<br>(유량조사단) |

K-water의 경우 부서별 업무에 따라 국가 규정·지침 등을 고려, 별도 품질관리 지침·매뉴얼 수립·운영하고 있으며 부서별 업무 단위로 구성됨에 따라 데이터 생애주기 전 과정에 일괄 적용 가능한 체계화된 규정·기준이 미비한 실정이다. 물정보 분야와 경영 기술 분야에 있어서 규정 및 기준현황은 아래 표와 같다.

표 20 물정보/경영기술 분야 규정 및 기준 현황

| 구분   | 규정·기준 현황                              |
|------|---------------------------------------|
| 수자원  | 댐관리규정, 실시간수문자료품질관리 기준, 수문조사업무매뉴얼 등 9개 |
| 수도   | 수도시설운영관리규정, 수도설비 사업장 운영관리기준 등 4개      |
| 지하수  | 지하수법, 국가지하수관측망 설치·운영 등 국가법·기준 적용      |
| 경영재무 | 인사, 교육, 법무, 회계 자산, 예산 분야 등 46개        |
| 기술안전 | 건설관리, 안전보건, 건설공사 분야 등 30개             |

특히 데이터의 품질관리 수준에 대한 정의·관리 미흡하며 부서별 품질관리를 수행 중이나, 검·보정된 데이터가 어떤 규정에 의해 어떤 검·보정 절차를 거쳤는지에 대해 누구나 이해할 수 있는 공통기준 없으며 물리적 의미를 고려한 활용·분석 단계의 품질관리에 대한 정확한 개념조차 정립되어 있지 않은 실정 (필요성에 대해서는 대체로 공감)이다. 이를 개선하고자 현재는 부서별 품질관리 기준을 바탕으로 다양한 검·보정 기법을 자체 개발하여 적용 중이며 수자원분야의 품질관리는 통합물관리처 HDIMS, 수도 : 수도기술처 DQMS, 연구원 : K-PICK(수도분야), K-WQMS(수도, 댐분야)로 개발중이거나 실제 적용하여 운영하고 있다. 이를 살펴보면 분야별로 오·결측 탐지·검보정 알고리즘 및 적용 대상·방법이 상이하며 수자원의 경우 모든 데이터의 오·결측을 룰 베이스로 탐지, 동일 기법을 활용하여 보정 (강수량·수위 자동보정 및 수동 검·보정을 위한 알고리즘 지원)하고 있으며, 수도의 경우 본사로 전송되는 데이터만 대상으로 룰 베이스로 오결측을 탐지, 필요시에만 수동 보정 시행 (관리자가 직접 검·보정 하여 값 입력)하고 있다.

물분야에 있어 데이터 관리 현황을 살펴보면 수자원의 경우 비상시 현장 단독 운영이 가능하도록 모든 데이터는 현장 DB(FA)에 저장·관리하되, 품질관리는 본사 DB(FA)에서 시행하고 있으며 현황은 아래 표와 같다.

표 21 수자원 분야 데이터 관리 현황

| 구 분    | 현장 (FA)          | 본사 (FA)    | 본사 (OA)    | 활용 (OA)            |
|--------|------------------|------------|------------|--------------------|
| 시스템/DB | DB               | HDAPS (DB) | HDAPS (DB) | Tech DB            |
| 관리 항목  | 1·10·30분, 1시간 자료 | 좌동         | 좌동         | 10·30분·1시간, 일 및 통계 |
| 검보정대상  | -                | HDIMS(자체)  | HDIMS(외부)  | -                  |
| 관리 부서  | 현장 관리단           | 통합물관리처     | 통합물관리처     | 정보관리처              |

데이터 검·보정은 현장 및 통합물관리처에서 자체 기준에 따라 시행하고 있으며 수문·유역 조사는 공용 DB(Tech 등) 미 저장으로 활용에 제약사항으로 작용되고 있다. 수도의 경우 효율적 유지관리를 위해 모든 데이터는 본부 DB(FA)에 저장·관리하고 품질관리 역시 본부 DB(FA)를 대상으로 시행하고 있으며 현황은 아래 표와 같다

표 22 수도 분야 데이터 관리 현황

| 구 분    | 현장 (FA)     | 본부 (FA)          | 본사 (FA)   | 본사 (OA)  |
|--------|-------------|------------------|-----------|----------|
| 시스템/DB | iWater (DB) | RWIS (DB)        | RWIS (DB) | TRANNDDB |
| 관리 항목  | 1분 (6개월)    | 1·15분, 1시간, 일 자료 | 좌동        | 좌동       |
| 검보정대상  | -           | DQMS             | -         | -        |
| 관리 부서  | 현장 관리단      | 수도기술처            | 수도기술처     | 정보관리처    |

데이터 검보정은 현장에서 자체 기준에 따라 시행하고 있다. 수질 실험 데이터는 항목에 따라 LIMS(실험실관리시스템)또는 기술 DB(TECH, OA)에 직접 입력되어 저장 활용하고 있으며 관측데이터는 분야별 수자원·수도 체계에 따라 저장·관리하고 있다. 지하수의 경우 국가 시스템(GIMS)을 통해 수집(지하수정보센터) 후, 기술 DB(Tech) 저장·활용하고 있으며 외부(지자체) 데이터는 행정망을 통해 수집, 기술 DB(Tech) 저장하고 있다. 또한 센서 시험성적서, 상세 제원, 검보정 결과 등은 미 전산화 (담당부서에서 문서형태로 보관) 상태이며 수질 실험 장비의 시험성적서, 검보정 이력 및 결과 등은 환경부 및 KOLAS 규정에 따라 LIMS에 통합 저장·관리 중이다. K-water의 경우 시스템 업무처리를 위해 44개 정보시스템, 4개 DB 운영 중이며 아래 표와 같이 알 수 있으며 지속적으로 증가 추세에 있어 비슷한 시스템과 알고리즘은 통합하는 것도 검토하는 것이 필요하다.

표 23 정보시스템 현황

| DB    | 합 계    | MIRS<br>(행정용) | TECH<br>(기술용) | INFOS<br>(지방용) | WLIB/기타<br>(설계용) |
|-------|--------|---------------|---------------|----------------|------------------|
| 시스템수  | 44     | 20            | 10            | 12             | 2                |
| 데이터용량 | 4.98TB | 3,000GB       | 1,300GB       | 600GB          | 80GB             |

범용 알고리즘을 타 분야로 확대하기 위해서는 모든 ‘데이터’ 자체를 효과적으로 생산, 관리하고 유기적으로 융복합하여 활용하기 위한 통합 체계가 필요하며 각 부서의 고유 권한을 유지하면서 시너지가 발생할 수 있게 통합되어 관리 개선하는 것이 필요하다고 판단 된다.

시계열 데이터를 분석 하는 방법중 최근 많이 활용되고 있는 인공지능(AI) 기법으로는 딥러닝(Deep Learning)과 선형회귀등 다양한 통계 기법을 이용하는데 여기서 딥러닝이란 데이터 ‘학습(Learning)’을 통해 예측, 연산, 분석 등의 작업을 수행하는 것을 말한다. 이러한 학습방법은 ‘지도학습(Supervised Learning)’과 비지도학습(Unsupervised Learning)’으로 분류된다. 지도학습은 문제와 문제에 대응되는 답을 대량으로 제시한 뒤, 특정 문제에 대한 답을 도출하는 방식이다. 이와 반대로 비지도 학습은 정답이 제시되지 않은 상태에서 문제만을 가지고 학습을 진행하는 것이다. 문제와 답이 정확하게 일대일 대응되지 않는 산업계 분야에서는 비지도 패턴의 분석 상황이 우세하기 때문에 분석 목적과 데이터의 특성에 맞는 학습방법을 택하는 것이 중요하다.

# 제 4 장 K-WQMS 프로그램

## 4.1 K-WQMS 개요

### 4.1.1 K-WQMS 특징

K-WQMS(K-Water data Quilty Management Software)는 관측 데이터베이스(ODM)을 기반으로 구축된 물정보 분석 DB (KDM)가 보유한 시계열 데이터의 품질관리를 목표로 개발되었다. 물정보 분석 DB와 연결되기 위해서는 사용자의 PC에 K-water의 보안(접근 통제, 포트허용) 설정되어야 한다. 물정보 분석 DB는 물리적으로 DAON DB서버에 구축되어 있으며, 정보관리처에서 내부 업무망(OA)의 운영을 목적으로 관리되므로, 외부망이나 FA망에 직접적인 접근은 불가능하다.

프로그램은 선형 연구과제 「통합물정보 품질관리 알고리즘 개발 및 활용 기반 구축」에서 개발된 (수자원)우량, 수위, (수도)우량, 압력 데이터의 검·보정 알고리즘을 내장하고 있다. 기존 'Matlab'언어로 개발되었던 알고리즘은 'R'언어(R-3.4.4)로 재개발 되었으며, 'RDOTNET'기술을 활용한 DLL형태로 배포하였다.

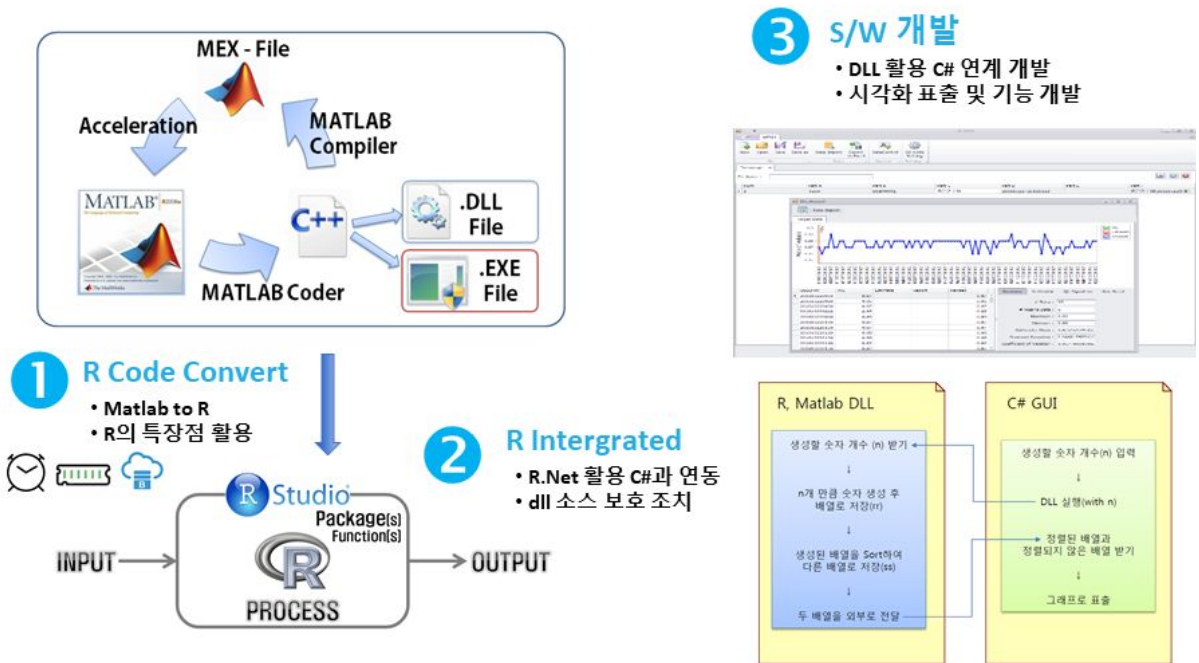


그림 22 SW 구성도

### 4.1.2 K-WQMS 구성

품질관리 시스템 메뉴 구성은 메인항목으로 품질관리, 자동검보정이 있으며 세부 품질관리 항목은 작업목록, 자료조회, 품질수행, 검정결과, 보정결과, 신뢰도, 보정값 업로드, 데이터 출력으로 구성되어 있다. 처음 구동 시에는 작업목록과 자료조회 메뉴만 활성화되고, 검·보정이 완료되면 나머지 메뉴도 활성화 된다.

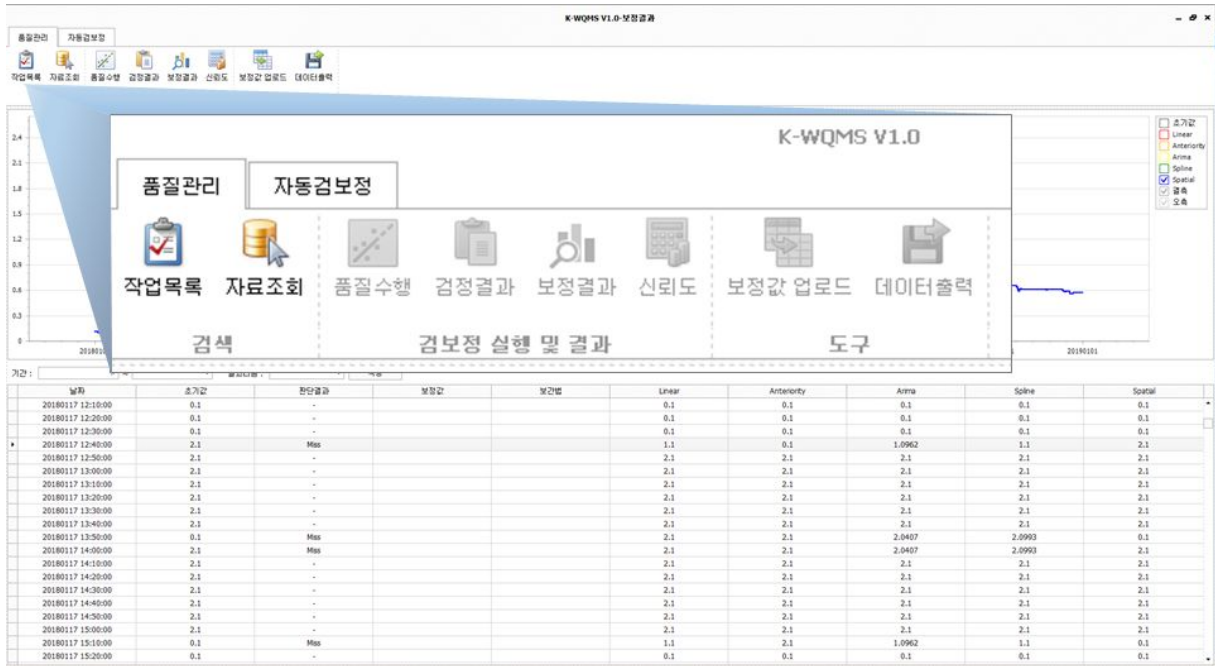


그림 23 메뉴 구성도

- 검색 그룹
  - 작업목록 : 최근 작업한 목록 표시
  - 자료조회 : 물정보 분석 DB(KDM)의 데이터 목록 검색 및 데이터 조회
- 검·보정 실행 및 결과 그룹
  - 품질수행 : 조회된 데이터의 표출과 검·보정 매개변수 설정 및 실행
  - 검정결과 : 조회된 데이터의 검정결과로 차트와 그리드 표출
    - 보정결과 : 보정 알고리즘 별 보정값의 차트 및 그리드 표출
  - 신뢰도 : 보정 알고리즘 별 신뢰도 결과 표출
- 도구 그룹
  - 보정값 업로드 : 선택된 보정값의 레벨을 선택하여 KDM에 업로드
  - 데이터 출력 : 보정결과를 Excel로 출력

## 4.2 K-WQMS 프로그램 기능

### 4.2.1 자료조회 기능

프로그램의 특징으로 서술된 바와 같이 데이터를 조회하기 위해서는 먼저 사용자의 PC에 DAON DB연결을 위한 보안설정(접근통제, 포트허용)이 선행되어야 한다.

자료조회 기능을 클릭하면 하단의 그림과 같이 데이터의 구분, 검색조건을 설정하여 데이터 목록을 검색할 수 있는 기능과 해당 데이터의 조회 기간을 설정하고 데이터를 조회하는 기능을 제공한다.

| RST_ID | 유역명     | 유역코드   | 관측소유형 | 관측소명   | 항목명  | 데이터유형 | 측정주기 | 시간 |
|--------|---------|--------|-------|--------|------|-------|------|----|
| 24637  | 대청댐     | 300804 | 우량관측소 | 대청     | 우량   | 중분    | 10   |    |
| 24638  | 대청댐     | 300804 | 우량관측소 | 대청     | 우량   | 중분    | 15   |    |
| 24639  | 대청댐     | 300804 | 우량관측소 | 대청     | 우량   | 중분    | 30   |    |
| 24640  | 대청댐     | 300804 | 우량관측소 | 대청     | 누가우량 | 누적    | 10   |    |
| 24641  | 대청댐     | 300804 | 우량관측소 | 대청     | 누가우량 | 누적    | 15   |    |
| 24642  | 대청댐     | 300804 | 우량관측소 | 대청     | 누가우량 | 누적    | 30   |    |
| 30631  | 무주남대천하류 | 300303 | 우량관측소 | 무주(대청) | 우량   | 중분    | 10   |    |
| 30632  | 무주남대천하류 | 300303 | 우량관측소 | 무주(대청) | 우량   | 중분    | 15   |    |
| 30633  | 무주남대천하류 | 300303 | 우량관측소 | 무주(대청) | 우량   | 중분    | 30   |    |
| 30634  | 무주남대천하류 | 300303 | 우량관측소 | 무주(대청) | 누가우량 | 누적    | 10   |    |
| 30635  | 무주남대천하류 | 300303 | 우량관측소 | 무주(대청) | 누가우량 | 누적    | 15   |    |
| 30636  | 무주남대천하류 | 300303 | 우량관측소 | 무주(대청) | 누가우량 | 누적    | 30   |    |

그림 24 검색 및 조회 화면 예시

물정보 분석DB는 데이터를 크게 수자원, 수도, 기상으로 구분하고 있으며, 관측 데이터 베이스 표준인 ODM의 구조를 따르기 때문에 유역, 관측소, 항목, 데이터유형, 주기, 단위 등의 데이터를 각각 별개의 Table로 관리하고 있어 데이터와 관련된 모든 정보를 한 번에 조회하기 어려운 단점이 있다. 이를 극복하고 데이터의 주요 정보를 연계하여 표출하기 위해 'VC\_WQMT\_SEL\_CODE'라는 View를 별도로 구성하고 해당 View로부터 자료조

회를 실행한다. 같은 지점의 동일 항목 데이터라도 데이터 유형이나 측정주기가 다를 수 있는데, 해당 View는 그림과 같이 사용자가 원하는 데이터를 상세하게 선택하여 조회할 수 있도록 구성되어 있으며 'VC\_WQMT\_SEL\_CODE'의 구조는 아래와 같다.

```

CREATE OR REPLACE FORCE VIEW kdm.vc_wqmt_sel_code (ctgry,
                                                    iem_nm,
                                                    rm_cntnt,
                                                    ctgry_type,
                                                    rst_id,
                                                    time_intrvl_vu,
                                                    time_unit,
                                                    basin_cd,
                                                    site_id,
                                                    site_nm,
                                                    addr,
                                                    site_lngt,
                                                    site_la,
                                                    altd_vu,
                                                    site_ty_mng_trml_nm,
                                                    data_unit,
                                                    vu_ty_mng_trml_nm,
                                                    basin_dc,
                                                    site_info
                                                    )
AS
SELECT base."CTGRY", base."IEM_NM", base."RM_CTNT", base."CTGRY_TYPE",
       base."RST_ID", base."TIME_INTRVL_VU", base."TIME_UNIT",
       base."BASIN_CD", base."SITE_ID", base."SITE_NM", base."ADDR",
       base."SITE_LNGT", base."SITE_LA", base."ALTD_VU",
       base."SITE_TY_MNG_TRML_NM", base."DATA_UNIT",
       base."VU_TY_MNG_TRML_NM", basin.basin_dc, base.site_info
FROM (SELECT a.gnrl_ctgry_mng_trml_nm AS ctgry, e.iem_nm, a.rm_cntnt,
            a.data_ty_mng_trml_nm AS ctgry_type, a.rst_id,
            a.time_intrvl_vu, f.unit_nm AS time_unit, c.basin_cd,
            a.site_id, c.site_nm, c.site_info, c.addr, c.site_lngt,
            c.site_la, c.altd_vu, c.site_ty_mng_trml_nm,
            g.unit_nm AS data_unit, a.vu_ty_mng_trml_nm
FROM tm_od_rst a,
     tm_od_site c,
     tm_od_iem e,
     tc_od_unit f,
     tc_od_unit g
WHERE a.site_id = c.site_id
     AND a.iem_id = e.iem_id
     AND a.time_unit_id = f.unit_id
     AND a.unit_id = g.unit_id) base
LEFT JOIN
tm_od_basin basin ON base.basin_cd = basin.basin_cd;

```

## 4.2.2 품질관리 기능

검색된 자료와 품질관리를 위한 매개변수를 설정하는 화면으로 데이터가 가지고 있는 기본적인 통계정보를 파악하고 검색된 관측소별 지역적 계절적 특성을 반영한 매개변수 선정을 통한 품질관리 수행은 데이터 품질 수준을 향상 시킬 수 있다. 자료조회 화면에서 임의로 첫 번째 데이터인 '대청' 지역의 '대청' 관측소의 10분 주기 우량 증분 데이터를 조회하면 아래 그림과 같이 데이터가 조회된다.

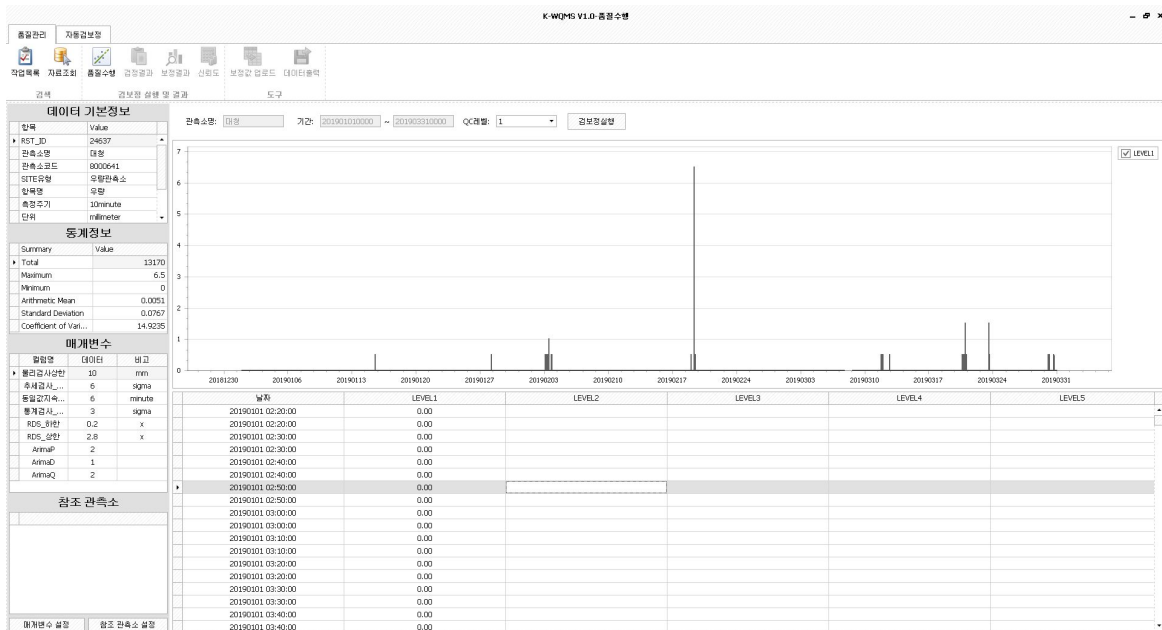


그림 25 품질 수행 기능 예시 화면

품질수행 메뉴에서는 데이터 품질관리를 수행에 참고할 수 있는 데이터 기본정보와 통계정보를 제공한다. 또한 검·보정 알고리즘에 적용될 매개변수의 기본값과 공간 검·보정을 수행할 시에 참조할 관측소를 설정할 수 있는 기능을 제공한다. 화면 좌측 하단의 매개변수 설정 버튼을 클릭하면, 최근 3년 치의 해당 데이터를 미리 분석한 매개변수 정보를 관측소 정보와 매칭시켜서 설정할 수 있다. 또한 공간 검·보정을 위해 해당 관측소와 가장 가까운 관측소 데이터를 참조시킬 수 있다. 위 그림에서 화면 왼쪽 하단의 '참조 관측소 설정'을 누르면 아래와 그림과 같은 참조관측소 설정 화면이 팝업 생성된다.

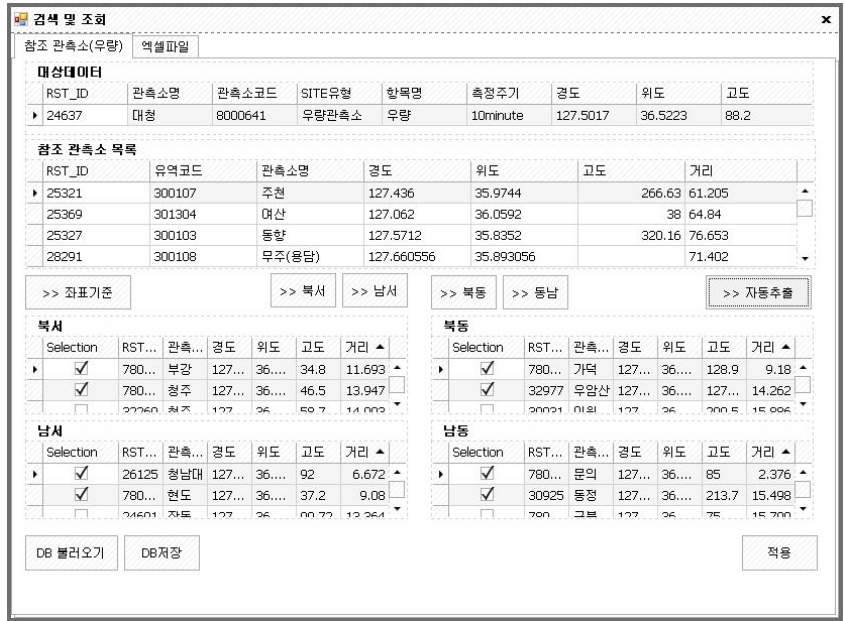


그림 26 관측소 설정 화면

우량 데이터의 경우, 사용자는 인근의 관측소를 원하는 방향에 맞춰 2개씩 참조시킬 수 있으나 '자동추출' 기능을 활용하면 위·경도 값으로 거리를 계산하여 가장 가까운 관측소를 각 방향에 2지점씩 자동으로 참조시켜 준다.

아래 그림을 보면 '매개변수 설정' 버튼을 눌렀을 때 기본값이 추출된 물리검사 상한과 Arima P, D, Q의 값이 바뀐 것과 '참조 관측소 설정'에서 참조 관측소를 선택했을 때의 변화된 화면을 확인할 수 있다.



그림 27 매개변수 설정 기능

상단의 그림과 같이 매개변수와 참조관측소 설정이 완료되면 차트 상단에 있는 '검보정실행' 버튼을 클릭하여 데이터 검정과 보정을 동시에 실행한다. 보정의 경우 검정한 데이터가 입력자료로 참조 되도록 알고리즘이 구성되어 있다. 검·보정 알고리즘 수행이 완료되면 '검정결과', '보정결과', '신뢰도' 메뉴가 활성화 되고 '검정결과' 화면으로 자동 전환 된다.



그림 28 검보정 실행

#### 4.2.3 검정결과 기능

검정결과 화면에서는 초기값과 데이터의 검정결과로 결측, 오측을 차트와 그리드로 표출하여준다. 오측의 경우에는 '물리검사', '추세검사', '지속검사', '공간검사' 중 어떠한 검정항목에서 오측으로 판별되었는지 확인이 가능하다. 또한 전체 데이터 중에서 오·결측된 데이터의 비율도 쉽게 확인할 수 있다.

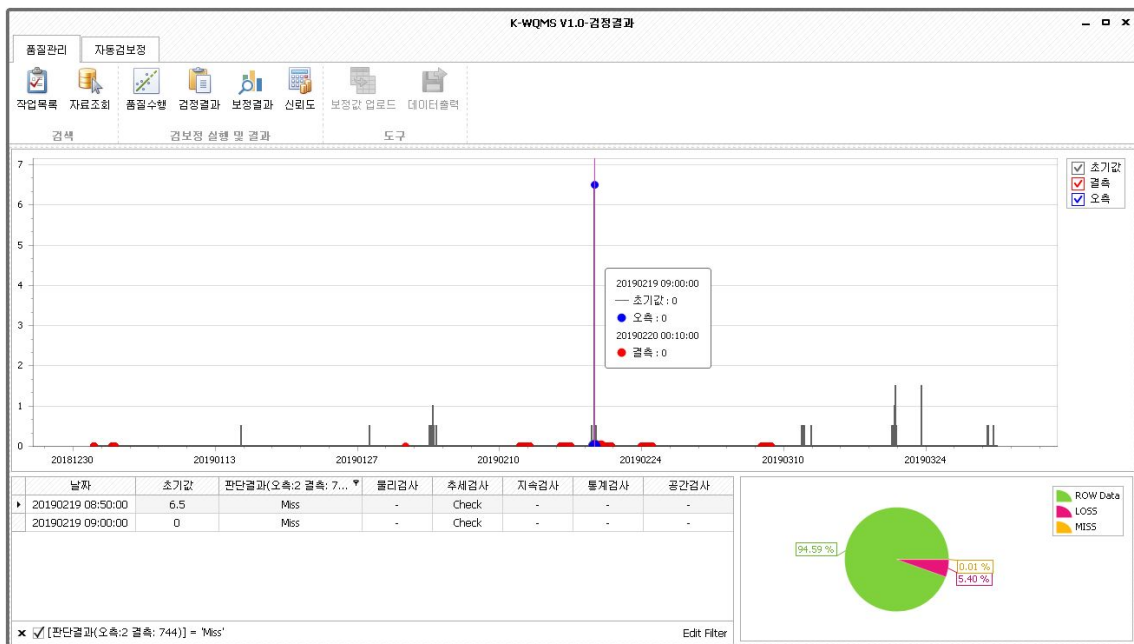


그림 29 검정결과 예시

또한 마우스를 원하는 시간 지점에 올려놓으면, 해당 시간의 초기값과 오·결측 상황을 상세히 나타내 준다.

그리드의 컬럼명 부분에는 필터링 기능이 제공되어 위의 그림과 같이 오측(Miss)된 데이터만 표출할 수 있으며, 오측으로 판별한 검정항목에는 'Check'라는 메시지를 나타내어 오측으로 판별한 사유 확인이 가능하다.

#### 4.2.4 보정결과 기능

보정결과에서는 검정 알고리즘의 판단결과를 기반으로, 오측 및 결측 데이터에 대해서 'Linear', 'Anteriority', 'Ariam', 'Spline', 'Spatial' 등 5가지의 보정 알고리즘을 사용해 보정한 결과값을 표출해준다. 차트에서는 초기값과 5개의 보정값, 오·결측 현황을 표출하고, 상세한 내용을 보기 위해 마우스를 원하는 위치에 올려놓으면 해당 시간의 자세한 정보를 확인 할 수 있다. 하단의 그리드는 차트와 연계하여 시간별 보정 결과값을 나타낸다. 부수적인 기능으로 차트 항목은 체크박스 선택으로 표출 유무를 결정한다.

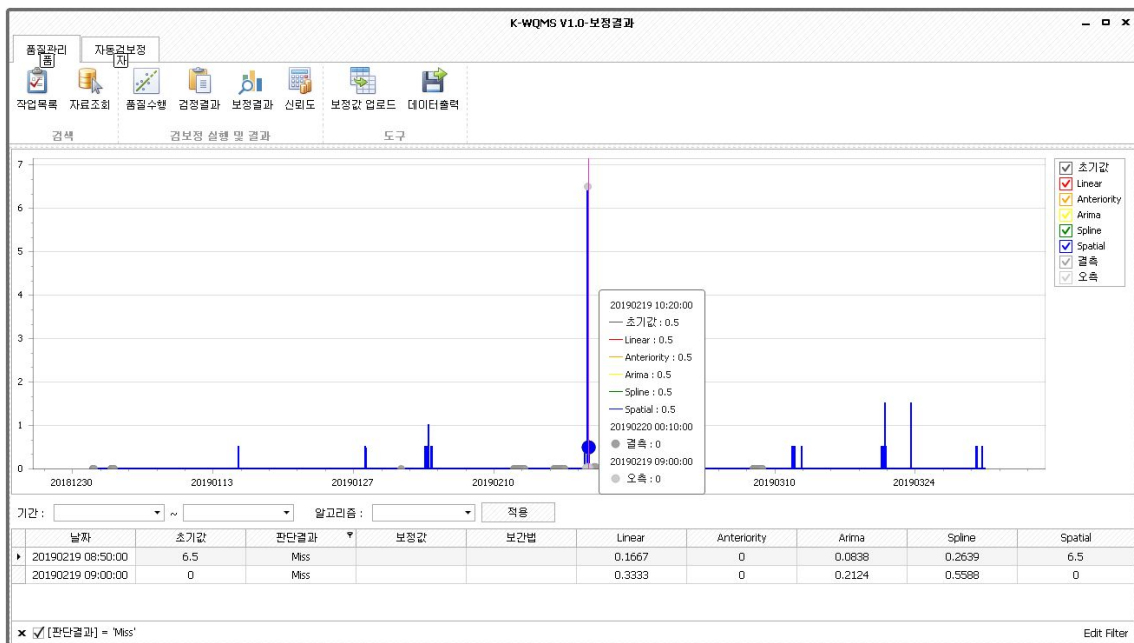


그림 30 보정결과 예시

사용자는 오·결측 데이터를 5가지의 보정값 중 1개의 값을 선택할 수 있다. 그리드에서 원하는 보정알고리즘의 값을 선택하고 더블클릭하면 보정값으로 선택되고, 어떤 보간법의 결과인지도 함께 기재된다. 보정해야 하는 데이터가 많은 경우에는 기간과 알고리즘을 선택하여 일괄 보정도 가능하며, 보정값과 보간법이 일괄로 기재된다.

기간 : [ ] ~ [ ] 알고리즘 : [ ] 적용

| 날짜                | 초기값 | 판단결과 | 보정값    | 보정방법   | Linear | Anteriority | Arma   | Spline | Spatial |
|-------------------|-----|------|--------|--------|--------|-------------|--------|--------|---------|
| 20190219 08:50:00 | 6.5 | Miss | 0.1667 | Linear | 0.1667 | 0           | 0.0838 | 0.2639 | 6.5     |
| 20190219 09:00:00 | 0   | Miss | 0.2124 | Arma   | 0.3333 | 0           | 0.2124 | 0.5588 | 0       |

기간 : 2019-02-19 ~ 2019-12-19 알고리즘 : Spline 적용

| 날짜                | 초기값 | 판단결과 | 보정값    | 보정방법   | Linear | Anteriority | Arma   | Spline | Spatial |
|-------------------|-----|------|--------|--------|--------|-------------|--------|--------|---------|
| 20190219 08:50:00 | 6.5 | Miss | 0.2639 | Spline | 0.1667 | 0           | 0.0838 | 0.2639 | 6.5     |
| 20190219 09:00:00 | 0   | Miss | 0.5588 | Spline | 0.3333 | 0           | 0.2124 | 0.5588 | 0       |

그림 31 보정 결과 선택 예시

#### 4.2.5 신뢰도 평가 기능

신뢰도 평가 기능은 데이터가 5가지의 알고리즘을 통해 보정되었을 때의 얼마만큼의 정확성을 가지고 있는지 제공된 오차(RMSE)와 Nash-Sutcliffe Efficiency (NSE) 계수 등의 통계 기법을 적용한 수치로 나타낸다. 검색 데이터 셋을 기준으로 결측 및 오측을 잭 나이프(Jack-Knife) 방법을 통해 발생시키게 하고 각 알고리즘별로 보정을 한 후, 원본 데이터와 비교하여 그 정확도와 높은 순위를 제시하여 품질 의사결정자의 객관적 판단을 지원해준다. 결측과 오측 그리고 오결측 통합 적용시 결과를 제공하여 각 상황별 판단의 기준을 제공하고 있으며, 신뢰도 분석은 해당 알고리즘을 총 3회에 걸쳐 수행하여 도출된 값의 평균값이 제시되고 있다.

| 통계지표       | 산정방법   | 통계지표                    | 산정방법   |
|------------|--|-------------------------|--|
| CC (상관계수)  | $\frac{\sum_{i=1}^n (O_i - \bar{O})(CM_i - \bar{CM})}{\sqrt{\sum_{i=1}^n (O_i - \bar{O})^2 \sum_{i=1}^n (CM_i - \bar{CM})^2}}$ | MAE (절대오차평균)            | $\frac{\sum_{i=1}^n  O_i - CM_i }{N}$                                    |
| IoA (일치계수) | $1 - \frac{\sum_{i=1}^n (O_i - CM_i)^2}{\sum_{i=1}^n ((O_i - \bar{O}) +  CM_i - \bar{O} )^2}$                                  | RMSE (최소 제공된 오차)        | $\sqrt{\frac{1}{n} \sum_{i=1}^n (O_i - CM_i)^2}$                         |
| Bias (편의)  | $\frac{1}{n} \sum_{i=1}^n (O_i - CM_i)$  | NSE (Nash-Sutcliffe 계수) | $1 - \frac{\sum_{i=1}^n (O_i - CM_i)^2}{\sum_{i=1}^n (O_i - \bar{O})^2}$ |

| 오결측 전체 신뢰도  | RMSE   | NSE      | RANK |
|-------------|--------|----------|------|
| Linear      | 0.0581 | 98.5419  | 3    |
| Anteriority | 0.0874 | 96.0519  | 5    |
| Arma        | 0.0502 | 99.0921  | 2    |
| Spline      | 0.0712 | 97.4183  | 4    |
| Spatial     | 0.0000 | 100.0000 | 1    |

| 결측 적용 신뢰도   | RMSE   | NSE      | RANK |
|-------------|--------|----------|------|
| Linear      | 0.0420 | 99.3969  | 2    |
| Anteriority | 0.0551 | 99.0232  | 5    |
| Arma        | 0.0468 | 99.2567  | 4    |
| Spline      | 0.0436 | 99.3518  | 3    |
| Spatial     | 0.0000 | 100.0000 | 1    |

| 오측 적용 신뢰도   | RMSE   | NSE      | RANK |
|-------------|--------|----------|------|
| Linear      | 0.0665 | 97.6477  | 3    |
| Anteriority | 0.1038 | 92.9448  | 5    |
| Arma        | 0.0510 | 98.9182  | 2    |
| Spline      | 0.0829 | 95.3982  | 4    |
| Spatial     | 0.0000 | 100.0000 | 1    |

그림 32 신뢰도 평가 결과 예시

#### 4.2.6 보정 결과값 업로드

데이터를 보정한 후에는 다시 분석 DB에 업로드 한다. 보정된 데이터는 수정이 된 이후이기 때문에 품질레벨을 선택하여 DB에 저장해야 한다. 보정된 데이터는 보정한 사유(오측, 결측)과 보정값, 보정한 알고리즘을 함께 저장한다.

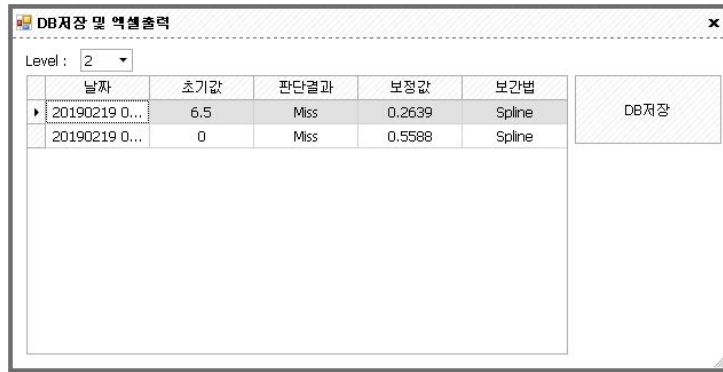


그림 33 보정값 업로드 기능 예시

#### 4.2.7 데이터 출력

데이터의 보정이 끝나면 사용자는 보정한 결과를 다시 엑셀로 저장할 수 있다. 저장되는 데이터는 보정결과 하단의 그리드 데이터가 선택된다.

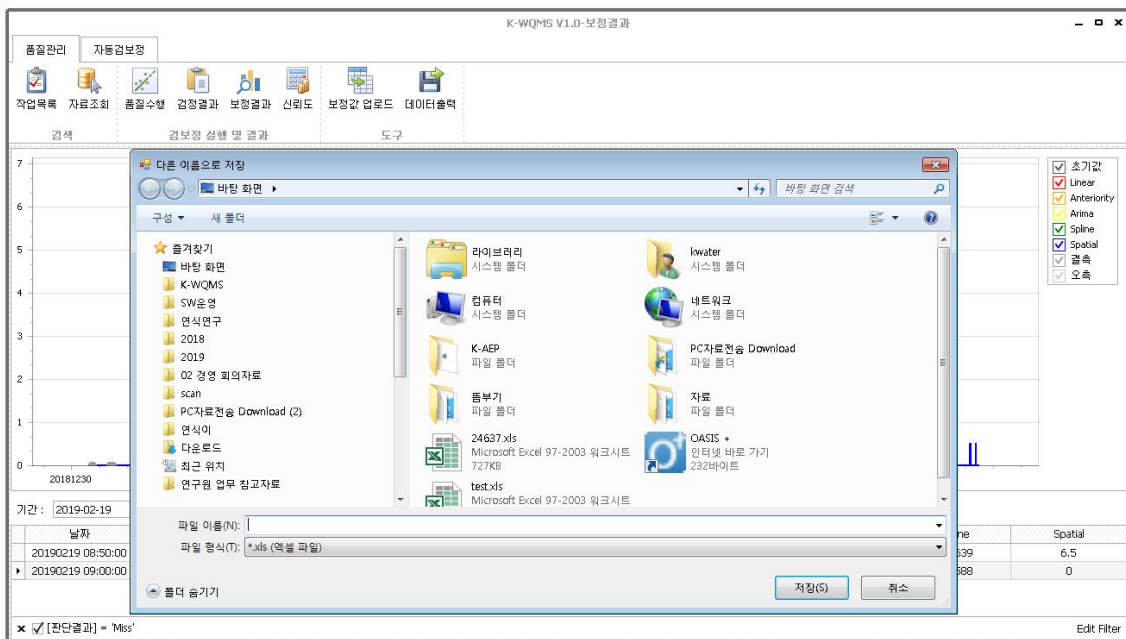


그림 34 데이터 출력 예시

## 4.3 물정보 분석 표준 DB

### 4.3.1 ODM(Observation Data Model)

선진국에서는 최신 IT 및 센서기술의 발달로 취득된 방대한 정보를 실시간 또는 능동적 활용을 위한 관측 데이터베이스를 개발하여 활용 중에 있다. 미국의 대학·정부기관 연합체(CUAHSI)에서 실시간 데이터의 저장을 위한 데이터베이스를 개발하였으며, 웹서비스 및 GIS 기반의 공간정보시스템(ArcHydro)에 연계 활용을 목적으로 수문관측 정보의 저장을 위한 DB(ODM1.1) 및 웹서비스(WaterOneFlow), 공유포맷(WaterML) 등을 제공하고 있다.



CUAHSI는 웹서비스를 통하여 89개 출처, 73만개의 관측소 데이터를 ODM 구조의 DB로 축적, 무료로 제공중이며, 관측 데이터베이스(ODM1.1)는 다양한 지구관측 연구 프로젝트(EarthChem<sup>1</sup>, CZO<sup>2</sup>, IOOS<sup>3</sup>) 등에서 활용되고 있다

'14년 대설, 집중호우 등 총 23회의 자연재해가 발생하였고, '15년 이상고온, 적은 강수량으로 역대 3위의 가뭄을 기록하는 등 급격한 기후변화 현상에 대비하기 위해서는 방대한 데이터를 잘 관리하고 국가 물 관리에 효과적으로 활용하는 것이 중요하다. 또한 신뢰도 높은 물 정보를 생산하고, 분야별 데이터관리에서 통합관리로의 전환을 통한 정확한 의사결정 정보 제공이 필요하다.

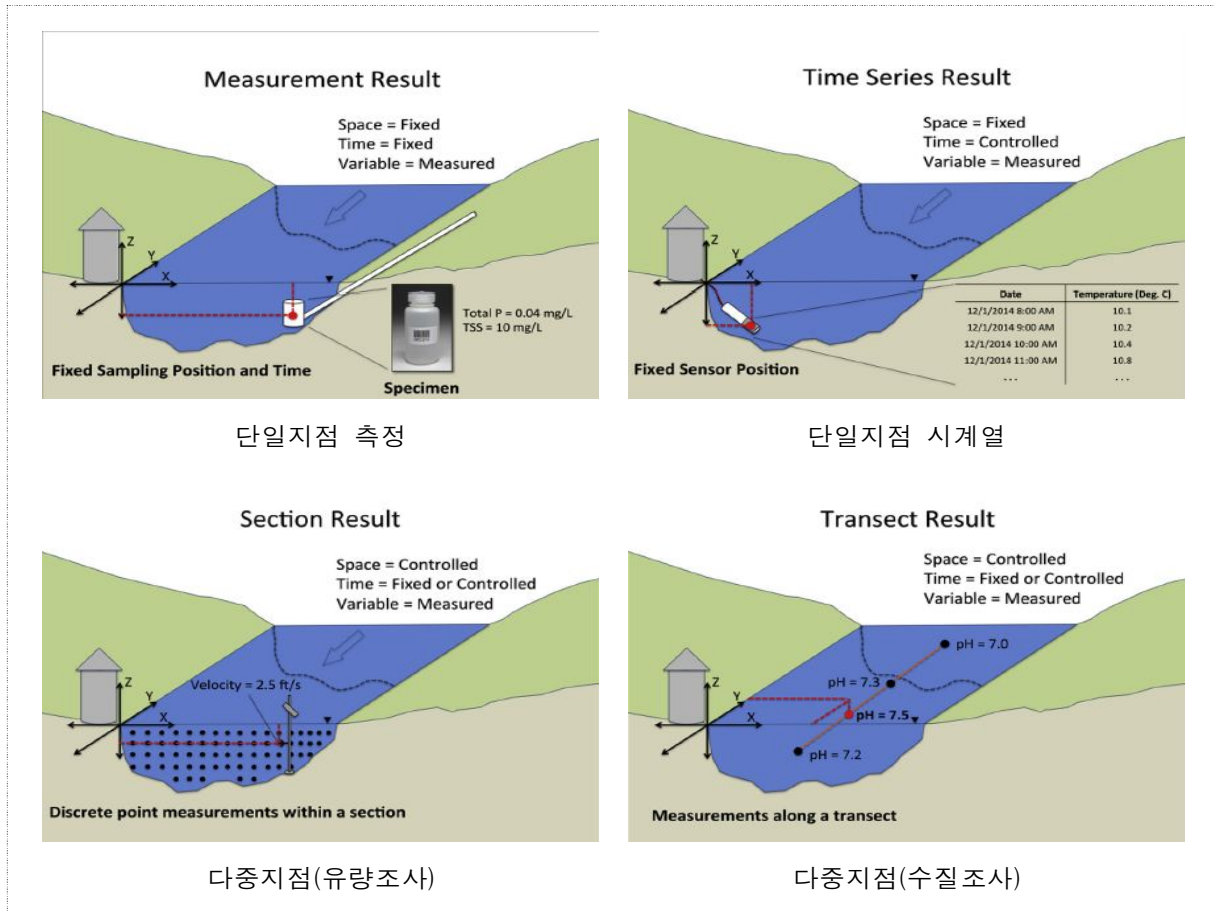
ODM(Observation Data Model)이란 관측 데이터베이스 모델로써 오스틴의 CUAHSI 워크숍(Maidment, 2005)에서 처음 발표, 그 후 1.1버전이 WATERS 네트워크<sup>4</sup>)에서 테스트 및 발표(Horsburgh et al., 2008) 되었다. 다양한 지구관측 프로젝트에서 단일지점의 시계열·측정 데이터에 적합한 데이터베이스 모델인 ODM 1.1에 여러가지 기능을 추가하여 활용하였고, 확장 및 수정된 부분을 정리하여 ODM 2버전이 개발되었다. ODM 1.1 버전은 시계열 데이터 관리에 적합, ODM 2 버전은 다양한 관측데이터의 수용이 가능하다.

1) EarthChem : 지구의 화학적 데이터의 검색·분석을 위한 관리 시스템 구축 프로젝트

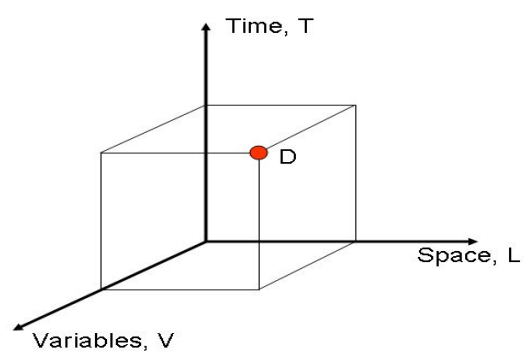
2) CZO(Critical Zone Observatory) : 지구의 화학·물리·지질·생물학적 과정을 이해하기 위한 목적으로 9개 기관의 공동연구 프로젝트

3) IOOS(Integrated Ocean Observing System) : 바다, 해안, 큰 저수지에 대한 데이터의 수집·전송·분석을 제공하기 위한 미국 연방-지역의 파트너쉽

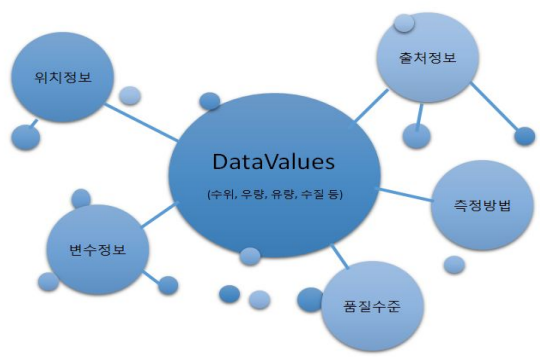
4) WATERS 네트워크 : 개발된 정보시스템을 시험할 수 있도록 구성된 Test-Bed 네트워크



ODM은 기본적으로 관측데이터의 기본속성(시간, 공간, 항목)이 2차 속성(출처, 품질, 측정방법) 정보들과의 유기적 관계로 연결되어 있다. 별도로 저장 관리되던 다양한 항목 별(수위·유량·유량 등), 시간별(1분·시간·일 등), 분야별(수도·수자원 등)로 상이하게 관리되던 데이터를 단일 테이블에 저장으로 접근성 향상되었으며, 별도의 정제과정 없이 단순 검색을 통하여 표준화된 속성정보를 갖는 데이터 확보가 가능하여 신속하고 능동적인 데이터 분석이 가능하다.

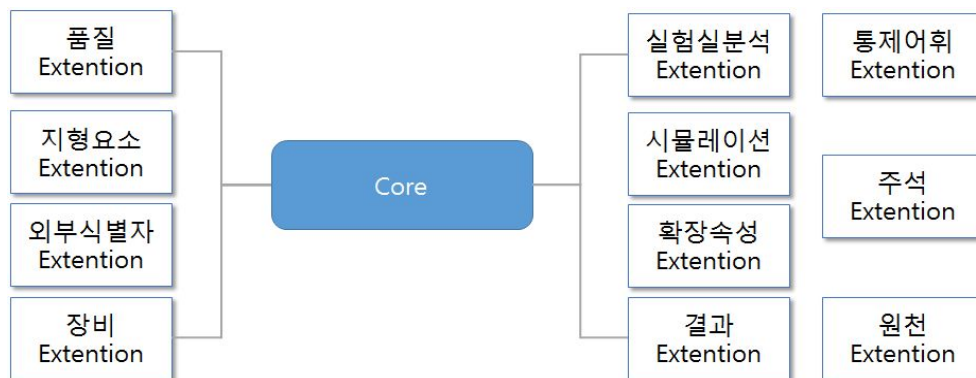


[관측데이터의 기본속성]



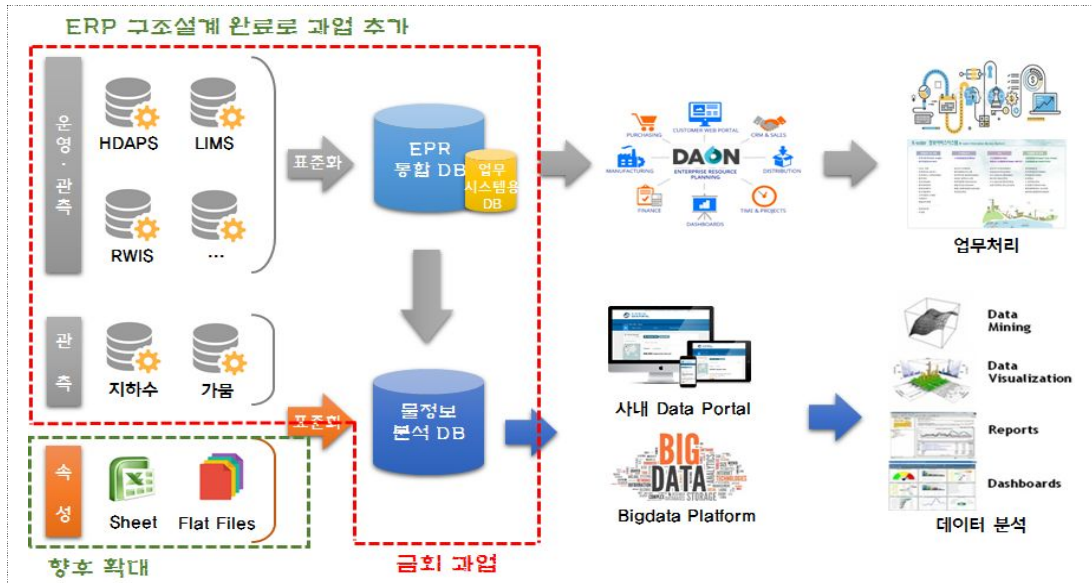
[ODM 구조 개념도]

또한, 연구목적으로는 품질관리가 되지 않은 원시데이터 필요, 운영자는 품질관리된 활용 가능한 가공데이터 필요한데, 가공 전·후 데이터를 동일 테이블에 저장하여 사용자가 원하는 데이터 제공이 가능하다. ODM 2는 ODM 1.1을 보완하여 개발된 관측 데이터베이스 모델이며, 12개의 주제영역과 132개 테이블로 구성되어 있다. 모든 관측에 대한 보편적인 내용을 담고 있는 하나의 'Core'와 특정 목적을 위한 많은 'Extention'을 갖는 모듈형으로 구성된다. Core와 결과, 통제어휘 Extention으로 관측데이터를 수용할 수 있으며 데이터의 특성에 따라 품질, 실험실분석, 장비 등의 Extention 모듈을 선택적으로 사용하는 구조로 사용되지 않는 Extention은 Core와 분리 가능하며, 외부 식별자 및 확장속성 Extention을 두어, 데이터베이스의 확장성이 제공된다.

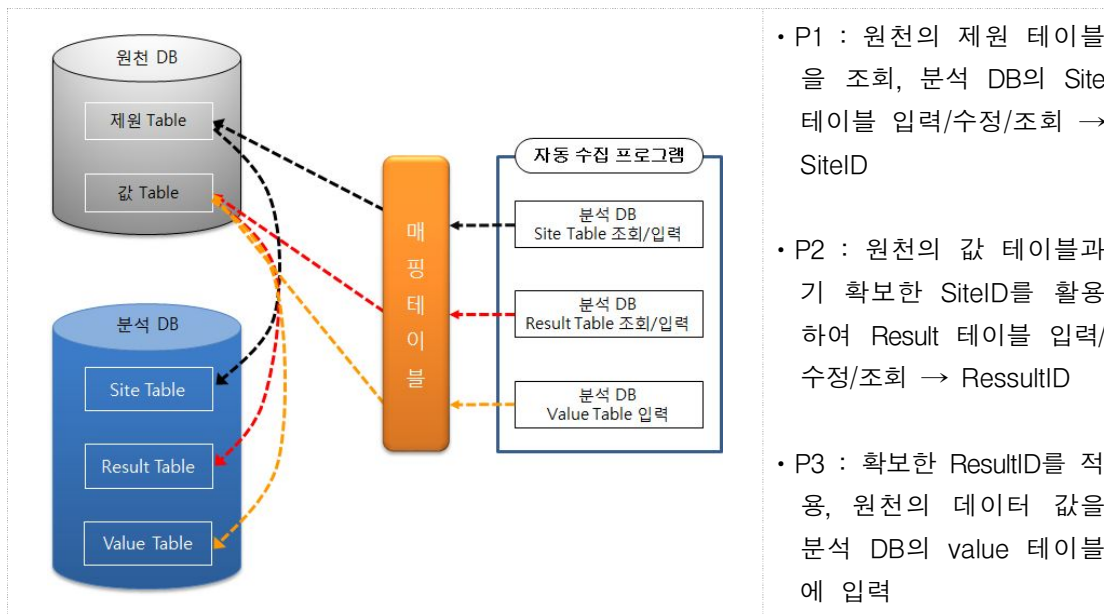


#### 4.3.2 KDM(K-water observation Data Model)

KDM은 ODM을 기반으로 한 K-water형 데이터 모델로써 기존 모델을 보완하여 구축한(표준 적용 등) DB 이다. 원천 DB 현황 및 데이터간의 연관관계를 고려, 일부 테이블이 수정되었으며, 데이터 연관관계를 고려, 품질관리는 Value 테이블로 변경(Result → Value 테이블)하였고, 원천 DB와 분석 DB의 연결고리를 쉽게 찾기 위해 Result 테이블에 태그(코드) 추가되었다.



원천 데이터는 ERP DB에 분산되어 수집되어 있는바 데이터의 자동수집을 위한 DB프로그램이 작성되어 일정한 주기로 저장 되어있는 데이터를 통합한다.



- P1 : 원천의 제원 테이블을 조회, 분석 DB의 Site 테이블 입력/수정/조회 → SiteID
- P2 : 원천의 값 테이블과 기 확보한 SiteID를 활용하여 Result 테이블 입력/수정/조회 → ResultID
- P3 : 확보한 ResultID를 적용, 원천의 데이터 값을 분석 DB의 value 테이블에 입력

## 참 고 문 헌

1. 김지원, “다양한 딥러닝 알고리즘과 활용”, 정보과학회지, 제33권, 제8호, pp.25-31, 2015.
2. 안성만, “딥러닝의 모형과 응용사례”, 지능정보연구, 제22권, 2호, pp.127-142, 2016.
3. 이수지, “순차적 데이터의 회귀분석을 위한 순환신경망 훈련방법 연구”, 석사논문, 2016.
4. 정대훈, “실시간 교통 데이터 활용 프로그램 개발”, 교통개발연구원 연구총서 01-07, 2001
5. 조한일, “수질 데이터 통합관리 기술과 과제”, 계장기술, 2008
6. 김형섭, 김충수 “수문자료 품질관리시스템의 시험유역 적용”, 대한토목학회 , 2005.
7. <http://deeplearning4j.org/kr/>
8. <https://devblogs.nvidia.com/parallelforall/deep-learning-nutshell-sequence-learning/>
9. <http://colah.github.io/posts/2015-08-Understanding-LSTMs/>
10. Heeseok Joeng, “Comparative study on dust pollution on air quality in Korea and China through query search”, sungkyunkwan University, 2015
11. Ramanarayanan, T. S., J. R. Williams, W. A. Dugas, L. M. Hauck, and A. M. S. McFarland. 1997. Using APEX to identify alternative practices for animal waste management. ASAE Paper No.972209. St. Joseph, Mich.: ASAE.
12. Wei Dai, "Improving Data Quality Through Deep Learning and Statistical Models" Advances in intelligent systems and computing 2017

# 실용화 추진 계획서

|             |                  |   |                |              |       |          |      |
|-------------|------------------|---|----------------|--------------|-------|----------|------|
| 과제번호        | G19019           | 구 분   | 프로그램개발, 운영방법개선 |              |       | 계획<br>년도 | 2021 |
| 연구수행        | 연구과제명            | 물정보 표준 DB와 연계한 품질관리 적용기술 개발   |                |              |       |          |      |
|             | 연구수행자            | 김종립, 정지영, 최연식, 김대욱, 김용섭   |                |              |       |          |      |
|             | 연구기간             | 2019 4. 1 ~ 2019 9. 31  |                |              |       |          |      |
|             | 예산현황             | 100% 집행   | 집행액: 천원        | 예산액: 6,000천원 |       |          |      |
|             | 연구결과             | 데이터 오결측 검출 프로그램 및 알고리즘 고도화  |                |              |       |          |      |
| 실용화<br>추진계획 | 제 목              | 품질관리 통합관리 시스템 구축  |                |              |       |          |      |
|             | 수 행 자            | 김종립, 정지영, 최연식, 김용섭  |                |              |       |          |      |
|             | 적용대상<br>부서 및 담당자 | 물종합정보처 구자호 차장   |                |              |       |          |      |
|             | 내 용              | <ul style="list-style-type: none"> <li>- 목 적 : K-water 물정보 표준과 연계된 생애주기별 데이터 품질 통합관리 프로그램 개발</li> <li>- 적용장소 : 물종합정보처</li> <li>- 기대효과 : 전사 품질관리 지침과 연계된 데이터 공인 절차 도구로써 활용, 품질수준 확보를 위한 기술적 기반 제공</li> <li>- 실용화 방안 확산계획 : 시범구축 및 운영안정화 → 향후 확대시행</li> </ul> |                |              |       |          |      |
|             | 추진계획             | 실용화 기간 : 2021. 1 ~ 2021. 10   |                |              |       |          |      |
|             |                  | 1/4분기   | 2/4분기          | 3/4분기        | 4/4분기 | 계        |      |
|             |                  | 30%   | 30%            | 40%          | %     | 100%     |      |
| 소요예산        | 없음               |   |                |              |       |          |      |

## 참 여 연 구 원 현 황

| 연구원   | 참 여 내 역  | 비 고   |
|-------|--|---|
| 김 종 립 | <ul style="list-style-type: none"> <li>○ 연구업무총괄 (연구책임자)</li> <li>○ 참여분야 : 연구과제 총괄</li> <li>○ 참여내역<br/>제1장 서 론<br/>제2장 통합물정보 품질관리 알고리즘 분석<br/>제3장 통합물정보 품질관리 알고리즘 개선</li> </ul> | <ul style="list-style-type: none"> <li>- '19.1 ~ '19.12</li> <li>- K-water 연구원<br/>연구지원처</li> <li>- 042-870-7682</li> </ul> |
| 정 지 영 | <ul style="list-style-type: none"> <li>○ 참여분야 : 알고리즘 고도화</li> <li>○ 참여내역<br/>제2장 통합물정보 품질관리 알고리즘 분석<br/>제3장 통합물정보 품질관리 알고리즘 개선</li> </ul>                                      | <ul style="list-style-type: none"> <li>- '19.1 ~ '19.12</li> <li>- K-water 연구원<br/>연구지원처</li> <li>- 042-870-7334</li> </ul> |
| 최 연 식 | <ul style="list-style-type: none"> <li>○ 참여분야 : 시스템 설계 및 개발</li> <li>○ 참여내역<br/>제4장 K-WQMS 프로그램</li> </ul>   | <ul style="list-style-type: none"> <li>- '19.1 ~ '19.12</li> <li>- K-water 연구원<br/>연구지원처</li> <li>- 042-870-7338</li> </ul> |
| 김 대 욱 | <ul style="list-style-type: none"> <li>○ 참여분야 : 적용성 평가 및 개선 도출</li> <li>○ 참여내역<br/>제2장 통합물정보 품질관리 알고리즘 분석</li> </ul>   | <ul style="list-style-type: none"> <li>- '19.1 ~ '19.12</li> <li>- K-water 연구원<br/>연구지원처</li> <li>- 042-870-7685</li> </ul> |
| 김 용 섭 | <ul style="list-style-type: none"> <li>○ 참여분야 : 시스템 설계 및 개발</li> <li>○ 참여내역<br/>제4장 K-WQMS 프로그램</li> </ul>   | <ul style="list-style-type: none"> <li>- '19.1 ~ '19.12</li> <li>- K-water 연구원<br/>연구지원처</li> <li>- 042-870-7337</li> </ul> |

# 서 지 자 료

|   |                |   |         |
|---|----------------|---|---------|
| (1) 출판물 고유번호<br>2019-TE-RR-410-1984   | (2) 사업분류       | (3) 발행일                                   |         |
| (4) 제목/부제<br>물정보 표준 DB와 연계한 품질관리 알고리즘 적용기술 개발   |                | (5) 연구기간<br>2019.1.1. - 2019.12.31        |         |
| (6) 연구수행기관<br>한국수자원공사, K-water연구원 연구지원처   |                | (7) 연구 수행자(소속)<br>김종립, 정지영, 최연식, 김대욱, 김용섭 |         |
| (8) 수행기관 주소<br>대전광역시 유성구 전민동 462-1  |                | (9) 연구의뢰기관 및 주소                           |         |
| (10) 공동 수행기간  |                | (11) 계약 또는 인가번호                           |         |
| <p>(12) 초록</p> <p>4차 산업혁명이 기반이 되는 표준 데이터에 대한 품질관리를 통해 신뢰성 높은 데이터 제공으로 新부가가치 창출 기반 확보 및 대국민 신뢰도 제고가 필요한 시점이다. 이에 따라 표준화 DB를 구축하고 이와 연계된 계측된 수자원, 수도분야 뿐만아니라 지하수 등 품질관리 알고리즘의 개발 및 고도화를 통한 고품질 데이터를 활용한 기능정보기술 및 의사결정 시스템 구축을 수행함</p> |                |   |         |
| (13) 키-워드<br>데이터, 표준 DB, 품질관리, 검보정, 알고리즘, S/W   |                |   |         |
| (14) 사업고유번호   |                |   |         |
| (15) 공개여부<br>아니요  | (16) 총면수<br>52 | (17) 발행부수<br>-                            | (18) 가격 |

# BIBLIOGRAPHIC DATA SHEET

|  |                                   |   |                   |
|--|-----------------------------------|---|-------------------|
| <b>(1) Report ID</b><br>2019-TE-RR-410-1984  | <b>(2) Project Classification</b> | <b>(3) Report Date</b><br>Dec. 31, 2019                                     |                   |
| <b>(4) Title/Subtitle</b><br>Development of quality control algorithm for standard database of water information   |                                   | <b>(5) Research Period</b><br>Jan. 1, 2019 ~ Dec. 31, 2019                  |                   |
| <b>(6) Performing Organization</b><br>R&D Promotion & Support Dept KWCI, K-water   |                                   | <b>(7) Author(s)</b><br>J.R. Kim. J.Y. Jung, Y.S. Choi. D.W. King, Y.S. Kim |                   |
| <b>(8) Performing Organization Address</b>   |                                   | <b>(9) Sponsoring Agency and address</b>                                    |                   |
| <b>(10) Co-performing Organization(s)</b>  |                                   | <b>(11) Contract or Authorization</b>                                       |                   |
| <b>(12) Abstracts</b><br><br>It is time to create new values and increase public confidence by providing reliable information through data quality management, which is the foundation of the Fourth Industrial Revolution.<br><br>Therefor, it establishes a standardized DB and secures high quality data and establishes a decision-making system by developing and upgrading quality control algorithm such as measured water resources, water and groundwater |                                   |   |                   |
| <b>(13) Keyword</b><br>Data, K-water Standard DB, Quality Control, Algorithm, Software   |                                   |   |                   |
| <b>(14) Project ID</b>   |                                   |   |                   |
| <b>(15) Security</b><br>NO   | <b>(16) No. of Pages</b><br>52    | <b>(17) Circulation</b><br>XXX  | <b>(18) Price</b> |

### 주의사항

본 보고서의 저작권은 한국수자원공사에 있으므로, 본 자료의 내용을 허가없이 인용, 복사, 전재, 재배포를 할 경우 법적 제재를 받을 수 있습니다.

Copyright © 2019  
by  
Korea Water Resources Corporation

발행인 : 한국수자원공사 사장 이 학 수

발행일 : 2019. 12.

발행처 : 한국수자원공사 K-water연구원

대전직할시 유성구 왕골고갯길 49

Tel. (042) 870-7682

한국수자원공사는 친환경 기업윤리 이념을 구현하기 위하여 2019. 12. 31 『환경경영』을 선포하고, 모든 보고서 발간에 『재생용지 및 무코팅 방식』을 사용하고 있습니다.