

연구보고서

GIS 기반 지반조건 예측의 지역 PoC 분석

남양주 왕숙 지역의 실제 시추자료와
심도별 앙상블 모델에 관한 후향적 연구

2026. 10.

SubSight-AI

기존 자료 기반 지층조건 예측의 기술 타당성을 검토한 소규모 분석보고서
분석 자료: 조사 등록 기록 70건 · 층서 구간 383개 · 공-심도 평가 258개

목차

초록 및 연구 목적	3
제1장 연구 지역과 자료	4
제2장 데이터 명세	5-6
제3장 모델 및 검증 방법	7-8
제4장 분석 결과	9-12
제5장 논의 및 결론	13
참고문헌 및 재현 자료	14

연구 범위

본 보고서는 제안서의 전체 기능 중 기존 자료 통합과 지층조건 예측의 선행 기술검토에 해당한다. 제안서가 다루는 조사 최적설계·지하수·환경·경제성·플랫폼 전체의 실증과 구분하여 아래 범위로 결과를 제시한다.

표 1. 제안 내용과 이번 분석의 대응

제안 구성	이번 분석	검토 수준
좌표·심도 기준 자료 통합	70개 기록·383개 층서 및 수위 연결	자료 구축 확인
심도별 지층조건 분포 추정	5m의 공간정보 활용 효과	탐색적 기술 PoC 확인
다층 융합의 추가 예측 효과	지형 RF 비교; 수위는 관측 표시	전체 추가 효과 미입증
암선·지하수의 정밀 확률예측	이번 지층 범주 분석과 구분	별도 검증 필요
조사 최적화·경제성·종료규칙	정보가치·비용 절감 실험 미수행	이번 연구 범위 밖
3D·위성·BIM/GIS 플랫폼	구현·통합시험·사용성 검증 미포함	후속 개발 범위

평가 결과를 읽는 기준

Brier 감소는 지층 분포 예측의 평균 제곱오차 감소이다. 이를 정확도·시추공 절감률·시간 절감률로 해석하지 않는다. 5가지 분할은 동일 지역 기록에 대한 민감도 검토이다. 근거의 적용 범위는 왕숙의 분석 자료와 평가한 심도이며, 독립 현장 일반화는 후속 과제로 남는다.

초록

기존 지반조사 자료를 GIS에서 결합하여 미조사 위치의 지층조건을 사전에 파악할 수 있는지 검토하였다. 남양주 왕숙 지역의 3개 블록의 4개 자료 출처에서 확보한 조사 등록 기록 70건과 층서 구간 383개를 이용하고, 지표 아래 2·5·8·10m의 원관측 범위에 해당하는 258개 공-심도 사례를 평가하였다. 지역 지층분포, 주변 시추 기반 직접 공간확률, 좌표 및 지형 랜덤포레스트, 심도별 적응 혼합 등 6개 모델을 비교하였다. 출처별 분리와 4가지 공간 분할에 모두 25m 학습-검증 거리 완충을 적용하였다. 직접 공간확률의 5m Brier는 지역분포 대비 18.07~27.99% 감소하였다. 학습자료 내부에서 공간 비중을 정하는 보수적 혼합은 전체 심도 통합 Brier를 0.95~3.32% 줄였다. 다만 직접 공간확률의 출처별 5m log loss는 악화되었고, 보수적 혼합의 소수 지층 hard 재현율은 개선되지 않았다. 이 사례는 특정 심도의 공간정보 활용과 깊이별 혼합 설계에 대한 기술적 PoC 근거를 제공한다. 동일 지역 자료의 후향적 탐색으로서 독립 현장 일반화, 다층 자료 융합의 추가 성능 및 조사시간 절감까지 입증한 것은 아니다.

주제어 GIS, 시추자료, 지층조건 예측, 공간 검증, 심도별 양상불, PoC

연구 배경과 목적

본 연구의 출발점은 기존 시추·지질·지형 자료를 위치 기준으로 연결하여, 지반조사 전에 참고할 조건과 자료 공백을 제시할 수 있다는 기획 가설이다. 예측 대상은 주어진 위치와 심도의 지층 범주 분포이며, 다음 시추 위치를 정답 좌표처럼 맞히는 문제가 아니다. 위치별 분포를 실제 주변 층서와 함께 조회할 수 있다면 후속 검토를 지원할 가능성이 있으나, 그 가능성과 실제 업무시간 절감은 별도로 검증해야 한다.

연구질문은 ① 주변 시추의 공간정보가 위치를 사용하지 않는 지역분포보다 지층 확률오차를 줄이는가, ② 그 효과가 깊이에 따라 달라질 때 적응 혼합이 전체 오차를 완화하는가, ③ DSM 파생변수가 좌표 모델에 추가 이점을 주는가이다. 앞선 탐색에서 5m의 개선을 관찰한 자료를 다시 분석한 것이므로, 이번 결과를 독립적인 확증 실험으로 해석하지 않는다.

제1장 연구 지역과 자료

왕숙 A-1·A-2·A-24 관련 자료 중 공간적으로 밀집한 70개 등록 기록을 분석하였다. 출처별 기록 수는 F3228 22건, F3205 16건, F3677 13건, F3234 19건이다. 관측 평면좌표의 범위는 약 284×582m이다. E/N 원문 좌표로 거리를 계산하고 GIS 표시는 명목 EPSG:5186에서 EPSG:4326으로 변환한 좌표를 사용하였다. 원출처별 공식 좌표기준의 독립 확인은 후속 품질 검수로 남는다.

표 2. 분석 자료와 사용 범위

자료	수량·범위	분석상 역할
조사 기록	70건	3개 블록의 4개 자료 출처; 독립 물리공 수는 별도 미확정
층서 구간	383개	원문 경계 깊이와 층명 보존
지하수	70건	GL 아래 2.01~4.32m; 측정일 미확인
지질도	1:50,000 의정부 도폭	모든 관측 지점이 Qa; 이번 지역 내 변별력 없음
지형	Copernicus GLO-30 DSM	고도·경사·국지 기복·주변 대비 고도 4변수
평가 심도	GL-2·5·8·10m	70·70·70·48개; 총258개 공-심도 관측

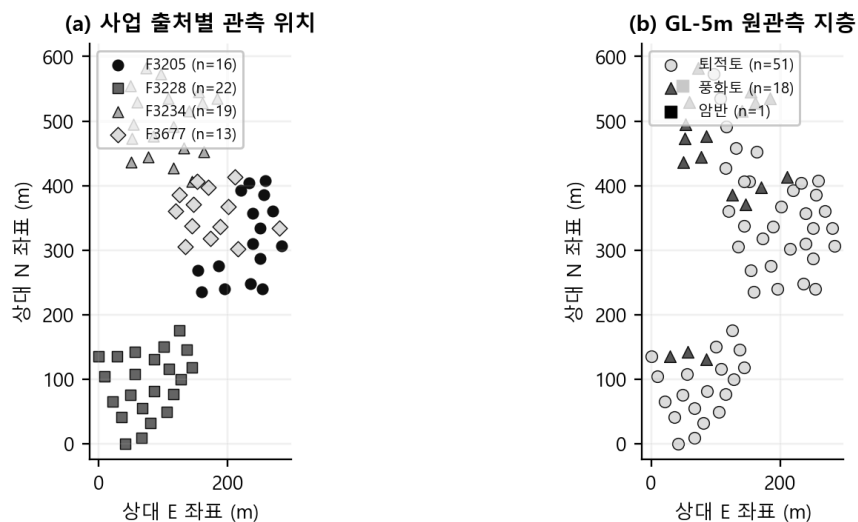


그림 1. 원문 평면좌표에 따른 관측 분포와 5m 지층. 좌표 원점은 E=214,835.284m, N=565,385.733m이다. (b)는 원관측이며 모델 예측 지도가 아니다.

자료는 3개 블록에 대응하는 4개 포털 출처이다. F3228은 A-2BL(22건), F3205는 A-1BL(16건), F3677은 A-1 추가 자료(13건), F3234는 A-24BL(19건)이다. 특히 A-1의 두 자료 묶음을 2개 독립 사업으로 간주하지 않는다. 출처별 분리는 포털 자료 묶음 단위의 분리이다.

지질도는 의정부 1:50,000 수치지질도[5], 지형은 Copernicus GLO-30 DSM[6]을 사용하였다. DSM은 지표면 높이를 나타내므로 관측공의 GL 또는 원문 지표고와 직접 같다고 두지 않았다.

제2장 데이터 명세

2.1 식별 체계와 필드 정의

등록기록을 공통 연결 단위로 하고 원문 층서·수위, 지형 파생변수, 모델 출력, 숨긴 관측 평가를 별도 계층으로 관리하였다. 공변만으로 연결하지 않고 출처 ID를 결합한 등록키를 사용하였다. 아래는 주요 필드이며 전체 경로·자료형·단위·결측 규칙·계보는 부속 데이터 명세에 기록하였다.

표 3. 주요 데이터 필드와 해석

필드	단위·형식	자료 역할	처리 및 해석
registration_key / id	문자열	70건 공통 연결키	등록기록 단위; 물리공 독립성은 별도 확인
project / hole	출처 ID / 공번	4개 자료 묶음	A-1 두 출처를 독립 사업으로 간주하지 않음
E / N	m	원문 평면좌표	명목 EPSG:5186; 원사업 공식 기준 추가 확인
longitude / latitude	degree	GIS 표시용 변환값	EPSG:4326; 위치측량 정밀도 보증 아님
top_m / bottom_m	GL 아래 m	383개 층서의 경계	[top,bottom)로 라벨 판정; 범위 밖 결측
layer_name / stage_index	문자열 / 0~4	원층명 / 파생 범주	풍화토와 잔적토는 stage 2로 집계
raw thickness	m 또는 공란	원문 두께 별도 보존	9공란·2층돌; 경계 라벨과 분리
groundwater depth	GL 아래 m	70건 과거 관측	2.01~4.32m; 양수는 지표 아래 깊이
water measurement date	날짜 / null	확정된 값 0건	완료일 70건을 측정일로 대체하지 않음
ground EL	m	원문 지표고	공통 수직기준 미확정; 모델 입력 제외
SPT/N	시험 수치	구조화 시험행 미확보	현장 시험 미실시를 뜻하지 않음
terrain[0:4]	m / ° / m / m	DSM 파생 4변수	고도·경사·기복·주변 대비 고도
observed_stage	0~4	258개 공·심도 정답	원관측 범위에서만 생성
probabilities[0:5]	합이 1인 비율	모델별 5클래스 분포	보정된 신뢰도와 구분
support / nearest distance	bool / m	GIS 셀의 자료 지원	표시 보류 기준; 정확도 등급 아님

2.2 관측값과 파생값의 구분

원문 좌표·층명·경계·수위는 관측 계층에, 범주 라벨·DSM 파생값·거리·자료 밀도는 파생 계층에, 클래스 분포·λ·지원 여부는 모델 계층에 보존하였다. 종료 깊이, 해당 공의 지층경계·수위·지표고, 원관측 정답을 숨긴 위치의 입력으로 사용하지 않았다.

부속 파일: 데이터명세.csv, 데이터명세.json, 자료출처_목록.csv. 각 필드의 미확보 상태는 null 또는 명시적 상태값으로 유지하며 0으로 대체하지 않는다.

2.3 지층 범주와 평가 관측 범위

표 4. 심도별 실제 클래스 구성

심도	n	매립토	퇴적토	풍화토	풍화암	암반
2.0m	70	3	67	0	0	0
5.0m	70	0	51	18	0	1
8.0m	70	0	0	20	11	39
10.0m	48	0	0	7	4	37

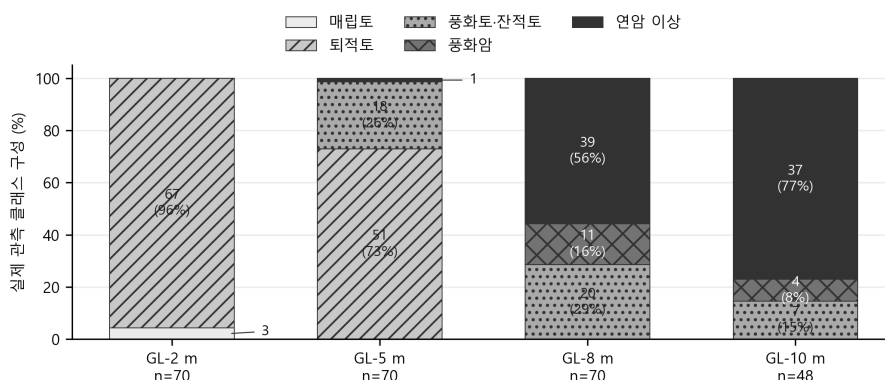


그림 2. 심도별 원관측 지층의 구성. 10m 라벨이 지원되는 48개 기록만 집계하였다. 종로 깊이 10.0m인 18건과 10m 미만인 4건은 반개구간 규칙상 지원 범위 밖이며 임의 지층으로 보충하지 않았다.

원문 층명은 매립토·퇴적토·풍화토·풍화암·암반의 5범주로 정리하였다. stage 2는 풍화토와 잔적토를 포함하고, 암반 범주(stage 4)는 연암·보통암을 포함하며 풍화암과 분리한다. 잔적토 1구간은 이번 평가 심도에 걸리지 않아 5m 풍화토 18건의 집계에는 영향을 주지 않는다. GL 깊이의 반개구간 [상부, 하부]에 속할 때만 라벨을 부여하였다.

2.4 결측·오류·시점의 처리

표 5. 자료 품질 상태와 처리

항목	확인 상태	처리
원문 층두께	9개 공란·2개 총돌 값	원문 필드 보존; 라벨은 GL 경계 사용
두께 총돌 위치	F3228:H-10의 2개 층서값	음수 두께와 경계차 불일치를 별도 기록
지하수 시점	수위 70건·측정일 확정 0건	보고서 완료일을 측정일로 대체하지 않음
지표고 기준	공통 수직기준 독립 확인 미완료	GL-DSM과 임의 결합하지 않음
SPT/N	보관 추출자료에 구조화 수치 미확보	시험 미실시로 해석하지 않음
공의 독립성	좌표 5m 이내 중복 0쌍	좌표 분리만으로 물리공 독립성 인증하지 않음

383개 층서의 원문 대응과 70개 수위 연결을 별도로 대조하였다. 자료 밀도와 가까운 관측 거리는 자료 분포를 설명하는 값이며, 측정 품질이나 모델의 확률 보정을 대신하지 않는다.

제3장 모델 및 검증 방법

3.1 비교 기준과 예측 모델

대조군은 각 학습 집합의 심도별 클래스 빈도이다. 따라서 무작위 추측보다 강한 기준이며, 자료를 이미 확보한 상태에서 개별 위치정보가 추가로 주는 이득을 평가한다. 실제 실무자의 조사계획과 비교한 연구는 아니다. 직접 공간확률은 해당 심도 라벨이 관측된 학습공 가운데 가까운 8개를 거리제곱 역수로 가중한다.

$$p_{\text{mix}}(c \mid x, z) = (1 - \lambda_z) p_{\text{prior}}(c \mid z) + \lambda_z p_{\text{spatial}}(c \mid x, z) \quad (1)$$

표 6. 고정된 비교 모델

모델	주요 입력·계산	비교 목적
지역 지층분포	학습공의 심도별 클래스 빈도	위치를 사용하지 않는 대조군
직접 공간확률	같은 심도 주변 8공; 거리제곱 역수 가중	조건 분포의 공간 변화 탐색
좌표 RF	E/N·심도; 200개 트리	비선형 공간·심도 관계 비교
좌표·지형 RF	좌표 RF + DSM 파생변수 4개	지형 추가정보의 효과 비교
최소오차 혼합	지역분포 + 공간확률; 내부 Brier 최소	덜 보수적인 혼합 대안
보수적 혼합(주)	내부 최소오차 + 1표준오차 이내 최소 공간비중	심도별 공간 활용의 안정화

RF는 랜덤포레스트 [3]다. 트리 200개, 최소 잎 크기 4, max_features=1.0, 시드 20261005로 고정했다. 학습공별 심도 관측의 총 가중치를 1로 맞췄다. DSM 변수는 중심 고도, 3×3 경사, 11×11 기복 및 주변 평균 대비 고도이다.

3.2 학습자료 내부의 혼합 선택

λ 후보는 {0, 0.25, 0.5, 0.75, 1}, 공간 방식 후보는 8공 거리제곱 가중, 16공 거리 가중, 100m Gaussian으로 고정하였다. 각 외부 학습 집합 안에서 자료 출처별 내부 검증을 수행했다. 최소오차 혼합은 내부 평균 Brier가 가장 작은 후보를, 보수적 혼합은 그 최소값과 출처 fold 표준오차의 합 이내인 후보 중 가장 작은 λ 를 선택한다. 내부 유효 출처 fold가 2개 미만이면 지역분포로 돌아간다. 숨긴 공의 라벨은 후보 선택에 입력하지 않았다.

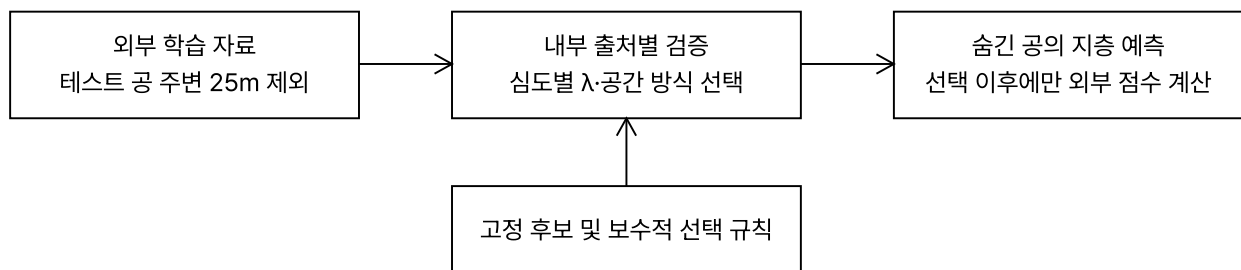


그림 3. 학습자료 내부의 모델 선택과 외부 평가를 분리한 절차. 외부 정답을 λ 선택에 사용하지 않았다.

3.3 외부 분할과 성능 지표

출처별 검증은 4개 출처를 하나씩 제외한다. 보조 검증은 N, E, E+N, E-N의 순서로 관측을 정렬한 뒤 각 방향에서 4개 비슷한 표본 수의 구역으로 나눈다. 모든 내부·외부 검증에서 테스트 공과 25m 이내의 학습공을 제외하고 한 공의 모든 심도는 함께 분리한다. 이는 공간·그룹 의존성을 고려한 검증 원칙[2]을 적용한 설계다. 5가지 분할은 같은 표본의 민감도 검토이며 독립 반복실험이 아니다.

$$\text{Brier} = (1/n) \sum_{i=1 \dots n} \sum_{c=1 \dots 5} [p_{ic} - 1(y_i=c)]^2 \quad (2)$$

주 지표는 5클래스 제공오차 합의 평균인 Brier[1]이며 범위는 0~2, 작을수록 좋다. 감소율은 $100 \times (\text{대조 Brier} - \text{모델 Brier}) / \text{대조 Brier}$ 이다. 보조 지표로 log loss, 정확도, 실제 존재 클래스만 평균한 balanced accuracy를 사용한다. log loss의 0확률은 계산할 때만 10^{-12} 로 절삭한다. 원 확률을 보정하거나 바꾸지는 않는다.

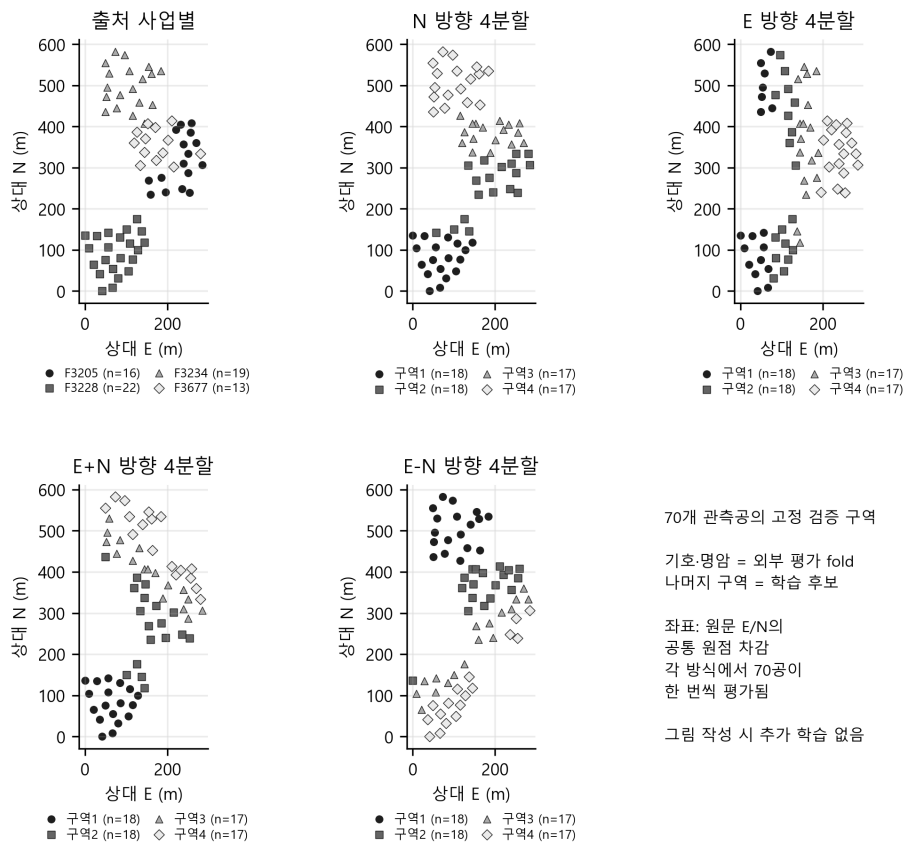


그림 4. 동일한 70개 관측에 적용한 출처별·남북·동서·두 대각선 분할. 공간 구역은 동일 면적이 아니라 관측 수가 비슷하도록 구성하였다. 그림은 분할 소속을 표시하며 25m 제외 후 학습 점수와 구분한다.

출처별 4fold의 테스트 기록은 16·22·19·13건이고 평가 심도 사례는 59·83·69·47건이다. 5방식×4fold는 20개 외부 fold를 구성한다. 같은 위치의 여러 심도를 독립 관측으로 분할하거나 학습자료와 테스트 자료 양쪽에 배치하지 않았다.

제4장 분석 결과

4.1 전체 심도 통합 성능

동일한 258개 공-심도 관측에서 보수적 혼합의 Brier는 모든 분할의 지역분포보다 낮았다. 출처별 검증은 0.395240에서 0.387494로 1.96%, 남북 검증은 0.392605에서 0.387429로 1.32% 줄었다. 나머지 방향에서는 0.95~3.32% 감소했다. 직접 공간확률을 전 심도에 동일하게 적용하면 전체 오차는 오히려 커졌다.

표 7. 전체 심도 Brier 비교

모델	출처별	남북	동서	대각45°	대각135°
지역 지층분포	0.3952	0.3926	0.3781	0.3716	0.3985
직접 공간확률	0.4126	0.3971	0.4512	0.4025	0.4032
좌표 RF	0.4360	0.4045	0.4302	0.4303	0.3949
좌표·지형 RF	0.4349	0.4181	0.4149	0.4138	0.4168
최소오차 혼합	0.3909	0.3860	0.3640	0.3484	0.4031
보수적 혼합(주)	0.3875	0.3874	0.3745	0.3592	0.3947

분할별 $n=258$. 같은 분할의 대조군과 비교한다. 방향에 따라 학습 집합이 달라 대조군 점수도 달라진다. 표의 전체 점수는 공-심도 관측을 합산한 평균이다.

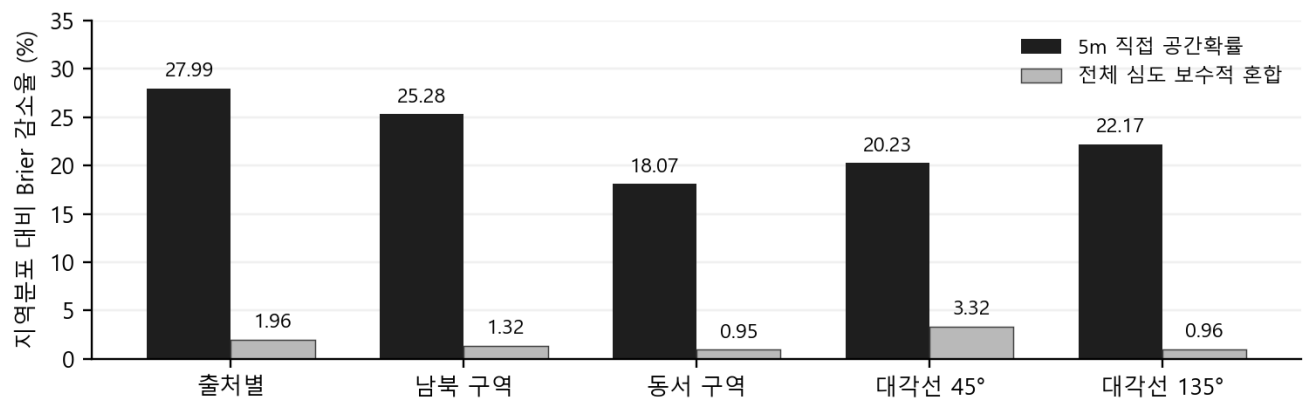


그림 5. 지역분포 대비 Brier 감소율. 5m 직접 공간확률과 전체 심도 보수적 혼합은 서로 다른 모델·평가 범위이다. 5m의 사업 자료 출처별 log loss 악화와 분류 성능의 한계는 뒤의 보조 지표에서 함께 제시한다.

주모델의 전체 정확도와 균형 정확도는 두 주요 분할에서 대조군과 같았다. 따라서 주모델의 개선은 확률오차의 소폭 완화로 해석한다. 동일 자료에서 최저 외부 점수 모델만 다시 골라 채택하지 않았다.

4.2 5m 직접 공간확률의 개선

5m에서는 직접 공간확률이 5가지 분할 모두 지역분포보다 낮은 Brier를 보였다. 출처별 27.99%, 남북 25.28%, 동서 18.07%, 대각선 45° 20.23%, 대각선 135° 22.17% 감소했다. 이는 이 자료에서 5m 조건의 공간 변화가 예측에 참고할 정보를 제공한다는 근거다. 아래 결과는 보조 공간모델이며 전체 심도 주모델의 효과와 구별한다.

표 8. 5m의 지역분포와 직접 공간확률

분할	지역분포	직접 공간	감소율	보수적 혼합
출처별	0.4874	0.3510	27.99%	0.4588
남북 구역	0.4930	0.3684	25.28%	0.4739
동서 구역	0.4440	0.3638	18.07%	0.4308
대각선 45°	0.4203	0.3353	20.23%	0.3749
대각선 135°	0.4910	0.3821	22.17%	0.4769

표 9. 대조군 대비 심도별 Brier 감소율

심도	n	공간 출처별	공간 남북	혼합 출처별	혼합 남북
2.0m	70	-10.25%	+4.63%	+0.00%	+0.00%
5.0m	70	+27.99%	+25.28%	+5.86%	+3.87%
8.0m	70	-27.27%	-18.65%	+0.00%	+0.00%
10.0m	48	-9.10%	-12.48%	+0.00%	+0.00%

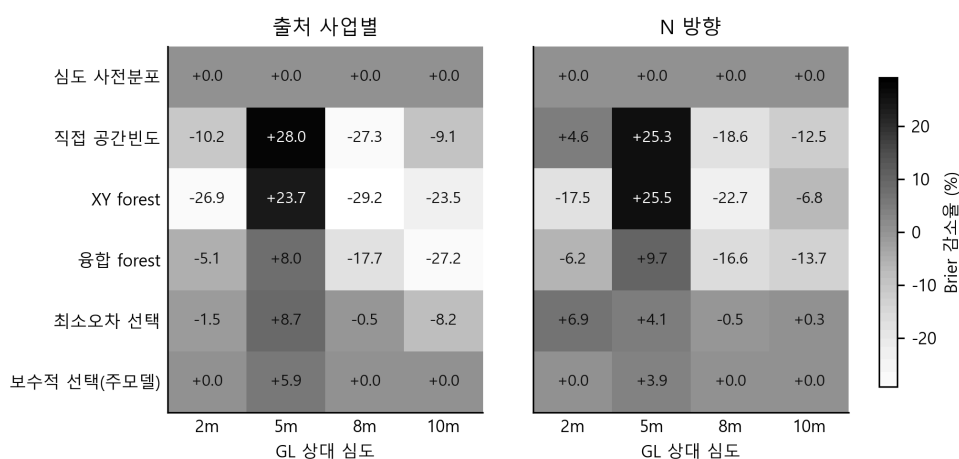


그림 6. 대조군 대비 모델별·심도별 Brier 감소율. 양수는 개선, 음수는 악화이며 모든 모델과 심도를 보존했다. 5m 공간 활용의 효과를 전체 심도의 우수성으로 확대하지 않는다.

직접 공간확률은 8·10m에서 두 주요 분할 모두 악화되었다. 보수적 혼합은 외부 평가의 2·8·10m를 지역분포와 같게 유지했으며, 전체 개선은 5m에서 나온 것이다. 전체자료 내부 검증의 배포 λ 는 5m에서 0.25, 나머지에서 0이었다. 외부 평가에는 각 학습 집합에서 다시 선택한 비중을 사용했다.

4.3 소수 지층과 분류 성능

표 10. 5m의 보조 지표와 소수 지층 재현

분할	모델	log loss	정확도	균형 정확도	소수 검출
출처별	지역분포	1.0925	72.86%	33.33%	0/19
출처별	직접 공간확률	2.2851	77.14%	47.28%	10/19
출처별	보수적 혼합	1.0549	72.86%	33.33%	0/19
남북 구역	지역분포	1.1041	72.86%	33.33%	0/19
남북 구역	직접 공간확률	0.8999	71.43%	35.08%	2/19
남북 구역	보수적 혼합	1.0777	72.86%	33.33%	0/19

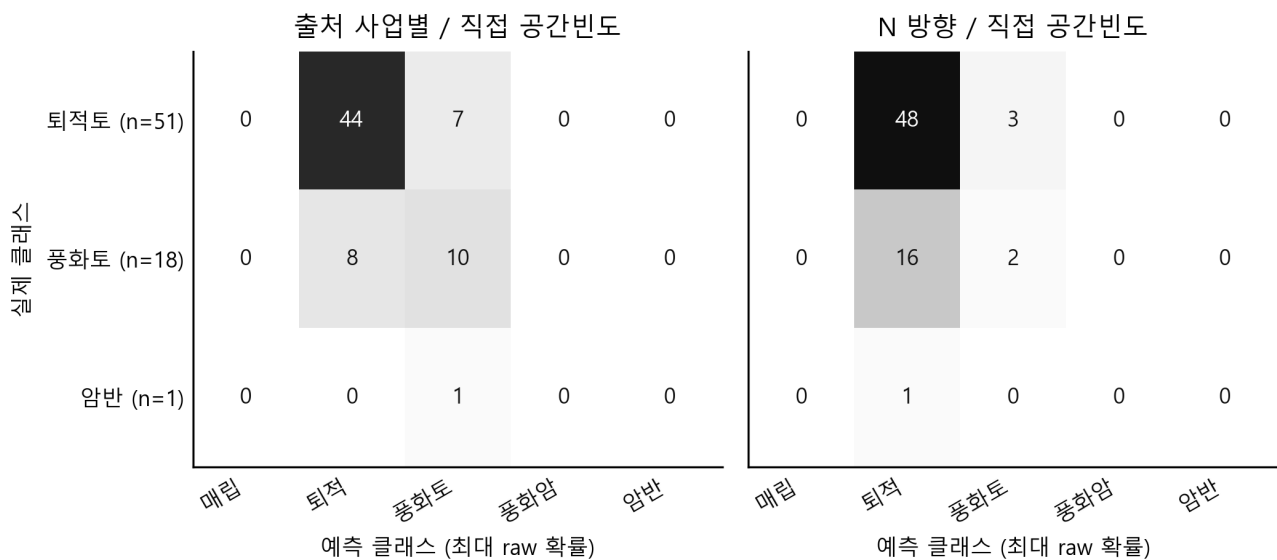


그림 7. 5m 직접 공간확률의 혼동행렬. 행은 실제 지층, 열은 argmax 예측 지층이며 수치는 관측 수이다. 소수 지층의 재현은 출처 분리와 남북 구역 분리에서 크게 달라졌다.

5m 실제 클래스는 퇴적토 51건, 풍화토 18건, 암반 1건이다. 소수 검출은 퇴적토 외 19건 중 예측 클래스가 실제 클래스와 같은 수로, 직접 공간확률은 출처별 10/19, 남북 2/19였다. 풍화토만의 분포 18과 혼용하지 않는다. 보수적 혼합의 소수 검출은 두 분할 모두 0/19였다.

직접 공간확률의 출처별 정확도는 72.86%에서 77.14%로 높아졌지만, 남북 정확도는 72.86%에서 71.43%로 낮아졌다. 보수적 혼합은 정확도와 균형 정확도가 지역분포와 같다. Brier 개선이 드문 지층 탐지나 일관된 위치 순위의 유용성을 자동으로 보장하지는 않는다.

4.4 확률 보정과 표본 불확실성

5m 직접 공간확률의 출처별 log loss는 1.0925에서 2.2851로 악화되었다. 일부 실제 클래스에 0확률을 부여한 과신 문제가 Brier 개선과 함께 존재한다. 원 확률의 합이 1이고 평균 Brier가 낮더라도, 보정된 신뢰도나 실제 발생확률로 인증된 것은 아니다.

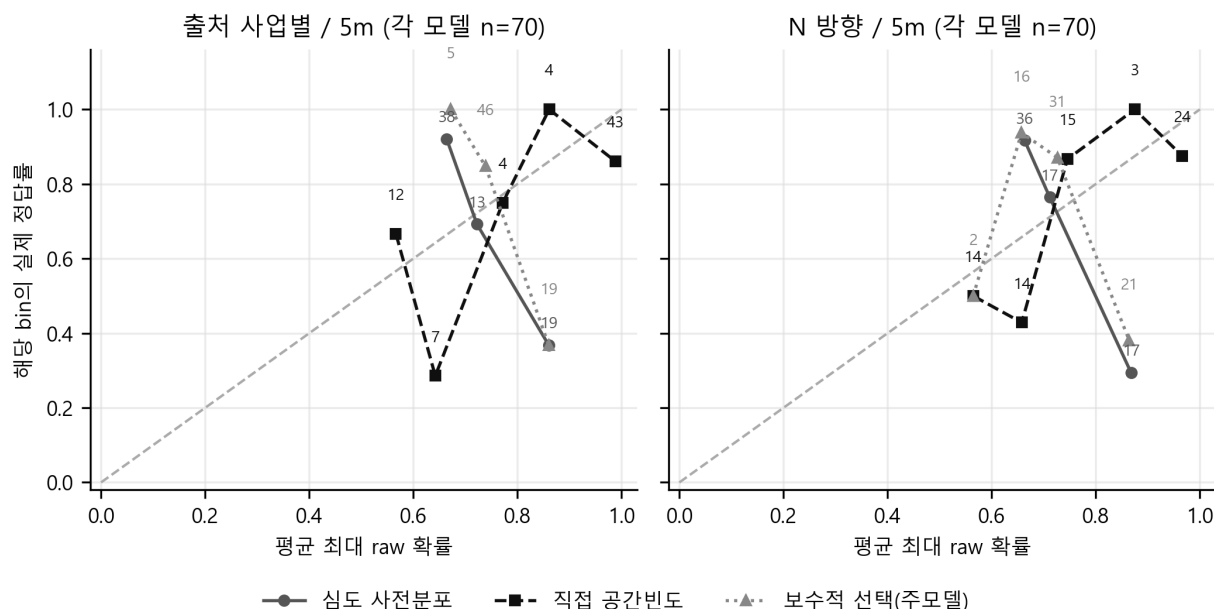


그림 8. 5m top-label reliability 점검. 각 구간에서 모델이 가장 높게 부여한 확률과 실제 argmax 정확도를 비교한다. 표시 n은 구간의 표본 수이며 대각선에 가까울수록 이 진단에서 일치한다. 소표본 후향 점검으로서 보정 완료를 뜻하지 않는다.

표 11. 주모델 전체 Brier 절대 감소의 기술적 bootstrap 구간

분할	재표집 단위	절대 감소	95% 구간
출처별	70개 관측그룹	0.007746	[0.002815, 0.011946]
출처별	4개 출처 자료 묶음	0.007746	[0.000000, 0.014074]
남북 구역	70개 관측그룹	0.005176	[-0.001141, 0.010766]
남북 구역	4개 출처 자료 묶음	0.005176	[0.000734, 0.011754]

같은 단위의 모든 심도를 묶어 고정된 후향 예측에서 5,000회 재표집한 기술적 구간이다. 모델 재학습과 전체 선택 절차를 반복한 구간이 아니다. 남북 관측그룹 재표집 구간은 0을 포함한다. 4개 출처 자료 묶음의 작은 수와 재사용 표본 때문에 독립 일반화 성능의 유의성 확증으로 해석하지 않는다.

그림의 reliability 진단과 log loss의 실제 클래스 확률은 서로 다른 측면을 평가한다. 최상위 확률의 평균 일치만으로 모든 지층 클래스의 확률이 보정됐다고 해석해서는 안 된다.

제5장 논의 및 결론

이 사례에서 확인된 기여는 공간정보를 무조건 많이 결합하는 것보다 공간적 변별력이 나타나는 심도를 찾아 활용하고, 불리한 심도는 지역분포로 축소하는 설계가 필요하다는 점이다. 5m Brier 개선의 방향별 일관성은 기존 시추자료 기반 조건 지도와 후속 예측 모델 개발을 추진할 근거가 된다. 다만 이것은 지층 분포 추정의 탐색적 기술 PoC다. 독립 사업에서 사용할 수 있는 보정된 예측이나 현장 의사결정 도구의 완성을 뜻하지 않는다.

좌표·지형 RF는 전체 Brier에서 모든 분할의 지역분포보다 불리했다. 지질이 모두 Qa인 작은 구역이라는 제한도 있어, 지질·지형·지하수의 다층 융합이 추가 성능을 높인다고 결론내릴 수 없다. 지하수 70건, 실제 층서 및 관측 밀도는 분석·조회 근거로 연결했지만 지하수와 지표고는 숨긴 위치의 예측 입력으로 쓰지 않았다. 현재 자료의 조회 가능성과 새 변수의 예측 기여는 구분해야 한다.

또한 등록 기록이 독립 물리공임을 완전히 인증하지 않았고, 원문 전사 및 좌표·수직기준에 남은 불확실성이 있다. DSM은 건물·수목 등을 포함하는 지표면 모델이며[6] 현재 측량 지반고와 같지 않다. 이전 분석에서 노출된 지역·심도를 재사용했으므로 5m의 긍정적 결과에는 탐색적 선택의 한계가 남는다. 25m 완충은 고정된 검증 설정이며 지층의 상관거리를 추정하여 독립성을 보증한 것은 아니다. 20m 표시 격자는 20m 지반 정확도를 보증하지 않는다.

결론적으로, 왕숙 지역의 실데이터에서 특정 심도 지층조건 추정에 공간정보를 활용할 기술적 가능성과 심도별 양상불의 후속 개발 타당성을 확인했다. 다음 검증은 새로운 사업의 숨긴 지층, 확률 보정 및 소수 지층 탐지, 실제 검토자의 시간·누락·출처 오류를 평가해야 한다. Three.js·위성·GIS 플랫폼의 구현·통합시험·시연은 이 연구의 결과 범위에 포함하지 않는다.

표 12. 연구질문에 대한 판단

연구질문	현재 판단	적용 범위
공간정보가 조건 추정에 도움이 되는가	5m에서 Brier 개선 확인	왕숙 자료의 후향적 민감도 검토
깊이별 혼합이 전체 오차를 완화하는가	Brier 소폭 감소; 분류 성능 동일	평가한 2·5·8·10m 통합
지형 융합에 추가 성능이 있는가	이번 실험에서 우월성 미입증	현재 DSM 및 RF 구성
조사시간·비용을 절감하는가	이번 연구에서 측정하지 않음	현장·사용자 검증 필요

본 연구의 결론은 조건 지도와 지층 확률모델을 후속 개발할 제한된 기술 근거의 확보이다. 새로운 자료, 보정된 확률, 설계조건과 전문가 판단을 결합하는 단계는 별도의 실증을 거쳐야 한다.

참고문헌 및 재현 자료

1. 재현성과 자료 관리

입력은 dataset.json(70개 기록), 평가 명세·분할·예측은 output/gis-fusion-v2/model에 보존되어 있다. map_fusion_v2.py → analyze_fusion_v2.py → audit_fusion_v2.py 순서로 재현한다. 별도 검토는 7,740개 모델-분할 예측과 주모델 예측 1,290개를 별도 코드로 재계산하여 일치를 확인했다. 보고서 부속 전체_성능표.csv는 6모델×5분할×5평가범위의 150행을 제공한다. 생성에 사용한 파일 해시는 report-lineage.json에 기록했다. 원문 접근·이용조건은 원기관을 따른다.

2. 참고문헌 및 자료 출처

1. Brier, G. W. (1950). Verification of forecasts expressed in terms of probability. *Monthly Weather Review*, 78(1), 1-3. [American Meteorological Society 원문 서지](#).

2. Roberts, D. R., et al. (2017). Cross-validation strategies for data with temporal, spatial, hierarchical, or phylogenetic structure. *Ecography*, 40(8), 913-929. [doi:10.1111/ecog.02881](#).

3. Breiman, L. (2001). Random Forests. UC Berkeley 원저자 공개 원고. [원고 PDF](#).

4. 국토지반정보포털. 왕숙 관련 4개 사업의 시추 주상도 및 보고서 표제. [geoinfo.or.kr](#). 원문 연결과 확보 계보는 부속 자료 참조.

5. 고희재, 송교영. (2015). 수치지질도 5만축척 의정부. 한국지질자원연구원. [doi:10.22747/data.20231113.5242](#).

6. Copernicus Data Space Ecosystem. Copernicus DEM GLO-30 자료 설명. [공식 DSM 설명](#). 자료 설명 확인일: 2026-10-05.

7. 김대운. SubSight-AI 제안서 원본. 내부 제안자료. 본 연구는 기존 자료 통합과 지층조건 예측의 일부 구성요소를 검토하며 제안서 전체의 실증 결과로 제시하지 않는다.

3. 주요 파일과 자료 기준

표 13. 부속 자료와 재현상 역할

파일·폴더	내용
전체_성능표.csv	6모델×5분할×5범위의150행 전체 지표
data-specification/	필드 경로·형식·단위·결측·입력 역할 및 출처
report-lineage.json	이 보고서 작성에 사용한 자료의 SHA-256
output/gis-fusion-v2/model/	평가명세·fold·7,740개 후향 예측·모델 선택
output/gis-fusion-v2/mapping/	70개 관측·383개 층서 연결의 품질 검수
scripts/write_wangsuk_poc_paper.py	고정 분석 산출물로 표·그림·본문 생성

원문 데이터는 확보한 시점의 보관 자료를 기준으로 한다. 공식 웹페이지 설명 확인일은 2026-10-05이며 이 날짜를 원관측 측정일로 사용하지 않는다. 자료계보 검수는 파일 동일성을 확인한 것으로 원자료의 측량·시험 품질을 인증하는 절차가 아니다.